REPORT

PROTEIN DESIGN

Principles for designing proteins with cavities formed by curved β sheets

Enrique Marcos,[1,2,3]* Benjamin Basanta,[1,2,4]* Tamuka M. Chidyausiku,[1,2,4] Yuefeng Tang,[5,6] Gustav Oberdorfer,[1,2,7] Gaohua Liu,[5,6] G. V. T. Swapna,[5,6] Rongjin Guan,[5,6] Daniel-Adriano Silva,[1,2] Jiayi Dou,[2,4,8] Jose Henrique Pereira,[9,10] Rong Xiao,[5,6] Banumathi Sankaran,[10] Peter H. Zwart,[10] Gaetano T. Montelione,[5,6,11] David Baker[1,2,12]†

Active sites and ligand-binding cavities in native proteins are often formed by curved β sheets, and the ability to control β-sheet curvature would allow design of binding proteins with cavities customized to specific ligands. Toward this end, we investigated the mechanisms controlling β-sheet curvature by studying the geometry of β sheets in naturally occurring protein structures and folding simulations. The principles emerging from this analysis were used to design, de novo, a series of proteins with curved β sheets topped with α helices. Nuclear magnetic resonance and crystal structures of the designs closely match the computational models, showing that β-sheet curvature can be controlled with atomic-level accuracy. Our approach enables the design of proteins with cavities and provides a route to custom design ligand-binding and catalytic sites.

Ligand-binding proteins with curved β sheets surrounding the binding pocket—as in the nuclear transport factor 2 (NTF2)–like, β-barrel, and jelly-roll folds—play key roles in molecular recognition, metabolic pathways, and cell signaling. To date, approaches to designing small molecule–binding proteins and enzymes have started by searching for native protein scaffolds with ligand-binding pockets having roughly the right geometry and then redesigning the surrounding residues to optimize interactions with the small molecule. Although this approach has yielded new binding proteins and catalysts (*1*–*5*), it is not optimal: There may be no naturally occurring scaffold with a pocket with the correct geometry, and introduction of mutations in the design process may change the pocket structure (*6*, *7*). Building de novo proteins with custom-tailored binding sites could be a more effective strategy, but this remains an outstanding challenge (*8*–*11*). De novo protein design has recently focused on proteins with ideal backbone structures (*12*–*16*) [straight helices, uniform β strands, and short loops; see (*17*) for an exception] and optimal core side-chain packing, but the binding pockets of naturally occurring proteins lie on concave surfaces formed by nonideal features, such as kinked helices, curved β sheets, or long loops. The design of proteins with concave surfaces requires examination of how such irregular structural features can be programmed into the amino acid sequence.

We begin by analyzing how classic (*18*, *19*) β bulges (irregularities in the pleating of edge strands) and register shifts (local termination of strand pairing) coupled with intrinsic β-strand geometry induce curvature in antiparallel β sheets (*20*, *21*). We quantify the curvature of an edge strand making an antiparallel pairing with a second strand by the bend angle (Fig. 1A). The absolute value of the bend angle (α) at residue i is the angle between vectors from the α-carbon Cα(i) atom to Cα(i − 2) and Cα(i + 2). The bend angle sign is a function of the relative orientation of a vector $\vec{c}$ describing the concave face of the edge strand (Fig. 1A, left), a vector $\vec{s}_{21}$ between the edge strand and the second strand (Fig. 1A, right), and a vector $\vec{s}_1$ along the edge strand direction (Fig. 1A, right). We analyzed the bend angle of two-stranded antiparallel β sheets in naturally occurring protein structures and in Rosetta folding simulations (Fig. 1B and fig. S1) and found that uniform strands tend to have positive bend angles (because of steric interactions between paired β strands) (fig. S2), whereas strands containing β bulges tend to have negative bend angles (owing to the different hydrogen bond pairing of β bulges) (Fig. 1B and figs. S2 and S3). For β sheets of three strands or more, we found that the type of strand pairing determines the magnitude of β-sheet curvature (Fig. 1C). In uniform three-stranded antiparallel β sheets, the bend directions of the two edge strand segments point in opposite directions, constraining the bend angle of the inner strand to close to zero and flattening the β sheet (Fig. 1D, top). In contrast, in three-stranded β sheets with a β bulge in one of the edge strands, the two edge strand segments bend in the same direction, leading to increased overall bending of the β sheet (Fig. 1D, middle). In uniform β sheets, register shifts enhance bending by terminating pairing between strand segments that would otherwise have opposite bending directions and flatten the β sheet (Fig. 1D, bottom). β-Sheet curvature can hence be programmed by combining β bulges and register shifts. For example, a number of naturally occurring proteins contain a three-stranded β-sheet core with β bulge–derived curvature complemented by additional strands with register shifts propagating the curvature (Fig. 1E).

Using these relationships between β bulges, register shifts, and the direction and magnitude of β-sheet curvature, we designed six protein folds (labeled from A to F, Fig. 2) inspired by the naturally occurring cystatin and NTF2-like superfamilies with a four-stranded antiparallel β sheet, β bulges at the edge strands, and strand lengths ranging between 10 and 14 residues. The width of the β-sheet central base (along the strand direction) is controlled by the relative position between β bulges (folds A, D, and B have central bases of increasing width), whereas the depth (perpendicular to the strand direction) is controlled by the number of strand pairs (folds C, D, E, and F increase the depth of folds A and B by adding on two extra strands) (Fig. 2). The lengths of the two arms flanking the relatively flat center of the base are controlled by the strand lengths and β-bulge positions (folds D, E, and F have arms of increasing length). We complemented the β sheets with one (fold A), three (folds B, C, and D), or four (folds E and F) α helices to form overall cone-shaped structures (as shown in Fig. 2, folds B, C, and D have wide cone bases, whereas fold E partially occludes the cone base with the fourth helix), which provide starting points for designing small molecule–binding sites with entrances at the base of the cone.

We constructed the protein backbones with a stepwise Monte Carlo fragment assembly protocol (*22*) that sequentially adds elements of secondary structure (strands and helices), β bulges, and loops (*23*). Hairpins were designed with two-residue loops following the ββ rule (*12*), which requires that β bulges be at even and odd positions from the following and previous hairpin loops, respectively (owing to the offset in side-chain directionality of β bulges) (fig. S3). We then carried out RosettaDesign calculations (*24*) to favor amino acid identities and side-chain conformations with

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [3]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain. [4]Graduate Program in Biological Physics, Structure, and Design, University of Washington, Seattle, WA 98195, USA. [5]Center for Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA. [6]Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA. [7]Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria. [8]Department of Bioengineering, University of Washington, Seattle, WA 98195, USA. [9]Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, Berkeley, CA 94720, USA. [10]Joint BioEnergy Institute, Emeryville, CA 94608, USA. [11]Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 USA. [12]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. *These authors contributed equally to this work. †Corresponding author. Email: dabaker@u.washington.edu

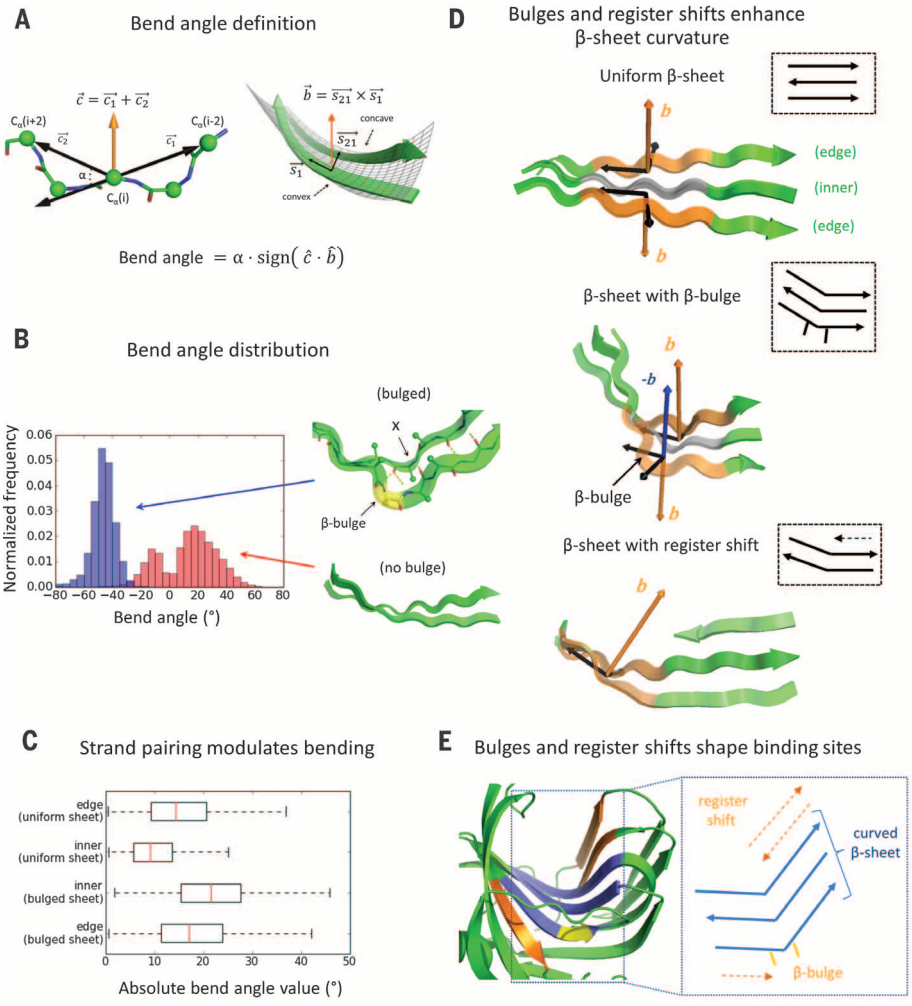

Fig. 1. Rules for β-sheet curvature design. (**A**) Bend angle definition. (**B**) Distribution of bend angles for strand pairs formed by uniform (red) and bulged (blue) strands. The local hydrogen bonding and offset in side-chain directionality at the β-bulge position are shown. The bulge and the residue following donate two backbone hydrogen bonds to the same residue X. (**C**) Bend angle (absolute value) box plots of strands with different pairing types in native three-stranded β sheets. The edge strand distribution in the bulged β-sheet case (bottom) is for the strand that does not contain the bulge. (**D**) Representation of the $\vec{b}$ vector in edge strand pairs for three types of three-stranded β sheets. The $-\vec{b}$ vector is also shown for the sheet with β bulge (middle) to indicate the natural bend direction resulting from a negative bend angle. (**E**) On the left, cartoon representation of the binding site formed by a curved β sheet in a native xylanase (PDB entry 2B45). The curved three-stranded β-sheet core is shown in blue, the β bulge in yellow, and the extra strands in orange. On the right, schematic representation of strand pairings in this curved β sheet.

low energy, tight packing, and high sequence-structure compatibility. We hypothesized that β-bulge positions could be specified at the sequence level solely by changing the normal alternating pattern of polar and hydrophobic amino acids [more complex patterns are observed in native structures (*19*, *25*) (fig. S4)]—in a β bulge, unlike regular strands, two successive residues point in the same direction (Fig. 1B). We relied on side-chain packing to drive strand bending in strands without β bulges (*26*) (fig. S5). Loops were designed with a bias toward sequence profiles obtained from protein fragments with similar backbone torsion angles (*23*). We characterized the folding energy landscape of the designs by Rosetta ab initio structure prediction calculations (*27*, *28*) preceded by a fast initial screen to eliminate designs incapable of folding even with local bias toward the native structure (*23*). We chose for experimental characterization designs with funnel-shaped energy landscapes ranging between 74 and 120 amino acids (table S1) (design names are dcs_X_n, where "dcs" stands for designed curved β sheet, "X" the fold type, and "n" the design number; if disulfide bonds are present, a "_ss" suffix is included). Blast searches (*29*, *30*) indicated that the designed sequences had weak or no similarity to native proteins [E values ranging from 0.00002 (for two of the nine fold D designs) to >10] (table S2); TM-align searches (*31*) identified structures with global fold similarity but little sequence similarity (E values >10, except for the two designs of fold D with low E value, where the top Blast hit was reidentified) and differences in the relative orientation of secondary structure elements and loop connections (fig. S6).

We obtained synthetic genes encoding 37 designs, expressed the proteins in *Escherichia coli*, and purified them by affinity chromatography. Thirty-three of the designs had far-ultraviolet circular dichroism (CD) spectra at 25°C characteristic of α,β proteins and were monomeric by size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) (Fig. 3, figs. S7 to S12, and table S1). Thirty-one of the designs have a melting temperature (T_m) above 95°C, and 24 unfold cooperatively in guanidinium chloride (GdmCl). Two-dimensional ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectra suggest that 12 designs fold into well-ordered structures. Fold E designs, which have a long C-terminal helix as a lid capping the cone base, were the most stable (with T_m > 95°C and denaturation midpoints up to 6 M GdmCl at 25°C) (fig. S11 and table S3). Fold F designs also were thermostable, but their noncooperative unfolding and poor HSQC spectra (fig. S12) suggest imperfect design of the short C-terminal helix interaction with the long arm.

We reasoned that when designing function into these de novo scaffolds, the proximity between the active site and the protein core could compromise protein stability, and we explored two additional stabilization strategies that would preserve pocket accessibility: disulfide bonds and homodimer design. We designed disulfide bonds in positions peripheral to the cone base of folds C and D (*23*). For six of eight designs characterized, disulfide bonds enhanced protein expression, folding cooperativity, and stability by up to 8 kcal mol^{-1} (fig. S13 and table S3). We designed homodimers of fold E designs with shape-complementary low-energy interfaces formed by the convex face of the curved β sheet (*23*). Nine designs with deep global energy minima at the designed dimer configuration in docking calculations were selected for experimental characterization, and three were found to form soluble dimers; the best-expressed design (dcs_E_4_dim9) is 1.4 kcal mol^{-1} more stable than the original monomer (fig. S14).

High-resolution structural analysis is essential for evaluation of the accuracy of computational designs and quickly becomes the bottleneck in protein design studies. During the course of our efforts to solve structures of the designs, we found that the cooperativity of chemical denaturation with GdmCl was a better predictor of rigid core formation, as indicated by nuclear magnetic resonance (NMR) HSQC spectra or the ability to crystallize, than was thermostability (fig. S15)—a number of thermostable designs with low cooperativity in chemical denaturation profiles had molten globulelike side-chain packing. This

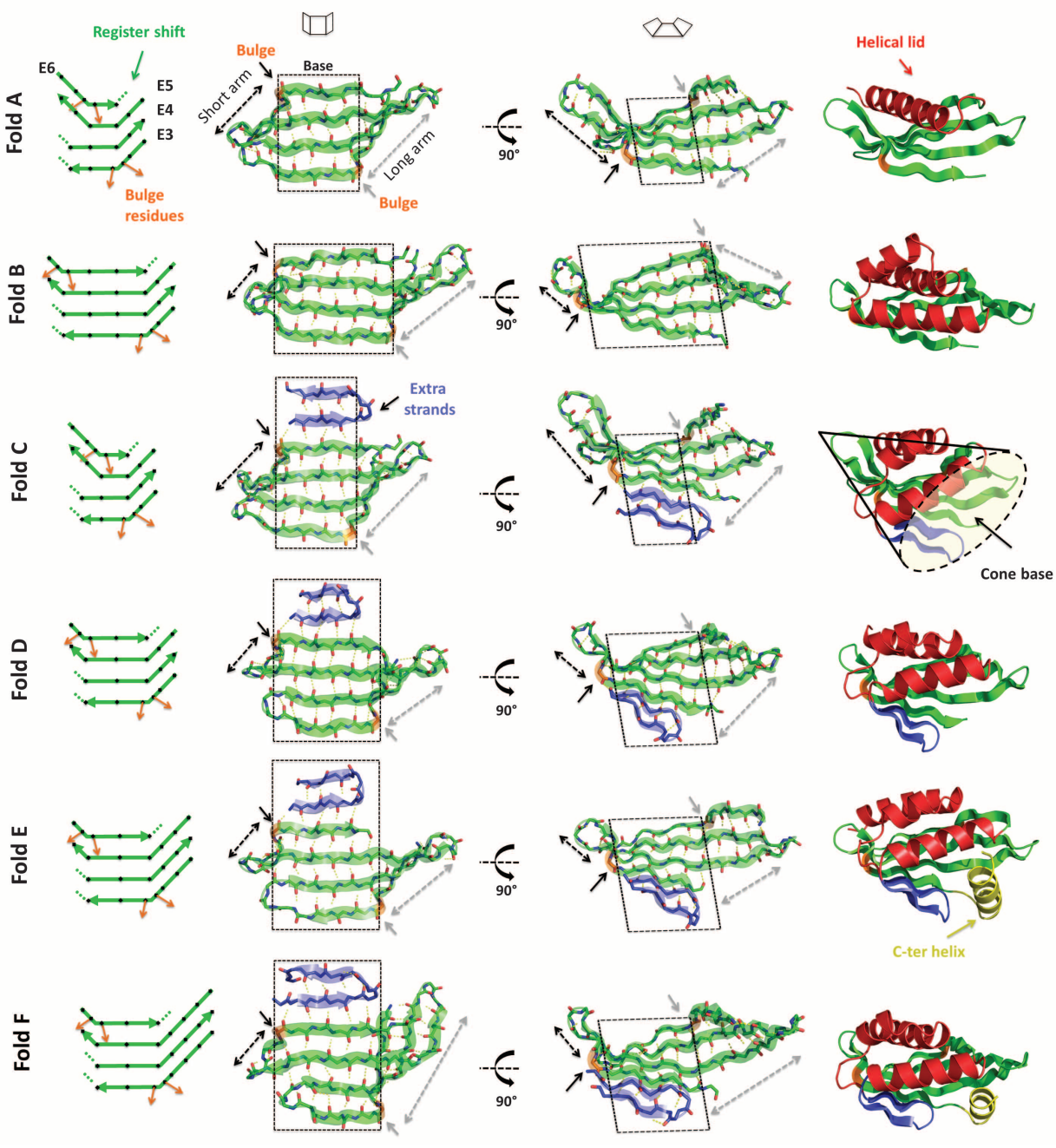

Fig. 2. Designed β sheets and folds. On the left, diagrams of the four-stranded antiparallel β sheets. Black diamonds represent residues with side chains pointing to the convex face of the β sheet, and orange arrows highlight the β-bulge offset in side-chain directionality. Dotted lines show the local termination of strand pairing because of the register shift between paired strands. Second and third columns show two views of the designed β sheets. Black and gray dashed arrows show the length of the short and long arms, respectively, that emerge from the flat central base (highlighted by a black dashed square). On the right, examples of each designed protein fold containing four-stranded antiparallel β sheets (green), helical lids (red), extra strands (blue), and a C-terminal helix capping the pocket entrance (yellow). The concave base of these conical folds is well suited for small molecule–binding site design.

relationship allows the focusing of structure characterization on the designs with the best-defined structures.

We solved the structures of nine designs by NMR spectroscopy or x-ray crystallography. These experimental structures span five different folds (from A to E) (Fig. 4) and are in close agreement with the computational models [Cα root mean square deviations (RMSDs) from 1.0 to 2.1 Å]. The overall β-sheet curvatures were accurately recapitulated, and β-bulge positions were as predicted, in accordance with our hypothesis about local encoding of β bulges. Crystal contacts in the structures of dcs_C_1_ss, dcs_D_2, dcs_E_3, dcs_E_4, and dcs_A_4 support the idea that β bulges minimize edge-to-edge strand pairing (*32*): Hydrogen bond pairing is restricted to the regular strand segments (fig. S16).

The experimental structures for folds A (dcs_A_3 by NMR, Fig. 4A; dcs_A_4 by x-ray crystallography, fig. S17) and B (dcs_B_2 by NMR, Fig. 4B) are in close agreement with the design models in the core of the β sheet and the helices. The designed side-chain packing between the tips of the two β-sheet arms and the helix was better recapitulated in dcs_A_3 and dcs_A_4 than in dcs_B_2 (compare Figs. 4A and 4B, right insets),

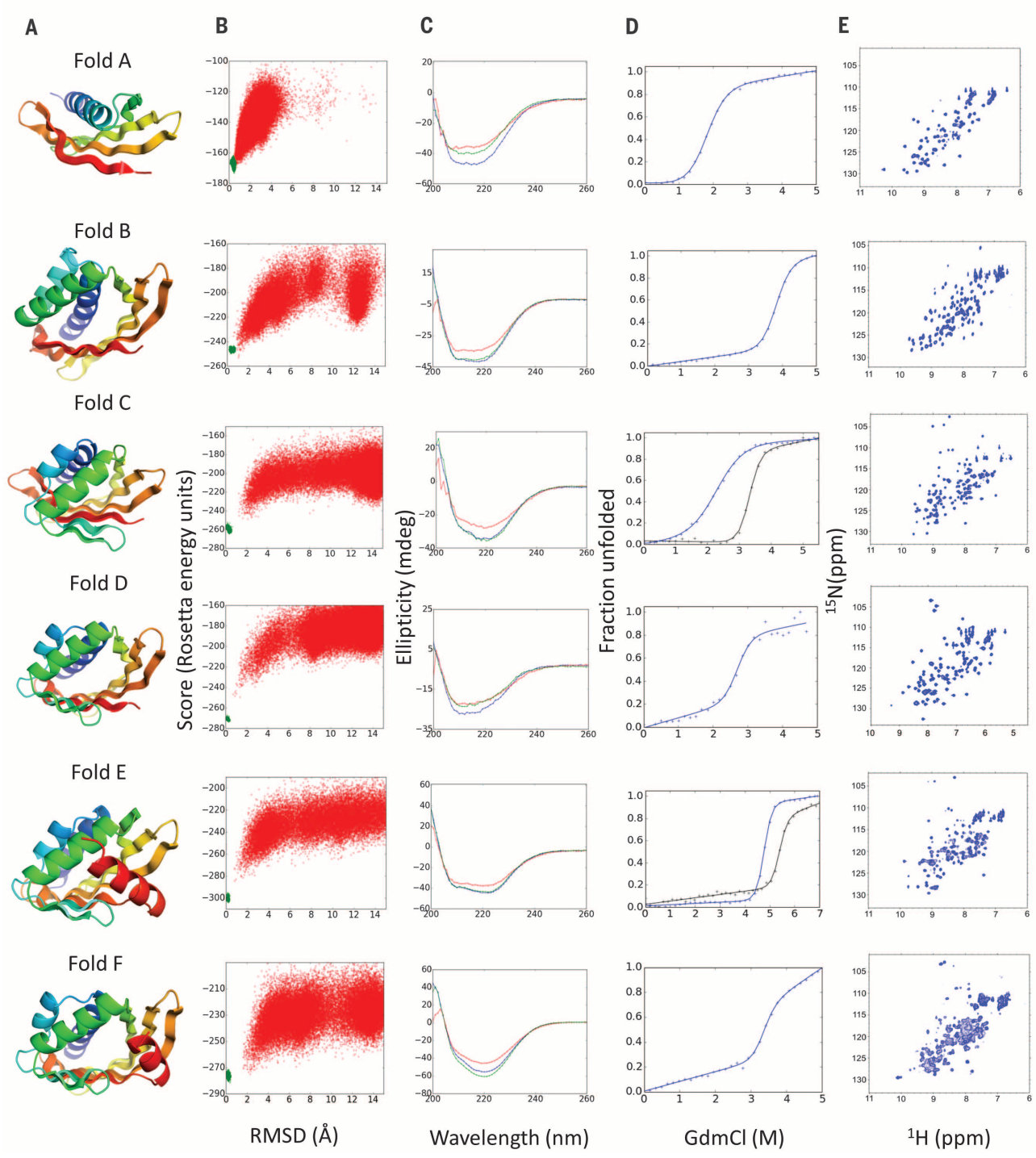

Fig. 3. Experimental characterization of designed proteins for each fold. (**A**) Examples of design models for each fold. (**B**) Folding energy landscapes generated by ab initio structure prediction calculations. Each dot represents the lowest energy structure identified in an independent trajectory starting from an extended chain (red dots) or from the design model (green dots); the *x* axis is the Cα root mean square deviation (RMSD) from the designed model; the *y* axis, the Rosetta all-atom energy. (**C**) Far-ultraviolet circular dichroism spectra (blue, 25°C; red, 95°C; and green, 25°C after cooling). (**D**) Chemical denaturation with GdmCl monitored with circular dichroism at 220 nm and 25°C. For folds C and E, the denaturation curves for designs stabilized by a disulfide bond or a dimer interface are shown in black lines. (**E**) ^{1}H-^{15}N HSQC spectra obtained at 25°C.

where the long arm is more twisted in the NMR structure than in the design model; full control over β-sheet geometry in these folds likely requires control over side-chain packing between the β sheet and the helical lid.

The crystal structures of folds C and D (Fig. 4, C and D) are very close to the design models, with designed aromatic-packing and hydrogen-bonding interactions bridging the protein core and the cone base; a designed disulfide bond is also correctly recapitulated (Fig. 4C, bottom inset). The two crystal structures for fold E monomeric designs also closely match the design models (Fig. 4, E and F), with the cone base capped by the C-terminal helix in two different orientations. A

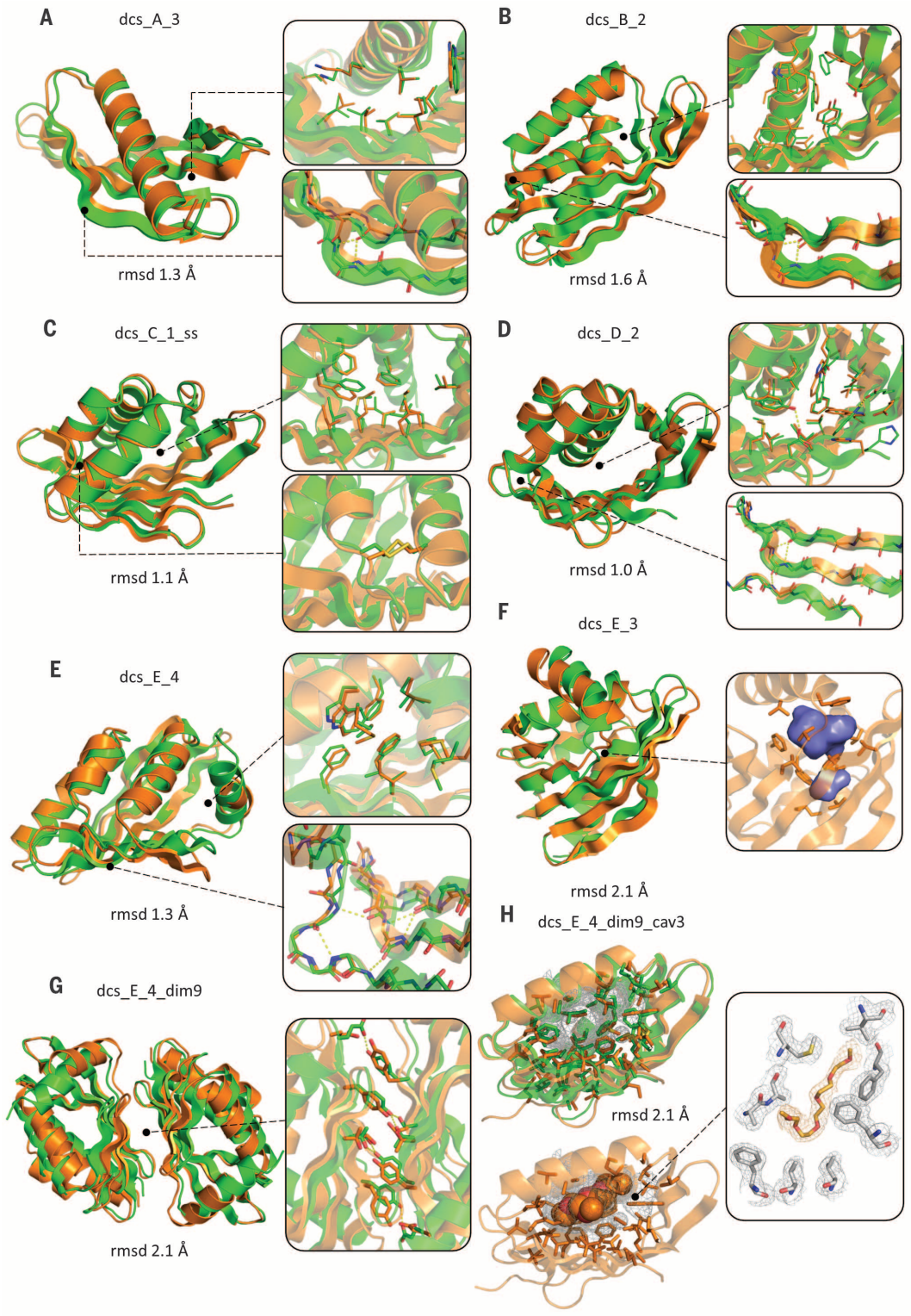

Fig. 4. Experimentally determined structures of designed proteins. In each panel, the experimental structure and the design model are superimposed and colored in orange and green, respectively. Insets show comparisons of side-chain rotamers, β-bulge geometry, and cavities, and designed side-chain and β-bulge hydrogen bonds are shown in yellow dashed lines. The RMSD calculated over all Cα atoms is shown in each panel. (**A**) dcs_A_3 and (**B**) dcs_B_2 were solved by NMR (comparisons utilized the lowest-energy NMR model). (**C**) dcs_C_1_ss (3.0 Å resolution) with designed disulfide bond in inset. (**D**) dcs_D_2 (2.0 Å resolution). (**E**) dcs_E_4 (2.9 Å resolution). (**F**) dcs_E_3 (3.1 Å resolution); an internal hydrophobic cavity forms in both the design and the crystal structure (volume, 192 $Å^3$). (**G**) dcs_E_4_dim9 (2.4 Å resolution); the interface aromatic-stacking and hydrogen-bonding interactions are very similar in the crystal structure and design model (right inset). (**H**) dcs_E_4_dim9_cav3 (1.8 Å resolution). A large (520 $Å^3$) cavity is filled with a pentaethylene glycol molecule in the crystal structure (bottom left; electron density map is on right and superimposition with design model, top left). The C-terminal helix and the dimer interface are not shown for better visualization of the cavity.

buried cavity designed in one of these (dcs_E_3) expands toward the cone base in the crystal structure (Fig. 4F and fig. S21). We explored the ability of the fold C and D designs to support cavities by reducing the size and increasing the polarity of side chains at the cone base (fig. S18). Five of the nine redesigns tested (with up to 19 mutations) were soluble and monomeric (figs. S19 and S20 and table S1).

The crystal structure of the designed homodimer dcs_E_4_dim9 closely matches the computational model over both the individual subunits and the designed β-sheet interface (Fig. 4G and fig. S22). We designed large cavities by truncating side chains at the cone base (figs. S18 and S20). The crystal structure of one such design revealed a large (520 $Å^3$) cavity very similar to that in the design model, lined by the curved β sheet (Fig. 4H) (a pentaethylene glycol fills the cavity). This is the largest de novo designed cavity to date and illustrates how large core-packing vacancies can be programmed by designing curved β sheets topped by helices.

The nine NMR and crystal structures show that β-sheet curvature can be accurately programmed with the principles that we have identified. The designed proteins exhibit a rich combination of structural features: curved β sheets with β bulges and register shifts, loops of variable length, helices, disulfide bonds, β-sheet interfaces, and cavities. Stability can be increased by incorporating disulfide bonds and homodimer interfaces without interfering with the accessibility of potential binding pockets, and this allows substitution of large hydrophobics by smaller or polar residues to line solvent-exposed clefts and buried cavities.

Computational methods have been used to design enzyme catalysts by defining an ideal active site ("theozyme") and then searching for placements of the theozyme in native protein scaffolds. This approach has yielded catalysts for a number of chemical reactions, including reactions not catalyzed by naturally occurring enzymes, but the activities are quite low. This likely results from two shortcomings in the design strategy: The detailed theozyme geometry cannot be perfectly realized in any preexisting scaffold, and the sequence changes introduced in the design process can produce unpredictable changes in structure (*6*, *7*). Our de novo design framework should make it possible to overcome both limitations by custom designing backbones for the reaction of interest.

REFERENCES AND NOTES

1. C. E. Tinberg *et al.*, *Nature* **501**, 212–216 (2013).
2. D. Röthlisberger *et al.*, *Nature* **453**, 190–195 (2008).
3. L. Jiang *et al.*, *Science* **319**, 1387–1391 (2008).
4. J. B. Siegel *et al.*, *Science* **329**, 309–313 (2010).
5. S. Rajagopalan *et al.*, *Nat. Chem. Biol.* **10**, 386–391 (2014).
6. F. Richter *et al.*, *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
7. L. Giger *et al.*, *Nat. Chem. Biol.* **9**, 494–498 (2013).
8. N. H. Joh *et al.*, *Science* **346**, 1520–1524 (2014).
9. A. R. Thomson *et al.*, *Science* **346**, 485–488 (2014).
10. L. Doyle *et al.*, *Nature* **528**, 585–588 (2015).
11. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, *Nat. Chem.* **8**, 837–844 (2016).
12. N. Koga *et al.*, *Nature* **491**, 222–227 (2012).
13. P.-S. Huang *et al.*, *Science* **346**, 481–485 (2014).
14. Y. R. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5478–E5485 (2015).
15. T. J. Brunette *et al.*, *Nature* **528**, 580–584 (2015).
16. P.-S. Huang *et al.*, *Nat. Chem. Biol.* **12**, 29–34 (2016).
17. T. M. Jacobs *et al.*, *Science* **352**, 687–690 (2016).
18. J. S. Richardson, E. D. Getzoff, D. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2574–2578 (1978).
19. A. W. Chan, E. G. Hutchinson, D. Harris, J. M. Thornton, *Protein Sci.* **2**, 1574–1590 (1993).
20. C. Chothia, *J. Mol. Biol.* **163**, 107–117 (1983).
21. F. R. Salemme, *Prog. Biophys. Mol. Biol.* **42**, 95–133 (1983).
22. A. Leaver-Fay *et al.*, *Methods Enzymol.* **487**, 545–574 (2011).
23. Materials and methods are available as supplementary materials at the *Science* website.
24. B. Kuhlman *et al.*, *Science* **302**, 1364–1368 (2003).
25. P. Craveur, A. P. Joseph, J. Rebehmed, A. G. de Brevern, *Protein Sci.* **22**, 1366–1378 (2013).
26. K. Fujiwara, S. Ebisawa, Y. Watanabe, H. Fujiwara, M. Ikeguchi, *BMC Struct. Biol.* **15**, 21 (2015).
27. C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, *Methods Enzymol.* **383**, 66–93 (2004).
28. P. Bradley, K. M. S. Misura, D. Baker, *Science* **309**, 1868–1871 (2005).
29. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389–3402 (1997).
30. C. Camacho *et al.*, *BMC Bioinformatics* **10**, 421 (2009).
31. Y. Zhang, J. Skolnick, *Nucleic Acids Res.* **33**, 2302–2309 (2005).
32. J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2754–2759 (2002).

ACKNOWLEDGMENTS

We thank L. Carter for assistance with SEC-MALS and protein production; J. Nguyen, A. Young-Seug, Z. Wang, M. Bick, S. Jayaraman, and P. O'Connell for assistance in x-ray crystallography; all Baker lab members for discussions; and Rosetta@Home volunteers for computing resources used in ab initio structure prediction calculations. Work carried out at the Baker laboratory was supported by the Howard Hughes Medical Institute and the Defense Threat Reduction Agency (funding HDTRA 1-11-1-0041). X-ray diffraction data were collected at the National Synchrotron Light Source with beamline X4C [Brookhaven National Laboratory, Upton, NY, U.S. Department of Energy (DOE)] and the Advance Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA, DOE). The Berkeley Center for Structural Biology is supported in part by the NIH; National Institute of General Medical Sciences (NIGMS), NIH; and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the DOE under contract no. DE-AC02-05CH11231. E.M. was supported by a Marie Curie International Outgoing Fellowship (FP7-PEOPLE-2011-IOF 298976). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from the Ministry of Economy, Industry, and Competitiveness (Government of Spain). G.O. was supported by a Marie Curie International Outgoing Fellowship (332094 ASR-CompEnzDes FP7-PEOPLE-2012-IOF). D.-A.S. is a Latin American PEW postdoctoral fellow and Mexican National Council of Science and Technology (CONACYT) postdoctoral fellow and acknowledges their support. This work was supported as a Community Outreach Activity of NIGMS Protein Structure Initiative grant U54 GM094597 (to G.T.M). Coordinates and structure factors have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank with the accession codes 5KPH (dcs_A_3), 4R80 (dcs_A_4), 5KPE (dcs_B_2), 5TS4 (dcs_C_1_ss), 5L33 (dcs_D_2), 5TPJ (dcs_E_3), 5TRV (dcs_E_4), 5TPH (dcs_E_4_dim9), and 5U35 (dcs_E_4_dim9_cav3). NMR data have been deposited in the Biological Magnetic Resonance Data Bank with the accession codes 30139 (dcs_A_3) and 30128 (dcs_B_2). E.M. and D.B. designed the research; E.M. developed the β-sheet design principles and set up the design method; E.M., B.B., and T.M.C. carried out the design calculations, protein expression, and biophysical characterization; E.M., B.B., and T.M.C. crystallized proteins from folds C, D, and E; J.H.P. crystallized dcs_C_1_ss; Y.T. solved the NMR structures of dcs_A_3 and dcs_B_2 with help from G.L.; R.G. solved the crystal structure of dcs_A_4; G.O. and B.B. solved the crystal structures of dcs_C_1_ss, dcs_D_2, dcs_E_3, dcs_E_4, dcs_E_4_dim9, and dcs_E_4_dim9_cav3; G.V.T.S. collected HSQC data; R.X. prepared isotope-enriched protein samples for NMR structure determination; D.-A.S. and E.M. developed the biased ab initio folding protocol; J.D. designed and experimentally characterized cavity-creating mutations in dcs_D_2; B.S. and P.H.Z. collected and analyzed crystallographic data; and E.M., B.B., G.T.M., and D.B. wrote the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6321/201/suppl/DC1
Materials and Methods
Figs. S1 to S22
Tables S1 to S7
References (*33*–*65*)
Data files

5 August 2016; accepted 2 December 2016
10.1126/science.aah7389