

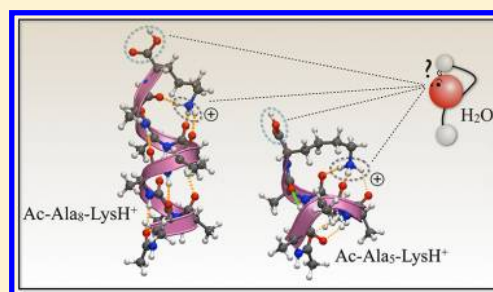
# Water Adsorption at Two Unsolvated Peptides with a Protonated Lysine Residue: From Self-Solvation to Solvation

Sucismita Chutia, Mariana Rossi, and Volker Blum\*

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Theory Department, D-14195 Berlin, Germany

## Supporting Information

**ABSTRACT:** We study the initial steps of the interaction of water molecules with two unsolvated peptides: Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>. Each peptide has two primary candidate sites for water adsorption near the C-terminus: a protonated carboxyl group and the protonated ammonium group of LysH<sup>+</sup>, which is fully hydrogen-bonded (self-solvated) in the absence of water. Earlier experimental studies have shown that H<sub>2</sub>O adsorbs readily at Ac-Ala<sub>5</sub>-LysH<sup>+</sup> (a non-helical peptide) but with a much lower propensity at Ac-Ala<sub>8</sub>-LysH<sup>+</sup> (a helix) under the same conditions. The helical conformation of Ac-Ala<sub>8</sub>-LysH<sup>+</sup> has been suggested as the origin of the different behavior. We here use first-principles conformational searches (all-electron density functional theory based on a van der Waals corrected version of the PBE functional, PBE+vdW) to study the microsolvation of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> with one to five water molecules and the monohydration of Ac-Ala<sub>8</sub>-LysH<sup>+</sup>. In both cases, the most favorable water adsorption sites break intramolecular hydrogen bonds associated with the ammonium group, in contrast to earlier suggestions in the literature. A simple thermodynamic model yields Gibbs free energies  $\Delta G^0(T)$  and equilibrium constants in agreement with experiments. A qualitative change of the first adsorption site does not occur. For few water molecules, we do not consider carboxyl deprotonation or finite-temperature dynamics, but in a liquid solvent, both effects would be important. Exploratory *ab initio* molecular dynamics simulations illustrate the short-time effects of a droplet of 152 water molecules on the initial unsolvated conformation, including the deprotonation of the carboxyl group. The self-solvation of the ammonium group by intramolecular hydrogen bonds is lifted in favor of a solvation by water.



## 1. INTRODUCTION

**1.1. Hydrogen Bonds, Microsolvation, and Competing Hydration Sites for Ac-Ala<sub>n</sub>-LysH<sup>+</sup>.** The creation, destruction, and recreation of hydrogen bond networks is a fundamental process that governs inorganic chemical processes in solution and much of biochemistry alike. Indeed, the very energy scale of hydrogen bond formation is what enables life as we know it under ambient conditions. Much lower temperatures would lead to static, frozen hydrogen bond networks only, whereas much higher temperatures would render the energy scale of hydrogen bonding irrelevant. In detail, however, a broad range of *specific* hydrogen bond energies is active throughout biology to control the molecular-scale processes of life: From the relatively static structural hydrogen bonds in Watson–Crick pairs in DNA or protein secondary structure (helices, sheets) to the dynamic, changing hydrogen bond networks of a liquid solvent.

In order to understand and control processes in a biochemical environment with specificity, it is of paramount importance to obtain a quantitative picture of the energy scales of and the competition between different hydrogen bonds. For instance, a peptide solvated by water can exhibit multiple different hydrophilic sites. The actual hydrogen bond network formed will be a result of the competition between (i) intramolecular (self-solvating) hydrogen bonds, (ii) solvent–solute interactions at different hydrophilic sites including the

possibility of more or less statically embedded, structural water molecules, and (iii) solvent–solvent interactions. Today, it is possible to model many of these processes atomistically in a computer, but for a truly predictive modeling, the subtle balance between the individual hydrogen bond patterns must be captured precisely. This balance, however, is difficult to decompose into its parts quantitatively in a full and dynamic solvent environment. In contrast, a much more quantitative reference picture can be obtained by focusing first on the formation of individual hydrogen bonds in steps of one solvent molecule at a time. This successive “microsolvation” of an initially unsolvated solute molecule by individual water molecules is a key technique to connect benchmark spectroscopic experiments and theory on equal footing in a precisely controlled environment (vacuum). Experimentally, the range of systems studied includes amino acids (neutral, protonated, and capped) and their derivative<sup>1–9</sup> peptides,<sup>10–23</sup> sugars and other biomolecules (e.g., refs 24 and 25 and references therein). Likewise, many theoretical studies have focused on the hydration of neutral,<sup>26–37</sup> protonated,<sup>1,38–41</sup> and deprotonated<sup>35</sup> individual amino acids, of infinite periodic models of the peptide backbone,<sup>42,43</sup> and very recently of a long

Received: October 4, 2012

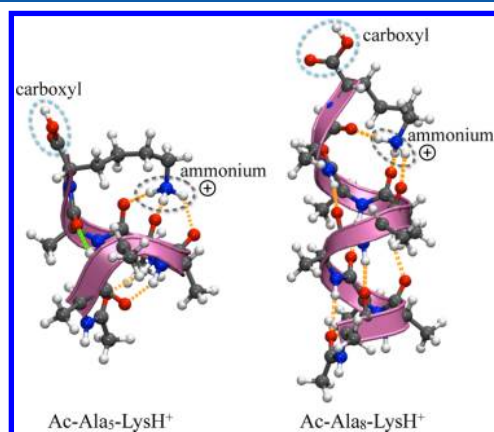
Revised: November 21, 2012

Published: November 21, 2012

finite polypeptide, capped neutral Ala<sub>15</sub>.<sup>44</sup> Some trends for amino acids or small peptides, established from such studies are as follows (see, e.g., refs 17 and 20 and references therein): (1) The water molecules are stabilized close to the charged sites in the amino acid (or peptide). (2) For individual hydration sites, the water binding energy (BE) decreases with increasing hydration for amino acids. Approximately constant BEs result for larger peptides with multiple hydration sites.<sup>17</sup> A recent, impressive experimental spectroscopic fingerprint study of up to 50 water molecules adsorbed at a decapeptide (gramicidin S) demonstrated that here, indeed, the self-solvated charged ammonium sites also are affected first.<sup>45</sup>

Our work focuses specifically on the first-principles prediction of the structure and energetics of the initial hydration steps of two peptides Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>, containing 80 and 110 atoms (without water), respectively. Both peptides have been studied extensively in experiments without any solvent,<sup>14,46–50</sup> and in microsolvation experiments that determined equilibrium constants for the initial water adsorption step.<sup>13,14,20</sup> Addressing peptides of these sizes allows us to focus on an important conceptual feature: The presence of two different, well separated and competing candidate sites for solvation, i.e., the C-terminal carboxyl group (protonated in the gas phase) and the protonated<sup>49,51</sup> ammonium group at the end of the Lys side chain.

The unsolvated structures of both peptides can be seen in Figure 1, established by benchmark experiments<sup>14,46,48,50</sup> and



**Figure 1.** Unsolvated ground state structures of the peptide molecules Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> as determined in refs 46 and 47. The (charged) ammonium and the carboxyl groups, which are candidate sites for initial hydration, are marked by dashed enclosures. In Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, the bond highlighted in green runs counter to what would be the normal helix dipole.

recent first-principles theory.<sup>46,47</sup> The structure of Ac-Ala<sub>8</sub>-LysH<sup>+</sup> is  $\alpha$ -helical, while the shorter Ac-Ala<sub>5</sub>-LysH<sup>+</sup> is not a helix. Instead, its N terminus is connected to the backbone near the C terminus by a hydrogen bond which runs counter to what would be the normal helix dipole (highlighted in green in Figure 1). The N terminus itself is protected by the acetyl group and contains no unsolvated hydrogen bonding sites. At the C terminus, the structure of the protonated ammonium group is very similar (but not identical) in either case, folding back to connect to the carbonyl groups of the backbone. The ammonium group is thus fully self-solvated, in contrast to the C-terminal carboxyl group.

Both the ammonium group and the carboxyl group are candidate sites for an initial hydration, but what is not settled is which site is preferred. An early empirical force field based simulation (ref 52, Figure 4) shows a H<sub>2</sub>O location near the ammonium group of Ac-Ala<sub>20</sub>-LysH<sup>+</sup> but without breaking any intramolecular hydrogen bonds. However, in the same paper, water is shown inserted into just such an ammonium-self-solvating intramolecular bond for the *doubly protonated* peptide. A later study by the same group<sup>13</sup> reiterates the conclusion that H<sub>2</sub>O does not break the self-solvating H bonds near the ammonium group for helical Ac-Ala<sub>n</sub>-LysH<sup>+</sup> ( $n = 8–20$ ). Another force field guided structure model was presented by Liu, Wyttbach, and Bowers,<sup>14</sup> showing the first H<sub>2</sub>O molecule at the carboxyl group instead. What is clear from these considerations is that assigning the correct hydration site from several candidates is a significant, not yet settled challenge, which we address in this paper.

A striking observation is that the measured gas phase equilibrium constants for H<sub>2</sub>O adsorption drop significantly when going from short to long peptides along the series Ac-Ala<sub>n</sub>-LysH<sup>+</sup> ( $n = 4–20$ ). Kohtani and Jarrold<sup>13</sup> attribute this decrease to a structural difference of the H<sub>2</sub>O adsorption sites at short non-helical peptides in comparison to long helical peptides, placing the onset of helix formation with helix length at  $n = 8$ . Liu and co-workers,<sup>14</sup> on the other hand, invoke the increasing dipole of a helix with length as a potential reason. In the present work, we address this question. Although the energy terms involved are subtle, we conclude that neither a drastic adsorption site change nor electrostatics are to blame. Instead, the adsorbed H<sub>2</sub>O appears to have an adverse impact on the vibrational free energy of a helix; i.e., a finite- $T$  entropic effect makes the difference.

**1.2. Structure of This Paper.** The layout of our paper is as follows.

In section 2, we address the basic methodological aspects. Section 2.1 describes the level of theory used (all-electron density-functional theory in the Perdew–Burke–Ernzerhof<sup>53</sup> generalized-gradient approximation with a correction<sup>54</sup> for van der Waals dispersion interactions) and our strategy for the conformational search. In section 2.2, we summarize the expressions used to calculate water binding energies, Gibbs free energies, and equilibrium constants. Section 2.3 covers the *ab initio* molecular dynamics protocol, and section 2.4 quantifies the accuracy of our first-principles approach by benchmark calculations for microsolvated protonated methylamine<sup>12,14,55,56</sup> and valine.<sup>6,19</sup>

Section 3 contains all our key results. Section 3.1.1 summarizes the nonhydrated conformations, and section 3.1.2 illustrates the conformational search for monohydrated Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. In sections 3.1.3, 3.1.4, and 3.1.5, we address the optimum monohydrated conformations for the non-helical conformational ground state, the lowest-energy helical conformer, and the energetics of alternative hydration sites, respectively, for Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. Section 3.2 then covers the monohydrated structures and hydration energies and enthalpies of Ac-Ala<sub>8</sub>-LysH<sup>+</sup>. For both Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>, the first H<sub>2</sub>O molecules are preferentially inserted into the intramolecular hydrogen bonds of the ammonium group, contrary to the earlier assertions.

Our calculated finite-temperature equilibrium constants for water adsorption at both peptides are in close agreement with earlier experiments (sections 3.1.3 and 3.2). Most importantly, our results show that the drop in the water adsorption

propensity from Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> originates from the vibrational contribution, which leads to a sufficiently large quantitative adsorption energy difference of  $\sim 0.04$  eV = 1 kcal/mol. Regarding the zero-temperature adsorption enthalpies, we find values that are comparable in magnitude to other calculations in the literature but that appear to disagree with an earlier experimental extrapolation to zero temperature.<sup>14</sup> One possible origin of this discrepancy is the assumed behavior of the entropy with temperature (in experiment, in theory, or both), but a definitive resolution is not within the scope of our work.

In section 3.3, we address the formation of the initial “nuclear” hydrogen bond networks of two to five water molecules with Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. In section 3.4, we present exploratory *ab initio* molecular dynamics (AIMD) simulations to connect our findings to the solvation of the same peptide by a droplet of 152 water molecules at finite temperature. The assessed time scale (up to 20 ps) reveals that the expected deprotonation of the carboxyl group occurs almost immediately. The self-solvation of the ammonium group is lifted both for non-helical and helical conformers.

Finally, section 4 summarizes the key findings and concludes our paper.

The structures (Cartesian coordinates) of all molecular conformations shown and discussed in our work are provided as Supporting Information in xyz format. Any structure is thus available for easy visualization with standard molecular viewing programs.

## 2. METHODS

**2.1. Level of Theory and Conformational Search.** Our goal is to identify the most stable hydrated conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> and their energy hierarchy with the accuracy of quantum-mechanical first principles. Our first-principles method of choice is density-functional theory (DFT) using the Perdew–Burke–Ernzerhof (PBE) GGA functional<sup>53</sup> with van der Waals corrections included via a  $C_6/R^6$  term.<sup>54</sup> We refer to this combination as “PBE+vdW” throughout the text. Importantly, the  $C_6$  coefficients are here derived from the self-consistent electron density in a nonempirical way. This level of theory has previously been shown to yield an accurate representation of the potential energy surface for alanine-based peptides<sup>46,57,58</sup> as well as for water clusters.<sup>59</sup> For the competing ammonium and carboxyl hydration sites of interest here, we provide additional benchmarks in section 2.4.

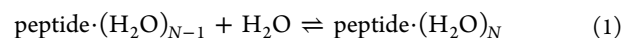
Since the full conformational space is huge, a direct search based on DFT is currently prohibitive. To narrow down the space of candidate conformations, we begin with a force-field (FF) guided search, using the empirical OPLS-AA<sup>60</sup> force field and a basin hopping search method implemented in the TINKER package.<sup>61</sup> These initial searches typically occur within a 25 kcal/mol ( $\approx 1$  eV) energy window using 15 torsional modes. We do not employ any constraints in the searches unless explicitly stated otherwise.

Next, we reoptimize a wide range of the energetically most favorable conformers found during the FF search. We typically relax from around a few hundred to a thousand lowest energy FF conformers in DFT. Specific numbers for each case are given in the relevant sections of the text. The DFT calculations are performed using the FHI-aims<sup>62</sup> program package for an accurate, all electron description based on numeric atom-centered orbitals. Initially, “light” settings for integration grids, the electrostatic potential, and the basis sets in FHI-aims are

used to relax the FF-generated candidate conformers. In a compact notation (a comprehensive technical explanation is given in ref 62), light settings comprise basis sets up to the *tier1* level, basis functions extended up to 5 Å from each nucleus, and a Hartree potential multicenter expansion up to a maximum angular momentum of  $l_{\max} = 4$ . The lowest-energy conformers found in this way (of the order of 10 for each separate search) are then postrelaxed with “tight” computational settings, using accurate *tier2* basis sets,<sup>62</sup> basis functions up to 6 Å from each nucleus, and  $l_{\max} = 6$  for the Hartree potential. The relative conformational energies obtained by calculations with light and tight settings typically differ by less than 0.01 eV. These energies, which correspond to the electronic energies at the local minima of the PBE+vdW potential energy surface (PES), are denoted by “PES min.” or “ $E_{\text{PES}}$ ” in the text and relevant tables.

The distinct structural minima of the peptides at the PBE+vdW level are classified into families according to their hydrogen bond (H-bond) patterns. This is done by means of a simple script that assigns H-bonds between (N–)H and (C–)O in each conformer by a distance criterion ( $d_{\text{OH}} < 2.5$  Å) and compares patterns for all the conformers. This classification is robust in the sense that it might allow more than one PES minimum structure with the same overall pattern to be classified into one family, rather than being too detailed and separating the conformational space artificially into many families. Thus, members of a particular family may differ in the exact orientation of the LysH<sup>+</sup> side chain or the COOH group near the C-terminus. In the case of the hydrated peptides, the exact positions where the water molecule binds to the peptide give rise to separate families, even if the peptide conformation itself is found to be the same. We compare this classification to a more conventional, root-mean-square deviation (RMSD) based separation criterion<sup>63</sup> in section 3.1.2 below.

**2.2. Binding Energies, Gibbs Free Energies, and Equilibrium Constants.** We are interested in the reactions (equilibrium or otherwise)



where  $\text{peptide} \cdot (\text{H}_2\text{O})_N$  denotes a specific peptide with  $N$  adsorbed water molecules (here,  $N = 1-5$ ). Step-wise binding energies (reaction energies), with or without thermal and/or entropic contributions, are defined as

$$\begin{aligned} \text{BE} = & E(\text{peptide} \cdot (\text{H}_2\text{O})_{N-1}) + E(\text{H}_2\text{O}) \\ & - E(\text{peptide} \cdot (\text{H}_2\text{O})_N) \end{aligned} \quad (2)$$

Thus,  $\text{BE} > 0$  means that binding is favorable.

In practice, BE can refer to different objects: to total energy differences  $\Delta E$ , to Helmholtz free energy differences  $\Delta F$ , or to Gibbs free energy differences  $\Delta G$ . By convention,  $\Delta G$ ,  $\Delta F$ , or  $\Delta E$  for eq 1 have the opposite sign as the BE in eq 2. Experiments often focus on equilibrium reaction conditions, where  $\Delta G = 0$  by definition. On the other hand, theory yields  $\Delta E$ ,  $\Delta F$ , or  $\Delta G$ , for instance, through the approximations defined below, for any conditions (equilibrium or otherwise). The indiscriminate use of equilibrium and nonequilibrium quantities together can lead to considerable confusion. Here, we therefore spell out in detail the pertinent equations which we use.

The experiments of interest here<sup>13,14,17,20</sup> yield equilibrium constants  $K_N$  as their direct output:

$$K_N = \frac{I_N}{I_{N-1}} \frac{p_0}{p(\text{H}_2\text{O})} \quad (3)$$

$I_N$  and  $I_{N-1}$  are measures (e.g., intensities in a mass spectrometer) of the particle numbers per volume of the peptide complexes peptide·(H<sub>2</sub>O)<sub>N</sub> and peptide·(H<sub>2</sub>O)<sub>N-1</sub>, respectively.  $p(\text{H}_2\text{O})$  is the water partial pressure, and  $p_0$  is a reference pressure (typically, standard ambient pressure  $p_0 = 1.01325 \times 10^5$  Pa) to make  $K_N$  dimensionless. In equilibrium,  $\Delta G = 0$ , and  $K_N$  is defined as

$$0 = \Delta G = \Delta G_N^0 + k_B T \ln K_N \quad (4)$$

( $\Delta G$  is given per individual reaction step, i.e., per particle).  $\Delta G_N^0$  is the Gibbs free energy difference of the reaction for the mixture at the reference conditions (no equilibrium at those conditions).

There are now two paths to connect experiment to theory: Either, use temperature dependent measurements to connect the experimental  $K_N$  back to  $\Delta G_N^0(T = 0)$ , and hence to the zero-temperature enthalpy of binding,  $\Delta H^0(T = 0)$ ,<sup>14,17,20</sup> or use the “forward route”: Assuming ideal mixtures (ideal gas law), predict approximate equilibrium conditions; i.e., find the mixtures  $I_N$ ,  $I_{N-1}$ ,  $p(\text{H}_2\text{O})$  for which  $\Delta G = 0$  based on calculated total energies. We here pursue the latter path. Furthermore, we can also compute  $\Delta G \neq 0$  for any nonequilibrium mixture.

The Helmholtz energy  $F$  is formally given by

$$\mathbf{F} = U - TS \quad (5)$$

where  $U$  is the internal energy of the system,  $T$  is the temperature, and  $S$  the entropy. It is related to the Gibbs free energy through the relation

$$\mathbf{G} = U + PV - TS = \mathbf{F} + PV \quad (6)$$

Since  $PV = Nk_B T$  for an ideal gas of  $N$  particles, the Gibbs free energy per molecule is

$$G = F + k_B T \quad (7)$$

Since our reaction removes one particle, the change in  $G$  per reaction step at constant temperature  $T$  is given by

$$\Delta G = \Delta F - k_B T \quad (8)$$

For systems of ideal polyatomic gases, the Helmholtz free energy per molecule can be computed through

$$F = F_{\text{trans}} + F_{\text{int}} + E_{\text{PES}} \quad (9)$$

$F_{\text{int}}$  is the contribution due to the internal degrees of freedom consisting of rotations and vibrations, and  $F_{\text{trans}}$  is the translational part of the free energy. Assuming that the harmonic approximation for the intramolecular potential energy surface holds, and neglecting any rotational–vibrational coupling, the internal free energy is given by

$$F_{\text{int}} = F_{\text{vib}} + F_{\text{rot}} \quad (10)$$

$$F_{\text{vib}} = \sum_i^{3M-6} \left[ \frac{\hbar\omega_i}{2} + k_B T \ln(1 - \exp^{-\hbar\omega_i/k_B T}) \right] \quad (11)$$

$$F_{\text{rot}} = -k_B T \ln \left[ \pi^{1/2} \left( \frac{2k_B T}{\hbar^2} \right)^{3/2} \sqrt{I_x I_y I_z} \right] \quad (12)$$

$\omega_i$  is the frequency of the normal modes of vibration,  $M$  is the number of atoms in the molecule or complex, and  $I_x$ ,  $I_y$ , and  $I_z$

are the moments of inertia along the three directions.  $T = 0$  defines the zero-point energy (ZPE) corrections used below.

The translational part of the free energy,  $F_{\text{trans}}$ , captures the impact of the pressure in an ideal gas of a given molecule. We have<sup>64</sup>

$$F_{\text{trans}} = -k_B T \left[ \ln \left( \frac{mk_B T}{2\pi\hbar^2} \right)^{3/2} + \ln \frac{k_B T}{p} + 1 \right] \quad (13)$$

where  $p$  is the partial pressure and  $m$  is the mass of the molecule or complex. We now have all the pieces together to relate  $K$  to  $\Delta G$  and write the mass-action law explicitly. For our reaction, we find

$$\Delta G = \Delta E_{\text{PES}} + \Delta F_{\text{int}} - k_B T \ln \left[ \left( \frac{2\pi\hbar^2}{k_B T} \frac{m_{\text{peptide}\cdot(\text{H}_2\text{O})_N}}{m_{\text{peptide}\cdot(\text{H}_2\text{O})_{N-1}} m_{\text{H}_2\text{O}}} \right)^{3/2} \cdot \frac{p_0}{k_B T} \right] + k_B T \ln \left[ \frac{I_N}{I_{N-1}} \frac{p_0}{p(\text{H}_2\text{O})} \right] \quad (14)$$

For equilibrium conditions  $\Delta G = 0$ , the first three terms equate to  $\Delta G_N^0$  and the last term defines  $K_N$  as in eqs 3 and 4. Then,  $\Delta G_N^0$  carries no pressure dependence, and neither does  $K_N$ .

**2.3. Ab Initio Molecular Dynamics Protocol.** In addition to stepwise microhydration of the peptides, we carry out *ab initio* molecular dynamics (AIMD) simulations of some of the structures hydrated with 152 water molecules forming a large cluster (vacuum outside). We begin by placing the isolated or microsolvated conformer in a box of TIP3P<sup>65</sup> water molecules using the XYZEDIT program in TINKER.<sup>61</sup> The surrounding water structure is then minimized in the OPLS-AA<sup>60</sup> force field. The resulting conformation of the peptide plus surrounding water molecules is used as the starting point for our AIMD calculations. The velocities of the AIMD run are randomly initialized using a Maxwell–Boltzmann distribution. This is followed by the actual AIMD calculation with the PBE+vdW functional and light settings for up to 20 ps using the Bussi–Donadio–Parrinello thermostat<sup>66</sup> at 300 K and a thermostat relaxation time of 0.1 ps. The time step used is 1 fs.

**2.4. Accuracy of Our Computational Approach: Hydration of Methylamine and Valine.** The purpose of this paper is not just to identify overall hydration energies but rather the much more subtle competition between protonated ammonium and carboxyl sites, and the formation of a “nuclear” hydrogen bond network by up to five water molecules. To capture this balance, numerous effects must be accurately represented: Electrostatics and the polarizability of individual molecules or groups of atoms, hydrogen bonds and their cooperativity, dispersion interactions, etc. In addition, subtle numerical errors must be avoided, such as the basis set superposition errors (BSSE)<sup>67</sup> that are often associated with a finite basis set.

As stated above, we employ DFT at the PBE+vdW<sup>54</sup> level for our predictions. The numerical implementation used is the FHI-aims code and tight computational settings with accurate numerical *tier2* basis sets,<sup>62</sup> implying sub-meV (per molecule) total energy uncertainties from the electrostatic and integration grid settings used.<sup>62,68</sup> An important feature of the numerical basis sets used here is that both the near-nuclear behavior and wave function tails are accurately represented. As a result, BSSE are negligible in practice for standard DFT.<sup>62</sup> This absence of

BSSE is critical not just for intermolecular energies, where a counterpoise (CP) correction<sup>67</sup> can capture much of the error, but also for intramolecular conformational energies of large peptides. For the latter, no unique CP correction formalism exists. In our approach, intramolecular conformational energies are also essentially BSSE-free at the outset.<sup>62</sup> Benchmarks for the accuracy of the *full* approach have been given in ref 54 for weakly bonded systems, for alanine-based peptides in refs 57 and 58, or for water clusters in ref 59. To validate our full computational approach for the systems of interest here, we provide an additional comparison against *experimental* benchmark data (stepwise hydration energies) that have been used as points of reference in earlier publications.<sup>6,12,14,19</sup> We address three cases: protonated methylamine (one, two, and three H<sub>2</sub>O), the protonated valine amino acid (one H<sub>2</sub>O), and protonated valine with 2 H<sub>2</sub>O.

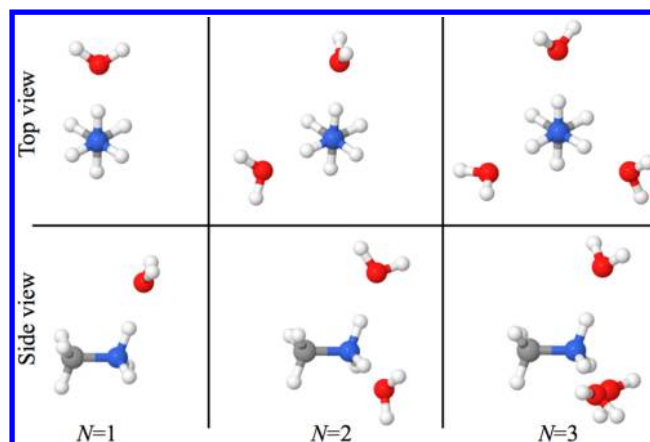
**Protonated Methylamine (1, 2, and 3 H<sub>2</sub>O).** In Table 1, we list the calculated sequential water binding energies ( $N = 1-3$ )

**Table 1. Stepwise Hydration Energies  $N = 1, 2, 3$  for Protonated Methylamine in eV, Calculated at the PBE+vdW Level of Theory, “tight” Computational Settings, and *tier2* Basis Sets as Used in the Remainder of This Paper<sup>a</sup>**

	$N = 1$	$N = 2$	$N = 3$
stepwise BE, PBE+vdW (this work)	0.784	0.644	0.548
CP correction (this work)	-0.003	-0.002	-0.002
stepwise BE, experiment (ref 56)	0.815	0.633	0.538
stepwise BE, experiment (ref 55)	0.729	0.633	0.533
stepwise BE, DFT-B3LYP (ref 12)	0.733	0.603	0.516

<sup>a</sup>ZPE corrections are added; a counterpoise correction for BSSE was *not* added. The CP correction for each case is given separately. For comparison, we include the experimental<sup>55,56</sup> and calculated<sup>12</sup> reference data (DFT-B3LYP level of theory<sup>69</sup>), as given in Table 1 of ref 14.

to the simplest possible protonated ammonium group, methylamine (CH<sub>3</sub>NH<sub>3</sub><sup>+</sup>), compared to reference data (DFT-B3LYP calculated values and experiments) given in ref 14. This case demonstrates our accuracy for the absolute binding energies at a single hydration site (ammonium). Our conformers were obtained by placing water molecules successively at each proton, and then fully relaxing each conformer until any remaining total energy gradients were below 10<sup>-4</sup> eV/Å. Vibrational frequencies and the ZPE were then calculated by an accurately converged finite-difference approach. All resulting conformers are shown in Figure 2. The ZPE-corrected sequential binding energies match precisely the decrease in H<sub>2</sub>O binding energy with increasing number of water molecules seen in experiment. The experiments<sup>55,56</sup> do not quite agree on the first water binding energy, and our calculated value deviates by approximately 1 kcal/mol (0.043 eV) from either value; for the second and third hydration energy, the agreement is clearly within less than 0.5 kcal/mol. The calculations in ref 12 are performed with a different density functional (B3LYP), numerical procedure, etc., and show the approximate level of agreement between different levels of theory for small systems (in large systems, the absence of dispersion interactions would be a problem in straightforward B3LYP calculations). Regarding our own level of theory, PBE+vdW, the benchmark demonstrates the accuracy that can be expected compared to experimental values. As seen in a separate line in the same table, the CP correction would



**Figure 2.** Top and side view of sequentially hydrated methylamine conformers used for the benchmark of Table 1.

amount to 2–3 meV (less than 0.1 kcal/mol). Indeed, in explicit tests for our larger monohydrated peptides, we find CP corrections of the same magnitude, a few meV per water molecule at most. In general, we therefore do not need a CP correction.

**Protonated Valine, 1 H<sub>2</sub>O.** This is a case in which protonated ammonium and carboxyl groups compete. Specifically, ref 19 shows by vibrational spectroscopy that the first water molecule adsorbs at the ammonium group (Figure 5 of ref 19, conformer A). A calculated DFT-B3LYP energy difference to the lowest-energy conformation with H<sub>2</sub>O at the carboxyl group is given in the same figure (conformer C),  $\Delta E = 5$  kJ/mol (0.05 eV); i.e., conformer A is more stable. Our present approach (PBE+vdW, FHI-aims “tight”, *tier2*) also predicts the experimentally observed conformer A to be the more stable one, by  $\Delta E = 0.020$  eV (ZPE included). Note that there is no reason to expect exact agreement between the two different levels of theory employed. The important point is the correct prediction of the experimental preference for the hydration at the ammonium site.

**Protonated Valine, 2 H<sub>2</sub>O.** A yet more subtle test of the site-competitive microsolvation in this system is that of two water molecules hydrating the protonated valine amino acid. In ref 19, the experimentally verified conformer (labeled A in Figure 6 of ref 19) has both water molecules hydrating the ammonium group. However, the calculated lowest-energy conformer (labeled B in Figure 6 of ref 19 and 0.015 eV lower in DFT-B3LYP) has one H<sub>2</sub>O attached to the ammonium group and another to the carboxyl group. For the same system and density functional, Gao and co-workers recomputed the energy difference as 0.052 eV (ref 6, Table 3), also in favor of conformer B. The experimental energy hierarchy is correctly reproduced at the MP2 level of theory, where A is lower than B by approximately 0.043 eV.<sup>6</sup> We have computed the energy difference for the same doubly hydrated conformers. Similar to DFT-B3LYP, the PBE+vdW functional arrives at an energy difference of 0.067 eV in favor of conformer B. The use of a hybrid density functional (PBE0<sup>70</sup>+vdW) at the exact same geometries reduces the difference to 0.026 eV, still in favor of conformer B.

The point of addressing dihydrated valine is to clearly delineate not just the successes but also the current limits of quantum-mechanical production methods. It would be too simple to declare that any higher-level theory (e.g., MP2) is the answer: For instance, MP2 overestimates the  $C_6$  coefficients of

long-range dispersion interactions,<sup>71</sup> which are critical in the very systems (larger peptides) that we here address. One would expect the “gold standard” of quantum chemistry, coupled-cluster theory with singles, doubles, and perturbative triples (CCSD(T)), to provide systematically improved answers, but such computations are currently unaffordable for systems of 80 and 110 atoms (Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>, respectively).

In our own study (below), we find the ammonium site to be hydrated first for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>. Luckily, the possible slight overestimation of the water binding energy at the carboxyl group of dihydrated valine would not affect this result (correcting the overestimation would make ammonium even more favorable). Achieving a general accuracy of  $\sim 1$  kcal/mol (0.043 eV) or better for hydrogen bonding energies is still a significant challenge. Nonetheless, the above benchmarks indicate that our level of agreement with experiments for the hydration sites of interest here is in this range.

### 3. RESULTS AND DISCUSSION

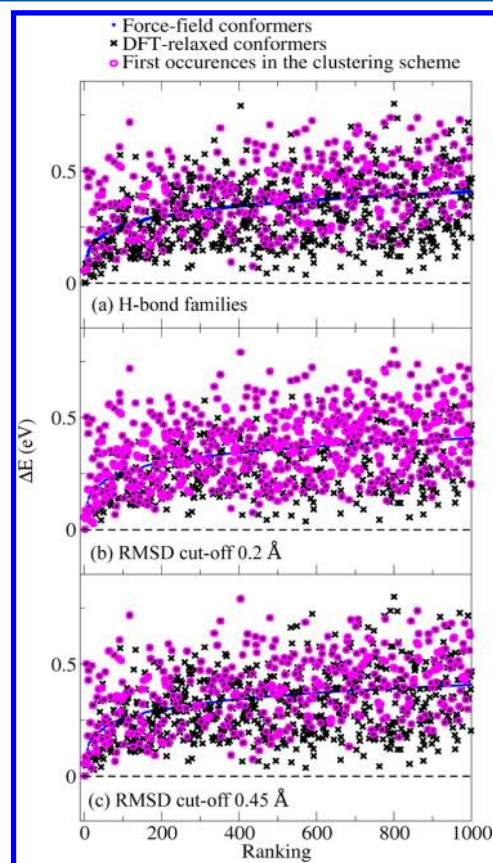
**3.1. Monohydration of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>.** *3.1.1. Non-Hydrated Conformations.* Detailed descriptions of the non-hydrated low-energy conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> can be found in refs 46 and 58. We here focus on the conformers denoted by “g-1” (see Figure 1), “ $\alpha$ -1”, and “ $\alpha$ -2”. In all three conformers, the lysine NH<sub>3</sub><sup>+</sup> (ammonium) group is fully self-solvated by H-bonding to three or even four carbonyl groups of the backbone. The C-terminal carboxyl group is not involved in any H-bonds. We briefly summarize the backbone features, denoting the respective residues by superscript numbers starting from the N terminus here and in the rest of the paper.

The lowest energy conformer is called “g-1”. It exhibits a particular feature where a H-bond, O(Ala<sup>5</sup>)–H(Ala<sup>2</sup>), connects a C-terminal carbonyl group to an N-terminal amide group. This H-bond thus goes against what would be the orientation in a helix. We shall see that the back-bonded O(Ala<sup>5</sup>) removes a potential hydrogen bonding site that is most favorable in an actual helical conformer.

The lowest-energy nonhydrated helical structure without water, called  $\alpha$ -1, occurs 0.10 eV higher in the PBE+vdW potential energy surface (PES). It has a bifurcated bond starting at the Ac termination between O(Ace)–H(Ala<sup>3</sup>) (3<sub>10</sub>-like, i.e., a 10-atom loop) and O(Ace)–H(Ala<sup>4</sup>) ( $\alpha$ -like, i.e., a 13-atom loop). This is followed by a H-bond between O(Ala<sup>1</sup>)–H(Ala<sup>5</sup>) ( $\alpha$ -like). The conformer that is next higher in energy is another helix, denoted  $\alpha$ -2. It occurs just 0.01 eV above  $\alpha$ -1. The H-bond at the N terminus O(Ace)–H(Ala<sup>3</sup>) is 3<sub>10</sub>-like, followed by a bifurcated bond between O(Ala<sup>1</sup>)–H(Ala<sup>4</sup>) (3<sub>10</sub>-like) and O(Ala<sup>1</sup>)–H(Ala<sup>5</sup>) ( $\alpha$ -like).

*3.1.2. Search for Hydrated Conformers.* The first step in the microsolvation of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> is the hydration by a single water molecule. The input structure for the unconstrained FF search, i.e., where any peptide structure is possible as allowed by the FF parameters, is generated by placing a water molecule in the vicinity of the most likely site for hydration of the isolated peptide conformer g-1, the ammonium group. Tests with other, randomly selected sites yielded the same outcome. Of the order of 10<sup>6</sup> structures are found within an energy window of 25 kcal/mol ( $\approx 1$  eV) employed during the search. The lowest 1000 FF structures are relaxed in FHI-aims using light settings. The lowest energy conformer belonging to each of the lowest 19 distinct H-bond families are then relaxed using tight settings to finalize the energy differences and hierarchies.

The correlation between the PBE+vdW (light settings, fully relaxed) predicted energy hierarchy and the FF energy hierarchy is shown in Figure 3a. Conformers found in the



**Figure 3.** Correlation between the force-field and PBE+vdW (relaxed) energy hierarchies for monohydrated Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. The reference energy (zero) corresponds to the lowest energy conformer in either energy function. Blue dots (monotonically increasing line): the lowest 1000 force field conformers ordered by their energy. Black crosses: the energies of the corresponding conformers relaxed with PBE+vdW. Magenta circles: the first occurrences, in the ranking scheme used, of the PBE+vdW (light) relaxed conformers (a) belonging to a given H-bond family, (b) having similar structures with RMSD cutoff distance 0.20 Å and (c) having similar structures with RMSD cutoff distance 0.45 Å.

OPLS-AA search (blue dots) have been ranked according to their energy hierarchy and appear as a monotonically increasing line. The black crosses show the relative energies of the corresponding conformers obtained after relaxation with the PBE+vdW functional. The reference energy (zero) corresponds to the lowest energy conformer in each energy function (FF and PBE+vdW). The lowest energy PBE+vdW conformer is obtained from the second lowest force-field conformer, but the spread between FF and DFT is large for higher-lying conformers. In the plots, several low-lying DFT conformers occur later in the FF ranking, indicating that the FF PES does not reflect the more accurate PBE+vdW PES. It is therefore a central question to ascertain just how many of the “late” low-lying DFT structures reflect substantially new conformers.

In Figure 3, we address this question for three different criteria to define specific “families” or clusters of structurally similar conformers. In Figure 3a, we highlight by magenta circles the first occurrences of the members of the “H-bond

families" (339 families) used through most of this work. For comparison, parts b and c of Figure 3 show the structural grouping obtained by the more conventional root-mean-square deviation (RMSD) criterion to characterize "clusters" of structures that are conformationally similar.<sup>63</sup> Specifically,

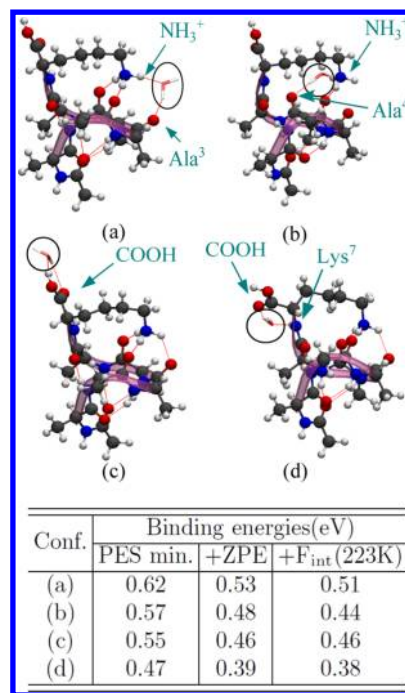
$$\text{RMSD} = \sqrt{\frac{1}{M_i} \sum_i \delta_i^2} \quad (15)$$

where  $\delta_i$  is the distance between  $M_i$  pairs of equivalent atoms, where we only include the heavy atoms (C, N, O) in the sum. Each cluster is delineated by a cutoff distance, such that no two structures in the same cluster can differ by more than this cutoff. In Figure 3b, magenta circles indicate the first occurrences of the members of RMSD clusters with a tight cutoff of 0.20 Å (634 clusters). In Figure 3c, a less rigid cutoff of 0.45 Å (388 clusters) is used. An interesting difference between all three schemes occurs already for the first two FF conformers. Their H-bond pattern as defined in section 2.1 is the same, and thus we include them in the same family in Figure 3a. FF rank 2 has the lower energy in DFT—the difference is in the exact orientation of the C-terminal carboxyl group and the conformation of the lysine side chain. However, even a cutoff of 0.45 Å puts them into different RMSD clusters. Because we wish to capture primarily H-bond patterns, the H-bond family assignment is therefore better suited for our particular purpose.

Irrespective of the criterion used, Figure 3 clearly shows that among the first 1000 conformers, most of the "late" newly found low energy data points up to a relative PBE+vdW energy of 0.2 eV are not entirely new conformers. Instead, they are FF generated candidate conformers that relax to PBE+vdW structures that have been seen before. This observation gives us confidence that the low energy PBE+vdW part of the conformational space is faithfully represented.

A final note concerns the protonation state of the carboxyl group. In a full H<sub>2</sub>O environment at neutral pH and ambient conditions, the carboxyl group should be deprotonated (we show this explicitly in section 3.4). For the unsolvated peptide, the carboxyl group is the known location of its extra proton.<sup>49,51</sup> As H<sub>2</sub>O molecules are added, they offer additional protonation sites, and a crossover of the proton to H<sub>2</sub>O is possible.<sup>72</sup> Such a crossover would not necessarily be captured in the FF-guided list of conformers and the following *local* structure optimization in PBE+vdW. For the monohydration of "our" peptides, we have explored some possible protonation changes from carboxyl to H<sub>2</sub>O explicitly, but none yielded a favorable outcome. Still, our *global* searches must be considered with the cautionary statement that a protonated carboxyl group was effectively assumed. We shall return to this question in section 3.4.

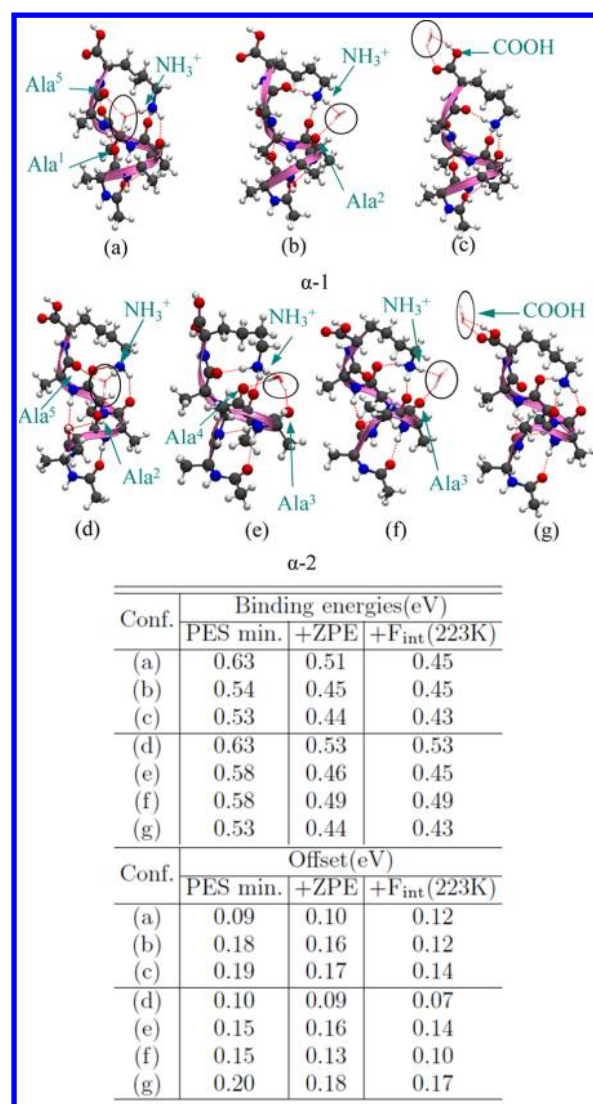
**3.1.3. *g*-1-Like Conformers.** Figure 4 shows the four lowest-energy hydration sites for the *g*-1 conformer of Ac-Ala<sub>3</sub>-LysH<sup>+</sup>. The binding energies of the configurations in order of decreasing binding energy are shown in the table below subfigures a–d, along with the respective BEs corrected for zero point and free energies (PES energy and internal degree of freedom contributions, eqs 10, 11, and 12 only) at  $T = 223$  K, the experimental temperature used in ref 13. The other experimentally relevant reference temperature is  $T = 260$  K (ref 14). Here and in Figures 5 and 8–10, free energy differences up to and including the internal free energy part are the same within 0.01 eV at  $T = 223$  K and  $T = 260$  K, so only  $T = 223$  K



**Figure 4.** Monohydrated instances of the *g*-1 conformer of Ac-Ala<sub>3</sub>-LysH<sup>+</sup> identified in the unconstrained search. (a–d) The lowest-energy water binding sites at the *g*-1 conformer, which include the global lowest-energy conformer (a). Table: H<sub>2</sub>O binding energies (PES local minima, ZPE corrected, and corrected for the internal free energy (eqs 10, 11, and 12) at  $T = 223$  K).

is shown. The situation is different for terms that include  $\Delta F_{\text{trans}}$ , the  $T$  dependence of which is significant. Thus, we report both temperatures for equilibrium reference Gibbs energy differences  $\Delta G_1^0(T)$  in Table 2.

The global minimum energy (DFT-PBE+vdW) hydrated structure found in our entire search is shown in Figure 4a. The first water molecule is inserted by disrupting the intramolecular H-bond between O(Ala<sup>3</sup>)–H(NH<sub>3</sub><sup>+</sup>). It acts both as a hydrogen bond donor and acceptor but participates in only two H-bonds. The next lowest energy site in Figure 4b would also "break" an intramolecular H-bond at the ammonium group, between H(NH<sub>3</sub><sup>+</sup>) and O(Ala<sup>4</sup>). Only then do we find hydration sites that are associated with the carboxyl group, in Figure 4c and d. No intramolecular H-bonds are broken for the latter two sites. We find a small but consistent energy difference of 0.05–0.07 eV (1.2 kcal/mol or more) between the favored hydration at the ammonium site and the less favorable site at the carboxyl group at zero and at finite temperature. This hydration site is consistent with the recent fingerprint study of a much larger decapeptide with two ammonium groups.<sup>45</sup> It does, however, disagree with the earlier, empirical force field based monohydration models of Ac-Ala<sub>*n*</sub>-LysH<sup>+</sup>. These models predicted either a hydration at the carboxyl group ( $n = 4$ )<sup>14</sup> or an attachment of H<sub>2</sub>O near the ammonium group but without breaking a self-solvating intramolecular bond ( $n = 20$ ).<sup>13</sup> On the basis of our benchmark findings in section 2.4 and the potentially considerably larger errors of empirical force fields compared to a first-principles approach, we conclude that the self-solvation of the ammonium group in a vacuum environment is partially lifted already upon contact with the first solvent molecule.



**Figure 5.** Monohydrated instances of the  $\alpha$ -1 and  $\alpha$ -2 conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> identified in the unconstrained search. (a–c) Lowest-energy water binding sites at  $\alpha$ -1. (d–g) Lowest-energy water binding sites at  $\alpha$ -2. Table: Upper part: H<sub>2</sub>O binding energies (PES local minima, ZPE corrected, and corrected for *internal* free energy at 223 K). The binding energy values for each conformer are calculated with respect to the unsolvated structures of its own H-bond family, i.e.,  $\alpha$ -1 or  $\alpha$ -2. Lower part: Energy offset of each depicted conformer from the global minimum energy monohydrated conformer found, Figure 4a.

In Table 2, we collect the calculated reference Gibbs free energy differences  $\Delta G_1^0(T)$  for the most stable water adsorption site at three different temperatures:  $T = 0$  K and the two temperatures at which the earlier experiments were performed,  $T = 223$  K<sup>13</sup> and  $T = 260$  K.<sup>14</sup> The table also contains data for Ac-Ala<sub>8</sub>-LysH<sup>+</sup>, which we will address in section 3.2.

The calculated values were obtained according to eq 14, taking atmospheric pressure as the reference pressure  $p_0$ . In addition, Table 2 also contains the corresponding measured  $\Delta G_1^0(T)$  values from ref 14 (for Ac-Ala<sub>4</sub>-LysH<sup>+</sup>, Ac-Ala<sub>6</sub>-LysH<sup>+</sup>, and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>) and ref 13 (for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>). In ref 13,  $\Delta G_1^0(T)$  values are not given directly but rather in the form of measured equilibrium constants. Using eq 4, we can convert the measured equilibrium constants  $K_1$  at  $T = 223$  K in Figure 2 of ref 13. The resulting experimental value is  $\Delta G_1^0(223 \text{ K}) \approx -0.20 \pm 0.02$  eV, where the error bar denotes

**Table 2.** Calculated  $\Delta G_1^0(T)$  (in eV, and corresponding to a Reference Pressure of  $p_0 = 1.01325 \times 10^5 \text{ Pa} = 760 \text{ Torr}$ ) for Monohydration of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> Compared to Literature Data<sup>a</sup>

peptide	method	$T = 0 \text{ K}$	$T = 223 \text{ K}$	$T = 260 \text{ K}$
This Work				
Ac-Ala <sub>5</sub> -LysH <sup>+</sup>	Theory/PBE +vdW	-0.53	-0.24	-0.19
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Theory/PBE +vdW	-0.51	-0.20	-0.14
Kohtani and Jarrold <sup>13</sup>				
Ac-Ala <sub>5</sub> -LysH <sup>+</sup>	Expt.		$-0.20 \pm 0.02$	
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Expt.		-0.15	
Liu et al. <sup>14</sup>				
Ac-Ala <sub>4</sub> -LysH <sup>+</sup>	Expt.			-0.15
Ac-Ala <sub>6</sub> -LysH <sup>+</sup>	Expt.			-0.13
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Expt.			-0.11
Ac-Ala <sub>4</sub> -LysH <sup>+</sup>	Expt./extrapolated	-0.30		
Ac-Ala <sub>6</sub> -LysH <sup>+</sup>	Expt./extrapolated	-0.25		
Ac-Ala <sub>8</sub> -LysH <sup>+</sup>	Expt./extrapolated	-0.21		
Ac-Ala <sub>4</sub> -LysH <sup>+</sup>	Theory/AMBER	-0.45		

<sup>a</sup>The results of Kohtani and Jarrold<sup>13</sup> were converted from  $K_1$  equilibrium constants that were read from Figure 2 of their work. The error bar given for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> indicates the uncertainty of our conversion due to the symbol size used in Figure 2 of Kohtani and Jarrold, not due to their experiment. Results from Liu et al.<sup>14</sup> include measured  $\Delta G_1^0(T)$  values at  $T = 260$  K, extrapolations to  $T = 0$  K, as well as a calculated value at  $T = 0$  K (AMBER force field, for adsorption at the carboxyl group).

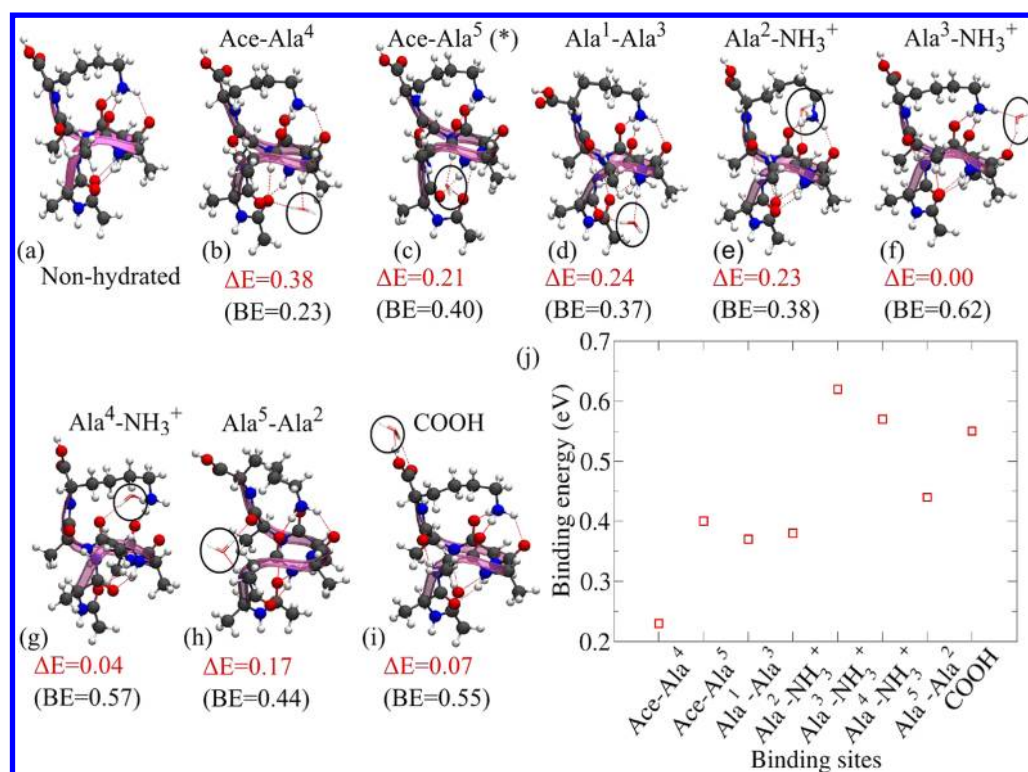
the uncertainty of our conversion (symbol size in the figure), not of the original experiment. Finally, we also include  $\Delta G_1^0(T)$  values for  $T = 0$  K from ref 14, which were obtained by extrapolating the slope of the measured  $\Delta G_1^0(T)$  (at finite  $T$ ) to zero.

What is encouraging is that our calculated  $\Delta G_1^0(T)$  values at the experimental temperatures are in rather close agreement with the measured values. The calculated  $\Delta G_1^0(223 \text{ K}) = -0.24$  eV is below the experimental value by  $\sim 0.04$  eV. Similarly, the calculated  $\Delta G_1^0(T)$  at  $T = 260$  K is 0.04–0.06 eV below the experimental  $\Delta G_1^0(T)$  values of Liu et al.<sup>14</sup> for Ac-Ala<sub>4</sub>-LysH<sup>+</sup> and Ac-Ala<sub>6</sub>-LysH<sup>+</sup>. Thus, the PBE+vdW overestimation of the H<sub>2</sub>O binding energy at finite  $T$  amounts to  $\sim 0.04$  eV. A systematic discrepancy of this magnitude is easily compatible with our computed benchmarks in section 2.4.

What is more puzzling is the comparison of the calculated and experimental (extrapolated) binding energy at  $T = 0$  K. For Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, we find a ZPE-corrected H<sub>2</sub>O adsorption energy of  $\sim -0.53$  eV. This value is in line with typical water adsorption energies at exposed charged sites,<sup>17,20</sup> but for the self-solvated ammonium groups of Ac-Ala<sub>4</sub>-LysH<sup>+</sup> and Ac-Ala<sub>6</sub>-LysH<sup>+</sup>, Liu et al.<sup>14</sup> arrive at much smaller  $\Delta H^0$  values,  $\Delta H^0 = -0.30$  and  $-0.25$  eV, respectively, by extrapolation of measured  $G_1^0(T)$  ( $T \approx 260$  K) to  $T = 0$  K.

The question is whether experiment or theory, or both, are responsible for this discrepancy. Regarding the calculated adsorption enthalpy at  $T = 0$  K, our experience suggests that





**Figure 6.** Systematic scan of all intramolecular hydrogen bonds of the g-1 conformer as monohydration sites, from the N terminus to the C terminus. Conformers were taken from the unconstrained search where available. A (\*) symbol marks hydrogen bonds which were never observed as hydration sites in the search. Here, the H<sub>2</sub>O adsorption structure was manually constructed and then fully relaxed. (a–i) Relaxed structures, intramolecular bond at which H<sub>2</sub>O was inserted, PES energy difference  $\Delta E$  to the lowest energy conformer (red) in eV, and corresponding H<sub>2</sub>O BE in eV (black, in brackets). (j) The BE as a function of the H-bond carbonyl distance from the N terminus.

even very pessimistic assumptions do not allow one to explain the discrepancy as a flaw of the level of theory used here. In fact, the adsorption energies found by us are roughly in line with similar calculations in the literature. A  $T = 0$  force field calculated H<sub>2</sub>O adsorption energy at the carboxyl group reported for Ac-Ala<sub>4</sub>-LysH<sup>+</sup> in ref 14 ( $-0.45$  eV) is of the same magnitude as our value for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> (PBE+vdW, ZPE corrected:  $-0.46$  eV in Figure 4). For water adsorption at a neutral Ac-Ala<sub>15</sub>-NHCH<sub>3</sub> peptide, using the X3LYP<sup>73</sup> density functional without any dispersion corrections, and including vibrational corrections for  $T = 298$  K, Marianski and Dannenberg report a water adsorption energy of  $-0.39$  eV.<sup>44</sup> If anything, the differences to our own case (charged site, dispersion interactions included, and zero temperature) all point toward a lower (stronger), not higher (weaker), adsorption energy for us.

The main difference between theory and experiment is the entropy term  $T\Delta S_1^0$  which is determined in ref 14 and used to extrapolate  $\Delta G_1^0(T)$  to  $T = 0$  K. Our calculated  $\Delta S_1^0$  values assume ideal mixing and are a factor of 2–3 larger than those reported by Liu et al. However, the approximate slope between the different experimental  $\Delta G_1^0(T)$  in the literature, of Liu et al. ( $T = 260$  K) for Ac-Ala<sub>4</sub>-LysH<sup>+</sup> and Ac-Ala<sub>6</sub>-LysH<sup>+</sup> and of Kohtani and Jarrold ( $T = 223$  K) for Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, is actually not far from the slope of our predicted  $\Delta G_1^0(T)$  with  $T$ . (Our results underestimate the experimental  $\Delta G_1^0(T)$  by the same amount of  $\sim 0.04$ – $0.06$  eV; see above.) Ultimately, the shape of  $\Delta S_1^0(T)$  leads to two sources of uncertainty that we cannot quantify in this paper:

- (1) We cannot comment on the original experimental extrapolation of ref 14. However, we believe that this is a straightforward procedure, limited only by the experimentally accessible  $T$  range.
- (2) Our calculated  $G_1^0(T)$ , based on the harmonic approximation and assuming ideal mixing (section 2.2), display a close to linear slope between  $T = 0$  and  $T = 260$  K, just like the shape that was assumed in the experimental extrapolation from finite  $T$  to  $T = 0$  K. It is in principle conceivable that this shape assumption does not hold over the entire  $T$  range. The actual entropy  $\Delta S_1^0(T)$  could deviate significantly from the harmonic approximation and/or ideal mixing in some temperature range, leading to a slope of  $G_1^0(T)$  that is not constant. However, we have no presently affordable way to capture the full anharmonic and nonideal entropy with a first-principles method and assess whether or not this is indeed the case.

Summarizing Table 2 for Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, we highlight again the close agreement between the experimental and calculated  $\Delta G_1^0(T)$  values at the experimental temperatures. Regarding the water binding energy at  $T = 0$  K, we have reason to believe that the calculated magnitude is correct, but we cannot speculate whether the experimental  $T \rightarrow 0$  K extrapolation, a deviation of  $\Delta S_1^0(T)$  from linearity, or both, are responsible for the apparent disagreement of  $T = 0$  K theory and the extrapolated experimental enthalpy.

**3.1.4.  $\alpha$ -Helix-Like Conformers.** We next investigate the impact of different peptide conformations on the hydration site. In Figure 5, we show the various hydration sites on the helices

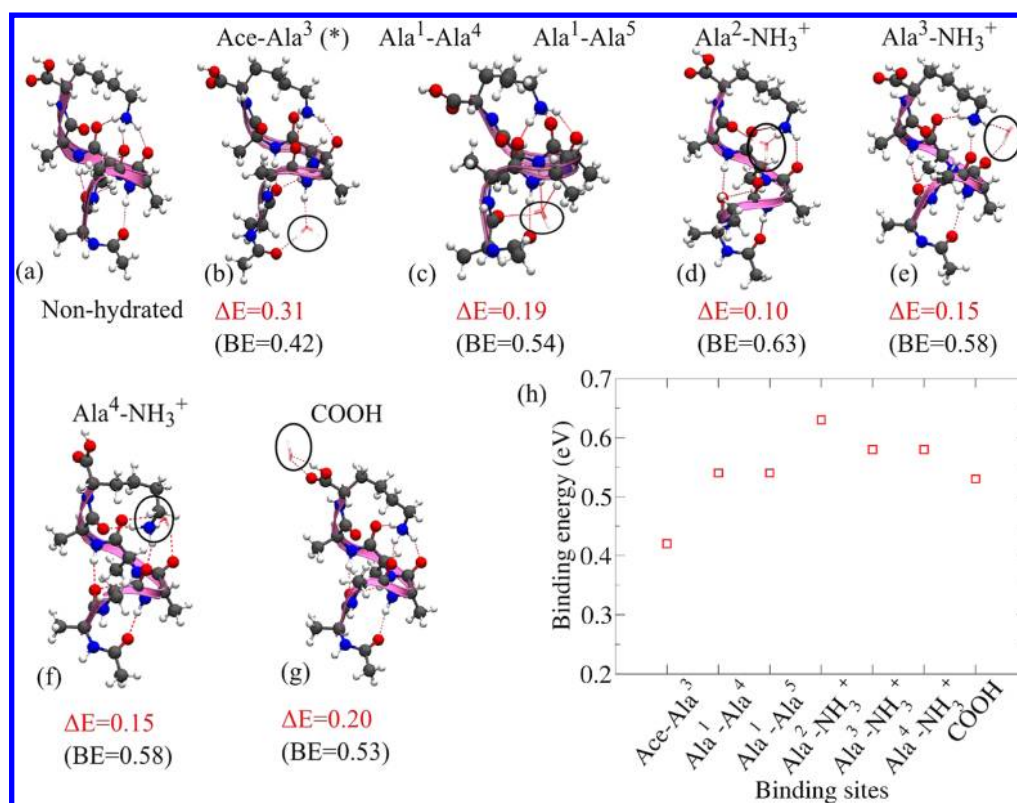


Figure 7. Same as Figure 6 but for the  $\alpha$ -2 conformer as the hydrated structure.

of the isolated case,<sup>46</sup>  $\alpha$ -1 and  $\alpha$ -2, identified from monohydration along with the water binding energies. In our search, the first non-g-1 monohydrated conformer occurs around 0.09 eV higher in energy than the optimum structure ( $\Delta F_{\text{int}}(223 \text{ K}) = 0.12 \text{ eV}$ ). It is the  $\alpha$ -1 helix of ref 46. The first water molecule is here inserted by forming three hydrogen bonds: to O(Ala<sup>1</sup>), O(Ala<sup>5</sup>), and H(NH<sub>3</sub><sup>+</sup>). This location disrupts again one self-solvating intramolecular hydrogen bond of the ammonium group of the unsolvated structure—that between H(NH<sub>3</sub><sup>+</sup>) and O(Ala<sup>1</sup>). In the g-1 structure, however, O(Ala<sup>5</sup>) is not available at all, as it is bonded back to the -H(Ala<sup>2</sup>) amide group. Thus, this triply hydrogen-bonded hydration “pocket” is a distinctive feature of the helical conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>.

The binding energy relative to the unsolvated  $\alpha$ -1 conformer is very similar to that of the first water at the g-1 conformer at  $T = 0$  but slightly lower at  $T = 223 \text{ K}$  (0.45 eV for Figure 5a vs 0.51 eV for Figure 4a). The other two low-energy monohydrated conformers of  $\alpha$ -1 again disrupt a single H-bond of the ammonium group (Figure 5b) or attach H<sub>2</sub>O to the carboxyl group (Figure 5c). While the  $T = 0$  energy clearly sets the lowest-energy conformer apart, this location is apparently unfavorable from a vibrational point of view. At  $T = 223 \text{ K}$ , we predict essentially the same H<sub>2</sub>O binding free energy for all three locations.

The other helical conformer,  $\alpha$ -2, is almost iso-energetic and occurs only 0.005 eV higher in energy. Again, the most energetically favorable H<sub>2</sub>O location is triply hydrogen-bonded between O(Ala<sup>2</sup>), O(Ala<sup>5</sup>), and H(NH<sub>3</sub><sup>+</sup>). One ammonium self-solvating H-bond is disrupted, and the O(Ala<sup>5</sup>) carbonyl group is involved. What is different are the details of the C-terminus of  $\alpha$ -2, allowing for an apparently more stable H<sub>2</sub>O placement also at  $T = 223 \text{ K}$ : For both helical conformers, the

binding site of Figure 5d is clearly stabilized over other hydration patterns of the helical conformers, including Figure 5e and f, which break self-solvating H-bonds of the ammonium group, and Figure 5g, with H<sub>2</sub>O placed at the carboxyl group.

In short, as in the case of the g-1 conformer, we find that the preferred adsorption sites in the helical conformers are broken hydrogen bonds of the self-solvated ammonium group, not the carboxyl group.

**3.1.5. Hydration at Other Intramolecular Hydrogen Bonds.** The lowest-energy monohydration of the g-1 and  $\alpha$ -1 conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> is similar in that H<sub>2</sub>O breaks a self-solvating hydrogen bond of the ammonium group in either case. These are, however, not the only hydration sites available. At the very least, others would be relevant in a fully solvated environment. Below, we shall see that the specific g-1 and helical conformations display some distinct differences in their hydration behavior.

In Figures 6 and 7, we illustrate this point for the g-1 and  $\alpha$ -2 conformers, respectively. We show the structures and binding energies that arise from a hypothetical H<sub>2</sub>O placement in or near each intramolecular hydrogen bond of the unsolvated conformers, starting from the N terminus and ending with the carboxyl group. Where available, the appropriate low-energy conformations were taken from the unconstrained search results described in sections 3.1.2 and 3.1.3. For the other cases, representative conformers were constructed by inserting a H<sub>2</sub>O molecule “by hand” in the respective hydrogen bond. All structures are fully relaxed.

What is similar in both cases is that the highest BE occurs at the ammonium group, and the lowest at the N-terminal acetyl group, but there is a striking difference. We find high binding energies for all three backbone residues that are connected to NH<sub>3</sub><sup>+</sup> (Ala<sup>2</sup>-Ala<sup>4</sup>) in the  $\alpha$ -2 case, and the highest binding

energy for H<sub>2</sub>O insertion at Ala<sup>2</sup>. In contrast, only the Ala<sup>3</sup> and Ala<sup>4</sup> sites are favorable in the global minimum conformer, g-1. H<sub>2</sub>O insertion at Ala<sup>2</sup>, on the other hand, is less favorable by ~0.2 eV. It is conceivable that a naive H<sub>2</sub>O insertion in this location would have to force open the “inverted” H-bond Ala<sup>5</sup>-Ala<sup>2</sup> of g-1. We thus expect this H-bond to be a stabilizing factor of the LysH<sup>+</sup> self-solvation, as long as the overall conformational pattern stays intact.

The  $\alpha$ -2 conformer is apparently also more flexible regarding the insertion of H<sub>2</sub>O in the “helix” hydrogen bonds (those connected to Ala<sup>1</sup>). We note that a greater flexibility of the “helical” conformers is already apparent in the unsolvated state. Here, the entropy contribution from significantly softer low-frequency vibrations leads to a systematic stabilization of helical H-bond networks over more compact ones.<sup>47,58</sup>

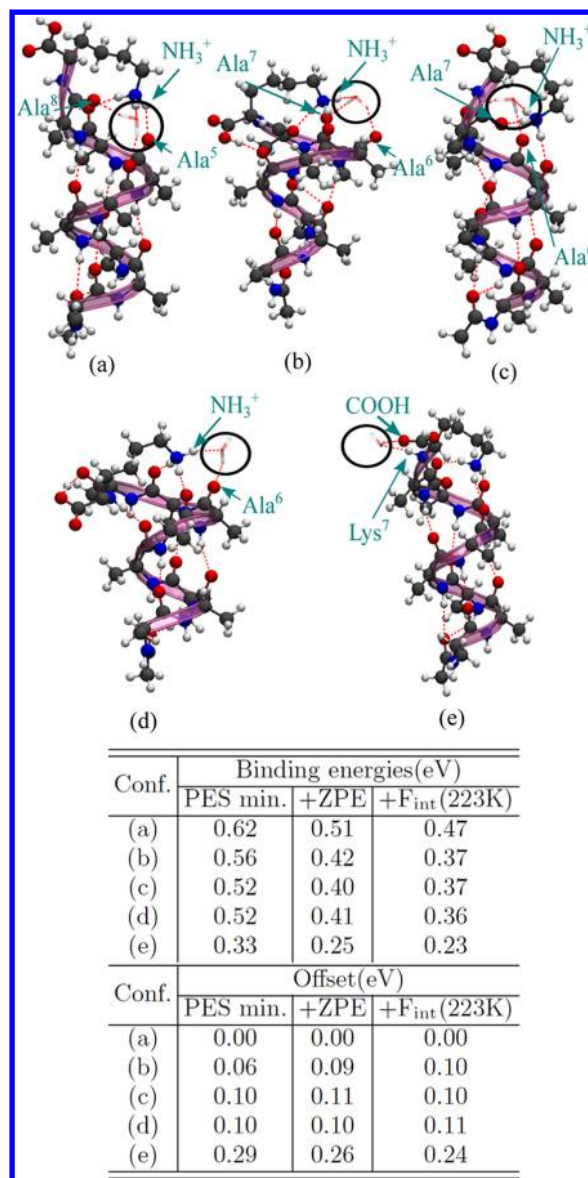
**3.2. Monohydration of Ac-Ala<sub>8</sub>-LysH<sup>+</sup>.** We next investigate the conformational space of the monohydrated peptide Ac-Ala<sub>8</sub>-LysH<sup>+</sup> using the methods outlined in section 2. The force field part of our exhaustive unconstrained conformer search yielded a total of 271 959 candidate conformers, out of which 147 were fully relaxed in DFT.

Figure 8 shows the lowest-energy monohydrated conformers. Two of these conformers (subfigures a and c, the lowest and third-lowest energy conformer) retain the helical structure of the unsolvated peptide.<sup>47</sup> In both cases, H<sub>2</sub>O forms three H-bonds: Two to carbonyl groups of the backbone, and one to the ammonium group. The latter, in turn, retains only two self-solvating H-bonds to the backbone—the third is broken by H<sub>2</sub>O. What is switched is the exact bond pattern of the backbone carbonyls. In conformer a, H<sub>2</sub>O is connected to residues 5 and 8, while NH<sub>3</sub><sup>+</sup> connects to 5 and 6. In conformer c, the connection is exactly opposite. The H<sub>2</sub>O location in conformer a is exactly the same as in the lowest-energy conformer of  $\alpha$ -2-helical Ac-Ala<sub>5</sub>-LysH<sup>+</sup> (Figure 5d). We suggest that this lowest-energy site could be generic for helical Ac-Ala<sub>n</sub>-LysH<sup>+</sup>.

Conformers b and d are not purely  $\alpha$ -helical. Both share the feature of a  $\pi$ -helical hydrogen bond loop right next to the C-terminus, as well as a carboxyl group that is hydrogen-bonded to a backbone carbonyl. Conformer b has a bifurcated  $\alpha$ -3<sub>10</sub>-helical bond next and is otherwise 3<sub>10</sub>-helical toward the N-terminus. Conformer d is  $\alpha$ -helical except for the  $\pi$ -helical bond near the C-terminus. What both conformers have in common is again a breaking of a self-solvation ammonium hydrogen bond by H<sub>2</sub>O. An attachment of H<sub>2</sub>O to the carboxyl group yields a much less favorable binding energy—the lowest-energy conformer with H<sub>2</sub>O in this location is shown in subfigure e.

To complete our assessment of potential H<sub>2</sub>O locations, we again attempted to systematically break each hydrogen bond of the unsolvated Ac-Ala<sub>8</sub>-LysH<sup>+</sup> helix in order to insert H<sub>2</sub>O, as done in section 3.1.5 for Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. Indeed, it is much less favorable or even impossible to insert H<sub>2</sub>O into one of the actual  $\alpha$ -helical backbone H-bonds (at least not without disrupting the remaining H-bond network as well). Thus, the actual helical part of the peptide is quite stable against disruption by the solvent, a degree of stability that most likely helps protect helices in solution as well.

We next return to the comparison of calculated and experimentally measured  $\Delta G_1^0(T)$  in Table 2. Just as for Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, the agreement for Ac-Ala<sub>8</sub>-LysH<sup>+</sup> is remarkable at the temperatures of both earlier experiments ( $T = 223$  K<sup>13</sup> and  $T = 260$  K<sup>14</sup>). For both molecules, the calculated  $\Delta G_1^0(T)$  values are slightly lower (stronger binding) by approximately



**Figure 8.** (a–d) Lowest-energy monohydrated conformers of Ac-Ala<sub>8</sub>-LysH<sup>+</sup>. (e) Lowest-energy identified conformer with H<sub>2</sub>O connected to the carboxyl group. Table: Upper part: Binding energies at the PES, zero-point corrected binding energies, and binding free energies at  $T = 223$  K are reported for each conformer shown. Lower part: Energy offset of each conformer with respect to the lowest-energy one shown in part a.

0.04 eV (1 kcal/mol), which could, for instance, reflect a slight systematic overestimation of the H<sub>2</sub>O binding energy by the PBE+vdW functional. We also see that the experimental extrapolation to  $T = 0$ <sup>14</sup> does not match the theoretically predicted zero-point corrected water adsorption energy (PBE+vdW: 0.51 eV), exactly as in the case of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. As already stated in section 3.1.3, in our view, the magnitude of the discrepancy is too large to be explained by a systematic error of the theory at  $T = 0$ .

We can now pinpoint the reason for the experimentally observed decrease of the water adsorption propensity at Ac-Ala<sub>8</sub>-LysH<sup>+</sup> (helical) compared to Ac-Ala<sub>5</sub>-LysH<sup>+</sup> (non-helical). Remarkably, the calculated  $\Delta G_1^0(T)$  difference between water adsorbed at Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and at Ac-Ala<sub>8</sub>-LysH<sup>+</sup>, also 0.04 eV, matches *exactly* the difference needed to explain the decrease in

H<sub>2</sub>O adsorption propensity seen in Figure 2 of ref 13. For both peptides, the H<sub>2</sub>O binding energy at the PES (no zero-point corrections) is exactly the same, 0.62 eV (compare Figures 4 and 8). The pressure dependence through  $F_{\text{trans}}$  is practically the same. The entire difference thus results from the internal (vibrational and rotational) parts of the free energies of both peptides. On the basis of this observation, we come to the conclusion that neither a pronounced adsorption site difference<sup>13</sup> nor the electrostatics of the growing helix dipole<sup>14</sup> are responsible for the decreasing H<sub>2</sub>O adsorption propensity for Ac-Ala<sub>*n*</sub>-LysH<sup>+</sup> with increasing length *n*. Instead, the subtle changes of the vibrational energy and entropy upon H<sub>2</sub>O adsorption at the C terminus of a helix are sufficient to explain the experimental difference quantitatively, at least for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup>.

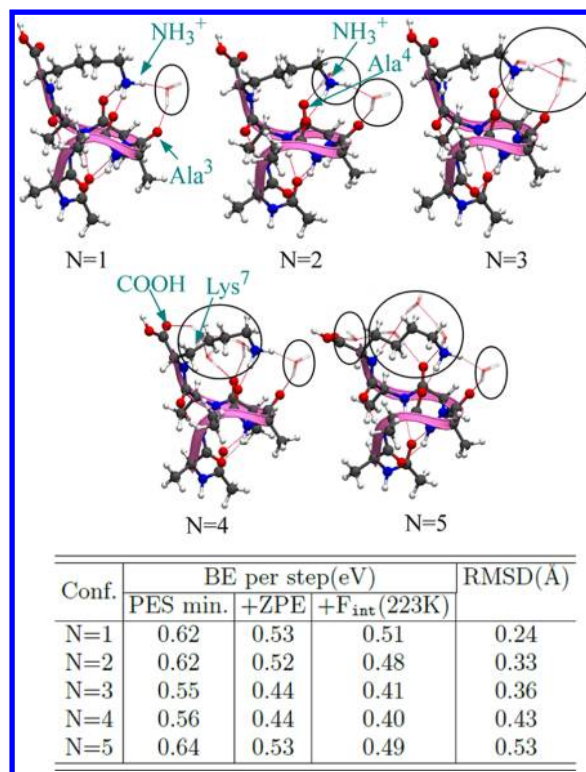
**3.3. Oligo-Hydration of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>: 2–5 Adsorbed H<sub>2</sub>O Molecules.** The first H<sub>2</sub>O molecule adsorbed at Ac-Ala<sub>5</sub>-LysH<sup>+</sup> breaks a self-solvating intramolecular H-bond of the ammonium termination (sections 3.1.3 and 3.1.4). The key questions during the next steps of hydration (addition of further H<sub>2</sub>O molecules) are

- (1) Does the lowest-energy conformation of the peptide backbone change, or does it remain the unsolvated ground state, *g*-1?
- (2) Is the original ammonium self-solvation by H-bonds to the backbone completely lifted?
- (3) Does the carboxyl group play a role as a lowest-energy adsorption site?
- (4) At which point does the formation of the second solvation shell set in, in which a H<sub>2</sub>O molecule is no longer directly H-bonded to the peptide?

The details of the conformational searches are the same as in the case of monohydration: water molecules are placed randomly around the isolated *g*-1 peptide to create input structures for unconstrained searches, where any peptide conformation is allowed, with the FF. Of the order of 10<sup>6</sup> structures were found, the lowest 200 of which for each hydration stage were relaxed within DFT in FHI-aims using light settings.

**3.3.1. *g*-1-Like Conformers.** Figure 9 summarizes the lowest energy microsolvated conformations of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> with *N* H<sub>2</sub>O molecules where *N* = 1–5, along with the positions where the waters bind. To answer question 1 above first: The optimum structure at each hydration stage retains the *g*-1 intramolecular H-bond pattern. Changes in the conformation of the *g*-1 peptide during the microhydration can be measured by the RMSD (eq 15) of the hydrated structure (all C, N, and O atoms) with respect to the isolated structure. The RMSD values are 0.24, 0.33, 0.36, 0.43, and 0.53 Å, respectively. Thus, there is only a very limited change of the ground state conformer except regarding the degree of the ammonium self-solvation.

The water binding energy per microsolvation step and the respective BEs corrected for zero point and free energies at 223 K are shown in the table in the bottom part of Figure 9. Notwithstanding the remarks above, there is no strong or systematic decrease of the stepwise binding energy (unlike, e.g., for methylamine in section 2.4<sup>12,14,55,56</sup> or for protonated<sup>3</sup> and neutral<sup>5,6</sup> amino acids). The microsolvation energetics found in our work is thus closer to that of the larger protonated peptide bradykinin, a molecule with several good hydration sites.<sup>17</sup>



**Figure 9.** Lowest energy microsolvated structures of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> with up to five water molecules, along with their respective binding energies per microsolvation step (PES local minima, ZPE corrected, and corrected for internal free energy at 223 K). The lowest energy structure in each case has the *g*-1 conformation. The RMSD values of hydrated peptide (all C, N, and O atoms) with respect to the nonhydrated *g*-1 conformer are also shown.

In the doubly hydrated peptide, the second water molecule is inserted between H(NH<sub>3</sub><sup>+</sup>) and O(Ala<sup>4</sup>) by breaking another intramolecular bond. Thus, only a single ammonium peptide bond remains (the one involving O(Ala<sup>2</sup>)). The stepwise solvation energy in this second step is the same as for the first H<sub>2</sub>O—apparently, both molecules benefit similarly from their attachment to the ammonium site. The triply hydrated structure is very similar to the *N* = 2 structure, but there is a structural rearrangement of the water molecules so that they are also bonded to each other. Upon the addition of the fourth water molecule, the COOH site is hydrated for the first time, accompanied by another rearrangement of the water molecules. Two waters are now bonded to O(Ala<sup>4</sup>), one forming a bridge with H(NH<sub>3</sub><sup>+</sup>) and the other with H(Lys<sup>7</sup>). Finally, the fifth H<sub>2</sub>O molecule does not bind to the peptide but only to the remaining water molecules. It thus reflects the onset of a formal second solvation shell, as far as lowest-energy conformations are concerned. Interestingly, the stepwise binding energy in this step increases over the preceding steps, indicating that the *N* = 3 and *N* = 4 solvation steps (which bridges to the carboxyl group) could not form ideal H-bond networks. Throughout these steps, one hydrogen in NH<sub>3</sub><sup>+</sup> is always bonded in an intramolecular bond to O(Ala<sup>2</sup>); i.e., the self-solvation of the ammonium group is not completely lifted. This observation is consistent with the already significantly reduced monohydration energy of this site in section 3.1.5 (BE, PES: 0.44 eV compared to 0.62 eV for the optimum position, Figure 6). In general, the order of the first four hydration steps appears to reflect the energetic ordering of the comparable monohydration

binding sites. The latter is thus, to some extent, predictive, as seen for microsolvated amino acids by previous authors<sup>39,40</sup>—until the formation of an actual H<sub>2</sub>O–H<sub>2</sub>O H-bond network sets in.

As in the case of the monohydrated peptides, we have again attempted to cross-check the protonation state of the carboxyl group—in the present case, by short (~18 ps) AIMD trajectories, beginning from the lowest energy conformers with four and five H<sub>2</sub>O molecules, respectively. In these trial runs, the proton remains attached to the carboxyl group, as in our extended conformational searches.

**3.3.2.  $\alpha$ -Helix-Like Conformers.** We find the  $\alpha$ -helical conformers by singling out all peptides with  $\alpha$ -1 and  $\alpha$ -2 patterns among the FF conformers in the unconstrained searches and relaxing them in PBE+vdW. The lowest  $\alpha$ -helical conformer found from the unconstrained microsolvation searches with (H<sub>2</sub>O)<sub>N</sub>, N = 2–5, is  $\alpha$ -2. It is 0.10 eV above g-1 in the PES, as in the unsolvated case.<sup>46</sup> The positions of the water molecules for each N are shown in Figure 10 along with the BE per microsolvation step. They faithfully follow the trend set up by monohydration. The first water molecule binds

directly to the charged ammonium group at the C-terminus and O(Ala<sup>2</sup>) and O(Ala<sup>5</sup>) by disrupting a single intramolecular H-bond. The second water molecule is inserted between a second H(NH<sub>3</sub><sup>+</sup>) and O(Ala<sup>3</sup>) by breaking another intramolecular H-bond. The third water molecule is inserted between the remaining H(NH<sub>3</sub><sup>+</sup>) and O(Ala<sup>4</sup>), thus breaking the intramolecular H-bond network around the charged ammonium group completely. Thus, in the case of a helix, it is apparently significantly easier than for the more compact g-1 conformer to lift the entire self-solvation of the ammonium group by inserting H<sub>2</sub>O molecules.

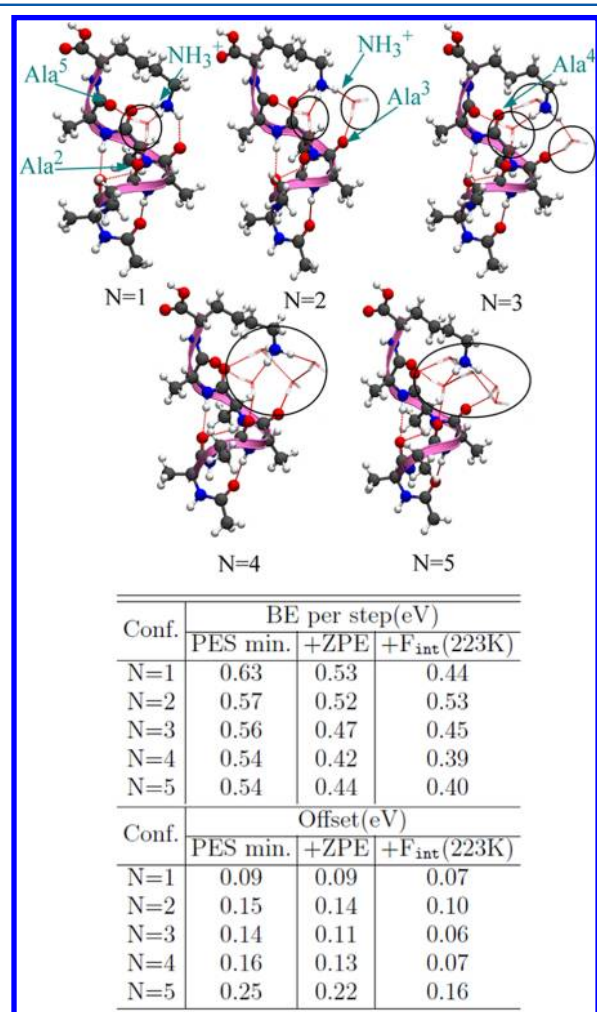
The fourth and fifth water molecules bind to the water molecules already present around the ammonium group, and no further intramolecular bonds are broken. A hydration of the carboxyl group is evidently less favorable than the insertion of H<sub>2</sub>O molecules near the already detached and now flexible lysine arm with the charged ammonium group. With the exception of the first (highest) stepwise hydration energy, all others are of similar magnitude. In total, the case of  $\alpha$ -2 suggests that lifting the self-solvation of the ammonium group by H<sub>2</sub>O will be a general trend for all helical conformers of Ac-Ala<sub>n</sub>-LysH<sup>+</sup> (the lowest-energy unsolvated state for n ≥ 7<sup>46,47,49</sup>), since their C-terminus structures are generally similar. It is the peculiar, more compact bonding pattern of the g-1 conformer that keeps the ammonium group at least partially self-solvated.

In summary, the answers to our questions 2, 3, and 4 above are conformer-dependent. For the lowest-energy conformer, g-1, the ammonium self-solvation is not completely lifted by up to 5 H<sub>2</sub>O molecules, the carboxyl group is partially solvated, and a “second solvation shell” forms. For the  $\alpha$ -2 conformer, at least, the answers are exactly the other way around.

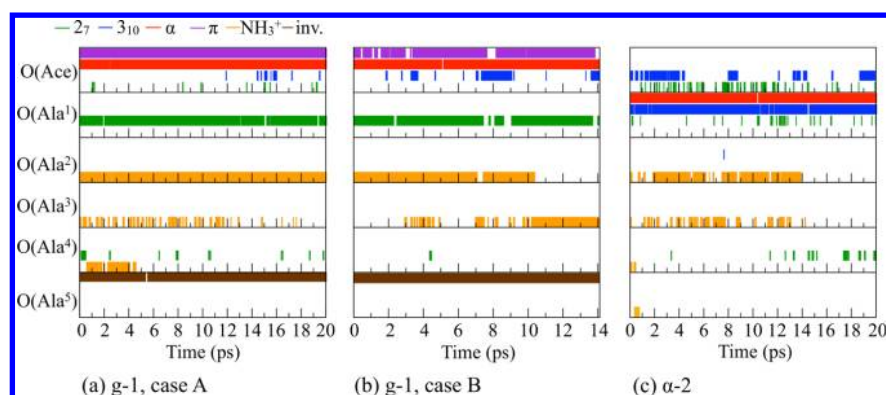
**3.4. *Ab Initio* Molecular Dynamics in a Droplet: Toward Solvation.** A key question is how the insights gained from a microsolvation study can help to understand similar peptide structural motifs in a full solvent environment. In the present section, we connect to a more solvent-like, dynamic environment of 152 H<sub>2</sub>O molecules for Ac-Ala<sub>5</sub>-LysH<sup>+</sup> by short (up to 20 ps) AIMD trajectories (T ≈ 300 K). Importantly, these simulations do not reflect a bulk solvent environment. 152 H<sub>2</sub>O molecules correspond more closely to a water nanodroplet. For instance, we observe a tendency of the water droplet to desolvate the N terminus of the peptide and to aggregate around the positively charged C terminus instead. Also, the time scale is not sufficient for any statistically significant statements regarding the peptide conformation.

Still, the exploratory AIMD trajectories described below allow us to make some key connections between the microsolvated case and the expected behavior in a bulk solvent. What we see is the following:

- The short-term stability of the self-solvated ammonium group is indeed conformation-dependent and different for g-1 and  $\alpha$ -2.
- In both cases, we expect the ammonium group to eventually become detached from the backbone carbonyl groups.
- The proton at the carboxyl group is detached almost immediately and becomes solvated in H<sub>2</sub>O, as would be expected for bulk water at neutral pH and ambient temperature and pressure. This is a major difference to the oligo-hydrated case.



**Figure 10.** Lowest-energy microsolvated conformers for one to five adsorbed H<sub>2</sub>O molecules at the  $\alpha$ -2 conformer of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. Table: Upper part: Step-wise binding energies (PES minimum energy value, ZPE corrected, and including internal free energy contributions at 223 K). Lower part: Energy differences of the depicted conformers to the global minimum g-1 conformer at each hydration stage.

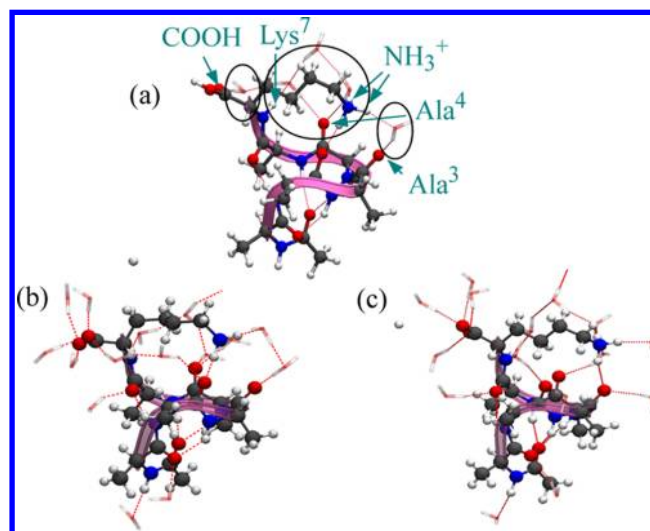


**Figure 11.** Evolution of the intramolecular hydrogen bond networks of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> for three different starting structures in an aggregate of 152 H<sub>2</sub>O molecules. The colored sticks indicate the presence or absence of hydrogen bonds at each carbonyl group, labeled by residues beginning from the N terminus (Ace). Only hydrogen bonds to NH groups are counted. The “presence” of a hydrogen bond is indicated if the CO–HN distance is less than 2.5 Å. Different colors indicate different types of hydrogen bonds (2<sub>7</sub> loop, 3<sub>10</sub> loop, 4<sub>13</sub> loop ( $\alpha$ ), 5<sub>16</sub> loop ( $\pi$ ), or the “inverted” Ala<sup>5</sup>-Ala<sup>2</sup> connection found in the g-1 conformer). Starting points: (a) Isolated g-1 conformer, surrounded by 152 H<sub>2</sub>O molecules (case A). (b) Pentahydrated g-1 conformer, surrounded by an additional 147 H<sub>2</sub>O molecules (case B). (c) Isolated  $\alpha$ -2 conformer, surrounded by 152 H<sub>2</sub>O molecules.

We also find that the specific H<sub>2</sub>O placements in the microsolvated lowest-energy structures are *not* rigid in the dynamic environment; i.e., they do not correspond to what could be called “structural” water sites.

**3.4.1. Ac-Ala<sub>5</sub>-LysH<sup>+</sup>: g-1 Conformer.** We consider two cases for the g-1 conformer of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. In the first case A, we use the isolated nonhydrated conformer and surround it with an aggregate of 152 water molecules. The simulation is run for 20 ps. In the second case B, we start with the (g-1)·(H<sub>2</sub>O)<sub>5</sub> structure found during microsolvation, surrounded by an additional 147 water molecules, making again a total of 152 water molecules in the aggregate. We run the simulation for 14 ps in this case. Figure 11 shows the evolution of the intramolecular H-bonds of the carbonyl oxygen atoms during the two AIMD simulations. The presence or absence of a stick indicates whether or not a given intramolecular H-bond exists at a given point in time at a given carbonyl residue. Relevant types of hydrogen bonds are a seven-membered ring (2<sub>7</sub>), a ten-membered ring (3<sub>10</sub>), a thirteen-membered ring ( $\alpha$ ), a 16-membered ring ( $\pi$ ), a connection between a backbone carbonyl group and the ammonium group (NH<sub>3</sub><sup>+</sup>), or the signature “inverted” hydrogen bond of g-1. The structure at the end of each run can be seen in Figure 12b and c, together with the microsolvated starting point structure for case B in Figure 12a. For both starting points A and B, the essential structure elements of g-1 (inverted H-bond at Ala<sup>5</sup>, bifurcated H-bond at Ace terminus, 2<sub>7</sub>-type bond at Ala<sup>1</sup>) remain intact.

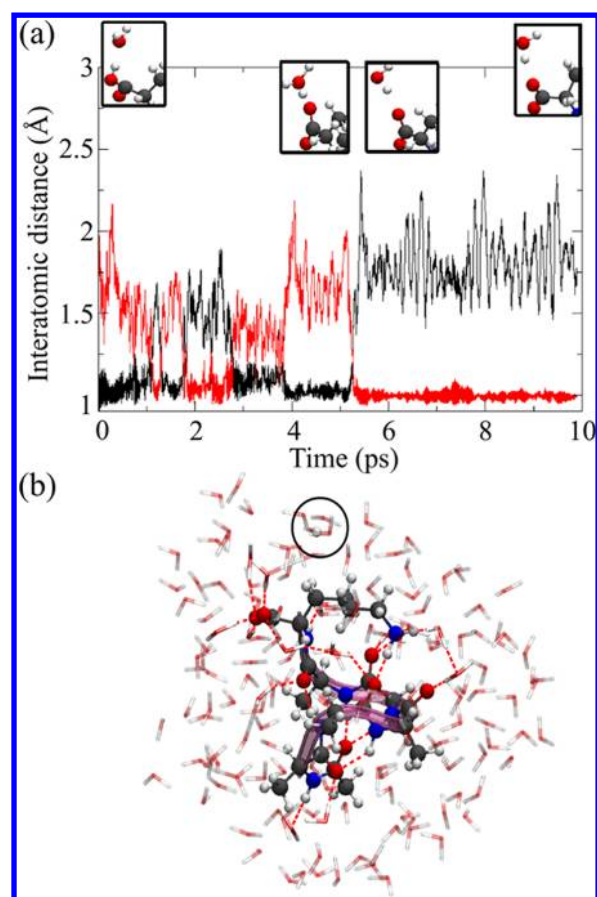
In Figure 11a (case A, unsolvated starting structure), we see how the H-bonds involving the ammonium group are broken to form hydration sites. The H-bond with O(Ala<sup>3</sup>) is broken first, followed by that with O(Ala<sup>4</sup>). The O(Ala<sup>2</sup>) H-bond, however, is not broken over the course of the simulation, consistent with the relatively high stability that we see in the microsolvation case. Figure 11b shows the evolution when starting from the microsolvated structure (case B). Since the initial intramolecular H-bond network is similar to that developed at the end of run A, the two runs may well be viewed as approximately connected. In case B, however, the most stable self-solvating H-bond of the ammonium group O(Ala<sup>2</sup>) is disrupted. At the end of the run, the ammonium group is only connected to the backbone by a single proton. We may thus expect that a longer run would eventually detach the ammonium group from the backbone carbonyls completely.



**Figure 12.** Comparison of the first solvation shells in Ac-Ala<sub>5</sub>-LysH<sup>+</sup> in (a) the microsolvated (g-1)·(H<sub>2</sub>O)<sub>5</sub> structure at 0 K, (b) case A, (g-1)·(H<sub>2</sub>O)<sub>152</sub> after 20 ps of AIMD, and (c) case B, the microsolvated (g-1)·(H<sub>2</sub>O)<sub>5</sub> conformer in an aggregate of 147 water molecules after 14 ps of AIMD at 300 K.

The primary conclusion is that the self-solvation of the ammonium group is indeed disrupted by the droplet at finite  $T$ , irrespective of the presence of the particularly stabilizing conformation g-1. Since the bonding pattern changes over time, we also do not see the same individual H<sub>2</sub>O molecules as in the microsolvated case reside in the broken intramolecular H-bonds.

Perhaps the most important conceptual difference to the microsolvated case is that the carboxyl group loses its proton almost immediately in the AIMD simulation of the 152-molecule water droplet. In case A, the proton detaches itself already after  $\sim 5$  ps. In case B, the deprotonation occurs after just below 10 ps. In both simulations, what is left behind is essentially the neutral, zwitterionic form of the peptide, and one extra proton that is effectively solvated by the surrounding H<sub>2</sub>O. An illustration is shown in Figure 13 for case A. Subfigure a shows the time evolution of the first proton jumps from the carboxyl to a nearby H<sub>2</sub>O molecule, indicated by the respective H–O distances over the course of the simulation. Subfigure b



**Figure 13.** Proton transfer between COOH and the H<sub>2</sub>O aggregate during the AIMD simulation of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> surrounded by 152 H<sub>2</sub>O molecules (case A). (a) H–O interatomic distances as a function of simulation time. Black: (COO)–proton distance. Red: Distance between proton and the H<sub>2</sub>O molecule where it eventually becomes attached. (b) Final geometry of the entire cluster after a simulation time of 20 ps. The location of an excess proton after several further proton jumps between H<sub>2</sub>O molecules is marked by a circle.

shows the final cluster geometry of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>. The excess proton is not the same as the one detached from the carboxyl group but rather a different one after several proton jumps between H<sub>2</sub>O molecules. Qualitatively, we have thus crossed the threshold between the few-water microsolvated gas-phase peptide and a protonation behavior that is closer to what is expected in bulk H<sub>2</sub>O at ambient temperature, pressure, and neutral pH.

**3.4.2. Ac-Ala<sub>5</sub>-LysH<sup>+</sup>:  $\alpha$ -2 Conformer.** The 20 ps AIMD simulation starting from the  $\alpha$ -2 conformer of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> was carried out by surrounding the isolated nonhydrated conformer with an aggregate of 152 water molecules. Figure 11c shows the time evolution of the intramolecular hydrogen bonds during simulation at 300 K. The bifurcated backbone H-bond attached to the carbonyl group at Ala<sup>1</sup> remains stable over the entire simulation, indicating that the essence of the  $\alpha$ -2 conformation remains intact. Consistent with the lowest-energy microsolvated conformations, the self-solvating H-bonds or the ammonium group are all disrupted within a short period of time, 14 ps in Figure 11c. Unlike in the g-1 case, the microsolvation study does not indicate any particular stability of one of these bonds, and this is reflected here.

Similar to the g-1 case, the protonation state of the carboxyl group changes. There is a spontaneous proton transfer between

the COOH group and one of the surrounding water molecules at around 4 ps. The proton is later recaptured.

For all three AIMD simulations, we emphasize again that the solvent shell which we have created here is not yet bulk H<sub>2</sub>O. Also, the observed deprotonation and the disruption of the ammonium self-solvation are not necessarily ground state ( $T = 0$ ) effects. Finally, our observations are derived from isolated, very short AIMD trajectories. Nonetheless, the trajectories are not without value: They illuminate the path from the low-energy oligo-hydrated conformers to the expected phenomena in a solvent. The deprotonation of the carboxyl group must happen somewhere between a few H<sub>2</sub>O molecules at low  $T$  and many H<sub>2</sub>O molecules at approximately ambient temperature. Pinpointing the exact crossover point and mechanism (entropy or enthalpy) in equilibrium as a function of water cluster size and  $T$  would be an extremely sensitive gauge for our current quantitative grasp of peptides in an environment, both in experiment and theory.

#### 4. CONCLUSION

The polypeptides Ac-Ala<sub>5</sub>-LysH<sup>+</sup> and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> are well studied benchmark systems in gas-phase experiments<sup>48–50</sup> and theory,<sup>46,47</sup> particularly with respect to microsolvation.<sup>13,14</sup> Their hallmark is the possible competition for hydration of the protonated C-terminal carboxyl site, and the fully self-solvated, protonated ammonium group of the lysine residue. Our first-principles analysis reveals the preferred monohydration sites for both peptides, the lowest-energy oligo-hydration conformers of Ac-Ala<sub>5</sub>-LysH<sup>+</sup>, and a connecting path to a solvating environment of 152 water molecules.

Our primary conclusions are as follows:

(1) The intramolecular hydrogen bonds of the self-solvated ammonium group are the most stable hydration sites.

(2) The carboxyl group is not a strongly preferred hydration site. It is somewhat competitive and hydrated by the fourth adsorbed H<sub>2</sub>O for the g-1 conformer of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> but not at all competitive for the helical conformers.

(3) We achieve excellent agreement between calculated and experimental values of  $\Delta G_1^0(T)$  (equivalent to equilibrium constants) at  $T = 223$  K and  $T = 260$  K. An overestimation of the water binding energy by  $\sim 0.04$  eV (1 kcal/mol) is seen. We also capture quantitatively the experimentally observed decrease in H<sub>2</sub>O adsorption propensity at Ac-Ala<sub>5</sub>-LysH<sup>+</sup> (non-helical) and Ac-Ala<sub>8</sub>-LysH<sup>+</sup> (helical). According to our calculations, the decrease is entirely due to modified internal free energy contributions (harmonic vibrational free energy) due to the specific H<sub>2</sub>O adsorption site at the LysH<sup>+</sup> termination.

(4) There is a strong dependence of the actual mono- and oligohydration patterns on the peptide conformation but not so much of the hydration energetics. The peculiar H-bond pattern of the g-1 conformer of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> lends significant stability to one of the self-solvating ammonium H-bonds, and thus to the termination as a whole.

(5) For helical conformers, the self-solvation of the ammonium group succumbs to hydration already for three adsorbed H<sub>2</sub>O molecules; i.e., all three of its H-bonds are broken first.

(6) Exploratory AIMD simulations of Ac-Ala<sub>5</sub>-LysH<sup>+</sup> with 152 water molecules indicate the connections between the microsolvated state and the more fully solvated conformers:

(a) The relatively larger stability of the ammonium self-solvation for g-1 than for the helical conformer is reflected in

the simulations. Ultimately, the ammonium self-solvation is most likely lifted in both cases.

(b) The crossover from the protonated to the deprotonated state of the carboxyl group is readily observed in the presence of many H<sub>2</sub>O but not in tests for the microsolvated state up to only five H<sub>2</sub>O molecules. Pinpointing this crossover precisely as a function of the number of H<sub>2</sub>O molecules and temperature would be a valuable benchmark for present experiments and theory alike.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The structures (Cartesian coordinates) of all molecular conformations shown and discussed in our work are provided as Supporting Information in xyz format. Any structure is thus available for easy visualization with standard molecular viewing programs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [blum@fhi-berlin.mpg.de](mailto:blum@fhi-berlin.mpg.de).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We gratefully acknowledge many helpful and constructive discussions with Dr. Carsten Baldauf and Dr. Gert von Helden, both at Fritz Haber Institute (FHI) of the Max Planck Society. Likewise, we thank Prof. Matthias Scheffler (FHI) for his continued support during this research.

## ■ REFERENCES

- (1) Rožman, M.; Srzić, D.; Klasinc, L. *Int. J. Mass Spectrom.* **2006**, *253*, 201–206.
- (2) Ebata, T.; Hashimoto, T.; Ito, T.; Inokuchi, Y.; Altunso, F.; Brutschy, B.; Tarakeshwar, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4783–4791.
- (3) Wincel, H. *Chem. Phys. Lett.* **2007**, *439*, 157–161.
- (4) Blom, M. N.; Compagnon, I.; Polfer, N. C.; von Helden, G.; Meijer, G.; Suhai, S.; Paizs, B.; Oomens, J. *J. Phys. Chem. A* **2007**, *111*, 7309–7316.
- (5) Gao, B.; Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2009**, *113*, 9995–10000.
- (6) Gao, B.; Wyttenbach, T.; Bowers, M. T. *J. Am. Chem. Soc.* **2009**, *131*, 4695–4701.
- (7) Prell, J. S.; Chang, T. M.; O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 7811–7819.
- (8) Fricke, H.; Schwing, K.; Gerlach, A.; Unterberg, C.; Gerhards, M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3511–3521.
- (9) Biswal, H. S.; Loquais, Y.; Tardivel, B.; Gloaguen, E.; Mons, M. *J. Am. Chem. Soc.* **2011**, *133*, 3931–3942.
- (10) Zwier, T. S. *J. Phys. Chem. A* **2001**, *105*, 8827–8839.
- (11) Snoek, L. C.; Kroemery, R. T.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2130–2139.
- (12) Liu, D.; Wyttenbach, T.; Barran, P. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2003**, *125*, 8458–8464.
- (13) Kohtani, M.; Jarrold, M. F. *J. Am. Chem. Soc.* **2004**, *126*, 8454–8458.
- (14) Liu, D.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass Spectrom.* **2004**, *236*, 81–90.
- (15) Liu, D.; Wyttenbach, T.; Carpenter, C. J.; Bowers, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 3261–3270.

- (16) Carcabal, P.; Kroemer, R. T.; Snoek, L. C.; Simons, J. P.; Bakker, J. M.; Compagnon, I.; Meijerd, G.; von Helden, G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4546–4552.
- (17) Wyttenbach, T.; Liu, D.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 221–232.
- (18) Mercier, S. R.; Boyarkin, O.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.; Rothlisberger, U.; Rizzo, T. R. *J. Am. Chem. Soc.* **2006**, *128*, 16938–16943.
- (19) Kamariotis, A.; Boyarkin, O. V.; Mercier, S. R.; Beck, R. D.; Bush, M. F.; Williams, E. R.; Rizzo, T. R. *J. Am. Chem. Soc.* **2006**, *128*, 905–916.
- (20) Wyttenbach, T.; Bowers, M. T. *Chem. Phys. Lett.* **2009**, *480*, 1–16.
- (21) Stearns, J.; Seaby, C.; Boyarkin, O.; Rizzo, T. *Phys. Chem. Chem. Phys.* **2009**, *11*, 125–132.
- (22) Zhu, H.; Blom, M.; Compagnon, I.; Rijs, A. M.; Roy, S.; von Helden, G.; Schmidt, B. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3415–3425.
- (23) Schwing, K.; Reyheller, C.; Schaly, A.; Kubik, S.; Gerhards, M. *ChemPhysChem* **2011**, *12*, 1981–1988.
- (24) Mayorkas, N.; Rudić, S.; Cocinero, E. J.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18671–18678.
- (25) Simons, J. P.; Jockusch, R. A.; CarCabal, P.; Hüinig, I.; Kroemer, R. T.; Macleod, N. A.; Snoek, L. C. *Int. Rev. Phys. Chem.* **2005**, *24*, 489–531.
- (26) Ahn, D.-S.; Park, S.-W.; Jeon, I.-S.; Lee, M.-K.; Kim, N.-H.; Han, Y.-H.; Lee, S. J. *Phys. Chem. B* **2003**, *107*, 14109–14118.
- (27) Chaudhari, A.; Sahu, P. K.; Lee, S.-L. *J. Mol. Struct.: THEOCHEM* **2004**, *683*, 115–119.
- (28) Lee, K.-M.; Park, S.-W.; Jeon, I.-S.; Lee, B.-R.; Ahn, D.-S.; Lee, S. *Bull. Korean Chem. Soc.* **2005**, *26*, 909–912.
- (29) Leung, K.; Rempe, S. B. *J. Chem. Phys.* **2005**, *122*, 184506.
- (30) Aikens, C. M.; Gordon, M. S. *J. Am. Chem. Soc.* **2006**, *128*, 12835–12850.
- (31) Rodziewicz, P.; Doltsinis, N. L. *ChemPhysChem* **2007**, *8*, 1959–1968.
- (32) Chuchev, K.; BelBruno, J. J. *J. Mol. Struct.: THEOCHEM* **2008**, *850*, 111–120.
- (33) Bachrach, S. M.; Nguyen, T. T.; Demoin, D. W. *J. Phys. Chem. A* **2009**, *113*, 6172–6181.
- (34) Rai, A. K.; Fei, W.; Lu, Z.; Lin, Z. *Theor. Chem. Acc.* **2009**, *124*, 37–47.
- (35) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 632–638.
- (36) Mullin, J. M.; Gordon, M. S. *J. Phys. Chem. B* **2009**, *113*, 8657–8669.
- (37) Berhanu, W. M.; Mikhailov, I. A.; Masunov, A. E. *J. Mol. Model.* **2010**, *16*, 1093–1101.
- (38) Michaux, C.; Wouters, J.; Jacquemin, D.; Perpète, E. A. *Chem. Phys. Lett.* **2007**, *445*, 57–61.
- (39) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 7702–7705.
- (40) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 9896–9902.
- (41) Michaux, C.; Wouters, J. *J. Phys. Chem. B* **2008**, *112*, 2430–2438.
- (42) Ireta, J. *J. Chem. Theory Comput.* **2011**, *7*, 2630–2637.
- (43) Ireta, J. *Int. J. Quantum Chem.* **2012**, *112*, 3612–3617.
- (44) Marianski, M.; Dannenberg, J. J. *J. Phys. Chem. B* **2012**, *116*, 1437–1445.
- (45) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. *Science* **2012**, *336*, 320–323.
- (46) Rossi, M.; Blum, V.; Kupser, P.; von Helden, G.; Bierau, F.; Pagel, K.; Meijer, G.; Scheffler, M. *J. Phys. Chem. Lett.* **2010**, *1*, 3465–3470.
- (47) Rossi, M.; Blum, V.; Scheffler, M. Impact of vibrational entropy on the stability of unsolvated peptide helices with increasing length. 2012, arXiv:1208.6133v1. arXiv.org e-Print archive. <http://arxiv.org/abs/1208.6133v1> (accessed Dec 2, 2012).



- (48) Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974–12975.
- (49) Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 3494–3501.
- (50) Jarrold, M. F. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1659–1671 (see also references therein).
- (51) Kaleta, D. T.; Jarrold, M. F. *J. Phys. Chem. B* **2001**, *105*, 4436–4440.
- (52) Kohtani, M.; Jarrold, M. *J. Am. Chem. Soc.* **2002**, *124*, 11148–11158.
- (53) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (54) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.
- (55) Meot-Ner, M. *J. Am. Chem. Soc.* **1984**, *106*, 1265–1272.
- (56) Lau, Y. K.; Kebarle, P. *Can. J. Chem.* **1981**, *59*, 151–155.
- (57) Tkatchenko, A.; Rossi, M.; Blum, V.; Ireta, J.; Scheffler, M. *Phys. Rev. Lett.* **2011**, *106*, 118102.
- (58) Rossi, M. Ab initio study of alanine-based polypeptide secondary-structure motifs in the gas phase. Ph.D. thesis, Fritz-Haber-Institut der Max-Planck Gesellschaft, TU-Berlin, 2011; <http://opus.kobv.de/tuberlin/volltexte/2012/3429/>.
- (59) Santra, B. Density-Functional Theory Exchange-Correlation Functionals for Hydrogen Bonds in Water. Ph.D. thesis, Fritz-Haber-Institut der Max-Planck Gesellschaft, TU-Berlin, 2010; <http://opus.kobv.de/tuberlin/volltexte/2010/2804/>.
- (60) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (61) Ponder, J. TINKER - Software Tools for Molecular Design. <http://dasher.wustl.edu/tinker/>. In this work, we used versions 4.2 and 5.1 of the program and the force-fields' versions distributed within the package.
- (62) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
- (63) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377–395.
- (64) McQuarrie, D. A. *Statistical Mechanics*, 1st ed.; University Science Books: Sausalito, CA, 2000.
- (65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (66) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (67) Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (68) Havu, V.; Blum, V.; Havu, P.; Scheffler, M. *J. Comput. Phys.* **2009**, *228*, 8367–8379.
- (69) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (70) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6171.
- (71) Tkatchenko, A.; Robert A. DiStasio, J.; Head-Gordon, M.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 094106.
- (72) Xie, H.-B.; Jin, L.; Rudić, S.; Simons, J. P.; Gerber, R. B. *J. Phys. Chem. B* **2012**, *116*, 4851–4859.
- (73) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673–2677.