Proteins *in vacuo*: A molecular dynamics study of the unfolding behavior of highly charged disulfide-bond-intact lysozyme subjected to a temperature pulse

C. T. Reimann,[1,*] I. Velázquez,[2] M. Bittner,[1] and O. Tapia[2]
[1]*Division of Ion Physics, Department of Materials Science, Uppsala University, Box 534, S-751 21 Uppsala, Sweden*
[2]*Department of Physical Chemistry, Uppsala University, Box 532, S-751 21 Uppsala, Sweden*
(Received 11 June 1999; revised manuscript received 11 August 1999)

Molecular dynamics simulations were used to interpret a variety of experimental data on highly charged disulfide-bond-intact lysozyme *in vacuo*. The simulation approach involved submitting a model of the protein [Reimann, Velázquez, and Tapia, J. Phys. Chem. B **102**, 9344 (1998)] in a given charge state to a 3-ns-long heat pulse (usually at 500 K) followed by cooling or relaxation for 1 ns back to room temperature (293 K). This treatment yielded a charge threshold around $Q_0=8+$ for obtaining significant unfolding, as indicated by an enhancement in collision cross section and conformer length. The collision cross sections and lengths theoretically obtained, along with the threshold charge state for initiating unfolding, were compatible with experimental results on lysozyme *in vacuo*. The unfolded, highly elongated conformations obtained for $Q\geqslant 9+$ displayed a significant level of non-native β-sheet content which appeared to be additionally stabilized by charge self-solvation. [S1063-651X(99)06312-6]

PACS number(s): 87.15.Aa, 87.15.He

INTRODUCTION

The physicochemical properties of biomolecules have usually been considered in the context of the solution-phase environment in order to ensure a strong biological relevance. However, developments in biological mass spectrometry over the last ten years have led inexorably to a consideration of the properties of such molecules *in vacuo*. Probing the structure—unfolding, relaxing, and refolding—and dynamical behavior of biomolecules *in vacuo* opens a new window for assessing the specific influences manifested by the forces *intrinsic* to the biomolecule: the effects of the solvent are factored out in such studies.

This report is focused on the unfolding of highly charged protein ions *in vacuo*. Such unfolding effects have been monitored in a ''physical'' way via collision cross section measurements [1,2] and energetic surface imprints [3], and in a ''chemical'' way by measuring hydrogen-deuterium (H/D) exchange [4] and charge-stripping rates [5,6]. Such measurements each characterize any identifiable conformational substrate ensemble of ions by a single number, i.e., cross-sectional area, length, reaction rate, etc. rather than giving an atomically-resolved picture. (However, H/D studies, when augmented by fragmentation techniques, will provide an increasingly competitive level of resolution [7].)

Lysozyme ions *in vacuo* have been studied experimentally at low structural resolution by means of gas-phase basicity measurements [5], ion-drift mobility (IDM) [8], and energetic surface imprinting [9]. Clearly, although neither experimental conditions nor the precise results obtained are identical between the different experiments, the data show that disulfide-bond-intact lysozyme is present either as a compact, possibly nativelike form, or else as a distinctly different fluxional form.

The lack of atomically-resolved experimental data on protein ions *in vacuo* motivates the use of molecular dynamics (MD), which can model the behavior of such species at the atomic level. Such a strategy has been applied to both peptide [10–14] and protein ions [15–20]. Due to the enormous difference in granularity between experiments and theory, there is a need for comparison between results coming from these two levels of description.

Recently, we employed MD to simulate the response of native disulfide-bond-intact lysozyme to weak centrifugal forces [15] and to Coulomb repulsion [16]. The native structure appeared as quite robust against Coulomb-repulsion-driven denaturation, in qualitative agreement with the experimental observation of putative nativelike conformers [5,8]. When unfolding was initiated, conformers characterized by lengths $\approx 3\times$ and collision cross sections 50–80% greater than those of the native structure were the result, in good agreement with experimental results pertaining to unfolded lysozyme conformers [5,8,9]. However, too high a charge state, $Q\geqslant 15+$ [16,19], was required to initiate this unfolding. We noted earlier [16] that this can reflect a failure to take into account the role of increased internal energy in allowing the system to overcome an effective attractive barrier against unfolding.

The primary objective of this report is to document a more complete MD simulation scheme in which heating of the *in vacuo* protein model produced a response of the highly charged disulfide-bond-intact lysozyme more in line with available experimental results. In these MD studies, we show that the system threshold for unfolding lies around $Q_0=8+$, leading to significant enhancement in collision cross section and conformer length. For a range of charge states, extended conformers resulted with collision cross sections and lengths which are roughly compatible with experimental results on lysozyme *in vacuo* for the same charge state. A specific pre-

*Author to whom correspondence should be addressed. Present address: Department of Analytical Chemistry, Lund University, Box 124, S-221 00 Lund, Sweden. FAX: +46 46-222-4544; Electronic address: curt.reimann@analykem.lu.se

diction of this study is that the elongated conformers obtained are characterized by a significant level of non-native β-sheet content.

METHODS

The potentials, charging models, and seeding structures are those reported in our previous work [16]. Hen-egg-white lysozyme [21] has 11 arginines, six lysines, one histidine, and an *N* terminus. Thus, the maximum charge state for protonation is around $Q=19+$, assuming that all these basic sites would be accessible. A very simple charging model [16–19], ''statistical'' charging, distributes a desired total charge state evenly over just these basic sites on the protein. This scheme exaggerates the amount of Coulomb repulsion while weakening charge self-solvation interactions. In contrast to our earlier works, here we use a dielectric constant of $\varepsilon=1\varepsilon_o$ (where ε_0 is the permittivity of free space). This is a valid choice when all partial charges are explicitly included in the system. An energy-minimized x-ray structure was used as a seed for MD simulations of temperature-driven unfolding. Atom velocities were assigned randomly based on a Maxwell-Boltzmann distribution to achieve any desired temperature, maintained by a Berendsen thermostat [22]. A typical unfolding MD run submitted the protein model to a temperature of 500 K for 3 ns. For most charge states, several unfolding MD runs were carried out. For these repeated runs, the random number seed used to initiate assignment of the atomic velocities was changed. Separate relaxation MD runs were carried out on selected unfolded structures by reducing the temperature back to $T=293$ K. The atomic velocities were rerandomized for these 1-ns-long relaxations. The simulations were carried out with GROMOS87 [23].

Trajectories were analyzed according to standard structural descriptors like root-mean-square deviations (RMSD's) and radii of gyration (R_{gyr}). Lengths, characterizing the conformational elongation, were calculated as the maximum atom-atom distance considering *all* atoms. For comparison with the results of energetic surface imprinting experiments [9], which yield a projected length distribution of randomly oriented surface-impacting protein ions, MD trajectories were divided into ''snapshots.'' Each snapshot conformation was randomly reoriented and projected onto a plane, and the maximum length between any two atoms calculated in this plane. Histograms of these projected lengths were constructed. This procedure fairly takes into account the influence of fluctuations and nonlinear conformations on the projected length distributions.

For comparison with results of IDM [24,25] experiments on lysozyme [8], orientationally averaged collision (momentum-transfer) cross sections $\langle\sigma\rangle$ were calculated. We used an adjusted [26] implementation of the hard-sphere projection approximation (cf. Ref. [27]) which we described earlier [16]. The projection method ignores multiple scattering and thus provides lower limits for the collision cross sections; true values can be somewhat larger [28].

The calculation of $\langle\sigma\rangle$ for one conformational snapshot is laborious, and therefore whole trajectories cannot be analyzed in this way. Rather, conformational snapshots were extracted periodically, randomly reoriented, projected onto a plane, and σ calculated for each projection. Then an average of σ was taken over 1000 snapshots [29]. In this way, an estimate of $\langle\sigma\rangle$ was obtained for the last 1 ns of a 3-ns trajectory, giving a fair estimate of the effects of conformational fluctuations on $\langle\sigma\rangle$.

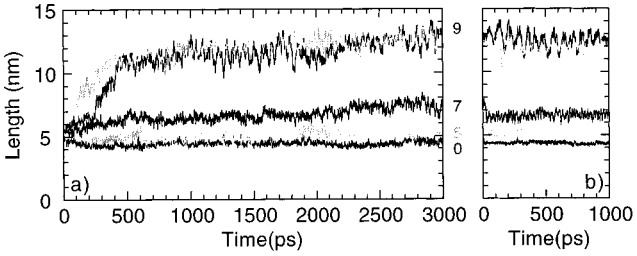

FIG. 1. Conformer length as a function of time for disulfide-bond-intact lysozyme in the indicated charge states. (a) Simulations carried out at $T=500$ K using a native seeding structure. (b) Simulations carried out at $T=293$ K using selected unfolded conformers near the end of the MD trajectories featured in (a) as seeding structures.

The secondary structure content of simulated proteins was analyzed with the MOLMOL software [30] as well as custom software. A hydrogen bond was presumed to exist when the hydrogen-donor-acceptor angle is less than 35° and the hydrogen-acceptor distance is less than 0.24 nm [31]. In the present situation, self-solvation was also monitored. Since this is a nondirectional interaction between a net charge and other partial charges, a simple distance criterion was used: that the charge-charge distance be less than 0.365 nm for the state of charge self-solvation [32].

RESULTS

Length profiles vs time are shown in Fig. 1(a) for $T=500$ K and several charge states Q. The $Q=0$ (neutral protein, NP) structures contracted slightly over time, and the $Q=5+$ structures maintained a length (5.2 ± 0.3 nm averaged over the last 1 ns of the trajectory) slightly greater than that of the native conformation (5.0 nm). However, the final structures obtained for $Q=0$ and $5+$ were *not* native: they are characterized by RMSD's approximately equal to 0.4 nm. By $Q=7+$, a certain degree of unfolding was achieved: the conformer length averaged over the last 700 ps of the trajectory was 7.3 ± 0.3 nm. For $Q=9+$ and $Q=11+$, dramatic unfolding was achieved. $Q=9+$ yielded lengths of 12.6 ± 0.6 nm and $Q=11+$ yielded lengths of 12.8 ± 0.5 nm, both averaged over the last 300 ps. Slightly greater lengths (13.8 ± 0.2 nm) were achieved in our previous simulations of $Q=19+$ at $T=293$ K [16].

Several 1-ns-long relaxation trajectories were carried out on unfolded structures obtained after ≈3 ns of temperature-pulse denaturation. It is apparent from Fig. 1(b) that on reducing the temperature to $T=293$ K, slight or minimal reductions in conformer length occurred. In particular, $Q=9+$ and $Q=11+$ maintained elongated structures characterized by lengths 12.5 ± 0.6 and 11.6 ± 0.5 nm, respectively. For $Q=9+$, a considerable degree of fluctuation occurred, showing ''bending'' or ''flapping'' at a frequency of ≈12 GHz. A similar effect occurred for $Q=11+$. (The errors in the lengths are standard deviations; considering peak-to-peak fluctuations, the lengths achieved after relaxation of

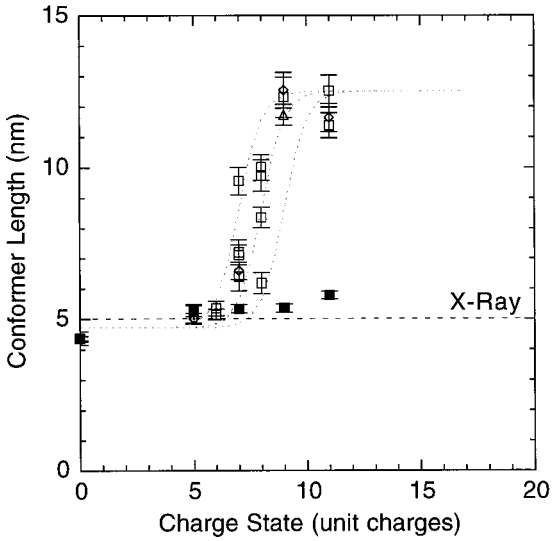

FIG. 2. Conformer lengths and standard deviations extracted from the last 1 ns of MD trajectories for $T=293$ K (■), $T=500$ K (□), and $T=293$ K after cooling from 500 K (◇). One relaxation run is included which uses unit charges (△); for details, see the text. The dashed line indicates the length of the x-ray structure of lysozyme. The dotted lines serve to guide the eye to a ''swathlike'' unfolding pathway. They are curves of the sigmoidal form $a+b/[1+\exp(-2.0(Q-Q_0))]$, where Q is the charge state and Q_0 is 7+, 8+, or 9+.

$Q=9+$ and 11+ are compatible.)

Conformer lengths averaged over the last 1 ns of all the MD trajectories are presented in Fig. 2. Several trajectories were calculated at $T=273$ K using the x-ray structure as a seed (solid squares). The $Q=0$ (NP) case provides a model for the ''*in vacuo* native state'' or IVNS described elsewhere [17]. It was shorter than the x-ray structure; as pointed out earlier [15], this is due to the collapse or closure of a potion of the active site of lysozyme (resides 43–51; see Fig. 3 in Ref. [15]). At 293 K, $Q\geqslant 5+$ drove a slight elongation with respect to the x-ray structure, but RMSD values remained mostly under 0.2 nm, indicating equivalence with the x-ray structure.

Considering both the unfolding runs carried out at 500 K (open squares) and the relaxation runs (open diamonds, Fig. 2), a ''swathlike'' unfolding transition was observed to occur, centered around $Q=8+$. For most of the charge states, extra unfolding runs were carried out with different initial velocity distributions: varying unfolding outcomes were observed. Still, however, most of the results roughly occupied the unfolding swath.

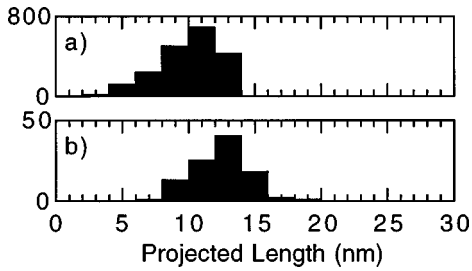

FIG. 3. Histograms of projected lengths of lysozyme conformers. (a) Projected length distribution for random orientations of each of 1000 conformational ''snapshots'' acquired from the 1-ns $Q=9+$ $T=293$-K relaxation run after unfolding [see Fig. 1(b)]. (b) Projected length distribution deduced experimentally for $Q=9+$ from Ref. [9].

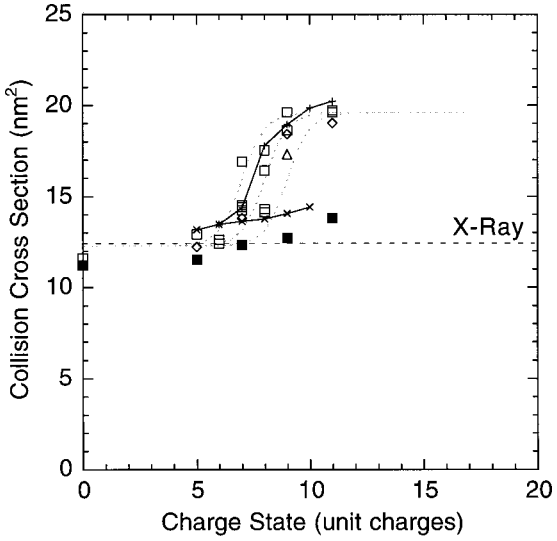

FIG. 4. Orientationally averaged collision cross sections $\langle\sigma\rangle$ estimated from the last 1-ns of MD trajectories for $T=293$ K (■), $T=500$ K (□), and $T=293$ K after cooling from 500 K (◇). One relaxation run is included which uses unit charges (△); for details see the text. (×) and (+) show experimental results for disulfide-intact lysozyme acquired under ''gentle'' and ''violent'' injection conditions, respectively [8]. The dashed line indicates the value of $\langle\sigma\rangle$ appropriate for the x-ray structure of lysozyme. The dotted lines serve to guide the eye. They are curves of the sigmoidal form $A+B/[1+\exp(-2.0(Q-Q_0))]$, where Q is the charge state and Q_0 is 7+, 8+, or 9+ (see also the caption to Fig. 2).

A projected length histogram representative of the low temperature relaxed $Q=9+$ MD trajectory from Fig. 1(b) appears in Fig. 3(a). The histogram peaks at about 11 nm and is skewed, and few projections have lengths below about 5 nm.

Collision cross sections $\langle\sigma\rangle$ calculated over the last 1 ns of all MD trajectories are presented in Fig. 4. For $T=273$ K and an x-ray structure seed (solid squares), the $Q=0$ (NP) case displayed a smaller cross section than the x-ray structure, again due to the closure of a portion of the active site of lysozyme [15]. $\langle\sigma\rangle$ gradually increased with increasing Q, and exceeded slightly the value characterizing the x-ray structure by $Q=9+$.

When MD trajectories were carried out at $T=500$ K, even tightly bound low-charge-state structures like $Q=0$ (NP) and $Q=5+$ unraveled to an extent (achieving RMSD values slightly higher than 0.4 nm). The rate of change of $\langle\sigma\rangle$ with respect to Q peaked somewhere between $Q=7+$ and 9+ (Fig. 4, open squares). The final results for $Q=9+$ and 11+ were quite similar, and manifested highly unfolded conformers with $\langle\sigma\rangle\approx 19$ nm^2.

As seen in Fig. 4, reducing the temperature to 273 K resulted in perceptible decreases in $\langle\sigma\rangle$ for the unfolded conformers (open diamonds). In particular, for $Q=0$ (NP) the final value of $\langle\sigma\rangle$ was indistinguishable from that of the

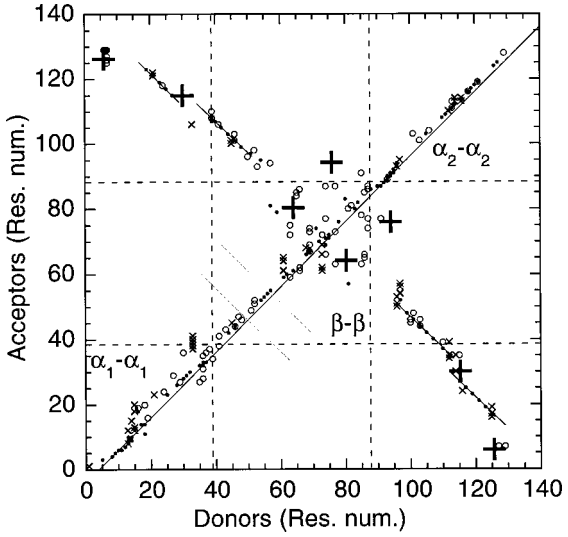

FIG. 5. Map of hydrogen donor-acceptor relationships for $Q=9+$ at $T=293$ K after cooling from 500 K. ''Res. num.'' is the residue number, ranging from 1 (N terminus) to 129 (C terminus). (●) Backbone donors and backbone acceptors, i.e., secondary structure. The line with ''slope'' +1 marks α helices. (○) Any donor-acceptor pairs *except* backbone to backbone. (×) Self-solvation of charged side chains. (+) Locations of disulfide bridges (included even though they are covalent bonds, i.e., they do not have anything to do with hydrogen bonds). The two grey lines indicate the pattern of β sheets characteristic of native lysozyme. The solid black lines with ''slope'' −1 indicate the presence of non-native β sheets.

IVNS; and for $Q=5+$, half of the enhancement in $\langle\sigma\rangle$ upon heating was eliminated after cooling. However, on a fractional basis, conformers with $Q=9+$ and 11+ hardly relaxed at all on reducing the temperature. For $Q=9+$, $\langle\sigma\rangle$ was 18.5 nm^2. When $\langle\sigma\rangle$ values obtained from all the unfolding and relaxation trajectories are gathered together, the unfolding behavior is again seen as variable but tended to occupy a swath centered at $Q=8+$.

A final main result summarizing information about hydrogen bonding and self-solvation appears in Fig. 5 for $Q=9+$ after heating and cooling. Traces of helical forms were apparent, as were hydrogen bonds and charge self-solvations involving amino acid residues located within four or five residues on the polypeptide backbone. In the β domain, for residue numbers 60–87, a set of hydrogen bonds appeared to link portions of the polypeptide chain separated by up to 20 or 25 residues, preserving a globular, intraconnected β domain character noted earlier [18]. The most striking observation is of two non-native β sheets linking residues 15–55 and 90–125.

Representative conformers observed after heating or relaxation are shown for different charge states in Figs. 6 and 7. (For $Q=9+$, a helix near the N terminus appears to be reconstituted on cooling; actually the helix fluctuated in and out of existence during both the 500- and 293-K trajectories.)

The long non-native β sheets seem to be a pervasive feature of the unfolded, elongated lysozyme conformers. We wondered whether the presence of intact disulfide bonds, which seem to ''pin'' the β strands together at three key points (Fig. 6), would promote the maintenance of the β

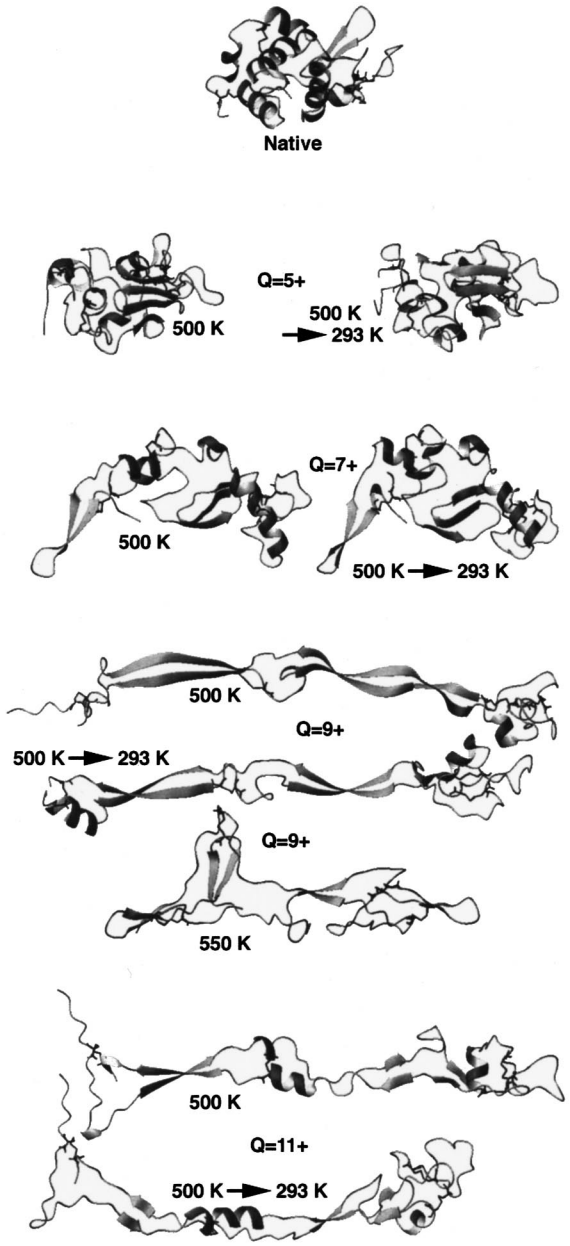

FIG. 6. Conformer snapshots. Structures obtained after 3 ns of MD at $T=500$ K and after a further 1 ns of MD at $T=293$ K are shown. (An extra snapshot from a trajectory carried out at $T=550$ K is also shown; the unfolding process was ''trapped'' in a somewhat shorter, wider structure.) Secondary structure is assigned by MOLMOL [30], from which the ribbons diagrams also come.

sheets. To investigate this, we broke the disulfide bridges in an unfolded conformer with $Q=9+$ which had been heated to 500 K and then cooled to 293 K, capping the cysteines with hydrogen and doing cycles of energy minimization according to the disulfide-bond-reduced topology. Then additional MD was performed for 1 ns. The conformer maintained an average length of 12.4 ± 0.9 nm, $\langle\sigma\rangle$ was 18.6 nm^2, and the non-native β sheets continued their existence. As concluded previously, the presence of disulfide bonds profoundly affects the unfolding pathway [15]. The interesting result presented here is that, once unfolding is achieved, the conformation is very stable—even without the disulfide bridges.

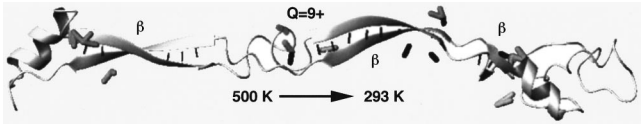

FIG. 7. The relaxed $Q=9+$ conformation showing all stabilizing bonding for three β sheets (β). Thin black lines: backbone-to-backbone hydrogen bonds spanning the two strands. Fat black cylinders: other hydrogen bonds spanning the two strands. Fat grey cylinders: self-solvation bonds spanning the two strands. All other bonds are omitted for clarity.

DISCUSSION

In the simulations presented above, standard assumptions were made about the seeding structure and the temperatures involved. Moreover, a ''statistical'' distribution of fractional charges was used, rather than specifically placing a smaller number of unit charges on the protein model. Finally, the time scales available in simulations are limited by 3–6 orders of magnitude compared with current experimental approaches. Any or all of these assumptions may constrain the relevance of the simulation results (although they may still yield a good qualitative description of what a highly charged protein *can* do *in vacuo*). The results are now discussed in light of the approximations made in the modeling. Following that a brief comparison of the simulation results to the results of experiments [5,8,9] is made.

A. Discussion of basic assumptions of the model

X-ray structure as seed. In several theoretical efforts carried out for interpreting experimental results on lysozyme ions *in vacuo*, the native structure is assumed to be a valid starting point [5,33]. Here this choice is made based on the general experience that disulfide-bond-intact lysozyme is a robust protein which is highly tolerant of harsh acid-solvent conditions [34]. Even if the conformation is distorted during the electrospray ionization (ESI) process, the use of a native starting conformation would serve to measure the disruptive effects of the Coulomb field as extended MD trajectories are calculated. Since there is strong evidence that biomolecular complexes [35], including viruses [36], can survive ESI and passage through a vacuum, the use of native structures to gauge deformation is a reasonable choice.

System temperature. Several authors took $T=500$ K to be a suitable temperature at which to computationally denature lysozyme [37] and other proteins [38]. This assumption is retained here. 500 K is a fairly extreme temperature in reality, but in MD, the need to confine the simulation to a few nanoseconds serves as a motivation for using such high temperatures to speed up the unfolding process [39]. However, to further test the generality of the unfolded lysozyme structures obtained for $Q=9+$, the native seeding structure was subjected to various temperature jumps. For $T=525$ and 1000 K, lengths of 11.1 ± 0.6 and 11.4 ± 0.8 nm, respectively, were obtained, averaged over the last 1 ns of the unfolding trajectories. These values can be compared to 12.6 ± 0.6 nm obtained for 500 K. The lengths attained were roughly independent of temperature [40], suggesting that once temperature-pulse-induced unfolding sets in, lengths of 11–12 nm are generally achieved by disulfide-bond-intact lysozyme for $Q=9+$ (Figs. 1 and 2).

For $T=550$ K, the final length achieved after 3 ns of integration was only about 8 nm (Fig. 6). Inspection of the final geometry revealed that the end of the conformation containing the N and C termini was trapped in a globular conformation which contained an ''extra'' β sheet oriented perpendicular to the long (longitudinal) axis of the conformer. Trajectories repeated at a *single* temperature with *different* initial velocity assignments can manifest rather different unfolding patterns and end points [38]. However, in the present study, the concerted forces engendered by Coulomb repulsion seem to bias the $Q=9+$ system to achieving fairly elongated structures, in spite of the inherently stochastic nature of the temperature pulse.

Dielectric constant. A strong rationale for using $\varepsilon=1\varepsilon_o$ in molecular mechanics and MD modeling was given in note (37) of Ref. [41]. We have used both $\varepsilon=1\varepsilon_o$ and $\varepsilon=2\varepsilon_o$ in previous work [16], in which we showed that the Coulomb repulsion associated with $Q=9+$ is unable to drive unfolding at $T=293$ K for either value of the dielectric constant [16]. Thus the point demonstrated in the present work that temperature jumps are required to affect gross unfolding of the $Q=9+$ state is *not* dependent upon the value assumed for ε [16].

Charging scheme. Another question concerns the use of a smooth distribution of fractional charges applied to all the basic sites on the protein, rather than the use of a more limited number of unit charges. We used the smooth or ''statistical'' distribution to avoid having to select which particular basic side chains would be charged [33,42], and also to deal with the ''mobile protons'' problem [43]. The profile of Coulomb forces in the statistical scheme is more ''concerted,'' and generally biases the system to unfold more smoothly and quickly to a representative final state [19]. However, the scheme also leads to greater Coulomb repulsion [16]. Therefore, it is important to check whether the unfolded structures remain unfolded even if the Coulomb repulsion is reduced due to the application of a smaller number of unit charges at specific protonation sites. For this purpose, the system was roughly checked by neutralizing the *N* terminus and the basic side chains *except* for residues [33,44] Arg_n, where $n=14$, 21, 45, 68, 73, 112, 114, 125, and 128, which were each provided with a net *unit* charge. Next, the system was equilibrated at T=500 K for 200 ps, and then relaxed for 1 ns at $T=293$ K. During this 1-ns relaxation trajectory the length reduced slightly to 11.6 ± 0.3 nm, and $\langle\sigma\rangle$ was reduced to 17.3 nm^2 (open triangle in Figs. 2 and 4). Secondary structural patterns, particularly the long nonnative β sheets, were preserved.

Length of the relaxation trajectories. To investigate whether the relaxation MD trajectories were sufficiently long to rule out later collapse, we extended the relaxation trajectory for $Q=9+$ [Fig. 1(b)] to a total elapsed time of 2 ns. The conformer length averaged over the last 1 ns of the relaxation trajectory was 12.4 ± 0.4 nm, a slight decease within the standard deviation of the fluctuations relative to the length quoted above. $\langle\sigma\rangle$ was 18.4 nm^2, also essentially unchanged. Therefore, the results for the $Q=9+$ charge scheme were not significantly altered on lengthening the MD trajectories.

Reproducibility. To gauge the overall reproducibility of the results, multiple unfolding MD trajectories were calculated for $T=500$ K using different initial distributions of atomic velocities. Some results are shown in Figs. 2 and 4 (open squares). A considerable variation in outcome is seen, particularly for $Q=7+$ and $8+$. The Coulomb repulsion due to $Q=8+$ seems to be at the threshold for gross unfolding, and therefore the range of behavior shows a stochastic nature and a high sensitivity to the initial conditions. (At $T=273$ K, the same behavior was observed for charge states around $Q=15+$ [19].) In spite of the variability, sigmoidal curves included in Figs. 2 and 4 roughly bracket all the MD results between unfolding charge thresholds $Q_0 \approx 8 \pm 1$.

The reproducibility of the relaxation trajectories was also gauged. For example, another 1-ns-long relaxation trajectory seeded with an alternative conformation extracted from near the end of the $T=500$ K unfolding run [Fig. 1(b)] yielded an average length of 12.6 ± 0.6 nm and an $\langle \sigma \rangle$ value of 18.2 nm^2, in good agreement with the primary results quoted above.

B. Comparison to experimental data acquired *in vacuo*

In the ESI process and the transfer of protein ions to the vacuum, as well as in subsequent physical treatment of the ions, heating and cooling occur, and the study reported here overcomes the lack of a heating pulse in our previous theoretical study [16]. Now we compare our present theoretical results with the results of experiments with disulfide-bond-intact lysozyme (DI-LYZ) *in vacuo* [5,8,9].

Energetic surface imprinting. These experiments provide the visual impression of elongated, cigar-shaped conformers for DI-LYZ with $Q=9+$, produced directly in the ESI process [9]. The histogram of projected lengths of protein ions impacting a surface in random orientation, deduced from modeling of experimental results [9], is shown in Fig. 3(b) and compares well with the MD results of Fig. 3(a) for the same charge state. The MD results suggest that the rounded peak, as well as the skewed distribution which, however, does not reach zero projected length, are characteristic of a conformation which possesses a finite width and moreover undergoes dynamic ''bow string'' and ''corkscrew'' type oscillations (frequency ≈ 12 GHz) while traversing space.

Collision cross sections. The collision cross sections of DI-LYZ conformers in various charge states have been measured by the IDM technique [8]. In this technique, the ions may be treated ''mildly'' (low-energy injection into the drift bath gas) or ''harshly'' (high-energy injection, which implies subjecting the ion to a high temperature pulse, followed by cooling back to room temperature). Experimental cross sections $\langle \sigma \rangle_{\text{expt}}$ appear in Fig. 4. Under mild conditions, $\langle \sigma \rangle_{\text{expt}}$ gently increases with Q, very similar to the room temperature MD results acquired by seeding with the x-ray structure (also shown in Fig. 4). This similarity suggests a compatibility between experimental conformers identified as being compact, and theoretical conformers identified as being essentially native.

Under harsh conditions, i.e., when the ions are given excess internal energy (and then cooled back to room temperature), a dominant ensemble of ions appears characterized by greatly enhanced $\langle \sigma \rangle_{\text{expt}}$ values for $Q>7+$. The MD results obtained with $t=500$ K and also after cooling various unfolded conformers back to 293 K occupy an unfolding swath which is consistent with the general trend of the experimental data acquired under harsh conditions [45].

Under the experimental conditions realized in IDM, the charge-state threshold for achieving a grossly unfolded conformation is $Q=8+$ (see Fig. 4 in Ref. [8]). This is precisely the charge state at which our unfolding MD trajectories appear to display the greatest sensitivity to the initial conditions (Fig. 4). The MD results thus imply a charge-state threshold for Coulomb-driven unfolding which is compatible with that observed experimentally.

Experimentally, charge states $Q \geqslant 8+$ are produced directly in the ESI process, whereas charge states $Q \leqslant 7+$ are produced by charge stripping from ions with charge state in the range $8 \leqslant Q \leqslant 11+$. Charge stripping was not modeled in the present study. However, if the higher charge-state ions retain a nativelike conformation under mild conditions, as strongly suggested by the MD results, it is plausible that the charge-stripped ions would also be nativelike. Thus, the assumption of a nativelike seed for $Q \leqslant 7+$ was maintained.

Gas-phase basicity. Measurements of gas-phase basicity, for DI-LYZ at $Q=9+$, imply the existence of compact conformers compatible with the native structure [5]. Our MD results (see also Ref. [16]) imply that the x-ray structure is stable *in vacuo* in the absence of perturbations, rationalizing the experimental observation. The experimental results also indicate the existence of partly unfolded conformers [5]. These structures may correspond to the ones obtained by MD after heating followed by cooling back to room temperature.

In gas-phase basicity measurements, conformers for $Q<9+$ are characterized by intermediate gas-phase basicity values considered consistent with partial denaturation and a more open structure [5]. However, this is in direct contradiction to IDM experiments, in which the results for $Q<9+$ are dominated by compact conformers [8]. This difference cannot be accounted for in the present work, but it could simply reflect varying experimental conditions. In fact, our MD simulations provide evidence for either compact or somewhat unfolded conformers for $Q \geqslant 5+$, depending on how the protein system is heated. Another possibility is that charge stripping (used to produce charge states $Q \leqslant 7+$) somehow produces a compact, misfolded structure characterized by gas-phase basicity values higher than that of the native conformation. The study reported here does not address such a situation.

Coexistence of differing conformational states. Under some circumstances, different conformers are observed experimentally to coexist. Let us first focus on the situation for $Q=9+$. For this charge state, gas-phase basicity measurements give clear evidence for the coexistence of nativelike *and* partially unfolded lysozyme. IDM allows the experimental conditions to be finely gradated. With mild treatment, a narrow conformer distribution results, but there is a ''shoulder'' on the mobility peak which indicates incipient unfolding. With harsh treatment, two peaks are obtained: a weak ''nativelike'' peak and a strong ''unfolded'' peak. A more detailed IDM spectrum for $Q=10+$ [8] shows that even the region *between* the nativelike and denatured peaks displays signal. This indicates the presence of metastable states which are distinctly non-native but which are not completely un-

folded. Our MD results for $Q=8+$ and $9+$ are qualitatively consistent with this in that not all MD runs achieved the same degree of unfolding.

Our MD results do not completely track with the experimental data. For $Q>5+$, we never observe a complete failure of the protein to initiate at least a degree of unfolding during a temperature jump, i.e., we cannot account for the weak nativelike peak observed in IDM when most of the ions are unfolded due to energetic injection into the drift tube buffer gas. This may ultimately be due to the bias induced by a too-strong Coulomb repulsion associated with the ''statistical'' charge distribution used in our model. Alternatively, a certain fraction of ions may refold (or misfold) on relaxation to form compact structures over the long time scale of the experiments, which cannot be addressed with MD simulations.

CONCLUSIONS

In this work, we employ the MD techniques to investigate the unfolding of highly charged disulfide-bond-intact lysozyme *in vacuo*. Using a simplified charging scheme which was previously demonstrated to have too high a charge-state threshold for unfolding at room temperature [16], and incorporating only a temperature jump to 500 K, the charge-state threshold for unfolding was reduced to a reasonable value, $Q_0=8\pm1$. Moreover, at a gross structural level, reasonable agreement was obtained between the results of experimental and MD simulation procedures, even after MD relaxation back to room temperature. The conformers obtained in the present MD study are thus hypothesized to be a good model of the structures experimentally observed.

In both unfolding and relaxation studies, the sheer extent of the protein's conformational space makes it difficult to ensure an exhaustive theoretical search. As a partial remedy to this problem, our MD simulations were carried out with the aid of a smooth, or ''statistical'' charge distribution, which gives an enhanced and more concerted repulsion. The use of such a distribution can be seen as a tool [19] to achieve computationally tractable unfolding, just as high-temperature pulses [37,38] and biased MD approaches [46] have been used. Alternatively, Monte Carlo techniques applied to simplified polymer models [47–49] may allow much longer time scales to be accessed, and many additional charging schemes to be considered. Such studies have already demonstrated a charge threshold for achieving open, elongated backbone conformations for polyelectrolytic [47] and polyampholytic [48,49] polymers [50]. In the future, a broad sampling of the resulting structures could be subjected to full-atom-resolved MD to fill in the structures at fine grain detail.

The MD results described in the present work have high structural resolution and therefore they provide complementary and unique information. A specific prediction of our results concerns the formation of non-native β sheets. These appeared as a pervasive characteristic of the disulfide-bond-intact lysozyme system *in vacuo*, perhaps because the location of the disulfide bonds constrains the mode and final degree of unfolding The actual existence of non-native β sheets could be tested by carrying out experiments at a finer level of detail [7]. Under some conditions, lysozyme in solution phase is prone to aggregation [51], and the extensive formation of non-native β sheets may provide a suitable aggregation mechanism [52]. This behavior suggests a connection between the protein in condensed phase and *in vacuo* environments.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Swedish Board for Technical Sciences (TFR) and the Swedish Board for Natural Sciences (NFR) for financial support. The authors would also like to thank Gustavo Arteca for stimulating discussions. C.T.R. thanks Sven Oscarsson for technical computing support.

[1] V. Nesatiy, Y.-L. Chen, B. A. Collings, and D. J. Douglas, Rapid Commun. Mass Spectrom. **12**, 40 (1998).
[2] K. B. Shelimov and M. F. Jarrold, J. Am. Chem. Soc. **118**, 10 313 (1996).
[3] P. A. Sullivan, C. T. Reimann, J. Axelsson, S. Altmann, A. P. Quist, and B. U. R. Sundqvist, J. Am. Soc. Mass Spectrom. **7**, 329 (1996).
[4] D. Suckau, Y. Shi, S. C. Beu, M. W. Senko, J. P. Quinn, F. M. Wampler III, and F. W. McLafferty, Proc. Natl. Acad. Sci. USA **90**, 790 (1993).
[5] D. S. Gross, P. D. Schnier, S. E. Rodriguez-Cruz, C. K. Fagerquist, and E. R. Williams, Proc. Natl. Acad. Sci. USA **93**, 3143 (1996).
[6] C. J. Cassady and S. R. Carr, J. Mass Spectrom. **31**, 247 (1996).
[7] F. W. McLafferty, Z. Q. Guan, U. Haupts, T. D. Wood, and N. L. Kelleher, J. Am. Chem. Soc. **120**, 4732 (1998).
[8] S. J. Valentine, J. G. Anderson, A. D. Ellington, and D. E. Clemmer, J. Phys. Chem. **101**, 3891 (1997).
[9] C. T. Reimann, P. A. Sullivan, J. Axelsson, A. P. Quist, S. Altmann, P. Roepstorff, I. Velázquez, and O. Tapia, J. Am. Chem. Soc. **120**, 7608 (1998).
[10] T. Wyttenbach, G. von Helden, and M. T. Bowers, J. Am. Chem. Soc. **118**, 8355 (1996).
[11] D. S. Gross and E. R. Williams, J. Am. Chem. Soc. **118**, 202 (1996).
[12] I. A. Kaltashov and C. Fenselau, Proteins: Struct., Funct., Genet. **27**, 165 (1997).
[13] R. R. Hudgins, M. A. Ratner, and M. F. Jarrold, J. Am. Chem. Soc. **120**, 12 974 (1998).
[14] S. C. Henderson, S. J. Valentine, A. E. Counterman, and D. E. Clemmer, Anal. Chem. **71**, 291 (1999).
[15] C. T. Reimann, I. Velázquez, and O. Tapia, J. Phys. Chem. B **102**, 2277 (1998).
[16] C. T. Reimann, I. Velázquez, and O. Tapia, J. Phys. Chem. B **102**, 9344 (1998).
[17] G. A. Arteca, I. Velázquez, C. T. Reimann, and O. Tapia, Phys. Rev. E **59**, 5981 (1999).
[18] I. Velázquez, Ph.D. dissertation, Faculty of Science and Technology, Uppsala University, Acta Universitatis Uppsaliensis,

1999 (unpublished).
[19] G. A. Arteca, C. T. Reimann, I. Velázquez, and O. Tapia (unpublished).
[20] Y. Mao, J. Woenckhaus, J. Kolafa, M. A. Ratner, and M. F. Jarrold, J. Am. Chem. Soc. **121**, 2712 (1999).
[21] K. P. Wilson, B. A. Malcolm, and B. W. Matthews, J. Biol. Chem. **267**, 10 842 (1992).
[22] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[23] W. F. van Gunsteren and H. J. C. Berendsen, Groningen Molecular Simulation (GROMOS) Library Manual, Biomos, Nijenborgh 16, 9747 AG Groningen, The Netherlands (1987).
[24] Ph. Dugourd, R. R. Hudgins, D. E. Clemmer, and M. F. Jarrold, Rev. Sci. Instrum. **68**, 1122 (1997).
[25] K. B. Shelimov, D. E. Clemmer, R. R. Hudgins, and M. F. Jarrold, J. Am. Chem. Soc. **119**, 2240 (1997).
[26] Our implementation yields an experimentally consistent value of $\langle\sigma\rangle = 1.28$ nm^2 for C_{60} [24], and a value of $\langle\sigma\rangle = 7.93$ mm^2 for BPTI (5 pti, A form, all atoms included), 3.5% too high [25]. Our implementation was also modified in one regard: we have revised downward the collision distances used for the ''united atoms''—CH, —CH_2, and —CH_3 to 0.276, 0.287, and 0.300 nm, respectively. The new collision distances were taken from full-atom hard-sphere projection approximations for these groups extracted from the Brookhaven protein data base entry for the protein BPTI (5 pti). Thus for the native or x-ray structure of lysozyme, $\langle\sigma\rangle = 12.4$ nm^2 instead of the value 12.7 nm^2 we obtained previously [16]. This exceeds the estimate provided in Ref. [8] by 5%. The method employed in Ref. [8] uses ''a porosity factor that takes into account the density of the protein'' which is not included in our implementation.
[27] G. von Helden, M.-T. Hsu, N. Gotts, and M. T. Bowers, J. Phys. Chem. **97**, 8182 (1993).
[28] A. A. Shvartsburg and M. F. Jarrold, Chem. Phys. Lett. **261**, 86 (1996).
[29] Dividing a portion of a MD trajectory into 5000 ''snapshots'' instead of 1000 led to $\langle\sigma\rangle$ changing by less than 0.5%.
[30] R. Koradi, M. Billeter, and K. Wüthrich, J. Mol. Graphics **14**, 51 (1996).
[31] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).
[32] W. Saenger, *Principles of Nucleic Acid Structure*, edited by C. R. Cantor, Springer Advanced Texts in Chemistry (Springer-Verlag, New York, 1984).
[33] M. Miteva, P. A. Demirev, and A. D. Karshikoff, J. Phys. Chem. B **101**, 9645 (1997).
[34] M. Buck, S. E. Radford, and C. M. Dobson, Biochemistry **32**, 669 (1993).
[35] M. Przybylski and M. O. Glocker, Angew. Chem. Int. Ed. Engl. **35**, 806 (1996).
[36] G. Siuzdak, J. Mass Spectrom. **33**, 203 (1998).
[37] P. H. Hünenberger, A. E. Mark, and W. F. van Gunsteren, Proteins: Struct., Funct., Genet. **21**, 196 (1995).
[38] T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[39] A. Li and V. Daggett, J. Mol. Biol. **257**, 412 (1996).
[40] For $T = 1000$ K, $\langle\sigma\rangle$ was higher than for the lower temperatures. This is because the secondary structure was completely destroyed, and the polypeptide chain ''bowed out'' in a sort of ''Figure 8,'' exposing a much greater fraction of the atoms for collisions. In this case, $\langle\sigma\rangle$ and conformer length are sensitive to different aspects of unfolding.
[41] I. A. Kaltashov and C. C. Fenselau, J. Am. Chem. Soc. **117**, 9906 (1995).
[42] P. D. Schnier, D. S. Gross, and E. R. Williams, J. Am. Soc. Mass Spectrom. **6**, 1086 (1995).
[43] A. Dongré, J. L. Jones, A. Somogyi, and V. H. Wysocki, J. Am. Chem. Soc. **118**, 8365 (1996).
[44] J. D. Carbeck, J. C. Severs, J. Gau, Q. Wu, R. D. Smith, and G. M. Whitesides, J. Phys. Chem. **102**, 10 596 (1998).
[45] $\langle\sigma\rangle$ values obtained from MD trajectories are often lower than the experimental values $\langle\sigma\rangle_{\text{expt}}$. This may reflect the approximate nature of the method of estimating the collision cross section of simulated conformers, which tends to slightly underestimate $\langle\sigma\rangle$ [26,28].
[46] M. Marchi and P. Balline, J. Chem. Phys. **110**, 3697 (1999).
[47] A. V. Dobrynin, M. Rubinstein, and S. P. Obukhov, Macromolecules **29**, 2974 (1996).
[48] S. Wolfling and Y. Kantor, Phys. Rev. E **57**, 5719 (1998).
[49] Y. Kantor and M. Kardar, Europhys. Lett. **27**, 643 (1994).
[50] The Monte Carlo (MC) studies presented in Refs. [47–49] modeled the behavior of polymers in solution, not *in vacuo*. Furthermore, no secondary structure building propensity was included in the polymer model. Unfolded structures obtained by the MC method look like ''beads on a necklace.'' However, since no covalent linkages between polymer monomers nonadjacent in the polymer backbone were included, the unfolded structures do not directly resemble the ones obtained in the study reported here of disulfide-bond-intact lysozyme.
[51] D. R. Booth, M. Sunde, V. Bellotti, C. V. Robinson, W. L. Hutchinson, P. E. Fraser, P. N. Hawkins, C. M. Dobson, S. E. Radford, C. C. F. Blake, and M. B. Pepys, Nature (London) **385**, 787 (1997).
[52] A. L. Fink, Folding Des. **3**, R9 (1998).