



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Using physical features of protein core packing to distinguish real proteins from decoys

Citation for published version:

Grigas, AT, Mei, Z, Treado, JD, Levine, ZA, Regan, L & O'Hern, CS 2020, 'Using physical features of protein core packing to distinguish real proteins from decoys', *Protein Science*, vol. 25, no. 9, pp. 1931-1944. <https://doi.org/10.1002/pro.3914>

Digital Object Identifier (DOI):

[10.1002/pro.3914](https://doi.org/10.1002/pro.3914)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Protein Science

Publisher Rights Statement:

This is the peer reviewed version of the following article: Grigas, AT, Mei, Z, Treado, JD, Levine, ZA, Regan, L, O'Hern, CS. Using physical features of protein core packing to distinguish real proteins from decoys. *Protein Science*. 2020; 29: 1931–1944. <https://doi.org/10.1002/pro.3914>, which has been published in final form at <https://doi.org/10.1002/pro.3914>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Using physical features of protein core packing to distinguish real proteins from decoys

Alex T. Grigas^{1,2}, Zhe Mei^{2,3}, John D. Treado^{2,4}, Zachary A. Levine^{5,6}, Lynne Regan⁷, and Corey S. O’Hern^{1,2,4,8,9}

¹*Graduate Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, 06520, USA*

²*Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut, 06520, USA*

³*Department of Chemistry, Yale University, New Haven, Connecticut 06520, USA*

⁴*Department of Mechanical Engineering and Materials Science, Yale University, New Haven, Connecticut 06520, USA*

⁵*Department of Pathology, Yale University, New Haven, Connecticut 06520, USA*

⁶*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, 06520*

⁷*Institute of Quantitative Biology, Biochemistry and Biotechnology, Centre for Synthetic and Systems Biology, School of Biological Sciences, University of Edinburgh*

¹*Graduate Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, 06520, USA*

⁸*Department of Physics, Yale University, New Haven, Connecticut 06520, USA*

⁹*Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA*

Manuscript Pages: 19

Total Manuscript Figures: 7

Total Manuscript Tables: 2

Supporting Information Pages: 14

Total Supporting Information Figures/Tables: 15/3

Abstract: The ability to consistently distinguish real protein structures from computationally-generated model decoys is not yet a solved problem. One route to distinguish real protein structures from decoys is to delineate the important physical features that specify a real protein. For example, it has long been appreciated that the hydrophobic cores of proteins contribute significantly to their stability. We used two sources to obtain datasets of decoys to compare with real protein structures: submissions to the biennial CASP competition, in which researchers attempt to predict the structure of a protein only knowing its amino acid sequence, and also decoys generated by 3DRobot, which have user-specified global root-mean-squared deviations from experimentally determined structures. Our analysis revealed that both sets of decoys possess cores that do not recapitulate the key features that define real protein cores. In particular, the model structures appear more densely packed (because of energetically unfavorable atomic overlaps), contain too few residues in the core, and have improper distributions of hydrophobic residues throughout the structure. Based on these observations, we developed a feed-forward neural network, which incorporates key physical features of protein cores, to predict how well a computational model recapitulates the real protein structure without knowledge of the structure of the target sequence. By identifying the important features of protein structure, our method is able to rank decoy structures with similar accuracy to that obtained by state-of-the-art methods that incorporate many additional features. The small number of physical features makes our model interpretable, emphasizing the importance of protein packing and hydrophobicity in protein structure prediction.

Significance: A common problem in the prediction of a protein’s three-dimensional structure from its amino acid sequence is that one can create low-energy computational models that do not match the experimentally determined structure. Here, we identify several features related to packing constraints that differentiate real protein structures from incorrect models. We subsequently use these features to develop a neural network, which is able to distinguish real protein structures from computationally-generated, incorrect structures.

Availability: https://github.com/agrigas115/core_packing_score

Keywords: protein decoys | hydrophobic core | protein structure prediction | protein design

1 Introduction

It remains a grand challenge of biology to design proteins that adopt user-specified structures and perform user-specified functions. Although there have been significant successes,¹⁻¹¹ the field is still not at the point where we can robustly achieve this goal for any application.¹² An inherent problem in protein structure prediction and design is that it is extremely difficult to distinguish between computational models that are apparently low energy,¹³ but which are different from the real, experimentally determined structures.¹⁴⁻¹⁶ This problem is known as “Decoy Detection”. For example, in recent Critical Assessment of protein Structure Prediction (CASP) competitions, in which researchers attempt to predict the three-dimensional (3D) structure of a protein, based on its amino acid sequence, many groups produced impressively accurate predictions for certain targets (Fig. 1 (A)). However, for most targets there is a wide spread of prediction accuracy across the submissions from different groups. (Note that even groups that perform well on average have a large standard deviation for their predictions, see Fig. S2 in the Supporting Information.)

In recognition of this issue, there is a subcategory in CASP, Estimation of Model Accuracy (EMA), in which researchers aim to rank order the submitted models according to their similarity to the backbone of the target structure. The challenge is that researchers must develop a scoring function for determining model accuracy, yet they do not have access to the target structure.¹⁷⁻²³ Although EMA methods are improving,²⁴⁻³⁵ they are still unable to consistently rank models submitted to CASP in terms of their similarity to the target structure.²³

We used two datasets for protein decoys: predicted structures submitted to CASP and decoys generated using 3DRobot.³⁶ Protein decoys from CASP were selected for two important reasons: 1. The decoys represent a large sampling of the types of decoys created from many different protein structure prediction methods; and 2. Decoys generated by protein structure prediction methods are the main test case for the application of decoy detection. Unfortunately, the CASP dataset is limited in sequence diversity, compared to the diversity found in experimentally solved protein structures. Therefore, to capture a larger sampling of sequence space, we generated decoys from a high resolution x-ray crystal structure database using the 3DRobot algorithm.³⁶ The range of sizes of the proteins and a measure of the backbone accuracy of the decoys are plotted in Fig. 1 (A) and (C). (See Materials and Methods for a detailed description of the computationally-generated decoy datasets.) While there are few near-native structures in the CASP dataset, there are many structures in the 3DRobot dataset that can be superimposed on the x-ray crystal target structure with root-mean-square deviations in the C_α positions $< 0.5 \text{ \AA}$.

The protein core has long been known to determine protein stability and provide the driving force for folding.³⁷⁻⁴⁵ Additionally, in our previous work, we have found that several features of core packing are universal among well-folded experimental structures, such as the repacking predictability of core residue side chain placement, core packing fraction, and distribution of core void space.⁴⁶⁻⁵¹ This work suggests that analysis of core residue placement and packing in proteins more generally should be effective in determining the accuracy of protein decoys. Indeed, software to assess x-ray crystal structure model quality often calculate interatomic overlaps,^{53,54} the RosettaHoles software uses defects in interior void space to differentiate between high resolution x-ray crystal structures and protein decoys,⁵² VoronoiQA scores protein decoys using a statistical potential based on Voronoi contact areas,³⁴ and many other decoy detection methods attempt to incorporate predictions of solvent accessibility.^{28-30,35,55} Nevertheless, a minimal set of physical features that can determine protein decoy accuracy has not yet been identified.

We demonstrate, that for recent CASP competition predictions, as well as 3DRobot decoys, we can determine protein decoy accuracy solely by identifying the structures that place the correct residues in the protein core. We also show that only predicted structures that place core residues

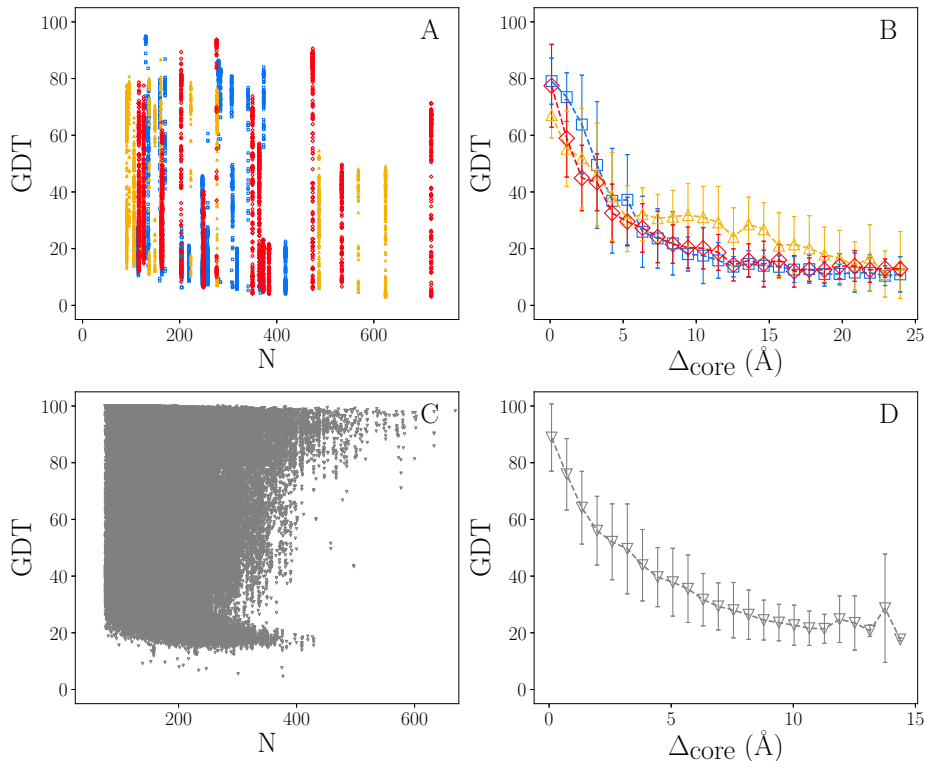


Figure 1: (A) Scatter plot of the Global Distance Test (GDT) score, which gives the average percentage of C_{α} atoms that is within a given cutoff distance to the target (averaged over four cutoff distances), versus the number of residues N in the target structure for free-modeling submissions to CASP11 (blue squares), CASP12 (orange triangles), and CASP13 (red diamonds). (B) GDT plotted versus the root-mean-square deviations (RMSD) in \AA among C_{α} atoms of core residues defined in the target (Δ_{core}). The symbols represent the average in each Δ_{core} bin and the error bars represent one standard deviation. (C) Scatter plot of GDT versus N for the decoys generated by the 3DRobot algorithm (grey inverted triangles). (D) GDT versus Δ_{core} showing the average and standard deviation in each Δ_{core} bin for the 3DRobot decoys.

accurately, measured using the root-mean-squared deviation of the C_{α} atoms of solvent inaccessible residues (i.e. $\Delta_{\text{core}} < 1\text{\AA}$), can achieve high Global Distance Test (GDT) scores (GDT $\gtrsim 70$) (Fig. 1 (B) and Fig. 1 (D)), where GDT ranges from 0 to 100 and 100 is a perfect match to the target structure.⁵⁶ Computational models with a larger GDT are ‘more native-like’, whereas models with a lower GDT are ‘more decoy-like.’ Motivated by these observations, we then analyzed several important attributes of the *cores* of both experimentally-observed and computationally generated protein structures. We demonstrate that 92% of all structures in the high resolution x-ray crystal structure database fall within the packing feature cutoffs we selected and that a majority of all accurate decoy structures do as well. Using these results, we developed a decoy detection method based on only six principal features of protein packing that are independent of the target structure. Our method is more effective than many of the methods in the CASP13 EMA and is comparable to the other decoy detection methods that we tested in this work. Moreover, all of the methods used in the CASP13 EMA employ a far greater number of features than we do.⁵⁷ For example, one of the top performing methods in the CASP13 EMA, ProQ3²⁹ uses 336 different features. Similarly, other successful methods, such as 3DCNN,⁵⁸ use all of the atomic coordinates of the protein structures as

input features to machine learning methods for decoy detection. While CASP EMA often focuses on predicting only the best model in a group, here we focus on predicting the continuous quantity, GDT, for computational models as a means of assessing the importance of protein core packing for determining the global structure. The effectiveness of the small number of features in our approach highlights the importance of core residues, which take up $\lesssim 10\%$ of globular proteins on average, and packing constraints in determining the global structure of proteins.

2 Results

First, we identify several key features that distinguish high resolution x-ray crystal structures and computationally-generated decoys, such as the average core packing fraction, core overlap energy, fraction of residues positioned in the core, the distribution of the packing fraction of hydrophobic residues throughout the protein, and the average hydrophobicity of the core residues. We then show how these features can be used to predict the GDT of CASP submissions and 3DRobot decoys, independent of knowing the target structure.

The distribution of packing fraction ϕ of core residues in proteins whose structures are determined by x-ray crystallography occur over a relatively narrow range, with a mean of 0.55 and a standard deviation of 0.02.^{46,48,51} We define core residues as those with small values of the relative solvent accessible surface area, $rSASA \leq 10^{-3}$. (See the Materials and Methods section for a description of the database of high resolution protein x-ray crystal structures and definition of rSASA.) In contrast, we find that many of the CASP submissions and 3DRobot decoys possess core residues with packing fractions that are much higher than those in experimentally determined proteins structures. One way to achieve such an un-physically high packing fraction would be to allow large atomic overlaps. We therefore analyzed the side-chain overlap energy for core residues, using the purely repulsive Lennard-Jones inter-atomic potential,

$$U_{\text{RLJ}} = N_a^{-1} \sum_{i,j} \frac{\epsilon}{72} \left(1 - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)^2 \Theta \left(1 - \frac{r_{ij}}{\sigma_{ij}} \right), \quad (1)$$

where the sum is taken over all side-chain atoms i and all other atoms not part of the same residue j , N_a is the number of atomic overlaps, ϵ defines the energy scale, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, σ_i is the diameter of atom i , r_{ij} is the distance between the centers of spherical atoms i and j , and $\Theta(x)$ is the Heaviside step function, which is 1 when $x > 0$ and is 0 when $x \leq 0$. For high resolution x-ray crystal structures, half of core residues have an overlap energy of zero; the remaining half of the residues have very small overlap energies with an average value of $U_{\text{RLJ}}/\epsilon \approx 10^{-4}$ (Figs. 2 (A) and (C)). In contrast, the model structures in the CASP and 3DRobot datasets include some extremely high energy residues, with $U_{\text{RLJ}}/\epsilon \sim 10^{16}$. The absence of data points in the lower right-hand corner of Fig. 2 (A) and (B) clearly highlights that artificially high packing fractions are only found when the overlap energy is high. Additionally, while large atomic overlaps can lead to high packing fractions, there are also many cases of significantly overpacked residues with relatively small overlap energies near the upper bound of x-ray crystal structures. Further, there are residues with overlap energies many orders of magnitude above those for x-ray crystal structures, but with packing fractions near those expected for x-ray crystal structures. In Fig. 2 (C), we show the frequency distribution of packing fractions for core residues with $U_{\text{RLJ}} = 0$. The differences in peak heights reflect how much more likely it is for core residues from x-ray crystal structures of proteins to have zero overlap energy compared to those in the CASP submissions and 3DRobot decoys. However, the tail of large packing fractions in Fig. 2 (B) is smaller for 3DRobot structures than that for the CASP datasets.

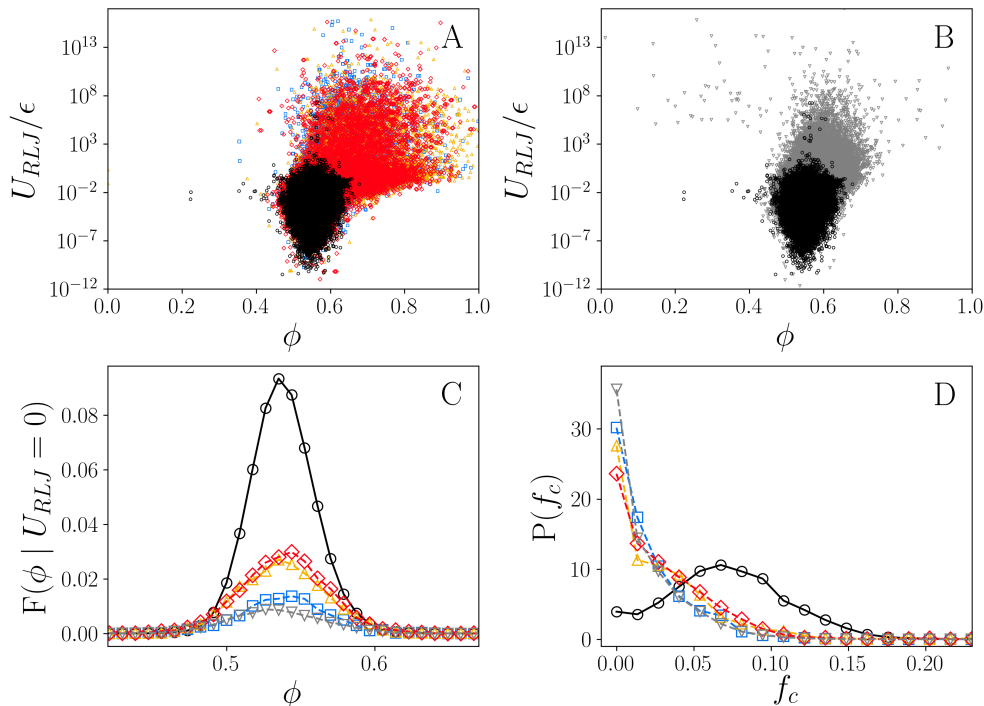


Figure 2: Three packing features of high resolution x-ray crystal structures (black circles) and free-modeling submissions to CASP11 (blue squares), CASP12 (orange triangles), and CASP13 (red diamonds), as well as 3DRobot decoys (grey inverted triangles). (A) Purely repulsive Lennard-Jones potential energy U_{RLJ} that measures the overlap of core residue ($rSASA \leq 10^{-3}$) sidechain atoms versus packing fraction ϕ comparing CASP decoys and x-ray crystal structures, excluding CYS residues. (B) U_{RLJ} versus ϕ comparing 3DRobot decoys and x-ray crystal structures. (C) Frequency distribution of the packing fraction $F(\phi | U_{RLJ} = 0)$ for core residues with zero overlap energy. (D) Probability distribution $P(f_c)$ of the fraction of core residues f_c ($rSASA \leq 10^{-3}$).

These results demonstrate that individual core residues in the computational models submitted to CASP and generated by 3DRobot are often overpacked (with packing fractions above those expected for x-ray crystal structures). We then asked whether core overpacking is related to the number of residues in the core relative to the number of residues in the protein. In Fig. 2 (D), we plot the probability that a structure, either computationally-generated or experimentally-determined, has a given fraction of its total number of residues in the core ($rSASA \leq 10^{-3}$). It is clear from this plot that computationally-generated models often have too few residues in the core. Additionally, the 3DRobot-generated structures are slightly more likely than the CASP structures to have no core residues. Thus, the computationally-generated models not only possess cores with un-physically high packing fraction and overlap energy, but they also, typically, have a smaller fraction of residues in the core compared to x-ray crystal structures of proteins.

After investigating the packing of amino acids in protein cores, we then looked at features to describe their hydrophobicity. Many CASP models and 3DRobot decoys have too few residues in the core; how does this affect the distribution of hydrophobic residues outside of the core? We examined the degree to which the packing fractions of all hydrophobic residues in a given protein deviate from the expected distribution from high resolution x-ray crystal structures.^{59,60} (See Fig. 3 (A).) Specifically, we measured the Kullback-Leibler (KL) divergence (D_{KL}) between the

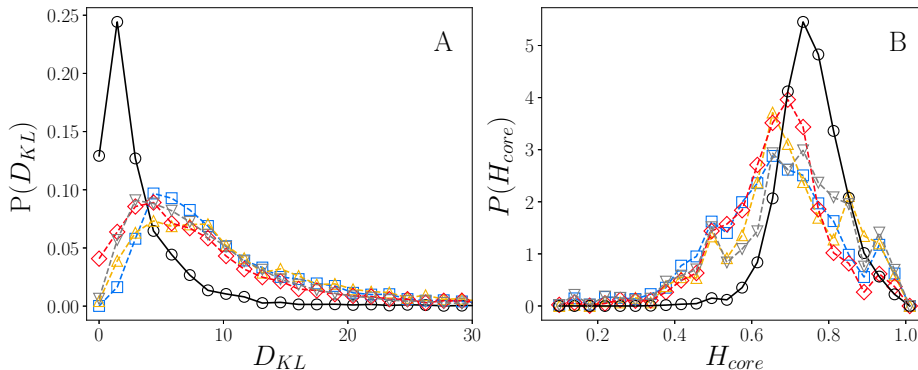


Figure 3: Two additional packing features of high resolution x-ray crystal structures (black circles) and free-modeling submissions to CASP11 (blue squares), CASP12 (orange triangles), and CASP13 (red diamonds), as well as 3DRobot decoys (grey inverted triangles). (A) Probability distribution $P(D_{KL})$ of the Kullback-Leibler divergence D_{KL} from the distribution of the packing fractions of all hydrophobic residues in high resolution x-ray crystal structures. (B) Probability distribution of the average hydrophobicity H_{core} of core residues (with $rSASA \leq 10^{-3}$).

overall distribution of packing fractions of hydrophobic residues from a database of high resolution x-ray crystal structures, and each individual structure’s packing fraction distribution for all its hydrophobic residues in that database.⁶¹ (See Supporting Information and Fig. S5 for more details.) Additionally, we measured the D_{KL} for all computational models against the distribution from the database of high resolution x-ray crystal structures. We find that the distribution of packing fractions of hydrophobic residues for each individual experimentally-observed protein structure is similar to the full distribution, whereas the distributions for the computationally-generated structures differ significantly from the experimentally observed distribution.

Finally, we also measured the average hydrophobicity of core residues (Fig. 3 (B)) for the computationally-generated decoys and x-ray crystal structures. As there are many hydrophobicity scales H for amino acids in the literature, we considered seven of them, normalized them between 0 and 1, and took the average for each residue.⁶² We find that many decoy structures in both the CASP and 3DRobot datasets have core residues (with $rSASA \leq 10^{-3}$) that are less hydrophobic (with $H_{core} < 0.5$) than any protein x-ray crystal structure, which typically have cores with $H_{core} \gtrsim 0.6$. In the above discussion, we examined the packing features of free-modeling submissions to CASP, which are predicted only from the amino acid sequence. However, we found the same results for CASP structures in the template-based modeling submissions, which are generated using given template structures. (See Fig. S6 in Supporting Information.) We therefore included both free-modeling and template-based modeling structures together in the analyses of CASP data below.

Before developing a predictive model for decoy detection, we investigated the correlation between the accuracy of backbone placement and correct identification of core residues. In Fig. 4, we plot the average GDT versus the fraction $f_{correct}$ of the predicted core residues ($rSASA \leq 10^{-1}$) that are core residues in the target structure. This plot shows that there is a strong correlation between the accuracy of backbone placement and correct identification of the core residues. In particular, when $f_{correct} \rightarrow 1$, the average GDT $\gtrsim 80$. However, one does not know the correct set of core residues at the time of the prediction. Yet, the core residues should share the features shown in Figs. 2 and 3. Therefore, we should be able to predict the GDT of a model based upon

how well the core properties and the distribution of the hydrophobic residues match those of high resolution x-ray crystal structures.

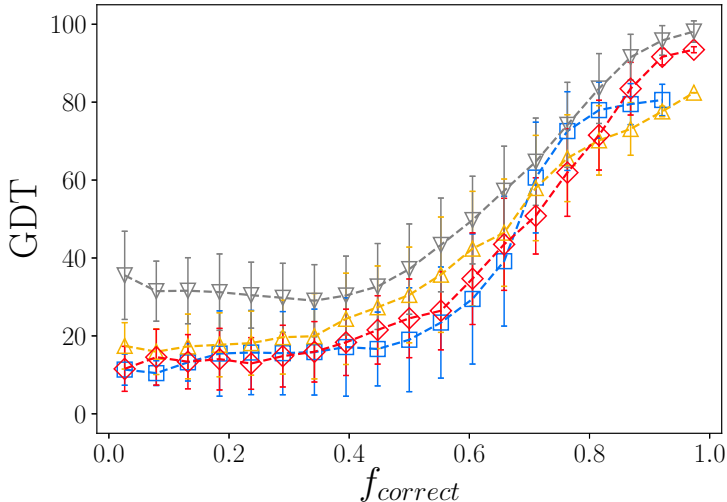


Figure 4: The average GDT of CASP predictions that correctly identify each given fraction of near core residues with $rSASA \leq 10^{-1}$, $f_{correct}$, for CASP11 (blue squares), CASP12 (orange triangles), CASP13 (red diamonds) and 3DRobot (grey inverted triangles) structures. Error bars represent one standard deviation.

The simplest approach to test this hypothesis is to filter the decoy datasets by identifying those predictions that fall within the range of the packing features sampled by proteins in the x-ray crystal structure database. In Fig. 5, we plot the fraction f_{pass} of decoy structures within a given GDT bin whose packing features fall within the ranges found for x-ray crystal structures for all five of the packing features. We used the following cutoffs: $0.5 \leq \langle \phi \rangle \leq 0.62$, $\langle U_{RLJ}/\epsilon \rangle \leq 10^2$, $f_c \geq 0.02$, $D_{KL} \leq 15$, and $H_{core} \geq 0.5$. These cutoffs capture 92% of the x-ray crystal structures, and nearly all of the high quality predictions in the CASP and 3DRobot decoy datasets. If this approach provided a perfect classifier, f_{pass} versus GDT would be sigmoidal with a steep rise at a given GDT threshold that separates “good” and “bad” predictions. In the case of the 3DRobot decoys, the curve is less linear and, for example, only 15% of 3DRobot structures with a GDT of 40 fall within these cutoffs. However, in CASP13, nearly 55% of predictions with a GDT of 40 fall within the cutoffs for the x-ray crystal structure packing features. Thus, there is an abundance of CASP predictions with reasonably sized cores containing hydrophobic residues that are well-packed with only small atomic overlaps, but possess the wrong core residues, resulting in low values for the GDT. Decoys with well-packed hydrophobic cores that do not contain the correct core residues from the target do not occur frequently for 3DRobot decoys, which may be expected, as the 3DRobot algorithm perturbs an existing x-ray crystal structure that has the correct core residues.

A common test of a classifier is to calculate the Receiver Operating Characteristic (ROC) curve, which compares the false positive and true positive rates for classifying good and bad protein predictions (defined over a range of GDT thresholds). To do this, we first need to report the probability that a prediction is either good or bad. We therefore defined a decoy score by assuming a Gaussian at the mean values of the packing features obtained from the x-ray crystal structures with the width defined by the standard deviation of the x-ray crystal structure packing features. Evaluating the Gaussian for each of the five features and then averaging over the features provides a score that ranges from 0 to 1, quantifying how similar a decoy is to a typical x-ray crystal

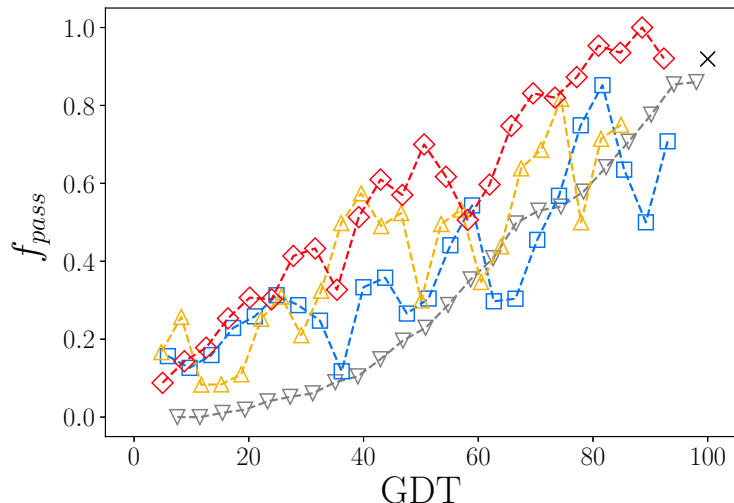


Figure 5: Fraction of decoys f_{pass} in a GDT bin that are within the cutoffs for the x-ray crystal structure packing features for submissions to CASP11 (blue squares), CASP12 (orange triangles), and CASP13 (red diamonds) and 3DRobot structures (grey inverted triangles). The fraction of x-ray crystal structures that fall within the packing feature cutoffs is represented as an x (at $f_{\text{pass}} = 0.92$).

structure. Additionally, we need to assign classes for good and bad predictions. There is no set cutoff of GDT, therefore, we measured the area under the ROC curve (AUC) over a range of GDT cutoffs. (See Fig. S7 (B) in Supporting Information.) A classifier with an AUC of 0.5 offers random discrimination and a classifier with an AUC of 1 offers perfect discrimination. Averaging the AUC for GDT cutoffs between 40-70 resulted in average AUC values of 0.7 for CASP11, 0.67 for CASP12, 0.8 for CASP13, and 0.83 for 3DRobot, matching the expected performance based on Fig. 5.

Many decoy detection methods employ features that evaluate sidechain placement to determine the backbone accuracy. However, the sidechain atoms in the computational model may not be placed in the optimal conformation given the placement of the backbone. Therefore, several decoy detection methods apply collective sidechain repacking before evaluating decoy structures.^{28,29,58} Since our method involves features based on sidechain placement, we also investigated the changes in the features after applying collective sidechain repacking using SCWRL,⁶³ which has been used in other decoy detection methods.⁵⁸ Our preliminary results after applying collective sidechain repacking with SCWRL show no significant changes in the packing features or classification performance in differentiating good from bad computational models. (See Figs. S8-S10 in Supporting Information)

While we have shown that many of the predicted structures submitted to CASP and generated by 3DRobot do not recapitulate the packing properties of high resolution protein x-ray crystal structures, and that drawing *ad hoc* cutoffs on the packing features based on x-ray crystal structures offers reasonable discrimination between good and bad predictions, we have not yet developed a model that can accurately predict the GDT based only on a protein’s packing features. To do this, we developed a neural network based on the five packing-related features in Figs. 2 and 3, plus the number, N , of residues in the protein, to construct the GDT function. (We included N to account for larger fluctuations in packing properties that occur for small N ⁴⁶). We implemented a self-normalizing feed-forward neural network with six layers, dropout, and regularization.⁶⁴ We

selected the logcosh error in GDT as the loss function. (For more details, see the Supporting Information.) We used submissions from CASP11, CASP12, and a database of high resolution x-ray crystal structures,^{59,60} as well as a random sampling of 3DRobot structures (where the number of decoys matched the number of x-ray crystal structures) as the training data with five-fold cross validation. The GDT model generated by the neural network was then tested on CASP13 submissions. Overall, the training set included 21,019 structures and the test set included 5,532 structures. Our GDT model achieved a Pearson correlation of 0.72 ± 0.003 , a Spearman correlation of 0.72 ± 0.002 , a Kendall Tau of 0.53 ± 0.001 , and an average absolute error of 15.2 ± 0.4 GDT. This model was then extended to include the same features as above, as well as the dependence of several of the features (f_c , $\langle\phi\rangle$, $\sigma(\phi)$, $\langle U_{RLJ}\rangle$, H_{core} , and $\sigma(H_{\text{core}})$) on rSASA, i.e. with bins $10^{-3} \leq \text{rSASA} < 10^{-2}$ and $10^{-2} \leq \text{rSASA} < 10^{-1}$ as well as $\text{rSASA} < 10^{-3}$. This new model includes 20 packing features. Training and testing the neural network in the same manner as above resulted in slightly improved statistics for the predicted GDT: a Pearson correlation of 0.75 ± 0.008 , a Spearman correlation of 0.76 ± 0.006 , a Kendall Tau of 0.56 ± 0.005 , and an average absolute error of 12.9 ± 0.4 .

For comparison, we tested five other methods (VoroMQA,³⁴ SBROD,²⁷ 3DCNN,⁵⁸ ProQ2,²⁸ and ProQ3²⁹) against CASP13. (See Table 1.) There are two main approaches to decoy detection: Consensus methods and single-ended methods. Consensus methods take a pool of models as input and then output a score for each model, using the consensus of the pool of models to score the structures, whereas single-end methods take only the single model as input to generate a score. We developed a single-ended method, and therefore we only compared our results to those from other single-ended methods. Single-ended methods can be grouped into several classes: 1) methods based on statistical potentials, 2) machine learning methods based on summarizing features of the model structures, and 3) machine learning methods based on the atomic coordinates of the model structures. Therefore, we tested top-ranking methods from each of these categories: A machine learning (ridge regression) score based on backbone orientation (SBROD), a statistical potential based on Voronoi contact areas (VoroMQA), a machine learning method based on summarizing features (ProQ2/ProQ3), and a machine learning method based on atomic coordinates (3DCNN). Additionally, even though our model was trained for regression, the predicted GDT values can be treated as confidence of the quality of whether a decoy falls into two classes (i.e. either good or bad). We can therefore draw an ROC curve and calculate the AUC. The results for the ROC AUC are plotted over a range of GDT cutoffs that define the two classes in Fig. 6 (A). Overall, our model achieves similar results to the other methods tested. For further comparison, the best absolute GDT loss reported in the CASP13 EMA competition was 7 GDT and the average GDT loss across all methods was 15.⁵⁷

We also tested the core/near-core packing method on the 3DRobot decoy dataset. We split the x-ray crystal structure database into fifths and included four-fifths of the x-ray crystal structures and their corresponding 3DRobot structures in the training set, along with CASP11 and CASP12 decoys, and tested on the remaining 3DRobot structures. This procedure ensured there were no 3DRobot structures in the test set that were based on an x-ray crystal structure in the training set, resulting in five training sets with 44,927 structures each and test sets with 6,026 structures each. We then performed five-fold cross validation on each of these five sets of training data. Our core/near-core model with 20 packing features achieved similar performance on the 3DRobot dataset as it did on the CASP13 dataset with a Pearson correlation of 0.8 ± 0.008 , a Spearman correlation of 0.79 ± 0.008 , a Kendall Tau of 0.6 ± 0.008 , and an average absolute error of 13.7 ± 0.6 . The results for VoroMQA, SBROD, 3DCNN, ProQ2, and ProQ3 are shown in Table 2. The ROC AUC, calculated over a range of GDT cutoffs, is plotted in Fig. 6 (B) for each method. For scatter plots of the predicted scores versus the true GDT for both CASP13 and 3DRobot decoy datasets,

see Figs. S13 and S14 in the Supporting Information.

Here, we focus on predicting GDT, since we want to determine whether specifying the key properties of the protein core is sufficient to determine its global properties. However, there are several other methods for quantifying a ground truth score of the accuracy of a computational model. In contrast to global alignment methods, such as GDT, the Local Distance Difference Test (LDDT)⁶⁵ evaluates the local accuracy of distances between heavy atoms in a given residue and heavy atoms in nearby residues, which can then be averaged to obtain a global accuracy score. Therefore, we also tested our method by training on GDT and predicting LDDT and vice versa, as well as evaluated the other tested methods against LDDT. Overall, we found very similar results when testing on LDDT. (See Fig. S15 and Tables S1-S3 in the Supporting Information.)

Method	Pearson	Spearman	Kendall Tau	Avg Error	AUC
Cutoffs	-	-	-	-	0.7
Core Packing	0.72	0.72	0.53	15.2	0.85
Core\Near-Core Packing	0.75	0.75	0.56	12.9	0.89
VoroMQA	0.76	0.78	0.58	17.2	0.9
SBROD	0.8	0.8	0.58	17.24	0.9
3DCNN	0.82	0.82	0.63	12	0.94
ProQ2	0.8	0.82	0.63	27.2	0.93
ProQ3	0.83	0.84	0.63	17.7	0.95

Table 1: Performance of all of the tested methods on the CASP13 dataset. To estimate an average error for VoroMQA, SBROD, ProQ2, and ProQ3, the predicted scores were normalized so that they ranged from 0 to 1. The AUC depends on the cutoff that defines a good versus a bad prediction. Thus, the AUC values were averaged over GDT cutoffs from 40 to 70.

Method	Pearson	Spearman	Kendall Tau	Avg Error	AUC
Cutoffs	-	-	-	-	0.83
Core\Near-Core Packing	0.8	0.79	0.61	13.2	0.91
VoroMQA	0.87	0.87	0.69	14.3	0.95
SBROD	0.81	0.81	0.61	17.6	0.93
3DCNN	0.93	0.93	0.77	18	0.98
ProQ2	0.76	0.78	0.58	14.8	0.91
ProQ3	0.74	0.75	0.55	15.6	0.9

Table 2: Performance of all of the tested methods on the 3DRobot decoy dataset. VoroMQA, SBROD, ProQ2, and ProQ3 return scores that do not range from 0 to 1. To estimate an average error, the predicted scores were normalized so that they fall within 0 to 1. The AUC values were averaged over GDT cutoffs from 40 to 70.

We also investigated the importance of each feature in the core/near-core packing model. To do this, we randomly permuted the values of a given feature after training. This procedure decorrelates each structure with its feature value to effectively remove that feature from the model. In Fig. 7, we display the Pearson correlation between the predicted and actual GDT following feature permutations when tested on CASP13 using the core/near-core packing model, averaged over the five-fold cross validation. Including the near-core (rSASA-dependent) features resulted in very small changes in the Pearson correlation and therefore for visualization, only the core features are plotted. The two largest changes that result from permutation of a single feature come from permuting either

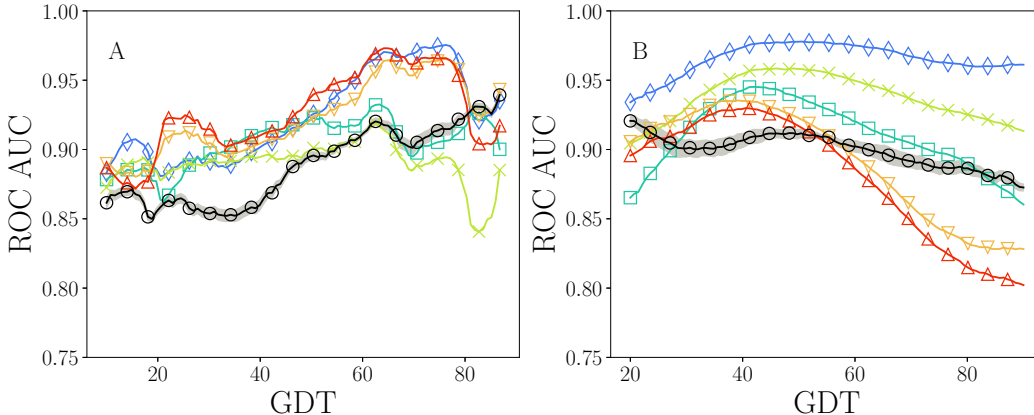


Figure 6: Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for different cutoffs of GDT that define a good versus bad prediction in the (A) CASP13 and (B) 3DRobot decoy datasets for our core/near-core packing model (average (black circles) and standard deviation (grey ribbon) from cross-validation), VoroMQA (green X), SBROD (cyan squares), 3DCNN (blue diamonds), ProQ2 (orange inverted triangles), and ProQ3 (red triangles).

the fraction of core residues or the KL divergence from the hydrophobic residue packing distribution, leading to Pearson correlations of 0.62 and 0.44, respectively. Also, permuting both of these features together leads to the largest pair-wise drop in the Pearson correlation to 0.24. These results indicate that the most important pair of features to include in protein decoy detection are the fraction of core residues and packing fraction distribution of hydrophobic residues. The packing fraction and overlap energy of core residues are slightly less important features. We believe this is because including the wrong residue in the core will give rise to a low GDT (Fig. 4), even if the packing fraction and overlap energy of the misplaced residues are typical of those for core residues in high resolution protein x-ray crystal structures.

3 Discussion

We have identified several important features characterizing protein packing that allow us to distinguish protein decoys from experimentally realizable structures. We developed a machine learning model, using a self-normalizing feed-forward neural network on a small number of packing features, that is able to predict the GDT of CASP13 and 3DRobot structures with high accuracy and without knowledge of the target structures. In addition to developing a highly predictive model, this work also demonstrates the importance of the core and packing constraints for protein structure prediction and points out potential improvements to current prediction methods by properly modeling protein cores. In terms of predicting the GDT and differentiating between good versus bad computationally-generated structures, our model achieved results ranging from slightly worse to as good as other methods, depending on the evaluation metric and test set, using significantly fewer features. For example, ProQ3 employs 336 features, VoroMQA and SBROD include features for each residue in each protein, and 3DCNN uses all of the protein coordinates as input features. Importantly, the machine learning model we developed can be used to identify protein decoys beyond those generated by CASP and 3DRobot. For example, molecular dynamics (MD) simulations are often used to analyze thermal fluctuations in folded proteins. To what extent do the protein conformations sampled in such MD simulations recapitulate the packing properties of

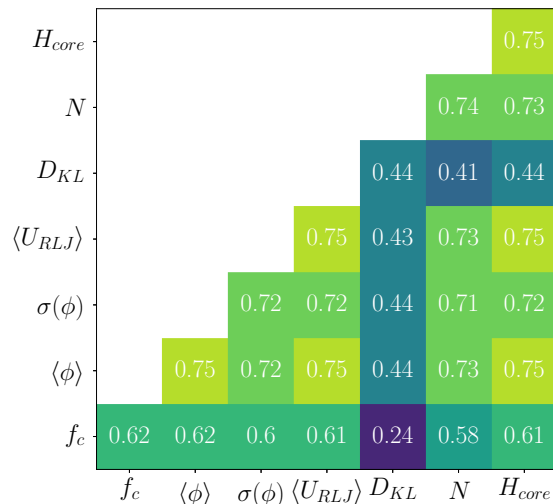


Figure 7: Pearson correlation coefficients between the predicted and actual GDT of CASP13 structures following permutations of single features (along the diagonal) and pairs of features (for the off-diagonal components). The color ranges from purple (minimum) to yellow (maximum) corresponding to the values of Pearson correlation coefficient.

experimentally observed protein structures⁶⁶? The model developed here can be used in concert with MD simulations to filter out unphysical conformations, which will have low values of GDT, without using knowledge of the experimentally observed protein structure. In future studies, we will develop guided molecular dynamics simulations in which we select the frame with the largest predicted GDT from a given trajectory, spawn a large number of simulations with this initial structure and random initial velocities, run the trajectories for a specified time, select the frame with the largest predicted GDT, and continue this process for a given number of cycles. This simulation technique is designed to identify native-like structures, even though the experimental x-ray crystal structure is not known. Additionally, our model can be used to assist protein design methods by selecting designs that are more likely to be experimentally realized. We also showed that when the correct residues are placed in the core, the GDT approaches 100. This result suggests that MD simulations that constrain the correct residues to be in the core may perform better in protein structure prediction applications.

Our analysis has also identified important differences between the CASP and 3DRobot decoys. We found that the CASP datasets include a number of false positives, i.e. decoys with packing features that are similar to those in x-ray crystal structures, but possess a low GDT. However, we do not find many false positives for the 3DRobot decoy dataset. This result is emphasized in the scatter plots in Fig. S14 in the Supporting Information, where there are few structures with a low GDT, but are predicted to be a good structure by any method. Thus, the 3DRobot algorithm does not tend to create well-packed cores with the wrong core residues. Additionally, we note that the accuracy of all methods tested differs when tested on the two computationally-generated decoys datasets. For example, VoromQA, SBROD, 3DCNN, and our packing method all performed better on the 3DRobot test set than on the CASP13 test set. In contrast, for ProQ2 and ProQ3, the performance decreased when testing against 3DRobot-generated decoys compared to CASP decoys. Thus, future studies are necessary to understand the important differences between decoy datasets and also understand why certain methods perform differently on different decoy datasets.

We expect future improvements to our basic model will increase its accuracy. For example, we have shown that the identification of core residues is one of the most important aspects for determining a predicted structure’s accuracy. Thus, we will also implement recurrent neural networks to predict the rSASA values for each residue.^{67,68} This model can then be concatenated with the model developed here. Additionally, machine learning methods can be used to predict the particular f_c for each amino acid sequence to estimate the difference, Δf_c , between the prediction and the actual value. In future studies, we will employ local machine learning and data representation techniques, such as graph neural networks, to assess the structure of each residue individually in computational models. It will also be informative to study in greater depth cases where there are large deviations in GDT. For example, investigating examples of high predicted GDT, but low actual GDT (or *vice versa*) will provide key insights into native protein structures.

4 Materials and Methods

Datasets

In the main text, we show results for the free-modeling CASP submissions. The corresponding results for template-based modeling data are provided in the Supporting Information. (See Fig. S6.) For the decoy datasets, we examined CASP11 (2014),⁶⁹ CASP12 (2016),⁷⁰ and CASP13 (2018)¹⁴ downloaded from the `predictioncenter.org` data archive. Each target in the competitions has a corresponding experimental structure. We selected targets with an x-ray crystal structure under a resolution cutoff. A cutoff of ≤ 2.0 Å was used in the cases of CASP11 and CASP12, however; a cutoff of ≤ 2.7 Å was used for CASP13, as very few protein targets fell under ≤ 2.0 Å. These cutoffs resulted in datasets of 6,576 predictions based on 22 target proteins for CASP11, 4,197 predictions based on 12 target proteins for CASP12, and 5,532 predictions based on 14 target proteins for CASP13. For the x-ray crystal structure dataset, we compiled a dataset of 5,123 x-ray crystal structures culled from the PDB using PISCES^{59,60} with resolution ≤ 1.8 Å, a sequence identity cutoff of 20%, an R-factor cutoff of 0.25, and lengths greater than 75 residues. The packing features of the targets from all three CASP datasets fall within the range of the features from structures in the high resolution x-ray crystal structure database. (See Figs. S3 and S4 in Supporting Information)

While the CASP submissions represent decoys generated from state-of-the-art folding algorithms, and thus offer the most important application of decoy detection methods, there are also algorithms designed to create protein decoys from a reference structure. These methods are helpful in decoy detection as they offer an independent dataset that has the same diversity in target sequences and structures as the x-ray crystal structure dataset. As 3DRobot only takes a single chain as input, we only generated 3DRobot decoys from x-ray crystal structures with a single chain. We generated ~ 14 decoy structures for each x-ray crystal structure with a constraint on the C_α RMSD ≤ 10 Å.³⁶ The 3DRobot algorithm generates decoys from a reference structure by first threading the structure through a PDB library using TM-align⁷¹ and then running Monte Carlo Replica Exchange simulation using the I-Tasser Protocol,⁷² while optimizing for overall compactness and hydrogen bonding in the structure. We obtained a dataset of 30,569 3DRobot decoys based on 2,243 target proteins. The GDT for both the CASP and 3DRobot decoy datasets was measured using the Local-Global Alignment method.⁵⁶ The LDDT for both CASP and 3DRobot decoy datasets was measured using the OpenStructure framework.⁶⁵ See https://github.com/agrigas115/core_packing_score for a full list of all structures used in this study. Hydrogens were added to all structures using RE-DUCE.⁷³

rSASA

To identify core residues, we measured each residue’s solvent accessible surface area (SASA). To calculate SASA, we use the NACCESS software package,⁷⁴ which implements an algorithm originally proposed by Lee and Richards.⁷⁵ To normalize the SASA, we take the ratio of the SASA within the context of the protein (SASA_{context}) and the SASA of the same residue extracted from the protein structure as a dipeptide (Gly-X-Gly) with the same backbone and side-chain dihedral angles:

$$\text{rSASA} = \frac{\text{SASA}_{\text{context}}}{\text{SASA}_{\text{dipeptide}}}. \quad (2)$$

Core residues are classified as those that have rSASA $\leq 10^{-3}$. In Fig. 4, “near-core” residues are those with rSASA $\leq 10^{-1}$.

Packing Fraction

A characteristic measure of the packing efficiency of a system is the packing fraction. The packing fraction of residue μ is

$$\phi_{\mu} = \frac{\nu_{\mu}}{V_{\mu}}, \quad (3)$$

where ν_{μ} is the non-overlapping volume of residue μ and V_{μ} is the volume of the Voronoi cell surrounding residue μ . The Voronoi cell represents the local free space around the residue. We calculate the non-overlapping residue volume with a grid-based volume estimation. To calculate the Voronoi tessellation for a protein structure, we use the surface Voronoi tessellation, which defines a Voronoi cell as the region of space in a given system that is closer to the bounding surface of the residue than to the bounding surface of any other residue in the system. We calculate the surface Voronoi tessellations using the POMELO software package.⁷⁶ This software approximates the bounding surfaces of each residue by triangulating points on the residue surfaces. We find that using ~ 400 points per atom, or ~ 6400 surface points per residue, gives an accurate representation of the Voronoi cells and the results do not change if more surface points are included. Large overlaps between particles can make it difficult to calculate Voronoi tessellations for collections of non-spherical particles. Therefore, we compared the core packing fraction for x-ray crystal structures and computational decoys using the surface as well as the radical Voronoi tessellation.⁷⁷ We found that for x-ray crystal structures, the packing fractions obtained using the two methods deviate by $\lesssim 1\%$. While there are larger deviations in the packing fraction between the two methods for the computational models, the deviations are less than 6%, and the volume of the Voronoi cells is not double-counted. (See Fig. S11 in Supporting Information.)

Other Decoy Detection Methods

To compare the performance of our model that focuses on packing features to other models, we implemented several other single-ended decoy detection methods. VoromQA³⁴ is a statistical potential based on Voronoi contact areas (Accessed: <http://bioinformatics.ibt.lt/wtsam/voromqa/help/standalone>), SBROD²⁷ is a machine learning score (ridge regression) on backbone orientation (Accessed: <https://gitlab.inria.fr/grudin/sbrod>), 3DCNN⁵⁸ is a Convolutional Neural Network that uses extracted voxels based on atomic coordinates as features (Accessed: https://github.com/ishidalab-titech/3DCNN_MQA), and ProQ2²⁸ and ProQ3²⁹ are Support Vector Machine methods that use the secondary structure, solvent accessibility, Rosetta energy terms, and atomic coordinates as features (Accessed: <https://bitbucket.org/ElofssonLab/proq3/src/master/>).

Supporting Information

The Supporting Information contains additional results in support of the analyses presented in the main text. First, we describe the datasets of protein structures used in the main text, as well as compare the fluctuations in GDT for different computationally generated structures submitted to CASP from a single group for a given target sequence to those across different groups. We also provide additional details on the calculation of the Kullback-Leibler Divergence (D_{KL}) of the packing fraction probability distribution of hydrophobic residues. Second, we show that the packing properties of the target x-ray crystal structures from CASP are within the expected fluctuations of the properties from the larger dataset of high resolution x-ray crystal structures. In the main text, we characterized the packing features (Figs. 2 and 3) for the free modeling CASP submissions. Here, we compare the packing properties of predicted structures in the free and template-based modeling approaches. We include details about the simple decoy scoring method that used cutoffs based on x-ray crystal structure feature ranges. We also show preliminary results of the packing features before and after collective sidechain repacking. In addition, we show calculations of the packing fraction for core residues obtained using both the surface and radical Voronoi tessellation methods. We then elaborate on the machine learning approach we used to predict the GDT of the computationally generated structures, including the methodology for training and validation, and show the scatter plots of the decoy scores versus GDT for each decoy detection method tested on the CASP13 and 3DRobot decoy structures. Finally, we compare the results of the decoy detection method when using GDT versus LDDT as the ground truth score.

References

- [1] Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* 302:1364–1368.
- [2] Butterfoss GL, Kuhlman B (2006) Computer-Based Design of Novel Protein Structures. *Annu Rev Bioph Biom* 35:49–65.
- [3] Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, DeGrado WF (2007) Computational Design of Peptides that Target Transmembrane Helices. *Science* 315:1817–1822.
- [4] Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De Novo Computational Design of Retro-Aldol Enzymes. *Science* 319:1387–1391.
- [5] Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houlston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A, Arrowsmith CH, Baker D (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 357, 168–175
- [6] Regan L, DeGrado W (1988) Characterization of a helical protein designed from first principles. *Science* 241:976–978.
- [7] Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O’Neil KT, DeGrado WF (1995). Protein Design: A Hierarchic Approach. *Science* 270:935–941.
- [8] Lanci CJ, MacDermaid CM, Kang SG, Acharya R, North B, Yang X, Qiu XJ, DeGrado WF, Saven JG (2012) Computational design of a protein crystal. *PNAS* 109:7304–7309.

- [9] Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble alpha-helical barrels. *Science* 346:485–488.
- [10] Dawson WM, Rhys GG, Woolfson DN (2019) Towards functional de novo designed proteins. *Curr Opin Chem Biol* 52:102–111.
- [11] Main ER, Xiong Y, Cocco MJ, D’Andrea L, Regan L (2003) Design of Stable Alpha-Helical Arrays from an Idealized TPR Motif. *Structure* 11:497–508.
- [12] Baker D (2019) What has de novo protein design taught us about protein folding and biophysics? *Protein Sci* 28:678–683.
- [13] Zhang Y (2009) Protein structure prediction: When is it useful? *Curr Opin Struct Biol* 19:145–155.
- [14] Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J (2019) Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins Struct Funct Bioinf* 87:1011–1020.
- [15] Robustelli P, Piana S, and Shaw DE (2018) Developing a molecular dynamics force field for both folded and disordered protein states. *PNAS* 115:E4758–E4766.
- [16] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How Fast-Folding Proteins Fold. *Science* 334: 517–520.
- [17] Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A (2007) Assessment of predictions in the model quality assessment category. *Proteins Struct Funct Bioinf* 69:175–183.
- [18] Cozzetto D, Kryshtafovych A, Tramontano A (2009) Evaluation of CASP8 model quality predictions. *Proteins Struct Funct Bioinf* 77: 157–166.
- [19] Kryshtafovych A, Fidelis K, Tramontano, A (2011) Evaluation of model quality predictions in CASP9. *Proteins Struct Funct Bioinf* 79: 91–106.
- [20] Kryshtafovych A, Barbato A, Fidelis K, Monastyrskyy B, Schwede T, Tramontano A (2014) Assessment of the assessment: Evaluation of the model quality estimates in CASP10. *Proteins Struct Funct Bioinf* 82:112–126.
- [21] Kryshtafovych A, Barbato A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A (2016). Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. *Proteins Struct Funct Bioinf* 84:349–369.
- [22] Kryshtafovych A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A (2018) Assessment of model accuracy estimations in CASP12. *Proteins Struct Funct Bioinf* 86:345–360.
- [23] Cheng J, Choe MH, Elofsson A, Han KS, Hou J, Maghrabi AHA, McGuffin LJ, Menendez-Hurtado D, Olechnovic K, Schwede T, Studer G, Uziela K, Venclovas C, Wallner B (2019) Estimation of model accuracy in CASP13. *Proteins Struct Funct Bioinf* 87:1361–1377.
- [24] Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524.

- [25] Zhang J, Zhang Y (2010) A Novel Side-Chain Orientation Dependent Potential Derived from Random-Walk Reference State for Protein Fold Selection and Structure Prediction. *PLoS One* 5:1–13.
- [26] Lu M, Dousis A, Ma J (2008) OPUS-PSP: An Orientation-dependent Statistical All-atom Potential Derived from Side-chain Packing. *J Mol Biol* 376:288–301.
- [27] Karasikov M, Pages G, Grudinin S (2018) Smooth orientation-dependent scoring function for coarse-grained protein quality assessment. *Bioinformatics* 35:2801–2808.
- [28] Ray A, Lindahl E, Wallner B (2012) Improved model quality assessment using ProQ2. *BMC Bioinf* 13:224.
- [29] Uziela K, Shu N, Wallner B, Elofsson A (2016) ProQ3: Improved model quality assessments using Rosetta energy terms. *Sci Rep* 6:33509.
- [30] Benkert P, Biasini M, Schwede T (2010) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350.
- [31] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303.
- [32] Zhou H, Skolnick J. (2011) GOAP: A Generalized Orientation-Dependent, All-Atom Statistical Potential for Protein Structure Prediction. *Biophys J* 101:2043–2052.
- [33] Zhou H, Zhou, Y. (2009) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci* 11:2714–2726.
- [34] Olechnovic K, Venclovas C. (2017) VoroMQA: Assessment of protein structure quality using interatomic contact areas. *Proteins Struct Funct Bioinf* 85:1131–1145.
- [35] Cao R, Bhattacharya D, Hou J, Cheng J (2016) DeepQA: improving the estimation of single protein model quality with deep belief networks. *BMC Bioinf* 17:495.
- [36] Deng H, Jia Y, Zhang Y (2015) 3DRobot: automated generation of diverse and well-packed protein structure decoys. *Bioinformatics* 32:378–387.
- [37] Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- [38] Richards FM, Lim WA (1993) An analysis of packing in the protein folding problem. *Q Rev Biophys* 26:423–498.
- [39] Munson M, Regan L, O’Brien R, Sturtevant JM (1994) Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci* 3:2015–2022.
- [40] Munson M, Balasubramanian S, Fleming KG, Nagi AD, O’Brien R, Sturtevant JM, Regan L (1996) What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci* 5:1584–1593.
- [41] Willis MA, Bishop B, Regan L, Brunger AT (2000) Dramatic Structural and Thermodynamic Consequences of Repacking a Protein’s Hydrophobic Core. *Structure* 8:1319–1328.

- [42] Dalal S, Balasubramanian S, Regan L (1997) Transmuting alpha-helices and beta-sheets. *Fold Des* 2:R71–R79.
- [43] Dalal S, Regan L (2000) Understanding the sequence determinants of conformational switching using protein design. *Protein Sci* 9:1651–1659.
- [44] Regan L, Caballero D, Hinrichsen MR, Virrueta A, Williams DM, O’Hern CS (2015) Protein design: Past, present, and future. *Pept Sci* 104:334–350.
- [45] Richards FM (1977) Areas, volumes, packing, and protein structure. *Annu Rev Biophys Bioeng* 6:151–176.
- [46] Gaines JC, Smith WW, Regan L, O’Hern CS (2016) Random close packing in protein cores. *Phys Rev E* 93:032415.
- [47] Treado JD, Mei Z, Regan L, O’Hern CS (2019) Void distributions reveal structural link between jammed packings and protein cores. *Phys Rev E* 99:022416.
- [48] Gaines JC, Acebes S, Virrueta A, Butler M, Regan L, O’Hern CS (2018) Comparing side chain packing in soluble proteins, protein-protein interfaces and transmembrane proteins. *Proteins Struct Funct Bioinf* 86(5):581–591.
- [49] Caballero D, Virrueta A, O’Hern CS, Regan L (2016) Steric interactions determine side-chain conformations in protein cores. *PEDS* 29:367–376.
- [50] Gaines J, Virrueta A, Buch D, Fleishman S, O’Hern CS, Regan L (2017) Collective repacking reveals that the structures of protein cores are uniquely specified by steric repulsive interactions. *PEDS* 30:387–394.
- [51] Gaines JC, Clark AH, Regan L O’Hern, CS (2017) Packing in protein cores. *J Phys: Condens Matter* 29:293001.
- [52] Sheffler W, Baker, D (2009) RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* 18:229–239.
- [53] Colovos C and Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci*, 2:1511–1519.
- [54] Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, 26:283–291.
- [55] Cheng J, Li J, Wang Z, Eickholt J, Deng X (2012) The MULTICOM toolbox for protein structure prediction. *BMC Bioinf* 13:65.
- [56] Zemla A (2003) LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 31:3370–3374.
- [57] Won J, Baek M, Monastyrskyy B, Kryshchak A, Seok C (2019) Assessment of protein model structure accuracy estimation in CASP13: Challenges in the era of deep learning. *Proteins Struct Funct Bioinf* 87:1351–1360.
- [58] Sato R, Ishida T (2019) Protein model accuracy estimation based on local structure quality assessment using 3D convolutional neural network. *PLoS One* 14:1–15.

- [59] Wang G, Dunbrack RL (2003). PISCES: A protein sequence culling server. *Bioinformatics* 19:1589–1591.
- [60] Wang G, Dunbrack, RL (2005) PISCES: Recent improvements to a PDB sequence culling server. *Nucleic Acids Res* 33:W94–W98
- [61] Kullback S, Leibler RA (1951) On Information and Sufficiency. *Ann Math Stat* 22:79–86.
- [62] Smith WW, Ho PY, O’Hern CS (2014) Calibrated Langevin-dynamics simulations of intrinsically disordered proteins. *Phys Rev E* 90:042709.
- [63] Krivov GG, Shapovalov MV, Dunbrack RL (2009) Improved prediction of protein side-chain conformations with scwrl4. *Proteins Struct Funct Bioinf*, 77:778–795.
- [64] Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Advances in Neural Information Processing Systems 30 Curran Associates, Inc. 971–980.
- [65] Mariani V, Biasini M, Barbato A, Schwede T (2013) IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29:2722–2728.
- [66] Mei Z, Treado JD, Grigas AT, Levine ZA, Regan L, O’Hern CS (2020) Analyses of protein cores reveal fundamental differences between solution and crystal structures. *Proteins Struct Funct Bioinf*, 1–8.
- [67] Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Yang Y, Zhou Y (2015) Improving prediction of secondary structure, local backbone angles and solvent accessible surface area of proteins by iterative deep learning. *Sci Rep* 5:11476.
- [68] Kaleel M, Torrisi M, Mooney C, Pollastri G (2019) PaleAle 5.0: prediction of protein relative solvent accessibility by deep learning. *Amino Acids* 51:1289–1296.
- [69] Moulton J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in Round XI. *Proteins Struct Funct Bioinf* 84:4–14.
- [70] Moulton J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2018) Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins Struct Funct Bioinf* 86:7–15.
- [71] Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309.
- [72] Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738.
- [73] Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285:1735–1747
- [74] Hubbard SJ, Thornton JM, (1993) Naccess <http://wolf.bms.umist.ac.uk/naccess/>.
- [75] Lee B, Richards F (1971) The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400.

- [76] Weis S, Schonhoferer PWA, Schaller FM, Schroter M, Schroder-Turk GE (2017) Pomelo: a tool for computing Generic Set Voronoi Diagrams of Aspherical Particles of Arbitrary Shape. EPJ Web of Conferences 140:06007
- [77] Rycroft, C (2009) Voro++: a three-dimensional Voronoi cell library in C++.