Cellular and Molecular Life Sciences

Structural determinants of protein folding

Tse Siang Kang · R. Manjunatha Kini

Abstract The last several decades have seen an explosion of knowledge in the field of structural biology. With critical advances in spectroscopic techniques in examining structures of biomacromolecules, in maturation of molecular biology techniques, as well as vast improvements in computation prowess, protein structures are now being elucidated at an unprecedented rate. In spite of all the recent advances, the protein folding puzzle remains as one of the fundamental biochemical challenges. A facet to this empiric problem is the structural determinants of protein folding. What are the driving forces that pivot a polypeptide chain to a specific conformation amongst the vast conformation space? In this review, we shall discuss some of the structural determinants to protein folding that have been identified in the recent decades.

Keywords Protein folding · Folding models · Folding propensity · Structural determinants · Posttranslational modifications · Secondary structures

T. S. Kang
The Scripps Research Institute, 10550 North Torrey Pines Road
GAC 1200, La Jolla, CA 92037, USA

T. S. Kang
Department of Pharmacy, National University of Singapore,
18 Science Drive 4, Block S4, Singapore 117543, Singapore

R. M. Kini (✉)
Department of Biological Sciences,
National University of Singapore, 14 Science Drive 4,
Block S3 #03-17, Singapore 117543, Singapore
e-mail: dbskinim@nus.edu.sg

Introduction

Protein folding is one of the most fundamental and yet most intriguing questions of structural biology. As aptly described by Honig et al., the protein folding puzzle is a "classical challenge in which the problem is easy to state, existence of an answer is apparent, and yet the solution remains elusive" [1]. Simply stated, how does a linear polypeptide chain of scrambled amino acid residues fold into a native distinctive conformation with pinpoint accuracy?

The impetus for in vitro protein folding research stems from two of the most well-known theories of protein folding: Anfinsen et al. [2, 3] on the folding of ribonuclease A, and the Levinthal Paradox by Cyprus Levinthal [4, 5]. In 1969, Anfinsen et al. demonstrated that ribonuclease A can spontaneously refold into its native conformation in the in vitro setting, without assistance from biological systems or other biomacromolecules (Fig. 1). These findings led to the next question: is this native conformation achieved by a random search of the vast conformational space that is available to the linear polypeptide? It was subsequently pointed out by Cyprus Levinthal that such a random search for the native conformational state of a protein would require an infinitesimally long time. For example, a 100-residue protein taking a mere 10^{-11} s per configuration, will take 10^{52} years to search through the entire set of possible conformations [5, 6]! This is in stark contrast to the millisecond time scale that a protein typically takes to fold to its native conformation. Therefore, it is apparent that there exist specific folding intermediates or pathways through which the proteins can adopt the final destined conformation. The implications from these theories triggered extensive research on the identification of folding intermediates [7], real-time folding studies [8], computation modeling and simulations [9, 10], and energy landscape

Fig. 1 Spontaneous refolding of bovine ribonuclease. Christian Anfinson et al. demonstrated that native bovine ribonuclease (PDB ID 7RSA) denatured with 8 M urea and 2-mercaptoethanol spontaneously refolds to the fully active refolded conformation upon removal of the denaturing reagents. Oxidation of the fully denatured polypeptide in the presence of 8 M urea in the absence of sulfhydryl group containing 2-mercaptoethanol results in a "scrambled" mixture of misfolded disulfide isomers. Subsequent removal of urea and addition of 2-mercaptoethanol facilitated disulfide rearrangement of the misfolded protein to the native conformation and restoration of full enzymatic activity [2, 3]
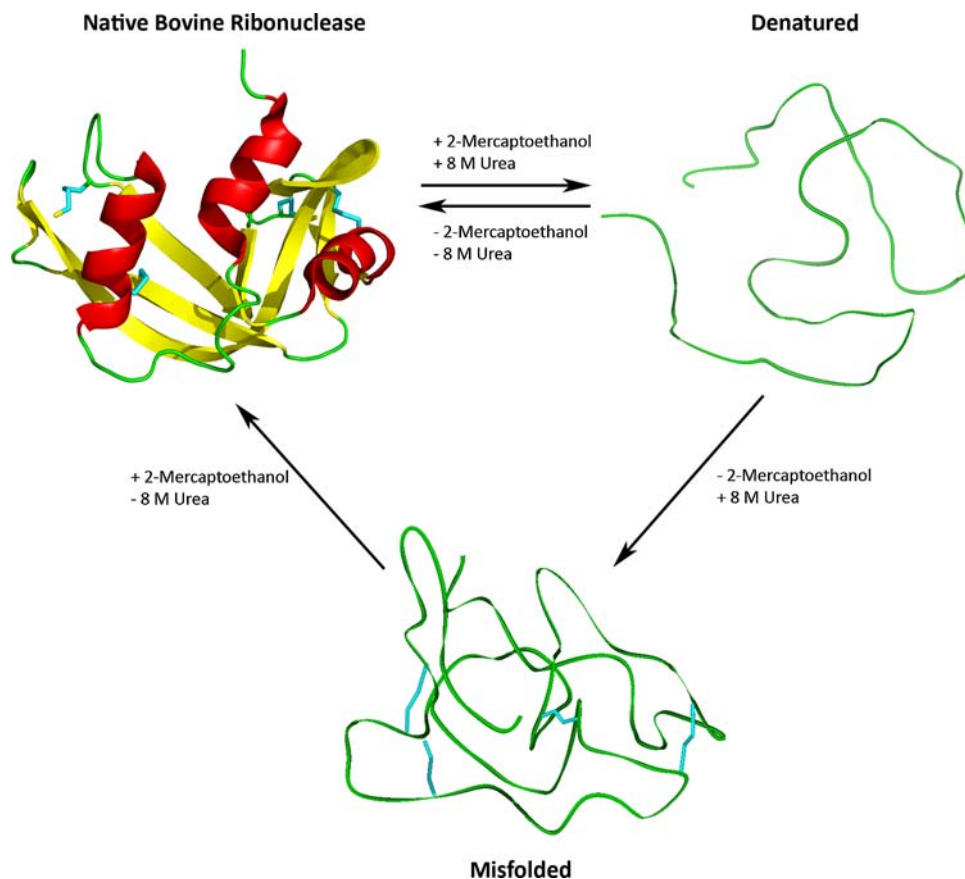
analysis [11]. From these pioneering works, several models on mechanism of protein folding were conceptualized [12].

One of the first models of protein folding was the nucleation growth model, in which a predominant structural nucleus is formed, and the secondary and tertiary structures of the protein propagate from this nucleus [13]. However, this mechanism was subsequently less favored due to the inability to account for folding intermediates present in folding pathways [14]. The next theoretical mechanism for protein folding was suggested by Oleg Ptitsyn, who proposed that protein folding occurs through a sequential on-pathway process with the formation of secondary structures in an initial "molten globular" state, with each stage of the dynamic protein folding process stabilized by the former stage [14, 15]. This model was subsequently formulated as the diffusion–collision model and the related framework model, which theorized that structural elements rapidly assemble independently before "colliding" to coalesce and stabilize by interactions between secondary structures to form the final tertiary form [16, 17]. A third model of protein folding was the hydrophobic collapse model, which proposed that the collapse of the primary structure around a hydrophobic core reduces the scope of search for the native conformation [18]. Subsequently, it was suggested that secondary and tertiary structures in the framework model were formed in

conjunction with the hydrophobic collapse of the polypeptide chain, leading to unification of the two folding models to the nucleation–condensation model [19, 20]. However, the various models are not without problems. For example, the hydrophobic interactions in the hydrophobic collapse model may indicate difficulties in structural reorganization of the polypeptide chain along the folding pathway. Further, it is unlikely that the unfolded state of a polypeptide exists in the form of a completely unstructured chain [21]. By examining the denatured states of a small model protein, Shortle et al. [22] further proposed that long range ordering of the polypeptide segments could very well occur before emergence of defined structural elements of the protein.

One of the critical advances in the understanding of folding pathways was spearheaded through the development of Φ-value analysis by Fersht et al. [23]. For this, the kinetics of mutants, in which deliberate perturbations were made to specific interactions in the folding pathway, were measured so as to examine the folding pathways concerned. In effect, these Φ-values are structural indices for specific residues in the folding intermediate [24]. Numerous recent innovative experimental applications or strategies have subsequently been devised to aid the understanding of the protein folding problem. These strategies include, but are not limited to, NMR spectroscopy [25, 26], single molecule

studies with Förster resonance energy transfer (FRET) or atomic force microscopy [27, 28], UV Raman spectroscopy [29], computational approaches and molecular dynamics simulations [30, 31], as well as theoretical analysis [32]. Through these methods, novel information that was previously inaccessible could now be obtained and even utilized in the reengineering of ultrafast folding proteins [33]. Complemented with classical strategies such as the Φ-values, these methods have been instrumental in the recent advancement of knowledge of protein folding pathways.

Although much of the specifics to folding pathways suggested by these folding models are still ambiguous, it is now known that folding occurs via a series of folding intermediates, through either a two-state (nucleation–condensation model) or three-state (diffusion–collision model) mechanism [24, 34, 35]. However, there are indications that the two-state and three-state folding models are, in fact, extremes of a common folding mechanism [21, 35], and the fundamental difference between the two routes lies in the relative stability of the folding intermediates [36]. Indeed, these two extremes of the folding mechanism have been demonstrated on single protein models, indicative of a unifying mechanism of protein folding that incorporates aspects of both two-state and three-state models [37]. Variations between proteins in following a single pathway or multiple different pathways could also be attributed to differences in the number of nucleation "foldon" motifs, as well as conservation of transition state structures within the final protein structures [38, 39]. Numerous reviews pertaining to the various experimental and theoretical aspects of the protein folding problem are available, and will not be further detailed in this review.

Although much advancement of knowledge has been made over the four decades since research into the protein folding problem intensified, this biochemical puzzle is far from being solved. It is now apparent that the various proposed models of protein folding are not likely to operate in a mutually exclusive fashion, and that the folding mechanisms are likely to vary between proteins [40]. Whilst these proposed folding models may be conceptually different, a commonality seen across the models is the formation of secondary structure elements.

Despite the fact that several groups have been successful in designing small protein molecules to specific conformations, prediction of the structure of native structures from the amino acid sequence remains a challenge [14, 41]. It is now clear that the various associated interactions arising from the functional side chains, albeit a few weaker than others, contribute to the protein stability and protein folding process [12, 42, 43]. These forces include electrostatic interactions [44], hydrogen bonding interactions [45], van der Waals interactions [46], and hydrophobic interactions [42]. Through the use of an all-atom ab initio simulation, Yang et al. demonstrated that it is likely a combination of the multiple interacting forces that drives the protein folding process [43]. Indeed, these interacting forces are the fundamental interactions on which most, if not all, of the structural determinants of protein folding rely. These contributing forces of interaction implicated in protein folding were reviewed nearly two decades ago [42, 47].

We now have a much better grasp of the mechanisms of protein folding, and the possibility of overcoming this challenge seems increasingly promising [12]. To an extent, we also understand the various in vivo [48] and in vitro [49, 50] parameters that influence the folding profile and thermodynamic stability of these sensitive biomacromolecules. The classical experiment by Christian Anfinsen reiterates the point that the native conformation of a protein is likely to be dictated by the primary structure of the molecule [2, 51]. Therefore, it is interesting to examine the structural determinants of the polypeptide chain that influence the folding pathway and final confirmation of the mature protein. In this review, we shall discuss some of the currently known or suggested structural determinants of protein folding.

Letters of the protein language alphabet

With merely 20 amino acids in the standard polypeptide alphabet, nature has conjured an impressive array of proteins. Over 50,000 known protein structures across the various genera have been deposited in the protein data bank (http://www.rcsb.org), and an exponentially higher number of distinct protein sequences with unique three-dimensional structures remain to be elucidated [52, 53]. Such diverse permutations arise from the assorted properties of the amino acid residues with structurally and biochemically distinct side chains.

It is the diverse physicochemical properties of these structural building blocks that endow distinctive characteristics to a polypeptide chain. Further, the functional side chains of the amino acids are at least as important as the backbone atoms in laying the platform for the multiple interacting forces critical for structure, function, and stability of proteins. Indeed, the side chain "appendages" to the polypeptide backbone have been so intriguing that a sidechain-only model of protein folding has been previously proposed [54]. Through the analysis and extrapolation of information derived from databases and literature on protein folding and conformation, it was also previously suggested that certain amino acids exert specific influence on the local conformation [55]. We shall discuss how the physicochemical properties of the various amino acids preferentially result in specific structural conformations.

Propensity of forming secondary structures

The right-handed α-helix is the most commonly observed secondary structure in proteins with 3.6 residues per turn and translation of 1.5 Å per residue [56]. The geometry of a classical α-helix involves the carbonyl group hydrogen bonding to the amide proton of the $i + 4$ residue of the polypeptide, with side chains projecting away from the helix backbone.

The propensity of a polypeptide to form an α-helix has been shown to be dependent on various factors [57]. The most prominent factor is the amino acid constituents in the helix-forming segment, and the consequent interactions that are sterically permitted. As early as the 1960s, several other groups presented predictive methodologies for forecasting potential secondary structures based on the characteristics of the various residues [10, 58, 59]. Although specific potentials of all 20 amino acids in forming secondary structures were not consistently predicted, the most favorable and least favorable helix formers were generally in agreement. For example, aliphatic residues with branched Cβ side chains (Val, Ile, Thr) were deemed unfavorable due to the large volume of the functional group that results in steric hindrances [60]. On the other hand, alanine residues have long been recognized to possess a high helix-forming propensity. Polyalanine, polyglutamine, polylysine, and polyaspartate peptides were also shown to be able to form stable helices in solution [61]. The alignment of periodic repeats of multiple backbone hydrogen bonds running parallel to the helix axis results in a net cumulative dipole along the termini of the helix [60]. In contrast, glycines have the highest helix destabilizing effect due to the lack of a side chain functional group that results in inherent flexibility of the peptide bond [62]. Due to the lack of an amide proton, the imino acid residue proline has also been noted to be a helix breaker [63]. In view of its unique physicochemical properties, proline will be further discussed in the next section.

In addition to the hydrogen bond interactions of backbone atoms, formation of helix-stabilizing hydrogen bond interactions between side chains has also been noted to contribute to the helix formation and propagation. One of the early examples identified was the helix-stabilizing interactions between Glu2–Arg10 and Phe8–His12 in the N-terminal helical segment of RNase S that reinforced the multitude of hydrogen-bonds of the helix [64]. Incidentally, the exact interactions were also observed in the isolated corresponding C-peptide fragment of RNase S, indicating the polypeptide's capability to autonomously nucleate into a helix [8]. Conversely, an unfavorable side chain interaction such as His12–Asp14 in the S-peptide of RNase could present as a helix stop signal by preventing the propagation of the helix beyond residue 13 [65].

Further work by Shoemaker et al. on C-peptide variants indicated that the charge–dipole interaction of constituent charged residues with the macrodipole of the helix contributes to the enthalpy change for helix formation: negatively charged residues such as glutamate and aspartate occur preferentially near the positive N-terminal dipole of the helix, while positively charged residues such as lysine, arginine, and histidine commonly reside at the negatively charged C-terminal dipole of the helix [64, 66]. These findings were supported by similar observations in other helical peptides [67].

In the bid to decipher the protein folding problem, one of the fundamental questions is to comprehend the factors that influence short amino acid sequences to fold into various conformations in solution. Early predictive algorithms involved basic parameters such as nucleation and elongation factors [68], and subsequently incorporated contributions of side chain interactions to improve the prediction accuracy of helix formation [69]. However, these algorithms were still unable to predict helix-coil transition in solution. The comparatively most accurate predictive algorithm for helix formation of various residues in non-repetitive sequences is the AGADIR algorithm developed by Muñoz et al. [70–72] The AGADIR algorithm incorporated several of the contributions discussed above, including specific amino acid-dependent entropic cost, hydrogen bonds interactions between main chain and side chains, as well as capping and dipole effects. Through correlation with NMR analysis, AGADIR was shown to be capable of accurately predicting helical behavior of peptides under various experimental conditions [70].

The other secondary structure element commonly observed in protein structures is the β-sheet. The distinctive feature of the β-sheet is the ability to form either parallel, or anti-parallel, sheets capable of bringing residues that are distant in sequence into close proximity of one another. Unlike the periodic backbone hydrogen bonding in α-helices, the backbone hydrogen bonding pattern in β-sheets is dependent on the orientation of the β-strands [73].

The propensity of a polypeptide chain to form a β-sheet is also influenced by the specific amino acid constituents. Through the statistical analysis of known protein structures and experimental measurements of individual amino acids in the "host–guest" method initially utilized for analyzing helices [62], residues with Cβ-branched side chains or aromatic side chains were noted to be most favorable in the formation of β-sheets [10, 73]. Although these findings could indicate these amino acids' importance and tendency of forming β-sheets, the frequent observation of β-sheets in hydrophobic cores of proteins suggests their concurrent importance in hydrophobic interactions [73]. As with α-helices, flexible or disorder-promoting residues (glycine and proline) are similarly disfavored in the formation of

β-sheet structures. However, statistical analyses have shown that these flexible residues and certain residues with small polar side chains, such as asparagine, aspartic acid, and serine, are favorable for forming the β-turns by satisfying supporting hydrogen bonding requirements of the main chain amide [73–75]. Residues that resulted in stabilizing interactions were also more frequently located in β-sheet structures. For example, Phe-Phe, Phe-Tyr, Glu-Arg, and Glu-Lys interactions were identified by experimental and statistical methods to occur most frequently in β-sheets, while Thr-Val and Thr-Trp interactions were the least frequently observed interacting pair [73].

The importance of specific amino acid composition in determining secondary structure elements can be further illustrated by intrinsically unstructured proteins. Unlike globular proteins that typically form hydrophobic cores, intrinsically disordered proteins possess a comparatively lower proportion of "order-promoting" bulky hydrophobic residues and aromatic residues, but are significantly richer in "disorder-promoting" residues including proline and glycine [76]; the reverse of the requirements for forming α-helical and β-sheeted elements discussed earlier.

In addition to influencing the specific hydrogen-bonding network within the respective secondary structure element, distal side chain interactions of constituting residues could also have a significant influence on the final structure of the protein. The classical illustration would be the amphipathic characteristic of α-helices, which exhibit predominant nonpolar side chains on one surface and polar side chains on the opposite surface of the helical cylinder [77]. Such a polarized arrangement of residues with opposing physico-chemical properties formed the basis of tertiary structural frameworks such as the α-helical coiled coil, in which the specific heptad repeat motif of hydrophobic residues at helix interfaces allowed meshing of side chains in a knob-into-hole packing of helical bundles [78]. While most methodical analysis of protein folding were based on independent folding domains, the vast majority of eukaryotic proteins are likely to possess a multiple independent folding domain that could also exhibit cooperativity in the folding process [79].

Imino acid proline

Of the 20 standard amino acids that are found in nature, proline is without doubt one of the most unusual: the fusion of the backbone with the side chain atom forms a unique pyrolidine ring, and at the same time eliminates the amide proton. Both of these features confer conformation limitation and distinctive structural features to the polypeptide chain in which the imino residue resides [80]. The first consequence of the pyrolidine ring of proline residues is the introduction of a kink to the polypeptide backbone [60, 81].

While placement of a proline at the N-terminal of an α-helix highly stabilizes the helix, insertion of prolines at any other positions are generally considered to be highly disruptive to the formation and propagation of α-helices or β-sheets, as well as having possible structural impacts on the conformation of the protein. For this reason, statistical data have shown that proline residues are most frequently located within the first helical turn as a helix stabilizer, or immediately at the C-terminal of the helix as a capping residue [10, 80, 81].

The second structural consequence of proline's pyrolidine ring arises from the lack of amide proton which impairs the imino acid's participation in formation of classical secondary structural elements due to the inability to act as hydrogen bond donor to carbonyl groups of other residues. However, the $C\delta H_2$ protons of proline can form weaker hydrogen bonds with amide carbonyls of $i - 2$ or $i - 4$ residues in a helix, justifying for certain degrees of tolerance to proline residues in a number of globular and transmembrane proteins [82]. In other cases, conserved proline residues have also been identified in the middle of helices, possibly upholding specific structural or functional roles by inducing kinks in the length of the helices concerned [83].

In order to illustrate the structural implications of proline on protein folding, we will discuss a few examples of animal toxins, in which the influence of various structural features either in isolation or in combination can be clearly visualized on these highly conserved and compact structural frameworks [84]. In this case, the disulfide-pairing arrangement and consequent conformation of maurotoxin, a 34-mer potassium channel blocker isolated from scorpion venom, was shown to be drastically influenced by the presence of proline residues [85]. Despite possessing a conserved cysteine framework of the four-disulfide-bridged scorpion toxin, maurotoxin adopts a novel cysteine-pairing pattern (C1–C5, C2–C6, C3–C4, and C7–C8) that is distinct from the classical cysteine-pairing framework (C1–C5, C2–C6, C3–C7, and C4–C8) seen in Pi1 and HsTx1 toxins, the other two members of the same family. Sequence analysis of maurotoxin revealed two prolines flanking the non-conventional C3–C4 disulfide bridge. Substitution of these proline residues with alanines resulted in a conformational switch to the conventional cysteine pairing seen in Pi1 and HsTx1 toxins.

As a further illustration on the role of proline as a structural determinant of protein folding, the conserved proline residue in the first intercysteine loop of ImI α-conotoxin was previously identified by our group to have profound implications on the folding of this 12-mer peptide toxin. α-conotoxins and χ/λ-conotoxins share similar cysteine arrangement framework of 4 cysteine residues, but natively fold with distinct disulfide pairing patterns into the
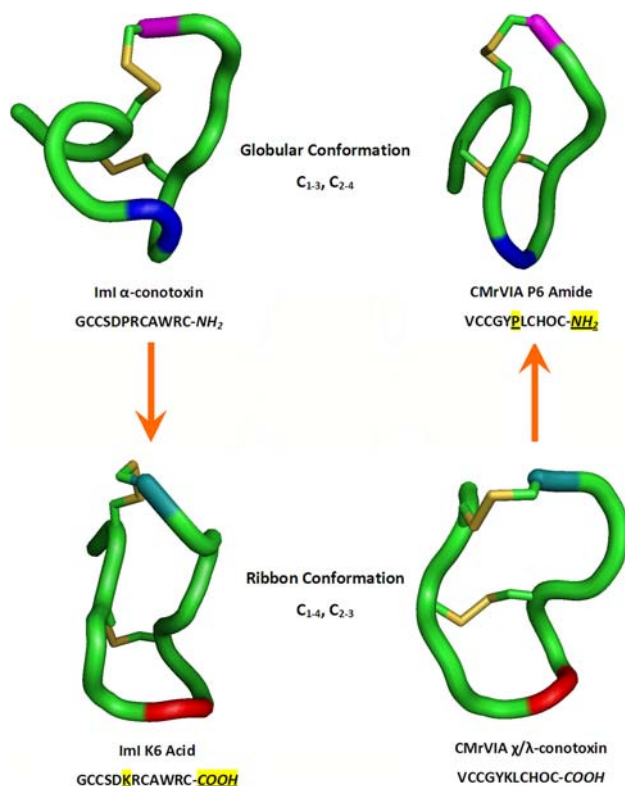
Fig. 2 Role of C-terminal amidation and proline residue on the folding of ImI α-conotoxin and CMrVIA χ/λ-conotoxin. Proline and lysine residues manipulated within the first intercysteine loop of the conotoxins are represented in *blue* and *red*, respectively. C-terminal amide and carboxylic acid are presented in *magenta* and *teal*, respectively

globular (C1–C3, C2–C4) and ribbon conformation (C1–C4, C2–C3), respectively [86, 87]. By substituting Pro6 with a conserved lysine residue from the corresponding position in χ/λ-conotoxins, it was possible to switch the native globular conformation of ImI α-conotoxin to the non-native ribbon conformation. These findings were corroborated by switching CMrVIA χ/λ-conotoxin's native ribbon conformation to the non-native globular conformation with a reciprocal substitution of Lys6 with a proline residue (Fig. 2) [88]. It was thus proposed that proline's tendency to bring two segments of the peptide backbone into juxtaposition acted as a conformational switch resulting in the preferential disulfide connectivity seen in ImI α-conotoxin. However, it is important to note at this juncture that changes of conformation might not necessarily imply changes in the folding pathway.

A similar importance of proline was also observed in the folding of larger proteins such as cytochrome P450, in which alanine substitutions of Pro30 and Pro33 in the conserved proline-rich region in CYP2C11 resulted in mutants that were misfolded and biologically inactive [89]. The purpose of the proline-rich region was hypothesized to

function as a hydrophobic knob for partial insertion of the protein into the membrane and to create a stable nucleus for initiation of protein folding.

Another unique structural characteristic of proline residues is the comparatively higher intrinsic propensity to adopt either the *trans* or *cis* peptide bond conformation with the preceding residue in the polypeptide chain [80]. It has been suggested that the ratio of the *cis:trans* prolyl peptide bond may be influenced by the residue following the proline residue and immediate environment of the proline residue [80]. While the precise reason for adopting the *cis* prolyl peptide bond are varied and not necessarily definitive, the increased propensity of proline residues to adopt the *cis* peptide bond is nonetheless significant, considering the fact that the majority of peptide bonds exist in the *trans* conformation [90].

Structurally, *cis-trans* isomerization of the prolyl peptide bond has been shown to be important for a variety of proteins such as Stefin B, a 11 kDA protein which forms amyloid fibrils by domain swapping between two dimers. The domain swapping involves extensive intermolecular contact as well as concurrent *trans-cis* isomerization. By substituting prolines for serines, it was shown that the conserved *trans* Pro74 in the monomeric and dimeric Stefin B is involved in the isomerization into the *cis* conformation to facilitate tetramerization that is crucial for amyloid fibril formation [91]. Functionally, the rate of *cis-trans* isomerization of prolyl peptide bond aids in the regulation of a variety of biological processes. For example, the rate of isomerization Pro213 in the hinge subdomain of N1–N2 domain of the gene-3-protein of phage fd has been shown to act as a kinetic block to the domain reassembly and conformational switch of the gene-3-protein, thereby regulating the rate of phage infection [92, 93]. The work by Kamen et al. [94] also indicated that prolyl *cis-trans* isomerization is implicated in the folding kinetics of proteins such as pectate lyase C.

Prolyl *cis-trans* isomerization is also a critical conversion in numerous cellular processes and is catalyzed by the four families of peptidyl prolyl *cis-trans* isomerases (PPIases). The rate of catalysis of prolyl *cis-trans* isomerization by these ubiquitous enzymes in turn act as molecular timers of cellular processes [95]. In fact, two of these PPIases are targets of immunosuppressant drugs [96, 97].

In this section, we have discussed the general propensities of the various amino acids in forming secondary structural elements. However, departures from the predicted propensities are not uncommon. For example, identical sequence segments have been identified in both β-sheeted and α-helical structures [98]. Peptide fragments termed "chameleon sequences" have also been reported in which the peptide segments are essentially ambivalent and the conformations of the segments concerned are largely

dependent on the local structure of the proteins [99]. In the earlier sections, we mentioned that the conformational restraining proline residue is a well-recognized helix breaker. Interestingly, polypeptide segments containing consecutive proline residues have also been reported as adopting a distinct nonconventional secondary structure termed the poly-L-proline helix (PPII) that presents as an all-*trans* helix with a left-handed turn. Unlike canonical α-helices, PPII helices are not held together by hydrogen bonds between i and $i + 4$ residues, but merely because of the steric constraints and physical rigidity of prolines in the helices. This further explains the left-handedness of PPII helices compared to right-handed canonical α-helices. The PPII helices have been reported to be important in diverse structural and biological functions including the formation of collagen [100], signal transduction, transcription, and immunological reactions [101].

Sequence-specific side chain aromatic interactions, such as π–π, cation–π, amide–π, and aryl–sulfur interactions, have also been implicated to be important contributors to various secondary structures such as β-turns [102, 103]. Engineered non-native aromatic side chain interactions have been shown to divert the conformation populations of ubiquitin into a complex folding landscape [104]. In other cases, specific steric arrangement of charged residues can also facilitate specific interaction and consequent conformational change. For example, the binding pockets of calcium binding proteins are coordinated by a bipyramidal or octahedral arrangement of negative charged residues, to which calcium ions facilitate conformational changes by overcoming repulsion of similar charges [105]. The importance of specific interaction can be further exemplified by the variety of exogenous stimuli such as the presence of metal ions, pesticides, variation in pH, temperature, protein concentration, and organic solvents that can result in partial protein unfolding and consequently amyloidogenesis [106, 107].

In summary, these observations indicate that, although propensities of the various amino acids can influence the structural characteristics and folding profile of the polypeptides, the context in which the residues are located, and the consequent differences in microenvironment and the addition of exogenous stimuli, can affect the protein folding.

Posttranslational modifications and folding

Posttranslational modification of proteins is nature's strategy for widening the diversity of the repertoire of polypeptides assembled from the standard list of 20 natural amino acids. Currently, over 200 kinds of posttranslational modifications have been documented, and they may be broadly classified as (1) covalent addition/ modifications of functional group to side chains, and (2) hydrolytic cleavage of peptide bonds [53]. Most of these posttranslational modifications result in functional consequences. In addition, several forms of posttranslational modifications have structural implications for the folding and conformation of the proteins. In the first part of this section, we shall discuss the various representative types of posttranslational modifications that bear direct implications to protein folding. In the remaining parts, we will touch on several forms of posttranslational modifications that are important for modifying local conformation of the tertiary structure and biological functions of proteins, but are not generally known to impact global protein folding. This is especially since some of these modifications occur only after the protein folding process has been completed in the endoplasmic reticulum.

It should be emphasized that this section is not intended to be encyclopedic on the plethora of possible posttranslational modifications, but rather, to illustrate with selected examples on the significance and the possible impact of these modifications on protein folding (Table 1).

Oxidation-reduction

One of the most important and commonly observed posttranslational modifications of protein is the oxidation of the free thiol group of cysteine residues and the consequent formation of disulfide bridges. Formation of disulfide bridges is influenced by the proximity and accessibility of the thiol groups on the polypeptide chain, and catalyzed by protein disulfide isomerase and Dsb family of proteins in eukaryotic and prokaryotic systems, respectively [108]. These posttranslational covalent linkages play critical roles in the structural stability of proteins [109]. Consequently, disulfide engineering had been used as an approach to stabilize proteins [110]. Disulfide linkages are also directly implicated in biological functions, such as catalytic disulfide bridges in oxidoreductases [111] and allosteric disulfide bridges that regulate protein function by triggering conformational changes [112]. In addition, the kinetics of the protein folding process have also been suggested to be influenced by the relative location of disulfide bridges to the folding nucleus [110].

Disulfide-rich proteins are frequently chosen as models for studies of the in vitro protein folding process due to the possibility of trapping disulfide folding intermediates, thereby allowing the reconstitution of folding pathways. Further, disulfide connectivity can possibly improve the overall folding rate of proteins by the formation of productive disulfide bridges that bring about the necessary interactions for the formation of a stable conformation. In other cases, the hydrophobic disulfide bridge may also be

Table 1 Examples on the significance and the possible impact of these modifications on protein folding

Posttranslational modification	Consequence	Examples	Structural implications	References
Protein folding				
Oxidation-reduction	Formation of disulfide bridges	Insulin superfamily, animal venoms, BPTI	Structural stability	[114, 117, 126]
			Putative role in improving protein folding rate	
			Cooperativity in folding by multiple disulfide bridges	
	Oxidation of methionine to methionine sulfoxide/ sulfone	Transthyretin, λ-repressor protein, calmodulin, thrombomodulin	Disruption of native non-covalent interactions	[130–134]
Glycosylation	Covalent attachment of carbohydrate moiety to conserved glycosylation sites	Lysosomal hydrolases, coagulation factors, antibodies, erythropoietin, contalukin-G	Augment solubility and stability profile	[135, 136, 148, 149, 163, 238]
			Promote folded conformation by reducing conformational space	
			Formation of new interactions between glycan and polypeptide	
Hydroxylation	Stereospecific hydroxylation of proline residue	Collagen, conotoxins	Stabilization of trans isomer of hydroxyprolyl peptide bond	[160, 164]
			Improve rate of oxidative folding in α-conotoxin model	
Carboxyamidation	Substitution of C-terminal carboxyl group with amide group	Venom proteins, as well as peptide hormones such as vasopressin, oxytocin, and calcitonin	Modification to electrostatic interactions	[163, 169, 239]
γ-Carboxylation	Carboxylation of glutamic acids	Osteocalcin, conantokins, as well as various other calcium binding proteins including coagulation factors and clotting inhibitors	Facilitate calcium-induced folding	[163, 177, 178]
Protein conformation				
Phosphorylation	Covalent attachment of phosphoryl group to serine, threonine, and tyrosine in eukaryotes, histidine and aspartate in prokaryotes	Glycogen phosphorylase, isocitrate dehydrogenase, as well as a large number of proteins involved in signal transduction and numerous critical cellular processes	Change in protein's tertiary conformation due to the dianionic phosphoryl group participating in hydrogen bonds	[187]
Methylation/acetylation	Covalent attachment of methyl or acetyl group to ε-amino group of lysines or arginines	Histone proteins	Augment specific cation–π interactions	[190, 191]
Farnesylation	Attachment of farnesyl group to CAAX C-terminal motif	(SNARE) Ykt6 protein	Increase the α-helical content and stability of protein	[194]
Nitration	Attachment of nitrate group to tyrosine	Mn superoxide dismutase, α-synuclein, tyrosine hydroxylase, tau protein	Modification to hydrophobic and electrostatic interactions	[193]
Sulfation	Transfer of sulfuryl group to tyrosine	Glycoprotein Ib$\alpha\alpha$	Contribution of negative charges	[240]
Citrullation	Conversion of arginine into citrulline residue	Fillagrin, trichohyalin, histone proteins	Charge reversal results in the disruption of electrostatic interactions and hydrogen bonds	[241]

Table 1 continued

Posttranslational modification	Consequence	Examples	Structural implications	References
Polyglutamylation	Elongation of polyglutamate chain at specific glutamic residues of the protein	α-tubulin, β-tubulin, nucleosome assembly proteins NAP1 and NAP2	Possible roles in the conformation and interaction of target proteins	[242]
Polyglycylation	Covalent attachment of multiple glycyl units to γ-carboxyl group of glutamate	α-tubulin, β-tubulin	Possible roles in the conformation and interaction of target proteins	[243]

involved in the formation of hydrophobic clusters that are necessary as the folding nucleus to initiate the folding process [110, 113]. The presence of multiple disulfide bridges has also been demonstrated to have a potential positive influence on the folding. For example, in several small disulfide-rich proteins, such as BPTI and proinsulin, the interactions between the three sets of disulfide bridges resulted in cooperativity of folding by lowering the conformational entropy necessary for interaction [114–116].

The insulin superfamily consists of small disulfide-containing globular proteins that share a common structural motif, and they have been extensively studied for the contribution of disulfide bridges to the structure and folding of proteins [117]. The insulin molecule, for example, consists of two chains that are linked by one intrachain disulfide bridge, A6–A11, and two interchain bonds, A7–B7 and A20–B19. It was shown that deletion of the intrachain disulfide bridge negated the α-helical N-terminal domain of the A–chain and the activity of insulin, while deletion of interchain disulfide A7–B7 resulted in more severe structural disturbances. However, deletion of either pair of disulfide bridges did not compromise the in vitro refolding of the proinsulin molecule, suggesting that these connectivities were non-critical in the folding of insulin molecule [118, 119]. In contrast, the interchain A20–B19 disulfide bond has been noted to be critical in stabilizing the folding intermediate [120]. Deletion of this critical disulfide bridge significantly disrupted the ordered secondary structure and increased the proinsulin's susceptibility to proteolysis [119]. At the other extreme of the spectrum, multiple disulfide bridges have occasionally been noted to be redundant, with folding and structural stability maintained with a reduced number of native disulfide bridges [121, 122].

Due to the conformational restraint exerted by disulfide bridges, permutations of cysteine connectivity in cysteine-rich polypeptides almost invariably result in differing conformations. As such, disulfide bond connectivities are typically considered to be non-promiscuous; a trait that is critical for the accurate disulfide connectivity and conformation of numerous disulfide-rich proteins and peptides.

However, there are examples in which one cysteine framework can result in more than a single possible disulfide connectivity. For example, the insulin-like growth factor I (IGF I), which is a member of the insulin superfamily discussed earlier, shares homologous structural motifs with other members of the same superfamily such as porcine insulin precursor and relaxin [120]. However, unlike other members of the superfamily, IGF I folds into two distinct conformations [123]. It was subsequently determined that sequence 1–4 and residue 9 of the B chain of IGF I were critical in determining the disulfide connectivity of the protein [124]. A further example of multiple disulfide isoforms arising from single cysteine template is that of the epidermal growth factor-like (EGF-like) domains of thrombomodulin, in which the fifth EGF-like domain exhibits a disulfide bonding pattern distinct from other EGF or EGF-like domains [125].

Non-conventional disulfide linkages have also been observed in other classes of cysteine rich peptides such as conotoxins derived from the venom of the marine predatory cone snails. Conotoxins are classified based on their highly conserved cysteine framework and their distinct biological activity. Unlike the majority of conotoxins that exhibit strict fidelity to their respective disulfide connectivity distinct biological activities, conotoxins with the fourteen-cysteine, six-cysteine and four-cysteine conotoxin scaffolds can each adopt at least two distinct conformations with differently connected disulfide bridges [126]. In our earlier work, we have shown that the conserved Pro6 and C-terminal amidation are instrumental in determining the disulfide connectivity of ImI α-conotoxin (Fig. 2) [88, 127]. In the case of GI α-conotoxin, Zhang and Snyder [128] demonstrated that the propensity of the peptide toxin to fold into the native globular conformation is largely dependent on the loop sizes of the two intercysteine loops. Similarly, the disulfide connectivity of charybdotoxin from the venom of the scorpion, *Leiurus quinquestriatus*, was influenced by the position of the cysteine residues and intercysteine loop spaces [122]. In an engineered barnase structure, the inclusion of disulfide bridges reduced the rate of protein unfolding [129]. Taken together, these findings

show that disulfide bonds not only play an important role in the structure and function but also the folding process of the proteins.

The other residue of the 20 standard amino acids that is prone to oxidation is methionine, which also contains a polarizable sulfur atom. The thioether side chain of methionine may be oxidized to methionine sulfoxide or methionine sulfone, and has been implicated in several diseases. This includes Alzheimer's disease and Parkinson's disease, as well as various disease states relating to oxidative damage. Conversely, oxidation of the amyloidogenic transthyretin prevents the formation of fibril formation by disrupting the aggregation interface [130]. Oxidized methionine has been demonstrated to affect the local environment within the protein, and has also been utilized to stabilize the denatured state of monomeric λ-repressor protein, and to disrupt the conformational stability of calmodulin and thrombomodulin [131–133]. These observations indicate that oxidation of methionine potentially destabilizes non-covalent interactions that are normally critical in maintaining the stability of certain proteins. As a corollary to this, it is highly possible that methionine oxidation can potentially influence protein folding. An illustration of the influence of methionine oxidation on protein folding was demonstrated in the work by Dado et al., in which the oxidation of specific methionine residues in a designed 18-mer peptide resulted in the transition of the α-helical conformation towards the β-strand [134].

N-linked glycosylation

N-linked glycosylation is another common form of post-translational modification, with nearly 50% of all human proteins being glycosylated [135]. Understandably, glycosylation must play an integral role in the trafficking, folding and function of eukaryotic proteins. The glycosylation process essentially occurs in three stages, in which the first phase occurs while the nascent polypeptide is still in the translocon complex and is associated to the endoplasmic reticulum-associated degradation (ERAD) system. Aside from participation in the folding of the polypeptide, the glycan indirectly influences the protein folding process by serving as an intracellular sorting signal on the folding status of the protein [136]. The second phase largely concerns the trimming of the glycan for direct targeting of the proteins, as in the case of lysosomal hydrolases [137]. The final stage of modification is effected in the target site of the protein [136]. Of particular interest in this context is the influence of glycosylation on the folding of the protein.

The carbohydrate moiety contributes to the folding of the protein in several ways. Firstly, glycosylation has been known to improve the overall stability of proteins [138],

and has been utilized to improve the circulation half-life and activity of erythropoietin analogues [139] as well as selective delivery of enzymes into lysosomes of target tissues [140–144]. The large hydrophilic carbohydrate group exerts a biophysical influence on the polypeptide by increasing the solubility of the protein and augmenting the stability of the protein by covalently attaching to the protein's surface. In silico examination of protein models also illustrates that the protein stabilizing effect stems from the destabilization of the unfolded state by the glycan moiety [145]. Secondly, although the glycan moiety does not directly result in secondary structure formation, the bulk of the relatively large glycan moiety promotes the folded conformation by decreasing the available conformational space, as well as the mobility of the unfolded conformation [136, 146, 147]. For example, sequential deglycosylation of human $\beta2$ glycoprotein I resulted in an increased change in the secondary structure of the native protein as determined by circular dichroism, reiterating the steric hindrance imposed by the carbohydrate moiety [148]. Thirdly, the addition of the large glycan moiety resulted in the formation of new interactions between the carbohydrates and amino acids and/or stabilized the protein fold by counterbalancing an unfavorable clustering of charges [149]. In addition to the global effect of improving solubility of the polypeptide, interactions between the N-acetyl group of glycans and the polypeptide have also been shown to affect the local conformation of the peptide backbone by promoting β-turns [150].

Examination of a glycosylated loop heptadecapeptide derived from the α subunit of the neuromuscular nicotinic acetylcholine receptor of *Torpedo californica* revealed that glycosylation of the conserved glycosylation site shifted the *cis-trans* equilibrium of a constituent proline residue towards the *trans* isomer, and the dithiol/disulfide equilibrium towards the oxidized species [151]. These observations demonstrate that steric and electrostatic hindrances conferred by the glycan can significantly affect the folding and structure of the protein. In addition to influencing the folding of proteins into secondary and tertiary structures, glycosylation has also been demonstrated to affect the quaternary assembly of protein monomers into higher order assemblies. For example, a bulky N-linked heptasaccharide binds to Asn17 of *Erythrina corallodendron* lectin, and forces the dimer to adopt a quaternary structure that is structurally non-identical to other non-glycosylated homologous lectins [152].

However, not all glycoproteins are equally dependent on the glycosylation for their folding and secretion [145]. While most eukaryotic proteins require glycosylation at some point in the folding process [153], the folding rate and pathway of several examples have been reported to be unaffected by the glycosylation state [154–156].

Further, not all glycosylation sites within a polypeptide chain possess the same bearing on the folding process. For example, of the five glycosylation sites on glucocerebrosidase, only four are *N*-glycosylated. Of these, only Asn19 was shown to be truly critical for the enzyme's function, while deglycosylation to the other sites had comparable catalytic efficiency [140]. Similar trends were also observed in influenza hemagglutinin as well as in several lysosomal hydrolases such as α-galactosidase and α-glucosidase, in which only one of the several attached glycans is truly essential for activity [157–159]. Therefore, although the specific mechanics for predicting critical glycosylation sites that are deemed critical for the folding pathway of proteins are still ambiguous, the contribution of these covalently linked glycans to the structure and function of proteins are beyond doubt.

Hydroxylation

Although hydroxylation of amino acids is generally not as common as the other forms of posttranslational modifications, it is the most frequently observed form of posttranslational modification seen in the human, in the form of hydroxyproline in collagen [160]. Collagen is ubiquitously found in the structural tissue of vertebrates with an exceedingly high proportion of hydroxylated proline that forms the characteristic Xaa-Yaa-Gly triad repeats in its tropocollagen triple helices coiled coil. The stereoselective hydroxylation results in a (2*S*,4*R*)-4-hydroxyproline (Hyp) in the Yaa position of the triad repeat of the collagen's primary structure [161]. While it was previously suggested that the hydroxyproline stabilizes the collagen triple helix via bridging water molecules [162], it was demonstrated through *O*-methylation of hydroxyproline that such conformational stability conferred by the hydroxylation was contributed by the "stereoelectronic" properties of the stereospecific hydroxyl group that promotes the proper conformation for triple helix formation [160, 161].

Apart from playing critical roles in the structural framework of collagen in vertebrates, hydroxylation of proline residues has been shown to have important influence on protein folding. For example, hydroxyproline is commonly observed across several families of conotoxins derived from the marine predatory cone snails [163]. While the functional roles of the hydroxylation remains uncertain, work by Lopez-Vera et al. demonstrated that hydroxylation of proline residues can have an important influence on protein folding rates [164]. In contrast, hydroxylation of proline not only improved the rate of oxidative folding of the α-conotoxin model, but also improved the in vitro folding rate of the ω-MVIIC that is notoriously difficult to fold [164, 165]. It has been suggested that the electron-withdrawing effect of the hydroxyl group stabilizes the

trans isomer of the hydroxyproline about the peptide bond, thereby contributing to the improved folding propensity [161]. It was further proposed that proline hydroxylation and disulfide formation might occur concurrently and cooperatively within the endoplasmic reticulum [126, 166]. In another example, hydroxylation contribute to the bioactivity of μ-conotoxin GIIIA, but was not observed to be beneficial to the folding of the peptide toxin [164].

Carboxyamidation

Carboxyamidation has been noted to be an important form of posttranslational maturation of polypeptides, especially in the cases of peptide hormones [167–169]. While there are several examples in which C-terminal amidation of the peptides do not result in significant functional disparity [169, 170], non-amidated forms of several peptides have been demonstrated to suffer significant reduction in biological potency [171–175].

Despite the fact that carboxyamidation results in a subtle substitution of a negatively charged carboxylic terminal with an amide group at the distal end of the polypeptide chain, such modification has been shown to be sufficient to result in conformational changes of the peptide's conformation. For example, in the case of eumenine mastoparan-AF (EMP-AF) isolated from the wasp venom, although both the native amidated and the synthetic deamidated forms possess amphipathic helical structures, the amidated peptide exhibited a well-defined helical structure with enhanced rigidity. In contrast, the C-terminal region of the deamidated analog exhibited higher degrees of conformational freedom due to charge repulsion and hydrogen bond disruption at the C-terminal segment of the synthetic peptide variant [169]. Similar consequence of conformational disruption of C-terminal region was also noted in the deamidated form of the crustacean hyperglycemic hormone Pej-SGP-I, in which the native amidated form was reported to be significantly more potent [171].

C-terminal amidation has also been observed in several families of the conotoxins derived from the marine predatory cone snails [163]. Unlike the previous examples in which the deamidated analogues exhibit conformational disruption at the C-terminal region, we have earlier shown that C-terminal amidation potentially determines the disulfide connectivity to specific families of these cysteine-rich peptide toxins, acting as a conformational switch. For example, the native ImI α-conotoxin folds predominantly into the globular conformation with a specific C_{1-3},C_{2-4} disulfide connectivity. In contrast, the synthetic carboxylate analog folded predominantly into the non-native ribbon conformation with C_{1-4},C_{2-3} disulfide connectivity and the consequent reduction of biological potency [88, 127]. In contrast, C-terminal amidation did not seem to influence

the folding of other families of conotoxins [176]. These findings indicate that, although not consistently applicable, the modification to the electrostatic interactions brought about by C-terminal amidation can substantially influence the final conformations of proteins.

γ-Carboxyglutamic acid

γ-Carboxylation of glutamic acid is an important posttranslational modification that is particularly pertinent for hemostasis. Specifically, posttranslational carboxylation of specific glutamate residues is indispensible in the sequences of the several coagulation factors, osseous tissues, or procoagulant factors derived from venom sources [177, 178]. These γ-carboxyglutamate residues (Gla) function by acting as binding sites for calcium ions, which are a ubiquitous component of critical cellular processes [179].

Posttranslationally γ-carboxylated glutamic acids are usually observed in specific calcium-binding proteins. Despite the wide tissue distribution of the vitamin K-dependent carboxylase that is responsible for this modification, Gla residues are generally less commonly implicated to the structure or function of other proteins [180]. However, this unusual modification has been shown to be present in several peptide toxins derived from *Conus* venom, and, in the case of conantokins, multiple Gla residues are usually present within the short peptide sequences [181, 182]. It has been demonstrated that the binding specificity of Gla residues to calcium ions facilitates the folding of several conantokins into their mature conformations. For example, addition of calcium ions resulted in a significant increase yield and rate of folding of the "spasmodic peptide" derived from *C. textile* [183]. Similar calcium-induced folding was also noted in several conantokins, and is particularly significant considering the fact that the majority of conantokins lack disulfide bridges, and the presence of Gla residues is the predominant driving force for the peptides to adopt their mature helical conformation and/or self-assembly into dimeric structures [184, 185]. In the case of conantokin-P which possesses a single disulfide bridge, the addition of calcium ions to the folding buffer accelerated the oxidative folding of the peptide, reinforcing the determining influence of Gla residues on the folding of these small peptides.

Posttranslational modifications affecting conformation of mature proteins

In addition to the various forms of posttranslational modifications discussed above, a variety of other forms of modifications, though not directly implicated in protein folding, have been shown to influence the conformation and the consequent functions of mature proteins. One of the most pervasive forms of posttranslational modification is the phosphorylation of proteins executed by the ∼500 specific kinases, providing an intracellular switch to diverse cellular processes [52, 53]. By covalently attaching a tetrahedral phosphoryl group to amino acids' neutral hydroxyl side chain, phosphorylation brings about conformational changes to the proteins' tertiary structures or assembly interfaces, thereby effecting cellular functions and processes [186]. Such conformational changes are possibly brought about by the dianionic phosphoryl group in physiological pH that can participate in hydrogen bonds, such as those with main-chain nitrogens of an α-helix or with the guanidinium side chain of arginine residues [187].

Methylation of lysine in histone proteins has been known to be involved in the chromatin condensation by interaction of the methylated lysine with three aromatic side chains. Specifically, the methylated histone tail of histone H3 inserts as a β-strand into the β-sandwich chromodomain of HP1 protein [188]. Based on this finding, a β-hairpin peptide corresponding to the histone tail was synthesized and found to be stabilized by the specific cation–π interaction between Trp2–Lys9 [189]. Further work by Hughes et al. [190, 191] indicated that N-methylation of the lysine or arginine residues augment the magnitude of this interaction, resulting in an exceptionally stable β-hairpins. Methylation of the ε-amino group of lysine altered the interactions with this polar side chain by conferring hydrophobic interactions, as well as generating a more polarized methylammonium group for cation–π interaction [188]. Another posttranslational modification known to influence the lysine residue of histone proteins is the acetylation of the ε-amino side chain [192]. Although acetylation of lysine alters the geometry of interaction and neutralizes the positive charge of lysine's protonated side chain that is essential for cation–π interaction described earlier, amide–π interaction was found to be at least as strong as the cation–π interaction between lysine and the aromatic ring [103]. Taken together, methylation and acetylation of lysine on histone proteins illustrate the specificity of histone-DNA interaction, as well as provide insights on protein folding.

Other forms of posttranslational modifications include nitration of tyrosine residues, lipidation by farnesylation, citrullation of arginine residues, and polyglutamylation and polyglycylation of glutamate residues (Table 1). Nitration of tyrosine residues modifies the hydrophobic and electrostatic interactions of the target protein, thereby disrupting the conformation and activity of the protein [193]. It is for this reason that nitration of tyrosine residues has been implicated in a variety of neurodegenerative diseases. Lipidation by farnesylation of the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) Ykt6 protein was shown to increase the

α-helical content of the 23 kDa protein as well as to stabilize the final conformation [194].

A minimalist approach

Inferences from Christian Anfinsen's theory of protein folding suggested that the amino acid sequence of a polypeptide encodes for sufficient information required for the native folding of the protein. However, recent research indicates that, as much as specific-sequence information is necessary, a significant degree of sequence degeneration and redundancy exists in the amino acid syllabus and only a fraction of the encoded information is likely to be necessary for fully defining the native protein structure [195, 196]. Studies on the site-directed mutagenesis of various proteins have concurred with the theory of redundancy of protein sequence. For example, by selecting residues that were not directly involved in structural or functional roles, the exhaustive analysis of the *lac* repressor protein indicated that mutations made at nearly half the 142 sites on the protein were phenotypically silent [195, 197–199]. Tolerance to single or multiple substitutions was also seen in the Arc repressor protein [200], bovine pancreatic trypsin inhibitor [201, 202], and RNA-binding protein ROP [203], among others. In addition to sequence redundancy, several forms of posttranslational modifications discussed earlier also exhibited certain degrees of redundancy.

In addition to the redundancy of sequence information, it was also observed that there was a generally lower degree of conservation of surface-exposed residue side chain as compared to the buried nonpolar residues, with the former group generally being more tolerant of sequence substitution [195]. Such a difference in evolutionary conservation reiterates the importance of the nonpolar or neutral residues that make up the hydrophobic core structure of proteins [204]. Although some exceptions exist, the segregation of hydrophobic and hydrophilic residues into the buried and surface exposed positions of the final structure generally fulfils the requirement of protein solubility and facilitates the folding process of the protein [205]. Clearly, the amphipathic pattern of hydrophobic/hydrophilic residue distributions must represent some of the protein folding instructions encoded in the polypeptide [195]. To this, two general models of protein folding have been derived. The first concept of "jigsaw puzzle model" was proposed by Chothia et al. for α-helices in which the close packing and complementarity of specific side chains define the basic architecture of the backbone of the protein [206]. This model was initially supported by work examining interactions concerning hydrophobic interfaces which indicated that changes in charge and volume of residues located within the core of proteins are generally destabilizing [207, 208].

However, a number of combinations of amino acids have been substituted into core sequences and yet resulted in viable proteins [209–213]. The second model, termed "oil-droplet model", posits that periodic binary distribution of hydrophobic and hydrophilic residues permits the formation of structural elements by hydrophobic interactions, without the need for specific inter-residue contacts [214, 215]. Nonetheless, drastic changes in shapes or volumes of the hydrophobic residues concerned can still result in changes in conformation of the tertiary structure [216]. Taken together, the hydrophobicity described by the "oil-droplet model" is recognized to be important. However, interdigitation of core residue side chains, though critical in selected models, may not necessarily be as strictly dictated as by the "jigsaw puzzle model". Although each of the two models of protein folding can only provide a limited perspective to the folding process, they are, nonetheless, sufficient for the design and generation of empiric secondary structure elements.

One of the early designs of de novo protein using the binary distribution of hydrophilic and hydrophobic residues was accomplished by Kamtekar et al., in which a four-helix bundle protein was empirically designed [214]. The degeneracy of sequence was demonstrated by using phenylalanine, leucine, isoleucine, methionine, and valine for the nonpolar residues, while polar residues were fulfilled by using glutamine, glutamate, aspartate, asparagine, lysine, or histidine. Through the use of merely 11 out of the 20 available natural amino acid residues, the designed protein correctly folded into the intended tertiary conformation, and exhibited a native-like hydrophobic/hydrophilic residue distribution. Further work by Schafmeister et al. showed that it is possible to design a larger four-helix bundle using just seven amino acids, with the designed protein yielding crystal structure of 2.9 Å resolution [217]. The concept of a minimized alphabet for protein folding was subsequently extended to the design of a small β-sheeted protein, the SH3 domain, in which five amino acids (isoleucine, lysine, glutamate, alanine, and glycine) were sufficient for coding the 53-residue domain [218]. In a more recent article by Taylor et al., it was demonstrated that the enzyme AroQ chorismate mutase could be simplified to binary modules of four polar and four nonpolar residues [219]. It was also subsequently demonstrated that a fully functional variant of the enzyme could be constructed with a mere nine-letter amino acid alphabet [220]. Thus, the possibility of designing basic secondary structure elements using a binary approach reaffirms the importance of hydrophobic interaction in the protein folding pathway. It also suggests the possibility of designing larger and more complex protein architecture. In a bold reverse engineering attempt by Silverman et al. [221], the prototypical $(\beta/\alpha)_8$ TIM barrel domain of the triosephosphate enzyme was shown to exhibit
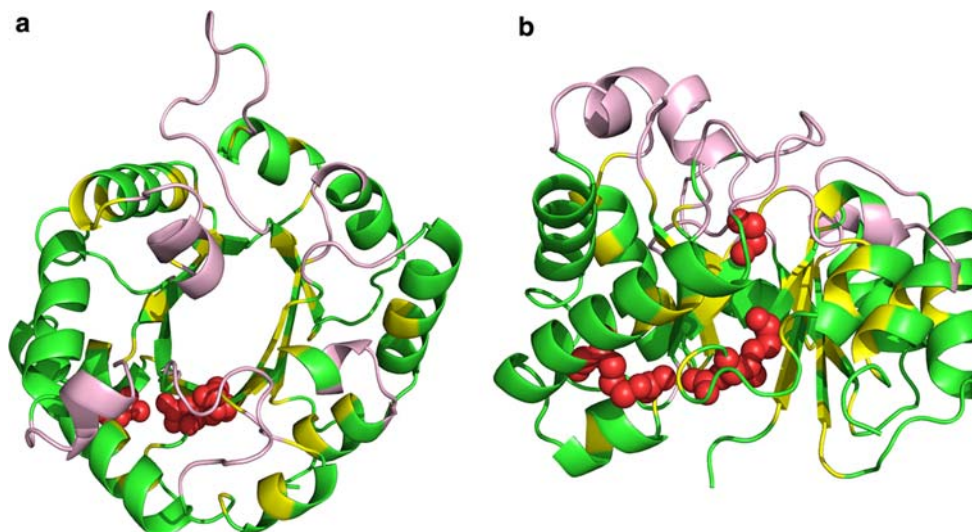
Fig. 3 Mutagenesis analysis of yeast triosephosphate isomerase (TIM)(PDB ID 7TIM). A total of 182 out of the 250 residues of the TIM protein were systematically mutated with one of the 7-letter amino acid alphabet (FVLAKEQ) reduced from the original 20-letter alphabet. Conservative substitutions were made based on either structural features or phylogenetic conservation between various TIM analogues. The top and side views of the TIM molecule are shown in **a** and **b**, respectively. Residues highlighted in *green* represent the 97 singly mutated positions that resulted in minimal disruption of functional activity of the protein as well as the 45 residues of the original protein that contain letters of the 7-letter alphabet. *Segments in yellow* represent the 36 positions that resulted in intermediate loss of protein activity, while residues represented by *red spheres* are the four immutable residues R189, D227, G209, and G228. *Segments in pale pink* are residues not examined in the mutagenesis analysis

sequence degeneracy and could be encoded with a reduced amino acid alphabet. It was determined that 142 out of the 182 examined sites on the enzyme's ~250-residue sequence could be mutated to one of the seven-letter amino acid alphabet and result in viable enzymes (Fig. 3).

Having discussed the propensities of the various amino acids in influencing protein folding, the influence of post-translational modifications, and the redundancy of sequence information, it should not be neglected that a plethora of short sequence motifs are possible with the various permutation and combinations of these residues. Specific sequence motifs have been defined for various local conformational motifs. These include, but are not limited to, "tyrosine corner" in the Greek key β-barrel [222], consensus sequences resulting in Type I β-turns [74], stacking of aromatic residues with a proline residue to form Type VI β-turn [223], packing motifs for transmembrane helices [224], and proline-rich motifs in cytochrome P450s [89], as well as numerous others [55, 81]. These specialized folding motifs exert specific structural and biological properties that fall beyond the scope of the current review.

Challenges of protein folding

Thus far, we have discussed some of the structural features of proteins that have been reported to influence the folding of the protein. While several of the features described were inherent properties of the basic amino acid sequence and concur with the theory proposed by Christian Anfinsen at a time when understanding of protein folding was still at its infancy, several structural determinants were made post-translationally to the polypeptide chain. However, it is still not possible to conclusively claim universal applicability for these structural "determinants". Indeed, it is the uniqueness of each polypeptide's amino acid sequence and the subsequent maturation steps that make the protein folding problem particularly challenging to decipher. A classical illustration of the distinctiveness of proteins can be seen in amyloidogenesis: while specific proteins have been shown to be prone to fibrillation, proteins that are unrelated to any known disorder can also be converted into aggregates that exhibit similar structural features to amyloid fibrils [106, 107, 225]. It is important to emphasize that determinants discussed in this review are merely a fraction of the knowledge demanded of the intricately complex picture of protein folding. In essence, the determinants of correct folding (or in the case of amyloidogenesis, protein misfolding) are neither completely understood nor can they be simplistically limited to structural features on the sequence. Several other aspects of protein folding that are at least as important include the thermodynamics and kinetics of protein folding, folding intermediates and pathways, and computation prediction [7–11].

A principal goal of understanding the structural determinants of protein folding is essential to appreciate the "language of proteins". From a clinical perspective, such knowledge could facilitate the understanding of

conformational diseases [226], in which the misfolding of endogenous proteins resulted in pathological consequences [227–229]. From a biochemical perspective, one of the ultimate aims of protein design would be the design of tailored protein structures, with precisely tailored biological functions. Although this has yet to be achieved, the understanding of the structural determinants coupled with increased computational capabilities has facilitated the design of various de novo small proteins with specific secondary structures intended in mind [230–233]. In the work by Kuhlman et al. [232], the accuracy of fold of the designed protein could even be predicted to an impressive atomic-level of accuracy! By reducing the degeneracy of encoding residues, several natural proteins including the prototypical $(\beta/\alpha)_8$ barrel structure have been successfully redesigned with a much smaller subset of the amino acid alphabet [217–221]. Another successful approach of protein design utilized the specific coordination of negatively charged residues to form calcium-inducible folding of the target protein [105]. The current computational prowess also led to the reverse design of a zinc-free "zinc finger" domain [234]. An alternative approach to successful de novo design utilized extensive combinatorial libraries instead of the classical rational design of sequence space, through which the optimized sequence for a desired trait or conformation was identified by binary patterning [215]. Thus, it has also been recently suggested that the current state of technology can possibly facilitate the de novo design of proteins with native-like properties of up to or surpassing 100 residues by using first principles [235].

The protein folding problem is a complex topic that has consumed several decades of research. With the recent advances in structural biology and theoretical and computational approaches, as well as the various biophysical techniques, including a variety of fast spectroscopic assays that allows monitoring of the folding process in real-time, it is now often possible to accurately predict the three-dimensional structures of small proteins through template-based models [236, 237]. While it is still a challenge to accurately predict the structures of larger, multi-domain proteins from their empiric amino acid sequences [14, 41, 236], the recent advances in the various aspects of the protein folding suggest that the solution to this grand biochemical challenge may one day be in sight.

References

1. Honig B, Cohen FE (1996) Adding backbone to protein folding: why proteins are polypeptides. Fold Des 1:R17–R20

2. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

3. Anfinsen CB, Haber E, Sela M, White FH Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA 47:1309–1314

4. Levinthal C (1968) Are there protein folding pathways? J Chim Phys 65:44–45

5. Levinthal C (1969) How to fold graciously. In: DeBrunner JTP, Munck E (Eds) Mossbauer spectroscopy in biological systems. University of Illinois Bulletin, vol 67, pp 22–24

6. Karplus M (1997) The Levinthal paradox: yesterday and today. Fold Des 2:S69–S75

7. Creighton TE (1986) Disulfide bonds as probes of protein folding pathways. Methods Enzymol 131:83–106

8. Baldwin RL (2008) The search for folding intermediates and the mechanism of protein folding. Annu Rev Biophys 37:1–21

9. Fine R, Dimmler G, Levinthal C (1991) FASTRUN: a special purpose, hardwired computer for molecular simulation. Proteins 11:242–253

10. Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13:222–245

11. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc Natl Acad Sci USA 89:8721–8725

12. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys 37:289–316

13. Wetlaufer DB (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci USA 70:697–701

14. Fersht AR (2008) From the first protein structures to our current knowledge of protein folding: delights and scepticisms. Nat Rev Mol Cell Biol 9:650–654

15. Ptitsyn OB (1995) How the molten globule became. Trends Biochem Sci 20:376–379

16. Kim PS, Baldwin RL (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51:459–489

17. Karplus M, Weaver DL (1994) Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci 3:650–668

18. Ptitsyn O (1996) How molten is the molten globule? Nat Struct Biol 3:488–490

19. Fersht AR (1995) Optimization of rates of protein folding: the nucleation–condensation mechanism and its implications. Proc Natl Acad Sci USA 92:10869–10873

20. Fersht AR (1997) Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7:3–9

21. Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V, Fersht AR (2003) Unifying features in protein-folding mechanisms. Proc Natl Acad Sci USA 100:13286–13291

22. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489

23. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224:771–782

24. Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Gianni S (2008) Mechanisms of protein folding. Eur Biophys J 37:721–728

25. Dyson HJ, Wright PE (2005) Elucidation of the protein folding landscape by NMR. Methods Enzymol 394:299–321

26. Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622

27. Borgia A, Williams PM, Clarke J (2008) Single-molecule studies of protein folding. Annu Rev Biochem 77:101–125

28. Forman JR, Clarke J (2007) Mechanical unfolding of proteins: insights into biology, structure and folding. Curr Opin Struct Biol 17:58–66

29. Balakrishnan G, Weeks CL, Ibrahim M, Soldatova AV, Spiro TG (2008) Protein dynamics from time resolved UV Raman spectroscopy. Curr Opin Struct Biol 18:623–629

30. Schaeffer RD, Fersht A, Daggett V (2008) Combining experiment and simulation in protein folding: closing the gap for small model systems. Curr Opin Struct Biol 18:4–9

31. Dokholyan NV (2006) Studies of folding and misfolding using simplified models. Curr Opin Struct Biol 16:79–85

32. Caflisch A (2006) Network and graph analyses of folding free energy surfaces. Curr Opin Struct Biol 16:71–78

33. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. Curr Opin Struct Biol 14:76–88

34. Brockwell DJ, Radford SE (2007) Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol 17:30–37

35. Gianni S, Ivarsson Y, Jemth P, Brunori M, Travaglini-Allocatelli C (2007) Identification and characterization of protein folding intermediates. Biophys Chem 128:105–113

36. Sanchez IE, Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J Mol Biol 325:367–376

37. Gianni S, Geierhaas CD, Calosci N, Jemth P, Vuister GW, Travaglini-Allocatelli C, Vendruscolo M, Brunori M (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci USA 104:128–133

38. Zarrine-Afsar A, Larson SM, Davidson AR (2005) The family feud: do proteins with similar structures fold via the same pathway? Curr Opin Struct Biol 15:42–49

39. Lindberg MO, Oliveberg M (2007) Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol 17:21–29

40. Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding? Trends Biochem Sci 28:18–25

41. Englander SW, Mayne L, Krishna MM (2007) Protein folding and misfolding: mechanism and principles. Q Rev Biophys 40:287–326

42. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155

43. Yang JS, Chen WW, Skolnick J, Shakhnovich EI (2007) All-atom ab initio folding of a diverse set of proteins. Structure 15:53–63

44. Cho JH, Raleigh DP (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure. J Mol Biol 359:1437–1446

45. Mirsky AE, Pauling L (1936) On the structure of native, denatured, and coagulated proteins. Proc Natl Acad Sci USA 22:439–447

46. Chen J, Stites WE (2001) Packing is a key selection factor in the evolution of protein hydrophobic cores. Biochemistry 40:15280–15289

47. Richards FM, Lim WA (1993) An analysis of packing in the protein folding problem. Q Rev Biophys 26:423–498

48. Marin M (2008) Folding at the rhythm of the rare codon beat. Biotechnol J 3:1047–1057

49. Sauer RT, Milla ME, Waldburger CD, Brown BM, Schildbach JF (1996) Sequence determinants of folding and stability for the P22 Arc repressor dimer. Faseb J 10:42–48

50. Barrow CJ, Yasuda A, Kenny PT, Zagorski MG (1992) Solution conformations and aggregational properties of synthetic amyloid beta-peptides of Alzheimer's disease. Analysis of circular dichroism spectra. J Mol Biol 225:1075–1093

51. Haber E, Anfinsen CB (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. J Biol Chem 236:422–424

52. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

53. Walsh CT (2006) Posttranslational modification of proteins. Roberts, Greenwood Village, CO

54. Sali A, Shakhnovich E, Karplus M (1994) How does a protein fold? Nature 369:248–251

55. Wilson DR, Finlay BB (1997) The 'Asx-Pro turn' as a local structural motif stabilized by alternative patterns of hydrogen bonds and a consensus-derived model of the sequence Asn-Pro-Asn. Protein Eng 10:519–529

56. Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA 37:205–211

57. Feyereisen C, Morcellet M, Loucheux C (1977) Preferential and absolute adsorption to poly[N5-(3-hydroxypropyl)-L-glutamine]

in water/2-chloroethanol solvent mixtures. Macromolecules 10:485–488

58. Schiffer M, Edmundson AB (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys J 7:121–135

59. Kotelchuck D, Scheraga HA (1969) The influence of short-range interactions on protein onformation. II. A model for predicting the alpha-helical regions of proteins. Proc Natl Acad Sci USA 62:14–21

60. Creighton TE (1993) Proteins: structures and molecular properties. Freeman, New York

61. Marqusee S, Robbins VH, Baldwin RL (1989) Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci USA 86:5286–5290

62. Chakrabartty A, Schellman JA, Baldwin RL (1991) Large differences in the helix propensities of alanine and glycine. Nature 351:586–588

63. Altmann KH, Wojcik J, Vasquez M, Scheraga HA (1990) Helix-coil stability constants for the naturally occurring amino acids in water. XXIII. Proline parameters from random poly (hydroxybutylglutamine-co-L-proline). Biopolymers 30:107–120

64. Shoemaker KR, Fairman R, Kim PS, York EJ, Stewart JM, Baldwin RL (1987) The C-peptide helix from ribonuclease A considered as an autonomous folding unit. Cold Spring Harb Symp Quant Biol 52:391–398

65. Kim PS, Baldwin RL (1984) A helix stop signal in the isolated S-peptide of ribonuclease A. Nature 307:329–334

66. Marqusee S, Baldwin RL (1987) Helix stabilization by Glu−⋯Lys + salt bridges in short peptides of de novo design. Proc Natl Acad Sci USA 84:8898–8902

67. Takahashi S, Kim EH, Hibino T, Ooi T (1989) Comparison of alpha-helix stability in peptides having a negatively or positively charged residue block attached either to the N- or C-terminus of an alpha-helix: the electrostatic contribution and anisotropic stability of the alpha-helix. Biopolymers 28:995–1009

68. Zimm BH, Brag JK (1959) Theory of phase transition between helix and random coil in polypeptide chains. J Chem Phys 31:526

69. Finkelstein AV, Badretdinov AY, Ptitsyn OB (1991) Physical reasons for secondary structure stability: alpha-helices in short peptides. Proteins 10:287–299

70. Munoz V, Serrano L (1994) Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol 1:399–409

71. Munoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J Mol Biol 245:275–296

72. Munoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. J Mol Biol 245:297–308

73. Smith C, Regan L (1997) Construction and design of β-sheets. Acc Chem Res 30:153–161

74. Lee J, Dubey VK, Longo LM, Blaber M (2008) A logical OR redundancy within the Asx-Pro-Asx-Gly type I beta-turn motif. J Mol Biol 377:1251–1264

75. Chou PY, Fasman GD (1978) Empirical predictions of protein conformation. Annu Rev Biochem 47:251–276

76. Williams RM, Obradovi Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker AK (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac Symp Biocomput 89–100

77. DeGrado W, Kezdy F, Kaiser E (1981) Design, synthesis and characterization of a cytotoxic peptide with melittin-like activity. J Am Chem Soc 103:679–681

78. Lupas AN, Gruber M (2005) The structure of alpha-helical coiled coils. Adv Protein Chem 70:37–78

79. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J (2007) The folding and evolution of multidomain proteins. Nat Rev Mol Cell Biol 8:319–330

80. MacArthur MW, Thornton JM (1991) Influence of proline residues on protein conformation. J Mol Biol 218:397–412

81. Senes A, Engel DE, DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr Opin Struct Biol 14:465–479

82. Chakrabarti P, Chakrabarti S (1998) C–H…O hydrogen bond involving proline residues in alpha-helices. J Mol Biol 284:867–873

83. Barlow DJ, Thornton JM (1988) Helix geometry in proteins. J Mol Biol 201:601–619

84. Kini RM (2002) Molecular moulds with multiple missions: functional sites in three-finger toxins. Clin Exp Pharmacol Physiol 29:815–822

85. Carlier E, Fajloun Z, Mansuelle P, Fathallah M, Mosbah A, Oughideni R, Sandoz G, Di Luccio E, Geib S, Regaya I, Brocard J, Rochat H, Darbon H, Devaux C, Sabatier JM, de Waard M (2001) Disulfide bridge reorganization induced by proline mutations in maurotoxin. FEBS Lett 489:202–207

86. Arias HR, Blanton MP (2000) Alpha-conotoxins. Int J Biochem Cell Biol 32:1017–1028

87. Balaji RA, Ohtake A, Sato K, Gopalakrishnakone P, Kini RM, Seow KT, Bay BH (2000) λ-Conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. Isolation and characterization from the venom of Conus marmoreus. J Biol Chem 275:39516–39522

88. Kang TS, Radic Z, Talley TT, Jois SD, Taylor P, Kini RM (2007) Protein folding determinants: structural features determining alternative disulfide pairing in alpha- and chi/lambda-conotoxins. Biochemistry 46:3338–3355

89. Kemper B (2004) Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. Toxicol Appl Pharmacol 199:305–315

90. Lorenzen S, Peters B, Goede A, Preissner R, Frommel C (2005) Conservation of cis prolyl bonds in proteins during evolution. Proteins 58:589–595

91. Jenko Kokalj S, Guncar G, Stern I, Morgan G, Rabzelj S, Kenig M, Staniforth RA, Waltho JP, Zerovnik E, Turk D (2007) Essential role of proline isomerization in stefin B tetramer formation. J Mol Biol 366:1569–1579

92. Martin A, Schmid FX (2003) A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. J Mol Biol 331:1131–1140

93. Eckert B, Martin A, Balbach J, Schmid FX (2005) Prolyl isomerization as a molecular timer in phage infection. Nat Struct Mol Biol 12:619–623

94. Kamen DE, Woody RW (2002) Identification of proline residues responsible for the slow folding kinetics in pectate lyase C by mutagenesis. Biochemistry 41:4724–4732

95. Lu KP, Finn G, Lee TH, Nicholson LK (2007) Prolyl cis-trans isomerization as a molecular timer. Nat Chem Biol 3:619–629

96. Takahashi N, Hayano T, Suzuki M (1989) Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. Nature 337:473–475

97. Harding MW, Galat A, Uehling DE, Schreiber SL (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. Nature 341:758–760

98. Kabsch W, Sander C (1984) On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. Proc Natl Acad Sci USA 81:1075–1078

99. Ambroggio XI, Kuhlman B (2006) Design of protein conformational switches. Curr Opin Struct Biol 16:525–530

100. Siligardi G, Drake AF (1995) The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. Biopolymers 37:281–292

101. Rath A, Davidson AR, Deber CM (2005) The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. Biopolymers 80:179–185

102. Waters ML (2004) Aromatic interactions in peptides: impact on structure and function. Biopolymers 76:435–445

103. Hughes RM, Waters ML (2006) Effects of lysine acetylation in a beta-hairpin peptide: comparison of an amide-pi and a cation-pi interaction. J Am Chem Soc 128:13586–13591

104. Rea AM, Simpson ER, Meldrum JK, Williams HE, Searle MS (2008) Aromatic residues engineered into the beta-turn nucleation site of ubiquitin lead to a complex folding landscape, non-native side-chain interactions, and kinetic traps. Biochemistry 47:12910–12922

105. Li S, Yang W, Maniccia AW, Barrow D Jr, Tjong H, Zhou HX, Yang JJ (2008) Rational design of a conformation-switchable Ca(2+)- and Tb(3+)-binding protein without the use of multiple coupled metal-binding sites. FEBS J 275:5048–5061

106. Uversky VN (2008) Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res 5:260–287

107. Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta 1698:131–153

108. Sevier CS, Kaiser CA (2002) Formation and transfer of disulphide bonds in living cells. Nat Rev Mol Cell Biol 3:836–847

109. Mamathambika BS, Bardwell JC (2008) Disulfide-linked protein folding pathways. Annu Rev Cell Dev Biol 24:211–235

110. Abkevich VI, Shakhnovich EI (2000) What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioninformatics analysis. J Mol Biol 300:975–985

111. Nakamura H (2005) Thioredoxin and its related molecules: update 2005. Antioxid Redox Signal 7:823–828

112. Schmidt B, Ho L, Hogg PJ (2006) Allosteric disulfide bonds. Biochemistry 45:7429–7433

113. Yokota A, Izutani K, Takai M, Kubo Y, Noda Y, Koumoto Y, Tachibana H, Segawa S (2000) The transition state in the folding-unfolding reaction of four species of three-disulfide variant of hen lysozyme: the role of each disulfide bridge. J Mol Biol 295:1275–1288

114. Creighton TE (1995) Protein folding. An unfolding story. Curr Biol 5:353–356

115. Narhi LO, Hua QX, Arakawa T, Fox GM, Tsai L, Rosenfeld R, Holst P, Miller JA, Weiss MA (1993) Role of native disulfide bonds in the structure and activity of insulin-like growth factor 1: genetic models of protein-folding intermediates. Biochemistry 32:5214–5221

116. Thornton JM (1981) Disulphide bridges in globular proteins. J Mol Biol 151:261–287

117. Guo ZY, Qiao ZS, Feng YM (2008) The in vitro oxidative folding of the insulin superfamily. Antioxid Redox Signal 10:127–139

118. Weiss MA, Hua QX, Jia W, Chu YC, Wang RY, Katsoyannis PG (2000) Hierarchical protein "un-design": insulin's intra-chain disulfide bridge tethers a recognition alpha-helix. Biochemistry 39:15429–15440

119. Chang SG, Choi KD, Jang SH, Shin HC (2003) Role of disulfide bonds in the structure and activity of human insulin. Mol Cells 16:323–330

120. Hua QX, Mayer JP, Jia W, Zhang J, Weiss MA (2006) The folding nucleus of the insulin superfamily: a flexible peptide model foreshadows the native state. J Biol Chem 281:28131–28142

121. Zhu Q, Liang S, Martin L, Gasparini S, Menez A, Vita C (2002) Role of disulfide bonds in folding and activity of leiurotoxin I: just two disulfides suffice. Biochemistry 41:11488–11494

122. Drakopoulou E, Vizzavona J, Neyton J, Aniort V, Bouet F, Virelizier H, Menez A, Vita C (1998) Consequence of the removal of evolutionary conserved disulfide bridges on the structure and function of charybdotoxin and evidence that particular cysteine spacings govern specific disulfide bond formation. Biochemistry 37:1292–1301

123. Hober S, Forsberg G, Palm G, Hartmanis M, Nilsson B (1992) Disulfide exchange folding of insulin-like growth factor I. Biochemistry 31:1749–1756

124. Huang QL, Zhao J, Tang YH, Shao SQ, Xu GJ, Feng YM (2007) The sequence determinant causing different folding behaviors of insulin and insulin-like growth factor-1. Biochemistry 46:218–224

125. White CE, Hunter MJ, Meininger DP, Garrod S, Komives EA (1996) The fifth epidermal growth factor-like domain of thrombomodulin does not have an epidermal growth factor-like disulfide bonding pattern. Proc Natl Acad Sci USA 93:10177–10182

126. Bulaj G, Olivera BM (2008) Folding of conotoxins: formation of the native disulfide bridges during chemical synthesis and biosynthesis of Conus peptides. Antioxid Redox Signal 10:141–155

127. Kang TS, Vivekanandan S, Jois SD, Kini RM (2005) Effect of C-terminal amidation on folding and disulfide-pairing of alpha-conotoxin ImI. Angew Chem Int Ed Engl 44:6333–6337

128. Zhang RM, Snyder GH (1991) Factors governing selective formation of specific disulfides in synthetic variants of alpha-conotoxin. Biochemistry 30:11343–11348

129. Clarke J, Fersht AR (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. Biochemistry 32:4322–4329

130. Maleknia SD, Reixach N, Buxbaum JN (2006) Oxidation inhibits amyloid fibril formation of transthyretin. FEBS J 273:5400–5406

131. Chugha P, Sage HJ, Oas TG (2006) Methionine oxidation of monomeric lambda repressor: the denatured state ensemble under nondenaturing conditions. Protein Sci 15:533–542

132. Gao J, Yin DH, Yao Y, Sun H, Qin Z, Schoneich C, Williams TD, Squier TC (1998) Loss of conformational stability in calmodulin upon methionine oxidation. Biophys J 74:1115–1134

133. Wood MJ, Becvar LA, Prieto JH, Melacini G, Komives EA (2003) NMR structures reveal how oxidation inactivates thrombomodulin. Biochemistry 42:11932–11942

134. Dado GP, Gellman SH (1993) Redox control of secondary structure in a designed peptide. J Am Chem Soc 115:12609–12610

135. Wong CH (2005) Protein glycosylation: new challenges and opportunities. J Org Chem 70:4219–4225

136. Helenius A, Aebi M (2004) Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem 73:1019–1049

137. Kornfeld S (1986) Trafficking of lysosomal enzymes in normal and disease states. J Clin Invest 77:1–6

138. Sola RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. J Pharm Sci 98:1223–1245

139. Sinclair AM, Elliott S (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci 94:1626–1635

140. Berg-Fussman A, Grace ME, Ioannou Y, Grabowski GA (1993) Human acid beta-glucosidase. N-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. J Biol Chem 268:14861–14866

141. Kobayashi T, Honke K, Jin T, Gasa S, Miyazaki T, Makita A (1992) Components and proteolytic processing sites of arylsulfatase B from human placenta. Biochim Biophys Acta 1159:243–247

142. Millat G, Froissart R, Maire I, Bozon D (1997) IDS transfer from overexpressing cells to IDS-deficient cells. Exp Cell Res 230:362–367

143. Scriver CR (2001) The metabolic & molecular bases of inherited disease. McGraw-Hill, New York

144. Zhao KW, Faull KF, Kakkis ED, Neufeld EF (1997) Carbohydrate structures of recombinant human alpha-L-iduronidase secreted by Chinese hamster ovary cells. J Biol Chem 272:22758–22765

145. Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proc Natl Acad Sci USA 105:8256–8261

146. Wyss DF, Wagner G (1996) The structural role of sugars in glycoproteins. Curr Opin Biotechnol 7:409–416

147. Wormald MR, Dwek RA (1999) Glycoproteins: glycan presentation and protein-fold stability. Structure 7:R155–R160

148. Walsh MT, Watzlawick H, Putnam FW, Schmid K, Brossmer R (1990) Effect of the carbohydrate moiety on the secondary structure of beta 2-glycoprotein. I. Implications for the biosynthesis and folding of glycoproteins. Biochemistry 29:6250–6257

149. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam AR, Smolyar A, Reinherz EL, Wagner G (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. Science 269:1273–1278

150. O'Connor SE, Imperiali B (1996) Modulation of protein structure and function by asparagine-linked glycosylation. Chem Biol 3:803–812

151. Rickert KW, Imperiali B (1995) Analysis of the conserved glycosylation site in the nicotinic acetylcholine receptor: potential roles in complex assembly. Chem Biol 2:751–759

152. Shaanan B, Lis H, Sharon N (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. Science 254:862–866

153. Helenius A (1994) How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. Mol Biol Cell 5:253–265

154. Hickman S, Kornfeld S (1978) Effect of tunicamycin on IgM, IgA, and IgG secretion by mouse plasmacytoma cells. J Immunol 121:990–996

155. Landolfi NF, Rich RR, Cook RG (1985) Differential glycosylation requirements for the cell surface expression of class I molecules. J Immunol 134:423–430

156. Shackelford DA, Strominger JL (1983) Analysis of the oligosaccharides on the HLA-DR and DC1 B cell antigens. J Immunol 130:274–282

157. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A (1997) The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. J Cell Biol 139:613–623

158. Hermans MM, Wisselaar HA, Kroos MA, Oostra BA, Reuser AJ (1993) Human lysosomal alpha-glucosidase: functional characterization of the glycosylation sites. Biochem J 289:681–686

159. Ioannou YA, Zeidner KM, Grace ME, Desnick RJ (1998) Human alpha-galactosidase A: glycosylation site 3 is essential for enzyme solubility. Biochem J 332:789–797

160. Kotch FW, Guzei IA, Raines RT (2008) Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues. J Am Chem Soc 130:2952–2953

161. Raines RT (2006) 2005 Emil Thomas Kaiser Award. Protein Sci 15:1219–1225

162. Brodsky B, Ramshaw JA (1997) The collagen triple-helix structure. Matrix Biol 15:545–554

163. Buczek O, Bulaj G, Olivera BM (2005) Conotoxins and the posttranslational modification of secreted gene products. Cell Mol Life Sci 62:3067–3079

164. Lopez-Vera E, Walewska A, Skalicky JJ, Olivera BM, Bulaj G (2008) Role of hydroxyprolines in the in vitro oxidative folding and biological activity of conotoxins. Biochemistry 47:1741–1751

165. Kubo S, Chino N, Kimura T, Sakakibara S (1996) Oxidative folding of omega-conotoxin MVIIC: effects of temperature and salt. Biopolymers 38:733–744

166. Koivunen P, Salo KE, Myllyharju J, Ruddock LW (2005) Three binding sites in protein-disulfide isomerase cooperate in collagen prolyl 4-hydroxylase tetramer assembly. J Biol Chem 280:5227–5235

167. Martinez A, Treston AM (1996) Where does amidation take place? Mol Cell Endocrinol 123:113–117

168. Merkler DJ (1994) C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. Enzyme Microb Technol 16:450–456

169. Sforca ML, Oyama S Jr, Canduri F, Lorenzi CC, Pertinhez TA, Konno K, Souza BM, Palma MS, Ruggiero Neto J, Azevedo WF Jr, Spisni A (2004) How C-terminal carboxyamidation alters the biological activity of peptides from the venom of the eumenine solitary wasp. Biochemistry 43:5608–5617

170. Yasin B, Lehrer RI, Harwig SS, Wagar EA (1996) Protegrins: structural requirements for inactivating elementary bodies of Chlamydia trachomatis. Infect Immun 64:4863–4866

171. Katayama H, Ohira T, Aida K, Nagasawa H (2002) Significance of a carboxyl-terminal amide moiety in the folding and biological activity of crustacean hyperglycemic hormone. Peptides 23:1537–1546

172. Mor A, Nicolas P (1994) The NH2-terminal alpha-helical domain 1–18 of dermaseptin is responsible for antimicrobial activity. J Biol Chem 269:1934–1939

173. Sandvik AK, Dockray GJ (1999) Biological activity of carboxy-terminal gastrin analogs. Eur J Pharmacol 364:199–203

174. Ali MF, Soto A, Knoop FC, Conlon JM (2001) Antimicrobial peptides isolated from skin secretions of the diploid frog, Xenopus tropicalis (Pipidae). Biochim Biophys Acta 1550:81–89

175. Meredith J, Ring M, Macins A, Marschall J, Cheng NN, Theilmann D, Brock HW, Phillips JE (1996) Locust ion transport peptide (ITP): primary structure, cDNA and expression in a baculovirus system. J Exp Biol 199:1053–1061

176. Fuller E, Green BR, Catlin P, Buczek O, Nielsen JS, Olivera BM, Bulaj G (2005) Oxidative folding of conotoxins sharing an identical disulfide bridging framework. FEBS J 272:1727–1738

177. Shearer MJ (1992) Vitamin K metabolism and nutriture. Blood Rev 6:92–104

178. Hauschka PV, Lian JB, Gallop PM (1975) Direct identification of the calcium-binding amino acid, gamma-carboxyglutamate, in mineralized tissue. Proc Natl Acad Sci USA 72:3925–3929

179. Howard JB, Nelsestuen GL (1975) Isolation and characterization of vitamin K-dependent region of bovine blood clotting factor X. Proc Natl Acad Sci USA 72:1281–1285

180. Suttie JW (1985) Vitamin K-dependent carboxylase. Annu Rev Biochem 54:459–477

181. McIntosh JM, Olivera BM, Cruz LJ, Gray WR (1984) Gamma-carboxyglutamate in a neuroactive toxin. J Biol Chem 259:14343–14346

182. Rigby AC, Lucas-Meunier E, Kalume DE, Czerwiec E, Hambe B, Dahlqvist I, Fossier P, Baux G, Roepstorff P, Baleja JD, Furie BC, Furie B, Stenflo J (1999) A conotoxin from Conus textile with unusual posttranslational modifications reduces presynaptic Ca2+ influx. Proc Natl Acad Sci USA 96:5758–5763

183. Bulaj G, Buczek O, Goodsell I, Jimenez EC, Kranski J, Nielsen JS, Garrett JE, Olivera BM (2003) Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. Proc Natl Acad Sci USA 100(Suppl 2):14562–14568

184. Gowd KH, Twede V, Watkins M, Krishnan KS, Teichert RW, Bulaj G, Olivera BM (2008) Conantokin-P, an unusual conantokin with a long disulfide loop. Toxicon 52:203–213

185. Chen Z, Blandl T, Prorok M, Warder SE, Li L, Zhu Y, Pedersen LG, Ni F, Castellino FJ (1998) Conformational changes in conantokin-G induced upon binding of calcium and magnesium as revealed by NMR structural analysis. J Biol Chem 273:16248–16258

186. Latzer J, Shen T, Wolynes PG (2008) Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. Biochemistry 47:2110–2122

187. Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101:2209–2242

188. Jacobs SA, Khorasanizadeh S (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science 295:2080–2083

189. Tatko CD, Waters ML (2004) Comparison of C-H…pi and hydrophobic interactions in a beta-hairpin peptide: impact on stability and specificity. J Am Chem Soc 126:2028–2034

190. Hughes RM, Waters ML (2005) Influence of N-methylation on a cation-pi interaction produces a remarkably stable beta-hairpin peptide. J Am Chem Soc 127:6518–6519

191. Hughes RM, Waters ML (2006) Arginine methylation in a beta-hairpin peptide: implications for Arg-pi interactions, DeltaCp(o), and the cold denatured state. J Am Chem Soc 128:12735–12742

192. Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. EMBO J 16:2054–2060

193. Reynolds MR, Berry RW, Binder LI (2007) Nitration in neurodegeneration: deciphering the "Hows" "nYs". Biochemistry 46:7325–7336

194. Pylypenko O, Schonichen A, Ludwig D, Ungermann C, Goody RS, Rak A, Geyer M (2008) Farnesylation of the SNARE protein Ykt6 increases its stability and helical folding. J Mol Biol 377:1334–1345

195. Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions. Science 247:1306–1310

196. Plaxco KW, Riddle DS, Grantcharova V, Baker D (1998) Simplified proteins: minimalist solutions to the 'protein folding problem'. Curr Opin Struct Biol 8:80–85

197. Miller JH (1979) Genetic studies of the lac repressor. XI. On aspects of lac repressor structure suggested by genetic experiments. J Mol Biol 131:249–258

198. Miller JH, Coulondre C, Hofer M, Schmeissner U, Sommer H, Schmitz A, Lu P (1979) Genetic studies of the lac repressor. IX. Generation of altered proteins by the suppression of nonsense mutations. J Mol Biol 131:191–222

199. Miller JH, Schmeissner U (1979) Genetic studies of the lac repressor. X. Analysis of missense mutations in the lacI gene. J Mol Biol 131:223–248

200. Brown BM, Sauer RT (1999) Tolerance of Arc repressor to multiple-alanine substitutions. Proc Natl Acad Sci USA 96:1983–1988

201. Islam MM, Sohya S, Noguchi K, Yohda M, Kuroda Y (2008) Crystal structure of an extensively simplified variant of bovine pancreatic trypsin inhibitor in which over one-third of the residues are alanines. Proc Natl Acad Sci USA 105:15334–15339

202. Kuroda Y, Kim PS (2000) Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. J Mol Biol 298:493–501

203. Munson M, O'Brien R, Sturtevant JM, Regan L (1994) Redesigning the hydrophobic core of a four-helix-bundle protein. Protein Sci 3:2015–2022

204. Baldwin RL (1986) Temperature dependence of the hydrophobic interaction in protein folding. Proc Natl Acad Sci USA 83:8069–8072

205. Bashford D, Chothia C, Lesk AM (1987) Determinants of a protein fold. Unique features of the globin amino acid sequences. J Mol Biol 196:199–216

206. Chothia C, Levitt M, Richardson D (1981) Helix to helix packing in proteins. J Mol Biol 145:215–250

207. Kellis JT Jr, Nyberg K, Sali D, Fersht AR (1988) Contribution of hydrophobic interactions to protein stability. Nature 333:784–786

208. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183

209. Gassner NC, Baase WA, Matthews BW (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci USA 93:12155–12158

210. Lim WA, Sauer RT (1991) The role of internal packing interactions in determining the structure and stability of a protein. J Mol Biol 219:359–376

211. Baldwin EP, Hajiseyedjavadi O, Baase WA, Matthews BW (1993) The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. Science 262:1715–1718

212. Pielak GJ, Auld DS, Beasley JR, Betz SF, Cohen DS, Doyle DF, Finger SA, Fredericks ZL, Hilgen-Willis S, Saunders AJ (1995) Protein thermal denaturation, side-chain models, and evolution: amino acid substitutions at a conserved helix–helix interface. Biochemistry 34:3268–3276

213. Axe DD, Foster NW, Fersht AR (1996) Active barnase variants with completely random hydrophobic cores. Proc Natl Acad Sci USA 93:5590–5594

214. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685

215. Moffet DA, Hecht MH (2001) De novo proteins from combinatorial libraries. Chem Rev 101:3191–3203

216. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262:1401–1407

217. Schafmeister CE, LaPorte SL, Miercke LJ, Stroud RM (1997) A designed four helix bundle protein with native-like structure. Nat Struct Biol 4:1039–1046

218. Riddle DS, Santiago JV, Bray-Hall ST, Doshi N, Grantcharova VP, Yi Q, Baker D (1997) Functional rapidly folding proteins from simplified amino acid sequences. Nat Struct Biol 4:805–809

219. Taylor SV, Walter KU, Kast P, Hilvert D (2001) Searching sequence space for protein catalysts. Proc Natl Acad Sci USA 98:10596–10601

220. Walter KU, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet. J Biol Chem 280:37742–37746

221. Silverman JA, Balakrishnan R, Harbury PB (2001) Reverse engineering the (beta/alpha)8 barrel fold. Proc Natl Acad Sci USA 98:3092–3097

222. Hemmingsen JM, Gernert KM, Richardson JS, Richardson DC (1994) The tyrosine corner: a feature of most Greek key beta-barrel proteins. Protein Sci 3:1927–1937

223. Yao J, Dyson HJ, Wright PE (1994) Three-dimensional structure of a type VI turn in a linear peptide in water solution. Evidence

for stacking of aromatic rings as a major stabilizing factor. J Mol Biol 243:754–766

224. Moore DT, Berger BW, DeGrado WF (2008) Protein–protein interactions in the membrane: sequence, structural, and biological motifs. Structure 16:991–1001

225. Gsponer J, Vendruscolo M (2006) Theoretical approaches to protein aggregation. Protein Pept Lett 13:287–293

226. Carrell RW, Lomas DA (1997) Conformational disease. Lancet 350:134–138

227. Conn PM, Ulloa-Aguirre A, Ito J, Janovick JA (2007) G protein-coupled receptor trafficking in health and disease: lessons learned to prepare for therapeutic mutant rescue in vivo. Pharmacol Rev 59:225–250

228. Brady RO (2006) Enzyme replacement for lysosomal diseases. Annu Rev Med 57:283–296

229. Santucci R, Sinibaldi F, Fiorucci L (2008) Protein folding, unfolding and misfolding: role played by intermediate States. Mini Rev Med Chem 8:57–62

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

231. Kortemme T, Ramirez-Alvarado M, Serrano L (1998) Design of a 20-amino acid, three-stranded beta-sheet protein. Science 281:253–256

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

233. Hecht MH, Richardson JS, Richardson DC, Ogden RC (1990) De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. Science 249:884–891

234. Magliery TJ, Regan L (2004) Combinatorial approaches to protein stability and structure. Eur J Biochem 271:1595–1608

235. Baltzer L, Nilsson H, Nilsson J (2001) De novo design of proteins—what are the rules? Chem Rev 101:3153–3163

236. Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A (2007) Critical assessment of methods of protein structure prediction-Round VII. Proteins 69(Suppl 8):3–9

237. Honig B (1999) Protein folding: from the levinthal paradox to structure prediction. J Mol Biol 293:283–293

238. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 24:1241–1252

239. Bradbury AF, Smyth DG (1991) Peptide amidation. Trends Biochem Sci 16:112–115

240. Dong J, Ye P, Schade AJ, Gao S, Romo GM, Turner NT, McIntire LV, Lopez JA (2001) Tyrosine sulfation of glycoprotein I(b)alpha. Role of electrostatic interactions in von Willebrand factor binding. J Biol Chem 276:16690–16694

241. Tarcsa E, Marekov LN, Mei G, Melino G, Lee SC, Steinert PM (1996) Protein unfolding by peptidylarginine deiminase. Substrate specificity and structural relationships of the natural substrates trichohyalin and filaggrin. J Biol Chem 271:30709–30716

242. Janke C, Rogowski K, van Dijk J (2008) Polyglutamylation: a fine-regulator of protein function? 'Protein modifications: beyond the usual suspects' review series. EMBO Rep 9:636–641

243. MacRae TH (1997) Tubulin post-translational modifications—enzymes and their mechanisms of action. Eur J Biochem 244:265–278