

De novo design of single domain antibodies against
CD24, a tumor-associated glycoprotein

Matthew Kidd

Iris Thompson

Research 2020-2021

American Heritage

Abstract

The purpose of this investigation was to use various methods to engineer novel nanobodies against the peptide core of CD24, a tumor-associated glycoprotein. The hypothesis was as follows: If 200 nanobody structures from the protein data bank were docked against CD24, and a docking pose with high shape complementarity underwent computational affinity maturation, then the mutant nanobody will exhibit binding to the CD24 peptide that is more stable during Molecular Dynamics Simulation than the parent nanobody. The model of the peptide was generated using PEPstrMOD, docking was performed with PatchDock, affinity maturation was performed with mmCSM-AB, and MD simulations were run in QwikMD. Ultimately, anti-HIV VHH A12 was determined to have high shape complementarity with the epitope GTSSNS. The mutation K95W was predicted by mmCSM-AB to improve binding affinity by 2.22 kcal/mol, likely due to the formation of a hydrogen bond. After performing a 5ns MD simulations of both the mutant and parent nanobody in complex with CD24, analysis in PyMol showed that while the epitope GTSSNS was unstable during both simulations, the epitope showed a significantly smaller RMSD when bound to the mutant as opposed to the control(RMSD of <7.6Å vs >11.6 Å), and it is likely that more favorable interactions can be designed. Ultimately, this investigation may contribute to the discovery of anti-CD24 single domain antibodies that can be used in cancer therapy, and the methods used could in the future be used to design single domain antibodies against other targets.

Introduction

The purpose of this investigation is to design single domain antibodies that bind to the CD24 peptide via various rational design methods. Then, Molecular Dynamics Simulations will be used to determine if the designed nanobodies can stably bind to the CD24 peptide over time, with the control being the unmutated nanobody in complex with the peptide, and the experimental group being the mutant nanobody in complex with the peptide. The independent variable will be the mutations introduced, and the dependent variable will be the RMSD of the bound peptide epitope over the course of the simulation. It is hoped that this investigation will result in novel cancer drug candidates, as CD24 is a known “don’t eat me” signal, and expand upon current methods of rational antibody design.

CD24 is a sialoglycoprotein usually expressed on the surface of B lymphocytes and differentiating neuroblasts. (Barkal, et.al, 2019) It is connected to the plasma membrane via a GPI anchor, and is composed of a ≈ 30 amino acid peptide to which several carbohydrates are covalently linked. CD24 is overexpressed in 70% of cancers, especially triple-negative breast cancer, ovarian cancer, and pancreatic cancer, all of which are notorious for their low survival rates. According to *Frontiers in Genetics*, after examining breast cancer cells from 1082 patients, “CD24 amplification remained the most significant and independent predictor for worse overall survival(HR=1.88, P=0.015).” (Zhang,et. al, 2019). This is in part because CD24 acts as an anti-phagocytic “don’t eat me” signal, signaling macrophages not to engulf malignant cells when it interacts with macrophage receptor Siglec-10.(Barkal, et.al, 2019) Drugs and toxin conjugates that bind CD24 could thus be used in cancer treatment, and drugs that specifically inhibit the interaction between CD24 and macrophage Siglec-10 could be the basis of novel cancer immunotherapies.

One class of molecules that could be engineered to bind to CD24 are antibodies, proteins generated by the immune system to neutralize foreign pathogens. A typical murine or human IgG antibody is composed of two heavy chains and two light chains, and binds to a given epitope of an antigen via two pairs of variable VH and VL domains. Furthermore, each of those variable domains have three CDR loops that form the antibody's paratope. Single domain antibodies, or nanobodies, are proteins based on the VHH domain of camelid heavy-chain only antibodies. Nanobodies are characterized by having a less hydrophobic framework 2 than human VH domains, and an extended CDR3.(Bannas, et.al, 2019) As drug molecules, nanobodies have several potential advantages over typical mAbs, including higher solubility in the bloodstream, and higher tissue and tumor penetration due to their smaller size. As of now, no nanobodies against CD24 currently exist, making CD24 a worthwhile target for nanobody engineering.

Unfortunately, current methods of antibody engineering have several limitations. The most common methods are immunization and directed evolution. Immunization involves raising antibodies against an antigen of interest in animals, followed by the immortalization of their B cells. Directed evolution involves the generation of a large library of protein variants in phages or yeast cells, followed by screening of the library for binders. Not only are both of these methods expensive and laborious, but the antibodies generated also often have to be optimized for improved binding affinity, lower immunogenicity, and better pharmacokinetic properties.(Cannon, et. al, 2019) Finally, with these methods it can be very difficult to obtain antibodies that bind to a specific epitope.(Sormanni, et. al, 2015) This has relevance for CD24, the different epitopes of which have different function roles. For example, one of the carbohydrates displayed by CD24 is sialyl-T antigen(E, M., et. al, 2009), which contains either an alpha 2-6 or 2-3 linked sialic acid, and likely mediates interactions with Siglec-10.

These limitations have led to a growing interest in Rational Antibody Design, in which antibodies are engineered using computational methods and knowledge of protein structure. In the past decade, most attempts at Rational Antibody Design, beyond antibody humanization, have involved *in silico* affinity maturation, in which a computational model of an antibody binding to its antigen is used to identify mutations that can improve its binding affinity. This has been done both with experimental and theoretical models of protein binding. Researchers at Tianjin University were able to use homology modeling and molecular docking to identify mutations that improved the binding affinity of anti-CD47 nanobody 87.4 fold.(Cheng et. al, 2019) Researchers at Cornell were able to improve the binding of a peptide-binding nanobody by an order of magnitude via homology modeling, docking, MD simulations, and optimization of the nanobody's electrostatic interactions.(Mahajan, et. al, 2018)

A challenge much more difficult than computational affinity maturation is the *de novo* design of antibodies, in which antibodies that bind to the desired epitope are designed from the ground up. Computational tools such as OptMAVEN have been created to guide this design process(Maranas, et.al, 2018). Integrating several different software applications, OptMAVEN positions the antigen of interest in the antibody site, and through CDR recombination and computational affinity maturation, attempts to generate antibodies that bind the desired epitope, and uses a scoring function to rank the designs by binding energy. Unfortunately, this scoring function, like other scoring functions involved in molecular docking, can be very inaccurate, especially when considering charged interactions. Furthermore, even if the designed antibodies show detectable binding, they may suffer from poor stability and other poor pharmacokinetic properties. Finally, OptMAVEN is not optimized for single-domain antibodies. Other antibody design software have similar limitations, prompting a search for other Rational Design methods.

At the University of Freiburg, researchers were trying to engineer haptens (fluorescein) binding nanobodies, something that is difficult with conventional methods of antibody engineering. (Wagner, et.al, 2018) To do this, they grafted the CDRs from the VH of an anti-fluorescein antibody onto a nanobody scaffold and conducted affinity maturation with phage display. As a result of the initial complementarity between the chimeric nanobody and fluorescein, the final nanobody specifically bound to fluorescein with nanomolar affinity. Taking a *de novo* approach, researchers at Cambridge in collaboration with Israeli scientists tried to engineer antibodies that bind the functional epitope of cytokine interleukin-17A. (Nimrod, et. al, 2018) First, they docked several antibody structures from the protein data bank with a homology model of the antigen, and identified an antibody with high shape complementarity. Finally, after analyzing the effects of different mutations on the binding energy and stability of the antibody with MD simulations, they designed a yeast antibody display library and screened it for binders. X-ray crystallography confirmed the resulting antibody specifically bound to the epitope of interest, and their method may be applicable to other targets.

The hypothesis of the experiment is as follows: If 200 nanobody structures from the protein data bank were docked against a model of the CD24 peptide, and a docking pose with high shape complementarity were refined and underwent computational affinity maturation, then the mutant nanobody will exhibit binding to the CD24 peptide that is more stable during MD simulation than the parent nanobody. This hypothesis is substantiated by the fact that similar processes, as described above, have been used to engineer antibodies and nanobodies in the past. Finally, this investigation could at the very least help guide the process of engineering anti-CD24 nanobodies that can be used in cancer therapy, and the methods used could potentially be employed to design antibodies and nanobodies against other targets in the future.

Method

Materials

1. Computer with a NVIDIA GPU
2. Software: VMD, NAMD, PyMol
3. Web Servers: PatchDock, PDB, FireDock, PEPstrMOD, mmCSM-AB

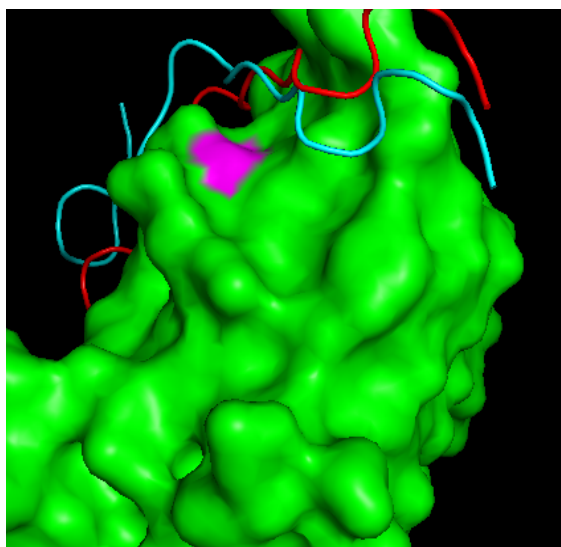
Protocol

1. Using the PEPstrMOD web server, generate a model of the following peptide in a vacuum: SETTTGTSSNSSQSTNSGLAP. Download the model as a PDB file.
2. Go to the Protein Data Bank web server. From the PDB, download over 200 structural models as PDB files containing proteins classified as being VHH domains.
3. Discard any nanobody structures that do not have a resolution of at least 3Å.
4. Extract each nanobody chain from the downloaded files with PyMol.
5. Using the PatchDock web server, dock each nanobody against the CD24 peptide using the antibody-antigen design mode. Specify the nanobody as the receptor and the CD24 peptide as the ligand. Make sure all results are sent to an accessible email address.
6. Analyze the results for each nanobody structure. Download from the server any docking poses that have a PatchDock score of over 8000, a binding surface area over 1000, and an atomic contact energy below -100.
7. Analyze the remaining docking poses in PyMol. Delete any docking poses where most of the contacts that CD24 has with the nanobody are with the framework residues.
8. Using the FireDock web server, perform flexible refinement of the remaining docking poses. Record the atomic contact energies predicted by FireDock afterward.
9. Analyze the refined complexes in PyMol. Compare the RMSD of the peptide in the initial docking pose with the RMSD of the peptide in the refined docking pose for each

complex. Select a docking pose with a change in RMSD of less than 5Å and negative atomic contact energy for computational design.

10. Upload the selected refined complex to the mmCSM-AB web server. Screen for point mutations in the CDRs predicted to improve binding affinity using Prediction Mode.
11. Record all the mutations predicted to improve binding affinity by over 1 kcal/mol. Of the mutations that are proximal to the CD24 peptide, select the mutation or set of mutations that has the largest predicted improvement in binding affinity for further analysis.
12. Upload to QwikMD the peptide nanobody complex selected in step 9. Using the Easy Run feature of QwikMD, set up a NAMD MD simulation specifying that there be an implicit solvent, NaCl ions, a temperature of 27 C, and 5 ns of simulation. There should be 1.242 ns of equilibration followed by 5ns of production simulation.
13. Repeat step 12, but this time before running the simulation introduces the mutation selected in step 11 using VMD.
14. Analyze the production simulation trajectories of both simulation in PyMol, and create a separate PDB file for every 500th frame for both the control and experimental group(each representing the state of the nanobody and peptide after one more nanosecond of production simulation).
15. Align each of the control frames in PyMol with the unmutated complex after equilibration, and record the RMSD of the bound peptide epitope at each nanosecond.
16. Repeat step 15 but with the experimental frames and the mutated complex.

Results



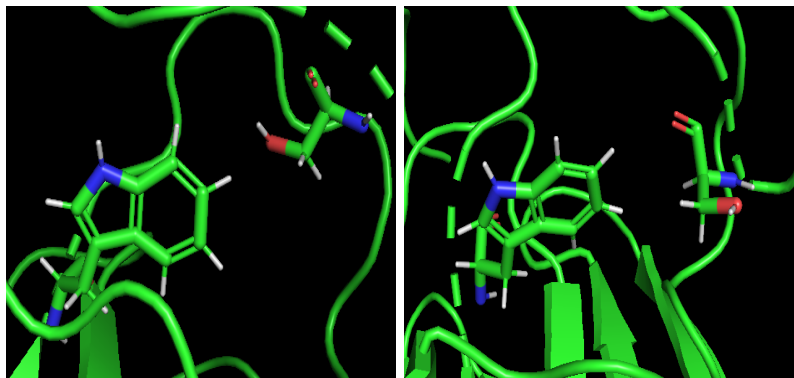
11/10/2020 by Matthew Kidd, PyMol, Refinement of docking pose with FireDock



11/14/2020 by Matthew Kidd, PyMol, Control



11/17/2020 by Matthew Kidd, PyMol, Mutant



0ns(After equilibration)

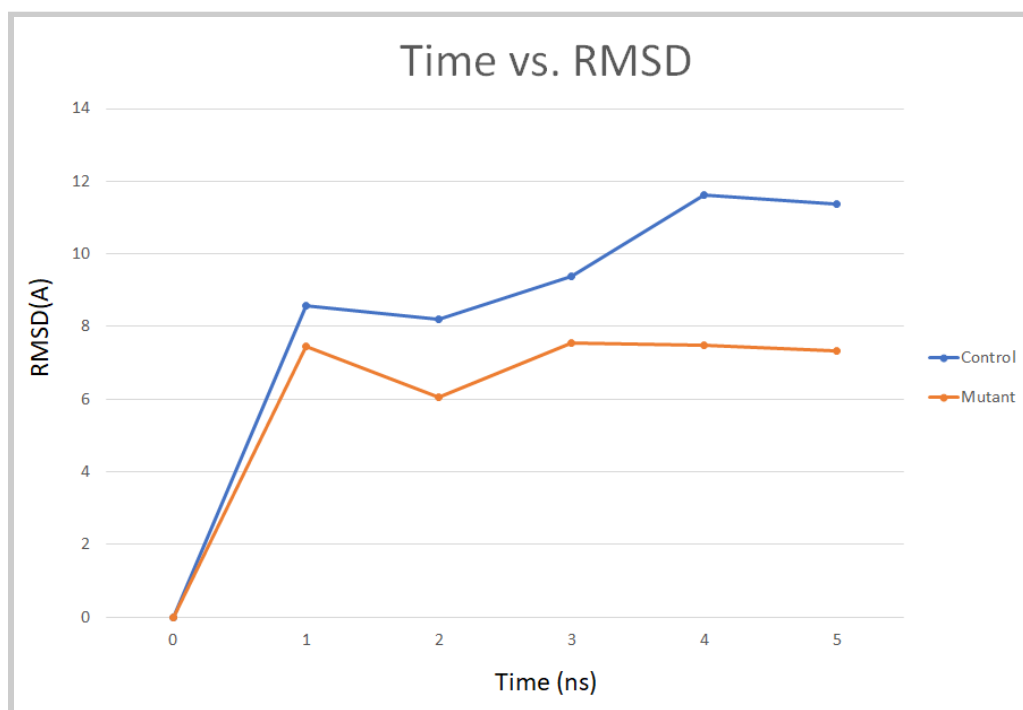
After 5ns

11/17/2020 by Matthew Kidd, Independent variable is the amino acid at position 95 of the Nb, PyMol
Tryptophan(W) at position 95 remains proximal to the serine at position 8 of the peptide, indicating a hydrogen bond

PDB ID	Geometric Score	Interface Area	PatchDock ACE	FireDock ACE	RMSD(A)
1OI0	8262	1252.6	-166.12	-2.67	17.622
3P9W	8040	1096.6	-182.35	-4.28	17.124
3R0M	8092	1090.4	-167.11	-0.82	4.244
4WEU	8530	1021	-125.43	-0.62	26.11

Time(ns)	Control(A)	Mutant(A)
0	0	0
1	8.575	7.443
2	8.202	6.045
3	9.368	7.534
4	11.613	7.491
5	11.369	7.323

Mutations	$\Delta\Delta G(\text{kcal/mol})$
K95W	2.22
P100ED	1.54
S100FD	1.18
A94F	1.07
A94H	1.12
A94N	1.3
P100EN	1.17
A94R	2.39
P100ER	1.9
S100FR	1.77
A94W	2.54
P98W	1.03
P100EW	1.18
P100EY	2.15
L99Y	1.23
A94Y	1.71



Discussion

Conclusion

In conclusion, the results of this experiment support the hypothesis that if several nanobody structures from the protein data bank were docked against CD24, and a docking pose with high shape complementarity underwent computational affinity maturation, then the mutant nanobody would exhibit binding to the CD24 peptide that is more stable during MD simulations than that of the parent nanobody. It was found that the CDR3 of sdAb A12 had high shape complementarity to the CD24 protein core. During the 5ns MD simulation of sdAb A12 in complex with CD24, the bound epitope GTSSNS had a Root Mean Square Deviation over 11 Angstrom in the latter part of the simulation. However, when the Lysine 95 was mutated to tryptophan, the RMSD of GTSSNS did not exceed 7.6 Angstroms, indicating it is more stably bound than the control, albeit likely less stable than is necessary for detectable binding *in vitro*.

Furthermore, while the epitope GTSSNS appeared to unbind from the unmutated nanobody, it appeared to remain bound to the mutant nanobody over the course of the simulation. Upon visual analysis in PyMol, this appears to be because of a hydrogen between the introduced tryptophan residue and a serine of the CD24 peptide, as the two amino acid residues remain proximal to one another during the entire simulation. Overall, the results show that computational affinity maturation of nanobodies with high shape complementarity to a desired epitope is a promising rational design strategy, and the web server used in this investigation, mmCSM-AB, can help guide this design process.

Applications

As cancer continues to kill millions of people around the globe every year, there is a continued search for novel treatments to improve outcomes for patients. Given that CD24 is overexpressed in so many malignancies, stable, non-immunogenic, and soluble protein with high tumor penetrability that specifically targets CD24 could have several applications in cancer therapy. This includes but is not limited to: being part of a Chimeric Antigen receptor for T-Cell Immunotherapy, a cytotoxin conjugate, or serving as a drug itself if it can inhibit interaction between CD24 and Siglec 10. If the top nanobody design generated in this study underwent further affinity maturation *in silico* or *in vitro*, it likely that it could serve as such a protein. Finally, the methods used in this investigation could guide the *de novo* design of single domain antibodies against other disease causing targets.

Limitations

At the beginning of this investigation, I had no experience in antibody design or computational biology. As such, it took several months for me to learn enough about the research topic to generate reliable data. Among the things I had to gain a better understanding were molecular docking, protein structure refinement, and setting up Molecular Dynamics Simulations, the latter of which I had difficulty with for months before I came across QwikMD. Furthermore, the simulations for the experimental and control group not only needed to be set up correctly, but took 2 days each to run on my desktop computer. As such, I was limited in the amount of data I could collect, and was only able to analyze the effect of one mutation during the investigation. Furthermore, I used an implicit solvent model during the simulations, which is faster but not as accurate as an explicit solvent model, and I was only to run the simulations for 5 nanoseconds.

Error Analysis

I am not aware of any random errors during the investigation. Among the systematic errors would be that during the investigation I used an implicit solvent model as opposed to an explicit solvent model which is more accurate, and that I only simulated 5 nanoseconds as opposed to 20 nanoseconds, which is closer to the ideal amount of time that is needed to properly study a phenomenon like protein binding.

Future Research

In the future, I would like to expand on this project in several areas. First of all, I would like to use a better refinement protocol than FireDock for refinement of docking poses. I would also like to try to design a greater number of favorable “hotspot” interactions between the nanobodies and CD24 than I could in this investigation, such that the nanobody designs generated will be ready for study in vitro. I would also like to run the MD simulations for longer(20ns) and use an explicit solvent model. Finally, the CD24 epitope that was targeted in this investigation was GTSSNS, or more broadly the peptide core of CD24. However it is well known that the carbohydrates that are displayed by CD24, including Sialyl-T antigen, help mediate interactions between CD24 and proteins such as macrophage receptor siglec-10, an interaction which allows cancer cells to evade phagocytosis. Thus, in future research I would like to finish designing nanobodies against the peptide core of CD24, design nanobodies that bind to Sialyl-T antigen, and test all of the designs in vitro, or use them as the basis for targeted libraries of protein variants which can be screened for binders to either epitope of this glycoprotein.

References

- Arbabi-Ghahroudi, M. (2017). Camelid Single-Domain Antibodies: Historical Perspective and Future Outlook. *Frontiers in Immunology*.
<https://www.frontiersin.org/articles/10.3389/fimmu.2017.01589/full>
- Bannas, P., Hambach, J., & Koch-Nolte, F. (2017). Nanobodies and Nanobody-Based Human Heavy Chain Antibodies As Antitumor Therapeutics. *Frontiers in Immunology*, (22).
<https://doi.org/10.3389/fimmu.2017.01603>
- Barkal, A. A., Brewer, R. E., Markovic, M., Kowarsky, M., & Barkal, S. A. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature*, 572, 392-396. <https://doi.org/10.1038/s41586-019-1456-0>
- Cannon, D. A., Shan, L., Du, Q., Shirnian, L., & Rickert, K. W. (2019). Experimentally guided computational antibody affinity maturation with de novo docking, modelling and rational design. *PLOS COMPUTATIONAL BIOLOGY*.
<https://doi.org/10.1371/journal.pcbi.1006980>
- Cheng, X., Wang, J., Kang, G., & Hu, M. (2019). Homology Modeling-Based in Silico Affinity Maturation Improves the Affinity of a Nanobody. *International Journal of Molecular Sciences*, 20(17), 4187. <https://doi.org/10.3390/ijms20174187>
- Chowdhury, S., Castro, S., Coker, C., Hinchcliffe, T. E., & Arpaia, N. (2019). Programmable bacteria induce durable tumor regression and systemic antitumor immunity. *Nat Med*, 25(7), 1057-1063. <https://doi.org/10.1038/s41591-019-0498-z>
- E, M., X, Z., X, S., Y, L., M, K., M, F., & PG, W. (2008). Analysis of Recombinant CD24 Glycans by MALDI-TOF-MS Reveals Prevalence of Sialyl-T Antigen. *American Journal of Biomedical Sciences*, 1(1), 1-11. <https://doi.org/10.5099/aj090100001>

- Fleishman, S. J., Corn, J. E., Strautch, E.-M., & Whitehead, T. A. (2011). Hotspot-Centric De Novo Design of Protein Binders. *Journal of Molecular Biology*, 413(5), 1047-1062.
<https://doi.org/10.1016/j.jmb.2011.09.001>
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- Karanicolas, J., Corn, J. E., Chen, I., Spiegel, C. P., & Liu, D. R. (2011). A De Novo Protein Binding Pair By Computational Design and Directed Evolution. *Molecular Cell*, 42(2), 250-260. <https://doi.org/10.1016/j.molcel.2011.03.010>
- Mahajan, S. P., Meksiriporn, B., Waraho-Zhmayev, D., Weyant, K. B., & Kocer, I. (2018). Computational affinity maturation of camelid single-domain intrabodies against the nonamyloid component of alpha-synuclein. *Sci Rep*.
<https://doi.org/10.1038/s41598-018-35464-7>
- Maranas, C. D., Allan, M. F., & Chowdhury, R. D. (2018). OptMAVEN-2.0: De novo Design of Variable Antibody Regions against Targeted Antigen Epitopes. *Antibodies*, 7(3).
<https://doi.org/10.3390/antib7030023>
- Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., & Wolfson, H. J. (2008). FireDock: a web server for fast interaction refinement in molecular docking†. *Nucleic Acids Research*, W229-W232. <https://doi.org/10.1093/nar/gkn186>
- Myung, Y., Pires, D. E. V., & Asher, D. B. (2020). mmCSM-AB: guiding rational antibody engineering through multiple point mutations. *Oxford Academic Nucleic Acids Research*, 48(W1), W125-W131. <https://doi.org/10.1093/nar/gkaa389>
- Nimrod, G. (2018). Computational Design of Epitope Specific Functional Antibodies. *Cell Reports*, 25(8), 2121-2131. <https://doi.org/10.1016/j.celrep.2018.10.081>

- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., & Tajkhorshid, E. (2005). Scalable Molecular Dynamics with NAMD. *J Comput Chem*, 26(16), 1781-1802.
<https://doi.org/10.1002%2Fjcc.20289>
- Ribeiro, J. V., Bernardi, R. C., Rudack, T., Stone, J. E., Phillis, J. C., Freddolino, P. L., & Schulten, K. (2016). QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. *Nature Research*. <https://doi.org/10.1038/srep26536>
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, W363-W367. <https://doi.org/10.1093/nar/gki481>
- Singh, S., Singh, H., Tuknait, A., & Chaudhary, K. (2015). PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol Direct*.
<https://doi.org/10.1186%2Fs13062-015-0103-4>
- Sormanni, P., Aprile, F. A., & Vendruscolo, M. (2015). Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *PNAS*, 112(32), 9902-9907.
<https://doi.org/10.1073/pnas.1422401112>
- Wagner, H. J., Wherle, S., Weiss, E., & Cavallari, M. (2018). A Two-Step Approach for the Design and Generation of Nanobodies. *International Journal of Molecular Sciences*, 19(11). <https://doi.org/10.3390/ijms19113444>
- Yang, E. Y., & Shah, K. (2020). Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Frontiers in Oncology*, (23). <https://doi.org/10.3389/fonc.2020.01182>
- Zhang, P., Zheng, P., & Liu, Y. (2019). Amplification of the CD24 Gene Is an Independent Predictor for Poor Prognosis of Breast Cancer. *Frontiers in genetics*.
<https://doi.org/10.3389/fgene.2019.00560>