Bioprocessing India 2024
Oral Presentation

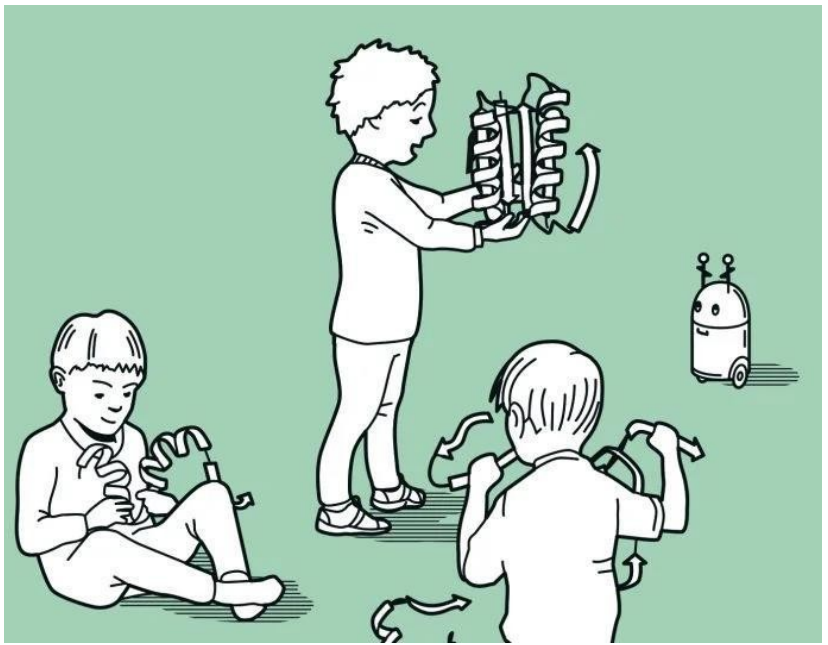
rAbDesFlow: *In silico* engineering of recombinant antibodies for onco-therapeutics

Sowmya Ramaswamy Krishnan[1], Divya Sharma[1], Yasin Nazeer[1], Mayilvahanan Bose[2], Thangarajan Rajkumar[2], Guhan Jayaraman[1], Narayanan Madaboosi[1], M. Michael Gromiha[1]

Department of Biotechnology, IIT Madras
Department of Molecular Oncology, Cancer Institute (WIA)

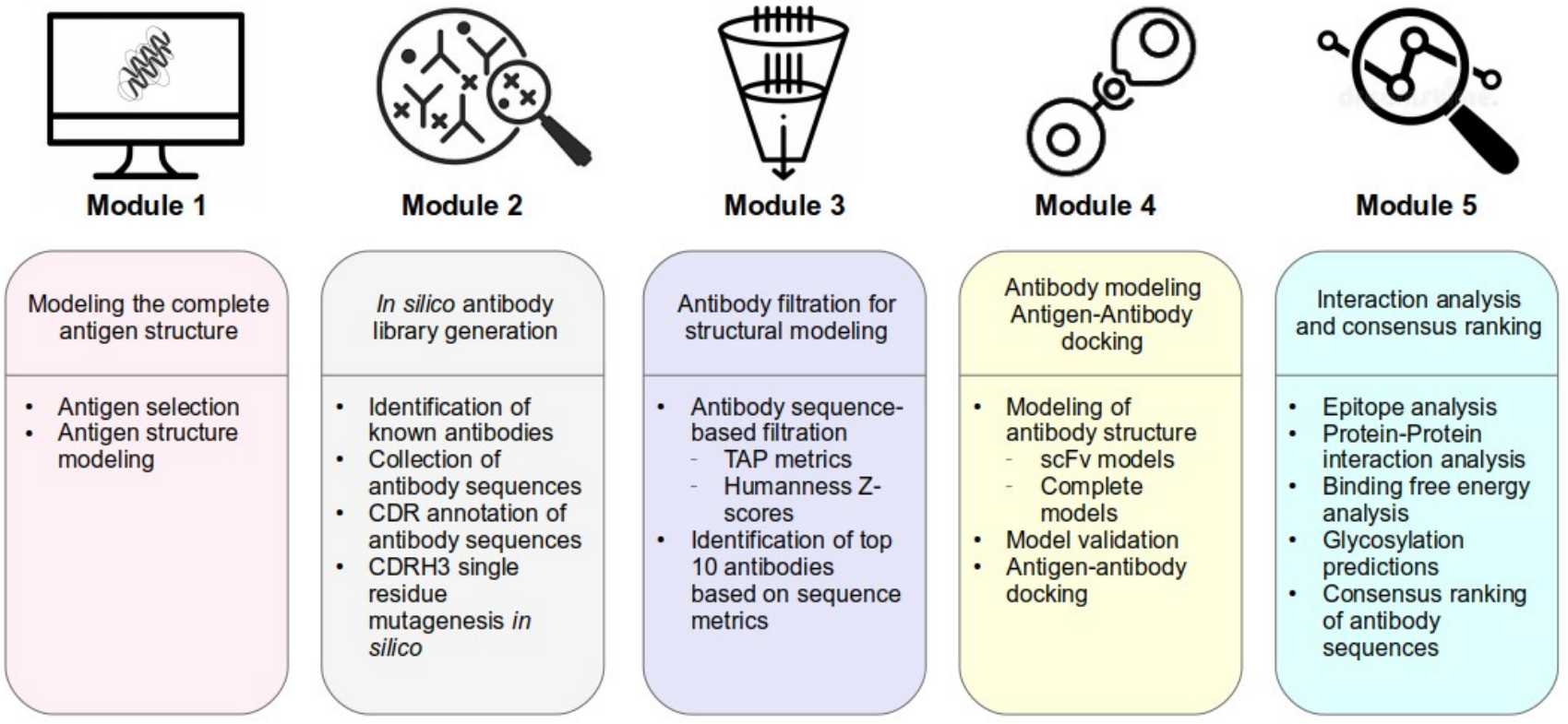
In silico protein engineering

- Proteins can be both therapeutic targets and therapeutics themselves.
- Biologics such as antibodies are proteins that act as selective recognizers of disease-specific antigens to trigger host immune response.
- Emergence of *in silico* protein modelling tools such as AlphaFold and RosettaFold have significantly helped pharma in understanding the sequence-structure-function relationship of proteins.



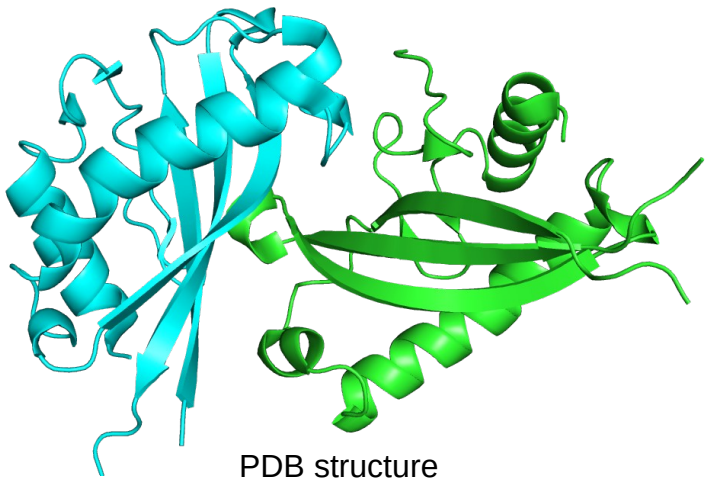
rAbDesFlow pipeline

- Recombinant Antibody Design Workflow (**rAbDesFlow**) is a sequence-based pipeline for designing antibodies for a given antigen.
- Assumes that a few known antibodies with moderate binding affinity to the antigen are already known (not entirely *de novo*).

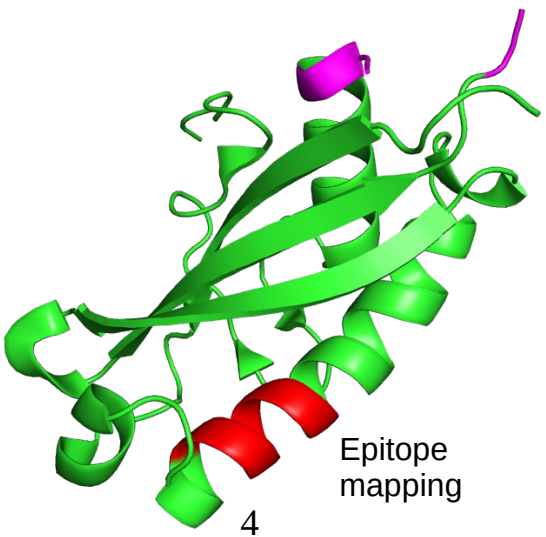


CA125 antigen

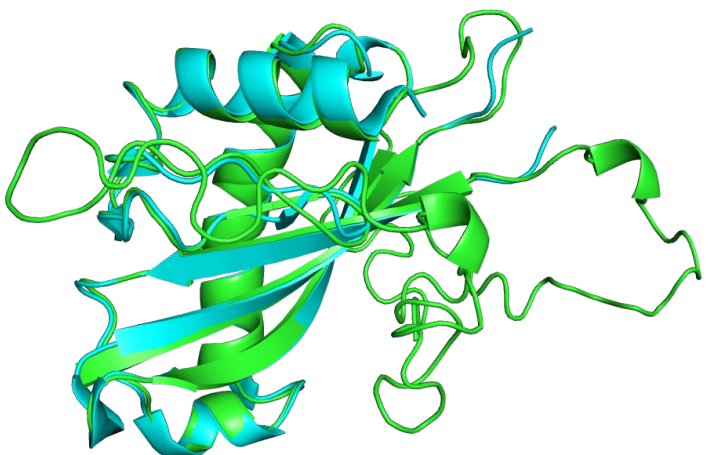
- CA125 – Cancer Antigen 125 or Mucin-16 is an ovarian cancer marker protein (~14,000 residues long)
- Transmembrane protein with a long extracellular region consisting of 60-70 tandem repeats of a SEA (Sperm protein, Enterokinase and Agrin) domain (isoforms vary based on repeat length) made of 196 residues.
- Previously considered SEA domain epitopes for antibody design identified from literature (purple and red) – Examples: OC-125, 5E11, M11 etc.
- UniProt ID: Q8WXI7 (MUC16_HUMAN)
- Human SEA domain partial structure (residues 35 to 160) solved to 1.69 Å resolution (PDB ID: 7SA9)
- Complete structure modelled using ITASSER and AlphaFold-2.0 for antibody design – High RMSD due to long disordered region (Many prolines in N & C terminal regions)



PDB structure



Epitope mapping



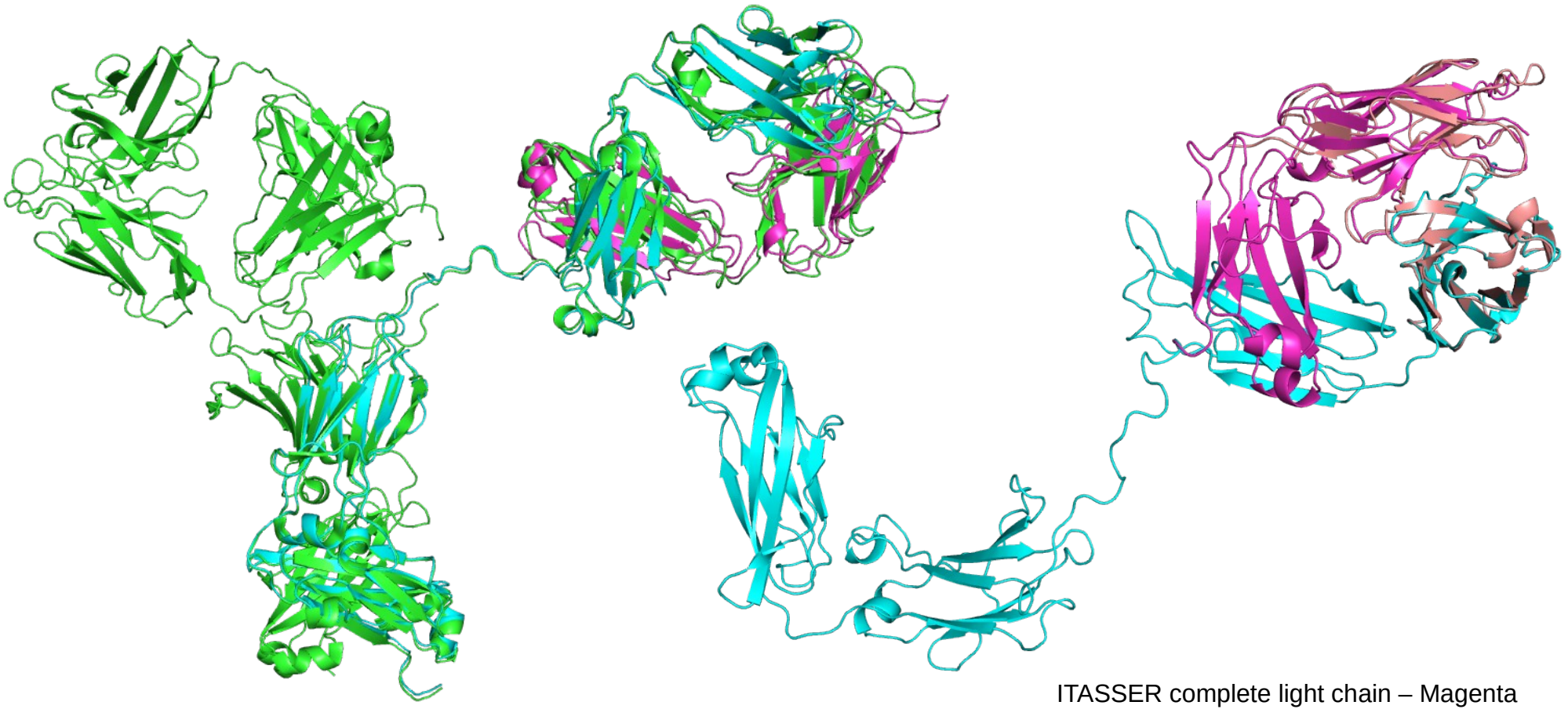
Model overlap

Enumeration of antibody sequences

- The CDR-H3 region (heavy chain variable region 3) – most variable region of existing antibodies – was used to design new antibody sequences.
- Existing CA-125 antibody sequences (Sofituzumab, Abagovomab, anti-CA125-196-14) were collected and CDR regions (L1, L2, L3, H1, H2, H3) were mapped using Kabat numbering scheme using the ANARCI package (Oxford).
- All possible single residue substitutions of the CDR-H3 region sequence was generated using custom program (200 substitutions).
- Therapeutic Antibody Profiler (TAP) metrics from Oxford group were calculated for all 200 substitutions – only antibody sequences passing all 5 metrics (ranges defined for human monoclonal antibodies) were used further (128/200).
- Further the humanness score (H-score) of antibodies calculated from UCL (http://www.bioinf.org.uk/abs/shab/) was calculated.
- Top 10 antibodies with highest heavy chain H-score were considered for detailed analysis.

Information	Value
Final set of mutations	D96G, D96P, D97G, D97P, D97R, D97T, D97Y, Y98G, D99G, M102F
Range of heavy chain H-score	-1.098 to -1.133 (WT = -1.142)
Heavy chain length	118 residues (CDRH3 = SDDYDYGMDY)
Light chain length	112 residues
Fab + Fc region length (1 monomer)	668 residues
Total antibody length	2 monomers = 668*2 = 1336 residues (~150 kDa)

Antibody structure modelling – ScFv region vs complete antibody

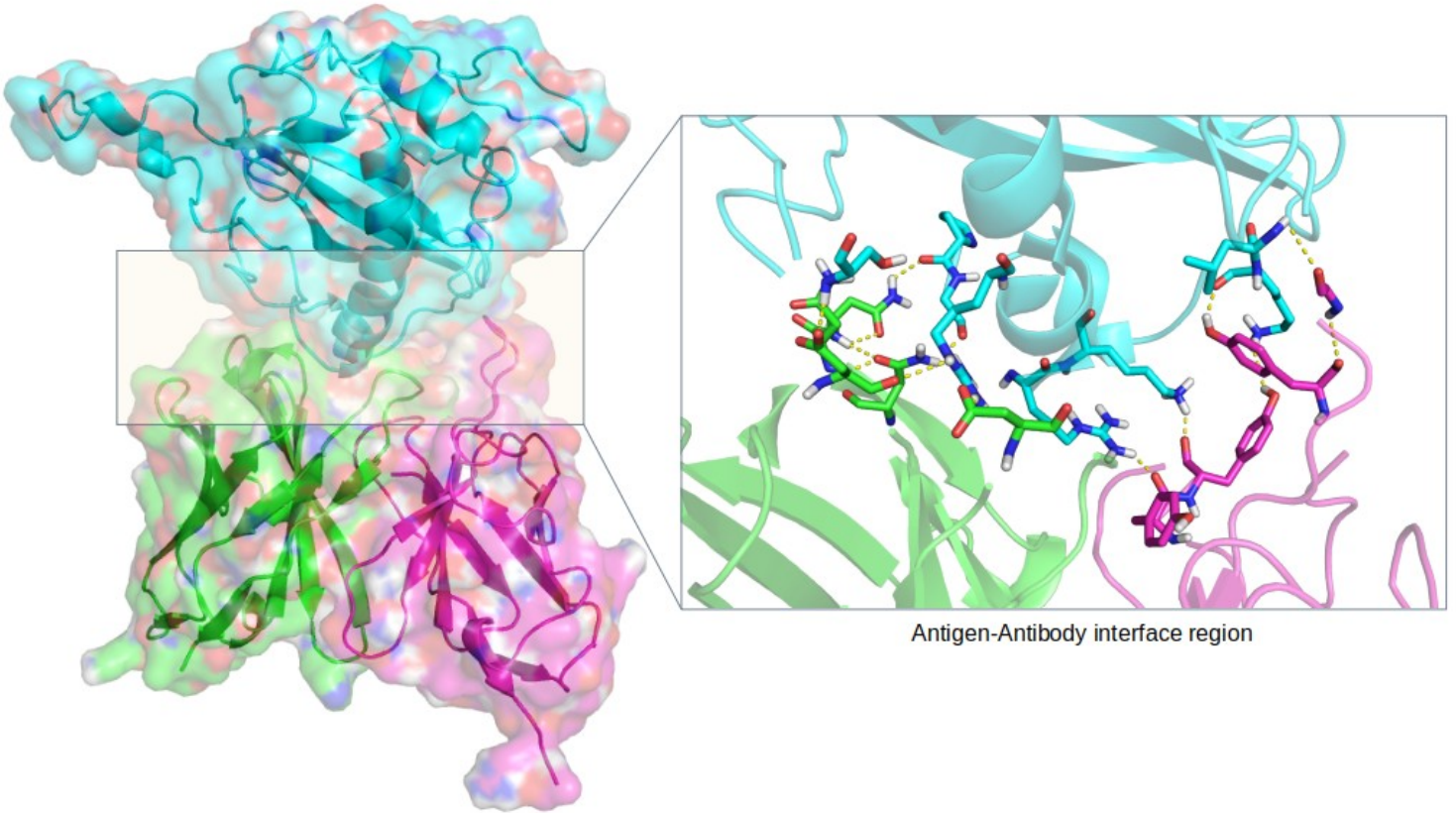


Complete antibody model from ITASSER (Cyan and Magenta)
PDB template antibody structure (Green)

ITASSER complete light chain – Magenta
ITASSER complete heavy chain – Cyan
ROSIE model for ScFv region - Pink

Antigen-Antibody docking with ClusPro

- ClusPro antibody docking mode was used to dock the ScFv model of antibody to SEA domain of CA125 (ITASSER model).
- 30 poses obtained from ClusPro after clustering models.
- Cluster with maximum number of models was identified and cluster representative used as final model.
- Protein-Protein interactions between antigen and antibody analyzed with RING v3.0 standalone tool.
- Interacting residues used to identify epitope and paratope for each antibody.
- Epitopes found to be discontinuous in sequence.

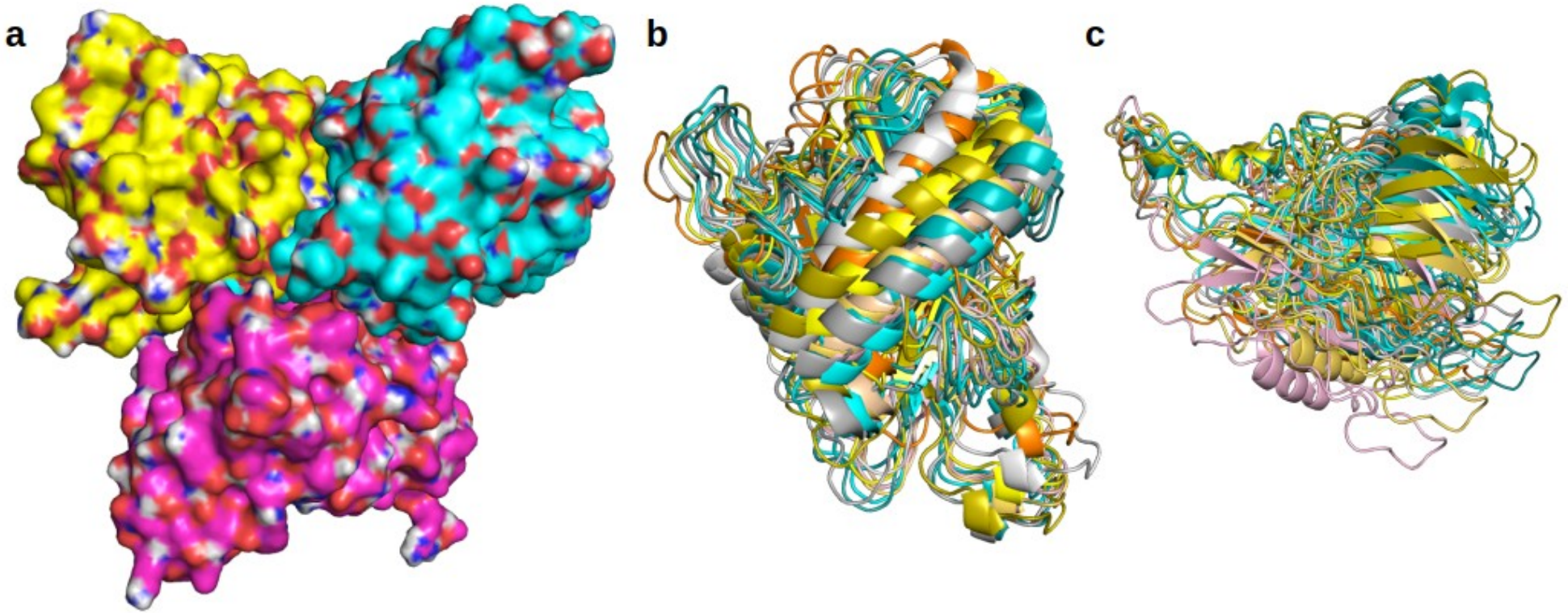


Antigen-Antibody interface region

Wild-type antigen-antibody complex obtained from ClusPro (Antigen in cyan, Antibody light chain in green, Antibody heavy chain in magenta) – Visualization generated using PyMOL

Antigen-Antibody orientations possible

- Two predominant orientations of the Antigen and Antibody were identified by aligning all 30 poses from ClusPro for each antigen-antibody complex and further alignment of antigen orientations between the different antibodies.
- The two predominant antigen orientations among all docking poses compared are highlighted below (b and c).



PPI analysis results – Possible orientation 1

Antibody	Mutation	Hydrogen bonds			Ionic	Cation-Pi	Pi-Pi stacking	Disulfide bridges	van der Waals
		MC-MC	MC-SC	SC-SC					
WT	-	0	11	9	1	0	2	0	13
ab1_26	D96G	0	4	**14**	**2**	0	2	0	13
ab1_33	D96P	0	6	9	1	0	0	0	**19**
ab1_46	D97G	0	11	**10**	0	0	2	0	**21**
ab1_53	D97P	0	5	**11**	1	0	**3**	0	**18**
ab1_55	D97R	0	5	8	0	0	**3**	0	**14**
ab1_57	D97T	0	5	9	0	0	2	0	11
ab1_60	D97Y	0	5	4	1	0	0	0	10
ab1_66	Y98G	0	4	6	0	0	0	0	12
ab1_86	D99G	0	9	**11**	0	0	**3**	0	**19**
ab1_145	M102F	0	7	7	1	0	2	0	**19**

*MC – Mainchain; SC – Sidechain

PPI analysis results – Possible orientation 2

Antibody	Mutation	Hydrogen bonds			Ionic	Cation-Pi	Pi-Pi stacking	Disulfide bridges	van der Waals
		MC-MC	MC-SC	SC-SC					
WT	-	1	7	11	3	0	0	0	17
ab1_26	D96G	0	5	9	3	0	0	0	17
ab1_33	D96P	0	**8**	7	1	0	0	0	**18**
ab1_46	D97G	0	6	**12**	1	0	0	0	**19**
ab1_53	D97P	0	2	6	1	0	0	0	12
ab1_55	D97R	0	4	7	2	0	0	0	14
ab1_57	D97T	1	**9**	**12**	2	0	0	0	**22**
ab1_60	D97Y	0	7	6	2	0	0	0	**19**
ab1_66	Y98G	0	2	9	2	0	0	0	16
ab1_86	D99G	0	4	8	2	0	0	0	15
ab1_145	M102F	0	8	**12**	2	0	0	0	**21**

*MC – Mainchain; SC – Sidechain

Antibody prioritization (Possible orientation 1)

- Binding free energy and Kd of antigen-antibody complex structures were predicted using the PRODIGY server (Bonvin lab).
- **Antibody sequence 46 (D97G) was found to be the best** in comparison witht the wild-type antibody.
- Overall, substitutions involving replacement of aspartic acids (D96, D97 and D99) with G, P, R, T and Y were found to be favorable (8/10).
- Replacement with Glycine is most frequent among the top 10 antibodies considered (4/10).
- Replacement of D97 leads to maximum improvement in antigen-antibody interaction profile among the other residues considered in the CDRH3 region (5/10).
- Top 3 ranked sequences match the previous results obtained.

Antibody	Mutation in CDRH3 region	PRODIGY predicted K_d (M) at 25°C	PRODIGY energy (kcal/mol)	Overall rank
WT	-	2.2E-09	-11.8	4
ab1_26	D96G	2.0E-09	**-11.9**	**3**
ab1_33	D96P	3.7E-09	-11.5	5
ab1_46	D97G	3.9E-10	**-12.8**	**1**
ab1_53	D97P	1.4E-08	-10.7	7
ab1_55	D97R	1.9E-08	-10.5	9
ab1_57	D97T	1.1E-08	-10.8	6
ab1_60	D97Y	5.8E-08	-9.9	11
ab1_66	Y98G	3.4E-08	-10.2	10
ab1_86	D99G	1.1E-09	**-12.2**	**2**
ab1_145	M102F	1.3E-08	-10.7	8

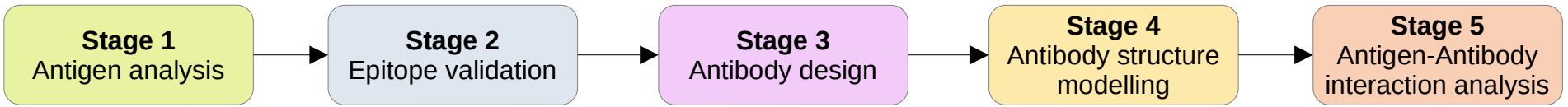
Antibody prioritization (Possible orientation 2)

- Binding free energy and Kd of antigen-antibody complex structures were predicted using the PRODIGY server (Bonvin lab).
- **Antibody sequence 33 (D96P) was found to be the best** in comparison witht the wild-type antibody.
- Overall, substitutions involving replacement of aspartic acids (D96, D97 and D99) with G, P, R, T and Y were found to be favorable (8/10).
- Replacement with Glycine is most frequent among the top 10 antibodies considered (4/10).
- Replacement of D97 leads to maximum improvement in antigen-antibody interaction profile among the other residues considered in the CDRH3 region (5/10).
- Ranking is reshuffled significantly compared to the last results.

Antibody	Mutation in CDRH3 region	PRODIGY predicted K_d (M) at 25°C	PRODIGY energy (kcal/mol)	Overall rank
WT	-	6.8E-11	-13.9	4
ab1_26	D96G	1.7E-09	-11.9	9
ab1_33	D96P	2.8E-11	**-14.4**	**3**
ab1_46	D97G	1.3E-11	**-14.9**	**1**
ab1_53	D97P	2.6E-08	-10.4	11
ab1_55	D97R	1.1E-08	-10.9	10
ab1_57	D97T	1.8E-10	-13.3	5
ab1_60	D97Y	8.3E-10	-12.4	7
ab1_66	Y98G	8.5E-10	-12.4	8
ab1_86	D99G	3.0E-10	-13.0	6
ab1_145	M102F	1.5E-11	**-14.8**	**2**

Phase 2 of rAbDesFlow: Adaptation to under-explored antigens

- Two more cancer-specific human antigens namely, Adipsin and IGFBP-2 are also to be tested with targeted antibodies. However, neither of these antigens have any known antibodies reported in literature.
- Hence, a modified workflow is to be considered to derive the antibody sequences.



- Antigen structure modelling
- B-cell epitope prediction
- Solvent accessibility prediction
- Glycosylation prediction
- Finalization of epitopes

- Gathering PPIs involving antigen and other human proteins (STRING)
- Modelling of interacting proteins
- PPI interface extraction
- Overlap analysis between PPI interface and identified epitopes

- Development of a generative model using Antigen+epitope as input
- Inference of antigen-specific antibodies from (pre-)trained model

- IgFold-derived scFv modelling of designed sequences
- CDR-H3 loop optimization using ROSIE server

- Antigen-antibody complex structure prediction through docking programs
- Interaction analysis
- Binding affinity prediction
- Prioritization of top-scoring antibodies for experimental validation

References

- Dunbar, J.; Deane, C. M. (2016). ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*. 32(2), 298-300. doi: 10.1093/bioinformatics/btv552
- Baran, D.; Pszolla, M. G.; Lapidoth, G. D.; Norn, C.; Dym, O.; Unger, T.; Albeck, S.; Tyka, M. D.; Fleishman, S. J. (2017). Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci. U. S. A*. 114(41), 10900-10905. doi: 10.1073/pnas.1707171114
- Charkhchi, P.; Cybulski, C.; Gronwald, J.; Wong, F. O.; Narod, S. A.; Akbari, M. R. (2020). CA125 and Ovarian Cancer: A Comprehensive Review. *Cancers (Basel)*. 12(12): 3730. doi: 10.3390/cancers12123730
- Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*. 12(1), 7-8. doi: 10.1038/nmeth.3213
- Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. 596(7873), 583-589. doi: 10.1038/s41586-021-03819-2
- Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R. et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 373(6557), 871-876. doi: 10.1126/science.abj8754
- Raybould, M. I. J.; Marks, C.; Lewis, A. P.; Shi, J.; Bujotzek, A.; Taddese, B.; Deane, C. M. (2020). Thera-SAbDab: the Therapeutic Structural Antibody Database. *Nucleic Acids Res*. 48(D1), D383-D388. doi: 10.1093/nar/gkz827
- Raybould, M. I. J.; Marks, C.; Krawczyk, K.; Taddese, B.; Nowak, J.; Lewis, A. P.; Bujotzek, A.; Shi, J.; Deane, C. M. (2019). Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci. U. S. A*. 116(10): 4025-4030. doi: 10.1073/pnas.1810576116
- Ahmad, Z. A.; Yeap, S. K.; Ali, A. M.; Ho, W. Y.; Alitheen, N. B. M.; Hamid, M. (2012). scFv antibody: principles and clinical application. *Clin. Dev. Immunol*. 980250. doi: 10.1155/2012/980250
- Clementel, D.; Conte, A. D.; Monzon, A. M.; Camagni, G. F.; Minervini, G.; Piovesan, D.; Tosatto, S. C. E. (2022). RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res*. 50(W1), W651-W656. doi: 10.1093/nar/gkac365
- Xue, L. C.; Rodrigues, J. P.; Kastritis, P. L.; Bonvin, A. M.; Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*. 32(23), 3676-3678. doi: 10.1093/bioinformatics/btw514
- White, B.; Patterson, M.; Karnwal, S.; Brooks, C. L. (2022). Crystal structure of a human MUC16 SEA domain reveals insight into the nature of the CA125 tumor marker. *Proteins*. 90(5): 1210-1218. doi: 10.1002/prot.26303
- Marcos-Silva, L.; Narimatsu, Y.; Halim, A.; Campos, D.; Yang, Z.; Tarp, M. A.; Pereira, P. J. B.; Mandel, U.; Bennett, E. P.; Vakhrushev, S. Y.; Levery, S. B.; David, L.; Clausen, H. (2014). Characterization of binding epitopes of CA125 monoclonal antibodies. *J. Proteome Res*. 13(7): 3349-59. doi: 10.1021/pr500215g
- Marcos-Silva, L.; Ricardo, S.; Chen, K.; Blixt, O.; Arigi, E.; Pereira, D.; Høgdall, E.; Mandel, U.; Bennett, E. P.; Vakhrushev, S. Y.; David, L.; Clausen, H. (2015). A novel monoclonal antibody to a defined peptide epitope in MUC16. *Glycobiology*. 25(11): 1172-82. doi: 10.1093/glycob/cwv056

Acknowledgements

- **Guide**: Prof. M. Michael Gromiha (Protein Bioinformatics Lab), Department of Biotechnology, IIT Madras
- **IIT Madras**: Ms. Divya Sharma, Prof. Narayanan Madaboosi Srinivasan, Dr. Yasin Nazeer, Prof. Guhan Jayaraman
- **Cancer Institute**: Dr. Mayilvahanan Bose and Prof. Thangarajan Rajkumar
- Anonymous reviewers from *Antibody Therapeutics* (Oxford)
- **Funding agency**: DST-SERB (IPA/2021/000147)

JOURNAL ARTICLE

rAbDesFlow: a novel workflow for computational recombinant antibody design for healthcare engineering

Sowmya Ramaswamy Krishnan, Divya Sharma, Yasin Nazeer, Mayilvahanan Bose, Thangarajan Rajkumar, Guhan Jayaraman, Narayanan Madaboosi ✉, M Michael Gromiha ✉ Author Notes

Antibody Therapeutics, Volume 7, Issue 3, July 2024, Pages 256–265, https://doi.org/10.1093/abt/tbae018

Published: 08 July 2024 **Article history** ▾