

# Design and characterization of a protein fold switching network

Received: 14 July 2022

Accepted: 13 January 2023

Published online: 26 January 2023

Check for updates

Biao Ruan<sup>1,6</sup>, Yanan He<sup>2,6</sup>, Yingwei Chen<sup>1,6</sup>, Eun Jung Choi<sup>1</sup>, Yihong Chen<sup>2</sup>, Dana Motabar <sup>1,3</sup>, Tsega Solomon<sup>2,4</sup>, Richard Simmerman<sup>1</sup>, Thomas Kauffman<sup>2,4</sup>, D. Travis Gallagher<sup>2,5</sup>, John Orban <sup>2,4</sup> & Philip N. Bryan <sup>1,2</sup>

To better understand how amino acid sequence encodes protein structure, we engineered mutational pathways that connect three common folds (3 $\alpha$ ,  $\beta$ -grasp, and  $\alpha/\beta$ -plait). The structures of proteins at high sequence-identity intersections in the pathways (nodes) were determined using NMR spectroscopy and analyzed for stability and function. To generate nodes, the amino acid sequence encoding a smaller fold is embedded in the structure of an ~50% larger fold and a new sequence compatible with two sets of native interactions is designed. This generates protein pairs with a 3 $\alpha$  or  $\beta$ -grasp fold in the smaller form but an  $\alpha/\beta$ -plait fold in the larger form. Further, embedding smaller antagonistic folds creates critical states in the larger folds such that single amino acid substitutions can switch both their fold and function. The results help explain the underlying ambiguity in the protein folding code and show that new protein structures can evolve via abrupt fold switching.

There have been remarkable advances recently in the ability to predict the tertiary structure of a protein from its primary amino acid sequence<sup>1,2</sup> as well as to design amino acid sequences that encode stable, unique protein structures<sup>3</sup>. It is also well-established, however, that some proteins have a propensity for two completely different, but well-ordered, conformations<sup>4–12</sup>. Better insight into the ambiguity of the protein folding code would lead to a better understanding of how proteins evolve, how mutation is related to disease, and how function can be annotated to sequences of unknown structure<sup>13–27</sup>. If the protein folding code were truly understood, it would be possible both to predict and design proteins that undergo profound switches in conformation. There has been significant progress in understanding natural proteins that switch folds<sup>11</sup> and predicting natural fold-switching proteins from amino acid sequence data<sup>25</sup>. Designing proteins at the interface between different folds has been possible<sup>7,28–30</sup> but still presents a formidable challenge. It has been particularly challenging to design monomeric proteins that switch fold without a change in quaternary structure, and a better understanding is needed about how a

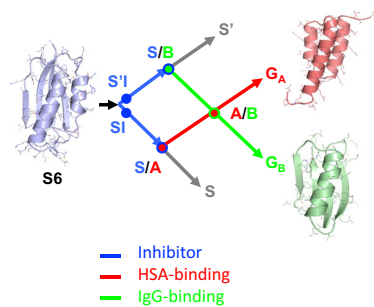
very limited subset of intra-protein interactions can tip the balance from one fold and function to another<sup>29,31,32</sup>.

Our goal here was to engineer monomeric proteins that are in a critical state between two distinct folds. To do this we chose three well-studied protein folds and designed a series of sequences such that each sequence is compatible with two sets of native interactions. Two of these folds are from *Streptococcal* Protein G which contains two types of domains that bind to serum proteins in blood: the G<sub>A</sub> domain binds to human serum albumin (HSA)<sup>33,34</sup> and the G<sub>B</sub> domain binds to the constant (Fc) region of IgG<sup>35,36</sup>. The third protein is S6, a component of the 30S ribosomal subunit of *Thermus thermophilus*<sup>37–41</sup>. For simplicity, the S6 fold is referred to as an S-fold, the G<sub>A</sub> fold as an A-fold, and the G<sub>B</sub> fold as a B-fold. These proteins share no significant sequence homology and are representative of three of the ten most common folds: the S-fold is a thioredoxin-like  $\alpha/\beta$  plait; the A-fold is a homeodomain-like 3 $\alpha$ -helix bundle; and the B-fold is a ubiquitin-like  $\beta$  grasp<sup>42</sup>.

Figure 1 depicts a network of high-identity sequence intersections (nodes) that connect the three folds. The arrows in Fig. 1 show a

<sup>1</sup>Potomac Affinity Proteins, 11305 Dunleith Pl, North Potomac, MD 20878, USA. <sup>2</sup>Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA. <sup>3</sup>Department of Bioengineering, University of Maryland, College Park, MD 20742, USA. <sup>4</sup>Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. <sup>5</sup>National Institute of Standards and Technology and the University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA. <sup>6</sup>These authors contributed equally: Biao Ruan, Yanan He, Yingwei Chen.

e-mail: [jorban@umd.edu](mailto:jorban@umd.edu); [pbryan@potomac-affinity-proteins.com](mailto:pbryan@potomac-affinity-proteins.com)



**Fig. 1 | Overview of engineered nodes in the S6, G<sub>A</sub>, and G<sub>B</sub> networks.** S6 is the origin sequence in the engineering process. SI and S'I are separate nodes and are loop-size variants of the S-fold, both having protease inhibitor functions. The SI branch of the mutational path leads to a node with the A-fold and HSA binding function. The S'I branch of the path leads to a node with the B-fold and IgG binding function. The S/A node (blue and red circles) includes proteins S<sub>a1</sub>, S<sub>a2</sub>, A<sub>1</sub>, and A<sub>2</sub>. The S/B node (blue and green circle) includes proteins S<sub>b3</sub>, S<sub>b4</sub>, S<sub>b5</sub>, B<sub>3</sub>, and B<sub>4</sub>. The A and B paths themselves intersect at an A/B node (green and red circles) at which A- and B-folds are nearly iso-energetic and bifunctional. The S and S' branches continue and connect with many other natural sequences in the  $\alpha/\beta$  plait super-fold family.

network originating with the natural S6 sequence. Circles represent nodes in the network at which structural and/or functional switches occur. The SI and S'I nodes are branch points and lead down diverging sequence pathways, one leading to a node with the A-fold (S/A) and one to a node with the B-fold (S/B). Intersecting mutational pathways lead from S/A to the native G<sub>A</sub> protein and S/B to the native G<sub>B</sub> protein. At this intersection (A/B), an A-fold switches to a B-fold.

Proteins around the A/B node have been extensively characterized in our earlier work<sup>29,31,32</sup>. Here we determine that both G<sub>A</sub> and G<sub>B</sub> can switch into a third fold ( $\alpha/\beta$ -plait) and show that these three folds and four functions (HSA-binding, IgG-binding, protease inhibition, and RNA-binding) can be connected in a network that avoids unfolded and functionless states. We describe how these nodes were engineered, determine key structures using NMR spectroscopy, and analyze stability and binding function. The ability to design and characterize nodes connecting three common small folds suggests that fold switching may be an intrinsic feature of the protein folding code and is important in the evolution of protein structure and function.

## Results

### Designing a functional switch from ribosomal protein to protease inhibitor

The S6 ribosomal protein is structurally homologous to subtilisin protease inhibitors known as prodomains (Fig. 2a, b)<sup>43,44</sup>. Prodomain-type inhibitors have two binding surfaces with the protease. One surface comprises the last nine C-terminal amino acids of the inhibitor which bind in the substrate binding cleft of the protease (Fig. 2b). A second, more dynamic surface is formed between two subtilisin helices and the large surface of the  $\beta$ -sheet in the  $\alpha/\beta$ -plait topology of the inhibitor (Fig. 2b)<sup>45–47</sup>. As a result, the S6 protein could be converted into a subtilisin inhibitor protein of the same overall fold (denoted SI) by replacing its nine C-terminal amino acids with residues optimized to bind in the substrate binding cleft of subtilisin. This replacement results in new contacts between the SI  $\beta$ -sheet and the subtilisin surface helices (Fig. 2b).

The SI-protein is 99 amino acids in length and has a 10 residue loop between  $\beta$ 2 and  $\beta$ 3. However, there are many natural variations in the length of loops in the conserved  $\alpha/\beta$ -plait topology<sup>48</sup>. Therefore, we also engineered a 91 amino acid version of the S-fold (denoted S'I), which resembles the topology of natural prodomain inhibitors (Supplementary Fig. 1). Specifically, the S'I inhibitor has a longer loop connecting  $\beta$ 1 to  $\alpha$ 1 and a shorter turn connecting  $\beta$ 2 to  $\beta$ 3 (Fig. 2b).

The SI and S'I proteins were expressed and purified by binding to a protease column<sup>49</sup>. The CD spectra were compared to the native S6 protein (Supplementary Fig. 1). Inhibition constants ( $K_i$ ) were measured using an engineered RAS-specific subtilisin protease and the peptide substrate QEEYSAM-AMC<sup>49</sup>. SI and S'I inhibit the RAS-specific protease with  $K_i$  values of 200 and 60 nM, respectively (Supplementary Table 1). The details of the competitive inhibition assay are described in the “Methods” section. The results demonstrate that a ribosomal protein can be converted into a protease inhibitor with minor modification (and without a fold switch). In addition, however, the SI and S'I proteins also facilitated engineering subsequent switches to new folds and functions by linking each of the S-, A-, and B- folds to easily measured binding functions: protease inhibition (S or S'-fold); HSA-binding (A-fold, Fig. 2e)<sup>50</sup>; and IgG binding (B-fold, Fig. 2f)<sup>51</sup>.

### Designing fold switches

In previous work, we created sequences that populate both A- and B-folds by threading the A-sequence through the B-fold, finding a promising alignment, and then using phage-display selection to reconcile one sequence to both folds<sup>29,52,53</sup>. Here the approach is conceptually similar, except that we use Rosetta<sup>54</sup> as a computational design tool to test compatible mutations rather than phage display. The design process is as follows:

- Thread the A- or B- sequence through both SI and S'I-fold types.
- Identify alignments that minimize the number of catastrophic interactions.
- Design mutations to resolve unfavorable interactions in clusters of 4–6 amino acids using Pymol<sup>55</sup> and energy minimize using Rosetta-Relax<sup>54</sup>.
- Optimize protein stability in the S-fold by computationally mutating amino acids at non-overlapping positions. Repeat energy minimization and evaluation with Rosetta-Relax.
- To reduce uncertainties involved in computational design, conserve original amino acids whenever possible.

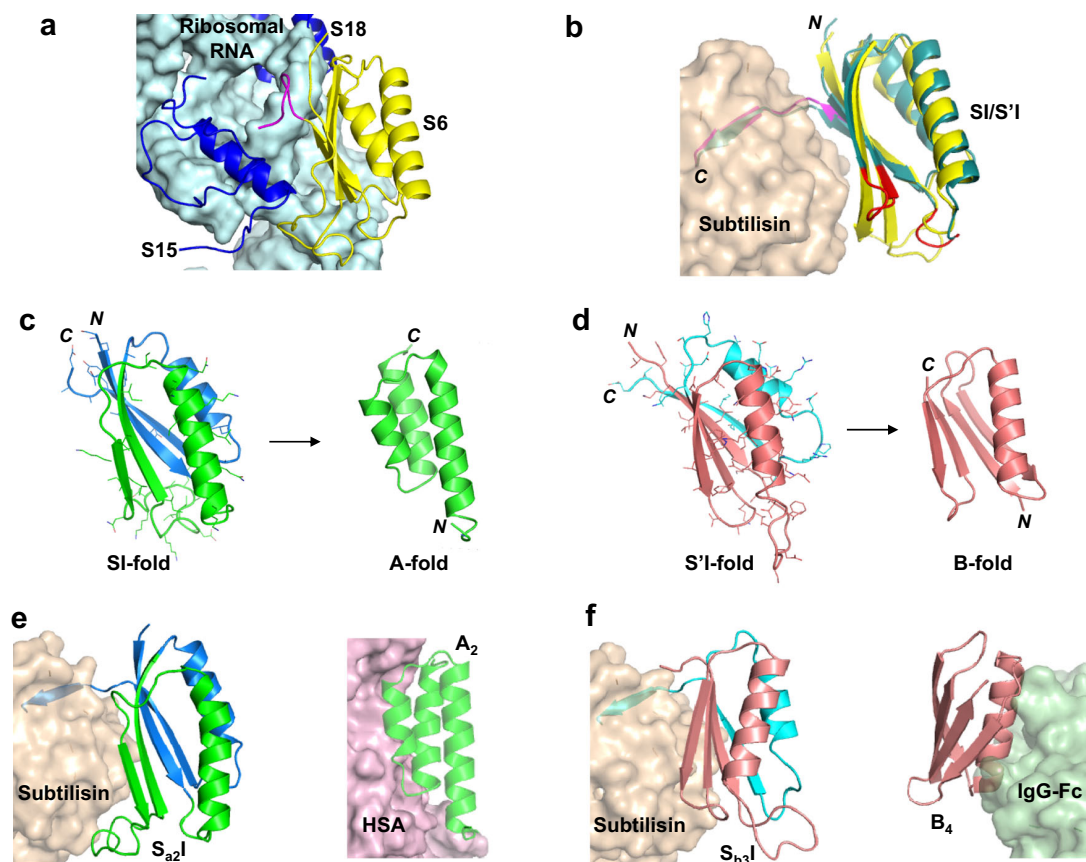
There is no reason to assume that this method is optimal. We are just applying a practicable scheme for engineering sequences compatible with two sets of native interactions and then evaluating structure, stability, and function. Initial designs were refined based on structural analysis with NMR, thermodynamic analysis of unfolding, and functional analysis using binding assays, as described below. All designed proteins were expressed in *E. coli* and purified to homogeneity as described in the “Methods” section.

### Designing a switch from $\alpha/\beta$ -plait protease inhibitor to 3 $\alpha$ HSA-binding protein

Alignment of the 56 amino acid HSA-binding, A-fold with the 99 amino acid SI-fold and subsequent mutation to resolve catastrophic interactions produced low-energy switch candidates denoted S<sub>a1</sub> and A<sub>1</sub>. The exact sequence of A<sub>1</sub> is embedded in S<sub>a1</sub> at positions 11–66 such that the  $\alpha$ 1 helices are structurally aligned (Fig. 3a, Supplementary Fig. 2A). Their final computational models were generated by Rosetta using the Relax application. The Relax protocol searches the local conformational space around an experimentally determined structure and is used only to evaluate whether the designed mutations have favorable native interactions within that limited conformational space. The designed models of S<sub>a1</sub> and A<sub>1</sub> are very similar in energy compared to the respective relaxed native structures (Supplementary Fig. 3 and Source data files).

### Structural analysis of A<sub>1</sub> and S<sub>a1</sub>

Overall, the 3 $\alpha$ -helical bundle topology of A<sub>1</sub> is very similar to the G<sub>A</sub> parent structure from which it was derived<sup>56</sup>. The sequence-specific chemical shift assignments for A<sub>1</sub> (Fig. 3b) were utilized to calculate a 3D structure with CS-Rosetta (Fig. 3c, Table 1). Our previous studies



**Fig. 2 | Summary of switches in structure and function.** **a** Structure of the S6 protein (yellow), RNA (light blue), and S15 and S18 proteins (blue) in the 30S ribosome (PDB 1FKA [<https://doi.org/10.2210/pdb1FKA/pdb>], ref. <sup>43</sup>). C-terminal amino acids of S6 are in magenta. **b** Subtilisin (wheat) is shown in complex with a model of the SI-inhibitor (yellow). The C-terminal nine amino acids of SI are shown in magenta. These positions were mutated in native S6 to generate affinity for subtilisin. The S'1-inhibitor (teal) is also shown with the altered loops in red. The subtilisin used in the modeling was the engineered RAS-specific protease. **c** The  $S_{a1}$  protein (blue and green) was generated from SI by mutating the 45 positions (mutant side chains shown with sticks). Deletion of C-terminal amino acids (blue)

switches  $S_{a1}$  into the A-fold (green). **d** The  $S_{b3}$  protein (rose and cyan) was generated from S'1 by mutating the 67 positions (mutant side chains shown with sticks). Deletion of C-terminal amino acids (cyan) or point mutation will switch  $S_{b3}$  into a B-fold (rose). **e** Model of  $S_{a2}1$  (green and blue) bound to subtilisin (wheat). Model of  $A_2$  (based on  $A_1$  structure) bound to HSA (violet). The HSA complex used PDB 2VDB (ref. <sup>50</sup>) as the template. **f** Model of  $S_{b3}$  (rose and cyan) bound to subtilisin (wheat). Model of  $B_4$  (rose) bound to Fc (mint). The Fc complex used PDB 1FCC (ref. <sup>51</sup>) as the template. The subtilisin used in the modeling and inhibition measurements was the engineered RAS-specific protease PDB 6UAO (ref. <sup>49</sup>).

indicated close correspondence of CS-Rosetta and de novo structures for A- and B-folds with high sequence identity<sup>57</sup>. The N-terminal residues 1–4 and the C-terminal residues 53–56 are disordered in the structure, consistent with <sup>1</sup>H-<sup>15</sup>N steady-state heteronuclear NOE data (Fig. 3e). Likewise,  $S_{a1}$  has the same overall  $\beta\alpha\beta\alpha\beta$ -topology as the parent S6 structure (Fig. 3d, Table 2). The backbone chemical shifts (Supplementary Fig. 4) were used in combination with main chain inter-proton NOEs (Supplementary Fig. 4) to determine a three-dimensional structure utilizing CS-Rosetta (PDB 7MNI). The conformational ensemble shows well-defined elements of secondary structure at residues 2–10 ( $\beta_1$ ), 16–32 ( $\alpha_1$ ), 40–44 ( $\beta_2$ ), 59–67 ( $\beta_3$ ), 73–81 ( $\alpha_2$ ), and 86–92 ( $\beta_4$ ). The principal difference from the native structure is that the  $\beta_2$ -strand is seven amino acids shorter in  $S_{a1}$  than in S6. Heteronuclear NOE data show overall consistency with the structure, indicating that the long loop between the  $\beta_2$ - and  $\beta_3$ -strands from residues 45–58 is more flexible than other internal regions of the polypeptide chain (Fig. 3e).

### Comparison of $A_1$ and $S_{a1}$ structures

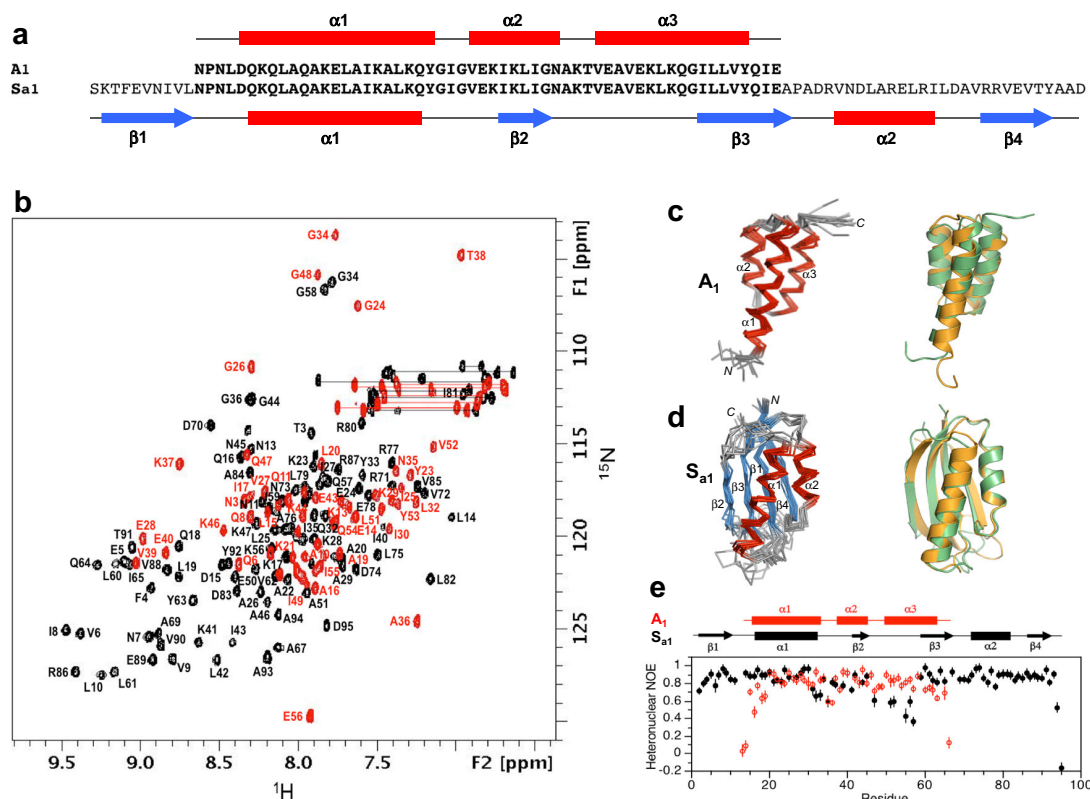
Although the 56 amino acid sequence of  $A_1$  is 100% identical to residues 11–66 of  $S_{a1}$ , a significant fraction of the backbone undergoes changes between the two structures. Most notably, while the  $\alpha_1$  helices in both  $A_1$  and  $S_{a1}$  are similar in length, the regions corresponding to

the  $\alpha_2$  and  $\alpha_3$  helices of  $A_1$  form the  $\beta_2$  and  $\beta_3$  strands of  $S_{a1}$  (Fig. 4a). Core amino acids in the  $\alpha_1$ -helix of  $A_1$  correspond with residues that also contribute to the core of  $S_{a1}$ . However, the  $\alpha_1$ -helix in  $S_{a1}$  contacts an almost entirely different set of residues (Fig. 4b). For example, amino acids L51, Y53, and I55 in the C-terminal tail of  $A_1$  do not have extensive contact with  $\alpha_1$  but the corresponding residues in  $S_{a1}$  (L61, Y63, and I65) form close core interactions with  $\alpha_1$  as part of the  $\beta_3$ -strand. Most of the other core residues contacting the  $\alpha_1$ -helix of  $S_{a1}$  are outside the 56 amino acid region coding for the  $A_1$  fold. These include F4, V6, I8, and L10 from the  $\beta_1$ -strand; A67 from the  $\beta_3$ -strand; V72, L75, and L79 from the  $\alpha_2$ -helix; and V85 from the loop between the  $\alpha_2$ -helix and the  $\beta_4$ -strand. Two additional residues, V88 and V90 ( $\beta_4$ ) also contribute significantly to the core but do not contact  $\alpha_1$ . Thus, except for the original topological alignment of the  $\alpha_1$ -helices, the cores of the  $3\alpha$  and  $\alpha/\beta$ -plait folds are largely non-overlapping. In total, approximately half of the residues participating in the  $S_{a1}$  core are not present in the  $A_1$  sequence.

### Energetics of unfolding for $A_1/S_{a1}$

Far-UV CD spectra were measured for  $S_{a1}$  and  $A_1$  and their thermal unfolding profiles were determined by measuring ellipticity at 222 nm versus temperature (Fig. 5 and Supplementary Fig. 5).  $S_{a1}$  has a





**Fig. 3 | Structure and dynamics of  $A_1$  and  $S_{a1}$ .** **a** Sequence alignment of  $A_1$  and  $S_{a1}$ , which are 100% identical over the 56 amino acid A-region. **b** Overlaid two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $S_{a1}$  (black) and  $A_1$  (red) with backbone amide assignments. Spectra were recorded at 25 and 5 °C, respectively. **c** Ensemble of 10 lowest energy CS-Rosetta structures for  $A_1$  (left panel). Superposition of the  $A_1$  structure (green) with the parent  $G_A$  fold (orange) (right panel). **d** Ensemble of 10

lowest energy CS-Rosetta structures for  $S_{a1}$  (left panel). Superposition of  $S_{a1}$  (green) with the parent S6 fold (orange) (right panel). **e** Backbone dynamics in designed proteins. Plot of  $\{^1\text{H}\}$ - $^{15}\text{N}$  steady state heteronuclear NOE values at 600 MHz versus residue for  $A_1$  (red) and for  $S_{a1}$  (black). Each set of heteronuclear NOEs was obtained from a single experiment. Errors were estimated based on the measured background noise level.

$T_M$  of -100 °C and an estimated  $\Delta G_{\text{folding}}$  of -5.3 kcal/mol at 25 °C (Fig. 5b, Supplementary Table 1)<sup>58</sup>. The  $\Delta G_{\text{folding}}$  of the parent S6 is -8.5 kcal/mol<sup>40</sup>. The Rosetta energy of the  $S_{a1}$  design is similar to that of the native sequence (Supplementary Fig. 3).  $A_1$  has a  $T_M$  of 65 °C and a  $\Delta G_{\text{folding}}$  = -4.0 kcal/mol at 25 °C<sup>58</sup> (Fig. 5a, Supplementary Table 1). The  $\Delta G_{\text{folding}}$  of the parent  $G_A$  is -5.6 kcal/mol<sup>59,60</sup>. The Rosetta energy of the  $A_1$  design is slightly more favorable than for the native sequence (Supplementary Fig. 3).

### HSA binding

Initial engineering of the fold switch was carried out without consideration of preserving function. As a result,  $A_1$  does not have detectable HSA binding affinity because two amino acids in the binding interface were mutated. Significant HSA-binding is recovered, however, when the surface mutations, E28Y and K29Y, are made in  $A_1$  (denoted  $A_2$ ). These mutations do not appear to affect the structure of  $A_1$  (Supplementary Fig. 5) but result in HSA binding of  $K_D \leq 1 \mu\text{M}$  (Supplementary Table 1). This was determined by measuring binding to immobilized HSA as described in the “Methods” section.

### Protease inhibition

$S_{a1}$  does not bind protease because C-terminal amino acids were not preserved in its design. It can be converted into a protease inhibitor, however, by replacing its three C-terminal amino acids (AAD) with DKLYRAL (denoted  $S_{a1I}$ ). A version of  $S_{a1I}$  was also made that contains the exact 56 amino acid  $A_2$  sequence by making E38Y, K39Y mutations (denoted  $S_{a2I}$ ).  $S_{a1}$ ,  $S_{a1I}$ , and  $S_{a2I}$  are

similar in structure by CD analysis (Supplementary Fig. 5). The inhibition constant of  $S_{a2I}$  with the engineered subtilisin was determined to be 50 nM as described in the “Methods” section (Supplementary Table 1). Thus, a stable A-fold with HSA-binding function can be embedded within a 99 amino acid S-fold with protease inhibitor function (Fig. 2c, e). It should be noted that all HSA contact amino acids are preserved in both the  $A_2$  and  $S_{a2I}$  sequences, but the three-dimensional topology necessary to form the HSA contact surface occurs only in the A-fold<sup>50</sup>. Nevertheless,  $S_{a2I}$  was observed to bind weakly to HSA ( $K_D \sim 100 \mu\text{M}$ , Supplementary Table 1). This weak affinity suggests that some  $S_{a2I}$  molecules may populate the 3 $\alpha$  fold even though the  $\alpha/\beta$ -plait fold strongly predominates.

### Designing a switch from $\alpha/\beta$ -plait protease inhibitor to $\beta$ -grasp IgG-binding protein

In designing an S- to B-fold switch, we used two topological alignments. The first was between S1- and B-folds, where the  $\beta_1$  strands of each fold were aligned (Supplementary Figs. 2B and 6A). The second alignment was between S1- and B-folds, where the long loop between  $\beta_2$  and  $\beta_3$  in S1 was shortened in S1 to be more consistent with natural protease inhibitors. In this scheme, the  $\alpha_1\beta_3\beta_4$  topology of the B-fold was aligned with the  $\alpha_1\beta_2\beta_3$  topology of the S1-fold (Fig. 6a, Supplementary Fig. 2C).

### Design and characterization of $B_1$ , $S_{b1}$ , $B_2$ , and $S_{b2}$

In the first approach, alignment of the  $\beta_1$ -strands of the B-fold and the S-fold and subsequent mutation to resolve catastrophic interactions

**Table 1 | Structure statistics for A<sub>1</sub>, B<sub>1</sub>, and B<sub>4</sub>**

	A <sub>1</sub>	B <sub>1</sub>	B <sub>4</sub>
<b>A. Experimental chemical shift inputs</b>			
<sup>13</sup> C $\alpha$	55	56	56
<sup>13</sup> C $\beta$	51	50	53
<sup>13</sup> CO	54	55	53
<sup>15</sup> N	54	55	54
<sup>1</sup> H <sub>N</sub>	54	55	54
<sup>1</sup> H $\alpha$	–	55	–
<b>B. RMSDs to the mean structure (Å)</b>			
Over all residues			
Backbone atoms	1.73 ± 0.47	0.86 ± 0.22	0.85 ± 0.34
Heavy atoms	2.27 ± 0.53	1.37 ± 0.29	1.43 ± 0.45
Secondary structures <sup>a</sup>			
Backbone atoms	0.90 ± 0.29	0.67 ± 0.22	0.62 ± 0.25
Heavy atoms	1.55 ± 0.42	1.12 ± 0.25	1.33 ± 0.41
<b>C. Measures of structure quality (%)</b>			
Ramachandran distribution			
Most favored	98.78 ± 1.64	92.31 ± 3.35	92.36 ± 2.71
Additionally allowed	1.22 ± 1.64	7.69 ± 3.36	7.64 ± 2.71
Generously allowed	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Disallowed	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
<b>D. Backbone RMSDs to the parent structure<sup>b</sup> (Å)</b>			
Over all residues			
Backbone atoms	2.48	0.62	0.64
Secondary structures			
Backbone atoms	1.21	0.57	0.49
<b>E. PDB/BMRB codes</b>			
PDBDEV	00000083	00000084	00000085
BMRB	50,907	50,910	50,909

<sup>a</sup>The secondary elements used were as follows: A<sub>1</sub>, residues 5–23, 27–35, 39–53; B<sub>1</sub>, residues 2–8, 13–19, 23–37, 43–46, 51–55; B<sub>4</sub>, residues: 2–8, 13–19, 23–37, 42–46, 51–55.

<sup>b</sup>The parent structure for A<sub>1</sub> is G<sub>A</sub> (PDB 2FS1). The parent structure for B<sub>1</sub> and B<sub>4</sub> is G<sub>B</sub> (PDB 1PGA). RMSDs were calculated by superimposing the mean structure from the NMR ensemble with either the mean structure (in the case of 2FS1) or the X-ray structure (in the case of 1PGA) for the parent.

produced low-energy switch candidates denoted B<sub>1</sub> and S<sub>B1</sub>. The exact sequence of B<sub>1</sub> is embedded in S<sub>B1</sub> at positions 4–59 (Supplementary Fig. 6A). The computational models of B<sub>1</sub> and S<sub>B1</sub> show relatively small increases in energy compared to the corresponding relaxed native structures (Supplementary Fig. 3). The NMR structure of B<sub>1</sub> displayed a  $\beta\beta\alpha\beta\beta$  topology identical to that of the parent B-fold, with a backbone RMSD of -0.6 Å (Supplementary Fig. 6B, C). The topology of S<sub>B1</sub> is not the same as the parent S6 structure, however, and instead has a fold similar to that of B<sub>1</sub> (Supplementary Figs. 6B, D, and 7, PDB 7MQ4). Introducing 13 mutations into S<sub>B1</sub> generated a protein denoted S<sub>B2</sub> (Supplementary Fig. 8). S<sub>B2</sub> contains four  $\beta$ -strands and two  $\alpha$ -helices and has the general features of the parent S-fold (Supplementary Fig. 9, PDB 7MN2). The 56 amino acid version of S<sub>B2</sub> (denoted B<sub>2</sub>) has a significantly higher Rosetta energy than B<sub>1</sub>, however, and is presumably unfolded (Supplementary Fig. 3). Thus, neither the B<sub>1</sub>/S<sub>B1</sub> nor B<sub>2</sub>/S<sub>B2</sub> protein pairs resulted in high identity sequences with different folds. Nonetheless, B<sub>1</sub> is 80% identical to the corresponding embedded region in the S-folded protein S<sub>B2</sub> (Supplementary Fig. 9A). The structures of B<sub>1</sub>, S<sub>B1</sub>, and S<sub>B2</sub> are described further in the Supplement and Tables 1 and 2.

### Design of S<sub>B3</sub> and B<sub>3</sub>

To improve the design of the S-to-B switch we aligned the B-fold with the S' inhibitor fold and chose an alignment that creates a topological match between  $\alpha\beta\beta\beta\beta$  in B and  $\alpha\beta\beta\beta\beta$  in S' (Supplementary Fig. 2C). Mutation to resolve deleterious interactions

**Table 2 | Structure statistics for S<sub>a1</sub>, S<sub>b1</sub>, S<sub>b2</sub>, and S<sub>b3</sub>**

	S <sub>a1</sub>	S <sub>b1</sub>	S <sub>b2</sub>	S <sub>b3</sub>
<b>A. Experimental restraint inputs</b>				
NOE restraints				
Sequential ( $i-j=1$ )	89	–	35	40
Medium range ( $1 < i-j \leq 5$ )	35	–	31	10
Long range ( $i-j > 5$ )	92	–	67	66
Hydrogen bond restraints	88	–	82	83
TALOS dihedral angle restraints	–	–	–	91
Total NOE restraint inputs	304	–	215	290
PRE restraints	41			
<b>B. Experimental chemical shift inputs</b>				
<sup>13</sup> C $\alpha$	88	79	83	–
<sup>13</sup> C $\beta$	86	70	76	–
<sup>13</sup> CO	69	67	72	–
<sup>15</sup> N	88	69	76	–
<sup>1</sup> H <sub>N</sub>	88	69	76	–
<sup>1</sup> H $\alpha$	76	45	61	–
<b>C. RMSDs to the mean structure (Å)</b>				
Over all residues <sup>a</sup>				
Backbone atoms	2.77 ± 0.82	5.47 ± 1.86	2.34 ± 0.60	2.46 ± 0.64
Heavy atoms	3.51 ± 0.85	6.32 ± 1.86	3.00 ± 0.62	3.29 ± 0.61
Secondary structures <sup>b</sup>				
Backbone atoms	1.07 ± 0.23	3.79 ± 1.50	1.08 ± 0.24	0.68 ± 0.14
		(0.71 ± 0.23) <sup>c</sup>		
Heavy atoms	1.84 ± 0.37	4.37 ± 1.42	1.78 ± 0.32	1.42 ± 0.25
		(1.24 ± 0.30) <sup>c</sup>		
<b>D. Measures of structure quality (%)</b>				
Ramachandran distribution				
Most favored	86.46 ± 4.07	92.35 ± 2.38	92.19 ± 1.83	86.30 ± 2.38
Additionally allowed	13.54 ± 4.07	7.54 ± 2.60	7.81 ± 1.83	10.52 ± 2.68
Generously allowed	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	1.47 ± 1.54
Disallowed	0.00 ± 0.00	0.12 ± 0.38	0.00 ± 0.00	1.96 ± 1.10
<b>E. Backbone RMSDs to the parent structure (Å)<sup>d</sup></b>				
Over all residues				
Backbone atoms	2.37	0.49	6.16	11.67
Secondary structures				
Backbone atoms	0.88	0.41	3.39	3.20
<b>F. PDB/BMRB codes</b>				
PDB	7MN1	7MQ4	7MN2	7MP7
BMRB	30,901	30,905	30,902	30,904

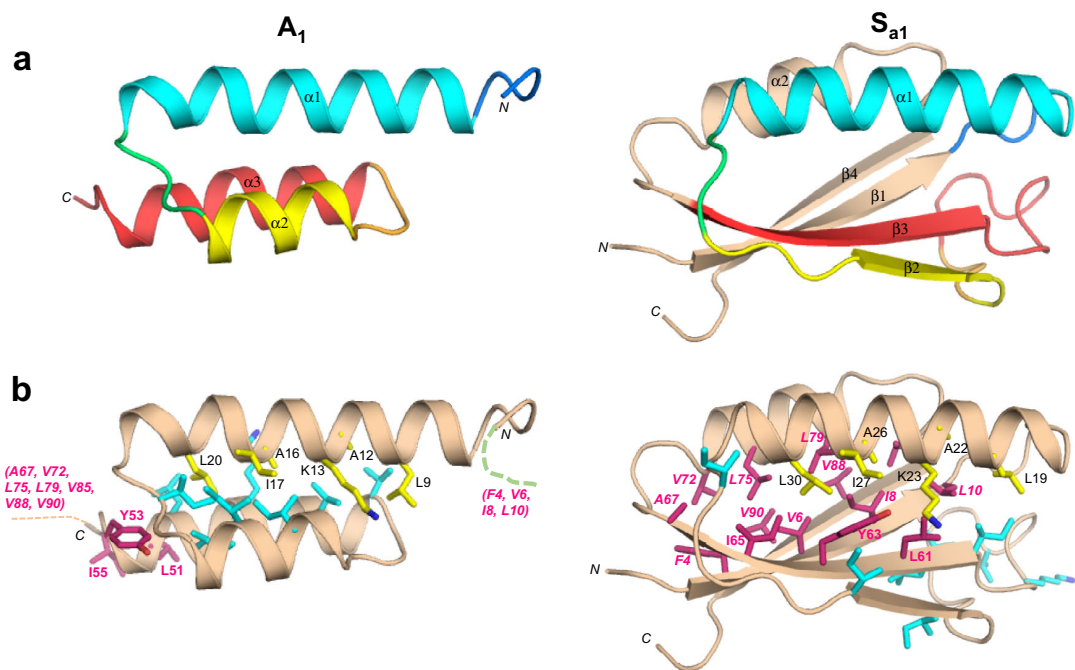
<sup>a</sup>Over all residues used as follows: S<sub>a1</sub>, 1–95; S<sub>b1</sub>, 4–85; S<sub>b2</sub>, 1–93; S<sub>b3</sub>, 1–87.

<sup>b</sup>The secondary elements used are as follows: S<sub>a1</sub>, residues 2–10, 16–32, 40–44, 59–67, 72–81, 86–92; S<sub>b1</sub>, residues 5–12, 17–24, 27–41, 46–50, 55–59, 73–83; S<sub>b2</sub>, residues 2–9, 23–32, 43–48, 59–65, 71–80, 85–91; S<sub>b3</sub>, residues 4–10, 24–37, 40–46, 51–57, 62–71, 79–85.

<sup>c</sup>RMSDs for S<sub>b1</sub> minus the putative  $\alpha$ 2 region: residues 5–12, 17–24, 27–41, 46–50, 55–59.

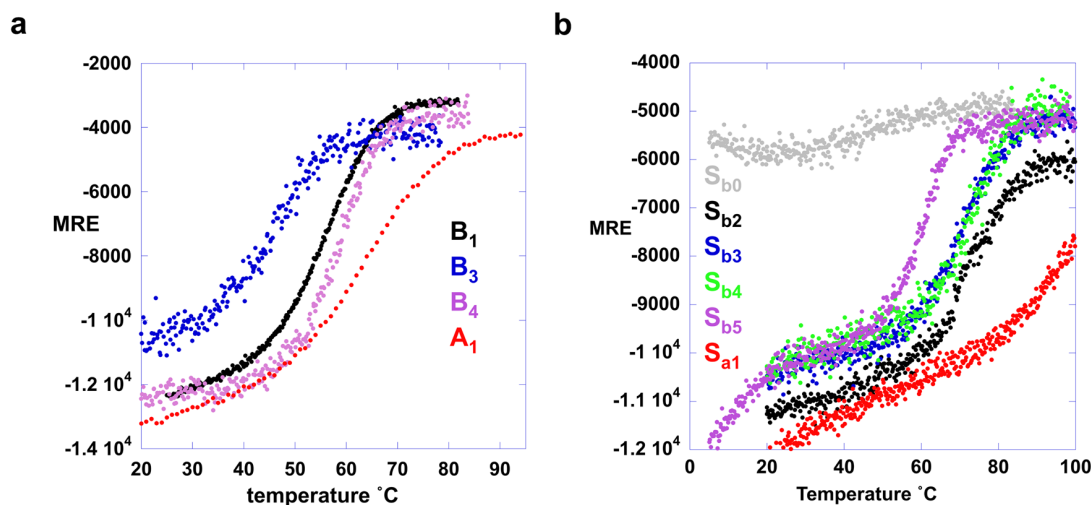
<sup>d</sup>The parent structure for S<sub>a1</sub>, S<sub>b2</sub>, and S<sub>b3</sub> is PDB 1RIS. The parent structure for S<sub>b1</sub> is PDB 1PGA. In this case, the structure alignment is over the 56 amino acid B-region of S<sub>b1</sub>.

in this alignment produced low-energy switch candidates denoted B<sub>3</sub> and S<sub>B3</sub> (Supplementary Fig. 10). The exact sequence of B<sub>3</sub> is embedded in S<sub>B3</sub> at positions 1–56. The energy of the computational model for S<sub>B3</sub> is slightly more favorable than the relaxed native structure. The designed model of B<sub>3</sub> shows relatively small



**Fig. 4 | Structural differences between the 100% sequence identical regions of  $A_1$  and  $S_{a1}$ .** **a** Main chain comparisons. (Left panel) CS-Rosetta structure of  $A_1$  with color coding for secondary structured elements. (Right panel) Corresponding color-coded regions mapped onto the CS-Rosetta structure of  $S_{a1}$ , illustrating changes in backbone conformation. Regions outside the 56 amino acid sequence of  $A_1$  are shown in wheat. **b** Side chain comparisons. (Left panel) Residues contributing

to the core of  $A_1$  from the  $\alpha_1$ -helix (yellow), and from other regions (cyan). The non- $\alpha_1$  core residues from  $S_{a1}$  (pink) do not overlap with the  $A_1$  core (see text for further details). (Right panel) Residues contributing to the core of  $S_{a1}$  from the  $\alpha_1$ -helix (yellow), and most of the other participating core residues (pink). The non- $\alpha_1$  core residues from  $A_1$  are also shown (cyan), highlighting the low degree of overlap.



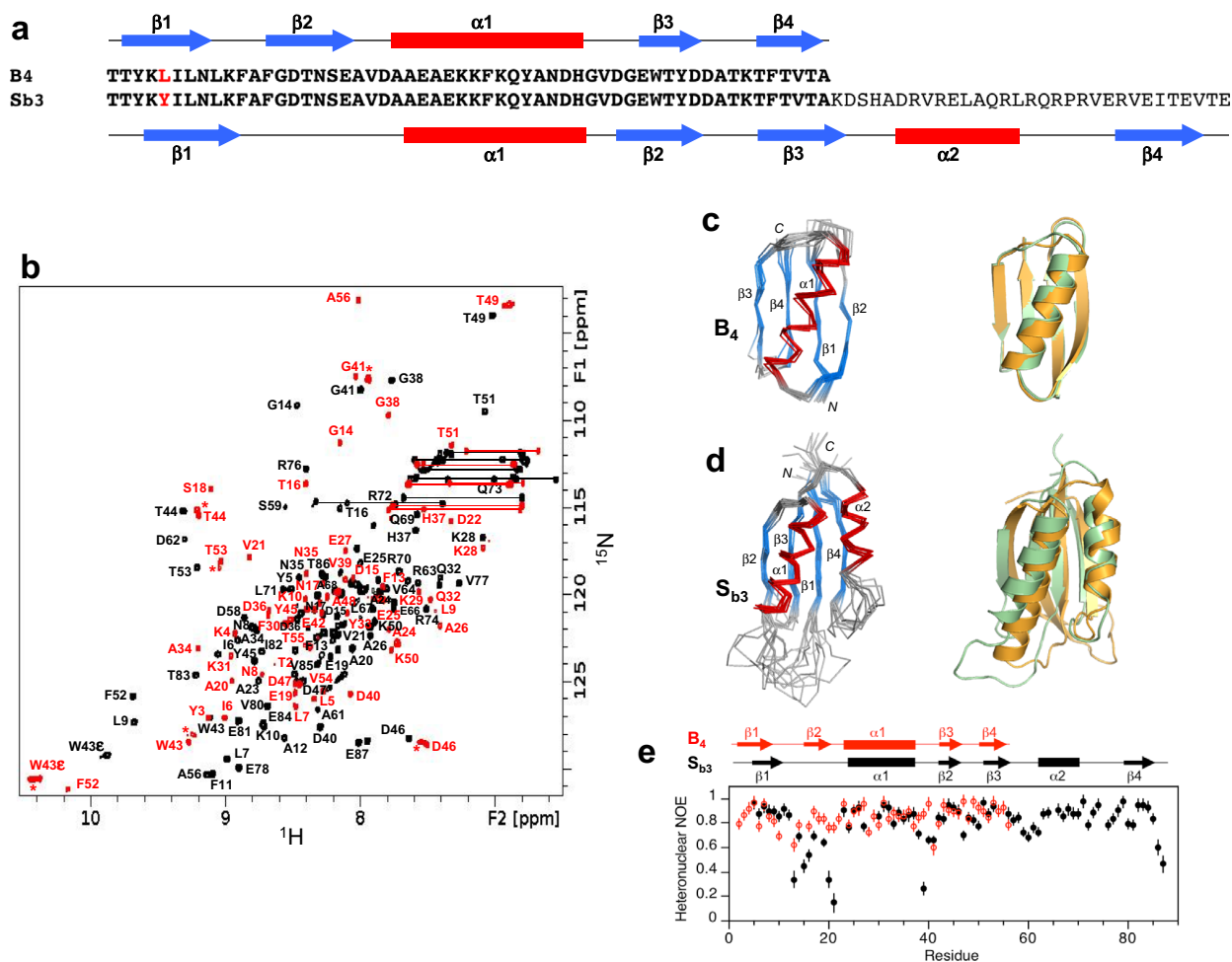
**Fig. 5 | CD melting curves.** **a** Ellipticity at 222 nm plot versus temperature for A- and B-variations. **b** Ellipticity at 222 nm plot versus temperature for S-variations.  $S_{b0}$  is a low-stability variant (F7V) of  $S_{b1}$  used to measure the temperature dependence of the unfolded state. Source data are provided as a Source Data file.

increases in energy compared to the relaxed native structure (Supplementary Fig. 3).

### Structural analysis of $S_{b3}$ and $B_3$

NMR-based structure determination indicated that  $S_{b3}$  has a  $\beta\alpha\beta\alpha\beta$  secondary structure and an S-fold topology (Fig. 6a, b, d, PDB 7MP7). Ordered regions correspond with residues 4–10 ( $\beta_1$ ), 24–37 ( $\alpha_1$ ), 42–46 ( $\beta_2$ ), 51–56 ( $\beta_3$ ), 62–70 ( $\alpha_2$ ), and 79–85 ( $\beta_4$ ). Comparison of  $S_{b3}$  with the parent S-fold indicates that the  $\beta_1/\alpha_2/\beta_4$  portion of the fold is similar in both. In contrast, the  $\beta_1$ - $\alpha_1$  loop is longer in  $S_{b3}$  (13 residues) than in the parent S-fold (5 residues), while  $\alpha_1$ ,  $\beta_2$ , the  $\beta_2$ - $\beta_3$  loop, and

$\beta_3$  are all shorter than in the parent (Fig. 6d). Consistent with the  $S_{b3}$  structure, the 13 amino acid  $\beta_1$ - $\alpha_1$  loop is highly flexible (Fig. 6e). We also expressed and purified a truncated protein corresponding to the embedded B-fold, the 56 amino acid version of  $S_{b3}$  (denoted  $B_3$ ). The 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $B_3$  at 5 °C and low concentrations (<20  $\mu\text{M}$ ) was consistent with a predominant, monomeric B-fold (Supplementary Fig. 11) but showed significant exchange broadening at 25 °C, indicative of low stability (see below). Presumably, the low stability is due to the less favorable packing of Y5 in the core of the B-fold compared with a smaller aliphatic leucine. However, additional, putatively oligomeric, species were also present for which relative



**Fig. 6 | Structure and dynamics of  $S_{b3}$  and  $B_4$ .** **a** Sequence alignment of  $B_4$  and  $S_{b3}$ , differing by one residue (L5Y) over the 56 amino acid B-region. **b** Overlaid two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $S_{b3}$  (black) and  $B_4$  (red) with backbone amide assignments. Spectra were recorded at 25 °C. The A56 peak is an aliased signal. Peaks labeled with an asterisk decrease in relative intensity as the  $B_4$  concentration is lowered, indicating the presence of a weakly associated putative dimer in addition to monomer. **c** Ensemble of 10 lowest energy CS-Rosetta structures for  $B_4$  (left

panel). Superposition of the  $B_4$  structure (green) with the parent  $G_B$  fold (orange) (right panel). **d** Ensemble of 10 lowest energy CS-Rosetta structures for  $S_{b3}$  (left panel). Superposition of  $S_{b3}$  (green) with the parent  $S_6$  fold (orange) (right panel). **e** Plot of  $^1\text{H}$ - $^{15}\text{N}$  steady-state heteronuclear NOE values at 600 MHz versus residue for  $B_4$  (red) and  $S_{b3}$  (black). Each set of heteronuclear NOEs was obtained from a single experiment. Errors were estimated based on the measured background noise level.

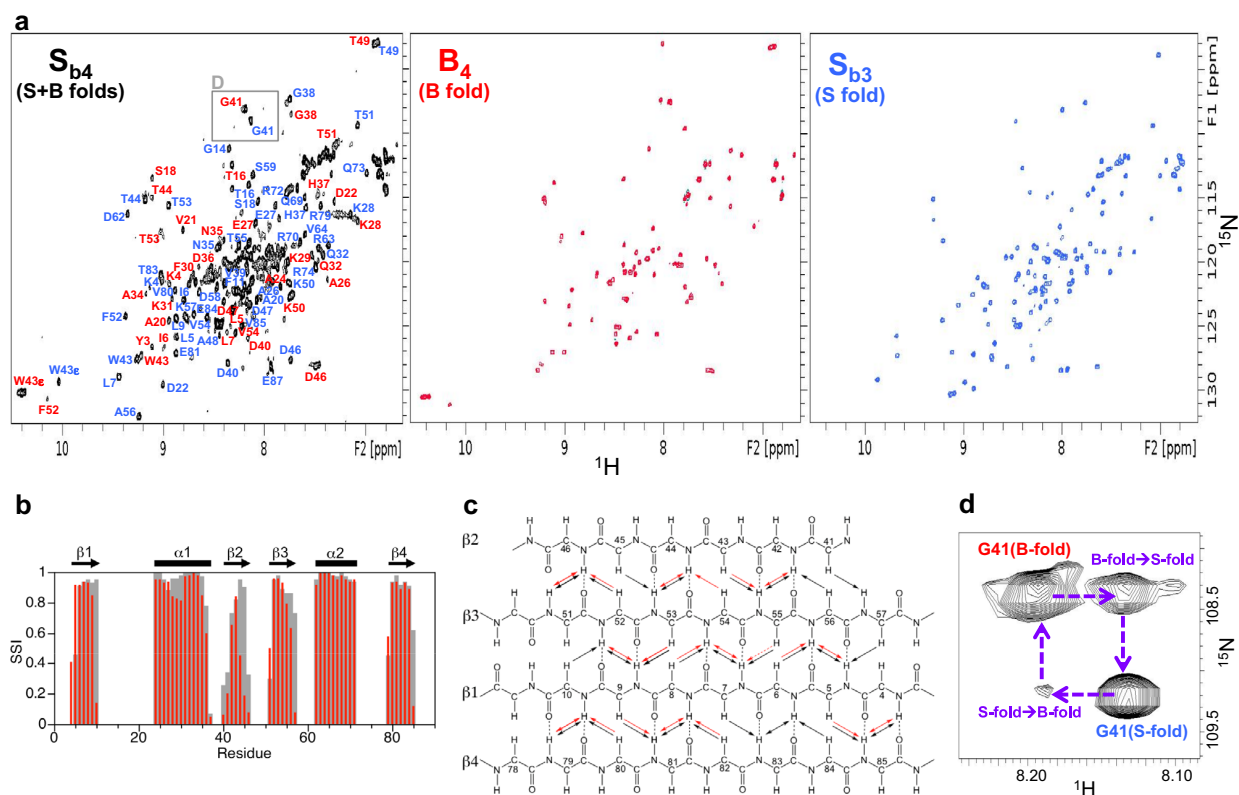
peak intensities increased with increasing protein concentration. Due to its relatively low stability and sample heterogeneity,  $B_3$  was not analyzed further structurally.

#### Design and analysis of point mutations that switch the fold of $S_{b3}$

We used the NMR structure of  $S_{b3}$  to design a point mutation, tyrosine 5 to leucine (Y5L), that would stabilize the embedded B-fold without compromising native contacts in the S-fold (Supplementary Figure 10). This mutant was therefore expected to shift the population to the B-fold. Two mutants were prepared, a Y5L mutant of  $S_{b3}$  (denoted  $S_{b4}$ ) and a Y5L mutant of  $B_3$  (denoted  $B_4$ ).  $B_4$  is indeed more stable than  $B_3$  (Fig. 5a, Supplementary Table 1). Assignment and structure determination of  $B_4$  showed its topology to be identical to the parent B-fold (Fig. 6b, c). At concentrations above 100  $\mu\text{M}$ ,  $B_4$  displayed a tendency for weak self-association similar to that seen for  $B_3$ . For  $S_{b4}$ , the HSQC spectrum exhibited approximately twice the number of amide cross-peaks relative to  $S_{b3}$  (Fig. 7a), suggesting that S- and B-states were populated simultaneously. This was confirmed by the NMR assignment and also a comparison of the HSQC spectra for  $S_{b4}$ ,  $B_4$ , and  $S_{b3}$ . A significant fraction of the  $S_{b4}$  backbone amide signals (~50 peaks) closely matched those of  $B_4$ , indicating the presence of a B-state (Supplementary Fig. 12A–C). The

close matching of these peaks is presumably because residues 1–56 in the B-state of  $S_{b4}$  are identical in sequence to  $B_4$ . The largest amide shift perturbations between the B-state of  $S_{b4}$  and  $B_4$  occur for residues proximal to the C-terminus of the B-fold, such as G41, where  $S_{b4}$  has additional residues and  $B_4$  does not. Many of the  $S_{b4}$  signals also matched well with  $S_{b3}$ , although the degree of similarity was not as extensive as with  $B_4$  (Supplementary Fig. 12D–F). More significant amide chemical shift differences between the S-state of  $S_{b4}$  and  $S_{b3}$  are likely due to the Y5L mutation, which is a relatively large change located adjacent to the core. To resolve these ambiguities, backbone resonance assignments were made for the S-state of  $S_{b4}$  (Fig. 7a, <https://doi.org/10.13018/BMR51719> see the “Methods” section for details). Comparison of  $S_{b4}$  S-state assignments with  $S_{b3}$  indicated that most of the larger amide shift perturbations were in the  $\beta 1$  and  $\beta 4$  strands. Secondary shift analysis showed that the pattern of secondary structure elements for the S-state of  $S_{b4}$  is similar to that of  $S_{b3}$  (Fig. 7b). Inter-proton NOE analysis indicated that the arrangement of the  $\beta$ -strands is also similar (Fig. 7c). Together, these results show that  $S_{b4}$  populates both S- and B-folds approximately equally at 25 °C. Moreover, a ZZ-exchange spectrum demonstrated that the S- and B-states of  $S_{b4}$  are in slow conformational exchange on the NMR timescale (Fig. 7d).





**Fig. 7 |  $S_{b4}$  is an equilibrium mixture of S- and B-states.** **a** Two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $S_{b4}$  (left),  $B_4$  (center), and  $S_{b3}$  (right) at 25 °C. In the  $S_{b4}$  spectrum, backbone amide resonance assignments are shown for the S-state (blue) and the B-state (red). **b** Confidence levels of ordered secondary structure regions for the S-state of  $S_{b4}$  at 30 °C (red) and for  $S_{b3}$  at 25 °C (gray) from chemical shifts using

TALOS-N. The secondary structure elements of  $S_{b3}$ , determined from the three-dimensional structure, are shown above the plot. **c** Summary of long-range backbone NOEs from 3D  $^{15}\text{N}$ -NOESY data for  $S_{b3}$  at 25 °C (black) and the S-state of  $S_{b4}$  at 30 °C (red). **d** Gly41 region of the 300 ms ZZ-exchange spectrum for  $S_{b4}$  at 25 °C, showing exchange peaks between the S- and B-states.

Finally, we designed a mutation of leucine 67 to arginine (L67R) in  $S_{b4}$  to destabilize the S-fold without changing the sequence of the embedded B-fold. The mutant is denoted as  $S_{b5}$  (Supplementary Fig. 10). This was expected to shift the population to the B-fold. The 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $S_{b5}$  indicates that the L67R mutation does indeed destabilize the S-fold, with the loss of S-type amide cross-peaks and the concurrent appearance of a new set of signals indicating a switch to a B-fold. The superposition of the spectrum of  $S_{b5}$  with that of  $B_4$  shows that the new signals in  $S_{b5}$  largely correspond with the spectrum of  $B_4$  (Supplementary Fig. 13). Thus, the L67R mutation shifts the equilibrium from the S-fold to the B-fold. The additional signals (-25–30) in the central region of the HSQC spectrum that are not detected in  $B_4$  are presumably due to the disordered C-terminal tail of  $S_{b5}$ . The C-terminal tail of  $S_{b5}$  does not appear to interact extensively with the B-fold, as evidenced by few changes in chemical shifts or peak intensities in the B-region of  $S_{b5}$  compared with  $B_4$ .

### Structural comparison of $S_{b3}$ and $B_4$

The aligned amino acids 1–56 of  $S_{b3}$  and  $B_4$  have 98% sequence identity, the only difference being an L5Y mutation in  $S_{b3}$  (Fig. 6a). The global folds of  $S_{b3}$  and  $B_4$  have large-scale differences, however (Fig. 8a, Supplementary Fig. 4). The  $\beta$ 1-strands, while similar in length, are in opposite directions in  $S_{b3}$  and  $B_4$ . The  $\beta$ 1-strand forms a parallel-stranded interaction with  $\beta$ 4 in  $B_4$ , but an antiparallel interaction with the corresponding  $\beta$ 3-strand in  $S_{b3}$ . Whereas residues 9–20 form the 6-residue  $\beta$ 1– $\beta$ 2 turn and the 6-residue  $\beta$ 2-strand of  $B_4$ , these same amino acids constitute the end of  $\beta$ 1 and 10 residues of the largely disordered  $\beta$ 1– $\alpha$ 1 loop in  $S_{b3}$ . The remainder of the B-region is topologically similar, with the  $\alpha$ 1/ $\beta$ 3/ $\beta$ 4 structure in  $B_4$  matching the  $\alpha$ 1/ $\beta$ 2/ $\beta$ 3 structure in  $S_{b3}$ .

Overall, however, the order of H-bonding in the 4-stranded  $\beta$ -sheets is quite different, with  $\beta$ 2 $\beta$ 3 $\beta$ 1 $\beta$ 4 in  $S_{b3}$  and  $\beta$ 3 $\beta$ 4 $\beta$ 1 $\beta$ 2 in  $B_4$ .

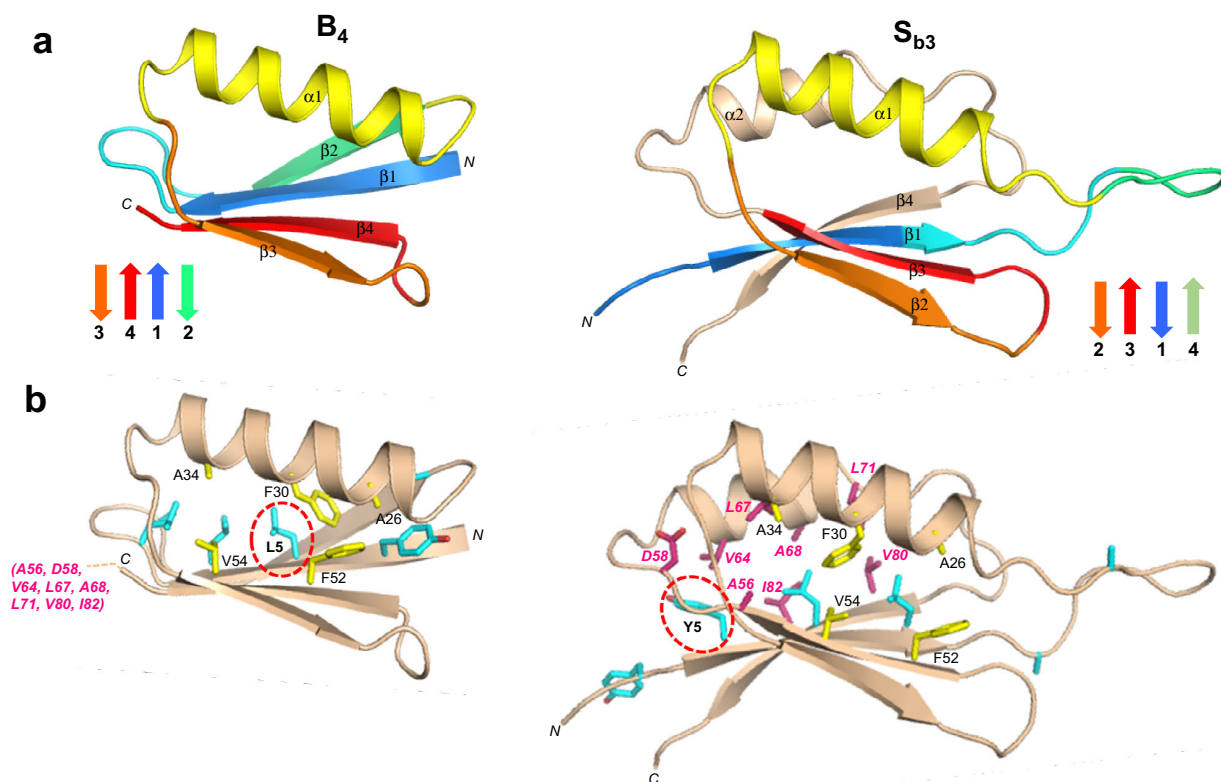
The main core residues of  $B_4$  consist of Y3, L5, L7, and L9 from  $\beta$ 1, A26, F30, and A34 from  $\alpha$ 1, and F52 and V54 from  $\beta$ 4 (Fig. 8b). In  $S_{b3}$ , the topologically equivalent regions of the core are A26, F30, and A34 from  $\alpha$ 1, and F52 and V54 from  $\beta$ 3. Residues Y5, L7, and L9 from the  $\beta$ 1 strand of  $S_{b3}$  also form part of the core, but with different packing from  $B_4$  due to the reverse orientation of  $\beta$ 1. Residues A12 and A20, which contribute to the periphery of the core in  $B_4$ , are solvent accessible in the  $\beta$ 1– $\alpha$ 1 loop of  $S_{b3}$ . Most of the remaining core residues of  $S_{b3}$  come from outside of the B-region and include amino acids from  $\beta$ 3 (A56),  $\alpha$ 2 (V64, L67, A68, L71), and  $\beta$ 4 (V80 and I82).

### Energetics of unfolding for $B_3/S_{b3}$ , $B_4/S_{b4}$ , and $S_{b5}$

Far-UV CD spectra were measured for  $B_3$ ,  $B_4$ ,  $S_{b3}$ ,  $S_{b4}$ , and  $S_{b5}$  and their thermal unfolding profiles were determined by measuring ellipticity at 222 nm versus temperature (Fig. 5, Supplementary Fig. 10, Supplementary Table 1). As described above, the predominant form of  $S_{b3}$  is an S-fold. CD and NMR analyses show that  $B_3$  is predominantly a B-fold with a  $\Delta G_{\text{folding}}$  of  $-1.2$  kcal/mol at 25 °C<sup>58</sup>. From the NMR analysis, it appears that the B-fold is in equilibrium with putatively dimeric states. This creates a situation in which the B-fold is both temperature-dependent and concentration-dependent. The predominant form at 5 °C and  $\leq 18$   $\mu\text{M}$  is the B-fold, however. The low stability and concentration-dependent behavior of  $B_3$  may indicate that some propensity for S-type conformations could persist in the 56-residue protein.

$S_{b4}$  has a temperature unfolding profile very similar to  $S_{b3}$  (Fig. 5) even though both S- and B- are approximately equally populated at 25 °C in  $S_{b4}$  (Fig. 7). This shows that the Y5L mutation results in two folds that





**Fig. 8 | Structural differences in the high (~98%) sequence identity regions of  $B_4$  and  $S_{b3}$ .** **a** Main chain comparisons. (Left panel) CS-Rosetta structure of  $B_4$  with secondary structure elements color coded. (Right panel) Corresponding color-coded regions mapped onto the CS-Rosetta structure of  $S_{b3}$ , showing changes in backbone conformation. Regions outside the 56 amino acid sequence of  $B_4$  are shown in wheat. **b** Side chain comparisons. (Left panel) Residues contributing to

the core of  $B_4$  from  $\alpha 1/\beta 3/\beta 4$  (yellow), and from other regions (cyan). The non- $\alpha 1/\beta 2/\beta 3$  core residues from  $S_{b3}$  (pink) do not overlap with the  $B_4$  core (see text for further details). (Right panel) Residues contributing to the core of  $S_{b3}$  from  $\alpha 1/\beta 2/\beta 3$  (yellow), and the other participating core residues (pink). The non- $\alpha 1/\beta 2/\beta 3$  core residues from  $B_4$  are also shown (cyan). The single L5Y amino acid difference between  $B_4$  and  $S_{b3}$  is highlighted.

are almost isoenergetic and both thermodynamically stable relative to the unfolded state. Further, because S- and B-folds are in equilibrium and approximately equally populated, the free energy of switching to the B-fold from the S-fold ( $\Delta G_{B\text{-fold}/S\text{-fold}}$ ) is  $\sim 0$  kcal/mol at 25 °C. The switch equilibrium reflects the influence of the antagonistic B-fold on the S-fold population in  $S_{b4}$ , where the leucine at residue 5 helps stabilize the alternative B-state at the expense of the S-state. Thermal denaturation by CD shows that  $B_4$  has a  $\Delta G_{\text{folding}} = -4.1$  kcal/mol at 25 °C<sup>58</sup>. The thermal unfolding profile of  $S_{b5}$  shows a low-temperature transition with a midpoint  $-10$  °C and a major transition with a midpoint of  $-60$  °C (Fig. 5b). The NMR analysis indicates that the major transition is unfolding of the B-fold. Thus, the arginine at 67 in  $S_{b5}$  makes the B-fold more favorable by making the S-fold unfavorable, consistent with the change in population from mixed to B-fold observed by NMR.

### Protease inhibition

The  $S_{b3}$  protein is closely related to S1 but lacks inhibitor function because C-terminal amino acids were changed in the design of the switch. It can be converted into a protease inhibitor, however, by altering C-terminal amino acids VTE to DKLYRAL. This mutant is denoted  $S_{b3}I$ .  $S_{b3}$  and  $S_{b3}I$  appear similar in structure by CD analysis (Supplementary Fig. 10). The  $K_I$  for  $S_{b3}I$  with the engineered subtilisin was determined to be 50 nM (Supplementary Table 1).

### IgG binding

Binding to IgG was determined for  $B_3$  and  $S_{b3}I$  (Supplementary Table 1).  $B_3$  and  $S_{b3}I$  bound to IgG Sepharose with  $K_D \leq 1 \mu\text{M}$  and  $10 \mu\text{M}$ , respectively. Presumably,  $S_{b3}I$  has significant IgG-binding

activity because the  $\alpha 1\beta 3$  IgG binding surface of the B-fold is largely preserved in the S-fold. Thus,  $S_{b3}I$  is a dual-function protein with both IgG-binding and protease inhibitor functions (Fig. 2f).

### Discussion

The entire network of intersecting pathways between the S-, A-, and B-folds is summarized in Fig. 9. The first node on the pathway is a functional switch from RNA binding protein to protease inhibitor without a fold switch. The  $\alpha/\beta$  plait is a common fold, and proteins with this basic topology include many different functions<sup>42</sup>. Engineering the SI and S1 nodes illustrates how protease inhibitor function can arise in the  $\alpha/\beta$  plait topology with a few mutations. Replacing only C-terminal amino acids in the S6 protein creates interaction with the substrate binding cleft of the protease (Fig. 2a, b). This C-terminal interaction plus adventitious contact between the  $\beta$ -sheet surface of the  $\alpha/\beta$  plait and two  $\alpha$ -helices in the protease result in protease inhibition in the 50 nM range. Based on the structure of S6 in the 30S complex, the C-terminal modification may not have major effects on binding interactions with ribosomal RNA and the S15 protein (Fig. 2a)<sup>43</sup>. Thus, the transition from RNA binding protein to protease inhibitor likely is uninterrupted. An insertion in the  $\beta 1$ - $\alpha 1$  loop and a deletion  $\beta 2$ - $\beta 3$  loop in the SI-inhibitor creates a topology that more closely resembles natural prodomain-type inhibitors<sup>44,46,61</sup> and creates an  $\alpha 1\beta 2\beta 3$  motif in the S'-fold that is similar to the  $\alpha 1\beta 3\beta 4$  motif of the B-fold. This topological similarity brings the S1 closer to an intersection with the B-fold. Thus, SI and S1 nodes are both functional switches and branch points for switching the S-fold into the A- and B-folds, respectively.



That is,  $\Delta G_{B\text{-fold}/S\text{-fold}}$  of  $S_{b4}$  is  $\sim 0$  kcal/mol at 25 °C. One further substitution that destabilizes the S-fold (L67R) shifts the population of  $S_{b5}$  to a stable B-fold ( $\Delta G_{B\text{-fold}/S\text{-fold}} \leq -5$  kcal/mol) (Fig. 9).

The existence of nodes between folds has implications for the evolution of new functions. In the case of the S/A node, all contact amino acids for HSA exist within the S-fold of the protease inhibitor  $S_{a2l}$  albeit in a cryptic topology. Deletion of amino acids 67–99 ( $A_2$ ) results in loss of inhibitor function and a fold switch from  $\alpha/\beta$  plait to  $3\alpha$ . Acquisition of HSA binding activity ( $K_D < 1 \mu\text{M}$ ) results from unmasking the cryptic HSA binding amino acids via the fold switch (Fig. 2e). This level of binding affinity could be biologically relevant since the concentration of HSA in serum is  $>500 \mu\text{M}$ <sup>63</sup>. In the case of the S/B node, the  $\alpha 1\beta 3$  motif contains all IgG contact amino acids and  $S_{b3l}$  has some affinity for both IgG ( $K_D = 10 \mu\text{M}$ ) and protease ( $K_i = 50 \text{ nM}$ ). In this case, the Y5L mutation ( $S_{b4}$ ) or a deletion of 57–91 ( $B_4$ ) causes a fold switch from  $\alpha/\beta$  plait to the  $\beta$ -grasp and results in tighter IgG binding ( $K_D \leq 1 \mu\text{M}$ ) (Fig. 2f). This level of binding affinity could also be biologically relevant since the concentration of IgG in serum is  $>50 \mu\text{M}$  (or  $>100 \mu\text{M}$  Fc binding sites)<sup>64</sup>. We have previously shown that an A-fold with HSA binding function can be switched to a B-fold with IgG-binding function via single amino acid substitutions that switch the folds and unmask cryptic contact amino acids for the two ligands<sup>29,32</sup>.

In conclusion, it was possible to connect three common folds in a network of high-identity nodes that form critical points between two folds. As in other complex systems, a small change in a protein near a critical point can have a “butterfly effect” on how the folds are populated. This property of the protein folding code means that proteins with multiple folds and functions can exist in highly identical amino acid sequences. This suggests that the evolution of new folds and functions sometimes can follow uninterrupted mutational pathways.

## Methods

### Mutagenesis, protein expression and purification

Mutagenesis was carried out using Q5® Site-Directed Mutagenesis Kits (NEB).  $G_A$  and  $G_B$  variants were cloned into a vector (pH0720) encoding the sequence:

MEAVDANSLA QAKEAAIKEL KQYIGDKYI KLINNAKTVE  
GVESLKNEIL KALPTEGSGN TIRVIVSVDK AKFNPHEVLG IGGHIVYQFK  
LIPAVVDVDP ANAVGKLLKMM PGVEKVEFDH QYRGL

as an N-terminal fusion domain<sup>56</sup>. Cell growth was carried out by auto-induction<sup>29,65</sup>. Cells were harvested by centrifugation at  $3750 \times g$  for 20 min and lysed by sonication on ice in 0.1 M KPi, pH 7.2. Cellular debris was pelleted by centrifugation at  $10,000 \times g$  for 15 min. Supernatant was clarified by centrifugation at  $45,000 \times g$  for 30 min. Proteins were purified using a second generation of the affinity-cleavage tag system employed previously to purify switch proteins<sup>29,66</sup>. The second-generation tag results in high-level soluble expression of the switch proteins and also enables the capture of the fusion protein by binding tightly to an immobilized processing protease via the C-terminal EFDHQYRGL sequence. Loading and washing were at 5 mL/min for a 5 mL *Im-Prot* column using a running buffer of 20 mM KPi, pH 6.8. The amount of washing required for high purity depends on the stickiness of the target protein and how much of it is bound to the column. We typically wash with 10 column volumes (CV) of wash solution followed by 3 CV 0.5 M NaCl and then -10 CV running buffer. This can be repeated as necessary. The 0.5 M NaCl shots are repeated until the amount of absorbance released with each high-salt shot becomes small and constant. All the high-salt solution is washed out before initiating the cleavage. The target protein was cleaved from the *Im-Prot* column by injecting 15 mL of imidazole solution (0.1 mM) at 1 mL/min, 22 °C. The cleaved protein typically elutes as a sharp peak in 2–3 CV. The purified protein was then concentrated to 0.2–0.3 mM, as

required for NMR analysis. The columns were regenerated by injecting 15 mL of 0.1 N  $\text{H}_3\text{PO}_4$  (0.227 mL concentrated phosphoric acid (85%) per 100 mL) at a flow rate of  $\sim 1$  CV/min. The wash solution was neutralized immediately after stripping. The purification system is available from Potomac Affinity Proteins.

Protease inhibitor proteins were purified by binding to *Im-Prot* media and then stripping off the purified inhibitor with 0.1 N  $\text{H}_3\text{PO}_4$ . Samples were then immediately neutralized by adding 1/10 volume 1 M  $\text{K}_2\text{HPO}_4$ .

### Rosetta calculations

Rosetta energies of all designed structures were generated using the Slow Relax routine<sup>54</sup>. 1000 decoys were calculated for each design. PDB coordinates and energy parameters for the lowest energy decoy for each design are included as supplemental files.

### Circular dichroism (CD)

CD measurements were performed in 100 mM KPi, pH 7.2 with a Jasco spectropolarimeter, model J-1100 with a Peltier temperature controller. Quartz cells with path lengths of 0.1 and 1 cm were used for protein concentrations of 3 and 30  $\mu\text{M}$ , respectively. The ellipticity results were expressed as mean residue ellipticity,  $[\theta]$ , deg  $\text{cm}^2 \text{dmol}^{-1}$ . Ellipticities at 222 nm were continuously monitored at a scanning rate of  $0.5^\circ/\text{min}$ . Reversibility of denaturation was confirmed by comparing the CD spectra at 20 °C before melting and after heating to 100 °C and cooling to 20 °C.

### Measuring HSA and IgG binding affinity

Affinity of proteins to HSA and IgG was determined by their retention on the immobilized ligands. HSA and rabbit IgG were immobilized by reaction with NHS-activated Sepharose 4 Fast Flow (Cytiva) according to the manufacturer's instructions. The concentration of immobilized HSA was 100  $\mu\text{M}$ . The concentration of immobilized IgG was 50  $\mu\text{M}$  (i.e. 100  $\mu\text{M}$  Fc binding sites). Generally, 0.2 mL of a 5  $\mu\text{M}$  solution of the test protein was injected into a 5 mL column at a flow rate of 0.5 mL/min. Determination of binding affinity assumes that binding is in rapid equilibrium such that the elution volume is proportional to the fraction of test protein bound to 100  $\mu\text{M}$  of binding sites. Proteins that are completely retained after 20 column volumes (CV) are assessed to have  $K_D \leq 1 \mu\text{M}$ . Completely retained proteins are stripped from the column with 0.1 N  $\text{H}_3\text{PO}_4$  at the end of the run.

### Measuring protease inhibition

Competitive inhibition constants ( $K_i$ ) were determined using the fluorogenic peptide substrate QEEYSAM-AMC (7-amino-4-methylcoumarin) purchased from AnaSpec Inc. and a highly specific, engineered protease known as RASProtease(I)<sup>49</sup>. Competitive inhibition constants ( $K_i$ ) were measured by determining the  $K_{M(\text{apparent})}$  in the presence of 0, 50, and 100 nM of each inhibitor protein. The reactions were carried out in 100 mM KPi, 10 mM imidazole, 0.005% tween-20, pH 7.0 at 25 °C with 1 nM RASProtease(I). The QEEYSAM-AMC concentrations used to determine  $K_M$  and  $K_{M(\text{apparent})}$  were 0.1, 0.5, 1, 2, 5, and 10  $\mu\text{M}$ . Initial rates were determined with a BioTek Synergy MT fluorescence microplate reader (Ex: 360/40, Em: 460/40) by measuring the release of the fluorescent AMC group via hydrolysis of the amide bond. Highly pure ( $\geq 98\%$ ) protease and inhibitor proteins were used for all kinetic experiments.

### NMR spectroscopy

Isotope-labeled samples were prepared at 0.2–0.3 mM concentrations in 100 mM potassium phosphate buffer (pH 7.0) containing 5%  $\text{D}_2\text{O}$ . NMR spectra were collected using Topspin3.6.1 software on Bruker AVANCE III 600 and 900 MHz spectrometers fitted with Z-gradient  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple resonance cryoprobes. Standard double and triple resonance experiments (HNCACB, CBCA(CO)NH, HNCOC, HN(CA)CO, and



HNHA) were utilized to determine main chain NMR assignments. Inter-proton distances were obtained from 3D  $^{15}\text{N}$ -edited NOESY and 3D  $^{13}\text{C}$ -edited NOESY spectra with a mixing time of 150 ms. NmrPipe<sup>67</sup> was used for data processing and analysis was done with Sparky<sup>68</sup>. Two-dimensional  $\{^1\text{H}\}$ - $^{15}\text{N}$  steady-state heteronuclear NOE experiments were acquired with a 5 s relaxation delay between experiments. Errors in heteronuclear NOEs were estimated based on the background noise level. Chemical shift perturbations were calculated using  $\Delta\delta_{\text{total}} = ((W_{\text{H}}\Delta\delta_{\text{H}})^2 + (W_{\text{N}}\Delta\delta_{\text{N}})^2)^{1/2}$ , where  $W_{\text{H}}$  is 1,  $W_{\text{N}}$  is 0.2, and  $\Delta\delta_{\text{H}}$  and  $\Delta\delta_{\text{N}}$  represent  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift changes, respectively. For PRE experiments on  $\text{S}_{\text{b1}}$ , single-site cysteine mutant samples were incubated with 10 equivalents of (1-oxyl-2,2,5,5-tetramethylpyrrolidine-3-methyl) methanethiosulfonate (MTSL, Santa Cruz Biotechnology) at 25 °C for 1 h and completion of labeling was confirmed by MALDI mass spectrometry. Control samples were reduced with 10 equivalents of sodium ascorbate. Backbone amide peak intensities of the oxidized and reduced states were analyzed using Sparky. Three-dimensional structures were calculated with CS-Rosetta3.2 using experimental backbone  $^{15}\text{N}$ ,  $^1\text{H}_{\text{N}}$ ,  $^1\text{H}_{\alpha}$   $^{13}\text{C}_{\alpha}$ ,  $^{13}\text{C}_{\beta}$ , and  $^{13}\text{C}_{\text{O}}$  chemical shift restraints and were either validated by comparison with experimental backbone NOE patterns ( $\text{A}_{\text{1}}$ ,  $\text{B}_{\text{1}}$ ,  $\text{B}_{\text{4}}$ ,  $\text{S}_{\text{b1}}$ ) or directly employed interproton NOEs ( $\text{S}_{\text{a1}}$ ,  $\text{S}_{\text{b2}}$ ) or PREs ( $\text{S}_{\text{b1}}$ ) as additional restraints. One thousand CS-Rosetta structures were calculated from which the 10 lowest energy structures were chosen. For  $\text{S}_{\text{b3}}$ , CS-Rosetta failed to converge to a unique low-energy topology, producing an approximately even mixture of S- and B-type folds despite the chemical shifts and NOE pattern indicating an S-fold. In this case, CNS1.6<sup>9</sup> was employed to determine the structure<sup>56</sup>, including backbone dihedral restraints from chemical shift data using TALOS-N<sup>70</sup>. The backbone resonances for the S-state of  $\text{S}_{\text{b4}}$  were assigned using triple resonance methods as above, under conditions where the S-state is more favorably populated (30 °C, 100 mM KPi, 200 mM sodium chloride, pH 7.0). Amide assignments were then transferred to the two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $\text{S}_{\text{b4}}$  at 25 °C in 100 mM KPi, pH 7.0. Inter-proton NOEs for the S-state of  $\text{S}_{\text{b4}}$  were obtained at the 30 °C/high salt condition, employing a 3D  $^{15}\text{N}$ -edited NOESY spectrum with a 150 ms mixing time. A two-dimensional ZZ-exchange  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum was recorded on  $\text{S}_{\text{b4}}$  using a mixing time of 300 ms (25 °C, 100 mM KPi, pH 7.0)<sup>71,72</sup>. Protein structures were displayed and analyzed utilizing PROCHECK-NMR<sup>73</sup>, MOLMOL<sup>74</sup> and PyMol (Schrodinger)<sup>55</sup>.

### Reporting summary

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

### Data availability

The NMR structures generated in this study have been deposited in the PDB: [<https://doi.org/10.2210/pdb7MNI/pdb>]; [<https://doi.org/10.2210/pdb7MQ4/pdb>]; [<https://doi.org/10.2210/pdb7MN2/pdb>]; [<https://doi.org/10.2210/pdb7MP7/pdb>]; [[https://pdb-dev.wwpdb.org/entry.html?PDBDEV\\_00000083](https://pdb-dev.wwpdb.org/entry.html?PDBDEV_00000083)]; [[https://pdb-dev.wwpdb.org/entry.html?PDBDEV\\_00000084](https://pdb-dev.wwpdb.org/entry.html?PDBDEV_00000084)]; [[https://pdb-dev.wwpdb.org/entry.html?PDBDEV\\_00000085](https://pdb-dev.wwpdb.org/entry.html?PDBDEV_00000085)]. NMR Assignments have been deposited in the BMRB: [<https://doi.org/10.13018/BMR30901>]; [<https://doi.org/10.13018/BMR30902>]; [<https://doi.org/10.13018/BMR30904>]; [<https://doi.org/10.13018/BMR30905>]; [<https://doi.org/10.13018/BMR50907>]; [<https://doi.org/10.13018/BMR50909>]; [<https://doi.org/10.13018/BMR50910>]; [<https://doi.org/10.13018/BMR51719>]. The structures referenced in this paper are publicly available in the PDB: [<https://doi.org/10.2210/pdb1FKA/pdb>]; [<https://doi.org/10.2210/pdb2VDB/pdb>]; [<https://doi.org/10.2210/pdb1FCC/pdb>]; [<https://doi.org/10.2210/pdb6UAO/pdb>]; [<https://doi.org/10.2210/pdb2LHC/pdb>]; [<https://doi.org/10.2210/pdb1IRIS/pdb>]. Source data are provided with this paper. Design models are provided as files in the source data. Source data are provided with this paper.

### References

- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
- Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
- Ambroggio, X. I. & Kuhlman, B. Design of protein conformational switches. *Curr. Opin. Struct. Biol.* **16**, 525–530 (2006).
- Bryan, P. N. & Orban, J. Proteins that switch folds. *Curr. Opin. Struct. Biol.* **20**, 482–488 (2010).
- Dishman, A. F. et al. Evolution of fold switching in a metamorphic protein. *Science* **371**, 86–90 (2021).
- Wei, K. Y. et al. Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proc. Natl Acad. Sci. USA* **117**, 7208–7215 (2020).
- Anderson, W. J., Van Dorn, L. O., Ingram, W. M. & Cordes, M. H. Evolutionary bridges to new protein folds: design of C-terminal Cro protein chameleon sequences. *Protein Eng. Des. Sel.* **24**, 765–771 (2011).
- Burmann, B. M. et al. An  $\alpha$  helix to  $\beta$  barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150**, 291–303 (2012).
- Kulkarni, P. et al. Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. *Protein Sci.* **27**, 1557–1567 (2018).
- Dishman, A. F. & Volkman, B. F. Design and discovery of metamorphic proteins. *Curr. Opin. Struct. Biol.* **74**, 102380 (2022).
- Alberstein, R. G., Guo, A. B. & Kortemme, T. Design principles of protein switches. *Curr. Opin. Struct. Biol.* **72**, 71–78 (2022).
- Rackovsky, S. Nonlinearities in protein space limit the utility of informatics in protein biophysics. *Proteins* **83**, 1923–1928 (2015).
- Chen, S. H., Meller, J. & Elber, R. Comprehensive analysis of sequences of a protein switch. *Protein Sci.* **25**, 135–146 (2016).
- Li, W., Kinch, L. N., Karplus, P. A. & Grishin, N. V. ChSeq: A database of chameleon sequences. *Protein Sci.* **24**, 1075–1086 (2015).
- Wolynes, P. G. Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* **119**, 218–230 (2015).
- Holzgräfe, C. & Wallin, S. Smooth functional transition along a mutational pathway with an abrupt protein fold switch. *Biophys. J.* **107**, 1217–1225 (2014).
- Scheraga, H. A. & Rackovsky, S. Homolog detection using global sequence properties suggests an alternate view of structural encoding in protein sequences. *Proc. Natl Acad. Sci. USA* **111**, 5225–5229 (2014).
- Ha, J. H. & Loh, S. N. Protein conformational switches: from nature to design. *Chemistry* **18**, 7984–7999 (2012).
- Yadid, I., Kirshenbaum, N., Sharon, M., Dym, O. & Tawfik, D. S. Metamorphic proteins mediate evolutionary transitions of structure. *Proc. Natl Acad. Sci. USA* **107**, 7287–7292 (2010).
- Lichtarge, O. & Wilkins, A. Evolution: a guide to perturb protein function and networks. *Curr. Opin. Struct. Biol.* **20**, 351–359 (2010).
- Rollins, N. J. et al. Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **51**, 1170–1176 (2019).
- Sikosek, T., Chan, H. S. & Bornberg-Bauer, E. Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. *Proc. Natl Acad. Sci. USA* **109**, 14888–14893 (2012).
- Chen, N., Das, M., LiWang, A. & Wang, L. P. Sequence-based prediction of metamorphic behavior in proteins. *Biophys. J.* **119**, 1380–1390 (2020).
- Porter, L. L. & Looger, L. L. Extant fold-switching proteins are widespread. *Proc. Natl Acad. Sci. USA* **115**, 5968–5973 (2018).
- Bedford, J. T., Poutsma, J., Diawara, N. & Greene, L. H. The nature of persistent interactions in two model  $\beta$ -grasp proteins reveals the



- advantage of symmetry in stability. *J. Comput. Chem.* **42**, 600–607 (2021).
27. Sykes, J., Holland, B. R. & Charleston, M. A. A review of visualisations of protein fold networks and their relationship with sequence and function. *Biol. Rev. Camb. Philos. Soc.* <https://doi.org/10.1111/brv.12905> (2022).
  28. Ambroggio, X. I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161 (2006).
  29. Alexander, P. A., He, Y., Chen, Y., Orban, J. & Bryan, P. N. A minimal sequence code for switching protein structure and function. *Proc. Natl Acad. Sci. USA* **106**, 21149–21154 (2009).
  30. Davey, J. A., Damry, A. M., Goto, N. K. & Chica, R. A. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280–1285 (2017).
  31. He, Y., Chen, Y., Alexander, P., Bryan, P. N. & Orban, J. NMR structures of two designed proteins with high sequence identity but different fold and function. *Proc. Natl Acad. Sci. USA* **105**, 14412–14417 (2008).
  32. He, Y., Chen, Y., Alexander, P. A., Bryan, P. N. & Orban, J. Mutational tipping points for switching protein folds and functions. *Structure* **20**, 283–291 (2012).
  33. Falkenberg, C., Bjorck, L. & Akerstrom, B. Localization of the binding site for streptococcal protein G on human serum albumin. Identification of a 5.5-kilodalton protein G binding albumin fragment. *Biochemistry* **31**, 1451–1457 (1992).
  34. Frick, I. M. et al. Convergent evolution among immunoglobulin G-binding bacterial proteins. *Proc. Natl Acad. Sci. USA* **89**, 8532–8536 (1992).
  35. Myhre, E. B. & Kronvall, G. Heterogeneity of nonimmune immunoglobulin Fc reactivity among gram-positive cocci: description of three major types of receptors for human immunoglobulin G. *Infect. Immun.* **17**, 475–482 (1977).
  36. Reis, K. J., Ayoub, E. M. & Boyle, M. D. P., Streptococcal Fc receptors. II. Comparison of the reactivity of a receptor from a group C streptococcus with staphylococcal protein A. *J. Immunol.* **132**, 3098–3102 (1984).
  37. Lindberg, M. O., Haglund, E., Hubner, I. A., Shakhnovich, E. I. & Oliveberg, M. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc. Natl Acad. Sci. USA* **103**, 4083–4088 (2006).
  38. Haglund, E., Lindberg, M. O. & Oliveberg, M. Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J. Biol. Chem.* **283**, 27904–27915 (2008).
  39. Haglund, E. et al. The HD-exchange motions of ribosomal protein S6 are insensitive to reversal of the protein-folding pathway. *Proc. Natl Acad. Sci. USA* **106**, 21619–21624 (2009).
  40. Haglund, E. et al. Trimming down a protein structure to its bare foldons: spatial organization of the cooperative unit. *J. Biol. Chem.* **287**, 2731–2738 (2012).
  41. Lindahl, M. et al. Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249–1254 (1994).
  42. Day, R., Beck, D. A., Armen, R. S. & Daggett, V. A consensus view of fold space: combining SCOP, CATH, and the Dali Domain Dictionary. *Protein Sci.* **12**, 2150–2160 (2003).
  43. Schluenzen, F. et al. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **102**, 615–623 (2000).
  44. Gallagher, T. D., Gilliland, G., Wang, L. & Bryan, P. The prosegment-subtilisin BPN' complex: crystal structure of a specific foldase. *Structure* **3**, 907–914 (1995).
  45. Tangrea, M. A. et al. Stability and global fold of the mouse pro-hormone convertase 1 pro-domain. *Biochemistry* **40**, 5488–5495 (2001).
  46. Tangrea, M. A., Bryan, P. N., Sari, N. & Orban, J. Solution structure of the pro-hormone convertase 1 pro-domain from *Mus musculus*. *J. Mol. Biol.* **320**, 801–812 (2002).
  47. Sari, N. et al. Hydrogen-deuterium exchange in free and prodomain-complexed subtilisin. *Biochemistry* **46**, 652–658 (2007).
  48. Orengo, C. A. & Thornton, J. M. Alpha plus beta folds revisited: some favoured motifs. *Structure* **1**, 105–120 (1993).
  49. Chen, Y. et al. Engineering subtilisin proteases that specifically degrade active RAS. *Commun. Biol.* **4**, 299 (2021).
  50. Lejon, S., Frick, I. M., Bjorck, L., Wikstrom, M. & Svensson, S. Crystal structure and biological implications of a bacterial albumin binding module in complex with human serum albumin. *J. Biol. Chem.* **279**, 42924–42928 (2004).
  51. Sauer-Eriksson, A. E., Keywegt, G. J., Uhlen, M. & Jones, T. A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265–278 (1995).
  52. Alexander, P. A., Rozak, D. A., Orban, J. & Bryan, P. N. Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code. *Biochemistry* **44**, 14045–14054 (2005).
  53. Alexander, P. A., He, Y., Chen, Y., Orban, J. & Bryan, P. N. The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl Acad. Sci. USA* **104**, 11963–11968 (2007).
  54. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
  55. Delano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA, 2002).
  56. He, Y. et al. Structure, dynamics, and stability variation in bacterial albumin binding modules: implications for species specificity. *Biochemistry* **45**, 10102–10109 (2006).
  57. Shen, Y. et al. De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci.* **19**, 349–356 (2010).
  58. Chen, Y. et al. Rules for designing protein fold switches and their implications for the folding code. Preprint at [bioRxiv](https://doi.org/10.1101/2021.05.18.444643) <https://doi.org/10.1101/2021.05.18.444643> (2021).
  59. Rozak, D. A., Orban, J. & Bryan, P. N. G148-GA3: a streptococcal virulence module with atypical thermodynamics of folding optimally binds human serum albumin at physiological temperatures. *Biochim. Biophys. Acta* **1753**, 226–233 (2005).
  60. He, Y., Chen, Y., Rozak, D. A., Bryan, P. N. & Orban, J. An artificially evolved albumin binding module facilitates chemical shift epitope mapping of GA domain interactions with phylogenetically diverse albumins. *Protein Sci.* **16**, 1490–1494 (2007).
  61. He, Y. et al. Solution NMR structure of a sheddase inhibitor prodomain from the malarial parasite *Plasmodium falciparum*. *Proteins* **80**, 2810–2817 (2012).
  62. Alexander, P., Fahnestock, S., Lee, T., Orban, J. & Bryan, P. Thermodynamic analysis of the folding of the Streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. *Biochemistry* **31**, 3597–3603 (1992).
  63. Chien, S.-C., Chen, C.-Y., Lin, C.-F. & Yeh, H.-I. Critical appraisal of the role of serum albumin in cardiovascular disease. *Biomark. Res.* **5**, 31 (2017).
  64. Gonzalez-Quintela, A. et al. Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clin. Exp. Immunol.* **151**, 42–50 (2008).
  65. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
  66. Ruan, B., Fisher, K. E., Alexander, P. A., Doroshko, V. & Bryan, P. N. Engineering subtilisin into a fluoride-triggered processing protease

- useful for one-step protein purification. *Biochemistry* **43**, 14539–14546 (2004).
67. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
  68. Goddard, D. & Kneller, D. G. SPARKY 3 Vol. 3 (University of California, San Francisco, 2004).
  69. Brunger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D (Biol. Crystallogr.)* **54**, 905–921 (1998).
  70. 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).
  71. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of <sup>15</sup>N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**, 727–734 (1994).
  72. Montelione, G. T. & Wagner, G. 2D Chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *J. Am. Chem. Soc.* **111**, 3096–3098 (1989).
  73. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
  74. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph. Model.* **14**, 51–55 (1996).

## Acknowledgements

This work was supported by National Institutes of Health Grant GM62154 (to P.B. and J.O.) and 5R44GM126676 (to P.B.). The NMR facility is supported by the University of Maryland, the National Institute of Standards and Technology, and a grant from the W. M. Keck Foundation. We also thank Drs. Nese Sari and Louisa Wu for critically reading the manuscript and for many thoughtful comments. Mention of commercial products does not imply recommendation or endorsement by NIST.

## Author contributions

Protein design: Yw.C., B.R., E.C., J.O., P.B.; Performed thermodynamic and binding analyses: B.R., Yw.C., D.M., R.S., P.B.; Performed dynamic light scattering experiments: T.G.; Performed NMR experiments/

structural analysis: Y.H., Yh.C., T.S., T.K., J.O.; Wrote the paper: J.O. (NMR and structural analysis), Yw.C., B.R., P.B. (remaining sections).

## 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/s41467-023-36065-3>.

**Correspondence** and requests for materials should be addressed to John Orban or Philip N. Bryan.

**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 permissions 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) 2023