Identification of the Precursor Cluster in the Crystallization Solution of Proteinase K Protein by Molecular Dynamics Methods

Yuliya V. Kordonskaya 1,*, Vladimir I. Timofeev 2, Margarita A. Marchenkova 2 and Petr V. Konarev 1,2

Abstract: It is known that precursor clusters appear in solution prior to protein crystallization. For proteinase K, as it was found by SAXS, such clusters are dimers, but the accuracy of the method did not allow for determining its type. In this work, the behavior of six possible types of precursor clusters was simulated by the molecular dynamics technique. Stability analysis revealed the most probable type of dimer formed in the proteinase K solution before its crystallization. SAXS data modelling also supported the MD calculations. The dynamics of this precursor cluster was modeled at three temperatures: 20, 30, and 40 °C. At 40 °C, an abnormal increase in the stability of the thermophilic proteinase K was found.

Keywords: protein crystallography; molecular modeling; molecular dynamics; precursor clusters

1. Introduction

The importance of obtaining protein crystals is well known; they are necessary for using the method of X-ray diffraction analysis, which provides the greatest amount of information about the protein structure. At the same time, until recently, the crystallization of proteins is a difficult problem solved by a very laborious trial and error method.

The first step of the protein crystallization is nucleation. The classical theory of nucleation states that initial nuclei have the same structure and properties as the new emerging phase. The nucleus is a critical cluster that has the same order found at the end of the crystal formation.

Recently, the three-stage model of protein crystallization, in which it is preceded by the formation of an intermediate phase of precursor clusters, has become the most popular model [1,2]. As it is shown in a number of works [3–7], such clusters are 3D fragments of the crystal structure, which were experimentally detected by small-angle X-ray scattering (SAXS) and neutrons (SANS) for several proteins. Thus, for tetragonal lysozyme, precursor clusters are octamers [3]; for thermolysin, they are hexamers [8]; for aminotransferase—dodecamers [9]; and for proteinase K—dimers [10].

At the same time, precursor clusters of aminotransferase are unambiguously distinguished, but for the tetragonal modification of lysozyme, there are two types of octamers which are possible (A and B), and for thermolysin, there are four types of possible hexamers. For proteinase, there are six types of dimers that can act as precursor clusters (denoted as A, B, C, D, E, and F in Figure 1).
Figure 1. Possible types of precursor clusters for proteinase K protein crystals (dimers A–F).

It is interesting to note that both dimers and octamers were found in lysozyme solution, while only dimers were formed in proteinase K solution when studying the growth of protein crystals of the same system (P4_12_2). Thus, the final configuration of the crystal did not affect the type of precursor clusters, although certain similarities (the presence of dimers) were observed in the structure of lysozyme and proteinase K solutions at the pre-crystallization stage in the case of the tetragonal system.

The most stable oligomer can be determined using molecular dynamics calculations. Thus, for the tetragonal modification of lysozyme, it was shown that octamer “A” is stable, while octamer “B” quickly decomposes into oligomers of a lower order [11]. In the present work, similar studies of the precursor clusters’ stability were performed using the molecular dynamics technique of the proteinase K. For this protein, the SAXS method established that the precursor clusters are dimers [10]. Moreover, it was shown in [10] that the type of oligomer does not depend on the precipitant. In this work, we studied dimers formed in a crystallization solution using sodium nitrate as a precipitant (concentration was 0.5 M). For the defined precursor cluster, calculations were performed for several temperatures (at 20, 30, and 40 °C).

2. Materials and Methods

Preparation of oligomer models. Molecular models of possible growth units of proteinase K crystals were isolated from the tetragonal crystal structures with PDB ID: 7A68. The proteinase K crystal belongs to space group P4_12_2 with unit cell parameters a = b = 68.32 Å, c = 108.1 Å, α = β = γ = 90.00°. Using PyMOL program [12], the appropriate symmetry operators were applied to reconstruct a fragment of the protein crystal lattice and the coordinates of proteinase K’s dimers were obtained. As a result, six (A–F) different dimer models were built. Sodium ions bound with proteinase K (2 ions per molecule) were retained in the structure, and crystalline water was removed.

Calculation of the amino acid residues ionization states at pH 8.0 (in accordance with the pH values of corresponding crystallization solutions) was performed using PROPKA server (Version 2.0.0 [13]). The residues SER 15, SER 101, SER 140, SER 143, SER 191, SER 197, ASP 207, SER 216, SER 219, SER 247, and ASN 276 in 7A68 PDB file exist in the crystal in two conformations with the same occupancy factors (0.5) but only one conformation was retained.
Molecular dynamics protocol. The calculations were performed via the GROMACS 2021 software package [14]. Molecular dynamics were modeled in the Amber ff99SB-ILDN field [15], because it contains refined torsion potentials for some groups of atoms.

Each dimer was placed in the center of a cubic simulation box. The dimensions of the boxes were set in such a way that the minimum distance between their edge and any protein atom was 1 nm, forbidding the undesirable interactions between protein atoms caused by the periodic boundary effect. To fill the boxes with an explicit solvent, a 4-site water model designed to use Ewald summation methods (TIP4P-Ew [16]) was chosen. In order to reproduce a solution containing a precipitant, water molecules were replaced by sodium and nitrate ions so that the salt concentration in the box was 0.5M. Because there were no parameters for NO$_3^-$ ion in the ff99SB-ILDN force field, its 3D structure was obtained from PDBeChem (code: NO$_3$) and converted from .pdb to .itp format using the ACPYPE protocol [17] to generate the ion’s topology. The total charge of each box was neutralized by adding a negligible number of chloride ions (four), as it applies the PME algorithm for calculating long-range electrostatic interactions.

Before each start of the productive MD calculations, the energy of the systems was minimized by the steepest descent method (50,000 steps) until the force acting on any atom became less than 1000 kJ/(M·nm$^{-2}$) to eliminate atomic overlaps that existed in the crystal structures and that are produced by the addition of water. Then, the boxes were thermostated for 100 ps by the modified Berendsen (V-rescale) method [18] in NVT-ensemble and barostated for 100 ps by the Parrinello–Raman algorithm [19] in NPT-ensemble.

The productive MD simulation was conducted in an isothermal-isobaric ensemble using the V-rescale thermostat and the Parrinello–Raman barostat (at 1 atm) again. Integration was performed using a standard leap-frog algorithm [20] with the time step set at 2 fs. The simulations were conducted using three-dimensional periodic boundary conditions. Noncovalent interactions were considered only for atoms located within a radius of 1 nm. The long-range electrostatic interactions were processed by the smooth particle mesh Ewald (PME) summation method [21] with cubic interpolation and grid spacing in Fourier space of 0.16 nm. The dimers’ bond lengths were constrained using the LINCS algorithm [22].

Each of the calculated trajectories lasts 100 ns. The dynamics of each of the six different proteinase K (A–F) dimers were calculated once at 20 °C, whereas independent simulations of the dimer “E” were performed three times for each temperature value (20, 30, and 40 °C).

Before the trajectory analysis, the gmx trjconv command with the –pbc nojump flag was run to return atoms that had jumped across the box, so that all molecules remained whole. The reference protein structure for the structural alignment of the trajectories was taken from the file with the atomic positions after equilibration. Molecules’ trajectories were fitted to the reference ones by performing the command gmx trjconv with the –fit rot + trans flag, eliminating the protein parallel transfers and rotations around the axis, because such movements are irrelevant while considering protein stability. RMSF, RMSD, and R$_g$ were computed by using the commands gmx rmsf, gmx rms, and gmx gyrate, respectively.

3. Results
3.1. Characterization of the Dimers’ Stability at 20 °C

The RMSF graphs of the C$_\alpha$ atoms characterize the stability of the dimers. Because RMSF is defined as the fluctuation of each atom around its time-average position, it can be interpreted as a measure of protein flexibility. The more mobile an atom is, the higher its RMSF value.

Figure 2 and Table 1 demonstrate the RMSF of proteinase K dimers: F, B, D, A, C, and E (in ascending order of stability). It is obvious that dimer E is the most stable as its RMSF curve (pink) is the lowest in Figure 2 and averaged RMSF value, 0.15 nm, is the smallest, which is 1.6 times less than that of the next most stable dimer (C) with RMSF of 0.24 nm.
A standard measure of the average distance between coordinates is RMSD, so we used RMSD of all Cα atoms to examine the change in the dimer structures over the simulations. While the RMSF represents the fluctuations of each atom’s mean position during the whole dynamics, the RMSD demonstrates the deviation of all atoms at once as a function of time.

From Figure 3, it follows that the structure of the dimer E (pink) remained the most similar to the initial one (as in the crystal) throughout the whole simulation, while dimer F (red) immediately began to undergo significant transformations. Other molecules’ RMSD values are practically the same by the end of the dynamics (100 ns). Although the main trends preserve, as was expected, there are still slight differences between RMSF and RMSD results. For instance, the second in stability was dimer C and the third was dimer A, as RMSF data revealed (0.24 and 0.28 nm, respectively). However, according to RMSD plots, dimer C had begun to change its structure already in 8 ns and its significant transformations occurred after 35 ns, while dimer A experienced noticeable rearrangements only after 45 ns. Thus, despite dimer C being more stable than dimer A, dimer C changed its form faster than dimer A.

Table 1. RMSF of proteinase K dimers averaged over all Cα atoms.

<table>
<thead>
<tr>
<th>Dimer</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average RMSF, nm</td>
<td>0.28</td>
<td>0.37</td>
<td>0.24</td>
<td>0.36</td>
<td>0.15</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Figure 2. RMSF of six different proteinase K dimers (A–F) with the precipitant in a solution at 20 °C.

Figure 3. RMSD of six different proteinase K dimers (A–F) with the precipitant in a solution at 20 °C.
The protein compactness was evaluated by the radius of the gyration ($R_g$) characteristic. This is defined as the root-mean-square average of the distance of all atoms from the center of mass of the protein.

Figure 4 shows that the volumes of the dimers F, D, and E decreased in the course of their dynamics. Moreover, dimer F thickened quite sharply (from 57 to 62 ns), whereas dimer E became only slightly more compact. The sizes of the dimers A and C, on the contrary, moderately rose. The radius of gyration of dimer B almost did not change despite undergoing moderate fluctuations throughout the dynamics. Although dimer E began to diminish fast enough in 95 ns, it still has the most stable volume.

![Figure 4. Radius of gyration of six different proteinase K dimers (A–F) with the precipitant in a solution at 20 °C.](image)

Although the dimers (especially A, C, E, and F) have close $R_g$ values by the end of the simulation, visual inspection of the superimposed structures (performed by alignment tool implemented in PyMol [12]) revealed that none of them transformed into the other type.

Ultimately, because dimer E of proteinase K is the most stable and retains its original structure and size the best of all, it is the most probable dimer type to form in a solution before crystallization. Therefore, dimer E is the precursor cluster for proteinase K, so we continue to examine only this type of dimer in the following studies.

3.2. The Stability of the Dimers at 20, 30, and 40 °C

The values of the RMSF, RMSD, and $R_g$ were averaged over three independent simulations for each amino acid residue in the plots below. Therefore, each curve represents three independent trajectories modelled at a definite temperature.

Figure 5 illustrates that dimer E is the most stable at 20 °C (blue curve), which is the temperature at which the crystallization experiment was performed. However, the protein is more flexible at 30 than 40 °C (green and red curves, respectively).

Figure 6 demonstrates that the proteinase K structure remained almost the same as the initial (crystalline) one at 20 °C though it began to transform about 95 ns. At 30 °C, dimer E became remarkably mobile already at 40 ns while at 40 °C, it started to change later—only at 70 ns. Nevertheless, dimer structures changed to the same extent (approximately 0.7 nm) at both temperatures (30 and 40 °C) by the end of the simulations.

It is observed from Figure 7 that the dimer E almost retained its initial size at 20 °C. At 30 °C, it thickened sharply at 50 ns whereas the protein first became more compact at 77 ns and then increased its volume compared to the original one at 40 °C. Although the dimer volume decreased at 30 °C and grew at 40 °C by the end of the simulations, it changed to the same extent at both temperatures as in the case of RMSD.
3.2. The Stability of the Dimers at 20, 30, and 40 °C

The values of the RMSF, RMSD, and R-factor were averaged over three independent simulations. For each curve, the RMSF values were averaged over three independent trajectories. Figure 5 illustrates that dimer E is the most stable at 20 °C. This consistency is most likely due to the nature of the protein itself rather than its crystallization. However, the protein first became more compact at 30 °C and grew at 40 °C. As can be seen from Table 2, the volume of the proteinase K dimer E changed to the same extent at both temperatures as in the case of RMSD.

Figure 5. RMSF of proteinase K dimer E at 20, 30, and 40 °C. For each curve, the RMSF values were averaged over three independent simulations.

Figure 6. RMSD of proteinase K dimer E at 20, 30, and 40 °C. For each curve, the RMSD values were averaged over three independent simulations.

Figure 7. Radius of gyration of proteinase K dimer E at 20, 30, and 40 °C. For each curve, the Rg values were averaged over three independent simulations.
To check the consistency of the proteinase K dimers to experimental SAXS data [10], we have modelled a monomer–dimer equilibrium using the program OLIGOMER [23]. As can be seen from Table 2, the best fits (with the lowest discrepancy values $\chi^2$) were obtained for the mixture of monomers and dimers E. The volume fractions of dimers (v) increased with the temperature decrease from 30 to 10 °C.

Table 2. SAXS analysis using a monomer–dimer mixture for experimental SAXS data of proteinase K in the presence of 1 M NaNO\textsubscript{3} [10]. The most stable (according to RMSF values) proteinase K dimers (A, C, and E) were employed (PDB ID: 7A68). The theoretical curves for the monomeric and dimeric components were calculated using the program CRY SOL [24]. The best fit qualities ($\chi^2$) and volume fractions of proteinase K dimers (v) obtained by the program OLIGOMER [23] are reported.

| $T$, °C | Monomer + Dimer A | | | Monomer + Dimer C | | | Monomer + Dimer E | |
|---|---|---|---|---|---|---|---|
| | $\chi^2$ | v, % | | | $\chi^2$ | v, % | | | $\chi^2$ | v, % |
| 30 | 1.30 | 6.6 | | | 1.36 | 6.1 | | | 1.18 | 6.5 |
| 20 | 0.75 | 7.4 | | | 0.76 | 6.9 | | | 0.71 | 7.3 |
| 10 | 0.78 | 8.7 | | | 0.80 | 8.2 | | | 0.74 | 8.5 |

Comparison of the MD results with experimental SAXS data in Table 3 shows that they partially correlate: with an increase in temperature from 20 to 30 °C, dimers lose their stability and their volume fraction decreases. However, when it is increased by another 10 °C (to 40 °C), the dimers stabilize, although they do not reach the level of stability observed at 20 °C. This inconsistency is most likely due to the fact that proteinase K can be active in the temperature range of 37–60 °C. Thus, the relative stability of the dimer at 40 °C is probably due to the nature of the protein itself rather than its crystallization conditions.

Table 3. Comparison of different proteinase K characteristics obtained by SAXS and MD methods. The fit qualities ($\chi^2$) and volume fractions of monomers–dimers E (%) for proteinase K SAXS data in the presence of 1 M NaNO\textsubscript{3} [10] are obtained using the program OLIGOMER [23]. RMSF and RMSD values are averaged over three independent simulations and all C\textalpha atoms.

<table>
<thead>
<tr>
<th>$T$, °C</th>
<th>SAXS Data</th>
<th>RMSF, nm</th>
<th>RMSD, nm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monomer, %</td>
<td>Dimer E, %</td>
<td>Dimer E</td>
</tr>
<tr>
<td>40</td>
<td>—</td>
<td>—</td>
<td>0.24</td>
</tr>
<tr>
<td>30</td>
<td>1.18</td>
<td>93.5</td>
<td>6.5</td>
</tr>
<tr>
<td>20</td>
<td>0.71</td>
<td>92.7</td>
<td>7.3</td>
</tr>
<tr>
<td>10</td>
<td>0.74</td>
<td>91.5</td>
<td>8.5</td>
</tr>
</tbody>
</table>

4. Discussion

The data obtained experimentally by the SAXS method for proteinase K show that only dimers appear in a solution with a precipitant in the temperature range of 10–30 °C. According to the results of the work presented in Figures 2 and 3, it can be concluded that dimer E is the most stable and best retains the original structure; therefore, its formation in solution before the crystallization of proteinase K is the most probable. SAXS data modelling using a monomer–dimer mixture also supports this choice (Table 2). Therefore, dimer E is the precursor cluster of proteinase K.

MD results are in accordance with SAXS data at 20 and 30 °C as the stability of dimer E decreases with the temperature growth. However, at 40 °C, the precursor cluster begins to increase its flexibility later than at 30 °C, which is probably due to the fact that this enzyme is active and the protein molecule itself is stable up to 60 °C.

5. Conclusions

In the present work, using the MD and SAXS methods, it was determined which of the six possible dimers is the precursor cluster (building block) of proteinase K during the growth of its crystals.
Modeling of the proteinase K precursor cluster behavior at different temperatures has shown that molecular dynamics methods make it possible to trace the spatial and temporal changes in the internal structure of a crystal-forming dimer upon heating to temperatures far from proteinase K denaturation. The combination of MD and SAXS techniques seems to be optimal for revealing and characterizing the formation of precursor clusters in crystallization solutions.

**Author Contributions:** Conceptualization, V.I.T.; methodology, V.I.T.; validation, Y.V.K. and P.V.K.; formal analysis, Y.V.K. and P.V.K.; investigation, Y.V.K., P.V.K., V.I.T. and M.A.M.; resources, M.A.M.; data curation, Y.V.K. and P.V.K.; writing—original draft preparation, Y.V.K.; writing—review and editing, Y.V.K., V.I.T., M.A.M. and P.V.K.; visualization, Y.V.K.; supervision, M.A.M.; project administration, M.A.M.; funding acquisition, M.A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Higher Education grant No 075-15-2021-1362 in a part of calculating proteinase oligomers stability by molecular dynamics and within the State assignment FSRC [Crystallography and Photonics] RAS in a part of analyzing the results of computer modeling and determination proteinase precursor cluster.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work has been performed using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, http://ckp.nrcki.ru/ (accessed on 27 March 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

9. Marchenko, M.A.; Konarev, P.V.; Rakitina, T.V.; Timofeev, V.I.; Boikova, A.S.; Dyakova, Y.A.; Ilina, K.B.; Korzhenevskiy, D.A.; Yu Nikolaeva, A.; Pisarevsky, Y.V.; et al. Dodecamers derived from the crystal structure were found in the pre-crystallization solution of the transaminase from the thermophilic bacterium Thermobaculum termrenum by small-angle X-ray scattering. *J. Biomol. Struct. Dyn.* 2020, 38, 2939–2944. [CrossRef]
13. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann
 Chem.* **2005**, *26*, 1701–1718. [CrossRef]
 Phys.* **1995**, *103*, 8577–8592. [CrossRef]
 Chem.* **1997**, *18*, 1463–1472. [CrossRef]
23. Konarev, P.V.; Volkov, V.V.; Sokolova, A.V.; Koch, M.H.J.; Svergun, D.I. PRIMUS: A Windows PC-based system for small-angle