ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

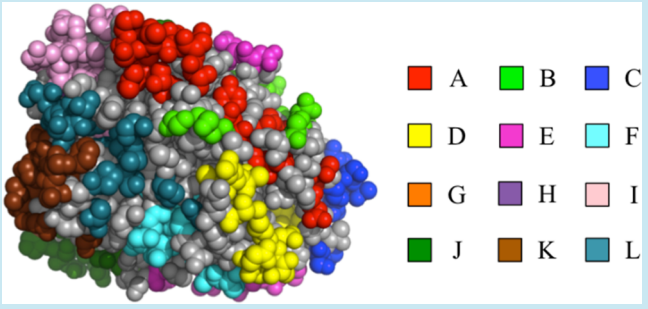
Hypocrea jecorina Cellobiohydrolase I Stabilizing Mutations Identified Using Noncontiguous Recombination

Matthew A. Smith, Claire N. Bedbrook, Timothy Wu, and Frances H. Arnold*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States

Supporting Information

ABSTRACT: Noncontiguous recombination (NCR) is a method to identify pieces of structure that can be swapped among homologous proteins to create new, chimeric proteins. These "blocks" are encoded by elements of sequence that are not necessarily contiguous along the polypeptide chain. We used NCR to design a library in which blocks of structure from *Hypocrea jecorina* cellobiohydrolase I (Cel7A) and its two thermostable homologues from *Talaromyces emersonii* and *Chaetomium thermophilum* are shuffled to create 531,438 possible chimeric enzymes. We constructed a maximally informative subset of 35 chimeras to analyze this library and found that the blocks contribute additively to the stability of a chimera. Within two highly stabilizing blocks, we uncovered six single amino acid substitutions that each improve the stability of *H. jecorina* cellobiohydrolase I by 1−3 °C. The small number of measurements required to find these mutations demonstrates that noncontiguous recombination is an efficient strategy for identifying stabilizing mutations.



KEYWORDS: *CBHI, Cel7A, cellulase, protein recombination, thermostability*

Highly thermostable cellulases are desirable for the production of sugars from cellulosic substrates. Thermotolerant mixtures of fungal cellulases have been shown to degrade cellulose faster at elevated temperatures than mixtures from mesophilic fungi.[1] At elevated temperatures, cellulolytic processes can benefit from reduced contamination and viscosity of the biomass slurry as well as increased cellulase hydrolysis rates.

The success of the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*) as an industrial cellulase producer derives from its ability to secrete cellulases at up to 100 g/L. Cellobiohydrolase I (CBHI, Cel7A) is one of the most important cellulase components. Removal of the *cbh1* gene reduces the cellulolytic activity of the fungus by 70% and the total secreted protein by 40%.[2] Low expression levels and altered glycosylation patterns,[3,4] however, make this enzyme difficult to engineer in heterologous expression systems.

There have been various efforts to engineer improved CBHI variants, including screening random mutants,[5] engineering disulfide bonds,[6] and DNA shuffling.[7] In addition, we have sought to enhance CBHI stability through protein recombination[8] and predictive methods.[9] The most stable CBHI enzymes from these latter works have more than 150 mutations from *H. jecorina* CBHI, which could adversely affect the high titers of secreted protein in fungal expression systems. We therefore sought to stabilize the *H. jecorina* CBHI by making minimal mutations to its native sequence.

We recently introduced a method for noncontiguous protein recombination[10] that identifies elements of structure ("blocks") that can be shuffled among homologous proteins. Unlike previous SCHEMA recombination libraries that swap elements of sequence,[11] these elements of structure are not necessarily contiguous polypeptide sequences. Here we show how noncontiguous recombination can be used to efficiently identify stabilizing mutations that have been incorporated into CBHI homologues in nature.

Swapping structural blocks among *H. jecorina* CBH1 and two thermostable homologues from *Talaromyces emersonii* and *Chaetomium thermophilum*, we analyze a subset of CBHIs from a library containing more than 500,000 possible chimeric sequences. We predict the thermostabilities of all library members using data from a maximally informative subset of just 32 chimeras (and 3 parents) and identify several blocks that are predicted to stabilize *H. jecorina* CBHI. Searching within these blocks, we find six single amino acid substitutions that stabilize *H. jecorina* CBHI by more than 1 °C. One previously undiscovered mutation improves its thermostability by 3 °C.

RESULTS AND DISCUSSION

Noncontiguous Protein Recombination Library Design. We wish to shuffle elements of sequence among homologous proteins to create a library of chimeras highly enriched in functional sequences. A good metric for the functional impairment of a chimeric protein is its SCHEMA disruption,[12] which is the number of non-native residue−

Received: February 14, 2013
Published: May 20, 2013

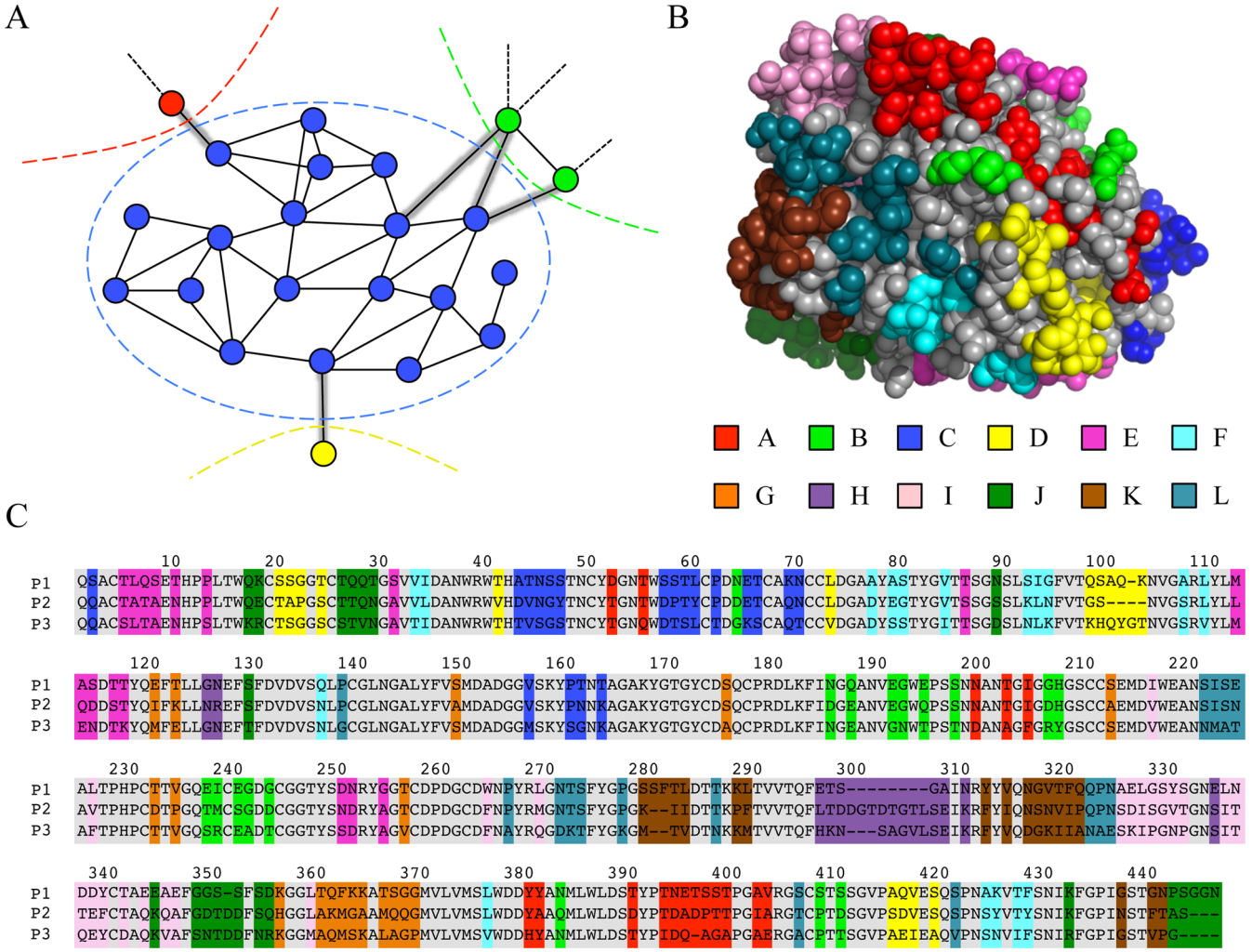


Figure 1. Noncontiguous recombination library design. (A) A graph view of the blue block and neighboring residues. Nodes represent residues, and edges represent residue−residue contacts. Colored, dashed lines define the graph partitions for each block. Contacts to residues from other blocks (highlighted) will be broken upon recombination. (B) The 12-block design displayed on the structure of P2 (1Q9H.pdb). Each block (labeled A to L) is represented by a different color, and conserved residues are in gray. (C) The 12-block design displayed on the numbered sequence alignment of the catalytic domains of the three parental enzymes.

residue contacts formed in the recombined sequence. We have used this metric previously to design recombination libraries that shuffled contiguous blocks of sequence.[11,13] Recombination of structural elements can be significantly less disruptive than recombining sequence elements. We recently presented a method for finding the optimal structural blocks for any given set of parent proteins, based on a graph partitioning algorithm.[10]

For NCR, we create a graph from the non-native residue−residue contacts, with nodes corresponding to residues and edges corresponding to non-native contacts. NCR minimizes the SCHEMA disruption by identifying minimal cuts that partition the graph.[10] We partition the graph with hmetis,[14,15] a suite of graph partitioning tools. Residues are assigned to blocks based on how nodes are assigned to partitions. Blocks can have noncontiguous sequences but will be contiguous pieces of structure in 3 dimensions. Shuffling these blocks generates a library of noncontiguous chimeras.

As parental enzymes we chose the catalytic domains of three fungal CBHI cellulases: *H. jecorina* CBHI (P1), *T. emersonii* CBHI (P2), and *C. thermophilum* CBHI (P3). *T. emersonii* CBHI has one fewer disulfide bond than *H. jecorina* CBH1 and *C. thermophilum* CBH1, which each have 10. To ensure that unpaired cysteines do not appear in the chimeras, we mutated P2 to include the missing cysteine pair (G4C, A72C). This extra disulfide bond is known to increase the stability of P2.[6] These three cellulase catalytic domains were used in a previous study to create a contiguous block SCHEMA recombination library.[8]

For ease of identifying stabilizing point mutations within a block, we divided mutations among blocks so each block contained only a small number of mutations. We also required the blocks to be of equal size to ensure a fair comparison of block stability contributions. We designed a 12-block library where each block contained approximately 18 nonconserved residues (see Materials and Methods). The design has an average SCHEMA disruption (number of disrupted contacts) of 24.8 and an average of 83.4 mutations from the closest parent. Most non-native residue−residue contacts are sequestered within blocks (Figure 1A), which increases the fraction of the library that is likely to be folded and functional. While almost all the blocks are contiguous pieces of structure (Figure 1B), they each comprise many fragments of the polypeptide chain (Figure 1C).

dx.doi.org/10.1021/sb400010m | *ACS Synth. Biol.* 2013, 2, 690−696

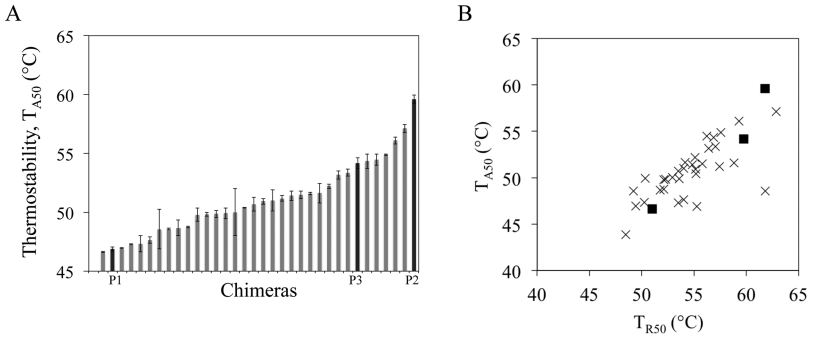


Figure 2. Thermostabilities of a maximally informative subset of the library. (A) T_{A50}: The elevated temperature at which a chimera's activity is half its maximum. Measurements were performed in duplicate. The parental enzymes are highlighted. (B) A plot of the elevated temperature at which an enzyme loses half its activity (T_{A50}) against the incubation temperature at which an enzyme loses half its (unincubated) activity (T_{R50}). The parental cellulases are highlighted with black squares. While most of the T_{R50} and T_{A50} measurements are similar, several cellulases have significantly higher T_{R50} than T_{A50}.

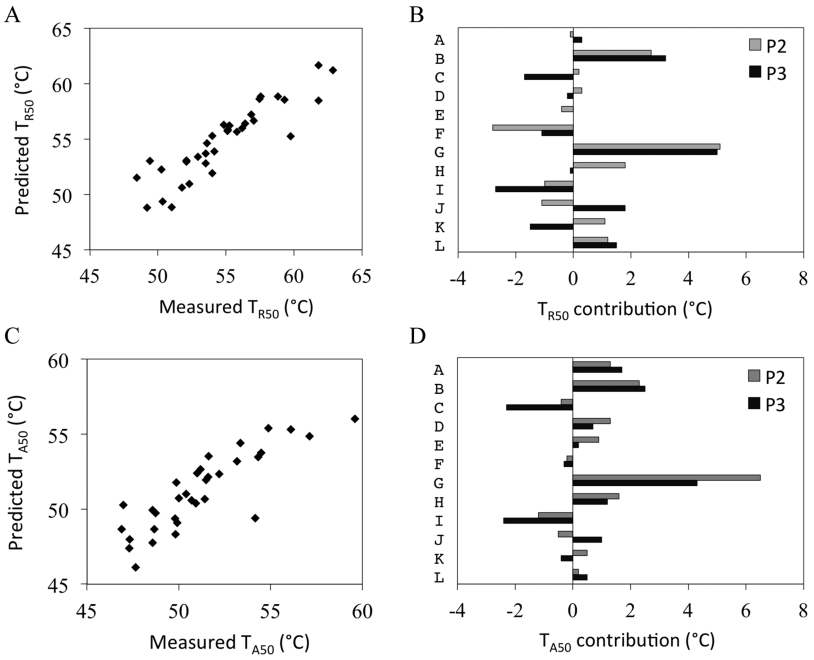


Figure 3. The thermostability of a chimera can be predicted with a simple linear model that sums the contributions from each block. (A) A linear thermostability model trained on the T_{R50}s of the chimeras accurately predicts the measured values (r^2 = 0.81). Blocks G and B from *T. emersonii* and *C. thermophilum* are predicted to be significantly stabilizing relative to those blocks from *H. jecorina*. (B) The predicted T_{R50} contributions of each block from parents P2 and P3 relative to parent P1. (C) A linear thermostability model trained on the T_{A50}s of the chimeras accurately predicts the measured values (r^2 = 0.74). (D) The predicted T_{A50} contributions of each block from parents P2 and P3 relative to parent P1.

Some groups of residues only have SCHEMA contacts with one another and not with the rest of the protein. These disconnected "sub-blocks" can belong to any block without altering the SCHEMA disruption, and they appear separate from the rest of the block. Blocks A, D, E, and J contain disconnected sub-blocks and thus contain several separate pieces of structure.

Stabilities of an Informative Subset. Our 3-parent, 12-block library contains more than half a million chimeras. The nature of noncontiguous recombination makes it very difficult to construct these chimeras with traditional cloning techniques. Because it is neither feasible nor necessary to synthesize and analyze the entire library, we selected a highly informative subset of 35 chimeras to construct and characterize (Supplementary Table 1 in the Supporting Information). These chimeras were chosen to maximize mutual information about the sequences (see Materials and Methods), as described previously for a library of chimeric arginases.[16] At the same time, the chosen sequences had low SCHEMA disruption in order to enrich the chimera subset in functional sequences. To the C-terminus of each chimeric catalytic domain we added the linker and carbohydrate binding module from *H. jecorina* CBHI.

Of the 35 chimeras synthesized, 32 (91%) were expressed with detectable levels of activity. We quantified the thermostabilities of the 32 chimeras and three parents using two measures. We define T_{R50} as the incubation temperature at which an enzyme loses half its (unincubated) activity. We incubated the enzymes at a range of temperatures for 10 min without substrate and measured the residual activities. These T_{R50}s are plotted in Supplementary Figure 1 in the Supporting Information. Most of the chimeras have stabilities that lie between those of the parents, but several were more stable than the most stable parent, P2.

dx.doi.org/10.1021/sb400010m | *ACS Synth. Biol.* 2013, 2, 690−696

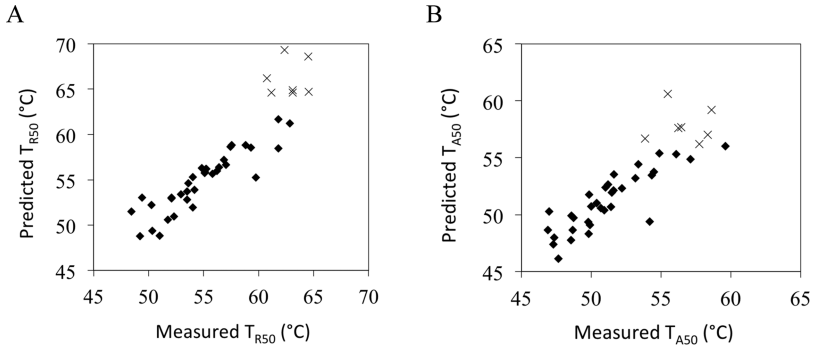


Figure 4. The thermostability models identify stable CBHI chimeric cellulases in the library. (A) Predicted T_{R50} against measured T_{R50} for seven chimeras predicted to have high stabilities (crosses). The original data used to train the model are represented as filled diamonds. (B) Predicted T_{A50} against measured T_{A50} for seven chimeras predicted to have high stabilities (crosses). The original data used to train the model are shown as filled diamonds.

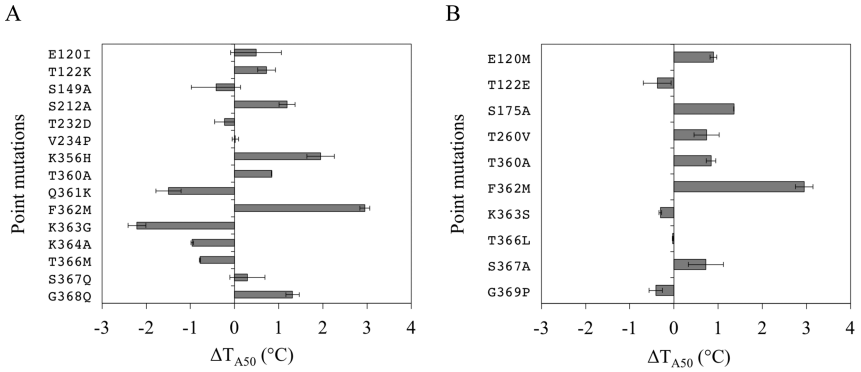


Figure 5. The effect on *H. jecorina* CBHI thermostability (T_{A50}) for a series of point mutations from two of the most stabilizing blocks. (A) Block G, parent P2. (B) Block G, parent P3. Two of the mutations (T360A and F362M) are present in both blocks. F362M is stabilizing by 3 °C.

We define T_{A50} as the elevated temperature at which an enzyme loses half its activity measured at its optimum temperature. We ran a 2 h activity assay at a range of temperatures and measured the total enzyme activities. Whereas T_{R50} is a measure of enzyme tolerance to thermal stress, T_{A50} measures an enzyme's ability to function at elevated temperature.

Values of T_{A50} are plotted in Figure 2A. T_{R50}s and T_{A50}s are correlated for the chimeras (Figure 2B), but there are some outliers where the T_{R50} greatly exceeds the T_{A50}. Even though the enzymes are incubated in 1 mM DTT, these are cases where the CBHIs are able to refold and regain activity once the temperature is reduced for the assay (Supplementary Figure 2 in the Supporting Information).

Modeling Thermostability. We have previously shown that contiguous blocks of sequence contribute additively to the stabilities of chimeras and that these stabilities are predictable with simple additive block models trained on a small sample of a library.[17,18] Here we used the same linear regression model to demonstrate that contiguous blocks of structure (with non-contiguous blocks of sequence) also contribute additively to the stabilities of recombined enzymes. We constructed predictive models of T_{R50} and T_{A50} based on the sequences of the 32 functional chimeras and three parental cellulases (see Materials and Methods). As shown in Figure 3A, the T_{R50} model accurately predicts the stabilities of the library sample (r^2 = 0.81). This model provides the predicted contributions of each structural block to T_{R50} (Figure 3B). Similarly, we trained a model that fits the T_{A50} stability data (r^2 = 0.74, Figure 3C). The predicted block contributions to T_{A50} are shown in Figure 3D. In both models, block G appears to be highly stabilizing to parent P1 when taken from either parent P2 or P3. There are two mutations common to P2 and P3 in this block, T360A and F362M.

With the stability models constructed from this highly informative sample set, we can predict the T_{R50}s and T_{A50}s of all the untested chimeras in the library. We correctly identified seven chimeras from the library expected to have both high T_{R50}s and T_{A50}s (Figure 4A and B, Supplementary Table 2 in the Supporting Information). While two of the predicted chimeras had T_{R50}s 2 °C higher than the most stable parent (P2), none of the chimeras had T_{A50}s above the most stable parent.

Stabilizing Point Mutations. We wish to stabilize *H. jecorina* CBHI (P1) with minimal disruption to its amino acid sequence. Using linear regression, we have identified two highly stabilizing blocks, block G from P2 and block G from P3. We placed each of these blocks in place of block G in *H. jecorina* CBH1 and found they were indeed stabilizing, improving *H. jecorina* CBHI's T_{R50} by 1.7 and 1.1 °C, respectively (Supplementary Table 3 in the Supporting Information). Given that a single block is made up of a combination of stabilizing and destabilizing mutations, we wanted to identify the individual amino acids that have the most significant positive contribution to stability. Similar to the approach we used on Cel6 cellulases,[19] we searched these blocks for individual mutations that stabilize P1 by substituting each of the 23 point mutations into P1 and measuring the T_{A50} (Figure 5A,B). Most of the amino acid substitutions have only a slight effect on *H. jecorina* CBHI thermostability (less than 1 °C). Of

dx.doi.org/10.1021/sb400010m | *ACS Synth. Biol.* 2013, 2, 690−696

the remaining mutations, three are destabilizing and six are stabilizing. One of the stabilizing mutations, F362M, present in both P2 and P3, is stabilizing by a full 3 °C. This mutation allows *H. jecorina* CBHI to retain higher levels of activity at elevated temperatures (Figure 6).

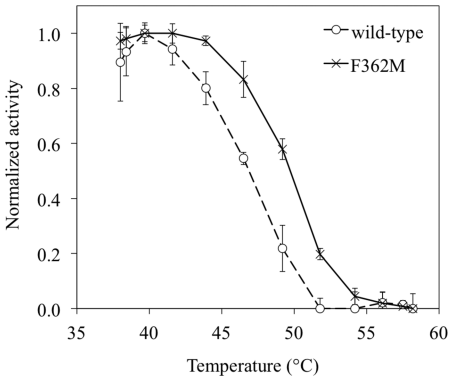


Figure 6. Effect of the mutation F362M (×) on the activity of *H. jecorina* CBHI (○) in a 90 min assay on MUL over a range of temperatures, performed in quadruplicate. To account for differences in levels of secreted cellulase, CBHI activity was normalized by the activity at 40 °C.

We created seven mutants that were combinations of the most stabilizing mutations. None of these combination mutants was more stable than the F362M single mutant (Supplementary Table 4 in the Suppporting Information).

SUMMARY AND CONCLUSIONS

We have utilized a new noncontiguous recombination method to design a library with more than 500,000 sequences enriched in functional family 7 cellobiohydrolases. NCR identifies swappable elements of structure that are not necessarily contiguous pieces of polypeptide. Because library designs with contiguous sequence elements are a small subset of the large number of possible noncontiguous designs, this approach identifies libraries that disrupt fewer SCHEMA contacts and therefore contain more functional chimeric proteins than contiguous block design algorithms such as RASPP.[20] Indeed, 39 of the 42 synthesized chimeras were functional even though on average they had 79 mutations from the closest parent.

By measuring the thermostabilities of a maximally informative subset of 32 chimeras, we showed that the structural blocks identified by NCR contribute additively to protein stability. Furthermore, we used a block-additive stability model to correctly identify several stable chimeras in the library.

H. jecorina CBHI, *T. emersonii* CBHI, and *C. thermophilum* CBHI differ at 213 residues. It is difficult to identify the stabilizing mutations from just analyzing the sequences, and it would be cumbersome to construct and test all 273 single mutants. The "divide and conquer" method we present first measures the stabilities of functional groups of mutations (blocks) and then identifies stabilizing single mutants within the most stable blocks. With a small number of experimental measurements, we identified two blocks that significantly stabilize *H. jecorina* CBHI and uncovered a single amino acid substitution that stabilizes this important industrial enzyme by 3 °C.

Despite much previous work on stabilizing *H. jecorina* CBHI, the F362M mutation has not been described previously. The mutation is located close to the surface of the protein facing inward, and the sulfur atom is proximal to the sulfur of another methionine residue. The enhanced stability may come from interaction of these two residues, possibly a hydrogen bond if one of the methionines is oxidized to methionine sulfoxide.

Surprisingly, combining this mutation with other stabilizing mutations did not produce an even more more stable cellulase. The stability contributions of the single mutations are nonadditive, which contrasts with the highly additive stability contributions of the SCHEMA NCR blocks. The observed nonadditivity may be caused by unfavorable interactions between pairs of mutations and suggests that the stabilizing effects of blocks include higher-order interactions among groups of mutations.

NCR identifies elements of structure that, when swapped, preserve protein function.[10] Splitting the CBHI structure into a relatively large number of equally sized blocks and swapping these structural elements with stable homologues has proven to be an efficient strategy to search for stabilizing mutations. While we tested all 23 single mutations from the two most stabilizing blocks, the most stabilizing single mutation was present in both blocks. A method of prioritizing point mutations within a stable block, such as using consensus mutagenesis, may further improve the speed with which valuable mutations are identified.

MATERIALS AND METHODS

Noncontiguous Recombination. PROMALS3D[21] was used to create a structure-based sequence alignment of the catalytic domains from *H. jecorina* CBHI (P1), *T. emersonii* CBHI (P2), and *C. thermophilum* CBHI (P3). Residues that have (non-hydrogen) atoms closer than 4.5 Å are considered to be in contact with one another. All residue−residue contacts were identified in PDB structure 1Q9H.pdb chain A. Contacts not conserved among the three parent enzymes form the SCHEMA contact map.

Designing libraries that minimize the average number of SCHEMA contacts in the resulting chimeras was reformulated as a graph partitioning problem. The SCHEMA contact map was transformed into a graph with each node representing a nonconserved residue and each weighted edge representing an average SCHEMA contact between two residues. Residues were assigned to blocks such that the sum of weighted edges between blocks was minimized. For the 12-block library designs, the hmetis graph partitioning suite[14,15] was used to perform a series of 12-way partitions of the SCHEMA contact map. A library design was chosen with an average SCHEMA energy (number of disrupted contacts) of 24.8 and an average of 83.4 mutations from the closest parent. Residues 41, 175, 197, 199, 202, and 442 have no SCHEMA contacts and were not partitioned into blocks; we assigned these residues to blocks D, G, B, A, A, J, respectively, based on their spatial proximity to those blocks. The C-terminal linker and carbohydrate binding module from *H. jecorina* CBHI was appended to each chimera.

Optimal Experimental Design. A greedy algorithm was employed to find a subset of sequences from the library with low SCHEMA disruption and maximized mutual information, as described.[16] Due to computational constraints, the informative set of chimeras was identified from 50,000 randomly chosen chimeras with a SCHEMA disruption below 30, rather than the entire library. This optimized experimental design was carried out with the Submodular Function Optimization Matlab toolbox.[22]

Gene Synthesis. The chimeric CBHI genes were optimized for expression in *Saccharomyces cerevisiae* and synthesized by DNA2.0 (Menlo Park, CA, USA).

Protein Expression. The genes encoding parental and chimeric CBHI catalytic domains were cloned into the yeast expression vector Yep352/PGT91-1-*αss* with an N-terminal His_6 tag and the *H. jecorina* CBHI linker and cellulose binding domain attached to the C-terminus. These vectors were transformed into yeast strain YDR483W BY4742 (*Matα hus3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 Δkre2*, ATCC No. 4014317) as described[23] and plated on synthetic dropout-uracil medium with 10 g/L agar. The plates were incubated for 2 days at 30 °C. 5 mL of synthetic dropout-uracil medium was inoculated by a single yeast colony from a plate and incubated for 1 day at 30 °C, with shaking at 250 rpm. Cultures were expanded at a 1:10 ratio into either 10 or 50 mL of yeast peptone dextrose (YPD) medium (10 g of yeast extract, 20 g of peptone, 20 g of dextrose) and incubated for 2 days at 30 °C, with shaking at 250 rpm. The cells were pelleted by centrifugation at 5000*g* for 10 min, and the supernatant, containing the secreted cellulases, was decanted and separated through a 0.20 μm pore size conical filter unit from Nalgene (Rochester, NY, USA). The supernatant was concentrated up to 4-fold using Vivaspin 20 spin columns with a 30 kDa MWCO PES membrane from GE Healthcare (Little Chalfont, U.K.) and stored at 4 °C with 0.02% sodium azide and 1 mM phenylmethanesulfonyl-fluoride.

Thermostability Residual Activity Assay (T_{R50} Measurement). In a 96-well PCR plate, 100 μL of supernatant was added to 25 μL of 625 mM sodium acetate, pH 4.8 with 5 mM dithiothreitol (DTT), giving a final concentation of 125 mM sodium acetate, pH 4.8 and 1 mM DTT, as described.[9] The plate was incubated in a gradient thermocycler for 10 min at a range of temperatures and then cooled to 4 °C. To each well was added 25 μL of 1.8 mM 4-methylumbelliferyl lactopyranoside (MUL) from Sigma-Aldrich (St. Louis, MI, USA) dissolved in 18% DMSO and 125 mM sodium acetate. The heat-treated cellulases were incubated in a thermocycler for 90 min at 45 °C. The reaction was quenched by adding 150 μL of 1 M Na_2CO_3, and cellulase activity was quantified by measuring the fluorescence of released 4-methylumbelliferone with excitation at 364 nm and emission at 445 nm.

Thermostability Activity Assay (T_{A50} Measurement). In a 96-well PCR plate, 100 μL of supernatant was added to 25 μL of 625 mM sodium acetate, pH 4.8, and 25 μL of 1.8 mM 4-methylumbelliferyl lactopyranoside (MUL) dissolved in 18% DMSO and 125 mM sodium acetate. The plate was incubated in a gradient thermocycler for 90 min at a range of temperatures and then cooled to 4 °C. The reaction was quenched by adding 150 μL of 1 M Na_2CO_3, and cellulase activity was quantified by measuring the fluorescence of released 4-methylumbelliferone with excitation at 364 nm and emission at 445 nm.

Linear Regression. Stability models for T_{R50} and T_{A50} were constructed as described previously[17] and trained using Matlab's "regress" function.

■ ASSOCIATED CONTENT

Supporting Information

Stability measurements and amino acid sequences of the chimeras. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: frances@cheme.caltech.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank John McIntosh and Mary Farrow for useful discussions in the preparation of this manuscript. This work was supported by the Institute for Collaborative Biotechnologies through a grant [W911NF-09-D-0001] from the U.S. Army Research and The National Central University, Taiwan, though a Cooperative Agreement for Energy Research Collaboration. M.A.S. is supported by a Resnick Sustainability Institute fellowship.

■ ABBREVIATIONS

NCR, noncontiguous recombination; CBHI, cellobiohydrolase I; MUL, 4-methylumbelliferyl lactopyranoside; DTT, dithiothreitol

■ REFERENCES

(1) Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J., and Siika-Aho, M. (2007) Thermostable enzymes in lignocellulose hydrolysis. *Adv. Biochem. Eng. Biotechnol. 108*, 121−145.

(2) Suominen, P. L., Mantyla, A. L., Karhunen, T., Hakola, S., and Nevalainen, H. (1993) High frequency one-step gene replacement in Trichoderma reesei. II. Effects of deletions of individual cellulase genes. *Mol. Gen. Genet. 241*, 523−530.

(3) Boer, H., Teeri, T. T., and Koivula, A. (2000) Characterization of Trichoderma reesei cellobiohydrolase Cel7A secreted from Pichia pastoris using two different promoters. *Biotechnol. Bioeng. 69*, 486−494.

(4) Jeoh, T., Michener, W., Himmel, M. E., Decker, S. R., and Adney, W. S. (2008) Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Biotechnol Biofuels 1*, 10.

(5) Voutilainen, S., Boer, H., and Alapuranen, M. (2009) Improving the thermostability and activity of Melanocarpus albomyces cellobiohydrolase Cel7B. *Appl. Microbiol. Biotechnol. 83*, 261−272.

(6) Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010) Expression of Talaromyces emersonii cellobiohydrolase Cel7A in Saccharomyces cerevisiae and rational mutagenesis to improve its thermostability and activity. *Protein Eng. Des. Sel. 23*, 69−79.

(7) Dana, C. M., Saija, P., Kal, S. M., Bryan, M. B., Blanch, H. W., and Clark, D. S. (2012) Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol. Bioeng. 109*, 2710−2719.

(8) Heinzelman, P., Komor, R., Kanaan, A., Romero, P. A., Yu, X., Mohler, S., Snow, C., and Arnold, F. H. (2010) Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng. Des. Sel. 23*, 871−880.

(9) Komor, R. S., Romero, P. A., Xie, C. B., and Arnold, F. H. (2012) Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng. Des. Sel. 25*, 827−833.

(10) Smith, M. A., Romero, P. A., Wu, T., Brustad, E. M., and Arnold, F. H. (2013) Chimeragenesis of distantly-related proteins by noncontiguous recombination. *Protein Sci. 22*, 231−238.

(11) Otey, C. R., Landwehr, M., Endelman, J. B., Hiraga, K., Bloom, J. D., and Arnold, F. H. (2006) Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol. 4*, e112.

(12) Voigt, C. A., Martinez, C., Wang, Z.-G., Mayo, S. L., and Arnold, F. H. (2002) Protein building blocks preserved by recombination. *Nat. Struct. Biol. 9*, 553−558.

(13) Meyer, M., Hochrein, L., and Arnold, F. H. (2006) Structure-guided SCHEMA recombination of distantly related *β*-lactamases. *Protein Eng. Des. Sel. 19*, 563−570.

(14) Karypis, G., Aggarwal, R., Kumar, V., and Shekhar, S. (1997) Multilevel hypergraph partitioning: application in VLSI domain, in *Proceedings of the 34th annual Design Automation Conference*, pp 526−529, ACM Press, New York, New York, USA.
(15) Karypis, G., and Kumar, V. (2000) Multilevel k-way hypergraph partitioning. *VLSI Des. 11*, 285−300.
(16) Romero, P., Stone, E., Lamb, C., Chantranupong, L., Krause, A., Miklos, A., Hughes, R., Fechtel, B., Ellington, A. D., Arnold, F. H., and Georgiou, G. (2012) SCHEMA-designed variants of human arginase I and II reveal sequence elements important to stability and catalysis. *ACS Synth. Biol. 1*, 221−228.
(17) Li, Y., Drummond, D. A., Sawayama, A. M., Snow, C. D., Bloom, J. D., and Arnold, F. H. (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol. 25*, 1051−1056.
(18) Romero, P. A., Krause, A., and Arnold, F. H. (2013) Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl. Acad. Sci. U.S.A. 110*, E193−E201.
(19) Heinzelman, P., Snow, C. D., Smith, M. A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F. H. (2009) SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J. Biol. Chem. 284*, 26229−26233.
(20) Endelman, J., Silberg, J., Wang, Z., and Arnold, F. H. (2004) Site-directed protein recombination as a shortest-path problem. *Protein Eng. Des. Sel. 17*, 589−594.
(21) Pei, J., Kim, B.-H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res. 36*, 2295−2300.
(22) Krause, A. (2010) SFO: A Toolbox for Submodular Function Optimization. *J. Mach. Learn. Res. 11*, 1141−1144.
(23) Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F. H. (2009) A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U.S.A. 106*, 5610−5615.