

# ZEPPI: proteome-scale sequence-based evaluation of protein-protein interaction models

Barry Honig (✉ [bh6@cumc.columbia.edu](mailto:bh6@cumc.columbia.edu))

Columbia University

Haiqing Zhao

Columbia University

Diana Murray

Columbia University

Donald Petrey

Columbia University

---

## Article

**Keywords:** Protein-protein interactions, protein structure, coevolution, protein complex prediction

**Posted Date:** September 18th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3289791/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

# ZEPPI: proteome-scale sequence-based evaluation of protein-protein interaction models

Haiqing Zhao<sup>1</sup>, Diana Murray<sup>1</sup>, Donald Petrey<sup>1</sup>, and Barry Honig<sup>1,2,3,4,\*</sup>

1. Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA
2. Department of Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, New York, NY 10032, USA
3. Department of Medicine, Columbia University, New York, NY 10032, USA
4. Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10027, USA

\*Correspondence email: bh6@columbia.edu

## Abstract

We introduce ZEPPI (Z-score Evaluation of Protein-Protein Interfaces), a framework to evaluate structural models of a complex based on sequence co-evolution and conservation involving residues in protein-protein interfaces. The ZEPPI score is calculated by comparing metrics for an interface to those obtained from randomly chosen residues. Since contacting residues are defined by the structural model, this obviates the need to account for indirect interactions. Further, although ZEPPI relies on species-paired multiple sequence alignments, its focus on interfacial residues allows it to leverage quite shallow alignments. ZEPPI performance is evaluated through applications to experimentally determined complexes and to decoys from the CASP-CAPRI experiment. ZEPPI can be implemented on a proteome-wide scale as evidenced by calculations on millions of structural models of dimeric complexes in the *E. coli* and human interactomes found in the PrePPI database. A number of examples that illustrate how these tools can yield novel functional hypotheses are provided.

## Key words

Protein-protein interactions, protein structure, coevolution, protein complex prediction

## Introduction

The past decade has seen continuing developments in the prediction of protein-protein interactions (PPIs). One can trace these advances to the use of amino acid coevolution to predict inter-residue contacts<sup>1,2</sup>. These methods have been used to predict protein structures<sup>3-5</sup> and, more recently, to predict interaction partners and interfacial residues involved in PPIs<sup>6-9</sup>. The underlying premise is that functional interactions between two residues will result in their coevolution, which should be reflected in species-paired multiple sequence alignments (MSAs) of putative orthologues and detectable through mutual information (MI) based metrics between the two positions in the alignment. A complication is that the correlation between two residue positions  $i$  and  $j$ , i.e., two columns in the MSA, may result from an indirect coupling of  $i$  and  $j$  through their interaction with a third residue  $k$ . To solve this problem, methods such as Direct Coupling Analysis (DCA)<sup>3,10</sup>, sparse inverse covariance (PSICOV)<sup>11</sup>, EVcouplings<sup>6,8</sup>, and Gremlin<sup>4,7</sup>, have been developed. However, these methods rely on the availability of deep MSAs and thus have almost exclusively been applied to bacterial systems. In contrast, as we demonstrate below, ZEPPI can be applied on a genome-wide scale to eukaryotic proteomes with relatively shallow MSAs.

AlphaFold-Multimer<sup>12</sup> (AFM) has fundamentally changed the landscape of the prediction of structures of multi-protein complexes. There have been continuing improvements in AFM-based methods<sup>13,14</sup> as is evident from the substantial progress in the recent CASP-CAPRI<sup>15,16</sup> experiment<sup>17</sup>. An underlying problem for MSA-based methods is that, for a heterodimeric pair, it is generally necessary to carry out a species-based matching of the two query sequences which limits application to eukaryotic organisms due to the relatively limited number of sequences available for a paired MSA. Recently, RoseTTAFold/AlphaFold, was used to screen 4.3 million potential yeast PPIs with alignments containing >200 sequences and proteins with < 1500 amino acids per pair<sup>18</sup>. This was enabled in part by the large number of fungal genomes available but, as the authors point out, by the extensive amount of structural information embedded in RoseTTAFold<sup>18</sup>. However, applying deep learning to predict whether and how two proteins interact for entire eukaryotic proteomes remains computationally challenging.

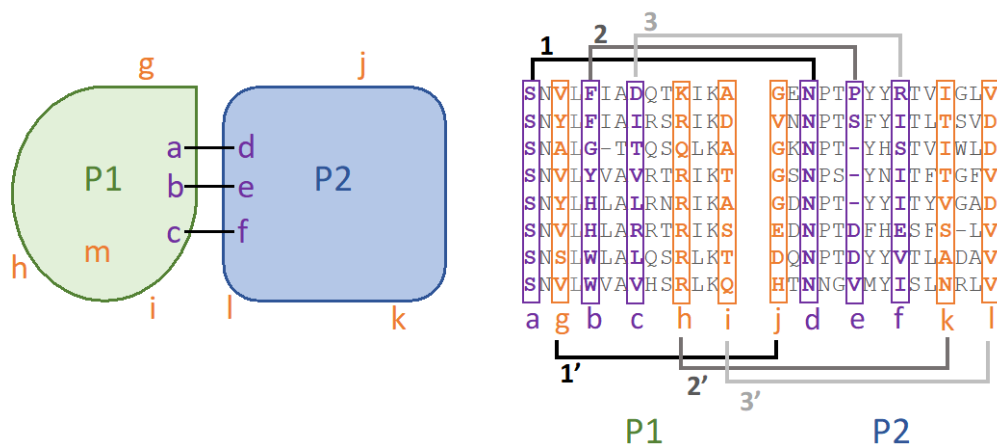
Docking-based methods predict models of protein dimers based on the structures of the constituent monomers<sup>19-22</sup> but have not been applied on a proteome-wide scale or to predict *whether* two proteins interact. Template-based modeling<sup>23</sup> is an alternate approach where the structures of individual proteins are superimposed on structurally similar proteins that appear in a complex present in the PDB<sup>24</sup>. In a series of papers, we reported the PrePPI (Predicting Protein-Protein Interactions) algorithm and database<sup>25-27</sup> that rely on template-based modeling and, through a highly efficient scoring function, leverage structural information on a truly proteome-wide scale. For example, PrePPI effectively screens the ~200 million possible pairwise combinations of human proteins which, in practice, amounts to billions of possible interactions among full-length proteins and protein domains. Based on a false positive rate of <0.005, 1.3 million high confidence predictions appear in the PrePPI online database with 370K corresponding to direct binary interactions<sup>27</sup>.

Here we use PrePPI predicted complexes in the *E. coli* and human proteomes to examine the extent to which simple evolution-based metrics are informative even in those cases for which the

multiple sequence alignment (MSA) depth is shallow. Once an interface is defined, we expect that MI calculations alone would be sufficient, even for eukaryotic proteins, as the deep MSAs required for DCA analysis would no longer be necessary. Our method, ZEPPI (for Z-score Evaluation of Protein-Protein Interfaces), uses paired MSAs to determine coevolutionary information across interfaces but also leverages sequence conservation which provides an additional signal as to the reliability of a predicted interface. An essential feature of ZEPPI is the comparison of evolutionary metrics derived from MSA positions corresponding to residues in predicted interfaces versus positions corresponding to randomly chosen residues.

Our focus on interfacial residues leads to a significant speedup in the evaluation of dimeric complexes that allows us to apply ZEPPI on a proteome-wide scale. Similarly, our finding that DCA is not needed for evaluating heterodimeric complexes effectively removes the need for deep species-paired MSAs. As shown below ZEPPI is extremely effective in distinguishing correct from incorrect protein-protein interfaces as indicated by tests on PDB structures and on a CASP-CAPRI benchmark set<sup>15,28</sup>. Most notably there is a strong inverse correlation between ZEPPI scores and false positive rates (FPRs) for PrePPI predictions thus providing strong support for the reliability of ZEPPI's efficacy and applicability to proteome-wide interactomes. We use a combined PrePPI/ZEPPI screen to identify a large number of novel interactions that do not appear in any database. A number of examples are discussed below.

## Results



**Figure 1. Schematic of the ZEPPI algorithm.** Two proteins, P1 and P2, form a complex with three interfacial contacts between residues **a** and **d**, **b** and **e**, and **c** and **f** (purple; left), respectively labeled as **1**, **2**, **3** in the right panel. Various evolutionary metrics (see text) are calculated from the corresponding columns in the paired alignment (purple; right). “Fake interfacial contacts” are generated between randomly chosen surface residues outside the interface, shown by **1'**, **2'** and **3'** on the right panel. When the number of surface residues outside the interface does not exceed the number of interfacial residues, buried residues (**m**) are considered as well. The same metrics are calculated from the corresponding columns in the MSA (orange; right) for each of 100 samples.

## ZEPPI overview

Figure 1 summarizes the ZEPPI algorithm. The procedure starts with a structural model of a complex between proteins P1 and P2 (left panel) and a paired MSA (right panel). Contacting interfacial residues are identified; in this case, P1-a contacts P2-d, P1-b contacts P2-e, and P1-c contacts P2-f. A paired MSA is created and used to calculate the following metrics for each of the four interfacial contacts as described in Methods: mutual information (MI), conservation (Con), direct coupling (DCA), and average product corrected (APC)<sup>29</sup> metric for each. The resulting values are then averaged over the interfacial contacts yielding 6 metrics. In addition, the highest single-contact score for each metric, denoted as “top”, is retained, resulting in a total of twelve metrics that characterize a predicted interface. Columns in the MSA corresponding to contacting interfacial residues are colored in purple. For example, the residues in columns P1-a and P2-d are almost completely conserved and would give a strong Con signal but a weak MI signal. Columns P1-b and P2-e show no obvious Con or MI signal but P1-c and P2-f show a clear MI signal.

The next step is to carry out the same procedure for a set of randomly chosen surface residues that are not in the interface. These are denoted in orange and are treated as if they were interfacial so that, for example, the metrics calculated between columns P1-a and P2-d are replaced by those between P1-g and P2-j. Each contact in the real interface is replaced in this way by fake contacts as indicated in the figure. Note that when the number of surface residues outside the interface is less than the number of residues in the interface, buried residues, e.g. P1-m, are included in sampling. This occurs for < 10% of PPIs evaluated. A hundred fake interfaces with corresponding values for the twelve metrics are generated in this way. A Z-score for the predicted interface is then calculated for each of the metrics based on the values for the real interface as compared to the values obtained for the 100 fake interfaces. In practice, based on results for PDB complexes (see next section) DCA is only used for homodimers where it makes an important contribution and where the computational cost is small given the relatively small number of homodimers in a proteome.

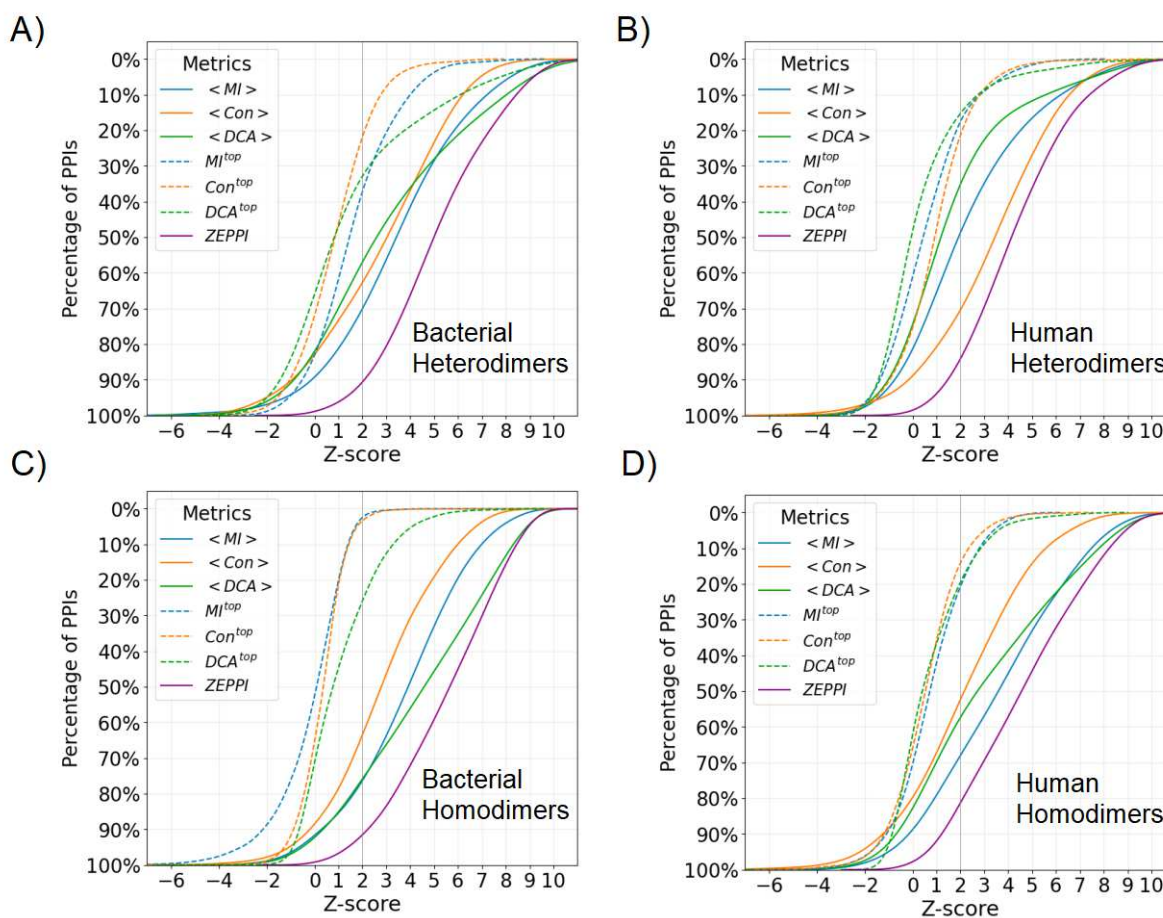
## Testing ZEPPI on PDB complexes

Dimeric PDB complexes were collected from the first bioassembly, as defined in the PDB structure file, for both bacterial and human complexes<sup>24</sup>. We tested performance with both prokaryotic and eukaryotic proteomes which, overall, have very different MSA depths. As described in Methods, complexes were selected based on resolution, chain length and the requirement that the proteins in the complex are from the same species. In total, 279 bacterial heterodimers, 247 human heterodimers, 3976 bacterial homodimers, and 977 human homodimers, for a total of 5,479 dimer structures, were obtained. For each complex we calculated the 12 metrics.

Figure S1 and S2 plot, for each of the twelve metrics, the fraction of PPIs with a Z-score above the threshold denoted along the x-axis. For example, in Figure S1A, at a Z-score of 2, the metric for average APC-corrected MI,  $\langle MI_{APC} \rangle$ , by itself recovers 60% of bacterial heterodimers whereas integrating all metrics (Figure 2) recovers 80%. Overall, it is evident that, for heterodimers, the

APC correction improves performance relative to raw (uncorrected) metrics for MI and DCA but not for Con (Figure S1, A and B). In contrast, for homodimers, the APC corrected Con metric is more effective than the corresponding raw metric (Figure S1, C and D). Further, choosing the top value for each metric is less effective than choosing the value averaged over the entire interface (dashed curves versus solid curves, Figures S1-S2). This is not unexpected since all contacts identified in PDB complexes are presumed to be correct and likely contribute to the total score. However, this is not necessarily the case with docked and predicted complexes as depicted below.

Figure 2 contains similar plots to those reported in Figures S1 and S2 but, for a given metric, the higher value of raw versus APC-corrected metric is chosen for each complex. The ZEPPI curve (purple) is generated by choosing the metric with the highest score for each complex. For the remainder of the paper, the ZEPPI score for a given complex corresponds to the maximum value from among the complex's 12 metrics.

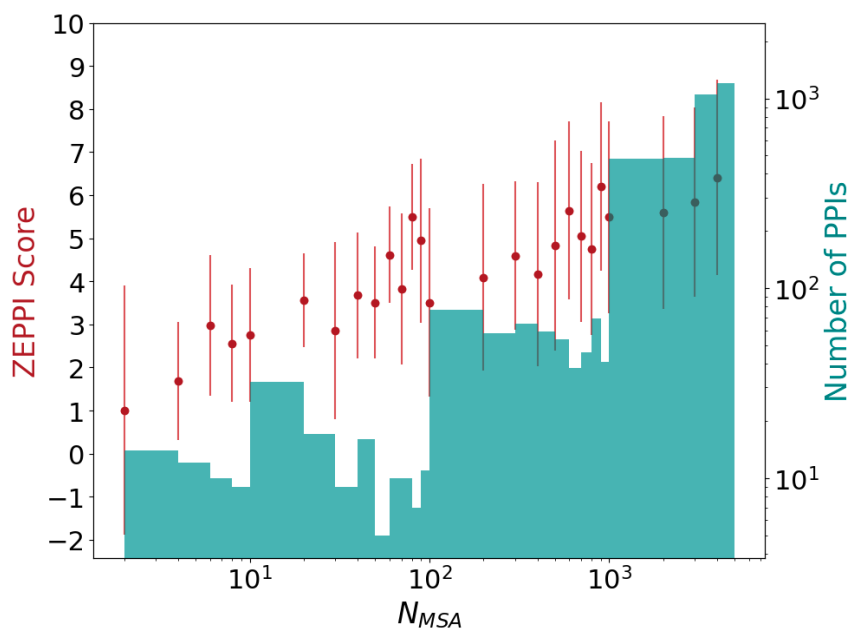


**Figure 2. Percentage of PDB PPIs as a function of Z-score.** Colors and line types in the legend indicate curves for different metrics each of which corresponds to the maximum of the raw and APC values for a given PPI. The mean and top metric of all interface contacts are denoted as <> ,

and <sup>top</sup>, respectively. The ZEPPI score for a given PPI is the largest Z-score among all metrics and the curve for all PPIs is shown in purple. See Supplementary File 1.

Figure 2 contains an evaluation of ZEPPI on PDB complexes. Of course, in all cases ZEPPI outperforms any individual metric since it “chooses” the best performing metric for each complex. For bacterial heterodimers (Figure 2A) MI is the best performing metric although DCA is slightly better at high Z-scores and Con performance is similar to both MI and DCA. In contrast, Con is clearly the most important metric for human heterodimers. We suggest that the difference between human and bacteria is the greater coevolutionary divergence underlying bacterial MSAs as opposed to eukaryotic MSAs. Of note, the overall ZEPPI performance is very similar for bacteria and human with, in both cases, about 65% of the complexes having a ZEPPI score > 4.

For homodimers, the two coevolutionary metrics perform the best with DCA performing better than MI at high Z-values for bacteria while MI is, overall, the best performer for human. The improved performance of coevolution for homodimers is likely due to the fact that MSA sequence depth for homodimeric complexes is much larger (reflecting two copies of a single protein) than for heterodimers. It is interesting that in all cases, MI performance is comparable to or better than that of DCA, except for bacterial homodimers which are associated with the deepest MSAs. But even in that case the differences manifest only at high Z-scores. Most importantly, since DCA contributes very little for human complexes and, given its need for deep MSAs and the extra computer time required in its use, below we only use DCA for homodimers. Of note, Bitbol has reported that, using an iterative pairing algorithm, MI alone performs at least as well as DCA in the sequence-based identification of protein-protein interaction partners<sup>9</sup>.



**Figure 3. Effect of MSA depth on ZEPPI score for PDB dimers.** The ZEPPI score is plotted against MSA depth,  $N_{MSA}$ , where each red dot and error bar correspond to the average and standard deviation of the ZEPPI score for the PPIs in a given bin of  $N_{MSA}$  values. A histogram of

the numbers of PPIs in each bin is shown in green. Data are plotted on a log scale for  $N_{\text{MSA}}$  and the number of PPIs.  $N_{\text{MSA}}$  is the depth of the paired MSAs after checking the coverage of surface residues (see Methods and Supplementary File 1.).

*Effect of MSA depth* – Figure 3 plots ZEPPI score (red dots) versus the sequence depth of the MSAs ( $N_{\text{MSA}}$ ). The histogram (green) displays the number of interfaces as a function of  $N_{\text{MSA}}$ . On average, ZEPPI score is seen to increase with increasing MSA depth although there are examples where ZEPPI scores  $> 2$  are obtained for very shallow MSA depths ( $N_{\text{MSA}} < 10$ ). Most of these result from significant sequence conservation of interfacial residues but there are cases where even MI yields a significant signal. Although these few cases may well be statistical anomalies, there are many high-scoring interfaces of relatively shallow depths with values of  $N_{\text{MSA}}$  in the range of 10 to 100. Note that  $N_{\text{MSA}}$  here is a raw number that does not include the low-weighting of redundant sequences. It is generally accepted that, for most applications,  $N_{\text{MSA}}$  should be at least the sum of the number of residues in each protein<sup>7,30–32</sup> and is typically taken to be greater than 500 or 1000 for predictions of protein-protein interactions<sup>7,31</sup>. Our results highlight the success of ZEPPI in leveraging even shallow MSAs, made possible by the evaluation of interfacial residues in experimentally determined structures.

#### Test on CAPRI benchmark decoys

We tested the performance of ZEPPI in differentiating good versus poor models in a widely used decoy set, `score_set`<sup>28</sup>, which was derived from targets from the CASP-CAPRI experiment. The `score_set` contains docking models predicted by 47 different groups for proteins from bacteria, yeast, vertebrates and artificial design. We considered 13 widely studied targets which, overall, have 18,538 corresponding decoys: 10% represent docking predictions of acceptable, medium or high quality based on CAPRI-defined criteria. We combine these three categories and annotate the group as “acceptable+”, whereas the remaining 90% are annotated as “incorrect.” Even though two of the targets, T53 and T54 contain designed proteins, they both have MSAs with  $N_{\text{MSA}}$  values of 2110 and 198, respectively. Table S1 reports MSA depth for all targets along with the number of acceptable+ and incorrect decoys, and the area under the ROC curve (AUROC) for each target. In contrast to the results with PDB complexes, the top metrics contribute to the ZEPPI score to a greater extent likely because the interfaces for acceptable+ models have inaccuracies. It is clear from the table that shallow MSA depths ( $<100$ ) can produce good AUROCs, particularly for T47 which has an AUROC of 0.93 and  $N_{\text{MSA}}$  of only 24.

Figure S3 plots the percentage of all models that have a given ZEPPI score in each of the four categories across targets. There is a clear distinction between acceptable+ and incorrect decoys with essentially 90% of the acceptable+ models having Z-scores  $> 2$ . Nevertheless, some incorrect decoys do have high Z-scores and some correct decoys have low Z-scores. For example, T40 (Table S1, AUROC = 0.66) is derived from a trimeric complex between a bovine protein and two copies of the same plant protein that bind in different locations. Only one interface is considered in the decoy set but the other forms a second interface complicating the sampling of non-interacting surface residues in creating fake interfaces. This issue that does not affect docking algorithm performance but compromises ZEPPI performance.



Table 1 compares ZEPPI performance across targets to that of other methods<sup>33–38</sup>, most of which are based on deep learning. The data for other methods was taken from Table S8 of Réau *et al.*<sup>34</sup> (see also Methods). ZEPPI, despite not involving training, is essentially tied as the top performer as measured by AUROC and is the best performer based on top 100 Success Rate. However, ZEPPI is outperformed by a number of other methods as measured by top 1 and top 5 Success Rates. Based on these criteria, iScore is the best performer. Of note, AUROC is affected by the distribution of false and true positives in a list of predictions while Success Rate depends on the number of good predictions at the top of the list. Success Rates are central to CASP-CAPRI rankings while ROC curve performance may be more important in asking whether a particular prediction is correct.

**Table 1. Performance of different scoring methods on CAPRI decoys.** AUROC in this table is averaged over values for each of 13 targets. Success Rates of Top N indicates the number of targets where there are acceptable or better predictions in the Top N predictions.

Method	AUROC	Success Rates		
		Top1	Top5	Top100
ZEPPI	0.72 ± 0.13	2/13	2/13	12/13
HADDOCK	0.57 ± 0.23	2/13	3/13	9/13
iScore	0.68 ± 0.21	5/13	6/13	9/13
DeepRank	0.64 ± 0.19	1/13	1/13	9/13
DOVE	0.56 ± 0.14	1/13	2/13	10/13
GNN-DOVE	0.63 ± 0.16	1/13	6/13	8/13
DeepRank-GNN	0.72 ± 0.19	1/13	5/13	10/13

### Evaluating PrePPI PPI models with ZEPPI

In recent work we reported PrePPI calculations for the human and *E. coli* interactomes represented by models for the full-length sequences and constituent domains<sup>27</sup>. A structural modeling score, SM, was trained on the HINT high-quality literature-curated (HINT-HQ-LC) dataset for human PPIs. HINT-HQ-LC datasets are designed to contain high confidence binary interactions<sup>39</sup>. ROC curves were reported for testing the PrePPI human and *E. coli* models on the human and *E. coli* HINT-HQ-LC datasets using 10-fold cross-validation. These yielded AUROC values of 0.83 and 0.88, respectively, thus, attesting to the overall high-quality of the predictions.

Figure S4 displays violin plots for the range of ZEPPI scores for PrePPI predictions in different bins of false positive rates (FPR). For PrePPI predictions of higher confidence (lower FPR), the median ZEPPI score is larger. These results provide a strong consistency check in that better

structural models as defined by PrePPI produce stronger evolutionary signals as measured by ZEPPI. For bacterial heterodimers (Figure S4A), at FPR <math>10^{-4}</math>, the percentage of predicted PPIs with a ZEPPI score > 2, 3, 4 is 94%, 81% and 67%, respectively. The comparable numbers for PDB structures (see discussion of Figure 2) are 95%, 85% and 71% suggesting that PrePPI's highest confidence predictions have ZEPPI scores close to those of PDB structures. Performance deteriorates as FPR increases but there are still many good ZEPPI scores for higher FPR values. The distributions in Figure S4 demonstrate that high and low ZEPPI scores are obtained in all FPR bins suggesting ZEPPI score can be used as an additional evidence source for prioritizing PrePPI models.

**Table 2. Number of proteins, PPIs and novel predictions for different combinations of PrePPI FPRs and ZEPPI-scores for *E. coli*. and Human. See Supplementary Files 2-4.**

**A). *E. coli***

PrePPI FPR ( $\leq$ )	0.05		0.01		0.001		0.0001	
ZEPPI Score ( $\geq$ )	-	4	-	4	-	4	-	4
# PPIs	303,212	38,062	71,151	17,098	10,336	6,002	3,151	2,355
# Proteins	3,941	3,671	3,557	3,030	2,464	1,993	1,605	1,350
# PPIs in DBs	38,916	9,069	14,580	6,090	5,289	3,528	2,386	1,837
# Novel PPIs	264,296	28,993	56,571	11,008	5,047	2,474	765	518

**B). Human**

PrePPI FPR ( $\leq$ )	0.05		0.01		0.001		0.0001	
ZEPPI Score ( $\geq$ )	-	4	-	4	-	4	-	4
# PPIs	6,209,528	1,002,052	1,271,323	228,321	130,447	30,572	11,896	7,392
# Proteins	16,780	15,987	13,903	11,781	6,358	5,250	2,882	2,441
# PPIs in DBs	463,971	130,426	148,461	54,131	20,605	10,762	6,293	4,409
# Novel PPIs	5,745,557	871,626	1,122,862	174,190	109,842	19,810	5,603	2,983

The *E. coli* structural interactome

Table 2A lists the number of *E. coli* PPIs (out of the 5.4 million for which a model can be built) and the number of proteins that comprise these interactions for different FPRs and ZEPPI scores. At FPR <math>0.01</math> PrePPI predicts 71K PPIs involving 3.5K proteins, and these numbers are significantly decreased when more stringent PrePPI FPRs and ZEPPI scores are applied. 2.3K PPIs satisfy the highly restrictive criteria of FPR <math>0.0001</math> and ZEPPI score > 4.

Table 2A also lists the overlap of ZEPPI-filtered PrePPI predictions with PPIs annotated in experimental databases (DBs). Any PPI that appears in the listed databases (see Methods) is considered whether or not the interaction is likely to be direct or indirect so as to determine the number of truly novel PPIs that our methods predict. At the most stringent end of the scale (FPR <math>0.0001</math>, ZEPPI score > 4) 518 novel predictions are made. On the other hand, as an example, there are 21,000 novel predictions made for FPR <math>0.05</math> and ZEPPI score > 4 suggesting that

using ZEPPI may facilitate the discovery of meaningful predictions that might be missed based on PrePPI alone.

#### The human structural interactome

Table 2B presents results for the PrePPI-predicted human interactome that parallel those for *E. coli* (Table 2A). In contrast to PrePPI results reported recently, which are based on both structural and non-structural evidence, Table 2B reports data for SM only, i.e. domain-domain structure-based predictions. A total of 1.3M PPIs are predicted with an FPR <0.01 which is an overly tolerant criterion. This number is reduced to only 130K for FPR <0.001 and only 12K for FPR <0.0001<sup>27</sup>. ZEPPI provides an alternate filter; for example ZEPPI = 4 reduces the number of predictions to 228K, 31K and 7K for FPR <0.01, 0.001 and 0.0001, respectively.

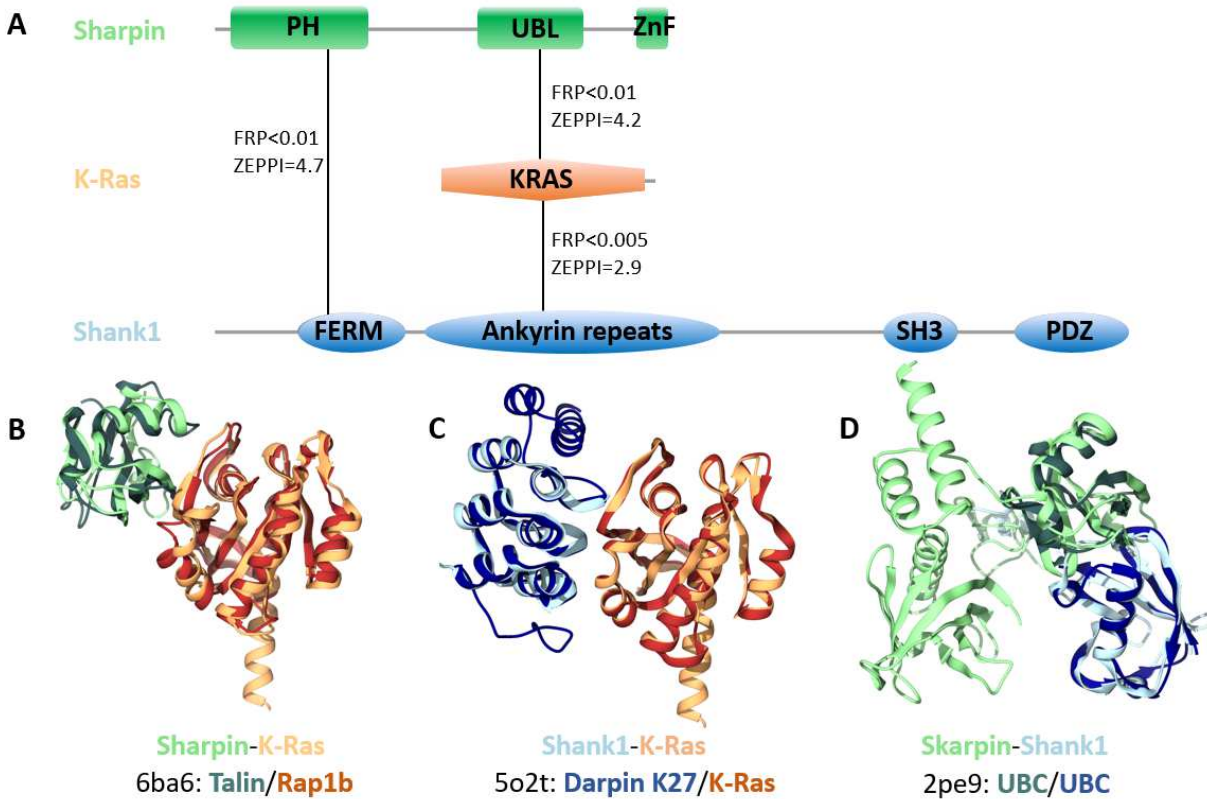
As is the case for *E. coli* (Table 2A), most PrePPI predictions do not appear in any experimental database nor in STRING<sup>40</sup> which includes many PPIs inferred from sequence relationships (collectively, “PPIs in DBs”). Although PrePPI provides structural models for many experimentally determined interactions, its value is also in hypothesis generation as many of its predictions are novel. At the highest confidence level (FPR <0.0001, ZEPPI>4), there are 2983 novel human PPI predictions. ZEPPI can be used to discriminate predictions at different PrePPI confidence levels, as indicated in the following examples.

#### Biological applications of ZEPPI/PrePPI

*Distinguishing among homologs:* An issue with PrePPI and other PPI prediction methods is that they encounter difficulties in predicting binding specificity when closely related homologs are involved. For example, based on a relatively small number of templates, there are many predicted interactions for the small GTPase K-Ras with GTPases, GAPs, GEFs, and other signaling proteins. Based on the X-ray complex for H-Ras/Grb14<sup>41</sup> (PDB ID: 4k81), PrePPI makes predictions (FPR ≤ 0.005) for K-Ras interactions with Grb7, Grb11, and Grb14. However, ZEPPI is significant (Z = 3.5) for only KRAS-Grb7<sup>42</sup>, the one for which there is evidence of an interaction (in DBs as defined above). This is a case where the structural models are too similar to be distinguished from one another but where there is a clear sequence signal that ZEPPI detects among interfacial residues.

*A possible role for K-Ras in synaptic signaling:* As shown in Figure 4A, PrePPI predicts interactions among K-Ras, Sharpin (the Shank-interacting protein-like 1) and Shank1 (the SH3 and multiple ankyrin repeat domains protein 1). Structural predictions are shown for 1) the Sharpin ubiquitin-like (UBL) domain and K-Ras (Figure 4B); 2) K-Ras and Sharpin ankyrin repeats (Figure 4C); and 3) Sharpin PH domain and Shank1 FERM domain (Figure 4D). Sharpin has previously been shown to interact with Shank1 and both co-localize at synaptic sites in mature neurons<sup>43</sup>. Altogether, our predictions (Figure 4) and the related experimental evidence suggest a novel role for K-Ras in synaptic signaling. Indeed, a recent study<sup>44</sup> found that mutant K-Ras increases synaptic transmission in inhibitory neurons, while it promotes the cell death of excitatory neurons.

*Secreted peptide fragments in the pancreas:* Chymotrypsin-like elastase family member 1 (CELA1) is a secreted elastase with high pancreatic expression. Recent studies have implicated peptides produced from the amyloid precursor protein (APP) in metabolic diseases<sup>45,46</sup>. In particular, human pancreatic islet cells process APP to release secreted fragments of APP (sAPP). The CELA1-APP model (Figure S5) suggests a pancreatic-specific mechanism for the production of sAPP.



**Figure 4. High-confidence PPIs in synaptic signaling.** A) PrePPI and ZEPPI predict interactions among Sharpin (the Shank-interacting protein-like 1, colored in green), the small GTPase K-Ras (orange), and Shank1 (the SH3 and multiple ankyrin repeat domains protein 1, colored in blue). Solid lines between protein domains denote the domains involved in the PPIs which are depicted as backbone ribbons. In B-D), the darker colored ribbons represent the chains from the PDB PPI template and are defined below the query protein names: B) Sharpin-K-Ras; C) K-Ras-Shank1; and D) Sharpin-Shank1. In all but two cases (KRAS-6ba6:B and KRAS-5o2t:A), the pairwise sequence identities between the queries and the respective template chains is less than 25%.

*Role of Cystatins in tumorigenesis:* Cystatins are inhibitors of cysteine peptidases. In tumor development and cancer progression the balance between cystatins and cysteine peptidases may be disrupted<sup>47</sup>. Cathepsin F (CTSF) was observed to have an anti-tumor effect in lung adenocarcinoma (LUAD)<sup>48</sup> whereas Cystatin-SN (CST1) promotes the epithelial-mesenchymal transition in LUAD cells<sup>49</sup>. The CST1-CTSF model (Figure S6) suggests that the mechanism of

action of CST1, which is highly expressed in LUAD<sup>49</sup>, may be to inhibit the anti-tumorigenic activity of CTSF.

## Discussion

Here we have introduced ZEPPI, a novel method that uses species-paired MSAs as a basis for scoring predicted models of protein-protein interfaces. ZEPPI's central feature involves the analysis of evolutionary information involving only contacting residues in a 3D structural model. The relatively limited number of residues to be analyzed results in a major reduction in computer time required to evaluate a model. Moreover, ZEPPI extracts signals from shallow MSAs enabled in part by its reliance on sequence conservation as well as mutual information. Deep learning methods implicitly leverage both sources of information but since most analyze entire sequences they are more computationally intensive.

In addition to validation on crystal structures, ZEPPI was tested on thirteen CASP-CAPRI targets and its performance was found to be comparable to or better than other interface evaluation approaches. We note that evolutionary information has been used for some time in the evaluation of docking models but generally in combination with other evidence sources, such as statistical propensities for surface residues to be in protein interfaces. ZEPPI differs from these approaches in its combined use of mutual information and conservation within interfaces and, especially, in its method of calculating Z-scores through the comparison of metrics for positions in the MSAs corresponding to interfacial residues versus positions in the MSAs corresponding to randomly chosen residues outside an interface. Our results on both PDB and CASP-CAPRI complexes demonstrate that ZEPPI provides a computationally efficient and highly effective measure of interface quality that can easily be combined with other sources of evidence.

To demonstrate its computational efficiency, we have applied ZEPPI to 5.4 million *E. coli* PPI interfaces and to 6.2 million (FPR <0.05) human PPI interfaces predicted by PrePPI. As suggested by the results in Table 2, filtering PrePPI predictions by ZEPPI scores has the potential to increase the reliability of high confidence predictions while identifying low confidence PrePPI predictions that are worthy of further consideration. An immediate application of ZEPPI is its integration into the PrePPI algorithm with the goal of combining evolutionary signals with a method based entirely on 3D structure. The integration should prove to be quite valuable, especially in applications to the human proteome and other eukaryotic organisms where available sequence information supports alignments of relatively shallow depth.

The vignettes provided above indicate the ability of ZEPPI to aid in the discovery of novel and potentially important functional hypotheses. In this regard, ZEPPI/PrePPI can be viewed as a hypothesis generating method that could be followed up with slower structure prediction methods ranging from docking to AF-multimer to methods based on their combination as evidenced from the most recent CASP-CAPRI experiment<sup>17</sup>. Of course, alternatively, ZEPPI can be used independent of PrePPI to evaluate any predicted model of a complex. For example, it could be used to distinguish direct physical from indirect interactions in experimentally determined multi-protein complexes.

The combination of PrePPI with ZEPPI suggests a general approach to the proteome-wide prediction of whether and how two proteins interact. ZEPPI can be used to score PrePPI models and thus provide an independent metric for their accuracy. In addition, if one is interested in evaluating specific PPIs, predicted for example by deep learning or docking-based methods, the post-prediction application of ZEPPI would appear to offer an efficient and accurate option. However, in genome-wide applications, we suggest using structure-based approaches to provide interactome-wide yes/no answers along with 3D models and then turning to increasingly accurate deep learning methods for a more limited set of interactions of particular interest.

## Methods

### Selecting bacterial and human PDB dimer structures

Taxonomy and UniProtKB summary files for all PDB chains were downloaded from the Structure Integration with Function, Taxonomy and Sequence (SIFTS) project<sup>50</sup>. From the SIFTS PDB chain taxonomy file, PDB chains that correspond to only one taxonomy ID were selected and then filtered to bacterial and human PDB chains, respectively. The taxonomy list of bacteria was collected from searching both the UniProt proteome<sup>51</sup> and the NCBI Taxonomy databases<sup>52</sup>. The union of the two searches provided 521,897 unique bacteria taxonomy IDs.

From the SIFTS PDB chain UniProt file, PDB files with only two UniProt IDs for heterodimers and one UniProt ID for homodimers with both chains longer than 30 amino acids are selected. PDBs that have any single chain mapped to  $\geq 2$  UniProt IDs are excluded to avoid fusion or chimera proteins. Structure resolution information is obtained through the PDB API service<sup>24</sup>. PDBs that are protein-only as the polymer entity type, and either from X-ray with resolution  $\leq 4$  Å or from EM with resolution  $\leq 4.5$  Å are selected. NMR structures are not used. Further, through reading the PDB file header, PDBs where the oligomer state of the first BioAssembly (BIOMOLECULE annotations) defined as “DIMERIC” by either the author or software with resolved sequence lengths longer than 30 amino acids are selected. Different PDB dimer structures for the same UniProt ID pairs are collapsed by keeping the structures with better structural resolution or longer chain-concatenated length (at least twice as long). Lastly, to remove closely related homologous protein pairs, we compared the pairwise sequence identities and removed sequence redundant structures where both protein sequences have 90% sequence identity with another structure. The detailed pipeline is provided in the supplemental information.

### Defining protein surface and protein-protein interface

The accessible surface area (ASA) of residues for individual chains A and B, and their complex AB are obtained using our in-house program of surfv<sup>53</sup>. An interface is defined as long as the buried ASA is larger than zero. The interface between proteins A and B consists of contacting residues where the distance between any heavy atoms is less than 6.0 Å. All the residue indices from the PDB are updated after mapping the PDB sequences to their full UniProt sequences using hhalgn of the hh-suite package<sup>54</sup>.

### Generating random protein-protein interfaces

The positions of interface residues on proteins A and B in the concatenated MSA are replaced, one by one, with positions for randomly chosen surface residues of the same protein as indicated in Figure 1. If one protein has more interface residues than surface residues that are not on the interface, the sampling pool goes to the entire protein sequence. To ensure statistical significance of the Z-score calculations, 100 random interfaces are generated for each protein-protein interface.

### Generating and pairing MSAs

To avoid biased sequence sampling due to over-studied model species, we carried out homolog sequence search on 5,090 representative proteomes that were carefully curated and selected in EggNog 5.0<sup>55</sup>. This database includes 4,445 prokaryotic reference genomes selected from original 25,038 bacteria genomes, and 477 eukaryotic genomes. Homologous sequences are searched using Jackhmmer (hmmer-3.2.1)<sup>56</sup> with 5 iterations and the default E-value of 0.001. In the final outputted multiple sequence alignment, only the sequence with highest identity to the query is kept as the representative sequence for each species.

The MSAs of two proteins, p1 and p2, are paired based on the shared common species. Sequence rows that cover less than 50% of surface residue positions of p1 or p2 are excluded from the paired MSA. MSA columns, either for interface residue or surface residue positions, that have more than 50% gaps are excluded.

### Calculating mutual information, conservation, DCA, and their APC-corrected terms

For two positions (**a**, **b**) in the paired MSA, their mutual information (MI) is calculated through Eq. 1, where *x* and *y* denote their amino acid type and the gap is treated as the 21st amino acid type or state. The  $p(x)$  and  $p(y)$  are the frequencies of a certain amino acid type and  $p(x, y)$  is the frequency of a pair of different amino acid types. The conservation score between two positions (**a**, **b**) is defined through the complement of their normalized joint entropy  $S(a, b)$  (Eq. 2). The direct coupling information is calculated through the mean field DCA method which is based on the maximum-entropy model<sup>3</sup>. The final direct coupling information is quantified using a similar definition as in the mutual information except  $p^{(dir)}(x, y)$  involves only the isolated direct coupling strength of (**a**, **b**) from the DCA calculations (Eq. 3).

$$MI(a, b) = - \sum_{x \in \{1..21\}} \sum_{y \in \{1..21\}} p(x, y) \ln \frac{p(x, y)}{p(x) p(y)} \quad (\text{Eq. 1})$$

$$Con(a, b) = 1 - \frac{S(a, b)}{2 * \log_2 21} \quad (\text{Eq. 2})$$

$$DI(a, b) = - \sum_{x \in \{1..21\}} \sum_{y \in \{1..21\}} P^{(dir)}(x, y) \ln \frac{P^{(dir)}(x, y)}{p(x) p(y)} \quad (\text{Eq. 3})$$

The average product correction (APC) is applied to measurements as denoted throughout the text. Taking MI as an example, the APC term between position (**a**, **b**) (from p1 and p2, respectively) is calculated as the product of the average MI signal of position **a** with positions in p2, and position **b** with positions in p1, between interfacial residues on both proteins, then

normalized by the average measurement of all protein to the other (Eq. 4). The APC-corrected MI is given by Eq. 5, and the same correction correspondingly applies to Con, and DCA scores.

$$APC_{MI}(a, b) = \frac{\langle MI(a, y) \rangle * \langle MI(b, x) \rangle}{\langle MI(x, y) \rangle} \quad (\text{Eq. 4})$$

$$MI_{APC}(a, b) = MI(a, b) - APC_{MI}(a, b) \quad (\text{Eq. 5})$$

#### Calculating Z-scores of the interface

For each interface contact of a given interface between protein p1 and p2 and the generated 100 random interfaces, the following six measurements are calculated: mutual information, conservation, direct coupling information and their corresponding APC-corrected terms. Of all the interface contacts, we choose the top and the mean as the representative metric for each measurement, denoted as  $MI^{\text{top}}$  and  $\langle MI \rangle$ , for example. The Z-score of the 12 metrics are then calculated for the given interface versus the generated random interfaces. The larger Z-score of the raw metric versus its APC-corrected metric is taken as the Z-score for this metric. The maximum of all metrics is taken as the final ZEPPI score.

#### Building the *E. coli* experimental PPI database

The experimental database of *E. coli* PPIs is integrated from several major resources including Interactome3D<sup>57</sup>, HINT<sup>39</sup>, APID<sup>58</sup>, STRING<sup>40</sup> and Ecocyc<sup>59</sup>, as well as previously known large-scale *E. coli* PPI high-throughput screening with experimental methods such as APMS<sup>60</sup> and Y2H<sup>61</sup>. Another well-known experimental database BioGrid<sup>62</sup> is not included due to the lack of *E. coli* (NCBI Taxonomy ID: 83333) PPIs included. Before their integration, each database was pre-processed by selecting only *E. coli* K12 proteins (proteome size: 4391) and sorting the uniprot IDs for each pair of PPIs. During the integration, redundant PPIs were removed. Note that Interactome3D also includes homology-modeled PPIs and the STRING database has many inferred PPIs, which are not determined by direct physical interaction experiments but inferred by other methods such as gene-related methods or species PPI transfer. By excluding these two contributions, we also built a purely experimental PPI database of *E. coli* based on direct physical experiments. In all, there are 565,007 PPIs in the integrated experimental database set and 45,634 PPIs in the physical experimental PPI dataset.

In summary, the integrated experimental database set includes: all HINT binary and complex PPIs (updates of 2021/11), all APID PPIs (updates of 2021/11), all Interactome3D PPIs (updates of 2021/11), all STRING PPIs (v11.5), the gold standard dataset used in Zhang and coworkers' Threpp work<sup>63</sup>, the high throughput experimental PPI set from Threpp (Table S1), the EcoCyc PPIs downloaded Cong *et al.*<sup>64</sup>(Table S5), the Y2H PPI set from Rajagopala *et al.*<sup>61</sup> (Supplementary Table 2), the high-confidence and median-confidence APMS PPI set from Babu *et al.*<sup>60</sup> (Supplementary Table 2). For the physical experimental PPI dataset, only physical links in the STRING database with experimental score >0 are included; only the PDB subset of Interactome3D is included; the other datasets remain the same as in the integrated experimental database.



### Building the human experimental PPI database

The integrated human experimental PPI database consists of the following resources: all HINT binary and complex PPIs, all Interactome3D PPIs, all APID PPIs, all STRING PPIs (v11.5), all BioGrid<sup>62</sup> PPIs, all HURI<sup>43</sup> PPIs, and the HC-2016 set from PrePPI<sup>26</sup>. In total, there are 6,068,248 PPIs collected from the above-mentioned experimental databases with the large majority derived from STRING.

### **Code and data availability**

Code for ZEPPI method with tutorial examples is available at <https://github.com/honig-lab/ZEPPI>.

The following Supplementary Files are available at FigShare:

<https://doi.org/10.6084/m9.figshare.c.6800502.v1>:

Supplementary File 1. ZEPPI results for

- 1A: Bacterial PDB heterodimer complexes
- 1B: Human PDB heterodimer complexes
- 1C: Bacterial PDB homodimer complexes
- 1D: Human PDB homodimer complexes

Supplementary File 2. ZEPPI results for

- 2A: PrePPI-AF predictions for E. coli PPIs, at FPR  $\leq 0.001$
- 2B: PrePPI-AF predictions for human PPIs, at FPR  $\leq 0.001$

Supplementary File 3. ZEPPI results for PrePPI-AF predictions for E. coli PPIs, at FPR  $\leq 0.05$ .

Supplementary File 4. ZEPPI results for PrePPI-AF predictions for human PPIs, at FPR  $\leq 0.05$ .

### **Acknowledgments**

This work was supported by the National Institute of Health (grant R35-GM139585).

## References:

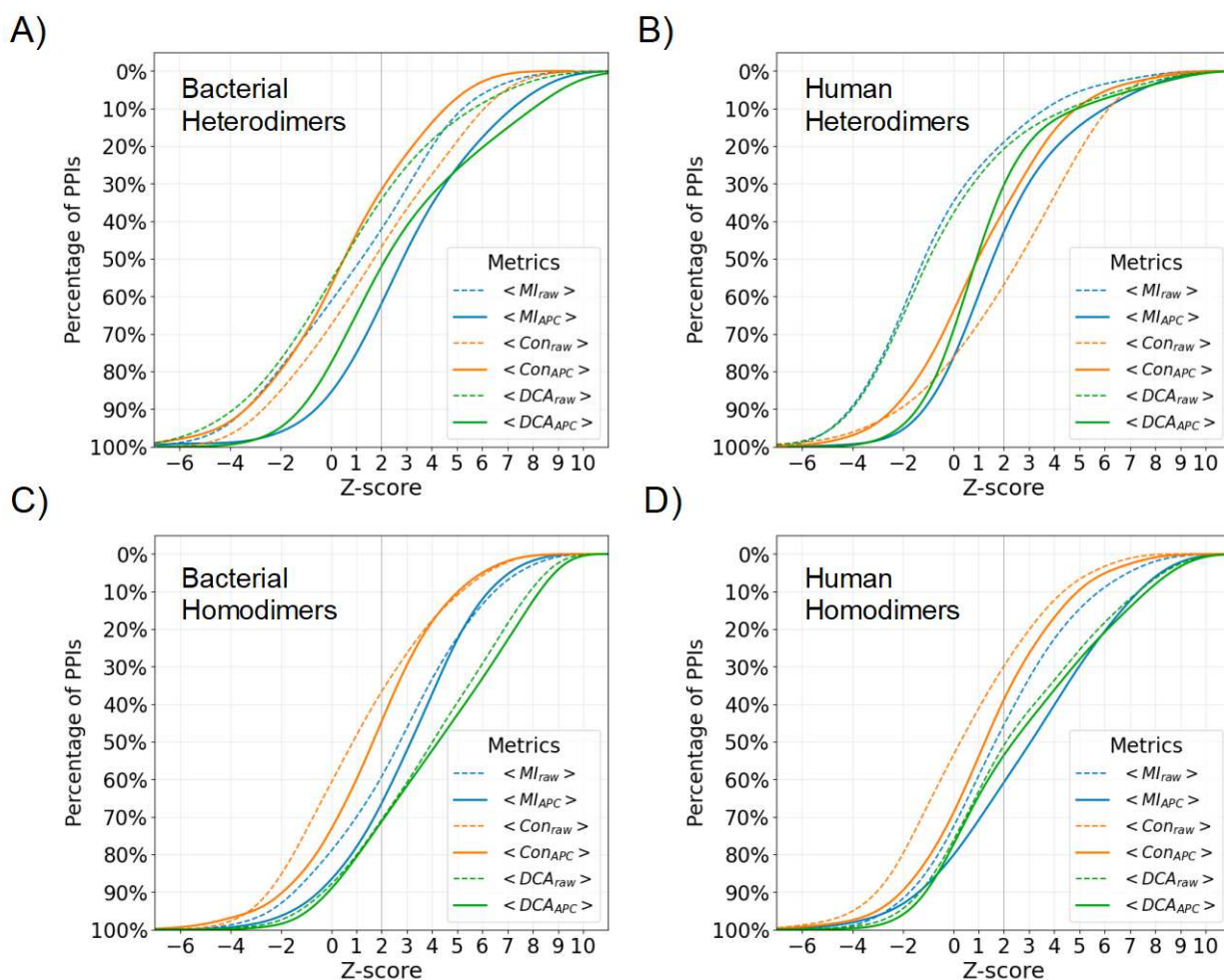
1. Göbel, U., Sander, C., Schneider, R. & Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Bioinformatics* **18**, 309–317 (1994).
2. Shoemaker, B. A. & Panchenko, A. R. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Computational Biology* vol. 3 595–601 Preprint at <https://doi.org/10.1371/journal.pcbi.0030043> (2007).
3. Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* (2011) doi:10.1073/pnas.1111471108.
4. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era (Proceedings of the National Academy of Sciences of the United States of America (2013) 110, 39 (15674-15679) DOI:10.1073/pnas.1314045110). *Proc Natl Acad Sci U S A* **110**, 18734–18735 (2013).
5. Marks, D. S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, (2011).
6. Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, 1–45 (2014).
7. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, 1–21 (2014).
8. Green, A. G. *et al.* Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat Commun* 1–12 (2021) doi:10.1038/s41467-021-21636-z.
9. Bitbol, A. F. Inferring interaction partners from protein sequences using mutual information. *PLoS Comput Biol* **14**, (2018).
10. Weigt, M., White, R. A., Szurmant, H., Hoch, J. A. & Hwa, T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* **106**, 67–72 (2009).
11. Jones, D. T., Buchan, D. W. A., Cozzetto, D. & Pontil, M. PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28**, 184–190 (2012).
12. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. (2022) doi:10.1101/2021.10.04.463034.
13. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **13**, 1–11 (2022).
14. Gao, M., Nakajima An, D., Parks, J. M. & Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun* **13**, (2022).
15. Janin, J. *et al.* CAPRI: A critical assessment of PRedicted interactions. *Proteins: Structure, Function and Genetics* **52**, 2–9 (2003).

16. Kryshchuk, A. *et al.* New prediction categories in CASP15. *Proteins: Structure, Function and Bioinformatics* 1–8 (2023) doi:10.1002/prot.26515.
17. Elofsson, A. Progress at protein structure prediction, as seen in CASP15. *Curr Opin Struct Biol* **80**, 102594 (2023).
18. Humphreys, I. *et al.* Computed structures of core eukaryotic protein complexes. *Science* (1979) **374**, (2021).
19. Vakser, I. A. Protein-protein docking: From interaction to interactome. *Biophysical Journal* vol. 107 1785–1793 Preprint at <https://doi.org/10.1016/j.bpj.2014.08.033> (2014).
20. Barradas-Bautista, D., Rosell, M., Pallara, C. & Fernández-Recio, J. Structural Prediction of Protein–Protein Interactions by Docking: Application to Biomedical Problems. in *Advances in Protein Chemistry and Structural Biology* vol. 110 203–249 (Academic Press Inc., 2018).
21. Pierce, B. G. *et al.* ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).
22. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat Protoc* **12**, 255–278 (2017).
23. Petrey, D. *et al.* Template-based prediction of protein function. *Current Opinion in Structural Biology* vol. 32 33–38 Preprint at <https://doi.org/10.1016/j.sbi.2015.01.007> (2015).
24. Berman, H. M. *et al.* *The Protein Data Bank*. *Nucleic Acids Research* vol. 28 <http://www.rcsb.org/pdb/status.html> (2000).
25. Zhang, Q. C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **490**, 556–560 (2012).
26. Garzón, J. I. *et al.* A computational interactome and functional annotation for the human proteome. *Elife* **5**, 1–27 (2016).
27. Petrey, D., Zhao, H., Trudeau, S., Murray, D. & Honig, B. PrePPI: A structure informed proteome-wide database of protein-protein interactions. *J Mol Biol* 168052 (2023) doi:10.1016/j.jmb.2023.168052.
28. Lensink, M. F. & Wodak, S. J. Score\_set: A CAPRI benchmark for scoring protein complexes. *Proteins: Structure, Function and Bioinformatics* **82**, 3163–3169 (2014).
29. Dunn, S. D., Wahl, L. M. & Gloor, G. B. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **24**, 333–340 (2008).
30. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
31. Mehrabiani, K. M., Cheng, R. R. & Onuchic, J. N. Expanding Direct Coupling Analysis to Identify Heterodimeric Interfaces from Limited Protein Sequence Data. *Journal of Physical Chemistry B* **125**, 11408–11417 (2021).
32. Haldane, A. & Levy, R. M. Influence of multiple-sequence-alignment depth on Potts statistical models of protein covariation. *Phys Rev E* **99**, 1–15 (2019).
33. Renaud, N. *et al.* DeepRank: a deep learning framework for data mining 3D protein-protein interfaces. *Nat Commun* **12**, 1–8 (2021).

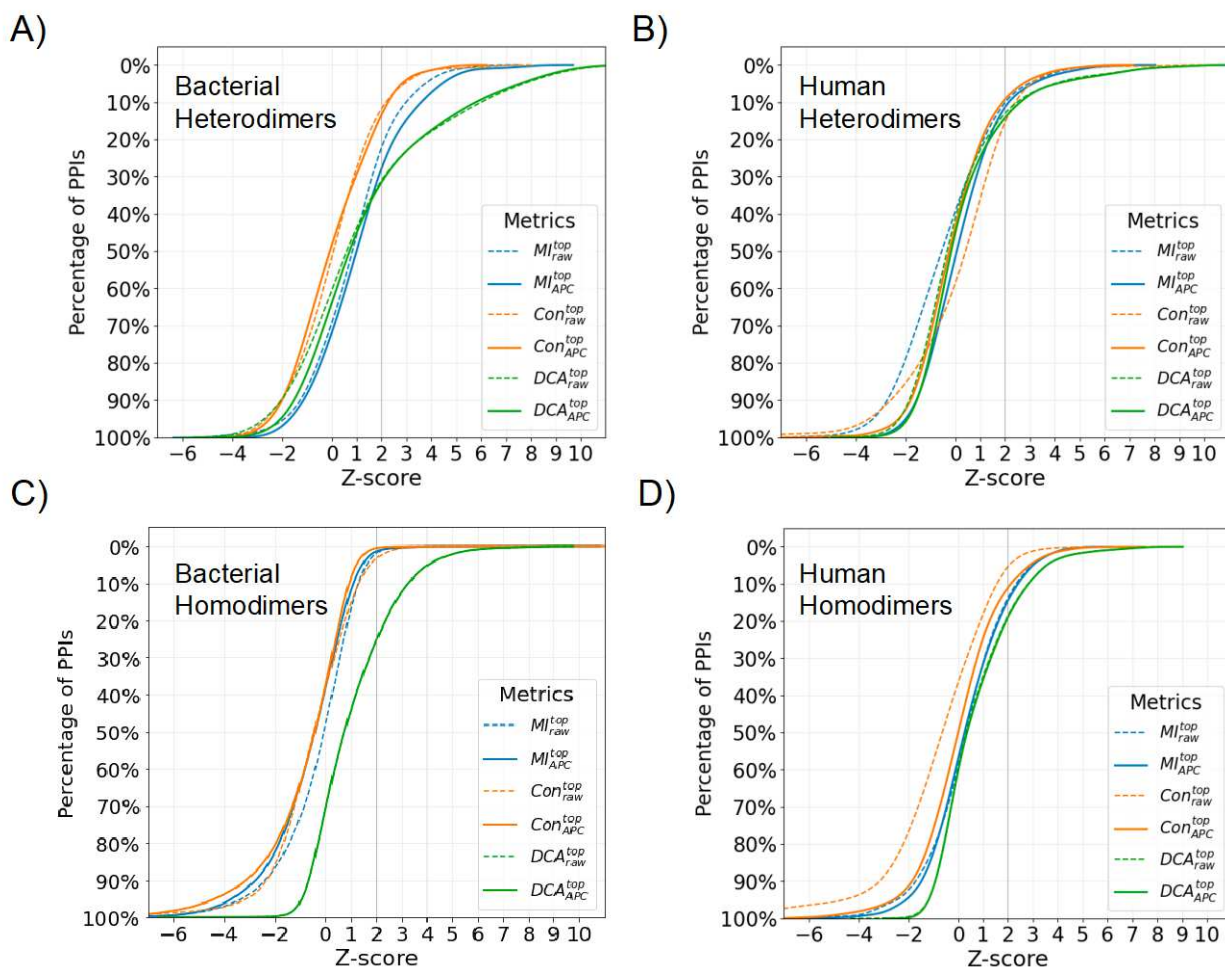
34. Réau, M., Renaud, N., Xue, L. C. & Bonvin, A. M. J. J. DeepRank-GNN: a graph neural network framework to learn patterns in protein-protein interfaces. *Bioinformatics* **39**, 1–8 (2023).
35. Geng, C. *et al.* Structural bioinformatics iScore: a novel graph kernel-based function for scoring protein-protein docking models. *Bioinformatics* **36**, 112–121 (2019).
36. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731–1737 (2003).
37. Wang, X., Flannery, S. T. & Kihara, D. Protein Docking Model Evaluation by Graph Neural Networks. *Front Mol Biosci* **8**, (2021).
38. Wang, X., Terashi, G., Christoffer, C. W., Zhu, M. & Kihara, D. Protein docking model evaluation by 3D deep convolutional neural networks. *Bioinformatics* **36**, 2113–2118 (2020).
39. Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* **6**, (2012).
40. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612 (2021).
41. Qamra, R. & Hubbard, S. R. Structural Basis for the Interaction of the Adaptor Protein Grb14 with Activated Ras. *PLoS One* **8**, 1–9 (2013).
42. Chen, K. *et al.* Methylation-associated silencing of miR-193a-3p promotes ovarian cancer aggressiveness by targeting GRB7 and MAPK/ERK pathways. *Theranostics* **8**, 423–436 (2018).
43. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
44. Ryu, H. H. *et al.* Neuron type-specific expression of a mutant KRAS impairs hippocampal-dependent learning and memory. *Sci Rep* **10**, 1–13 (2020).
45. Guo, Y., Wang, Q., Chen, S. & Xu, C. Functions of amyloid precursor protein in metabolic diseases. *Metabolism* **115**, 154454 (2021).
46. Kulas, J. A., Puig, K. L. & Combs, C. K. Amyloid precursor protein in pancreatic islets. *Journal of Endocrinology* **235**, 49–67 (2017).
47. Breznik, B., Mitrović, A., T. Lah, T. & Kos, J. Cystatins in cancer progression: More than just cathepsin inhibitors. *Biochimie* **166**, 233–250 (2019).
48. Song, L. *et al.* Expression signature, prognosis value and immune characteristics of cathepsin F in non-small cell lung cancer identified by bioinformatics assessment. *BMC Pulm Med* **21**, 1–17 (2021).
49. Yang, J. *et al.* Cystatin SN promotes epithelial-mesenchymal transition and serves as a prognostic biomarker in lung adenocarcinoma. *BMC Cancer* **22**, 1–18 (2022).
50. Velankar, S. *et al.* SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res* **41**, (2013).
51. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).
52. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res* **40**, (2012).

53. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Structure, Function, and Bioinformatics* **11**, 281–296 (1991).
54. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**, 173–175 (2012).
55. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309–D314 (2019).
56. Eddy, S. R. *HMMER User's Guide Biological sequence analysis using profile hidden Markov models*. <http://hmmer.org> (2020).
57. Mosca, R., Céol, A. & Aloy, P. Interactome3D: Adding structural details to protein networks. *Nat Methods* **10**, 47–53 (2013).
58. Alonso-López, Di. *et al.* APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database* **2019**, (2019).
59. Keseler, I. M. *et al.* The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
60. Babu, M. *et al.* Global landscape of cell envelope protein complexes in Escherichia coli. *Nat Biotechnol* **36**, 103–112 (2018).
61. Rajagopala, S. V. *et al.* The binary protein-protein interaction landscape of escherichia coli. *Nat Biotechnol* **32**, 285–290 (2014).
62. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res* **47**, D529–D541 (2019).
63. Gong, W. *et al.* Integrating Multimeric Threading With High-throughput Experiments for Structural Interactome of Escherichia coli. *J Mol Biol* **433**, 166944 (2021).
64. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science (1979)* **365**, 185–189 (2019).

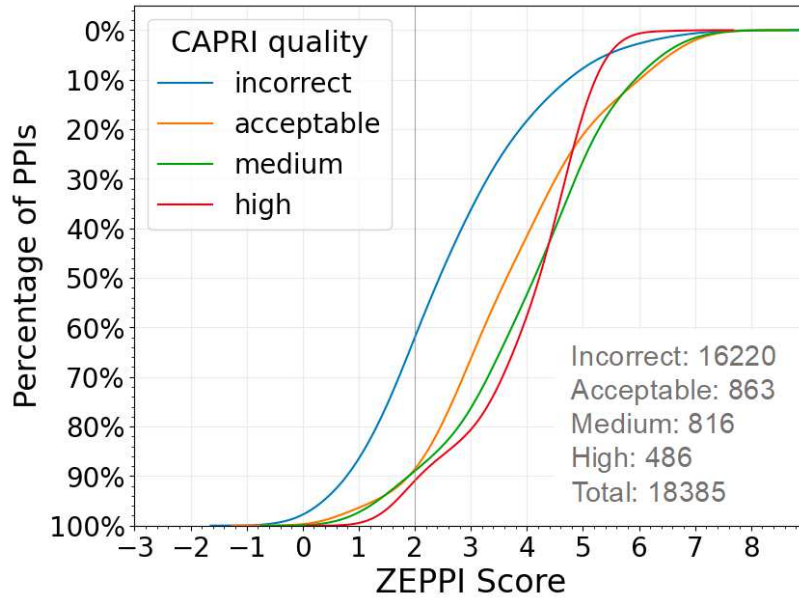
## Supplemental Information:



**SI Figure S1. Percentage of PDB PPIs as a function of Z-score for raw and APC-corrected metrics averaged over interface contacts.** Colors and line type defined in the legend indicate curves for different metrics

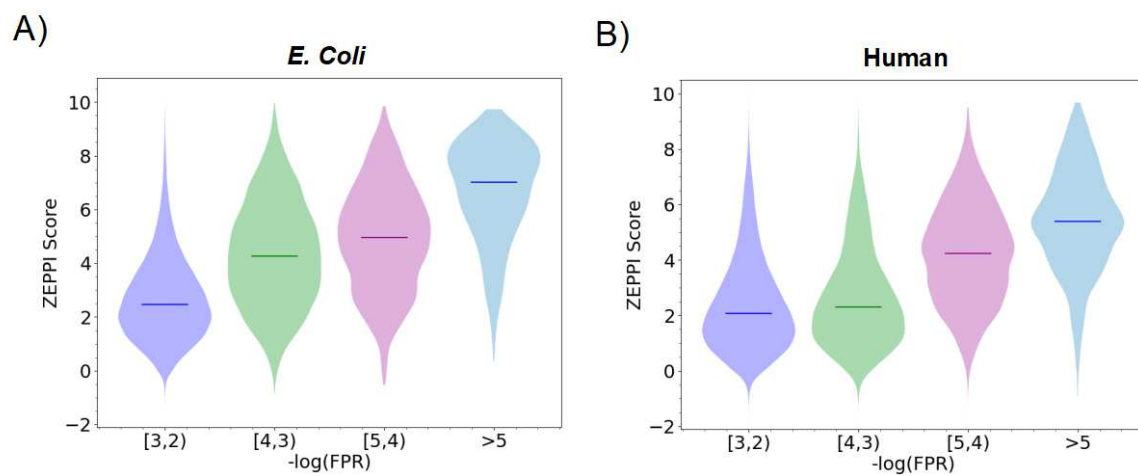


**SI Figure S2. Percentage of PDB PPIs as a function of Z-score for the interface contact with the top value for a given metric. Colors and line type defined in the legend indicate curves for different metrics.**

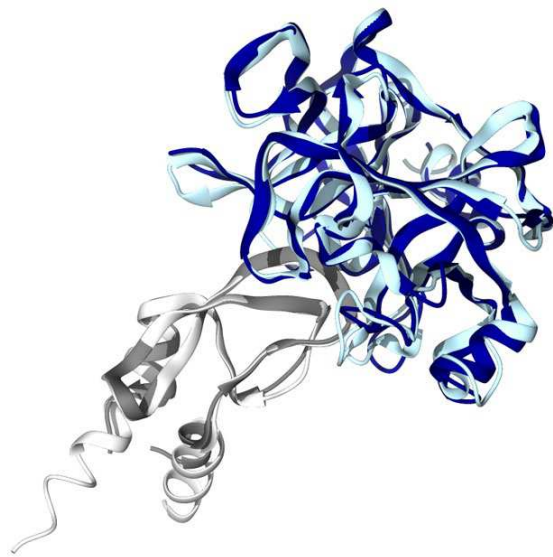


**SI Figure S3. Percentage of CAPRI structures and decoys having a given ZEPPI score.** Percentages are plotted along the y-axis for four classes of CAPRI models defined in the legend. The total number of models in each class is indicated in the text at the lower left.





**SI Figure S4. Relationship between the ZEPPI score and FPR of PrePPI-predicted PPIs .** FPR ranges are indicated below each set of color-coded violin charts, where the median ZEPPI scores are shown as a bar.



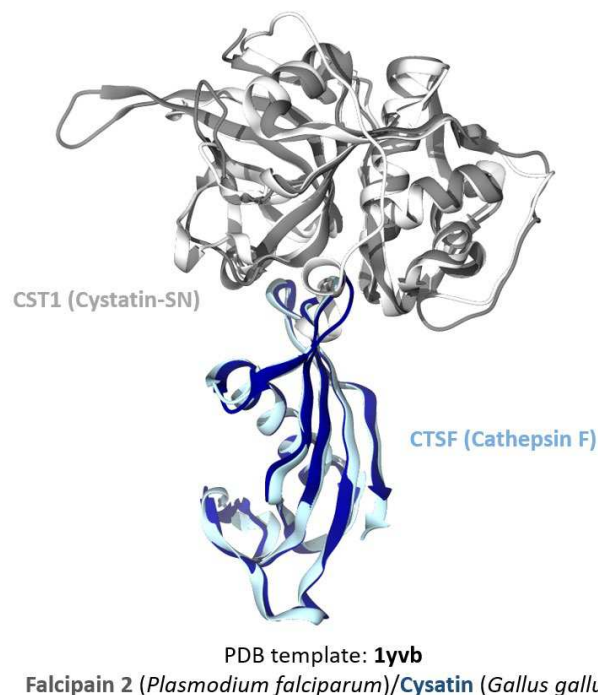
APP (Amyloid-beta precursor protein)

CELA1 (Chymotrypsin-like elastase family member 1)

PDB template: **4bnr**

Trypsin inhibitor (*Bos Taurus*)/Trypsin (*Astacus leptodactylus*)

**SI Figure S5. Interaction model for Amyloid-beta precursor protein, APP(294-341) (light gray), and Chymotrypsin-like elastase family member 1, CELA1(19-258) (light blue).** The template for modeling (PDB ID: 4bnr) is a complex of Trypsin inhibitor from *Bos Taurus* (dark gray) and Trypsin from *Astacus leptodactylus* (dark blue). PrePPI FPR < 0.0001, and ZEPPI score = 5.1. The pairwise sequence identities between queries and template chains are 26% and 33%, respectively.



**SI Figure S6. Interaction model for Cystatin-SN, CST1(35-141) (light gray), and Cathepsin F, CTSF(244-484) (light blue).** The template for modeling (PDB ID: 1yvb) is a complex of Falcipain 2 from *Plasmodium falciparum* (dark gray) and Cystatin from *Gallus gallus* (dark blue). PrePPI FPR < 0.005, and ZEPPI score = 5.9. The pairwise sequence identities between queries and template chains are 19% and 32%, respectively.

**SI Table S1. Properties and ZEPPI performance of each CAPRI target.**

ID	$N_{MSA}$	$\langle N_{IFR} \rangle$	$N_{\geq accpt.}$	$N_{incorrect}$	AUROC (ZEPPI)
Target29	623	64	167	1916	0.926
Target30	155	42	2	1341	0.550
Target32	62	70	15	584	0.687
Target35	161	65	3	496	0.528
Target37	142	48	99	1401	0.542
Target39	267	58	4	1396	0.843
Target40	55	61	588	1592	0.664
Target41	27	50	371	829	0.755
Target46	1396	56	24	1675	0.651
Target47	24	53	611	440	0.930
Target50	88	3	133	1318	0.734
Target53	2112	55	130	1270	0.742
Target54	198	49	19	1381	0.763

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [flatHonigepec.pdf](#)
- [flatHonigrs.pdf](#)