International Journal of
Molecular Sciences

MDPI

Article

A Statistical Analysis of the Sequence and Structure of Thermophilic and Non-Thermophilic Proteins

Zahoor Ahmed, Hasan Zulfiqar, Lixia Tang and Hao Lin *

School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China
* Correspondence: hlin@uestc.edu.cn

Abstract: Thermophilic proteins have various practical applications in theoretical research and in industry. In recent years, the demand for thermophilic proteins on an industrial scale has been increasing; therefore, the engineering of thermophilic proteins has become a hot direction in the field of protein engineering. However, the exact mechanism of thermostability of proteins is not yet known, for engineering thermophilic proteins knowing the basis of thermostability is necessary. In order to understand the basis of the thermostability in proteins, we have made a statistical analysis of the sequences, secondary structures, hydrogen bonds, salt bridges, DHA (Donor–Hydrogen–Accepter) angles, and bond lengths of ten pairs of thermophilic proteins and their non-thermophilic orthologous. Our findings suggest that polar amino acids contribute to thermostability in proteins by forming hydrogen bonds and salt bridges which provide resistance against protein denaturation. Short bond length and a wider DHA angle provide greater bond stability in thermophilic proteins. Moreover, the increased frequency of aromatic amino acids in thermophilic proteins contributes to thermal stability by forming more aromatic interactions. Additionally, the coil, helix, and loop in the secondary structure also contribute to thermostability.

Keywords: thermophilic proteins; proteins sequence; secondary structure; hydrogen bonds; salt bridges

check for updates

Citation: Ahmed, Z.; Zulfiqar, H.; Tang, L.; Lin, H. A Statistical Analysis of the Sequence and Structure of Thermophilic and Non-Thermophilic Proteins. *Int. J. Mol. Sci.* **2022**, *23*, 10116. https://doi.org/10.3390/ijms231710116

Academic Editor: Jean-Marc Lancelin

Received: 29 July 2022
Accepted: 31 August 2022
Published: 4 September 2022

1. Introduction

Proteins are large biomolecules containing one or more long chains of amino acid residues. Enzymes are complex proteins that are involved in life-essential processes like DNA replication, transcription, translation, metabolism, and signal transduction [1,2]. Enzymes can also carry out chemical transformations, which makes them valuable for industrial applications as biocatalysts [3,4]. In the early 2000s, biocatalysts were used for the synthesis or resolution of optically active intermediates [3,5]. Since then, biocatalysts have gradually evolved as applicable tools for the large-scale synthesis and manufacturing of chemicals; thus the demand for biocatalysts is increasing [6,7].

At present, biocatalysts are extensively used in the pharmaceutical, food, animal nutrition, cosmetics, and beverage industries [8–10]. In addition, the use of biocatalysts has also entered the detergent, textile, pulp, and paper industries, and into organic synthesis, natural gas conversion, and the biofuel industries [11,12]. Common biocatalysts used on an industrial scale include proteinase and protease for food processing, α-amylase and xylanases in paper bleaching, cellulase and lipase in polymer breakdown, pullulanase as detergents, L-haloacid dehalogenase for chiral halo-carboxylic acid production, and *Sulfolobus solfataricus* γ-Lactamase for the synthesis of the γ-bicyclic lactam, which is an important building block for the anti-HIV compound abacavir [13–16]. Biocatalysts have more advantages with respect to sustainability, process efficiency, exceptional product selectivity, and lower environmental and physiological toxicity when compared to traditional catalysts [17,18]. Therefore, currently, biocatalysts are preferred to traditional catalysts, but there are still several factors that limit their application on an industrial scale, including biocatalyst stability. In fact, the stability of biocatalysts has received attention from many scholars [18–22]. As we know, a higher temperature can improve the efficiency of enzyme catalysis. However, proteins are temperature-sensitive and denature at high temperatures, which hinders the wide application of enzymes in the industry [17]. Using enzymes with high thermal stability to solve this shortcoming is the key to the application of enzymes on an industrial scale. Therefore, using enzymes in thermophiles is a means to solve the problem [18]. In addition, recent advancements in the protein engineering field have made protein engineering facile and have drawn the interest of researchers to engineer thermostable enzymes for industrial use [23–28].

Some microorganisms in nature have been seen to survive in severe environmental and thermodynamic conditions, and their biological growth is most ideal between 50 and 100 °C [29]. The organisms living in such harsh conditions of increased temperatures are generally termed thermophiles. The molecular machinery of the thermophiles is developed to withstand and function at high temperatures [30]. These thermophiles produce proteins that are capable to maintain their structure and activity at high temperatures [31]. The question of how these thermophilic proteins remain stable at such high temperatures has attracted more and more attention. In recent years, researchers have focused on discovering the sequence and structural features of thermophilic proteins. This finding is critical for the theoretical description of the principle behind protein thermal stability [30]. In addition, the discovery of relevant factors also helps to design heat-resistant proteins/enzymes that can meet the requirements of industrial processes.

We designed this study with a view toward the importance of principles behind the stability of thermophilic proteins at high temperatures. We obtained thermophilic proteins from our previous study on thermophilic proteins [32] and searched for their non-thermophilic orthologous. We preferred thermophilic and non-thermophilic orthologous pairs with the optimum growth temperature (OGT) difference of >20 °C, as the higher OGT difference between a thermophilic protein and its non-thermophilic orthologous could give us clear reasons for thermostability in proteins. Finally, we obtained 10 thermophilic proteins and their non-thermophilic orthologous. The structures of these proteins were obtained from the protein database (PDB), and their sequences, secondary structures, hydrogen bonds, bond lengths, bond angles, and salt bridges were analyzed. In the analysis, we found that the polar amino acids glutamic acid, histidine, lysine, arginine, tyrosine, and aromatic amino acids were slightly more frequent in the thermophilic proteins. Moreover, in the secondary structure, the percentage of the coil, helix, and sheet in thermophilic proteins was higher, while the turn percentage in thermophilic proteins was lower. Subsequently, the number of hydrogen bonds and salt bridges of thermophilic proteins increased. Compared with non-thermophilic proteins, the DHA angle in thermophilic protein was wider and the bond length was shorter. The following sections describe the analysis in detail

2. Results and Discussion

To understand the important factors that maintain the thermostability in protein, 10 thermophilic proteins and their non-thermophilic orthologous were collected to investigate the effects of their sequence, secondary structure, hydrogen bond, salt bridge, bond length, bond angle, and aromaticity value on thermal stability in proteins.

The primary amino acid composition (AAC) of a protein imparts specific properties to the protein molecule [32–35]. Our previous studies have shown that there are significant differences in AAC between thermophilic and non-thermophilic proteins [36,37], which suggests that AAC is the main basis of protein thermostability. Therefore, we analyzed the AAC. Figure 1 shows the frequency of amino acids (AAs) in the thermophilic and non-thermophilic proteins.

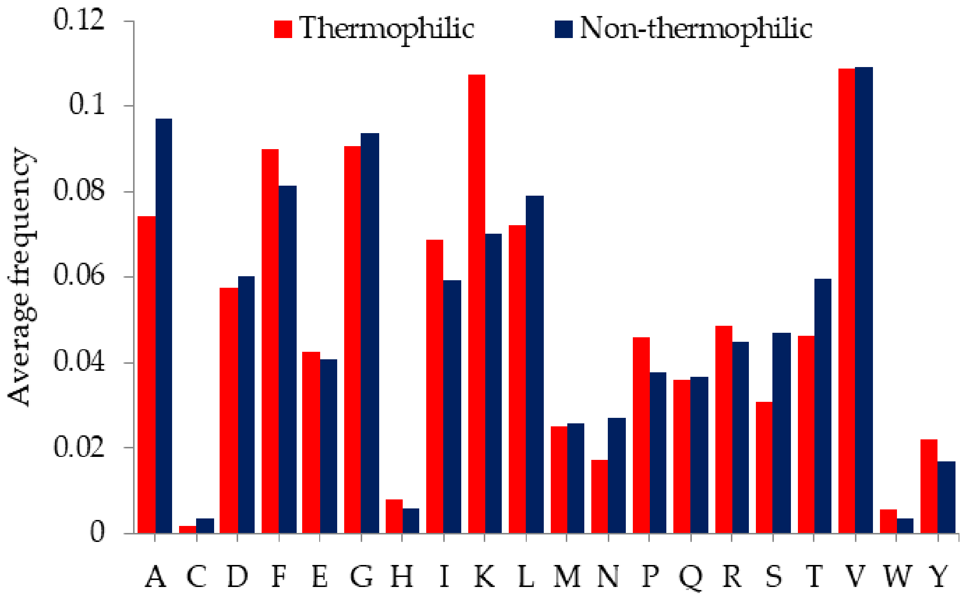

Figure 1. The average frequencies of amino acids in thermophilic proteins and their non-thermophilic orthologous.

As shown in Figure 1, among polar AAs, glutamic acid (E), histidine (H), lysine (K), arginine (R), and tyrosine (Y) have higher frequencies in thermophilic proteins. Polar amino acids R and Y are long-side chain amino acids [38]. Due to their long-side chain, these AAs contribute to the formation of hydrogen bonds, salt bridges, and other long- and short-range interactions to stabilize the protein structure. Moreover, R and Y are abundantly present in the binding hotspot of protein interactions. Hence, it seems that R and Y have similar contributions to the binding and folding of proteins and hinder the unfolding of proteins at elevated temperatures [38–40]. In addition, the guanidium group in R involves in the formation of salt bridges that resist thermal denaturation of proteins [38,41–43]. Polar amino acid K contains a side chain with a positive charge and forms stable electrostatic interactions with nearby negatively charged groups, and offers stability to the protein structure [44]. It has been reported that amino acid E easily forms interactions in protein to stabilize its structure [45–47]. The higher frequency of these polar AAs in thermophilic proteins infers that these AAs contribute to the formation of hydrogen bonds, salt bridges, and other stable interactions to resist the thermal denaturation of proteins at elevated temperatures [48]. Pace et al. [49] also pointed out that the long-range interactions of polar AAs buried in proteins that are not bounded by hydrogen bonding have other interaction forces, such as van der Waals interactions, which stabilize the protein structure. The increased frequency of these polar AAs suggests that they may be the cause to maintain the protein structural stability.

Other polar AAs, including threonine (T), glutamine (Q), asparagine (N), and serine (S), have a lower frequency in thermophilic proteins. Amino acids T and S could interact with water molecules at high temperatures and increase instability in protein molecules [50,51]. Moreover, S has been reported to impair hydrophobic interactions between beta strands [52]. Increased temperature can also cause chemical alterations in AAs. Amino acids Q and N undergo deamidation at elevated temperatures, which imparts an extra negative charge on residues and alters the protein interactions that affect the folding and activity of proteins [53]. The low frequency of these polar AAs in thermophilic proteins suggests that they may be one of the factors that destroy thermal stability in proteins.

Among nonpolar AAs, proline (P) is more common in thermophilic proteins. P has a more rigid structure and plays a role in reducing the entropy of the main chain, resisting the protein unfolding, and stabilizing the loop structure. Due to its hydrophobicity, P interacts with hydrophobic residues on the core and surface of protein molecules, thus preventing the protein from unfolding at elevated temperatures and maintaining protein activity [54]. Isoleucine (I) is also found to be more frequent in thermophilic proteins than in non-thermophilic proteins. Previous studies have also reported the frequent occurrence of I in thermophilic proteins; however, the exact mechanism by which I is used to contribute to thermostability in protein is still not clear [55].

Other nonpolar AAs, including alanine (A), cystine (C), glycine (G), and leucine (L), are less frequent in thermophilic proteins. G and A are amino acids with short-side chains. These AAs form a flexible, rather than a rigid, mechanism. Since their side chains are too short, these AAs form fewer short-range interactions and fail to form long-range interactions to stabilize the protein structure at high temperatures [54]. Amino acid C is easy to deamidate or oxidize at high temperatures, which changes the charge on residues, disturbs the interaction in protein, and affects the folding of the protein [53]. It has been reported that the amino acid L existing in the protein core does not easily form van der Waals and other interactions, resulting in poor thermal stability in the protein [54]. It can be inferred that these unstable amino acids are avoided in thermophilic proteins in order to maintain structural stability and activity at high temperatures. The nonpolar AAs methionine (M) and valine (V) have almost the same frequency in thermophilic and non-thermophilic proteins, indicating that these amino acids have no significant effect on thermostability.

Moreover, we also used aromaticity values to analyze aromatic AAs in thermophilic and non-thermophilic proteins. Aromaticity is the relative frequency of aromatic AAs [56]. Figure 2 shows the aromaticity value of thermophilic proteins and their non-thermophilic orthologous. As the figure shows, except for nitrogen regulatory protein, thioredoxin, and chemotaxis protein CheW, most of the thermophilic proteins have high aromaticity, which indicates that aromatic AAs are preferred in thermophilic proteins. Aromatic AAs form stable aromatic interactions, which contribute to thermal stability. Anderson et al. and Serrano et al. [57,58] have reported that a pair of aromatic interactions contribute 0.5 to 1.4 kcal/mol energy, which means that the increase in aromaticity in thermophilic proteins helps to endow proteins with thermal stability. It is also confirmed by protein engineering methods that the introduction of aromatic clusters on the surface of protein can improve stability in the protein. Kannan et al. [59] have analyzed aromatic clusters in 26 thermophilic proteins and their non-thermophilic orthologous. They found that thermophilic proteins have higher aromatic clusters. These aromatic clusters were able to produce pairwise interactions, which may be crucial to hinder the thermal denaturation of the protein structure.

The secondary structure is a folded structure formed by hydrogen bonds between partially positive hydrogen atoms and partially negative nitrogen atoms in the backbone [60–62]. Common secondary structure elements include coil, helix, sheet, and turn. The secondary structures of all 10 pairs of proteins are shown in Figures 3 and 4, and their percentage is represented in Table 1.

Table 1. Percentage of secondary structures in thermophilic (Ther) and non-thermophilic (non-Ther) proteins.

	Helix (%)		**Sheet (%)**		**Coil (%)**		**Turn (%)**	
	Ther	**Non-Ther**	**Ther**	**Non-Ther**	**Ther**	**Non-Ther**	**Ther**	**Non-Ther**
Hsp40 Chaperones	6.67	6.38	40.00	31.91	48.89	43.62	4.44	18.09
Nitrogen Regulatory Protein	31.58	17.86	31.58	32.14	24.21	30.36	12.63	19.64
Cold Shock Protein	0.00	4.48	45.45	50.75	45.45	29.85	9.09	14.93
DNA-Binding Protein HU	47.06	51.32	22.35	25.00	24.71	9.21	5.88	14.47
Ribosome-Binding Factor A	41.51	38.89	16.04	14.81	33.96	32.41	8.49	13.89
NusG	19.21	0.00	24.86	43.55	37.29	35.48	18.64	20.97
Protein RecA	38.30	46.82	20.18	16.76	33.04	26.88	8.84	9.54
Thioredoxin	31.43	29.31	21.90	22.41	22.86	26.72	23.81	21.55
CheW	9.27	13.17	37.75	34.13	42.38	36.53	10.60	16.17
Adenylate Kinase	52.22	49.07	14.78	14.49	22.66	22.43	10.34	14.02

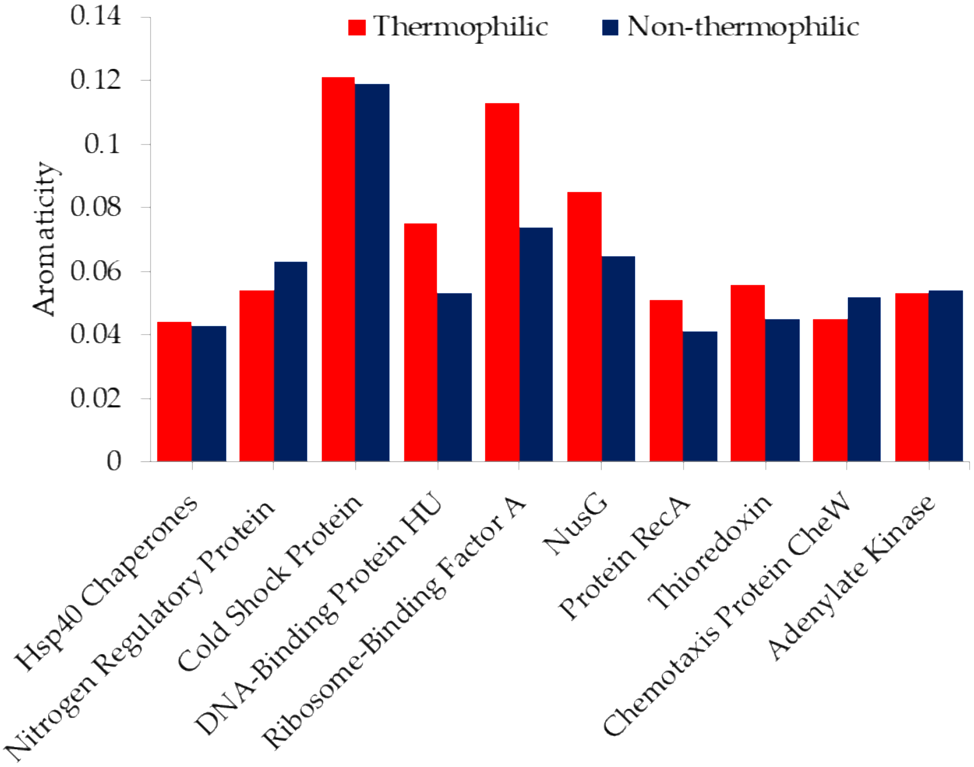

Figure 2. Aromaticity values of thermophilic proteins and non-thermophilic orthologous.

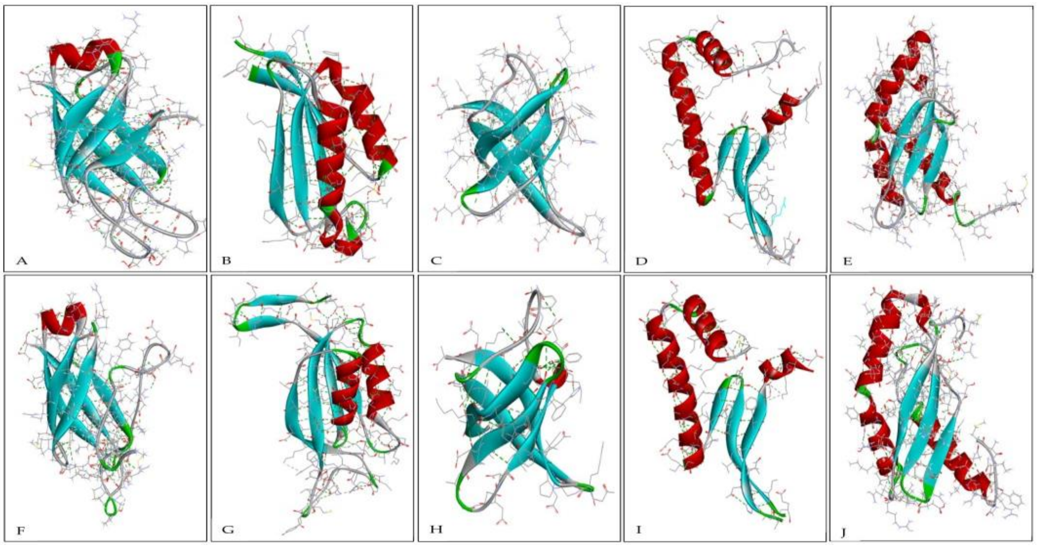

Figure 3. (**A**–**E**) show the structure of thermophilic protein Hsp40 chaperones, nitrogen regulatory protein, cold shock protein, DNA-binding protein HU and ribosome-binding factor A, respectively. (**F**–**J**) shows the structure of their non-thermophilic orthologs, respectively. Coil structure is shown in gray, helix in red, sheet in blue, and turn in green. Conventional hydrogen bonds are represented by the dotted green line, carbon-hydrogen bonds by the light green dotted line, and salt bridges by the dotted yellow line.

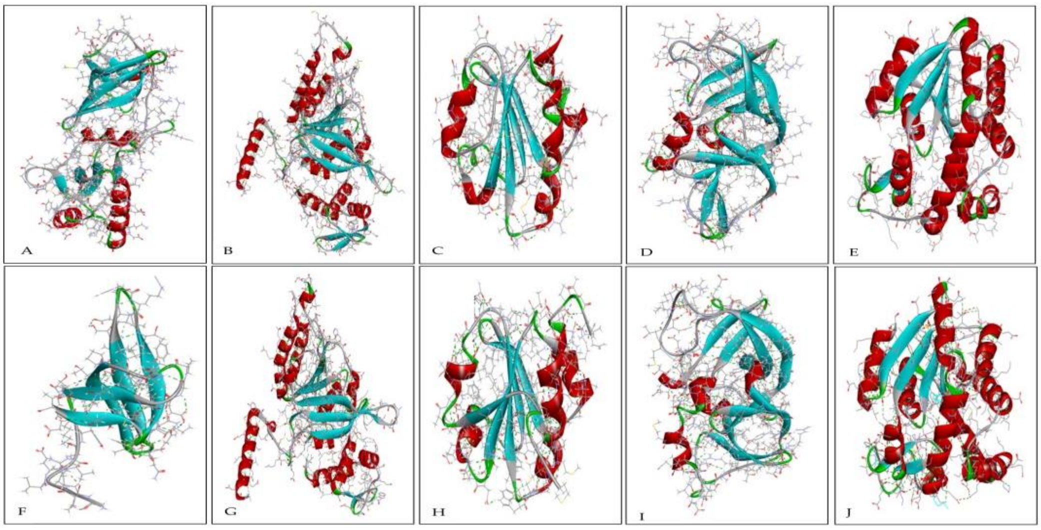

Figure 4. (**A**–**E**) show the structure of thermophilic protein transcription antitermination protein NusG, protein RecA, thioredoxin, chemotaxis protein CheW, and adenylate kinase, respectively. (**F**–**J**) shows the structure of their non-thermophilic orthologs, respectively. Coil structure is shown in gray, helix in red, sheet in blue, and turn in green. Conventional hydrogen bonds are represented by the dotted green line, carbon-hydrogen bonds by the light green dotted line, and salt bridges by the dotted yellow line.

A helix is a structure formed by hydrogen bonding between every fourth amino acid in a way that makes the side chain of residues directed outward and away from the helical axis, thus allowing the charged residues of the helix to form stable interactions with other elements and resulting in greater stability in the protein structure [63]. The secondary structure analysis showed that most thermophilic proteins, including Hsp40 chaperones, nitrogen regulatory protein, ribosome-binding factor A, transcription antitermination protein NusG, thioredoxin, and adenylate kinase, have a higher helix percentage as compared with their non-thermophilic orthologous (Figure 5A).

Sheet structure consists of two different regions of a polypeptide chain arranged side by side and connected by hydrogen bonds. In our analysis, it was found that the thermophilic proteins Hsp40 chaperones, ribosome-binding factor A, RecA, chemotaxis protein CheW, and adenylate kinase have a higher percentage of sheet structures (Figure 5B). The helix and sheet structures maximize the hydrogen bonding groups of the polypeptide and also allow the protein chains to be buried in the hydrophobic core, making it more compact. The compact protein structure is capable of hindering thermal denaturation [63–65]. It is inferred that the increase in the ratio of the helix and sheet structures contributes to thermostability in proteins.

In our analysis, thermophilic proteins have a higher percentage of coil structure than their non-thermophilic orthologous with the exception of nitrogen regulatory protein and thioredoxin (Figure 5C). A higher percentage of coil structure in thermophilic proteins is also reported in the literature [66]. Moreover, the percentage of the turn structure in non-thermophilic proteins except thioredoxin is higher than that of thermophilic proteins (Figure 5D). A turn is considered to allow a change of direction in protein chains. The lower percentage of turns in thermophilic proteins suggests that turns may not be conducive to thermostability in proteins, and the higher percentage of turns in non-thermophilic proteins may contribute to their non-thermostability.

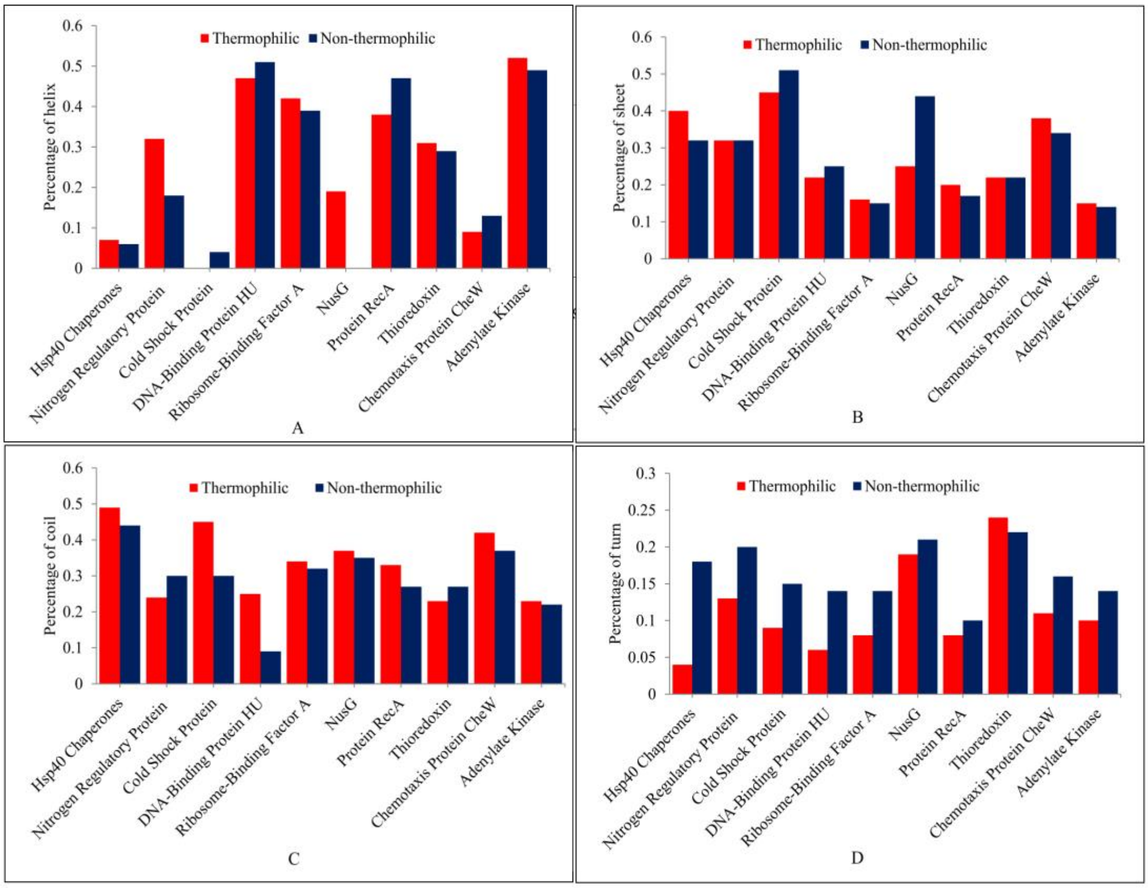

Figure 5. Figure (**A**) shows the percentage of the helix, (**B**) shows the percentage of the sheet, (**C**) shows the percentage of the coil, and (**D**) shows the percentage of the turn structure in thermophilic proteins and their non-thermophilic orthologous.

Hydrogen bonds are crucial for the stability of proteins by providing resistance against denaturation [47,64,67]. We analyzed the hydrogen bonds between thermophilic and non-thermophilic proteins (Table 2). The results showed that the thermophilic proteins including nitrogen regulatory protein, DNA-binding protein HU, ribosome-binding factor A, transcription antitermination protein NusG, and protein RecA have a ratio of hydrogen bonds > 0.5. The ratio of hydrogen bonds in Hsp40 chaperones is exactly 5. The increased number of hydrogen bonds in thermophilic protein implies that hydrogen bonds play some roles in protein thermostability. However, in other thermophilic proteins, such as cold shock protein, thioredoxin, CheW, and adenylate kinase, the ratio of hydrogen bonds was <50, indicating that hydrogen bonds are not the key factor in maintaining thermal stability.

Table 2. The ratio of hydrogen bonds and salt bridges in thermophilic protein.

Proteins	Hydrogen Bond Ratio	Salt Bridge Ratio
Hsp40 Chaperones	0.50	0.65
Nitrogen Regulatory Protein	0.70	0.63
Cold Shock Protein	0.44	1.00
DNA-Binding Protein HU	0.52	0.86
Ribosome-Binding Factor A	0.55	1.00
NusG	0.55	0.60
Protein RecA	0.66	0.43
Thioredoxin	0.48	0.31
CheW	0.38	0.25
Adenylate Kinase	0.48	0.40

Our analysis showed that most of the thermophilic proteins have a higher ratio of salt bridges than their non-thermophilic orthologous. In the present study, the thermophilic proteins Hsp40 chaperones, nitrogen regulatory protein, cold shock protein, DNA-binding protein HU, ribosome-binding factor A, and transcription antitermination protein NusG have a ratio of salt bridges >5, while the other thermophilic proteins including protein RecA, thioredoxin, CheW, and adenylate kinase showed a ratio of salt bridges <5 (Table 2). Salt bridges are electrostatic interactions between oppositely charged groups. Salt bridge is also an important factor contributing to thermostability in proteins. Salt bridges increase thermostability in proteins by the heat capacity change of unfolding ΔC_p [68–70].

In our analysis to elucidate the factors contributing to protein thermostability, we found some surprising results. The increase in aromaticity is beneficial to thermostability in proteins. However, we found that some thermophilic proteins, including nitrogen regulatory protein, thioredoxin, and chemotaxis protein CheW, have fewer aromaticity values than their non-thermophilic orthologous. Although chemotaxis protein CheW has a lower aromaticity value, the percentage of the secondary structures coil and sheet is higher, which may provide thermostability to this protein. Similarly, compared with their non-thermophilic orthologous, nitrogen regulatory protein has more hydrogen bonds and salt bridges, a shorter bond length, a wider DHA angle (Table 3), and a higher helix percentage, and thioredoxin has a short bond length and a higher helix structure, which may contribute to their thermostability.

Table 3. Average bend length and DHA angle in thermophilic and non-thermophilic proteins.

Proteins	Average Bond Length (Å)		Average DHA Angle	
	Thermophilic	Non-Thermophilic	Thermophilic	Non-Thermophilic
Hsp40 Chaperones	2.41	2.40	135.20	134.41
Nitrogen Regulatory Protein	2.94	3.08	109.53	106.45
Cold Shock Protein	2.25	2.99	135.85	108.01
DNA-Binding Protein HU	3.00	3.00	108.30	109.00
Ribosome-Binding Factor A	2.36	2.21	139.29	134.75
NusG	2.34	2.34	135.82	136.37
Protein RecA	3.02	3.04	109.23	108.02
Thioredoxin	2.24	2.31	141.92	142.46
CheW	2.30	2.36	133.03	142.06
Adenylate Kinase	3.04	3.05	109.44	108.51

We also found that thermophilic cold shock protein has no helix structure, while its non-thermophilic orthologous has 4% of the helix. However, thermophilic cold shock protein has more salt bridges, a shorter bond length, a wider DHA angle (Table 3), a higher aromaticity value, and more coil structures, which are favorable for imparting thermostability and may compensate for the absence of helix structure. Similarly, thermophilic proteins RecA, DNA-binding protein HU, and chemotaxis protein CheW also have fewer helix structures in them as compared with their non-thermophilic orthologous. However, thermophilic RecA has more hydrogen bonds, a slightly short bond length, and more coils and sheets, which may compensate for the reduction helix, thereby providing thermostability to this protein. CheW has a short bond length and a greater proportion of coil and sheet. The DNA-binding protein HU has a higher ratio of hydrogen bonds, a higher aromaticity value, and a greater proportion of coil. These factors may contribute to their thermostability.

In addition, thermophilic cold shock protein, transcription antitermination protein NusG, and the DNA binding protein HU have a lower sheet content. However, cold shock protein has more salt bridge, a shorter bond length, and a higher percentage of coil structure; transcription antitermination protein NusG has more hydrogen bond, a higher aromaticity value, more coil, and more helix structure; and the DNA binding protein HU has a higher aromaticity value and more coils in its secondary structure; all of which favor thermostability. The nitrogen regulatory protein and thioredoxin have the same proportion of sheets when compared with their non-thermophilic orthologous. This implies that sheet structure does not contribute to their thermostability. However, nitrogen regulatory protein and thioredoxin have more polar AAs, a shorter bond length, and a higher percentage of helix structure, factors which may contribute to their thermostability.

Moreover, thermophilic proteins including nitrogen regulatory protein and thioredoxin showed a slightly lower proportion of coil. However, thermophilic nitrogen regulatory protein contains more polar AAs; therefore, the number of hydrogen bonds and salt bridges is greater when compared with its non-thermophilic orthologous. These factors may lead to differences in thermostability between thermophilic nitrogen regulatory protein and its non-thermophilic orthologous. In thermophilic thioredoxin, the bond length is shortened and the proportion of helical structure is increased, which may be responsible for its thermostability.

Hydrogen bond analysis showed that thermophilic proteins, including cold shock protein, thioredoxin, chemotaxis protein CheW, and adenylate kinase, have fewer ratios of hydrogen bonds than their non-thermophilic orthologous. Although the ratio of hydrogen bonds in thermophilic proteins cold shock protein, chemotaxis protein CheW, and thioredoxin is lower, the bond length is shorter than in their non-thermophilic orthologous. It has been reported that a hydrogen bond with a shorter bond length is more stable than one with a wider bond length [71], implying that a shorter bond length may make the hydrogen bonds more stable. Hence these proteins are more thermally stable than their non-thermophilic orthologous. However, for the thermophilic adenylate kinase, not only is the ratio of hydrogen bonds small, but also the bond length is slightly greater than in their non-thermophilic orthologous, which is unfavorable for thermostability. Our analysis showed that thermophilic adenylate kinase has more helix, coil, and sheet, which may compensate for the smaller ratio of hydrogen bonds in this protein.

In salt bridge analysis, thermophilic proteins thioredoxin, adenylate kinase, RecA, and chemotaxis protein CheW showed a smaller ratio of salt bridges than their non-thermophilic orthologous. In thioredoxin, the helix structure percentage is higher than in its non-thermophilic orthologous, and in adenylate kinase, helix, coil, and sheet structure are greater; these secondary structures may be the factors leading to their thermostability. Thermophilic chemotaxis proteins CheW and RecA showed a slightly shorter bond length and an increased percentage of coil and sheet in their secondary structures than did their non-thermophilic orthologous, which may contribute to their thermostability.

3. Materials and Methods

3.1. Data Collection

Thermophilic proteins were collected from our previous study on thermophilic proteins [32]. We searched non-thermophilic orthologous using BLAST (Basic Local Alignment Search Tool). We preferred the thermophilic and non-thermophilic orthologous pairs with a high difference in optimum growth temperature (OGT), which could provide us with a clear cause for thermostability. For obtaining thermophilic and non-thermophilic protein pairs with more OGT differences, we considered thermophilic proteins with OGT > 60 °C and their non-thermophilic with OGT < 40 °C to keep the OGT difference at least 20 °C between thermophilic proteins and their non-thermophilic orthologous. As a result, 10 pairs were obtained, which were listed in Table 4. All the analyses were performed on these data. The analyses included sequence-based analysis and structure-based analysis.

Table 4. Thermophilic proteins and their non-thermophilic orthologous.

Protein Name	PDB ID	Organism Name	OGT
Hsp40 chaperones	6PRP	*Thermus thermophilus*	80
Hsp40 chaperones	6PQM	*Escherichia coli*	37
Nitrogen regulatory protein	2EG1	*Aquifex aeolicus*	85
Nitrogen regulatory protein	1PIL	*Escherichia coli*	37
Cold shock protein	1G6P	*Thermotoga maritima*	80
Cold shock protein	1CSP	*Bacillus subtilis*	25–35
DNA-binding protein HU	5EKA	*Thermus thermophilus*	85
DNA-binding protein HU	1MUL	*Escherichia coli*	37
Ribosome-binding factor A	2KZF	*Thermotoga maritima*	90
Ribosome-binding factor A	1KKG	*Escherichia coli*	37
Transcription antitermination protein NusG	2LQ8	*Thermotoga maritima*	80
Transcription antitermination protein NusG	2MI6	*Mycobacterium tuberculosis*	30–32
Protein RecA	3HR8	*Thermotoga maritima*	80
Protein RecA	4OQF	*Mycobacterium tuberculosis*	32
Thioredoxin	1RQM	*Alicyclobacillus acidocaldariu*	60–65
Thioredoxin	2L4Q	*Mycobacterium tuberculosis*	30–32
Chemotaxis protein CheW	1K0S	*Thermotoga maritima*	80
Chemotaxis protein CheW	2HO9	*Escherichia coli*	37
Adenylate kinase	2RGX	*Aquifex aeolicus*	85
Adenylate kinase	4K46	*Photobacterium profundum*	15

3.2. Sequence-Based Analysis

The sequence-based analysis included the occurrence frequency of amino acids (AA) and the relative occurrence frequency of aromatic amino acids (aromaticity) [72–76].

3.2.1. Occurrence Frequency of Amino Acids

In order to find sequence-based differences, we calculated the occurrence frequency of AA in thermophilic and non-thermophilic proteins [77]. The occurrence frequency of AA is the frequency of 20 amino acids in a protein sequence, which is given by:

$$f(t) = \frac{N(t)}{N}, t \in \{A, C, D, \ldots, Y\} \tag{1}$$

where $f(t)$ is the frequency of amino acid t, $N(t)$ is the number of amino acid t present in the protein sequence, and N is the length of the protein sequence [78].

3.2.2. Aromaticity

Aromaticity is a relative occurrence of aromatic amino acids (phenylalanine, tyrosine, and tryptophan) in a protein. The aromaticity value of a protein can be calculated by the formula given below:

$$Aromaticity = \sum_{i=1}^{20} \gamma_i f_i \tag{2}$$

where f_i represents the relative frequency of amino acid i, γ_i is taken as 1 when the amino acid is aromatic, and γ_i is taken as 0 when the amino acid is not aromatic amino acid [56].

3.3. Structure-Based Analysis

The structure-based analysis included the analysis of the secondary structure of proteins, namely coil, loop, helix, and turn, and the analysis of hydrogen bonds and salt bridges and their bond length and bond angle. To visualize and analyze the secondary structure of the thermophilic and non-thermophilic proteins, a discovery studio visualizer was used, and the percentages of coils, sheets, helices, and turns were calculated [79] using the following formula:

$$p(sc) = \frac{nAAs(sc)}{N}, sc \in \{helix, coil, sheet, coil\} \tag{3}$$

where $p(sc)$ represents the percentage of secondary structure sc, $nAAs(sc)$ represents the number of amino acids in secondary structure sc, and N represents the total number of amino acids.

In addition, the ratios of hydrogen bonds and salt bridges were calculated by the following formula:

$$ratio = \frac{number(Ther)}{[number(Ther) + number(non - Ther)]} \tag{4}$$

where *number(Ther)* and *number(non-Ther)* are the numbers of hydrogen bonds or salt bridges in the thermophilic proteins and non-thermophilic proteins, respectively. Moreover, to compare the strength of hydrogen bonds and salt bridges between thermophilic and non-thermophilic proteins, the average bond angle and average bond length were also calculated.

4. Conclusions

Enzymes are excellent biocatalysts and have many applications in scientific research and industry. In recent years, the use of biocatalysts on an industrial scale has increased. Biocatalysts are more sustainable, efficient, selective, and less environmentally and physiologically toxic compared to traditional chemical catalysts. However, industrial processes are carried out at higher temperatures, and stability of biocatalysts at such temperatures is a major concern. The best way to solve this problem is to use enzymes produced by thermophiles or to design thermostable enzymes.

Thermophiles are organisms that can survive at elevated temperatures. Thermophiles produce thermally stable proteins. Understanding thermal stability in these proteins is essential to theoretically describe the principles behind protein thermostability as well as to design the thermostable proteins/enzymes that can meet the demand of industrial processes. To arrive at the principles behind protein thermostability, we analyzed the sequences, secondary structures, hydrogen bonds, salt bridges, bond lengths, bond angles, and aromaticity values of 10 thermophilic proteins and their non-thermophilic orthologous. The analysis shows that the frequencies of polar AAs glutamic acid, histidine, lysine, arginine, and tyrosine are higher in thermophilic proteins, which may provide thermostability to protein through the formation of hydrogen bonds, salt bridges, and other long- and short-range interactions. Moreover, the nonpolar AA proline is more common in thermophilic proteins. The rigid structure of the proline may play a role in reducing the entropy of the main chain and in resisting the protein's unfolding. In addition, thermophilic proteins have a higher frequency of aromatic AAs which form aromatic interactions. The aromatic interactions are also valuable for providing thermostability to the proteins. In the secondary structure, the increase in the proportion of coil, helix, and sheet structure has an important contribution to thermostability in the proteins. In addition, an increased number of hydrogen bonds and salt bridges, their shorter bond length, and their wider bond angle also offer thermal stability to the protein structure. In some proteins, all the above-mentioned factors work individually, but in other proteins, these factors work together in a subtle combination. Therefore, it can be concluded that the basis of thermostability is complex and is affected by different factors alone or in different combinations.

Author Contributions: Conceptualization, H.L.; methodology, Z.A.; formal analysis, Z.A.; data curation, Z.A.; writing—original draft preparation, Z.A.; writing—review and editing, H.L.; visualization, Z.A. and H.Z.; supervision, H.L. and L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grants from the Basic Research Program of the Sichuan Science and Technology Foundation (2021YJ0223) and by Medico-Engineering Cooperation Funds from the University of Electronic Science and Technology of China (zygx2021yglh202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, S.; Yang, X.; Yang, S.; Zhu, M.; Wang, X. Technology prospecting on enzymes: Application, marketing and engineering. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209017. [CrossRef] [PubMed]
2. Cao, Y.; Li, X.; Ge, J. Enzyme Catalyst Engineering toward the Integration of Biocatalysis and Chemocatalysis. *Trends Biotechnol.* **2021**, *39*, 1173–1183. [CrossRef] [PubMed]
3. Hauer, B. Embracing nature's catalysts: A viewpoint on the future of biocatalysis. *ACS Catal.* **2020**, *10*, 8418–8427. [CrossRef]
4. Wu, S.; Snajdrova, R.; Moore, J.C.; Baldenius, K.; Bornscheuer, U.T. Biocatalysis: Enzymatic synthesis for industrial applications. *Angew. Chem. Int. Ed.* **2021**, *60*, 88–119. [CrossRef]
5. Sheldon, R.A.; Brady, D. Broadening the scope of biocatalysis in sustainable organic synthesis. *ChemSusChem* **2019**, *12*, 2859–2881. [CrossRef]
6. Chapman, J.; Ismail, A.E.; Dinu, C.Z. Industrial applications of enzymes: Recent advances, techniques, and outlooks. *Catalysts* **2018**, *8*, 238. [CrossRef]
7. Atalah, J.; Cáceres-Moreno, P.; Espina, G.; Blamey, J.M. Thermophiles and the applications of their enzymes as new biocatalysts. *Bioresour. Technol.* **2019**, *280*, 478–488. [CrossRef] [PubMed]
8. Ravindran, R.; Jaiswal, A.K. Enzymes in Bioconversion and Food Processing. In *Enzymes in Food Technology*; Springer: Cham, Switzerland, 2018; pp. 19–40.
9. Guerrand, D. Lipases industrial applications: Focus on food and agroindustries. *OCL—Oilseeds Fats Crops Lipids* **2017**, *24*, D403. [CrossRef]
10. Liu, X.; Kokare, C. Microbial Enzymes of Use in Industry. In *Biotechnology of Microbial Enzymes*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 267–298.
11. Sanchez, S.; Demain, A.L. Useful Microbial Enzymes—An Introduction. In *Biotechnology of Microbial Enzymes*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 1–11.
12. Satterfield, C.N. *Heterogeneous Catalysis in Industrial Practice*; McGraw-Hill: New York, NY, USA, 1991.
13. Littlechild, J.A. Enzymes from extreme environments and their industrial applications. *Front. Bioeng. Biotechnol.* **2015**, *3*, 161. [CrossRef]
14. Synowiecki, J. Some applications of thermophiles and their enzymes for protein processing. *Afr. J. Biotechnol.* **2010**, *9*, 7020–7025.
15. Buchsbaum, C.; Schmidt, M.U. Rietveld refinement of a wrong crystal structure. *Acta Crystallogr. Sect. B Struct. Sci.* **2007**, *63*, 926–932. [CrossRef] [PubMed]
16. Taylor, S.J.; McCague, R.; Wisdom, R.; Lee, C.; Dickson, K.; Ruecroft, G.; O'Brien, F.; Littlechild, J.; Bevan, J.; Roberts, S.M. Development of the biocatalytic resolution of 2-azabicyclo [2.2. 1] hept-5-en-3-one as an entry to single-enantiomer carbocyclic nucleosides. *Tetrahedron Asymmetry* **1993**, *4*, 1117–1128.
17. Singh, R.; Kumar, M.; Mittal, A.; Mehta, P.K. Microbial enzymes: Industrial progress in 21st century. *3 Biotech* **2016**, *6*, 1–15.
18. Liszka, M.J.; Clark, M.E.; Schneider, E.; Clark, D.S. Nature versus nurture: Developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3*, 77–102. [PubMed]
19. Siddiqui, K.S. Some like it hot, some like it cold: Temperature dependent biotechnological applications and improvements in extremophilic enzymes. *Biotechnol. Adv.* **2015**, *33*, 1912–1922. [PubMed]
20. Liu, Q.; Xun, G.; Feng, Y. The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnol. Adv.* **2019**, *37*, 530–537.
21. Ao, C.; Yu, L.; Zou, Q. Prediction of bio-sequence modifications and the associations with diseases. *Brief. Funct. Genom.* **2021**, *20*, 1–18. [CrossRef]
22. Shang, Y.; Gao, L.; Zou, Q.; Yu, L. Prediction of drug-target interactions based on multi-layer network representation learning. *Neurocomputing* **2021**, *434*, 80–89. [CrossRef]
23. Loladze, V.V.; Ibarra-Molero, B.; Sanchez-Ruiz, J.M.; Makhatadze, G.I. Engineering a thermostable protein via optimization of charge—Charge interactions on the protein surface. *Biochemistry* **1999**, *38*, 16419–16423.
24. Asial, I.; Cheng, Y.X.; Engman, H.; Dollhopf, M.; Wu, B.; Nordlund, P.; Cornvik, T. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat. Commun.* **2013**, *4*, 2901.
25. Rigoldi, F.; Donini, S.; Redaelli, A.; Parisini, E.; Gautieri, A. Review: Engineering of thermostable enzymes for industrial applications. *APL Bioeng.* **2018**, *2*, 011501. [PubMed]

26. Liu, R.; Liang, L.; Lacerda, M.P.; Freed, E.F.; Eckert, C.A. Advances in Protein Engineering and Its Application in Synthetic Biology. In *New Frontiers and Applications of Synthetic Biology*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 147–158.
27. Li, Y.; Cirino, P.C. Recent advances in engineering proteins for biocatalysis. *Biotechnol. Bioeng.* **2014**, *111*, 1273–1287. [PubMed]
28. Chandler, P.G.; Broendum, S.S.; Riley, B.T.; Spence, M.A.; Jackson, C.J.; McGowan, S.; Buckle, A.M. Strategies for Increasing Protein Stability. In *Protein Nanotechnology*; Springer: Cham, Switzerland, 2020; pp. 163–181.
29. Vieille, C.; Zeikus, G.J. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1–43. [PubMed]
30. Sterpone, F.; Melchionna, S. Thermophilic proteins: Insight and perspective from in silico experiments. *Chem. Soc. Rev.* **2012**, *41*, 1665–1676. [PubMed]
31. Pucci, F.; Rooman, M. Physical and molecular bases of protein thermal stability and cold adaptation. *Curr. Opin. Struct. Biol.* **2017**, *42*, 117–128.
32. Ahmed, Z.; Zulfiqar, H.; Khan, A.A.; Gul, I.; Dao, F.-Y.; Zhang, Z.-Y.; Yu, X.-L.; Tang, L. iThermo: A Sequence-Based Model for Identifying Thermophilic Proteins Using a Multi-Feature Fusion Strategy. *Front. Microbiol.* **2022**, *13*, 790063.
33. Zhang, D.; Xu, Z.-C.; Su, W.; Yang, Y.-H.; Lv, H.; Yang, H.; Lin, H. iCarPS: A computational tool for identifying protein carbonylation sites by novel encoded features. *Bioinformatics* **2021**, *37*, 171–177.
34. Tang, H.; Zhao, Y.W.; Zou, P.; Zhang, C.M.; Chen, R.; Huang, P.; Lin, H. HBPred: A tool to identify growth hormone-binding proteins. *Int. J. Biol. Sci.* **2018**, *14*, 957–964. [CrossRef]
35. Ao, C.; Zhou, W.; Gao, L.; Dong, B.; Yu, L. Prediction of antioxidant proteins using hybrid feature representation method and random forest. *Genomics* **2020**, *112*, 4666–4674. [CrossRef]
36. Zulfiqar, H.; Yuan, S.-S.; Huang, Q.-L.; Sun, Z.-J.; Dao, F.-Y.; Yu, X.-L.; Lin, H. Identification of cyclin protein using gradient boost decision tree algorithm. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4123–4131.
37. Lin, H.; Chen, W. Prediction of thermophilic proteins using feature selection technique. *J. Microbiol. Methods* **2011**, *84*, 67–70. [PubMed]
38. Kumar, S.; Tsai, C.-J.; Nussinov, R. Factors enhancing protein thermostability. *Protein Eng.* **2000**, *13*, 179–191. [PubMed]
39. Clackson, T.; Wells, J.A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383–386. [PubMed]
40. Bogan, A.A.; Thorn, K.S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1–9. [PubMed]
41. Jeffrey, G.A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer Science & Business Media: Berlin, Germany, 2012.
42. Russell, R.J.; Ferguson, J.M.; Hough, D.W.; Danson, M.J.; Taylor, G.L. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* **1997**, *36*, 9983–9994. [PubMed]
43. Vogt, G.; Woell, S.; Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **1997**, *269*, 631–643.
44. De Farias, S.T.; Bonato, M.C.M. Preferred codons and amino acid couples in hyperthermophiles. *Genome Biol.* **2002**, *3*, 1–18. [CrossRef]
45. Querol, E.; Perez-Pons, J.A.; Mozo-Villarias, A. Analysis of protein conformational characteristics related to thermostability. *Protein Eng. Des. Sel.* **1996**, *9*, 265–271.
46. Haney, P.; Konisky, J.; Koretke, K.; Luthey-Schulten, Z.; Wolynes, P. Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus Methanococcus. *Proteins Struct. Funct. Bioinform.* **1997**, *28*, 117–130.
47. Li, W.; Zhou, X.; Lu, P. Structural features of thermozymes. *Biotechnol. Adv.* **2005**, *23*, 271–281.
48. Gromiha, M.M.; Pathak, M.C.; Saraboji, K.; Ortlund, E.A.; Gaucher, E.A. Hydrophobic environment is a key factor for the stability of thermophilic proteins. *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 715–721.
49. Pace, C.N.; Fu, H.; Fryar, K.L.; Landua, J.; Trevino, S.R.; Schell, D.; Thurlkill, R.L.; Imura, S.; Scholtz, J.M.; Gajiwala, K. Contribution of hydrogen bonds to protein stability. *Protein Sci.* **2014**, *23*, 652–661. [PubMed]
50. Trevino, S.R.; Scholtz, J.M.; Pace, C.N. Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* **2007**, *366*, 449–460. [PubMed]
51. Mattos, C. Protein-water interactions in a dynamic world. *Trends Biochem. Sci.* **2002**, *27*, 203–208. [PubMed]
52. Nishio, Y.; Nakamura, Y.; Kawarabayasi, Y.; Usuda, Y.; Kimura, E.; Sugimoto, S.; Matsui, K.; Yamagishi, A.; Kikuchi, H.; Ikeo, K.; et al. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens. *Genome Res.* **2003**, *13*, 1572–1579. [PubMed]
53. Catanzano, F.; Barone, G.; Graziano, G.; Capasso, S. Thermodynamic analysis of the effect of selective monodeamidation at asparagine 67 in ribonuclease A. *Protein Sci.* **1997**, *6*, 1682–1693.
54. Sælensminde, G.; Halskau, Ø.; Jonassen, I. Amino acid contacts in proteins adapted to different temperatures: Hydrophobic interactions and surface charges play a key role. *Extremophiles* **2009**, *13*, 11–20.
55. Kumwenda, B.; Litthauer, D.; Bishop, Ö.T.; Reva, O. Analysis of protein thermostability enhancing factors in industrially important thermus bacteria species. *Evol. Bioinform.* **2013**, *9*, 327–342.
56. Lobry, J.; Gautier, C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* **1994**, *22*, 3174–3180. [CrossRef]
57. Serrano, L.; Bycroft, M.; Fersht, A.R. Aromatic-aromatic interactions and protein stability: Investigation by double-mutant cycles. *J. Mol. Biol.* **1991**, *218*, 465–475.

58. Anderson, D.E.; Hurley, J.H.; Nicholson, H.; Baase, W.A.; Matthews, B.W. Hydrophobic core repacking and aromatic—Aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117 → Phe. *Protein Sci.* **1993**, *2*, 1285–1290. [CrossRef] [PubMed]
59. Kannan, N.; Vishveshwara, S. Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13*, 753–761. [CrossRef] [PubMed]
60. Liu, D.; Li, G.; Zuo, Y. Function determinants of TET proteins: The arrangements of sequence motifs with specific codes. *Brief. Bioinform.* **2019**, *20*, 1826–1835. [CrossRef]
61. Xu, B.F.; Liu, D.Y.; Wang, Z.R.; Tian, R.X.; Zuo, Y.C. Multi-substrate selectivity based on key loops and non-homologous domains: New insight into ALKBH family. *Cell. Mol. Life Sci.* **2021**, *78*, 129–141. [CrossRef] [PubMed]
62. Zulfiqar, H.; Sun, Z.-J.; Huang, Q.-L.; Yuan, S.-S.; Lv, H.; Dao, F.-Y.; Lin, H.; Li, Y.-W. Deep-4mCW2V: A sequence-based predictor to identify N4-methylcytosine sites in *Escherichia coli*. *Methods* **2022**, *203*, 558–563. [CrossRef] [PubMed]
63. Yakimov, A.; Afanaseva, A.; Khodorkovskiy, M.; Petukhov, M. Design of stable α-helical peptides and thermostable proteins in biotechnology and biomedicine. *Acta Nat.* **2016**, *8*, 70–81. [CrossRef]
64. Hubbard, R.E.; Haider, M.K. Hydrogen bonds in proteins: Role and strength. *eLS* **2010**. [CrossRef]
65. Gromiha, M.M.; Oobatake, M.; Sarai, A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys. Chem.* **1999**, *82*, 51–67. [CrossRef]
66. Szilágyi, A.; Závodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. *Structure* **2000**, *8*, 493–504. [CrossRef]
67. Vieira, D.S.; Degreve, L. An insight into the thermostability of a pair of xylanases: The role of hydrogen bonds. *Mol. Phys.* **2009**, *107*, 59–69. [CrossRef]
68. Chan, C.-H.; Yu, T.-H.; Wong, K.-B. Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS ONE* **2011**, *6*, e21624. [CrossRef] [PubMed]
69. Lee, C.-W.; Wang, H.-J.; Hwang, J.-K.; Tseng, C.-P. Protein thermal stability enhancement by designing salt bridges: A combined computational and experimental study. *PLoS ONE* **2014**, *9*, e112751. [CrossRef] [PubMed]
70. Missimer, J.H.; Steinmetz, M.O.; Baron, R.; Winkler, F.K.; Kammerer, R.A.; Daura, X.; Van Gunsteren, W.F. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci.* **2007**, *16*, 1349–1359. [CrossRef] [PubMed]
71. Jeffrey, G.A.; Jeffrey, G.A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, NY, USA, 1997; Volume 12.
72. Kurata, H.; Tsukiyama, S.; Manavalan, B. iACVP: Markedly enhanced identification of anti-coronavirus peptides using a dataset-specific word2vec model. *Brief. Bioinform.* **2022**, *23*, bbac265. [CrossRef] [PubMed]
73. Manavalan, B.; Patra, M.C. MLCPP 2.0: An Updated Cell-penetrating Peptides and Their Uptake Efficiency Predictor. *J. Mol. Biol.* **2022**, *434*, 167604. [CrossRef] [PubMed]
74. Basith, S.; Lee, G.; Manavalan, B. STALLION: A stacking-based ensemble learning framework for prokaryotic lysine acetylation site prediction. *Brief. Bioinform.* **2022**, *23*, bbab376. [CrossRef]
75. Charoenkwan, P.; Nantasenamat, C.; Hasan, M.M.; Manavalan, B.; Shoombuatong, W. BERT4Bitter: A bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics* **2021**, *37*, 2556–2562. [CrossRef]
76. Malik, A.; Subramaniyam, S.; Kim, C.B.; Manavalan, B. SortPred: The first machine learning based predictor to identify bacterial sortases and their classes using sequence-derived information. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 165–174. [CrossRef]
77. Zheng, L.; Liu, D.Y.; Yang, W.; Yang, L.; Zuo, Y.C. RaacLogo: A new sequence logo generator by using reduced amino acid clusters. *Brief. Bioinform.* **2021**, *22*, bbaa096. [CrossRef]
78. Chen, Z.; Zhao, P.; Li, F.; Leier, A.; Marquez-Lago, T.T.; Wang, Y.; Webb, G.I.; Smith, A.I.; Daly, R.J.; Chou, K.-C. iFeature: A python package and web server for features extraction and selection from protein and peptide sequences. *Bioinformatics* **2018**, *34*, 2499–2502. [CrossRef]
79. Zheng, L.; Liu, D.Y.; Li, Y.A.; Yang, S.Q.; Liang, Y.C.; Xing, Y.Q.; Zuo, Y.C. RaacFold: A webserver for 3D visualization and analysis of protein structure by using reduced amino acid alphabets. *Nucleic Acids Res.* **2022**, *50*, W633–W638. [CrossRef] [PubMed]