A Fragment-Growing Method for the Discovery and Analysis of Novel Scaffolds for Neuraminidase Inhibitor Design

Howard Li
Pacific Rim Experiences for Undergraduates Program, 2011
University of California, San Diego
Mentors: Dr. Wilfred Li (UCSD), Dr. Jung-Hsin Lin (NTU)

1. Abstract

Neuraminidase is a key target protein for the development of antiviral drugs. This study proposes and investigates a scaffold-based, fragment-growing method to identify novel molecular scaffolds for Group 1 neuraminidase inhibition. In addition to re-identifying the scaffolds of known hits, this study identifies adamantane ($C_{10}H_{16}$) as a new, potential scaffold for development into an effective N1 inhibitor. Analysis of the derived spectrum of new chemical species with a high predicted binding affinity revealed consistent structural trends and suggested a new set of key bindable residues in the sialic acid pocket.

2. Introduction

The cause of several global pandemics within the last century, the influenza virus continues to be a major threat to human health[1, 2]. In response to the constant and urgent need for novel antiviral drugs, a number of key targets have come under the interest of the scientific community for the development of novel antiviral drugs. Among them is neuraminidase (NA), for which the drugs Zanamivir (Relenza) and Oseltamivir (Tamiflu) were developed as commercial inhibitors.[3, 4] Neuraminidase is a crucial viral surface protein in the influenza reproductive cycle. After new viral particles are synthesized in an infected host cell, the new viruses must undergo a budding process. Neuraminidase is required for catalyzing the cleaving of viral particles from host glycoproteins and releasing them so that they may attack other hosts.

Docking programs such as AutoDock 4 and AutoDock Vina attempt to predict the binding affinity between a protein target and ligand by considering physical factors such as electrostatic and hydrophobic interactions.[5, 6] These energetic calculations can be used as an important tool in computer-aided, structure-based drug design.

To exploit the ability of docking programs to predict protein-ligand interactions for the discovery of novel protein inhibitors, fragment-growing applications generate new compounds through the assembly of smaller molecular fragments.[7] Doing so allows for the exploration of chemical space via a set of fragments significantly smaller than the massive compound libraries used for virtual screening.[8] AutoGrow uses AutoDock Vina's ability to quickly model, compute, and quantify a ligand's affinity for a protein binding pocket to discover novel compounds using an evolution-based algorithm. Starting with a "parent" ligand, the program randomly modifies the molecule with substituents from a fragment library. Using AutoDock Vina as a scoring function, modifications that lead to greater binding affinity are preserved in the next generation. This cycle of random modification and selection continues until an effective inhibitor evolves out of AutoGrow's simulated process of "natural selection," leading to novel structures uniquely tailored through evolution to bind a target of interest.[7]

This study seeks to use AutoGrow to demonstrate a fragment-growing method, herein called "growth-screening," to identify potential scaffolds that are most appropriate for binding to the N1 neuraminidase active site, look for trends in fragment attachment about core scaffolds, and highlight key residues in the neuraminidase binding pocket that are most available for ligand interaction.

3. Methods

3.1. Receptor Preparation

The crystal structure for N1 neuraminidase (PDB ID: 3CKZ)[9] was first downloaded and stripped of its water molecules and a co-crystallized Zanamivir ligand using UCSF Chimera.[10] The calcium ion cofactor was left in place. Hydrogen atoms were then added through the pdb2pqr webserver, using an AMBER forcefield and output naming scheme.[11] PROPKA was used to assign protonation states at pH = 7. As a control, the structure for bacterial leucine transporter membrane protein (PDB ID: 2A65)[12] was also prepared using an identical protocol.

3.2. Scaffold Preparation & Population

Molecular scaffolds for fragment-growing evolution were derived from hits from previous screening trials. Based on the results of a previous virtual screening study using N1 neuraminidase ensemble structures and relaxed complex scheme rescoring, a set of hit compounds were identified from the NCI diversity set as potential neuraminidase inhibitors.[13] Among them were NSC 109836, NSC 211332, NSC 350191, and Zanamivir - four top hits that occupied the sialic acid binding site. These 3D structures were downloaded from the NCI database and edited using MarvinSketch, where substituent groups were removed and replaced with a singly bonded hydrogen atom while leaving the core ring structure intact. In addition, several geometric scaffolds were chosen: cyclohexane (chair), cyclohexane (boat), cyclopentane (envelope), cycloheptane, adamantane, cis-decalin, and trans-decalin. The core scaffolds of leucine and an alternative ring conformation of Zanamivir was also included in the initial core scaffold population.

3.3. AutoGrow Evolutions

For each core scaffold, three triplicate AutoGrow evolutions were run in parallel using the Opal2 Dashboard web service application.[14] Scaffolds were evolved for 8 generations using the small-fragment library of 46 fragments. Mutations were generated by converting a randomly selected hydrogen atom on both the scaffold and a random fragment into a single bond. The grid box was centered at the sialic acid binding site and generously sized to accommodate for scaffold growth - typically 14 x 14 x 14 angstroms or 23 x 23 x 23 angstroms, for smaller and larger scaffolds, respectively. Each generation consisted of 10 carryovers, 20 crossovers, and 20 mutants. The maximum atom number for each ligand evolution varied between 60 and 200, depending on the size of the initial scaffold.

3.4. Visualization & Analysis

Resulting ligands and top binding poses were visualized in MarvinSketch, PyMol and Chimera. Protein-ligand interactions were analyzed using MOE.[15, 16, 17.]

4. Results & Discussion

4.1. Core scaffold evolutions

All core scaffolds evolved within the N1 sialic acid binding pocket showed a general increase in binding affinity, with top endpoint free energy of binding between -8 and -11 kcal/mol. The evolutions of a select group of scaffolds representative of the range of resulting evolutionary trajectories are plotted in **Figure 1**.

Here, results from growth-screening using AutoGrow were able to reveal the ability of certain scaffolds to evolve into high-affinity inhibitors that was not readily apparent prior to evolution. Particularly, the free energy of binding of the adamantane scaffold according to an initial AutoDock Vina docking was -5.3 kcal/mol, lower in magnitude than the initial core

scaffold affinities of trans-decalin (-5.8 kcal/mol), cis-decalin (-5.8 kcal/mol), and NSC 109836 (-7.4 kcal/mol). After the 8-generation triplicate evolutions, however, the adamantane scaffold was revealed be the scaffold with the most potential for N1 sialic acid pocket binding. The final free energy of binding was -10.67± 0.21 kcal/mol - significantly higher than those of trans-decalin (-9.47 ± 0.12 kcal/mol), cis-decalin (-9.40 ± 0.26 kcal/mol), and NSC 109836 (-10.13 ± 0.29 kcal/mol).

Thus, these growth-screening results not only agree with previous screening studies by re-identifying the scaffolds of NSC 109836 and Zanamivir as effective scaffolds for neuraminidase inhibition, but also suggest that the adamantane scaffold may provide an even better geometrical foundation for the development of novel drugs.

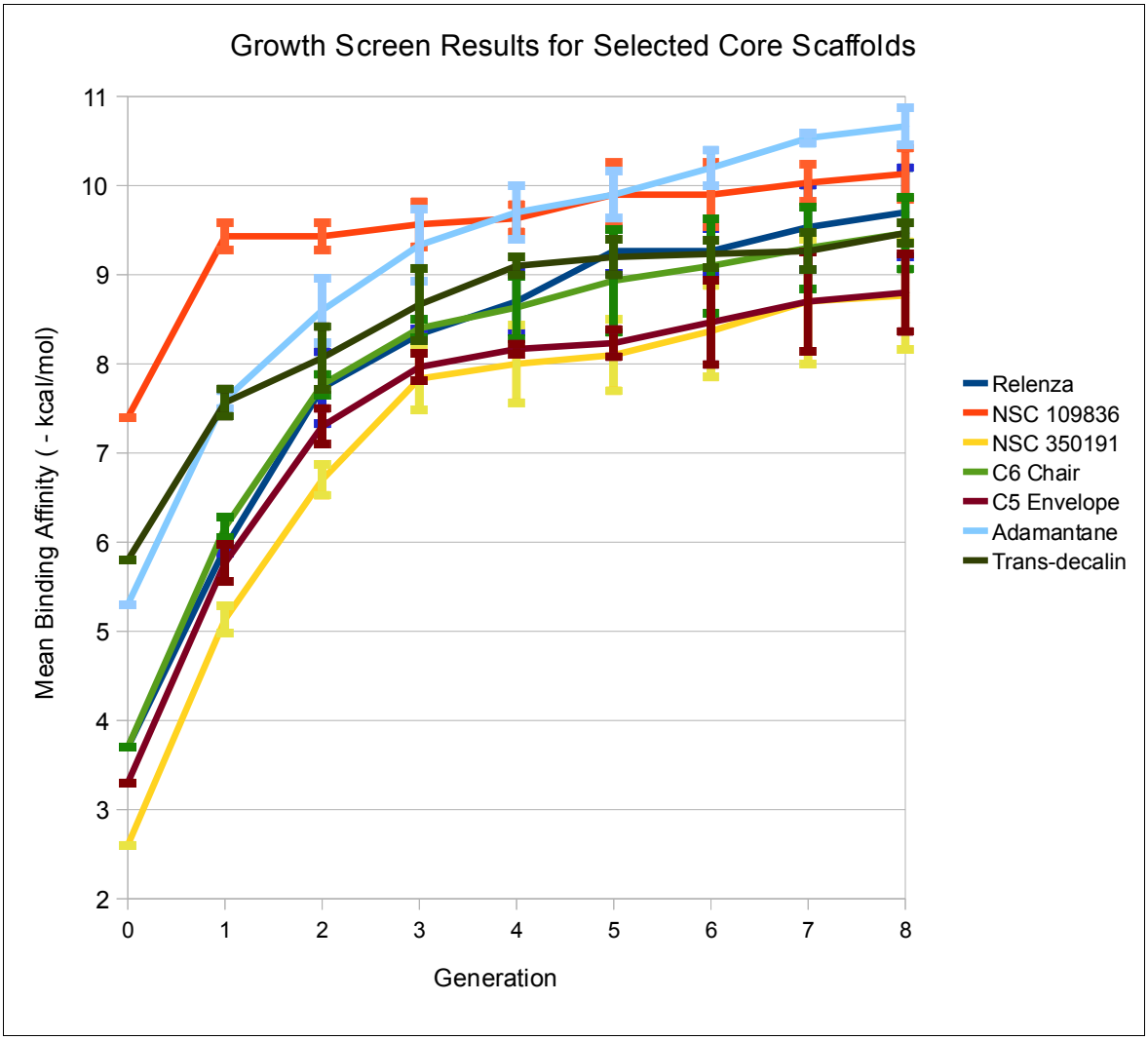


Figure 1. Evolutionary trajectories of selected core scaffold evolutions. Results of triplicate parallel evolutions were plotted for each scaffold as the mean free energy of binding for every generation ± 1 standard deviation. Results from an initial core scaffold docking using AutoDock Vina were used for Generation 0.

When a control evolution series was performed using the active site of the bacterial leucine transporter protein, evolutions were unsuccessful. Binding affinities showed little to no improvement over 8 generations. Furthermore, the free energy of binding generally remained

extremely positive (up to + 20 kcal/mol) throughout all 8 generations. This may suggest that the scaffold geometries identified through the growth screening method are uniquely suited for N1 sialic acid pocket binding. This may also reflect the relatively high ligand specificity of the leucine transporter protein compared to that of the neuraminidase active site.

The fact that *all* scaffolds used in this study (including arbitrarily selected scaffolds such as cycloheptane, cyclohexane, cyclopentane, and a distorted conformation of the Zanamivir scaffold) were able to evolve into ligands with relatively high affinity for the sialic acid binding pocket, however, calls into question the true specificity of scaffold geometries. The high evolvability of scaffolds for N1 active site binding might be due to a variety of other factors such as the shape and size of the binding pocket, the overall size of the core scaffold, and the composition of the fragment library, rather than the unique geometric structure of the scaffold itself. Furthermore, the core scaffolds show convergent evolutionary trajectories. When allowed to evolve independently over 8 parallel generations, the initially diverse set of molecular scaffolds actually became more *similar* in binding ability, rather than more different. **Figure 2** plots the standard deviation of the mean binding affinities of the entire set of scaffolds for each generation, which steadily decreases to a value of ~0.6 kcal/mol as the evolutions progress. This suggests that the final endpoint binding affinities of the evolved ligands may be largely influenced not so much by the geometry of the initial scaffold as by an intrinsic property of the receptor pocket itself that reflects the "bindability" (and perhaps, *druggability*[18]) of the sialic acid binding site.

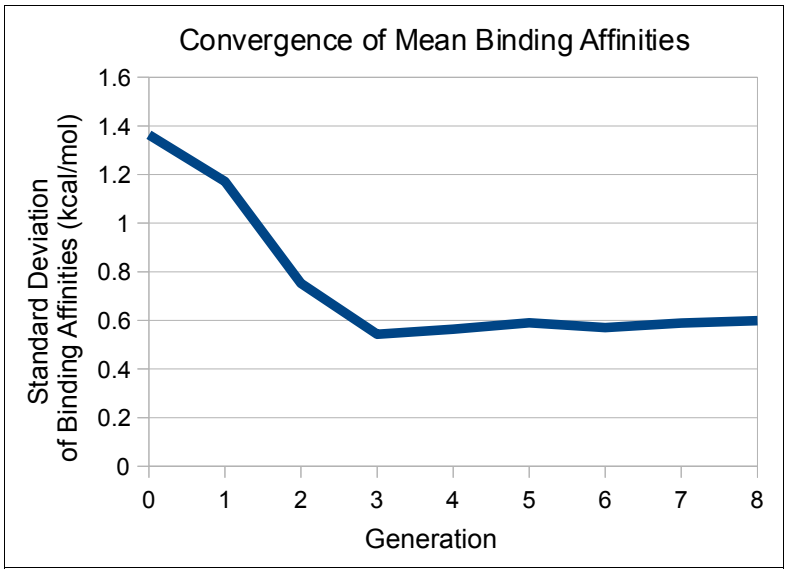


Figure 2. AutoGrow evolution causes binding affinities of ligands converge. Standard deviation of binding affinities decreases steadily to a value of ~0.6 kcal/mol as evolution progresses, suggesting the convergent nature of fragment-growth evolution. A diverse set of structurally distinct core scaffolds, when allowed to evolve and mutate on independent, parallel tracks, become more similar, rather than more different, in binding affinity.

4.2. Fragment analysis

In addition to a quantitative analysis of binding affinities, evolved ligands were also analyzed for structural trends. Comparing the top ligands from the final generations across all three parallel evolutions, it was found that even when scaffolds were allowed to evolve on independent, parallel tracks, analogous fragments were generally attached at similar sites on the core scaffold.

Adamantane, the most evolutionarily successful scaffold in this study, showed remarkably similar growth patterns across all three parallel evolutions. The top nine ligands (three from each parallel evolution) were analyzed. The fragments are listed in **Figure 3**.

H Index	Trial 1 Lig. 1	Trial1 Lig. 19	Trial 1 Lig. 3	Trial 2 Lig. 19	Trial 2 Lig. 26	Trial 2 Lig. 22	Trial 3 Lig. 1	Trial 3 Lig. 2	Trial 3 Lig. 19
1	$R\text{-}SO_4$	$R\text{-}SO_4$	$R\text{-}SO_4$					$R\text{-}SO_4$	$R\text{-}SO_4$
2		OH		$R\text{-}SO_4$	$R\text{-}SO_4$	$R\text{-}SO_4$	$R\text{-}SO_4$		
3					NH_2			$CH(OH)NH_2$	$CH(OH)NH_2$
4			$[CH(OH)]_2CH_3$						
5	OH			NHOH			$CH(CH_3)CH_2OH$		
6									
7				CH_2CH_3	$(CH_2)_2OH$	$(CH_2)_3OH$			
8	CH_2SO_3		CH_2SO_3						
9									

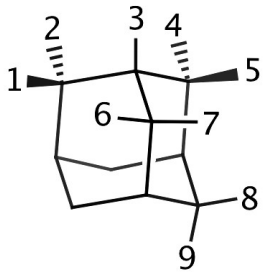


Figure 3. Fragment analysis of high N1-affinity adamantane derivatives. Relevant hydrogen indices were labelled, corresponding to possible fragment-attachment positions. The table lists attached fragments at each site for the top three ligands in each of three parallel evolutions. Sites corresponding to the same carbon atom on the adamantane scaffold are grouped together with a boldface outline. Ligands from the same evolutionary trajectory (molecular "relatives") are grouped with a red outline.

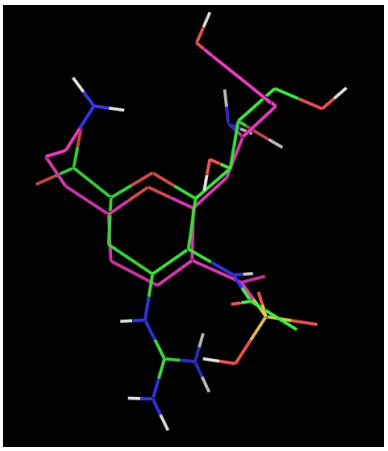
General trends include a sulfate-containing branch ($R\text{-}SO_4$) attached to the carbon center corresponding to positions 1 or 2 in all top evolved ligands, a NH_2-containing branch in position 3, and an OH-containing branch in positions 4 or 5. Throughout all three independent evolutions, only 3 carbon atoms (out of the 10 possible carbon atoms that comprise the initial scaffold) were repeatedly targeted.

To further validate this method, fragment-attachment patterns in derivatives of the Zanamivir scaffold were compared to the original structure of Zanamivir. When the binding poses of top ligands were visualized, it was noted that the distribution of fragment attachment positions did not cluster together at specific sites as was observed with adamantane. An explanation for this observation could be based on the much smaller size of the Zanamivir scaffold compared to that of adamantane, which allowed the Zanamivir scaffold to move about and reposition itself within the binding pocket throughout the evolutions to accommodate for new fragment attachments. The more sterically hindered adamantane did not have such mobility, resulting in a relatively conserved scaffold pose throughout evolutions and thus, distinct trends in fragment-attachment positions.

However, when ligands from the Zanamivir core scaffold evolutions were selected based on their core scaffold binding poses, it was found that a ligand whose core scaffold adopted a binding pose similar to the experimental binding pose of Zanamivir had nearly identical fragment attachment sites as the original Zanamivir ligand. **Figure 4** lists the fragments for Trial 2, Ligand 3. Here, the results show that the evolved ligand exactly replicated 3 out of 4 attachment sites on Zanamivir. Furthermore, at positions 6 and 7, where both R and S

positions were available for attachment, the evolved ligand reproduced the stereochemical configurations of Zanamivir exactly. This suggests that the bioactivity of Zanamivir may be strongly based in the stereochemical decoration of these chiral centers.

H Index	Zanamivir	Trial 2, Ligand 3
1	COOH	$CH_2CH_2CH_2NH_2$
2		
3	$NHCH(NH_2)_2$	
4		
5		
6	$NHC(=O)CH_3$	$CH(CH_3)SO_4$
7	$CH(OH)CH(OH)CH_2OH$	$C(NH_2)CH_2CH(CH_3)CH_2OH$
8		



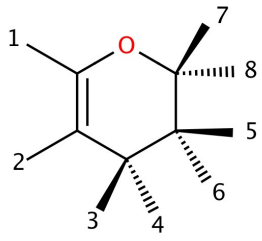


Figure 4. Fragment analysis of Zanamivir core scaffold derivatives. All 8 possible fragment-attachment positions were labeled and assigned index numbers. The evolved ligand exactly replicated 3 out of 4 attachment sites. Remarkably, for the fragments at positions 6 and 7, the random fragment-growing process correctly replicated not only the same carbon of attachment, but also the exact stereochemical positions. Also shown is the binding pose of the evolved ligand, superimposed upon the experimental binding pose of Zanamivir.

The consistencies observed in fragment-attachment patterns for both adamantane and the Zanamivir core scaffold suggests that a fragment-analysis of growth-screening results can be used to help profile the utility of molecular scaffolds for protein inhibition – how can we best utilize a particular scaffold for protein pocket binding? In what ways (what types of fragments, at which positions) must a scaffold be decorated so that it serves as an effective inhibitor? Answers to such questions may aid future efforts in developing effective protein inhibitors – particularly those employing pharmacophore-based approaches.[19]

4.3 Binding Residue Analysis

To analyze the binding modes of ligands evolved from the core scaffolds, ligand-protein interactions were calculated and visualized in MOE. Residues targeted by the top ligands for each separate evolution series were tabulated to see which residues in the sialic acid pocket were most accessible for ligand binding. **Figure 5** shows data from a binding mode analysis of ligands derived from the core scaffolds of NSC 109836, NSC 211332, Zanamivir, and cyclohexane.

From this analysis, the residues most often targeted (> 12 binding events) by AutoGrow-generation ligands were Arg 118, Ser 179, Glu 277, Arg 292, Tyr 347, Arg 371, and Tyr 406. Interestingly, Glu 276, a residue that has been for both catalysis and the activity of current inhibitors, was one of the lowest scoring residues in this binding-event analysis, with only 2 binding events. The neighboring Glu 277 residue, however, scored over six times higher, with 13 binding events. Of the highest scoring residues, Arg 118, Arg 292, and Arg 371 are currently targeted by known neuraminidase inhibitors such as Zanamivir, Oseltamivir (active form),

Laninamivir, and Peramivir.[20-21.] The others, Arg 118, Ser 179, Glu 277, Tyr 347, and Tyr 406 may be new residues that might be targeted by medicinal chemists for the development of novel antiviral drugs.

Results from this type of analysis may not only point to future directions in the design and discovery of novel neuraminidase inhibitors, but also predict which viral mutations may most likely yield drug-resistant strains. The number of accessible residues in the pocket might also help characterize how viable a protein pocket is as a potential drug-target.

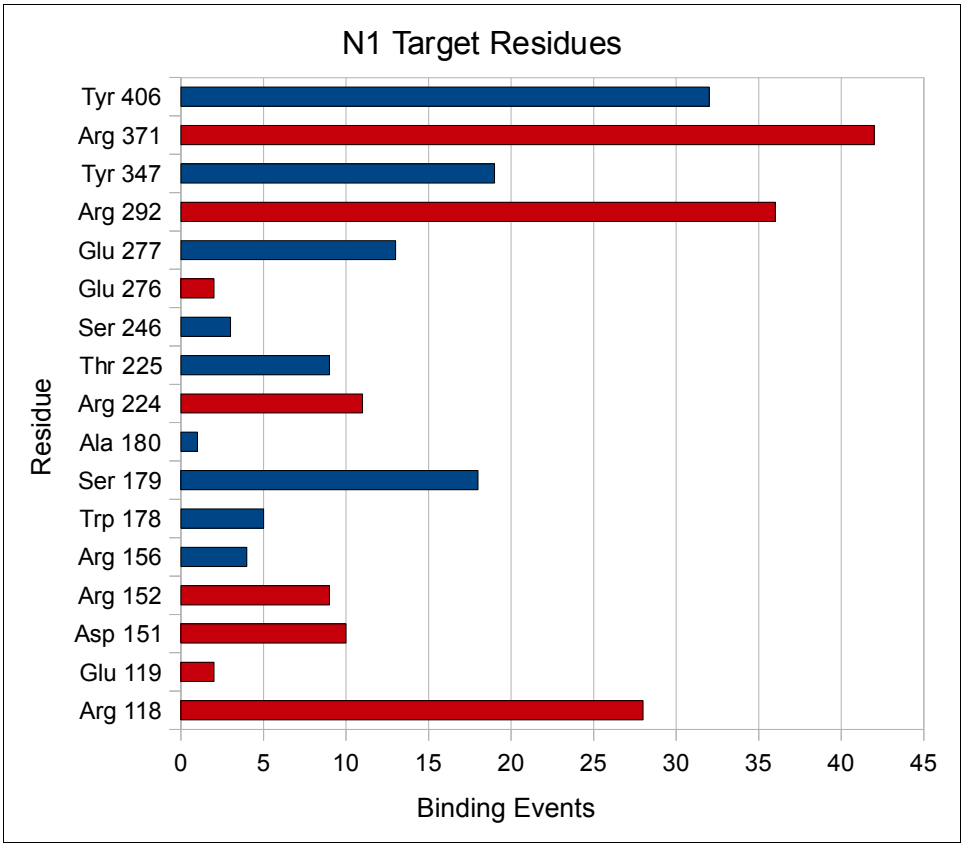


Figure 5. Key residues indicated by AutoGrow growth-screens. For each triplicate evolution for NSC 109836, NSC 211332, Zanamivir, and cyclohexane core scaffolds, the binding modes of the top 3 ligands were visualized. Every instance in which a ligand interacted with a specific residue was recorded as a "binding event." Residues targeted by current neuraminidase inhibitors (Zanamivir, Oseltamivir, Peramivir, and Laninamivir)[20-21] are highlighted in red.

5. Conclusion

The growth-screening method investigated in this study assumes that the chemical geometry of a molecular scaffold largely determines its potential to be developed into an effective inhibitor and attempts to test this potential through fragment-based evolution via fragment-growing applications such as AutoGrow. In this study, growth-screening using AutoGrow revealed the adamantane structure to be one of the highest-potential scaffolds. However, because no scaffold identified in this study showed a remarkable *inability* to evolve into a ligand with high predicted binding affinity, this study is unable to attest to ability of growth-screening to screen *out* compounds. All scaffolds investigated, according to the method used in this study, appeared to be realtively good scaffolds, with adamantane being the best. Nevertheless, the consistency of fragment-attachment trends suggests that growth-screening methods may yield valuable, reproducible directions for effective drug design.

Furthermore, growth-screening generates a wealth of information that allows for a unique analysis of both the structural characteristics of effective ligands and the bindability of the receptor pocket.

There are, however, several important limitations to this method. First, the accuracy and relevance of results depend heavily on the accuracy of the docking algorithms and receptor structures that are used, both of which are rarely without flaw. Second, the attempt to quantify a scaffold's "potential" to develop into an effective ligand rests on the assumption that the fragment library and growth algorithm allow each scaffold fair and adequate access to diverse chemical space; otherwise, important and viable structures of potential inhibitors may be systematically overlooked. Finally, as the products of random fragment growth, resulting ligands may have little to no synthetic feasibility. However, their analysis may still yield clues for inhibitor design.

While the results of this study are based solely on AutoGrow and AutoDock Vina, future work may investigate the use of other fragment-growing algorithms and scoring functions. Rescoring and redocking with other docking algorithms and scoring functions would help strengthen the validity of results. Alternative receptor, scaffold, and ligand preparation techniques should also be investigated, along with a wider range of initial molecular scaffolds and controls. Of course, only experimental validation through *in vitro* or *in vivo* studies can truly demonstrate the ability of proposed compounds to be effective; an experimental evaluation of adamantane derivatives matching the fragment-attachment trends reported in this study would be ideal.

6. Acknowledgements

This research was conducted under the invaluable mentorship and hospitality of Dr. Jung-Hsin Lin (National Taiwan University) and Dr. Wilfred Li (UCSD) as part of the Pacific Rim Experiences for Undergraduates Program (PRIME), made possible through the vision, dedication, and hard work of Dr. Gabriele Wienhausen, Dr. Peter Arzberger, Teri Simas, Jim Galvin, Tricia Taylor, Jennifer Choy, and Anafrancesca Comunale. Special thanks are due to Yu-Hsuan Chen, Jui-Chih Wang, Che-Chia Chang, and Ting-Rong Chern of the NTU Pharmacoinformatics Group as well as Jane Ren and Wendy Fong at the UCSD National Biomedical Computation Resource (NBCR) for their friendship and extrordinary patience in offering technical guidance. This project also received support from the National Science Foundation, IOSE-0710726, and from the UCSD Eureka! Summer Undergraduate Research Scholarship.

7. References

1. Sambhara S, Poland GA. (2010) H5N1 Avian Influenza: Preventive and Therapeutic Strategies Against a Pandemic. *Annual Review of Medicine.* 61: 187-98.
2. Stanekova Z, Vareckova E. (2010) Converved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virology.* 7: 351.
3. Meindl P, Bodo G, Palese P, Schuman J, Tuppy H. (1974) Inhibition of neuraminidase activity by derivatives of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid. *Virology.* 2: 457-63.
4. Lew W, Chen X, Kim CU. (2000) Discovery and development of GS 4101 (oseltamivir): an orally active influenza neuraminidase inhibitor. *Current Medicinal Chemistry.* 7(6): 663-72.
5. Morris GM et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry.* 30(16): 2785-81.

6. Trott O, Olson AK. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry* 31: 455-461
7. Durrant JD, Amaro RE, McCammon JA. (2009). AutoGrow: A Novel Algorithm for Protein Inhibitor Design. *Chemical Biology & Drug Design.* 73(2): 168-78.
8. Hajduk PJ, Greer J. (2007) A decade of fragment-based drug design: strategic advancements and lessons learned. *Nature Reviews. Drug Discovery.* 6(3): 211-9.
9. Collins PJ et al. (2008) Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature.* 453: 1258-61.
10. Pettersen EF et al (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 25: 1605-12.
11. Dolinsky TJ et al (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research.* 32: W665-7.
12. Yamashita A et al. (2005) Crystal structure of a bacterial homologue of Na+/Cl-dependent neurotransmitter transporters. *Nature*. 437: 215-23.
13. Cheng et al. (2008). Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase. *Journal of Medicinal Chemistry.* 51: 3878-94.
14. Krishnan S, Clementi L, Ren J, Papadopoulos P, Li WW. (2009) Design and Evaluation of Opal2: A Toolkit for Scientific Software as a Service. *2009 Congress on Services*. Services: 709-16.
15. Csizmadia F. (2000). Jchem: Java Applets and Modules Supporting Chemical Database Handling from Web Browsers. *J. Chem. Inf. Comput. Sci.* 40(2): 323-4.
16. DeLano: The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002
17. Vilar S, Cozza G, Moro S. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery
18. Owens J. (2007) Determining Druggability. *Nature Reviews. Drug Discovery.* 6: 187.
19. Pickett SD et al. (1996) Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ). *J. Chem. Inf. Comput. Sci.* 36(6): 1214-23.
20. Smith BJ et al. (2001) Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Science*. 10(4): 689-96.
21. Yen HL et al. (2006) Importance of Neuraminidase Active-Site Residues to the Neuraminidase Inhibitor Resistance of Influenza Viruses. *Journal of Virology*. 80(17): 8787-95.