



Article

# Rational Design of Adenylate Kinase Thermostability through Coevolution and Sequence Divergence Analysis

Jian Chang <sup>1</sup>, Chengxin Zhang <sup>2,3</sup>, Huaqiang Cheng <sup>1</sup> and Yan-Wen Tan <sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Surface Physics, Multiscale Research Institute of Complex Systems, Department of Physics, Fudan University, Shanghai 200433, China; 13110190056@fudan.edu.cn (J.C.); huaqiangch@gmail.com (H.C.)

<sup>2</sup> School of Life Science, Fudan University, Shanghai 200433, China; zcx@umich.edu

<sup>3</sup> Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

\* Correspondence: ywtan@fudan.edu.cn

**Abstract:** Protein engineering is actively pursued in industrial and laboratory settings for high thermostability. Among the many protein engineering methods, rational design by bioinformatics provides theoretical guidance without time-consuming experimental screenings. However, most rational design methods either rely on protein tertiary structure information or have limited accuracies. We proposed a primary-sequence-based algorithm for increasing the heat resistance of a protein while maintaining its functions. Using adenylate kinase (ADK) family as a model system, this method identified a series of amino acid sites closely related to thermostability. Single- and double-point mutants constructed based on this method increase the thermal denaturation temperature of the mesophilic *Escherichia coli* (*E. coli*) ADK by 5.5 and 8.3 °C, respectively, while preserving most of the catalytic function at ambient temperatures. Additionally, the constructed mutants have improved enzymatic activity at higher temperature.

**Keywords:** protein engineering; thermal stability; primary sequence; sequence profile comparison; adenylate kinase



**Citation:** Chang, J.; Zhang, C.; Cheng, H.; Tan, Y.-W. Rational Design of Adenylate Kinase Thermostability through Coevolution and Sequence Divergence Analysis. *Int. J. Mol. Sci.* **2021**, *22*, 2768. <https://doi.org/10.3390/ijms22052768>

Academic Editor: Sergio F. Sousa

Received: 1 February 2021

Accepted: 5 March 2021

Published: 9 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Protein thermostability improvement is one of the most essential and sought after protein engineering aspects [1]. Increasing the thermostability of an enzyme can boost the total turnover number and extend the shelf life of derived products [2]; thermostable proteins in drugs or vaccines can reduce the cost of storage and transportation [3]; high thermostability can help crystallize unstable proteins and thus determine their structures [4]. Directed evolution and rational design are two approaches for protein engineering. Thermal stability improvement is one of the many tasks that directed evolution can readily accomplish. However, directed evolution can be laborious and costly [5]. By contrast, rational design shows the extent to which we understand the functioning principles of proteins. When applied properly, rational design can be time-saving and lower the barrier for thermal stability engineering.

Most of the stability rational design methods require protein tertiary structural information. Structural information can be used to calculate surface charge distribution, backbone torsion, spatial distance between residues, side chain volumes, and solvent accessibility for improved thermostability [6–8]. When the protein structure is unavailable, options for effective protein engineering are limited. Among few sequence-based stability engineering methods, the improved configurational entropy (ICE) algorithm required only primary sequences of homologous proteins for thermostability improvement design through minimizing the local structural entropy of the secondary structure [9,10]. However, because of the lack of consideration for long-range side-chain contacts of residues that are

far apart along the primary sequence, ICE might stabilize local structures at the cost of interrupting other noncovalent interactions such as favorable ion pairs, hydrogen bonds, or hydrophobic interactions that link distant parts of the protein backbone [11]. Consensus method hypothesized that residues contributing to catalysis or stability were more likely to be conserved during the evolution [1]. This approach was operated by replacing targets' residue with the most frequent one among homologues. It has the advantage of screening for improved thermal stability in a single step with just sequence information. Unfortunately, after multiple residue substitutions, consensus proteins may not retain their biological activity or be unable to fold accurately [12]. In the consensus method, the correlation among residue pairs are disregarded, especially for poorly-conserved sites [1]. In addition, the consensus method usually ignores the difference between mesophiles and thermophiles, which may lead to biased results. For example, a consensus method may favor an amino acid type due to its high frequency among mesophilic proteins in the family, even though the residue type that actually stabilizes the protein may be a different amino acid type found in a smaller number of thermophiles.

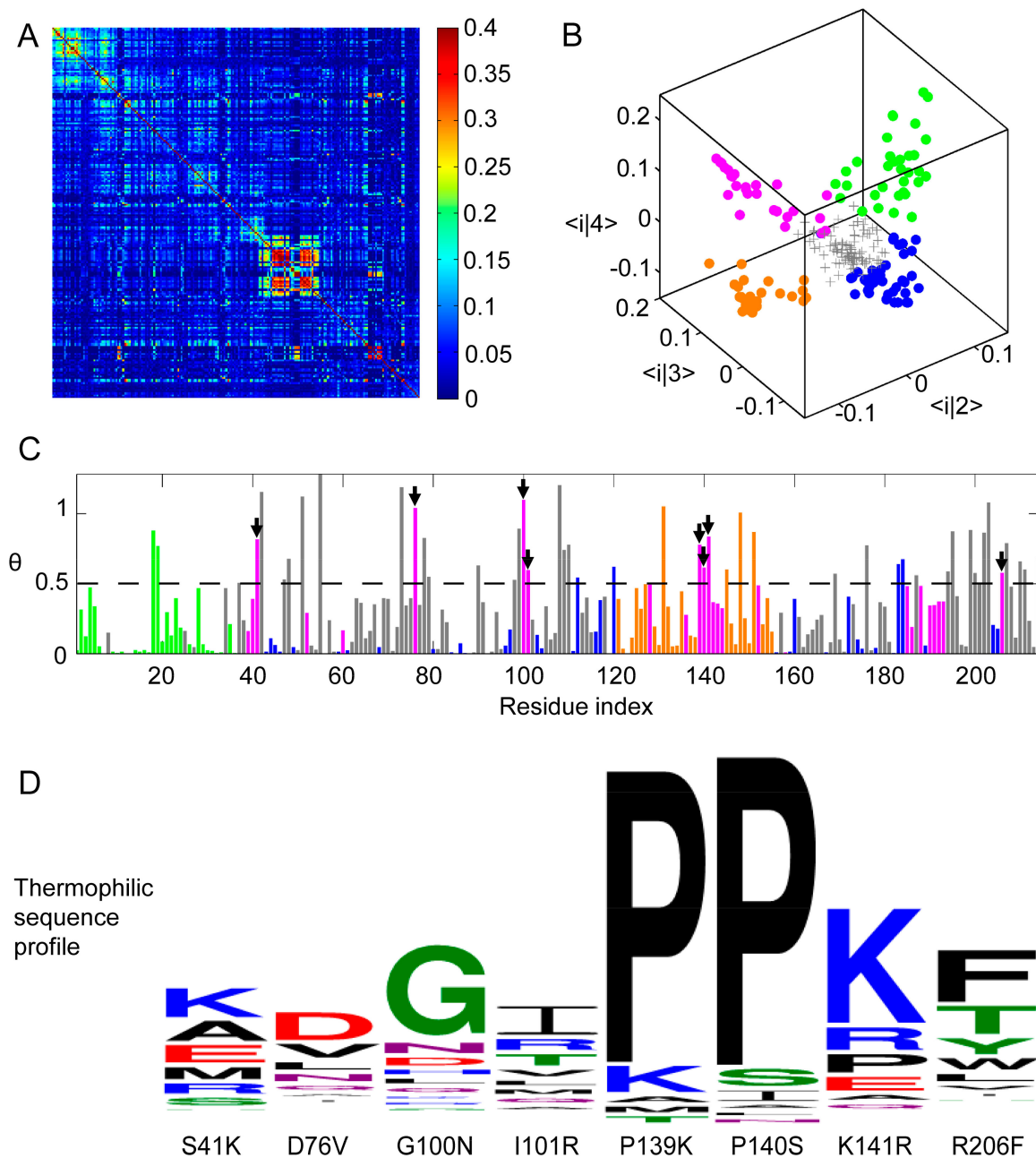
Nature has presented abundant evolutionary traces with potential clues for thermal stability engineering. Thermophiles living in extremely hot environments have proteins that are largely homologous to that of mesophiles. These homologous proteins execute the same function as their mesophilic counterparts, although with different sequences and considerably higher working temperatures. However, using thermophilic homolog sequences directly for thermostability improvement would result in poor activity at ambient temperature.

In this study, we designed a novel primary-sequence-based bioinformatics method to generate thermostable variants from mesophilic enzyme, *E. coli* ADK. ADK is a phosphotransferase enzyme involved in catalyzing the interconversion of adenine nucleotides—adenosine triphosphate (ATP), adenosine monophosphate (AMP), and adenosine diphosphate (ADP). A number of previous studies have provided various results for thermostability improvement of this model protein [7,9,13–15]. Therefore, we had chosen mesophilic *E. coli* ADK as the engineering target to examine our method. Our primary-sequence-based bioinformatics method could identify residues that are crucial for structural stability without drastically changing their functionality. Firstly, we applied the residue correlation analysis (RCA) coevolution algorithm to analyze evolutionary information embedded in multiple sequence alignment (MSA) to detect thermostability-related residues [16]. Secondly, protein sector method proposed by Halabi et al. was applied to categorize residues according to their intrinsic correlations [17]. Then, amino acid compositions of thermophile and mesophile were compared to identify the sector that are related to structural stability. Lastly, residues within the chosen sector were mutated to amino acids suggested by the thermophilic sequence profile. In this work, single- and double- point mutations by replacing original residues with amino acids having different chemical properties were characterized and achieve up to 5.5 and 8.3 °C higher melting temperature ( $T_m$ ) than that of the wild type, which successfully validated our method.

## 2. Results

### 2.1. Protein Sectors Derived from Amino Acid Sequences

The foundation of our method is an MSA of the protein of interest. Through sequence alignment, a  $9203 \times 214$  matrix of amino acid distribution was obtained within the ADK protein family, where 9203 and 214 are the number of homologous sequences and the number of amino acid positions in *E. coli* ADK, respectively. The details about the MSA of ADK family were displayed in the Supplementary Text (Text S1). This MSA contained the information of residue transfer at corresponding sites in species evolution and was used as the input of the RCA, which outputs a  $214 \times 214$  matrix for the evolutionary coupling strength of all residue pairs (Figure 1A).



**Figure 1.** Residue Correlation Analysis (RCA) of Protein Sectors in ADK Family. **(A)** Heat map representation of RCA matrix  $r_{ij}$  for MSA of ADK.  $X$ - and  $y$ -axis coordinates both correspond to residue indices of ADK from *E. coli*, which has 214 residues. **(B)** Three dimensional scatter plot of the 214 residues in the space formed by the three eigenvectors of the second, the third, and the fourth largest eigenvectors. Each data point represented one position. After eliminating randomized background residues (gray), the rest of the positions could be clustered into four sectors, colored green, blue, orange, and magenta. **(C)** The relative entropy angle  $\theta$  of thermophilic and mesophilic sequence profile at each position. Bars were colored by sectors. Residues whose angles were larger than 0.5 in the magenta sector were selected for further mutation (marked by arrow heads). **(D)** Amino acid distribution of thermophilic sequences at the chosen mutation positions. The first series of single-point mutations were S41K, D76V, G100N, I101R, P139K, P140S, K141R, and R206F.

To reduce the noise in the RCA, eigen-decomposition is performed to generate 214 eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{214}$  (ranked in the descending order of value). Among them,  $\lambda_1, \lambda_2, \lambda_3$ , and  $\lambda_4$  were considerably larger than the remaining eigenvalues. According to the protein sector method [17], the eigen-decomposition components with eigenvalues lower

than those generated from a randomized RCA can be attributed to a limited sampling of sequences (statistical noise). Furthermore, the largest component,  $\lambda_1$ , arises from the phylogenetic effect within the ADK family (historical noise). After removing all noises, the raw  $r_{ij}$  value for residue pair  $i$  and  $j$  can be approximated as the sum of three eigen-decomposition components related to  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ . The mode 2–4 in eigen-decomposition contain information critical for the structure and function of ADK and can be used to approximate the global correlation matrix between the residues of the ADK family (See Supplemental Figure S1A,B).

$$r_{ij} \approx \sum_{k=2}^4 v_k \lambda_k v_k^T \quad (1)$$

According to protein sector method [17], in the three-dimensional eigenspace representation formed by three modes, residues with similar characteristics will gather into clusters. In our case, all 214 ADK residues can be plotted in this eigenspace (Figure 1B). After eliminating residues (plotted as the gray plus signs) with eigenvalues smaller than the boundary set by randomized RCA eigenvalues, the remaining residues can be classified into four clusters using k-mean clustering (Figure 1B), which are plotted respectively in green, orange, blue, and magenta. Among them, the green and orange sectors reveal obvious continuity in primary sequence while residues in the blue and magenta sectors locate on the whole sequence in a dispersed manner (Figure 1C).

## 2.2. Thermostability-Related Sector Identification and Mutant Design

To determine sites that should be mutated, sequence entropy analysis was applied using the subset of 256 and 3535 homologous ADK protein sequences which were confirmed to be encoded by thermophilic and mesophilic organisms according to credible published literatures or databases as listed in the Supplementary File. The remaining subset of 5412 out of all 9203 sequences were not associated with information on the optimal growth temperatures of their host organisms and therefore excluded from the sequence entropy analysis. The details about how to classify the thermophilic and mesophilic species were exhibited in the Supplementary Text (Text S2). For each position, a pair of relative entropy vectors can be obtained, which contained the information of amino acid constitutions of thermophilic and mesophilic proteins. The relative entropy angle  $\theta$  represents the difference between the two sequence profiles. Intuitively, the greater the  $\theta$ , the more divergent was the amino acid composition at that position of two profiles. In this study, sites with low conservation scores but high  $\theta$  values were chosen for further mutation because highly conserved positions were crucial for preserving molecular function, whereas we hypothesized that sites critical for stabilizing the molecular structure were highly differentiated during adaptation to various environments.

Four sectors were classified based on an RCA, but not all of them might be correlated with ADK thermal stability. To determine which sector might be the most relevant to protein's stability, we combined results obtained from the RCA and the relative entropy angle  $\theta$  from sequence entropy analysis and finally chose the magenta sector to be mutated for further thermal stability improvement. The reason is the following: among all four sectors, the magenta one had the maximum average  $\theta$  value (0.4727) and the blue one had the minimum average  $\theta$  value (0.1342). Furthermore, both green and orange sectors were located closely along the one-dimensional sequence, suggesting that they could form connected domain or subdomain structures and play essential roles in the catalytic function. To examine how mutations in green, blue, and orange sectors influence the enzymatic function and protein stability of ADK, we have performed control experiments on the mutants from these three sectors. We picked two mutations from each sector according to the same procedure as in the magenta sector. Six mutants—Q18K, F19R, S43T, M96L, R131A, and V148K were generated and characterized. Their corresponding assay and circular dichroism (CD) experiment results were displayed in the Supplementary Materials (See Figure S2), which indicated that the blue sector was highly conserved for ADK's enzymatic function and had little to do with structure stability while the green and orange

sectors might be both responsible for the enzymatic activity and protein thermostability. Therefore, they were excluded from subsequent considerations.

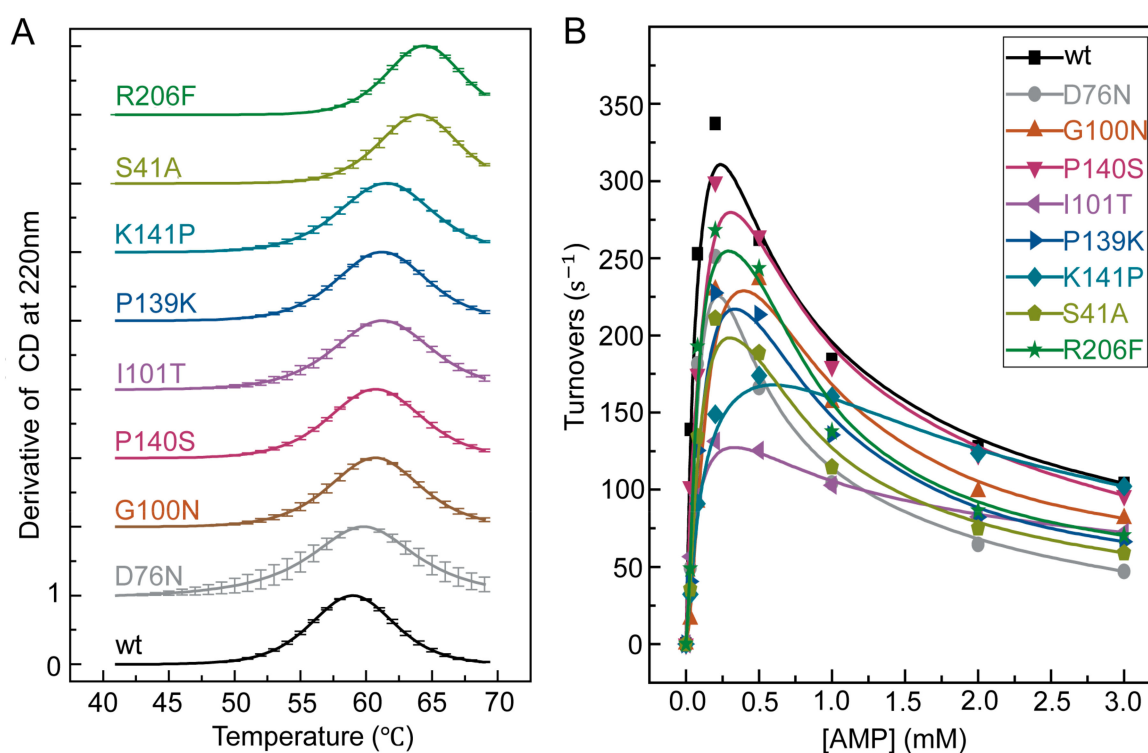
Combining the RCA and sequence entropy analysis, residues in the magenta sector with the eight highest relative entropy angle  $\theta$  (least conserved) were selected for mutations (Figure 1C). Up to now, we have already recognized those protein sites responsible for thermal stability but it remained unclear which amino acids should replace the original ones. In our initial attempt, residues at those chosen sites were mutated to the corresponding amino acids that were the most frequent in thermophile ADK MSA (Figure 1D). If the same amino acid was the most frequent in both thermophile and mesophile ADK, we moved to the second most frequent one. Eight mutants of *E. coli* ADK (ADK wt) were thus generated: S41K, D76V, G100N, I101R, P139K, P140S, K141R, and R206F.

### 2.3. Thermostability Characterization of Mutants

We examined the thermostability variations of the eight mutants with CD T-melt measurements. The results showed that, compared with the  $T_m$  of the wild type, four of the eight mutants—G100N, P139K, P140S, and R206F—had  $T_m$  increase of 1.8 °C, 2.3 °C, 1.8 °C, and 5.5 °C, respectively (Table 1). However, the others, S41K, D76V, I101R, and K141R, had their  $T_m$  decrease by 0.7–5.4 °C and were therefore mutated to the next most common residues in the thermophile: S41A, D76N, I101T, and K141P. What's more, compared with the first four mutants (S41K, D76V, I101R, K141R), all new residues in second mutations (S41A, D76N, I101T, K141P) had side chains with apparently different properties. For example, lysine (K) in S41K (first mutation) was positively charged, whereas alanine (A) in S41A (second mutation) held a hydrophobic side chain. CD results of S41A, D76N, I101T, and K141P exhibited consistent  $T_m$  rises: 5.1 °C, 0.9 °C, 2.3 °C, and 2.6 °C (Table 1). Considering all successful mutation cases (G100N, P139K, P140S, R206F, S41A, D76N, I101T, K141P), we suggested that the strategy for choosing mutation amino acid might be by replacing the original residues by amino acids with apparently different properties. However, this condition was optional rather than mandatory. In our amino acid selection strategy, the amino acid distribution of thermophilic sequences was the major consideration. In our primary-sequence-based method, the lack of 3D protein structural information limited our recognition on whether selection of particular amino acid type would work well or not. In future work, our method might be combined with 3D structural information to improve the accuracy of single-point mutation. In summary, we have obtained thermal stable ADK variants for all eight sites as the first step for the proof of principle for our method (Figure 2A).

**Table 1.**  $T_m$  of ADK wt and all mutants, ranked in ascending order.

Protein	Melting Temperature (°C)
ADK wt	59.4 ± 0.1
D76N	60.3 ± 0.1
G100N	61.2 ± 0.1
P140S	61.2 ± 0.1
I101T	61.7 ± 0.1
P139K	61.7 ± 0.0
K141P	62.0 ± 0.1
S41A	64.5 ± 0.2
R206F	64.9 ± 0.1
R206F/P140S	65.3 ± 0.2
R206F/G100N	65.9 ± 0.2
R206F/P139K	66.4 ± 0.1
R206F/S41A	67.7 ± 0.3



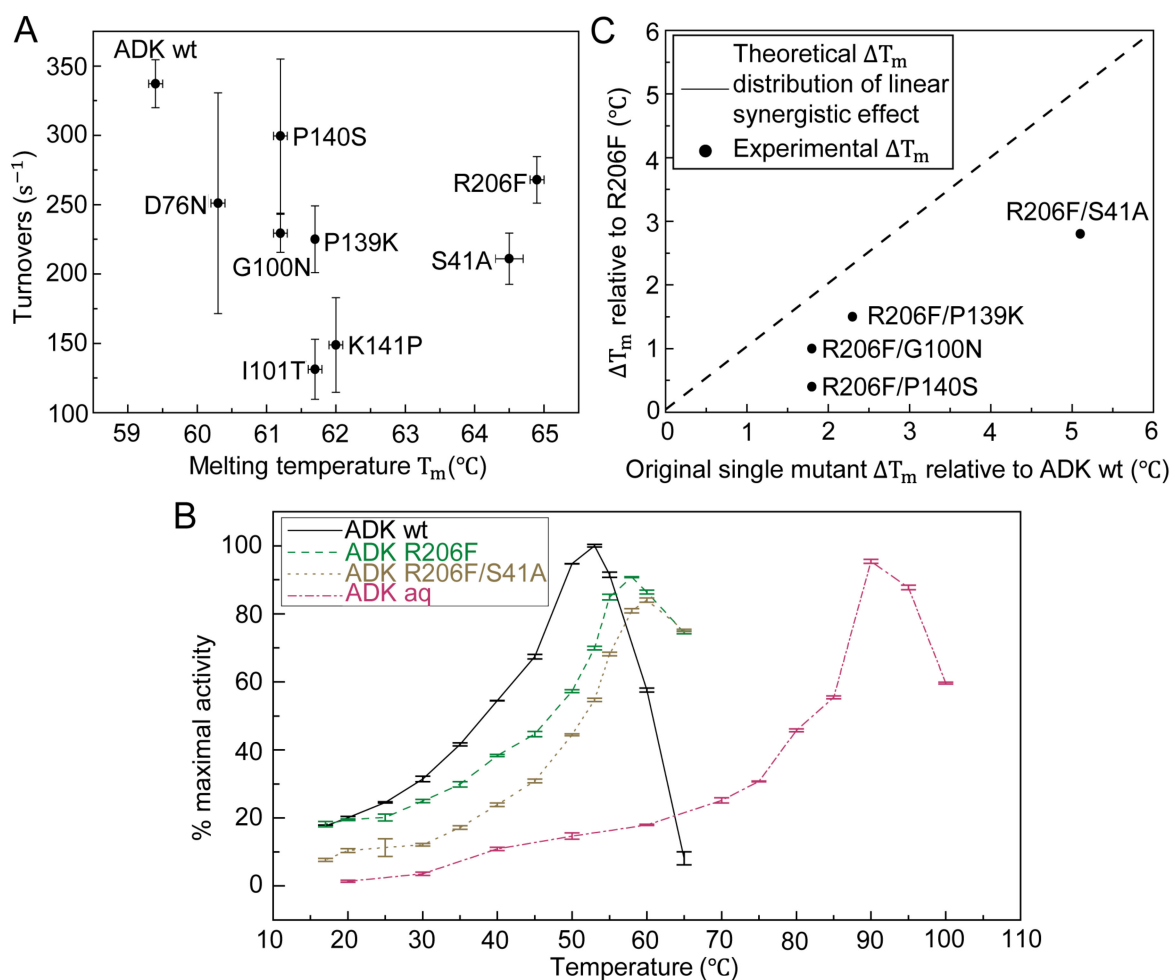
**Figure 2.** Thermal stability and enzymatic activity of single-point mutants characterized by circular dichroism (CD) spectroscopy and room temperature enzymatic activity assay. (A) CD differential signal at 222nm of wild type and mutants S41A, D76N, G100N, I101T, P139K, P140S, K141A, K141P, and R206F from 40 °C to 70 °C. All data were normalized to a scale of zero to one. To be distinguished from each other, each curve was offset by one along the  $y$ -axis. (B) Forward enzymatic activity of wild type and eight mutants at 25 °C. [ATP] = 1 mM, [magnesium acetate ( $\text{MgOAc}_2$ )] = 2 mM.

As a bi-substrate enzyme involved in catalyzing the reversible phosphoryl transfer reaction:  $\text{Mg}^{2+} \cdot \text{ATP} + \text{AMP} \xrightleftharpoons{\text{ADK}} \text{Mg}^{2+} \cdot \text{ADP} + \text{ADP}$ , ADK follows a random bi bi reaction mechanism [18]. However, for *E. coli* ADK, the reaction curve of varying AMP at constant ATP does not follow Michaelis–Menten kinetics. Previously, such a trend has been attributed to an AMP inhibition effect. However, in our earlier study, we have found that it was originated from a magnesium activation effect [19]. Since the overall enzymatic model is rather complicated, we will use the forward turnover rate at a representative ATP, AMP, and  $\text{Mg}^{2+}$  concentrations instead of the commonly used catalytic rate,  $k_{\text{cat}}$ , to represent the engineered ADKs' catalytic function in the current study.

In the enzymatic forward activity assay at room temperature (Figure 2B), these thermostable mutants had turnover rates at [ATP] = [AMP] = 1 mM that were 56–98% of that of the wild type among which the most stable variant, R206F, retained 75% of the activity (Table S1). For the mutants with the lowest turnover rates, I101T and K141P, they seemed to have lost their magnesium activation effect [19] instead of losing the overall enzymatic activities. Since the optimal reaction temperatures of mutants are shifted as shown in later, the room temperature reaction rate of the mutants are not fully reflective of their activities. Nonetheless, the preservation of room temperature activity shown in Figure 2B is still an important metric to confirm the success of enzyme design, as our goal is to improve the thermal stability of the enzyme without significantly jeopardizing its activity at its original working temperature.

To examine whether the stabilizing effect of single-point mutation is additive, we then constructed four double mutants based on the template R206F because it had the maximum  $T_m$  among all single-point mutants. S41A, G100N, P139K, and P140S were chosen as the second mutation sites since they not only had relatively high  $T_m$ s but also preserved better enzymatic activities (Figure 3A). Finally, four double mutants R206F/S41A, R206F/G100N,

R206F/P139K, and R206F/P140S were constructed and characterized. Compared with R206F, all four mutants presented increased  $T_m$  but the increments are lower than the linear summation of  $\Delta T_m$ s from two corresponding single mutants (Figure 3C). R206F/S41A was the most prominent double-mutation variant with the highest  $T_m$  rise of 8.3 °C (Table 1). However, almost all double mutants showed further decrease in enzymatic activities compared with respective single mutants (Figure S3A–D).



**Figure 3.** Correlation between melting temperature and enzymatic activity of wt and selected mutants of ADK. (A) Forward reaction activities (at 0.2 mM AMP and 1 mM ATP) versus melting temperature of ADK wild type and single mutants. (B) Temperature dependence of enzyme activity for ADK from *E. coli*, single mutant R206F, double mutant R206F/S41A and ADK from *Aquifex aeolicus* in the direction of ADP formation. The absolute value of 100% maximal activity was calculated as  $458 \pm 4 s^{-1}$ . (C) Four double mutants, R206F/S41A, R206F/G100N, R206F/P139K, R206F/P140S, all displayed a moderate  $T_m$  increment lower than the sum of respective  $T_m$  rise from two related individual mutants. The synergistic effect lowered as the  $\Delta T_m$  sum increased. The dash line indicated the theoretical  $\Delta T_m$  distribution of double mutants compared to R206F provided that the synergistic effect was linear.

Although the aforementioned experiments separately characterized the catalytic activity at ambient temperature and thermostability, we shall investigate whether and how the activity of our mutants differs from naturally occurring ADK at medium to high temperature. To this end, we performed temperature-dependent enzymatic activity experiments shown in Figure 3B, where the turnover rate profiles at  $[ATP] = 2.5$  mM and  $[AMP] = 2.5$  mM of ADK wt and *Aquifex aeolicus* ADK (ADK aq) showed that they had the maximum activity at approximately 50 °C and 90 °C, respectively. To prevent inaccurate enzyme activity measurement due to insufficient substrates, we increased the concentrations of ATP and AMP to 2.5 mM. The result was consistent with the results from a previous study [13]. At

the industrial level, most enzymes worked at high temperatures (approximately 60 °C) where both ADK wt and ADK aq would lose the majority of their enzymatic function [20]. In our design, two selected variants, the most stable single mutant R206F and double mutant R206F/S41A, showed shifted peak catalytic activities at elevated temperatures (58–65 °C) compared with mesophilic and thermophilic natural enzymes, which validates our thermal stability modification.

### 3. Discussion

In this work, we used the RCA-based method to analyze evolutionary differences in the MSA of thermophiles and mesophiles of a model enzyme ADK. Following the concept of protein sectors [17], we isolated the group of residues that might be related to structural integrity and stability (See the magenta sector in Figure 4A). In the eigenspace of the RCA, we calculated the inner product of the vector formed through thermophilic and mesophilic MSAs to derive the relative entropy angle  $\theta$ . This result revealed residues that are the most different between thermophiles and mesophiles. We chose eight residues with the highest relative entropy angle  $\theta$  values, namely S41, D76, G100, I101, P139, P140, K141, and R206, and constructed eight thermally stable single-point mutants. The maximum  $T_m$  increment of favorable single-point mutations can be as high as 5.5 °C. Meanwhile, the enzymatic activity of the most stabilizing mutant retains 75% of the performance at [ATP] = [AMP] = 1 mM compared with the wild-type ADK.

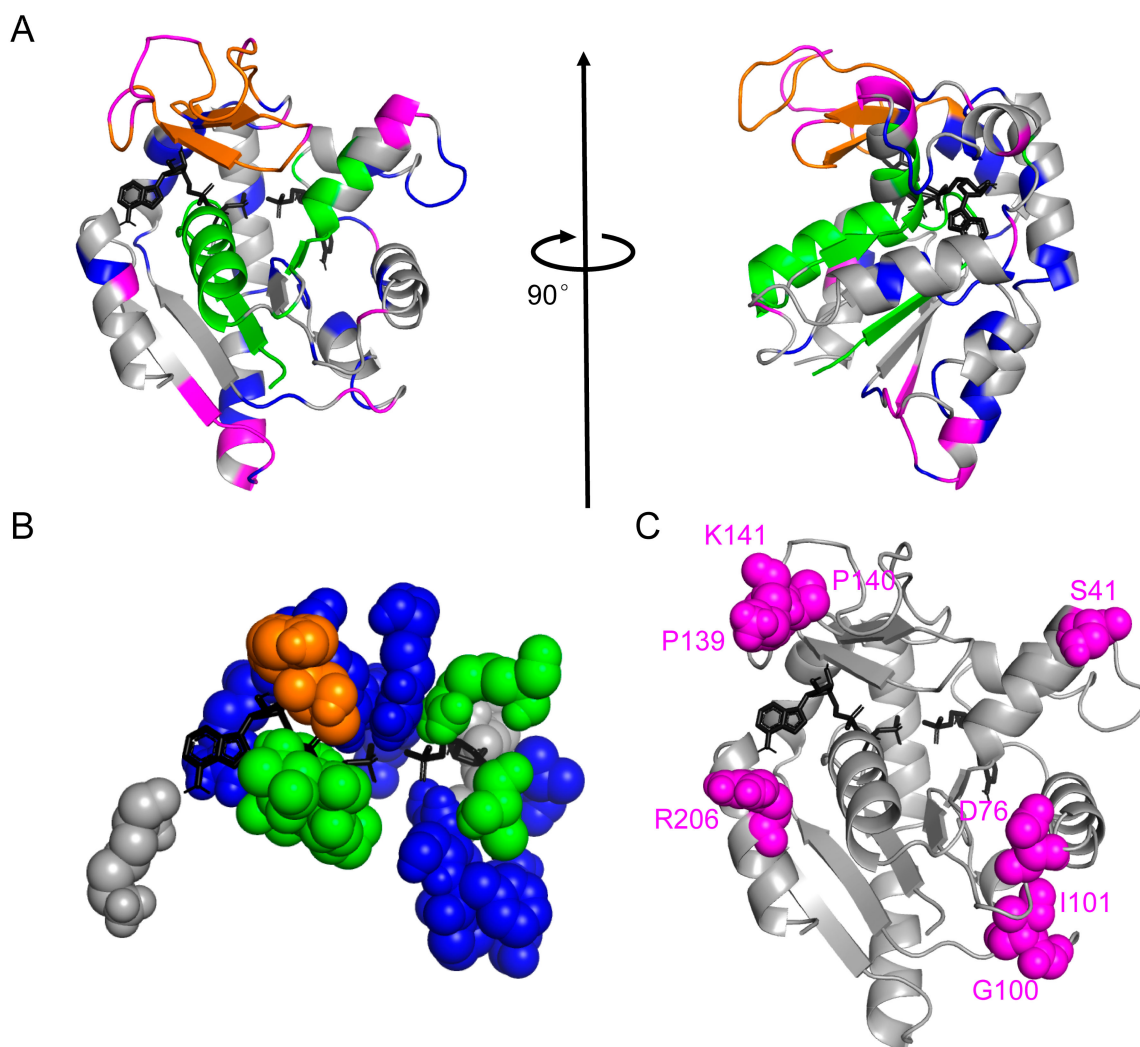
The prediction of the magenta sector using our primary-sequence-based method could be additionally justified by two structural aspects since the structure of ADK has been solved. First, only the magenta sector had no residues directly participating in substrate binding (see Figure 4B), which minimized the possibility of hindering the enzymatic function. In addition, results (Table S2) showed that except for K57, V132, and K200, residues in direct contact with substrates usually had a small average  $\theta$  value (0.0141), indicating their high evolutionary conservation. Second, the residues in the magenta sector were mainly located at the ends or turns of secondary structure elements (see Figure 4C). From this point of view, the extra stability obtained through our bioinformatics method might result from the stabilization effect of these residues on helices and sheets of ADK. It provides a clue to further optimize this algorithm by considering secondary structure information to and leading to a stabilized overall structure. Nonetheless, experimental structure determination will be necessary to exactly confirm the mechanisms behind the stabilization effect of these residues.

With thermostability elevation, enzymatic activity loss from mutants was also observed. This is a common issue in previous ADK design attempts [15,21,22]. In our case, we attributed this to that *E. coli* ADK requires a large magnitude of conformational change for substrates to enter into the reaction center and release of product [23]. Thus, mutations that stabilized the backbone also made it less flexible to accommodate the conformational change, hindering substrate turnover.

So far, our work revealed that appropriate single-point mutations could generate thermostable ADK mutants. However, multiple residue substitutions might be able to further improve the thermostability. For example, Euiyoung Bae redesigned the ADK from *Bacillus subtilis* by mutating as many as 26 residues and finally obtained a thermally stable variant with a  $\Delta T_m$  of 12.5 °C [9]. Moreover, different thermostability-improved mutants, which were derived from various stabilization methods, such as directed evolution, comparative study, or theoretical computation, could be combined together to form one construct with a higher  $T_m$  [21]. Taking all above cases into account, we went beyond single residue substitution and constructed a set of double mutants using the eight single-mutated variants. This was to examine, through our method, whether mutants with more point substitutions could achieve a greater thermal stability gain and show additive effects in  $T_m$ . As expected, four double mutants (R206F/S41A, R206F/G100N, R206F/P139K, and R206F/P140S) showed improved thermal stabilities compared with their original single-point mutants, which confirmed the synergistic effect of two stabilizing single mutations.



In addition, all double mutants displayed a  $T_m$  increment lower than the summation of the respective  $\Delta T_m$  s from the two original individual mutants. Furthermore, our results demonstrated that the additive effect weakened as the sum of two  $\Delta T_m$  increased (Figure 3C), which is consistent with the literature [21].



**Figure 4.** Mutations sites chosen by the relative entropy angle  $\theta$ . (A) Residues colored according to our protein sectors color scheme on the “closed” conformation of ADK (PDB id: 1ANK chain A). ATP analog and AMP (both colored black) were shown in sticks. The structure on the right was obtained by rotating the diagram on the left by 90 degrees clockwise around the vertical axis. (B) Residues directly participating in substrate binding in *E. coli* ADK. All sectors except the magenta one were involved in ADK’s function execution. (C) Eight selected mutation sites were marked as magenta spheres in the tertiary structure of ADK.

The protein sector method was previously applied to identify activity-related regions within specific proteins, but thermal stability improvement was rarely achieved. Only one report mentioned that 23 covariant residues found in GH1  $\beta$ -glucosidases were confirmed using statistical coupling analysis as one protein sector. This sector could be further divided into two subsectors by using alanine-screening mutagenesis experiments, one is related to catalytic activity regulation, whereas the other is related to thermal stability [24]. Thus, the single-protein-sector architecture proposed by the previous study was too rough to accurately categorize residues. Moreover, due to the lack of statistical insight in terms of what amino acid type should a position be mutated into, mutation effects of all 23 positions were characterized by mutating into the same amino acid type (alanine). Consequently,

no insight was proposed on how to read just the amino acid type for six positions whose mutation caused  $T_m$  reduction. By contrast, in our study, we present a detailed protein sector assignment together with rational selection of both residue position and amino acid types for thermostable mutant construction.

Studies have shown that various techniques were complementary for the thermal stability improvement of the same protein, where different stabilizing methods generated different mutants with few overlapping sites [21]. The consensus technique considered those well-conserved sites while our method focused on the correlation in whole sequences and the divergence between mesophilic and thermophilic profile. To obtain more stable ADK mutants, integrating our method with other techniques, such as a consensus-based method, might be a promising optimization scheme for further improvements.

Among various developed computational methods to predict how amino acid substitutions influence protein stability, the most accurate ones are still tertiary structure based [25]. Most existing programs, such as Rosetta [8], FoldX [26], and ABACUS [27], are based on the optimization of physical or statistical folding free energy functions ( $\Delta G^0$ ) to determine appropriate amino acid types for different positions of given protein backbone structures. Almost all factors contributing to protein stability such as van der Waals, electrostatics, solvation, hydrogen bonding, disulfide bonding, backbone torsion, side-chain torsion and so on are totally taken into account in the construction of energy function, which guarantees reasonable accuracy of the prediction of protein modifications. However, the accuracies of these energy functions are contingent on the availability of high-resolution experimental structures of target proteins with little consideration of the dynamics of the backbone structure. These requirements limit their applications to proteins without an experimental structure. Additionally, these methods do not leverage the rich information of homologous protein sequences, which are far more readily available than experimental structures, considering only 50,000 (0.03%) of 180 million protein sequences in the UniProt database have structures experimentally solved in the Protein Data Bank database [28]. Although studies have attempted to combine the physical energy function with an additional energy term accounting for amino acid composition at individual positions of sequence profile [14,29], their profile terms lack pairwise evolutionary coupling information, which our study shows is crucial for a successful design.

Compared with structure-based predictors, developed sequence-based methods are few in number and applied usually in the absence of experimental structure [30]. I-Mutant2.0 quantifies the  $\Delta G^0$  change plus or minus of single residue substitution by just considering the nearest sequence neighbors when only the sequence is available [31]. It is trained and tested on a data set derived from the Thermodynamic Database for Proteins and Mutants. Then I-Mutant2.0 can be applied to any target protein. The principles of another three methods, iPTREE-STAB, MuPRO, and EASE-MM, are almost the same as I-Mutant2.0, with just different training procedures [32–34]. The ICE method can improve the protein thermostability by minimizing target protein's local structural entropy in the sequence alignment of the target and another reference [9]. These five sequence-based methods present one obvious advantage of just requiring information from one or two protein sequences. However, they also have the disadvantage of just considering the local sequence information and ignoring the long-range correlation among residue pairs. In contrast, our method effectively utilizes the pairwise evolutionary coupling information embedded in the multiple sequence alignment of homologous proteins, which can be readily collected from corresponding databases.

In summary, as our approach is solely based on evolutionary information, a future research direction may focus on a combination of evolutionary information with physical/statistical energy terms from structure-based methods. Such explicit consideration of physical interaction in the protein structure should in principle further boost the success rate of our current design efforts solely based on evolutionary signals from MSA.

## 4. Materials and Methods

### 4.1. Primary-Sequence-Based Thermostability Design

Briefly, our sequence-based protein redesign algorithm consists of the following steps.

1. Construct a multiple sequence alignment for the target protein family.
2. Apply RCA to the MSA to obtain a correlation matrix.
3. Identify protein sectors through the eigen-decomposition of the correlated matrix.
4. Calculate the relative entropy angle  $\theta$  from the difference between mesophiles and thermophiles. The protein sector with the largest average  $\theta$  should be picked out for further selection of mutation sites.
5. Replace residues whose  $\theta$  exceed a certain threshold with the corresponding amino acids which were the most common among thermophiles or whose side chains had distinct properties.
6. Characterize the mutants by circular dichroism measurement and enzymatic activity assay.
7. Consider constructing double-point mutations consisting of above characterized single-point mutation sites.

Details of the whole process will be introduced step by step in the following sections.

### 4.2. Sequence Alignment and Curation

Protein sequences of the ADK family (PF00406) were obtained from Pfam database version 32.0 [35]. Sequences with domain organization labeled as “ADK, ADK\_lid”, or “ADK” were retrieved. Fragments or entries that do not have the function of catalyzing the reversible phosphate group transfer between AMP and ATP according to UniProt database were removed from the pool. MSA of the remaining 9203 sequences was performed with Clustal Omega 1.2.2 using Hidden Markov Model (HMM) of PF00406 family as seed HMM [36]. The alignment output was projected onto ADK wt (P69441/KAD\_ECOLI) indices such that only columns corresponding to the residues of ADK from *E. coli* were preserved. Proteins were further classified into “thermophilic”, “mesophilic”, and “un-recognized” categories according to the optimal growth temperature range of their host organisms as reported by National Center for Biotechnology Information Genome [37], UniProt and so on. Details of all data sources are listed in the Supplementary Information.

### 4.3. Coevolution Analysis

A modified version of the RCA was applied to locate residues related to different functions. Consider position  $i$  of an MSA with  $M$  sequences and  $N$  positions, the substitution event  $X_{ikl}$  where the amino acid at row  $k$  was mutated to amino acid at row  $l$  could be scored using BLOSUM50 scoring matrix. If either sequence  $k$  or sequence  $l$  but not both of them contained a gap at position  $i$ ,  $X_{ikl}$  would be assigned the gap penalty  $-8$ . If both sequences  $k$  and  $l$  contained a gap,  $X_{ikl}$  would be assigned the gap-to-gap score zero. Afterward, the correlation between positions  $i$  and  $j$  could be quantified using Pearson correlation coefficient between all substitution events at position  $i$  and all substitutions at position  $j$ :

$$r_{ij} = [2/(M(M-1))] \sum_{k=1}^{M-1} \sum_{l=k+1}^M \left[ (X_{ikl} - \langle X_i \rangle) (X_{jkl} - \langle X_j \rangle) / (\sigma_i \sigma_j) \right] \quad (2)$$

$\langle X_i \rangle$  was the average for all substitution scores at position  $i$ , whereas  $\sigma_i$  was the standard deviation. Thus, an  $N \times N$  square matrix denoting RCA correlations could be generated from the MSA with  $M$  sequences and  $N$  positions.  $N$  eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_N$ , ranking in a descending order, could be computed on which eigen-decomposition could be performed for the correlation matrix. The corresponding eigenvectors can be denoted as  $v_1, v_2, \dots, v_N$ , respectively.

Only the top-ranking eigen components were correlated to “real” information in the correlation matrix, while the rest were insignificant noises. Specifically, among the 214 eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{214}$  (ranked in descending order of values), most components could be attributed to limited sampling of sequences (statistical noise) and the largest one corresponding to  $\lambda_1$  (13.29) arose from the phylogenetic effect within ADK family (historical noise) [17]. Therefore, only a few eigenvectors were required as an approximation for protein sector analysis of the original RCA matrix. In our work, the summation of three eigen-decomposition components related to  $\lambda_2$  (6.67),  $\lambda_3$  (4.97),  $\lambda_4$  (4.35) exhibited the features resembling the original correlation matrix (Figure S1).

The resulting three-component-approximation matrix was then used to cluster residues in the eigen-space representation. The boundary of noises determined by repeating the same calculations on a randomized ADK MSA and comparing the distributions of the two sets of eigenvalues. Specifically, we randomly shuffled the order of amino acids at each position in the alignment to eliminate all correlations between positions while preserving the amino acid frequencies at each position. 10 trials of randomization would give the average histograms of elements in eigenvector 2–4, respectively (Figure S4A–C, bottom panels). The Gaussian fit of the three histograms could yield the positional weight boundaries, which were used to rule out the residue with weak correlation. More specifically, residues with all three position weights from eigenvector 2–4 which were smaller than the boundaries were considered to have no significant correlation (Figure S4A–C, upper panels). We then deleted them and classified remaining residues into four clusters by K-means clustering algorithm. Finally, four sector representation of the RCA matrix was obtained.

#### 4.4. Sequence Entropy Analysis

The relative sequence entropy of thermophilic/mesophilic sequence profile was used to determine residues related to thermostability. Considering position  $i$  at an MSA with  $M$  sequences, its relative entropy vector  $\vec{D}_i$  could be defined as:

$$\vec{D}_i = [D_i^a, D_i^c, \dots, D_i^y, D_i^{gap}] \quad (3)$$

$D_i^a$ , the relative entropy of amino acid  $a$  at position  $i$ , was defined as the logarithm of binomial probability of choosing the observed number of  $a$  at  $i$ :

$$D_i^a = \log \left[ (q^a)^{Mf_i^a} (1 - q^a)^{M(1-f_i^a)} M! / ((Mf_i^a)! (M(1-f_i^a))!) \right] \approx f_i^a \log(f_i^a / q^a) + (1 - f_i^a) \log((1 - f_i^a) / (1 - q^a)) \quad (4)$$

$f_i^a$  was the frequency of amino acid  $a$  at position  $i$ , and  $q^a$  was the mean frequency of amino acid  $a$  among all proteins [17]. The “Mean frequency” of gaps  $q^{gap}$  was defined as the gap frequency in the trimmed MSA. The mean frequency of all other 20 amino acids was normalized by multiplying  $(1 - q^{gap})$ .  $|D_i^a|$  represented the difference of the probability distribution of amino acid type  $a$  from background probability at position  $i$  of the protein family. The relative entropy vectors  $\vec{D}_i$  were separately calculated for thermophilic proteins and for mesophilic proteins.  $\vec{D}_i^{thermo}$  and  $\vec{D}_i^{meso}$  denoted the relative entropy vectors for the thermophilic and mesophilic sequence profiles, respectively, for the  $i$ th position. The angle between these two vectors represented how divergent the encoding of amino acids at position  $i$  was for thermophilic sequences and mesophilic sequences. We defined it as the relative entropy angle  $\theta$ , and it can be calculated as follows:

$$\theta_i = \cos^{-1} \left[ \left( \vec{D}_i^{thermo} \cdot \vec{D}_i^{meso} \right) / \left( \left| \vec{D}_i^{thermo} \right| \left| \vec{D}_i^{meso} \right| \right) \right] \quad (5)$$

We hypothesized that residues that were essential for thermostability would score higher in this  $\theta$  parameter. Considering the correlations between ADK residues, we selected the protein sector with largest average  $\theta$  as the mutation target. Then residues with  $\theta$  value

above one custom threshold in this sector were mutated from their mesophilic type to the most common amino acid in that position for thermophilic sequences.

#### 4.5. Materials

Tris(hydroxymethyl)aminomethane was purchased from Sangon Biotech (Shanghai) Co., Ltd. (Shanghai, China). KCl, MgOAc<sub>2</sub>, and bovine serum albumen (BSA) were obtained from Sinopharm Chemical Reagent Co., Ltd. (Beijing, China). ATP, AMP, and Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were obtained from BBI Life Sciences (Shanghai, China). Pyruvate kinase/lactic dehydrogenase (PK/LDH) enzymes from rabbit muscle and P<sup>1</sup>,P<sup>5</sup>-di(adenosine-5') pentaphosphate pentasodium salt (Ap5A) were from Sigma-Aldrich(Shanghai)Trading Co.,Ltd. (Shanghai, China). Nicotinamide adenine dinucleotide disodium salt (NADH) was from Roche Diagnostics GmbH (Mannheim, Germany). Phosphoenolpyruvate (PEP) was from Alfa Aesar, Thermo Fisher Scientific (Waltham, MA, USA). All the aforementioned reagents were used without further purification.

#### 4.6. ADK Purification

The wild-type ADK from *E. coli* strain K12 was used as the starting sequence in this study. The wild-type ADK and its mutants were purified in a manner similar to that described earlier [23]. The protein was first purified with a HisTrap HP histidine-tagged nickel affinity chromatography column (GE Healthcare, Chicago, Illinois, USA). Further purification was performed with a HiTrap Q HP anion-exchange chromatography column and a Superdex 200 Increase small-scale size-exclusion chromatography column (both were from GE Healthcare). The purification results of ADK wt, R206F, and R206F/S41A were shown in the Supplementary Material (Figure S5).

#### 4.7. Room Temperature Forward Activity Assay of ADK

All enzymatic assays were performed with the assay buffer: 100 mM Tris (pH by acetic acid), pH 7.5, 100 mM KCl, 0.08 mg/mL (1.2 μM) BSA, and 0.6 mM TCEP. ADP production was monitored with coupling enzymes PK and LDH. Each ADP produced caused the oxidation of one NADH to NAD<sup>+</sup> and was detected through the absorbance change at 340 nm [38]. The coupling reagent concentrations used were 5 mM PEP and 0.8 mM NADH with 2 μL PK/LDH (600–1000 units/mL PK and 900–1400 units/mL LDH) per 100 μL reaction. ADK's forward reaction rates were assayed at a fixed ATP concentration, and AMP concentrations varied from 0 to 3 mM, with MgOAc<sub>2</sub> concentrations at 2 mM. All the measurements were obtained in 96-well plates and read with a microplate reader (Molecular Devices SpectraMax M2) at room temperature (~25 °C). All the rates reported in this work are initial velocities, which are measured at the beginning of the reaction, when the product concentration is increasing linearly under known concentrations of enzyme and substrate. Each experiment was repeated at least three times and averaged to determine error bars. All error bars are reported to 1 standard deviation.

#### 4.8. CD Temperature Melt Measurement

The  $T_m$ s of wild type ADK and mutants were measured with CD spectrometer (Applied Photophysics qCD Chirascan plus, Surrey, UK). ADK sample (10 μM) in phosphate buffer (60 mM) with pH 7.4 was monitored at 222 nm to obtain the CD trajectory. The temperature range from 30 °C to 70 °C is scanned at a rate of 1 °C/min.  $T_m$  is analyzed by using the modified Sigmoid function contained in the subsidiary software (CDNN2.1). All measurements were repeated three times to generate their standard deviations.

#### 4.9. Temperature-Dependent Enzymatic Activity Assay

ADK wt was a mesophilic enzyme catalyzing the following reversible reaction:  $Mg^{2+} \cdot ATP + AMP \xrightleftharpoons{ADK} Mg^{2+} \cdot ADP + ADP$ . Activity assays at various temperatures were implemented in the direction of ADP formation by using a modified end-point method as described previously [38]. Coupling buffer, ATP, and AMP solutions were

prepared in advance. The final assay system consisted of 100 mM Tris, pH 7.5, 100 mM KCl, 1.2  $\mu$ M BSA, 0.8 mM TCEP, 2 mM MgOAc<sub>2</sub>, 2.5 mM ATP, 2.5 mM AMP, and 5 nM ADK. Before the reaction, the mixture of 38  $\mu$ L coupling buffer, 20  $\mu$ L ATP, and 20  $\mu$ L AMP was heated to the desired temperature for 2 min. A preheated ADK solution of 22  $\mu$ L (22.72 nM in 100 mM Tris, pH 7.5, 100 mM KCl) was blended with the mixture to initiate the reaction, and it lasted for 1 min. The enzyme reaction was terminated immediately at the end of 1 min by adding 1 mM Ap5A and cooled on a bath of iced water for 2 min. The amount of ADP produced was quantified by the ADP-involved oxidation of NADH to NAD<sup>+</sup> through the addition of 2.5  $\mu$ L PK/LDH, 5 mM PEP, and 1 mM NADH at room temperature for 10 min. The enzymatic activity can be calculated from the absorbance at 340 nm measured with ultraviolet–visible spectroscopy (Agilent Technologies 8453/G1103A, Santa Clara, CA, USA). For each temperature, three independent experiments were performed with identical conditions simultaneously.

## 5. Conclusions

In conclusion, we developed a bioinformatics method using homologous primary sequences to improve the thermal stability of proteins and applied it to rationally redesign improved thermally stable variants of ADK wt. Our approach is directly applicable to any protein family with adequate thermophilic and mesophilic primary sequence information. Single- and double-point mutants constructed based on this method increase the thermal denaturation temperature of the mesophilic *E. coli* ADK by 5.5 and 8.3 °C, respectively, while preserving most of the catalytic function at ambient temperatures. The combination of our method with other theoretical methods is expected to provide more efficient guidance for stability modification in the future. In addition, our method might help to understand the intrinsic mechanism of species evolution when adapting to changing environmental conditions.

**Supplementary Materials:** Supplementary Materials can be found at <https://www.mdpi.com/1422-0067/22/5/2768/s1>.

**Author Contributions:** Conceptualization, Y.-W.T., J.C. and C.Z.; methodology, J.C. and C.Z.; data curation, J.C. and C.Z.; software, J.C. and C.Z.; validation, Y.-W.T. and J.C.; formal analysis, J.C. and C.Z.; investigation, J.C. and C.Z.; writing—original draft, Y.-W.T., J.C. and C.Z.; writing—review and editing, Y.-W.T., J.C., H.C. and C.Z.; visualization, J.C.; resources, Y.-W.T.; supervision, Y.-W.T.; project administration, Y.-W.T.; funding acquisition, Y.-W.T., J.C. and C.Z. contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China grants (11274076 and 21773039) and Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), all to Y.-W.T.

**Institutional Review Board Statement:** Our study did not involve humans or animals.

**Informed Consent Statement:** Our study did not involve humans.

**Data Availability Statement:** Data is contained within the article or Supplementary Material.

**Acknowledgments:** This work was supported in part by ZJLab, all to Y.-W.T. We would like to thank Song Song and Chongjun Ma for supervising C.Z. We also would like to thank Mengyuan Sun, Ke Xu, Yingqi Lou, Qiuyu Tao, and Jialin Yang for technical assistance. We thank Guanghong Zuo for fruitful discussion.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

ADK	adenylate kinase
ADK aq	<i>Aquifex aeolicus</i> ADK
ADK wt	<i>E. coli</i> ADK
ADP	adenosine diphosphate
AMP	adenosine monophosphate
Ap5A	P <sup>1</sup> ,P <sup>5</sup> -Di(adenosine-5') pentaphosphate pentasodium salt
ATP	adenosine triphosphate
BSA	bovine serum albumen
CD	circular dichroism
<i>E. coli</i>	<i>Escherichia coli</i>
HMM	hidden Markov model
ICE	improved configurational entropy
MgOAc <sub>2</sub>	magnesium acetate
MSA	multiple sequence alignment
NADH	nicotinamide adenine dinucleotide disodium salt
PEP	phosphoenolpyruvate
PK/LDH	pyruvate kinase/lactic dehydrogenase
RCA	residue correlation analysis
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
<i>T<sub>m</sub></i>	melting temperature

## References

- Magliery, T.J. Protein Stability: Computation, Sequence Statistics, and New Experimental Methods. *Curr. Opin. Struct. Biol.* **2015**, *33*, 161–168. [CrossRef]
- As, B.; Jk, B.; Mj, A. Status of Protein Engineering for Biocatalysts: How to Design an Industrially Useful Biocatalyst. Available online: <https://pubmed.ncbi.nlm.nih.gov/21115265/> (accessed on 14 July 2020).
- Wang, G.; Cao, R.-Y.; Chen, R.; Mo, L.; Han, J.-F.; Wang, X.; Xu, X.; Jiang, T.; Deng, Y.-Q.; Lyu, K.; et al. Rational Design of Thermostable Vaccines by Engineered Peptide-Induced Virus Self-Biomineralization under Physiological Conditions. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7619–7624. [CrossRef]
- Deller, M.C.; Kong, L.; Rupp, B. Protein Stability: A Crystallographer's Perspective. *Acta Cryst. F Struct. Biol. Commun.* **2016**, *72*, 72–95. [CrossRef] [PubMed]
- Cobb, R.E.; Sun, N.; Zhao, H. Directed Evolution as a Powerful Synthetic Biology Tool. *Methods* **2013**, *60*, 81–90. [CrossRef] [PubMed]
- Schweiker, K.L.; Zarrine-Afsar, A.; Davidson, A.R.; Makhatazde, G.I. Computational Design of the Fyn SH3 Domain with Increased Stability through Optimization of Surface Charge–Charge Interactions. *Protein Sci.* **2007**, *16*, 2694–2702. [CrossRef] [PubMed]
- Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and Accurate Predictions of Protein Stability Changes upon Mutations Using Statistical Potentials and Neural Networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25*, 2537–2543. [CrossRef] [PubMed]
- Alford, R.F.; Leaver-Fay, A.; Jeliazkov, J.R.; O'Meara, M.J.; DiMaio, F.P.; Park, H.; Shapovalov, M.V.; Renfrew, P.D.; Mulligan, V.K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048. [CrossRef]
- Bae, E.; Bannen, R.M.; Phillips, G.N. Bioinformatic Method for Protein Thermal Stabilization by Structural Entropy Optimization. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9594–9597. [CrossRef]
- Chan, C.-H.; Liang, H.-K.; Hsiao, N.-W.; Ko, M.-T.; Lyu, P.-C.; Hwang, J.-K. Relationship between Local Structural Entropy and Protein Thermostability. *Proteins Struct. Funct. Bioinform.* **2004**, *57*, 684–691. [CrossRef]
- Moon, S.; Bannen, R.M.; Rutkoski, T.J.; Phillips Jr., G.N.; Bae, E. Effectiveness and Limitations of Local Structural Entropy Optimization in the Thermal Stabilization of Mesophilic and Thermophilic Adenylate Kinases. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 2631–2642. [CrossRef]
- Sternke, M.; Tripp, K.W.; Barrick, D. Consensus Sequence Design as a General Strategy to Create Hyperstable, Biologically Active Proteins. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11275–11284. [CrossRef]
- Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E.Z.; Kern, D. Linkage between Dynamics and Catalysis in a Thermophilic-Mesophilic Enzyme Pair. *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–949. [CrossRef]
- Huang, X.; Zheng, W.; Pearce, R.; Zhang, Y. SSiPE: Accurately Estimating Protein–Protein Binding Affinity Change upon Mutations Using Evolutionary Profiles in Combination with an Optimized Physical Energy Function. *Bioinformatics* **2020**, *36*, 2429–2437. [CrossRef]

15. Burlacu-Miron, S.; Perrier, V.; Gilles, A.-M.; Pistotnik, E.; Craescu, C.T. Structural and Energetic Factors of the Increased Thermal Stability in a Genetically Engineered *Escherichia Coli* Adenylate Kinase. *J. Biol. Chem.* **1998**, *273*, 19102–19107. [[CrossRef](#)] [[PubMed](#)]
16. Göbel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated Mutations and Residue Contacts in Proteins. *Proteins Struct. Funct. Bioinform.* **1994**, *18*, 309–317. [[CrossRef](#)] [[PubMed](#)]
17. Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774–786. [[CrossRef](#)] [[PubMed](#)]
18. Ainsworth, S.; Macfarlane, N. A Kinetic Study of Rabbit Muscle Pyruvate Kinase. *Biochem. J.* **1973**, *131*, 223–236. [[CrossRef](#)]
19. Tan, Y.-W.; Hanson, J.A.; Yang, H. Direct Mg<sup>2+</sup> Binding Activates Adenylate Kinase from *Escherichia Coli*. *J. Biol. Chem.* **2009**, *284*, 3306–3313. [[CrossRef](#)]
20. Rigoldi, F.; Donini, S.; Redaelli, A.; Parisini, E.; Gautieri, A. Review: Engineering of Thermostable Enzymes for Industrial Applications. *APL Bioeng.* **2018**, *2*, 011501. [[CrossRef](#)] [[PubMed](#)]
21. Moon, S.; Jung, D.; Phillips Jr., G.N.; Bae, E. An Integrated Approach for Thermal Stabilization of a Mesophilic Adenylate Kinase. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 1947–1959. [[CrossRef](#)]
22. Perrier, V.; Burlacu-Miron, S.; Bourgeois, S.; Surewicz, W.K.; Gilles, A.-M. Genetically Engineered Zinc-Chelating Adenylate Kinase From *Escherichia Coli* with Enhanced Thermal Stability. *J. Biol. Chem.* **1998**, *273*, 19097–19101. [[CrossRef](#)] [[PubMed](#)]
23. Hanson, J.A.; Duderstadt, K.; Watkins, L.P.; Bhattacharyya, S.; Brokaw, J.; Chu, J.-W.; Yang, H. Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18055–18060. [[CrossRef](#)]
24. Tamaki, F.K.; Textor, L.C.; Polikarpov, I.; Marana, S.R. Sets of Covariant Residues Modulate the Activity and Thermal Stability of GH1  $\beta$ -Glucosidases. *PLoS ONE* **2014**, *9*, e96627. [[CrossRef](#)] [[PubMed](#)]
25. Pucci, F.; Bernaerts, K.V.; Kwasigroch, J.M.; Rومان, M. Quantification of Biases in Predictions of Protein Stability Changes upon Mutations. *Bioinformatics* **2018**, *34*, 3659–3665. [[CrossRef](#)] [[PubMed](#)]
26. Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More than 1000 Mutations. Available online: <https://pubmed.ncbi.nlm.nih.gov/12079393/> (accessed on 14 July 2020).
27. Xiong, P.; Wang, M.; Zhou, X.; Zhang, T.; Zhang, J.; Chen, Q.; Liu, H. Protein Design with a Comprehensive Statistical Energy Function and Boosted by Experimental Selection for Foldability. *Nat. Commun.* **2014**, *5*, 1–9. [[CrossRef](#)]
28. The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)]
29. Pearce, R.; Huang, X.; Setiawan, D.; Zhang, Y. EvoDesign: Designing Protein-Protein Binding Interactions Using Evolutionary Interface Profiles in Conjunction with an Optimized Physical Energy Function. Available online: <https://pubmed.ncbi.nlm.nih.gov/30851277/> (accessed on 14 July 2020).
30. Structure-Based Mutant Stability Predictions on Proteins of Unknown Structure. *J. Biotechnol.* **2012**, *161*, 287–293. [[CrossRef](#)] [[PubMed](#)]
31. Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306–W310. [[CrossRef](#)]
32. Huang, L.-T.; Gromiha, M.M.; Ho, S.-Y. IPTREE-STAB: Interpretable Decision Tree Based Method for Predicting Protein Stability Changes upon Mutations. *Bioinformatics* **2007**, *23*, 1292–1293. [[CrossRef](#)]
33. Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins Struct. Funct. Bioinform.* **2006**, *62*, 1125–1132. [[CrossRef](#)]
34. Folkman, L.; Stantic, B.; Sattar, A.; Zhou, Y. EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. Available online: <https://pubmed.ncbi.nlm.nih.gov/26804571/> (accessed on 20 October 2020).
35. Punta, M.; Coghill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. The Pfam Protein Families Database. *Nucleic Acids Res.* **2012**, *40*, D290–D301. [[CrossRef](#)] [[PubMed](#)]
36. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)] [[PubMed](#)]
37. Tatusova, T.; Ciufu, S.; Fedorov, B.; O'Neill, K.; Tolstoy, I. RefSeq Microbial Genomes Database: New Representation and Annotation Strategy. *Nucleic Acids Res.* **2014**, *42*, D553–D559. [[CrossRef](#)] [[PubMed](#)]
38. Bae, E.; Phillips, G.N. Structures and Analysis of Highly Homologous Psychrophilic, Mesophilic, and Thermophilic Adenylate Kinases. *J. Biol. Chem.* **2004**, *279*, 28202–28208. [[CrossRef](#)] [[PubMed](#)]