Available online at www.sciencedirect.com

ScienceDirect

Mendeleev Commun., 2023, **33**, 225–227

Mendeleev Communications

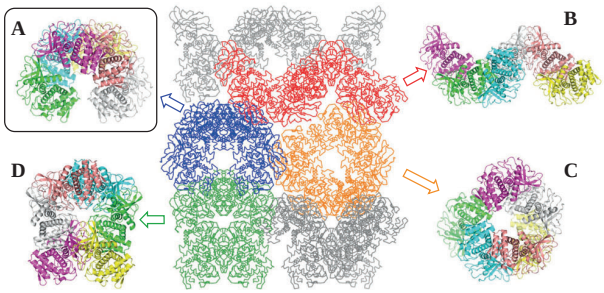
Identification of the precursor cluster in thermolysin crystallization solution by molecular dynamics methods

Yuliya V. Kordonskaya,*[a,b] Vladimir I. Timofeev,[b] Yulia A. Dyakova,[a] Margarita A. Marchenkova,[a] Yury V. Pisarevsky,[b] Svetlana Yu. Silvestrova[c] and Mikhail V. Kovalchuk[a,b]

[a] *National Research Center 'Kurchatov Institute', 123182 Moscow, Russian Federation. E-mail: yukord@mail.ru*
[b] *A. V. Shubnikov Institute of Crystallography, FSRC 'Crystallography and Photonics', Russian Academy of Sciences, 119333 Moscow, Russian Federation*
[c] *A. S. Loginov Moscow Clinical Scientific Center, 111123 Moscow, Russian Federation*

DOI: 10.1016/j.mencom.2023.02.024

An analysis of the crystal structure of thermolysin revealed four possible precursor clusters (hexamers) of its crystal. Using the method of molecular dynamics and plots of root mean square fluctuation, root mean square deviation and radius of gyration, the most stable hexamer, which is a precursor cluster, was determined. The importance of the established structure of the thermolysin precursor cluster for determining the mechanism of crystal formation is shown.



Keywords: protein crystallization, molecular modeling, molecular dynamics, precursor cluster, thermolysin crystal, crystallization solution.

Recently, opportunities have opened up to significantly accelerate the search for protein crystallization conditions by controlling the oligomeric composition of protein solutions, since it has been established that nucleation is preceded by the formation of precursor clusters that is special 3D fragments of the crystal structure. The presence and concentration of these fragments can be established, for example, using small-angle X-ray scattering (SAXS). Detection of pre-crystallization oligomers and measurement of their concentration make it possible not to wait weeks or months to find out whether protein crystals are formed, but to immediately and accurately determine the possibility of protein crystal formation. According to the SAXS data, it was found that for lysozyme, precursor clusters of a tetragonal crystal are octamers,[1] for proteinase K, dimers,[2] and for aminotransferase, dodecamers.[3] In the case of thermolysin, the SAXS method showed that hexamers are such building blocks.[4] It should be noted that not only hexamers, but also dimers, which are fragments of hexamers, were found in the thermolysin crystallization solution.

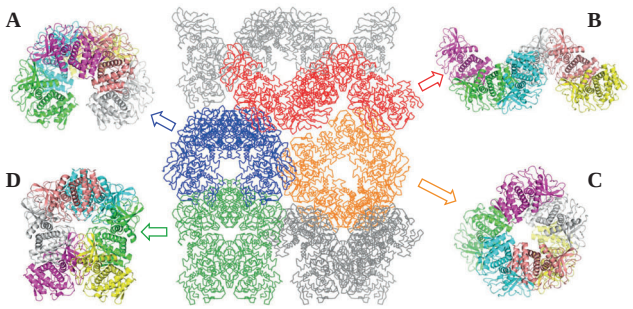


Figure 1 A fragment of a hexagonal thermolysin crystal (in the middle) and four possible types of crystal precursor clusters (hexamers **A**–**D**) extracted from it.

The SAXS method allows one to determine the size of the cluster, on the basis of which it is possible to establish the number of protein molecules that make up the precursor cluster, but does not provide information on the exact structure. However, it is important to know the structure of the oligomer in order to create a crystallization model. One of the most effective methods for determining the structure of a precursor cluster is molecular dynamics (MD). This is a universal tool that allows one to solve problems of various scales, from studying chemical reactions and their mechanisms[5] to determining the stability of molecules.[6] For the tetragonal modification of lysozyme, MD showed that one of the two possible octamers decomposes and, therefore, cannot be a precursor cluster, while the other is stable.[6] Similarly, using a combination of SAXS and MD methods, it was determined which of the six possible dimers is the precursor cluster of the proteinase K crystal.[7]

Previously, using SAXS, it was experimentally found that the precursor cluster of a thermolysin crystal consists of six molecules of this protein.[4] In this work, hexamer models were generated based on the structure of hexagonal thermolysin crystals deposited in the Protein Data Bank (PDB ID: 3DNZ) using the PyMol software.[8] An analysis of this structure made it possible to identify four possible precursor clusters **A**–**D** (Figure 1) from which a crystal can be built.

For each of these 3D fragments, MD calculations were conducted.[†]

[†] The protonation states of amino acid residues at pH 6.0 (according to experimental conditions) were determined using the PROPKA server (version 2.0.0).[9] All calculations were performed in the Amber ff99SB-ILDN

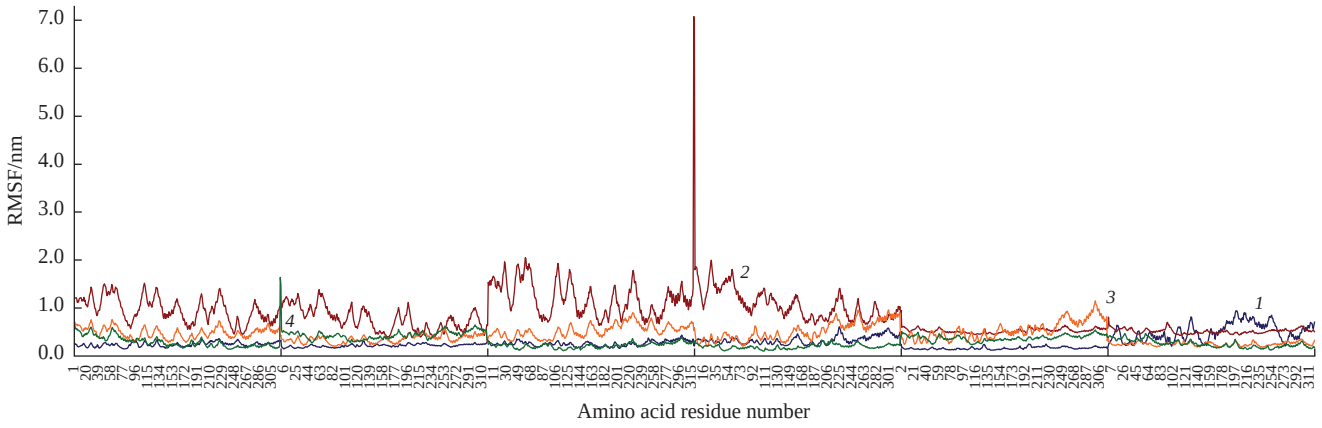


Figure 2 RMSF of C_α atoms of four different thermolysin hexamers (*1*) **A**, (*2*) **B**, (*3*) **C** and (*4*) **D** in solution with a precipitant at 20 °C for the last 25 ns of the trajectories.

The RMSF graphs of the C_α atoms characterize the stability of the molecules. Figure 2 demonstrates the RMSF of thermolysin hexamers **B**, **C**, **D** and **A**, listed in order of increasing stability. The RMSF values were calculated for the last 25 ns of the trajectories, when all the hexamers had confidently reached their equilibrium state. Obviously, hexamer **A** (see Figure 2, curve *1*) is the most stable, since its RMSF curve is located in Figure 2 below the others. On the contrary, hexamer **B**, which has the highest solvent-accessible surface area (see Figure 1), is the most unstable, with two amino acid residues reaching RMSF values of up to 2.4 nm according to Figure 2. Moreover, a visual inspection of the trajectory revealed that these residues (valine and lysine) broke away from hexamer **B**, and it itself dissipated into oligomers of a lower order.

The standard measure of the average distance between coordinates is RMSD, so we used the RMSD of all C_α atoms to examine the change in molecular structures during the simulation. It follows from Figure 3 that the structure of hexamer **A** remained the most similar to the initial one (as in the crystal) throughout the entire simulation, while hexamer **B** immediately began to undergo significant transformations. The RMSD of hexamers **C** and **D** are approximately the same, although slight differences are observed: hexamer **C** apparently does not finish changing after 100 ns, while hexamer **D** seems to achieve its equilibrium state.

Protein compactness was evaluated by such a characteristic as the radius of gyration (R_g). Figure 4 shows that the volumes of all hexamers increased in the course of their dynamics. However, the size of hexamer **A** almost returned to its original value by 73 ns, while hexamer **B** had the largest initial radius of gyration and underwent the most noticeable transformations.

From a comparison of the RMSF, RMSD and R_g plots for different hexamers, it follows that they are in good agreement: hexamer **A** is the most stable (RMSF) and retains the original structure (RMSD) and size (R_g) best of all; *vice versa*, hexamer **B** is the least rigid, the most voluminous and undergoes the most significant rearrangements; hexamers **C** and **D** take on intermediate values. Visual inspection of the trajectories also confirmed the conclusions drawn from the analysis of the plots.

The R_g value of the most stable hexamer **A** derived from MD simulations varies from 4.1 to 4.3 nm (see Figure 4), whereas the experimental R_g value obtained using SAXS[4] takes values in the range of 2.2–2.4 nm. It should be noted that for polydisperse systems, the R_g values determined by the SAXS method[4] do not

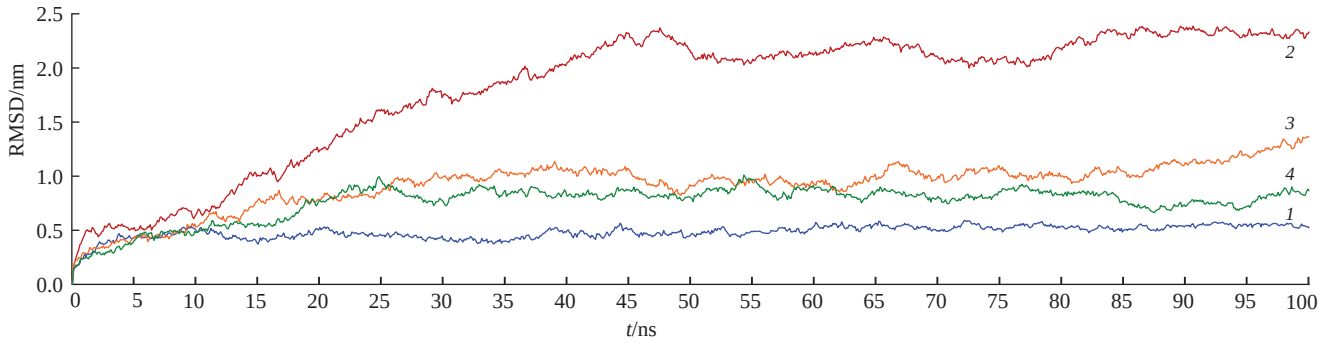


Figure 3 RMSD of four different thermolysin hexamers (*1*) **A**, (*2*) **B**, (*3*) **C** and (*4*) **D** in solution with a precipitant at 20 °C.

force field[10] using GROMACS (version 2021)[11] and applying 3D periodic boundary conditions. Since there are no parameters for NH_4^+ and SO_4^{2-} ions in the ff99SB-ILDN force field, the ammonium 3D structure was taken from PDBeChem (code: NH4) and the sulfate structure was obtained from the PLMD (Peptide Ligand Molecular Dynamics) python module[12] and was converted from the format .mol2 to .pdb format using the Antechamber algorithm.[13] Each hexamer was placed in the center of a cubic simulation box. The minimum distance between the box wall and any protein atom was 1 nm. The remaining space of the box was filled with the TIP4P-Ew water model.[14] The edge of the box was 14.4, 19.9, 14.3 and 15.2 nm for hexamers **A**, **B**, **C** and **D**, respectively. Some water molecules were replaced with ammonium and sulfate ions so that the salt concentration in the box was 0.75 M, according to the crystallization conditions.

The energy of each system was minimized using the steepest descent algorithm (50 000 steps) so that the force acting on any atom did not exceed 1000 kJ mol^{-1} nm^{-2}. Then, NVT and NPT equilibrations were performed using the modified Berendsen (V-rescale)[15] and Parrinello–Rahman methods,[16] respectively (for 100 ps each). The time integration step was set equal to 2 fs, the temperature and pressure were 293 K and 1 atm, respectively. Productive MD was calculated in the NPT ensemble using a modified Berendsen thermostat and a Parrinello–Rahman barostat. Integration was carried out according to the standard leap-frog algorithm.[17] The duration of each trajectory was 100 ns. RMSF (root mean square fluctuation), RMSD (root mean square deviation) and R_g (radius of gyration) were computed using the commands *gmx rmsf*, *gmx rms* and *gmx gyrate*, respectively.

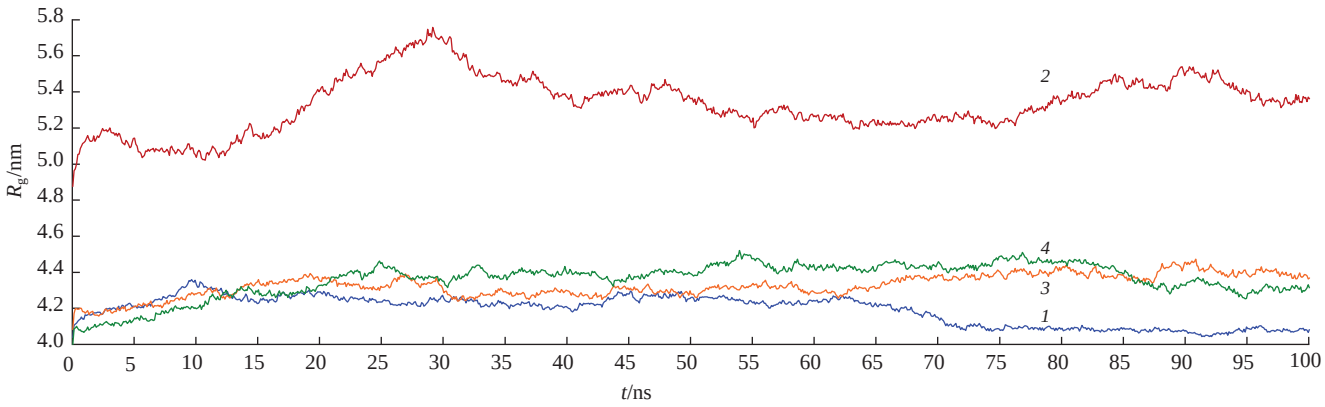


Figure 4 Radius of gyration of four different thermolysin hexamers (*1*) **A**, (*2*) **B**, (*3*) **C** and (*4*) **D** in solution with a precipitant at 20 °C.

correspond to an individual particle (hexamer), but to the average over the entire ensemble of monomers, dimers and hexamers. The volume fraction of thermolysin dimers in solution is only 4–8%, while for hexamers, it is even less: 1–3%. Thus, monomers with R_g = 2.1 nm (calculated from the X-ray crystal structure) make the largest contribution to the R_g value obtained using SAXS, and the presence of oligomers only slightly increases the R_g value from 2.1 to about 2.3 nm. Therefore, although it is not possible to determine the type of hexamer in the crystallization solution based on R_g data alone (because hexamers **A**, **C** and **D** have similar R_g values), the results obtained by MD and SAXS methods are still consistent and seem plausible, since the modeled R_g value of hexamer **A** (~4.2 nm) is well above the average one (~2.3 nm) determined by the SAXS method.[4]

According to the MD results, hexamer **A** is the most probable type of hexamer formed in solution before thermolysin crystallization. Therefore, we assumed that hexamer **A** is the precursor cluster of a hexagonal thermolysin crystal. The results greatly complement the experimental data, since SAXS only provided an approximate size and shape of the thermolysin pre-crystallization cluster,[4] but the precursor cluster model has now been established. Moreover, the molecular dynamics approach to identifying the building blocks of crystals was tested on one more protein (in addition to lysozyme[1] and proteinase K[2]).

According to the SAXS investigation,[4] only dimers and hexamers were observed in the thermolysin crystallization solution, the latter being proposed as a crystal precursor cluster. However, only two types of hexamers (**A** and **B**) were tested in that investigation,[4] and no differences were reported between the data processing results for these two structures. In this work, all possible types of hexamers (**A**–**D**) were considered and the most probable one was determined using a specific molecular dynamics approach, which was successfully used for similar purposes for several proteins (lysozyme[1] and proteinase K[2]), thereby gradually developing.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation directly (grant no. 075-15-2021-1363, contract no. 208 EP) and within the framework of the State Assignment of the FSRC ‘Crystallography and Photonics’ of the RAS and was carried out using the computing resources of the federal center for collective use ‘Complex for Simulation and Data Processing for Mega-science Facilities’ at the NRC ‘Kurchatov Institute’ (http://ckp.nrcki.ru/).

Online Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi: 10.1016/j.mencom.2023.02.024.

References

1 M. V. Kovalchuk, A. E. Blagov, Y. A. Dyakova, A. Yu. Gruzinov, M. A. Marchenkova, G. S. Peters, Y. V. Pisarevsky, V. I. Timofeev and V. V. Volkov, *Cryst. Growth Des.*, 2016, **16**, 1792.
2 A. S. Boikova, Yu. A. D’yakova, K. B. Il’ina, P. V. Konarev, A. E. Kryukova, M. A. Marchenkova, Yu. V. Pisarevskii and M. V. Koval’chuk, *Crystallogr. Rep.*, 2018, **63**, 865 (*Kristallografiya*, 2018, **63**, 857).
3 M. A. Marchenkova, P. V. Konarev, T. V. Rakitina, V. I. Timofeev, A. S. Boikova, Y. A. Dyakova, K. B. Ilina, D. A. Korzhenevskiy, A. Y. Nikolaeva, Y. V. Pisarevsky and M. V. Kovalchuk, *J. Biomol. Struct. Dyn.*, 2020, **38**, 2939.
4 M. V. Kovalchuk, A. S. Boikova, Y. A. Dyakova, K. B. Ilina, P. V. Konarev, A. E. Kryukova, M. A. Marchenkova, Y. V. Pisarevsky and V. I. Timofeev, *J. Biomol. Struct. Dyn.*, 2019, **37**, 3058.
5 I. V. Polyakov, M. G. Khrenova, B. L. Grigorenko and A. V. Nemukhin, *Mendeleev Commun.*, 2022, **32**, 739.
6 Yu. V. Kordonskaya, V. I. Timofeev, Yu. A. Dyakova, M. A. Marchenkova, Yu. V. Pisarevsky, D. D. Podshivalov and M. V. Kovalchuk, *Crystallogr. Rep.*, 2018, **63**, 947 (*Kristallografiya*, 2018, **63**, 902).
7 Y. V. Kordonskaya, V. I. Timofeev, M. A. Marchenkova and P. V. Konarev, *Crystals*, 2022, **12**, 484.
8 L. Schrödinger and W. DeLano, *PyMOL*, 2020, https://www.pymol.org/pymol.html.
9 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32** (suppl_2), W665.
10 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950.
11 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701.
12 M. Gruber, *PLMD: Peptide Ligand Molecular Dynamics*, 2022, https://github.com/MathiasGruber/plmd.
13 A. W. Sousa da Silva and W. F. Vranken, *BMC Res. Notes*, 2012, **5**, 367.
14 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665.
15 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.
16 M. Parrinello and A. Rahman, *J. Chem. Phys.*, 1982, **76**, 2662.
17 W. F. Van Gunsteren and H. J. C. Berendsen, *Mol. Simul.*, 1988, **1**, 173.

Received: 20th June 2022; Com. 22/6932