Lawrence Berkeley National Laboratory
Lawrence Berkeley National Laboratory

Title
Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance

Permalink
https://escholarship.org/uc/item/5mx5597j

Author
Zhang, Tao

Publication Date
2011-06-30

eScholarship.org Powered by the California Digital Library
University of California

Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance

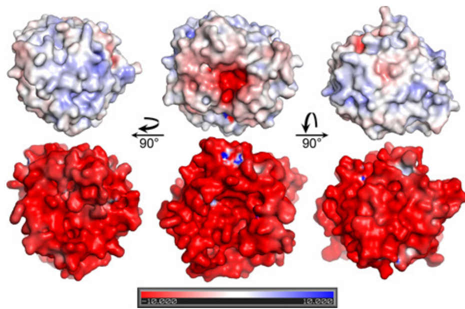
Tao Zhang[1], Supratim Datta[2], Jerry Eichler[3], Natalia Ivanova[1], Seth D. Axen[1], Cheryl A. Kerfeld[1, 4], Feng Chen[1], Nikos Kyrpides[1], Philip Hugenholtz[1,2,5], Jan-Fang Cheng[1], Kenneth L. Sale[2], Blake Simmons[2*] and Eddy Rubin[1*]

(1) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598; (2) The Joint BioEnergy Institute, Emeryville, CA 94608; (3) Dept. of Life Sciences, Ben Gurion University of the Negev, P.O. Box 653, Beersheva 84105, Israel; (4) Department of Plant and Microbial Biology, UC Berkeley, CA 94704; (5) Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia..

*Correspondences should be addressed to Eddy Rubin (EMRubin@lbl.gov) and Blake Simmons (basimmo@sandia.gov).

Keywords: halophilic cellulase, ionic liquids, thermostable, *Halorhabdus utahensis*.

Table of contents entry



Hu-CBH1, a haloalkaliphilic and thermostable cellulase, likely contains negatively charged acidic amino acids on the surface and is resistant to ionic liquids.

Abstract

Some ionic liquids (ILs) have been shown to be very effective solvents for biomass pretreatment. It is known that some ILs can have a strong inhibitory effect on fungal cellulases, making the digestion of cellulose inefficient in the presence of ILs. The identification of IL-tolerant enzymes that could be produced as a cellulase cocktail would reduce the costs and water use requirements of the IL pretreatment process. Due to their adaptation to high salinity environments, halophilic enzymes are hypothesized to be good candidates for screening and identifying IL-resistant cellulases. Using a genome-based approach, we have identified and characterized a halophilic cellulase (*Hu*-CBH1) from the halophilic archaeon, *Halorhabdus utahensis*. *Hu*-CBH1 is present in a gene cluster containing multiple putative cellulolytic enzymes. Sequence and theoretical structure analysis indicate that *Hu*-CBH1 is highly enriched with negatively charged acidic amino acids on the surface, which may form a solvation shell that may stabilize the enzyme, through interaction with salt ions and/or water molecules. *Hu*-CBH1 is a heat tolerant haloalkaliphilic cellulase and is active in salt concentrations up to 5 M NaCl. In high salt buffer, *Hu*-CBH1 can tolerate alkali (pH 11.5) conditions and, more importantly, is tolerant to high levels (20% w/w) of ILs, including 1-allyl-3-methylimidazolium chloride ([Amim]Cl). Interestingly, the tolerances to heat, alkali and ILs are found to be salt-dependent, suggesting that the enzyme is stabilized by the presence of salt. Our results indicate that halophilic enzymes are good candidates for the screening of IL-tolerant cellulolytic enzymes.

Introduction

Plant cell walls are composed of crystalline cellulose entangled with hemi-cellulose and lignin, forming a complex matrix rendering plant biomass largely inaccessible to cellulolytic enzymes in the native state. Current methods of lignocellulosic biofuel production typically involve disrupting plant cell walls using high temperatures and/or corrosive chemicals to liberate the polysaccharides and generate a product that is more accessible to hydrolytic saccharification. These pretreatments are costly, inefficient and, in certain cases, are environmentally toxic. It is, therefore, necessary to improve pretreatment methods.

Ionic liquids (ILs) represents a promising solution to the problem of recalcitrant biomass. ILs are nonvolatile salts, typically with melting points under 100 °C, and some ILs can efficiently solubilize cellulose, hemicellulose and lignin from plant biomass under moderate temperatures [1-6]. The regeneration of cellulose from ILs can be achieved by adding an anti-solvent, such as water or ethanol, into the solution [7-9]. ILs can be recycled for new rounds of pretreatment. It has been shown that regenerated cellulose after IL pretreatment has reduced crystallinity, and is thus easier for cellulolytic enzymes to access [10-11].

Several improvements are needed in the ionic liquid pretreatment process technology before it is cost effective with other pretreatments that are based on the pulp and paper processing technologies that utilize dilute acids and bases. One of the most important areas for cost reduction is reducing the number of washes required after IL pretreatment. Unfortunately, commercial fungal cellulases are inhibited by some ILs [8,12-13] and, therefore, require extensive washing after IL pretreatment. Therefore, it is crucial

to identify IL-resistant enzymes for digesting cellulose in the presence of ionic liquids to decrease the number of washes required and increase the yields of monomeric sugars.

It has been suggested that ILs inhibit enzymatic activity by disrupting hydrogen bonding and hydrophobic interactions and depriving the water hydration shell of the protein [14-18]. This is similar to the denaturing effect caused by salt on mesophilic proteins. Although it's not clear if salt and ionic liquids denature proteins in identical ways, both create an environment characterized by low-water activity and high ionic strength [14,19]. Microbes living in extremely high salt environments can possess a cytoplasm containing > 3 M salt. Accordingly, such organsims have evolved a unique mechanism to compete with salt for water. In high salt concentrations, proteins contain an excessive number of negatively charged acidic amino acids on their surface, while at the same time having only few basic amino acids and a low hydrophobic amino acid content [20-23]. Among these, the negative charges are the most prominent feature [24-27] and are thought to keep the protein soluble in a high salt solution either by forming a hydrated ion network with cations or by preventing the formation of protein aggregation through electrostatic repulsive charges at the protein surface [25,28-31]. Theoretically, positive charges on protein surface may have similar effect on protein stability in high salt environments as negative charges. Yet, in nature, majority of the halophilic proteins are enriched with acidic amino acids on the protein surface, suggesting that negatively charged proteins are under positive selection in halophilic microorganisms. In addition, reduced surface area is also important for the protein to remain folded and require less water to form a hydration shell. All of these salt adaptation strategies could be used for enzymes to resist ILs.

We have identified a gene cluster that contains multiple cellulolytic enzymes from the halophilic archaeon, *Halorhabdus utahensis*. We cloned and expressed one cellobiohydrolase in a different haloarchaeal host, *Haloferax volcanii*. We named this gene as *Halorhabdus utahensis* CBH1, or *Hu*-CBH1, in short. Using cellulase activity assay, we found that this enzyme is a haloalkaliphilic and heat tolerant cellobiohydrolase. The protein is enriched in acidic amino acids and presents strong negative charges on its surface. Interestingly, we determined that salt is essential for the stability and function of the protein and that it can tolerate up to 20% (w/w) of ILs, including 1-ethyl-3-methylimidazolium acetate ([Emim]Ac), 1-ethyl-3-methylimidazolium chloride ([Emim]Cl), 1-butyl-3-methylimidazolium chloride ([Bmim]Cl) and 1-allyl-3-methylimidazolium chloride ([Amim]Cl).

Results

Halorhabdus utahensis contains multiple cellulase genes

Halorhabdus utahensis is an obligatory halophilic archaeon that requires 27% NaCl for optimal growth [32]. Sequence analysis of its genome revealed that 44 putative glycosyl hydrolases (GH), including cellulases, are encoded (Iain Anderson, et al., manuscript in preparation)[33]. Given the native hypersaline environment of *Hrd. utahensis*, these GHs may include strong candidates for salt-tolerant cellulases. The organism could be cultured on xylan, but not cellulose as substrate, raising the question of whether these cellulase genes are indeed functional [34-35] and/or perhaps incorrectly classified as cellulases.

Seven of the putative cellulase genes are found in a 37 kb genomic locus together with sequences predicted to encode two xylanases, a mannase, a pectinase, a sugar-specific transcription regulator, and three uncharacterized proteins (Table 1). These genes are organized in head to tail orientation (Figure 1), similar to genes in bacterial operons. Except for the transcription regulator, these genes share three conserved domains, namely a catalytic domain, a fibronectin domain 3 (FN-3) and an Ig-like domain (Supplementary Figure 1). These domains are present in both cellulase and non-cellulase genes in the gene cluster. The catalytic domain is less conserved, suggesting that these enzymes serve diverse functions. However, the FN-3 and Ig-like domains are more conserved than the catalytic domain (Supplementary Figure 2), In bacteria, FN-3 and Ig-like domains have often been found in glycosyl hydrolase enzymes, suggesting that members in the gene cluster may be involved in carbohydrate metabolic pathways [36-39]. The existence of conserved domains among these genes suggests that the gene cluster was created by gene duplications from one parental gene, and that the functional differences of these genes were acquired later, primarily through mutation of the catalytic domain.

Enrichment of acidic amino acids is a common feature of halophilic proteins. Analysis of the amino acid sequences revealed that proteins from the cellulolytic enzyme gene cluster are enriched with acidic amino acids (Supplementary Figure 3). This indicates that these genes have adapted to the high salt environment and are not a result of any recent lateral gene transfer of cellulolytic enzymes from mesophilic organisms. The calculated pI of these proteins is around 4, indicating that these proteins are negatively charged under physiological conditions. To study the function of these genes, we selected *Hu*-CBH1 (Huta-2387) for our initial work.

The Hu-CBH1 is a secreted cellobiohydrolase

Structure prediction of *Hu*-CBH1 indicates that the negatively charged amino acids are predominantly present on the protein surface (Figure 2, bottom). On the contrary, the distribution of acidic amino acids in a mesophilic cellulase homologue from *Erwinia chrysanthemi* is restricted to the catalytic pocket region (Figure 2, top). *Hu*-CBH1 bearing a C-terminal 6 His-tag was cloned and expressed in *Haloferax volcanii*, a culturable moderate halophile that is widely use for heterogeneous expression of halophilic proteins. The cells were cultured in a medium containing 20% NaCl, which is required for proper folding of halophilic proteins. The N-terminal region of the *Hu*-CBH1 protein contains the double arginine-based signature of a TAT secretion pathway signal peptide [40] (Supplementary Figure 2). The 709 amino acid sequence predicts a protein of 76 kDa, which was purified from culture medium by binding of the protein to Ni-NTA beads. The secreted protein migrates as a single band of ~90 kDa (Figure 3. A). The identity of the protein was confirmed by mass spectrometry analysis (see *Experimental*). The apparent slow migration of the protein in SDS PAGE gel was likely caused by the excessive amount of acidic amino acids in the protein[41]. Purified *Hu*-CBH1 showed comparable enzyme activity to fungal cellulase from *T. reesei*, suggesting that *Hu*-CBH1 is an active enzyme in this organism (Figure 3. B).

Enzyme activity on carboxymethyl cellulose (CMC) substrate was also detected in both cell lysate and culture medium. About 39% of the crude enzyme activity was present in culture medium, while 59% of the crude enzyme activity was present in the cell lysate (data not shown). In the cell lysate, the active enzyme was present exclusively

in the soluble fraction (Supplementary Figure 4). Therefore, the *Hu*-CBH1 is a soluble protein when expressed in the non-native halophilic host. We have tested the substrate specificity of the enzyme and found that it was only reactive to p-nitrophenyl-beta-D-cellobioside and CMC, suggesting that it is a cellobiohydrolase (Supplementary Table 1).

Hu-CBH1 is a salt-dependant thermal tolerant cellulase.

Many halophilic enzymes require minimum 2 M salt for proper function. However, *Hu*-CBH1 showed a different level of salt-dependence at different temperatures. At 37 ^{o}C, the optimal salt concentration of *Hu*-CBH1 is 0.25 M NaCl (Figure 4. A). However, at 80 ^{o}C, the optimal salt concentration is 5 M NaCl (Figure 4. B). In low salt buffer, although the enzyme is active at low temperature, its activity was reduced with increasing temperature. This suggests that the enzyme is not stable under low salt conditions. By contrast, in a high salt buffer, activity was stimulated by elevation of temperature but dropped quickly beyond the optimal temperature. The extreme thermal stability is closely correlated with salt concentration, suggesting that salt may help to reinforce the protein structure, probably through forming a salt-ion hydration shell to "lock" the protein in a properly folded structure.

Hu-CBH1 is an alkaliphilic cellulase

Since the surface of *Hu*-CBH1 is negatively charged at physiological pH, we further tested the effect of pH on enzyme activity. In 2M NaCl, the optimal pH of the enzyme was at 9.5, but could tolerate a pH up to 11.5 (Figure 5). However, the enzyme was very sensitive to low pH, and was completely inactive at pH 4.5. The observed

enzyme activity at high pH suggests that maintaining the negative charges on the protein surface is important for enzyme function. Interestingly, some alkaliphilic cellulase proteins are also enriched in acidic amino acids (Supplementary Figure 5). Therefore, negatively charged amino acids are selected for proteins in both halophilic and alkaliphilic organisms.

Hu-CBH1 is resistant to ionic liquids

Given that *Hu*-CBH1 is stabilized by salt and is tolerant of high temperatures, we further tested its ability to resist [Emim]Ac, [Emim]Cl, [Bmim]Cl and [Amim]Cl, common choices of ILs for cellulose solvation [3-4,42-45]. In 20% of [Emim]Ac or [Emim]Cl, *T. reesei* cellulase was partially inhibited (Figure 6. A). However, in 20% of [Bmim]Cl or [Amim]Cl, *T. reesei* cellulase was almost completely inhibited (Figure 6. A). However, *Hu*-CBH1 activity was unchanged and, sometimes, slightly stimulated in present of 20% of one of these ILs. Therefore, the resistance to ILs is a general character of *Hu*-CBH1. It should be noted that the activity of *Hu*-CBH1 in solutions of ILs was also observed to be dependent on salt. In 0.25 M NaCl, *Hu*-CBH1 was completely inhibited by 20% [Amim]Cl (Figure 6. B), suggesting that the enzyme was not stable in low salt buffer. Interestingly, in high salt buffer, the enzyme activity was not inhibited, but rather stimulated by the presence of [Amim]Cl. This may indicate that [Amim]Cl and NaCl have different effects in terms of protein folding and activity. The reason for which is not clear and requires further investigation. We have tested the upper limit of [Amim]Cl that *Hu*-CBH1 could tolerate in the presence of 2 M salt. However, further increasing [Amim]Cl concentration in 2M salt buffer caused precipitation of NaCl and hence

sharply reduced enzyme activity. In 30% [Amim]Cl, only ~60% of *Hu*-CBH1 activity was observed. In 40% of [Amim]Cl, the enzyme was completely inactivated. As salt is important for enzyme stability, reducing salt in the reaction might have destabilized the enzyme and reduced its activity in the presence of higher concentrations of ILs.

Discussion

Certain microbes are known to thrive in hypersaline environments. Culturing methods have been used to identify cellulose degrading salt tolerant microbes from salt mines and salt lakes [46-48]. However, none of the cellulase genes were identified. Indeed, detailed characterization of these genes and their products is extremely difficult. Recently, salt-tolerant cellulases have been found by functional screening of recombinant genomic libraries [49-50]. The primary limitation of these studies is that halophilic enzymes may not fold properly when expressed in *E. coli*, and as such are not detectable by functional screening [51-52]. Our study represents the first known case of using a sequence homology-oriented approach for the prediction of halophilic cellulases, and the expression of these salt-tolerant genes in a halophilic host. This strategy enhances the possibility of identifying true obligatory halophilic enzymes.

Like other halophilic proteins, expression products of the cellulolytic gene cluster in *Hrd. utahensis* are all enriched with acidic amino acids, suggesting that these enzymes are specifically adapted to tolerate and function in the high salt environment. Although their homologies to known collections of fungal glycosylhydrolases are, as expected, low, the operon-like structure of the gene cluster increases the confidence of gene prediction. In addition, the presence of a fibronectin 3 domain in these genes further indicates that

they may indeed be involved in lignocellulose metabolism. In bacteria, the FN3 domain is only found in extracellular glycosylhydrolase proteins [36]. It has been shown that this domain is not directly involved in cellulose hydrolysis but rather enhances disruption of the surface of crystalline cellulose and enhances saccharification efficiency [37,53]. The function of the conserved Ig-like domain is unknown, but it may be important for salt-tolerant enzyme function [38-39]. More studies are required for elucidating its function. It remains to be determined whether some of these proteins may form a protein complex for cellulose degradation, similar to cellulosome [54].

Given the importance of the salt hydration shell on protein stability, it is natural that the protein needs to maintain its negative charges in order to resist the denaturing environment created by high salt concentrations. Interestingly, this strategy has also been observed in alkaliphilic proteins that are also enriched with acidic amino acid residues [55-57] and exhibit improved salt-tolerance and salt-dependent thermal stability. This suggests that converged evolution has made these proteins both salt- and alkali- tolerant and as such, could be strong candidates for the discovery of enzymes that can tolerate high concentration of ILs. A recent study demonstrated that hyperthermophilic cellulases are also resistant to ILs, suggesting that there are different mechanisms for extremophiles to maintain their stability in extreme environments [42].

A halophilic host provides an ideal environment for halophilic protein folding. However, growing halophiles takes a long time and requires high levels of salt in the medium. These are technical issues to be overcome before it is possible to produce halophilic enzymes on the large scale. We have expressed *Hu*-CBH1 in *E. coli*. The recombinant protein formed insoluble aggregates (data not shown). This may be due to

improper folding of protein in the low salt growth conditions required for *E. coli*. It has been reported that aggregated proteins can be rescued by denaturing the protein in 6 M guanidine hydrochloride or urea followed by renaturing the protein through dilution into salt solutions [58]. It would be interesting to see if such approach can be used to refold *Hu*-CBH1 proteins produced by *E. coli*. Nevertheless, understanding the mechanism of salt-tolerance may help to engineer novel cellulolytic enzymes adapted to ionic liquid environments.

Experimental

Cloning and expression of Hu-CBH1 in a halophilic host.

The *Hu*-CBH1 coding region was amplified by PCR, using genomic DNA of *Halorhabdus utahensis* (DSM 12940) as template. The coding sequences of six histidines followed by a stop codon were introduced in-frame to the 3'-end of the gene through one of the PCR primers. The PCR products were cloned into the *Haloferax volcanii* pJAM202 shuttle vector [59]. The plasmids were first transformed and propagated in *Escherichia coli* cells. The fidelity of the cloned DNA was validated by Sanger sequencing of the plasmids. Later, Hu-CBH1-containing plasmids were transformed into *Hfx. volcanii* strain WR-340 cells, following the procedure described previously [60]. Positive clones were screened by PCR. The cells were then cultured in liquid halophile culture medium, containing NaCl (206 g), $MgSO_4*7H_2O$ (37 g), $MnCl_2$ (75 mg/liter) (1.7 ml), 1M Tris-HCl (pH 7.2) (50 ml), KCl (3.7 g), yeast extract (3 g), tryptone (5 g) and $CaCl_2*2H_2O$ (10%) (5 ml) per liter of liquid medium, supplemented with 0.5 μg/ml of

novobiocin (final concentration). Cells were cultured at 40 °C with shaking for 6 to 10 days till the cells were confluent.

Recombinant protein purification.

Haloferax volcanii cells from 200 ml culture were collected by centrifugation. The supernatant was saved for purifying secreted proteins. The cell pellets were resuspended in 15 ml extraction buffer, containing 2 M NaCl and 10 mM Tris-HCl (pH 7). Cells were broken by acoustic homogenization using a Covaris S220 ultra-sonicator. Cell lysates were spun down by centrifugation at 12,000xg for 15 minutes. The supernatants were separated from cell debris and used for enzyme activity assays.

Hu-CBH1 proteins were purified from culture medium by incubation with Ni-NTA beads (Qiagen). Proteins were eluted with an elution buffer containing 250 mM imidazole, 2 M NaCl and 10 mM Tris-HCl (pH 7). Purified enzymes were quantified using Bradford Protein Assay Kit from Bio-Rad (Cat#500-0001). The yield of purified proteins was 35 μg per 200 ml of culture medium supernatant. The purity and molecular weight of the protein were assessed by electrophoresis in a 7.5% SDS polyacrylamide gel and Coomassie staining.

To confirm the identity of the purified protein, the 90kDa band was excised from the gel and subject to trypsin digestion and mass spectrometry analysis. A total of 10 unique peptides identical to the *Hu*-CBH1 protein sequence were identified, namely 1) DGNLIKDPDGNTVTLR, 2) GVNIADPK, 3) RINETAQAR, 4) GMTATQVIDMLTDESNGWYPR, 5) GVYCIIDYHR, 6)

DVQWAEGQDGPVNTELQDEVDMFWDTVAPR, 7) PVMFSR, 8) SALEQYR, 9) GANGGEEDEFIFDMGGAR, and 10) LNFWQGGSSTLEIEEIR.

Enzyme activity assay

Hu-CBH1 enzymes (2 μg) purified from culture medium or *Trichoderma reesei* cellulase (Sigma, #C8546) (2 μg) were incubated with 50 μl of 0.5% soluble carboxymethyl cellulose (CMC) in 200 μl reaction buffer, containing 2 M NaCl and 10 mM Tris-HCl (pH 7.0), at 37 °C for 30 minutes. The specific activities of enzymes were determined by measuring the amount of glucose released through enzyme digestion using the DNS assay [61]. The enzyme activity was calculated as μmols of glucose released by 1 mg of enzyme per minute.

Lysate supernatants of *Hu*-CBH1 expressing cells were used to determine enzyme activities in response to changes of temperature, pH and different concentrations of salt and ionic liquids. Briefly, lysate supernatants (50 μl) were mixed with 0.5% soluble carboxymethyl cellulose (CMC) (50 μl) and appropriate buffers (100 μl). Salt and ILs concentrations and pH were adjusted as described in the figure legends. The reactions were incubated for 1 hour, at different temperatures as also described in the figure legends. The amounts of glucose released were determined by the DNS assay. Enzyme activity assays were performed in triplicate for each experimental condition tested. Averaged readings were used to calculate the glucose released by the enzymes. The mock reaction mixtures (without incubation), containing lysate supernatants (50 μl), 0.5% CMC substrates (50 μl) and buffers (100 μl), were used as background controls for each experimental condition tested. The amount of ’glucose’ present in the mock reaction

mixture was determined by the DNS assay. Background control readings of glucose were subtracted from the enzyme reaction readings to determine the actual amount of glucose released from enzyme hydrolysis reactions. Lysate supernatants of *Haloferax volcanii* cells transformed with control vector showed no cellulase activity.

1-Ethyl-3-methylimidazolium acetate ([Emim]Ac) (#689483), 1-Ethyl-3-methylimidazolium chloride ([Emim]Cl) (#30764), 1-butyl-3-methylimidazolium chloride ([Bmim]Cl) (#94128) , 1-Allyl-3-methylimidazolium chloride ([Amim]Cl) (#43961) were purchased from Sigma and used as received.

Structure of Hu-CBH1

The theoretical structure of the cellulase domain of *Hu*-CBH1 from *Halorhabdus utahensis* was modeled using the SWISS-MODEL [62-64] server, an automated homology modeling program. The *Pseudoalteromonas haloplanktis* cellulase (PDB: 1TVN) was selected as the template using "Automated Mode". Residues 64 to 382 of *Hu*-CBH1 aligned to residues 1 to 267 of the template structure with 20.3% sequence identity.

Electrostatics of the model were generated by first removing heteroatoms from the PDB coordinate files for the homology model of the *Hu*-CBH1 cellulase domain and for the structure of the cellulase (PDB: 1EGZ) from *Erwinia chrysanthemi*. PQR files were generated from the PDB files with PARSE as the forcefield using the program PDB2PQR program [64]. The electrostatics were then calculated in PyMOL [65] using the APBS [66] plug-in with the PDB and PQR files as input. Surface electrostatics were visualized by coloring the PyMOL surface model [67] by the potential on the solvent accessible surface, using -10 $k_B T/e$ and +10 $k_B T/e$ as the Low and High surface potentials, respectively.

Conclusions

Our results showed that *Hu*-CBH1 is clearly a halophilic cellulase. It can function in high salt, it remains active at high temperature, and can tolerate up to 20% (w/w) [Amim]Cl. Structure prediction suggested that the surface of this protein is enriched in acidic amino acids. Our data supports the hypothesis that this negative surface charge may interact with the hydrated salt cations present in the solution to form a hydration shell. This makes the proteins resistant to heat, although neither *Hrd. utahensis* nor *Hfx. volcanii* is typically heat resistant. Indeed, we found that *Hu*-CBH1 is a salt-dependant enzyme and that its activity can be sustained and stimulated by salt at high temperatures. These results indicate that enzymes isolated from hypersaline environments are strong candidates for the development of IL-tolerant enzymes and cocktails capable of efficiently liberating monomeric sugars from IL-pretreated biomass. Furthermore, it has been recently reported that recovery of ILs from cellulose/ILs solution can be achieved by adding aqueous kosmotropic salt solutions, such K_3PO_4 [43,68]. This offers the possibility of integrating IL pretreatment and enzymatic saccharification using halophilic enzymes with minimal washing and high rates of ionic liquid recovery.

Acknowledgments

We thank Bonnie Baxter, Len Pennacchio, Rajat Sapra and Rex Malmstrom for helpful suggestions and discussions, and Yukun (Kelvin) Zhang for critical reading of the manuscript. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under

Contract No. DE-AC02-05CH11231. The work conducted by the U.S. Department of Energy Joint BioEnergy Institute is supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231. J.E. is supported by the Israel Science Foundation (grant 30/07).

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

FIGURES and TABLES

Table 1

The Gene IDs from 0 to 14 are assigned based on the position of the genes from beginning to the end of the gene cluster. All genes are in same orientation in the cluster. The predicted protein sequences were used to search NCBI NR protein database. The best hit of the BLAST search was assigned to each gene. Sequence similarity between the predicted genes to their best hits is below or close to 50%.

Locus Tag	Internal ID	BLAST hits
Huta_2386	0	Transcriptional regulator (sugar-specific)
Huta_2387	1	Cellulase (GH 5)
Huta_2388	2	Cellulase (GH 5)
Huta_2389	3	Hypothetical protein.
Huta_2390	4	Uncharacterized protein
Huta_2391	5	Beta-1,4-xylanase
Huta_2392	6	Beta-1,4-xylanase
Huta_2393	7	Cellulase (GH 5)
Huta_2394	8	Cellulase
Huta_2395	9	Cellulase
Huta_2396	10	Endo-beta-mannanase
Huta_2397	11	Pectate lyase
Huta_2398	12	Cellulase (GH 5)
Huta_2399	13	Cellulase (GH 9)
Huta_2400	14	Hypothetical protein

Figure 1. The *Halorhabdus utahensis* genome contains a single gene cluster encoding cellulolytic enzymes with conserved domains. The gene cluster contains 1 sugar specific transcription regulator (in red), 7 cellulase (in green), 2 xylanase (in blue), 1 mannanase (in yellow), 1 pectate lyase (in pink) and 3 proteins with unknown function (in white).

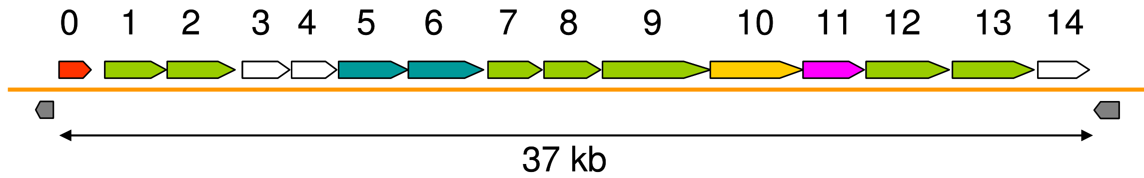


Figure 2. Acidic amino acids are highly enriched in halophilic proteins present in the gene cluster. The *Hu*-CBH1 (gene-1) protein surface is extensively covered by negatively charged amino acids. Electrostatics of the cellulase (neutral protein) from *Erwinia chrysanthemi* (PDB:1EGZ; top) and the homology model of the cellulase domain of *Hu*-CBH1 (acidic protein) from *Halorhabdus utahensis* (bottom).

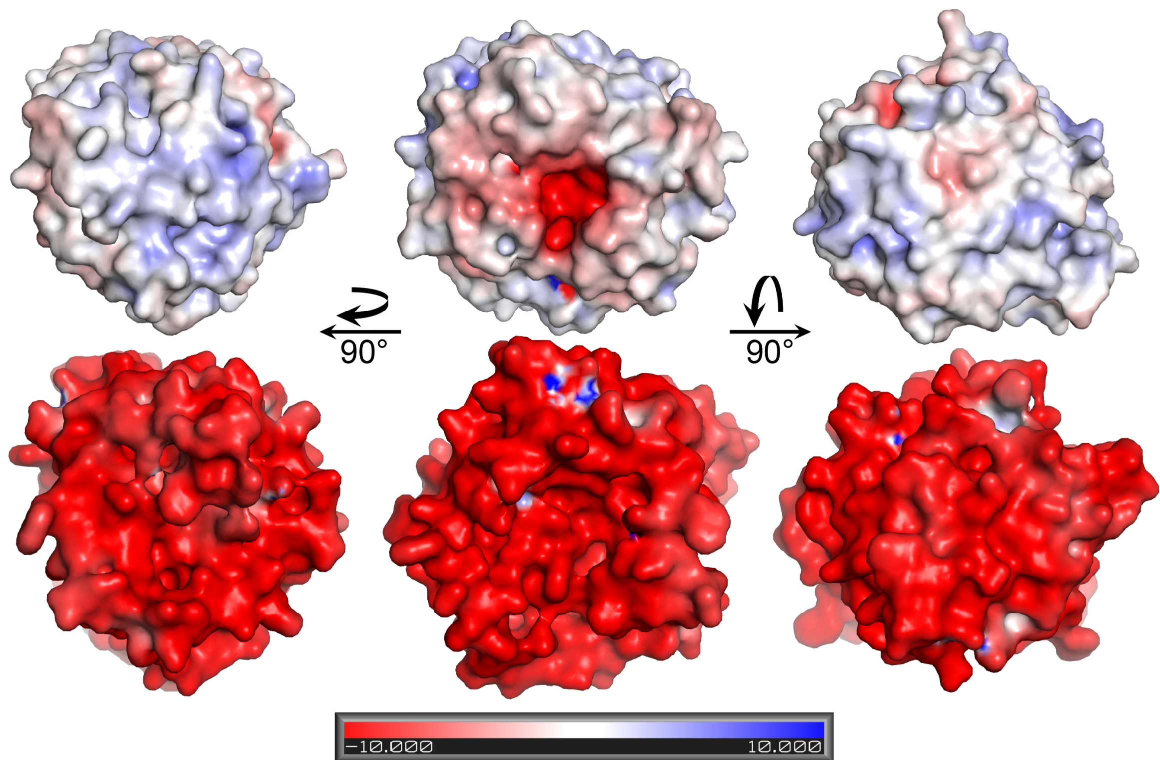


Figure 3. ***Hu*-CBH1 is a secreted protein with cellulase activity.**

A. *Hu*-CBH1 protein bearing a polyhistidine tag was purified from culture medium using Ni-NTA resin and analyzed by SDS-PAGE. The protein migrated as a ~90 kDa band in the gel. The size of protein ladder is in kDa. **B.** Purified recombinant *Hu*-CBH1 protein and *T. reesei* cellulase were incubated with carboxymethyl cellulose (CMC) in 2 M NaCl and 10 mM Tris-HCl (pH 7.0) at 37 °C for 30 minutes. The specific activity was measured by the DNS assay, and quantified as μmol of glucose produced per mg of enzyme per minute.

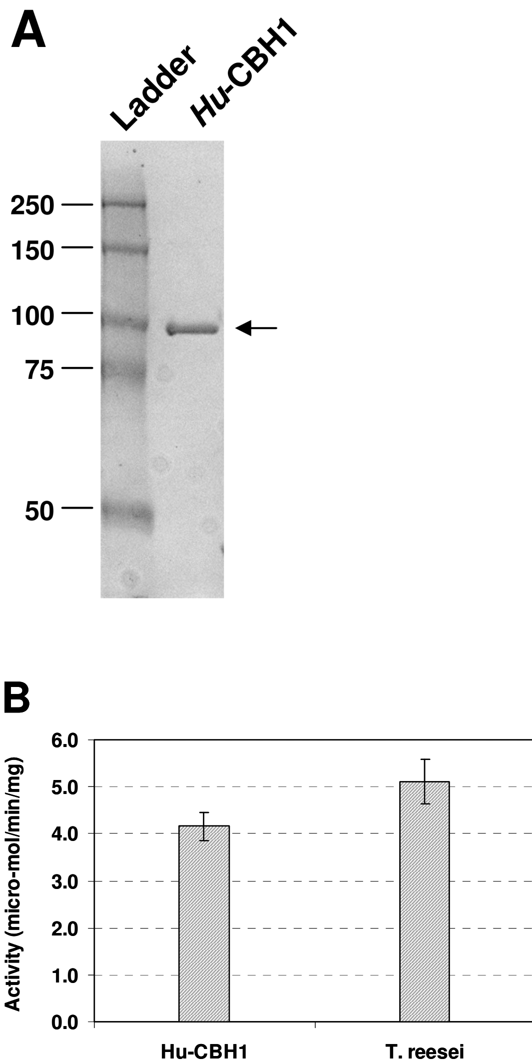


Figure 4. *Hu*-CBH1 activity and stability are regulated by salt.

A. *Hu*-CBH1 was used in a CMC assay conducted in different concentrations of NaCl buffer and 10 mM Tris-HCl (pH 7.0), at 37 °C for 1 hour. The enzyme activity in reaction containing 2 M NaCl is set as 100%. **B.** *Hu*-CBH1 was used in a CMC assay conducted in different concentrations of NaCl and 10 mM Tris-HCl (pH 7.0) at different temperatures for 1 hour. The activity of reaction in 2 M NaCl at 37 °C was set as 100%.

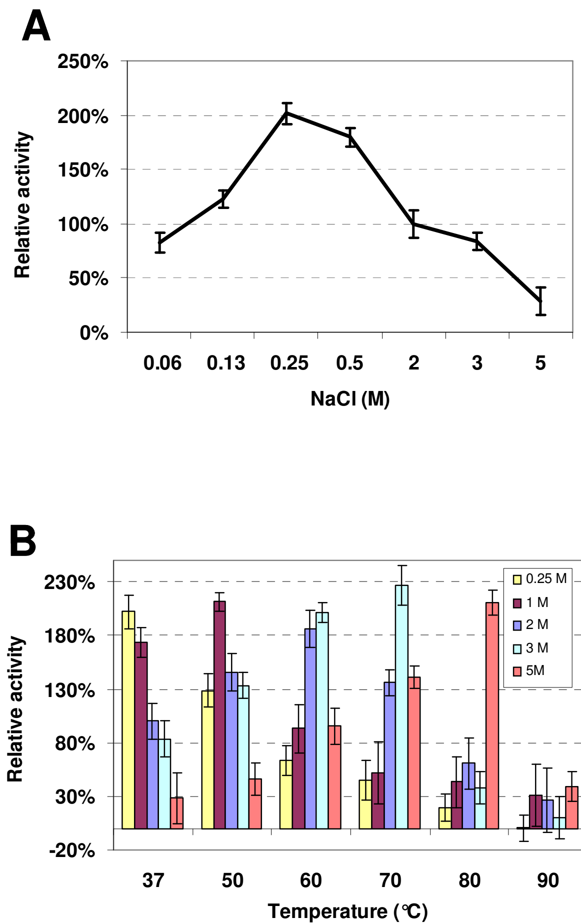


Figure 5. *Hu*-CBH1 requires high pH for its activity. *Hu*-CBH1 was used in a CMC assay containing 2 M NaCl, at 37 °C, at different pH values for 1 hour. The activity detected at pH 7.5 was set as 100%.

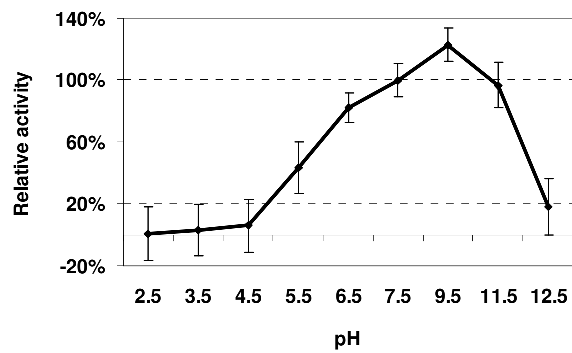
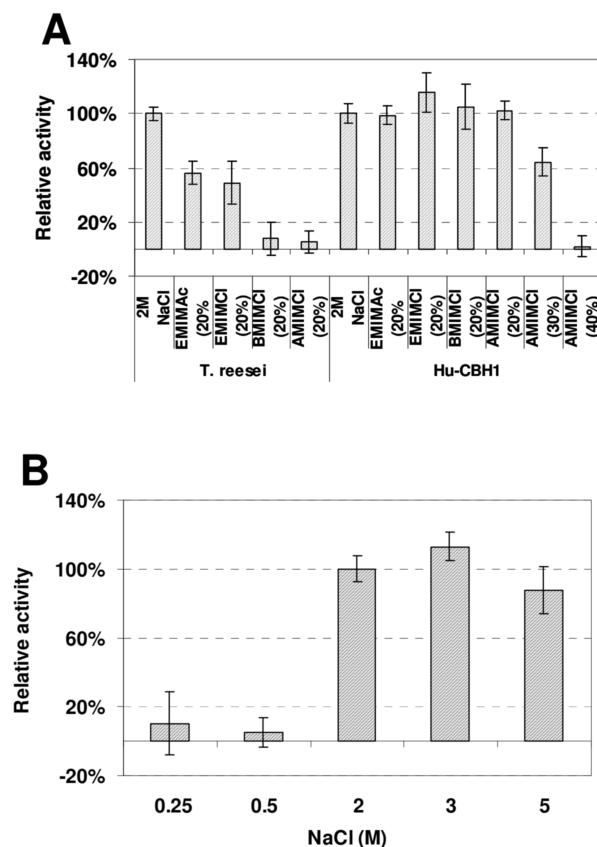


Figure 6. ***Hu*****-CBH1 is resistant to high concentrations of ILs in the presence of salt.** **A.** *T. reesei* cellulase and the *Hu*-CBH1 were are incubated with CMC substrate at 37 °C for 1 hour in the presence of 2 M NaCl and 10 mM Tris-HCl (pH 7.0), with or without addition of 20% of [Emim]Ac, [Emim]Cl, [Bmim]Cl or 20%, 30% and 40% of [Amim]Cl**.** The activities in reactions without ILs were set as 100% for both enzymes. **B.** *Hu*-CBH1 was incubated with CMC substrates and 20% [Amim]Cl, in the presence of 0.25, 0.5, 2, 3 and 5 M NaCl and 10 mM Tris-HCl (pH 7.0), at 37 °C for 1 hour. The activity of the reaction performed in 2 M NaCl was set as 100%.



References

(1) Wasserscheid, P.; Keim, W. *Angew Chem Int Ed Engl* **2000**, *39*, 3772.

(2) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J Am Chem Soc* **2002**, *124*, 4974.

(3) Kilpelainen, I.; Xie, H.; King, A.; Granstrom, M.; Heikkinen, S.; Argyropoulos, D. S. *J Agric Food Chem* **2007**, *55*, 9142.

(4) Zavrel, M.; Bross, D.; Funke, M.; Buchs, J.; Spiess, A. C. *Bioresour Technol* **2009**, *100*, 2580.

(5) Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol Bioeng* **2009**, *104*, 68.

(6) Fort, D. A.; Remsing, R. C.; Swatloski, R. P.; Moyna, P.; Moyna, G.; Rogers, R. D. *Green Chemistry* **2007**, *9*, 63.

(7) Fort, D. A.; Swatloski, R. P.; Moyna, P.; Rogers, R. D.; Moyna, G. *Chem Commun (Camb)* **2006**, 714.

(8) Zhao, H.; Jones, C. L.; Baker, G. A.; Xia, S.; Olubajo, O.; Person, V. N. *J Biotechnol* **2009**, *139*, 47.

(9) Lee, S. H.; Doherty, T. V.; Linhardt, R. J.; Dordick, J. S. *Biotechnol Bioeng* **2009**, *102*, 1368.

(10) Dadi, A. P.; Varanasi, S.; Schall, C. A. *Biotechnol Bioeng* **2006**, *95*, 904.

(11) Dadi, A. P.; Schall, C. A.; Varanasi, S. *Appl Biochem Biotechnol* **2007**, *137-140*, 407.

(12) Kamiya, N.; Matsushita, Y.; Hanaki, M.; Nakashima, K.; Narita, M.; Goto, M.; Takahashi, H. *Biotechnol Lett* **2008**, *30*, 1037.

(13) Turner, M. B.; Spear, S. K.; Huddleston, J. G.; Holbrey, J. D.; Rogers, R. D. *Green Chemistry* **2003**, 443.

(14) Park, S.; Kazlauskas, R. J. *Curr Opin Biotechnol* **2003**, *14*, 432.

(15) Bose, S.; Armstrong, D. W.; Petrich, J. W. *J Phys Chem B* **2010**, *114*, 8221.

(16) Moniruzzaman, M.; Kamiya, N.; Goto, M. *Org Biomol Chem* **2010**, *8*, 2887.

(17) Zhao, H.; Olubajo, O.; Song, Z.; Sims, A. L.; Person, T. E.; Lawal, R. A.; Holley, L. A. *Bioorg Chem* **2006**, *34*, 15.

(18) Constatinescu, D.; Herrmann, C.; Weingartner, H. *Phys Chem Chem Phys* **2010**, *12*, 1756.

(19) van Rantwijk, F.; Sheldon, R. A. *Chem Rev* **2007**, *107*, 2757.

(20) Paul, S.; Bag, S. K.; Das, S.; Harvill, E. T.; Dutta, C. *Genome Biol* **2008**, *9*, R70.

(21) Lanyi, J. K. *Bacteriol Rev* **1974**, *38*, 272.

(22) Hutcheon, G. W.; Vasisht, N.; Bolhuis, A. *Extremophiles* **2005**, *9*, 487.

(23) Fukuchi, S.; Yoshimune, K.; Wakayama, M.; Moriguchi, M.; Nishikawa, K. *J Mol Biol* **2003**, *327*, 347.

(24) Frolow, F.; Harel, M.; Sussman, J. L.; Mevarech, M.; Shoham, M. *Nat Struct Biol* **1996**, *3*, 452.

(25) Tadeo, X.; Lopez-Mendez, B.; Trigueros, T.; Lain, A.; Castano, D.; Millet, O. *PLoS Biol* **2009**, *7*, e1000257.

(26) Coquelle, N.; Talon, R.; Juers, D. H.; Girard, E.; Kahn, R.; Madern, D. *J Mol Biol* **2010**, *404*, 493.

(27) Zaccai, G. *Philos Trans R Soc Lond B Biol Sci* **2004**, *359*, 1269.

(28) Zaccai, G.; Cendrin, F.; Haik, Y.; Borochov, N.; Eisenberg, H. *Journal of Molecular Biology* **1989**, *208*, 491.

(29) Ebel, C.; Costenaro, L.; Pascu, M.; Faou, P.; Kernel, B.; Proust-De Martin, F.; Zaccai, G. *Biochemistry* **2002**, *41*, 13234.

(30) Mevarech, M.; Frolow, F.; Gloss, L. M. *Biophys Chem* **2000**, *86*, 155.

(31) Ebel, C.; Faou, P.; Kernel, B.; Zaccai, G. *Biochemistry* **1999**, *38*, 9039.

(32) Waino, M.; Tindall, B. J.; Ingvorsen, K. *Int J Syst Evol Microbiol* **2000**, *50 Pt 1*, 183.

(33) Wu, D.; Hugenholtz, P.; Mavromatis, K.; Pukall, R.; Dalin, E.; Ivanova, N. N.; Kunin, V.; Goodwin, L.; Wu, M.; Tindall, B. J.; Hooper, S. D.; Pati, A.; Lykidis, A.; Spring, S.; Anderson, I. J.; D'Haeseleer, P.; Zemla, A.; Singer, M.; Lapidus, A.; Nolan, M.; Copeland, A.; Han, C.; Chen, F.; Cheng, J. F.; Lucas, S.; Kerfeld, C.; Lang, E.; Gronow, S.; Chain, P.; Bruce, D.; Rubin, E. M.; Kyrpides, N. C.; Klenk, H. P.; Eisen, J. A. *Nature* **2009**, *462*, 1056.

(34) Waino, M.; Ingvorsen, K. *Extremophiles* **2003**, *7*, 87.

(35) Bakke, P.; Carney, N.; Deloache, W.; Gearing, M.; Ingvorsen, K.; Lotz, M.; McNair, J.; Penumetcha, P.; Simpson, S.; Voss, L.; Win, M.; Heyer, L. J.; Campbell, A. M. *PLoS One* **2009**, *4*, e6291.

(36) Little, E.; Bork, P.; Doolittle, R. F. *J Mol Evol* **1994**, *39*, 631.

(37) Kataeva, I. A.; Seidel, R. D., 3rd; Shah, A.; West, L. T.; Li, X. L.; Ljungdahl, L. G. *Appl Environ Microbiol* **2002**, *68*, 4292.

(38) Liu, H.; Pereira, J. H.; Adams, P. D.; Sapra, R.; Simmons, B. A.; Sale, K. L. *FEBS Lett* **2010**, *584*, 3431.

(39) Pereira, J. H.; Sapra, R.; Volponi, J. V.; Kozina, C. L.; Simmons, B.; Adams, P. D. *Acta Crystallogr D Biol Crystallogr* **2009**, *65*, 744.

(40) Natale, P.; Bruser, T.; Driessen, A. J. M. *Bba-Biomembranes* **2008**, *1778*, 1735.

(41) Vyazmensky, M.; Barak, Z.; Chipman, D. M.; Eichler, J. *Comp Biochem Physiol B Biochem Mol Biol* **2000**, *125*, 205.

(42) Datta, S.; Holmes, B.; Park, J. I.; Chen, Z. W.; Dibble, D. C.; Hadi, M.; Blanch, H. W.; Simmons, B. A.; Sapra, R. *Green Chemistry* **2010**, *12*, 338.

(43) Shill, K.; Padmanabhan, S.; Xin, Q.; Prausnitz, J.; Clark, D. S.; Blanch, H. W. *Biotechnol Bioeng* **2010**.

(44) Wu, J.; Zhang, J.; Zhang, H.; He, J.; Ren, Q.; Guo, M. *Biomacromolecules* **2004**, *5*, 266.

(45) Zhu, S. D.; Wu, Y. X.; Chen, Q. M.; Yu, Z. N.; Wang, C. W.; Jin, S. W.; Ding, Y. G.; Wu, G. *Green Chemistry* **2006**, *8*, 325.

(46) Vreeland, R. H.; Piselli, A. F., Jr.; McDonnough, S.; Meyers, S. S. *Extremophiles* **1998**, *2*, 321.

(47) Rohban, R.; Amoozegar, M. A.; Ventosa, A. *J Ind Microbiol Biotechnol* **2009**, *36*, 333.

(48) Rees, H. C.; Grant, S.; Jones, B.; Grant, W. D.; Heaphy, S. *Extremophiles* **2003**, *7*, 415.

(49) Voget, S.; Steele, H. L.; Streit, W. R. *J Biotechnol* **2006**, *126*, 26.

(50) Gao, Z.; Ruan, L.; Chen, X.; Zhang, Y.; Xu, X. *Appl Microbiol Biotechnol* **2010**, *87*, 1373.

(51) Taupin, C. M.; Hartlein, M.; Leberman, R. *Eur J Biochem* **1997**, *243*, 141.

(52) Cendrin, F.; Chroboczek, J.; Zaccai, G.; Eisenberg, H.; Mevarech, M. *Biochemistry* **1993**, *32*, 4308.

(53) Zhou, W.; Irwin, D. C.; Escovar-Kousen, J.; Wilson, D. B. *Biochemistry* **2004**, *43*, 9655.

(54) Bayer, E. A.; Lamed, R.; White, B. A.; Flint, H. J. *Chem Rec* **2008**, *8*, 364.

(55) Hakamada, Y.; Hatada, Y.; Koike, K.; Yoshimatsu, T.; Kawai, S.; Kobayashi, T.; Ito, S. *Biosci Biotechnol Biochem* **2000**, *64*, 2281.

(56) Endo, K.; Hakamada, Y.; Takizawa, S.; Kubota, H.; Sumitomo, N.; Kobayashi, T.; Ito, S. *Appl Microbiol Biotechnol* **2001**, *57*, 109.

(57) Hirasawa, K.; Uchimura, K.; Kashiwa, M.; Grant, W. D.; Ito, S.; Kobayashi, T.; Horikoshi, K. *Antonie Van Leeuwenhoek* **2006**, *89*, 211.

(58) Blecher, O.; Goldman, S.; Mevarech, M. *Eur J Biochem* **1993**, *216*, 199.

(59) Kaczowka, S. J.; Maupin-Furlow, J. A. *J Bacteriol* **2003**, *185*, 165.

(60) Cline, S. W.; Lam, W. L.; Charlebois, R. L.; Schalkwyk, L. C.; Doolittle, W. F. *Can J Microbiol* **1989**, *35*, 148.

(61) Ghose, T. K. *Pure & Appl. Chem.* **1987**, *59*, 257.

(62) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195.

(63) Kiefer, F.; Arnold, K.; Kunzli, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Res* **2009**, *37*, D387.

(64) Peitsch, M. C. *Bio-Technol* **1995**, *13*, 723.

(65) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Research* **2004**, *32*, W665.

(66) Schrödinger, L., *Version 1.2r2*.

(67) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *P Natl Acad Sci USA* **2001**, *98*, 10037.

(68) Gutowski, K. E.; Broker, G. A.; Willauer, H. D.; Huddleston, J. G.; Swatloski, R. P.; Holbrey, J. D.; Rogers, R. D. *J Am Chem Soc* **2003**, *125*, 6632.