PNAS
www.pnas.org

Supplementary Information for
Protein sequence design by conformational landscape optimization.

Christoffer Norn, Basile I. M. Wicky, David Juergens, Sirui Liu, David Kim, Doug Tischer, Brian Koepnick, Ivan Anishchenko, Foldit Players, David Baker, Sergey Ovchinnikov

Corresponding Authors: David Baker, Sergey Ovchinnikov
Emails: dabaker@uw.edu, so@fas.harvard.edu

This PDF file includes:

Supplementary Figures

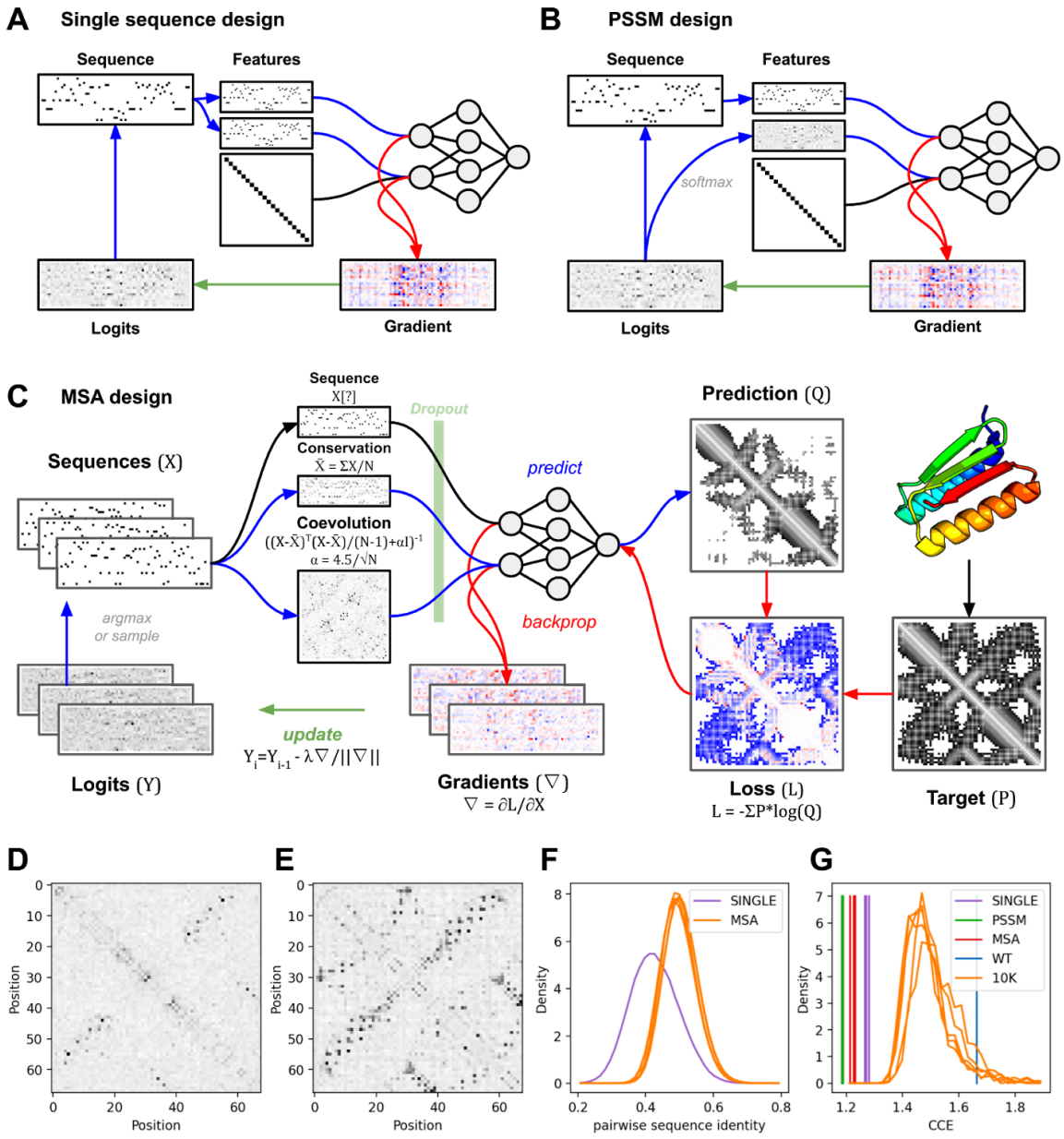

Fig. S1. The inputs to trRosetta include the tiled one-hot encoded query sequence, conservation and coevolution matrix. For single sequence design (**A**) the conservation channel is substituted by the sequence, and the coevolution channel by an identity matrix. For PSSM design (**B**), the conservation channel is the softmax of the logits. For MSA design (**C**), conservation is the mean of all the sequences, and coevolution is the inverse covariance of the sequences. To prevent overfitting on the first sequence or any specific features, we only backpropagate through the conservation and coevolution channels, randomly selecting a different sequence for the sequence channel, and included a dropout layer (80%) after tiling. (**D**) We often find that only a subset of contacts coevolve for independently generated sequences, while for MSA design (**E**), all contacts have a covariance signal. (**F**) The diversity of designed sequences (pairwise sequence identity) and (**G**) cross-entropy scores for: Single sequence design, MSA design, PSSM design and [W]ild[T]ype sequence for a given backbone (PDB:6MRR). The orange distributions are the CCE scores ($-logP(structure|sequence)$) of the 10,000 MSA designed sequences when scored using the single-sequence predictor from (**A**).

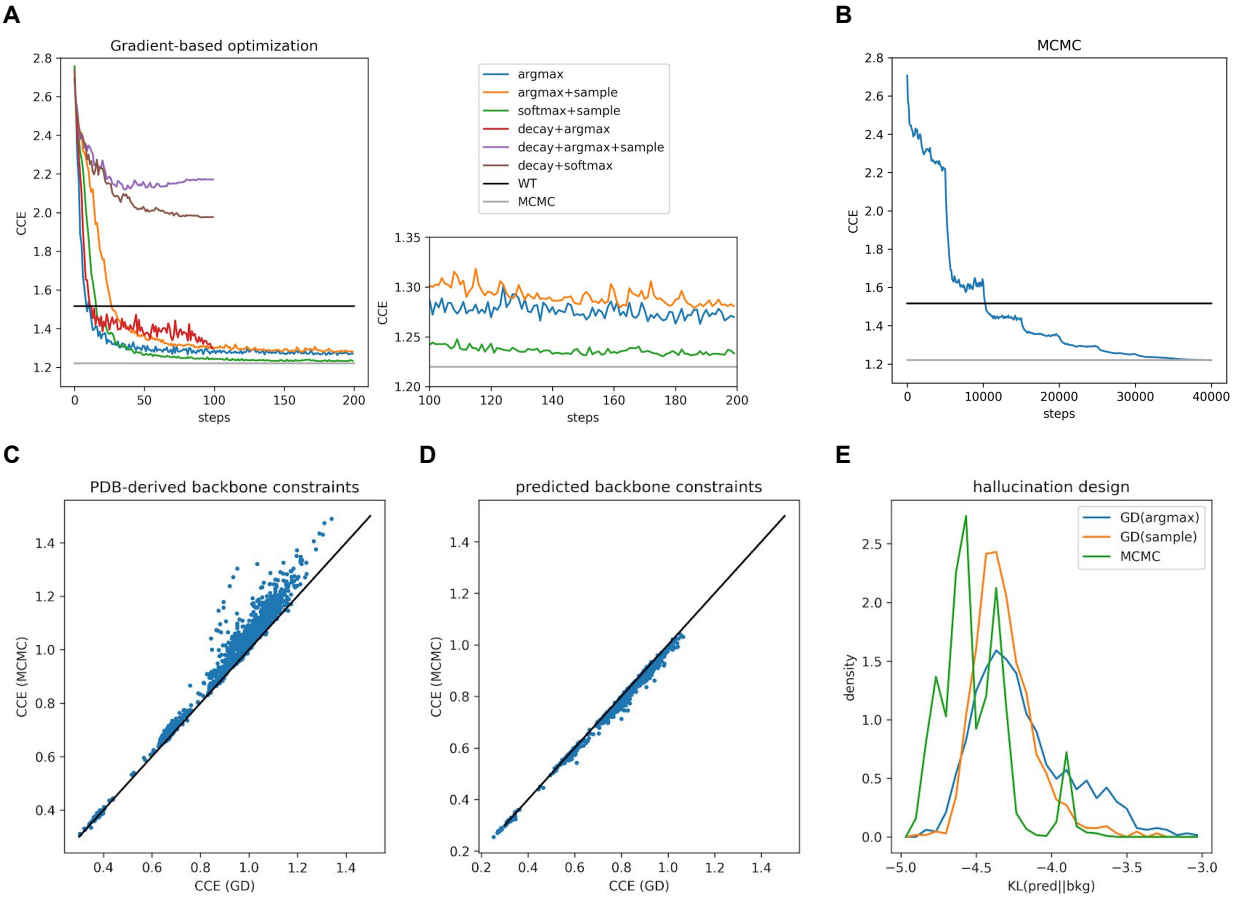

Fig. S2. Comparing gradient descent (GD) based optimization to Markov chain Monte Carlo (MCMC). (**A-B**) Optimization trajectories for Top7 (PDB:1QYS) fixed backbone design. (**A**) For gradient-based optimization, we tried six different optimization schemes. We find that a constant learning rate ($\sqrt{length}$) requires a smaller number of steps to converge compared to decay-based minimization (as described in the main text). During optimization, we also experimented with adding noise from a softmax-gumbel distribution (a.k.a. sampling). For the reported CCE ($-logP(structure|sequence)$), each sequence along the trajectory was re-scored with sampling disabled and argmax enabled. The black line indicates the CCE of the wildtype sequence, and the grayline the average CCE at the end of the MCMC trajectories for the same backbone. (**B**) Average of twenty independent MCMC runs. (**C-E**) We compare GD to MCMC on a larger set of proteins designed with the MCMC protocol (15). For (**C**) we designed a new sequence to match constraints derived from the Rosetta-relaxed backbone. For all cases, the GD method is able to find a sequence that is similar or lower (in terms of CCE loss) than what was hallucinated by trRosetta using the MCMC approach. (**D**) If we instead try to design a sequence to match the predicted backbone constraints of the hallucinated design, GD is able to find an alternative sequence to match these constraints with similar or slightly worse CCE. (**E**) For newly hallucinated designs, the MCMC protocol is able to sample sequences with lower KL divergences. Interestingly, GD with sampling generates sequences that have lower KL divergences more frequently than when using argmax discretization.

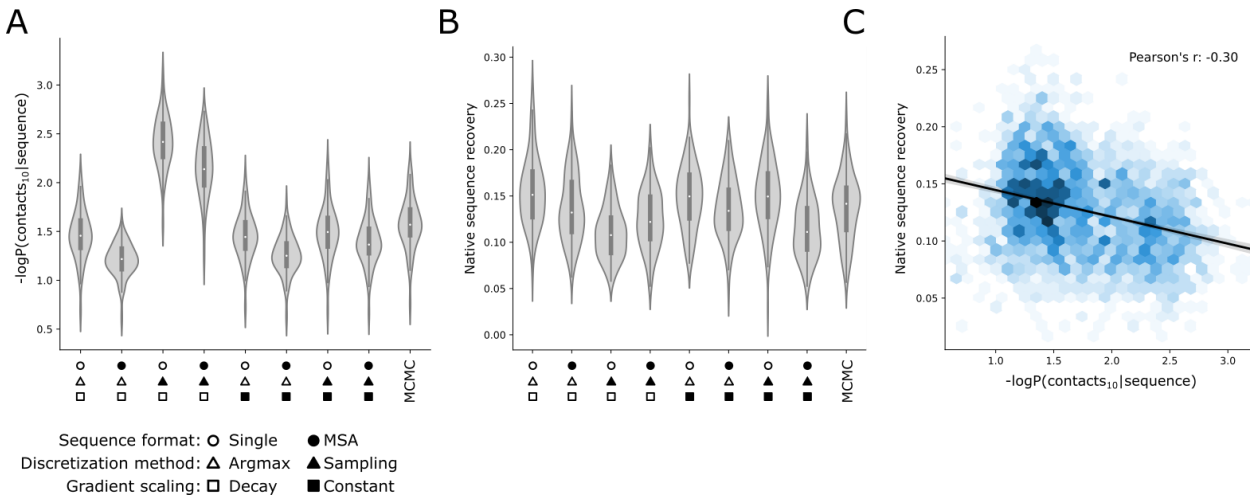

Fig. S3 Comparison between optimization methods for fixed backbone design on a larger dataset. TrRosetta designed sequences for a subset of the native proteins described in the Supplementary Methods with different optimization settings. Because of GPU memory limitations, we only designed the 103 native proteins that were less than or equal to 139 amino acids. (**A**) We computed $-logP(contacts|sequence)$ as a function of different optimization settings using a distance cutoff of 10 Å. We performed gradient decay with an exponent of 2. When the gradient scaling was kept constant, we used a learning rate of 1.0. (**B**) Native sequence recovery with the same optimization settings. (**C**) The $-logP(contacts|sequence)$ values and native sequence recovery are inversely correlated, meaning the closer a design's predicted probability distributions are to the native's, the better the native sequence recovery. Data are for the designs of all 103 native proteins using all the optimization settings reported in Table S1.

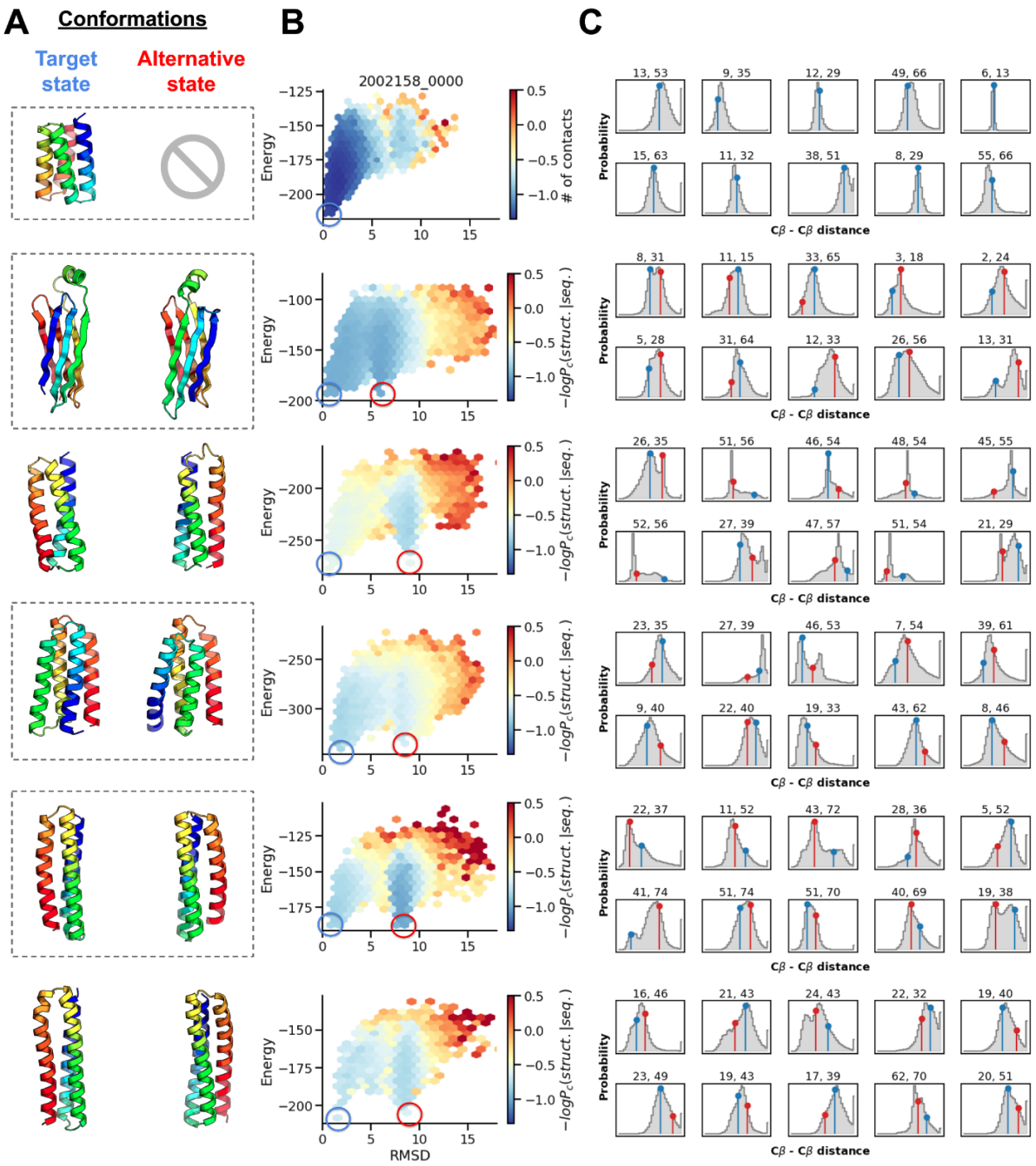

Fig. S4 Examples of distograms from trRosetta. (**A**) Structures (target state and alternative state) of Foldit-player designs as determined by *ab initio* folding (boxed structures are the same as in Fig. 2)) (**B**) *Ab initio* folding funnels for the designs are shown on the left. The positions of the designed and alternative structures are shown by blue and red circles respectively. (**C**) Examples of outputs from trRosetta (*Cβ-Cβ* distance predictions) for specific *i,j* pairs (indicated above each plot). The actual distances observed in the target state and the alternative state are shown in blue and red respectively. These examples were randomly selected from a set of *i,j* pairs filtered as follows: *i*) predicted to be in contact by trRosetta (*Cβ-Cβ* distance < 20 Å), *ii*) differing by at least 2.5 Å between target and alternative state, and *iii*) having significant probabilities for both states (ratio of probability < 5). An analysis of bimodality for these structures is shown in Fig. S5.

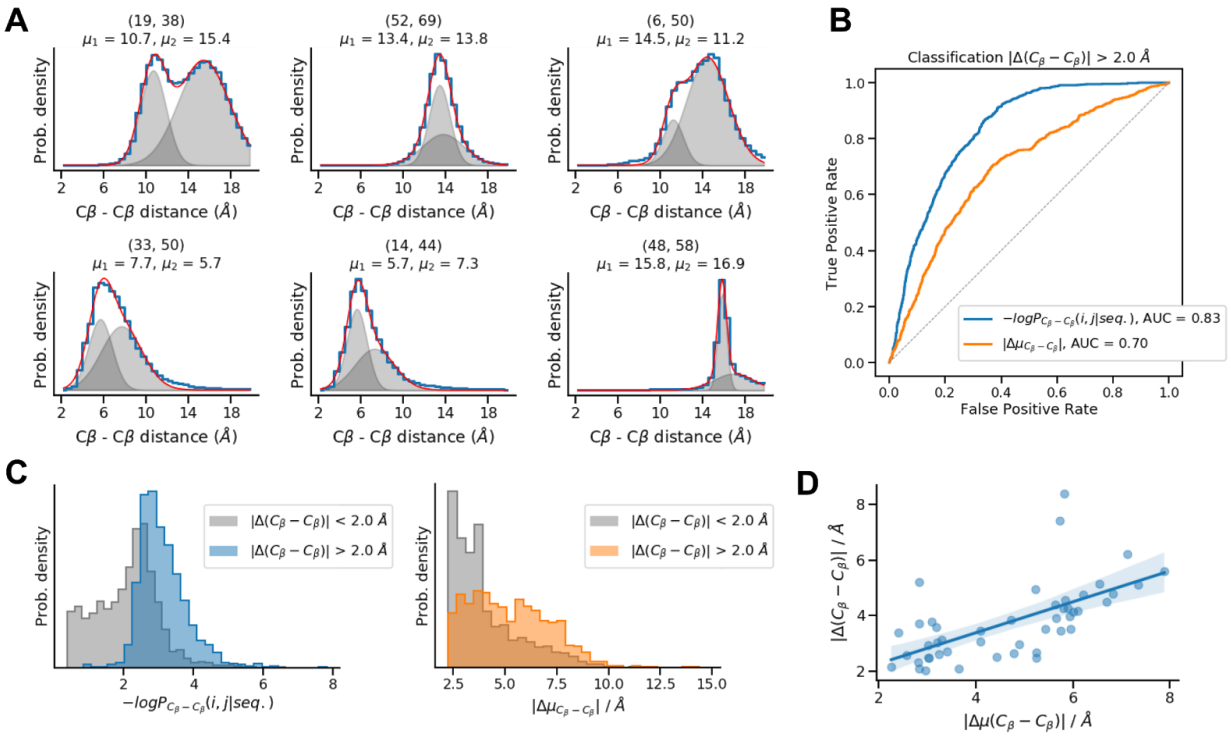

Fig. S5 Bimodality in trRosetta *Cβ-Cβ* predictions is indicative of an alternative conformation. Each individual *i,j* *Cβ-Cβ* prediction was analyzed for the bimodality of its distribution, and the results used to predict the presence of an alternative conformation at this position. (**A**) Examples of individual *i,j* (indicated above each plot) *Cβ-Cβ* predictions. The probability density outputted by trRosetta is shown in blue. The distributions were fitted to a double Gaussian function (shown in red, with the individual Gaussian functions shown in grey). The mean of each Gaussian function (μ) is reported above the plot. (**B**) Receiver operating characteristic (ROC) curves for the classification of *i,j* pairs where the absolute difference in *Cβ-Cβ* distance between the designed and alternative state (|ΔCβ-Cβ|) is greater than 2.0 Å. The reduced probability for the *i,j* distance observed in the designed structure ($-logP_{C\beta-C\beta}(i,j \mid seq.)$) is the best predictor for the presence of an alternative state (AUC = 0.81), consistent with the dilution of probability observed for the full structures (Fig. 2 and Fig. S4). The absolute difference between the means of the individual Gaussian functions ($|\Delta\mu_{C\beta-C\beta}|$) is also capable of predicting some of the *i,j* pairs with alternative conformations (AUCs of 0.7). This suggests that bimodal predictions for individual *i,j* pairs can be indicative of an alternative structure. (**C**) Distributions for the two metrics used in (**B**), split at the |ΔCβ-Cβ| cutoff value used for computing the ROC curves (2.0 Å). The dataset for this analysis was composed of *i,j* pairs from the five structures with alternative minima shown in Fig. S4. Only pairs that were predicted to be in contact and separated by at least seven residues in primary sequence were considered for the analysis: 3651 distributions (18.9% of total), with 31% having |ΔCβ-Cβ| > 2.0 Å. (**D**) For *i,j* pairs where trRosetta is confident about the bimodality of the distribution, the absolute difference in the fitted means ($|\Delta\mu_{C\beta-C\beta}|$) correlates with the absolute difference in *Cβ-Cβ* distance between the designed and alternative state (|ΔCβ-Cβ|): R^2 = 0.40, p-value = 2e-6 (N = 47). Confidence in the bimodality of the distribution was defined as having a ratio between the individual Gaussian integrals below 5 (*i.e.* two peaks with appreciable probability each), and sharp distributions (*i.e.* standard deviations below 3 Å).

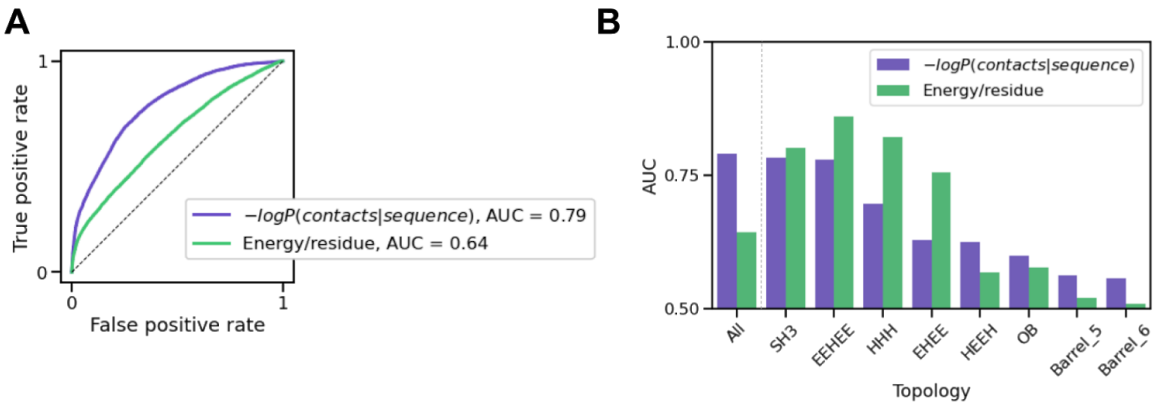

Fig. S6. TrRosetta recognises protein stability across topologies, but Rosetta is often better at classifying intra-topology differences**.** TrRosetta ($-logP(contacts|sequence)$, purple) and Rosetta (energy/residue, green) scores were used to predict protease resistance (a proxy for folding stability, threshold = 0.5, 22% of the dataset) for 30,159 designs spanning 8 topologies. (**A**) Receiver operating characteristic (ROC) curves showing the classification of protease stability data across all topologies. Area under the curve (AUC) values are indicated in the legend. (**B**) Comparison of AUC values for tRosetta and Rosetta classifications of stabilities across all designs (left of the dashed line, same as A), and *within* topologies (right of the dashed line).

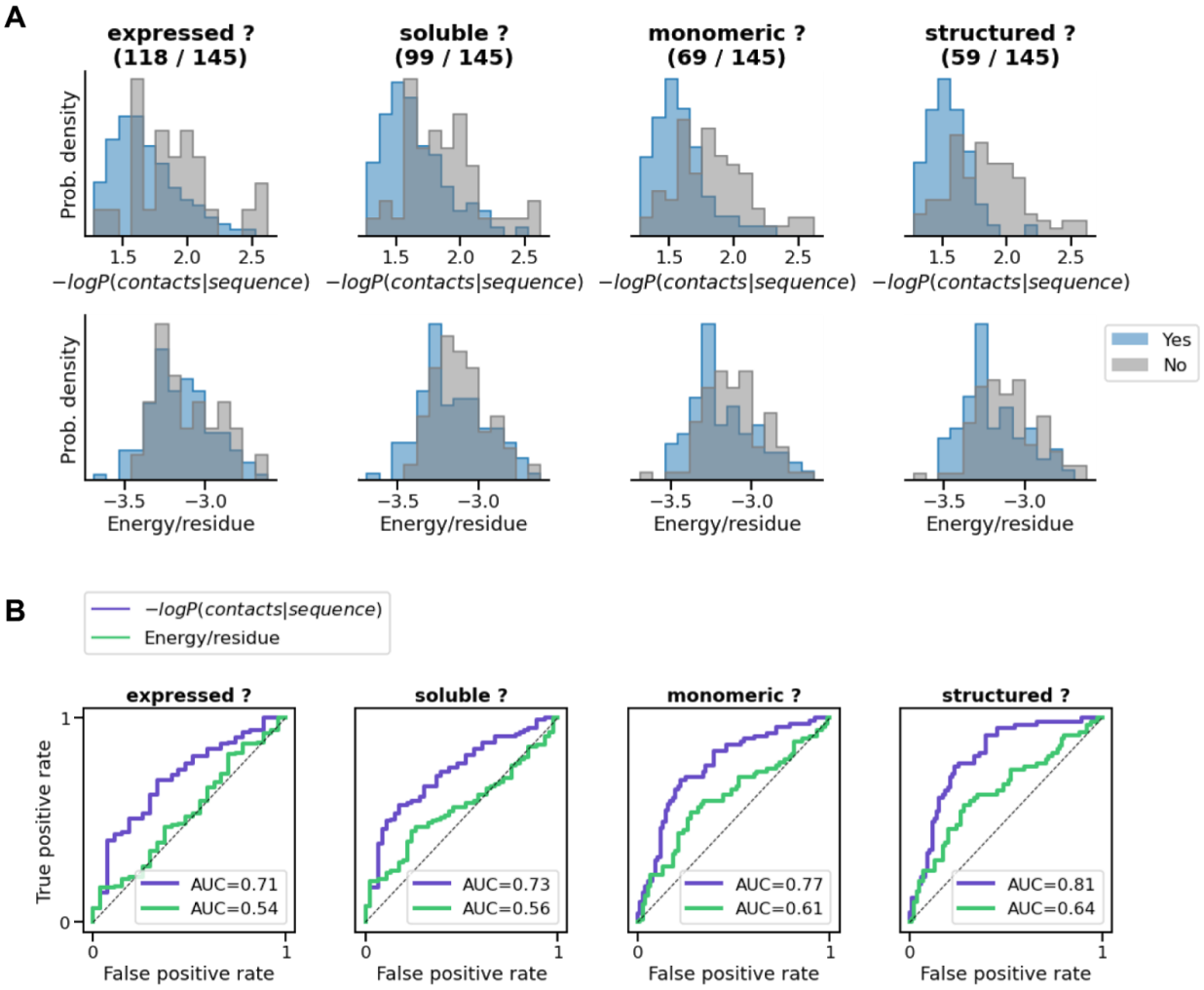

Fig S7. TrRosetta score is better at classifying experimental success than Rosetta. For 145 Foldit-player designs that were experimentally characterized (16), success was assessed at different stages: protein expression, solubility, oligomeric state, and presence of secondary structure after purification. (**A**) Histograms showing the distributions of success or failure at each experimental stage as assessed with trRosetta (top) and Rosetta (bottom). (**B**) The same data represented as ROC curves. AUC values for trRosetta (purple), and Rosetta (green) are shown on the plots.

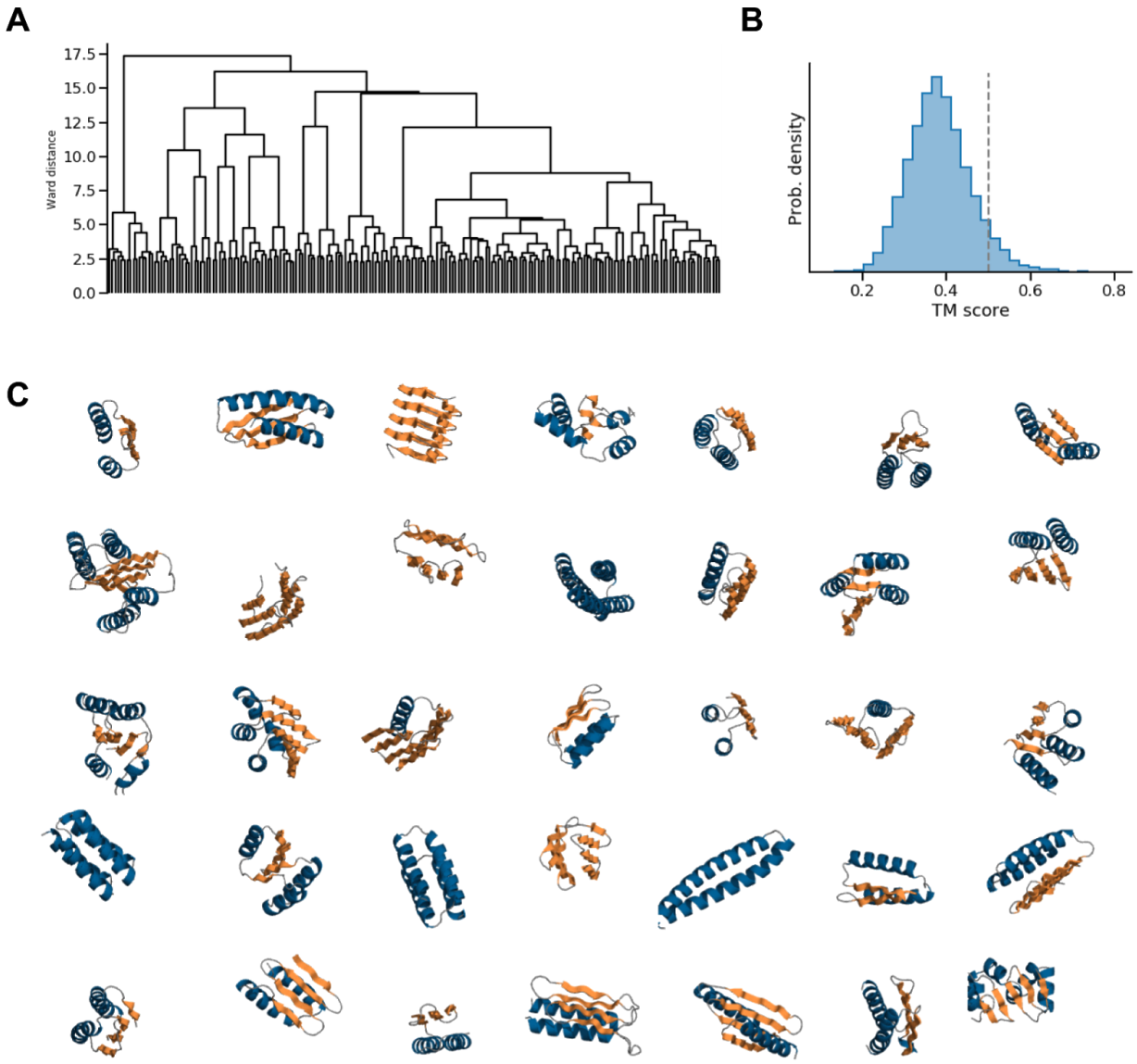

Fig S8. Structural clustering of Foldit players designs. (**A**) 5288 designs were clustered into 200 bins based on their pairwise TM-align scores (34). One design per bin was used for re-design. (**B**) Distribution of pairwise TM-align scores for the 200 selected designs (mean = 0.38) shows that they are structurally diverse. The threshold for homology (0.5) is indicated by a dashed line. (**C**) Examples of structures from the 200 selected designs, spanning all-α, all-β, and mixed α/β folds. Helices, sheets, and loops are shown in blue, orange and grey respectively.

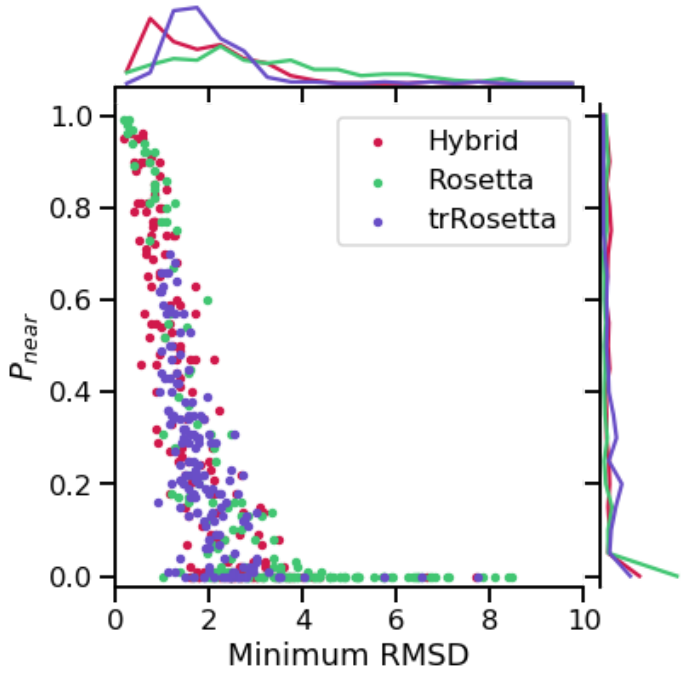

Fig S9. TrRosetta improves sampling of conformations near the target conformation. The energy landscape quality (*Pnear*) and RMSD of the lowest RMSD decoy were computed for the folding energy landscapes of 165 backbones designed with trRosetta, Rosetta or the Hybrid (trRosetta PSSM + Rosetta) method. In contrast to trRosetta-based methods (trRosetta and Hybrid), decoys from *ab initio* folding of Rosetta-designed sequences rarely sampled the target backbone conformations.

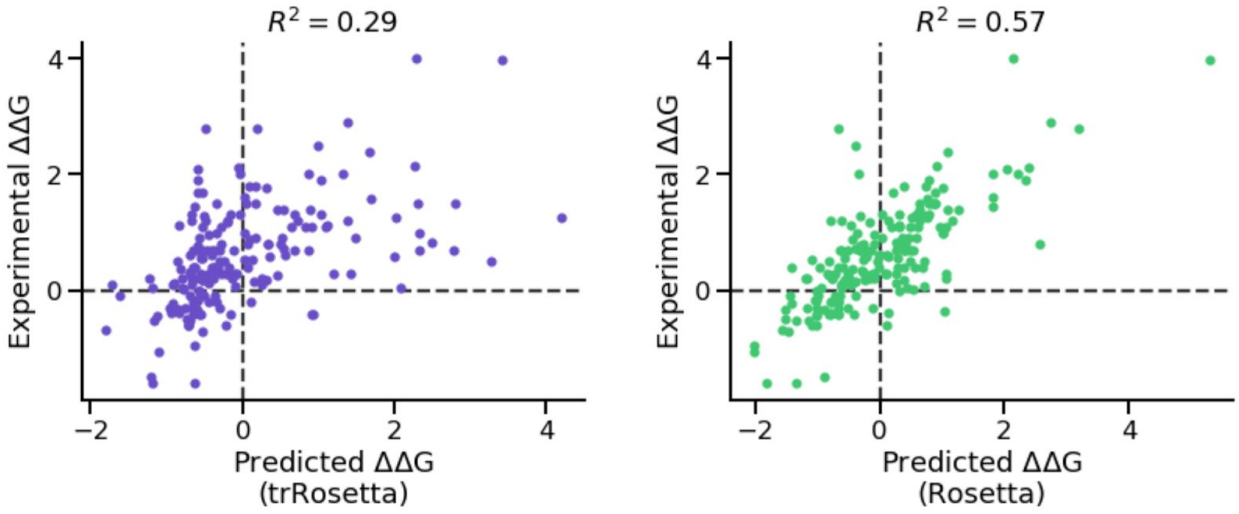

Fig S10. The Rosetta all-atom energy function (right) is better at predicting the thermodynamics of single point mutations than trRosetta (left). Based on a curated set of single point mutations from (35), we sub-selected the 15% best-scored sites with respect to $-logP\,(mutation\ environment\ contacts|wildtype\ sequence)$, where *mutation environment contacts* is defined as the subset of *i,j* pairs encompassing the contacts between the mutated position and the positions within 10 Å of it (*Cβ-Cβ* distances). The effect of mutations was computed as the difference $logP\,(mutation\ environment\ contacts|mutated\ sequence) - logP\,(mutation\ environment\ contacts|wildtype\ sequence)$. Predicted Rosetta ΔΔGs were taken from (36).

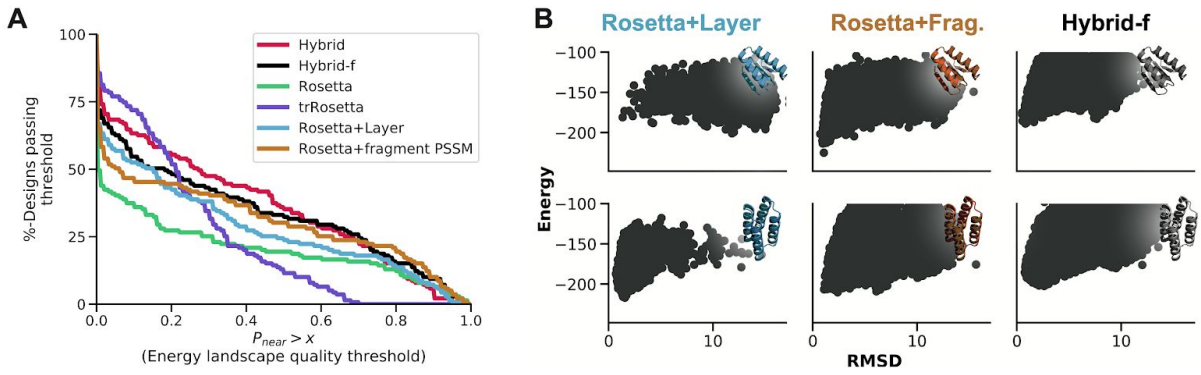

Fig. S11. (A) Performance curves for other state-of-the-art methods that address the lack of energy landscape-awareness of Rosetta. Compared to Rosetta design (all amino acids possible at all positions), LayerDesign restricts amino acid choices by burial, which improves the outcome, but still underperforms compared to the Hybrid method. The fragment PSSM method guides sequence design by generating sequence profiles using fragments from known structures that match the target, and biases the energy function during design using this PSSM. The Hybrid method performs slightly better on average (although not significantly) than the fragment PSSM method. However, we note that the fragment PSSM method has an inherent advantage in this benchmark, as both fragment-based sequence profile generation and *ab initio* folding collect fragments from the same database. As such, the resulting sequences are more likely to encode fragments needed for reaching the target structure in *ab initio* folding calculations. The Hybrid method was not significantly different from Hybrid-f, but was significantly better than all other methods (paired t-test, p-value < 0.05). The Hybrid-f method was not significantly different from either the Hybrid or Rosetta+fragment PSSM method, but was significantly better than all other methods (paired t-test, p-value < 0.05). (**B**) Examples of energy landscapes for two Foldit-player generated backbones, each designed with Hybrid-f, Rosetta with LayerDesign, and Rosetta with fragment-based PSSM bias. The energy landscapes for the same Foldit-player backbones designed with Rosetta, Hybrid and trRosetta can be found in Fig. 3C.

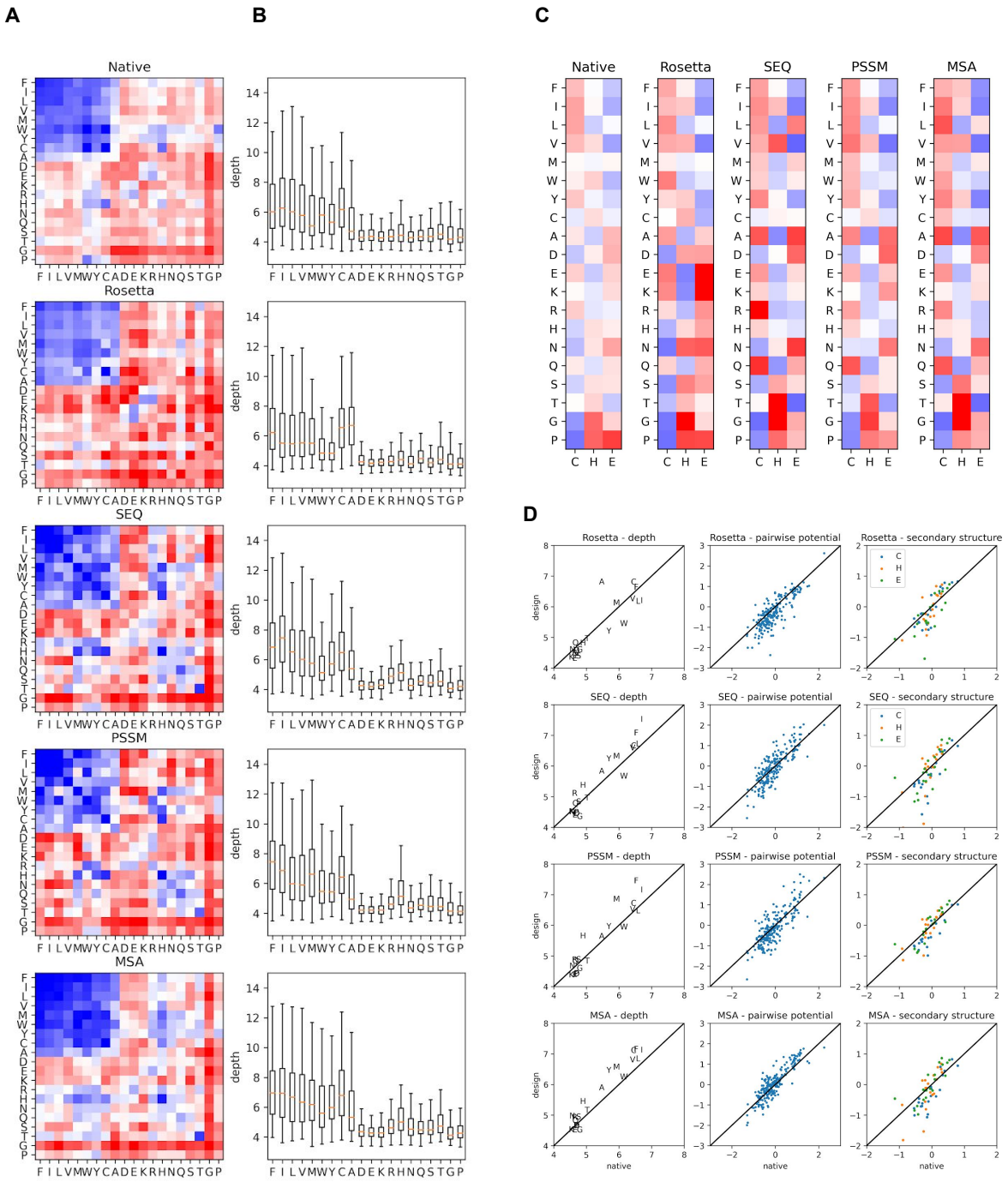

Fig. S12. TrRosetta designs recapitulate statistics from the PDB. For the 209 native proteins from the dataset described in Supplementary Methods, we designed new sequences with trRosetta using single-sequence, PSSM and MSA mode. (**A**) For each design, we threaded the sequence onto the original backbone, and repacked the side-chains using SCWRL4 (37), with contact defined as the minimal distance between any two heavy atoms less than 5.0 Å apart. The protocol recapitulates the pairwise potential, defined as $log(P(a,\ b \mid contact) / (P(a) \cdot P(b))$. (**B**) For each amino acid type, we computed the distribution of depth (using DEPTH (33)). (**C**) Amino acid propensities for each secondary structure, computed as $log(P(aa \mid ss) / P(aa))$. For (**A**), (**C**), the scale for the color is -1.5 to 1.5. The results are summarized in (**D**).

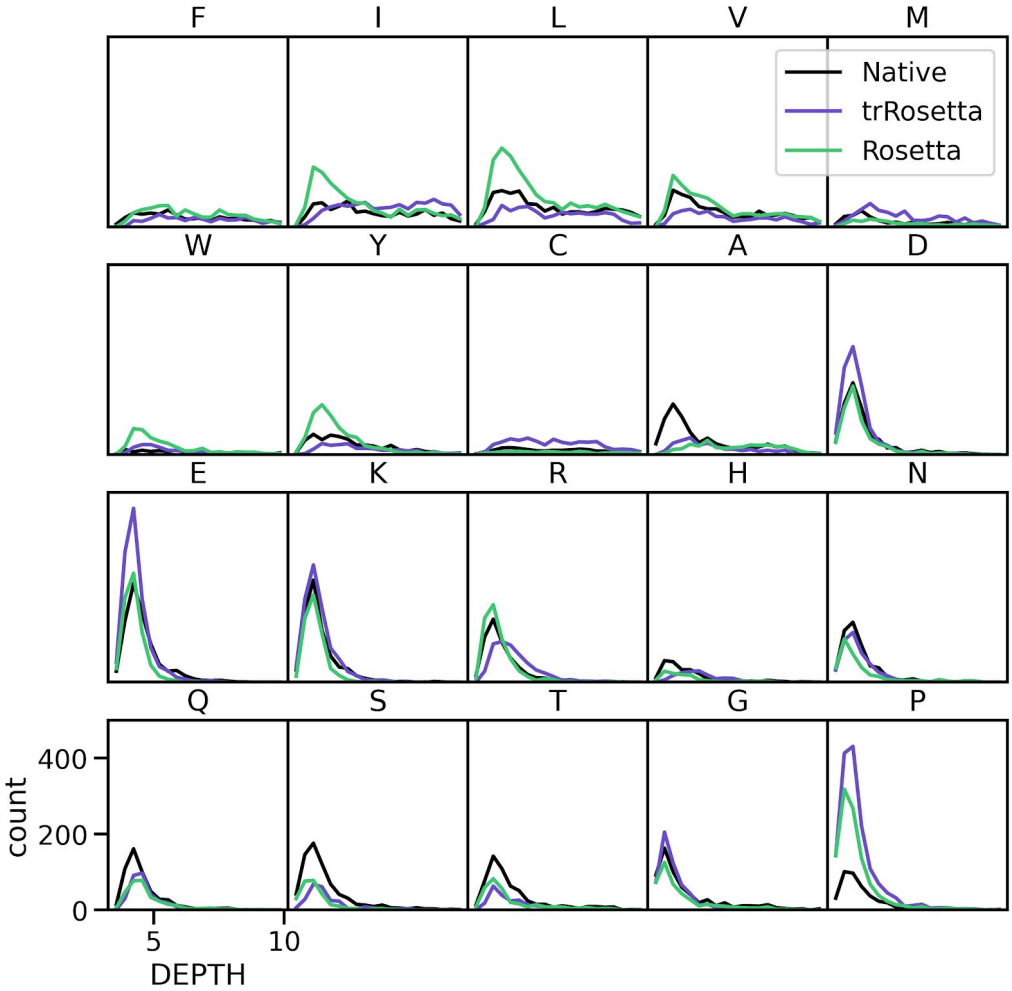

Fig. S13. DEPTH (33) distributions for single sequence design on a set of native protein backbones using Rosetta or trRosetta. The distributions of the original sequences are shown for comparison.

Table S1. TrRosetta designed sequences for a subset of the native proteins described in the Supplementary Methods with different optimization settings. Because of GPU memory limitations, we only designed the 103 native proteins that were less than, or equal to, 139 amino acids. Median cross entropy (10 Å contact cutoff), and median native sequence recovery are shown.

Gradient Scaling	Discretization method	Sequence format	Learning rate	Decay rate	Median -logP(contacts|sequence)	Median native sequence recovery
decay	argmax	MSA	n/aN	1	1.21	0.13
decay	argmax	MSA	n/a	2	1.22	0.13
decay	argmax	single	n/a	1	1.47	0.15
decay	argmax	single	n/a	2	1.46	0.15
decay	sample	MSA	n/a	1	1.95	0.14
decay	sample	MSA	n/a	2	2.14	0.12
decay	sample	single	n/a	1	2.34	0.11
decay	sample	single	n/a	2	2.42	0.11
constant	argmax	MSA	0.01	n/a	1.27	0.13
constant	argmax	MSA	0.10	n/a	1.21	0.14
constant	argmax	MSA	1.00	n/a	1.25	0.13
constant	argmax	MSA	10.00	n/a	1.85	0.08
constant	argmax	single	0.01	n/a	1.51	0.15
constant	argmax	single	0.10	n/a	1.50	0.15
constant	argmax	single	1.00	n/a	1.44	0.15
constant	argmax	single	10.00	n/a	2.24	0.08
constant	sample	MSA	0.01	n/a	2.54	0.11
constant	sample	MSA	0.10	n/a	1.54	0.15
constant	sample	MSA	1.00	n/a	1.37	0.11
constant	sample	MSA	10.00	n/a	1.69	0.09
constant	sample	single	0.01	n/a	2.47	0.11
constant	sample	single	0.10	n/a	2.05	0.14
constant	sample	single	1.00	n/a	1.49	0.15
constant	sample	single	10.00	n/a	2.30	0.08
n/a	MCMC	single	n/a	n/a	1.57	0.14

Supplementary Methods

Generation of trRosetta-based PSSMs

To generate position-specific scoring matrices (PSSMs) for trRosetta-guided Rosetta design (the hybrid method), we optimized trRosetta sequence profiles for Foldit-player designed backbones over 100 gradient descent steps using argmax discretization and constant gradient scaling. We generated 100 sequences for each Foldit-player design following the method described in Fig. 1B, S1A, and from those, computed log-odds scores by dividing the position-specific amino acid probabilities with their mean probabilities across all sites and taking the natural log. For the faster method (hybrid-f, see below), we designed a single sequence (100 gradient descent steps), and stored the probabilities ($\hat{Y}$) directly and from those computed natural log-odds as stated above.

Hyperparameter optimization for fast PSSM generation (hybrid-f method)

To make the hybrid design method computationally less expensive, we experimented with the generation of PSSMs directly from $\hat{Y}$, instead of computing one hundred individual sequences. This method decreases PSSM generation time by about a 100-fold, with only minor reduction in the quality of the resulting energy landscapes (Fig. S11). This method required hyperparametrization, which we describe below.

We began by modifying Eq. 3, by adding two additional loss terms calibrated to maximize sequence recovery and reproduce the distribution of site-specific entropies observed in sequence alignments of native proteins, effectively biasing and regularizing the fitted amino acid probability distributions:

$$Loss_{hybrid-f} = Loss + \sum_{a=1}^{20} \sum_{i}^{L} \widehat{Y}_{ai}\, w_a + w_H D_{KL}(H_{native} || H_{design}) \qquad \text{(S1)}$$

Here $\widehat{Y}_{ai}$ is the probability of the a^{th} canonical amino acid at residue position i (Eq. 4a), w_a is the bias loss weight associated with amino acid a, $D_{KL}(H_{native} || H_{design})$ is the KL-divergence between a reference distribution of site-entropies obtained from a set of native protein sequence alignments and the distribution of site-entropies for the designed protein (see below), and w_H is the weight of the KL-divergence term. The second term maximizes the agreement between position-specific amino acid frequencies (sequence profile, $\widehat{Y}$) of natural proteins and those designed by trRosetta. The third term pushes the distribution of single-site Shannon entropies in a designed sequence profile to match the distribution of entropies seen in PSSMs derived from alignments of native proteins.

To build a training set for the optimization, we collected native structures and their corresponding sequence alignments from Hiranuma *et al.* (38), and reduced the redundancy of the set so that no structure had more than 30% sequence identity to any protein in the trRosetta training dataset. This resulted in 226 proteins of which 209 could be designed with trRosetta (sequence length/memory limitations). Native sequence profiles were computed for these proteins with PSI-BLAST (39, 40). Next, to fit the weights w_a and w_H , we used Nelder-Mead simplex optimization (41) to minimize a loss function, $Loss_{NM}$, composed of two terms.

$$Loss_{NM} = \sum_{i=1}^{n} D_{KL}(\widehat{Y}_{native} || \widehat{Y}_{design})/n + 2D_{KL}(C_{native} || C_{design}) \qquad \text{(S2)}$$

where $\widehat{Y}$ is a sequence profile as defined in Eq. 4a, *n* is the number of proteins in the training set, and *C* is the average amino acid composition across *n* sequence profiles. The first term is the mean per-site KL-divergence between a set of native sequence profiles and profiles generated by outputting the softmax-logits from trRosetta. The second term was included to ensure that designed sequences maintained near-native mean amino acid frequencies, and is the KL-divergence between the mean amino acid composition of all profiles in the training set, and the mean amino acid composition of their respective trRosetta designs. We ran the minimization process on the *n* = 209

proteins for approximately 3 days using 20 RTX 2080 GPUs in parallel. We did this hyperparameter optimization for the three design modes of trRosetta: SEQ, PSSM, and MSA. The hyperparameters can be found on GitHub.

Generation of fragment-based PSSMs

To improve the local sequence-structure relationship it can be helpful to restrict and bias sequence design by sequence profiles derived from natural sequences. This has been successfully applied to a range of design problems where native protein families were available (42–45). For *de novo* structures and structure remodeling, where natural sequence alignments are unavailable, fragment-based approaches have been used instead (22, 46). Here, we used the approach from Brunette *et al.* (22), which generates sequence profiles by taking the sequences of matching native structure fragments for each 9-residue-long sliding window in the design. To account for sequence stretches where only a few native fragments are found, we add a small pseudocount to the observed amino acid frequencies (f_a):

$$p_a = \frac{\alpha \cdot f_a + \beta \cdot g_a}{\alpha + \beta}$$

where α is the total number of observed amino acids for that site, β is a constant pseudocount of 50, and g_a is the pseudo frequency computed as:

$$g_b = \sum_{a=1}^{20} f_a \cdot q(b|a)$$

where *q(b|a)* is the conditional probability of amino acid *b* given *a* in the BLOSUM62 matrix (47). Finally, we computed log-odds ratio scores from the pseudocount-corrected frequencies and the background frequencies of each amino acid as in BLOSUM62:

$$s_{fragment} = ln(p_a / p_{a,\,background})$$

Hyperparameter optimization for hybrid-f and fragment PSSM design methods

To improve the performance curves of the hybrid-f and fragment-based PSSM design methods, we hyperparameterized the cutoff value for the log-odds scores and weight applied during design (see below). These hyperparameters were sampled on a grid, and the values corresponding to the Pareto front with respect to energy per residue and average fragment quality were selected. The hyperparameters used in the hybrid method were not optimized, and standard values were used instead.

Restricting and biasing Rosetta sequence design

Sequence design with Rosetta was performed with FastDesign (28, 31), using the Rosetta all-atom energy function (32) with beta16_nostab weights. We applied three different methods for restricting and biasing amino acid choices during FastDesign. In LayerDesign (21), amino acid identities were restricted based on the site's surface exposure, secondary structure type and helix capping, generally disallowing hydrophobic amino acids on the surface, and hydrophobic amino acids in the core. For the two other methods (fragment-based PSSMs and trRosetta-based PSSMs), we disallowed amino acids according to their log-odds scores ($s_{hybrid-f} > -0.5$, $s_{hybrid} > -1$, $s_{fragment} > -0.125$), and biased the energy function with these PSSMs during design ($w_{hybrid} = -0.3$, $w_{hybrid-f} = -1.5$, $w_{fragment} = -1.75$).

Bootstrap analysis of the differences in Pearson correlation coefficients

In Fig. 3A. we estimated the p-value for the null-hypothesis $R^2_{Rosetta} \geq R^2_{trRosetta}$. This was done using the bootstrap analysis (48). Specifically, we computed the number of times $R^2_{Rosetta}$ was greater than or equal to $R^2_{trRosetta}$ using random sampling with replacement (sample size = 30,159), taking 10,000 bootstrap samples. For both computed p-values, the correlation coefficient for Rosetta was never higher than that of trRosetta.

List of Foldit players

Alan Coral	Jamie Quinn	Owen Yin
Alex J. Bubar	Jane Norrgard	Pascal Wolfteich
Alexander Boykov	Jason Truong	Paul Gummersall
Alexander Uriel Valle Pérez	Jasper A. Diderich	Paweł Tłuścik
Alison MacMillan	Jeffrey Michael Canfield	Peter Gajar
Allen Lubow	Jeffrey Photakis	Peter John Triggiani IV
Andrea Mussini	Jesse David Slone	Rajarshi Guha
Andrew Cai	Joanna Madzio	Renton Braden Mathew Innes
Andrew John Ardill	Joanne Mitchell	Ricky Buchanan
Aniruddha Seal	John Charles Stomieroski	Robert Gamble
Artak Kalantarian	John H. Mitch	Robert Leduc
Barbara Failer	Johnathan Robert Altenbeck	Robert Spearing
Belinda Lackersteen	Jonas Schinkler	Rodrigo Luccas Corrêa dos Santos
Benjamin Chagot	Jonathan Barak Weinberg	Gomes
Beverly R. Haight	Joshua David Burbach	Roger D. Estep
Bora Taştan	João Carlos Sequeira da Costa	Ryan DeWitt
Boris Uitham	Juan Francisco Bada Juarez	Ryan Moore
Brandon G. Roy	Jón Pétur Gunnarsson	Scott G. Shnider
Breno Renan de Melo Cruz	Kathleen Diane Harper	Scott J. Zaccanelli
Brian Echols	Keehyoung Joo	Sergey Kuznetsov
Brian Edward Lorenz	Keith T. Clayton	Sergio Burillo-Sanz
Bruce Blair	Kenneth E. DeFord	Seán Mooney
Bruno Kestemont	Kevin F. Scully	Sidoruk Vasiliy
C. D. Eastlake	Kevin M. Gildea	Slava S. Butkovich
Callen Joseph Bragdon	Kirk J. Abbey	Spencer Bruce Hudson
Carl Vardeman	Kristen Lee Kohli	Spencer Len Pote
Carlo Salerno	Kyle Stenner	Stephen Phillip Denne
Casey Comisky	Kálmán Takács	Steven A. Schwegmann
Catherine Louise Hayman	LaVerne L. Poussaint	Sumanth Ratna
Catherine R Landers, MD	Larry C. Manalo Jr.	Susan C. Kleinfelter
Cathy Zimov	Larry C. Withers	Thomas Bausewein
Charles David Coleman	Lilium Carlson	Thomas J. George
Charles Robert Painter	Linda Wei	Tobias Scherf de Almeida
Christopher Ince Jr.	Luke Ryan Fisher	Ulas Yeginer
Conor Lynagh	Lynn Carpenter	Walter Barmettler
Dmitrii Malaniia	Ma Ji-hwan	Warwick Robert Pulley
Douglas Craig Wheeler	Manuel Ricci	William Scott Wright
Douglas Robertson	Marcus Anthony Belcastro	Willyanto
Dr. Vera Simon	Marek Leniec	Wyatt Lansford
Emanuele Chisari	Marie Hohmann	Xavier Hochart
Eric Lim Jit Kai	Mark Thompson	Yoan Anthony Skander Gaiji
Farah Rezae	Matthew A. Thayer	Yuriy Lagodich
Ferenc Lengyel	Matthias Gaebel	Vivier Christian
Flavian Tabotta	Michael D. Cassidy	
Franco Padelletti	Michael Fagiola	
Frisno Boström	Michael Lewis	
Gary O. Gross	Michael Pfützenreuter	
George McIlvaine	Michael Simon	
Gil Beecher	Moamen M. Elmassry	
Gregory T. Hansen	Noah Benevides	
Guido de Jong	Norah Kathleen Kerr	
Harald Feldmann	Nupur Verma	
Jami Lynne Borman	Oak Shannon	

SI References

34. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
35. E. H. Kellogg, A. Leaver-Fay, D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins Struct. Funct. Bioinforma.* **79**, 830–838 (2011).
36. H. Park, *et al.*, Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
37. G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778–795 (2009).
38. N. Hiranuma, H. Park, I. Anishchanka, M. Baek, D. Baker, Improved protein structure refinement guided by deep learning based accuracy estimation. *bioRxiv*, 2020.07.17.209643 (2020).
39. S. Altschul, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
40. T. Oda, K. Lim, K. Tomii, Simple adjustment of the sequence weight algorithm remarkably enhances PSI-BLAST performance. *BMC Bioinformatics* **18**, 288 (2017).
41. J. A. Nelder, R. Mead, A Simplex Method for Function Minimization. *Comput. J.* **7**, 308–313 (1965).
42. G. D. Lapidoth, *et al.*, AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins Struct. Funct. Bioinforma.* **83**, 1385–1406 (2015).
43. D. Baran, *et al.*, Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci.* **114**, 10900–10905 (2017).
44. A. Goldenzweig, *et al.*, Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **63**, 337–346 (2016).
45. O. Khersonsky, *et al.*, Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **72**, 178-186.e5 (2018).
46. R. Netzer, *et al.*, Ultrahigh specificity in a network of computationally designed protein-interaction pairs. *Nat. Commun.* **9**, 5286 (2018).
47. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**, 10915–10919 (1992).
48. B. Efron, Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **7**, 1–26 (1979).