PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Modeling protein–small molecule conformational ensembles with PLACER

Ivan Anishchenko[a,b,1], Yakov Kipnis[a,b,c,1], Indrek Kalvet[a,b,c,1], Guangfeng Zhou[a,b], Rohith Krishna[a,b], Samuel J. Pellock[a,b], Anna Lauko[a,b,d], Gyu Rie Lee[a,b,c], Linna An[a,b], Justas Dauparas[a,b], Frank DiMaio[a,b], and David Baker[a,b,c,2]

Affiliations are included on p. 8.

Edited by Brian K. Shoichet, University of California, San Francisco, CA; received January 3, 2025; accepted September 22, 2025 by Editorial Board Member William F. DeGrado

Modeling the conformational heterogeneity of protein–small molecule interactions is important for understanding natural systems and evaluating designed systems but remains an outstanding challenge. We reasoned that while residue-level descriptions of biomolecules are efficient for de novo structure prediction, for probing heterogeneity of interactions with small molecules in the folded state, an entirely atomic-level description could have advantages in speed and generality. We developed a graph neural network called PLACER (protein-ligand atomistic conformational ensemble resolver) trained to recapitulate correct atomic positions from partially corrupted input structures from the Cambridge Structural Database and the Protein Data Bank; the nodes of the graph are the atoms in the system. PLACER accurately generates structures of diverse organic small molecules given knowledge of their atom composition and bonding. When given a description of the larger protein context, it builds up structures of small molecules and protein side chains for protein–small molecule docking. Because PLACER is rapid and stochastic, ensembles of predictions can be readily generated to map conformational heterogeneity. In enzyme design efforts described here and elsewhere, we find that using PLACER to assess the accuracy and preorganization of the designed active sites results in higher success rates and higher activities; we obtain a preorganized retroaldolase with a k_{cat}/K_M of 11,000 $M^{-1}min^{-1}$, considerably higher than any pre–deep learning design for this reaction. We anticipate that PLACER will be widely useful for rapidly generating conformational ensembles of small molecule and small molecule–protein systems and for designing higher activity preorganized enzymes.

machine learning | enzyme design | ligand docking

Significance

Capturing the conformational diversity of protein–small molecule interactions is crucial for gaining insights into natural and designed systems. We have developed a graph neural network, PLACER (protein-ligand atomistic conformational ensemble resolver), that generates conformational ensembles of protein–small molecule systems using an entirely atomic representation. PLACER ensemble generation provides a rapid means to map protein–small molecule conformational landscapes in ligand docking and to assess the accuracy and preorganization of designed active sites, which can substantially increase the success of enzyme design efforts.

Interactions of proteins with nucleic acids (NAs), small organic and inorganic molecules, and metals are critical to biological function, but atomistic modeling of such interactions and their conformational heterogeneity remains a challenging problem. Deep learning (DL)-based small molecule docking tools like DiffDock (1) improve on earlier methods such as Glide (2) and GNINA/SMINA, but in the high accuracy regime, the difference in performance is not pronounced, and performance also drops substantially on unseen receptors (1, 3, 4). DL methods have also been devised to generate small molecule conformations from their chemical structure (5–7); when trained on the synthetic GEOM-DRUGS dataset (8) of drug-like molecules computed with semiempirical QM methods, diffusion generative models like DiffDock (1) and Torsional Diffusion (7) show best-in-class performance on the hold-out test set. However, these approaches model specific classes of interaction partners, limiting the ability to model the full spectrum of protein functions, and the input features can differ depending on the type of input molecules, which could limit the ability of the networks to distill general physicochemical principles. AlphaFold2 (AF2) (9) and RoseTTAFold (RF) (10) enabled atomically accurate structure prediction of proteins and protein–protein complexes using sequences and structures of evolutionary-related proteins as inputs. These methods have recently been extended by AF3 and related approaches to modeling the structures of protein–NA complexes and more general biomolecular systems (11–13) by supplementing tokens representing residues with full atomic representations of the nonprotein components of the system.

We reasoned that rapid and accurate modeling of structure and conformational heterogeneity could be achieved by an atom centric approach to predicting the structures of small molecules and peptides in isolation and in the context of a protein binding or active site defined at the backbone level. Networks such as AF2, AF3, and RF achieve accurate structure prediction using a primarily residue-level description of biomolecules and multiple sequence information to help infer contacts. At the atomic level close to the native structure, the

Author contributions: I.A. and D.B. designed research; I.A., Y.K., I.K., G.Z., R.K., S.J.P., A.L., J.D., and F.D. performed research; I.A., Y.K., I.K., G.Z., S.J.P., A.L., G.R.L., and L.A. analyzed data; F.D. and D.B. supervised research; and I.A., Y.K., I.K., and D.B. wrote the paper.

Competing interest statement: D.B. is a cofounder and shareholder of Vilya, an early-stage biotechnology company that has licensed the provisional patent. A provisional patent (Application No. 63/535,404) covering the PLACER network (formerly known as ChemNet) presented in this paper has been filed by the University of Washington. D.B., I.A., G.Z., R.K., and F.D. are inventors on this patent.

This article is a PNAS Direct Submission. B.K.S. is a guest editor invited by the Editorial Board.

[1]I.A., Y.K., and I.K. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: dabaker@uw.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2427161122/-/DCSupplemental.

Published November 4, 2025.

complexity grows considerably, and evolutionary information is less relevant. Hence, we reasoned that an appropriate starting point for an ensemble generation method would not be the amino acid sequence of a protein but rather the coordinates of the overall protein backbone surrounding a binding or catalytic pocket, along with an atomic-level description of the bonded geometry of the interacting small molecules and amino acid side chains. Such a network would, by construction, not be useful for structure prediction from sequence, but could be very useful for modeling the conformations of small molecules and constrained peptides both in isolation and in a protein binding site along with the conformations of the interacting sidechains. Since the protein backbone structure is input, the calculations could be faster than AF2, AF3, and RF, enabling rapid binding site refinement and evaluation. We reasoned that a network that generated predictions stochastically could have the advantage of enabling rapid generation of predicted ensembles for modeling systems with a distribution of possible states (which is difficult with AF2, AF3, and RF), and for evaluating the extent of preorganization of designed molecules and functional sites.

Guided by this reasoning we set out to develop a stochastic deep neural network called PLACER (protein-ligand atomistic conformational ensemble resolver) for atomistic modeling of small molecules and small molecule–protein interactions (Fig. 1*A*). In many applications, a reliable structure of the target protein can be generated by AF2 (9) or RF (10), or obtained from the PDB, and location of the putative interacting region on the protein surface is also known; the task is then to dock a small molecule into the region of interest, while adjusting conformations of the protein side chains and the molecule itself. We framed the learning problem as a structure denoising task and trained PLACER to recapitulate the correct atom positions from partially corrupted input structures provided that all the chemical information about the system being modeled is known from the start. We customized the corruption strategy to the application of interest. In the case of the protein–small molecule docking (Fig. 1*A*), the inputs to the network include the protein backbone coordinates, the amino acid sequence with side chain coordinates randomly initialized around the respective C-alpha atoms, and the small molecule chemical structure with atomic positions randomly initialized in the vicinity of the putative binding site.

At input, the molecular system is converted to a chemical graph with nodes representing individual heavy atoms (hydrogen atoms are not modeled to reduce computation cost) and edges representing chemical bonds between atoms (Fig. 1*C*). This representation is uniform across molecule types. Each node in the network carries information about the atom type and its 3D coordinates which are initially corrupted. The network is tasked to iteratively denoise the input coordinates and to estimate uncertainties in the atom positions of the output model structure (Fig. 1*D*). PLACER has a three-track architecture inspired by RF (10). After the initial embedding of the 1D and 2D features, they are passed to a block which iteratively updates the embeddings and the 3D structure. In the iteration block, the atom neighbor graph is first constructed: For every atom, the 32 closest neighbors are picked in equal proportions based on both spatial proximity and proximity in the chemical graph. 2D pair features are then projected by a feed-forward adapter layer to edge embeddings, and, together with the 1D features, the atom neighbor graph and the current atomic 3D structure serve as inputs to the SE3-Transformer network (14) which updates the 3D coordinates and the 1D embeddings. Information about the chiral centers is communicated to the network via type-1 (vector) features (Fig. 1*E*). Features in the 2D track undergo pair-to-pair updates with bias from structure (15). Atom and atom pair confidence prediction heads branch off the 1D and 2D tracks, respectively, completing the iteration block. The fully trained network consists of eight iteration blocks with shared weights.

PLACER is trained using a combination of structure and confidence prediction losses applied after every iteration. The primary structure loss is all-atom frame-aligned point error $\mathrm{FAPE}_{\mathrm{allatom}}$, which is an extension of the original FAPE from ref. 9 to arbitrary molecules (Fig. 1*F*). The confidence of the modeled structures is evaluated on a per atom and per atom pair basis. Atom–atom pair accuracies are estimated using the distance signed error approach introduced in ref. 16. On the per-atom level, we predict all-atom lDDT scores as in AF2 (9). The network also predicts deviations in atomic positions σ_i (in Å) in the generated model relative to the reference structure (Fig. 1*G*). These predicted deviations can then be combined over a subset of atoms of interest (e.g., a small molecule) to give the predicted RMSD $\mathrm{pRMSD} = \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right)^{1/2}$.

Tuning Network Architecture on Experimental Structures of Small Molecules

While small rigid molecules have known three-dimensional structures, larger organic molecules can have unique ground states that are very time consuming to predict with computational methods, hence, accurate prediction of general molecular structure is an important challenge. We explored network architecture and training hyperparameters on crystal structures of chemicals from the Cambridge Structural Database (17). This also helped ensure that the network architecture can handle diverse chemistry and considerably reduced the architecture exploration time (PLACER training on the PDB until convergence takes ×10 longer). The training task was to predict the small molecule conformation observed in the crystal from the chemical structure and randomly initialized starting coordinates (Fig. 2*A*). The training and validation datasets consisted of 226,684 and 7,116 examples of organic nonpolymeric small molecules, respectively (further details can be found in *SI Appendix*, section 1.2), and confidence heads were omitted. As PLACER is a denoising network, structurally diverse samples of molecular conformations can be generated by running the network multiple times with different random initialization of the input coordinates. Training of PLACER on CSD structures was done in two stages: in the first stage, four iteration blocks were used and $\mathrm{FAPE}_{\mathrm{allatom}}$ was the only loss term. In the second stage, the number of iterations was increased to eight and bonded geometry loss terms were added to enhance the local quality of predicted structures (red vs. violet in Fig. 2 *B* and *C*). Reducing the number of iterations to two or replacing the FAPE loss with a combination of coordinate and distance RMSD losses both result in predictions of substantially lower quality (orange and blue in Fig. 2 *B* and *C*). Performance also degraded when the 2D inputs did not include the bond separation feature (an integer counting the number of covalent bonds between any two atoms in the chemical graph) but only included a binary feature flagging that the two atoms are immediately connected by a bond. Fully trained PLACER generates the correct 3D structures of molecules as complex as macrocycles with over 50 atoms (Fig. 2*D*) with sub-Å accuracy, including peptidic macrocycles (18) shown in the top row of Fig. 2*D*.

Modeling Protein–Small Molecule Interactions with PLACER

To train PLACER on the structures from the PDB (proteins plus small molecules, rather than small molecules alone, as in the case of the CSD) we parsed them into chemical graphs (Fig. 1*C*) using

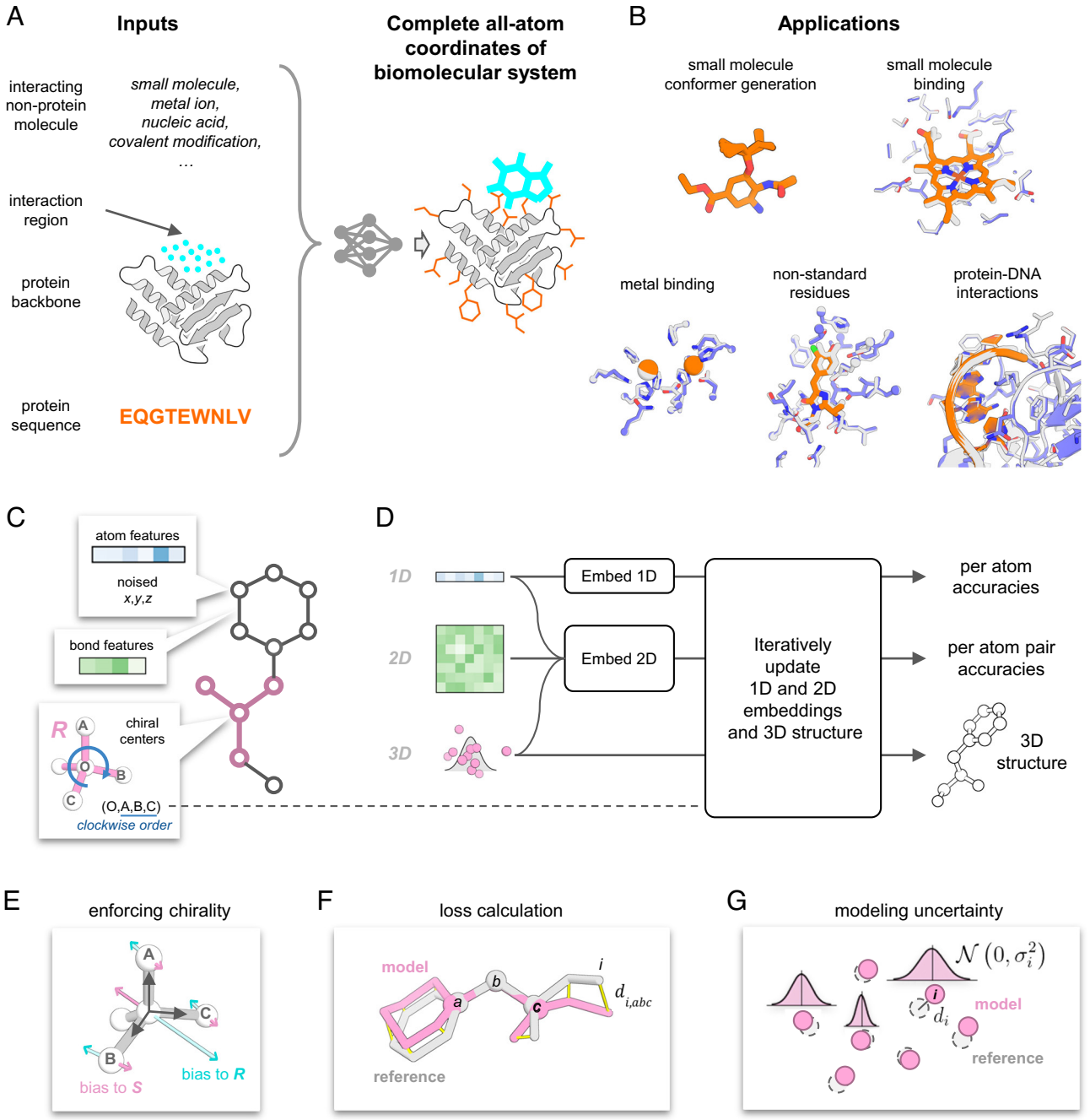

Fig. 1. Overview of PLACER. (*A*) PLACER is a denoising neural network which takes at input a partially corrupted protein structure and the chemical structure (but not the coordinates) of any interacting molecules, and predicts the all-atom structure of the complex, as well as the uncertainties in the atom positions in the generated model. (*B*) PLACER can be used for a wide range of tasks including docking of small molecules and metals to a protein target, modeling nonstandard residues, and predicting side chain conformations of amino acid residues and nucleotides at the protein–DNA interface. Shown are X-ray structures (in gray) superimposed with PLACER models (in blue and orange). (*C*) At input, the molecular system is represented by an annotated graph where nodes are individual atoms and edges are chemical bonds between atoms. Information about chiral centers is supplied to the network as (O, A, B, and C) tuples where O is the central pyramidal or tetrahedral atom and its neighboring atoms A, B, and C are ordered clockwise. (*D*) PLACER is a three-track network that iteratively updates 1D and 2D embeddings and the 3D structure, producing at each iteration a refined atomic structure model and estimating uncertainties in atom placements. (*E*) The triple product $V = \vec{e}_A \cdot \vec{e}_B \times \vec{e}_C$ of the three unit vectors pointing from the central chiral atom to its neighbors (gray arrows) is a pseudoscalar that differs in sign for the *R* and *S* configurations: For ideal tetrahedral geometry, $V_R = \frac{-4}{3\sqrt{3}}$ and $V_S = \frac{+4}{3\sqrt{3}}$. By comparing V in the nonideal geometry of the modeled structure to the ideal values V_R or V_S and taking gradients w.r.t. atom coordinates, one gets biasing vectors $\vec{f}_{R/S} = \nabla_{\vec{r}} \left(V - V_{R/S}\right)^2$ showing the directions in which atoms should be moved around in order to recreate the desired configuration. (*F*) All-atom FAPE is calculated by aligning the model and the reference structures on every three respective bonded atoms *a*, *b*, *c* and calculating the deviations in atom positions between the aligned structures. $\text{FAPE}_{\text{allatom}}$ is then the mean over all atoms and all superimpositions. Atom–atom distances are clamped at 10 Å. (*G*) Assuming that uncertainties in atom positions in the modeled structure are normally distributed, we let the dedicated head of the network predict the variances σ_i^2 for every atom in the system to recapitulate actual deviations d_i. These variances are learned during training by maximizing the likelihood $N(d_i; 0, \sigma_i^2)$.

the Chemical Component Dictionary (19) that provides detailed chemical description for all residue and small molecule components found in the PDB. We reasoned that although some molecules in the PDB may be nonbiological (e.g., solvents) and their interactions with the rest of the structure may be nonspecific, these interactions could still carry information about

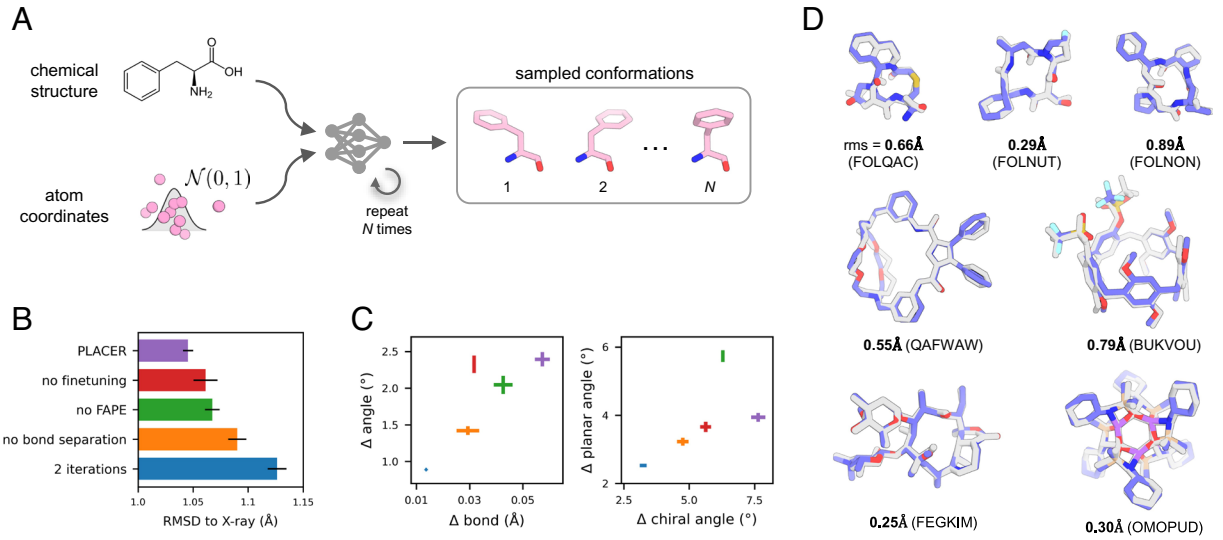

Fig. 2. Modeling complex small molecules. (*A*) PLACER predicts the 3D structure of a small molecule from its chemical structure and randomly initialized atomic coordinates. Running PLACER multiple times with different random initializations of atom positions yields a diverse set of molecular conformations. (*B*) Ablating PLACER features negatively affects performance as measured by the RMSD in atom coordinates between the model and the reference X-ray structure after the two are optimally superimposed. (*C*) Quality of the local geometry in models produced by the five networks from panel *B*. Shown are mean absolute errors in bonds, bonded, planar, and chiral angles computed from the model and reference structures. Validation set of 7,116 small molecules from the CSD was used for panels *B* and *C*. A single conformer was generated for every validation example; the error bars represent SD over five independent validation runs. (*D*) Examples of seven macrocyclic molecules (18) for which PLACER generates the correct structure. Shown are the best RMSD models (in blue; out of 1,000 generated) overlaid with the experimental crystal structures (gray). CCDC database identifiers shown in parentheses.

physicochemical preferences at molecular interfaces and thus be informative for the network training; only water molecules were excluded. Training and validation sets (112,828 and 7,090 examples, respectively) were compiled from higher resolution (<2.5 Å) structures deposited to the PDB before January 12, 2023. During training, the input structures were cropped to at most 600 heavy atoms and centered around a randomly selected atom. The non-backbone-atom coordinates of all individual connected molecular components in the crop were collapsed to their connected backbone atom (if residue or NA), or to a randomly selected atom (ligands), and corrupted with Gaussian noise (σ = 1.5 Å). From these starting points, PLACER was trained to recapitulate the atomic coordinates of all atoms in the cropped regions. For a detailed description of the training procedure, see *SI Appendix, Supplementary Methods*.

We used PLACER to generate ensembles of small molecule poses in the pocket of the target protein (Fig. 3*A*) by repeated runs with different random initialization of the input coordinates. For each run, we chose one ligand atom at random, added Gaussian noise (σ = 1.5 Å) to its coordinates, collapsed all other ligand atoms onto that atom (these starting positions from multiple network runs are shown in Fig. 3 *B* and *C*), and added uncorrelated Gaussian noise (σ = 1.5 Å) to the resulting coordinates of all ligand atoms—exactly matching the initialization used during training and ensuring that neither the internal structure of the ligand nor its orientation in the pocket were inferrable from the inputs to the network. Analyzing the resulting ensembles reveals that PLACER is not very sensitive to the initial placement of the ligand: Multiple different starting positions yield near native predictions (Fig. 3*B*), and these positions cover the entire space sampled at input (Fig. 3*C*). We also observed that the predicted pRMSD score calculated over the ligand atoms allows selection of more accurate models from the sampled pool (Fig. 3*E*), with the best scoring models closely matching the experimental structure [ligand RMSD = 0.53 Å and 0.77 Å for heme (20) and cortisol (21), respectively, Fig. 3*D*].

To assess PLACER performance in a more realistic scenario of docking against non-native protein conformations, we used a standard test set of 65 drug targets from the Astex non-native set (22). Each of these targets has a cocrystal structure with a drug molecule, as well as a number of structures either without the drug or with a different small molecule cocrystallized, totaling to 1,112 non-native structures over the 65 targets. The docking results highlight the importance of both sampling and scoring aspects of the network (Fig. 3*F*): The success rate of generating and selecting a near-native conformation of the docked small molecule increases if more models are generated. Among the three confidence prediction metrics tested, the pRMSD score shows the best power of selecting near-native conformations (Fig. 3*G*, top 3 bars), with higher docking success rates than Vina (23), GOLD (22), and GalaxyDock (24) (Fig. 3*G*, bars in the middle; number were taken from respective references). Compared to the best performing Rosetta GALigandDock method (25). PLACER performance is higher in the lower accuracy regime (% of complexes with ligand RMSD < 2 Å, 82.4% vs. 73.6%), but is behind in the high-accuracy regime (RMSD < 1 Å, 41.8% vs. 51.6%). PLACER performance is still notable, however, because unlike the other methods, the network was not specifically trained on the non-native protein–small molecule docking task. PLACER recreates both the conformations of the small molecule and the protein side chains from scratch, while the other methods tested rely largely on the input protein coordinates. Improved performance can be obtained by combining PLACER with physics-based methods: minimizing the docks generated by the network using the generalized Rosetta force field and estimating binding free energies of the minimized structures gives +7.3% increase in docking success rate at <1 Å and no notable change at <2 Å; likely due to improved sampling. Astex benchmark performance showed no significant decline when ligands with Tanimoto similarity ≥ 0.5 to the training set were excluded, indicating that data leakage is not substantial (*SI Appendix*, Table S3).

We further compared PLACER with AlphaFold3 and RF All-Atom on the PoseBusters benchmark (26). This benchmark

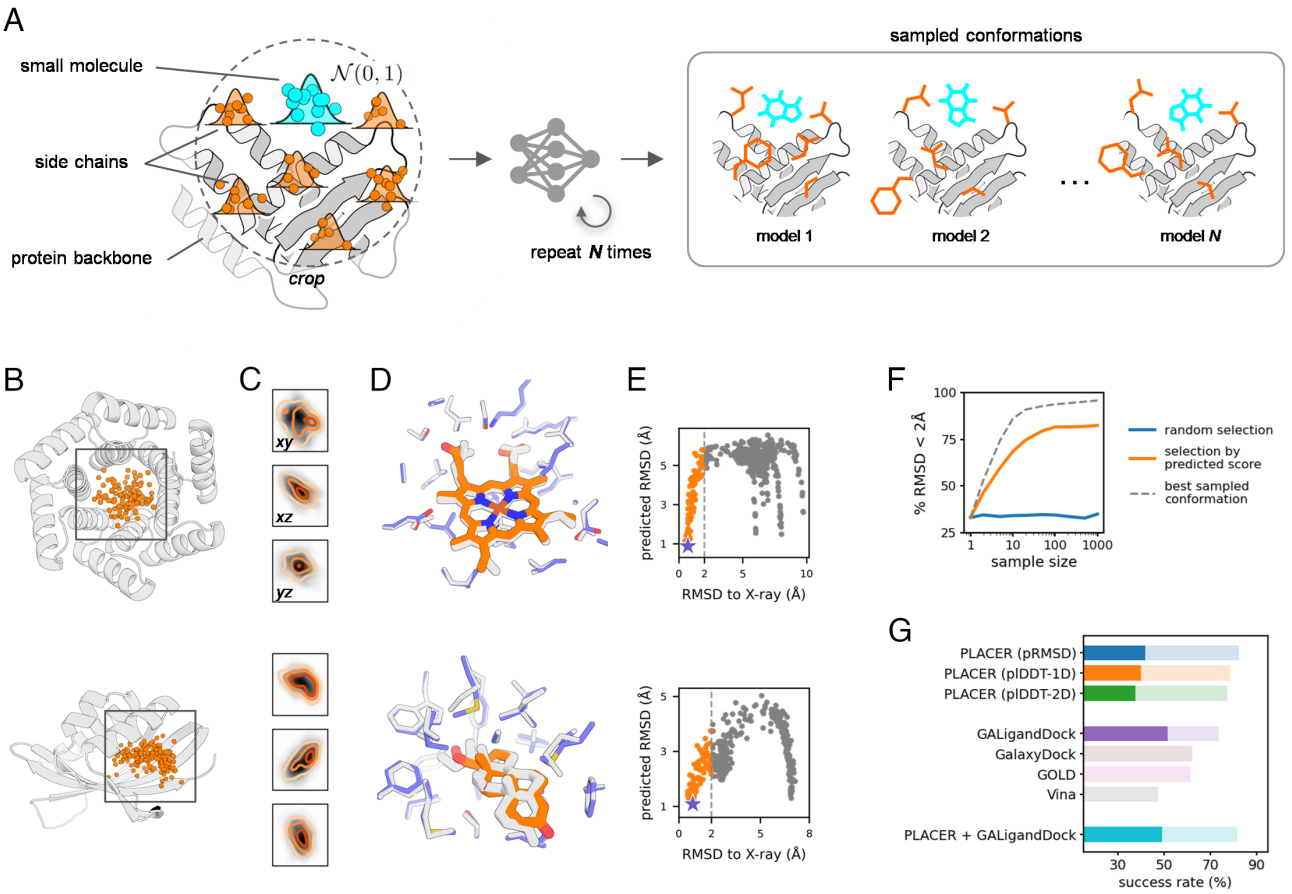

Fig. 3. Modeling protein–small molecule interactions with PLACER. (*A*) Starting with the protein backbone and the coordinates of the small molecule randomly initialized in the vicinity of the binding site (cyan dots), PLACER predicts protein side chain conformations (which are initially randomized around the respective C-alpha atoms, orange dots) and the structure and the placement of the small molecule relative to the target protein. By repeating this process multiple times, PLACER generates a structurally diverse set of docks. (*B* and *C*) Sampled starting positions of the small molecule for the two de novo designed small molecule binders. Panel *C* shows projections of the sampled starting points onto the three coordinate planes. Shades of gray show density of all sampled positions; orange contours indicate density of points which resulted in docks with ligand RMSD < 2 Å. The two densities largely overlap, suggesting that PLACER is not very sensitive to the initial placement of the ligand in the binding pocket. (*D*) Poses with the lowest pRMSD scores (blue/orange) closely match the experimental structure (gray). (*E*) Models with lower pRMSD scores are funneled toward the native conformation demonstrating the discriminative power of the predicted score. (*F*) Increasing the sample size and rescoring PLACER models by the pRMSD score (orange curve) greatly increases the chance of picking a better model compared to a random selection (blue curve). Scoring by pRMSD is however still not optimal (orange vs. gray dashed curve). (*G*) Among the three accuracy scores predicted by PLACER, pRMSD shows best performance (the three bars at the top). Bright and pale bars show docking success rates in the Astex non-native set for ligand RMSD < 1Å and 2Å, respectively. The four bars in the middle show performance of the widely used docking tools Vina, GOLD, GalaxyDock, and Rosetta GALigandDock. Sampling docks with PLACER and minimizing and rescoring them using GALigandDock increases docking success rate by 7.3% (ligand RMSD < 1Å) over PLACER alone.

consists of 428 protein–ligand complexes from the PDB, and for PLACER, the prediction task was solely to predict the correct ligand binding orientation, given a known ligand binding site and protein backbone structure. On the 118 complexes absent from PLACER's training and validation sets, PLACER achieved 84% success at <2 Å RMSD, better than AlphaFold3 (77%) and RF All-Atom (38%) (the latter two solve the harder problem of de novo prediction of protein structure and small molecule docking; *SI Appendix*, Fig. S6). At the stricter <1 Å threshold, however, AlphaFold3 remained superior (58% vs. 51%).

Assessment of Enzyme Active Site Design Accuracy and Preorganization

A critical challenge in de novo enzyme design is to come up with amino acid sequences that not only encode the target backbone structure but also position the catalytic sidechains and the reaction substrate such that the key sidechain functional groups make the hydrogen bonding and electrostatic interactions with the transition state and each other that are essential to catalysis. As 0.5 Å deviations in distance and 30-degree deviations in angle can have large effects on hydrogen bonding energies, the level of accuracy in positioning required is quite high. Furthermore, the designed active site should be preorganized in the absence of substrate with the catalytic sidechains largely held in place by intraprotein interactions to reduce the entropic cost of binding and ensure that all catalytic groups are properly positioned; such preorganization is a notable feature of natural enzyme active sites. The most general way to evaluate such preorganization is molecular dynamics simulations (27, 28), but as sidechain reconfiguration can occur on the microsecond time scale, such simulations are computationally expensive and cannot be applied to large numbers of designs. Approximate discrete methods such as the Rosetta Rotamer Boltzmann method (29) estimate preorganization at the individual sidechain level, but are limited by the rotamer approximation and the inability to model coupled and concerted movements of side chains and small molecules. We reasoned that PLACER could enable assessment of the accuracy and preorganization of designed active sites by generating ensembles of models of small molecules and interacting sidechain placements much more rapidly than molecular dynamics simulations and without the limitations of the Rotamer Boltzmann method.

We applied PLACER to the RA95 series of previously designed retro-aldolase (Fig. 4*A*) enzymes, and directed evolution generated improved versions of these, for which high-resolution crystal structures have been determined (30–32). For every retro-aldolase intermediate shown in Fig. 4*A*, we ran PLACER 50 times and analyzed the structural diversity of the active site lysine (and covalent adducts of this lysine) through the mean predicted RMSD (pRMSD) calculated over side chain atoms and excluding backbone N, Cα, C, O. As illustrated in Fig. 4 *B* and *C* and S5, PLACER generated ensembles are highly varied for the lower activity initial computational designs, indicating a lack of preorganization, and become increasingly well ordered for the more active evolved versions. This suggests that lack of preorganization was a major shortcoming of early enzyme design efforts [similar conclusions were reached using considerably more expensive MD simulations (33)] and that PLACER provides a rapid means to assess this, and thus guide enzyme design efforts.

We further tested the use of PLACER prospectively to guide protein design by making a new round of designed retro-aldolases using DL–generated NTF2-like folds (21, 34) that previously yielded high activity de novo luciferases (35). We started from an active site description based on the lysine-centered catalytic tetrad found in the most active of the evolved retroaldolases (32) (*SI Appendix*, Fig. S9). This active site contains an intricate network of hydrogen bonds keeping the catalytic residues in place

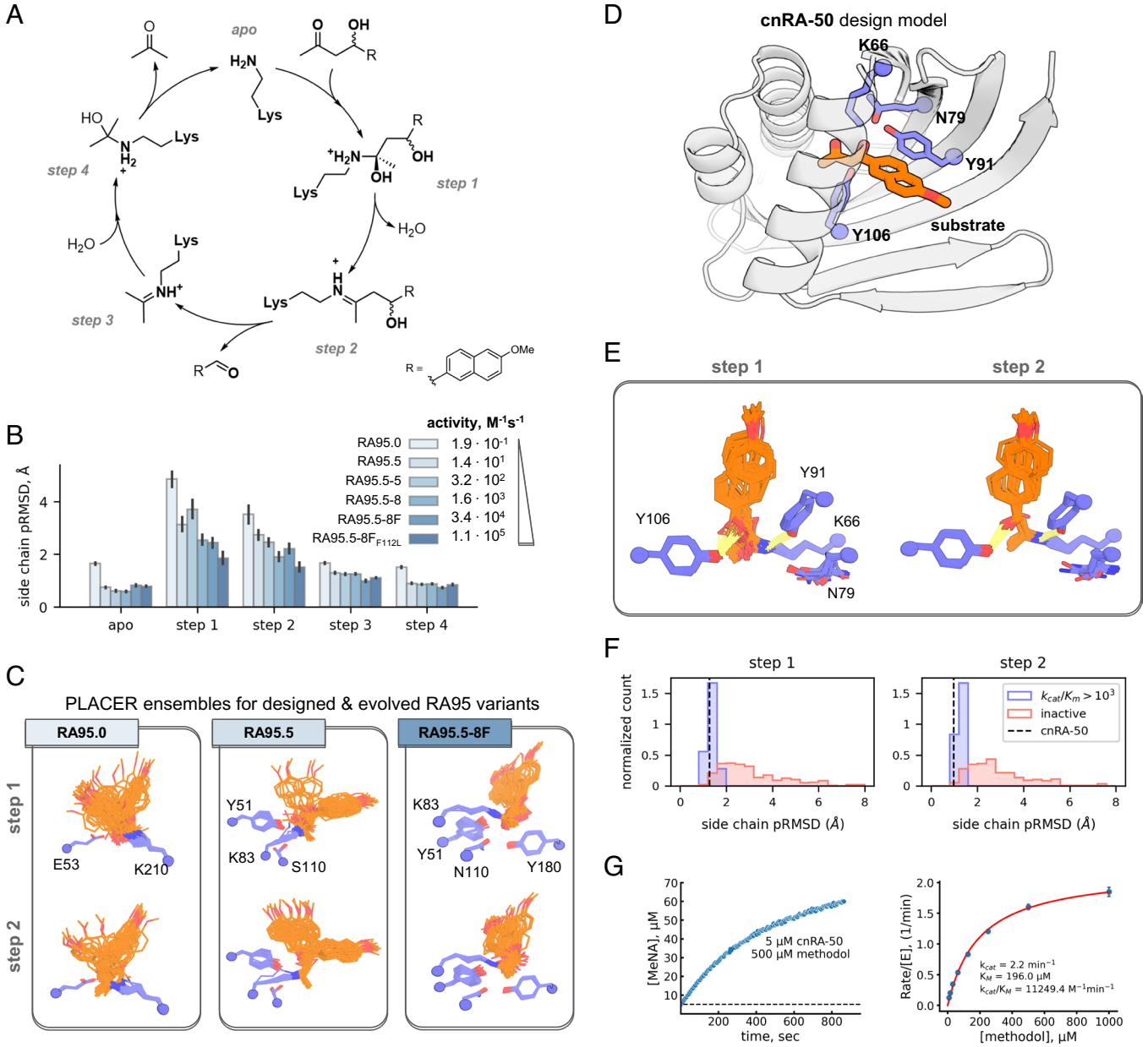

Fig. 4. Selecting designs with preorganized catalytic residues increases activity. (*A*) Mechanism of the retro-aldol reaction with the key intermediates shown. (*B*) Five intermediate states of the RA reaction with methodol were modeled by PLACER for the RA95 series of designs to probe preorganization of the active site lysine and its modifications. (*C*) Examples of PLACER ensembles for active sites of the three previously published retro-aldolases with increasing activity show higher degree of preorganization in steps 1 and 2 along the reaction pathway. (*D*) Design model of the most active variant, cnRA-50, with the catalytic tetrad highlighted in blue and substrate in orange. (*E*) PLACER generated ensembles of lysine–methodol conjugates of catalytic steps 1 and 2 of the active site of cnRA-50. (*F*) PLACER active site preorganization ensemble metrics (*p*RMSD of conjugated lysine atoms, except backbone) in steps 1 and 2 enrich for selecting more active de novo designed retro-aldolase enzymes. (*G*) Retro-aldol reaction catalyzed by cnRA-50: time-course following the formation of 6-methoxy-2-naphthaldehyde (*Left*), and Michaelis–Menten plot (*Right*) obtained from initial velocities.

(Fig. 4*D*); we reasoned that PLACER might be helpful to overcome this difficulty by assisting in identification of preorganized active sites. Designs were generated using RosettaMatch (36) and LigandMPNN (37); full details of the design strategy are provided in the supplemental methods. The activities of the 320 designs were first evaluated in an in vitro transcription translation system (IVTT), enabling rapid identification of promising variants without the need for protein purification. The most active of these were expressed and purified in *Escherichia coli*, and their k_{cat}/K_M values determined by measuring activity as a function of substrate concentration. We then examined how PLACER preorganization correlated with activity over all of the tested designs, focusing on the conformational flexibility of the catalytic lysine conjugated to reaction intermediates in the ensembles at each step in the reaction pathway (Fig. 4*E*). We found that the designs with the highest k_{cat}/K_M values were more preorganized than the low activity designs from the initial IVTT screen (Fig. 4*F* and *SI Appendix*, Figs. S13–S15). The most active design, cnRA-50, displayed one of the highest preorganization levels according to PLACER and had a k_{cat}/K_M of 11,000 M^{-1} min^{-1} (Fig. 4*G*), much higher than earlier computational designs prior to directed evolution, and comparable to recent designs made using RFdiffusion and proteinMPNN (38).

Conclusions

PLACER provides a versatile and rapid approach for generating conformational ensembles of molecules both in isolation and in the context of a protein binding site given the sequence of the protein and the positions of the backbone atoms. Unlike AF3, RF All-Atom, and other protein structure prediction methods, PLACER does not predict protein backbone structure—the benefit is that calculations are considerably more rapid, enabling stochastic generation of conformational ensembles. The use of a consistent atom-level representation for all interactions enables facile extension beyond biomolecules to macrocycles and other complex small molecules. The ability to rapidly generate conformational ensembles for protein–small molecule assemblies should be of considerable utility for computational enzyme design and small molecule binder design efforts: The accuracy with which the intended active sites are recapitulated and the extent of preorganization of the key catalytic/interacting sidechain functional groups can be readily assessed. PLACER assessment of active site accuracy and preorganization considerably improves discovery success rates for multistep serine hydrolases (39) and Zn-dependent metallohydrolases (40), and we describe here the design of retroaldolases with much higher activities than pre–DL designs for this reaction, prior to experimental optimization. We anticipate that PLACER-based ensemble generation will be broadly useful for modeling the structures of complex nonprotein molecules both in isolation and in a protein context, and for evaluating enzyme and protein–small molecule binder designs quite generally.

Materials and Methods

Training on the PDB. Training and validation sets were compiled from structures deposited to the Protein Data Bank (41) before January 12, 2023. We selected X-ray and EM structures with resolution 2.5 Å and better and filtered out entries having ≥10% of unresolved atoms, as well as entries with ≥20 nonstandard residues across all polypeptide chains in the asymmetric unit. These filters reduced the size of the set by 33% from 200,713 to 135,172. We additionally removed entries which share similarity to the test set [we used Astex Non-Native Set (22) as our primary small molecule docking benchmark set] either on the protein or the small molecule levels. A PDB structure was discarded if any of its protein chains had ≥30% sequence identity to any of the protein chains in the test set. Structures containing small molecules with ≥80% Tanimoto similarity to the small molecules in the test set were also excluded. We used Open Babel (42) to first compute the small molecule FP4 fingerprints and then to calculate the Tanimoto coefficients. The remaining 119,916 PDB structures were randomly split into training (112,828) and validation (7,090) sets such that none of the protein chains in the two sets share >30% sequence identity.

During training and validation, we do not split the PDB entries into individual chains or molecules but operate on the whole asymmetric units. We directly parse PDBx/mmCIF files keeping all molecular content; only water molecules are excluded. The Chemical Component Dictionary (19) is used to construct chemical graphs for individual residues and small molecules observed in a PDB entry; nodes in a chemical graph represent individual atoms (only nonhydrogen atoms are considered) and edges represent chemical bonds between atoms. The residue-level graphs are joined together into a single chemical graph for the whole PDB entry. Edges are added to connect subgraphs representing individual residues within a polymer chain (proteins, DNA, and RNA), as well as are edges parsed from the struct_conn data records of the PDBx/mmCIF file. Atomic coordinates are parsed from the PDBx/mmCIF file and are saved on a corresponding node of the chemical graph.

PDB structures were cropped to at most 600 heavy atoms to fit the GPU memory. First, we split all atoms in the input structure into four classes (protein, NA, small molecule, and metal), then randomly choose one of these classes with ratios 1:1:5:1, and finally sample a random heavy atom from the selected class to be the center of the cropped region. The crop is defined as the collection of residues and small molecules in the PDB structure which are closest in space to the residue or the small molecule the crop center belongs to. Atom coordinates in the crop undergo corruption. First, we collect all atoms in the 8-hop neighborhood of the crop center in the chemical graph and add them to the corruption set. Backbone atoms in polypeptide (N, CA, and C), polyribo- and polydeoxyribonucleotide chains (O3′, C3′, C4′, C5′, O5′, and P) which are not part of the corruption set are kept fixed with only a minor Gaussian noise ($\sigma = 0.1$ Å) added to the atom coordinates to make the network more robust to small displacements in atomic coordinates of the backbone atoms (43). All the remaining nonbackbone atoms are included in the corruption set. The chemical graph corresponding to the crop is then split into connected components. If a connected component has fixed backbone atoms, then the corrupted atoms in that component are initialized with the coordinates of the closest backbone atom in the graph (based on the shortest path in the graph). Otherwise, a random atom is picked from the component, Gaussian noise with $\sigma = 1.5$ Å is added to its coordinates, and all other atoms reachable from the selected atom are collapsed to it. Finally, Gaussian noise with $\sigma = 1.5$ Å is applied to all atoms in the corruption set.

For more information, see *SI Appendix*, section 1.1.

Training on the CSD. The PDB-derived datasets were complemented by the crystal structures of small molecules deposited to the Cambridge Structural Database (CSD v5.43; November 2021) (17). We used structures with R-factor 7.5% and better and no disorder and excluded polymeric and organo-metallic entries; the number of heavy atoms per small molecule was limited to the 3 to 80 range. Only molecules which could be successfully parsed by Open Babel were retained. CSD structures with multiple molecules in the asymmetric unit were split into individual molecules, followed by merging of identical molecules with the same canonical SMILES string into one entry for training and validation. Upon merging, we kept 3D coordinates of all the instances to account for alternative conformations during training. Training and validation splits were done using 75% Tanimoto similarity cutoff, yielding 226,684 and 7,116 examples, respectively. During training, molecules were sampled with frequencies inversely proportional to the number of similar molecules in the set at Tanimoto similarity 75%.

Astex Non-Native Benchmark. To evaluate PLACER's performance in ligand docking tasks, we applied it on the Astex non-native benchmark set, consisting of 65 target molecules, placed into their non-native protein structures, with a total of 1,112 unique ligand–protein pairs (22). PLACER was run on each protein–ligand pair, generating 1,000 models. The performance of PLACER was evaluated by ranking the 1,000 models based on three confidence metrics (pRMSD, pLDDT, and pLDDT-pDE), and using RMSD of the highest confidence model as the result.

We further explored the impact of applying physics-based minimization on the PLACER-produced models. For this, we used the minimization protocol available in GALigandDock (25) on all of the produced 1,000 models. The produced models were ranked based on the computed dG metric, and the RMSD of the best model used as the final result.

For more information, see *SI Appendix*, section 2.

PoseBusters Benchmark. We evaluated PLACER's docking performance against AlphaFold3 (12) and RF All-Atom (11) using the 428 protein–ligand complexes from the PoseBusters benchmark (as implemented in the 2023 preprint) (26). Predictions were based on the corresponding crystal structure mmCIF files, with target ligands initialized by randomizing their ground-truth coordinates with the same procedure as used during training (Training on the PDB). Known buffer components and crystallization additives, as listed in the AlphaFold3 study (12), were excluded, while relevant cofactors within 600 atoms of the ligand were retained with their crystallographic coordinates approximately fixed. For each complex, PLACER generated 100 models.

Ligand similarity to the PLACER training set was assessed using Tanimoto coefficients [ECFP4 fingerprints, OpenBabel (42)]. Of the 118 PoseBusters complexes absent from the PLACER training and validation sets, 59 contained ligands with TC ≤ 0.7 forming a more challenging benchmark subset (*SI Appendix*, Table S5 and Fig. S5). Further details are provided in *SI Appendix*, section 3.

Computational De Novo Design of Retroaldolases in NTF2 Fold. Structurally diverse set of 10,000 backbones in NTF2 fold was generated using RF joint inpainting (RFjoint2) (44). We used XML-Matcher protocol (36) to identify locations in NTF2 fold compatible with the placement of the catalytic residues forming retro-aldolase catalytic tetrad (32). Protein backbones with preinstalled catalytic residues were designed in the presence of the substrate using LigandMPNN (37). Self-consistency of the sequence and desired scaffold was determined using AF2 (9) (in single sequence prediction mode, using model 4 ptm and three recycles). AF2 models with lDDT higher than 85 formed a pool of designs subjected to in silico selection strategies including or skipping PLACER-assisted selection.

For more information, see *SI Appendix*, section 5.

Application of PLACER for Design Selection. We used PLACER to assist in picking the most promising designs from the same pool of 17,810 variants relies on a set of geometrical and uncertainty metrics from generating an ensemble of 50 structures with PLACER, predicting interactions between the ligand and the active site residues in the context of AF2 predicted backbones. The obtained metrics included distances between polar atoms of the theozyme side chains and uncertainty (pRMSD) in polar atom positions computed over the ensemble of 50 models. To identify a smaller subset of PLACER metrics, we trained a simple logistic regression model to classify designs as active or inactive on a set of designs characterized in the preliminary experiment performed to identify the focus theozyme configuration.

IVTT-Based Retroaldolase Activity Testing Assay. We developed a protocol for medium-throughput semiquantitative retro-aldolase activity testing of the designs using IVTT. We obtained linear duplex DNA, encoding protein of interest, flanked by appropriately spaced T7 promoter and T7 terminator sequences as eBlocks™ gene fragments from IDT, and by mixing it directly with 2 μL components of PurExpress IVTT system produced enough protein to screen for active variants. Genes of several control proteins with a range of activities previously determined in purified form were included with each batch of tested designs to assist in ranking activity of the new variants, and help to compare experiments run under different experimental conditions. It should be noted that in this format of the experiment, concentration of the soluble protein remains unknown, and therefore, activity level should be treated as apparent activity defined as a product of the amount of soluble protein and intrinsic activity of each variant. Typically, a batch of 90 designs and 6 controls were used to evaluate performance of a particular computational workflow.

Characterization of Individual Variants. To get more accurate and quantitative data describing catalytic activities of the designs and validate IVTT-based assay, we cloned, expressed in *E. coli*, purified by IMAC (Ni-NTA) and size-exclusion chromatography, and determined Michaelis–Menten parameters for 24 designs with the highest apparent activities. Additionally, we purified and tested under similar experimental conditions three previously characterized proteins possessing retro-aldolase activity. This group of control proteins consists of published RA95.5-8F (32), RA110-4.6 (45), and RAβb-16.2 (46) retro-aldolases.

Data, Materials, and Software Availability. Code and neural network weights are available for download on GitHub (https://github.com/baker-laboratory/PLACER) (47). The model files and sequences of designed retroaldolases are available for download at Zenodo (https://doi.org/10.5281/zenodo.14591000) (48). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Luki Goldschmidt and Kandise VanWormer for maintaining the computational and wet lab resources at the Institute for Protein Design. We thank Dr. Declan Evans and Dr. Florence Hardy for helpful conversations during the development of the method and Madison A. Kennedy for reading and editing earlier drafts of the manuscript. We thank Microsoft for generous donation of Azure Compute Credits and Perlmutter Grant NERSC Award BER-ERCAP0022018 for access to the Perlmutter high-performance computing resources. This work was supported by a gift from Microsoft (R.K., D.B., I.A., and J.D.). This work was funded by HHMI (D.B., I.K., and G.R.L.), the Schmidt Futures program (F.D.), the Open Philanthropy Project Improving Protein Design Fund (I.K, G.R.L., Y.K., S.J.P., and J.D.), the Audacious Project at the Institute for Protein Design (L.A. and A.L.), the Washington Research Foundation's Innovation Fellows Program (G.R.L.), Postdoctoral Fellowship Award (S.J.P.), and Translational Research Fund (L.A.), Defense Threat Reduction Agency HDTRA1-19-1-0003 (S.J.P. and A.L.) and HDTRA1-21-1-0007 (G.Z. and I.A.), the National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services, under Contract No.: 75N93022C00036 (I.A.), NIH and/or the National Institute of General Medical Sciences (NIGMS) Award (T32GM008268) (A.L.), and the NSF CHE-2226466 (F.D.).

Author affiliations: [a]Department of Biochemistry, University of Washington, Seattle, WA 98105; [b]Department of Biochemistry, Institute for Protein Design, University of Washington, Seattle, WA 98105; [c]HHMI, University of Washington, Seattle, WA 98105; and [d]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA 98105

1. G. Corso, H. Stärk, B. Jing, R. Barzilay, T. Jaakkola, DiffDock: Diffusion steps, twists, and turns for molecular docking. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2210.01776 (Accessed 3 January 2025).
2. T. A. Halgren *et al.*, Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759 (2004).
3. H. Stärk, O.-E. Ganea, L. Pattanaik, R. Barzilay, T. Jaakkola, EquiBind: Geometric deep learning for drug binding structure prediction. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2202.05146 (Accessed 3 January 2025).
4. W. Lu *et al.*, TANKBind: Trigonometry-aware neural networks for drug-protein binding structure prediction. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.06.06.495043 (Accessed 3 January 2025).
5. O. Ganea *et al.*, Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Adv. Neural Inf. Process. Syst.* **34**, 13757–13769 (2021).
6. M. Xu *et al.*, GeoDiff: A geometric diffusion model for molecular conformation generation. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2203.02923 (Accessed 3 January 2025).
7. B. Jing, G. Corso, J. Chang, R. Barzilay, T. Jaakkola, Torsional diffusion for molecular conformer generation. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2206.01729 (Accessed 3 January 2025).
8. S. Axelrod, R. Gómez-Bombarelli, GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **9**, 185 (2022).
9. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
10. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
11. R. Krishna *et al.*, Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science* **384**, eadl2528 (2024).
12. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* **630**, 493–500 (2024).
13. Z. Qiao, W. Nie, A. Vahdat, T. F. Miller III, A. Anandkumar, State-specific protein–ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **6**, 195–208 (2024).
14. F. Fuchs, D. Worrall, V. Fischer, M. Welling, Se (3)-transformers: 3D roto-translation equivariant attention networks. *Adv. Neural Inf. Process. Syst.* **33**, 1970–1981 (2020).
15. M. Baek *et al.*, Efficient and accurate prediction of protein structure using RoseTTAFold2. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.05.24.542179 (Accessed 3 January 2025).
16. N. Hiranuma *et al.*, Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat. Commun.* **12**, 1340 (2021).
17. C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, The Cambridge Structural Database. *Acta Crystallogr. B, Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
18. P. J. Salveson *et al.*, Expansive discovery of chemically diverse structured macrocyclic oligoamides. *Science* **384**, 420–428 (2024).

19. J. D. Westbrook *et al.*, The chemical component dictionary: Complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics* **31**, 1274–1278 (2015).
20. I. Kalvet *et al.*, Design of heme enzymes with a tunable substrate binding pocket adjacent to an open metal coordination site. *J. Am. Chem. Soc.* **145**, 14307–14315 (2023).
21. G. R. Lee *et al.*, Small-molecule binding and sensing with a designed protein family. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.11.01.565201 (Accessed 3 January 2025).
22. M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn, C. W. Murray, Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model.* **48**, 2214–2225 (2008).
23. V. Y. Tanchuk, V. O. Tanin, A. I. Vovk, G. Poda, A new, improved hybrid scoring function for molecular docking and scoring based on AutoDock and AutoDock Vina. *Chem. Biol. Drug Des.* **87**, 618–625 (2016).
24. M. Baek, W.-H. Shin, H. W. Chung, C. Seok, GalaxyDock BP2 score: A hybrid scoring function for accurate protein-ligand docking. *J. Comput. Aided Mol. Des.* **31**, 653–666 (2017).
25. H. Park, G. Zhou, M. Baek, D. Baker, F. DiMaio, Force field optimization guided by small molecule crystal lattice data enables consistent sub-Angstrom protein–ligand docking. *J. Chem. Theory Comput.* **17**, 2000–2010 (2021).
26. M. Buttenschoen, G. M. Morris, C. M. Deane, PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2024).
27. R. V. Rakotoharisoa *et al.*, Design of efficient artificial enzymes using crystallographically enhanced conformational sampling. *J. Am. Chem. Soc.* **146**, 10001–10013 (2024).
28. A. Broom *et al.*, Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).
29. S. J. Fleishman, S. D. Khare, N. Koga, D. Baker, Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20**, 753–757 (2011).
30. E. A. Althoff *et al.*, Robust design and optimization of retroaldol enzymes. *Protein Sci.* **21**, 717–726 (2012).
31. L. Giger *et al.*, Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
32. R. Obexer *et al.*, Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
33. A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catal.* **7**, 8524–8532 (2017).
34. E. Marcos *et al.*, Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017).
35. A.H.-W. Yeh *et al.*, De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
36. B. D. Weitzner, Y. Kipnis, A. G. Daniel, D. Hilvert, D. Baker, A computational method for design of connected catalytic networks in proteins. *Protein Sci.* **28**, 2036–2041 (2019).
37. J. Dauparas *et al.*, Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **22**, 717–723 (2025).
38. M. Braun *et al.*, Computational design of highly active de novo enzymes. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.08.02.606416 (Accessed 3 January 2025).
39. A. Lauko *et al.*, Computational design of serine hydrolases. *Science* **388**, eadu2454 (2025).
40. D. Kim *et al.*, Computational design of metallohydrolases. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.11.13.623507 (Accessed 3 January 2025).
41. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
42. N. M. O'Boyle *et al.*, Open Babel: An open chemical toolbox. *J. Cheminf.* **3**, 33 (2011).
43. J. Dauparas *et al.*, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
44. D. E. Kim *et al.*, Parametrically guided design of beta barrels and transmembrane nanopores using deep learning. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2425459122 (2025).
45. R. Obexer *et al.*, Active site plasticity of a computationally designed retro-aldolase enzyme. *ChemCatChem* **6**, 1043–1050 (2014).
46. Y. Kipnis *et al.*, Design and optimization of enzymatic activity in a de novo β-barrel scaffold. *Protein Sci.* **31**, e4405 (2022).
47. I. Anishchenko, I. Kalvet, PLACER. Github. https://github.com/baker-laboratory/PLACER. Deposited 28 January 2025.
48. I. Kalvet, Y. Kipnis, I. Anishchenko, D. Baker, Retroaldolase designs from the PLACER manuscript. Zenodo. https://doi.org/10.5281/ZENODO.14591000. Accessed 22 September 2025.