FEBS Letters 584 (2010) 3431–3435

journal homepage: www.FEBSLetters.org

Molecular simulations provide new insights into the role of the accessory immunoglobulin-like domain of Cel9A

Hanbin Liu [a,c], Jose Henrique Pereira [b,d], Paul D. Adams [b,d], Rajat Sapra [a,c], Blake A. Simmons [a,c], Kenneth L. Sale [a,c,*]

[a] Deconstruction Division, Joint BioEnergy Institute, Emeryville, CA, United States
[b] Technology Division, Joint BioEnergy Institute, Emeryville, CA, United States
[c] Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, CA, United States
[d] Lawrence Berkeley National Laboratories, Berkeley, CA, United States

ARTICLE INFO

Article history:
Received 30 April 2010
Revised 13 June 2010
Accepted 28 June 2010
Available online 4 July 2010

Edited by Robert B. Russell

Keywords:
Glycosylase-GH9
Cel9A
Immunoglobulin-like domain
Computational modeling
Molecular dynamic

ABSTRACT

Cel9A from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* belongs to the subfamily E1 of family 9 glycoside hydrolases, many members of which have an N-terminal Ig-like domain followed by the catalytic domain. The Ig-like domain is not directly involved in either carbohydrate binding or biocatalysis; however, deletion of the Ig-domain promotes loss of enzymatic activity. We have investigated the functional role of the Ig-like domain using molecular dynamics simulations. Our simulations indicate that residues within the Ig-like domain are dynamically correlated with residues in the carbohydrate-binding pocket and with key catalytic residues of Cel9A. Free energy perturbation simulations indicate that the Ig-like domain stabilizes the catalytic domain and may be responsible for the enhanced thermostability of Cel9A.

Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

1. Introduction

The growing importance of renewable fuels in displacing fossil fuels, providing energy security and reducing the risks associated with global warming, has spurred intense research into methods for conversion of lignocellulosic, non-food feedstocks, such as corn stover and switchgrass, into potentially low-carbon biofuels [1]. Existing biomass pretreatment processes typically rely on some combination of chemical and mechanical treatments at high temperatures and extremes of pH to break down the plant cell walls and liberate polysaccharides [2]. The next step in the biofuel conversion process involves the addition of enzymes, generally called cellulases and hemicellulases, to hydrolyze the liberated polysaccharides into monomeric fermentable sugars. It is desired to generate cellulases that can operate efficiently in the pretreatment chemical environments, such as extremes of pH and temperature, in order to reduce costs associated with the current approach of treating the effluent before saccharification. Thus, protein engineering for enhanced activity of cellulases that can tolerate higher temperatures and extremes in pH is an important field of ongoing research within the biofuel and enzymology scientific communities. There are examples of extremophilic cellulases in nature that possess unique structural components that may serve to stabilize the cellulase while maintaining activity at higher temperatures or extremes in pH. A fundamental understanding of the role of these auxiliary domains will provide insight into how cellulases are stabilized while maintaining activity at higher temperatures. These structural motifs may serve as a template for further enzyme engineering efforts of cellulases isolated from other environments and families.

Members of the glycoside hydrolases (GH) family have a modular architecture composed of one or two catalytic domains connected to several kinds of accessory domains. These accessory proteins include carbohydrate-binding domains (CBMs) [3], which enhance the association of the catalytic domains with insoluble carbohydrates, immunoglobulin-like (Ig-like) domains [4], and fibronectin-like domains [5]. While there is increasing structural information on the existence of these domains, their function and role in the enzymatic cycle remain open questions. Improved understanding of the structural and functional relationship between cellulases and these accessory domains provides information that may suggest an approach to rationally engineering

* Corresponding author at: Deconstruction Division, Joint BioEnergy Institute, Emeryville, CA, United States.
E-mail address: Klsale@lbl.gov (K.L. Sale).

0014-5793/$36.00 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2010.06.041

cellulases for industrial hydrolysis of cellulose in extreme temperature and pH environments.

For example, Cel9A from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius* (Aa_Cel9A) (Fig. 1-top) has a temperature optimum of 70 °C and a pH optimum of 5.5 [6]. Aa_Cel9A belongs to the subfamily E1 of family 9 of glycoside hydrolases, many members of which have an N-terminal Ig-like domain followed by the catalytic domain (CD). The function of the Ig-like domain has not been determined; however, deletion of the Ig-like domain results in a complete loss of enzymatic activity in a related cellobiohydrolase, CbhA from *Clostridium thermocellum* [7]. Our experiments also show that deletion of the Ig-like domain results in significantly decreased recombinant protein expression.

We have investigated the role of the Ig-like domain in stabilizing Aa_Cel9A and the correlation of its motions with motions of the catalytic domain using molecular dynamics (MD) simulations using the recently solved crystal structure of Aa_Cel9A [8] and an in silico Aa_Cel9A mutant in which the Ig-like domain was removed. Using MD simulation we have investigated: (1) the presence or absence of correlated motions between active site atoms and those within the Ig-like domain, (2) the effect that the Ig-like domain has on the orientation of catalytic residues in the active site, and (3) the contributions of the Ig-like domain to protein stability.

2. Theory and method

Aa_Cel9A is represented in atomic detail using the AMBER force field ff03 [9]. Initial coordinates for our simulations were taken from the recently solved 2.3 Å structure (PDB ID 3EZ8) of Aa_Cel9A [8]. Simulations were carried out both the wild-type Aa_Cel9A and on a truncated wild-type Cel9A structure, in which the IG domain residues (residues 7–87) were deleted from the wild-type structure. The protein was immersed in a pre-equilibrated solvent box of TIP3P water molecules [10], which extended to 10A beyond the outermost protein atoms on both sides of the *x*, *y*, *z*-axes. The SHAKE method was used to constrain all bonds involving hydrogen atoms. The cutoff radius of non-bond interactions was set to be 12 Å. The Particle Mesh Ewald summation method [11] was used to calculate the electrostatic potential. The entire simulation system of WT Aa_Cel9A contained 528 amino acid residues, 2 Ca^{2+} ion, 1 Zn^{2+} ion, 15 Na^{+} ions and 13746 water molecules, for a total of 49289 atoms. The truncated Aa_Cel9A contained 448 amino acid residues, 2 Ca^{2+} ions, 1 Zn^{2+} ion, 15 Na^{+} ions and 11156 water molecules, for a total of 40267 atoms. The charges of all the chargeable residues were set to their ionizable states at pH 7. Each system was energy minimized for 1000 steps and then equilibrated for 200 ps, in the NVT ensemble, over which the temperature was increased gradually up to 300 K. This was followed by another 500 ps of unconstrained dynamics using a Nose–Hoover constant pressure (P = 1 bar) and temperature (NPT) simulations were carried. The time step is chosen to be 2 fs. After 1 ns equilibrium state, 10 ns simulations were carried out in production runs and data are collected for analysis. The coordinates were saved every 500 steps (1 ps). In total, 10000 frames were saved for further analysis.

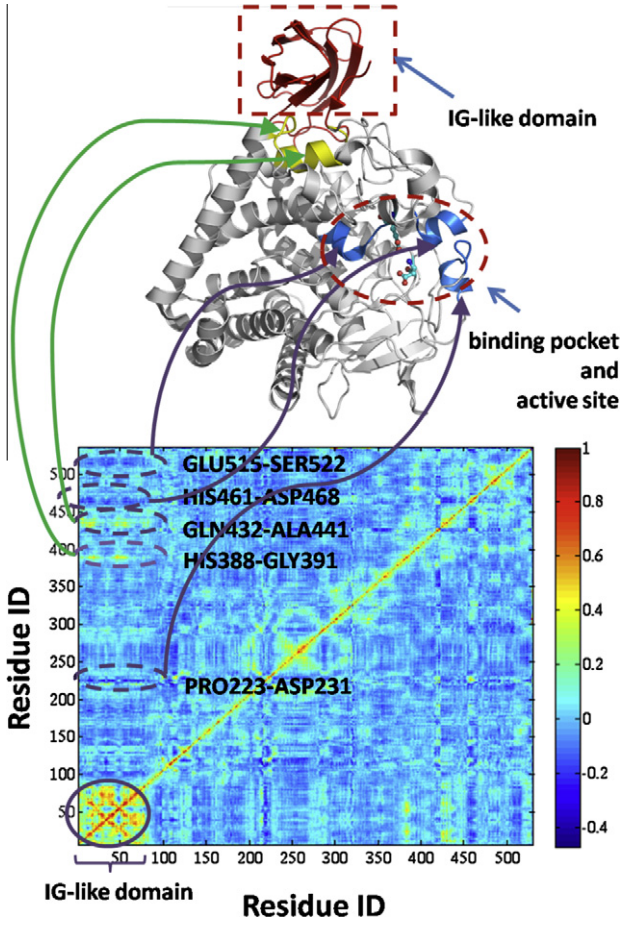


Fig. 1. Crystal structure of Aa Cel9A (top, generated by using Pymol [21] program) and map of residue correlations of wt-Cel9A based on molecular dynamics simulations (bottom).

The PTRAJ Module of AMBER 9 package was used to analyze our MD simulation results. For each trajectory, the amino acid residues in every frame of the MD trajectory were aligned to the heavy atom positions of their crystal structure to remove the overall translation and rotation. Later, the Root mean-square deviations (RMSD) of main chain atoms (Ca, C, N) in the catalytic domain (residue 81-to in full protein) were calculated.

The residue cross-correlation maps for WT Aa_Cel9A and the catalytic domain of Aa_Cel9A were also calculated using the PTRAJ module in AMBERTOOLS. The normalized fluctuation covariance matrix C_{ij} is defined as

$$C_{ij} = \frac{\langle \Delta r_1 \cdot \Delta r_j \rangle}{\left(\langle \Delta r_i \rangle^2 \langle \Delta r_i \rangle^2 \right)^{1/2}} \quad (1)$$

where i and j are any two residues, Δr_i and Δr_j are displacement vectors of i and j. If C_{ij} = 1 the fluctuations of i and j are completely correlated (same period and same phase), if C_{ij} = −1 the fluctuations of i and j are completely anti-correlated (same period and opposite phase) and if C_{ij} = 0 the fluctuations of i and j are not correlated.

Energy landscape theory provides a framework for the description of the kinetic and thermodynamic mechanisms of protein folding [12,13]. Because of the importance of adequately sampling the potential energy surface, coarse grained models (also called simplified models) have been used in simulations of protein folding [14–17]. Here a coarse grained force field within the Molaris [18,19] software package was used to explore the free-energy landscape of the WT Aa_Cel9A and Ig truncated mutant. A brief description of the refined model system follows. In the refined model, the potential energy surface of the simplified model is written as

$$U_{simplified} = U_{main} + U_{main-side} + U_{side-side} + U^{self}_{solvation} \quad (2)$$

where U_{main} describes the potential energy for the main chain, which is a standard part of the MOLARIS software package. $U_{side-side}$ describes the interaction between the side chains and is based on an "8–6" potential. The $U_{main-side}$ term describes the interaction between the effective side chains and the main-chain atoms, and $U^{self}_{solvation}$ accounts for the change in the solvation energy of each of these groups upon moving from water to its protein site.

The free energy surface was evaluated using the FEP/US method as in previous study [19] as a function of radius of gyrations (R_g) and sorting the results in two well-defined parameters, the R_g and RMSD of the catalytic domains. The starting points for the free-energy landscapes were taken as relaxed structures of the simplified model after 200 ps relaxation. A force constant of 100 kcal/mol/Å was applied to the R_g of the catalytic domain only and the catalytic domain was forced to unfold by increasing its R_g along 21 frames each of which was simulated for 60 ps at 300 K and with a step size of 1 fs. It is worth pointing out that in the two sets of simulations on both the WT and the truncated mutant, the force constants are only applied to the catalytic domain. Thus the comparison of the free energy of unfolding is between the catalytic domains. For the case of the WT the interactions with the Ig-like domain are present, and for the truncated mutant these interactions are absent.

3. Results

The role of molecular motions and in particular correlated molecular motions in Aa_Cel9A function was investigated by calculating the correlation coefficients among all pairs of residues in WT Aa_Cel9A across the MD simulation. The residue correlation map, along with the crystal structure of Aa_Cel9A, is shown in Fig. 1. The degree of correlation refers to the correlation between fluctuations of a pair of atoms around the average structure. The correlation values vary from −1 (anti-correlated) to 0 (completely uncorrelated) to 1 (fully correlated).

The residue cross-correlation map provides a view of considerable through-space correlations that are not obvious *prima facie* from analysis of a static crystal structure. In our study, the major focus is the long-range dynamic correlation between the Ig-like domain residues and the cel9A catalytic domain. From Fig. 1, there are several residues in catalytic domain that have strong positive correlation ($C_{ij} > 0.4$, also see the supporting Table S1 in Supplementary result) with Ig-like domain residues, including GLN432, ARG433, ALA435, ASP436, GLY440, ALA441, HIS388 PRO389, PHE390 and GLY391 (colored yellow in Aa_Cel9A 3D structure). This indicates that the Ig-like domain and those residues move in concert in WT Aa_Cel9A. This observation is consistent with previous studies of crystal structures, because these residues are part of interface between residues in the Ig-like domain and residues in the catalytic domain [8]. These results also indicate that this region is composed of a rigid protein–protein binding surface.

As seen in Fig. 1, several regions exhibited anti-correlated motions ($C_{ij} < -0.2$, also see the supporting Table S1). The amino acid residues in these regions are PRO223, LEU224, ASP225, THR226, ARG227, PRO228, GLU229, ASP230, ASP231, HIS461, HIS462, PRO464, ALA467, ASP468, GLU515, VAL516, ALA517, TRP520, ASN521 and SER522. Regions 223–231 and 461–468 represent two loops in the WT Aa_Cel9A structure. Structurally these two loops are very close to each other even though they are distant in amino acid sequence space. The observation of anti-correlated motions between these two loops and the Ig-like domain indicates that the global motion of Ig-like domain and catalytic domain are linked and may be required for binding specificity and activity. Alignment of the crystal structure of Cel9A with similar cellulases shows that the residues located at 223–231 and 461–468 are likely involved in the cellulose binding, either conferring specificity or promoting activity. It has also been speculated from protein crystallography data [20] that the Ig-like domain common to glycosyl transferases and galactose oxidase might be involved in carbohydrate binding in these enzymes. Our finding of correlated motions between the Ig-like domain and these putative cellulose-binding loops provides compelling evidence for the existence of concerted domain motions as a mechanism for controlling Aa_Cel9A activity.

Based on the crystal structure and earlier biochemical studies of homologs, it is proposed that GLU515 within Aa_Cel9A acts as a proton donor in the inverted hydrolysis reaction (Scheme 1) [8]. The proton on the GLU515 side chain initially forms an H-bond with the O of the glycosidic bond between two glucose monomers. The proton is then transferred to the O atom in a concerted transition state (Scheme 1, middle).

In this reaction, proper orientation of the GLU515 side chain is critical for enzyme function. To investigate the effect of the presence of the Ig-like domain on GLU515 conformations, we calculated its side chain χ_1 and χ_2 angles over the course of our last 4 ns MD trajectories (4000 trajectories) of both the WT and the truncated Aa_Cel9A mutant. The angular distributions are plotted in Fig. 2. The χ_1 and χ_2 angles in the crystal structure of WT Cel9A are −163° and 168°, respectively.

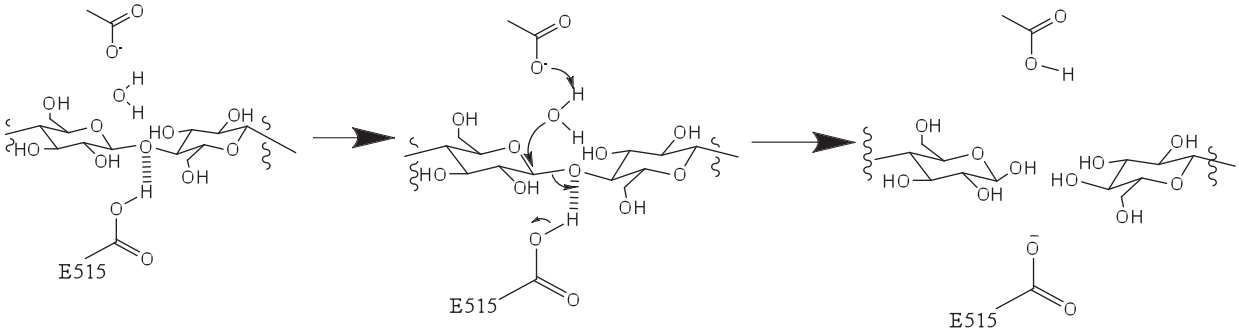
As seen in Fig. 2A, the χ_1 angle in WT Aa_Cel9A has a single peak in its distribution, with an angular distribution centered on −160°. However, in the truncated mutant, the χ_1 angle distribution shows a bimodal pattern with the peaks centered on −160° and −80°. As shown in Fig. 2B, the χ_2 angle for WT GLU515 has a broader range of available conformations and is more flexible in the case of the truncated mutant. MD simulations show that in the truncated mutant the catalytic residue GLU515 resides adopts conformations. Because the orientation of GLU515 is essential for catalytic activity, the increased conformational heterogeneity of the GLU515 side-chain in the mutant may in part be responsible for the loss of catalytic activity in the absence of the Ig-like domain.

It is also interesting to note that correlated motions also exist between the Ig-like domain and the region containing residues 515–522, an α-helical region in the active site. As discussed, residue GLU515 has intrinsic preferences for certain backbone dihedral values that place the side chain in an orientation that is required for catalysis. Our results showing correlated motions between the Ig-like domain and the 515–522 helix imply that these long-range correlated motions may affect catalysis by controlling the GLU515 side-chain orientations and may help explain why removal of the Ig-like domain reduces overall enzyme activity.

From the residue correlation maps, the Ig-like domain residues themselves are strongly positively correlated with each other. To determine the extent to which the high correlation values are overestimated due to the Ig-like "dangling" around the larger catalytic domain, an MD simulation was performed on the Ig-like domain only. While the correlation patterns are mainly retained, the correlations in the free Ig-like domain are smaller and thus were over estimated in the cross-correlation analysis of the full-length Cel9A.

Another possibility for the presence of the Ig-like domain at the N-terminus of the Aa_Cel9A is to stabilize the catalytic domain by limiting motions that lead to unfolding of the protein. To study whether the catalytic domain is stabilized by the presence of the Ig-like domain, we used a simplified folding model and Umbrella Sampling/free energy perturbation (UM/FEP) [19] to study the unfolding free energy landscapes of the Aa_Cel9A catalytic domain both with (WT) and without its Ig-like domain (mutant). The free energy of the protein configurations was explored as a function of two well defined parameters: the radius of gyration (R_g) and the root mean square deviation (RMSD). Fig. 3 plots the contours of the free energy required to move from one (RMSD, R_g) conformation to another and represents the free energy of unfolding. The RMSD is calculated only on the catalytic domain aligned to the crystal structure of the Cel9A catalytic domain.

Comparison of the two profiles also shows that for each (R_g RMSD) pair the free energy of unfolding is higher for the WT than for the mutant, indicating that more energy is required to increase the radius of gyration (unfold) of the WT Aa_Cel9A catalytic



Scheme 1. Hydrolytic reaction catalyzed by Aa_Cel9A. E515 is the proton donor in the hydrolysis reaction.

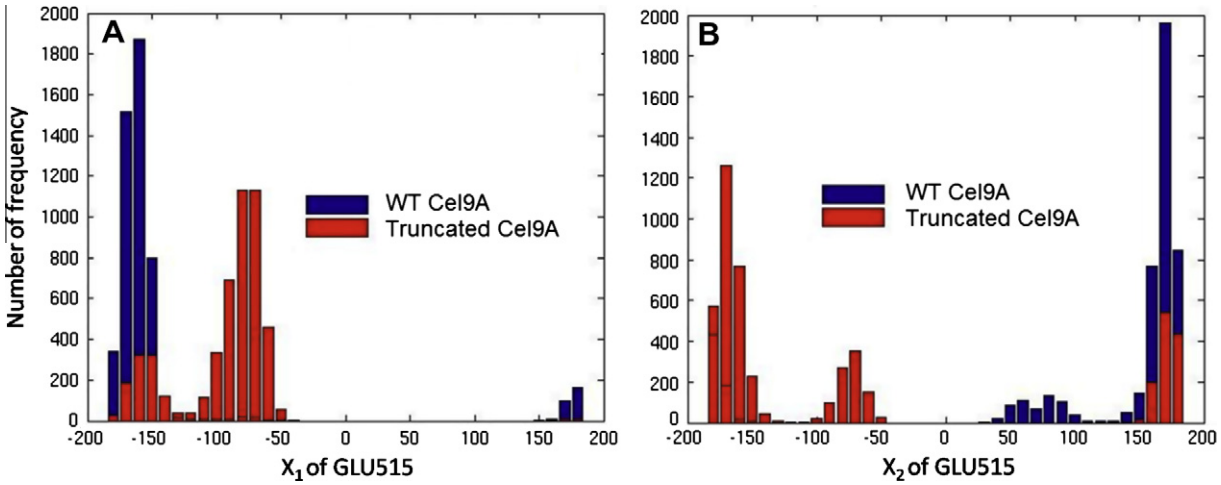


Fig. 2. GLU515 side chain χ angle distribution for the WT and mutant Aa_Cel9A without the Ig-like domain.

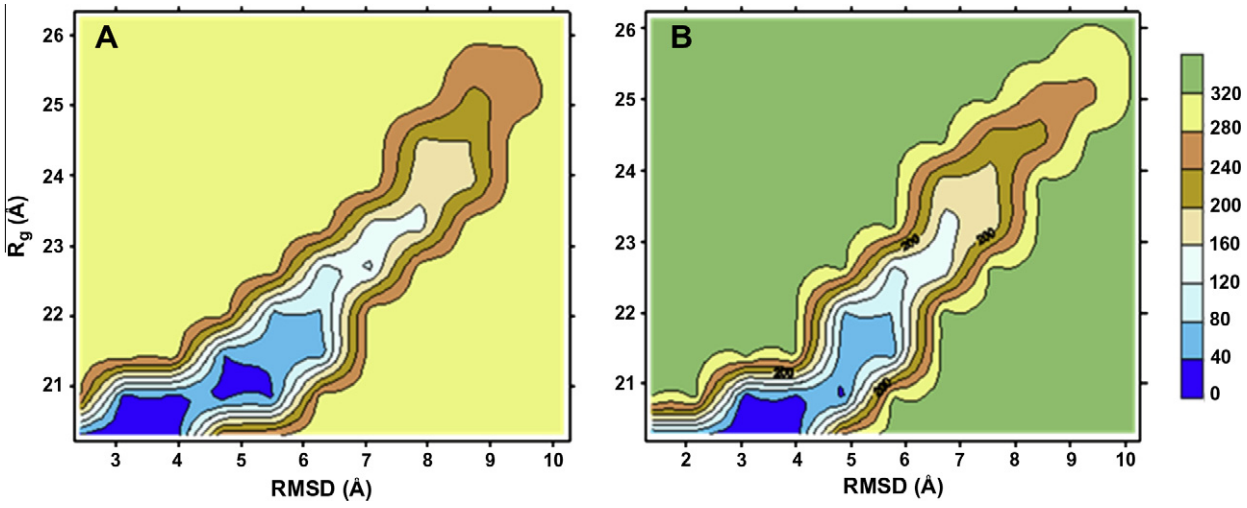


Fig. 3. Free energy landscapes for the truncated Aa Cel9A mutant without it's Ig-like domain (A) and the WT Aa_Cel9A (B). The units are in kcal/mol.

domain (~400 residues) with the Ig-like domain attached than it does to unfold the truncated Aa_Cel9A. The surfaces of both the WT and the truncated mutant have similar ranges of backbone RMSD in their folded structure, indicating that the unfolding free energy profiles are measures of the energy required to move among similar states, and are thus different due to the presence or absence of the Ig-like domain. The differences between the unfolding free energy landscapes suggest that the N-terminal Ig-like domain stabilizes the fold of the Aa_Cel9A catalytic domain, suggesting the Ig-like helps maintain the correct fold of the catalytic domain and may facilitate the thermostability of Aa_Cel9A.

4. Discussion

In summary, we have employed computational modeling to gain molecular level insights into the role of the N-terminus Ig-like

domain in the recently solved structure of Cel9A from the thermoacidophilic bacterium *A. acidocaldarius*. Experimental studies of truncated mutants [20] of a Cel9A homolog and crystal structures analysis of Aa_Cel9A [8] suggest that one function of the Ig-like domain might be to stabilize the catalytic domain. We applied two approaches (standard MD simulations and biased unfolding simulations) to model WT Aa_Cel9A and a mutant without the Ig-like domain. Our results indicate that the Ig-like domain (1) stabilizes the catalytic domain, playing an important role in keeping the catalytic domain folded in the correct conformation; (2) has strongly correlated internal motions, indicating that at higher temperatures the presence of the Ig-like domain enhances stability and enables the recovery of its properly folded conformation after perturbation, which may allow it to serve as a conduit for release of the energy and stresses caused by the hydrolysis reaction; (3) has correlated motions with several regions in the catalytic domain, including the key catalytic residue GLU515 and two loop regions (residues 223–231, and residues 461–468) that are responsible to cellulose binding.

This work suggests that one of the critical functions of the Ig-like domain is to ensure that the catalytic domain is properly folded, and to assists in maintaining the proper orientation of key active site and catalytic residues. The correlated motion between the Ig-like domain and the catalytic domain controls the motions of the two loops (PRO223 to ASP 231) and (HIS461 to ASP468), which are colored blue in Fig. 3, and are involved in carbohydrate binding. This may be a general feature of Ig-like domain in the GH family. These results may enable the engineering of cellulases that are more robust and stable under industrial processing conditions.

Acknowledgments

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. This research utilized the resources of the National Energy Research Scientific Computing Center (NERSC). We thank Drs. Maite Roca and Arieh Warshel for stimulating and informative discussion.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.06.041.

References

[1] Simmons, B.A., Loque, D. and Blanch, H.W. (2008) Next-generation biomass feedstocks for biofuel production. Genome Biol. 9, 242.1–242.6.
[2] Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. and Foust, T.D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315, 804–807.
[3] Tomme, P. et al. (1998) Characterization and affinity applications of cellulose-binding domains. J. Chromatogr. B Biomed. Sci. Appl. 715, 283–296.
[4] Juy, M., Amrt, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Beguin, P. and Aubert, J.-P. (1992) Three-dimensional structure of a thermostable bacterial cellulase. Nature 357, 89–91.
[5] Kataeva, I.A., Seidel 3rd, R.D., Shah, A., West, L.T., Li, X.L. and Ljungdahl, L.G. (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. Appl. Environ. Microbiol. 68, 4292–4300.
[6] Eckert, K., Ernst, H.A., Schneider, E., Larsen, S. and Lo Leggio, L. (2003) Crystallization and preliminary X-ray analysis of *Alicyclobacillus acidocaldarius* endoglucanase CelA. Acta Crystallogr. D Biol. Crystallogr. 59, 139–141.
[7] Kataeva, I.A., Uversky, V.N., Brewer, J.M., Schubot, F., Rose, J.P., Wang, B.C. and Ljungdahl, L.G. (2004) Interactions between immunoglobulin-like and catalytic modules in *Clostridium thermocellum* cellulosomal cellobiohydrolase CbhA. Protein Eng. Des. Sel. 17, 759–769.
[8] Pereira, J.H., Sapra, R., Volponi, J.V., Kozina, C.L., Simmons, B. and Adams, P.D. (2009) Structure of endoglucanase Cel9A from the thermoacidophilic *Alicyclobacillus acidocaldarius*. Acta Crystallogr. D 65, 744–750.
[9] Case, D.A., Darden, T.A., Cheatham, T.E. IIIIII, Simmerling, C.L. and Wang, J., Eds., (2006). AMBER, 9th ed, University of California, San Francisco.
[10] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.
[11] Darden, T., York, D. and Pedersen, L. (1993) Particle Mesh Ewald – an N.Log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092.
[12] Onuchic, J.N., LutheySchulten, Z. and Wolynes, P.G. (1997) Theory of protein folding: the energy landscape perspective. Annu. Rev. Phys. Chem. 48, 545–600.
[13] Shea, J.E., Onuchic, J.N. and Brooks, C.L. (2002) Probing the folding free energy landscape of the src-SH3 protein domain. Proc. Natl. Acad. Sci. USA 99, 16064–16068.
[14] Levitt, M. and Warshel, A. (1975) Computer-simulation of protein folding. Nature 253, 694–698.
[15] Bryngelson, J.D. and Wolynes, P.G. (1987) Spin-glasses and the statistical-mechanics of protein folding. Proc. Natl. Acad. Sci. USA 84, 7524–7528.
[16] Dill, K.A. (1990) Dominant forces in protein folding. Biochemistry 29, 7133–7155.
[17] Shakhnovich, E., Abkevich, V. and Ptitsyn, O. (1996) Conserved residues and the mechanism of protein folding. Nature 379, 96–98.
[18] Lee, F.S., Chu, Z.T. and Warshel, A. (1993) Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the polaris and enzymix programs. J. Comput. Chem. 14, 161–185.
[19] Roca, M., Liu, H., Messer, B. and Warshel, A. (2007) On the relationship between thermal stability and catalytic power of enzymes. Biochemistry 46, 15076–15088.
[20] Bork, P., Holm, L. and Sander, C. (1994) The immunoglobulin fold – structural classification, sequence patterns and common core. J. Mol. Biol. 242, 309–320.
[21] DeLano, W.L. (Eds.) (2002). The PyMOL molecular Graphic System. in: The PyMOL Molecular Graphic System, pp. on World Wide Web http://www.pymol.org.