Research Article

BIOMEDICAL Journal of Scientific & Technical Research

ISSN: 2574 -1241

DOI: 10.26717/BJSTR.2021.35.005636

Supplementary Data

Section 1

PDIA1 andPDIA6 were tested as potential partners of JAK 2 deploying PRISM, it yielded a very clear unambiguous response indicating there is no interaction between PDIA1 and PDIA6 with JAK2 .However, ERP44 when tested, revealed three potential interactions with JAK2 through three protein domains as shown in below table which had the following PDB Ids :2j23 (Malasseziasympodialis Thioredoxin) found in Malassezia sympodialis, a yeast involved in the pathogenesis of atopic eczema, Limacher, et al. [51], 10BB (alpha-glucosidase A, found in Thermotoga maritime a bacterium Lodge, et al. [53], and 3ZRJ (Clp V N-domain with Vip B peptide) found in a bacterium Vibrio cholerae Lenherr, et al. [50]. When these three interacting proteins were, searched using Uniprot, PDB, and SWISS prot databases, it revealed that, none of them were reported to be found in Homo sapiens. T herefore proteins that gave these interactions were not considered for further analysis. However TXND5 gave one interaction with JAK2 through CDK5 protein which is reported known to be present in in Homo sapiens Zho et al. 2002; Mapelli, et al. [30,54] (Supplementary Table 1).

Table 1.

PDB ID	Protein Name	Organism	Energy Score
2j23	Malasseziasympodialis Thioredoxin	Malasseziasympodialis	-21.11
10BB	alpha-glucosidase A	Thermotoga maritime	-3.15
3ZRJ	Clp V N-domain with Vip B peptide	Vibrio cholerae	-0.82

Section 2

RMSD (Root Mean Square Deviation) calculation was used which is a well-accepted method to measure the difference between protein structuers or complexes. Efrat et al. 2015. In order to assess the accuracy of the obtained results two measuers of RMSD had been used. The Ligand RMSD (Lrms) and the interface RMSD (lrms). The ligand RMSD calculation was was done based on the superposition of the bound and the unbound receptors and calculating the pair-wise rms distance of the relavantlignad C alpha atom. The interface RMSD is calculated over C alpha atoms of the interface residues after finding the best superposition of the bound and the unbound interfaces. For the Irms calculation, an interface residue is defined as a residue with at least one atom within 10 A° of any atom of the docking partner Andrusier, et al. [23]. In this case we have accepted the default threshold value on PRISSM as the maximal rmsd is 2 A° for Irms Tuncbag, et al. [3].

The RMSD is calculated according to the following equation Mashiach, et al. 2015.

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n} ||\upsilon_i - u_i||^2}$$

Where n is the number of atoms in the compared molecules, vi is the position of the ith atom of the first molecule, and ui is the position of the corresponding atom in the second molecule. In this work, we evaluated the results by two types of RMSD measurements:

LRMSD: The RMSD between the predicted location of the ligand and its location in the native complex. The calculation was performed on Cα atoms of the ligand after superimposing the receptor molecules in the native complex and in the predicted complex.

IRMSD: The RMSD between the interface Cα atoms in the predicted complex structure and in the native complex structure after superimposing the two interfaces. The interface includes all the residues that contain an atom within 10 Å of the other interacting protein in the structure of the native complex.

Section 3

The statistical analysis validate the accuracy of the PRISM with respect to the STRING Data base which analyses the protein-protein interactions based on annotated data on the literature. This is in a way that test the integrity of PRISM prediction which is solely based on structural properties using a tool which uses a completely different method to predict PPI. The F measure, tests how close the predictions generated by PRISM and STRING to each other statistically.

Here, we have defined TP, #FP, #FN, #TN, precision, recall, and the F-measure, which we used to evaluate the prediction results:

TP = Number of predicted PPIs that were also found in STRING (true-positive),

FP = Number of predicted PPIs that were not in STRING (false-positive),

FN = Number of PPIs not predicted by the system even though the pair was found to interact in STRING (false-negative)

TN = Number of negative predictions that were also not found in STRING (true-negative).

Precision, recall, and the F-measure are represented as follows

$$\Pr ecision = \frac{tp}{tp + fp}$$

$$\mathrm{Re}\, call = \frac{tp}{tp + fn}$$

TP = 56;

FP=186;

FN= 81;

TN=1273

Precision = 0.231

Recall = 0.409

F measure = 0.296

Since the calculated F measure 0.296 is less than the tabulated F value for infinite degrees of freedom which is 1.000 it can be concluded that the predictions that are generated through PRISM are comparable with STRING at 95% confidence interval Ohue et al. 2013.

Section 4

The intermediate proteins were selected according to the ranking based on a calculated energy score which is based on the approximation of the binding free energy function. Aggregated free energy also known as interface energy is sum of Solvation Energy, Electrostatic Energy, Internal Energy, Hydrogen and Disulphide bonds, Pi stacking of aliphatic interactions and finally van der Waals interactions Andrusier, et al. [23] The binding free energy is the change in the free energy of the system, which occurs upon complex formation.

$$\Delta G = G^C - \left(G^R + G^L\right)$$

Where Gis the free energy of the receptor-ligand complex. GR and G Lare the free energies of the un-complexed receptor and ligand, respectively. The binding score for the candidates ranking is an approximation of the binding-free energy function. Many docking methods calculate only Gc which is enough for ranking, because GR and G Lare constant for all the candidates.

The added deformation energy term approximates the energy required for deforming the unbound backbone structure of the flexible protein according to the calculated linear combination of the chosen relevant normal modes. This term is specified in below equation

$$E_{derorm} - \sum \left(V_i^{freq} \right)^2 | V_i^{amp} |,$$

Where V I freqdenotes the frequency of the ith normal mode and V ampdenotes its amplitude. The deformation energy term, Edeform, is added to the interface energy function with a weight of λ = 0.05. In this analysis we considered benchmark as 1.0 and the only protein that exceeded the benchmark and qualified as intermediate protein was CDK 5 with a fire dock score of 1.81. None of the other proteins from template data set was suitable enough to exceed the benchmark in terms of fire dock score. Therefore CDK 5 was an unambiguous choice. However few other proteins which did not exceed the benchmark but with marginal scores below the benchmark are shown in the table below along with the value for CDK 5 for clarity (Supplementary Table 2).

Table 2.

Protein	PDB ID	Fire Dock Energy Score
CDK 5	1UNH	1.81
Crystal Structure of Cdk5: p25 in complex with an	3O0S	0.89
ATP analogue		
CDK2- with diaminopyrimidine inhibitor	2FVD	0.86
Cyclin-dependent kinase 5 activator 1 protein	1JST	0.86