8-31-2017

Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation.

Lars Konermann

Molecular Dynamics Simulations on Gas Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation

Lars Konermann*

Department of Chemistry, The University of Western Ontario, London, Ontario, N6A 5B7, Canada

* To whom correspondence should be addressed. Telephone: (519) 661-2111 ext. 86313.

Email: konerman@uwo.ca

ABSTRACT: Molecular dynamics (MD) simulations have become a key tool for examining the properties of electrosprayed protein ions. Traditional force fields employ static charges on titratable sites, whereas in reality protons are highly mobile in gas phase proteins. Earlier studies tackled this problem by adjusting charge patterns during MD runs. Within those algorithms proton redistribution was subject to energy minimization, taking into account electrostatic and proton affinity contributions. However, those earlier approaches described (de)protonated moieties as point charges, neglecting charge solvation which is highly prevalent in the gas phase. Here we describe a mobile proton algorithm that considers the electrostatic contributions from all atoms, such that charge solvation is explicitly included. MD runs were broken down into 50 ps fixed-charge segments. After each segment the electrostatics were reanalyzed, and protons were redistributed. Challenges associated with computational cost were overcome by devising a streamlined method for electrostatic calculations. Avidin (a 504 residue protein complex) maintained a native-like fold over 200 ns. Proton transfer and side chain rearrangements produced extensive salt bridge networks at the protein surface. The mobile proton technique introduced here should pave the way towards future studies on protein folding, unfolding, collapse, and subunit dissociation in the gas phase.

Introduction

Electrospray ionization (ESI) allows the transfer of proteins and protein complexes from solution into the gas phase, thereby making them amenable to mass spectrometry (MS), ion mobility spectrometry, and related techniques.[1] "Native" ESI employs non-denaturing solutions and keeps collisional heating to a minimum.[2-9] These experiments aim to preserve solution-phase conformations and interactions *in vacuo*, exploiting the fact that electrosprayed proteins can get kinetically trapped in solution-like structures under gentle conditions.[10-12] However, the mechanistic basis of this trapping remains poorly understood. Also, it is not clear how long solution-like structures can survive, and how gaseous protein ions respond to perturbations.[13-17] For example, collisions with background gas can induce collapse, unfolding, and subunit ejection.[18-21] Protons in gas phase proteins are highly mobile, such that protein structural transitions will be accompanied by charge migration.[22-25]

The following discussion will focus on $[M + zH]^{z+}$ protein ions produced by positive ion ESI. A key factor that governs the behavior of these ions is the number of excess protons z. Low protonation states favor compact gas phase structures, while elevated z values promote electrostatically driven unfolding.[26, 27] It has often been assumed that $[M + zH]^{z+}$ ions are protonated at z basic sites (Lys, Arg, His, and N-termini), while all acidic sites (Asp, Glu, C-termini) are neutral (-COOH).[18, 28-31] However, recent studies strongly suggest that electrosprayed proteins contain zwitterionic moieties, where negative charges on acidic sites exist in salt bridges.[32-38]

Conventional structure determination methods such as X-ray crystallography and NMR spectroscopy are restricted to condensed-phase investigations. Atomistic insights into the conformations of macromolecules in the gas phase are difficult to obtain. Hence, molecular dynamics (MD) simulations have become an essential tool for probing the

structure and dynamics of gaseous proteins.[9, 11, 17, 30, 39-43] Typical MD force fields do not allow for the formation or dissociation of covalent bonds, rendering them incapable of dealing with proton migration. Considering the high proton mobility in electrosprayed ions,[22-25] the lack of mobile charges represents a significant limitation. Computational methods capable of incorporating mobile protons include QM/MM,[24, 44, 45] *ab initio* MD[25, 46-48] and DFT/MD.[49] Unfortunately, the computational cost of these advanced techniques precludes their application to large biomolecules. Simplified proton hopping algorithms exist, but most of these are geared towards solution simulations.[50, 51]

Thachuk et al.[31] proposed an interesting mobile proton model for gas phase MD simulations. Their method successfully described proton hopping during the dissociation of collisionally heated protein complexes, which culminated in the ejection of highly charged single subunits. Unfortunately, that study[31] only considered protonation of basic sites, not allowing for the existence of salt bridges.[32-37] Also, Thachuk's work[31] employed a coarse-grained force field. Coarse grained models use simplified representations of molecular structures, which are useful for studies on very large systems in bulk solution.[52] Gas phase protein simulations, however, are readily amenable to atomistic force fields that capture structural features in much greater detail.

Building on Thachuk's work,[31] we recently developed a mobile proton MD method that (i) employed an atomistic force field and (ii) allowed for protonation changes at all acidic and basic sites ($NT^{0/+}$, $Lys^{0/+}$, $Arg^{0/+}$, $His^{0/+}$, $Asp^{0/-}$, $Glu^{0/-}$, and $CT^{0/-}$, where NT and CT represent N- and C-termini). MD runs were broken down into brief fixed-charge segments. Protons were redistributed after each segment, and the subsequent segment was started with a newly optimized charge pattern. Proton migration was subject to minimization of the energy $E_{tot} = E_{Coul} + E_{PAint}$, where electrostatic interactions between N titratable sites with charges Q_k and Q_l and distances R_{kl} are given by

$$E_{Coul} = \frac{1}{4\pi\varepsilon_0} \sum_{k=1}^{N} \sum_{l=k+1}^{N} \frac{Q_k \times Q_l}{R_{kl}} \qquad (1)$$

The second contribution to E_{tot} arises from the intrinsic proton affinities ($PA_{int}(x) > 0$) of the $x = 1 \ldots N$ titratable sites, with $\delta_x = 1$ for protonated sites, and $\delta_x = 0$ for deprotonated sites.

$$E_{PA\,\mathrm{int}} = -\sum_{x=1}^{N} PA_{\mathrm{int}}(x) \times \delta_x \qquad (2)$$

Although Thachuk's[31] and our[53] earlier models performed quite well, they both suffer from conceptual deficiencies. Equation 1 represents a point charge approximation that assigns a charge of +1, 0, or -1 to one single atom within each titratable residue. Non-titratable residues are ignored (Figure 1A). In reality, all residues exhibit complex patterns of atomic charges, regardless of their protonation state (Figure 1B). The description of such patterns using a point charge approximation with simplistic R_{kl} distances is problematic. More importantly, both approaches[31, 53] neglect intramolecular charge solvation effects which are highly prevalent in macromolecular gas phase ions.[16, 29, 54]

To address the aforementioned deficiencies it would be desirable to extend the electrostatic expression of equation 1 from a few point charges ($N \approx 100$) to all the thousands of atoms that constitute a protein. The computational cost associated with the N^2 scaling of equation 1 renders such all-atom treatments challenging, keeping in mind that E_{Coul} has to be calculated countless times during a simulation run.

This work proposes an approach that addresses the issues outlined above. We develop a method for redistributing protons in gas phase proteins by minimizing E_{tot} during MD runs. The algorithm includes electrostatic contributions from all atoms, thereby treating solvation effects in a comprehensive fashion. Streamlined electrostatic calculations keep the computational cost low, despite increasing the number of charge sites by two orders of

magnitude. We will first examine aspects related to proton affinity and salt bridge stability. Then it will be illustrated how the algorithm can be applied to mobile proton simulations on avidin, a protein complex that has been widely used as ESI-MS model system.[20, 55]

Methods

Mobile Proton Simulations: General Concepts. While typical MD force fields are optimized for solution conditions, several of them have also been shown to perform well for gas phase protein simulations.[9, 11, 17, 30, 39-43, 53, 56-58] The current work employs the OPLS all-atom (OPLS/AA) force field[59] which is one of the most commonly used choices for gas phase applications.[40, 41, 53, 56-58] Also, OPLS/AA contains parameters for all titratable sites in their charged and neutral forms, which is an essential feature for the algorithm used here.[40] The X-ray coordinates 5chk[60] served as starting structure for simulations on avidin. Missing residues (Ala1, Arg2, Arg124, Thr125, Gln126) were added using Pymol (Schrödinger), thereby generating a symmetric homo-tetramer. We employed the Gromacs 5.1 MD suite of programs[61] with GPU acceleration in a vacuum environment without cut-offs for electrostatic or Lennard-Jones interactions. The integration step length was 2 fs, and the temperature was controlled using the Nosé-Hoover thermostat with a coupling constant of 0.1 ps.[62] Neutral His was modeled as Nε2-H tautomer.[53, 63]

Protein X-ray structures were initially subjected to steepest descent energy minimization, followed by 100 ps of equilibration at 1 K. Subsequent 200 ns production runs were conducted at 300 K. Each of these runs was broken down into 50 ps segments, in accordance with typical proton hopping times.[46, 53] After each 50 ps segment the electrostatics were reanalyzed in accordance with the current conformation, and the protein was subjected to proton redistribution using the algorithm described below. This was followed by energy

minimization to mitigate potential clashes associated with the new charge pattern, prior to the subsequent 50 ps segment. New velocities (with a new pseudo-random seed) were assigned from a Maxwell-Boltzmann distribution at the start of each segment, thereby eliminating artifacts associated with the accumulation of angular momentum. The alternation between proton redistribution and MD segments was coordinated by a bash script.

The mobile proton model used here relies on the premise that protons are highly mobile in gas phase protein ions,[22-25] and that these H^+ will redistribute spontaneously such that the energy E_{tot} is minimized.[18, 31, 53] We employed a steepest descent search algorithm that systematically executes single-proton moves involving termini, as well as all acidic and basic side chains. The search commences with an occupied protonation site; the proton from this site is tentatively moved to all possible vacant sites. A new configuration is retained if one of these moves produces a lower E_{tot}. The search algorithm then turns to another occupied site and tentatively moves its proton to all possible acceptor sites, and so on. Once each occupied site has been examined the cycle is repeated. To eliminate any possible bias the order of donor and acceptor sites is chosen at random for each cycle. The repetition ends once a complete cycle does not produce a lower E_{tot}. For avidin this convergence criterion was typically reached after ~20000 single proton moves.

Electrostatic Calculations. The key novel aspect implemented in this work is a treatment of electrostatic interactions that goes beyond the point charge approximation of equation 1. Instead, we consider the overall energy

$$E_{tot} = V_{Coul} + E_{PAint} \qquad (3)$$

where V_{Coul} includes the contributions from all n atoms in the protein. The goal of the all-atom mobile proton algorithm is to identify the charge pattern that minimizes E_{tot}. We define

$$v_{ij} = \frac{1}{4\pi\varepsilon_0} q_i q_j / r_{ij} \qquad (4)$$

which reflect the electrostatic interaction between two atomic charges q_i and q_j that are separated by distance r_{ij}. The corresponding overall electrostatic energy is

$$V_{Coul} = \sum_{i=1}^{n} \sum_{\substack{j=i+1 \\ I \neq J}}^{n} v_{ij} \qquad (5)$$

where the indices i, j signify individual atoms. I and J refer to residues, regardless whether they are titratable or not. $I = J$ terms are treated by subtracting PA_{int} values for those residues that are protonated (as in equation 2).[24,58]

The computational cost of equations 4 and 5 would render mobile proton simulations prohibitively slow for large proteins, where electrostatics have to be calculated for ten thousands of charge patterns for each of the thousands of different conformations. We developed a strategy that allows such repetitive V_{Coul} calculations to be performed in a manner that is much more efficient. This new method starts with calculation of the hypothetical energy W_{Coul}, which represents the sum of all conceivable electrostatic interactions that might be encountered for arbitrary charge patterns, taking into account the protonated and de-protonated forms of all titratable sites. For calculating W_{Coul} it is conceptually assumed that the protein simultaneously possesses two versions of each residue. For non-titratable residues every twinned pair has identical atom positions and atomic charges. For residues bearing a titratable site the two twins also share the same atomic positions, but they differ in their atomic charges, *i.e.*, the protonated form ("prot") is used for one twin, and the deprotonated form ("deprot") for the other one. Atomic charges q_i for each atom are defined in the OPLS/AA force field.[59] Matrix calculations are facilitated by assuming that the extra hydrogen of a protonated residue is also present in its deprotonated

twin. These dummy hydrogens carry zero charge. The protein sequence is then duplicated, *i.e.*, for a protein consisting of K residues one obtains a virtual $2K$ sequence that comprises all protonated residues and their deprotonated twins: residue$_1$(prot) ~ residue$_2$(prot) ~ … ~ residue$_K$(prot); residue$_{K+1}$(deprot) ~ residue$_{K+2}$(deprot) ~ … ~ residue$_{2K}$(deprot). This sequence corresponds to a string of $2n$ atom coordinates and charges (x_i, y_i, z_i, q_i), yielding

$$W_{Coul} = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{2n} \sum_{j=i+1}^{2n} v_{ij} \qquad (6)$$

Figure S2A illustrates how W_{Coul} is computed for Gly-Gly-Gly. While these calculations are in progress one also determines the matrix elements

$$V_{IJ} = \frac{1}{4\pi\varepsilon_0} \sum_{i \in I} \sum_{j \in J} v_{ij} \qquad (7)$$

which describe the electrostatic interactions between atoms in residues I and J. Both I and J cover the range from 1 to $2K$. In case of atom overlap ($r_{ij} = 0$) and for interactions involving a dummy proton we define $v_{ij} = 0$. Note that $W_{Coul} = \Sigma\Sigma V_{IJ}$.

After each 50 ps MD segment, W_{Coul} and all of the V_{IJ} terms are calculated *once*, and they remain stored in memory for ready access. The actual Coulombic energy of the system V_{Coul} is then obtained as

$$V_{Coul} = W_{Coul} - \sum V_{IJ} \qquad (8)$$

This equation allows V_{Coul} to be calculated for any possible protein charge z and for any proton distribution by judiciously choosing the V_{IJ} terms that constitute $-\Sigma V_{IJ}$. The following criteria apply: (i) W_{Coul} considers every non-titratable residue twice. To eliminate these surplus contributions $-\Sigma V_{IJ}$ has to include all V_{IJ} terms involving non-titratable residues in the

"deprot" region of the W_{Coul} matrix. (ii) Protonated and deprotonated forms of the same residue cannot interact with each other, implying that the corresponding terms have to become part of $-\Sigma V_{IJ}$ as well. (iii) If a certain residue is protonated, there can be no contributions involving its deprotonated form (and vice versa).

For any given protein conformation, equations 7 and 8 require significant up-front computation, but once these are complete the calculation of V_{Coul} for arbitrary charge patterns is trivial. In Figures S1, S2 this method is illustrated by applying it to the zwitterionic species Gly^+~Gly~Gly^-, demonstrating that V_{Coul} obtained from equations 4/5 and from equations 7/8 is identical.

Implementation Details. The algorithm outlined above was implemented in Fortran. All-atom charge assignments for the twinned sequence were performed using a library assembled from Gromacs-generated topology files. After reading this library into memory, the code possesses the option to redefine charges such that a point charge pattern analogous to that used in earlier work[53] was obtained. Within this point charge model all atom charges are set to zero, and charges of $+e$ or $-e$ are placed only on atoms with the following OPLS/AA atom names:[59] "N" for NT^+, "NZ" for Lys^+, "NH2" for Arg^+, "NE" for His^+, "OD2" for Asp^-, "OE2" for Glu^-, and "C" for CT^-. With this option, the user can decide whether to run the proton redistribution code using the old point charge model that is based on equations 1, 2,[53] or the new all-atom model that considers equations 6-8. In either case, intrinsic proton affinities were dealt with as described in equation 2, using the following PA_{int} values (in kJ mol^{-1}).[31, 53, 64] NT: 960, Arg: 1029, Lys: 937, His: 958, Asp^-: 1450, Glu^-: 1450, CT^-: 1450.

Results and Discussion

Computational Cost. The goal of this study was to develop an improved mobile proton algorithm for MD studies on gaseous proteins. As in our earlier work,[53] the simulations employed an atomistic force field (OPLS/AA),[59] rather than a coarse grained model. The novel aspect introduced here is that proton redistribution is subject to the electrostatic contributions arising from *all* atoms, thereby taking into account the distributed nature of charge in all residues, as well as intramolecular solvation. This is in contrast to previous work[18, 31, 53] where solvation was neglected and titratable sites were modeled as point charges (Figure 1). It is instructive to consider the computational cost associated with the transition from a point charge model to an all-atom strategy for a typical protein such as avidin.

Proton redistribution *via* E_{tot} minimization requires the electrostatic energy to be calculated after each single proton test move. For a point charge model, the number of $Q_k Q_k / R_{kl}$ terms that have to be computed for a E_{tot} minimization with 20000 test moves and ~100 titratable sites is on the order of $20000 \times (0.5 \times 100^2) = 10^8$. Such point charge algorithms converge within a few minutes on standard desktop workstations.[31, 53]

When extending this simple approach to all atoms, $q_i q_j / r_{ij}$ has to be calculated for ~10000 charge sites, resulting in $20000 \times (0.5 \times 10000^2) = 10^{12}$ electrostatic terms. The computing time for a single E_{tot} minimization within such a "brute force" strategy would be roughly one month. Considering that a production run requires thousands of such E_{tot} minimizations (one after each 50 ps MD segment), it becomes clear that it would be impossible to use an all-atom strategy that involves calculating all electrostatic terms after each test move.[31, 53] For a 200 ns run of the type discussed below, the time spent on electrostatic calculations would be a staggering 4000 months (~300 years, assuming the same hardware is used as for the estimate in the previous paragraph).

The strategy developed here (equations 6-8) overcomes this problem. The algorithm starts with computation of W_{Coul} which contains roughly $(0.5 \times 20000^2) = 2 \times 10^8$ $q_i q_j / r_{ij}$ terms. After W_{Coul} and the associated V_{LJ} values are stored in memory, the electrostatic energy V_{Coul} can be calculated for any charge pattern simply by using a few subtraction steps (equation 8). Hence, despite increasing the number of charge sites by two orders of magnitude, the computational cost of the all-atom algorithm proposed here is comparable to that of previously used point charge approaches.[18, 31, 53] For clarity, it is emphasized again that W_{Coul} and its associated V_{LJ} values are recalculated after each 50 ps MD segment, such that the electrostatics always reflect the current protein conformation.

Charge Solvation. We illustrate the importance of charge solvation using the helical peptide Gln_6~Lys^+~Gln_6. The system was allowed to relax during a 1 ns MD run at 200 K, while keeping both termini neutral. In the initial conformation (not shown) all side chains pointed away from the helix axis. After 1 ns the Lys^+ side chain was bent inwards, allowing for extensive charge solvation. Specifically, the Lys-NH_3^+ group was found to interact with two carbonyl dipoles on Gln side chain. Also, the helix opened up slightly, allowing for additional solvation by a backbone carbonyl (Figure 2A). For comparison we can consider another peptide of the same fold, where all Gln side chains were altered to methyl groups (Ala_6~Lys^+~Ala_6, Figure 2B). The absence of side chain carbonyls in this second scenario implies that charge solvation is significantly reduced.

For examining the implications of charge solvation we consider the proton affinity (*PA*) of the Lys side chain in the two scenarios, assuming a rigid peptide conformation. *PA* of a gas phase species M is defined as negative enthalpy difference of the reaction[29]

$$M + H^+ \rightarrow [M + H]^+ \qquad PA = -\Delta H \qquad (9)$$

When applying the all-atom algorithm it is found that $PA = -(E_{tot}[M+H]^+ - E_{tot}[M])$, yielding PA = 1230 kJ mol^{-1} for the species in Figure 2A. The less extensively solvated peptide in Figure 2B has PA = 1054 kJ mol^{-1}. Point charge models[31, 53] ignore all of the uncharged residues,[24,53] such that $PA = PA_{int}$(Lys) = 937 kJ mol^{-1} in both cases.

Unlike equation 9, proton redistribution within gaseous proteins does not involve the binding of free protons. Nonetheless, the example considered here illustrates how solvation can affect internal charge patterns. In the context of a large protein Figure 2A would represent a Lys residue that is likely to capture a mobile proton. For Figure 2B there would be a high probability of having the proton transferred elsewhere, *i.e.*, to a site with more favorable solvation and/or higher PA_{int}. Simple point charge algorithms[18, 31, 53] are oblivious of such differences in local environment. Thus, the inclusion of charge solvation provides a more accurate picture of charge patterns in the gas phase.

Deprotonated Sites and Salt Bridges. Recent studies suggest that $[M + zH]^{z+}$ ions can contain zwitterionic motifs in the form of salt bridges.[32-38] This is in contrast to the traditional view that all carboxylates will be neutralized due to their high PA_{int}, yielding neutral pairs (*e.g.*, $Lys^+/Asp^- \rightarrow Lys^0/Asp^0$).[18, 28-31] We will consider a simple system to examine under what conditions a salt bridge is energetically preferred over the neutralized form. We conducted a point charge analysis and an all-atom treatment for Ala_6~Lys~Ala~Glu~Ala_6. The zwitterionic tautomer was subjected to energy minimization and 2 ps of MD at 200 K, yielding a structure where the Lys^+/Glu^- side chains are in close contact (Figure 3A). Is this salt-bridged form stable, or will H^+ transfer produce Lys^0/Glu^0 (Figure 3B)? The zwitterion is prevalent if $\Delta E_{tot} = E_{tot}(Lys^+/Glu^-) - E_{tot}(Lys^0/Glu^0) < 0$.[53] The point charge approximation (equations 1 and 2) results in

$$\Delta E_{tot} = 513 \text{ kJ mol}^{-1} - \frac{e^2}{4\pi\varepsilon_0 R} \qquad (10)$$

where R is the distance between the two charges. For Figure 3 with R = 0.26 nm it was found that ΔE_{tot} = -18 kJ mol^{-1}, *i.e.*, the Lys^+/Glu^- tautomer is predicted to be stable within the point charge approximation. The all-atom treatment yields

$$\Delta E_{tot} = 513 \text{ kJ mol}^{-1} + V_{Coul}(Lys^+ / Glu^-) - V_{Coul}(Lys^0 / Glu^0) \qquad (11)$$

which corresponds to ΔE_{tot} = +34 kJ mol^{-1}. Hence, the all-atom model predicts conversion of the Lys^+/Glu^- salt-bridge to the Lys^0/Glu^0 tautomer.

For both models the survival chances of zwitterionic motifs increases as the distance between acid/base pairs decreases, because small distances between unlike charges provide favorable electrostatic interactions. The point charge model and the all-atom approach can nonetheless yield contradicting predictions. For Figure 3 the all-atom model does not support retention of the salt bridge. However, the ΔE_{tot} values considered here are quite small compared to typical solvation-induced PA changes (Figure 2). It will be seen below that zwitterionic motifs can be highly prevalent within the all-atom model, as long as the system offers close acid/base proximity as well as favorable solvation.

Optimizing the Charge Pattern of a Static Protein Structure. Avidin is a homo-tetramer with a total of 504 residues and 120 titratable sites (Figure 4A, Figure S3). During native ESI it produces $[M + 16H]^{16+}$ ions.[20, 55] We initially tested the all-atom proton distribution algorithm by applying it to the static X-ray structure.[60] Three 16+ input charge patterns were examined. The first pattern has charges 16/0/0/0 on subunits [a]-[d] (Figure 4B). In addition, two 4/4/4/4 input patterns were examined (Table 1). All three scenarios produced the same

optimized 4/4/4/4 pattern, with a $Arg100^{+}/Asp109^{-}$ salt bridge in each subunit (Table 1, Figure 4C).

The energy changes associated with charge redistribution are different for the three initial scenarios. By design, ΔE_{tot} is always negative, reflecting the fact that the algorithm searches for the energetically most favorable outcome (Figure 5A). Scenarios (1) and (2) are characterized by $\Delta V_{Coul} < 0$, as electrostatic repulsion causes protons to spread out over the protein surface (Figure 5B). Changes in E_{PAint} are small because proton hopping occurs primarily between basic sites which share similar PA_{int} values (Figure 5C). A different behavior is seen for scenario (3), which starts with numerous deprotonated sites (Table 1). These conditions give rise to $\Delta V_{Coul} > 0$ because proton redistribution eliminates numerous negative charges, thereby causing the loss of favorable +/- interactions (Figure 5B). Instead, proton migration for scenario (3) is driven by $R\text{-}COO^{-}$ protonation ($\Delta E_{PAint} < 0$, Figure 5C).

The data of Table 1 confirm that the all-atom proton redistribution algorithm can successfully navigate the vast parameter space associated with E_{tot} minimization, producing a physically reasonable 4/4/4/4 pattern for the symmetrical avidin X-ray structure. The fact that different starting patterns produce the same solution indicates that the algorithm is quite insensitive to the initial conditions.

Mobile Proton MD Simulations. The dynamics of avidin $[M + 16H]^{16+}$ were modeled by subjecting the protein to 200 ns MD runs, with proton redistribution in 50 ps intervals using the all-atom algorithm. Three simulations were conducted at 300 K, mimicking the gentle thermal environment experienced under native ESI-MS conditions.[2-8] The charge patterns obtained after 200 ns for the three runs were 3/4/4/5, 3/4/4/5, and 3/5/4/4. Because the overall features seen in all simulations were quite similar we will only consider the first run in more detail. Additional data are summarized in Tables S1 and S2.

Starting with the X-ray coordinates and the optimized 16+ charge pattern (see previous section), the internal proton distribution underwent dramatic changes during the MD run. The number of deprotonated sites increased from 4 (Asp^-) at t = 0 to 29 sites (Asp^-, Glu^-, as well as three deprotonated C-termini) at t = 200 ns (Figure 4D). The internal charge distributions within the subunits were not perfectly symmetrical and underwent some fluctuations during the simulation window (Table 1). For example, at around $t \approx 100$ ns, the complex transiently adopted a 4/3/3/6 charge pattern that persisted for almost 50 ns. Under the gentle simulation conditions used here (300 K) this relatively high degree of charge asymmetry was not associated with major alterations in the overall structure of the complex. Interestingly, the existence of such asymmetric charge patterns has been postulated recently, as scenarios of this type may play a role at the onset of dissociation processes that take place upon collisional heating.[35]

In our MD simulations all deprotonated sites formed salt bridges with at least one Arg^+ or Lys^+ side chain. Formation of these contacts was mediated by side chain orientational alterations. For example, in the initial X-ray structure titratable sites adjacent to Asp86 do not strongly interact with each other (Figure 4A). In contrast, the t = 200 ns MD structure exhibits an intricate salt-bridged cluster comprising $Arg2^+$, $Asp86^-$, $Arg87^+$, $Lys90^+$, $Glu91^-$ and Arg 122^+ (Figure 4E). The charged side chains are packed such that a number of NH···OC hydrogen bonds are formed. Charged clusters analogous to that depicted in Figure 4E were also found in other locations throughout the protein (Figure 4D). Despite these major changes in side chain orientation, the overall secondary, tertiary, and quaternary structure of the protein was largely preserved during the 200 ns gas phase simulations, with a backbone rmsd of 2.5 Å relative to the X-ray coordinates (Figure 4F).

Energy changes during the 200 ns simulation window are summarized in Figure 5. V_{Coul} undergoes a steep ~1.5×10^4 kJ mol^{-1} decline during the first few nanoseconds, caused

by the formation of new salt bridges. This process is concomitant with an E_{PAint} increase due to proton migration that produces vacant R-COO$^-$ moieties. Hence, proton migration is accompanied by compensation between E_{PAint} and V_{Coul}, where unfavorable proton affinity contributions are balanced by the formation of electrostatic +/- contacts. These opposing energetic trends are most dramatic during the first nanosecond, after which they continue at a much slower pace with $E_{tot} \approx const$ (Figure 5A).

Point Charge Results for Avidin. When applied to the avidin X-ray structure, the point charge model produced a charge pattern different from that seen for the all-atom algorithm (Table S1). The former generated zwitterionic clusters with participation of Arg26, Glu28, and His50 (Figure S4A), in contrast to the Arg100$^+$/Asp109$^-$ salt bridges seen in all-atom mode (Figure 4C). In 200 ns MD runs the point charge model produced compact avidin gas phase structures (Figures S4B and S4C) resembling those discussed above for the all-atom model. Owing to their high PA_{int},[38] Arg residues are preferred protonation sites under both conditions, while the locations of negative charges are quite different for the two models (Table S2). These differences reflect the fact that the point charge model is insensitive to intramolecular solvation, and that it uses a crude description of charge sites (Figure 1). This is in contrast to the all-atom model, where electrostatics are treated with a much higher level of detail. Charge patterns predicted by the all-atom model will therefore be more reliable.

Conclusions

It remains a contentious question to what extent electrosprayed protein ions retain their solution structures.[13-17, 58] One problem is a lack of information regarding the charge pattern and the presence of salt bridges in the gas phase. Most techniques that are capable of probing

these issues are limited to small systems,[33, 54, 65, 66] leaving large complexes that are of key interest for many biological studies out of reach.

In the past, algorithms of different complexity have been used for predicting energetically favorable charge patterns within $[M + zH]^{z+}$ macromolecular ions. The simplest approaches considered only the electrostatic repulsion between protonated basic sites, which were assumed to be point charges.[18] More advanced methods took into account the intrinsic proton affinities of these sites.[31] The next level was the inclusion of –COOH moieties as possible proton donors, thereby allowing for the presence of salt bridges.[53] The all-atom model introduced here takes the additional step of incorporating intramolecular charge solvation, a factor that is known to be highly prevalent in gas phase proteins.[16, 29, 54]

To our knowledge, the current work marks the first time that an all-atom mobile proton model has been applied for simulating the structural dynamics of a large biomolecular complex *in vacuo*. Previous efforts in this direction may have been hampered by the seemingly staggering computational cost of all-atom electrostatic calculations. We devised a simple solution to this problem which allows the energies of arbitrary charge patterns to be determined via a computationally inexpensive subtractive strategy.

Consistent with earlier gas phase studies[40, 41, 53, 56-58] the current work employed the OPLS/AA force field, although OPLS/AA was originally designed for solution simulations.[59] However, the architecture of our mobile proton algorithm is not uniquely tied to OPLS/AA. Minor adjustments will allow the algorithm to be used with other force fields, including future parametrizations that might be specifically optimized for the gas phase.

Our MD conditions aimed to mimic a native ESI environment where gaseous proteins are produced in low charge states, while keeping collisional heating at a minimum. Consistent with experiments[2-8] our simulations revealed that proteins can maintain solution-like conformations under these conditions. Kinetic trapping[10-12] of these conformers is

promoted by an intricate network of salt bridges that are reinforced by side chain hydrogen bonds. Our results indicate that the formation of such salt bridge networks via rapid side chain reorientation and proton hopping is an intrinsic property of gaseous proteins. While the presence of salt bridges in the gas phase has been proposed in earlier,[32-38] the large number of these contacts predicted by our all-atom model is nonetheless surprising.

The all-atom mobile proton approach introduced here paves the way towards future MD studies on various processes that involve charge migration. For example, our laboratory is currently working towards applications related to collision-induced dissociation (CID)[19] and surface-induced dissociation (SID)[21] of protein complexes, collapse and compaction events,[18, 20] collisionally triggered unfolding and refolding,[26] and conformational changes caused by charge reduction.[27] Several of these processes have previously been simulated using either static charges[18, 26] or charge hopping algorithms that did not consider solvation.[31, 53] It is hoped that the mobile proton strategy of this work will provide the foundation for a better mechanistic understanding of biomolecular behavior in the gas phase.

Supporting Information. *Figure S1:* Electrostatic calculations for Gly^{+}~Gly~Gly^{-}. *Figure S2:* Electrostatic calculations for Gly^{+}~Gly~Gly^{-} using W_{Coul}. *Figure S3:* Avidin sequence and titratable sites. *Figure S4:* Mobile proton results obtained using the point charge approximation. *Table S1:* Optimized X-ray charge patterns. *Table S2:* Optimized 200 ns charge patterns. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments. Funding for this work was provided by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 217080-2013).

References

1. Fenn, J. B. Electrospray wings for molecular elephants (nobel lecture). *Angew. Chem. Int. Ed.* **2003,** *42*, 3871-3894.
2. Leney, A. C.; Heck, A. J. R. Native mass spectrometry: What is in the name? *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 5-13.
3. Morrison, L. J.; Brodbelt, J. S. 193 nm ultraviolet photodissociation mass spectrometry of tetrameric protein complexes provides insight into quaternary and secondary protein topology. *J. Am. Chem. Soc.* **2016,** *138*, 10849-10859.
4. Fatunmbi, O.; Abzalimov, R. R.; Savinov, S. N.; Gershenson, A.; Kaltashov, I. A. Interactions of haptoglobin with monomeric globin species: Insights from molecular modeling and native electrospray ionization mass spectrometry. *Biochemistry* **2016,** *55*, 1918-1928.
5. Konijnenberg, A.; Butterer, A.; Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta* **2013,** *1834*, 1239-1256.
6. Kaddis, C. S.; Loo, J. A. Native protein ms and ion mobility: Large flying proteins with esi. *Anal. Chem.* **2007,** *79*, 1779-1784.
7. Sharon, M. How far can we go with structural mass spectrometry of protein complexes? *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 487-500.
8. Ashcroft, A. E. Mass spectrometry and the amyloid problem - how far can we go in the gas phase? *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 1087-1096.
9. Landreh, M.; Marklund, E. G.; Uzdavinys, P.; Degiacomi, M. T.; Coincon, M.; Gault, J.; Gupta, K.; Liko, I.; Benesch, J. L. P.; Drew, D., et al. Integrating mass spectrometry with md simulations reveals the role of lipids in na+/h+ antiporters. *Nat. Commun.* **2017,** *8*.
10. Hendricks, N. G.; Julian, R. R. Leveraging ultraviolet photodissociation and spectroscopy to investigate peptide and protein three-dimensional structure with mass spectrometry. *Analyst* **2016,** *141*, 4534-4540.
11. Deng, L.; Broom, A.; Kitova, E. N.; Richards, M. R.; Zheng, R. B.; Shoemaker, G. K.; Meiering, E. M.; Klassen, J. S. Kinetic stability of the streptavidin-biotin interaction enhanced in the gas phase. *J. Am. Chem. Soc.* **2012,** *134*, 16586-16596.
12. Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H. From solution to the gas phase: Stepwise dehydration and kinetic trapping of substance p reveals the origin of peptide conformations. *J. Am. Chem. Soc.* **2013,** *135*, 19147-19153.
13. Skinner, O. S.; McLafferty, F. W.; Breuker, K. How ubiquitin unfolds after transfer into the gas phase. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1011-1014.
14. Wyttenbach, T.; Bowers, M. T. Structural stability from solution to the gas phase: Native solution structure of ubiquitin survives analysis in a solvent-free ion mobility–mass spectrometry environment. *J. Phys. Chem. B* **2011,** *115*, 12266-12275.
15. Oh, M. I.; Consta, S. Stability of a transient protein complex in a charged aqueous droplet with variable ph. *J. Phys. Chem. Lett.* **2017,** *8*, 80-85.
16. Warnke, S.; von Helden, G.; Pagel, K. Protein structure in the gas phase: The influence of side-chain microsolvation. *J. Am. Chem. Soc.* **2013,** *135*, 1177-1180.
17. Chen, S.-H.; Russell, D. H. How closely related are conformations of protein ions sampled by im-ms to native solution structures? *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 1433-1443.
18. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. Charge-state dependent compaction and dissociation of protein complexes: Insights from ion mobility and molecular dynamics. *J. Am. Chem. Soc.* **2012,** *134*, 3429-3438.

19. Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion mobility–mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed.* **2007,** *46*, 8001-8004.
20. Pacholarz, K. J.; Barran, P. E. Distinguishing loss of structure from subunit dissociation for protein complexes with variable temperature ion mobility mass spectrometry. *Anal. Chem.* **2015,** *87*, 6271-6279.
21. Harvey, S. R.; Yan, J.; Brown, J. M.; Hoyes, E.; Wysocki, V. H. Extended gas-phase trapping followed by surface-induced dissociation of noncovalent protein complexes. *Anal. Chem.* **2016,** *88*, 1218-1221.
22. Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **1996,** *118*, 8365-8374.
23. Boyd, R. K.; Somogyi, Á. The mobile proton hypothesis in fragmentation of protonated peptides: A perspective. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 1275-1278.
24. Li, J. Y.; Lyu, W. P.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P. Proton dynamics in protein mass spectrometry. *J. Phys. Chem. Lett.* **2017,** *8*, 1105-1112.
25. Cautereels, J.; Blockhuys, F. Quantum chemical mass spectrometry: Verification and extension of the mobile proton model for histidine. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1227-1235.
26. Shelimov, K. B.; Jarrold, M. F. Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding. *J. Am. Chem. Soc.* **1997,** *119*, 2987-2994.
27. Laszlo, K. J.; Munger, E. B.; Bush, M. F. Folding of protein ions in the gas phase after cation -to-anion proton-transfer reactions. *J. Am. Chem. Soc.* **2016,** *138*, 9581-9588.
28. Loo, J. A.; Edmonds, C. G.; Udseh, H. R.; Smith, R. D. Effect of reducing disulfide-containing proteins on electrospray ionisation mass spectra. *Anal. Chem.* **1990,** *62*, 693-698.
29. Schnier, P. D.; Gross, D. S.; Williams, E. R. Electrostatic forces and dielectric polarizability of multiply protonated gas-phase cytochrome c ions probed by ion/molecule chemistry. *J. Am. Chem. Soc.* **1995,** *117*, 6747-6757.
30. Mao, Y.; Woenckhaus, J.; Kolafa, J.; Ratner, M. A.; Jarrold, M. F. Thermal unfolding of unsolvated cytochrome c: Experiment and molecular dynamics simulations. *J. Am. Chem. Soc.* **1999,** *121*, 2712-2721.
31. Fegan, S. K.; Thachuk, M. A charge moving algorithm for molecular dynamics simulations of gas-phase proteins. *J. Chem. Theory Comput.* **2013,** *9*, 2531-2539.
32. Yoo, H. J.; Wang, N.; Zhuang, S. Y.; Song, H. T.; Hakansson, K. Negative-ion electron capture dissociation: Radical-driven fragmentation of charge-increased gaseous peptide anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.
33. Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. Infrared spectroscopy of arginine cation complexes: Direct observation of gas-phase zwitterions. *J. Phys. Chem. A* **2007,** *111*, 11759-11770.
34. Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational effects in protein electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **2016,** *35*, 111-122.
35. Ogorzalek Loo, R. R.; Loo, J. A. Salt bridge rearrangement (sabre) explains the dissociation behavior of noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 975-990.
36. Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 604-613.

37. Breuker, K.; Brüschweiler, S.; Tollinger, M. Electrostatic stabilization of a native protein structure in the gas phase. *Angew. Chem. Int. Ed.* **2011,** *50*, 873-877.
38. Bonner, J. G.; Lyon, Y. A.; Nellessen, C.; Julian, R. R. Photoelectron transfer dissociation reveals surprising favorability of zwitterionic states in large gaseous peptides and proteins. *J. Am. Chem. Soc.* **2017,** *139*, 10286-10293.
39. Beveridge, R.; Migas, L. G.; Payne, K. A. P.; Scrutton, N. S.; Leys, D.; Barran, P. E. Mass spectrometry locates local and allosteric conformational changes that occur on cofactor binding. *Nat. Commun.* **2016,** *7*.
40. Marchese, R.; Grandori, R.; Carloni, R.; Raugei, S. A computational model for protein ionization by electrospray based on gas-phase basicity. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1903-1910.
41. Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; van der Spoel, D. A direct comparison of protein structure in the gas and solution phase: The trp-cage. *J. Phys. Chem. B* **2007,** *111*, 13147-13150.
42. Khakinejad, M.; Kondalaji, S. G.; Tafreshian, A.; Valentine, S. J. Gas-phase hydrogen-deuterium exchange labeling of select peptide ion conformer types: A per-residue kinetics analysis. *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 1115-1127.
43. Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. Amyloid β-protein monomer structure: A computational and experimental study. *Protein Sci.* **2006,** *15*, 420-428.
44. Goyal, P.; Qian, H. J.; Irle, S.; Lu, X. Y.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular simulation of water and hydration effects in different environments: Challenges and developments for dftb based models. *J. Phys. Chem. B* **2014,** *118*, 11007-11027.
45. Barnes, G. L.; Hase, W. L. Energy transfer, unfolding, and fragmentation dynamics in collisions of n-protonated octaglycine with an h-sam surface. *J. Am. Chem. Soc.* **2009,** *131*, 17185-17193.
46. Marx, D.; Chandra, A.; Tuckerman, M. E. Aqueous basic solutions: Hydroxide solvation, structural diffusion, and comparison to the hydrated proton. *Chem. Rev.* **2010,** *110*, 2174-2216.
47. Iyengar, S. S.; Day, T. J. F.; Voth, G. A. On the amphiphilic behavior of the hydrated proton: An ab initio molecular dynamics study. *Int. J. Mass Spectrom.* **2005,** *241*, 197-204.
48. Burlet, O.; Orkiszewski, R. S.; Ballard, K. D.; Gaskell, S. J. Charge promotion of low-energy fragmentations of peptide ions. *Rapid Commun. Mass Spectrom.* **1992,** *6*, 658-662.
49. Nakai, H.; Sakti, A. W.; Nishimura, Y. Divide-and-conquer-type density-functional tight-binding molecular dynamics simulations of proton diffusion in a bulk water system. *J. Phys. Chem. B* **2016,** *120*, 217-221.
50. Lill, M. A.; Helms, V. Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (q-hop md). *J. Chem. Phys.* **2001,** *115*, 7993-8005.
51. Peng, Y. X.; Swanson, J. M. J.; Kang, S. G.; Zhou, R. H.; Voth, G. A. Hydrated excess protons can create their own water wires. *J. Phys. Chem. B* **2015,** *119*, 9212-9218.
52. Marrink, S. J.; Tieleman, D. P. Perspective on the martini model. *Chem. Soc. Rev.* **2013,** *42*, 6801-6822.
53. Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L. Collision-induced dissociation of electrosprayed protein complexes: An all-atom molecular dynamics model with mobile protons. *J. Phys. Chem. B* **2016,** *120*, 5114-5124.
54. Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S. On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput. Biol.* **2010,** *6*.

55. Bornschein, R. E.; Niu, S.; Eschweiler, J.; Ruotolo, B. T. Ion mobility-mass spectrometry reveals highly-compact intermediates in the collision induced dissociation of charge-reduced protein complexes. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 41-9.
56. Fegan, S. K.; Thachuk, M. Suitability of the martini force field for use with gas-phase protein complexes. *J. Chem. Theory Comput.* **2012,** *8*, 1304-1313.
57. Cadenas, E.; Boveris, A.; Chance, B. Low-level chemiluminescence of hydroperoxide-supplemented cytochrome c. *Biochem. J.* **1980,** *187*, 131-140.
58. Ly, T.; Julian, R. R. Elucidating the tertiary structure of protein ions in vacuo with site specific photoinitiated radical reactions. *J. Am. Chem. Soc.* **2010,** *132*, 8602-8609.
59. Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free energies of hydration and pure liquid properties of hydrocarbons from the opls all-atom model. *J. Phys. Chem.* **1994,** *98*, 13077-13082.
60. Strzelczyk, P.; Bujacz, G. Crystal structure and ligand affinity of avidin in the complex with 4'-hydroxyazobenzene-2-carboxylic acid. *J. Mol. Struct.* **2016,** *1109*, 232-238.
61. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1–2*, 19-25.
62. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31*, 1695-1697.
63. Sudmeier, J. L.; Bradshaw, E. M.; Haddad, K. E. C.; Day, R. M.; Thalhauser, C. J.; Bullock, P. A.; Bachovchin, W. W. Identification of histidine tautomers in proteins by 2d h-1/c-13(delta 2) one-bond correlated nmr. *J. Am. Chem. Soc.* **2003,** *125*, 8430-8431.
64. Bienkowski, T.; Swider, P.; Blaziak, K.; Danikiewicz, W. Proton affinities of the anions of aromatic carboxylic acids measured by kinetic method. *Int. J. Mass Spectrom.* **2014,** *357*, 29-33.
65. Jockusch, R. A.; Price, W. D.; Williams, E. R. Structure of cationized arginine (arg center dot m+, m = h, li, na, k, rb, and cs) in the gas phase: Further evidence for zwitterionic arginine. *J. Phys. Chem. A* **1999,** *103*, 9266-9274.
66. Ahn, S.; Cong, X.; Lebrilla, C. B. Zwitterion formation in gas-phase cyclodextrin complexes. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 166-175.

Figure Captions

Figure 1. Internal charge distributions of amino acid residues Lys^+, Arg^+, Glu^-, and Gln. (A) Simplified representations used for earlier mobile proton studies employing point charge models.[31, 53] (B) OPLS/AA charges used for the all-atom electrostatic calculations of this work. Partial charges (0.06) for hydrogen atoms in C-H bonds were omitted to reduce cluttering. Element coloring: C, green; N, blue; O, red; H, white.

Figure 2. Intramolecular charge solvation in small gas phase peptides. (A) Gln_6~Lys^+~Gln_6. The Lys ammonium group is solvated by two side chain (sc) carbonyl groups and one backbone (bb) carbonyl. This charge solvation conveys a high proton affinity. (B) Ala_6~Lys^+~Ala_6. Charge solvation is less favorable due to the absence of side chain carbonyls, resulting in a lower proton affinity.

Figure 3. Gas phase peptide Ala_6~Lys~Ala~Glu~Ala_6 as (A) zwitterionic Lys^+/Glu^- species and (B) as neutralized Lys^0/Glu^0 tautomer. A point charge approximation predicts the prevalence of the former, whereas all-atom electrostatics favor the latter. The results obtained for this particular system do not imply that all-atom electrostatics generally disfavor salt bridges (for details, see text).

Figure 4. Mobile proton results for $[avidin + 16H]^{16+}$ in all-atom mode. (A) X-ray structure 5chk.[60] Subunits are depicted in different colors. Basic (blue) and acidic (red) sites are highlighted. (B) Manually selected initial pattern, with charges 16/0/0/0 on subunits [a]/[b]/[c]/[d]. Heavy atoms of positive (blue) and negative (red) residues are shown as spheres. (C) Optimized 4/4/4/4 charge pattern of the X-ray structure after proton redistribution. (D) Protein structure and 3/4/4/5 charge pattern obtained after a 200 ns mobile

proton MD simulation. (E) Close-up of a salt bridge cluster in subunit [b] at t = 200 ns. (F) Comparison of X-ray coordinates and 200 ns MD structure.

Figure 5. Energy changes during a typical 200 ns mobile proton MD simulation for gas phase [avidin + 16H]$^{16+}$. (A) Overall energy E_{tot}. (B) Electrostatic contribution V_{Coul}. (C) Proton affinity contribution E_{PAint}. At t = 0 the MD run starts with the X-ray structure of the protein (dashed line). Shown at the dotted line are the energies of three initial charge patterns (1), (2) and (3), defined in Table 1.

Table 1. Charge patterns in subunits [a]-[d] of [avidin + 16H]$^{16+}$. The table illustrates three input scenarios, the optimized pattern for the X-ray structure, and the pattern for four MD time points obtained in one particular run. Each line shows the net charge and the charge pattern in a given subunit. Also shown is the sequence of titratable sites in single-letter code. All data were produced using the all-atom electrostatic model.

	NT	R	K	K	D	R	E	E	K	E	H	E	K	R	K	E	D	R	K	E	K	R	D	D	D	K	R	R	R	CT
initial pattern (1)																														
[a]	16	0	1	1	1	0	1	0	0	1	0	0	0	1	1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	1	0
[b]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[d]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
initial pattern (2)																														
[a]	4	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[b]	4	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[c]	4	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[d]	4	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
initial pattern (3)																														
[a]	4	0	1	1	1-1	1-1-1	1-1	0-1	1	1	1-1-1	1	1-1	1	1-1-1-1	1	1	1	1-1											
[b]	4	0	1	1	1-1	1-1-1	1-1	0-1	1	1	1-1-1	1	1-1	1	1-1-1-1	1	1	1	1-1											
[c]	4	0	1	1	1-1	1-1-1	1-1	0-1	1	1	1-1-1	1	1-1	1	1-1-1-1	1	1	1	1-1											
[d]	4	0	1	1	1-1	1-1-1	1-1	0-1	1	1	1-1-1	1	1-1	1	1-1-1-1	1	1	1	1-1											
Charge pattern of X-ray structure after optimization																														
[a]	4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0-1	0	0	1	1	0		
[b]	4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0-1	0	0	1	1	0		
[c]	4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0-1	0	0	1	1	0		
[d]	4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0-1	0	0	1	1	0		
***t* = 50 ns**																														
[a]	3	0	0	0	0	0	1-1	0	1	0	0-1	1	1	0	0-1	1	1	0	0	1	0-1-1	0	0	1	1-1					
[b]	4	0	1	0	0	0	1-1-1	1	0	0	0	0	1	1-1-1	1	1-1	0	1	0-1-1	1	1	1	1	1-1					
[c]	4	0	0	0	0	0	1-1-1	1	0	0-1	1	1	1	0-1	1	1-1	0	1-1	0-1	1	0	1	1	0					
[d]	5	0	0	1	0	0	1-1	0	0-1	1	0	1	1	0	0	0	0	1-1	0	1	0-1-1	1	1	1	1-1					
***t* = 100 ns**																														
[a]	4	0	1	0	0	0	1-1	0	1	0	0	0	0	1	0	0-1	1	1	0	0	1-1-1-1	1	0	1	1-1					
[b]	3	0	1	0	0	0	0-1-1	1	0	0	0	0	1	1-1-1	1	1-1	0	1	0-1-1	1	1	1	1	1-1					
[c]	3	0	0	0	0	0	1-1-1	1	0	0-1	1	1	1-1-1	1	1-1	0	1-1	0-1	1	0	1	1	0							
[d]	6	0	0	1	0	0	1-1	0	0	0	1	0	1	1	0	0	0	1	0-1	0	1	0-1-1	1	1	1	0	0			
***t* = 150 ns**																														
[a]	3	0	1	0	0	0	1-1	0	1	0	0-1	1	1	0	0-1	0	1	0	0	1-1	0-1	0	0	1	1-1					
[b]	4	0	1	1	0	0	1-1-1	1	0	0-1	0	1	1-1-1	1	1-1	0	1	0-1-1	1	1	1	1	1-1							
[c]	4	0	0	0	0	0	1-1-1	1	0	0-1	1	1	1	0-1	1	1-1	0	1-1	0-1	1	0	1	1	0						
[d]	5	0	1	1	0	0	1-1	0	0-1	1	0	1	1	0	0-1	0	1-1	0	1	0-1-1	1	1	1	1-1						
***t* = 200 ns**																														
[a]	3	0	1	0	0	0	1-1	0	1	0	0-1	1	1	0	0-1	0	1	0	0	1-1	0-1	0	0	1	1-1					
[b]	4	0	1	1	0	0	1-1-1	1	0	0-1	0	1	1-1-1	1	1-1	0	1	0-1-1	1	1	1	1	1-1							
[c]	4	0	0	0	0	0	1-1-1	1	0	0-1	1	1	1	0-1	1	1-1	0	1-1	0-1	1	0	1	1	0						
[d]	5	0	1	1	0	0	1-1	0	0-1	1	0	1	1	0	0-1	0	1-1	0	1	0-1-1	1	1	1	1-1						
	NT	R	K	K	D	R	E	E	K	E	H	E	K	R	K	E	D	R	K	E	K	R	D	D	D	K	R	R	R	CT

Figure 1

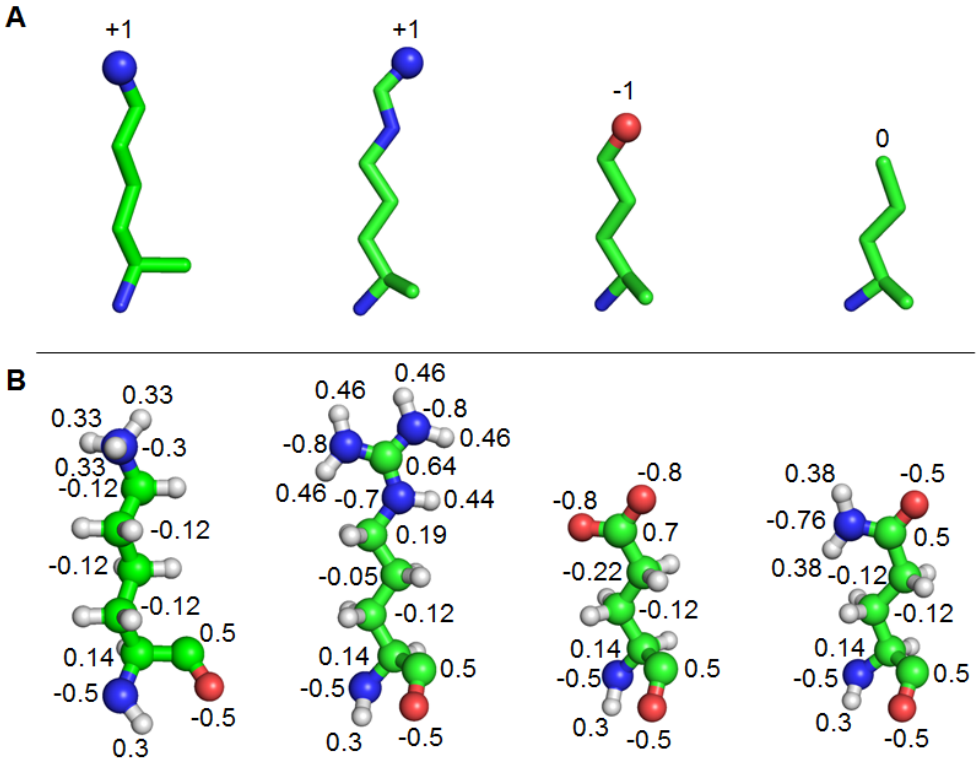

A
+1
+1
-1
0
B
0.33
0.33
-0.3
0.33
-0.12
-0.12
-0.12
-0.12
0.5
0.14
-0.5
-0.5
0.3
0.46
0.46
-0.8
-0.8
0.46
0.64
0.46
-0.7
0.44
0.19
-0.05
-0.12
0.14
0.5
-0.5
0.3
-0.5
-0.8
-0.8
0.7
-0.22
-0.12
0.14
0.5
-0.5
0.3
-0.5
0.38
-0.5
-0.76
0.5
0.38
-0.12
-0.12
0.14
0.5
-0.5
0.3
-0.5

Figure 2

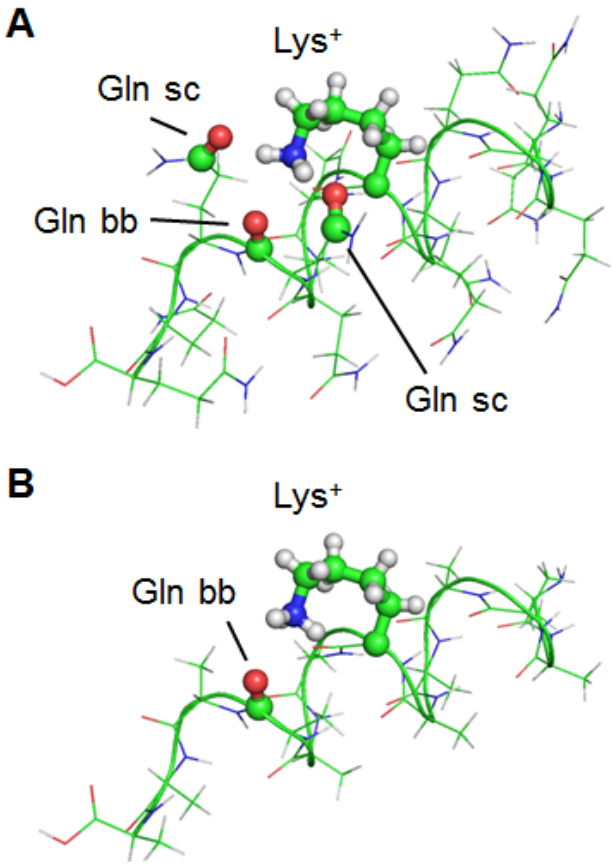


Figure 3

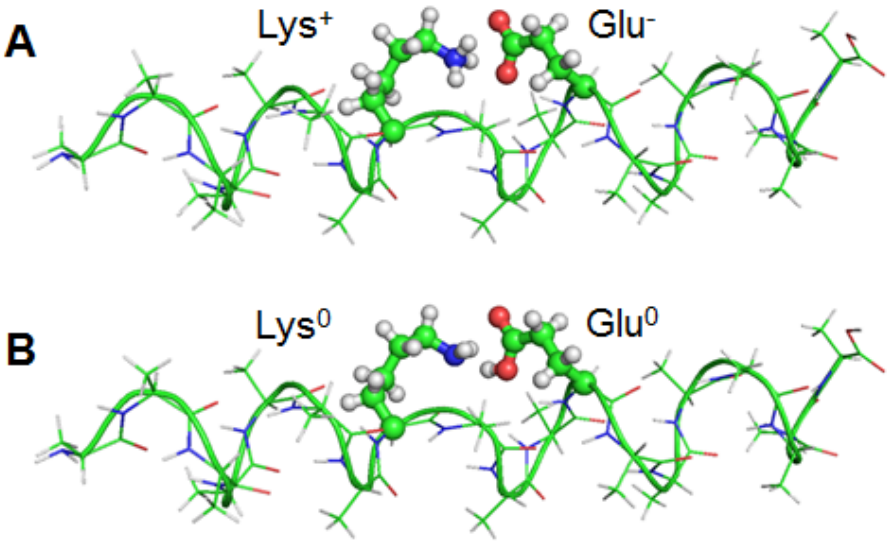


Figure 4

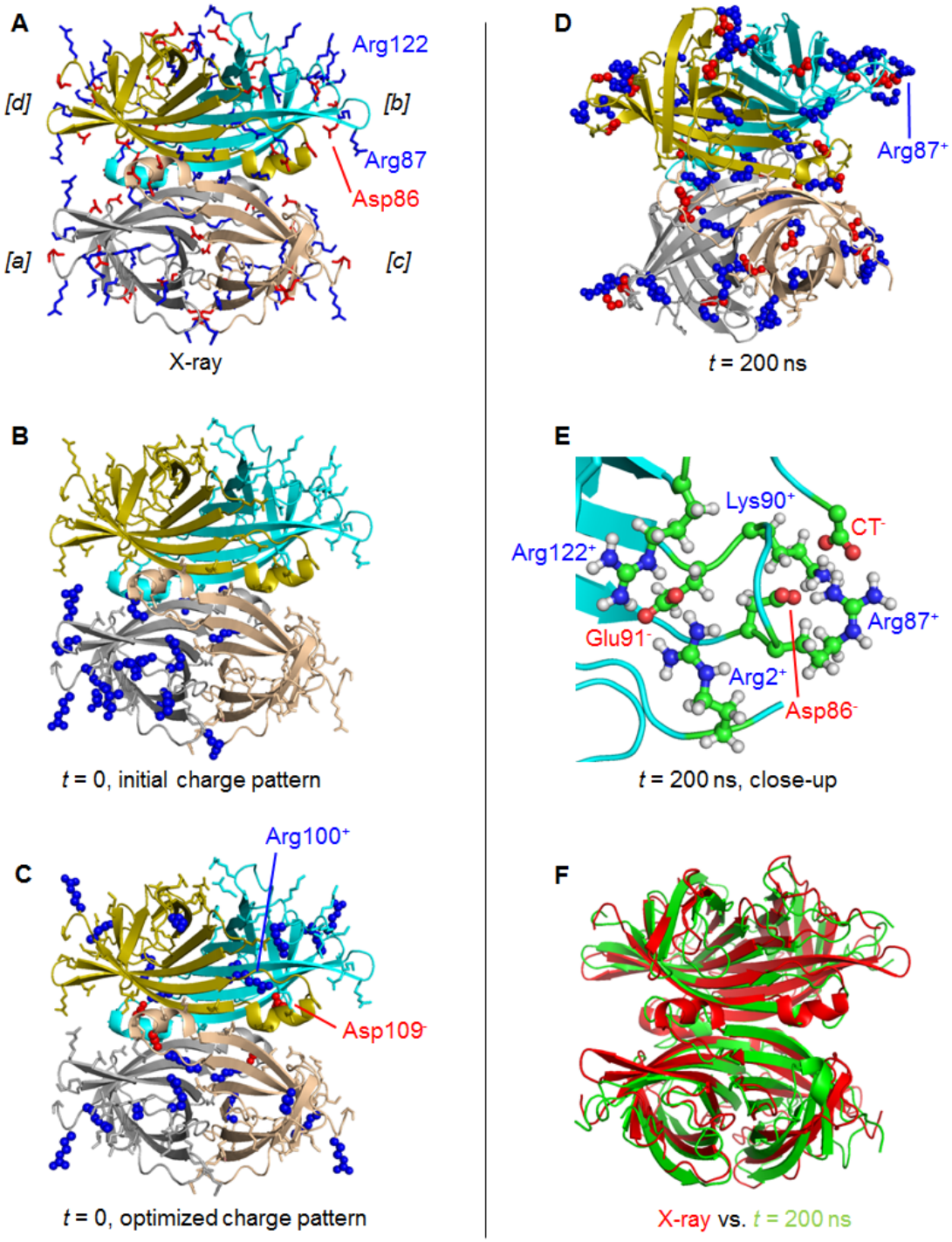
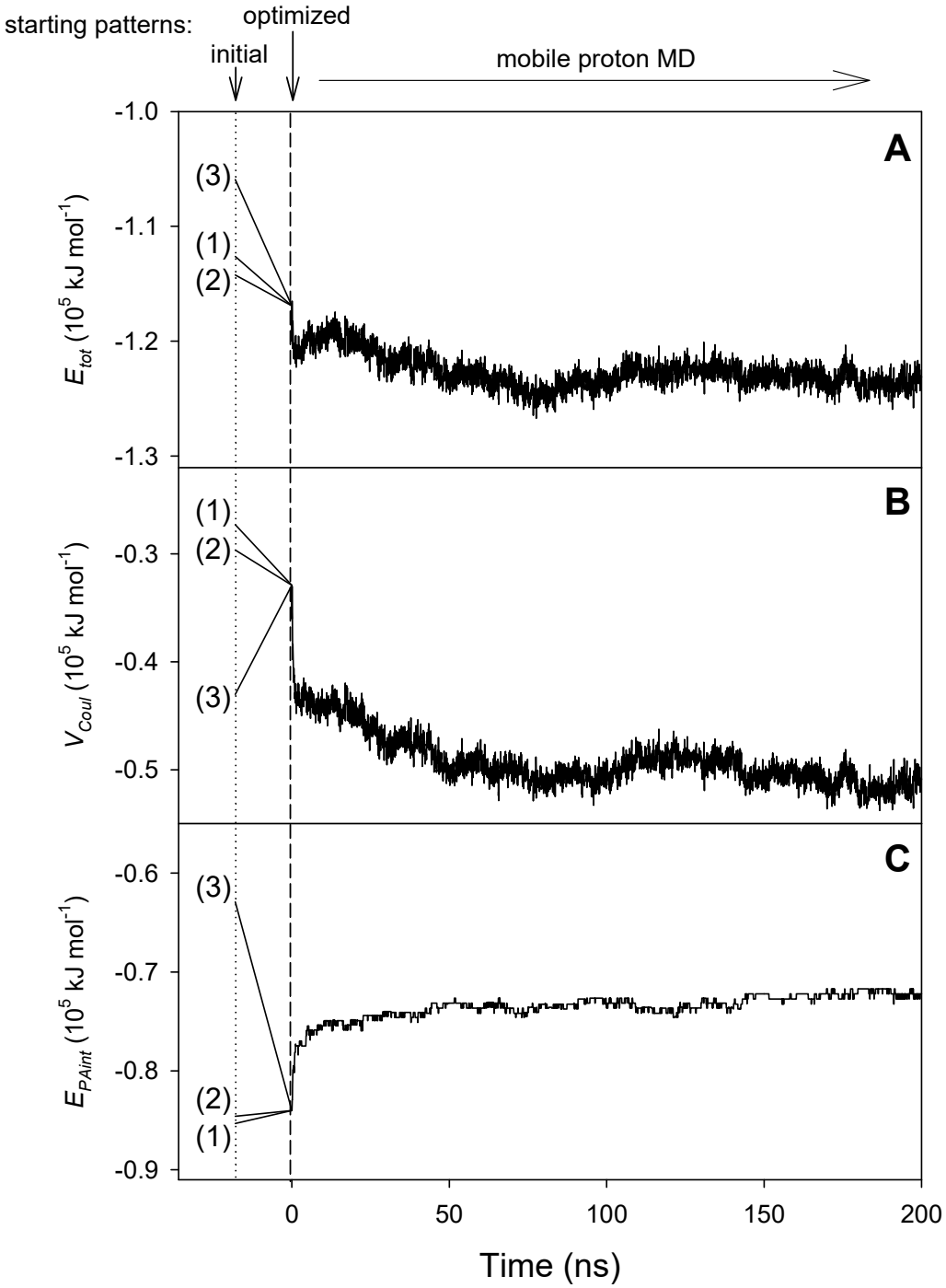

A
Arg122
[d]
[b]
Arg87
Asp86
[a]
[c]
X-ray
B
t = 0, initial charge pattern
C
Arg100+
Asp109-
t = 0, optimized charge pattern
D
Arg87+
t = 200 ns
E
Lys90+
CT-
Arg122+
Arg87+
Glu91-
Arg2+
Asp86-
t = 200 ns, close-up
F
X-ray vs. t = 200 ns

Figure 5


starting patterns:
optimized
initial
mobile proton MD
A
B
C
(1)
(2)
(3)
E_{tot} (10^5 kJ mol^{-1})
V_{Coul} (10^5 kJ mol^{-1})
E_{PAint} (10^5 kJ mol^{-1})
-1.0
-1.1
-1.2
-1.3
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
0
50
100
150
200
Time (ns)

TOC Graphic