Published in final edited form as:
*Nat Chem Biol*. 2015 March ; 11(3): 221–228. doi:10.1038/nchembio.1739.

# Local and macroscopic electrostatic interactions in single α-helices

**Emily G. Baker**[1], **Gail J. Bartlett**[1], **Matthew P. Crump**[1], **Richard B. Sessions**[2], **Noah Linden**[3], **Charl F. J. Faul**[1], and **Derek N. Woolfson**[1,2,*]

[1]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK

[2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8 1TD, UK

[3]School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

## Abstract

The non-covalent forces that stabilise protein structures are not fully understood. One way to address this is to study equilibria between unfolded states and α-helices in peptides. For these, electrostatic forces are believed to contribute, including interactions between: side chains; the backbone and side chains; and side chains and the helix macrodipole. Here we probe these experimentally using designed peptides. We find that both terminal backbone-side chain and certain side chain-side chain interactions (*i.e.*, local effects between proximal charges, or interatomic contacts) contribute much more to helix stability than side chain-helix macrodipole electrostatics, which are believed to operate at larger distances. This has implications for current descriptions of helix stability, understanding protein folding, and the refinement of force fields for biomolecular modelling and simulations. In addition, it sheds light on the stability of rod-like structures formed by single α-helices that are common in natural proteins including non-muscle myosins.

Folded proteins are stabilised by many weak intramolecular forces between thousands of atoms, including: hydrogen bonds, van der Waals' contacts, and Coulombic interactions between formal charges (*i.e.*, salt bridges).[1] Because of the interdependence of non-covalent interactions, and the resulting cooperativity of protein folding, it is extremely difficult to disentangle contributions from these various components. For example, in terms of backbone interactions alone, $CO_i \rightarrow NH_{i+4}$ hydrogen bonds are regarded as the major factor stabilising α-helices.[2] Whilst this remains unchallenged, recent studies have identified additional contributions from $CO_i \rightarrow CO_{i+1}$ $n \rightarrow \pi^*$ interactions,[3] which can "compete" with

*To whom correspondence should be addressed: D.N.Woolfson@bristol.ac.uk.

**Author contributions**

EGB, CFJF and DNW designed the research; EGB made the synthetic peptides and performed the CD spectroscopy; MPC and EGB collected and analysed the NMR data; GJB, EGB and DNW performed the bioinformatics; EGB and RBS constructed and analysed atomistic models for the peptides; NL performed and analysed the helix-dipole calculations; EGB, GJB and DNW wrote the paper. All authors reviewed and contributed to the manuscript.

The authors declare no competing financial interests.

**Dedication:** DNW would like to dedicate this paper to the memory of Prof. Dudley H. William FRS, an inspirational scientist, mentor and person.


Europe PMC Funders Author Manuscripts

traditional hydrogen bonding.[4] Thus, it is important to develop continually, and to challenge our understanding of the fundamental forces underpinning protein folding, and of biomolecular interactions in water more generally. Improved understanding will lead to better modelling, prediction and design of biomolecular structures and assemblies *ab initio*, and with these many improved applications in the basic and applied biosciences will follow.

One way forward is to establish straightforward experimental models that isolate certain interactions, or reduce the complexity of their contexts. An example is the study of helix⇆coil transitions in short peptides. This began with observations that certain peptide fragments of natural proteins exhibit considerable α-helicity,[5,6] and developed to the *de novo* design of model peptides that behave systematically with rational changes to sequence.[7] The latter led to the dissection of various non-covalent interactions, and to sequence-to-structure relationships also apparent in fully folded natural proteins, including: α-helical propensity scales;[8-10] specific backbone-side chain interactions;[11] and pairwise and higher-order side-chain interactions.[12-15]

To balance the configurational entropy lost on folding into an α-helix, a number of electrostatic forces probably contribute to favourable enthalpies of helix folding, including: backbone $CO_i \rightarrow NH_{i+4}$ hydrogen bonds; hydrogen bonds and salt bridges between side chains spaced i➔i+3 and i➔i+4;[16] backbone-side chain interactions at both helical termini;[17] and through-space electrostatic interactions between charged side chains and the so-called helix macrodipole[18] (Fig. 1). The latter is believed to arise from the alignment of individual dipoles of backbone peptide units parallel to the major axis of the α-helix. This has local consequences: it results in partial positive and negative charges at the *N*- and *C*-termini, respectively,[19,20] which leads to side-chain preferences and ligand-binding phenomena at these sites.[21] However, the evidence and roles for the macroscopic effects are less clear, and the energetic contributions to structure and stability by both surface-exposed salt bridges[22-24] and the α-helix macrodipole[25] remain unresolved.

Most recently, the field has come full circle with observations of natural long single α-helices that span 30 – 200 residues. These self-stabilising structures do not require tertiary or quaternary interactions to persist. Originally noted in the calmodulin-binding protein caldesmon[26] and the non-muscle myosins VI,[27] VII,[28] and X[29], these single α-helices (SAHs) are predicted to occur widely.[30-32] Their hallmark is alternating blocks of four negatively charged glutamic acid (Glu, E) and four positively charged arginine (Arg, R) or lysine (Lys, K) residues.[33] They are also dubbed charged single α-helix (CSAH) domains,[31] or E-R/K α-helical motifs[34] and are proposed to form rigid rods or spacers to separate other protein domains.

Given this revived interest in single α-helices, but the incomplete understanding of the forces that stabilise them, we designed a series of peptides comprising predominantly Glu and Lys to prise apart different contributions to helix stability. Our first designs were consistent with preceding reports. However, further iterations conflicted with current thinking for the contributions from both salt-bridge and macrodipole interactions. Specifically, certain side-chain arrangements between Glu and Lys are favoured over others,

and these dominate any manifestation of the macroscopic helix dipole; indeed, for these systems, there is no need to invoke macroscopic effects and local descriptions suffice.

## RESULTS

### Peptide design – iteration 1

We followed foregoing studies of *de novo*[35] and natural SAHs[29] to design two highly helical peptides, **$(E_4K_4)_3$** and **$(K_4E_4)_3$** (Table 1). In these, alternating blocks of four Glu and four Lys, which both have favourable α-helical propensities,[10] were used to match the 3.6-residues per turn of the α-helix. This maximises potentially favourable i➔i+4 E➔K and K➔E inter-side-chain salt bridges (Fig. 1), whilst minimising unfavourable inter-side-chain electrostatic interactions (Supplementary Results, Supplementary Table 1). *N.B.*, in these two designs, both i➔i+4 E➔K and K➔E interactions are possible, but one potentially outnumbers the other by 3:2 in each case. In addition, all four types of side-chain-terminus interaction are represented, *e.g.* blocks of four Glu or Lys at both the *N*- and *C*- termini of the peptides (Fig. 1). We used **$(E_2K_2)_6$** as a control peptide of the same length, which maximises repulsive i➔i+4 E➔E and K➔K pairs (Supplementary Table 1).[35]

Circular dichroism (CD) spectroscopy showed that **$(E_2K_2)_6$** was largely unfolded; and that both **$(E_4K_4)_3$** and **$(K_4E_4)_3$** had high mean residue ellipticities at 222 nm ($MRE_{222}$), which translated to 74% and 62% helix, respectively (Fig. 2, Supplementary Fig. 28a&b and Table 1). **$(E_4K_4)_3$** was expected to be more helical than **$(K_4E_4)_3$** because: it better matches the formal charges of the side chains and the partial charges of backbone in the *N*- and *C*-terminal turns (Fig. 1); and, the $E_i$➔$K_{i+4}$ arrangement counters the helix macrodipole better. This continued over a range of peptide lengths for **$(E_4K_4)_n$** and **$(K_4E_4)_n$**, where n = 1 – 4, with the former being the more helical in each case (Fig. 2c).

As secondary structure can be stabilised by peptide association, we used sedimentation equilibrium experiments to determine peptide oligomeric states. These indicated that both **$(E_4K_4)_3$** and **$(K_4E_4)_3$** were monomeric in solution (Supplementary Fig. 30 and Supplementary Table 2).

Next, we used high-resolution $^1$H-NMR NOESY and TOCSY experiments to probe the structure in **$(E_4K_4)_3$** and **$(K_4E_4)_3$** in detail (Fig. 3 and Supplementary Figs. 31-34). Sequence-specific assignments were made for all backbone HN resonances, and the majority of the Hα resonances for both peptides. From Δδ and NOE measurements, **$(E_4K_4)_3$** was helical throughout with contiguous, sequential $HN_i$-$HN_{i+1}$ NOEs from Gly1 through Trp27 (Fig. 3g). Whereas, for **$(K_4E_4)_3$** the *N*-terminal residues Lys2 – Lys4 appeared disordered (Fig. 3h); otherwise, the helical structure continued through the sequence from Lys5 to Trp27. This fraying in **$(K_4E_4)_3$** seems best explained by the mismatched charge of the side chains of Lys2, 3 and 4 with the $\delta^+$ *N*-terminus; although the *C*- terminus is folded despite the negatively charged Glu residues at this $\delta^-$ end.

### Peptide design – iteration 2

To reduce complications from end effects—*i.e.*, local electrostatic interactions between charged side chains and the partially charged termini—we synthesised peptides in which

$(E_4K_4)_3$ and $(K_4E_4)_3$ were flanked at one or both sides by tetra-alanine ($Ala_4$, $A_4$) blocks (Table 1). Alanine was chosen for its high helix propensity,[10] small size and charge neutrality. Although we anticipated some influence of these blocks on overall helicity of the peptides, we reasoned that these properties would best serve the purpose of homogenising end effects. The resulting six peptides, **$A_4(E_4K_4)_3$**, **$(E_4K_4)_3A_4$**, **$A_4(E_4K_4)_3A_4$**, **$A_4(K_4E_4)_3$**, **$(K_4E_4)_3A_4$** and **$A_4(K_4E_4)_3A_4$**, showed the following responses to the flanks in CD spectroscopy (Fig. 2, and Supplementary Figs. 28c – f and 29c – f).

First, the percent helix in all of the peptides increased compared with the parents (Fig. 2d). This is almost certainly due in part to the increased lengths of the peptides and the choice of Ala as the flanking residues. It is also fully consistent with the $Ala_4$ flanks limiting the terminal fraying/unfolding of the peptides.[36-38] Second, changes in helicity were more pronounced for the $(K_4E_4)_3$-based peptides. This seems best explained simply by the removal of fraying due to the block of unfavourable positive charge from the helical *N*-termini. Consistent with this, within the $(K_4E_4)_3$ series the favourable impact of the flanks was largest at the *N*-terminus.

However, a final response was unexpected, and is less readily explained by current thinking. Comparing the overall helicities of the parent and doubly flanked peptides **$(E_4K_4)_3$** > **$(K_4E_4)_3$**, but **$A_4(K_4E_4)_3A_4$** > **$A_4(E_4K_4)_3A_4$** (Table 1 and Fig. 2). This runs contrary to most reports of side chain-side chain interactions between Glu and Lys in the context of the helix macrodipole. Current understanding is that $E_i$➔$K_{i+4}$ should be more stabilising than the reverse combination because the side chains align energetically favourably with the helix macrodipole.[7,39] However, we find the opposite: **$A_4(K_4E_4)_3A_4$**, which has more potential $K_i$➔$E_{i+4}$ pairings than $E_i$➔$K_{i+4}$, is the more helical.

Further NMR measurements on **$A_4(K_4E_4)_3A_4$** confirmed contiguous sequential $HN_i$-to-$HN_{i+1}$ NOEs from Ala3 through Trp35 (Supplementary Fig. 34). Thus, the entire central $(K_4E_4)_3$ block is α-helical. Moreover, length-dependence experiments for the flanked peptides revealed that the $(K_4E_4)_n$ blocks were consistently more helical than $(E_4K_4)_n$ blocks (Fig. 2c). *N.B.*, sedimentation-equilibrium studies confirmed that both **$A_4(E_4K_4)_3A_4$** and **$A_4(K_4E_4)_3A_4$** were monomeric in solution (Supplementary Fig. 30 and Supplementary Table 2).

### $K_i$➔$E_{i+4}$ salt bridges are enthalpically better

The additional stabilising effect of $K_i$➔$E_{i+4}$ pairings compared with $E_i$➔$K_{i+4}$ was also evident in thermal unfolding followed by CD spectroscopy. The $E_4K_4$-based peptides showed broad and approximately linear transitions from predominantly helical to gradually less helical states, including the most folded **$A_4(E_4K_4)_3A_4$** (Fig. 2b and Supplementary Figs. 29c&e). This is typical of short *de novo* α-helical peptides.[7] By contrast, the $K_4E_4$-based peptides had sharper more-sigmoidal transitions (Fig. 2b and Supplementary Figs. 29d&f), and some had measurable midpoints ($T_M$) with **$A_4(K_4E_4)_3A_4$** having the highest. Though there are examples of such behaviour,[40] it is unusual and suggests cooperativity in the transitions, and that more enthalpically favourable interactions are being made in the folded state of this peptide.

We fitted the thermal unfolding data to the Gibbs-Helmholtz equation (Fig. 2b and Supplementary Fig. 29). This assumed $\Delta C_p = 0$, and that helix⇆coil transitions are two-state processes. The latter is almost certainly incorrect as the "folded states" are likely ensembles of unfolded, folded and partly folded species;[40] however, we found no evidence for alternative "unfolded" states by CD spectroscopy at least, *i.e.* no polyproline helices or β-structure. Caveats aside, the analysis is valuable because the gradients of the fits reflect the enthalpy change accompanying unfolding. For our designed peptides, the enthalpies of folding were all exothermic (Supplementary Table 7). Moreover, the enthalpies for the $(K_4E_4)_n$-based peptides were higher than for the $(E_4K_4)_n$ analogues. This suggests that $K_i$➔$E_{i+4}$ salt bridges are somehow enthalpically better than $E_i$➔$K_{i+4}$ interactions, and cooperate better to induce and maintain structure in **$A_4(K_4E_4)_3A_4$**.

## Helical propensities and salt-bridges in folded proteins

There have been extensive analyses of residue preferences and pairings in α-helices,[12,41,42] but our results raised issues to be addressed afresh. Therefore, we analysed the conformations and interactions of Glu and Lys side chains in α-helices from the PDB (Fig. 4).[43]

First, we tested for specific side-chain-backbone interactions at the helical termini. Consistent with previous studies[9,36-38] and our experiments, this analysis gave favourable propensities for Glu at helical *N*-termini and Lys at the *C*-termini of 1.24 and 1.26, respectively (Supplementary Table 8). However, using HBPLUS,[44] we found that only small fractions (9% for Glu at *N*-termini, and 2.5% for Lys at *C*-termini) of these made side-chain-backbone hydrogen bonds (Supplementary Fig. 36). In the context of the many potential hydrogen bonds with water, this suggests strongly that the experimental and statistical preferences for Glu at *N*-termini are best attributed to proximal through-space electrostatic interactions between the γ-carboxylate and the $\delta^+$ ends of nearby peptide-bond dipoles, rather than by specific atomic contacts; and the preferences for Lys at *C*-termini by similar interactions between the ε-amino moiety and the $\delta^-$ charge.

Regarding potential $E_i$➔$K_{i+4}$ and $K_i$➔$E_{i+4}$ salt bridges, we found that more $K_i$➔$E_{i+4}$ pairs (21%) formed salt bridges than $E_i$➔$K_{i+4}$ (11%); and that the salt-bridged $K_i$➔$E_{i+4}$ pairs were more structurally conserved, with a lower RMSD value of 1.08 Å, *vs.* 1.79 Å for those made by $E_i$➔$K_{i+4}$; calculated over all side-chain atoms, (Supplementary Fig. 37). To understand this, we compared the $\chi_1,\chi_2$ distributions for Glu-Lys salt bridges (Fig. 4) and all possible $\chi_1,\chi_2$ combinations for Glu and Lys in i➔i+3 and i➔i+4 pairings in α-helices (Supplementary Table 9).

Glu has three dominant $\chi_1,\chi_2$ combinations: *gauche*$^-$,*trans* ($g^-$ *t*, 36%); *trans*,*trans* (*tt*, 25%); and *gauche*$^-$,*gauche*$^-$ ($g^-g^-$, 19%), (Figs. 4a&d). However, when paired with Lys only certain combinations of these led to salt bridges: For $E_i$➔$K_{i+4}$ pairs, salt bridges were only made with Glu in the second-favoured conformer (*tt*) or a fourth minor conformation ($tg^+$, 10%), (Fig. 4a,e&f and Supplementary Table 9). Having to fix the side chain in these less-favoured rotamers incurs an entropic penalty. Nevertheless, the more-flexible side chain of Lys complemented these Glu conformers to account for 92% (23/25) of the $E_i$➔$K_{i+4}$ salt-bridge interactions (Figs. 4c,e&f). By contrast, 75% of the salt-bridged $K_i$➔$E_{i+4}$ pairs had

the same paired conformation, with Glu in its preferred $\chi_1,\chi_2$ conformation and Lys in its second-favoured conformer (*tt*, 32%; which was only marginally less preferred than $g^- t$ (39%)), (Figs. 4b,d,i). Hence, the entropic penalty of freezing out this conformer is near the minimum possible.

Thus, compared with $E_i$➔$K_{i+4}$ salt bridges, the preferred $K_i$➔$E_{i+4}$ salt bridges appear to sample fewer conformations and use more-favourable $\chi_1\chi_2$ pairs for each residue. We posit that this leads to $K_i$➔$E_{i+4}$ interactions that are enthalpically better, because the side chains are predisposed in conformations to facilitate salt bridges. This is not to say that sequences dominated by $E_i$➔$K_{i+4}$ arrangements will necessarily be less helical; from our data and the literature this is patently not the case. However, as discussed below, the mechanism of helix stability is subtly different.

## Uncoupling ($E_4K_4$) and ($K_4E_4$) blocks

The peptide designs described above can make both $E_i$➔$K_{i+4}$ and $K_i$➔$E_{i+4}$ interactions. Thus, it is not possible to ascribe absolutely either the percent-helix or thermal unfolding properties to one i➔i+4 interaction type. To address this, we designed further sequences with isolated $E_4K_4$ or $K_4E_4$ blocks flanked by $A_4$ blocks; namely, **$A_4(E_4K_4)A_4$**, **$A_4(K_4E_4)A_4$**, **$A_4(E_4K_4)A_4(E_4K_4)A_4$** and **$A_4(K_4E_4)A_4(K_4E_4)A_4$** (Table 1). In these, only one type of i➔i+4 E➔K or K➔E interaction can be made. The shorter peptides were surprisingly helical (Fig. 2c, Table 1, and Supplementary Fig. 28c&d), with **$A_4(K_4E_4)A_4$** having the higher α-helicity. Moreover, the thermal-unfolding curve for this peptide was sharper and more sigmoidal than for **$A_4(E_4K_4)A_4$** (Supplementary Fig. 29c&d). The longer peptides with the additional, but uncoupled, $E_4K_4$ or $K_4E_4$ blocks were highly helical at 83% and 97%, respectively. (*N.B.*, some of the structure is likely due to the central $Ala_4$ block). Both peptides had sigmoidal thermal-denaturation curves with measurable midpoints, but the K➔E arrangement had the sharper transition and the higher $T_M$ (Table 1 and Supplementary Fig. 29g). This confirms that $K_i$➔$E_{i+4}$ potential salt bridges contribute more to helical stability than $E_i$➔$K_{i+4}$ pairs, despite the apparent misalignment with the helix macrodipole.

## Consideration of E/$K_i$➔K/$E_{i+3}$ interactions

Our designs could also make i➔i+3 interactions (Supplementary Table 1). Therefore, we examined such pairs in α-helices of the PDB. Interestingly, more $E_i$➔$K_{i+3}$ pairs (23.4%) made salt bridges than the $K_i$➔$E_{i+3}$ pairs (15.6%) (Supplementary Table 8b). Also, for the $E_i$➔ $K_{i+3}$ pairs there was one dominant rotamer combination that accounted for 77% of $E_i$➔$K_{i+3}$ salt bridges made; whilst, for $K_i$➔$E_{i+3}$ pairs multiple conformers made salt bridges, with two accounting for 60% of the interactions (Supplementary Fig. 38). In addition, we found that $E_i$➔$K_{i+3}$ salt bridges were more structurally conserved, with lower a RMSD (1.09 Å) than for $K_i$➔$E_{i+3}$ salt bridges (2.27 Å) (Supplementary Fig. 37). These trends are the opposite of those observed for the i➔i+4 pairs, which suggest that $E_i$➔$K_{i+3}$ salt bridges might be more helix stabilising than $K_i$➔$E_{i+3}$.

To test this, we synthesised **$A_4(E_3K_3)_4A_4$** and **$A_4(K_3E_3)_4A_4$**. Both peptides were highly helical and showed clear thermal unfolding transitions (Table 1 and Supplementary Figs.

39b&d). However, they were less helical than the $E_4K_4$ and $K_4E_4$ analogues, **$A_4(E_4K_4)_3A_4$** and **$A_4(K_4E_4)_3A_4$**, illustrating that i➔i+3 interactions are less favourable than i➔i+4. In addition, and as predicted above, the trend was switched compared with $E_4K_4$ and $K_4E_4$, *i.e.*, **$A_4(E_3K_3)_4A_4$** was more helical and had a higher $T_M$ than **$A_4(K_3E_3)_4A_4$** (Table 1). Nonetheless, and perhaps counter-intuitively, the folding of **$A_4(K_3E_3)_4A_4$** was the more exothermic (Supplementary Table 7). We suggest that this is because **$A_4(K_3E_3)_4A_4$** can make more of the higher-enthalpy K➔E i➔i+4 interactions (Supplementary Table 1).

## The role of the helix macrodipole

Whilst $K_i$➔$E_{i+4}$ salt bridges make enthalpically better interactions and lead to more-folded α-helical peptides than $E_i$➔$K_{i+4}$, this arrangement of positively charged lysine blocks *N*-terminal to negatively charged glutamate blocks is predicted to be unfavourable with respect to the helix macrodipole. Therefore, we calculated the electrostatic potential ($V_E$) of a model α-helix in vacuum, and at varying distances (x) from the central helix axis (z) by approximating partial dipole charges as point charges (Fig. 5a and Supplementary Fig. 40). In this model, the electrostatic potential for an individual peptide dipole was ≈ ± 0.7 V (Fig. 5b). For x = 0 in an extended helix, *i.e.* directly along its axis, we found that the central peptide-bond dipoles cancelled, leaving a negligible macrodipole at the termini ($V_E \approx \pm 1.5$ V, Fig. 5c). At x = 2.5 Å, *i.e.* where the majority of backbone atoms lie, the residual macrodipole is overshadowed by local electrostatics from these atoms, with $V_E \approx \pm 10$ V (Fig. 5d). At x = 5 Å, *i.e.* approximately at the termini of side chains, the effect is similar to x = 0, but with the macrodipole diminished further ($V_E \approx \pm 0.6$ V, Fig. 5e). Indeed, at this distance the macrodipole is approximated by partial point charges at the termini of the helix (Fig. 5f). Thus, and even in the absence of solvation effects, the influence of the macrodipole on the stability of central K➔E interactions is predicted to be minimal.

## Correlation with helix-prediction programs

Considerable effort has gone into capturing the data from helix⇆coil transitions in computational methods to predict helix content from peptide sequence, with AGADIR probably being the most popular.[45] We made AGADIR predictions for both the percentage helix and thermal unfolding behaviours of our designs, and compared these with our experimental data (Supplementary Figs. 41&42). At first sight, the predictions appeared good. However, there were some deviations from the observed values, particularly for the longer and $Ala_4$-flanked variants. Moreover, the thermal unfolding data were not modelled well by AGADIR, which tended to over predict stabilities. This paper provides new data that could be modelled to improve these predictions and to aid peptide designs.

## Biological context

Our observations that certain side-chain arrangements are more stabilising than others, *i.e.* $K_i$➔$E_{i+4}$ > $E_i$➔$K_{i+4}$ and $E_i$➔$K_{i+3}$ > $K_i$➔$E_{i+3}$, may help understand the behaviour of SAHs. To examine this in the natural sequences, we counted all of the potential D/E➔K/R and K/R➔D/E pairs spaced i➔i+3/4 in sequences from non-homologous proteins assigned as SAHs from CD spectroscopy and/or electron microscopy; we then normalised the data with expected rates (Supplementary Fig. 43). Firstly, as anticipated,[10] we found that aspartic acid

(D) hardly featured in the sequences or pairings. Secondly, the normalised frequencies of oppositely charged pairs were much greater than those for like-charge pairs, by an average factor of 7 for i➔i+4, and 3 for i➔i+3. Third, except for the E➔R combinations, which had similar observed/expected ratios, the propensities of oppositely charged i➔i+4 pairs were greater than those for the corresponding i➔i+3 pairs. Fourthly, and consistent with our model-peptide studies, there were more $R/K_i \rightarrow E_{i+4}$ pairs than the reverse orientation. Fifthly, the data for i➔i+3 pairings in natural SAHs were less clear-cut, although the pairs with overall highest propensity were $E_i \rightarrow R_{i+3}$, and $K_i \rightarrow E_{i+3}$. Finally, there was a consistent slight preference for Arg-based pairings over those involving Lys (apart from $K_i \rightarrow E_{i+3} > R_i \rightarrow E_{i+3}$), which may reflect arginine's potential to participate simultaneously in two salt bridges.[15]

## DISCUSSION

We have found that: (1) short peptides of ≈20 – 30 residues comprising alternating blocks of $Glu_4$ and $Lys_4$ formed highly α–helical structures; (2) flanking these with $Ala_4$ blocks increased helical content and revealed context-dependent phenomena; for example, (3) Glu residues in the *N*-terminal turn strongly promoted α–helix formation, and Lys in these positions disrupted it; whereas, (4) similar effects at the *C*-terminus were less pronounced; (5) central $K_i \rightarrow E_{i+4}$ arrangements made enthalpically better side-chain salt-bridge interactions than $E_i \rightarrow K_{i+4}$ pairs; as a result, (6) although central $E_i \rightarrow K_{i+4}$ and $K_i \rightarrow E_{i+4}$ arrangements both favoured helix formation, the resulting helical structures were different in nature, particularly in how they responded to temperature; and (7) i➔i+3 potential salt-bridge interactions were weaker than those made by i➔i+4 arrangements. We place these results in context as follows.

### 1. Local interactions between side-chain and backbone moieties at helical termini

It is well established, in peptides[9,36-38] and proteins,[11] that certain side chains interact favourably with polar backbone groups and/or partial charges at the *N*- and *C*-termini of α-helices.[25] Our first peptide designs confirmed this: peptides with *N*-terminal blocks of lysine were less folded and more frayed than those with blocks of glutamate at that end. Curiously, however, in protein structures of the PDB, we found very few examples of formal hydrogen-bonded or other atom-atom contacts between side chains and backbone atoms at the termini. This indicates that such interactions are ephemeral, if there at all, and that more-general electrostatic interactions are at play between side chains and the partial charge at the *N*-terminus. Moreover, for our *de novo* peptides, these effects were much more pronounced at the *N*-terminus than at the *C*-terminus. This is readily understood as side chains emerge from the helical backbone and point towards the *N*-terminus (Fig. 1).

### 2. Local interactions between side-chain groups in the central regions of helices

To remove the end effects, we flanked the Glu/Lys-based peptide sequences with tetra-alanine units. This led to a surprising discovery: i➔i+4 K➔E pairs stabilised α-helical structure more than the reversed E➔K orientation. This result was consistent across all of our peptides, and was supported by the analysis of the PDB. The latter revealed that $K_i \rightarrow E_{i+4}$ pairs make more formal salt-bridges than $E_i \rightarrow K_{i+4}$ pairs. Moreover, the

mechanisms through which such salt bridges are formed appears different in the two cases: whereas, those in $K_i \rightarrow E_{i+4}$ pairs employ a small number of sterically favoured side-chain conformations, salt-bridges made by $E_i \rightarrow K_{i+4}$ arrangements use many more and less-favoured conformations. We argue, that this leads to enthalpically better $K_i \rightarrow E_{i+4}$ salt bridges, and entropically favoured $E_i \rightarrow K_{i+4}$ interactions. This was manifest in the thermal unfolding transitions of different peptides: those with more $K_i \rightarrow E_{i+4}$ arrangements had sharper, and in some cases sigmoidal, thermal unfolding curves; whereas, peptides dominated by $E_i \rightarrow K_{i+4}$ pairs showed broader, less-defined transitions.

These findings are noteworthy as there is disagreement in the literature over which pairing is the more stabilising, with all possible stances being adopted: $E_i \rightarrow K_{i+4} > K_i \rightarrow E_{i+4}$,[7] *vice versa*,[12] and that the two interactions are equivalent.[16] We believe that our combined experimental and bioinformatics studies help resolve this.

### 3. Non-local interactions between side chains and the macroscopic helix dipole

Our finding that $K_i \rightarrow E_{i+4}$ pairs are more stabilising than $E_i \rightarrow K_{i+4}$ arrangements has a further implication. This result is contrary to that expected from consideration of the macroscopic helix dipole alone.[7] Organising $K_i \rightarrow E_{i+4}$ (+ve −ve) pairs in an α-helical conformation should require work against the helix macrodipole, which runs from a partially positively charged *N*-terminus to a partially negatively charged *C*-terminus; *i.e.*, it will cost free energy and disfavour the helical state. In contrast, folding of $E_i \rightarrow K_{i+4}$ pairs should counter the dipole and therefore be favourable. That we do not see this effect—indeed, we observe the opposite—requires explaining.

The most straightforward explanation is that the macrodipole does not operate and is attenuated out where the charged moieties of the side chains reside. That is, even in the absence of solvation and dielectric effects, the charged moieties of side chains such as glutamic acid and lysine are not expected to make significant Coulombic interactions with the macrodipole. That said, they are expected to interact with the terminal partial charges arising from the half-dipoles organised at the *N*- and *C*-termini. Thus the local, terminal effects of the helix dipole are not in doubt; indeed, these are well established, and our results and conclusion in point *1* above support this. What we do question, however, is the need to invoke helix-dipole arguments beyond this, *i.e.*, any macrodipole effect of the α-helix; we argue that the local description of the helix dipole, and how this affects side-chain preferences, binding of co-factors in proteins and so on, should be sufficient for explanations of experimental data and theoretical treatments.

In addition to contributing to fundamental understanding of helix stability, we envisage that our systematic set of designed peptides, the inventory of potential interactions within these, and the experimental data that we provide on helix content and thermal stability, will provide a useful resource to computational chemists and biochemists developing empirical and theoretical methods for analysing and predicting helix-forming peptides and protein fragments, and force fields for protein modelling more generally. Finally, in the context of synthetic biology, the fully folded peptides that we present may find use as linkers and spacers of tuneable length and thermal unfolding behaviour to combine protein domains in defined ways.[46,47]

# METHODS

## General methods

Rink amide ChemMatrix™ resin was purchased from PCAS Biomatrix Inc (St-Jean-sur-Richelieu, Canada). Fmoc protected amino acids, peptide grade DMF and HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) were obtained from ACTG Bioproducts (Hessle, UK). All other reagents were acquired from Fisher Scientific (Loughborough, UK). Peptide concentrations were determined by absorbance at 280 nm using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and employing Trp $\varepsilon_{280}$ = 5690 $M^{-1}$ $cm^{-1}$ and Tyr $\varepsilon_{280}$ = 1280 $M^{-1}$ $cm^{-1}$.

## Peptides synthesis

Peptides were synthesised on a 0.1 mmole scale using a Liberty CEM microwave-assisted automated synthesiser (CEM corporation, Mathews, NC, U.S.A.) with standard Fmoc chemistry (deprotection of Fmoc using 20% piperidine in DMF) and HBTU coupling (0.45 M HBTU as activator, 1M DIPEA in DMF as activator base) on a rink amide ChemMatrix™ resin. Each coupling and deprotection step was performed twice to minimise amino-acid deletions from the sequence, in addition to extra washings between steps.

Acetylations of *N*-termini were performed in DMF using 0.25 mL acetic anhydride and 0.5 mL pyridine. The resin was washed with DCM before peptides were cleaved by treatment with TFA (trifluoroacetic acid) / triisopropylsilane / water (90:5:5 vol%) for 3 hours under agitation. The cleavage mixtures were extracted by filtration and the resin rinsed with a further 5 mL of TFA before the total volume obtained was reduced to <5 mL under a flow of nitrogen. Crude peptides were precipitated in cold diethyl ether (~50 mL) and centrifuged to produce a pellet, which was then dissolved in 50:50(vol%) acetonitrile:water and freeze-dried to yield a fine white solid.

## Peptide purification

Peptides were purified by reverse phase HPLC using a JASCO chromatography system with a KYA TECH HiQ C18 column (150 μm × 10 cm) and linear gradients of various buffers depending on the peptide sequence. **$(E_4K_4)_4$**, **$(K_4E_4)_4$**, **$A_4(E_4K_4)_3A_4$**, **$A_4(K_4E_4)_3A_4$**, **$A_4(E_4K_4)_3$**, **$(E_4K_4)_3A_4$**, **$A_4(K_4E_4)_3$**, **$(K_4E_4)_3A_4$**, **$A_4(E_4K_4)A_4(E_4K_4)A_4$**, **$A_4(K_4E_4)A_4(K_4E_4)A_4$**, **$A_4(E_3K_4)_3A_4$** and **$A_4(K_3E_3)_4A_4$** were purified using linear gradients of water (buffer A, 13 mM ammonium acetate) and 20:80 (vol%) water:acetonitrile (buffer B, 13 mM ammonium acetate). Suitable fractions were freeze-dried and re-purified using linear gradients of water (buffer A, 0.1 vol% TFA) and acetonitrile (buffer B, 0.1 vol% TFA). All other peptides were purified directly using linear gradients of water (buffer A, 0.1 vol% TFA) and acetonitrile (buffer B, 0.1 vol% acetonitrile).

## Peptide characterisation

Peptide identities were confirmed by mass spectrometry (matrix: α-cyano-4-hydroxycinnamic acid) using an Applied Biosystems 4700 Proteomics Analyser MALDI-TOF instrument in reflector mode. Linear mode was used for **$(K_4E_4)_4$**. Peptide purity was

confirmed using a JASCO chromatography system with analytical C18 (Phenomenex Kinetex. 5 μm, 100 × 4.6 mm) or C8 (GraceVydac 5 μm, 250 × 4.6 mm) HPLC columns to >95% purity for all peptides.

## Circular dichroism (CD) spectroscopy

CD spectra and thermal denaturation curves were recorded once for each peptide at 100 μM peptide concentration, pH 7.4 in phosphate buffered saline (PBS), comprising $Na_2HPO_4$ (8.2 mM), $KH_2PO_4$ (1.8 mM), NaCl (137 mM) and KCl (2.7 mM). Data were recorded using a JASCO-815 spectropolorimeter fitted with a Peltier temperature controller and baseline corrected. Spectra were recorded at 5 °C from 260 – 190 nm, with a data pitch and bandwidth of 1 nm, and a scan speed of 100 nm/min. Fraction helix (%) = $100(([\theta]_{222} - [\theta]_{coil}) / (-42{,}500(1-(3/n) - [\theta]_{coil}))$. Where $[\theta]_{coil} = 640–45T = 415$ deg $cm^2$ $dmol^{-1}$ $res^{-1}$ at 5 °C, and n is the number of peptide bonds including the *N*-terminal acetyl.[48,49] Thermal denaturation curves were obtained by heating from 0 – 85 °C at 40 °C/hr and monitoring the absorbance at 222 nm over 1 °C increments (16 sec delay, and 16 sec response times), with CD spectra being taken at 5 °C intervals. The midpoints of thermal transitions ($T_M$) were taken as the maximum of the first derivative of the thermal denaturation curves. Thermal unfolding curves were also fitted to the Gibbs-Helmholtz equation to give the calculated mid-point ($T_M$) and enthalpies (ΔH) of folding. These fits assumed $\Delta C_p = 0$: and used the following equations in Sigma Plot to fit the observed MRE ($MRE_{obs}$) to the predicted MRE ($MRE_{pred}$) as a function of temperature:[50] (1) $T = T_{(°C)}+273.15$ and $T_M = T_{M(°C)}+273.15$, to convert temperatures to Kelvin; (2) $K = \exp((\Delta H/(1.987*T))*((T/T_m)-1))$, to calculate the equilibrium constant (*K*) at any temperature; (3) $F = K/(1+K)$, to calculate fractions folded (*F*) at any temperature; and (4) $MRE_{pred} = ((MRE_{hel}-MRE_{unf})*F)+MRE_{unf}$, where $MRE_{hel}$ = MRE of 100% helical peptide, and $MRE_{unf}$ = MRE of unfolded peptide. Starting parameters were as follows: $\Delta H = -20$ kcal $mol^{-1}$; $T_m = 30$ °C, $MRE_{hel} = -32{,}000$ deg $cm^2$ $dmol^{-1}$ $res^{-1}$, and $MRE_{unf} = -5000$ deg $cm^2$ $dmol^{-1}$ $res^{-1}$.

## Analytical Ultracentrifugation

AUC sedimentation-equilibrium experiments were conducted once for each of: **$(E_4K_4)_3$**, **$(K_4E_4)_3$**, **$A_4(E_4K_4)_3A_4$**, and **$A_4(K_4E_4)_3A_4$** at 88 μM peptide concentration (110 μL) in PBS at 20 °C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti rotor, with a cell comprising an aluminium centre piece and sapphire windows. The reference channel contained 120 μL of PBS buffer. The samples were centrifuged at speeds ranging 40 – 60 krpm, and the absorbance recorded across the cell at a radial distance of 5.8 – 7.3 cm. The absorbance was recorded after 8 hrs at each speed, followed by another scan after 1 hr to check the sample had reached equilibrium before moving onto the next speed. Data were fitted to single-ideal species using Ultrascan[51] and 99% confidence limits obtained by Monte Carlo analysis of the fits.

## NMR spectroscopy

Each of **$(E_4K_4)_3$**, **$(K_4E_4)_3$**, **$(K_4E_4)_4$** and **$A_4(K_4E_4)_3A_4$** were prepared at 1 mM peptide concentration in PBS, and the pH adjusted to pH 7.4 using 10 mM NaOH. Samples were then lyophilised and reconstituted in the appropriate volume of water/$D_2O$ (90:10 vol%) to

give 1 mM of peptide at pH 7.4. 330 μL of each sample was measured in a $D_2O$-matched Shigemi tube. High-resolution 2D NOESY and TOCSY spectra were recorded once for **$(E_4K_4)_3$**, **$(K_4E_4)_3$**, **$(K_4E_4)_4$** and **$A_4(E_4K_4)_3A_4$**, with 250 ms mixing times at 600 MHz on a Varian VNMRS or INOVA spectrometer (University of Bristol) equipped with a triple resonance cryogenically cooled probe-head and room-temperature probe, respectively. Spectral widths of 7000 Hz, 8192 complex points in $f_2$ and 800 complex points in $f_1$ were used for both the NOESY and TOSCY spectra. Due to signal overlap at 600 MHz data for **$(K_4E_4)_3$**, **$(K_4E_4)_4$** and **$A_4(K_4E_4)_3A_4$**, spectra were also recorded once at 900 MHz on a Varian INOVA spectrometer (University of Birmingham) equipped with a triple-resonance cryogenically cooled probe-head. Spectral widths of 8000 Hz, 8192 complex points in $f_2$ and 1024 complex points in $f_1$ were used for both the NOESY and TOCSY spectra.

Δδ values were calculated as the difference in the observed chemical shifts and the sequence-corrected random-coil chemical shifts. The latter were generated using the chemical shifts for intrinsically disordered proteins javascript (http://www1.bio.ku.dk/english/research/pv/sbin_lab/staff/MAK/randomcoil/script/).[52,53]

## Bioinformatics

Bioinformatics analyses were performed using in-house Perl scripts on α-helices from a non-redundant set (≤30% pairwise sequence identity) of 2775 sub-1.6 Å resolution X-ray crystal structures collated from the RCSB Protein Data Bank (PDB)[43] using PISCES.[54] Only helices of 12 amino acids or longer and as identified by Promotif,[55] which uses a modified algorithm[56] for protein secondary structure assignment, were considered.

The distribution of hydrogen bonds made by Glu, Lys and Asn, were identified using Promotif[55] at various positions within the α-helices. Hydrogen bonds were identified using HBPlus[44] and categorised according to hydrogen bond donor type as follows: no hydrogen bond; ≥1 hydrogen bond to water; ≥1 hydrogen bond to a side chain atom; and ≥1 hydrogen bond to a main-chain atom.

Identification of interacting residue pairs within helices were identified as follows: The first four residues of each helix were classed as *N*-terminal; residues at least 4 positions in sequence away from the *N*- and *C*-termini were considered ‘central’; and the last four residues were classed as *C*-terminal. Numbers of attractive $E_i \rightarrow K_{i+4}$, $K_i \rightarrow E_{i+4}$, $E_i \rightarrow K_{i+3}$, and $K_i \rightarrow E_{i+3}$ and repulsive $E_i \rightarrow E_{i+4}$, $K_i \rightarrow K_{i+4}$, $E_i \rightarrow E_{i+3}$, and $K_i \rightarrow K_{i+3}$ pairs were identified. Expected numbers of pairs were estimated using the occurrence of each residue in the whole dataset. Salt bridges were considered to be formed if the distance between Lys Nζ and the centroid of (Glu Oε1, Oε2) was ≤ 4 Å. For the repulsive pairs, the mean shortest distance between side-chain atoms was recorded, along with the percentage of repulsive residue pairs with sub-4 Å distances between them.

$\chi_1\chi_2$ Side-chain rotamer distributions of for both salt-bridging and non-salt-bridged central $E/K_i \rightarrow E/K_{i+3/4}$ pairs were categorised as follows: *t*, $\chi > 120°$ or $\chi < -120°$; $g^+$, $0° < \chi < 120°$; $g^-$, $-120° < \chi < 0°$ for Glu and Lys residues in all α-helices. Theoretical rotamer combinations were modelled in PyMOL (www.pymol.org) and salt-bridge potential

assigned if Glu Oε1/Oε2…Lys Nζ < 4 Å. Rotamer combinations were identified using Promotif.[55]

### Calculating electrostatic potentials

A 32 residue polyalanine α-helix was aligned along its long axis, z, (x,y = 0,0), so that the N-terminal atom was at z = 0 (all distances are in Angstroms). The relative partial charges of each dipole were approximated by single point charges on atoms as follows: HN, −0.307; N, −0.621; C, +0.582; O, −0.574.[57] Overall charge neutrality was preserved with the following point charges for side chain atoms: Cα, +0.250; Hα, +0.085; Cβ, −0.416; and Hβ, +0.129. The electrostatic potential in vacuum $V_E(x,0,z)$, (i.e. Coulomb's potential) at a distance (x) from the helix axis (y = 0) was calculated in volts using Mathematica scripts to sum the electrostatic potentials arising from all of the atomically centred point charges.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We are grateful to the EPSRC of the UK for a studentship to EGB, and the EPSRC/NSF (EP/J001430) for a grant to DNW that funded GJB. DNW holds a Royal Society Wolfson Research Merit Award. We thank Mr Chris Wood for a script to align helical axes; Profs. Michelle Peckham and Adrian Mulholland, and members of the Woolfson and Faul groups for helpful discussions; and Sara Whittaker and HWB·NMR at the University of Birmingham for access to the Wellcome Trust-funded 900 MHz spectrometer.

## References

1. Dill KA. Dominant forces in protein folding. Biochemistry. 1990; 29:7133–7155.
2. 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]
3. Bartlett GJ, Choudhary A, Raines RT, Woolfson DN. n➔π* interactions in proteins. Nat. Chem. Biol. 2010; 6:615–620. [PubMed: 20622857]
4. Bartlett GJ, Newberry RW, VanVeller B, Raines RT, Woolfson DN. Interplay of hydrogen bonds and n➔π* interactions in proteins. J. Am. Chem. Soc. 2013; 135:18682–18688. [PubMed: 24256417]
5. Bierzynski A, Kim PS, Baldwin RL. A salt bridge stabilizes the helix formed by isolated C-peptide of RNase-A. Proc. Natl. Acad. Sci. USA. 1982; 79:2470–2474. [PubMed: 6283528]
6. Kim PS, Baldwin RL. A helix stop signal in the isolated S-peptide of ribonuclease-A. Nature. 1984; 307:329–334. [PubMed: 6694731]
7. Marqusee S, Baldwin RL. Helix stablization by $Glu^{-}$…$Lys^{+}$ salt bridges in short peptides of *de novo* design. Proc. Natl. Acad. Sci. USA. 1987; 84:8898–8902. [PubMed: 3122208]
8. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of α-helices. Science. 1988; 240:1648–1652. [PubMed: 3381086]
9. Doig AJ, Baldwin RL. N- and C-capping preferences for all 20 amino acids in α-helical peptides. Protein Sci. 1995; 4:1325–1336. [PubMed: 7670375]
10. Pace NC, Scholtz MJ. A helix propensity scale based on experimental studies of peptides and proteins. Biophys. J. 1998; 75:422–427. [PubMed: 9649402]
11. Penel S, Hughes E, Doig AJ. Side-chain structures in the first turn of the α-helix. J. Mol. Biol. 1999; 287:127–143. [PubMed: 10074412]

12. Iqbalsyah TM, Doig AJ. Anticooperativity in a Glu–Lys–Glu salt bridge triplet in an isolated α-helical peptide. Biochemistry. 2005; 44:10449–10456. [PubMed: 16060653]
13. Spek EJ, Bui AH, Lu M, Kallenbach NR. Surface salt bridges stabilize the GCN4 leucine zipper. Protein Sci. 1998; 7:2431–2437. [PubMed: 9828010]
14. Mayne L, et al. Stabilizing effect of a multiple salt bridge in a prenucleated peptide. J. Am. Chem. Soc. 1998; 120:10643–10645.
15. Olson CA, Spek EJ, Shi ZS, Vologodskii A, Kallenbach NR. Cooperative helix stabilization by complex Arg-Glu salt bridges. Proteins: Structure Function and Genetics. 2001; 44:123–132. [PubMed: 11391775]
16. Scholtz JM, Qian H, Robbins VH, Baldwin RL. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. Biochemistry. 1993; 32:9668–9676. [PubMed: 8373771]
17. Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL. Tests of the helix dipole model for stabilization of α-helices. Nature. 1987; 326:563–567. [PubMed: 3561498]
18. Armstrong KM, Baldwin RL. Charged histidine affects α-helix stability at all positions in the helix by interacting with the backbone charges. Proc. Natl. Acad. Sci. USA. 1993; 90:11337–11340. [PubMed: 8248249]
19. Wada A. The α-helix as an electric macro-dipole. Adv. Biophys. 1976
20. Hol WGJ, van Duijnen PT, Berendsen HJC. The α-helix dipole and the properties of proteins. Nature. 1978; 273:443–446. [PubMed: 661956]
21. Chakrabarti P. An assessment of the effect of the helix dipole in protein structures. Protein Eng. 1994; 7:471–474. [PubMed: 8029204]
22. Daopin S, Sauer U, Nicholson H, Matthews BW. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. Biochemistry. 1991; 30:7142–7153. [PubMed: 1854726]
23. Lumb KJ, Kim PS. Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. Science. 1995; 268:436–439. [PubMed: 7716550]
24. Lavigne P, Sönnichsen FD, Kay CM, Hodges KJ. Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. Science. 1996; 271:1136–1138. [PubMed: 8599093]
25. Aqvist J, Luecke H, Quiocho FA, Warshel A. Dipoles localized at helix termini of proteins stablize charges. Proc. Natl. Acad. Sci. USA. 1991; 88:2026–2030. [PubMed: 2000410]
26. Wang CL, et al. A long helix from the central region of smooth muscle caldesmon. J. Biol. Chem. 1991; 266:13958–13963. [PubMed: 1856225]
27. Spink BJ, Sivaramakrishnan S, Lipfert J, Doniach S, Spudich JA. Long single α-helical tail domains bridge the gap between structure and function of myosin VI. Nat. Struct. Mol. Biol. 2008; 15:591–597. [PubMed: 18511944]
28. Yang Y, et al. A FERM domain autoregulates Drosophila myosin 7a activity. Proc. Natl. Acad. Sci. USA. 2009; 106:4189–4194. [PubMed: 19255446]
29. Knight PJ, et al. The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. J. Biol. Chem. 2005; 280:34702–34708. [PubMed: 16030012]
30. Peckham M, Knight PJ. When a predicted coiled coil is really a single α-helix, in myosins and other proteins. Soft Matter. 2009; 5:2493–2503.
31. Süveges D, Gáspári Z, Tóth G, Nyitray L. Charged single α-helix: A versatile protein structural motif. Proteins: Structure Function and Bioinformatics. 2009; 74:905–916. [PubMed: 18712826]
32. Gaspari Z, Suveges D, Perczel A, Nyitray L, Toth G. Charged single alpha-helices in proteomes revealed by a consensus prediction approach. Biochim. Biophys. Acta. 2012; 1824:637–646. [PubMed: 22310480]
33. Peckham M. Coiled coils and SAH domains in cytoskeletal molecular motors. Biochem. Soc. Trans. 2011; 39:1142–1148. [PubMed: 21936779]
34. Sivaramakrishnan S, Spink BJ, Sim AYL, Doniach S, Spudich JA. Dynamic charge interactions create surprising rigidity in the ER/K α-helical protein motif. Proc. Natl. Acad. Sci. USA. 2008; 105:13356–13361. [PubMed: 18768817]
35. Lyu PCC, Gans PJ, Kallenbach NR. Energetic contribution of solvent-exposed ion-pairs to α-helix structure. J. Mol. Biol. 1992; 223:343–350. [PubMed: 1731079]

36. Cochran DAE, Penel S, Doig AJ. Effect of the N1 residue on the stability of the α-helix for all 20 amino acids. Protein Sci. 2001; 10:463–470. [PubMed: 11344315]
37. Cochran DAE, Doig AJ. Effect of the N2 residue on the stability of the α-helix for all 20 amino acids. Protein Sci. 2001; 10:1305–1311. [PubMed: 11420432]
38. Iqbalsyah TM, Doig AJ. Effect of the N3 residue on the stability of the α-helix. Protein Sci. 2004; 13:32–39. [PubMed: 14691219]
39. Walter S, Hubner B, Hahn U, Schmid FX. Destabilization of a protein helix by electrostatic interactions. J. Mol. Biol. 1995; 252:133–143. [PubMed: 7666425]
40. Scholtz JM, et al. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. Proc. Natl. Acad. Sci. USA. 1991; 88:2854–2858. [PubMed: 2011594]
41. Fernández-Recio J, Sancho J. Intrahelical side chain interactions in α-helices: poor correlation between energetics and frequency. FEBS Lett. 1998; 429:99–103. [PubMed: 9657391]
42. de Sousa MM, et al. Amino acid pair- and triplet-wise groupings in the interior of α-helical segments in proteins. J. Theor. Biol. 2011; 271:136–144. [PubMed: 21130100]
43. Berman HM, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
44. McDonald IK, Thornton JM. Satisfying hydrogen-bonding potential in proteins. J. Mol. Biol. 1994; 238:777–793. [PubMed: 8182748]
45. Lacroix E, Viguera AR, Serrano L. Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. J. Mol. Biol. 1998; 284:173–191. [PubMed: 9811549]
46. Sivaramakrishnan S, Spudich JA. Systematic control of protein interaction using a modular ER/K α-helix linker. Proc. Natl. Acad. Sci. USA. 2011; 108:20467–20472. [PubMed: 22123984]
47. Baboolal TG, et al. The SAH domain extends the functional length of the myosin lever. Proc. Natl. Acad. Sci. USA. 2009; 106:22193–22198. [PubMed: 20018767]
48. Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. Biopolymers. 1991; 31:1463–1470. [PubMed: 1814498]
49. Myers JK, Pace CN, Scholtz JM. A direct comparison of helix propensity in proteins and peptides. Proc. Natl. Acad. Sci. USA. 1997; 94:2833–2837. [PubMed: 9096306]
50. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. 2006; 1:2527–2535. [PubMed: 17406506]
51. Gorbet G, et al. A parametrically constrained optimization method for fitting sedimentation velocity experiments. Biophys. J. 2014; 106:1741–1750. [PubMed: 24739173]
52. Kjaergaard M, Brander S, Poulsen FM. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. J. Biomol. NMR. 2011; 49:139–149. [PubMed: 21234644]
53. Kjaergaard M, Poulsen FM. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. J. Biomol. NMR. 2011; 50:157–165. [PubMed: 21604143]
54. Wang GL, Dunbrack RL. PISCES: a protein sequence culling server. Bioinformatics. 2003; 19:1589–1591. [PubMed: 12912846]
55. Hutchinson EG, Thornton JM. PROMOTIF - A program to identify and analyze structural motifs in proteins. Protein Sci. 1996; 5:212–220. [PubMed: 8745398]
56. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
57. Thomas A, Milon A, Brasseur R. Partial atomic charges of amino acids in proteins. Proteins: Structure Function and Bioinformatics. 2004; 56:102–109. [PubMed: 15162490]

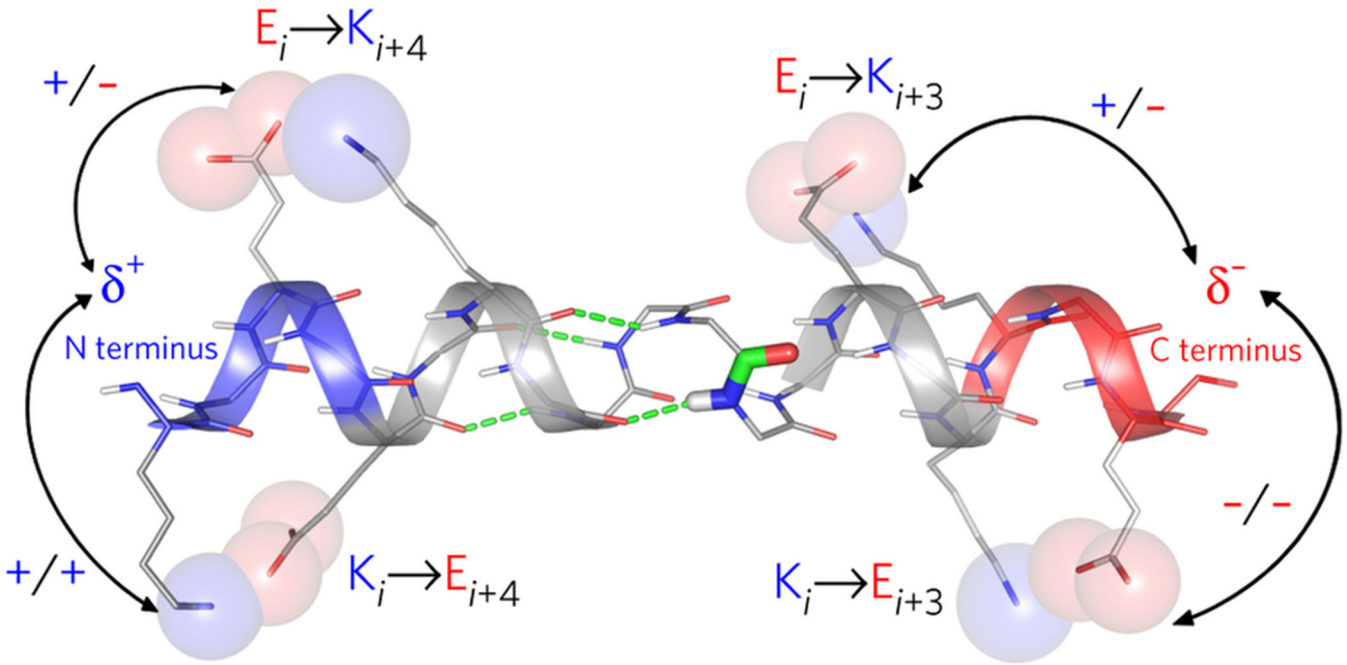


**Figure 1. Electrostatic interactions in the α-helix**

Cartoon representation of the α-helix showing the following: $CO_i \rightarrow NH_{i+4}$ hydrogen bonds (green dashes); the peptide-bond dipole (thick sticks, centre), many copies of which align to give the helix macrodipole; clusters of backbone NH and CO groups, which give rise to local $\delta^+$ and $\delta^-$ charges at the *N*- (blue) and *C*- termini (red), respectively; interactions between charged side chains and the helix terminal charges (arrows); and salt bridges between charged side chains spaced i➔i+3 and i➔i+4 apart (labelled spheres).

Europe PMC Funders Author Manuscripts

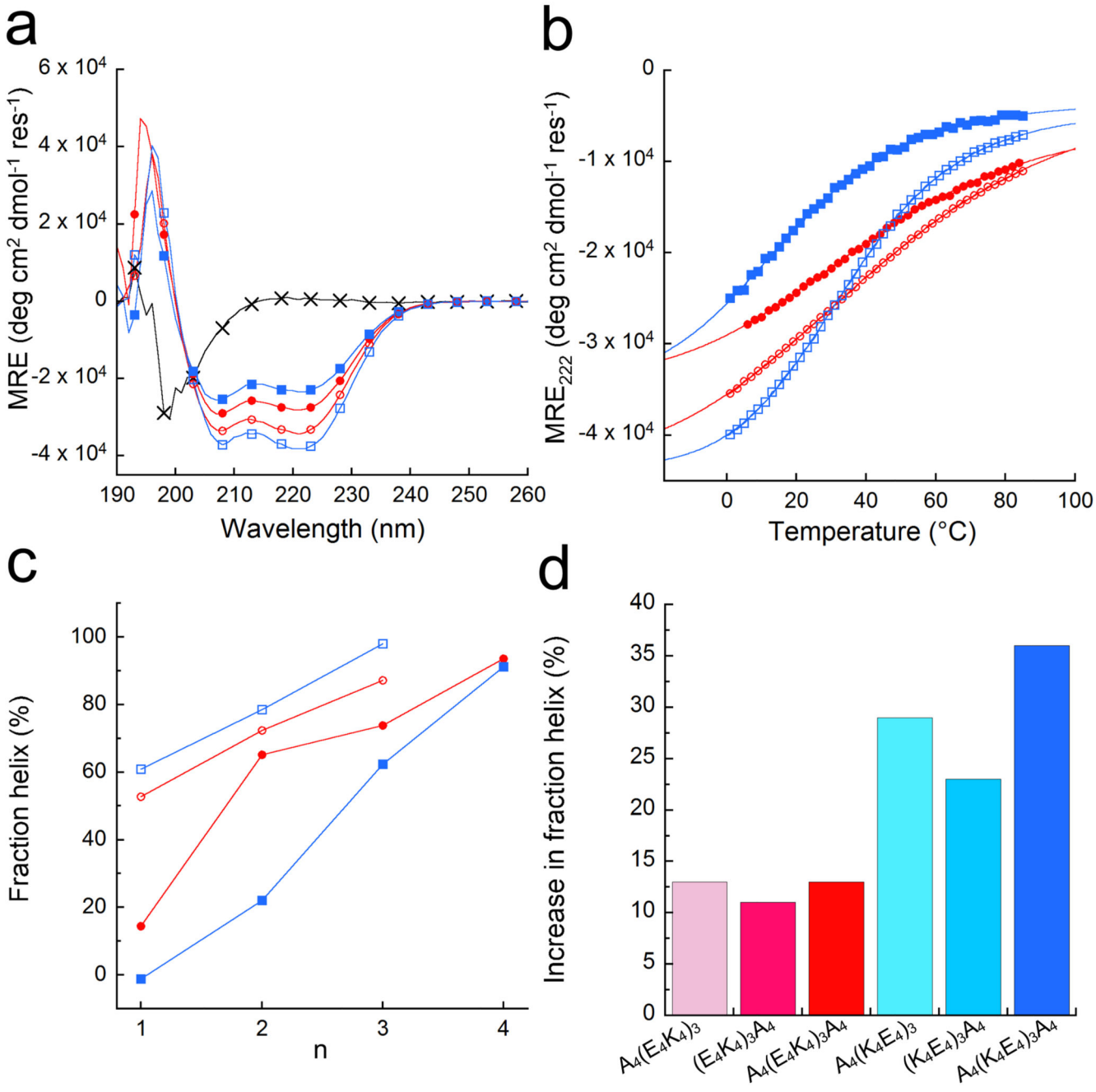


**Figure 2. Helicities of the designed peptides in solution**

(**a**) CD spectra at 5 °C of: **$(E_2K_2)_6$** (black crosses), **$(E_4K_4)_3$** (red circles), **$(K_4E_4)_3$** (blue squares), **$A_4(E_4K_4)_3A_4$** (red open circles), and **$A_4(K_4E_4)_3A_4$** (blue open squares). (**b**) Thermal unfolding curves followed by the change in $MRE_{222}$ for all except **$(E_2K_2)_6$**, with the symbols and colours from (**a**) preserved. Fits to the Gibbs-Helmholtz equation are shown by solid lines. (**c**) Fraction helix as a function of peptide length for the $(E_4K_4)_n$-(red) and $(K_4E_4)_n$-based (blue) peptides with (open symbols) and without (filled symbols) tetra-alanine flanks (data points are joined by lines to guide the eye). (**d**) Increase in fraction

helicities at 5 °C of the alanine-flanked peptides **$A_4(E_4K_4)_3$**, **$(E_4K_4)_3A_4$** and **$A_4(E_4K_4)_3A_4$** (red colours) relative to **$(E_4K_4)_3$**; and for **$A_4(K_4E_4)_3$**, **$(K_4E_4)_3A_4$** and **$A_4(K_4E_4)_3A_4$** (blue colours) relative to **$(K_4E_4)_3$**. Conditions: 100 μM peptide concentration in PBS 137 mM NaCl, pH 7.4.



Europe PMC Funders Author Manuscripts

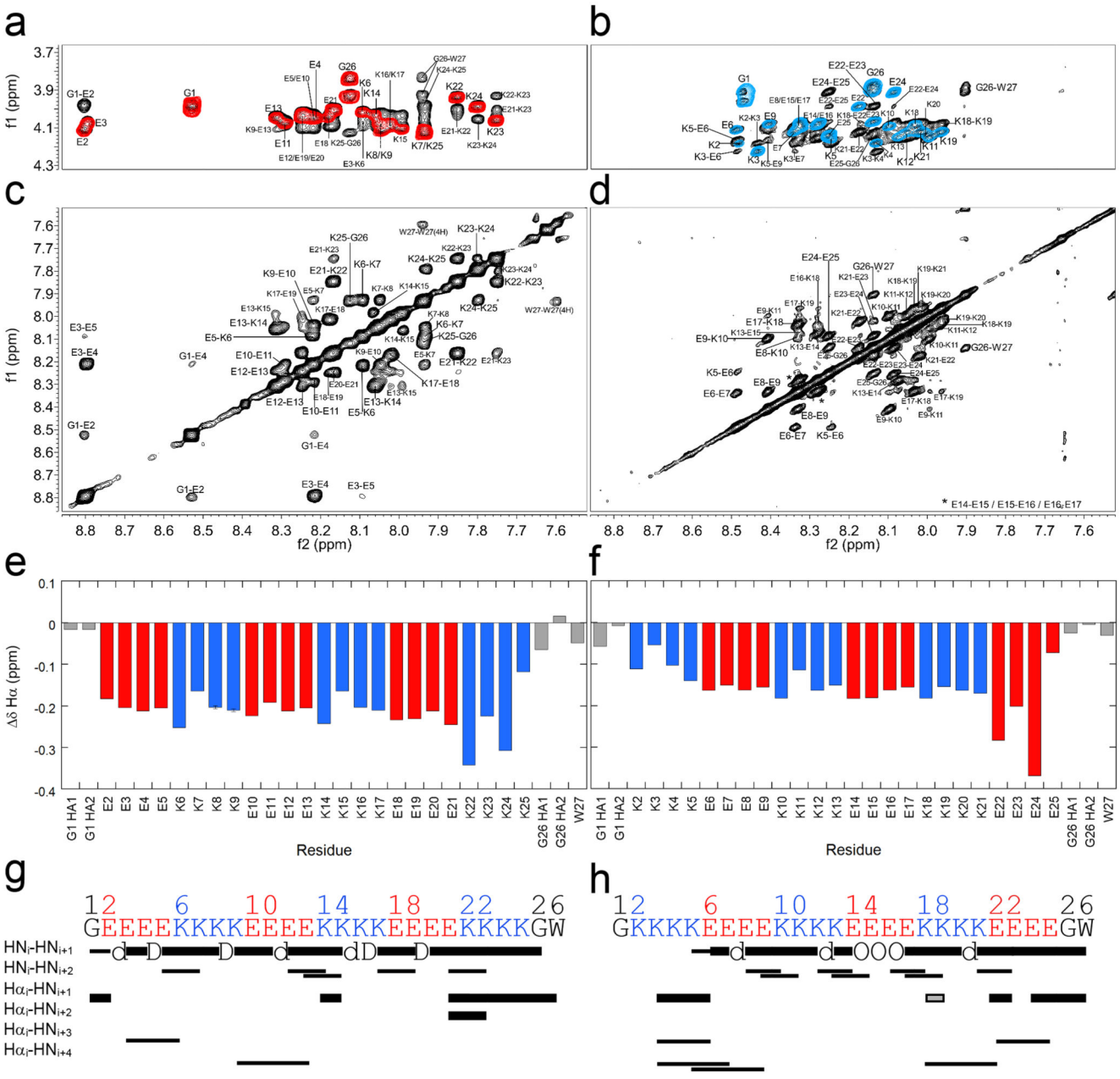


**Figure 3. Locating α-helical structure by NMR spectroscopy**

$^1$H-NMR spectra for **$(E_4K_4)_3$** (**a&c**) and **$(K_4E_4)_3$** (**b&d**). (**a&b**) Overlaid fingerprint regions of the TOCSY (coloured contours: red for **$(E_4K_4)_3$**, and blue for **$(K_4E_4)_3$**) and NOESY (black contours) spectra. Note the larger range of HN shifts for **$(E_4K_4)_3$**. (**b&d**) NOESY spectra showing the amide regions and large number of sequential $HN_i$-$HN_{i+1}$ NOEs. Note the first NOE of this type for **$(K_4E_4)_3$** is between Lys-5 and Glu-6 (K5/E6 circled in **d**). (**e&f**) Δδ plots for the backbone Hα chemical shifts of **$(E_4K_4)_3$** and **$(K_4E_4)_3$**, respectively. Key: Glu, red; Lys, blue; Gly and Trp, grey. Δδ values are the difference in sequence-corrected random coil chemical shifts from those observed. Δδs are greater for **$(E_4K_4)_3$** than **$(K_4E_4)_3$** on average, consistent with a more folded α-helix. (**g&h**) Connectivity diagrams




Europe PMC Funders Author Manuscripts

for **$(E_4K_4)_3$** and **$(K_4E_4)_3$**, respectively. Note there are only a few identifiable Hα-HN NOEs due to significant peak overlap in this region (**a&b**). Key: thick bars, medium intensity NOEs (1.8 – 3.5 Å); thin bars, weak NOEs (1.8 – 5.0 Å); grey, ambiguous; d, NOE hidden by diagonal; D, degenerate; O, overlapped. Conditions: 1 mM peptide concentration in PBS (137 mM NaCl) with 10% $D_2O$, pH 7.4 at 5 °C; **$(E_4K_4)_3$** 600 MHz, and **$(K_4E_4)_3$** 900 MHz.



Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

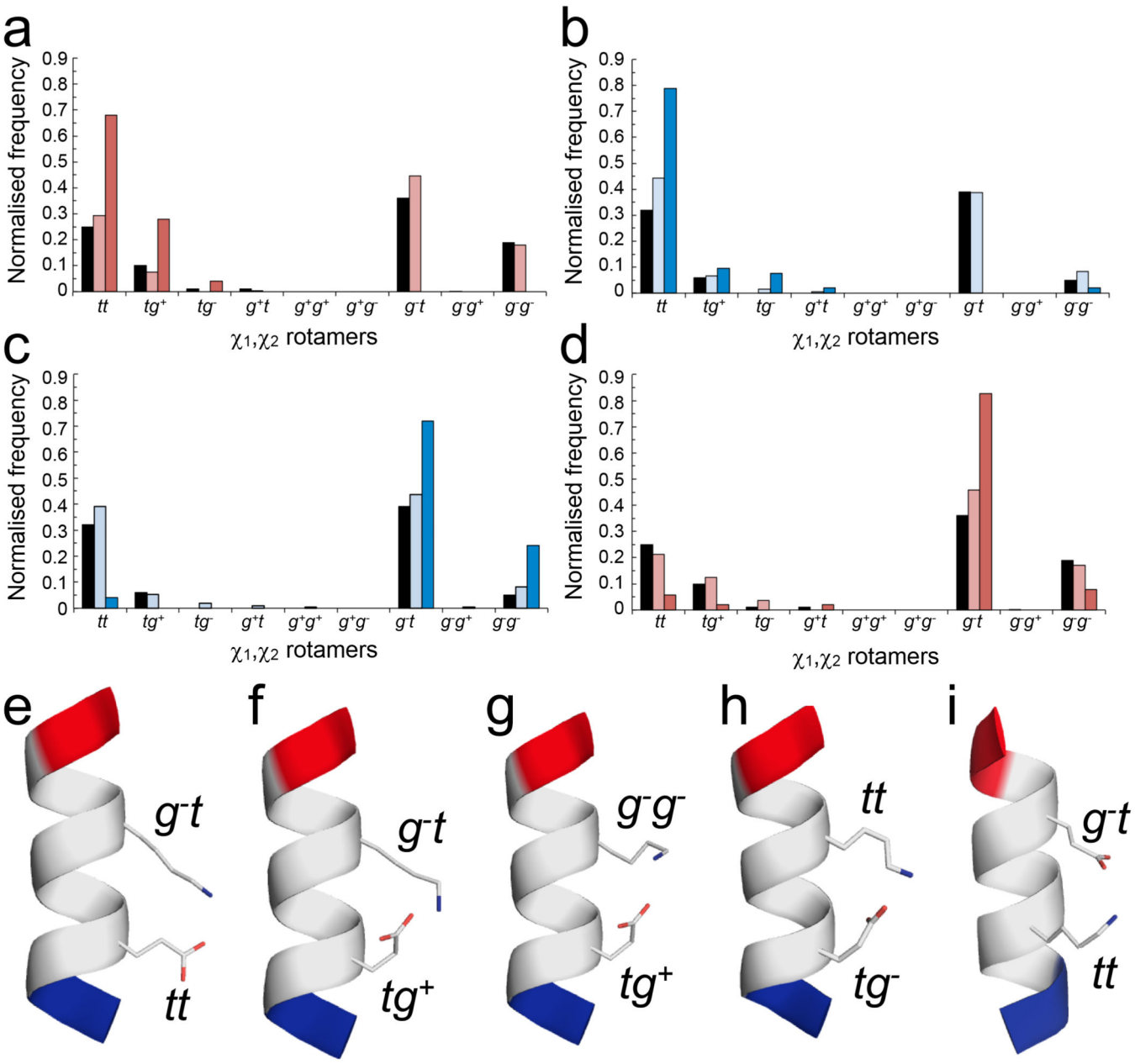


**Figure 4. Side-chain interactions observed in α-helices from the Protein Data Bank**

$\chi_1,\chi_2$ distributions and conformers for $E_i \rightarrow K_{i+4}$ and $K_i \rightarrow E_{i+4}$ pairs in high-resolution X-ray crystal structures. (**a – d**) Normalised frequencies of preferred $\chi_1,\chi_2$ angles for Glu (**a&d**) and Lys (**b&c**) residues in $E_i \rightarrow K_{i+4}$ (**a&c**) and $K_i \rightarrow Ei_{+4}$ (**b&d**) pairs. Key: red for Glu, blue for Lys; black bars indicate the frequency of each rotamer found in all α-helices; pale bars indicate pairs where no salt bridge is made; and dark bars indicate pairs where a salt bridge is formed. (**e – i**) Examples of salt-bridging pairs. For the $E_i \rightarrow K_{i+4}$ pairs, there are two dominant rotamer combinations, with Glu (*tt; tg*$^+$) plus Lys (*g*$^-$*t*) (**e&f**, 17 examples (68%) and 6 examples (24%), respectively), and two minor combinations (**g&h**, 1 example of each (8%)) with Glu;Lys (*tg*$^+$*;g*$^-$*g*$^-$) and Glu;Lys (*tg*$^-$*; tt*). Whereas for the $K_i \rightarrow E_{i+4}$ pairs there is

one clearly preferred conformation (Glu, $g^{-}t$; Lys, *tt*) (**i**, 39 examples (75%)). Only the $g^{-}t;g^{-}t$ combination would be better, but inspection of molecular models revealed that with $\chi_1 = g^{-}$ for Lys takes the ε-amino group too far from the γ-carboxylate of Glu to form a salt bridge. Examples were taken from the PDB as follows: (**e**) 1kqp A246-A250; (**f**) 1moq A481-A485; (**g**) 3n0u A177-A181; (**h**) 2r75 A141-A145; (**i**) 1egw A30-A34. Images were generated with PyMol (www.pymol.org).



Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

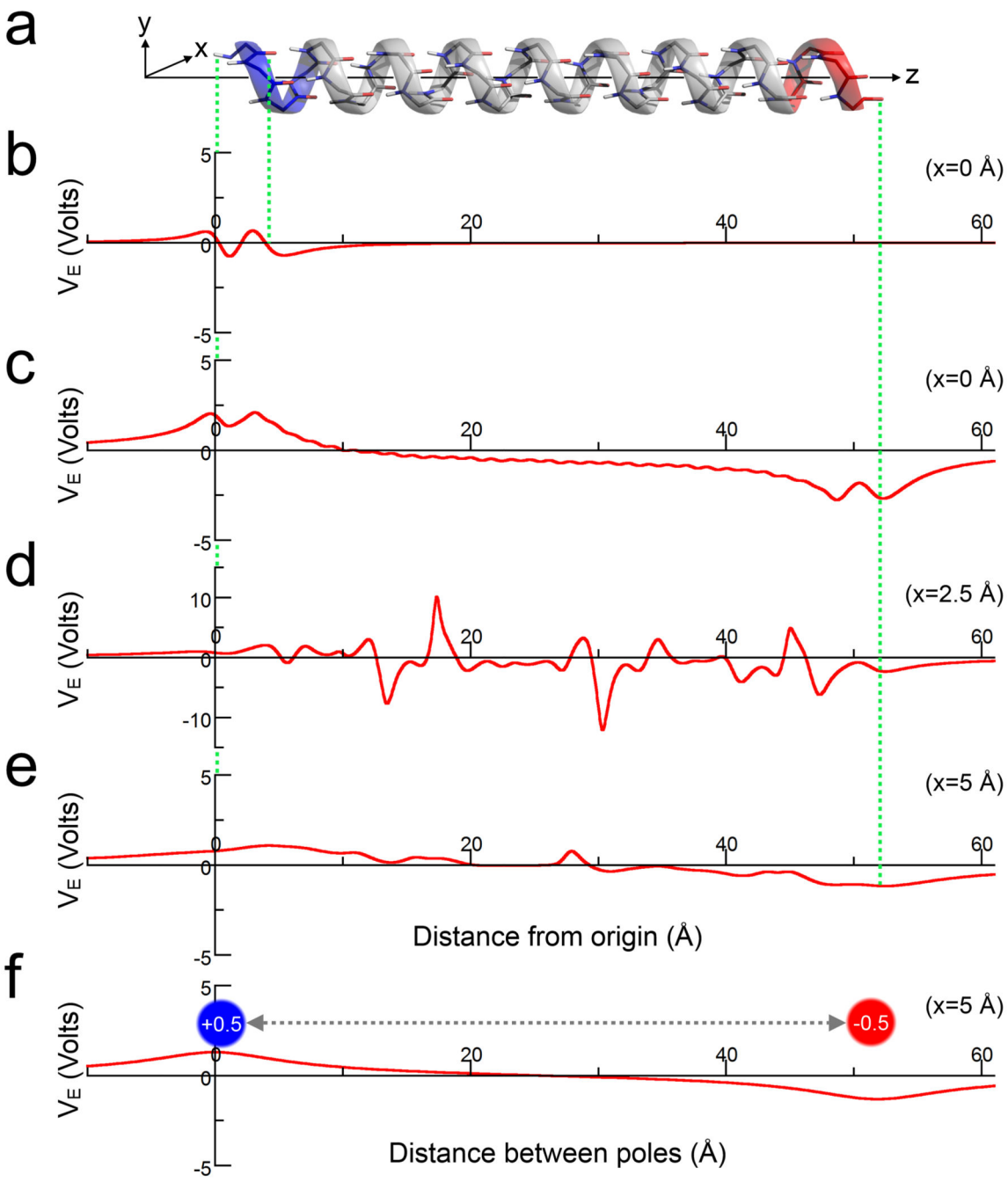


**Figure 5. Electrostatic potential of a model α-helix**

(**a**) A 32-residue α-helix with its long axis aligned along z. (**b**) The electrostatic potential ($V_E$) of a single residue from the helix at x = 0 Å. (**c**-**e**) The electrostatic potential along the entire helix at varying distances (x) from the helix axis: (**c**) x = 0 Å; (**d**) x = 2.5 Å; and (**e**) x = 5.0 Å. Note, the increased limits for the $V_E$-axis in part (**d**). (**f**) The electrostatic potential arising from point charges of +0.5 and −0.5 at positions equivalent to the *N*- and *C*-terminal

helical atoms on the z-axis, respectively. The electrostatic potentials in all parts (**b-f**) were calculated in vacuum.



Europe PMC Funders Author Manuscripts

## Table 1

### Peptide designs

Names, sequences, mean residue ellipticities at 222 nm ($MRE_{222}$) and 5 °C, corresponding percent helicities, and midpoints of thermal denaturation curves ($T_m$) for the peptides used in this study. Conditions: phosphate buffered saline (PBS, 137 mM NaCl), pH 7.4, and 100 μM peptide concentration. The purity and identities of the peptides were confirmed by HPLC and MALDI-TOF mass spectrometry (Supplementary Figs. 1 – 27).

| Peptide Name | Sequence[1] | $MRE_{222}$ | Fraction Helix (%) | $T_m$ (°C) |
|---|---|---|---|---|
| $(E_2K_2)_6$ | Ac–GEEKKEEKKEEKKEEKKEEKKEEKKGYY–$NH_2$ | 813 | −1 | |
| $(EK)_{12}$ | AC–GEKEKEKEKEKEKEKEKEKEKEKEKGW–$NH_2$ | −8,136 | 22 | |
| $(KE)_{12}$ | AC–GKEKEKEKEKEKEKEKEKEKEKEKEGW–$NH_2$ | −4,180 | 12 | |
| $(E_4K_4)_4$ | AC–GEEEEKKKKEEEEKKKKEEEEKKKKEEEEKKKKGW–$NH_2$ | −36,454 | 94 | |
| $(E_4K_4)_3$ | AC–GEEEEKKKKEEEEKKKKEEEEKKKKGW–$NH_2$ | −27,885 | 74 | |
| $(E_4K_4)_2$ | AC–GEEEEKKKKEEEEKKKKGW–$NH_2$ | −23,374 | 65 | |
| $(E_4K_4)$ | AC–GEEEEKKKKGW–$NH_2$ | −4,219 | 14 | |
| $(K_4E_4)_4$ | AC–GKKKKEEEEKKKKEEEEKKKKEEEEKKKKEEEEGW–$NH_2$ | −35,479 | 91 | 28 |
| $(K_4E_4)_3$ | AC–GKKKKEEEEKKKKEEEEKKKKEEEEGW–$NH_2$ | −23,493 | 62 | |
| $(K_4E_4)_2$ | AC–GKKKKEEEEKKKKEEEEGW–$NH_2$ | −7,666 | 22 | |
| $(K_4E_4)$ | AC–GKKKKEEEEGW–$NH_2$ | 816 | −1 | |
| $A_4(E_4K_4)_3A_4$ | AC–GAAAAEEEEKKKKEEEEKKKKEEEEKKKKAAAAGW–$NH_2$ | −33,934 | 87 | |
| $A_4(E_4K_4)_2A_4$ | AC–GAAAAEEEEKKKKEEEEKKKKAAAAGW–$NH_2$ | −27,313 | 72 | |
| $A_4(E_4K_4)A_4$ | AC–GAAAAEEEEKKKKAAAAGW–$NH_2$ | −18,885 | 53 | |
| $A_4(K_4E_4)_3A_4$ | AC–GAAAAKKKKEEEEKKKKEEEEKKKKEEEEAAAAGW–$NH_2$ | −38,166 | 98 | 33 |
| $A_4(K_4E_4)_2A_4$ | AC–GAAAAKKKKEEEEKKKKEEEEAAAAGW–$NH_2$ | −29,687 | 78 | 21 |
| $A_4(K_4E_4)A_4$ | AC–GAAAAKKKKEEEEAAAAGW–$NH_2$ | −21,847 | 61 | |
| $A_4(E_4K_4)_3$ | AC–GAAAAEEEEKKKKEEEEKKKKEEEEKKKKGW–$NH_2$ | −33,288 | 87 | |
| $(E_4K_4)_3A_4$ | AC–GEEEEKKKKEEEEKKKKEEEEKKKKAAAAGW–$NH_2$ | −32,668 | 85 | |
| $A_4(K_4E_4)_3$ | AC–GAAAAKKKKEEEEKKKKEEEEKKKKEEEEGW–$NH_2$ | −35,177 | 91 | 25 |
| $(K_4E_4)_3A_4$ | AC–GKKKKEEEEKKKKEEEEKKKKEEEEAAAAGW–$NH_2$ | −32,796 | 85 | 22 |
| $A_4(E_4K_4)A_4(E_4K_4)A_4$ | AC–GAAAAEEEEKKKKAAAAEEEEKKKKAAAAGW–$NH_2$ | −31,779 | 83 | 24 |
| $A_4(K_4E_4)A_4(K_4E_4)A_4$ | AC–GAAAAKKKKEEEEAAAAKKKKEEEEAAAAGW–$NH_2$ | −37,155 | 97 | 29 |
| $(E_3K_3)_4$ | AC–GEEEKKKEEEKKKEEEKKKEEEKKKGWW–$NH_2$ | −25,428 | 67 | |
| $(K_3E_3)_4$ | AC–GKKKEEEKKKEEEKKKEEEKKKEEEGWW–$NH_2$ | −12,229 | 33 | |
| $A_4(E_3K_3)_4A_4$ | AC–GAAAAEEEKKKEEEKKKEEEKKKEEEKKKAAAAGWW–$NH_2$ | −31,869 | 82 | 21 |
| $A_4(K_3E_3)_4A_4$ | AC–GAAAAKKKEEEKKKEEEKKKEEEKKKEEEAAAAGWW–$NH_2$ | −30,287 | 78 | 14 |

[1] Amino acids are represented by standard one-letter codes: A, alanine; E, glutamic acid; G, glycine; and K, lysine. To remove complicating terminal formal charges, and to ameliorate end effects in the helices, the E/K regions in all of the peptides for this study were flanked by glycine residues, and capped with acetyl (Ac) and amide groups ($NH_2$) at their *N*-and *C*-termini, respectively. *C*-terminal tryptophan (W) or tyrosine (Y) residues were included for concentration determination.



Europe PMC Funders Author Manuscripts