Molecular Dynamics Characterization of Protein Crystal Contacts in Aqueous Solutions

Giuseppe Pellicane
Dipartimento di Fisica, Università degli Studi di Messina, Contrada Papardo, 98166 Messina, Italy

Graham Smith
Henry Wellcome Laboratory for Biogerontology, Newcastle University, Newcastle upon Tyne, NE4 5PL, United Kingdom

Lev Sarkisov*
Institute for Materials and Processes, School of Engineering and Electronics, University of Edinburgh, Edinburgh, EH9 3JL, United Kingdom

(Received 2 October 2008; published 10 December 2008)

We employ nonequilibrium molecular dynamics simulation to characterize the effective interactions between lysozyme molecules involved in the formation of two hydrophobic crystal contacts. We show that the effective interactions between crystal contacts do not exceed a few kT, the range of the attractive part of the potential is less than 4 Å, and, within this range, there is a significant depletion of water density between two protein contacts. Our findings highlight the different natures of protein crystallization and protein recognition processes.

DOI: 10.1103/PhysRevLett.101.248102 PACS numbers: 87.15.km, 87.15.kr, 87.15.ap, 64.70.qd

The possible role and nature of anisotropic interactions in protein crystallization has been the subject of intense investigation. Indeed, the effort of the recent decades to understand protein interactions in aqueous solution and to describe quantitatively the phase diagram by means of isotropic models has proven to be an elusive task [1–8]. Globular proteins are not perfectly spherical, and their surface is structurally and energetically heterogeneous, leading to anisotropic protein-protein interactions.

No unified picture of these anisotropic interactions has been developed to date. A popular direction of research associates anisotropic contributions in water-mediated (solvation) protein-protein interactions with hydrophobic regions (patches) on the surfaces of the proteins [9,10]. Often, a ''patch-patch interaction'' term is introduced in the model to incorporate these effects [2,4,6,7,11–13].

Although the role of hydrophobic interactions is well established in protein-protein *recognition*, quantitative details of these interactions in crystallization processes remain largely unknown, which limits options for systematic construction of accurate models of protein solutions. Development of the quantitative description of patch-patch interactions poses several challenges. The first of them is associated with the very definition of a hydrophobic patch. Studies of the protein surfaces show that hydrophobic and hydrophilic residues on the surface of a protein form rather complex patterns [14]. Relatively hydrophobic regions can be identified and usually have complex shape and curvature [14]. Furthermore, two hydrophobic surfaces must be geometrically complementary to each other when they associate to form a patch-patch contact. These features are common for interactions involved in a very specific ''key and lock'' mechanism of protein recognition.

In this Letter, we use fully atomistic computer simulations to directly measure interactions involved in protein crystallization and generate some insights into their nature and quantitative details. In particular, we report clear evidence that their range is correlated with trapping/release of water molecules and their depth only slightly exceeds $1kT$. We assume that, if protein crystallization is governed by hydrophobic patch-patch interactions, some of these interactions will survive upon crystallization in the form of the actual crystal contacts. We explore interactions associated with the crystal contacts of lysozyme C. This system has been extensively studied over the years, and its phase behavior is well established [15]. The crystal structure of tetragonal hen-egg white lysozyme has been recently reported with a resolution of 0.94 Å using x-ray diffraction (entry 1iee in the Protein Data Bank) [16]. We employ the PISA (protein interfaces, surfaces, and assemblies) database to identify possible crystal contacts between two molecules in this crystal [17]. Each contact is characterized by the number of atoms involved in the interface, the surface area of the contact, and an estimate of the free energy gain upon the contact formation. PISA returns eight protein-protein contacts for the 1iee lysozyme. Three of the contacts have negative free energy and are also hydrophobic, although one of them is only very weakly hydrophobic. The other two contacts are selected for this study: namely, the one with the surface area of 51.2 Å^2 and 6 atoms of 4 residues participating in the contact (we label it contact A) and another one with the surface area 432.5 Å^2 and 48 atoms of 11 residues participating in the contact (contact B).

In order to calculate the potential of mean force (PMF) for these crystal contacts, we implement a protocol re-

cently proposed by Kosztin and co-workers and originally devised for water confined in single wall carbon nanotubes [18]. The protocol is based on the Crooks transient fluctuation theorem, which is a generalization of the celebrated Jarzynski equality [19]. Briefly, two proteins forming a particular crystal contact are placed in a simulation box containing water and ions. One of the proteins is fixed in space, whereas the other one is mobile. A stiff harmonic spring is attached to the center of mass of the mobile protein, and the latter is moved along the axis connecting the crystal contact areas. Mutual orientation of the proteins is preserved via a system of restraints. This pulling takes place either away from the original crystal contact (forward pulling F) or towards the fixed protein starting from some initial separation distance (reverse pulling R). As follows from the Crooks fluctuation theorem, within the stiff spring regime, the free energy difference ΔF in the system biased by the guiding harmonic potential is a good approximation for the PMF ΔU of the *unbiased* system: $\Delta U \cong \Delta F = \frac{\bar{W}_F - \bar{W}_R}{2}$, where $\bar{W}_F$ and $\bar{W}_R$ are the average work done by the spring in F pullings and R pullings, respectively.

The initial configuration of two proteins forming a crystal contact is generated by the PISA database [17]. Two proteins are placed at a desired distance between the contact areas (0 Å for forward pulling and 10 Å for the reverse pulling) and solvated with about 15 000 water molecules. Each lysozyme molecule carries a net charge of $+7e$, which corresponds to neutral pH conditions. Sodium and chloride ions are added to maintain the electroneutrality of the system and the ionic strength around $0.21M$. We refer the reader to additional information provided regarding the simulation details and how positional and rotational restraints are implemented [20].

Pullings are performed along the axis defined as follows. Atoms participating in the contact are identified. The principal axes of inertia are calculated for the participating atoms by assigning to them fictitious masses, proportional to the extent of their accessibility to the solvent. The fractional area exposed by the interface atoms to the solvent is calculated by means of the PISA database [17]. Two axes of inertia form a plane of contact, whereas the axis perpendicular to this plane constitutes the reaction coordinate and is easily visualized [20]. Twenty-five F-R pullings are performed for contact A for a total of 0.5 μs of simulation, and 100 F-R pullings are performed for contact B for a total of 2 μs of simulation [21]. W_z is calculated along each trajectory, and $\Delta U(z)$ is estimated from the averages $\bar{W}_{F\text{-}R}(z)$.

The results of computer simulations are presented in Fig. 1. The PMFs feature a short attraction region (not exceeding 4 Å in range), and the depth of the potential is about 3 kJ/mol (or about $1.2kT$) for the smaller contact A. The larger contact has not reached the expected potential plateau at larger separations; however, with the observed trend continuing, the depth of the potential should not exceed 4 kJ/mol (about $1.5kT$). Convergence of the results as a function of the number of sampled trajectories is also shown in Fig. 1; we note that the rate of convergence is much slower for contact B because of a significantly larger number of atoms (degrees of freedom) involved. The estimated error of the final result for both contacts is below 1 kJ/mol. This error can be further reduced simply by systematic accumulation of additional trajectories; however, we observe that, for the highest number of pullings, additional trajectories introduce only minor changes in the PMF profiles (Fig. 1). Examination of the presented data for contact B reveals a repulsive barrier of less than $1kT$ in height, in qualitative agreement with the Derjaguin-Landau-Verwey-Overbeek interaction potential (DLVO) theory at low-intermediate ionic strengths [15]. We note here that contact B contains three positive and two negative residues [17]. Therefore, the presence of this barrier, as a result of ion screening of the electrostatic repulsion between two charged contact areas, is expected from both a linearized Poisson-Boltzmann treatment of electrostatic interactions [15] and from an explicit consideration of charge correlations in a model with discrete point charges on the protein surface [8]. On the other hand, contact A has no polar residues, and, accordingly, the related PMF is monotonically increasing towards zero from the minimum.

Furthermore, it is interesting to compare these findings to the recent studies by Lund and co-workers on a coarse-grained model of lysozyme in aqueous solution, where the

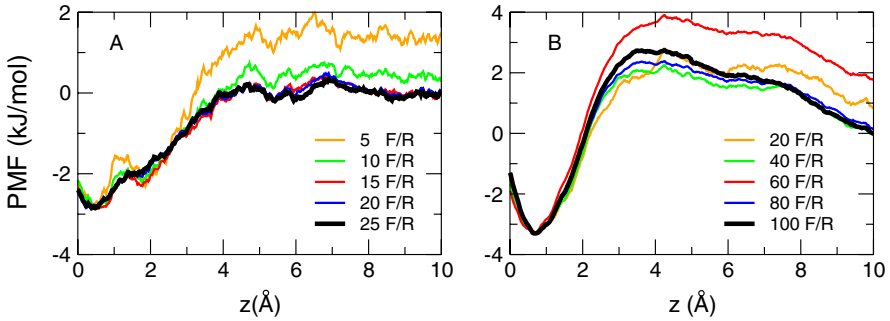

FIG. 1 (color online). Potential of mean force calculated for two crystal contacts A and B averaged over 25 (A) and 100 (B) forward and reverse trajectories. The process of convergence as the trajectories are accumulated is also shown.

solvent was treated as a dielectric continuum [22]. The orientation averaged PMF in their model has many features similar to the PMF calculated here. The magnitude of the attractive well depth in the work of Lund and co-workers did not exceed $1.5kT$, which compares well with our fixed-orientation measurements. However, the range of the effective interactions is much longer (30 Å) in their studies as it can be expected after averaging over the orientational degrees of freedom. It was further speculated that the observed higher concentration of anions in the vicinity of hydrophobic surfaces leads to more efficient screening of the repulsive electrostatic interactions between the proteins and stronger effective attraction between hydrophobic regions [22]. In this Letter, we also investigate the density profile of different solvent species in the cylinder surrounding the separation axis (the diameter of the cylinder is selected so that its cross-sectional area is equal to the area of the contact as reported by PISA) and observe a similar effect. As can be seen from the inset for the right graph in Fig. 2, for contact *B*, the region between two proteins has a systematically enhanced concentration of chloride ions compared to the bulk concentration, whereas the concentration for sodium ions is systematically depleted in the contact region. This is drastically different from the picture for contact *A*, where in the contact area most of the time we observe no ions of either type. Occasionally, a few ions may appear in the region; however, the error associated with such fluctuation is very significant (the inset for the left graph in Fig. 2). These results prompt more detailed studies of ion effects under different conditions and for other salts (NaI). Furthermore, Fig. 2 shows the reduced density profiles for water in between two proteins averaged over forward pulling trajectories. An important feature of these profiles is the density change taking place over the range of 3–4 Å from the contact, for both contacts. The attractive part of the PMF has a clear correlation with the length scale of a water molecule, and therefore we believe the attractive contribution to the PMFs in Fig. 2 is, at least partially, mediated by the desolvation of the interface as two proteins form a crystal contact. The significance of this contribution has been recently assessed by Vekilov *et al.* [23]. However, it is important to note that other factors, such as, for example, van der Waals forces, may also play a role and operate on the same length scale.

In summary, we have explicitly measured orientation restricted potentials of mean force for two lysozyme molecules in aqueous solution. These potentials characterize the effective interactions involved in the formation of two specific crystal contacts, and we observe that they are quite weak and their well depth does not exceed a couple of kT. Hence, these results point to a nonspecific nature of protein crystal contacts in agreement with several previous studies [24,25]. This is further supported by the well established protein crystal polymorphism: For example, pancreatic ribonuclease can crystallize in a number of space groups, with almost the entire surface capable of participating in a crystal contact [26]. However, our observations do not support a picture of protein crystallization as a purely stochastic process as has been also suggested by Carugo and Argos [25]. Indeed, it is quite evident that the magnitude of the observed interactions and even their shape *do* depend on the nature of the interfaces participating in the contact formation, thus leading to a slight anisotropy of protein interactions. Larger surfaces involving positively charged residues develop features, qualitatively predicted even by the DLVO theory, such as screened repulsion dominating at longer distances. The results reported here also confirm that the enhanced concentrations of anions do accompany formation of sufficiently large hydrophobic contacts. This effect is expected to influence the phenomenon of protein aggregation and ultimately protein crystallization. Finally, we reported quantitative evidence that the release of water molecules is an important driving force of

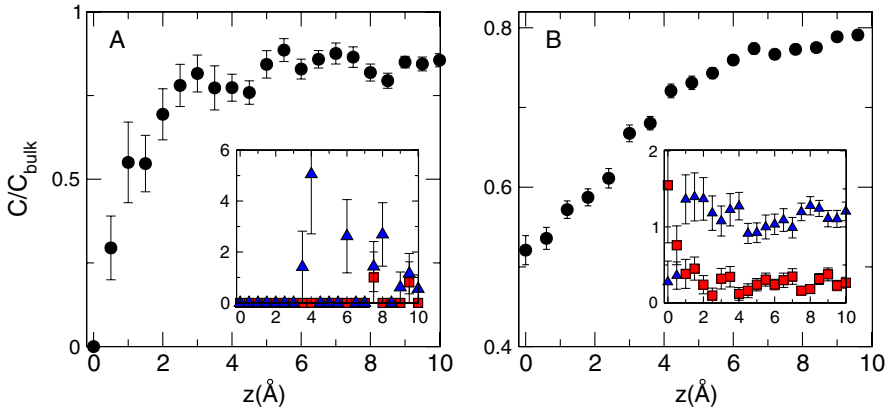

FIG. 2 (color online). Reduced concentration C/C_{bulk} profiles for water and ions (the insets) as a function of proteins separation z(Å) averaged over forward pullings (similar trends are observed for reverse pullings and are not shown for clarity). The left graph corresponds to contact *A*, and the right graph corresponds to contact *B*. Water is shown in black circles, whereas chloride ion concentration is shown in blue triangles and sodium ion concentration in red squares.

crystal contact formation, and this process depends on the nature, geometry, and size of the surfaces involved. We believe that systematic studies of protein crystal contacts with a fully atomistic detail, such as the ones employed in this Letter, may help to construct and calibrate accurate coarse-grained models of proteins in solution.

This work was carried out under the HPC-Europa + + project (Project No. 211437), with the support of the European Community—Research Infrastructure Action of the FP7. We also acknowledge Edinburgh Compute and Data Facility (ECDF), where the majority of the calculations were performed, and the PI2S2 Project managed by the Consorzio COMETA, a project cofunded by the Italian Ministry of University and Research (MIUR). We thank Professor C. Caccamo and Professor P. V. Giaquinta for useful discussions. G. S. acknowledges helpful discussions with Dr. S. Pal.

*Lev.Sarkisov@ed.ac.uk

[1] D. F. Rosenbaum *et al.*, J. Chem. Phys. **111**, 9882 (1999).
[2] R. P. Sear, J. Chem. Phys. **111**, 4800 (1999).
[3] R. A. Curtis, H. W. Blanch, and J. M. Prausnitz, J. Phys. Chem. B **105**, 2445 (2001).
[4] A. Lomakin, N. Asherie, and G. B. Benedek, J. Chem. Phys. **104**, 1646 (1996); Proc. Natl. Acad. Sci. U.S.A. **96**, 9465 (1999); N. Asherie, Methods **34**, 266 (2004).
[5] M. L. Broide, T. M. Tominc, and M. D. Saxowsky, Phys. Rev. E **53**, 6325 (1996); G. Pellicane, D. Costa, and C. Caccamo, J. Phys. Chem. B **108**, 7538 (2004).
[6] H. Liu, S. K. Kumar, and F. Sciortino, J. Chem. Phys. **127**, 084902 (2007).
[7] X. Y. Song, Phys. Rev. E **66**, 011909 (2002).
[8] E. Allahyarov *et al.*, Phys. Rev. E **67**, 051404 (2003).
[9] S. Jones and J. M. Thornton, J. Mol. Biol. **272**, 121 (1997).
[10] S. Leikin *et al.*, Annu. Rev. Phys. Chem. **44**, 369 (1993).
[11] F. Carlsson, M. Malmsten, and P. Linse, J. Phys. Chem. B **105**, 12 189 (2001).
[12] A. Shiryayev, X. F. Li, and J. D. Gunton, J. Chem. Phys. **125**, 024902 (2006); N. Wentzel and J. D. Gunton, J. Phys. Chem. B **112**, 7803 (2008).
[13] J. K. Cheung *et al.*, Biophys. J. **92**, 4316 (2007).
[14] P. Lijnzaad, H. J. C. Berendsen, and P. Argos, Proteins: Struct. Funct. Genet. **26**, 192 (1996); **25**, 389 (1996); P. Lijnzaad and P. Argos, *ibid.* **28**, 333 (1997).
[15] M. Muschol and F. Rosenberger, J. Chem. Phys. **103**, 10 424 (1995); **107**, 1953 (1997).
[16] C. Sauter *et al.*, Acta Crystallogr. Sect. D **57**, 1119 (2001).
[17] E. Krissinel and K. Henrick, in *Computational Life Sciences: Proceedings of the First International Symposium, CompLife 2005*, Lect. Notes Comput. Sci. Vol. 3695 (Springer, New York, 2005), p. 163; E. Krissinel and K. Henrick, J. Mol. Biol. **372**, 774 (2007).
[18] I. Kosztin, B. Barz, and L. Janosi, J. Chem. Phys. **124**, 064106 (2006).
[19] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[20] See EPAPS Document No. E-PRLTAO-101-048850 for the simulation details, schematic visualization of the simulation protocol, visualization of proteins in the initial configuration, and details regarding the restraints imposed in the system. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[21] Each trajectory (either forward or reverse) required about 80 hours of computational time using parallel runs on two dual-core Intel Woodcrest Xeon 5160 3.0 GHz equipped with low latency interconnection (Infiniband).
[22] M. Lund, P. Jungwirth, and C. E. Woodward, Phys. Rev. Lett. **100**, 258105 (2008).
[23] P. G. Vekilov *et al.*, Acta Crystallogr. Sect. D **58**, 1611 (2002).
[24] J. Janin and F. Rodier, Proteins: Struct. Funct. Genet. **23**, 580 (1995).
[25] O. Carugo and P. Argos, Protein Sci. **6**, 2261 (1997).
[26] M. P. Crosio, J. Janin, and M. Jullien, J. Mol. Biol. **228**, 243 (1992).