nature chemical biology

ARTICLE

PUBLISHED ONLINE: 11 JULY 2010 | DOI: 10.1038/NCHEMBIO.406

$n \rightarrow \pi^*$ interactions in proteins

Gail J Bartlett[1,5], Amit Choudhary[2,5], Ronald T Raines[3*] & Derek N Woolfson[1,4*]

Hydrogen bonds between backbone amides are common in folded proteins. Here, we show that an intimate interaction between backbone amides also arises from the delocalization of a lone pair of electrons (*n*) from an oxygen atom to the antibonding orbital (π^*) of the subsequent carbonyl group. Natural bond orbital analysis predicted significant $n \rightarrow \pi^*$ interactions in certain regions of the Ramachandran plot. These predictions were validated by a statistical analysis of a large, non-redundant subset of protein structures determined to high resolution. The correlation between these two independent studies is striking. Moreover, the $n \rightarrow \pi^*$ interactions are abundant and especially prevalent in common secondary structures such as α-, 3_{10}- and polyproline II helices and twisted β-sheets. In addition to their evident effects on protein structure and stability, $n \rightarrow \pi^*$ interactions could have important roles in protein folding and function, and merit inclusion in computational force fields.

Electron delocalization is a unifying theme behind many fundamental concepts of chemistry, including resonance and hyperconjugation. An especially important biological manifestation of electron delocalization is the hydrogen bond, which involves the delocalization of the lone pair (n) of the hydrogen bond acceptor over the antibonding orbital (σ^*) of the hydrogen bond donor[1–4]. Nearly 75 years ago, Mirsky and Pauling wrote: "The importance of the hydrogen bond in protein structure can hardly be overemphasized"[5]. We report on a contributor to protein structure, termed the "$n \rightarrow \pi^*$ interaction," that arises from electron delocalization analogous to that of the hydrogen bond.

The extent of overlap between donor and acceptor orbitals is governed by their relative spatial orientation. Contrary to the expectations of valence shell electron pair repulsion (VSEPR) theory, the two lone pairs of divalent oxygen do not occupy equivalent orbitals[6] that resemble "rabbit ears"[7,8]. For example, the lone pairs of a carbonyl oxygen are not in sp^2 hybrid orbitals but instead reside in non-degenerate s-rich and p-rich orbitals. The s-rich lone pair (n_s) has ~60% s and 40% p character, whereas the p-rich lone pair (n_p) has ~100% p character. In an amide oxygen, n_s is aligned with the σ bond of the carbonyl group (**Fig. 1a**), and n_p is orthogonal to the π bond of that group (**Fig. 1b**).

We postulate that an α-helix avails itself of both of these lone pairs. In particular, n_s delocalizes over the antibonding orbital (σ^*) of an N-H bond to form a hydrogen bond between residues i and i+4 in a polypeptide chain (**Fig. 1c**)[9,1–4]. Likewise, n_p is poised to delocalize over the antibonding orbital (π^*) of the i+1 carbonyl group (**Fig. 1d**). Model studies investigating peptides, peptoids and related small molecules have revealed that the ensuing $n \rightarrow \pi^*$ interaction can alter conformational preferences[10–20].

Here, we perform a thorough survey of protein conformational space for significant $n \rightarrow \pi^*$ interactions. First, we use calculations to anticipate the main chain dihedral angles that give rise to significant $n \rightarrow \pi^*$ interactions. We then perform a statistical analysis of a large, non-redundant subset of high-resolution protein crystal structures in the Protein Data Bank (PDB[21]) to search for such interactions. The results of these calculations and structural analyses indicate that $n \rightarrow \pi^*$ interactions are abundant and widespread in proteins.

RESULTS

Prevalence of $n \rightarrow \pi^*$ interactions

We sought to identify regions of protein conformational space that could contain significant $n \rightarrow \pi^*$ interactions (**Fig. 2**). We did so by creating and analyzing a computational library of allowed conformations. This library was generated with a tetramer of alanine residues capped at its N terminus with an acetyl group and at its C terminus with a methylamino group ($AcAla_4NHMe$). The conformations calculated by natural bond orbital (NBO) methods to have favorable $n \rightarrow \pi^*$ interaction energies ($E_{n \rightarrow \pi^*}$; **Supplementary Table 1**) are depicted in a Ramachandran-type plot in **Figure 2b**. Briefly, two broken elliptical regions are favored, one in each half of the ψ range. The uppermost region stretches from the allowed through the partially allowed β region and into the α_L conformations; the other spans the allowed and partially allowed α region and the disallowed regions in the lower-right quadrant of the plot.

The NBO analyses predict what protein conformations could give rise to $n \rightarrow \pi^*$ interactions. Do actual proteins avail themselves of those conformations and thus engage in $n \rightarrow \pi^*$ interactions? To answer this question, we generated a subset of high-resolution (≤1.6 Å) and non-redundant (≤30% pairwise sequence identity) X-ray crystal structures from the PDB. Within this set, $n \rightarrow \pi^*$ interactions between backbone carbonyl groups of consecutive residues (residue pairs) were identified using the operational definition given in **Figure 2a**. At distances ≤3.2 Å, the van der Waals surfaces of the two carbonyl groups interpenetrate, putting these contacts in the realm of quantum mechanics rather than classical electrostatics[16,20]; the allowed angles are reminiscent of the Bürgi-Dunitz trajectory for nucleophilic attack on a carbonyl carbon (~109°). Residue pairs that tested positive for $n \rightarrow \pi^*$ interactions are depicted in red on a Ramachandran plot (**Fig. 2c**). For comparison, the ϕ,ψ angles for all other residue pairs from the whole subset are depicted in black. As might be expected, the observed $n \rightarrow \pi^*$ interactions occur in allowed regions of Ramachandran space (**Fig. 2c**). What is remarkable, however, is that the favored regions for $n \rightarrow \pi^*$ interactions anticipated by the computational analyses correspond to those identified in the PDB search. The correlation is not exact because the hot spots identified by our computational analyses report on $n \rightarrow \pi^*$ interactions only, without

[1]School of Chemistry, University of Bristol, Bristol, United Kingdom. [2]Graduate Program in Biophysics, University of Wisconsin–Madison, Madison, Wisconsin, USA. [3]Departments of Biochemistry and Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA. [4]Department of Biochemistry, University of Bristol, Bristol, United Kingdom. [5]These authors contributed equally to this work. *e-mail: rtraines@wisc.edu or d.n.woolfson@bristol.ac.uk

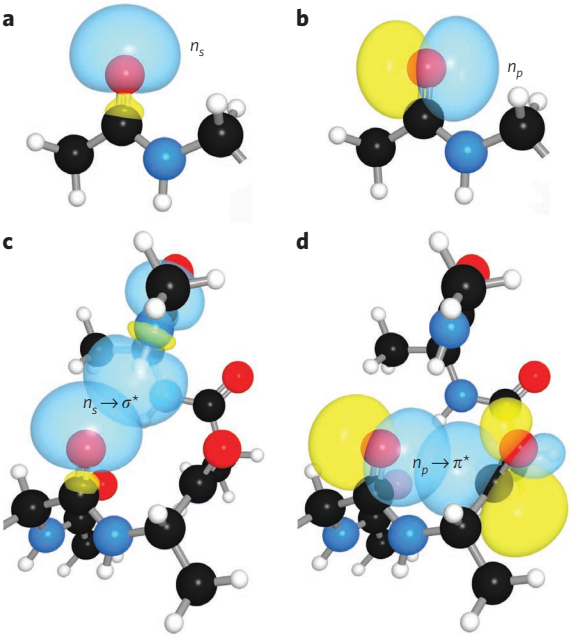

Figure 1 | Kinship of hydrogen bonds and $n\rightarrow\pi^*$ interactions in an α-helix. (**a**) *s*-rich lone pair of an amide oxygen. (**b**) *p*-rich lone pair of an amide oxygen. (**c**) $n_s\rightarrow\sigma^*$: hydrogen bond in an α-helix. (**d**) $n_p\rightarrow\pi^*$: $n\rightarrow\pi^*$ interaction in an α-helix.

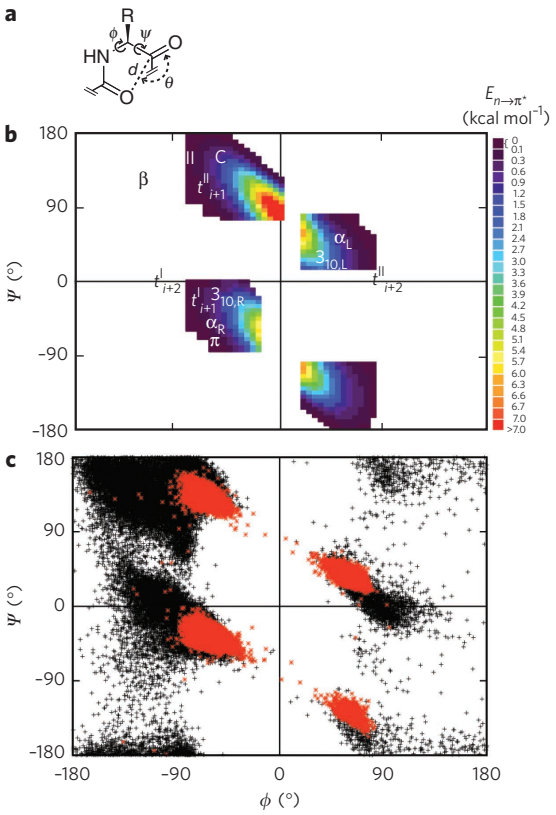

Figure 2 | Ramachandran plots of $n\rightarrow\pi^*$ interactions. (**a**) Definitions of dihedral angles ϕ (C'_{i-1}–N_i–$C^\alpha{}_i$–C'_i) and ψ (N_i–$C^\alpha{}_i$–C'_i–N_{i+1}), distance (*d*) and planar angle θ. Criteria for an $n\rightarrow\pi^*$ interaction in the crystallographic analyses: $d \leq 3.2$ Å; $99° \leq \theta \leq 119°$. (**b**) Computational data showing the energy of an $n\rightarrow\pi^*$ interaction for 836 ϕ,ψ angles. Common secondary structures are indicated. (**c**) Crystallographic data showing all residues in 1,731 high-resolution structures (black) and those that meet the criteria in panel **a** (red). Nineteen of those residues had $d \leq 2.60$ Å (**Supplementary Table 2**).

any steric considerations, and therefore cover regions of conformational space that are disallowed in protein structures.

Thus, both the computational and the crystallographic analyses suggest widespread and abundant $n\rightarrow\pi^*$ interactions in the allowed regions of the Ramachandran plot. Notably, we found $n\rightarrow\pi^*$ interactions in all 1,731 of the selected protein crystal structures, with proportions in the range 5–97% (mean: 34%). A breakdown of these data by secondary structure is detailed in **Table 1**. $n\rightarrow\pi^*$ interactions were especially prevalent in right- and left-handed α-helices but were less common in β-strands. One-quarter of the residues in left-handed polyproline II (P_{II}) helices participate in $n\rightarrow\pi^*$ interactions. A similar fraction was found in residues not in regular secondary structures, on the right-hand side of the Ramachandran plot (**Table 1** and **Fig. 2**).

Significance of Xaa-Pro dipeptides

The mean ϕ angle for the residues on the left-hand side of the Ramachandran plot—the side occupied by right-handed α-helices and β-strands—involved in $n\rightarrow\pi^*$ interactions is –63.3° ± 5.6° (**Fig. 2c**), which is close to the distribution of ϕ angles for proline residues (–65.0° ± 10.6°). This correlation, as well as prior results with a peptidic model system[11,16], suggested that proline could be particularly suited for engaging in an $n\rightarrow\pi^*$ interaction when found at the acceptor (*i*+1) position and led us to examine Xaa-Pro dipeptides (where Xaa is any amino acid), using Xaa-Ala and Xaa-Gly as control pairs. We found that Xaa-Pro pairs contributed significantly to the prevalence of $n\rightarrow\pi^*$ interactions in both α-helices and β-strands (**Table 1**). For α-helices, the amino acid preference for $n\rightarrow\pi^*$ interaction decreased in the order proline > alanine > glycine > all other amino acids, with almost all the proline residues participating in $n\rightarrow\pi^*$ interactions. For β-strands in which $n\rightarrow\pi^*$ interactions were found, the amino acid preference decreased in the order proline >> glycine > all, with a ten-fold preference for proline over glycine. In P_{II} helices, the distribution was less and differently skewed, all = alanine > proline >> glycine; we return to this important point below. In regions without regular secondary structure, a similar distribution was observed, with

Table 1 | Potential $n\rightarrow\pi^*$ interactions in different types of secondary structure

	α-helix		β-sheet		P_{II} helix		Other secondary structures	
Dipeptide	Total number of pairs	Those with potential $n\rightarrow\pi^*$ interactions	Total number of pairs	Those with potential $n\rightarrow\pi^*$ interactions	Total number of pairs	Those with potential $n\rightarrow\pi^*$ interactions	Total number of pairs	Those with potential $n\rightarrow\pi^*$ interactions
Xaa-Ala	11,384	9,367 (82%)	2,786	80 (2.8%)	24	6 (25%)	185	47 (25%)
Xaa-Gly	2,713	2,188 (81%)	2,036	67 (3.3%)	9	0	156	9 (5.7%)
Xaa-Pro	544	518 (95%)	527	193 (37%)	65	12 (18%)	324	63 (12%)
Xaa-All	88,124	63,792 (72%)	38,657	1,166 (3%)	227	58 (25%)	2,588	427 (16%)

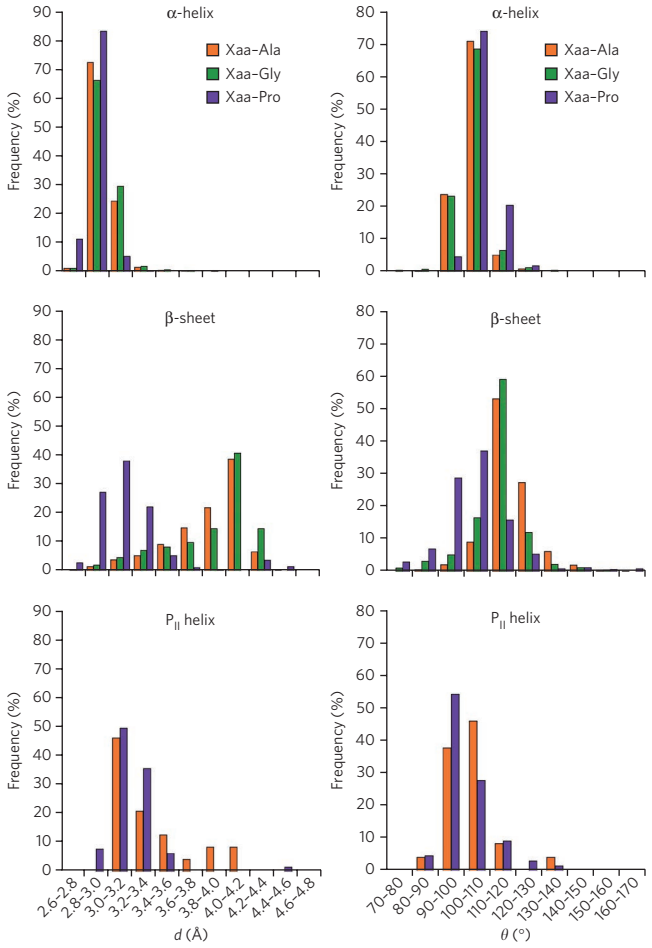

Figure 3 | Histograms of *d* and *θ* values for Xaa-Ala, Xaa-Gly and Xaa-Pro dipeptides in α-helices, β-sheets and P_{II} helices. Mean values (± s.d.) of *d* and *θ* were as follows. α-helix: Xaa-Ala, $d = 2.97 \pm 0.10$ Å, $\theta = 102.9 \pm 4.7°$; Xaa-Gly, $d = 3.01 \pm 0.17$ Å, $\theta = 103.0 \pm 7.6°$; Xaa-Pro, $d = 2.89 \pm 0.10$ Å, $\theta = 107.2 \pm 6.5°$. β-sheet: Xaa-Ala, $d = 3.87 \pm 0.32$ Å, $\theta = 118.1 \pm 9.3°$; Xaa-Gly, $d = 3.89 \pm 0.36$ Å, $\theta = 112.7 \pm 11.3°$; Xaa-Pro, $d = 3.17 \pm 0.33$ Å, $\theta = 103.8 \pm 13.5°$. P_{II} helix: Xaa-Ala, $d = 3.16 \pm 0.33$ Å, $\theta = 97.7 \pm 9.5°$; Xaa-Pro, $d = 3.01 \pm 0.21$ Å, $\theta = 95.6 \pm 8.7°$. Mean values of *d* for Xaa-Pro were smaller than those for Xaa-Ala and Xaa-Gly in each secondary structure, according to Student's *t*-test ($P < 0.05$, one-tailed test, number of observations as in **Table 1**). Mean values of *θ* for Xaa-Pro were larger than those for Xaa-Ala and Xaa-Gly in α-helices but smaller than those for Xaa-Ala and Xaa-Gly in β-sheets, according to Student's *t*-test ($P < 0.05$, one-tailed test, number of observations as in **Table 1**). The number of Xaa-Gly dipeptides in P_{II} helices was too small for their inclusion.

proline engaging in proportionally fewer $n \rightarrow \pi^*$ interactions than alanine and all other residues but more than glycine.

Detailed geometry of $n \rightarrow \pi^*$ interactions

We examined in detail the $O_{i-1} \cdots C_i$ distance (*d*, **Fig. 2a**) as well as the $O_{i-1} \cdots C_i = O_i$ angle (*θ*, **Fig. 2a**) for Xaa-Pro, Xaa-Gly and Xaa-Ala dipeptides in protein structures. For α-helices, β-strands, P_{II} helices and runs of residues with no secondary structure, we found that the mean $O_{i-1} \cdots C_i$ distance in Xaa-Pro was shorter by ~0.1 Å than the same distance in Xaa-Gly and Xaa-Ala (**Fig. 3**). In α-helices, the mean $O_{i-1} \cdots C_i = O_i$ angle was 107.2° ± 6.5° for Xaa-Pro motifs, compared with 103.0° ± 7.6° for Xaa-Gly and 102.9° ± 4.6° for Xaa-Ala. This wider angle is close to the Bürgi-Dunitz trajectory. In β-strands, the mean $O_{i-1} \cdots C_i = O_i$ angle was 103.8° ± 13.5° for Xaa-Pro, compared with 118.1° ± 9.3° for Xaa-Ala and 112.7° ± 11.3° for Xaa-Gly.

Insufficient data were available for Xaa-Gly pairs in P_{II} helices because of the requisite non-redundancy of the dataset. Many collagen structures, which consist of three intertwined glycine- and proline-rich P_{II} helices[15], were removed because of similarity between their Pro-Hyp-Gly repeats. To address this issue explicitly, we examined 15 high-resolution (≤1.6 Å) structures of collagen-like peptides. We found the mean value of *d* in Xaa-Pro dipeptides to be shorter than in Xaa-Gly by ~0.1 Å, similar to the difference found in α-helices and β-strands. The mean $O_{i-1} \cdots C_i = O_i$ angles were, however, shallower (80.9° ± 2.2° for Xaa-Gly; 89.4° ± 2.4° for Xaa-Pro) than that for an optimal $n \rightarrow \pi^*$ interaction, which precluded their designation as $n \rightarrow \pi^*$ interactions in the original search (**Table 1**).

Continuity of $n \rightarrow \pi^*$ interactions

To identify only definitive cases of $n \rightarrow \pi^*$ interactions, we used a conservative operational definition in our PDB search (**Fig. 2a**). The value of $E_{n \rightarrow \pi^*}$ is, however, a continuum that depends on the values of *d* and *θ*. The NBO analyses indicate significant conformational stabilization offered by the $n \rightarrow \pi^*$ interaction, even at angles outside the range $99° \leq \theta \leq 119°$ (**Fig. 4a**). In other words, significant $n \rightarrow \pi^*$ interactions are likely to be more abundant than implied by the constrained PDB data depicted in **Figures 2c** and **3**. For example, the mean $O_{i-1} \cdots C_i = O_i$ angle in P_{II} helices was only slightly shallower than 99° for Xaa-Ala (97.7° ± 9.5°) and Xaa-Pro (95.6° ± 8.7°) dipeptides.

Predetermination of $n \rightarrow \pi^*$ interactions?

Do $n \rightarrow \pi^*$ interactions stabilize protein structures, or are their signature $O_{i-1} \cdots C_i$ distances and $O_{i-1} \cdots C_i = O_i$ angles merely an artifact of protein topology? To address this question, we analyzed the relationship between *d* and *θ* for all residue pairs in the 1,731 high-resolution structures used to generate **Figures 2c** and **3**. We found that α-helices, the most prevalent secondary structure in the dataset, are dominated by values of $d \leq 3.2$ Å and *θ* near 100° (**Fig. 4b**), as expected from the data shown in **Figure 3**. The tendency of *θ* to be near 100° when $d \leq 3.2$ Å is also evident in residue pairs that are not in an α-helix (meaning that they are largely unconstrained by secondary structure; **Fig. 4c**). The value of *θ* varies widely, however, when $d > 3.2$ Å, which is an unexpected result if the distances and angles are topological artifacts. In other work, some of us have discovered that in the absence of an $n \rightarrow \pi^*$ interaction, the donor group neither makes a short contact with the acceptor ($d > 3.2$ Å) nor resides along the Bürgi-Dunitz trajectory ($\theta \approx 125°$)[20]. These results suggest that $n \rightarrow \pi^*$ interactions are not an artifact of protein topology.

Protein structures with many or few $n \rightarrow \pi^*$ interactions

Some of the 1,731 structures that we examined had an inordinately high or low fraction of residues in $n \rightarrow \pi^*$ interactions, according to the criteria in **Figure 2a**. These structures are listed and depicted in **Supplementary Table 5**. In eight structures, >80% of the residues participate in $n \rightarrow \pi^*$ interactions. These are small α-helical proteins; five comprise single α-helices, three have helix-turn-helix structures. Notably, almost all of them are coiled-coil proteins, and the extensive $n \rightarrow \pi^*$ interactions could enhance the rigidity of those coiled coils. At the other end, 39 structures have <10% of the residues in $n \rightarrow \pi^*$ interactions. In these structures, almost all the secondary structure is β-sheet, and only four structures have any residues in α-helices. For these structures, the proportion of residues not in regular secondary structures is slightly greater than in the entire dataset (37% versus 34%). Proteins in this group include streptavidin, tumor necrosis factor-α and a variety of carbohydrate-binding proteins.

We searched the PDB for runs of four or more $n \rightarrow \pi^*$ interactions not contained within common secondary structural elements in proteins. This search returned only eight hits from a possible

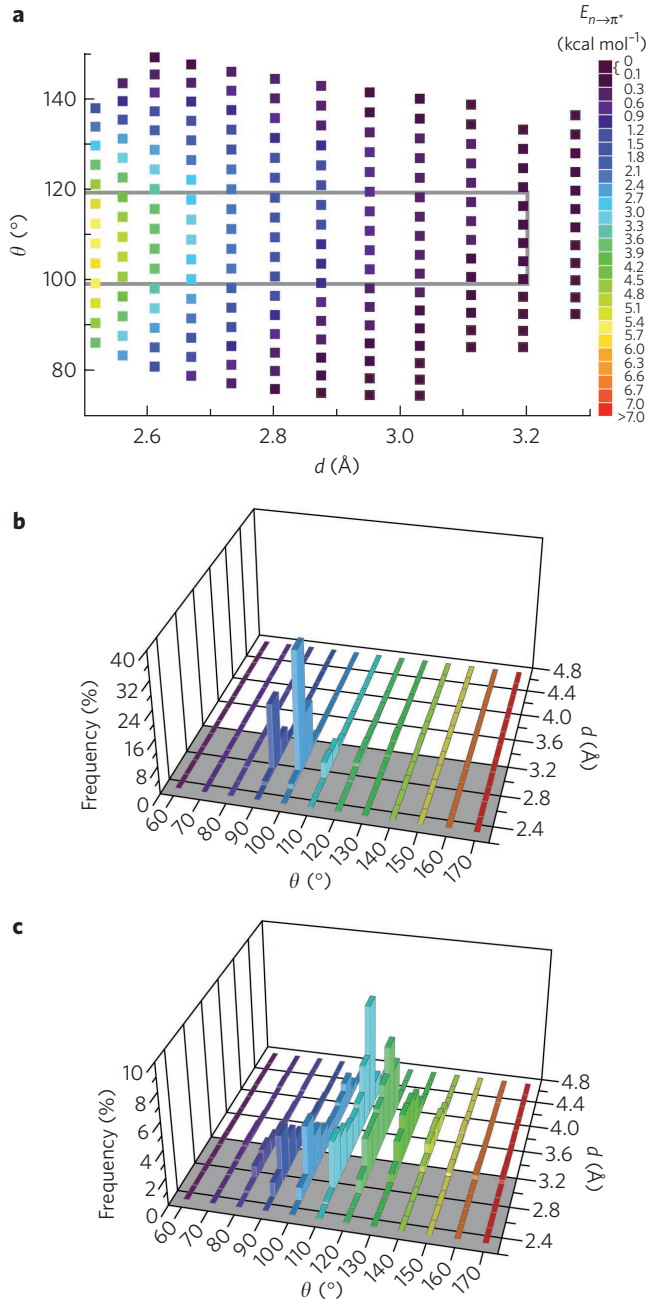

Figure 4 | Values of *d* and *θ* in $n\to\pi^*$ interactions. (**a**) Energy of the $n\to\pi^*$ interactions for values of *d* and *θ* in the right-handed helical region of the Ramachandran plot ($-180 \le \phi \le 0$; $-180 \le \psi \le 0$). The criteria in **Figure 2a** are obeyed within the gray rectangle. (**b**) Frequency of values of *d* and *θ* among residue pairs (132,149) in an α-helix (**Supplementary Table 3**). (**c**) Frequency of values of *d* and *θ* among residue pairs (221,469) not in an α-helix (**Supplementary Table 4**). Combined, the plots in **b** and **c** include all residue pairs (353,618) in 1,731 high-resolution structures; however, there are more data plotted here than are listed in **Table 1**, as the latter were culled for four-residue sequences, whereas the data used in this figure were from residue pairs. The frequencies of residue pairs with $d \le 3.2$ Å (**Fig. 2a**) emerge from the gray plane.

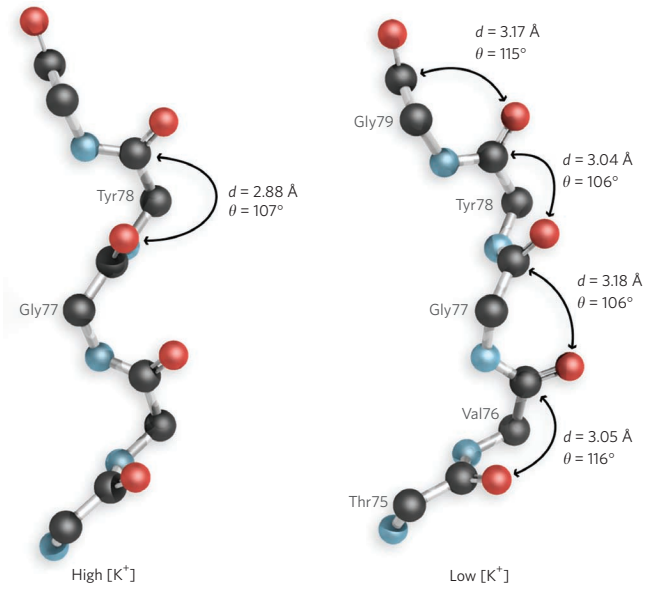

Figure 5 | Potential $n\to\pi^*$ interactions in the selectivity filter of the KcsA K$^+$ channel. A single chain of the tetrameric channel structure determined at a high K$^+$ concentration (PDB code 1k4c) was superposed on that determined at a low K$^+$ concentration (PDB code 1k4d). Superpositions were carried out by the Secondary Structure Matching service at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm).

7,833 sets of four residues. Detailed examination showed that these hits were located between α-helices or β-strands, except for one that formed part of an iron-binding site (residues 246–251 inclusive in cytosine deaminase, PDB code 1ra0 (ref. 22), of which His246 is a ligand to the iron ion). Most notably, however, in a wider search of the PDB, we found a run of four $n\to\pi^*$ interactions in the selectivity filter of the potassium ion channel from *Streptomyces lividans* (KcsA; **Fig. 5**)[23], whereas $n\to\pi^*$ interactions were nearly absent in the structure of the filter region determined at high K$^+$ concentration.

DISCUSSION

$n\to\pi^*$ interactions in quadrants of the Ramachandran plot

Our computational and structural analyses reveal abundant and widespread $n\to\pi^*$ interactions in three quadrants of the Ramachandran plot. This finding has implications for common secondary structures, as follows.

The $\phi = 0$ to $-180°$, $\psi = 0$ to $-180°$ quadrant houses the most common right-handed helices, such as α-, 3_{10}- and π-helices. The stabilization offered by the $n\to\pi^*$ interaction to these secondary structures as determined by NBO analyses is 0.5 kcal mol^{-1} for an α-helix ($\phi,\psi = -57°,-47°$), 1.3 kcal mol^{-1} for a 3_{10}-helix ($-49°,-26°$), and 0.1 kcal mol^{-1} for a π-helix ($-57°,-70°$). For comparison, experiments have revealed that the $n\to\pi^*$ interaction in *N*-formyl-L-proline methyl ester (FmProOMe) (which has an ester carbon rather than amide as the acceptor) has $E_{n\to\pi^*}$ ~0.7 kcal mol^{-1} (ref. 11).

The existence of an $n\to\pi^*$ interaction in α-helices is supported by considerable evidence. Two diagnostic signatures of the quantum mechanics that underlie the $n\to\pi^*$ interaction[16] have been observed in α-helices: pyramidalization of the acceptor carbonyl group[24] and polarization of its π-electron cloud[25]. The origin of these perturbations was previously unknown. Moreover, detailed analyses of peptides have demonstrated that $n\to\pi^*$ interactions provide significant conformational stabilization at $d \le 3.2$ Å[10,11,16,20]. Such distances are encountered in 72% of α-helical residues (**Table 1**).

We hypothesize that, in addition to stabilizing the α-helix, the $n\to\pi^*$ interaction could have an important role in α-helix formation. The nucleation of an α-helix is energetically unfavorable: forming the characteristic $i\to i+4$ hydrogen bond necessitates the restriction of at least six main chain dihedral angles and a consequent loss in conformational entropy, and the enthalpy of a helix-coil transition has been estimated to be ~0.9 kcal mol^{-1} per

residue[26]. Additionally, considerable enthalpic cost is incurred because the amidic dipoles of adjacent residues repel each other in an α-helical conformation[27]. The $n\rightarrow\pi^*$ interaction, which operates between the *i* and *i*+1 residues, could mitigate these energetic penalties. We note that, consistent with this proposal, proline is a strong helix initiator[28–31] and has a high propensity for engaging in the $n\rightarrow\pi^*$ interaction (**Table 1**)[10–13,16]. Moreover, the purported intermediacy of the 3_{10}-helix[32–35] and its prevalence at α-helix termini suggest that the strong $n\rightarrow\pi^*$ interaction within a 3_{10} helix (1.3 kcal mol^{-1}) could have an important role in α-helix formation. The short $O_{i-1}\cdots C_i$ distance induced by an $n\rightarrow\pi^*$ interaction (**Fig. 1d**) causes lateral compaction of the α-helix that could align the hydrogen bond donor and acceptor groups for a stronger $i\rightarrow i+4$ hydrogen bond (**Fig. 1c**).

Neither parallel nor antiparallel β-strands of the $\phi = 0$ to −180°, $\psi = 0$ to 180° quadrant show significant $n\rightarrow\pi^*$ interactions. In these extended structures, adjacent carbonyl groups are too distal for an intimate interaction ($d > 3.2$ Å). Nonetheless, a significant proportion of β-sheets have a twist with $d < 3.2$ Å. It is plausible that the $n\rightarrow\pi^*$ interaction provides a consequential conformational stabilization to these twisted β-strands, which then form twisted β-sheets. Occasionally, β-strands have an amplified right-handed twist resulting in local disruption of the β-sheet structure. Such β-bulges are involved in the dimerization of immunoglobulin domains and can assist in enclosing the active sites of proteins[36–39]. Two common types of β-bulges, the G1 and wide types, adopt ϕ and ψ dihedral angles indicative of considerable $n\rightarrow\pi^*$ interactions. In particular, of all sets of three residues comprising G1 bulges in our dataset identified using PROMOTIF[40], the 25% that partake in $n\rightarrow\pi^*$ interactions have constrained mean (ϕ, ψ) angles (residue X: −68.0° ± 5.5°, 130.4° ± 6.6°; residue 1: 58.4° ± 17.6°, 36.4° ± 11.0°; residue 2: −68.7° ± 6.8°, 136.9° ± 6.5°) that place them in energetic 'hot spots' on the Ramachandran plot. The contribution of $n\rightarrow\pi^*$ interactions within P_{II} helical strands to the conformational stability of the collagen triple helix has been described previously[15].

β-turns have been implicated in protein folding[41,42]. Our computational analyses indicate that $n\rightarrow\pi^*$ interactions confer conformational stability to the *i*+1 residues in common type I and type II β-turns (**Fig. 2b**). As in an α-helix (**Fig. 1c**,**d**), the donor oxygen participates in a hydrogen bond (here, $i\rightarrow i+3$) along with the $n\rightarrow\pi^*$ interaction. We note that the *i*+1 residues of β-turns, like all of the residues in a P_{II} helix, adopt their conformation without the aid of local or intra–secondary structure hydrogen bonds. Accordingly, the $n\rightarrow\pi^*$ interaction could have a special role to play in the stability of these structures.

Although they occur less frequently than their right-handed counterparts in proteins, left-handed helices of the $\phi = 0$ to 180°, $\psi = 0$ to 180° quadrant have distinctive structural and functional roles in substrate specificity, cofactor binding and protein-protein interactions[43]. For example, the left-handed α-helical conformation of MPK-3 is critical for the binding and subsequent negative regulation of mitogen-activated protein kinase (MAPK)[44]. The $n\rightarrow\pi^*$ interaction could be important in the stability and folding of the left-handed α-helices and 3_{10} helices.

Possible functional roles and other implications

Could $n\rightarrow\pi^*$ interactions contribute to protein function? Their prevalence in canonical secondary structures (**Table 1**) implies a broad impact. We note, however, an especially intriguing situation in a non-canonical structure. The selectivity filter of the potassium ion channel from *S. lividans* (KcsA[23]) has a run of $n\rightarrow\pi^*$ interactions, but only in the structure determined at low K^+ concentration (**Fig. 5**). The filter is occluded in this 'closed' state, and only two dehydrated K^+ ions are present in the channel. At high K^+ concentration, $n\rightarrow\pi^*$ interactions do not appear to play a significant role, and four dehydrated K^+ ions are evident within the channel. These observations suggest that main chain oxygens in the potassium ion channel can participate in $n\rightarrow\pi^*$ interactions or bind to K^+ ions, but not both.

Our results suggest several areas for future investigation. For example, we note that the formation of $n\rightarrow\pi^*$ interactions could be cooperative, as the polarization of an acceptor $C'_{i+1} = O_{i+1}$ group makes O_{i+1} into a better donor. We speculate that the conduit provided by $n\rightarrow\pi^*$ interactions could facilitate the tunneling of electrons through proteins[45]. Finally, we urge the inclusion of $n\rightarrow\pi^*$ interactions in relevant computational force fields.

METHODS

Computational analyses. Geometry optimization and frequency calculation for an $AcAla_4NHMe$ model peptide were performed at the B3LYP/6-31+G* level of theory with all the amide bonds locked in the *trans* conformation[46]. Frequency calculations of the optimized structure yielded no imaginary frequencies. The library of allowed conformations was generated by varying the ϕ and ψ dihedral angles systematically in increments of 5°. Natural bond orbital (NBO) analyses were performed at the B3LYP/6-311+G(2*d*,*p*) level of theory on each conformation. The NBO method transforms a calculated wavefunction into a localized form, which corresponds to the lone pair and bond elements of the Lewis structure[47]. The stabilization afforded by the $n\rightarrow\pi^*$ interaction, $E_{n\rightarrow\pi^*}$, was estimated by using second-order perturbation theory as implemented in NBO 5.0. Orbital depictions were generated with NBOView 1.1.

PDB analyses. A subset of high-resolution (≤1.6 Å) and non-redundant (pairwise sequence identity ≤30%) protein X-ray crystal structures was collated from the PDB using PISCES[48] with data downloaded on 18 December 2007. The following operational definition was used to identify unambiguous examples of $n\rightarrow\pi^*$ interaction in these proteins: $C'_{i-1} = O_{i-1}\cdots C'_i = O_i$ contacts in which (i) the distance between the amide oxygen (O_{i-1}) and subsequent amide carbon (C'_i) was less than the sum of their van der Waals radii ($d \leq r_O + r_C = 3.2$ Å; **Fig. 2**), and (ii) the angle between O_{i-1} and $C'_i = O_i$ was $99° \leq \theta \leq 119°$ (**Fig. 2**). Distributions of these short contacts were studied as a function of the main chain dihedral angles ϕ and ψ. α-helix and β-sheet assignments were made with PROMOTIF[40], which uses modified Kabsch and Sander criteria[49] to assign secondary structure. P_{II} helices were assigned with the method of Stapley[50], which assigns regions of at least four residues in length to be P_{II} helices if their ϕ and ψ angles fall within angle constraints particular to this secondary structural class and do not fall into any other structural class according to the Kabsch and Sander criteria[49]. Residue pairs were only considered to belong to a structural class if they were residues X and Y in sequence WXYZ in which all four amino acids were assigned to that structural class, with the exception of the data shown in **Figure 4b,c**, in which only residue pairs (like X and Y) were considered.

Received 8 February 2010; accepted 6 June 2010; published online 11 July 2010

References

1. Isaacs, E.D., Shukla, A., Platzman, P.M., Hamann, D.R., Barbiellini, B. & Tulk, C.A. Covalency of the hydrogen bond in ice: a direct x-ray measurement. *Phys. Rev. Lett.* **82**, 600–603 (1999).
2. Weinhold, F. & Landis, C.R. *Valency and Bonding: A Natural Bond Orbital Donor–Acceptor Perspective* (Cambridge University Press, Cambridge, UK, 2005).
3. Weinhold, F. Resonance character of hydrogen-bonding interactions in water and other H-bonded species. *Adv. Protein Chem.* **72**, 121–155 (2005).
4. Khaliullin, R.Z., Cobar, E.A., Lochan, R.C., Bell, A.T. & Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A* **111**, 8753–8765 (2007).
5. Mirsky, A.E. & Pauling, L. On the structure of native, denatured, and coagulated proteins. *Proc. Natl. Acad. Sci. USA* **22**, 439–447 (1936).
6. Gray, H.B. *Electrons and Chemical Bonding* (W.A. Benjamin, New York, 1965).
7. Raber, D.J., Raber, N.K., Chandrasekhar, J. & Schleyer, P.v.R. Geometries and energies of complexes between formaldehyde and first- and second-row cations. A theoretical study. *Inorg. Chem.* **23**, 4076–4080 (1984).
8. Laing, M. No rabbit ears on water. *J. Chem. Educ.* **64**, 124–128 (1987).
9. Pauling, L., Corey, R.B. & Branson, H.R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205–211 (1951).
10. DeRider, M.L. *et al.* Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J. Am. Chem. Soc.* **124**, 2497–2505 (2002).
11. Hinderaker, M.P. & Raines, R.T. An electronic effect on protein structure. *Protein Sci.* **12**, 1188–1194 (2003).

12. Horng, J.-C. & Raines, R.T. Stereoelectronic effects on polyproline conformation. *Protein Sci.* **15**, 74–83 (2006).
13. Hodges, J.A. & Raines, R.T. Energetics of an $n\rightarrow\pi^*$ interaction that impacts protein structure. *Org. Lett.* **8**, 4695–4697 (2006).
14. Gao, J. & Kelly, J.W. Toward quantification of protein backbone–backbone hydrogen bonding energies: an energetic analysis of an amide-to-ester mutation in an α-helix within a protein. *Protein Sci.* **17**, 1096–1101 (2008).
15. Shoulders, M.D. & Raines, R.T. Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958 (2009).
16. Choudhary, A., Gandla, D., Krow, G.R. & Raines, R.T. Nature of amide carbonyl–carbonyl interactions in proteins. *J. Am. Chem. Soc.* **131**, 7244–7246 (2009).
17. Dai, N. & Etzkorn, F.A. *Cis–trans* proline isomerization effects on collagen triple-helix stability are limited. *J. Am. Chem. Soc.* **131**, 13728–13732 (2009).
18. Gorske, B.C., Stringer, J.R., Bastian, B.L., Fowler, S.A. & Blackwell, H.E. New strategies for the design of folded peptoids revealed by a survey of noncovalent interactions in model systems. *J. Am. Chem. Soc.* **131**, 16555–16567 (2009).
19. Pal, T.K. & Sankararamakrishnan, R. Quantum chemical investigations on intraresidue carbonyl–carbonyl contacts in aspartates of high-resolution protein structures. *J. Phys. Chem. B* **114**, 1038–1049 (2010).
20. Jakobsche, C.E., Choudhary, A., Raines, R.T. & Miller, S.J. $n\rightarrow\pi^*$ interaction and $n)(\pi$ Pauli repulsion are antagonistic for protein stability. *J. Am. Chem. Soc.* **132**, 6651–6653 (2010).
21. Berman, H., Henrick, K., Nakamura, H. & Markley, J.L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **35**, D301–D303 (2007).
22. Mahan, S.D., Ireton, G.C., Knoeber, C., Stoddard, B.L. & Black, M.E. Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy. *Protein Eng. Des. Sel.* **17**, 625–633 (2004).
23. Zhou, Y., Morais-Cabral, J.H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K^+ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
24. Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. Pyramidalization of backbone carbonyl carbon atoms in proteins. *Protein Sci.* **9**, 2038–2042 (2000).
25. Lario, P.I. & Vrielink, A. Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. *J. Am. Chem. Soc.* **125**, 12787–12794 (2003).
26. Makhatadze, G.I. Thermodynamics of α-helix formation. *Adv. Protein Chem.* **72**, 199–226 (2006).
27. Yang, A.-S. & Honig, B. Free energy determinants of secondary structure formation: I. α-Helices. *J. Mol. Biol.* **252**, 351–365 (1995).
28. Tanaka, S. & Scheraga, H.A. Statistical mechanical treatment of protein conformation. I. Conformational properties of amino acids in proteins. *Macromolecules* **9**, 142–159 (1976).
29. Toniolo, C., Bonora, G.M., Mutter, M. & Pillai, V.N.R. Linear oligopeptides. 78. The effect of the insertion of a proline residue on the solution conformation of host peptides. *Makromol. Chem.* **182**, 2007–2014 (1981).
30. Altmann, K.-H., Wojcik, J., Vasquez, M. & Scheraga, H.A. 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 (1990).
31. Yun, R.H., Anderson, A. & Hermans, J. Proline in α-helix: stability and conformations studied by dynamics simulation. *Proteins* **10**, 219–228 (1991).
32. Venkatachalapathi, Y.V. & Balaram, P. An incipient 3_{10} helix in Piv–Pro–Pro–Ala-NHMe as a model for peptide folding. *Nature* **281**, 83–84 (1979).
33. Tobias, D.J. & Brooks, C.L. III. Thermodynamics and mechanism of α-helix initiation in alanine and valine peptides. *Biochemistry* **30**, 6059–6070 (1991).
34. Sheinerman, F.B. & Brooks, C.L. III. 3_{10} helices in peptides and proteins as studied by modified Zimm–Bragg Theory. *J. Am. Chem. Soc.* **117**, 10098–10103 (1995).
35. Monticelli, L.P., T.D. & Colombo, G. Mechanism of helix nucleation and propagation: Microscopic view from microsecond time scale MD simulation. *J. Phys. Chem. B* **109**, 20064–20067 (2005).
36. Richardson, J.S., Getzoff, E.D. & Richardson, D.C. The β-bulge: a common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. USA* **75**, 2574–2578 (1978).
37. Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. Domain association on immunogloblin molecules. The packing of variable domains. *J. Mol. Biol.* **186**, 651–663 (1985).
38. Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K. & Stuart, D.I. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* **360**, 232–239 (1992).
39. Chan, A.W.E., Hutchinson, E.G., Harris, D. & Thornton, J.M. Identification, classification, and analysis of β-bulges in proteins. *Protein Sci.* **2**, 1574–1590 (1993).
40. Hutchinson, E.G. & Thornton, J.M. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).
41. Lewis, P.N., Momany, F.A. & Scheraga, H.A. Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc. Natl. Acad. Sci. USA* **68**, 2293–2297 (1971).
42. Zimmerman, S.S. & Scheraga, H.A. Local interactions in bends of proteins. *Proc. Natl. Acad. Sci. USA* **74**, 4126–4129 (1977).
43. Novotny, M. & Kleywegt, G.J. A survey of left-handed helices in protein structures. *J. Mol. Biol.* **347**, 231–241 (2005).
44. Farooq, A. *et al.* Solution structure of ERK2 binding domain of MAPK phosphatase MKP-3: Structural insights into MKP-3 activation by ERK2. *Mol. Cell* **7**, 387–399 (2001).
45. Gray, H.B. & Winkler, J.R. Electron flow through proteins. *Chem. Phys. Lett.* **483**, 1–9 (2009).
46. Frisch, M.J. *et al.* Gaussian 03, Revision C.02 (Gaussian, Inc., Wallingford, Connecticut, USA, 2004).
47. Weinhold, F. Natural bond orbital methods. in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P.v.R. *et al.*) **3**, 792–1811 (John Wiley & Sons, Chichester, UK, 1998).
48. Wang, G. & Dunbrack, R.L. PISCES: A protein sequence culling server. *Bioinformatics* **19**, 1589–1591 (2003).
49. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
50. Stapley, B.J. & Creamer, T.P. A survey of left-handed polyproline II helices. *Protein Sci.* **8**, 587–595 (1999).

Acknowledgments

We thank J. Spencer, J. Harvey, A. Mulholland, B. Bromley, M.D. Shoulders, B.R. Caes, C.N. Bradford, M.J. Palte and E. Moutevelis for helpful discussions. Financial support was provided by the University of Bristol and the Biotechnology and Biological Sciences Research Council of the United Kingdom (grant D003016) to D.N.W. and the US National Institutes of Health grant R01 AR044276 to R.T.R.

Author contributions

D.N.W. and R.T.R. conceived the project. G.J.B. and D.N.W. designed the PDB analyses; G.J.B. performed the PDB analyses. A.C. and R.T.R. designed the computational analyses; A.C. performed the computational analyses. All of the coauthors wrote and edited the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to D.N.W. or R.T.R.