

NOVEL COMPUTATIONAL EPIGENETIC-DRIVEN THERAPEUTIC TARGETING

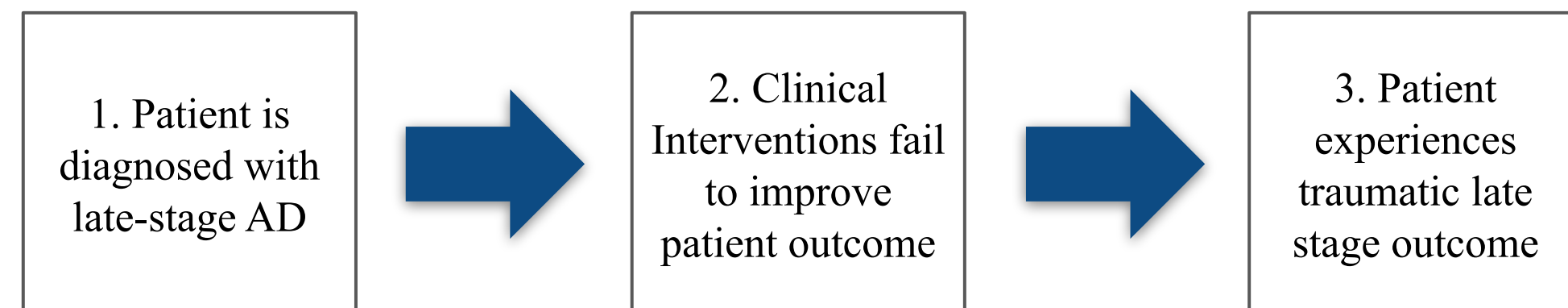
Modulating Tau Pathology: Identifying a Novel Inhibitor for the Suv39h1 Protein Using a Multi-Omic, Attention-Based Neural Network Approach to Treat Alzheimer's Disease

Om Kherde, Reedy High School, Frisco, TX, United States of America

INTRODUCTION

Problems

- Current **computational epigenetic drug discovery techniques** have not **pinpointed specific genetic alterations** that exacerbate diseases like Alzheimer's Disease (AD).
- Therapeutics that target Alzheimer's Disease have seen some of the **lowest failure rates out of any other disease**, and the disease **affects nearly 500,000 additional people every year**.
- An **efficient methodology** has not been developed to capture potential drug target leads in drug discovery through epigenetic mechanisms.

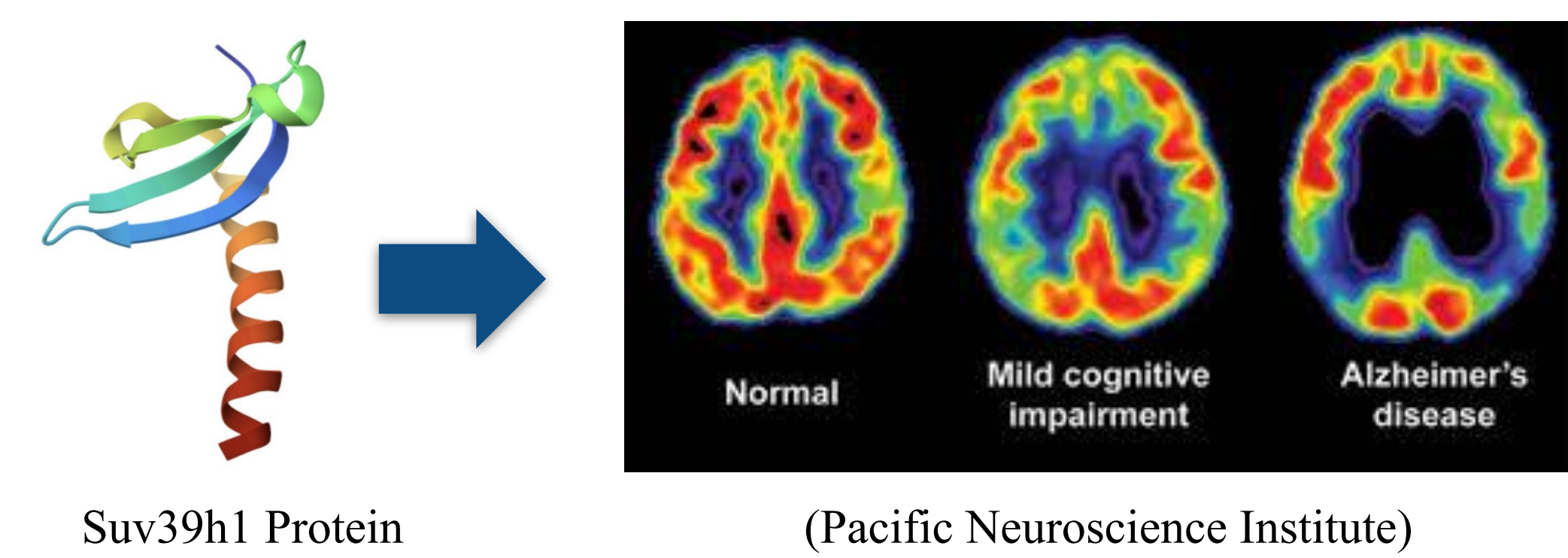


Purpose

- This study aims to **develop a novel Computer Aided Drug Design (CADD) approach** to experimentally pinpoint and validate Suv39h1's involvement in AD.

What is Suv39h1?

- Epigenetic factors like **Euchromatic histone-lysine N-methyltransferase 1 (EHMT1)** and **EHMT2** have already shown to inhibit Amyloid- β aggregation in AD (DeTure and Dickinson, 2019). These genes are located on Histone 3 Lysine 9 (H3K9), the same histone that Suv39h1 modifies.
- Suv39h1 is a gene that codes for evolutionarily-conserved protein containing an **N-terminal chromodomain** and a **C-terminal SET domain**.
- The **protein** that Suv39h1 encodes acts as a histone methyltransferase that **trimethylates lysine 9** of histone H3, which results in **transcriptional gene silencing (DNA Modulation and Histone Methylation)**.
- DNA and Histone Methylation** modulation are crucial in Alzheimer's disease (AD) development.
- Suv39h1 creates gradients for **histone methylation** on histone 3 (h3k9me2)



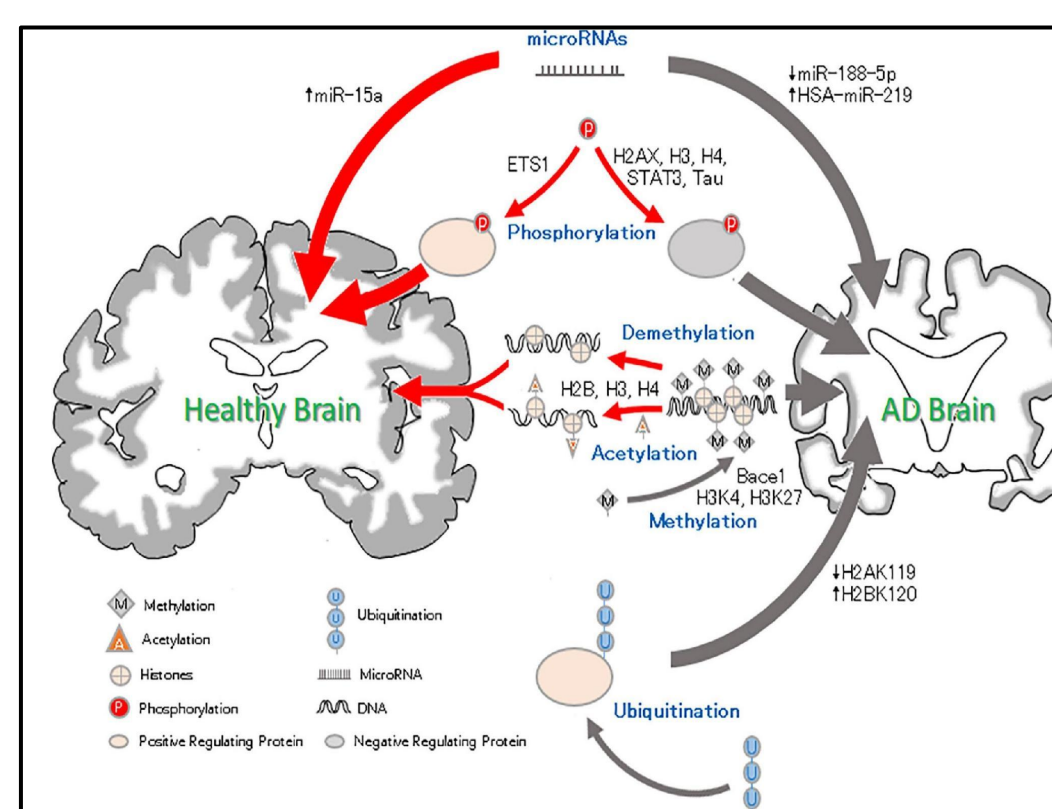
RESEARCH OBJECTIVES

Research Question

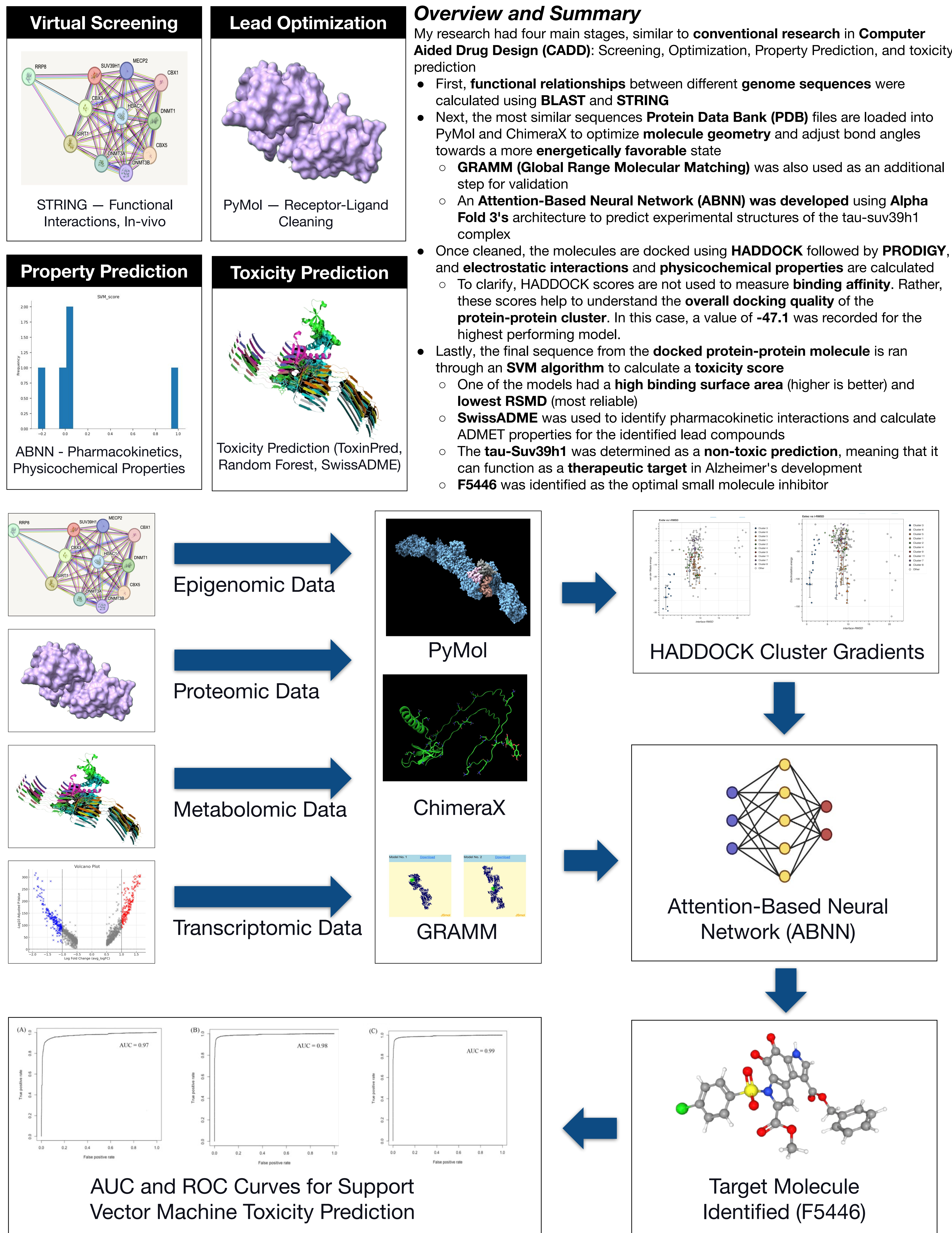
Can Suv39h1 be used as a therapeutic target in Alzheimer's disease, and what is the potential of novel inhibitors in modulating its role in neurodegeneration?

Claim

Suv39h1 is associated with Alzheimer's Disease through histone methylation and DNA modulation. This research develops a novel computational methodology to exponentially improve the efficiency of the Computer Aided Drug Design (CADD).



SCIENTIFIC METHODOLOGY



PROCEDURE

Step 1: Virtual Screening (Target Identification)

- To find a target, we first need to browse through associated mechanisms in Alzheimer's Disease, in which we find **histone methylation** and DNA alteration as a cause of AD development.
- One of these molecules is **EHMT1** (Euchromatic histone-lysine N-methyltransferase 1). In fact, inhibition of EHMT1 rescues **synaptic and cognitive activity in diseased mouse models** (Zheng, et. al, 2019)
- We can then run a **sequence alignment process** using BLAST and STRING to find functional relationships. For STRING, use a **Protein-Protein Interaction (PPI) enrichment value** of $1.17 \cdot 10^{-11}$ for the NETWORK analysis. The PPI Enrichment value is used to determine whether a set of **protein interactions** within a network result **from chance or through functional relationships**. If interactions are caused by functional associations, a **low p-value** is produced ($p < 0.05$).
 - Thus, there is a **very strong** functional relationship within the **Suv39h1 translated protein**.
 - For BLAST, configure the **Blast Protein Server** using the Suv39h1 genome information from the **NCBI database**.
- Thus, we have found a target (Suv39h1)
- Differential Gene Expression** from the Single-cell and Spatial RNA-Seq Database for Alzheimer's Disease (ssRNA) was ran, and p-value, and log fold change was used to determine whether the gene was significantly **upregulated or downregulated in Alzheimer's disease** compared to control samples in mouse models (**in-vivo component**).

Step 2: Molecular Docking (Lead Optimization)

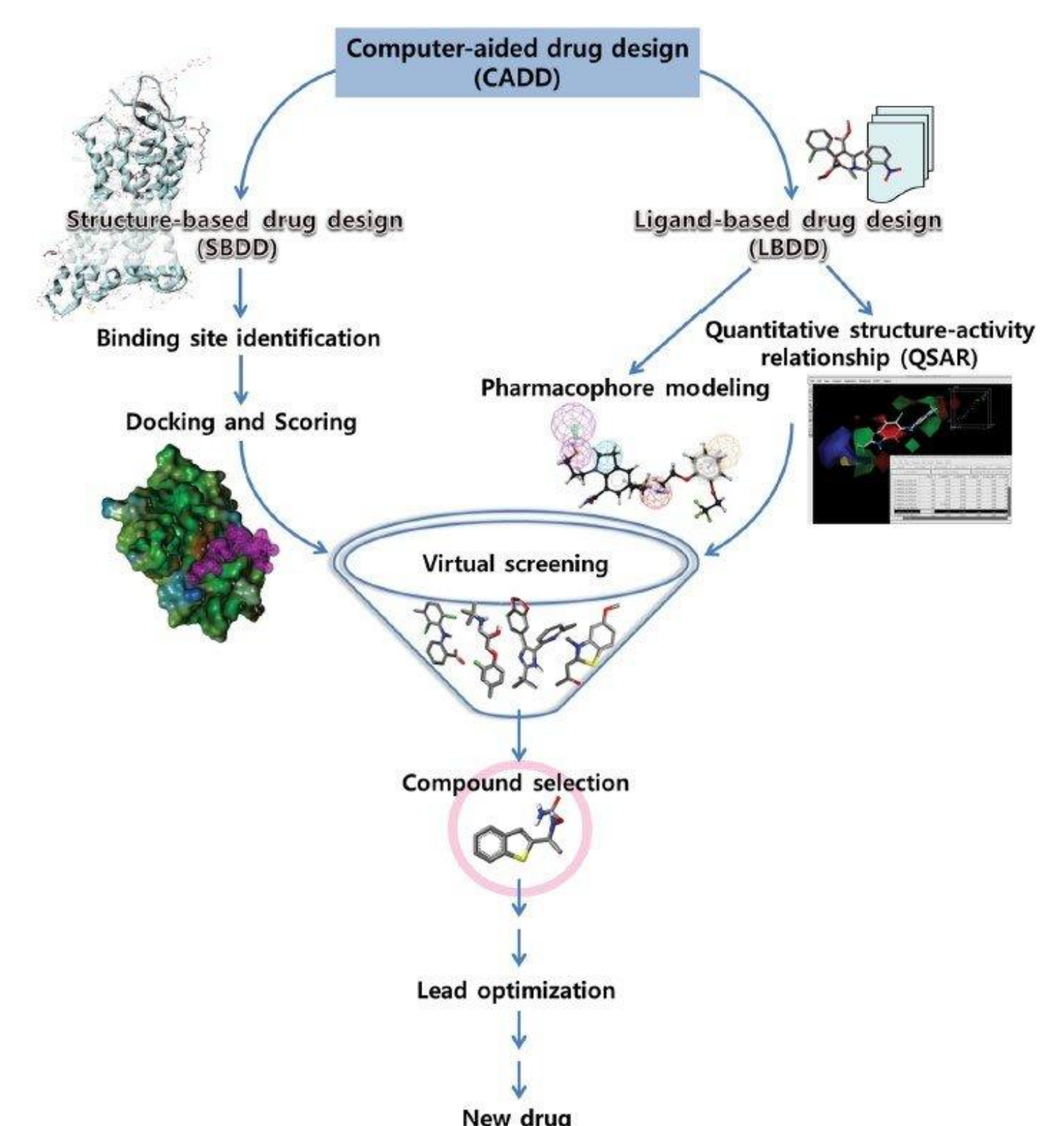
- Now that we have established our target, we will use **HADDOCK** to parse through the **123 possible clusters** and select the **top 10 docking configurations**.
 - Selection Criteria: **HADDOCK score** (lower is better), **Electrostatic Energy**, **Desolvation Energy**, and **Cluster Size**.
- We download the best structure (top-most), and load the structure into **PyMOL** and **ChimeraX**
- We can use the **Wizard tool** to "clean" the structure and optimize **molecular geometry** by adding hydrogens, optimizing hydrogen bonds, and removing water molecules and ligands.
- Save the **cleaned structure** as a multi-model PDB file.
- Use the **GRAMM tool** to **validate PyMOL** and ChimeraX findings
- An **Attention-Based Neural Network (ABNN)** was developed to predict experimental protein folding of the Suv39h1-F5446 complex, relying on Alpha Fold 3's Deep Learning integration.

Step 3: Property Prediction

- Upload the **multi-model PDB file** into the **PRODIGY network**, set temperature to **25 degrees Celsius**, select chain A for interactor 1, and select chain B for interactor 2.
- Run **PRODIGY tool** to calculate binding affinity (in kcal/mol) and ΔG .
- Extract physicochemical properties from previous HADDOCK job and determine **Van Der Waals interactions**, **RMSD**, **molecular weight**, and other properties (Lipinski's rule of five)

Step 4: Toxicity Prediction

- Lastly, run the sequence in the **ToxinPred** web server, use the **first 40 nucleotides** as the search **query**
- PyRx** was used as a cross validation tool to further study pharmacokinetic properties
- The F5446 SMILES sequence was extracted from the PubChem Library, and the **SwissADME** tool analyzed drug likeness and ADMET properties.



MATHEMATICAL MODELING

$$V_{ij}(r) = \frac{1}{\alpha\sigma_{ij}^6 + r^6} \left(\frac{4\epsilon_{ij}\sigma_{ij}^{12}}{\alpha\sigma_{ij}^6 + r^6} - 4\epsilon_{ij}\sigma_{ij}^6 \right)$$

- Above is a modified **Lennard-Jones potential equation** to generate the 6 most **energetically favorable** docking interactions. The **Vakser Lab** used this equation using these parameters. $V(r)$ represents the **potential energy** as a function of the **interatomic distance** r .
 - ϵ = depth of the **potential well** (strength of attraction)
 - σ = distance at which the **potential energy** is zero
 - r = distance between the two **interacting particles**.
- The equation accounts for attractive van der Waals forces and repulsive Pauli repulsion forces.

$$U(r) = \epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Lennard-Jones Potential Equation

- Above is a standard **Lennard-Jones potential equation**. $U(r)$ represents the potential energy as a function **interatomic distance** r .
 - ϵ = depth of the **potential well** (strength of attraction).
 - σ = distance at which the **potential energy** is zero
 - r = distance between the **two interacting particles**
 - $\left(\frac{\sigma}{r} \right)^{12}$ = repulsive term (Pauli repulsion, due to electron cloud overlap)
 - $\left(\frac{\sigma}{r} \right)^6$ = attractive term (van der Waals dispersion forces)

These equations are frequently **used in molecular dynamics in modeling**, allowing me to use these functions to **rank the configurations** produced by the **GRAMM tool**.

ADDITIONAL CHARTS

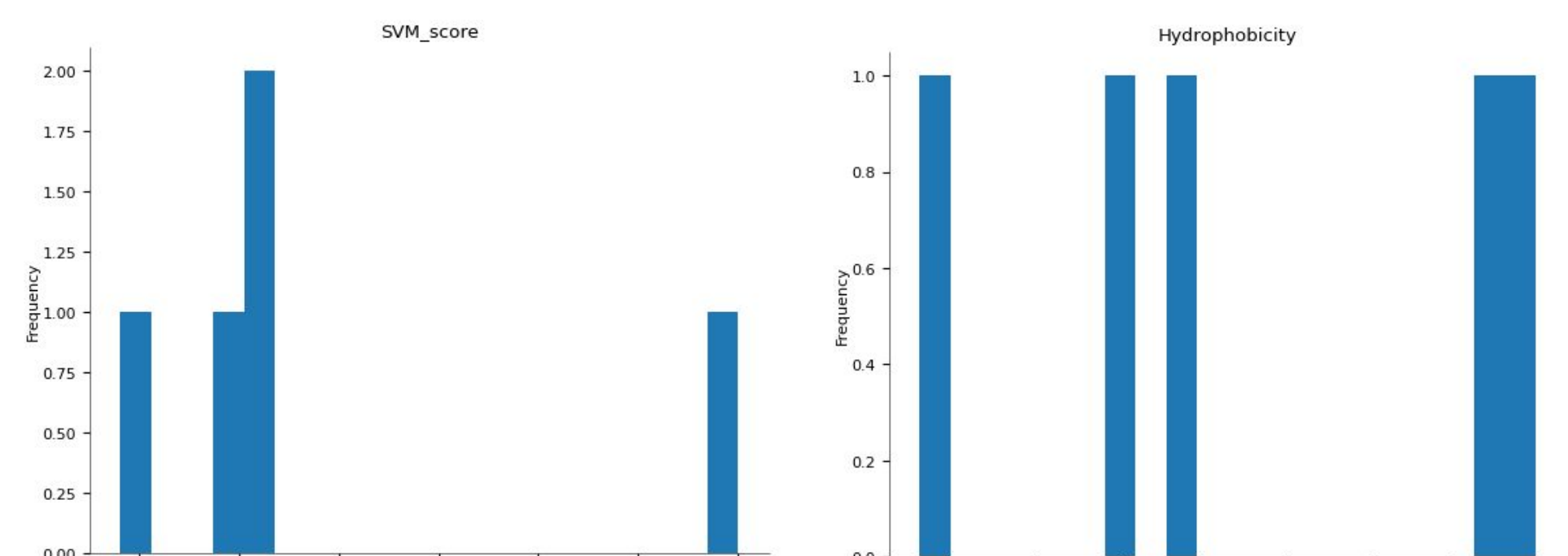


Figure 1: SVM Score vs Frequency

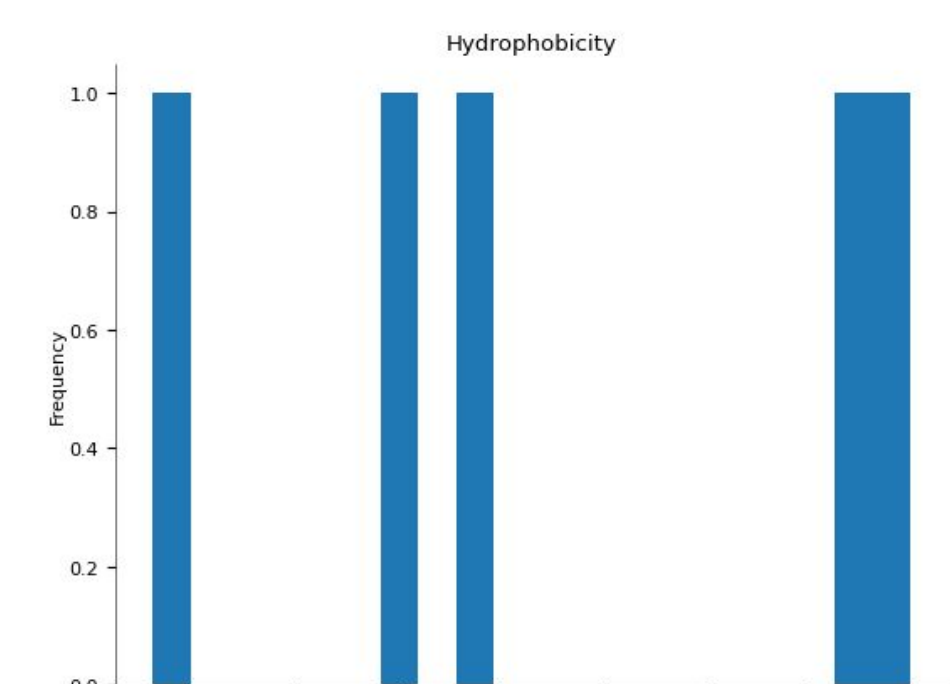


Figure 2: Hydrophobicity vs Frequency

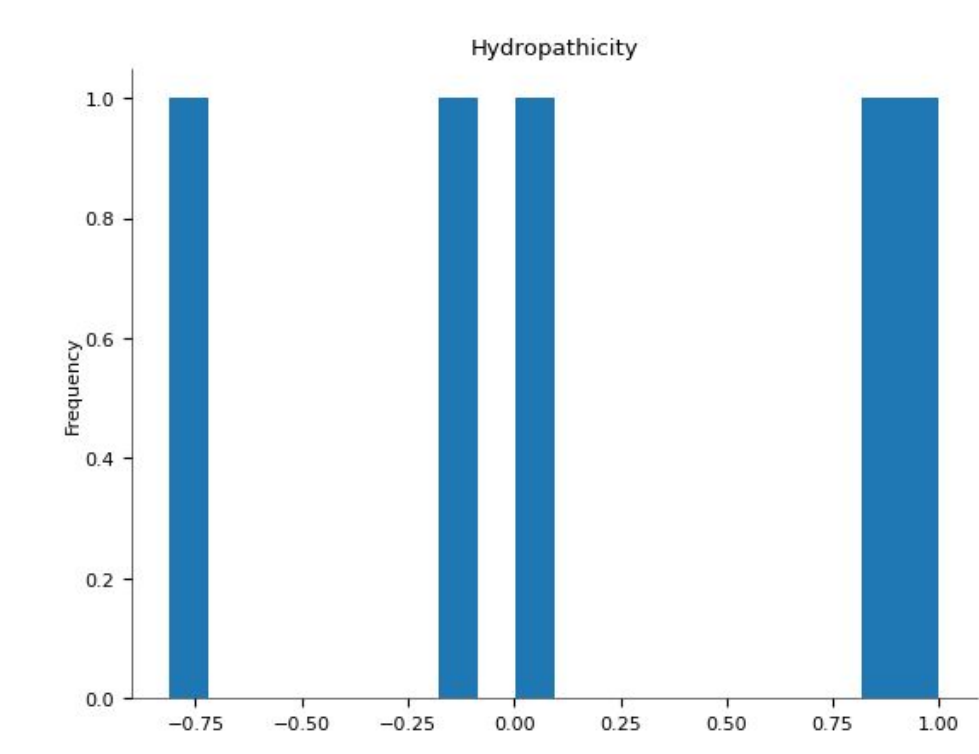


Figure 3: Hydropathicity vs Frequency

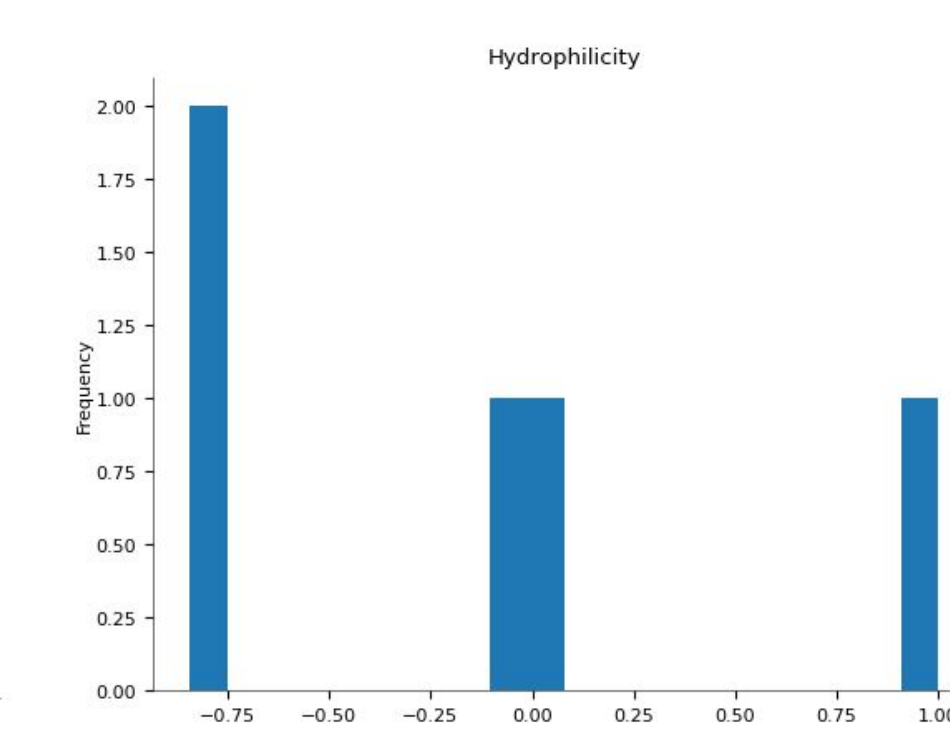


Figure 4: Hydrophilicity vs Frequency

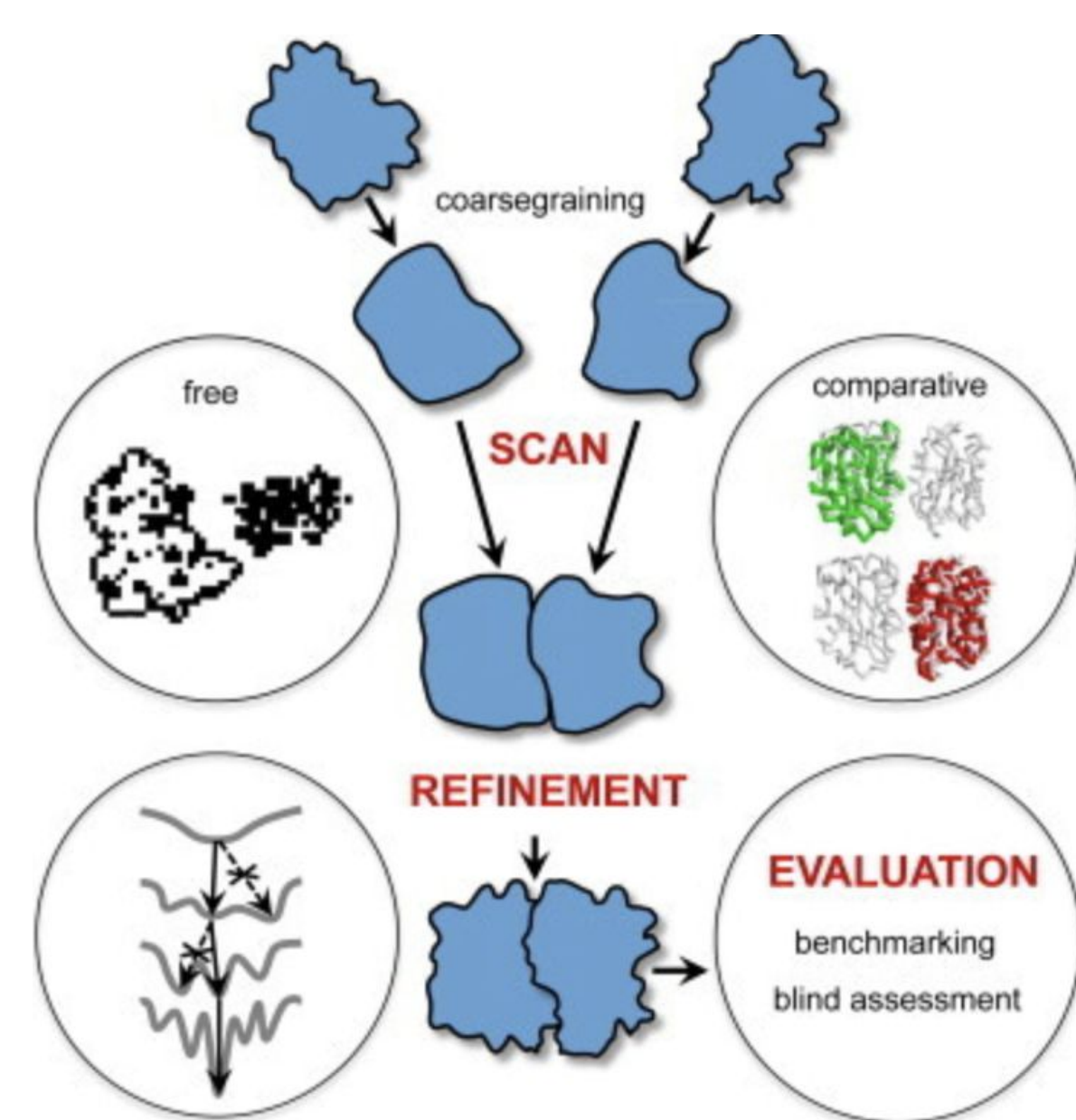


Figure 5: Protein Docking Methodology for Global Range Molecular Matching (GRAMM). (Vakser, 2014)*

GRAMM uses a subset of **Computer Aided Drug Design (CADD)** called **Structure-Based Drug Design (SBDD)**. First, **GRAMM** coarse grains the molecules, scans them, and docks them to ultimately calculate the physicochemical **properties**. Then, using the **Vakser equation**, **GRAMM** scores each model.

RESULTS

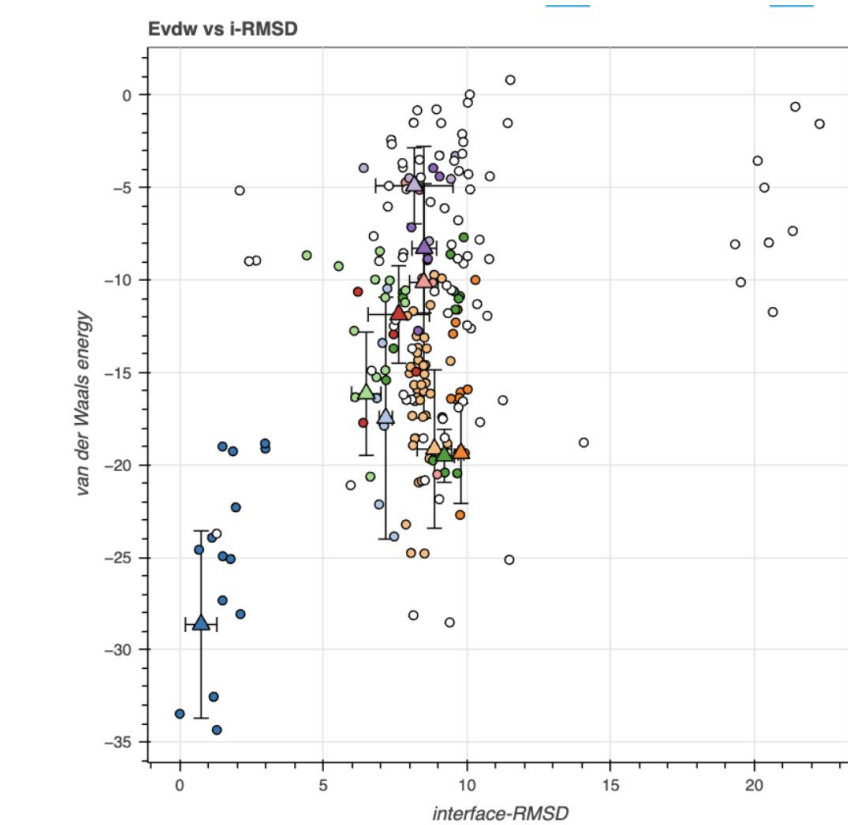


Figure 6: Van der Waals Energy (EVDW) vs Interface Root Mean Square Deviation (I-RMSD) in top 10 HADDOCK Clusters

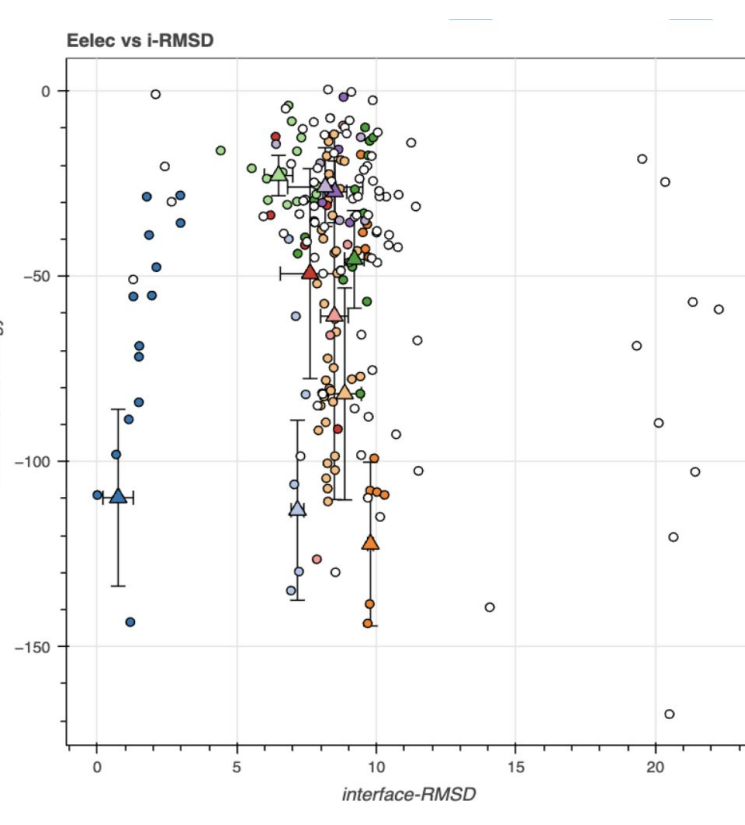


Figure 7: Electrostatic Energy (Eelec) vs Interface Root Mean Square Deviation (I-RMSD) in top 10 HADDOCK Clusters

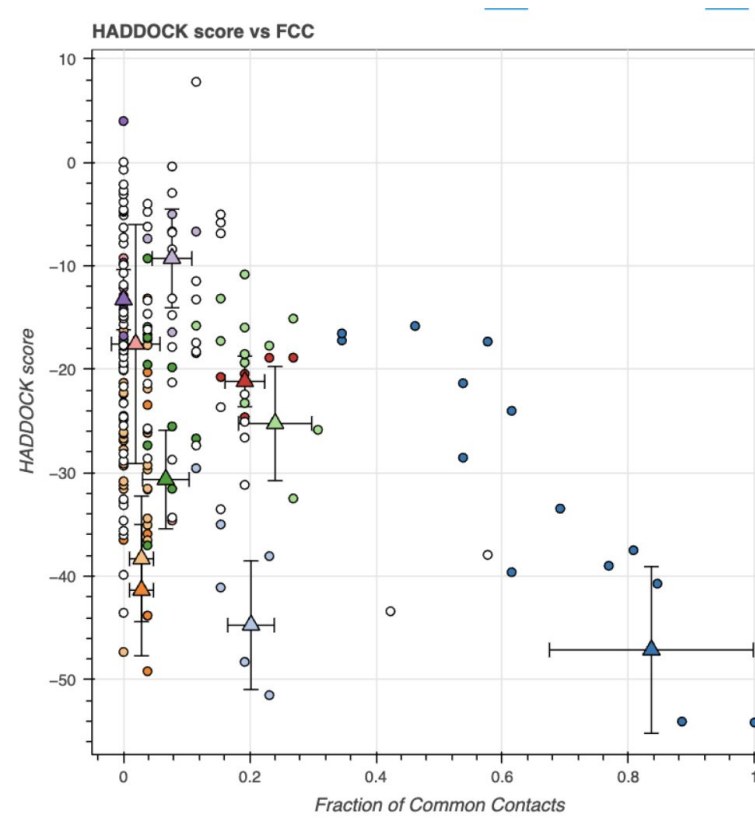


Figure 8: HADDOCK Score vs Fraction of Common Contacts (FCC) in top 10 HADDOCK Clusters

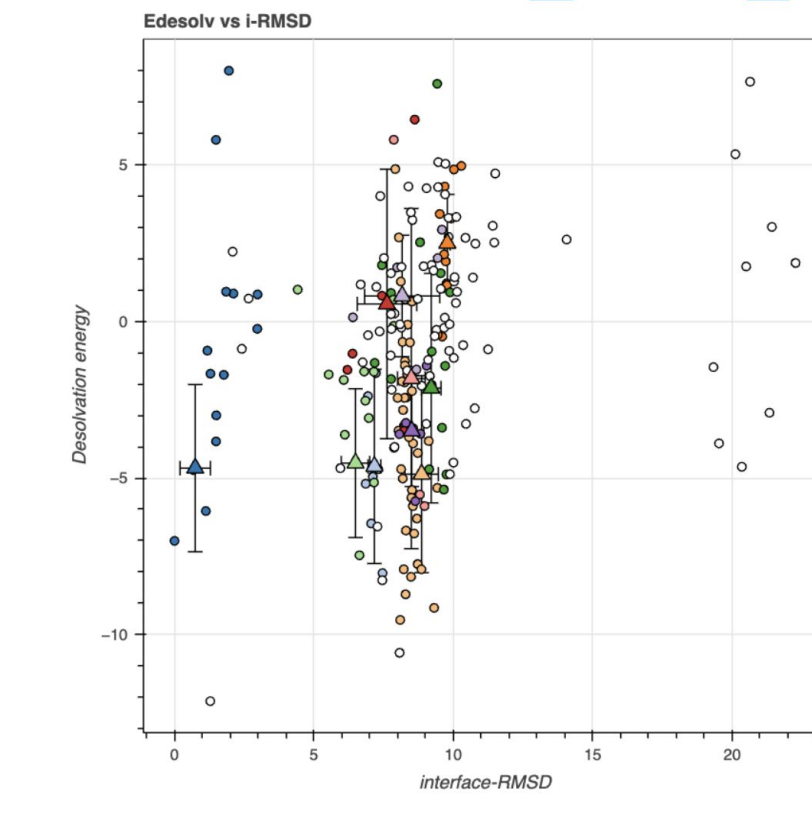


Figure 9: Desolvation Energy (Edesolv) vs Interface Root Mean Square Deviation (I-RMSD) in top 10 HADDOCK Clusters

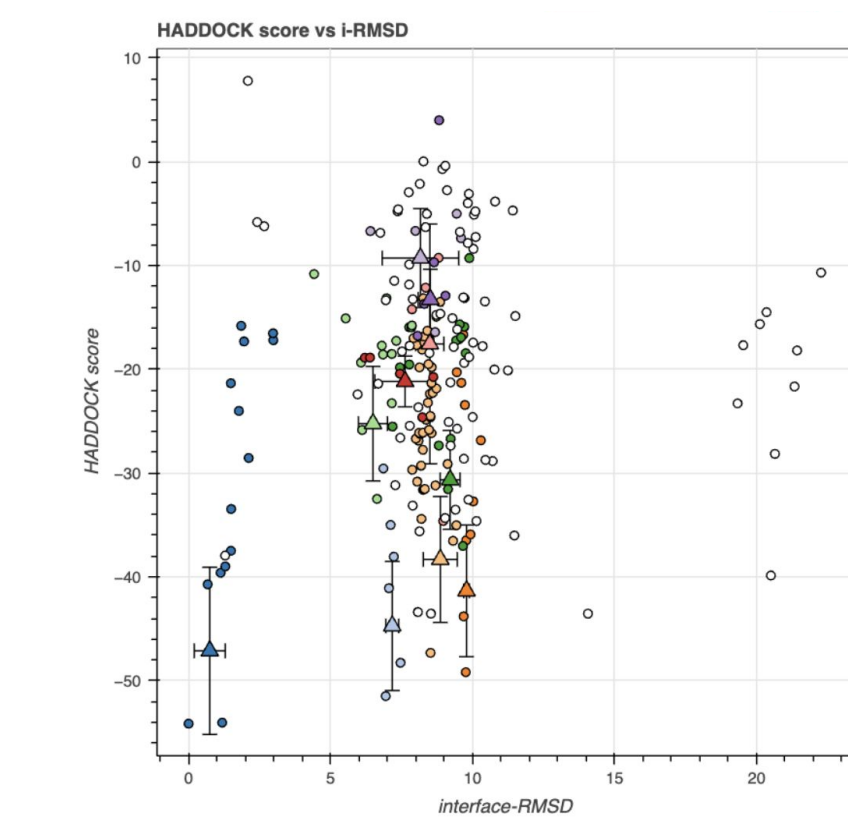


Figure 10: Desolvation Energy (Edesolv) vs Interface Root Mean Square Deviation (I-RMSD) in top 10 HADDOCK Clusters

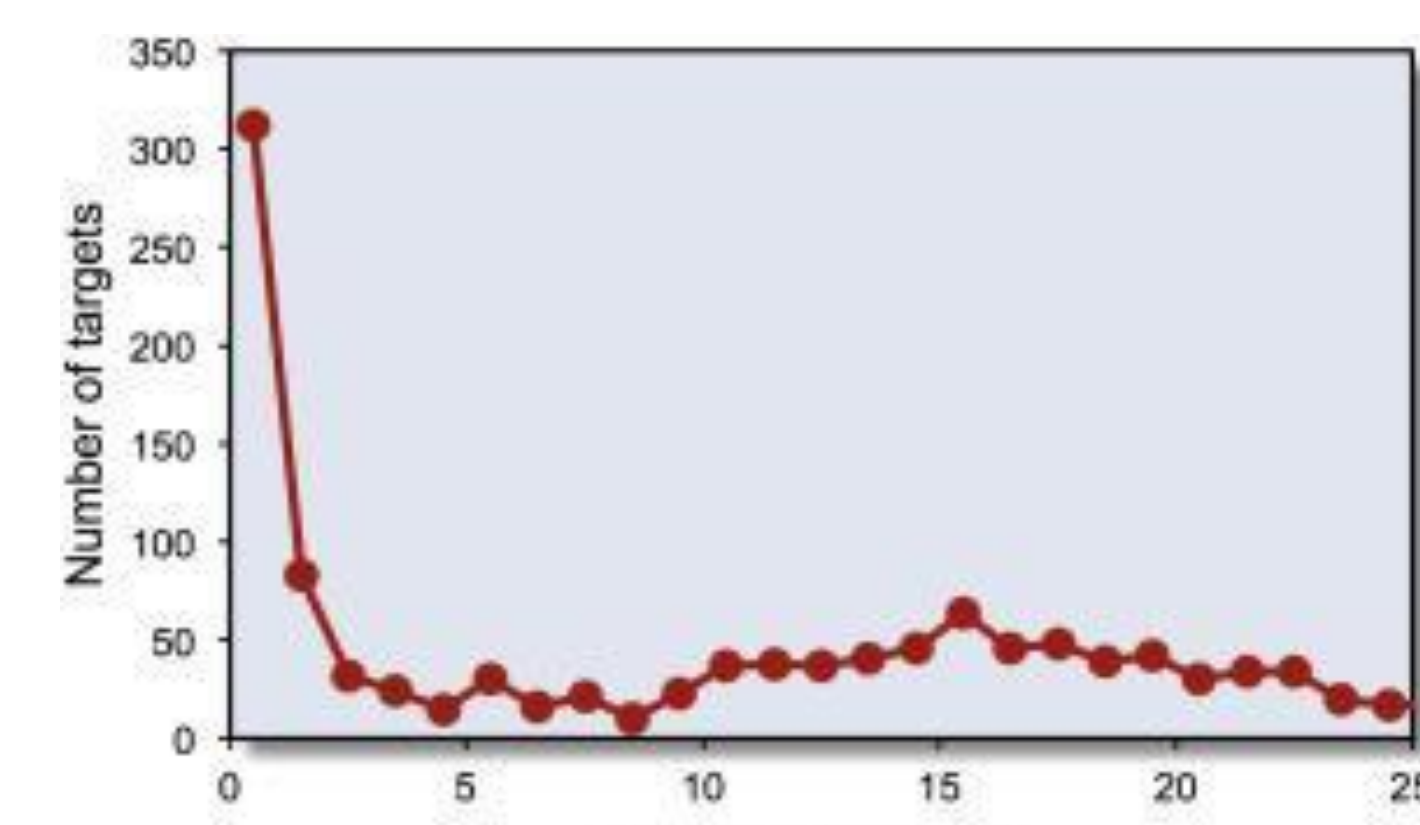


Figure 11: Interface Root Mean Square Deviation (I-RMSD) vs Number of Targets (Vakser)*

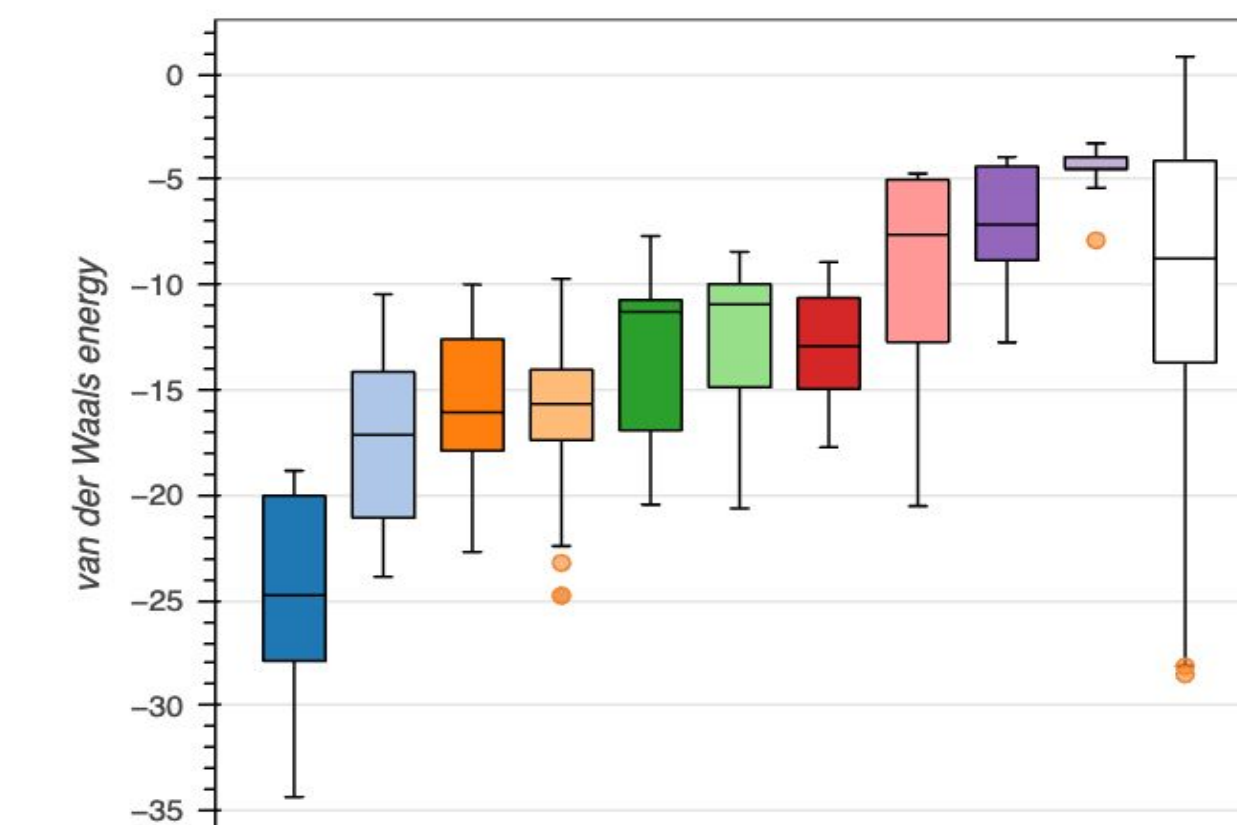


Figure 12: Box & Whisker Plot between Cluster # vs van der Waals Energy

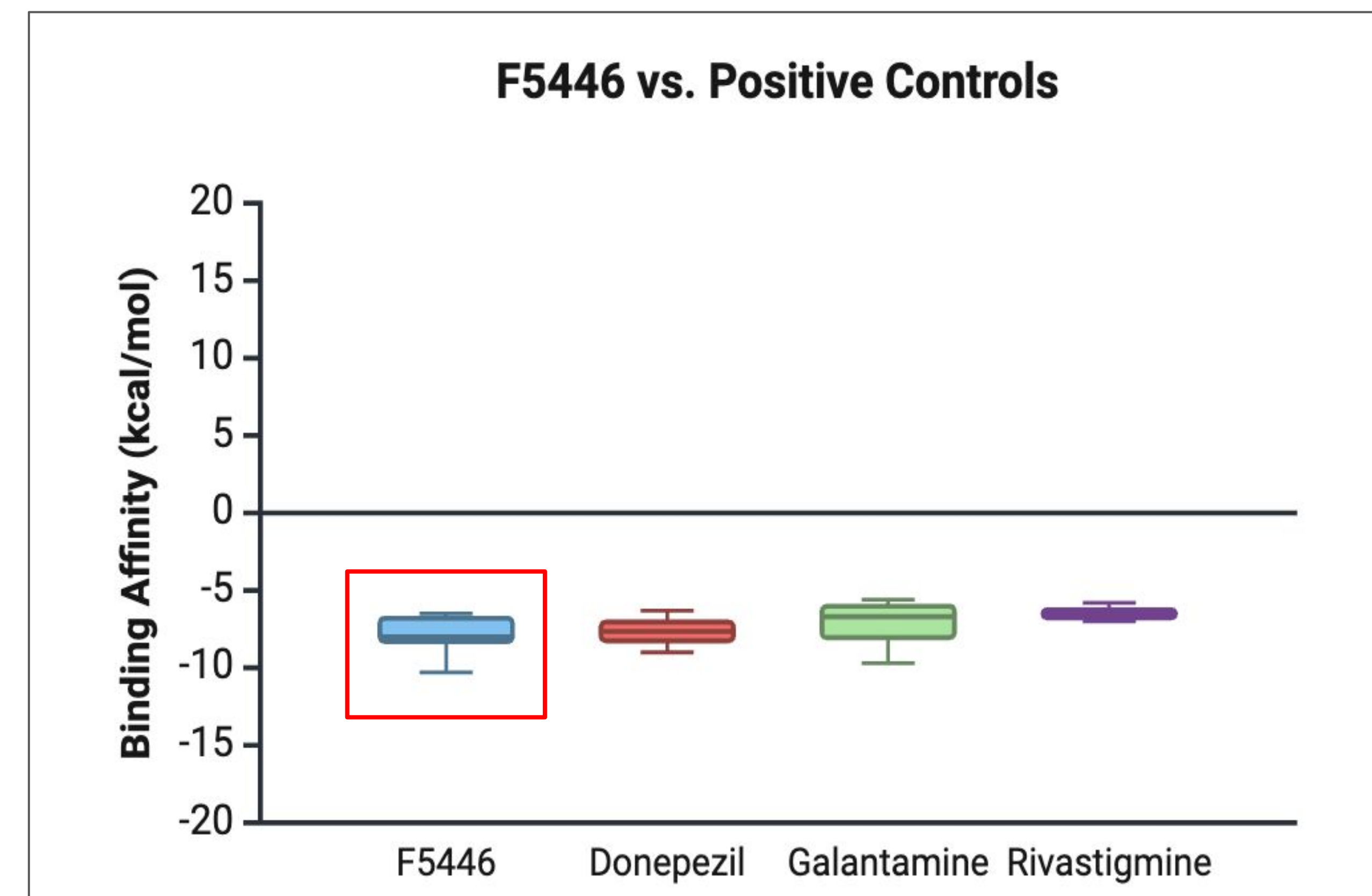


Figure 13: Binding Affinity (lower is better) vs Current Alzheimer's Medications

Original Protein	Protein Name	Sequence Length	Accession Number	Protein Name	Sequence Length	Accession Number	Protein Name	Sequence Length	Accession Number
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696
ADAM10	ADAM10	1012	P18696	ADAM10	1012	P18696	ADAM10	1012	P18696

Figure 14: ToxinPred Support Vector Machine Results (Lipinski's Rule Of Five)

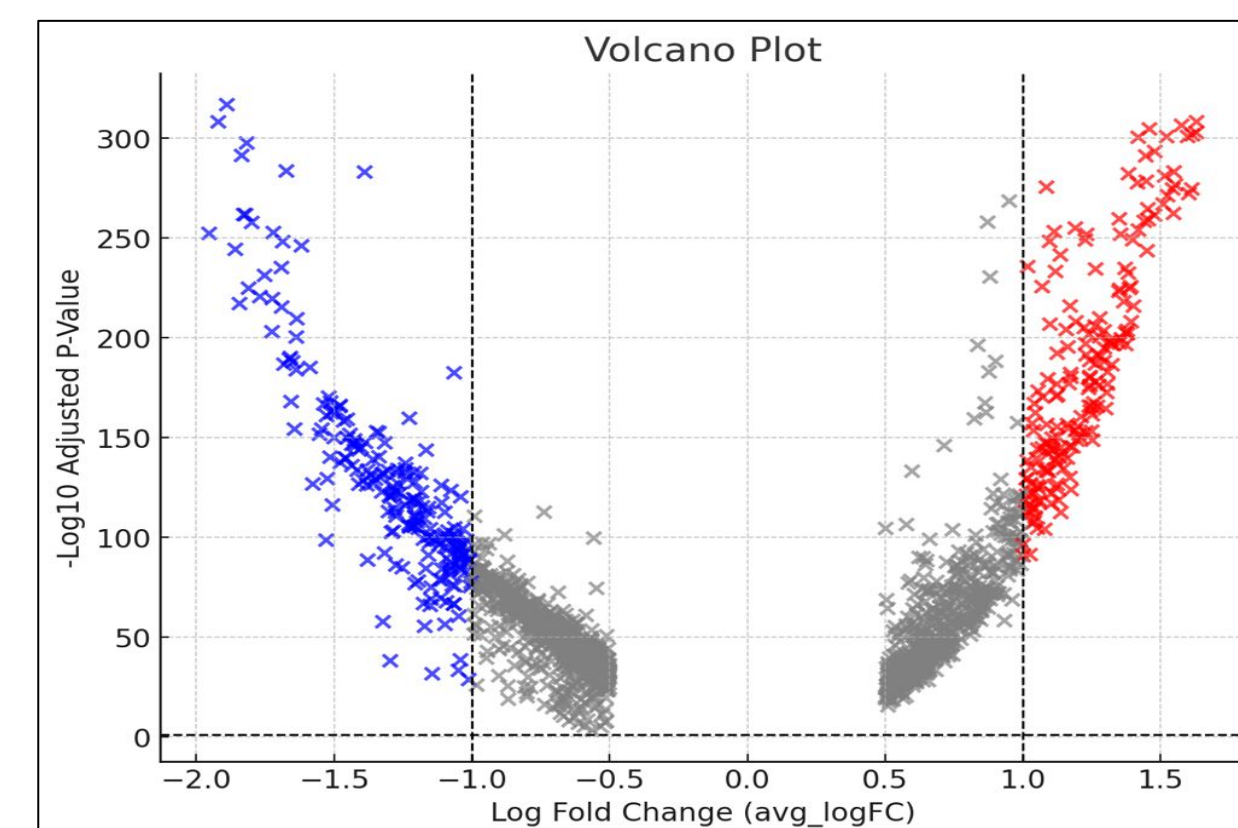


Figure 15: Differential Gene Expression Log fold Change vs -Log10 Adjusted P-Value

STATISTICAL ANALYSIS

Index	SVM_score	Hydrophobicity	Hydropathicity	Hydrophilicity	Mol_wt
SVM_score	1.0	-0.019807	0.066956	0.058834	-0.236617
Hydrophobicity	-0.019807	1.0	0.892232	-0.841943	-0.264747
Hydropathicity	0.066956	0.892232	1.0	-0.810596	-0.163919
Hydrophilicity	0.058834	-0.841943	-0.810596	1.0	-0.050469
Mol_wt	-0.236617	-0.264747	-0.163919	-0.050469	1.0

Figure 16: Correlation Coefficient Matrix Between Physicochemical Properties

An in-depth statistical analysis was used to supplement the **data-driven approach** taken in this study. Atoms, bonds, and residues were counted to compare **average mutant peptide** data with the **“normal” peptide**. The results were then analyzed using basic statistics involving **mean** and **standard deviation (std.)**, single **t-tests**, and **t-tests** between **positions**. The t-tests between Suv39h1 and Tau positions were used to assess relationships between **molecular combinations** of the **Suv39h1 structure**. The reasoning behind using this test was to compare differences between **NCBI molecular data** and **ToxinPred mutant sequences**. For each sequence, the **p-value** was evaluated to identify **statistical significance**. The **p-value threshold** used determined any value less than 0.05 to be **statistically significant** ($p < 0.05$). Lastly, the **Pearson correlation coefficient** was calculated using a correlation matrix. From the matrix, the most notable relationships were between **hydropathicity** and **hydrophobicity**, **Hydrophobicity** vs. **Hydrophilicity**, and weak correlations overall in **SVM classification** (1.0, -0.019807, 0.066956, 0.058834, -0.236617). This is expected, as the SVM of Suv39h1 already concluded the **epigenetic factor** was **non-toxic**. The generated **parameter values** extracted from omics **NCBI data** suggest a **non-toxic sequence peptide of Suv39h1**.

STATISTICAL ANALYSIS, cont'd

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

The Root Mean Square Deviation (RMSD) is the **average distance** between the atoms of superimposed molecules. In this study,

the RMSD measures how **stable Suv39h1's structure** is when bound to the drug candidate. RMSD is measured in Angstroms, a fundamental unit in structural biology which 10^{-10} m. A **low RMSD** (under 2 Angstroms), the **binding is stable**. If the RMSD is higher (above 3 Angstroms), the binding may cause **structural deformation**. In this study, a modified measure of RMSD was used called **interface-RMSD**. The **I-RMSD** focuses on the binding area (active site) rather than the docking as a whole, which As seen in the results, the best performing structure is cluster 3, which consistently had the **lowest I-RMSD value**. In this equation, δ_i is the distance between atom i and a reference structure or the N equivalent atoms.

$$FCC = \frac{\text{Number of common contacts between two structures}}{\text{Total number of contacts in the reference structure}}$$

The **Fraction of Common Contacts** (FCC) is also used as a measure of overlap between the two structures used in docking. In this case as well, **Cluster 3** outperforms each of the other clusters, which translates to a highly **optimized structure**.

CONCLUSIONS

- The research
 - Identified the **first association between Alzheimer's Disease and Suv39h1**
 - Identified a novel inhibitor to treat Alzheimer's Disease (F5446)** MAPT (microtubule associated protein tau) and Suv39h1 ligand-receptor complex is **-8.3 kcal/mol**, which is **moderately strong**
 - This validates the usage of Suv39h1 as a **drug target**
 - Opens up potential for further **in-vitro** and **in-vivo** validation
 - The image below visualizes the top 3 models from GRAMM
- This methodology decreases costs significantly, as wet-lab studies often cost hundreds of thousands, sometimes even millions of dollars
- F5446 performs better than current Alzheimer's medications including Donepezil, Galantamine, Rivastigmine, which were used as positive controls for the study.

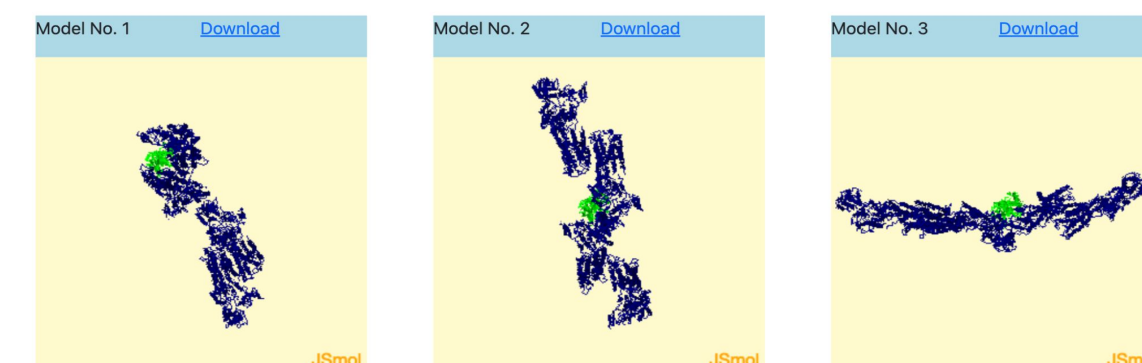


Figure 17: Top 3 Constructed Protein-Protein Models.

FUTURE RESEARCH

I am **extremely excited** about the **future of this work**. To build on this work, I will continue this research by doing the following:

- Apply **Neural Ordinary Differential Equations (nODE)** to develop a more **streamlined** and **automatic** system for **in-silico drug discovery** by creating a multi-model **database-scanning** system
- Adapt my research towards **other diseases** to streamline drug candidate screening
- Test **ABNN** model on multiple demographics to test for bias.
- Construct **novel drugs in-vitro** in a wet-lab setting to apply techniques such as **Isothermal Titration Calorimetry (ITC)** to the study
 - Currently was not possible due to **limitations** in laboratory access

KEY REFERENCES

Bhatti, G. K., Reddy, A. P., Reddy, P. H., & Bhatti, J. S. (2020). Lifestyle Modifications and Nutritional Interventions in Aging-Associated Cognitive Decline and Alzheimer's Disease. *Frontiers in Aging Neuroscience*, 11, 369. <https://doi.org/10.3389/fnagi.2019.00369>

Chen, Z. R., Huang, J. B., Yang, S. L., & Hong, F. F. (2022). Role of Cholinergic Signaling in Alzheimer's Disease. *Molecules*, 27(6), 1816. <https://doi.org/10.3390/molecules27061816>

DeTure, M. A., & Dickson, D. W. (2019). The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration*, 14(1), 32. <https://doi.org/10.1186/s13024-019-0333-5>

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein-protein docking approach based on biochemical and/or biophysical information. *Journal of the American Chemical Society*, 125(6), 1731-1737. <https://doi.org/10.1021/ja026939x>

Gao, Xiaodie, et al. "Epigenetics in Alzheimer's Disease." *Frontiers*, Frontiers, 24 May 2022. <https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2022.911635/full>

Gupta, Sudheer, et al. "[PDF] In Silico Approach for Predicting Toxicity of Peptides and Proteins | Semantic Scholar." Semantic Scholar, Sept. 2013. <https://www.semanticscholar.org/paper/In-Silico-Approach-for-Predicting-Toxicity-of-and-Gupta-Kapoor/d9b0b942/09953df552524ee0352b36e08a93c9>

Honorato, R. V., Koukos, P. I., Jimenez-Garcia, B., Tsaregorodtsev, A., Veriato, M., Giachetti, A., Rosato, A., & Bonvin, A. M. J. J. (2021). Structural biology in the clouds: The WebMR-EOSC ecosystem. *Frontiers in Molecular Biosciences*, 8, 729513. <https://doi.org/10.3389/fmolb.2021.729513>

Prajapati, Prakash, et al. "Drug Discovery Process. Figure 2 Computer Aided ...". ResearchGate, Dec. 2017. https://www.researchgate.net/figure/Drug-Discovery-Process-Figure-2-Computer-Aided-Drug-Design_fig2_322065181

Vakser, I. A. (2014). Protein-protein docking: from interaction to intermolecular. *Biophysical Journal*, 107(8), 1785-1793. <https://doi.org/10.1016/j.bpj.2014.08.033>

Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastriis, P. L., Karaca, E., Melquiond, A. S. J., Van Dijk, M., De Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428(4), 720-725. <https://doi.org/10.1016/j.jmb.2015.09.014>