Supplementary Material

Results on the performance of JET and ET on the Huang dataset (homodimers, heterodimers and transients) are reported. JET and ET have been executed without (Tables 1, 2, 3) and with (Tables 4, 5, 6) clusterisation. The addition of information on physical-chemical properties of residues is evaluated as well as the importance of iterating JET to extract a set of reliable interface residues (Tables 7, 8, 9). For the proteins in the Huang dataset and the extra proteins discussed in the article, we report the size and the number of sequences retrieved with PSI-BLAST after filtering (Tables 10-13). Propensity values of residues are given in Table 14. Information on how to run the program are described below. Results of the large scale analysis of JET on the Kanamori dataset (signal transduction proteins, enzymes, inhibitors, antibodies, antigens and others) are reported (Tables 15-20).

How to run JET and default values of its parameters

Installation note : JET has to be installed with ClustalW and naccess. PSI-BLAST can be installed too but this is not a requirement.

List of default values of parameters stored in the config file used in the system :

PSI-BLAST : the program can be called on the server or on local. To reach similar outputs on the server and on local, we coded a local call to PSI-BLAST with the option -t 2 setting the composition-based score adjustment method conditioned on sequence properties, for improving accuracy in finding true positive matches ; attention, because -t 2 is a PSI-BLAST option and not a JET option).

Notice that a PSI-BLAST threshold = 0.005 (this is a statistical significance threshold to include a sequence in the model used by PSI-BLAST to create the PSSM in the next iteration) is used for both server and local calls.

- gap opening cost = 11
- gap extension =3
- e-value = 10^{-5} for extracting sequences with PSI-BLAST. The automatic adjustment of this default value has been discussed in the text, in case not sufficiently many sequences have been retrieved ;
- number of sequences retrieved = 5000
- database used for PSI-BLAST search = nr
- matrix used in PSI-BLAST to fetch homologues = Blosum62
- number of iterations for PSI-BLAST = 3
- output format = text (corresponding to the local PSI-BLAST option `-T F`)

Software location :

- clustalW location
- naccess location
- psiblast location

Sequence filtering :

- alignments issued with PSI-BLAST are selected based on their length which has to satisfy a min cut-off of 80% and a max cut-off of 110% of the reference sequence length ; this ensures overlapping of all aligned sequences ;
- a min cut-off of 20% and a max cut-off of 98% of sequence identity between retrieved sequences and reference sequence ;

Tree construction :

- alignment matrix = Blosum62. An automatic selection for another matrix (Gonnet and HSDM) under suitable conditions has been discussed in the text.
- msaNumber n fixes the number of alignments (trees) at $n \geq 1$; -1 asks JET to compute itself the best value depending on the number of retrieved sequences.
- seqNumber n fixes the number of sequences aligned for each tree ; -1 asks JET to compute itself the best value depending on the number of retrieved sequences.

- coverage traces : trace levels are computed for a maximum of 95% of the residues of the reference sequence starting from highest values ;

Examples : the options msaNumber and seqNumber have been implemented in such a way that the set of retrieved (and filtered) sequences will be covered (by random selection) in the best way, whenever possible. The automatic handling of trees by JET (corresponding to the value -1) generates at most 50 trees of at most 50 sequences. For instance :

-msaNumber 1 -seqNumber 50 creates 1 tree of 50 sequences.
-msaNumber 3 -seqNumber -1 creates 3 trees with $N/3$ sequences.
-msaNumber -1 -seqNumber 50 creates $N/50$ trees (with a minimum number of 2 trees).
-msaNumber 1 -seqNumber -1 creates 1 tree with all retrieved (and filtered) sequences.
-msaNumber -1 -seqNumber -1 creates $\sqrt{N}$ alignments of $\sqrt{N}$ sequences.

Clustering :

- radius of probe used for accessible surface detection $= 1.4\mathring{A}$;
- minimum percentage of accessible surface for a residue to be considered accessible $= 5\%$;
- minimum accessible surface for an atom to be considered accessible $= 1\mathring{A}^2$;
- maximum distance between residues to aggregate them in clusters $= 5\mathring{A}$;
- mean coverage of clusters (relative to protein surface), calculated by JET (-1) as a function of surface size using the curve in Figure 4, or provided by the user ;

Calculation of protein interacting sites :

- percentage variation of accessible surface of a residue computed within a protein complex and a single protein has to be higher than 10% for the residue to belong to the interaction site ;
- ligand set to yes (no) to take into account ligands in the PDB database for the evaluation of predicted interactions ;
- homologousPDB set to yes (no) to add interface residues of homologous structures for the evaluation of predicted interactions ;

Command lines and options to exploit JET functioning modes

Command lines :
`java jet.JET [option]`

Mandatory :
-c config-file : file containing all values of parameters discussed above
-i input-file : input pdb file or directory with all input pdb files. These files must match to the pattern pdbCode_chain.pdb
-o output-directory : directory where JET output files will be generated
-p type-of-program {AIJCR} : A to compute accessibility of residues and atoms, I to compute interface residues if the pdb input file is a complex, J to launch JET analysis, C to launch the clustering algorithm, and R to evaluate jet results according to real interface residues (I analysis needed)

Optional :
-l log-file : file containing characteristics of the JET analysis for each protein (pdb code, length of the protein, number of retrieved sequences in identity classes)
-b blast-file : PSI-BLAST input file or directory containing PSI-BLAST input files used by a large scale JET analysis. These files must match to the pattern pdbCode_chain.psiblast
-w pdb_code : pdb code of a structural complex or of a protein that the user wants to analyze. The pdb file corresponding to the pdb code is retrieved on the pdb database site found in the config file
-f fasta-file : fasta input file or directory fasta input files used by the JET analysis. These files must match to the pattern pdbCode_chain.fasta
-m merging-option {T|F} : if "T", several input pdb files with same pdb code are merged in one pdb file. If "F", no merging is done. This option merges pdb files (containing different chains of a complex) and allows

calculation of real interfaces ; an example of accepted pdb file names : 1apm_A.pdb et 1apm_B.pdb

-s coverage-threshold $]0.0, 0.5[$: mean coverage of clusters computed by JET

-n nb (1,50) : if $nb > 1$, iJET runs with nb iterations. For $nb = 1$, basic JET is run

-t threshold (1,50) : in iterative mode (see option -n) residues which appear a number of times $\geq$ threshold are selected

-a type-of-analysis {1|2} : 1 for JET analysis based on conservation properties (trace), 2 for JET analysis based on conservation (trace) and physical-chemical properties (pc)

-h : online help

-d accessType {chain|complex} : accessibility computed on a chain (chain) or on a complex (complex) ; in this latter case certain residues in the complex will be inaccessible because belonging to the complex interface ; attention here, because residues at the complex interface are not necessarily inaccessible.

-r retrieving-method {input ;server ;local} : input for input file (assume the use of option -b or -f) ; local for local psiblast analysis (the command to run local psiblast must appear in the config-file) ; server for server psiblast analysis (web address of the psiblast server must appear in the config-file).

-g pdb-results-file (name1,name2, ...) : list of names separated by commas in parenthesis. Names are those of the results columns in the results files {tr ;freq ;pc ;trace ;clusters ;axs ;surfAxs ;percentSurfAxs ;inter ;atomAxs ;atomSurfAxs}. Columns results selected with these names are written in pdb file format (temperature factor column of the pdb file containing value of selected column) and could be viewed in rasmol/VMD/etc. Example : (trace,clusters) if the user wants to have trace and clusters results in pdb format.

JET output.

JET outputs several main files collecting results :

1. JET prediction (<filename>_jet.res). For each residue in the reference sequence S, it lists : residue name r_j, position j in the chain, chain name, trace $d(j)$, trace frequency in the generated trees (number of residues with an average trace value $trace(j) \neq 0$/ number of trees), propensity value, trace*frequency, accessibility (0 if not accessible, 1 if accessible), mixed trace for clustered residues.

If iJET is run instead, with $i \neq 1$, then the file <filename>_jet.res contains some extra columns. For each residue in the reference sequence S, it lists : residue name r_j, position j in the chain, chain name, trace $d(j)$, trace frequency in the generated trees (number of residues with an average trace value $trace(j) \neq 0$/ number of trees), propensity value, trace*frequency for all runs, maximal trace value obtained in all runs, mixed trace for clustered residues calculated for each run, number of runs that the residue appeared in a cluster.

2. NACCESS output (<filename>_axs.res). It is rearranged in a new format : residue name, position in the chain, chain name, whether or not it is an accessible residue, the accessible surface a_j, the fraction of the residue accessible surface.

The file <filename>_atomAxs.res contains information on atomic accessibility. Columns are : atom code, atom position, whether or not it is accessible, the fraction of atom accessible surface.

3. Clusters information (<filename>_clusters.res). It contains : residue name, position, chain, mixed trace value if the residue belongs to a cluster, 0 otherwise. This information is included in <filename>_jet.res.

4. Characteristics of the computation (caracTest.dat) : protein name, chain name, enzyme compounds, names of homologous PDB structures, protein size, number of retrieved sequences after filtering that occur in the four identity classes. This file can be found in the same directory where the program JET is located and collects the history of all PDB files that have been executed.

Example for the protein 1APM, chain E :
```
    >1apm :E
    EnzymeCompound        ATP ADP GDP PHA GNP ADP
    HomologousPDB        1atp 1jbp
    size        341
    partition        370 ;242 ;12 ;131
```

JET also generates PDB files where the "temperature factor" column is replaced with specific information depending on user defined options to visualize properties such as surface clusters and trace significance with some visualization program like VMD, rasmol or others.

	JET trace results								ET trace results							
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1bnc :A	30.6	9.5	33.8	1.45	81.8	2.9	69.8	4.4	33.3	10.9	34.8	1.49	80.9	3.3	69.8	5.1
1daa :A	48.5	26.8	69.6	2.23	90.4	12.1	77.4	16.7	53.0	30.9	74.5	2.39	91.8	13.9	79.7	19.2
1dpg :A	41.7	19.4	43.8	1.87	83.6	5.9	73.8	9.1	41.7	20.2	45.5	1.94	84.7	6.2	74.7	9.5
1ecp :B	36.6	13.4	76.9	1.58	89.5	12.7	63.7	13.1	39.0	16.4	84.2	1.73	93.0	15.6	66.7	16.0
1efu :A	31.2	7.3	37.5	1.31	79.1	2.9	65.3	4.2	37.7	11.9	42.0	1.46	79.1	4.8	67.2	6.8
1efu :B	45.2	21.9	60.0	1.94	86.5	9.8	73.7	13.5	53.4	29.7	69.6	2.25	89.6	13.3	78.4	18.4
1frp :A	35.3	14.3	63.8	1.68	87.8	8.8	67.9	10.9	38.8	18.7	73.3	1.93	91.4	11.5	71.4	14.2
1fuq :A	61.1	36.9	41.3	2.53	83.0	7.2	79.5	12.1	66.7	41.6	43.4	2.66	83.0	8.1	80.4	13.6
1gdh :A	29.0	9.1	41.7	1.46	83.7	3.6	68.0	5.2	26.1	5.8	36.7	1.28	82.0	2.3	66.0	3.3
1ges :A	46.1	22.4	65.4	1.95	87.7	11.3	73.7	15.1	47.8	23.3	65.5	1.95	87.2	11.8	74.0	15.6
1glq :A	34.9	14.4	45.5	1.7	84.7	5.2	71.4	7.7	37.2	14.2	43.2	1.62	82.2	5.2	70.2	7.6
1gpm :B	30.2	6.1	29.2	1.25	77.7	1.9	66.7	2.8	32.6	8.2	31.1	1.33	78.1	2.5	67.5	3.8
1hur :A	43.2	22.0	67.9	2.04	89.8	11.0	74.2	14.6	43.2	22.7	70.4	2.11	90.9	11.4	75.0	15.2
1hyh :A	27.4	4.0	57.1	1.17	80.5	3.8	54.6	3.9	29.1	4.5	57.6	1.18	79.7	4.3	55.0	4.4
1ids :A	30.3	10.2	62.5	1.51	87.1	7.2	63.5	8.4	22.7	2.6	46.9	1.13	81.7	1.8	57.2	2.1
1ies :B	21.1	2.3	64.0	1.12	84.2	3.0	48.1	2.6	22.4	4.3	70.8	1.24	87.7	5.8	50.4	4.9
1leh :A	3.7	-16.2	1.9	0.19	78.3	-1.8	70.8	-3.2	7.4	-12.2	3.8	0.38	79.1	-1.3	72.0	-2.4
1mas :A	34.6	19.5	58.1	2.29	91.5	6.6	77.1	9.9	28.8	11.3	41.7	1.64	86.3	3.8	71.7	5.7
1mld :A	49.2	25.5	54.7	2.08	85.5	9.1	75.9	13.4	54.2	30.1	59.3	2.25	86.7	10.8	78.1	15.9
1nhk :L	33.3	9.7	46.2	1.41	81.1	4.7	65.5	6.3	36.1	12.5	50.0	1.53	82.4	6.1	67.3	8.2
1nqv :A	29.9	6.4	74.1	1.27	85.4	8.9	53.0	7.4	31.3	7.0	75.0	1.29	85.4	9.8	53.9	8.2
1oro :A	48.9	22.3	51.2	1.84	82.1	8.6	72.8	12.4	53.3	27.4	57.1	2.06	84.6	10.5	75.9	15.2
1osj :A	42.4	19.6	49.1	1.86	84.2	7.0	73.2	10.4	45.5	22.7	52.6	1.99	85.3	8.1	74.8	12.0
1pky :A	25.9	2.8	18.4	1.12	77.5	0.6	69.0	0.9	27.8	5.3	20.3	1.23	78.5	1.0	70.2	1.7
1poy :1	29.7	11.5	26.2	1.64	84.0	2.2	75.3	3.7	29.7	10.2	24.4	1.53	82.5	2.0	74.0	3.3
1qor :A	37.5	18.0	45.7	1.92	86.1	5.6	74.6	8.5	37.5	19.3	48.8	2.06	87.8	6.0	75.8	9.1
1rah :B	31.3	10.0	53.6	1.47	84.5	5.7	65.2	7.3	22.9	3.2	42.3	1.16	82.1	1.8	60.6	2.3
1scu :D	44.3	20.4	74.5	1.86	89.8	13.7	71.6	16.4	46.8	22.5	77.1	1.92	90.7	15.0	73.1	18.0
1scu :E	37.0	15.3	46.9	1.71	84.1	5.8	71.2	8.4	39.5	16.1	46.4	1.69	82.7	6.1	70.8	8.8
1set :A	40.0	18.1	41.1	1.82	83.3	5.3	73.6	8.1	42.7	19.8	42.1	1.87	82.9	5.8	73.9	8.9
1sft :A	41.0	19.6	61.8	1.92	88.0	9.3	72.9	12.6	39.8	17.3	56.9	1.77	85.7	8.2	70.9	11.1
1tph :1	54.8	33.1	62.2	2.53	89.1	10.8	80.7	16.3	52.4	30.2	57.9	2.36	87.6	9.8	78.9	14.8
1xik :A	25.8	10.2	43.2	1.65	88.0	3.6	71.7	5.3	19.4	5.0	35.3	1.35	87.4	1.8	69.6	2.6
2cst :A	32.5	12.1	45.0	1.59	84.4	4.8	69.7	6.8	31.3	10.6	42.6	1.51	83.4	4.2	68.7	6.0
2eip :A	18.8	-3.5	10.3	0.84	77.2	-0.5	70.0	-0.9	25.0	2.7	13.8	1.12	78.1	0.4	71.5	0.7
2hhm :A	39.0	19.3	42.1	1.98	85.5	5.2	75.6	8.2	41.5	20.2	41.5	1.95	84.2	5.5	75.1	8.6
2pcd :B	29.9	5.3	66.7	1.25	85.3	6.0	53.8	5.6	24.7	4.1	63.6	1.2	84.0	4.6	52.5	4.3
2pcd :N	19.4	1.4	79.4	1.08	86.0	4.0	37.0	2.1	20.9	2.3	82.9	1.13	88.0	6.5	38.6	3.4
2pol :A	49.2	24.5	45.6	1.99	82.5	7.3	74.9	11.2	46.0	20.9	42.0	1.83	81.1	6.2	73.1	9.6
3lad :A	39.1	16.9	54.2	1.76	85.3	7.5	71.0	10.4	44.3	20.8	58.0	1.88	85.7	9.3	72.9	12.8
3mde :A	48.1	23.6	53.6	1.96	84.4	8.9	74.5	12.9	51.9	26.4	55.6	2.03	84.4	9.9	75.5	14.4
6gsv :A	46.5	23.2	52.6	2	85.0	8.3	74.8	12.2	39.5	16.2	44.7	1.7	82.5	5.8	71.2	8.6
8cat :A	39.2	17.4	53.4	1.8	85.6	7.3	71.9	10.3	34.2	11.9	45.6	1.54	82.8	5.0	68.4	7.1
1gp1 :A	28.6	13.1	28.6	1.85	87.0	2.4	77.9	4.1	0.0	0.0	0.0	0	100.0	0.0	86.8	0.0

TAB. 1 – Comparison of ET and JET traces results for homodimers of the Huang benchmark. Protein `1gp1 :A` has been evaluated because it belongs to the dataset but its performance has not been included in the average (denoted `All`) reported in the article because no answer is obtained with ET. Structure `1rvv` in the Huang dataset has been replaced by structure `1nqv` which contains only 5 chains instead of 30 ; this change makes no difference in JET/ET prediction.

Heterodimers

pdbCode	JET trace results								ET trace results							
	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1all :A	42.2	17.0	57.6	1.68	83.7	8.9	69.5	11.7	42.2	18.6	61.3	1.78	86.0	9.7	71.0	12.7
1all :B	48.9	24.4	66.7	2	87.8	12.2	74.8	16.3	40.0	14.8	52.9	1.59	82.2	7.4	68.1	9.9
1hcg :A	26.9	10.7	25.9	1.66	85.7	2.0	76.5	3.3	19.2	3.0	18.5	1.18	84.3	0.5	74.1	0.9
1hcg :B	31.6	6.1	46.2	1.24	78.1	3.6	60.8	4.5	26.3	0.8	38.5	1.03	75.0	0.5	56.9	0.6
1luc :A	37.8	16.7	33.3	1.79	82.7	3.8	74.4	6.2	46.7	26.8	43.8	2.35	86.3	6.1	78.9	10.0
1luc :B	50.0	29.0	48.9	2.38	86.5	7.5	79.0	11.9	47.8	28.2	50.0	2.43	87.6	7.3	79.5	11.6
1scu :D	44.3	20.4	74.5	1.86	89.8	13.7	71.6	16.4	46.8	22.5	77.1	1.92	90.7	15.0	73.1	18.0
1scu :E	37.0	15.3	46.9	1.71	84.1	5.8	71.2	8.4	39.5	16.1	46.4	1.69	82.7	6.1	70.8	8.8
1tco :A	17.3	2.1	37.1	1.14	85.8	1.0	63.5	1.4	17.3	3.0	39.4	1.21	87.1	1.4	64.3	1.9
1tco :B	26.7	5.9	51.6	1.28	83.1	4.0	60.4	4.7	35.0	12.9	63.6	1.58	86.5	8.7	65.8	10.3
1tco :C	31.0	9.7	47.4	1.45	83.3	4.7	66.3	6.3	31.0	10.8	50.0	1.53	85.0	5.2	67.4	7.0
1tcr :A	34.6	11.8	45.0	1.51	82.1	5.0	68.0	7.0	42.3	19.5	55.0	1.85	85.4	8.2	72.6	11.6
1tcr :B	37.7	17.6	51.3	1.88	86.5	6.6	73.2	9.6	32.1	11.5	42.5	1.56	83.7	4.3	69.6	6.3
1ubs :A	55.6	31.9	45.5	2.35	84.0	7.7	78.5	12.3	61.1	36.9	48.9	2.53	84.7	8.9	80.1	14.3
1ubs :B	39.2	21.3	45.5	2.19	87.7	5.6	77.6	8.8	43.1	24.4	47.8	2.31	87.7	6.4	78.5	10.1
1wdc :A	24.1	-0.9	81.3	0.96	70.0	-5.0	31.3	-1.6	24.1	-0.9	81.3	0.96	70.0	-5.0	31.3	-1.6
1wdc :B	36.0	13.3	60.0	1.58	85.4	8.1	66.7	10.1	40.0	15.8	62.5	1.65	85.4	9.6	68.2	11.9
1wdc :C	40.9	16.8	56.3	1.7	84.3	8.3	69.9	11.1	47.7	22.9	63.6	1.92	86.5	11.3	73.7	15.2
2pcd :B	25.9	5.3	66.7	1.25	85.3	6.0	53.8	5.6	24.7	4.1	63.6	1.2	84.0	4.6	52.5	4.3
2pcd :N	19.4	1.4	79.4	1.08	86.0	4.0	37.0	2.1	20.9	2.3	82.9	1.13	88.0	6.5	38.6	3.4
8atc :A	37.1	18.9	34.2	2.03	85.5	3.8	77.4	6.4	37.1	19.4	35.1	2.09	86.1	3.9	77.9	6.5
8atc :B	37.3	16.8	73.1	1.82	90.8	11.3	69.3	13.5	35.3	15.6	72.0	1.79	90.8	10.5	68.5	12.5
9atc :A	42.1	22.8	39.0	2.18	85.6	5.0	77.8	8.2	39.5	21.1	38.5	2.15	86.2	4.6	77.8	7.6
9atc :B	36.5	13.8	63.3	1.61	86.3	9.0	66.7	10.9	36.5	13.8	63.3	1.61	86.3	9.0	66.7	10.9

Tab. 2 – Comparison of ET and JET traces results for heterodimers of the Huang benchmark

Transients

pdbCode	JET trace results								ET trace results							
	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1apm :E	52.2	33.4	50.0	2.78	88.6	7.3	82.0	12.0	60.9	41.3	56.0	3.12	89.5	9.1	84.4	14.9
1efu :A	31.2	7.3	37.5	1.31	79.1	2.9	65.3	4.2	37.7	11.9	42.0	1.46	79.1	4.8	67.2	6.8
1efu :B	45.2	21.9	60.0	1.94	86.5	9.8	73.7	13.5	53.4	29.7	69.6	2.25	89.6	13.3	78.4	18.4
1g3n :A	15.8	-4.8	33.3	0.77	75.7	-3.6	49.8	-4.1	15.8	-5.2	32.7	0.75	75.0	-4.0	49.4	-4.5
1g3n :B	12.5	-4.2	15.0	0.75	82.3	-1.0	68.3	-1.7	4.2	-15.0	4.3	0.22	77.1	-3.8	62.5	-6.0
1g3n :C	27.0	6.1	27.8	1.29	80.7	1.7	69.2	2.6	16.2	-3.5	17.6	0.82	79.3	-1.0	65.7	-1.5
1got :A	44.2	21.4	30.2	1.94	81.1	3.9	75.4	6.7	51.2	29.1	36.1	2.31	83.3	5.4	78.3	9.1
1got :B	20.2	1.3	39.6	1.07	81.9	0.8	59.1	1.0	20.2	2.1	41.3	1.12	83.1	1.2	59.8	1.6
1got :G	30.2	4.4	86.7	1.17	86.7	12.5	44.8	6.5	32.6	6.7	93.3	1.26	93.3	19.2	48.3	9.9
1k9o :E	36.0	22.3	40.9	2.63	90.4	4.1	82.0	6.9	32.0	19.6	40.0	2.58	91.2	3.6	82.0	6.1
1k9o :I	12.5	-3.7	4.7	0.77	83.5	-0.2	79.2	-0.5	12.5	-5.6	4.2	0.69	81.5	-0.4	77.4	-0.7
1rrp :A	20.0	-1.7	51.4	0.92	76.0	-2.2	44.7	-1.9	26.7	4.3	66.7	1.19	83.1	5.5	51.6	4.8
1rrp :B	32.7	12.0	73.9	1.58	89.8	10.6	63.1	11.2	30.8	11.0	72.7	1.55	89.8	9.6	62.2	10.3
1rrp :C	22.4	0.6	56.7	1.03	79.0	0.8	47.8	0.7	28.9	5.8	68.8	1.25	83.9	7.0	53.6	6.3
1ugh :E	44.8	31.2	59.1	3.28	93.2	6.8	84.5	11.2	55.2	39.0	61.5	3.42	92.4	8.6	85.7	14.1
1ugh :I	11.1	-9.2	21.4	0.55	73.8	-5.9	49.3	-7.2	0.0	0.0	0.0	0	100.0	0.0	60.9	0.0
1ytf :A	48.2	22.3	75.0	1.86	89.2	15.1	72.7	18.0	50.0	24.1	77.8	1.93	90.4	16.3	74.1	19.4
1ytf :B	27.3	1.2	100.0	1.05	100.0	26.1	30.4	2.3	22.7	-3.4	83.3	0.87	0.0	-73.9	21.7	-6.4
1ytf :D	26.5	2.0	75.0	1.08	80.0	4.5	42.9	2.7	27.9	3.5	79.2	1.14	83.3	7.8	44.9	4.8

Tab. 3 – Comparison of ET and JET traces results for transients of the Huang benchmark

5

Homodimers																
	JET with clustering								ET with clustering							
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1bnc :A	19.4	8.4	41.2	1.76	91.5	2.6	74.7	3.9	30.6	9.1	33.3	1.43	81.4	2.8	69.5	4.3
1daa :A	48.5	32.0	91.4	2.94	97.9	14.5	82.5	19.9	51.5	31.7	81.0	2.6	94.5	14.3	81.1	19.7
1dpg :A	21.4	11.1	48.6	2.08	93.1	3.4	76.3	5.2	36.9	16.6	42.5	1.81	84.7	5.1	73.5	7.8
1ecp :B	43.9	21.3	94.7	1.94	97.7	20.3	71.4	20.8	37.8	17.0	88.6	1.81	95.3	16.2	67.3	16.6
1efu :A	35.1	19.4	64.3	2.24	92.1	7.8	75.7	11.1	37.7	12.3	42.6	1.48	79.6	5.0	67.5	7.1
1efu :B	26.0	12.9	61.3	1.98	92.6	5.8	72.0	8.0	52.1	28.7	69.1	2.23	89.6	12.9	78.0	17.8
1frp :A	30.6	15.9	78.8	2.08	95.0	9.7	70.5	12.0	35.3	16.5	71.4	1.88	91.4	10.1	70.1	12.6
1fuq :A	55.6	37.7	50.8	3.12	89.5	7.4	84.0	12.3	66.7	42.5	45.0	2.76	84.1	8.3	81.3	13.9
1gdh :A	18.8	3.9	36.1	1.26	86.6	1.6	67.2	2.2	23.2	4.5	35.6	1.24	83.1	1.8	66.0	2.6
1ges :A	32.2	19.3	84.1	2.5	96.9	9.8	75.1	13.0	46.1	23.3	67.9	2.02	89.0	11.8	74.6	15.7
1glq :A	67.4	45.7	82.9	3.1	94.9	16.7	87.6	24.4	37.2	16.1	47.1	1.76	84.7	5.9	72.0	8.6
1gpm :B	19.8	7.8	38.6	1.66	90.5	2.4	74.0	3.7	32.6	9.5	32.9	1.41	79.9	2.9	68.8	4.4
1hur :A	54.5	29.5	72.7	2.18	89.8	14.8	78.0	19.7	40.9	22.0	72.0	2.16	92.0	11.0	75.0	14.6
1hyh :A	18.8	5.5	68.8	1.41	91.9	5.2	56.3	5.3	27.4	4.0	57.1	1.17	80.5	3.8	54.6	3.9
1ids :A	42.4	16.6	68.3	1.65	86.0	11.8	67.9	13.8	22.7	3.2	48.4	1.17	82.8	2.3	57.9	2.7
1ies :B	25.0	3.9	67.9	1.19	84.2	5.3	50.4	4.5	22.4	4.3	70.8	1.24	87.7	5.8	50.4	4.9
1leh :A	0.0	-13.3	0.0	0	85.2	-1.5	76.8	-2.6	3.7	-15.1	2.0	0.2	79.5	-1.7	72.0	-3.0
1mas :A	57.7	32.8	58.8	2.32	86.3	11.2	79.0	16.6	26.9	11.3	43.8	1.72	88.2	3.8	72.7	5.7
1mld :A	44.1	27.5	70.3	2.67	93.3	9.9	80.4	14.5	52.5	29.8	60.8	2.31	87.9	10.6	78.6	15.7
1nhk :L	11.1	-10.7	16.7	0.51	73.0	-5.2	52.7	-7.0	33.3	10.6	48.0	1.47	82.4	5.2	66.4	6.9
1nqv :A	49.3	18.8	94.3	1.62	95.8	26.3	68.7	21.9	29.9	7.2	76.9	1.32	87.5	10.1	53.9	8.4
1oro :A	51.1	28.3	62.2	2.24	88.0	10.9	77.8	15.7	53.3	28.0	58.5	2.11	85.5	10.8	76.5	15.6
1osj :A	25.8	10.6	44.7	1.69	88.6	3.8	72.0	5.6	43.9	21.5	51.8	1.96	85.3	7.7	74.4	11.4
1pky :A	18.5	5.8	23.8	1.45	88.4	1.1	76.9	1.9	27.8	8.0	23.1	1.41	81.8	1.6	72.9	2.6
1poy :1	21.6	10.8	32.0	2	91.2	2.1	80.1	3.5	27.0	9.7	25.0	1.56	84.5	1.9	75.3	3.1
1qor :A	50.0	35.6	82.4	3.47	96.7	11.1	85.6	16.9	37.5	21.4	55.3	2.33	90.6	6.7	78.0	10.2
1rah :B	43.8	15.7	56.8	1.56	81.0	9.0	67.4	11.4	20.8	2.7	41.7	1.15	83.3	1.5	60.6	1.9
1scu :D	50.6	29.3	95.2	2.37	98.3	19.6	79.2	23.5	45.6	22.7	80.0	1.99	92.4	15.2	73.6	18.2
1scu :E	27.2	11.9	48.9	1.78	89.3	4.5	72.2	6.5	37.0	16.0	48.4	1.76	85.0	6.1	71.9	8.8
1set :A	28.0	14.8	47.7	2.12	91.1	4.3	76.9	6.7	40.0	19.0	42.9	1.9	84.5	5.5	74.5	8.5
1sft :A	48.2	30.0	85.1	2.65	96.0	14.2	80.6	19.3	39.8	20.0	64.7	2.01	89.7	9.5	73.6	12.9
1tph :1	54.8	35.5	69.7	2.84	92.2	11.5	83.0	17.4	50.0	28.9	58.3	2.37	88.4	9.4	78.9	14.2
1xik :A	37.1	19.8	56.1	2.14	89.7	7.0	75.9	10.4	19.4	5.9	37.5	1.43	88.6	2.1	70.5	3.1
2cst :A	28.9	14.3	55.8	1.98	91.0	5.6	73.5	8.1	28.9	9.2	41.4	1.47	83.9	3.6	68.4	5.2
2eip :A	18.8	-5.1	9.7	0.79	75.4	-0.7	68.5	-1.3	25.0	4.2	14.8	1.2	79.8	0.6	73.1	1.0
2hhm :A	9.8	2.0	26.7	1.26	92.8	0.5	75.1	0.8	41.5	22.3	45.9	2.16	86.8	6.0	77.2	9.5
2pcd :B	37.6	12.0	78.0	1.47	88.0	13.6	61.3	12.8	24.7	4.7	65.6	1.24	85.3	5.3	53.1	5.0
2pcd :N	26.6	2.8	82.2	1.12	84.0	7.8	41.8	4.1	18.7	2.3	83.9	1.14	90.0	6.4	37.6	3.4
2pol :A	23.8	7.4	33.3	1.46	85.8	2.2	71.6	3.4	46.0	21.7	43.3	1.89	82.1	6.4	73.8	9.9
3lad :A	28.7	17.7	80.5	2.61	96.9	7.9	75.9	10.9	41.7	20.0	59.3	1.92	87.2	8.9	73.2	12.3
3mde :A	53.2	37.3	91.1	3.34	98.0	14.0	85.8	20.4	51.9	28.5	60.6	2.22	87.3	10.7	77.7	15.6
6gsv :A	53.5	33.2	69.7	2.64	91.7	11.9	81.6	17.5	37.2	16.4	47.1	1.78	85.0	5.9	72.4	8.6
8cat :A	35.0	21.2	75.0	2.53	95.1	8.9	77.3	12.5	31.7	11.4	46.3	1.56	84.6	4.8	68.9	6.8
1gp1 :A	61.9	41.3	46.4	3.01	87.0	7.5	83.1	12.8	0.0	0.0	0.0	0	100.0	0.0	86.8	0.0

TAB. 4 – Comparison of ET and JET performance after clustering for homodimers of the Huang benchmark

Heterodimers																
	JET with clustering								ET with clustering							
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1all :A	37.8	12.6	51.5	1.5	81.4	6.6	66.4	8.6	42.2	20.1	65.5	1.91	88.4	10.5	72.5	13.8
1all :B	51.1	20.7	56.1	1.68	80.0	10.4	70.4	13.8	40.0	15.6	54.5	1.64	83.3	7.8	68.9	10.4
1hcg :A	30.8	7.3	20.5	1.31	77.9	1.3	70.5	2.3	19.2	3.0	18.5	1.18	84.3	0.5	74.1	0.9
1hcg :B	68.4	11.6	44.8	1.2	50.0	6.9	56.9	8.6	26.3	4.7	45.5	1.22	81.3	2.8	60.8	3.5
1luc :A	44.4	27.1	47.6	2.56	88.8	6.2	80.6	10.1	46.7	26.8	43.8	2.35	86.3	6.1	78.9	10.0
1luc :B	60.9	43.9	73.7	3.59	94.4	11.3	87.5	18.0	47.8	29.5	53.7	2.61	89.3	7.6	80.8	12.1
1scu :D	50.6	29.3	95.2	2.37	98.3	19.6	79.2	23.5	45.6	22.7	80.0	1.99	92.4	15.2	73.6	18.2
1scu :E	27.2	11.9	48.9	1.78	89.3	4.5	72.2	6.5	37.0	16.0	48.4	1.76	85.0	6.1	71.9	8.8
1tco :A	21.3	5.7	44.4	1.36	87.1	2.7	65.7	3.7	14.7	1.6	36.7	1.12	87.7	0.8	63.9	1.1
1tco :B	45.0	18.2	67.5	1.68	85.4	12.2	69.1	14.6	31.7	10.9	61.3	1.52	86.5	7.3	64.4	8.7
1tco :C	55.2	14.7	44.4	1.36	66.7	7.1	62.9	9.6	31.0	13.1	56.3	1.73	88.3	6.3	69.7	8.5
1tcr :A	46.2	19.9	52.2	1.76	82.1	8.4	71.4	11.8	40.4	19.2	56.8	1.91	87.0	8.1	73.1	11.4
1tcr :B	45.3	19.0	47.1	1.72	80.9	7.1	71.1	10.4	32.1	11.5	42.5	1.56	83.7	4.3	69.6	6.3
1ubs :A	44.4	31.5	66.7	3.44	94.7	7.6	84.9	12.2	61.1	38.0	51.2	2.64	86.0	9.1	81.2	14.7
1ubs :B	3.9	3.1	100.0	4.82	100.0	0.8	80.1	1.3	39.2	21.7	46.5	2.24	88.2	5.7	78.0	9.0
1wdc :A	40.7	1.7	88.0	1.04	70.0	9.1	45.3	2.8	24.1	-0.9	81.3	0.96	70.0	-5.0	31.3	-1.6
1wdc :B	30.0	9.5	55.6	1.47	85.4	5.8	64.4	7.2	36.0	13.3	60.0	1.58	85.4	8.1	66.7	10.1
1wdc :C	47.7	18.4	53.8	1.63	79.8	9.1	69.2	12.2	47.7	22.9	63.6	1.92	86.5	11.3	73.7	15.2
2pcd :B	37.6	12.0	78.0	1.47	88.0	13.6	61.3	12.8	24.7	4.7	65.6	1.24	85.3	5.3	53.1	5.0
2pcd :N	26.6	2.8	82.2	1.12	84.0	7.8	41.8	4.1	18.7	2.3	83.9	1.14	90.0	6.4	37.6	3.4
8atc :A	34.3	21.3	44.4	2.64	91.3	4.3	81.7	7.2	37.1	20.3	37.1	2.21	87.3	4.1	78.8	6.8
8atc :B	58.8	27.3	75.0	1.87	86.8	18.3	75.6	21.9	35.3	17.2	78.3	1.95	93.4	11.5	70.1	13.8
9atc :A	39.5	23.9	45.5	2.54	89.7	5.2	80.7	8.6	39.5	22.0	40.5	2.26	87.4	4.8	78.8	7.9
9atc :B	53.8	25.8	75.7	1.92	88.8	16.8	75.0	20.3	34.6	13.4	64.3	1.63	87.5	8.7	66.7	10.6

TAB. 5 – Comparison of ET and JET performance after clustering for heterodimers of the Huang benchmark

Transients																
	JET with clustering								ET with clustering							
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1apm :E	52.2	41.2	85.7	4.77	98.1	9.0	89.8	14.8	58.7	39.9	56.3	3.13	90.0	8.8	84.4	14.4
1efu :A	35.1	19.4	64.3	2.24	92.1	7.8	75.7	11.1	37.7	12.3	42.6	1.48	79.6	5.0	67.5	7.1
1efu :B	26.0	12.9	61.3	1.98	92.6	5.8	72.0	8.0	52.1	28.7	69.1	2.23	89.6	12.9	78.0	17.8
1g3n :A	15.8	0.0	43.2	1	84.1	0.0	54.5	0.0	15.8	-4.8	33.3	0.77	75.7	-3.6	49.8	-4.1
1g3n :B	12.5	-21.7	7.3	0.37	60.4	-5.4	50.8	-8.7	4.2	-14.2	4.5	0.23	78.1	-3.5	63.3	-5.7
1g3n :C	40.5	17.9	38.5	1.79	82.2	4.9	73.3	7.7	16.2	-2.4	18.8	0.87	80.7	-0.7	66.9	-1.0
1got :A	44.2	32.2	57.6	3.7	94.0	5.9	86.2	10.0	48.8	28.5	37.5	2.41	85.0	5.3	79.3	8.9
1got :B	17.0	6.0	57.1	1.54	92.5	3.5	64.6	4.4	18.1	3.1	44.7	1.21	86.9	1.8	61.4	2.3
1got :G	74.4	10.6	86.5	1.17	66.7	30.5	72.4	15.8	30.2	6.1	92.9	1.25	93.3	17.5	46.6	9.0
1k9o :E	52.0	37.7	56.5	3.64	92.6	6.9	86.3	11.7	32.0	20.8	44.4	2.86	92.6	3.8	83.2	6.5
1k9o :I	18.8	3.7	7.5	1.24	85.1	0.2	81.1	0.4	12.5	-3.7	4.7	0.77	83.5	-0.2	79.2	-0.5
1rrp :A	24.4	0.2	56.4	1.01	76.0	0.3	47.2	0.2	26.7	4.9	68.6	1.23	84.5	6.2	52.2	5.5
1rrp :B	46.2	19.1	80.0	1.71	89.8	16.9	69.4	17.9	30.8	11.0	72.7	1.55	89.8	9.6	62.2	10.3
1rrp :C	25.0	0.4	55.9	1.01	75.8	0.4	47.8	0.4	28.9	5.8	68.8	1.25	83.9	7.0	53.6	6.3
1ugh :E	75.9	51.0	55.0	3.05	86.4	11.2	84.5	18.4	55.2	39.6	64.0	3.55	93.2	8.7	86.3	14.3
1ugh :I	29.6	-18.2	24.2	0.62	40.5	-11.7	36.2	-14.2	0.0	0.0	0.0	0	100.0	0.0	60.9	0.0
1ytf :A	57.1	29.8	84.2	2.09	92.8	20.1	78.4	24.0	50.0	24.8	80.0	1.99	91.6	16.7	74.8	20.0
1ytf :B	54.5	-2.0	92.3	0.97	0.0	-43.5	52.2	-3.8	20.4	-3.5	81.8	0.86	0.0	-76.1	19.6	-6.6
1ytf :D	38.2	4.6	78.8	1.14	76.7	10.3	50.0	6.3	26.5	3.0	78.3	1.13	83.3	6.8	43.9	4.2

TAB. 6 – Comparison of ET and JET performance after clustering for transients of the Huang benchmark

Homodimers															
	JET with clustering - conservation+PC properties								Iterated JET - conservation+PC properties							
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1bnc :A	22.2	7.6	35.6	1.52	87.7	2.3	72.4	3.5	22.2	7.6	35.6	1.52	87.7	2.3	72.4	3.5
1daa :A	57.6	39.2	97.4	3.13	99.3	17.7	86.3	24.4	57.6	39.2	97.4	3.13	99.3	17.7	86.3	24.4
1dpg :A	35.7	20.7	55.6	2.37	91.3	6.3	78.3	9.7	33.3	19.4	56.0	2.39	92.0	5.9	78.3	9.1
1ecp :B	36.6	18.1	96.8	1.98	98.8	17.3	68.5	17.7	34.1	16.9	96.6	1.98	98.8	16.1	67.3	16.5
1efu :A	28.6	17.4	73.3	2.55	95.8	7.0	76.5	10.0	26.0	15.9	74.1	2.58	96.3	6.4	76.1	9.1
1efu :B	31.5	16.3	63.9	2.07	92.0	7.3	73.3	10.1	30.1	16.6	68.8	2.22	93.9	7.4	74.2	10.3
1frp :A	34.1	18.5	82.9	2.18	95.7	11.3	72.3	14.0	32.9	18.2	84.8	2.24	96.4	11.1	72.3	13.8
1fuq :A	61.1	40.6	48.5	2.97	87.4	7.9	83.1	13.2	55.6	37.4	50.0	3.06	89.2	7.3	83.7	12.2
1gdh :A	37.7	18.6	56.5	1.97	88.4	7.5	73.9	10.6	29.0	13.6	54.1	1.89	90.1	5.5	72.6	7.8
1ges :A	32.2	18.7	80.4	2.39	96.0	9.5	74.6	12.6	26.1	15.3	81.1	2.41	96.9	7.7	73.1	10.3
1glq :A	62.8	38.6	69.2	2.59	89.8	14.1	82.6	20.6	53.5	36.7	85.2	3.19	96.6	13.4	85.1	19.6
1gpm :B	22.1	7.5	35.2	1.51	87.6	2.3	72.4	3.5	20.9	7.7	36.7	1.58	89.0	2.3	73.2	3.6
1hur :A	63.6	28.0	59.6	1.79	78.4	14.0	73.5	18.7	63.6	29.5	62.2	1.87	80.7	14.8	75.0	19.7
1hyh :A	14.5	1.6	54.8	1.12	88.6	1.5	52.5	1.6	15.4	2.5	58.1	1.19	89.4	2.3	53.3	2.4
1ids :A	31.8	16.1	84.0	2.02	95.7	11.4	69.2	13.4	31.8	15.5	80.8	1.95	94.6	11.0	68.6	12.8
1ies :B	34.2	7.1	72.2	1.26	82.5	9.5	54.9	8.2	35.5	8.5	75.0	1.31	84.2	11.3	56.4	9.7
1leh :A	37.0	14.9	16.7	1.67	79.5	1.6	75.3	3.0	3.7	-12.9	2.2	0.22	82.0	-1.4	74.2	-2.6
1mas :A	42.3	23.3	56.4	2.22	88.9	7.9	77.1	11.8	40.4	22.3	56.8	2.24	89.5	7.6	77.1	11.3
1mld :A	47.5	31.4	77.8	2.95	95.2	11.2	82.6	16.5	37.3	24.8	78.6	2.98	96.4	8.9	80.8	13.1
1nhk :L	38.9	2.5	35.0	1.07	64.9	1.2	56.4	1.7	44.4	9.9	42.1	1.29	70.3	4.8	61.8	6.5
1nqv :A	38.8	12.7	86.7	1.49	91.7	17.8	60.9	14.8	41.8	14.0	87.5	1.5	91.7	19.5	62.6	16.3
1oro :A	60.0	31.0	57.4	2.07	82.9	11.9	76.5	17.2	57.8	31.9	61.9	2.23	86.3	12.2	78.4	17.7
1osj :A	33.3	15.3	48.9	1.85	87.5	5.5	73.2	8.1	31.8	15.0	50.0	1.89	88.6	5.4	73.6	7.9
1pky :A	18.5	3.6	20.4	1.24	85.8	0.7	74.8	1.2	18.5	3.3	20.0	1.22	85.5	0.7	74.5	1.1
1poy :1	18.9	5.9	23.3	1.46	88.1	1.1	77.1	1.9	18.9	7.2	25.9	1.62	89.7	1.4	78.4	2.3
1qor :A	41.1	27.9	74.2	3.13	95.6	8.7	82.6	13.3	42.9	29.3	75.0	3.16	95.6	9.1	83.1	13.9
1rah :B	45.8	8.7	44.9	1.23	67.9	5.0	59.8	6.3	39.6	10.8	50.0	1.38	77.4	6.2	63.6	7.9
1scu :D	51.9	30.1	95.3	2.38	98.3	20.1	79.7	24.1	53.2	31.3	97.7	2.44	99.2	21.0	80.7	25.1
1scu :E	30.9	13.9	50.0	1.82	88.3	5.3	72.5	7.6	28.4	12.5	48.9	1.78	88.8	4.7	72.2	6.8
1set :A	30.7	18.4	56.1	2.49	93.0	5.3	79.0	8.3	30.7	19.6	62.2	2.76	94.6	5.7	80.2	8.8
1sft :A	51.8	31.3	81.1	2.52	94.3	14.8	80.6	20.1	53.0	34.4	91.7	2.85	97.7	16.3	83.3	22.1
1tph :1	71.4	46.3	69.8	2.84	89.9	15.1	85.4	22.7	69.0	45.1	70.7	2.88	90.7	14.7	85.4	22.1
1xik :A	33.9	19.9	63.6	2.43	93.1	7.1	77.6	10.4	32.3	18.8	62.5	2.39	93.1	6.6	77.2	9.8
2cst :A	16.9	7.3	50.0	1.77	93.4	2.9	71.8	4.1	16.9	8.7	58.3	2.07	95.3	3.4	73.1	4.9
2eip :A	56.3	26.3	23.1	1.87	73.7	3.7	71.5	6.5	50.0	22.3	22.2	1.81	75.4	3.1	72.3	5.5
2hhm :A	7.3	0.1	21.4	1.01	92.8	0.0	74.6	0.0	7.3	0.6	23.1	1.09	93.4	0.2	75.1	0.2
2pcd :B	42.4	13.6	78.3	1.47	86.7	15.4	63.1	14.5	41.2	17.4	92.1	1.73	96.0	19.8	66.9	18.5
2pcd :N	20.1	3.7	90.3	1.23	94.0	10.4	39.7	5.5	20.1	3.2	87.5	1.19	92.0	8.9	39.2	4.7
2pol :A	30.2	12.3	38.8	1.69	85.8	3.7	73.1	5.7	28.6	12.6	40.9	1.79	87.7	3.7	74.2	5.8
3lad :A	38.3	23.2	78.6	2.55	95.3	10.4	77.7	14.3	39.1	25.2	86.5	2.81	97.3	11.2	79.4	15.5
3mde :A	46.8	31.9	85.7	3.14	97.1	12.0	83.3	17.4	46.8	31.9	85.7	3.14	97.1	12.0	83.3	17.4
6gsv :A	51.2	31.5	68.8	2.61	91.7	11.3	81.0	16.6	48.8	31.0	72.4	2.74	93.3	11.1	81.6	16.4
8cat :A	21.7	13.0	74.3	2.51	96.8	5.5	74.6	7.7	19.2	11.5	74.2	2.5	97.2	4.8	74.1	6.8
1gp1 :A	57.1	31.4	34.3	2.22	80.0	5.7	76.5	9.7	61.9	35.4	36.1	2.34	80.0	6.5	77.2	10.9

TAB. 7 – Left : JET with clustering based on conservation and physical-chemical properties (PC). Right : JET iterated 10 times with selection of clustered residues which have been obtained by a consensus of 7 distinguished runs. The proteins listed are all the homodimers in the Huang benchmark.

| | **Heterodimers** | | | | | | | | | | | | | | |
| | **JET with clustering - conservation+PC properties** | | | | | | | | **Iterated JET - conservation+PC properties** | | | | | | | |
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1all :A	46.7	19.9	60.0	1.75	83.7	10.4	71.0	13.7	44.4	20.8	64.5	1.88	87.2	10.9	72.5	14.3
1all :B	66.7	35.6	71.4	2.14	86.7	17.8	80.0	23.7	62.2	32.6	70.0	2.1	86.7	16.3	78.5	21.7
1hcg :A	57.7	31.8	34.9	2.23	80.0	5.9	76.5	10.0	38.5	16.2	27.0	1.73	80.7	3.0	74.1	5.1
1hcg :B	78.9	22.1	51.7	1.39	56.3	13.1	64.7	16.5	78.9	22.1	51.7	1.39	56.3	13.1	64.7	16.5
1luc :A	48.9	30.3	48.9	2.63	88.3	6.9	81.0	11.3	46.7	28.9	48.8	2.63	88.8	6.6	81.0	10.7
1luc :B	67.4	46.4	66.0	3.21	91.0	12.0	86.2	19.1	65.2	44.2	63.8	3.11	90.4	11.4	85.3	18.2
1scu :D	51.9	30.1	95.3	2.38	98.3	20.1	79.7	24.1	53.2	31.3	97.7	2.44	99.2	21.0	80.7	25.1
1scu :E	30.9	13.9	50.0	1.82	88.3	5.3	72.5	7.6	28.4	12.5	48.9	1.78	88.8	4.7	72.2	6.8
1tco :A	33.3	14.2	56.8	1.74	87.7	6.9	70.0	9.3	26.7	11.0	55.6	1.7	89.7	5.3	69.1	7.2
1tco :B	50.0	23.2	75.0	1.86	88.8	15.6	73.2	18.6	43.3	21.9	81.3	2.02	93.3	14.7	73.2	17.6
1tco :C	58.6	14.8	43.6	1.34	63.3	7.1	61.8	9.6	58.6	17.0	45.9	1.41	66.7	8.2	64.0	11.1
1tcr :A	34.6	16.9	58.1	1.95	89.4	7.1	73.1	10.0	36.5	18.8	61.3	2.06	90.2	8.0	74.3	11.2
1tcr :B	49.1	25.9	57.8	2.11	86.5	9.7	76.3	14.1	47.2	29.6	73.5	2.69	93.6	11.1	80.9	16.2
1ubs :A	58.3	41.7	67.7	3.5	93.3	10.0	86.6	16.1	61.1	43.9	68.8	3.55	93.3	10.5	87.1	17.0
1ubs :B	11.8	-2.1	17.6	0.85	85.6	-0.5	70.3	-0.9	11.8	2.0	25.0	1.21	90.8	0.5	74.4	0.8
1wdc :A	48.1	4.4	92.9	1.1	80.0	23.8	53.1	7.4	66.7	5.7	92.3	1.09	70.0	30.9	67.2	9.7
1wdc :B	58.0	20.9	59.2	1.56	75.6	12.7	68.9	15.8	58.0	25.4	67.4	1.78	82.9	15.5	73.5	19.3
1wdc :C	47.7	13.1	45.7	1.38	71.9	6.5	63.9	8.7	45.5	15.4	50.0	1.51	77.5	7.6	66.9	10.2
2pcd :B	42.4	13.6	78.3	1.47	86.7	15.4	63.1	14.5	41.2	17.4	92.1	1.73	96.0	19.8	66.9	18.5
2pcd :N	20.1	3.7	90.3	1.23	94.0	10.4	39.7	5.5	20.1	3.2	87.5	1.19	92.0	8.9	39.2	4.7
8atc :A	42.9	26.0	42.9	2.55	88.4	5.3	80.8	8.8	40.0	25.1	45.2	2.68	90.2	5.1	81.7	8.4
8atc :B	54.9	25.0	73.7	1.83	86.8	16.8	74.0	20.1	56.9	23.8	69.0	1.72	82.9	16.0	72.4	19.1
9atc :A	42.1	21.4	36.4	2.03	83.9	4.7	76.4	7.7	42.1	23.7	41.0	2.29	86.8	5.2	78.8	8.5
9atc :B	48.1	14.7	56.8	1.44	76.3	9.6	65.2	11.6	46.2	16.6	61.5	1.56	81.3	10.8	67.4	13.1

TAB. 8 – Left : JET with clustering based on conservation and physical-chemical properties (PC). Right : JET iterated 10 times with selection of clustered residues which have been obtained by a consensus of 7 distinguished runs. The proteins listed are all the heterodimers in the Huang benchmark

| | **Transients** | | | | | | | | | | | | | | | |
| | **JET with clustering - conservation+PC properties** | | | | | | | | **Iterated JET - conservation+PC properties** | | | | | | | |
pdbCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1apm :E	45.7	34.7	75.0	4.17	96.7	7.6	87.5	12.5	43.5	33.3	76.9	4.28	97.1	7.3	87.5	12.0
1efu :A	28.6	17.4	73.3	2.55	95.8	7.0	76.5	10.0	26.0	15.9	74.1	2.58	96.3	6.4	76.1	9.1
1efu :B	31.5	16.3	63.9	2.07	92.0	7.3	73.3	10.1	30.1	16.6	68.8	2.22	93.9	7.4	74.2	10.3
1g3n :A	11.9	-2.3	36.4	0.84	84.1	-1.7	52.8	-2.0	11.9	-2.3	36.4	0.84	84.1	-1.7	52.8	-2.0
1g3n :B	25.0	-15.0	12.5	0.63	56.3	-3.7	50.0	-6.0	20.8	-16.7	11.1	0.56	58.3	-4.2	50.8	-6.7
1g3n :C	37.8	19.8	45.2	2.1	87.4	5.4	76.7	8.5	32.4	16.2	42.9	1.99	88.1	4.4	76.2	6.9
1got :A	44.2	28.6	44.2	2.84	89.7	5.3	82.6	8.9	41.9	26.6	42.9	2.75	89.7	4.9	82.2	8.3
1got :B	26.6	6.5	49.0	1.32	83.8	3.8	62.6	4.8	21.3	3.5	44.4	1.2	84.4	2.1	61.0	2.6
1got :G	74.4	14.1	91.4	1.23	80.0	40.3	75.9	20.9	62.8	11.1	90.0	1.21	80.0	31.7	67.2	16.4
1k9o :E	64.0	43.5	48.5	3.12	87.5	8.0	83.9	13.5	60.0	44.5	60.0	3.86	92.6	8.2	87.6	13.8
1k9o :I	12.5	-5.6	4.2	0.69	81.5	-0.4	77.4	-0.7	18.8	1.8	6.7	1.1	83.1	0.1	79.2	0.2
1rrp :A	17.8	-0.2	55.2	0.99	81.7	-0.3	46.0	-0.3	16.7	-0.7	53.6	0.96	81.7	-0.9	45.3	-0.8
1rrp :B	53.8	21.4	77.8	1.66	86.4	18.9	71.2	20.1	51.9	24.0	87.1	1.86	93.2	21.1	73.9	22.5
1rrp :C	23.7	1.9	60.0	1.09	80.6	2.4	49.3	2.1	25.0	3.3	63.3	1.15	82.3	4.0	50.7	3.6
1ugh :E	75.9	51.6	56.4	3.13	87.1	11.3	85.1	18.6	75.9	53.5	61.1	3.39	89.4	11.8	87.0	19.3
1ugh :I	18.5	-17.7	20.0	0.51	52.4	-11.4	39.1	-13.9	22.2	4.8	50.0	1.28	85.7	3.1	60.9	3.8
1ytf :A	58.9	28.0	76.7	1.9	88.0	18.9	76.3	22.6	57.1	29.1	82.1	2.04	91.6	19.6	77.7	23.4
1ytf :B	65.9	-1.5	93.5	0.98	0.0	-32.6	63.0	-2.8	61.4	-1.7	93.1	0.97	0.0	-37.0	58.7	-3.2
1ytf :D	38.2	4.6	78.8	1.14	76.7	10.3	50.0	6.3	32.4	3.8	78.6	1.13	80.0	8.6	46.9	5.2

TAB. 9 – Left : JET with clustering based on conservation and physical-chemical properties (PC). Right : JET iterated 10 times with selection of clustered residues which have been obtained by a consensus of 7 distinguished runs. The proteins listed are all the transients in the Huang benchmark

Homodimers					
pdb code	size	20-39	40-59	60-79	80-98
1bnc :A	433	0	774	136	50
1daa :A	277	854	46	2	0
1dpg :A	485	778	86	2	1
1ecp :A	237	403	158	45	47
1efu :A	364	0	53	783	105
1efu :B	282	120	284	59	7
1frp :A	321	267	206	43	26
1fuq :A	456	161	528	229	18
1gdh :A	320	936	12	0	1
1ges :A	448	576	192	113	30
1glq :A	209	582	25	13	40
1gp1 :A	184	71	47	21	27
1gpm :B	499	5	487	169	32
1hur :A	180	163	400	144	159
1hyh :A	297	860	35	0	1
1ids :A	198	462	405	35	28
1ies :A	174	217	258	29	44
1leh :A	364	22	140	28	3
1mas :A	308	497	11	37	2
1mld :A	313	299	384	40	49
1nhk :L	143	59	642	181	1
1nqv :A	154	349	525	12	1
1oro :A	213	416	167	78	37
1osj :A	345	261	679	0	4
1pky :A	464	478	410	24	30
1poy :1	323	708	92	24	5
1qor :A	326	714	207	18	10
1rah :B	153	63	30	33	8
1scu :D	288	220	279	296	105
1scu :E	388	104	512	119	18
1set :A	202	793	29	0	2
1sft :A	421	717	110	2	3
1tph :1	245	222	542	91	66
1xik :A	340	170	17	46	50
2cst :A	411	403	302	3	30
2eip :A	168	195	240	73	35
2hhm :A	272	855	43	7	18
2pcd :B	200	145	80	6	12
2pcd :N	233	86	71	35	22
2pol :A	366	459	153	46	17
3lad :A	472	312	616	20	14
3mde :A	385	748	177	21	24
6gsv :A	217	463	46	108	13
8cat :A	498	160	677	35	41

TAB. 10 – Size (amino-acids number) and number of sequences retrieved by PSI-BLAST for all homodimer proteins in the Huang database. Number of sequences are considered after filtering and they are reported for each sequence identity class.

Heterodimers					
pdb code	size	20-39	40-59	60-79	80-98
1all :A	160	219	14	10	12
1all :B	161	187	40	3	21
1hcg :A	229	598	51	14	11
1hcg :B	51	320	109	8	2
1luc :A	326	231	40	7	11
1luc :B	320	65	59	3	3
1scu :D	288	220	279	296	105
1scu :E	388	104	512	119	18
1tco :A	352	508	165	44	62
1tco :B	169	691	50	19	29
1tco :C	107	81	690	37	37
1tcr :A	202	354	73	27	2
1tcr :B	236	771	59	113	14
1ubs :A	257	617	75	52	97
1ubs :B	389	5	616	41	172
1wdc :A	64	52	235	19	5
1wdc :B	142	665	141	1	12
1wdc :C	152	621	160	7	2
2pcd :B	200	145	80	6	12
2pcd :N	233	86	71	35	22
8atc :A	310	683	158	40	34
8atc :B	146	63	27	38	7
9atc :A	310	697	151	40	23
9atc :B	146	62	27	36	8

Tab. 11 – Size (amino-acids number) and number of sequences retrieved by PSI-BLAST for all heterodimer proteins in the Huang dataset. Number of sequences are considered after filtering and they are reported for each sequence identity class.

Transients					
pdb code	size	20-39	40-59	60-79	80-98
1apm :E	341	355	295	12	124
1efu :A	364	0	53	783	105
1efu :B	282	120	284	59	7
1g3n :A	293	254	471	25	9
1g3n :B	155	896	1	0	8
1g3n :C	233	269	0	0	5
1got :A	326	184	313	207	12
1got :B	339	556	68	44	88
1got :G	58	143	11	14	6
1k9o :E	223	206	406	141	37
1k9o :I	376	845	9	1	14
1rrp :A	204	718	3	56	69
1rrp :B	134	134	114	27	3
1rrp :C	180	709	0	20	87
1ugh :E	223	78	512	9	13
1ugh :I	82	32	0	0	0
1ytf :A	180	100	43	40	73
1ytf :B	46	18	39	6	1
1ytf :D	100	11	51	2	0

TAB. 12 – Size (amino-acids number) and number of sequences retrieved by PSI-BLAST for all transient proteins in the Huang dataset. Number of sequences are considered after filtering and they are reported for each sequence identity class.

pdb code	size	20-39	40-59	60-79	80-98
1all :A	160	219	14	10	12
1all :B	161	187	40	3	21
1apm :E	341	355	295	12	124
1grn :A	191	15	520	224	73
1grn :B	197	472	23	3	10
1leh :A	364	22	140	28	3
1shc :A	195	3	7	28	5
1ycr :A	85	0	14	6	11
2cjk :A	167	316	175	5	5
2ktq :A	528	475	139	0	13
2pol :A	366	459	153	46	17
2ptc :E	220	271	352	178	19

TAB. 13 – Size (amino-acids number) and number of sequences retrieved by PSI-BLAST for all proteins illustrated in the article. Number of sequences are considered after filtering and they are reported for each sequence identity class.

Propensity Values									
W	I	F	L	C	M	V	Y	P	A
2.19	1.41	2.21	1.56	1.42	1.46	1.25	1.76	1.03	0.85
T	H	G	S	Q	N	E	D	K	R
0.7	1.3	0.9	0.8	1.0	0.9	0.7	0.7	0.6	1.1

TAB. 14 – Propensity values describing physical-chemical properties of residues at the interface as estimated in (Nagi and Braun 2007). A value ≥ 1 suggests that a residue most likely belongs to an interface rather than outside of it.

Kanamori dataset - Enzymes
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1a4y	A	B	24	300	28	49	32.9	91.5	46.2	80.8	13	0.28
1acb	E	I	26	154	32	13	66.7	82.8	44.8	80	25.8	0.428
1ava	A	C	6	296	36	65	8.5	89.2	14.3	74.9	10.4	-0.03
1avw	A	B	16	230	25	25	39	90.2	39	83.1	13.9	0.292
1azz	A	D	9	223	39	15	37.5	85.1	18.8	81.1	16.8	0.168
1azz	A	C	18	202	34	31	36.7	85.6	34.6	77.2	18.2	0.218
1bqq	M	T	25	195	28	41	37.9	87.4	47.2	76.1	18.3	0.275
1bvn	P	T	32	289	28	21	60.4	91.2	53.3	86.8	16.2	0.49
1cdk	A	I	17	203	16	15	53.1	92.7	51.5	87.6	13.1	0.452
1clv	A	I	20	277	9	32	38.5	96.9	69	87.9	8.6	0.455
1cse	E	I	26	164	30	9	74.3	84.5	46.4	83	24.5	0.492
1d6r	A	I	21	144	35	18	53.8	80.4	37.5	75.7	25.7	0.301
1dfj	E	I	17	321	22	56	23.3	93.6	43.6	81.2	9.4	0.22
1dhk	A	B	20	364	12	61	24.7	96.8	62.5	84	7	0.322
1dpj	A	B	19	176	10	51	27.1	94.6	65.5	76.2	11.3	0.306
1dtd	A	B	13	184	33	24	35.1	84.8	28.3	77.6	18.1	0.183
1eai	A	C	25	142	47	21	54.3	75.1	34.7	71.1	30.6	0.254
1eja	A	B	19	136	35	17	52.8	79.5	35.2	74.9	26.1	0.279
1f34	A	B	18	245	29	63	22.2	89.4	38.3	74.1	13.2	0.144
1f7z	A	I	23	135	34	10	69.7	79.9	40.4	78.2	28.2	0.407
1fak	H	I	23	144	30	14	62.2	82.8	43.4	79.1	25.1	0.394
1fle	E	I	20	126	44	18	52.6	74.1	31.2	70.2	30.8	0.224
1gl0	E	I	34	136	18	13	72.3	88.3	65.4	84.6	25.9	0.586
1hia	B	I	18	92	48	13	58.1	65.7	27.3	64.3	38.6	0.188
1hx1	A	B	15	290	44	29	34.1	86.8	25.4	80.7	15.6	0.185
1i4o	A	C	9	135	16	9	50	89.4	36	85.2	14.8	0.342
1jiw	I	P	29	334	26	25	53.7	92.8	52.7	87.7	13.3	0.461
1jlt	A	B	29	110	22	48	37.7	83.3	56.9	66.5	24.4	0.236
1jtd	A	B	14	301	27	43	24.6	91.8	34.1	81.8	10.6	0.188
1jtg	A	B	14	240	12	48	22.6	95.2	53.8	80.9	8.3	0.257
1kig	H	I	17	152	30	21	44.7	83.5	36.2	76.8	21.4	0.261
1ppf	E	I	18	139	30	15	54.5	82.2	37.5	77.7	23.8	0.32
1slu	A	B	14	193	36	29	32.6	84.3	28	76.1	18.4	0.159
1stf	E	I	24	143	37	22	52.2	79.4	39.3	73.9	27	0.287
1tbr	H	R	18	153	35	60	23.1	81.4	34	64.3	19.9	0.051
1tmq	A	B	22	307	30	39	36.1	91.1	42.3	82.7	13.1	0.29
1toc	B	R	16	170	31	63	20.3	84.6	34	66.4	16.8	0.058
1ugh	E	I	21	162	12	35	37.5	93.1	63.6	79.6	14.3	0.375
2btc	E	I	26	124	25	13	66.7	83.2	51	79.8	27.1	0.455
2sic	E	I	18	198	30	19	48.6	86.8	37.5	81.5	18.1	0.319
3bth	E	I	21	139	34	13	61.8	80.3	38.2	77.3	26.6	0.353
4htc	H	I	16	125	39	47	25.4	76.2	29.1	62.1	24.2	0.017

TAB. 15 – iJET is evaluated on enzymes of the Kanamori benchmark, for $i = 9$. Columns correspond to chain names, number of true and false positive, number of true and false negatives, evaluation scores, JET coverage, Matthews' correlation coefficient (MCC). TP, TN, FP and FN are defined as the sum of the corresponding values computed on the two chains.

Kanamori dataset - Inhibitors												
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1a4y	A	B	24	300	28	49	32.9	91.5	46.2	80.8	13	0.28
1ava	A	C	6	296	36	65	8.5	89.2	14.3	74.9	10.4	-0.03
1avw	A	B	16	230	25	25	39	90.2	39	83.1	13.9	0.292
1bqq	M	T	25	195	28	41	37.9	87.4	47.2	76.1	18.3	0.275
1d6r	A	I	21	144	35	18	53.8	80.4	37.5	75.7	25.7	0.301
1df9	B	C	18	151	33	18	50	82.1	35.3	76.8	23.2	0.281
1dfj	E	I	17	321	22	56	23.3	93.6	43.6	81.2	9.4	0.22
1dhk	A	B	20	364	12	61	24.7	96.8	62.5	84	7	0.322
1f7z	A	I	23	135	34	10	69.7	79.9	40.4	78.2	28.2	0.407
1fak	H	I	23	144	30	14	62.2	82.8	43.4	79.1	25.1	0.394
1g73	A	D	9	162	52	6	60	75.7	14.8	74.7	26.6	0.2
1jlt	A	B	29	110	22	48	37.7	83.3	56.9	66.5	24.4	0.236
1k9o	E	I	18	331	54	23	43.9	86	25	81.9	16.9	0.235
1ppf	E	I	18	139	30	15	54.5	82.2	37.5	77.7	23.8	0.32
1sgp	E	I	19	103	38	14	57.6	73	33.3	70.1	32.8	0.256
1tbr	H	R	18	153	35	60	23.1	81.4	34	64.3	19.9	0.051
2sic	E	I	18	198	30	19	48.6	86.8	37.5	81.5	18.1	0.319
3bth	E	I	21	139	34	13	61.8	80.3	38.2	77.3	26.6	0.353

Tab. 16 – iJET (with $i = 9$) is evaluated on inhibitors of the Kanamori dataset.

| **Kanamori dataset - Signal transduction** |
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1a2k	A	D	28	170	40	14	66.7	81	41.2	78.6	27	0.4
1agr	A	E	26	304	22	14	65	93.3	54.2	90.2	13.1	0.538
1avz	B	C	18	66	48	7	72	57.9	27.3	60.4	47.5	0.23
1b6c	A	B	16	237	56	24	40	80.9	22.2	76	21.6	0.165
1bdj	A	B	7	109	62	11	38.9	63.7	10.1	61.4	36.5	0.016
1blx	A	B	11	263	53	38	22.4	83.2	17.2	75.1	17.5	0.051
1buh	A	B	12	201	53	19	38.7	79.1	18.5	74.7	22.8	0.132
1c1y	A	B	18	114	44	11	62.1	72.2	29	70.6	33.2	0.263
1c4z	A	D	7	312	49	31	18.4	86.4	12.5	79.9	14	0.041
1cdm	A	B	20	71	18	14	58.8	79.8	52.6	74	30.9	0.374
1cmx	A	B	31	139	34	15	67.4	80.3	47.7	77.6	29.7	0.426
1ds6	A	B	28	204	35	30	48.3	85.4	44.4	78.1	21.2	0.326
1du3	A	D	11	130	39	30	26.8	76.9	22	67.1	23.8	0.035
1du3	A	F	5	126	49	27	15.6	72	9.3	63.3	26.1	-0.102
1eay	A	C	22	89	45	7	75.9	66.4	32.8	68.1	41.1	0.329
1es7	C	D	12	87	45	27	30.8	65.9	21.1	57.9	33.3	-0.03
1es7	A	D	6	102	51	12	33.3	66.7	10.5	63.2	33.3	0
1ev2	A	E	23	198	13	40	36.5	93.8	63.9	80.7	13.1	0.378
1f51	A	E	23	159	47	24	48.9	77.2	32.9	71.9	27.7	0.227
1f5q	A	B	7	299	54	54	11.5	84.7	11.5	73.9	14.7	-0.038
1flt	V	Y	13	105	37	17	43.3	73.9	26	68.6	29.1	0.144
1foe	A	B	34	341	28	33	50.7	92.4	54.8	86	14.2	0.446
1fq1	A	B	17	265	53	21	44.7	83.3	24.3	79.2	19.7	0.218
1fqk	A	B	26	278	26	19	57.8	91.4	50	87.1	14.9	0.463
1g3n	A	C	10	306	46	43	18.9	86.9	17.9	78	13.8	0.057
1g4u	R	S	15	304	41	36	29.4	88.1	26.8	80.6	14.1	0.168
1g4y	B	R	2	110	50	47	4.1	68.8	3.8	53.6	24.9	-0.266
1gcq	B	C	15	34	46	13	53.6	42.5	24.6	45.4	56.5	-0.035
1got	A	B	22	421	43	44	33.3	90.7	33.8	83.6	12.3	0.242
1hcf	B	Y	13	122	36	28	31.7	77.2	26.5	67.8	24.6	0.084
1hcf	A	Y	6	143	49	11	35.3	74.5	10.9	71.3	26.3	0.061
1he1	A	C	26	160	23	24	52	87.4	53.1	79.8	21	0.397
1he8	A	B	10	570	62	22	31.2	90.2	13.9	87.3	10.8	0.148
1i2m	A	B	16	283	32	53	23.2	89.8	33.3	77.9	12.5	0.151
1i4d	A	D	11	237	29	24	31.4	89.1	27.5	82.4	13.3	0.194
1ib1	A	E	30	220	36	37	44.8	85.9	45.5	77.4	20.4	0.309
1ibr	A	B	14	335	41	69	16.9	89.1	25.5	76	12	0.071
1jdp	A	H	0	246	8	20	0	96.9	0	89.8	2.9	-0.049
1jma	A	B	2	231	20	37	5.1	92	9.1	80.3	7.6	-0.037
1k90	A	D	22	366	26	102	17.7	93.4	45.8	75.2	9.3	0.163
1lfd	A	B	19	120	47	10	65.5	71.9	28.8	70.9	33.7	0.281
1lfd	B	C	11	119	58	7	61.1	67.2	15.9	66.7	35.4	0.172
1qav	A	B	16	84	49	26	38.1	63.2	24.6	57.1	37.1	0.011
1qbk	B	C	18	709	45	88	17	94	28.6	84.5	7.3	0.139
1qmz	A	B	12	295	40	61	16.4	88.1	23.1	75.2	12.7	0.052
1rrp	A	B	28	128	34	82	25.5	79	45.2	57.4	22.8	0.052
1tnr	A	R	8	169	45	29	21.6	79	15.1	70.5	21.1	0.005
1tx4	A	B	29	184	36	23	55.8	83.6	44.6	78.3	23.9	0.363
1vrk	A	B	20	82	20	18	52.6	80.4	50	72.9	28.6	0.325
1wq1	G	R	40	259	23	27	59.7	91.8	63.5	85.7	18.1	0.528
1www	V	Y	11	117	38	26	29.7	75.5	22.4	66.7	25.5	0.047
1www	W	Y	7	136	39	13	35	77.7	15.2	73.3	23.6	0.091
1zbd	A	B	22	152	37	30	42.3	80.4	37.3	72.2	24.5	0.217

TAB. 17 – iJET (with $i = 9$) is evaluated on signal transduction proteins of the Kanamori dataset.

Kanamori dataset - Antibodies												
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1a2y	B	C	3	110	63	20	13	63.6	4.5	57.7	33.7	-0.159
1a2y	A	C	2	115	57	14	12.5	66.9	3.4	62.2	31.4	-0.124
1bgx	H	T	10	651	81	76	11.6	88.9	11	80.8	11.1	0.005
1bgx	L	T	7	679	73	60	10.4	90.3	8.8	83.8	9.8	0.007
1bj1	H	W	0	166	70	32	0	70.3	0	61.9	26.1	-0.219
1bzq	A	L	10	123	40	47	17.5	75.5	20	60.5	22.7	-0.073
1dee	D	G	14	140	61	14	50	69.7	18.7	67.2	32.8	0.137
1dqj	B	C	3	198	47	20	13	80.8	6	75	18.7	-0.044
1dqj	A	C	0	200	52	21	0	79.4	0	73.3	19	-0.14
1e6j	H	P	1	280	55	25	3.8	83.6	1.8	77.8	15.5	-0.09
1egj	A	H	0	203	46	18	0	81.5	0	76	17.2	-0.123
1eo8	A	H	3	316	64	30	9.1	83.2	4.5	77.2	16.2	-0.057
1fbi	H	X	8	200	61	24	25	76.6	11.6	71	23.5	0.012
1fc2	C	D	7	148	39	21	25	79.1	15.2	72.1	21.4	0.034
1fe8	A	H	0	221	71	22	0	75.7	0	70.4	22.6	-0.148
1fe8	A	L	0	232	61	26	0	79.2	0	72.7	19.1	-0.145
1fj1	B	F	0	351	24	21	0	93.6	0	88.6	6.1	-0.06
1fj1	A	F	0	342	58	26	0	85.5	0	80.3	13.6	-0.101
1fns	A	H	0	224	56	22	0	80	0	74.2	18.5	-0.134
1fsk	A	C	8	241	50	21	27.6	82.8	13.8	77.8	18.1	0.078
1g9m	G	H	5	335	60	20	20	84.8	7.7	81	15.5	0.031
1jhl	A	H	1	103	67	19	5	60.6	1.5	54.7	35.8	-0.22
1jps	H	T	0	262	39	36	0	87	0	77.7	11.6	-0.125
1jtp	A	L	11	111	56	24	31.4	66.5	16.4	60.4	33.2	-0.017
1mlc	B	E	2	211	48	23	8	81.5	4	75	17.6	-0.078
1ncc	L	N	0	363	61	25	0	85.6	0	80.8	13.6	-0.096
1ncc	H	N	1	346	46	27	3.6	88.3	2.1	82.6	11.2	-0.065
1osp	L	O	0	344	38	18	0	90.1	0	86	9.5	-0.07
1osp	H	O	0	331	47	21	0	87.6	0	83	11.8	-0.086
1qkz	A	H	8	140	57	19	29.6	71.1	12.3	66.1	29	0.005
1t83	B	C	4	238	49	17	19	82.9	7.5	78.6	17.2	0.013
1wej	F	L	1	202	52	13	7.1	79.5	1.9	75.7	19.8	-0.074
1wej	F	H	2	204	51	15	11.8	80	3.8	75.7	19.5	-0.05
2hmi	B	D	0	475	31	24	0	93.9	0	89.6	5.8	-0.054
2jel	H	P	2	167	52	23	8	76.3	3.7	69.3	22.1	-0.115
2vir	B	C	0	289	59	33	0	83	0	75.9	15.5	-0.132

Tab. 18 – iJET (with $i = 9$) is evaluated on antibodies of the Kanamori dataset.

Kanamori dataset - Antigens												
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1adq	A	H	1	261	66	22	4.3	79.8	1.5	74.9	19.1	-0.1
1bgx	H	T	10	651	81	76	11.6	88.9	11	80.8	11.1	0.005
1bgx	L	T	7	679	73	60	10.4	90.3	8.8	83.8	9.8	0.007
1bj1	H	W	0	166	70	32	0	70.3	0	61.9	26.1	-0.219
1bzq	A	L	10	123	40	47	17.5	75.5	20	60.5	22.7	-0.073
1dqj	B	C	3	198	47	20	13	80.8	6	75	18.7	-0.044
1dqj	A	C	0	200	52	21	0	79.4	0	73.3	19	-0.14
1dzb	A	X	8	192	35	32	20	84.6	18.6	74.9	16.1	0.044
1e0o	A	B	19	205	23	22	46.3	89.9	45.2	83.3	15.6	0.359
1e6j	H	P	1	280	55	25	3.8	83.6	1.8	77.8	15.5	-0.09
1e96	A	B	14	199	39	12	53.8	83.6	26.4	80.7	20.1	0.279
1eo8	A	H	3	316	64	30	9.1	83.2	4.5	77.2	16.2	-0.057
1fbi	H	X	8	200	61	24	25	76.6	11.6	71	23.5	0.012
1fe8	A	L	0	232	61	26	0	79.2	0	72.7	19.1	-0.145
1fns	A	H	0	224	56	22	0	80	0	74.2	18.5	-0.134
1fsk	A	C	8	241	50	21	27.6	82.8	13.8	77.8	18.1	0.078
1g9m	G	H	5	335	60	20	20	84.8	7.7	81	15.5	0.031
1g9m	C	G	3	293	38	51	5.6	88.5	7.3	76.9	10.6	-0.067
1hez	A	E	13	153	47	18	41.9	76.5	21.7	71.9	26	0.143
1jhl	A	H	1	103	67	19	5	60.6	1.5	54.7	35.8	-0.22
1jtp	A	L	11	111	56	24	31.4	66.5	16.4	60.4	33.2	-0.017
1mlc	B	E	2	211	48	23	8	81.5	4	75	17.6	-0.078
1mlc	A	E	1	215	51	13	7.1	80.8	1.9	77.1	18.6	-0.067
1ncc	L	N	0	363	61	25	0	85.6	0	80.8	13.6	-0.096
1ncc	H	N	1	346	46	27	3.6	88.3	2.1	82.6	11.2	-0.065
1t83	B	C	4	238	49	17	19	82.9	7.5	78.6	17.2	0.013
1t83	A	C	0	233	54	22	0	81.2	0	75.4	17.5	-0.127
1wej	F	H	2	204	51	15	11.8	80	3.8	75.7	19.5	-0.05
2jel	H	P	2	167	52	23	8	76.3	3.7	69.3	22.1	-0.115
2vir	B	C	0	289	59	33	0	83	0	75.9	15.5	-0.132

TAB. 19 – iJET (with $i = 9$) is evaluated on antigens of the Kanamori dataset.

							Kanamori dataset - Others					
pdbCode	chain	chain	TP	TN	FP	FN	Sen	Spe	PPV	Acc	Cov	MCC
1aip	A	D	7	398	35	17	29.2	91.9	16.7	88.6	9.2	0.163
1aip	A	C	10	372	37	41	19.6	91	21.3	83	10.2	0.109
1ak4	A	D	11	154	57	16	40.7	73	16.2	69.3	28.6	0.096
1aro	L	P	7	577	66	42	14.3	89.7	9.6	84.4	10.5	0.034
1atn	A	D	2	360	46	38	5	88.7	4.2	81.2	10.8	-0.058
1axi	A	B	15	207	30	43	25.9	87.3	33.3	75.3	15.3	0.146
1b41	A	B	14	284	59	35	28.6	82.8	19.2	76	18.6	0.097
1bml	A	C	4	313	52	89	4.3	85.8	7.1	69.2	12.2	-0.122
1bp3	A	B	18	228	25	44	29	90.1	41.9	78.1	13.7	0.222
1cd9	A	B	0	225	44	34	0	83.6	0	74.3	14.5	-0.147
1cic	B	C	0	268	66	20	0	80.2	0	75.7	18.6	-0.117
1d5m	A	C	4	253	32	36	10	88.8	11.1	79.1	11.1	-0.013
1dkg	A	D	9	325	55	38	19.1	85.5	14.1	78.2	15	0.041
1dx5	I	M	8	199	51	24	25	79.6	13.6	73.4	20.9	0.036
1ebd	B	C	7	344	14	19	26.9	96.1	33.3	91.4	5.5	0.254
1efu	A	B	26	401	21	56	31.7	95	55.3	84.7	9.3	0.339
1f60	A	B	36	280	31	47	43.4	90	53.7	80.2	17	0.363
1fak	H	T	5	228	53	19	20.8	81.1	8.6	76.4	19	0.014
1fak	L	T	17	159	41	35	32.7	79.5	29.3	69.8	23	0.117
1gh6	A	B	2	263	38	34	5.6	87.4	5	78.6	11.9	-0.068
1hyr	B	C	0	262	45	28	0	85.3	0	78.2	13.4	-0.119
1hyr	A	C	0	251	45	33	0	84.8	0	76.3	13.7	-0.133
1i1r	A	B	1	303	38	28	3.4	88.9	2.6	82.2	10.5	-0.067
1i5k	A	C	8	47	18	16	33.3	72.3	30.8	61.8	29.2	0.055
1i7w	A	B	34	274	3	64	34.7	98.9	91.9	82.1	9.9	0.495
1icf	A	I	18	120	46	28	39.1	72.3	28.1	65.1	30.2	0.103
1im3	A	D	5	238	32	31	13.9	88.1	13.5	79.4	12.1	0.02
1im9	A	D	0	317	43	30	0	88.1	0	81.3	11	-0.102
1ira	X	Y	11	251	32	77	12.5	88.7	25.6	70.6	11.6	0.016
1itb	A	B	6	245	44	84	6.7	84.8	12	66.2	13.2	-0.108
1j7v	L	R	10	235	40	25	28.6	85.5	20	79	16.1	0.121
1kac	A	B	21	157	47	17	55.3	77	30.9	73.6	28.1	0.261
1kcg	B	C	4	156	52	15	21.1	75	7.1	70.5	24.7	-0.025
1kcg	A	C	2	145	55	24	7.7	72.5	3.5	65	25.2	-0.146
1lpb	A	B	7	289	51	32	17.9	85	12.1	78.1	15.3	0.025
1nrn	H	R	12	121	19	8	60	86.4	38.7	83.1	19.4	0.388
1qo3	A	C	1	250	68	25	3.8	78.6	1.4	73	20.1	-0.116
1sbb	A	B	0	298	44	28	0	87.1	0	80.5	11.9	-0.105
1t7p	A	B	18	504	36	24	42.9	93.3	33.3	89.7	9.3	0.323
1ycs	A	B	15	229	36	20	42.9	86.4	29.4	81.3	17	0.25
2pcc	A	B	7	232	60	23	23.3	79.5	10.4	74.2	20.8	0.02
3c98	A	B	9	457	38	95	8.7	92.3	19.1	77.8	7.8	0.014

TAB. 20 – iJET (with $i = 9$) is evaluated on a list of non-classified proteins (others) of the Kanamori dataset.

Proteins discussed in the text - with clustering															
	JET on conservation								JET on conservation and pc properties							
PDBCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1all :A	46.7	18.4	56.8	1.7	81.4	9.6	69.5	12.7	44.4	20.0	62.5	1.8	86.0	10.5	71.8	13.8
1all :B	51.1	23.0	60.5	1.8	83.3	11.5	72.6	15.3	66.7	36.3	73.2	2.2	87.8	18.1	80.7	24.2
1apm :E	52.2	41.2	85.7	4.8	98.1	9.0	89.8	14.8	43.5	33.7	80.0	4.5	97.6	7.4	87.9	12.1
1grn :A	81.8	54.2	45.0	3.0	82.1	9.7	82.1	16.5	68.2	48.2	51.7	3.4	88.6	8.6	85.5	14.6
1grn :B	65.5	47.2	70.4	3.6	93.2	11.6	87.8	18.6	62.1	43.7	66.7	3.4	92.4	10.7	86.4	17.2
1leh :A	3.8	-10.4	2.6	0.3	84.6	-1.1	76.7	-2.0	37.0	14.9	16.7	1.7	79.5	1.6	75.3	3.0
1shc :A	65.0	39.1	29.5	2.5	79.3	5.2	77.6	9.2	70.0	47.1	35.9	3.1	83.3	6.3	81.8	11.1
1ycr :A	58,8	19,7	37,0	1,5	67,3	6,4	65,2	9,7	76,5	30,1	40,6	1,6	63,5	9,8	66,7	14,8
2cjk :A	40.0	27.6	82.4	3.2	97.1	9.5	82.5	14.1	37.1	24.0	72.2	2.8	95.1	8.2	80.3	12.3
2ktq :A	43.5	32.3	45.5	3.9	93.1	4.3	87.3	7.6	39.1	27.4	39.1	3.3	91.9	3.6	85.8	6.4
2pol :A	25.4	10.5	39.0	1.7	88.2	3.1	73.8	4.8	30.2	13.8	42.2	1.8	87.7	4.1	74.5	6.3
2ptc :E	61.9	44.4	46.4	3.5	89.2	6.7	85.6	11.7	71.4	56.4	62.5	4.8	93.5	8.5	90.6	14.8
average	49.6	28.9	50.1	2.6	86.4	7.1	79.2	11.1	**53.4**	**32.5**	**53.3**	**2.9**	**87.1**	**8.0**	**80.4**	**12.3**

Tab. 21 – JET performance on a set of proteins discussed individually in the text. Predictions based on conservation and on both conservation and physical-chemical signals are evaluated. Bold characters indicate best performance. The evaluations that we present correspond to the best performance over the 10 runs used to evaluate iJET in Table 22, where the "best performance" corresponds to the highest PPV and Acc values.

Proteins discussed in the text - with clustering															
	ET								iJET : conservation and pc properties							
PBDCode	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc	Sen	ScSen	PPV	ScPPV	Spe	ScSpe	Acc	ScAcc
1all :A	42.2	20.1	65.5	1.9	88.4	10.5	72.5	13.8	44.4	20.8	64.5	1.9	87.2	10.9	72.5	14.3
1all :B	40.0	15.6	54.5	1.6	83.3	7.8	68.9	10.4	62.2	32.6	70.0	2.1	86.7	16.3	78.5	21.7
1apm :E	58.7	39.9	56.3	3.1	90.0	8.8	84.4	14.4	43.5	33.3	76.9	4.3	97.1	7.3	87.5	12.0
1grn :A	40.9	23.7	36.0	2.4	87.0	4.2	80.0	7.2	77.3	51.1	44.7	2.9	82.9	9.1	82.1	15.5
1grn :B	34.5	18.8	43.5	2.2	89.0	4.6	78.2	7.4	62.1	43.0	64.3	3.3	91.5	10.6	85.7	17.0
1leh :A	3.7	-15.1	2.0	0.2	79.5	-1.7	72.0	-3.0	3.7	-12.2	2.3	0.2	82.8	-1.3	74.9	-2.4
1shc :A	70.0	50.0	41.2	3.5	86.7	6.7	84.7	11.8	60.0	34.7	27.9	2.4	79.3	4.6	77.1	8.2
1ycr :A	41.2	16.5	41.2	1.7	80.8	5.4	71.0	8.1	70.6	27.1	40.0	1.6	65.4	8.9	66.7	13.4
2cjk :A	54.3	32.4	63.3	2.5	89.2	11.1	80.3	16.5	34.3	21.9	70.6	2.8	95.1	7.5	79.6	11.2
2ktq :A	60.9	41.3	36.4	3.1	85.9	5.5	83.0	9.7	41.3	30.6	45.2	3.9	93.4	4.1	87.3	7.2
2pol :A	46.0	21.7	43.3	1.9	82.1	6.4	73.8	9.9	28.6	12.6	40.9	1.8	87.7	3.7	74.2	5.8
2ptc :E	38.1	28.1	50.0	3.8	94.2	4.2	86.9	7.4	71.4	57.1	65.2	5.0	94.2	8.6	91.3	15.0
average	44.2	24.4	44.4	2.3	86.3	6.1	78.0	9.5	**49.9**	**29.4**	**51.0**	**2.7**	**86.9**	**7.5**	**79.8**	**11.5**

Tab. 22 – iJET is compared to ET on a set of proteins discussed individually in the text. Bold characters indicate best performance.