Biophysical Chemistry 230 (2017) 36–44

Contents lists available at ScienceDirect

Biophysical Chemistry

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

Ultracompact states of native proteins

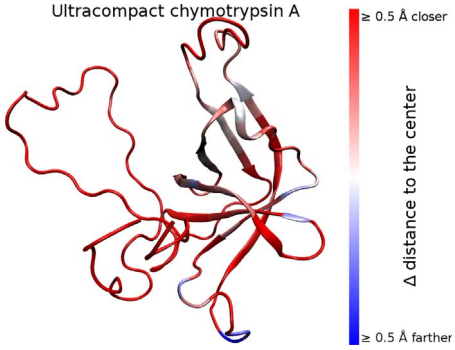
Leandro Grille Coronel[a], Juan P. Acierno[b], Mario R. Ermácora[a,b,*]

[a] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Saenz Pena 352, B1876BXD Bernal, Pcia. de Buenos Aires, Argentina
[b] Instituto Multidisciplinario de Biología Celular, Conicet, Calle 526 y Camino Gral. Belgrano, B1906APO La Plata, Buenos Aires, Argentina

HIGHLIGHTS

- Inspection of circa 20000 X-ray structures confirmed cryogenic protein contraction.
- Ultracompact native states found by statistical analysis of radius of gyration
- Ultracompact states have shorter van der Waals contacts and hydrogen bonds.
- Ultracompact states have more van der Waals interactions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Protein conformation
Cryocooling
X-ray crystallography
Radius of gyration
Protein glass transition
Protein energy landscape
Protein compaction
Protein intramolecular interactions

ABSTRACT

A statistical analysis of circa 20,000 X-ray structures evidenced the effects of temperature of data collection on protein intramolecular distances and degree of compaction. Identical chains with data collected at cryogenic ultralow temperatures (≤160 K) showed a radius of gyration (R_g) significantly smaller than at moderate temperatures (≥240 K). Furthermore, the analysis revealed the existence of structures with a R_g significantly smaller than expected for cryogenic temperatures. In these ultracompact cases, the unusually small R_g could not be specifically attributed to any experimental parameter or crystal features. Ultracompaction involves most atoms and results in their displacement toward the center of the molecule. Ultracompact structures on average have significantly shorter van der Waals and hydrogen bonds than expected for ultralow temperature structures. In addition, the number of van der Waals contacts was larger in ultracompact than in ultralow temperature structures. The structure of these ultracompact states was analyzed in detail and the implication and possible causes of the phenomenon are discussed.

1. Introduction

All the actual knowledge on the function of proteins is firmly grounded in the atomic description of the native state, which is generally defined in terms of atomic coordinates from X-ray diffraction data. Moreover, since the seminal contributions of Linus Pauling [1], protein folding theory and modeling rely on the energy of non covalent interactions estimates from inter atomic geometry and distances obtained from diffraction data.

A frequently overlooked aspect is the temperature dependence of diffraction data. Most of the structures in the RCSB Protein Data Bank (PDB; http://www.rcsb.org/; [2]) have been solved with data collection at cryogenic temperatures (≤200 K), after flash-freezing protein crystals. Although there is little doubt that this experimental condition permits a faithful representation of the native state at higher temperatures, the detailed effects of cryocooling on non covalent inter

* Corresponding author at: Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Saenz Pena 352, B1876BXD Bernal, Buenos Aires, Argentina
E-mail address: ermacora@unq.edu.ar (M.R. Ermácora).

http://dx.doi.org/10.1016/j.bpc.2017.08.004
Received 2 August 2017; Accepted 16 August 2017
Available online 18 August 2017

atomic distances need to be further investigated.

Several early studies assessed the effects of cryogenic temperatures on the atomic mobility in protein crystals. These studies confirmed that the atomic mean-square displacements are greatly reduced at temperatures below 200 K [3,4]. In addition, Hartman et al. reported in 1982 that the overall structure of metmyoglobin at 80 K was very similar to that at 300 K, but the volume of molecule was smaller [5]. According to these authors, the decrease in volume was accompanied by a shortening of non covalent intra molecular distances.

A few years latter, the volume contraction of metmyoglobin at cryogenic temperatures was further examined by Frauenfelder et al. [6]. It was found that the protein atoms were displaced toward the center of the molecule by 0.16 Å on average. Most importantly, most atoms were similarly involved in the displacement. The thermal compaction of metmyoglobin atoms was captured by the radius of gyration (R_g), which showed a contraction of 0.21 Å between 290 and 80 K. The authors also concluded that the compaction was not the result of an obliteration of the larger cavities within the molecule, but resulted from a general closer packing of all atoms.

In the early nineties, Earnest et al. [7] compared 120 and 295 K structures of rat trypsin, finding a reduction in the unit cell dimensions accompanied by significant decreases of 1.2% in molecular surface area and 0.2% in R_g at the lower temperature. Similarly, Tilton et al. [8] analyzed the structures of ribonuclease A at nine different temperatures ranging from 98 to 320 K, showing that the protein expands slightly (0.4% per 100 K) with increasing temperature and that this expansion was linear. Most inter atomic distances were involved in the change and this was evidenced by a significant change in R_g linearly related to temperature.

More recently, a survey of 15 crystal structures [9] showed that, on average and compared with room temperature, these proteins contracted by 1–2% at ~100 K. The unit cell also contracted on average 4–5% in volume. Accordingly, the average *Rg* of the compacted proteins was 0.53% smaller than at room temperature. Cryocooling also increased the number of intramolecular van der Waals contacts. A subsequent analysis of 30 crystal structure cryo-room-temperature pairs essentially confirmed the above trend in structure parameters and, in addition, showed that the cryogenic structures have superior packing compared to the isomorphous high-resolution room-temperature structures [10].

The structural cryocooling effects raised the interest on the dynamic of protein. Several independent experimental techniques provided insights into a peculiar transition centered at 200–220 K. Above 200 K internal protein motions could not be modeled only as harmonic motions of individual atoms. Instead, collective motions of groups of atoms superimposed to simple vibrations had to be included in the models. This extra, high-order mobility above 200 K was invoked to explain the strong temperature dependence of the mean-square atomic displacements.

The characteristics of the broad transition between a temperature regime dominated by atom centered harmonic motions and another that included in addition correlated motions of groups of atoms were reminiscent of a liquid–glass transition, and it was termed the 'glass transition' in protein dynamics [11–17]. It has also been proposed that an additional protein transition takes place at about 110 K, correlated to a cryogenic phase transition of water from a high-density amorphous to a low-density amorphous state [18].

Binding and functional experiments across the glass transition temperature provided ground to the concept that conformational flexibility and adjustment are necessary for protein function [15]. Recent confluent methodological advances in X-ray crystallography, NMR and computer simulations are beginning to reveal the structural details of protein conformational dynamics at high resolution [19]. These advances make possible dynamic structural biology studies at atomic resolution, across many orders of magnitude of timescales, and at temperatures in the 100–300 K range, linking conformational variation to function.

The application of the above methodological advances enabled a recent study of the conformational dynamics of cyclophilin A from 100 K to room temperature [20]. The authors report that many alternative conformations in cyclophilin A are populated only at 240 K and above, and others remain populated or become populated at 180 K and below. These results suggest a conformational heterogeneity between 180 and 240 K, involving thermal deactivation and solvent-driven attenuation of protein motions in the crystal.

Although most of the crystallographic evidence for the existence of the glass transition was from the differential linear behavior of mean-square atomic displacements – which combines static and dynamic information – such transition can be captured by a purely static variable. Indeed, a biphasic behavior of R_g as a function of temperature was reported by Teeter et al.[14]: cambrin R_g remains constant from 100 to 160 K and increases linearly from 160 to 293 K.

Recently, we showed that the compactness of the native state may be enhanced by protein engineering and established a new lower limit to the compactness of the Class-A β-lactamase fold [21]. In this work, we reported a 1.7-Å resolution X-ray structure of *Bacillus licheniformis* exo-small penicillinase mutant in which phenylalanine replaces wild-type tryptophan residues. The structure revealed no qualitative conformational changes compared with thirteen previously reported structures of the same protein, but it had a significantly smaller R_g. The importance of this finding is twofold. First, it suggests that temperature may not be the only factor involved in unusual protein compaction. Second, it shows that the subject can be further investigated by statistical analysis of the PDB structures.

In this work, we undertook a statistical survey of the PDB looking for unusually compact forms of protein structures. We will show that protein thermal contraction at cryogenic temperature is a generalized phenomenon. Furthermore, our analysis will show the existence of structures with a degree of compaction well beyond that attributable to normal thermal effects. The impact of ultracompaction on non covalent inter atomic distances will be also established.

2. Materials and methods

2.1. The analyzed set of protein chains

An initial list of protein chain IDs in the PDB was downloaded from PISCES (pdbaanr; 2015; [22]). The list includes 65,195 chain classes. Each chain class includes several experimental realizations of the same sequence in one or more PDB entries. The total number of chains in the initial set was 249,185, from 95,503 PDB entries. For instance, the hemoglobin W37A chain class includes four chains: 1A01 B, 1A01 D, 1A0 W B, and 1A0 W D from two different PDB entries (1A01 and 1A0A). All four hemoglobin chains have identical sequence and their structures can be considered experimental replicates of the same chain in different contexts.

The initial set was cleaned as follows: (*i*) only chains structures with a resolution better than 3.0 Å were retained, (*ii*) chains with missing atom coordinates, geometrical inconsistencies or other experimental anomalies were discarded, and (*iii*) chain classes including < 20 members were eliminated.

The final working set consisted of 631 chain classes, each with 20 or more replicates of structures from the same sequence chain. Summing all the classes, the sample space contained 19,393 chains from 7114 PDB entries.

2.2. Calculation of R_g

R_g describes the shape and size of a molecule by computing the dispersion of the individual atoms about either the mass or the geometrical center. In this work, R_g was calculated about the geometrical center, considering only main chain heavy atoms and disregarding mass