



RESEARCH ARTICLE

Statistical Analysis of Ion Mobility Spectrometry. I. Unbiased and Guided Replica-Exchange Molecular Dynamics

Fabien Chirot,^{1,2,5} Florent Calvo,^{1,3} Florian Albrieux,^{1,2,3} Jérôme Lemoine,^{1,2}
Yury O. Tsybin,⁴ Philippe Dugourd^{1,3}

¹Université de Lyon, Villeurbanne, France

²CNRS, UMR 5280, ISA, Lyon, France

³CNRS, UMR 5579, LASIM, Lyon, France

⁴Biomolecular Mass Spectrometry Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

⁵Institut des Sciences Analytiques, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

Abstract

Achieving (bio)macromolecular structural assignment from the interpretation of ion mobility spectrometry (IMS) experiments requires successful comparison with computer modeling. Replica-exchange molecular dynamics simulations with suitable force fields not only offer a convenient framework to locate relevant conformations, especially in the case of multiple-funnel energy landscapes, but they are also well suited to statistical analyses. In the present paper, we discuss two extensions of the method used to improve its efficiency in the context of IMS. Two doubly-protonated polyalanines $[RA_4XA_4K + 2H]^{2+}$ with $X=V$ and D appear as favorable cases for which the calculated collision cross-section distributions naturally agree with the measurements, providing reliable candidate structures. For these compounds, a careful consideration of other order parameters based on the weighted histogram method resolves several otherwise hidden underlying conformational families. In the case of a much larger peptide exhibiting bistability, assignment is more difficult but could be achieved by guiding the sampling with an umbrella potential using the square gyration radius as the biasing coordinate. Applied to triply protonated bradykinine, the two presented methods indicate that different conformations compatible with the measurements are very close in energy.

Key words: Ion Mobility, Molecular Dynamics, Structural Assignment, Biomolecules, Umbrella Sampling, WHAM

Introduction

Mass spectrometry (MS) is a powerful analytical tool for the structural analysis of peptides and proteins. The

combination of ion mobility spectrometry (IMS) with MS gives a great insight into the three-dimensional conformation, the folding mechanisms, or the way such biomolecules assemble together in covalent or noncovalent complexes [1–5]. IMS measurements are usually analyzed by calculating cross-sections for unsolvated trial conformations obtained from computational chemical methods. This entails two main bottlenecks on the theoretical side. Quantum chemistry methods, such as Hartree-Fock or density-functional theory still struggle to achieve energetic and structural accuracy even

Electronic supplementary material The online version of this article (doi:10.1007/s13361-011-0281-y) contains supplementary material, which is available to authorized users.

Correspondence to: Fabien Chirot; e-mail: fabien.chirot@univ-lyon1.fr

Received: 1 July 2011
Revised: 10 October 2011
Accepted: 16 October 2011
Published online: 1 December 2011

for relatively small peptides [6], a limitation caused by the variety of the types of interactions (covalent, ionic, dispersion, polarization) within such compounds. Molecular mechanics force fields, although not chemically as accurate [7], are then a necessary alternative to produce sets of candidate structures. Unfortunately, for systems containing more than typically a hundred atoms, the energy landscapes are rugged and difficult to sample ergodically by conventional simulation methods, even using simple force fields [8]. This difficulty is particularly critical in molecular dynamics (MD), because its time scales lie far below the folding times of most peptides, but of course Monte Carlo (MC) methods are affected as well, notably in the context of global optimization. Thus, the prediction of secondary and tertiary structures of peptides by unbiased simulations still stands as a very active goal of computational biochemistry. Enhanced sampling can be achieved by simplifying the system, using for instance coarse-graining procedures in the molecular model itself [9], but a constant effort in recent years has focused on designing computational strategies to accelerate the exploration of configurational space. Prominent methods such as temperature accelerated dynamics [10], metadynamics [11], or replica-exchange MD [12] are examples of refinements over traditional MD that are considerably more efficient, but without necessarily involving many additional parameters.

Interpretation of IMS measurements thus usually proceeds by comparing the collision cross-sections obtained for various candidate structures based on molecular modeling. However, multiple and possibly inter-converting structures may share the same ion mobility, which renders the assignment of structures more difficult. Additional theoretical information about the energies or free energies of these structures helps, but for large and charged systems the accuracy of the underlying force field may be questionable. A more stringent comparison between experiment and theory should be made at a statistical level because a distribution with a finite broadening is measured. Calculating a meaningful statistical distribution of collision cross-sections from molecular simulations is significantly more demanding than performing a series of individual cooling trajectories, especially if one hopes to get the sampling right.

Our first goal in the present paper is to illustrate how replica-exchange molecular dynamics, combined with proper statistical data analysis based on histogram reweighting, can be used to calculate collision cross-section distributions that are comparable with experimental distributions. In the case of the small polyalanines $[\text{RA}_4\text{VA}_4\text{K} + 2\text{H}]^{2+}$ and $[\text{RA}_4\text{DA}_4\text{K} + 2\text{H}]^{2+}$, this agreement will be exploited to get deeper insight into the conformations that match the measurements for which additional order parameters reveal several structural families.

Our second goal is concerned with the case of larger peptides, for which conformational sampling is much more challenging. Here, molecular modeling should not be as ambitious as above in trying to reproduce an entire distribution of collision cross-sections. Yet we still would like to obtain some structural information that matches the measured values, especially when the experiment reveals bi-stable or competing conformations. Straightforward simulations, even improved

with replica-exchange or other schemes, may fail simply because the landscape was still too rugged, or because the force field may not be accurate enough. It is then tempting to guide the sampling toward the experimentally measured region by biasing the potential energy surface with an additional explicit but non-physical contribution, the simplest form of which being an umbrella potential [13]. The main rationale behind this approach is to reduce the computational time dedicated to regions of the configurational space that are incompatible with experimental observations while driving the sampling toward more appropriate places. The practical implementation of this idea requires a suitable order parameter that mimics the collision cross-section without being computationally involved. We have found that the squared gyration radius satisfactorily meets this purpose. The performance of this biasing strategy is illustrated here on a variant of a sub-unit of the M2 trans-membrane protein from influenza virus A (M2TMP), which displays a gas phase bi-stable structure with a competition between helices and more compact structures [14].

Finally, we have applied the two above-mentioned methods to triply-protonated bradykinine. Although not as large as the former M2TMP sub-unit, this peptide was recently found by Pierson *et al.* [15] to exhibit, upon specific activation conditions, three main features in their ion mobility spectra still awaiting structural assignment. This system thus appears as a good benchmark to test the ability of the present techniques to yield candidate conformations.

The paper is organized as follows. The next section focuses on the polyalanine-based peptides, for which replica-exchange molecular dynamics can properly sample the canonical ensemble and account for the possible competition between conformational families. In particular, additional order parameters from the configurations sampled provide a wealth of information hidden from the mere distributions of collision cross-sections, revealing various conformational families. The second section discusses the effects of guiding the sampling toward the experimentally measured distributions, and how this bias can even be used to improve the original sampling. The application of the two methods to bradykinine is discussed in the last section. We finally summarize and give some concluding remarks.

Distributions of Cross-Sections in Polyalanines

Unbiased Replica-Exchange Molecular Dynamics

Standard replica-exchange, or parallel tempering, was originally introduced to enhance barrier crossing in complex energy landscapes such as spin glasses [16, 17]. Its application to molecular systems is more recent [12], but the method has since become widespread and is implemented in most biomolecular software packages. Briefly, replica-exchange consists of propagating simultaneously the various trajectories (referred to as replicas) at different temperatures or, more generally, different ensemble-generating parameters. Periodical-

ly, some exchanges of configurations between trajectories are attempted, one exchange being accepted typically based on a Metropolis criterion. The idea is then that high temperature replicas will explore broader regions of the energy landscape, and the exchange process can communicate the various basins of attraction to lower temperature replicas. The method is particularly suitable for multiple funnel landscapes [18], however, it becomes less efficient in systems with numerous degrees of freedom because the probability of a successful exchange drastically drops unless a larger number of replicas are used.

An explicit assumption in parallel tempering simulations is that individual trajectories must sample the required statistical ensemble that is consistent with the exchange criterion. The Berendsen thermostat, for instance, while maintaining kinetic energies around prescribed values, does not rigorously sample the canonical ensemble. Here we used the Nosé-Hoover technique [19], with thermostat masses chosen following the recommendations of references [20]. The simulations consisted of 32 replicas spanning the temperature range $100 \text{ K} \leq T \leq 800 \text{ K}$, temperatures being distributed according to a geometric progression in order to maximize the overlap between successive energy distributions. Each trajectory was propagated for several consecutive series of 10 ns with a time step of 1 fs. The Amber99 force field [21] was used to model the interactions.

Statistical Analysis

Our primary objective was to assist the interpretation of IMS experiments by calculating distributions of collision cross-section Ω for a representative ensemble of conformations. Besides several approximate methods [22, 23], we have used the more accurate but also more demanding trajectory method of Mesleh and coworkers [24], which consists in simulating directly the collision between the molecular target and the impinging gas atom, eventually obtaining Ω after suitable averaging on orientations and impact parameters. The calculated structures may not be accurately characterized by Ω alone, but complementary insight on calculated conformations can be provided by additional structural quantities. The other properties may of course be obtained on-the-fly, but it is advantageous to post-process the structural analysis using weighted histograms in the so-called WHAM method [25]. The original WHAM procedure aims at combining the multidimensional histograms $h(T, E, \chi)$, where T , E , and χ are the temperature, potential energy, and property of interest, respectively, into the continuous canonical averages $\langle \chi \rangle(T)$. The usual WHAM approach follows an iterative process to evaluate the partition functions at all temperatures, providing a relative weight of each trajectory used to interpolate the final estimate of $\langle \chi \rangle(T)$. It is thus a natural complement to replica-exchange simulations, since both methods need the energy distributions to overlap in order to be efficient.

Our present implementation of weighted histograms differs from the original WHAM method, as it builds directly upon the formerly developed histogram technique of Ferrenberg and Swendsen [26], without involving iterative procedures. The

multidimensional histograms $h(T, E, \chi)$ are first reduced into energy-only histograms $H(T, E)$ as

$$H(T, E) = \sum_{\chi} h(T, E, \chi) \quad (1)$$

and from $H(T, E)$ the microcanonical entropy $S(E)$ can be numerically calculated by minimizing exactly a least-square estimator [27]. Once S is known, all thermal properties are straightforwardly obtained as continuous functions of temperature through Laplace transformations of E and its powers. Averages and distributions involving the structural parameter χ still require explicit interpolation as

$$\langle \chi \rangle(T) = \frac{1}{Q(T)} \sum_{E, \chi} \chi h(T, E, \chi) e^{S(E) - \frac{E}{k_B T}} \quad (2)$$

$$Q(T) = \sum_{E, \chi} h(T, E, \chi) \frac{e^{S(E) - \frac{E}{k_B T}}}{M(E)} \quad (3)$$

$$M(E) = \sum_{T, \chi} h(T, E, \chi) \quad (4)$$

where k_B is the Boltzmann constant. The previous treatment assumes that χ and T are independent, which in the case of the collision cross-section Ω is not entirely valid, since T enters the velocity of impinging gas atoms in the trajectory method [24]. However, by comparing to the direct average at the replica temperatures, we have checked that the error on $\langle \Omega \rangle(T)$ was negligible.

Comparison to Experimental Data: From Cross-Section Distributions to Arrival-Time Distributions

IM-MS measurements were carried out using a custom-built ion mobility spectrometer, the experimental setup of which is described elsewhere [28]. Briefly, the ions are formed in an electrospray ion source and transferred to a cylindrical ion trap [29] via an ion funnel. Ion bunches from the trap are periodically injected directly in a 1 m-long drift tube filled with 12 Torr of helium and maintained at 293 K. The duration of an ion pulse is 1 ms at the injection. At the end of the tube the ions enter a vacuum chamber through a 0.7 mm diameter aperture. The ions are finally conveyed via two ion funnels to the time-of-flight mass spectrometer (microTOF_Q; Bruker-Daltonics, Bremen, Germany) where they are detected as a function of their time of flight (mass-to-charge ratio) and drift time (collision cross-section).

At a given electric field E , the ions travel across the drift tube at a constant velocity v_D , and the arrival time, or drift time in the mobility tube, t_D is related to this field through:

$$t_D = \frac{L}{v_D} = \frac{L}{KE} \quad (5)$$

where K defines the ion mobility and L is the length of the drift tube. Under the experimental conditions, K is inversely

proportional to the orientationally-averaged diffusion cross-section Ω [30]:

$$K = \frac{3}{16} \frac{ze}{N} \left(\frac{1}{m} + \frac{1}{M} \right)^{\frac{1}{2}} \left(\frac{2\pi}{k_B T} \right)^{\frac{1}{2}} \frac{1}{\Omega} \quad (6)$$

In Equation 6, ze is the charge of the ion, N the buffer gas density, and m and M are the masses of the helium atom and the ion, respectively. As a direct consequence, the most compact conformers have the shortest drift times.

In practice, the drift time t_D is measured for various drift voltages V and the ion mobility is obtained from the slope of t_D as a function of V^{-1} . The cross-section can then be derived using Equations 5 and 6. The experimental precision on cross-sections is estimated to be no more than 1%.

In order to be directly compared with the experimental data, the simulated cross-section distributions were converted into arrival time distributions (ATDs) by using Equations 5 and 6 to calculate the drift time t_D corresponding to each calculated cross-section. The ATD for a single isomer was modeled through the convolution of the injected ion pulse with the Fick's law for the diffusion in the tube [31]. The expression of the ions flux at time t at the end of the tube for isomers with average drift time t_D and injected at time t_0 is then:

$$s(t, t_D) = \frac{v_D}{\sqrt{4\pi D t_D}} \int_{-\infty}^{+\infty} I(t_0) \exp\left(-v_D^2 \frac{(t_D + t_0 - t)^2}{4D t_D}\right) dt_0 \quad (7)$$

where $I(t_0)$ represents the initial ion pulse, and $D = Kk_B T/q$ the diffusion constant at temperature T for ions with charge q and mobility K [30]. The global ATD was then calculated as a sum of the contributions of the identified isomers weighted by their abundance $C(t_D)$ in the simulated cross-section distribution:

$$S(t) = \sum_{t_D} C(t_D) s(t, t_D) \quad (8)$$

This convolution procedure was applied to the distributions directly extracted from a MD run as well as to the distributions calculated through histogram reweighting.

Application to Polyalanines

Two alanine-based peptides were used as model systems to evaluate the potentiality of WHAM to assist structure elucidation from IMS measurements. We chose the two systems as representatives of the two main structural families identified in reference [28] upon a series of variants of the sequence RA₄XA₄K, X standing for one of the 20 naturally occurring amino acids. Most of the variants were found to form α -helices at room temperature, and shorter helices were assigned to the peptides for which the central amino acid side chain is involved in the solvation of the charges, as for X=D (further denoted AlaD). Most of the

other variants, including X=V (further denoted AlaV) form longer helices [28]. Figure 1 displays the ATDs recorded at room temperature for AlaV and AlaD using a drift field of 770 V.m⁻¹ and a drift gas pressure of 12 Torr. The figure also displays the simulated ATDs obtained from the cross-sections of the structures sampled from a REMD run at 304 K (referred to as raw REMD) and the distributions resulting from the WHAM procedure described above.

For the two considered peptides, the agreement between experimental and simulated ATDs is very satisfactory. Both the position and the shape of the distributions are well reproduced by the two simulation procedures. By reproducing the shape of the experimental ATDs, the REMD simulations quantitatively account for the extent of conformational flexibility consistent with the observed peaks. As already shown in reference [28], different conformers may actually all lie within a single IMS peak. However, the width alone is compatible with the prediction from the Fick's law, hence the contribution of isomer spreading to the calculated distribution cannot be inferred and the diffusion broadening appears as the limiting factor for the resolving power of the instrument. We demonstrate in the following that, based on a satisfactory agreement with experimental ATDs, the statistical analysis of computational results can be used to probe possible isomer coexistence within a single ATD peak.

Figure 2 displays the evolution of the cross-section distributions for AlaV and AlaD as a function of temperature, as inferred using the above described WHAM procedure. For both systems, the calculated cross-section distribution consists of a single peak in the entire temperature range. As temperature increases, the average cross-section first decreases. This trend mainly reflects the intrinsic temperature dependence of the cross-section, rather than any conformational change, as the high cross-sections at low temperatures are due to the lower velocities of the impinging neutral atoms of the colder gas. To illustrate this effect, the evolution of the cross-section calculated for the lowest-lying isomer identified in the simulation has been plotted in Figure 2 as a function of temperature, without considering any structural change and fixing the geometry as the global minimum conformer. The evolution of the statistically averaged cross-section follows the same trend at low temperatures ($T < 150$ K for AlaV, and $T < 300$ K for AlaD), but significantly differs at higher temperatures, thus revealing structural changes. For $T = 500$ – 600 K the average cross-section displays an inflexion corresponding to the temperature-driven unfolding of the peptides. Concomitantly, the cross-section distribution significantly broadens due to the wider conformational landscape accessible in the unfolded state.

To summarize, from the two cross-section maps in Figure 2, the only visible structural transition occurs in the 500–600 K range and corresponds to the unfolding of the peptide. However, other order parameters reveal more subtle changes. The average squared gyration radius $\langle Rg^2 \rangle$ is not

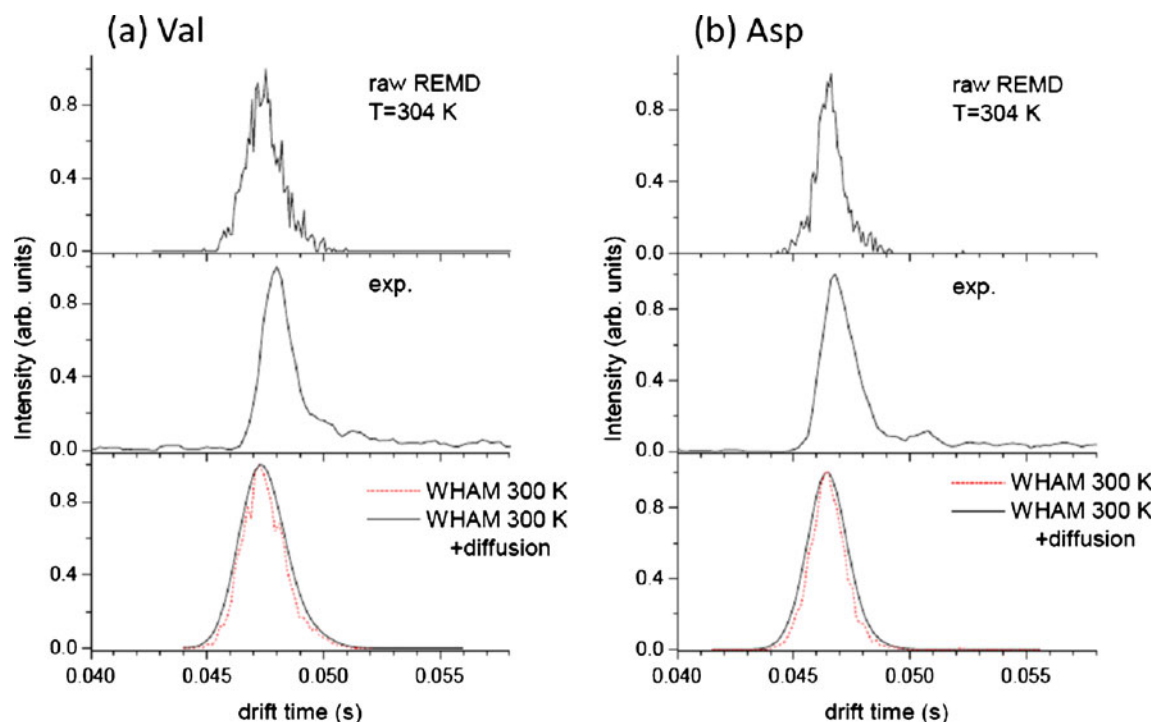


Figure 1. Experimental and calculated arrival drift time distributions for the polyanalines **(a)** $[\text{RA}_4\text{VA}_4\text{K}+2\text{H}]^{2+}$ and **(b)** $[\text{RA}_4\text{DA}_4\text{K}+2\text{H}]^{2+}$. The top panels show the raw MD data from the simulated trajectories closest to 300 K, the middle panels show the experimental distributions, and the bottom panels show the ATD obtained after histogram reweighting (dashed red lines) and further including the contribution of diffusional broadening during the drift (solid black lines)

expected as a global indicator to carry much more detailed information than the diffusion cross-section does. However, in the case of AlaV, its evolution with temperature surprisingly differs from the variations of Ω (see Figure 3). Apart from the above mentioned increase at $T=500\text{--}600\text{ K}$, the evolution of $\langle R_g^2 \rangle$ is also non-monotonic at low temperature. The square gyration radius distribution for AlaV at 200 K is plotted in Figure 3. It shows three main peaks at 28, 31, and 36.5 \AA^2 while the corresponding cross-section distribution displays a single peak with a slight shoulder. In the corresponding conformational families, some of the structural features of the lowest-lying isomers discussed in reference [28] are conserved. Namely, all of them are α -helices and the charge on Arg1 is solvated by the C-terminus carboxyl group. The two most compact conformers only differ by the orientation of the side chain of Arg1, and both show the same solvation scheme for the charged side chain of Lys11. The latter feature is not conserved in the most extended family. This leads to the unfolding of the C-terminus part and some shifting of the helix towards the N-terminus side. This structure, which lacks the highly favorable interaction between the charge on Lys11 and the electric dipole carried by the helix, might be an intermediate for the transition between the two other conformational families. This conformational variety found from the distribution in gyration radius could not be resolved from the cross-section distribution alone, the slight shoulder visible on the high cross-section side (see inset in Figure 2a) being insufficient as an indicator.

In the case of AlaD, no clear feature is visible on the evolution of $\langle R_g^2 \rangle$ with temperature, except for the increase at the unfolding range of $T=500\text{--}600\text{ K}$ (see inset in Figure 4). More insightful is the evolution of the distance between the N_ζ atom of Lys11 and the backbone carbonyl of Ala10 ($d_{N_\zeta\text{-CO}10}$). As displayed in Figure 4, this average distance is about 6.7 \AA at $T < 500\text{ K}$, which corresponds to the charge solvation scheme identified for the lowest-lying isomer. As discussed in reference [28], the interaction between the charge on Arg1 and the side chain carboxyl group of Asp leads to an α -helix shorter than the one assigned to AlaV. As a consequence, the solvation of the charge on Lys11 seems to be optimal if achieved by the backbone carbonyls of Ala7 and Ala8. However, at higher temperatures $d_{N_\zeta\text{-CO}10}$ strongly drops. As illustrated in Figure 4 by the probability distribution of $d_{N_\zeta\text{-CO}10}$ at 500 K, this corresponds to the formation of a new structural family where the charge on Lys11 is much closer to the carbonyl group of Ala10, $d_{N_\zeta\text{-CO}10}$ reaching about 2.7 \AA . The corresponding structures have a longer helix, similar to the one observed in AlaD. But Arg1 no longer interacts with the central side chain or with the C-terminus carboxyl. Moreover, in order to preserve the stabilizing interaction between the helix dipole and Lys11, charge solvation here has to involve Ala10, leading to shorter values of $d_{N_\zeta\text{-CO}10}$.

As a conclusion, the present WHAM-based strategy allows to identify structures possibly hidden in a single ATD peak. Such an approach would probably benefit to the interpretation of the experimental results obtained from the

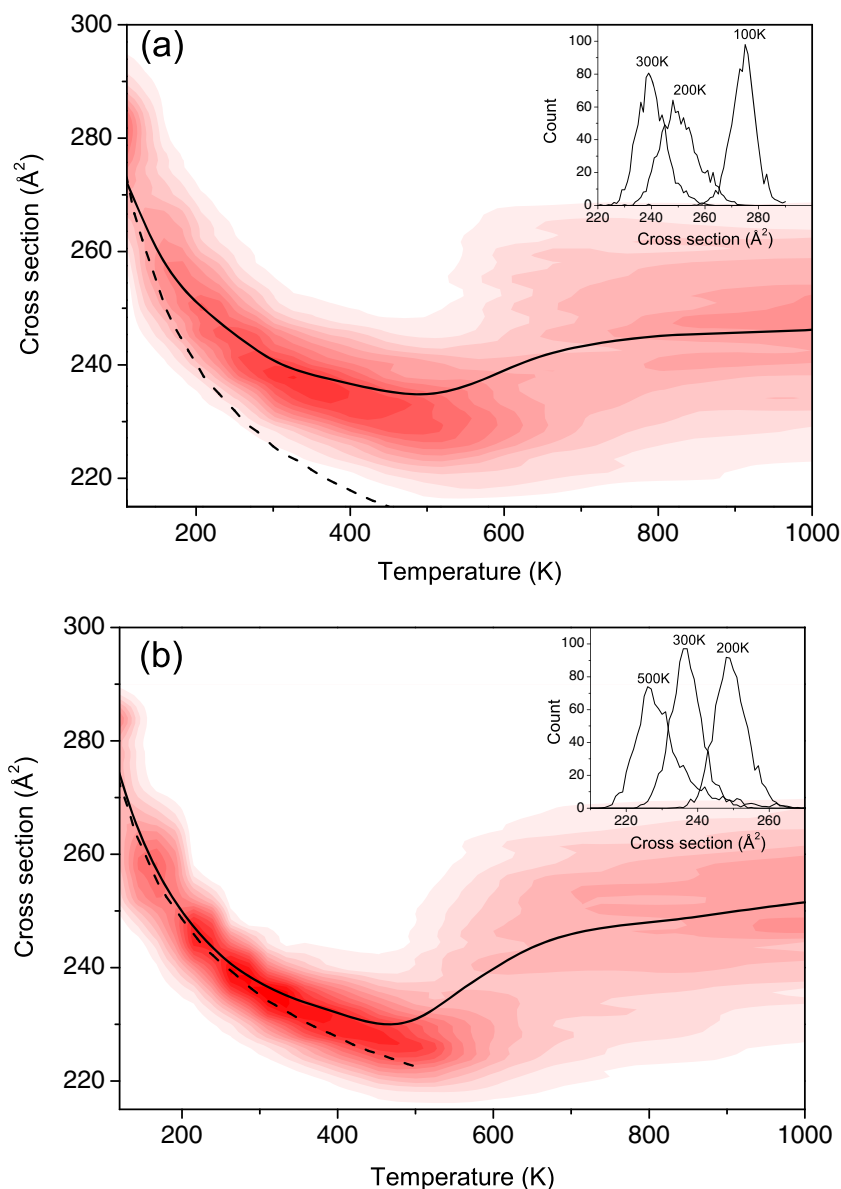


Figure 2. Average cross-section (solid line) and contour plot of its distribution versus temperature, calculated for **(a)** $[\text{RA}_4\text{VA}_4\text{K}+2\text{H}]^{2+}$ and **(b)** $[\text{RA}_4\text{DA}_4\text{K}+2\text{H}]^{2+}$ after histogram reweighting. The dashed lines illustrate the intrinsic temperature dependence of the cross-section for the (frozen) lowest energy structure, and the insets highlight the distributions at chosen temperatures

coupling of local conformational probes to IMS taken as a global conformational probe [32–37].

Improving Structural Assignment for Larger Peptides: M2TMP

Biasing the Sampling

The 11 amino acid long polyanilines investigated in the previous section were sufficiently small for their energy landscape to be successfully sampled by replica-exchange molecular dynamics. In this section, we deal with much larger molecules having 400 atoms or more, which can adopt myriads of conformations. The REMD simulations were first

performed under similar conditions as for the polyanilines, and the number of successful exchanges turned out to be satisfactory again despite a poorer overlapping between energy distributions. However, as will be seen below, and even after repeating the cycles of 10 ns per trajectory several times, we could not get reasonable agreement with the measured collision cross-sections. Performing systematic quenches to locate the putative lowest-energy structures, and restarting the REMD simulations from them did not significantly improve this agreement.

This led us to attempt another strategy consisting in guiding the sampling towards the desired regions of configurational space observed in the experimental distributions. Biasing sampling toward unfavored parts of the

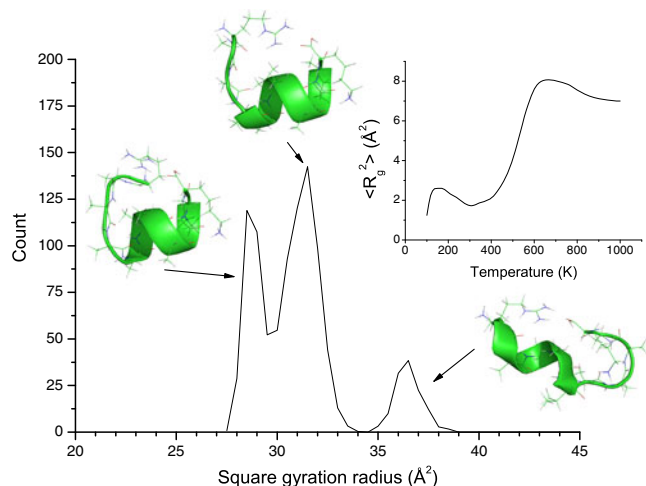


Figure 3. Distribution of square gyration radius calculated for $[\text{RA}_4\text{VA}_4\text{K}+2\text{H}]^{2+}$ at 200 K from histogram reweighting. Three typical conformations are depicted next to the corresponding peaks. The inset shows the variations of the thermally averaged square gyration radius versus temperature

landscape is an old idea in simulation, which probably dates back to umbrella sampling [13]. Here the active use of experimental information in the simulation meets several purposes. First, helping to reach an agreement can provide the structural information that is missing from direct, unsuccessful trajectories. Second, and even though the bias may be unphysical, it may correct for deficiencies of the original, unbiased potential energy surface. Third, facilitating exploration near the experimentally relevant regions may also favor the discovery of important minima that were hardly accessible by plain REMD.

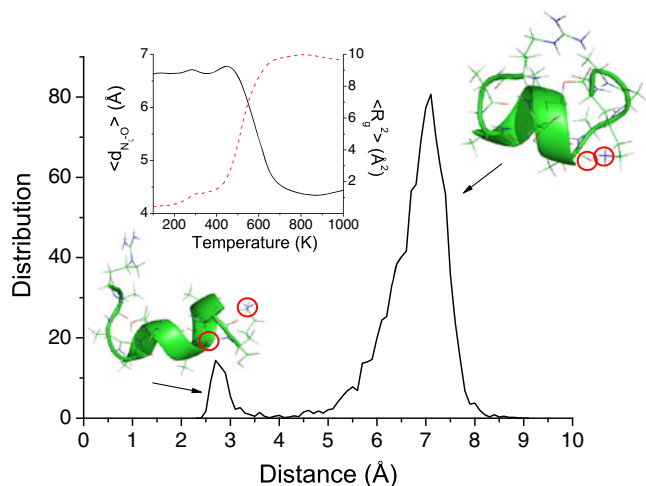


Figure 4. Distribution of the distance $d_{\text{Nz-O}}$ between the N_z atom of Lys11 and the backbone oxygen atom of Ala10 in $[\text{RA}_4\text{DA}_4\text{K}+2\text{H}]^{2+}$, calculated at 500 K from histogram reweighting. Two typical conformations are depicted next to the corresponding peaks, with the N_z and O atoms circled in red. The inset shows the variations of this thermally averaged distance (black solid line) and the average square gyration radius (red dashed line)

At this stage, it should be mentioned that a broad variety of sophisticated methods are available to systematically bias sampling based on specific order parameters [10, 11, 38, 39].

In our case, one difficulty arises due to the heavy computational cost associated with the calculation of Ω for a given conformation, which for the unpolarizable Amber force field exceeds energy and gradient evaluation by about three orders of magnitude (numbers obtained to get converged values of cross-sections for various peptides on an Intel Opteron single processor.)

This prevents acting on Ω directly, however, a possible substitute for the collision cross-section is the squared gyration radius R_g^2 , which has the same dimensions and is expected to behave similarly except for different sensitivity to detailed conformation. The evaluation of R_g^2 being straightforward, we implemented a simple umbrella potential \tilde{V} in the REMD simulation with form

$$\tilde{V}(\vec{R}, R_0^2) = \frac{1}{2}k(R_g^2 - R_0^2)^2 \quad (9)$$

R_0^2 being a target square gyration radius, k a spring constant chosen after several trials in such a way that a 25% deviation in R_g^2 is equivalent to $3Nk_B T_0 / 2$ energy when $T_0 = 100$ K.

The target value R_0^2 is experimentally unknown, but for a variety of molecules ranging from three to more than 400 atoms, we could find a satisfactory correlation between R_g^2 and Ω (see [Supplementary Material](#)). In practice, R_g^2 could thus be estimated from calculated cross-sections by the simple linear relation $R_g^2 \approx 6\Omega$.

Results

The computational protocol described above was applied to a variant of M2TMP-(22–46). In a previous study [14], a series of variants of this peptide were investigated using combined IMS and REMD simulations. Part of the recorded ATDs was consistent with elongated helices, while the other part was identified as originating from more compact structures. For one of the variants, denoted G13L (sequence: SSDPLVVAASIIILHLILWILDRL), the ATD turned out to be clearly bimodal, as shown in Figure 5. Simulations with the Amber force field have been carried out for this peptide, assuming protonation on the N-terminus and on the His and Arg residues side chains. Unbiased REMD simulations fail in reproducing the observed doubly peaked drift time distributions, as the calculated distributions fall between the two experimental peaks. From this first series of simulations, the lowest-energy structure happened to be α -helical over 43% of its sequence with a kink around Leu15, while remaining unstructured on its N-terminus part. The structures sampled from additional REMD trajectories at 300 K are essentially similar, but some of them are found to have larger cross-sections, with an α -helix extending towards the N-

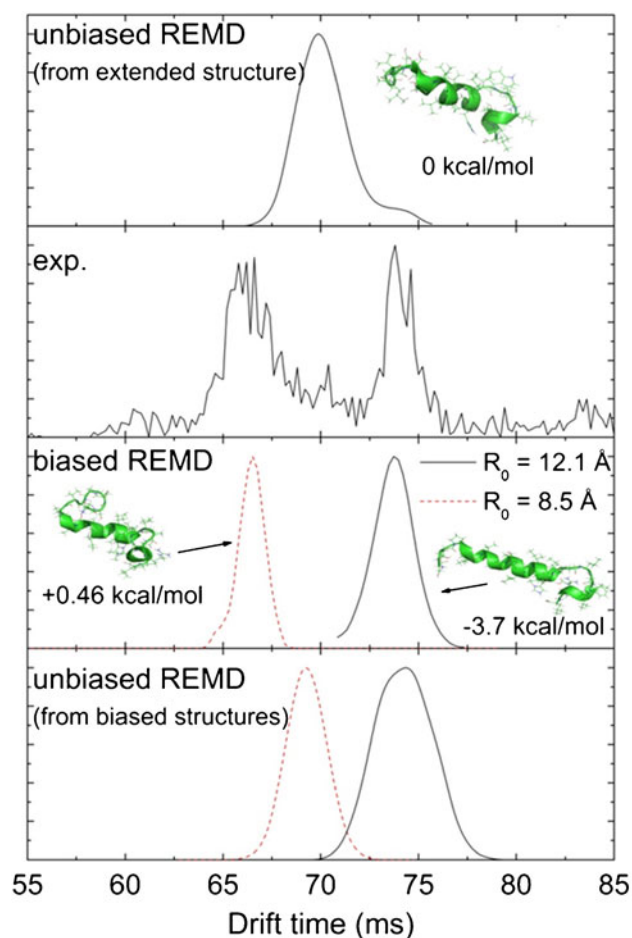


Figure 5. Experimental and calculated arrival drift time distributions for the G13L variant. The upper panel shows the results of unbiased REMD simulations at 300 K, the second panel is the experimental distribution, the third panel shows the two distributions obtained from biased REMD simulations with either $R_0=8.5\text{ \AA}$ (black solid line) or $R_0=12.1\text{ \AA}$ (red dashed line), the distributions obtained from an unbiased REMD run initiated with the structures identified from the biased simulation, with $R_0=8.5\text{ \AA}$ (black solid line) or $R_0=12.1\text{ \AA}$ (red dashed line), are displayed in the bottom panel. The three lowest-energy conformations found in each simulation are depicted near the corresponding distribution, with the value of the unbiased energy

terminus part. The contribution of these isomers is visible on the calculated ATD as a shoulder on the longer drift times side (see Figure 5). Structures with drift times compatible with the experimental values could actually be identified in the simulation but only as rare conformations.

We then turned to biasing the simulations, and initiated the trajectories using instantaneous structures from the first unbiased REMD run. These conformations were selected according to their cross-sections in agreement with the experimental values. The ATDs extracted from this biased simulations are displayed below the experimental data in Figure 5. The target values used for the biasing potential were $R_0=12.1$ and 8.5 \AA , respectively. The ATDs from the

biased simulations are in much better agreement with experimental results than unbiased data, yielding peaks centered at 67 and 73 ms, respectively. Interestingly, and although the ATDs differ, the structures obtained in the biased simulations are only mildly affected with respect to the low-energy, unbiased conformations. More precisely, guiding sampling towards more compact structures induces folding at the N-terminal end, the main part of the peptide essentially keeping its conformation without the bias. Guiding towards more extended structures also preserves a large part of the conformational arrangement. In both situations, only the N-terminus part is modified to adopt an α -helical structure, and the kink around Leu15 is preserved. The robustness of the latter conformational feature upon biasing the energy landscape highlights the possible importance of those secondary elements in stabilizing the global peptide structure.

The structures extracted from the biased simulation are not necessarily relevant for the realistic, unbiased force field. However, it is highly instructive to treat them as new candidate structures after removing the biasing term, hence they were locally optimized for comparison with the previously determined putative global minima. The relative energies resulting from these optimizations, given in Figure 5, show that the simulation guided towards higher cross-sections yields a structure that is actually slightly lower in energy than the minimum identified from the initial unbiased run. Such a small energy difference should, of course, be taken with some caution because it may lie outside the accuracy of the force field. However, and although we do not claim that this lower energy minimum is the most stable structure on the unbiased landscape, it is valuable to note that the inclusion of experimental information on the sampling process actually leads to better structural correlation.

Biased simulations thus provide candidate structures that are compatible with experimental data. However the shape of the ATDs extracted from the biased calculations should not be directly compared with the measurements because of the undesired and uncontrolled effects of the guiding potential. In order to obtain a more realistic picture of the conformational flexibility at 300 K around the conformers identified from biased simulations, additional unbiased REMD runs were initiated from these conformers and conducted following the same protocol as described in ‘Unbiased Replica-Exchange Molecular Dynamics’, but using only seven replicas in the 200–500 K range. The corresponding ATDs are displayed in Figure 5. The distribution for the most extended conformer is still centered around 73 ms but it is wider than the one from the guided simulation. This broadening better reflects the flexibility of the peptide and is in fairly good agreement with the experimental data. In contrast, the ATD for the more compact conformer is significantly shifted towards longer drift times and it is almost reproduced by the distribution obtained from the initial unbiased simulation. This is not surprising since the relative energy of the compact conformer was found to be above the initial minimum, and it seems that no sufficiently high barrier was present to prevent isomerization in this case. However, we keep in mind that the overall relevance

of the Amber force field may be questionable for such multiply-charged systems, so absolute energies may not be fully comparable. Unfortunately, polarizable models would be much more time-consuming and would probably not be feasible at a significant statistical level. The present work still shows that despite limitations inherent to the underlying force field, the use of a guiding potential allows a much more efficient exploration of the conformational landscape of complex systems and helps in providing realistic candidates for structural assignment.

Application to Triply-Protonated Bradykinin

We finally address one case of a relatively modest peptide for which ion mobility spectra obtained by Pierson *et al.* [15] have recently revealed conformational polystability, triply protonated bradykinine. In its doubly charged form, bradykinine stands as a calibration system for IMS, but the situation appears more complex in the triply charged state. In the experiments carried out by Pierson *et al.*, a variety of conformations was produced by electrospray ionization. Activation of selected isomers led to a trimodal distribution of cross-sections, independent of the initial selected conformation.

Standard (unbiased) REMD simulations were carried out for this system, the protonation sites being naturally taken as the most basic sites, namely at the N-terminus amine and on the two arginine side chains. In a first attempt, series of conventional replica-exchange MD trajectories, followed by systematic local minimizations provided a rough estimate of the lowest-energy structure, subsequently used to generate larger statistical samples at finite temperature. The computational details for these simulations are the same as those used for polyalanines, with the exception that the mobility calculations did not include the intrinsic dependence of the gas temperature on the collision cross-sections, but only the temperature dependence of conformations. This was motivated by the activation technique used in the experiments of reference [15], which could have a heating effect on the peptide. After processing by the weighted histogram method, we obtain distributions of collision cross-sections as a function of temperature. The distributions from the samples at 200, 300, and 500 K are shown in Figure 6. For these unbiased simulations, the results at room temperature indicate a main peak near 288 \AA^2 , in agreement with one of the experimental peaks within 3 \AA^2 . However, at 200 K, an additional peak clearly emerges corresponding to a more compact conformation near 278 \AA^2 . No other significant signature is found, even by considering the conformations sampled at 500 K that turn out to be more flexible, the distribution of cross-sections being essentially broader. The stable conformations corresponding to the two peaks in the distributions, also depicted in Figure 6, differ more in their side chains than in backbone arrangements, differences in intramolecular hydrogen bonds leading to a minor 0.6 kcal/mol gap. Although the peak at 288 \AA^2 is in very satisfactory agreement with measurements [15], the present

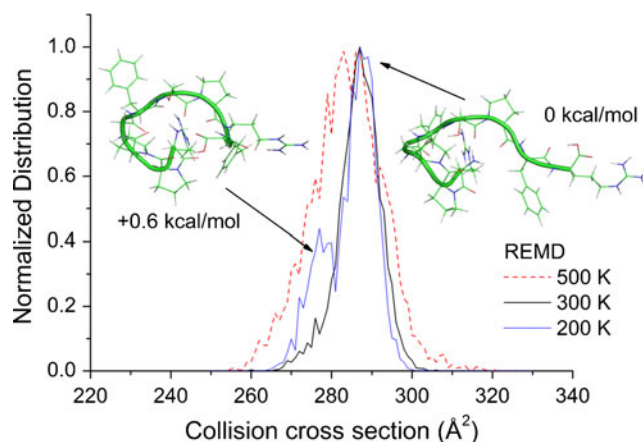


Figure 6. Distributions of collision cross-sections calculated at 300 K for triply-protonated bradykinine, using configurational samples obtained from unbiased replica-exchange molecular dynamics trajectories at 200, 300, and 500 K. Two low-energy conformations corresponding to the main peak and its left side shoulder are depicted

simulations provide no evidence for stable structures at the other experimental collision cross sections of 269 and 305 \AA^2 .

The failure of standard REMD simulations to locate all conformations that are compatible with experiments suggests to biasing the sampling along the lines of the previous section, by applying a series of guiding potentials, aiming in the vicinity of the experimental cross sections. However, even after taking larger values for the spring constant of the umbrella potential, we found it difficult to reproduce the measurements, hence the target gyration radii were given significantly lower or higher values than those expected based on simple linear correlation. We show in Figure 7 two distributions of cross sections we managed to obtain by employing $R_0=5.9 \text{ \AA}$ and $R_0=6.6 \text{ \AA}$. Upon quenching without the biasing potential, the stable

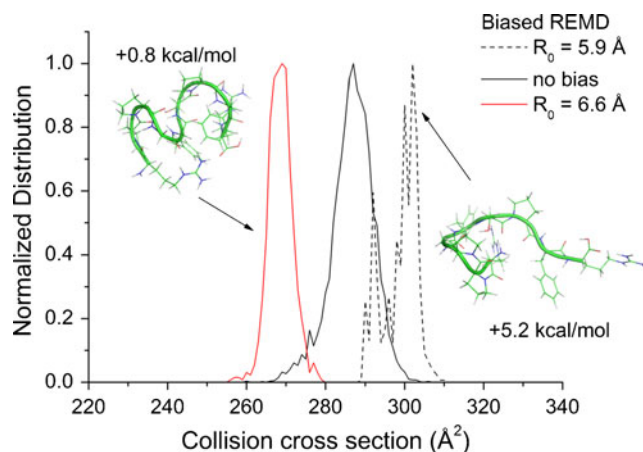


Figure 7. Distributions of collision cross-sections obtained at 300 K for triply-protonated bradykinine from replica-exchange molecular dynamics with different target biasing potentials corresponding to $R_0=5.9$ and 6.6 \AA , and from unbiased trajectories. The lowest-energy conformations are depicted next to the corresponding peak

conformations resulting from these samples lie in a rather narrow energy range, within about 5 kcal/mol from each other. Interestingly enough, and as was the case for the M2TMP subunit, several of these biased runs were able to locate conformations slightly more stable (by 0.5–1 kcal/mol) than the structure initially considered for the unbiased REMD trajectories.

The distributions of cross-sections determined for the biased simulations are shown in Figure 7. By mere construction, they are now in better agreement with the experiment, and the quenching procedure provides some insight into possible structural candidates that are both low in energy and have a collision cross section compatible with the measured value. Two such structures are shown in Figure 7 next to the corresponding peaks. The most compact isomer at 269 \AA^2 is only 0.8 kcal/mol higher in energy than the putative global minimum, but its backbone exhibits a clearly different pattern, with more hydrogen bonds but also more strain. In addition, this compact conformation allows the solvation of the Arg1 side chain by the C-terminus carboxyl group. The extended conformation near 305 \AA^2 , on the other hand, looks very similar to the most stable conformer except in the different arrangement of arginine side chains. This conformational similarity suggests that the two structures belong to the same funnel of the energy landscape, other isomers with comparable collision cross-sections being much higher in energy.

Conclusions

Ion mobility spectrometry is a powerful method for determining the global structure of gas-phase biomolecules, but it is bound to successful comparison with computed structures. Once a good agreement is reached, these theoretical structures can be exploited beyond the mere production of ATDs. The example of polyalanine-based peptides illustrates the capability of the combined REMD and WHAM analyses to reveal the distinct conformational families that may be hidden behind a single peak in ATD distributions. This confirms the great potential of the method as a practical tool to evaluate various conformers of a flexible molecule, without assuming any particular geometry beforehand, an impossible task based on experimental information alone.

In the case of more complex systems, for which the prediction of candidate structures is highly non trivial, we have shown how experimental measurements can be exploited to guide sampling towards physically relevant conformations. In particular, it was shown on the example of M2TMP variants and triply protonated bradykinine that such a biasing procedure can actually contribute in locating stable conformers on the original, unbiased energy landscape, and even to help improve the putative global minimum. This could be extremely useful from a computational perspective, for instance to refine the force field, or more practically to design two-stage assignment strategies where the bias itself

is self-consistently refined until agreement with experiment is reached. Guiding the simulation thus appears to be a good way to enforce conformational exploration towards regions of interest. Additionally, at least when approximate reaction coordinates such as the gyration radius are used in place of the diffusion cross-section, the biasing strategy has a low computational cost and is easy to carry out. Obviously, the efficiency of the sampling and biasing methods is naturally limited by the accuracy of the force field, which in particular neglects the polarization effects and the possible proton migration. Reactive approaches remain unpractical to combine with extensive simulations, but in the future they should be used more systematically.

In even larger peptides, it is not clear that the replica-exchange method will still be successful because of the larger number of replicas needed to obtain a satisfactory overlap between energy distributions. Flat-histogram biasing methods will then provide a possibly efficient alternative for producing sets of conformations that are broadly distributed in the range of reaction coordinate [11, 39]. However, the practical implementation of such biases requires fine-tuning of a larger number of parameters, which for the present biomolecules could turn out as a significant effort. In the purpose of systematically employing the present framework for a broader class of molecular compounds, it would also be useful to assess [40] and improve the quality of the biasing order parameter, by finding more accurate geometric quantities that correlate better with the diffusion cross-section than the gyration radius, which remains a phenomenological isotropic measure. A natural multidimensional extension could consider higher order moments of the gyration tensor, which measure sphericity or prolateness. Convex hulls or other global shape indicators could also be used. Such extensions will be explored in a forthcoming contribution. Finally, the present approach could also be of interest when using solution phase or crystal data as a guide for the calculations [4].

References

1. Jarrold, M.F.: Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **32**, 360–367 (1999)
2. Barran, P.E., et al.: Is it biologically relevant to measure the structures of small peptides in the gas-phase? *Int. J. Mass Spectrom.* **240**, 273–284 (2005)
3. Bohrer, B.C., Merenbloom, S.I., Koeniger, S.L., Hilderbrand, A.E., Clemmer, D.E.: Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **1**, 293–327 (2008)
4. Politis, A., Park, A.Y., Hyung, S.-J., Barsky, D., Ruotolo, B.T., Robinson, C.V.: Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS ONE* **5**, e12080 (2010)
5. Canon, F., et al.: Folding of a Salivary Intrinsically Disordered Protein upon Binding to Tannins. *J. Am. Chem. Soc.* **133**, 7847–7852 (2011)
6. Rodriguez C.F., Orlova G., Guo Y., Li X., Siu C.K., Hopkinson A.C., Siu K.W.M.: Gaseous Bradykinin and Its Singly, Doubly, and Triply Protonated Forms: A First-Principles Study. *J. Phys. Chem. B* **110**, 7528–7537 (2011)
7. Strittmatter, E.F., Williams, E.R.: Structures of Protonated Arginine Dimer and Bradykinin Investigated by Density Functional Theory: Further Support for Stable Gas-Phase Salt Bridges. *J. Phys. Chem. A* **104**, 6069–6076 (2011)

8. Wales, D.J.: Energy Landscapes. Wales D.J. Cambridge University Press, Cambridge (2003)
9. Zhou, J., Thorpe, I.F., Izvekov, S., Voth, G.A.: Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach. *Biophys. J.* **92**, 4289–4303 (2007)
10. Maragliano, L., Vanden-Eijnden, E.: A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem. Phys. Lett.* **426**, 168–175 (2006)
11. Laio, A., Parrinello, M.: Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002)
12. Sugita, Y., Okamoto, Y.: Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **314**, 141–151 (1999)
13. Torrie, G.M., Valleau, J.P.: Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **23**, 187–199 (1977)
14. Albrieux F., Ben Hamidane H., Calvo F., Chirot F., Tsybin Y.O., Antoine R., Lemoine J., Dugourd P.: Structural Preferences of Gas-Phase M2TMP Monomers upon Sequence Variations. *J. Phys. Chem. A* **115**, 4711–4718 (2011)
15. Pierson, N.A., Valentine, S.J., Clemmer, D.E.: Evidence for a Quasi-Equilibrium Distribution of States for Bradykinin [M+3H]³⁺ Ions in the Gas Phase. *J. Phys. Chem. B* **114**, 7777–7783 (2010)
16. Swendsen, R.H., Wang, J.-S.: Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **57**, 2607–2609 (1986)
17. Geyer, C. J. In: *Computing Science and statistics: Proceedings of the 23 rd Symposium on the Interface*. In: Keramidis, E.K. (eds.) Fairfax Station: Interface Foundation, 1991; p. 156
18. Neirotti, J.P., Calvo, F., Freeman, D.L., Doll, J.D.: Phase Changes in 38-Atom Lennard-Jones Clusters. I. A Parallel Tempering Study in the Canonical Ensemble. *J. Chem. Phys.* **112**, 10340–10349 (2000)
19. Frenkel, D., Smit, B.: *Understanding Molecular Simulation*. Acad. Press, NY (1996)
20. Martyna, G.J., Klein, M.L., Tuckerman, M.: Nosé-Hoover Chains: The Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992)
21. Wang, J., Cieplak, P., Kollman, P.: How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000)
22. von Helden, G., Hsu, M.T., Gotts, N., Bowers, M.T.: Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity. *J. Phys. Chem.* **97**, 8182–8192 (1993)
23. Shvartsburg, A.A., Jarrold, M.F.: An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **261**, 86–91 (1996)
24. Mesleh, M.F., Hunter, J.M., Shvartsburg, A.A., Schatz, G.C., Jarrold, M.F.: Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **100**, 16082–16086 (1996)
25. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., Kollman, P.A.: The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **13**, 1011–1021 (1992)
26. Ferrenberg, A.M., Swendsen, R.H.: Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195–1198 (1989)
27. Poteau, R., Spiegelmann, F., Labastie, P.: Isomerisation and Phase Transitions in Small Sodium Clusters. *Z. Phys. D* **30**, 57–68 (1994)
28. Albrieux F., Calvo F., Chirot F., Vorobyev A., Tsybin Y.O., Lepère V., Antoine R., Lemoine J., Dugourd P.: Conformation of Polyalanine and Polyglycine Dications in the Gas Phase: Insight from Ion Mobility Spectrometry and Replica-Exchange Molecular Dynamics. *J. Phys. Chem. A* **114**, 6888–6896 (2010)
29. Albrieux, F., Antoine, R., Chirot, F., Lemoine, J., Dugourd, P.: Ion trajectory simulations in a high-pressure cylindrical ion trap. *Eur. J. Mass Spectrom.* **16**, 557–565 (2010)
30. Revercomb, H.E., Mason, E.A.: Theory of Plasma Chromatography, Gaseous Electrophoresis—Review. *Anal. Chem.* **47**, 970–983 (1975)
31. Hudgins, R.R., Dugourd, P., Tenenbaum, J.M., Jarrold, M.F.: Structural Transitions in Sodium Chloride Nanocrystals. *Phys. Rev. Lett.* **78**, 4213–4216 (1997)
32. Polfer, N.C., Bohrer, B.C., Plasencia, M.D., Paizs, B., Clemmer, D.E.: On the Dynamics of Fragment Isomerization in Collision-Induced Dissociation of Peptides. *J. Phys. Chem. A* **112**, 1286–1293 (2008)
33. Fromherz, R., Ganteför, G., Shvartsburg, A.A.: Isomer-Resolved Ion Spectroscopy. *Phys. Rev. Lett.* **89**, 083001 (2002)
34. Misaizu, F., Hori, N., Tanaka, H., Komatsu, K., Furuya, A., Ohno, K.: Isomer-Selected Photoreactions of Gas-Phase Cluster Ions. *Eur. Phys. J. D* **52**, 59–62 (2009)
35. Joly L., Antoine R., Albrieux F., Ballivian R., Broyer M., Chirot F., Lemoine J., Dugourd P., Greco C., Mitric R., Bonacic-Koutecky V.: Optical and Structural Properties of Copper–Oxytocin Dications in the Gas Phase. *J. Phys. Chem. B* **113**, 11293–11300 (2009)
36. Pouilly J.C., Lecomte F., Nieuwjaer N., Manil B., Schermann J.P., Desfrancois C., Gregoire G., Ballivian R., Chirot F., Lemoine J., Calvo F., Antoine R., Dugourd P.: Combining Ion Mobility Mass Spectrometry and Infrared Multiphoton Dissociation Spectroscopy to Probe the Structure of Gas-Phase Vancomycin–Ac2LKDADA Noncovalent Complex. *Int. J. Mass Spectrom.* **297**, 28–35 (2010)
37. Lee, S., Zucker, S.M., Webber, N., Valentine, S.J., Clemmer, D.E., Reilly, J.P.: An Ion Mobility-Linear Ion Trap Hybrid Mass Spectrometer for Analysis of Biological Molecules. Proceedings of the Annual ASMS Conference, Denver (2011)
38. Darve, E., Pohorille, A.: Calculating Free Energies Using Average Force. *J. Chem. Phys.* **115**, 9169–9183 (2001)
39. Calvo, F.: Sampling Along Reaction Coordinates with the Wang-Landau Method. *Mol. Phys.* **100**, 3421–3427 (2002)
40. Calvo, F., Poulain, P.: Transitions between Secondary Structures in Isolated Polyalanines. *Eur. Phys. J. D* **51**, 15–23 (2009)