

**Title: Designing Drug-Like Molecules using
Bioinformatics**

By

Natalie Ghigliotty

Middlesex County College
Edison, NJ

CHM 240 (Research in Chemistry)

Research Advisor: Phalguni Ghosh, Associate Chair,
NASC, MCC

12/20/16

Abstract:

The Drug Discovery process is a very time consuming, expensive process. It may take one-billion dollars and twenty years to bring a drug to market and making it available for people to use. Any approach that can be found to cut down on the time and money spent during the drug discovery process is valuable. The goal of this research was to build inhibitors for the human protein, Dipeptidyl Peptidase IV, also known as DPP4. DPP4 is a protein found in humans that is responsible for problems related to Type-2 diabetes. The goal here was not only to build inhibitor of this protein, but to create entirely new molecules that have never been tested in DPP4 to our knowledge before. This research is open-ended, meaning that someone may pick up where we left off and attempt to synthesize these molecules in the lab, and eventually test them in DPP4 for their inhibition activity. This research is valuable, there is a great deal of people around the world today struggling with Diabetes, so any way we are able to improve on research and development that is being done today is useful.

Introduction:

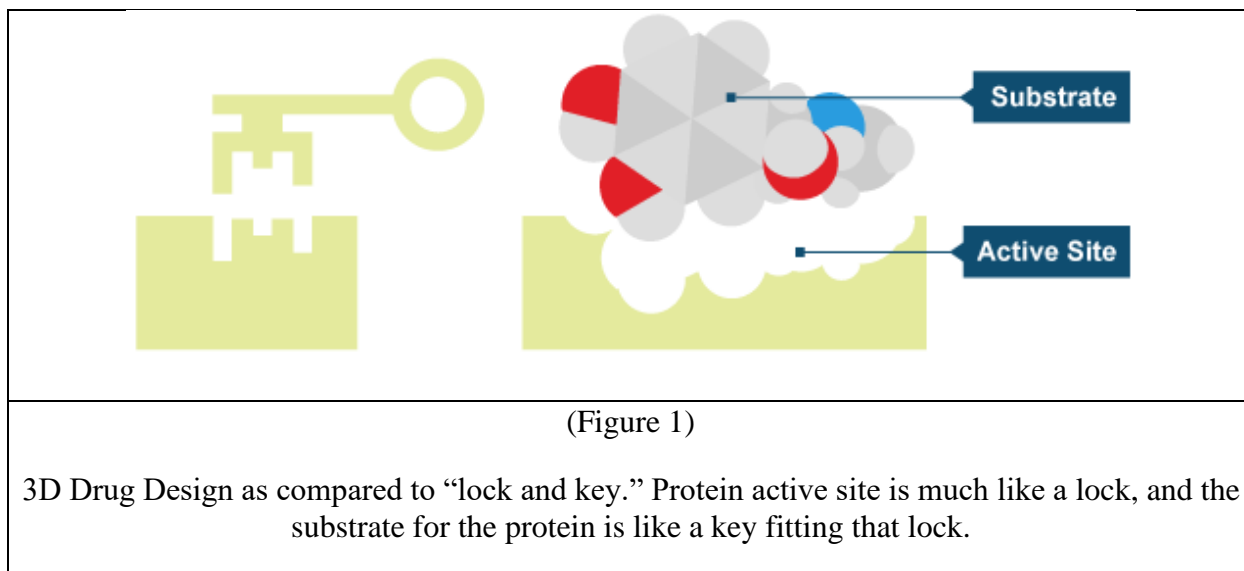
This research cannot be completed without some basic knowledge about the molecules that of are interest, or without some tools to help conduct the project. This is where the world of Bioinformatics becomes so incredibly useful. Bioinformatics is the field of developing methods and software to help people better understand all the information on biological molecules. These tools combine technology, biology, computer science, and mathematics in such a way that allows the user to better understand biological information.

In order to conduct our research, we utilized several Bioinformatic tools. The first tool used was the Protein Data Bank, otherwise known as the PDB, to extract the 3D crystal structures of the target protein and other molecules being studied. Another tool I used was UCSF Chimera, which is a free software program that is used to view 3D macromolecular structures. I used Chimera to view molecules inside the protein pocket and study their molecular interactions. Next, I used ChemDraw software that was provided by Middlesex County College to create small molecules in 2D and 3D. Last, I used online websites, SwissTargetPrediction and SwissDock, which are run by the Swiss Institute of Bioinformatics, to check the binding characteristics and interactions of my small designed molecules. Each of these will be discussed later in detail.

The Drug Discovery Process is a very long, expensive process. Today's big pharmaceutical companies follow the same basic steps in bringing a drug to market. The first step in this process would be to identify the target protein. It is essential that researchers know exactly what protein in the body it is they are trying to deliver a drug to. Once that is successfully done, researchers would then move on to the Lead Discovery phase where they synthesize some molecules that have been identified as the best fit molecules to interact with the target protein.

Next, researchers will move on to the Lead Optimization/Medicinal Chemistry phase, where they are constantly working on modifying and adjusting the lead molecules with hopes that these changes will produce better results in their binding abilities to the protein. Here, chemists will also synthesize the molecule. Once successful in this stage, chemists will move on to cell-free binding assay, In Vitro studies, where the synthesized molecules are tested in binding assays outside and inside cells. If the compound shows reasonable activity in cell-free and cell-based, they will move on to In Vivo studies where the molecule is introduced to small animals like mice and rats. Once the researchers find this to be successful, the next step would be Pre-Clinical trials, where wide-range testing is done to determine the safety of the molecule and determine the safe dose for first-in man study. This research resides in the Lead Optimization phase, where we have created some molecules and are attempting to make accurate changes to them to produce better results in their interaction abilities with the target protein.

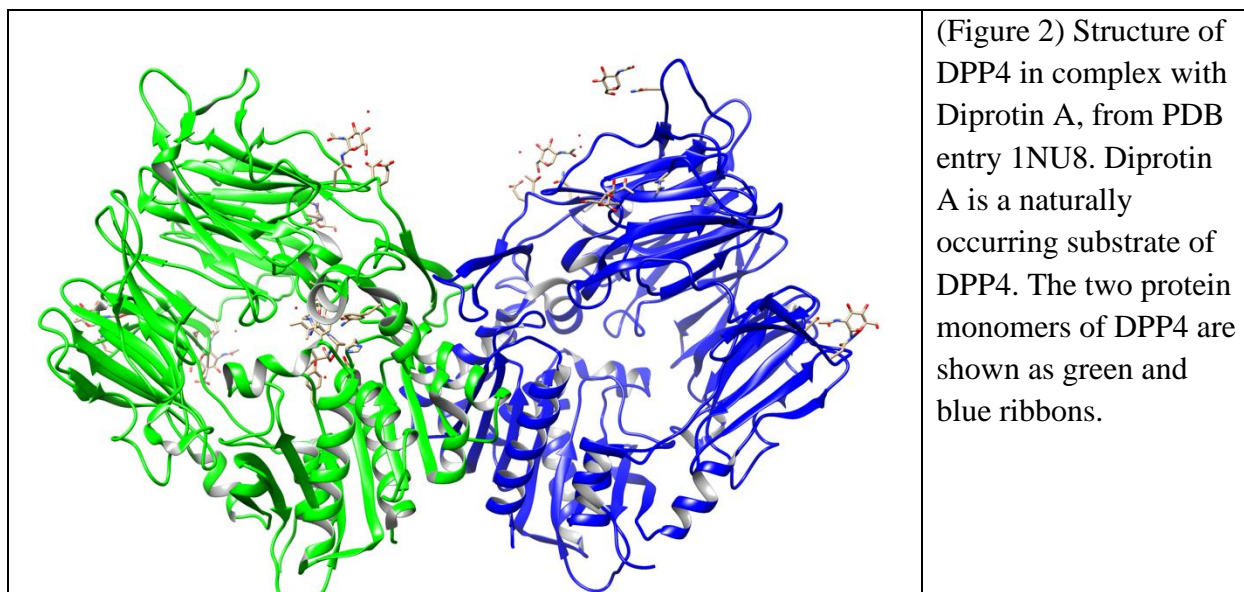
Using the Bioinformatic approach, inhibitor design is being done through 3D Structure Based Design method. By using this approach, it is clear exactly how small molecules will interact with the target protein, DPP4. This allows chemists to make accurate changes to the scaffold of a ligand because Bioinformatic tools provide for a richer picture of ligand-receptor interactions. We often referred to this as the “lock and Key” method. If given a lock, it is fairly easy to design some key to fit that lock. 3D Structure based drug design is analogous to that, where the protein’s active site is the lock and a small molecule is the inhibitor, or key, that we want to fit in that lock (Figure 1).



Traditionally, big pharmaceutical companies and drug companies would use the 2D structure-activity relationship (SAR) approach, where thousands and thousands of molecules are synthesized and analyzed, which requires much effort and time. This method does not provide information about how the small molecule they just made interacts with their target protein,

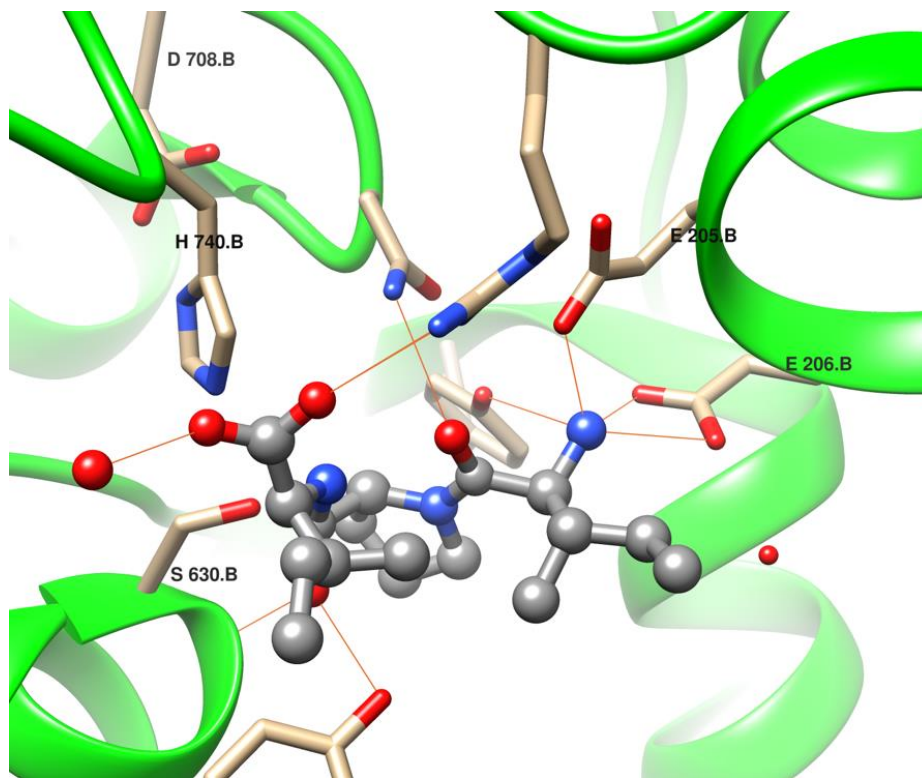
which is something 3D Structure Based Drug Design allows for. This is what makes 3D design efficient. In our case, we were able to find the structure of our protein target, Dipeptidyl Peptidase IV using the Protein Data Bank, a bioinformatic tool, and thoroughly understand the nature of the active site pocket to create some molecule that will inhibit it.

Dipeptidyl Peptidase IV is a human protein found on the surface of cells. It naturally occurs as a dimer, where two identical sub-units complexed together to make one unit (Figure 2). A naturally occurring process that happens in humans' bodies is that when we eat, cells signal to our bodies to release incretin hormones, which in turn signals the release of insulin to lower blood glucose levels. However, when the incretin hormones are released, DPP4 will inactivate some to ensure homeostasis, maintaining only the necessary amount of incretin hormones present in our blood stream. For someone who has Type-2 Diabetes, however, this becomes problematic because their bodies are not able to use insulin properly. So, when the incretin hormones are made inactive by DPP4, there is no longer stimulation to release insulin, causing a decrease in insulin levels and dangerously high levels of blood glucose. Therefore, drug companies have created drug inhibitors, like Januvia, to block the enzymatic action of DPP4, which in turn will stop the degradation of the incretin hormones, which ultimately allows insulin and blood glucose levels to stabilize.



Dipeptidyl Peptidase IV will only clip peptides in our bodies that have an Alanine or Proline present at the 2nd (N-Terminal) position on the substrate. DPP4 inactivates the incretin hormones since they have Alanine present at the second position on the hormone. During the enzymatic reaction, the incretin hormone will enter the pocket of DPP4, and acidic residues on DPP4 active site, Glutamates 205 and 206, will hydrogen-bond to the free-amino acid end of the hormone. This formation of hydrogen bonds will lock the hormone into place and prepare it for

the inactivation reaction (Figure 3). DPP4 residue Serine is made more chemically active by nearby Histidine and Aspartate, creating a catalytic triad of amino acids on the protein. This allows Serine to perform the cleavage reaction, cutting the hormone into two pieces. Now, what is left is two inactivated pieces of the hormone, which get released back into the blood stream, and another incretin hormone can enter the protein pocket and the same reaction will happen, and so on. Drug companies have developed medication that will block the degradation of incretin hormones by DPP4. For example, drugs like Januvia specifically block DPP4 to help people with Type-2 Diabetes better manage their disease.

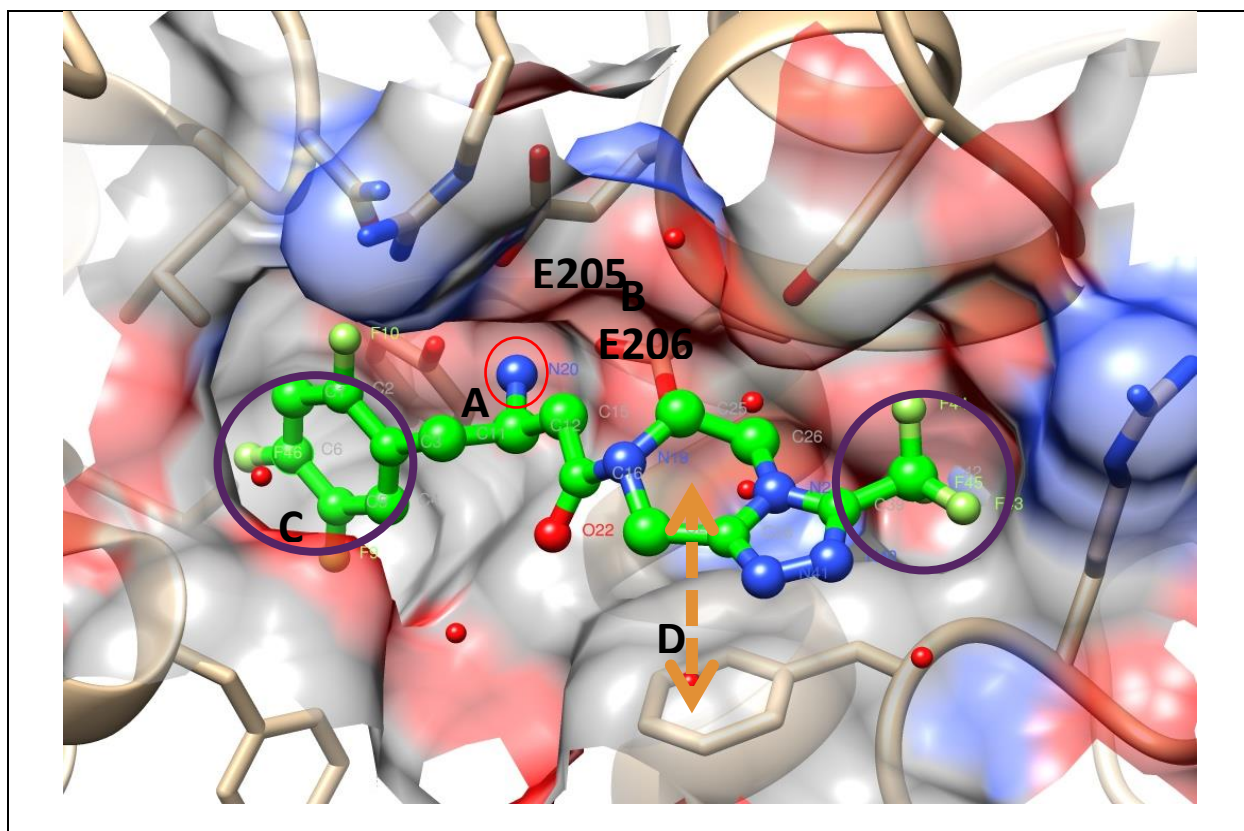


(Figure 3)

Active site of DPP4; DPP4 in complex with Diprotin A, a naturally occurring substrate of the protein. DPP4 breaks down peptides through a cleavage reaction, performed by Serine 630. Diprotin A, shown in grey ball-and stick model, hydrogen bonded to GLU 205 and 206 of DPP4 in PDB entry 1NU8. DPP4 shown in green ribbon, residues in beige stick. Diprotin A has Proline present at 2nd position, which is why it can be cleaved by DPP4.

Using the Protein Data Bank and Chimera, we studied the 3D structure of Januvia in the active site of DPP4 by fetching PDB ID 1X70 in Chimera (Figure 4). After studying this picture

for some time, we were able to make several conclusions on what was necessary for designing inhibitors of DPP4. We can see that inhibitors like Januvia mimic the free-amino acid end of incretin hormones, or in other words these inhibitors will have a basic nitrogen (Figure 4A, red circle around blue atom) that will form the essential hydrogen bonds with acidic residues on the protein, GLU205 and 206 (Figure 4B). These inhibitors will bind to the protein and block the active site, preventing the clipping and inactivation of incretin hormones. The colored surface behind Januvia shows the hydrophobic protein pocket. Looking at the pocket, we can see that there are many smaller hydrophobic pockets (Figure 4C-purple circles), which allows for bulkier groups on the small molecule in the pocket. Something else that should be noted is the stacking effects happening between protein residues and the small ligand (Figure 4D-orange arrow), which will happen because it is energetically preferred. So, moving forward, we know that to design a new molecule to inhibit DPP4, we will need a nitrogen that will allow the molecule to bind to the protein. We also know that bulky hydrophobic groups on the small molecules will allow for a good fit in the pocket of the protein.



(Figure 4)

PDB ID 1X70, Januvia in ball-and-stick model bound to DPP4 shown in surface and stick representations. (A) Basic nitrogen of Januvia, hydrogen bonded to (B) E205, E206. (C) Hydrophobic pockets in the DPP4 active site; (D) stacking interaction between Januvia and DPP4 residue.

Experimental Section:

Tools used:

1. RCSB PRB (www.rcsb.org): to study and download 3D structure of DPP4-Ligand complexes
2. Chimera: for visualization of 3D structures of macromolecules and to prepare the designed molecules for docking
3. ChemDraw 2D: to draw designed molecule to import to ChemBio 3D to generate 3D structure of ligand
4. ChemBio 3D: to minimize the structure of the designed small molecule and to create its SMILES string. SMILES string is a molecular description of the connectivity between atoms in a small molecule. It can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules
5. SwissTargetPrediction: An online tool that can predict the targets of a small molecule using a combination of 2D and 3D similarity measures. It compares the query molecule to a library of several thousand compounds active on selected targets from different species
6. SwissDock: A web-based tool that predicts the molecular interactions that may occur between a target protein and a small molecule

Figure 5 shows the steps we took to design a small molecule. First, we would search the PDB for entries of small molecules in complex with DPP4, then we would open several of those structures in Chimera and superimpose them, viewing the structures of the small molecules in the pocket of the protein.

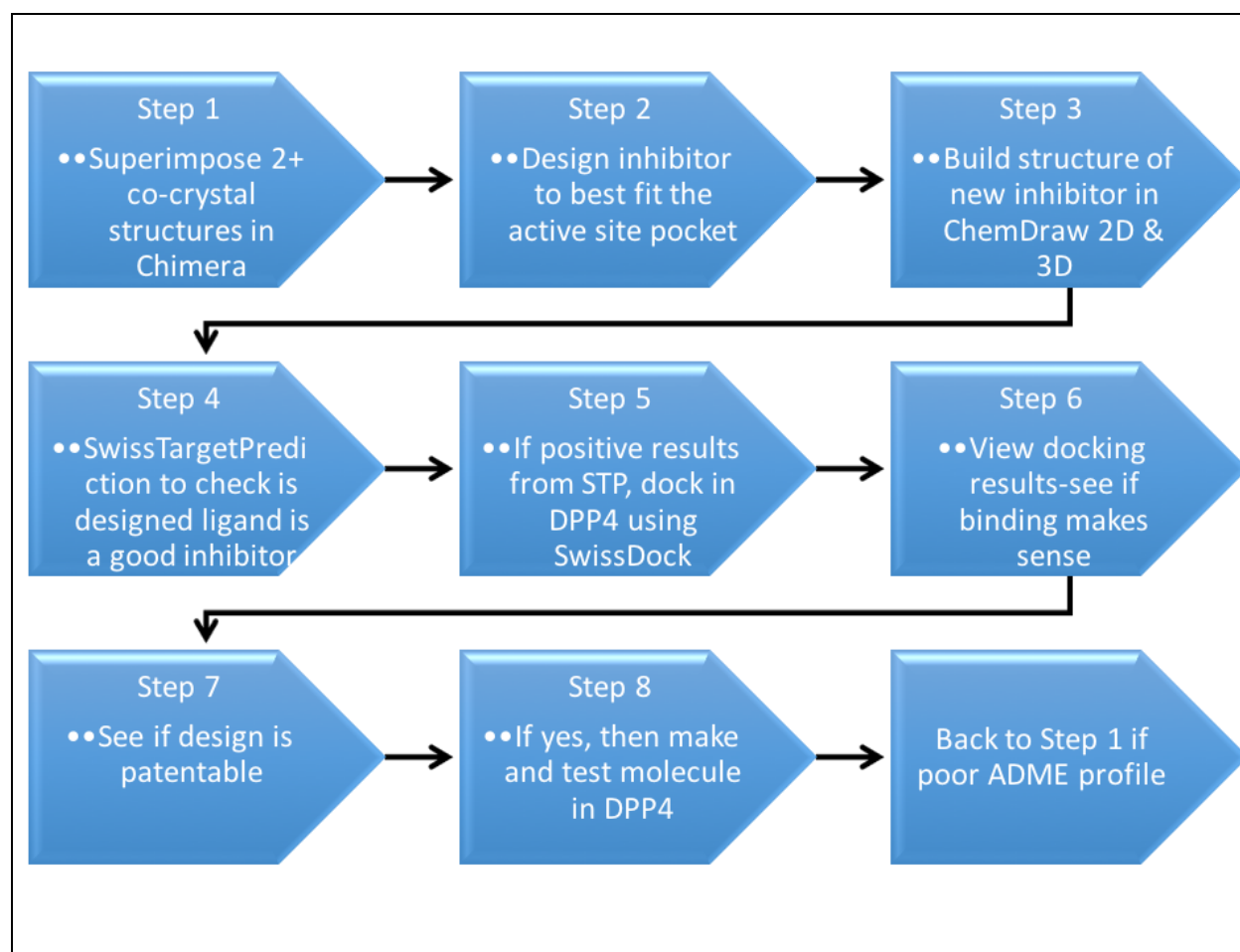
After observing these molecules in the protein pocket, we would be able to design some new small molecule. Then, we would have to build the designed molecule. We would do so using ChemDraw 2D and 3D, bioinformatic tools mentioned earlier. Once the 2D and 3D structures of the molecule were created, we would then use SwissTargetPrediction to runs checks to ensure the designed molecule does indeed bind and interact with our target protein.

Pending positive results from SwissTargetPrediction, we would move forward and dock the small designed molecule to DPP4 using SwissDock, another bioinformatic tool which predicts the intermolecular interactions between the small molecule and target protein.

Once docking is completed, we would view and analyze these results to verify that the binding mode makes sense. This is where our research is completed, however if it were to be continued the next step would be to see if the design is patentable.

If it is indeed patentable, the next step is to synthesize the designed molecule in the lab and test it in DPP4 for its inhibition activity. For a molecule to become a drug, there is some criteria that is considered. Researchers will examine the ADME profile of the molecule, where the absorption, distribution, metabolism, and excretion of the molecule is being studied.

This process altogether is iterative, so if unwanted results were found at any step in the design process, the molecule would undergo some earlier and some modifications would be made with hopes of better results. For example, if at Step 4 we got poor results from SwissTargetPrediction, we would go back to Step 2 and make some design changes to the molecule. Or if the molecule had a poor ADME profile, again it would go back to an earlier step and some changes would be made to the molecule.



(Figure 5)

Process followed in designing new drug-like molecules to inhibit DPP4

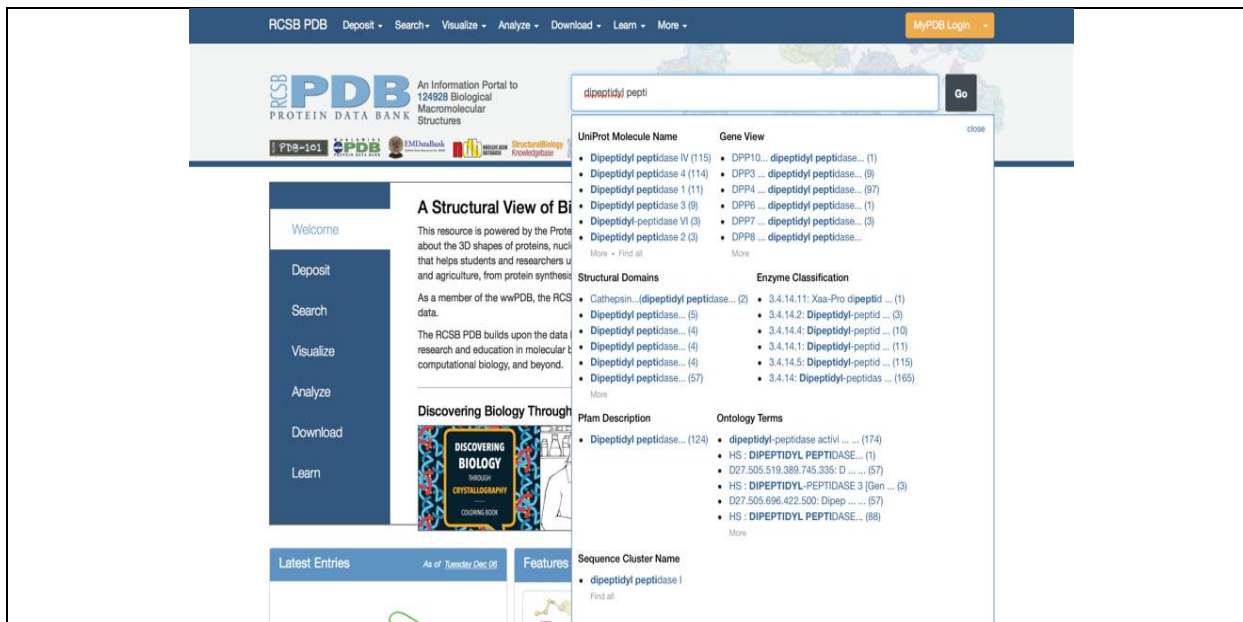
Results and Discussion

Designing Molecules

Natalie 1

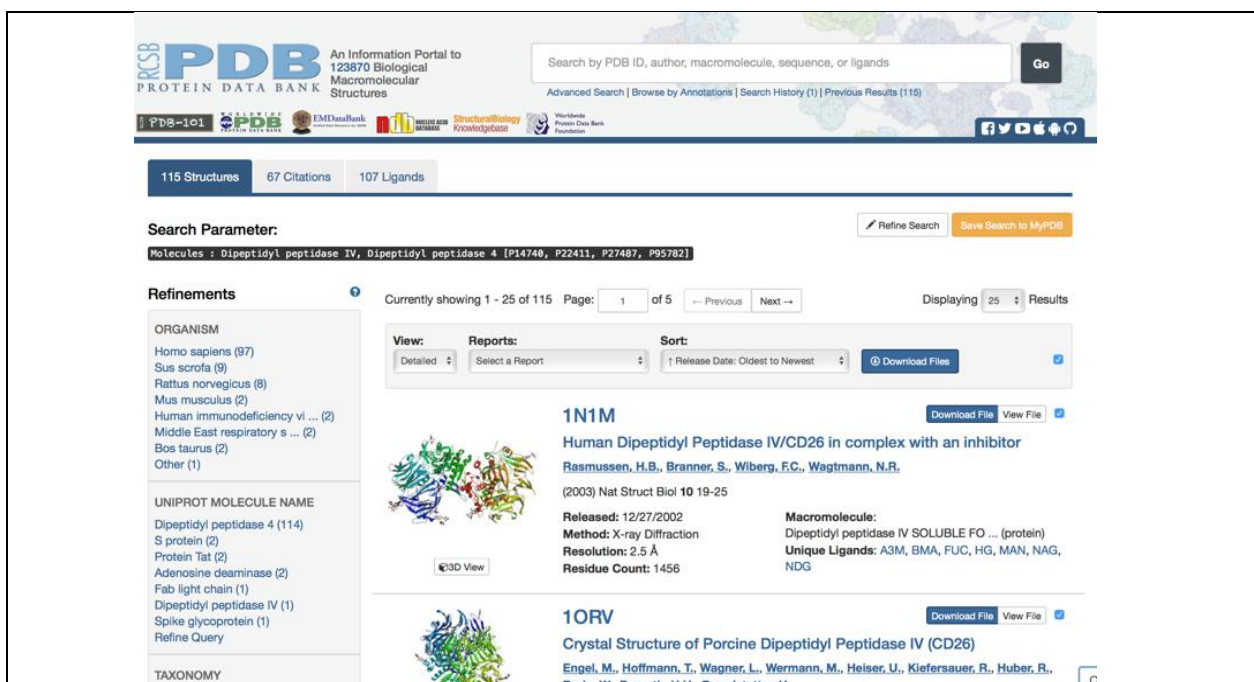
Using the Protein Data Bank

To utilize the Protein Data Bank, the user must simply have access to the Internet, and enter in the web address, which is RSCB.org. Once the homepage is open, the user can simply type in some keywords into the search box, which is located at the top right-hand corner of the homepage. The user can look up macromolecules, proteins, ligands, and so on. For our purposes, I would search for my target protein, Dipeptidyl Peptidase IV. Immediately, the website will show some suggested searches based on what I am typing into the search box (Figure 6), and you can see that Dipeptidyl Peptidase IV is the first thing that is appearing as a suggested search. If the user clicks on that blue link, they will be brought to a page that lists all entries involving DPP4 (Figure 7). Looking near the top left-hand side of the page, under the PDB logo, it shows that there is a total of 115 structures of DPP4 archived on the PDB. Going across to the right, it also shows that there are 67 citations with DPP4, and a total of 107 ligands archived on the PDB that are known to interact with DPP4. In clicking on this tab, the website will show a page which lists each ligand individually. In selecting a ligand, the page will show what PDB entries the ligand is present in, interacting with DPP4 (Figure 8). This was very important for our research; this page listing these ligands that interact with DPP4 is what we would reference back to when we are designing a new molecule to inhibit DPP4.



(Figure 6)

The PDB will suggest searches based on keywords being entered in the search bar. Here, DPP4 is the first suggested search, showing 115 structures archived on the PDB.



(Figure 7)

Clicking on the “DPP4” link from Figure 6, this is the resulting page. The top, left-hand corner of the page shows that, again, there are 115 structures of DPP4 archived on the PDB, 67 citations including DPP4, and 107 ligands archived that are known to interact with DPP4.

The screenshot shows the RCSB PDB website interface. At the top, there are navigation links: Deposit, Search, Visualize, Analyze, Download, Learn, and More. A search bar is present with the text "Search by PDB ID, author, macromolecule, sequence, or ligands". Below the search bar, there are statistics: 115 Structures, 67 Citations, and 107 Ligands. The search parameter is "Molecules : Dipeptidyl peptidase IV, Dipeptidyl peptidase 4 [P14740, P22411, P27407, P25702]". The results are displayed in a table with columns for Ligand Structure, ID / Formula / Name, and Structures with Specific Ligands.

Ligand Structure	ID / Formula / Name	Structures with Specific Ligands
	007 C12 H17 N 1-METHYLAMINE-1-BENZYL-CYCLOPENTANE	1 PDB Structures contains 007 (ZBUA ...)
	008 C20 H22 F N3 O2 (S)-2-((R)-3-AMINO-4-(2-FLUORO-PHENYL)-BUTYRYL)-1,2,3,4-TETRAHYDRO-ISQUINOLINE-3-CARBOXYLIC ACID AMIDE	1 PDB Structures contains 008 (ZBUC ...)
	01T C22 H30 N2 O2 [5-(aminomethyl)-6-(2,2-dimethylpropyl)-2-ethyl-4-(4-methylphenyl)pyridin-3-yl]acetic acid	1 PDB Structures contains 01T (BO95 ...)

(Figure 8)

Clicking on the tab that showed “107 ligands” interacting with DPP4, this is the resulting page. This page lists each ligand that is known to interact with DPP4, as well as what PDB entries they are found in. This page was important in our research; this is what we would refer to for ideas when designing new molecules.

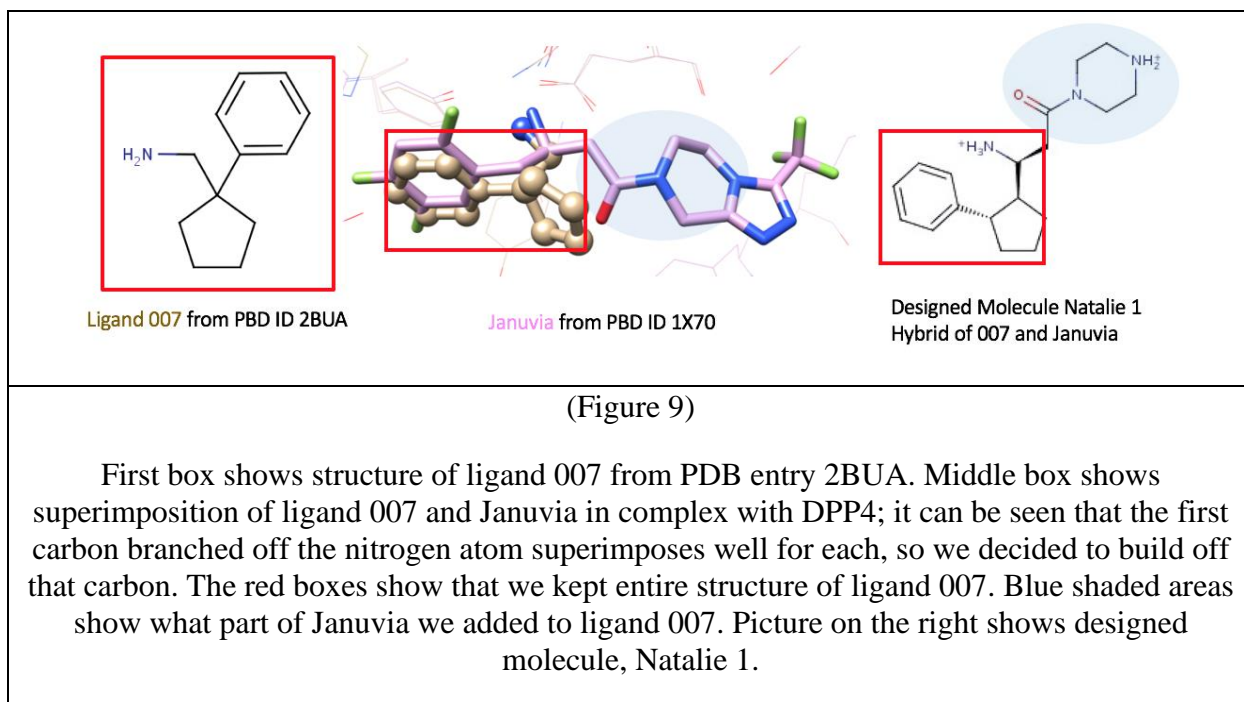
Using UCSF Chimera for Viewing/Designing Molecules

First, the user must download UCSF Chimera online and install the program. Once that is complete, Chimera can be used to extract the 3D structures of molecules from the PDB for viewing purposes. First, the user must launch Chimera. Once open, select “File” in the toolbar, and then select “Fetch by ID.” A small window will pop up with several search options; because we are pulling structures from the protein data bank, it is important Chimera is “fetching” from the PDB. Once “PDB” is selected, the user can search for the entry of interest. For the purposes of this research, one of the first entries I studied was PDB ID 1X70, which is DPP4 in complex with Sitagliptin, otherwise known as Januvia. Chimera will automatically open the 3D structure of DPP4 with Januvia in the protein pocket.

Next, we want to superimpose another small ligand in complex with DPP4 with Januvia; from that point we can move forward in creating some designed molecule. Following the same protocols in “fetching” Januvia in complex with DPP4, we will fetch another molecule from the

PDB. In designing Natalie 1, we fetched PDB ID 2BUA, which has a small ligand, 007, in complex with DPP4. Once the working view is refined with only one chain of DPP4 in complex with ligand for each ID, the complexes can then be superimposed.

From this superimposition, we decided to keep the entire structure of ligand 007 and build off that molecule using a part of Januvia. Ultimately we designed Natalie 1, a hybrid of ligand 007 and Januvia. This design process can be seen in Figure 9. Based on the superimposition, it was clear that the first carbon branched off the basic nitrogen atom on each atom superimposed almost perfectly; so, we decided to keep the entire ligand 007 and branch from that very carbon atom. To that atom, we added some parts of Januvia.



The next step in our design process is to build the molecule, which we did using ChemDraw 2D and ChemBio 3D software.

Using ChemDraw Software to Build Molecules

ChemDraw 2D and ChemBio 3D were software programs provided by Middlesex County College that I used to create the 2D and 3D structures of my molecules. Once we had the design of the molecule thought up, we then used ChemDraw 2D to draw the structure of that molecule. Once the 2D version was drawn up, we would then save that ChemDraw 2D file and open that file up using ChemBio 3D. This would immediately create the 3D version of that molecule.

From here, we would minimize the energy of the molecule, where the software would create the most stable form of the molecule. We would do so by highlighting the molecule, then going to “Calculations” in the toolbar, selecting “MM2,” then selecting “Minimize energy.” You will see the program adjusting the molecule, ultimately leaving the user with the most stable conformation of that molecule. Once that is done, the user can then save the .Mol2 file which will be used later in docking.

The last thing we used ChemBio 3D for is saving the designed molecule’s SMILES string. The SMILES string is a unique identifier for any molecule; just like people have social security numbers to identify them, molecules all have a SMILES string. To save the molecule’s string, the structure must first be selected using the selection tool. Then, from the Edit menu in the toolbar, select “Copy As” and choose SMILES. The string will automatically copy to your clipboard; good practice is to paste this smiles string into a Word Document and saving it along with other files related to that molecule. The SMILES string brings us to the next phase in our designing a molecule, which is testing the molecule to ensure that it does indeed interact with our target protein, DPP4. We can do this using SwissTargetPrediction.

Using SwissTargetPrediction

SwissTargetPrediction is another bioinformatic tool which we used to run checks to ensure the designed molecule does indeed interact with DPP4, the target protein. The tool can be accessed through the Internet at Swisstargetprediction.ch (Figure 10). The SMILES string from ChemBio 3D just gets pasted into the box provided, and within seconds the tool will provide a report showing what proteins the molecule is predicted to interact with.

Figure 10: SwissTargetPrediction landing page. Here, the target organism can be selected, and the SMILES string is pasted into the given box. Then select “Submit”

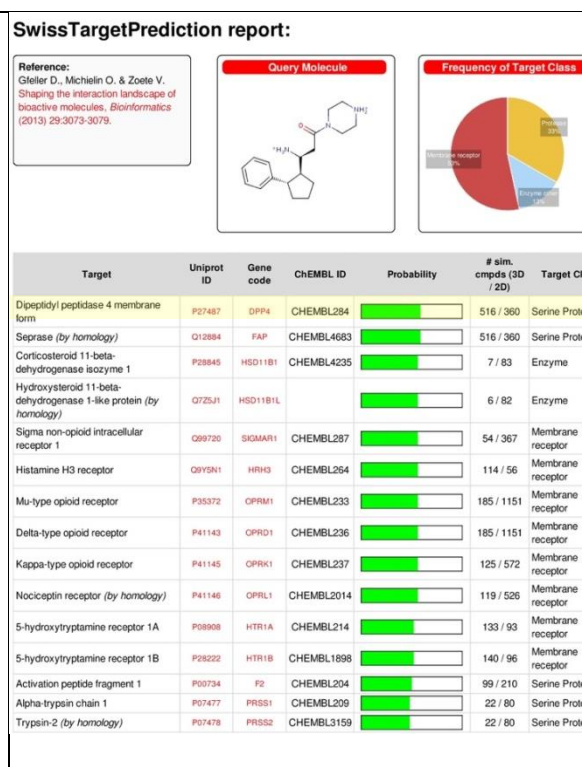


Figure 11: Generated SwissTargetPrediction of Natalie 1 for DPP4. DPP4 is the first protein showing predicted interaction, with just above 50% probability of interacting

The report provided by SwissTargetPrediction will list each protein the ligand is predicted to interact with, as well as the probability of interaction (green bars in Figure 11). Based on the results for Natalie 1 (Figure 11), it is worth moving forward with the molecule and docking it in DPP4 because it is indeed predicted to interact with a good probability. The small molecule is also predicted to interact with other proteins; if this molecule were a drug these other interactions could cause side effects, something that isn't desirable in a drug.

Using SwissDock to Dock Molecules

If the SwissTargetPrediction report looks promising, the next step in the design process is to dock the designed molecule in DPP4 to view the predicted intermolecular interactions, which is done using SwissDock. SwissDock can be accessed at Swissdock.ch, then selecting the “Submit Docking” tab. From this page, the user must input information necessary for the tool to “run the docking” (Figure 12).

(Figure 12)

SwissDock “Submit Docking” page where information about docking job is input by the user. Once all information is entered, then job will begin by selected “Submit Docking”

A file of the protein must be prepared prior to running docking. To do so, a PDB entry containing the target protein must be opened in Chimera and saved as a .PDB file. For example, entry 1X70 that is Januvia in complex with DPP4 would work fine. Once the file is open in Chimera, every nonstandard ligand must be deleted. In the toolbar, go to “select,” then hover over “residue” and finally select “all nonstandard.” Chimera will automatically highlight all small molecules or atoms that do not belong to the protein structure. Then, with those parts highlighted, go to “actions” in the toolbar, hover over “atoms/bonds” and select “delete.” Now, the only thing on file will be the protein, DPP4. In the toolbar, select “tools” then hover over “structure editing” and perform the following; first, select “AddH.” A small window will pop up, press “Ok.” Next, select “Add Charge.” Again, a small window will pop up, simply select “Ok.” Last, select “Dock Prep.” A small window will pop up, every option in that window should be checked. Then, press “Ok,” and then “Ok” again on the next window. Chimera will now prompt the user to save the file as a .Mol2 file. Once saved, the user must open this .Mol2 file and save it again, this time as a .PDB file.

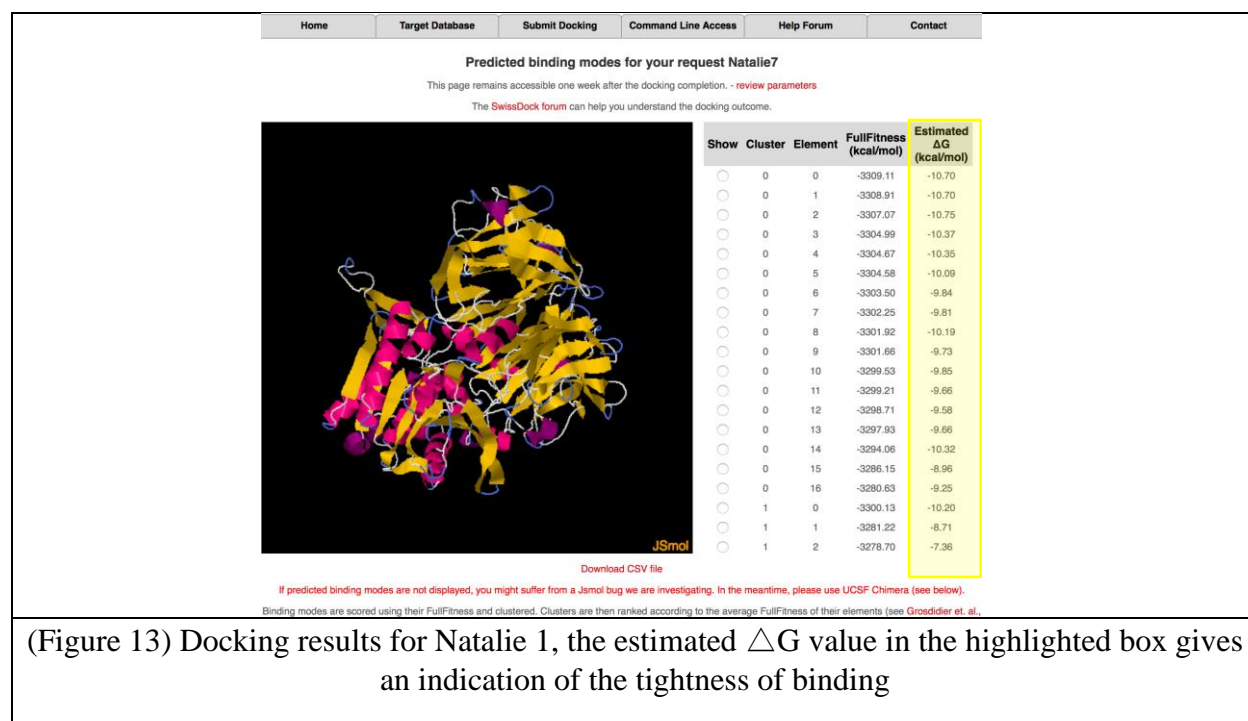
Under *Target Selection*, the .PDB file of the protein must be uploaded; this file is created using Chimera. SwissDock will run checks on the file to ensure it is ready for docking, then will issue a green checkmark when the file is deemed ready for docking. Under *Ligand selection*, the .Mol2 file of the minimized structure of Natalie 1 is uploaded, SwissDock will again run checks on the file.

Under *Description*, a job name must be created, such as “Natalie1,” and an email address should be entered so the website can inform the user of both the starting and completion of a job.

Then, “show extra parameters” must be click, which will then show an area in which additional information can be entered.

Here, the *Docking type* should be set to “Accurate.” To *Define the region of interest*, some background work must be completed first. The user needs to identify the boundary and location of the pocket for docking. Next, identify the center of the pocket, or the x-y-z coordinates of the protein pocket, this way we can clearly define the site at which docking should be happening. For DPP4, these coordinates are 39, 51, and 37, respectively. This position corresponds to the basic nitrogen atom on the small ligand, like Januvia, which is locked in by Glutamates 205 and 206. We found these coordinates using Chimera and the PDB. In chimera, we can locate this nitrogen atom on Januvia in the pocket of the protein. In doing so, it can be found that this nitrogen is identified as “N20” for entry 1X70. Going back to the PDB and opening the PDB file for entry 1X70 and searching for “N20” in ligand “715,” which is Januvia, the coordinates are given. Next, SwissDock must be informed what the size of the protein pocket is. For DPP4, setting the x-y-z sizes to 5 will work. Finally, under *Flexibility*, the user must identify some level of flexibility SwissDock should run the job under. This should be set to 3 angstroms, allowing for a moderate level of flexibility for both the protein residues and the ligand. If the binding pocket is not clarified, SwissDock will attempt to run “blind docking,” where it will dock the small molecule all over the protein, rather than just at the active site, or pocket of the protein. It is important that this information is given before running docking.

Once this information is inputted, the job can be submitted. Typically, jobs will take a few hours to be completed. Upon completion, the user should receive an email with a link to the docking results (Figure 13).

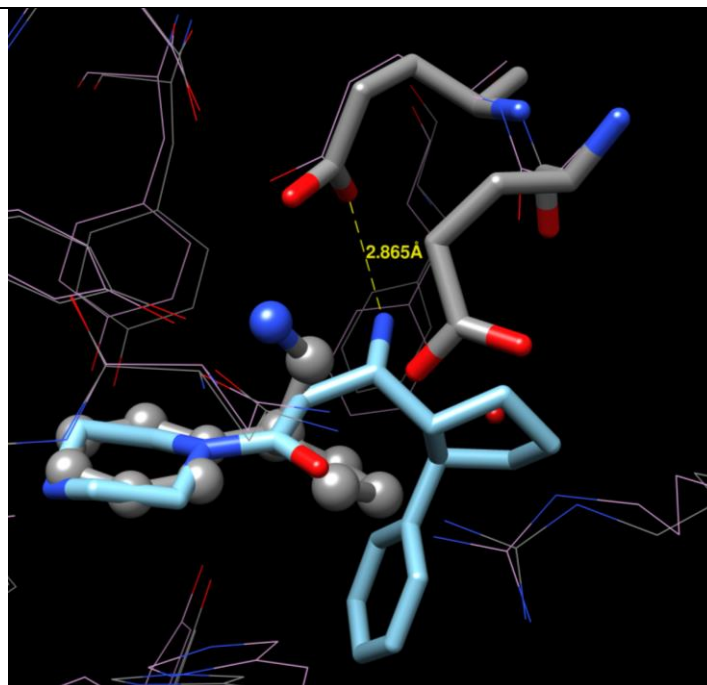


Looking at this page, the lines show the many different poses, or predicted orientations of the small molecule in the active site of the protein. From this webpage, the user can interactively select different poses and see the ligand moving in the pocket of the protein from pose to pose, although Chimera is certainly preferred for viewing and analyzing poses. There can be hundreds of poses predicted for any given docking job; selected one best-fit pose may seem intimidating but that is not the case. The first criteria considered for narrowing down good poses is the Estimated ΔG value. This value tells us about the stability of the binding mode; the more negative this value, the better the binding mode. This negative value indicates a release of energy, so the greater the release of energy indicates the more binding affinity. Next, the results should be analyzed further using Chimera. From this webpage, there is a red link that reads “Launch UCSF Chimera to visualize predicted binding modes,” which will automatically open the job’s docking results in Chimera.

Using UCSF Chimera to analyze Docking Results

To analyze docking results, the job should be superimposed with structures that were used to design the molecule. For Natalie 1, we superimposed the docking job with ligand 007 from PDB entry 2BUA to analyze the results. Next, each pose had to be considered in the pocket of the protein to see which does make sense, and to ensure that whatever pose is chosen, that it does indeed make the essential hydrogen-bonds with residues GLU 205 and 206 on the protein.

After analyzing Natalie 1 superimposed with ligand 007, results showed that the best-fit pose in the pocket of the protein had a ΔG value of -7.081, which is not great. It was also found that only one of the two essential hydrogen-bonds for inhibition is being made, which means this is not a tight bind to the protein. In conclusion, this was a promiscuous ligand that was predicted to bind to DPP4 with good probability, yet it does not bond tightly to the protein. This molecule has much room for improvement. The best fit pose for Natalie 1 in the pocket of DPP4 can be seen below, in Figure 14, where Natalie 1 is shown in blue and ligand 007 is shown in grey ball and stick.

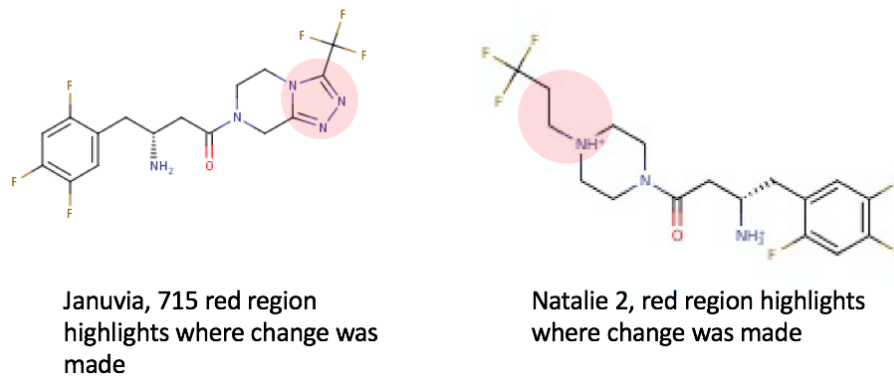


(Figure 14)

Best docked pose of Natalie 1, shown in blue stick, superimposed with ligand 007, grey ball and stick, from PDB entry 2BUA.

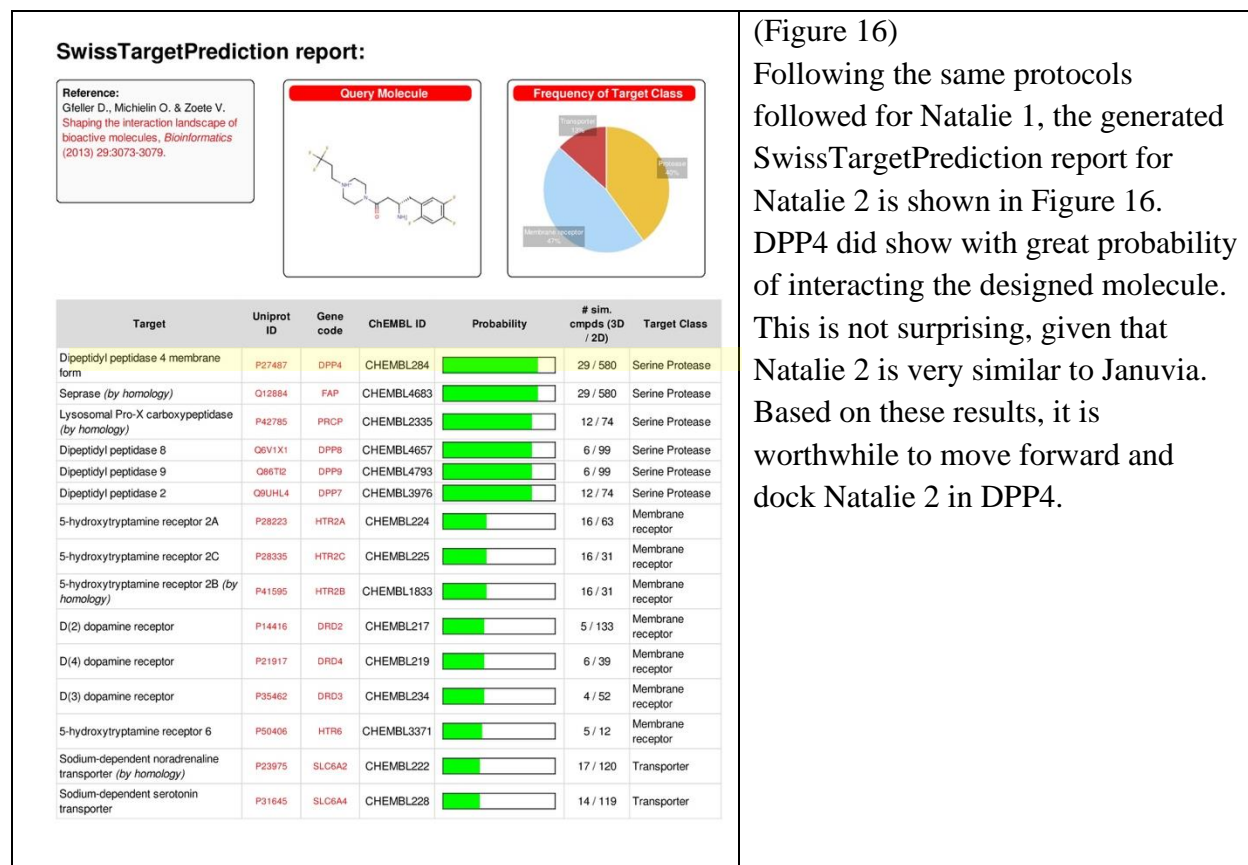
Natalie 2

Natalie 2 was a simple design, based on making a small change to the structure of Januvia. We simply changed one ring structure within the ligand, making it a carbon chain instead. The design process can be seen below in Figure 15.



(Figure 15)

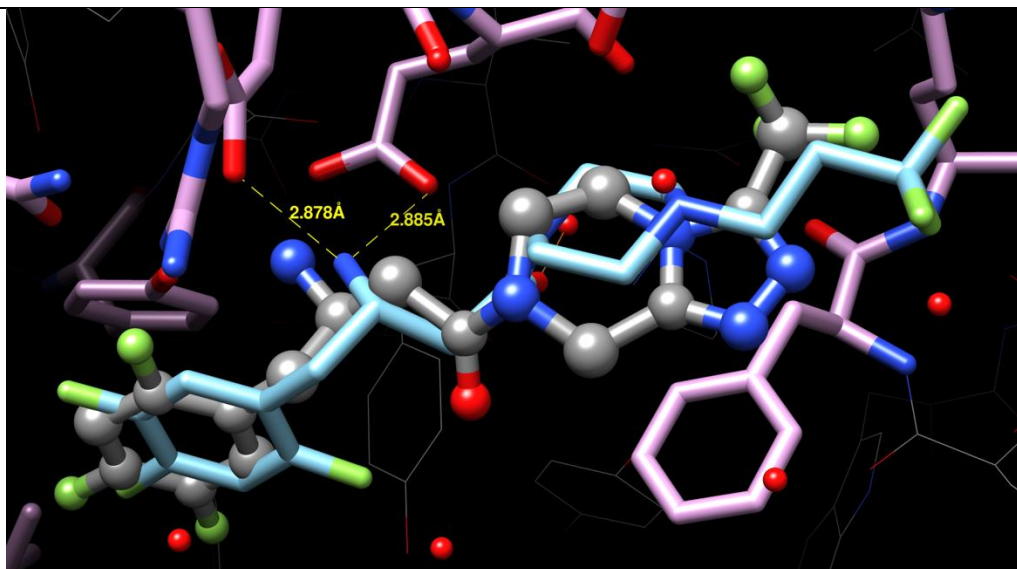
Red shaded area shows where change was made to Januvia to create Natalie 2



(Figure 16)

Following the same protocols followed for Natalie 1, the generated SwissTargetPrediction report for Natalie 2 is shown in Figure 16. DPP4 did show with great probability of interacting the designed molecule. This is not surprising, given that Natalie 2 is very similar to Januvia. Based on these results, it is worthwhile to move forward and dock Natalie 2 in DPP4.

The docking results with Natalie 2 superimposed with Januvia in complex with DPP4 can be seen below in Figure 17. In analyzing the results, one pose was chosen as the best fit for DPP4. The molecule does indeed make both necessary hydrogen-bonds to GLU205 and 206 on the protein, and has a very reasonable ΔG value of -10.77. Of course, though, some changes can be made to this molecule; perhaps a change going forward would be to make this molecule less like Januvia, designing it again, but based on a different ligand.

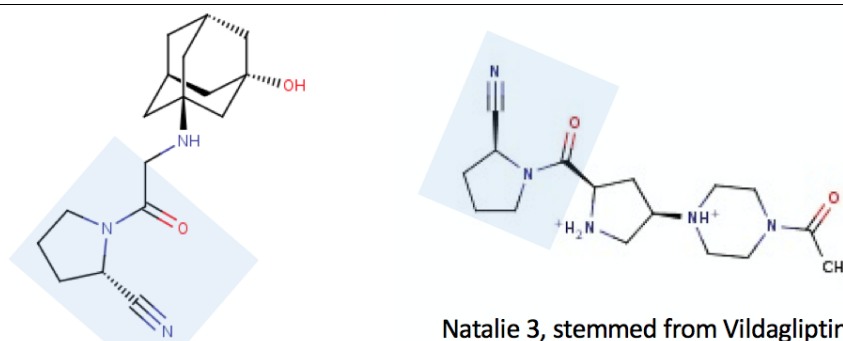


(Figure 17)

Best docking pose of Natalie 2 (blue stick) superimposed with Januvia (grey ball and stick) from PDB entry 1X70 in complex with DPP4

Natalie 3

Natalie 3 was designed based on the structure of Vildagliptin, which is a drug inhibitor of DPP4 marketed under the name Galvus. The design process can be seen below in Figure 18.

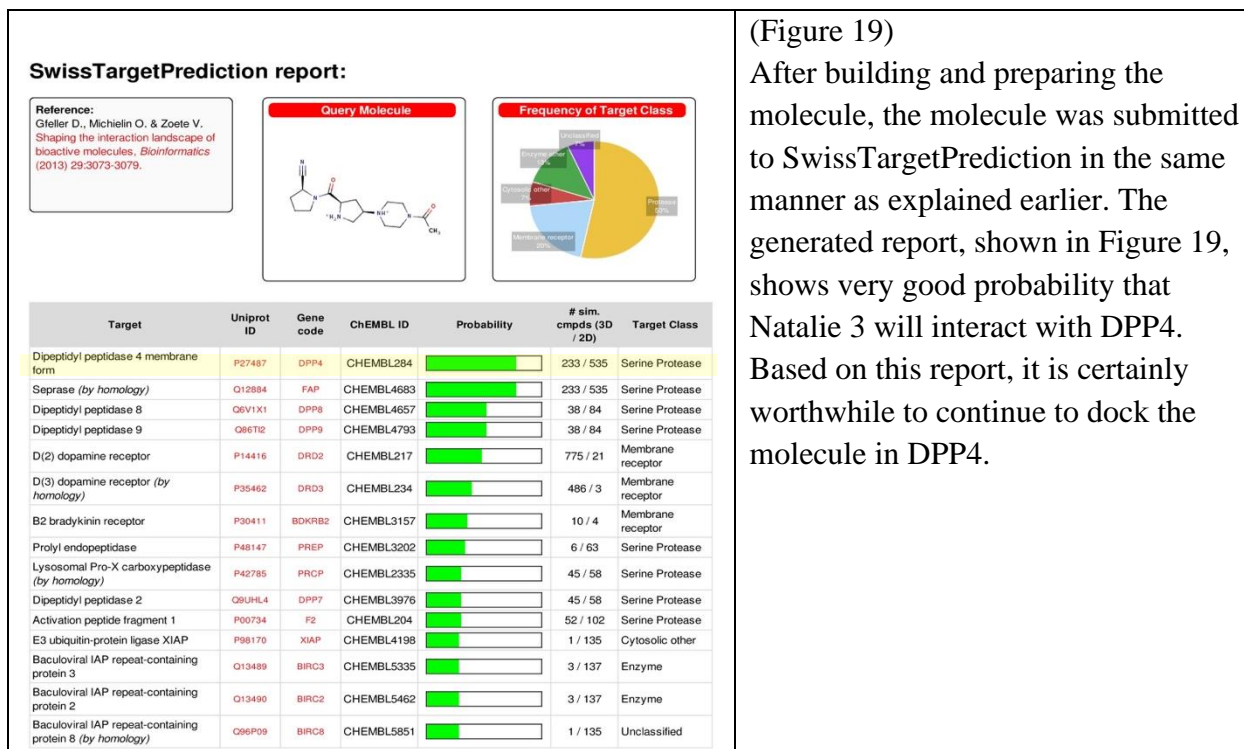


Blue box highlights what part of Vildagliptin (LF7) extracted for new molecule

Natalie 3, stemmed from Vildagliptin

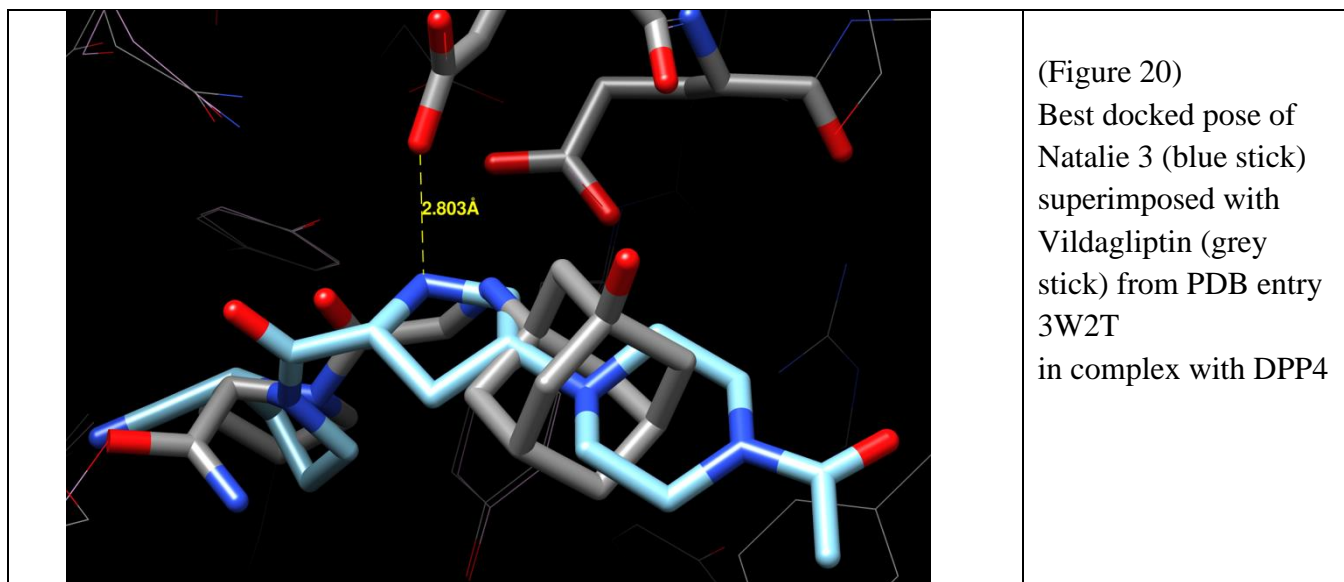
(Figure 18)

Blue shaded area shows area of Vildagliptin, PDB entry 3W2T, used to create designed molecule, Natalie 3



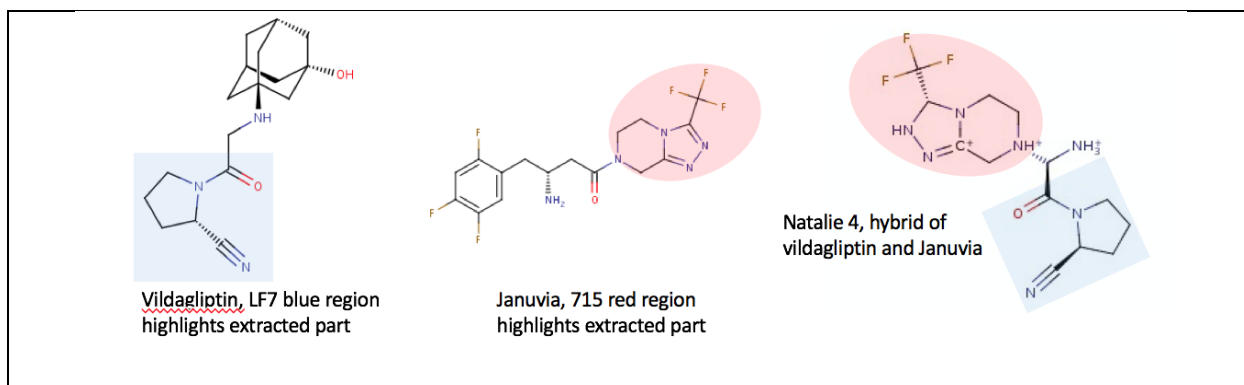
(Figure 19)
 After building and preparing the molecule, the molecule was submitted to SwissTargetPrediction in the same manner as explained earlier. The generated report, shown in Figure 19, shows very good probability that Natalie 3 will interact with DPP4. Based on this report, it is certainly worthwhile to continue to dock the molecule in DPP4.

To analyze the docking results, Natalie 3 was superimposed with Vildagliptin in Chimera. The results are shown below in Figure 20. This molecule only makes one of the two necessary hydrogen-bonds to DPP4, binding to GLU 205 but not GLU 206. This pose has a ΔG value of -7.685, which isn't great. Based on these results, there is certainly room for improvement in the design of this molecule.



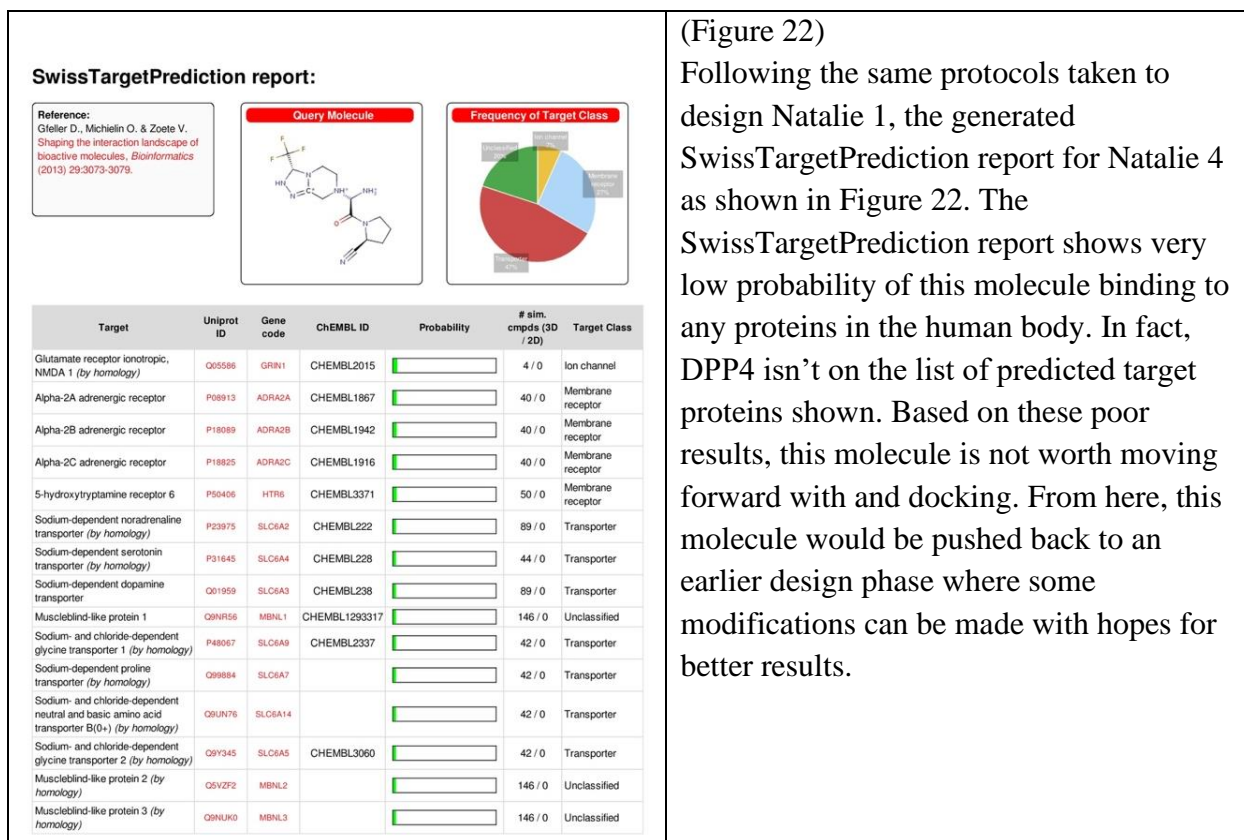
Natalie 4

Natalie 4 was mistakenly designed using 2D structure design, rather than the 3D structure based drug design that this research is centered around. This molecule was designed based on the 2D structures of known binding “Gliptins.” Because this was done through 2D structure design, there was no superimposition of 3D structures in Chimera, as was done with the other designed molecules. Figure 21 shows the design process. The red and blue highlighted parts of the two drugs were joined together to make Natalie 4.



(Figure 21)

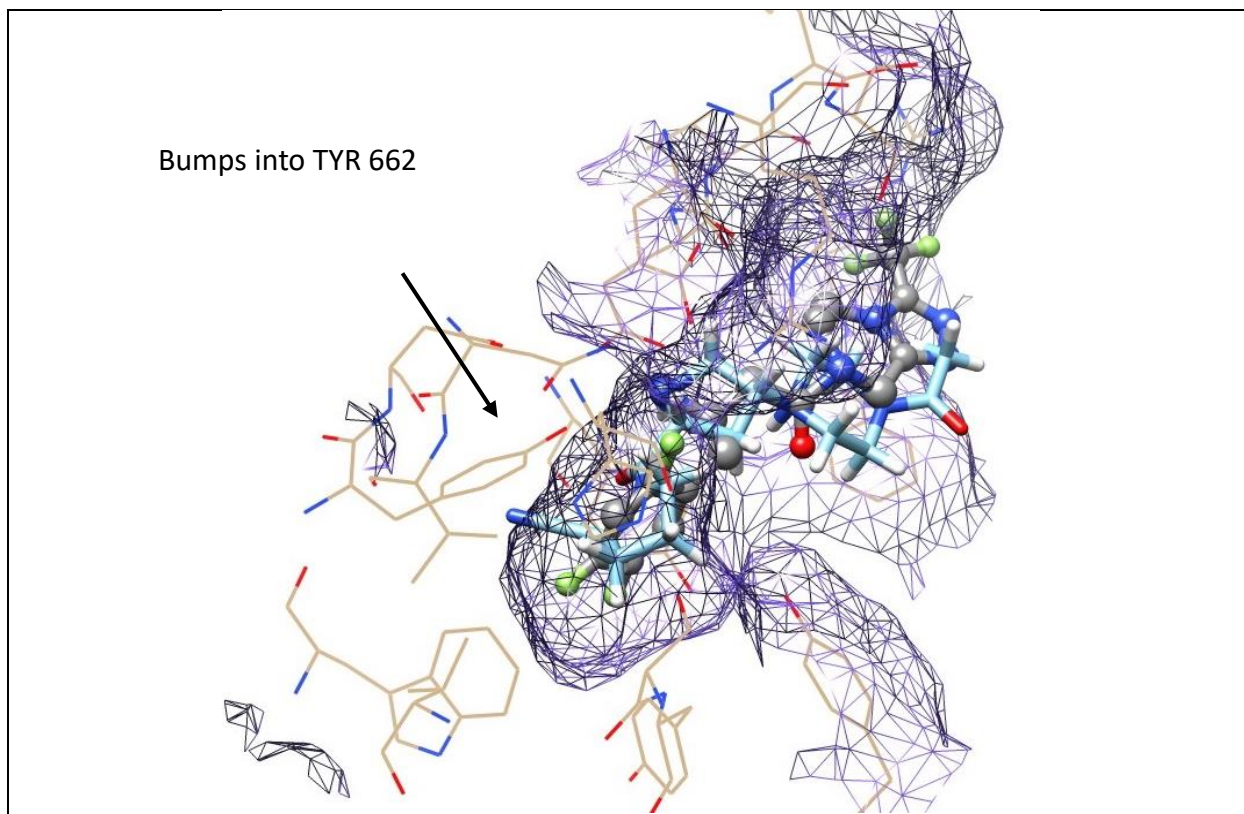
Blue shaded area show part of Vildagliptin from PDB entry 3W2T used in design. Red shaded area shows part of Januvia, PDB entry 1X70, used in design. Natalie 4 is a hybrid of Vildagliptin and Sitagliptin



(Figure 22)

Following the same protocols taken to design Natalie 1, the generated SwissTargetPrediction report for Natalie 4 as shown in Figure 22. The SwissTargetPrediction report shows very low probability of this molecule binding to any proteins in the human body. In fact, DPP4 isn't on the list of predicted target proteins shown. Based on these poor results, this molecule is not worth moving forward with and docking. From here, this molecule would be pushed back to an earlier design phase where some modifications can be made with hopes for better results.

Figure 23 below shows why Natalie 4 failed. Looking at the picture below from Chimera, the pocket of the protein is lined with a dark mesh, and the protein residues are shown in beige sticks. Looking at the pocket of the protein, a small atom can be seen poking through; this tells us that the molecule does not fit in the pocket of the protein, in fact it is bumping into residue TYR 662 on the protein. Knowing this, perhaps a modification to the molecule that would allow it to be more flexible would better its chances of docking to DPP4.

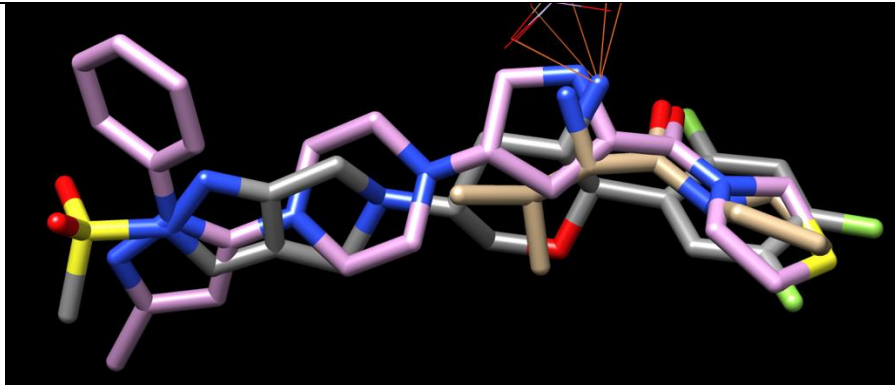


(Figure 23)

Beige sticks are the residues on the protein, DPP4. Purple mesh area shows the lining of the protein pocket. Natalie 4 is superimposed with Januvia in the protein pocket. Arrow points to part of Natalie 4 that is sticking through the pocket (nitrile group) and bumping into residue TRY 662 on DPP4. Natalie 4 would never fit in the active site of DPP4.

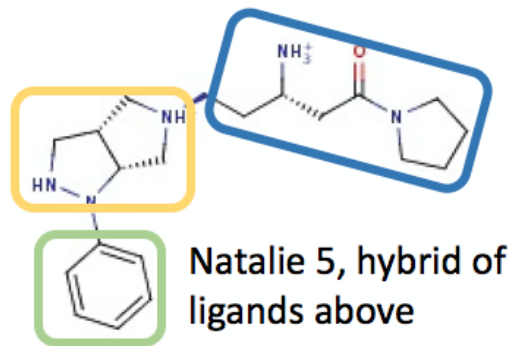
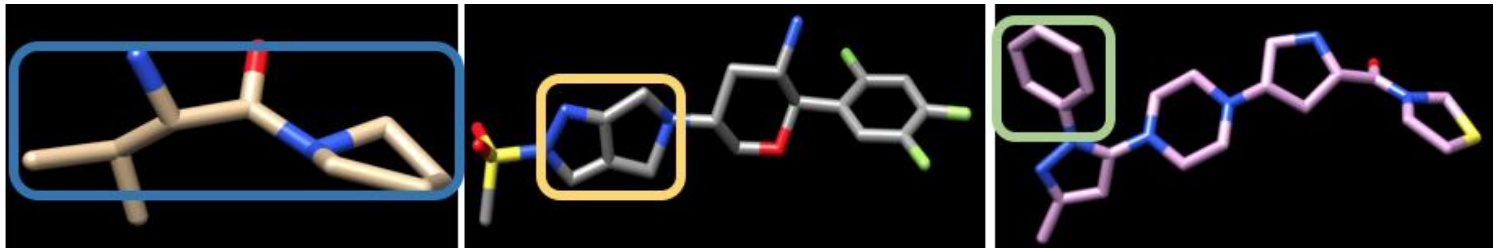
Natalie 5

Natalie 5 was designed based on the superimposition of 3 small ligands from the following PDB entries: 1N1M, 4PNZ, and 3VJK. From this original superimposition (Figure 24), the design for Natalie 5 was created; small pieces of each ligand were cut and pasted together to create Natalie 5, a hybrid of 3 molecules. The design process can be seen below in Figure 25.



(Figure 24)

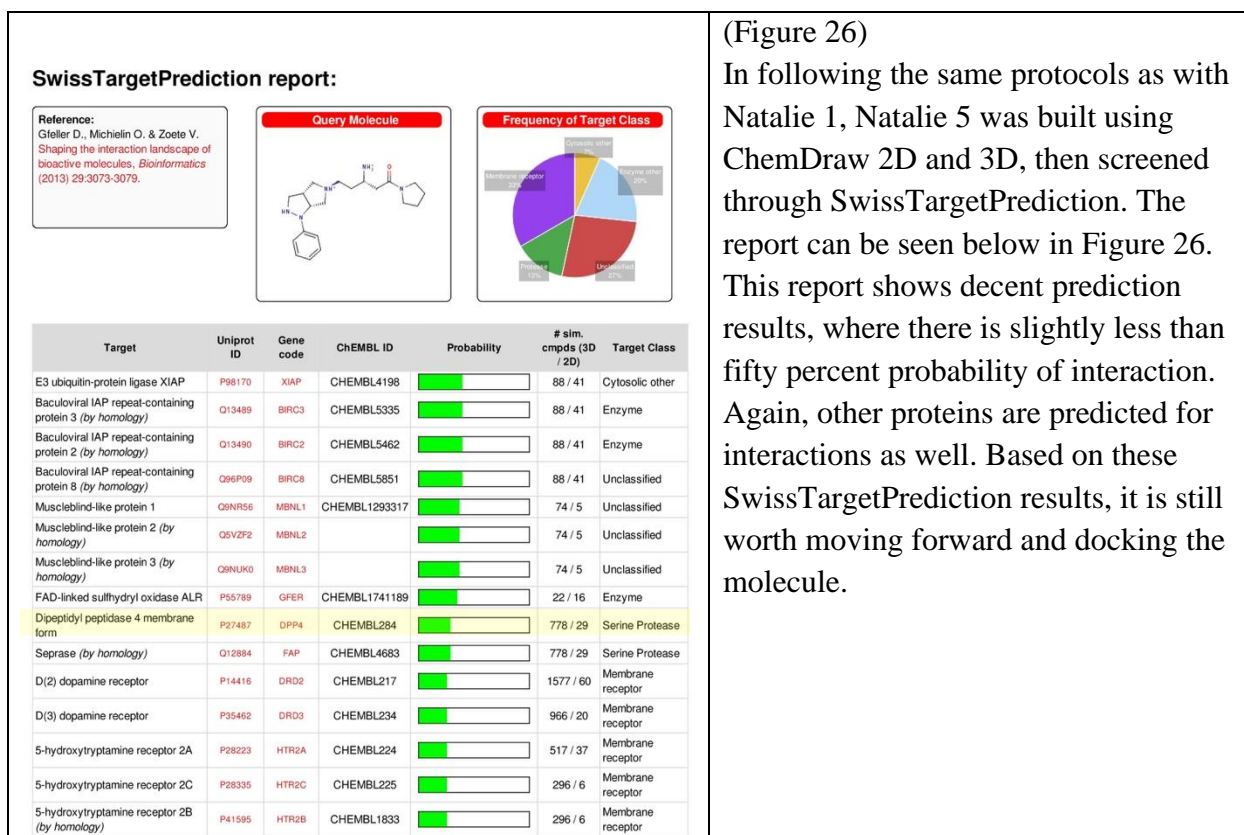
Superimposition of PDB IDs 1N1M (beige stick), 4PNZ (grey stick), 3VJK (pink stick) in Chimera.



(Figure 25)

Left to right: 1N1M, 4PNZ, 3VJK

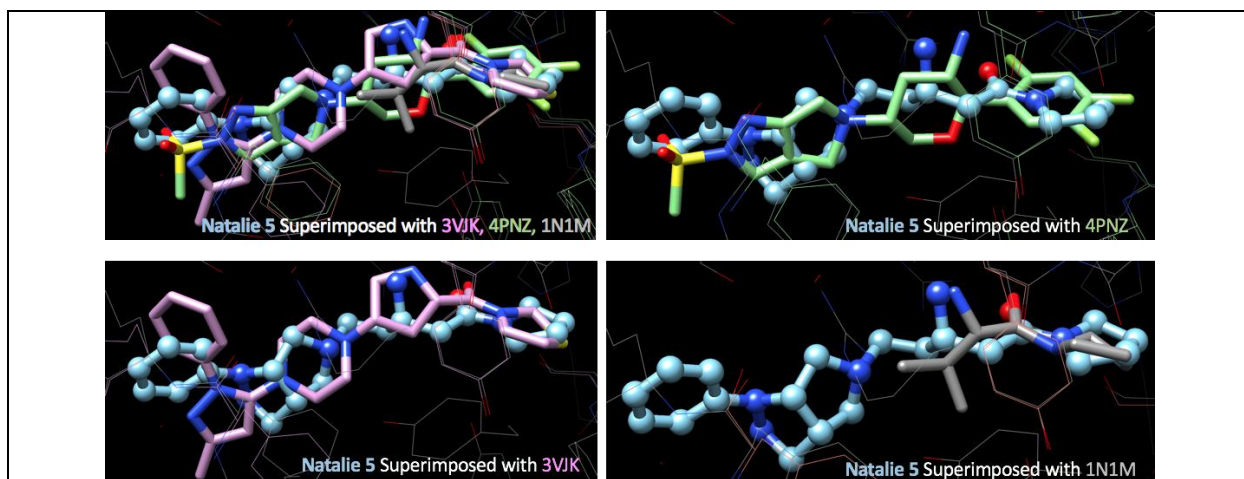
Color coded parts show what parts of each ligand were taken to create Natalie 5



(Figure 26)

In following the same protocols as with Natalie 1, Natalie 5 was built using ChemDraw 2D and 3D, then screened through SwissTargetPrediction. The report can be seen below in Figure 26. This report shows decent prediction results, where there is slightly less than fifty percent probability of interaction. Again, other proteins are predicted for interactions as well. Based on these SwissTargetPrediction results, it is still worth moving forward and docking the molecule.

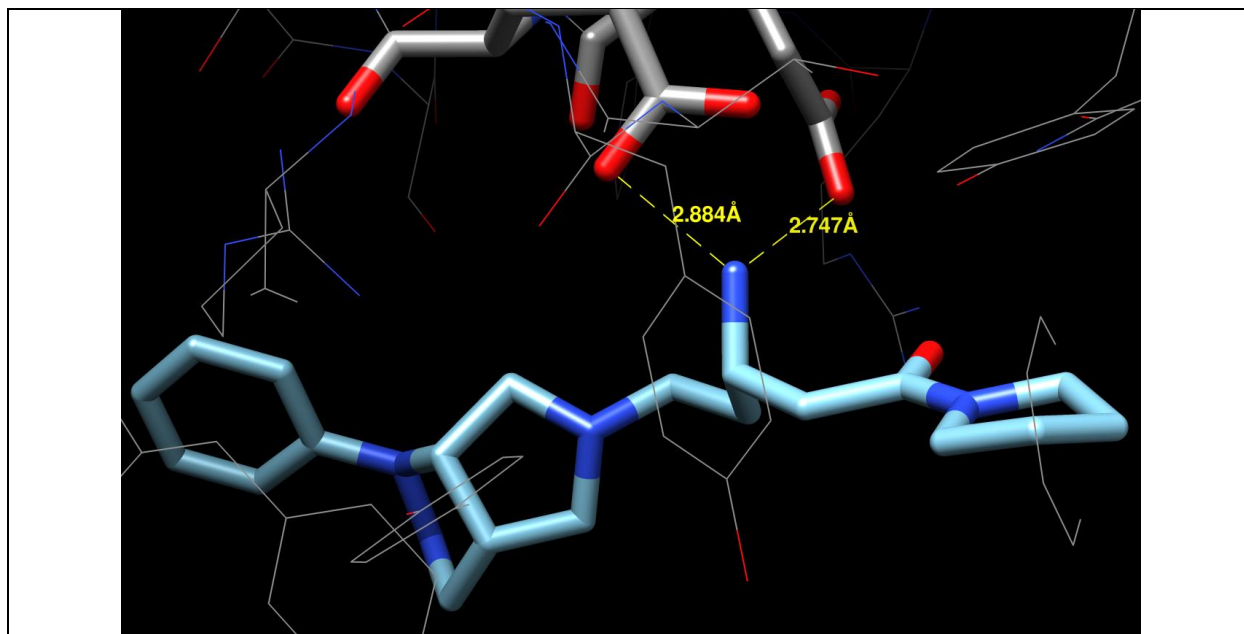
The docking results superimposed with the original PDB entry ligands can be seen below in Figure 27. The chosen pose superimposed with each individual ligand shows the binding mode is as expected for DPP4.



(Figure 27)

Best docked pose of Natalie 5 (blue ball and stick) superimposed with original ligands. Superimposition with each individual ligand shows good results for the designed molecule

In further analyzing the docking pose, it was found that both essential GLU 205 and 206 hydrogen-bonds are being formed with a ΔG value of -10.072. Despite sub-par results from SwissTargetPrediction, this molecule had great docking results; the binding mode does make sense and there is a tight bind. There is always room for improvement, though. Results are shown below in Figure 28.

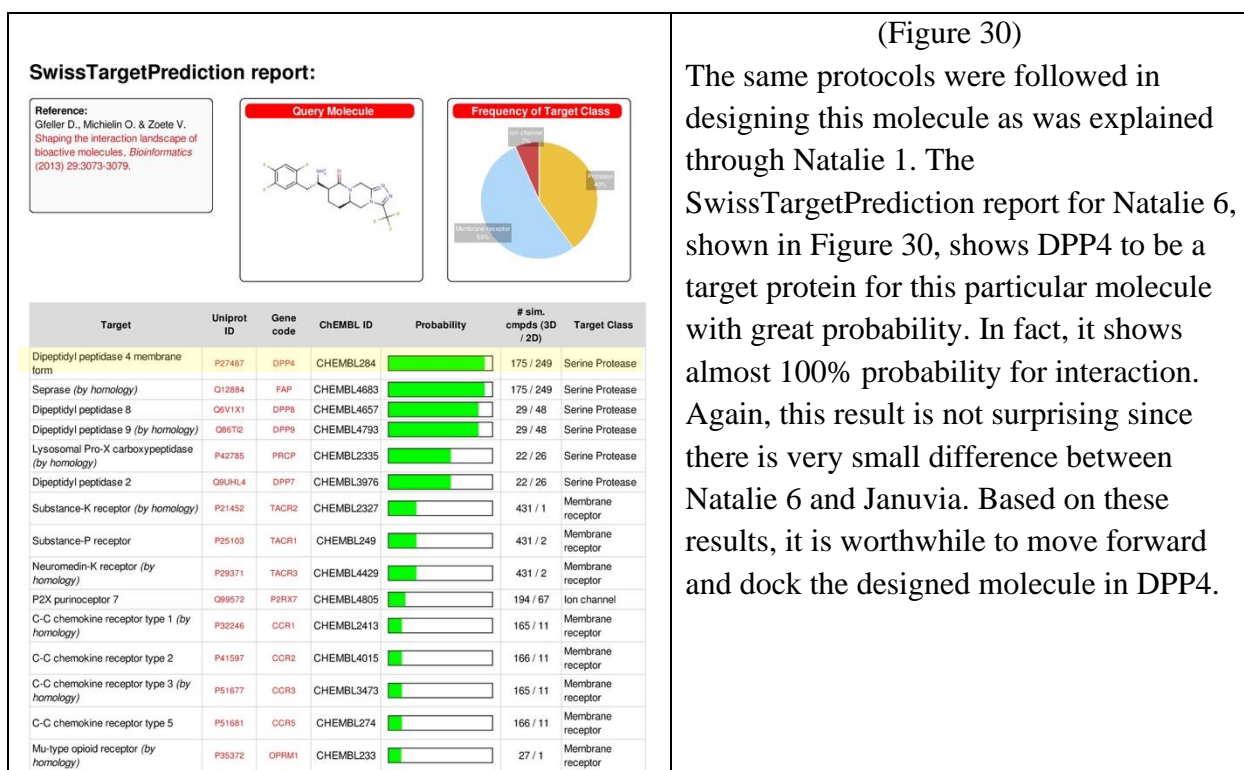
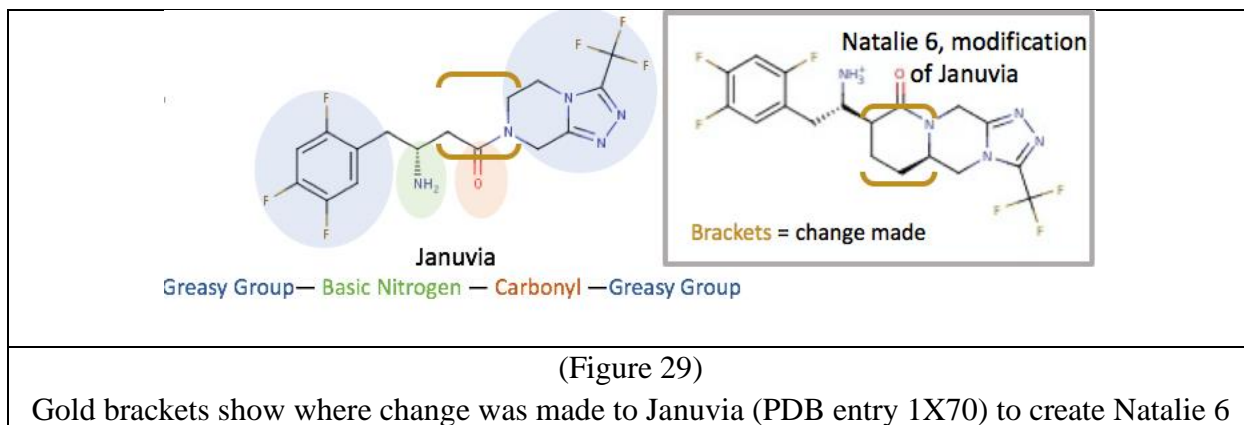


(Figure 28)

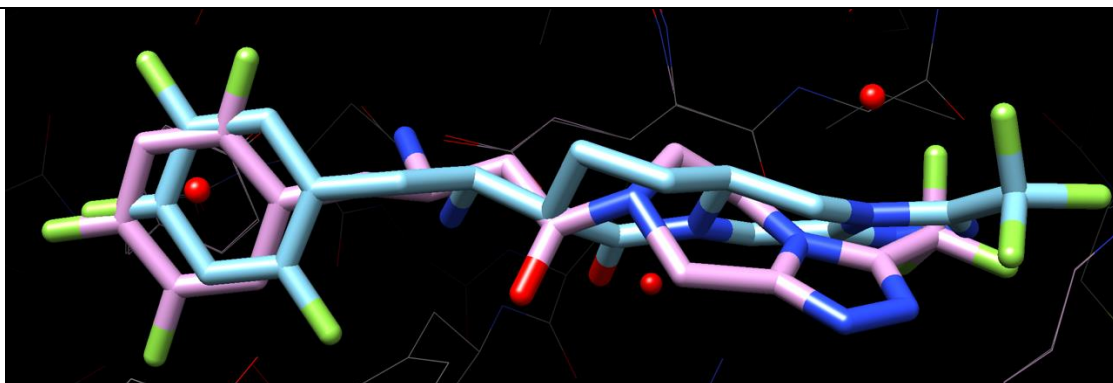
Natalie 5 docking results. Both essential hydrogen-bonds with GLU 205 and 206 are being made. Results show good binding for Natalie 5.

Natalie 6

Natalie 6 was designed in a simple, yet accurate fashion. The idea was to simply make a small a small change to the inside of Januvia. This was similar to the design approach taken in designing Natalie 2. The design process can be seen below in Figure 29.



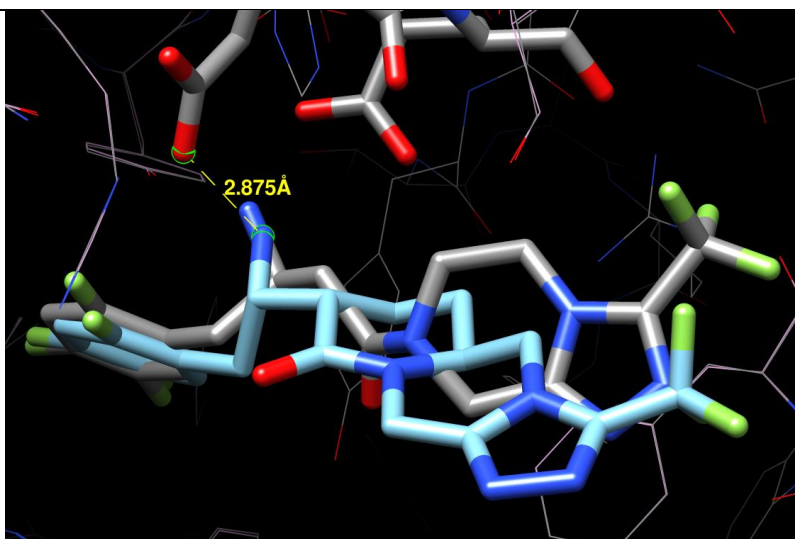
After docking was completed using SwissDock, Natalie 6 was superimposed with Januvia superimposed with DPP4 to check its binding mode. Figure 31 below shows that superimposition.



(Figure 31)

Best docking pose of Natalie 6 (blue stick) superimposed with Januvia (pink stick) in complex with DPP4.

Figure 32 below shows, again, Natalie 6 superimposed with Januvia in complex with DPP4. Here, it is shown that Natalie 6 only makes one of the two necessary hydrogen-bonds necessary for a tight bind to the protein. This binding has a ΔG value of -10.303. So, although the SwissTargetPrediction report had great results, docking shows otherwise. There is room for improvement moving forward.

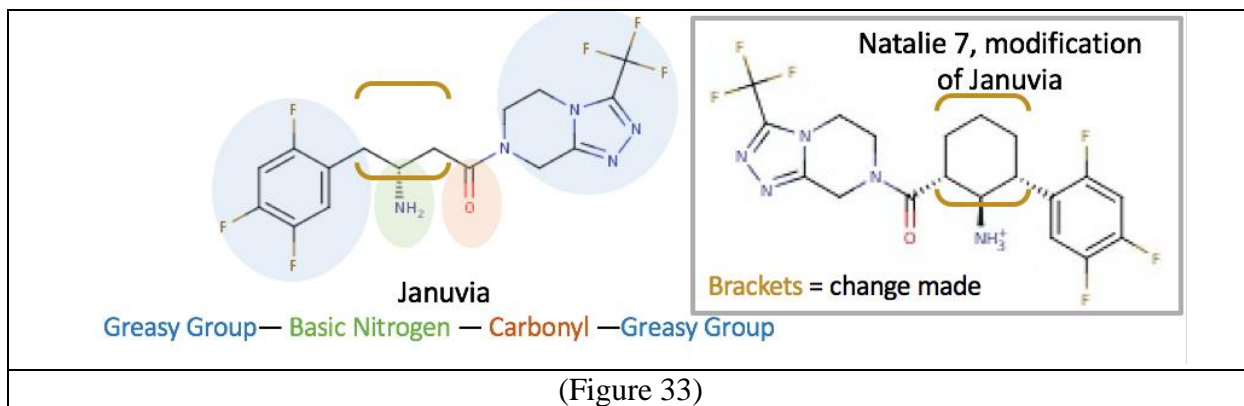


(Figure 32)

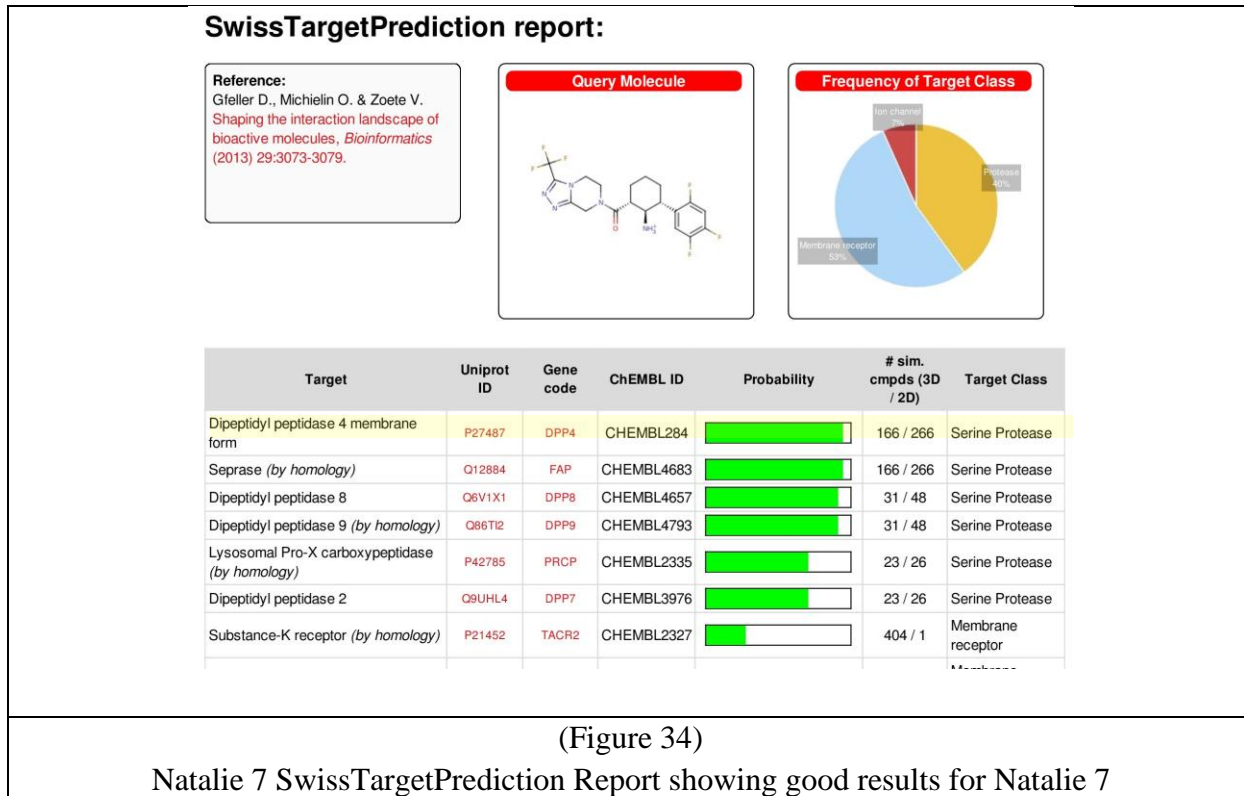
Natalie 6 docking results. Superimposition shows good results for deigned molecule, as was expected. However, only one of two essential hydrogen-bonds are being made. Not tight binding.

Natalie 7

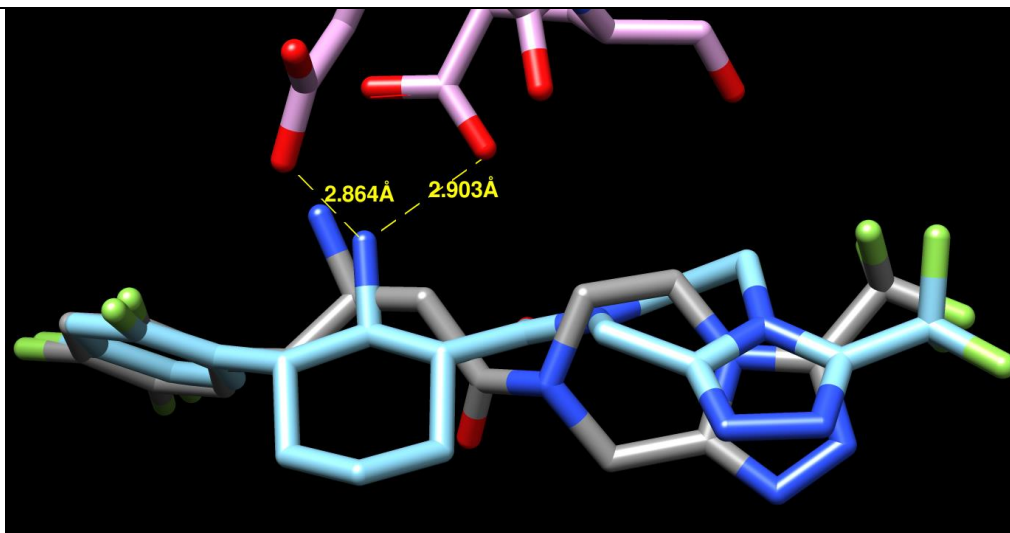
The approach taken to design Natalie 7 was the same as Natalie 6; to make small changes to the structure of Januvia while maintaining its overall structure, ensuring good docking results. The design process can be seen in Figure 33.



A section of the SwissTargetPrediction report for Natalie 7 can be seen in Figure 34. DPP4 is the first protein predicted for interaction, with almost an absolute probability of interaction.



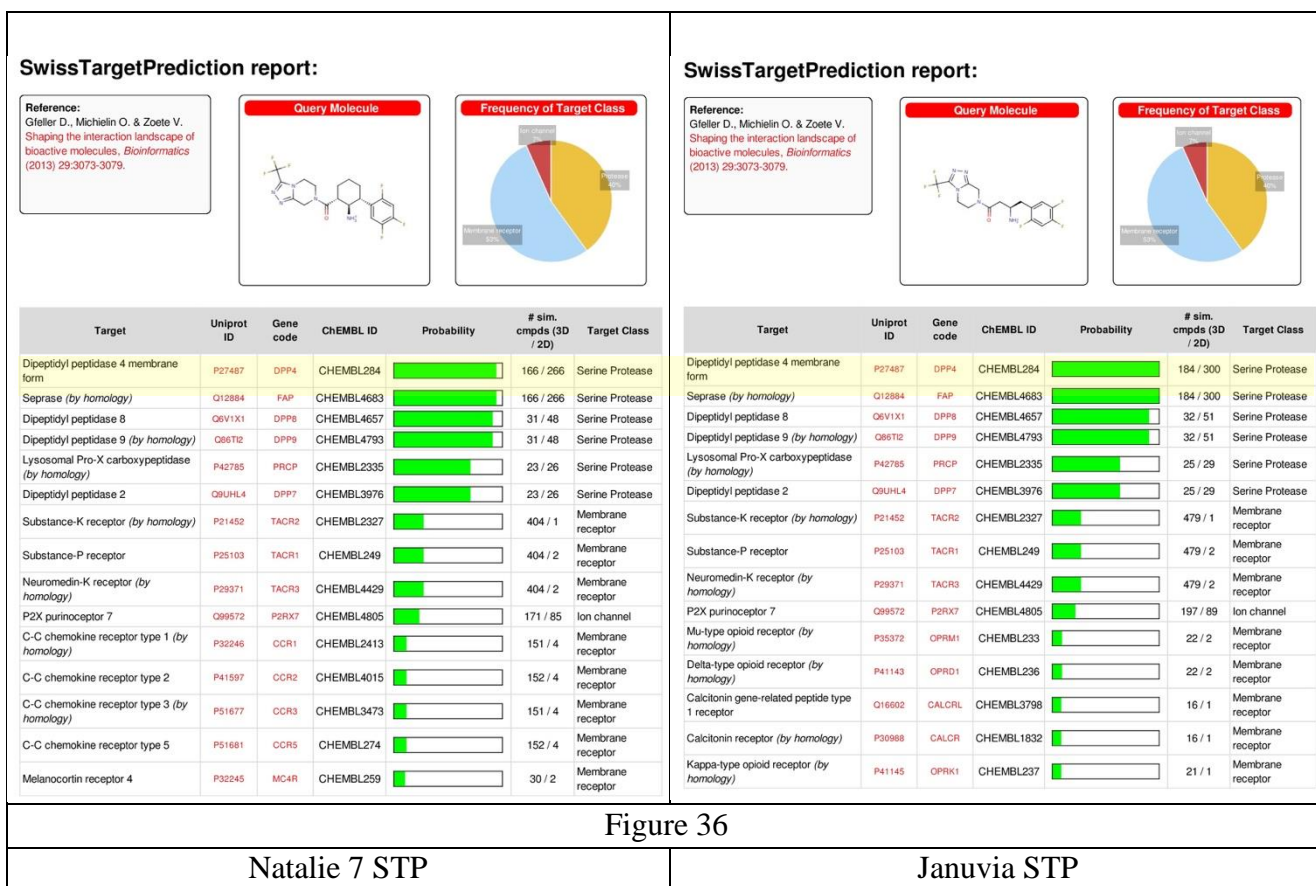
The docking results for Natalie 7 can be seen below in Figure 35, superimposed with Januvia. As was predicted, the results were very well, showing a ΔG value of -10.751 and both essential hydrogen-bonds to GLU 205 and 206 are being made.



(Figure 35)

Natalie 7 Docking Results. Natalie 7 (blue stick) superimposed with Januvia (grey stick) from PDB entry 1X70. Both essential hydrogen-bonds are being made to GLU 205 and 206, making for tight binding to DPP4.

Although these results are great, the next step would be to compare these results to that of Januvia, to see if the small modification improved or did not improve on Januvia's binding abilities. To do so, each molecule's SwissTargetPrediction report was compared, in Figure 36.

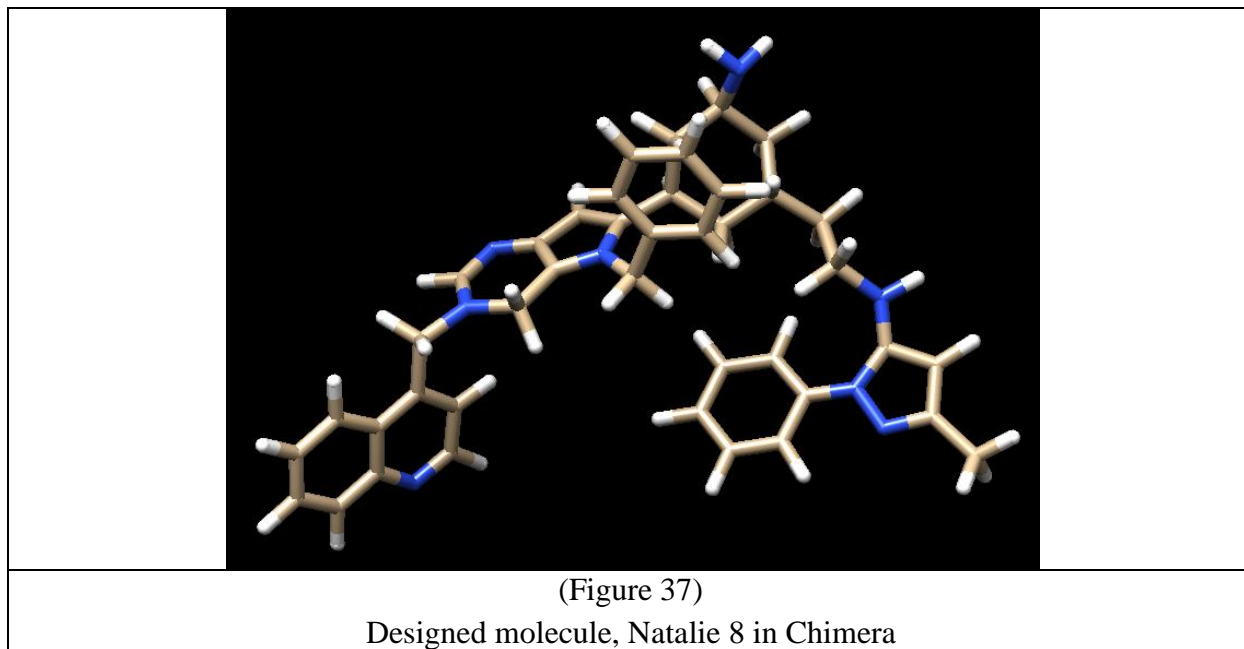


In comparing these SwissTargetPrediction reports, the small modification made to Januvia decreased the probability of interacting with DPP4. So, although this molecule produced great results, it does not improve on that of Januvia.

In comparing results between Natalie 6 and Natalie 7, it is clear that the small change made to each made for seemingly identical SwissTargetPrediction reports. This means the predicting probability of interaction with DPP4 for each was essentially the same. However, docking results show otherwise. Results for Natalie 6 showed good superimposition with Januvia, however the designed molecule did not make both essential hydrogen-bonds to GLU 205 and 206 on the protein, so binding wasn't great. For Natalie 7, though, the docking showed better results. Superimposition with Januvia showed great positioning in the active site of DPP4, and Natalie 7 does indeed make both hydrogen-bonds with GLU 205 and 206 on the protein. For this reason, it seems as though the small change made to Januvia to create Natalie 7 made for better docking and binding than the change made to create Natalie 6 did.

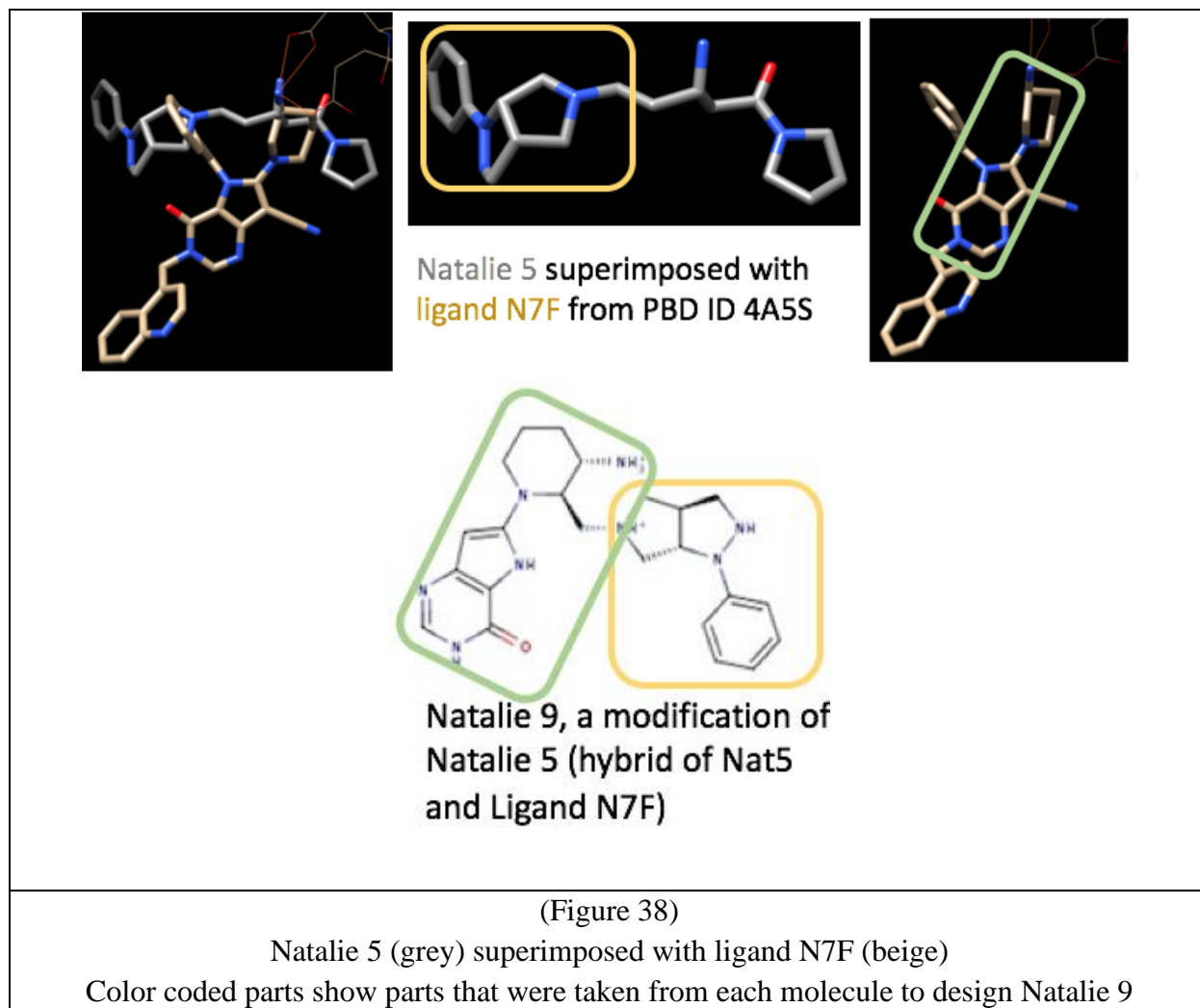
Natalie 8

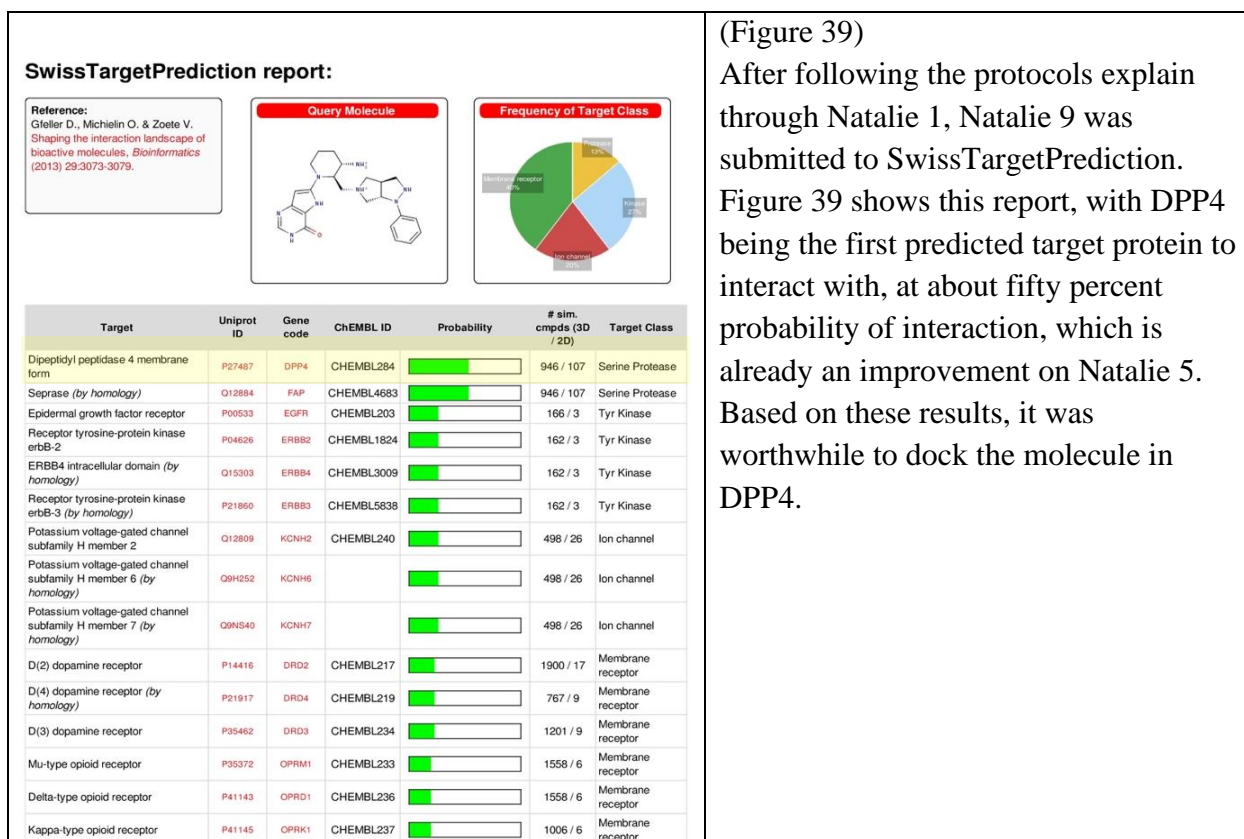
Natalie 8 was designed as an attempt to modify an earlier molecule, however, this molecule showed poor results from the beginning, not making it past the SwissTargetPrediction report. After several attempts to improve on Natalie 8, we decided to leave the molecule be for the time being. It is our hopes that in the future we can touch back on Natalie 8 and attempt to fix its design again, hopefully being able to get the molecule to the docking phase. Below, Figure 37 shows Natalie 8.



Natalie 9

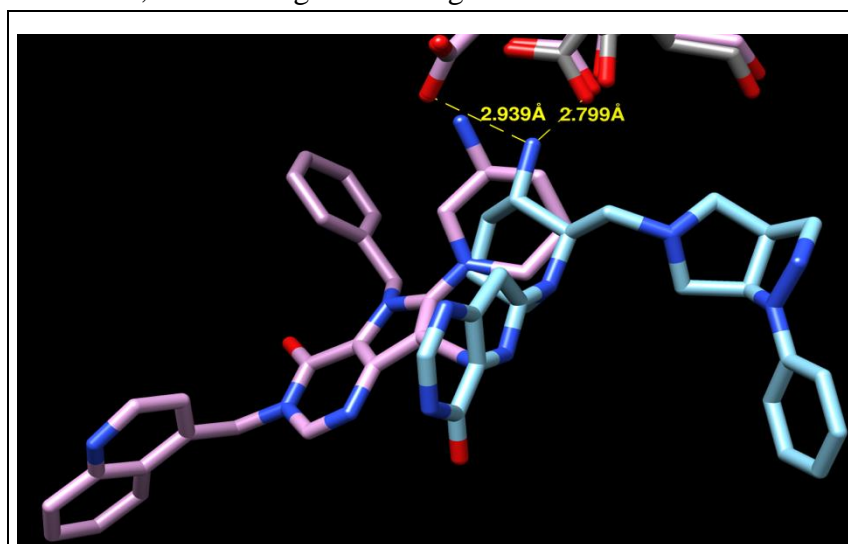
Natalie 9 was designed with the desire to improve on Natalie 5. This new design was based on the superimposition of Natalie 5 and ligand N7F from PDB ID 4A5S. The design process is shown below in Figure 38.





(Figure 39)
After following the protocols explain through Natalie 1, Natalie 9 was submitted to SwissTargetPrediction. Figure 39 shows this report, with DPP4 being the first predicted target protein to interact with, at about fifty percent probability of interaction, which is already an improvement on Natalie 5. Based on these results, it was worthwhile to dock the molecule in DPP4.

Building and preparing the molecule for docking was done in the same manner as all other molecules. Figure 40 below shows the best fit pose of Natalie 9 after analyzing docking results superimposed with ligand N7F. As seen in the photo, Natalie 9 dose make both necessary hydrogen-bonds to GLU 205 and 206 to ensure a tight bind to DPP4. This bind has a ΔG value of -10.170, which is a good binding.



(Figure 40)
Natalie 9 (blue stick) superimposed with ligand N7F from PDB entry 4A5S (pink stick)

Conclusion:

In designing a total of nine molecules, there were some successes and failures. Through the failures, it was learned that designing molecules using 2D approach is not optimal, as it does not consider the 3D structures of the molecules in the active site of the protein. By taking the Bioinformatics approach and using 3D Structure Based Drug Design to design new inhibitors of Dipeptidyl Peptidase IV, we have gotten more accurate results. Also, in following this approach we were able to cut down on the discovery process, as there were less molecules that needed to be designed and tested. Rather, Bioinformatic tools lead us to make accurate design choices, as opposed to 2D Structure Based Drug Design.

References:

- Di Costanzo, L., Ghosh, S., Zardecki, C., and Burley, S.K. 2016. Using the tools and resources of the rcsb protein data bank. *Curr. Protoc. Bioinform.*, 55:1.9.1-1.9.35. doi:10.1002/cpbi.13
- Hiramatsu, H., Yamamoto, A., Kyono, K., Higashiyama, Y., Fukushima, C., Shima, H., Sugiyama, S., Inaka, K., and Shimizu, R. 2005. The crystal structure of human dipeptidyl peptidase IV (DPPIV) complex with diprotin A. *Biological Chemistry*, 385(6): 561-564. doi:10.1515/BC.2004.068.
- International Diabetes Federation,. (2015). *Diabetes: Facts and Figures*. Retrieved from <http://www.idf.org/about-diabetes/facts-figures>
- International Diabetes Federation,. (2013). *2013 Diabetes Worldwide Statistics*.
- International Diabetes Federation,. (2014). *IDF Diabetes Atlas, Sixth Addition: 2014 Update*.
- Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. 2012. Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.*, 7(2): 146-157. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3151162/>
- PDB ID 1NU8: Thoma, R., Löffler, B., Stihle, M., Huber, W., Ruf, A., Hennig, M. (2003). Structural Basis of Proline-Specific Exopeptidase Activity as Observed in Human Dipeptidyl Peptidase-IV. *Structure*, 11(8): 947-59. doi: 10.1016/S0969-2126(03)00160-6.
- PDB ID 1N1M: Rasmussen, H.B., Branner, S., Wiberg, F.C., Wagtmann, N.R.(2002). Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analogue. *Nat.Struct.Biol.* 10: 19-25 doi: 10.1038/nsb882
- PDB ID 1X70: Kim, D., Wang, L., Beconi, M., Eiermann, G.J., Fisher, M.H., He, H., Hickey, G.J., Kowalchick, J.E., Leiting, B., Lyons, K., Marsilio, F., McCann,

M.E., Patel, R.A., Petrov, A., Scapin, G., Patel, S.B., Roy, R.S., Wu, J.K., Wyvratt, M.J., Zhang, B.B., Zhu, L., Thornberry, N.A., Weber, A.E.(2005). (2R)-4-Oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(2,4,5-trifluorophenyl)butan-2-amine: a Potent, Orally Active Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes. *J.Med.Chem.*, 48: 141-151 doi: 10.1021/jm0493156.

- PDB ID 2BUA: Nordhoff, S., Cerezo-Galvez, S., Feurer, A., Hill, O., Matassa, V.G., Metz, G., Rummey, C., Thiemann, M., Edwards, P.J.(2006). The Reversed Binding of B-Phenethylamine Inhibitors of Dpp-IV. X-Ray Structures and Properties of Novel Fragment and Elaborated Inhibitors. *Bioorg.Med.Chem.Lett.* 16: 1744 doi:10.1016/j.bmcl.2005.11.103.
- PDB ID: 3VJK: Yoshida, T., Akahoshi, F., Sakashita, H., Kitajima, H., Nakamura, M., Sonda, S., Takeuchi, M., Tanaka, Y., Ueda, N., Sekiguchi, S., Ishige, T., Shima, K., Nabeno, M., Abe, Y., Anabuki, J., Soejima, A., Yoshida, K., Takashina, Y., Ishii, S., Kiuchi, S., Fukuda, S., Tsutsumiuchi, R., Kosaka, K., Murozono, T., Nakamaru, Y., Utsumi, H., Masutomi, N., Kishida, H., Miyaguchi, I., Hayashi, Y.(2012). Discovery and preclinical profile of teneligliptin (3-[(2S,4S)-4-[4-(3-methyl-1-phenyl-1H-pyrazol-5-yl)piperazin-1-yl]pyrrolidin-2-ylcarbonyl]thiazolidine): A highly potent, selective, long-lasting and orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *Bioorg.Med.Chem.* 20: 5705-5719 doi: 10.1016/j.bmc.2012.08.012.
- PDB ID 3W2T: Nabeno, M., Akahoshi, F., Kishida, H., Miyaguchi, I., Tanaka, Y., Ishii, S., Kadowaki, T.(2013). A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors in the active site. *Biochem.Biophys.Res.Commun.* 434: 191-196 doi: 10.1016/j.bbrc.2013.03.010.
- PDB ID 4A5S: Sutton, J.M., Clark, D.E., Dunsdon, S.J., Fenton, G., Fillmore, A., Harris, N.V., Higgs, C., Hurley, C.A., Krintel, S.L., Mackenzie, R.E., Duttaroy, A., Gangl, E., Maniara, W., Sedrani, R., Namoto, K., Ostermann, N., Gerhartz, B., Sirockin, F., Trappe, J., Hassiepen, U., Baeschlin, D.K.(2012). Novel Heterocyclic Dpp-4 Inhibitors for the Treatment of Type 2 Diabetes. *Bioorg.Med.Chem.Lett.* 22: 1464 doi: 10.1016/j.bmcl.2011.11.054
- PDB ID 4PNZ: Biftu, T., Sinha-Roy, R., Chen, P., Qian, X., Feng, D., Kuethe, J.T., Scapin, G., Gao, Y.D., Yan, Y., Krueger, D., Bak, A., Eiermann, G., He, J., Cox, J., Hicks, J., Lyons, K., He, H., Salituro, G., Tong, S., Patel, S., Doss, G., Petrov, A., Wu, J., Xu, S.S., Sewall, C., Zhang, X., Zhang, B., Thornberry, N.A., Weber, A.E.(2014). *J.Med.Chem.* 57: 3205-3212 doi: 10.1021/jm401992e.
- Pissarnitski, D.A., Zhao, Z., Cole, D., Wu, W.-L., Domalski, M., Clader, J.W., Scapin, G., Voigt, J., Soriano, A., Kelly, T., Powles, M.A., Yao, Z., and Burnett, D.A. 2016. Scaffold-hopping from xanthenes to tricyclic guanines: A case study of dipeptidyl peptidase 4 (DPP4) inhibitors. *Bioorganic & Medicinal Chemistry*, 24(21): 5534-5545. doi: 10.1016/j.bmc.2016.09.007

- Statistics about diabetes. (2016, April 1). Retrieved from <http://www.diabetes.org/diabetes-basics/statistics/?loc=db-slabnav>
- SwissDock, a protein-small molecule docking web service based on EADock DSS. Fast docking using the CHARMM force field with EADock DSS.
- SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014 Jul;42 Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V

Acknowledgements:

- Research Advisor: Dr. Phalguni Ghosh
- Dr. Steven Rowley and Dr. Brian Lavey
- Dr. Sutapa Ghosh (PDB, Rutgers University, Piscataway NJ)
- Dr. Parag Muley, Chair Natural Sciences
- Dr. Linda Scherr, Dean Natural Sciences
- All of my peers
- Bristol-Meyers Squibb for financial support
- Douglas Reardon, David Rosen for their help
- Dr. Haim Baruh