JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 31, Number 8, 2024
© Mary Ann Liebert, Inc.
Pp. 769–781
DOI: 10.1089/cmb.2024.0554

Protocol

Combining Complementarity and Binding Energetics in the Assessment of Protein Interactions: EnCPdock—A Practical Manual

GARGI BISWAS,[1] DEBASISH MUKHERJEE,[2] and SANKAR BASU[3]

ABSTRACT

The combined effect of shape and electrostatic complementarities (Sc, EC) at the interface of the interacting protein partners (PPI) serves as the physical basis for such associations and is a strong determinant of their binding energetics. EnCPdock (https://www.scinetmol.in/EnCPdock/) presents a comprehensive web platform for the direct conjoint comparative analyses of complementarity and binding energetics in PPIs. It elegantly interlinks the dual nature of local (Sc) and nonlocal complementarity (EC) in PPIs using the complementarity plot. It further derives an AI-based $\Delta G_{binding}$ with a prediction accuracy comparable to the *state of the art*. This book chapter presents a practical manual to conceptualize and implement EnCPdock with its various features and functionalities, collectively having the potential to serve as a valuable protein engineering tool in the design of novel protein interfaces.

Keywords: complementarity, complementarity plot, free energy of binding, protein–protein interaction, structure-based thermodynamics.

1. INTRODUCTION

Design of novel therapeutic agents and the engineering of proteins with desired functionalities are interconnected forefront of modern biomedical research. On one hand, understanding the thermodynamics and kinetics of protein–protein interactions (PPIs) is essential in identifying and validating potential drug targets (Feng et al., 2017), whereas on the other hand, protein–protein-binding energetics play a pivotal role in rational protein engineering, interface design, and modulation of PPIs (Sable and Jois, 2015). Apart from the applications, the interaction between protein complexes is also crucial for deciphering their roles in cellular processes, diseases, mechanisms, and signal transduction pathways (Keskin et al., 2008). A significant portion of the data concerning protein–protein-binding energetics is encoded within the structural features of protein complexes (Zhang et al., 2012). The three-dimensional arrangement of proteins within these complexes reveals critical

[3] Department of Microbiology, Asutosh College, University of Calcutta, Kolkata, India.
[1] Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel.
[2] Institute of Molecular Biology gGmbH (IMB), Mainz, Germany.

insights into the nature and specificity of their interactions, offering a rich source of information for understanding the energetic basis of their associations (Bryant et al., 2022).

The interplay between protein–protein-binding energetics and the structural characteristics of PPI complexes has been elucidated with numerous computational approaches. Within the realm of molecular dynamics simulations, a prominent technique extensively used is the physical effective energy function (PEEF). PEEF relies on theoretically derived interparticle forces that encompass all atoms within a given structure of protein complexes (Lazaridis, 2000; Lazaridis and Karplus, 1997). The parameters of PEEF are typically obtained from small molecule crystal and solvation data, as well as *ab initio* calculations (Brooks et al., 1983; MacKerell et al., 1998). However, because of the absence of parametrization from actual protein structures, the PEEF approach encountered challenges in accurately identifying native protein folds (Lazaridis and Karplus, 1997). More specifically, as the PEEF is derived from atomic models, it frequently exhibits a rugged energy surface that lacks a smooth descent when approaching the native state (Lazaridis and Karplus, 1999).

A statistical effective energy function (SEEF) addresses the limitations of PEEF by using parametrization from a database of known protein structures to extract statistics related to pair contacts and surface area burial (Sippl, 1995; Simons et al., 1999). This enables the determination of "pseudo-potentials" for protein structures or PPIs. SEEF offers advantages over PEEF, including a smoother energy landscape and reduced sensitivity to small perturbations (Simons et al., 1999). Moreover, its statistical nature allows for the inclusion of all known and potential physical effects, enhancing its robustness (Thomas and Dill, 1996). However, SEEF may exhibit a lower discriminatory power because of this very robustness (Simons et al., 1997).

Among other methodologies, a widely used approach involves combining molecular mechanics energy (MM) with solvation-free energy and configurational entropy (Chen et al., 2016). The molecular MM incorporates bond, angle, dihedral, electrostatic, and van der Waals energy in the gas phase. Conformational entropy is typically computed from normal-mode analysis based on a set of conformational snapshots obtained from molecular dynamics simulations. Solvation-free energy, on the contrary, is determined by calculating the change in free energy associated with transferring a molecule from an ideal gas to a solvent at a specific pressure and temperature (Duarte Ramos Matos et al., 2017). This process considers alterations in solvent accessible surface area (SASA) and electrostatic interactions between the solute and solvent. The electrostatics part can be determined using either the Generalized Born (GB) model (Gohlke and Case, 2004) or by solving the finite difference Poisson–Boltzman (PB) equation (Brown and Muchmore, 2006), which leads to the MM/PBSA and MM/GBSA approaches, respectively. Although both methods share the entropic, SASA, and molecular mechanics components, their treatment of electrostatics differs based on the charge model, force field, radius parameter in the continuum solvent model, and solvent dielectric constant. Generally, MM/PBSA outperforms MM/GBSA in predicting protein–protein-binding free energies (Chen et al., 2016). However, it is crucial to note that MM/PBSA's sensitivity to the dielectric constant of the solute necessitates careful calibration based on the charge distribution of the binding interface in PPI complexes (Chen et al., 2016).

The usage of artificial intelligence to predict protein–protein-binding affinities is a recent development. Many approaches focus on determining the changes in binding free energy resulting from one or multiple mutations in PPI complexes. For instance, mmCSM-PPI (Rodrigues et al., 2021), Geo-PPI (Liu et al., 2021), TopNetTree (Wang et al., 2020), and PPI-affinity (Romero-Molina et al., 2022) use extra-tree, gradient-boosting trees and support vector machine (SVM) algorithms to achieve a Pearson's correlation coefficients (r) of 0.75, 0.52, 0.79, and 0.78, respectively, between predicted and experimental data (ΔΔG) for the SKEMPI 2.0 database (Jankauskaite et al., 2019), thereby predicting changes in binding affinity upon mutations. In the study conducted by Romero-Molina et al., the PPI-affinity method demonstrated a general applicability for predicting the binding free energy of diverse PPI complexes (Romero-Molina et al., 2022). The PPI affinity method achieved an *r* value of 0.62 between experimental and predicted binding free energies for a training dataset comprising 833 PPI complexes (Test set 1). However, the *r* value dropped to 0.50 when evaluated on a separate hold-out test dataset consisting of 90 PPI complexes (Test set 2) (Romero-Molina et al., 2022). Furthermore, the performance of PP affinity was compared with other previously available methods to predict the protein–protein-binding affinity on both Test set 1 and Test set 2. PRODIGY, another method that uses an SEEF approach, exhibited an *r* value of 0.74 on Test set 1 (on which it was trained), but its performance declined with an *r* value of 0.31 on Test set 2, indicating potential overfitting toward the benchmark dataset (Xue et al., 2016). In addition, DFIRE (Liu et al., 2004), CP_PIE (Ravikant and Elber, 2010), and ISLAND (Abbasi et al., 2020) displayed *r* values of 0.10, −0.10, and 0.27, respectively, on the hold-out dataset (Test set 2). It is noteworthy that all of these available methods use an SEEF in predicting the protein–protein-binding affinity. Relatively newer approaches of predicting binding affinity and assessing thermal stability in proteins have also explored analysis of interfacial contact networks and the spatial organization of hydrophobic and charged residues (Desantis et al., 2022; Vangone and Bonvin, 2015).

On the contrary, EnCPdock (Biswas et al., 2023), trained on a dataset comprising 3200 PPI complexes with binding free energies calculated using FoldX (Blanco et al., 2018), used a support vector regression approach. Cross validation of EnCPdock yielded a maximum correlation (r_{max}) value of 0.745 between the target function (ΔG_{FoldX_norm}) and the predicted output ($\Delta G_{EnCPdock\_norm}$), with a corresponding maximum balanced accuracy (BACC) score of 0.833. Furthermore, EnCPdock's performance was evaluated on two independent datasets, namely the Affinity benchmark dataset and the "SKEMPI + PROXiMATE–merged" datasets, comprising 106 and 236 binary complexes, respectively. It achieved correlation coefficients of 0.45 and 0.52, respectively, between the predicted $\Delta G_{EnCPdock\_norm}$ and the actual binding free energies for these datasets. Furthermore, EnCPdock offers more than just an AI-predicted $\Delta G_{binding}$; it also provides essential information such as surface and electrostatic complementarities (Sc, EC), surface area estimates, and other high-level structural descriptors used as input feature vectors. In addition, EnCPdock delivers a binary PPI complex mapping in the complementarity plot (CP) (https://en.wikipedia.org/wiki/Complementarity_plot) (Basu et al., 2012) and generates interactive molecular graphics of the atomic contact network at the interface, along with a contact map for further analysis.

This comprehensive platform facilitates the direct visualization and analysis of specific native interactions (contacts) contributing to binding, offering insights into their stability or transience across a library of mutants. EnCPdock further furnishes individual feature trends and relative probability estimates (Pr_{fmax}) of the obtained feature scores, providing a valuable tool for targeted protein interface design and aiding researchers in identifying structural defects, irregularities, and sub-optimality for subsequent redesign. Combining its wide array of features and applications, EnCPdock stands out as a unique online tool that will undoubtedly benefit structural biologists and researchers across related disciplines. Its capabilities offer valuable support in studying protein interactions and facilitating the design of dockable peptides, making it an invaluable resource for the scientific community.

2. MATERIALS

EnCPdock was developed using several external programs for various tasks. The "sc" program, a part of the CCP4 package (Winn et al., 2011), was used to quantify the shape complementarity at protein–protein interfaces—measured by directly implementing the original shape correlation statistic (Sc) formulated by Lawrence and Colman (Lawrence and Colman, 1993). Sc was designed based on the cumulative alignment of the nearest neighboring dot surface points (unit normal vectors) of the interacting molecular (Connolly) surfaces (Connolly, 1983) at protein–protein interfaces (binary complexes). On the contrary, the EC function measures the complementarity of surface electrostatic potential at the protein–protein interacting surfaces, arising from the distribution of atomic partial charges across the whole molecular complex. For this purpose, the same molecular (Connolly) surfaces were constructed using EDTSurf (Xu and Zhang, 2009) (at 20 dots/Å^2), and the surface electrostatic potentials on these dot surface points were computed by iteratively solving the PB equation by the finite difference method of DelPhi (Li et al., 2012) implementing its smoothed Gaussian dielectric function (Li et al., 2013). EC was then computed as the negative correlation of appropriately chosen troughs of surface electrostatic potential values—as detailed in its original and adapted formulations (Basu et al., 2012; McCoy et al., 1997). The "sc" method, in spite of being somewhat more computationally expensive, was chosen over and above other shape descriptors in docked protein complexes (e.g., region-based 3D Zernike descriptors [Venkatraman et al., 2009]) because of its rigorous use of dot surface points, numerical range [−1, 1] that aligns with EC and for being traditionally used in the CPs (Basu et al., 2012; Basu, 2017; Biswas et al., 2023).

To map a PPI complex based on its $\{Sc, EC\}$ values (treated as an ordered pair), EnCPdock used the docking scoring version [CP_{dock} (Basu, 2017)] of the two-dimensional CP (Basu et al., 2012). The CP serves as a visual aid to validate the structural accuracy of atomic models, applicable to both folded globular proteins

(Basu et al., 2014, 2012) and protein–protein interfaces (Basu, 2017; Basu et al., 2021; Biswas et al., 2023). The CP_{dock} version of the plot represents Sc and EC of the protein–protein complex attained at their interface on the X-axis and Y-axis, respectively. For training, EnCPdock implemented a support vector regression machine with a radial basis function kernel, distributed as SVM^{light} (Joachims, 2002). The binding free energy ($\Delta G_{binding}$) of the PPI complexes in the training and test datasets was determined using the standalone version (v.4) of FoldX (http://foldxsuite.crg.eu/) (Schymkowitz et al., 2005), which follows a "fragment-based strategy" using fragment libraries similar to the "fragment assembly simulated annealing" technique for protein structure prediction (Kandathil et al., 2018; Simons et al., 1997). Atoms that underwent a net change (nonzero) in SASA upon binding were identified as atoms at the protein–protein interface, wherein the ΔASA was calculated by NACCESS (Hubbard and Thornton, 1993) with a probe size of 1.4 Å, representing the hydrodynamic radius of water.

3. METHODS

3.1. Input Feature Vectors

EnCPdock's training involves the use of 13 input feature vectors, which serve as high-level, fine-grained structural descriptors for the overall protein complex or the protein–protein interface. These 13 structural features used in building EnCPdock can be broadly categorized into four groups. The first group comprises complementarity descriptors (Sc, EC), followed by accessibility descriptors (nBSA, nBSAp, nBSAnp, fracI) in the second group. The third group encompasses interfacial contact network descriptors (Ld, ACI, $slope_{dd}$, $Yinter_{dd}$, CCp_{dd}), whereas the fourth group consists of size descriptors (logN, log_{asp}). The SVM predictor is trained on the combination of all different (13) features, and, any correction or compensation for a disadvantage because of one or more features is done inherently by rest of the features. The definitions of each of these features are as follows.

3.1.1. Complementarity descriptors. First and foremost, the parameter Sc, known as shape (or surface) complementarity (Lawrence and Colman, 1993), assesses the extent of topographical correlation (or, conjointness) between the molecular surfaces of two proteins at their interface. The interface is defined as the region where both proteins interact, remaining shielded from the solvent. When the Sc value is 1, it signifies that the molecular surfaces mesh precisely, indicating a strong (perfect) correlation. Conversely, an Sc value of 0 indicates that the surfaces of interest are not at all topographically correlated, whereas a negative Sc is indicative of anticorrelation, often resulting from short contacts. Sc is calculated using the following formula:

$$S(a,b) = n_a.n_b e^{-w.d^2_{ab}}; Sc = median\{S\}$$

where n_a and n_b are two unit normal normal vectors (outwardly and inwardly oriented, respectively) corresponding to the two nearest neighboring dot surface points taken orderly (target → neighbor), located at dot surface points a and b on the two interacting surfaces, whereas d_{ab} is the distance between the two points. The specified parameter (*S[a,b]*) is computed for every nearest neighboring ordered pair of points (*a,b*) located on the interacting surfaces contributed by the two partner proteins, whereas w is a scaling constant, traditionally set to 0.5 (Lawrence and Colman, 1993). The ultimate shape complementarity value (for a specified target → neighbor pair of interacting surfaces) is determined as the median of this distribution for its left-skewness. The semi-empirical correlation statistic (Sc) is so designed that the effect of the short range van der Waals forces is precisely captured in a threshold-dependent manner ($S[a,b] \sim 0$ at $d_{ab} \sim 3.5$ Å^2 even for perfectly aligned unit normal vectors: $n_a.n_b$=1), by and large, accounting only for the relative alignment of the unit normal vectors originating from the proximal nearest neighboring points (Banerjee et al., 2003; Lawrence and Colman, 1993).

Complementarity at macro-molecular (e.g., protein) interfaces duels well beyond the local shape effects (captured by Sc) wherein the nonlocal complementarity is electrostatic in nature. To that end, EC is yet another crucial feature that measures the extent of anticorrrelation in surface electrostatic potentials at the two interacting surfaces arising from the distribution of atomic partial charges across the whole molecular complex (McCoy et al., 1997). A positive value of EC (trending to +1) indicates a good match (anticorrelation) in surface electrostatic potentials between the two interacting surfaces, indicating strong complementarity. Conversely, a negative value of EC (trending to −1) suggests a similar (and hence not complementary) surface electrostatic potentials of the two partners at their interface. Mathematically, EC is taken as the negative of the Pearson's correlation of corresponding surface electrostatic potentials (Biswas et al., 2023; McCoy et al., 1997), represented by the following expression:

$$EC = -\frac{\sum_{i=1}^{N}\left(\varphi(i)-\overline{\varphi}\right)\left(\varphi'\ (i)-\overline{\varphi}'\ \right)}{\sqrt{\left(\sum_{i=1}^{N}\left(\varphi(i)-\overline{\varphi}\right)^2\sum_{i=1}^{N}\left(\varphi'\ (i)-\overline{\varphi'}\right)^2\right)}}$$

where $\varphi(i)$ is the potential realized on the ith interfacial dot surface point because of its own atoms and $\varphi'(i)$ because of the atoms of the partner (protein) molecule, $\overline{\varphi}$ and $\overline{\varphi}'$ are the mean potentials of $\varphi(i)$ i= 1, 2, . . ., N, and $\varphi'(i)$ i= 1, 2, . . ., N, respectively (for a given interfacial surface consisting of a total of N dot surface points).

Both Sc and EC are thus correlation functions having similar trends, and identical ranges [−1, 1], and, are not necessarily reciprocative in nature. Hence, in both cases, the correlations are computed twice (once by taking each interacting surface as the target and its partner as the neighbor) followed by taking their arithmetic mean (Lawrence and Colman, 1993; McCoy et al., 1997).

3.1.2. Accessibility descriptors. Upon the formation of a protein complex, certain regions of the molecular surfaces from both partner proteins become less accessible to the solvent because of their interaction. To quantify this change in SASA, we employ the concept of normalized buried surface area (*nBSA*) using the following formula:

$$nBSA = \frac{\sum_{i=1}^{A}\Delta ASA(i)+\sum_{i=1}^{B}\Delta ASA(i)}{\sum_{i=1}^{A+B}\Delta ASA(i)}$$

Here, $\sum_{i=1}^{A}\Delta ASA(i)$ and $\sum_{i=1}^{B}\Delta ASA(i)$ represent the net change in SASA for all the atoms of protein A and B, respectively, before and after the complex formation. The resulting value is then normalized by the total change in SASA that occurs when partners A and B form the complex. During the computation of *nBSA*, the alteration in SASA [$\Delta ASA(i)$] is taken into account for all types of atoms. However, to determine the distinct contributions from polar and nonpolar atoms, we separately calculate the changes in SASA for each group, which are termed nBSA_p and nBSA_{np}, respectively.

Another significant accessibility-based feature in the context of PPIs is fracI. This feature can be defined as the ratio of number of interfacial residues to the total number of residues contributed by both interacting protein partners (PPI) (as follows).

$$fracI = \frac{N_{intres}}{N_{tot}}$$

Here, N_{intres} represents the number of interfacial residues, whereas N_{tot} corresponds to the total number of residues within the protein–protein complex.

3.1.3. Interfacial contact network descriptors. In addition to the complementarity and accessibility-based features, there are also network-based features. When a receptor and a ligand form a complex, it results in a contact interaction between the interfacial residues contributed by both partners at their interface. The criteria for residues being in contact with each other involve having any non-hydrogen atom from one residue within a distance of 4 Å from that of another residue. Residues are connected with a link if they are contributed by the two interacting partners and are in contact with each other. By observing the contact map, one can visualize the interactions between residues from the two interacting PPI. A crucial factor concerning the contact network of a PPI complex is the link density (*Ld*). This parameter is defined as the ratio of the actual number of contacts between the receptor and the ligand to the theoretical maximum number of contacts that can occur between them. To elaborate, if the receptor and the ligand have N_1 and N_2 interfacial residues in physical contact with one or more residues from their partner molecules, and N_{icnt} is the count of interchain inter-residue contacts formed at the receptor–ligand interface, then *Ld* can be expressed as follows:

$$Ld = \frac{N_{icnt}}{N_1 \times N_2}$$

It is to be noted that by definition, the PPI contact networks would be bi-partite. Although *Ld* provides insight into the overall number of contacts formed among the maximum possible contacts, the average contact intensity (ACI) (defined by the following expression) assesses the average strength of contact interactions for each inter-residue interchain link formed at the interface. To calculate the average strength, the number of interatomic (inter-residue, interchain) contacts *atcon(i)* formed by each (*i*th) link was summed up and divided by the total number of interfacial links (N_{icnt}) formed in the bi-partite PPI network.

$$ACI = \frac{\sum_{i}^{N_{icnt}} atcon(i)}{N_{icnt}}$$

It is worth noting that when the receptor and the ligand have N_1 and N_2 interfacial residues in physical contact with one or more residues from their partner molecules, the resulting adjacency matrix will consist of $N_1 \times N_2$ elements. If two residues, i from the receptor and j from the ligand, actually form a contact, the (i, j)-th element of the adjacency matrix will be 1; otherwise, it will be 0. In brief, the adjacency matrix provides a visual representation of which residues from the receptor form contact with specific residues from the ligand. If a particular residue (or node) from the receptor establishes contact with n residues from the ligand individually, then the degree (or connectivity) of that specific residue or node from the receptor will be n. When we plot the frequencies converted to a log scale (Y-axis) against the degrees converted to a log scale (X-axis), a degree distribution graph for a PPI complex is obtained. According to the power law, this degree distribution profile can be represented by a straight line, characterized by a specific slope ($slope_{dd}$) and a specific intercept ($Yinter_{dd}$). These two parameters, $slope_{dd}$ and $Yinter_{dd}$, hold significance in interpreting the patterns in the degree distribution graph for PPI complexes. To obtain the expected y-value (Y_{exp}; logarithmic frequency value) for each observed x-value (logarithm of degree for each node), we use the relation provided below.

$$Y_{exp} = slope_{dd} X_{obs} + Yinter_{dd}$$

The corresponding observed ordinate (Y_{obs}) is also obtained. The Pearson's correlation coefficient between Y_{exp} and Y_{obs}, termed as CCp_{dd}, serves as an interesting feature in understanding the degree distribution profile of protein contact networks. The CCp_{dd} can be calculated using the following formula,

$$CCp_{dd} = \frac{cov(Y_{exp}, Y_{obs})}{\sigma(Y_{exp})\sigma(Y_{obs})}$$

Here, $cov(Y_{exp}, Y_{obs})$ represents the covariance between parameters Y_{exp} and Y_{obs}, whereas $\sigma(Y_{exp})$ and $\sigma(Y_{obs})$ are variance of the expected and observed ordinates.

3.1.4. Size descriptors. Two distinct features were constructed based on the chain lengths of the interacting protein partners. The molecule with the longer chain is referred to as the receptor, whereas the one with the shorter chain is designated as the ligand. If the chain length of the receptor is represented by lenR and that of the ligand molecule by lenL, the logarithm of their combined length is denoted as logN, and the logarithm of their length ratio is termed as $\log_{asp}$ (analogous to an aspect ratio). These two parameters, calculated using the following formulas, together formulas, together takes into account the relative and absolute sizes (in terms of chain length—which could in turn be interpreted as molecular mass) of the molecular complex.

$$logN = log_{10}(lenR + lenL)$$

$$log_{asp} = log_{10}\left(\frac{lenR}{lenL}\right)$$

3.2. Training and Performance

EnCPdock was trained on a dataset comprising 3200 binary PPI complexes with high resolution (better than or equal to 2 Å) crystal structures, curated from the RCSB PDB (Berman et al., 2000) database. The details of the curation can be found in the original EnCPdock article (Biswas et al., 2023). The free energy of binding ($\Delta G_{binding}$) for each complex was calculated using the standalone version (v.4) of FoldX (Blanco et al., 2018). Importantly, the $\Delta G_{binding}$ value was normalized by the number of interfacial residues in each case. Input feature vectors were computed for these PPI complexes using a combination of external (as detailed above) and in-built programs written in FORTRAN90, PERL (v5.26.1), PYTHON3.6 along with BASH scripts—as per requirement. With the FoldX-derived and normalized $\Delta G_{binding}$ value (taken as the target function) and the 13 input feature vectors, EnCPdock was trained using the radial basis function kernel of the SVM^{light} module, with 10-fold cross validation to predict normalized $\Delta G_{binding}$ for each PPI complex in the database. A maximum correlation of $r = 0.745$ was obtained between the predicted $\Delta G_{binding}$ and FoldX derived $\Delta G_{binding}$ during cross validation, and the 90 models[1] that exhibited correlation of r = 0.745, were further utilized in the independent validation.

The independent validation was performed on three independent datasets with experimentally determined binding free energy: an affinity benchmark dataset (Kastritis et al., 2011; Vreven et al., 2015) containing 106 PPI complexes and another dataset combining PROXiMATE (Jemimah et al., 2017) and SKEMPI (Jankauskaite et al., 2019) consisting of 236 PPI complexes. The binding free energy for each complex in these datasets was predicted using the 90 models, and the median of the 90 predicted values was taken as the predicted binding free energy from EnCPdock. The correlation coefficients for the predicted binding free energy and the experimental binding free energy (obtained from the database) were found to be 0.48 and 0.63 for the affinity benchmark and "SKEMPI + PROXiMATE-merged" datasets, respectively. These correlation coefficients in predicting $\Delta G_{binding}$ are comparable to other *state-of-the-art* methods for predicting binding free energy-as detailed in the original EnCPdock article (Biswas et al., 2023). In addition to the binding free energy, EnCPdock provides other valuable information about the binding and interface properties of the PPI complex—to be discussed in subsequent sections.

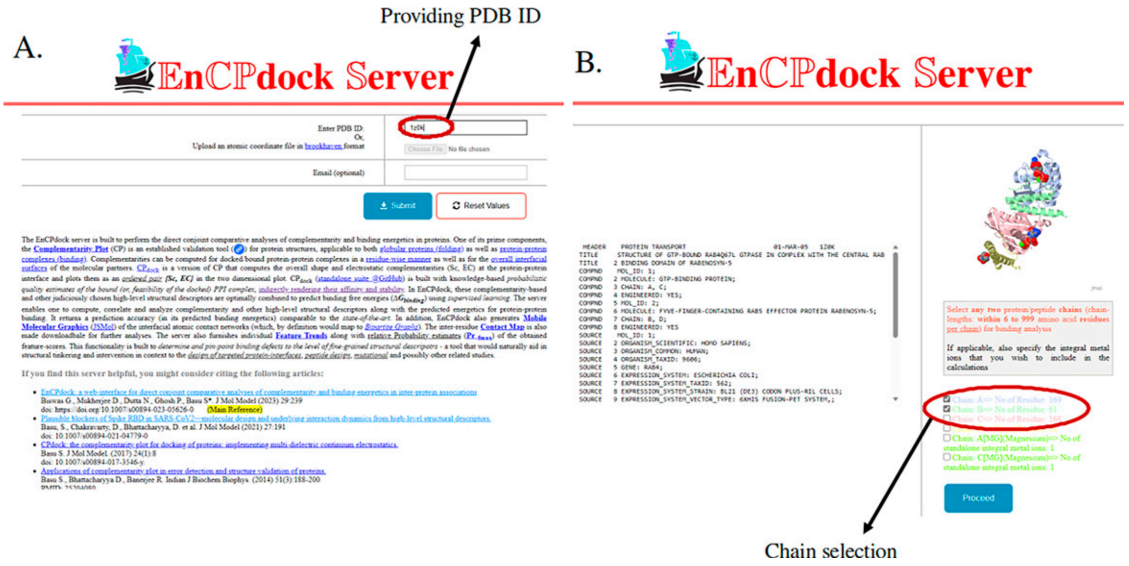
3.3. Output Features

The EnCPdock server serves as a user-friendly web interface, available at https://scinetmol.in/EnCPdock/, designed for the comprehensive analysis of complementarity and binding energetics in protein associations. To utilize the EnCPdock server, one must first obtain a PDB structure of a binary PPI complex. Users have the option to fetch experimental structures (of protein–protein/peptide complexes) directly from the RCSB PDB database (Fig. 1A) by providing their 4-letter PDB Ids or alternatively upload their own coordinate file(s) written in the brookhaven (PDB) format (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/manual/manappb.html). As the provided PPI complex might consist of multiple chains, it becomes essential to carefully select the chain names as they are identified in the provided PDB file (Fig. 1B). We recommend users view the structure in any molecular viewer to accurately identify the correct chains and provide this information to the EnCPdock server at the next step. Upon providing all the necessary information, the EnCPdock server will process the data, which may take some time (~ 1–2 min on average). Subsequently, the server will furnish its analysis of the given PPI complex in four distinct panels. In the forthcoming section, we will delve into each of the four panels separately, illustrated with an example PPI complex.

3.3.1. Scores and plots. The default first panel in the EnCPdock server provides essential insights into the complementarity and binding energetics of the given PPI complex. On the right-hand side of this panel, users can find the Sc and EC values hit by the specified complex (Fig., 2, item 1) along with the scores obtained for the other input feature vectors. In addition, the mapping (location) of the query structure on the CP_{dock} is presented (Fig. 2, item 2). CP_{dock} (Basu, 2017) is constructed based on a probabilistic representation of preferred amino acid side-chain orientations, delineated into three regions on the plot: "probable," "less probable," and "improbable." These regions are color-coded with "purple," "mauve," and "sky-blue," respectively (Basu et al., 2014). For the provided example PPI complex (PDB ID: 1Z0K), an Sc value of 0.699 is indicated, which is considered very good, whereas the EC value of 0.119 is characterized as moderate since both Sc and EC range from −1 to +1. Notably, the location of the given PPI complex on CP_{dock} is depicted as a black point, positioned in the "probable" region of the plot. This placement suggests a favorable structural configuration in terms of complementarity (local and non-local—combined) for the provided complex.

Furthermore, on the left-hand side of this default first panel, users can find the values of the 13 input feature vectors tabulated (Fig. 2, item 3), encompassing various categories such as complementarity-based, surface area-based, degree distribution profile-based, and length-based features. For instance, in the given example (PDB ID: 1Z0K), the attained *nBSA* value is 0.166, indicating approximately 16.6% decrease in solvent-accessible surface area upon binding (complexation). Likewise, the $nBSA_p$ and $nBSA_{np}$ values (0.155, 0.175, respectively) imply the fraction of polar and nonpolar surface getting buried upon complexation. Furthermore, the fracI value of 0.217 signifies that around 21.7% of the total residues from both

[1] Top 9 sets hitting the same highest correlation of r=0.745 × 10 models for the ten-fold cross validation (for each set)



Enter PDB ID: 1z0k
Or,
Upload an atomic coordinate file in brookhaven format
Choose File No file chosen

Email (optional)

Submit
Reset Values

The EnCPdock server is built to perform the direct conjoint comparative analyses of complementarity and binding energetics in proteins. One of its prime components, the **Complementarity Plot** (CP) is an established validation tool for protein structures, applicable to both globular proteins (folding) as well as protein-protein complexes (binding). Complementarities can be computed for docked/bound protein-protein complexes in a residue-wise manner as well as for the overall interfacial surfaces of the molecular partners. CP_{dock} is a version of CP that computes the overall shape and electrostatic complementarities (Sc, EC) at the protein-protein interface and plots them as an *ordered pair {Sc, EC}* in the two dimensional plot. CP_{dock} (standalone suite @GitHub) is built with knowledge-based *probabilistic quality estimates of the bound (or, feasibility of the docked) PPI complex*, indirectly rendering their affinity and stability. In EnCPdock, these complementarity-based and other judiciously chosen high-level structural descriptors are optimally combined to predict binding free energies ($\Delta G_{binding}$) using *supervised learning*. The server enables one to compute, correlate and analyze complementarity and other high-level structural descriptors along with the predicted energetics for protein-protein binding. It returns a prediction accuracy (in its predicted binding energetics) comparable to the *state-of-the-art*. In addition, EnCPdock also generates **Mobile Molecular Graphics** (JSMol) of the interfacial atomic contact networks (which, by definition would map to *Bipartite Graphs*). The inter-residue **Contact Map** is also made downloadbale for further analyses. The server also furnishes individual **Feature Trends** along with relative Probability estimates (**Pr.** $_{fmax}$) of the obtained feature-scores. This functionality is built to *determine and pin-point binding defects to the level of fine-grained structural descripotrs* - a tool that would naturally aid in structural tinkering and intervention in context to the *design of targeted protein-interfaces, peptide design, mutational* and possibly other related studies.

If you find this server helpful, you might consider citing the following articles:

- EnCPdock: a web-interface for direct conjoint comparative analyses of complementarity and binding energetics in inter-protein associations
Biswas G., Mukherjee D., Dutta N., Ghosh P., Basu S*. J Mol Model (2023) 29:239
doi: https://doi.org/10.1007/s00894-023-05626-0 (Main Reference)
- Plausible blockers of Spike RBD in SARS-CoV2—molecular design and underlying interaction dynamics from high-level structural descriptors.
Basu, S., Chakravarty, D., Bhattacharyya, D. et al. J Mol Model (2021) 27:191
doi: 10.1007/s00894-021-04779-0
- CPdock: the complementarity plot for docking of proteins: implementing multi-dielectric continuum electrostatics.
Basu S. J Mol Model. (2017) 24(1):8
doi: 10.1007/s00894-017-3546-y.
- Applications of complementarity plot in error detection and structure validation of proteins.
Basu S., Bhattacharyya D., Banerjee R. Indian J Biochem Biophys. (2014) 51(3):188-200
PMID: 25204080

```
HEADER    PROTEIN TRANSPORT                       01-MAR-05   1Z0K
TITLE     STRUCTURE OF GTP-BOUND RAB4Q67L GTPASE IN COMPLEX WITH THE CENTRAL RAB
TITLE    2 BINDING DOMAIN OF RABENOSYN-5
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: GTP-BINDING PROTEIN;
COMPND   3 CHAIN: A, C;
COMPND   4 ENGINEERED: YES;
COMPND   5 MOL_ID: 2;
COMPND   6 MOLECULE: FYVE-FINGER-CONTAINING RAB5 EFFECTOR PROTEIN RABENOSYN-5;
COMPND   7 CHAIN: B, D;
COMPND   8 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606;
SOURCE   5 GENE: RAB4;
SOURCE   6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE   8 EXPRESSION_SYSTEM_STRAIN: BL21 (DE3) CODON PLUS-RIL CELLS;
SOURCE   9 EXPRESSION_SYSTEM_VECTOR_TYPE: 6XHIS FUSION-PET SYSTEM,;
```

Select **any two** protein/peptide **chains** (chain-lengths: **within 6 to 999** amino acid **residues** per chain) for binding analysis

If applicable, also specify the integral metal ions that you wish to include in the calculations

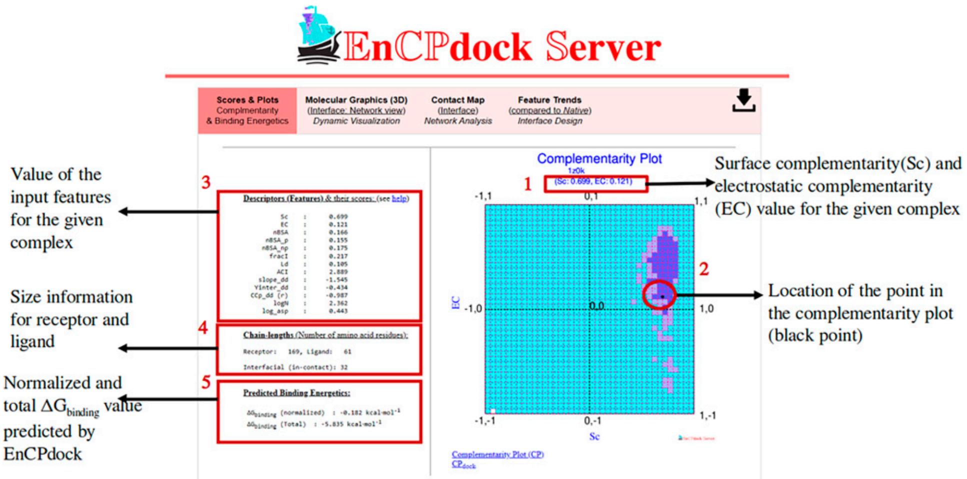
☑ Chain: A=> No of Residue: 169
☑ Chain: B=> No of Residue: 61
☐ Chain: C=> No of Residue: 168
☐
☐ Chain: A[MG](Magnesium)=> No of standalone integral metal ions: 1
☐ Chain: C[MG](Magnesium)=> No of standalone integral metal ions: 1

Proceed

FIG. 1. Workflow for utilizing the EnCPdock server. **(a)** Fetching or uploading input coordinate file**(s)** for the desired protein complex. **(b)** Selection of a pair of chains (binary interaction) from the complex.

chain A and chain B are interfacial residues. Among the network parameters, the *Ld* value of 0.105 indicates that 10.5% of all possible contacts between the interfacial residues of the two interacting chains have actually formed in the query PPI complex. In addition, the ACI value of 2.889 reveals that, on average, ∼3 interatomic contacts had formed for each inter-residue interchain link in the given PPI complex.

Moreover, from the degree distribution profile of the PPI complex, the obtained values for slope_{dd} and Yinter_{dd} are −1.545 and −0.545 for the given PPI complex, respectively. Together, the slope and the Y-intercept of the degree distribution profile reflect the extent of scale freeness (Choromański et al., 2013) of networks (i.e., preferential attachments of new nodes to already existing high-degree nodes in a network) suggestive of hub-like nodes (attractant) in the network. Specifically, a negative slope_{dd} of around this magnitude (∼ 1.6) signatures for approaching scale freeness in power-law (degree) distributions (i.e., $\gamma \sim$ 2–3 in $Y = k.X^{-\gamma}$). This may be practically relevant for targeted interface design. Using these slope_{dd} and


EnCPdock Server
Scores & Plots
Complmentarity
& Binding Energetics
Molecular Graphics (3D)
(Interface: Network view)
Dynamic Visualization
Contact Map
(Interface)
Network Analysis
Feature Trends
(compared to Native)
Interface Design
Complementarity Plot
1z0k
(Sc: 0.699, EC: 0.121)
1
2
3
4
5
-1,1
0,1
1,1
-1,0
0,0
1,0
-1,-1
0,-1
1,-1
EC
Sc
Complementarity Plot (CP)
CPdock
Value of the input features for the given complex
Size information for receptor and ligand
Normalized and total ΔG binding value predicted by EnCPdock
Surface complementarity(Sc) and electrostatic complementarity (EC) value for the given complex
Location of the point in the complementarity plot (black point)

Descriptors (Features) & their scores: (see help)

```
       Sc   :    0.699
       EC   :    0.121
     nBSA   :    0.166
   nBSA_p   :    0.155
  nBSA_np   :    0.175
    fracI   :    0.217
       Ld   :    0.105
      ACI   :    2.889
 slope_dd   :   -1.545
 Yinter_dd  :   -0.434
 CCp_dd (r) :   -0.987
     logN   :    2.362
  log_asp   :    0.443
```

Chain-lengths (Number of amino acid residues):

```
Receptor:   169, Ligand:   61

Interfacial (in-contact): 32
```

Predicted Binding Energetics:

$\Delta G_{binding}$ (normalized) : -0.182 kcal·mol^{-1}
$\Delta G_{binding}$ (Total) : -5.835 kcal·mol^{-1}

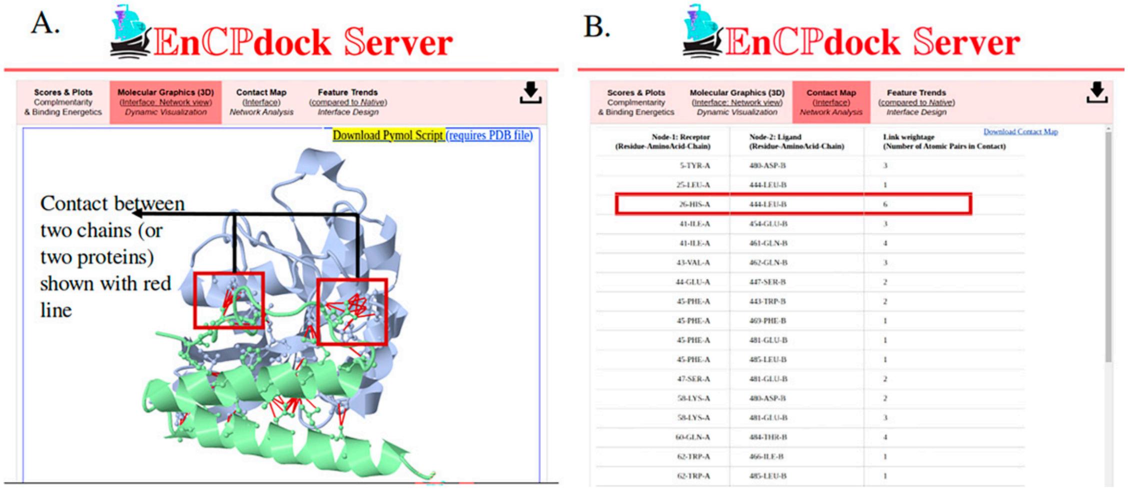
FIG. 2. Overview of EnCPdock analysis regarding complementarity and binding energetics for a PPI complex. 1. The server provides the values of Sc and EC for the given protein–protein interaction (PPI) complex, 2. The location of the query PPI complex is indicated in the complementarity plot, 3. The server presents the calculated values of all input feature vectors for the given complex, 4. Information regarding chain lengths and the number of interfacial residues is included, 5. The normalized and total binding free energy predicted by EnCPdock for the given PPI complex is also provided. EC, electrostatic complementarity; PPI, protein–protein interaction; Sc, shape complementarity.

Yinter_{dd} values, one can determine an expected ordinate (Y_{exp}) for each observed abscissa (X_{obs}) and thereby a goodness of fit (CCp_{dd}; –0.987 for the given complex) between the expected and the observed ordinates (Y_{exp}, Y_{obs}). Furthermore, the logN and $\log_{asp}$ values are found to be 2.362 and 0.443, respectively, providing additional insights into the absolute and comparative size of the binding partners in the given PPI complex. From this, the intuitive shape of the binding partners (analogous to an aspect ratio) and the complex can be guessed. In addition, the "Scores and Plots" panel also directly returns the individual chain lengths of the receptor and the ligand and the number of interfacial residues (item 4, Fig. 2). Finally, EnCPdock provides the normalized and total $\Delta G_{binding}$ values (–0.182 and –5.835 kcal.mole^{-1} for the given complex)—the latter of which can simply be obtained by multiplying the former with the total number of interfacial residues (item 5, Fig. 2).

3.3.2. Molecular graphics, contact maps, and feature trends. In the second panel dedicated to molecular graphics, one can interactively explore the interaction between the two partners using JSmol. Both chains can be displayed in different colors, with protein chains shown in a cartoon representation and interface residues represented with ball and stick models. The visualization includes red lines connecting atoms in contact with each other (Fig. 3A). The structure is easily rotatable and can be oriented for optimal viewing. To obtain high-resolution still images of specific orientations of the complex, one may right-click on the window and follow the options: File > Export > Export PNG image.

The contact map panel illustrates interactions between residues from the interacting PPI (Fig. 3B). In the marked portion of the figure, one can observe that six atoms from residue number 26 in chain A (receptor) interact with any heavy atoms from residue number 444 in chain B (ligand). Notably, residue 26 in chain A is histidine, whereas residue 444 in chain B is lysine. The right-most column indicates the total number of atomic pairs in contact with each other from the given two residues (Fig. 3B). Adjacency matrices can then be derived from the contact map for subsequent network analyses.

The fourth panel (Fig. 4) showcases the feature trends for each high-level structural descriptor utilized as input feature vectors during EnCPdock training. It presents their respective native kernel density distributions and calculates the relative probabilities of each feature score in relation to the event with the highest observed frequency for that specific feature (Pr_{fmax}). These relative probability estimates provide insights into whether an input PPI complex is categorized as regular (common) or terminal (rare) cases based on the acquired feature scores for each descriptor. This functionality is purposely designed to facilitate various tasks such as structural tinkering, intervention, targeted design of protein interfaces, mutational studies, and peptide design.



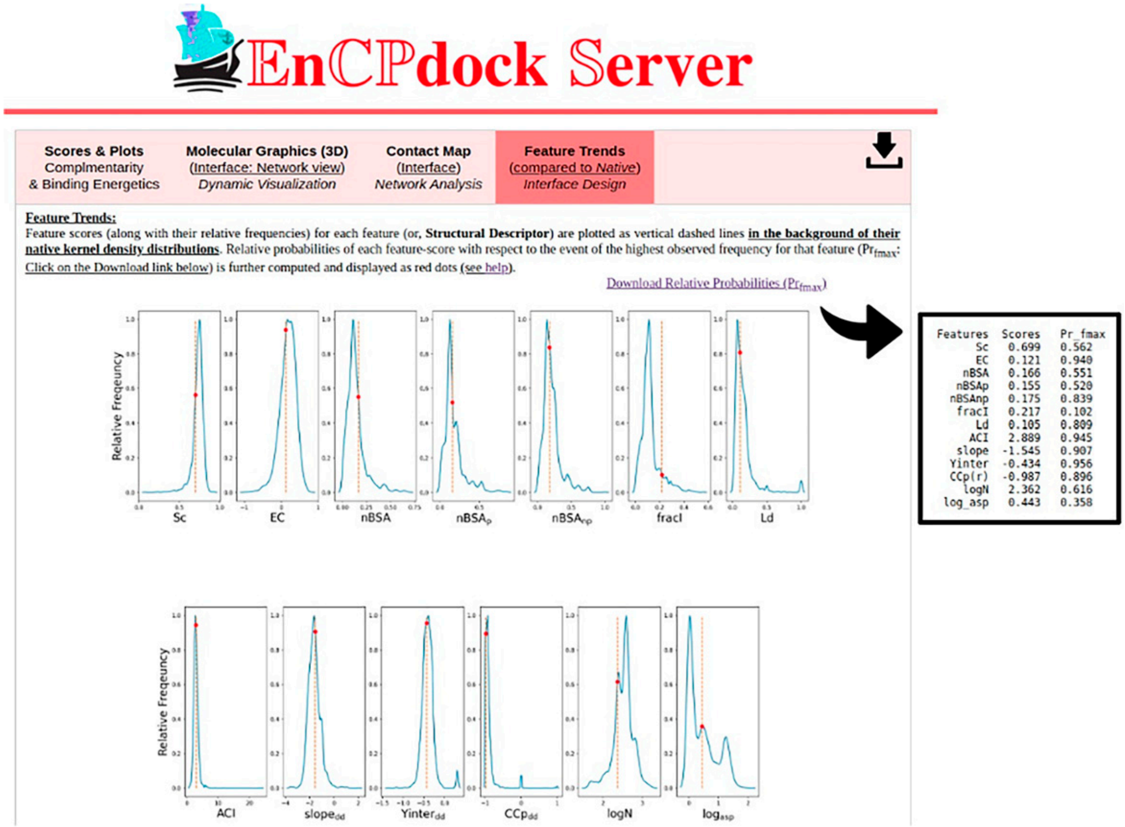
Node-1: Receptor (Residue-AminoAcid-Chain)	Node-2: Ligand (Residue-AminoAcid-Chain)	Link weightage (Number of Atomic Pairs in Contact)
5-TYR-A	480-ASP-B	3
25-LEU-A	444-LEU-B	1
26-HIS-A	444-LEU-B	6
41-ILE-A	454-GLU-B	3
41-ILE-A	461-GLN-B	4
43-VAL-A	462-GLN-B	3
44-GLU-A	447-SER-B	2
45-PHE-A	443-TRP-B	2
45-PHE-A	469-PHE-B	1
45-PHE-A	481-GLU-B	1
45-PHE-A	485-LEU-B	1
47-SER-A	481-GLU-B	2
58-LYS-A	480-ASP-B	2
58-LYS-A	481-GLU-B	3
60-GLN-A	484-THR-B	4
62-TRP-A	466-ILE-B	1
62-TRP-A	485-LEU-B	1

FIG. 3. Visualization of protein surface interaction and contact map. **(a)** Two interacting protein molecules are depicted in different colors. The interfacial residues are represented with ball and stick models, and interatomic interchain links between residues are shown in red lines. **(b)** The contact map is presented in a tabulated form. The first and second columns display the nodes (in <"residue number"–"residue type"–"chain ID"> format) coming from the first and second chains (alphanumerically sorted), respectively, that are in contact (i.e., connected by a link). The third column indicates the total number of interatomic contacts (i.e., contact intensity) for each inter-residue interchain (i.e., interfacial) link.

Moreover, a download button is conveniently placed on the top right tab bar, allowing users to easily download the entire OUTPUT folder (zipped) for local analyses and local visualization purposes.

4. CONCLUSION

In conclusion, EnCPdock was developed with the primary aim of creating an extensive web interface for conducting comprehensive comparative analyses of physicochemically relevant high-level structural descriptors, with a particular focus on complementarity, and the binding energetics of interacting PPI. With this broad objective in mind, the current version of the web server provides detailed interface properties of binary PPI complexes, encompassing complementarity and other high-level structural features. In addition, it offers predictive capabilities for the free energies of binding, including both average interfacial contribution and total values, derived from atomic coordinates. Furthermore, the web server allows users to generate mobile molecular graphics using JSmol, enabling them to explore the interfacial atomic contact network and access the contact map of the interface. Moreover, users have the opportunity to analyze trends of individual features (Sc, Ld, etc.) against their native (kernel density) distributions. This analytical capability proves to be beneficial for structural tinkering and intervention, applicable to the comparison of docked poses and the interface design of targeted complexes. For demonstrative case studies presenting specific applications of EnCPdock (for example, in probing peptide-binding specificity and mutational effects), the readers are requested to read the original EnCPdock article (Biswas et al., 2023). In summary, EnCPdock presents itself as a powerful tool in the field of structural bioinformatics, empowering researchers to conduct in-depth investigations into protein–protein/peptide interactions and their associated binding energetics. Its user-friendly interface, comprehensive analyses, and visualization options make it a



Feature Trends:
Feature scores (along with their relative frequencies) for each feature (or, **Structural Descriptor**) are plotted as vertical dashed lines **in the background of their native kernel density distributions**. Relative probabilities of each feature-score with respect to the event of the highest observed frequency for that feature (Pr$_{fmax}$: Click on the Download link below) is further computed and displayed as red dots (see help).

Download Relative Probabilities (Pr$_{fmax}$)

```
Features  Scores   Pr_fmax
      Sc   0.699   0.562
      EC   0.121   0.940
    nBSA   0.166   0.551
   nBSAp   0.155   0.520
  nBSAnp   0.175   0.839
   fracI   0.217   0.102
      Ld   0.105   0.809
     ACI   2.889   0.945
   slope  -1.545   0.907
  Yinter  -0.434   0.956
  CCp(r)  -0.987   0.896
    logN   2.362   0.616
 log_asp   0.443   0.358
```

FIG. 4. Relative probability estimates for input feature vectors in the given protein–protein interaction (PPI) complex. The red point represents the relative probability estimate for the 13 input feature vectors in the given PPI complex. The corresponding value on the X-axis is indicated by an orange line extending from the point to the X-axis. The background depicts the relative probability density distribution for the same 13 features in the training dataset.

valuable asset for advancing our understanding of protein interactions and supporting various applications, including drug discovery and protein interface engineering efforts.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Modesto Redrejo Rodríguez, Biochemistry Department, School of Medicine, Universidad Autónoma de Madrid (Spain) who first invited us to write this protocol article as a book chapter to be included in a series which was unfortunately discontinued because of lack of quality submissions.

AUTHORS' CONTRIBUTIONS

S.B. conceptualized the problem, G.B. wrote the first draft of the article, and D.B. provided technical support wherever required. S.B. and G.B. together edited the article and addressed the reviewers comments. All authors read and approved the final version of the article.

AUTHOR DISCLOSURE STATEMENT

The authors declare no conflict of interest.

FUNDING INFORMATION

The project was self-funded.

REFERENCES

Abbasi WA, Yaseen A, Hassan FU, et al. ISLAND: In-silico proteins binding affinity prediction using sequence information. BioData Min 2020;13(1):20; doi: 10.1186/s13040-020-00231-w

Banerjee R, Sen M, Bhattacharya D, et al. The jigsaw puzzle model: Search for conformational specificity in protein interiors. J Mol Biol 2003;333(1):211–226.

Basu S. CPdock: The complementarity plot for docking of proteins: Implementing multi-dielectric continuum electrostatics. J Mol Model 2017;24(1):8; doi: 10.1007/s00894-017-3546-y

Basu S, Bhattacharyya D, Banerjee R. Self-complementarity within proteins: Bridging the gap between binding and folding. Biophys J 2012;102(11):2605–2614; doi: 10.1016/j.bpj.2012.04.029

Basu S, Bhattacharyya D, Banerjee R. Applications of complementarity plot in error detection and structure validation of proteins. Indian J Biochem Biophys 2014;51(3):188–200.

Basu S, Chakravarty D, Bhattacharyya D, et al. Plausible blockers of spike RBD in SARS-CoV2—molecular design and underlying interaction dynamics from high-level structural descriptors. J Mol Model 2021;27(6):191; doi: 10.1007/s00894-021-04779-0

Berman HM, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res 2000;28(1):235–242; doi: 10.1093/nar/28.1.235

Biswas G, Mukherjee D, Dutta N, et al. EnCPdock: A web-interface for direct conjoint comparative analyses of complementarity and binding energetics in inter-protein associations. J Mol Model 2023;29(8):239; doi: 10.1007/s00894-023-05626-0

Blanco JD, Radusky L, Climente-González H, et al. FoldX accurate structural protein-DNA binding prediction using PADA1 (protein assisted DNA assembly 1). Nucleic Acids Res 2018;46(8):3852–3863; doi: 10.1093/nar/gky228

Brooks BR, Bruccoleri RE, Olafson BD, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4(2):187–217; doi: 10.1002/jcc.540040211

Brown SP, Muchmore SW. High-throughput calculation of protein-ligand binding affinities: Modification and adaptation of the mm-pbsa protocol to enterprise grid computing. J Chem Inf Model 2006;46(3):999–1005; doi: 10.1021/ci050488t

Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun 2022;13(1):1265; doi: 10.1038/s41467-022-28865-w

Chen F, Liu H, Sun H, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. Phys Chem Chem Phys 2016;18(32):22129–22139; doi: 10.1039/c6cp03670h

Choromański K, Matuszak M, Mię̧kisz J. Scale-free graph with preferential attachment and evolving internal vertex structure. J Stat Phys 2013;151(6):1175–1183; doi: 10.1007/s10955-013-0749-1

Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983;16(5):548–558; doi: 10.1107/S0021889883010985

Desantis F, Miotto M, Di Rienzo L, et al. Spatial organization of hydrophobic and charged residues affects protein thermal stability and binding affinity. Sci Rep 2022;12(1):12087; doi: 10.1038/s41598-022-16338-5

Duarte Ramos Matos G, Kyu DY, Loeffler HH, et al. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database. J Chem Eng Data 2017;62(5):1559–1569; doi: 10.1021/acs.jced.7b00104

Feng Y, Wang Q, Wang T. Drug target protein-protein interaction networks: A systematic perspective. Biomed Res Int 2017;2017:1289259; doi: 10.1155/2017/1289259

Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 2004;25(2):238–250; doi: 10.1002/jcc.10379

Hubbard SSJ, Thornton JJM. “NACCESS”, Computer Program. Dep Biochem Mol Biol Univ Coll Lond 1993.

Jankauskaite J, Jiménez-García B, Dapkunas J, et al. SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics 2019;35(3):462–469; doi: 10.1093/bioinformatics/bty635

Jemimah S, Yugandhar K, Michael Gromiha M. PROXiMATE: A database of mutant protein–protein complex thermodynamics and kinetics. Bioinformatics 2017;33(17):2787–2788; doi: 10.1093/bioinformatics/btx312

Joachims T. Learning to Classify Text Using Support Vector Machines. Springer 2002.; doi: 10.1007/978-1-4615-0907-3

Kandathil SM, Garza-Fabre M, Handl J, et al. Improved fragment-based protein structure prediction by redesign of search heuristics. Sci Rep 2018;8(1):13694; doi: 10.1038/s41598-018-31891-8

Kastritis PL, Moal IH, Hwang H, et al. A structure-based benchmark for protein–protein binding affinity. Protein Sci 2011;20(3):482–491; doi: 10.1002/pro.580

Keskin O, Gursoy A, Ma B, et al. Principles of protein–protein interactions: What are the preferred ways for proteins to interact? Chem Rev 2008;108(4):1225–1244; doi: 10.1021/cr040409x

Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234(4):946–950; doi: 10.1002/bip.360340711

Lazaridis T. Effective energy functions for protein structure prediction. Curr Opin Struct Biol 2000;10(2):139–145; doi: 10.1016/S0959-440X(00)00063-4

Lazaridis T, Karplus M. “New View” of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278(5345):1928–1931; doi: 10.1126/science.278.5345.1928

Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1999;288(3):477–487; doi: 10.1006/jmbi.1999.2685

Li L, Li C, Sarkar S, et al. DelPhi: A comprehensive suite for DelPhi software and associated resources. BMC Biophys 2012;5(1):9; doi: 10.1186/2046-1682-5-9

Li L, Li C, Zhang Z, et al. On the dielectric “Constant” of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. J Chem Theory Comput 2013;9(4):2126–2136; doi: 10.1021/ct400065j

Liu S, Zhang C, Zhou H, et al. A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. Proteins Struct Funct Bioinforma 2004;56(1):93–101; doi: 10.1002/prot.20019

Liu X, Luo Y, Li P, et al. Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. PLoS Comput Biol 2021;17(8):e1009284; doi: 10.1371/journal.pcbi.1009284

MacKerell AD, , Jr, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102(18):3586–3616; doi: 10.1021/jp973084f

McCoy AJ, Chandana Epa V, Colman PM. Electrostatic complementarity at protein/protein interfaces. J Mol Biol 1997;268(2):570–584; doi: 10.1006/jmbi.1997.0987

Ravikant DVS, Elber R. PIE-efficient filters and coarse grained potentials for unbound protein-protein docking. Proteins 2010;78(2):400–419; doi: 10.1002/prot.22550

Rodrigues CHM, Pires DEV, Ascher DB. mmCSM-PPI: Predicting the effects of multiple point mutations on protein–protein interactions. Nucleic Acids Res 2021;49(W1):W417–W424; doi: 10.1093/nar/gkab273

Romero-Molina S, Ruiz-Blanco YB, Mieres-Perez J, et al. PPI-affinity: A web tool for the prediction and optimization of protein–peptide and protein–protein binding affinity. J Proteome Res 2022;21(8):1829–1841; doi: 10.1021/acs.jproteome.2c00020

Sable R, Jois S. Surfing the protein-protein interaction surface using docking methods: Application to the design of PPI inhibitors. Molecules 2015;20(6):11569–11603; doi: 10.3390/molecules200611569

Schymkowitz J, Borg J, Stricher F, et al. The FoldX web server: An online force field. Nucleic Acids Res 2005;33(Web Server issue):W382–388; doi: 10.1093/nar/gki387

Simons KT, Kooperberg C, Huang E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268(1):209–225; doi: 10.1006/jmbi.1997.0959

Simons KT, Ruczinski I, Kooperberg C, et al. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34(1):82–95; doi: 10.1002/(sici)1097-0134(19990101)34:1<82::aid-prot7>3.0.co;2-a

Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995;5(2):229–235; doi: 10.1016/0959-440x(95)80081-6

Thomas PD, Dill KA. Statistical potentials extracted from protein structures: How accurate are they? J Mol Biol 1996; 257(2):457–469; doi: 10.1006/jmbi.1996.0175

Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. Elife 2015;4: e07454; doi: 10.7554/eLife.07454

Venkatraman V, Yang YD, Sael L, et al. Protein-protein docking using region-based 3D zernike descriptors. BMC Bioinformatics 2009;10(1):407; doi: 10.1186/1471-2105-10-407

Vreven T, Moal IH, Vangone A, et al. Updates to the integrated protein–protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. J Mol Biol 2015;427(19):3031–3041; doi: 10.1016/j.jmb.2015.07.016

Wang M, Cang Z, Wei G-W. A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation. Nat Mach Intell 2020;2(2):116–123; doi: 10.1038/s42256-020-0149-6

Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011;67(Pt 4):235–242; doi: 10.1107/S0907444910045749

Xu D, Zhang Y. Generating triangulated macromolecular surfaces by Euclidean distance transform. PLoS One 2009; 4(12):e8140; doi: 10.1371/journal.pone.0008140

Xue LC, Rodrigues JP, Kastritis PL, et al. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bioinformatics 2016;32(23):3676–3678; doi: 10.1093/bioinformatics/btw514

Zhang QC, Petrey D, Deng L, et al. Structure-based prediction of protein–protein interactions on a genome-wide scale. Nature 2012;490(7421):556–560; doi: 10.1038/nature11503

Address correspondence to:
Dr. Sankar Basu
Department of Microbiology
Asutosh College
University of Calcutta, 92
Shyama Prasad Mukherjee Rd
Bhowanipore, Kolkata 700026
India

E-mails: nemo8130@gmail.com;
sankarchandra.basu@asutoshcollege.in