농학박사학위논문

열 안정성 향상을 위한 아데닐레이트 카이네이즈 구조적 특성 연구

Structural Optimization for Thermal Stabilization of Adenylate Kinases

2016년 8월

서울대학교 대학원

농생명공학부 응용생명화학전공

문 소 진

A Dissertation for the Degree of Doctor of Philosophy

Structural Optimization for Thermal Stabilization of Adenylate Kinases

August 2016

Sojin Moon

Applied Life Chemistry Major

Department of Agricultural Biotechnology

Seoul National University

Abstract

Thermally stable proteins are desirable in many industrial and laboratory settings. Numerous approaches have been developed to redesign proteins for higher thermal stability. However, these approaches are sometimes unpredictable and bring no significant change. In this study, I systematically tested the effects of multiple stabilization techniques, thereby suggesting an integrated approach for improving protein thermal stability. Using a mesophilic adenylate kinase from *Bacillus subtilis* (AKmeso) as a model protein, I identified mutations important for thermal stability, based on various techniques including local structural entropy optimization, structure-guided mutagenesis, and molecular evolution. Variants of adenylate kinase (AK) harboring single or multiple mutations were generated and the changes in their thermal stabilities were measured. Compared to AKmeso, the AK variants displayed higher thermal denaturation midpoints with retained enzymatic activity at high temperatures. In addition, I solved crystal structures of representative AK variants to confirm the effects of the generated mutations affecting the thermal stability. These results provide a unique example of the integrated and systematic approaches for protein thermal stabilization.

Key words : adenylate kinase, crystal structure, local structural entropy, mutagenesis, protein engineering, thermal stability

***Student Number*: 2010-23449**

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AK	adenylate kinase
Ap_5A	P^1,P^5-di(adenosine 5')-pentaphosphate
CD	circular dichroism
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
NMR	nuclear magnetic resonance spectroscopy
PDB	protein data bank
RMSD	root-mean-square deviation
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
T_m	thermal denaturation midpoint

Introduction

Protein thermal stabilization

A function of protein depends on its three dimensional structure, which is often disrupted by external factors such as heat, extreme changes in pH, or addition of chemical denaturants (Wolynes et al., 1995). In native state protein, a proper protein fold sustains its whole structure (Ooi, 1994).

Most proteins have evolved to carry out their function at their optimal temperatures. However, proteins losing their structures at high temperature cannot function properly (Mirsky and Pauling, 1936; Oobatake and Ooi, 1993). This phenomenon is called thermal denaturation, which occurs at temperature higher than optimal range of a protein (Khechinashvili et al., 1995; Karplus, 2011; Mallamace et al., 2016).

Thermal denaturation midpoint (T_m) value is an intrinsic property that represents the stability of a protein. When the proteins possess higher stability, they have relatively thermal stability at elevated temperatures. In general, thermostable proteins have unique features such as highly compact structures, abundance of non-covalent intramolecular and intermolecular interactions such as electrostatic interactions, hydrophobic effects, and hydrogen bonds (Pauling et al., 1951; Jaenicke, 2000). Various studies aimed to increase the protein thermal stability have been performed on the basis of these features. In addition, search for identifying other factors that might contribute to the thermal stability have been carried out.

Importance of protein stability

Protein thermal stabilization is an important subject in several areas (Vieille and Zeikus, 2001; Schoemaker et al., 2003; Eijsink et al., 2004; Turner et al., 2007; Unsworth et al., 2007; Daniel and Danson, 2010; Liszka et al., 2012; Adrio and Demain, 2014). In manufacturing protein-based pharmaceuticals, processing at high temperatures accelerates reaction rate and reduces microbial contamination. An increased rate of catalysis allows development for more efficient and robust catalysis in industrial processes. Thus, thermally stable proteins can increase the productivity of biosynthesis in a large scale.

Additionally, various incurable diseases such as Alzheimer's disease and bovine spongiform encephalopathy are caused by loss of particular protein folding. These proteins play important roles in entire cell stability, thereby affecting the survival of species.

Increasing thermo-stability of therapeutic proteins, the effectiveness of drug could be applied in gene therapy. In laboratory settings, thermally stable proteins are easier both to store and handle at room temperature, and are often required in experiments as thermally stable DNA polymerases are used of in polymerase chain reaction. Thus, researchers have carried out many studies about improving thermally stable proteins by using various techniques.

Furthermore, the demand for enhanced protein stability in other fields, such as food, and alternative fuels, are rapidly increasing. Accordingly, numerous approaches have been developed to identify the factors influencing the thermal stability, and to redesign target proteins to be more thermally stable.

Adenylate kinase

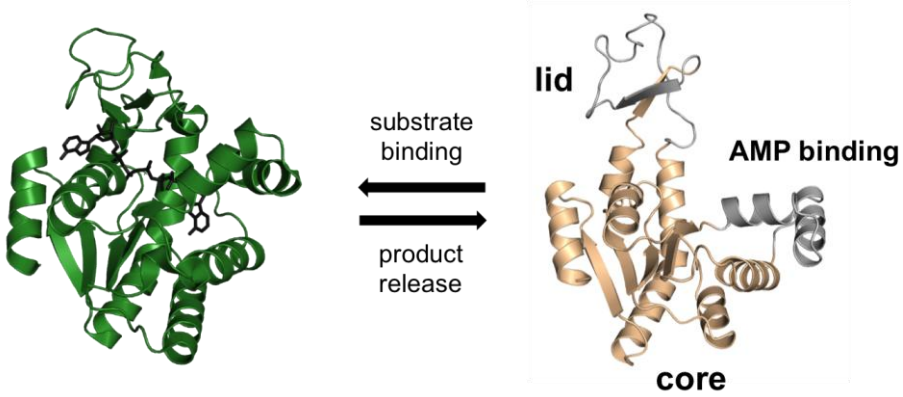
Adenylate kinase (AK) catalyzes reversible conversions between ATP/AMP and two ADP molecules, maintaining equilibrium of cellular adenylate levels. AK was first identified as myokiase from rabbit muscle (Sidney and Herman, 1943). AK is a key protein involved in regulation of cell survival and cell death by controlling the adenylate levels (Counago and Shamoo, 2005). In human, dysfunctional AKs are connected to several diseases, such as heart failure, metabolic diseases, and neurodegenerative diseases (Corrons et al., 2003; Dzeja et al., 2007).

AK has been reported to be comprised of three representative domains: core, lid, and AMP binding domain (Schulz et al., 1990). The AMP binding domain is the site of AMP binding, whereas the lid domain covers ATP within the conserved P-loop. The core domain is relatively rigid except two loops that binds substrate. This domain plays an important role in the overall thermal stability (Bae and Phillips, 2006). Inhibition of AK has been studied in various sources, and Ap_5A is known as the most effective inhibitor binding to the substrate-binding site (Lienhard and Secemski, 1973; Reinstein et al., 1990).

Structure of AK has dramatic conformational change when it is inhibited by bisubsstrate analog Ap_5A (Fig. 1). Additionally, a cofactor Mg^{2+} ion activates phosphoryl transfer in the active site of AK. Kerns et al. (2015) reported the result of Mg^{2+} induced conformational change by molecular dynamics simulations. The conformational change of AK structure in presence Mg^{2+} accelerates opening of the lid domain during catalysis.

Figure 1. Conformational change of adenylate kinase

Crystal structures of AK represent both open (*E.coli* AK) (Muller et al., 1996) and closed (*Bacillus* AK) (Bae and Phillips, 2004) form. *Bacillus* AK show the bound inhibitor Ap_5A molecule as a black stick. *E.coli* AK consists its three domains: core, lid, and AMP binding domain.


substrate
binding
product
release
lid
AMP binding
core

In addition, NMR study (Wolf-Watz et al., 2004) and ensemble-averaged fluorescence transfer experiments (Sinev et al., 1996) showed the movements of open and closed forms of AK. Formoso et al. (2015) reveals drug target related to a key role during conformational transitions based on molecular dynamics simulations of AK.

AKs have been reported to have two different oligomeric states so far; monomeric and trimeric. In eubacteria, AKs were only found as monomeric states (Bae and Phillips, 2004; Thach and Lee, 2014), whereas AKs isolated from archaebacteria exist as trimeric forms (Criswell et al., 2003). Also, crystal structures of AKs have been represented closed form complexed with natural substrates or the inhibition of Ap_5A or open form without their substrates (Vonrhein et al., 1998; Criswell et al., 2003; Thach et al., 2014).

Bacillus adenylate kinase

As an experimental model, AKs from *Bacillus* are used in this study. Crystal structures of *Bacillus* AKs have been determined using X-ray crystallography (Fig. 2). In addition, AK has been used as the subject of several studies related to thermal stabilization using various methods. (Bae and Phillips, 2004; 2005; 2006; Counago et al., 2006; Bae et al., 2008; Bannen et al., 2008; Counago et al., 2008; Miller et al., 2010).

The present study uses AKs isolated from a thermophile *Bacillus stearothermophilus* (AKthermo), a mesophile *Bacillus subtilis* (AKmeso), and a psychrophile *Bacillus globisporus* (AKpsychro) (Berry and Phillips, 1998; Bae and Phillips, 2004). AKthermo, AKmeso and AKpsychro have many similar properties.

They have the same number of amino acids (217 amino acids), a high sequence identity (~70%), and similar overall structures (Figs. 2 and 3). AKpsychro, AKmeso, and AKthermo have very different thermal stability, whereas their amino acids are highly conserved (Table 1). Consequently, these little differences lead to great discrepancy in protein thermostability. Therefore, the *Bacillus* AKs are selected as a model to study the mechanism towards increased thermal stability.

Figure 2. Structures of *Bacillus* AKs

Structures of wild-type AKs from *Bacillus* are (A) AKthermo, (B) AKmeso, and (C) AKpsychro in red, green, and blue, respectively. Wild-type AKs show the bound inhibitor Ap_5A molecule as a black stick. PDB entry: (A) 1ZIO, (B) 1P3J, and (C) 1S3G.


A
B
C

Figure 3. Sequence alignment of *Bacillus* AKs

Sequence alignment of wild-type AKs shows conserved residues as black boxes.

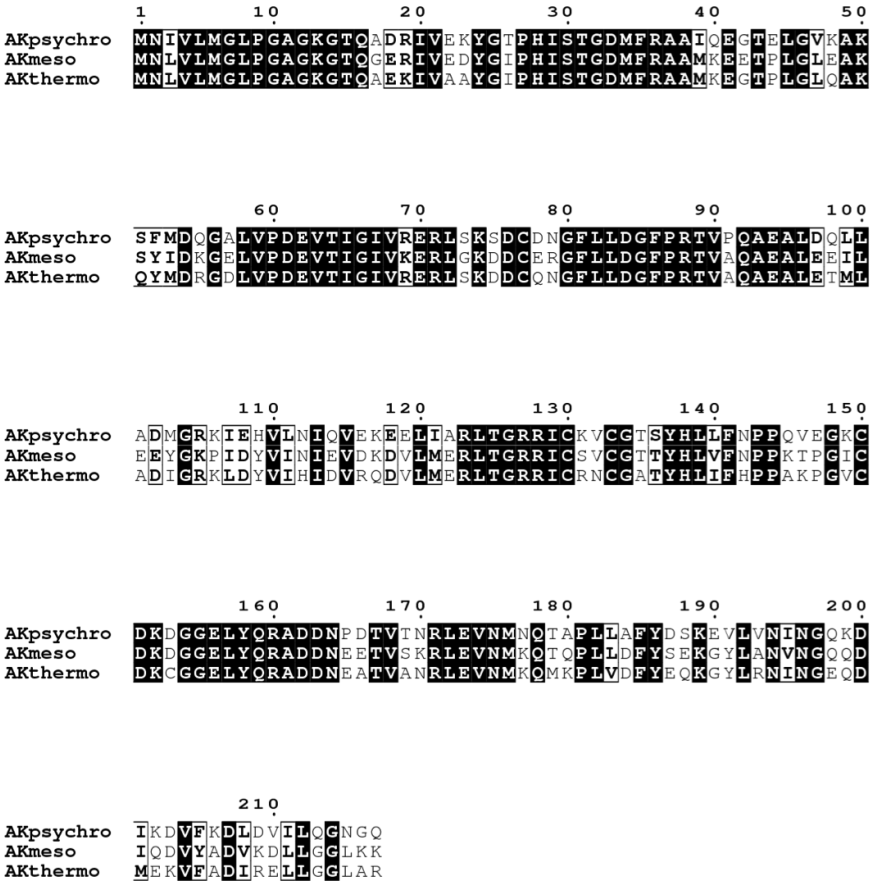

AKpsychro
AKmeso
AKthermo

Table 1. T_m values of wild-type AKs

AK	Source organism	T_m (°C)
AKpsychro	*Bacillus globisporus*	43.3
AKmeso	*Bacillus subtilis*	47.6
AKthermo	*Bacillus stearothermophilus*	74.5[a]

[a]From Glaser et al. 1992.

Previous approaches on *Bacillus* AKs thermostabilization

Here, I provide previous approaches to increase thermal stability of *Bacillus* AKs. There are three strategies for improving protein stability: comparative method, experimental evolution method, and computational method.

Firstly, a comparative method allows designing for improving protein stability by comparing homologues sequences and/or structures. (Vieille and Zeikus, 2001; Eijsink et al., 2004; Korkegian et al., 2005; Razvi and Scholtz, 2006; Kiss et al., 2013). In this method, point mutations to the protein sequence are designed to introduce stabilizing structural elements such as ion pairs, hydrogen bonds, and hydrophobic interactions.

Although there are many successful cases, this method requires detailed analysis of the target protein structure to identify the necessary mutations, making the method difficult to be generalized and automated. However, the modified structural features often have unpredictable results in proteins. Therefore, it is difficult to develop a generalized stabilization strategy based on the comparative study alone.

A comparison of the wild-type AK structures revealed a hydrophobic residue that exists only in AKthermo, which could potentially contribute to thermal stability through favorable hydrophobic interactions (Bae and Phillips, 2004) (Fig. 4 and Table 2). In the crystal structure of AKthermo, the side chain of Met179 fits into a hydrophobic pocket formed by three other hydrophobic residues (Met6, Gly7, and Leu8). Although these three hydrophobic residues are conserved in the wild-type AKs, Met179 is substituted with Thr179 in AKmeso.

Figure 4. Hydrophobic packing around a pocket formed by Met6, Gly7 and Leu8 in AKthermo

AKthermo (red), AKmeso (green), and AKpsychro (blue) are shown as sticks. Water molecules are represented as red spheres. PDB entry: 1ZIO (AKthermo), 1P3J (AKmeso), and 1S3G (AKpsychro).

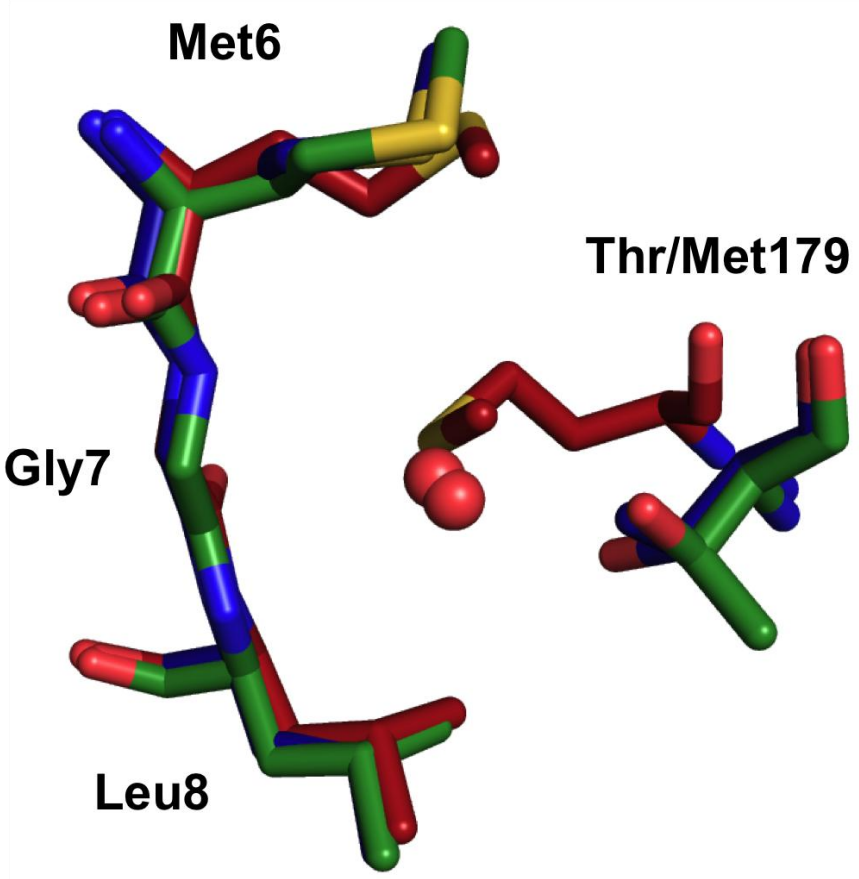

Met6
Thr/Met179
Gly7
Leu8

Table 2. Effect of residue 179 substitution on surface accessibility

AK	Amino acid at position 179	Accessible molecular surface area (Å^2)			
		Met6	Gly7	Leu8	Total
AKpsychro	Thr	3.1	5.3	8.6	17.0
AKmeso	Thr	4.1	6.3	12.3	22.7
AKthermo	Met	1.2	1.2	2.2	4.6

In the crystal structure of AKmeso, a water molecule is found nearby Thr179 and the hydrophobic pocket that the water molecule destabilizes the protein structure. Based on these observations, I decide to mutate Thr179 to methionine in the AKmeso to determine its thermal stabilization effect.

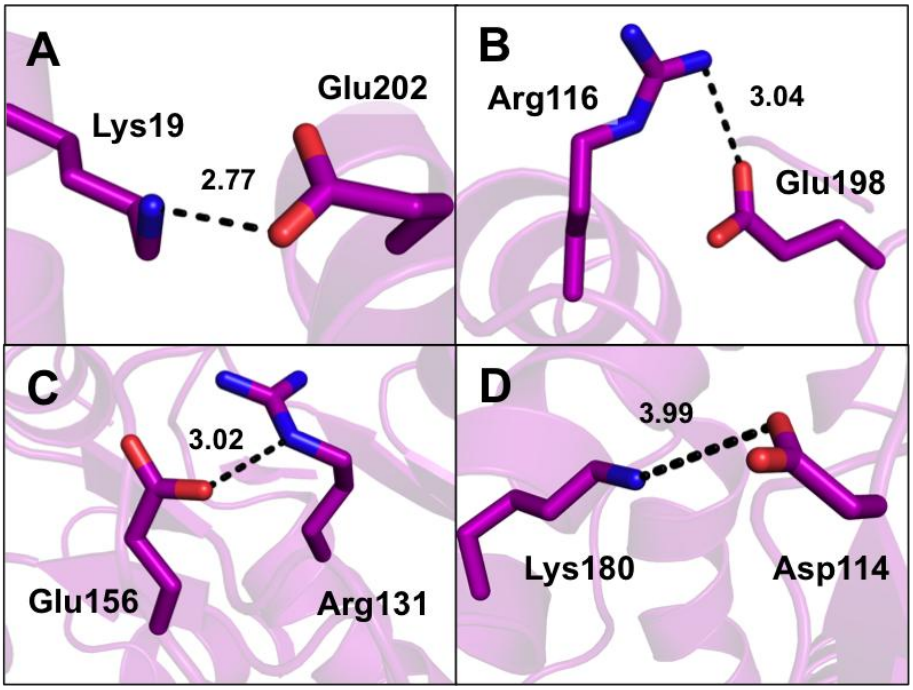
Additionally, four ion pairs of AKthermo are also identified that connect distant regions (>10 residues) of the polypeptide: Lys19-Glu202, Arg116-Glu198, Arg131-Glu156, and Lys180-Asp114 (Bae and Phillips, 2004) (Fig. 5). As Arg131-Glu156 is not located within the core domain, and the Lys180-Asp114 pair is most likely an artifact of crystallization (Bae and Phillips, 2005), therefore I select two ion pairs (Lys19-Glu202 and Arg116-Glu198) as candidates for mutational analysis in AKmeso. Indeed, mutation of these two ion pairs has previously been shown to increase the thermal stability of AKmeso when introduced individually (Bae and Phillips, 2005).

Next, experimental evolution method is a nonstructure-based approach that related to natural selection. Experimental evolution has been applied along with *in vitro* directed evolution and *in vivo* molecular evolution (Wintrode and Arnold, 2000; Eijsink et al., 2005; Schmid, 2011; Traxlmayr and Obinger, 2012; Socha and Tokuriki, 2013; Lane and Seelig, 2014; Currin et al., 2015; Packer and Liu, 2015). These methods are applied to evaluate desired traits such as increased thermal stability.

In directed evolution, gene encoding the target protein is diversified to generate a library of variants using error-prone PCR and DNA shuffling. The variants are then screened for the presence of the desired property. To obtain the best variants, multiple rounds of diversification and selection are performed.

Figure 5. Ion pairs connecting distant regions of a polypeptide in AKthermo

(A) Lys19-Glu202, (B) Arg116-Glu198, (C) Arg131-Glu156, and (D) Lys180-Asp114 are unique ion pairs found in AKthermo compared to AKmeso. They are connecting distant regions of a polypeptide (more than ten residues apart) and their closest oppositely charged atoms are within a 4 Å distance cutoff. PDB entry: 1ZIO.


A
Lys19
2.77
Glu202
B
Arg116
3.04
Glu198
C
3.02
Glu156
Arg131
D
3.99
Lys180
Asp114

Molecular evolution refers to an experimental process in which a target is mutated *in vivo* due to continuous exposure to a specific environmental stress. These methods require no structural information for a given target or comparison to homologues, but may require extensive experimentation and substantial laboratory resources.

Shamoo and coworkers replace the AK of thermophile *B. stearothermophilus* with that of *B. subtilis* and performed experimental evolution under the pressure of increasing growth temperatures (Counago and Shamoo, 2005; Counago et al., 2006). They discover a strain containing two mutations (Gln16Leu and Gln199Arg) in the AK gene, which resulted in increased thermal stability (Counago et al., 2006). In the crystal structures, the two amino acid substitutions appear to improve hydrophobic packing and facilitate the formation of an additional electrostatic network, respectively (Counago et al., 2006; Miller et al., 2010) (Fig. 6).

Lastly, the computational method is a scientific computing approach using computatiional programs for designing stable proteins. The method has been used for designing stable proteins (Damborsky and Brezovsky, 2009; Pantazes et al., 2011; Kiss et al., 2013; Shirke et al., 2016). The developed applications range from redesigning existing protein domains to generating novel protein structures with high stability (Dahiyat and Mayo, 1997; Kuhlman et al., 2003; Korkegian et al., 2005; Dantas et al., 2007). In some cases, this computational approach requires detailed structural information of targets when combined with experimental methods (Kiss et al., 2013; McConnell et al., 2013).

Computing structural entropy, which is one of computational method, calculated the average structural entropy of a query sequence (Chan et al., 2004).

The method is based on structural information derived from structures on the PDB (Berman et al., 2000). Local structural entropy (LSE) describes the extent of conformational heterogeneity in short protein sequences (Chan et al., 2004). By carefully examining structures deposited on the PDB, the LSE of certain amino acid tetramers is calculated to compile a database of 20^4 possible tetramer sequences. In this process, the tetramers that appeared in various secondary structures are assigned higher LSE values than those always adopting a single secondary structural element.

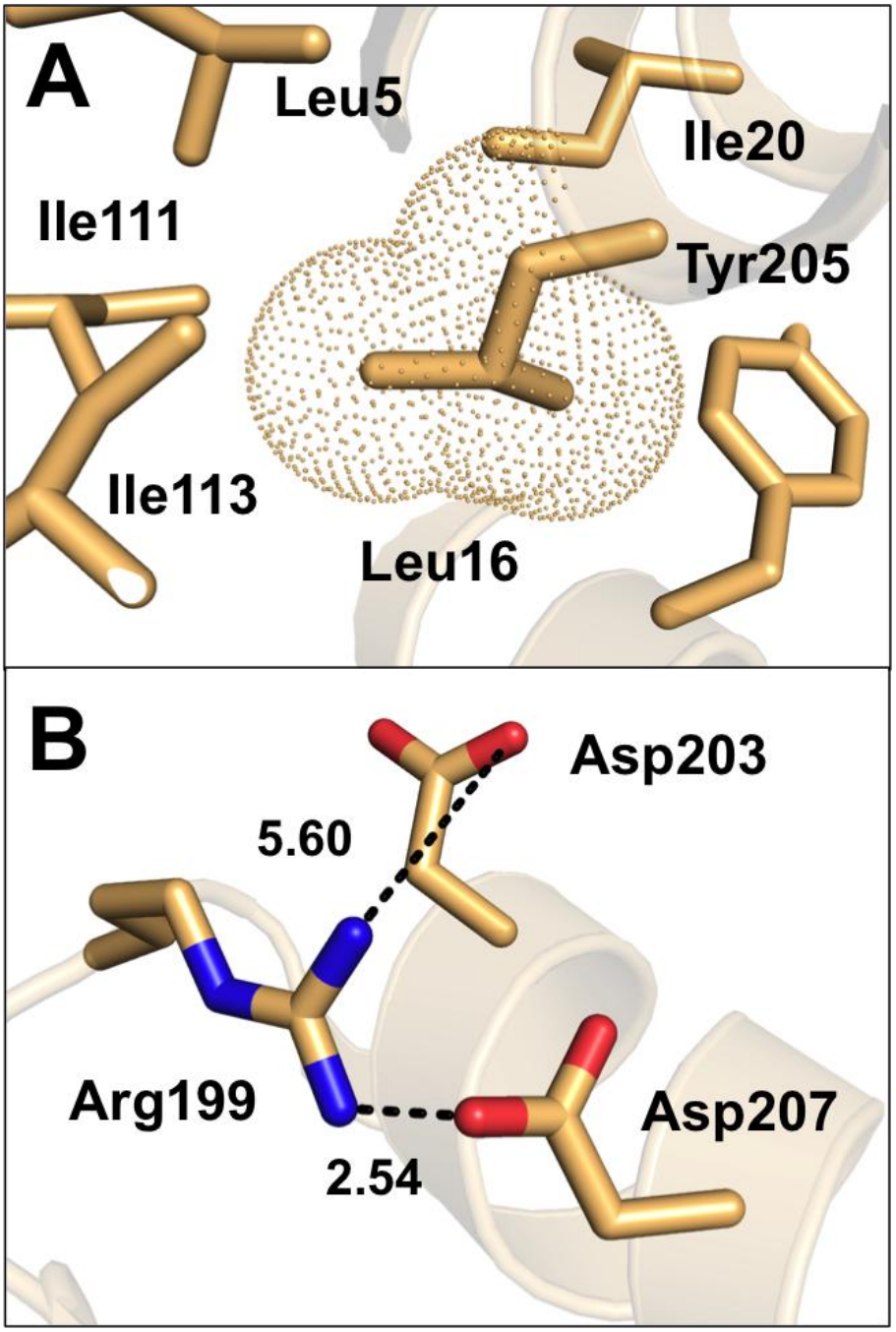
The relationship between structural entropy and the thermal stability of proteins has been studied. Chan and coworkers, who originally developed the concept of LSE, observe a linear relationship between average LSE and T_m values in many protein families. Additionally, Phillips group shows LSE optimization for protein stabilization (Bae et al., 2008; Bannen et al., 2008). Reducing the LSE through the introduction of mutations can result in fewer conformational states, sometimes leading to a more stable structure.

In this computational approach, a sequence alignment is performed using a target and many homologous proteins to identify conserved and/or major amino acids in particular positions. This technique is based on the assumption that these "consensus" amino acids are crucial for maintaining the structure shared by homologous proteins and therefore contribute more significantly to stability than do nonconsensus amino acids.

Although detailed structural information or substantial experimental efforts may not be necessary, this bioinformatic approach requires a large number of homologous sequences and often arbitrary constraints when the sequence alignments do not clearly indicate dominant amino acids at given positions.

Figure 6. Improved hydrophobic packing and electrostatic interaction by mutations identified by molecular evolution

(A) Gln16Leu and (B) Gln199Arg are two point mutations increased thermal stability in AKmeso. PDB entry: 2OSB.


A
Leu5
Ile20
Ile111
Tyr205
Ile113
Leu16
B
Asp203
5.60
Arg199
Asp207
2.54

Stable AKmeso variants were generated through optimization of the LSE using *Bacillus* AKs (Bae et al., 2008). AKmeso was substituted several of its amino acid residues with those from AKpsychro, resulting in three LSE optimized AK variants (Fig. 7). The variants were selectively substituted residues in AKmeso with those found in AKpsychro such that the resulting amino acid sequence had the optimal (lowest) LSE value. Only residues in the core domain (residues 1–30, 61–126, and 165–217) are substituted because this central domain almost exclusively controls the overall stability of AK (Bae and Phillips, 2006).

The LSE-optimized AK variant, AKlse1, contains 23 "AKpsychro-like" substitutions relative to wild-type AKmeso. AKlse2 is only slightly less optimized than AKlse1, but contained the largest number (26) of residue substitutions of the top 20 LSE-optimized AK variants. AKlse3 is considerably less LSE-optimized than AKlse1 and AKlse2 and contained only 10 substitutions. AKlse3 is designed to avoid any "AKthermo-like" substitutions, in which the resulting amino acids would be identical to those in AKthermo, although the substitutions are originally chosen based on comparison with AKpsychro.

Despite using a less thermally stable homologue (AKpsychro) for these residue substitutions, the LSE-optimized AK variants (AKlse1, AKlse2, and AKlse3) exhibited considerably higher T_m than that of the AKmeso.

These three variants contained 10 to 23 substituted residues from AKpsychro yet displayed increase in T_m up to 11.3°C (Table 3).

Table 3. T_m values and average LSE values of AKmeso, AKthermo and LSE-optimized AK variants

	AKmeso	AKlse1	AKlse2	AKlse3	AKthermo
T_m (°C)	46.4	57.7	56.2	50.1	74.5[a]
ΔT_m (°C)[b]	-	11.3	9.8	3.7	28.1
Average LSE	1.4938	1.4083	1.4085	1.4603	1.4323

[a]From Glaser et al. 1992.
[b]Difference to T_m of AKmeso

Figure 7. Sequence alignment of wild-type AKs and LSE-optimized AK variants

Sequences of wild-type AKs and LSE-optimized AK variants are aligned. The core domain (residues 1–30, 61–126, and 165–217) is indicates as a black line.

```
            1         10         20         30         40         50
AKpsychro   MNIVLMGLPGAGKGTQADRIVEKYGTPHISTGDMFRAAIQEGTELGVKAK
AKmeso      MNLVLMGLPGAGKGTQGERIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKthermo    MNLVLMGLPGAGKGTQAEKIVAAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKlse1      MNIVLMGLPGAGKGTQAERIVEKYGIPHISTGDMFRAAMKEETPLGLEAK
AKlse2      MNIVLMGLPGAGKGTQAERIVEKYGIPHISTGDMFRAAMKEETPLGLEAK
AKlse3      MNLVLMGLPGAGKGTQGERIVEKYGIPHISTGDMFRAAMKEETPLGLEAK

                      60         70         80         90        100
AKpsychro   SFMDQGALVPDEVTIGIVRERLSKSDCDNGFLLDGFPRTVPQAEALDQLL
AKmeso      SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKthermo    QYMDRGDLVPDEVTIGIVRERLSKDDCQNGFLLDGFPRTVAQAEALETML
AKlse1      SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
AKlse2      SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEELL
AKlse3      SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL

                     110        120        130        140        150
AKpsychro   ADMGRKIEHVLNIQVEKEELIARLTGRRICKVCGTSYHLLFNPPQVEGKC
AKmeso      EEYGKPIDYVINIEVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKthermo    ADIGRKLDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKlse1      EEMGRPIDYVINIQVDKEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKlse2      EEMGRPIDYVINIQVDKEELIARLTGRRICSVCGTTYHLVFNPPKTPGIC
AKlse3      EEMGKPIDYVINIQVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC

                     160        170        180        190        200
AKpsychro   DKDGGELYQRADDNPDTVTNRLEVNMNQTAPLLAFYDSKEVLVNINGQKD
AKmeso      DKDGGELYQRADDNEETVSKRLEVNMKQTQPLLDFYSEKGYLANVNGQQD
AKthermo    DKCGGELYQRADDNEATVANRLEVNMKQMKPLVDFYEQKGYLRNINGEQD
AKlse1      DKDGGELYQRADDNEETVTKRLEVNMKQTAPLLAFYDSKEVLVNVNGQQD
AKlse2      DKDGGELYQRADDNEETVTKRLEVNMKQTAPLLAFYDSKEVLVNVNGQQD
AKlse3      DKDGGELYQRADDNEETVTKRLEVNMKQTAPLLDFYDEKGYLVNVNGQQD

                     210
AKpsychro   IKDVFKDLDVILQGNGQ
AKmeso      IQDVYADVKDLLGGLKK
AKthermo    MEKVFADIRELLGGLAR
AKlse1      IQDVFADVKVILGGLKQ
AKlse2      IQDVFADVKVILGGLKQ
AKlse3      IQDVYADLKVLLGGLKQ
```

Materials and Methods

Cloning

AKmeso variants genes were generated of the *Bacillus* AKs using mismatched primers in polymerase chain reaction by Du-kyo Jung. Among the AKmeso variants, AKv18 gene was redesigned to a new construct for successful crystallization screening. C-terminus truncated AKv18 (212 amino acid residues) was amplified with mutagenic primers. The forward and reverse oligonucleotides used were 5'-CAT ATG AAC ATT GTG CTG ATG GGT-3' and 5'-CTC GAG TTA CAG AAT CAC ATC CAC ATC-3', respectively. The underlined sequences are restriction enzyme sites, NdeI and XhoI, respectively. The truncated construct digested the restriction enzyme was cloned into pET-28a(+) vector (Novagen) containing C-terminal hexa-histidine tag.

Genes of LSE-optimized AK variants were prepared commercially or by polymerase chain reaction using mismatched primers. AKlse1, AKlse2, and AKlse3 were designed and generated by Euiyoung Bae, Ryan M. Bannen, and George N. Phillips Jr. (Bae et al., 2008). I will name AKlse1 to AKv8. AKlse4, AKlse4m1, AKlse5, and AKlse6 were designed and generated by Ryan M. Bannen, Thomas J. Rutkoski, and George N. Phillips Jr. In addition, two AKlse4 mutants (AKlse4m2 and AKlse4m3) were mutated using mismatched primers. (Table 4) The synthetic AKlse4 gene in a pET-11a (Novagen) vector was used as a template for generated the AKlse4 mutants. The mutagenic PCR primers were generated based on a protein sequence of a template using PrimerX (http://www.bioinformatics.org/primerx/).

Table 4. Primers used in the mutagenesis for AKlse4

Name	Direction	Primer sequence
His109Tyr	Forward	5'-GAAATGGGCCGTAAACTGGAAT<u>A</u>TGTAATTCATATCGAAGTG-3'
	Reverse	5'-CACTTCGATATGAATTACAT<u>A</u>TTCCAGTTTACGGCCCATTTC-3'
Arg193Val	Forward	5'-GATAGCAAAGAAGTGCTG<u>GTG</u>AACGTCAACGGCCAGCAG-3'
	Reverse	5'-CTGCTGGCCGTTGACGTT<u>CAC</u>CAGCACTTCTTTGCTATC-3'
Leu211Ile	Forward	5'-TAAAGATCTGCGTGAA<u>ATT</u>CTGCAGGGCCTGGCGC-3'
	Reverse	5'-GCGCCAGGCCCTGCAG<u>AAT</u>TTCACGCAGATCTTTA-3'

The genes were amplified by PCR using the mutagenic primer. PCR was carried out for 30 cycles (denaturation for 10 sec at 98°C, annealing for 5 sec at 57°C - 66°C, elongation for 8 min at 72°C). The PCR products were treated with DpnI then transformed into *Escherichia coli* DH5α stain.

Taq DNA polymerase was purchased from Takara (Japan) and Bioneer (Korea). Restriction enzymes and T4 DNA ligase were purchase from Takara (Japan).

Protein expression and purification

AK genes were transformed into *Escherichia coli* BL21 (DE3) strain. The transformed strain was cultured overnight using Luria broth (LB) medium with ampicillin (50 μg/ml) or kanamycin (50 μg/ml) with shaking at 37°C. The pre-incubaed cells (7 ml) were inoculated into 700 ml of LB medium containing appropriate antibiotics with shaking at 37°C until optical density at 600 nm reached 0.7. AK proteins were then induced for 6 *h* by the addition of IPTG at a final concentration of 0.5 m*M*.

The AK cells were harvested by centrifugation at 5500 g for 8 min at 4°C. After discaring the supernatant, the cell pellets were resuspended in a purification buffer containing 50 m*M* KCl, 10 m*M* $MgCl_2$, 20 m*M* Tris-HCl at pH 7.8. The resuspended cells were lysed using a sonicator for 4 min at 30% amplitude. The cell lysates were centrifuged at 35,000 g for 50 min and filtered to remove denatured proteins and cell debris.

The filtered sample was loaded onto a column containing Affi-Gel blue resin (Bio-Rad, Hercules, CA). The column was washed with the purification buffer, and bound proteins were eluted in applying a linear gradient of NaCl (up to 2 *M*) (Fig. S1). The eluted sample was concentrated using filter device

and filtered with a 0.2 μm filter for injection onto a size exclusion chromatography column. The AK protein was further purified using a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare) equilibrated with a size exclusion chromatography buffer (200 m*M* NaCl, 50 m*M* KCl, 10 m*M* $MgCl_2$, 20 m*M* Tris-HCl at pH 7.8) (Fig. S2). Collected fractions were dialyzed against 10 m*M* HEPES-NaOH at pH 7.0, and concentrated for crystallization. The proteins were judged pure by 12% SDS-PAGE (Fig. S3), and their concentrations were determined by the absorbance at 280 nm.

The cell of AKv18, containing histidine-tagged protein, was harvested by centrifugation, resuspended in the purification buffer (10 m*M* $MgCl_2$, 500 m*M* NaCl, 20 m*M* imidazole, 10% (v/v) glycerol, 5 m*M* β-mercaptoethanol, 20 m*M* Tris at pH 7.8) with 0.25 m*M* phenylmethanesulfonylfluoride and lysed mechanically. For affinity chromatography, supernatant was loaded onto HisTrap HP column (GE Healthcare) and eluted with a linear gradient of 20-500 m*M* imidazole. Eluted protein was further purified by size-exclusion chromatography as described for the other AK proteins.

Competent cells were purchased from Enzynomics (Korea). LB medium, ampicillin, and kanamycin were purchased from Duchefa (Haarlem, The Netherlands). IPTG was purchased from EMD-Calbiochem (San Diego, CA). Other chemicals were purchased from Sigma-Aldrich.

Determination of T_m values

T_m values of AKs were estimated by CD. All CD experiments were performed using Chirascan Plus spectrometer (Applied Photophysics Ltd, Leatherhead, Surrey, UK). Samples of AKs at 41 μ*M* in 10 m*M* potassium phosphate at pH 7.0 were prepared and degassed for 10 min at 20°C. The temperature was

steadily increased from 20°C to 90°C at a rate of 1°C/min, while measuring the absorbance at 220 nm.

Estimated T_m from CD spectra was analyzed by the protocol developed by John and Weeks using SigmaPlot (Systat Software) (John and Weeks, 2000). Thermodynamic information can be extracted using the van't Hoff relationship:

$$\frac{d \ln K}{d(\frac{1}{T})} = \frac{-\Delta H_{vH}}{R} \quad (1)$$

The van't Hoff relationship between the equilibrium constant K and the signal was obtained during denaturation. The equilibrium constant for this reaction is given by

$$K = \frac{D}{N} = \frac{f}{1-f} \quad (2)$$

From the fraction signal alter during thermal denaturation, the fraction denatured protein *f* can be acquired algebraically.

By taking the natural logarithm of both sides of equation 2 and differentiating with respect to (1/T) to obtain:

$$\frac{d \ln K}{d(\frac{1}{T})} = \frac{1}{f}\left(\frac{df}{d\left(\frac{1}{T}\right)}\right) - \frac{1}{1-f}\left(\frac{-df}{d(\frac{1}{T})}\right) \quad (3)$$

Substituting equation 1 and collecting terms gives

$$\frac{df}{d(\frac{1}{T})} = \frac{-\Delta H_{vH}}{R} f(1-f) \quad (4)$$

where *f* is a function of both T_m and ΔH_{vH}.

$$\frac{d(\text{signal})}{dT} = A\, f(1-f)(T^2) \quad (5)$$

where *d*(signal)/*dT* can be the algebraic derivative of any denaturation-sensitive signal in arbitrary units and *A* indicates a scaling factor.

John and Weeks use the relationship $d(1/T) = dT/T^2$.

T_m and ΔH_{vH} are acquired by substituting into equation 5,

$$f = \frac{K}{K+1} \tag{6}$$

and, the integrated form of the van't Hoff equation for unimolecular reactions.

$$K = \exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right] \tag{7}$$

Thermal denaturation curves are applied to equations 5-7 using only three flexible parameters, the physical constants T_m and ΔH_{vH}.

For data analysis, temperature unit was converted from Celsius to Kelvin (273.15 + temperature) and then the smooth data were obtained using CD signal and temperature (Kelvin). The smooth data were derived using temperature for x column, and signal data for y column. The constant, T_m, and enthalpy change were obtained by substituting derivative x and y into equation 8.

$$y = A \times \frac{\exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m} - \frac{1}{x}\right)\right]}{\left\{\exp\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m} - \frac{1}{x}\right)\right] + 1\right\}^2} \times x^2 \tag{8}$$

where A: constant, R: gas constant, x: temperature (Kelvin), y: signal data, T_m: melting temperature, ΔH_{VH}: van't Hoff enthalpy.

Activity assay

The activities of AKs were determined at various temperatures by measuring the rate of ATP formation. The enzymatic reaction at each temperature was

started by addition of 10 μL of 0.24 μg/mL AK to a 0.2 mL reaction mixture (1 m*M* ADP, 1 m*M* glucose, 0.4 m*M* $NADP^+$, 2 m*M* $MgCl_2$, 100 m*M* KCl, 50 m*M* HEPES-NaOH pH 7.4). After 5 min incubation, the reaction was stopped by adding 40 μL of 4 m*M* Ap_5A and keeping the mixture in ice. The amount of ATP produced by the reaction was measured adding the mixture including ATP-dependent $NADP^+$ reduction for 20 min using hexokinase and glucose-6-phosphate dehydrogenase in 50 m*M* HEPES-NaOH at pH 7.4 at room temperature. After 20 min, the absorbance of NADPH then was measured by spectrophotometer at 340 nm. The experiment was performed from 35°C to 95°C incrementing 10°C.

Crystallization

AKv8 crystals were grown at 20°C by the hanging-drop vapor diffusion method from 10 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH pH 7.0) mixed with an equal volume (2 μℓ) of reservoir solution (1.4 *M* trisodium citrate at pH 6.0) (Fig. S4). These crystals were cryoprotected in the reservoir solution supplemented with 20% (v/v) ethylene glycol and flash-frozen in liquid nitrogen.

Crystals of the His-tagged AKv18 were grown at 37°C by the hanging-drop vapor diffusion method from 10 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) mixed with an equal volume (2 μℓ) of reservoir solution (1.6 *M* ammonium citrate at pH 8.0) (Fig. S4). The crystals were cryoprotected in the reservoir solution supplemented with 25% (v/v) ethylene glycol and flash-frozen in liquid nitrogen.

AKlse2 crystals were grown at 20°C by the hanging-drop vapor diffusion method from 10 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) mixed with an equal amount of reservoir solution (1.9 *M* ammonium citrate at pH 6.0) (Fig. S5). The crystals were cryoprotected in the reservoir solution supplemented with 30% (v/v) glycerol and flash-frozen in liquid nitrogen.

AKlse4 crystals were grown at 20°C by the hanging-drop vapor diffusion method from 30 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) mixed with an equal amount of reservoir solution (32% (w/v) polyethylene glycol 4,000, 250 m*M* $MgCl_2$, 100 m*M* Tris at pH 8.5) (Fig. S5). The crystals were cryoprotected in the reservoir solution supplemented with 20% (v/v) glycerol and flash-frozen in liquid nitrogen.

AKlse4m1 crystals were grown at 20°C by the hanging-drop vapor diffusion method with 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) (Fig. S6). AKlse4m1 (30 mg/mL) was mixed with an equal amount of a reservoir solution containing 18% (w/v) polyethylene glycol 3,350, 100 m*M* lithium sulfate, and 100 m*M* Bis-Tris pH at 5.5.

Crystals of AKlse4m2 (18 mg/mL) were obtained from a reservoir solution including 22% (w/v) polyethylene glycol 3,350 and 200 m*M* calcium chloride (Fig. S6). The crystals were cryoprotected in the reservoir solutions supplemented with 20% (v/v) glycerol and flash-frozen in liquid nitrogen.

AKlse5 crystals were grown at 20°C by the hanging-drop vapor diffusion method from 10 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) mixed with an equal amount of reservoir solution (26.5% (w/v) polyethylene glycol 4,000, 150 m*M* magnesium acetate, 100 m*M* sodium cacodylate at pH 6.5) (Fig. S7). The crystals were cryoprotected in the

reservoir solution supplemented with 10% (v/v) glycerol and flash-frozen in liquid nitrogen.

AKlse6 crystals were grown at 20°C by the hanging-drop vapor diffusion method from 13 mg/mL protein and 4 m*M* Ap_5A in a buffer (10 m*M* HEPES-NaOH at pH 7.0) under the same conditions as those used for obtaining AKlse5 crystals (Fig. S7). The crystals were cryoprotected in the reservoir solution supplemented with 10% (v/v) glycerol and flash-frozen in liquid nitrogen.

X-ray data collection and structure determination

Diffraction data were kindly collected at the beamline 5C (for AKv3) by Du-kyo Jung in this laboratory and the beamline 7A (for AKv8 and AKv18) at the Pohang Accelerator Laboratory (South Korea) at 100 K. The diffraction images were processed using HKL2000 (Otwinowski and Minor, 1997) (for AKv3 and AKv18) and iMOSFLM (Battye et al., 2011) (for AKv8). The AKmeso crystal structure (PDB entry 1P3J) (Bae and Phillips, 2004) was used as a starting model for molecular replacement phasing of AKv3 and AKv8 in PHASER (McCoy et al., 2007) by Du-kyo Jung. A solution for AKv18 was found using the AKv8 structure.

Diffraction data of AKlse2, AKlse4, AKlse5, and AKlse6 were collected at 100 K at the beamline 7A of the Pohang Accelerator Laboratory. Diffraction images were processed with HKL2000. PHASER was used for molecular replacement phasing for all crystals. The structure of AKlse1 was used as a search model for AKlse2 and AKlse4. Solutions for AKlse5 and AKlse6 were found using the structure of AKthermo.

Diffraction data for AKlse4m1 and AKlse4m2 were collected at 100 K at beamlines 7A and 5C, respectively, of the Pohang Accelerator Laboratory. The diffraction images were processed using HKL2000. AK structures (PDB entry 4MKG and 1ZIO, respectively) (Berry and Phillips, 1998) were used as starting models for molecular replacement phasing in PHASER for AKlse4m1 and Molrep for AKlse4m2.

The final structures were completed using alternate cycles of manual fitting in COOT (Emsley and Cowtan, 2004) and refinement by REFMAC5 (Murshudov et al., 1997). In the AKv3, AKv8, and AKv18 structures, TLS refinement and Babinet scaling were also used in refinement. The stereochemical quality of the final models were assessed using MolProbity (Chen et al., 2010).

Structure analysis

For structural analysis, the last five C-terminal residues (residues 213–217) were excluded since they were not resolved well in the crystal structure of AKmeso (Bae and Phillips, 2004). The PDB entries of crystal structures of AKpsychro, AKmeso, AKlse1, and AKthermo used for the analysis are 1S3G, 1P3J, 4MKG, and 1ZIO, respectively. Two oppositely charged residues were identified as an ion pair if their closest oppositely charged atoms were within specific cutoff distances (4 or 6 Å). The side-chain oxygen atoms of Asp and Glu residues were considered negatively charged and the side-chain nitrogen atoms of Arg and Lys residues were considered positively charged. For thermal stable AK variants, positive charged atoms also included His residues in addition to the Arg and Lys residues.

Hydrogen bonds were defined using WHAT IF (Rodriguez et al., 1998). Accessible and buried molecular surface areas were calculated with WHAT IF using a probe radius of 1.4 Å. Nitrogen and oxygen atoms were considered polar, and carbon and sulfur atoms as apolar.

Data deposition

The atomic coordinates and structure factors for AKv3, AKv8, AKv18, AKlse2, AKlse4, AKlse5, AKlse6, AKlse4m1 and AKlse4m2 were deposited in the RCSB PDB (Bannen et al., 2008) under the accession codes 4MKF, 4MKG, 4MKH, 4QBF, 4QBG, 4QBH, 4QBI, 4TYP, and 4TYQ, respectively.

Results

Part I

Integrated Approach for Thermal Stabilization of AKmeso

The present study describes the effects of several application methods for thermal stabilization using *Bacillus* AKs. I generated AK variants with enhanced thermal stability. To test the additivity of mutations in different backgrounds for accomplishing more significant thermal stabilization, I initially applied these methods. In addition, I combined the individual stabilizing factor for accomplishing more significant thermal stabilization. Furthermore, I analysed thermal stabilities of AK variants in order to confirm effect of introduced mutations. Identifying effects of individual stabilizing mechanisms in thermal stabilization, crystal structures of representative AK variants were determined.

Identification of stabilizing mutations for AKmeso

AKmeso was designed for improving thermal stability by introducing identified stabilizing factors based on the three approaches. Comparative study optimized structural features such as adding ion pairs or hydrogen bonds and increasing hydrophobic packing by substitutions of amino acids. Experimental evolution was generated selections of targets having adaptation to extreme environments. Computational study compared sequences of proteins based on optimization features of targets.

Mutated residues were only located in core domain (residues 1–30, 61–126, and 165–217) of AKmeso. In the AKmeso, the negatively charged side chain of the Asp23 residue was located adjacent to the positively charged Lys209, while AKpsychro contains the Lys23-Asp209 ion pair (Fig. 8). After the 23 LSE substitutions (AKlse1), I chose a Lys209Asp mutation to recover the ion pair (Lys23-Asp209) in the AKpsychro.

Generation AKmeso variants for thermally stable AKs

AKmeso was redesign by introducing mutations leading to higher overall stability of the protein. As no overlap was seen between any of the 31 identified mutations (Fig. 9). In the process of designing the AK variants, mutations for the positively and negatively charged residues comprising the Lys19-Glu202 and Arg116-Glu198 ion pairs were introduced simultaneously. Although both Gln16Leu and Gln199Arg mutations were selected based on the molecular evolution experiment, they were used separately in some variants due to the distance between residues and the difference in the proposed mechanisms of stabilization.

The residue 23 AKpsychro-like mutations identified by the LSE optimization, including the Asp23Lys mutation, were added together, after which the Lys209Asp mutation necessary for reconstituting the Lys23-Asp209 ion pair was introduced. Mutations introduced into each of the designed AK variants are listed in Table 5. These variants can be divided into two groups depending on whether or not they contain the LSE mutations. Seven variants (AKv1 though AKv7) were designed without the LSE substitutions while the remaining 11 variants (AKv8 though AKv18) were constructed using AKv8 as their template.

While two of the AK variants, AKv4 and AKv8, have been described previously (Counago et al., 2006; Bae et al., 2008; Miller et al., 2010), they were also included in this study in order to examine them under equivalent experimental conditions using the same instruments.

Together, the panel of AK variants contained 1–31 mutated residues, relative to the wild-type AKmeso sequence (Fig. 9 and Table 5). This combination of mutations provided a unique opportunity not only to compare the effectiveness of different stabilization techniques, but also to test the additivity of mutations in different backgrounds for accomplishing more significant thermal stabilization.

Measurement of thermal stability of AKmeso variants

To systematically test and compare the effectiveness and additivity of each stabilizing mutation, I employed CD spectroscopy to examine the thermal stabilities of AKmeso variants (Fig. 10). All enzymes were expressed and purified in the same manner, and their denaturation curves were obtained and analyzed using identical protocols.

The T_m values for each AK are listed in Table 5. All 18 AK variants exhibited higher T_m values than that of AKmeso with the increases varying from 2.5°C (for AKv1) to 26.6°C (for AKv18). The mutations were identified by the comparative study resulted in increased T_m values when introduced into AKmeso. AKv1, in which Thr179 was substituted to a Met residue, displayed higher T_m value than AKmeso by 2.5°C while the two unique ion pairs (Lys19-Glu202 and Arg116-Glu198) found in AKthermo increased the T_m of AKmeso by 4.3°C in AKv2.

Figure 8. Ion pairs between residues 23 and 209 in wild-type AKs

(A) AKmeso and (B) AKpsychro contain ion pairs between residues 23 and 209. PDB entry: (A) 1P3J and (B) 1S3G.

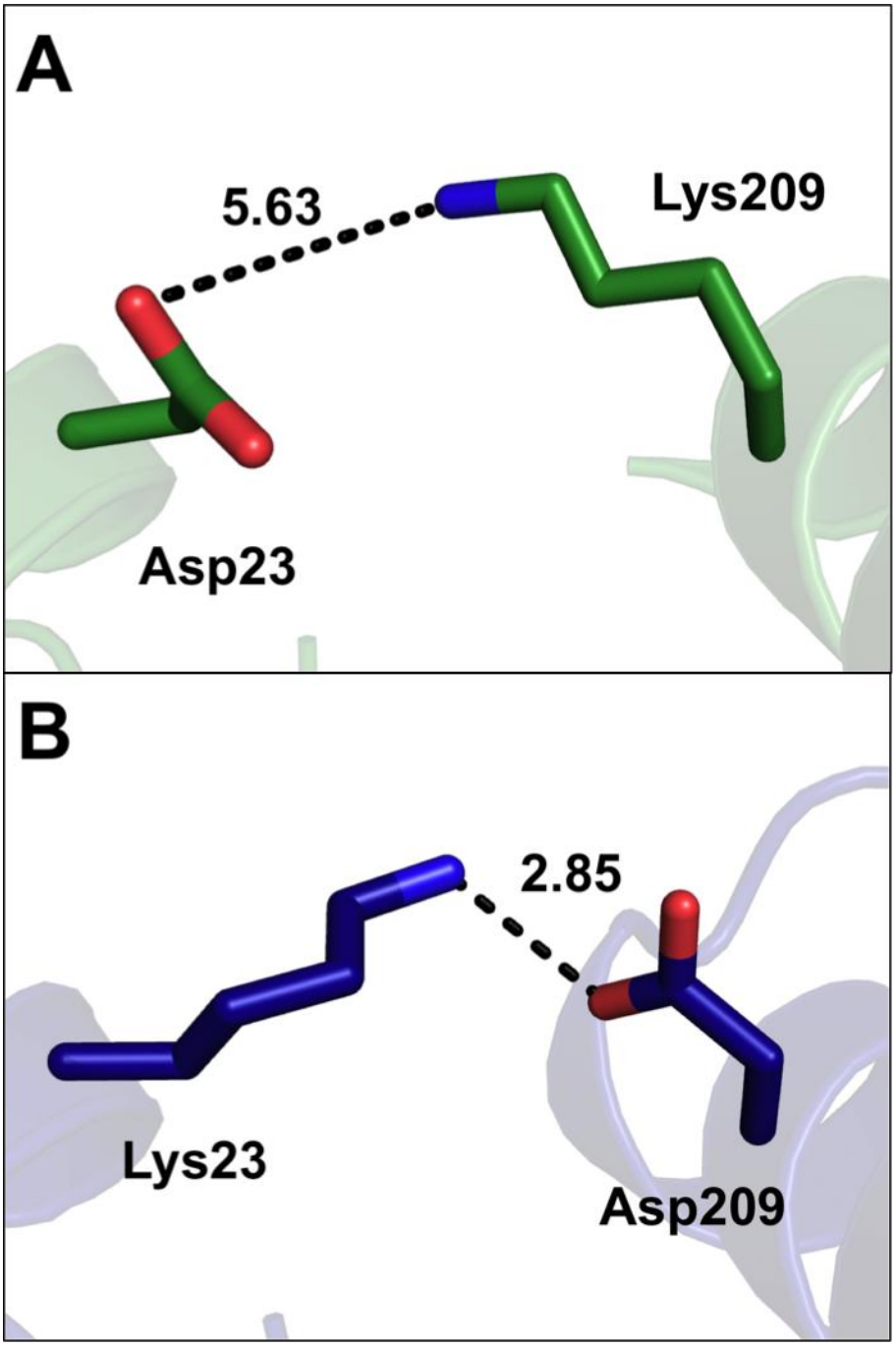

A
5.63
Lys209
Asp23
B
2.85
Lys23
Asp209

Figure 9. Sequence alignment of AK variants

Sequences of wild-type AKs and three representative AK variants: AKv3, AKv8, and AKv18 are represented. Positions of stabilizing mutations identified by comparative study, molecular evolution, and LSE optimization are marked with red, black, and blue triangles, respectively. The site of the Lys209Asp mutation necessary for the formation of an ion pair after LSE optimization is also indicated by a yellow triangle. Secondary structure elements are indicated based on the structure of AKv18.

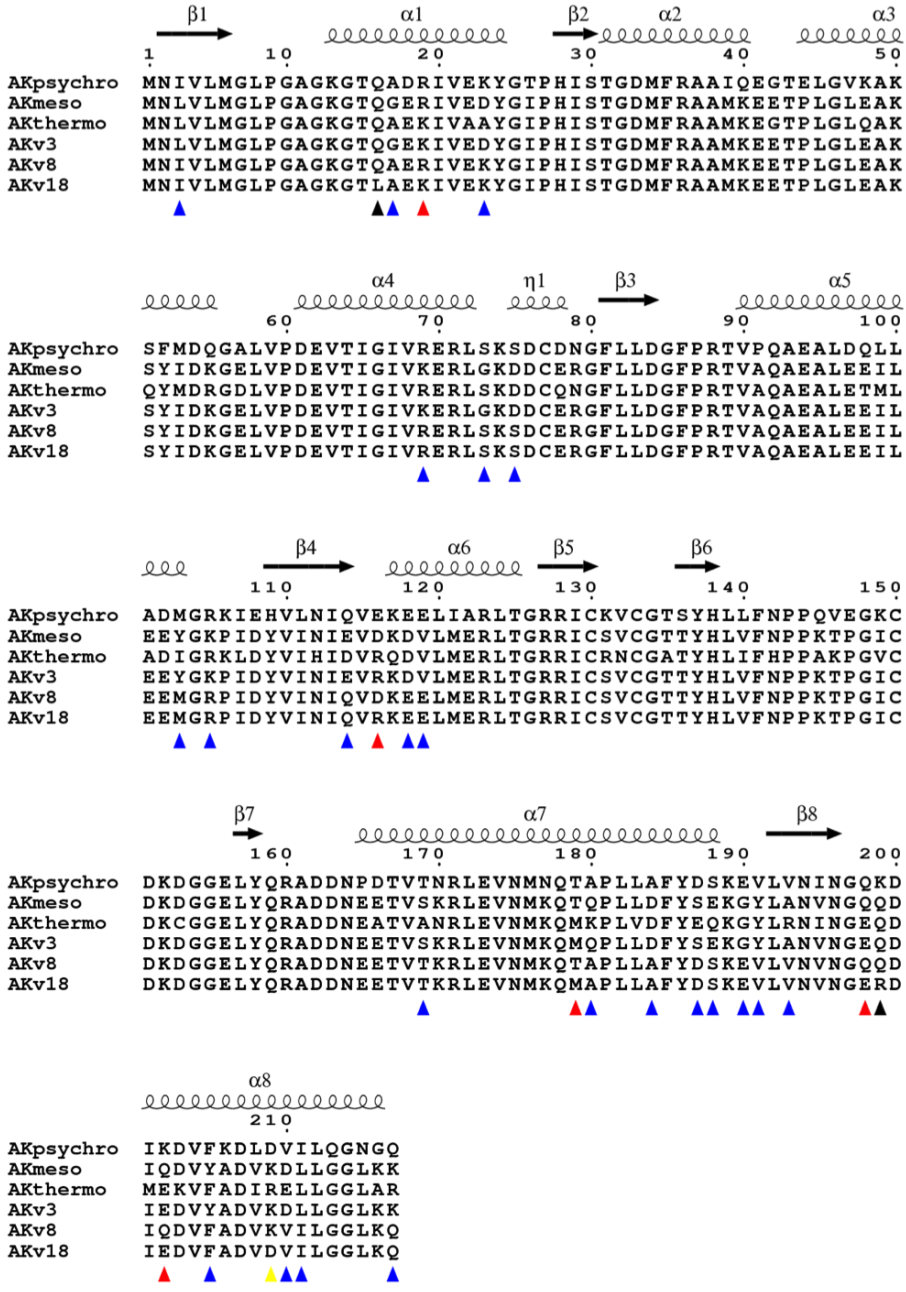

β1 α1 β2 α2 α3
1 10 20 30 40 50
AKpsychro MNIVLMGLPGAGKGTQADRIVEKYGTPHISTGDMFRAAIQEGTELGVKAK
AKmeso MNLVLMGLPGAGKGTQGERIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKthermo MNLVLMGLPGAGKGTQAEKIVAAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKv3 MNLVLMGLPGAGKGTQGEKIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKv8 MNIVLMGLPGAGKGTQAERIVEKYGIPHISTGDMFRAAMKEETPLGLEAK
AKv18 MNIVLMGLPGAGKGTLAEKIVEKYGIPHISTGDMFRAAMKEETPLGLEAK
α4 η1 β3 α5
60 70 80 90 100
AKpsychro SFMDQGALVPDEVTIGIVRERLSKSDCDNGFLLDGFPRTVPQAEALDQLL
AKmeso SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKthermo QYMDRGDLVPDEVTIGIVRERLSKDDCQNGFLLDGFPRTVAQAEALETML
AKv3 SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKv8 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
AKv18 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
β4 α6 β5 β6
110 120 130 140 150
AKpsychro ADMGRKIEHVLNIQVEKEELIARLTGRRICKVCGTSYHLLFNPPQVEGKC
AKmeso EEYGKPIDYVINIEVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKthermo ADIGRKLDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKv3 EEYGKPIDYVINIEVRKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKv8 EEMGRPIDYVINIQVDKEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKv18 EEMGRPIDYVINIQVRKEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
β7 α7 β8
160 170 180 190 200
AKpsychro DKDGGELYQRADDNPDTVTNRLEVNMNQTAPLLAFYDSKEVLVNINGQKD
AKmeso DKDGGELYQRADDNEETVSKRLEVNMKQTQPLLDFYSEKGYLANVNGQQD
AKthermo DKCGGELYQRADDNEATVANRLEVNMKQMKPLVDFYEQKGYLRNINGEQD
AKv3 DKDGGELYQRADDNEETVSKRLEVNMKQMQPLLDFYSEKGYLANVNGEQD
AKv8 DKDGGELYQRADDNEETVTKRLEVNMKQTAPLLAFYDSKEVLVNVNGQQD
AKv18 DKDGGELYQRADDNEETVTKRLEVNMKQMAPLLAFYDSKEVLVNVNGERD
α8
210
AKpsychro IKDVFKDLDVILQGNGQ
AKmeso IQDVYADVKDLLGGLKK
AKthermo MEKVFADIRELLGGLAR
AKv3 IEDVYADVKDLLGGLKK
AKv8 IQDVFADVKVILGGLKQ
AKv18 IEDVFADVDVILGGLKQ

Figure 10. Thermal denaturation curves of AKmeso and AK variants

T_m values of AKmeso and AKmeso variants were measured by circular dichroism at increasing temperaures.

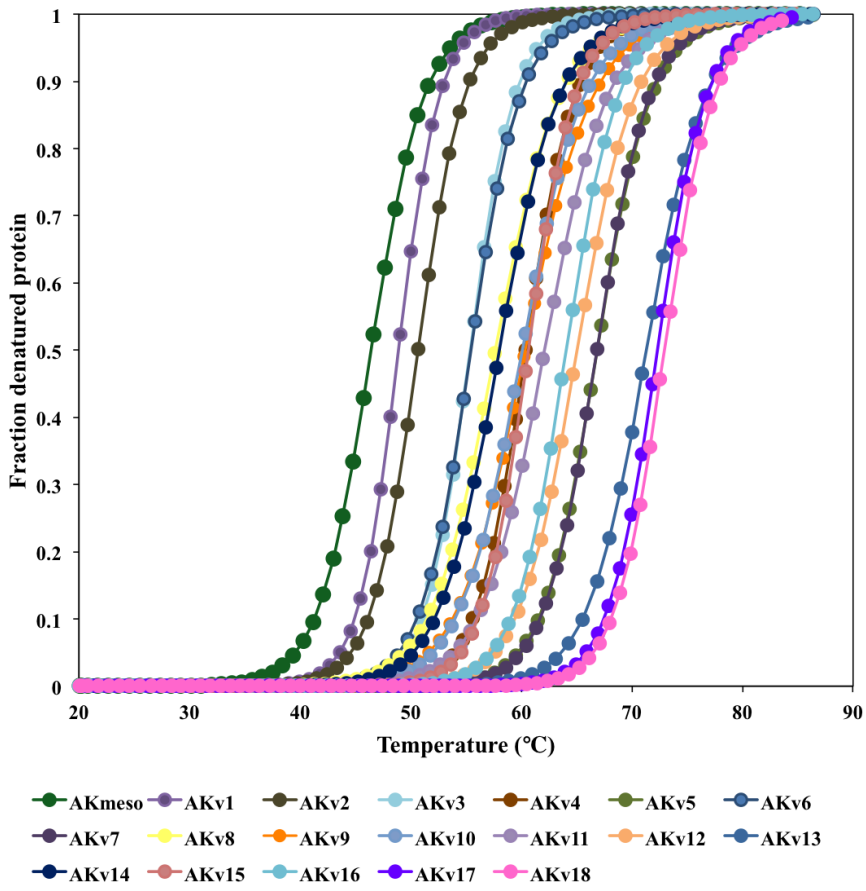

Fraction denatured protein
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
20
30
40
50
60
70
80
90
Temperature (°C)
AKmeso
AKv1
AKv2
AKv3
AKv4
AKv5
AKv6
AKv7
AKv8
AKv9
AKv10
AKv11
AKv12
AKv13
AKv14
AKv15
AKv16
AKv17
AKv18

Table 5. Mutations and T_m values of AK variants

AK	Comparative study			Molecular evolution		LSE optimization	Ion pair recovery	T_m (°C)	ΔT_m (°C)[b]
	T179M	R19K-Q202E	D116R-Q198E	Q16L	Q199R	23 mutations[a]	K209D		
AKpsychro	N/A	N/A	N/A	N/A	N/A	N/A	N/A	42.7 ± 0.1	- 3.7
AKmeso	N/A	N/A	N/A	N/A	N/A	N/A	N/A	46.4 ± 0.4	0.0
AKv1	+[c]	–[d]	–	–	–	–	–	48.9 ± 0.2	2.5
AKv2	–	+	+	–	–	–	–	50.7 ± 0.1	4.3
AKv3	+	+	+	–	–	–	–	55.4 ± 0.1	9.0
AKv4	–	–	–	+	+	–	–	60.4 ± 0.1	14.0
AKv5	+	+	+	+	–	–	–	66.8 ± 0.2	20.4
AKv6	+	+	+	–	+	–	–	55.5 ± 0.1	9.1
AKv7	+	+	+	+	+	–	–	66.8 ± 0.2	20.4
AKv8	–	–	–	–	–	+	–	57.7 ± 0.2	11.3
AKv9	+	–	–	–	–	+	–	60.3 ± 0.3	13.9
AKv10	–	+	+	–	–	+	–	60.1 ± 0.2	13.7
AKv11	+	+	+	–	–	+	–	62.1 ± 0.2	15.7
AKv12	–	–	–	+	+	+	–	65.1 ± 0.2	18.7
AKv13	+	+	+	+	–	+	–	71.3 ± 0.3	24.9
AKv14	–	–	–	–	–	+	+	58.0 ± 0.2	11.6
AKv15	–	+	+	–	–	+	+	60.6 ± 0.1	14.2
AKv16	+	+	+	–	–	+	+	64.1 ± 0.1	17.7
AKv17	+	+	+	+	–	+	+	72.3 ± 0.2	25.9
AKv18	+	+	+	+	+	+	+	73.0 ± 0.2	26.6

[a]23 LSE mutations: L3I, G17A, D23K, K69R, G73S, D75S, Y103M, K105R, E114Q, D118E, V119E, S169T, Q180A, D184A, S187D, E188S, G190E, Y191V, A193V, Y205F, D210V, L211I, K217Q
[b]Difference to T_m of AKmeso
[c]Mutation introduced
[d]Mutation not introduced

Furthermore, AKv3 which contained both Thr179Met and substitutions of ion pairs, showed an increase in T_m of 9.0°C relative to AKmeso. This value was greater than the sum of T_m increases conferred by the hydrophobic substitution or the ion pair introduction alone, indicating a synergistic effect on the thermal stability of AKmeso.

The two mutations discovered in the molecular evolution (Gln16Leu and Gln199Arg) showed the most profound effects on the thermal stability of AKmeso. Combined, these mutations led to a 14.0°C increase in T_m of the AKv4 variant. Even though there were not examined individually, the Gln16Leu mutation provided greater contribution to protein stability than the Gln199Arg mutation. The introduction of individual Gln16Leu or Gln199Arg mutations into AKv3 (resulting in variants AKv5 and AKv6, respectively) led to a more substantial change in T_m for AKv5 compared with AKv6.
However, AKv7 that contained the hydrophobic substitution, ion pair substitutions, and the two mutations in the molecular evolution, appeared to not a synergistic effect compared with AKv5.

Meanwhile, AKv8, which contains all 23 of the LSE mutations, displayed a considerable increase (11.3°C) in T_m relative to AKmeso. However, one of the LSE substitutions (Asp23Lys) resulted in a loss of electrostatic attraction connecting distant regions of the polypeptide, and subsequent addition of the Lys209Asp mutation restored this interaction. Despite its predicted effect, the Lys209Asp mutation showed only negligible increase (0.3°C) in thermal stability of AKv14 though more significant results were observed in other backgrounds.

For example, the AKv11 and AKv16 variants are identical except for the Lys209Asp mutation, which conferred a 2.0°C increase in the stability of AKv16. The most stable AK variant, AKv18, displayed higher enzymatic activity at elevated temperatures (65–85°C) relative to AKmeso, indicating the retention of catalytic function after the collective stabilizing mutations (Fig. 11).

Crystal structures of AKmeso variants

To verify the structural basis for the increased thermal stabilities of AK variants, I determined crystal structures of two representative AK variants, AKv8, and AKv18, to resolutions of 1.45, and 1.50 Å, respectively (Fig. 12). The crystal structure of AKv3 was determined to 1.70 Å resolution by Du-kyo Jung. Data collection and refinement statistics are summarized in Table 6.

The asymmetric unit of the AKv3 structure contained two chains while only one chain was found for the structures of AKv8 and AKv18. As expected, the chain folds of the AK variants were essentially identical to those of other AKs. The structures revealed the co-crystallized inhibitor Ap_5A in the active site of the proteins, along with the characteristic AK domain arrangement composed of the core (residues 1–30, 61–126, and 165–217), AMP binding (residues 31–60), and lid (residues 127–164) domains (Bae and Phillips, 2004). The RMSD values of Ca atoms between AKmeso and each of the three AK variants ranged from 0.6 to 1.9 Å.

Table 6. Data collection and refinement statistics[a] for structures of three representative AK variants

	AKv3	AKv8	AKv18
Space group	$P2_1$	$P2_12_12_1$	$P2_1$
Unit cell parameters (Å)	a=34.4, b=75.0, c=77.2, β=100.9°	a=44.6, b=61.3, c=88.0	a=44.2, b=55.2, c=46.4, β=110.6°
Wavelength (Å)	1.0000	0.9795	0.9793
Data collection statistics			
Resolution range (Å)	50.00-1.70 (1.76-1.70)	44.62-1.45 (1.53-1.45)	50.00-1.50 (1.55-1.50)
Number of reflections	42262 (4194)	43604 (6272)	33460 (3318)
Completeness (%)	99.9 (99.8)	100.0 (100.0)	99.7 (100.0)
R_{merge}[b]	0.072 (0.515)	0.062 (0.561)	0.064 (0.653)
Redundancy	7.4 (7.3)	12.7 (13.0)	3.7 (3.6)
Mean I/σ	26.4 (4.4)	23.9 (4.7)	18.9 (2.1)
Refinement statistics			
Resolution range (Å)	50.00-1.70	44.62-1.45	43.41-1.50
R_{cryst}[c]/R_{free}[d] (%)	15.9/20.3	17.0/20.0	17.4/21.8
RMSD bonds (Å)	0.021	0.025	0.027
RMSD angles (deg)	2.252	2.567	2.775
Average *B* factor (Å^2)	16.68	13.62	17.87
Number of water molecules	471	321	135
Ramachandran favored (%)	99.1	99.1	99.1
Ramachandran allowed (%)	0.9	0.9	0.9

[a]Values in parentheses are for the highest-resolution shell.

[b]$R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

[c]$R_{cryst} = \Sigma_h | |F_{obs}| - |F_{calc}| | / \Sigma_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d]R_{free} was calculated as R_{cryst} using 5% of the randomly selected unique reflections that were omitted from structure refinement.

Figure 11. Temperature dependence of enzymatic activity for AKmeso and the most stable AK variant, AKv18

Catalytic activity of AKs is determined at various temperatures in the direction of ATP formation.

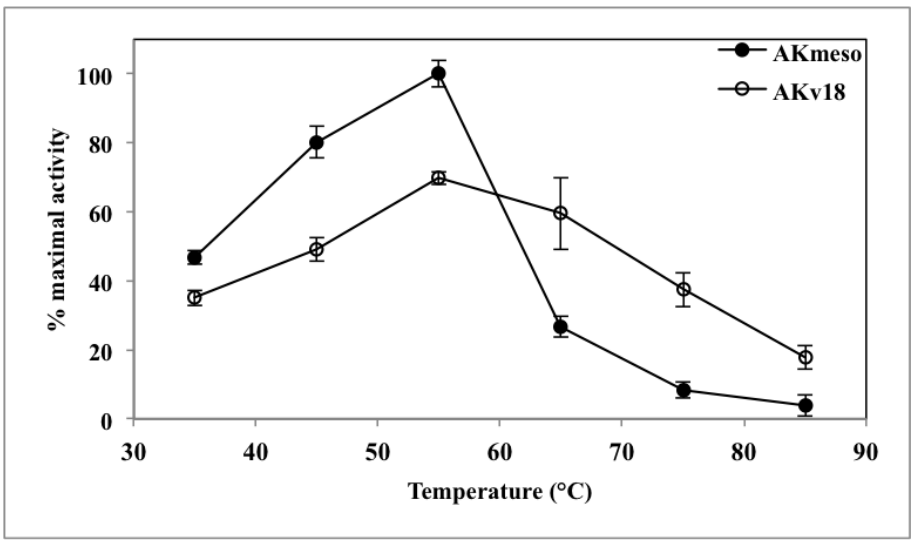
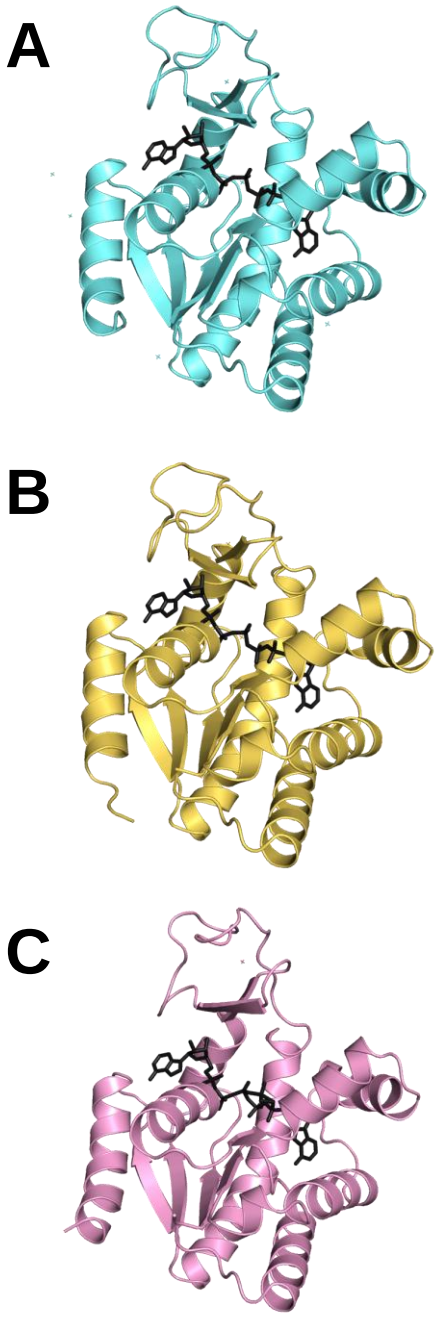

AKmeso
AKv18
% maximal activity
100
80
60
40
20
0
30
40
50
60
70
80
90
Temperature (°C)

Figure 12. Crystal structures of three representative AK variants

(A) AKv3, (B) AKv8, and (C) AKv18 indicate cyan, yellow, and pink, respectively. The bound inhibitor Ap_5A molecules are shown as black sticks. For AKv3, chain A is shown.


A
B
C

Structural analysis of AKv3

AKv3 contained the Thr179Met mutation introduced to improve hydrophobic interactions with the pocket formed by residues Met6, Gly7, and Leu8. Despite some heterogeneity observed between the two chains in the asymmetric unit, the crystal structure of AKv3 revealed an overall decreased solvent accessibility for the hydrophobic pocket caused by the Thr179Met substitution, suggesting stabilization of the protein's hydrophobic core by this mutation (Fig. 13 and Table 7).

In AKv3, oppositely charged residues (Lys19-Glu202 and Arg116-Glu198) were also introduced to facilitate favorable electrostatic interactions bridging distant regions of the polypeptide. In the crystal structure, each pair of oppositely charged atoms were found to be within short distances of each other indicating successful formation of the intended ion pairs (Table 8 and Fig. 14).

While the Arg116-Glu198 ion pairs were tightly maintained in both chains of the structure, one of the Lys19-Glu202 ion pairs displayed a relatively longer distance between oppositely charged atoms than the other. This difference was likely due to the presence of a bound calcium ion added as a component of the crystallization solution, indicating the difficulty of identifying ion pairs based solely upon the crystal structure of a protein.

Table 7. Effect of residue substitution at the position 179 on surface accessibility

AK	Amino acid at position 179	Accessible molecular surface area (Å^2)			
		Met6	Gly7	Leu8	Total
AKmeso	Thr	4.1	6.3	12.3	22.7
AKv3 (A/B)[a]	Met	3.2 (1.6/4.8)	3.6 (1.8/5.3)	4.3 (3.8/4.8)	11.1 (7.2/14.9)
AKv18	Met	5.1	4.9	5.5	15.5
AKthermo	Met	1.2	1.2	2.2	4.6

[a]Average of the two chains in asymmetric unit was reported with values in parentheses for chains A and B, respectively.

Figure 13. Hydrophobic packing around the pocket formed by residues Met6, Gly7, and Leu8 in AKv3

AKv3 contains the Thr179Met mutation to improve hydrophobic interactions. Residues Met6, Gly7, and Leu8 in chain A are shown.

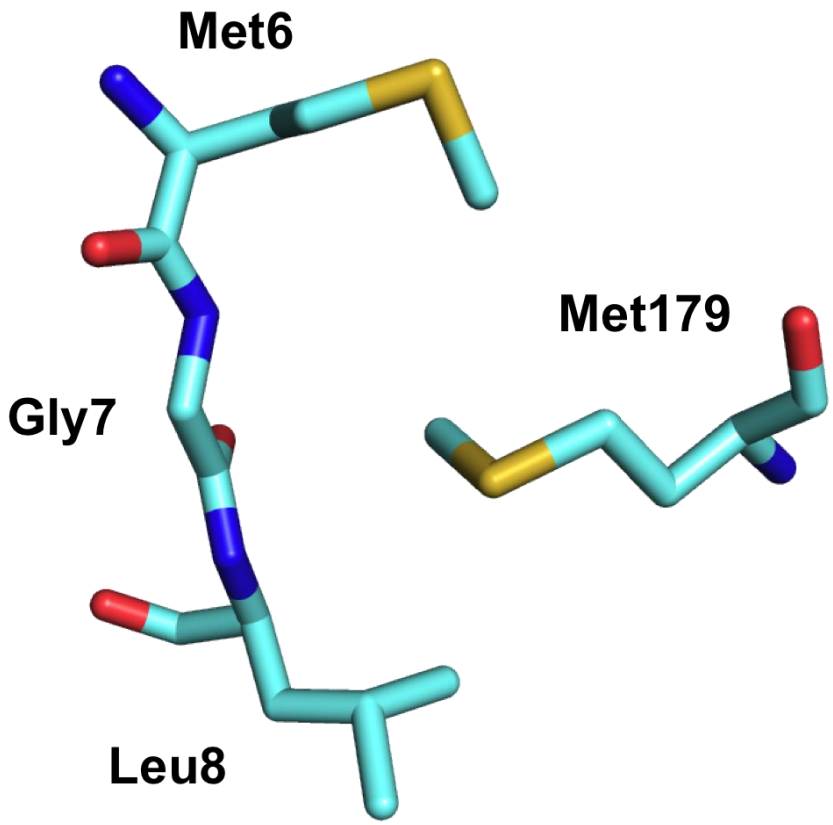

Met6
Met179
Gly7
Leu8

Figure 14. Ion pairs connecting distant regions of a polypeptide in AKv3

(A) Lys19-Glu202 and (B) Arg116-Glu198 are closely located in chain A of AKv3.

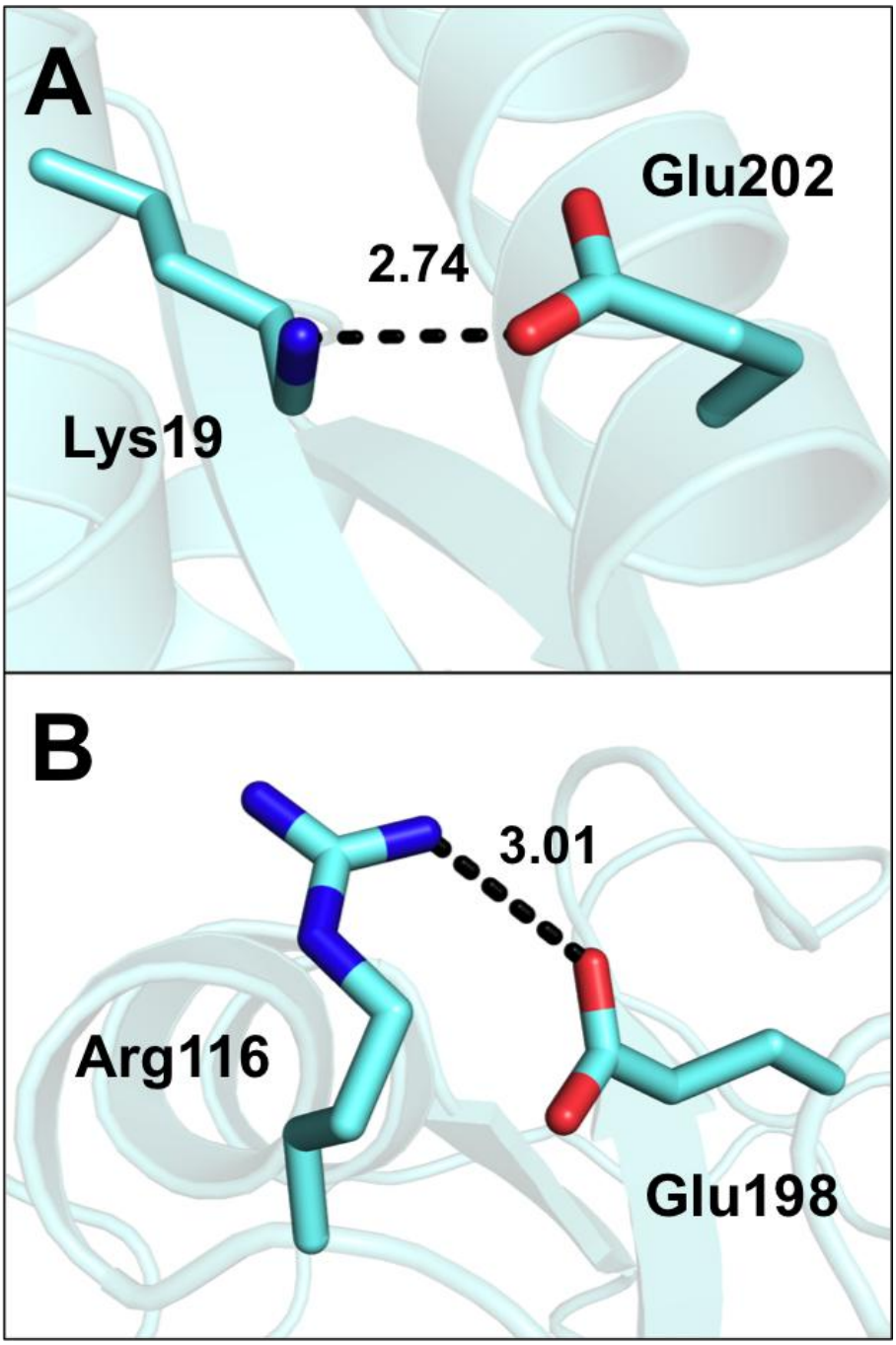

A
Glu202
2.74
Lys19
B
3.01
Arg116
Glu198

Table 8. Distances between oppositely charged residues of ion pairs

AK	Distance (Å)	
	Lys19-Glu202	Arg116-Glu198
AKv3 (A/B)[a]	3.82 (2.74/4.89)	2.98 (3.01/2.95)
AKv18	3.30	2.88
AKthermo	2.77	3.04

[a]Average of the two chains in asymmetric unit was reported with values in parentheses for chains A and B, respectively.

Structural analysis of AKv8

To identify possible structural mechanisms for thermal stabilization conferred by the LSE optimization, I compared the crystal structure of AKv8, the LSE-optimized variant, with that of AKmeso. Several structural features including the numbers of ion pairs, the number of hydrogen bonds, and the magnitude of buried and accessible molecular surface areas were calculated for the core domains (Table 9). In this comparison, it was difficult to identify specific structural features that might be responsible for the stability difference between AKv8 and AKmeso.

AKv8 had more ion pairs than AKmeso with a distance cutoff of 6.0 Å but contained fewer pairs when the cutoff distance was reduced to 4.0 Å. When considering only the ion pairs connecting distant regions of the protein (more than 10 residues), AKv8 possessed a smaller number of ion pairs than AKmeso for both distance cutoff values. It was also difficult to determine whether hydrogen bonds play an important role in the thermal stability as AKv8 contained only one more hydrogen bond than AKmeso (Table 9), which is less than 1% of its total hydrogen bonds. Compared to AKmeso, AKv8 had more apolar buried molecular surface area (Table 9), which is often used as a measure of hydrophobic interaction.

However, the difference (41.8 $Å^2$) was too small to establish apolar buried surface area as a major determinant of thermal stability of the AKs. In the crystal structure of AKv3 whose asymmetric unit contained two independent polypeptide chains, the difference in the apolar buried molecular surface area between the two chains amounts to 33.9 $Å^2$ (Table 9).

Table 9. Structural features[a] of wild-type AKs and AK variants

	AKpsychro	AKmeso	AKv3 (A/B)[b]	AKv8	AKv18	AKthermo
T_m (°C)	42.7	46.4	55.4	57.7	73.0	74.5[c]
Number of ion pairs (<4.0 Å)	6	5	7 (7/7)	3	6	8
Number of ion pairs connecting distant regions (<4.0 Å)[d]	3	3	3.5 (4/3)	1	4	4
Number of ion pairs (<6.0 Å)	14	10	17.5 (17/18)	12	18	14
Number of ion pairs connecting distant regions (<6.0 Å)[d]	6	5	8.5 (8/9)	4	6	6
Number of hydrogen bonds	154	147	145.5 (144/147)	148	151	157
Number of apolar atoms	707	717	721	717	721	718
Aplor buried molecular surface area (Å^2)	3227.5	3293.6	3378.3 (3395.2/3361.3)	3335.4	3372.6	3419.7
Apolar accessible molecular surface area (Å^2)	1326.3	1278.4	1253.0 (1234.6/1271.3)	1294.5	1339.5	1237.9
Number of polar atoms	410	414	412	412	410	411
Polar buried molecular surface area (Å^2)	1654.5	1645.7	1629.9 (1628.2/1631.5)	1613.7	1619.6	1675
Polar accessible molecular surface area (Å^2)	1003.7	1005.8	1046.8 (1067.8/1025.7)	1056.3	1023.6	1014.7

[a]Calculated for core domain (residues 1-30, 61-126 and 165-212)

[b]Average of the two chains in asymmetric unit was reported with values in parentheses for chains A and B, respectively.

[c]From Glaser et al., 1992.

[d]Ion pairs connecting distant regions of a polypeptide (more than ten residues)

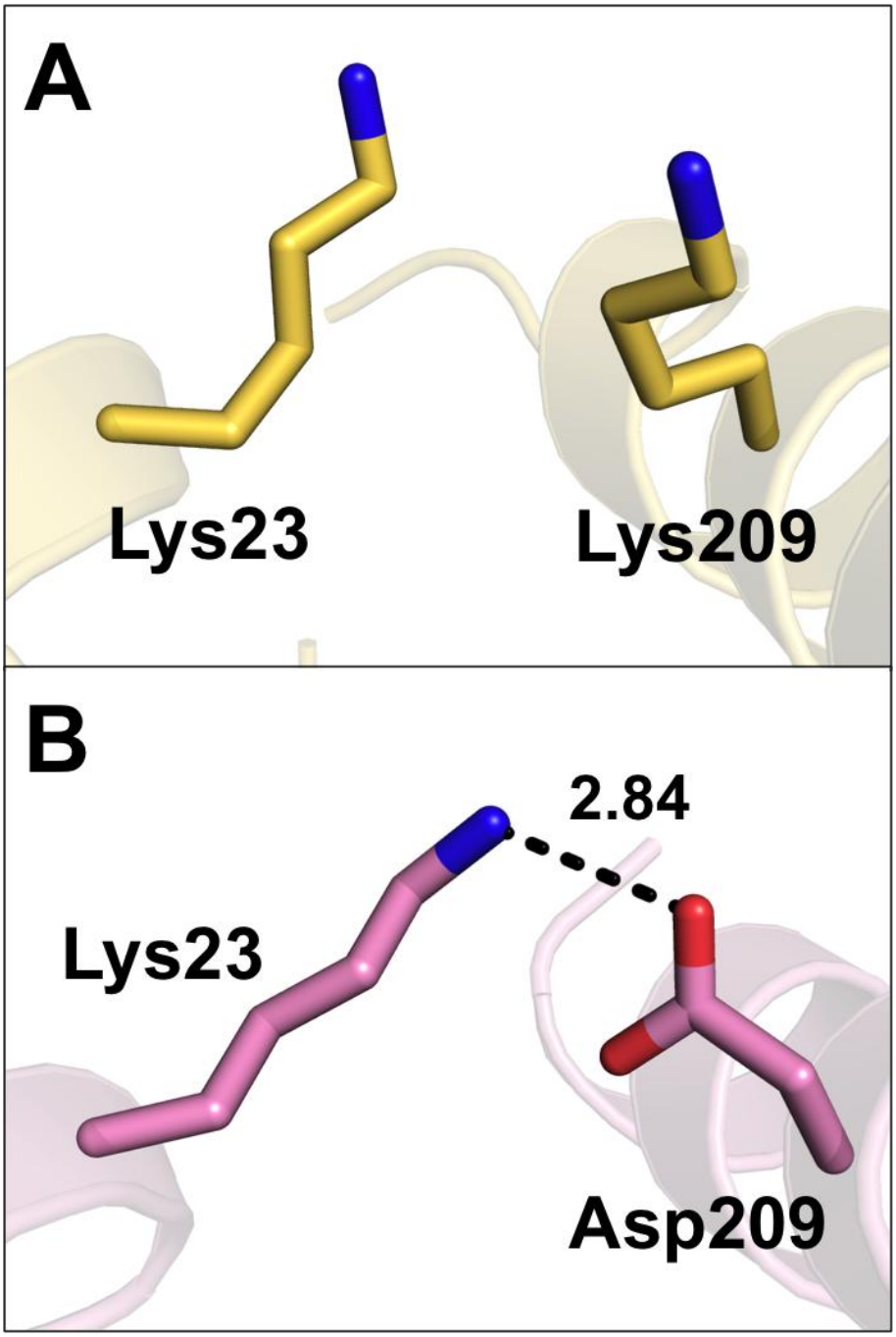
Moreover, AKv3 had more apolar buried surface area than AKv8 despite its lower thermal stability, highlighting the difficulty in correlating the total apolar buried surface area with a protein's thermal stability. A "residue-by-residue" comparison was also made for the 23 LSE substitutions in AKv8 to identify the formation of stabilizing interactions caused by the mutations.

In AKv8, the two ion pairs (Arg69-Asp70 and Lys117-Glu118) were involved at least one of the 23 mutated residues, and these ion pairs are not present in AKmeso. However, it was unlikely that the two ion pairs contribute significantly to the thermal stability because they occur between nearby residues, and the involved mutations (Lys69Arg and Asp118Glu) did not alter the original charge. Rather, the LSE substitutions removed the Asp23-Lys209 ion pair, which connects distant regions of the polypeptide in AKmeso (Fig. 15).

Taken together, it was unclear which structural factors increased the thermal stability of AKv8. However, this increased stability was unlikely to be the result of unexpected stabilizing interactions between distant regions of the polypeptide.

Figure 15. Loss and recovery of electrostatic attraction between residues 23 and 209

Residues 23 and 209 are shown in (A) AKv8 and (B) AKv18. LSE substitutions remove Asp23-Lys209 ion pair in AKmeso. In AKv18, ion pair between residues 23 and 209 is recovered by Lys209Asp mutation.


A
Lys23
Lys209
B
2.84
Lys23
Asp209

Structural analysis of AKv18

In addition, I also solved the crystal structure of AKv18, the most stable of the 18 AK variants and the variant containing the greatest number of mutations compared with AKmeso. The mutated residues identified by the comparative study and molecular evolution, appear to participate in favorable electrostatic and hydrophobic interactions.

Residues from the two ion pairs (Lys19-Glu202 and Arg116-Glu198) remained in close proximity (Table 8 and Fig. 16) while the two molecular evolution mutations (Gln16Leu and Gln199Arg) contributed to improved hydrophobic packing and the formation of an electrostatic network, respectively (Fig. 17).

The Thr179Met mutation continued to exhibit improved hydrophobic contact with a pocket made by Met6, Gly7, and Leu8 (Table 7). However, the side chain of Met179 was located farther from the pocket in AKv18 (Fig. 18), resulting in the exposure of residues Met6 and Gly7 compared with the AKv3 structure.

This difference is likely an artifact of the cryoprotection used in X-ray data collection since an ethylene glycol molecule added as a cryoprotecting reagent was found near the hydrophobic pocket and appears to displace the Met179 residue. It was also difficult to identify the structural basis for the LSE stabilization in the AKv18 structure. Neither the comparison of structural features nor the “residue-by-residue” inspections were able to provide any straightforward answers, similar to what was seen for the AKv8 variant.

At minimum, I was able to confirm that thermal stabilization by the LSE optimization was not the result of unexpected electrostatic or hydrophobic interactions. The Lys209Asp mutation introduced in this variant was successful in restoring the ion pair destroyed by the LSE substitutions. This newly reconstructed the ion pair, Lys23-Asp209, was tightly maintained in the structure of AKv18, with a distance between the oppositely charged atoms of 2.8 Å (Fig. 15).

Figure 16. Ion pairs connecting distant regions of a polypeptide in AKv18

AKv18 has additional two ion pairs: (A) Lys19-Glu202 and (B) Arg116-Glu198.

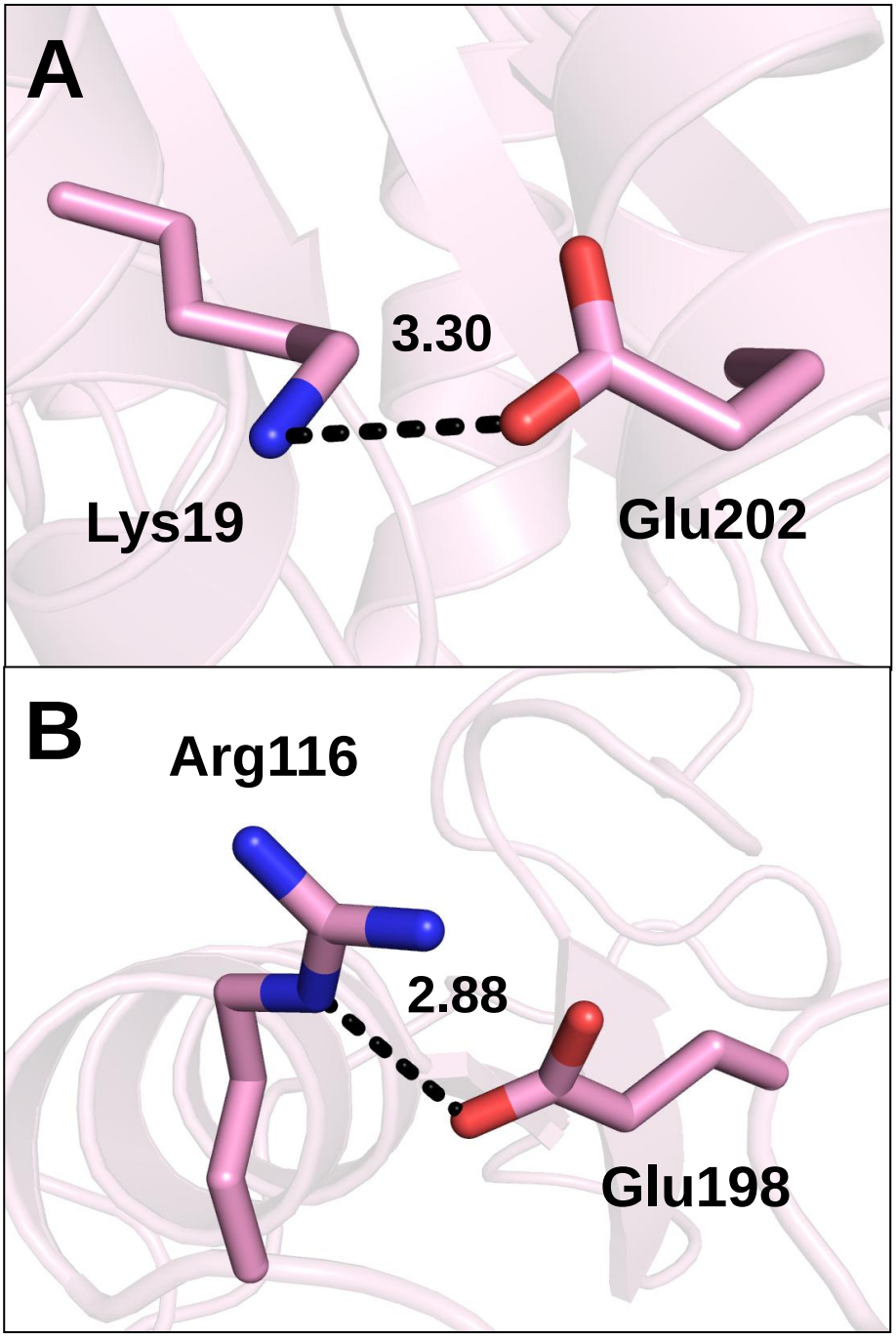

A
3.30
Lys19
Glu202
B
Arg116
2.88
Glu198

Figure 17. Effect of mutations identified by molecular evolution in AKv18

Hydrophobic packing and electrostatic interactions are improved by (A) Gln16Leu and (B) Gln199Arg mutations.

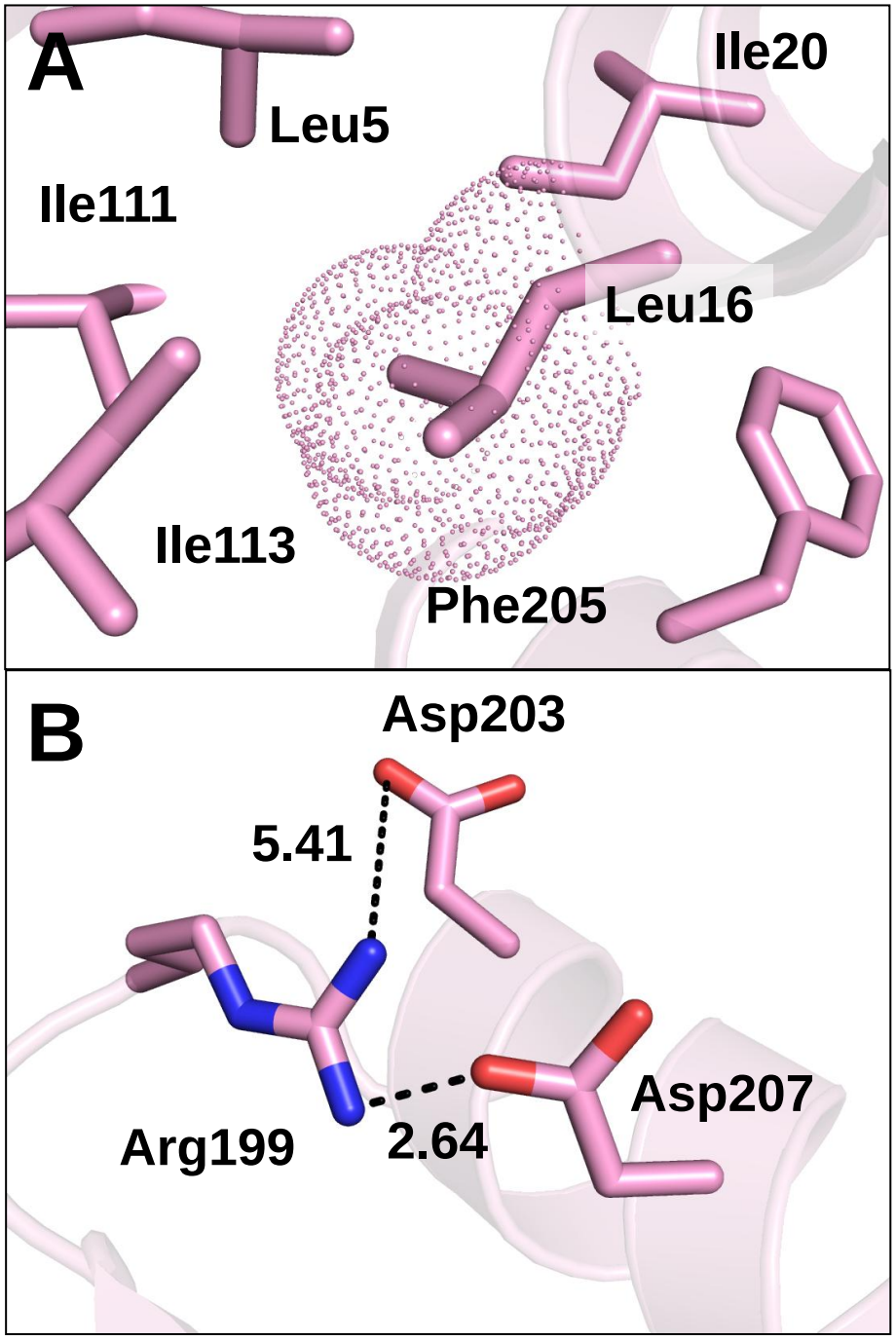
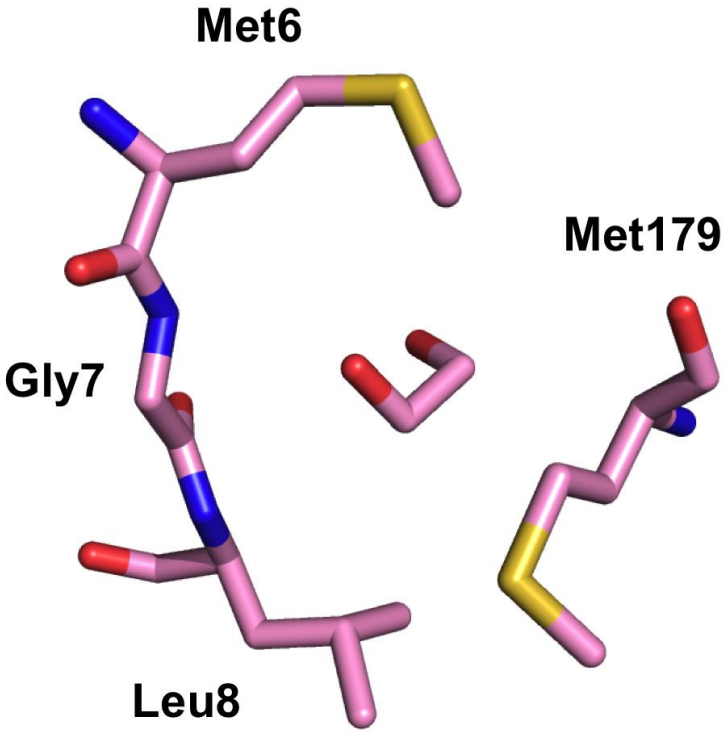

A
Leu5
Ile20
Ile111
Leu16
Ile113
Phe205
B
Asp203
5.41
Asp207
Arg199
2.64

Figure 18. Hydrophobic packing around the pocket formed by residues Met6, Gly7, and Leu8 of AKv18

AKv18 contains the Thr179Met mutation to improve hydrophobic interactions. An ethylene glycol molecule as a cryoprotecting reagent is represented in AKv18.


Met6
Met179
Gly7
Leu8

Part II

Effectiveness and Limitation of LSE Optimization for Thermal Stabilization of AKs

In the previous study, LSE was incorporated as part of a protein stabilization method applied to AKmeso. I demonstrated that thermal stabilization by LSE optimization was comparable with that obtained using comparative study and experimental evolution. In this study, I designed additional LSE-optimized AK variants and evaluated their thermal stabilities to understand effectiveness of LSE in protein stability and its use for thermal stabilization. Crystal structures of LSE-optimized AK variants were determined to identify the structural bases of thermal stabilization by LSE optimization. The results illustrated the effectiveness and the limitation of LSE optimization as a thermal stabilization approach in proteins.

Thermal stabilization by LSE optimization was effective but limited in mesophilic AK

The LSE optimization of AKmeso was performed previously by substituting several of its amino acid residues with those from AKpsychro, resulting in three LSE-optimized AK variants (Fig. 7). Another LSE-optimized variant, AKlse4, was designed using AKmeso as a template (Fig. 19 and Table 10) by Thomas J. Rutkoski.

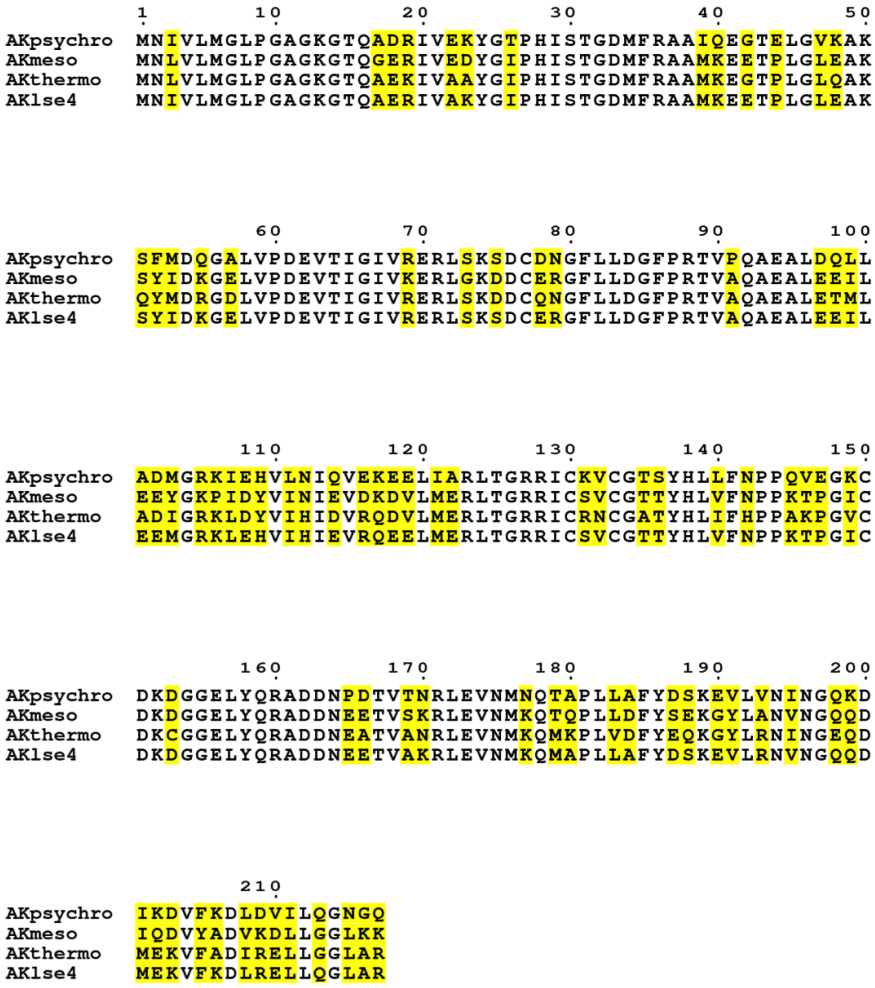
Table 10. Design strategy and properties of wild-type AKs and four LSE-optimized AK variants

	AKmeso	AKlse1[a]	AKlse2[a]	AKlse3[a]	AKlse4[b]	AKthermo
Template	-	AKmeso	AKmeso	AKmeso	AKmeso	-
Substitutions from	-	AKpsychro	AKpsychro	AKpsychro	AKpsychro , AKthermo	-
Additional design constraint/ strategy	-	None	The most substitutions in the top 20 most optimized variants	No 'AKthermo-like' substitutions	None	-
Number of substitutions	-	23	26	10	38	-
Average LSE[c]	1.4938	1.4083	1.4085	1.4603	1.3696	1.4323
T_m (°C)	46.4	57.7	56.2	50.1	56.6	74.5[d]
ΔT_m (°C)[e]	-	11.3	9.8	3.7	10.2	-

[a]Designed previously in Bae et al., 2008
[b]Designed by Thomas J. Rutkoski
[c]Calculated for core domain (residues 1-30, 61-126, and 165-212)
[d]From Glaser et al., 1992
[e]Difference from T_m of AKmeso

Figure 19. Sequence alignment of wild-type AKs and AKlse4

The amino acid sequences of three wild-type AKs (AKpsychro, AKmeso, and, AKthermo) and another LSE-optimized AK variant (AKlse4) are aligned together. Variable residues are highlighted in yellow.


1 10 20 30 40 50
AKpsychro MNIVLMGLPGAGKGTQADRIVEKYGTPHISTGDMFRAAIQEGTELGVKAK
AKmeso MNLVLMGLPGAGKGTQGERIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKthermo MNLVLMGLPGAGKGTQAEKIVAAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKlse4 MNIVLMGLPGAGKGTQAERIVAKYGIPHISTGDMFRAAMKEETPLGLEAK
60 70 80 90 100
AKpsychro SFMDQGALVPDEVTIGIVRERLSKSDCDNGFLLDGFPRTVPQAEALDQLL
AKmeso SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKthermo QYMDRGDLVPDEVTIGIVRERLSKDDCQNGFLLDGFPRTVAQAEALETML
AKlse4 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
110 120 130 140 150
AKpsychro ADMGRKIEHVLNIQVEKEELIARLTGRRICKVCGTSYHLLFNPPQVEGKC
AKmeso EEYGKPIDYVINIEVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKthermo ADIGRKLDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKlse4 EEMGRKLEHVIHIEVRQEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
160 170 180 190 200
AKpsychro DKDGGELYQRADDNPDTVTNRLEVNMNQTAPLLAFYDSKEVLVNINGQKD
AKmeso DKDGGELYQRADDNEETVSKRLEVNMKQTQPLLDFYSEKGYLANVNGQQD
AKthermo DKCGGELYQRADDNEATVANRLEVNMKQMKPLVDFYEQKGYLRNINGEQD
AKlse4 DKDGGELYQRADDNEETVAKRLEVNMKQMAPLLAFYDSKEVLRNVNGQQD
210
AKpsychro IKDVFKDLDVILQGNGQ
AKmeso IQDVYADVKDLLGGLKK
AKthermo MEKVFADIRELLGGLAR
AKlse4 MEKVFKDLRELLQGLAR

To more significantly reduce the LSE of AKmeso, residue substitutions were allowed from both AKpsychro and AKthermo. This resulted in a much larger search space ($2^{44} \cdot 3^{19}$ vs. 2^{53}) in which to identify an optimal sequence.

AKmeso contains a total of 63 residues in the core domain that differ from those of either AKpsychro or AKthermo, including 19 positions where amino acids are different in all three of the wild-type AKs. Consequently, AKlse4, the most LSE-optimized variant generated in this process, contained more residue substitutions and a lower average LSE value than those of the three previously designed variants (Table 10).

Thermal stabilization was evaluated experimentally for AKlse2, AKlse3 and AKlse4 using CD spectroscopy (Fig. 20). T_m values of the four LSE-optimized AK variants (AKlse1 to AKlse4) and their template, AKmeso, which were obtained using the same instrumentation and experimental conditions, are listed in Table 10. As expected, AKlse2, AKlse3, and AKlse4 were more thermally stable than AKmeso, with T_m values 9.8°C, 3.7°C, and 10.2°C greater than that of AKmeso, respectively.

A strong correlation was observed between the four LSE optimized AK variants and AKmeso, as indicated by a plot of T_m as a function of average LSE (Fig. 21). While AKlse4 exhibited a higher T_m value than AKmeso, the extent of stabilization was not as significant as expected based on its LSE value (Table 10). Despite its considerably lower LSE value, which implies greater thermal stabilization, AKlse4 exhibited a smaller T_m increase (10.2°C) than AKlse1 (11.3°C), weakening the overall correlation. As its LSE was reduced, the thermal stability of AKmeso seemed to increase only to a certain degree, suggesting that thermal stabilization by LSE optimization was limited.

Structural analysis revealed a strong correlation between LSE and apolar buried surface area

The crystal structure of AKlse1 was previously compared with that of AKmeso, but no clear structural mechanism of LSE optimization (Table 9). In order to identify that the increased thermal stability of AKmeso by LSE optimization, the crystal structures of three additional LSE-optimized AK variants (AKlse2, AKlse3, and AKlse4) were determined at 1.80, 1.53, and 1.37 Å, respectively (Fig. 22). Crystal structure of AKlse3 was determined by Ryan M. Bannen. Data collection and refinement statistics are summarized in Table 11.

The asymmetric unit of AKlse3 contained two chains while only one chain was found in the structures of AKlse2 and AKlse4. The three variants were crystallized with the inhibitor Ap_5A in their active sites as AKmeso and AKlse1. The chain folds of the four variants and AKmeso are essentially identical (Figs. 2 and 22). The RMSD values of Cα atom positions between AKmeso and each of the four LSE-optimized variants range from 0.6 to 2.1 Å.

To determine the molecular origins of the different thermal stabilities of the four LSE-optimized variants and AKmeso, their crystal structures were compared with focus on the core domain where residue substitutions for LSE optimization were allowed. Several structural features of their core domains were calculated and were listed in Table 12, including the number of ion pairs, the number of hydrogen bonds, and the buried and accessible molecular surface areas. Among the five AKs, AKlse4, its LSE was reduced to the largest extent, contains the most ion pairs and hydrogen bonds. This variant also had the highest apolar buried surface area, which is often used as a measure of the degree of hydrophobic interactions.

Figure 20. Thermal denaturation curves of LSE-optimized AK variants

T_m values of LSE-optimized AK variants were measured by circular dichroism at increasing temperaures.

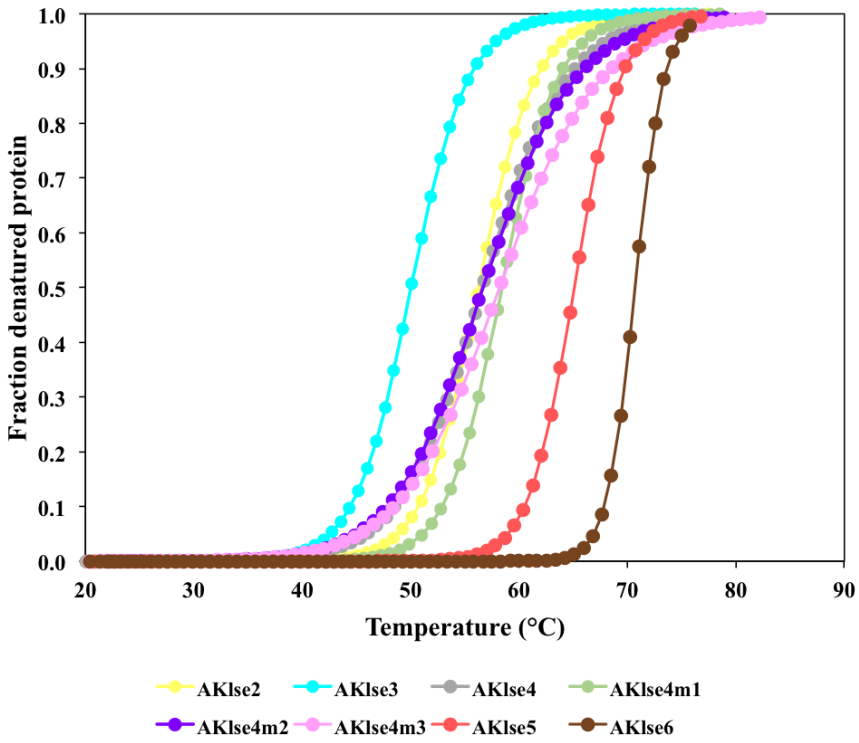

Fraction denatured protein
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
20
30
40
50
60
70
80
90
Temperature (°C)
AKlse2
AKlse3
AKlse4
AKlse4m1
AKlse4m2
AKlse4m3
AKlse5
AKlse6

Figure 21. A plot of T_m versus average LSE for AKmeso and LSE-optimized AK variants

Data for the five AKs are represented by different colors: AKmeso (black), AKlse1 (red), AKlse2 (yellow), AKlse3 (blue), AKlse4 (green). Trend lines and R^2 values are also shown.

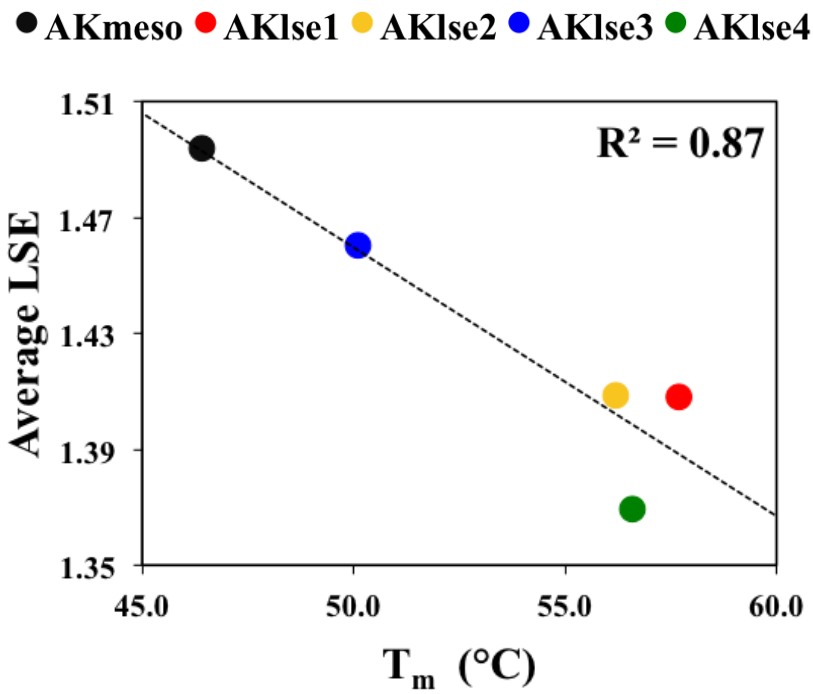

AKmeso
AKlse1
AKlse2
AKlse3
AKlse4
R² = 0.87
Average LSE
1.51
1.47
1.43
1.39
1.35
45.0
50.0
55.0
60.0
T m (°C)

Figure 22. Overall structures of four LSE-optimized AK variants

Crystal structures of four LSE-optimized AK variants: AKlse1, AKlse2, AKlse3, and AKlse4 indicate forest, limon, teal, and green, respectively. The bound inhibitor Ap_5A molecules are shown as stick model. PDB entry: 3DL0 (AKlse3).

AKIse1

AKIse2

AKIse3

AKIse4

Table 11. Data collection and refinement statistics[a] for structures of AKlse2 and AKlse4

	AKlse2	AKlse4
Space group	$P2_12_12_1$	$C2$
Unit cell parameters (Å)	a=39.3, b=47.2, c=109.3	a=68.8, b=71.0, c=45.6, β=95.1°
Wavelength (Å)	0.9795	0.9795
Data collection statistics		
Resolution range (Å)	50.00-1.80 (1.86-1.80)	50.00-1.37 (1.42-1.37)
Number of reflections (measured/unique)	135728/19418	165077/44220
Completeness (%)	99.9 (100.0)	97.0 (95.1)
R_{merge}[b]	0.107 (0.752)	0.047 (0.720)
Redundancy	7.0 (7.1)	3.7 (3.6)
Mean I/σ	15.6 (4.0)	14.5 (2.4)
Refinement statistics		
Resolution range (Å)	50.00-1.80	49.32-1.37
R_{cryst}[c]/R_{free}[d] (%)	17.1/22.2	17.9/22.0
RMSD bonds (Å)	0.019	0.026
RMSD angles (deg)	2.332	2.51
Average B factor (Å^2)	20.02	17.91
Number of water molecules	144	232
Ramachandran favored (%)	99.1	100
Ramachandran allowed (%)	0.9	0

[a]Values in parentheses are for the highest-resolution shell.
[b]$R_{merge} = \Sigma_h\Sigma|I_i(h) - \langle I(h)\rangle| / \Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection.
[c]$R_{cryst} = \Sigma_h||F_{obs}|-|F_{calc}||/\Sigma_h|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[d]R_{free} was calculated as R_{cryst} using 5% of the randomly selected unique reflections that were omitted from structure refinement.

Table 12. Structural features of wild-type AKs and four LSE-optimized AK variants

	AKmeso	AKlse1	AKlse2	AKlse3[a]	AKlse4	AKthermo
Average LSE[b]	1.4938	1.4083	1.4085	1.4603	1.3696	1.4323
T_m (°C)	46.4	57.7	56.2	50.1	56.6	74.5[c]
Number of ion pairs[d]	4 (2)	2 (0)	4 (2)	3.5 (1)	7 (1)	6 (3)
Number of hydrogen bonds[d]	147 (41)	147 (36)	145 (38)	143.5 (36.5)	154 (39)	157 (43)
Number of apolar atoms	717	717	715	720	726	718
Aplor buried molecular surface area (Å^2)	3294	3335	3332	3313	3360	3420
Apolar accessible molecular surface area (Å^2)	1278	1295	1287	1330	1321	1238
Number of polar atoms	414	412	410	409	419	411
Polar buried molecular surface area (Å^2)	1646	1614	1634	1607	1705	1675
Polar accessible molecular surface area (Å^2)	1006	1056	1039	1057	1055	1015

[a]Aveage of the two chains in asymmetric unit
[b]Calculated for core domain (residues 1-30, 61-126 and 165-212)
[c]From Glaser et al., 1992.
[d]Values in parentheses are for interactions connecting distant regions (more than ten residues in a polypeptide).

However, such stabilizing features were not always the least in AKmeso, which had the highest LSE value and was used as a template for generating the LSE-optimized variants. For example, AKmeso did not have the fewest ion pairs or hydrogen bonds. In fact, the numbers of ion pairs and hydrogen bonds found in AKmeso were greater than or equal to those identified in the other three LSE-optimized variants (AKlse1, AKlse2, and AKlse3).

To identify the structural basis of thermal stabilization by LSE optimization more precisely, several plots were generated that describe the correlations between LSE and the structural features calculated based on the five structures (Fig. 23). Although the most LSE-optimized variant, AKlse4, revealed the most ion pairs and hydrogen bonds among the five AKs, correlations between the LSE and the numbers of ion pairs and hydrogen bonds were relatively weak.

Conversely, a strong correlation was observed between the average LSE and the apolar buried molecular surface area. This indicated that more apolar surface area becomes buried as the LSE was optimized. This was not entirely due to an increase in the number of apolar atoms upon LSE reduction because the correlation between the number of apolar atoms and LSE was poor (R^2 = 0.22) while the apolar buried surface area was almost perfectly correlated (R^2 = 0.98) with the average LSE value. No such strong correlation with LSE was observed with other structural features. These data suggest that LSE reduction may be closely related to the burial of apolar surface area, and that thermal stabilization via LSE optimization was a result of optimizing hydrophobic interactions.

LSE optimization overlooks and can damage stabilizing noncovalent interactions that connect distant regions of a polypeptide

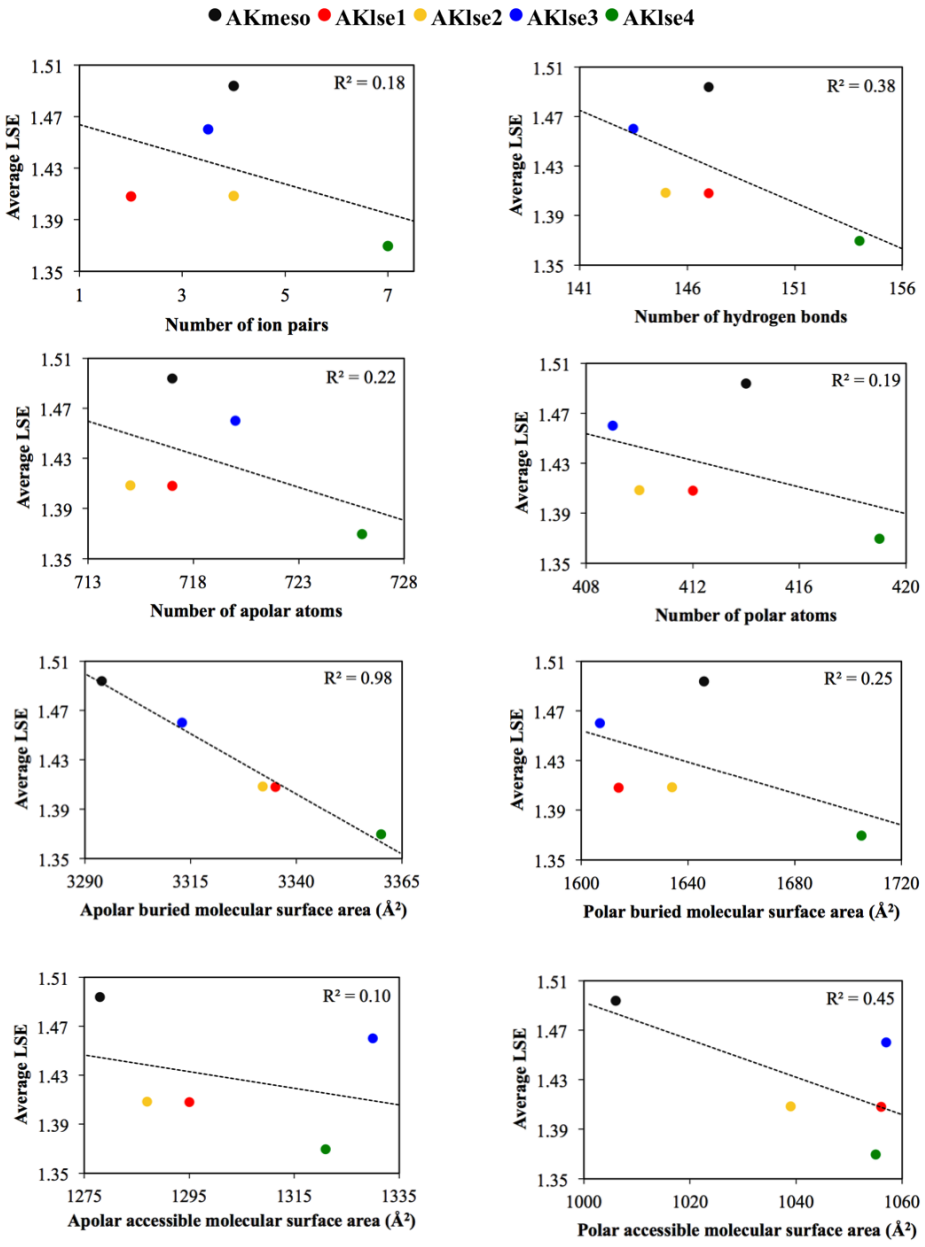
Although AKlse4 was designed to have the lowest LSE value among the four LSE-optimized AKmeso variants (Table 10), its T_m value was experimentally determined as 1.1°C lower than that of AKlse1, and only slightly higher (0.4°C) than that of AKlse2. This seems to indicate that an upper limit exists for the thermal stabilization of AKmeso by LSE optimization.

The results of structural comparisons were also complicated because the structure of AKlse4 appears to contain more stabilizing structural features, such as ion pairs, hydrogen bonds, and apolar buried surface area, than the other LSE-optimized AKmeso variants (Table 12). Because LSE describes conformational heterogeneity only in short stretches of a protein sequence, mutations designed to optimize LSE may disrupt noncovalent interactions that connect distant regions of a polypeptide. Such connections may be important for the overall stability of AKs (Bae and Phillips, 2004; 2005).

Therefore, additional structural analyses were performed, focusing on stabilizing structural features that involve residues that are distant in sequence but adjacent in three-dimensional space. When considering the interactions between two amino acid residues separated by more than 10 residues in a single polypeptide, AKlse4 was no longer the variant with the most ion pairs and hydrogen bonds. According to this definition for identifying interactions that work to connect distant regions, AKlse4 contains only one ion pair and 39 hydrogen bonds, both of which are fewer than those found in its template, AKmeso (Table 12).

Figure 23. Comparison of average LSE and structural features of four LSE-optimized AK variants and AKmeso

Data for the five AKs are represented by different colors: AKmeso (black), AKlse1 (red), AKlse2 (yellow), AKlse3 (blue), AKlse4 (green). Trend lines and R^2 values are also shown.


AKmeso AKlse1 AKlse2 AKlse3 AKlse4
Average LSE
R² = 0.18
Number of ion pairs
R² = 0.38
Number of hydrogen bonds
R² = 0.22
Number of apolar atoms
R² = 0.19
Number of polar atoms
R² = 0.98
Apolar buried molecular surface area (Å²)
R² = 0.25
Polar buried molecular surface area (Å²)
R² = 0.10
Apolar accessible molecular surface area (Å²)
R² = 0.45
Polar accessible molecular surface area (Å²)

Three such ion pairs (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) were found in the core domain of AKthermo (Bae and Phillips, 2004) but were not introduced into AKlse4 (Figs. 5 and 24), although residue substitutions were allowed from the thermophilic AK in the design of the AKlse4 sequence. This was because introducing AKthermo residues comprising these three ion pairs would result in unfavorable changes in LSE. In AKlse4, the three ion pairs of AKthermo were lost by residue mutations to uncharged amino acids or due to distances greater than the cutoff (4 Å) between the two oppositely charged residues.

A comparison of LSE-optimized AK variant structures also revealed an example in which hydrophobic interactions between distant regions in a given protein sequence were weakened due to residue substitutions designed to lower the LSE. While AKlse1 and AKlse2 make tight the hydrophobic contacts within 4 Å at Tyr109, Val193, and Ile211, AKlse4 were mutated to His, Arg, and Leu residues, respectively (Fig. 25 and Table 13).

These LSE-reducing mutations disrupted the hydrophobic packing around the residues and decreased the apolar buried surface area (Table 13). Thus, the observed limit of LSE optimization on the thermal stabilization of AKmeso may result from ignorance with regard to the stabilizing effect of noncovalent interactions between distant regions of a polypeptide.

To experimentally validate this hypothesis, three mutants of AKlse4 were designed in which the lost noncovalent interactions were recovered. In the first AKlse4 mutant, AKlse4m1, the three AKthermo ion pairs (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) were introduced.

In AKlse4m2, His109, Arg193, and Leu211 of AKlse4, were mutated to Tyr, Val, and Ile residues. In AKlse4m3, His109 and Arg193 of AKlse4, were mutated to Tyr and Val residues. These residues were formation of the hydrophobic interactions in AKlse1 and AKlse2 (Fig. 25 and Table 14).

By designing three AKlse4 mutants and testing their thermal stabilities, I experimentally demonstrated that residue substitutions for reducing LSE at the expense of stabilizing bridging interactions can decrease overall thermal stability. In the CD spectrum (Fig. 20), the thermal stabilities of these mutants were measured to determine whether these interactions confer additional thermal stabilization (Table 14). The AKlse4 mutants displayed higher T_m values than AKlse4, supporting this hypothesis.

Figure 24. Loss of ion pairs during LSE optimization in AKlse4

Three ion pairs of AKthermo (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) are found in the core domain, but they are not introduced into AKlse4 during LSE optimization. There are residues in AKlse4: (A) Arg19-Glu202, (B) Arg116-Gln198, and (C) Ala180-Glu114.

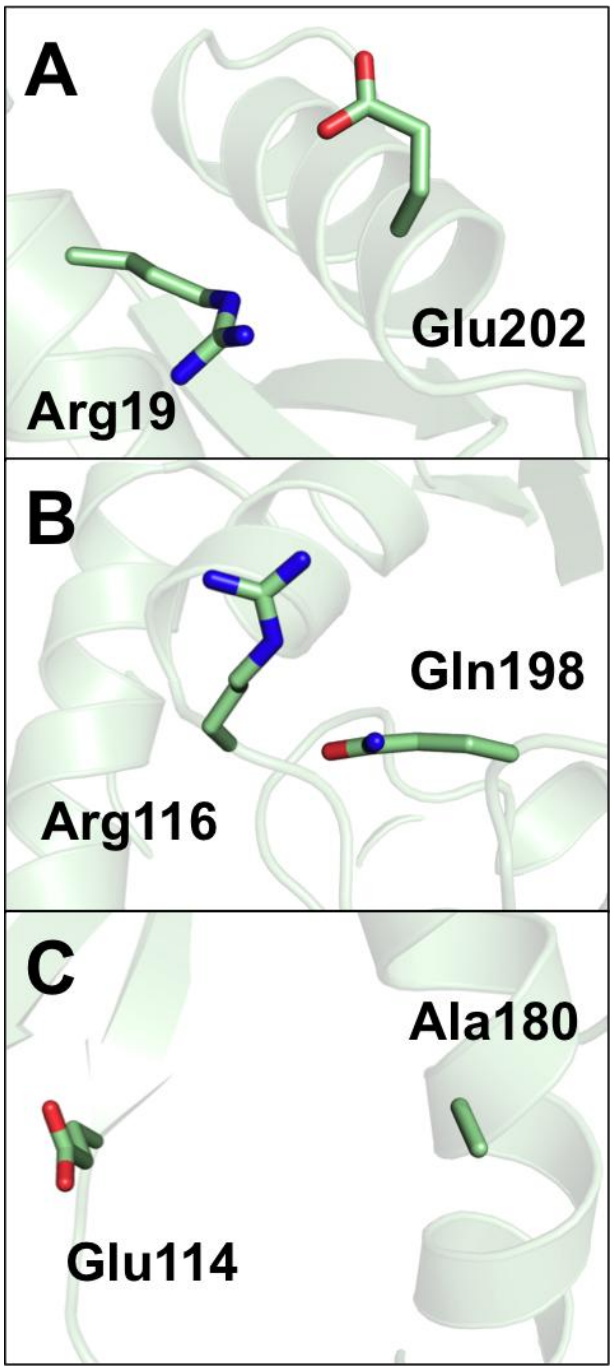

A
Glu202
Arg19
B
Gln198
Arg116
C
Ala180
Glu114

Figure 25. Effect of LSE-optimized mutations on hydrophobic interactions

Tyr109, Val193, and Ile211 residues make tight hydrophobic contacts in (A) AKlse1. In AKlse4 (B), the three residues are substituted to His109, Arg193, and Leu211, respectively, during LSE optimization. Van der Waals surfaces of the three residues are represented for side-chain atoms.

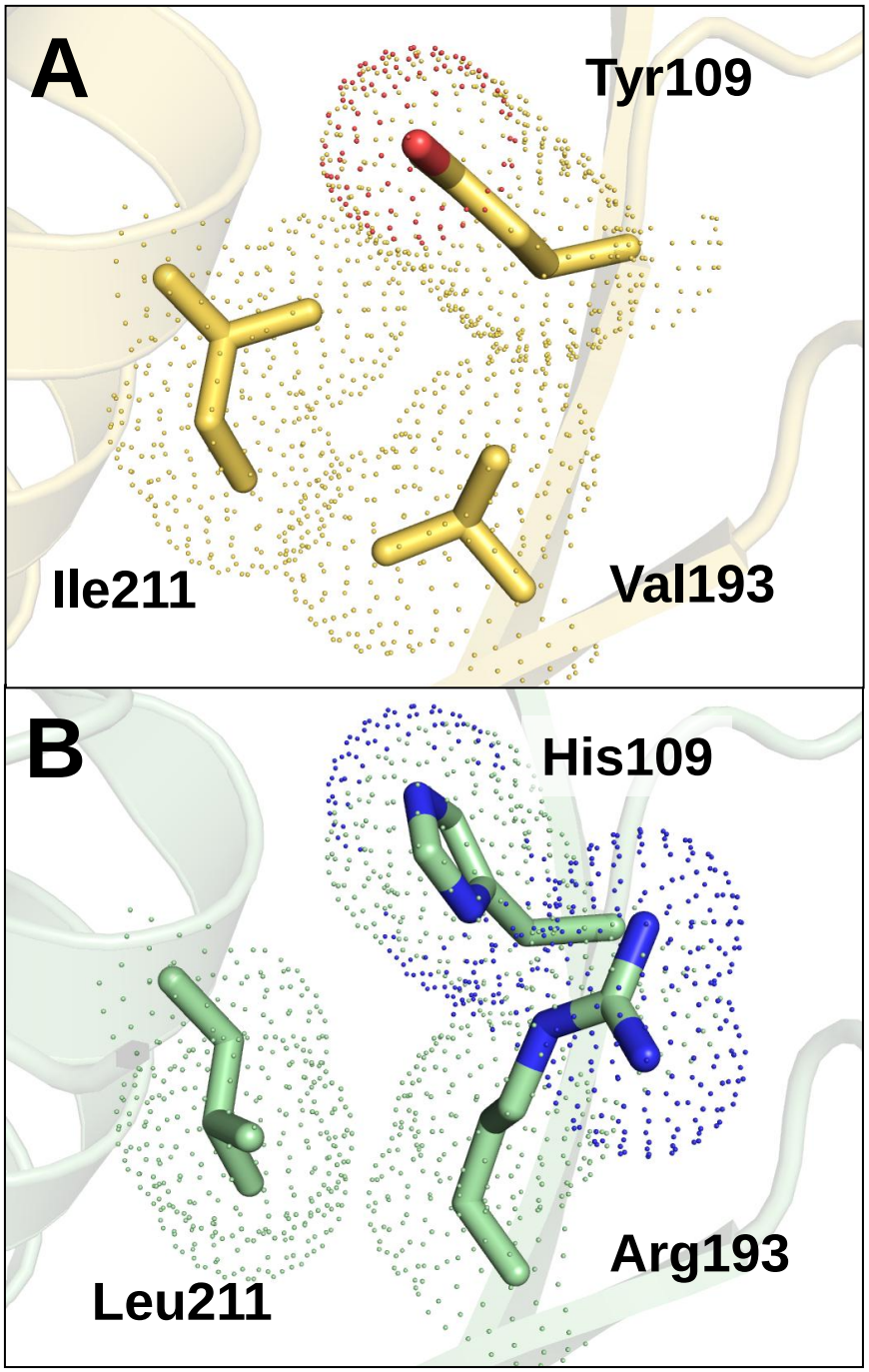

A
Tyr109
Ile211
Val193
B
His109
Leu211
Arg193

Table 13. Effect of LSE-optimized residue substitutions on apolar buried surface area of residues 109, 193, and 211

	Apolar buried molecular surface area (Å^2)						
	Residue 109		Residue 193		Residue 211		Total
AKlse1	Tyr	34	Val	23	Ile	43	100
AKlse2	Tyr	33	Val	22	Ile	41	96
AKlse3[a]	Tyr	34	Val	27	Leu	37	98
AKlse4	His	27	Arg	21	Leu	36	84

[a]Average of the two chains in asymmetric unit.

Structure determination of AKlse4 mutants

The AKlse4-based mutants, AKlse4m1 and AKlse4m2, were exhibited considerable increases in their thermal stabilities compared with their template (AKlse4). To verify the structural basis of thermal stabilization of AKlse4m1 and AKlse4m2 relative to that of AKlse4, the crystal structures were determined of them to resolutions of 2.90 and 1.65 Å, respectively (Fig. 27). Data collection and refinement statistics are summarized in Table 15.

The asymmetric unit of AKlse4m1 was included four polypeptide chains while two chains were identified in the structure of AKlse4m2. The chain folds of AKlse4m1 and AKlse4m2 were, essentially identical to those of other AKs, including AKlse4 (Figs. 2, 22, and 27). They display the characteristic AK domain arrangement with the bound inhibitor Ap_5A in the active site.

The RMSD values of Cα atoms between AKlse4 and each of the two mutants, AKlse4m1 and AKlse4m2, ranged from 1.2 to 2.0 Å. Thus, it was most likely that the differences in thermal stability between the AKlse4 mutants and AKlse4, result from relatively small structural changes caused by the mutations introduced to form stabilizing interactions connecting distant regions of the polypeptide.

Figure 26. Sequence alignment of AKlse4 mutants

Sequences of wild-type AKs, AKlse4, and AKlse4 mutants are aligned. Mutated residues for ion pairs and hydrophobic contacts indicate blue and red triangles, respectively. Secondary structures indicate based on the crystal structure of AKlse4.

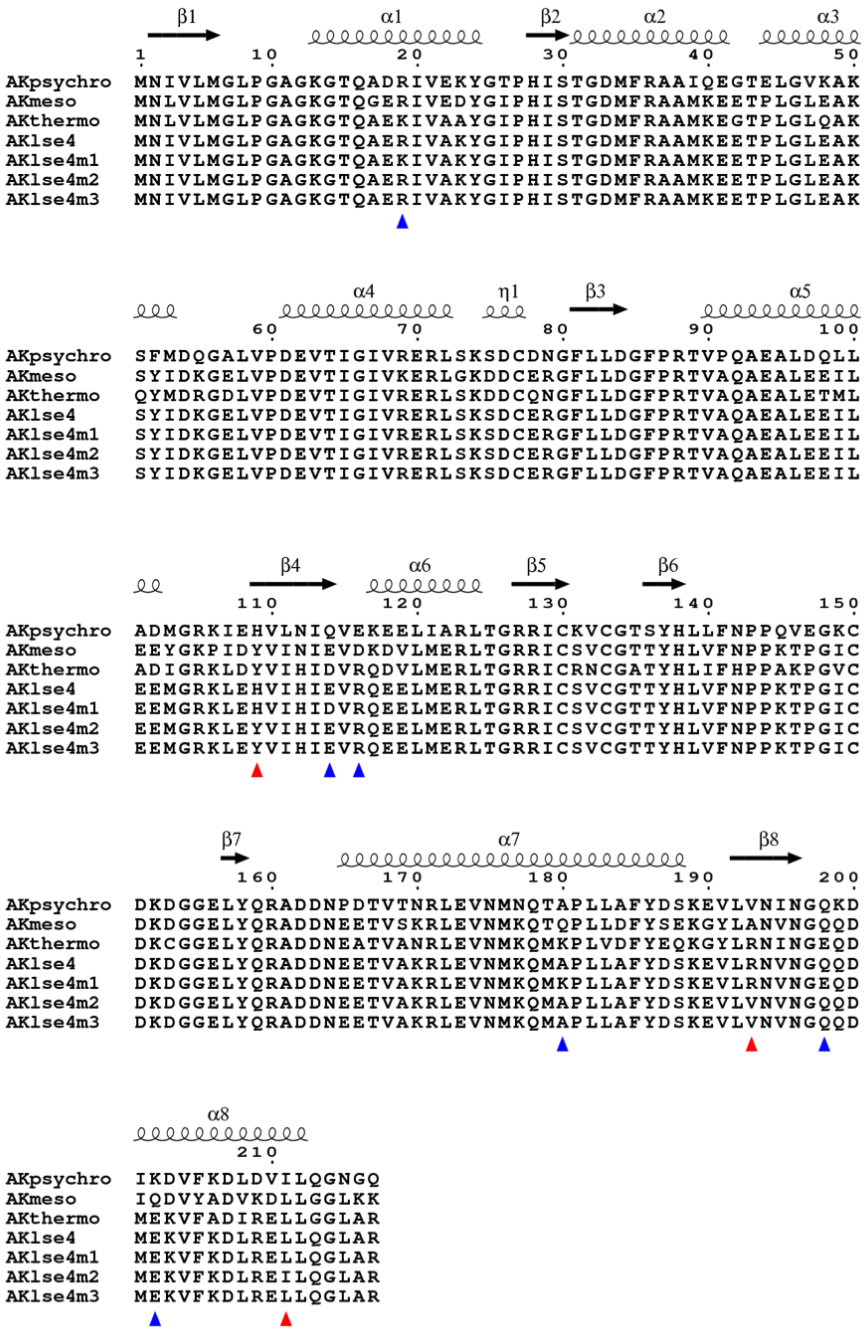
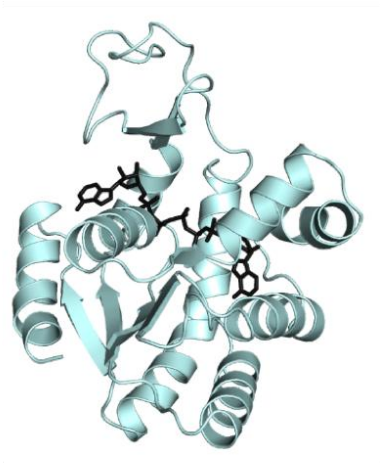

β1 α1 β2 α2 α3
1 10 20 30 40 50
AKpsychro MNIVLMGLPGAGKGTQADRIVEKYGTPHISTGDMFRAAIQEGTELGVKAK
AKmeso MNLVLMGLPGAGKGTQGERIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKthermo MNLVLMGLPGAGKGTQAEKIVAAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKlse4 MNIVLMGLPGAGKGTQAERIVAKYGIPHISTGDMFRAAMKEETPLGLEAK
AKlse4m1 MNIVLMGLPGAGKGTQAEKIVAKYGIPHISTGDMFRAAMKEETPLGLEAK
AKlse4m2 MNIVLMGLPGAGKGTQAERIVAKYGIPHISTGDMFRAAMKEETPLGLEAK
AKlse4m3 MNIVLMGLPGAGKGTQAERIVAKYGIPHISTGDMFRAAMKEETPLGLEAK
α4 η1 β3 α5
60 70 80 90 100
AKpsychro SFMDQGALVPDEVTIGIVRERLSKSDCDNGFLLDGFPRTVPQAEALDQLL
AKmeso SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKthermo QYMDRGDLVPDEVTIGIVRERLSKDDCQNGFLLDGFPRTVAQAEALETML
AKlse4 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
AKlse4m1 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
AKlse4m2 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
AKlse4m3 SYIDKGELVPDEVTIGIVRERLSKSDCERGFLLDGFPRTVAQAEALEEIL
β4 α6 β5 β6
110 120 130 140 150
AKpsychro ADMGRKIEHVLNIQVEKEELIARLTGRRICKVCGTSYHLLFNPPQVEGKC
AKmeso EEYGKPIDYVINIEVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKthermo ADIGRKLDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKlse4 EEMGRKLEHVIHIEVRQEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKlse4m1 EEMGRKLEHVIHIDVRQEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKlse4m2 EEMGRKLEYVIHIEVRQEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKlse4m3 EEMGRKLEYVIHIEVRQEELMERLTGRRICSVCGTTYHLVFNPPKTPGIC
β7 α7 β8
160 170 180 190 200
AKpsychro DKDGGELYQRADDNPDTVTNRLEVNMNQTAPLLAFYDSKEVLVNINGQKD
AKmeso DKDGGELYQRADDNEETVSKRLEVNMKQTQPLLDFYSEKGYLANVNGQQD
AKthermo DKCGGELYQRADDNEATVANRLEVNMKQMKPLVDFYEQKGYLRNINGEQD
AKlse4 DKDGGELYQRADDNEETVAKRLEVNMKQMAPLLAFYDSKEVLRNVNGQQD
AKlse4m1 DKDGGELYQRADDNEETVAKRLEVNMKQMKPLLAFYDSKEVLRNVNGEQD
AKlse4m2 DKDGGELYQRADDNEETVAKRLEVNMKQMAPLLAFYDSKEVLVNVNGQQD
AKlse4m3 DKDGGELYQRADDNEETVAKRLEVNMKQMAPLLAFYDSKEVLVNVNGQQD
α8
210
AKpsychro IKDVFKDLDVILQGNGQ
AKmeso IQDVYADVKDLLGGLKK
AKthermo MEKVFADIRELLGGLAR
AKlse4 MEKVFKDLRELLQGLAR
AKlse4m1 MEKVFKDLRELLQGLAR
AKlse4m2 MEKVFKDLREILQGLAR
AKlse4m3 MEKVFKDLRELLQGLAR

Figure 27. Structures of AKlse4 mutants

AKlse4m1 and AKlse4m2 indicate yellow and cyan, respectively. Bound Ap_5A inhibitor molecules are shown as stick model in black.

AKIse4m1



AKIse4m2

Table 14. Design strategy of AKlse4 mutants

	Design strategy	T_m (°C)	ΔT_m (°C)[a]
AKlse4	None	56.6	-
AKlse4m1[b]	Three ion pairs from AKthermo[c]	58.4	1.8
AKlse4m2	Hydrophobic contacts from AKlse1[d]	58.2	1.6
AKlse4m3	Hydrophobic contacts from AKlse1[e]	56.7	0.1

[a]Difference from T_m of AKlse4
[b]Designed by Thomas J. Rutkoski
[c]Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114
[d]Tyr109, Val193, and Ile211
[e]Tyr109 and Val193

Table 15. Data collection and refinement statistics[a] for structures of AKlse4 mutants

	AKlse4m1	AKlse4m2
Space group	$P2_1$	$P1$
Unit cell parameters (Å)	a=43.641, b=123.338, c=86.693, β=98.28°	a=39.003, b=48.790, c=54.832, α=90.00°, β=97.34°, γ=95.83°
Wavelength (Å)	0.97934	0.97960
Data collection statistics		
Resolution range (Å)	50.00-2.90 (3.00-2.90)	50.00-1.65 (1.71-1.65)
Number of reflections (measured/unique)	18989 (1878)	45746 (4395)
Completeness (%)	94.0 (94.1)	95.7 (92.3)
R_{merge}[b]	0.164 (0.611)	0.074 (0.198)
Redundancy	3.5 (3.6)	1.9 (1.9)
Mean I/σ	5.9 (2.3)	8.4 (4.6)
Refinement statistics		
Resolution range (Å)	50.00-2.90	50.00-1.65
R_{cryst}[c]/R_{free}[d] (%)	28.98 / 32.88	17.75 / 21.86
RMSD bonds (Å)	0.010	0.021
RMSD angles (deg)	1.303	2.246
Average B factor (Å^2)	42.461	18.729
Number of water molecules	13	332
Ramachandran favored (%)	96.86	99.77
Ramachandran allowed (%)	3.14	0.23

[a]Values in parentheses are for the highest-resolution shell.
[b]$R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.
[c]$R_{cryst} = \Sigma_h | |F_{obs}| - |F_{calc}| | / \Sigma_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[d]R_{free} was calculated as R_{cryst} using 5% of the randomly selected unique reflections that were omitted from structure refinement.

Introduction of ion pairs connecting distant regions of AKlse4m1

In the process of generating AKlse4m1 using AKlse4 as a template, mutations were introduced in order to construct three ion pairs (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) in AKthermo (Fig. 5). These ion pairs were originally found in the structure of AKthermo and connect distant regions (>10 residues) in the AKthermo sequence. In AKlse4, the ion pairs were not tightly maintained or lost by residue mutations to uncharged amino acids.

In the structure of AKlse4m1, oppositely charged side chains of the Lys19-Glu202 and Arg116-Glu198 ion pairs were found within close proximity to each other (Fig. 28), indicating that these two ion pairs were tightly maintained and serve to connect distant regions of the AKlse4m1 polypeptide. The average distances between the positively and negatively charged atoms in the Lys19-Glu202 and Arg116-Glu198 ion pairs are 4.21 and 3.17 Å, respectively, for the four independent chains in the crystal structure of AKlse4m1.

Conversely, the Lys180-Asp114 ion pair did not seem to be formed in AKlse4m1 since the distance between the two oppositely charged atoms was greater than 9 Å in all four of the AKlse4m1 chains. In a molecular dynamics simulation of AKthermo, the Lys180-Asp114 ion pair was not maintained within a short distance, suggesting that the formation of the Lys180-Asp114 ion pair in AKthermo might be a crystallographic artifact caused by the packing of neighboring molecules in the crystal lattice (Bae and Phillips, 2005).

Thus, the structure of AKlse4m1 showed that two ion pairs (Lys19-Glu202 and Arg116-Glu198) of the three ion pairs were successfully introduced into an LSE-optimized template by structure-guided mutagenesis.

This result suggests that these two ion pairs likely contribute to the increased thermal stability of AKlse4m1 compared to that of AKlse4.

In AKmeso variants, the introduction of the Lys19-Glu202 and Arg116-Glu198 ion pairs increased the T_m value of AKmeso (Table 5). In contrast, the Lys180-Asp114 ion pair did not have any effect on thermal stability when incorporated into AKmeso (Bae and Phillips, 2005).

Optimization of hydrophobic contacts for thermal stabilization in AKlse4m2

In AKlse4m2, the His109, Arg193, and Leu211 of AKlse4 were replaced with more hydrophobic amino acids (Tyr109, Val193, and Ile211) to optimize the hydrophobic interactions between distant polypeptide regions. Tyr, Val, and Ile residues were selected for mutation based on sequence alignments with wild-type AKs (Fig. 26). In the crystal structure of AKlse4m2 (Fig. 29), the hydrophobic side chain atoms of the three mutated residues make close hydrophobic contacts over relatively short distances (<4.5 Å). This suggests that the enhanced thermal stability of AKlse4m2, compared to AKlse4, results from the improved hydrophobic packing of residues from distant regions of the polypeptide.

Figure 28. Ion pairs connecting distant regions of a polypeptide in AKlse4m1

The structure of AKlse4m1 shows two ion pairs (A) Lys19-Glu202 and (B) Arg116-Glu198 by structure-guided mutagenesis.

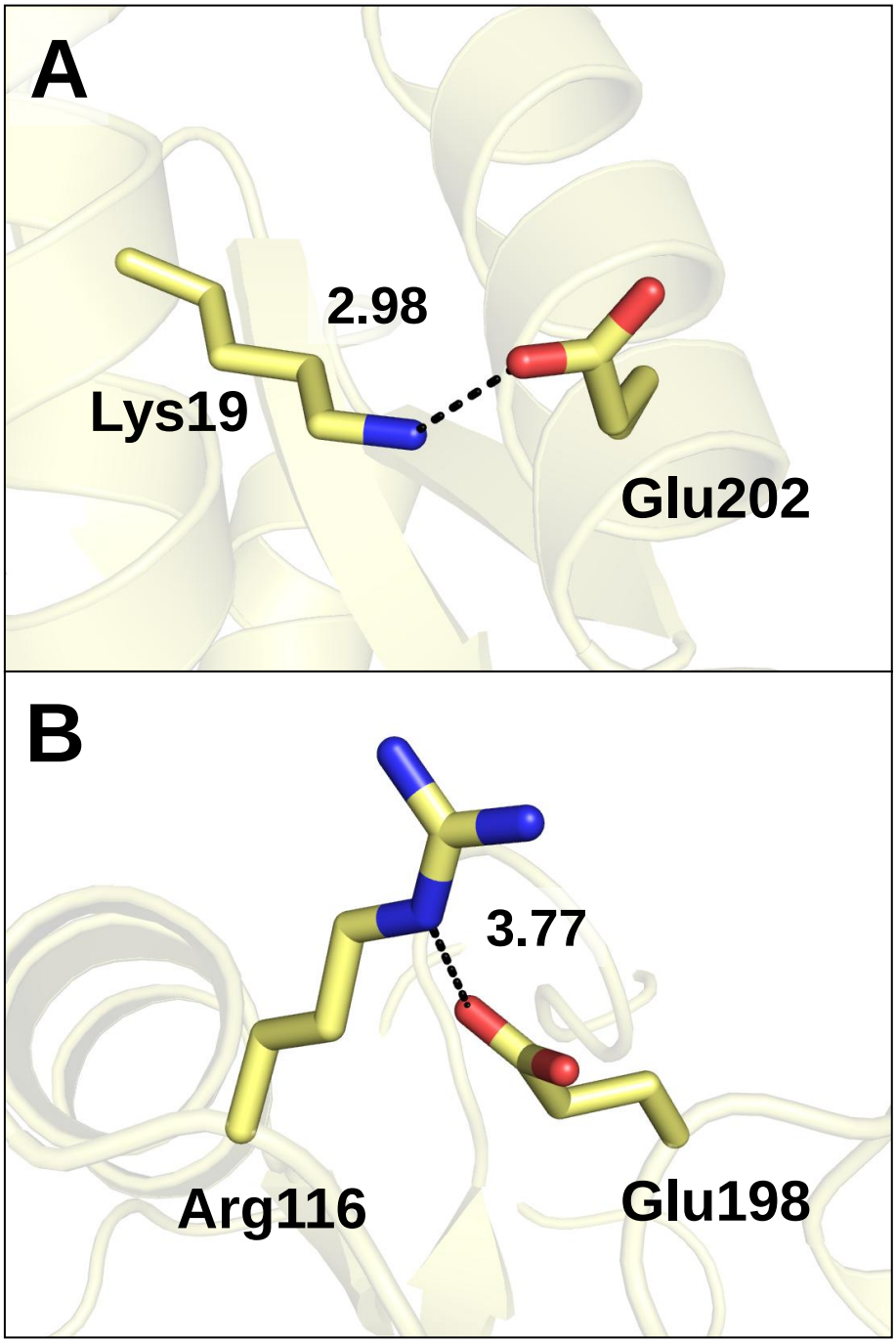
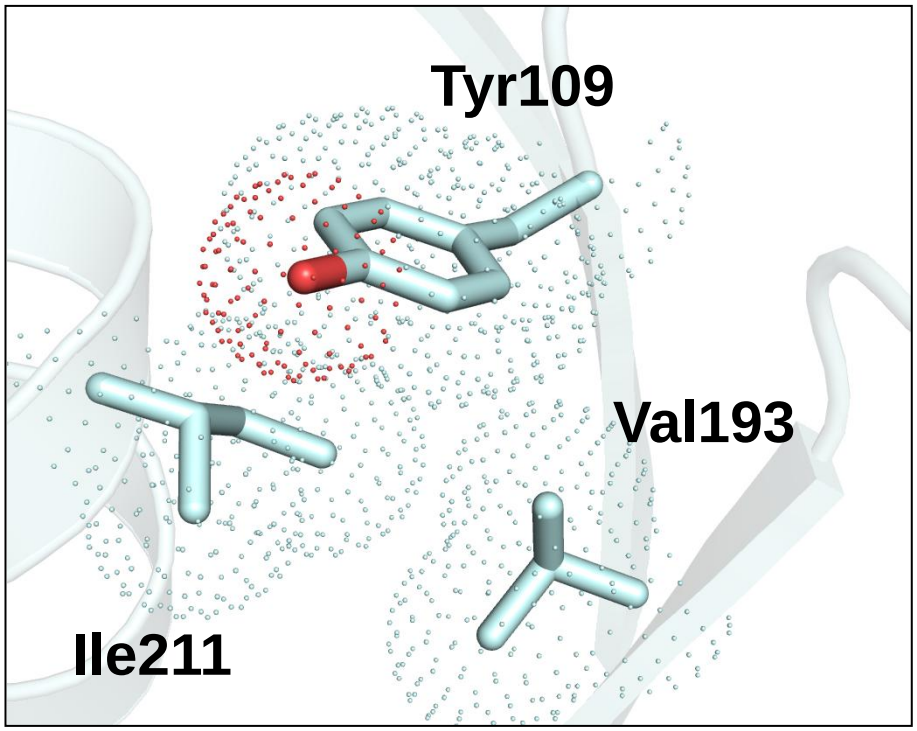

A
2.98
Lys19
Glu202
B
3.77
Arg116
Glu198

Figure 29. Optimizing hydrophobic packing for AKlse4m2

AKlse4m2 indicate improved hydrophobic packing by optimizing of stabilizing interactions. Tyr109, Val193, and Ile211 make tight hydrophobic contacts in AKlse4m2. For side-chain atoms of the three residues, van der Waals surfaces are represented.


Tyr109
Val193
Ile211

Additionally, I also designed another AKlse4 mutant, AKlse4m3, and tested its thermal stability by measuring its T_m value (Fig. 26 and Table 14). In AKlse4m3, the His109 and Arg193 of AKlse4 were replaced with Tyr and Val residues, respectively, as in AKlse4m2, but Leu211 was not changed. Despite the two mutated residues, the T_m value of AKlse4m3 (56.7°C) was only 0.1°C higher than that of AKlse4 (Table 14). This result indicated that these mutations did not increase the thermal stability of AKlse4m3 and suggests that an Ile-to-Leu substitution can disrupt optimal hydrophobic packing.

LSE optimization of thermophilic AK resulted in decreased thermal stability

Although the four LSE-optimized AK variants and the AKlse4 mutants exhibited considerable increases in their thermal stabilities compared with AKmeso, none of them was more thermally stable than AKthermo, a naturally evolved thermophilic homologue of AKmeso. Despite high sequence identity (74.5%), AKmeso and AKthermo display disparate thermal stabilities (Glaser et al., 1992). The difference between their T_m values is more than 25°C.

In an attempt to produce a more stable AK variant than the natural thermophilic homologue, Thomas J. Rutkoski designed an LSE-optimized AK variant using AKthermo as a template and residue substitutions allowed from AKmeso. The resulting lowest-LSE variant, AKlse5, contained 21 substituted residues in the core domain (Table 16 and Fig. 30).

Table 16. Design strategy of AKlse5 and AKlse6

	AKmeso	AKlse5[a]	AKlse6[a]	AKthermo
Template	-	AKthermo	AKthermo	-
Substitutions from	-	AKmeso	AKmeso	-
Additional design strategy	-	None	Charged residues excluded for substitutions	-
Number of substitutions	-	21	15	-
Average LSE[b]	1.4938	1.3905	1.4068	1.4323
T_m (°C)	46.4	65.1	70.7	74.5[c]
ΔT_m (°C)[d]	-	-9.4	-4.5	-

[a]Designed by Thomas J. Rutkoski

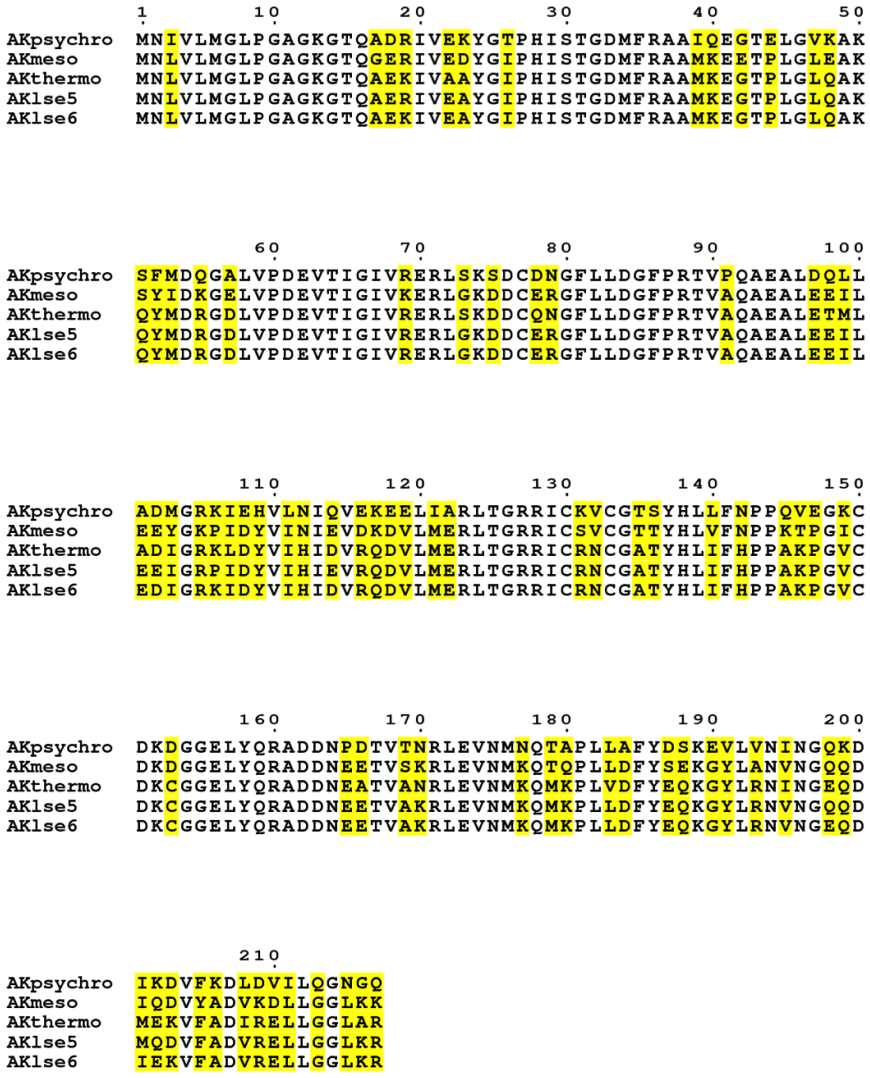
[b]Calculated for core domain (residues 1-30, 61-126 and 165-212)

[c]From Glaser et al., 1992.

[d]Difference from T_m of the respective template AK (AKthermo)

Figure 30. Sequence alignment of wild-type AKs, AKlse5, and AKlse6

The amino acid sequences of three wild-type AKs (AKpsychro, AKmeso, and, AKthermo) and two LSE-optimized AK variants (AKlse5 and AKlse6) are aligned together. Variable residues are highlighted in yellow.


1 10 20 30 40 50
AKpsychro MNIVLMGLPGAGKGTQADRIVEKYGTPHISTGDMFRAAIQEGTELGVKAK
AKmeso MNLVLMGLPGAGKGTQGERIVEDYGIPHISTGDMFRAAMKEETPLGLEAK
AKthermo MNLVLMGLPGAGKGTQAEKIVAAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKlse5 MNLVLMGLPGAGKGTQAERIVEAYGIPHISTGDMFRAAMKEGTPLGLQAK
AKlse6 MNLVLMGLPGAGKGTQAEKIVEAYGIPHISTGDMFRAAMKEGTPLGLQAK
60 70 80 90 100
AKpsychro SFMDQGALVPDEVTIGIVRERLSKSDCDNGFLLDGFPRTVPQAEALDQLL
AKmeso SYIDKGELVPDEVTIGIVKERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKthermo QYMDRGDLVPDEVTIGIVRERLSKDDCQNGFLLDGFPRTVAQAEALETML
AKlse5 QYMDRGDLVPDEVTIGIVRERLGKDDCERGFLLDGFPRTVAQAEALEEIL
AKlse6 QYMDRGDLVPDEVTIGIVRERLGKDDCERGFLLDGFPRTVAQAEALEEIL
110 120 130 140 150
AKpsychro ADMGRKIEHVLNIQVEKEELIARLTGRRICKVCGTSYHLLFNPPQVEGKC
AKmeso EEYGKPIDYVINIEVDKDVLMERLTGRRICSVCGTTYHLVFNPPKTPGIC
AKthermo ADIGRKLDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKlse5 EEIGRPIDYVIHIEVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
AKlse6 EDIGRKIDYVIHIDVRQDVLMERLTGRRICRNCGATYHLIFHPPAKPGVC
160 170 180 190 200
AKpsychro DKDGGELYQRADDNPDTVTNRLEVNMNQTAPLLAFYDSKEVLVNINGQKD
AKmeso DKDGGELYQRADDNEETVSKRLEVNMKQTQPLLDFYSEKGYLANVNGQQD
AKthermo DKCGGELYQRADDNEATVANRLEVNMKQMKPLVDFYEQKGYLRNINGEQD
AKlse5 DKCGGELYQRADDNEETVAKRLEVNMKQMKPLLDFYEQKGYLRNVNGQQD
AKlse6 DKCGGELYQRADDNEETVAKRLEVNMKQMKPLLDFYEQKGYLRNVNGEQD
210
AKpsychro IKDVFKDLDVILQGNGQ
AKmeso IQDVYADVKDLLGGLKK
AKthermo MEKVFADIRELLGGLAR
AKlse5 MQDVFADVRELLGGLKR
AKlse6 IEKVFADVRELLGGLKR

CD spectroscopy was used to measure the thermal stability of AKlse5 (Fig. 20). Surprisingly, the thermal stability of AKlse5 was not improved relative to its template, AKthermo, but rather exhibited a substantial decrease (9.4°C) in T_m despite a significantly lower LSE (Table 16).

Based on the analysis of the LSE optimization of AKmeso, I suspected that this unexpected decrease in thermal stability of AKlse5 might be the result of damage to important noncovalent stabilizing interactions between distant regions. One way to circumvent this problem would be to exclude residues involved in such interactions from LSE-optimizing substitutions.

To test this possibility, another LSE-optimized AKthermo variant, AKlse6, was generated in which all formally charged residues (Arg, Lys, Asp, and Glu) were excluded from substitutions. As a result, AKlse6 contained a higher LSE value and fewer residue substitutions than did AKlse5 (Table 16 and Fig. 30). In the design of AKlse6, Thomas J. Rutkoski intended to maintain bridging ion pairs by leaving out charged residues while optimizing the burial of apolar surface area.

However, the T_m value of AKlse6 remained lower than that of AKthermo, although the extent of destabilization was reduced compared with AKlse5 (Table 16). To understand the structural basis of these decreased thermal stabilities, the crystal structures of AKlse5 and AKlse6 were determined to resolutions of 1.68 and 1.67 Å, respectively (Fig. 31). Data collection and refinement statistics are summarized in Table 17. I noted that the beta angles in the unit cell parameters were close to 90°. However, processing the data in higher symmetries significantly increased R_{merge}, suggesting that the monoclinic space group was correct.

Table 17. Data collection and refinement statictics[a] for structures of AKlse5 and AKlse6

	AKlse5	AKlse6
Space group	$P2_1$	$P2_1$
Unit cell parameters (Å)	a=36.1, b=75.4, c=82.0, β=90.0°	a=36.8, b=76.5, c=85.1, β=90.2°
Wavelength (Å)	0.9793	0.9793
Data collection statistics		
Resolution range (Å)	50.00-1.68 (1.74-1.68)	50.00-1.67 (1.73-1.67)
Number of reflections (measured/unique)	358723/49458	202169/54822
Completeness (%)	97.6 (95.7)	100.0 (100.0)
R_{merge}[b]	0.074 (0.728)	0.094 (0.617)
Redundancy	7.3 (7.1)	3.7 (3.7)
Mean I/σ	22.8 (2.7)	12.8 (2.2)
Refinement statistics		
Resolution range (Å)	50.00-1.68	50.00-1.67
R_{cryst}[c]/R_{free}[d] (%)	17.8/22.4	16.8/20.8
RMSD bonds (Å)	0.02	0.025
RMSD angles (deg)	2.186	2.428
Average *B* factor (Å^2)	18.59	22.04
Number of water molecules	401	537
Ramachandran favored (%)	98.8	98.4
Ramachandran allowed (%)	1.2	1.6

[a]Values in parentheses are for the highest-resolution shell.

[b]$R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

[c]$R_{cryst} = \Sigma_h | |F_{obs}| - |F_{calc}| | / \Sigma_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d]R_{free} was calculated as R_{cryst} using 5% of the randomly selected unique reflections that were omitted from structure refinement.

Figure 31. Overall structures of AKlse5 and AKlse6

Crystal structures of two LSE-optimized AK variants, AKlse5 and AKlse6 indicate grey and chocolate, respectively. The bound inhibitor Ap_5A molecules are shown as stick model.

AKlse5

AKlse6

Both AKlse5 and AKlse6 contain two molecules in their asymmetric units and were structurally similar to other AKs with regard to chain folding, the Ap_5A binding, and domain arrangement (Figs. 2, 22, and 31). Several stabilizing structural features were compared between the two LSE-optimized AKthermo variants and AKthermo to identify the molecular mechanism for the unexpected destabilization following LSE optimization (Table 18).

Among the three structures, AKlse5 contained the fewest ion pairs and hydrogen bonds. This was also true when counting only those interactions that serve to connect distant regions of the polypeptide. AKlse6, which was generated by excluding formally charged residues from LSE-optimizing substitution, contains more ion pairs than AKthermo while the number of those that bridge distant regions were identical in AKlse6 and AKthermo. AKlse6 contained fewer hydrogen bonds than AKthermo, under all conditions, but the decrease in the number of hydrogen bonds was reduced relative to AKlse5.

Structural comparisons also indicated that LSE optimization of AKthermo decreased the apolar buried surface area (Table 18). This was surprising because in the experiment for AKmeso, the reduction in LSE increased the buried apolar surface area (Table 9).

In addition, the two LSE-optimized AKthermo variants (AKlse5 and AKlse6) contained more apolar atoms than AKthermo (Table 18). The difference in the buried apolar surface area between AKlse5, the most LSE-optimized variant, and AKthermo was 82 $Å^2$, greater than the largest increase (66 $Å^2$ between AKlse4 and AKmeso) observed in the AKmeso optimizations.

Table 18. Structural features of AKlse5 and AKlse6

	AKmeso	AKlse5[a]	AKlse6[a]	AKthermo
Average LSE[b]	1.4938	1.3905	1.4068	1.4323
T_m (°C)	46.4	65.1	70.7	74.5[c]
Number of ion pairs[d]	4 (2)	4 (2)	8 (3)	6 (3)
Number of hydrogen bonds[d]	147 (41)	141.5 (37.5)	185 (39)	157 (43)
Number of apolar atoms	717	726	727	718
Aplor buried molecular surface area ($Å^2$)	3294	3338	3340	3420
Apolar accessible molecular surface area ($Å^2$)	1278	1305	1371	1238
Number of polar atoms	414	419	417	411
Polar buried molecular surface area ($Å^2$)	1646	1633	1670	1675
Polar accessible molecular surface area ($Å^2$)	1006	1117	1109	1015

[a]Average of the two chains in asymmetric unit

[b]Calculated for core domain (residues 1-30, 61-126 and 165-212)

[c]From Glaser et al., 1992.

[d]Values in parentheses are for interactions connecting distant regions (more than ten residues in a polypeptide).

In fact, the buried apolar surface area calculated for AKthermo was larger than that of AKlse4, which displayed the greatest buried apolar surface area of the AKmeso variants (Table 9), suggesting highly optimized hydrophobic packing in AKthermo. Thus, the substitutions designed to reduce LSE likely disrupt local and/or global hydrophobic interactions in AKthermo, resulting in relative thermal destabilization.

Discussion

In this study, I tested the effectiveness of several stabilization principles by designing a series of AK variants and generating substituted residues in the core domain. Previous chimeric AKs study of Bae and Phillips (2006) revealed that the overall stability of the AKs is closely related to the core domains (residues 1–30, 61–126, and 165–217), and the catalytic activity is controlled by the other regions of the AKs. Accordingly, only residues in the core domain were considered for mutation.

Additionally, one of the 23 AKpsychro-like substitutions which is identified by LSE optimization is Asp23Lys in AKlse1 (Bae et al., 2008). Despite LSE optimization the mutation is expected to break an electrostatic attraction between oppositely charged residues bridging distant regions of the polypeptide. Switching the charges at this position could potentially destabilize the protein structure by removing favorable electrostatic interactions.

After cloning the AK variants, I measured their thermal stabilities using CD spectroscopy. To ensure the validity of my comparisons, it is important to perform all experiments in the same manner, especially when determining T_m values as these measurements have been shown to vary substantially depending on the instruments and the experimental conditions used. A careful examination of the experimental procedures is therefore necessary when comparing thermal stabilities of proteins based on their T_m values. Considering the identical experimental conditions and the large number of stable AK variants tested in this study, I believe that present data provide a

unique opportunity for comparing the effectiveness and additivity of different thermal stabilization mechanisms identified using a variety of techniques.

In addition to T_m value, which I used to compare thermal stabilities of the AK variants, the free energy of maximal stability can also be used as a measure of protein stability. A correlation between these two values has been previously observed, indicating that an increase in T_m tends to be associated with an increase in the free energy of maximal stability (John and Weeks, 2000). However, it is still possible that a protein with a higher free energy value can undergo thermal denaturation at a lower temperature. Thus, I believe that T_m can better serve as a practical measure of protein thermal stability and my interpretation based on the T_m values of the AK variants was valid.

Although the AK variants were designed to exhibit higher thermal stability, they shared higher sequence identities with the mesophilic AK than with AKthermo, the naturally-evolved thermophilic homologue (Table 19). These observations suggest that the basis for increased thermal stability in AK variants may differ from that of AKthermo and that my collective stabilization strategy was more efficient than natural evolution at least in terms of the number of mutations required to achieve thermal stability.

The AK variants can be classified into two groups. The first seven variants (AKv1 to AKv7) were constructed using AKmeso as a template while the other variants were produced based on the LSE-optimized variant, AKv8. I was able to generate several pairs of variants, which were created by introducing equivalent mutations into the two different templates. Five such pairs are produced: AKv1-AKv9, AKv2-AKv10, AKv3-AKv11, AKv4-AKv12, and AKv5-AKv13.

Table 19. Sequence identities of AK variants to wild-type AKs

AK	Sequence identity (%)	
	To AKmeso	To AKthermo
AKmeso	-	74.5
AKthermo	74.5	-
AKv1	99.5	75.1
AKv2	98.2	76.5
AKv3	97.7	77.0
AKv4	99.1	73.7
AKv5	97.2	76.5
AKv6	97.2	76.5
AKv7	96.8	76.0
AKv8	89.4	73.3
AKv9	88.9	73.3
AKv10	87.6	75.1
AKv11	88.5	72.4
AKv12	88.9	73.3
AKv13	87.1	75.6
AKv14	87.1	75.1
AKv15	86.6	75.1
AKv16	86.6	75.6
AKv17	86.2	75.1
AKv18	86.7	74.7

Within each pair, the introduced mutations result in lower thermal stabilization in the AKv8 template (49–96%) than in AKmeso (Table 5), highlighting the difficulty of protein thermal stabilization at elevated temperatures. The results present here provide an opportunity for comparing the effects of different thermal stabilization techniques since all of the T_m measurements in this study are performed under identical conditions.

AKv3, AKv4, and AKv8 were contained mutations identified based on comparative study (Bae and Phillips, 2005), molecular evolution (Counago et al., 2006), and LSE optimization (Bae et al., 2008), resulting in T_m increases of 9.0, 14.0, and 11.3°C, respectively, relative to AKmeso (Table 5). From these results it appears that molecular evolution is the most efficient method as it produced the greatest increase in thermal stability with the smallest number of mutations. However, from a practical point of view, molecular evolution may have disadvantages. It often requires development of new screening methods for each protein and can be both expensive and labor-intensive compared with other techniques. In contrast, LSE optimization only requires the sequence information of a target and its homologue.

Although I selected only two mutations (Gln16Leu and Gln199Arg), based on the results of the molecular evolution study, several other less stabilizing mutations have also been found, including Thr179Ile, Ala193Val, Gly213Glu, and Gly214Arg (Counago et al., 2006). Some of these residues were also identified by other methods, including LSE optimization, which also produce the Ala193Val mutation.

Based upon the comparative study, Thr179 is mutated to Met, resulting in improved hydrophobic packing. Furthermore, replacement of surface Thr residues with more hydrophobic amino acids has been shown to confer

stabilizing effects in other target proteins (Schmid, 2011). Direct comparison of several AK structures failed to identify clear correlation between structural features and thermal stabilities (Table 9). For detecting ion pairs and hydrogen bonds and for calculating surface areas, it seems that crystallization artifacts lead to both false positives and negatives.

These examples highlight the difficulty of identifying ion pairs, especially those existing at the surface of the molecule where interactions with crystallization reagents or neighboring molecules in the crystal lattice can affect side chain conformation. One should therefore be careful when making definitive predictions regarding a protein's thermal stability based solely on small differences in the structural features.

I had also tried to identify the structural basis of thermal stabilization conferred by LSE optimization by directly comparing AK structures. Since LSE optimization considers only the local context of a protein's sequence, it is possible that the process could break stabilizing interactions among distant regions of a polypeptide.

These results indicate that for more significant thermal stabilization, it is important to consider both local stabilizing factors as well as global interactions among distant regions of a polypeptide. The significant thermal stabilization obtained in this study is the result of multiple stabilizing mutations identified by using a number of different procedures.

The AK variants were produced using only a single stabilization technique also showed considerable increases (9.0–14.0°C) in their thermal stabilities (Table 5). However, the largest gains are observed when mutations are introduced together. This suggests that the effects of different stabilization methods can be additive and do not significantly interfere with one another.

This could be just a coincidence, but it may also result from the fact that the LSE optimization only considers local structural features while the other mutations primarily improved noncovalent interactions connecting distant regions of the polypeptide. This is also supported by a poor correlation (R^2 = 0.17) between the average LSE and T_m values of the 18 variants (Fig. 32).

Additionally, I designed LSE-optimized AK variants with enhanced stability based on measures of LSE (Bae et al., 2008). Basis on the results, I confirm that LSE optimization can be an effective method for protein thermal stabilization. However, the current study also reveals cases in which LSE optimization resulted in limited or even negative effects on protein thermal stability. These results indicate that the current study provide a unique opportunity to examine both effectiveness and the limitations of LSE optimization as a strategy for protein thermal stabilization.

To elucidate the structural basis of thermal stabilization by LSE optimization, I analyzed the crystal structures of four LSE-optimized AKmeso variants and their template, AKmeso. Because the LSE values of the four variants are optimized to various extents, this study allows the examination of correlations between LSE and specific stabilizing structural features instead of simpler one-to-one comparisons.

The most striking result obtained from structural analyses of the AKmeso variants is that the strong inverse correlation between LSE and the apolar buried surface area (R^2 = 0.98), indicating that more apolar surface area becomes buried as LSE is reduced. However, this is likely not due to changes in the number of apolar atoms, which correlates weakly (R^2 = 0.22) with LSE.

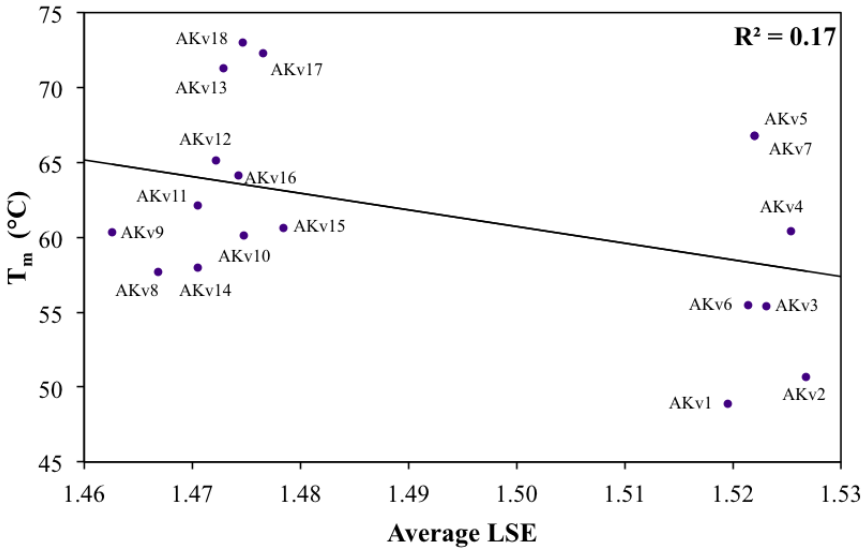
Substitutions made for the purpose of reducing LSE may increase the apolar buried surface area, which is often considered a measure of the degree of hydrophobic interactions. Thus, enhanced thermal stabilization is ultimately achieved by optimizing the hydrophobic packing of proteins.

The optimization of hydrophobic interactions has been recognized as an important stabilization strategy in a number of thermally stable AKs. Among the three wild-type AKs, AKthermo contains the highest apolar buried surface area (Bae and Phillips, 2004). Structural comparisons of wild-type AKs also identified a specific AKthermo residue (Met179) that is involved in hydrophobic contacts with conserved residues (Bae and Phillips, 2004) and was later found to contribute to thermal stabilization when substituted into AKmeso (Table 5).

Experimental evolution of AKmeso for higher thermal stability also generate several stable variants containing mutations (Q16L, T179I, and A193V) that resulted in better hydrophobic packing (Counago et al., 2006; Miller et al., 2010). Interestingly, one of these mutations, A193V, is also generated in the LSE optimization of AKmeso and is included in each of the three LSE-optimized variants (AKlse1, AKlse2, and AKlse3) (Bae et al., 2008).

Figure 32. A plot of T_m versus average LSE for 18 AK variants

In 18 AK variants, a poor correlation between average LSE and thermal stability (T_m value) are observed. AKv5 and AKv7 have the same average LSE value.


R² = 0.17
AKv18
AKv17
AKv13
AKv12
AKv16
AKv11
AKv9
AKv15
AKv10
AKv8
AKv14
AKv5
AKv7
AKv4
AKv6
AKv3
AKv1
AKv2
75
70
65
60
55
50
45
1.46
1.47
1.48
1.49
1.50
1.51
1.52
1.53
Tm (°C)
Average LSE

A recent computational study also shows that AKmeso could be thermally stabilized for its hydrophobic core (Howell et al., 2014). Two of the mutations (L3I and L211I) identified in this computational protein design process are also found in the LSE optimization of AKmeso (AKlse1 and AKlse2) (Bae et al., 2008).

In the LSE optimization of AKmeso, the resulting thermal stabilization was effective but limited. Although all four LSE-optimized AKmeso variants yielded higher T_m values than that of their template, AKmeso, the observed increases in T_m were not completely proportional to the decreases in LSE. AKlse4, the most LSE-optimized AKmeso variant, had a significantly lower LSE value than either AKlse1 or AKlse2, but these three AKmeso variants exhibited similar T_m values (Table 10).

These results indicate a stability limit for LSE-optimized AKmeso variants. It is also puzzling to note that the AKlse4 structure contained the largest number of stabilizing structural features, such as ion pairs, hydrogen bonds, and apolar buried surface area, among the variants (Table 12). The importance of noncovalent interactions that serve to connect distant regions is also recognized in my previous studies of AKs.

In structural comparisons of three wild-type AKs, such interactions are identified most often in AKthermo (Bae and Phillips, 2004). Several of these interactions are later found to contribute to overall stability when substituted into AKmeso (Bae and Phillips, 2005). In the process of generating AKlse1, I encountered a situation in which LSE optimization damaged the noncovalent interactions bridging distant regions. When residues in AKmeso were substituted with those of AKpsychro for the purpose of reducing LSE, an ion pair between residues 23 and 209 connecting the N- and C-terminal regions of

AK was broken, and a mutation introduced into AKlse1 for reconstituting the bridging electrostatic interaction resulted in additional stabilization (Fig. 15). It is interesting to note that a stability limit is also recognized in another thermal stabilization trial of AKmeso.

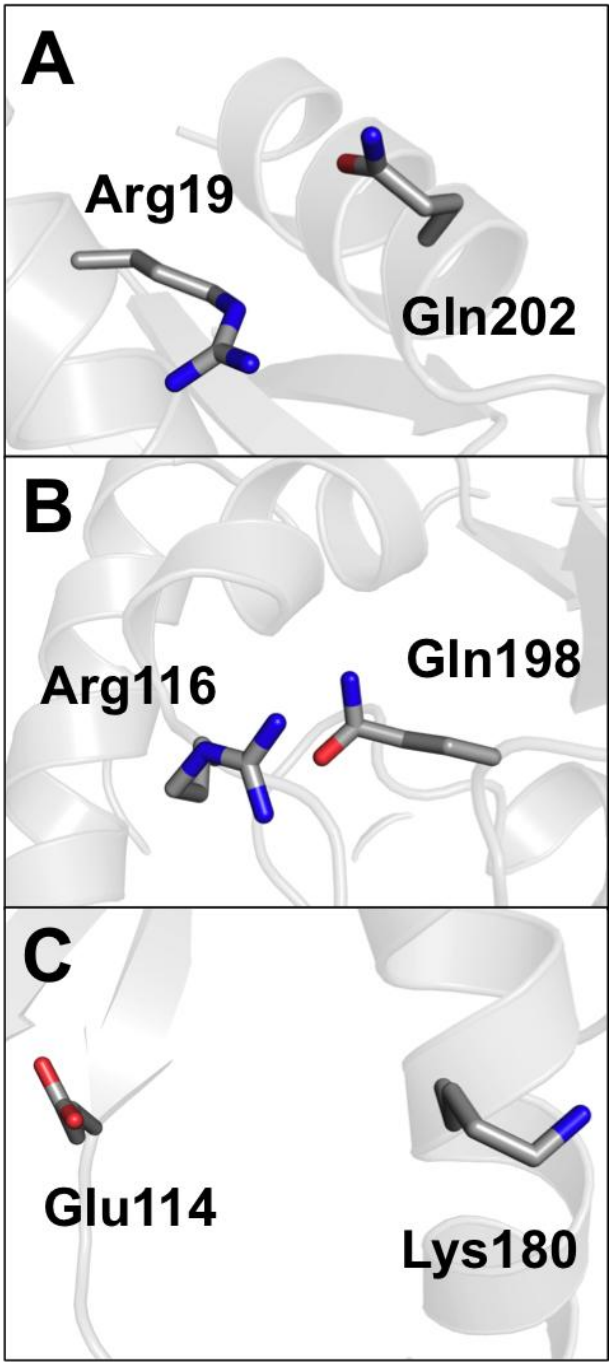
In the computational study (Howell et al., 2014), AKmeso variants are stabilized for repacking the hydrophobic core. The variants also display an upper thermal limit that could be exceeded only by introducing additional electrostatic interactions. This finding is consistent with the results of these studies assuming that the thermal stabilization of AKmeso by LSE optimization resulted from the optimization of hydrophobic interactions.

Additionally, the introduction of three ion pairs into AKlse4, the most LSE-optimized AKmeso variant, increased its thermal stability. Previously, AKlse1 was stabilized to different extents by adding one to three ion pairs (Table 5). While the LSE optimization of AKmeso generated more thermally stable variants, reducing the LSE of AKthermo decreased its thermal stability. The T_m of AKlse5, in which 21 residues were substituted with those of AKmeso, was 9.4°C lower than that of AKthermo despite its considerably lower LSE value (Table 16).

I speculated that the destabilization encountered during the reduction of the LSE of AKthermo to generate AKlse5 was the result of destruction of three ion pairs (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) that served to connect distant regions of AKthermo (Fig. 33). To test this hypothesis, I generated another AKthermo variant, AKlse6, by excluding formally charged residues from the substitution pool during LSE optimization.

Figure 33. Loss of ion pairs during LSE optimization in AKlse5

AKlse5 is results of destruction of three ion pair (Lys19-Glu202, Arg116-Glu198, and Lys180-Asp114) that served to connect distant regions of AKthermo. There are residues in AKlse5: (A) Arg19-Gln202, (B) Arg116-Gln198, and (C) Lys180-Glu114.


A
Arg19
Gln202
B
Arg116
Gln198
C
Glu114
Lys180

Although the thermal stability of AKlse6 was improved relative to that of AKlse5, its T_m remained lower than that of its template, AKthermo. This suggests that the loss of the ion pairs is not solely responsible for the observed decrease in thermal stability.

Structural analysis revealed that both AKlse5 and AKlse6 contained significantly reduce apolar buried surface areas compared with their template, AKthermo, despite increases in the number of apolar atoms (Table 18). This suggests that LSE optimization might disrupt not only stabilizing ion pairs, but also favorable hydrophobic interactions in AKthermo. It seems that AKthermo, like many other natural thermophilic proteins, already possesses highly organized networks of stabilizing structural features such as ion pairs and hydrophobic interactions that are easily damaged by substituting certain residues.

Thus, to increase the thermal stability of thermophilic proteins, it is likely necessary to carefully introduce additional stabilizing features while minimizing the destruction of existing ones. This, however, is difficult if detailed structural information is not available for the target proteins.

Despite its limitations, I believe that LSE optimization can still be an effective and efficient method for the thermal stabilization of proteins. As shown in the case of AKlse1, LSE optimization can significantly increase the thermal stability of a mesophilic target (AKmeso) by more than 10°C using only the sequence information from one homologue (AKpsychro).

Other stabilization approaches require detailed structural information of the target or a large number of homologous sequences. LSE optimization may also be useful when combined with other techniques.

LSE-optimized sequences can serve as a foundation to which stabilizing mutations identified by other methods are introduced. Alternatively, residues involved in important stabilizing interactions can be excluded from pool of available mutations used in LSE optimization.

The T_m of AKmeso was increased by 26.6°C using LSE optimization simultaneously with other approaches, including structure-guided mutagenesis and molecular evolution. Additionally, the thermal stability of AKlse4 was increased by additionally introducing noncovalent interactions identified in structure-based analyses. These results also highlight the importance and complementarity of local conformational stability and global interactions in protein thermal stability.

References

Adrio JL and Demain AL (2014) Microbial enzymes: tools for biotechnological processes. *Biomolecules* 4, 117-139.

Bae E, Bannen RM and Phillips GN, Jr. (2008) Bioinformatic method for protein thermal stabilization by structural entropy optimization. *Proc Natl Acad Sci U S A* 105, 9594-9597.

Bae E and Phillips GN, Jr. (2004) Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J Biol Chem* 279, 28202-28208.

Bae E and Phillips GN, Jr. (2005) Identifying and engineering ion pairs in adenylate kinases. Insights from molecular dynamics simulations of thermophilic and mesophilic homologues. *J Biol Chem* 280, 30943-30948.

Bae E and Phillips GN, Jr. (2006) Roles of static and dynamic domains in stability and catalysis of adenylate kinase. *Proc Natl Acad Sci U S A* 103, 2132-2137.

Bannen RM, Suresh V, Phillips GN, Jr., Wright SJ and Mitchell JC (2008) Optimal design of thermally stable proteins. *Bioinformatics* 24, 2339-2343.

Battye TG, Kontogiannis L, Johnson O, Powell HR and Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst D* 67, 271-281.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.

Berry MB and Phillips GN, Jr. (1998) Crystal structures of Bacillus stearothermophilus adenylate kinase with bound Ap5A, Mg2+ Ap5A, and Mn2+ Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg2+ and Mn2+. *Proteins* 32, 276-288.

Chan CH, Liang HK, Hsiao NW, Ko MT, Lyu PC and Hwang JK (2004) Relationship between local structural entropy and protein thermostability. *Proteins* 57, 684-691.

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* 66, 12-21.

Corrons JL, Garcia E, Tusell JJ, Varughese KI, West C and Beutler E (2003) Red cell adenylate kinase deficiency: molecular study of 3 new mutations (118G>A, 190G>A, and GAC deletion) associated with hereditary nonspherocytic hemolytic anemia. *Blood* 102, 353-356.

Counago R, Chen S and Shamoo Y (2006) In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol Cell* 22, 441-449.

Counago R and Shamoo Y (2005) Gene replacement of adenylate kinase in the gram-positive thermophile Geobacillus stearothermophilus disrupts adenine nucleotide homeostasis and reduces cell viability. *Extremophiles* 9, 135-144.

Counago R, Wilson CJ, Pena MI, Wittung-Stafshede P and Shamoo Y (2008) An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. *Protein Eng Des Sel* 21, 19-27.

Criswell AR, Bae E, Stec B, Konisky J and Phillips GN, Jr. (2003) Structures of thermophilic and mesophilic adenylate kinases from the genus Methanococcus. *J Mol Biol* 330, 1087-1099.

Currin A, Swainston N, Day PJ and Kell DB (2015) Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem Soc Rev* 44, 1172-1239.

Dahiyat BI and Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* 278, 82-87.

Damborsky J and Brezovsky J (2009) Computational tools for designing and engineering biocatalysts. *Curr Opin Chem Biol* 13, 26-34.

Daniel RM and Danson MJ (2010) A new understanding of how temperature affects the catalytic activity of enzymes. *Trends Biochem Sci* 35, 584-591.

Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, Isern NG et al. (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J Mol Biol* 366, 1209-1221.

Dzeja PP, Bast P, Pucar D, Wieringa B and Terzic A (2007) Defective metabolic signaling in adenylate kinase AK1 gene knock-out hearts compromises post-ischemic coronary reflow. *J Biol Chem* 282, 31366-31372.

Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B et al. (2004) Rational engineering of enzyme stability. *J Biotechnol* 113, 105-120.

Eijsink VG, Gaseidnes S, Borchert TV and van den Burg B (2005) Directed evolution of enzyme stability. *Biomol Eng* 22, 21-30.

Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D* 60, 2126-2132.

Glaser P, Presecan E, Delepierre M, Surewicz WK, Mantsch HH, Barzu O et al. (1992) Zinc, a novel structural element found in the family of bacterial adenylate kinases. *Biochemistry* 31, 3038-3043.

Howell SC, Inampudi KK, Bean DP and Wilson CJ (2014) Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure* 22, 218-229.

Jaenicke R (2000) Stability and stabilization of globular proteins in solution. *J Biotechnol* 79, 193-203.

John DM and Weeks KM (2000) van't Hoff enthalpies without baselines. *Protein Sci* 9, 1416-1419.

Karplus M (2011) Behind the folding funnel diagram. *Nat Chem Biol* 7, 401-404.

Khechinashvili NN, Janin J and Rodier F (1995) Thermodynamics of the temperature-induced unfolding of globular proteins. *Protein Sci* 4, 1315-1324.

Kiss G, Celebi-Olcum N, Moretti R, Baker D and Houk KN (2013) Computational enzyme design. *Angew Chem Int Ed Engl* 52, 5700-5725.

Korkegian A, Black ME, Baker D and Stoddard BL (2005) Computational thermostabilization of an enzyme. *Science* 308, 857-860.

Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL and Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364-1368.

Lane MD and Seelig B (2014) Advances in the directed evolution of proteins. *Curr Opin Chem Biol* 22, 129-136.

Lienhard GE and Secemski, II (1973) P1,P5-Di(adenosine-5')pentaphosphate, a potent multisubstrate inhibitor of adenylate kinase. *J Biol Chem* 248, 1121-1123.

Liszka MJ, Clark ME, Schneider E and Clark DS (2012) Nature versus nurture: developing enzymes that function under extreme conditions. *Annu Rev Chem Biomol Eng* 3, 77-102.

Mallamace F, Corsaro C, Mallamace D, Vasi S, Vasi C, Baglioni P et al. (2016) Energy landscape in protein folding and unfolding. *Proc Natl Acad Sci U S A* 113, 3159-3163.

McConnell AD, Spasojevich V, Macomber JL, Krapf IP, Chen A, Sheffer JC et al. (2013) An integrated approach to extreme thermostabilization and affinity maturation of an antibody. *Protein Eng Des Sel* 26, 151-164.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

Miller C, Davlieva M, Wilson C, White KI, Counago R, Wu G et al. (2010) Experimental evolution of adenylate kinase reveals contrasting strategies toward protein thermostability. *Biophys J* 99, 887-896.

Mirsky AE and Pauling L (1936) On the Structure of Native, Denatured, and Coagulated Proteins. *Proc Natl Acad Sci U S A* 22, 439-447.

Muller CW, Schlauderer GJ, Reinstein J and Schulz GE (1996) Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 4, 147-156.

Murshudov GN, Vagin AA and Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst D* 53, 240-255.

Oobatake M and Ooi T (1993) Hydration and heat stability effects on protein unfolding. *Prog Biophys Mol Biol* 59, 237-284.

Ooi T (1994) Thermodynamics of protein folding: effects of hydration and electrostatic interactions. *Adv Biophys* 30, 105-154.

Otwinowski Z and Minor W (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276, 307-326.

Packer MS and Liu DR (2015) Methods for the directed evolution of proteins. *Nat Rev Genet* 16, 379-394.

Pantazes RJ, Grisewood MJ and Maranas CD (2011) Recent advances in computational protein design. *Curr Opin Struct Biol* 21, 467-472.

Pauling L, Corey RB and Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* 37, 205-211.

Razvi A and Scholtz JM (2006) Lessons in stability from thermophilic proteins. *Protein Sci* 15, 1569-1578.

Reinstein J, Vetter IR, Schlichting I, Rosch P, Wittinghofer A and Goody RS (1990) Fluorescence and NMR investigations on the ligand binding properties of adenylate kinases. *Biochemistry* 29, 7440-7450.

Rodriguez R, Chinea G, Lopez N, Pons T and Vriend G (1998) Homology modeling, model and software evaluation: three related resources. *Bioinformatics* 14, 523-528.

Schmid FX (2011) Lessons about protein stability from in vitro selections. *Chembiochem* 12, 1501-1507.

Schoemaker HE, Mink D and Wubbolts MG (2003) Dispelling the myths-biocatalysis in industrial synthesis. *Science* 299, 1694-1697.

Schulz GE, Muller CW and Diederichs K (1990) Induced-fit movements in adenylate kinases. *J Mol Biol* 213, 627-630.

Shirke AN, Basore D, Butterfoss GL, Bonneau R, Bystroff C and Gross RA (2016) Toward rational thermostabilization of Aspergillus oryzae cutinase: Insights into catalytic and structural stability. *Proteins* 84, 60-72.

Sidney CP and Herman KM (1943) The role of myokinase in transphosphorylations *J Biol Chem* 148, 117-126.

Sinev MA, Sineva EV, Ittah V and Haas E (1996) Domain closure in adenylate kinase. *Biochemistry* 35, 6425-6437.

Socha RD and Tokuriki N (2013) Modulating protein stability - directed evolution strategies for improved protein function. *FEBS J* 280, 5582-5595.

Thach TT and Lee S (2014) New crystal structures of adenylate kinase from Streptococcus pneumoniae D39 in two conformations. *Acta Cryst F* 70, 1468-1471.

Thach TT, Luong TT, Lee S and Rhee DK (2014) Adenylate kinase from Streptococcus pneumoniae is essential for growth through its catalytic activity. *FEBS Open Bio* 4, 672-682.

Traxlmayr MW and Obinger C (2012) Directed evolution of proteins for increased stability and expression using yeast display. *Arch Biochem Biophys* 526, 174-180.

Turner P, Mamo G and Karlsson EN (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb Cell Fact* 6, 9.

Unsworth LD, van der Oost J and Koutsopoulos S (2007) Hyperthermophilic enzymes-stability, activity and implementation strategies for high temperature applications. *FEBS J* 274, 4044-4056.

Vieille C and Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65, 1-43.

Vonrhein C, Bonisch H, Schafer G and Schulz GE (1998) The structure of a trimeric archaeal adenylate kinase. *J Mol Biol* 282, 167-179.

Wintrode PL and Arnold FH (2000) Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv Protein Chem* 55, 161-225.

Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ and Kern D (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol* 11, 945-949.

Wolynes PG, Onuchic JN and Thirumalai D (1995) Navigating the folding routes. *Science* 267, 1619-1620.

Supplementary Materials

Figure S1. Affinity chromatogram of AKlse2

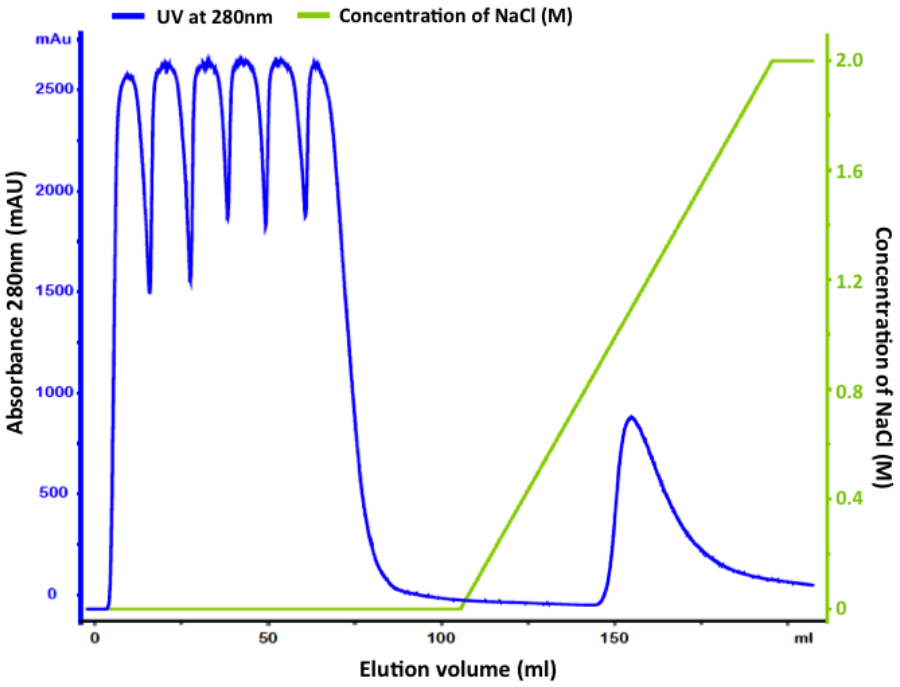


Figure S2. Size exclusion chromatogram of AKlse2

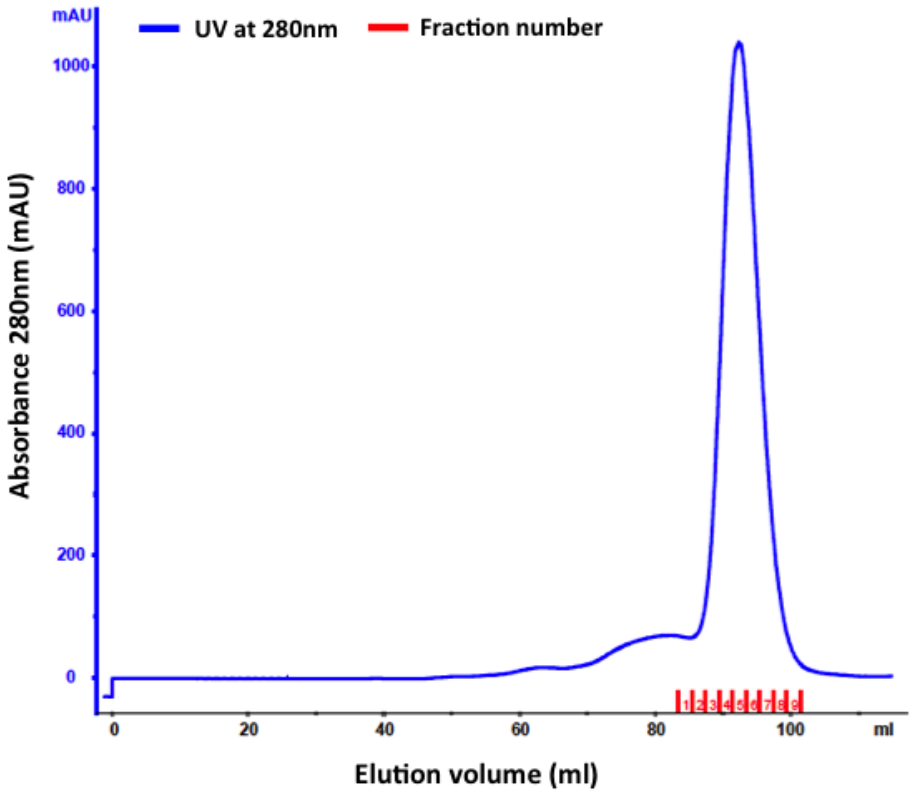
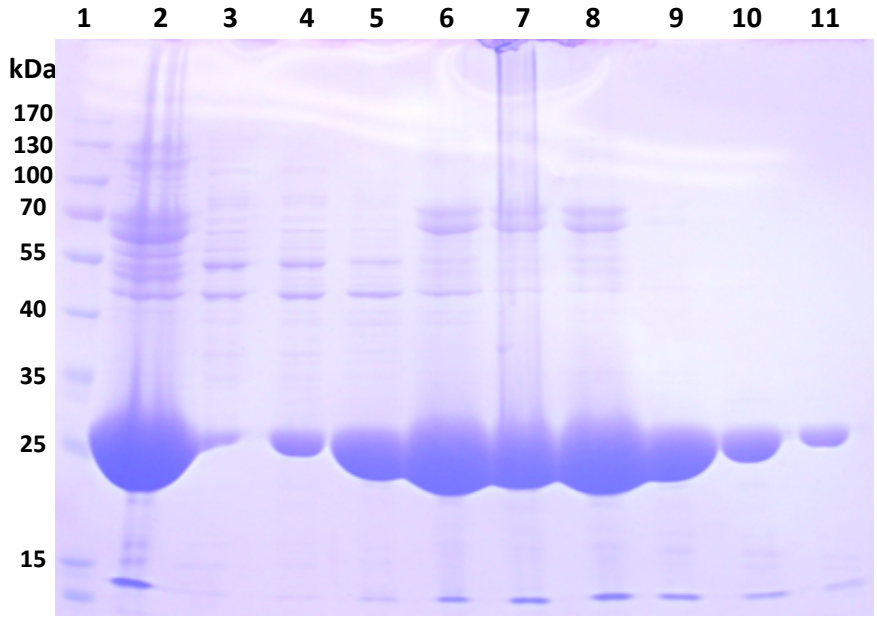
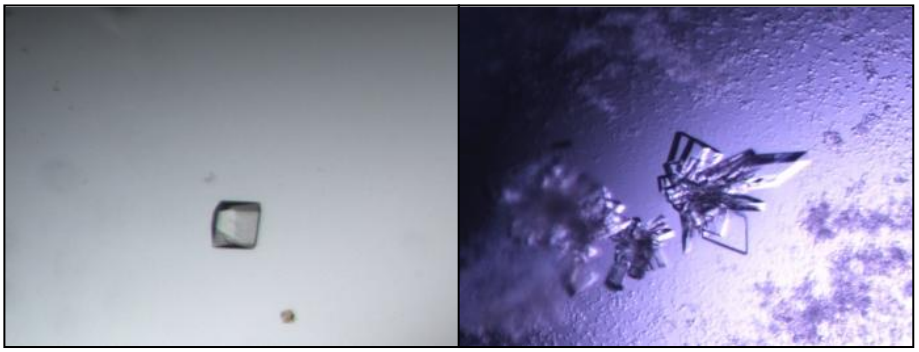


Figure S3. SDS-PAGE results of size exclusion chromatogram fractions



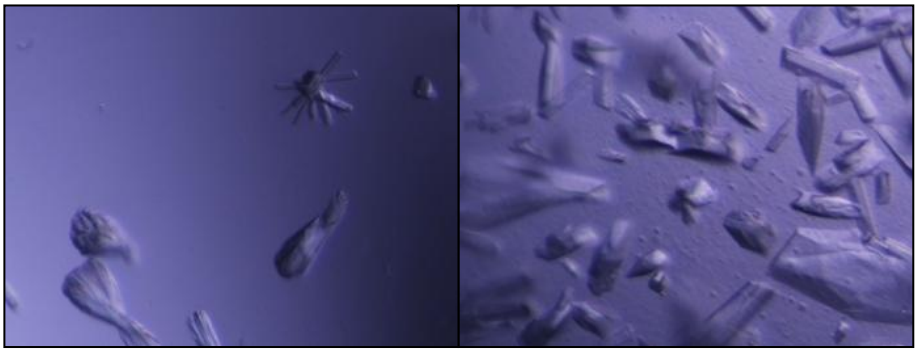
Lane 1: protein marker (Fermentas), lane 2: sample load,
lane 3: fraction #1 shown in Fig. S2, lane 4: fraction #2 shown in Fig. S2,
lane 5: fraction #3 shown in Fig. S2, lane 6: fraction #4 shown in Fig. S2,
lane 7: fraction #5 shown in Fig. S2, lane 8: fraction #6 shown in Fig. S2,
lane 9: fraction #7 shown in Fig. S2, lane 10: fraction #8 shown in Fig. S2,
lane 11: fraction #9 shown in Fig. S2

Figure S4. Crystals of AKv8 and AKv18



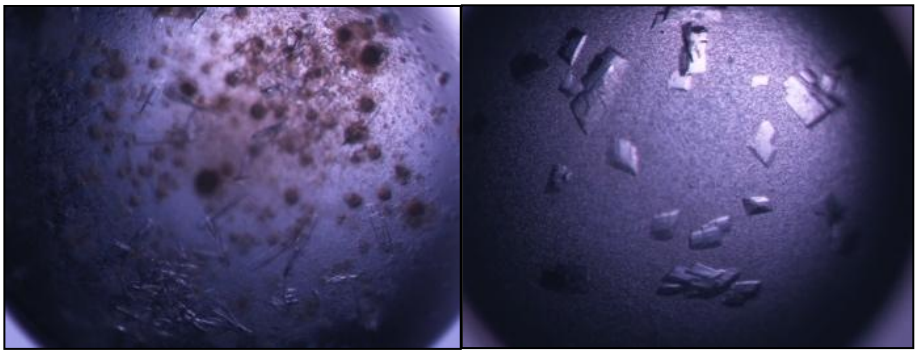
AKv8 **AKv18**

Figure S5. Crystals of AKlse2 and AKlse4



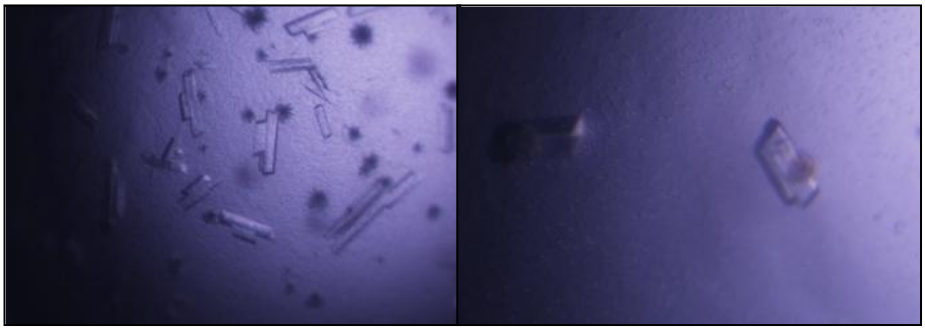
AKlse2 **AKlse4**

Figure S6. Crystals of AKlse4m1 and AKlse4m2



AKlse4m1 **AKlse4m2**

Figure S7. Crystals of AKlse5 and AKlse6



AKlse5 **AKlse6**

Abstract in Korean

서울대학교 대학원

농생명공학부 응용생명화학전공

문소진

열에 안정한 단백질은 반응의 효율을 증가시키는 측면과 효소 저장 및 운반에 이용되는 비용 절감 등의 산업적, 실험적 환경에 필수적이다. 따라서, 이를 위한 단백질 재 디자인 방법들이 오랜 기간동안 다양한 시도들이 연구되고 있다. 그럼에도 불구하고, 지금까지 알려진 방법들은 이의 결과들이 예측이 힘들고, 재디자인 결과가 열 안정성에 크게 영향을 주지 않는 문제점을 가지고 있다. 본 연구자는, 여러가지 열 안정성에 관련된 기술들의 효과를 테스트 하였다. 이를 바탕으로, 개별적이고 통합적인 방법으로 단백질을 재디자인하여 어떠한 메커니즘이 단백질 열 안정성에 영향을 미치는 역할을 하는지 제공한다. 본 연구의 모델로서, 중온성 아데닐레이트 카이네이즈 (AKmeso) 에 바이오인포메틱스 방법인 local structural entropy 최적화, 삼차원적 구조를 바탕으로한 돌연변이 생성, 분자생물학적인 진화법을 적용하였다. 이 방법들을 개별적과 통합적으로 적용하여, 아데닐레이트 카이네이즈 돌연변이 (AK variants) 를 생성하고, 각 AK variant 의 열 안정성을 측정하였다. 재 디자인된 AK variants 는 야행형인 AKmeso 보다 향상된 열 안정성을 가지며, 높은 온도에서 효소 활성을 나타냈다. 호열성인 아데닐레이트 카이네이즈와 유사한 열 안정성을 가지는 AK variant 인 AK v18 을 포함하여, 디자인에 관련된 대표적인 AK variants 의 단백질 결정을 이용한 삼차원적 구조 규명 하였다. 이를 통하여, AK variants 의 단백질 구조를 확인하였고, AK variant 에 적용된 열 안정성 관련 방법의 효과를 분석하였다. 이러한 실험의 결과를 통하여, 아데닐레이트 카이네이즈의 열 안정성과 관련된 메커니즘의 역할를 확인할 수 있었다. 이를 바탕으로 열에 안정한 단백질을 디자인 함에

있어서, 여러 방법들을 개별적 및 통합적으로 이용하였을 때, 보다 효과적으로 적용시킬 수 있는 방법을 제시하고 있다.

주요어 : 구조적 엔트로피, 단백질 구조, 단백질 공학, 돌연변이 유발, 아데닐레이트 카이네이즈, 열 안정성

학 번 : 2010-23449