








# Design of buried charged networks in artificial proteins

Mona Baumgart<sup>1</sup>, Michael Röpke <sup>1,3</sup>, Max E. Mühlbauer <sup>1,3</sup>, Sam Asami <sup>1</sup>, Sophie L. Mader <sup>1</sup>, Kai Fredriksson<sup>1</sup>, Michael Groll <sup>1</sup>, Ana P. Gamiz-Hernandez <sup>1,2</sup> & Ville R. I. Kaila <sup>1,2</sup>✉

Soluble proteins are universally packed with a hydrophobic core and a polar surface that drive the protein folding process. Yet charged networks within the central protein core are often indispensable for the biological function. Here, we show that natural buried ion-pairs are stabilised by amphiphilic residues that electrostatically shield the charged motif from its surroundings to gain structural stability. To explore this effect, we build artificial proteins with buried ion-pairs by combining directed computational design and biophysical experiments. Our findings illustrate how perturbation in charged networks can introduce structural rearrangements to compensate for desolvation effects. We validate the physical principles by resolving high-resolution atomic structures of the artificial proteins that are resistant towards unfolding at extreme temperatures and harsh chemical conditions. Our findings provide a molecular understanding of functional charged networks and how point mutations may alter the protein's conformational landscape.

<sup>1</sup>Center for Integrated Protein Science Munich (CIPSM) at the Department Chemie, Technische Universität München, Lichtenbergstraße 4, 85748 Garching, Germany. <sup>2</sup>Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden. <sup>3</sup>These authors contributed equally: Michael Röpke, Max E. Mühlbauer. ✉email: [ville.kaila@dbb.su.se](mailto:ville.kaila@dbb.su.se)

Protein structures are universally folded by shielding hydrophobic amino acids from the aqueous environment to reach a global free energy minimum<sup>1</sup>. However, buried charged networks provide central functional elements responsible for the biological activity in many proteins<sup>2–4</sup>. In complex I, a redox-driven proton pump essential for cell respiration, a 200 Å network of buried ion-pairs and charged residues provide functional elements responsible for the enzyme's proton pumping machinery<sup>2,5</sup>. Similarly, the heat shock protein 90, an essential molecular chaperone in the eukaryotic cell, employs an extended network of charged amino acids to regulate the enzymatic activity and the chaperone function<sup>3,6</sup>.

Charged residues buried within the low-dielectric protein core ( $\epsilon_{\text{protein}} = 4–10$ ) are thermodynamically destabilised relative to the high-dielectric aqueous environment ( $\epsilon_{\text{aq}} = 80$ ) by the (Born) desolvation free energy associated with moving charged groups towards lower dielectric surroundings. This effect can in part be compensated by the electrostatic interactions between oppositely charged residues, but recent studies<sup>4,7</sup> suggest that each ion-pair destabilises the overall native protein fold by up to 5 kcal mol<sup>-1</sup>. This raises the important question on how proteins with large extended networks with hundreds of charged residues<sup>2,5</sup> are structurally and thermodynamically stabilised. To address this puzzling question without the potential bias of natural protein scaffolds, we introduce here buried ion-pairs into the hydrophobic core of artificial helical bundle scaffolds<sup>8,9</sup> by combining directed computational design and molecular simulations with biophysical, spectroscopic, and structural experiments. These experimentally well-behaved, minimal de novo models do not structurally resemble highly intricate natural proteins, but despite their simple structure, four-helical bundles have been used to probe, e.g., oxidoreductase<sup>9,10</sup>, oxygen-binding<sup>8</sup>, light-capturing<sup>11</sup> and ion-transport properties<sup>12</sup>. Here they have been used to deduce interaction principles of buried ion-pairs without the complicated effect of complex natural surroundings based on rather simple biophysical and computational experiments.

## Results and discussion

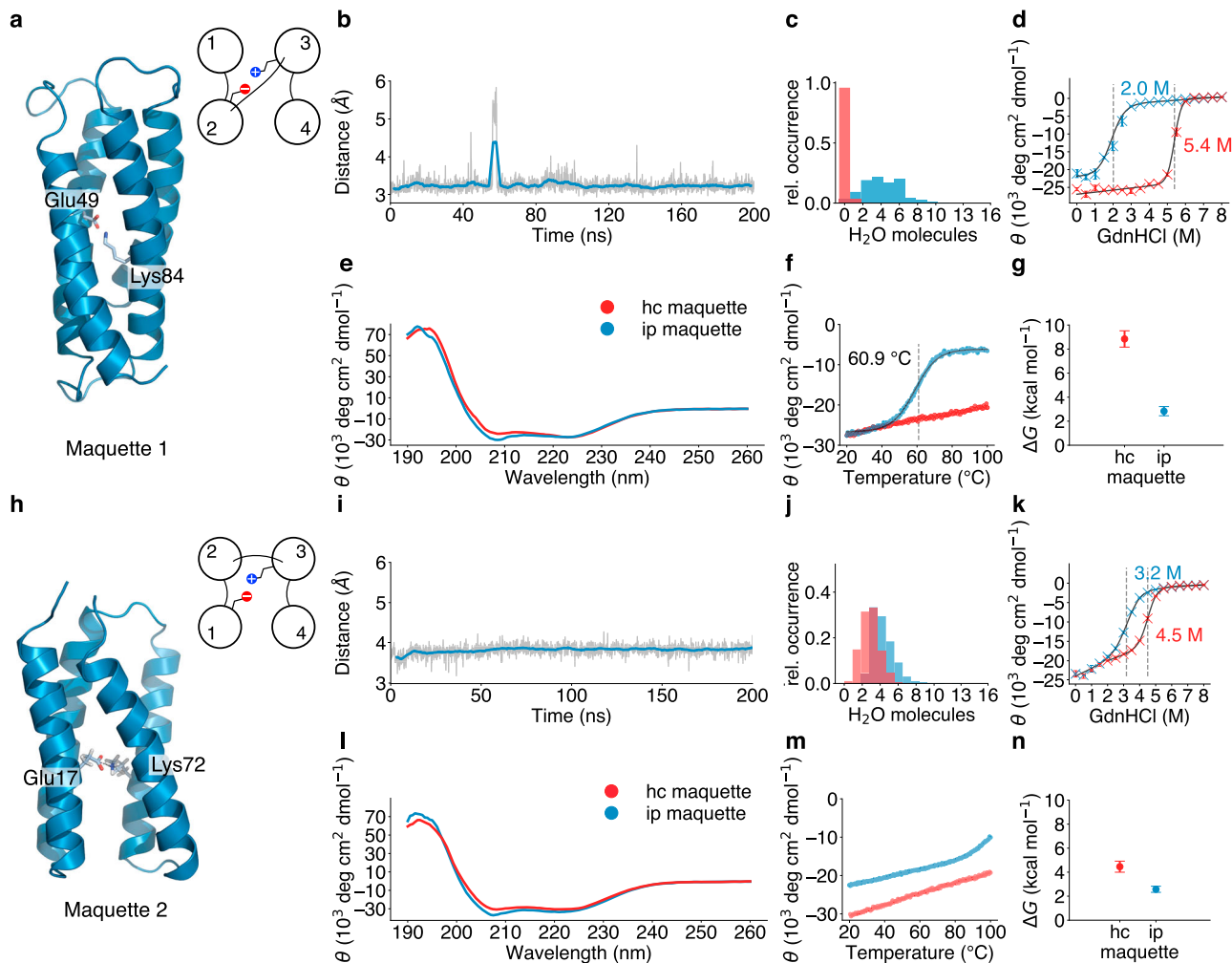
**Structural stability of buried ion-pairs in artificial bundle proteins.** To study how the protein stability is affected by buried charged residues, we introduced a glutamate (Glu)/lysine (Lys) ion-pair into the hydrophobic protein core of an artificial 4 $\alpha$ -helical protein scaffold<sup>10</sup> at positions 49 and 84 (F49E and F84K), where helix 1 forms contacts with helices 2 and 3 (see Fig. 1a, Methods and Supplementary Table 1). Atomistic molecular dynamics (MD) simulations of this *Maquette 1* model suggest that the overall structure remains stable on the simulation timescale (Fig. 1b), and energetically favours the charged (Glu/Lys<sup>+</sup>) state over the neutral (Glu<sup>0</sup>/Lys<sup>0</sup>) state (Supplementary Fig. 1d). During the MD simulations, a few water molecules transiently interact with the buried charges and partially destabilise the hydrophobic core relative to the *Maquette 1* model with a non-polar core (Fig. 1c). To further characterise the properties of the constructs, we expressed the proteins in *E. coli*, and studied their stability using circular dichroism (CD) spectroscopy and chemical unfolding experiments. The proteins remain  $\alpha$ -helical, as indicated by the minima at 208 and 222 nm (Fig. 1e). However, consistent with findings on ion-pairs in natural proteins<sup>4,7</sup>, the introduced charges decrease both the melting temperature by  $\Delta T_m > 40^\circ\text{C}$  and the protein overall stability by  $\Delta\Delta G \approx 5$  kcal mol<sup>-1</sup> relative to our model with an intact hydrophobic core (Fig. 1d–g).

Removal of the Glu or Lys counter charges decreases the protein stability by around 3 kcal mol<sup>-1</sup> based on unfolding experiments (Supplementary Figs. 1, 2a). However, despite the

well-packed protein models in our MD simulations and a two-state folding transition in the chemical unfolding experiments, our nuclear magnetic resonance (NMR) experiments show heteronuclear single quantum coherence (HSQC) spectra that are not too well-dispersed and contain some overlapping signals (Supplementary Fig. 2f), an effect that may arise from the highly symmetric structure of these models, and/or involvement of possible molten-globule like states<sup>10</sup>. Although helix dipole-sidechain charge interactions may affect the protein stability<sup>13–15</sup>, we find here that inverting the polarity of the ion-pair in the *Maquette 1* model only leads to a minor shift in the melting temperature ( $\Delta\Delta T_m = 0.1^\circ\text{C}$ ) (Supplementary Fig. 1f, 2a). Although we currently lack experimental structural data to draw definite conclusions about the buried ion-pairs in the *Maquette 1* models, our data support that buried charges decrease the overall protein stability similarly to what has been described for natural proteins<sup>4,7</sup> and consistent with the data for other de novo protein models (see below).

To improve the accuracy of the design, we next employed another artificial protein scaffold (*Maquette 2*) with a resolved core scaffold<sup>11</sup>, but a different helical topology (helix 1 interacting with helices 2 and 4) as compared to the *Maquette 1* model (Fig. 1h, *inset*). Insertion of the ion-pair at positions 17 and 72 (F17E/F72K relative to *Maquette 2*/hc) yields a stable structure in the MD simulations (Fig. 1i, Supplementary Fig. 3), but similar to the *Maquette 1* model, the charged element forms transient interactions with a few water molecules during the simulations (Fig. 1j). The ion-pair destabilises the protein by  $\Delta\Delta G \approx 2$  kcal mol<sup>-1</sup> in chemical unfolding experiments (Fig. 1k, n), suggesting an improved stability relative to *Maquette 1*, but an overall destabilisation relative to the construct with the non-polar core. Moreover, the overall protein stability remains unaffected when the charged residues are introduced into the more water accessible region of the bundle (I101E/F50K) (Supplementary Fig. 3c).

**Characterisation of buried ion-pairs.** The *Maquette 2* proteins have well-dispersed HSQC spectra, but the introduced ion-pair leads to significant chemical shift perturbations (Fig. 2c, Supplementary Fig. 2b), indicating possible structural rearrangements to compensate for the electrostatic penalty. The ion-pair is supported by NH<sub>3</sub>-selective heteronuclear in-phase single quantum coherence spectroscopy (HISQC) measurements at 275 K that reveal slow proton exchange of the shielded sidechains. This NMR signal further disappears upon substitution of either Lys72 or Glu17 with phenylalanine (Fig. 2d), whereas the intensity of the peak indicates dynamic flexibility consistent with our MD simulations (Fig. 2b, Supplementary Fig. 3). Based on 3D-NMR experiments at 293 K, we resolved a structural model of *Maquette 2* with the introduced charged element (Fig. 2a, Supplementary Fig. 2c, Supplementary Table 2). The model reveals a subtle bending of helices 1 and 3 that could arise from the thermodynamic cost associated with the introduced charged element (Figs. 1n and 2a). However, the mixed open/closed ion-pair conformations as indicated by NH<sub>3</sub>-selective HISQC spectrum (Fig. 2d) and MD simulation data (Fig. 1h, Supplementary Fig. 7b), are not reflected in the model building, which favours dissociation of the ion-pair in the structure calculation (see “Methods” section). Nevertheless, the MD simulations suggest that the closed ion-pair conformation can easily form in the NMR-models by a <1 Å inward motion of helices 1 and 3. These conformational changes are also qualitatively supported by structural modelling (Fig. 2a, Supplementary Fig. 2c). Taken together, our data suggest that the ion-pair introduces some dynamical flexibility and subtle structural rearrangements in the



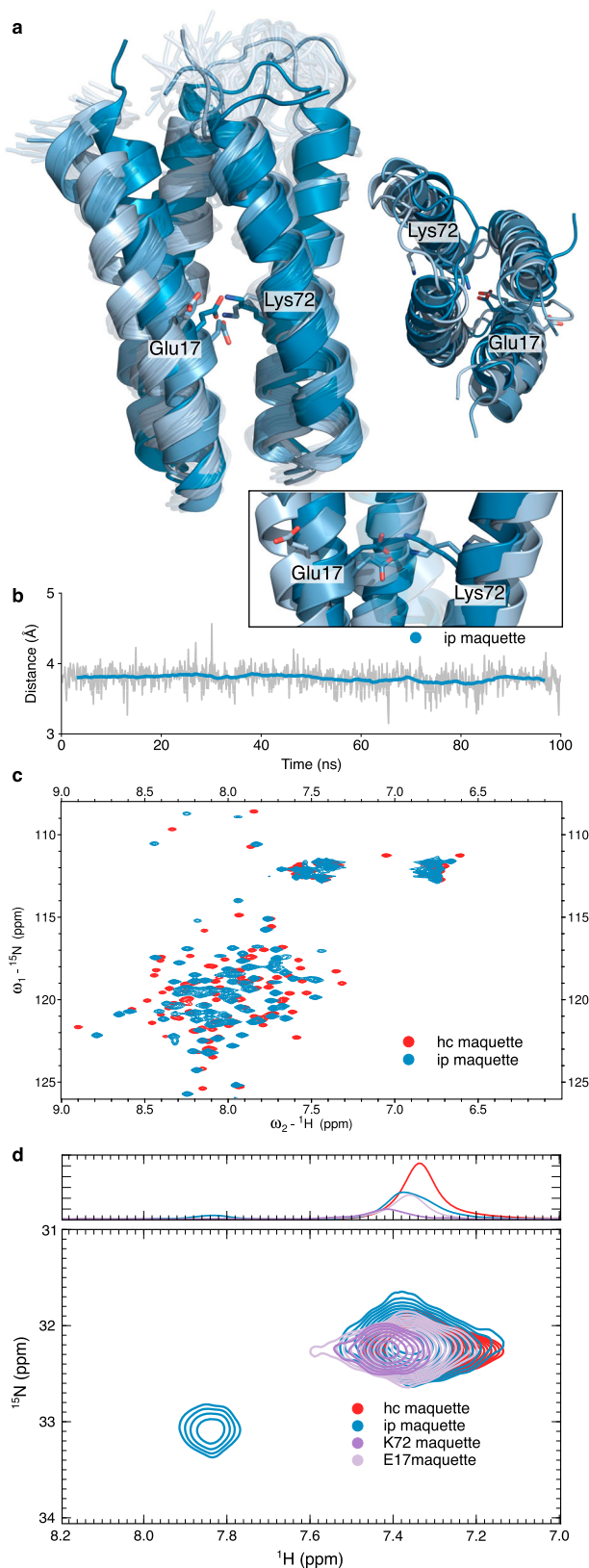
**Fig. 1** Structural stability of buried ion-pairs in artificial bundle proteins. **a, h** Structural snapshot of protein models; **b, i** distances between introduced ion-pairs (the shaded line represents the actual data, the coloured line is a moving average over 50 frames), and **c, j** number of buried water molecules during MD simulations within 9 Å of helices 1–4; **e, l** CD spectra at 20 °C of protein *Maquettes* 1 and 2 with a hydrophobic core (hc) and introduced ion-pair (ip); **f, m** melting curves measured by following the temperature dependence of the 222-nm trough in the CD spectra; **d, k** chemical unfolding with GdnHCl up to 8 M monitored by CD spectroscopy. Data points are presented as mean values of triplicates ( $n = 3$  independent experiments) with  $\pm$ standard deviation shown as error bars; and **g, n** thermodynamic stability ( $\Delta G$ ) values calculated from the chemical unfolding experiments, with error bars computed from the standard deviation of the chemical unfolding profiles.

bundle scaffold, which might enable exploration of both open and closed ion-pair conformations (Fig. 2a, Supplementary Fig. 3). Such dynamic flexibility coupled to subtle conformational changes are critical for the function of ion-pairs in natural proteins<sup>5,6</sup>.

**Analysis and design of buried charged elements in proteins.** To better understand the charge-compensation principles, we next developed a computational design approach, where we directed the residue search based on atomistic molecular simulations, by explicitly modelling both solvation and electrostatic effects underlying the thermodynamic stability of the ion-pairs. To this end, we optimised the Glu17-Lys72 ion-pair (F17E/F72K) by introducing random *in silico*-point mutations in the protein surroundings, and directed the residue search using a Metropolis Monte Carlo (MC) criterion towards an increased interaction between the ion-pair and the protein structure as an optimisation target. At the same time, we selected for structures with an increased overall protein stability (Fig. 3a, b, see Methods section).

In our directed design, we find that polar and charged residues that form hydrogen-bonded contacts with the buried ion-pair are

introduced in 2230 of 2838 ( $\approx 80\%$ ) designed constructs probed within the first solvation sphere of the charged element (Fig. 3c). In addition to other charged residues, Gln/Asn, Tyr, and Ser/Thr constitute 20% of the sampled substitutions. These mutations insert amphiphilic elements around the charged centre (Fig. 3c) that could enhance the protein stability. In statistical analysis of ca. 180,000 ion-pairs from around 6000 resolved membrane protein structures (Fig. 3d, Supplementary Figs. 4 and 5), we find similar charge-stabilising motifs in natural proteins as predicted in the computational design. Interestingly, we find a 10–30% increased probability of observing charge-stabilising residues (Gln, Asn, Tyr, Ser, Thr) within the first solvation sphere of the ion-pair in the natural proteins (Fig. 3d, Supplementary Figs. 4 and 5), closely resembling the distribution predicted by our directed design (Fig. 3c). Moreover, tyrosine residues that have an overall slightly higher natural abundance, are somewhat over-represented within the membrane proteins dataset (Supplementary Fig. 5b), possibly as they comprise both bulky non-polar/aromatic properties that can be used for packing the protein core or form cation- $\pi$  interactions with positively charged residues, but also a polar hydroxy headgroup that can form hydrogen



**Fig. 2 Biophysical and structural characterisation of ion-pair in *Maquette 2*.**

**a** NMR-structural model of *Maquette 2* with an ion-pair at positions 17 and 72 (F17E/F72K) modelled in both closed/open conformations (in dark/pale blue). **b** Dynamics of the Glu17 and Lys72 ion-pair modelled based on the NMR structure. The shaded line represents the actual data, the coloured line is a moving average over 50 frames. **c** Heteronuclear single quantum coherence (HSQC) NMR spectrum of *Maquette 2* with Glu17-Lys72 ion-pair (ip, F17E/F72K, in blue), and hydrophobic core (hc, in red). **d**  $\text{NH}_3$ -selective HSQC experiment at 275 K, indicative of a protected sidechain with slow proton exchange. Glu17-Lys72 ion-pair (ip, F17E/F72K, in blue), hydrophobic core (hc, in red), and single residue substitutions (K72 *maquette* (E17F) in purple); E17 *maquette* (K72F), in pale purple.

residues<sup>18</sup>, are also clearly pronounced in the distribution (Fig. 3d, Supplementary Fig. 4). We note that the buried Trp68 might stabilise the Glu17-Lys72 ion-pair in the *Maquette 2* model (F17E/F72K, Supplementary Fig. 2c). The ion-pair distributions further suggest that glutamine residues seem to favour Arg-Asp ion-pairs, whilst Lys-Glu ion-pairs favour asparagine residues (Supplementary Fig. 5a), possibly due to steric constraints in natural proteins.

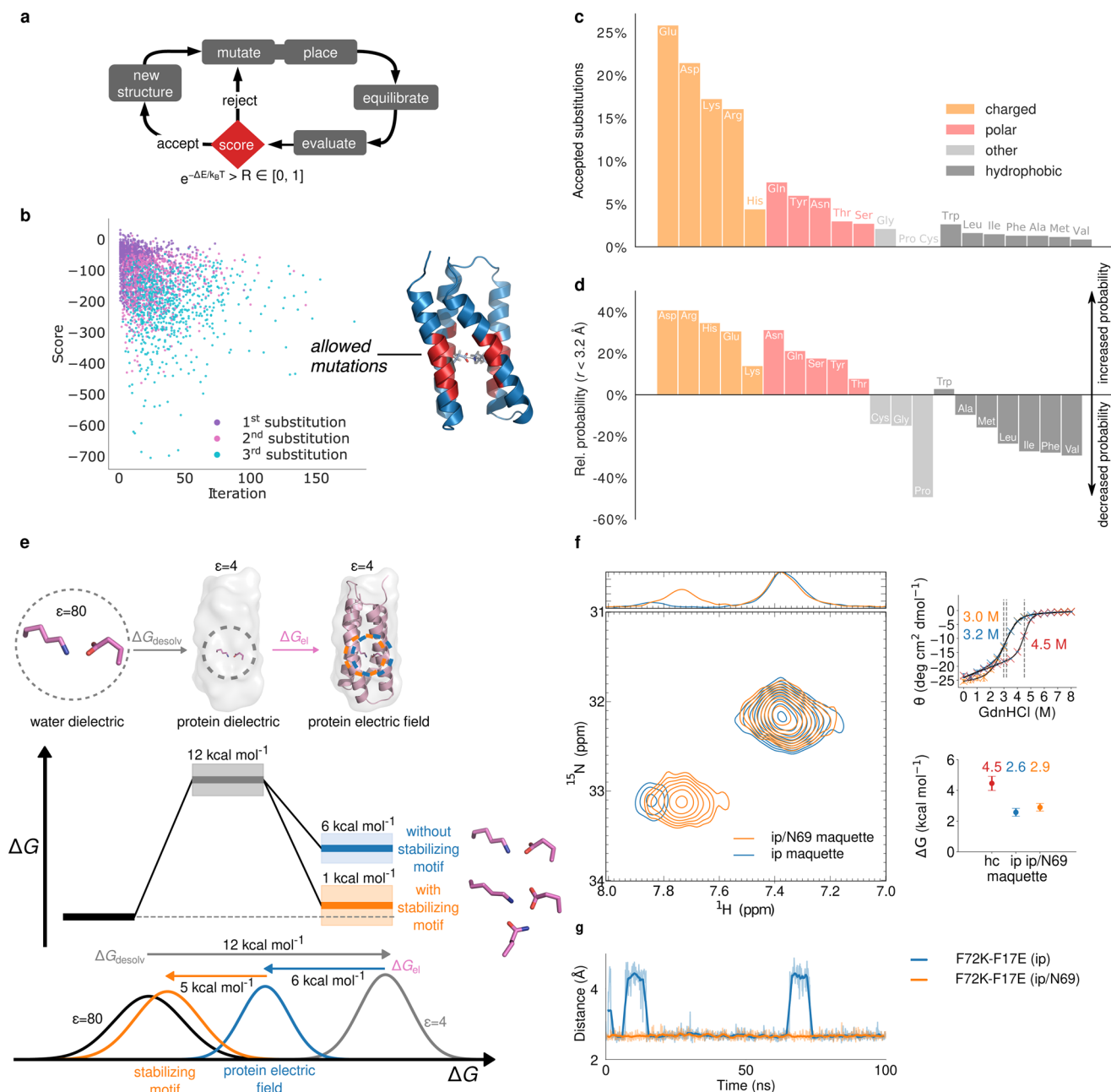
To understand the physical basis of the shielding motifs, we further studied the designed charged elements by electrostatic calculations. For the unshielded ion-pair, the Born desolvation penalty ( $\Delta G = +12 \text{ kcal mol}^{-1}$ ) becomes partially compensated by the electrostatic interaction with the protein surroundings ( $\Delta G = -6 \text{ kcal mol}^{-1}$ ), leading to an overall destabilising effect in the model calculations (Fig. 3e, see Supplementary Information). The partially uncompensated charges expose a non-uniform electric field, which could be responsible for the partial wetting of the protein core during the MD simulations and the overall reduced protein stability (Fig. 1). However, the shielding element provides a clear electrostatic compensation in the model systems that results in enhanced stability (Fig. 3e). The thermodynamic cost of embedding a charge-shielding Gln or Asn residue next to an ion-pair within the protein core is nearly isoenergetic in the model calculations (Supplementary Fig. 2e).

Inspired by these findings, potential charge-stabilising motifs were introduced in the ion-pair surroundings of the *Maquette 2* (Supplementary Fig. 2d). We find that introduction of Asn at position 69 (*Maquette 2/V69N*) leads to a small, but statistically significant enhancement of the protein stability in chemical unfolding experiments (Fig. 3f). Moreover, in  $\text{NH}_3$ -selective HSQC measurements, a significant increase in the population of the closed ion-pair is observed, suggesting that the asparagine could stabilise the closed ion-pair conformation (Fig. 3f), as also qualitatively supported by MD simulations (Fig. 3g). The HSQC spectrum of *Maquette 2/V69N* indicates that the protein is well-folded, with a downfield shift of most peaks relative to the ion-pair model (Supplementary Fig. 2g). To further investigate the conformational flexibility of the *Maquette 2* models, we performed heteronuclear NOE (hetNOE), as well as longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation experiments<sup>19</sup>. The hetNOE and relaxation data of the hydrophobic core, ion-pair, and charge-stabilised ion-pair constructs are very similar (Supplementary Fig. 8), indicating that all models have similar dynamics on the ps-ns timescale, maintain a similar fold on the studied conditions, and are likely to have overall similar structures.

bonds. We note, however, that certain amino acids may have special preferred locations in the protein structure<sup>16</sup>. Moreover, charged elements are pronounced with an 10–40% increased probability<sup>17</sup>, whereas hydrophobic residues occur with a 10–60% reduced probability around the ion-pair. Interestingly, tryptophan residues that can form cation- $\pi$  interactions with, e.g., Lys

#### Rational design of an ultra-stable buried charged network.

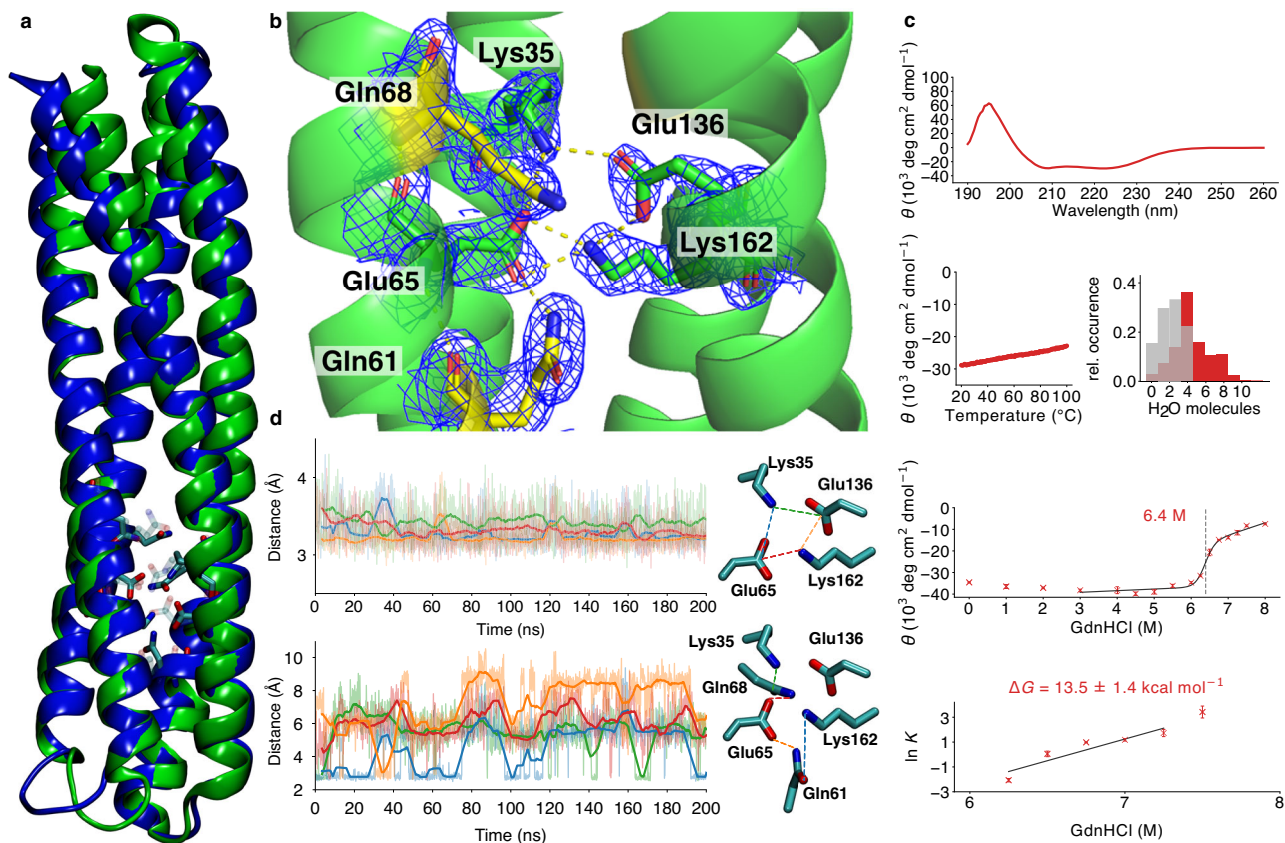
To test the physical (Fig. 3d, e) and evolutionary (Fig. 3c) insights, we created a charged network, with a double ion-pair surrounded by four charge-stabilising Gln-motifs embedded



**Fig. 3 Directed design, evolutionary conservation, and biophysical analysis of buried charged elements in proteins.** **a** Directed computational design algorithm employed to optimise buried charge elements. **b** Sampled energy-score during directed design (see “Methods” section). Inset: *Maquette 2* model with area of introduced substitutions (in red). **c** Distribution of substituted amino acids in 1000 directed simulations. **d** Relative probability of amino acids surrounding ca. 182,000 natural ion-pairs from ~6000 structures. **e** Electrostatic calculations on the energetics of the ion-pair with and without charge-stabilising motif. **f**  $\text{NH}_3$ -selective HISQC experiment at 275 K of *Maquette 2* with charge-shielded ion-pair. Right: chemical unfolding with GdnHCl monitored by CD spectroscopy of the *Maquette 2*-protein with stabilising shielding motif (top), and thermodynamic stability ( $\Delta G$ ) values calculated from the chemical unfolding experiments (bottom). Data points of the chemical unfolding profiles are presented as mean values of triplicates ( $n = 3$  independent experiments) with  $\pm$ standard deviation shown as error bars. **g** MD simulations of *Maquette 2* models at 275 K. The figure shows distances (Glu/OE2-Lys/NZ) between the ion-pair without (in blue) and with charged-stabilising Asn69 (in orange). The shaded line represents the actual data, and the coloured line is a moving average over 50 frames.

within a ca. 70 Å-long  $\alpha$ -helical *Maquette*-protein framework (*Maquette 3*, Fig. 4c, d, see Extended Methods section), originally a  $\text{Zn}^{2+}$ -site binding scaffold (PDB ID: 5VJT), that provides a larger non-polar interaction surface relative to the minimal 4 $\alpha$ -helical *Maquette* models 1 and 2. The designed construct is stable in MD simulations with intact ion-pairs and minimal water penetration (Fig. 4a, Supplementary Fig. 6), and the pronounced

minima at 208 and 222 nm in the CD spectrum support its  $\alpha$ -helical secondary structure (Fig. 4c). Remarkably, the *Maquette 3* model is resistant towards unfolding at >100 °C and 6 M guanidine chloride concentrations, indicating a high overall protein stability, despite the large buried charge density. To validate the underlying molecular structure behind the high stability, we resolved the x-ray structure of the *Maquette 3* model at 1.85 Å



**Fig. 4 Rational design of a buried ultra-stable charged motif.** **a** Structural snapshots from the in silico model of the *Maquette 3* protein with a charge-shielded buried ion-pair network (in blue) and the resolved x-ray structure (in green). **b** Closeup of the designed charged network in the 1.85 Å resolved x-ray structure. **c** CD spectrum at 20 °C of *Maquette 3* (top); melting curves measured by following the temperature dependence of the 222-nm trough from CD spectroscopy (left); count of buried water molecules 5 Å around the ion-pair/Asn surroundings (grey) and water molecules within 5 Å of the complete protein (in red) (right); chemical unfolding with GdnHCl up to 8 M monitored by CD spectroscopy; thermodynamic stability inferred from the chemical unfolding experiments (bottom). Data points of the chemical unfolding profile are presented as mean values of triplicates with  $\pm$ standard deviation as error bars. **d** MD simulation of the ion-pair dynamics. Top: Distances between Lys35-Glu65 (blue), Lys35-Glu136 (green), Glu65-Lys162 (red), and Glu136-Lys162 (orange). Bottom: Distances between Lys35-Gln68 (green), Gln61-Glu65 (orange), Gln61-Lys162 (blue), and Glu65-Gln68 (red). The measured atomic distances are depicted to the right of the plots. See Supplementary Fig. 6 for further details. The shaded line represents the actual data, the coloured line is a moving average over 50 frames.

resolution (Fig. 4a, b, Supplementary Fig. 6, Supplementary Table 3). The  $2F_o - F_c$  electron density map closely resembles the designed in silico model (backbone RMSD (in silico/x-ray) = 1.7 Å), with a charged core forming tight ion-pairs between Lys162/Glu65 and Lys35/Glu136 at distances of 2.7–2.9 Å. The charged residues are stabilised by hydrogen bonds to Gln61/68/169/132 (Fig. 4b, d, Supplementary Fig. 6c, e) that dynamically exchange in the MD simulations (Fig. 4d, Supplementary Fig. 7c). Although other charge-stabilising residues, such as tyrosine and histidine, are often present around natural ion-pairs (Supplementary Figs. 4 and 5), these residues were too bulky to be inserted around the charged network in the current minimal *Maquette 3* model. Intriguingly, the ion-pairs remain tightly sealed from water molecules in the crystal structure, consistent with the high stability indicated by our unfolding experiments and MD simulation (Fig. 4c, Supplementary Fig. 6b). Similar to the *Maquette 1* models, the HSQC spectrum of *Maquette 3* (Supplementary Fig. 6f) shows several overlapping and less sharp signals possibly due to the higher internal symmetry of the model. We could therefore not use NMR to study these models further. Nevertheless, to quantify the stabilising effect of the charge-shielding residues surrounding the ion-pairs, we further substituted two of

the glutamines with alanine residues (Q61A, Q68A). Our chemical unfolding measurements suggest that these substitutions indeed destabilise the protein by around 5 kcal mol<sup>-1</sup>, supporting the proposed charge-shielding effect (Supplementary Fig. 6g). Taken together, the *Maquette 3* design supports the proposed electrostatic principles in which amphiphilic residues shield charged networks from the non-polar surroundings to gain thermodynamic stability.

In summary, we have characterised here the energetics of buried ion-pairs using artificial protein design in combination with computational, biophysical, and structural experiments. We identified that amphiphilic shielding motifs stabilise the buried charged networks by a fine-tuned balance between desolvation and electrostatic effects. The motifs could represent an evolutionary prerequisite for buried ion-pairs, as they are found in a majority of the analysed natural protein structures, and independently supported by our directed computational design. Our de novo proteins built based on these physical principles, resulted in ultra-stable charged networks that sustain harsh chemical and physical conditions. Our findings may provide a basis for understanding the structural stability of extended charged networks in molecular gates, switches, and allosteric

regulation sites in different protein machineries<sup>5,6,20,21</sup>, and how disease-causing mutations may destabilise the protein structure and result in an altered conformational landscape<sup>22–24</sup>.

## Methods

**In silico modelling of Maquette models.** An atomic model of *Maquette 1* was built in silico with  $\alpha$ -helical secondary structure restraints placed on the residues 1–27, 36–62, 71–97, and 106–130 based on PsiPred<sup>25</sup>. The helices were packed to maximise the non-polar interaction within the core, in which Phe-15 and Phe-21 were restrained to form a stacking interaction with the nearest anti-symmetric helix, following computational modelling principles applied before<sup>26</sup>. After the initial restrained relaxation, loops were built to connect the helices, followed by restrained minimisation and MD simulations using the CHARMM36 force field<sup>27</sup> at  $T = 300$  K and using a 1 fs integration timestep. The *Maquette 1* models were built and relaxed using CHARMM<sup>28</sup>, followed by MD simulations (see below). The *Maquette 2* models were constructed based on scaffold structures resolved in ref. <sup>11</sup>. The *Maquette 3* model was built in silico based on a helical coiled-coil template (PDB ID: 5VJT). First, all prosthetic groups were removed from the template, and glycine and alanine residues were replaced by valine and isoleucine residues, respectively, to stepwise maximise the non-polar packing, followed by a short (20 ns) MD simulation to test its stability. To further maximise the non-polar packing, hydrophobic residues at positions H10F, H111F, H125F, H139F, and the charged network was modelled manually by substitutions E162K, E35K, I61Q, H68Q, I132Q, Y169Q according to the charge-shielding principles described in Fig. 3e. The final sequence of the *Maquette 3* model is shown in Supplementary Table 1.

**Molecular simulations.** Classical MD simulations of the designed proteins were performed by solvating the protein in a TIP3P water box and 150 mM NaCl concentration. MD simulations were performed in duplicates for 200 ns (Supplementary Fig. 7) at  $T = 310$  K using a 2 fs integration timestep and the CHARMM36 force field<sup>27</sup>, and modelling the long-range electrostatics using the particle-mesh Ewald approach. Solvation free energies were estimated using Poisson-Boltzmann (PB) continuum electrostatic calculations with MC sampling. Protein residues were modelled as atomic point charges, and the protein surroundings were treated as polarisable medium with  $\epsilon = 4$  that was embedded in a water dielectric of  $\epsilon = 80$ . The linearized-PB equations were solved numerically using the adaptive PB-solver (APBS) and Karlsberg<sup>29,30</sup>. Visual Molecular Dynamics (VMD)<sup>31</sup> and PyMOL<sup>32</sup> were used for visualisation and analysis, MD simulations were performed using NAMD<sup>33</sup>. See Supplementary Information for summary of the performed simulations (Supplementary Fig. 6, Supplementary Table 4).

**Directed computational design.** Charged elements in the artificial protein models were optimised by introducing random mutations into the *Maquette* structures, followed by 10 ps classical MD simulations in explicit solvent. The protein models were scored by calculating the self-interaction of the ion-pair (weight = 0.48) and its interaction with the remaining protein surroundings (weight = 0.48) and all nonbonded interactions within the protein (weight = 0.02). The sum of the weights were normalized to 1.0. The substitutions were accepted or rejected based on weighted interaction scores using a Metropolis Monte Carlo criterion. All substitutions were subjected to 5000 steepest-gradient minimisation steps, followed by 5 ps restrained relaxation with harmonic protein backbone restraints with force constants of 1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The protein interactions were modelled using the CHARMM36 force field<sup>27</sup>, and the MD relaxation was performed using a 1 fs integration timestep at  $T = 310$  K using NAMD<sup>33</sup>. The directed computational design approach was implemented in Python in our in-house code PACMAN (Packing with MC-assisted networks).

**Statistical analysis of ion-pairs.** Statistical analysis of naturally occurring ion-pairs was carried out based on structures available in the orientations of proteins in membranes (OMP) database<sup>34</sup>. From the 7230 available structures, 6045 with a resolution <3.5 Å were included. The protein surface was calculated using the MSMS program<sup>35</sup>, and ion-pairs were identified using the Salt Bridges plugin of VMD<sup>31</sup>. All residues within a 15 Å sphere of the identified ion-pairs were counted and probabilities for their occurrence within a certain distance to the centre-of-mass of the ion-pair was calculated as,

$$p^{aa}(r_{\text{thr}}) = \frac{N_{\text{thr}}^{aa}}{\sum_i^{AA} N_{\text{thr}}^i} \quad (1)$$

where  $aa$  is the identified amino acid,  $r_{\text{thr}}$  is the threshold radius,  $AA$  the full set of amino acids, and  $N_{\text{thr}}^{aa}$  is the number of amino acids within a given threshold radius.

**Protein expression and purification.** Genes for the computationally designed proteins were synthesized and cloned into pET21a(+) vector (Genscript). F17E, F72K, 1ip/I43Y, 1ip/I39Q, 1ip/R76S, and 1ip/L91H of *Maquette 2* were produced using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's manual and the primers listed in Supplementary Table 5.

The C-terminal His<sub>6</sub>-tagged proteins were overexpressed in *E. coli* BL21 (DE3) for 4 h at 37 °C after induction with 0.5 mM IPTG at OD<sub>600</sub> 0.7. After sonication, the lysate was applied on a Ni-NTA affinity column. The His-tag was cleaved by a Tobacco Etch Virus protease, and the proteins were purified by size exclusion chromatography using a preparative Superdex 75 pg HiLoad 16/600 column (GE Healthcare) with a flow rate of 1.0 mL min<sup>-1</sup>. See Supplementary Table 1 for amino acid sequences of all artificial proteins and Supplementary Table 6 for DNA sequences. The buffers used during protein purification are summarised in Supplementary Table 7.

**Circular dichroism spectroscopy.** CD spectra and melting curves were measured on a Jasco J-715 system combined with a cooling-heating unit using a 1-mm-path quartz cuvette. Protein concentrations of around 5–10  $\mu$ M were used for measurements at 20 °C. For the determination of melting curve, samples were heated up from 20 to 100 °C, while the alpha-helical CD-signal at 222 nm was monitored. The protein stability was also studied by guanidine-hydrochloride (GdnHCl) or urea induced unfolding with GdnHCl/urea concentrations up to 8 M. Folding free energies were obtained from linear least-square fits of  $-RT \ln K$  that were extrapolated to 0 M GdnHCl. The apparent equilibrium constant  $K$  was determined from the fraction of the denatured protein,  $f_D = (y - y_N) / (y_D - y_N)$ , where  $y$  is the CD raw signal, and  $y_N$  and  $y_D$  are the signals for the native and denatured states, respectively. Error bars were computed from the standard error of the least-square fits from the average unfolding profiles.

**Nucleic magnetic resonance spectroscopy.** Heteronuclear NMR spectra were recorded using <sup>15</sup>N- and <sup>15</sup>N,<sup>13</sup>C-labelled protein at 20 °C in a buffer with 50 mM sodium phosphate, 100 mM NaCl, and 1 mM EDTA at pH 7.5. NMR experiments were performed on Bruker Avance III spectrometers operating at a <sup>1</sup>H Larmor frequency of 600 MHz (14.09 T) and 950 MHz (22.31 T) using a CPTCI triple-resonance cryoprobe and the Bruker TopSpin v. 3.5 software. Sequence-specific backbone and side-chain assignment were obtained using 3D HNCACB, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(CCO)NH, CC(CO)NH, HC(C)H/(H)CCH TOCSY, <sup>15</sup>N/<sup>13</sup>Cali/<sup>13</sup>Caro-edited <sup>1</sup>H,<sup>1</sup>H NOESY (120 ms mixing time) as well as 2D (HB)CB(CGCD)HD. 2D <sup>1</sup>H,<sup>15</sup>N HSQC spectra were recorded for the detection of Lys <sup>15</sup>NH<sub>3</sub> groups<sup>36</sup>. <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N chemical shifts were referenced via DSS and spectra were assigned using Sparky<sup>37</sup>. <sup>15</sup>N  $T_1$ ,  $T_{1\rho}$ , and steady-state hetNOE experiments were collected at an external magnetic field of 11.74 T (<sup>1</sup>H Larmor frequency of 500 MHz) and 293 K. Temperature compensation was employed, according to Lakomek et al.<sup>38</sup> The <sup>15</sup>N rf amplitude for the  $T_{1\rho}$  spin-lock was set to 1.5 kHz. For the steady-state hetNOE experiment a saturation time of 4 s was employed.  $T_2$  was calculated according to

$$\frac{1}{T_2} = \frac{R_{1\rho}}{\sin^2\theta} - \frac{R_1}{\tan^2\theta} \quad (2)$$

with

$$R_{1\rho} = 1/T_{1\rho}, \quad (3)$$

$$R_1 = 1/T_1, \quad (4)$$

$$\tan\theta = \frac{\omega_1}{\Omega}. \quad (5)$$

$\omega_1$  is the amplitude of the <sup>15</sup>N spin-lock field and  $\Omega$  is the <sup>15</sup>N resonance offset from the spin-lock carrier frequency. The experimental error was set to two times the standard deviation of the spectral noise. Uncertainties were estimated by 1000 MC runs.

**NMR structure calculation.** Automated NOE assignment was conducted in CYANA<sup>39,40</sup> to generate restraints and calculate de novo structures of the *Maquette 2* construct with ion-pair at Glu17/Lys72. The closed ion-pair conformation was modelled by restraining lower and upper distance limits between Glu17-OE1 and Lys72-NZ of 2.8 and 3.0 Å, respectively, as guided by the MD simulations. The calculations with torsion angle dynamics were started from 200 random structures, employing 20,000 torsion angle dynamics steps. In addition to distance restraints, dihedral angle restraints derived from chemical shifts were obtained using TALOS-N<sup>41</sup>. The restrained energy minimisation of the consensus bundle of 20 conformers in explicit solvent against the AMBER force field was carried out using OPALp<sup>42,43</sup>. The statistics of the structure calculations is given in the Supplementary Table 2.

**X-ray crystallography.** Crystals of the *Maquette 3* construct were grown at 20 °C using the sitting drop vapour diffusion method. Drops contained a 1:2 mixture of protein solution (40 mg ml<sup>-1</sup> protein) and reservoir solution (100 mM HEPES buffer pH 7.5, 1.5 M potassium phosphate). The crystals were cryo-protected with 25% glycerine. Diffraction data were collected at the beamline X06SA at the Paul Scherrer Institute, SLS, Villigen, Switzerland ( $\lambda = 1.0$  Å). Evaluation of reflection intensities and data reduction were performed with the program package XDS (v. January 31, 2020)<sup>44</sup>. Refinement of the initial model was carried out with the coordinates of the in silico model of *Maquette 3* using REFMAC<sup>45</sup>, followed by

model building using COOT<sup>46</sup> and CCP4 v. 7.0<sup>47</sup>. See Supplementary Table 3 for further details.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The structures have been deposited to the Protein Data Bank, PDB IDs: 6ZOC [rcsb.org/structure/6zoc], 6Z35 [rcsb.org/structure/6z35]. The NMR assignments are deposited in the BMRB under the entry number 34518 [bmrb.io/data\_library/summary/?bmrblD=534518]. Other data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

### Code availability

The pacman code is available under MIT license at <https://github.com/KailaLab/pacman-protein-design>.

Received: 30 May 2020; Accepted: 19 February 2021;

Published online: 25 March 2021

### References

- Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
- Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
- Ali, M. M. U. et al. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).
- Robinson, A. C., Castaneda, C. A., Schlessman, J. L. & Garcia-Moreno, E. B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl Acad. Sci. USA* **111**, 11685–11690 (2014).
- Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, 20170916 (2018).
- Mader, S. L. et al. Conformational dynamics modulate the catalytic of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410 (2020).
- Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno, E. B. Charges in the hydrophobic interior of proteins. *Proc. Natl Acad. Sci. USA* **107**, 16096–16100 (2010).
- Koder, R. L. et al. Design and engineering of an O<sub>2</sub> transport protein. *Nature* **458**, 305–309 (2009).
- Robertson, D. E. et al. Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432 (1994).
- Farid, T. A. et al. Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **9**, 826–833 (2013).
- Polizzi, N. F. et al. De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
- Joh, N. H. et al. De novo design of a transmembrane Zn<sup>2+</sup>-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
- Hwang, J.-L. & Warshel, A. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature* **334**, 270–272 (1988).
- Robinson, A. C., Schlessman, J. L. & García-Moreno, E. B. Dielectric properties of a protein probed by reversal of a buried ion pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018).
- Sali, D., Bycroft, M. & Fersht, A. R. Stabilization of protein structure by interaction of  $\alpha$ -helix dipole with a charged side chain. *Nature* **335**, 740–743 (1988).
- Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).
- Bush, J. & Makhatazde, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins* **79**, 2027–2032 (2011).
- Dougherty, D. A. Cation- $\pi$  interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* **271**, 163–168 (1996).
- Wagner, G. The importance of being floppy. *Nat. Struct. Biol.* **2**, 255–257 (1995).
- Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
- Di Luca, A., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl Acad. Sci. USA* **114**, E6314–E6321 (2017).
- Fiedorczuk, K. & Sazanov, L. A. Mammalian mitochondrial complex I structure and disease-causing mutations. *Trends Cell Biol.* **28**, 835–867 (2018).
- Boczek, E. E. et al. Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl Acad. Sci. USA* **112**, E3189–E3198 (2015).
- Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
- Buchan, D. W. A. & Jones, D. T. The PSIREN protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
- Gámiz-Hernández, A. P., Kieseritzky, G., Galstyan, A. S., Demir-Kavuk, O. & Knapp, E. W. Understanding properties of cofactors in proteins: redox potentials of synthetic cytochromes b. *Chem. Phys. Chem.* **11**, 1196–1206 (2010).
- Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
- Brooks, B. R. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1174–1178 (2009).
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
- Kieseritzky, G. & Knapp, E. W. Optimizing pKa computation in proteins with pH adapted conformations. *Proteins* **71**, 1335–1348 (2008).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
- The PyMol Molecular Graphics System. Version 1.8 Schrödinger LLC (2015).
- Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
- Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
- Sanner, M. F., Olson, A. J. & Spehner, J.-C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
- Iwahara, J., Jung, Y. S. & Clore, G. M. Heteronuclear NMR spectroscopy for lysine NH<sub>3</sub> groups in proteins: Unique effect of water exchange on 15N transverse relaxation. *J. Am. Chem. Soc.* **129**, 2971–2980 (2007).
- Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
- Lakomek, N. A., Ying, J. & Bax, A. Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012).
- Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015).
- Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
- Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
- Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85 (2003).
- Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147 (2000).
- Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2184–2195 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, 486–501 (2010).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Cryst.* **D67**, 235–242 (2011).

### Acknowledgements

We acknowledge Prof. Michael Sattler and Prof. Franz Hagn for providing us access to the Bavarian NMR Facility, Astrid König for technical assistance in crystallisation trials, and Dr. Sina Kazemi for helpful discussions regarding NMR structure calculations. This work was supported by the collaborative research centre, SFB1035 (B12) and the Centre for Integrated Protein Science Munich (CIPSM). This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program/grant agreement no. 715311. VRIK is supported by the Knut and Alice Wallenberg Foundation. Computational resources were provided by the Leibniz-Rechenzentrum (LRZ), SuperMuc (project: pr27xu), and the Swedish National Infrastructure for Computing (SNIC) at PDC partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

### Author contributions

M.B. expressed, purified, and characterised the designed proteins, and performed computational and biophysical characterisation. M.B., M.R., M.E.M., S.L.M., A.P.G.H., and V.R.I.K. performed molecular simulations and design. M.R., M.E.M., S.L.M., and V.R.I.K. developed analytic computational tools. M.B. and S.A. performed NMR experiments and calculated NMR structure ensembles. M.B., M.G., and V.R.I.K. resolved the x-ray structures. A.P.G.H. performed initial computational model building. M.B., M.R., M.E.M., S.A., S.L.M., K.F., M.G., A.P.G.H., and V.R.I.K.



analysed results. V.R.I.K. designed and directed the project, performed initial computational design and experimental characterisation, and wrote the manuscript with input from all authors.

### Funding

Open access funding provided by Stockholm University.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21909-7>.

**Correspondence** and requests for materials should be addressed to V.R.I.K.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021