DISSERTATION

ENGINEERING STABILIZED ENZYMES VIA COMPUTATIONAL DESIGN AND IMMOBILIZATION

Submitted by

Lucas B. Johnson

Department of Chemical and Biological Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2016

Doctoral Committee:

Advisor: Christopher Snow

Kenneth Reardon
Christie Peebles
Olve Pearson

ABSTRACT

ENGINEERING STABILIZED ENZYMES VIA COMPUTATIONAL DESIGN AND IMMOBILIZATION

The realm of biocatalysis has significantly matured beyond ancient fermentation techniques to accommodate the demand for modern day products. Enzymatically produced goods already influence our daily lives, from sweeteners and laundry detergent to blood pressure medication and antibiotics. Protein engineering has been a major driving force behind this biorevolution, yielding catalysts that can transform non-native substrates and withstand harsh industrial conditions. Although successful in many regards, computational design efforts are still limited by the crude approximations employed in searching a complex energy landscape. Advancements in protein engineering methods will be necessary to develop our understanding of biomolecules and accelerate the next generation of biotechnology applications.

Our work employs a combination of computational design and simulation to achieve improved enzyme stability. In the first example, an enzyme used in the production of cellulosic biofuels was redesigned to remain active at high temperature. An initial approach involving consensus sequence analysis, predicted point mutation energy, and combinatorial optimization resulted in a sequence with reduced stability and activity. However, by using recombination methods and molecular dynamics simulations, we were able to identify specific mutations that had a stabilizing or destabilizing effect, and we successfully isolated mutations that benefited enzyme stability. Our iterative approach demonstrated how common design failures could be

overcome by careful interpretation and suggested methods for improving future computational design efforts.

In the second example, a cellulase was designed to have a high net charge via selected surface mutagenesis. “Supercharged” cellulases were experimentally characterized in various ionic liquids to assess the effect of high ion concentration on enzyme stability and activity. The designed enzymes also provided an opportunity to systematically probe the protein-solvent interface. Molecular dynamics simulations showed how ions influenced protein behavior by inducing minor unfolding events or by physically blocking the active site. Contradictory to previous reports, charged mutations only appeared to alter the affinity of anions and did not significantly change the binding of cations at the protein surface. Understanding the different modes of enzyme inactivation could motivate targeted design strategies for engineering protein resilience in ionic solvents.

In addition to the discussed computational design methods, immobilization strategies were identified for capturing enzymes within porous protein crystals. Immobilization offers a generic approach for improving enzyme stability and activity. Our preliminary studies involving horseradish peroxidase and other enzymes suggested protein scaffolds could be employed as an effective immobilization material. Co-immobilizing multiple enzymes within the porous material led to improved product yield via exclusion of off-pathway reactions. Although future studies will be required to assess the potential capabilities of this immobilization strategy in comparison to other materials, preliminary results suggest protein crystals offer a favorable, controlled environment for immobilizing enzymes.

The diversity of approaches presented in this thesis emphasizes that there are many options for engineering enzyme stability. Extending the lessons learned from our cellulase

engineering to the greater field of rational protein design promotes the concept of biomolecules as designable entities. By establishing the shortcomings of our designs and suggesting routes for improvement, we anticipate our design methods and immobilization strategies will procure continued interest from the biotechnology community. The toolsets we developed for cellulases can be directly transferred to other enzymes and have the potential to impact a range of protein engineering applications.

ACKNOWLEDGEMENTS

Reflecting on the last five years of my graduate student life has made me eternally grateful for the path that has led me to this point. When my undergraduate mentors Dr. Michael Misovich and Dr. Graham Peaslee encouraged me to begin this process, I had no idea what lay ahead. Fortunately, the guidance and support of many people at Colorado State University has allowed me to learn and develop as a student. Although this thesis carries my name, I attribute the culmination of my work to the support of others.

First and foremost, I would like to thank my advisor, Dr. Christopher Snow, for continuously teaching and encouraging me. Taking on a young student with no prior computational or biology experience for a *computational biology* project was more than a leap of faith. However, with his guidance, I have grown to appreciate an entirely new field. His dedication to his work and his enthusiasm for research provide inspiration for students and faculty alike.

I gratefully acknowledge the guidance and support provided by my committee members Dr. Kenneth Reardon, Dr. Christie Peebles, and Dr. Olve Peersen. Their teaching both in the classroom and in the lab has helped me to grow and develop as a scientist.

Many thanks are owed to my peer, Thaddaus Huber. His knowledge and expertise has constantly influenced my teaching, writing, presenting, and experimental planning. As the first students in a newly established lab, we encountered countless daily struggles together and were forced to resolve issues as a team. Sorry Tad, but the next time the -80^{o}C freezer goes out, you will have to find someone else to call at midnight.

I thank all of the past and present graduate students in the Snow lab, including Mark Lunt, Jacob Sebesta, Ann Kowalski, and Luke Hartje, for their assistance with experimental projects and their review and feedback regarding my work. All have worked together as a team, and I have enjoyed their shared company both in lab and in life.

I have been fortunate to work with several talented undergraduate students, and their many hours spent in lab have generated most of the data presented in this document. Many thanks to: Luke Minardi, for his time spent in our lab as an REU; Lucas Gintner, for learning how to properly prepare media and for skillfully executing many molecular biology techniques; Sehoo Park, for his years of hard work and dedication to our lab; and Karina Appel, Jarad Yost, Eli McPherson, Carolyn Keating, Stacey Zintgraff, Brisco Arechederra, and the many other undergraduate students who have either directly or indirectly influenced this work.

I would like to acknowledge the collaborators at CSU and elsewhere who have helped me appreciate new ideas and gain new skills. Thanks are due to: Dr. Steve Albers and Ethan Dusto, for establishing a joint high school teaching module; Dr. Brian Heinze, Dr. Devon Osbourne, and all of the employees at OptiEnz Sensors for teaching me what it means to work in a biotechnology company; Dr. Jacob Meyers, for providing an opportunity to practice my skills in an entirely new field; Dr. Erik Nordwald, for his helpful questions and discussions regarding cellulase engineering; Lei Wang and members of Dr. Chuck Henry's lab, for assisting with the fabrication of microfluidic devices.

Numerous colleagues have helped me survive the engineering coursework and offered me advice on research. I appreciate the help and support of Dr. Mona Mirsiaghi, Dr. Steve Albers, Dr. Yi Ern Cheah, Dr. Justin Sweeley, Dr. Todd Zurlinden, and all of the students who have shared in my graduate school experience.

I am also grateful for the faculty and staff in our building and across the department who have provided support, including Claire Lavelle, Barbara Gibson, Tim Gonzalez, Denise Morgan, Ellen Brennan-Pierce, and Marilyn Gross.

I thank the Colorado Center for Biorefining and Bioproducts (C2B2) and the Sustainable Bioenergy Development Center for providing financial support.

Last but not least, I would like to thank my friends and family for providing the love and encouragement needed to make it through all of the challenges of graduate school. Without the support of my wife, Nicole, I would not be where I am today.

DEDICATION

For my mother, who would be proud to see this degree completed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. ENZYMES AS USEFUL INDUSTRIAL CATALYSTS

1.1. PROTEINS: MORE THAN WHAT YOU EAT

People often consider proteins from a macroscopic perspective. For example, proteins might be broadly viewed as a component of food that is necessary for human health. When viewed at the molecular scale, however, it becomes readily apparent that proteins are complex macromolecules. Each protein is made up of hundreds or even thousands of amino acid building blocks, each of which forms distinct chemical interactions to give the protein its unique structure and function.

As we learn more about proteins and how they interact with their surroundings, it is only natural that we begin to consider how to harness their useful properties. Enzymes, or proteins that catalyze chemical reactions, can aid in the synthesis of countless chemicals and small molecules. The commercial use of enzymes has become more diverse[1] and widespread over the last few decades, and the market for specialty enzymes is expected to exceed $4 billion in the near future.[2] Enzymatically produced goods already influence our daily lives, from food and beverages to advanced medicines.[3] A review from one of the leading enzyme producers, Novozymes, demonstrates the diverse range of potential applications, including dishwasher detergent, sweeteners, paper processing, and textile treatment.[4] Many Novozymes products contain cellulases, or specialized enzymes capable of depolymerizing cellulose chains. Laundry detergents commonly use protease and cellulase mixtures in place of phosphates to create a more environmentally friendly cleaning solvent. Cellulases are also used to treat textiles and give clothes a “stonewashed” appearance. Bioenergy is one of the largest potential markets for cellulases: enzyme mixtures can be used to perform a key hydrolysis step in the production of

biofuels. In all of these cases, enzymes provide a commercial benefit over traditional chemical processes by reducing waste, increasing efficiency, improving selectivity, or minimizing environmental impact.

1.2. PROTEIN STABILITY

Despite having many potential applications, the utility of enzymes is often limited by their native stability. The word 'stability' is an all-encompassing term that can relate to enzyme fitness in a number of ways.[5] The thermodynamic stability of an enzyme depends upon the difference in free energy between the folded and unfolded states (Figure 1.1). Numerous driving forces, such as the hydrophobic effect or the formation of hydrogen bonds,[6] minimize the free energy of a protein relative to its unfolded state. Although thermodynamic stability can be measured using calorimetry or other experimental approaches, free energy values can not be determined on an absolute scale and must be reported as the difference between folded and unfolded states (ΔG).

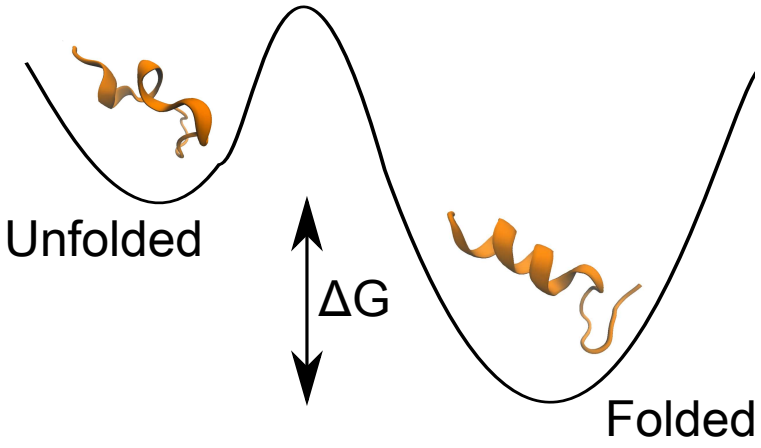

FIGURE 1.1: THERMODYNAMIC STABILITY
Hypothetical two-state energy model showing the difference in free energy between the folded and unfolded states of a protein.

Alternatively, protein stability can be defined based on tolerance to high temperatures. The thermostability of an enzyme indicates the range of temperatures over which it remains folded or active. Thermostability can be expressed using a variety of metrics, most of which are related on a temperature scale. For example, the melting temperature of a protein (T_m) defines the temperature at which half of the enzyme is reversibly unfolded.

Stability can also be defined based on unfolding rates or retained activity over time. Kinetic stability parameters might be reported as the preincubation temperature that decreases enzyme activity by 50% over a set period of time (T_{50}) or as the observed rate of inactivation ($k_{d,obs}$). Additional metrics based on enzyme lifetime or optimum operating temperature could also be used to indicate kinetic stability.

Chemical denaturants, such as urea or organic solvents, can influence enzyme stability as well. The denaturant concentration that unfolds half of the protein molecules in solution ($C_{1/2}$) provides an analogous chemical comparison to the thermal melting point (T_m). In broad terms, any change in solution conditions, including pH or salt concentration, can be used to empirically define the stability of an enzyme. In these cases, stability might be generically reported in terms of optimum activity or relative changes in activity.

On a molecular level, loss of enzyme activity can be caused by a variety of interactions: active site inhibition, modification of catalytic residues, or global protein denaturation, to name a few. Small perturbations in secondary structure can also cause large changes in activity by initiating protein aggregation or other irreversible unfolding processes. Correlating these molecular-level interactions with changes in macroscopic properties can be exceedingly difficult but is necessary to facilitate rational protein engineering.

1.3. PROTEIN ENGINEERING

Most enzymes are adapted to perform best near physiological conditions. High temperatures, organic solvents, or extreme pH conditions encountered in industrial processes can quickly denature or inactivate enzymes. To expand the scope and potential usefulness of enzymes as industrial biocatalysts, they must be engineered or selectively modified to perform in non-native conditions. Fortunately, protein engineering has made significant progress in recent years, and there are now a number of effective strategies for improving enzyme stability.[7]

1.3.1. IMMOBILIZATION, CROSSLINKING, AND CHEMICAL MODIFICATION

Chemical modification was one of the first approaches used for altering a protein's properties.[8] Introducing certain chemical groups on the surface of a protein has been shown to improve solubility in a variety of solvents. PEGylation, or the attachment of polyethylene glycol to the protein surface, can enhance enzyme solubility in organic solvents.[9,10] Other amphiphilic modifiers have improved enzyme tolerance to organic solvents as well, as demonstrated with polyoxyethylene lauryl ether modification of catalase.[11] Changing the surface charge of a protein via acetylation or succinylation has also been shown to improve enzyme performance in ionic liquids.[12–14] Although chemical modifications are intended to target specific functional groups on the protein surface, it can be difficult to control the exact modification sites. Undesired changes near the active site of the enzyme can result in reduced activity.

Immobilization is another common approach for stabilizing enzymes. Attachment to a scaffold or secondary material can help preserve enzymatic activity at high temperatures, in organic solvents, at extreme pH, or even in proteolytic conditions.[15] Numerous other benefits make immobilization favorable for industrial processes. Immobilized enzymes can be recovered

and reused in multiple reactions. Removing the catalyst can result in a more pure product and reduce the need for downstream separations. Recycling immobilized enzymes can lead to higher productivity and reduce operating costs. In addition, multiple enzymes can be immobilized on the same matrix to reduce diffusion limitations, promote substrate channeling, and facilitate coenzyme recycling.

There are many commercial processes that rely on enzyme immobilization as a stabilization strategy. For example, aminoacylase from *Aspergillus oryzae* was immobilized on a modified cellulose matrix and used in the production of L-amino acids via the Tanabe process.[15] Immobilization of penicillin amidase on a polymer matrix facilitated enzyme usage in the penicillin production process.[16] The immobilized amidase retained >60% activity even after 800 reuses. Similarly, lipase from *Candida antarctica* has been immobilized on a variety of textile-based materials or acrylic resins for use in biodiesel production.[17]

The type of material used for immobilization can vary tremendously. Historically, many types of natural materials were used for immobilization. Functionalized biopolymers such as cellulose, starch, agarose, or chitosan can provide useful immobilization scaffolds.[15] More recently, synthetic organic polymers or porous acrylic resins have gained in popularity. Hydrogels formed from polyvinyl acohol or other copolymers with an amphiphilic network can also be used for encapsulation. Some of the cheapest supports are made from inorganic particles, including alumina, silica, or zeolites.[15] Immobilization on inorganic nanoparticles or other types of metals has the added benefit of facilitating magnetic separation[18] without requiring filtration or centrifugation steps.

Even in the absence of a carrier material, bifunctional crosslinking agents such as glutaraldehyde can be used to covalently link protein molecules and increase particle size.

Crosslinking in the presence of albumin or other proteins can create a gelatinous matrix that tethers enzymes. Alternatively, enzymes can be crosslinked directly to form crosslinked enzyme aggregates[19] or crosslinked enzyme crystals.[20] These techniques can yield particles with high activity per unit mass and have been used in many industrial applications.[21] Furthermore, the microporous nature of crosslinked crystals allows them to be used as nanoscale filtration devices for macromolecules.[20]

The drive to specifically localize enzymes with nanometer-scale precision has led to a variety of biomolecular immobilization scaffolds. Attaching enzymes to DNA,[22] viruses, or protein cages[23] can create nanoscale reactors for sensing, imaging, or drug delivery applications. Porous protein crystals, which will be discussed further in Chapter 5, offer an alternative for localizing macromolecules within a biological matrix.

The diverse range of available immobilization strategies provides many options for stabilizing enzymes. Unfortunately, no single immobilization strategy is perfect. Immobilization often causes reduced specific activity due to restricted enzyme dynamics or diffusion limitations. Behavior can also vary from protein to protein, necessitating customized protocols for each enzyme. Although immobilization has proven successful in many applications, alternative design strategies will continue to play an important role in enhancing enzyme stability and retaining enzyme activity.

1.3.2. BIOPROSPECTING AND DISCOVERY

Sometimes the discovery of naturally stable enzymes can yield industrially worthy catalysts.[24,25] Organisms that have adapted to harsh conditions, such as geothermal hot pools[26] or saline environments,[27] rely upon extremely stable cellular components to support their growth.

Thermophilic organisms have long been sought as sources of stable enzymes. For example, a number of endoglucanases have been isolated from extremophiles,[28] including one that maintains activity above 100°C.[29] Xylanases, dehydrogenases, and other enzymes with temperature optimums above 100°C have also been isolated from thermophiles.[30]

As the genomes of extremophilic organisms become more readily available, the identification of genes through metagenomic screening will become more feasible.[31,32] Computational screening can be used to identify previously uncharted genes and reduce time spent in lab on heterologous expression and characterization. Comparisons between mesophilic enzymes and their thermophilic counterparts provides further insight regarding common stabilization mechanisms.[33–37] In general, however, bioprospecting is not a systematic design tool, and widespread usage is often limited by the need for the fortuitous discovery of different enzymes.

1.3.3. DIRECTED EVOLUTION

Although natural evolution is too slow to be used as a design tool, a similar accelerated process can be used to adapt proteins in the laboratory. The concept of directed evolution was popularized by Frances Arnold[38] and Willem Stemmer[39] in the 1990s and is now commonly used for many types of proteins.[40,41] Because no *a priori* knowledge is required, directed evolution techniques can be used to modify almost any gene. At the start of the process, mutations are randomly introduced via error prone polymerase chain reaction (PCR), gene shuffling, or site saturation mutagenesis.[40] The selection of an appropriate method usually relies upon the desired level of genetic diversity and the available screening capacity. To assess the mutated variants, genes are transcribed and translated to form proteins. *In vivo* screening via yeast or phage

display[42] can drastically increase screening capacity, but most enzymes are assessed *in vitro* using a multi-well plate format.[40] Enzyme thermostability is one of the simplest fitness challenges to screen for: measuring retained activity after exposure to high temperature can quickly identify improved variants. However, the danger of a single-criterion selection processes is that properties not being screened for can be unfavorably changed; screening for thermostability alone, for example, might result in reduced activity at low temperatures or altered enzyme selectivity. If improved variants are identified in the first round of directed evolution, an iterative process can be employed to consecutively accumulate mutations until the desired design goal is achieved.

Directed evolution has been successful in improving enzyme stability in a number of cases. One of the most impressive examples is the modification of a xylanase by site saturation mutagenesis: nine mutations achieved a stability benefit of more than 35^{o}C, increasing T_m from 61^{o}C to 96^{o}C.[43] Another example involving β-lactamase evolution via error-prone PCR and gene shuffling led to an increase in temperature optimum from 35^{o}C to 65^{o}C.[44] Recombination of homologous genes in a study of subtilisins from *Bacillus* species led to improved activity under high pH conditions, at high temperature, and in organic solvents.[45] Performing multiple screening steps identified variants with the best combination of properties. Error-prone PCR has also been effective in improving enzyme stability in ionic liquids. The fitness of a dehydrogenase from *Candida boidinii* was improved in 1,3-dimethylimdazolium dimethyl phosphate by introducing mutations at the dimer interface and near the active site.[46]

The power of directed evolution makes it a popular choice for engineering stabilized enzymes. However, the degree of success is most often determined by the availability of resources: screening large libraries takes both time and money. Reducing library size using

targeted mutagenesis strategies or rational design can increase screening efficiency and lead to improved success rates.

1.3.4. RATIONAL DESIGN

Unlike directed evolution, rational design methods rely upon structural models and energy functions to predict protein stability. Numerous types of interactions can be formed between neighboring amino acids and with the surrounding solvent. Van der Waals' interactions, ionic interactions, hydrogen bonds, pi-pi stacking, bond stretching, torsion potential, and water entropy are some of the terms that determine protein stability,[47] many of which are incorporated in computational energy functions.[48]

Targeted design strategies often capitalize on local interactions to improve enzyme stability. The formation of disulfide bonds, salt bridges, helix capping, and entropy reduction are all common strategies for selectively engineering improved stability.[47] In one example of semi-rational protein engineering, the combination of designed and homologous mutations successfully increased the optimum temperature of a protease from 74^{o}C to 95^{o}C.[49] Multiple design iterations similarly increased T_m from 83^{o}C to 106^{o}C for an α-amylase enzyme.[50] Although impressive, these drastic changes might present a skewed perspective on the potential of targeted design strategies. Large stability impacts are only realized if the mutation sites play an important role in the protein unfolding process. Identifying these key sites can be difficult, and researchers are often forced to use less specific approaches where mutations are dispersed across the protein sequence.

Delocalized design strategies can incorporate mutations at the protein core, on the protein surface, at protein interfaces, or across the whole protein sequence. The challenge with this

approach is identifying an appropriate design search space. Complete sampling of all possible combinations is only feasible for short peptide sequences. For a typical protein containing ~300 amino acids, the search space grows to greater than 10^{390} possibilities. Information from previous experiments or from consensus sequence alignments[51,52] can be incorporated into the design process to reduce the number of candidate mutations per site and create a search space of a more tractable size.

The dynamic nature of proteins further complicates design efforts. Proteins are constantly moving and undergoing small structural changes in solution. Capturing this behavior *in silico* can be difficult. It is not possible to search all possible protein conformations, so protein engineers must rely upon reasonable approximations. Current computational protein design strategies can be classified into three main categories based on the types of approximations employed: traditional methods limit movement to discrete rotameric positions for amino acid side chains, flexible backbone methods extend the limits of dynamic behavior to include perturbations in secondary structure, and molecular dynamics (MD) simulations provide a full representation of protein dynamics based on Newton's equations of motion (Figure 1.2).

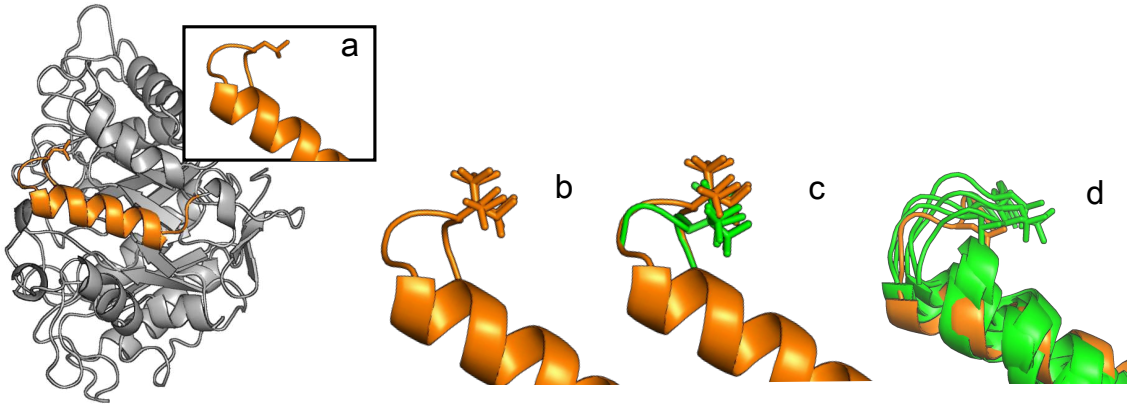

FIGURE 1.2: APPROXIMATIONS IN COMPUTATIONAL PROTEIN DESIGN STRATEGIES
(a) Cartoon depiction of a protein structure with a single leucine amino acid shown as sticks. **(b)** Discrete rotameric positions employed in a traditional computational design search space. **(c)** Backbone flexibility introduced via loop closure algorithms. **(d)** Full range of motion provided by molecular dynamics simulations.

Steve Mayo and coworkers first implemented traditional 'rotameric' design strategies in the late 1990s.[53,54] Sequences were threaded onto an existing protein structure and amino acid side chains were allowed to adopt discrete conformations based on a rotamer library.[55] Early attempts to improve protein stability included modification of a small hyperthermostable Streptococcal protein,[54] design of globular proteins with increased T_m values,[56] and core repacking of spectrin peptides.[57] Redesign of a growth hormone and cytokine analog in the early 2000s achieved thermostability increases of 16^{o}C and 13^{o}C, respectively.[58,59] One of the first computational designs targeting an enzyme (yeast cytosine deaminase) improved the apparent melting temperature by 10^{o}C without reducing catalytic efficiency.[60]

David Baker and others have made modifications to the original rotameric search strategy to design more than just protein stability. Examples involving the design of new catalytic activity[61,62] or even the design of a novel protein fold[63] demonstrate the potential of these computational approaches. Incorporating additional scoring criteria, such as the quality of core packing, can further improve the accuracy of design methods.[64] However, computational approaches are still limited by the accuracy of energy function predictions and the extent of sampling. In several published cases (and likely many unpublished cases), traditional rotameric computational design methods have failed to identify sequences with improved stability.[65,66]

To improve the success rate of computational design, some researchers have implemented algorithms that allow flexibility in the protein backbone. Templates can be generated using nuclear magnetic resonance data, homologous crystal structures, normal mode analysis, loop closure algorithms or 'backrub' strategies, snapshots from molecular dynamics simulations, or other strategies that introduce variability in the protein backbone.[67,68] Incorporating backbone flexibility has been shown to improve stability predictions relative to

traditional fixed backbone design algorithms.[69] Murphy and coworkers used flexible backbone design to modify a small four-helical peptide and successfully increased T_m from 91°C to 142°C.[70] Unfortunately, the use of flexible backbone design is generally limited to small proteins; adding degrees of freedom for backbone flexibility quickly increases the size of the search space and can cause the corresponding computational expense to become prohibitive.

An extreme example of computational expense is realized in molecular dynamics simulations. MD simulations provide the most realistic representation of protein behavior, but simulation timescales are limited to the nanosecond or microsecond range for most computational resources. Sampling many sequences from a design search space using MD is prohibitively expensive; therefore, MD is typically used as a screening or inspection tool rather than as combinatorial design tool. Identification of dynamic domains in a target protein can indicate regions where disulfide bonds or other stabilizing features might be beneficial. Gao *et al.* used this approach to introduce mutations in the most dynamic domain of cocaine esterase and improved the melting temperature by ~7°C.[71] A comparable study by Janssen and coworkers designed a disulfide bond in haloalkane dehalogenase that increased protein thermostability by ~5°C.[72] By using an iterative design method incorporating computational energy predictions, traditional computational design methods, and MD screening steps, Janssen and coworkers increased the thermostability of a hydrolase from ~50°C to ~85°C.[73] This impressive change in protein stability (an increase of ~35°C) demonstrates the potential benefit of combining multiple design strategies.

1.3.5. COMBINED APPROACHES

Comparisons between multiple design strategies suggest that no single approach is best. Even though they are divergent strategies and often yield different final solutions, directed evolution and computational design can achieve similar results.[74] In practice, strategies that combine multiple engineering techniques can be most successful. Directed evolution is frequently used to mutate computational designs and further improve enzymatic activity or stability.[75–77] Proteins can even be evolved to be better suited for immobilization.[78] Other combined approaches use naturally thermostable sequences as starting templates for directed evolution or computational studies.[79] Designed or evolved enzymes can also be immobilized to provide an additional stability benefit or to facilitate enzyme reuse/recycling.

An entirely new class of computational design methods has evolved to complement directed evolution efforts. Often referred to as library design, combinatorial design, or data-driven design,[80,81] these approaches seek to minimize library size by targeting key sites or by avoiding unfavorable interactions. Structural models or sequence analysis can provide useful guidelines for narrowing a design search space and minimizing time spent on evolving or screening a target protein. Depending on the design goal, library-based methods typically fall into two categories: localized or delocalized design.

Localized design strategies focus on a specific region within a protein. This can be useful for design goals related to substrate specificity, coenzyme binding, or protein interfaces. In these scenarios, it makes sense to target the specific site of interest rather than randomly introducing mutations at allosteric positions. Structural analysis is first used to identify key mutation sites, and then computational tools are employed to rationally select the most suitable degenerate codons.[81–83] While the common *NNK* codon provides full sequence diversity by encoding all 20

amino acids, selecting a biased degenerate codon favors a subset of mutations that are best matched for each specific design site. Picking appropriate degenerate codons reduces library size and can minimize time spent on screening.

Delocalized properties, such as protein stability or solubility, can be designed using consensus sequence analysis or structure-guided recombination approaches. Consensus-based methods are a semi-rational version of saturation mutagenesis: comparing sequences from multiple homologous proteins identifies common mutations that are likely to be stabilizing.[52] This approach can yield a restricted set of candidate mutations that are distributed across multiple sites. In contrast, recombination-based approaches are analogous to gene shuffling. By utilizing structural information, candidate block boundaries or recombination sites are optimized for chimera complementarity. An algorithm called SCHEMA has demonstrated the potential of this approach. SCHEMA libraries have been designed for β-lactamase,[84] cytochrome P450,[85,86] arginase,[87] β-glucosidase,[88] and several families of cellulases[89–94] with some success.

With so many tools at the disposal of protein engineers, selecting an appropriate strategy can be a daunting task. Balancing tradeoffs between risk and reward, time and money, or logic and perseverance can determine the path and ultimate success of a protein engineering project. It should be reiterated that no single strategy is always optimal: protein engineers must be willing to consider multiple approaches in parallel or in combination to achieve a desired design goal. The numerous examples cited here demonstrate the power of modern design tools and provide motivation for a researcher who is seeking to engineer a stabilized enzyme.

1.4. TARGET SELECTION

To begin any engineering problem, a starting template or design target must be selected. In many industrial applications, the target enzyme has already been well defined, or, at the very least, the engineer has been given a specific design goal. In an academic environment, design goals can be more ambiguous. Generic stabilization strategies could be applied to any enzyme, and the tradeoff between the conveniences of simple model proteins must be balanced against the usefulness of more industrially relevant proteins.

For most of the design strategies discussed in this thesis, endoglucanase E1 from *Acidothermus cellulolyticus* was used as the starting template. Although E1 is a moderate size protein (~40.2 kDa, 358 amino acids) with a complex α/β_8-barrel fold,[95,96] it has been well studied and is considered by many as a 'gold-standard' of cellulases. E1 was originally isolated and characterized from *A. cellulolyticus* in the late 1980s,[97,98] and the full genome of this species has since been sequenced.[99] E1 has been expressed in bacteria,[100,101] yeast,[102] and numerous plant species,[103] demonstrating its robustness and potential transferability between expression platforms. It is a naturally thermostable enzyme with a T_m of ~84^oC.[104] High native stability provides both a favorable starting template and a challenging design goal. Previous computational studies have made comparisons between the stability of E1 and other homologues within the Cel5A family.[105,106] Other studies have investigated the energetics of substrate binding[107] and catalysis.[108] E1 is known to have high activity on crystalline cellulose and can act synergistically with other enzymes.[109,110]

Since structural information is a prerequisite for rational design, it was important to select a target enzyme with available structural information. The structure of E1 has been determined both with[111] and without[101] bound substrate analogues, and numerous other homologues have

been structurally characterized (Figure 1.3, Table 1.1). Mutational studies provided additional insight regarding key mutation sites and were used to further guide our design efforts.[112,113]

TABLE 1.1: STRUCTURAL INFORMATION FROM E1 HOMOLOGUES

Species	PDB Code(s)	Sequence Identity[a]	T_{opt}[b] (^oC)	Additional Information	References
Acidothermus cellulolyticus	1ece, 1vrx	99-100% (100%)	81-83	Mutagenesis studies, reduced product inhibition, bound substrate, QM	97-113
Pyrococcus horikoshii	2zum, 3axx, 3qho, 3qhm, 3qhn, 4dm1, 4dm2	45-46% (75%)	95-100	Binding domain studies, bound substrate, disulfide bond engineering, thermostability studies	114–120
Bacillus agaradhaerens	3a3h, 4a3h, 5a3h, 6a3h, 7a3h	29% (49%)	37-40	Bound substrate, reaction coordinate	121,122
Bacillus sp.	1lf1	29% (47%)	40-45	Mechanistic studies	123–125
Thermoascus aurantius	1gzj, 1h1n	25-27% (43%)	70-75		126–128
Erwinia chrysanthemi	1egz	26% (59%)	31-37	Mutagenesis, secretion pathway	129,130
Podospora anserina	3ziz, 3zm8	26% (72%)	NA	Mannanase, structural info on CBD	131
Trichoderm reesei (Hypocrea jecorina)	1qno, 1qnp, 1qnr, 1qns, 3qr3	23% (75%)	NA	Mannanase/cellulase	132–134
Thermobifida fusca	2cks	21% (89%)	70-75	Bound substrate	Unpublished
Clostridium thermocellum	1cec, 4im4	21% (61%)	55-70	Diverse activity studies	135
Clostridium cellulovorans	3ndy	20% (68%)	NA	Structural info on CBD	Unpublished
Thermotoga maritima	3amg, 3aof, 3azr, 3mmu	20% (52%)	95	Multiple bound substrates, mutagenesis studies, CBD engineering, gene expression	136–140
Fervidobacterium nodosum	3rjy	19% (79%)	80	Bound substrate	141,142

Abbreviations: Not available (NA), quantum mechanics (QM), molecular dynamics (MD), cellulose binding domain (CBD), Protein Data Bank (PDB).
[a]Percent identity (query coverage) based on Basic Local Alignment Search Tool (blast.ncbi.nlm.nih.gov).
[b]Temperature ranges based on reported values in Badieyan *et al.*[105] and Yeoman *et al.*[143]

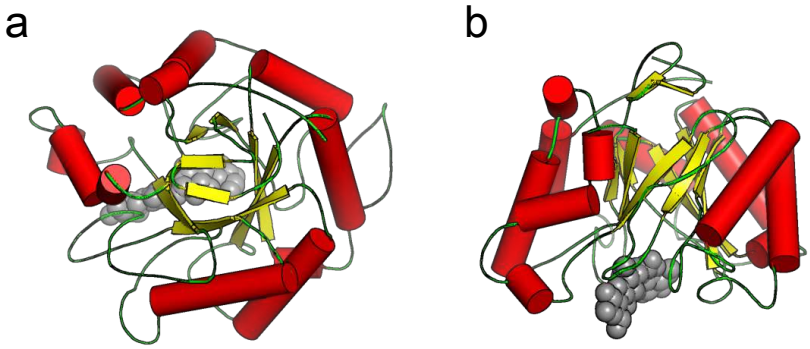

FIGURE 1.3: CARTOON DEPICTION OF ENDOGLUCANASE E1
(a) The top view and **(b)** side view of E1 show the helices (red), β-sheets (yellow), and loops (green) that make up the α/β_8-barrel fold. Bound cellotetraose substrate analogue is shown as spheres (grey) to indicate the binding pocket. Structure is based on Protein Data Bank structure 1ECE.

1.5. THESIS OVERVIEW

Despite the many significant advances that have been made in protein engineering, computational design methods are still far from perfect. Most often, a library of candidate designs is assessed with the hope that at least one design will be successful. Although this approach can yield impressive results, it does little to advance the reliability of design methods. Understanding both successful and failed designs will be necessary to improve current methods. In **Chapter 2**, an approach is presented for dissecting and interpreting a computational design. The design is broken into pieces and recombined with the starting template to assess the impact of small groups of mutations. This recombination approach identified regions of the design that were severely destabilizing or nearly neutral. By analyzing molecular dynamics simulations, single mutations were isolated that had a negative, positive, or neutral impact on stability. Combining two favorable mutations resulted in a protein sequence with higher stability than the original design or the starting template.

An alternative approach for designing proteins is to target the interface between the protein surface and the surrounding solvent. Aqueous ionic liquid solvents can have dramatic6

effects on protein stability and solubility, yet little is understood regarding their molecular interactions with proteins. In **Chapter 3**, we experimentally characterize the activity and stability of E1 in a wide array of aqueous ionic liquids. Computationally designed enzymes with high net surface charge were used to systematically probe the effects of electrostatic interactions. In **Chapter 4**, the same designed enzymes were simulated using molecular dynamics to provide insight regarding potential inactivation mechanisms. Cation and anion interactions at the protein surface, within the protein core, and near the binding pocket provided candidate explanations for why reduced activity was seen in enzymatic assays.

In **Chapter 5**, the focus of the thesis shifts to a different engineering strategy: immobilization. A new protein-based material is assessed for its ability to immobilize proteins of varying size. Enzymes were immobilized using non-covalent interactions and physical barriers. Horseradish peroxidase and other enzymes maintained activity inside the crystal environment. Confocal microscopy confirmed that enzymes, small molecule substrates, and fluorescent products could diffuse throughout the porous scaffold. Furthermore, two-enzyme pathways could detect glucose and appeared to be isolated from off-pathway enzymes in the surrounding solution. The control and precision of porous protein crystals makes them a promising immobilization material for future mesoscale applications.

2. DISCRIMINATING BETWEEN STABILIZING AND DESTABILIZING PROTEIN DESIGN MUTATIONS

Lucas B. Johnson,[a] Lucas P. Gintner,[a] Sehoo Park,[a] Christopher D. Snow[a]

This chapter is derived in part from an article published in *Protein Engineering Design and Selection* on June 15, 2015, and is reproduced here with permission (Appendix A).

[a]Author affiliation:

Colorado State University

1370 Campus Delivery

Fort Collins, CO 80523-1370

USA

*Corresponding author contact information:

Christopher.Snow@ColoState.edu

Phone: (970) 491-5276

Fax: (970) 491-7369

2.1. SUMMARY

Accuracy of current computational protein design (CPD) methods is limited by inherent approximations in energy potentials and sampling. These limitations are often used to qualitatively explain design failures; however, relatively few studies provide specific examples or quantitative details that can be used to improve future CPD methods. Expanding the design method to include a library of sequences provides data that is well suited for discriminating between stabilizing and destabilizing design elements. Using thermophilic endoglucanase E1 from *Acidothermus cellulolyticus* as a model enzyme, we computationally designed a sequence with 60 mutations. The design sequence was rationally divided into structural blocks and recombined with the wild type sequence. Resulting chimeras were assessed for activity and thermostability. Surprisingly, unlike previous chimera libraries, regression analysis based on one and two body effects was not sufficient for predicting chimera stability. Analysis of molecular dynamics simulations proved helpful in distinguishing stabilizing and destabilizing mutations. Reverting to the wild type amino acid at destabilized sites partially regained design stability, and introducing predicted stabilizing mutations in wild type E1 significantly enhanced thermostability. The ability to isolate stabilizing and destabilizing elements in computational design offers an opportunity to interpret previous design failures and improve future CPD methods.

2.2. INTRODUCTION

The field of computational protein design (CPD) has seen an amazing array of successes over the last decade, ranging from the design of novel enzymatic activity[61,62,144] to the prediction and stabilization of new protein folds.[63] However, successful designs are often members of a

larger set that includes unsuccessful variants; CPD methods are far from perfect. Demonstrating the vast potential of CPD is the first step in advancing the field, but continued effort will be necessary to achieve improved design reliability and facilitate widespread usage. Creating reliable design methods will require careful consideration of both design successes and failures. Expanding a single design sequence into a library of candidate sequences is a useful strategy for explaining design failures and identifying shortcomings in the design method.

Previous computational design methods have generated libraries using multiple starting scaffolds[145] or reaction state intermediates.[146] Although libraries based on energy scores are effective, inaccuracies in the underlying energy function can lead to biased or limited sequence diversity. In contrast, recombining homologous sequences produces a diverse and interpretable library without depending on an all-atom energy function. Site-specific recombination between blocks of homologous parent proteins has previously generated stabilized chimeric variants.[87,90,147–149] In cellobiohydrolase II, for example, recombination of three parent sequences resulted in 31 stabilized chimeras and further aided in identifying a single highly stabilizing mutation.[147,150] Hypothetically, dissecting computational designs using a similar recombination approach can produce a library that is both maximally folded and interpretable.

While library-scale recombination can indicate regions of stability or instability within a design, it can be difficult to further rationalize the impact of individual mutations within each region. Assessing computational designs in a dynamic, solvated environment provides additional hypotheses regarding the role of specific mutations. Molecular dynamics (MD) simulations capture the flexibility, repacking, and local unfolding events that are not well represented in the traditional CPD process. Assessing structural changes via MD can provide some indication whether the proposed design is likely to remain folded, identify sites where unfolding events are

most likely to occur, and generate hypotheses for why certain mutations behave differently in experiments compared to initial *in silico* predictions.

To demonstrate the potential of our method, we selected endoglucanase E1 (Cel5A) from *Acidothermus cellulolyticus* as a model enzyme. E1 was discovered and promoted by researchers at the National Renewable Energy Lab.[97,109,151] The high synergistic activity of E1 indicated prospective applicability in cellulosic fuel production,[110] and a wealth of previous research on E1 and its homologues provided both structural and mutagenesis information for guiding the design process.[101,111–113] Baker *et al.* (2005) have previously shown that surface residues near the substrate binding cleft only marginally impacted E1 activity. Nonetheless, a single mutation, Tyr245Gly, significantly reduced product inhibition.[113] For our work, selecting a naturally thermostable enzyme provided both a stable starting scaffold and a challenging thermostability design target. Achieving further improvements in cellulase thermostability has potential benefits in high-temperature process conditions, including reduced contamination, faster reaction rates, and longer enzyme lifetime.[24,152]

The organization of this paper tracks the progression of our design strategy (Figure 2.1), which consisted of five steps: (1) computational design of an initial starting sequence, (2) optimized recombination with a native sequence, (3) experimental assessment of stability and activity, (4) parameter fitting via linear regression, and (5) molecular dynamics analysis. Library-scale design assessment identified a number of mutations in the computational design that significantly impacted E1 stability. Discriminating between stabilizing and destabilizing mutations enhanced understanding of the design methodology and ultimately led to an improved E1 variant.

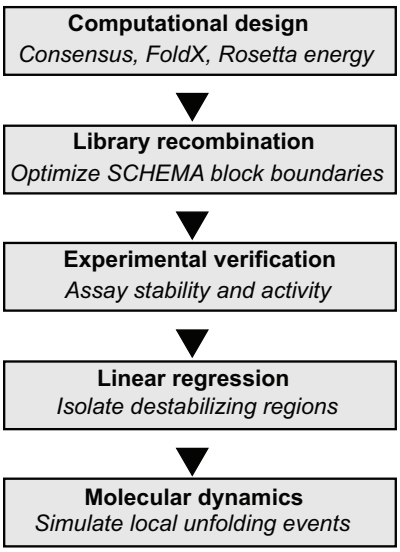

FIGURE 2.1: STEPS IN DESIGN CYCLE
Creating and dissecting a design encompasses a series of steps. First the initial design is generated and recombined with the wild type sequence to form chimeras. Experimental stability and activity values can then be used to train a regression algorithm and isolate the impact of each block. Molecular dynamics simulations provide further details regarding the predicted stability of individual point mutations.

2.3. MATERIALS AND METHODS

2.3.1. COMPUTATIONAL DESIGN

In the initial design phase, Basic Local Alignment Search Tool[153] was used to identify homologous sequences with greater than 35% sequence identity to E1. A palette of candidate mutations was designed to include the consensus (most common) amino acid at each site, then the second most common amino acid, etc. until 90% of aligned sequences were included (Figure 2.2a).

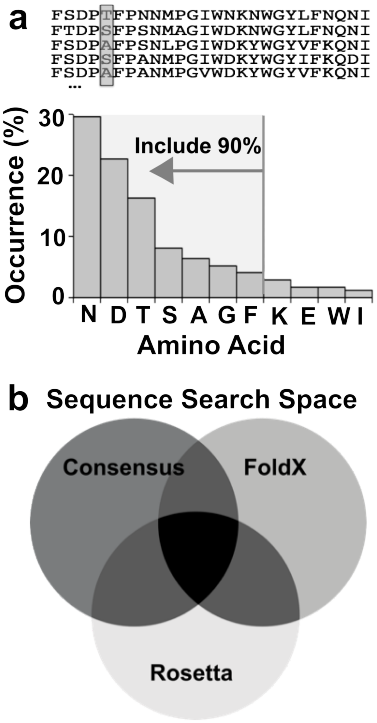

FIGURE 2.2: CONSERVATIVE DESIGN STRATEGY
The computational protein design process included **(a)** alignment of homologous sequences. At each site, 90% of sequence diversity seen in the homologues was included in the design palette. **(b)** Candidate consensus mutations were compared to predicted FoldX energy scores. A favorable set of mutations was identified from the overlapping search space using the Rosetta energy function.

Relative free energy changes (ΔΔG) were predicted for every point mutation using FoldX.[154] Calculations were performed on multiple structures of E1 (two chains each from Protein Data Bank structures 1ECE and 1VRX), and the predicted FoldX ΔΔG values were used to reduce the set of allowed mutations to those predicted to be neutral or stabilizing (ΔΔG≤0). Combinatorial optimization using the FasterPacker search algorithm as implemented in SHARPEN[155] identified a set of 62 mutations with improved Rosetta all-atom energy score.[48] Visual inspection and detailed local repacking calculations (Dunbrack rotamers ±1 standard deviation for all residues near site of interest) flagged two of the mutations, Gly221Ala and Asn228Asp, as potentially problematic. In both cases the proposed mutation was rejected due to a significant change in packing and reduced hydrogen bond energy. The design sequence had a total of 60 mutations

(83.2% sequence identity relative to wild type E1, alignment shown in Figure 2.3). Key enzymatic and structural features were held fixed during the design process, including catalytic residues Glu162 and Glu282, residues conserved within the glycoside hydrolase family 5 (Arg62, His116, Asn161, His238, Tyr240, Trp319) and disulfide cysteines (Cys34/Cys120 and Cys168/Cys171).

2.3.2. BLOCK RECOMBINATION

Block boundaries or recombination sites for the chimera library were identified using a modified version of the SCHEMA algorithm.[156] Specified parent sequences included the wild type sequence, consensus design, and two alternate preliminary designs. Each parent was divided into four blocks, creating a hypothetical library of 4^4 or 256 chimeras. Although only a subset of 16 chimeras was studied in this work, additional parent sequences were included to ensure recombination sites would be compatible with potential future designs. To facilitate library construction via sticky end overlap, a constraint was added to restrict candidate recombination sites to regions of four conserved bases, allowing alternate codons if necessary. Native contacts were defined based on intra-monomer contacts in Protein Data Bank structure 1ECE.

FIGURE 2.3: ALIGNED WILD TYPE AND DESIGNED SEQUENCES
Wild type E1 sequence (WT) and computationally design sequence (CD) aligned with conserved residues highlighted in red. Secondary structural elements are represented above the sequences: spiral indicates α-helix and arrow indicates β-sheet.

2.3.3. EXPERIMENTAL ACTIVITY AND STABILITY ASSAYS

For experimental verification, a codon-optimized gene encoding the designed sequence was synthesized via commercial gene synthesis (GenScript, New Jersey, USA). The wild type E1 gene was obtained as a kind gift from Dr. Joel Kaar (University of Colorado Boulder, Colorado, USA). Each gene was cloned into pET21_NESG expression vector using Gibson assembly.[157] Chimera block sequences were amplified from parent sequences via polymerase chain reaction, combined using Golden Gate assembly,[158] and purified using gel electrophoresis. Gel-extracted chimera genes were cloned into pET21_NESG expression plasmid using Gibson assembly. Each sequence was verified at Colorado State University core sequencing facility using T7-promoter and T7-termination primers. All cloning steps were performed in *E. coli* DH5α cells. For expression, *E. coli* BL21 (DE3) pLySs cells were grown overnight in 50 ml of Luria-Bertoni (LB) media supplemented with 100 μg/ml ampicillin. Large-batch expression cultures were inoculated at a 1:20 dilution ratio and grown for 2 hours at 37^{o}C and 250 rpm before inducing with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Overnight (~16 hour) induction at 30^{o}C followed by centrifugation at 3,000 rpm and sonication gave soluble yields of 2-5 mg protein per liter of culture. A C-terminal 6xHis tag allowed protein purification via Ni-NTA affinity on a HisTrap HP column (GE Healthcare) in HisTrap Buffer A (50 mM sodium acetate, 0.5 M NaCl, 100 mM imidazole, 0.1% Tween20, 10% glycerol, pH 8.0). Bound fractions were eluted using a linear gradient of HisTrap Buffer B (HisTrap Buffer A with imidazole concentration increased to 1 M). Recovered fractions were buffer exchanged into Assay Buffer (50 mM sodium acetate, pH 5.0) on HiPrep 26/10 Desalting column (GE Healthcare). Correct size and relative protein purity were verified via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Activity assays were performed on a soluble cellulose analogue, p-nitrophenyl-β-d-cellobioside (pNPC). Enzymatic cleavage of the β-1,4 linkage in pNPC released 4-nitrophenol, allowing measurement of accumulated product via absorbance at 415 nm wavelength. Specific activity was assessed in Assay Buffer using 0.5-25 μg protein (enough to generate final product absorbance between 0.1 and 1.0 AU) and 20 μg pNPC. Reactions were stopped after 30 minutes at 50^{o}C using Stop Buffer (2 M sodium acetate, 0.1 M glycine, pH 12.5). Protein concentration was measured using the Bradford Assay with albumin standards, and measurement error estimates were adjusted to account for potential impurities in purified protein samples. Temperature optimum assays were performed the same as specific activity assays, but with varying temperature. For retained activity assays, a 30-minute pre-incubation period in Assay Buffer with temperatures ranging from 40 to 80^{o}C was used to partially denature the enzyme. Retained activity was measured after an additional 30-minute, 50^{o}C incubation period with 10-25 μg pNPC. Reactions were stopped with a 2:1 ratio of Stop Buffer. An inverse sigmoidal dose-response curve was fit using SciPy (www.scipy.org) and used to predict T_{50} values (defined as the pre-incubation temperature that resulted in 50% decreased activity).

A fluorescent dye (SYPRO Orange, Sigma Aldrich #S5692) was used to monitor changes in protein structure induced by thermal denaturation. Fluorescence was measured near optimum dye excitation/emission wavelength of 470/570 nm using the VIC filter on a Realplex2 Mastercycler (Eppendorf). A linear heating gradient of 2^{o}C/min was used to irreversibly denature protein samples. Each sample contained 1-10 μg protein, 2.5 μl 125x SYPRO orange dye, and enough Assay Buffer to bring the final volume to 20 μl. The Clarke-Fersht equation[159] was fit to T_m data using SciPy. Apparent T_m values (not thermodynamic T_m due to irreversible unfolding) were defined as the temperature at which 50% of the protein was unfolded. This corresponds to

the $[D]_{50\%}$ fit parameter in the original Clarke-Fersht equation.[159] All assays were performed in triplicate, and standard deviations for T_m and T_{50} values were approximated using a bootstrap method (1000 samples with replacement).

2.3.4. REGRESSION ANALYSIS

Linear regression analysis was performed using a Python numpy linear algebra package (www.python.org). A fit parameter was included for each design block (1-body terms) or pair of blocks (2-body terms). Test sets consisted of all possible combinations of 14 chimera T_m values. Because all T_m values were subtracted from wild type T_m, fitting parameters to a test set of 14 chimeras allowed assessment of library variability and goodness of fit via prediction of the remaining chimera T_m value. Python was also used to calculate the distribution of all single block and two block substitution differences.

2.3.5. MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics (MD) simulations were performed using GROMACS.[160] Each chimera structure was initially energy minimized using FasterPacker rotamer optimization in SHARPEN. Structures were solvated using TIP3P solvent with sodium chloride ions added to a concentration of 100 mM. Equilibration included 5000 energy minimization steps each *in vaccuo* and in solvent, 10 ps of position-restrained equilibration, 2 ps MD with canonical NVT ensemble at the final simulation temperature, and 2 ps MD under fixed pressure NPT ensemble. For each chimera, production MD simulations were run for 50 ns at 298 K or 450 K with a time step of 2 fs and OPLS-AA force field.[161] Particle Mesh Ewald electrostatics with Verlet cutoff scheme, velocity rescale temperature coupling, Parrinello-Rahman pressure coupling, and periodic

boundary conditions were used for all simulations. Trajectories were analyzed using GROMACS (g_rms, g_rmsf) or custom Python scripts based on SHARPEN software.

2.4. RESULTS

2.4.1. COMPUTATIONAL DESIGN

Candidate mutations were limited to those that appeared to be favorable or "meta-consensus" neutral in both consensus sequence alignments and FoldX energy predictions. Sequence/orientation combinations were evaluated based on an all-atom Rosetta energy function (Figure 2.2b). Restricting the design space to a set of consensus mutations was intended to create a more conservative sequence search space. The Rosetta energy function was used to identify candidate low-energy sequences from within this search space. The design sequence identified with the lowest Rosetta energy score contained a total of 60 mutations relative to wild type E1 (83.2% sequence identity). The overall Rosetta Energy Unit (REU) score improved significantly from -686 REU to -736 REU. The largest differences between wild type and design energy terms were repulsive van der Waals interactions (reduced by 15 REU in the design), solvation energy (15 REU more favorable in the design), and secondary structure phi/psi dependent probability (10 REU more favorable in the design). The net hydrogen bond interactions remained nearly constant. Mutations were distributed both internally and on the protein surface (37 mutations had >20% solvent exposed surface area) and avoided key catalytic and substrate binding residues.

2.4.2. BLOCK RECOMBINATION

A SCHEMA-based algorithm was used to identify chimera block boundaries that minimally impacted native E1 structure. Wild type E1 and computational designs provided

parent sequences for defining structural contacts. Optimum crossover sites at amino acids Ile100, Pro230, and Asn274 dissected the protein into four blocks (Figure 2.4), giving a library average SCHEMA energy <E> of 2.1 and an average chimera mutation level <m> of 28.2 (Figure 2.5). The specified block boundaries led to a distribution of fifteen design mutations in Block 1, eighteen mutations in Block 2, nine mutations in Block 3, and eighteen mutations in Block 4. Dividing two parent sequences into four blocks each generated a library of 2^4 or 16 chimeras. In contrast to previous studies with much larger libraries,[85,90,148,150] our library could be comprehensively screened to maximize information available for regression analysis. Complete screening is desirable for interpreting two-body or higher order stability effects.

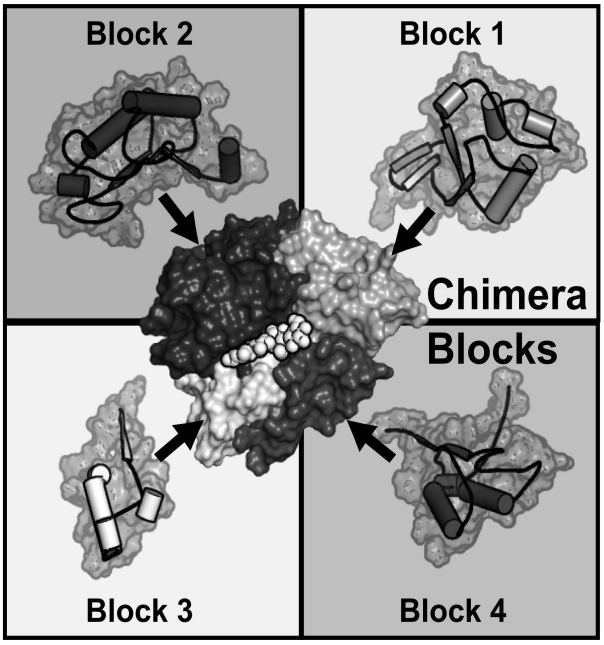

FIGURE 2.4: E1 DIVIDED INTO CHIMERA BLOCKS
A bottom view shows how E1 was divided into four blocks for recombination. Block 1 (light grey) included residues 1-100, Block 2 (black) included residues 101-230, Block 3 (white) included residues 231-274, and Bock 4 (dark grey) included residues 275-358. Bound substrate analogue cellotetraose is shown as white spheres. Coordinates are based on Protein Data Bank structure 1ECE.

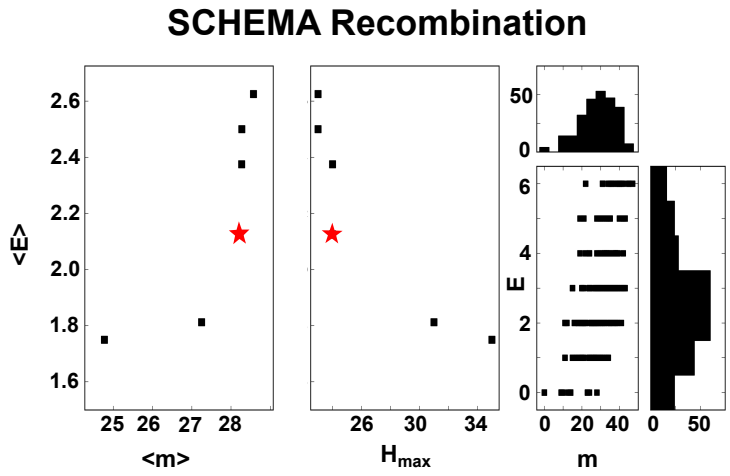

FIGURE 2.5: IDENTIFYING OPTIMUM BLOCK BOUNDARIES
Library design simultaneously optimized multiple parameters: library average SCHEMA energy <E> was minimized, library average mutation level <m> was maximized, and maximum block-block Hamming distance H_{max} was minimized. Each point on the scatter plot represents a candidate library; a red star indicates the library selected for this study. The distribution of properties for each chimera in the library is shown on the far right. The total number of chimeras is 256 (not 16) because two additional hypothetical designs were included in the library design calculations.

2.4.3. EXPERIMENTAL STABILITY AND ACTIVITY ASSAYS

Activity and stability were experimentally verified using substrate analogue p-nitrophenyl-β-d-cellobioside (pNPC) and fluorescent dye SYPRO orange (Table 2.1). The chimera naming scheme indicated whether each block was from parent 1 (wild type) or parent 2 (design). For example, 1211 denoted a chimera with blocks 1, 3, and 4 from wild type E1 and block 2 from the design sequence.

Each chimera was purified using immobilized metal ion affinity chromatography to reduce assay variability. Replicate T_m measurements were typically within 1°C (chimeras 1111, 1112, 1121, 1122, 2111, 2222); however, T_m measurements for chimeras 1221, 2112, 2122, and 2211 were more variable (up to 5°C). Notably, chimera 2212 displayed two distinct unfolding peaks at 51°C and 70°C. The lower temperature peak better matched T_m values predicted from regression analysis. Most chimeras with low specific activity had a correspondingly low melting temperature. Because T_{50} values could not be determined for some chimeras with low activity,

T_m measurements provided a more complete data set for performing regression analysis. Despite being distinct properties, stability (T_m) and retained activity (T_{50}) were correlated (R^2=0.98 for dataset of 9 points- excluding 3 outliers).

TABLE 2.1: EXPERIMENTAL VALIDATION OF ACTIVITY AND STABILITY

Chimera	T_m[a] (^oC)	T_{50}[a,b] (^oC)	Specific Activity[a] (μmol/mg/min)
1111	84.9 ± 0.4	79.0 ± 0.1	70,000 ± 1,000
1112	76.9 ± 0.5	71.3 ± 0.1	11,800 ± 100
1122	72.4 ± 0.1	61.5 ± 0.1	8,900 ± 300
1222	48.7 ± 0.1	NA	7,300 ± 200
1212	52.0 ± 0.1	NA	870 ± 80
1221	69.3 ± 0.3	65.7 ± 0.2	7370 ± 90
1211	53.6 ± 0.1	NA	150 ± 10
1121	80.1 ± 0.1	73.3 ± 0.1	59,000 ± 1,000
2111	81.7 ± 0.2	75.9 ± 0.4	80,000 ± 1,000
2211	47.4 ± 0.1	61.2 ± 0.1	90 ± 10
2221	77.2 ± 0.1	61.2 ± 0.1	350 ± 30
2121	68.2 ± 0.2	65.9 ± 0.6	9,700 ± 80
2212	50.9 ± 1.6	NA	140 ± 20
2112	69.3 ± 0.2	65.7 ± 0.1	2,030 ± 60
2122	63.0 ± 0.2	62.7 ± 0.6	2,630 ± 60
2222	65.2 ± 0.4	63.2 ± 0.1	50 ± 10
2222-D130S	69.2 ± 0.2	NA	70 ± 10
2222-F143A	65.7 ± 0.2	NA	330 ± 30
1111-SB	87.3 ± 0.1	83.0 ± 0.1	53,000 ± 500

[a]Error values represent one standard deviation for each calculated fit parameter or activity value.
[b]Activity of some chimeras was too low to accurately predict T_{50} values (NA).

2.4.4. REGRESSION ANALYSIS

The simplest form of analysis was based on single-block substitutions (2111, 1211, 1121, 1112). This approach indicated Block 2 was severely destabilizing ($\Delta T_m^{1111\rightarrow1211}$ = -31.3^{o}C) while Blocks 1 and 3 had a minimal impact on stability ($\Delta T_m^{1111\rightarrow2111}$ = -3.2^{o}C and $\Delta T_m^{1111\rightarrow1121}$ = -4.8^{o}C). Block 4 had a moderate destabilizing effect ($\Delta T_m^{1111\rightarrow1112}$ = -8.0^{o}C). One-body regression extended this basic analysis by averaging the impact of each block substitution across

the full chimera library. A simple one-body model was effective at capturing a large degree of stability variation for previous chimera libraries[150] despite ignoring possible block-block interactions. Using a leave-one-out technique for our library of 15 ΔT_m values, a training set of 14 ΔT_m values was used to fit four regression parameters and predict the excluded ΔT_m value for each set. Average, minimum, and maximum parameter values demonstrated moderate variability across the 15 training sets (Table 2.2). As with the monomeric chimeras, one-body regression analysis indicated Block 2 was most destabilizing (Block 2 parameter = -19.4°C). Block 3 appeared to be beneficial in many chimeras (average Block 3 parameter = +0.6°C) but did not improve stability when introduced independently into E1 ($\Delta T_m^{1111\rightarrow1121}$ = -4.8°C). The 16 chimeras showed a surprising range of blockwise stability impacts (Table 2.2). For example, incorporating Block 1 decreased stability by 0-15°C in the majority of chimeras, but in two cases Block 1 significantly increased stability ($\Delta T_m^{1221\rightarrow2221}$ = +7.9°C and $\Delta T_m^{1222\rightarrow2222}$ = +16.5°C). Similarly, for Block 2 the stability impact ranged from -31.3°C ($\Delta T_m^{1111\rightarrow1211}$) to +9.0°C ($\Delta T_m^{2121\rightarrow2221}$). In principle, this variability could stem from inconsistent measurement of T_m values or from interactions between sets of blocks. Every attempt was made to validate experimental measurements via replicate assays and comparison to alternate stability measurements (T_{50} values).

When one-body regression still did not accurately predict T_m values, we expanded the analysis to account for higher order effects. Including pairwise or two-body effects should account for cases in which block-block interactions contribute to stability. Ideally SCHEMA recombination would reduce non-native block interface interactions (only two mutations in Block 1 and four mutations in Block 4 were within 2 Å of block-block interfaces), but long-range interactions could still lead to two-body effects. For our relatively small library, including

six additional parameters to account for two-body effects risked over fitting the data. However, we were able to conclude first that two-body parameters were even more variable than one-body parameters (Table 2.2) and second that two-body effects did not drastically improve the predictive capability of regression. In the presence of two-body fitting, including additional parameters also showed that Block 2 alone was very destabilizing (Block 2 two-body parameter = -32.4^{o}C) and that flanking design blocks recovered some of this stability loss (Blocks 1-2 parameter = +11.6 ^{o}C and Blocks 2-3 parameter = +20.7 ^{o}C).

TABLE 2.2: PARAMETERS FOR LINEAR REGRESSION

Parameter	One-body				Two-body			
	Value (^{o}C)	St dev (^{o}C)	Min (^{o}C)	Max (^{o}C)	Value (^{o}C)	St dev (^{o}C)	Min (^{o}C)	Max (^{o}C)
Block 1	-4.7	1.3	-7.0	-1.8	-10.7	3.9	-24.1	-6.4
Block 2	-19.4	1.4	-21.6	-16.4	-32.4	1.6	-35.3	-28.5
Block 3	+0.6	1.2	-1.7	3.3	-5.2	1.5	-8.6	-1.8
Block 4	-10.8	1.2	-12.8	-7.5	-5.2	2.0	-8.4	0.0
Blocks 1-2					+11.6	3.2	6.7	18.3
Blocks 1-3					+4.6	3.2	-0.3	11.3
Blocks 1-4					+2.3	3.1	-2.6	8.9
Blocks 2-3					+20.7	2.9	16.5	26.6
Blocks 2-4					0.0	3.0	-4.3	5.8
Blocks 3-4					-7.4	3.0	-11.9	-1.6

2.4.5. MOLECULAR DYNAMICS SIMULATIONS

The first step in setting up useful MD simulations was identifying an appropriate temperature for a 50 ns simulation timescale.[162] Incrementally increasing temperature from 298 K to 500 K identified a maximally informative temperature near 450 K (Figure 2.6). Lower temperatures did not induce unfolding (protein maintained low RMSD) and higher temperature required a reduced step size to prevent explosion. Comparison between room temperature (298K) and elevated temperature (450K) simulations for each chimera helped distinguish

between unfolding events and normal thermal fluctuations. Simulating all 16 chimeras at both low and high temperature resulted in more than 1.6 μs of simulated data for analysis.

As indicated by root mean square fluctuation (RMSF), some design regions had altered flexibility relative to wild type E1 (Figure 2.7). These differences guided our initial study of residue interactions. However, it should be noted that focusing on RMSF alone might not be enough to describe stability, as rigidity does not always correlate with increased thermostability.[163] Furthermore, minor differences in residue RMSF might correspond to significant unfolding events.

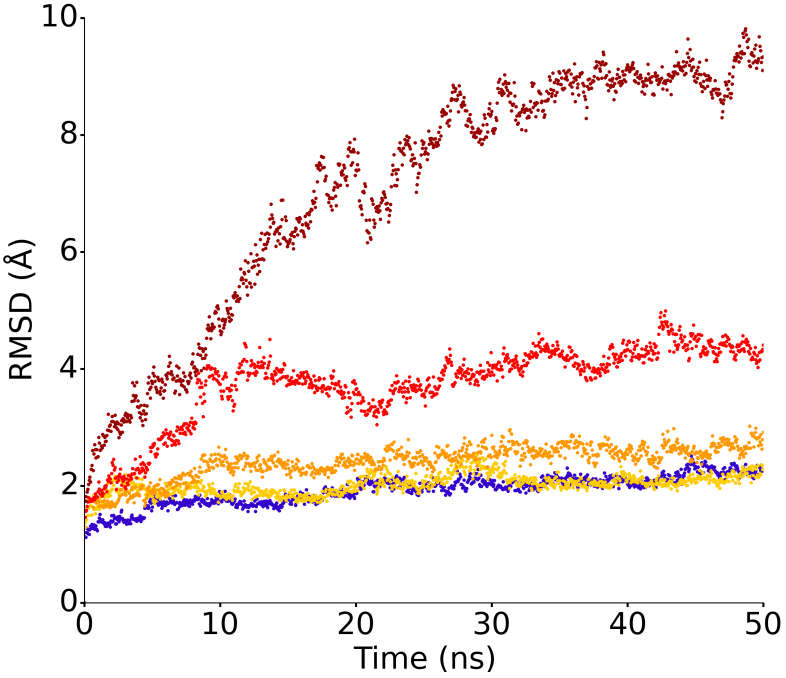

FIGURE 2.6: RMSD FOR WILD TYPE E1 AT VARYING TEMPERATURES
All-atom root mean square deviation (RMSD) for wild type E1 molecular dynamics simulations at varying temperatures: 298K (blue), 400K (yellow), 425K (orange), 450K (red), 500K (dark red).

FIGURE 2.7: PROTEIN ROOT MEAN SQUARE FLUCTUATION
Root mean square fluctuations were averaged across three 50 ns molecular dynamics simulations at 450 K for wild type E1 (solid black line), computationally designed E1 (solid grey line), and chimera 1211 (dashed grey line). Specific regions (shaded rectangles) were targeted for further analysis using site-directed mutagenesis.

One example of RMSF-guided analysis can be demonstrated in Region 1 (residues 16 to 22). This highly flexible region contained three mutations: Ala16Ser, Asn18Gly, and Val19Arg. Neighboring residue 356 also included a design mutation (Asp356Glu). Visual inspection of MD simulations showed Arg19 and Glu356 moved close together to form stable hydrogen bond interactions (Figure 2.8a,b). This salt bridge locked the C-terminus in place and decreased flexibility near residue 19, leading to an overall decrease in RMSF in Region 1.

Based on this predicted behavior, we introduced both mutations in wild type E1 to observe how the salt bridge would affect stability. The double point mutant (1111-SB) showed a remarkable 4.0^{o}C increase in T_{50} stability and 2.4^{o}C increase in $T_{m.}$ To the best of our knowledge, this is the most thermostable variant of E1 engineered to date. Thermostable celulase homologues from *Pyrococcus horikoshii* and *Thermotoga maritima* may have a higher stability (reported T_m values range from 92^{o}C to 102^{o}C [119,120,164]) but are quite distinct from E1 (45% and 20% sequence identity to E1, respectively). The benefits of improved thermostability become

evident when assays are performed at high temperatures. For example, the ~10^{o}C shift in optimum operating temperature for 1111-SB relative to 1111 (Figure 2.9) allowed the design to have a five-fold higher specific activity than wild type at 85^{o}C (72,500 ± 2,400 μmol/mg/min compared to 13,400 ± 400 μmol/mg/min, respectively).

FIGURE 2.8: SIMULATED BEHAVIOR AT MUTATION SITES

Example MD-guided mutagenesis studies. **(a)** Design mutations Val19Arg and Asp356Glu (black) improved stability relative to wild type E1 (light grey). **(b)** Favorable hydrogen bonding distance was maintained between Arg19 and Glu356 (black) for the majority of a 50 ns MD simulation. Wild type E1 lacking this salt bridge (light grey) displayed much greater N-terminal flexibility. **(c)** The distance between catalytic residues Glu162 and Glu282 was used as a proxy for active site rearrangement. This distance increased significantly in MD simulations for chimera 1211 (black) compared to wild type E1 (light grey). Grey spheres indicate the proximity of the bound substrate. **(d)** Unfolding events that led to increased catalytic residue distance were more common in MD simulations of chimera 1211 (black) than wild type E1 (light grey). **(e)** Phe143 was solvent exposed in the initial computational design (light grey) but reoriented to bury additional surface area during MD simulations of chimera 1211 (black). The buried state disrupted backbone hydrogen bonds in the α-helix between residues 139-143 and 142-146. **(f)** Fractional solvent accessible surface area and the sum of backbone hydrogen bond distances were selected as order parameters for monitoring unfolding near residue 143. For chimera 1211 (black), Phe143 spent a fraction of simulation time in the "buried" state. Ala143 in wild type E1 (light grey) did not display this partial unfolding behavior.

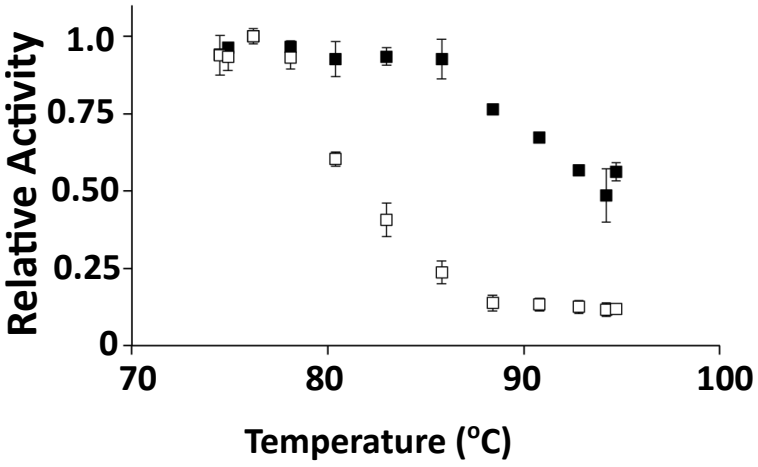

FIGURE 2.9: OPTIMUM OPERATING TEMPERATURE
Temperature optimum profiles for wild type E1 (1111, white squares) and the salt bridge design (1111-SB, black squares) demonstrate the shift in thermostability for the salt bridge mutant. Error bars indicate standard deviations from triplicate assays.

Stabilizing mutations Val19Arg and Asp356Glu provided an example of how computational design can correctly predict stability. Both FoldX and Rosetta scored these mutations as favorable. The average FoldX ΔΔG across all four scaffolds was -0.56 kcal/mol for Val19Arg and -0.13 kcal/mol for Asp356Glu. Similarly, the Rosetta energy score decreased by 0.37 REU for Val19Arg and 0.30 REU for Asp356Glu, primarily due to improved LK-solvation energy and reduced backbone dependent strain. In the consensus sequence alignments, Val19Arg was seen in 8.5% of aligned sequences and Asp356Glu was seen in 22.2% of aligned sequences. In a best-case scenario, two-body regression analysis would have identified the interaction between Blocks 1 and 4 without necessitating MD. Although the Block 1-4 parameter (+2.3^{o}C) indicated interactions between the blocks might be favorable, the effect appeared minimal relative to other two-body parameters (Table 2.2).

In contrast to the stabilizing mutations seen in Region 1, significantly destabilizing mutations were present in Region 2 (residues 120-134). RMSF for Region 2 increased slightly for the full sequence design 2222 and significantly for chimera 1211 (Figure 2.7). Closer

inspection showed that increased RMSF might have been due to reorientation of a loop region (residues 116-133), emphasizing the danger of neglecting alternate states in fixed-backbone design algorithms. Although design mutations Gln123Asn, Ser130Asp, and Ser131Gln were predicted to be favorable on the fixed backbone scaffold, alternate conformations were preferred during MD simulations. Any of these three mutations could have caused instability in Region 2; we selected Ser130Asp for further characterization based on its large fluctuation in MD simulations. Mutating back to the wild type Ser amino acid at site 130 in the design improved T_m by 4.0^{o}C. This accounted for ~20% recovery of stability despite being only one of sixty mutations present in the 2222 design sequence.

Instability in this region might also have directly influenced enzymatic activity. Loop reorientation allowed catalytic residue Glu162 to move away from the active site during MD simulations. The minimum heavy atom contact distance between catalytic residues Glu162 and Glu282 (EE_{min}) was used as a proxy for measuring active site orientation (Figure 2.8c). In chimera 1211, EE_{min} was significantly higher than wild type E1 throughout the simulation (Figure 2.8d). Similar localized unfolding was also seen in several different chimeras that contained design block 2 (1212, 1222, 1122, 2221), indicating the unfolding event may not have been unique to the simulation or the chimera. Previous studies have shown that orientation of catalytic residues in MD simulations can be correlated with enzymatic activity.[165,166] In our case, active site rearrangement could help explain some of the activity loss in E1 chimeras, but replicate simulations would be required to more quantitatively predict enzyme activity.

As a further test of MD capabilities, a single isolated point mutation was flagged for analysis. A mutation from alanine to phenylalanine (Ala143Phe) present in the unstable region of Block 2 showed unusual behavior during MD simulations. The original minimized design

structure placed the hydrophobic ring on the surface of the protein: Phe143 was 26% solvent exposed based on MSMS calculations.[167] During the course of the simulation, the phenylalanine reoriented to bury additional surface area (Figure 2.8e). The torsional stress induced by this reorientation disrupted the backbone hydrogen bonds and caused the helix to begin to unfold. Metrics based on fractional solvent accessible surface area (SASA) of residue 143 and the sum of O-H backbone hydrogen bond distances between residues 139-143 and 142-146 partially discriminated between the folded (exposed) and unfolded (buried) states (Figure 2.8f). Fractional SASA remained near 0.25 for wild type E1 but decreased to near zero during the Phe143 ring reorientation in 1211. Similarly, the combined O-H distances fluctuated between 3.8Å and 5.8 Å for 1111 (close to the ideal single hydrogen bond distance of 1.95 Å [168]), while the same hydrogen bonds were completely lost (>4 Å between O and H atoms) during the partial unfolding of 1211. The localized unfolding induced by the Phe143 reorientation was variable in other chimera simulations, making it difficult to predict the relative contribution of mutation Ala143Phe to overall protein stability.

Mutation Ala143Phe was assessed experimentally by studying the reverse mutation in the design sequence (2222-F143A). The reverse mutation had a marginal impact on stability (T_m increase of 0.5^{o}C), suggesting the original Ala143Phe mutation was slightly destabilizing. In contrast to the experimental results, all initial computational design steps predicted Ala143Phe to be favorable. Attractive van der Waals interactions with surrounding hydrophobic residues (Leu70, Leu94, Leu144) caused the Rosetta energy function to score this mutation as favorable (overall decrease of 0.43 REU). Furthermore, FoldX predicted the ΔΔG of Ala143Phe to be favorable on all four scaffolds (-0.6 kcal/mol, -0.74 kcal/mol, -0.84 kcal/mol, and -1.06 kcal/mol for scaffolds 1VRX.A, 1VRX.B, 1ECE.A, and 1ECE.B, respectively). In the consensus sequence

alignment, phenylalanine was the most commonly seen amino acid at position 143 (present in 43 out of 180 sequences). Other common mutations were methionine (39 sequences) and tryptophan (37 sequences), but alanine was present in only 12 sequences at this site. Although *Xanthomonas* species accounted for more than half (24 total) of the aligned sequences that contained phenylalanine at position 143, glycoside hydrolases from at least 6 other species were included in this set, indicating the BLAST results were not unrealistically skewed by highly similar sequences from the same species.

Agreement between consensus sequent alignments, FoldX energy analysis, and Rosetta energy scoring was not enough to correctly select a stabilizing mutation at site 143. Accurately predicting mutational stability remains a difficult challenge. By dissecting design methodology flaws down to the level of specific mutations, we provide challenging training examples for subsequent design methodology improvement.

2.5. DISCUSSION

In principle, employing independent design strategies (consensus sequence alignment, FoldX) would be expected to reduce the probability of introducing destabilizing mutations. Komor *et al.*[169] have previously shown how a combined consensus-FoldX strategy can identify stabilizing mutations in a cellobiohydrolase enzyme. Despite taking precautions in our initial design process, the computationally designed sequence displayed both reduced activity and stability relative to the native sequence. Wild type E1 set a high stability standard with a T_m of 84.9^{o}C; in contrast, the designed sequence T_m was reduced to 65.2^{o}C. Similarly, the specific activity was reduced from 70,000 μmol/mg/min to 50 μmol/mg/min. Given that the design included 60 mutations, it proved challenging to generate specific hypotheses about the cause of

instability. Recombination with wild type E1 created a library of 14 additional chimeras, all of which displayed reduced stability relative to wild type E1 (Table 2.1). In one instance (chimera 2111), specific activity was higher than wild type E1 at 50^{o}C. This effect may have been caused by a shift in optimum operating temperature. If specific activity assays were performed at a higher temperature, the more stable wild type sequence would likely be favored. The detailed analysis outlined for our E1 library provides an example of how researchers can overcome CPD deficiencies, improve initial designs, and potentially extend the lessons learned to yield improvements in future computational methods.

Narrowing a large set of mutations down to a handful of impactful mutations requires a technique that spans multiple levels of detail. Initially, it can be helpful to identify broad regions or blocks of a protein that positively or negatively impact stability. Previous reports indicated SCHEMA recombination followed by one-body linear regression accurately predicted chimera stability.[150] By extending this approach to include a computational design as a starting parent, we estimated the relative stability contributions from four separate design blocks. All four of the blocks were destabilizing when introduced independently into the wild type sequence. Destabilizing mutations within each block masked any potential stabilizing mutations. We originally selected four blocks to make complete library construction and characterization more feasible. This choice turned out to be fortuitous; only by characterizing the stability of nearly the entire library could we confidently assert the existence of "higher-order" stability effects. In contrast to previous libraries, one-body regression did not accurately predict overall chimera stability for our l6-member library. The impact of each block varied based on the context of the surrounding blocks. Higher order effects (beyond one and two body effects) were clearly present. The reason why the effects were large in this case relative to previous libraries is not yet known.

Although there were significant differences between the topology of E1 and the cellobiohydrolase II studied by the Arnold lab,[147] both proteins were of comparable size and consisted of a similar αβ-barrel motif. A variety of other protein folds have also been studied using the SCHEMA approach, including cellobiohydrolase I,[90] β-lactamase,[148] and P450;[170] therefore, it appears topology is not a limiting factor in determining block modularity. The number of mutations per block in our library was comparable or even lower than previous studies due to the high sequence identity of the E1 parent sequences. In addition, the average SCHEMA energy <E> remained low for our library (Figure 2.5) despite selecting only four blocks. The contact map for E1 (not shown) indicated E1 chimeras had a comparable ratio of intra- to inter-block contacts as that of Heinzelman *et al.* (2009a). Although the underlying cause of the higher order coupling remains unclear, the unusual behavior of this library makes it a challenging test case for understanding protein stability. In combination with other systematic library stability measurements, the data reported herein could prove highly rewarding in training future design methods.

While regression broadly aided in identifying destabilizing regions of the design, it proved difficult to refine the regression analysis to a residue level interpretation. All mutations were computationally predicted to be favorable, and visual inspection had already been used to eliminate any severe clashes or noticeably poor interactions. Some mutations switched hydrogen bonding partners, but the net number of hydrogen bonds remained nearly constant. Regression analysis alone was not sufficient for isolating stability impacts; instead, more detailed dynamic behavior was studied in a solvated environment using molecular dynamics (MD).

The predictive capability of MD was demonstrated for three test cases (Figure 2.8). In Example 1, an isolated salt bridge was identified that significantly increased E1 thermostability.

Mutations Val19Arg and Asp356Glu were present in the initial design, but their beneficial impact was not identified via regression analysis. MD simulations showed the salt bridge formed stable hydrogen bond interactions and locked the C-terminus in place. Although salt bridges are a common stabilizing feature in thermophilic proteins,[171] rationally introducing a salt bridge can be difficult due to competing interactions with solvent. Conserved hydrogen bond interactions in MD simulations (even at elevated temperature) provided a good indication that Arg19 and Glu356 might form a stabilizing salt bridge. When tested in lab, these two mutations increased thermostability (T_{50}) by 4.0^{o}C relative to wild type E1. In Example 2, a contrasting scenario was presented where initial design mutations led to decreased stability and activity. A single mutation (Ser130Asp) was hypothesized to influence stability based on fluctuations seen in MD simulations. When the reverse mutation was introduced into the design sequence (2222-D130S), T_m increased by 4^{o}C (Table 2.1). Example 3 provided a limiting test case where an Ala143Phe mutation only marginally impacted overall stability. Burial of Phe143 hydrophobic surface area caused partial helix unfolding during MD simulations, but the unfolding event was not nearly as severe as in Example 2. Experimental assessment of the reverse mutant 2222-F143A showed only a slight change in stability ($\Delta T_m^{2222 \rightarrow F143A} = +0.5^{o}C$).

These examples demonstrate the potential benefit of including MD in the computational design process. The ability to discriminate between stable and unstable design mutations could improve the success rate of CPD by salvaging failed initial designs. Alternatively, MD could be used as a prescreening strategy to predict design behavior before proceeding to the lab scale.

2.6. CONCLUSION

The poor stability and activity of our initial E1 design served as a reminder that design methods will need continued improvement to achieve more accurate predictions. For our design, computational search algorithms failed to identify the more stable wild type sequence from the larger sequence search. Discrepancies between predicted and measured stability values demonstrated how small point-mutation inaccuracies can be compounded and eventually lead to an unfavorable design sequence. Our design method reiterates the challenges associated with CPD and stresses that detailed analysis will be required to incrementally progress towards successful full sequence design.

3. CHARACTERIZATION OF SUPERCHARGED CELLULASE ACTIVITY AND STABILITY IN IONIC LIQUIDS

Lucas B. Johnson,[a] Sehoo Park,[a] Lucas P. Gintner,[a] Christopher D. Snow[a]

This chapter is derived in part from an article published in the *Journal of Physical Chemistry B: Enzymatic* on July 12, 2016, and is reproduced here with permission (Appendix B).

[a]Author affiliation:

Colorado State University

1370 Campus Delivery

Fort Collins, CO 80523-1370

USA

*Corresponding author contact information:

Christopher.Snow@ColoState.edu

Phone: (970) 491-5276

Fax: (970) 491-7369

3.1. SUMMARY

Ionic liquids (ILs) have many potential benefits in biochemical processes, including improved substrate or product solubility, increased enzyme selectivity, and higher yield. Varying ion substituents allow ILs to be tuned or optimized to achieve a specific goal. Unfortunately, optimization based on a single design criterion can have undesirable side effects on other process components. For example, hydrophilic ILs capable of efficiently dissolving biomass often inhibit enzymatic activity during hydrolysis. A panel of nine different aqueous ILs was selected for this study to systematically assess which factors contribute to the loss of enzyme activity. The activity of endoglucanase E1 from *Acidothermus cellulolyticus* steadily decreased in higher concentrations of ILs, especially in the presence of the common cellulose dissolving solvent 1--ethyl-3-methylimidazolium acetate. The impact of most other ILs could be rationalized via the Hofmeister series. Enzyme behavior was further probed by rationally modifying the surface charge of E1. Variants were computationally designed to have positively or negatively charged surfaces and assessed for activity in ILs. Surprisingly, positive supercharging maintained wild type activity levels in ILs, while negative supercharging drastically reduced activity. Discrepancy between stability and activity measurements for some ILs indicated active site inhibition or other unique inactivation mechanisms might be crucial components to consider in future studies.

3.2. INTRODUCTION

Salts with melting temperatures below 100^{o}C, termed ionic liquids (ILs), offer a variety of benefits in industrial processes. Much like organic solvents, ILs can be used to adjust polarity or hydrophobicity in a reaction system and alter the solubility of substrates and products.[172] ILs

can also enhance stereoselectivity or facilitate phase separation.[173] Among other applications, ILs have been proposed as a method for pretreating cellulosic biomass prior to enzymatic hydrolysis. Pioneering work by Rogers and coworkers first demonstrated that neat ILs could dissolve cellulose,[174] and many imidazolium-based ILs have since been tested on lignocellulosic substrates.[175] ILs that have performed exceptionally well with industrially relevant substrates include 1-ethyl-3-methylimidazolium acetate ([Emim]OAc), 1-butyl-3-methylimidzolium chloride ([Bmim]Cl), 1-allyl-3-methylimidzaolium chloride ([Amim]Cl), and 1,3-dimethylimidazolium dimethylphosphate ([Dmim]DMP).[175]

Unfortunately, it remains difficult to find an IL that dissolves biomass without detrimentally impacting enzyme activity. The high hydrogen bond basicity, or thermodynamic tendency to act as a hydrogen bond acceptor, typically makes cellulose-dissolving ILs incompatible with enzymes.[176] Early studies indicated cellulase activity was drastically reduced in ILs.[177] In recent years, numerous other enzymes have been assessed for activity and stability across a wide range of ILs.[175,178–181] Various strategies have been proposed for preserving enzymatic activity[180,182] but a generic stabilization approach remains elusive.

As early as the 19th century, the role of ions in stabilizing or destabilizing proteins was categorized via the Hofmeister series.[183] Enzymes typically prefer large, low charge density, kosmotropic anions over small, high charge density, chaotropic anions.[178,179,184,185] This trend is reversed for cations due to the "matching kosmotropicity" effect.[186] Although the Hofmeister series provides useful guidance, examples exist where this classification scheme fails to apply.[178] In addition to ion kosmotropicity, solvent properties such as viscosity, polarity, hydrophobicity, nucleophilicity, or hydrogen bond basicity can be important.[179,187–189] To date there is little consensus regarding how water, IL, and enzymes interact in aqueous-IL solutions. Solvated ions

can influence enzymes in a variety of ways, include stripping of surface water molecules, changing of protein structure and dynamics, modification of surface pH, or preferential stabilization of the unfolded state. [172,178,190,191]

Modifying the surface charge of a protein inherently influences its stability, solubility, and aggregation propensity[192–194] and is likely to play a key role in moderating IL-enzyme interactions. Kaar and coworkers experimentally demonstrated the importance of electrostatic interactions by chemically modifying surface charge groups on cellulase, lipase, papain, and chymotrypsin.[12–14] Chemical modification via acetylation or succinylation improved resistance against denaturing ILs [Bmim]Cl and [Emim]ethylsulfate. Furthermore, molecular dynamics simulations indicated surface charge significantly altered ion interactions at the enzyme surface.[195] Other reports have noted a correlation between native enzyme charge and halotolerance,[196–198] providing additional evidence that electrostatic interactions can strongly influence protein behavior in ionic solutions.

Systematically altering protein charge via mutagenesis could provide a direct route for probing the IL-enzyme interface. Although the "supercharging" concept introduced by Liu and coworkers[199] has not previously been employed in conjunction with ionic liquids, it provides an extreme example for studying enzyme charge modification. Supercharging imparts a high net charge, either positive or negative, on the protein surface.[192–194] Although successful in many regards, supercharged variants can display decreased binding affinity or loss of enzymatic activity.[199] Ideally, careful selection of charged mutation sites minimizes these negative impacts.

To build upon previous supercharging studies, we set out to explore IL inactivation mechanisms for supercharged variants of endoglucanase (E1) from *Acidothermus cellulolyticus*. E1 is an industrially relevant enzyme with high native stability that has been well characterized

through mutagenesis and structural studies.[97,101,104,109,111–113,151] A single mutation near the active site (Tyr245Gly) has previously been shown to reduce product inhibition[101] and was therefore included in wild type and design sequences. Positively charged (pE1) and negatively charged (nE1) variants were computationally designed to have altered surface electrostatics. Each enzyme variant was tested against a panel of imidazolium-based ILs (Figure 3.1) in aqueous solutions to assess specific ion interactions.

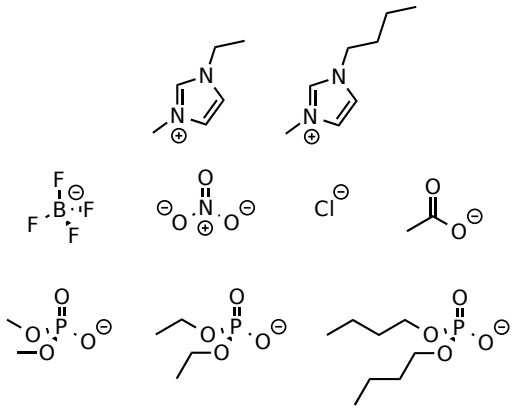

FIGURE 3.1: IONIC LIQUIDS USED IN EXPERIMENTAL ASSAYS
Top row, left to right: cations 1-ethyl-3-methylimidazolium ([Emim]) and 1-butyl-3-methylimidazolium ([Bmim]). Middle row: anions tetrafluoroborate (BF_4), nitrate (NO_3), chloride (Cl), and acetate (OAc). Bottom row: anions dimethylphosphate (DMP), diethylphosphate (DEP), and dibutylphosphate (DBP).

3.3. MATERIALS AND METHODS

3.3.1. COMPUTATIONAL PROTEIN DESIGN

The computational method used to selectively introduce charged mutations on the surface of E1 closely resembled the Rosetta approach demonstrated by the Kuhlman lab. [200,201] Candidate mutation sites were selected based on solvent exposed surface area calculations[167] and evaluated using the Rosetta energy function[48] as implemented in SHARPEN.[155] Key catalytic residues (Glu162/Glu282), binding residues conserved within glycoside hydrolase family 5 (Arg62, His116, Asn161, His238, Tyr240, Trp319), disulfide bonds (Cys34/Cys120,

Cys168/171), and prolines were held fixed and not allowed to mutate during the design process. Either positively charged (Lys, Arg) or negatively charged (Glu, Asp) mutations were allowed at each design site, and a selective energy bias was applied to adjust the number of mutations and achieve a desired net charge (Figure 3.2). Based on previous supercharging reports,[199–201] a net change in charge of ~20 e^- units was targeted for E1 designs. Supernegative nE1 contained a total of 24 mutations (93.3% sequence identity to wild type E1), of which 13 were initially polar, 6 were nonpolar, 2 were positively charged, and 3 were negatively charged (Asp to Glu mutations). For pE1, 26 candidate mutation sites were identified. Site 165 had previously been characterized as unfavorable for positively charged mutations[112] and was therefore left as wild type aspartic acid. In addition, site 324 was deemed too close to the active site upon visual inspection and was left as wild type aspartic acid. Therefore, the final pE1 design contained 24 mutations (93.3% sequence identity to wild type E1), of which 13 were initially polar sites, 6 were nonpolar, 1 was negatively charged, and 4 were positively charged (Lys to Arg mutations). More than half (14 of 24) of the mutation sites were common to both pE1 and nE1. The full design sequences are available in Table 3.1.

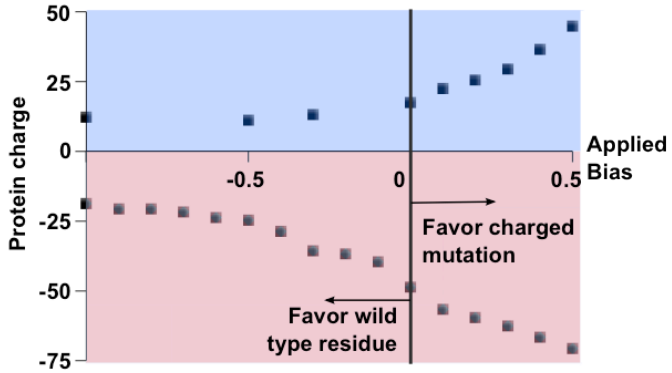

FIGURE 3.2: SUPERCHARGING DESIGN BIAS
A selective energy bias was applied during the design search to favor or disfavor charged mutations. A positive bias favored the mutation at a given site, while a negative bias disfavored the proposed mutation. A bias of -0.3 Rosetta Energy Units (REU) was used to design nE1, and a bias of -0.1 REU was used to design pE1. Coloring denotes net positive charge (blue) and net negative charge (red).

TABLE 3.1: SUPERCHARGED DESIGN SEQUENCES

Wild type E1 amino acid sequence
MAGGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVVHGLWSRDYRSMLDQIKSLG YNTIRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHR PDCSGQSALWYTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWGCGDPS IDWRLAAERAGNAVLSVNPNLLIFVEGVQSYNGDSYWWGGNLQGAGQYPVVLNVPNR LVYSAHDYATSVGPQTWFSDPTFPNNMPGIWNKNWGYLFNQNIAPVWLGEFGTTLQST TDQTWLKTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTVDTDKDGY LAPIKSSIFDPVHHHHHHH
Superpositive (pE1) amino acid sequence
MAGGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVVHGLWSRDYRSMLDQIRSLG YNTIRLPYSDDILKPGTRPNSIKFKRMNQDLQGLTSLQVMDRIVAYAGKIGLRIILDRHRP DCSGQSALWYTRKVSEATWISDLQRLAQRYKGNPTVVGFDLHNEPHDPACWGCGDPSI DWRLAAERAGNAVLSVNPNLLIFVEGVQSYNGDSYWWGGNLQGARKYPVVLNVPNR LVYSAHDYATSVGPQTWFKDPTFPNNMPRIWNRNWGYLFKQRIAPVWLGEFGTTLRST TDQTWLRTLVQYLRPTRQYGADSFQWTFWSWNPDSGDTGGILKDDWKTVDKDKDGY LRPIKSSIFRPVHHHHHHH
Supernegative (nE1) amino acid sequence
MDGGGYWHTDGREILDANNVPVRIAGINWFGFETCNYVVHGLWSRDYRSMLDQIKSL GYNTIRLPYSDDILKPGTMPNSINFDDMNQDLQGLTSLQVMDEIVAYAGEIGLRIILDRH RPDCSGESALWYTDDVSEDTWISDLQALAQRYKGNPTVVGFDLHNEPHEPACWGCGD PSIDWRLAAERAGNAVLSVNPNLLIFVEGVQSYNGDSYWWGGNLQGAEQYPVVLNVP NRLVYSAHDYATSVGPQTWFDDPTFPNNMPGIWNDNWGYLFEDNIAPVWLGEFGTTL QSTTDQTWLKTLVQYLRPTEQYGADSFQWTFWSWNPESGDTGGILKEDWEEVDEDKD GYLDPIKSSIFDPDHHHHHHH

3.3.2. ENZYME EXPRESSION AND PURIFICATION

Supercharged sequences were constructed via commercial gene synthesis (DNA2.0, California, USA). Expression and purification steps followed those previously reported.[104] Briefly, proteins were expressed in *E. coli* BL21 (DE3) pLySs cells using a glucose/lactose induction system, lysed using sonication, and purified via immobilized metal affinity chromatography. Soluble protein yields ranged from 10 to 30 mg protein per liter culture. All variants except pE1 were buffer exchanged into 50 mM sodium acetate pH 5; pE1 was buffer exchanged and stored in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

pH 7.5 for improved stability. Proteins were assessed for relative size and purity using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, Figure 3.3).

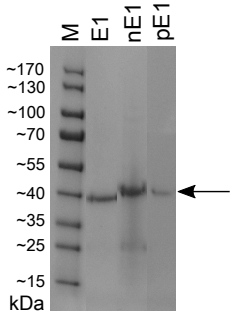

FIGURE 3.3: SDS-PAGE ANALYSIS
Protein size and purity were assessed via SDS-PAGE after immobilized metal affinity chromatography. From left to right, gel lanes contained marker (M), wild type (E1), supernegative (nE1), and superpositive (pE1). The arrow marks the expected E1 size at ~41 kDa.

3.3.3. ACTIVITY AND STABILITY ASSAYS

Expressed enzymes were tested for activity using the fluorescent substrate p-nitrophenol-β-d-cellobioside (pNPC). Assays were performed at 50ºC for 30 minutes in 125 mM sodium acetate pH 5 with 200 ug pNPC. Similarly, salt and IL assays were performed at 50ºC, pH 5 for 30 minutes. Retained activity (T_{50}) was measured after a 30 minute preincubation temperature ranging from 40 to 80ºC. Curves were fit to T_{50} data using an inverse sigmoidal dose-response curve, and T_{50} was defined as the temperature that resulted in 50% decreased activity. Optimum pH conditions were assessed by combining enzyme and 200 mM buffer at various pH: glycine pH 2, sodium citrate pH 3 to 4, sodium acetate pH 5, 2-(N-morpholino)ethanesulfonic acid (MES) pH 6, HEPES pH 7 to 8, and N-Cyclohexyl-2-aminoethanesulfonic acid (CHES) pH 9.

Stability assays were performed using the fluorescent dye SYPRO Orange (Sigma Aldrich #S5692). Fluorescence measurements were taken at excitation/emission wavelengths of 470/570 nm using the VIC filter on a Realplex2 Mastercycler (Eppendorf). Temperature was gradually increased from 25ºC to 98ºC using a linear temperature gradient of 2ºC/min. Each

sample was prepared with 1-10 μg protein, 2.5 μl 125x SYPRO orange dye, and 10wt% IL and was brought to a final volume of 20 μl with 200 mM sodium acetate pH 5. Denaturation curves were fit based on the Clarke-Fersht equation.[159] Apparent T_m values (not thermodynamic due to irreversible unfolding) indicated the temperature at which 50% of the protein was unfolded. Standard deviations of T_m fit parameters were based on triplicate measurements.

3.3.4. MATERIALS

All ionic liquids were purchased from Sigma Aldrich and stored in a desiccator. Prior to use, each IL was diluted to 50wt% concentration with 200 mM sodium acetate and titrated to pH 5 using hydrochloric acid. Careful adjustment of pH was necessary for isolating ion effects from pH effects. When pH was not fixed, enzymes were most active in acidic ILs (near pH 5) and inactive in basic ILs.

3.4. RESULTS AND DISCUSSION

3.4.1. COMPUTATIONAL DESIGN

The net charge of E1 was systematically altered via introduction of charged surface mutations. Positively charged (Lys/Arg) or negatively charged (Asp/Glu) mutations were introduced at sites that avoided key catalytic residues and scored well after combinatorial rotamer optimization with the Rosetta energy function.[48] Net charge was shifted from -9.3 (E1) to +11.7 (pE1) or -32.2 (nE1) as approximated by Protein Calculator v3.4 (protcalc.sourceforge.net). The ratio of primary amines to carboxylic acids was 10:35, 15:34, and 8:56 for wild type E1, pE1, and nE1, respectively. The negative design achieved a comparable amine:acid ratio (0.14 vs. 0.13) as the succinylated form of chymotrypsin in previous studies.[12]

Both pE1 and nE1 designs had drastically altered surface potential as assessed by Delphi Poisson-Boltzmann electrostatics calculations (Figure 3.4).[202]

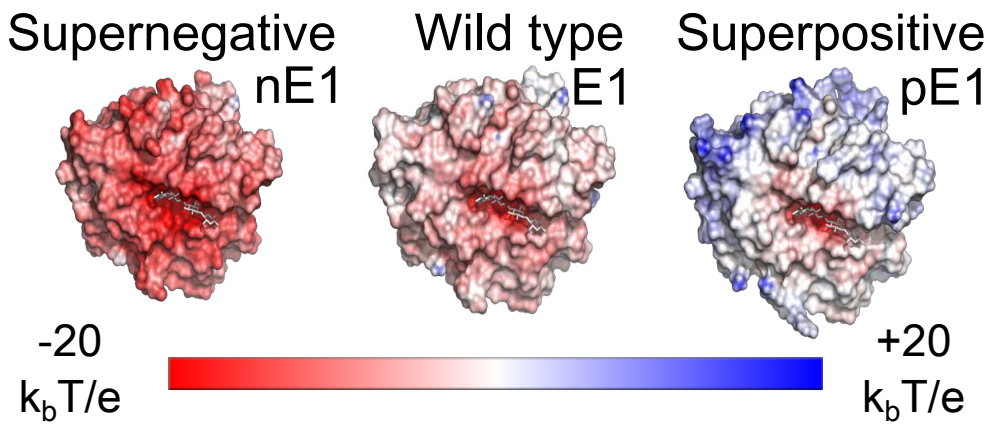

FIGURE 3.4: ELECTROSTATIC POTENTIAL MAP FOR E1, PE1, NE1
The structural model for E1 was adopted from Protein Data Bank structure 1ECE; structures for pE1 and nE1 were homology models. Bound substrate cellotetraose is shown in stick representation to indicate the substrate binding pocket. Color bar denotes the electrostatic potential range.

3.4.2. IONIC LIQUID IMPACT: HOFMEISTER SERIES

E1 was tested against a panel of seven anions and two cations (Figure 3.1). [Bmim] and [Emim] were selected to represent common cellulose dissolving cations, and anions were selected to have varying charge density, kosmotropicity, and hydrophobicity. Activity was assessed in aqueous IL solutions with concentration ranging from 0-25wt%. In general, the impact of ILs on E1 activity followed the Hofmeister series[183]: E1 preferred bulky kosmotropic anions such as dialkylphosphates (DMP/DEP) over chaotrophic anions such as nitrate (NO_3) or tetrafluoroborate (BF_4). Among 25wt% [Emim] ILs, E1 activity increased as follows: $BF_4 < NO_3 < Cl \approx DMP < DEP$ (Figure 3.5, 3.6).

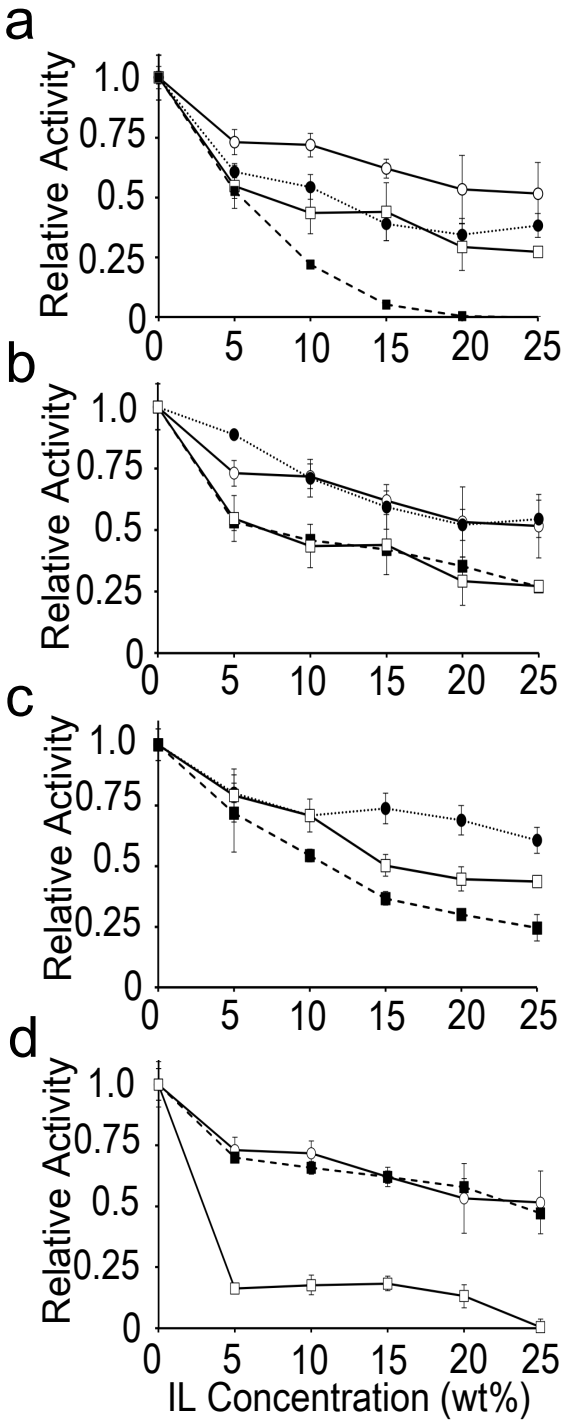

FIGURE 3.5: RETAINED ENDOGLUCANASE ACTIVITY IN IONIC LIQUIDS
(a) E1 activity in [Emim]Cl (open circle), [Emim]NO_3 (filled circle), [Emim]BF_4 (open square), and [Emim]OAc (filled square). **(b)** E1 activity in ILs containing [Emim] and alkyl phosphate anions: dimethyl phosphate (open square), diethyl phosphate (filled circle), and dibutyl phosphate (filled square). **(c)** E1 activity in [Emim]Cl (open circle), [Emim]BF_4 (open square), [Bmim]Cl (filled squre), and [Bmim]BF_4 (filled circle). **(d)** Enzyme activity in [Emim]Cl for E1 (open circle), pE1 (filled square), and nE1 (open square).

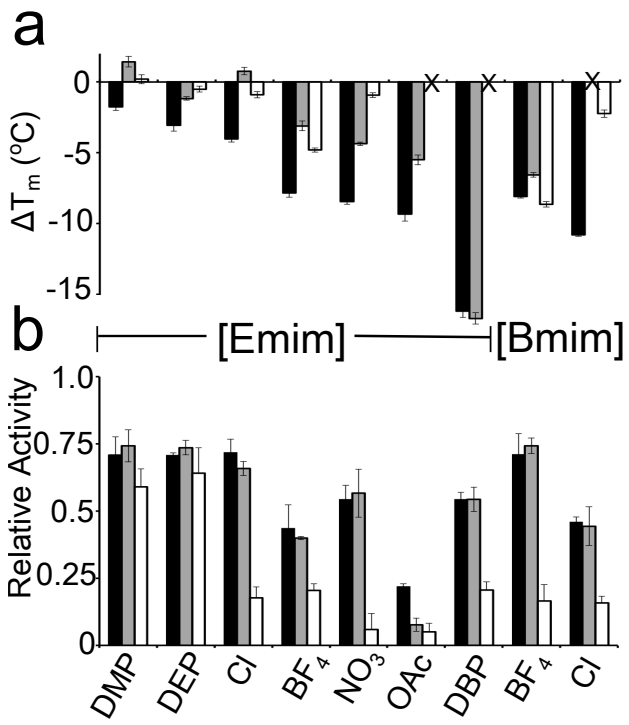

FIGURE 3.6: PROTEIN STABILITY IN IONIC LIQUIDS
(a) ΔT_m values indicate the shift in melting temperature in the presence of 10wt% IL. Temperatures are relative to the respective T_m of each enzyme in buffer ($T_{m,E1}$=84.3±0.4°C, $T_{m,pE1}$=74.4±0.2°C, $T_{m,nE1}$=59.0±0.1°C). T_m values could not be determined for some enzyme/IL combinations (marked by X). Cations (middle label) and anions (bottom label) correspond to both panels A and B. **(b)** Relative enzyme activity in 10wt% IL. Shading denotes E1 (black), pE1 (grey), or nE1 (white). Error bars indicate the standard deviation of triplicate measurements.

The exceptions to the Hofmeister trend were acetate and dibutylphosphate ([Emim]OAc and [Emim]DBP). The long alkyl chain of DBP likely caused unique hydrophobic interactions with E1 (hydrophobicity will be discussed in the next section). It is not well understood why [Emim]OAc severely inactivated the enzymes. Despite differing native stabilities, E1, pE1, and nE1 all had <10% activity in 15wt% [Emim]OAc (Figure 3.7a). As mentioned in Li *et al.*,[203] [Emim]OAc ions could interrupt interactions with the substrate through competitive inhibition, chemically altered catalytic residues, or substrate modifications. Inactivation effects for other cellulases appeared to be consistent across multiple different substrates,[164,204–206] alleviating some concern regarding unique substrate modifications. Although evidence of cation inhibition

has been seen in computational studies of imidazolium ILs,[190,207–209] it fails to explain why [Emim]OAc was more inactivating than [Emim]Cl or other ILs containing [Emim] cations. In E1 assays, similar concentrations of sodium acetate or [Emim]Cl were not nearly as inactivating as [Emim]OAc (Figure 3.7b). The choice of buffer did influence the inhibitory effect of [Emim]OAc at low concentrations (Figure 3.7c), but in all cases E1 had <10% activity in 15wt% [Emim]OAc. Activity of other Cel5A enzymes in [Emim]OAc has varied significantly (0-79% activity in 15% [Emim]OAc) and appeared to correlate with thermostability.[164] However, if activity loss was due entirely to protein denaturation, one might expect activity to continue to decay over time. E1 and pE1 activity levels were significantly reduced in initial activity assays, but there was no additional activity loss after 6 days incubation in [Emim]OAc (Figure 3.8). Furthermore, other ILs such as [Emim]NO_3 which were almost as destabilizing as [Emim]OAc (ΔTm of -8.5±0.2^{o}C and -9.3±0.5^{o}C, respectively) did not completely inactivate E1 (Figure 3.6). Unexpected E1 behavior reiterates the difficulty in predicting the numerous interactions between enzymes, cations, anions, water, substrates, products, and buffer.[210] Future computational simulations with [Emim]OAc could be beneficial in better understanding this particular system.

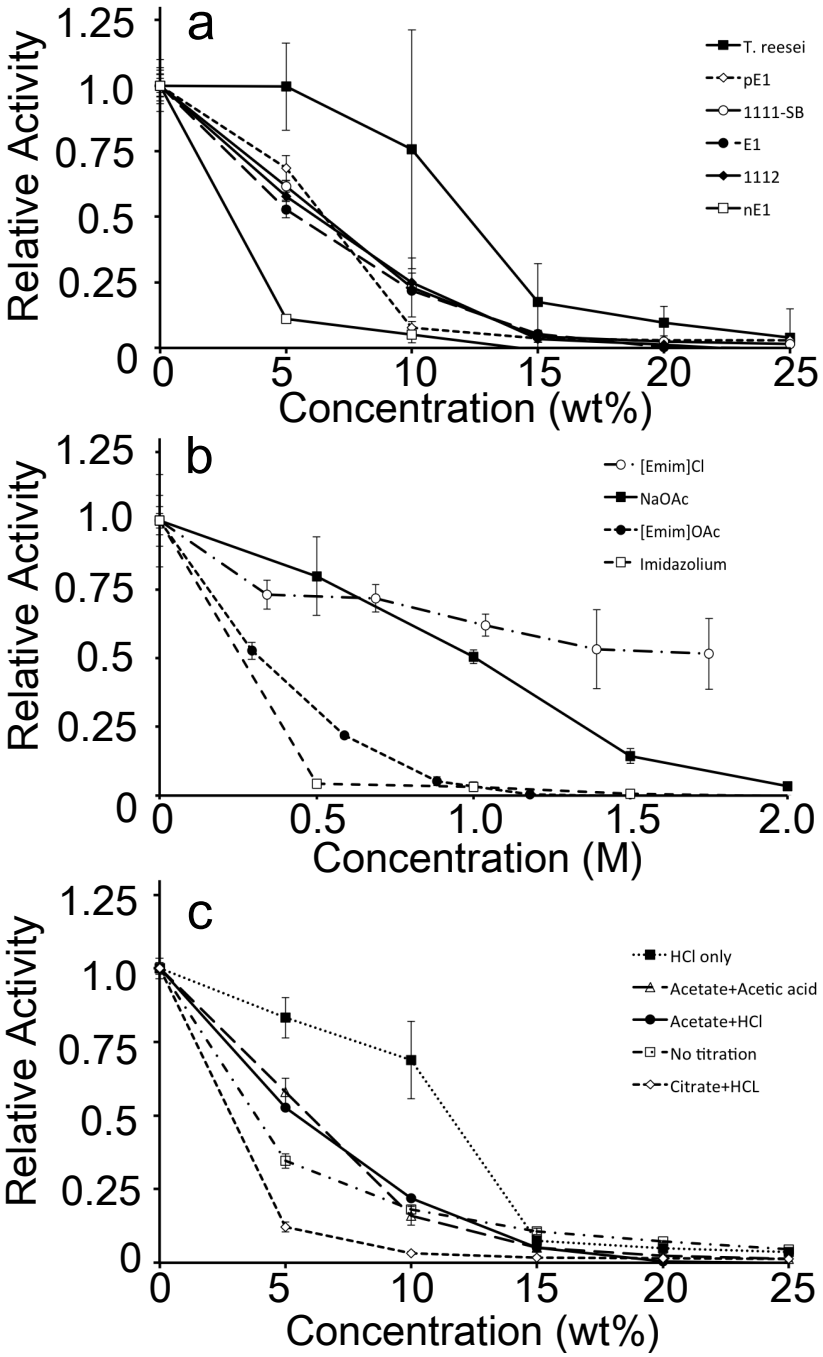

FIGURE 3.7: DISSECTING [EMIM]OAC DENATURING EFFECTS
(a) Activity of *T. reesei* cellulase extract (Sigma C2730) and E1 homologues in [Emim]OAc. Numbered chimeras named based on Johnson *et al.*[104] **(b)** Inhibition of E1 activity could not be isolated to the IL cation or anion independently. **(c)** Choice of buffer or titration method influenced activity at low concentrations but was indistinguishable at 15wt% [Emim]OAc. Error bars indicate standard deviation of triplicate assays.

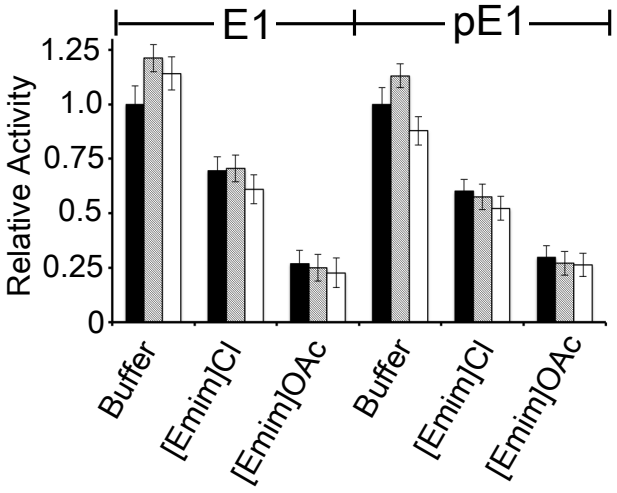

FIGURE 3.8: LONG TERM ACTIVITY FOR E1 AND PE1
Residual activities of E1 and pE1 were measured after 0 days (black), 3 days (grey), or 6 days (white) incubation in 50 mM sodium acetate pH 5 at 50°C with and without 10wt% [Emim]Cl or [Emim]OAc. Activities were measured 30 minutes after the addition of substrate. Although ILs rapidly decreased activity in initial assays, no further activity loss occurred over time for E1 or pE1.

3.4.3. IONIC LIQUID IMPACT: HYDROPHOBICITY

Ion hydrophobicity played an evident role in moderating enzyme activity. Previous studies have indicated shorter alkyl chains were preferable for stability.[211] In E1 assays, decreasing cation tail length by two carbons (ethyl [Emim] vs. butyl [Bmim]) significantly increased activity for chloride-based ILs. Relative E1 activity was ~52% for 25wt% [Emim]Cl and ~27% for [Bmim]Cl (Figure 3.5b). These activity levels directly corresponded with enzyme stability. Compared to E1 stability in buffer only (T_m=84.3±0.4°C), 10wt% [Emim]Cl was only moderately destabilizing (T_m=80.3±0.2°C) while 10wt% [Bmim]Cl was significantly destabilizing (T_m=76.2±0.1°C).

Variations in anion hydrophobicity showed that a long hydrophobic tail had a significant impact on activity and stability. E1 maintained moderate activity levels in 25wt% [Emim]DMP and [Emim]DEP (44% and 61%, respectively), while activity in 25wt% [Emim]DBP was reduced to ~25% (Figure 3.5c). Stability values correspondingly decreased with longer alkyl tail

lengths: T_m values were 82.5±0.3°C, 81.2±0.4°C, and 68.1±0.4°C for E1 in 10wt% [Emim]DMP, [Emim]DEP, and [Emim]DBP, respectively (Figure 3.6).

One exception to the hydrophobicity trend occurred with ILs composed of [Emim] or [Bmim] cations and tetrafluoborate anions. In this case, solvent coupling effects were clearly present (i.e. cation and anion effects were not independent). Unlike other [Bmim] ILs, E1 had higher activity in the presence of the longer butyl chain [Bmim]: relative activity was ~55% in [Bmim]BF_4 and ~27% in [Emim]BF_4 at 25wt% concentration (Figure 3.5b). Kosmotropicity failed to rationalize this effect. According to the law of matching water affinity,[178] chaotropic [Emim] cations and chaotropic BF_4 anions would be expected to interact more strongly with each other than with the surrounding solvent, leading to higher than predicted enzyme activity. Instead, the combination of chaotropic [Emim] and chaotropic BF_4 ions reduced enzyme activity more than kosmotropic [Bmim] and chaotropic BF_4 ions.

3.4.4. SUPERCHARGED DESIGNS: SPECIFIC ACTIVITY AND THERMOSTABILITY

To further probe the interactions between ILs and protein surfaces, two variants of E1 (superpositive pE1 and supernegative nE1) were created with significantly altered charge. The specific activities of E1 and pE1 were comparable (23,000±6,000 and 23,000±5,000 μmol substrate/mg protein/min, respectively), while the activity of nE1 was significantly reduced (1,300±900 μmol substrate/mg protein/min). Thermostability measurements in aqueous solution showed that nE1 was also severely destabilized (T_{50}=50.6±0.1°C), whereas pE1 was only moderately destabilized (T_{50}=68.7±0.1°C, Figure 3.9a). For comparison, the T_{50} for E1 was 79.6±0.1°C. Similarly, E1 and pE1 had a longer lifetime than nE1: both E1 and pE1 maintained >80% activity after six days incubation at 50°C (Figure 3.8), while nE1 was completely

inactivated in ≤1 day (data not shown). Despite having differing surface charge and altered pH activity profiles, both supercharged variants had maximum activity near pH 5 (Figure 3.9b). All activity assays, including IL assays, were performed at pH 5.

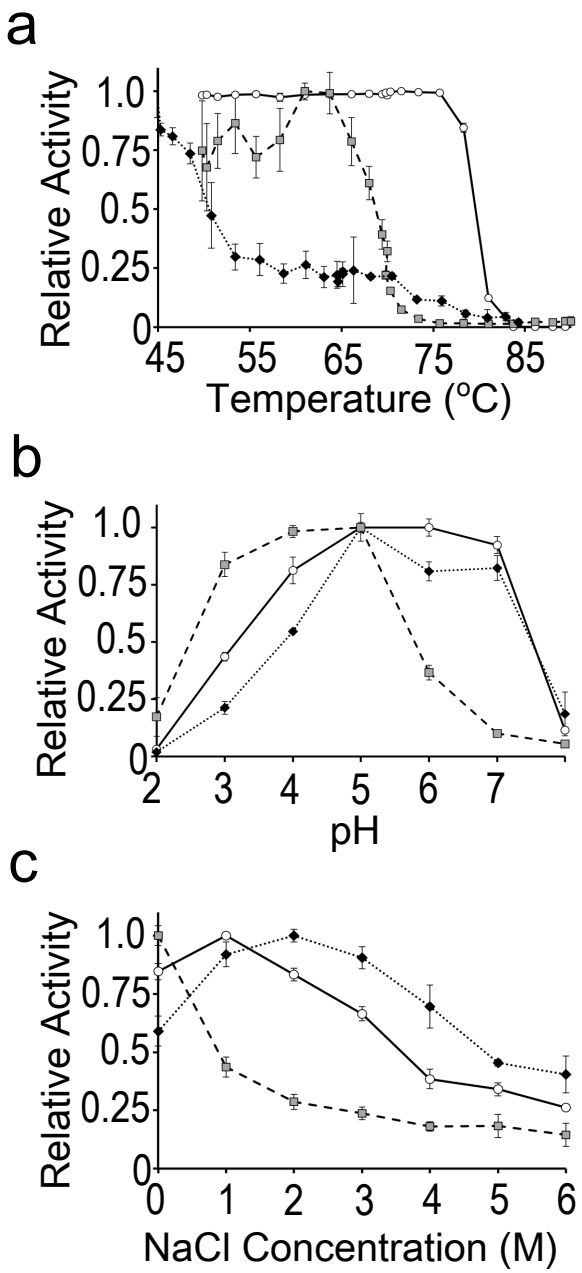

FIGURE 3.9: CHARACTERIZATION OF SUPERCHARGED DESIGNS
(a) Retained activity measured after a 30 minute incubation period provided T_{50} thermostability estimates. **(b)** Designs were evaluated across a range to identify the optimum pH. **(c)** Activity varied significantly in high concentrations of sodium chloride. For all three panels, shading denotes E1 (open circle), pE1 (grey square), and nE1 (black diamond). Error bars indicate the standard deviation of triplicate measurements.

3.4.5. SUPERCHARGED DESIGNS: SALT DEPENDENCE

Enzyme activity was measured in varying concentrations of sodium chloride to assess the impact of small monovalent ions (Figure 3.9c). Wild type E1 was tolerant of NaCl up to ~2 M, while pE1 steadily lost activity in increasing NaCl. Activity of nE1 increased in high concentrations of sodium chloride (up to 2 M). This difference could be caused by the distribution of ions at the protein's surface: a positive surface is likely to attract chloride anions that disrupt native hydrogen bonds. In contrast, the negative surface of nE1 would be more likely to repel chloride anions and preserve activity. Despite this observation, basing a hypothesis purely on direct ion interactions could be misleading. As noted previously, nE1 was significantly less thermostable than E1 and pE1. High salt concentrations might "rescue" unfavorable mutations by screening out electrostatic interactions. If this were true, a less aggressive supercharging strategy, or one that distributed negatively charged mutations differently, might be capable of producing a more active supernegative design. It is also important to consider the proximity of mutations to the active site. Although there were no mutations near the catalytic site (within 15 Å) for pE1, mutations Gln123Glu and Asp165Glu in nE1 were both within 10 Å (α-carbon distance) of key catalytic residue Glu162. In addition, mutation Asp324Glu in nE1 was along the substrate binding cleft and could have altered hydrogen bonding with the substrate.

3.4.6. SUPERCHARGED DESIGNS: ACTIVITY IN IONIC LIQUIDS

Despite having differing thermostabilities and salt dependencies, E1 and pE1 were essentially indistinguishable in IL activity assays (Figure 3.6b). All ILs caused a steady decrease in activity from 0-25wt% IL concentration with the most inactivating being [Emim]OAc and the least inactivating [Emim]DEP. For pE1 in 25wt% [Emim] ILs, the trend of increasing activity

was: OAc < DBP < BF_4 < NO_3 < Cl < DMP < DEP. For nE1, activity decreased rapidly in nearly all conditions tested (excluding [Emim]DMP and [Emim]DEP). It was unclear whether activity loss was due to the lower base enzyme stability or the negative charge of nE1. In previous studies, the chemical addition of negative charge groups led to enhanced activity in ILs.[12–14] Poor IL resistance for nE1 might be caused by the supercharging design method, the native enzyme stability or protein fold, or the properties of the ILs themselves. The relatively high activity of nE1 in some cases (>50% activity in 25wt% [Emim]DEP) indicated cations alone were not responsible for activity loss; anions clearly played an important role in influencing the bulk solvent or enzyme properties.

3.4.7. SUPERCHARGED DESIGNS: STABILITY IN IONIC LIQUIDS

Supercharged design stability varied significantly across the panel of ILs. For example, pE1 showed a slight increase in T_m in [Emim]DMP and [Emim]Cl (Figure 3.6a), but incubation in ILs with longer alkyl chains caused large stability decreases (T_m decreased by 6.6±0.2°C and 16.7±0.4°C in [Bmim]BF_4 and [Emim]DBP, respectively). Notably, several ILs decreased E1 stability but had a less severe impact on pE1. E1 stability decreased by more than 7°C in [Emim]BF_4, [Emim]NO_3, and [Emim]OAc, whereas pE1 stability was only shifted by ~5°C or less in these ILs (Figure 3.6a).

The destabilizing effect upon addition of IL was much more varied for nE1. [Emim]DMP, [Emim]DEP, and [Emim]Cl had only a minor impact on nE1 stability, changing T_m values by 0.2±0.3°C, -0.5±0.2°C, and -0.9±0.2°C, respectively. In some ILs, such as [Emim]OAc and [Emim]DBP, nE1 unfolding occurred even at room temperature, creating a broad SYPRO fluorescence peak with no discernable T_m. For other cases, stability and activity

effects appeared to be anti-correlated: ILs such as [Emim]NO_3 and [Bmim]Cl were only mildly destabilizing (Figure 3.6a) but severely deactivating (Figure 3.6b). This discrepancy implied inactivation mechanisms other than structural unfolding might be present. Active site inhibition provides one candidate explanation for this effect, although other unique inactivation mechanisms could also be present.

3.5. CONCLUSION

Many hydrophilic ionic liquids can have detrimental impacts on enzyme activity. In order to better understand why cellulases rapidly lost activity in ILs, supercharged variants of endoglucanase E1 were tested across a panel of nine different ILs. Surprisingly, anion kosmotropicity did not always correlate with enzyme activity: acetate was a clear exception to this rule. All enzyme variants tested had ≤10% activity in 15wt% [Emim]OAc. It is not yet well understood why [Emim]OAc had such a large negative impact on activity. Molecular dynamics simulations could provide additional insight; follow-up computational studies are currently in progress.

Significant activity loss in all ILs, including *neutral or stabilizing* ILs [Emim]Cl and [Emim]DMP, indicated alternate inactivation pathways were likely present. Previous MD simulations have suggested cations can bind near the active site and reduce enzyme activity.[190,207–209] Unfortunately, aromatic residues or negatively charged residues where cations are likely to bind are often key elements for cellulase activity, potentially creating a fundamental conflict. Simultaneously addressing active site inhibition and enzyme stability could complicate future design efforts but might be necessary for generating IL-tolerant enzymes suitable for combined bio-IL platforms.

4. MOLECULAR DYNAMICS SIMULATIONS OF CELLULASE HOMOLOGUES IN AQUEOUS 1-ETHYL-3-METHYLIMMIDAZOLIUM CHLORIDE

Lucas B. Johnson,[a] *Christopher D. Snow*[a]

This chapter is derived in part from an article published in the *Journal of Biomolecular Structure and Dynamics* on July 8, 2016, available online: http://www.tandfonline.com/10.1080/07391102.2016.1204364.

[a]Author affiliation:

Colorado State University

1370 Campus Delivery

Fort Collins, CO 80523-1370

USA

*Corresponding author contact information:

Christopher.Snow@ColoState.edu

Phone: (970) 491-5276

Fax: (970) 491-7369

4.1. SUMMARY

Pretreating biomass using ionic liquids (ILs) can decrease cellulose crystallinity and lead to improved hydrolysis. However, cellulase activity is often reduced in even low concentrations of ILs, necessitating complete washing between pretreatment and hydrolysis steps. To better understand how ILs interact with enzymes at the molecular scale, endoglucanase E1 from *Acidothermus cellulolyticus* was simulated in aqueous 1-ethyl-3-methylimidazolium chloride ([Emim]Cl). Homologues with differing surface charge were also simulated to assess the role of electrostatic interactions between the enzyme and the surrounding solvent. Chloride anions interacted with the enzyme surface via Coulomb or hydrogen bond interactions while [Emim] cations primarily formed hydrophobic or ring stacking interactions. Cations strongly associated with the binding pocket of E1, potentially inhibiting the binding of substrate molecules. At elevated temperatures, cations also disrupted native hydrophobic contacts and caused some loss of secondary structure. These observations suggested that both cations and anions could influence enzyme behavior and that denaturing and inhibitory interactions might both be important in aqueous IL systems.

4.2. INTRODUCTION

Many commercial goods, including fuels, pharmaceuticals, and cosmetics, are produced using catalytically active biomolecules. Naturally occurring or specially designed enzymes can achieve fast catalytic rates and a high degree of selectivity.[212] The addition of organic or non-aqueous solvents can further improve biochemical reactions by enhancing the solubility of substrates or products.[173] In particular, low melting point ionic liquids (ILs) can be used to adjust solvent polarity or hydrophobicity without inducing a high vapor pressure.[172] ILs have garnered

significant interest for their unusual capacity to dissolve cellulose and lignin, the major components of biomass.[213,214] Pretreating biomass with ILs can improve downstream hydrolysis steps and lead to increased yields.[215] Unfortunately, many cellulases are inactivated by low concentrations of ILs,[175] necessitating excessive washing between pretreatment and hydrolysis steps.

Although the fortuitous discovery of new enzymes can lead to improved stability and activity in ILs,[216,217] a rational approach would be beneficial in creating a more generalized stabilization strategy. Rational design requires a detailed understanding of protein structure and dynamics, and computational simulations are invaluable in predicting this behavior at the molecular scale. Many studies involving cellulose and other carbohydrates have focused on substrate interactions in ILs, but to date only a few computational studies have investigated how ILs influence cellulase enzymes.[106,207] Other proteins, including lipase,[190,195,218–221] protease,[195,222,223] lysozyme,[224] monooxygenase,[209,225] luciferase,[226] xylanase,[208] ubiquitin,[227] and zinc finger proteins,[228] have been simulated in a range of ILs and provide many hypotheses regarding candidate inactivation mechanisms. In neat ILs or highly concentrated mixtures, surface solvation can play an important role in moderating enzyme activity. Denaturation via loss of hydrogen bonds or hydrophobic contacts can also lead to decreased activity. Experimental studies suggest these interactions depend on various IL properties such as polarity, viscosity, kosmotropicity, hydrophobicity, or nucleophilicity.[172,180,191,229] Understanding the interplay between multiple inactivation mechanisms could help isolate important design criteria and guide rational design efforts.

Surface charge modification is one strategy that has led to improved enzyme activity in ILs. Motivated by the apparent correlation between negative surface charge and halophilicity,[217]

Kaar and coworkers chemically modified the charge of cellulase, lipase, papain, and chymotrypsin to enhance stability and activity in ILs.[12–14,230] Changes in protein dynamics and solvent interactions were experimentally measured via nuclear magnetic resonance and fluorescence quenching assays. Furthermore, computational simulations confirmed that charge modification of lipase and chymotrypsin altered ion interactions near the enzyme surface.[195] Exclusion of chloride anions from the enzyme surface appeared to benefit enzyme activity and stability in these cases.

In contrast, several recent studies have focused on cation interactions as a potential inactivation mechanism. Li *et al.* demonstrated via simulation of cellobiohydrolase I from *Trichoderma reesei* that 1-butyl-3-methylimidazolium ([Bmim]) cations could infiltrate into the cellulose binding tunnel and potentially inhibit substrate interactions.[207] Others have noted similar behavior of imidazolium cations in simulations containing different proteins and force fields.[190,208,209] Burney and coworkers also noted strong ion binding interactions in xylanase simulations[218] and experimental assays indicated 1-ethyl-3-methylimidazolium acetate ([Emim]OAc) behaved like a competitive inhibitor.[203] In addition, directed evolution studies of other cellulases identified mutations near the active site that improved IL resistance.[231,231] Taken together, these results emphasize that IL cations might directly influence enzyme activity by blocking or modifying the active site.

In a computational study of Cel5A homologues, Jaeger and Pfaendtner proposed different inactivation mechanisms for each simulated enzyme.[106] A mesophilic cellulase from *Trichoderma viride* had comparable root mean square deviation in water as in 15% [Emim]OAc on the 500 ns simulation timescale. In this case, activity loss was attributed to small structural changes near the binding pocket. Interestingly, a more thermophilic homologue from

Thermotoga maritima had extensive loss of secondary structure and was likely inactivated due to protein denaturation. The most thermostable homologue from *Pyrococcus horikoshii* displayed no significant structural variation in any of the simulations, and the authors speculated that aggregation or higher order effects not evident in molecular dynamics simulations could play an important role.

The multitudes of different IL interactions emphasize how cations and anions can have varying impacts on enzyme activity. To better understand the protein/IL interface, we selected a thermostable cellulase from glycoside hydrolase family 5 (GH5) for further characterization. Endoglucanase E1 from *Acidothermus cellulolyticus* is a well studied cellulase in the GH5 family that has been characterized both experimentally[97,101,109,111–113,151] and computationally.[104,105] The high native stability of E1 ($T_m \approx 84^oC$)[232] makes it a promising candidate for assessing IL tolerance; studies of other Cel5A homologues in ILs have suggested thermostability and IL tolerance might be correlated.[164,205] In our previous work, E1 was experimentally characterized in the presence of eight different aqueous ILs.[232] ILs largely followed Hofmeister series prediction: E1 maintained the highest relative activity in ILs with kosmotropic anions and the lowest relative activity in ILs with chaotropic anions. E1 variants with high net surface charge were also assayed against the panel of ILs. Despite having >90% sequence identity, the activity and stability of these homologues varied significantly: positive supercharging maintained comparable activity levels as wild type while negative supercharging drastically reduced activity. Surprisingly, decreased stability in certain ILs did not always correlate with decreased activity, suggesting alternate inhibition modes were important.

In the present study, computational simulations of the E1 homologues were used to generate hypotheses regarding candidate inactivation mechanisms. For consistency, the same

naming scheme was employed as in Johnson *et al.*[232] with E1 referring to the native endoglucanase, pE1 referring to a positively charged variant, and nE1 referring to a negatively charged variant. All three homologues shared a common $(\alpha/\beta)_8$ barrel structure (Figure 4.1). Key catalytic residues (Glu162, Glu282) located on the C-terminal end of β-strands 4 and 7,[96,233] substrate binding residues along the active site cleft (Arg62, His116, Asn161, His238, Tyr240, Trp319), and disulfide residues (Cys34/Cys120, Cys168/171) were conserved across all three enzymes. Simulations were performed at ambient and high temperatures in aqueous 1-ethyl-3-methylimidazolium chloride ([Emim]Cl) to systematically assess solvent interactions and structural changes at the molecular scale.

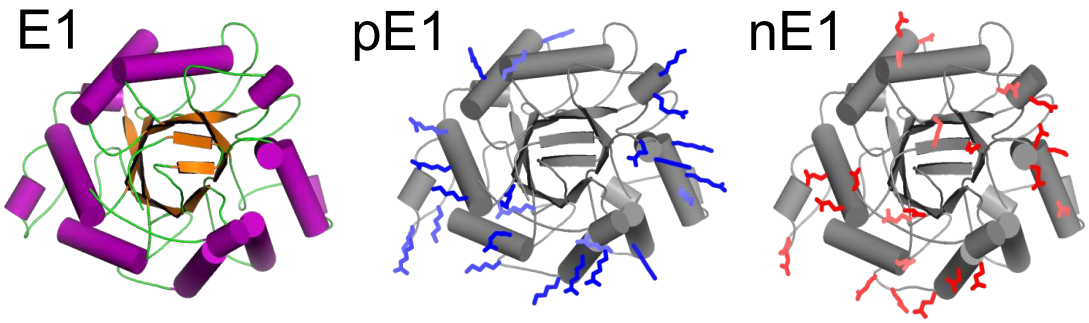

FIGURE 4.1: STRUCTURE OF E1, PE1, AND NE1
Top view of wild type endoglucanase E1 (left), positively charged pE1 (middle), and negatively charged nE1 (right). Colors depict $(\alpha/\beta)_8$ barrel structure for E1: α-helices (purple), β-sheets (orange), and loops (green). Mutations relative to E1 are shown as sticks for pE1 (blue) and nE1 (red). Structures are based on Protein Data Bank structure 1VRX for E1 and homology models for pE1 and nE1.

4.3. METHODS

Parameters for ionic liquid [Emim]Cl (Figure 4.2) were adopted from the generic force field developed by Lopes and coworkers.[234] Despite known limitations such as underestimated diffusion times and overestimated heat of vaporization,[235] this force field has been widely used to represent IL behavior.[221,222,236–241] In our simulations, the Lopes parameters accurately reproduced the density of pure [Emim]Cl (Figure 4.3). For protein atoms, parameters were taken

from the Optimized Potential for Liquid Solvents All Atom (OPLS-AA) force field.[161] OPLS-AA accurately reproduces experimental data for organic solvents[242] and is reasonably accurate in predicting protein dynamics.[243] Most importantly, the OPLS-AA force field is deemed compatible with the generalized IL parameters developed by Lopes *et al.*[234]

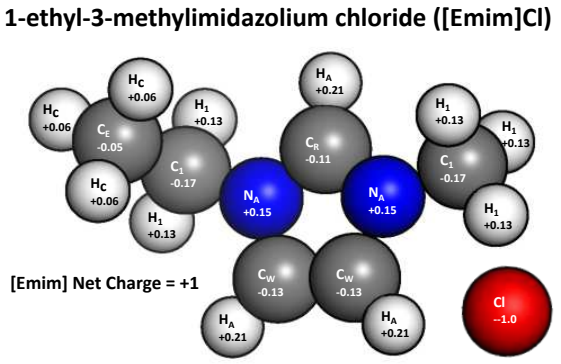

FIGURE 4.2: [EMIM]CL FORCE FIELD PARAMETERS
Cartoon representation of [Emim] cation chloride anion showing atom naming scheme and partial charges as defined by Lopes et al. Color code: carbon (grey), hydrogen (white), nitrogen (blue), chlorine (red).

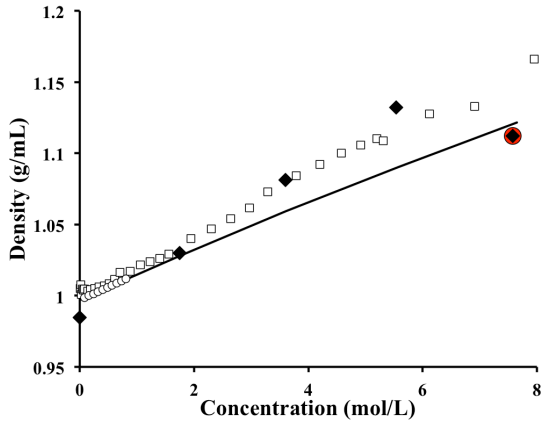

FIGURE 4.3: SIMULATED AND EXPERIMENTAL DENSITIES OF AQUEOUS ILS
Although no experimental densities were available that exactly matched the conditions reported in our study (0.26-3.6 M [Emim]Cl, 300-450 K), several publications provided useful comparison points. Tang *et al.* published data for dilute (up to 0.8 M) solutions of [Emim]Cl in water at 298.15 K (open circles),[244] and Wu *et al.* published densities of a similar IL (1-*allyl*-3-methylimidazolium chloride) across a broad range (0 to 8 M) at 298.15 K (open squares).[245] The line represents predicted densities based on a linear extrapolation from the Tang *et al.* data. Density values calculated from isothermal-isobaric ensemble molecular dynamics simulations at 300 K (black diamonds) followed the same trend as the experimental data. Standard deviations were less than 0.004 g/mL for triplicate simulations at each condition. Despite the simulation temperature of 300 K being below the expected melting temperature of ~353 K, the density of neat [Emim]Cl (1.112± 0.003 g/mL) predicted from our MD simulations closely matched the density of 1.112 g/mL for (liquid) [Emim]Cl at 353 K (red circle) reported in the Sigma Aldrich material safety data sheet.

For each IL-protein simulation, a box with one protein molecule and the appropriate number of [Emim]Cl ions was generated using PACKMOL.[246] The starting structure of E1 was taken from Protein Data Bank structure 1VRX[101] while nE1 and pE1 structures were modeled in SHARPEN[155] with side chain orientations determined by the optimization algorithm FasterPacker. Additional IL cations or anions were added as necessary to balance net charge, making molarity calculations approximate for each simulation. For non-IL simulations containing sodium chloride ions, the initial box was set up using built-in GROMACS features.[160] Each 512 cubic nm box was solvated with TIP3P water molecules.[247] Energy minimization consisted of 10,000 steps of steepest descent minimization *in vaccuo* and again in solvent. Equilibration involved a gradual temperature increase over 40 ps from 1 K to the final simulation temperature using the velocity rescale thermostat,[248] followed by 200 ps at 1 bar using Parrinello-Rahman pressure coupling.[249] One 250 ns simulation and five 50 ns simulations were performed for each condition. Post-simulation analysis was performed using Gromacs tools or custom scripts based on the SHARPEN platform. All-atom root mean square deviation (RMSD) was calculated relative to the equilibrated starting structure after alignment of backbone atoms. Similarly, root mean square fluctuation (RMSF) was calculated after performing backbone alignment and averaged across the last 40 ns of each simulation. Cumulative ion distributions were averaged across the last 40 ns of each simulation using the g_rdf function in Gromacs. Each ion distribution was calculated relative to the protein surface. Distribution functions were calculated as the derivative of the cumulative sum and normalized to correct for the varying number of ions.[250] Salt bridges were initially identified using a Visual Molecular Dynamics Salt Bridges plugin[251] with a default cutoff distance of 3.2 Å between oxygen atoms in acidic residues

and nitrogen atoms in basic residues. Hydrogen bond potential was then evaluated for each salt bridge pair based on Rosetta energy scoring[48] of trajectory snapshots.

4.4. RESULTS AND DISCUSSION

4.4.1. SELECTION OF SIMULATION CONDITIONS

Before beginning production simulations, it was important to identify the simulation time, temperature, and IL concentrations that were most informative. Simulating the system in atomistic detail was computationally expensive and could only sample a fraction of the timescale used in experimental assays. Previous studies have suggested that increasing the simulation temperature can accelerate the rate of unfolding without significantly altering the unfolding pathway[162] and can improve correlations between simulated RMSD and experimental thermostability values.[105] Increasing simulation temperature for thermostable E1 systematically increased both RMSD (Figure 4.4c,d) and RMSF (Figure 4.5c,d) while preserving overall residue level trends. For example, RMSF of the surface loop containing residues 245-258 increased from an average of ~1.2 Å at 300 K to ~5.6 Å at 450 K in 1.32 M (18.9wt%) [Emim]Cl. Other regions that had high RMSF also tended to increase with temperature, preserving the overall RMSF "footprint" of peaks and valleys (Figure 4.5c,d). In general, there was little difference in RMSF when averaged across the last 40 ns of 250 ns simulations compared to the last 40 ns of 50 ns simulations (Figure 4.5a,b). Unique RMSF peaks, such as the one seen at residue 308 in the 50 ns NaCl simulation (Figure 4.5a) were likely due to stochastic unfolding events that occurred early within the simulation. Large structural changes commonly occurred within the first 50 ns for simulations performed at 450 K. At 1.32 M (18.9wt%) [Emim]Cl and 450 K, protein RMSD increased to >6 Å within the first 35 ns and exceeded 11 Å

by the end of a 250 ns simulation (Figure 4.4b). In contrast, protein RMSD remained low (<2.5 Å) at 300 K (Figure 4.4a,b), demonstrating that the simulation was stable and that the protein retained most of its native structure. Despite the relatively high viscosity of IL/water mixtures, the simulation timescale was expected to be sufficient for solvent equilibration.[195] Ion distribution functions were nearly identical when averaged across short 50 ns simulations or long 250 ns simulations. Simulations performed at low temperature (300 K) thus provided a useful model for studying protein-solvent interactions while simulations performed at high temperature (450 K) provided an understanding of how ILs might influence protein denaturation.

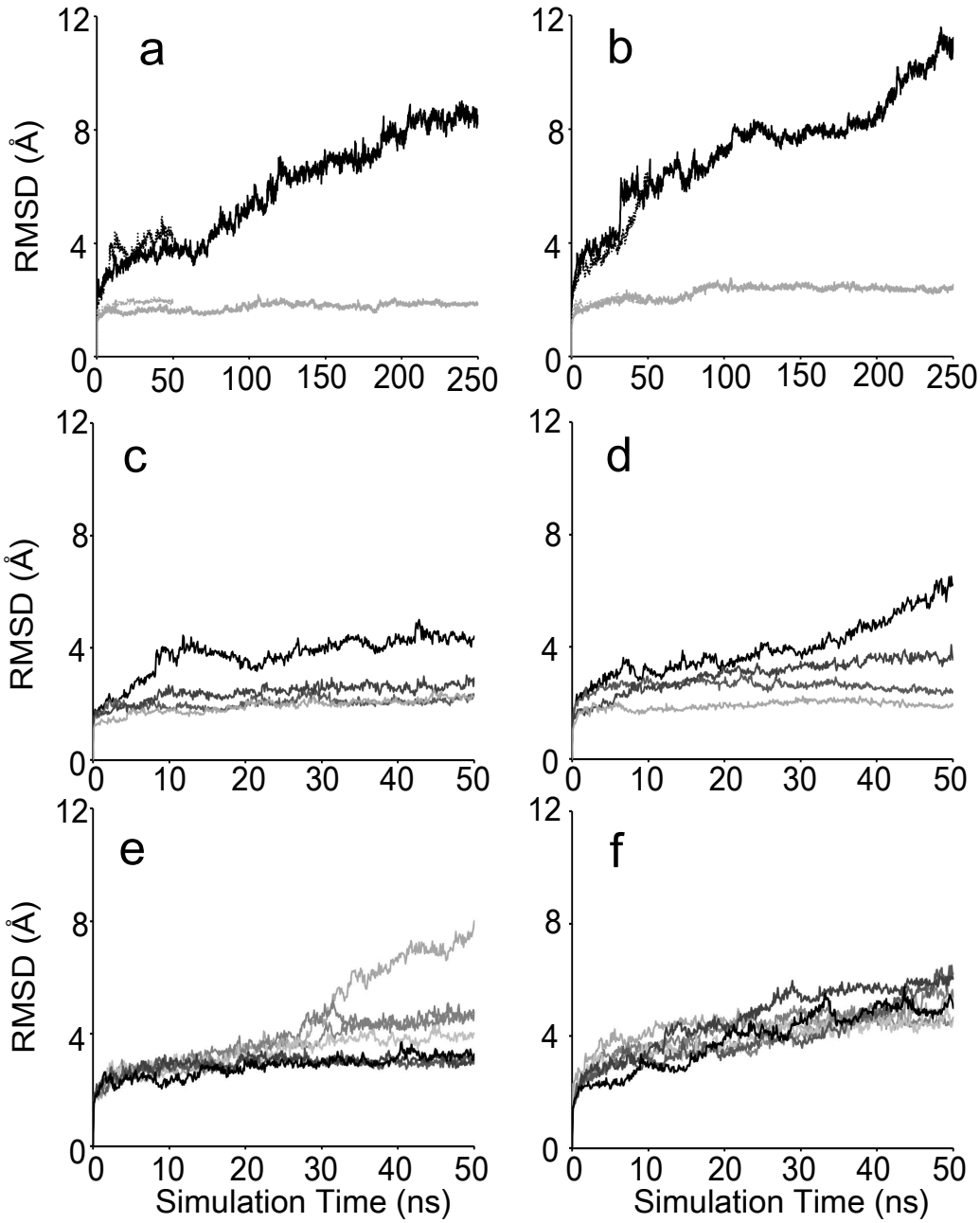

Figure 4.4: ROOT MEAN SQUARE DEVIATION IN MD SIMULATIONS
All-atom root mean square deviation (RMSD) for E1 with **(a)** 0.1 M NaCl or **(b)** 1.32 M (18.9wt%) [Emim]Cl at simulated temperatures of 300 K (grey) and 450 K (black). Short 50 ns and long 250 ns simulations are shown for each condition. Increasing temperature systematically increased RMSD for both **(c)** NaCl and **(d)** [Emim]Cl. Shading denotes simulation temperatures of 300 K, 400 K, 425 K, and 450 K; light grey corresponds with the lowest temperature (300 K) and black corresponds with the highest temperature (450 K). Changing the salt concentration for **(e)** NaCl and **(f)** [Emim]Cl had a less direct impact. Concentrations of 0.26 M, 0.52 M, 0.78 M, 1.05 M, 1.32 M, 1.75 M, and 3.6 M are shown for NaCl and [Emim]Cl with the lowest concentration shaded light grey and the highest concentration shaded black.

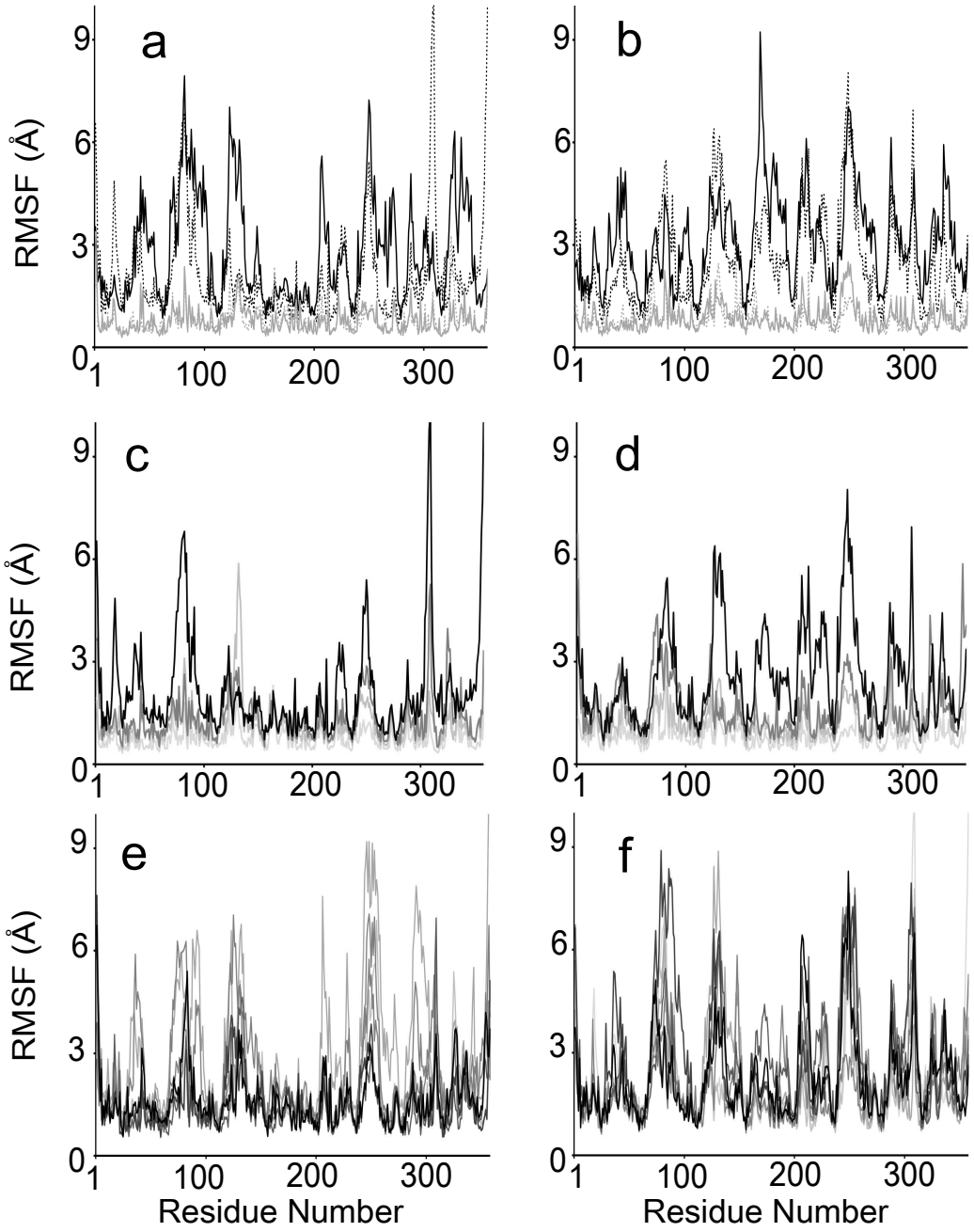

FIGURE 4.5: ROOT MEAN SQUARE FLUCTUATION IN MD SIMULATIONS

All-atom root mean square fluctuation (RMSF) for E1 with **(a)** 0.1 M NaCl or **(b)** 1.32 M (18.9wt%) [Emim]Cl at simulated temperatures of 300 K (grey) and 450 K (black). RMSF was averaged across the last 40 ns of 250 ns simulations (solid line) or 50 ns simulations (dotted line). Increasing temperature systematically increased RMSF for both **(c)** NaCl and **(d)** [Emim]Cl. Shading denotes simulation temperatures of 300 K, 400 K, 425 K, and 450 K; light grey corresponds with the lowest temperature (300 K) and black corresponds with the highest temperature (450 K). Changing the salt concentration for **(e)** NaCl and **(f)** [Emim]Cl had a less direct impact. Concentrations of 0.26 M, 0.52 M, 0.78 M, 1.05 M, 1.32 M, 1.75 M, and 3.6 M are shown for NaCl and [Emim]Cl with the lowest concentration shaded light grey and the highest concentration shaded black.

The concentration of ions in solution would be expected to have a large influence on solvent-protein interactions. Previous reports have suggested high concentrations of ILs can dramatically increase viscosity[252] and decrease protein fluctuations.[208,218] However, we observed no clear correlation between IL concentration and protein RMSF in our simulations (Figure 4.5f). Higher IL concentrations (up to 3.6 M [Emim]Cl) increased RMSF in some regions and decreased RMSF in other regions. Protein RMSD also remained fairly consistent when [Emim]Cl concentration was increased from 0.26 M (3.8wt%) to 3.6 M (50wt%) (Figure 4.4f). This behavior was markedly different in simulations containing NaCl. On average, RMSD and RMSF for E1 increased up to ~0.5 M NaCl and then steadily decreased at higher concentrations (Figure 4.4e, 4.5e). This bell-shaped dependency corresponded with experimentally measured E1 activities in NaCl.[232] Intermediate concentrations of 0.1 M NaCl and 1.32 M (18.9wt%) [Emim]Cl were selected for final production simulations. Experimental studies at similar IL concentrations have shown cellulases can maintain activity[232] and even achieve improved cellulose hydrolysis rates.[253]

4.4.2. PROTEIN UNFOLDING AT HIGH TEMPERATURE

nE1 and pE1 homologues had significantly lower thermal stability than E1 in experimental assays,[232] but the mechanism of destabilization was not well understood. Differences in stability might have been caused by intra-protein, inter-protein, or protein-solvent interactions. Simulating a single protein molecule with periodic boundary conditions failed to address higher order interactions between protein molecules; however, it did provide an opportunity to study local unfolding events and ion interactions at the protein surface. For simulations performed at an elevated temperature of 450 K, the average RMSD at the end of five

replicate 50 ns simulations ($RMSD_{50}$) was 6.9±0.8 Å, 5.8±0.9 Å, and 5.4±0.5 Å for nE1, E1, and pE1, respectively (Figure 4.6). This did not directly correspond with experimental assays: thermostability (apparent T_m) values were reported as 59.0±0.1°C, 84.3±0.4°C, 74.4±0.2°C for nE1, E1, and pE1, respectively.[232] However, the relative change in $RMSD_{50}$ between 0.1 M NaCl and 1.32 M (18.9wt%) [Emim]Cl simulations was consistent with experimental results. pE1 had a higher thermal stability in 0.69 M (10wt%) [Emim]Cl than buffer; correspondingly, pE1 $RMSD_{50}$ decreased in 1.32 M (18.9wt%) [Emim]Cl relative to 0.1 M NaCl simulations. The trend was reversed for E1 and nE1. Thermal stability decreased in the presence of [Emim]Cl in experimental assays, and $RMSD_{50}$ increased by ~1.6 Å in 1.32 M [Emim]Cl simulations relative to 0.1 M NaCl simulations (Figure 4.6).

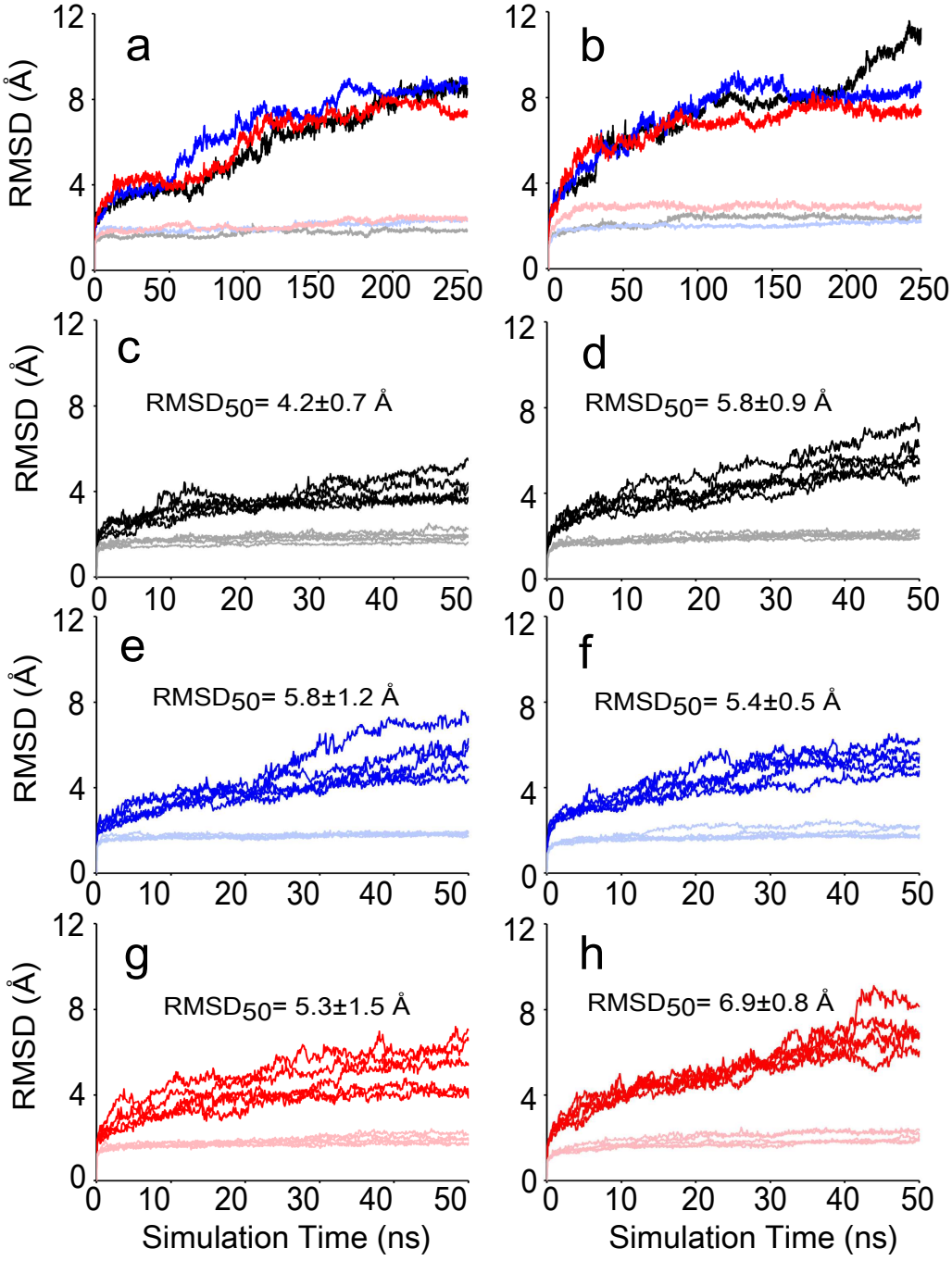

FIGURE 4.6: IONIC LIQUID INTERACTIONS AT THE PROTEIN SURFACE
E1 (black), pE1 (blue), and nE1 (red) were simulated for 250 ns at 300 K (faint color) and 450 K (darker color) in **(a)** 0.1 M NaCl and **(b)** 1.32 M (18.9wt%) [Emim]Cl. In addition, replicate simulations were performed for 50 ns: **(c)** E1 in 0.1 M NaCl, **(d)** E1 in 1.32 M [Emim]Cl, **(e)** pE1 in 0.1 M NaCl, **(f)** pE1 in 1.32 M [Emim]Cl, **(g)** nE1 in 0.1 M NaCl, and **(h)** nE1 in 1.32 M [Emim]Cl. $RMSD_{50}$ indicates the RMSD at 50 ns averaged across five simulations; standard deviations are reported for each case.

Although RMSD provided a useful metric for tracking global protein changes, local unfolding events provided more detailed insight regarding candidate denaturation mechanisms. Numerous small unfolding events contributed to the overall RMSD. One of the more severe cases was selected to discuss in detail. In this example, an [Emim] ion intercalated between a surface loop and the core of the protein and caused reduced hydrophobic packing between residues Trp249 and Ile262 (Figure 4.7). RMSF in this loop region (residues 245 to 258) was consistently higher than background thermal fluctuation for E1, pE1, and nE1, indicating it could be a common unfolding pathway. Unfolding occurred in the same region in 0.1 M NaCl simulations, but the rate and severity of unfolding was exacerbated in the presence of [Emim]Cl. For E1 simulated at 450 K in 3.6 M [Emim]Cl, insertion of an [Emim] cation between Trp249 and Ile262 occurred early (~200 ps). This disruption in native packing appeared to trigger a series of unfolding events throughout the 50 ns simulation. Following the numbered steps in Figure 4.7, the cation first bound to the protein (Step 1), and the surface loop containing Trp249 reoriented to become more solvent exposed (Step 2). The unfolded state was partially stabilized by a hydrogen bond between a chloride anion and Ser251. Notably, the unfolded loop region was adjacent to the Tyr245Gly mutation included in all sequences for reduced product inhibition.[101] Introducing glycine at position 245 could have influenced loop flexibility; however, in 50 ns simulations containing the tyrosine reverse mutation (Gly245Tyr), the behavior of the surface loop appeared comparable to E1 (data not shown). Changes in secondary structure also occurred in the adjacent α-helix containing residues 255 to 275 (Figure 4.7, Step 3). By the end of the simulation, a neighboring β-turn region containing residues 203 to 212 had begun to unfold (Step 4), causing significant loss in secondary structure. Tracking this single stochastic event demonstrated how ILs might disrupt native packing and destabilize enzymes. Interestingly, NMR

studies of lipase in [Bmim]Cl[230] also identified ion binding sites in flexible, solvent exposed loop regions, indicating these features might be good design targets for future stabilization strategies.

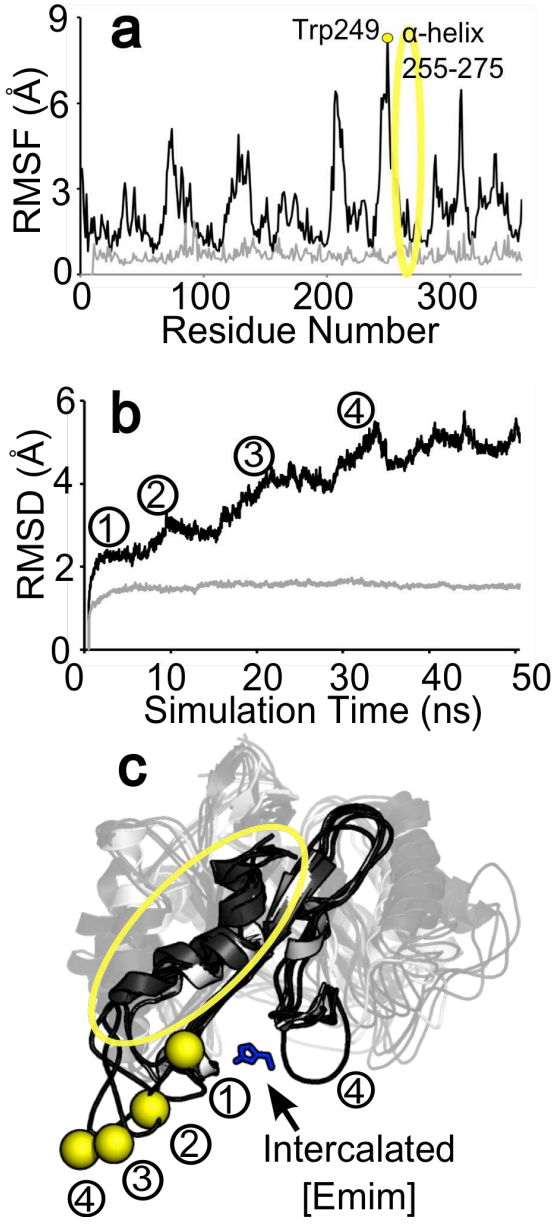

FIGURE 4.7: MD UNFOLDING EVENTS FOR E1 IN [EMIM]CL
Unfolding events tracked via **(a)** root mean square fluctuation and **(b)** root mean square deviation for E1 in 3.6 M [Emim]Cl simulated at 450 K (black) or 300 K (grey). Labeled residues and circled numbers correspond to **(c)** snapshots taken during the simulation. Protein Data Bank structure 1VRX is shown in white for reference. The α-carbon of Trp249 (yellow spheres) deviated significantly from its initial position. Intercalation of [Emim] cation (blue sticks) occurred early in the simulation (Step 1). Loss of secondary structure (Steps 2,3) was evident in the surface helix containing residues 255 to 275 (highlighted by yellow ellipse). Late time points (Step 4) also showed unfolding of a neighboring β-turn (residues 203-212, bottom right).

Salt bridge interactions could also account for some of the difference in relative protein stability. In previous experimental studies, the introduction of a single salt bridge (Arg19, Glu356) increased stability by ~2°C in buffer[104] and in 0.69 M [Emim]Cl or 0.59 M [Emim]OAc (unpublished data). However, when simulated in [Emim]Cl, residues Arg19 and Glu356 formed hydrogen bond interactions in <6% of the snapshots, suggesting these mutations might provide IL stability benefits via solvent interactions rather than salt bridge formation. The number of salt bridges also varied between homologues nE1 and pE1. On average, pE1 formed ~3 more salt bridges than nE1 and ~7 more than E1. Despite having numerous surface charges, nE1 had a low primary amine to carboxylic acid ratio and could not form as many favorable salt bridge interactions as pE1. Analysis of MD simulations showed that 1.32 M [Emim]Cl may have disrupted weaker salt bridge interactions, but the number of strong hydrogen bond forming salt bridges (defined as having a favorable Rosetta hydrogen bond score for >25% of the simulation) remained approximately constant in 0.1 M NaCl and 1.32 M [Emim]Cl.

4.4.3. ION INTERACTIONS AT THE PROTEIN SURFACE

Introducing charged mutations on the surface of E1 significantly altered interactions with the surrounding solvent. For all three homologues, the distribution function (i.e. differentiated cumulative sum) of chloride ions displayed a sharp peak ~1.9 Å from the enzyme surface and a smaller peak ~4.3 Å from the enzyme surface (Figure 4.8a). In contrast, the particle density of [Emim] cations was highest ~3.5 Å away from the protein surface (Figure 4.8b). This behavior was consistent with previous [Bmim]Cl simulations[195] and suggested ion density gradients formed around the enzyme. Small, high charge density anions penetrated closest to the protein surface, while bulky, low charge density cations interacted at a slightly longer distance.

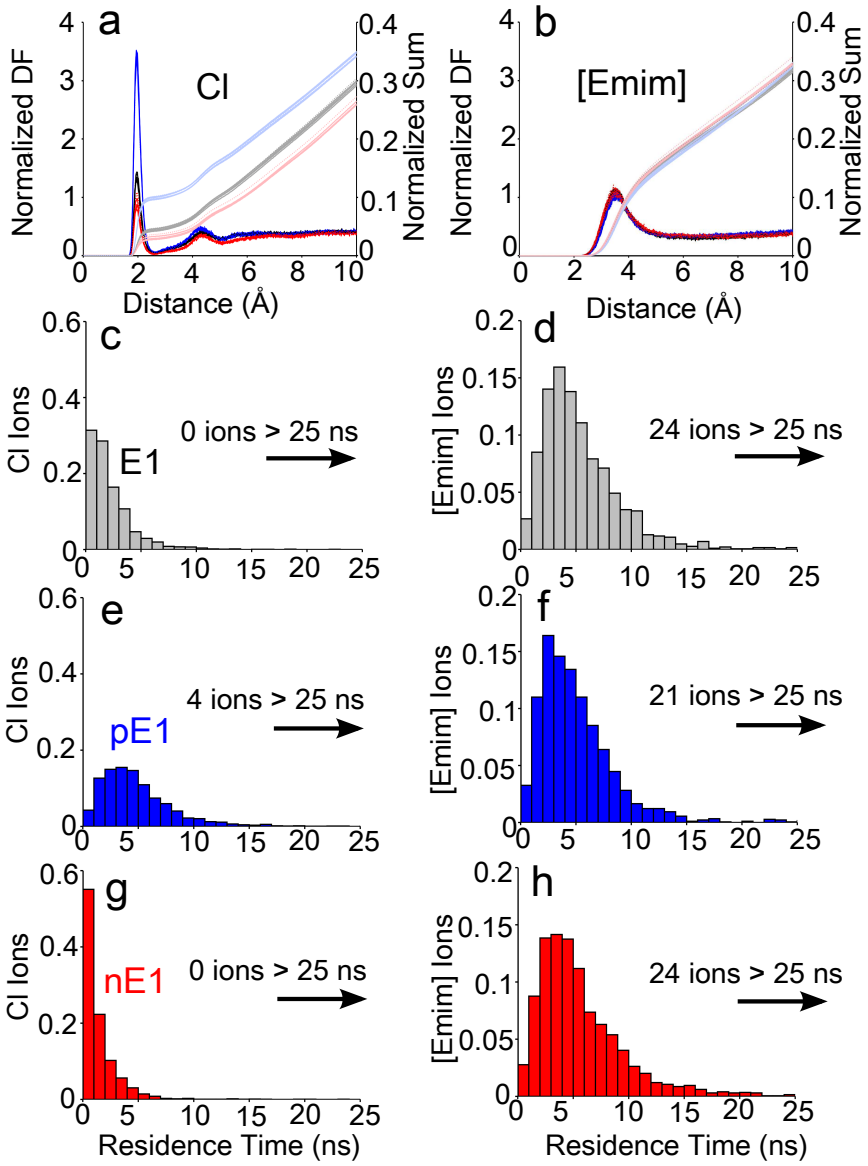

FIGURE 4.8: IONIC LIQUID INTERACTIONS AT THE PROTEIN SURFACE
[Emim]Cl interactions with E1 (grey), pE1 (blue), and nE1 (red). Normalized distribution functions (dark color) and cumulative sums (light color) are shown for **(a)** Cl anions and **(b)** [Emim] cations relative to the protein surface. Distribution functions were normalized to account for varying number of ions and were nearly identical for 250 ns simulations (dotted line) or five replicate 50 ns simulations (solid lines). The normalized distribution of ion binding times, or time each ion spent within 2.5 Å of the protein surface, are shown for Cl and [Emim] ions interacting with **(c,d)** E1, **(e,f)** pE1, and **(g,h)** nE1. Distributions represent aggregate residence times across five replicate 50 ns simulations. All simulations shown were performed at 300 K with 1.32 M (18.9wt%) [Emim]Cl.

The intensity of these interactions varied between enzymes. In general, chloride anions strongly associated with the positive surface of pE1 and weakly associated with the negative surface of nE1. The peak particle density of anions, measured at a distance of 1.9 Å from the

protein surface, was approximately threefold higher for pE1 than nE1. Similarly, the distribution of ion binding times shifted based on protein charge, with chloride being less likely to remain bound (shorter average residence time) for nE1 than either E1 or pE1 (Figure 4.8 c,e,g). Chloride anions with long residence times (10-35 ns) typically favored protein sites containing positively charged (Arg/Lys) or hydroxyl (Ser/Thr) moieties. In particular, residues Arg11 and Arg304 for E1, Ser352 and Ser353 for pE1, and Arg11 and Arg304 for nE1 strongly bound chloride anions in multiple simulations (Figure 4.9). At 300 K, these interactions did not result in protein denaturation; however, at higher simulation temperatures, anion binding might trigger protein unfolding or stabilize unfolded intermediate states. Both charge-charge interactions and hydrogen bond interactions appeared to be important in moderating anion affinity to the enzyme surface.

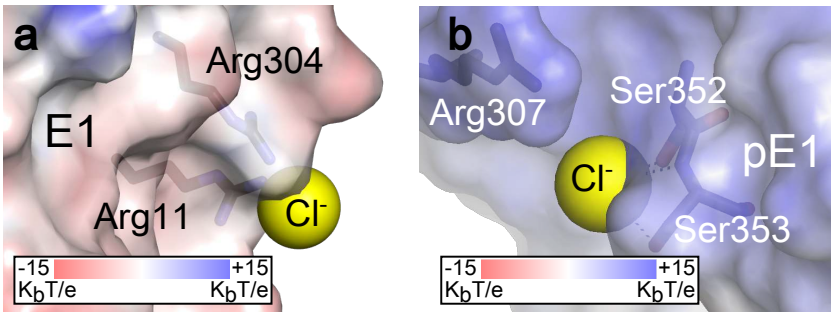

FIGURE 4.9: ANION ELECTROSTATIC AND HYDROGEN BOND INTERACTIONS
(a) Chlorine anions commonly bound near positively charged residues, especially residues Arg304 and Arg311 (shown as sticks) in E1. The displayed snapshot was taken at time t=20 ns from a simulation of E1 in [Emim]Cl. **(b)** Interactions with Ser residues (shown as sticks) also resulted in long anion residence times, potentially due to hydrogen bond interactions (denoted by dashes). The displayed snapshot was taken at t=40 ns from a simulation of pE1 in [Emim]Cl. Simulations from both (a) and (b) were performed in 1.32 M (18.9wt%) [Emim]Cl at 300 K. Protein surface coloration was based on electrostatic potentials calculated using Delphi[202] with a solvent dielectric constant of ε=64[254] and Parse force field parameters.[255]

The trend for [Emim] cations was much less distinct. When normalized to account for varying number of cations, there was little to no difference in the distribution function for the three enzyme variants. The lower than expected electrostatic effect suggested alternate binding

interactions were important for [Emim] cations. Previous reports have indicated enzyme inhibition could not be entirely explained by electrostatic interactions,[179] and docking studies have suggested imidazolium cations bound primarily via hydrophobic interactions.[256] A few MD studies have proposed non-electrostatic binding interactions as well. Most notably, Klähn *et al*. found IL anions interacted with the surface of lipase B via Coulomb interactions and cations interacted with the protein core primarily via van der Waals interactions.[220] Jaegar and Pfaendtner also noted that acetate and ethyl sulfate anions were attracted to positively charged residues in a family 11 xylanase, but [Emim] cations did not show specific affinity for negatively charged residues. Instead, [Emim] cations bound near the active site for the majority of the 500 ns simulation.[208]

Similar cation interactions were seen in simulations of E1. Histograms of [Emim] binding times (Figure 4.8) indicated surface residence times ranged from shorter than 1 ns to longer than 25 ns. By selecting individual cations and tracking their trajectories with 20 ps resolution, it was readily apparent that different types of interactions were formed. Freely diffusing cations only associated with the protein surface for short intervals (Figure 4.10a), while weakly attractive interactions resulted in more sustained binding events (Figure 4.10b). Some [Emim] ions strongly associated with the protein surface, remaining bound within 2.5 Å for almost the entirety of the 50 ns simulation (Figure 4.10c). There were several strongly associated [Emim] cations in every simulation (Figure 4.8). The average number of long-term cation binding events (>25 ns residence time) averaged across five replicate 50 ns simulations was 4.8±1.3, 4.2±0.8, and 4.8±1.7 for E1, pE1, and nE1, respectively. Despite having different surface charges, the region near the active site where cations commonly bound was largely conserved between homologues. Mutations present in nE1 (Gln123Glu, Asp165Glu, Asp324Glu) did not drastically

alter the electrostatics of the binding pocket. [Emim] cations which bound along the substrate binding cleft appeared to form hydrophobic or ring stacking interactions with aromatic residues. Cation residues remained bound within 2.5 Å for >50% of the simulation time at conserved GH5 amino acids His238, Tyr240, Glu282, and Trp319. Although no quantitative statement could be made regarding competitive inhibition without explicitly simulating substrates and products, comparisons showed cation binding within the active site occurred more frequently and for longer duration than other surface sites. A sampling of [Emim] states near aromatic residue His238 provided a qualitative depiction of the strong anisotropy in binding orientation due to favorable stacking interactions (Figure 4.11).

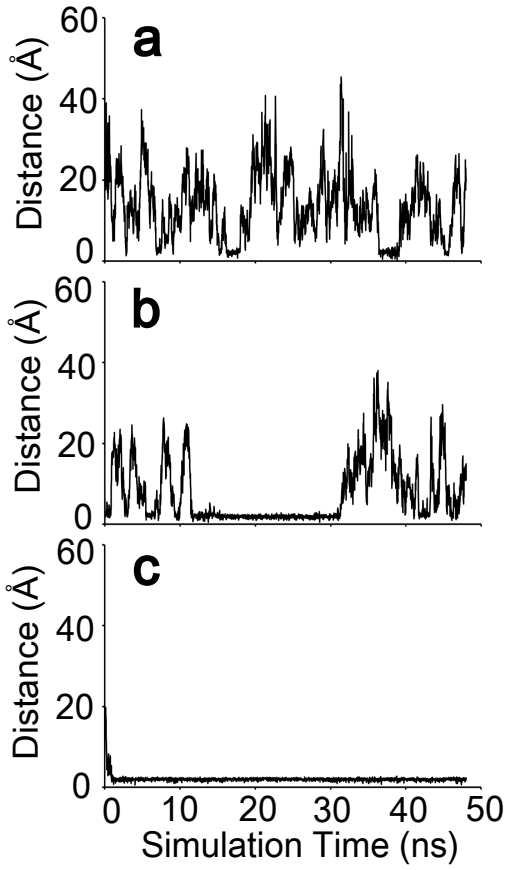

FIGURE 4.10: DETAILED [EMIM] BINDING TO THE PROTEIN SURFACE
Tracking individual trajectories of [Emim] cations indicated several types of interactions were formed: **(a)** minimal interactions with the enzyme surface resulted in free diffusion, **(b)** hydrophobic or aromatic interactions typically led to medium length binding events of 10-20 ns, or **(c)** strong interactions near the active site trapped ions for 40-50 ns.

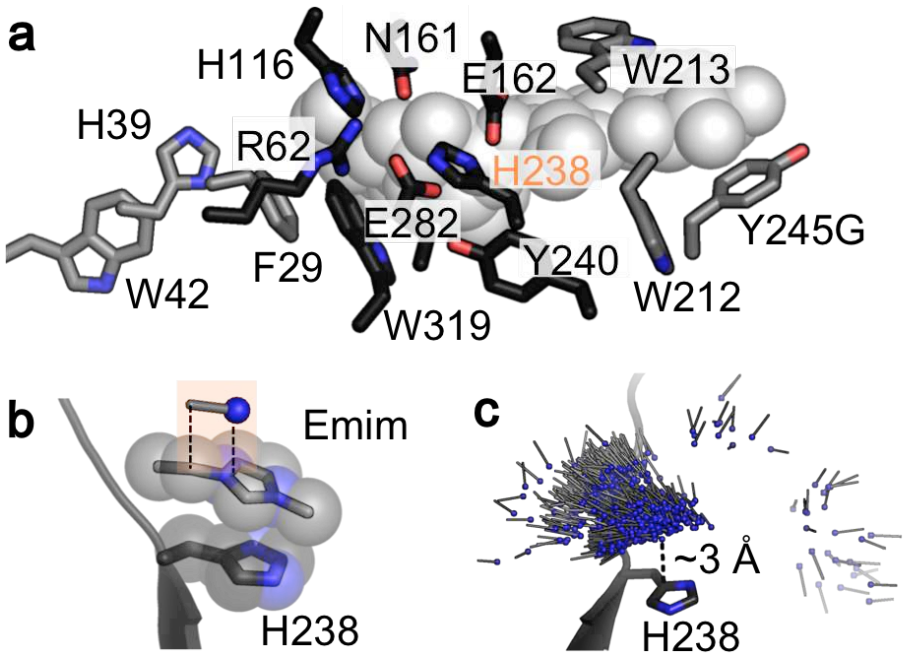

FIGURE 4.11: DETAILED [EMIM] BINDING TO THE PROTEIN SURFACE
(a) A close up of the active site shows key catalytic and substrate binding residues. Residues conserved within glycoside hydrolase family 5 are colored black. Grey spheres indicate the alignment of cellotetraose substrate analogue from Protein Data Bank structure 1ECE. **(b)** A representative [Emim] snapshot shows the alignment of an [Emim] cation relative to His238. **(c)** An example trajectory of a single [Emim] ion is depicted with snapshots taken every 20 ps. Each line represents the vector from aliphatic carbon C_1 (grey) to imidazolium nitrogen N_A (blue sphere) as depicted in the inset of (b). For this example, E1 was simulated in 1.32 M (18.9wt%) [Emim]Cl at 300 K.

4.5. CONCLUSION

Despite having many potential benefits in biotechnology, the utility of ionic liquids is often limited by enzyme incompatibility. Cellulase enzymes in particular can suffer from severe inactivation in ILs.[175] Simulating the behavior of endoglucanase homologues in [Emim]Cl provided a better understanding of potential inactivation mechanisms. In low temperature simulations, binding of IL molecules to the protein surface only caused minor conformational changes. Chloride anions bound primarily via electrostatic interactions while [Emim] cations appeared to bind via a combination of hydrophobic and ring stacking interactions. The introduction of negatively charged surface mutations in nE1 did reduce anion affinity. However, in contrast, the introduction of positively charged surface mutations in pE1 had little to no effect

on cation affinity. Cations strongly bound near the active site of E1, potentially inhibiting substrate interactions. [Emim] cations also appeared to destabilize proteins by disrupting native packing, and high temperature simulations sampled several candidate unfolding pathways.

Molecular dynamics simulations indicated ionic liquids could interact with enzymes through a variety of mechanisms, but the impact of these interactions will likely be unique for each IL/enzyme system. For thermostable E1 in low concentrations of [Emim]Cl, cation interactions were clearly important and influenced both protein denaturation and active site inhibition. Increased anion affinity did not correlate with decreased activity[232] for this particular system. These results might be different for a less stable protein or a more strongly denaturing IL. The properties of ILs can vary significantly based on ion composition, and their respective interactions with proteins should be studied on a case-by-case basis. Future design efforts will likely need to simultaneously address protein stability and active site inhibition in order to improve E1 activity in [Emim]Cl.

5. ENZYME IMMOBILIZATION IN POROUS PROTEIN CRYSTALS

Lucas B. Johnson,[a] Sehoo Park,[a] Thaddaus Huber,[a] Christopher D. Snow[a]

This chapter has not previously been published at the time of submission but may be published during the allowed embargo period. Terms and conditions will be subject to the journal of publication.

[a]Author affiliation:

Colorado State University

1370 Campus Delivery

Fort Collins, CO 80523-1370

USA

*Corresponding author contact information:

Christopher.Snow@ColoState.edu

Phone: (970) 491-5276

Fax: (970) 491-7369

5.1. SUMMARY

Carboxysomes, cellulosomes, and other impressive scaffolds seen in Nature demonstrate the many advantages of ordered biological systems. However, achieving similar levels of structure and control can be exceedingly difficult for *in vitro* assemblies. The interconnected solvent channels found in some porous protein crystals provide a template for selectively capturing and ordering guest molecules. To demonstrate the utility of this approach, stabilized crystals made from a *Campylobacter jejuni* protein were loaded with fluorescently labeled albumin, peroxidase, reductase, and other guest proteins. Macromolecule diffusion and retention within the crystal scaffold were assessed using confocal microscopy. Enzymes immobilized within the crystal via non-covalent interactions maintained their activity in single-crystal and bulk-assay formats. Furthermore, coating the surface of the crystals with a crosslinked albumin hydrogel successfully sequestered a multi-enzyme pathway and excluded off-pathway enzymes present in the surrounding solution. The potential applications of enzyme-laden crystals as sensing devices, delivery capsules, or microreactors motivate future development of this technology.

5.2. INTRODUCTION

Structured biological systems have evolved to provide a variety of benefits in living cells. Designed or natural compartments can be used to balance metabolic flux, prevent off pathway reactions, or isolate toxic intermediates.[257,258] Creating structured scaffolds *in vitro* can yield similar benefits for isolated systems. Enhancements such as substrate channeling,[259] protection from protease degradation,[260] and coenzyme recycling[261] make scaffolds particularly appealing for immobilizing enzymatic pathways. Biological scaffolds created from nucleic acids,[22] viruses

and protein cages,[23] or other supramolecular assemblies can position macromolecules with nanometer scale precision and have myriad applications in imaging, sensing, and drug delivery.[262]

Porous protein crystals offer an alternative immobilization matrix with several advantageous features. Like other protein-based materials, the presence of numerous amino acid functional groups permits scaffold modification or attachment of guest molecules. The structured biological matrix is also capable of spanning multiple size scales without sacrificing resolution and control. Crystal diameters can range from hundreds of nanometers to hundreds of micrometers, yet the ordered matrix can be analyzed with atomic precision using x-ray diffraction.[263]

Diverse protein crystals have been soaked with cofactors, drugs, or other small molecules in x-ray diffraction experiments; however, only highly porous protein crystals with a connected diffusive network are capable of accommodating macromolecular guests. Our previous work with a putative periplasmic polyisoprenoid-binding protein from *Campylobacter jejuni* (CJ) identified a hierarchical pore network that could selectively bind and load gold nanoparticles[264] or fluorescent proteins.[265] When crystallized in high salt concentrations, CJ forms large (13 nm) axial pores and small (~3 nm) perpendicular pores (Figure 5.1) with a repeating P622 space group. The porous CJ network mimics mesoporous structures commonly seen in nature[266] and permits diffusion of substrates, enzymes, and products.

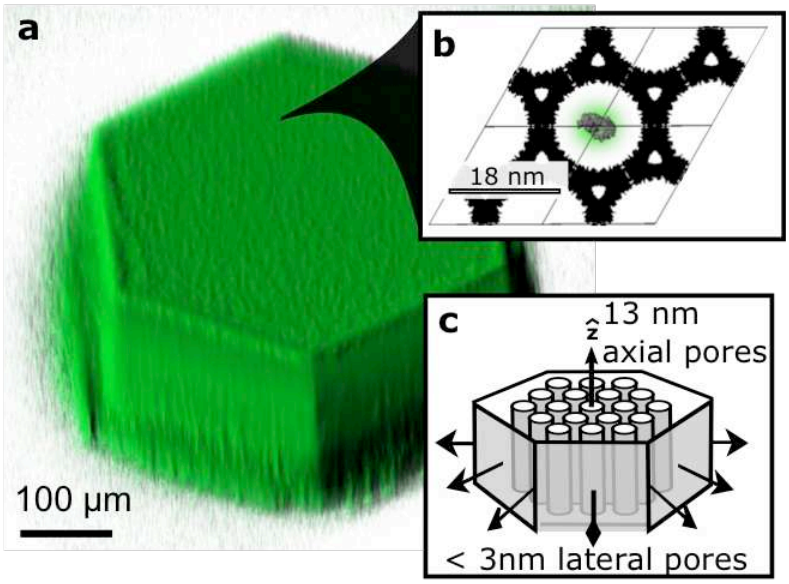

FIGURE 5.1: FLUORESCENT ENZYME LOADING, UNIT CELL RECONSTRUCTION, AND AXIAL PORE NETWORK

(a) Volume reconstruction from image stack obtained via confocal microscopy. **(b)** Representation of four unit cells from a CJ crystal (Protein Data Bank code 2FGS) with a single HRP molecule (green) modeled within the central nanopore. **(c)** Schematic of a hexagonal protein crystal showing the direction of axial and lateral pores.

Unlike crosslinked enzyme crystals or crosslinked enzyme aggregates,[15] enzymes that are soaked into or bound within the CJ scaffold remain accessible to small molecule substrates and are not locked in a single conformation. To demonstrate the utility of CJ crystals in binding multiple enzymes, we engineered host-guest crystalline matrices containing horseradish peroxidase (HRP), glucose oxidase (GOX), xylose reductase (XR), glycerol dehydrogenase (GDH), bovine serum albumin (BSA), or other proteins (Figure 5.2). Analysis via confocal microscopy confirmed that fluorescently tagged proteins could load throughout CJ crystals and form functional microreactors.

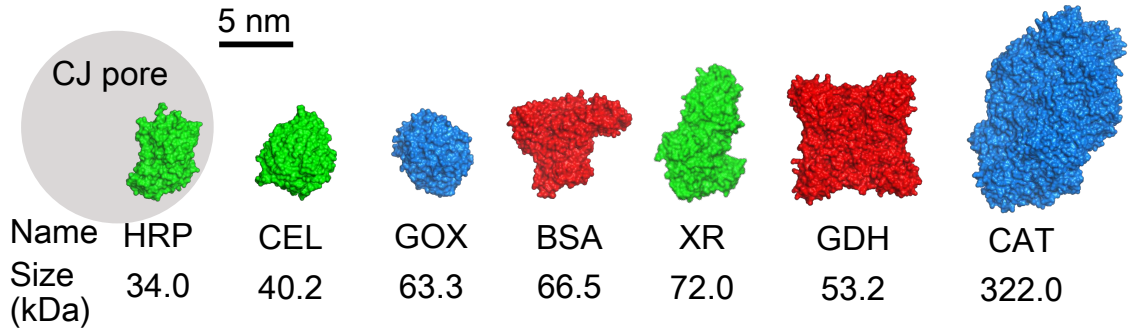

FIGURE 5.2: SURFACE REPRESENTATION OF PERTINENT GUEST PROTEINS USED FOR CJ CRYSTAL SOAKING

From left to right: horseradish peroxidase (HRP), cellulase (CEL), glucose oxidase (GOX), bovine serum albumin (BSA), xylose reductase (XR), glycerol dehydrogenase (GDH), and catalase (CAT). Coloring denotes the fluorescent probe used for imaging. Models were based on Protein Data Bank structures 2ATJ,[267] 1ECE,[111] 1CF3,[268] 3V03,[269] 1JEZ,[270] 4MCA,[271] and 4AUE[272] (homology model constructed via SWISS-MODEL[273]), respectively. The biologically relevant unit is shown for each case; all proteins were predicted to be monomeric except XR (dimer), GDH (tetramer), and CAT (tetramer). The grey circle indicates the average diameter (13 nm) of the central nanopore in CJ crystals.

5.3. MATERIALS AND METHODS

5.3.1. PROTEIN EXPRESSION AND PURIFICATION

Horseradish peroxidase (HRP, Pierce™ #31490), HisProbe™-HRP conjugate (hHRP, Pierce™ #15165), glucose oxidase from *Aspergillus niger* (GOX, SigmaAldrich #G7141), and catalase from *Aspergillus niger* (CAT, SigmaAldrich #C3515) were purchased from commercial suppliers and used without further purification.

A codon optimized gene encoding the *Campylobacter jejuni* protein (CJ, GenBank ID CJ0420) was cloned into pSB3 vector as decribed previously.[265] Codon optimized versions of xylose reductase from *Candida tenuis* (XR, GenBank ID AF074484) and glycerol dehydrogenase from *Serratia marcescens* (GDH, GenBank ID WP_060445076.1 with mutations V15A, F33L, R53Q, D153S, S154V, N159K, M284V, E327Q, A360T) genes were synthesized as gBlocks Gene Fragments (Integrated DNA Technologies, Coralville, IA, USA) and cloned into pSB3 expression vector using ligase independent cloning.[274] CJ, XR, and GDH were

expressed in *E. coli* strain BL21(DE3) using 1 mM IPTG induction at 30°C for 16 hours followed by cell lysis via sonication. Target proteins were purified from cell lysate via immobilized metal affinity chromatography and assessed via sodium dodecyl sulfate polyacrylamide gel electrophoresis.

5.3.2. PROTEIN CRYSTALLIZATION AND CROSSLINKING

CJ was crystallized via sitting drop vapor diffusion by mixing 1 µl of ~15 mg/mL purified protein with 1 µl 90% Tadsimate, 10% glycerol at pH 6.5. Tadsimate is a modified version of Tacsimate™ (Hampton Research) that lacks primary amines.[265] After overnight growth, CJ crystals were crosslinked using glyoxal to improve crystal stability. Glyoxal crosslinking steps consisted of a 30 minute wash in crystallization buffer, a two hour crosslink at room temperature in crystallization buffer (pH 7.5) supplemented with 1% glyoxal and 25 mM borane dimethylamine complex (DMAB), and a two hour quench in 1 M hydroxylamine with 100 mM DMAB. This series of steps was found to optimize crystal stability and pore accessibility in our previous work.[265] In scenarios where pore *inaccessibility* was desired (e.g. containment of GOX or exclusion of CAT), crystals were briefly swished through a solution of 1 mg/mL bovine serum albumin (BSA) with 1% glutaraldehyde to create a surface hydrogel matrix. The time each crystal was incubated in BSA/glutaraldehyde determined the extent of surface coating (Figure 5.3).

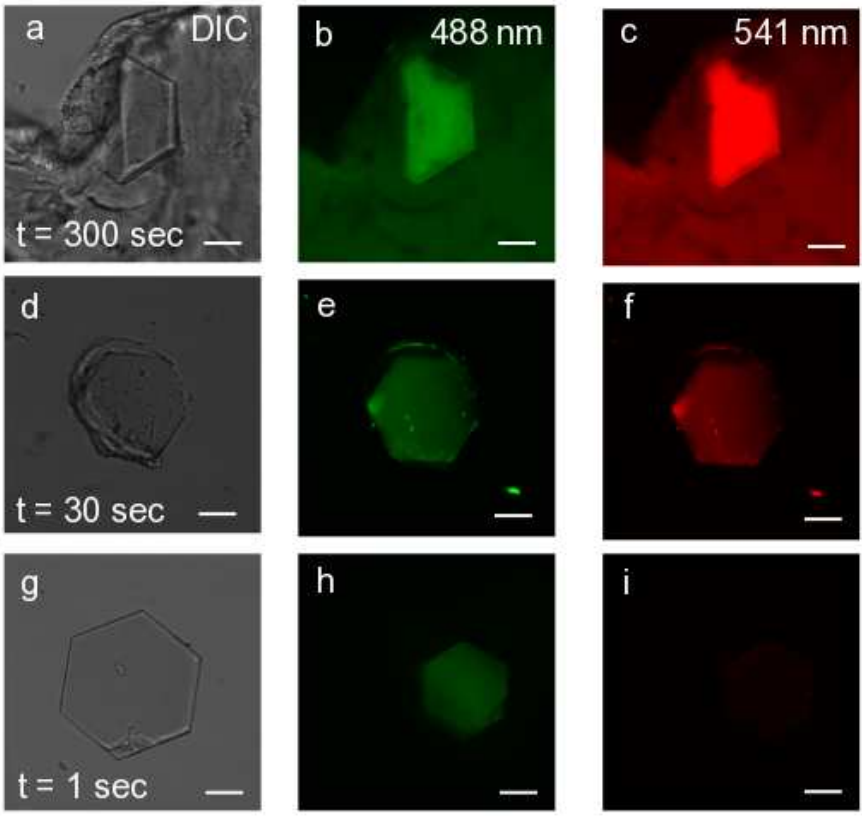

FIGURE 5.3: SURFACE COATING OF CJ CRYSTALS
Crystals were soaked in BSA/glutaraldehyde solutions for **(a-c)** 300 sec, **(d-f)** 30 sec, or **(g-i)** 1 sec. The hydrogel matrix was visualized via autofluorescence at 488 nm and 561 nm wavelength excitation.[275] Scale bars (lower right) are 100 μm.

5.3.3. CRYSTAL LOADING AND MICROSCOPY

Enzymes were fluorescently labeled using blue (CF™405S succinimidyl ester, Biotium #92110), green (NHS-fluorescien, ThermoFisher #46410), or red (CF™594 SE, Biotium #92132) dyes. Enzyme labeling colors are represented in Figure 5.2: green was used for XR, CEL, HRP, and hHRP; blue for GOX and CAT; red for GDH and BSA. Prior to soaking in enzymes, each crosslinked CJ crystal was incubated overnight in BufferA (10% glycerol, 50 mM HEPES, 300 mM NaCl, pH 7.5) supplemented with 10 mM $ZnSO_4$. Protein crystals were then looped into ~1 mg/mL enzyme solutions in BufferA (HRP, hHRP, CAT, BSA), 100 mM

potassium phosphate pH 7.5 (XR, GDH), or 50 mM sodium acetate pH 5 (CEL) and allowed to equilibrate at room temperature for at least one hour.

To permit fluid flow during imaging, crystals were entrapped in a custom polydimethylsiloxane (PDMS) chamber with a gradient channel depth (Figure 5.4). The microfluidic device was cast from layers of Duck EZ Start Crystal Clear Packaging Tape stacked on a silicon substrate. The polymeric matrix was made by mixing PDMS prepolymer (SYLGARD 184 ®) and curing agent in a 10:1 ratio and curing at 80^{o}C for 30 minutes. Crystals were looped into the chamber and covered with a glass coverslip for imaging. Fluid flow was controlled via pipet insertion at the inlet or outlet hole.

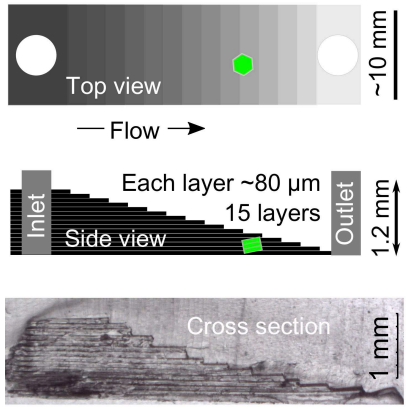

FIGURE 5.4: SCHEMATIC OF MICROFLUIDIC DEVICES USED TO CAPTURE PROTEIN CRYSTALS
(a) Top view showing direction of fluid flow and physically entrapped protein crystal (green). **(b)** Side view depicting channel with gradient depth. **(c)** Microscopic view of PDMS cross-section.

Loaded CJ crystals were imaged on a Nikon Eclipse Ti spinning-disk confocal microscope with an Andor iXon Ultra 897U EMCCD camera. Blue, green, and red fluorescent dyes were imaged using 405 nm, 488 nm, or 561 nm excitation wavelengths, respectively. Laser intensity was set to 20% for enzyme detection or 10% for product detection. All images were collected with a 100 ms exposure time and three-frame image averaging.

5.3.4. ENZYME ACTIVITY ASSAYS

The activity of hHRP was assessed on the substrate AmplexRed (ThermoFisher #A12222) by tracking resorufin product fluorescence at 561 nm wavelength excitation. For each reaction, a single enzyme-laden crystal was incubated with 50 μM AmplexRed and 100 μM H_2O_2 (or 20 μM AmplexRed and 10 mM glucose for hHRP/GOX assays). To assess the extent of GOX/hHRP channeling, varying concentrations of CAT were added for some *in vitro* and *in crystallo* assays. Activity profiles were generated by averaging fluorescence intensity across a 100 μm x 100 μm square in the center of each crystal.

XR and GDH activity was assessed in bulk format using five medium size crystals (~200 μm diameter) per well. Prior to the start of each assay, crystals were soaked overnight in 10 mM $ZnSO_4$, loaded for at least one hour in ~1 mg/mL enzyme solution, and washed for two minutes in BufferA to remove residual surface bound enzyme. Conditions were optimized for maximum activity as follows: GDH, 1 mM NAD^+ and 100 mM glycerol in 100 mM K_2PO_4 pH 9 at 60°C; XR, 1 mM NADH and 100 mM xylose in 100 mM K_2PO_4 pH 7.5 at 25°C; combined XR/GDH, 1 mM NADH, 1 mM NAD^+, 100 mM glycerol, 100 mM xylose in 100 mM K_2PO_4 pH 7.5 at 25°C. NADH absorbance at 340 nm was monitored over the course of the 60 min assays. Product formation was confirmed using a Shimadzu high performance liquid chromatography (HPLC) system with a RID-10A refractive index detector. An Aminex HPX-87H column was used for product separation with a mobile phase of 35% acetonitrile, 0.5 mM sulfuric acid and run conditions of 0.5 mL/min at 65°C.

Cellulase assays were also performed in bulk format using five crystals per well. Crystals were loaded as described above and then incubated with 0.5 mM p-nitrophenol β-d-cellobioside (pNPC) in 50 mM sodium acetate pH 5. For stability studies, enzyme-laden crystals were

incubated at 100°C for two minutes prior to the addition of substrate. Activity assays were run at 65°C for 30 min and then stopped by the addition of 2 M sodium acetate, 0.1 M glycine, pH 12.5 followed by subsequent 0.22 μm filtering. Formation of the soluble yellow product p-nitrophenol was measured at 415 nm wavelength absorbance.

5.4. RESULTS AND DISCUSSION

5.4.1. LOADING AND RELEASE OF FLUORESCENTLY TAGGED PROTEINS

Our previous work with fluorescent proteins[265] confirmed that macromolecular guests could diffuse into porous CJ crystals. However, macromolecular diffusion is likely to depend on size, charge, or other properties that can vary from protein to protein. To confirm that our proteins of interest could diffuse into crosslinked CJ crystals, each was tagged with a fluorescent probe and monitored via confocal microscopy. Bovine serum albumin (BSA), a moderate size protein, diffused throughout a ~80 μm thick CJ crystal on the timescale of ~20 minutes (Figure 5.5). Viewing the crystal from a side view (i.e. perpendicular to the $\hat{z}$ axis) distinctly showed how BSA diffused through the axial crystal pores.

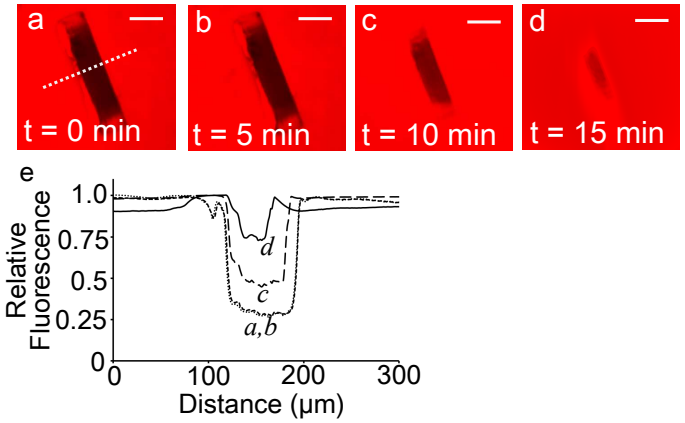

FIGURE 5.5: LOADING OF FLUORESCENTLY TAGGED BSA INTO A CJ CRYSTAL
(a-d) Side view (perpendicular to the $\hat{z}$ axis; major axial pores are in the plane of the page) of BSA diffusing into a ~80 μm thick crystal over a 15 min period. **(e)** Fluorescence intensity profiles along the dotted line in panel *a* normalized relative to the background solution intensity. Scale bars (upper right) are 100 μm.

Although not shown here, other protein guests displayed similar loading behavior. Fluorescent imaging of a ~40 μm thick CJ crystal after approximately one hour of soaking in hHRP indicated that enzymes had diffused throughout the entire crystal. Fluorescence from the tagged enzyme guest was at least as bright on the center plane as on the crystal surface (Figure 5.6).

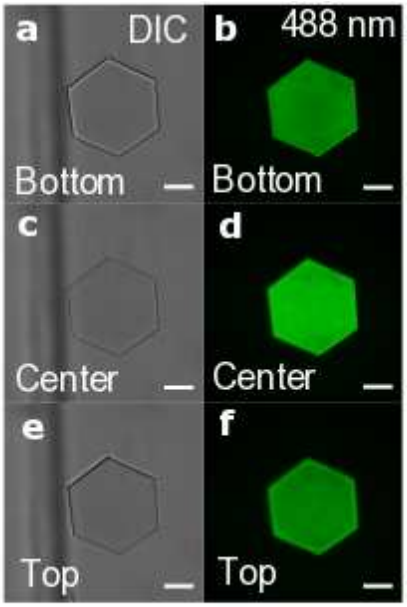

FIGURE 5.6: FLUORESCENTLY TAGGED HHRP LOADED THROUGHOUT A CJ CRYSTAL
Brightfield and fluorescence images were taken at **(a,b)** the bottom plane, **(c,d)** the middle plane, and **(e,f)** the top plane of a ~40 μm thick crystal. Scale bars (lower right) are 100 μm.

Enzyme retention would be expected to vary based on the type and relative magnitude of non-covalent interactions formed with the crystal scaffold. For fluorescent proteins, metal-mediated binding resulted in the longest retention times.[265] To take advantage of this with our enzymes, a variant of HRP with a modified tridentate chelator was used for activity assays. Although standard HRP and His-Probe™ HRP conjugate (hHRP) initially followed similar exponential decay unloading profiles, some hHRP appeared to stay bound within the crystal even after multiple days of washing in BufferA (Figure 5.7). The initial decrease in fluorescence over

the first ~30 minutes can likely be attributed to the diffusion of unbound (i.e. not strongly coordinated via a metal ion) hHRP molecules. After the initial fluorescence decay period, enzyme diffusion out of the crystal appeared to be minimal. Indeed, crystals loaded with hHRP maintained activity even after four days of incubation in BufferA.

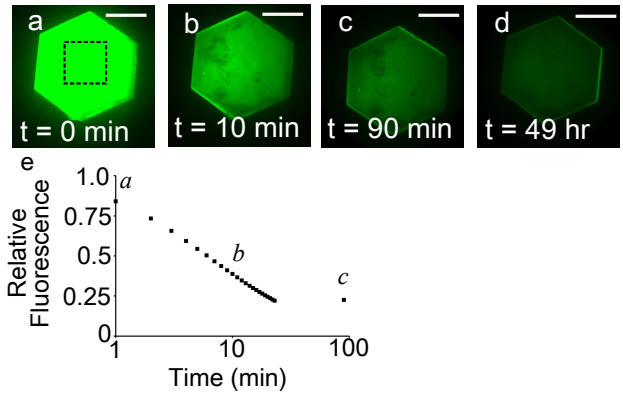

FIGURE 5.7: RETENTION OF FLUORESCENTLY TAGGED HHRP MOLECULES OVER TIME
Fluorescence at the center plane of the crystal indicates the presence of hHRP after **(a)** 0 min, **(b)** 10 min, or **(c)** 90 min of washing. **(d)** Some fluorescence was maintained for greater than two days. **(e)** The average fluorescence intensity (measured within the dashed box indicated in panel *a*) initially decreased exponentially but appeared to stabilize after ~30 min. Scale bars (upper right) are 100 μm.

Shared metal coordination by histidine tags (His-tags) can also result in high affinity.[276] In theory, any appropriately sized enzyme with an exposed His-tag could be captured within the crystal by interacting with the CJ His-tags that line the pore walls. To demonstrate this more general binding strategy, multiple His-tag bearing enzymes were soaked into the crystal scaffold. Measurable cellulase (CEL) activity on soluble substrate pNPC suggested successful metal-mediated retension of His-tag bearing enzymes within CJ crystals. Xylose reductase (XR) and glycerol dehydrogenase (GDH) enzymes could also be simultaneously loaded into CJ crystals as assessed by confocal microscopy (Figure 5.8). When each His-tag bearing enzyme was loaded individually, fluorescence was only observed at the appropriate excitation wavelength. However, when both enzymes were loaded at the same time, fluorescence was clearly evident at both 488 nm and 561 nm excitation wavelengths. Interestingly, GDH tended to form aggregates at high

zinc concentrations in the vicinity of the crystal surface. Although zinc helps coordinate glycerol in the binding pocket[271] and is likely necessary for enzyme activity, high concentrations appeared to decrease protein solubility. We attribute aggregation to the aforementioned propensity of His-tags to dimerize in the presence of zinc; each GDH tetramer can bind up to four neighbors via metal coordination.

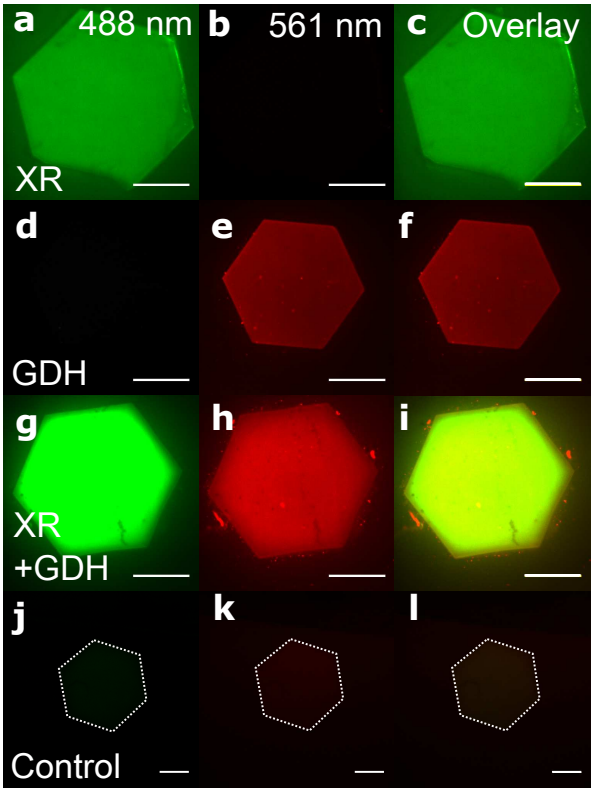

FIGURE 5.8: LOADING OF FLUORESCENTLY TAGGED XR AND GDH ENZYMES
(a-c) When XR was loaded independently, fluorescence was only evident with 488 nm excitation. **(d-f)** When GDH was loaded independently, fluorescence was only evident with higher wavelength (561 nm) excitation. **(g-i)** Both enzymes could be loaded simultaneously and imaged at their respective wavelengths. **(j-l)** An example control crystal (indicated by the dotted hexagon) shows autofluorescence of the CJ scaffold was not the primary source of fluorescence. Scale bars (lower right) are 100 μm for each image.

While metal-mediated coordination might be sufficient for some enzymes, the majority of useful enzymes do not possess a specific binding motif. Trapping macromolecular guests via physical encapsulation would provide the most generic strategy for enzyme immobilization. To this end, CJ crystals were briefly soaked in a BSA/glutaraldehyde solution to create a

semipermeable layer. Long soaking times created an extensive hydrogel matrix (Figure 5.3), while short soak times of ~1 sec appeared to block axial diffusion without visibly changing the crystalline scaffold. Preliminary diffusion studies suggested that more glucose oxidase (GOX) remained entrapped within a BSA-soaked crystal than a pre-BSA control crystal upon washing in BufferA (Figure 5.9). However, it should be noted that the fluorescence of GOX varied significantly between the BSA and non-BSA crystals (Figure 5.9a vs. 5.9c), preventing comparison of absolute fluorescence. Evidence of retained GOX activity after washing suggested the BSA encapsulation strategy was successful, but further trials will be required.

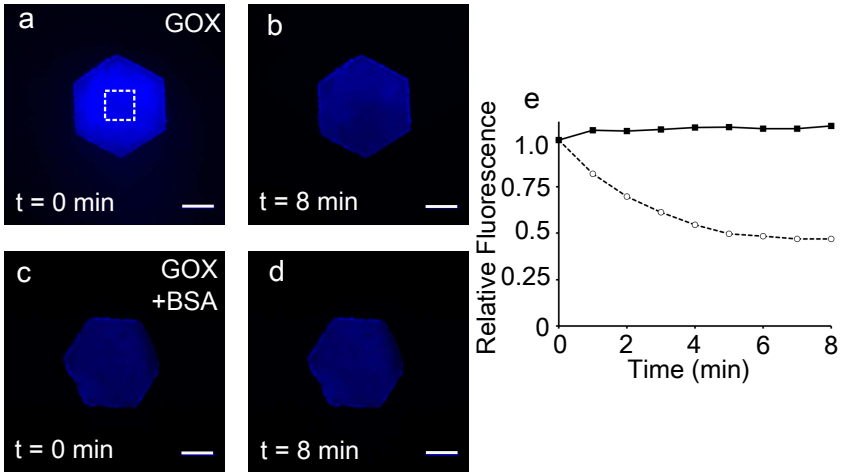

FIGURE 5.9: ENTRAPPING GOX BY COATING THE CRYSTAL SURFACE WITH BSA
Fluorescence of tagged GOX is shown **(a)** before and **(b)** after soaking in BufferA for 8 min without modifying the crystal and **(c,d)** with a BSA/glutaraldehyde encapsulation matrix. **(e)** The relative fluorescence (averaged across the dotted square in panel *a*) remained higher for the coated crystal than for the non-coated crystal. Scale bars (lower right) are 100 μm.

5.4.2. ENZYMATIC ACTIVITY ASSAYS

As an initial proof of concept, the model enzyme horseradish peroxidase (hHRP) was loaded into crosslinked CJ crystals. Resorufin product formation was clearly evident when solution conditions were changed from BufferA to assay buffer (50 μM AmplexRed, 100 μM H_2O_2). The internal region of the crystal turned red (resorufin excited at 561 nm) within a few

seconds. The fluorescent red product continued to be generated and diffused throughout the surrounding solution over a period of ~2 min (Figure 5.10). hHRP-laden crystals retained high activity after multiple two hour rinse cycles in BufferA (Figure 5.11). Even after four days incubation in excess BufferA, a significant fraction of hHRP remained bound and active within the crystal (data not shown).

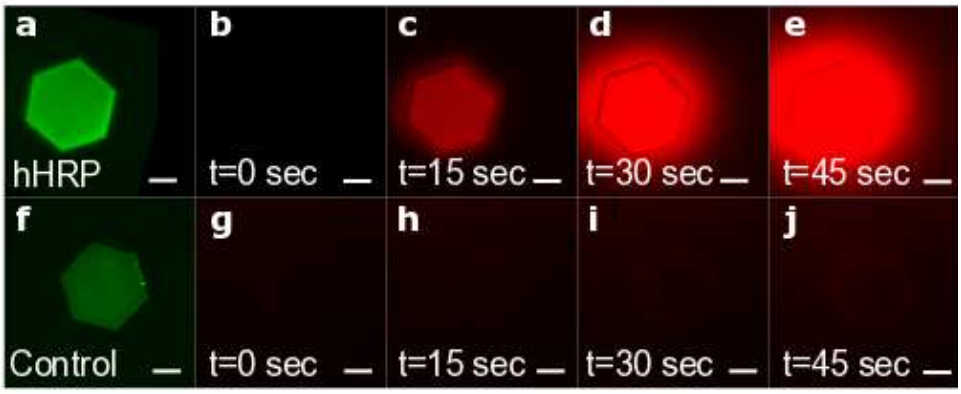

FIGURE 5.10: HHRP ACTIVITY IN A CJ CRYSTAL
(a) Fluorescence of the tagged hHRP enzyme (green) loaded in a crystal. **(b-e)** Formation of the fluorescent resorufin product over a 45 sec time period. **(f)** When imaged under the same settings, an empty control crystal had minimal green autofluorescence and **(g-j)** did not have any significant activity. Scale bars (lower right) are 100 µm.

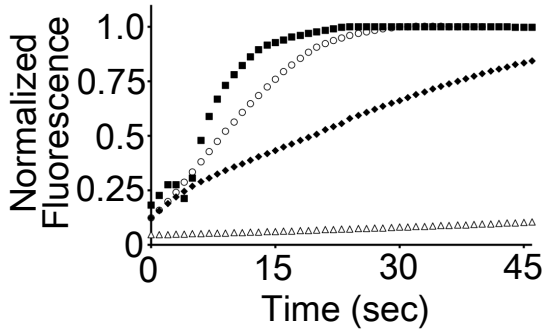

FIGURE 5.11: RETAINED HHRP ACTIVITY AFTER MULTIPLE WASH CYCLES
Normalized fluorescence profiles indicate the formation of the resorufin product inside the crystal during the first assay (black squares), after a single two hour wash cycle (open circles), and after two wash cycles (black diamonds). All had significantly higher activity than an empty control crystal (open triangles).

To demonstrate the potential utility of protein crystal microreactors as sensors, a combined two-enzyme pathway was loaded into a single crystal. Sequential soaking in hHRP followed by GOX resulted in a co-loaded crystal that had detectable activity in a 10 mM glucose

solution (Figure 5.12). Much like the single-enzyme hHRP crystal, resorufin product formation was visible within a few seconds and continued to increase within the crystal throughout the two minute assay. Product diffusion into the surrounding solution appeared to be slowed but not fully inhibited by the BSA/glutaraldehyde coating. Alternative semipermeable membranes or different coating techniques could be selected to optimize substrate and product diffusion for a given enzymatic pathway.

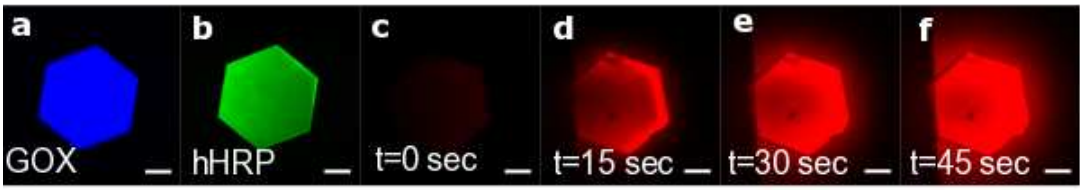

FIGURE 5.12: GOX/HHRP ENZYME PATHWAY
When a CJ crystal was loaded with **(a)** fluorescently tagged GOX (blue) and **(b)** fluorescently tagged hHRP (green), 10 mM glucose could be detected via **(c-f)** a two-step process that formed fluorescent product resorufin. Scale bars (lower right) are 100 μm.

5.4.3. CATALASE EXCLUSION

Distinct benefits can be realized by co-localizing multiple enzymes within a protein crystal. For example, by coating the surface of the crystal with a semipermeable BSA/glutaraldehyde layer, free enzymes in the surrounding solution can be excluded from the internal crystal microenvironment. This effect was demonstrated by soaking coated and non-coated protein crystals in a solution of fluorescently tagged catalase (CAT). A non-coated protein crystal had higher fluorescence inside the crystal than the surrounding solution, while a BSA/glutaraldehyde coated crystal had much lower fluorescence inside the crystal than the surrounding solution (Figure 5.13).

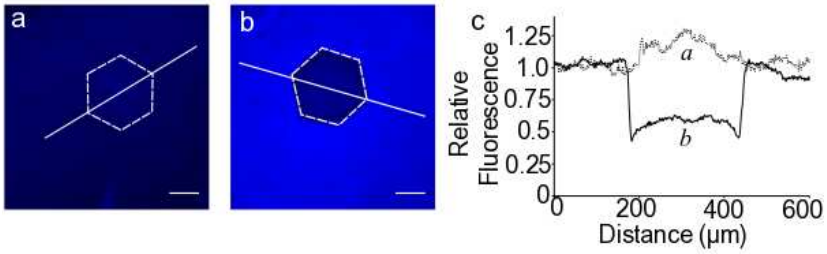

FIGURE 5.13: CAT LOADING OR EXCLUSION FROM CJ CRYSTALS
(a) Control CJ crystal soaked in fluorescently tagged CAT. **(b)** CJ crystal coated with BSA/glutaraldehyde matrix prior to soaking in tagged CAT. The crystal perimeter is denoted by a dashed hexagon; a white line indicates the profile analyzed in panel *c*. Scale bars (lower right) are 100 μm. **(c)** Profiles through each ~200 μm diameter crystal indicate the internal fluorescence relative to the surrounding solution.

Co-immobilizing GOX and hHRP within a coated protein crystal created a reaction pathway that appeared to be isolated from the surrounding CAT enzyme. Although CAT drastically decreased hHRP activity (<15% residual activity upon addition of 50 μg/mL CAT) in *in vitro* assays by scavenging the H_2O_2 intermediate, there was only a minimal change in activity for a GOX/hHRP-laden crystal in CAT-free or 50 μg/mL CAT solutions (Figure 5.14). This result suggested that the H_2O_2 intermediate was consumed by hHRP within the protein crystal before it was able to diffuse into the surrounding CAT solution. Slight variability in the initial fluorescence was caused by residual bound product from the previous assay. The poor solubility of resorufin may have slowed release from the protein crystal; extensive washing (two or more hours in BufferA) was required in order to decrease fluorescence to near background levels.

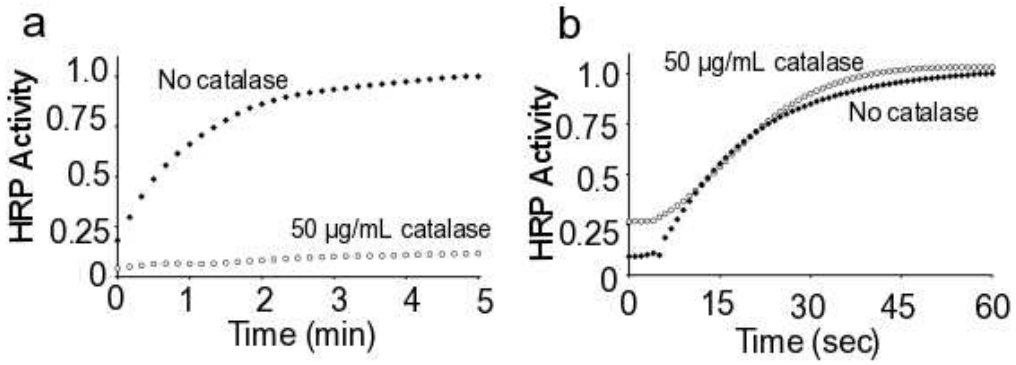

FIGURE 5.14: CAT EFFECT ON GOX/HHRP PATHWAY
(a) *In vitro* assays with and without added catalase showed how CAT scavenging of the H_2O_2 intermediate reduced hHRP activity. **(b)** When both hHRP and GOX were immobilized inside a CJ crystal, the semipermeable BSA/glutaraldehyde layer appeared to exclude CAT, and hHRP showed no decrease in activity.

5.5. CONCLUSION

Porous protein crystals provide a unique framework for selectively capturing and releasing guest molecules. By demonstrating the potential of CJ crystals in binding multiple macromolecular guests, this study lays the groundwork for future protein crystal immobilization studies. As methods for non-covalent or covalent binding are further developed, the list of potential molecular variations and applications will quickly grow. Studies involving coenzyme immobilization, complex multi-enzyme pathways, or controlled guest release provide just a few examples of how the proposed platform could be extended in scope and application.

6. CONCLUSIONS AND FUTURE WORK

The benefits of protein engineering have been realized in many biotechnology applications. However, computational protein design is still a relatively new technology and will need continued improvement in order to achieve higher success rates. Our work identifies potential flaws in current rational design methods and presents ways to reduce the number of design failures in the future. Detailed description of protein-solvent interactions in aqueous ionic liquids provides further insight for improving enzyme stability in non-traditional solvents. In addition, our proposed immobilization strategy involving porous protein crystals offers an opportunity to extend molecular scale studies in to the realm of mesoscale applications.

A few additional computational experiments could extend the usefulness of the proposed design methods. In Chapter 2, the conservative design strategy failed to improve the stability of a thermostable enzyme. However, it remains unclear whether these limitations were a result of the design methods themselves or of the high native stability of the starting template. Applying a similar approach to a mesostable enzyme could very well lead to significant improvements in stability.

Sampling of alternate states and solvent interactions were discussed as potential flaws in the original computational design. Incorporating backbone flexibility could improve the sampling of amino acid positions. We developed a computational strategy for varying backbone orientations at local sites; however, our approach has not yet been experimentally tested. Performing a side-by-side comparison of the three computational design strategies (rotameric, flexible backbone, and MD) for the same target enzyme could provide a useful comparison point for future studies.

Solvent interactions are more difficult to address in computational design algorithms. Current design methods rely upon implicit solvent approximations or crude solvation energy terms. An ideal approach would incorporate explicit solvent molecules, but their countless orientations and interactions prevent this approach from being feasible in combinatorial studies. Instead, it might prove beneficial to train current solvation terms based on short MD simulations of each point mutation. Sampling different states in the presence of explicit solvent molecules could improve the inherent shortcomings of rotamer libraries and provide an empirical bias for training energy functions.

Expanding the applicability of the presented computational methods from a single enzyme to a panel of candidate proteins would further justify claims regarding generalizability. Some modifications might be required to adapt the computational methods for alternate targets. For example, rather than relying on a predicted energy function, we replaced the scoring metric with apparent backbone 'flexibility' for MD-guided design of protein crystal interfaces. This approach has yet to be experimentally verified, but similar modified algorithms could extend the usefulness of MD to a range of new applications.

The results of supercharging experiments also suggested several future computational design modifications. Our initial strategy targeted a change in charge of ~20 e^-. Designing proteins with a higher or lower net charge would better define the limits of this approach. If introducing a large number of mutations appeared to have a negative impact, the design strategy could be modified to target only charge reversal sites (positive to negative or *vice versa* instead of neutral to charged mutations). Alternatively, as a control, our design strategies could be modified to reduce net protein charge. A recent example of a cellulose binding domain that lacks

charged residues[277] suggests some proteins could be robust enough for several types of charge modification strategies.

The MD simulations of E1 in ionic liquids (Chapter 4) generated several hypotheses for improving enzyme stability. The dynamic loop region of E1 that unfolded in the presence of ILs could be a good target for site saturation mutagenesis studies or other targeted stabilization approaches. Alternatively, mutations introduced at sites where cations strongly bound could reduce the inhibitory effect of ILs.

One of the least inactivating ILs, [Emim]Cl, was selected for our MD studies. However, enzyme behavior varied drastically across the panel of experimentally tested ILs, and simulations of different protein-IL systems could provide useful generic design rules. In particular, simulations that investigated the behavior of E1 in [Emim]OAc, a strongly inactivating IL, could be useful in guiding future IL design efforts.

To meet the long-term goal of creating an industrially relevant enzyme, improved E1 variants should be tested in a one-pot system that combines pretreatment and hydrolysis steps. All of our experimental activity studies involved soluble substrate analogues. Assaying on cellulosic biomass samples in the presence of ILs would provide a more useful standard for industrial applications.

Some of the results presented in Chapter 5 were based on preliminary experiments and will need additional trials prior to publication. For example, the release of GOX from coated crystals was only tested in a single case and still needs to be independently verified. Similarly, GOX/hHRP assays have only been performed for a single test case and should be assessed via additional trials, preferably with a varied glucose concentration to assess the minimum limit of detection. Comparisons between long-term HRP and hHRP binding could also provide support

for claims of metal-specific binding. Comparison between CJ crystal fluorescence prior to loading, at maximal HRP/hHRP loading, and after one week of washing in BufferA might help distinguish the benefit of metal-coordinated affinity. Testing the activity of the washed crystals could also help confirm the presence of residual bound enzyme.

Additional immobilization benefits could be realized by testing the crystal/enzyme complex under harsh conditions. For example, measuring activity pre- and post-lyophilization could provide an assessment of retained activity relative to free enzyme solutions. In theory, the CJ matrix could also impart a thermostability benefit for entrapped enzymes; however, CEL stability assays proposed in the "Methods" section of Chapter 5 have yet to be completed. If CEL stability assays are successful, CEL should be fluorescently labeled and imaged via confocal microscopy to confirm uniform crystal loading.

Quantifying the concentration of bound enzymes presents one of the largest remaining hurdles for protein crystal immobilization studies. Estimating the extent of enzyme loading would allow activity measurements to be presented on a per molecule basis. Without this information, it is difficult to compare the relative activity of immobilized enzymes to those in free solution. However, variability between crystals and across fluorescence tagging experiments complicated efforts to correlate fluorescence and enzyme concentration via a standard curve. In addition, fluorescence intensity depended on path length (crystal height or depth of solution) in confocal microscopy experiments, presumably due to significant excitation of fluorophores outside the confocal plane. Advanced microscopy techniques, such as two-photon excitation microscopy, could reduce the effects background fluorescence interference.

HPLC conditions reported in the "Methods" section of Chapter 5 were suitable for creating standard curves and detecting XR activity; however, the activity of GDH-soaked

crystals was not above background. In fact, the activity of GDH enzymes was relatively low even for *in vitro* assays. This could be attributed to NAD^+ coenzyme degradation or GDH aggregation. Poor recovery of GDH during purification and low specific activity levels further complicated bulk crystal activity assays. Increasing the number of crystals, the loaded enzyme concentration, or the specific enzyme activity might be necessary to achieve measurable GDH product formation. Alternatively, GDH from a different organism could be purchased from a commercial supplier and immobilized using the BSA/glutaraldehyde entrapment strategy.

Although CJ crystals have been extensively analyzed via x-ray diffraction in previous studies,[264,265] enzyme-laden crystals have not yet been analyzed using this technique. X-ray diffraction experiments could buttress multiple statements made in Chapter 5, including the claim that bound enzymes are free to adopt multiple orientations and that treatment with BSA/glutaraldehyde did not significantly alter the crystalline matrix.

As the challenges regarding enzyme immobilization are overcome, the realm of interesting experiments for porous protein crystals dramatically increases. Enzymes could be covalently attached or positioned at specific sites within the protein crystal to promote substrate channeling or coenzyme recycling. Crystals with varying morphology could be loaded with different enzymes to create sensors capable of detecting multiple analytes.[278] Enzymes could be selectively released for drug delivery applications. Many enzymes could be immobilized simultaneously to create complex pathways. The list of exciting ideas goes on and on, providing inspiration for the next generation of protein engineering graduate students.

7. BIBLIOGRAPHY

1. Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13,** 345–351 (2002).
2. Markets and Markets. Specialty Enzymes Market worth $4317.2 Million by 2018. Available at: http://www.marketsandmarkets.com/PressReleases/specialty-enzymes.asp. (Accessed: 7th May 2016)
3. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
4. Darnhus, T., Kaasgaard, S. & Olsen, H. S. Enzymes at Work. in **4th edition,** (2013).
5. Polizzi, K. M., Bommarius, A. S., Broering, J. M. & Chaparro-Riggers, J. F. Stability of biocatalysts. *Curr. Opin. Chem. Biol.* **11,** 220–225 (2007).
6. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J.* **10,** 75–83 (1996).
7. Liszka, M. J., Clark, M. E., Schneider, E. & Clark, D. S. Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **3,** 77–102 (2012).
8. DeSantis, G. & Jones, J. B. Chemical modification of enzymes for enhanced functionality. *Curr. Opin. Biotechnol.* **10,** 324–330 (1999).
9. Hernaiz, M. J., Sánchez-Montero, J. M. & Sinisterra, J. V. Influence of the nature of modifier in the enzymatic activity of chemical modified semipurified lipase from Candida rugosa. *Biotechnol. Bioeng.* **55,** 252–260 (1997).
10. Garcia, D., Ortega, F. & Marty, J.-L. Kinetics of thermal inactivation of horseradish peroxidase: stabilizing effect of methoxypoly (ethylene glycol). *Biotechnol. Appl. Biochem.* **27,** 49–54 (1998).
11. Jene, Q., Pearson, J. C. & Lowe, C. R. Surfactant modified enzymes: solubility and activity of surfactant-modified catalase in organic solvents. *Enzyme Microb. Technol.* **20,** 69–74 (1997).
12. Nordwald, E. M. & Kaar, J. L. Stabilization of enzymes in ionic liquids via modification of enzyme charge. *Biotechnol. Bioeng.* **110,** 2352–2360 (2013).
13. Nordwald, E. M., Brunecky, R., Himmel, M. E., Beckham, G. T. & Kaar, J. L. Charge engineering of cellulases improves ionic liquid tolerance and reduces lignin inhibition. *Biotechnol. Bioeng.* **111,** 1541–1549 (2014).
14. Nordwald, E. M. & Kaar, J. L. Mediating electrostatic binding of 1-butyl-3-methylimidazolium chloride to enzyme surfaces improves conformational stability. *J. Phys. Chem. B* **117,** 8977–8986 (2013).
15. Sheldon, R. A. Enzyme immobilization: The quest for optimum performance. *Adv. Synth. Catal.* **349,** 1289–1307 (2007).
16. Katchalski-Katzir, E. & Kraemer, D. M. Eupergit® C, a carrier for immobilization of enzymes of industrial potential. *J. Mol. Catal. B Enzym.* **10,** 157–176 (2000).
17. Tan, T., Lu, J., Nie, K., Deng, L. & Wang, F. Biodiesel production with immobilized lipase: a review. *Biotechnol. Adv.* **28,** 628–634 (2010).

18. Ngu, T. A., Li, Z. & others. Phosphotungstic acid-functionalized magnetic nanoparticles as an efficient and recyclable catalyst for the one-pot production of biodiesel from grease via esterification and transesterification. *Green Chem.* **16,** 1202–1210 (2014).
19. Schoevaart, R. *et al.* Preparation, optimization, and structures of cross-linked enzyme aggregates (CLEAs). *Biotechnol. Bioeng.* **87,** 754–762 (2004).
20. Häring, D. & Schreier, P. Cross-linked enzyme crystals. *Curr. Opin. Chem. Biol.* **3,** 35–38 (1999).
21. Nancy L. St. Clair, M. A. N. Cross-linked enzyme crystals as robust biocatalysts. J Am Chem Soc. *J. Am. Chem. Soc. - J AM CHEM SOC* **114,** 7314–7316 (1992).
22. Fu, J., Liu, M., Liu, Y. & Yan, H. Spatially-interactive biomolecular networks organized by nucleic acid nanostructures. *Acc. Chem. Res.* **45,** 1215–1226 (2012).
23. de la Escosura, A., Nolte, R. J. & Cornelissen, J. J. Viruses and protein cages as nanocontainers and nanoreactors. *J. Mater. Chem.* **19,** 2274–2278 (2009).
24. de Miguel Bouzas, T., Barros-Velázquez, J. & Gonzalez Villa, T. Industrial applications of hyperthermophilic enzymes: a review. *Protein Pept. Lett.* **13,** 645–651 (2006).
25. Haki, G. D. & Rakshit, S. K. Developments in industrially important thermostable enzymes: a review. *Bioresour. Technol.* **89,** 17–34 (2003).
26. Graham, J. E. *et al.* Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat. Commun.* **2,** 375 (2011).
27. Arfi, Y. *et al.* Characterization of salt-adapted secreted lignocellulolytic enzymes from the mangrove fungus Pestalotiopsis sp. *Nat. Commun.* **4,** 1810 (2013).
28. Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M. & Sani, R. K. Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour. Technol.* **128,** 751–759 (2013).
29. Bok, J.-D., Yernool, D. A. & Eveleigh, D. E. Purification, Characterization, and Molecular Analysis of Thermostable Cellulases CelA and CelB fromThermotoga neapolitana. *Appl. Environ. Microbiol.* **64,** 4774–4781 (1998).
30. Kumar, L., Awasthi, G. & Singh, B. Extremophiles: a novel source of industrially important enzymes. *Biotechnology* **10,** 121–135 (2011).
31. Brumm, P. J. Bacterial genomes: what they teach us about cellulose degradation. *Biofuels* **4,** 669–681 (2013).
32. D'haeseleer, P. *et al.* Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. *PloS One* **8,** e68465 (2013).
33. Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R. & Shoichet, B. K. Comparing the Thermodynamic Stabilities of a Related Thermophilic and Mesophilic Enzyme†. *Biochemistry (Mosc.)* **38,** 2570–2576 (1999).
34. Mingardon, F., Bagert, J. D., Maisonnier, C., Trudeau, D. L. & Arnold, F. H. Comparison of Family 9 Cellulases from Mesophilic and Thermophilic Bacteria. *Appl. Environ. Microbiol.* **77,** 1436 –1442 (2011).
35. Kumar, S., Tsai, C.-J. & Nussinov, R. Thermodynamic differences among homologous thermophilic and mesophilic proteins. *Biochemistry (Mosc.)* **40,** 14152–14165 (2001).
36. Szilágyi, A. & Závodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8,** 493–504 (2000).

37. Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev. MMBR* **65,** 1–43 (2001).
38. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci.* **90,** 5618–5622 (1993).
39. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. (1994).
40. Eijsink, V. G., G\a aseidnes, S., Borchert, T. V. & van den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **22,** 21–30 (2005).
41. Bommarius, A. S., Broering, J. M., Chaparro-Riggers, J. F. & Polizzi, K. M. High-throughput screening for enhanced protein stability. *Curr. Opin. Biotechnol.* **17,** 606–610 (2006).
42. Sieber, V., Plückthun, A. & Schmid, F. X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16,** 955–960 (1998).
43. Palackal, N. *et al.* An evolutionary route to xylanase process fitness. *Protein Sci.* **13,** 494–503 (2004).
44. Hecky, J. & Müller, K. M. Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of β-lactamase. *Biochemistry (Mosc.)* **44,** 12640–12654 (2005).
45. Ness, J. E. *et al.* DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* **17,** 893–896 (1999).
46. Carter, J. L., Bekhouche, M., Noiriel, A., Blum, L. J. & Doumèche, B. Directed evolution of a formate dehydrogenase for increased tolerance to ionic liquids reveals a new site for increasing the stability. *ChemBioChem* **15,** 2710–2718 (2014).
47. Eijsink, V. G. H. *et al.* Rational engineering of enzyme stability. *J. Biotechnol.* **113,** 105–120 (2004).
48. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383,** 66–93 (2004).
49. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G. & Eijsink, V. G. Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci.* **95,** 2056–2060 (1998).
50. Declerck, N. *et al.* Hyperthermostabilization of Bacillus licheniformis α-amylase and modulation of its stability over a 50 C temperature range. *Protein Eng.* **16,** 287–293 (2003).
51. Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta BBA-Protein Struct. Mol. Enzymol.* **1543,** 408–415 (2000).
52. Wijma, H. J., Floor, R. J. & Janssen, D. B. Structure-and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* **23,** 588–594 (2013).
53. Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **278,** 82–87 (1997).
54. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Mol. Biol.* **5,** 470–475 (1998).
55. Ponder, J. W. & Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193,** 775–791 (1987).

56. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332,** 449–460 (2003).
57. Ventura, S. *et al.* Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat. Struct. Mol. Biol.* **9,** 485–493 (2002).
58. Filikov, A. V. *et al.* Computational stabilization of human growth hormone. *Protein Sci.* **11,** 1452–1461 (2002).
59. Luo, P. *et al.* Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci.* **11,** 1218–1226 (2002).
60. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **308,** 857–860 (2005).
61. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).
62. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329,** 309–313 (2010).
63. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364 (2003).
64. Sheffler, W. & Baker, D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* **18,** 229–239 (2009).
65. Mooers, B. H. *et al.* Repacking the core of T4 lysozyme by automated design. *J. Mol. Biol.* **332,** 741–756 (2003).
66. Lazar, G. A., Desjarlais, J. R. & Handel, T. M. De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* **6,** 1167–1178 (1997).
67. Mandell, D. J. & Kortemme, T. Backbone flexibility in computational protein design. *Curr. Opin. Biotechnol.* **20,** 420–428 (2009).
68. Friedland, G. D. & Kortemme, T. Designing ensembles in conformational and sequence space to characterize and engineer proteins. *Curr. Opin. Struct. Biol.* **20,** 377–384 (2010).
69. Yin, S., Ding, F. & Dokholyan, N. V. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* **15,** 1567–1576 (2007).
70. Murphy, G. S. *et al.* Increasing Sequence Diversity with Flexible Backbone Protein Design: The Complete Redesign of a Protein Hydrophobic Core. *Structure* **20,** 1086–1096 (2012).
71. Gao, D. *et al.* Thermostable Variants of Cocaine Esterase for Long-Time Protection against Cocaine Toxicity. *Mol. Pharmacol.* **75,** 318–323 (2009).
72. Pikkemaat, M. G., Linssen, A. B., Berendsen, H. J. & Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng.* **15,** 185–192 (2002).
73. Wijma, H. J. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27,** 49–58 (2014).
74. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* **351,** 1160–1168 (2005).
75. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16,** 378–384 (2005).

76. Karanicolas, J. *et al.* A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Mol. Cell* **42,** 250–260 (2011).
77. Khersonsky, O. *et al.* Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 10358–10363 (2012).
78. Ansorge-Schumacher, M. B., Slusarczyk, H., Schümers, J. & Hirtz, D. Directed evolution of formate dehydrogenase from Candida boidinii for improved stability during entrapment in polyacrylamide. *Febs J.* **273,** 3938–3945 (2006).
79. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103,** 5869–5874 (2006).
80. Chaparro-Riggers, J. F., Polizzi, K. M. & Bommarius, A. S. Better library design: data-driven protein engineering. *Biotechnol. J.* **2,** 180–191 (2007).
81. Johnson, L., Huber, T. & Snow, C. in *Protein Design* (ed. Kohler, V.)
82. Mena, M. A. & Daugherty, P. S. Automated design of degenerate codon libraries. *Protein Eng. Des. Sel. PEDS* **18,** 559–561 (2005).
83. Patrick, W. M. & Firth, A. E. Strategies and computational tools for improving randomized protein libraries. *Biomol. Eng.* **22,** 105–112 (2005).
84. Meyer, M. M., Hochrein, L. & Arnold, F. H. Structure-guided SCHEMA recombination of distantly related β-lactamases. *Protein Eng. Des. Sel.* **19,** 563–570 (2006).
85. Landwehr, M., Carbone, M., Otey, C. R., Li, Y. & Arnold, F. H. Diversification of catalytic function in a synthetic family of chimeric cytochrome p450s. *Chem. Biol.* **14,** 269–278 (2007).
86. Li, Y. *et al.* A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* **25,** 1051–1056 (2007).
87. Romero, P. A. *et al.* SCHEMA Designed Variants of Human Arginase I & II Reveal Sequence Elements Important to Stability and Catalysis. *ACS Synth. Biol.* **1,** 221–228 (2012).
88. Smith, M. A., Romero, P. A., Wu, T., Brustad, E. M. & Arnold, F. H. Chimeragenesis of distantly-related proteins by noncontiguous recombination. *Protein Sci.* **22,** 231–238 (2013).
89. Chang, C.-J. *et al.* Exploring the Mechanism Responsible for Cellulase Thermostability by Structure-Guided Recombination. *PLOS ONE* **11,** e0147485 (2016).
90. Heinzelman, P. *et al.* Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng. Des. Sel. PEDS* **23,** 871–880 (2010).
91. Heinzelman, P. *et al.* SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J. Biol. Chem.* **284,** 26229–26233 (2009).
92. Heinzelman, P. *et al.* A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5610–5615 (2009).
93. Smith, M. A., Bedbrook, C. N., Wu, T. & Arnold, F. H. Hypocrea jecorina cellobiohydrolase I stabilizing mutations identified using noncontiguous recombination. *ACS Synth. Biol.* **2,** 690–696 (2013).
94. Smith, M. A. *et al.* A diverse set of family 48 bacterial glycoside hydrolase cellulases created by structure-guided recombination. *FEBS J.* **279,** 4453–4465 (2012).

95. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859 (1995).
96. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett.* **362,** 281–285 (1995).
97. Tucker, M. P., Mohagheghi, A., Grohmann, K. & Himmel, M. E. Ultra-Thermostable Cellulases From Acidothermus cellulolyticus: Comparison of Temperature Optima with Previously Reported Cellulases. *Nat. Biotechnol.* **7,** 817–820 (1989).
98. Thomas, S. R., Laymon, R. A. & Himmel, M. E. Gene coding for the E1 endoglucanase. (1996).
99. Barabote, R. D. *et al.* Complete genome of the cellulolytic thermophile Acidothermus cellulolyticus 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome Res.* **19,** 1033–1043 (2009).
100. Linger, J. G., Adney, W. S. & Darzins, A. Heterologous Expression and Extracellular Secretion of Cellulolytic Enzymes by Zymomonas mobilis. *Appl. Environ. Microbiol.* **76,** 6360–6369 (2010).
101. Baker, J. O. *et al.* Catalytically Enhanced Endocellulase CeI5A from Acidothermus cellulolyticus. in *Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals* 129–148 (2005).
102. Lindenmuth, B. E. & McDonald, K. A. Production and characterization of Acidothermus cellulolyticus endoglucanase in Pichia pastoris. *Protein Expr. Purif.* **77,** 153–158 (2011).
103. Yennamalli, R. M., Rader, A. J., Kenny, A. J., Wolt, J. D. & Sen, T. Z. Endoglucanases: insights into thermostability for biofuel applications. *Biotechnol. Biofuels* **6,** 1 (2013).
104. Johnson, L. B., Gintner, L. P., Park, S. & Snow, C. D. Discriminating between stabilizing and destabilizing protein design mutations via recombination and simulation. *Protein Eng. Des. Sel.* **28,** 259–267 (2015).
105. Badieyan, S., Bevan, D. R. & Zhang, C. Study and design of stability in GH5 cellulases. *Biotechnol. Bioeng.* **109,** 31–44 (2012).
106. Jaeger, V., Burney, P. & Pfaendtner, J. Comparison of three ionic liquid-tolerant cellulases by molecular dynamics. *Biophys. J.* **108,** 880–892 (2015).
107. Skopec, C. E., Himmel, M. E., Matthews, J. F. & Brady, J. W. Energetics for displacing a single chain from the surface of microcrystalline cellulose into the active site of Acidothermus cellulolyticus Cel5A. *Protein Eng.* **16,** 1005–1015 (2003).
108. Liu, J., Wang, X. & Xu, D. QM/MM study on the catalytic mechanism of cellulose hydrolysis catalyzed by cellulase Cel5A from Acidothermus cellulolyticus. *J. Phys. Chem. B* **114,** 1462–1470 (2009).
109. Baker, J. O. *et al.* A new thermostable endoglucanase, Acidothermus cellulolyticus E1. *Appl. Biochem. Biotechnol.* **45,** 245–256 (1994).
110. Baker, J. O. *et al.* Synergism between purified bacterial and fungal cellulases. in *ACS Symposium Series* **618,** 113–141 (1995).
111. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R. & Karplus, P. A. Crystal structure of thermostable family 5 endocellulase E1 from Acidothermus cellulolyticus in complex with cellotetraose. *Biochemistry (Mosc.)* **35,** 10648–10660 (1996).

112. McCarter, S. L. *et al.* Exploration of cellulose surface-binding properties of Acidothermus cellulolyticus Cel5A by site-specific mutagenesis. *Appl. Biochem. Biotechnol.* **98,** 273–287 (2002).
113. Rignall, T. R. *et al.* Effect of single active-site cleft mutation on product specificity in a thermostable bacterial cellulase. *Appl. Biochem. Biotechnol.* **98,** 383–394 (2002).
114. Kim, H.-W. & Ishikawa, K. Structure of hyperthermophilic endocellulase from Pyrococcus horikoshii. *Proteins Struct. Funct. Bioinforma.* **78,** 496–500 (2010).
115. Kim, H.-W., Kataoka, M. & Ishikawa, K. Atomic resolution of the crystal structure of the hyperthermophilic family 12 endocellulase and stabilizing role of the DxDxDG calcium-binding motif in Pyrococcus furiosus. *FEBS Lett.* **586,** 1009–1013 (2012).
116. Han-Woo, K. & Kazuhiko, I. Functional analysis of hyperthermophilic endocellulase from Pyrococcus horikoshii by crystallographic snapshots. *Biochem. J.* **437,** 223–230 (2011).
117. Kim, H.-W., Mino, K. & Ishikawa, K. Crystallization and preliminary X-ray analysis of endoglucanase from Pyrococcus horikoshii. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **64,** 1169–1171 (2008).
118. Kim, H.-W. & Ishikawa, K. The role of disulfide bond in hyperthermophilic endocellulase. *Extremophiles* **17,** 593–599 (2013).
119. Kang, H.-J., Uegaki, K., Fukada, H. & Ishikawa, K. Improvement of the enzymatic activity of the hyperthermophilic cellulase from Pyrococcus horikoshii. *Extremophiles* **11,** 251–256 (2007).
120. Kashima, Y., Mori, K., Fukada, H. & Ishikawa, K. Analysis of the function of a hyperthermophilic endoglucanase from Pyrococcus horikoshii that hydrolyzes crystalline cellulose. *Extremophiles* **9,** 37–43 (2005).
121. Davies, G. J. *et al.* Structure of the Bacillus agaradherans family 5 endoglucanase at 1.6 A and its cellobiose complex at 2.0 A resolution. *Biochemistry (Mosc.)* **37,** 1926–1932 (1998).
122. Davies, G. J. *et al.* Snapshots along an enzymatic reaction coordinate: analysis of a retaining β-glycoside hydrolase. *Biochemistry (Mosc.)* **37,** 11707–11713 (1998).
123. Hatada, Y. *et al.* Sequence of the gene for a high-alkaline mannanase from an alkaliphilic Bacillus sp. strain JAMB-750, its expression in Bacillus subtilis and characterization of the recombinant enzyme. *Extremophiles* **9,** 497–500 (2005).
124. Li, Y.-H., Ding, M., Wang, J., Xu, G. & Zhao, F. A novel thermoacidophilic endoglucanase, Ba-EGA, from a new cellulose-degrading bacterium, Bacillus sp. AC-1. *Appl. Microbiol. Biotechnol.* **70,** 430–436 (2006).
125. Shaw, A. *et al.* A novel combination of two classic catalytic schemes. *J. Mol. Biol.* **320,** 303–309 (2002).
126. Leggio, L. L. & Larsen, S. The 1. 62 A structure of Thermoascus aurantiacus endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett.* **523,** 103–108 (2002).
127. Parry, N. J. *et al.* Biochemical characterization and mode of action of a thermostable endoglucanase purified from Thermoascus aurantiacus. *Arch. Biochem. Biophys.* **404,** 243–253 (2002).
128. Van Petegem, F., Vandenberghe, I., Bhat, M. K. & Van Beeumen, J. Atomic resolution structure of the major endoglucanase from Thermoascus aurantiacus. *Biochem. Biophys. Res. Commun.* **296,** 161–166 (2002).

129. Chapon, V. *et al.* Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase cel5 (formerly EGZ) from Erwinia chrysanthemi. *J. Mol. Biol.* **310,** 1055–1066 (2001).
130. Bartoli-German, I., Haiech, J., Chippaux, M. & Barras, F. Informational Suppression to Investigate Structural Functional and Evolutionary Aspects of the Erwinia chrysanthemi Cellulase EGZ. *J. Mol. Biol.* **246,** 82–94 (1995).
131. Couturier, M. *et al.* Structural and biochemical analyses of glycoside hydrolase families 5 and 26 β-(1, 4)-mannanases from Podospora anserina reveal differences upon manno-oligosaccharide catalysis. *J. Biol. Chem.* **288,** 14624–14635 (2013).
132. Lee, T. M., Farrow, M. F., Arnold, F. H. & Mayo, S. L. A structural study of Hypocrea jecorina Cel5A. *Protein Sci.* **20,** 1935–1940 (2011).
133. Sabini, E. *et al.* The three-dimensional structure of a Trichoderma reesei-mannanase from glycoside hydrolase family 5. *Acta Crystallogr. D Biol. Crystallogr.* **56,** 3–13 (2000).
134. Nakazawa, H. *et al.* Characterization of the catalytic domains of Trichoderma reesei endoglucanase I, II, and III, expressed in Escherichia coli. *Appl. Microbiol. Biotechnol.* **81,** 681–689 (2008).
135. Schwarz, W. H., Gräbnitz, F. & Staudenbauer, W. L. Properties of a Clostridium thermocellum endoglucanase produced in Escherichia coli. *Appl. Environ. Microbiol.* **51,** 1293–1299 (1986).
136. Wu, T.-H. *et al.* Diverse substrate recognition mechanism revealed by Thermotoga maritima Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. *Biochim. Biophys. Acta BBA-Proteins Proteomics* **1814,** 1832–1840 (2011).
137. Arumugam Mahadevan, S., Gon Wi, S., Lee, D.-S. & Bae, H.-J. Site-directed mutagenesis and CBM engineering of Cel5A (Thermotoga maritima). *FEMS Microbiol. Lett.* **287,** 205–211 (2008).
138. Pereira, J. H. *et al.* Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic Thermotoga maritima. *J. Struct. Biol.* **172,** 372–379 (2010).
139. Chhabra, S. R., Shockley, K. R., Ward, D. E. & Kelly, R. M. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium Thermotoga maritima grown on glucan-and mannan-based polysaccharides. *Appl. Environ. Microbiol.* **68,** 545–554 (2002).
140. Chhabra, S. R. *et al.* Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium Thermotoga maritima. *J. Biol. Chem.* (2002).
141. Zheng, B. *et al.* Crystal Structure of Hyperthermophilic Endo-β-1, 4-glucanase. *J. Biol. Chem.* **287,** 8336–8346 (2012).
142. Wang, Y. *et al.* A novel thermostable cellulase from Fervidobacterium nodosum. *J. Mol. Catal. B Enzym.* **66,** 294–301 (2010).
143. Yeoman, C. J. *et al.* in *Advances in Applied Microbiology* (ed. Allen I. Laskin; Sima Sariaslani; Geoffrey M. Gadd) **Volume 70,** 1–55 (Academic Press, 2010).
144. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
145. Allen, B. D., Nisthal, A. & Mayo, S. L. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc. Natl. Acad. Sci.* **107,** 19838–19843 (2010).

146. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15,** 2785–2794 (2006).
147. Heinzelman, P. *et al.* A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci.* **106,** 5610 (2009).
148. Meyer, M. M. *et al.* Library analysis of SCHEMA-guided protein recombination. *Protein Sci.* **12,** 1686–1693 (2003).
149. Otey, C. R. *et al.* Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol.* **4,** e112 (2006).
150. Heinzelman, P. *et al.* SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J. Biol. Chem.* **284,** 26229–26233 (2009).
151. Himmel, M. E., Adney, W. S., Tucker, M. P. & Grohmann, K. Thermostable purified endoglucanas from acidothermus cellulolyticus ATCC 43068. (1994).
152. Margaritis, A. & Merchant, R. F. J. Thermostable cellulases from thermophilic microorganisms. *Crit. Rev. Biotechnol.* **4,** 327–399 (1986).
153. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
154. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J. Mol. Biol.* **320,** 369–387 (2002).
155. Loksha, I. V., Maiolo, J. R., 3rd, Hong, C. W., Ng, A. & Snow, C. D. SHARPEN-systematic hierarchical algorithms for rotamers and proteins on an extended network. *J. Comput. Chem.* **30,** 999–1005 (2009).
156. Silberg, J. J., Endelman, J. B. & Arnold, F. H. in *Methods in Enzymology* (ed. Dan E. Robertson and Joseph P. Noel) **388,** 35–42 (Academic Press, 2004).
157. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
158. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* **4,** e5553 (2009).
159. Clarke, J. & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry (Mosc.)* **32,** 4322–4329 (1993).
160. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).
161. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105,** 6474–6487 (2001).
162. Day, R., Bennion, B. J., Ham, S. & Daggett, V. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **322,** 189–203 (2002).
163. Jaenicke, R. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci.* **97,** 2962–2964 (2000).
164. Datta, S. *et al.* Ionic liquid tolerant hyperthermophilic cellulases for biomass pretreatment and hydrolysis. *Green Chem.* **12,** 338–345 (2010).

165. Kiss, G., Röthlisberger, D., Baker, D. & Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **19,** 1760–1773 (2010).
166. Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci.* **109,** 3790–3795 (2012).
167. Sanner, M. F., Olson, A. J. & Spehner, J.-C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38,** 305–320 (1996).
168. Baker, E. N. & Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44,** 97–179 (1984).
169. Komor, R. S., Romero, P. A., Xie, C. B. & Arnold, F. H. Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng. Des. Sel. PEDS* **25,** 827–833 (2012).
170. Otey, C. R. *et al.* Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol.* **4,** e112 (2006).
171. Elcock, A. H. The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284,** 489–502 (1998).
172. van Rantwijk, F. & Sheldon, R. A. Biocatalysis in ionic liquids. *Chem. Rev.* **107,** 2757–2785 (2007).
173. Kragl, U., Eckstein, M. & Kaftzik, N. Enzyme catalysis in ionic liquids. *Curr. Opin. Biotechnol.* **13,** 565–571 (2002).
174. Swatloski, R. P., Spear, S. K., Holbrey, J. D. & Rogers, R. D. Dissolution of cellose with ionic liquids. *J. Am. Chem. Soc.* **124,** 4974–4975 (2002).
175. Wahlström, R. M. & Suurnäkki, A. Enzymatic hydrolysis of lignocellulosic polysaccharides in the presence of ionic liquids. *Green Chem.* **17,** 694–714 (2015).
176. Zhao, H., Jones, C. L. & Cowins, J. V. Lipase dissolution and stabilization in ether-functionalized ionic liquids. *Green Chem.* **11,** 1128–1138 (2009).
177. Turner, M. B., Spear, S. K., Huddleston, J. G., Holbrey, J. D. & Rogers, R. D. Ionic liquid salt-induced inactivation and unfolding of cellulase from Trichoderma reesei. *Green Chem.* **5,** 443–447 (2003).
178. Yang, Z. Hofmeister effects: an explanation for the impact of ionic liquids on biocatalysis. *J. Biotechnol.* **144,** 12–22 (2009).
179. Zhao, H. Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids. *J. Mol. Catal. B Enzym.* **37,** 16–25 (2005).
180. Gao, W.-W., Zhang, F.-X., Zhang, G.-X. & Zhou, C.-H. Key factors affecting the activity and stability of enzymes in ionic liquids and novel applications in biocatalysis. *Biochem. Eng. J.* **99,** 67–84 (2015).
181. Moniruzzaman, M., Nakashima, K., Kamiya, N. & Goto, M. Recent advances of enzymatic reactions in ionic liquids. *Biochem. Eng. J.* **48,** 295–314 (2010).
182. Yu, X., Sun, Y., Xue, L., Huang, X. & Qu, Y. Strategies for improving the catalytic performance of an enzyme in ionic liquids. *Top. Catal.* **57,** 923–934 (2014).
183. Hofmeister, F. On the understanding of the effects of salts. *Arch Exp Pathol PharmakolLeipzig* **24,** 247–260 (1888).
184. Kumar, A. & Venkatesu, P. Does the stability of proteins in ionic liquids obey the Hofmeister series? *Int. J. Biol. Macromol.* **63,** 244–253 (2014).

185. Constantinescu, D., Weingärtner, H. & Herrmann, C. Protein denaturation by ionic liquids and the Hofmeister series: a case study of aqueous solutions of ribonuclease A. *Angew. Chem. Int. Ed.* **46,** 8887–8889 (2007).
186. Eggers, D. K. & Valentine, J. S. Crowding and hydration effects on protein conformation: a study with sol-gel encapsulated proteins1. *J. Mol. Biol.* **314,** 911–922 (2001).
187. Zhao, H. *et al.* Effect of kosmotropicity of ionic liquids on the enzyme stability in aqueous solutions. *Bioorganic Chem.* **34,** 15–25 (2006).
188. de Diego, T., Manjón, A. & Iborra, J. L. in *Production of Biofuels and Chemicals with Ionic Liquids* 275–301 (Springer, 2014).
189. Weingärtner, H., Cabrele, C. & Herrmann, C. How ionic liquids can help to stabilize native proteins. *Phys. Chem. Chem. Phys.* **14,** 415–426 (2012).
190. Klähn, M., Lim, G. S., Seduraman, A. & Wu, P. On the different roles of anions and cations in the solvation of enzymes in ionic liquids. *Phys. Chem. Chem. Phys.* **13,** 1649–1662 (2011).
191. Lai, J.-Q., Li, Z., Lü, Y.-H. & Yang, Z. Specific ion effects of ionic liquids on enzyme activity and stability. *Green Chem.* **13,** 1860–1868 (2011).
192. Hajduczki, A., Majumdar, S., Fricke, M., Brown, I. A. M. & Weiss, G. A. Solubilization of a Membrane Protein by Combinatorial Supercharging. *ACS Chem. Biol.* **6,** 301–307 (2011).
193. Lee, C. C., Perchiacca, J. M. & Tessier, P. M. Toward aggregation-resistant antibodies by design. *Trends Biotechnol.* **31,** 612–620 (2013).
194. Simeonov, P., Berger-Hoffmann, R., Hoffmann, R., Strater, N. & Zuchner, T. Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Eng. Des. Sel.* **24,** 261–268 (2010).
195. Burney, P. R., Nordwald, E. M., Hickman, K., Kaar, J. L. & Pfaendtner, J. Molecular dynamics investigation of the ionic liquid/enzyme interface: Application to engineering enzyme surface charge. *Proteins Struct. Funct. Bioinforma.* **83,** 670–680 (2015).
196. Ilmberger, N. *et al.* Metagenomic cellulases highly tolerant towards the presence of ionic liquids—linking thermostability and halotolerance. *Appl. Microbiol. Biotechnol.* **95,** 135–146 (2012).
197. Liang, C. *et al.* Cloning and characterization of a thermostable and halo-tolerant endoglucanase from Thermoanaerobacter tengcongensis MB4. *Appl. Microbiol. Biotechnol.* **89,** 315–326 (2011).
198. Zhang, T. *et al.* Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. *Green Chem.* **13,** 2083–2090 (2011).
199. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.* **129,** 10110 (2007).
200. Der, B. S. *et al.* Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability. *PloS One* **8,** e64363 (2013).
201. Miklos, A. E. *et al.* Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* **19,** 449–455 (2012).
202. Li, L. *et al.* DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **5,** 9 (2012).

203. Li, H. *et al.* Thermostabilization of extremophilic Dictyoglomus thermophilum GH11 xylanase by an N-terminal disulfide bridge and the effect of ionic liquid [emim] OAc on the enzymatic performance. *Enzyme Microb. Technol.* **53,** 414–419 (2013).
204. Wahlström, R., Rovio, S. & Suurnäkki, A. Partial enzymatic hydrolysis of microcrystalline cellulose in ionic liquids by Trichoderma reesei endoglucanases. *RSC Adv.* **2,** 4472–4480 (2012).
205. Wahlström, R., King, A., Parviainen, A., Kruus, K. & Suurnäkki, A. Cellulose hydrolysis with thermo-and alkali-tolerant cellulases in cellulose-dissolving superbase ionic liquids. *RSC Adv.* **3,** 20001–20009 (2013).
206. Wahlström, R., Rahikainen, J., Kruus, K. & Suurnäkki, A. Cellulose hydrolysis and binding with Trichoderma reesei Cel5A and Cel7A and their core domains in ionic liquid solutions. *Biotechnol. Bioeng.* **111,** 726–733 (2014).
207. Li, W., Wang, L., Zhou, R. & Mu, Y. Ionic liquid induced inactivation of cellobiohydrolase I from Trichoderma reesei. *Green Chem.* **17,** 1618–1625 (2015).
208. Jaeger, V. & Pfaendtner, J. Structure, dynamics, and activity of xylanase solvated in binary mixtures of ionic liquid and water. *ACS Chem. Biol.* **8,** 1179–1186 (2013).
209. Tee, K. L. *et al.* Ionic liquid effects on the activity of monooxygenase P450 BM-3. *Green Chem.* **10,** 117–123 (2008).
210. Yang, Z. *et al.* Importance of the ionic nature of ionic liquids in affecting enzyme performance. *J. Biochem. (Tokyo)* **145,** 355–364 (2009).
211. Attri, P., Venkatesu, P. & Kumar, A. Activity and stability of α-chymotrypsin in biocompatible ionic liquids: enzyme refolding by triethyl ammonium acetate. *Phys. Chem. Chem. Phys.* **13,** 2788–2796 (2011).
212. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
213. Brandt, A., Gräsvik, J., Hallett, J. P. & Welton, T. Deconstruction of lignocellulosic biomass with ionic liquids. *Green Chem.* **15,** 550–583 (2013).
214. Mora-Pale, M., Meli, L., Doherty, T. V., Linhardt, R. J. & Dordick, J. S. Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass. *Biotechnol. Bioeng.* **108,** 1229–1245 (2011).
215. Vancov, T., Alston, A.-S., Brown, T. & McIntosh, S. Use of ionic liquids in converting lignocellulosic material to biofuels. *Renew. Energy* **45,** 1–6 (2012).
216. Portillo, M. del C. & Saadeddin, A. Recent trends in ionic liquid (IL) tolerant enzymes and microorganisms for biomass conversion. *Crit. Rev. Biotechnol.* **35,** 294–301 (2015).
217. Xu, J., Xiong, P. & He, B. Advances in improving the performance of cellulase in ionic liquids for lignocellulose biorefinery. *Bioresour. Technol.* **200,** 961–970 (2016).
218. Burney, P. R. & Pfaendtner, J. Structural and Dynamic Features of Candida rugosa Lipase 1 in Water, Octane, Toluene, and Ionic Liquids BMIM-PF6 and BMIM-NO3. *J. Phys. Chem. B* **117,** 2662–2670 (2013).
219. Kim, H. S., Ha, S. H., Sethaphong, L., Koo, Y.-M. & Yingling, Y. G. The relationship between enhanced enzyme activity and structural dynamics in ionic liquids: a combined computational and experimental study. *Phys. Chem. Chem. Phys.* **16,** 2944–2953 (2014).

220. Klähn, M., Lim, G. S. & Wu, P. How ion properties determine the stability of a lipase enzyme in ionic liquids: A molecular dynamics study. *Phys. Chem. Chem. Phys.* **13,** 18647–18660 (2011).
221. Latif, M. A. M., Micaêlo, N. M. & Rahman, M. B. A. Influence of anion–water interactions on the behaviour of lipases in room temperature ionic liquids. *RSC Adv.* **4,** 48202–48211 (2014).
222. Latif, M. A. M., Tejo, B. A., Abedikargiban, R., Abdul Rahman, M. B. & Micaêlo, N. M. Modeling stability and flexibility of α-Chymotrypsin in room temperature ionic liquids. *J. Biomol. Struct. Dyn.* **32,** 1263–1273 (2013).
223. Micaêlo, N. M. & Soares, C. M. Protein structure and dynamics in ionic liquids. Insights from molecular dynamics simulation studies. *J. Phys. Chem. B* **112,** 2566–2572 (2008).
224. Ghosh, S., Parui, S., Jana, B. & Bhattacharyya, K. Ionic liquid induced dehydration and domain closure in lysozyme: FCS and MD simulation. *J. Chem. Phys.* **143,** 125103 (2015).
225. Sprenger, K. G., Choudhury, A., Kaar, J. L. & Pfaendtner, J. The Lytic Polysaccharide Monooxygenases ScLPMO10B and ScLPMO10C Are Stable in Ionic Liquids as Determined by Molecular Simulations. *J. Phys. Chem. B* (2016).
226. Ghaedizadeh, S. *et al.* Understanding the molecular behaviour of Renilla luciferase in imidazolium-based ionic liquids, a new model for the α/β fold collapse. *Biochem. Eng. J.* **105,** 505–513 (2016).
227. Haberler, M. & Steinhauser, O. On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study. *Phys. Chem. Chem. Phys.* **13,** 17994–18004 (2011).
228. Haberler, M., Schröder, C. & Steinhauser, O. Solvation studies of a zinc finger protein in hydrated ionic liquids. *Phys. Chem. Chem. Phys.* **13,** 6955–6969 (2011).
229. Moniruzzaman, M., Kamiya, N. & Goto, M. Activation and stabilization of enzymes in ionic liquids. *Org. Biomol. Chem.* **8,** 2887–2899 (2010).
230. Nordwald, E. M., Armstrong, G. S. & Kaar, J. L. NMR-guided rational engineering of an ionic-liquid-tolerant lipase. *ACS Catal.* **4,** 4057–4064 (2014).
231. Lehmann, C., Bocola, M., Streit, W. R., Martinez, R. & Schwaneberg, U. Ionic liquid and deep eutectic solvent-activated CelA2 variants generated by directed evolution. *Appl. Microbiol. Biotechnol.* **98,** 5775–5785 (2014).
231. Johnson, L. B., Park, S., Gintner, L. P. & Snow, C. D. Characterization of Supercharged Cellulase Activity and Stability in Ionic Liquids. *J. Mol. Catal. B Enzym.* **132**, 84-90 (2016).
233. Henrissat, B. *et al.* Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci.* **92,** 7090–7094 (1995).
234. Canongia Lopes, J. N., Deschamps, J. & Pádua, A. A. Modeling ionic liquids using a systematic all-atom force field. *J. Phys. Chem. B* **108,** 2038–2047 (2004).
235. Köddermann, T., Paschek, D. & Ludwig, R. Molecular dynamic simulations of ionic liquids: A reliable description of structure, thermodynamics and dynamics. *ChemPhysChem* **8,** 2464–2470 (2007).
236. Porter, A. R., Liem, S. Y. & Popelier, P. L. Room temperature ionic liquids containing low water concentrations—a molecular dynamics study. *Phys. Chem. Chem. Phys.* **10,** 4240–4248 (2008).

237. Qiao, B., Krekeler, C., Berger, R., Delle Site, L. & Holm, C. Effect of anions on static orientational correlations, hydrogen bonds, and dynamics in ionic liquids: A simulational study. *J. Phys. Chem. B* **112,** 1743–1751 (2008).
238. Kowsari, M. H., Alavi, S., Ashrafizaadeh, M. & Najafi, B. Molecular dynamics simulation of imidazolium-based ionic liquids. I. Dynamics and diffusion coefficient. *J. Chem. Phys.* **129,** 224508 (2008).
239. Kowsari, M. H., Alavi, S., Ashrafizaadeh, M. & Najafi, B. Molecular dynamics simulation of imidazolium-based ionic liquids. II. Transport coefficients. *J. Chem. Phys.* **130,** 014703 (2009).
240. Latif, M. A. M., Micaêlo, N. & Rahman, M. B. A. Solvation free energies in [bmim]-based ionic liquids: Anion effect toward solvation of amino acid side chain analogues. *Chem. Phys. Lett.* **615,** 69–74 (2014).
241. Monk, J., Singh, R. & Hung, F. R. Effects of pore size and pore loading on the properties of ionic liquids confined inside nanoporous CMK-3 carbon materials. *J. Phys. Chem. C* **115,** 3034–3042 (2011).
242. Jorgensen, W. L. & Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6665–6670 (2005).
243. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PloS One* **7,** e32131 (2012).
244. Tang, J., Li, S., Zhai, Q., Jiang, Y. & Hu, M. Measurements and Correlations of the Solid–Liquid Equilibrium of RbCl/CsCl+[C n mim] Cl (n= 2, 4, 6, 8)+ H2O Ternary Systems at T=(288.15, 298.15, and 308.15) K. *J. Chem. Eng. Data* **59,** 726–735 (2014).
245. Wu, D., Wu, B., Zhang, Y. M. & Wang, H. P. Density, Viscosity, Refractive Index and Conductivity of 1-Allyl-3-methylimidazolium Chloride+ Water Mixture†. *J. Chem. Eng. Data* **55,** 621–624 (2009).
246. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30,** 2157–2164 (2009).
247. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
248. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).
249. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).
250. Vrbka, L., Jungwirth, P., Bauduin, P., Touraud, D. & Kunz, W. Specific ion effects at protein surfaces: a molecular dynamics study of bovine pancreatic trypsin inhibitor and horseradish peroxidase in selected salt solutions. *J. Phys. Chem. B* **110,** 7036–7043 (2006).
251. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
252. Bose, S., Barnes, C. A. & Petrich, J. W. Enhanced stability and activity of cellulase in an ionic liquid and the effect of pretreatment on cellulose hydrolysis. *Biotechnol. Bioeng.* **109,** 434–443 (2012).

253. Xia, S., Baker, G. A., Li, H., Ravula, S. & Zhao, H. Aqueous ionic liquids and deep eutectic solvents for cellulosic biomass pretreatment and saccharification. *RSC Adv.* **4,** 10586–10596 (2014).
254. Li, G. *et al.* [CuCl n] 2- n Ion-Pair Species in 1-Ethyl-3-methylimidazolium Chloride Ionic Liquid- Water Mixtures: Ultraviolet- Visible, X-ray Absorption Fine Structure, and Density Functional Theory Characterization. *J. Phys. Chem. B* **114,** 12614–12622 (2010).
255. Sitkoff, D., Lockhart, D. J., Sharp, K. A. & Honig, B. Calculation of electrostatic effects at the amino terminus of an alpha helix. *Biophys. J.* **67,** 2251–2260 (1994).
256. Shu, Y., Liu, M., Chen, S., Chen, X. & Wang, J. New insight into molecular interactions of imidazolium ionic liquids with bovine serum albumin. *J. Phys. Chem. B* **115,** 12306–12314 (2011).
257. Chen, A. H. & Silver, P. A. Designing biological compartmentalization. *Trends Cell Biol.* **22,** 662–670 (2012).
258. Conrado, R. J., Varner, J. D. & DeLisa, M. P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* **19,** 492–499 (2008).
259. Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **134,** 5516–5519 (2012).
260. Zhao, Z. *et al.* Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion. *Nat. Commun.* **7,** (2016).
261. Fu, J. *et al.* Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* **9,** 531–536 (2014).
262. Vriezema, D. M. *et al.* Self-assembled nanoreactors. *Chem. Rev.* **105,** 1445–1490 (2005).
263. Berman, H. M. *et al.* Trendspotting in the Protein Data Bank. *FEBS Lett.* **587,** 1036–1045 (2013).
264. Kowalski, A. E. *et al.* Gold nanoparticle capture within protein crystal scaffolds. *Nanoscale* **8,** 12693–12696 (2016).
265. Huber, T. R. *et al.* Programmed Assembly of Host-Guest Protein Crystals. *(submitted)*
266. Trogadas, P., Nigra, M. M. & Coppens, M.-O. Nature-inspired optimization of hierarchical porous media for catalytic and separation processes. *New J. Chem.* **40,** 4016–4026 (2016).
267. Henriksen, A. *et al.* Structural interactions between horseradish peroxidase C and the substrate benzhydroxamic acid determined by X-ray crystallography. *Biochemistry (Mosc.)* **37,** 8054–8060 (1998).
268. Wohlfahrt, G. *et al.* 1.8 and 1.9 A resolution structures of the Penicillium amagasakiense and Aspergillus niger glucose oxidases as a basis for modelling substrate complexes. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 969–977 (1999).
269. Majorek, K. A. *et al.* Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol. Immunol.* **52,** 174–182 (2012).
270. Kavanagh, K. L., Klimacek, M., Nidetzky, B. & Wilson, D. K. The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from Candida tenuis. *Biochemistry (Mosc.)* **41,** 8785–8795 (2002).

271. Musille, P. & Ortlund, E. Structure of glycerol dehydrogenase from Serratia. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70,** 166–172 (2014).
272. Yuzugullu, Y. *et al.* Structure, recombinant expression and mutagenesis studies of the catalase with oxidase activity from Scytalidium thermophilum. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 398–408 (2013).
273. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).
274. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4,** 251–256 (2007).
275. Ma, X. *et al.* A Biocompatible and Biodegradable Protein Hydrogel with Green and Red Autofluorescence: Preparation, Characterization and In Vivo Biodegradation Tracking and Modeling. *Sci. Rep.* **6,** (2016).
276. Evers, T. H., Appelhof, M. A. M., Meijer, E. W. & Merkx, M. His-tags as Zn(II) binding motifs in a protein-based fluorescent sensor. *Protein Eng. Des. Sel.* **21,** 529–536 (2008).
277. Højgaard, C. *et al.* A Soluble, Folded Protein without Charged Amino Acid Residues. *Biochemistry (Mosc.)* **55,** 3949–3956 (2016).
278. Jang, E. & Koh, W.-G. Multiplexed enzyme-based bioassay within microfluidic devices using shape-coded hydrogel microparticles. *Sens. Actuators B Chem.* **143,** 681–688 (2010).

8. APPENDICES

APPENDIX A

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Aug 05, 2016

This Agreement between Lucas Johnson ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	3922680491003
License date	Aug 05, 2016
Licensed content publisher	Oxford University Press
Licensed content publication	Protein Engineering, Design and Selection
Licensed content title	Discriminating between stabilizing and destabilizing protein design mutations via recombination and simulation:
Licensed content author	Lucas B. Johnson, Lucas P. Gintner, Sehoo Park, Christopher D. Snow
Licensed content date	08/01/2015
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Engineering stabilized enzymes via computational design and immobilization
Publisher of your work	n/a
Expected publication date	Aug 2016
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Requestor Location	Lucas Johnson 125 2nd St FORT COLLINS, CO 80524 United States Attn: Lucas B Johnson
Publisher Tax ID	GB125506730
Billing Type	Invoice
Billing Address	Lucas B Johnson 125 2nd St FORT COLLINS, CO 80524 United States Attn: Lucas B Johnson
Total	0.00 USD
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

APPENDIX B

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 05, 2016

This Agreement between Lucas Johnson ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	3922681485102
License date	Aug 05, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Molecular Catalysis B: Enzymatic
Licensed Content Title	Characterization of supercharged cellulase activity and stability in ionic liquids
Licensed Content Author	Lucas B. Johnson,Sehoo Park,Lucas P. Gintner,Christopher D. Snow
Licensed Content Date	October 2016
Licensed Content Volume Number	132
Licensed Content Issue Number	n/a
Licensed Content Pages	7
Start Page	84
End Page	90
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Engineering stabilized enzymes via computational design and immobilization
Expected completion date	Aug 2016
Estimated size (number of pages)	158
Elsevier VAT number	GB 494 6272 12
Requestor Location	Lucas Johnson 125 2nd St FORT COLLINS, CO 80524 United States Attn: Lucas B Johnson
Total	0.00 USD

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available

version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above:
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.