RSC Advances

REVIEW

View Article Online
View Journal | View Issue

Cite this: RSC Adv., 2016, 6, 115252

Protein thermostability engineering

H. Pezeshgi Modarres,[ac] M. R. Mofrad[ab] and A. Sanati-Nezhad*[cd]

The use of enzymes for industrial and biomedical applications is limited to their function at elevated temperatures. The principles of thermostability engineering need to be implemented for proteins with low thermal stability to broaden their applications. Therefore, understanding the thermal stability modulating factors of proteins is necessary for engineering their thermostability. In this review, first different thermostability enhancing strategies in both the sequence and structure levels, discovered by studying the natural proteins adapted to different conditions, are introduced. Next, the progress in the development of various computational methods to engineer thermostability of proteins by learning from nature and introducing several popular tools and algorithms for protein thermostability engineering is highlighted. Further discussion includes the challenges in the field of protein thermostability engineering such as the protein stability–activity trade-off. Finally, how thermostability engineering could be instrumental for the design of protein drugs for biomedical applications is demonstrated.

Received 1st July 2016
Accepted 7th November 2016

DOI: 10.1039/c6ra16992a

www.rsc.org/advances

1. Introduction

Enzymes are the primary catalytic agents conducting chemical reactions within cells. Enzymes' functions have not been limited to cells but are used in a range of applications because of their favorable features such as high specificity and activity. However, natural enzymes cannot usually meet the needs of industrial conditions such as the harsh environment in chemical industries, because they are evolved to function in their native conditions. Protein engineering has revolutionized the application of naturally available enzymes for different applications and led to the development of commercially available enzymes. The aim of protein engineering is to improve different features of functional proteins, such as the stability and activity, to overcome their natural limitations. Stability engineering mainly develops longer stability at harsh conditions such as elevated temperatures, high pH values, and high concentrations of salts. In this review, thermostability engineering will be discussed, which is a special subset of protein engineering, with the focus on the engineering of proteins/enzymes to overcome the natural limitations of their stability against temperature.

Increasing the temperature by only a few degrees higher than the normal functioning temperature of proteins leads to unfolding and structural changes, and impacts their function.[1,2] In addition to the structural and functional changes, the hydrophobic amino acids (AAs), normally buried inside the protein structure, are exposed to solvents and hydrophobic AAs from other proteins and form aggregates, leading to irreversible unfolding.[3]

However, stability of proteins at higher temperatures has always been attractive for different reasons. Heat treatment methods can be used to purify enzymes after their production within mesophilic hosts. In addition, higher thermostability of proteins is associated with their stability to other destabilizing agents such as guanidinium hydrochloride and solvents. Finally, conducting an enzymatic chemical reaction is more favorable at higher temperatures, where a lower viscosity and a higher diffusivity and solubility of the substrate can increase the reaction yield and minimize the risk of contamination by mesophilic species. In addition, understanding thermostabilizing and destabilizing factors is important not only for protein thermostability engineering (PTE) with different applications, but also to figure out the origin of genetic diseases because pathogenic missense mutations result in misfolding and destabilizing effects.[4]

The thermostability engineering of natural enzymes has been implemented by assessing their stability response to temperature changes. Such assessment for naturally occurring proteins has extended our knowledge about the thermostabilizing factors used by nature over thousands of years to evolve enzymes for adapting the ever-changing environment. There are enzymes in nature, called extro-enzymes, that can function at extreme conditions such as high salt concentrations (halozymes), highly alkaline conditions (alkanozymes), as well

[a]Molecular Cell Biomechanics Laboratory, Departments of Bioengineering and Mechanical Engineering, University of California Berkeley, 208A Stanley Hall, Berkeley, CA 94720-1762, USA

[b]Physical Biosciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

[c]BioMEMS and Bioinspired Microfluidic Laboratory, Department of Mechanical and Manufacturing Engineering, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada, T2N 1N4

[d]Center for BioEngineering Research and Education, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada, T2N 1N4. E-mail: amir.sanatinezhad@ucalgary.ca; Tel: +1 403 220 7708

as at high temperature and pressure.[5,6] Hyper/thermophilic proteins have been isolated from both natural (such as hydrothermal vents) and industrial (such as geothermal power plants) sources.[7] It is worth noting that the stable temperature of proteins belonging to hyper and thermophilic organisms is not essentially the same as their optimum growth temperature (OGT). Although most of the enzymes characterized from these organisms possess the optimal activity at temperatures around the OGT of the organism, their cell bound and extracellular proteins such as saccharidases have optimal activity at temperatures higher than the OGT of the host microorganism. For example, the optimum activity of amylopullulanase occurs at 117 °C, whereas the OGT of its hosting microorganism (*Thermococcus litoralis*) is reported to be as high as 88 °C.[8] In addition, to study the behavior of protein at elevated temperatures, we should consider that although the upper limit of temperature for their survival is not definitely known, biomolecules such as metabolites and AAs become unstable at above 110 °C and key stabilizing interactions within protein structures, mainly hydrophobic interactions, are significantly weakened.[9] The development of accurate methods to engineer proteins viable at elevated temperatures broadens the application of enzymes considerably for different purposes.

2. Thermostabilization strategies

Different studies have been conducted to characterize the modulating factors of the thermostability of proteins. There are two primary resources of data needed to learn about thermostability: (1) naturally occurring sequences with diverse degrees of thermostability, and (2) the mutagenesis data.

Here, a set of thermostabilization strategies used by nature to enhance the thermostability of proteins is summarized. These strategies have been studied by comparing of proteins with different thermostabilities. Nevertheless, these strategies can be useful as starting points for engineering a target protein sequence to enhance its thermostability.[7] It should be noted that the comparative studies cannot provide universal rules of protein thermostability modulation, given the fact that the results of the comparative studies vary among protein families. However, they can be considered as the family specific thermostability governing modes.[10,11] The results of the systematic comparison have not been very successful for the thermostability assessment.[12] Furthermore, hyper/thermophilic and mesophilic homologs are highly similar because: (1) typically, their sequence similarity is in the range of 40–85% or higher,[13] (2) they have both superposable three-dimensional (3D) structures,[14] and (3) their catalytic mechanism is the same.[15] From now on in this review, the comparative results are summarized to two levels of the sequence and the structure. Our knowledge of these two analysis levels is used to design effective PTE strategies.

2.1. Sequence level comparison

Finding a correlation between the composition of AAs in protein sequences and their thermostability has been one of the primary approaches for engineering the thermal stability of proteins.[16] However, the rapid growth in DNA sequencing technologies has clarified that the difference in the composition of AAs in mesophilic and thermophilic sequences is not as significant as initially expected.[17] It was concluded that the thermostability modulating factors cannot be expressed only by identifying the differences in the amino acid composition.[17] The first research addressing the difference between mesophilic and thermophilic sequences indicated a higher hydrophobicity and residue volume, fewer uncharged polar residues, and higher charged AAs in thermophilic proteins.[18] Advances in genome sequencing technologies have been supplying more data at the sequence level for different types of micro-organisms, and have, therefore, provided more statistics for comparison purposes with higher resolutions. The hydrophobic content is higher in thermophiles compared with mesophiles, which in consequence increases both the hydrophobicity and rigidity of the protein.[19] In particular, isoleucine (Ile) and valine (Val) show higher frequencies in the thermophilic proteins.[19,20] Although some studies indicate that glycine (Gly) has a lower frequency in the thermophilic proteins because there are voids within the protein structure,[21] other reports show no difference in Gly presence in mesophilic and thermophilic proteins.[22] Proline (Pro) has been demonstrated to have a higher frequency in thermophilic proteins[22,23] so that an increase in Pro content has been used as a strategy to enhance protein thermostability.[24] On the other hand, methionine (Met) has a lower frequency in thermophilic proteins compared to that in mesophiles.[20,25] Several reports indicate that uncharged polar AAs, mainly serine (Ser), glutamine (Gln), cysteine (Cys), and threonine, are less frequent in thermophiles compared to mesophiles.[19,20,22] Cys, asparagine, and Gln are known as thermolabile AAs.[26] This explains their lower frequency in thermophilic proteins which minimizes the backbone cleavage, deamination, and oxidation at temperatures corresponding to thermophiles.[20,25] Therefore, the following replacement preference was reported from mesophile to thermophile: Met → Ala, Cys → Ala, Trp → Tyr, Met → Leu, Cys → Val, and Cys → Ile.[27] Generally, a higher frequency of charged AAs is reported in thermophilic proteins.[22,28] The most common explanation for such an increase in the charged amino acid content is making the salt bridges stabilizing weak points on the protein structure against high temperatures.[29–34] In a combinatorial work, the $E + K/Q + H$ ratio has been reported as an indicator for the thermostability of proteins. The protein is considered as hyper-thermophilic for the ratios above 4.5, mesophilic for the ratios below 2.5, and thermophilic for the ratios between 3.2 and 4.6.[35]

A relatively higher frequency is reported for aromatic AAs in thermophilic proteins.[20,22] However, it is worth noting that the difference in preference of AAs between mesophilic and thermophilic proteins vary from one protein family to other.[28,36,37] In another report exploring all possible amino acid combinations, the sum of frequencies of Ile, arginine (Arg), glutamic acid (Glu), Val, tyrosine (Tyr), tryptophan (Trp), and leucine (Leu) (IVYWREL) AAs presented a correlation coefficient with the OGT of the microorganisms as high as 0.93 in 86 proteomes.[38]

In addition to analysis of the amino acid composition, the coupling patterns of AAs were assessed in terms of thermostability analysis in the sequence level.[39] Denoting the [X*d*Z] as the coupling between AAs X and Z with a distance *d*, the following amino acid patterns have been suggested to fit mesophilic proteins: [C(−4)L], [C(−3)L], [C(−2)L], [C2L], [C3L], [D(−5)T], [D(−4)T], [E(−8)T], [E(−4)T], [E1Q], [E3T], [E4T], [G(−3)Q], [K(−4)T], [K2T], and [K3T]. Similarly, the amino acid patterns of [C(−2)P], [C1P], [C3C], [C4C], [C6C], [C7C], [K(−7)E], [K(−4)E], [K3E], [K4E], and [H(−4)V] were suggested to fit thermophilic proteins.[39]

AA physicochemical properties. A systematic analysis of the differences in 48 physicochemical properties of AAs between mesophilic and thermophilic proteins indicated that the shape factor and Gibbs free energy of the hydration for native proteins are the most important properties of AAs for thermostability labeling of proteins.[40]

Occurrence of AAs on different structural regions. In a report that studied the occurrence of AAs following their solvent accessible surface area (SASA) in thermophilic proteins, a higher relative occurrence was reported for Gln and Ile at the fully exposed state; Arg, Glu, and AAs with lower solvation energy; ion pairs located on the 3/10 helix at the exposed state; flexible AAs at the partially exposed states; alanine (Ala) residues located on the 3/10 helix; Arg and Glu at the buried state; and cation–π interactions at the well-buried states. However, a lower relative occurrence was reported for Ala, Ser, and Val at the exposed state, Ser at the partially exposed state, Met at the buried state, and aspartic acid (Asp) and Gly at the well-buried state.[12]

2.2. Structure level comparison

Comparison of protein structures provides more useful and practical help because it considers the meaningful pair and group interactions between the different classes of AAs that are physically interpretable.

Disulfide bridges. Disulfide bridges make strong interactions and can stabilize the protein by decreasing the entropy of the unfolding state.[41]

Hydrophobic interactions. The effect of hydrophobic interactions to the protein stability *via* escaping from water media and forming and preserving hydrophobic–hydrophobic interactions was first shown by Kauzmann.[42] The hydrophobic cores have not only been shown to be essential for the general protein stability[43,44] but also as a principal stabilizing strategy for the PTE.[42,45,46] The hydrophobicity content has also been proposed to be informative for discriminating between mesophilic and thermophilic proteins.[47] As well as comparisons between mesophilic and thermophilic proteins, several mutagenesis studies proved that the protein stability could be altered significantly by mutations in the hydrophobic cores.[48] Nevertheless, the mechanism of stabilization/destabilization of such mutations is not well understood.[48] Some reports proposed that alteration in the stability was dependent on factors such as changes in the transfer free energy and neighboring residues in the structure.[48] In particular, the loss of hydrophobicity and disturbance of the well packed core residue side chains have been suggested to contribute in destabilizing the mutations in the hydrophobic cores.[46] Many experimental and theoretical works have also addressed the effect of mutation size from the small to large (and *vice versa*) substitution.[44,49,50] It was indicated that both the large to small (*i.e.*, by introducing cavities)[48,51] and small to large (*i.e.*, by introducing unfavorable contacts)[48,52] mutations could result in protein instability. Mutations should make specific interactions and optimize a hydrophobic core, thus the "small to large mutations" strategy cannot be used as a universal rule. However, L to I mutation usually do not show alteration in the protein stability.[53] This could reveal the importance of side chain hydrophobicity *versus* packing.[53] To gain more insight into the thermodynamics, $\Delta\Delta G$ of the substitution of a hydrophobic core member was correlated with features of hydrophobic cores such as the local packing density,[44] structural perturbations in neighboring atoms,[54] and the number of neighboring methyl and methylene groups.[55] However, whereas this approach is effective for several specific proteins, the generalization of these correlations was shown to be difficult.[44] Somewhat simpler correlations between $\Delta\Delta G$ and features of hydrophobic AAs and the consequent cavities and interactions cannot be simply generalized to all other protein families.[44] Besides experimental studies, computational techniques have also been utilized to address the effect of hydrophobic core mutations on the protein thermostability.[56]

Packing. Although better packing has been shown to contribute to the thermostability enhancement,[57] the role of packing of the interior and exposed residues is still under debate. Whereas Pack and Yoo[58] found no distinct difference between packing of exposed residues in mesophilic and thermophilic proteins, Glyakina *et al.*[59] recently showed that the exposed residues are more closely packed in thermophilic proteins than in mesophilic ones.

Aromatic interactions. The contribution of aromatic interactions in the protein folding and stability is significant and has been discussed for a long time.[60,61] Different forms of aromatic interactions have been detected including π–π, cation–π, aryl–sulfur, and carbohydrate–π interactions.[62] The cluster formation of two or more aromatic AAs has also been shown to be prevalent with a significant role in protein stability.[63,64] For Phe–Phe interactions on locations i and i + 4 of a helix, a free energy of −0.1 to −0.8 kcal mol^{-1} is reported, which is the greatest among π–π interactions,[65,66] and is almost the same magnitude as hydrophobic interactions such as Leu–Tyr[67] or Phe–Met.[68,69] The stability effect of Phe–Phe interaction also depends on its location on the protein structure, showing more stabilizing impact on the C-terminus compared to the central regions of a helix.[66] Aromatic clusters are also required for well folded hairpin structures.[70,71] In fact, Phe–Phe interactions are among the most abundant cross-stand pair interactions on β-strands[72] and are suggested as key interactions in the formation of amyloids.[73] Trp–Trp cross-strand pair interactions on a beta-hairpin such as tryptophan zipper (Trpzip) strongly increase the stability to exhibit a melting point temperature (T_m) of up to 78 °C.[74] Aromatic interactions have also been utilized for inducing specificity in folded protein structures, showing

features such as a narrow thermal denaturation range.[75,76] However, the cation–π-interaction tendency has been reported to have a wide range of −0.8 to −1.2 kcal mol^{-1} for Trp–His with almost no stabilizing effect on the protein structure for Phe–Lys/Arg interaction.[62,77–79]

Salt bridges and hydrogen bonds. Stronger networks of salt bridges and hydrogen bonds play a central role in the higher thermostability of hyper/thermophilic proteins compared to mesophilic proteins.[80] It has been suggested that the salt bridges are not the driving force of the protein folding and stability because there exists a limited number of these bridges in the protein structure that are not highly conserved.[81] However, other researchers have indicated a significant stabilizing role for salt bridges.[82] For T4 lysozyme, the stabilizing effect was calculated to be around 3–5 kcal mol^{-1} for a single salt bridge.[83] Although salt bridges have a positive effect on the thermostability they do not have a principal role. Furthermore, a larger number of salt bridges are not able to explain the increased number of the positively charged residues compared to the negatively charged ones.[84] It is argued that the number of charged AAs in thermophilic proteins increases in order to make more salt bridges. However, this needs further discussion and more detailed investigation. It is suggested that the surface charged residue interactions with solvent molecules at higher temperatures increase the stability. In addition to the effect of solvent interaction, a "negative design" is suggested to explain why the charged AAs are present more in hyper/thermophilics, whereas they do not necessarily form salt bridges.[85] A negative design means that during the evolution, proteins are selected in a way that partially unfolded states are less favorable because of the clustering of residues with identical charges with a consequent repelling effect.[85,86]

Prolines and decreasing the entropy of unfolding. The reduction in entropy of the protein unfolding is suggested to act as one stabilizing factor.[87] Whereas Gly has the highest conformational entropy, Pro has limited conformational states, leading to a lower conformational entropy.[88] Therefore, Glu → Pro is expected to enhance the stability of proteins. This technique has been used in other studies.[89] The most significant stabilizing effect has been reported for mutations located on turns or the N caps of helices. However, this kind of mutation has less stabilizing effect if it introduced unfavorable interactions or removed favorable interactions such as H-bonds.[90]

Inter-subunit interactions and oligomerization. Ion pairs and hydrophobic interactions have been shown to enhance the thermostability *via* inter-subunit interactions.[91–93]

Helix dipole stabilization. Helix dipoles are shown to be effective on the protein stability.[94] Locating the negatively charged residues close to the N-terminal, and the positively charged ones close to the C-terminal of helixes is shown to have a stabilizing effect.[95]

Docking of the N and C termini, and anchoring of loose ends. Protein N and C terminals and loops show the highest B-factors. Therefore, stabilizing these regions with the rest of the protein *via* H-bonds and ion pairs can result in thermostabilization of the protein. For loops, shortening the loop length is another strategy for thermostabilization.[96]

Metal binding. Binding of metal ions is also shown to alter the thermostability of proteins.[97]

Extrinsic parameters. Although hyper/thermostable proteins are inherently stable at elevated temperatures, some environmental factors such as salts[98,99] and substrate[100] can increase the thermostability of intracellular enzymes.

Secondary structure propensity. The secondary structure propensity of AAs is also suggested to affect the protein thermostability.[101] However, such a relationship is not supported by all the reports.[81]

2.3. Quantification of protein thermostability

The thermostability of proteins is addressed by measuring the thermodynamic and kinetic stabilities. The thermodynamic stability is usually expressed by the unfolding free energy (ΔG) of proteins and their T_m, whereas the kinetic stability is expressed by the half-life time ($t_{1/2}$) for enzymes at a specific temperature.[7] The unfolding free energy is typically reported to be between 5 and 15 kcal mol^{-1} for globular mesophilic proteins at 25 °C. There is a small difference between the unfolding free energy of hyper/thermophilic and mesophilic homologs, which is typically between 5 and 20 kcal mol^{-1}. However, mutations in enzymes with change in Gibbs free energy ($\Delta\Delta G$) of 3 to 6.5 kcal mol^{-1} may increase the melting temperature up to 12 °C.[102] Theoretically, there are three approaches for increasing the thermodynamic stability of proteins by manipulating the ΔG–T curve by:[103] (1) shifts towards higher ΔG, (2) shifts towards higher T, and (3) flattening.[103] Different combinations of these three approaches are used by hyper/thermophilic proteins to achieve a higher thermodynamic stability although the most common approach is the shifting of the curve towards higher ΔG values.[103]

3. Protein thermostability engineering

Although the reaction rates, in general, double by increasing the reaction temperature by 10 °C (Q10 rule), hyper/thermophilic and mesophilic homologous enzymes usually have the same activity at their corresponding physiological conditions.[7] Nevertheless, the observation of lower catalytic efficiency of hyper/thermophilic enzymes, compared to mesophilic, at mesophilic conditions has strongly suggested that there is a trade-off between the activity and thermostability.[7] A number of protein engineering studies showed the possibility of increasing the thermostability without compromising the activity.[104]

Traditional experimental methods particularly the random mutations method have been employed to engineer proteins with low thermostability. Given the improvement in the knowledge about proteins, specifically the difference between categories of thermostability, computational techniques nowadays contribute to improve the thermostability. In this section, some commonplace experimental methods are introduced and because this review is mainly devoted to exploring *in silico* methods, the computational methods are investigated in depth.

3.1. Experimental methods for protein thermostability enhancement

The stabilization of proteins has usually been conducted *via* genetic protein engineering (by introducing appropriate mutations) or chemical modifications.[105] Site specific mutations are usually used when the structure of the protein is known. However, for directed evolution techniques,[106,107] the structure of the target protein is not needed.[108,109] In addition, the extension of the protein *via* N- or C-terminals using a random peptide technique to improve the thermostability of proteins can also be used.[110,111]

Chemical modification techniques do not introduce any mutation. These techniques have been considered less during the last decade compared to genetic strategies.[105] The proteins of interest can be chemically altered by the covalent attachment of proteins to water soluble polymers.[112] Alternatively, the surface of the protein can be chemically modified using chemical groups. For example, the side chains of surface lysines were modified using citraconic anhydrides and as a consequence, the lysines with positive charges were replaced with carboxyl groups with negative charges.[113] The details of these experimental techniques have been reviewed elsewhere.[114,115]

3.2. Computational protein thermostability engineering techniques

Because the experimental methods used for studying the protein stability are usually costly and time-consuming, computational techniques are appropriate alternatives to predict the function and activity of proteins.[116] The most common computational technique used for the engineering of protein thermostability is called rational engineering. In this method, a hot spot should be first detected in a protein structure, followed by engineering an appropriate mutation to improve the thermostability. Hot spots could be mechanically weak regions on the structure, detected by B-factors or molecular dynamic (MD) simulations, or cavities within the protein structure composed of hydrophobic residues. Stabilizing substitutes can stabilize the hot spots by either making salt bridges or improving the hydrophobic cores. The comparison studies are considered to be the origin of the methods that can improve the stability such as introducing a salt bridge at the appropriate point on the structure. Alternatively, the concept of stability predictors introduces another category of computational methods for thermostability engineering.

Sequence-based engineering. All the analyses presented in previous sections were based on knowledge about the structure of the protein of interest. However, for a lot of (interesting) protein sequences, there is no structure available. However, there are plenty of characterized and annotated protein sequences available in public databases. Therefore, any analytic tool that can provide hints about the enhancement of the thermostability of proteins out of such a big collection of information would be of great interest.[117] As an alternative to structure-based engineering, sequence-based engineering, known as data driven engineering, has recently been attracting attention. Specifically, advances in DNA sequencing technologies and an increasing number of characterized protein sequences have made this approach more attractive.[118] This approach uses all the available homologous sequences to propose thermostabilizing mutations.[118]

Consensus concept. To extract thermostabilizing mutations out of homologous sequences, the consensus concept (CC) was introduced and has been used in many studies as the primary sequence-based PTE technique.[119–127] This method has been applied successfully for different proteins including phytases,[128] repeat proteins,[129–133] β-lactamase,[134] endoglucanase,[123] fibronectin type III (FN3) domains,[135] fluorescent proteins,[136] penicillin G acylase,[137] glucose dehydrogenase,[138] and α-amino ester hydrolases.[139] The logic behind this method is simple: using a multiple sequence alignment (MSA), non-consensus residues are substituted by consensus ones.[118] This method was first introduced by Pantoliano *et al.*[140] and a few years later Steipe provided a statistical explanation by making an analogy to the thermodynamic canonical ensemble.[141,142] The number of sequences is not important in this method. A successful PTE has even been reported using four sequences.[142,143] Because the functional residues essential for the protein folding and enzyme activity belong to the consensus pool,[144] this method does not compromise the stability and catalytic activity.[126,144,145] Furthermore, the thermostabilized proteins that are engineered by this approach have shown enhanced stability against water miscible organic solvents and the high concentration of kosmotropic and chaotropic salts.[125]

However, the CC method does not guarantee that all the proposed individual mutations can increase the thermostability. Reports indicate that the mutations suggested by the CC method are usually composed of stabilizing, neutral, and destabilizing mutations, which eventually counterbalance each other and produce an overall stabilizing effect.[121,126,145] Nevertheless, removing the destabilizing mutations has been shown to increase the thermostability.[118,121,145] Therefore, it is important to formulate strategies to refine the CC results and more precisely detect stabilizing and destabilizing candidates among the proposed set of mutations. In the next sections, three CC refinement methods that have been used by different groups will be introduced.

Analysis of residues' coupling. An important assumption behind the CC method is that the functions of AAs in all the positions are independent, whereas in reality, this is not true because residues interact to form the optimum structure and work cooperatively to gain the overall required functions.[146] Studies on indole-3-glycerol phosphate synthase and anthranilate phosphoribosyltransferase enzymes confirm that positions on the protein that are highly correlated with other sites are important for protein stability and function[147] and the mutation of such positions should be omitted from the library. For example, Sullivan *et al.* found that mutation of the coupled location to other residues were less likely to stabilize the protein in triosephosphate isomerase from *Saccharomyces cerevisiae*.[148,149] They suggested the removal of statistically correlated sites as an effective strategy to eliminate destabilizing mutations from the library.[148] However, it is worth noting that the analysis of the AA coupling needs big MSAs and this can

limit the application of this method for those protein families with a limited number of sequences.[127,148,149]

Comparing sequences with higher thermostability. There are a few reports which describe taking advantage of comparing the thermostability of target proteins with their thermophilic homologues.[104,150] The primary benefit of this method underlies the fact that the target protein is compared with the known thermally stable sequences, which results in mutations with a higher chance of thermostability enhancement. In addition, the sequences extracted from the thermophilic species are tolerant to a set of harsh conditions, where the high temperature is only part of it. Therefore, the mutations found through this method are expected to improve not only the thermostability but also the resistance to other harsh conditions such as intense pH. Nevertheless, this approach suffers from shortage of the number of species isolated from hyper/thermophilic regions compared to mesophiles.[104,150,151]

Structure-guided CC. For further refinement of CC results and to increase the chance of selecting thermostabilizing mutations, the rational analysis of the protein structure is one of the most popular approaches,[119,120] specifically for a target sequence with a few available homologues or low sequence identities.[137–139,152] This method has been applied successfully for enzymes such as penicillin G acylase,[137] glucose dehydrogenase,[138] α-amino ester hydrolases,[139] and pullulanase[152] with a higher rate of thermostabilizing mutations detection compared to the conventional CC method. In all these studies, after building a mutation library using the traditional CC method, the function or stability disturbing mutations were eliminated using the following procedure. First of all, to decrease the risk of mutation on the function and activity of the proteins, mutation sites should be far away from the important residues such as those located at the active sites (6–10 Å away). Then, to preserve the secondary structure of the protein, the secondary structure propensity should be taken into account. For example, helix destabilizing substitutions were eliminated from the list of mutations. Finally, mutations that could disturb the existing salt bridges or hydrogen bonds were eliminated from the list. As an example, Polizzi *et al.*[137] decreased the number of mutations from 109 to 21 mutations using this method. Further complementary analysis can be added to this procedure for further refinement of the results using available experimental data such as B-factor analysis, analysis of the water-exposed surface, subunit interactions, and Ramachandran plots.[138,139,152]

Structure-based engineering. At high temperatures, mechanically weak regions of proteins are the most likely unfolding initiators. Therefore, locating these fragile regions and strengthening them by appropriate mutations has been widely used for the PTE.[153–160] In this approach, first, the weak points are found on the protein structure by flexibility/rigidity analysis and then strengthened by genetic modifications such as adding salt bridges,[161,162] introducing disulfide bridges,[163–165] or incorporating Pro residues.[154,166–169]

B-factor analysis. From the data available for proteins with known 3D structures, characterized using X-ray crystallography, the relative flexibility of atoms on the protein structure can be studied by looking at their B-factors. The B-factor analysis of crystallographic structures has been used in several thermostability engineering studies.[124,154,170–172] The B-values may vary significantly between proteins depending on the structural refinement and the crystal quality. To select the best 3D structure for the protein B-factor analysis, high-resolution structures without any molecules attached to the protein with the minimum number of non-resolved residues should be considered.[173] In cases with several available Protein Data Bank (PDB) files for a protein or different subunits of an individual protein, it is possible that different structures present different B-factor scales. Therefore, for making the comparison possible between different PDB files, the B-factors should be taken from the BDB database[174] or standardized using algorithms such as B-FITTER.[175]

The B-factor of the CA atom of a residue is usually considered as the representative of the residue and is used to find the most flexible residues or regions on the protein structure. However, the measured B-factors belong to the crystal state of the proteins and the crystal packing affects the dynamic information extracted from the B-factors. Therefore, the B-factor does not represent the dynamics of protein residues in solution.[104] In addition, the crystallographic data are measured at temperatures around 110 K under which the protein may not essentially represent the physiological state.[176]

Molecular dynamics (MD) simulations. Whereas the experimental investigation is challenging for studying phenomena such as protein folding and unfolding, MD simulations have been very helpful. The protocols and applications of MD simulations are detailed elsewhere.[177,178] MD simulations have been used for analysis of the unfolding pathway of proteins at temperatures higher than the working physiological temperature to study the thermostability affecting factors and detecting the weak points.[151,154,179,180] The MD method is the most widely used computational technique for thermostability analysis and engineering of a variety of enzymes including lipases,[181–183] esterases,[184,185] α-amylases,[186] amidases,[187] xylanases,[167,180,188] adenylate kinases,[189] adenylosuccinate synthetase,[190] carboxylesterases,[191,192] xylose isomerase,[193] β-fructosidases,[194] phytases,[195] and ligases.[104] MD simulations for the thermostability analysis usually consist of a sequence of steps. First, a 3D structure of a target protein is prepared. The structure is characterized using nuclear magnetic resonance spectroscopy or X-ray crystallography, and if not available, generated using homology modeling. After addition of explicit water molecules and neutralizing ions to the system, the energy minimization is used to remove atomic clashes. Then, the system is equilibrated at room temperature. Finally, the production simulations are run at room temperature or higher temperatures and the trajectory is saved over time for use in the subsequent analyses for example, flexibility analysis. The length of simulations varies in the range of 2–100 ns depending upon the nature of the protein and the computational power.[196] The time step is usually set to 2 ps in typical biological systems. However, to prevent the unforeseen protein collapse at higher temperatures it is more effective to set it to 1 fs.[104,151] In some studies, simulations are run and analyzed only at room temperature to

calculate the B-factor using root mean square fluctuations (RMSF) and to determine the overall flexibility of the protein using the root mean square deviation (RMSD).[163,197] However, in the majority of MD simulations for the thermostability analysis, simulations are run at different temperatures ranging from 298 K to 600 K.[151,154,179,180] An important advantage of MD simulation is that the flexibility analysis is not only possible for individual residues but also for a sub-domain of the protein. Such sub-domain analysis is not feasible using the B-factor analysis.[104,151] For example, using MD simulations, Wang *et al.*[151] detected an susceptible to unfold loop region that was stabilized by mutations, leading to the thermostability improvement. In addition to targeting the hot-spots for the PTE, MD simulations have also been used to understand the thermostabilizing factors and mechanisms. It is implemented by studying the flexibility, secondary structure, hydrogen bonds, salt-bridges, SASA, RMSF, RMSD, metal ion binding sites, and radius of gyration at different temperatures for the thermo-resistant proteins and mutants.[155,182,195,196,198–201]

Different MD simulation packages such as NAMD,[202] GROMACS,[203] CHARMM,[204] or Amber[205] have been used for thermostability analysis and engineering studies. However, thermosensitive analysis using MD simulations can only find hot spots, which are usually located on the surface of a protein. They cannot provide useful information about clusters of hydrophobic residues that are usually buried inside the protein structure. These hydrophobic residues are the key for protein folding and preserve the stability.

The thermostability simulations addressed previously have been categorized as atom simulations meaning that all the atoms in the system including the protein and solvent, which usually make up the biggest portion of the system, are included. The incorporation of such a large number of particles in a system causes limitations in the simulation time length (usually less than 100 ns for thermostability simulations) and does not allow phenomena that may happen at longer time scales to be observed. Coarse-grained (CG) force fields can be utilized to tackle this limitation. In CG models, a group of atoms are represented by a single particle and can significantly decrease the number of particles that are subjected to the simulation. CG models lead to longer simulations with more sampling of protein dynamics and the exploration of the thermal unfolding process. Although several studies have used atomic MD simulations, the use of the CG model is rarely reported in the field of thermostability analysis and engineering.[206] Kalimeri *et al.*[206] used a CG model called optimized potential for efficient protein structure prediction (OPEP) to study the thermostability of a protein for hundreds of nanoseconds. They showed that the number of sub-states visited by the hyperthermophilic one is larger compared to the mesophilic homolog. They also showed slower dynamics with more resilient behavior against the temperature increment. There is still a lack of evidence about the successful performance and application of CG models for thermostability studies. New CG force fields need to be tested for different proteins.

Constraint network analysis (CNA). Constraint network analysis (CNA) runs rigidity/flexibility analysis, models thermal unfolding, and computes the local and global stability indices. This method uses a simplified representation of protein residue interactions and could be used as an alternative to MD simulations that may take much time and computational resources.[207–209] However, as far as is known to the authors, this method was rarely used in thermostability studies.[210]

Floppy inclusion and rigid substructure topography (FIRST). The FIRST algorithm represents the protein structure as a set of constraints over covalent bonds and angles, hydrogen bonds, and hydrophobic interactions, and further uses the constraint analysis to identify the rigid and flexible regions.[211] Although the FIRST algorithm has shown successful results for the flexibility prediction for some proteins,[212–214] more protein families should be tested by this technique to assess its performance.[173,215]

Although algorithms such as FIRST and CNA are much faster than MD simulations for the detection of flexible and rigid regions on the protein structure, they are simplified methods and do not take into account important details such as the effects of explicit solvent molecules and long range electrostatic interactions that are considered in the atomic MD simulations. Such differences can result in differences in the measured flexibilities calculated using simplified methods and MD simulations.[216]

Disulfide by design. This program uses energy and geometry constraints, extracted from known protein structures containing disulfide bonds, to locate potential sites for Cys substitutions leading to the formation of the disulfide bond.[217] This algorithm has been used in several studies not only for the thermostability analysis of thermophilic proteins[218] but also for protein engineering to improve the thermostability.[219–223]

Rosetta design. This program finds the sequences of AAs that can fit into a given protein structure with optimal packing, hydrogen bonding, and hiding of hydrophobic residues. The program uses the Monte Carlo optimization and simulated annealing to search the possible sequence space.[224] It also contains a package to estimate changes in the protein stability *via* single and multiple mutations.[225] The Rosetta design has been used to find stabilizing mutations for weak points detected using either MD simulations[154] or B-factor analysis.[168,169] Korkegian *et al.*[226] showed that this program can be utilized without a weak point detection step for further thermostabilization of the proteins.

Databases. The databases play a crucial role in the design and evaluation of engineering algorithms and tools by providing the needed data.

ProTherm. ProTherm is the most frequently used database for design and assessment of the stability change predictors for proteins. ProTherm contains numerical values for thermodynamic parameters for protein unfolding including changes in Gibbs free energy ($\Delta\Delta G$), heat capacity (ΔC_P), enthalpy (ΔH), and melting temperature (ΔT_m) and their changes upon mutations. Based on the latest update, it contains 12 561 single, 12 561 double, and 1132 multiple mutations extracted from 1902 references for 1040 proteins. ProTherm[227] has been used to design several prediction algorithms and tools including FoldX,[228] Prethermut,[229] PoPMuSiC,[230] iPTREE,[231] FQ-STAB,[232]

WET-STAB,[233] AUTO-MUTE,[234] CUPSAT,[235] MUpro,[236] and I-Mutant.[237] However, for developing stability predictors that are based on machine learning techniques, one should be aware that the majority of available experimental data are biased.[238] Whereas there are abundant data for some frequently occurring substitutions (such as X → Ala, where X can be any of the other 19 AAs), the accumulated data is very limited for some substitutions.[238] In addition, the majority of the reported mutations in a database such as ProTherm (around 70%)[227] are destabilizing which causes another bias, answering the question: "why is the accuracy of prediction of the de-stabilizing mutations higher than that of stabilizing mutations".[4]

DSDBASE. DSDBASE together with AnalyCys[239] are databases that provide information not only for native disulfides on proteins but also for the residue pairs that can stereochemically form disulfide bridges. This database can be used for PTE by introducing disulfide bridges for the analysis of native disulfide bonds in the database, if available, or by mutating the suggested locations on the target protein structure.[240]

MODEL. MODEL is another database that provides MD simulation trajectories for different proteins. Currently, it contains simulations for 1595 structures covering around 40% of the PDB protein structures with the probability of observing results by chance (*e*-value) of less than 10^{-5}.[241] One can use this database for flexibility analysis without running expensive MD simulations.

3.3. Stability predictors

The main feature of predictors is that only the information needed about the protein is its sequence or structure. Although there are predictors that are based on the protein sequence, their accuracy is less than the structure-based predictors.[242–247] However, even the structure-based predictors still suffer from limited accuracy. This, together with the fact that they do not provide sufficient rationalization for the thermostability analysis (specifically machine learning based predictors), makes them difficult to use for thermostability engineering with high confidence. For an efficient engineering of proteins, predictors are needed that can accurately predict the effect of any mutation on stability of the protein, preferably with a reasonable answer to the question of "why the mutation makes the protein thermostable or thermo-unstable".[242–247] Such predictors can also be helpful for better understanding the mechanisms underlying genetic diseases.[116,248] Different computational tools have been developed to predict the effect of single or multiple mutations on the protein stability by predicting the difference in the free energy of unfolding or the ΔT_m between the wild type and the mutated protein. Such computational methods can be classified into two groups: methods that use energy functions[249–258] and methods that utilize the machine learning approach.[249,259–262] Table 1 presents a list of available predictors. In a systematic evaluation of several stability predictors upon the mutations, Potapov *et al.*[263] compared CC/PBSA, EGAD, FoldX, I-Mutant2.0, Rosetta, Hunter, and combinations of them to predict the effect of 2156 mutations on the stability of proteins. Interestingly, the results showed that even the combination of methods did not significantly enhance the power of predictions. In another study, Khan and Vihinen[4] showed that I-Mutant (utilizing protein structure information), D-Mutant, and FoldX gave the most reliable results with almost similar accuracies. However, it is worth noting that even for these three predictors, the accuracy level was moderate (60%), and designing tools with a higher accuracy are still challenging. Both these studies suggest that there is still opportunity for the development of more accurate stability predictors.

4. Challenges of thermostability engineering

4.1. Absence of a protein structure

Many of the rational PTE methods need the 3D structure of the target protein. In other words, without knowing the protein structure the existing valuable tools cannot be used for protein engineering purposes. However, for most of the protein sequences available, there is no structure available in the sequence databases. To overcome this problem, two approaches are applicable: using the structure prediction methods[264] and using sequences for the prediction of structural features such as flexibility.[265]

Over the years, many algorithms have been developed for protein structure prediction, and these have been comprehensively reviewed elsewhere.[264] Among the protein structure prediction methods, homology modeling is one of the most popular methods used in several thermostability analysis and engineering studies.[166,266–272] In homology modeling, a 3D structure is generated for a target sequence using the structure of its homologous sequence(s) and packages such as SWISS-MODEL[273] and Modeller.[274] The constructed structure can then be used as the starting point for the PTE,[166,271,275–278] for the analysis of mutants of the target protein to identify the mechanisms of thermostabilization,[104,272,279–281] or for studying the origin of the thermostability for hyper/thermophilic proteins.[282–286]

In cases with no available structure among homologous sequences, it is possible to predict the contact maps using the sequence alignment of homologs.[287–291] The knowledge about the contact maps is very important, specifically for the detection of hydrogen bonds, salt bridges, and hydrophobic cores. In addition to the contact map prediction, other methods and algorithms have been developed to predict the specific structural features of a given sequence. For example, it is possible to determine the flexibility of residues only using the sequence of residues for detecting the potential weak points,[265,292–294] predicting the secondary structure, for considering the secondary structure's propensity for thermostability engineering,[295–297] and predicting the disulfide forming residues on the protein sequence.[298–301]

4.2. The stability–activity trade-off and flexibility/rigidity of enzymes

Enzyme function modulating mutations are reported to be destabilizing in general. The effect of mutation on the protein can be dramatic, first for catalytic residues, then for substrate

Table 1 Protein stability change predictors upon mutations[a]

Name	Input	Mutation	Method	Website address	Ref.
FoldX	Str.	Single/multiple	PE	http://foldxsuite.crg.eu/	228
Prethermut	Str.	Single/multiple	ML	http://www.mobioinfor.cn/prethermut	229
MLI	Str./Seq.	Single	ML	http://www.prc.boun.edu.tr/appserv/prc/mlsta/server.php	330
PoPMuSiC-2.0	Str.	Single	PE	http://babylone.ulb.ac.be/popmusic	230
iPTREE-STAB	Seq.	Single	ML	http://bioinformatics.myweb.hinet.net/iptree.htm	231
FQ-STAB	Seq.	Single	Fuzzy query	http://bioinformatics.myweb.hinet.net/fqstab.htm	232
WET-STAB	Seq.	Double	ML	http://bioinformatics.myweb.hinet.net/wetstab.htm	233
AUTO-MUTE	Str.	Single	PE	http://proteins.gmu.edu/automute	234
CUPSAT	Str.	Single	PE	http://cupsat.tu-bs.de	235
MUpro	Str./Seq.	Single	ML	http://www.igb.uci.edu/servers/servers.html	236
I-Mutant	Seq.	Single	ML	http://folding.biofold.org/i-mutant/i-mutant2.0.html	237
SDM	Str.	Single	PE	http://www-cryst.bioc.cam.ac.uk/~sdm/sdm.php	331
mCSM	Str.	Single	Graphs	http://structure.bioc.cam.ac.uk/mcsm	332
DUET	Str.	Single	ML	http://structure.bioc.cam.ac.uk/duet	333
iStable	Str./Seq.	Single	ML	http://predictor.nchu.edu.tw/iStable	334
INPS	Seq.	Single	ML	http://inps.biocomp.unibo.it	335
NeEMO	Str.	Single	Networks	http://protein.bio.unipd.it/neemo/	336
ENCoM	Str.	Single	Normal mode analysis	http://bcb.med.usherbrooke.ca/encom	337
EASE-MM	Seq.	Single	ML	http://sparks-lab.org/server/ease	338
MAESTRO	Str.	Single/multiple	ML	https://biwww.che.sbg.ac.at/maestro/web	339
STRUM	Str.	Single	PE	http://zhanglab.ccmb.med.umich.edu/STRUM/	340

[a] The second column shows the inputs needed for the prediction. Str. and Seq. represent structure and sequence, respectively. The third column defines the mutations predicted using the predictor. The forth column shows the methods used for the design of the predictor. ML and PE represent machine learning and potential energy function, respectively.

binding residues, and finally for surface residues with a non-adaptive evolutionary change.[302] The arrangement of active site residues is naturally unfavourable because they are usually charged or polar when located within hydrophobic clefts,[303] and they often have unfavourable backbone angles.[304] Therefore, a mutation in catalytic residues often arises with a significant stability increment and at the same time, sacrificing the original function or activity.[305–307] Wang *et al.* and others[308–310] have shown that the mutations resulted in new functions are usually destabilizing. Nevertheless, it is necessary to remember that most mutations are destabilizing, whatever the functional/catalytic role of the residues.[311,312]

The extensive experimental data have suggested that mesophilic enzymes are more flexible than their thermophilic homologues at ambient temperatures. Thus, the direct correlation between the rigidity and thermostability has been the primary working hypothesis so far,[313,314] confirmed by the MD and B-factor analysis.[315] However, this hypothesis is not supported by some other studies which state that the stability and rigidity are not necessarily correlated.[315,316] The enzymes with a higher activity than thermophilic homologues are expected to combine the local flexibility of their active sites (responsible for their higher activity) with a higher overall rigidity (which is the origin of their higher thermostability).[317]

The thermophilic enzymes have a lower activity around ambient temperature when compared to psychrophilic enzymes.[318–320] The higher activity of enzymes has been linked to their higher flexibility of active site residues during their evolution and this may cause unfavorable geometry and interactions. Such flexibility in the active site results in the effective transition state binding and subsequently reduces the activation energy barrier. However, the stability of active site is optimized in thermophilic homologues at the cost of activity.[321] Following the transition state theory, there are two possibilities for increasing the activity of an enzyme: decreasing the activation enthalpy change (ΔH) or increasing the activation entropy change (ΔS). Structurally, the decreased flexibility of active sites in thermophilic proteins can be related to the ΔH increment by considering the increased number of thermodynamically favorable interactions such as salt bridges and hydrogen bonds.[320–322] On the other hand, the higher flexibility results in a bigger population of conformational states and therefore a higher ΔS.[323]

It is also possible to improve the thermostability and catalytic efficiency simultaneously[278,324,325] or at least keep the catalytic activity close to its original state.[226] Such evidence helped to elucidate the next generation of *in silico* protein engineering tools and algorithms: simultaneous improvement of

thermostability and activity. During the last decade, the primary concern of scientists has been minimizing the unfavourable consequences of thermal-stabilizing mutations on enzymatic activity and now it is time to improve both with a minimum number of mutations. Interestingly, some of the available computational techniques that contributed to improving the enzymatic activity are very similar in function to thermostability engineering. For example, Liu *et al.*[326] used statistical coupling analysis to improve the activity of isopentenyl phosphate kinase. However, they did not report the effect of activity improving mutations on the stability of the enzyme.

5. Thermostability of protein drugs

Protein drugs have been very attractive to the pharmaceutical industry. Indeed, compared to small molecule drugs, protein drugs have a higher activity and specificity.[327] The number of engineered proteins and peptides used as therapeutics is growing including an antagonistic variant of human growth hormone (hGH); Genentech's Somavert 1 (pegvisomant), Trimeris' Fuzeon 1 (enfuvirtide), an inhibitor of human immunodeficiency virus (HIV) fusion derived from the viral protein gp41; Amgen's Aranesp 1 (darbepoetin alfa), a hyperglycosylated variant of erythropoietin (EPO); a fully human monoclonal anti-TNF-α antibody; and Abbott's Humira 1 (adalimumab).[328,329] However, some of the proteins eligible to be considered as therapeutics do not show needed or optimal features essential for a therapeutic protein. Therefore, the improvement of their favourable features by protein engineering will provide a huge opportunity for the development of more efficient protein therapeutics. A variety of strategies including mutations, fusions, and chemical modifications have been developed to engineer different features of therapeutic proteins such as their stability, solubility, binding affinity, and oligomerization.[329] Next, mutation-based genetic engineering will be focussed on, with the aim of enhancing the stability of protein therapeutics and to further show how PTE can be helpful in the detection of such mutations. Table 2 shows some of the genetically engineered protein therapeutics.

The stability of protein drugs is not only important for their shelf life and efficiency during the treatment but also for the production process.[327] A drug–protein with a higher stability lasts longer, therefore a less frequent usage with a lower dosage is needed.[341,342] In addition, the expression level of stable recombinant proteins is often higher, and its function during the manufacturing process is retained, leading to lower costs.[341] Whereas small molecules can be taken orally, lack of stability does not allow them to be taken orally, thereby they are traditionally used by injection.[343–345] Instability of proteins against different conditions makes their application as therapeutic agents challenging. Although thermostability, as discussed in this review, plays a crucial role in the global stability of drug proteins, it is not the only factor.[342] Other factors impact on the protein stability, *e.g.*, physical and chemical factors such as pH, ionic strength, shearing, shaking, solvents, additives, pressure, protein concentration, oxidation, deamidation, hydrolysis, isomerization, succinimidation, non-disulfide crosslinking, and deglycosylation.[327,346] In addition, protein pharmaceuticals should be stored at low temperatures or be freeze dried to preserve them for an acceptable lifetime.[347] The weak thermostability of protein drugs is one of the main reasons which necessitates maintaining these drugs at low temperatures during the storage and transportation.[346,348] Furthermore, aggregate formation can be considered as a common consequence of proteins' instability which may be caused by a change in temperature. The increase in the stability of the protein drugs can minimize the risk of aggregation.[347] Therefore, exploring the stabilizing factors to improve their desired features is a major step in the development of protein drugs.[327] Among the different strategies, genetic engineering is focussed on for the thermal stabilization of protein drugs, given that their sensitivity to temperature is the most critical origin of instability.[347]

The genetic modifications that improve the conformational thermostability of proteins can increase the global stability of protein drugs and even improve other favorable features.[349] For example, the pharmaceutical application of wild-type hGH is limited because of its low stability as a solution.[350] The thermal stability of this protein was increased by 16 °C using six to ten mutations, selected using computational techniques, which could modify core interactions. As another example, the mutant granulocyte colony stimulating factor (G-CSF), with 10 to 14 substitutions, detected by computational techniques, showed not only an increase up to 13 °C in thermal stability but also preserved its biological activity with a prolonged shelf life.[351] Human fibroblast growth factor 1 (FGF1) is another successful example of therapeutic protein engineering. FGF1 is a promising therapeutic candidate because of its osteogenic, angiogenic, and wound healing properties. Using the semi-rational approach, FGF1's mutants showed an increase up to 27 °C in thermostability with a prolonged *in vivo* half-life and an enhanced protease resistance.[327,352] Finally, mutations in the G-

Table 2 Genetically engineered protein-based drugs available on the market (adapted from [327])

Name	Protein	Mutation	Disease	Company
Betaseron®	Interferon β (IFN-β)	Cys-Ser substitution	Multiple sclerosis therapy	Bayer
Humalog®	Insulin lispro	Pro28lysine (Lys), Lys29Pro	Diabetes	Eli Lilly
NovoLog®	Insulin aspart	Pro28Asp	Diabetes	Novo Nordisk
Proleukin® (aldesleukin)	Interleukin 2 (IL-2)	Cys-Ser substitution	Metastatic renal cell carcinoma, metastatic melanoma	Prometheus

CSF not only increased the thermostability but also improved the shelf life by 5–10 fold, thus preserving its bioactivity.[351]

Genetic engineering methods need special care to avoid unfavorable side effects for practical aspects of protein therapeutics *in vivo*. As a general consideration, mutations increasing the stability of proteins may also affect their activity drastically in a negative way as discussed in the section on protein stability function trade-off. The stability and activity may decrease simultaneously, for example for D58A mutation in ribonuclease T1.[353] However, it is still not feasible to deal with this challenge to increase and preserve the original activity.[104,312] In other words, achieving such mutations through the process of random mutation is very rare, whereas rational engineering approaches can aid significantly in the selection of the mutations with the lowest negative effects on the activity.[104] For proteins with a potential application as drugs, in particular, the effect of the mutation on the activity is not the only issue that should be taken into account during genetic engineering. Another important subject is the effect of mutations on immunogenicity. A higher similarity of the engineered proteins to the endogenous human proteins has been reported by avoiding T- and B-lymphocyte reactivity as a strategy to overcome the immunogenicity problem.[343] To deal with this problem, a method called humanization was successfully utilized.[354–358] Using the CC method that has been shown to be useful to improve the thermostability and other favorable features of the protein, one can build up a mutation library and add constraints that increase the probability of similarity to the human sequence from the selected mutations. This is recognized as a practical strategy for engineering the proteins towards a higher stability with a minimal risk of the immunogenicity. Finally, the solubility is another important feature of drug proteins that may be influenced by mutations, specifically for hydrophobic/hydrophilic exchange mutations.[347] Because the improvement of hydrophilic interactions on the surface and optimization of hydrophobic interactions within the protein core are among the main stability engineering strategies, it is important to make absolutely sure that the thermostability enhancing mutations do not inversely affect the solubility of the target proteins.

The replacement of the free Cys residues, CC process, and introducing the disulfides are reported as three successful genetic engineering approaches to increase the stability of protein drugs.[327,329,341,342,352,354,359–370] Among them, the replacement of free Cys is most likely to be the most popular approach.[327,329,366–370] Such mutation prevents the aggregation and instability of proteins caused by the intermolecular disulfide bonds. This genetic modification has been successfully applied for commercially available IL-2 and (IFN-β) (Table 2).[369] It is worth noting that although Cys replacements can increase the half-life of the protein, it may inversely decrease the thermostability.[329,366] Another commonly used method is the CC process.[341,342,352,354,359–363] This method was used in human fibroblast growth hormone-1 and resulted in a 27 °C increase in its thermostability.[352] It also had a longer half-life with a stronger mitogenic activity.[352] Introducing disulfide bridges[354,364,365] and optimization of the interaction of AAs located on the protein core[350] have also been reported to enhance the stability of the protein drugs.

Correlating the mutation-based modifications with the alterations induced to the different features of a protein is usually difficult. Approaches such as the CC not only improve the thermostability but also benefit other therapeutic functions of the proteins such as the solubility and the expression level.[342,359,360] If the sequence set of proteins, used to form the consensus sequence, is selected carefully from the human homologs, or if special care is taken to the preserve the residues conserved within the human homologs, it can overcome the immunogenicity issue by converging it with the humanization method.[354–358] In addition, methods that use comparison with more stable homologous sequences (such as comparing them with hyper/thermophilic sequences) are in fact comparing the sequence with sequences that are not only working at high temperatures but also under other harsh conditions, such as intense pH, high salt concentration, and in the presence of organic solvents. Therefore, mutations selected by comparing the target sequence with homologs with a higher thermostability can also increase the stability of the protein against other harsh conditions. Finally, the thermostability enhancing strategies that stabilize flexible regions of the protein structure can also improve other favorable features of the protein drugs such as reducing the proteolytic susceptibility.[371] As the protease cleavage sites are usually placed on the flexible regions of the protein, the flexibility decreasing mutations may make the proteolysis prone site to be no longer a match for the putative protease.[371] This concept is in complete synergy with the thermostabilizing strategies such as the thermo-sensitive region enhancement and in a good example shows how the thermostabilization by structural modifications can lead to the enhancement of other stability factors.

6. Conclusions

Given the progress in thermostability engineering of proteins, there are still significant challenges in the areas of activity trade-off, rigidity, and drug properties. There are pressing needs to develop complex models that can cover different desired features, satisfy them by a reasonable trade off, and enhance all the favorable properties of proteins at the same time. It is also necessary to develop new sophisticated experimental evidence to find correlations between different properties. So far, only single parameter evaluations such as the correlation between various structure/sequence properties and the individual biochemical features such as the thermostability or activity have been investigated. Finding the relationship between the structure/sequence properties and multiple biochemical features, such as the thermostability and activity together needs numerous experimental data. High-throughput techniques using robotic systems or microfluidics are expected to be helpful in this context.

Most individual prediction/engineering methods used for thermostability analysis suffer from some limitations. More comprehensive and global models should be developed to minimize the risks by mixing different models to cover the

weaknesses of each other but meanwhile taking advantage of them. Rational engineering methods are mostly based on the structures although there are fewer structures available compared to the sequences. Incorporation of accurate 3D structure predictors can partially solve this challenge.

All the discussions about protein drugs together indicate that the suggested mutations developed using thermostability engineering methods could be a good starting point for designing protein drug engineering libraries. To protect or improve the function, activity, immunogenicity, and other important factors of the protein drugs, more constraints should be applied to the initially designed mutation library. This leads to the importance of a new generation of computational algorithms that can simultaneously take into account all of these issues.

Acknowledgements

The authors acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Alberta Innovates Bio Solutions for their support to this research.

References

1 K. Ghosh and K. Dill, *Biophys. J.*, 2010, **99**, 3996–4002.
2 Y.-G. Park, M.-C. Jung, H. Song, K.-W. Jeong, E. Bang, G.-S. Hwang and Y. Kim, *J. Biol. Chem.*, 2016, **291**, 1692–1702.
3 D. Volkin and C. Middaugh, *Stability of protein pharmaceuticals, part A. Chemical and physical pathways of protein degradation*, Plenum Press, New York, NY, 1992, pp. 215–247.
4 S. Khan and M. Vihinen, *Hum. Mutat.*, 2010, **31**, 675–684.
5 A. Lewin, A. Wentzel and S. Valla, *Curr. Opin. Biotechnol.*, 2013, **24**, 516–525.
6 L. Tazi, D. P. Breakwell, A. R. Harker and K. A. Crandall, *Extremophiles*, 2014, **18**, 525–535.
7 C. Vieille and G. J. Zeikus, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 1–43.
8 S. H. Brown and R. M. Kelly, *Appl. Environ. Microbiol.*, 1993, **59**, 2614–2621.
9 R. Jaenicke, *Biochemistry*, 1998, **63**, 312–321.
10 G. Vogt, S. Woell and P. Argos, *J. Mol. Biol.*, 1997, **269**, 631–643.
11 S. Chakravarty and R. Varadarajan, *FEBS Lett.*, 2000, **470**, 65–69.
12 S. P. Pack, T. J. Kang and Y. J. Yoo, *Appl. Biochem. Biotechnol.*, 2013, **171**, 1212–1226.
13 C. Vieille, J. M. Hess, R. M. Kelly and J. G. Zeikus, *Appl. Environ. Microbiol.*, 1995, **61**, 1867–1875.
14 D. Maes, J. P. Zeelen, N. Thanki, N. Beaucamp, M. Alvarez, M. H. D. Thi, J. Backmann, J. A. Martial, L. Wyns and R. Jaenicke, *Proteins: Struct., Funct., Bioinf.*, 1999, **37**, 441–453.
15 M. W. Bauer and R. M. Kelly, *Biochemistry*, 1998, **37**, 17170–17178.
16 P. Argos, M. G. Rossmann, U. M. Grau, H. Zuber, G. Frank and J. D. Tratschin, *Biochemistry*, 1979, **18**, 5698–5703.
17 B. Gerald, *Int. J. Pept. Protein Res.*, 1994, **43**, 97–106.
18 X.-X. Zhou, Y.-B. Wang, Y.-J. Pan and W.-F. Li, *Amino Acids*, 2008, **34**, 25–33.
19 S. Chakravarty and R. Varadarajan, *FEBS Lett.*, 2000, **470**, 65–69.
20 S. Kumar, C.-J. Tsai and R. Nussinov, *Protein Eng.*, 2000, **13**, 179–191.
21 N. Panasik, J. E. Brenchley and G. K. Farber, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 2000, **1543**, 189–201.
22 S. P. Pack and Y. J. Yoo, *J. Biotechnol.*, 2004, **111**, 269–277.
23 M. Sadeghi, H. Naderi-Manesh, M. Zarrabi and B. Ranjbar, *Biophys. Chem.*, 2006, **119**, 256–270.
24 B. Van den Burg, G. Vriend, O. R. Veltman, G. Venema and V. G. Eijsink, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 2056–2060.
25 Z. Xu, Y. Liu, Y. Yang, W. Jiang, E. Arnold and J. Ding, *J. Bacteriol.*, 2003, **185**, 4038–4049.
26 F. Catanzano, G. Barone, G. Graziano and S. Capasso, *Protein Sci.*, 1997, **6**, 1682–1693.
27 M. M. Gromiha, M. Oobatake and A. Sarai, *Biophys. Chem.*, 1999, **82**, 51–67.
28 S. Trivedi, H. Gehlot and S. Rao, *Genet. Mol. Res.*, 2006, **5**, 816–827.
29 A. Szilágyi and P. Závodszky, *Structure*, 2000, **8**, 493–504.
30 S. Fukuchi, K. Yoshimune, M. Wakayama, M. Moriguchi and K. Nishikawa, *J. Mol. Biol.*, 2003, **327**, 347–357.
31 H. Nakashima, S. Fukuchi and K. Nishikawa, *J. Biochem.*, 2003, **133**, 507–513.
32 N. F. Saunders, T. Thomas, P. M. Curmi, J. S. Mattick, E. Kuczek, R. Slade, J. Davis, P. D. Franzmann, D. Boone and K. Rusterholtz, *Genome Res.*, 2003, **13**, 1580–1588.
33 K. Suhre and J.-M. Claverie, *J. Biol. Chem.*, 2003, **278**, 17198–17202.
34 Y. Tanaka, K. Tsumoto, Y. Yasutake, M. Umetsu, M. Yao, H. Fukada, I. Tanaka and I. Kumagai, *J. Biol. Chem.*, 2004, **279**, 32957–32967.
35 S. T. Farias and M. Bonato, *Genet. Mol. Res.*, 2003, **2**, 383–393.
36 T. Kawashima, N. Amano, H. Koike, S.-i. Makino, S. Higuchi, Y. Kawashima-Ohya, K. Watanabe, M. Yamazaki, K. Kanehori and T. Kawamoto, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 14257–14262.
37 O. Fütterer, A. Angelov, H. Liesegang, G. Gottschalk, C. Schleper, B. Schepers, C. Dock, G. Antranikian and W. Liebl, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9091–9096.
38 K. B. Zeldovich, I. N. Berezovsky and E. I. Shakhnovich, *PLoS Comput. Biol.*, 2007, **3**, 62–72.
39 H. K. Liang, C. M. Huang, M. T. Ko and J. K. Hwang, *Proteins: Struct., Funct., Bioinf.*, 2005, **59**, 58–63.
40 M. M. Gromiha, M. Oobatake and A. Sarai, *Biophys. Chem.*, 1999, **82**, 51–67.
41 A. A. Dombkowski, K. Z. Sultana and D. B. Craig, *FEBS Lett.*, 2014, **588**, 206–212.

42 H. Chirakkal, G. C. Ford and A. Moir, *Protein Eng.*, 2001, **14**, 161–166.
43 J. L. Gilmore, X. Yi, L. Quan and A. V. Kabanov, *Journal of Neuroimmune Pharmacology*, 2008, **3**, 83–94.
44 M. Bueno, L. A. Campos, J. Estrada and J. Sancho, *Protein Sci.*, 2006, **15**, 1858–1872.
45 M. M. Gromiha, M. C. Pathak, K. Saraboji, E. A. Ortlund and E. A. Gaucher, *Proteins*, 2013, **81**, 715–721.
46 J. G. B. Northey, A. A. Di Nardo and A. R. Davidson, *Nat. Struct. Biol.*, 2002, **9**, 126–130.
47 A. Banerji and I. Ghosh, *Eur. Biophys. J. Biophys. Lett.*, 2009, **38**, 577–587.
48 H. Kono, M. Nishiyama, M. Tanokura and J. Doi, *Protein Eng.*, 1998, **11**, 47–52.
49 J. M. Chen and W. E. Stites, *Biochemistry*, 2001, **40**, 14004–14011.
50 F. Pucci, R. Bourgeas and M. Rooman, *Sci. Rep.*, 2016, **6**, 23257.
51 J. T. Kellis, K. Nyberg and A. R. Fersht, *Biochemistry*, 1989, **28**, 4914–4922.
52 E. Baldwin, J. Xu, O. Hajiseyedjavadi, W. A. Baase and B. W. Matthews, *J. Mol. Biol.*, 1996, **259**, 542–559.
53 B. Anil, S. Sato, J. H. Cho and D. P. Raleigh, *J. Mol. Biol.*, 2005, **354**, 693–705.
54 A. M. Buckle, P. Cramer and A. R. Fersht, *Biochemistry*, 1996, **35**, 4298–4305.
55 D. E. Otzen, M. Rheinnecker and A. R. Fersht, *Biochemistry*, 1995, **34**, 13051–13058.
56 U. D. Priyakumar, *J. Biomol. Struct. Dyn.*, 2012, **29**, 961–971.
57 W.-t. Li, R. A. Grayling, K. Sandman, S. Edmondson, J. W. Shriver and J. N. Reeve, *Biochemistry*, 1998, **37**, 10563–10572.
58 S. P. Pack and Y. J. Yoo, *Int. J. Biol. Macromol.*, 2005, **35**, 169–174.
59 A. V. Glyakina, S. O. Garbuzynskiy, M. Y. Lobanov and O. V. Galzitskaya, *Bioinformatics*, 2007, **23**, 2231–2238.
60 G. Dong, C. Vieille, A. Savchenko and J. G. Zeikus, *Appl. Environ. Microbiol.*, 1997, **63**, 3569–3576.
61 K. Ishikawa, M. Okumura, K. Katayanagi, S. Kimura, S. Kanaya, H. Nakamura and K. Morikawa, *J. Mol. Biol.*, 1993, **230**, 529–542.
62 M. L. Waters, *Pept. Sci.*, 2004, **76**, 435–445.
63 R. Bhattacharyya, U. Samanta and P. Chakrabarti, *Protein Eng.*, 2002, **15**, 91–100.
64 A. Thomas, R. Meurisse and R. Brasseur, *Proteins: Struct., Funct., Bioinf.*, 2002, **48**, 635–644.
65 M. L. Waters, *Curr. Opin. Chem. Biol.*, 2002, **6**, 736–741.
66 S. M. Butterfield, P. R. Patel and M. L. Waters, *J. Am. Chem. Soc.*, 2002, **124**, 9751–9755.
67 S. Padmanabhan, M. Jimenez, D. Laurents and M. Rico, *Biochemistry*, 1998, **37**, 17318–17330.
68 A. R. Viguera and L. Serrano, *Biochemistry*, 1995, **34**, 8771–8779.
69 B. J. Stapley, C. A. Rohl and A. J. Doig, *Protein Sci.*, 1995, **4**, 2383–2391.
70 J. F. Espinosa and S. H. Gellman, *Angew. Chem.*, 2000, **112**, 2420–2423.
71 S. Honda, N. Kobayashi and E. Munekata, *J. Mol. Biol.*, 2000, **295**, 269–278.
72 C. K. Smith and L. Regan, *Acc. Chem. Res.*, 1997, **30**, 153–161.
73 E. Gazit, *FASEB J.*, 2002, **16**, 77–83.
74 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5578–5583.
75 J. J. Skalicky, B. R. Gibney, F. Rabanal, R. J. Bieber Urbauer, P. L. Dutton and A. J. Wand, *J. Am. Chem. Soc.*, 1999, **121**, 4941–4951.
76 V. Souza, C. Ikegami, G. Arantes and S. Marana, *FEBS J.*, 2016, **238**, 1124–1138.
77 C. D. Tatko and M. L. Waters, *Protein Sci.*, 2003, **12**, 2443–2452.
78 C. D. Andrew, S. Bhattacharjee, N. Kokkoni, J. D. Hirst, G. R. Jones and A. J. Doig, *J. Am. Chem. Soc.*, 2002, **124**, 12706–12714.
79 Z. Shi, C. A. Olson and N. R. Kallenbach, *J. Am. Chem. Soc.*, 2002, **124**, 3284–3291.
80 P. I. de Bakker, P. H. Hunenberger and J. A. McCammon, *J. Mol. Biol.*, 1999, **285**, 1811–1830.
81 K. A. Dill, *Biochemistry*, 1990, **29**, 7133–7155.
82 A. S. Thomas and A. H. Elcock, *J. Am. Chem. Soc.*, 2004, **126**, 2208–2214.
83 D. E. Anderson, W. J. Becktel and F. W. Dahlquist, *Biochemistry*, 1990, **29**, 2403–2408.
84 G. Saelensminde, O. Halskau Jr and I. Jonassen, *Extremophiles*, 2009, **13**, 11–20.
85 I. N. Berezovsky, K. B. Zeldovich and E. I. Shakhnovich, *PLoS Comput. Biol.*, 2007, **3**, e52.
86 O. Halskau Jr, R. Perez-Jimenez, B. Ibarra-Molero, J. Underhaug, V. Munoz, A. Martinez and J. M. Sanchez-Ruiz, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8625–8630.
87 B. Matthews, H. Nicholson and W. Becktel, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 6663–6667.
88 D. Sriprapundh, C. Vieille and J. G. Zeikus, *Protein Eng.*, 2000, **13**, 259–265.
89 T. Nakai, K. Okada, S. Akutsu, I. Miyahara, S.-i. Kawaguchi, R. Kato, S. Kuramitsu and K. Hirotsu, *Biochemistry*, 1999, **38**, 2413–2424.
90 K. Watanabe, T. Masuda, H. Ohashi, H. Mihara and Y. Suzuki, *Eur. J. Biochem.*, 1994, **226**, 277–283.
91 R. N. Z. A. Rahman, S. Fujiwara, H. Nakamura, M. Takagi and T. Imanaka, *Biochem. Biophys. Res. Commun.*, 1998, **248**, 920–926.
92 C. Vetriani, D. L. Maeder, N. Tolliday, K. S.-P. Yip, T. J. Stillman, K. L. Britton, D. W. Rice, H. H. Klump and F. T. Robb, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12300–12305.
93 K. E. Neet and D. E. Timm, *Protein Sci.*, 1994, **3**, 2167–2174.
94 H. Nicholson, W. Becktel and B. Matthews, *Nature*, 1988, **336**, 651–656.
95 A. Harada, H. Yagi, A. Saito, H. Azakami and A. Kato, *Biosci., Biotechnol., Biochem.*, 2007, **71**, 2952–2961.
96 W. Grabarse, M. Vaupel, J. A. Vorholt, S. Shima, R. K. Thauer, A. Wittershagen, G. Bourenkov, H. D. Bartunik and U. Ermler, *Structure*, 1999, **7**, 1257–1268.

97 K. Kojoh, H. Matsuzawa and T. Wakagi, *Eur. J. Biochem.*, 1999, **264**, 85–91.
98 S. Shima, D. A. Hérault, A. Berkessel and R. K. Thauer, *Arch. Microbiol.*, 1998, **170**, 469–472.
99 S. Shima, C. Tziatzios, D. Schubert, H. Fukada, K. Takahashi, U. Ermler and R. K. Thauer, *Eur. J. Biochem.*, 1998, **258**, 85–92.
100 V. Wilquet, J. A. Gaspar, M. van de Lande, M. Van de Casteele, C. Legrain, E. M. Meiering and N. Glansdorff, *Eur. J. Biochem.*, 1998, **255**, 628–637.
101 A. M. Facchiano, G. Colonna and R. Ragone, *Protein Eng.*, 1998, **11**, 753–760.
102 S. Kawamura, Y. Kakuta, I. Tanaka, K. Hikichi, S. Kuhara, N. Yamasaki and M. Kimura, *Biochemistry*, 1996, **35**, 1195–1200.
103 H. Nojima, A. Ikai, T. Oshima and H. Noda, *J. Mol. Biol.*, 1977, **116**, 429–442.
104 H. Pezeshgi Modarres, B. D. Dorokhov, V. O. Popov, N. V. Ravin, K. G. Skryabin and M. Dal Peraro, *Biochemistry*, 2015, 3076–3085.
105 C. Ó'Fágáin, *Enzyme Microb. Technol.*, 2003, **33**, 137–149.
106 K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5618–5622.
107 W. P. Stemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10747–10751.
108 B. Morawski, S. Quan and F. H. Arnold, *Biotechnol. Bioeng.*, 2001, **76**, 99–107.
109 V. G. Eijsink, S. Gåseidnes, T. V. Borchert and B. van den Burg, *Biomol. Eng.*, 2005, **22**, 21–30.
110 T. Matsuura, T. Yomo and I. Urabe, *Methods Mol. Biol.*, 2002, **182**, 221–230.
111 J.-H. Liu, C.-F. Tsai, J.-W. Liu, K.-J. Cheng and C.-L. Cheng, *Enzyme Microb. Technol.*, 2001, **28**, 582–589.
112 L. Gomez and R. Villalonga, *Biotechnol. Lett.*, 2000, **22**, 1191–1195.
113 K. Khajeh, H. Naderi-Manesh, B. Ranjbar, A. akbar Moosavi-Movahedi and M. Nemat-Gorgani, *Enzyme Microb. Technol.*, 2001, **28**, 543–549.
114 R. D. Socha and N. Tokuriki, *FEBS J.*, 2013, **280**, 5582–5595.
115 C. Ó'Fágáin, *Protein Chromatography*, 2017, 101–129.
116 J. Thusberg and M. Vihinen, *Hum. Mutat.*, 2009, **30**, 703–714.
117 H. J. Wijma, R. J. Floor and D. B. Janssen, *Curr. Opin. Struct. Biol.*, 2013, **23**, 588–594.
118 J. F. Chaparro-Riggers, K. M. Polizzi and A. S. Bommarius, *Biotechnol. J.*, 2007, **2**, 180–191.
119 K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa and A. S. Bommarius, *Biotechnol. J.*, 2006, **1**, 531–536.
120 E. Vazquez-Figueroa, J. Chaparro-Riggers and A. S. Bommarius, *ChemBioChem*, 2007, **8**, 2295–2301.
121 M. Lehmann, L. Pasamontes, S. F. Lassen and M. Wyss, *Biochim. Biophys. Acta*, 2000, **1543**, 408–415.
122 M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. van Loon and M. Wyss, *Protein Eng.*, 2002, **15**, 403–411.
123 M. Anbar, O. Gul, R. Lamed, U. O. Sezerman and E. A. Bayer, *Appl. Environ. Microbiol.*, 2012, **78**, 3458–3464.
124 J. K. Blum, M. D. Ricketts and A. S. Bommarius, *J. Biotechnol.*, 2012, **160**, 214–221.
125 E. Vazquez-Figueroa, V. Yeh, J. M. Broering, J. F. Chaparro-Riggers and A. S. Bommarius, *Protein Eng., Des. Sel.*, 2008, **21**, 673–680.
126 M. Lehmann and M. Wyss, *Curr. Opin. Biotechnol.*, 2001, **12**, 371–375.
127 B. T. Porebski and A. M. Buckle, *Protein Eng., Des. Sel.*, 2016, 1–7.
128 M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes and A. P. van Loon, *Protein Eng.*, 2000, **13**, 49–57.
129 A. Kohl, H. K. Binz, P. Forrer, M. T. Stumpp, A. Plückthun and M. G. Grütter, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1700–1705.
130 L. K. Mosavi, D. L. Minor and Z.-y. Peng, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16029–16034.
131 P. Forrer, H. K. Binz, M. T. Stumpp and A. Plückthun, *ChemBioChem*, 2004, **5**, 183–189.
132 K. W. Tripp and D. Barrick, *Structure*, 2003, **11**, 486–487.
133 T. Kajander, A. L. Cortajarena and L. Regan, *Protein Design*, 2006, 151–170.
134 N. Amin, A. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti and V. Schellenberger, *Protein Eng., Des. Sel.*, 2004, **17**, 787–793.
135 S. A. Jacobs, M. D. Diem, J. Luo, A. Teplyakov, G. Obmolova, T. Malia, G. L. Gilliland and K. T. O'Neil, *Protein Eng., Des. Sel.*, 2012, **25**, 107–117.
136 M. Dai, H. E. Fisher, J. Temirov, C. Kiss, M. E. Phipps, P. Pavlik, J. H. Werner and A. R. Bradbury, *Protein Eng., Des. Sel.*, 2007, **20**, 69–79.
137 K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa and A. S. Bommarius, *Biotechnol. J.*, 2006, **1**, 531–536.
138 E. Vázquez-Figueroa, J. Chaparro-Riggers and A. S. Bommarius, *ChemBioChem*, 2007, **8**, 2295–2301.
139 J. K. Blum, M. D. Ricketts and A. S. Bommarius, *J. Biotechnol.*, 2012, **160**, 214–221.
140 M. W. Pantoliano, M. Whitlow, J. F. Wood, S. W. Dodd, K. D. Hardman, M. L. Rollence and P. N. Bryan, *Biochemistry*, 1989, **28**, 7205–7213.
141 B. Steipe, B. Schiller, A. Pluckthun and S. Steinbacher, *J. Mol. Biol.*, 1994, **240**, 188–192.
142 B. Steipe, *Protein Eng.*, 2004, **388**, 176–186.
143 M. W. Pantoliano, M. Whitlow, J. F. Wood, S. W. Dodd, K. D. Hardman, M. L. Rollence and P. N. Bryan, *Biochemistry*, 1989, **28**, 7205–7213.
144 M. Lehmann, L. Pasamontes, S. F. Lassen and M. Wyss, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 2000, **1543**, 408–415.
145 M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. G. M. van Loon and M. Wyss, *Protein Eng.*, 2002, **15**, 403–411.
146 V. Durani and T. J. Magliery, *Methods Enzymol.*, 2013, **523**, 237–256.
147 S. Dietrich, N. Borst, S. Schlee, D. Schneider, J.-O. Janda, R. Sterner and R. Merkl, *Biochemistry*, 2012, **51**, 5633–5641.

148 B. J. Sullivan, T. Nguyen, V. Durani, D. Mathur, S. Rojas, M. Thomas, T. Syu and T. J. Magliery, *J. Mol. Biol.*, 2012, **420**, 384–399.
149 B. J. Sullivan, V. Durani and T. J. Magliery, *J. Mol. Biol.*, 2011, **413**, 195–208.
150 Z. H. Xiao, H. Bergeron, S. Grosse, M. Beauchemin, M. L. Garron, D. Shaya, T. Sulea, M. Cygler and P. C. K. Lau, *Appl. Environ. Microbiol.*, 2008, **74**, 1183–1189.
151 K. Wang, H. Luo, J. Tian, O. Turunen, H. Huang, P. Shi, H. Hua, C. Wang, S. Wang and B. Yao, *Appl. Environ. Microbiol.*, 2014, **80**, 2158–2165.
152 S.-f. Li, J.-y. Xu, Y.-j. Bao, H.-c. Zheng and H. Song, *J. Biotechnol.*, 2015, **210**, 8–14.
153 X. Huang, D. Gao and C. G. Zhan, *Org. Biomol. Chem.*, 2011, **9**, 4138–4143.
154 J. C. Joo, S. P. Pack, Y. H. Kim and Y. J. Yoo, *J. Biotechnol.*, 2011, **151**, 56–65.
155 Y. T. Meharenna and T. L. Poulos, *Biochemistry*, 2010, **49**, 6680–6686.
156 S. Kundu and D. Roy, *J. Mol. Graphics Modell.*, 2010, **28**, 820–827.
157 S. Kundu and D. Roy, *J. Mol. Graphics Modell.*, 2009, **27**, 871–880.
158 M. B. A. Rahman, R. A. Karjiban, A. B. Salleh, D. Jacobs, M. Basri, A. L. T. Chor, H. A. Wahab and R. N. Z. R. A. Rahman, *Protein Pept. Lett.*, 2009, **16**, 1360–1370.
159 V. Spiwok, P. Lipovova, T. Skalova, J. Duskova, J. Dohnalek, J. Hasek, N. J. Russell and B. Kralova, *J. Mol. Model.*, 2007, **13**, 485–497.
160 M. Purmonen, J. Valjakka, K. Takkinen, T. Laitinen and J. Rouvinen, *Protein Eng., Des. Sel.*, 2007, **20**, 551–559.
161 J. Chen, H. Yu, C. Liu, J. Liu and Z. Shen, *J. Biotechnol.*, 2013, **164**, 354–362.
162 B. Fei, H. Xu, Y. Cao, S. Ma, H. Guo, T. Song, D. Qiao and Y. Cao, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 457–464.
163 M. G. Pikkemaat, A. B. Linssen, H. J. Berendsen and D. B. Janssen, *Protein Eng.*, 2002, **15**, 185–192.
164 S. Badieyan, D. R. Bevan and C. Zhang, *Biotechnol. Bioeng.*, 2012, **109**, 31–44.
165 Q. A. T. Le, J. C. Joo, Y. J. Yoo and Y. H. Kim, *Biotechnol. Bioeng.*, 2012, **109**, 867–876.
166 J. Tian, P. Wang, S. Gao, X. Chu, N. Wu and Y. Fan, *FEBS J.*, 2010, **277**, 4901–4908.
167 J. C. Joo, S. Pohkrel, S. P. Pack and Y. J. Yoo, *J. Biotechnol.*, 2010, **146**, 31–39.
168 H. S. Kim, Q. A. T. Le and Y. H. Kim, *Enzyme Microb. Technol.*, 2010, **47**, 1–5.
169 S. J. Kim, J. A. Lee, J. C. Joo, Y. J. Yoo, Y. H. Kim and B. K. Song, *Biotechnol. Prog.*, 2010, **26**, 1038–1046.
170 H. S. Kim, A. T. L. Quang and Y. H. Kim, *Enzyme Microb. Technol.*, 2010, **47**, 1–5.
171 J. H. Zhang, Y. Lin, Y. F. Sun, Y. R. Ye, S. P. Zheng and S. Y. Han, *Enzyme Microb. Technol.*, 2012, **50**, 325–330.
172 A. Siglioccolo, R. Gerace and S. Pascarella, *Biophys. Chem.*, 2010, **153**, 104–114.
173 H. Yu and H. Huang, *Biotechnol. Adv.*, 2014, **32**, 308–315.
174 W. G. Touw and G. Vriend, *Protein Eng., Des. Sel.*, 2014, **27**, 457–462.
175 M. T. Reetz and J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891–903.
176 K. Teilum, J. G. Olsen and B. B. Kragelund, *Cell. Mol. Life Sci.*, 2009, **66**, 2231–2247.
177 A. Heinecke, W. Eckhardt, M. Horsch and H.-J. Bungartz, in *Supercomputing for Molecular Dynamics Simulations*, Springer, 2015, pp. 11–29.
178 P. E. Lopes, O. Guvench and A. D. MacKerell, *Molecular Modeling of Proteins*, 2015, 47–71.
179 F. Zhu, Y. Zhuang, B. Wu, J. Li and B. He, *Appl. Biochem. Biotechnol.*, 2016, **178**, 725–738.
180 M. Purmonen, J. Valjakka, K. Takkinen, T. Laitinen and J. Rouvinen, *Protein Eng., Des. Sel.*, 2007, **20**, 551–559.
181 M. B. Abdul Rahman, R. A. Karjiban, A. B. Salleh, D. Jacobs, M. Basri, T. Chor, A. Leow, H. A. Wahab and R. N. Z. Abd Rahman, *Protein Pept. Lett.*, 2009, **16**, 1360–1370.
182 R. A. Karjiban, M. B. A. Rahman, M. Basri, A. B. Salleh, D. Jacobs and H. A. Wahab, *Protein J.*, 2009, **28**, 14–23.
183 R. Abedi Karjiban, B. A. Rahman, A. Bakar Salleh, M. Basri, R. Noor Zaliha Raja Abd Rahman and A. L. T. Chor, *Protein Pept. Lett.*, 2010, **17**, 699–707.
184 S. D'Auria, P. Herman, J. R. Lakowicz, E. Bertoli, F. Tanfani, M. Rossi and G. Manco, *Proteins: Struct., Funct., Bioinf.*, 2000, **38**, 351–360.
185 P. L. Wintrode, D. Zhang, N. Vaidehi, F. H. Arnold and W. A. Goddard, *J. Mol. Biol.*, 2003, **327**, 745–757.
186 J. Fitter and J. Heberle, *Biophys. J.*, 2000, **79**, 1629–1636.
187 R. D. Sharma, A. M. Lynn, P. K. Sharma, *et al.*, High temperature unfolding of Bacillus anthracis amidase-03 by molecular dynamics simulations, *Bioinformation*, 2009, **3**(10), 430–434.
188 D. S. Vieira and L. Degreve, *Mol. Phys.*, 2009, **107**, 59–69.
189 E. Bae and G. N. Phillips, *J. Biol. Chem.*, 2005, **280**, 30943–30948.
190 S. Vemparala, S. Mehrotra and H. Balaram, *Biochim. Biophys. Acta, Proteins Proteomics*, 2011, **1814**, 630–637.
191 S. Kundu and D. Roy, *J. Mol. Graphics Modell.*, 2010, **28**, 820–827.
192 X. Yu, S. C. Sigler, D. Hossain, M. Wierdl, S. R. Gwaltney, P. M. Potter and R. M. Wadkins, *J. Mol. Model.*, 2012, **18**, 2869–2883.
193 W. Xu, P. Cai, M. Yan, L. Xu and P.-k. Ouyang, *Chin. J. Chem. Phys.*, 2009, **22**, 467–472.
194 Y. Mazola, O. Guirola, S. Palomares, G. Chinea, C. Menéndez, L. Hernández and A. Musacchio, *J. Mol. Model.*, 2015, **21**, 1–11.
195 K. Kumar, K. Patel, D. Agrawal and J. Khire, *J. Mol. Model.*, 2015, **21**, 1–13.
196 B. Singh, G. Bulusu and A. Mitra, *J. Phys. Chem. B*, 2015, **119**, 392–409.
197 T. Tu, H. Luo, K. Meng, Y. Cheng, R. Ma, P. Shi, H. Huang, Y. Bai, Y. Wang and L. Zhang, *Appl. Environ. Microbiol.*, 2015, **81**, 6938–6944.
198 X. Yin, Y. Yao, M. Wu, T. Zhu, Y. Zeng and Q. Pang, *Biochemistry*, 2014, **79**, 531–537.

199 I. A. Noorbatcha, A. Sultan, H. Salleh and A. Amid, *Protein J.*, 2013, **32**, 309–316.
200 N. J. Christensen and K. P. Kepp, *PLoS One*, 2013, **8**, e61985.
201 L. Chen, X. Li, R. Wang, F. Fang, W. Yang and W. Kan, *J. Biomol. Struct. Dyn.*, 2016, 1–14.
202 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
203 E. Lindahl, B. Hess and D. Van Der Spoel, *Molecular Modeling Annual*, 2001, **7**, 306–317.
204 B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels and S. Boresch, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
205 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
206 M. Kalimeri, P. Derreumaux and F. Sterpone, *J. Non-Cryst. Solids*, 2015, **407**, 494–501.
207 C. Pfleger, P. C. Rathi, D. L. Klein, S. Radestock and H. Gohlke, *J. Chem. Inf. Model.*, 2013, **53**, 1007–1015.
208 P. C. Rathi, D. Mulnaes and H. Gohlke, *Bioinformatics*, 2015, **31**, 2394–2396.
209 D. M. Krüger, P. C. Rathi, C. Pfleger and H. Gohlke, *Nucleic Acids Res.*, 2013, gkt292.
210 P. C. Rathi, K.-E. Jaeger and H. Gohlke, *PLoS One*, 2015, **10**, e0130289.
211 D. J. Jacobs, L. A. Kuhn and M. F. Thorpe, in *Rigidity Theory and Applications*, Springer, 2002, pp. 357–384.
212 A. Ahmed and H. Gohlke, *Proteins: Struct., Funct., Bioinf.*, 2006, **63**, 1038–1051.
213 A. Rader, *Phys. Biol.*, 2009, **7**, 016002.
214 P. C. Rathi, S. Radestock and H. Gohlke, *J. Biotechnol.*, 2012, **159**, 135–144.
215 I. B. Kuznetsov, *Curr. Protein Pept. Sci.*, 2009, **10**, 607–613.
216 H. Gohlke, L. A. Kuhn and D. A. Case, *Proteins: Struct., Funct., Bioinf.*, 2004, **56**, 322–337.
217 D. B. Craig and A. A. Dombkowski, *BMC Bioinf.*, 2013, **14**, 1.
218 X. Yin, D. Hu, J.-F. Li, Y. He, T.-D. Zhu and M.-C. Wu, *PLoS One*, 2015, **10**, e0126864.
219 S. Zhang, Y. Wang, X. Song, J. Hong, Y. Zhang and L. Yao, *J. Chem. Inf. Model.*, 2014, **54**, 2826–2833.
220 Z. Tan, J. Li, M. Wu and J. Wang, *Appl. Biochem. Biotechnol.*, 2014, **173**, 1752–1764.
221 M. Surzhik, A. Schmidt, E. Glazunov, D. Firsov and M. Petukhov, *Appl. Biochem. Microbiol.*, 2014, **50**, 118–124.
222 L. Liu, Z. Deng, H. Yang, J. Li, H.-d. Shin, R. R. Chen, G. Du and J. Chen, *Appl. Environ. Microbiol.*, 2014, **80**, 798–807.
223 H. Ding, F. Gao, D. Liu, Z. Li, X. Xu, M. Wu and Y. Zhao, *Enzyme Microb. Technol.*, 2013, **53**, 365–372.
224 Y. Liu and B. Kuhlman, *Nucleic Acids Res.*, 2006, **34**, W235–W238.
225 K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan and J. Meiler, *Biochemistry*, 2010, **49**, 2987–2998.
226 A. Korkegian, M. E. Black, D. Baker and B. L. Stoddard, *Science*, 2005, **308**, 857–860.
227 K. A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima and A. Sarai, *Nucleic Acids Res.*, 2004, **32**, D120–D121.
228 R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–387.
229 J. Tian, N. Wu, X. Chu and Y. Fan, *BMC Bioinf.*, 2010, **11**, 1.
230 Y. Dehouck, A. Grosfils, B. Folch, D. Gilis, P. Bogaerts and M. Rooman, *Bioinformatics*, 2009, **25**, 2537–2543.
231 L.-T. Huang, M. M. Gromiha and S.-Y. Ho, *Bioinformatics*, 2007, **23**, 1292–1293.
232 L.-T. Huang, L.-F. Lai and C.-C. Wu, *Open Struct. Biol. J.*, 2009, **3**, 143–148.
233 L.-T. Huang and M. M. Gromiha, *Bioinformatics*, 2009, **25**, 2181–2187.
234 M. Masso and I. I. Vaisman, *Bioinformatics*, 2008, **24**, 2002–2009.
235 V. Parthiban, M. M. Gromiha and D. Schomburg, *Nucleic Acids Res.*, 2006, **34**, W239–W242.
236 J. Cheng, A. Randall and P. Baldi, *Proteins: Struct., Funct., Bioinf.*, 2006, **62**, 1125–1132.
237 E. Capriotti, P. Fariselli, R. Calabrese and R. Casadio, *Bioinformatics*, 2005, **21**, ii54–ii58.
238 S. Kang, G. Chen and G. Xiao, *Protein Eng., Des. Sel.*, 2009, **22**, 75–83.
239 R. R. Thangudu, P. Sharma, N. Srinivasan and B. Offmann, *Proteins*, 2007, **67**, 255–261.
240 A. Vinayagam, G. Pugalenthi, R. Rajesh and R. Sowdhamini, *Nucleic Acids Res.*, 2004, **32**, D200–D202.
241 T. Meyer, M. D'Abramo, A. Hospital, M. Rueda, C. Ferrer-Costa, A. Pérez, O. Carrillo, J. Camps, C. Fenollosa and D. Repchevsky, *Structure*, 2010, **18**, 1399–1409.
242 L. Baltzer and J. Nilsson, *Curr. Opin. Biotechnol.*, 2001, **12**, 355–360.
243 J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan and F. H. Arnold, *Curr. Opin. Struct. Biol.*, 2005, **15**, 447–452.
244 D. N. Bolon, C. A. Voigt and S. L. Mayo, *Curr. Opin. Chem. Biol.*, 2002, **6**, 125–129.
245 G. L. Butterfoss and B. Kuhlman, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 49–65.
246 M. Lehmann and M. Wyss, *Curr. Opin. Biotechnol.*, 2001, **12**, 371–375.
247 B. van den Burg and V. G. Eijsink, *Curr. Opin. Biotechnol.*, 2002, **13**, 333–337.
248 S. Sunyaev, W. Lathe IIIrd and P. Bork, *Curr. Opin. Struct. Biol.*, 2001, **11**, 125–130.
249 E. Capriotti, P. Fariselli and R. Casadio, *Bioinformatics*, 2004, **20**(1), i63–68.
250 J. W. Pitera and P. A. Kollman, *Proteins*, 2000, **41**, 385–397.
251 C. Deutsch and B. Krishnamoorthy, *Bioinformatics*, 2007, **23**, 3009–3015.
252 Y. Li and J. Fang, *PLoS One*, 2012, **7**, e47247.
253 D. Gilis and M. Rooman, *Protein Eng.*, 2000, **13**, 849–856.
254 C. Magyar, M. M. Gromiha, G. Pujadas, G. E. Tusnady and I. Simon, *Nucleic Acids Res.*, 2005, **33**, W303–W305.
255 H. Zhou and Y. Zhou, *Protein Sci.*, 2002, **11**, 2714–2726.
256 H. Zhou and Y. Zhou, *Proteins*, 2004, **54**, 315–322.

257 R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–387.
258 V. Parthiban, M. M. Gromiha and D. Schomburg, *Nucleic Acids Res.*, 2006, **34**, W239–W242.
259 E. Capriotti, P. Fariselli, I. Rossi and R. Casadio, *BMC Bioinf.*, 2008, **9**(2), S6.
260 J. Cheng, A. Randall and P. Baldi, *Proteins*, 2006, **62**, 1125–1132.
261 Z. Dosztanyi, C. Magyar, G. Tusnady and I. Simon, *Bioinformatics*, 2003, **19**, 899–900.
262 B. Shen, J. Bai and M. Vihinen, *Protein Eng., Des. Sel.*, 2008, **21**, 37–44.
263 V. Potapov, M. Cohen and G. Schreiber, *Protein Eng., Des. Sel.*, 2009, **22**, 553–560.
264 M. Dorn, M. B. E. Silva, L. S. Buriol and L. C. Lamb, *Comput. Biol. Chem.*, 2014, **53**, 251–276.
265 A. G. de Brevern, A. Bornot, P. Craveur, C. Etchebest and J.-C. Gelly, *Nucleic Acids Res.*, 2012, **40**, W317–W322.
266 G. OmPraba, D. Velmurugan, P. Arumugam, V. Govindasamy and P. Kalaichelvan, *J. Biomol. Struct. Dyn.*, 2007, **25**, 311–319.
267 G. Manco, L. Camardella, F. Febbraio, G. Adamo, V. Carratore and M. Rossi, *Protein Eng.*, 2000, **13**, 197–200.
268 R. P. Singh, A. R. Rai, K. Roychoudhury and R. Dubey, *J. Appl. Sci. Environ. Sanit.*, 2011, **6**, 485–494.
269 J. Wong, *Molecular Modeling of Thermostable Endoglucanases*, ProQuest, 2007.
270 I. Gontia-Mishra, V. Kumar Singh, N. Tripathi, S. Sasidharan and S. Tiwari, *Biologia*, 2014, **69**, 1283–1294.
271 P. Farrokh, B. Yakhchali and A. A. Karkhane, *J. Mol. Microbiol. Biotechnol.*, 2014, **24**, 262–269.
272 G. Wang, R.-Y. Cao, R. Chen, L. Mo, J.-F. Han, X. Wang, X. Xu, T. Jiang, Y.-Q. Deng and K. Lyu, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7619–7624.
273 M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni and L. Bordoli, *Nucleic Acids Res.*, 2014, gku340.
274 B. Webb and A. Sali, *Current Protocols in Bioinformatics*, 2014, 5.6.1–5.6.32.
275 R. Feng, B. Liang, C. Hou, D. Han, L. Han, Q. Lang, A. Liu and L. Han, *Enzyme Microb. Technol.*, 2016, **84**, 78–85.
276 H. Kaneko, H. Minagawa and J. Shimada, *Biotechnol. Lett.*, 2005, **27**, 1777–1784.
277 S. B. Mabrouk, N. Aghajari, M. B. Ali, E. B. Messaoud, M. Juy, R. Haser and S. Bejar, *Bioresour. Technol.*, 2011, **102**, 1740–1746.
278 G. C. Mu, Y. Nie, X. Q. Mu, Y. Xu and R. Xiao, *Appl. Biochem. Biotechnol.*, 2015, **176**, 1736–1745.
279 B. S. Alipour, S. Hosseinkhani, S. K. Ardestani and A. Moradi, *Photochem. Photobiol. Sci.*, 2009, **8**, 847–855.
280 K. Rakesh, C. Nisha, S. Ranvir, K. Pushpender and K. Jagdeep, *Adv. Genet. Eng.*, 2015, **2015**, DOI: 2169-0111.1000126.
281 N. Akbulut, M. T. Öztürk, T. Pijning, S. İ. Öztürk and F. Gümüşel, *J. Biotechnol.*, 2013, **164**, 123–129.
282 B. Kumwenda, D. Litthauer, Ö. T. Bishop and O. Reva, *Evol. Bioinf.*, 2013, **9**, 327.
283 S. Sinchaikul, B. Sookkheo, S. Phutrakul, Y.-T. Wu, F.-M. Pan and S.-T. Chen, *Biochem. Biophys. Res. Commun.*, 2001, **283**, 868–875.
284 S.-G. Lee, S.-P. Hong, J. J. Song, S.-J. Kim, M.-S. Kwak and M.-H. Sung, *Appl. Environ. Microbiol.*, 2006, **72**, 1588–1594.
285 Y. Xiaoyan, M. Zhen, C. Dongwei, G. Xu, J. Zeyer, L. Shuangjiang and C. Jiang, *Chin. J. Chem. Eng.*, 2012, **20**, 52–61.
286 H. Wang, Y. Gong, W. Xie, W. Xiao, J. Wang, Y. Zheng, J. Hu and Z. Liu, *Appl. Biochem. Biotechnol.*, 2011, **164**, 1323–1338.
287 H. P. Sun, Y. Huang, X. F. Wang, Y. Zhang and H. B. Shen, *Proteins: Struct., Funct., Bioinf.*, 2015, **83**, 485–496.
288 J. Andreani and J. Söding, *Bioinformatics*, 2015, btv041.
289 P. Di Lena, K. Nagata and P. Baldi, *Bioinformatics*, 2012, **28**, 2449–2457.
290 J. A. Brown, V. Pensabene, D. A. Markov, V. Allwardt, M. D. Neely, M. Shi, C. M. Britt, O. S. Hoilett, Q. Yang and B. M. Brewer, *Biomicrofluidics*, 2015, **9**, 054124.
291 T. A. Hopf, C. P. Schärfe, J. P. Rodrigues, A. G. Green, O. Kohlbacher, C. Sander, A. M. Bonvin and D. S. Marks, *Elife*, 2014, **3**, e03430.
292 O. V. Galzitskaya, S. O. Garbuzynskiy and M. Y. Lobanov, *Bioinformatics*, 2006, **22**, 2948–2949.
293 A. Schlessinger, G. Yachdav and B. Rost, *Bioinformatics*, 2006, **22**, 891–893.
294 E. Cilia, R. Pancsa, P. Tompa, T. Lenaerts and W. F. Vranken, *Nucleic Acids Res.*, 2014, **42**, W264–W270.
295 P. Kountouris and J. D. Hirst, *BMC Bioinf.*, 2009, **10**, 1.
296 B. Borguesan, M. Inostroza-Ponta and M. Dorn, *J. Comput. Biol.*, 2016, DOI: 10.1089/cmb.2016.0074.
297 A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, *Nucleic Acids Res.*, 2015, gkv332.
298 F. Ferrè and P. Clote, *Nucleic Acids Res.*, 2006, **34**, W182–W185.
299 J. Yang, B.-J. He, R. Jang, Y. Zhang and H.-B. Shen, *Bioinformatics*, 2015, **31**, 3773–3781.
300 A. Yaseen and Y. Li, Accelerating knowledge-based energy evaluation in protein structure modeling with Graphics Processing Units, *J. Parallel Distr. Comput.*, 2012, **72**(2), 297–307.
301 A. Ceroni, A. Passerini, A. Vullo and P. Frasconi, *Nucleic Acids Res.*, 2006, **34**, W177–W181.
302 N. Tokuriki, *et al.*, *How protein stability and new functions trade off, PLoS Comput. Biol.*, 2008, **4**(2), e1000002.
303 A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, W. H. Freeman, New York, 1999.
304 D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman and J. Schrag, *Protein Eng.*, 1992, **5**, 197–211.
305 B. M. Beadle and B. K. Shoichet, *J. Mol. Biol.*, 2002, **321**, 285–296.
306 C. Garcia, C. Nishimura, S. Cavagnero, H. J. Dyson and P. E. Wright, *Biochemistry*, 2000, **39**, 11227–11237.
307 R. A. Nagatani, A. Gonzalez, B. K. Shoichet, L. S. Brinen and P. C. Babbitt, *Biochemistry*, 2007, **46**, 6688–6695.

308 X. Wang, G. Minasov and B. K. Shoichet, *J. Mol. Biol.*, 2002, **320**, 85–95.
309 Y. Chen, B. Shoichet and R. Bonnet, *J. Am. Chem. Soc.*, 2005, **127**, 5423–5434.
310 J. D. Bloom, S. T. Labthavikul, C. R. Otey and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 5869–5874.
311 R. Godoy-Ruiz, R. Perez-Jimenez, B. Ibarra-Molero and J. M. Sanchez-Ruiz, *J. Mol. Biol.*, 2004, **336**, 313–318.
312 H. Zhao and F. H. Arnold, *Protein Eng.*, 1999, **12**, 47–53.
313 G. Manco, E. Giosuè, S. D'Auria, P. Herman, G. Carrea and M. Rossi, *Arch. Biochem. Biophys.*, 2000, **373**, 182–192.
314 A. Gershenson, J. A. Schauerte, L. Giver and F. H. Arnold, *Biochemistry*, 2000, **39**, 4658–4665.
315 T. Lazaridis, I. Lee and M. Karplus, *Protein Sci.*, 1997, **6**, 2589.
316 G. Hernández, F. E. Jenney, M. W. Adams and D. M. LeMaster, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3166–3170.
317 A. Merz, T. Knöchel, J. N. Jansonius and K. Kirschner, *J. Mol. Biol.*, 1999, **288**, 753–763.
318 R. M. Daniel, M. J. Danson, D. W. Hough, C. K. Lee, M. E. Peterson and D. A. Cowan, *Protein Adapt. Extremophiles*, 2008, 1–34.
319 G. Feller, *J. Phys.: Condens. Matter*, 2010, **22**, 323101.
320 K. S. Siddiqui and R. Cavicchioli, *Annu. Rev. Biochem.*, 2006, **75**, 403–433.
321 K. S. Siddiqui, *Crit. Rev. Biotechnol.*, 2016, 1–14.
322 P. S. Low, J. L. Bada and G. N. Somero, *Proc. Natl. Acad. Sci. U. S. A.*, 1973, **70**, 430–432.
323 T. Lonhienne, C. Gerday and G. Feller, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 2000, **1543**, 1–10.
324 S. You, T. Tu, L. Zhang, Y. Wang, H. Huang, R. Ma, P. Shi, Y. Bai, X. Su and Z. Lin, *Biotechnol. Biofuels*, 2016, **9**, 124.
325 X. Duan, J. Chen and J. Wu, *Appl. Environ. Microbiol.*, 2013, **79**, 4072–4077.
326 Y. Liu, Z. Yan, X. Lu, D. Xiao and H. Jiang, *Sci. Rep.*, 2016, **6**, 24117.
327 A. Szlachcic, M. Zakrzewska and J. Otlewski, *Biotechnol. Adv.*, 2011, **29**, 436–441.
328 G. A. Lazar, S. A. Marshall, J. J. Plecs, S. L. Mayo and J. R. Desjarlais, *Curr. Opin. Struct. Biol.*, 2003, **13**, 513–518.
329 S. A. Marshall, G. A. Lazar, A. J. Chirino and J. R. Desjarlais, *Drug Discovery Today*, 2003, **8**, 212–221.
330 A. Özen, M. Gönen, E. Alpaydın and T. Haliloğlu, *BMC Struct. Biol.*, 2009, **9**, 1.
331 C. L. Worth, R. Preissner and T. L. Blundell, *Nucleic Acids Res.*, 2011, **39**, W215–W222.
332 D. E. Pires, D. B. Ascher and T. L. Blundell, *Bioinformatics*, 2014, **30**, 335–342.
333 D. E. Pires, D. B. Ascher and T. L. Blundell, *Nucleic Acids Res.*, 2014, gku411.
334 C.-W. Chen, J. Lin and Y.-W. Chu, *BMC Bioinf.*, 2013, **14**, 1.
335 P. Fariselli, P. L. Martelli, C. Savojardo and R. Casadio, *Bioinformatics*, 2015, **31**, 2816–2821.
336 M. Giollo, A. J. Martin, I. Walsh, C. Ferrari and S. C. Tosatto, *BMC Genomics*, 2014, **15**, 1.
337 V. Frappier, M. Chartier and R. J. Najmanovich, *Nucleic Acids Res.*, 2015, gkv343.
338 L. Folkman, B. Stantic, A. Sattar and Y. Zhou, *J. Mol. Biol.*, 2016, **428**, 1394–1405.
339 J. Laimer, J. Hiebl-Flach, D. Lengauer and P. Lackner, *Bioinformatics*, 2016, btv769.
340 L. Quan, Q. Lv and Y. Zhang, *Bioinformatics*, 2016, btw361.
341 I. Hwang and S. Park, *Drug Discovery Today*, 2008, **5**, e43–e48.
342 R. E. Kontermann, *Curr. Opin. Biotechnol.*, 2011, **22**, 868–876.
343 S. Frokjaer and D. E. Otzen, *Nat. Rev. Drug Discovery*, 2005, **4**, 298–306.
344 H. O. Alpar, S. Somavarapu, K. Atuah and V. Bramwell, *Adv. Drug Delivery Rev.*, 2005, **57**, 411–430.
345 A. Hussain, J. J. Arnold, M. A. Khan and F. Ahsan, *J. Controlled Release*, 2004, **94**, 15–24.
346 N. Rathore and R. S. Rajan, *Biotechnol. Prog.*, 2008, **24**, 504–514.
347 W. Wang, *Int. J. Pharm.*, 1999, **185**, 129–188.
348 W. Wang, S. Singh, D. L. Zeng, K. King and S. Nema, *J. Pharm. Sci.*, 2007, **96**, 1–26.
349 S. M. Malakauskas and S. L. Mayo, *Nat. Struct. Mol. Biol.*, 1998, **5**, 470–475.
350 A. V. Filikov, R. J. Hayes, P. Luo, D. M. Stark, C. Chan, A. Kundu and B. I. Dahiyat, *Protein Sci.*, 2002, **11**, 1452–1461.
351 P. Luo, R. J. Hayes, C. Chan, D. M. Stark, M. Y. Hwang, J. M. Jacinto, P. Juvvadi, H. S. Chung, A. Kundu and M. L. Ary, *Protein Sci.*, 2002, **11**, 1218–1226.
352 M. Zakrzewska, D. Krowarsch, A. Wiedlocha, S. Olsnes and J. Otlewski, *J. Mol. Biol.*, 2005, **352**, 860–875.
353 B. A. Shirley, P. Stanssensp, J. Steyaertfll and C. N. Pace, *Chemistry*, 1989, **264**, 11621–11625.
354 J. Caravella and A. Lugovskoy, *Curr. Opin. Chem. Biol.*, 2010, **14**, 520–528.
355 J. C. Almagro, S. Kodangattil and J. Li, *Making and Using Antibodies: A Practical Handbook*, 2013, p. 395.
356 M. J. Bernett, S. Karki, G. L. Moore, I. W. Leung, H. Chen, E. Pong, D.-H. T. Nguyen, J. Jacinto, J. Zalevsky and U. S. Muchhal, *J. Mol. Biol.*, 2010, **396**, 1474–1490.
357 M. Kügler, C. Stein, M. Schwenkert, D. Saul, L. Vockentanz, T. Huber, S. K. Wetzel, O. Scholz, A. Plückthun and A. Honegger, *Protein Eng., Des. Sel.*, 2009, **22**, 135–147.
358 A. Honegger, A. D. Malebranche, D. Röthlisberger and A. Plückthun, *Protein Eng., Des. Sel.*, 2009, **22**, 121–134.
359 C. S. Fishburn, *J. Pharm. Sci.*, 2008, **97**, 4167–4183.
360 S. Jevševar, M. Kunstelj and V. G. Porekar, *Biotechnol. J.*, 2010, **5**, 113–128.
361 N. Wang, W. F. Smith, B. R. Miller, D. Aivazian, A. A. Lugovskoy, M. E. Reff, S. M. Glaser, L. J. Croner and S. J. Demarest, *Proteins: Struct., Funct., Bioinf.*, 2009, **76**, 99–114.
362 L. Borras, T. Gunde, J. Tietz, U. Bauer, V. Hulmann-Cottier, J. P. Grimshaw and D. M. Urech, *J. Biol. Chem.*, 2010, **285**, 9054–9066.

363 M. Gebauer and A. Skerra, *Curr. Opin. Chem. Biol.*, 2009, **13**, 245–255.
364 D. Saerens, K. Conrath, J. Govaert and S. Muyldermans, *J. Mol. Biol.*, 2008, **377**, 478–488.
365 R. Gong, B. K. Vu, Y. Feng, D. A. Prieto, M. A. Dyba, J. D. Walsh, P. Prabakaran, T. D. Veenstra, S. G. Tarasov and R. Ishima, *J. Biol. Chem.*, 2009, **284**, 14203–14210.
366 J. F. Culajay, S. I. Blaber, A. Khurana and M. Blaber, *Biochemistry*, 2000, **39**, 7153–7158.
367 T. Arakawa, S. J. Prestrelski, L. O. Narhi, T. C. Boone and W. C. Kenney, *J. Protein Chem.*, 1993, **12**, 525–531.
368 J. Ishikawa, M. Yoshimura, T. Matsunashi, N. Tominaga, H. Teshima, A. Hiraoka, H. Nakamura, H. Shibata, T. Masaoka and F. Takaku, *Jpn. J. Clin. Oncol.*, 1991, **21**, 169–175.
369 D. Mark, S. Lu, A. Creasey, R. Yamamoto and L. Lin, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 5662–5666.
370 L. Wang, A. Gamez, H. Archer, E. E. Abola, C. N. Sarkissian, P. Fitzpatrick, D. Wendt, Y. Zhang, M. Vellard and J. Bliesath, *J. Mol. Biol.*, 2008, **380**, 623–635.
371 S. W. Pipe and R. J. Kaufman, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 11851–11856.