

Published in final edited form as:

*Nat Chem Biol.* 2010 August ; 6(8): 615–620. doi:10.1038/nchembio.406.

## $n \rightarrow \pi^*$ Interactions in Proteins

Gail J. Bartlett<sup>1,†</sup>, Amit Choudhary<sup>2,†</sup>, Ronald T. Raines<sup>3,\*</sup>, and Derek N. Woolfson<sup>1,4,\*</sup>

<sup>1</sup> School of Chemistry, University of Bristol, Bristol, BS8 1TS, United Kingdom

<sup>2</sup> Graduate Program in Biophysics, University of Wisconsin–Madison, Madison, WI 53706, USA

<sup>3</sup> Departments of Biochemistry and Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

<sup>4</sup> Department of Biochemistry, University of Bristol, Bristol, BS8 1TD, United Kingdom

### Abstract

Hydrogen bonds between backbone amides are common in folded proteins. Here, we show that an intimate interaction between backbone amides likewise arises from the delocalization of a lone pair of electrons ( $n$ ) from an oxygen atom to the antibonding orbital ( $\pi^*$ ) 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  $\alpha$ -,  $3_{10}$ -, and polyproline II helices, and twisted  $\beta$ -sheets. In addition to their evident effects on protein structure and stability,  $n \rightarrow \pi^*$  interactions could play 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 ( $\sigma^*$ ) of the hydrogen-bond donor<sup>1–4</sup>. Nearly 75 years ago, Mirsky and Pauling wrote: “The importance of the hydrogen bond in protein structure can hardly be overemphasized”<sup>5</sup>. 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<sup>6</sup> that resemble “rabbit ears”<sup>7,8</sup>. For example, the lone pairs of a carbonyl oxygen are not in

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence and requests for materials should be addressed to D.N.W. or R.T.R. [rtraines@wisc.edu](mailto:rtraines@wisc.edu); [d.n.woolfson@bristol.ac.uk](mailto:d.n.woolfson@bristol.ac.uk).

†Both authors contributed equally to this work.

### 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 co-authors wrote and edited the manuscript.

### Additional Information

Supplementary information accompany this paper at [www.nature.com/naturechemicalbiology/](http://www.nature.com/naturechemicalbiology/). Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>.

$sp^2$  hybrid orbitals, but 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  $\sigma$  bond of the carbonyl group (Fig. 1a), and  $n_p$  is orthogonal to the  $\pi$  bond of that group (Fig. 1b).

We postulate that an  $\alpha$ -helix avails both of these lone pairs. In particular,  $n_s$  delocalizes over the antibonding orbital ( $\sigma^*$ ) of an N–H bond to form a hydrogen bond between residues  $i$  and  $i+4$  in a polypeptide chain (Fig. 1c),<sup>1–4,9</sup>. Likewise,  $n_p$  is poised to delocalize over the antibonding orbital ( $\pi^*$ ) 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<sup>10–20</sup>.

Here, we perform a thorough survey of protein conformational space for significant  $n \rightarrow \pi^*$  interactions. First, we employ 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<sup>21</sup>) 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<sub>4</sub>NHMe). The conformations calculated by natural bond orbital (NBO) methods to have favorable  $n \rightarrow \pi^*$  interaction energies ( $E_{n \rightarrow \pi^*}$ ; Table S1) are depicted in a Ramachandran-type plot in Fig. 2b. Briefly, two broken elliptical regions are favored, one in each half of the  $\psi$  range. The uppermost region stretches from the allowed, through partially allowed  $\beta$ -region, and into the  $\alpha_L$  conformations; the other spans the allowed through partially allowed  $\alpha$ -region and 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 those conformations and thus employ  $n \rightarrow \pi^*$  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 Fig. 2a. At distances  $\leq 3.2$  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<sup>16,20</sup>; the allowed angles are reminiscent of the Bürgi–Dunitz trajectory for nucleophilic attack on a carbonyl carbon ( $\sim 109^\circ$ ). Residue pairs that tested positive for  $n \rightarrow \pi^*$  interactions are depicted in red on a Ramachandran plot, Fig. 2c. For comparison, the  $\phi, \psi$  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 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. Importantly, we found  $n \rightarrow \pi^*$  interactions in *all* 1731 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  $\alpha$ -helices, but were less common in  $\beta$ -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  $\phi$  angle for the residues on the left-hand side of the Ramachandran plot—*i.e.*, that occupied by right-handed  $\alpha$ -helices and  $\beta$ -strands—involved in  $n \rightarrow \pi^*$  interactions is  $-63.3^\circ \pm 5.6^\circ$  (Fig. 2c), which is close to the distribution of  $\phi$  angles for proline residues ( $-65.0^\circ \pm 10.6^\circ$ ). This correlation, as well as prior results with a peptidic model system<sup>11,16</sup>, 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  $\alpha$ -helices and  $\beta$ -strands (Table 1). For  $\alpha$ -helices, the amino-acid preference for  $n \rightarrow \pi^*$  interaction decreased in the order: Pro > Ala > Gly > all other amino acids, with almost all the proline residues participating in  $n \rightarrow \pi^*$  interactions. For  $\beta$ -strands in which  $n \rightarrow \pi^*$  interactions were found, the amino-acid preference decreased in the order: Pro  $\gg$  Gly > all, with a ten-fold preference for proline over glycine. In  $P_{II}$  helices, the distribution was less and differently skewed, all = Ala > Pro  $\gg$  Gly; we return to this important point below. In regions without regular secondary structure, a similar distribution was observed, with 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  $O_{i-1} \cdots C_j$  distance ( $d$ , Fig. 2a) as well as the  $O_{i-1} \cdots C_j = O_i$  angle ( $\theta$ , Fig. 2a) for Xaa–Pro, Xaa–Gly, and Xaa–Ala dipeptides in protein structures. For  $\alpha$ -helices,  $\beta$ -strands,  $P_{II}$  helices, and runs of residues with no secondary structure, we found that the mean  $O_{i-1} \cdots C_j$  distance in Xaa–Pro was shorter by  $\sim 0.1$  Å than the same distance in Xaa–Gly and Xaa–Ala (Fig. 3). In  $\alpha$ -helices, the mean  $O_{i-1} \cdots C_j = O_i$  angle was  $107.2^\circ \pm 6.5^\circ$  for Xaa–Pro motifs, compared with  $103.0^\circ \pm 7.6^\circ$  for Xaa–Gly and  $102.9^\circ \pm 4.6^\circ$  for Xaa–Ala. This wider angle is close to the Bürgi–Dunitz trajectory. In  $\beta$ -strands, the mean  $O_{i-1} \cdots C_j = O_i$  angle was  $103.8^\circ \pm 13.5^\circ$  for Xaa–Pro, compared with  $118.1^\circ \pm 9.3^\circ$  for Xaa–Ala and  $112.7^\circ \pm 11.3^\circ$  for Xaa–Gly.

Insufficient data were available for Xaa–Gly pairs in  $P_{II}$  helices due to the requisite non-redundancy of the dataset. Many collagen structures, which consist of three intertwined glycine- and proline-rich  $P_{II}$  helices<sup>15</sup>, were removed because of similarity through their Pro–Hyp–Gly repeats. To address this issue explicitly, we examined 15 high-resolution ( $\sim 1.6$  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  $\sim 0.1$  Å, similar to the difference found in  $\alpha$ -helices and  $\beta$ -strands. The mean  $O_{i-1} \cdots C_j = O_i$  angles were, however, shallower ( $80.9^\circ \pm 2.2^\circ$  for Xaa–Gly;  $89.4^\circ \pm 2.4^\circ$  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 employed 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  $\theta$ . The NBO analyses indicate significant conformational stabilization offered by the  $n \rightarrow \pi^*$  interaction, even at angles outside the range  $99^\circ \leq \theta \leq 119^\circ$  (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 Fig. 2c and 3. For example, the mean  $O_{i-1} \cdots C_F=O_i$  angle in P<sub>II</sub> helices was only slightly shallower than  $99^\circ$  for Xaa-Ala ( $97.7^\circ \pm 9.5^\circ$ ) and Xaa-Pro ( $95.6^\circ \pm 8.7^\circ$ ) 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  $O_{i-1} \cdots C_F=O_i$  angles merely an artifact of protein topology? To address this question, we analyzed the relationship between  $d$  and  $\theta$  for all residue pairs in the 1731 high-resolution structures used to generate Fig. 2c and 3. We found that  $\alpha$ -helices, the most prevalent secondary structure in the dataset, are dominated by values of  $d \approx 3.2$  Å and  $\theta$  near  $100^\circ$  (Fig. 4b), as expected from Fig. 3. The tendency of  $\theta$  to be near  $100^\circ$  when  $d \approx 3.2$  Å is also evident in residues pairs that are not in an  $\alpha$ -helix (*i.e.*, largely unconstrained by secondary structure; Fig. 4c). The value of  $\theta$  varies widely, however, when  $d > 3.2$  Å, 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 ( $\theta \approx 125^\circ$ )<sup>20</sup>. 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 1731 structures that we examined had an inordinately high or low fraction of residues in  $n \rightarrow \pi^*$  interactions, according to the criteria in Fig. 2a. These structures are listed and depicted in Table S5. In 8 structures, >80% of the residues participate in  $n \rightarrow \pi^*$  interactions. These are small  $\alpha$ -helical proteins; five comprise single  $\alpha$ -helices, three have helix-turn-helix structures. Interestingly, 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  $\beta$ -sheet, and only four structures have any residues in  $\alpha$ -helices. For these structures, the proportion of residues not in regular secondary structures is slightly greater than in the entire data set (37% versus 34%). Proteins in this group include streptavidin, tumor necrosis factor- $\alpha$ , 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 7833 sets of four residues. Detailed examination showed that these hits were located between  $\alpha$ -helices or  $\beta$ -strands, except for one that formed part of an iron-binding site (residues 246–251 inclusive in cytosine deaminase, PDB code 1ra0<sup>22</sup>, 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  $n \rightarrow \pi^*$  interactions in the selectivity filter of the potassium ion channel from *Streptomyces lividans* (KcsA; Fig. 5)<sup>23</sup>, while  $n \rightarrow \pi^*$  interactions were nearly absent in the structure of the filter region determined at high  $K^+$  concentration.

## Discussion

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

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

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

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

In addition to stabilizing the  $\alpha$ -helix, we hypothesize that the  $n \rightarrow \pi^*$  interaction could have an important role in  $\alpha$ -helix formation. The nucleation of an  $\alpha$ -helix is energetically unfavorable: forming the characteristic  $i \rightarrow i+4$  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  $\sim 0.9$  kcal/mol/residue<sup>26</sup>. Additionally, considerable enthalpic cost is incurred because the amidic dipoles of adjacent residues repel each other in an  $\alpha$ -helical conformation<sup>27</sup>. The  $n \rightarrow \pi^*$  interaction, which operates between the  $i$  and  $i+1$  residues, could mitigate these energetic penalties. Consistent with this proposal, we note that proline is a strong helix initiator<sup>28–31</sup>, and has a high propensity for engaging in the  $n \rightarrow \pi^*$  interaction (Table 1)<sup>10–13,16</sup>. Moreover, the purported intermediacy of the  $3_{10}$ -helix<sup>32–35</sup> and its prevalence at  $\alpha$ -helix termini suggest that the strong  $n \rightarrow \pi^*$  interaction within a  $3_{10}$  helix (1.3 kcal/mol) could have an important role in  $\alpha$ -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  $\alpha$ -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 anti-parallel  $\beta$ -strands of the  $\phi = 0$  to  $-180^\circ$ ,  $\psi = 0$  to  $180^\circ$  quadrant exhibit 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  $\beta$ -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  $\beta$ -strands, which then form twisted  $\beta$ -sheets. Occasionally,  $\beta$ -strands have an amplified right-handed twist resulting in local disruption of the  $\beta$ -sheet structure. Such  $\beta$ -bulges are involved in the dimerization of immunoglobulin domains and can assist in enclosing the active sites of proteins<sup>36–39</sup>. Two common types of  $\beta$ -bulges, the G1 and wide types, adopt  $\phi$  and  $\psi$  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<sup>40</sup>, the 25% that partake in  $n \rightarrow \pi^*$  interactions have constrained mean ( $\phi, \psi$ ) angles (Residue X:  $-68.0^\circ \pm 5.5^\circ, 130.4^\circ \pm 6.6^\circ$ ; Residue 1:  $58.4^\circ \pm 17.6^\circ, 36.4^\circ \pm 11.0^\circ$ ; Residue 2:  $-68.7^\circ \pm 6.8^\circ, 136.9^\circ \pm 6.5^\circ$ ) which place them in energetic ‘hotspots’ 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<sup>15</sup>.

$\beta$ -Turns have been implicated in protein folding<sup>41,42</sup>. Our computational analyses indicate that  $n \rightarrow \pi^*$  interactions confer conformational stability to the  $i+1$  residues in common type I and type II  $\beta$ -turns (Fig. 2b). As in an  $\alpha$ -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  $\beta$ -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  $180^\circ$ ,  $\psi = 0$  to  $180^\circ$  quadrant play distinctive structural and functional roles in substrate specificity, cofactor binding, and protein–protein interactions<sup>43</sup>. For example, the left-handed  $\alpha$ -helical conformation of MPK-3 is critical for the binding and subsequent negative regulation of mitogen-activated protein kinase (MAPK)<sup>44</sup>. The  $n \rightarrow \pi^*$  interaction could play an important role in the stability and folding of the left-handed  $\alpha$ -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 channel from *Streptomyces lividans* (KcsA<sup>23</sup>) 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<sup>45</sup>. 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<sub>4</sub>NHMe model peptide were performed at the B3LYP/6-31+G\* level of theory with all the amide bonds locked in the *trans* conformation<sup>46</sup>. Frequency calculations of the optimized structure yielded no imaginary frequencies. The library of allowed conformations was generated by varying the  $\phi$  and  $\psi$  dihedral angles systematically in increments of  $5^\circ$ . Natural bond orbital (NBO) analyses were performed at the B3LYP/6-311+G(2d,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<sup>47</sup>. 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 ( $\leq 1.6 \text{ \AA}$ ), and non-redundant (pairwise sequence identity  $\geq 30\%$ ) protein X-ray crystal structures was collated from the PDB using PISCES<sup>48</sup> 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 (1) 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 ( $d(O_{i-1} \cdots C'_i) < r_O + r_C = 3.2 \text{ \AA}$ ; Fig. 2); and (2) the angle between  $O_{i-1}$  and  $C'_i=O_i$  was  $99^\circ \leq \theta \leq 119^\circ$  (Fig. 2). Distributions of these short contacts were studied as a function of the main-chain dihedral angles  $\phi$  and  $\psi$ .  $\alpha$ -Helix and  $\beta$ -sheet assignments were made with PROMOTIF<sup>40</sup>, which uses modified Kabsch and Sander criteria<sup>49</sup> to assign secondary structure. PPII helices were assigned with the method of Stapley<sup>50</sup>, which assigns regions of at least four residues in length to be PPII helices if their  $\phi$  and  $\psi$  angles fall within angle constraints particular to this secondary structural class and do not fall into any structural class according to the Kabsch and Sander criteria<sup>49</sup>. 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 Fig. 4b and 4c, in which only residue pairs (*i.e.*, X and Y) were considered.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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 BBSRC of the UK (grant D003016) to D.N.W., and grant R01 AR044276 (NIH) to R.T.R.

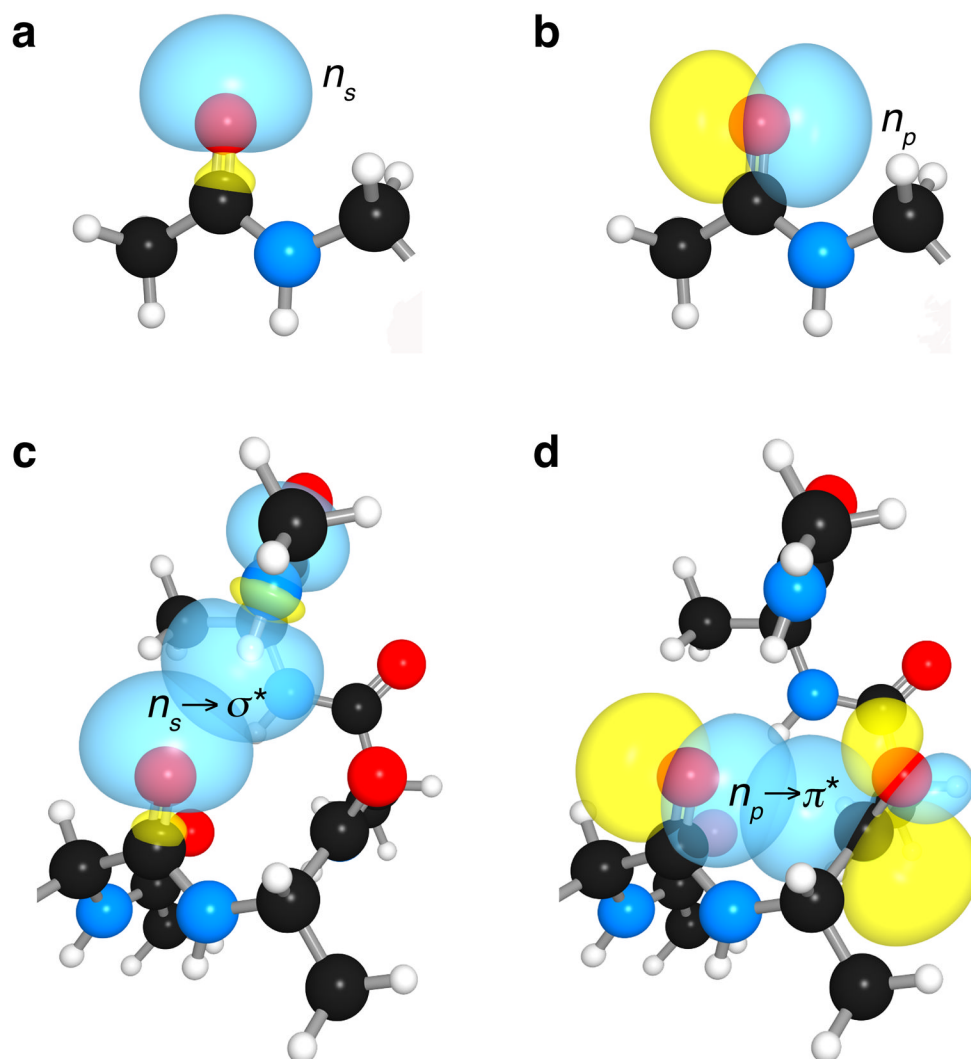
## References

1. Isaacs ED, Shukla A, Platzman PM, Barbiellini DR, Tulk CA. Covalency of the hydrogen bond in ice: A direct x-ray measurement. *Phys Rev Lett*. 1999; 82:600–603.
2. Weinhold, F.; Landis, CR. 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*. 2005; 72:121–155. [PubMed: 16581375]
4. Khaliullin RZ, Cobar EA, Lochan RC, Bell AT, Head-Gordon M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J Phys Chem A*. 2007; 111:8753–8765. [PubMed: 17655284]
5. Mirsky AE, Pauling L. On the structure of native, denatured, and coagulated proteins. *Proc Natl Acad Sci USA*. 1936; 22:439–447. [PubMed: 16577722]
6. Gray, HB. *Electrons and Chemical Bonding*. W. A. Benjamin; New York: 1965.
7. Raber DJ, Raber NK, Chandrasekhar J, Scheleyer PvR. Geometries and energies of complexes between formaldehyde and first- and second-row cations. A theoretical study. *Inorg Chem*. 1984; 23:4076–4080.
8. Laing M. No rabbit ears on water. *J Chem Educ*. 1987; 64:124–128.
9. Pauling L, Corey RB, Branson HR. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA*. 1951; 37:205–211. [PubMed: 14816373]
10. DeRider ML, 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*. 2002; 124:2497–2505. [PubMed: 11890798]

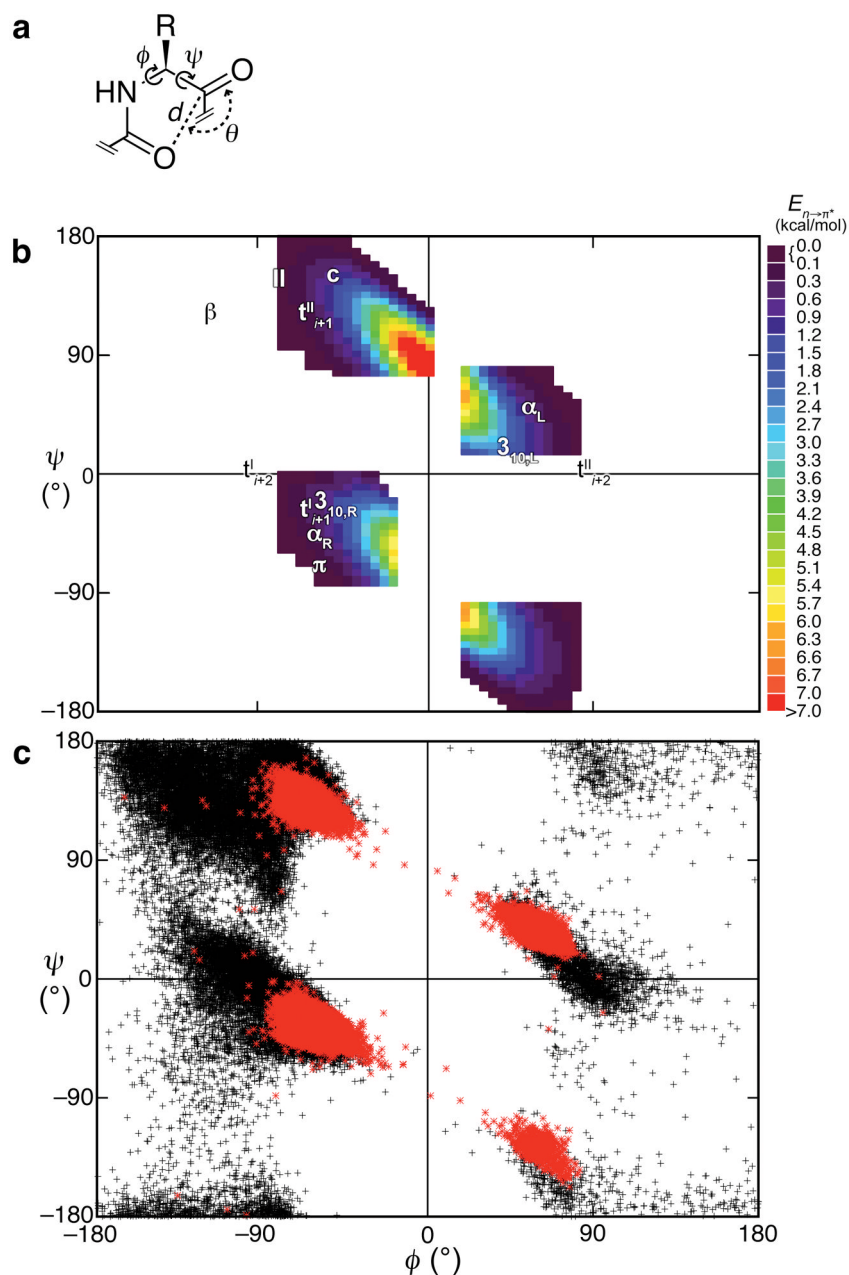
11. Hinderaker MP, Raines RT. An electronic effect on protein structure. *Protein Sci.* 2003; 12:1188–1194. [PubMed: 12761389]
12. Horng JC, Raines RT. Stereoelectronic effects on polyproline conformation. *Protein Sci.* 2006; 15:74–83. [PubMed: 16373476]
13. Hodges JA, Raines RT. Energetics of an  $n \rightarrow \pi^*$  interaction that impacts protein structure. *Org Lett.* 2006; 8:4695–4697. [PubMed: 17020280]
14. Gao J, Kelly JW. Toward quantification of protein backbone–backbone hydrogen bonding energies: An energetic analysis of an amide-to-ester mutation in an  $\alpha$ -helix within a protein. *Protein Sci.* 2008; 17:1096–1101. [PubMed: 18434500]
15. Shoulders MD, Raines RT. Collagen structure and stability. *Annu Rev Biochem.* 2009; 78:929–958. [PubMed: 19344236]
16. Choudhary A, Gandla D, Krow GR, Raines RT. Nature of amide carbonyl–carbonyl interactions in proteins. *J Am Chem Soc.* 2009; 131:7244–7246. [PubMed: 19469574]
17. Dai N, Etzkorn FA. *Cis–trans* proline isomerization effects on collagen triple-helix stability are limited. *J Am Chem Soc.* 2009; 131:13728–13732. [PubMed: 19725497]
18. Gorske BC, Stringer JR, Bastian BL, Fowler SA, Blackwell HE. New strategies for the design of folded peptoids revealed by a survey of noncovalent interactions in model systems. *J Am Chem Soc.* 2009; 131:16555–16567. [PubMed: 19860427]
19. Pal TK, Sankararamakrishnan R. Quantum chemical investigations on intraresidue carbonyl–carbonyl contacts in aspartates of high-resolution protein structures. *J Phys Chem B.* 2010; 114:1038–1049. [PubMed: 20039723]
20. Jakobsche CE, Choudhary A, Raines RT, Miller SJ.  $n \rightarrow \pi^*$  Interaction and  $n(\pi)$  Pauli repulsion are antagonistic for protein stability. *J Am Chem Soc.* 2010; 132:6651–6653. [PubMed: 20420451]
21. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* 2007; 35:D301–D303. [PubMed: 17142228]
22. Mahan SD, Ireton GC, Knoeber C, Stoddard BL, Black ME. Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy. *Protein Eng Des Sel.* 2004; 17:625–633. [PubMed: 15381761]
23. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R. Chemistry of ion coordination and hydration revealed by a  $K^+$  channel–Fab complex at 2.0 Å resolution. *Nature.* 2001; 414:43–48. [PubMed: 11689936]
24. Esposito L, Vitagliano L, Zagari A, Mazzarella L. Pyramidalization of backbone carbonyl carbon atoms in proteins. *Protein Sci.* 2000; 9:2038–2042. [PubMed: 11106179]
25. Lario PI, Vrielink A. Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. *J Am Chem Soc.* 2003; 125:12787–12794. [PubMed: 14558826]
26. Makhatadze GI. Peptide solvation and H-bonds. *Adv Protein Chem.* 2006; 72:199–226. [PubMed: 16581378]
27. Yang AS, Honig B. Free energy determinants of secondary structure formation: I.  $\alpha$ -Helices. *J Mol Biol.* 1995; 252:351–365. [PubMed: 7563056]
28. Tanaka S, Scheraga HA. Statistical mechanical treatment of protein conformation. I. Conformational properties of aminoacids in proteins. *Macromolecules.* 1976; 9:142–159. [PubMed: 1249985]
29. Toniolo C, Bonora GM, Mutter M, Pillai VNR. Linear oligopeptides. 78 The effect of the insertion of a proline residue on the solution conformation of host peptides. *Makromol Chem.* 1981; 182:2007–2014.
30. Altmann KH, Wojcik J, Vasquez M, Scheraga HA. Helix–coil stability constants for the naturally occurring amino acids in water. XXIII Proline parameters from random poly(hydroxybutylglutamine–co–L-proline). *Biopolymers.* 1990; 30:107–120. [PubMed: 2224046]
31. Yun RH, Anderson A, Hermans J. Proline in  $\alpha$ -helix: Stability and conformations studied by dynamics simulation. *Proteins.* 1991; 10:219–228. [PubMed: 1881878]



32. Venkatachalapathi YV, Balaram P. An incipient  $3_{10}$  helix in Piv-Pro-Pro-Ala-NHMe as a model for peptide folding. *Nature*. 1979; 281:83–84. [PubMed: 551271]
33. Tobias DJ, Brooks CL III. Thermodynamics and mechanism of  $\alpha$ -helix initiation in alanine and valine peptides. *Biochemistry*. 1991; 30:6059–6070. [PubMed: 2043644]
34. Sheinerman FB, Brooks CL III.  $3_{10}$  Helices in peptides and proteins as studied by modified Zimm–Bragg Theory. *J Am Chem Soc*. 1995; 117:10098–10103.
35. Monticelli L, PTD, Colombo G. Mechanism of helix nucleation and propagation: Microscopic view from microsecond time scale MD simulation. *J Phys Chem B*. 2005; 109:20064–20067. [PubMed: 16853593]
36. Richardson JS, Getzoff ED, Richardson DC. The  $\beta$ -bulge: A common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci USA*. 1978; 75:2574–2578. [PubMed: 275827]
37. Chothia C, Novotny J, Brucoleri R, Karplus M. Domain association on immunoglobulin molecules. The packing of variable domains. *J Mol Biol*. 1985; 186:651–663. [PubMed: 4093982]
38. Jones EY, Davis SJ, Williams AF, Harlos K, Stuart DI. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature*. 1992; 360:232–239. [PubMed: 1279440]
39. Chan AWE, Hutchinson EG, Harris D, Thornton JM. Identification, classification, and analysis of  $\beta$ -bulges in proteins. *Protein Sci*. 1993; 2:1574–1590. [PubMed: 8251933]
40. Hutchinson EG, Thornton JM. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci*. 1996; 5:212–220. [PubMed: 8745398]
41. Lewis PN, Momany FA, Scheraga HA. Folding of polypeptide chains in proteins: A proposed mechanism for folding. *Proc Natl Acad Sci USA*. 1971; 68:2293–2297. [PubMed: 5289387]
42. Zimmerman SS, Scheraga HA. Local interactions in bends of proteins. *Proc Natl Acad Sci USA*. 1977; 74:4126–4129. [PubMed: 270658]
43. Novotny M, Kleywegt GJ. A survey of left-handed helices in protein structures. *J Mol Biol*. 2005; 347:231–241. [PubMed: 15740737]
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*. 2001; 7:387–399. [PubMed: 11239467]
45. Gray HB, Winkler JR. Electron flow through proteins. *Chem Phys Lett*. 2009; 483:1–9. [PubMed: 20161522]
46. Frisch, MJ., et al. Gaussian 03, Revision C.02. Gaussian, Inc; Wallingford, Connecticut, USA: 2004.
47. Weinhold, F. Natural Bond Orbital Methods. In: Schleyer, PvR, et al., editors. *Encyclopedia of Computational Chemistry*. Vol. 3. John Wiley & Sons; Chichester, UK: 1998. p. 1792–1811.
48. Wang G, Dunbrack RL. PISCES: A protein sequence culling server. *Bioinformatics*. 2003; 19:1589–1591. [PubMed: 12912846]
49. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983; 22:2577–2637. [PubMed: 6667333]
50. Stapley BJ, Creamer TP. A survey of left-handed polyproline II helices. *Protein Sci*. 1999; 8:587–595. [PubMed: 10091661]

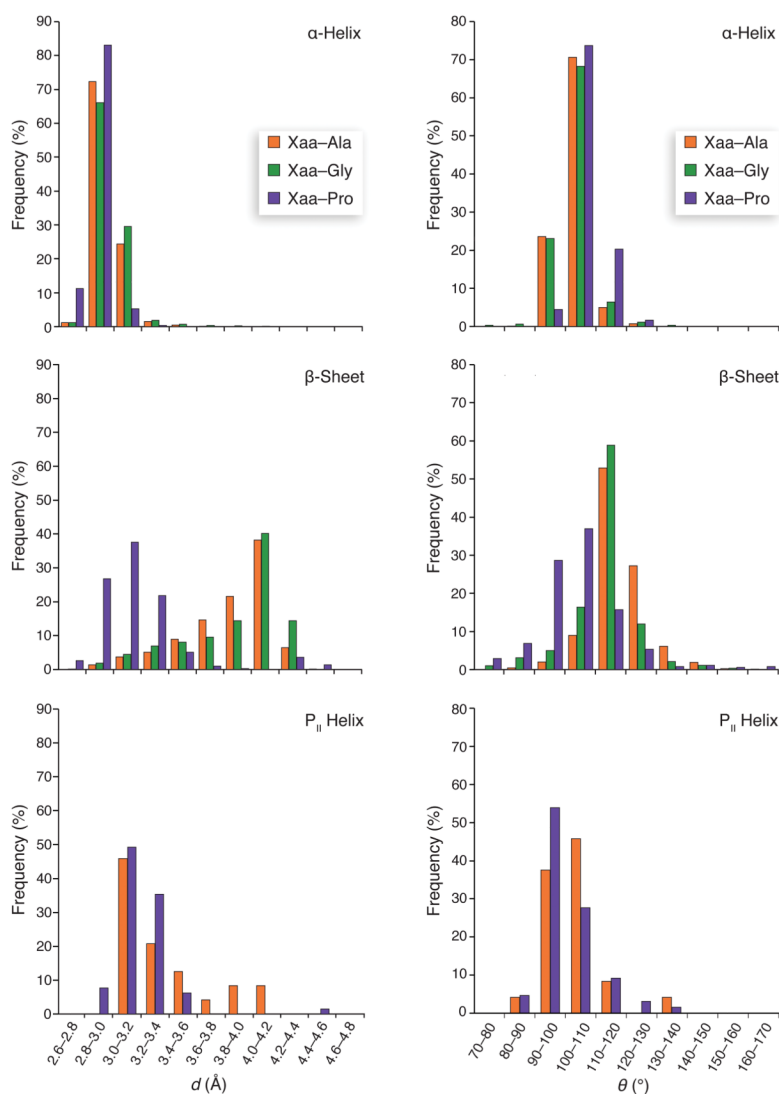


**Figure 1. Kinship of hydrogen bonds and  $n \rightarrow \pi^*$  interactions in an  $\alpha$ -helix**  
**a**,  $s$ -rich lone-pair of an amide oxygen. **b**,  $p$ -rich lone-pair of an amide oxygen. **c**,  $n_s \rightarrow \sigma^*$ : hydrogen bond in an  $\alpha$ -helix. **d**,  $n_p \rightarrow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



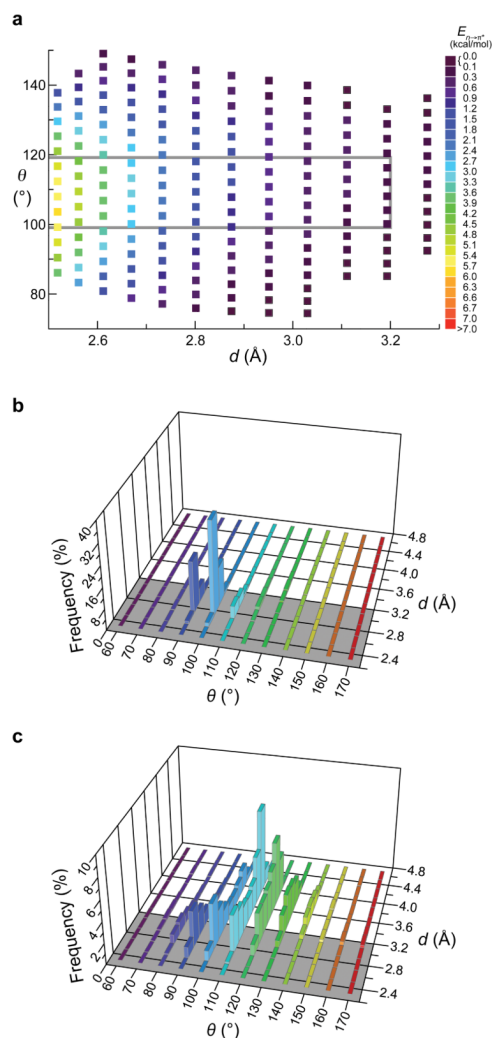
**Figure 2. Ramachandran plots of  $n \rightarrow \pi^*$  interactions**

**a**, Definitions of dihedral angles  $\phi$  ( $C'_{i-1}-N_i-C^\alpha_i-C'_i$ ) and  $\psi$  ( $N_i-C^\alpha_i-C'_i-N_{i+1}$ ), distance  $d$ , and planar angle  $\theta$ . Criteria for an  $n \rightarrow \pi^*$  interaction in the crystallographic analyses:  $d \leq 3.2 \text{ \AA}$ ;  $99^\circ \leq \theta \leq 119^\circ$ . **b**, Computational data showing the energy of an  $n \rightarrow \pi^*$  interaction for 836  $\phi$ ,  $\psi$  angles. Common secondary structures are indicated. **c**, Crystallographic data showing all residues in 1731 high-resolution structures (black) and those that meet the criteria in panel (a) (red). Nineteen of those residues had  $d \leq 2.60 \text{ \AA}$  (Table S2).



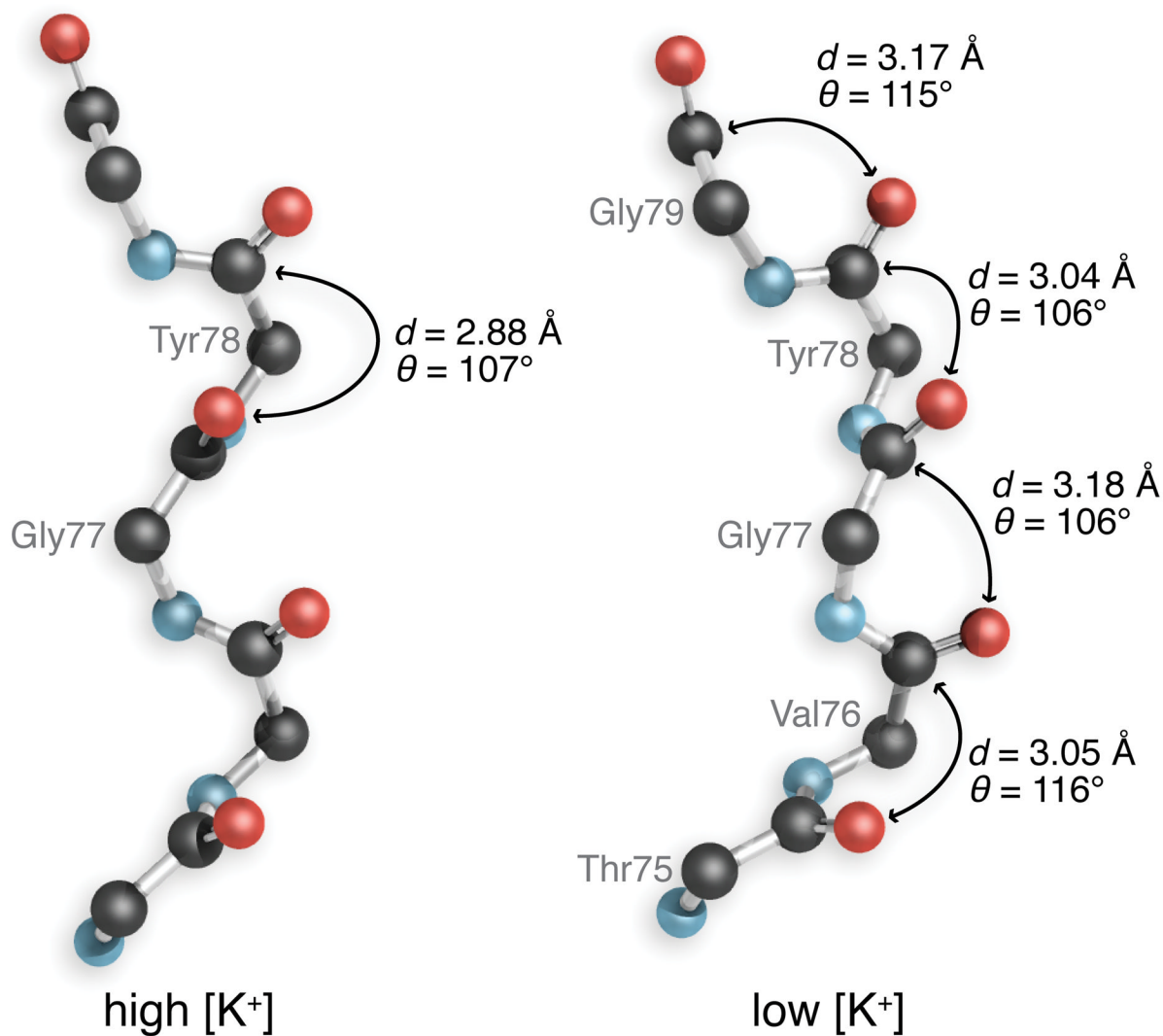
**Figure 3. Histograms of  $d$  and  $\theta$  values for Xaa-Ala, Xaa-Gly, and Xaa-Pro dipeptides in  $\alpha$ -helices,  $\beta$ -sheets, and  $P_{II}$  helices**

Mean values ( $\pm$  SD) of  $d$  and  $\theta$  were as follows.  $\alpha$ -Helix: Xaa-Ala,  $d = 2.97 \pm 0.10$  Å,  $\theta = 102.9 \pm 4.7^\circ$ ; Xaa-Gly,  $d = 3.01 \pm 0.17$  Å,  $\theta = 103.0 \pm 7.6^\circ$ ; Xaa-Pro,  $d = 2.89 \pm 0.10$  Å,  $\theta = 107.2 \pm 6.5^\circ$ .  $\beta$ -Sheet: Xaa-Ala,  $d = 3.87 \pm 0.32$  Å,  $\theta = 118.1 \pm 9.3^\circ$ ; Xaa-Gly,  $d = 3.89 \pm 0.36$  Å,  $\theta = 112.7 \pm 11.3^\circ$ ; Xaa-Pro,  $d = 3.17 \pm 0.33$  Å,  $\theta = 103.8 \pm 13.5^\circ$ .  $P_{II}$  helix: Xaa-Ala,  $d = 3.16 \pm 0.33$  Å,  $\theta = 97.7 \pm 9.5^\circ$ ; Xaa-Pro,  $d = 3.01 \pm 0.21$  Å,  $\theta = 95.6 \pm 8.7^\circ$ . 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  $\theta$  for Xaa-Pro were larger than those for Xaa-Ala and Xaa-Gly in  $\alpha$ -helices, but smaller than those for Xaa-Ala and Xaa-Gly in  $\beta$ -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.



**Figure 4. Values of  $d$  and  $\theta$  in  $n \rightarrow \pi^*$  interactions**

**a**, Energy of the  $n \rightarrow \pi^*$  interactions for values of  $d$  and  $\theta$  in the right-handed helical region of the Ramachandran plot (*i.e.*,  $-180^\circ \leq \phi \leq 0^\circ$ ;  $-180^\circ \leq \psi \leq 0^\circ$ ). The criteria in Fig. 2a are obeyed within the gray rectangle. **b**, Frequency of values of  $d$  and  $\theta$  among residue pairs (132149) in an  $\alpha$ -helix (Table S3). **c**, Frequency of values of  $d$  and  $\theta$  among residue pairs (221469) not in an  $\alpha$ -helix (Table S4). Combined, the plots in panels b and c include all residue pairs (353618) in 1731 high-resolution structures; *n.b.*, there are more data plotted here than are listed in Table 1, as the latter were culled for four-residue sequences, whereas as the data used in this figure were from residue pairs. The frequency of residue pairs with  $d$  3.2 Å (Fig. 2a) emerge from the gray plane.



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



Table 1

Potential  $n \rightarrow \pi^*$  interactions in different types of secondary structure.

Dipeptide	$\alpha$ -helix		$\beta$ -sheet		$P_{II}$ helix		Other secondary structures	
	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	11384	9367 (82%)	2786	80 (2.8%)	24	6 (25%)	185	47 (25%)
Xaa-Gly	2713	2188 (81%)	2036	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	88124	63792 (72%)	38657	1166 (3%)	227	58 (25%)	2588	427 (16%)