Structure-Based Identification of High-Affinity scFv Antibody Targeting YKL-40 as a Biomarker for Early Alzheimer's Disease Detection

DIYA RAJAN[1,2]

[1]Vernon Hills High School, Vernon Hills, IL
[2]Eigen Sciences, Apex, NC

Published December, 2026

Alzheimer's disease (AD) is a neurological disease where the brain shrinks due to the death of brain cells. Common symptoms of this disease include problems with memory and cognitive tasks. Patients suffering from this disease have elevated levels of the protein YKL-40, which is a secreted glycoprotein produced by inflammatory cells. A single-chain fragment variable (scFv) is a small unit of an antibody that is genetically engineered to bind to an antigen. We hypothesize that the scFv 5yd3 will bind to the YKL-40 protein and can be used to detect YKL-40 protein in the cerebrospinal fluid of AD patients. By obtaining the amino acid sequence of YKL-40 from the UniProt database, we were able to utilize it to create a 3D structure of YKL-40 using AlphaFold3, a machine learning based method. Next, we conducted a series of molecular docking experiments, in which we used 10 different scFv and attempted to bind them to YKL-40. These docking results were further compared to the predicted binding site obtained from P2rank and ScanNet web server. Finally, after receiving the results from the molecular docking, we analyzed the strength of those bonds through PLIP and PRODIGY. Following our analysis, scFv 9b6t was selected as a suitable candidate for detecting YKL-40 protein. In future studies, we will be performing amino acid mutations to enhance the antibody binding affinity. This current research will pave the way for future detection of AD, and possibly even treatment.

1. INTRODUCTION

Alzheimer's disease is a neurological disease, the most common form of dementia, that primarily affects people 65 years and older. [1] This disease is caused by the buildup of amyloid proteins, which leads to the death of brain cells and subsequent brain shrinkage. Symptoms of Alzheimer's include problems with memory, confusion, mood changes, and problems with cognitive tasks. [2] Worldwide, it is estimated that around 55 million people suffer from Alzheimer's and an estimated 6.9 million in the US. While there is no standard cure for this disease yet, there are some treatments to ensure a better quality of life. This includes Cholinesterase inhibitors, Disease-modifying immunotherapies, as well as cognitive therapy in order to limit memory loss. Biomarkers, characteristics of the body that one can actually measure, can be used to indicate the presence of disease, such as AD, as well as its progression. [3] They can be detected through methods like mass spectrometry, PCR, and others.

YKL-40, a secreted glycoprotein, is an example of one of these biomarkers. This protein is produced by inflammatory cancer cells. It is found in the brain, or cerebrospinal fluid, in elevated levels in patients with neurological and neurodegenerative cases. Elevated levels of YKL-40 can be observed in patients with Alzheimer's disease. It is a marker that can indicate inflammation in patients with neurological diseases.

As shown in Figure 1, patients with AD suffer from an accumulation of Amyloid-β plaques. This build up causes astrocytes, glial cells in the brain, to become reactive and have weakened function. It is also common for patients to suffer from the breakdown and death of neurons in the brain, which attributes to cognitive trouble. Shortly after, nueral inflammation occurs, the protein YKL-40 attempts to fix the neurons. We propose, that using scFv's we can locate these YKL-40 proteins and possible administer treatment.

Molecular docking simulations are used to research how small molecules interact with proteins, and is crucial in understanding how proteins interact. We used HDOCK in order to understand

Molecular docking simulations are used to research how small molecules interact with proteins. By using these simulations, it can help understand how proteins interact. It is also

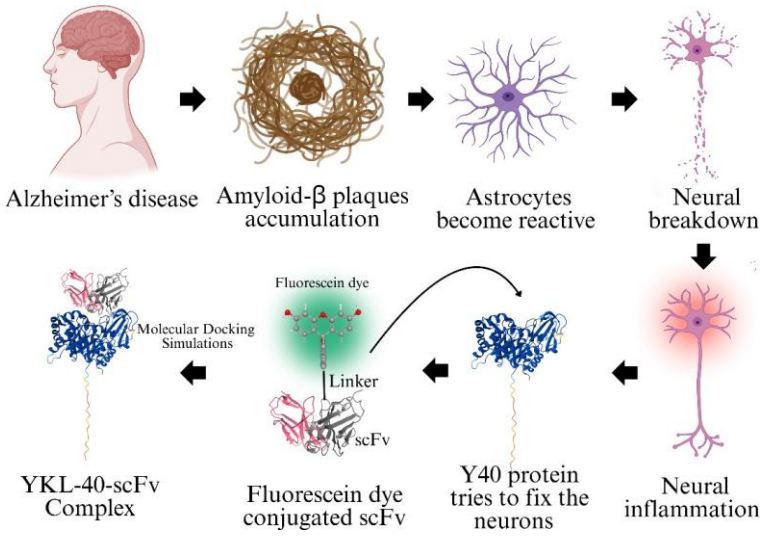


Fig. 1. Schematic representation of the AD formation and YKL40 inhibition using an antibody.

crucial to drug development, and provides an avenue to understand how drugs interact with their target. The applications of this technology include drug discovery and understanding protein interactions. Commonly used molecular docking software includes HDOCK, Autodock, and Glide.

Recently, Ali et al. Have identified that YKL-40 protein present in the cerebrospinal fluid (CSF) can be used in the detection in the early detection of AD. In this research we have used computational methods to identify antibodies that can binds to this protein and aid in the AD detection. From the simulations we have identified antibody 9B6t could be used for this purpose. These results pave the way towards a novel AD detection method and aid in combatting the global burden of the disease.

2. METHOD

Uniprot is a database of protein sequences and information about those proteins. [4] We used it to get the amino acid sequence for the protein YKL-40, which is an essential biomarker in this study. PROTTER is a tool that aids in visualizing proteins location inside the cell, allowing for further analysis and research. [5] We used it to create Figure 2(a), in which we were able to visualize YKL-40 and its location outside the cell. The Human Protein Atlas is a database that can map the location of proteins in cells, tissues, and organs. [6] We used this to understand where YKL-40 is secreted from. AlphaFold 3 is an AI program that generates predictions of a protein's structure based on its amino acid sequence. [7] We used this application to develop an image of the protein YKL-40's structure, based on its amino acid sequence, which we later used in experiments to investigate how different ligands bind to it. Molecular docking simulations are used to research how small molecules interact with proteins. We used it to determine how each scFv bonded with the YKL-40 protein and which bonds were the strongest. P2Rank is a machine learning application that predicts binding sites of ligands to proteins. [8] We used it to predict the binding site of the scFvs to the protein YKL-40, as shown in Figure 3a. ScanNet is a deep learning software that also predicts binding sites for ligand-protein interactions. [9] We used it to formulate an alternative view of the scFv and YKL-40 interaction, as shown in Figure 3b. PLIP is a software that analyzes the chemical interactions between macromolecules and ligands. [10] We used it to determine all the different interactions and bonds between each scFv and YKL-40. It listed any hydrophobic interactions, hydrogen bonds, salt bridges, and more. PRODIGY is a web server that calculates the binding affinity between a molecule and a protein. [11] We used it to determine which scFv bound the strongest to YKL-40, which we found to be 9b6t. RING is a software that identifies any non-covalent bonds in a ligand-protein structure. We used it to generate all the images of the various interactions, including hydrogen bonds, salt bridges, and others.

3. RESULTS

Protein Location and Surface Properties

Using Protter (Figure 2a), the amino acid sequence of YKL-40 was mapped, confirming that the protein is located outside of the cell, secreted into the extracellular space. Validation from the Human Protein Atlas (Figure 2b) supported these findings, demonstrating YKL-40 secretion from inflammatory cells into the periplasmic environment. This establishes YKL-40 as an accessible biomarker for detection in cerebrospinal fluid (CSF).

Binding Site Prediction

Binding site analysis was performed using **P2Rank** (Figure 3a) and **ScanNet** (Figure 3b). Both approaches identified consis-

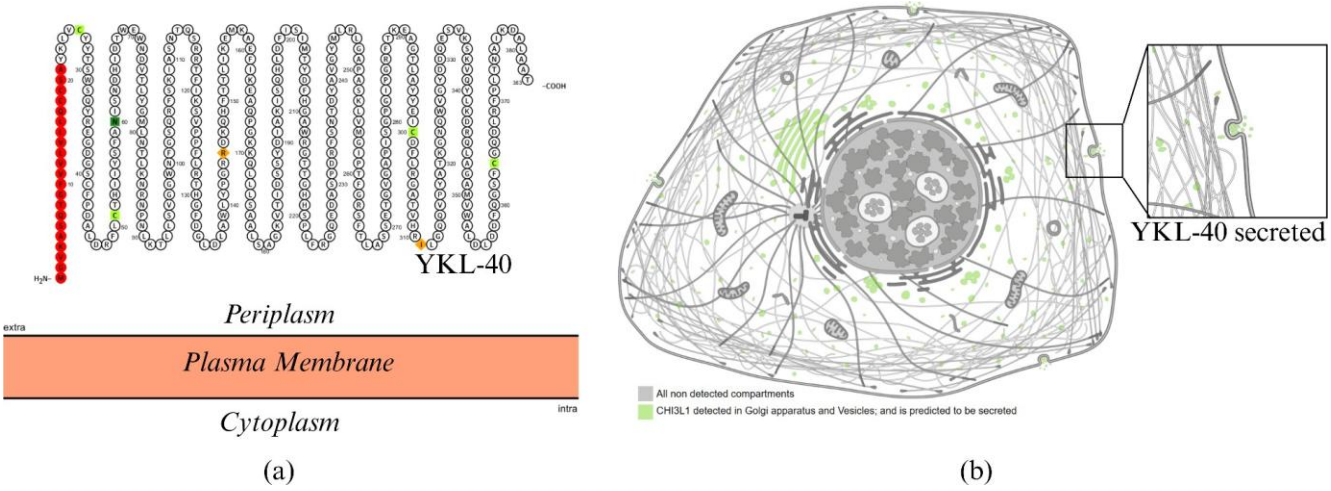


Fig. 2. (a) We used Protter to create an image of the amino acid sequence of YKL-40 to understand that this protein is located outside of the cell, in the periplasm. (b) This image was generated using the Human Protein Atlas and explains that YKL-40 is secreted into the periplasm, and the process in which it is released from the cell.

tent ligand binding regions highlighted in red/pink, providing independent validation of the probable YKL-40 scFv docking interface. The overlap of predictions across machine learning (P2Rank) and deep learning (ScanNet) increases confidence in the identified pocket.

Molecular Docking of scFvs to YKL-40

Docking simulations with HDOCK were performed for 10 candidate scFvs. Structural visualization in ChimeraX (Figure 4) showed clear binding orientations of scFvs to the YKL-40 surface, with YKL-40 represented in purple and scFvs in pink. Comparative docking revealed differences in binding stability across candidates.

Binding Affinity Analysis

Binding free energies calculated with PRODIGY (Figure 5) demonstrated that scFv 9b6t exhibited the strongest interaction with YKL-40, having the lowest binding energy of −19.9 kcal/mol. This suggests high binding affinity, making 9b6t the most promising diagnostic antibody fragment.

Interaction Profiling

Chemical interaction analysis using PLIP revealed that scFv–YKL-40 complexes contained multiple stabilizing interactions, including hydrogen bonds, hydrophobic interactions, and salt bridges (Figure 6). In particular, scFv 9b6t formed the most frequent and stable non-covalent interactions compared to the other candidates. RING analysis further confirmed the presence of extensive non-covalent bonds (Figure 7), supporting the predicted binding stability.

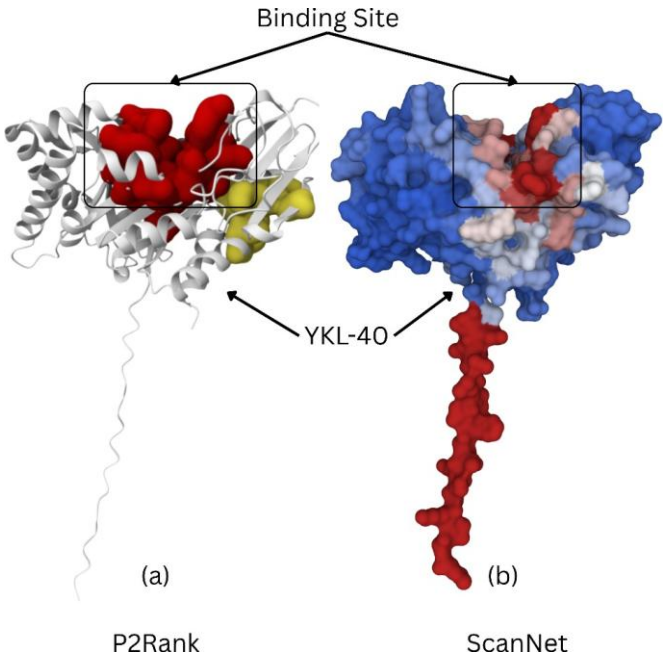
4. DISCUSSION

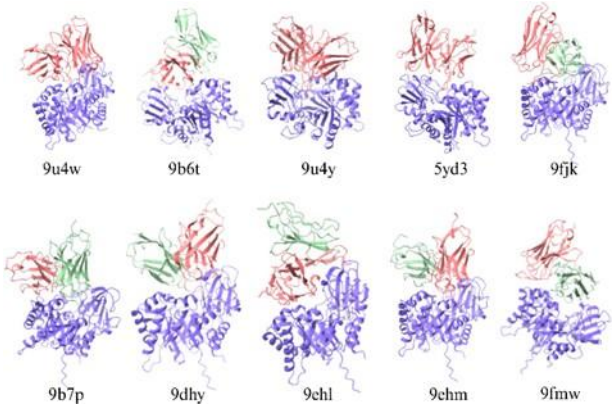
With the advancement of medical technologies, multiple approaches have been developed to aid in the detection and diagnosis of AD. These include the measurement of biomarkers in blood, urine, and cerebrospinal fluid (CSF), cognitive assessments, and neuroimaging techniques. [2] Despite these advances, current diagnostic tools remain inconclusive. A definitive diagnosis can only be confirmed post-mortem through the identification of hallmark neuropathological features, such as amyloid-β plaques and neurofibrillary tangles. Compounding this challenge, neurodegenerative processes associated with AD are estimated to begin 20–30 years before the onset of clinical symptoms, significantly limiting the window for early diagnosis and therapeutic intervention. [2]

Applications

This research demonstrates the potential of scFv 9b6t as a novel diagnostic tool for early Alzheimer's disease detection. By targeting YKL-40, which is elevated in the cerebrospinal fluid of AD patients, this antibody fragment can help identify the disease much earlier than current diagnostic approaches. Beyond early detection, scFv 9b6t can be applied to monitoring disease progression. Since YKL-40 levels change over time, repeated measurements using this scFv could allow clinicians to track disease severity and progression more accurately.

Another important application lies in broader neurological and inflammatory conditions. Because YKL-40 is also expressed in other disorders, this antibody has potential utility for biomarker discovery across multiple diseases and for evaluating therapeutic responses to interventions.





5. LIMITATIONS

A major limitation of this study is that YKL-40 detection relies on cerebrospinal fluid, which requires invasive lumbar puncture. This reduces the practicality of large-scale screening compared to blood-based diagnostics. Protein concentration variability presents another challenge. YKL-40 levels differ between patients, and without standardized clinical thresholds, the accuracy of diagnosis could be affected. Careful calibration will be necessary to ensure reproducibility. Finally, this work is entirely computational. Although docking and binding energy analyses provide valuable predictions, experimental validation is required to confirm affinity, specificity, and real-world diagnostic potential.

6. FUTURE WORK

Future studies will focus on experimental confirmation of the computational results. Laboratory assays such as ELISA, surface plasmon resonance (SPR), and bio-layer interferometry (BLI) will be essential to verify the binding affinity of 9b6t to YKL-40. Antibody engineering will also play a critical role. Through amino acid mutagenesis or affinity maturation, 9b6t could be optimized to enhance stability, reduce off-target binding, and improve diagnostic performance.

Additionally, future research should investigate the detection of YKL-40 in less invasive biofluids like blood or saliva, extend testing into preclinical animal models and clinical cohorts, and explore the therapeutic potential of scFvs to not only detect but also modulate YKL-40 activity in Alzheimer's disease.

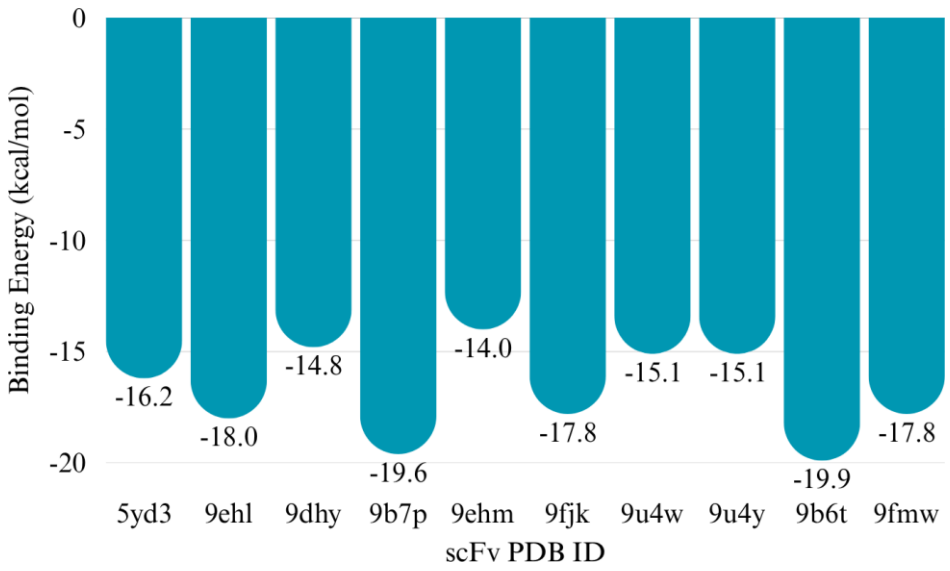


Fig. 5. Through the use of PRODIGY, a webserver that determines binding affinity, we were able to determine the binding energy for each YKL-40-scFv interaction. After analyzing the results, it is clear that 9b6t had the highest binding affinity, meaning it bound the strongest, since it had the lowest binding energy of -19.9.

7. CONCLUSION:

From this research, we are able to conclude that the scFv best suited for this treatment is 9b6t. After conducting molecular docking simulations, generating data on binding energy, and obtaining the different chemical interactions of each YKL-40 and scFv pairing, it is clear that 9b6t is the most appropriate. We are able to see this because from our analysis we know that 9b6t had the lowest binding energy with YKL-40, making it the best match. Since 9b6t and YKL-40 have the strongest bond, it is clear that 9b6t is the best choice.

REFERENCES

1. P. Scheltens, “Alzheimer’s disease,” The Lancet **397**, 1577–1590 (2021).
2. M. Kennison and E. Long, “The Long Journey of Alzheimer’s Disease,” Journal of Christian Nursing **35**, 218–227 (2018).
3. O. L. López and S. T. Dekosky, “Clinical symptoms in Alzheimer’s disease,” Handbook of Clinical Neurology **89**, 207–216 (2008).
4. T. U. Consortium, “UniProt: a hub for protein information,” Nucleic Acids Research **43**, 204–212 (2014).
5. U. Omasits, C. H. Ahrens, S. Müller, and B. Wollscheid, “Protter: interactive protein feature visualization and integration with experimental proteomic data,” Bioinformatics **30**, 884–886 (2013).
6. F. Pontén, K. Jirström, and M. Uhlen, “The Human Protein Atlas-a tool for pathology,” The Journal of Pathology **216**, 387–393 (2008).
7. J. Abramson, “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” Nature **630**, 493–500.
8. R. Krivák and D. Hoksza, “P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure,” Journal of Cheminformatics **10**, 39–39 (2018).
9. J. Tubiana, D. Schneidman-Duhovny, and H. J. Wolfson, “ScanNet:an interpretable geometric deep learning model for structure-based protein binding site prediction,” Nature Methods 19, 730–739.
10. S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, “PLIP: fully automated protein-ligand interaction profiler,” Nucleic Acids Research 43, 443–447 (2015).
11. A. Vangone and A. Bonvin, “PRODIGY: A Contact-based Predictor of Binding Affinity in Protein-protein Complexes,” Bio Protoc 7, 2124–2124 (2017).

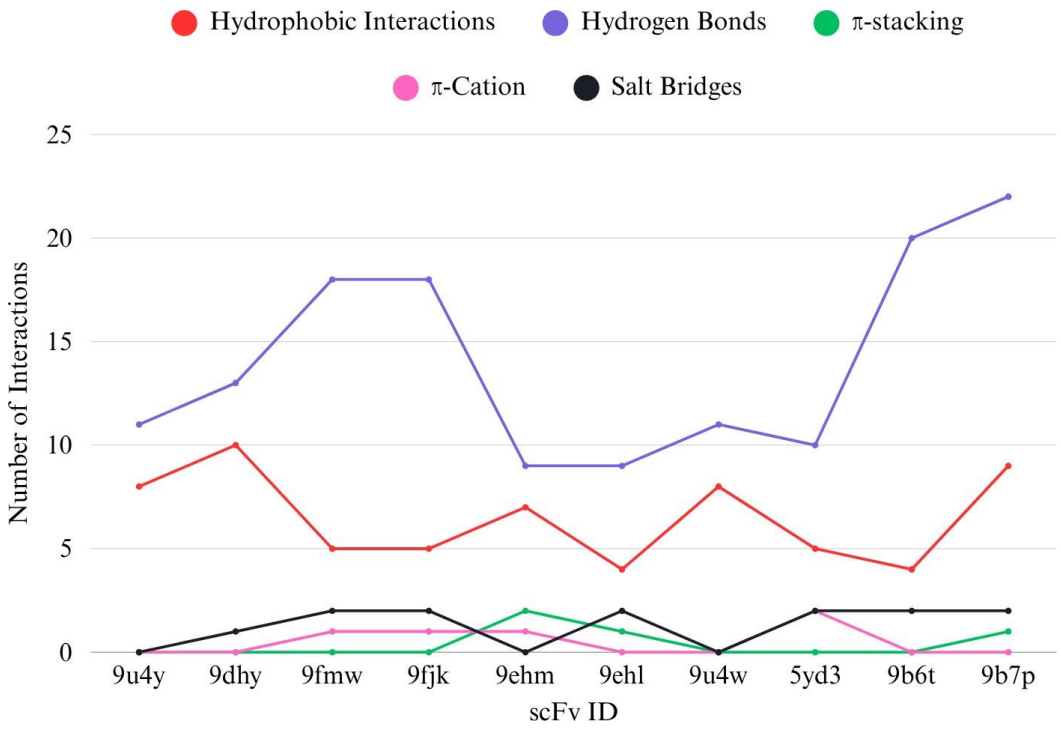


Fig. 6. YKL40-scfv interactions computed using plip.By using PLIP, a software that determines chemical interactions between YKL-40 and the scFv's, we were able to determine which interactions were present in each combination, and how often they occurred.

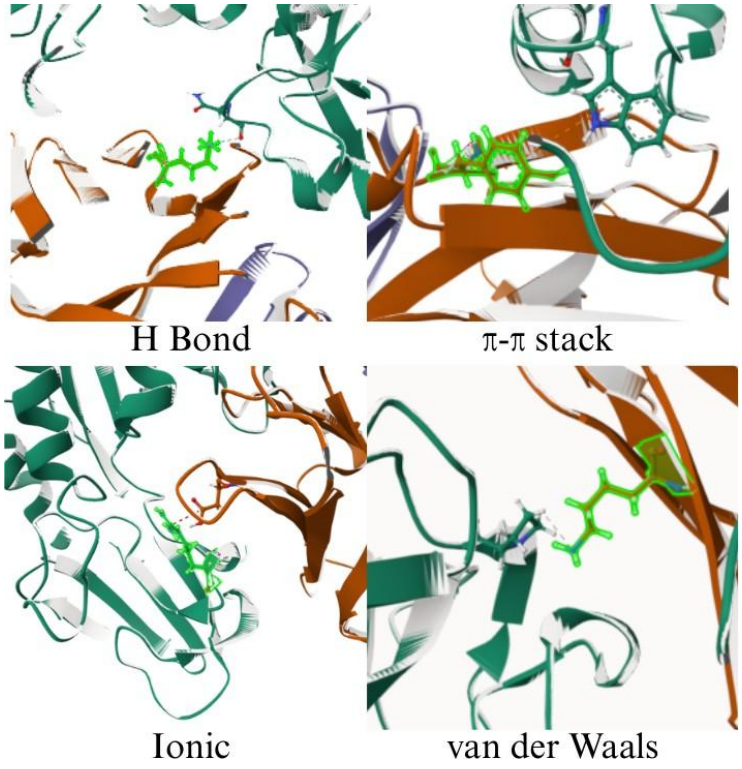


Fig. 7. Different bonds formed between the YKL40 and scFv structures.