International Journal of Mass Spectrometry 367 (2014) 10–15

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Short communication

An atomistic structure of ubiquitin +13 relevant in mass spectrometry: Theoretical prediction and comparison with experimental cross sections

Moshe Goldstein[a,c,*], Liron Zmiri[a], Elad Segev[d], Thomas Wyttenbach[e], R. Benny Gerber[a,b,1]

[a] The Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
[b] Department of Chemistry, University of California, Irvine, CA 92697, USA
[c] Department of Computer Science, Jerusalem College of Technology (JCT), Jerusalem 91160, Israel
[d] Department of Applied Mathematics, Holon Institute of Technology, Holon, Israel
[e] Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 22 January 2014
Received in revised form 24 March 2014
Accepted 22 April 2014
Available online 2 May 2014

Keywords:
Highly charged protein
Theoretical prediction
Mass spectrometry
Ubiquitin +13

ABSTRACT

The 3D structure of protein ions in the gas phase is presently not obtainable from experiment in atomic detail. Here we use a theoretical approach to determine the 3D structure of ubiquitin +13 (UBQ +13) in the absence of solvent. Global minimization of the UBQ +13 force field within the recently developed DEEPSAM algorithm yields a nearly linear overall geometry. Four helical segments are found in this full atomistic structure – three of them are 3_{10}-helices and one is an α-helix. The protein cross section computed for the predicted structure is in excellent accord with ion mobility experimental results of UBQ +13. This suggests that computational structure predictions together with (theoretical and experimental) cross section values can serve as a useful tool for determining the atomistic structures of charged proteins in the gas phase.

1. Introduction

Over the last two decades mass spectrometry (MS) evolved into a powerful structural analysis tool in the field of biochemistry [1]. However, whereas the primary structure of a bio-polymer (its sequence) is the same in the presence or absence of solvent, it is less clear how its tertiary structure (its conformation or folding) changes in the transition from solution to the solvent-free environment of the mass spectrometer. Several careful studies [2–5] indicate that solution-like structures have a chance to survive the electrospray ionization (ESI) process [6], a method frequently used in biochemistry applications to transfer an analyte from solution into the mass spectrometer under gentle conditions. Once desolvated, tightly folded protein conformations are found to be frozen in place for the duration of a typical MS experiment (milliseconds) in the absence of energizing collisions [7,8].

Methods attempting to deliver structural information beyond primary structure in the gas phase include ion mobility spectrometry (IMS) [9], electron capture dissociation (ECD) [10], and vibrational spectroscopic methods [11,12]. However, unlike X-ray crystallography and solution NMR methods, these gas-phase approaches are generally not able to resolve macromolecular structures in atomic detail. IMS experiments, for instance, yield a cross section for collisions with buffer gas molecules, a measure of the size or overall shape of the polyatomic ion. A theoretical first-principles approach, based on available force fields, to generate molecular structures combined with rigorous cross section calculations for comparison with IMS experimental data appears to be an appealing protocol to predict structures in atomic detail in this case.

Abbreviations: UBQ, ubiquitin; DEEPSAM, Diffusion Equation Evolutionary Programming Simulated Annealing Method; EP, evolutionary programming; DEM, Diffusion Equation Method; SA, simulated annealing; RMSD, average root mean square deviation; L-BFGS, low memory BFGS quasi Newton local minimization algorithm.

* Corresponding author at: The Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, SRAEL Jerusalem 91904, Israel. Tel.: +972 2 6586265/+972 54 5334316; fax: +972 2 6513742.
E-mail addresses: goldmosh@fh.huji.ac.il, goldmosh@jct.ac.il (M. Goldstein), lironz79@hotmail.com (L. Zmiri), elad1segev@gmail.com (E. Segev), wyttenbach@chem.ucsb.edu (T. Wyttenbach), benny@fh.huji.ac.il (R.B. Gerber).
[1] Tel.: +972 2 6585732; fax: +972 2 6513742.

http://dx.doi.org/10.1016/j.ijms.2014.04.013
1387-3806/

In contrast to native state proteins, biopolymers electrosprayed under denaturing solution conditions end up in a high charge state [13] and therefore, due to Coulomb repulsion, in a significantly extended conformation [2]. Theoretical work indicates that such a highly charged desolvated polypeptide chain rapidly loses any structural elements present in solution and may quickly assume a structure fully adjusted to the solvent-free environment [14]. Therefore, highly charged proteins are optimal candidates for a theoretical search of low-energy gas-phase structures which can subsequently be tested by a comparison with experimental data. Very little is known about gas-phase equilibrium structures of proteins. The increased helical content present in a nonpolar milieu such as in organic solvents or in membranes [15–17] suggests helices might be favored in the nonpolar environment of vacuum or a buffer gas such as nitrogen or helium.

Here we use DEEPSAM, a global minimization algorithm developed by us [18], to search for the global minimum on the potential energy surface of a 13-fold protonated ubiquitin molecule (UBQ +13) in vacuum. Ubiquitin is a good choice for this type of study since this protein has been studied extensively not only in the condensed phase by X-ray crystallography, solution NMR, and other methods, but also in the gas phase by various MS-based methods. Charge state +13 is typically the most highly charged ion observed in an ubiquitin mass spectrum [19]. IMS-MS data indicate a significant deviation of the overall shape of UBQ +13 from a globular structure with a cross section about twice that of the native state [20]. This extended structure might contain helical sections since additional ECD-MS data available for UBQ +13 are consistent with a largely α-helical structure [21]. In contrast, earlier theoretical work on UBQ +13 [14] shows an extended structure without any helical content at all. In this study, the starting structure of the MD run was the A-state of ubiquitin +13 and the final conformation of the calculation corresponded to a high-temperature structure which is not necessarily representative of a properly equilibrated room temperature geometry nor of the global minimum on the UBQ +13 potential energy surface. Hence, it is clear that in the UBQ +13 structure a balance must be achieved (a) between the Coulomb repulsive forces among excess charges and the attractive forces present in the Hydrogen bonds of the secondary structural elements, and (b) between the dihedral potentials which give compact structures. So, of particular interest is the question whether an extended gas-phase equilibrium structure still contains any secondary structure.

In this paper, Section 2 introduces the UBQ +13 system and the theoretical methods employed. Results are presented, analyzed, and discussed in Section 3, and the concluding remarks are in Section 4.

2. Model system and methods

In general, different charge states of a protein lead to different structures with different cross sections. Here we focus on the protein ubiquitin in its charge state +13. Specifically, this UBQ +13 system corresponds to the fully desolvated, 13-fold protonated ubiquitin ion modeled in the gas phase. In this highly charged state, essentially all basic sites are protonated and there is little ambiguity regarding the distribution of the 13 protons. By far the most likely charge distribution is that originally suggested by Williams et al. [23] which is also used here and in our previous work [14]. In this configuration, the N-terminus, PRO19, and all basic amino acids are positively charged, except LYS29. Protonation of LYS29 is found to be unfavorable due to its proximity to the charged LYS27 residue [23]. While some other charge distributions might occur, their population is likely very small, and they are therefore not considered here.

The initial conformation employed here for our calculations is the X-ray structure 1UBQ of the native state [22] and it was protonated to a total +13 charge as described above. Ubiquitin's native structure is well characterized both in the crystal and in solution and it contains 3.5 turns of α-helix at residues 23–34, a 3_{10}-helix at residues 56–59, a parallel β-sheet formed by two strands at residues 1–7 and 64–72, and an anti-parallel β-sheet formed by three strands at residues 10–17, 40–45, and 48–50 (see Fig. 1a).

DEEPSAM [18], an Evolutionary Programming (EP) [24] algorithm designed for the global minimization of potential energies, has previously been successfully applied to the prediction of the native structure of neutral peptides. As usual in this kind of algorithms, it works on a "population" (an ensemble) of structures. In order to be able to start the evolutionary search loop, the first step of this kind of algorithms is the generation of the initial population. Because of considerations of run time and computational resources' use, the population in a DEEPSAM run is composed of only five conformations. In order to generate the initial population, DEEPSAM relies on an MD run that starts from the UBQ's native structure (PDB's 1UBQ). The five conformations of the initial population are selected among those found along the trajectory of that MD run. A conformation is selected to be included in the initial population if the average root mean square deviation (RMSD) between it and all the other selected structures is at least 4 Å. This geometric dissimilarity yields an acceptably wide initial distribution of conformations over the potential energy surface. It is important to emphasize that the choice of the starting structure in this initial run is irrelevant: Any point on the multi-dimensional potential energy surface of UBQ +13 is equally appropriate as a starting structure of a DEEPSAM run. The only relevant criterion is the structural diversity of the output of this initial run. Once the initial population is generated, the DEEPSAM's search loop starts. Briefly, at each iteration step, a new population is generated as a result of the application of specially designed random combinations of three different algorithms. One of them is the PES smoothing component of the Diffusion Equation Method (DEM) [25], another is Simulated Annealing (SA) [26] with MD where MD plays a partial role in the search for the global minimum of the PES, and the L-BFGS [27] quasi-Newton local minimization method. This way, the population evolves toward an approximation of the PES global minimum. The force field parameterization used in this research was TINKER's [28] implementation of amber98 [29].

Two methods are used to evaluate collision cross sections of protein structures in helium theoretically. In the first model, the exact hard-sphere scattering (EHSS) model [30], every atom is treated as a hard sphere undergoing hard-sphere collisions with helium. Whereas the resulting cross section values can be converged to an error of just a few square Angstrom, the systematic error could be much larger due to the unrealistic hard-sphere potential employed in this model. This shortcoming can be overcome by propagating trajectories in a Lennard–Jones potential [31] and with great effort this is doable for a globular protein structure. However, for extremely extended structures it becomes prohibitively expensive to converge this type of calculation to a meaningful accuracy and this rigorous trajectory method is not used here. Instead, the second model used here is the recently developed projected superposition approximation (PSA) employing helium buffer gas [32]. This model is expected to yield more realistic results than EHSS while still keeping the computational effort at a reasonable level (see Appendix A for details).

3. Results and discussion

Fig. 3 shows the UBQ +13 structure corresponding to the global minimum on the potential surface found by DEEPSAM. This