

SOFTWARE

PPI-ID: Streamlining protein-protein interaction prediction through domain and SLiM mapping

Haley V. Goodwin^{1,2*}, Nigel S. Atkinson^{1,2*}

1 Department of Neuroscience, University of Texas at Austin, Austin, Texas, United States of America, **2** Waggoner Center for Alcohol & Addiction Research, University of Texas at Austin, Austin, Texas, United States of America

* nigela@utexas.edu (NSA); hvg246@mit.edu (HVG)



Abstract

AlphaFold-Multimer models protein complexes and facilitates protein-protein interaction (PPI) prediction. Mapping of protein interaction domains and motifs onto the 3D structure can lend credence to the model and provide insight into the function of a given interaction. Furthermore, limiting structure prediction to only the domains and motifs that are likely to interact can reduce the computational demand and produce a higher quality model. To satisfy these needs, we built the Protein-Protein Interaction Identifier (PPI-ID). PPI-ID maps interaction domains and motifs onto molecular structures and filters for those that are sufficiently close to interact. Once an interface is found, PPI-ID labels interacting amino acids. Given only sequences, PPI-ID predicts regions for AlphaFold-Multimer modeling, reporting potential interactions only when each protein has one-half of a paired sequence. Testing with known dimers confirms high accuracy of the tool.

OPEN ACCESS

Citation: Goodwin HV, Atkinson NS (2025) PPI-ID: Streamlining protein-protein interaction prediction through domain and SLiM mapping. *PLoS Comput Biol* 21(10): e1013062. <https://doi.org/10.1371/journal.pcbi.1013062>

Editor: Shugang Zhang, Ocean University of China, CHINA

Received: April 18, 2025

Accepted: October 5, 2025

Published: October 16, 2025

Copyright: © 2025 Goodwin, Atkinson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All program code and data have been archived under the FigShare DOI of <https://doi.org/10.6084/m9.figshare.28266584.v5>.

Funding: This work was funded by the National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism grant to N.S.A.

Introduction

Many protein-protein interaction (PPI) domains and interaction motifs have been archived in databases. Domain-domain interactions (DDIs) between large structured regions are important for stable protein multimerization, while short linear motifs (SLiMs), often located in intrinsically disordered regions (IDRs), are thought to mediate short-lived signaling interactions and to play a lesser role in stable protein interactions [1].

Systematic curation and annotation of known protein domains and motifs have enhanced our ability to understand protein interactions, signaling pathways, regulatory pathways, and disease mechanisms [2,3]. The shortest of these are SLiMs, which are 3–10 amino acids long and defined by regular expressions. SLiMs are difficult to physically study because of apparent weak affinity for and transient interaction with their targets [1]. In addition, prediction algorithms overpredict these sequences in proteins, reducing their utility to indicate protein function. When interpreting the

[grant number R21AA030833]. Funding for open access charge is from The University of Texas at Austin. N.S.A. received summer salary support from the NIH via the above grant. Stipend support for H.V.G. is from The University of Texas at Austin College of Natural Sciences in the form of an Advanced Research Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

meaning of a SLiM in a protein, it is best to have supporting evidence that the motif is important. This usually involves physical protein-protein interaction assays, evolutionary conservation analysis, mutation analysis, and, very recently, mapping of the motif onto models of interacting proteins [4].

AlphaFold-Multimer can be used to predict protein structure and interactions [5,6]. Mapping interaction domains and motifs onto an existing PPI model (top-down approach) can lend credence to the model or provide insight to the function of a given interaction. Analyzing primary protein sequences for domains and motifs can also be useful for selecting which protein regions should be modeled (bottom-up approach). Bret et al. and Lee et al. showed that multimeric predictions can be improved if one focuses on sequences involved in the interface formation, thereby decreasing confounding molecular contacts [7,8].

Mapping interaction domains and SLiMs onto proteins is conveniently achieved using the InterPro and ELM websites (<https://www.ebi.ac.uk/interpro/> and <http://elm.eu.org/>, respectively) [2,9]. However, neither InterPro nor ELM can confirm that two proteins contain an appropriately matched complement of domains and/or motifs in the appropriate positions to interact. Manually evaluating these criteria can become time consuming and subject to error, especially if a protein has a large number of domains and/or motifs or if many protein pairs need to be surveyed. For the top-down approach, it would be helpful if matched pairs could be displayed on a 3D molecular model and if potential interfaces could be filtered for domains and motifs close enough to interact. For the bottom-up approach, it would be ideal if a program reported matches only if each protein contained one-half of a pair of interaction sequences. To facilitate these approaches, we generated the Protein-Protein Interaction Identifier (PPI-ID). The program can be accessed at <http://ppi-id.biosci.utexas.edu:7215/>. The complete code is also found in FigShare at <https://doi.org/10.6084/m9.figshare.28266584.v5>.

Design and implementation

Required libraries

PPI-ID is written in R with a web interface generated using Shiny. The InterPro, ELM, and UniProt database APIs were accessed with *httr*. The *r3dmol* and *bio3d* packages were used to import, format, and display pdb information [10,11].

Database composition

The PPI-ID database of 40,535 unique domain-domain interactions (DDIs) was compiled from the 3did (2025 release) and DOMINE (2011 release) databases (<https://3did.irbbarcelona.org/index.php> and <https://manticore.niehs.nih.gov/cgi-bin/Domine>, respectively) [3,12]. 3did provides DDIs mapped in crystal structures. DOMINE provides a combination of crystal structure and high-confidence predicted DDIs. To ensure dataframe compatibility when combining the DDI data from both databases, only columns containing Pfam IDs were retained for the datasets. The two DDI datasets were then concatenated to one another, and all duplicates present in the dataset were removed. An instructional Quarto script for DDI database preparation is provided in Supporting information.

The database of domain-motif interactions (DMIs) was prepared from the 2024 release of the ELM Database [2]. No further cleaning or preparation for the DMI information was done. This resulted in a database of 399 DMIs based on experimentally determined structures (crystal, EM, and NMR determined structures). In this database, domain information is stored according to Pfam ID. Motif information is stored according to the ELM classifiers.

Extracting domain/motif information

UniProt and InterPro APIs are respectively used to fetch amino acid sequences from protein accession numbers and to search sequences for the presence of protein domains. Regular expression searches are used to identify the presence of SLiMs. Then, a number of custom functions check Pfam or ELM IDs against the compiled DDI/DMI databases to determine whether a pair of domains or a domain and a motif constitute a potential interaction. If so, the pair, along with the associated amino acid ranges for the domains or motif are documented and reported to the user.

DDI prediction. Within the ‘Predict from Accession’ tab, accession numbers are used to access the InterPro API (<https://www.ebi.ac.uk/interpro/api/entry/pfam/protein/uniprot/> [Accession Number]. Within the ‘Predict from Sequence’ tab, user-provided TSV files obtained from InterProScan are analyzed.

DMI prediction. Within the ‘Predict from Accession’ tab, accession numbers are used to access the InterPro or the UniProt API for domain and fasta information, respectively. The UniProt API URL used within PPI-ID to fetch appropriate amino acid sequences is <https://rest.uniprot.org/uniprotkb/> [Accession Number]. The protein sequence is searched for the presence of a SLiM using the `str_locate_all()` function from the *stringr* package.

Within the ‘Predict from Sequence’ tab, user-provided TSV files from InterProScan and ELM Predict are analyzed. The user can also obtain SLiM information from PPI-ID directly, using the ‘Get SLiM Information’ tab. PPI-ID is capable of predicting DDIs or DMIs, but not motif-motif interactions, because available databases do not document motif-motif interactions. If a user attempts to seek motif information for both protein inputs, PPI-ID will return an error and prompt for domain information for at least one protein.

Filter by contact distance

If the user has provided a pdb file of the protein complex, the table of predicted DDIs/DMIs can be filtered for contact distance (in Å). Contact distance filtering is done by the `filter_by_distance()` function, which uses the `atom.selection()` and `cmap()` functions from the *bio3d* library to select alpha carbons and determine whether DDIs/DMIs are within the user-provided contact distance.

PPI-ID validation

A total of 80 PPI complexes were folded using AlphaFold-Multimer (version 2.3.2), with 40 folds for DDI validation and 40 folds for DMI validation. AlphaFold-Multimer was run at the Texas Advanced Computing Center (TACC) using the reduced database. A total of 5 machine learning models were produced with a single pdb prediction per model.

For DDI validation, 40 PDB database entries were randomly selected from a collection of 14,972 crystal structures containing DDIs, as curated by 3did [3]. Only PDB database entries representing inter-protein interactions were selected. Entries containing synthetic proteins were excluded (because synthetic proteins are manmade and may contain modified amino acids and because UniprotKB does not currently have accession numbers for synthetic proteins). For bottom-up validation, the accession numbers were input into PPI-ID. For top-down validation, contact distance filters of 4–11 Å were used. In both cases, program output was checked to confirm that the interacting domains were correctly identified. To validate PPI-ID for the capacity to identify DMIs we used a dataset curated by Bret et al. [7] for the validation of AlphaFold2-Multimer structures. Bret et al. downloaded the July 3 2023 ELM database. They filtered for receptor ligand/receptor pairs using the “LIG” and “DOC” ELM class identifiers that had also been described in at least one PubMed

article. From this collection of 1884 entries, Bret et al. selected those that only exhibited one binding site in the ligand and also had a reference structure in the RCSB PDB database [13]. This resulted in a collection of 923 protein pairs. For PPI-ID DMI validation, 40 PDB database entries were randomly selected from this collection [7]. The previously described bottom-up and top-down procedures used for DDI validation were used. Validation data can be found in S1 and S2 Tables, respectively. The models used to validate PPI-ID are available from the online program under the About & Help tab.

Results

PPI-ID facilitates structural analyses of predicted dimeric models through molecular visualization, contact distance-based interface identification, and contact residue labeling. PPI-ID predicts potential interaction interfaces from protein sequence information using information from ELM, 3did, Interpro, and DOMINE databases [2,3,9,12]. In the top-down approach (Fig 1, right), PPI-ID graphically maps domain-domain interactions (DDIs) and domain-motif interactions (DMIs) onto the 3D model of the protein dimer only if each protein contains compatible interaction sequences. Further stringency can be applied to predicted top-down interactions through the identification of only those protein-protein interfaces that come within a specified distance of one another.

For the bottom-up approach (Fig 1, left), PPI-ID displays the position of DDI and DMI sequences in the primary amino acid sequence for two proteins only when each protein contains one-half of a pair of protein:protein interaction signature sequences. It expresses these as a table of the amino acid residue numbers from each protein that are predicted to interact. This information might then be used to select the amino acid residues to be used to model the interacting regions of the proteins.

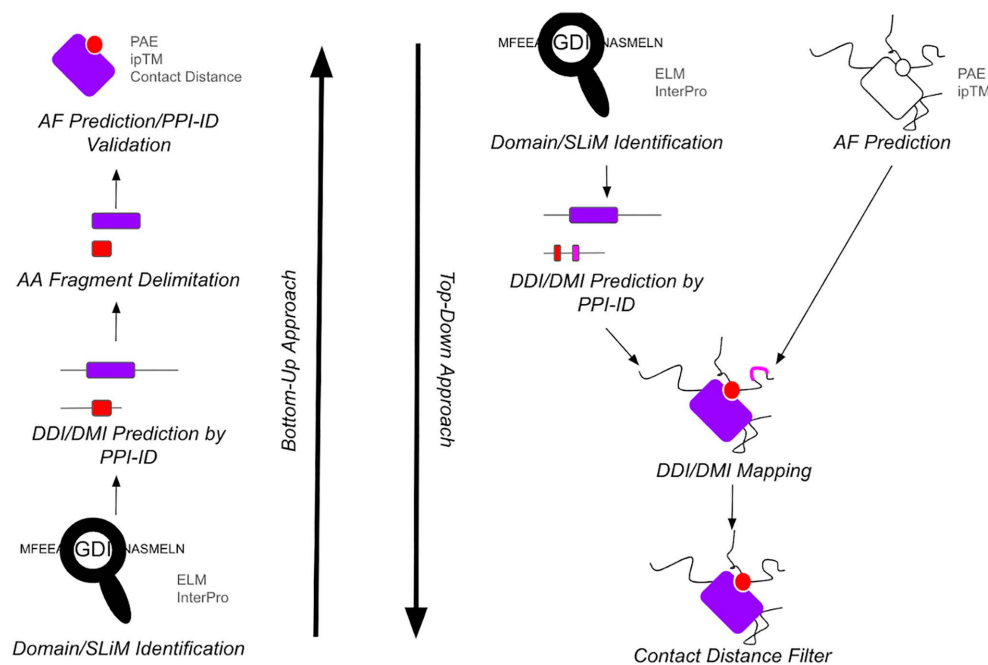


Fig 1. Summary of analyses by PPI-ID. *Left side:* The bottom-up approach searches two protein fasta files for domains and SLiMs in each protein that could facilitate a physical interaction between the two proteins. Output is presented in a way that facilitates cropping of the fasta files before using AlphaFold-Multimer which can favor the generation of a higher accuracy model of the region. *Right side:* In the top down approach, protein-protein interaction domains/SLiMs are mapped onto a 3D heterodimer model of two proteins and their positions visualized. Domains or SLiMs that are too far apart to interact can be filtered out thereby enriching for sequence pairs that have the greatest potential for a physical interaction.

<https://doi.org/10.1371/journal.pcbi.1013062.g001>

Case studies presented in this paper take advantage of the 'Predict from Accession' tab within PPI-ID. However, in the 'Predict from Sequence' tab, users are also able to upload domain information for a protein (obtained from InterProScan) and motif information for a protein (obtained using the 'Get SLiM Information' tab of PPI-ID). The benefit of uploading annotated sequence information in the 'Predict from Sequence' tab is that user predictions are not limited to proteins with associated UniProt accession numbers. As a result, users can perform computational PPI prediction and analysis with synthetic proteins, uncharacterized isoforms, and protein fragments. There is a detailed instructional video and sample pdb files under the About & Help tab of the program.

PPI prediction

PCNA and DNMT1 interaction, top-down example. For the top-down approach, PPI-ID presents a detected DMI between human proliferating cell nuclear antigen (PCNA) and DNA methyltransferase 1 (DNMT1) proteins (Fig 2). Many PCNA pairing partners contain a conserved motif called the PCNA-interacting protein (PIP) box [14–16]. One PIP box-containing protein that has been demonstrated to interact with PCNA is DNMT1 [17].

Protein-Protein Interaction Identifier

Predict from Accession Predict from Sequence Get SLiM Information About

Provide UniProt Accession Numbers for 2 Proteins Below:

P12004 P26358

Protein 1 Protein 2

Domain Domain

Motif Motif

Select for each protein if you would like to fetch domain or motif information. PPIID can predict domain-domain interactions or domain-motif interactions (at least one protein must have 'Domain' selected).

Enter

Download Results

Choose PDB file for Multimer

Browse... relaxed_model_1_multimer_v3_f

Upload complete

Show PDB Hide PDB

Revert Model to Original

Contact Distance Filter:

6

Show 10 entries Search:

Domain 1	SLiM 2	Range 1	Range 2	Residue 1	Residue 2
1	LIG_PCNA_PIPBox_1	127-254	164-173	254, 252, 127	164, 167, 172

Showing 1 to 1 of 1 entries Previous 1 Next

Fig 2. The top-down approach using the human PCNA (P12004) and DNMT1 (P26358) proteins. Contact Distance Filter was used to only show predicted interactions that come within 6 Å. The region selected in the table is displayed on the protein in green and yellow (PCNA and DNMT1, respectively). The Add Contact Labels button (not pictured) labels residues within the distance filter limit and represents them in ball-and stick form with carbon in gray, oxygen in red, and nitrogen in light blue.

<https://doi.org/10.1371/journal.pcbi.1013062.g002>

InterProScan sequence alignment shows the PCNA amino acid sequence contains 20 domain entries. DDI and DMI are stored by Pfam ID or ELM ID, so PPI-ID only analyzes entries using these identifiers. A search of the PCNA sequence identifies two Pfam domains: the PCNA C-terminal domain and the PCNA N-terminal domain. Regular expression searches on the DNMT1 amino acid sequence identifies 546 possible SLiMs. Manually assessing the compatibility of every possible domain:motif pairing would require the manual inspection of 1092 pairs.

PPI-ID automates the identification of potential interactions between the PCNA C-terminal domain and two PIP box motifs in DNMT1. When a 6 Å contact distance filter and contact residue labeling are applied, PPI-ID identifies and highlights residues likely important for mediating interactions between PCNA and DNMT1 – specifically Ala252 and Gly127 of PCNA which are known to hydrogen bond with Gln164 and Ala172 of DNMT1, respectively [17]. Furthermore, Ile167 of DNMT1, also identified by the contact filter, has been shown to plug into the hydrophobic binding pocket of PCNA [17].

In practice, when using the Top-Down approach we initially view the domains and or motifs contained in the proteins without setting a contact filter distance. The contact filter is then used to identify those that interact at biologically meaningful distances. To do so, the filter is incrementally varied, starting at 11 Å and working down to 4 Å in one Å steps. The maximum distance cutoff of 11 Å between alpha carbons is based on the simplified alpha-carbon coarse-grained elastic network model proposed by Jeong et al. [18]. Typically, we consider interactions greater than 11 Å to be less likely to be biologically meaningful. Furthermore, we stop our top-down search at 4 Å because we have never observed an interaction occurring at smaller distances. When using PPI-ID for the top-down approach, *it is important to refresh the browser window between each set of proteins to clear all of the internal variables from PPI-ID – otherwise PPI-ID will return erroneous information.*

DifA and cactus interaction, bottom-up example. For the bottom-up approach, PPI-ID displays the position of signature sequences for DDIs or DMIs that could contribute to an association between the two proteins. The *Drosophila* DifA and Cactus proteins are well-known interacting partners, with this interaction taking place between the Rel homology region dimerization domain of DifA and the ankyrin repeats of Cactus [19]. Fig 3 shows the result of submitting the accession numbers of DifA and Cactus proteins for DDI prediction. Namely, the DNA-binding and dimerization domains of the DifA Rel-homology region are identified as potential interactors with ankyrin repeat domains found within Cactus. This information can be used to delimit the protein sequences for AlphaFold-Multimer modeling, which enhances the accuracy of the structure [7,8].

PPI-ID validation

80 PPIs with experimentally determined structures were folded with AlphaFold-Multimer. 40 PPIs were used to validate DMI prediction accuracy, and the remaining 40 were used to validate DDI prediction accuracy. A detailed description of how these protein complexes were chosen can be found in Methods. The accuracy of PPI-ID top-down and bottom-up analyses is summarized in Table 1.

Top-down validation. Pairs of protein accession numbers and pdb files of predicted complexes were submitted to PPI-ID. A prediction was considered successful if the proper interaction interface was predicted by AlphaFold-Multimer, identified by PPI-ID, and confirmed by an experimentally-determined structure [7]. PPI-ID top-down validation yielded respective accuracy rates of 85% and 73% for DDIs and DMIs. Using a similar top-down approach, Bret et al. reported that AlphaFold-Multimer had a predictive performance of 52.7% [7]. The increased success of PPI-ID over the manual analysis performed by Bret et al. [7] must be a product of sampling error on our part when we selected our test cases.

We performed a sensitivity analysis to understand the relationship between contact filter distance and correct predictions. Correct predictions are those that match the experimentally determined protein-protein interface. Five and four Å distances produce the greatest number of positive identifications (S1 Fig). Because AlphaFold-Multimer folding accuracy degrades with very large proteins and large multiprotein complexes, we expect that the prediction capability of PPI-ID will likewise decline in parallel [7,8].

Protein-Protein Interaction Identifier

Predict from Accession | Predict from Sequence | Get SLiM Information | About

Provide UniProt Accession Numbers for 2 Proteins Below:

P98149 | Q03017

Protein 1: Domain Motif

Protein 2: Domain Motif

PPI-ID can predict domain-domain interactions or domain-motif interactions (at least one protein must have 'Domain' selected). If performing predictions for more than one protein-pairing, please refresh the browser between each prediction.

Enter

Download Results

Choose PDB file for Multimer

Browse... | No file selected

Show PDB | Hide PDB

Revert Model to Original

Show 10 entries

Search:

	Domain 1	Domain 2	Range 1	Range 2
1	Rel homology DNA-binding domain	Ankyrin repeat	80-250	298-324
2	Rel homology dimerisation domain	Ankyrin repeats (3 copies)	256-358	224-296
3	Rel homology dimerisation domain	Ankyrin repeats (3 copies)	256-358	342-419

Showing 1 to 3 of 3 entries

Previous | 1 | Next

Fig 3. The bottom-up approach using the DifA (P98149) and Cactus (Q03017) proteins, which are known to bind. The table shows four predicted interactions that take place between ankyrin repeats of Cactus and the Rel homology domain of DifA. Relative positions of each domain are shown.

<https://doi.org/10.1371/journal.pcbi.1013062.g003>

Table 1. Percentage of known DDIs and DMIs successfully identified. Forty protein pairs were used to validate DDI identification capability, and 40 protein pairs were used to validate DMI identification capability.

Interaction Type	Bottom-Up % Correct	Top-Down % Correct
Domain-Domain (DDIs)	98%	82.5%
Domain-Motif (DMIs)	98%	73%

<https://doi.org/10.1371/journal.pcbi.1013062.t001>

Bottom-up validation. Pairs of protein accession numbers were submitted to PPI-ID. A prediction was considered successful if the proper interaction interface was identified in PPI-ID output. PPI-ID bottom-up validation yielded respective accuracy rates of 98% for both DDIs and DMIs. The missing DDI and DMI identifications arose because of missing Pfam or SLiM identifiers in the InterPro and ELM databases.

We explored the cause of the reduced top-down accuracy compared to bottom-up accuracy. The failed identifications occurred even though the domains and motifs are present. They failed because the contact filter showed that they occurred at a non-biologically relevant distance (S1 Fig). These misses occurred because AlphaFold-Multimer produced a model that differed from the experimentally determined structure. This could be because the AlphaFold-Multimer model is an alternative conformation or more simply reflects an error by AlphaFold-Multimer. We know this to be the case because PPI-ID does identify the interaction in the experimentally determined structures within a reasonable distance. In addition, one should note the capacity of PPI-ID to correctly identify DMIs and DDIs is dependent on the quality and quantity of data in the InterPro and ELM databases.

Using PPI-ID with PAE viewer to enhance understanding of motifs in IDRs

Another AlphaFold product that can be evaluated to ascertain the quality of a PPI model is Predicted Aligned Error (PAE). PAE expresses the error in Å of the relative position of two amino acids. A low error within interdomain quadrants of PAE plots, which are quadrants 1 and 3 (top right and bottom left quadrants; Fig 4), suggests a physical interaction between the proteins [20]. The PAE Viewer web tool plots PAE values and lets one determine the residue numbers of all plotted points, a feature that complements PPI-ID (<https://subtiwiki.uni-goettingen.de/v4/paeViewerDemo>) [21]. If a PPI-ID predicted DDI or DMI were valid, one would expect the predicted interface to have a very low PAE score. In turn, one can determine whether regions of low PAE value house known interaction domains or motifs. Such outcomes would be supportive that the interaction is real.

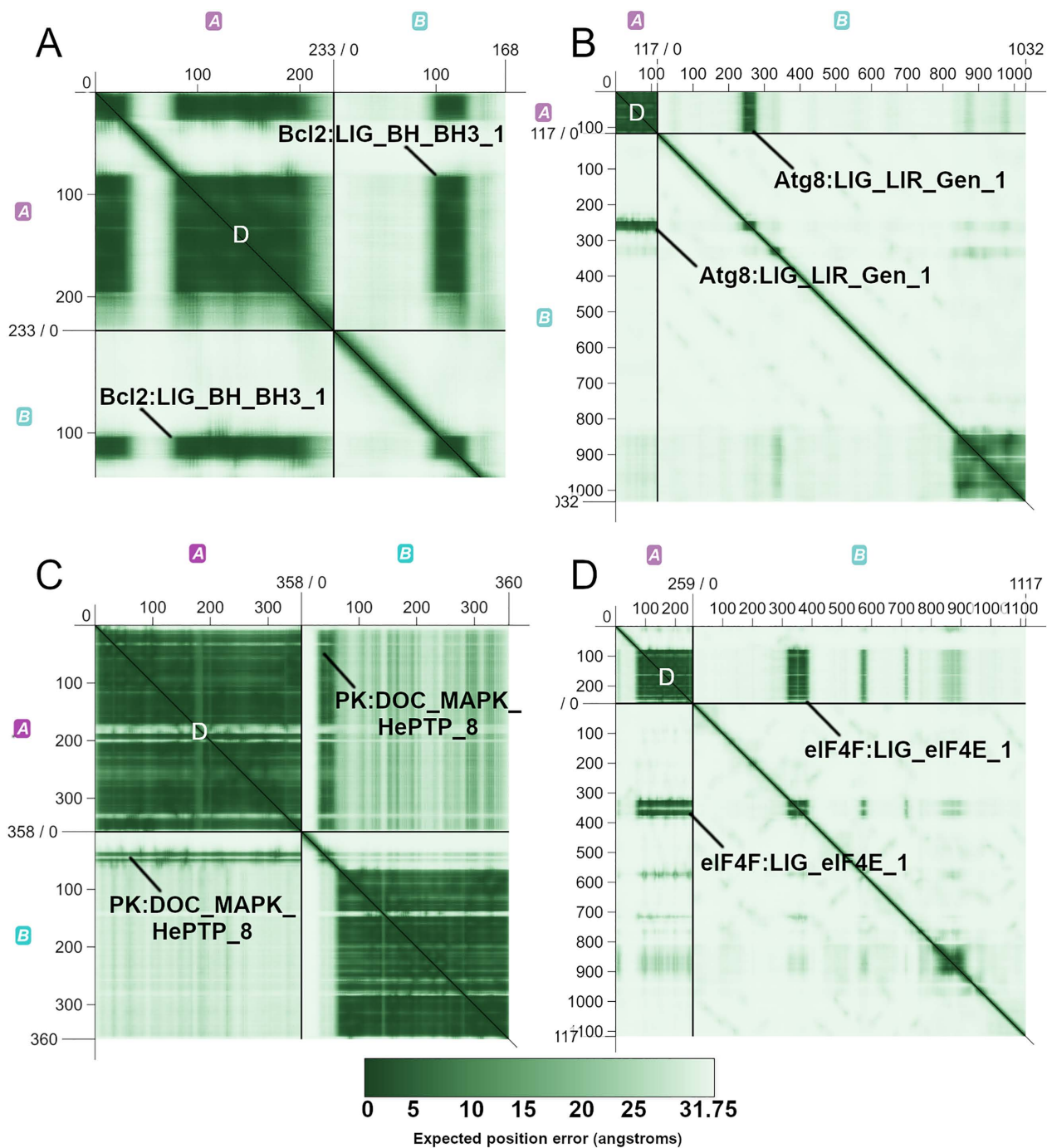
The protein pairs in Fig 4 were processed with PPI-ID, which mapped a SLiM inside an IDR of protein B and predicted that the SLiM is within 5 Å of the target domain within protein A (labeled in Fig 4). PPI-ID mapped interacting protein regions solely based on information about, and the distances between, domains and motifs. PPI-ID provided the residue numbers for the DMIs, and PAE Viewer provided residue numbers of the regions of low PAE within the IDR of protein B. We used these residue numbers to correlate the information from the PAE plot to the output from PPI-ID. In Fig 4, a DMI is predicted to occur at the sites of low relative positional error between the proteins, bolstering the evidence that PPI-ID provides a salient interpretation of the protein motifs documented within the SLiM database provided by ELM.

Applications and limitations

AlphaFold-Multimer is an important tool for advancing our understanding of how proteins interact [5,6]. PPI-ID is a program that can be run from its website or downloaded and run locally. It identifies potential interaction interfaces between proteins and characterizes interaction interfaces in AlphaFold-produced models. The top-down approach (identification of potential DDIs and DMIs that are in close proximity) performed less well than the bottom-up approach (simple search of protein sequences for compatible DDI and DMI pairs). It is likely that this occurred either because the model tested did not represent the exact conformation in which the interface interaction actually occurs or as a result of AlphaFold-Multimer error (we confirmed that PPI-ID could identify the interface in the experimentally determined structure). It is to be expected that the incidence of such failures will rise with large proteins, because both Bret et al. [7] and Lee et al. [8] have reported declines in AlphaFold-Multimer accuracy as protein sizes increase. Bret et al. [7] suggest that diminished accuracy in larger complexes arises from intramolecular contacts that mask correct domain/motif interaction.

Additionally, top-down performance is predicted to worsen when attempting to predict PPIs between two polypeptides that are part of a complex composed of more than two proteins. This is because we expect that multimeric protein complexes are synergistically held together by each constituent polypeptide chain, and missing polypeptides may lead to the absence of PPIs essential for proper complex assembly. Multiprotein complexes were certainly overrepresented in instances in which the DDI interface was not correctly identified, however, we also had 10 instances in which the DDI from proteins belonging to a multiprotein complex were correctly identified. Thus, PPI-ID performed better than our expectations with multiprotein complexes.

PPI-ID is a convenient tool for scanning a pair of proteins or a model of a protein dimer for protein domains and motifs that are annotated as interacting in well-curated standard databases. As such, it is a knowledge-based tool whose performance is capped by the completeness of the underlying DDI/DMI databases and the capacity to generate accurate models. That is, the veracity of PPI-ID mapping of DDIs and DMIs is dependent on the quality of the domain and motif databases and the quality of the protein model. Since domain and motif databases and protein modeling fidelity continuously improve, we expect PPI-ID performance to also improve over time. We recognize that for any given protein-protein interaction prediction, it would be best to confirm the interaction by a physical technique. However, interactive rapid screening, as that provided by PPI-ID, can help focus one on the most likely interaction candidates.



	Protein A	Protein B	SLiM	PPI-ID	PAE Viewer Q1	PAE Viewer Q3
A	Q07817	Q92934	LIG_BH_BH3_1	106-122	96-132	104-124
B	O95166	O95197	LIG_LIR_Gen_1	245-256	239-280	250-273
C	P63086	P35236	DOC_MAPK_HePTP_8	38-50	30-58	38-54
D	P48598	Q9VMA3	LIG_eIF4E_1	327-333	320-389	331-370

Fig 4. PPI-ID in conjunction with PAE Viewer enhances confidence in the protein interactions visualized in a model. Panels are PAE plots of DMI protein pairs. Within each plot, quadrant 1 (Q1) is the top right hand quadrant and quadrant 3 (Q3) is the bottom left quadrant. Q1 and Q3 represent interactions between a motif in protein B and a domain in protein A. The heat map represents the positional error between amino acid pairs measured in Å.

PAE Viewer reports residue numbers from positions on the plot. Using these, we correlated the outputs of PPI-ID and PAE Viewer. The PPI-ID contact distance filter was set to $\leq 5 \text{ \AA}$. SLiMs identified by PPI-ID are labeled. The predicted domain that the SLiM interacts with is the dark green square to the left of the labeled Q1 SLiM and above the labeled Q3 SLiM. The domains, identified by the white D in panels A-D, are the Bcl-2 apoptosis regulator domain (Bcl-2), autophagy protein Atg8 ubiquitin like (Atg8), protein kinase domain (PK), and eukaryotic initiation factor 4E domains (eIF4E), respectively. The table shows residue numbers of SLiMs identified by PPI-ID and the PAE Viewer identified residues with low PAE scores (dark stripes) in Q1 and Q3. The programs show good agreement in identification of the same sequences based on distinct criteria.

<https://doi.org/10.1371/journal.pcbi.1013062.g004>

PPI-ID was developed to screen through a list of PPI candidates and is particularly useful when used in conjunction with PAE Viewer [21]. Using both programs in tandem enables the correlation of interaction domains and motifs with regions of low positional error. Such analysis provides additional support for the idea that the identified domains and motifs of interest stabilize one another and drive a protein:protein association.

Another program focused on interpreting SLiMs is SLiMAN [22]. The unique capabilities of SLiMAN include interactome-level analysis of DMIs and the filtering for specific properties, e.g., disordered regions or post-translational modifications. On the other hand, the unique capabilities of PPI-ID include the prediction of both DDIs and DMIs, PPI visualization that is granular, rapid, and interactive, and filtering protein:protein interfaces based on their intermolecular distance. SLiMAN seems well suited for characterizing a known protein interactome while PPI-ID is best suited for providing evidence of whether protein pairs belong to the same interactome. We imagine that sequential use of SLiMAN and PPI-ID could have utility.

PPI-ID is capable of visualizing protein:protein interfaces, filtering interfaces by contact distance, and residue labeling. Furthermore, PPI-ID performs DDI and DMI prediction from protein sequence information. The identification of protein:protein interactions may extend beyond academic interest and have a practical impact on drug discovery and therapeutics development. While a protein encoded by a disease-causing gene might not be itself druggable, the identification of protein:protein pairing partners and the site of the interactions between the proteins might provide an alternate druggable target that could modulate the same pathway. PPI-ID is a compact and flexible tool that streamlines the processes of both PPI structural analysis and protein sequence delimitation for structural prediction.

Availability and future directions

All program code and data have been archived on FigShare at <https://doi.org/10.6084/m9.figshare.28266584.v5>. A running version of the application and all datasets used to train the program are at <http://ppi-id.biosci.utexas.edu:7215/>.

The PPI-ID database is a compilation of data from other protein interaction databases. At regular intervals the web version of the PPI-ID database will be updated as new information becomes available. The end user can update a local version of the PPI-ID database using the provided Quarto script.

Currently, the distances between interfaces is presented as the distance between alpha carbons. An envisioned future direction is an upgrade so that H bond and salt bridge distances can be displayed.

Supporting information

S1 Fig. Distribution of positive identifications and failed identifications of domain:motif interactions (DMI) and domain:domain interactions (DDI) in AlphaFold-Multimer test models (S1 and S2 Tables). Contact filter distance was varied from 3 to 15 angstroms between alpha carbons in integer steps. **Panels A and B)** Distances at which correct inter-protein interfaces were detected. Correct interfaces were determined from crystal structures. Each slice of the pie chart is identified with the contact filter distance at which the DMI or DDI was identified. Below the contact distance is the percent of the test cases that were detected at this filter setting. No interactions were detected below 4 angstroms and no correct interactions were observed above 15 angstroms. Panel A shows the distribution for correct DMI identifications and Panel B shows the distribution for correct DDI identifications. **Panels C and D)** For some models, the correct DMI or DDI was

not identified. On the X axis are the ranges, in angstroms, at which an interaction was detected. Failures were distributed over ranges shown. DMI failures occurred at contact filter settings of 4, 5, 5, 7, 21, 33, 39, 64, and 70 angstroms. DDI failures occurred at 14, 16, 18, 22, 27, and 49 angstroms.

(TIFF)

S1 Table. DMI validation information. In the Bottom-Up and Top-Down columns, a 1 corresponds with a successful prediction, and 0 corresponds with a failed prediction.

(XLSX)

S2 Table. DDI validation information. In the Bottom-Up and Top-Down columns, a 1 corresponds with a successful prediction, and 0 corresponds with a failed prediction.

(XLSX)

S1 Script. ppi-id_ddi_compilation.qmd is a Quarto script to recreate the database used by PPI-ID for the identification of domain:domain interactions (DDI).

(QMD)

Acknowledgments

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources that have contributed to the research results reported within this paper. URL: <http://www.tacc.utexas.edu>. The authors also thank Mr. James R. Derry for his assistance helping us navigate the university's web services.

References

1. Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci J Virtual Libr.* 2008;13:6580–603.
2. Kumar M, Michael S, Alvarado-Valverde J, Zeke A, Lazar T, Glavina J, et al. ELM-the Eukaryotic Linear Motif Resource-2024 Update. *Nucleic Acids Research.* 2024;52(D1):D442-55.
3. Mosca R, Céol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 2014;42(Database issue):D374-9. <https://doi.org/10.1093/nar/gkt887> PMID: [24081580](https://pubmed.ncbi.nlm.nih.gov/24081580/)
4. Strom JM, Luck K. Bias in, bias out - AlphaFold-Multimer and the structural complexity of protein interfaces. *Curr Opin Struct Biol.* 2025;91:103002. <https://doi.org/10.1016/j.sbi.2025.103002> PMID: [39938238](https://pubmed.ncbi.nlm.nih.gov/39938238/)
5. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: [34265844](https://pubmed.ncbi.nlm.nih.gov/34265844/)
6. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer [Internet]. 2021 [cited 2024 July 10]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034>
7. Bret H, Gao J, Zea DJ, Andreani J, Guerois R. From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nat Commun.* 2024;15(1):597. <https://doi.org/10.1038/s41467-023-44288-7> PMID: [38238291](https://pubmed.ncbi.nlm.nih.gov/38238291/)
8. Lee CY, Hubrich D, Varga JK, Schäfer C, Welzel M, Schumbera E, et al. Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. *Mol Syst Biol.* 2024;20(2):75–97.
9. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. *Nucleic Acids Res.* 2023;51(D1):D418-27.
10. Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics.* 2015;31(8):1322–4.
11. Grant BJ, Skjaerven L, Yao X-Q. The Bio3D packages for structural bioinformatics. *Protein Sci.* 2021;30(1):20–30. <https://doi.org/10.1002/pro.3923> PMID: [32734663](https://pubmed.ncbi.nlm.nih.gov/32734663/)
12. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* 2011;39(Database issue):D730-5. <https://doi.org/10.1093/nar/gkq1229> PMID: [21113022](https://pubmed.ncbi.nlm.nih.gov/21113022/)
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235–42. <https://doi.org/10.1093/nar/28.1.235> PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)

14. Moldovan G-L, Pfander B, Jentsch S. PCNA, the maestro of the replication fork. *Cell*. 2007;129(4):665–79. <https://doi.org/10.1016/j.cell.2007.05.003> PMID: [17512402](https://pubmed.ncbi.nlm.nih.gov/17512402/)
15. Dieckman LM, Freudenthal BD, Washington MT. PCNA Structure and Function: Insights from Structures of PCNA Complexes and Post-translationally Modified PCNA. In: MacNeill S, editor. *The Eukaryotic Replisome: a Guide to Protein Structure and Function* [Internet]. Dordrecht: Springer Netherlands; 2012 [cited 2025 Jan 5]. p. 281–99. (Subcellular Biochemistry; vol. 62). Available from: https://link.springer.com/10.1007/978-94-007-4572-8_15
16. Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*. 1994;79(7):1233–43. [https://doi.org/10.1016/0092-8674\(94\)90014-0](https://doi.org/10.1016/0092-8674(94)90014-0) PMID: [8001157](https://pubmed.ncbi.nlm.nih.gov/8001157/)
17. Jimenji T, Matsumura R, Kori S, Arita K. Structure of PCNA in complex with DNMT1 PIP box reveals the basis for the molecular mechanism of the interaction. *Biochem Biophys Res Commun*. 2019;516(2):578–83. <https://doi.org/10.1016/j.bbrc.2019.06.060> PMID: [31235252](https://pubmed.ncbi.nlm.nih.gov/31235252/)
18. Jeong JI, Jang Y, Kim MK. A connection rule for alpha-carbon coarse-grained elastic network models using chemical bond information. *J Mol Graph Model*. 2006;24(4):296–306. <https://doi.org/10.1016/j.jmgl.2005.09.006> PMID: [16289973](https://pubmed.ncbi.nlm.nih.gov/16289973/)
19. Malek S, Huang D-B, Huxford T, Ghosh S, Ghosh G. X-ray crystal structure of an I kappa B beta x NF-kappa B p65 homodimer complex. *J Biol Chem*. 2003;278(25):23094–100. <https://doi.org/10.1074/jbc.M301022200> PMID: [12686541](https://pubmed.ncbi.nlm.nih.gov/12686541/)
20. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, et al. AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for Over 214 Million Protein Sequences. *Nucleic Acids Res*. 2024;52(D1):D368-75.
21. Eifmann C, Stülke J. PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acids Res*. 2023;51(W1):W404-10.
22. Reys V, Pons JL, Labesse G. SLiMAn 2.0: meaningful navigation through peptide-protein interaction networks. *Nucleic Acids Res*. 2024;52(W1):W313–7.