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.

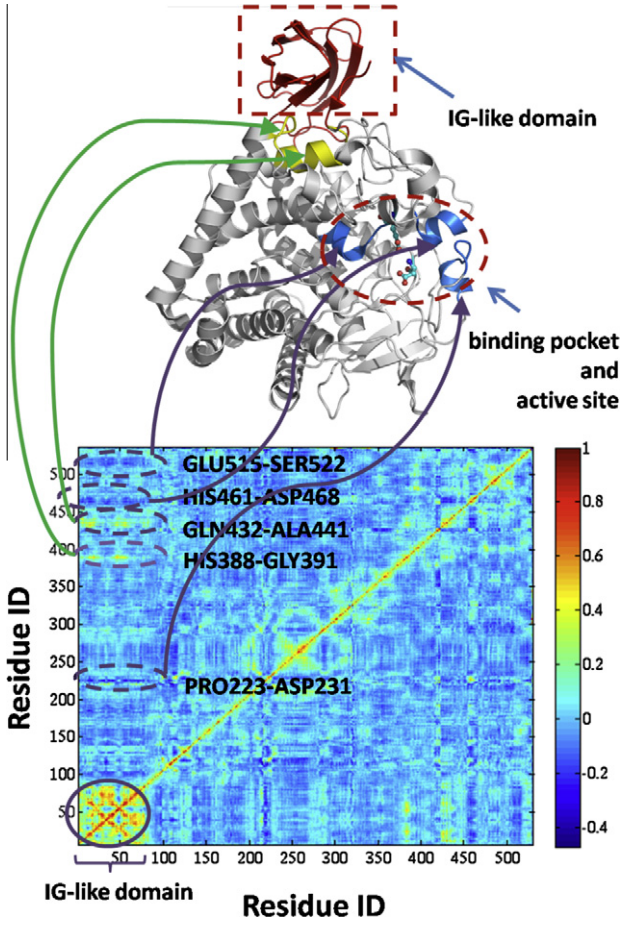

Fig. 1. Crystal structure of Aa Cel9A (top, generated by using Pymol [21] program) and map of residue correlations of wt-Cel9A based on molecular dynamics simulations (bottom).

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.