



Proteins in the gas phase

Tim Meyer,^{1*} Valérie Gabelica,² Helmut Grubmüller¹ and Modesto Orozco^{3,4}

Proteins are complex macromolecules that evolved over billions of years to be active in aqueous solution. Water is a key element that stabilizes their structure, and most structural studies on proteins have thus been carried out in aqueous environment. However, recent experimental approaches have opened the possibility to gain structural information on proteins from gas-phase measurements. The obtained results revealed significant structural memory in proteins when transferred from water to the gas phase. However, after several years of experimental and theoretical research, the nature of the structural changes induced by vaporization, the exact characteristics of proteins in the gas phase, and the physicochemical forces stabilizing dehydrated proteins are still unclear. We will review here these issues using both experimental and theoretical sources of information.

© 2012 John Wiley & Sons, Ltd.

How to cite this article:

WIREs Comput Mol Sci 2013, 3: 408–425 doi: 10.1002/wcms.1130

INTRODUCTION

Biomacromolecules such as proteins have evolved over billions of years to display optimal functional structures under biological conditions, which are characterized by a massive presence of water, the main solvent of life. Water is, for several reasons, a crucial element in defining protein structure. Current knowledge on protein folding points to the hydrophobic effect as the driving force responsible of proteins adopting the bioactive conformation.¹ Furthermore, water screens charged residues and reduces Coulombic interactions that would otherwise destabilize the native pattern of residue contacts. Water also defines the dynamics of proteins by acting as a lubricant for protein movements, and modulates protein–protein interactions.² Last but not least, water is a key reactant in many protein-mediated reactions, acting as

general acid/base compound, as a nucleophile, or as cocatalyst stabilizing the correct orientation of reactants at the active site of enzymes. We can summarize the importance of water saying that the majority of proteins would not function outside their aqueous environment.

Correlated to function is structure and experimental efforts to characterize protein structure are naturally carried out in environments as close as possible to the native one. Recent spectroscopic or spectrometric experimental setups have, however, raised the possibility to obtain low-resolution structural information on proteins in vacuum conditions. Promise exists that X-ray free-electron laser (X-FEL) diffraction will give the means to determine the structure of proteins in vacuum at atomic resolution.^{3,4} The main interest of spectrometric methods is the possibility to separate and then characterize each component present a complex mixture, but these methods will only be of biological interest if the solution-phase structure is preserved in the gas phase.

As noted above, basic physicochemical reasoning suggests that proteins should unfold in the gas phase due to the lack of hydrophobic effect and the enhancement of long-range Coulombic interactions. Quite surprisingly, mass spectrometry (MS) experiments have reported solid evidence that at least a part of the structure of not only proteins,⁵ but also nucleic acids^{6,7} remains upon vaporization.

*Correspondence to: meyerti@gmail.com

¹Theoretische und computergestützte Biophysik, Max-Planck-Institut für biophysikalische Chemie, Göttingen, Germany

²Physical Chemistry and Mass Spectrometry Laboratory, Department of Chemistry, University of Liège, Liège, Belgium

³Joint IRB - BSC Program on Computational Biology, Institute for Research in Biomedicine; Parc Científic de Barcelona, Barcelona, Spain

⁴Departament de Bioquímica i Biologia Molecular, Facultat de Biologia, Barcelona, Spain

DOI: 10.1002/wcms.1130

Furthermore, even noncovalent complexes of proteins with other proteins,⁸ small ligands,⁹ and nucleic acids¹⁰ can stay intact in the gas phase, and evidences have been reported that at least some proteins remain bioactive after vaporization.¹¹ All these exciting and counterintuitive experimental findings have raised a large interest in determining structural changes occurring in proteins when quickly transferred from aqueous conditions to the gas phase and to determine the structures of the proteins in the absence of water. The main questions concern the extent of structural change upon vaporization, and the possibility to infer solution structure from gas-phase measurements. We will review here current knowledge on the structure of proteins in the gas phase, on the physicochemical nature of the vaporization process, and on theoretical efforts that are now made to describe the structure of proteins in the gas phase. Advances in MS of small peptides, membrane proteins,⁵ and protein assemblies⁸ have been recently reviewed and will be, for brevity sake, excluded at this place.

PHYSICAL DESCRIPTION OF VAPORIZATION

We define vaporization as the process of transferring a molecule, in this case a protein, from solution to the gas phase in a very short period of time. Vaporization is typically done in the context of MS, a very powerful technique that combines high sensitivity, high mass resolving power, and high mass measurement accuracy. The mass measurement is based on the movement of charged species under the influence of electric or magnetic fields. Therefore, the molecules need to be analyzed as ions and not as neutral entities. The most widely used experimental approach to transfer macromolecules from solution to the gas phase is electrospray ionization (ESI),^{12,13} carefully applied it is 'soft' enough to maintain most protein interactions and is available on most commercial mass spectrometers. The sample is a solution containing the molecule (micromolar concentration range). Most protein MS analyses are performed in positive ion mode (i.e., protonated proteins), for sensitivity reasons, but producing negative ions of proteins (i.e., deprotonated proteins) by electrospray is also feasible.

Electrospray Ionization

The electrospray experimental setup is as follows: the solution is injected in a capillary tip which is placed in front of the entrance of the mass spectrometer,

without touching it. Between the tip and spectrometer entrance is either air or a pure, inert gas (usually nitrogen) at atmospheric pressure. A voltage is applied on the conducting tip. As a result, the ions (the charge carriers in solution) move downfield. If a positive voltage is applied on the tip compared to the counterelectrode, the cations in the solution tend to move toward the counterelectrode. Therefore, instead of forming a meniscus that ends up dripping from the tip, the solution forms a cone that ends up producing tiny droplets enriched in positive charge carriers. This process of jet formation and droplet emission has been studied macroscopically (Figure 1a, b),¹⁴ and modeled at the atomic level (see Figure 1c–g).¹⁵ In the modeling shown in Figure 1(c)–(g), neutral droplets consisting of 7150 formamide molecules and 900 dissolved NaI units were submitted to a uniform external electric field. The droplet deforms longitudinally, producing jets that break into smaller nanodroplets that contain excess positive charges. The inset of Figure 1(e) shows the emission of a single sodium cation surrounded by only a few solvent molecules. This process is called 'ion evaporation' (see below).

The idealized system modeled above was a neutral droplet, but the droplets emitted at the electrospray tip are charged droplets, which undergo friction with the ambient gas and evaporate as they move downfield toward the counterelectrode. When the charge density at the droplet surface becomes larger than the surface tension (this is called the Rayleigh limit), the droplets become unstable and emit a secondary jet of smaller droplets that carry away some of the excess charge (Figure 1b). The production of droplets containing a single macromolecule requires only a few generations of droplet fission (depending on the initial droplet size), and occurs in the microsecond timescale. Once a charged droplet containing one macromolecule is generated, it needs to be transformed in a fully desolvated macromolecule ion.

Charged Residue or Ion Evaporation Mechanism?

The exact mechanism by which this transition occurs is still a matter of debate, and two main models are proposed. In the *charged residue mechanism* (CRM), the droplet containing the macromolecule gradually evaporates until the very last solvent molecule departs. In the *ion evaporation mechanism* (IEM), the macromolecule can evaporate from the droplet, taking away some of the droplet's charges (protons or counterions), and leaving the solvent droplet.

The current understanding is that both mechanisms exist, and which one is at stake depends on the

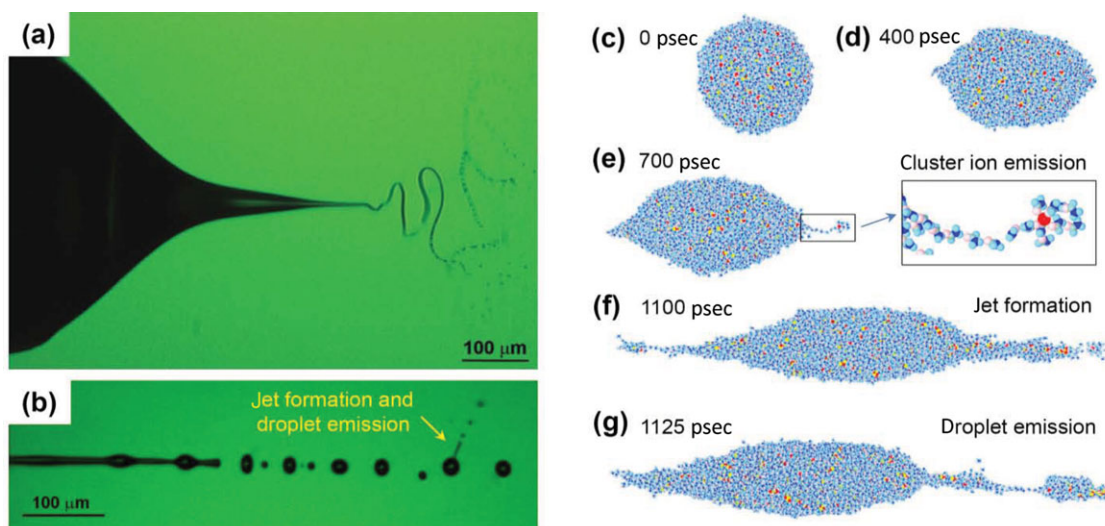


FIGURE 1 | Charged droplet emission in electrospray. (a) Photographs showing droplet emission from the electrospray tip in the burst mode. (b) Close-up on a charged filament that break up through varicose instabilities. Droplets of different sizes are produced. The beforelast droplet undergoes jet fission due to charge instability and produces secondary droplets.¹⁴ (c–g) Atomic configurations are taken from a molecular dynamics simulation of a 10 nm neutral droplet of NaI in formamide, following the application of a uniform external electric field of 0.625 V/nm. Na⁺ and I⁻ ions are shown as red and yellow spheres respectively, and formamide molecules are in blue.¹⁵ (Reproduced with permission from Ref 15. Copyright 2008, American Chemical Society.)

analyte.^{16,17} It is generally assumed that small ions (Na⁺, NH₄⁺) are produced via the IEM,¹⁸ following a process similar to that illustrated in Figure 1(e), where a single sodium ion was seen to escape from the parent droplet (see also Figure 2d). Actually, when this single ion carries away a few solvent molecules that will need to evaporate later, the distinction between evaporation of a single, solvated ion and a fission event that would produce a droplet constituted of a few solvent molecules, gets blurred¹⁹ because it actually depends on from how many solvent molecules we consider minimum to talk about a ‘droplet’.

It is also commonly assumed that large macromolecules that are initially folded in the droplets will be charged via the CRM (see Figure 2c): the solvent evaporates gradually, some extra charge carriers evaporate by the IEM, and the maximum number of charges remaining on the macromolecule is therefore fixed by the maximum number of charges bearable by a solvent droplet having the same size as the macromolecule. Experimentally, the CRM is supported by (1) the good correlation between the average charge state of globular proteins and the protein surface area^{20–22} and (2) by the fact that higher charge states can be obtained by using solvents with higher surface tension than water.²³

Note that within the CRM the maximum number of charges supported by a protein is, in principle, fixed by the Rayleigh limit,²⁴ and depends on its

accessible surface area. However, charging beyond the Rayleigh limit of globular proteins has also been observed,²⁵ which can only be rationalized by assuming that a population of macromolecules does not stay globularly folded, but gradually extends and protrudes out of the droplet. It is now confirmed that some high surface tension ‘supercharging’ agents actually denature the protein during the electrospray emission process,^{26,27} and the unfolding favors higher charge uptake.

In summary, vaporization of folded proteins seems to follow CRM, leading to a population of charged states. The most abundant state is determined by the Rayleigh limit of a droplet with the same surface area as the solvated protein, higher charged states reflect the presence of partially or completely unfolded conformers. If vaporization conditions become more drastic the relative importance of high charged peaks increases indicating that protein conformational ensemble has moved toward unfolded states, and as this happens the importance of desolvation via IEM increases.

A clear example of protein vaporization under different conditions is shown in Figure 2(a) and (b), which represents the electrospray mass spectra of myoglobin recorded from a solution where it is folded (pH = 7, Figure 2a) or unfolded (pH = 2, Figure 2b²⁸). Experimentally, both the average charge and the total ion current are much higher for the

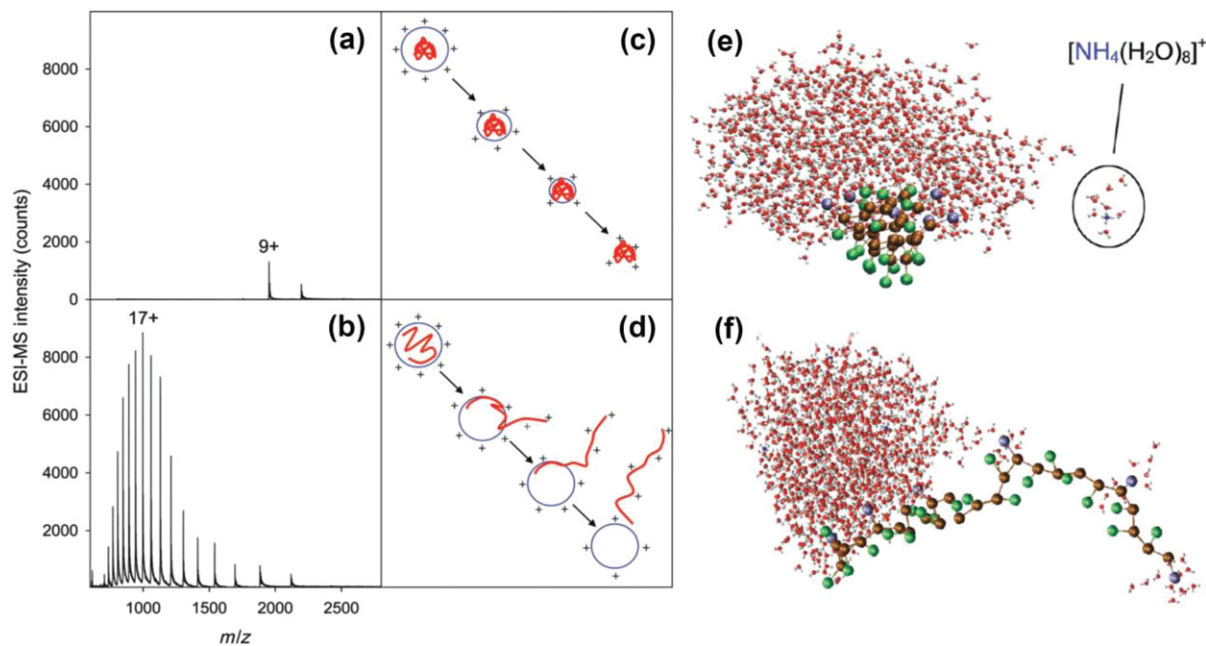


FIGURE 2 | Final production of desolvated macromolecule ions in electrospray. (a, b) representative electrospray mass spectra of a folded protein (panel a: myoglobin at pH 7) and of an unfolded protein (panel b: myoglobin at pH 2). (c) Cartoon illustrating the charged residue mechanism (CRM). (d) Cartoon illustrating the ion evaporation mechanism (IEM). (e, f) Simulation data from molecular dynamics runs for (e) a folded/hydrophobic polymer and (f) an unfolded/hydrophobic polymer, each in 1000 water molecules and excess NH_4^+ as charge carriers. (Reproduced with permission from Ref 28. Copyright 2012, American Chemical Society.)

unfolded protein than for the globular protein. This is interpreted as ion desolvation taking place via CRM (Figure 2c) and IEM (Figure 2d), respectively. Snapshots from molecular dynamics (MD) simulations show what happens to a folded hydrophobic polymer (Figure 2e) and to an initially unfolded hydrophobic polymer (Figure 2f). The folded polymer stays in the evaporating droplet and sometimes one small charge carrier is expelled (see the solvated ammonium ion in Figure 2e). In contrast, the unfolded polymer gradually extends and protrudes out of the droplet, taking away some of the charge carriers (in blue) and only a few water molecules, and thereby releasing the charge repulsion on the overall system. Hydrophilic polymers, however, tended to stay in the droplet, independently of whether they were initially folded or unfolded.

The Charged Residue Mechanism

We will center our following discussion on the nature of the structure obtained upon desolvation (vaporization) by the CRM, because this mechanism is more likely to maintain the structure of the solvated protein. Ensembles obtained by the IEM might represent the unfolded state of the protein in the gas phase, but such ensembles are unlikely to be useful to repre-

sent the nature of the unfolded state of the protein in aqueous solution.

Several aspects of the CRM process may influence the structure of the protein. A major issue is the evaluation of the temperature in the droplet. Evaporation is an endothermic process and, as vaporization proceeds, the droplet temperature decreases in a way that depends on solvent properties. For example, experiments in the presence of viscous additives such as glycerol have shown evidence of protein thermal denaturation,²⁷ attributed to higher droplet temperatures reached at the final desolvation steps.²⁹ In contrast, other additives such as imidazole were experimentally found to preserve some fragile native complexes,³⁰ which was attributed to evaporative cooling properties of these additives. In summary, there are uncertainties on the internal temperature of ions when just emitted from the droplets, and it clearly depends on the experimental conditions.

Even more uncertainties exist on the temperature of the fully desolvated protein. The ionization chamber is a zone in the instrument with pressures decreasing from atmospheric pressure to 10^{-5} mbar or lower. In this transition zone, electric fields carry the molecules toward the analyzer. Collisions with neutral molecules may transfer energy to the molecule, increasing its effective temperature. In zones where

collisions are sufficiently numerous, a quasithermal stationary regime with both activating and deactivating collisions can be established. The temperature of ions in these zones is equal to the gas temperature only if the axial component of the ion velocity due to the electric field is negligible compared to Brownian motion, and higher as the electrophoretic friction becomes significant.

Charge Location

A second major source of uncertainty in MS experiments is the nature of the charge state of the protein. The absolute charge of a protein ion is an integer number that can be determined experimentally without uncertainty even for large proteins. However, experiments do not indicate the placement of the charges among the acidic and basic sites of a protein. Locating charges in the protein is not trivial as there are usually many more available basic side chains than actual charges on a native protein, generating an enormous number of possible states compatible with a given total charge. For example, assuming a typical protein such as lysozyme (see Table 2) with $n = 19$ strongly basic sites (Arg, Lys, His, N-ter) and $q = 5$ protons to distribute among them, a total of $n!/((n - q)! \cdot q!) = 11,628$ states need to be considered to determine the final charge location, a number that would significantly increase if also less basic residues (Pro, Trp, Gln)³¹ or zwitterionic states³² were taken into account. Several of these potential states can be discarded as they would lead to very strong electrostatic repulsion, but many others are quasidegenerated and can coexist in the ensemble. Furthermore, as ions progress in the spectrometer, charge migration can happen to achieve more stable states with lower Coulombic repulsion, further complicating the definition of the chemical topology of the protein ion. A reasonable approach is to assume that the major charge state minimizes the Coulombic repulsion in the (known) solution structure^{33,34} or in a relaxed form of it.³⁵ Finding this state within the vast number of possible ones can be efficiently done by Monte Carlo algorithms.³⁴ In cases where several states are quasidegenerated in energy, an ensemble of different charge states should be considered.³⁶ The urge to find the lowest energy charge location is somewhat relieved by the finding that several low energy charge locations yield similar structural ensembles.^{35,37}

The main alternative to electrospray ionization are laser desorption ionization techniques where nanosecond-long laser pulses are used to transfer the sample into the vapor phase. Most widespread is the use of UV lasers in combination with crys-

talline, organic matrices that embed the analyte (protein) and whose absorption matches the wavelength of the laser, termed matrix-assisted laser desorption ionization or MALDI. The UV laser pulse is energetic enough to break covalent bonds of the matrix and generates a plume of sample, matrix, radicals, and ions.³⁸ Sample ions that are detectable by the MS device are formed by either the 'lucky survivor' or the gas-phase protonation model, depending on matrix type.³⁹ The 'lucky survivor model' becomes predominant with increasing molecular weight of the sample³⁹ and ionization is achieved by stochastic charge separation during the ablation process. The gas-phase protonation model assumes that ablation creates initially neutral cluster, which are then charged by gas-phase collisions with ions stemming from the disintegrated matrix.

UV-MALDI matrices are chemically quite distinct from the aqueous environment, which hampers their use in the analysis of intact noncovalently bound protein complexes.⁴⁰ Infrared laser tuned to the vibrational frequency of the O-H stretch around $3 \mu\text{m}$ allow to use ice⁴¹ and liquid water⁴² as matrices for desorption, the latter method is known as laser-induced liquid beam desorption and ionization (LILBID, Figure 3b), although alternative methods use liquid droplets instead of beams to minimize sample consumption.⁴²

In LILBID a fast-flowing (~ 40 millisecond⁻¹) liquid microjet of about $10 \mu\text{m}$ in diameter⁴³ is hit by a short (nanosecond) laser pulse that trigger a sudden (picosecond range) supercritical phase transition within the surface layer, inducing a shock wave that disperses the solution into hot, nanometer- to micrometer-sized droplets. These nanodroplets have sufficient thermal energy to evaporate water almost completely without further heating so that analyte-containing droplets may produce MS-competent proteins ions, the exact mechanism of ion formation is, however, still controversially discussed. The absolute charges of protein ions rarely exceeds one⁴³ to three,^{42,44} which is attributed to the fast ion recombination of high charge states in the context of a reduced dielectric constant in the expanding supercritical phase.⁴² Charged droplets that undergo incomplete ion recombination are the 'lucky survivors' that generate detectable protein ions. The 'liquid dispersion model' takes the heterogeneity of the system, the coexistence of liquid and supercritical water, explicitly into account.⁴³ Recent computer simulations have studied the desorption process in atomistic detail, and suggest local screening of salt ions and neutralizing ion fluxes via transient water bridges during droplet detachment as further important factors

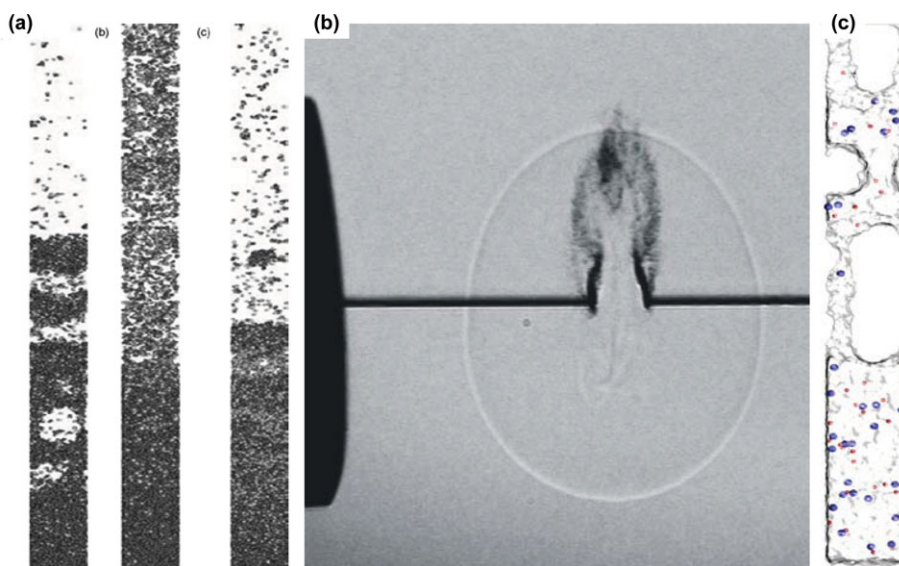


FIGURE 3 | Laser induced ablation from a solid crystalline support matrix (a) and from a liquid water beam (b). UV laser cause photothermal (a left) and photochemical (a middle, right) excitation, the latter leads to bond cleavage formation of ions and radicals. Computer simulations of the supercritical phase expansion observed in laser-induced liquid beam desorption and ionization revealed transient water bridges that facilitate ion recombination (c). [Figure 3(a) reproduced with permission from Ref 38. Copyright 2007, Elsevier. Figure 3(b) reproduced with permission of Prof. Bernd Abel. Figure 3(c) reproduced with permission from Ref 45. Copyright 2009, Dr. Frank Wiederschein.]

that limit the total charge of the protein ions to low count numbers (Figure 3c). The simulations further show a remarkably strong transient and reversible compression of the protein analyte by the shock wave briefly before desorption.⁴⁵

EXPERIMENTAL APPROACHES PROVIDING STRUCTURAL INFORMATION IN THE GAS PHASE

In addition to the mass measurement, mass spectrometers can nowadays perform numerous other operations on the ions: gas-phase electrophoresis [(more commonly named ‘ion mobility spectrometry’ (IMS)); collisional activation; reactions with electrons, ions, molecules, or photons. Describing each of these techniques in detail is beyond the scope of the present overview, but in a nutshell the general aim is to obtain structural information on each species, either at the level of the covalent structure (i.e., for a protein, the amino acid sequence) or at the level of the non-covalent structure (i.e., for a protein, the secondary, tertiary, or quaternary structure). Restricting the discussion to the techniques providing information on the noncovalent structure of ions in the gas phase, three categories of experimental techniques can be distinguished.

First, there are techniques relying on the reactivity of the protein or protein complex in the gas phase. They encompass ion–electron reactions [electron capture dissociation (ECD)]^{46–48} and ion–molecule reactions such as hydrogen/deuterium exchange (HDX-MS).^{49,50} In ECD, it is believed that cleavage occurs at all sites, but that fragments separate only if the internal energy of the ion is able to overcome the remaining interactions between them. Hence, fragments involving folded regions in the gas phase are actually not observed. The interpretation of HDX-MS in terms of solution-phase structure is more elusive. In solution, regions that are protected from the solvent exchange protons more slowly than exposed regions, but this is no longer true in the gas phase because a relay mechanism requiring some proximity between two exchangeable sites is at stake.⁵¹ To sum up, in addition to uncertainties regarding the difference between solution phase structures and gas-phase structures, there are doubts on the reaction mechanisms themselves, and hence on how to interpret the results. Thus, the above-mentioned techniques are good indicators that different conformations are indeed present⁵² and that their relative abundance changes in response to changes in the starting solution (indicating some memory of the solution phase structure in the gas-phase ones⁵³), but do not necessarily allow us to assign a structure to these conformations.

TABLE 1 | Timescales of Common Experimental Methods for Study of Gas-Phase Proteins

Structure Probing Method	Timescale of Probing
HDX-MS	10 millisecond–several seconds
Multiple-pulse dissociation (MPD) spectroscopy (IRMPD, UV-Vis)	1 millisecond–several seconds
Electron capture dissociation (ECD)	1 millisecond–1 second
Ion mobility spectrometry (IMS) ⁶⁹	1–10 millisecond
Single-pulse spectroscopy (IR, UV-Vis, photoelectron, fluorescence)	picosecond–nanosecond (depends on laser)
X-ray free electron laser (X-FEL) ³	2–100 femtosecond

A second category is that of gas-phase spectroscopy techniques, encompassing infrared spectroscopy,^{54,55} UV-visible (UV-Vis) spectroscopy,⁵⁶ fluorescence spectroscopy,^{57,58} and photoelectron spectroscopy.⁵⁹ The structural interpretation of experimental data is based on the matching between the experimental and the theoretical spectra expected for a proposed gas-phase structure. In particular, infrared spectroscopy performed at low temperature provides high-resolution vibrational data, which can therefore be confidently assigned to a single structure, but the real power of the technique has been recently demonstrated only for relatively small systems such as peptides and synthetic complexes.^{60–62}

A third category encompasses IMS techniques. The common principle is that ions of different shapes will drift with different velocities when placed in a chamber filled with gas and under the influence of an electric field. By definition, the mobility (K) is the proportionality factor between the steady state velocity and the electric field:

$$v = K E. \quad (1)$$

In the low field limit, the mobility is expressed by the Mason–Schamp equation⁶³:

$$K = \frac{3}{16} N_0 \frac{q}{N^2} \left(\frac{2\pi}{\mu kT} \right)^{\frac{1}{2}} \frac{1}{\Omega}, \quad (2)$$

where N is the gas number density, N_0 is the gas number density at standard temperature ($T_0 = 273.16$ K) and pressure ($p_0 = 760$ Torr), q is the ion charge, μ is the reduced mass ($\mu = mM/(m + M)$, where M is the mass of the gas and m is the mass of the ion), k is the Boltzmann constant, T is the gas temperature, and Ω is the collision cross-section (CCS). Drift tube IMS instruments operate in the low field limit, and the CCS is derived from first principles (all parameters in Eqs (1) and (2) are known or measured). Commercially available traveling wave IMS instruments

operate close to the low field limit and the CCS is obtained following instrument calibration with samples studied previously by drift tube IMS.

The last category contains single-particle X-ray diffraction. Femtosecond short and extremely bright X-ray pulses originating from free-electron lasers (XFEL) are shot at nanostructures like crystals, viruses, or single molecules and generate characteristic diffraction pattern.³ A recent study on nanocrystals of the photosystem I yielded diffractions at 8.5 Å resolution⁶⁴ and the authors expect that further increase in peak-brilliance and repetition rate, and shorter pulse lengths, will soon allow analysis of single molecules at atomic resolution. Single molecule imaging at nanometer resolution provides structural ensembles that represent all accessible states in a single experiment⁶⁵ and are, therefore, inherently richer in information than single crystal structures. The question whether or not these ensembles contain information on biologically relevant processes will be addressed in this review.

All these techniques probe the structure of the ions after desolvation but, as noted above, the hope is that this information is also relevant to characterize the sample in the initial solution. Modeling of experimentally relevant gas-phase protein structures should take everything into account that happens to the protein between the onset of desolvation and the moment their structure is probed by one of the above-mentioned techniques. Table 1 summarizes typical timescales of MS experiments. In addition to the time required for structure probing there are time delays between total desolvation and the beginning of the probing experiment. Another issue is the ion temperature, which equals the container temperature only if sufficient equilibration time is given. The time required to equilibrate depends on the pressure inside the container, and ranges from several milliseconds in multipole ion traps, where the pressure is around 10^{-3} mbar,⁶⁶ to several seconds in penning traps where the pressure is $<10^{-9}$ mbar.⁶⁷ Internal ion temperatures can, moreover, increase temporarily in those parts of

the instrument where ions undergo collisional activation with ambient gas.⁶⁸

The uncertainties on the vaporization temperature of the protein are carried further downstream into the mass spectrometer. Depending on instrument design, there can be high-pressure regions called ‘collision cells’ (typical pressure: 10^{-3} mbar, typical transit time: microseconds) where a few activating collisions will occur, changing the temperature of the protein. The internal energy distribution is not necessarily Boltzmann like, which renders the application of thermodynamic equations problematic. The ‘real’ temperature is therefore replaced by an ‘effective’ temperature which relates to a specific weighted average of the internal energy distribution and which describes the system’s behavior with respect to a particular observable. In general, the effective temperature of an ion increases when it is accelerated to cross a region of higher pressure in the mass spectrometer. This also holds for IMS (typical pressure: 1 mbar, typical transit time: milliseconds). Ions transiting from a low-pressure region to the mbar-pressure mobility cell will undergo collisional activation when they encounter the gas curtain opposing its displacement.^{70,71} Then, depending on mean ion velocity, the ion temperature inside the mobility cell lowers to a value either equal to the bath gas temperature (this is the case in custom-made drift tube ion mobility spectrometers) or higher than the bath gas temperature (this is the case of commercially available traveling wave ion mobility cells).

In summary, knowledge of instrument hardware and experimental parameters is required to estimate not only the protein structure during the structural probing experiment but also, because the goal is to relate that structure to the one initially present in solution, to estimate all sources of changes between the solution and the gas-phase probing. This implies knowing how much time ions spend at each temperature, during desolvation and after desolvation. Numbers available in the literature are scarce. Ion temperatures just after desolvation might range from 250 to 650 K (depending on activation or evaporative cooling, and on ion mass—high mass ions coming out at lower effective temperature than low-mass ions^{72,73}), but their temperatures in the source could increase due to subsequent collisional activation. Activation in collision cells could lead to high internal energies (probably non-thermal internal energy distributions). A recent paper by Robinson’s group⁷⁴ reported MD simulations with temperature gradients from 300 to 800 K, but for a short time (10 nanosecond), to model the dissociation of protein complexes in a collision cell. When it comes to modeling structures in ion

mobility experimental conditions, the tube’s temperature (usually room temperature) can be used, but then the challenge is the millisecond timescale of the experiment. An alternative approach followed by our groups is to perform very large MD simulations in a few discrete temperatures covering in principle a variety of experimental situations.^{35,75} In any case, we should recognize problems in defining the real experimental conditions to be then introducing as input in theoretical calculations.

THEORETICAL APPROACHES FOR THE STUDY OF PROTEIN STRUCTURE IN THE GAS PHASE

As described above, current experimental approaches yield only low-resolution data on structures of proteins and complexes in the gas phase, mostly CCS (see above), which are not by themselves informative enough as to fully characterize protein structure in the gas phase. This means that structural details need to be added from theoretical approaches, which in turn need to be validated by their ability to reproduce available low-resolution experimental information. This is typically done by back computing experimental observables, such as the CCS, from the suggested ensembles of structures obtained by simulation.^{76–78} As noted above, this process is not rigorously defined and some caution is necessary, since the same structural ensemble processed with different algorithms can provide quite different CCS (see Table 2⁷⁹).

The exact gas-phase structure of small, rigid molecules can be determined from high-level quantum mechanical calculations, but these techniques are in general not applicable to the study the conformational space of large and flexible macromolecules—such as proteins—on the nanosecond to multi-microsecond timescale. Classical mechanics thus needs to be used in conjunction with a sampling method, typically force field (FF) based MD, to study the structural properties of proteins in the gas phase.

Performing a MD simulation requires: (1) a starting configuration for the system to study, (2) a functional connecting the geometry of the protein with its energy (the FF), and (3) selection of the simulation conditions (temperature, pressure, and the statistical ensemble). Most MD simulations use the equilibrium geometry of the protein in solution, determined from X-ray or NMR data, as starting configuration. Depending on the question to be addressed (studying the desolvation process or the structure of the protein in the gas phase), the starting system might

TABLE 2 | Experimental and Computed Collision Cross-Sections of Selected Well-Studied Proteins (Data used from Tables 1 and 2 of Ref 79.)

Protein Name	MW (Da)	PDB ID	Collision Cross-section (Å ²)			
			Experimental (Charge)	Projection Approximation	Exact Hard Sphere Scattering	Trajectory Method
Melittin	2846	1MLT	544 (+3)	493	586	574
Bpti	6512	6PTI	770 (+4)	749	914	891
Ubiquitin	8565	1UBQ	791 (+4)	885	1088	1055
Cytochrome C	12355	1HRC	1217 (+5)	1056	1317	1310
Lysozyme	14305	1DPX	1300 (+5)	1172	1461	1468

contain or not solvent molecules, ions, and cosolutes, in addition to the protein. A common setup to study protein vaporization contains the protein in its most abundant ionic state in solution, immersed in a large drop of water. Finding of the most probable ionic state of all titrable residues is nontrivial, as discussed later. Numerous additives, ranging from salts, polar solvents,⁵⁸ polyols, and aminoacids⁸⁰ to organic detergents,⁸¹ have been used experimentally to stabilize the native structure during transfer into the gas phase and they can be introduced into the simulation. For small, volatile additives the stabilizing effect stems most probably from enhanced evaporative cooling^{30,82} whereas certain detergents such as acetyltrimethylammonium bromide are capable of encapsulating proteins and protein complexes in micelles.⁸³ Van der Spoel and colleagues⁸⁴ simulated micelle stabilization of the transmembrane domain OmpA171 embedded in a dodecylphosphocholine soluble myoglobin (Mb) embedded cetyl trimethylammonium bromide (CTAB),⁸⁵ which matches the experimental setup of Sharon et al.⁸³ Micelles were surrounded by water layers of varying thickness to mimic their gradual desolvation during the ESI process. During this process the Mb.CTAB micelles invert to form reverse micelles, as observed in the experiment. Both, normal and reverse micelles protected their content from the harsh vacuum environment and significantly reduced structural distortions due to vacuum exposure. This finding strongly supports the use of XFEL crystallography for the determination of membrane protein structures.

Simulations intended to reproduce desolvation and to study the structural properties of the gas-phase proteins are typically done assuming zero pressure. This favors water drifting away from the protein but assumes that transfer from normal pressure to vacuum happens instantaneously. In principle, the constant energy ensembles (NPE, NVE) seem the most natural ones to simulate MS experiments. However,

as noted above, water evaporation cools the droplet until evaporation stops at approximately 225 K,⁸⁶ impeding the removal of water molecules tightly bound to the protein and thus generating noise in the m/z signals. Experimentally, this is overcome by introducing external heat to keep the droplet temperature constant and drive the protein to complete desolvation. Simulation protocols that aim at reproducing the complete desolvation process were run at constant energy and starting from water shells of varying thickness,⁸⁷ or at constant temperature, where the droplet is thermally coupled to a heat bath (Meyer et al., to be published). Once the protein is desolvated most authors assume constant temperature conditions, using sometimes a range of effective temperatures,^{35,87} and take care that thermal energy is effectively translated into internal energy. Selection of effective temperature for MD simulation of spectrometric experiments is a rather arbitrary decision, but results are quite robust to changes in this parameter and simulations we performed at temperatures from 300 to 400 K provided similar results,^{35,75} in agreement with the thermostability of unsolvated proteins.⁸⁸

The decision on the FF to be used to simulate a protein in the gas phase or during the vaporization process is far from trivial. For efficiency reasons most FFs used in MD studies of proteins are pair-additive, ignoring explicit consideration of polarization effects, which are indirectly captured by increasing the magnitude of the atomic charges by around 20%. This approach is effective to capture the average polarizing effect of water, but will probably overestimate the magnitude of electrostatic interactions in the gas phase. The commonly used Tip3P water model has, for example, a fixed dipole of 2.35D,⁸⁹ which corresponds to the solution phase value but is significantly larger than the 1.85D measured for single *in vacuo* molecules.⁹⁰ Similarly, bonded-terms in proteins, particularly torsions, are typically adjusted during FF development to reproduce torsional barriers in

solution, and there is no guarantee on their validity in gas-phase conditions. To our knowledge, no gas-phase FF for proteins or nucleic acids has yet been developed, and thus most groups use standard solution FFs without any correction, or just scaling by 0.8 the atomic charges in neutral residues.⁹¹ Clearly, development of specific gas-phase FFs, ideally incorporating polarization terms, will increase dramatically our ability to represent the properties of proteins during the vaporization process and in the gas phase.

As noted above (see Table 1), vaporization happens in the microsecond timescale, whereas the desolvated protein travels across the spectrometer tube for periods ranging from several microseconds to milliseconds. Considering that Newton's equations of motion need to be solved every femtosecond, reproducing the travel of one single molecule along the spectrometer requires 10^{12} integration steps, which, considering the nature of the system, cannot be easily spread over multiple processors, limiting the use of massively parallel supercomputers and the computational efficiency of the calculation. In fact, the largest simulations of macromolecules in the gas phase are in the multi-microsecond range,⁷⁵ far from the required millisecond scale. Fortunately, inspection of such trajectories shows that most of the conformational transition happens very early in the trajectory and that after such a change (occurring in the nanoseconds scale) the protein becomes very rigid, suggesting that the large gap between simulated and experimental time might be not a dramatic problem. To improve the representation of the structural diversity in each ion peak, we found that better sampling is achieved when performing multiple sub-microsecond trajectories instead of a single multi-microsecond trajectory. The gas-phase structures that are experimentally detected correspond most likely to metastable conformations in which the protein is trapped during the desolvation process, and not the real absolute minima in the gas phase. If information on the native (solution) structure is wanted it is desirable to tune experimental setups toward minimum exposure times to vacuum prior to analysis, this minimizes unfolding and, in turn, will favor convergence between simulation and experiment.

SIMULATION OF PROTEINS IN THE GAS PHASE

Simulations of the Vaporization Process

Simulations of the vaporization process of folded proteins have demonstrated that it typically follows the CRM paradigm (Figure 1), with water evaporating

very fast at the beginning—both as small droplets or as individual molecules—and more slowly later, because the remaining water molecules are those tightly bound to the most polar groups of the protein. It is worth noting that temperatures that will yield very quickly to thermal unfolding in aqueous conditions do not destroy the structure of proteins *in vacuo*. The two reasons are clear⁹²: (1) in solution, hot water molecules hit the structure inducing local distortions that yields to protein unfolding, whereas in the gas phase these highly activated water molecules migrate to the vacuum during vaporization without affecting protein contacts; (2) in aqueous solution temporarily lost protein–protein interactions, for example, residue–residue hydrogen bonds are replaced by water–protein interactions, stabilizing open, unfolded states. This compensatory effect vanishes in the absence of water.

Clearly, the ionization state of every residue of the folded protein in water is well defined at a given pH, but at some point of the vaporization process acidic side chains capture and basic side chains yield protons to the water molecules (or additives) in the vicinity of the protein. It has been shown that few water molecules are sufficient to stabilize zwitterions in vacuum,⁹³ hence the proton transfer occurs only toward the end of the evaporation period when these charged sites become exposed and accessible to proton carriers. The exact kinetics and the thermodynamics of the process are unknown. For example, even in the common positive ion mode some Asp and Glu side chains may remain charged when stabilized by nearby positive charges and form salt bridges that significantly stabilize the native structure, which should be taken into account in simulations. For example, a recent theoretical study of three small systems (Trp-charge, bradikinin, angiotensin) suggests that zwitterions are common at low charge states.³² Thus, the definition of the set of residues that need to be neutralized or remain charged in the simulation has to be done with much caution.^{57,94} The standard approach is to protonate the required number of basic side chains in a way that the electrostatic repulsion of the starting conformation is minimized.^{34,35,95,96} Several charge states (i.e., alternative forms to distributed neutral and anionic acidic residues) might be followed simultaneously in parallel MD trajectories, but no change in the protonation pattern is allowed *a priori* during the classical trajectories. Proton shuttling can be addressed in *ab initio* MD simulations or QM/MM simulation; however, these approaches are computationally too costly to simulate tens of nanoseconds of the evaporation process. More efficient approaches include

the multi state empirical valence bond method,⁹⁷ the QHOP-MD⁹⁸ and related⁹⁹ methods, or the λ -dynamics approach¹⁰⁰ that was recently extended to explicit water constant pH simulations.¹⁰¹

The Gas-Phase Equilibrium State

It is clear that if a protein is kept in the gas phase for very long periods of time, it will reach its most stable vacuum conformation, a structure with no memory of that in solution. Water is, by definition, hydrophilic, whereas vacuum is perfectly hydrophobic, and thus the distribution of polar and apolar residues on the protein is expected to complete change upon transfer to vacuum.¹⁰² In solution, electrostatic interactions between charged surface residues, which constitute $\sim 1/3$ of a typical protein surface, are shielded by the solvent. These interactions, either attractive or repulsive, are greatly enhanced in the absence of solvent. For example the numerous salt bridges that can be formed by the KIX protein stabilize its structure only moderately in solution ($\Delta G^{\text{folding}} \sim -4.8 k_B T$) but significantly *in vacuo*⁴⁸ whereas ubiquitin, which is more stable in solution ($\Delta G^{\text{folding}} \sim -16.7 k_B T$) but has a lower salt bridge density, partially unfolds on the subsecond timescale in vacuum.¹⁰³ Typically, the protein net charge in solution differs from that in the gas phase, leading to a clear unbalance between electrostatic interactions. However, even for the few known proteins that maintain solution net charge in the gas phase, such as the Trp-Cage,^{94,104} changes of intraprotein interactions are important, as reflected, for example, in an increase in the number of internal hydrogen bonds and the loss of apolar interactions, leading to the exposure of apolar residues to the exterior.^{35,104} The tendency of relaxed gas-phase structure to maximize the number of favorable hydrogen bonds becomes especially clear looking at the helices, which are marginally stable in solution at room temperature but nevertheless resist without problems temperatures of at least 725 K when desolvated.⁸⁸ The aligned hydrogen bonds of helices generate large dipole moments and thus destabilize helices in the absence of counteracting charges, making charge and charge location dominant factors in secondary structure propensities *in vacuo*.^{88,105}

The *In Vacuo* Native Structure

The determination of the most stable conformation of proteins in the gas phase might be only of academic interest, as it most likely has little relationship with that of the protein in solution. Fortunately, the drift

time of the experiment is typically much shorter than the time required to evolve from the solution structure to the minimum energy gas-phase structure. For example, a small protein ion such as ubiquitin⁷ requires several hundred milliseconds to equilibrate to its more stable conformation in the gas phase,^{103,106} whereas a normal electrospray experiment takes a few milliseconds, meaning that the experiment is detecting a metastable conformation whose fold is probably similar to that of the solution structure. Low CCSs could in principal be obtained not only from native-like structures but also from similarly compact but nonnative conformations. Proton stripping experiments have shown that extended structures can indeed collapse to compact but not necessarily native-like structures; however, the observed *in vacuo* folding pathway is so rugged that a complete collapse would be frustrated under normal conditions.¹⁰⁷ Accordingly, experiments with proteins injected under folding and unfolding conditions exhibit quite different CCSs^{108–110} and clearly support the hypothesis of an enduring memory of aqueous structure in gas-phase structures detected experimentally. In the following, we will discuss the nature of this ‘meta-stable gas phase structure’ of proteins, which Arteca and Tapia¹¹¹ termed ‘*in vacuo* native structure’ (IVNS).

Figure 4 is hypothetical, but it is useful to summarize, in a very simplified way, the changes to the free energy surface of a protein upon transfer from aqueous solution (blue) into the gas phase (magenta, yellow). Native-like structures are typically compact with hydrophobic residues buried inside the core, whereas polar residues are exposed on the surface. Upon unfolding toward extended structures, hydrophobic residues become inevitably exposed, which is, due to the hydrophobic effect, highly unfavorable in aqueous solution (blue) but energetically not penalized in the gas phase (magenta, yellow) where the hydrophobic effect becomes effectively inverted. Whether or not extended structures get populated depends on the total charge on the proteins. The absence of water significantly strengthens existing electrostatic interactions such as hydrogen bonds and salt bridges and stabilizes the native-like IVNS. This stabilization is counteracted by repulsive Coulomb interactions. If the total charge is low (magenta line) the net effect of unshielded electrostatics is still stabilizing, whereas at high charge states (yellow line) the Coulomb repulsion (ΔE_Q) becomes dominant and destabilizes the IVNS from a point where the unfolding barrier can be overcome spontaneously. Within the extended structures one may distinguish between the

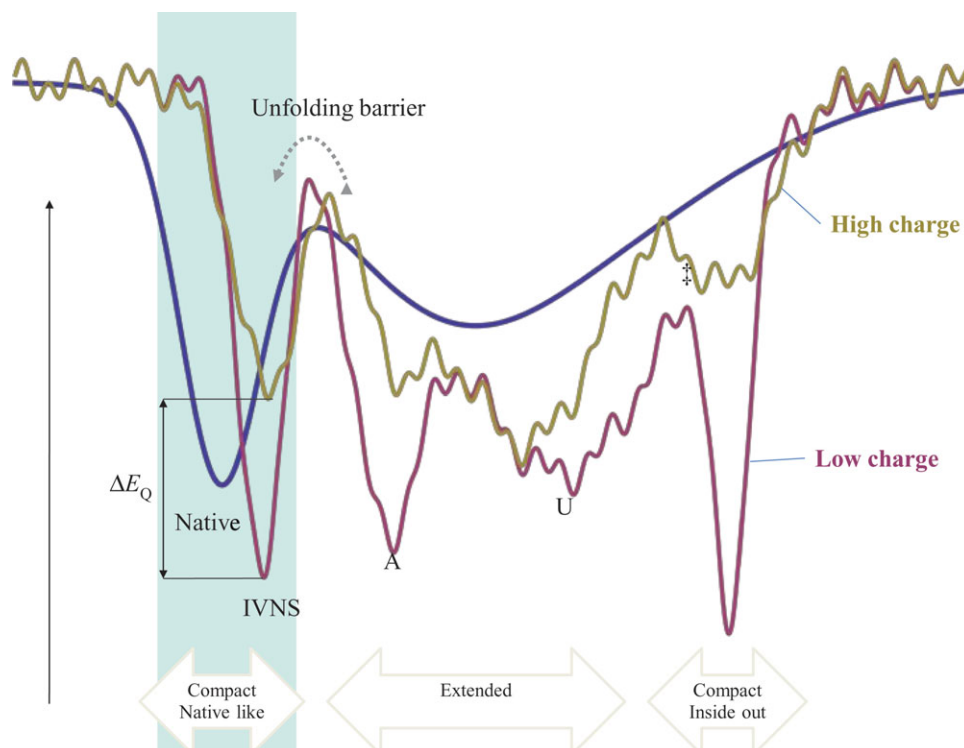


FIGURE 4 | Sketch of the expected changes to a protein's free energy surface upon transfer from aqueous (blue) to the gas phase (magenta, yellow). Extended (partially (A) and completely (U) unfolded) and reversed structures become accessible in absence of the hydrophobic effect. Whether or not they become populated depends on the charge state magenta vs. yellow) and whether the unfolding barrier can be surmounted. Coulomb repulsion (ΔE_Q) destabilizes compact structures [*in vacuo* native structure (IVNS), inside-out] and drives highly charged ions into extended states, low-charge ions may fold into the hypothetical inside-out conformation; however, the folding pathway (\ddagger) may be entropically blocked.

partially unfolded ones where only few key interactions are broken (labeled as 'A' in the figure), and the fully extended structures (labeled as 'U' in the figure). Low charge states remain partially folded since this preserves many favorable secondary interactions, whereas high charge states proceed to fully unfolded structures to maximize charge separation. Theoretical considerations suggest that so-called inside-out conformations where polar residues cluster inside the protein to maximize hydrogen bonds and salt bridges, whereas hydrophobic residues (and residues carrying unbalanced charges) reside on the protein surface, are energetically the most favorable for low charged ions, but the large barriers separating the IVNS from the compact inside-out conformation implies that the protein will not reach the later conformation until very extended periods of time (longer than the scale of the experiment, and much longer than simulation scale).

Experiments and computer simulations support the hypothetical picture drawn in Figure 4. Long MD trajectories of moderately charged protein ions suggest that there are around 30–37%

more intramolecular hydrogen bonds in vacuum than in solution.^{35,112} Furthermore, the absence of competing water molecules renders hydrogen bonds on the protein surface significantly more stable (~ 5 kcal/mol¹¹³), which leads to the observed extraordinary stability of helices in vacuum⁸⁸ and to the stabilization of the surface-collapsed IVNS. The network of hydrogen bonds and other electrostatic contacts traps hydrophobic residues that would otherwise escape to the exterior to gain entropy.^{102,111} ESI-IMS measurements, with drift times of ~ 0.1 millisecond, have identified ubiquitin⁹⁺ ions in the partially unfolded A state and ubiquitin¹¹⁺ ions in the extended-state U.^{103,106} ECD measurements after 0.2 and 5 seconds identified extended states also for the ubiquitin⁷⁺ ion and yielded information on the timescales required to overcome the unfolding barrier.¹⁰³ Recent replica exchange simulations, a simulation type designed to avoid kinetic traps,¹¹⁴ generated inside-out structures of the amyloid- β protein,¹¹⁵ giving strong support for the inside-out hypothesis. Whether well-defined inside-out structure can be generated in experiment remains

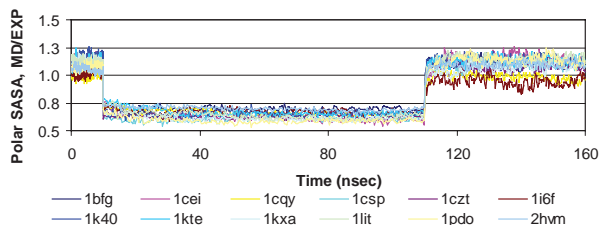


FIGURE 5 | Fast collapse of polar surface area upon transfer into vacuum (at 10 nanosecond) and close to complete recovery after rehydration (110 nanosecond). Structural changes occur within the first few nanoseconds after transfer to vacuum. (Reproduced with permission from Ref 35. Copyright 2009, Elsevier.)

to be seen. Evolution has sculptured the folding funnel the way that proteins can efficiently fold to their most favorable structure¹¹⁶ in water, solving the Levinthal paradox.¹¹⁷ There is no reason to assume such a smooth, funnel-shaped free energy surface in vacuum, in contrary, *in vacuo* folding experiments rather indicate a rough free energy landscape with significant barriers separating intermediate states from maximally compact structures.¹⁰⁷ The free energy landscape of a protein in vacuum resembles that of a frustrated system with conformers getting trapped in local energy minima.¹¹⁸

Systematic simulations of all domain metafolds³⁵ confirm that the IVNS is in most cases, and independently of fold, close enough to the solution structure so that fold-recognition algorithms such as MAMMOTH¹¹⁹ correctly assign solution phase domain architectures from the gas-phase measurements. Future XFEL experiments will produce atomistic structures of the IVNS. Explicit water MD simulations starting from the IVNS have been shown to full recover the solution structure within few nanoseconds³⁵ (see Figure 5), reinforcing the potential of atomistic gas-phase structures as sources of structural information for the aqueous protein. Very interestingly, low-resolution information derived from IMS experiments can provide useful structural information on the solution structure. For example, we demonstrated that use of CCS can help to filter structural models of single domains and of protein complexes, yielding to a significant enrichment in the quality of the models, which is only slightly below that obtained when small angle x-ray scattering data, obtained in solution, is used for filtering the theoretical models.¹²⁰ All these results reinforce our confidence in the information load of gas-phase structure and on their potential use to determine the solution structure.

It is clear that a protein unfolds when charging is increased. Ubiquitin (α/β) and Cytochrome C (all- α) ions do, for example, overcome the unfolding barrier if the charge density exceeds 0.18 charges per nm² surface area. Structures then progress to partially unfolded conformation whose structural characteristics are unknown, and whose stability depends on protein charge. Thus, ubiquitin⁺⁸ ions were stable on a 100 millisecond timescale, whereas +11 to +13 ions directly progress to the fully unfolded-state U^{106,121} in the same time window. Simulation of these partially folded and unfolded states and of the transitions between them represents a great challenge for future calculations as they occur on timescales that are beyond the current simulation capabilities. Generation of multiple (up to thousands) short simulations starting at different regions of the free energy landscape may cause relief. Long timescale kinetics can be reconstructed from these short trajectories by partitioning the sampled phase space into small conformational states and constructing a Markov state model between them.^{122,123} In this network representation of the protein's free energy landscape nodes correspond to basins (metastable states), and edges and edge weights denote the possibility and probability to transit between them.¹²⁴ The approach has been successfully applied to protein folding,¹²⁵ explorations of the *in vacuo* free energy landscape have, to our knowledge, not yet not been reported.

CONCLUSIONS

Recent mass spectrometric experiments, especially those using electrospray ionization, have generated an unexpected interest for understanding the structure of proteins in the gas phase. Clearly, one might expect the equilibrium structures of proteins in the absence of water differ markedly from those in solution, for example, with the hydrophobic core exposed, with many nonnative electrostatic interactions, and with extreme structural diversity due to many different structures being trapped within a rough and frustrated energy surface. However, experimental setups collect data very shortly after vaporization. This fact, combined with the roughness of the potential energy surface of proteins in the gas phase recently revealed by *in vacuo* MD simulations, fortuitously implies that the gas-phase structure detected experimentally has a very long-term memory of the solution structure. Therefore, low-resolution data obtained in gas-phase experiments should thus contain much more

information on the protein native state in solution than anticipated.

In light of these quite unexpected findings, future *in vacuo* structure determination techniques such as ultrashort X-ray laser pulses from XFEL experiments³ will not only provide a wealth of information on new tertiary and quaternary structures of proteins, but will also open up a new route to high-throughput structural proteomics and systems biology. Vice versa, these emerging techniques will also enhance the predictive power of new simulation and database-driven structure determination methods, both for isolated proteins and larger protein or RNA complexes. To this aim, more accurate FFs (in solution and *in vacuo*), as well as bet-

ter sampling and equilibration both on the simulation as well as on the experimental side will be essential.

From what we have learned from *in vacuo* MD simulations so far, it is very likely that *in vacuo* structures obtained from XFEL experiments will be efficiently refined by aqueous MD simulations and will thus yield accurate structural models of the protein in its native environment. In this sense, and despite the challenges inherent to the required tight interplay between experiment and theory, the synergy between these type of experiments and simulations is much more than just an added value—it is the main driving force of the emerging field of gas-phase structural biology.

ACKNOWLEDGMENTS

We would like to thank Prof. Bernd Abel for providing Figure 3(b) and Dr. Frank Wiederschein for providing Figure 3(c) and helpful discussions.

REFERENCES

1. Dill KA. Dominant forces in protein folding. *Biochemistry* 1990, 29:7133–7155.
2. Barron LD, Hecht L, Wilson G. The lubricant of life: a proposal that solvent water promotes extremely fast conformational fluctuations in mobile heteropolypeptide structure. *Biochemistry* 1997, 36:13143–13147.
3. Neutze R, Huldt G, Hajdu J, van der Spoel D. Potential impact of an X-ray free electron laser on structural biology. *Radiation Phys Chem* 2004, 71:905–916.
4. Chapman HN, Barty A, Bogan MJ, Boutet S, Frank M, Hau-Riege SP, Marchesini S, Woods BW, Bajt S, Benner WH, et al. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat Phys* 2006, 2:839–843.
5. Barrera NP, Robinson CV. Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes. *Annu Rev Biochem* 2011, 80:247–271.
6. Gabelica V, De Pauw E, Rosu F. Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry. *J Mass Spectrom* 1999, 34:1328–1337.
7. Rosu F, Gabelica V, Houssier C, Colson P, Pauw ED. Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 2002, 16:1729–1736.
8. Benesch JLP, Robinson CV. Biological chemistry: dehydrated but unharmed. *Nature* 2009, 462:576–577.
9. Yin S, Loo JA. Top-down mass spectrometry of supercharged native protein–ligand complexes. *Int J Mass Spectrom* 2011, 300:118–122.
10. Park AY, Robinson CV. Protein–nucleic acid complexes and the role of mass spectrometry in their structure determination. *Crit Rev Biochem Mol Biol* 2011, 46:152–164.
11. Fuerstenau SD, Benner WH, Thomas JJ, Brugidou C, Bothner B, Siuzdak G. Mass spectrometry of an intact virus. *Angew Chem Int Ed* 2001, 40:541–544.
12. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry. *Science* 1989, 246:64–71.
13. Fenn JB, Mann M, Meng CK, Wong SF. Electrospray ionization—principles and practice. *Mass Spectrom Rev* 1990, 9:37–70.
14. Nemes P, Marginean I, Vertes A. Spraying mode effect on droplet formation and ion chemistry in electrosprays. *Anal Chem* 2007, 79:3105–3116.
15. Luedtke WD, Landman U, Chiu YH, Levandier DJ, Dressler RA, Sok S, Gordon MS. Nanojets, electrospray, and ion field evaporation: molecular dynamics simulations and laboratory experiments. *J Phys Chem A* 2008, 112:9628–9649.
16. Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* 2000, 35:804–817.

17. Kebarle P, Peschke M. On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Anal Chim Acta* 2000, 406: 11–35.
18. Znamenskiy V, Marginean I, Vertes A. Solvated ion evaporation from charged water nanodroplets. *J Phys Chem A* 2003, 107:7406–7412.
19. Hogan CJ, Carroll JA, Rohrs HW, Biswas P, Gross ML. Combined charged residue-field emission model of macromolecular electrospray ionization. *Anal Chem* 2009, 81:369–377.
20. Fenn JB. Ion formation from charged droplets: roles of geometry, energy and time. *J Am Soc Mass Spectrom* 1993, 4:524–535.
21. Felitsyn N, Peschke M, Kebarle P. Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *Int J Mass Spectrom* 2002, 219:39–62s.
22. Hautreux M, Hue N, Du Fou de Kerdaniel A, Zahir A, Malec V, Lapr evote O. Under non-denaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface. *Int J Mass Spectrom* 2004, 231:131–137.
23. Iavarone AT, Jurchen JC, Williams ER. Supercharged protein and peptide ions formed by electrospray ionization. *Anal Chem* 2001, 73:1455–1460.
24. Fernandez De la Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal Chim Acta* 2000, 406:93–104.
25. Testa L, Brocca S, Grandori R. Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal Chem* 2011, 83:6459–6463.
26. Sterling HJ, Cassou CA, Trnka MJ, Burlingame AL, Krantz BA, Williams ER. The role of conformational flexibility on protein supercharging in native electrospray ionization. *Phys Chem Chem Phys* 2011, 13:18288–18296.
27. Sterling HJ, Kintzer AF, Feld GK, Cassou CA, Krantz BA, Williams ER. Supercharging protein complexes from aqueous solution disrupts their native conformations. *J Am Soc Mass Spectrom* 2012, 23:191–200.
28. Ahadi E, Konermann L. Modeling the behavior of coarse-grained polymer chains in charged water droplets: implications for the mechanism of electrospray ionization. *J Phys Chem B* 2012, 116:104–112.
29. Collette C, De Pauw E. Calibration of the internal energy distribution of ions produced by electrospray. *Crit Rev Biochem Mol Biol* 1998, 12:165–170.
30. Bagal D, Kitova EN, Liu L, El-Hawiet A, Schnier PD, Klassen JS. Gas phase stabilization of noncovalent protein complexes formed by electrospray ionization. *Anal Chem* 2009, 81:7801–7806.
31. Schnier PD, Gross DS, Williams ER. On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *J Am Soc Mass Spectrom* 1995, 6:1086–1097.
32. Marchese R, Grandori R, Carloni P, Raugi S. On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput Biol* 2010, 6.
33. Schnier PD, Gross DS, Williams ER. 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.
34. Miteva M, Demirev PA, Karshikoff AD. Multiply-protonated protein ions in the gas phase: calculation of the electrostatic interactions between charged sites. *J Phys Chem B* 1997, 101:9645–9650.
35. Meyer T, de la Cruz X, Orozco M. An atomistic view to the gas phase proteome. *Structure* 2009, 17:88–95.
36. Valentine SJ, Clemmer DE. Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase. *J Am Soc Mass Spectrom* 2002, 13:506–517.
37. Mao Y, Woenckhaus J, Kolafa J, Ratner MA, Jarrold MF. Thermal unfolding of unsolvated cytochrome c: experiment and molecular dynamics simulations. *J Am Chem Soc* 1999, 121:2712–2721.
38. Prasad M, Conforti PF, Garrison BJ, Yingling YG. Computational investigation into the mechanisms of UV ablation of poly(methyl methacrylate). *Appl Surf Sci* 2007, 253:6382–6385.
39. Jaskolla T, Karas M. Compelling evidence for lucky survivor and gas phase protonation: the unified MALDI analyte protonation mechanism. *J Am Soc Mass Spectrom* 2011, 22:976–988.
40. Heck AJR, van den Heuvel RHH. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* 2004, 23:368–389.
41. Berkenkamp S, Karas M, Hillenkamp F. Ice as a matrix for IR-matrix-assisted laser desorption/ionization: mass spectra from a protein single crystal. *PNAS* 1996, 93:7003–7007.
42. Morgner N, Kleinschroth T, Barth H-D, Ludwig B, Brutschy B. A novel approach to analyze membrane proteins by laser mass spectrometry: from protein subunits to the integral complex. *J Am Soc Mass Spectrom* 2007, 18:1429–1438.
43. Charvat A, Abel B. How to make big molecules fly out of liquid water: applications, features and physics of laser assisted liquid phase dispersion mass spectrometry. *Phys Chem Chem Phys* 2007, 9:3335.
44. Wattenberg A, Sobott F, Brutschy B. Detection of intact hemoglobin from aqueous solution with laser desorption mass spectrometry. *Crit Rev Biochem Mol Biol* 2000, 14:859–861.

45. Wiederschein F. *Investigation of Laser-Induced-Liquid-Beam-Ion-Desorption (LILBID) with Molecular Dynamics Simulations*. PhD Dissertation. Göttingen, Germany: University of Göttingen; 2009.
46. Breuker K, McLafferty FW. The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation. *Angew Chem Int Ed* 2005, 44:4911–4914.
47. Cooper HJ, Hakansson K, Marshall AG. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom Rev* 2005, 24:201–222.
48. 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.
49. Hemling ME, Conboy JJ, Bean MF, Mentzer M, Carr SA. Gas phase H/D exchange in electrospray ionization mass spectrometry as a practical tool for structure elucidation. *J Am Soc Mass Spectrom* 1994, 5:434–442.
50. Wood TD, Chorush RA, Wampler FM III, Little DP, O'Connor PB, McLafferty FW. Gas-phase folding and unfolding of cytochrome c cations. *Proc Natl Acad Sci USA* 1995, 92:2451–2454.
51. Wytenbach T, Bowers MT. Gas phase conformations of biological molecules: the hydrogen/deuterium exchange mechanism. *J Am Soc Mass Spectrom* 1999, 10:9–14.
52. Robinson EW, Williams ER. Multidimensional separations of ubiquitin conformers in the gas phase: relating ion cross sections to H/D exchange measurements. *J Am Soc Mass Spectrom* 2005, 16:1427–1437.
53. Wang F, Freitas MA, Marshall AG, Sykes BD. Gas-phase memory of solution-phase protein conformation: H/D exchange and Fourier transform ion cyclotron resonance mass spectrometry of the N-terminal domain of cardiac troponin C. *Int J Mass Spectrom* 1999, 192:319–325.
54. Eyler JR. Infrared multiple photon dissociation spectroscopy of ions in Penning traps. *Mass Spectrom Rev* 2009, 28:448–467.
55. Polfer N, Oomens J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom Rev* 2009, 28:468–494.
56. Antoine R, Dugourd P. Visible and ultraviolet spectroscopy of gas phase protein ions. *Phys Chem Chem Phys* 2011, 13:16494–16509.
57. Iavarone AT, Patriksson A, van der Spoel D, Parks JH. Fluorescence probe of Trp-cage protein conformation in solution and in gas phase. *J Am Chem Soc* 2007, 129:6726–6735.
58. Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* 2012, 2012:1–40.
59. Vonderach M, Ehrler OT, Matheis K, Karpuschkin T, Papalazarou E, Brunet C, Antoine R, Weis P, Hampe O, Kappes MM, Dugourd P. Probing electrostatic interactions and structural changes in highly charged protein polyanions by conformer-selective photoelectron spectroscopy. *Phys Chem Chem Phys* 2011, 13:15554–15558.
60. Garand E, Kamrath MZ, Jordan PA, Wolk AB, Leavitt CM, McCoy AB, Miller SJ, Johnson MA. Determination of noncovalent docking by infrared spectroscopy of cold gas-phase complexes. *Science* 2012, 335:694–698.
61. Nagornova NS, Rizzo TR, Boyarkin OV. Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide. *Science* 2012, 336:320–323.
62. Papadopoulos G, Svendsen A, Boyarkin OV, Rizzo TR. Conformational distribution of bradykinin [bk + 2 H]⁽²⁺⁾ revealed by cold ion spectroscopy coupled with FAIMS. *J Am Soc Mass Spectrom* 2012.
63. Revercomb HE, Mason EA. Theory of plasma chromatography/gaseous electrophoresis—a review. *Anal Chem* 1975, 47:970–983.
64. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, et al. Femtosecond X-ray protein nanocrystallography. *Nature* 2011, 470:73–77.
65. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 2010, 466:329–333.
66. Black DM, Payne AH, Glish GL. Determination of cooling rates in a quadrupole ion trap. *J Am Soc Mass Spectrom* 2006, 17:932–938.
67. Dunbar RC. BIRD (blackbody infrared radiative dissociation): evolution, principles, and applications. *Mass Spectrom Rev* 2004, 23:127–158.
68. McLuckey SA. Principles of collisional activation in analytical mass spectrometry. *J Am Soc Mass Spectrom* 1991, 3:599–614.
69. Wang B, Valentine S, Plasencia M, Raghuraman S, Zhang X. Artificial neural networks for the prediction of peptide drift time in ion mobility mass spectrometry. *BMC Bioinformatics* 2010, 11:182.
70. Morsa D, Gabelica V, De Pauw E. Effective temperature of ions in traveling wave ion mobility spectrometry. *Anal Chem* 2011, 83:5775–5782.
71. Merenbloom SI, Flick TG, Williams ER. How hot are your ions in TWAVE ion mobility spectrometry? *J Am Soc Mass Spectrom* 2012, 23:553–562.
72. Drahos L, Heeren RMA, Collette C, De Pauw E, Vêkey K. Thermal energy distribution observed in electrospray ionization. *J Mass Spectrom* 1999, 34:1373–1379.

73. Takats Z, Drahos L, Schlosser G, Vékey K. Feasibility of formation of hot ions in electrospray. *Anal Chem* 2002, 74:6427–6429.
74. Hall Z, Politis A, Bush MF, Smith LJ, Robinson CV. Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics. *J Am Chem Soc* 2012, 134:3429–3438.
75. Arcella A, Portella G, Ruiz ML, Eritja R, Vilaseca M, Gabelica V, Orozco M. Structure of triplex DNA in the gas phase. *J Am Chem Soc* 2012, 134:6596–6606.
76. Mesleh MF, Hunter JM, Shvartsburg AA, Schatz GC, Jarrold MF. Structural information from ion mobility measurements: effects of the long-range potential. *J Phys Chem* 1996, 100:16082–16086.
77. Wyttenbach T, von Helden G, Batka JJ Jr, Carlat D, Bowers MT. Effect of the long-range potential on ion mobility measurements. *J Am Soc Mass Spectrom* 1997, 8:275–282.
78. Bohrer BC, Mererblom SI, Koeniger SL, Hilderbrand AE, Clemmer DE. Biomolecule analysis by ion mobility spectrometry. In: *Annual Review of Analytical Chemistry*. Vol 1. Palo Alto, CA: Annual Reviews; 2008, 293–327.
79. Jurneczko E, Barran PE. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross-sections in the gas phase. *Analyst* 2011, 136:20–28.
80. Sun J, Kitova EN, Klassen JS. Method for stabilizing protein–ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal Chem* 2006, 79:416–425.
81. Barrera NP, Di Bartolo N, Booth PJ, Robinson CV. Micelles protect membrane complexes from solution to vacuum. *Science* 2008, 321:243–246.
82. Sun J, Kitova EN, Klassen JS. Method for stabilizing protein–ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal Chem* 2006, 79:416–425.
83. Sharon M, Ilag LL, Robinson CV. Evidence for micellar structure in the gas phase. *J Am Chem Soc* 2007, 129:8740–8746.
84. Friemann R, Larsson DSD, Wang Y, van der Spoel D. Molecular dynamics simulations of a membrane protein–micelle complex *in vacuo*. *J Am Chem Soc* 2009, 131:16606–16607.
85. Wang Y, Larsson DSD, van der Spoel D. Encapsulation of myoglobin in a cetyl trimethylammonium bromide micelle *in vacuo*: a simulation study. *Biochemistry* 2009, 48:1006–1015.
86. Caleman C, van der Spoel D. Temperature and structural changes of water clusters in vacuum due to evaporation. *J Chem Phys* 2006, 125:154508–154508–9.
87. Marklund EG, Larsson DSD, van der Spoel D, Patriksson A, Caleman C. Structural stability of electrosprayed proteins: temperature and hydration effects. *Phys Chem Chem Phys* 2009, 11:8069–8078.
88. Jarrold MF. Helices and sheets *in vacuo*. *Phys Chem Chem Phys* 2007, 9:1659.
89. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* 2001, 105:9954–9960.
90. Gao J, Luque FJ, Orozco M. Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solution. *J Chem Phys* 1993, 98:2975–2982.
91. Rueda M, Kalko SG, Luque FJ, Orozco M. The structure and dynamics of DNA in the gas phase. *J Am Chem Soc* 2003, 125:8007–8014.
92. Orozco M, Tirado-Rives J, Jorgensen WL. Mechanism for the rotamase activity of FK506 binding protein from molecular dynamics simulations. *Biochemistry* 1993, 32:12864–12874.
93. Blom MN, Compagnon I, Polfer NC, von Helden G, Meijer G, Suhai S, Paizs B, Oomens J. Stepwise solvation of an amino acid: the appearance of zwitterionic structures. *J Phys Chem A* 2007, 111:7309–7316.
94. Kjeldsen F, Silivra OA, Zubarev RA. Zwitterionic states in gas-phase polypeptide ions revealed by 157-nm ultra-violet photodissociation. *Chemistry* 2006, 12:7920–7928.
95. Mao Y, Ratner MA, Jarrold MF. Molecular dynamics simulations of the charge-induced unfolding and refolding of unsolvated cytochrome c. *J Phys Chem B* 1999, 103:10017–10021.
96. Patriksson A, Marklund E, van der Spoel D. Protein structures under electrospray conditions. *Biochemistry* 2007, 46:933–945.
97. Peng Y, Voth GA. Expanding the view of proton pumping in cytochrome c oxidase through computer simulation. *Biochim Biophys Acta* 2012, 1817:518–525.
98. Lill MA, 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.
99. Ahadi E, Konermann L. Ejection of solvated ions from electrosprayed methanol/water nanodroplets studied by molecular dynamics simulations. *J Am Chem Soc* 2011, 133:9354–9363.
100. Kong X, Brooks CL. λ -Dynamics: a new approach to free energy calculations. *J Chem Phys* 1996, 105:2414.
101. Donnini S, Tegeler F, Groenhof G, Grubmüller H. Constant pH molecular dynamics in explicit solvent with λ -dynamics. *J Chem Theory Comput* 2011, 7:1962–1978.

102. Wolynes PG. Biomolecular folding in vacuo!!!(?). *Proc Natl Acad Sci USA* 1995, 92:2426–2427.
103. Skinner O, McLafferty F, Breuker K. How ubiquitin unfolds after transfer into the gas phase. *J Am Soc Mass Spectrom* 2012, 23:1011–1014.
104. Patriksson A, Adams CM, Kjeldsen F, Zubarev RA, 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.
105. Dugourd P, Antoine R, Breaux G, Broyer M, Jarrold MF. Entropic stabilization of isolated β -sheets. *J Am Chem Soc* 2005, 127:4675–4679.
106. Wyttenbach T, Bowers MT. 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.
107. Shelimov KB, Jarrold MF. Conformations, unfolding, and refolding of apomyoglobin in vacuum: an activation barrier for gas-phase protein folding. *J Am Chem Soc* 1997, 119:2987–2994.
108. Li J, Taraszka JA, Counterman AE, Clemmer DE. Influence of solvent composition and capillary temperature on the conformations of electrosprayed ions: unfolding of compact ubiquitin conformers from pseudonative and denatured solutions. *Int J Mass Spectrom* 1999, 185–187:37–47.
109. Koeniger SL, Merenbloom SI, Sevugarajan S, Clemmer DE. Transfer of structural elements from compact to extended states in unsolvated ubiquitin. *J Am Chem Soc* 2006, 128:11713–11719.
110. Pierson NA, Chen L, Valentine SJ, Russell DH, Clemmer DE. Number of solution states of bradykinin from ion mobility and mass spectrometry measurements. *J Am Chem Soc* 2011, 133:13810–13813.
111. Artega GA, Tapia O. Structural transitions in neutral and charged proteins *in vacuo*. *J Mol Graph Model* 2001, 19:102–118.
112. van der Spoel D, Marklund EG, Larsson DSD, Coleman C. Proteins, lipids, and water in the gas phase. *Macromol Biosci* 2011, 11:50–59.
113. Ben-Tal N, Sitkoff D, Topol IA, Yang A-S, Burt SK, Honig B. Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *J Phys Chem B* 1997, 101:450–457.
114. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 1999, 314:141–151.
115. Baumketner A. Amyloid beta-protein monomer structure: a computational and experimental study. *Protein Sci* 2006, 15:420–428.
116. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. *Science* 1995, 267:1619–1620.
117. Levinthal C. How to fold gracefully. In: *Mossbauer Spectroscopy in Biological Systems*. Proceedings of a meeting held at Allerton House. Urbana: University of Illinois Press; 1969, 22–24.
118. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997, 48:545–600.
119. Ortiz AR, Strauss CEM, Olmea O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci* 2002, 11:2606–2621.
120. D'Abramo M, Meyer T, Bernadó P, Pons C, Recio JF, Orozco M. On the use of low-resolution data to improve structure prediction of proteins and protein complexes. *J Chem Theory Comput* 2009, 5:3129–3137.
121. Koeniger SL, Clemmer DE. Resolution and structural transitions of elongated states of ubiquitin. *J Am Soc Mass Spectrom* 2007, 18:322–331.
122. Bowman GR, Huang X, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* 2009, 49:197–201.
123. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *PNAS* 2009, 106:19011–19016.
124. Bowman GR, Pande VS. Protein folded states are kinetic hubs. *PNAS* 2010, 107:10890–10895.
125. Voelz VA, Bowman GR, Beauchamp K, Pande VS. Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1–39). *J Am Chem Soc* 2010, 132:1526–1528.