Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2020, **11**, 9480

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 1st June 2020
Accepted 30th July 2020

DOI: 10.1039/d0sc03060k

rsc.li/chemical-science

Increasing protein stability by engineering the n → π* interaction at the β-turn†

Bhavesh Khatri, [a] Puja Majumder, [a] Jayashree Nagesh, [b] Aravind Penmatsa *[a] and Jayanta Chatterjee *[a]

Abundant n → π* interactions between adjacent backbone carbonyl groups, identified by statistical analysis of protein structures, are predicted to play an important role in dictating the structure of proteins. However, experimentally testing the prediction in proteins has been challenging due to the weak nature of this interaction. By amplifying the strength of the n → π* interaction *via* amino acid substitution and thioamide incorporation at a solvent exposed β-turn within the **GB1** proteins and Pin 1 WW domain, we demonstrate that an n → π* interaction increases the structural stability of proteins by restricting the ϕ torsion angle. Our results also suggest that amino acid side-chain identity and its rotameric conformation play an important and decisive role in dictating the strength of an n → π* interaction.

Introduction

An array of noncovalent interactions including electrostatic forces, hydrogen bonds, van der Waals interactions and hydrophobic effects in a polypeptide chain dictate its three-dimensional structure and govern its folding.[1] In particular, owing to their high abundance, the noncovalent interactions originating from the backbone (main chain) atoms of a polypeptide chain,[2] including the classical hydrogen bonds,[3] C–H···O hydrogen bonds,[4] C5 hydrogen bonds[5] and n → π* interactions,[6] play a crucial role in stabilizing protein structures. The n → π* interaction originates from the donation of the lone pair (n) electron density of the carbonyl oxygen $(O)_i$ into the empty π* orbital of the adjacent carbonyl group $(C{=}O)_{i+1}$.[7–9] The distance ($d \leq 3.2$ Å) and angular criteria ($\theta = 109 \pm 10°$) defining an n → π* interaction are in agreement with the Bürgi–Dunitz trajectory for nucleophilic attack,[10] which along with the associated directionality $i \rightarrow i+1$ (N-term → C-term) (Fig. 1) is indicative of its possible role in folding and stabilization of protein secondary structures.[11–16]

Contribution of the n → π* interaction towards the stability of the protein structure was initially reported in collagen mimetics.[17] The enhanced thermostability of a collagen mimetic with the 4*R*-configured proline derivative compared to that with the 4*S*-configured proline derivative was attributed to the stronger n → π* interaction in the 4*R*-configured proline derivative with the *exo*-pucker of the pyrrolidine ring.[18] The finding was exquisitely substantiated later by the high-resolution crystal structure of the oligoproline PPII helix, where the n → π* interaction was favored by the C^{γ}-*exo* pucker and disfavored by the C^{γ}-*endo* pucker of the pyrrolidine ring.[19] Furthermore, the stability of this PPII helix in the absence of intramolecular hydrogen bonds and hydration emphasizes the role of the n → π* interaction in the structural stability of collagen.

For an idealized geometry, the n → π* interaction between amides contributes ~0.3 kcal mol^{-1},[20] which may seem moderate. However, given the ubiquity of carbonyl groups in a polypeptide chain, n → π* interactions could have a significant collective contribution towards the overall energetics of

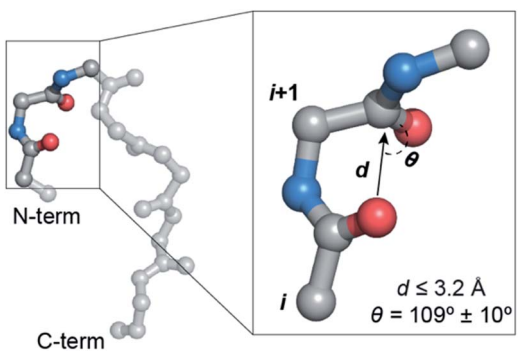


Fig. 1 A protein backbone depicting a $C{=}O_i \rightarrow C{=}O_{i+1}$ n → π* interaction with the distance (d) and angular (θ) criteria used in the crystallographic analyses.

[a]*Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India. E-mail: penmatsa@iisc.ac.in; jayanta@iisc.ac.in*

[b]*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore, India*

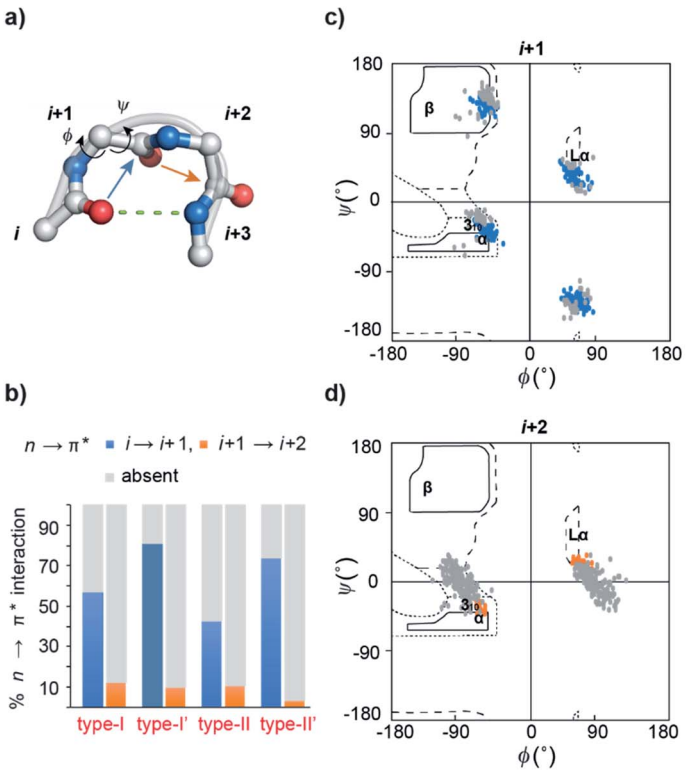
† Electronic supplementary information (ESI) available. The coordinates for the GB1 crystal structures have been deposited in the Protein Data Bank with the following accession codes 6L9B, 6L9D, 6LJI, 6L91. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/d0sc03060k

protein stability.[2] The distribution of the n → π* interaction obtained from analyses of protein crystal structures reveals that >70% of residues in α-helices, as opposed to <5% of residues in β-sheets engage in this interaction.[11,12] Furthermore, since one third of all amino acids in the random coil have torsion angles in the α-helical region,[21] the n → π* interaction might have an important role in restricting the conformational ensemble of unfolded proteins.[22] In this context, it is worth noting that random coils and turn regions of proteins show a high abundance of reciprocal n → π* interactions (back and forth donation between adjacent carbonyl pairs).[9]

The evidence of the n → π* interaction has been shown by microwave and IR spectroscopy in various small molecular systems.[23–27] However, despite enormous excitement in this area, so far experimental measurements of the energy of an n → π* interaction in proteins and its practical consequence on protein structural stability have been lacking. Therefore, we sought to engineer an n → π* interaction at the β-turn within a protein to understand its influence on the protein structure and its stability.

β-Turns (Fig. 2a) are the third most important protein secondary structure representing ~20% of all protein residues[28] having an important role in protein folding.[29–31] Furthermore, substituting non-proline residues with proline residues in the β-turn leads to increased stabilization of the turn[32] and enhanced protein stability.[33,34] The increased stability results from the decreased backbone conformational entropy of the denatured state due to the restricted rotation of the N–C^{α} bond, also known as the ϕ torsion angle. Since an n → π* interaction also restricts the ϕ angle of an amino acid residue,[11] we speculated that engineering an n → π* interaction at the β-turn would have direct consequence on the protein stability.

Here, by using bioinformatic analysis of the β-turn in proteins, we find an interplay between the conformational flexibility of the peptide backbone and the abundance of n → π* interactions at the two central residues, $i + 1$ and $i + 2$ of the β-turn. Through subsequent X-ray crystallography and computational analysis of synthetic **GB1** proteins with amino acid substitutions at the $i + 2$ residue of the β-turn, we show that amino acid side-chain identity and its rotameric conformation have a direct influence on the strength of an n → π* interaction. Gratifyingly, the thermal denaturation of the **GB1** proteins shows a good correlation between their stability and the strength of an n → π* interaction at the β-turn. Finally, we validate this observation in the Pin 1 WW domain, wherein by amplifying the strength of an n → π* interaction at the β-turn by thioamide incorporation, we could increase the thermal stability of the thioamidated Pin 1 WW domain.



e)

i+1	φ(°)	Δφ(°)	ψ(°)	Δψ(°)
βI	−58.6 ± 7.6 (−68.6 ± 8.7)	10.0	−32.2 ± 7.4 (−20.3 ± 11.9)	11.9
βII	−58.7 ± 6.6 (−58.0 ± 10.0)	0.7	128.2 ± 6.4 (136.0 ± 11.3)	7.8
βI'	54.1 ± 7.6 (58.1 ± 9.2)	4.0	38.5 ± 7.4 (38.8 ± 16.6)	0.3
βII'	59.8 ± 7.9 (58.1 ± 9.7)	1.7	−126.9 ± 6.5 (−128.5 ± 11.6)	1.6

f)

i+2	φ(°)	Δφ(°)	ψ(°)	Δψ(°)
βI	−67.5 ± 6.4 (−92.5 ± 12.5)	25.0	−30.4 ± 5.9 (0.4 ± 15.0)	30.8
βII	63.0 ± 6.4 (84.3 ± 14.6)	21.3	29.2 ± 4.9 (−0.5 ± 16.5)	29.7
βI'	64.9 ± 7.6 (79.8 ± 11.5)	14.9	28.5 ± 3.7 (4.6 ± 11.6)	23.9
βII'	−68.8 ± 2.2 (−92.4 ± 12.2)	23.6	−26.7 ± 5.5 (4.6 ± 13.9)	31.3

Fig. 2 (a) A β-turn (type-II') depicting two central residues $i + 1$ and $i + 2$ along with the C=O_i···HN_{i+3} hydrogen bond, C=O_i → C=O_{i+1} (blue arrow) and C=O_{i+1} → C=O_{i+2} (orange arrow) n → π* interactions. (b) The abundance of C=O_i → C=O_{i+1} and C=O_{i+1} → C=O_{i+2} n → π* interactions in the different turns derived from PDB analyses. Ramachandran plot of (c) $i + 1$ residues engaged in the C=O_i → C=O_{i+1} n → π* interaction, and (d) $i + 2$ residues engaged in the C=O_{i+1} → C=O_{i+2} n → π* interaction in the β-turns. The gray dots indicate the respective turn residues, which do not engage in the n → π* interaction. Mean torsion angle ± S.D. of (e) $i + 1$ and (f) $i + 2$ residues in the presence and (absence) of the n → π* interaction. $\Delta\phi$ and $\Delta\psi$ represent the difference between the mean torsion angles in the presence and absence of the n → π* interaction.

Results and discussion

n → π* interaction and conformational flexibility at the β-turn

Previous computational analyses predicted that n → π* interactions confer conformational stability to the *i* + 1 residue in common type I and type II β-turns, and thus have a special role to play in the stability of turns.[11] Therefore, we sought to examine the abundance of n → π* interactions and their possible correlation with the conformational flexibility of the peptide backbone in the β-turns.

By analyzing a non-redundant subset of high-resolution (≤2.0 Å) protein crystal structures in the Protein Data Bank (PDB), we curated 500 β-turns (identified using Promotif) representing the common type-I, type-II, type-I′, and type-II′ turns. Next, using the distance and angular criteria defining an n → π* interaction (Fig. 1), we determined the abundance of n → π* interactions at the *i* + 1 and *i* + 2 residues in the β-turns. We noted that 40–80% of the residues engage in a C=O$_i$ → C=O$_{i+1}$ n → π* interaction, whereas, only 3–12% of the residues are involved in the C=O$_{i+1}$ → C=O$_{i+2}$ n → π* interaction (Fig. 2b).

To identify the underlying cause of this behavior, we determined the torsion angles ϕ and ψ of *i* + 1 and *i* + 2 residues in all the β-turns and plotted them on the Ramachandran map. It was interesting to note the broader distribution of ϕ and ψ angles at the *i* + 2 residue (Fig. 2d) as opposed to the *i* + 1 residue (Fig. 2c). This is suggestive of restricted conformational freedom at the *i* + 1 residue, which is associated with the higher abundance of n → π* interactions at this site. We also calculated the difference in mean ϕ and ψ angles ($\Delta\phi$ and $\Delta\psi$) in the presence and absence of the n → π* interaction in the respective β-turns (Fig. 2e and f). The differences were significantly higher at the *i* + 2 residue in comparison to the *i* + 1 residue. This further indicates that the lower abundance of n → π* interactions at the *i* + 2 residue is associated with greater conformational flexibility of the peptide backbone.

Influence of the amino acid side-chain on the n → π* interaction

As β-turns are stabilized by the intramolecular hydrogen bond between C=O$_i$···HN$_{i+3}$ (Fig. 2a), the higher abundance of n → π* interactions at the *i* + 1 residue is perhaps linked with the conformational restriction of C=O$_i$ *via* hydrogen bonding. Thus, we surmised that the C=O$_i$ → C=O$_{i+1}$ n → π* interaction and the conformational space at the *i* + 1 position might be insensitive to amino acid substitution. Instead, the relatively flexible *i* + 2 residue of the β-turn (Fig. 2a), where neither the donor (n) C=O$_{i+1}$ nor the acceptor (π*) C=O$_{i+2}$ is constrained by the intramolecular hydrogen bond, is an ideal site to probe the role of the n → π* interaction in the protein structure and its stability. Additionally, the solvent exposure of β-turns allows for amino acid substitution and examining the influence of the amino acid side-chain on the n → π* interaction. Thus, we chose to engineer the loop L1 of the 56-residue immunoglobulin-binding domain B1 of the streptococcal

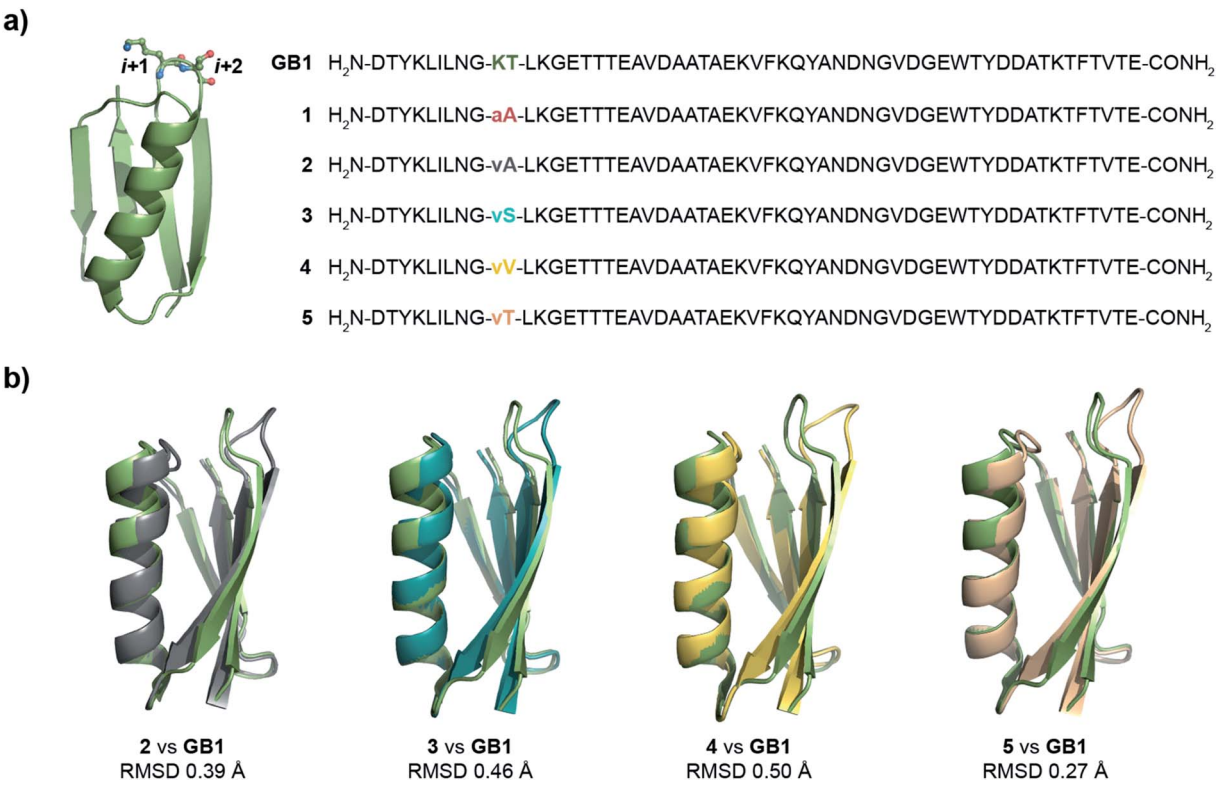


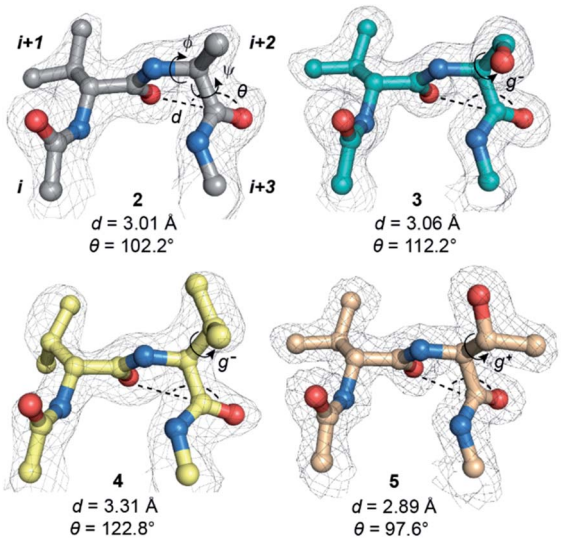
Fig. 3 (a) Crystal structure of **GB1** with –Lys–Thr– in loop L1 forming a type-I β-turn, which has been modified to a type-II′ β-turn in **1**–**5**. The single letter code in lower case indicates D-amino acid. (b) The backbone overlay of **2** (6L9B), **3** (6L9D), **4** (6LJI), and **5** (6L91) with **GB1** (2QMT).

protein G (**GB1**).[35–38] The solvent exposed loop L1 in wild type **GB1** is a type-I β-turn (Fig. 3a) with lysine at the $i + 1$ position and threonine at the $i + 2$ position that lacks a C═O_{i+1} → C═O_{i+2} n → π* interaction. However, to have a better control over the turn conformation, we decided to introduce a type-II′ β-turn.[39] Thus, we synthesized a **GB1** variant where –KT– in loop L1 was substituted with D-Ala–L-Ala (**1**) to induce a type-II′ β-turn (Fig. 3a).

Alanine was chosen due to its preference for an n → π* interaction[11] and high helix propensity[40] (preference of an amino acid to be in α-helices). Although thermal denaturation of **1** using variable temperature circular dichroism (CD) showed unfolding cooperativity similarly to **GB1** (Fig. S8 and S9†), multiple attempts to crystallize **1** remained unsuccessful. This is possibly a consequence of conformational flexibility introduced in the loop L1 by alanine substitution. Our earlier results indicated that the β-branched amino acid D-Val at the $i + 1$ site stabilizes a type-II′ β-turn more than D-Ala.[41] Thus, we synthesized **2**, with D-Val–L-Ala in loop L1 (Fig. 3a), which readily crystallized and X-ray diffraction data were collected to a maximum resolution of 1.9 Å. The structure of **2** overlaps closely with the tertiary structure of **GB1** (backbone RMSD 0.39 Å) (Fig. 3b), although with a significant displacement of loop L1. Gratifyingly, the D-Val C═O_{i+1} and L-Ala C═O_{i+2} in the type-II′ β-turn engage in an n → π* interaction, where the torsion angles of L-Ala at the $i + 2$ site (ϕ, $\psi = -59.7°, -42.2°$) are remarkably close to the mean torsion angles of a right-handed α-helix (ϕ, $\psi = -62°, -41°$)[42] (Fig. 4).

Encouraged by this result, we next incorporated serine with moderate helix propensity (**3**), valine (**4**) and threonine (**5**) with low helix propensity[40] at the $i + 2$ site of the type-II′ β-turn (Fig. 3a). With the decreasing helix propensity of amino acids in the order Ala > Ser > Val, we noted an increase in both d and θ between C═O_{i+1} → C═O_{i+2}, suggesting a gradual weakening of the n → π* interaction at the $i + 2$ residue (Fig. 4). Thus, to obtain a quantitative estimate of the n → π* interaction energy ($E_{n\rightarrow\pi^*}$) at the $i + 2$ residue in **2**, **3**, and **4**, we resorted to NBO analysis,[43] which clearly indicated a decreasing $E_{n\rightarrow\pi^*}$ in the order **2** > **3** > **4** (Fig. 4).

Despite the low helix propensity of threonine, we were surprised to note the shortest d and θ between C═O_{i+1}···C═O_{i+2} at the type-II′ β-turn in **5** with an $E_{n\rightarrow\pi^*}$ of 0.46 kcal mol^{-1} (Fig. 4). An overlay of the type-II′ β-turns of both the β-branched amino acids valine (**4**) and threonine (**5**) revealed a clear difference in the side-chain rotamer conformation (Fig. 5a). Valine in **4** crystallized in a *gauche*$^-$ (g^-) side-chain rotameric conformation, whereas threonine in **5** crystallized in a *gauche*$^+$ (g^+) conformation. From the statistical analyses of protein structures, it is known that β-branched amino acids favor the g^+ side-chain conformation over g^- in helices.[44–47] Thus, Thr25 in the α-helix of **GB1** with a g^+ conformation engages in an n → π* interaction, whereas Thr11 at the $i + 2$ residue in loop L1 with a g^- conformation lacks the n → π* interaction (Fig. 5b). Moreover, our dataset revealed that Thr with a g^- conformation at the $i + 2$ residue in the β-turns does not engage in an n → π* interaction (Table S4†). Therefore, despite the low helix propensity, the n → π* interaction in threonine at the type-II′ β-turn in **5** is a result of the altered side-chain rotamer conformation.



	i+1, i+2	ϕ_{i+2}, ψ_{i+2} (°)	$E_{n\rightarrow\pi^*}$	T_M (°C)
2	-D-Val-L-Ala-	−59.7, −42.2	0.25	72.2 ± 0.15
3	-D-Val-L-Ser-	−66.3, −28.9	0.20	70.5 ± 0.05
4	-D-Val-L-Val-	−85.1, −12.4	0.0	68.9 ± 0.01
5	-D-Val-L-Thr-	−58.8, −46.6	0.46	73.4 ± 0.05
	type-II′ β-turn	−80.0, 0.0		

Fig. 4 n → π* interaction at the $i + 2$ residue of the type-II′ β-turn in **GB1** variants characterized by the d and θ between C═O_{i+1}···C═O_{i+2}. The electron density map is contoured at 1.0σ. The $i + 2$ side-chain rotamer conformation is depicted in **3**, **4**, and **5**. Note the alteration in torsion angles of the $i + 2$ residue and its $E_{n\rightarrow\pi^*}$ (kcal mol^{-1}) derived from NBO analysis. The torsion angles of an ideal type-II′ β-turn are given for comparison. Also note the absence of the C═O_i···HN_{i+3} intramolecular hydrogen bond. The midpoint of the thermal transition (T_M ± S.D.) of the proteins was determined by variable temperature CD.

The Ramachandran plot of the $i + 2$ residue (Fig. 5c) in the type-II′ β-turn of **2**, **3**, **4**, and **5** revealed that, as the strength of the n → π* interaction increases, the torsion angles of an amino acid in a non-helical region in the absence of the stabilizing intramolecular hydrogen bond are gradually altered to occupy the right-handed α-helical region. Thus, proline with a high propensity to engage in an n → π* interaction[11] is a strong helix initiator.[48,49] Hence, our result further supports the crucial role of the n → π* interaction in helix nucleation, as hypothesized earlier.[11]

Implication of the n → π* interaction on protein stability

Since, an n → π* interaction results in a restricted ϕ torsion angle (Fig. 2e and f),[11] we sought to examine the influence of the n → π* interaction on the conformational stability of **2**, **3**, **4**, and **5**. The midpoint of the thermal transition (T_M), which is a measure of structural stability was determined by variable temperature CD (Fig. S10–S13†). **5** with the strongest C═O_{i+1} → C═O_{i+2} n → π* interaction displayed the maximum stability (T_M) and **4** with no detectable C═O_{i+1} → C═O_{i+2} n → π* interaction showed the least stability (Fig. 4). We were

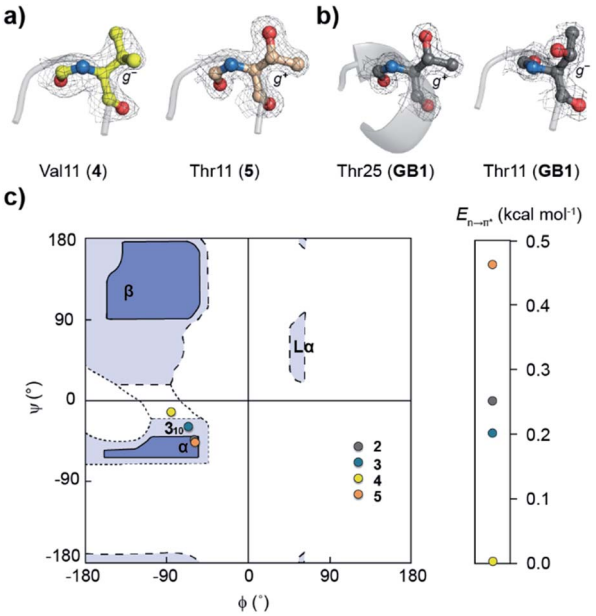


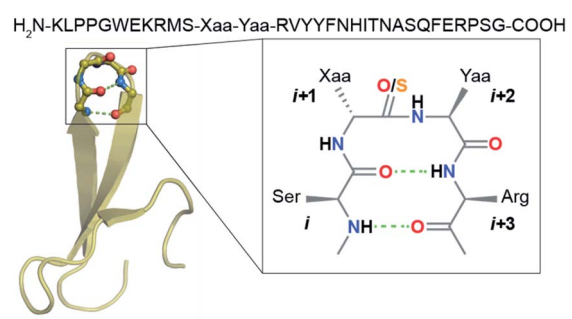
Fig. 5 (a) Side-chain rotamer conformation of the i + 2 residue in the type-II′ β-turn of **4** and **5**. (b) The side-chain rotamer conformation of Thr25 (α-helix) and Thr11 (loop L1) in **GB1**, showing the presence and absence of the n → π* interaction, respectively. The electron density map is contoured at 1.0σ. (c) Ramachandran plot of the i + 2 residue at the type-II′ β-turn in **GB1** variants.

surprised to note a very good correlation (Fig. S7†) between the T_M of these proteins and $E_{n\to\pi^*}$ between C=O_{i+1} → C=O_{i+2}, in the absence of the stabilizing C=$O_i\cdots$HN$_{i+3}$ hydrogen bond (Fig. 4). An n → π* interaction rigidifies the β-turn by reducing the conformational entropy at the i + 2 residue, which is presumably responsible for the increased stability of the protein in solution. However, as the amino acid side-chains at the i + 2 residue of the type II′ β-turn are different in **2**–**5**, there might be additional factors that contribute towards the stability of these proteins. Therefore, we adopted an orthogonal strategy to validate the role of the n → π* interaction in protein stability.

By employing a prolyl-based torsion balance system, Raines *et al.* have shown that a thioamide (C=S_i) engages in a stronger C=S_i → C=O_{i+1} n → π* interaction than amide C=O_i.[20,50] However, due to the longer C=S bond length (1.71 Å)[51] and larger van der Waals radius of sulfur (1.85 Å),[52] thioamide substitution perturbs the local secondary structure of proteins where the amide oxygen participates in a shorter hydrogen bond.[53–57] On the other hand, thioamide substitution at a site where the amide oxygen is involved in a longer hydrogen bond or is solvent exposed, leads to minimal perturbation of the secondary structure.[53,55,57–59] Therefore, we chose to substitute the solvent exposed C=O_{i+1} in the type-II′ β-turn of **2** by C=S_{i+1}. The NBO analysis of the C=O_{i+1} to C=S_{i+1} substituted type-II′ β-turn in **2**, **3**, **4**, and **5** clearly indicated a significant enhancement in $E_{n\to\pi^*}$, due to the amplified C=S_{i+1} → C=O_{i+2} n → π* interaction (Table S3†).

Thus, towards the synthesis of i + 1 thionated **GB1** (D-Val^t–L-Ala; **2a**) (the thionated residue is denoted by superscript "t"), we obtained a clean 46-mer polypeptide up to the L-Ala$_{i+2}$. However, on completion of the 56-mer **2a** on a solid support, following the acidolytic removal of protecting groups, the mass spectrum corresponded to a 45-mer fragment without the L-Ala$_{i+2}$ (Fig. S16B†). To circumvent the undesirable peptide cleavage, we coupled the tetrapeptide Fmoc-Asn(Trt)-Gly-D-Val^t-L-Ala-COOH and Fmoc-Asn(Trt)-Gly-D-Val-L-Ala-COOH onto two individual 45-mer polypeptides. After acidolytic cleavage, although we obtained the 49-mer oxo-polypeptide, the thio-tetrapeptide coupling repeatedly resulted in the 45-mer fragment without the L-Ala$_{i+2}$ (Fig. S16C and D†). This suggests a spontaneous acid catalyzed cleavage of the peptide bond C-terminal to L-Ala$_{i+2}$ in thioamidated **GB1**, **2a**.

With numerous failed attempts to synthesize **2a**, we focused towards the 32-mer Pin 1 WW domain, a three stranded β-sheet protein that shows a cooperative two-state folding.[60,61] The Pin 1 protein is amenable to loop modification that retains the global fold with alteration in its thermodynamic stability, making it an excellent model protein for structure-folding studies.[60] We selected a Pin 1 variant with a type-I′ β-turn in loop 1 and substituted the –Asn–Gly– with D-Val–L-Ala– (**6**) and D-Ala–L-Ala– (**7**) (Fig. 6a) to adopt a type-II′ β-turn that was confirmed by characteristic NOEs at the β-turn (Fig. S29†). Subsequently, we synthesized Pin 1 variants with thioamidation at the i + 1 site (D-Val^t–L-Ala; **6a** and D-Ala^t–L-Ala; **7a**). Remarkably, the acidolytic removal of the protecting groups to obtain **6a** resulted in both the desired product and the N- and C-terminal fragmented peptides resulting from the nucleophilic attack of D-Val C=S_{i+1} onto L-Ala C=O_{i+2} (Fig. 7a) as observed in **2a**.



b)

	i+1 Xaa	*i*+2 Yaa	T_M (°C)	ΔT_M (°C)	ΔG_f (kcal mol^{-1})	$\Delta\Delta G_f$
6	D-Val	L-Ala	67.8 ± 0.21	-	–2.87 ± 0.08	-
6a	D-Val^t	L-Ala	70.1 ± 0.18	+2.3	–3.69 ± 0.09	–0.8
7	D-Ala	L-Ala	65.6 ± 0.33	-	–1.93 ± 0.06	-
7a	D-Ala^t	L-Ala	66.7 ± 0.15	+1.1	–2.21 ± 0.05	–0.3

Fig. 6 (a) Crystal structure of the Pin 1 WW domain (1ZCN) with –Asn–Gly– in loop 1 forming a type-I′ β-turn that has been modified to form a type-II′ β-turn in **6** and **7**. The n → π* interaction at the i + 2 residue is amplified by the C=S_{i+1} substitution in **6a** and **7a**. (b) The midpoint of the thermal transition (T_M ± S.D.) was derived from variable temperature CD. The free energy of folding (ΔG_f) was obtained by fitting the guanidine hydrochloride denaturation (4 °C) curves to a two-state model. $\Delta\Delta G_f$ = **6a** − **6** and **7a** − **7**.

An identical fragmentation was reported by Heimgartner *et al.* during the aqueous acidolytic workup of the thioacylated Aib–Pro dipeptide (Fig. 7b), towards the synthesis of Ph-(C=S)-Aib-Pro-Aib-N(Me)Ph.[62] However, by bubbling HCl gas through the dipeptide in THF, the thiazolone intermediate could be characterized, which results from the nucleophilic attack of Ph C=S_i onto Aib C=O_{i+1}. To our excitement, the crystal structure of the final product Ph-(C=S)-Aib-Pro-Aib-N(Me)Ph revealed the C=S_i → C=O_{i+1} n → π* interaction, leading to a high degree of pyramidalization, $\Delta = 0.059$ Å at Aib C=O_{i+1}, a firm indicator of the n → π* interaction.[19] Thus, the directional ($i + 1 \rightarrow i + 2$) fragmentation observed in **2a** (Fig. S16†), **6a** (Fig. 7a) and **7a** (Fig. S17B†) is a chemical signature of the amplified n → π* interaction between C=S_{i+1} and C=O_{i+2}.

Next, we assessed the folding of Pin 1 variants **6**, **6a**, **7**, and **7a** in sodium phosphate buffer (pH 7.4). All the Pin 1 proteins showed the characteristic 227 nm maximum in the CD spectrum, indicating the presence of a folded protein with β-sheets (minimum centered around 215 nm) (Fig. S22A–S25A†). The virtually identical H^{α} chemical shift perturbation deduced from TOCSY and NOESY experiments indicated that a single atom substitution (O to S) at the solvent exposed C=O_{i+1} did not lead to major structural perturbation in **6a** and **7a** (Fig. S28†). We next performed thermal and chemical denaturation (Fig. S22–S25†) to understand the effect of the amplified n → π* interaction. The proteins showed a two-state unfolding and we were delighted to note that the C=S_{i+1} → C=O_{i+2} n → π* interaction enhanced the stability of **6a** by 0.8 kcal mol^{-1} and **7a** by 0.3 kcal mol^{-1} (Fig. 6b).

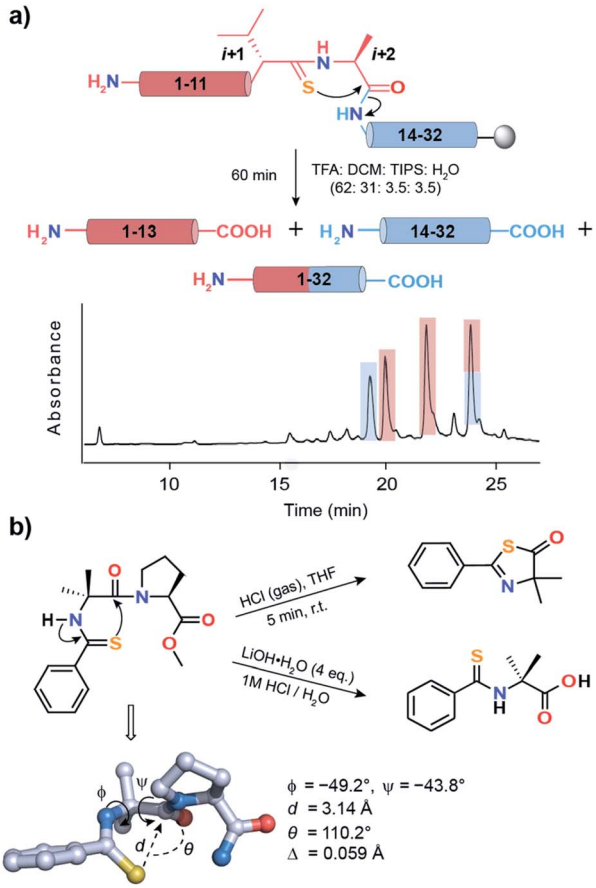


Fig. 7 (a) Acid-catalyzed cleavage of **6a** yielding two N-terminal fragments (1–13), one C-terminal fragment (14–32) and the desired product (1–32). The two N-terminal fragments possibly result from the racemization of the L-Ala_{i+2} due to keto–enol tautomerization of the thiazolone intermediate.[63] The polypeptides are color coded and are shown in the HPLC chromatogram. (b) Spontaneous acid-catalyzed cleavage of the thioacylated Aib–Pro dipeptide in aqueous solution.[62] The thiazolone intermediate was trapped by passing HCl gas and characterized. The crystal structure of Ph-(C=S)-Aib-Pro-Aib-N(Me)Ph clearly depicting the n → π* interaction between C=S_i and C=O_{i+1} of Aib.

The increased stability arises from the reduced conformational flexibility of the amino acid residue engaged in an n → π* interaction, a feature that is analogous to the ring constraint in proline, which restricts its conformational space compared to other amino acids and increases protein stability by reducing the entropy of the unfolded state.[64,65] An n → π* interaction restricts the conformational space of an amino acid residue with the adoption of torsion angles as depicted in the Ramachandran plot of the β-turn residues (Fig. 2c and d). This would also be expected in an amplified n → π* interaction by thioamide substitution. The adoption of such torsion angles is favorable at the $i + 1$ and $i + 2$ positions of a β-turn (Fig. 2a). Furthermore, since β-branched amino acids restrict the backbone conformation more than the unbranched residues,[65] the C=S_{i+1} → C=O_{i+2} n → π* interaction stabilizes **6a** more than **7a**.

Thus, our results in the Pin 1 WW domain re-emphasize the role of the amino acid side-chain in tuning the n → π* interaction energy. Not only the side-chain rotamer of the amino acid involved in an n → π* interaction dictates its strength (Fig. 4), the steric interactions imposed by the amino acid side-chain of the donor carbonyl $oxygen_{(i)}$ (C=S_{i+1} in this case) (Fig. 6) can also influence an n → π* interaction.

Conclusions

In summary, our bioinformatic analysis indicates that the reduced conformational freedom of the donor C=O_i by the intramolecular C=$O_i \cdots HN_{i+3}$ hydrogen bond in β-turns is associated with the high abundance of n → π* interactions at the $i + 1$ residue, whereas, the absence of the intramolecular hydrogen bond, constraining either the C=O_{i+1} or C=O_{i+2} results in conformational flexibility of the $i + 2$ residue, which could be restricted by introducing an C=O_{i+1} → C=O_{i+2} n → π* interaction. The experimental results at the $i + 2$ residue of the type-II′ β-turn in **GB1** variants suggest that amino acid side-chain identity and the rotamer conformation can modulate the strength of an n → π* interaction. Although, it is challenging to estimate the exact contribution of this energetically subtle interaction towards the global stability of the protein, we note that the altered rotamer conformation as a result of local structural changes can amplify/weaken an n → π* interaction affecting the backbone torsion angles (ϕ, ψ), thereby influencing its stability. With an enhanced n → π* interaction in the absence of the stabilizing intramolecular hydrogen bond, we

observe a clear shift of amino acid torsion angles (ϕ, ψ) from a non-helical to the right-handed α-helical region. It is worth noting that the $i \rightarrow i + 1$ directionality (N-term → C-term) associated with the n → π* interaction coincides with the formation of the productive helix nucleus at the N-terminus of a polypeptide,[66–68] highlighting an important contribution of the n → π* interaction towards helix nucleation. Furthermore, the recent report of a long-range n → π* interaction in stabilizing the α-helical conformation of a synthetic peptide in water, re-emphasizes the potential of this noncovalent interaction in engineering helical structures.[69]

To conclusively demonstrate the influence of the n → π* interaction on protein stability, we chose to amplify this weak noncovalent interaction by thioamide substitution. Since a strong n → π* interaction induces a "kink" in the polypeptide backbone by optimizing the ϕ, ψ torsion angles suitable for orbital overlap, and thereby reducing the conformational entropy at the β-turn, the thioamide substitution increased the protein stability. It is worth noting that thio-Gly465 in the natural protein methyl-coenzyme M reductase, which is suggested to stabilize the protein secondary structure near the active site, induces a kinked conformation (ϕ, $\psi = -68.5°$, $-47.2°$) by engaging in an n → π* interaction with C=O of Phe466 (Fig. S30†).[70,71] With the recent advancement in ribosome mediated incorporation of thioamide into proteins and polypeptides, thioamide substitution could be potentially utilized to stabilize turns and enhance protein stability,[72,73] aided by exogenous factors like salt concentration[11] and solvation by water molecules[69] that have been shown to influence the n → π* interaction in protein secondary structures.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

JC acknowledge DST-SERB for funding this work through the project EMR/2016/006193. The authors acknowledge the DBT-IISc partnership program phase-I, phase-II and IISc (MHRD) for funding this work. We also acknowledge funding for infrastructural support from the following programs of the Government of India: DST-FIST, UGC-CAS, and the DBT-IISc partnership program. The X-Ray diffraction facility for Macromolecular Crystallography at IISc is funded by DST-SERB grant IR/SO/LU/0003/2010-PHASE-II. BK thanks UGC for the research fellowship. JN thanks DST-SERB for funding under the Ramanujan Faculty Fellowship scheme, Prof. E. Arunan, IISc for the access to NBO6.0 and IPC clusters to carry out the NBO calculations and Compute Canada for the use of the Niagara supercomputer resources. AP is an Intermediate fellow of the DBT-Wellcome Trust India Alliance (IA/I/15/2/502063). Sunita Prakash is acknowledged for her tremendous support with the acquisition of LC-MS at the Proteomics facility in IISc. We thank Dr Amit Choudhary for stimulating discussions.

References

1 K. A. Dill, *Biochemistry*, 1990, **29**, 7133–7155.
2 R. W. Newberry and R. T. Raines, *ACS Chem. Biol.*, 2019, **14**, 1677–1686.
3 L. Pauling, R. B. Corey and H. R. Branson, *Proc. Natl. Acad. Sci. U. S. A.*, 1951, **37**, 205–211.
4 Z. S. Derewenda, L. Lee and U. Derewenda, *J. Mol. Biol.*, 1995, **252**, 248–262.
5 R. W. Newberry and R. T. Raines, *Nat. Chem. Biol.*, 2016, **12**, 1084–1088.
6 M. P. Hinderaker and R. T. Raines, *Protein Sci.*, 2003, **12**, 1188–1194.
7 S. K. Singh and A. Das, *Phys. Chem. Chem. Phys.*, 2015, **17**, 9596–9612.
8 R. W. Newberry and R. T. Raines, *Acc. Chem. Res.*, 2017, **50**, 1838–1846.
9 A. Rahim, P. Saha, K. K. Jha, N. Sukumar and B. K. Sarma, *Nat. Commun.*, 2017, **8**, 78.
10 H. B. Burgi, J. D. Dunitz and E. Shefter, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1974, **30**, 1517–1527.
11 G. J. Bartlett, A. Choudhary, R. T. Raines and D. N. Woolfson, *Nat. Chem. Biol.*, 2010, **6**, 615–620.
12 C. Fufezan, *Proteins*, 2010, **78**, 2831–2838.
13 C. E. Jakobsche, A. Choudhary, S. J. Miller and R. T. Raines, *J. Am. Chem. Soc.*, 2010, **132**, 6651–6653.
14 Z. S. Shi and N. R. Kallenbach, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3–4.
15 G. J. Bartlett, R. W. Newberry, B. VanVeller, R. T. Raines and D. N. Woolfson, *J. Am. Chem. Soc.*, 2013, **135**, 18682–18688.
16 H. R. Kilgore and R. T. Raines, *J. Am. Chem. Soc.*, 2018, **140**, 17606–17611.
17 L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 777–778.
18 M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines and J. L. Markley, *J. Am. Chem. Soc.*, 2002, **124**, 2497–2505.
19 P. Wilhelm, B. Lewandowski, N. Trapp and H. Wennemers, *J. Am. Chem. Soc.*, 2014, **136**, 15829–15832.
20 R. W. Newberry, B. VanVeller, I. A. Guzei and R. T. Raines, *J. Am. Chem. Soc.*, 2013, **135**, 7843–7846.
21 S. Hovmoller, T. Zhou and T. Ohlson, *Acta Crystallogr.*, 2002, **58**, 768–776.
22 N. J. Zondlo, *Nat. Chem. Biol.*, 2010, **6**, 567–568.
23 P. Deb, G. Y. Jin, S. K. Singh, J. Moon, H. Kwon, A. Das, S. Bagchi and Y. S. Kim, *J. Phys. Chem. Lett.*, 2018, **9**, 5425–5429.
24 C. Cabezas, J. L. Alonso, J. C. Lopez and S. Mata, *Angew. Chem., Int. Ed.*, 2012, **51**, 1375–1378.
25 S. K. Singh, K. K. Mishra, N. Sharma and A. Das, *Angew. Chem., Int. Ed.*, 2016, **55**, 7801–7805.
26 S. Blanco, J. C. Lopez, S. Mata and J. L. Alonso, *Angew. Chem., Int. Ed.*, 2010, **49**, 9187–9192.
27 Q. Gou, G. Feng, L. Evangelisti and W. Caminati, *Angew. Chem., Int. Ed.*, 2013, **52**, 11888–11891.
28 A. G. de Brevern, *Sci. Rep.*, 2016, **6**, 33191.

29 I. D. Kuntz, *J. Am. Chem. Soc.*, 1972, **94**, 4009–4012.
30 G. D. Rose, L. M. Gierasch and J. A. Smith, *Adv. Protein Chem.*, 1985, **37**, 1–109.
31 P. N. Lewis, F. A. Momany and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1971, **68**, 2293–2297.
32 H. E. Stanger and S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 4236–4237.
33 H. L. Fu, G. R. Grimsley, A. Razvi, J. M. Scholtz and C. N. Pace, *Proteins*, 2009, **77**, 491–498.
34 K. Watanabe, T. Masuda, H. Ohashi, H. Mihara and Y. Suzuki, *Eur. J. Biochem.*, 1994, **226**, 277–283.
35 T. Gallagher, P. Alexander, P. Bryan and G. L. Gilliland, *Biochemistry*, 1994, **33**, 4721–4729.
36 A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield and G. M. Clore, *Science*, 1991, **253**, 657–661.
37 Z. E. Reinert, G. A. Lengyel and W. S. Horne, *J. Am. Chem. Soc.*, 2013, **135**, 12528–12531.
38 H. L. Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson and C. M. Rienstra, *J. Phys. Chem. B*, 2007, **111**, 14362–14369.
39 B. L. Sibanda and J. M. Thornton, *Nature*, 1985, **316**, 170–174.
40 C. N. Pace and J. M. Scholtz, *Biophys. J.*, 1998, **75**, 422–427.
41 P. Lahiri, H. Verma, A. Ravikumar and J. Chatterjee, *Chem. Sci.*, 2018, **9**, 4600–4609.
42 D. J. Barlow and J. M. Thornton, *J. Mol. Biol.*, 1988, **201**, 601–619.
43 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899–926.
44 P. Chakrabarti and D. Pal, *Prog. Biophys. Mol. Biol.*, 2001, **76**, 1–102.
45 T. M. Gray and B. W. Matthews, *J. Mol. Biol.*, 1984, **175**, 75–81.
46 J. Janin and S. Wodak, *J. Mol. Biol.*, 1978, **125**, 357–386.
47 M. J. McGregor, S. A. Islam and M. J. Sternberg, *J. Mol. Biol.*, 1987, **198**, 295–310.
48 M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.*, 1991, **218**, 397–412.
49 J. S. Richardson and D. C. Richardson, *Science*, 1988, **240**, 1648–1652.
50 A. Choudhary, D. Gandla, G. R. Krow and R. T. Raines, *J. Am. Chem. Soc.*, 2009, **131**, 7244–7246.
51 M. R. Truter, *J. Chem. Soc.*, 1960, 997–1007.
52 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.
53 D. R. Artis and M. A. Lipton, *J. Am. Chem. Soc.*, 1998, **120**, 12200–12206.
54 R. M. Culik, H. Jo, W. F. DeGrado and F. Gai, *J. Am. Chem. Soc.*, 2012, **134**, 8026–8029.
55 R. W. Newberry, B. VanVeller and R. T. Raines, *Chem. Commun.*, 2015, **51**, 9624–9627.
56 A. Reiner, D. Wildemann, G. Fischer and T. Kiefhaber, *J. Am. Chem. Soc.*, 2008, **130**, 8079–8084.
57 C. R. Walters, D. M. Szantai-Kis, Y. T. Zhang, Z. E. Reinert, W. S. Horne, D. M. Chenoweth and E. J. Petersson, *Chem. Sci.*, 2017, **8**, 2868–2877.
58 J. H. Miwa, A. K. Patel, N. Vivatrat, S. M. Popek and A. M. Meyer, *Org. Lett.*, 2001, **3**, 3373–3375.
59 H. Verma, B. Khatri, S. Chakraborti and J. Chatterjee, *Chem. Sci.*, 2018, **9**, 2443–2451.
60 M. S. Ardejani, E. T. Powers and J. W. Kelly, *Acc. Chem. Res.*, 2017, **50**, 1875–1882.
61 M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele and J. W. Kelly, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10648–10653.
62 A. Budzowski, A. Linden and H. Heimgartner, *Helv. Chim. Acta*, 2008, **91**, 1471–1488.
63 J. H. Davies, R. H. Davis and R. A. G. Carrington, *J. Chem. Soc., Perkin Trans. 1*, 1972, 1983–1985.
64 U. Arnold, M. P. Hinderaker, J. Koditz, R. Golbik, R. Ulbrich-Hofmann and R. T. Raines, *J. Am. Chem. Soc.*, 2003, **125**, 7500–7501.
65 B. W. Matthews, H. Nicholson and W. J. Becktel, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 6663–6667.
66 A. Acharyya, Y. H. Ge, H. F. Wu, W. F. DeGrado, V. A. Voelz and F. Gai, *J. Phys. Chem. B*, 2019, **123**, 1797–1807.
67 L. Monticelli, D. P. Tieleman and G. Colombo, *J. Phys. Chem. B*, 2005, **109**, 20064–20067.
68 L. Pal, P. Chakrabarti and G. Basu, *J. Mol. Biol.*, 2003, **326**, 273–291.
69 H. N. Hoang, C. Y. Wu, T. A. Hill, A. D. de Araujo, P. V. Bernhardt, L. G. Liu and D. P. Fairlie, *Angew. Chem., Int. Ed.*, 2019, **58**, 18873–18877.
70 W. G. Grabarse, F. Mahlert, S. Shima, R. K. Thauer and U. Ermler, *J. Mol. Biol.*, 2000, **303**, 329–344.
71 D. D. Nayak, N. Mahanta, D. A. Mitchell and W. W. Metcalf, *eLife*, 2017, **6**, e29218.
72 R. Maini, L. M. Dedkova, R. Paul, M. M. Madathil, S. R. Chowdhury, S. X. Chen and S. M. Hecht, *J. Am. Chem. Soc.*, 2015, **137**, 11206–11209.
73 R. Maini, H. Kimura, R. Takatsuji, T. Katoh, Y. Goto and H. Suga, *J. Am. Chem. Soc.*, 2019, **141**, 20004–20008.