BIOINFORMATION
Discovery at the interface of physical and biological sciences

www.bioinformation.net
Volume 18(7)

Research Article

Received June 2, 2022; Revised July 31, 2022; Accepted July 31, 2022, Published July 31, 2022

DOI: 10.6026/97320630018604

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:
The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

Comments from readers:
Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Edited by P Kangueane
Citation: Kumar *et al.* Bioinformation 18(7): 604-612 (2022)

YAPPIS-Finder: A novel method for protein-protein interaction site predictions

Vicky Kumar[1], Ashita Sood[2], Anjana Munshi[3], Tarkeshwar Gautam[4] & Mahesh Kulharia[2,*]

[1]Centre for Computational Sciences, Central University of Punjab, Bathinda, Punjab, India; [2]Centre for Computational Biology and Bioinformatics, Central University of Himachal Pradesh, Dharamsala, Himachal Pradesh, India; [3]Department of Human Genetics and Molecular Medicine, School of Health Sciences, Central University of Punjab, Bathinda, India; [4]Department of Zoology, Kalindi College, University of Delhi, Delhi, India; *Corresponding author

Institution URL:
http://cup.edu.in/
http://www.cuhimachal.ac.in/
http://cup.edu.in/
https://www.kalindicollege.in/

Author contacts:
Vicky Kumar - E-mail: vickysibbal@gmail.com
Ashita Sood -E-mail: ashitasood01@gmail.com
Anjana Munshi - E-mail: anjanadurani@yahoo.co.in
Tarkeshwar Gautam - E-mail: tarukmc@gmail.com
Mahesh Kulharia - E-mail: kulharia@gmail.com; Tel: +91-99884 28856

Abstract:
We describe a multi parametric-approach, **YAPPIS-Finder**, for predicting the PPI sites on protein surface. A non-redundant database of comprised of 2,265 protein-protein interaction interfaces (PPIIs) involving 4,530 protein-protein interacting partners (PPIPs) and depicting the interaction between protein-chains of experimentally determined PPCs was used in designing the YAPPIS-Finder. Parametric score obtained on analyzing these 4,530 PPIPs with respect to their residue interface propensity, their hydrophobic content, and amount of solvation free energy associated with them provided the basis of YAPPIS-Finder. By applying YAPPIS-Finder on another dataset 4,290 PPIPs from 2,145 PPIIs, the optimal range of the parametric scores and protein-probe van der Waals energy of interaction was determined. Subsequently, taking the optimal range of PPIP parametric scores and threshold for protein-probe van der Waals energy of interaction into the consideration, the YAPPIS-Finder was tested on a blind dataset of 554 protein-chains and it was

found predicting 69.67% sites correctly. On predicting only one PPI site on each protein-chain, the YAPPIS-Finder found covering 22.91% of actually sites in the predicted site. Contrary to this, the sites predicted by SPPIDER covered 22.7% of actual sites. However, on predicting two PPI sites for each protein-chain, the percentage coverage of actual sites in the predicted sites by YAPPIS-Finder exceeded two-fold (i.e. 41.81%), thus making the YAPPIS-Finder a better method.

Keywords: YAPPIS-Finder, protein-protein interaction site, predictions

Background:

Proteins are the basic functional unit in cellular world of life **[1]**. They are genetically programmed to enact an array of molecular functions in response to biological events at cellular and system levels **[2, 3]**. The unquestionable roles that proteins play in executing various intra- and extra-cellular processes such as cell proliferation, differentiation, apoptosis, and signal transduction have drawn the attention of the scientific community to get their structural and functional insights. In various studies, the efficacy of proteins in the cellular environment is reported to be of short-range and inadequate to sustain life in isolation **[4].** More often, proteins interactively associate with other bio-molecules to form supramolecular assemblies responsible for molecular functioning in living organisms. Understanding the molecular phenomenon that triggers and maintains the complexity of such associations may improve the application of protein chemistry. Formation of protein-protein complexes (PPCs) are the outcome of one of such molecular association in which the interaction between two proteins is governed by formation of covalent and non-covalent associations between them **[5, 6]** . The covalent PPIs, although rare to see, are owed to sharing of electrons between the protein constituents **[7]**. While in the case of non-covalent interactions, formation of hydrogen bonds, ionic interactions, van der Waals interactions, or hydrophobic bonds, are the main contributing factors for proteins complexation **[8].** The presence of such binding factors help in associating two proteins, thereby, assists in maintenance of life. Therefore, development of an approach emphasizing on the protein dynamics in context of preference of binding partners, binding site location, functionality concomitant with the formed protein complexes, is the need of hour. Despite the availability of experimental techniques for identification of PPIs in abundance **[9] [10] [11]** recognition of binding sites for proteins like Wnt **[12,13]** (or similar hydrophobic protein) and Hedgehog **[14]** (membrane proteins), even though very much crucial, is difficult. Additionally, relatively high experimental cost and time-intensiveness nature of experimental techniques makes them less apt for application *en masse*. Therefore, development and application of computational methods, which are free from such problems, are growing rapidly. In the proposed work, we have presented an approach named **Y**et **A**nother **P**rotein **P**rotein **I**nteraction **S**ite-Finder (YAPPIS-Finder) for predicting the PPI sites on protein surface. A non-redundant database of comprised of 2,265 PPIIs (or 4,530 PPIPs) depicting the interaction sites between two protein-chains of experimentally determined PPCs was used in designing YAPPIS-Finder. Analysis of these 4,530 PPIPs was carried out to understand the PPI sites with respect to their residue interface propensity, their hydrophobic content, and amount of solvation free energy associated with them. Another dataset of 2,145 PPIIs (4,290 PPIPs) was used to train the proposed approach and subsequently the optimal range of PPIPs parametric scores was derived. The approach was tested on a blind dataset of 554 protein-chains and its performance was also compared with the SPPIDER. **[15]**

Materials and Methods:

A non-redundant database protein-protein interaction interfaces (NRDB)**[16]** depicting the actual PPI sites was analyzed with respect to residue interface propensity (RIP), hydrophobicity and solvation free energy to design the proposed computational approach for PPI sites prediction.

Non-redundant database of protein-protein interaction interfaces (NRDB) depicting actual interfaces:

A non-redundant database of PPIIs **[17]** demarcated from experimentally determined PPCs was used in proposed study. NRDB was designed considering the PDBs for which the information of structural classification of protein was available in the last manually curated SCOP version 1.75. In NRDB, the information of interacting interface was determined by considering the interatomic distance between the constituting atoms of two protein-chains in a PPC. Two atoms belonging to two different protein-chains of a PPC were said to be in contact and demarcated as an atomic contact pair (ACP) if the intervening distance between them was less than the sum of their van der Waals radii plus 1 Å as tolerance factor (Figure 1). The collection of ACPs between a pair of interacting protein-chains was referred as "**P**rotein-**P**rotein **I**nteraction **I**nterface" (PPII) and the collection of interacting atoms from individual interacting protein-chain were termed as the "**P**rotein-**P**rotein **I**nteracting **P**atch" (PPIP). Only the PPIIs with at least 20 ACPs were retained in the NRDB and a total of 2,265 PPIIs (4,530 PPIPs) from 1,931 PDB files were demarcated. These 2,265 PPIIs were representing 43,509 PPIIs and the same number of SCOP super family pairs.

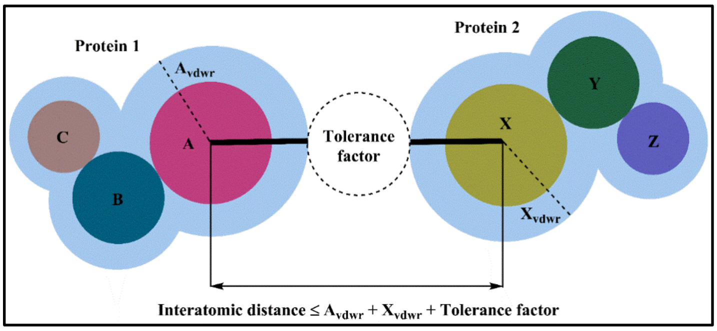

Protein 1
Protein 2
A_{vdwr}
C
A
B
Tolerance factor
X
Y
Z
X_{vdwr}
Interatomic distance ≤ A_{vdwr} + X_{vdwr} + Tolerance factor

Figure 1: Definition of atomic contact pair (ACP)

Analysis of PPIPs with respect to residue interface propensity (RIP)

All of the PPIPs from NRDB **[16]** were examined with respect to RIP and the cumulative RIP score for each PPIP was carried out by taking into the account the RIP scores of individual residues derived in our unpublished work. For each interacting atom in a PPIP, its individual RIP score was calculated by dividing the overall RIP score of the residue by number of atoms the residue, whose this interacting atoms is a part, does normally have (ignoring the atoms involved in peptide bond formation). This was followed by the summation of individual RIP score of all interacting atoms in the PPII to represent the cumulative RIP score of the PPIP (Eq. 1).

$$PPIP_{\text{Jʋ}} = \sum_{i=1}^{n} n_i * \text{Jʋ}_i \quad \textbf{Eq. 1}$$

Where n represents the total number of interacting atoms in the PPIP, n_i represents current interacting atom, and Jʋ_i represents the per-atom RIP of the residue under consideration.

Analysis of PPIPs with respect to hydrophobicity:

To determine the level of hydrophobicity (Φ) associated with PPIPs, the hydrophobicity scale for amino acids given by Hessa *et al.* 2005 was used **[18]**. For each interacting atom in the PPIP, its corresponding hydrophobicity score was obtained by dividing the overall residue hydrophobicity score by number of atoms the residue, whose this interacting atom is a part, does normally have (ignoring the atoms involved in peptide bond formation). At last, the hydrophobicity score of all interacting atoms in the PPIP was summed up in linear fashion to represent the hydrophobicity score of the PPIP (Eq. 2)

$$PPIP_{\Phi} = \sum_{i=1}^{20} n_i * \Phi_i \qquad \textbf{Eq. 2}$$

Where n_i represents the total number of atoms from amino acid i involved in interaction and Φ_i represents the per-atom hydrophobicity score of amino acid under consideration.

Analysis of PPIPs with respect to solvation free energy:

The solvation free energy of PPIP was calculated by taking into account the solvation energy scale for amino acids given by **[19]** White et al, 1996 for each interacting atom in PPIP, its corresponding solvation free energy score was calculated by dividing the overall solvation free energy of the residue by number of atoms the residue, whose this interacting atoms is a part, does normally have (ignoring the atoms involved in peptide bond formation). This was followed by the summation of solvation free energy score of all interacting atoms in the PPII to represent the solvation energy score of the PPIP (Eq. 3).

$$PPIP_{\omega} = \sum_{i=1}^{20} n_i * \omega_i \qquad \textbf{Eq. 3}$$

Where n_i represents the total number of atoms from amino acid i observed to be interacting in a PPIP and ω_i represents the per-atom solvation free energy score of residue under consideration.

Removal of outliers from the parametric scores of PPIPs and decomposition of parametric scores into sub-ranges:

The parametric scores obtained on analyzing the PPIPs from NRDB were set to provide foundation for the proposed scheme. However, the NRDB analysis revealed that the parametric scores for PPIPs contained a significant number of outliers in them. Therefore, a statistical approach of inter quartile range was adopted to remove the outliers. After outlier removal, the entire range of parametric score for each PPI sites parameter was divided into a number of bins with width calculated using Scott's rule (Eq. 4).

$$W = 3.49 * \sigma * N^{\frac{-1}{3}} \qquad \textbf{Eq. 4}$$

Where W is the width of the bin, σ is the standard deviation of the distribution of parametric scores, and N is the total number of PPIPs for which parametric scores was available.

Creation of the training and test dataset to implement the proposed approach:

Success of any prediction tool largely depends of quality of the datasets used in its designing. The dataset should be comprised of information related to both known interacting protein pairs (positive set) and non-interacting protein pairs (negative set). It is quite easy to obtain the experimental instances of PPI for the positive set while construction of negative set is not that much straightforward. In the proposed study, to develop a computational approach for PPI prediction, one training and one test set was designed considering the presence or absence of SCOP superfamily pair in NRDB proposed by our research group previously **[17]**. To design the training set, the SCOP **[20, 21]** superfamily pairs with their corresponding PPIIs in the NRDB were taken into the account. However, to remove the overlap between the NRDB and training set, for each SCOP **[20, 21]** superfamily pair covered in NRDB, a new PPII representative was selected from PPInS **[17]**. If there were more than one PPII available for a SCOP superfamily pair in PPInS, then the PPII with the largest number of ACPs was selected. If the PPII with the largest number of ACPs was the one which was already a part of NRDB **[16]** , then the PPII with second largest number of ACPs was selected for the training set. Following this strategy, a total of 2,145 PPIIs from 1,896 PDBs were selected as training set.

To design the test set, the SCOP superfamily pairs which were not covered in NRDB **[16]** were selected. For each such SCOP superfamily pair, the PPII with the largest number of ACPs was selected from PPInS as a part of the test set. In this way, no room was left for overlap between the training and test set and a total of 554 binary PPIIs from 277 PDBs were selected as test set.

Implementation of YAPPIS-Finder:

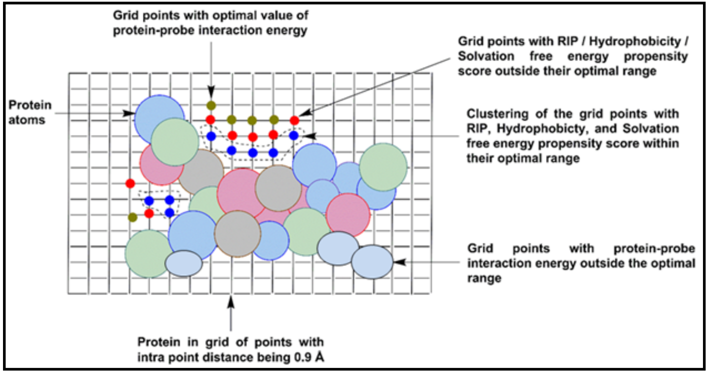

Grid points with optimal value of protein-probe interaction energy
Grid points with RIP / Hydrophobicity / Solvation free energy propensity score outside their optimal range
Protein atoms
Clustering of the grid points with RIP, Hydrophobicty, and Solvation free energy propensity score within their optimal range
Grid points with protein-probe interaction energy outside the optimal range
Protein in grid of points with intra point distance being 0.9 Å

Figure 2: YAPPIS-Finder algorithm

(i) Overview: The implementation of the proposed computational approach was completed in two phases; training and testing phase. During training phase, initially, using all of the PPIIs involved in NRDB **[16]** , the optimal range of PPI sites' parametric scores was calculated. The obtained ranges were then decomposed into multiple sub-ranges (called domains) using the method of interquartile range. This was followed by examining the surface of unbounded proteins extracted from the experimentally determined PPCs contained in training dataset. Taking two parameters at a time and their corresponding domain scores one at a time along with the threshold for protein-probe van der Waals energy of interaction, the protein surface was examined using an energy-based grid-oriented approach. The group of atoms with parametric scores within the range of current domain and protein-probe van der Waals interaction energy threshold were clustered into 15 clusters (created considering the spatial proximity of predicted interacting atoms). The obtained clusters were recognized as possible PPI sites and a predicted site score (T) representing the precision and coverage of predicted PPI sites against the actual sites was also calculated for each of them.

$$Precision = \frac{N_P}{N_A + N_P - N_C} \qquad \textbf{Eq. 5}$$

$$Coverage = \frac{N_A}{N_A + N_P - N_C} \qquad \textbf{Eq. 6}$$

$$\mathrm{T} = Precision * Coverage \qquad \textbf{Eq. 7}$$

This was followed by the identification of the parametric scores and protein-probe interaction energy conforming $\mathrm{T} \geq 0.25(25\%)$ against the actual PPI sites (demarcated on the basis ACP definition). The obtained parametric scores and threshold for protein-probe van der Waals interaction energy were designated as the optimal range of the parameters. This way, the optimal ranges of all of the parameters were calculated.

During the testing phase, the unbound forms of protein-chains extracted from the experimentally determined PPCs contained in testing dataset were examined using the similar strategy. The protein surface was examined corresponding to the optimal ranges of PPI sites parametric score and the optimal value of the van der Waals energy of interaction obtained during the training phase. For each PPI site parameter pair, the protein atoms with their parametric scores within the threshold values (i.e. optimal range) were clustered on the basis of geometric proximity and ranked using Eq. 5-7. Here too, only the predicted sites with $T \geq 0.25(25\%)$ were termed as the correctly predicted sites. Following this, the best ranked predicted sites were compared against the actual PPI sites to evaluate the prediction efficacy.

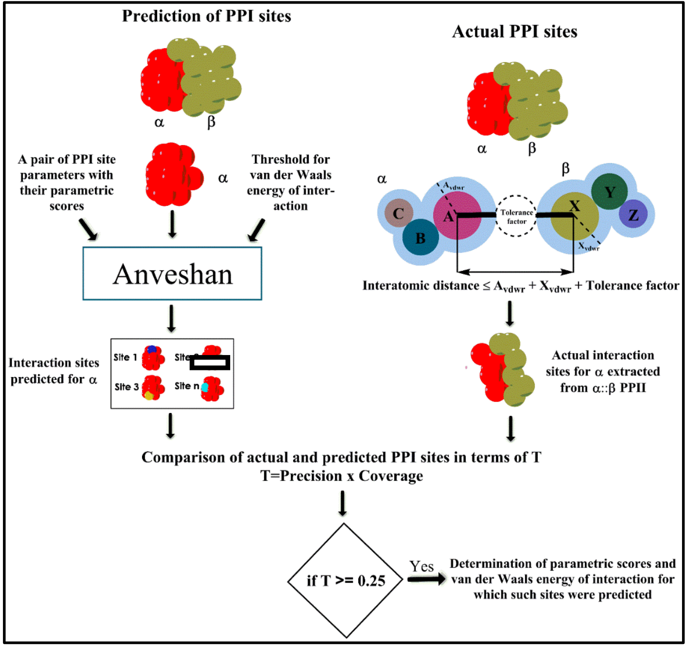

Prediction of PPI sites
Actual PPI sites
α
β
A pair of PPI site parameters with their parametric scores
Threshold for van der Waals energy of interaction
Anveshan
Interaction sites predicted for α
Site 1
Site 3
Site n
C
A
B
A_vdwr
Tolerance factor
X
Y
Z
X_vdwr
Interatomic distance ≤ A_vdwr + X_vdwr + Tolerance factor
Actual interaction sites for α extracted from α::β PPII
Comparison of actual and predicted PPI sites in terms of T
T=Precision x Coverage
if T >= 0.25
Yes
Determination of parametric scores and van der Waals energy of interaction for which such sites were predicted

Figure 3: Calculation of optimal range of PPI site parametric scores

YAPPIS-Finder algorithm:

Input: Protein-chains extracted from experimentally determined protein-protein complex

Process: The input protein-chain was placed in a three-dimensional box, which was divided into a cubic grid of resolution 0.9 Å. At each grid point, a methyl (-CH$_3$) probe was placed and van der Waals energy of interaction was calculated between the protein atom and the methyl probe. The process of calculating the protein-probe van der Waals interaction energy is described in detail in Laurie and Jackson **[22]**. The interaction energy was calculated using the GRID force field parameters as described in **[23]**. Grid points with a ''protein–probe interaction'' energy more favourable (negative) than a predetermined threshold were retained. For such grid points, three protein binding propensity scores (i.e. residue interface propensity score, solvation energy propensity score, and hydrophobicity propensity score) were calculated by considering the type of amino acid residue whose atoms are occluded from solvent exposure due to the predicted grid points. An amino acid is considered to be interacting with a grid point if at least one of its atoms is within 1.6 Å of the grid point. The overall protein binding propensity of the grid point, k, was defined as:

$$\text{Jʋ}_k = \frac{\sum_{i=1}^{i=20} n_i * \text{Jʋ}_i}{N} \qquad \textbf{Eq. 8}$$

$$\omega_k = \frac{\sum_{i=1}^{i=20} n_i * \omega_i}{N} \qquad \textbf{Eq. 9}$$

$$\Phi_k = \frac{\sum_{i=1}^{i=20} n_i * \Phi_i}{N} \qquad \textbf{Eq. 10}$$

where n_i is the number of atoms of a specific amino acid i within 1.6 Å of the grid point; N is the total number of atoms interacting with the grid point k; while Jʋ_i, ω_i and, Φ_i are the residue interface propensity, hydrophobicity propensity, and solvation free energy propensity, respectively, of the amino acid i under consideration. If Jʋ_k, ω_k, and Φ_k were falling in range of their parametric domains, then the residues (or their atoms) interacting with grid point k were demarcated as the interacting atoms. The interacting atoms were then clustered on the basis of their spatial proximity. A cluster is defined as the group of grid points wherein none of the grid points has its centre farther than 1.0 Å from the centre of the nearest grid point and demarcated as the putative PPI site.

Output: Clusters of atoms as the predicted PPI sites.

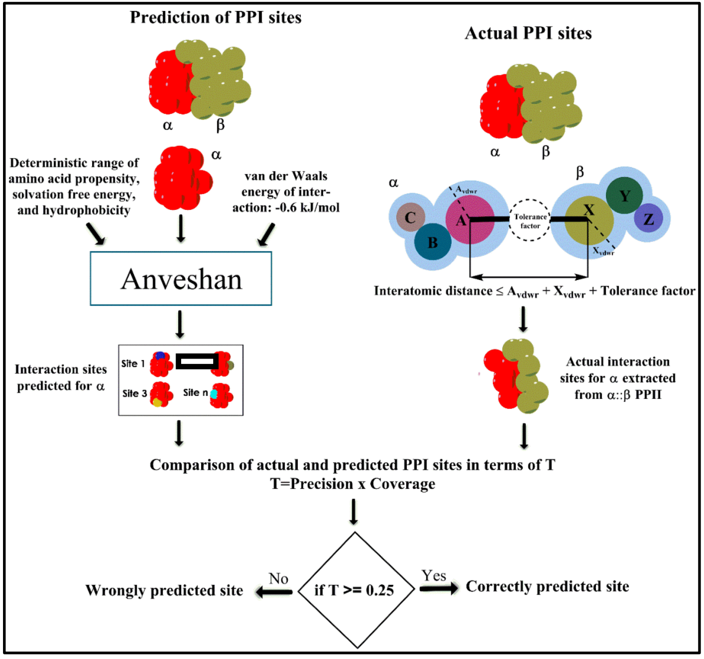

Prediction of PPI sites
Actual PPI sites
α
β
Deterministic range of amino acid propensity, solvation free energy, and hydrophobicity
van der Waals energy of interaction: -0.6 kJ/mol
Anveshan
Interaction sites predicted for α
Site 1
Site 3
Site n
C
A
B
A_vdwr
Tolerance factor
X
Y
Z
X_vdwr
Interatomic distance ≤ A_vdwr + X_vdwr + Tolerance factor
Actual interaction sites for α extracted from α::β PPII
Comparison of actual and predicted PPI sites in terms of T
T=Precision x Coverage
Wrongly predicted site
No
if T >= 0.25
Yes
Correctly predicted site

Figure 4: Prediction of predicted sites for the protein-chains from test datasets

Comparison of prediction power of SPPIDER and YAPPIS-Finder:

The prediction power of **YAPPIS-Finder** was also compared against the SPPIDER **[15]** which is one of most efficient approach (as claimed by its designer) available. The protein-chains from the test dataset were examined using SPPIDER to predict the PPI sites and prediction power of both these approaches, SPPIDER **[15]** and **YAPPIS-Finder**, was compared using Eq. 5-7.

Results and Discussion:

Optimization of protein-protein interaction sites parameters by running YAPPIS-Finder for the training dataset

To obtain the optimal range of PPI site parameters, the **YAPPIS-Finder** algorithm was applied on protein chains extracted from experimentally determined PPCs of training dataset as described in Step 4 and Figure 3.5. The predicted and actual sites were compared (demarcated using the definition of ACP) and overlap between the two was determined using Eq. 8 to 10. Subsequently, the values of parametric scores and threshold of van der Waals energy for interaction were retrieved for which, $T \geq 0.25(25\%)$.

The parametric scores obtained were termed as the optimistic range of the parameters under consideration.

Prediction of interaction sites for the test dataset proteins by YAPPIS-Finder and performance analysis was done in similar manner. Only predicted sites with $T \geq 0.25(25\%)$ were termed as the correctly predicted sites.

Table 1: Parametric scores after removal of statistical outliers

Parameter	Before outlier removal	After outlier removal
Residue interface propensity	0.15 to 495.87	0.15 to 21.56
Hydrophobicity	-5.8 to 525.05	-1.38 to 22.36
Solvation free energy	-5.67 to 277.38	-5.67 to 13.22

Prediction efficacy of YAPPIS-Finder and SPPIDER and comparison of their performance evaluation

The prediction efficacy of YAPPIS-Finder was examined by running it on a blind dataset of 554 protein-chains from the test set. For these 554 protein-chains, their actual PPI sites were demarcated using the definition of ACP described in Materials and Methods section this section as well as in **[17]**. Using YAPPIS-Finder with optimistic range of PPIP parametric scores and optimal values of protein-probe van der Waals energy of interaction, 10 PPI sites were predicted for each protein-chain. However, to minimize the false negatives, the sites with $T \geq 0.25(25\%)$ were termed as the correctly predicted sites. This way, a total of 385 sites were termed as the correctly predicted against the 554 actual sites giving us the prediction accuracy of 69.67%. When all these 554 protein-chains were given to SPPIDER server **[15]** for site prediction, total 529 sites were predicted by it. However, opposite to your approach where we have put filtering criteria of $T \geq 0.25(25\%)$ for a site to be considered as the correctly predicted site, the SPPIDER has even predicted the site with only one residue, which may be a result of false prediction as well.

Comparison between the performance evaluation of YAPPIS-Finder and SPPIDER

Following to the determination of prediction efficacy of the YAPPIS-Finder and SPPIDER revealed that SPPIDER, the sites predicted by both of these approaches were compared against the actual sites. On predicting only one PPI site for each protein-chain, the YAPPIS-Finder found covering 22.91% of actually sites in the predicted site. Contrary to this, the sites predicted by SPPIDER covered 22.7% of actual sites. However, on predicting two PPI sites for each protein-chain, the percentage coverage of actual sites in the predicted sites by YAPPIS-Finder exceeded two-fold (i.e. 41.81%), thus making the YAPPIS-Finder a superior approach.

Conclusion:

This paper describes the development of a novel, multiparameteric-method YAPPIS-Finder to identify the protein-protein interaction sites. The YAPPIS-Finder was tested on a set of above 500 protein-protein complexes. The ability of the method in identifying the protein-protein interaction sites has been investigated. The YAPPIS-Finder method included the information of solvation. Residue propensity, hydrophobicity and van der Waals interaction energy in its ability to identify near-precise region of protein-protein interactions whilst at the same time giving a higher degree of correlation in overlap between predicted and experimentally proved protein-protein interaction sites.

Acknowledgements

The authors are grateful to University Grants Commission (UGC), India, for providing financial assistance to VK in the form of UGC-NET JRF awards and major research grant (MRP) to MK. The authors are also grateful to the Central University of Himachal Pradesh, and Central University of Punjab, India, for providing the academic, administrative and infrastructural support to carry out and publish this work.

References:

[1] Shoshan-Barmatz V *et al. Mol. Aspects Med.* 2010 **227:**285. [PMID: 20346371]

[2] Sun Y & MacRae TH *Cell. Mol. Life Sci.* 2005 **62:**2460. [PMID: 16143830]

[3] Volkenshtein MV, *Molecules and life: an introduction to molecular biology*, Springer, 1970.DOI

[4] Ohno Y *et al. Comput. Phys. Commun.* 2014 **185:** 2575. https://doi.org/10.1016/j.cpc.2014.06.004

[5] Saio T *et al. J. Biomol. NMR.* 2010 **46:** 271. [PMID: 20300805]

[6] Swapna LS *et al. BMC Struct. Biol.* 2012 **12**:6. [PMID: 22554255]

[7] Ngounou Wetie AG *et al. Cell. Mol. Life Sci.* 2014 **71**:205. [PMID: 23579629]

[8] Gabius HJ *Pharm. Res.* 1998 **15:**23. [PMID: 9487542]

[9] Kulharia M *et al. J. Mol. Graph. Model.* 2009 **28:**297. [PMID: 19762259]

[10] Porollo A & Meller J, *Proteins.* 2007 **66:** 630. [PMID: 17152079]

[11] Zahiri J *et al. Curr Genomics.* 2013 **14:**397. [PMID: 24396273]

[12] Wong HC *et al. Nat. Struct. Biol.* 2000 **7:** 1178. [PMID: 11101902]

[13] Kahn M *Nat. Rev. Drug Discov.* 2014 **13:**513. [PMID: 24981364]

[14] Myers BR *et al. Dev. Cell.* 2013 **26**:346. [PMID: 23954590]

[15] Langdonc QK *et al. Mol. Biol. Evol.* 2018 **35:** 2835. [PMID: 30184140]

[16] Jolley KA *et al. Wellcome Open Res.* 2018 **3**:124 [PMID: 30345391]

[17] Kumar V *et al. Sci. Rep.* 2018 **1**:9 [PMID: 30127348]

[18] Hessa T *et al. Nature.* 2005 **433:** 377. [PMID: 15674282]

[19] Wimley WC *et al. Biochemistry.* 1996 **35:**5109. [PMID: 8611495]

[20] Andreeva A *et al. Nucleic Acids Res.* 2014 **42:**D310. [PMID: 24293656]

[21] Andreeva A *et al. Nucleic Acids Res.* 2020 **48:** D376. [PMID: 31724711]

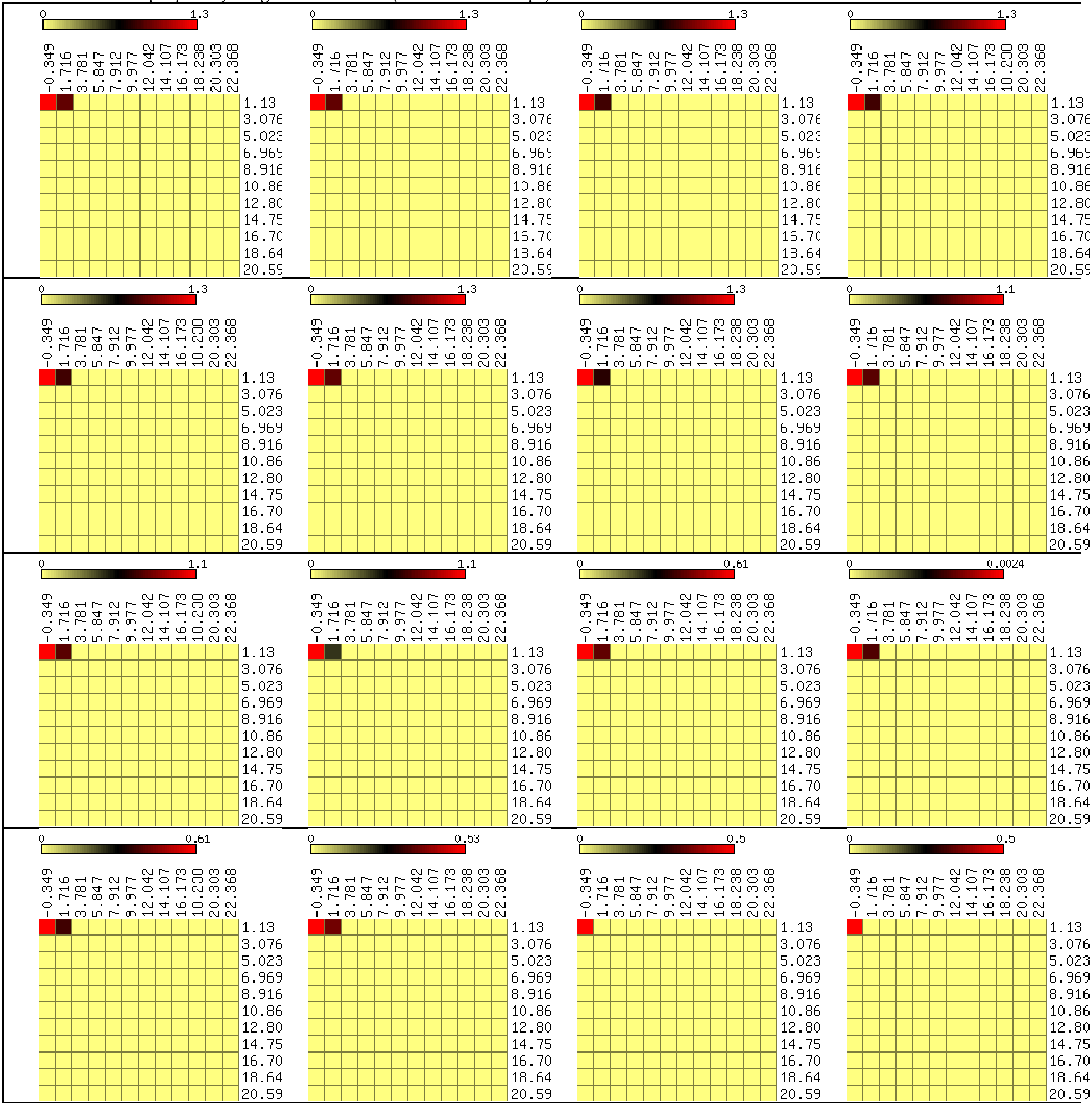
[22] Laurie ATR *et al. Bioinformatics* 2005 **21:**1908. [PMID: 15701681]

[23] Cruciani G & Watson KA *J. Med. Chem.* 1994 **37:**2589. [PMID: 8057302]

Hydrophobicity and Residue interface propensity

Hydrophobicity range: -0.349 to 22.368 (X-axis of heat maps)

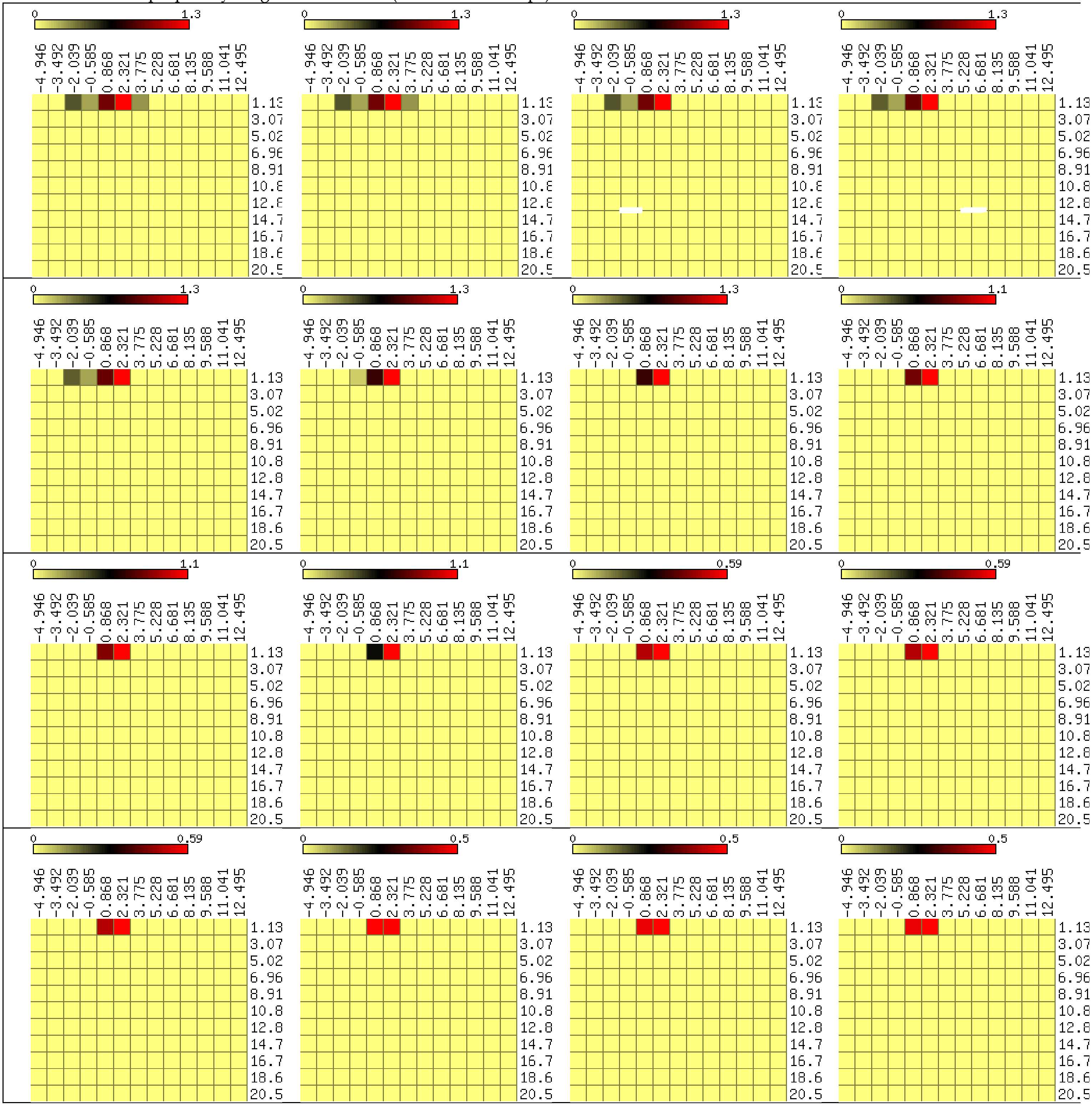
Residue interface propensity range: 1.13 to 20.596 (Y-axis of heat maps)



Solvation free energy and Residue interface propensity

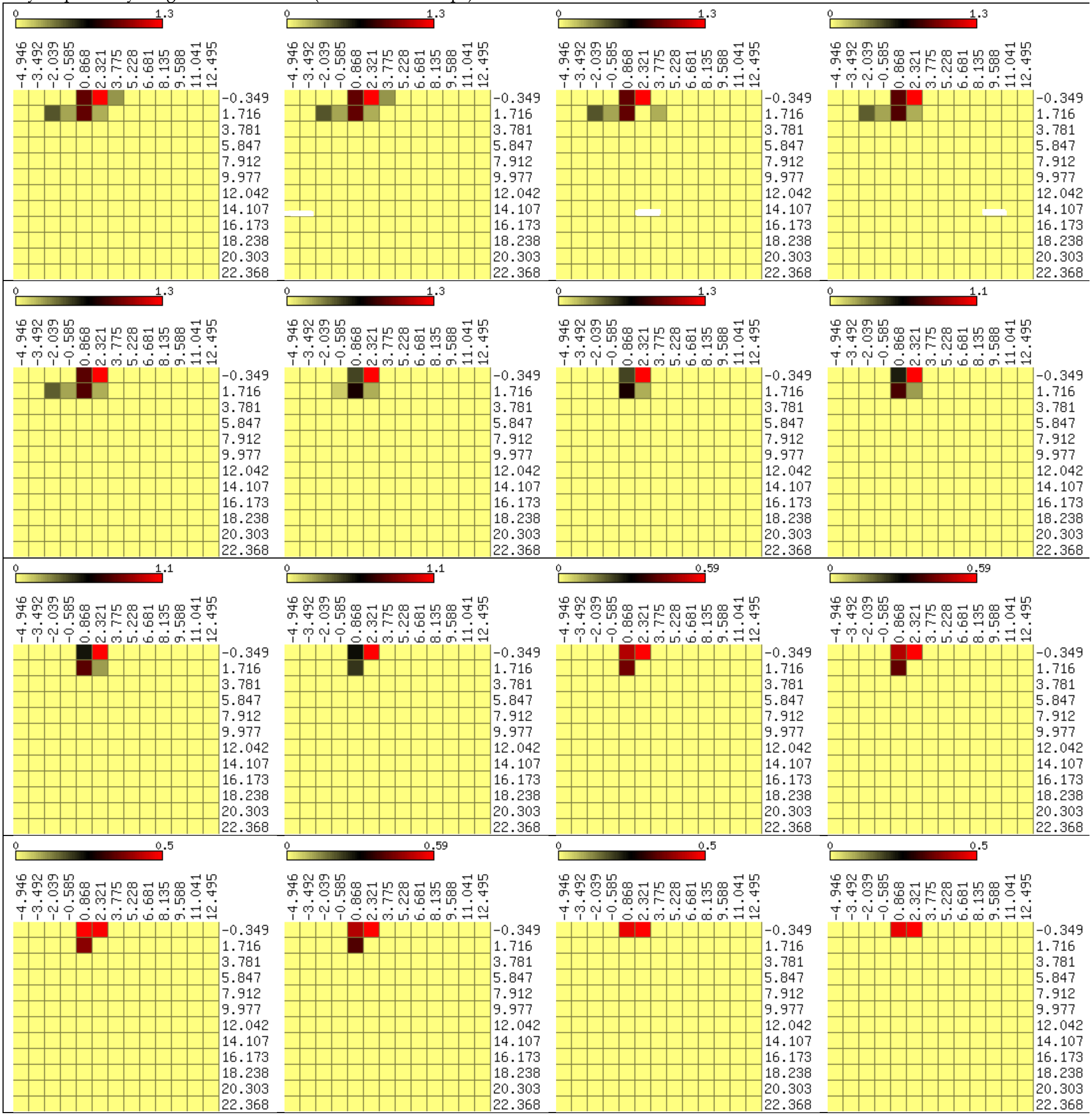
Solvation free energy range: -4.946 to 12.495 (X-axis of heat maps)

Residue interface propensity range: 1.13 to 20.596 (Y-axis of heat maps)



Solvation free energy and Hydrophobicity

Solvation free energy range: -4.946 to 12.495 (X-axis of heat maps)
Hydrophobicity range: -0.349 to 22.368 (Y-axis of heat maps)



Agro Informatics Society
BIOMEDICAL
INFORMATICS
BMIS
AGIS
Biomedical Informatics Society
since 2005
BIOINFORMATION
Discovery at the interface of physical and biological sciences
indexed in
PMC
PubMed
EBSCO
EMERGING SOURCES CITATION
WEB OF SCIENCE
Web of Science Group
Crossref
doi
ResearchGate
R^G
publons