

# SCIENTIFIC REPORTS



OPEN

## Characterization of long and stable *de novo* single alpha-helix domains provides novel insight into their stability

Marcin Wolny<sup>1,\*</sup>, Matthew Batchelor<sup>1,\*</sup>, Gail J. Bartlett<sup>2</sup>, Emily G. Baker<sup>2</sup>, Marta Kurzawa<sup>1</sup>, Peter J. Knight<sup>1</sup>, Lorna Dougan<sup>3</sup>, Derek N. Woolfson<sup>2,4,5</sup>, Emanuele Paci<sup>1</sup> & Michelle Peckham<sup>1</sup>

Naturally-occurring single  $\alpha$ -helices (SAHs), are rich in Arg (R), Glu (E) and Lys (K) residues, and stabilized by multiple salt bridges. Understanding how salt bridges promote their stability is challenging as SAHs are long and their sequences highly variable. Thus, we designed and tested simple *de novo* 98-residue polypeptides containing 7-residue repeats (AEEEXXX, where X is K or R) expected to promote salt-bridge formation between Glu and Lys/Arg. Lys-rich sequences (EK3 (AEEKKK) and EK2R1 (AEEKRR)) both form SAHs, of which EK2R1 is more helical and thermo-stable suggesting Arg increases stability. Substituting Lys with Arg (or *vice versa*) in the naturally-occurring myosin-6 SAH similarly increased (or decreased) its stability. However, Arg-rich *de novo* sequences (ER3 (AEEERRR) and EK1R2 (AEEKRR)) aggregated. Combining a PDB analysis with molecular modelling provides a rational explanation, demonstrating that Glu and Arg form salt bridges more commonly, utilize a wider range of rotamer conformations, and are more dynamic than Glu–Lys. This promiscuous nature of Arg helps explain the increased propensity of *de novo* Arg-rich SAHs to aggregate. Importantly, the specific K:R ratio is likely to be important in determining helical stability in *de novo* and naturally-occurring polypeptides, giving new insight into how single  $\alpha$ -helices are stabilized.

Naturally occurring single  $\alpha$ -helices (SAHs) are found in ~4% of proteins<sup>1–3</sup>, and have been commonly misidentified as coiled coils<sup>4</sup>. SAHs are not stabilized by a tertiary fold, remain monomeric and highly helical over a broad range of salt concentrations and pH, and exhibit a thermal denaturation profile that is only weakly cooperative<sup>2,5</sup>. They are stiff enough to replace the canonical lever in myosin<sup>6</sup>, can behave as a “constant force spring” when extended by small (<30 pN) forces<sup>5</sup>, and have a persistence length of ~15 nm (equivalent to ~100 residues of  $\alpha$ -helix)<sup>7</sup>. Many putative SAHs have now been identified in a wide range of proteins with diverse functions, typically ranging in length between 40 and ~200 residues<sup>1,2,7–9</sup>. However, only a few of these have been characterized experimentally and in detail<sup>4–6,9–12</sup>. Most recently, it has been shown that in some cases, a region of the SAH can mediate binding to other proteins<sup>13,14</sup>. Thus it is important to understand the underlying stability of these intriguing domains, in order to understand their contribution to protein function.

SAHs are likely to form when a sequence is rich in acidic and basic residues (Glu, E; Arg, R and Lys, K), lacks a hydrophobic seam, and contains many potential intrahelical interactions (salt bridges) between either E and K (E–K), or E and R (E–R). These pairings are spaced three or four residues apart (e.g. “E → K(+3)” or “K → E(+3)” where the K residue is 3 residues downstream or upstream from E respectively)<sup>2,4</sup>. However, understanding how such salt bridges promote the highly helical states of natural sequences is challenging, as SAHs have a range of different potential salt bridges in terms of sequence separation and residue type. Single E–K or E–R pairs of this type are known to promote the folding of monomeric  $\alpha$ -helices in short (<20 residue) alanine-based peptides<sup>15,16</sup>, and thus are assumed to stabilize long SAHs. Studies on short peptides have additionally suggested that K → E(+4)

<sup>1</sup>Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, UK. <sup>2</sup>School of Chemistry, University of Bristol, Cantock’s Close, Bristol, BS8 1TS, UK. <sup>3</sup>Astbury Centre for Structural Molecular Biology and School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, UK. <sup>4</sup>School of Biochemistry, University of Bristol, Biomedical Sciences Building, Bristol, BS8 1TD, UK. <sup>5</sup>BrisSynBio, University of Bristol, Life Sciences Building, Bristol, BS8 1TQ, UK. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.P. (email: m.peckham@leeds.ac.uk)



neighbouring charge–charge interactions, and has a high helix-propensity<sup>24</sup>. In addition, alanine is commonly found in known SAHs<sup>9</sup>, including those in caldesmon<sup>25</sup>, and the Kelch-motif family protein<sup>10</sup>.

The expressed and purified EK3 polypeptide exhibited the behaviour expected for a SAH. It was highly helical (>90% at 10 °C, 17 μM in 100 mM NaCl, Fig. 2a) and unfolded upon heating with a broad, non-cooperative transition (Fig. 2b), as shown by circular dichroism (CD) spectroscopy experiments. EK3 helicity remained high at high salt concentration (its helicity in 4 M NaCl was 60% of that in 100 mM NaCl (Fig. 2c)), and over a wide pH range (pH 2 to pH 10; Fig. 2d). EK3 is monomeric as shown by analytical ultracentrifugation (AUC) (Fig. 2e), and has an elongated shape as shown by size exclusion chromatography (SEC), where it eluted faster than a globular protein (ribonuclease A) of similar mass but slower than a dimeric coiled coil (15 H CC; ref. 5), which is elongated and has approximately double the mass of EK3 (Fig. 2f).

Surprisingly, the purified ER3 polypeptide did not behave as a SAH. It aggregated at low salt and neutral pH (100 mM NaCl, pH 7.4), as evidenced by the high turbidity of the solution (Fig. 2g) and was only soluble at low pH (<3.5). As a high proportion of the Glu residues will be protonated at low pH, ER3 is expected to have a net positive charge that allows it to be soluble. Although ER3 was highly helical (Fig. 2h) and had similar melting behaviour at pH 3.5 (Fig. 2i) to EK3 at pH 7.4 (Fig. 2b), we concluded that its lack of solubility at neutral pH was inconsistent with that of a SAH.

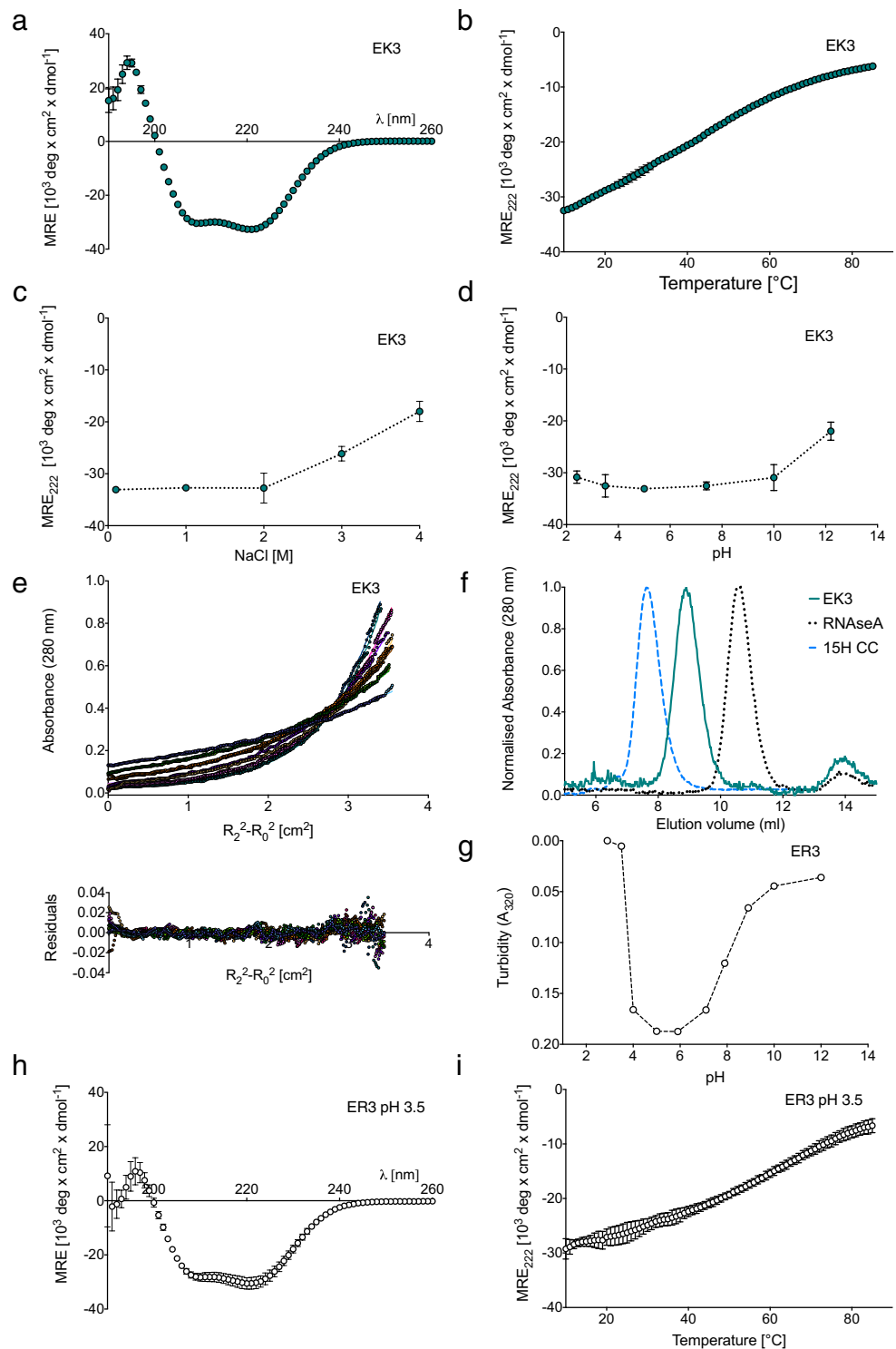
To further test the contribution of Arg to SAHs, we expressed and tested two additional *de novo* polypeptides, EK2R1 (AEEKRRK) and EK1R2 (AEEKRRR), in which either one or two Lys residues per repeat in EK3 were replaced by Arg (Fig. 1a). Replacing one Lys residue with Arg (EK2R1) resulted in a peptide that behaved as a SAH that was more helical (Fig. 3a) and thermally stable (Fig. 3b) compared to EK3. It was also more helical at high salt (4 M NaCl) than EK3 (Fig. 3c), and remained helical over a range of pH (Fig. 3d). AUC confirmed that EK2R1 was monomeric (Fig. 3e) and SEC showed it to be elongated but not oligomerized (Fig. 3f). In contrast, while replacing two Lys residues with Arg (EK1R2) resulted in a peptide that was soluble at pH 7.4 and highly helical (Fig. 3a), its more complex thermal unfolding behaviour suggested that it was not monomeric (Fig. 3b). Both AUC (Supplementary Fig. S1a) and SEC (Fig. 3f) showed that it was indeed oligomeric. Therefore, while EK2R1 behaves as a SAH, EK1R2 does not.

To explore the sequence requirements for *de novo* SAH formation, we designed two further polypeptides, EK1 and EK2 (Fig. 1). These had lower proportions of charged residues, and concomitant increases in helix-favouring alanine residues. EK2, which contained the repeat AEEAKKA (Fig. 1a), was highly helical (Supplementary Fig. S1b) but apparently dimerized. Its high helicity persisted up to 85 °C (Supplementary Fig. S1c) and in SEC, EK2 eluted between two known dimeric coiled-coil proteins of 25 kDa and 18.3 kDa (dimer mass, Supplementary Fig. S1d). However, an accurate molecular mass by AUC could not be determined due to protein precipitation (Supplementary Fig. S1e). We propose that the alanine residues of EK2 probably form a hydrophobic seam to promote dimerization into a coiled coil. EK1, which has AAEEAKA repeats (Fig. 1a), was soluble only in a very low ionic strength buffer (10 mM NaCl, 5 mM Tris, pH 7.4; Supplementary Fig. S1h). Under these conditions EK1 was helical (Supplementary Fig. S1f) but the thermal melt profiles indicated that it formed oligomeric species (Supplementary Fig. S1g). Thus, reducing the proportions of charged residues in this manner to form a polyalanine based construct is not compatible with the formation of a long SAH.

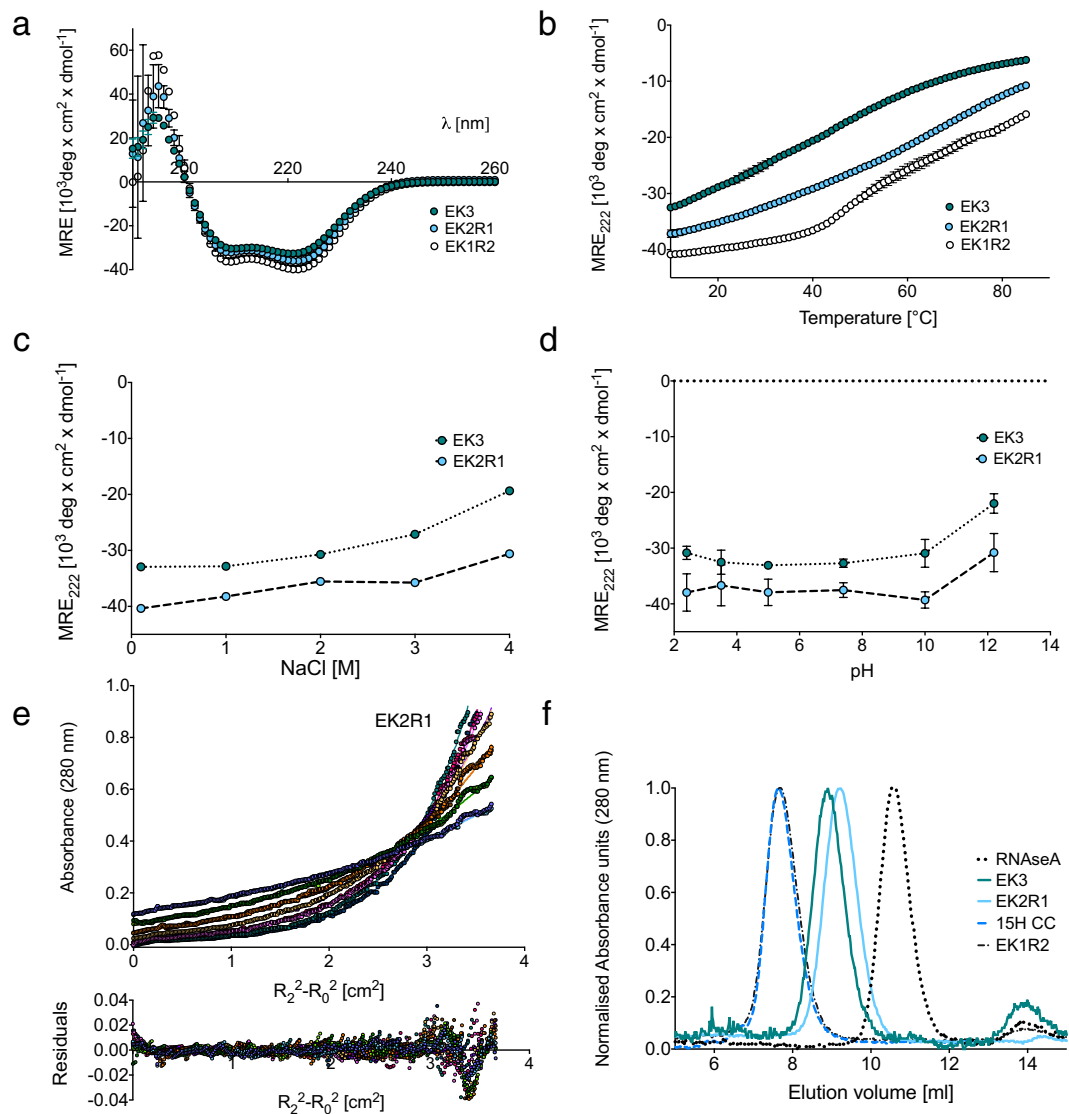
**Re-engineering natural SAHs predictably alters their properties.** The data above show that replacing a single Lys per repeat with Arg (EK3 to EK2R1) increases both the helicity and resistance to thermal unfolding. To test if this translates to a natural SAH, we substituted all the Lys residues in the SAH of myosin-6 with Arg (M6WT to M6R, Fig. 4a,b, which increases % Arg content from 19 to 30%), and, in a second construct, all of the Arg residues with Lys (M6K, Fig. 4a, which increases %K content from 11 to 30%). M6R was slightly more helical compared to M6WT, while M6K was slightly less helical (Fig. 4c), and M6R was significantly more helical than M6K (Fig. 4d). Similarly, the apparent order of thermal stability was M6R > M6WT > M6K (Fig. 4e). In SEC experiments M6R, M6WT and M6K eluted similarly, although the elution time increased with increasing Arg content (Fig. 4f) despite M6R having a larger mass and being more helical. The slower elution could arise from a smaller hydration shell for M6R compared to M6WT and M6K, or from stronger interactions of Arg with the column. M6WT, M6K and M6R were all found to be monomeric by AUC (Fig. 4g–i). Thus, modulating K/R content in a natural SAH has similar effects on helicity and thermal behaviour as discovered for the *de novo* SAHs, EK3 and EK2R1. Interestingly, increasing the Arg content in this case did not promote aggregation. This may be due to the non-repetitive nature of the sequence compared to *de novo* polypeptides, and more diverse content of amino acids other than E, K and R.

**E–R and E–K pairings have different properties in protein X-ray crystal structures.** The data presented above show that the AEEEXXX repeat pattern is suitable for designing model SAHs, with X = R increasing stability, albeit in small doses, and X = K important for solubility. What gives rise to the increased stability with respect to thermal unfolding in Arg-containing sequences? We addressed this by analysing salt bridges in α-helices of the PDB, and by MD simulations of our *de novo* polypeptide designs. Throughout, a salt bridge was assigned for an E–R (or E–K) pair in a helix if the centroid of Glu Oε1 and Oε2 atoms was <4 Å from any of the Nε, NH1 or NH2 atoms in Arg (or from the Nζ atom in Lys).

The PDB analysis showed that E–R pairs were more frequent than expected by chance from the joint probabilities of individual amino acids (Supplementary Table S1), as previously described for E–K<sup>17</sup>. Overall, the total number of E–R pairs are similar to those reported for E–K pairs (see Table S8 in ref. 17). The preference for E–R pairs occurred in the order E → R(+3) > R → E(+4) > E → R(+4) > R → E(+3) (observed/expected in Supplementary Table S1). In contrast, E–K pairs<sup>17</sup>, showed a preference for +4 pairs over +3 pairs in the order K → E(+4) > E → K(+4) > K → E(+3) ≈ E → K(+3).



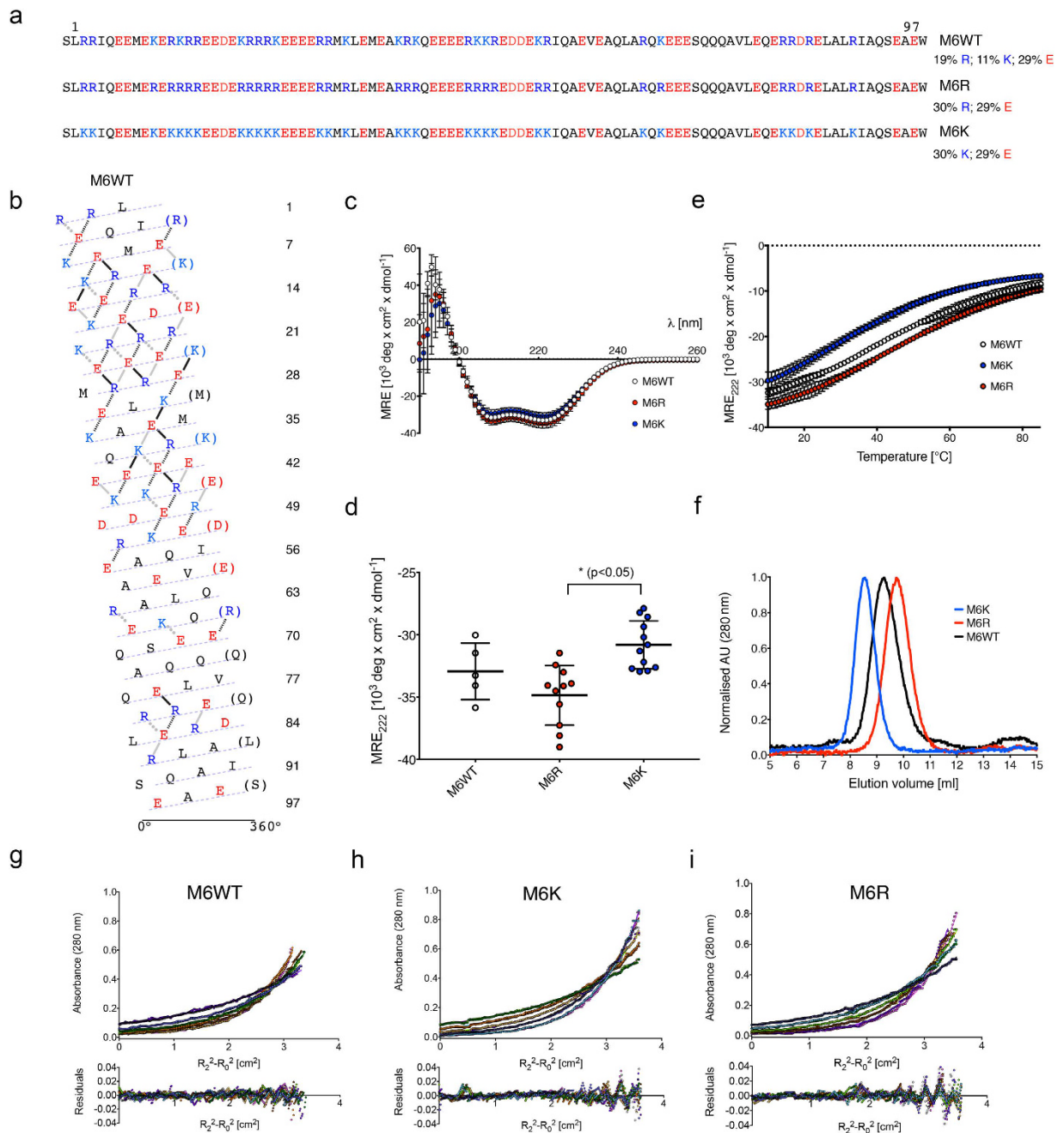
**Figure 2. Properties of EK3 and ER3.** (a) CD spectrum (pH 7.4, 10°C, 10–20  $\mu\text{M}$  *de novo* polypeptide in 100 mM NaCl) of EK3. (b) Thermal denaturation of EK3. The mean residue ellipticity (MRE) at 222 nm is shown as a function of temperature (10–85°C, pH 7.4, 100 mM NaCl). (c) MRE at 222 nm as a function of salt concentration (pH 7.4, 10°C), and (d) MRE at 222 nm as a function of pH (100 mM NaCl, 10°C). Mean ( $\pm$ S.D.) values are plotted. (e) AUC data for EK3 showing the data (circles) and single ideal species fits (solid lines), together with the residuals of the fits. Each colour corresponds to a different speed. The fit to the AUC data for EK3 suggests a mass of 13.3 kDa,  $1.13 \times$  that predicted (11.8 kDa). (f) Results for size exclusion chromatography of EK3. For comparison, the results for a coiled-coil protein (15 heptad construct from  $\beta$ -cardiac myosin-2 tail: 15H CC, with a dimer mass of 25 kDa) and a globular protein with a similar molecular mass (ribonuclease A/RNase A, 13.7 kDa) are also shown. (g) Solubility of ER3 (15  $\mu\text{M}$  in 100 mM NaCl) over a range of pH, as measured by turbidity measurements at 320 nm. (h) CD spectrum (pH 3.5, 10°C, 10–20  $\mu\text{M}$  *de novo* protein in 100 mM NaCl) of ER3. (i) Thermal denaturation of ER3 at pH 3.5. Conditions are the same as in (h).



**Figure 3. Properties of EK2R1 and EK1R2.** (a) CD spectra (pH 7.4, 10 °C, 10–20  $\mu$ M *de novo* polypeptide, 100 mM NaCl) for EK2R1 (light green) and EK1R2 (open circles). The data from Fig. 2 for EK3 (green) is repeated here to enable comparison. (plots show mean  $\pm$  S.D.) (b) Thermal denaturation of EK3, EK2R1 and EK1R2. The  $MRE_{222}$  values at each temperature are shown (mean  $\pm$  S.D.). Conditions as in (a). The profile for EK1R2 exhibits greater cooperativity than EK3. (c) MRE at 222 nm as a function of salt concentration (pH 7.4, 10 °C) for EK3 and EK2R1. Mean  $MRE_{222}$  values are plotted ( $\pm$  S.D.) (d) MRE at 222 nm as a function of pH (100 mM NaCl, 10 °C). Mean  $MRE_{222}$  values are plotted ( $\pm$  S.D.) (e) AUC data for EK2R1 showing the data (circles) and single ideal species fits (solid lines), together with the residuals of the fits. Each colour corresponds to a different speed. The fit to the AUC data for EK2R1 indicates a mass of 12.7 kDa,  $1.04\times$  that predicted (12.2 kDa). (f) Results for size exclusion chromatography of EK2R1 and EK1R2. For comparison, the results for EK3, a coiled-coil protein (15 heptad construct from  $\beta$ -cardiac myosin-2: 15 H CC) and a globular protein with a similar molecular mass (ribonuclease A/RNase A, 13.7 kDa) are also shown.

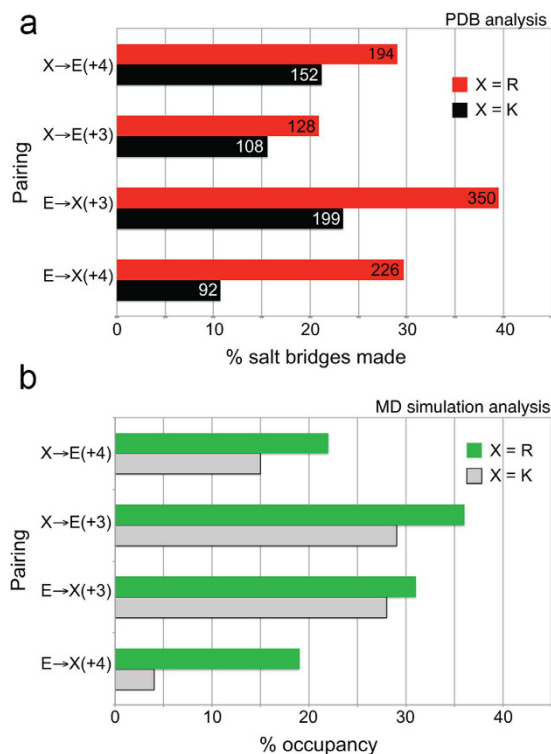
The analysis also showed that E–R pairs form salt bridges more frequently than E–K pairs (Fig. 5a, Supplementary Table S1). For example, 40% of E  $\rightarrow$  R(+3) pairings formed salt bridges, compared to 23% of E  $\rightarrow$  K(+3) pairings. Similar trends were observed for the remaining three types of pairings. E  $\rightarrow$  R(+3) pairs, being the most over-represented and the most likely to actually form salt bridges, are likely to be the most stabilizing of the four E–R pairing options. This contrasts with E–K pairs, where K  $\rightarrow$  E(+4) is thought to be the most stabilizing<sup>17</sup>.

Several specific side-chain rotamer conformations predominate for Arg and Glu in salt bridges. For individual amino acids (black bars in Supplementary Fig. S2), Arg has two preferred rotamer ( $\chi_1, \chi_2$ ) combinations,  $g^-t$  (38%) and  $tt$  (25%), and two minor conformers ( $tg^+$ , 10% and  $g^-g^-$ , 5%). These two major Arg rotamers are the same as those found previously for Lys:  $g^-t$  (39%) and  $tt$  (32%)<sup>17</sup>. Glu has three preferred  $\chi_1, \chi_2$  combinations,  $g^-t$  (36%),  $tt$  (25%) and  $g^-g^-$  (19%). In E  $\rightarrow$  R(+3) salt bridges, the  $tg^+/tt$  (E/R) combination (Supplementary Table S2)



**Figure 4. Properties of re-engineered myosin-6 SAHs.** (a) Sequences of the proteins M6WT, M6R and M6K, together with the percentage of Arg, Lys and Glu content. (b) Helical net plot<sup>2</sup> for the M6WT SAH. (c) CD spectra (pH 7.4, 10 °C, 10–20  $\mu$ M protein, 100 mM NaCl) for each of the three proteins. (d) Mean values ( $\pm$ S.D.) of repeat measurements for MRE<sub>222</sub> values at 10 °C. The mean values for M6R and M6K were significantly different ( $p < 0.05$ ). (e) Thermal denaturation. The MRE<sub>222</sub> values at each temperature are shown. (f) SEC of each of the three proteins. AUC data for (g) M6WT, (h) M6K and (i) M6R. Presentation as per Fig. 2e. The fits to the data indicate a mass of 12.8 kDa for M6WT: ( $1.02 \times$  calculated monomer mass of 12.5 kDa); 12.6 kDa for M6K ( $1.05 \times$  calculated monomer mass of 12.0 kDa) and 12.5 kDa for M6R ( $0.98 \times$  monomer mass of 12.8 kDa).

was the most prevalent. This combination has the second most-preferred conformation for Arg, and a disfavoured conformation for Glu (Supplementary Fig. S2).  $E \rightarrow R(+3)$  salt bridges also utilized other rotamer combinations  $g^-g^-/tt$ ,  $g^-g^-/g^-t$  and  $tg^+/g^-t$  (21%, 14% and 14%, respectively, Supplementary Table S2), which mostly draw on the two major Arg conformations and the more-preferred Glu conformation  $g^-g^-$ . Overall, the majority of the  $E \rightarrow R(+3)$  pairings had Glu in a less favoured conformation. The entropic penalty that this incurs is likely to be offset by the use of more-favourable Arg conformations and the multiple modes available to make profitable salt bridges.



**Figure 5. Structural analysis of helices in the PDB, and equilibrium properties at 300 K from MD simulations.** (a) PDB analysis. For each pairing, the number of salt bridges observed in helices in the PDB is shown on the bar. The bar length represents this number of salt bridges observed as a percentage of pairings identified from protein sequence (see Supplementary Table S1). The values for E–K pairings are taken from Baker *et al.*<sup>17</sup>. (b) MD simulation analysis. For each pairing mode, the bar length represents the occupancy of salt bridges observed as a percentage of all pairings from simulations of EK3, EK2R1, EK1R2 and ER3.

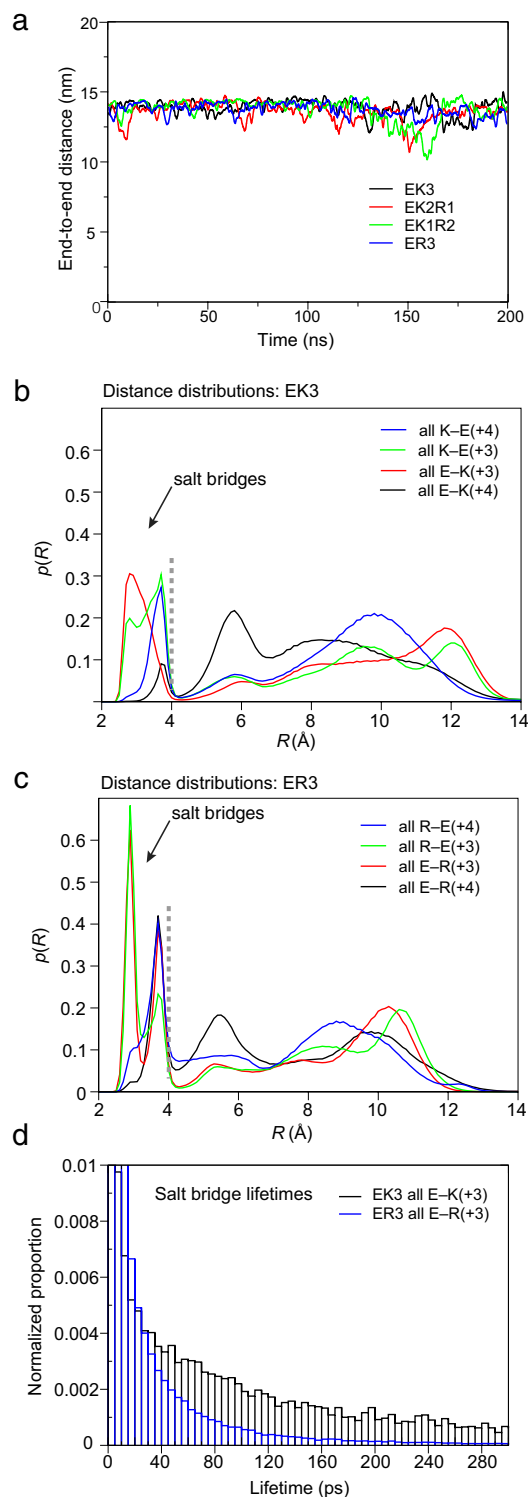
Rotamer combinations utilized by E–R pairs to form salt bridges were less dominated by a single combination, in contrast to E–K pairs<sup>17</sup>. The frequency of the favoured  $tg^+/tt$  rotamer combination in  $E \rightarrow R(+3)$  (47%) was lower than the dominant combination ( $g^-g^-/g^-t$ ) for  $E \rightarrow K(+3)$  (77%, Supplementary Table S2). Similar reductions in the predominance of a single rotamer combination were also found for the  $E \rightarrow R(+4)$  and  $R \rightarrow E(+4)$  pairs (Supplementary Table S2, Supplementary Fig. S2). The contribution of  $tt/g^-t$  reduced from 68% in  $E \rightarrow K(+4)$  to 34% in  $E \rightarrow R(+4)$ ; and 75% in  $K \rightarrow E(+4)$  to 56% in  $R \rightarrow E(+4)$  for  $g^-t/tt$ .  $R \rightarrow E(+3)$  and  $K \rightarrow E(+3)$  salt bridges were spread in similar fashion across four rotamer combinations without one particularly dominant contributor.

Taken together, the PDB analysis demonstrates that E–R pairs are more prevalent and display a greater range of rotamer conformations than E–K pairs to make salt bridges. The increased number of salt bridges made by E–R pairs is probably related to the Arg side chain being longer and the multi-dentate guanidinium group having more possibilities of interactions compared with the amino side chain of lysine. The increased variability in their rotamer conformations suggests that the E–R pairs are more dynamic than E–K, whilst still productively engaging in salt-bridge interactions. Indeed, structural superpositions of helices containing salt bridges (Supplementary Fig. S3) revealed significant variation in Glu and Arg side chain conformations for each of the four E–R arrangements.

#### Molecular dynamics simulations show different dynamics for E–K and E–R salt bridges.

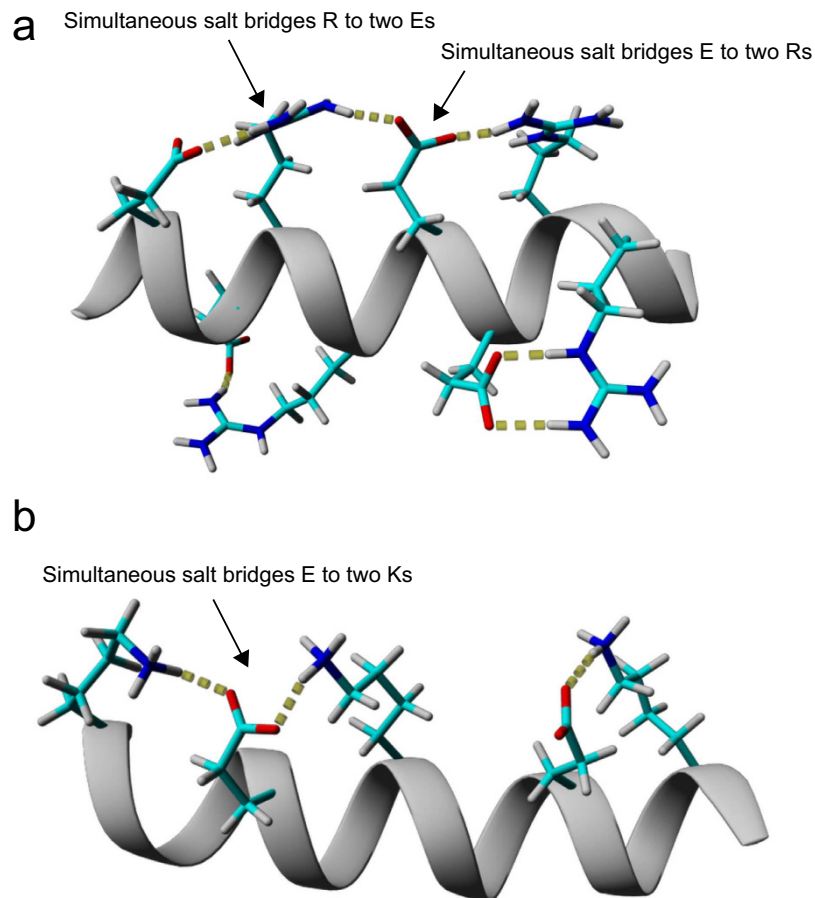
To investigate the dynamic behaviour of E–K and E–R salt bridges, we performed MD simulations. The *de novo* polypeptides all remained as near-complete continuous  $\alpha$ -helices ( $\geq 95\%$  helix), which were elongated (not bundled) throughout the 200 ns simulations (Fig. 6a). Clear transitions between states in which the side chains for residues in E–K or E–R pairs were either in close proximity (*i.e.*, forming a salt bridge) or well-separated (non-interacting) were observed during the simulations, as illustrated by the trajectory for K49 in EK3 (Supplementary Fig. S4). Calculating the distances between side chains in all available pairings throughout the simulation trajectories, showed salt bridge occurrence as distinct peaks in probability below 4 Å (Fig. 6b and c, Supplementary Fig. S5). This agrees well with the cut-off for salt bridge assignment of 4 Å used in the PDB database interrogation. Peaks observed at  $\sim 2.8$  Å and  $\sim 3.7$  Å arise from salt bridges that utilize different rotamer combinations.

MD simulations showed E–K salt bridges were less highly occupied than E–R (Fig. 5b, Supplementary Table S3).  $K \rightarrow E(+4)$  pairs were more likely to form salt bridges than  $E \rightarrow K(+4)$  pairs (Fig. 5b, Supplementary Table S3) as previously shown experimentally<sup>17</sup>, and in good agreement with the PDB analysis (Fig. 5a). The % salt bridge



**Figure 6. Characteristics of *de novo* proteins during MD simulations.** (a) Plots of the end-to-end distance for each protein during the simulation. (b) Normalized distance distributions in EK3 for all K  $\rightarrow$  E(+4), K  $\rightarrow$  E(+3), E  $\rightarrow$  K(+3) and E  $\rightarrow$  K(+4) pairings. Pairing distances are between the centre-of-mass (CoM) of the Glu O $\epsilon$  atoms and the N $\zeta$  atom in Lys. The plots were generated from histograms of all salt bridge distances for pairings of each type using 0.1  $\text{\AA}$  bins. The grey dashed line at 4  $\text{\AA}$  represents the cut-off point for considering a pairing as a salt bridge. Conformations with distances up to 4  $\text{\AA}$  (*i.e.* to the left side of this line) are considered to have formed a salt bridge. The peaks at 5.6–6.0  $\text{\AA}$  can be attributed to ‘indirect’ salt bridges separated by one water molecule, which are not included in the analysis<sup>46</sup>. (c) Normalized distance distributions in ER3 for all R  $\rightarrow$  E(+4), R  $\rightarrow$  E(+3), E  $\rightarrow$  R(+3) and E  $\rightarrow$  R(+4) pairings. The pairing distance is the smallest of the distances between the CoM of the Glu O $\epsilon$  atoms and each of the three side chain N atoms in Arg. (d) A normalized histogram plot of salt bridge lifetimes (5 ps bins), for all E  $\rightarrow$  K(+3) and E  $\rightarrow$  R(+3) salt bridges in EK3 and ER3, respectively.





**Figure 7. Salt bridge network formation in ER3 and EK3.** Example snapshots from simulations of ER3 (**a**) and EK3 (**b**) displaying salt bridges between selected residue side chains.

occupancy for  $R \rightarrow E(+4)$  and  $E \rightarrow R(+4)$  pairs was more similar (Fig. 5b, Supplementary Table S3), also in agreement with the PDB analysis (Fig. 5a). The only difference in trends we observed between the PDB results and the MD simulations is that the occupancy of  $R \rightarrow E(+3)$  and  $K \rightarrow E(+3)$  salt bridges is higher than might be expected from the observed % of salt bridges made for these pairs in the PDB (Fig. 5). These salt bridges may be underrepresented in the helices available in the PDB, which lacks high-resolution structures of SAHs. It is worth pointing out that the PDB analysis does not make any conclusions about the strength of salt bridges, as the presence of a salt bridge is defined through geometry as explained above (*i.e.* salt bridges are either present, or absent), while in MD simulations, occupancy can be used as a proxy for strength of charge–charge interaction.

Strikingly, while MD simulations showed that E–R salt bridges were more highly occupied, their average lifetimes were shorter than those for E–K (Fig. 6d, Supplementary Table S3). For example, the average lifetime for all  $E \rightarrow R(+3)$  salt bridges was only 30 ps compared to 159 ps for  $E \rightarrow K(+3)$  when averaged over all sequences. However, the number of  $E \rightarrow R(+3)$  salt bridge formation events was 10 times higher than for  $E \rightarrow K(+3)$  (Supplementary Table S3), accounting for the higher occupancy of E–R pairs. Simulations performed on the SAH from myosin-6 (M6WT) and its Lys- and Arg-only mutants (M6K and M6R) gave very similar results to those of the *de novo* sequences in terms of salt-bridge occupancies and lifetimes (Supplementary Table S4).

Simulations also show that simultaneous salt bridges involving Arg (particularly “E–R–E” networks) form more frequently than Lys (Fig. 7, Supplementary Table S5). This may also help to explain the higher contribution to stability that Arg provides. Arg simultaneously forms salt bridges in both directions along the helix ( $R \rightarrow E(-4)$  &  $E(+3)$  or  $R \rightarrow E(-3)$  &  $E(+4)$  combinations, *i.e.* “E–R–E”) up to 10% of the time (Supplementary Table S5). In contrast, Lys only interacts with two Glu residues that are both C-terminal to the Lys ( $K \rightarrow E(+3)$  &  $E(+4)$ ) (Supplementary Table S5, not shown in Fig. 7). Other simultaneous salt bridges ( $X \rightarrow E(-4)$  &  $E(+4)$  and  $X \rightarrow E(-3)$  &  $E(+3)$ , for  $X = K$  or  $R$ ) were not significantly populated, in agreement with the experimental finding of non-cooperativity in an alanine-based peptide with a  $K \rightarrow E(-4)$  &  $E(+4)$  triplet pattern<sup>18</sup>. Glu residues are also able to form simultaneous salt bridges with two Lys or Arg partners (Supplementary Table S4). With Lys, only those using the  $E \rightarrow K(-4)$  &  $K(+3)$  combination formed (7–10% in EK3, EK2R1 and EK1R2). However, with Arg these salt bridges tended to form in the same direction along the helix (~10% for both  $E \rightarrow R(-4)$  &  $R(-3)$  and  $E \rightarrow R(+3)$  &  $R(+4)$ ), although the  $E \rightarrow R(-4)$  &  $R(+3)$  combination was also populated (~5%). Intermediate occupancies were observed for simultaneous salt bridges of Glu to one Arg and one Lys in EK2R1 and EK1R2.

Overall, MD simulations show that E–R salt bridges form more frequently than E–K salt bridges but have significantly shorter lifetimes. E–R salt bridges are more varied in terms of conformational freedom than E–K salt bridges. Additionally, simultaneous salt bridges involving Arg (particularly “E–R–E” networks) tend to form more frequently than those involving Lys.

## Discussion

Here we have successfully designed, and tested a range of polypeptides to determine the relative contributions of Lys and Arg to the stability of SAHs. We determined that: (i) the only *de novo* polypeptides to exhibit the behaviour typical of SAHs were EK3 and EK2R1, with the inclusion of a single Arg residue in EK2R1 increasing helicity and stability; (ii) any further increase in the Arg content in the *de novo* polypeptides (EK1R2 and ER3) promoted aggregation; and (iii) substituting all the Lys residues with Arg, in the naturally occurring myosin-6 SAH increased helicity and stability with respect to thermal unfolding, while substituting Arg with Lys, had the opposite effect. Thus, Lys and Arg are not completely interchangeable but make distinct and significant contributions to the stability of SAHs.

The PDB analysis and MD simulations rationalize these results. In general, they reveal that E–R pairs are more likely to form salt bridges than E–K, by utilizing multiple rotamer conformations and more binding ‘modes’, but that the lifetimes of E–R salt bridges are shorter. The most stabilizing E–R pairing (E → R(+3)) uses a wider range of side chain combinations than those for the most stabilizing E–K pairing (K → E(+4)), (*i.e.* fewer side chain rotamers need to be fixed to form salt bridge interactions for E → R(+3)). This demonstrates that the multiple binding modes of E–R pairs are separated by marginal energy barriers and they rapidly interconvert, suggesting an important entropic contribution to the stability of the helical state. For E–K, salt bridges formed by K → E(+4) pairs use fewer side-chain rotamer combinations than other pairings, and these better-defined K → E(+4) salt bridges contribute a larger favourable enthalpy to the free-energy of helix formation<sup>17</sup>. Thus, EK2R1, in which a single Lys is substituted by Arg in each 7 residue repeat (or substituting in Arg for Lys in the SAH from myosin-6), results in a more helical and thermally stable SAH than EK3 (or M6WT). This is important, as the ratio of K:R in naturally occurring SAHs varies<sup>2</sup>. Therefore, the relative proportions of K and R in these domains are likely to have biological relevance, especially given the possible variety of functions for SAHs in biological systems<sup>2</sup>.

The more promiscuous nature of individual E–R salt bridges also helps to explain why increasing levels of Arg in *de novo* polypeptides increased their tendency to aggregate. Shorter salt bridge lifetimes, as well as the multi-dentate nature of the guanidinium group, are likely to increase the chances that E–R pairings will form between molecules, and not just along the helix of a single molecule, explaining the tendency for *de novo* polypeptides with high levels of Arg to aggregate. Moreover, the guanidyl group of Arg can exhibit weak hydration<sup>26</sup>, there can be significant pairing between guanidine-terminated side chains in polyarginine (but not between amine-terminated side chains in polylysine)<sup>27</sup>, and the Arg side chain can be hydrophobic above and below the plane of the guanidyl group, allowing the stacking of Arg residues<sup>28</sup>. Thus, Arg side chains may stack within or between proteins, promoting oligomerization<sup>29</sup>. These considerations will be important for the future design of *de novo* helical polypeptides. While Arg can increase the stability of SAHs, extensive Arg “patches” in long polypeptides should be avoided, at least in designs with a regular 7-residue repeat as used herein.

Reducing the number of charged residues per repeat unit, and replacing them with Ala to maintain a high helical propensity, was not successful for generating *de novo* SAHs. Exchanging pairs of charged residues in EK3 for alanine results in an amphipathic helix pattern, with a predominantly alanine-based face, which is likely to promote protein–protein association mediated by hydrophobic contacts<sup>22</sup>. Not surprisingly then, the majority of these *de novo* polypeptides formed multimeric helical complexes. Alacoils are naturally occurring anti-parallel coiled coils in which alanine is the predominant residue in either the ‘a’ or ‘d’ positions of the heptad sequence repeat, and the two helical strands are closely spaced compared to other coiled coils such as leucine zippers<sup>30</sup>. We suspect that the hydrophobic seam formed by alanine in EK2 results in a dimer structure such as this, its stability enhanced by inter- and/or intra-chain charged interactions outside the hydrophobic seam.

The MD simulations have shown that modelled SAHs (known to be helical experimentally) are kinetically (very) stable. Such remarkable kinetic stability makes it currently unviable to explore equilibrium properties of these sequences computationally, particularly using fully solvated models. Thus, here we have focused on the contributions of K versus R to their helical state, and do not consider the effect of substituting K with R on the unfolded state. However, we would argue that the unfolded state is not relevant for our interpretation. Unfolded states are likely to be dominated by expanded coil structures due to the high charge content and charge distribution pattern of the sequence<sup>31</sup>. Repulsion between like-charge residues will limit the accessible conformation space available to the unfolded state and thus limit the entropic benefit of unfolding. Despite this, there will exist an expanded range of possible stabilizing salt bridge interactions in unfolded, non-helical structures, with Arg again interacting more dynamically than Lys. However, we would argue that helical forms (Arg-rich sequences in particular) benefit more through positioning their side chains to avoid charge–charge repulsion, and in their ability to dynamically rearrange and make multiple simultaneous salt bridge pairings without the need for concurrent rearrangements of the backbone. This avoids the need for disruption to the hydrogen-bonded network that makes up the core of the helix, or greater solvent exposure of hydrophobic methylene groups.

The simple repeating sequences used in EK3 and EK2R1 allows them to be easily customized for many potential synthetic biology applications. These domains have the potential to be used as force sensors, helical spacers (*i.e.* inserted between two protein domains), and/or to modulate and report on protein–protein interactions in both *in vitro* and *in vivo* applications (as reviewed in ref. 3). Choosing Lys, or a mixture of Lys and Arg subtly alters stability to provide flexibility in design. Moreover, *de novo* polypeptides can be engineered to be any length, and Ala can be replaced with other residues (*e.g.* cysteine to allow fluorescent labelling).

In summary, we have designed, expressed and characterized long, highly stable model SAHs (*de novo* polypeptides), using just four amino acids, Ala, Glu, Lys and Arg. These simple designs have enabled us to gain

significant insight into the mechanism by which these domains are stabilized. We have discovered Lys and Arg are not completely interchangeable, in that E–R pairings are more likely to form salt bridges than E–K thus increasing the stability of SAHs, and that E–R salt bridges are more dynamic but their promiscuous nature could contribute to the aggregation of SAHs, when Arg is increased to high levels. These data suggest that naturally occurring SAHs are likely to have different properties, depending on their relative K:R content, and provide guidelines for engineering long customizable SAHs.

## Materials and Methods

**Expression constructs.** DNA sequences encoding model SAH domains were synthesized (GeneArt; GenScript) and subcloned into the pET28a SUMO vector (received as a kind gift from Dr Thomas Edwards) to introduce an N-terminal His-tag and SUMO fusion protein for increased expression and solubility as described<sup>5</sup>. Full sequences for these *de novo* polypeptides are provided (Fig. 1a). Each sequence contains an additional N-terminal serine residue carried over as a result of SUMO cleavage and a C-terminal tryptophan residue for UV absorbance concentration measurements. The full sequence for the SAH domain from myosin-6 (human, Uniprot ID Q9UM54, residues 926–1022) together with the mutants generated (in which Lys is substituted for Arg, and *vice versa*) is provided (Fig. 4).

**Protein expression and purification.** All proteins were expressed in *Escherichia coli* BL21 Rosetta 2 (Novagen) and purified using a Ni-NTA (cOMplete His-Tag Resin, Roche) affinity chromatography column. Bacterial pellets were re-suspended in ~10 ml of buffer A (300 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM imidazole, 0.1% Tween-20, 1 mM EDTA, 0.2 mM PMSE, 0.03% NaN<sub>3</sub>, pH 8.0 with NaOH) and sonicated on ice (Sonic Vibra-Cell sonicator, 50% amplitude, 6 cycles: 10 s on/off). Lysates were centrifuged (30,000 g, 20 min, 4 °C) and supernatants applied on pre-equilibrated gravity-flow columns (1 ml of resin). Columns were washed with 50 ml of buffer B (300 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 20 mM imidazole, 0.1% Tween-20, 1 mM EDTA, 0.2 mM PMSE, pH 8.0). Proteins were eluted in 8 × 1 ml fractions in buffer C (300 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 200 mM imidazole, 0.03% NaN<sub>3</sub>, 0.2 mM PMSE, pH 8.0) and analysed by SDS-PAGE (12% gels). Proteins were then dialyzed against 150 mM NaCl (300 mM in the case of EK1), 20 mM Tris-HCl, 1 mM DTT, pH 8.0 and proteolysed for 2 h at room temperature, using ULP1 recombinant SUMO protease in a substrate to enzyme ratio 100:1. SUMO protease is a recombinant fragment of ULP1 (Ubiquitin-like-specific protease 1) from *Saccharomyces cerevisiae*. It is highly specific for the SUMO fusion protein, recognizing the tertiary structure of SUMO rather than an amino acid sequence<sup>32,33</sup>. EK3, EK2R1, EK1R2 and EK2 were separated from SUMO on 5 ml Q sepharose columns using an AKTA system. Buffers used were: 20 mM Tris-HCl, pH 8.0, 0.03% NaN<sub>3</sub> (Buffer A); 1 M NaCl, 20 mM Tris-HCl, pH 8.0, 0.03% NaN<sub>3</sub> (Buffer B); salt gradient: 100–600 mM. The purest fractions were combined and concentrated resulting in a 1–2 mg/ml protein solution and a typical yield of 2.5–5 mg per litre of *E. coli* culture. Purified protein was dialyzed against 100 mM NaCl, sodium phosphate (7.7 mM Na<sub>2</sub>HPO<sub>4</sub>/2.3 mM NaH<sub>2</sub>PO<sub>4</sub>), pH 7.4, and snap-frozen in liquid nitrogen for long term storage at –80 °C. An alternative method of purification was used for ER3 and EK1, which showed a high level of aggregation upon removal of SUMO. Pellets of aggregated proteins were washed 3x with 300 mM NaCl, 20 mM Tris-HCl, pH 8.0 and re-suspended in 100 mM NaCl, 10 mM sodium citrate, pH 3.5 (ER3) or 10 mM NaCl, 5 mM Tris-HCl, pH 7.4 (EK1) and dialyzed. Due to the high level of aggregation we avoided freezing these proteins and only fresh preparations were used. For CD experiments at different pH, the following buffer solutions were used – pH 2.4: 100 mM NaCl, 10 mM glycine-HCl; pH 3.5 and pH 5.0: 100 mM NaCl, 10 mM sodium citrate/citric acid; pH 10: 100 mM NaCl, 10 mM Tris; pH 12: 100 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>-NaOH. The 15 heptad (15 H CC) and 11 heptad (11 H CC) coiled-coil fragments from human β-cardiac myosin-2 tail were expressed and purified as described previously<sup>34</sup>. Protein concentration was measured by absorption at 280 nm. Absorption coefficients were obtained from ProtParam software (<http://web.expasy.org/protparam/>). Protein concentrations used were in the range 10–40 μM.

**CD spectroscopy.** CD measurements were performed on an Applied Photo Physics Chirascan CD spectropolarimeter with a 0.1 cm path length quartz cuvette in buffers as specified in the Protein expression and purification section. Data were collected every 1 nm with a scan rate of 120 nm/min; for each measurement two scans were recorded. Data presented are averaged from at least two separate measurements of different protein preparations. Thermal measurements were performed in a temperature range from 10 to 85 °C with a 0.7 °C/min heating rate, data acquisition every 1 °C. The mean residue molar ellipticity (MRE) of proteins was calculated as described<sup>35</sup>. Here we use the units of deg × cm<sup>2</sup> × dmol<sup>-1</sup>, rather than the units deg × cm<sup>2</sup> × dmol<sup>-1</sup> res<sup>-1</sup>. The helical content of proteins was calculated from values of the amide n → π\* transition at 222 nm ([MRE<sub>222</sub>]), as previously described<sup>35</sup>.

**Size exclusion chromatography.** Size exclusion chromatography was used to estimate the shape of the *de novo* polypeptides, and in particular to determine how elongated they were<sup>36</sup>. This technique separates molecules on the basis of their molecular size, and the time it takes for these molecules to elute from the column is inversely correlated with their equivalent hydrodynamic radius (Stokes radius, R<sub>s</sub>)<sup>37</sup>. The R<sub>s</sub> for an elongated protein is larger than that for a globular protein of the same mass and hence the elongated proteins elute earlier from the column.

A GE Healthcare Tricorn 10/20 column was packed with Superdex 75 resin and calibrated using the GE Healthcare gel filtration calibration kit, which comprises albumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). The elution profiles of the *de novo* polypeptides of interest were obtained by injecting 200 μl of protein sample within a concentration range of 20–40 μM in column buffer (150 mM NaCl, 10 mM sodium phosphate, 0.03% NaN<sub>3</sub>, pH 7.4) onto the column at a flow rate of 0.5 ml/min, using an AKTA system. The column exclusion volume was 6.3 ml (obtained using dextran blue).

**Analytical ultracentrifugation.** Sedimentation-equilibrium experiments by analytical ultracentrifugation were performed in triplicate using a Beckman Optima XL-A analytical ultracentrifuge at 20 °C with an AN50 8-place rotor, and cells with epon 6-channel centrepieces and quartz windows. Samples were prepared in 100 mM NaCl, 7.7 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.3 mM NaH<sub>2</sub>PO<sub>4</sub> and 4.61 mM NaN<sub>3</sub>. The samples were centrifuged in the speed range 18,000–42,000 rpm and data collected in increments of 4,000 rpm. Data were fitted to single ideal species using Ultrascan II<sup>38</sup> and the confidence limits obtained by Monte Carlo analysis of the fits. Representative data for one channel for each sample are shown.

**E–R pairings in the PDB.** These data were culled from the same helix dataset as used previously for E–K pairings, and analysed in an analogous manner<sup>17</sup>. The dataset contains helices of 12 amino acids or longer detected among 2,775 sub-1.6 Å resolution X-ray crystal structures. Interactions involving any of the first four residues of each helix are classed as ‘N-terminal’; those just involving residues at least four positions in sequence away from the N and C termini are ‘central’ and those involving the last four residues are ‘C-terminal’. The numbers of R → E(+4), R → E(+3), E → R(+3) and E → R(+4) pairs were identified. Expected numbers of pairs were estimated using the occurrence of each residue in the whole dataset. A salt bridge was considered to be formed for a pair if the centroid of Glu Oε1 and Oε2 atoms was <4 Å from any of the Arg Nε, NH1 or NH2 atoms.

$\chi_1$ ,  $\chi_2$  side chain rotamer distributions for central salt bridge and non-salt bridge R → E(+4), R → E(+3), E → R(+3) and E → R(+4) pairs were categorized as follows:  $t$ ,  $\chi \geq 120^\circ$  or  $\chi \leq -120^\circ$ ;  $g^+$ ,  $0^\circ \leq \chi < 120^\circ$ ;  $g^-$ ,  $-120^\circ < \chi < 0^\circ$  for Glu and Arg residues in all helices. Theoretical rotamer combinations were modelled in PyMOL (<http://www.pymol.org>) and salt bridge potential assigned if the centroid of Glu Oε1 and Oε2 atoms was <4 Å from any of the Arg Nε, NH1 or NH2 atoms and no atoms were closer than 2.5 Å to avoid steric clashes. Rotamer combinations were identified using Promotif<sup>39</sup>. The procedure for the RMSD calculation uses the multi-structure fitting algorithm in ProFit (<http://www.bioinf.org.uk/software/profit/>).

**Modelling.** Explicit solvent modelling was performed using the CHARMM 36 force field<sup>40</sup> parameters with TIP3P water. EK3, ER3, EK2R1 and EK1R2 structures were built as perfect  $\alpha$ -helices, their N and C termini capped with acetyl (ACE) and methylamine (CT3) groups, respectively. Perfectly  $\alpha$ -helical conformers were created by setting internal dihedral angles to  $\Phi = -57^\circ$  and  $\Psi = -47^\circ$ . Structures were energy minimized for 1,000 steepest decent steps in vacuum using CHARMM<sup>41</sup>. Using VMD<sup>42</sup>, a 1.5 nm surround of water molecules (EK3: 15,310 water molecules; ER3: 17,350; EK2R1: 15,924; EK1R2: 13,505) and Na<sup>+</sup> and Cl<sup>-</sup> ions were added to neutralize the protein and give a NaCl concentration of ~150 mM. A further minimization (10,000 steps), 0–300 K heating protocol and short pre-equilibration was performed using NAMD<sup>43</sup> (100,000 steps). Data are taken from 200-ns simulations run using NAMD at 300 K. The timestep used was 2 fs and trajectory frames were recorded every 500 steps. Simulations were performed in an equivalent manner for the SAH domain from myosin-6 and its K-only and R-only mutants (M6WT, M6K, M6R).

Wordom<sup>44</sup> was used to analyse the simulation trajectories. The secondary structure of the protein was assigned for each timeframe using the DSSPcont criteria<sup>45</sup>. This was then used to calculate the helicity (or average helical fraction) of the protein overall. For the salt bridge analysis, the distance between Lys Nζ and the centroid of (Glu Oε1, Oε2) was calculated for each potential K → E(+4), K → E(+3), E → K(+3), and E → K(+4) pair, and the distances between each of the three Arg NH1/NH2/Nε and the centroid of (Glu Oε1, Oε2) were calculated for each potential R → E(+4), R → E(+3), E → R(+3), and E → R(+4) pair. As with the analysis of helices in the PDB, the definition of a salt bridge pairing at any frame of the trajectory required any of the resulting distances described to be less than 4 Å.

## References

- Gaspari, Z., Suveges, D., Perczel, A., Nyitrai, L. & Toth, G. Charged single  $\alpha$ -helices in proteomes revealed by a consensus prediction approach. *Biochim. Biophys. Acta* **1824**, 637–646 (2012).
- Peckham, M. & Knight, P. J. When a predicted coiled coil is really a single  $\alpha$ -helix, in myosins and other proteins. *Soft Matter* **5**, 2493–2503 (2009).
- Swanson, C. J. & Sivaramakrishnan, S. Harnessing the Unique Structural Properties of Isolated  $\alpha$ -Helices. *J. Biol. Chem.* **289**, 25460–25467 (2014).
- Knight, P. J. *et al.* The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J. Biol. Chem.* **280**, 34702–34708 (2005).
- Wolny, M. *et al.* Stable single  $\alpha$ -helices are constant force springs in proteins. *J. Biol. Chem.* **289**, 27825–27835 (2014).
- Baboolal, T. G. *et al.* The SAH domain extends the functional length of the myosin lever. *Proc. Natl. Acad. Sci. USA* **106**, 22193–22198 (2009).
- Sivaramakrishnan, S., Spink, B. J., Sim, A. Y., Doniach, S. & Spudich, J. A. Dynamic charge interactions create surprising rigidity in the ER/K  $\alpha$ -helical protein motif. *Proc. Natl. Acad. Sci. USA* **105**, 13356–13361 (2008).
- Simm, D., Hatje, K. & Kollmar, M. Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single  $\alpha$ -helices (SAH domains). *Bioinformatics* **31**, 767–769 (2015).
- Suveges, D., Gaspari, Z., Toth, G. & Nyitrai, L. Charged single  $\alpha$ -helix: a versatile protein structural motif. *Proteins* **74**, 905–916 (2009).
- Sivaramakrishnan, S. *et al.* Combining single-molecule optical trapping and small-angle x-ray scattering measurements to compute the persistence length of a protein ER/K  $\alpha$ -helix. *Biophys. J.* **97**, 2993–2999 (2009).
- Spink, B. J., Sivaramakrishnan, S., Lipfert, J., Doniach, S. & Spudich, J. A. Long single  $\alpha$ -helical tail domains bridge the gap between structure and function of myosin VI. *Nat. Struct. Mol. Biol.* **15**, 591–597 (2008).
- Wang, E. & Wang, C. L. (i, i + 4) Ion pairs stabilize helical peptides derived from smooth muscle caldesmon. *Arch. Biochem. Biophys.* **329**, 156–162 (1996).
- Samejima, K. *et al.* The INCENP Coil is a Single  $\alpha$  Helical (SAH) Domain that Binds Directly to Microtubules and is Important for CPC Localization and Function in Mitosis. *J. Biol. Chem.* (2015).
- Ulrich, A. K., Seeger, M., Schutze, T., Bartlick, N. & Wahl, M. C. Scaffolding in the Spliceosome via Single  $\alpha$  Helices. *Structure* **24**, 1972–1983 (2016).

15. Huyghues-Despointes, B. M., Scholtz, J. M. & Baldwin, R. L. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci.* **2**, 80–85 (1993).
16. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu<sup>-</sup> ... Lys<sup>+</sup> salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. USA* **84**, 8898–8902 (1987).
17. Baker, E. G. *et al.* Local and macroscopic electrostatic interactions in single  $\alpha$ -helices. *Nat. Chem. Biol.* **11**, 221–228 (2015).
18. Iqbalsyah, T. M. & Doig, A. J. Anticooperativity in a Glu-Lys-Glu salt bridge triplet in an isolated  $\alpha$ -helical peptide. *Biochemistry* **44**, 10449–10456 (2005).
19. Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**, 9668–9676 (1993).
20. Somme, R. F., Sivaramakrishnan, S., Baldwin, R. L. & Spudich, J. A. Helicity of short E-R/K peptides. *Protein Sci.* **19**, 2001–2005 (2010).
21. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
22. Banwell, E. F. *et al.* Rational design and application of responsive [ $\alpha$ ]-helical peptide hydrogels. *Nat. Mater.* **8**, 596–600, (2009).
23. Meuzelaar, H. *et al.* Solvent-Exposed Salt Bridges Influence the Kinetics of  $\alpha$ -Helix Folding and Unfolding. *J. Phys. Chem. Lett.* **5**, 900–904 (2014).
24. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**, 422–427 (1998).
25. Wang, C. L. *et al.* A long helix from the central region of smooth muscle caldesmon. *J. Biol. Chem.* **266**, 13958–13963 (1991).
26. Mason, P. E., Neilson, G. W., Dempsey, C. E., Barnes, A. C. & Cruickshank, J. M. The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. *Proc. Natl. Acad. Sci. USA* **100**, 4557–4561 (2003).
27. Vondrasek, J., Mason, P. E., Heyda, J., Collins, K. D. & Jungwirth, P. The molecular origin of like-charge arginine-arginine pairing in water. *J. Phys. Chem. B* **113**, 9041–9045 (2009).
28. Armstrong, C. T., Mason, P. E., Anderson, J. L. & Dempsey, C. E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci. Rep.* **6**, 21759, doi: 10.1038/srep21759 (2016).
29. Neves, M. A., Yeager, M. & Abagyan, R. Unusual arginine formations in protein function and assembly: rings, strings, and stacks. *J. Phys. Chem. B* **116**, 7006–7013, (2012).
30. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S. & Richardson, D. C. The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Sci.* **4**, 2252–2260 (1995).
31. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **110**, 13392–13397 (2013).
32. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5**, 865–876 (2000).
33. Muller, S., Hoegge, C., Pyrowolakis, G. & Jentsch, S. SUMO, ubiquitin's mysterious cousin. *Nat. Rev. Mol. Cell Biol.* **2**, 202–210 (2001).
34. Wolny, M. *et al.* Cardiomyopathy mutations in the tail of  $\beta$  cardiac myosin modify the coiled-coil structure and affect integration into thick filaments in muscle sarcomeres in adult cardiomyocytes. *J. Biol. Chem.* **288**, 31952–31962 (2013).
35. Greenfield, N. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
36. Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proc. Online* **11**, 32–51, doi: 10.1007/s12575-009-9008-x (2009).
37. Horiike, K., Tojo, H., Yamano, T. & Nozaki, M. Interpretation of the stokes radius of macromolecules determined by gel filtration chromatography. *J. Biochem.* **93**, 99–106 (1983).
38. Demeler, B. In *Modern Analytical Ultracentrifugation: Techniques and Methods* (eds D. J. Scott, S. E. Harding, & A. J. Rowe) 210–229 (Royal Society of Chemistry, (UK), 2005).
39. Hutchinson, E. G. & Thornton, J. M. PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).
40. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi(1)$  and (2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
41. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–38 (1996).
43. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
44. Seeber, M. *et al.* Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **32**, 1183–1194 (2011).
45. Carter, P., Andersen, C. A. & Rost, B. DSSPcont: Continuous secondary structure assignments for proteins. *Nucleic Acids Res.* **31**, 3293–3295 (2003).
46. Dzubiella, J. Salt-specific stability and denaturation of a short salt-bridge-forming  $\alpha$ -helix. *J. Am. Chem. Soc.* **130**, 14000–14007 (2008).

## Acknowledgements

M.W. and M.B. were supported by BBSRC grants BB/I007423/1 (to M.P., E.P., P.J.K. and L.D.) and (BB/M009114/1 to M.P. and P.J.K.); E.G.B. and D.N.W. are supported by a BBSRC/ERASynBio grant (BB/M005615/1); G.J.B. and D.N.W. are supported by the ERC (340764); and D.N.W. is a Royal Society Wolfson Research Merit Award holder (WM140008). L.D. supported is supported by ERC (258259). The Wellcome Trust (WT094232) funded the CD spectropolarimeter. We would like to acknowledge Amy Barker (Leeds) for performing a preliminary AUC analysis on one of the constructs.

## Author Contributions

M.W., M.B., P.J.K., L.D., E.P. and M.P. designed the constructs. M.W. and M.K. expressed and purified the proteins. M.W. characterized the proteins. M.B. and E.P. performed and analysed the MD simulations. G.J.B., E.G.B. and D.N.W. designed and performed the analysis of the PDB. E.G.B. made the AUC measurements. All authors commented on the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Wolny, M. *et al.* Characterization of long and stable *de novo* single alpha-helix domains provides novel insight into their stability. *Sci. Rep.* **7**, 44341; doi: 10.1038/srep44341 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017