Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models

Joseph L. Watson[‡1,2], David Juergens[‡1,2,3], Nathaniel R. Bennett[‡1,2,3], Brian L. Trippe[‡2,4], Jason Yim[‡2,5], Helen E. Eisenach[1,2], Woody Ahern[1,2,6], Preetham Venkatesh[1,2,7], Susana Vázquez Torres[1,2,7], Andrew J. Borst[1,2], Basile I. M. Wicky[1,2], Robert J. Ragotte[1,2], Lukas F. Milles[1,2], Alexis Courbet[1,2,8], William Sheffler[1,2], Jue Wang[1,2], Isaac Sappington[1,2,7], Samuel J. Pellock[1,2], Nikita Hanikel[1,2], Regina Barzilay[5], Tommi S. Jaakkola[5], Neil P. King[1,2], Frank DiMaio[1,2], Minkyung Baek[9], David Baker*[1,2,8]

‡Equal contribution
*To whom correspondence should be addressed

1. Department of Biochemistry, University of Washington, Seattle, WA 98105, USA
2. Institute for Protein Design, University of Washington, Seattle, WA 98105, USA
3. Molecular Engineering Graduate Program, University of Washington, Seattle, WA 98105, USA
4. Columbia University, Department of Statistics, New York, NY 10027, USA
5. Massachusetts Institute of Technology, Cambridge, MA 02139, USA
6. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA 98105, USA
7. Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA 98105, USA
8. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98105, USA
9. School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea

Abstract

Deep learning methods for protein design have shown considerable promise for sequence design, scaffolding functional sites[1,2], and building new monomers[3], cyclic oligomers[4], and antibody loops[5,6]. Despite this progress, a general framework for protein design that enables solution of a wide range of design challenges, including *de novo* binder design and design of higher order symmetric architectures, has yet to be described. Diffusion models[7,8] have had considerable success in image and language generative modeling, and have been applied to the protein monomer generation problem but with limited success likely due to the complexity of protein backbone geometry and sequence-structure relationships. Here we show that by utilizing powerful structure prediction methods, which have a deep understanding of protein sequence and structure, as the diffusion denoising network, we can improve on the state of the art for unconditional and topology constrained protein monomer design, protein and peptide binder design, symmetric oligomer design, *de novo* enzyme design and symmetric motif scaffolding for therapeutic and metal-binding protein design. We demonstrate the power and generality of the method, called RoseTTAFold Diffusion (RF*diffusion*), by experimentally characterizing hundreds of new designs. Highlights include a picomolar binder to parathyroid

hormone, considerably higher affinity than any previous computational designed binder prior to experimental optimization, and a series of not previously observed symmetric assemblies experimentally confirmed by electron microscopy.

Main

Denoising diffusion probabilistic models (DDPMs) have emerged as a powerful class of generative models to sample from complex data distributions[7,8]. DDPMs are trained to reconstruct data (for instance images or text) corrupted with varying amounts of added noise. After this training, new samples can be generated by feeding the model random noise and then refining it by iterative application of the trained denoising network[7,8]. The power of DPPMs is illustrated in the context of computer graphics by the generation of novel, photorealistic images in response to text prompts[9]. DDPMs have a number of qualities that naturally lends them to protein design: they (1) generate highly diverse outputs, due to the stochasticity of the inputs and subsequent denoising trajectory (2) can be guided at each step of the iterative data generation process towards specific design objectives, either through provision of conditioning information or through external "potentials", and (3) unlike methods that design proteins through generation or optimization of protein *sequences* alone[1,3,4,10,11], DDPMs can be formulated to generate protein *structures* directly, enabling more direct control over structural properties. Recent work has sought to adapt DDPMs for protein monomer design by conditioning on small protein "motifs"[2,6] or on secondary structure and adjacency ("fold") information[5]. While showing promise, these attempts have thus far had limited success in generating sequences that are predicted to fold to the intended structures *in silico*[2,12], and have not been tested experimentally.

We reasoned that improved diffusion models for protein design could be developed by taking advantage of the deep understanding of protein sequence and structure implicit in powerful structure prediction methods like AlphaFold2 (AF2) and RoseTTAFold (RF). The power of fine-tuning pretrained structure prediction networks for protein design is illustrated by an inpainting version of RoseTTAFold (called RF_{joint}[1]) that was trained to recover missing sequence and structure information. Experimental characterization showed that the method can scaffold a wide range of protein functional sites with atomic accuracy (Ref [1], see eLetter), but the approach fails to generate useful designs for more minimalist sites when there is insufficient topological information and, because it is deterministic, sampling of different input parameters is required to generate output diversity. We reasoned that by instead fine-tuning RoseTTAFold for use as the neural network in a diffusion based generative model, both problems could be overcome: because the starting point is random noise, each denoising trajectory yields a different solution, and because structure is built up progressively through many denoising iterations, much less starting information would likely be required.

We formulate the diffusion model in a manner well-suited to fine-tuning from pre-trained RoseTTAFold (Fig 1A). As in ref [5], at each timestep we predict the final protein structure given the current noised structure. We then generate the slightly denoised input to the next timestep via a noisy interpolation from the current (input) structure toward the predicted final structure. The correspondence between RoseTTAFold structure prediction and a RF*diffusion* denoising

step is highlighted in Fig. 1A: in either case, input sequence and structure information is transformed by the model into a prediction of native protein structure. During classical structure prediction with RoseTTAFold, structural inputs to the model come from homologous template structures, each of which have associated per-residue "confidence" values $\epsilon[0,1]$[13]. In RF*diffusion*, structural inputs are derived from the partially (de-)noised structure, and the confidence feature is reparameterized to represent the current denoising timestep, on which the model conditions its prediction (see methods). To generate noised protein structures for training or inference, we perform "forward" diffusion on some subset (potentially all) of the amino acids in a protein over backbone N-C_α-C frame translations and rotations. For translations, we perturb the C_α coordinates with 3D Gaussian noise. For rotations, we use Brownian motion on the manifold of rotation matrices, SO(3) (building on refs [14,15]). The noised structures are input to the network via the structure (3D) track of RF. While in this study we use RoseTTAFold as the basis for the denoising network architecture, our approach is quite general, and it should be possible to substitute in other structure prediction networks that manipulate 3D coordinates (AF2[12], Omegafold[17], ESMfold[18], etc.).

We explored two different strategies for training RF*diffusion*. Firstly, we trained it in a manner akin to "canonical" diffusion models, such that predictions at each timestep are independent of predictions at previous timesteps (as in previous work[2,5,6,12]). Secondly, we trained it allowing the model to condition on previous predictions between timesteps (as in self-conditioning[19], Fig. 1A bottom row). The latter strategy was inspired by the success of "recycling" (allowing the model to condition on previous predictions) in both AF2 and RF_{joint} Inpainting. We found that self-conditioning within RF*diffusion* dramatically improved performance on *in silico* benchmarks encompassing both conditional and unconditional protein design tasks (Fig. S1A, Table 1, methods). Fine-tuning RF*diffusion* from a pre-trained RF model was also crucial (Fig. S1B). For all *in silico* benchmarks in this paper, we use the AF2 structure prediction network[16] for validation and define *in silico* "success" as an RF*diffusion* output for which the AF2 structure predicted from a single sequence (1) has high confidence (mean predicted aligned error, pAE, < 5), (2) is globally within 2 Å backbone-RMSD of the designed structure, and (3) is within 1 Å backbone-RMSD on the scaffolded functional-site. Though more stringent than metrics described elsewhere[2,5,12,20] (TM score between design and subsequent structure prediction, see Fig. S2A-B), these metrics have been demonstrated to be good predictors of experimental success[1,4,21]. Because RoseTTAFold and AF2 are entirely different networks, AF2 serves as a reasonably independent arbitrator of the success of a design calculation. To design amino acid sequences that encoded the RF*diffusion*-generated backbones, we experimented both with simultaneous diffusive generation of sequence within RF*diffusion* (similar to refs [19,22]) and with simply using the rapid and robust ProteinMPNN network[23]. Benchmark tests showed that the latter approach was more powerful and more flexible (as multiple diverse sequences can be readily generated for each output structure) and we use this throughout the remainder of the paper. We generate 8 ProteinMPNN sequences per backbone, and select those predicted to fold to the target structure most accurately by AF2 (in line with previous work[2,12]).

Unconditional protein monomer generation

Unconstrained generation of diverse protein monomers is a longstanding challenge in protein design which is difficult to address with physically-based protein design methods due to the magnitude of the conformational sampling problem, and has been a primary test of deep learning based protein design approaches[2,3,5,10,12,24]. As illustrated in Fig. 1C-E, Fig. S3A, RF*diffusion* can readily generate complex protein structures from scratch with little overall structural similarity to any known protein structures, indicating considerable generalization beyond the PDB training set. The designs span a wide range of alpha-, beta- and mixed alpha-beta- topologies, with AF2 predictions very close to the design structure models for de novo designs with as many as 600 residues (we interestingly found that ESMFold[20] even more closely recapitulated the design structures - Fig. S1H, S2A, but given the experimental success in using AF2 for design validation[1,4,21], we used AF2 as the primary independent validation for the design challenges described in this study). RF*diffusion* generates plausible structures for even very large proteins, but these are difficult to validate *in silico* as they are likely beyond the single sequence prediction capabilities of AF2. The quality and diversity of designs that are sampled is inherent to the model, and does not require *any* auxiliary conditioning input (for example secondary structure information[5]). RF*diffusion* strongly outperforms Hallucination, the only experimentally validated deep learning approach for unconditional generation, with success rates for Hallucination deteriorating beyond 100 amino acids. RF*diffusion* is also far more compute efficient than unconstrained hallucination, requiring 2:20 minutes on an NVIDIA RTX A4000 GPU to generate a 100 residue structure compared to 8:20 for Hallucination. Computational efficiency can be further improved by taking larger steps at inference time, and by truncating trajectories early - an advantage of predicting the *final* structure at each timestep (Fig S2C-D). For design problems where a particular fold or architecture is desired (such as TIM barrels or cavity-containing NTF2s for small molecule binder and enzyme design[25,26]), we further fine-tuned RF*diffusion* to condition on (partial) input secondary structure and/or fold information, enabling rapid and accurate generation of diverse designs with the desired topologies or folds (Fig. S3B-D). *In silico* success rates were 42.5% and 54.1% for TIM barrels and NTF2s respectively (Fig. S3C).

Higher order oligomer design through denoising with explicit symmetrization

There is considerable interest in designing new higher order symmetric oligomers which can serve as vaccine platforms[27], delivery vehicles[28], and catalysts[29]. Cyclic oligomers have been generated using structure prediction networks by starting from a random sequence and carrying out a Monte Carlo search for sequences predicted to fold to the desired cyclic symmetry[4]. This “hallucination” approach fails with higher order dihedral, tetrahedral, octahedral, and icosahedral symmetries, likely because these architectures require multiple distinct sets of monomer-monomer interactions. We reasoned that this limitation could be overcome using our diffusion framework. As RF*diffusion* acts directly on amino acid coordinates (as opposed to input sequence tokens), it allows the explicit symmetrisation of the “denoising” process, and RF's equivariance properties with respect to global rotation of coordinate inputs ensures that the targeted symmetry is maintained in denoising predictions (see methods). We experimented with

arranging multiple copies of a starting random Gaussian monomer coordinate distribution with the desired symmetry as the input, and explicitly symmetrizing the denoising updates at each step (Fig. 1B, second row). For octahedral and icosahedral architectures, to reduce the computational cost and memory footprint, we only explicitly modeled the smallest subset of monomers required to generate the full assembly (in the icosahedral case, the subunits at the five-fold, three-fold, and two-fold symmetry axes).

We found that despite not being trained on symmetric inputs, RF*diffusion* was able to generate higher order symmetric oligomers with high *in silico* success rates (Fig. S4A), particularly when guided by an auxiliary inter- and intra- chain contact potential (Fig. S4B). As illustrated in Figure 2A,B,D and Fig. S4C, D, RF*diffusion*-generated cyclic (C3, C5, C6, C8, C10, C12), dihedral (D2, D3, D4, D5), tetrahedral, octahedral and icosahedral designs are nearly indistinguishable from AF2 predictions of the structures adopted by the designed sequences (for the full assemblies for the cyclic and dihedral designs, and trimeric substructures of the octahedral and icosahedral designs). These include a number of topologies not seen in nature, including two-layer beta strand barrels (Fig. 2A, bottom row) and complex mixed alpha/beta topologies (Fig. 2A). We selected 376 designs for experimental characterization, and found using size exclusion chromatography that at least 37 had oligomerization states closely consistent with the design models. We collected negative stain electron microscopy (nsEM) data on five of the 37 designs with the highest total molecular weights (ranging from 70-110 kilodaltons), and for four of the five distinct particles were evident with shapes resembling the design models (Fig 2C,D).

Averages of the electron microscopy data for the four designs are shown in Fig. 2D. All structures of all four assemblies are to our knowledge unprecedented in nature. HE0537 is a D4 octameric dihedral assembly resembling a dimer of tetramers with an overall rectangular prism shape (5x5x6 nanometers) formed by a largely alpha helical monomer. The electron microscopy images clearly indicate the rectangular prism shape in both top down and side views. HE0600 is a C12 dodecameric ring shaped assembly with an inner 24 stranded barrel buttressed by an outer ring of 12 helices. Electron microscopy micrographs reveal homogeneous ring shaped particles, and the resulting 2D class averages are consistent with the design models. HE0675 is a C8 octameric ring composed of an inner ring of 16 strands and an outer ring of 16 helices. In contrast to HE0600, the helices are not parallel to each other, instead forming a flower-like pattern. The electron microscopy individual particle images and corresponding 2D class averages are again closely consistent with the design model. HE0626 is a C6 hexameric ring composed of an inner ring of 18 strands and an outer ring of 18 helices. As in HE0675, the helices are packed in a flower-shaped arrangement, and nsEM micrographs and 2D class averages are again close to the design model (Fig. 2D). Importantly, we were also able to obtain a nsEM 3D reconstruction of HE0626 (Fig. 2E), where rigid-body docking of the design model demonstrated strikingly high agreement with the density map. This agreement empirically validates the design of the intended two ring architecture and the flower-shaped outer ring surface of this symmetric oligomer.

Functional-site scaffolding with RFdiffusion

We next investigated the use of RF*diffusion* for scaffolding protein structural motifs that carry out binding and catalytic functions, where the role of the scaffold is to hold the site in precisely the 3D geometry needed for optimal function. A number of deep learning methods have been recently developed to address this problem, including RF_{joint} Inpainting[1], constrained hallucination[1], and probabilistic diffusion networks[2,5,20]. To rigorously evaluate the performance of these methods in comparison to RF*diffusion* across a representative set of design challenges, we established an *in silico* benchmarking test comprising all functional site scaffolding design problems described in six recent publications[1,2,20,30–32] encompassing both deep learning-based and conventional design methodologies. There are 25 challenges in total, spanning a broad range of functional sites, including simple “inpainting” problems, viral epitopes, receptor traps, small molecule binding sites, binding interfaces and enzyme active sites. Full details of this benchmark are described in Table 1. RF*diffusion*, with no prior optimization on the problem set, outperforms Hallucination (where some preliminary optimization was used) and Inpainting in all but one design problem, and provides solutions to six problems for which hallucination and inpainting, even with the aid of ProteinMPNN, fail to generate successful designs under these conditions *in silico* (Fig. 3A-C). In 17/23 of the problems, RF*diffusion* generated successful solutions with higher success rates when noise was not added during the reverse diffusion trajectories.

Fine-tuning RFdiffusion for enzyme design from minimalist active site motifs

A grand challenge in protein design is the ability to scaffold minimalist descriptions of enzyme active sites (typically just a few single amino acids). While some *in silico* success has been reported previously[1], a general solution that can readily produce high-quality, orthogonally-validated outputs is not currently available. Following fine tuning on training examples involving scaffolding of the relative orientations and geometries of 2-3 residues close in Euclidean space, but discontinuous in sequence space, RF*diffusion* was able to scaffold enzyme active sites comprised of multiple sidechain and backbone functional groups with high accuracy and *in silico* success rates (Fig. 3D-F), illustrating the ease with which RF*diffusion* can be fine-tuned to solve problems beyond those in the original training set. While RF*diffusion* is currently unable to *explicitly* model bound small molecules (see conclusion), the substrate could be *implicitly* modeled using an auxiliary potential, which could be used to guide the generation of “pockets” around the active site (Fig. S5).

Symmetric functional-site scaffolding for metal mediated assemblies and antiviral therapeutics and vaccines

A number of important design challenges involve the scaffolding of multiple copies of a functional motif in symmetric arrangements. For example, many viral glycoproteins are trimeric, and symmetry matched arrangements of inhibitory domains can be extremely potent[33–36]. On the other hand, symmetric presentation of viral epitopes in an arrangement that mimics the virus

could induce new classes of neutralizing antibodies[37,38]. To explore this general direction, we sought to design trimeric multivalent binders to the SARS-CoV-2 spike protein. In previous work, flexible linkage of a design that binds to the ACE2 binding site on the receptor binding domain of the spike to a trimerization domain yielded a high-affinity inhibitor that had potent and broadly neutralizing antiviral activity in animal models[33]. Rigidly fusing or oligomerizing the binder could in principle improve its affinity for the target by reducing the entropic cost of binding while maintaining the avidity benefits from multivalency. We used RF*diffusion* to design C3 symmetric trimers which rigidly hold three binding domains (the “functional-site” in this case) so they exactly match the ACE2 binding sites on the SARS-CoV-2 spike protein trimer. Design models were confidently recapitulated by AF2 to both assemble as C3-symmetric oligomers, and to scaffold the AHB2 SARS-CoV-2 binder interface with sub-angstrom accuracy (Fig. 3G).

The ability to scaffold functional sites with any desired symmetry opens up new approaches to design protein-metal assemblies. Divalent metal ions exhibit distinct preferences for different metal coordination geometries - square planar (C4), tetrahedral, and octohedral - with ion-specific optimal sidechain-metal bond lengths. RF*diffusion* provides a general route to building up symmetric protein assemblies around such sites. As a first test of this, we sought to design square planar Nickel binding sites. We designed C4 protein assemblies with four central histidine imidazoles arranged in ideal Nickel binding geometry (Fig. 3H, left panel). Designs starting from two different C4-symmetric histidine functional sites showed high *in silico* design success (Fig. 3H, middle and right panel). The confident AF2 predictions of the histidine residues in the desired geometry strongly suggests these designs would indeed bind to Nickel, although further experimental evidence will be required to verify this.

De novo protein and peptide binder design

The design of high-affinity binders to target proteins is a grand challenge in protein design, with numerous therapeutic applications[39]. The ability to design binders *de novo* using the physically based Rosetta method was recently described[40], and subsequently, the utility of ProteinMPNN and AF2 for sequence design and design filtering respectively has improved design success rates[21]. However, experimental success rates are typically low, requiring many thousands of designs to be screened for each design campaign[40]. Further, this work relied on pre-specifying a particular set of protein scaffolds as the basis for the designs, inherently limiting the diversity and shape complementarity of possible solutions[40]. We reasoned that RF*diffusion* might be able address this challenge by directly generating diverse, and target-compatible protein binders. To our knowledge, no deep-learning method has yet demonstrated general experimental success in designing binders completely *de novo*.

It is often desirable to be able to specify specific sites on a target protein that binders should bind to. We therefore fine-tuned RF*diffusion* on protein complex structures, providing as input a subset of the interface residues the diffused chain binds (Fig. S6A, B, see methods). With this fine-tuned model, we were able to design putative binders confidently predicted by AF2 to bind their target[21]. These could be generated without any fold/topology information, with success rates several orders of magnitude higher than with our previous Rosetta based approach (Fig.

4A-B). To enable control over binder scaffold topology, we also fine-tuned a model to condition binder diffusion on secondary structure adjacency information[5] (Fig. S6C, D), and in cases where compatible folds for putative binders were known, this model typically further improved *in silico* success rates (Fig. 4B, bottom row).

An outstanding challenge in protein design is the design of binders to flexible helical peptides, which are challenging targets due to their general lack of structure in solution and therefore the entropic cost of binding in a rigid conformation. Experimental characterization revealed that RF*diffusion* had quite remarkable success on this problem. Designed binders to the apoptosis-related peptide Bim were found using biolayer interferometry to have single digit nanomolar affinity (Fig. 4C, D). Designed binders to parathyroid hormone (PTH) had even higher affinity: fluorescence polarization measurements indicated a dissociation constant of ~340pM (Fig. 4E, F). To our knowledge, this is the highest affinity binder to any target (protein, peptide, or small molecule) achieved directly by computational design with no experimental optimization.

Conclusion

RFdiffusion is a major improvement over current physically based and deep learning protein design methods over a wide range of protein design challenges. Substantial progress was recently made using Rosetta in designing binding proteins from target structural information alone, but this required testing tens of thousands of designs – with RF*diffusion* high affinity binders can now be identified in testing of dozens of targets. There has also been progress in scaffolding protein functional motifs using deep learning methods (hallucination, inpainting and diffusion), but hallucination becomes very slow for complex systems, inpainting fails when insufficient starting information is provided, and previous diffusion methods had quite low accuracy; our benchmark tests show that RF*diffusion* considerably outperforms all previous methods in the complexity of the motifs that can be scaffolded, the ability to precisely position sidechains (for catalysis and other functions), and the accuracy of motif recapitulation by AF2. For the classic unconstrained protein structure generation problem, RF*diffusion* readily generates novel protein structures with as many as 600 residues that are accurately predicted by AF2 (and ESMFold), far exceeding the complexity and accuracy achieved by previously described diffusion and other methods. The versatility and control provided by diffusion models enabled extension of RF*diffusion* unconditional generation to higher order architectures with any desired symmetry (hallucination methods are primarily limited to cyclic symmetries); experimental characterization of a subset of these designs using electron microscopy revealed overall all structures very similar to the design models which are without precedent in nature. Combining the accurate motif scaffolding with the ability to design symmetric assemblies, we were able to scaffold functional sites spanning multiple symmetrically arranged chains which has not been previously possible. Overall, the complexity of the problems solvable with RF*diffusion* and the robustness and accuracy of the solutions (as evaluated by similarity to AF2 predictions *in silico* and by experimental characterization) far exceeds what has been achieved previously.

The power and scope of RF*diffusion* can be extended in several directions. RF has recently been extended to nucleic and protein-nucleic acid complexes[13] which should enable diffusive modeling of protein-nucleic acid assemblies. Extension of RF to incorporate ligands should similarly enable diffusive generation of protein-small molecule complexes. The ability to customize RF*diffusion* to specific design challenges by addition of external potentials and by fine-tuning (as illustrated here for catalytic site scaffolding, binder-targeting and fold-specification) should enable protein design to achieve still higher levels of complexity moving forward, to approach and in some cases surpass what was achieved by natural evolution.

Figures

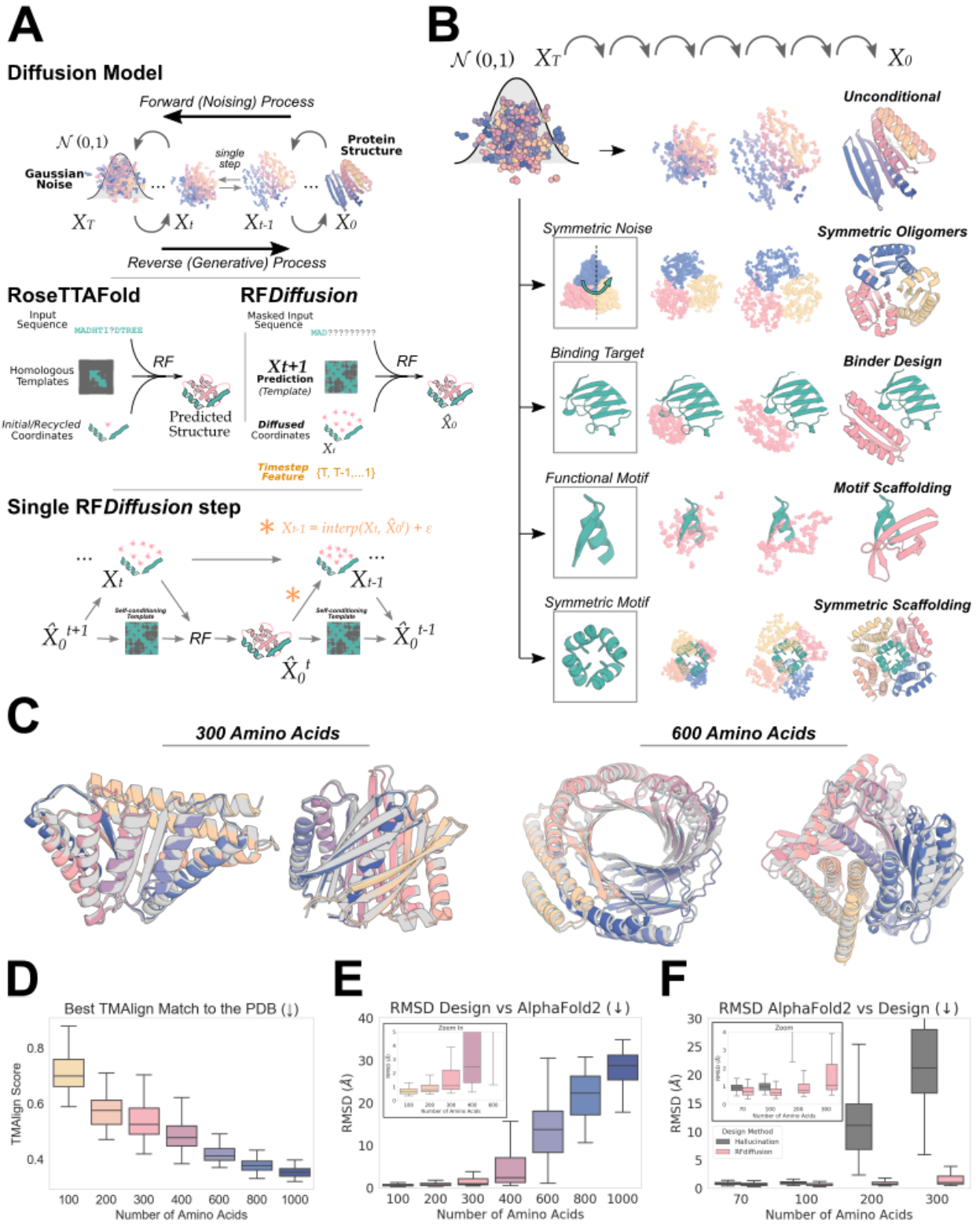

A
Diffusion Model
Forward (Noising) Process
Protein Structure
Gaussian Noise
single step
Reverse (Generative) Process
RoseTTAFold
RFDiffusion
Input Sequence
Homologous Templates
Initial/Recycled Coordinates
RF
Predicted Structure
Masked Input Sequence
Xt+1 Prediction (Template)
Diffused Coordinates
Timestep Feature
Single RFDiffusion step
B
Unconditional
Symmetric Noise
Symmetric Oligomers
Binding Target
Binder Design
Functional Motif
Motif Scaffolding
Symmetric Motif
Symmetric Scaffolding
C
300 Amino Acids
600 Amino Acids
D
Best TMAlign Match to the PDB (↓)
TMAlign Score
Number of Amino Acids
E
RMSD Design vs AlphaFold2 (↓)
RMSD (Å)
Number of Amino Acids
F
RMSD AlphaFold2 vs Design (↓)
Design Method
Hallucination
RFdiffusion
RMSD (Å)
Number of Amino Acids

Figure 1: Incorporating diffusion into RoseTTAFold addresses a broad range of protein design problems. A) Top panel: Diffusion models for protein design are typically trained to take a true protein structure with some degree of noise added to it, and to iteratively denoise this input back to the true protein structure. At inference time, the reverse (generative) process is employed, with random Gaussian coordinates input to the model, which iteratively refines this input until a final, novel protein sample is generated. Middle panel: Diffusion models can be incorporated into RoseTTAFold. RoseTTAFold (RF, left) is an SE(3)-equivariant protein structure prediction network that takes in sequence, template structures and initial coordinates, which it uses and refines to generate a protein structure prediction for those inputs . RF*diffusion* (right) is trained from a *pre-trained* RF network with minimal architectural changes. The input sequence is partially (or even fully) masked, and the template is derived from the model's previous prediction (self-conditioning, see methods). Finally, the input coordinates to the model are derived by "noising" the true protein structure to timestep "t" (following the process depicted in the top panel), with this timestep feature also provided to the model. The output from RF*diffusion*, just as in RF, is the prediction of the true protein structure (now denoted X_0). Bottom panel: RF*diffusion* can be used over many (typically 200) timesteps to iteratively refine random input coordinates to generate a novel designed protein. A single timestep of this iterative refinement is depicted here. At time "t", RF*diffusion* takes noised coordinates as input, along with a template representation of the model's X_0 prediction from the previous timestep (t+1), and the partially-masked true protein sequence (not depicted here). This template input allows RF*diffusion* to "self-condition" on its previous predictions, a concept introduced in this study. RF*diffusion* then predicts the X_0 structure, and the coordinate input to the model at the next time step (t-1) is generated by a noisy interpolation towards this X_0 prediction. These steps are repeated until a novel protein structure (X_0 at time=0) is generated. **B)** RF*diffusion* is of broad applicability to protein design, and can be employed in many design scenarios, described in this study. RF*diffusion* can generate protein samples without any conditioning information (top row), or can condition on: symmetric inputs to design symmetric oligomers (second row); a target structure to which it designs protein binders (third row); protein functional sites to which it designs scaffolds supporting their structure (fourth row); symmetric functional sites to design symmetric oligomers scaffolding these sites (bottom row). In each case, random gaussian noise, along with conditioning information, is input to RF*diffusion*, which iteratively refines that noise until a final protein is designed. **C)** RF*diffusion* can generate new monomeric proteins with no conditioning information. Two examples are shown here for two different lengths; 300 and 600 amino acids in length. Gray=design model; colors= AlphaFold2 (AF2) prediction. RMSD AF2 vs design (Å), left to right: 0.90, 0.98, 1.15, 1.67. **D)** Unconditional designs from RF*diffusion* are novel and not present in the training set. The closest match (highest TM score) to the protein databank (PDB) is plotted here. Designs are increasingly diverse with increasing length. **E)** Unconditional samples are closely re-predicted by AF2, an independent structure prediction network. Beyond approximately 400 amino acids, the recapitulation by AF2 deteriorates. **F)** RF*diffusion* significantly outperforms hallucination (with RoseTTAFold) at unconditional monomer generation. While hallucination is quite successful at generating designs up to 100 amino acids in length, beyond this length, in contrast to RF*diffusion*, success rates deteriorate significantly (as assessed by RMSD recapitulation by AF2).

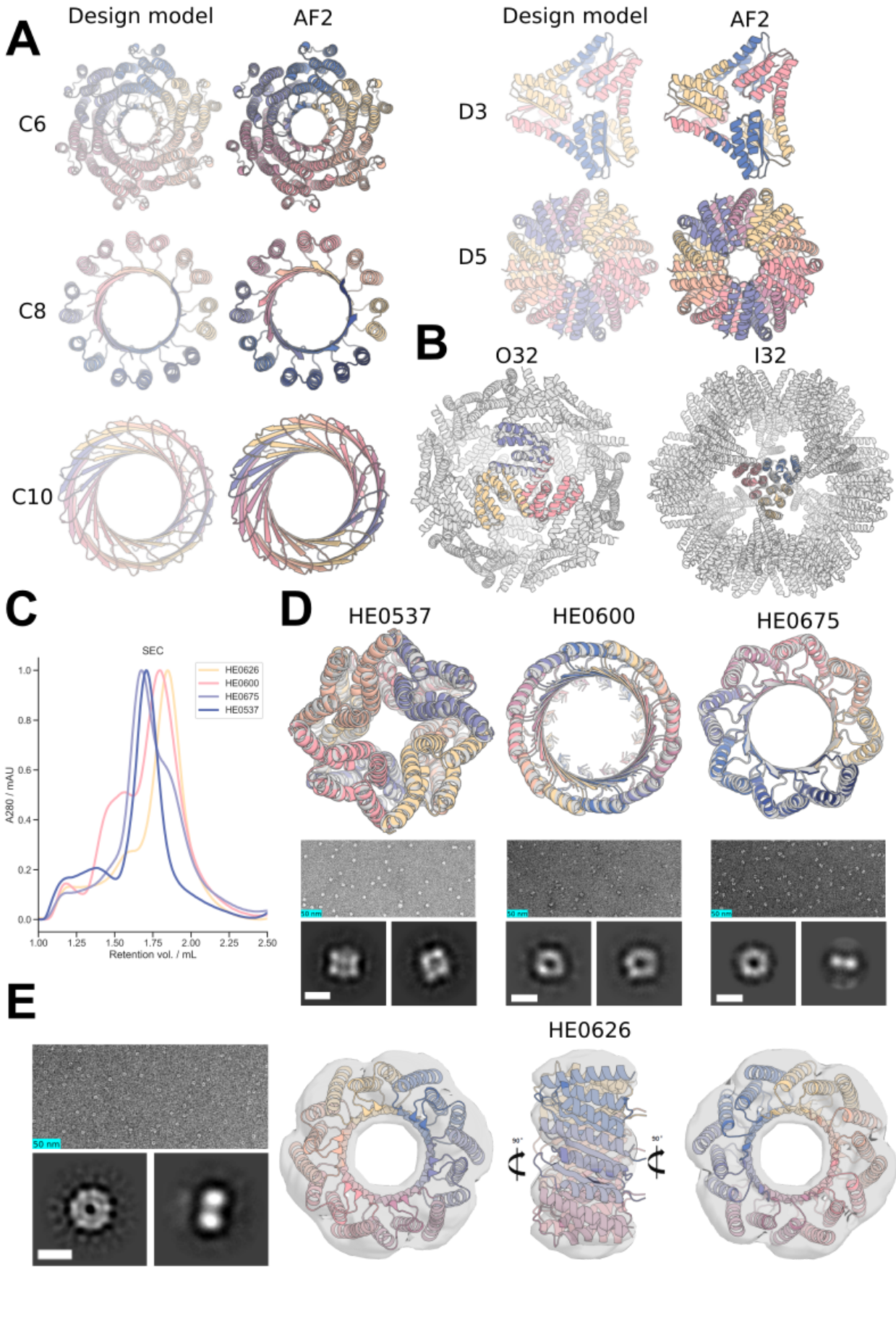

A
Design model
AF2
C6
C8
C10
Design model
AF2
D3
D5
B
O32
I32
C
SEC
HE0626
HE0600
HE0675
HE0537
A280 / mAU
Retention vol. / mL
D
HE0537
HE0600
HE0675
50 nm
E
50 nm
HE0626
90°
90°

Figure 2: Design and experimental validation of high-order symmetric oligomers. A) RFdiffusion generated cyclic and dihedral assemblies (left) compared to AF2 structure predictions based on the designed sequences (right); in all 5 cases they are nearly indistinguishable (backbone RMSDs vs AF2 for C6, C8, C10, D3, D5 are 1.04, 0.45, 0.60, 0.66, 0.72, respectively, with total amino acids 1200, 480, 600, 480, 1000, respectively). Symmetries are indicated to the left of the design models. **B)** Octohedral (left) and icosohedral (right) assemblies generated by RFdiffusion (gray). These structures are too large to be predicted by AF2 in their entirety; instead AF2 predictions for trimeric substructures are shown superimposed on the models (colors). **C)** Size exclusion chromatography of designs selected for experimental characterization, demonstrating major peaks at the expected elution volumes. **D)** Designed assemblies validated by single molecule electron microscopy averages. Top row: design models (gray) with superimposed AF2 predictions (color); as in A they are nearly indistinguishable (backbone RMSDs for HE0537, HE0600, HE0675 are 0.75, 0.88, 0.74, respectively) . Model symmetries from left to right are D4 (HE0537), C12 (HE0600), and C8 (HE0675). Middle row: representative fields of negative stained particles, demonstrating homogeneous samples. Bottom row: 2D averages of single particle images. The overall shapes are closely consistent with the design models. **E)** 3D reconstruction of design HE0627 (C6 symmetry) is closely consistent with the design model. Left: EM field and averages as in panel D; right: top, side, and bottom view of EM reconstructed density with superimposed design model. The density closely follows the contours of the design model.

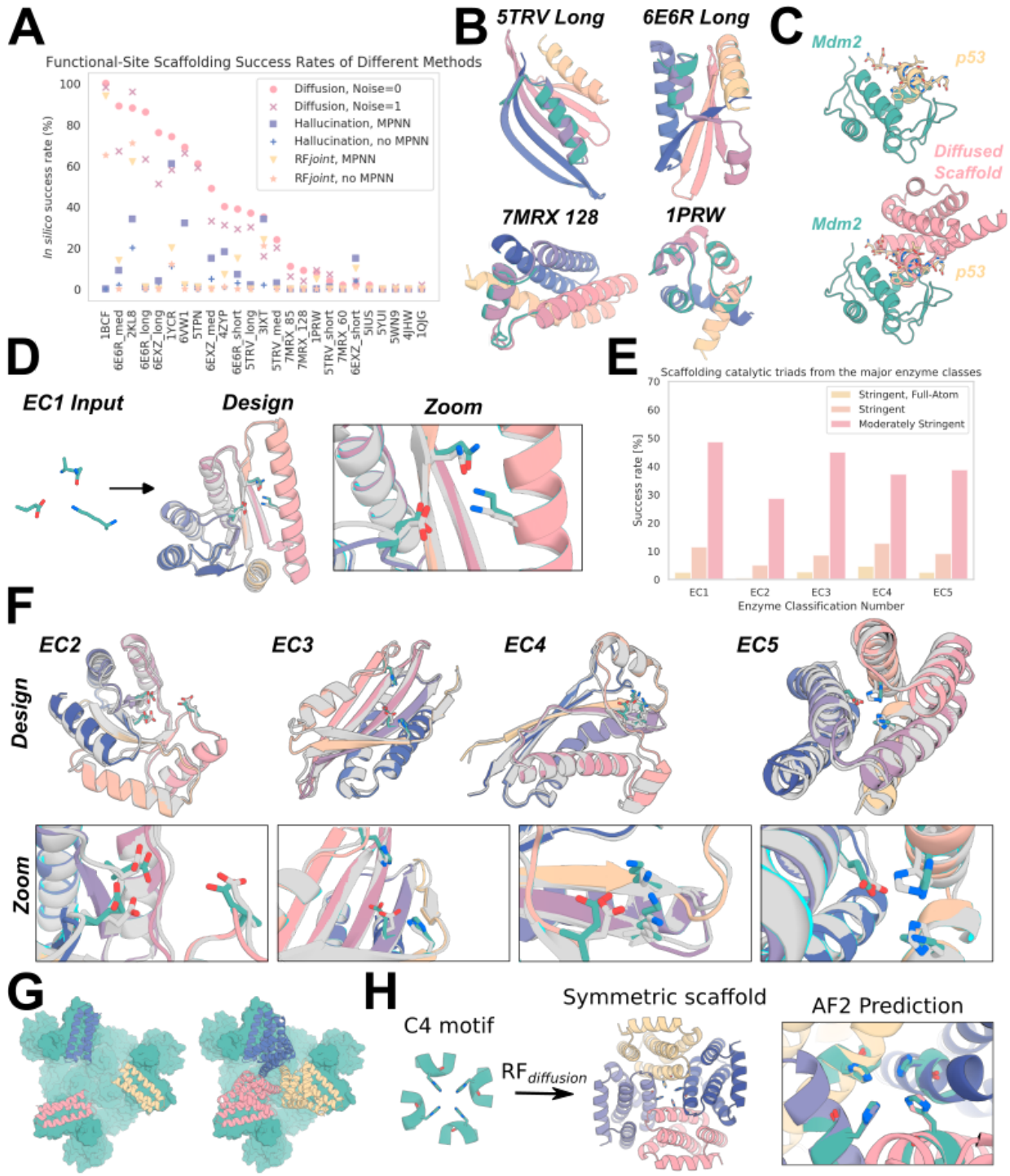
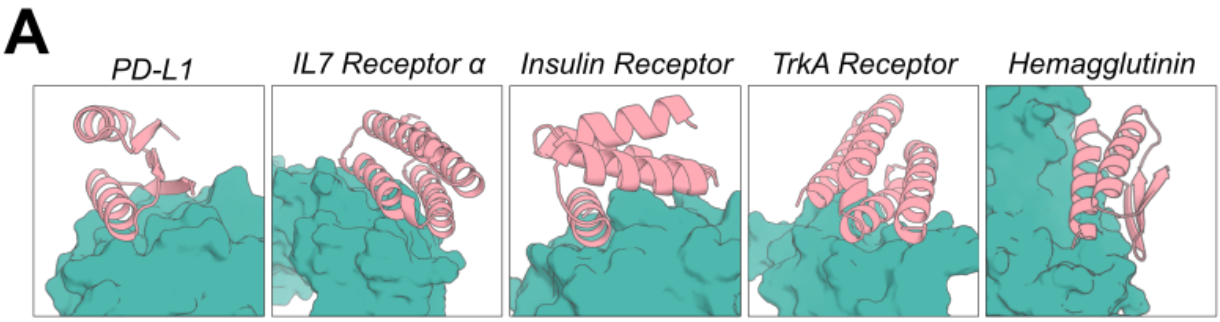


Figure 3 - Scaffolding of diverse functional-sites with RF*diffusion*. A) RF*diffusion* is state of the art across a diverse set of benchmark functional-site (“inpainting") problems. These 25 problems were collected from six recent papers, and encompass a broad range of functional sites, including enzyme active sites, binding interfaces and viral epitopes. Success was defined as AF2 RMSD to design model < 2 Å , AF2 RMSD to the native functional site (the “motif”) < 1 Å , and AF2 predicted alignment error (pAE) < 5, and the examples are ordered by success rate with RF*diffusion* (with noise scale = 0). 100 designs were generated per problem, with no

prior optimization on the benchmark set (some optimization was necessary and permitted for the hallucination data). RF*diffusion* solves 23/25 problems, and outperforms existing methods in 22/23 of these problems (RF*joint* and hallucination were similarly unable to generate solutions for the latter two problems). **B)** Four example designs for problems where RF*diffusion* significantly outperforms existing methods. Teal: native motif; colors: AF2 prediction of an RF*diffusion* design. Metrics (RMSD AF2 vs Design, RMSD AF2 vs native motif, AF2 pAE): 5TRV Long: 1.17 Å, 0.57 Å, 4.73; 6E6R Long: 0.89 Å, 0.27 Å, 4.56; 7MRX Long: 0.84 Å, 0.82 Å 4.32; 1PRW: 0.77 Å, 0.89 Å, 4.49. **C)** RF*diffusion* can scaffold the native p53 helix that binds to Mdm2, and can make additional contacts with the target. The designed scaffold (pink) is confidently predicted to interact with Mdm2 by AF2, and is predicted to scaffold the native p53 helix with atomic accuracy (Interaction pAE: 4.65, Monomer pAE: 4.93, AF2 Motif RMSD: 0.52 Å, AF2 vs design RMSD: 0.43 Å). *In silico* successful designs had, on average, 31% higher contacting surface area than the original helix. **D)** RF*diffusion* can be fine-tuned for specific and highly challenging design tasks, including the design of scaffolds supporting minimalist functional sites such as enzyme active sites. The input to RF*diffusion* is a few individual residues (left, in this case from the first enzyme class) and the network then designs a scaffold to these sites, which are often accurately repredicted by AF2 (middle and right, gray: design model; colors: AF2 prediction. Motif backbone RMSD 0.53 Å, Motif full-atom RMSD 1.05 Å, AF2 vs Design RMSD: 0.88 Å; AF2 pAE: 4.47). **E)** RF*diffusion* is able to scaffold a broad range of enzyme active sites from the five major enzyme classes. A random active site was selected from each of the five classes in the M-CSA database[41] designs generated with the fine-tuned RF*diffusion* model. Given the challenging nature of this problem, three degrees of stringency for success are reported: *Stringent, Full-Atom/Stringent, Backbone/Moderately Stringent*: AF2 vs design RMSD (backbone) < 2 Å/2 Å/3 Å; AF2 vs design Motif RMSD (backbone) < 1 Å/1 Å/1.5 Å, AF2 pAE < 5/5/7.5; AF2 vs design RMSD (full-atom) 1.5 Å/na/na. For all cases, designs were generated that passed our most stringent filters (EC1: 2.6%; EC2: 0.6%; EC3: 2.7%; EC4: 4.7%; EC5: 2.6%) **F)** An example (top row) and zoomed view (bottom row) of successful designs generated to the other four enzyme classes, demonstrating high-accuracy scaffolding of the active sites. Gray: design model, colors: AF2 model, Teal: motif structure prediction. Metrics (AF2 vs design backbone RMSD, AF2 vs design motif backbone RMSD, AF2 vs design motif full-atom RMSD, AF2 pAE): EC2: 0.93 Å, 0.50 Å, 1.29 Å, 3.51; EC3: 0.92 Å, 0.60 Å, 1.07 Å, 4.59; EC4: 0.93 Å, 0.80 Å, 1.03 Å, 4.41; EC5: 0.78 Å, 0.44 Å, 1.14 Å, 3.32. **G-H)** RF*diffusion* can scaffold symmetric functional sites. **G)** The SARS-CoV-2 spike protein is a C3-symmetric trimer. AHB2, a previously-described ACE2 mimic, can bind to a single spike protein subunit. To achieve higher affinity through multivalency, we symmetrized AHB2 around the C3 axis (left) and used RF*diffusion* to design a bespoke C3-symmetric oligomer to allow rigid scaffolding of the AHB2 interface in a position well-suited to interacting with all three spike subunits (right). Teal: SARS-CoV-2 structure (from PDB: 7JZL), colors: symmetrized AHB2 (left) and AF2 model of RF*diffusion* design (right). Metrics: AF2 pAE (monomer): 7.18; AF2 RMSD vs design (monomer/triple): 1.07 Å/1.28 Å; AF2 Motif RMSD (monomer/triple): 0.53 Å/2.36 Å. **H)** Nickel can be coordinated in a square-planar geometry. We generated C4 symmetric motifs scaffolding Histidine residues in positions ideal for coordinating nickel (left). RF*diffusion* generates C4 symmetric oligomers scaffolding these motifs.



B ***In Silico* Success Rate (%) [raw / +FastRelax]**

	Noise Scale	PD-L1	IL7 Receptor α	Insulin Receptor	TrkA Receptor	Hemagglutinin
Unconditional	0.0	39.3/47.6	9.1/18.8	**17.9**/17.9	14.7/12.1	4.2/3.9
	0.5	25.4/29.9	6.6/9.9	7.1/6.7	6.5/6.0	0.6/1.0
	1.0	10.7/14.4	2.1/4.4	2.8/3.2	1.5/1.6	0.1/0.1
Fold-Conditioned	0.0	44.0/46.0	11.7/**21.9**	10.6/10.8	**29.2**/24.2	**14.6**/13.9
	0.5	37.0/**49.0**	11.9/15.0	3.7/3.7	16.3/15.2	3.2/4.7
	1.0	22.0/26.0	5.3/10.7	2.5/3.0	6.6/7.2	0.2/2.3

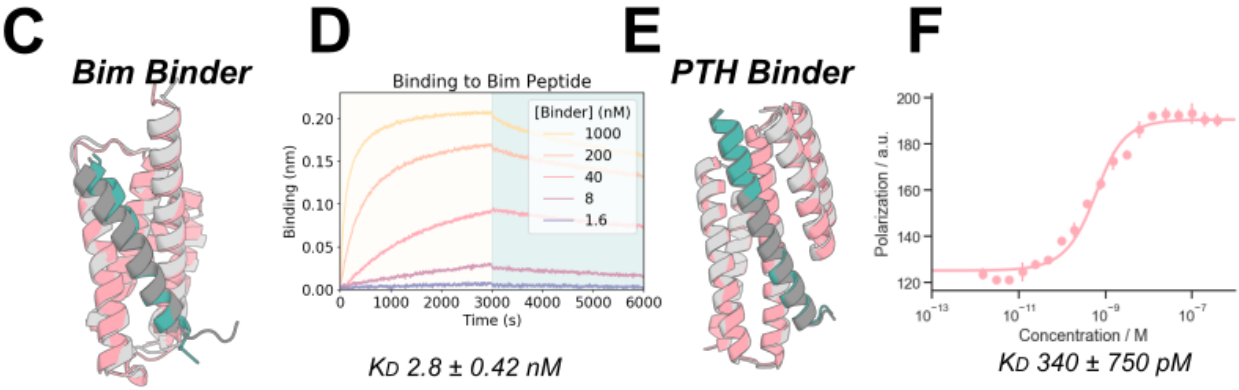


Figure 4: Design of protein and peptide binders. A-B) *De novo* binders were designed to five protein targets; PD-L1, IL7 Receptor α, Insulin Receptor, TrkA receptor and Influenza Hemagglutinin, and tested *in silico* with AF2 prediction. **A)** An example structure for each of the five targets, highlighting the diversity and complementarity of designs to their respective targets. AF2 models are shown (teal: target, pink: design). Metrics (Monomer pLDDT, Interaction pAE, Monomer RMSD AF2 vs Design): PD-L1: 87.9, 4.35, 0.56 Å ; IL7-Rα: 94.9, 7.33, 0.23 Å ; Insulin: 94.0, 4.84, 0.37 Å ; TrkA Receptor: 95.3, 4.62, 0.37 Å ; Hemagglutinin: 91.9, 9.20, 0.71 Å . **B)** Full *in silico* success rates for the protein binders designed to five targets. In each case, the best fold-conditioned results are shown (i.e. from the most target-compatible input fold), and the success rates at each noise scale are shown. In line with current best practice[21], we tested using Rosetta FastRelax[42] before designing the sequence with ProteinMPNN, but found that this did not systematically improve designs. Success is defined in line with current best practices[21]: AF2 pLDDT of the monomer > 80, AF2 interaction pAE < 10, AF2 RMSD monomer vs design < 1 Å . **C-F)** RF*diffusion* can design binders to helical peptides. **C)** Design model (gray) and AF2 prediction (colors) of an experimentally validated binder to the apoptosis-related peptide Bim. Teal: Bim peptide, Pink: designed binder. Metrics: RMSD AF2 vs Design: 1.14 Å ; interaction pAE: 4.18; Binder pLDDT: 94.0. **D)** The designed binder bound with 2.8nM affinity to Bim, as

measured by bio-layer interferometry. **E)** Design model (gray) and AF2 prediction (colors) of an experimentally validated binder to the helical peptide parathyroid hormone (PTH). Teal: PTH peptide, Pink: designed binder. Metrics: RMSD AF2 vs Design: 0.78 Å ; interaction pAE: 4.40; Binder pLDDT: 94.3. **F)** The design bound PTH with an exceptionally high affinity of 340pM, as measured by Fluorescence Polarization with TAMRA-labeled PTH peptide.

Supplementary Figures

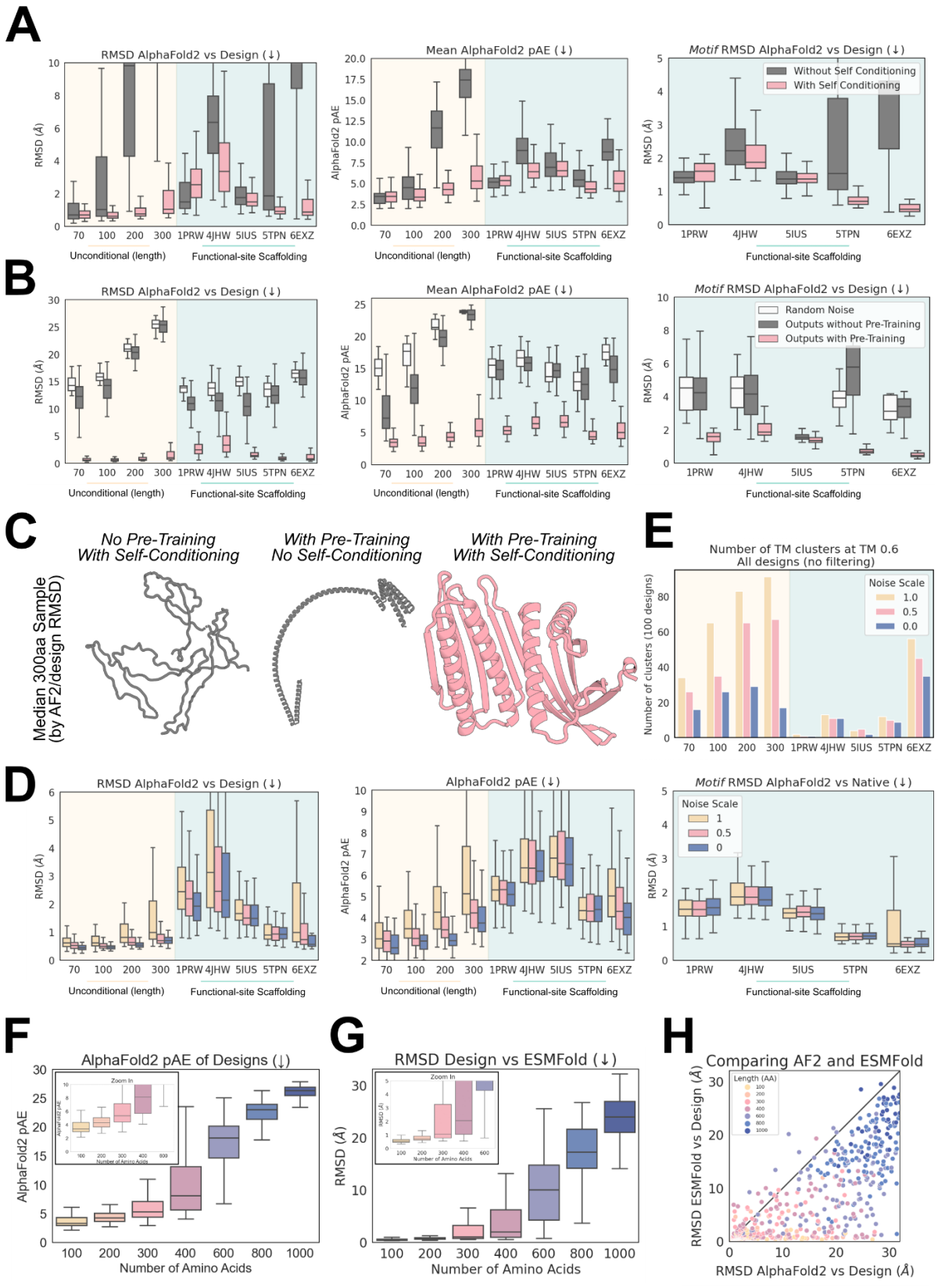


Figure S1: Self-conditioning and RF pre-training dramatically improve RF*diffusion* performance. A) Allowing the model to condition on its X_0 prediction at the previous timestep (see methods) improves designs. Designs with self-conditioning (pink) improves the recapitulation of designs by AF2 (left), the AF2 confidence in the prediction (middle) and the AF2 RMSD to the native motif (right). **B)** RF*diffusion* leverages the protein representations learned during RF pre-training. RF*diffusion* fine-tuned from pre-trained RF (pink) comprehensively outperforms a model trained for an equivalent amount of time, from untrained weights (gray). Indeed, training RF*diffusion* without pre-training showed minimal improvement as compared to generating ProteinMPNN sequences from random Gaussian-sampled coordinates (white). **C)** The median (by AF2 RMSD vs design) 300 amino acid unconditional sample highlighting the importance of self-conditioning and pre-training. Without pre-training, RF*diffusion* outputs bear little resemblance to proteins (gray, left). Without self-conditioning, outputs show characteristic protein secondary structures, but lack core-packing and ideality (gray, middle). With pre-training and self-conditioning, proteins are diverse and well-packed (pink, right). **D-E)** During the reverse (generation) process, the noise added at each step can be scaled (reduced). Reducing the noise scale comprehensively improves the quality of designs (particularly by AF2 RMSD vs design and AF2 pAE, **D)** left and middle). As expected however, this comes at the expense of diversity, with the number of unique clusters at a TM score cutoff of 0.6 reduced by reducing the noise (**E)**. **F-H)** RF*diffusion* (without reducing the added noise) can generate high quality large unconditional monomers. Designs are routinely accurately recapitulated by AF2 (see also Fig. 1E), with high confidence (**F**) for proteins up to approximately 400 amino acids in length. **G)** Further orthogonal validation of designs by ESMFold demonstrates the quality of unconditional RF*diffusion* designs. **H)** Recapitulation of the design structure is generally better with ESMFold compared with AF2. For each backbone, the best of 8 ProteinMPNN sequences is plotted here, so points are paired by backbone rather than sequence.

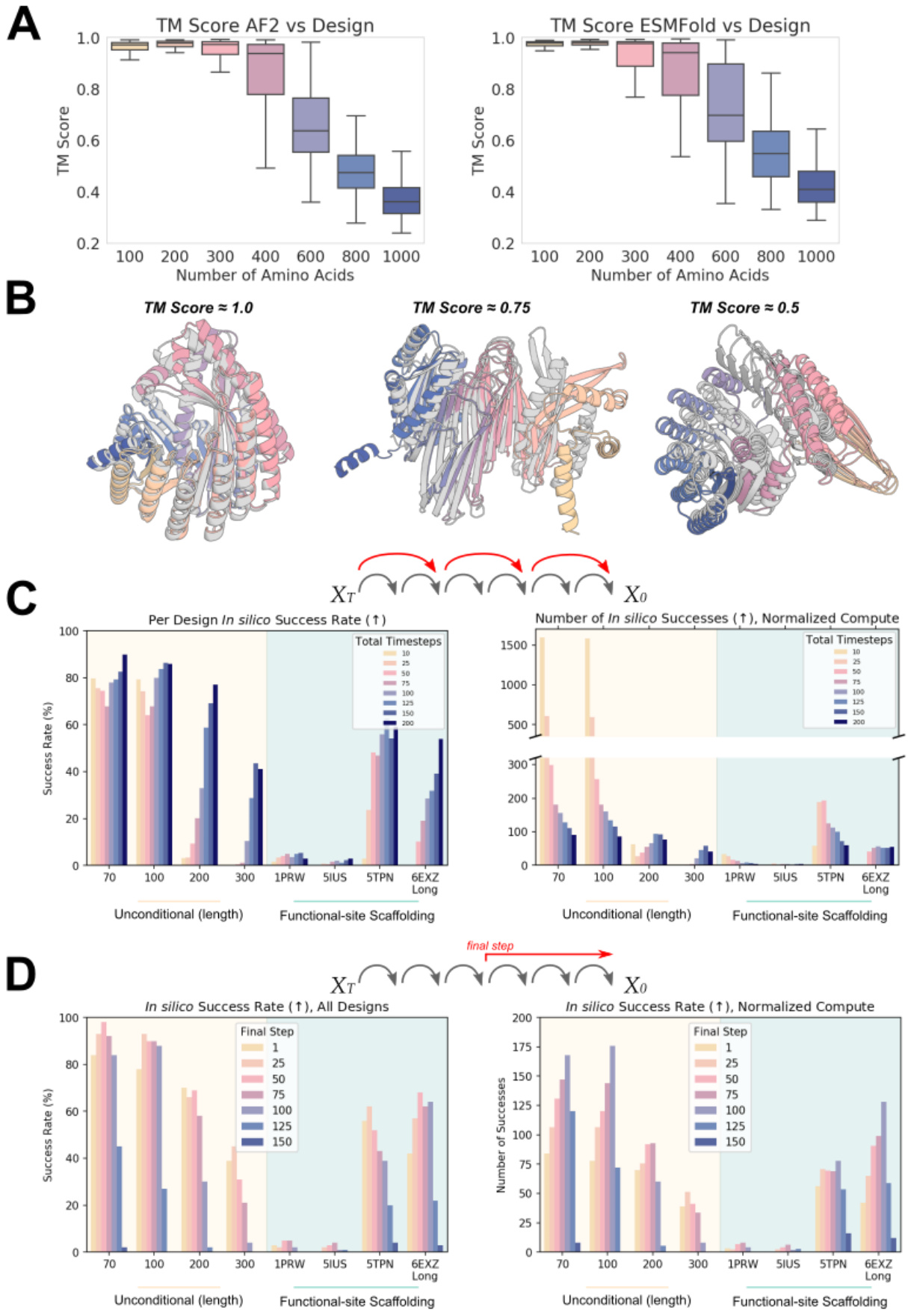

A
TM Score AF2 vs Design
TM Score ESMFold vs Design
TM Score
Number of Amino Acids
100 200 300 400 600 800 1000
B
TM Score ≈ 1.0
TM Score ≈ 0.75
TM Score ≈ 0.5
C
X_T
X_0
Per Design In silico Success Rate (↑)
Number of In silico Successes (↑), Normalized Compute
Total Timesteps
Success Rate (%)
70 100 200 300 1PRW 5IUS 5TPN 6EXZ Long
Unconditional (length)
Functional-site Scaffolding
D
final step
In silico Success Rate (↑), All Designs
In silico Success Rate (↑), Normalized Compute
Final Step
Number of Successes

Figure S2: Optimizing inference and improving metrics for *in silico* success. A-B) TM score between a design and a subsequent orthogonal prediction (e.g. AF2), has been previously used, typically with a threshold of > 0.5, as a metric for design success. **A)** By TM score, RF*diffusion* has high scores to both the AF2 (left) and ESMFold (right) predictions of the unconditional structures, with TM > 0.5 for a significant fraction of designs even up to 1000 amino acids in length. **B)** TM score is, however, much less stringent than RMSD alignment. Depicted here are three unconditional RF*diffusion* designs of 600 amino acids in length (gray), overlaid with the AF2 prediction (colors), with TM scores of 0.983, 0.757 and 0.506 respectively. While a TM score of 0.5 clearly shows some resemblance to the designed structure, it differs significantly and should not be classed as “successfully designed”. RMSD with a strict threshold (for example, 2 Å) is significantly more stringent. RMSDs for the displayed designs are 1.15 Å , 9.78 Å and 21.4 Å respectively. **C-D)** While RF*diffusion* is trained to generate samples over 200 timesteps, in many cases, trajectories can be shortened to improve computational efficiency. **C)** Bigger steps can be taken between timesteps at inference. While decreasing the number of timesteps typically reduces the per-design success rate (left plot), when normalized for compute budget (right plot), it is often more efficient to run more trajectories with fewer timesteps. For example, while generating 100 amino acid unconditional proteins, using a schedule with just 10 timesteps (as opposed to 200) allows the generation of 1584 *in silico* successful designs in the time taken to generate 86 successful designs with 200 timesteps. As problems get more challenging, however, this no longer remains the case (for example, fourth column, with generation of 300 amino acid designs). **D)** An alternative to taking larger steps is to stop trajectories early (possible because RF*diffusion* predicts X_0 at every timestep). In many cases, trajectories can be stopped at timestep 50-75 with little effect on the final success rate of designs (left plot), and when normalized by compute budget (right plot), success rates per unit time are typically higher generating more designs with early-stopping. For example, in the *6EXZ_Long* benchmarking motif-scaffolding problem, stopping trajectories at t=100 allows the generation of 128 *in silico* successful designs in the time it takes to generate 42 successful designs running full trajectories.

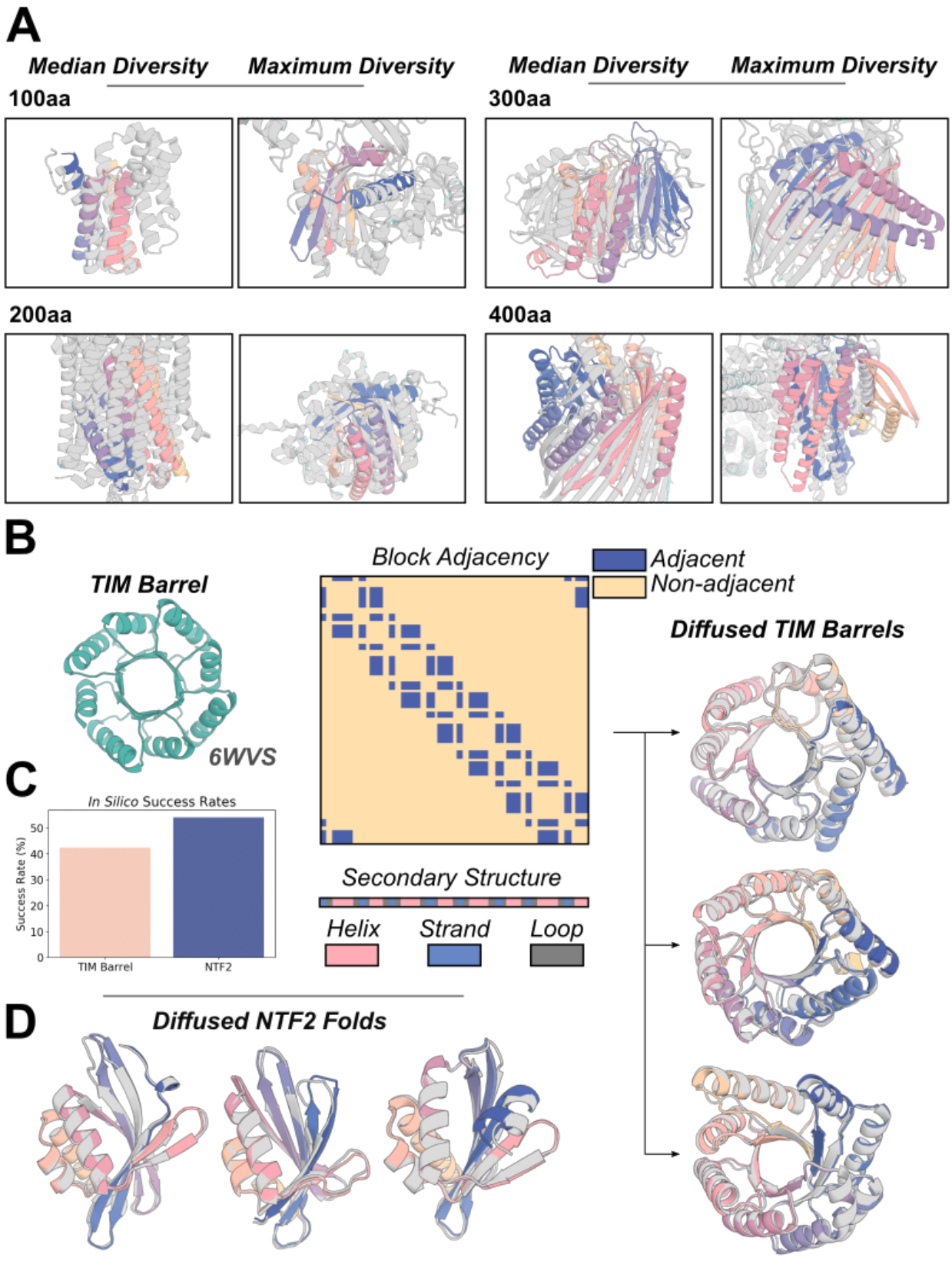

A
Median Diversity
Maximum Diversity
Median Diversity
Maximum Diversity
100aa
300aa
200aa
400aa
B
TIM Barrel
6WVS
Block Adjacency
Adjacent
Non-adjacent
Diffused TIM Barrels
C
In Silico Success Rates
Success Rate (%)
0
10
20
30
40
50
TIM Barrel
NTF2
Secondary Structure
Helix
Strand
Loop
D
Diffused NTF2 Folds

Figure S3: RF*diffusion* designs are novel without conditional information, or can be conditioned to generate specific folds. A) Example designs demonstrating extrapolation beyond the training set for generating novel folds. Gray: closest protein in the PDB by TM score, colors: RF*diffusion* design model, overlaid by TM alignment. For each protein length, the median and most diverse samples are shown. While for short proteins, designs typically show some similarity to known protein folds, with increasing length, designs become increasingly dissimilar to the PDB. TM score (closest PDB, TM score; median, most diverse): 100aa: 5WVE_A, 0.71; 4W5T_A, 0.59; 200aa: 4AV3_A, 0.58; 4CLY_A, 0.47; 300aa: 4PEW_B, 0.53; 4RDR_A, 0.46; 400aa: 4AIP_A, 0.49; 6R9T_A, 0.42. **B-D)** Designs can also be generated by conditioning on protein fold information. **B)** 6WVS is a previously-described *de novo* designed TIM barrel (left). A fine-tuned RF*diffusion* model can condition on 1D and 2D inputs representing this protein fold, specifically secondary structure (middle, bottom) and block adjacency information (middle, top, see methods). RF*diffusion* readily conditions on this information and generates a diverse set of TIM barrels (right). Gray: RF*diffusion* design, colors: AF2 prediction. **C)** TIM barrels are generated with an *in silico* success rate of 42.5% (left bar). Success incorporates AF2 metrics and a TM score vs 6WVS > 0.5. **C-D)** NTF2 folds are useful scaffolds for *de novo* enzyme design, and can also be readily generated with fold-conditioned RF*diffusion*. Designs are diverse (**D**) and designed with an *in silico* success rate of 54.1% (**C**, right bar). NTF2 fold design success also included both AF2 metrics and a TM score vs PDB: 1GY6 > 0.5. Gray: RF*diffusion* design, colors: AF2 prediction.

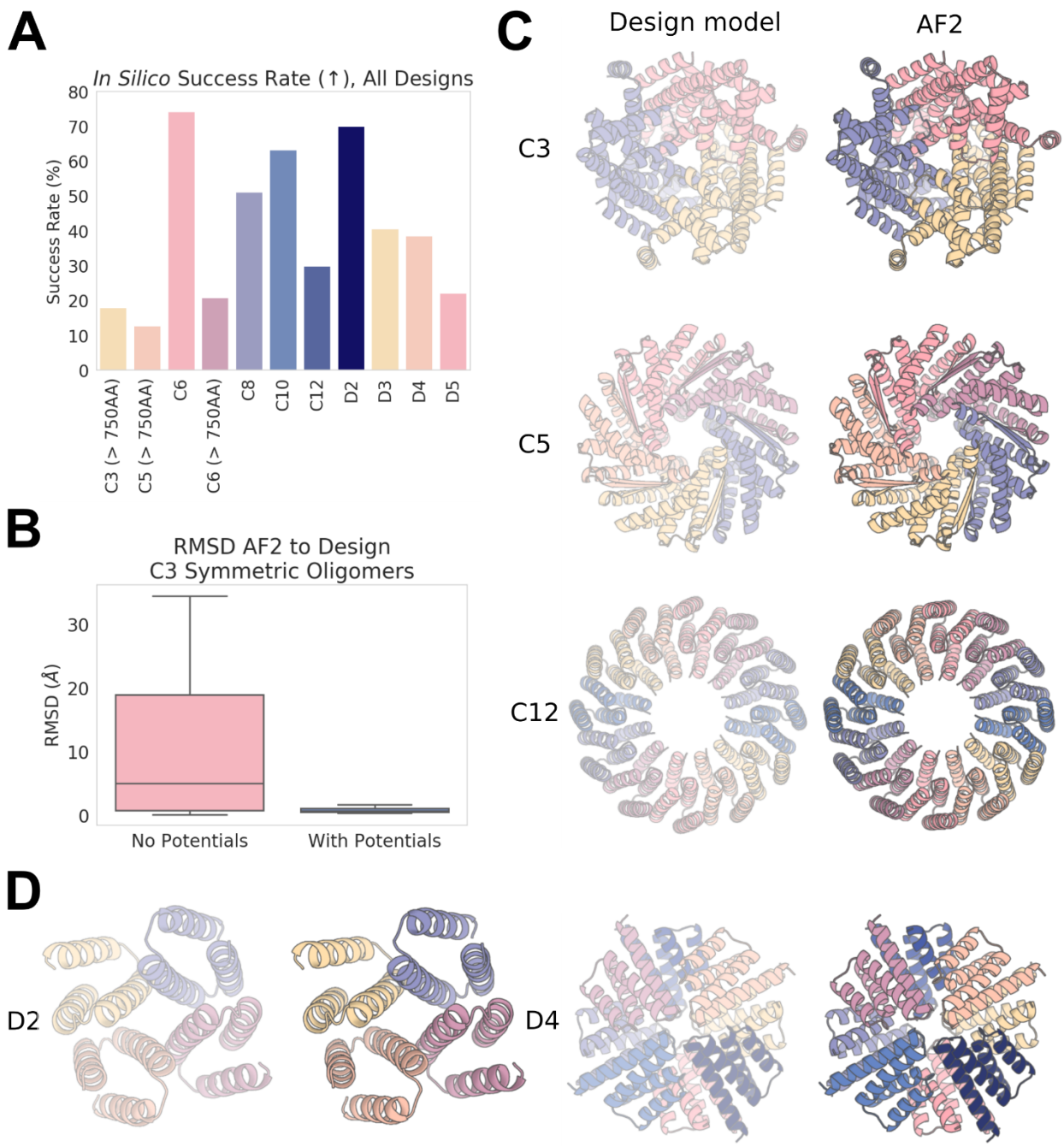


Figure S4: Symmetric oligomer design with RF*diffusion*. A) *In silico* success rates for symmetric oligomer designs of various cyclic and dihedral symmetries. Success is defined here as the proportion of designs for which AF2 yields a prediction from a single sequence that has mean pLDDT > 80 and backbone RMSD over the oligomer between the design model and AF2 < 2 Å. Note that 16 sequences per RF*diffusion* design were sampled. **B)** Box plots of the distribution of backbone RMSDs between AF2 and the RF*diffusion* design model with and without the use of external potentials during the trajectory. The external potentials used are the "inter-chain" contact potential (pushing chains together), as well as the "intra-chain" contact potential (making chains more globular). Using these potentials dramatically improves *in silico*

success. **C)** Additional examples of design models (left) against AF2 predictions (right) for C3, C5, C12, D2, and D4 symmetric designs with backbone RMSDs against their AF2 predictions of 0.82, 0.63, 0.79, 0.43, 0.78 with total amino acids 750, 900, 960, 240, 640.

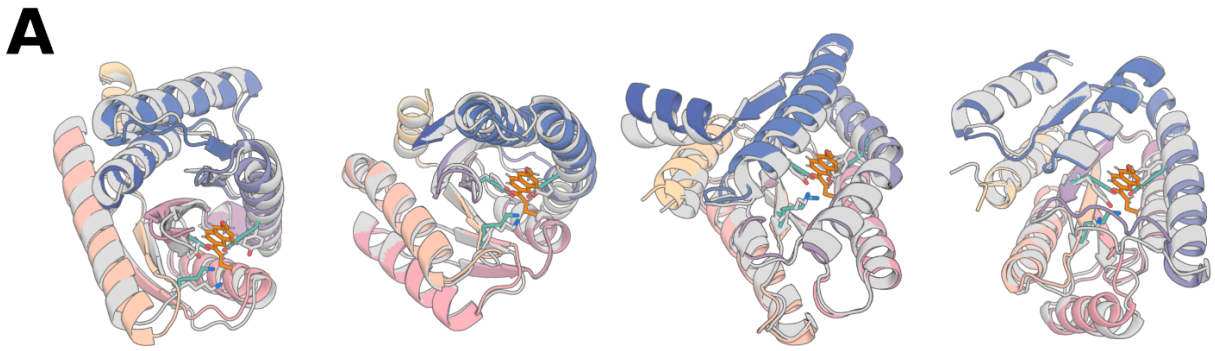


Figure S5: External potentials for generating pockets around substrate molecules. A) Enzymes generated from the triadic active site [TYR1051-LYS1083-TYR1180] of a retro-aldolase: PDB:5AN7. All designs shown here have AF2 RMSD to the native motif backbone < 1 Å, AF2 RMSD to the design model < 2 Å, and AF2 pAE < 5. The functional form and parameters used for the pocket potential are discussed in Methods section N4. In each case, the substrate is superimposed on the AF2 prediction of the catalytic triad. In all cases depicted here, RF*diffusion* designs pockets around the (implicitly-modeled) substrate.

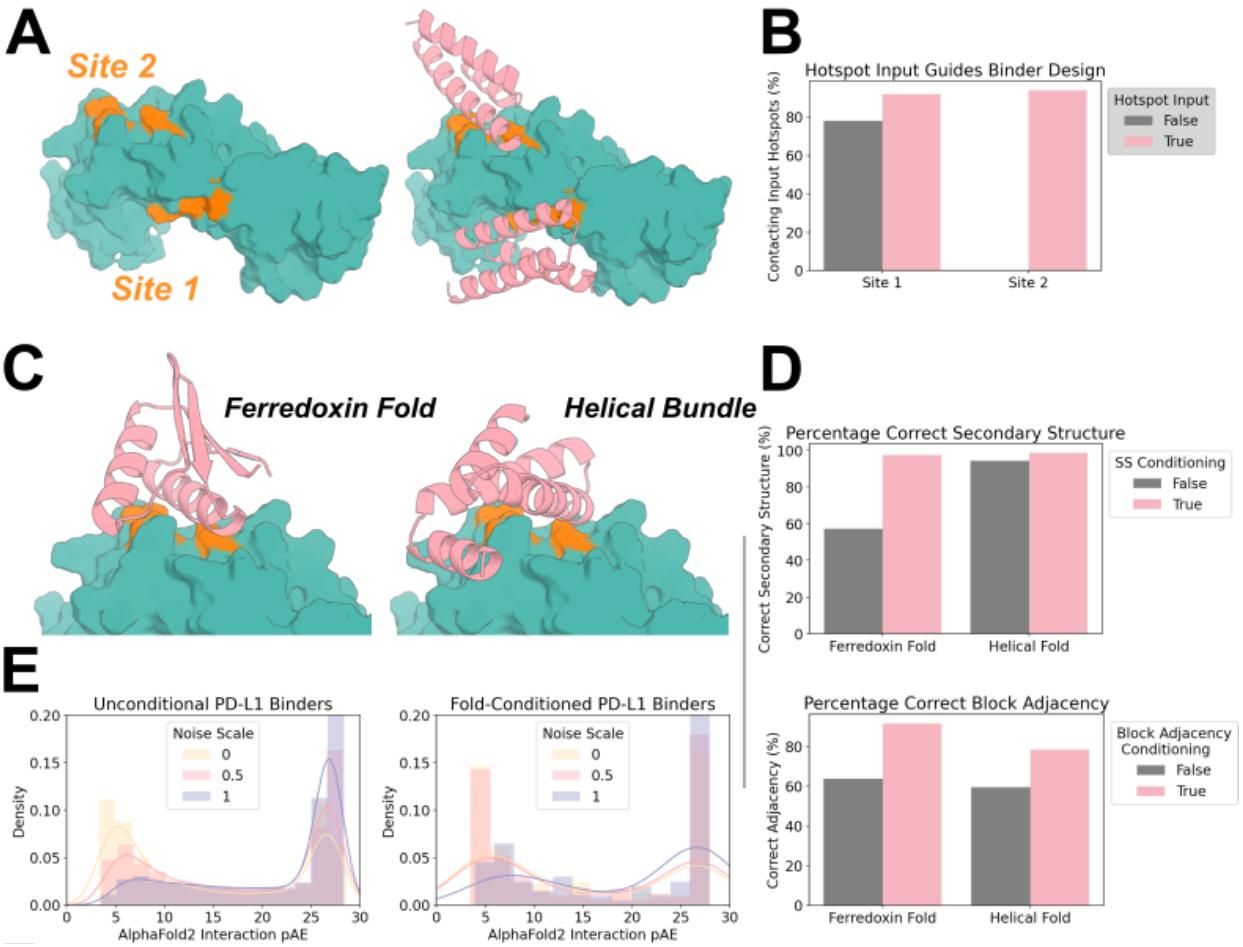


Figure S6: Targeted unconditional and fold-conditioned protein binder design. A-B) The ability to specify where on a target a designed binder should bind is crucial. Specific "hotspot" residues can be input to a fine-tuned RF*diffusion* model, and with these inputs, binders almost universally target the correct site. **A)** IL7-Rα (PDB: 3DI3) has two patches that are optimal for binding, denoted Site 1 and Site 2 here. For each site, 100 designs were generated (without fold-specification). **B)** Without guidance, designs typically target Site 1 (left bar, gray), with contact defined as Cα-Cα distance between binder and hotspot reside < 10 Å. Specifying Site 1 hotspot residues increases further the efficiency with which Site 1 is targeted (left bar, pink). In contrast, specifying the Site 2 hotspot residues can completely redirect RF*diffusion*, allowing it to efficiently target this site (right bar, pink). **C-D)** As well as conditioning on hotspot residue information, a fine-tuned RF*diffusion* model can also condition on input fold information (secondary structure and block-adjacency information - see methods). This effectively allows the specification of a (for instance, particularly compatible) fold that the binder should adopt. **C)** Two examples showing binders can be specified to adopt either a ferredoxin fold (left) or a particular helical bundle fold (right). **D)** Quantification of the efficiency of fold-conditioning. Secondary structure inputs was accurately respected (top, pink). Note that in this design target and target site, RF*diffusion* without fold-specification made generally helical designs (right, gray bar). Block adjacency inputs were also respected for both input folds (bottom, pink). **E)** Reducing the noise

added at each step of inference improves the quality of binders designed with RF*diffusion*, both with and without fold-conditioning. As an example, the distribution of AF2 interaction pAEs (known to indicate binding when pAE < 10) is shown for binders designed to PD-L1. In both cases, the proportion of designs with interaction pAE < 10 is high (blue curve), and improved when the noise is scaled by a factor 0.5 (pink curve) or 0 (yellow curve).

Name, Reference	Description	Input	Total Length	Sequence to be redesigned*
1PRW[1]	Double EF-hand motif	5-20,**A16-35**,10-25,**A52-71**,5-20	60-105	A16-19,A21,A23,A25,A27-30,A32-35,A52-55,A57,A59,A61,A63-66,A68-71
1BCF[1]	Di-iron binding motif	8-15,**A92-99**,16-30,**A123-130**,16-30,**A47-54**,16-30,**A18-25**,8-15	96-152	A19-25,A47-50,A52-53,A92-93,A95-99,A123-126,A128-129
5TPN[1]	RSV F-protein Site V	10-40,**A163-181**,10-40	50-75	A163-168,A170-171,A179,A189
5IUS[1]	PD-L1 binding interface on PD-1	0-30,**A119-140**,15-40,**A63-82**,0-30	57-142	A63,A65,A67,A69,A71,A72,A76,A79,A80,A82,A119,A120,A121,A122,A123,A125,A127,A129,A130,A131,A133,A135,A137,A138,A140
3IXT[31]	RSV F-protein Site II	10-40,**P254-277**,10-40	50-75	P255,P258-259,P262-263,P268,P271-272,P275-276
5YUI[1]	Carbonic anhydrase active site	5-30,**A93-97**,5-20,**A118-120**,10-35,**A198-200**,10-30	50-100	A93,A95,A97,A118,A120
1QJG[1]	Delta5-3-ketosteroid isomerase active site	10-20,**A38**,15-30,**A14**,15-30,**A99**,10-20	53-103	n/a
1YCR[1]	P53 helix that binds to Mdm2	10-40,**B19-27**,10-40	40-100	B17-18,B20-22,B24-25
2KL8[1,20]	*De novo* designed protein	**A1-7**,20,**A28-79**	79	n/a
7MRX_60[20]	Barnase ribonuclease inhibitor	0-38,**B25-46**,0-38	60	n/a
7MRX_85[20]		0-63,**B25-46**,0-63	85	n/a
7MRX_128[20]		0-122,**B25-46**,0-122	128	n/a
4JHW[30]	RSV F-protein Site 0	10-25,**F196-212**,15-30,**F63-69**,10-25	60-90	F196,F198,F203,F211-212,F63,F69
4ZYP[30]	RSV F-protein Site 4	10-40,**A422-436**,10-40	30-50	A422-427,A430-431,A433-436
5WN9[31]	RSV G-protein 2D10 site	10-40,**A170-189**,10-40	35-50	A170-175,A188-189
6VW1[1,32]	ACE2 interface binding SARS-CoV-2	***E400-510***/20-30,**A24-42**,4-10,**A64-82**,0-5†	62-83	A25-26,A29-30,A32-34,A36-42,A64-82
5TRV_short[2]	*De novo* designed protein	0-35,**A45-65**,0-35	56	n/a
5TRV_med[2]		0-65,**A45-65**,0-65	86	n/a
5TRV_long[2]		0-95,**A45-65**,0-95	116	n/a
6E6R_short[2]	Ferridoxin Protein	0-35,**A23-35**,0-35	48	n/a
6E6R_med[2]		0-65,**A23-35**,0-65	78	n/a
6E6R_long[2]		0-95,**A23-35**,0-95	108	n/a
6EXZ_short[2]	RNA export factor	0-35,**A28-42**,0-35	50	n/a
6EXZ_med[2]		0-65,**A28-42**,0-65	80	n/a
6EXZ_long[2]		0-95,**A28-42**,0-95	110	n/a

Table 1: A benchmarking set of recently published functional-site scaffolding problems. To benchmark RF*diffusion* at functional-site scaffolding, against existing methods, we generated a benchmark set encompassing problems described in six recent publications[1,2,20,30–32], which utilize a range of design methodologies to address these problems. For each problem, named by PDB accession (and, where applicable, the length of the designs to be generated, left column), we recapitulated the inputs as closely as possible with respect to details available in each publication. So that others can test methods on this benchmark, the exact input is specified in the third column. In bold, prefixed by a letter, are the inputs (chain, residues) from the PDB structure provided to the model (the "functional-site"). In non-bold text are the lengths that the different methods randomly sampled to generate good designs. The final lengths of the proteins were either specified by the input to the model, or were provided as constraints (for example, for 6EXZ_Long, the model could sample any N- and C-terminal length between 0 and 95 residues, but the total length of the output had to equal 110 amino acids). For each design challenge, 100 designs were generated, and, where ProteinMPNN was used, 8 sequences were designed, with the best sequence chosen for each backbone. *Both the RF*joint* and RoseTTAFold constrained hallucination approaches can simultaneously redesign sequences during generation, which can, in some cases, be helpful (if extracting the "functional-site" exposes hydrophobic residues which may subsequently end up as surface residues in the output designs, for example). Therefore, in this benchmark, these methods were allowed to redesign non-functional residues, listed in the right-most column. † This example is multi-chain generation (scaffolding a functional-site in the presence of a second chain). All methods benchmarked here can represent chain breaks (with large residue index jumps).

Materials & Methods

M. RF-diffusion as an SE(3) invariant generative model of protein structure:

In this section we describe in greater detail how we have repurposed RoseTTAFold (RF) as a generative model of protein structure.

Machine learning models for protein structure design must confront two major challenges to representing protein structures: (1) protein structure is most naturally represented by coordinates in a semantically arbitrary 3D coordinate system, yet (2) each amino acid which lives in this subspace has (effectively) two degrees of freedom (the ϕ and ψ backbone torsions) as opposed to the canonical six for a free rigid body. To navigate these challenges, most previous works on generative models of protein structure[24,43] have represented proteins as "maps" of pairwise distances between amino acids, followed by realizing chemically plausible 3D structures from these maps. However, given the remarkable representative power and accuracy of networks like AlphaFold2 (AF2) and RF which manipulate a "gas" of rigid bodies in 3D space in an SE(3) equivariant manner to produce a final 3D protein structure, we chose to formulate the protein generation task in a way that was compatible with this representation strategy. Moreover, design methods that directly parameterize structures in 3D are appealing for design because they allow specification of both rigid structural constraints such as the presence of functional motifs or existence of a desired symmetry by direct manipulations of structure[2,5].

We next give a brief overview of our diffusion modeling framework and how we have adapted it to protein structures in 3D. We then detail how we have applied it to the different components of our representation of structure. Lastly, we describe how we train conditional variants of the diffusion model for motif-scaffolding and generation with secondary structure constraints.

M1 Diffusion probabilistic modeling of protein structure

Our approach builds on denoising diffusion probabilistic models (DDPMs)[7,8]. We follow & adapt the conventions and notation set by [7], which we review here. DDPMs are a class of generative models based on a reversible, discrete-time diffusion process. The *forward process* starts with a sample $x_0 \sim q(x_0)$ from an unknown data distribution q. Noise is added at each step, to obtain a sequence of increasingly noisy samples x_t such that the final step $x_T \sim q(x_T)$ is indistinguishable from a *reference* distribution that has no dependence on the data. DDPMs approximate $q(x_0)$ with a second distribution $p_\theta(x_0)$ defined by transition distributions of the *reverse process* $p_\theta(x_{t-1} | x_t)$ at each t which are parameterized by a neural network. The neural network is trained such that $p_\theta(x_{t-1} | x_t)$ approximates $q(x_{t-1} | x_t)$. One then draws from $p_\theta(x_0)$ by first sampling from the reference distribution $x_T \sim p_\theta(x_T) \approx q(x_T)$, and then for each $t < T$ iteratively denoising by sampling $x_{t-1} \sim p_\theta(x_{t-1} | x_t)$ until $x_0 \sim p_\theta(x_0)$ is obtained.

In our case, we consider $q(x_0)$ to be a distribution over the structures of backbones of native proteins. We adopt the "residue-gas" representation of backbones used by both RF[44]. This representation consists of the 3D coordinates (z) of the central carbon (Cα) and 3x3 rotation matrices (r) representing the rigid-body orientation of each residue in a global reference frame, thereby additionally defining the coordinates of the N and C backbone atoms. $x = [z, r]$. We defined a forward process that applies noise independently both across residues and across these two components of residue geometry. We similarly model the reverse process transitions as independent across these components:

$$p_\theta(x_{t-1}|x_t) = p_\theta(z_{t-1}|x_t)p_\theta(r_{t-1}|x_t).$$

While $q(x_{t-1}|x_t)$ can in general be correlated across these different components of structure, standard practice has found it beneficial to ignore the correlation across dimensions in the reverse diffusion process[7]. Indeed, in the limiting regime where the number of steps in the forward process tends to infinity, and the forward process is viewed as a discretization of a continuous time diffusion process, one can see that the correlation between different dimensions is absent in the reverse process as well[45].

To address the challenge of the arbitrary reference frames we build on previous work[2], and seek to learn a distribution over protein structure that is invariant to global rotation; that is, we require that any protein structure is modeled as equally likely upon a rigid body rotation. More formally, this means that for any structure x_0 and rotation R we desire to have that $p_\theta(x_0) = p_\theta(R * x_0)$, where $R * x_0$ represents the structure obtained by rotating x_0 about the origin, $[0, 0, 0]$ by R (for each residue $R * x = [Rz, Rr]$).

Following others[2,46], we incorporate this invariance by (1) using a rotation invariant reference distribution ($p_\theta(x_T) = p_\theta(R * x_T)$) and (2) constraining the reverse diffusion model to be equivariant to rigid body rotations ($p_\theta(x_t|x_{t+1}) = p_\theta(R * x_t|R * x_{t+1})$). To this end, we leverage the geometric equivariance and invariance properties inherent to RoseTTAFold. In particular RF uses the SE(3)-transformer architecture[47] to provide equivariant updates to intermediate predictions of structure across recycling steps; we use these same input channels to obtain *equivariant* updates of Cα coordinates (M2) and rotations for each residue (M3).

M2 Training RosettaFold to denoise protein structures

Our approach to learning the reverse process transition is to train RoseTTAFold to denoise noisy protein structures. For each step of training, we first choose an example protein structure x_0 and a time step t uniformly at random between 1 and T, and then simulate the forward process to obtain $x_t \sim q(x_t|x_0)$. We next apply RoseTTAFold to obtain a prediction of the denoised structure, which we denote by $\hat{x}_\theta(x_t) = (\hat{z}_0, \hat{r}_0)$. We then compute a loss on this

output consisting of the squared Euclidean distance on the Cα coordinates and the square of a metric on the space of rotation matrices[48]. Algorithm 1 summarizes this procedure.

The approach above takes inspiration from Ho *et al*[7]. In particular, Ho *et al*[7] (section 3.2) comments that when the forward process consists of adding Gaussian noise, the training objective of minimizing the KL divergence of $q(x_t|x_{t-1})$ to $p_\theta(x_{t-1}|x_t)$ can be rewritten as a rescaling of the expected squared error of a prediction of x_0 from noisy observations x_t:

$$E_{x_0,x_t\sim q}[KL(q(x_t|x_{t-1})||p_\theta(x_{t-1}|x_t))] \propto E_{x_0,x_t\sim q}[||x_0 - \hat{x}_\theta(x_t)||^2] + c, \quad (1)$$

where c is a constant that does not depend on θ. Consequently when one minimizes the right-hand-side of equation (1), they maximize a weighted variational lower bound on the likelihood of the data[7] that is globally minimized only when each $p_\theta(x_{t-1}|x_t)$ matches $q(x_{t-1}|x_t)$, and $p_\theta(x_0)$ therefore matches the data-distribution. Although ref [7] found better performance in generative modeling of images when predicting the noise added in the forward process (rather than x_0), we reasoned that by predicting x_0 we could better leverage the inductive biases of RoseTTAFold pre-trained for structure prediction to produce realistic structures as in [1]. Additionally, although the Gaussian distribution and Euclidean distance are not well defined on the space of rotation matrices we reasoned that our approach was reasonable because noising process and metric on rotations we use are approximately Gaussian and Euclidean, respectively, and closely agrees at time steps t near zero [M4 for details]

In the following sections we show how to relate $\hat{x}_\theta(x_t)$ to approximations $p_\theta(x_{t-1}|x_t)$ of $q(x_{t-1}|x_t)$.

Algorithm 2 RF-Diffusion Sampling.

function SAMPLEREFERENCE(M)
  for $m = 1, \ldots, M$ **do**
    $z_T \sim \mathcal{N}(0, I_3)$
    $r_T \sim \text{Uniform}(SO(3))$
  end for
  $x_T = [(z_{T,1}, r_{T,1}), \ldots, (z_{T,M}, r_{T,M})]$
return x_T
end function

function SAMPLE(M)
  $x_T = \text{SampleReference}(M)$
  $\hat{x}_0 = \vec{0}$ ▷ Initialize self-conditioning
  for $t = T-1, \ldots, 0$ **do**
    $\hat{x}_0 = \text{RFDiffusion}_\theta(x_{t+1}, \hat{x}_0, t)$
    $x_t = \text{ReverseStep}(x_{t+1}, \hat{x}_0, t)$
  end for
return x_0
end function

function SAMPLESYMMETRIC($M, \mathfrak{R} = \{R_k\}_{k=1}^K$)
  $x_T^1 = \text{SampleReference}(M)$
  $\hat{X}_0 = [\vec{0}, \ldots, \vec{0}]$ ▷ Initialize self-conditioning
  for $t = T-1, \ldots, 0$ **do**
    $X_{t+1} = [R_1 * x_{t+1}^1, \ldots, R_K * x_{t+1}^1]$
    $\hat{X}_0 = \text{RFDiffusion}_\theta(X_{t+1}, \hat{X}_0, t)$
    $x_t = \text{ReverseStep}(X_{t+1}, \hat{X}_0, t)$
  end for
return $\hat{X}_0$
end function

M3 Details of forward and reverse diffusion of backbone residue translations:

In this subsection we describe our forward diffusion over backbone coordinates (z), and how we relate predictions $\hat{x}_\theta(x_t)$ of x_0 to our approximation $p_\theta(z_{t-1}|x_t)$ of $q(z_t|x_{t-1})$. Our development and notation follows ref [7]. We let $\beta_1, \beta_2, \cdots, \beta_T$ be scalars between 0 and 1 that define a variance schedule such that for each $t = 1, 2, \cdots, T$ the transition density of the forward process is $q(z_t \mid z_{t-1}) = N(z_t; \sqrt{1-\beta_t} z_{t-1}, \beta_t I_3)$. Define $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_t$. To sample z_t during training, rather than sampling $z_s|z_{s-1}$ from $s = 1$ all the way up to $s = t$, we draw z_t directly from its marginal distribution,

$$q(z_t|z_0) = N(z_t; \sqrt{\bar{\alpha}_t} z_0, (1 - \bar{\alpha}_t) I_3). \quad (2)$$

Given that $q(z_{t-1}|z_t, z_0) = N(z_{t-1}; \tilde{\mu}(z_t, z_0), \tilde{\beta}_t I_3)$

for $\tilde{\mu}(z_t, z_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} z_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t$ and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \approx \beta_t$, we choose to parameterize the reverse transitions by

$$p_\theta(z_t|x_{t-1}) = N(z_t; \mu_\theta(x_t), \beta_t I_3) \text{ for } \mu_\theta(x_t) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \hat{z}_\theta(x_t) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t,$$

where $\hat{z}_\theta(x_t)$ are the predicted Cα coordinates obtained from $\hat{x}_\theta(x_t)$. We choose β_t according to a linear variance schedule as in[24,49] with parameters $\beta_0 = 0.01$ and $\beta_T = 0.07$. We chose these parameters such that signal remaining in x_0 (as quantified by $\bar{\alpha}_t$) decayed slowly toward zero as t approaches T=200 (Fig. M1).

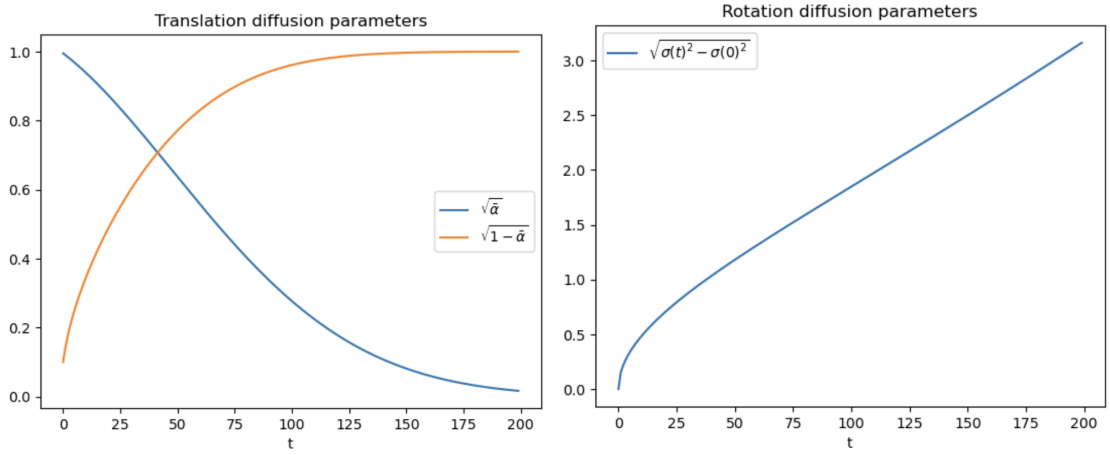


Figure M1: Noising schedules for translations and rotations associated with each residue.

M.4 Details of forward and reverse diffusion on backbone residue rotations

We model the remaining two backbone atoms (N and C) with a diffusion process on rigid body rotations that map an axis-aligned residue with idealized internal geometry (i.e. bond lengths and angle) to the positions of these atoms relative to central Cα. Specifically, for any backbone atom coordinates z_c, $z_{c\alpha}$ and z_N for any residue we may apply a Gram-Schmidt process to compute a 3x3 rotation matrix r with rows

$r[1] = (z_c - z_{c\alpha}) / ||z_c - z_{c\alpha}||$,

$r[2] = ((z_N - z_{c\alpha}) - (z_N - z_{c\alpha}) \cdot r[1])/||(z_N - z_{c\alpha}) - (z_N - z_{c\alpha}) \cdot r[1]||$, and

$r[3] = r[1] \times r[2]$.

where $\cdot$ and $\times$ are the dot and cross-products, respectively. 3D backbone coordinates can then be reconstructed by multiplication of idealized coordinates (with $z^*_{c\alpha}$ at the origin, $z^*_c - z^*_{c\alpha}$ along the x-axis, and $z^*_N - z^*_{c\alpha}$ in the xy-plane) by R:

$$[z_c, z_N, z_{c\alpha}] = R\,[z^*_c, z^*_N, z^*_{c\alpha}] + z_{c\alpha}\vec{1}_3 ,$$

where $\vec{1}_3 = [1, 1, 1]$.

So we can see that modeling the locations of backbone atoms is equivalent to modeling associated rotation matrices.

However, modeling rotation matrices introduces challenges not addressed by ref [7]; the space of 3x3 rotation matrices (known as the special orthogonal group of dimension 3, or SO(3)) is a compact Riemmanian manifold on which the typical Gaussian distribution is not well-defined and the so the associated techniques of [7] do not apply. To this end we adapt the approach of [14], who extend diffusion generative modeling to Riemannian manifolds. In brief, they build on the continuous-time diffusion framework[45] and define their forward diffusion as the Brownian motion on the manifold of interest. In the case of SO(3), this Brownian motion is described by the IGSO3 distribution.[15] In brief, the density of the IGSO3 distribution with respect to the uniform distribution on SO(3) is given by

$$IGSO3(r; \mu, \epsilon^2) = f(\omega(r\mu^\top)),\ for\ f(\omega) = \sum_{l=0}^{\infty} (2l + 1) e^{-l(l+1)\epsilon^2} \frac{\sin((l+\frac{1}{2})\omega)}{\sin(\omega/2)} , \quad (3)$$

where μ is 3x3 mean rotation matrix and $\omega(r)$ denotes the angle of rotation in radians associated with a rotation r (i.e. if r were written in the axis-angle parameterization). Notably, if a rotation matrix r_t evolves according to Brownian motion $dr_t = dB_t$ from initialization $r_0 = \mu$, then $r_t \sim IGSO3(r_t; \mu, 2t)$.

Accordingly, these dynamics motivate the following discrete forward noising process:

$q(r_t | r_{t-1}) = IGSO3(r_t; r_{t-1}, \epsilon^2 = 2(\sigma_t^2 - \sigma_{t-1}^2))$ and marginally

$q(r_t | r_0) = IGSO3(r_t; I_3, \epsilon^2 = 2\sigma_t^2)$

where σ_t^2 is a variance schedule for rotations [Figure M1]. In contrast to z_t, the forward process for r_t converges to the uniform distribution on $SO(3)$.

De Bortoli *et al*[14] (Theorem 1) prove that (up to error from discretization of the continuous time process) the reverse process transitions have the form,

$r_{t-1}|x_t \sim \exp\{\Delta r_t\} r_t$ for $\Delta r_t \sim IGSO3(\exp\{(\sigma_t^2 - \sigma_{t-1}^2) \cdot \nabla_{r_t} \log q(x_t)\}, 2(\sigma_t^2 - \sigma_{t-1}^2))$, (4)

where $\nabla_{r_t} \log q(x_t)$ denotes the "Stein score", that is the gradient with respect to r_t of the log density of the noised structure x_t according to the forward process at time t, and $\exp\{\cdot\}$ denotes the *exponential map* to SO(3) from the Lie algebra of SO(3) (the space in which the score is defined).

Equation 4 describes how one could sample from the reverse process using the IGSO(3) distribution based on the score of the forward process. One could in principle learn this score function directly by score matching training as described previously[14]. However, we instead rely on an approximation that directly leverages RoseTTAFold's ability to produce denoised structures when trained according to Algorithm 1. For a given t and r_t we note that we may write

$$\begin{aligned}
\nabla_{r_t} \log q(x_t) &= \mathbb{E}_q \left[\nabla_{r_t} \log q(x_t \mid x_0) \mid x_t \right] \\
&= \mathbb{E}_q \left[\nabla_{r_t} \log q(r_t \mid r_0) \mid x_t \right] \\
&\approx \nabla_{r_t} \log q(r_t \mid r_0 = \hat{r}_0) \\
&= \nabla_{r_t} \log \mathrm{IGSO3}(r_t; \hat{r}_0, 2\sigma_t^2),
\end{aligned} \quad (5)$$

where the first line is known as the denoising score matching identity[14,50], the second line obtains from the conditional independence structure of the forward process, the third line is an approximation that can be thought of as replacing $q(r_0|r_t)$ with a point mass on the noiseless rotation $\hat{r}_0$ predicted by RoseTTAFold, and the final line recognizes the approximation as the tractable IGSO3 density. In particular,

$\nabla_r \log IGSO3(r; \hat{r}, 2\epsilon^2) = r \log(r\hat{r}^T) / \omega(r\hat{r}^T) \cdot \frac{d}{d\omega} \log f(\omega, 2\epsilon^2)|_{\omega=\omega(r\hat{r}^T)}$ (6)

Where log is the logarithmic map from SO(3) to the Lie algebra of SO(3) [51], $\omega(r\hat{r}^T)$ is the angle of rotation associated with $r\hat{r}^T$,

and f is the IGSO3 density factor in equation (3). Notably, $r \log(r\hat{r}^T) / \omega(r\hat{r}^T)$ represents a rotation of unit length on the tangent space of SO(3) at r, and $\frac{d}{d\omega} \log f(\omega, 2\epsilon^2)|_{\omega=\omega(r\hat{r}^T)}$ is a scaling of this direction.

We reasoned that approximation in equation (5) may be reasonably accurate for two reasons. First, in the case of Gaussian DDPMs where optimizing to convergence would provide $\hat{z}_0(z_t) = E_q[z_0|x_t]$, this approximation holds exactly in the sense that $E_q[\nabla_{z_t} \log q(z_t|z_0)|x_t] = \nabla_{z_t} \log q(z_t|z_0 = \hat{z}_0(x_t))$ Though this does not hold with equality with the IGSO(3), because SO(3) is a Riemannian manifold and is therefore locally Euclidean the IGSO(3) closely resembles a Gaussian for small t. Second, again when t is near to zero, x_t will be close to an un-noised structure and, if the model is trained well, $q(r_0|x_t)$ will be concentrated around $\hat{r}_0$. Finally, we note that this rule has beneficial qualitative behavior -- as with the Gaussian score, the magnitude of $\nabla_{r_t} \log q(r_t|r_0 = \hat{r}_0(x_t))$ will grow roughly linearly with the distance between r_t and $\hat{r}_0$ (in terms of the geodesic distance on SO(3)). Consequently, this leads to larger steps when r_t is farther from $\hat{r}_0$.

In summary we approximate reverse transitions by $p_\theta(r_{t-1} \mid x_t) = IGSO3(\hat{r}_{t-1}; 2(\sigma_t^2 - \sigma_{t-1}^2))$, where $\hat{r}_{t-1} = \exp\{(\sigma_t^2 - \sigma_{t-1}^2) \nabla_{r_t} \log IGSO3(r_t; \hat{r}_0, 2\sigma_t^2)\} r_t$ with $\nabla_{r_t} \log IGSO3(r_t; \hat{r}_0, 2\sigma_t^2)$ computed as in Equation (6).

M.5 Self-Conditioning

Self-conditioning was introduced previously[19] where it was shown to dramatically improve text diffusion. Self-conditioning is closely related to Step Unrolled Denoising Autoencoders[52]; both methods perform a step of inference at training time to train the model to expect inputs of the same distribution that will be generated at inference time. We implement self-conditioning in the manner described in Chen *et al*[19], which we review here.

We wish to train a denoising function $f(x_t, \hat{x}_0, t)$. During training 50% of the time we train the model with $\hat{x}_0 = 0$. The other 50% of the time, the model first performs one step of inference to generate $\hat{x}_0 = f(x_t, 0, t)$, gradients are then turned on and the model is trained to estimate $f(x_t, \hat{x}_0, t)$. The training step with $\hat{x}_0 = 0$ is required during training so that the model can make accurate estimates of $\hat{x}_0 = f(x_t, 0, t)$ and provide valid inputs to the Self-Conditioning training step.

In RF*diffusion*, we input $\hat{x}_0$ through the template structure feature and we input x_t as input coordinates to the 3D track of RF. Inputting x_t as coordinates, as opposed to the distogram and anglegram used in the template structure feature, allows the network to keep the motif fixed in coordinate space.

M.6 Symmetric diffusion

As discussed in the main text, generating oligomeric assemblies obeying desired point-group symmetry constraints is a design goal. In what follows we describe how we have leveraged RFdiffusion to design symmetric oligomers.

Point group symmetries may be represented by a finite collection of rotation matrices that form a mathematical group with respect to matrix multiplication as the group operation[29]. For example, we may represent the cyclic symmetry group of order K by the set of rotation matrices that rotate increments of (360/K)° about the z-axis, $C_K = \{R_z^{(k/K)360^\circ}\}_{k=0}^{K-1}$. Analogous representations exist for all other point groups (including dihedral, tetrahedral, octahedral, and icosahedral). Without loss of generality, we set the first rotation to be the identity $R_1 = I_3$. We represent an oligomer with K monomer subunits each with M residues by $X = [x^1, \cdots, x^K]$ where each subunit k consists of the translations and rotations $x^k = ([z^k{}_1, \ldots, z^k{}_M], [r^k{}_1, \ldots r^k{}_M])$. Then, we say an oligomer obeys a point group symmetry $\mathfrak{R} = \{R_1, \ldots, R_K\}$, if

$X = [R_1 * x^1, \cdots, R_K * x^1]$ where $R * x^1 = ([Rx^1{}_1, \ldots, Rx^1{}_M], [Rr^1{}_1, \ldots, Rr^1{}_M])$ denotes the rotation of the monomer backbone structure by R.

Previous work has demonstrated some success generating designs with symmetry through hallucination with the inclusion of penalty terms on the deviation of predicted structures from the desired symmetry, but this work suffered from large computational cost (on the order of 1 GPU day per design) and low success rates, presumably due to the inability to precisely control the desired symmetry. We hypothesized that RF*diffusion* by contrast could provide improved control over symmetries in design by enforcing hard constraints during the *reverse process*.

In contrast to the hallucination approach, the desired symmetry is enforced from the beginning of the design trajectory and preserved throughout (Algorithm 2). Although exact symmetry is enforced through explicit symmetrization at each denoising step, we observe that RF*diffusion* provides predictions of the denoised oligomer structures that preserve the desired symmetry nearly exactly, even in the first denoising steps. This property of denoised predictions owes to the exact equivariance of RoseTTAFold with respect to global rotations and the approximate equivariance with respect to permutation (i.e. relabeling) of chains. In particular, in Section M.6II we provide a proposition that guarantees that rotation and permutation equivariance of a neural network are sufficient conditions for maintenance of point group symmetries of the neural network's output. In RoseTTAFold diffusion, exact rotation equivariance is inherited from the SE(3)-transformer architecture used in the structure module of RoseTTAFold[44]. Permutation

equivariance by contrast arises if the intermediate representations and outputs for each residue are unaffected by the ordering of chains. This is nearly the case with RF diffusion, with the exception that the RoseTTAFold pair representation contains directional sequence distance feature inputs for each pair of residues, clipped between -32 and 32 residues away; since oligomers are presented to RosettaFold by incrementing the sequence position index at the start of each chain[44], the sign of these features breaks exact permutation symmetry. However, we find empirically that deviation from exact symmetry in RF*diffusion* predictions is minimal even at the early steps.

M.6.II Proposition on rotation symmetry

We here provide a proposition that provides a mechanism by which predictions of denoised structures maintain the desired symmetry at each step. Here, we consider functions F: $[x_1, \ldots, x_K] \rightarrow [\hat{x}_1, \ldots, \hat{x}_K]$ that transform K rigid objects.

We first formally define rotation and permutation equivariance before stating the proposition.

Definition 1: F is rotation equivariant if for every $[x_1, \ldots, x_K]$ and every rotation matrix R, if $F([x_1, \ldots, x_K]) = [\hat{x}_1, \ldots, \hat{x}_K]$ then $F([Rx_1, \ldots, Rx_K]) = [R\hat{x}_1, \ldots, R\hat{x}_K]$.

Definition 2: F is permutation equivariant if for every $[x_1, \ldots, x_K]$ and every permutation $\sigma = [\sigma_1, \ldots, \sigma_K]$, if $F([x_1, \ldots, x_K]) = [\hat{x}_1, \ldots, \hat{x}_K]$ then $F([x_{\sigma_1}, \ldots, x_{\sigma_K}]) = [\hat{x}_{\sigma_1}, \ldots, \hat{x}_{\sigma_K}]$.

We now state the proposition.

Proposition 1: Consider any function F: $[x_1, \ldots, x_K] \rightarrow [\hat{x}_1, \ldots, \hat{x}_K]$ and point group symmetry $\mathfrak{R} = \{R_1, \ldots, R_K\}$. If F is both (1) rotation equivariant and (2) permutation equivariant, then F preserves symmetry; that is for any x, $F([R_1 x, \ldots, R_K x]) = [R_1\hat{x}, \ldots, R_K\hat{x}]$ for some $\hat{x}$.

Notably, Proposition 1 holds for any neural network satisfying the assumption on F above.

Supplementary section N:

Supplementary section M described the construction and RF-diffusion for unconditional generation of protein backbones. In this section we describe how we have leveraged RF-diffusion for generation subject to specific design criteria.

N1 Training RF*diffusion*

a) RoseTTAFold2 Architecture

RoseTTAFold2 (RF2) is an updated version of RoseTTAFold[44] with multiple architectural improvements: 1) use of a three-track architecture with initial coordinates from a template structure, 2) use of biased axial attention to update 2D pair features by considering geometric constraints between residues inferred from the current 3D structure, 3) communication between 1D, 2D, and 3D tracks through attention biasing, and 4) use of recycling that executes the network multiple times with the updated input embeddings based on outputs from the previous cycle. RF2 contains two major types of architecture blocks: main three-track blocks and the final structure refinement blocks. The 3-track blocks consist of layers of biased row and column attention over the 1D and 2D features , SE(3)-equivariant layers[47] to update 3D coordinates, and layers to communicate between 1D, 2D, and 3D features. The structure refinement blocks is based on SE(3)-equivariant network which gives refined 3D coordinates based on given 1D and 2D features.

b) RoseTTAFold2 Training

RF2 was trained based on mixture of dataset including 1) monomer/homo-oligomer structures in PDB, 2) hetero-oligomer structures in PDB, 3) AlphaFold2 structural models having pLDDT > 0.7, and 4) negative protein-protein interaction examples generated by random pairing. The training examples were sampled from each database with ratio of 2:1:4:1. The model was trained using the masked language model loss, distogram prediction loss, FAPE loss, accuracy estimation loss, and bond geometry loss, van der Waals energy loss. For the initial round of training, only the first four loss terms were used with crop size 256. After 200 epochs of initial round training, we performed fine-tuning with all the loss terms with crop size 384 for 100 epochs. The entire training took ~4 weeks of training using 64 V100 GPUs on Microsoft Azure. The training details are summarized in Table 2.

Table 2. Details for RoseTTAFold2 training

	Initial training	Fine-tuning
Crop size	256	384
Batch size	64	64
Loss function	$3.0*Loss_{MLM} + 1.0*Loss_{dist} + 10.0*Loss_{FAPE} + 0.1*Loss_{accuracy}$	$3.0*Loss_{MLM} + 1.0*Loss_{dist} + 10.0*Loss_{FAPE} + 0.1*Loss_{accuracy} + 0.1*Loss_{bond} + 0.1*Loss_{vdW}$
Learning rate & scheduling	0.001 Linear warm-up for first 1000 optimization steps, then decay learning rate by 0.95 after every 15000 optimization steps	0.0005 No warm-up. Decay learning rate by 0.95 after every 15000 optimization steps
Number of epochs	200	100

c) RF*diffusion* Training

RF*diffusion* is trained starting from the final RF2 weights. We do not implement loss ramping when training RF*diffusion*. We train RF*diffusion* with the following weights on each loss:

L2 Cα Displacement: 0.5
L2 Frame Displacement: 1
Distogram CCE: 1

RF*diffusion* trains to convergence when started from RF2 weights in ~5 epochs. Training takes ~3 days on 8 NVIDIA A100 GPUs.

N2: Conditional training for motif scaffolding

Our approach to scaffolding functional motifs with RF-Diffusion follows [2] and treats motif scaffolding as a conditional sampling problem. Notably, we consider partitioning the residues of a structure into the motif, and consider the remainder of the backbone as the scaffold that supports it. For a structure with L residues, we let M denote the set of indices corresponding to the motif and $\bar{M}$ be the remaining indices, such that the union of M and $\bar{M}$ is the set of indices up to L (i.e. $M \cup \bar{M} = \{1, \cdots, L\}$). We write x^M to denote the structure of the motif residues and $x^{\bar{M}}$ to be the scaffold residues such that we may write the whole protein structure as $x = [x^M, x^{\bar{M}}]$.

We build on previous work[1] demonstrating that RoseTTAFold may be trained to respect motif constraints provided as inputs through the template structure input features through retraining.

In the context of RF*diffusion*, inpainting training corresponds to learning the reverse process of a conditional generation task[53]. In particular, at training we learn

$p_\theta(x_{t-1}^{\bar{M}} | x_t^{\bar{M}}, x_0^{M})$.

To compute this we provide to RF*diffusion* the structure $x_t = [x^M, x_t^{\bar{M}}]$ at each step, and apply the RF*diffusion* denoising step only to the scaffold residues. For training and inference for motif scaffolding, we also condition the associated amino acid sequence and side chain torsion angles in addition to the backbone structure (provided through RoseTTAFold's template feature inputs). At training, losses are applied to both motif and scaffold residues to encourage RF*diffusion* to not move the motif.

N3: Fine-tuning RF*diffusion* on Protein Complexes and with fold information

The version of RF*diffusion* fine-tuned on protein complexes, henceforth referred to as RF*diffusion-PPI*, is trained starting from the base version of RF*diffusion* trained for 5 epochs. The model is shown monomer examples 50% of the time and complex examples 50% of the time. When the model is shown a complex example, only one side of the complex is noised, the other side is kept fixed (this is in keeping with established PPI design methods[21] where the target protein is kept fixed). When the model is shown a complex example the model is provided with the residue indices of 0-20% of the residues ("hotspot residues") in the interface on the fixed chain side (the interface is defined as all residues within 10 Å Cβ-Cβ distance of another chain), to permit targeting of the designed binder at inference time. During both complex and monomer training the model is provided with secondary structure 50% of the time and (independently) block adjacency information 50% of the time for the noised region. The junctions between blocks of secondary structure and their corresponding entries in the block adjacency matrix are masked during training, such that at inference time, one does not need to specify exact, *per residue* secondary structure and block adjacency matrices. Specifically, 0-75% of secondary structure (and corresponding adjacency, when provided), is masked, with this masking occurring over junctions in secondary structure (mask length 1-8 residues).

N4 Guiding RF*diffusion* inference with external potentials

In addition to the network's inbuilt ability to condition on structural motifs, the inference process can be steered by external potential functions to generate proteins which possess arbitrary desired properties, such as making contact with another protein chain or creating a desired concavity. Previous work has demonstrated that diffusion models can made to sample conditionally from $p_0(x_0|y)$ without retraining if given a classifier able to operate on noisy samples, $p_0(y|x_t)$ by moving in the direction of $\nabla_x \log p_t(x) + \nabla_x \log p_t(y|x)$ in the reverse step[8,45]. Using domain knowledge, it is relatively simple to construct heuristic approximations of $\nabla_x \log p_t(y|x) \approx \nabla_x P(x)$ for many protein conditional generation

objectives. When designing symmetric oligomers, we employ an inter- and intra-chain contact potential to promote the formation of contacts between subunits:

$$P_{sym}(X_{1...N}; d_0, r_c, n, m) = w_{inter}\left[\sum_{x_i \in X_k, x_j \in X_l; l \neq j} \text{Coord}(||x_i - x_j||_2^2; d_0, r_c, n, m)\right] + w_{intra}\left[\sum_{x_i \in X_k, x_j \in X_l; l = j} \text{Coord}(||x_i - x_j||_2^2; d_0, r_c, n, m)\right]$$

where the functional form of the attractive term is inspired by the differentiable coordination counter used in molecular dynamics:

$$\text{Coord}(r; d_0, r_c, n, m) = \frac{1 - (\frac{r - d_0}{r_c})^n}{1 - (\frac{r - d_0}{r_c})^m}$$

When designing enzymes, in addition to recapitulating the sidechain geometry of the active site, a pocket must be formed which has shape complementarity to the substrate. This condition can be captured effectively by a simple attractive-repulsive potential parameterized by the minimum distance between enzyme alpha-carbons and the substrate:

$$P_{enzyme}(X, S; d_0, r_c, n, m, r_0, m, p) = w_{attr}\left[\sum_{x_i \in X} \text{Coord}(\min_{s_j \in S} ||x_i - s_j||_2^2; d_0, r_c, n, m)\right] - w_{rep}\left[\sum_{x_i \in X} \text{Rep}(\min_{s_j \in S} ||x_i - s_j||_2^2; r_0, m, p)\right]$$

where:

$$\text{Rep}(r; r_0, p) = \max(0, \frac{|r_0 - r|^p}{p r_0^{p-1}})$$

decays smoothly from slope -1 at r=0 to 0 at r=r_0, penalizing clashes between the protein backbone and the substrate. As the noise β_t monotonically increases with t, at higher t the network is less sensitive to changes in X_t induced by a potential, so we scale of the potential by a monotonically increasing guide-scale g(t), such that:

$$P(X_t, t) = g(t) P(X_t)$$

The following hyperparameters for the potentials were empirically selected:

$$P_{sym} : w_{intra} = 0.2, w_{inter} = 2, n = 6, m = 12, d_0 = 8, r_c = 4, g(t) = \frac{t^2}{T^2}$$

$$P_{enzyme} : w_{attr} = 1, w_{rep} = 5, n = 6, m = 12, d_0 = 8, r_c = 2, g(t) = 1$$

N5: *In silico* experiments with RF*diffusion*

Unconditional benchmarking

To test RF*diffusion* on unconditional generation of monomers (Fig. 1C-F), we generated 100 designs for lengths 70, 100, 200, 300, 400, 600, 800 and 1000 amino acids. For each backbone, we generated 8 sequences with ProteinMPNN and subsequently predicted their structures with AF2 (or ESMFold - Fig. S1H). The best sequence (by alignment of the predicted structure to the design model) was taken for each backbone. We benchmarked against the recently-published RoseTTAFold Hallucination[1]. As some knowledge of how best to use RoseTTAFold for Hallucination is required, these samples were generated by the respective expert.

Conditional benchmarking

The full conditional benchmark is described in Table 1, and encompasses 25 design challenges from six recent publications[1,2,20,30–32]. RF*diffusion* was compared to RoseTTAFold Hallucination and RF*joint* Inpainting. While both Hallucination and Inpainting are able to generate sequences directly, for the fairest comparison, we also redesigned the sequence with ProteinMPNN, and took the best of 8 sequences per backbone. Both RF*joint* Inpainting and RF Hallucination are able to scaffold structure without sequence, so in cases where functional-site residues were not required for function, these methods were permitted to redesign the sequence of the non-functional residues, which is generally beneficial for design. Finally, as Hallucination requires some expert knowledge and empirical hyperparameter tuning, some exploration of the benchmark set was permitted, and these designs were generated by the respective expert.

Design of protein binders to rigid targets

To test the ability of RF*diffusion* to design *de novo* binders to rigid targets, we designed binders to five targets: PD-L1 (PDB: 5O45), IL7 Receptor α (PDB: 3DI3), Insulin Receptor (PDB: 4ZXB), TrkA Receptor (PDB: 1WW7) and Flu Hemagglutinin (PDB: 5VLI). We generated designs both with and without fold conditioning, with the folds used derived from scaffold sets typically used for Rosetta-based protein binder design. In all cases, we targeted binders, using input “hotspot” residues, to a specific site on the target protein. In line with current best practice[21], we tried using Rosetta FastRelax[42] before running a single ProteinMPNN, although we found that this was not systematically helpful for design success rates. For the five design cases, we generated several thousand designs. We classed a design as successful if it had AF2 pAE of interaction between binder and target < 10 (this has been shown to be highly indicative of design success), as well as RMSD between the designed binder and the AF2 prediction < 1 Å , and AF2 pLDDT > 80. Success rates are reported in figure 4B, and were several orders of magnitude higher than with traditional Rosetta binder design.

Bibliography

1. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
2. Trippe, B. L. *et al.* Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. (2022) doi:10.48550/ARXIV.2206.04119.
3. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
4. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
5. Anand, N. & Achim, T. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models. (2022) doi:10.48550/ARXIV.2205.15019.
6. Luo, S. *et al.* Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models. 13.
7. Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. Preprint at https://doi.org/10.48550/arXiv.2006.11239 (2020).
8. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N. & Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. Preprint at https://doi.org/10.48550/arXiv.1503.03585 (2015).
9. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M. Hierarchical Text-Conditional Image Generation with CLIP Latents. Preprint at http://arxiv.org/abs/2204.06125 (2022).
10. Jendrusch, M., Korbel, J. O. & Sadiq, S. K. AlphaDesign: A de novo protein design framework based on AlphaFold. 2021.10.11.463937 Preprint at https://doi.org/10.1101/2021.10.11.463937 (2021).
11. Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for

protein design. *Nat. Commun.* **13**, 4348 (2022).

12. Wu, K. E. *et al.* Protein structure generation via folding diffusion. (2022) doi:10.48550/arXiv.2209.15611.
13. Baek, M., McHugh, R., Anishchenko, I., Baker, D. & DiMaio, F. Accurate prediction of nucleic acid and protein-nucleic acid complexes using RoseTTAFoldNA. 2022.09.09.507333 Preprint at https://doi.org/10.1101/2022.09.09.507333 (2022).
14. De Bortoli, V. Convergence of denoising diffusion models under the manifold hypothesis. Preprint at http://arxiv.org/abs/2208.05314 (2022).
15. Leach, A., Schmon, S. M., Degiacomi, M. T. & Willcocks, C. G. Denoising Diffusion Probabilistic Models On SO(3) For Rotational Alignment. 8 (2022).
16. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
17. Wu, R. *et al.* High-resolution de novo structure prediction from primary sequence. 2022.07.21.500999 Preprint at https://doi.org/10.1101/2022.07.21.500999 (2022).
18. Lin, Z. *et al.* Language models of protein sequences at the scale of evolution enable accurate structure prediction. 31.
19. Chen, T., Zhang, R. & Hinton, G. Analog Bits: Generating Discrete Data using Diffusion Models with Self-Conditioning. Preprint at https://doi.org/10.48550/arXiv.2208.04202 (2022).
20. Lee, J. S. & Kim, P. M. ProteinSGM: Score-based generative modeling for de novo protein design. *bioRxiv* (2022) doi:10.1101/2022.07.13.499967.
21. Bennett, N. *et al.* Improving de novo Protein Binder Design with Deep Learning. 2022.06.15.495993 Preprint at https://doi.org/10.1101/2022.06.15.495993 (2022).
22. Shi, C., Wang, C., Lu, J., Zhong, B. & Tang, J. Protein Sequence and Structure Co-Design with Equivariant Translation. Preprint at https://doi.org/10.48550/arXiv.2210.08761 (2022).
23. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* eadd2187 (2022) doi:10.1126/science.add2187.

24. Anand, N. & Huang, P. Generative modeling for protein structures. in *Advances in Neural Information Processing Systems* vol. 31 (Curran Associates, Inc., 2018).
25. Basanta, B. *et al.* An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl. Acad. Sci.* **117**, 22135–22145 (2020).
26. Pan, X. *et al.* Expanding the space of protein geometries by computational design of de novo fold families. *Science* **369**, 1132–1136 (2020).
27. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **176**, 1420-1431.e17 (2019).
28. Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
29. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
30. Sesterhenn, F. *et al.* De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science* **368**, (2020).
31. Yang, C. *et al.* Bottom-up de novo design of functional proteins with complex structural features. *Nat. Chem. Biol.* **17**, 492–500 (2021).
32. Glasgow, A. *et al.* Engineered ACE2 receptor traps potently neutralize SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28046–28055 (2020).
33. Hunt, A. C. *et al.* Multivalent designed proteins neutralize SARS-CoV-2 variants of concern and confer protection against infection in mice. *Sci. Transl. Med.* **14**, eabn1252 (2022).
34. Silverman, J. *et al.* Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat. Biotechnol.* **23**, 1556–1561 (2005).
35. Detalle, L. *et al.* Generation and Characterization of ALX-0171, a Potent Novel Therapeutic Nanobody for the Treatment of Respiratory Syncytial Virus Infection. *Antimicrob. Agents Chemother.* **60**, 6–13 (2016).

36. Strauch, E.-M. *et al.* Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat. Biotechnol.* **35**, 667–671 (2017).
37. Boyoglu-Barnum, S. *et al.* Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).
38. Walls, A. C. *et al.* Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367-1382.e17 (2020).
39. Quijano-Rubio, A., Ulge, U. Y., Walkey, C. D. & Silva, D.-A. The advent of de novo proteins for cancer immunotherapy. *Curr. Opin. Chem. Biol.* **56**, 119–128 (2020).
40. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
41. Ribeiro, A. J. M. *et al.* Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **46**, D618–D623 (2018).
42. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
43. Lin, Z., Sercu, T., LeCun, Y. & Rives, A. Deep generative models create new and diverse protein structures. 17.
44. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
45. Song, Y. *et al.* Score-Based Generative Modeling through Stochastic Differential Equations. Preprint at http://arxiv.org/abs/2011.13456 (2021).
46. Xu, M. *et al.* GeoDiff: a Geometric Diffusion Model for Molecular Conformation Generation. Preprint at http://arxiv.org/abs/2203.02923 (2022).
47. Fuchs, F. B., Worrall, D. E., Fischer, V. & Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. Preprint at http://arxiv.org/abs/2006.10503 (2020).
48. Larochelle, P. M., Murray, A. P. & Angeles, J. A Distance Metric for Finite Sets of Rigid-Body

Displacements via the Polar Decomposition. *J. Mech. Des.* **129**, 883–886 (2007).

49. Dhariwal, P. & Nichol, A. Diffusion Models Beat GANs on Image Synthesis. Preprint at http://arxiv.org/abs/2105.05233 (2021).
50. Vincent, P. A Connection Between Score Matching and Denoising Autoencoders. *Neural Comput.* **23**, 1661–1674 (2011).
51. Solà, J., Deray, J. & Atchuthan, D. A micro Lie theory for state estimation in robotics. Preprint at https://doi.org/10.48550/arXiv.1812.01537 (2021).
52. Savinov, N., Chung, J., Binkowski, M., Elsen, E. & Oord, A. van den. Step-unrolled Denoising Autoencoders for Text Generation. Preprint at http://arxiv.org/abs/2112.06749 (2022).
53. Saharia, C. *et al.* Palette: Image-to-Image Diffusion Models. Preprint at http://arxiv.org/abs/2111.05826 (2022).
35. Thomas, Nathaniel, et al. "Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds." arXiv preprint arXiv:1802.08219 (2018).
36. Nikolayev, Dmitry I., and Tatjana I. Savyolov. "Normal distribution on the rotation group SO (3)." Textures and Microstructures 29 (1970).