Computational analysis of protein conformational heterogeneity

Kristen Rhinehardt and Ming Dong

ABSTRACT

In this paper, we applied the molecular dynamics (MD) simulations and used thermolysin as the system to study the overall protein dynamics and side chain dihedral angles across the Arrhenius break. Simulations were performed at a high temperature of 36°C which is above the previously observed Arrhenius break, and the lower temperature of 20°C which is below the Arrhenius break. We observed different protein dynamics and conformational heterogeneity of side chain dihedral angles of thermolysin at the two temperatures. Our results indicated that certain regions of thermolysin have a higher level of fluctuation at lower temperature. A temperature dependent dihedral angles were also observed at the two temperatures. The changes observed in the protein indicated key areas of temperature sensitivity within the protein.

Introduction

Protein dynamics allow significant movement while maintaining biological structure and function, which are governed by atomic interactions in macromolecules, e.g. hydrogen bonding, electrostatic, and van der Waals interactions. Previously, various experimental approaches, e.g. X-ray crystallography, Kinetic Isotope Effect (KIE), Hydrogen Deuteron Exchange Mass Spec (HDX-MS), and neutron scattering have been applied to study the relationship of protein conformational heterogeneity and their temperature dependence (Dong et al., 2020; Doster et al., 1989; Hu et al., 2014; Keedy, 2015; Keedy et al., 2015; Liang et al., 2004; Liu & Konermann, 2008; Petsko & Ringe, 1984; Réat et al., 1998; Tilton et al., 1992; Zaccai, 2000). Interestingly, B-factors which reflect the amplitudes of thermal fluctuation and crystal disorder in protein crystallography showed a biphasic transition below its optimal temperature (Keedy, 2015; Tilton et al., 1992). Remarkably, studies utilizing HDX-MS strongly suggests a direct correlation between the time scale of conformational fluctuations and the turnover number of the enzyme (Liu & Konermann, 2008). Myoglobin, studied with neutron scattering showed extra mobility above the 180 K (Doster et al., 1989). Notably, a biphasic transition was observed in the Arrhenius plot of Alcohol Dehydrogenase (ADH), where above the break a temperature dependent KIE and a higher activation energy were observed (Nagel et al., 2011, 2012). Similarly, another thermophilic enzyme, thermolysin from Bacillus thermoproteolyticus rokko showed an Arrhenius break at a temperature range of 27–30°C which is much lower than its optimal temperature of 70°C (Dong et al., 2020).

Besides in vitro experimental methods, computational methods have also been used to study the protein conformational ensembles and the dynamics of proteins within a range of temperatures (Bowman & Geissler, 2014; Childers & Daggett, 2018; Glass et al., 2010; Jephthah et al., 2019; Merkley et al., 2010; Michetti et al., 2017; Sang et al., 2020; Zhang & Bruice, 2007). In the molecular dynamics simulation studies on B. Stearothermophilus alcohol dehydrogenase (ht-ADH) ES complex, a temperature dependent conformational change was observed and it was concluded that the temperature dependence of active site conformations is caused by both long-range and short-range motions of the ES complex (Zhang & Bruice, 2007). It was also shown that at higher temperature another thermophilic enzymes, nitroreductase and thermophilic serine protease which retain their native conformations and have a lower global flexibility to a much greater degree than their respective mesophilic enzymes, which indicated a higher stability that benefits their functions at higher temperature (Sang et al., 2020).

Previously, we used enzyme Thermolysin from Bacillus thermoproteolyticus rokko, a heat-stable zinc metalloproteinase (Fujii et al., 1983) as our system and studied the temperature dependence of protein structure and function via ambient temperature crystallography analyzed by Ringer and kinetics (Dong et al., 2020; Lang et al., 2010). Thermolysin hydrolyzes peptide bonds on the amino side of bulky hydrophobic residues such as Leu, Ile, Val and Phe (Fujii et al., 1983; Inouye et al., 1996). From the kinetics results, the activation energy of thermolysin was shown to be lower at temperatures above the Arrhenius break than below the break.
indicating a better optimized protein conformational sampling (Dong et al., 2020). Interestingly, there were temperature dependent conformational changes of side chains which were uncovered by Ringer analysis (Lang et al., 2010). More specifically, there were temperature dependent dihedral angle distributions above and below the Arrhenius break which suggested a temperature dependent alternate conformations of side chains at various sites, e.g. Glu143 at the active site, Met120 close to the active site, Asn227 on the surface of the protein, that leads to various populations of functional conformational ensembles of thermolysin above and below the Arrhenius break (Figure 1) (Dong et al., 2020).

Within these three residues, the Glu143 is the general base that takes the proton from the active water molecule which initiates the nucleophilic attack on the carbonyl carbon of the peptide bond of the substrate. The Met120 has been shown to indirectly affect the catalytic efficiency (Menach et al., 2012). The Asn227 mutation N227D showed a 5-fold decrease in the catalytic efficiency (de Kreij et al., 2002).

In this paper, we applied the molecular dynamics (MD) simulations, studied this thermostable enzyme and compared the protein conformational heterogeneity at higher temperature of 36 °C and lower temperature of 20 °C, which are above and below the Arrhenius break respectively (Dong et al., 2020). Our results showed that overall thermolysin structure appeared with a quicker stabilization and regional lower fluctuations at the higher temperature. Alternate dihedral angles were also observed in the side chains of the targeted three residues.

Methods

MD simulation

Molecular dynamics simulations were performed using GROMACS (Abraham et al., 2016). The protein structures are obtained from the RCSB data bank (PDB:5T9I) the zinc was placed in the configuration before simulation. These were then converted into a GROMACS .gro file. The simulation space is then defined with the molecule centered in a decahedron shaped cell with a minimum distance of 1.5 nm from the box edge. For stability simulations all systems were solvated in 100 mM NaCl to establish the baseline behavior. Minimization of the molecular system was performed using the steepest descent method, and the potential energy is analyzed to make sure that the system reached its most stable configuration. This configuration is then constrained and equilibrated using a canonical ensemble (NVT) followed by an isothermal–isobaric (NPT) ensemble. The NVT equilibration is performed with the temperatures (20 °C and 36 °C) coupled by a Velocity Rescale thermostat, which is a modified Berendsen thermostat specific to GROMACS. The NPT thermo-dynamical ensemble was performed with the same velocity-rescale temperature coupling in addition to the Parrinello-Rahmen pressure coupling. As volume is no longer constrained, the Parrinello-Rahmen barostat allows the simulation cell box to change its shape accordingly (Parrinello & Rahman, 1981). Both NVT and NPT equilibrations were performed for a time duration of 1 ns. The fully equilibrated system is used as the starting configuration for the MD dynamic analysis/production simulation. The mdp file has temperature coupling groups was done as ‘Protein Non-Protein’. All MD simulations were modeled for 185 ns.

Post-processing analysis methods and tools are applied to visualize and quantify the configurations of the protein-aptamer systems studied using Visual Molecular Dynamics (VMD) software and GROMACS (Humphrey et al., 1996). Visual analysis, though used throughout the simulation process, provided the first look into conformational changes, anomalies and binding occurrences that will provide guidance into further quantitative analysis of the dynamics simulation results of molecular level changes. The Radius of Gyration (Rg) is a measurement of the compactness of a structure (de Gennes, 1979). The Rg decreases and increases as atoms in a molecule move together and apart. When the molecule is stable, the Rg stabilizes at a finite value with little fluctuation indicating the minimal atomistic movements within the stabilized structure. The root mean square deviation (RMSD) was calculated to determine the changes in the overall structure over time. The root mean square fluctuations (RMSF) the of atomic positions in the trajectory are fit to the initial reference structure to determine the changes in specific amino acids over time (Abraham et al., 2016). The RMSF was calculated from the last 50 ns of simulation. Further comparison to in vitro experiments (Dong et al., 2020) was completed using χ angle analysis via the gmx chi function in GROMACS.

Results and discussion

The Root-Mean-Square Deviation (RMSD) and Radius of Gyration (Rg) plots were used as indicators to investigate the overall changes of protein structures. The RMSD of the protein backbone from the lower temperature of 20 °C showed more fluctuations at 50 ns, 100 ns, 135 ns and 150 ns, while at the higher temperature of 36 °C it showed a gradual increase at a range of 25–50 ns before the system plateaus with little change (Figure 2, Table S1). In the gyration plot, the gyration function in GROMACS.

During the 185 ns period of simulation, where at higher
Figure 1. (A) The structure of thermolysin (colored in cyan), and the active site pocket residues including Glu143 together with Met120 and Asn227 (colored in light grey). The color scheme of residues is displayed by heteroatom. The bound product dipeptide Val-Lys is also displayed in the active site (from PDB:5T9I). (B) The close-up of the active site of thermolysin.

Figure 2. The backbone RMSD overlay of thermolysin at higher temperature of 36 °C (A) (yellow: simulation 1, orange: simulation 2, red: simulation 3), and lower temperature of 20 °C (B) (blue: simulation 1, cyan: simulation 2, purple: simulation 3).

Figure 3. The overlay of the average structure from the last 50 ns simulation from high temperature of 36 °C (A) (simulation 1: yellow, simulation 2: orange, simulation 3: red) and low temperature of 20 °C (B) (simulation 1: blue, simulation 2: cyan, simulation 3: purple) (PDB used: 5T9I).
temperature the overall protein structure reaches the conformational homogeneity quickly and becomes ‘settled’ very quickly.

The Root-Mean-Square Fluctuations (RMSF) of the side chains showed similar fluctuations at higher and lower temperatures (Figure 4A). However, there are residues in the loop regions, e.g. resi 32–38, 92–99, which have notably higher fluctuations (a cutoff of 1.5-fold higher fluctuation was used) at the lower temperature (Figure 4B and 4C). Repeated simulations (sim 2 and sim 3) showed across all simulations general regions of peaks were consistent and the intensity of the fluctuations varied marginally (Figure S2). RMSF reflects B-factor or temperature factor for simulation systems, which are used to identify the flexibility of side chains of the protein. The B-factor from the crystal structure was also mapped and used as a reference (Figure S3), where the regions of resi 32–38 and 92–99 did not though show an obvious higher temperature factor, which could be due to the static disorder and scattering factors (Caldararu et al., 2019). On the other hand, the more flexible regions of thermolysin at the lower temperature of 20°C agrees with the previous HDX-MS study on thermolysin that was performed at 22°C (Liu & Konermann, 2008). Contrarily, when the temperature is higher at 36°C the fluctuations of these residues stay less temperature sensitive. Thus, the higher fluctuation of these residues at the lower temperature indicated an elevated temperature sensitive conformational sampling in these regions at the low temperature.

To look more closely at the temperature influences on the side chain conformational samplings, we analyzed the dihedral angles of amino acids from various regions of the protein at high and low temperatures across the proposed Arrhenius break temperature (Dong et al., 2020). Residue Asn227 which is on the surface of the protein has its \( \chi_1 \) angle distributed at two angles across the Arrhenius break. Both at lower temperature and higher temperature, the \( \chi_1 \) angle of Asn227 is distributed at \( -80^\circ \) and \( -180^\circ \) (Figure 5A and 5D, Figure S4). How this alternate rotamer of Asn227 contributes to the lower activation energy and lower value of A factor at higher temperature is still unclear. However previous studies on thermolysin mutation N227D has shown a 5-fold less catalytic efficiency compared to the wild type, where the proximity of amine group of Asn227 to the active center was compromised (de Kreij et al., 2002). From our results, the Asn227 has two \( \chi_1 \) angles of \( -80^\circ \) and \(-180^\circ \), where the \(-80^\circ \) is likely the more catalytically favorable based on its proximity to the active site. Residue Met120 which is close to the active site also has two dihedral angles distributed across the Arrhenius break (Figure 5B and 5E, Figure S5). Interestingly, at higher temperature, the \( \chi_1 \) angle of Met120 is mainly distributed at \(-180^\circ \) and minorly distributed at \(-80^\circ \) through the simulation. Contrarily, at lower
temperature, the $\chi_1$ angle of Met120 are mainly distributed at angles of $-80^\circ$ and $-180^\circ$ first, and then the $-80^\circ$ distribution almost disappeared after 75 ns. The $-80^\circ$ from our results could be more catalytically favorable which will be consistent with previous studies on mutation L144S, where the mutation stabilized the $\chi_1$ angle $-80^\circ$ via hydrogen bond, showed a 10 times higher catalytic efficiency than wild type (Menach et al., 2012). Residue Glu143 which is the active residue that initiates the nucleophilic attack on water molecule in the thermolysin catalysis mechanism, also has two distributed populations of dihedral angles of $90^\circ$ and $180^\circ$ across the Arrhenius break (Figure 5C and 5E, Figure S6). At higher temperature, two repeated simulation runs showed the $\chi_1$ angle of Glu143 is mainly distributed at $-90^\circ$. On the other hand, at lower temperature, the $\chi_1$ angle of Glu143 in the repeated runs is mainly distributed at both $-180^\circ$ and $-90^\circ$ (Figure 5E). As we know, when at $-90^\circ$ the Glu143 side chain is closer to the catalytic zinc, it favors the catalysis process. This is consistent with the previously observed lower activation energy of thermolysin at the higher temperature (Dong et al., 2020). Furthermore, our results agree with the previous research which showed that residues populate alternate rotamers (Bowman & Geissler, 2014; Fraser et al., 2009; Lang et al., 2010).

We know that at high-temperature environments, thermophilic proteins possess enhanced thermal stability in order to maintain their normal biological functions. For example, previous studies on the temperature dependent heterogeneity of thermophilic enzyme and their mesophilic counterpart showed that the thermophilic enzyme adopts a more compact conformation and lower global flexibility at the higher temperature (Merkley et al., 2010; Sang et al., 2020). The work from this paper probed a thermophilic enzyme’s protein conformational heterogeneity with temperature via MD simulation method and analyzed the functional related dihedral angles distributions of the residues above and below the previously proposed Arrhenius break. The results showed that a thermophilic enzyme stays more compact and regional more stable in its optimized functional temperature range. This paper also showed that certain side chain rotamers have alternate conformations with a temperature dependence, which could affect the catalytic efficiency of thermolysin. For future studies, simulations with substrate analogues need to be included to further analyze the interactions of substrate analogues and the enzyme thermolysin.

**Acknowledgements**

Computational facilities of the Computational Molecular Engineering Lab and the VICAR Center at North Carolina A & T State University.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Ming Dong (http://orcid.org/0000-0002-9121-6020)
References

Abraham, M. J., van der Spoel, D., Lindahl, E., Hess, B., & GROMACS development team. (2016). GROMACS 5.1.2 documentation. http://manual.gromacs.org/documentation/5.1.2/index.html

Bowman, G. R., & Geissler, P. L. (2014). Extensive conformational heterogeneity within protein cores. Journal of Physical Chemistry B, 119, 6417–6423.

Caldararu, O., Kumar, R., Oksanen, E., Logan, T., & Ulf Ryde, D. (2019). Are crystallographic B-factors suitable for calculating protein conformational entropy? Physical Chemistry Chemical Physics: PCCP, 21(33), 18149–18160. https://doi.org/10.1039/c9cp02504a

Childers, M. C., & Daggett, V. (2018). Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles. The Journal of Physical Chemistry B, 122(26), 6673–6689. https://doi.org/10.1021/acs.jpcb.8b02144

de Gennes, P.-G. (1979). Scaling concepts in polymer physics. Cornell University Press. https://books.google.com/books/about/Scaling_Concepts_in_Polymer_Physics.html?id=Apzf12LyWvGUC

de Kreij, A., van den Burg, B., Venema, G., Vriend, G., Eijsink, V. G. H., & Nielsen, J. E. (2002). The effects of modifying the surface charge on the catalytic activity of a thermolysin-like protease. The Journal of Biological Chemistry, 277(18), 15432–15438. https://doi.org/10.1074/jbc.M200807200

Dong, M., Lauro, M. L., Koblish, T. J., & Bahnson, B. J. (2020). Conformational sampling and kinetics changes across a non-Arhenuis break point in the enzyme thermolysin. Structural Dynamics, 7, 014101.

Doster, W., Cusack, S., & Petry, W. (1989). Dynamical transition of myoglobin revealed by inelastic neutron scattering. Nature, 337(6209), 754–756. https://doi.org/10.1038/337754a0

Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D., & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. Nature, 462(7273), 669–673. https://doi.org/10.1038/nature08615

Fujii, M., Takagi, M., Imanaka, T., & Alba, S. (1983). Molecular cloning of a thermostable neutral protease gene from Bacillus steaothermophilus in a vector plasmid and its expression in Bacillus stearothermophilus and Bacillus subtilis. Journal of Bacteriology, 154(2), 831–837. https://doi.org/10.1128/jb.154.2.831-837.1983

Glass, D. C., Krishnan, M., Nutt, D. R., & Smith, J. C. (2010). Temperature dependence of protein dynamics simulated with three different water models. Journal of Chemical Theory and Computation, 6(4), 1390–1400. https://doi.org/10.1021/ct9006508

Hu, S., Sharma, S. C., Scouras, A. D., Soudackov, A. V., Carr, C. A. M., Hamnes-Schiller, S., Alber, T., & Klinman, J. P. (2014). Extremely elevated room-temperature kinetic isotope effects quantify the critical role of barrier width in enzymatic C-H activation. Journal of the American Chemical Society, 136(23), 8157–8160. https://doi.org/10.1021/ja502726s

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14(1), 33–38.

Inouye, K., Lee, S.-B., & Tonomura, B. (1996). Effect of amino acid residues at the cleavable site of substrates on the remarkable activation of thermolysin by salts. Biochemical Journal, 315(1), 133–138. https://doi.org/10.1014/bj.315133

Jephthah, S., Staby, L., Kragelund, B. B., & Skepö, M. (2019). Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? Journal of Chemical Theory and Computation, 15(4), 2672–2683. https://doi.org/10.1021/acs.jctc.9b01281

Keedy, D. A. (2015). Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. Elife, 4, 1–26. https://doi.org/10.7554/eLife.07574

Keedy, D. A., Fraser, J. S., & van den Bedem, H. (2015). Exposing hidden alternative backbone conformations in x-ray crystallography using gFit. PLoS Computational Biology, 11, 1–22.

Lang, P. T., Ng, H.-L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., Holton, J. M., & Alber, T. (2010). Automated electron-density sampling reveals widespread conformational polymorphism in proteins. Protein Science, 19(7), 1420–1431. https://doi.org/10.1002/pro.423

Liang, Z.-X., Tsigos, I., Lee, T., Bouriotis, V., Resing, K. A., Ahn, N. G., & Klinman, J. P. (2004). Evidence for increased local flexibility in psychrophilic alcohol dehydrogenase relative to its thermophilic homologue. Biochemistry, 43(46), 14676–14683. https://doi.org/10.1021/bi049004x

Liu, Y.-H. Y., & Konermann, L. (2008). Conformational dynamics of free and catalytically active thermolysin are indistinguishable by hydrogen/deuterium exchange mass spectrometry. Biochemistry, 47(24), 6342–6351. https://doi.org/10.1021/bi800463q

Menach, E., Yasukawa, K., & Inouye, K. (2012). Effects of site-directed mutagenesis of the loop residue of the N-terminal domain Gly117 of thermolysin on its catalytic activity. Journal of Biochemistry, 152(3), 231–239. https://doi.org/10.1093/jb/mvs064

Merkley, E. D., Parson, W. W., & Daggett, V. (2010). Temperature dependence of the flexibility of thermophilic and mesophilic flavoenzymes of the nitroreductase fold. Protein Engineering, Design & Selection: PEDS, 23(5), 327–336. https://doi.org/10.1093/protein/gzp090

Michetti, D., Brandsdal, B. O., Bon, D., Isaksen, G. V., Tiberti, M., & Papaleo, E. (2017). A comparative study of cold- and warm-adapted Endonucleases A using sequence analyses and molecular Dynamics simulations. PLoS One, 12(2), e0169586. https://doi.org/10.1371/journal.pone.0169586

Nagel, Z. D., Dong, M., Bahnson, B. J., & Klinman, J. P. (2011). Impaired protein conformational landscapes as revealed in anomalous Arrenius prefactors. Proceedings of the National Academy of Sciences of the United States of America, 108(26), 10520–10525. https://doi.org/10.1073/pnas.1104989108

Nagel, Z. D., Meadows, C. W., Dong, M., Bahnson, B. J., & Klinman, J. P. (2012). Active site hydrophobic residues impact hydrogen tunneling differently in a thermophilic alcohol dehydrogenase at optimal versus nonoptimal temperatures. Biochemistry, 51(20), 4147–4156. https://doi.org/10.1021/bi3001352

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics, 52(12), 7182–7190. https://doi.org/10.1063/1.328693

Petsko, G. A., & Ringe, D. (1984). Fluctuations in protein structure from X-ray diffraction. Annual Review of Biophysics and Bioengineering, 13, 331–371. https://doi.org/10.1146/annurev.bb.13.060184.001555

Réat, V., Patzel, H., Ferrand, M., Pfister, C., Oesterhelt, D., & Zaccai, G. (1998). Dynamics of different functional parts of bacteriorhodopsin: H2H labeling and neutron scattering. Proceedings of the National Academy of Sciences of the United States of America, 95(9), 4970–4975. https://doi.org/10.1073/pnas.95.9.4970

Sang, P., Liu, S. Q., & Yang, L. Q. (2020). New insight into mechanisms of protein adaptation to high temperatures: A comparative molecular dynamics simulation study of thermophilic and mesophilic subtilisin-like serine proteases. International Journal of Molecular Sciences, 21, 3128–3140.

Tilton, R. F., Dewan, J. C., & Petsko, G. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. Biochemistry, 31(9), 2469–2481. https://doi.org/10.1021/bi00124a006

Zaccai, G. (2000). How soft is a protein? A protein dynamics force constant measured by neutron scattering. Science (New York, N.Y.), 288(5471), 1604–1607. https://doi.org/10.1126/science.288.5471.1604

Zhang, X., & Bricu, T. C. (2007). Temperature-dependent structure of the E x s complex of Bacillus steaothermophilus alcohol dehydrogenase. Biochemistry, 46, 837–843.