**­­­­­­SUPPLEMENTARY MATERIALS**

CSM-AB: graph-based antibody-antigen binding affinity prediction

and docking scoring function

Yoochan Myung1,2, Douglas E.V. Pires1,2,3,\*, David B. Ascher1,2,4,\*

1 Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

2 Systems and Computational Biology, Bio21 Institute, University of Melbourne, Melbourne, Victoria, Australia

3 School of Computing and Information Systems, University of Melbourne, Melbourne, Victoria, Australia

4 School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

\*To whom correspondence should be addressed D.B.A. Tel: +61 90354794; Email: [david.ascher@unimelb.edu.au](mailto:david.ascher@unimelb.edu.au). Correspondence may also be addressed to D.E.V.P. [douglas.pires@unimelb.edu.au](mailto:douglas.pires@unimelb.edu.au).

**SUPPLEMENTARY METHODS**

## **Data collection and preparation**

We collected publically available protein-protein binding affinity information (*K*D) from PDBbind (v.2018) (Liu, et al., 2017) and SabDab (Mar/2019) (Dunbar, et al., 2014) databases and corresponding experimentally solved structures from the Protein Data Bank. Among 5,594 protein-protein complexes, 472 antibody-antigen structures including 375 Fab, 82 Nanobody and 12 scFv were identified using Chothia annotation (Dunbar and Deane, 2016) and collected as a training set (Supplementary Fig. S4). Prior to feature generation PDB structures were processed, HETATM molecules and alternative conformations and occupancies removed. Incomplete residues and missing atoms were repaired using the *RepairPDB* module of the FoldX (Schymkowitz, et al., 2005). Interface residues were identified based on a 5 angstrom cutoff distance and were further sorted into antibody-antibody (AbAb), antigen-antigen (AgAg) or antibody-antigen (AbAg).

To assess the performance of the machine learning models in predicting antibody-antigen binding affinity when small perturbations are introduced in the structures as a blind test, the 3D structures and the binding affinity changes (∆∆G) of 689 single-point and 301 multiple-point mutations were collected from mCSM-AB2 and mmCSM-AB data sets. As described in the earlier study (Myung, et al., 2020), the binding affinity and experimentally determined structures of single-point mutations were collected from AB-BIND, SKEMPI2.0 and PROXiMATE. Out of 905 mutations, we removed 216 redundant structures and kept 689 mutations as a single-point mutation blind test set. For the multiple-point mutation blind test set (Myung, et al., 2020), we prepared 301 data points by combining 242 multiple-point mutations collected from AB-BIND, SKEMPI2.0 and PROXiMATE and 59 multiple mutants from Barlow et al. All structures in the mutant datasets were non-redundant at a sequence-level (Supplementary Table S3).

We also assessed the ability of CSM-AB of identifying near-native structures, working as a docking scoring function. For this purpose, blind test sets were compiled. We collected docked poses of Ab-ag complexes from Dockground (Kundrotas, et al., 2018) unbound docking decoy set 2 and ZDOCK benchmark v4 (Hwang, et al., 2010). While Dockground provides one near-native model and 99 false-positive docked poses for each protein-protein complex, we re-evaluated the 100 poses of 28 Ab-ag complexes on their bound structures using DockQ score (Basu and Wallner, 2016). We removed 13 complexes that had no structures available with better or equal to *Acceptable* DockQ-CAPRI quality.

For the ZDOCK benchmark v4, we downloaded a benchmark v4 of ZDOCK3.0.2 IRAD algorithm using six-degree sampling. Ab-ag complexes overlapping with the Dockground data set were excluded. Out of 360,000 docked poses, we filtered the top 1,000 based on Cα RMSD and re-evaluated the predicted Ab-ag complexes using DockQ. Since each Ab-ag complex has different numbers of poses sorted by DockQ-CAPRI quality, we limited the maximum number of *High*, *Acceptable*, *Medium* and *Incorrect* poses to up to 15, 50, 50 and 50 poses, respectively. All docked poses were structurally non-identical to training structures showing the lowest interface RMSD (iRMSD) and ligand RMSD (LRMSD) of 1.88Å and 3.33Å for Dockground, and 0.64Å and 1.6Å for ZDOCK benchmark. The details of datasets used for training and blind test sets (Supplementary Table S3) are available at <http://biosig.unimelb.edu.au/csm_ab/datasets>.

## **Feature engineering**

***Graph-based signatures.*** The graph-based signature is a way to describe physicochemical properties and the geometry of 3D structures by encoding distance patterns labelled based on atom pharmacophores. While earlier approaches focused on the changes of graph-based signature between single amino acid (Myung, et al., 2020; Myung, et al., 2020) or ligand (Pires and Ascher, 2016) and its surrounding residues, CSM-AB was designed to consider all pharmacophores and atomic distance over interface residues. The pharmacophores consist of eight atom label types: Hydrophobic, Positive, Negative, Acceptor, Donor, Aromatic, Sulfur and Neutral. Atomic distances between interface residues and their surrounding residues were counted and represented as a cumulative distribution function for each atom pair type Ab-Ab, Ag-Ag or Ab-Ag, depending on whether they both belong to the antibody, to the antigen or one to the antibody and another to the antigen, respectively.

***Non-covalent interactions.*** Arpeggio (Jubb, et al., 2017) was used to calculate non-covalent interactions across interface residues. Using an in-house script, the interactions only related to interface residues were processed into Ab-Ab, Ag-Ag or Ab-Ag interactions.

***Complementary features.*** The distribution of residues per protein secondary structure type was computed using DSSP (Joosten, et al., 2011) for those occurring at the antibody-antigen interface. Additionally the solvent accessible surface area (SASA) was also calculated using FreeSASA (Mitternacht, 2016). We implemented four features: SASAAb, SASAAg, SASAAbAg, ∆SASAbinding by calculating SASA of the isolated antibody, the isolated antigen, the Ab-ag complex and the difference between SASA of the Ab-ag complex and individual structures.

## **Machine learning**

CSM-AB predictive models were trained and validated via different cross validation schemes, including leave-one-out, 5-fold, 10-fold and 20-fold cross-validation. Different supervised learning algorithms were assessed including, Ensemble methods (Adaboost, Extra Trees, Gradient Boosting, Random Forest), Gaussian Processes, Nearest Neighbors (KNeighbor), Neural network (Multi-layer Perceptron) of Scikit-learn 0.21.1 (Pedregosa, 2011). The number of trees (n\_estimators) was set to 300 for all Ensemble algorithms and default values were set for other parameters (Supplementary Table S4). A bottom-up greedy feature selection procedure was employed to reduce dimensionality and improve the performance of the models in predicting the binding affinity (∆G) and ranking of docked poses. The best performing model (Extra Trees) was selected based on Borda voting which can determine the rank of different metrics such as Pearson's correlation coefficient, RMSE and average docking rankings. The selected final model was deployed on the CSM-AB webserver for both binding affinity prediction and docking pose scoring.

## **Comparative study**

We used standalone scripts to get the binding affinity scores of LISA (Raucci, et al., 2018) and CIPS (Nadalin and Carbone, 2018). For those scores (AP\_DCOMPLEX, AP\_DFIRE2, AP\_PISA, AP\_T2, AP\_dDFIRE, CP\_DDG\_W, CP\_TB, ELE, FIREDOCK, FIREDOCK\_AB, HBOND2, LK\_SOLV, PYDOCK\_TOT, ROSETTADOCK, ZRANK and ZRANK2) compared in the recent study (Guest, et al., 2021) they were calculated via CCharPPI webserver (Moal, et al., 2015).

**Webserver**

The front-end and back-end frameworks of CSM-AB webserver were developed using Materialize-1.0.0 and Flask-1.0.2. For rendering 3D protein structure and Arpeggio interactions, NGLviewer (Rose, et al., 2018) was implemented in the result page of binding affinity prediction and scoring poses.

**SUPPLEMENTARY TABLES**

**Table S1**. Performance comparison of available methods on the binding affinity prediction and docking decoy data sets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Pearson's correlation coefficient**  (RMSE (Kcal/mol)) | | | **Number of *Top1* ranked  near-native poses** |
| **Methods** | **Training** | **Blind test #1**  **(mCSM-AB2)** | **Blind test #2**  **(mmCSM-AB)** | **Blind test #3**  **(Dockground)** |
| **CSM-AB** | **0.40**  **(1.71)** | **0.61**  **(1.68)** | **0.64**  **(1.75)** | **6** |
| CIPS | 0.25\* | 0.23\* | 0.52\* | 6 |
| LISA | 0.09\* | 0.16\* | 0.05\* | 0 |
| AP\_DCOMPLEX | 0.15\* | 0.02\* | 0.11\* | 0 |
| AP\_DFIRE2 | 0.13\* | 0.10\* | 0.16\* | 1 |
| AP\_PISA | 0.26\* | 0.27\* | 0.01\* | 2 |
| AP\_T2 | 0.17\* | 0.07\* | 0.22\* | 1 |
| AP\_dDFIRE | 0.15\* | 0.29\* | 0.31\* | 3 |
| CP\_DDG\_W | 0.09\* | 0.03\* | 0.12\* | 0 |
| CP\_TB | 0.16\* | 0.03\* | 0.30\* | 5 |
| ELE | 0.17\* | 0.21\* | 0.30\* | 3 |
| FIREDOCK | 0.23\* | 0.24\* | 0.60 | 1 |
| FIREDOCK\_AB | 0.23\* | 0.28\* | 0.57 | 0 |
| HBOND2 | 0.04\* | 0.16\* | 0.10\* | 0 |
| LK\_SOLV | 0.08\* | 0.05\* | 0.20\* | 0 |
| PYDOCK\_TOT | 0.24\* | 0.34\* | 0.48\* | 4 |
| ROSETTADOCK | 0.15\* | 0.09\* | 0.28\* | 0 |
| ZRANK | 0.22\* | 0.21\* | 0.50\* | 2 |
| ZRANK2 | 0.25\* | 0.19\* | 0.41\* | 0 |

**\*P-value < 0.05 (two-tailed); The statistical significance of the difference between Pearson's correlations of CSM-AB and 18 tools was evaluated using Fisher's r-to-Z transformation.**

**Table S2**. Performance of available methods on the ZDOCK benchmark v4 data set. Kendall's correlation coefficient was calculated for 19 methods using 13 antibody-antigen complexes, including two nanobody-antigen complexes (1KXQ and 2I25).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Kendall’s correlation coefficient (Rank across 19 tools)** | | | | | | | | | | | | | |
|  | **Bound form** | | | | | | **Unbound form** | | | | | | **All** | |
| **Methods** | **1BJ1** | **1FSK** | **1IQD** | **1KXQ** | **1NCA** | **2JEL** | **1AHW** | **1DQJ** | **1JPS** | **1MLC** | **2FD6** | **2I25** | **Mean** |
| **CSM-AB** | **0.59**  **(1)** | **0.45**  **(3)** | **0.48**  **(1)** | **0.40**  **(1)** | **0.42**  **(2)** | **0.38**  **(2)** | **0.00**  **(19)** | **0.05**  **(15)** | **0.18**  **(7)** | **0.41**  **(1)** | **0.03**  **(17)** | **0.27**  **(3)** | **0.43**  **(1)** |
| **LISA** | -*\** | 0.02 | 0.06 | 0.04 | 0.06 | 0.21 | 0.02 | 0.01 | 0.03 | 0.02 | 0.05 | 0.05 | 0.05 |
| **CIPS** | 0.03 | 0.25 | 0.17 | 0.40 | 0.03 | 0.04 | 0.08 | 0.07 | 0.03 | 0.24 | 0.42 | 0.18 | 0.16 |
| **AP\_DCOMPLEX** | 0.15 | 0.26 | 0.14 | 0.07 | 0.16 | 0.03 | 0.28 | 0.12 | 0.09 | 0.22 | 0.25 | 0.09 | 0.14 |
| **AP\_DFIRE2** | 0.19 | 0.46 | 0.43 | 0.28 | 0.42 | 0.35 | 0.39 | 0.00 | 0.29 | 0.28 | 0.00 | 0.15 | 0.27 |
| **AP\_PISA** | 0.10 | 0.21 | 0.25 | 0.04 | 0.08 | 0.16 | 0.05 | 0.14 | 0.09 | 0.05 | 0.29 | 0.10 | 0.12 |
| **AP\_T2** | 0.15 | 0.24 | 0.04 | 0.16 | 0.12 | 0.03 | 0.11 | 0.13 | 0.09 | 0.07 | 0.46 | 0.01 | 0.13 |
| **AP\_dDFIRE** | 0.13 | 0.42 | 0.36 | 0.32 | 0.36 | 0.30 | 0.37 | 0.09 | 0.24 | 0.33 | 0.07 | 0.03 | 0.24 |
| **CP\_DDG\_W** | 0.17 | 0.14 | 0.23 | 0.10 | 0.03 | 0.01 | 0.01 | 0.18 | 0.14 | 0.35 | 0.11 | 0.31 | 0.17 |
| **CP\_TB** | 0.21 | 0.28 | 0.20 | 0.35 | 0.31 | 0.38 | 0.12 | 0.23 | 0.22 | 0.23 | 0.30 | 0.25 | 0.26 |
| **ELE** | 0.19 | 0.41 | 0.48 | 0.01 | 0.43 | 0.43 | 0.25 | 0.26 | 0.22 | 0.32 | 0.09 | 0.22 | 0.26 |
| **FIREDOCK** | 0.30 | 0.42 | 0.37 | 0.24 | 0.24 | 0.30 | 0.07 | 0.12 | 0.06 | 0.26 | 0.14 | 0.30 | 0.22 |
| **FIREDOCK\_AB** | 0.16 | 0.27 | 0.19 | 0.00 | 0.01 | 0.25 | 0.07 | 0.05 | 0.09 | 0.09 | 0.23 | 0.21 | 0.13 |
| **HBOND2** | 0.21 | 0.36 | 0.35 | 0.20 | 0.22 | 0.32 | 0.31 | 0.04 | 0.27 | 0.13 | 0.02 | 0.01 | 0.20 |
| **LK\_SOLV** | 0.30 | 0.25 | 0.36 | 0.08 | 0.21 | 0.34 | 0.15 | 0.16 | 0.24 | 0.21 | 0.18 | 0.24 | 0.22 |
| **PYDOCK\_TOT** | 0.01 | 0.39 | 0.24 | 0.31 | 0.34 | 0.28 | 0.09 | 0.12 | 0.03 | 0.37 | 0.25 | 0.05 | 0.20 |
| **ROSETTADOCK** | 0.21 | 0.00 | 0.22 | 0.07 | 0.00 | 0.25 | 0.05 | 0.10 | 0.14 | 0.07 | 0.28 | 0.24 | 0.13 |
| **ZRANK** | 0.14 | 0.48 | 0.01 | 0.06 | 0.15 | 0.15 | 0.32 | 0.03 | 0.04 | 0.25 | 0.31 | 0.13 | 0.16 |
| **ZRANK2** | 0.19 | 0.12 | 0.23 | 0.06 | 0.06 | 0.29 | 0.09 | 0.10 | 0.11 | 0.03 | 0.29 | 0.23 | 0.14 |

**\*Unable to get the results on 1BJ1 due to unknown errors.**

**Table S3**. Number of data points for training and blind test. The number of data points for Dockground and Zdock benchmark v4 was calculated by summing docked poses for all complexes. Unique structures were determined by comparing the PDB ID for antibody-antigen complexes of each dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Training data** | **Single-point mutation** | **Multiple-point mutation** | **Dockground** | **ZDOCK benchmark v4**  **(decoys\_bm4\_zd3.0.2\_irad)** |
| Number of data points | 472 | 689 | 301 | 1500 | 1951 |
| Number of Ab-ag complexes | 472 | 49 | 30 | 15 | 13 |
| Number of same Ab-ag templates used for training | - | 47 | 29 | - | 7 |

**Table S4.** Hyperparameters used for the final CSM-AB model.

|  |  |
| --- | --- |
| **Hyperparameter name** | **Value** |
| n\_estimators | 300 |
| max\_depth | None |
| min\_samples\_split | 2 |
| min\_samples\_leaf | 1 |
| max\_features | auto |
| random\_state | 1 |

**SUPPLEMENTARY FIGURES**

Chart, scatter chart

Description automatically generated

**Figure S1**. Predictive performance of CSM-AB under different validation schemes including leave-one-out cross-validation (top-left), 5-fold (top-right), 10-fold (bottom-left) and 20-fold (bottom-right) cross validation, demonstrating consistency and robustness of the method. The red (triangle) and black (circle) data points indicate datasets with and without 10% outliers, respectively. Metrics in red represent performance on the whole data set and in black on the 90% best performing predictions

Chart, scatter chart

Description automatically generated

**Figure S2.** Performance of CSM-AB on predicting *binding affinity (∆G)*, derived from single-point (left)  and multiple mutation (right) data sets, respectively. The red datapoints and metric represent 10% outliers and performance on the 90% predictions. The black datapoints and performance on datasets without 10% outliers is shown in black.

A screenshot of a computer

Description automatically generated with low confidence

**Figure S3.** Outlier distribution per complex for single-point (top) and multiple mutations (bottom). The high mutation number and strong binders accounted for a substantial number of outliers in 3BDY and 3L5X complexes, respectively.

Chart, histogram

Description automatically generated

**Figure S4.** Distribution of experimental binding affinities for the different data sets used in CSM-AB (training and blind tests on left-hand side) and distribution of antibody types.

# References

Basu, S. and Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS One* 2016;11(8):e0161879.

Dunbar, J. and Deane, C.M. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* 2016;32(2):298-300.

Dunbar, J.*, et al.* SAbDab: the structural antibody database. *Nucleic Acids Res* 2014;42(Database issue):D1140-1146.

Guest, J.D.*, et al.* An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure* 2021.

Hwang, H.*, et al.* Protein-protein docking benchmark version 4.0. *Proteins* 2010;78(15):3111-3114.

Joosten, R.P.*, et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res* 2011;39(Database issue):D411-419.

Jubb, H.C.*, et al.* Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol* 2017;429(3):365-371.

Kundrotas, P.J.*, et al.* Dockground: A comprehensive data resource for modeling of protein complexes. *Protein Sci* 2018;27(1):172-181.

Liu, Z.*, et al.* Forging the Basis for Developing Protein-Ligand Interaction Scoring Functions. *Acc Chem Res* 2017;50(2):302-309.

Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* 2016;5:189.

Moal, I.H., Jimenez-Garcia, B. and Fernandez-Recio, J. CCharPPI web server: computational characterization of protein-protein interactions from structure. *Bioinformatics* 2015;31(1):123-125.

Myung, Y., Pires, D.E.V. and Ascher, D.B. mmCSM-AB: guiding rational antibody engineering through multiple point mutations. *Nucleic Acids Res* 2020;48(W1):W125-W131.

Myung, Y.*, et al.* mCSM-AB2: guiding rational antibody design using graph-based signatures. *Bioinformatics* 2020;36(5):1453-1459.

Nadalin, F. and Carbone, A. Protein-protein interaction specificity is captured by contact preferences and interface composition. *Bioinformatics* 2018;34(3):459-468.

Pedregosa, F.a.V., Ga{\"e}l and Gramfort, Alexandre and Michel, Vincent and Thirion, Bertrand and Grisel, Olivier and Blondel, Mathieu and Prettenhofer, Peter and Weiss, Ron and Dubourg, Vincent and others. Scikit-learn: Machine learning in Python. *Journal of machine learning research}* 2011;12:2825-2830.

Pires, D.E. and Ascher, D.B. CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res* 2016;44(W1):W557-561.

Raucci, R., Laine, E. and Carbone, A. Local Interaction Signal Analysis Predicts Protein-Protein Binding Affinity. *Structure* 2018;26(6):905-915 e904.

Rose, A.S.*, et al.* NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* 2018;34(21):3755-3758.

Schymkowitz, J.*, et al.* The FoldX web server: an online force field. *Nucleic Acids Res* 2005;33(Web Server issue):W382-388.