

國立交通大學

生物資訊及系統生物研究所

博士論文

同藥理同源途徑蛋白質對多標靶抑制劑之研發

Pharmapathlogs for Multitarget Inhibitors



研究生：許凱程

指導教授：楊進木 教授

中華民國一百年八月

同藥理同源途徑蛋白質對多標靶抑制劑之研發

Pharmapathlogs for Multitarget Inhibitors

研究生：許凱程

Student : Kai-Cheng Hsu

指導教授：楊進木

Advisor : Jinn-Moon Yang



A Thesis Submitted to Institute of Bioinformatics and Systems Biology
National Chiao Tung University in partial Fulfillment of the Requirements
for the Degree of Ph.D. in
Bioinformatics and Systems Biology

August 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年八月

同藥理同源途徑蛋白質對多標靶抑制劑之研發

研究生：許凱程

指導教授：楊進木博士

國立交通大學 生物資訊及系統生物研究所 博士班

摘 要

「單一藥物針對單一標靶，治療單一疾病」是過去數十年藥物發展的主要概念，這個概念促使研究者發展具有高專一性的藥物，然而這個概念逐漸變得不適用於治療疾病上，主要原因之一是單一標靶藥物容易受到蛋白質結合位上的突變之影響，而導致抗藥性的產生。因此，開發一個能尋找多標靶藥物之新策略，藉由抑制多個標靶來降低抗藥性之產生並且增加療效，將對藥物開發提供重要的價值。

針對此議題，我們提出一個新的概念「同藥理同源途徑蛋白質(pharmapathlog)」來尋找多標靶藥物。同藥理同源途徑蛋白質是一群具有下述特性的蛋白質：這些蛋白質(1)是同反應途徑上的同源蛋白質；(2)具有相似的核心結合環境；(3)能夠被相同的化合物所抑制。在生物體反應途徑上，因為一個蛋白質的產物是其下游蛋白質的受質，使得這些蛋白質在結合位上具有相似的物理化學特性及構型。此外，同源蛋白質常在結合位上具有高度保留性的區域，針對這些區域設計的藥物能降低產生抗藥性之機率。根據此概念，我們發展新的策略「以同藥理同源途徑蛋白質為基礎之藥物篩選(pharmapathlog-based screening strategy)」用以尋找多物種間之同藥理同源途徑蛋白質及其核心結合環境，當化合物能同時符合此核心結合環境時將有潛力成為這些蛋白質之多標靶藥物，可增加療效並且降低產生抗藥性的機率。

我們應用此策略於尋找細菌及病毒之抑制劑，包括結核桿菌、幽門螺旋桿菌和流行性感冒病毒。在抗菌研究上，成功找出三個多標靶抑制劑 ($IC_{50} < 10.0 \mu M$) 能同時抑制幽門螺旋桿菌中的莽草酸激酶和莽草酸去氫酶，此三個抑制劑也能抑制結核桿菌中的莽草酸激酶 ($IC_{50} < 10.0 \mu M$)。此外，我們發現三個能抑制流行性感冒病毒 H1N1 與 H5N1 神經胺酸酶之新型抑制劑 ($IC_{50} 4\sim 20 \mu M$)，實驗結果顯示此三個抑制劑能有效抑制抗藥性之神經胺酸酶 (H274Y 和 I222R)，並且不引起細胞毒性，提供對抗抗藥性病株的一個良好起始點。我們也設計了五個瑞樂沙衍生物，這些衍生物位於 150 cavity 中且有良好的抑制效果 ($IC_{50} < 10.0 nM$)。另外，統計結果顯示同藥理同源途徑蛋白質中的核心結合環境為高度保留性之區域，可以用來設計不易產生抗藥性之藥物。這些實驗結果顯示同藥理同源途徑蛋白質之藥物篩選的概念將有助於尋找多標靶抑制劑，我們相信此策略對於藥物研究開發具有極大價值。

Pharmapathlogs for Multitarget Inhibitors

Student : Kai-Cheng Hsu

Adviser : Dr. Jinn-Moon Yang

Institute of Bioinformatics and Systems Biology

National Chiao Tung University

ABSTRACT

The concept of "one-drug, one-target, one-disease" has been dominant to drug development strategy in the past decades. This strategy induces researchers to develop inhibitors with high specificity. However, the strategy is increasingly becoming inappropriate. One of the major reasons is that single-target inhibitors often lose potency because of even one residue mutation, leading to drug resistance. Therefore, developing a new strategy to discover multitarget inhibitors, which decrease probability of drug resistances and enhance therapeutic potency by inhibiting multiple targets, provides a great value for drug design.

To address the issue, we proposed a new concept "pharmapathlog" to discover multitarget inhibitors. Pharmapathlog are a group of proteins that satisfy the following properties: (1) they are protein **orthologs** in the same **pathway**; (2) they share comparable core binding environments; (3) they can be inhibited by the same compounds. Proteins in the same pathway may share similarities in physical-chemical properties and shapes in their binding sites because a product of one enzyme is a substrate of the next enzyme in a series of catalytic reactions. Furthermore, orthologous proteins often share conserved core binding environments during evolution, providing an opportunity to develop multitarget inhibitors to target these conserved regions for reducing the probability of drug resistance. Based on the new concept, we developed a "pharmapathlog-based screening strategy" to identify pharmapathlogs in the same pathways across multiple species and their core binding environments by using site-moiety maps. A compound highly agreeing with the core binding environments of pharmapathlogs could simultaneously inhibit the multiple proteins of pharmapathlogs.

To verify the utility of the pharmapathlog-based screening strategy, we applied this strategy to identify new inhibitors for bacteria and virus, including *Helicobacter pylori*, *Mycobacterium tuberculosis*, and influenza virus. Based on the strategy, three multitarget inhibitors simultaneously inhibiting shikimate dehydrogenase and shikimate kinase of *Helicobacter pylori* with low IC_{50} values ($<10.0 \mu\text{M}$) were discovered. The three inhibitors also showed inhibitory effects ($IC_{50} <10 \mu\text{M}$) for shikimate kinase of *Mycobacterium tuberculosis*. Subsequently, the strategy was successfully used to discover three new inhibitors with low IC_{50} values (4~20 μM) for H1N1 and H5N1 neuraminidases, and design five zanamivir derivatives located at the 150-cavity with IC_{50} values in the <10 nanomolar range. Our experimental results showed that the three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 neuraminidase without causing apparent

cytotoxicity, suggesting a starting point to combat drug-resistant strains. In addition, we found that core binding environments of pharmapathlogs are highly conserved, suggesting that targeting the core binding environments is useful to avoid drug-resistance. These experimental results show that the concept of pharmapathlogs is useful to discover multitarget inhibitors. We believe that the new strategy is useful to design new drugs toward human diseases.



Acknowledgement

經由很多人的努力及幫助使我能夠完成這份論文。首先我必須感謝我的指導教授楊進木老師，他對研究的熱情感染了我，使我能在學習的路程積極向上，當我遇到瓶頸或對問題絕望時他總是會說：『來討論一下。』，討論完之後總是會豁然開朗。除了在研究專業上的教導與訓練之外，楊老師沒有什麼事情是解決不了的態度更令人值得學習，這也激勵我能走得更遠且不輕言放棄。還要感謝他所提供的學習環境、研究資源以及出國的機會，使得我在求學時能夠更加順利，也要感謝他經常提供對外合作的機會，增廣我的見聞。

接著我要感謝我的口試委員，包含所內我的指導教授楊進木教授、黃鎮剛教授，所外的王雯靜教授、徐祖安教授、郭盛助教授、吳永昌教授、鄭添祿教授。感謝每位教授在百忙之中抽空擔當我的口試委員並且評鑑我的論文，以及在口試期間的對我研究所提供的寶貴建議，有了他們的指教與建言才使得這本論文能更臻完美。

我也感謝實驗室的全體同仁與我走過這一段路。特別感謝與我同組的彥甫、伸融、御哲、志達、凌婷、彥超、宣人，常常與我一起討論與工作到半夜，在新竹與你們吃飯的時候總是非常開心的。感謝章維在研究與生活上給我的幫助，也感謝俊辰、其樺、一原、宇書、峻宇、怡馨在研究上給的建議。感謝士中平日在生活上及行政上的幫助，讓我能輕鬆很多。因為實驗室的同仁使得我的求學過程更豐富更多采多姿。

再來我要感謝合作的夥伴給予的實驗資料以及研究建議。感謝王雯靜教授與鄭文琪博士在研究幽門螺旋桿菌之抑制劑上所提供的實驗資料，常常看到鄭文琪博士來生科實驗館做一整天的實驗。感謝徐祖安教授、洪慧貞博士及方明瑜小姐在研究流行性感冒病毒之抑制劑上所提供的協助，每次都必須麻煩他們花很多時間做實驗。感謝林俊成教授與林建宏先生在合成抗流感病毒之抑制劑所提供的資料。因為這些人的努力使得我能夠驗證我電腦預測及分析的結果，使得論文能更加完整。

最後我想特別感謝我父母與睿瑜。我的父母從小就提供了良好的環境讓我無後顧之憂地學習，總是無怨無悔地照顧我，即使在成績不理想時也不曾責罵過我，溫暖的家是我堅持下去的力量。睿瑜總是與我分享生活中的每一件事，我們一起度過快樂與悲傷，甚至在研究上也能互相幫忙，在每個星期的努力後，與她在一起能讓我充完電，充滿精力面對挑戰，希望我們未來能一直這樣走下去。僅將此論文獻給我這些敬愛的人以及幫助我的人。

Contents

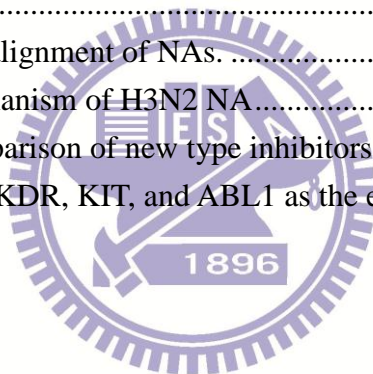
摘要.....	i
ABSTRACT.....	ii
Acknowledgement.....	iv
Contents.....	v
List of figures.....	vii
List of tables.....	ix
Chapter 1. Introduction.....	1
1.1 Background.....	1
1.2 Current state of computational drug design.....	5
1.3 Thesis overview.....	7
Chapter 2. Site-moiety map to recognize interaction preferences between protein pockets and compound moieties.....	10
2.1 Introduction.....	11
2.2 Methods and Materials.....	12
2.2.1 Site-moiety map, anchor and pocket.....	15
2.2.2 Data sets.....	15
2.2.3 Main procedure of constructing a site-moiety map.....	16
2.2.4 Input and output of the SiMMap server.....	18
2.3 Results and Discussion.....	20
2.3.1 Thymidine kinase.....	20
2.3.2 Estrogen receptor.....	22
2.3.3 Discussion.....	22
2.4 Conclusions.....	30
Chapter 3. Pharmapathlogs for discovering multitarget inhibitors of shikimate pathways using site-moiety maps.....	31
3.1 Introduction.....	32
3.2 Methods and Materials.....	33
3.2.1 Overview of pharmapathlog-based screening strategy.....	33
3.2.2 Preparations of protein structures and screening databases.....	37
3.2.3 Computational screening and establishment of site-moiety maps.....	39
3.2.4 Identification of core anchors, pharmapathlogs, and multitarget inhibitors.....	42
3.3 Results and Discussion.....	44
3.3.1 Site-moiety map of shikimate dehydrogenase.....	44
3.3.2 Site-moiety maps of shikimate kinases.....	48
3.3.3 Core anchors of pharmapathlogs.....	51

3.3.4	New multitarget inhibitors	56
3.3.5	Specific site and inhibitors for shikimate dehydrogenase.....	58
3.4	Conclusions	59
3.5	Acknowledgments	60
Chapter 4.	Pharmapathlogs for optimizing and identifying neuraminidase inhibitors	61
4.1	Introduction	62
4.2	Methods and Materials	64
4.2.1	Overview of pharmapathlog-based screening strategy	64
4.2.2	Dataset preparation.....	66
4.2.3	Main procedure for identifying new type inhibitors and lead optimization 69	
4.3	Results and Discussion.....	73
4.3.1	Site-moiety maps of H1N1, H5N1, and H3N2 NAs	73
4.3.2	Optimization process of zanamivir derivatives	79
4.3.3	Identified novel inhibitors	86
4.3.4	Proposed binding mechanism of novel inhibitors	88
4.3.5	Advantages of pharmapathlog-based screening strategy	93
4.4	Conclusions	94
4.5	Acknowledgments	95
Chapter 5.	Conclusions	96
5.1	Summary	96
5.2	Major contributions	97
5.3	Future works.....	98
List of publications	101
References	102

List of figures

Figure 1.1. Concept of pharmapathlogs using protein orthologs in the shikimate pathways as the example.....	3
Figure 2.1. Overview of the SiMMap server for the site-moiety map using herpes simplex virus type-1 thymidine kinase and 1000 docked compounds as the query..	14
Figure 2.2. The SiMMap server analysis results using estrogen receptor and 1000 docked compounds as the query.	19
Figure 2.3. The relationships between the site-moiety map and 15 co-crystallized ligands of TK	24
Figure 2.4. The relationships between the site-moiety map and 22 co-crystallized ligands of ER.	25
Figure 2.5. Comparison of SiMMap with GEMDOCK.	26
Figure 3.1. Overview of pharmapathlog-based screening strategy for identifying multitarget inhibitors.....	35
Figure 3.2. Discovery of multitarget inhibitors for inhibiting multiple proteins in a pathway. .	36
Figure 3.3. Establishment of a site-moiety map for a target binding site using shikimate dehydrogenase of <i>Helicobacter pylori</i> as the example.	41
Figure 3.4. Identification of core anchors through alignment of site-moiety maps	44
Figure 3.5. Site-moiety map of shikimate dehydrogenase.....	47
Figure 3.6. Importance of anchors in biological functions	48
Figure 3.7. Site-moiety maps of shikimate kinases of <i>Helicobacter pylori</i> and <i>Mycobacterium tuberculosis</i>	50
Figure 3.8. Importance of anchors in biological functions using shikimate kinase of <i>Helicobacter pylori</i> as the example.....	51
Figure 3.9. Core anchors of HpSDH, HpSK, and MtSK.	53
Figure 3.10. The division process for deciding core anchor residues, anchor residues, binding site residues, and other residues.	55
Figure 3.11. Conserved score distribution of core anchor residues, anchor residues, binding site residues, and other residues.....	55
Figure 3.12. New multitarget inhibitors identified by the pharmapathlog-based screening strategy	57
Figure 3.13. Specific site and inhibitors for shikimate dehydrogenase.	59
Figure 4.1. Overview of the pharmapathlog-based screening strategy for optimizing lead compounds and identifying novel inhibitors	65
Figure 4.2. Main steps for constructing a site-moiety map.....	66
Figure 4.3. Sequence comparison between H1N1 and H3N2 NA 150-loops	68

Figure 4.4. Modeled structure of H3N2 NA with a 150-open form.....	69
Figure 4.5. Site-moiety maps of NA pharmacophores.....	73
Figure 4.6. Relationship between core anchors and moieties of known inhibitors.....	76
Figure 4.7. Conserved hydrogen-bonding interactions between water atoms and pockets of the H2 and H3 core anchors	78
Figure 4.8. Core anchors of NA pharmacophores for lead optimization processes of zanamivir, oseltamivir, and peramivir.....	80
Figure 4.9. Zanamivir, oseltamivir, and peramivir analogues for verifying core anchors of pharmacophores on lead optimization processes	81
Figure 4.10. Pearson's correlation coefficient between moiety energies and IC ₅₀ values of compounds.....	82
Figure 4.11. Lead optimization process of zanamivir derivatives for extending to the 150-cavity.	84
Figure 4.12. Discovered new type inhibitors	86
Figure 4.13. Effects of influenza virus in cytopathic effect inhibition assays of novel inhibitors.	88
Figure 4.14. Multiple sequence alignment of NAs.	89
Figure 4.15. Proposed new mechanism of H3N2 NA.....	91
Figure 4.16. Binding mode comparison of new type inhibitors.....	92
Figure 5.1. Pharmacophores using KDR, KIT, and ABL1 as the example.....	100



List of tables

Table 2.1. The relationship between the anchors and moieties of 15 co-crystallized ligands for TK.....	21
Table 2.2. Comparing SiMMap with other methods on thymidine kinase and estrogen receptor by false-positive rates	27
Table 2.3. The mapping between the anchors and active and typical compounds for TK.....	28
Table 2.4. The mapping between the anchors and active and typical compounds for ER.....	29
Table 3.1. Properties of compound databases.	38



Chapter 1. Introduction

1.1 Background

The concept of "one-disease, one-target, one-drug" has been the dominating drug development strategy in the past decades^{1,2}. This strategy induces researchers to develop inhibitors with high specificity. For example, in anti-influenza drug development, neuraminidase is considered a valid target, and two drugs, zanamivir and oseltamivir, have been reported³⁻⁵. However, these single-target inhibitors may easily lose their effectiveness due to even one amino acid mutation in binding sites of target proteins, leading to drug resistance. For instance, some influenza strains that are resistant to oseltamivir have been reported since a single residue mutates⁶. Another example is tetracycline, which is a broad spectrum antibiotic. Tetracycline loses its potency in *Helicobacter pylori* (*H. pylori*) because of a single triple-base-pair substitution⁷. Particularly in antibiotics, the increasing emergence of multiple-antibiotic-resistant superbugs causes a great concern in the world⁸⁻¹⁰, revealing the insufficiency of the single-target strategy. Therefore, developing a new strategy to discover multitarget inhibitors, which decrease probability of drug resistances by inhibiting multiple targets, provides a great value for drug design.

Proteins may share many similarities in physical-chemical properties and shapes in their binding sites despite low sequence or structural homology. For example, proteins in the same pathways contain comparable core binding environments because a product of one enzyme is a substrate of the next enzyme in a series of catalytic reactions. Hence, it is possible to design a multitarget inhibitor to simultaneously inhibit proteins in the same pathways by targeting their core binding environments. Furthermore, orthologous proteins often share conserved core binding environments during evolution, providing an opportunity to develop inhibitors to target

these conserved regions for reducing the probability of drug resistance and increasing hit rate. Recently, the concept of polypharmacology, which means that a drug binds multiple target proteins, has been proposed to design drugs¹¹⁻¹³. In general, proteins with high sequence or structure similarity could be considered to be bound by the same compounds. However, designing these multitarget inhibitors is still a challenging task since these proteins often lack structural and sequence homology^{14,15}, resulting in a difficulty for extracting core binding environments among these proteins. Therefore, a new strategy for extracting core binding environments without relying on sequences of structures will be useful for discovering multitarget inhibitors.

To address these issues, we propose a new strategy, called pharmapathlog-based screening strategy, to discover multitarget inhibitors (Fig. 1.1). Pharmapathlogs are a group of proteins that satisfy the following properties: (1) they are protein orthologs in the same pathway; (2) they share comparable core binding environments; (3) they can be inhibited by the same compounds. To extract core binding environments of protein binding sites, we developed the SiMMap server for generation of site-moiety maps¹⁶. A site-moiety map consists of anchors for a protein binding site. An anchor, presenting a key binding environment, includes three essential elements: (1) binding pockets, which are parts of the binding site, with conserved interacting residues; (2) moiety preferences; (3) interaction type (electrostatic, hydrogen-bonding, or van der Waals). A site-moiety map is able to present the relationship between the moiety preferences and the physico-chemical properties of the binding site through anchors. Hence, protein orthologs sharing comparable core anchors (core binding environments) in the same pathway could be considered pharmapathlogs, and the core anchors of pharmapathlogs can be used to identify multitarget inhibitors. A compound that agrees with the core anchors is often able to simultaneously inhibit the multiple targets. In addition, the moiety preferences of the core anchors can guide lead optimization processes.

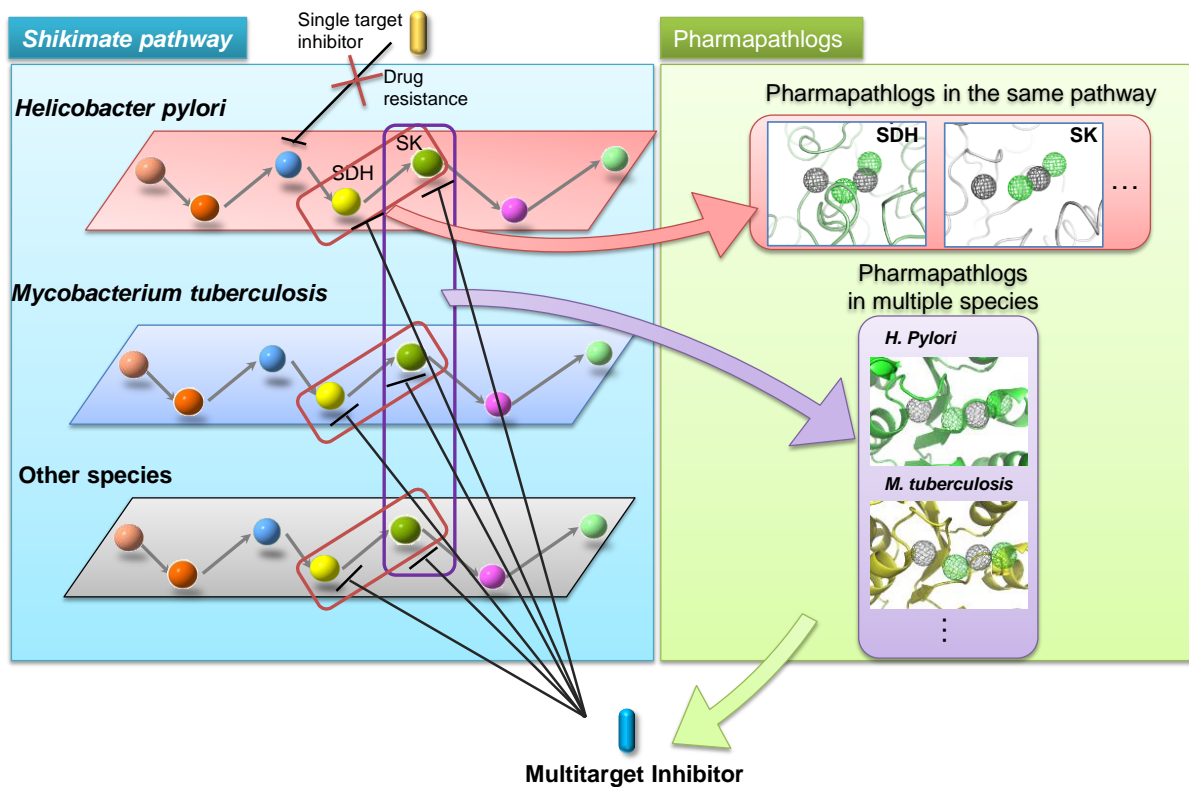


Figure 1.1. Concept of pharmapathlogs using protein orthologs in the shikimate pathways as the example. Orthologous proteins sharing similar binding environments in the same pathways can be considered pharmapathlogs. For example, shikimate dehydrogenase (SDH) and shikimate kinase (SK) are adjacent proteins in the shikimate pathway. The two proteins share two key hydrogen binding environments (green sphere) and two van der Waals binding environments (grey sphere), and can be regarded as pharmapathlogs in the pathway. In addition, orthologous SKs in *Helicobacter pylori* (*H. pylori*) and *Mycobacterium tuberculosis* (*M. tuberculosis*) are regarded as pharmapathlogs in multiple species because of their comparable binding environments. The consensus binding environments among pharmapathlogs are core binding environments and can be used to find multitarget inhibitors agreeing with the core regions. Multitarget inhibitors have good therapeutic effectiveness and reduce probability of resistant mutations, whereas single target inhibitors often lose effectiveness when target residues mutate.

To verify the utility of the pharmapathlog-based screening strategy, at first, we applied this strategy to identify new multitarget inhibitors for shikimate pathway of *H. pylori* and *Mycobacterium tuberculosis* (*M. tuberculosis*), which are human pathogens and causes peptic ulcer disease and chronic infectious disease, respectively¹⁷⁻¹⁹. The shikimate pathway containing seven proteins is an attractive target pathway for drug development because the pathway is absent in human²⁰. By use of this strategy, we successfully discovered three multitarget inhibitors with low IC₅₀ values (<10.0 μM) for simultaneously inhibiting shikimate dehydrogenase and shikimate kinase by collaborating with Dr. Wen-Ching Wang and Dr. Wen-Chi Cheng of National Tsing Hua University (NTHU). Subsequently, we applied the strategy to discover three new inhibitors with low IC₅₀ values (4~20 μM) for H1N1 and H5N1 neuraminidases, and design five zanamivir derivatives with IC₅₀ values in the <10 nanomolar range by collaborating with Dr. John T.A. Hsu and Dr. Hui-Chen Hung of National Health Research Institutes, and Dr. Chun-Cheng Lin and Mr. Chien-Hung Lin of NTHU. Our experimental results showed that the three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 neuraminidase without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains. The experimental results showed that the concept of pharmapathlogs is useful to discover multitarget inhibitors. We believe that the new strategy is useful to discover and optimize new lines of inhibitors toward human diseases.

The pharmapathlog-based screening is a general strategy for drug development, and can be extend to other human diseases and drug-resistant pathogens. A study showed that developing a drug costs 15 years and US\$800 million on average²¹. The high cost and lengthy development time reveals the insufficiency of the traditional strategy in developing drugs for combating rapidly emerging diseases, such as malaria, tuberculosis, cholera, and avian flu. Once drug-resistant pathogens emerge, current drug treatments may be ineffective. As a result,

the pharmapathlog-based screening, which is different to currently used single-target approaches, has great potential because of the following advantages: 1) High success rate. The new strategy simultaneously considers multiple target proteins for discovering inhibitors, providing an additive opportunity to discover true hits against diseases. 2) Reduction of drug resistance. The probability of drug resistant mutations arising in all targets is extremely low. 3) High treatment efficiency. Multitarget inhibitors inhibit multiple targets; therefore, these drugs increase the efficiency of therapy and are useful to treat complexity of diseases. 4) Reduction of cost and time. Based on above reasons, we believe our research results are helpful for the drug development process.

1.2 Current state of computational drug design

Virtual screening is an efficient and promising strategy in drug discovery^{2,22,23}. In virtual screening, thousands of compounds are ranked according to their binding affinities predicted by scoring functions, and top-ranked compounds are then tested by experiments. There are three general classes of scoring functions, including force-field-based methods^{24,25}, empirical methods^{26,27}, and knowledge-based methods^{28,29}. Force-field-based scoring functions are derived from molecular mechanics force-fields such as van der Waals potentials and Columbic interactions. Empirical scoring functions measure binding affinities by summing up terms that describe physical contributions, such as hydrogen bonding, van der Waals forces, and hydrophobic contacts. Generally, empirical scoring functions have simplified energy descriptors based on physical properties and then reduce computational cost in virtual screening. Knowledge-based scoring functions are derived from energy-like functions by considering the distributions of interatomic distances in a set of crystal structures of protein–ligand complexes. Furthermore, regression techniques are often applied to scoring functions, and coefficients of descriptors are derived from a set of protein-ligand complexes with

experimental binding affinities. As a result, these scoring functions (*e.g.*, X-SCORE³⁰, ChemScore²⁷, SCORE³⁰, DrugScore²⁹, and PLD³¹) usually perform well in the prediction of binding affinity.

As the number of protein structures increases rapidly, virtual screening approaches is becoming important and helpful in lead discovery^{2,22,23,32}. Based on the various scoring functions, many popular programs (*e.g.*, GEMDOCK³³, DOCK³⁴, AutoDock³⁵, and GOLD²⁵) were designed for virtual screening, and were successfully applied to identify lead compounds for target proteins. However, hit rates of these programs remained intensive because of the incomplete understandings of ligand binding mechanisms in protein-ligand interactions^{2,22,23}. For example, most of the scoring functions often lack consideration for key binding environments, such as pharmacophore spots, metal ions, and conserved residues. Pharmacophore spots are the spatial arrangement of compound moieties that are responsible for biological activity³⁶. Metal-ligand interactions stabilize ligands to facilitate catalysis³⁷. Conserved residues interacting with bound compounds often play important roles for biological functions. For instance, catalytic residues, which polarize substrates and thereby stabilize transition states³⁸, are evolutionarily conserved. These key binding environments are essential for ligand binding and biological functions but ignored in most computational methods. In addition, most of these docking programs^{25,34,39} use energy-based scoring methods, which are often biased toward both the selection of high molecular weight compounds and charged polar compounds^{40,41}.

Recently, some approaches have been proposed to derive key binding environments of protein-ligand interfaces (*e.g.*, pharmacophore spots) from known compounds^{40,42,43}. These approaches apparently increase the chance to identify active compounds. However, they are often unable to be applied for new targets, which have no known active compounds. Currently, some approaches used the scoring functions to discover multitarget inhibitors by virtual

screening^{13,44}. Wei *et al.* discovered multitarget inhibitors for the human leukotriene A4 hydrolase and the human nonpancreatic secretory phospholipase A2 by a pharmacophore-based method^{45,46}. These approaches also contained the disadvantages as we mentioned in the previous part, leading low prediction accuracy or limiting discovery of new target inhibitors. Therefore, the more powerful techniques for identifying these key binding environments and multitarget inhibitors provide a great potential value for drug design.

In this thesis, we presented the SiMMap server to infer the key binding environments by a site-moiety map in protein-ligand interfaces. The server provides pocket-moiety interaction preferences (anchors) including binding pockets with conserved interacting residues, moiety preferences, and interaction type. We verified the site-moiety map on three therapeutic targets, including thymidine kinase, and estrogen receptors of antagonists and agonists. Experimental results showed that an anchor is often a hot spot and the site-moiety map is useful to identify active compounds for these targets. In addition, we applied site-moiety maps to extract core binding environments (core anchors) of pharmapathlogs for shikimate pathway and neuraminidases of influenza A, and several inhibitors were identified. We believe that site-moiety maps are able to provide biological insights and are useful for drug discovery and lead optimization.

1.3 Thesis overview

The thesis is organized as follows. In Chapter 2, to identify pharmapathlogs, which are ortholog proteins sharing comparable binding environments in the same pathways, we developed the SiMMap server to infer key binding environments of binding sites via site-moiety maps. A site-moiety map describes the relationship between the moiety preferences and the physico-chemical properties of the binding site through anchors. An anchor includes three

essential elements including binding pockets with conserved interacting residues; moiety preferences; and interaction type. We provided initial validation of the site-moiety map on three targets, thymidine kinase, and estrogen receptors of antagonists and agonists. Experimental results show that an anchor is often a hot spot and the site-moiety map can help to assemble potential leads by optimal steric, hydrogen-bonding, and electronic moieties. When a compound highly agrees with anchors of site-moiety map, this compound often activates or inhibits the target protein.

In Chapter 3, we presented the pharmapathlog-based screening strategy by combing site-moiety maps. The concept of strategy is to simultaneously screen multiple protein orthologs in the same pathways cross pathogens, and extract conserved binding environments (core anchors) of these proteins for discovering multitarget inhibitors. Proteins sharing similar binding environments (core anchors) are considered pharmapathlogs, and could be inhibited by the same inhibitors. In this study, we applied site-moiety maps to describe the core binding environments, called core anchors, which present conserved binding pockets with specific physico-chemical properties, similar moieties of ligands, and consensus interaction types, all of which are essential to perform biological functions during species evolution. Hence, the proteins with comparable core anchors can be regarded as pharmapathlogs, and the core anchors of pharmapathlogs can be used to identify multitarget inhibitors. Then this strategy was applied to identify multitarget inhibitors of shikimate pathways for *M. tuberculosis* and *H. pylori*, which are human pathogens and causes peptic ulcer disease and chronic infectious disease, respectively¹⁷⁻¹⁹. By use of the strategy, we successfully discovered three multitarget inhibitors with low IC₅₀ values (<10.0 μM) for simultaneously inhibiting shikimate dehydrogenase and shikimate kinase of in the shikimate pathway. The preliminary results show that the pharmapathlog-based screening strategy is useful to discover multitarget inhibitors.

In Chapter 4, we applied the pharmapathlog-based screening strategy to identify and

optimize inhibitors of neuraminidases. The neuraminidase has been considered an attractive target for the treatment of influenza infection^{3,4}. By use of the strategy, we designed five derivatives simultaneously inhibiting H1N1 and H5N1 neuraminidases with IC₅₀ values in the <10 nanomolar range. The derivatives could be new type inhibitors located at the 150-cavity, which is adjacent to the sialic acid binding site. Moreover, we found three novel inhibitors with IC₅₀ values <10 μM through the strategy. Our experimental results reveals the three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 NAs without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains.

In the final chapter, we summarized the results of this thesis, and then discuss the future works. Currently, we propose a new concept of "pharmacologs" by extending the concept of "pharmapathlogs". Pharmacologs are a group of proteins sharing comparable binding environments in multiple pathways, and can be bound the same compounds. The major difference between pharmapathlogs and pharmacologs is that pharmacologs could be constituted by targets of multiple disease-related pathways, whereas pharmapathlogs only contain targets in a pathway. We believe that this new concept is able to identify multitarget inhibitors of multiple disease-related pathways, and then enhance therapeutic efficacy for human diseases.

Chapter 2. Site-moiety map to recognize interaction

preferences between protein pockets and compound


moieties

Identifying pharmapathlogs, which are protein orthologs sharing similar core binding environments in the same pathway and can be inhibited by the same compounds, is a challenging task. Because proteins in the same pathways often lack structural and sequence homology^{15,47}, it is inapplicable to use sequence or structure alignment methods to find the core binding environments within these proteins. For example, in the shikimate pathway, the sequence identity between shikimate dehydrogenase (SDH) and shikimate kinase (SK) is 8.3%, and the root mean square deviation between SD and SK structures is 4.8Å. To address this issue, we developed a new server, called SiMMap⁴⁶, to describe key binding environments of protein binding sites. Based on the SiMMap server, pharmapathlogs could be identified through comparing key binding environments of proteins binding sites without considering sequence and structure similarities. In addition, consensus binding environment among these proteins are regarded as core binding environments, which can be applied to find multitarget inhibitors.

In this chapter, I cooperated with our laboratory members, including Dr. Yang, Yen-Fu Chen, Shen-Rong Lin, and Yu-Chi Huang, to develop the SiMMap server. The SiMMap server statistically derives site-moiety map with several anchors, which describe the relationship between the moiety preferences and physico-chemical properties of the binding site, from the interaction profiles between query target protein and its docked (or co-crystallized) compounds. Each anchor includes three basic elements: a binding pocket with conserved interacting

residues, the moiety composition of query compounds, and pocket-moiety interaction type (electrostatic, hydrogen-bonding, or van der Waals). We provide initial validation of the site-moiety map on three targets, thymidine kinase, and estrogen receptors of antagonists and agonists. Experimental results show that an anchor is often a hot spot and the site-moiety map can help to assemble potential leads by optimal steric, hydrogen-bonding, and electronic moieties. When a compound highly agrees with anchors of site-moiety map, this compound often activates or inhibits the target protein. We believe that the site-moiety map is useful to extract the core binding environments (core anchors) and find pharmapathlogs. The SiMMap web server is available at <http://simfam.life.nctu.edu.tw/>. The results were published in *Nucleic Acids Research*.

2.1 Introduction



As the number of protein structures increases rapidly, structure-based drug design and virtual screening approaches are becoming important and helpful in lead discovery^{2,22,23,32}. A number of docking and virtual screening methods^{25,33,34,39} have been utilized to identify lead compounds, and some success stories have been reported⁴⁸⁻⁵¹. However, identifying lead compounds by exploiting thousands of docked protein-compound complexes is still a challenging task. The major weakness of virtual screenings is likely due to incomplete understandings of ligand binding mechanisms and the subsequently imprecise scoring algorithms^{2,22,23}.

Most docking programs^{25,34,39} use energy-based scoring methods which are often biased toward both the selection of high molecular weight compounds and charged polar compounds^{40,41}. These approaches generally cannot identify the key features (*e.g.*, pharmacophore spots) that are essential to trigger or block the biological responses of the target

protein. Although pharmacophore techniques⁴² have been applied to derive the key features, these methods require a set of known active ligands that were acquired experimentally. Therefore, the more powerful techniques for post-screening analysis to identify the key features through docked compounds and to understand the binding mechanisms provide a great potential value for drug design.

To address these issues, we presented the SiMMap server to infer the key features by a site-moiety map describing the relationship between the moiety preferences and the physico-chemical properties of the binding site. According to our knowledge, SiMMap is the first public server that identifies the site-moiety map from a query protein structure and its docked (or co-crystallized) compounds. The server provides pocket-moiety interaction preferences (anchors) including binding pockets with conserved interacting residues; moiety preferences; and interaction type. We verified the site-moiety map on three targets, thymidine kinase, and estrogen receptors of antagonists and agonists. Experimental results show that an anchor is often a hot spot and the site-moiety map is useful to identify active compounds for these targets. We believe that the site-moiety map is able to provide biological insights and is useful for drug discovery and lead optimization.

2.2 Methods and Materials

Figure 2.1 presents an overview of the SiMMap server for identifying the site-moiety map with anchors, describing moiety preferences and physico-chemical properties of the binding site, from a query protein structure and docked compounds. The server first uses checkmol (<http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm>) to recognize the compound moieties and utilizes GEMDOCK³³ to generate a merged protein-compound interaction profile (Fig. 2.1B), including electrostatic (E), hydrogen-bonding (H) and van der Waals (V) interactions.

According to this profile, we infer anchor candidates by identifying the pockets with significant interacting residues and moieties with Z-score ≥ 1.645 . The neighbor anchor candidates, which are the same interaction type and the distances between their centers are less than 3.5\AA , are grouped into one anchor. These anchors form the site-moiety map describing interaction preferences between compound moieties and the binding site of the query (Figs. 2.1C and 2.1D). Finally, this server provides graphic visualization for the site-moiety map; anchors with moiety structures and compositions; pocket-moiety interactions; and the relationship between anchors and moieties of query compounds.



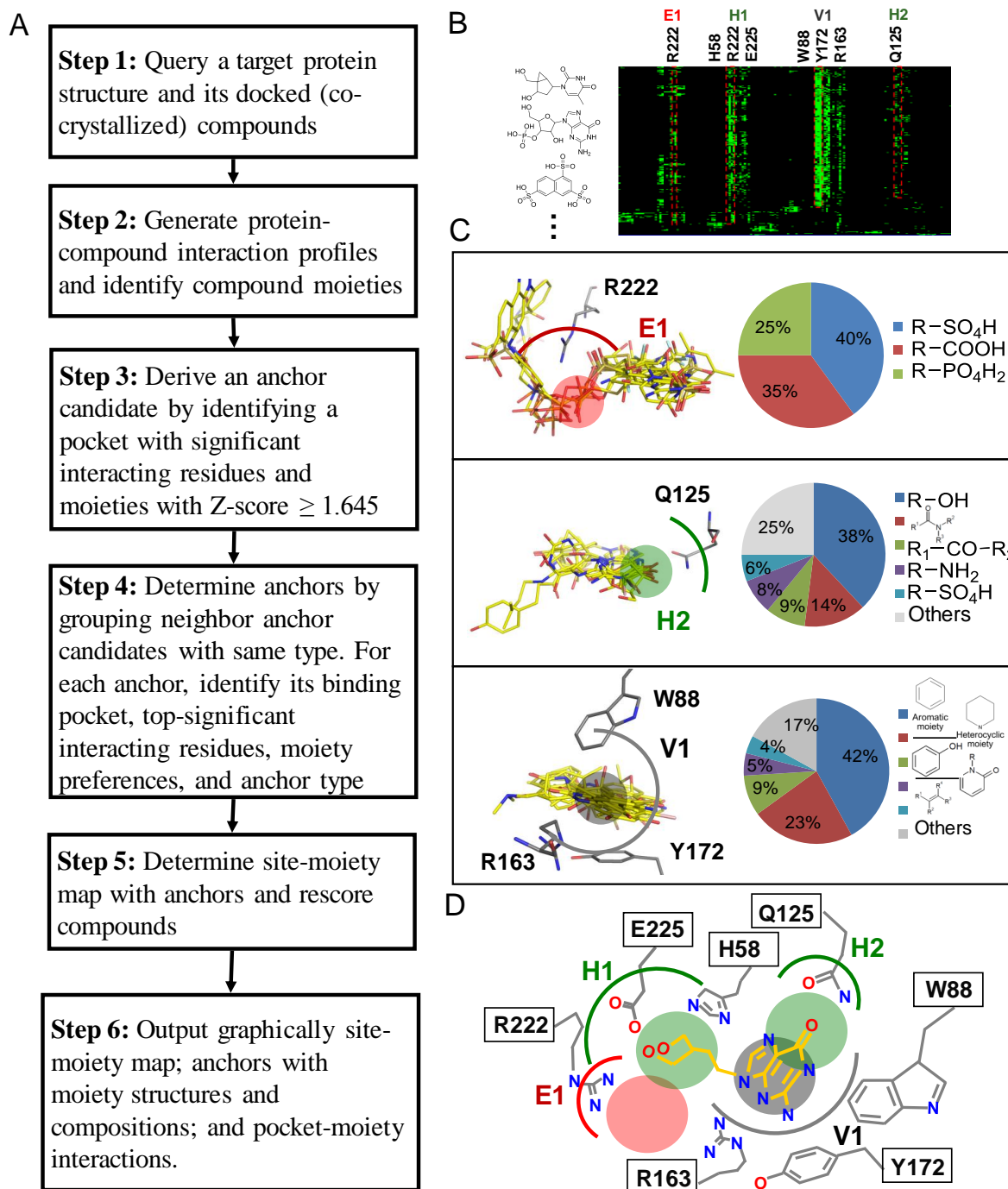


Figure 2.1. Overview of the SiMMap server for the site-moiety map using herpes simplex virus type-1 thymidine kinase (TK) and 1000 docked compounds as the query. (A) Main procedure. The Z-score cutoff was set to 1.645, which was commonly used in statistics (95% confidence level). (B) The merged protein-compound interaction profile. A cell is colored by green if there is an interaction (*e.g.*, electrostatic, hydrogen-bonding, or van der Waals interaction) between a compound and a residue (C) The pocket-moiety interaction preferences of three anchors: E1

(electrostatic), H2 (hydrogen-bonding), and V1 (van der Waals). Each anchor consists of a binding pocket with conserved interacting residues, the moiety composition and anchor type; (D) The site-moiety map with four anchors.

2.2.1 Site-moiety map, anchor and pocket

The anchor (pocket-moiety interaction preference) is the core of a site-moiety map. An anchor possesses three essential elements: (1) a binding pocket with conserved interacting residues and specific physico-chemical properties; (2) moiety preferences of the pocket; (3) pocket-moiety interaction type (E, H, or V). An anchor can be considered as "key features" for representing the conserved binding environment element or a "hot spot" which involves biological functions. In addition, we regard a binding pocket, which consists of several residues significantly interacting to compound moieties, as a part of the binding site. The binding pocket often possesses specific physico-chemical properties and geometric shape to bind preferred moieties. The site-moiety map, which can help to assemble potential leads by optimal steric, hydrogen-bonding, and electronic moieties, is useful for drug discovery and understanding biological mechanisms.

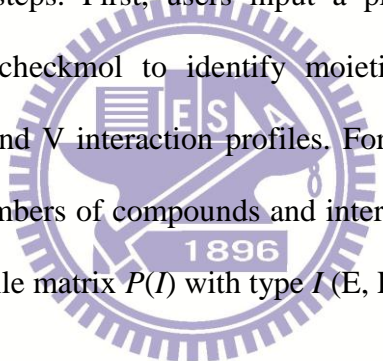
2.2.2 Data sets

To describe and evaluate the utility of the SiMMap server, we tested the server on three target proteins for virtual screening. These proteins are herpes simplex virus type-1 thymidine kinase (TK, PDB code 1kim⁵²), estrogen receptor α for antagonists (ER, PDB code 3ert⁵³), and estrogen receptor α for agonists (ERA, PDB code 1gwr⁵⁴). Each compound set consists of 10 known active ligands and 990 compounds selected randomly from Available Chemical

Directory (ACD) proposed by Bissantz *et al.*⁵⁵. Currently, the docked conformations of these 1000 compounds were generated by the in-house developed GEMDOCK program³³ which is comparable to some docking methods (*e.g.*, DOCK, FlexX, and GOLD) on the 100 protein-ligand complexes and some screening targets^{33,40}. In addition, GEMDOCK has been successfully applied to identify inhibitors and binding sites for some targets^{51,56,57}.

2.2.3 Main procedure of constructing a site-moiety map

The SiMMap server performs six main steps for a query (Fig. 2.1A). Here, we used TK as an example for describing these steps. First, users input a protein structure and its docked compounds. The server used checkmol to identify moieties of docked compounds and GEMDOCK to generate E, H and V interaction profiles. For each profile, the matrix size is $N \times K$ where N and K are the numbers of compounds and interacting residues of query protein, respectively. An interaction profile matrix $P(I)$ with type I (E, H, or V) is represented as


$$P(I) = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1K} \\ p_{2,1} & p_{2,2} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{NK} \end{bmatrix}$$

, where $p_{i,j}$ is a binary value for the compound i interacting to the residue j (Fig. 2.1B). For H and E profiles, $p_{i,j}$ is set to 1 (green) if an atom pair between the compound i and the residue j forms hydrogen-bonding or electrostatic interactions, respectively; conversely, the interaction is set to 0 (black). For van der Waals (vdW) interaction, an interaction is set to 1 when the energy is less than -4 (kcal/mol).

SiMMap identified consensus interactions between residues and compound moieties with similar physical-chemical properties through the profiles. For each interacting residue (a

column of the matrix $P(I)$ (Fig. 2.1B), we used Z-score value to measure the interacting conservation between this residue and moieties. Z-score values are often used to evaluate statistical significances. Here, we used to measure how significant the consensus interactions are in the protein-compound interaction profiles. High Z-score values indicate that the interactions are highly consensus. The standard deviation (σ) and mean (μ) were derived by random shuffling 1,000 times in a profile. The Z-score of the residue j is defined as

$$Z_j = \frac{f_j - \mu}{\sigma}, \text{ where } f_j \text{ is the interaction frequency and given as } f_j = \sum_{i=1}^N \frac{P_{ij}}{N}.$$

Spatially neighbor interacting residues and moieties with statistically significant Z-score ≥ 1.645 were referred as an anchor candidate. The Z-score threshold is set to 1.645 (95% confidence level) using Student's t -test. Neighbor anchor candidates, which are spatially overlapped and the same anchor type, were clustered as an anchor and the anchor center is the weighted geometric center of their interacting compound moieties. Here, two anchors were merged if the distance of two anchor centers is less than 3.5 Å. In each anchor, top three residues with the highest Z-score values were regarded as key residues forming a binding pocket. For each anchor, we identified its moieties of docked compounds according to the moiety library derived from checkmol, and calculated the moiety composition (Fig. 2.1C). These anchors form the site-moiety map (Fig. 2.1D) of the query.

SiMMap can be applied to identify active compounds for structure-based virtual screening. One of weaknesses of virtual screening is likely incomplete understanding of the chemistry involved in ligand binding and the subsequently imprecise scoring algorithms. When a compound highly agrees with the anchors of the site-moiety map, this compound often activates or inhibits the target. The SiMMap server scores a compound by combining predicted binding energy of GEMDOCK and the anchor score between the map and the compound. The SiMMap score, $S(i)$, for a compound i is defined as

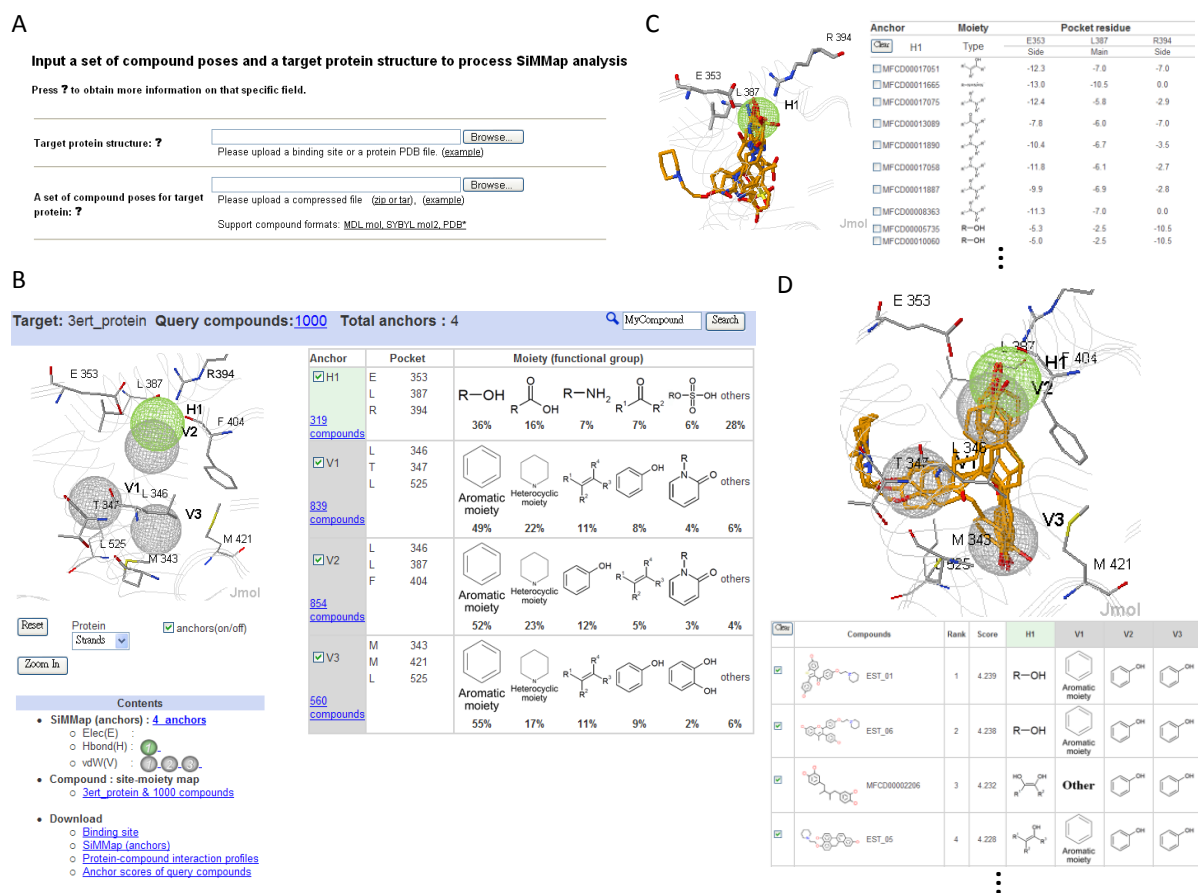
$$S(i) = \sum_{a=1}^n AS_a(i) + (-0.001) \frac{E(i)}{M^{0.5}} \quad (3.1)$$

, where $AS_a(i)$ is the anchor score of compound i in the anchor a , n is the number of anchors, $E(i)$ is the docked energy of compound i , and M is the atom number of compound i . The anchor score is set to 1 when the compound i agrees the moiety preference of the anchor a . Here, the anchor score and the term $M^{0.5}$ are useful to reduce the deleterious effects of selecting high molecular weight compounds⁵⁸. Based on SiMMap scores, we can obtain new ranks of query compounds.

2.2.4 Input and output of the SiMMap server

SiMMap is an easy-to-use web server (Fig. 2.2). Users input a protein structure without ligands in PDB format and its docked or co-crystallized compounds in MDL mol, SYBYL mol2, or PDB format (Fig. 2.2A). These docked compounds should be generated by any external docking methods (*e.g.*, DOCK, FlexX, GOLD and GEMDOCK) before users uploaded these compounds. Typically, the SiMMap server yields a site-moiety map within 5 minutes if the number of query compounds is less than 100. This server provides the graphic visualization of the site-moiety map and anchors elements, including a binding pocket with interacting residues, moiety compositions and structures, numbers of involved compounds, and anchor types (Fig. 2.2B). For each anchor, this server shows docked conformations of compounds and the detailed atomic interactions between pocket residues and moieties (Fig. 2.2C). In addition, SiMMap shows the new rank and compound moiety structures fitting the anchors for each query compound (Fig. 2.2D). SiMMap uses two open source tools for graphic visualization: Jmol (<http://www.jmol.org/>) for displaying three-dimensional protein and compound structures with anchors and OASA (http://bkchem.zirael.org/oasa_en.html) for visualizing compound

structures. The server allows users to download the anchor coordinates in the PDB format; interaction profiles; new ranks and anchor scores of query compounds.



2.3 Results and Discussion

2.3.1 Thymidine kinase

The SiMMap server inferred the site-moiety map of TK. This map consisted of four anchors (*i.e.*, E1, H1, H2, and V1 ([Fig. 2.1D](#))) and the moiety composition and conserved interacting residues of each anchor ([Fig. 2.1C](#)). The E1 anchor possesses a binding pocket with residue R222, and three moiety types (*i.e.*, sulfuric acid monoester (40%), carboxylic group (35%) and phosphoric acid monoester (25%)) derived from 57 compounds. The E1 includes the phosphate moiety of ATP and its residue R222 playing a major role to interact with the substrate^{59,60}. Furthermore, the H1 anchor is a polar pocket with three residues (H58, R222, and E225) which often form hydrogen bonds with polar moiety types among 308 compounds, for example, hydroxyl group (22%), carboxylic acid (8%), ketone (8%), ether (7%), and carboxylic amide (7%). The H2 anchor consists of the residue Q125 and 157 moieties divided into five major moiety types, including hydroxyl group (38%), carboxylic amide (14%), ketone (9%), amine (8%) and sulfuric acid monoester (6%). Finally, the V1 anchor has a binding pocket with residues W88, R163, Y172 and bulky moieties, such as aromatic ring (42%), heterocyclic group (23%), phenol (9%), and oxohetarene (5%).

The preferred moiety types of an anchor are suitable groups interacting to conserved residues of the binding pocket. The moiety preference is able to guide the suggestion of functional group substitutions for lead structures. For example, the moiety preferences of these four anchors ([Fig. 2.1D](#)) cover the moiety types derived from 15 TK co-crystal ligands ([Table 2.1](#)). In addition, these compounds contain carboxylic amide or amine groups in the H1 anchor. This result shows that the pocket-moiety interactions of these 15 complexes are highly consistent with the pocket-moiety interaction preferences.

Table 2.1. The relationship between the anchors and moieties of 15 co-crystallized ligands for

TK

PDB code	Compound structure	Anchor			
		E1	H1	H2	V1
3vtk		$\text{RO}-\overset{\text{O}}{\underset{\text{OH}}{\text{P}}}-\text{OH}$	R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1vtk		$\text{RO}-\overset{\text{O}}{\underset{\text{OH}}{\text{P}}}-\text{OH}$	R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1p7c		$\text{RO}-\overset{\text{O}}{\underset{\text{OH}}{\text{P}}}-\text{OH}$	R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1of1			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1ki6			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1ki8			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1e2k			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1ki7			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1ki4			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1kim			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1e2p			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	
1ki3			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	Aromatic moiety
1ki2			R-OH	$\text{R}-\overset{\text{O}}{\underset{\text{R}_2}{\text{N}}}-\text{R}_1$	Aromatic moiety
1qhi			R-OH	R-NH ₂	Aromatic moiety
2ki5			R-OH	R-NH ₂	Aromatic moiety

2.3.2 Estrogen receptor

We used estrogen receptor (ER), a therapeutic target for osteoporosis and breast cancer⁶¹, as the second example. Based on 1000 docked compounds and ER, the SiMMap server identifies four anchors (H1, V1, V2, and V3) and provides moiety preferences and compositions in these anchors (Fig. 2.2B). The H1 anchor comprises three residues (E353, L387, and R394) and five main moiety types: hydroxyl group (36%), carboxylic acid (16%), amine (7%), ketone (7%), and sulfuric acid monoester (6%) summarized from 319 compounds. Furthermore, three residues (L346, T347, and L525) and 839 compounds are involved in the V1 anchor, preferring five moiety types (*i.e.*, aromatic ring (49%), heterocyclic group (22%), alkenes (11%), phenol (8%), and oxohetarene (4%)). The anchor V2 is a hydrophobic pocket containing L346, F404, and L387, and the former two residues are highly conserved⁶². These hydrophobic residues interact with aromatic ring (52%), heterocyclic group (23%), phenol (12%), alkenes (5%), and oxohetarene (3%). Finally, aromatic rings (55%), heterocyclic groups (17%), alkenes (11%), and phenols (9%) summarized from 560 compounds often form vdW contacts with the long side chains of M343, M421, and L525 in the anchor V3. The ring groups of antagonists are often stabilized by the side chains of M343, L346, T347, L387, M421, and L525. In this case, most selective estrogen receptor modulators of ER (*e.g.*, EST_01 (raloxifene), EST_06 (LY-326315,) and EST_05 (EM-343)) agree with these four anchors (Fig. 2.2D).

2.3.3 Discussion

Anchors identified by the SiMMap server often contain key pockets and moieties. To initially validate the anchors for biological mechanisms (*e.g.*, ligand binding and catalysis mechanisms), we selected 15 TK and 22 ER co-crystallized ligands (Fig. 2.3, Fig. 2.4, and Table 2.1). The

corresponding moieties of these co-crystallized ligands were highly matched the anchors derived from 1000 docked compounds (10 known active ligands and 990 randomly selected compounds described in Data sets). The site-directed mutagenesis shows that the conserved interacting residues of the anchors are often essential for ligand binding and catalysis mechanisms. For example, the positive charged residue R222 in E1 interacts with the phosphate group of TK substrates for phosphorylation⁶³ (Figs. 2.3A and 2.3B). The site-directed mutagenesis indicates that Q125 in H2 is essential for the substrate specificity⁵⁹ and the triple mutant ,H58L/M128F/Y172F (H1 and V1), shows the drug resistance to the compound acyclovir⁶⁴. In addition, the hydrogen-bonding interaction between E225 and the hydroxyl group of the substrates is able to help stabilize the LID region for the catalytic reaction⁶⁴. For ER target, 22 ER co-crystallized ligands contain three consistent moieties that are hydroxyl group and aromatic rings (Fig. 2.4). The hydroxyl group forms hydrogen bonds with R394 and E353 in H1, and the aromatic ring yields vdW contacts with L346, L387, and F404 in V2. The other consistent aromatic ring forms vdW contacts with L346, T347, and L525 in V1. These results show that an anchor is often a hot spot and involved in biological functions.

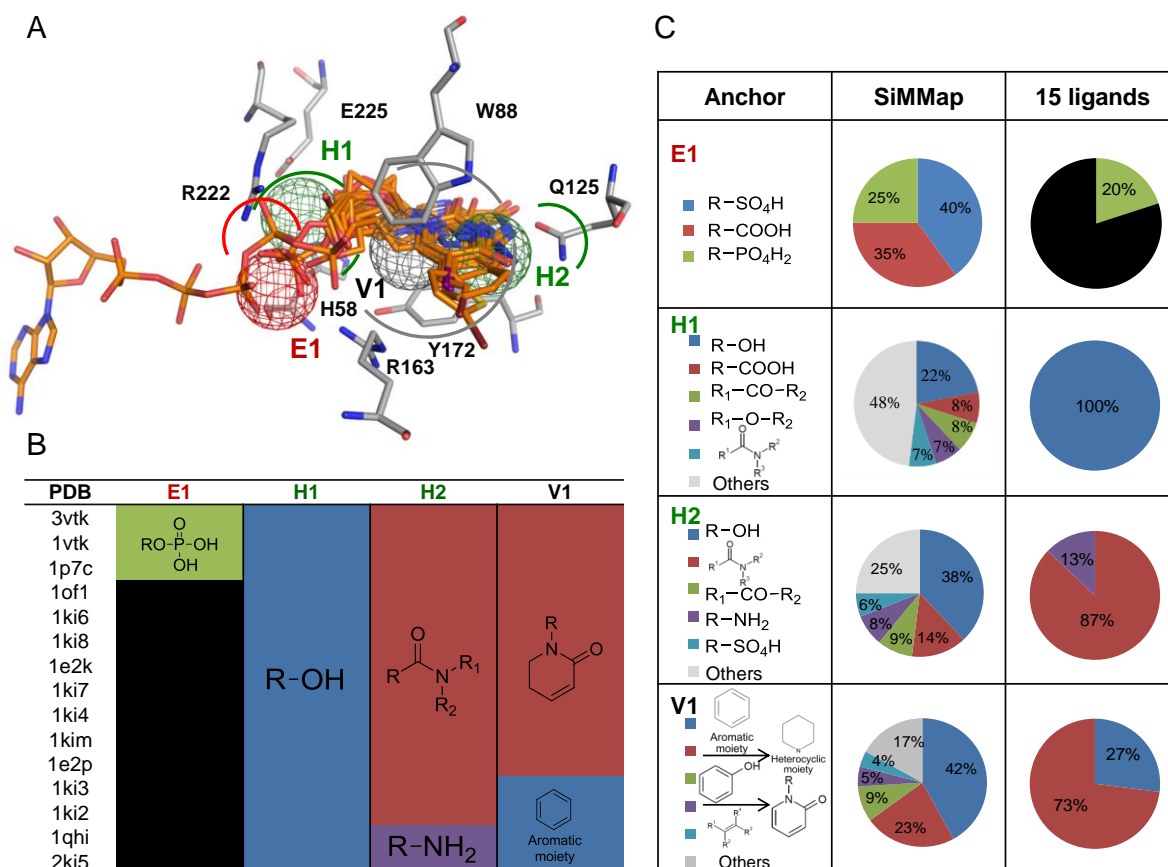


Figure 2.3. The relationships between the site-moiety map and 15 co-crystallized ligands of TK. (A) The mapping between four inferred anchors (binding pocket with conserved interacting residues) and these 15 ligands in the active site. (B) The moieties of these 15 ligands in each anchor. (C) The moiety compositions of 1000 docked compounds (SiMMap) and these 15 ligands.

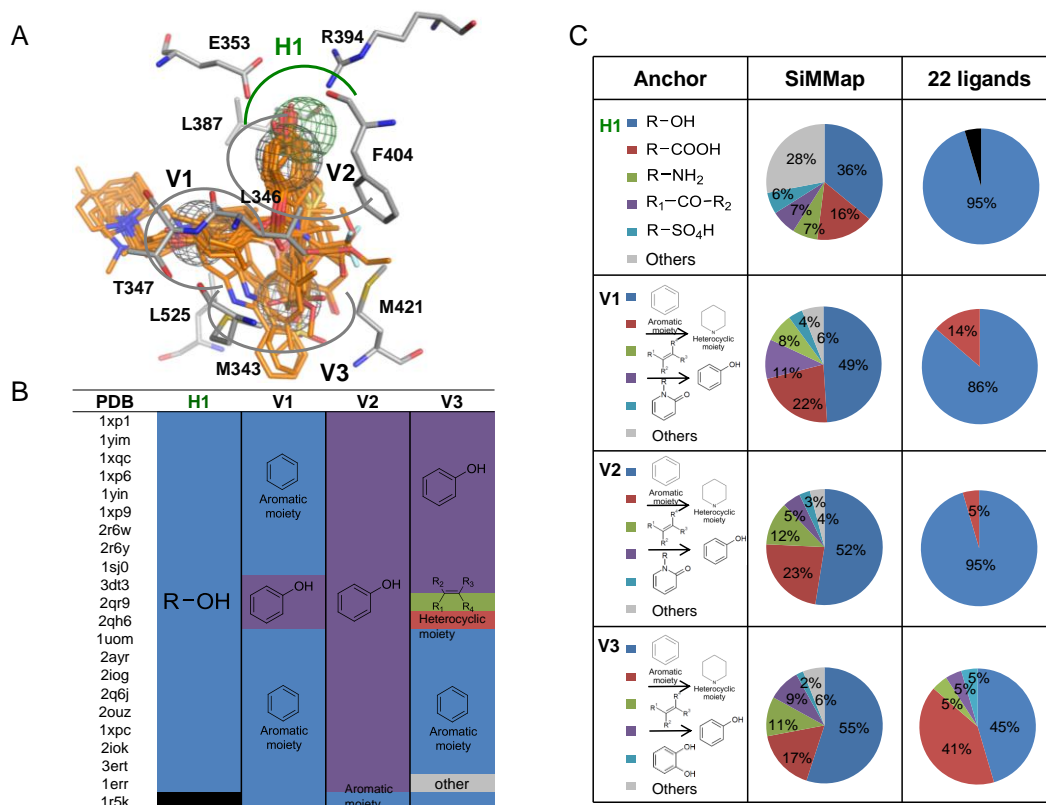


Figure 2.4. The relationships between the site-moiety map and 22 co-crystallized ligands of ER. (A) The mapping between four inferred anchors (binding pocket with conserved interacting residues) and these 22 ligands in the active site. (B) The moieties of these 22 ligands in each anchor. Black cell presents that the moiety of the compound does not agree with the anchor H1. (C) The moiety compositions of 1000 docked compounds (SiMMap) and these 22 ligands.

To provide initial validation of the SiMMap server for virtual screening, we selected TK, ER, and ERA with 1000 compounds as test sets. First, we compared the accuracies of SiMMap with those of GEMDOCK on these three targets based on true positive rates (Fig. 2.5). SiMMap, combining anchor scores and docking energies (Equation (3.1)), outperforms GEMDOCK on these cases. We then compared SiMMap with other three programs (DOCK, FlexX, and GOLD) on TK and ER sets. All approaches were tested using the same proteins and compound sets (Table 2.2). When the true positive rate was 90%, the false positive rates were 6.8% (SiMMap), 25.5% (DOCK), 13.3% (FlexX), and 9.1% (GOLD) for TK and were 1.1% (SiMMap), 17.4% (DOCK), 70.9% (FlexX), and 8.3% (GOLD) for ER.

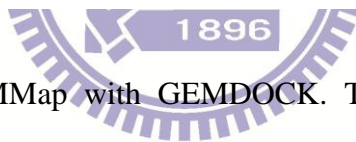
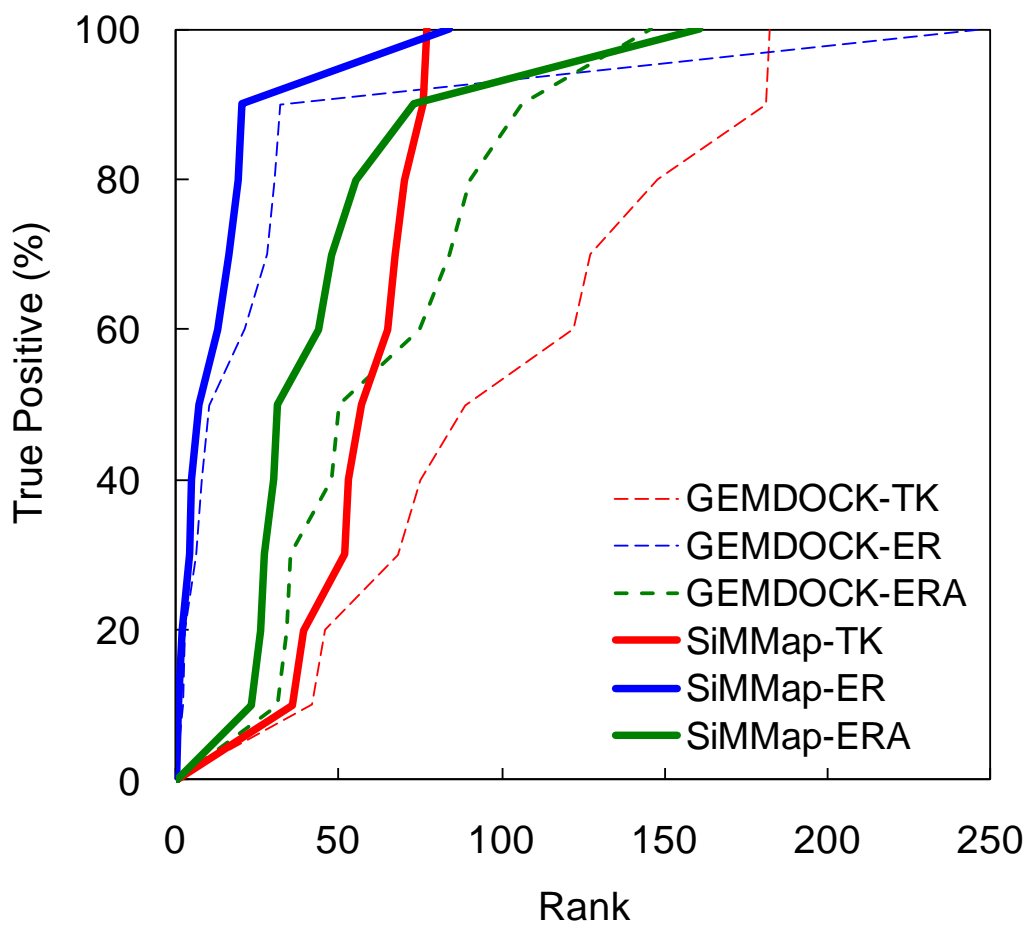


Figure 2.5. Comparison of SiMMap with GEMDOCK. The SiMMap server (solid lines) consistently outperforms GEMDOCK (dash lines) on three targets: TK (red), ER (blue) and ERA (green).

Table 2.2. Comparing SiMMap with other methods on thymidine kinase and estrogen receptor by false-positive rates

True positive rate (%)	Thymidine kinase (TK)				Estrogen receptor (ER)			
	SiMMap	DOCK ^a	FlexX ^a	GOLD ^a	SiMMap	DOCK ^a	FlexX ^a	GOLD ^a
80	6.3 ^b	23.4	8.8	8.3	1.1	13.3	57.8	5.3
90	6.8	25.5	13.3	9.1	1.1	17.4	70.9	8.3
100	6.8	27	19.4	9.3	7.5	18.9	NA	23.4

^a Summarized from Bissantz *et al.*⁵⁵

^b The false-positive rate (%) is obtained based on 1000 compounds.

The compound, which agrees with anchors of the site-moiety map, is often able to activate or inhibit the target protein (Tables 2.1, 2.3 and 2.4). In addition, the anchor score (i.e. $AS(i)$ defined in Equation (3.1)) of SiMMap can be used to reduce the ill-effect of the energy-based scoring methods which are often biased toward both the selection of high molecular weight compounds and charged polar compounds^{40,41}. For example, according to the SiMMap scores (Equation (3.1)), the ranks of MFCD0005750 (adenylic acid), MFCD0005753 (deoxyadenylic acid) and MFCD0005763 (3'-guanylic acid) are 1, 3, and 9, respectively. These three compounds are thymidine analogs and agree with the four anchors of TK (Fig. 2.1 and Table 2.1). For the top ranks of ER, MFCD0002206 (masoprocol) and MFCD00012748 were also the analogs of the active compounds (Table 2.4). The anchor score of SiMMap was helpful to reduce the highly polar compounds (*e.g.*, MFCD00011393 and MFCD00003569 in TK; MFCD00004690 and MFCD00013089 in ER) whose anchor scores are low. The anchor score of SiMMap can easily combine with other energy-based scoring functions.

Table 2.3. The mapping between the anchors and active and typical compounds for TK

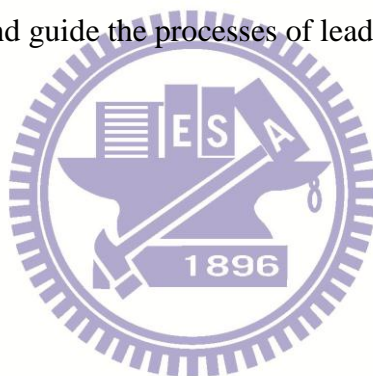
Compound	Structure	GEMDOCK rank	SiMMap rank	SiMMap score	E1	H1	H2	V1
TK_01		89	57	3.253		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_02		182	76	3.233		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_03		127	67	3.243		$R^1-C(=O)-N(R^2)(R^3)$	$R^1-C(=O)-N(R^2)(R^3)$	
TK_04		122	65	3.244		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_05		46	39	3.262		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_06		42	36	3.263		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_07		75	53	3.255		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_08		68	52	3.256		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_09		181	77	3.233		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
TK_10		148	70	3.239		R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
MFCD00005750		8	1	4.283	$RO-P(=O)(OH)_2$	R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
MFCD00005753		19	3	4.271	$RO-P(=O)(OH)_2$	R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
MFCD00005763		87	9	4.253	$RO-P(=O)(OH)_2$	R-OH	$R^1-C(=O)-N(R^2)(R^3)$	
MFCD00011393		4	384	1.295	$RO-P(=O)(OH)_2$			
MFCD00003569		9	118	2.281			$R^1-C(=O)-N(R^2)(R^3)$	Other

Table 2.4. The mapping between the anchors and active and typical compounds for ER

Compound	Structure	GEMDOCK rank	SiMap rank	SiMap score	H1	V1	V2	V3
EST_01		2	1	4.239	R-OH	 Aromatic moiety		
EST_02		32	19	4.216	R-OH	 Aromatic moiety		
EST_03		28	16	4.217	R-OH			
EST_04		8	5	4.226	R-OH	 Aromatic moiety		 Aromatic moiety
EST_05		6	4	4.228		 Aromatic moiety		
EST_06		3	2	4.238	R-OH	 Aromatic moiety		
EST_07		21	13	4.218		 Aromatic moiety		Other
EST_08		10	7	4.225	R-OH	Other		Other
EST_09		30	20	4.216	R-OH	 Aromatic moiety		 Aromatic moiety
EST_10		246	84	4.193	R-OH	 Aromatic moiety		
MFCD00002206		4	3	4.232		Other		
MFCD00012748		17	11	4.221	R-OH	Other		
MFCD00004690		5	154	3.23	R-OH		Other	Other
MFCD00013089		25	617	2.218				

2.4 Conclusions

This work demonstrates the utility and feasibility of the SiMMap server for statistically inferring the site-moiety map describing the relationship between the moiety preferences and physico-chemical properties of the binding site. Our experimental results show that the site-moiety map is useful to reflect biological functions and identify active compounds from thousands of compounds. In addition, the site-moiety map can guide to assemble potential leads by optimal steric, hydrogen-bonding, and electronic moieties. We believe that the SiMMap server is able to provide the biological insights of protein-ligand binding models, enrich the screening accuracy, and guide the processes of lead optimization.



Chapter 3. Pharmapathlogs for discovering multitarget

inhibitors of shikimate pathways using site-moiety maps

Multitarget inhibitors of pharmapathlogs can enhance therapeutic treatments by inhibiting several disease-related proteins in the same pathways, and reduce probabilities of drug-resistances. However, identifying pharmapathlogs and their multitarget inhibitors are still challenging tasks since these proteins often lack structural and sequence homology. In this chapter, we used site-moiety map developed in the previous chapter to infer pharmapathlogs. Subsequently, we proposed a pharmapathlog-based screening strategy to identify multitarget inhibitors of pharmapathlogs. This strategy was initially tested in discovering multitarget inhibitors of *Helicobacter pylori* (*H. pylori*) and *Mycobacterium tuberculosis* (*M. tuberculosis*), which are human pathogens and causes peptic ulcer disease and chronic infectious disease. By uses of the strategy, three inhibitors that simultaneously showed good inhibition abilities (IC₅₀ values <10 μM) to shikimate dehydrogenase and shikimate kinase in the shikimate pathway were discovered with collaborations with Dr. Wen-Ching Wang and Dr. Wen-Chi Cheng of National Tsing Hua University. Furthermore, we found that residues of core anchors are more conserved than those of the other regions, revealing that core anchors of pharmapathlogs are often essential for catalysis or substrate binding during evolution. This suggests that designing inhibitors targeting the core anchors could decrease probabilities of drug-resistances. These experimental results reveal that the pharmapathlog-based strategy could be useful to infer pharmapathlogs and their multitarget inhibitors. We believe that the strategy can be further applied for drug design of human diseases.

3.1 Introduction

Pharmapathlogs, which are a group of protein orthologs in the same pathway that share similar binding environments and can be inhibited by the same compounds, provide clues to discover multitarget inhibitors. Therapeutics using multitarget inhibitors are more effective and less vulnerable to drug-resistances than those using single-target inhibitors. For example, cosalane, an antiretroviral drug, has potent antiviral activity by simultaneously inhibit multiple targets of HIV-1 proteins, including gp120, integrase, protease, and reverse transcriptase⁶⁵⁻⁶⁷. However, current processes of discovering multitarget inhibitors are often serendipitous, and binding mechanisms of these inhibitors may be retrospectively understood⁶⁸. One of the main reasons is that target proteins of multitarget inhibitors often lack structural and sequence homology¹⁴. In addition, the low sequence similarity of proteins in the same pathway makes discovering pharmapathlogs difficult as well¹⁵. As a result, a new strategy for inferring pharmapathlogs and their multitarget inhibitors will be valuable for drug design.

Here, we proposed the pharmapathlog-based screening strategy to infer pharmapathlogs and discover their multitarget inhibitors by applying site-moiety maps described in the previous chapter¹⁶. A site-moiety map can present and characterize key binding environments of a protein binding site by using anchors. An anchor contains three crucial elements, including conserved interacting residues constituting a binding pocket (*i.e.*, a part of the binding site), the preference of moieties (*i.e.*, functional groups), and pocket-moiety interaction type [electrostatic (E), hydrogen-bonding (H), or van der Waals (V)]. Therefore, we are able to infer pharmapathlogs by identifying proteins sharing consensus binding environments (core anchors) instead of using structure or sequence similarities. Furthermore, core anchors of pharmapathlogs could be used to find multitarget inhibitors by screening compounds.

Current emergence of antibiotic-resistant bacteria causes a great concern in the world⁸⁻¹⁰.

For example, heavy antibiotic use generates drug-resistant *M. tuberculosis* strains. *M. tuberculosis* is one of the most persistent human disease and has a high death toll between 1.6 and 2 million fatalities per year^{69,70}. Another example is drug-resistant *H. pylori* strains, which were resulted from overuse and misuse of antibiotics⁷¹. *H. pylori* is a human pathogen that causes peptic ulcer disease and gastric cancer, and infects about half the human population in the world^{17,18,72}. Therefore, discovering new antibiotics for *M. tuberculosis* and *H. pylori* is an urgent need.

To verify the utility of the pharmapathlog-based screening strategy, we applied this strategy to identify new antibiotics for shikimate pathway of *H. pylori* and *M. tuberculosis*. The shikimate pathway containing seven proteins is an attractive target pathway for drug development because the pathway is absent in human²⁰. By use of the strategy, we identified pharmapathlogs in the pathway, and successfully discovered three new multitarget inhibitors with low IC₅₀ values (<10.0 μM) for simultaneously inhibiting shikimate dehydrogenase (SDH) and shikimate kinase (SK) in the shikimate pathway. The experimental results show that the pharmapathlog-based screening strategy is useful to infer pharmapathlogs and their multitarget inhibitors. We believe that the new strategy is useful to discover new lines of antibiotics toward drug-resistant bacteria and enhance human health.

3.2 Methods and Materials

3.2.1 Overview of pharmapathlog-based screening strategy

The concept of the pharmapathlog-based screening strategy is to simultaneously screen multiple orthologous proteins in the same pathway and extract core binding environments among these proteins. In this study, we applied anchors of site-moiety maps to describe the key

binding environments of protein binding sites. Consensus anchors (core anchors) of multiple binding sites present conserved binding pockets with specific physico-chemical properties, similar moieties of ligands, and consensus interaction types, all of which are essential to perform pathway functions during evolution. The following criteria are considered: (1) the screening targets are protein ortho**logs** in the same **path**way; (2) they share comparable core binding environments (core anchors). Compounds agreeing with core anchors could simultaneously inhibit these screening targets.

Figure 3.1 presented the main steps of the pharmapathlog-based screening strategy to discover multitarget inhibitors of the shikimate pathway for *H. pylori* and *M. tuberculosis*. The pathway contains seven proteins for biosynthesis of aromatic compounds²⁰ and four structures of them are available. In this study, we initially selected two adjacent proteins in the pathway, SDH and SK, as the screening targets to verify the strategy (Fig. 3.1A). For the targets, including *H. pylori* SDH (HpSDH), *H. pylori* SK (HpSK), and *M. tuberculosis* SK (MtSK), their site-moiety maps were generated by screening compounds from public databases. The site-moiety maps revealed these targets shared comparable core anchors and can be considered pharmapathlogs (Fig. 3.1B). Subsequently, the core anchors were applied to select potential multitarget inhibitors from public compound databases (Fig. 3.1C). Finally, compounds that simultaneously agreed with the core anchors in the targets were selected for bioassay (Fig. 3.1D).

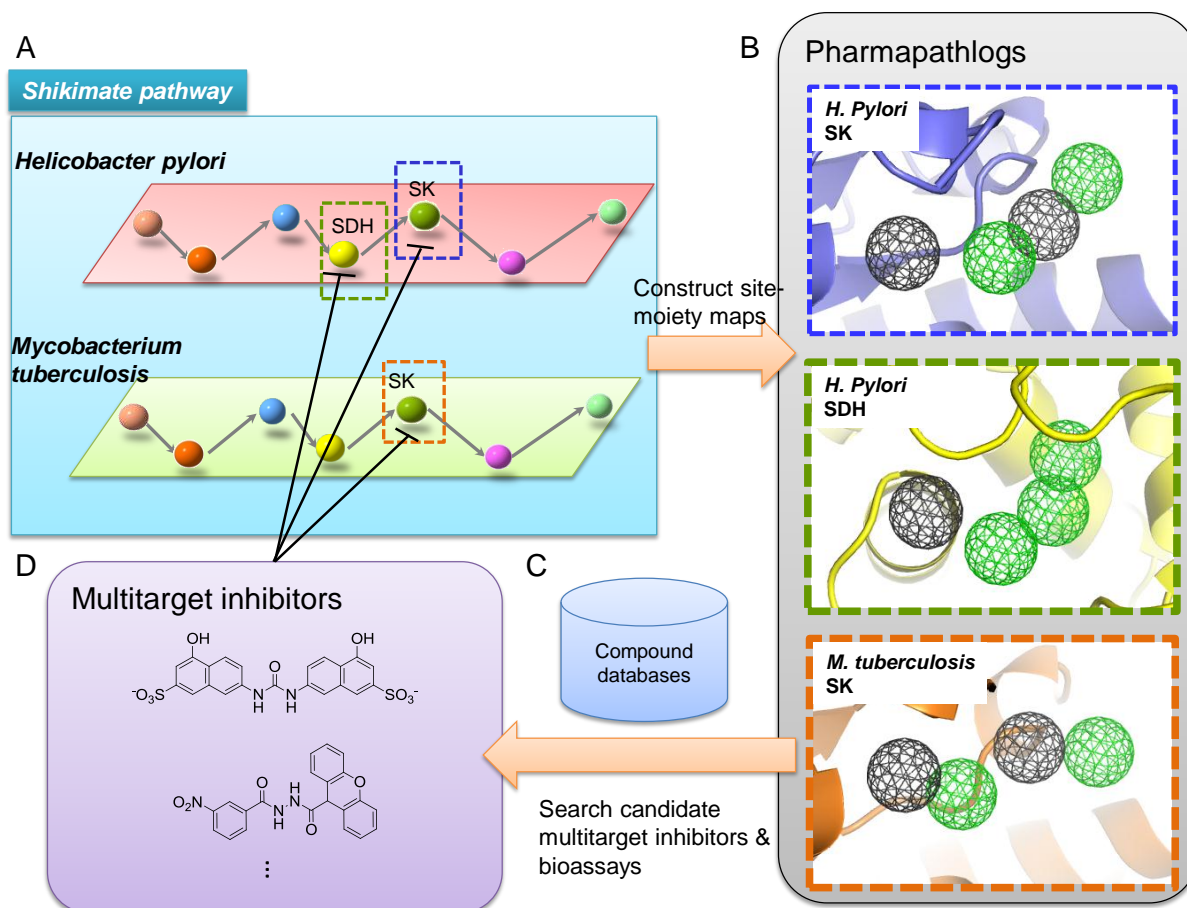


Figure 3.1. Overview of pharmapathlog-based screening strategy for identifying multitarget inhibitors. (A) Shikimate pathways of *Helicobacter pylori* (*H. pylori*) and *Mycobacterium tuberculosis* (*M. tuberculosis*). In this study, we used shikimate dehydrogenase (SDH) and shikimate kinase (SK) as the screening targets. (B) Pharmapathlogs and their core anchors. SDH and SK shared comparable core anchors despite of low sequence identity. Hydrogen-bonding and van der Waals anchors are colored green and grey, respectively. (C) Multitarget inhibitor selection from public compound databases. Inhibitors matching these core anchors of these targets were selected. (D) Identified multitarget inhibitors that simultaneously inhibited multiple targets in the shikimate pathways of *H. pylori* and *M. tuberculosis*.

Proteins in the same pathway may share many similarities in physical-chemical properties and shapes to interact similar ligands for a series of catalytic reactions. Because a product of one enzyme is a substrate of the next enzyme, it is possible to design multitarget inhibitors to

simultaneously inhibit several targets. For example, in the shikimate pathway, seven proteins catalyze several metabolites to synthesize chorismate⁷³. SDH and SK are the fourth and fifth enzymes (Fig. 3.2A). SDH converts 3-dehydroshikimate to shikimate, and SK subsequently converts shikimate to shikimate 3-phosphate^{73,74} (Fig. 3.2B). Although the sequence identity between SDH and SK is 8.3% and the root mean square deviation (RMSD) between SDH and SK structures is 4.8Å, some binding environments of the two proteins are conserved for the binding of these metabolites. Therefore, their core binding environments provide an opportunity to design multitarget inhibitors to simultaneously inhibit them (Fig. 3.2C).

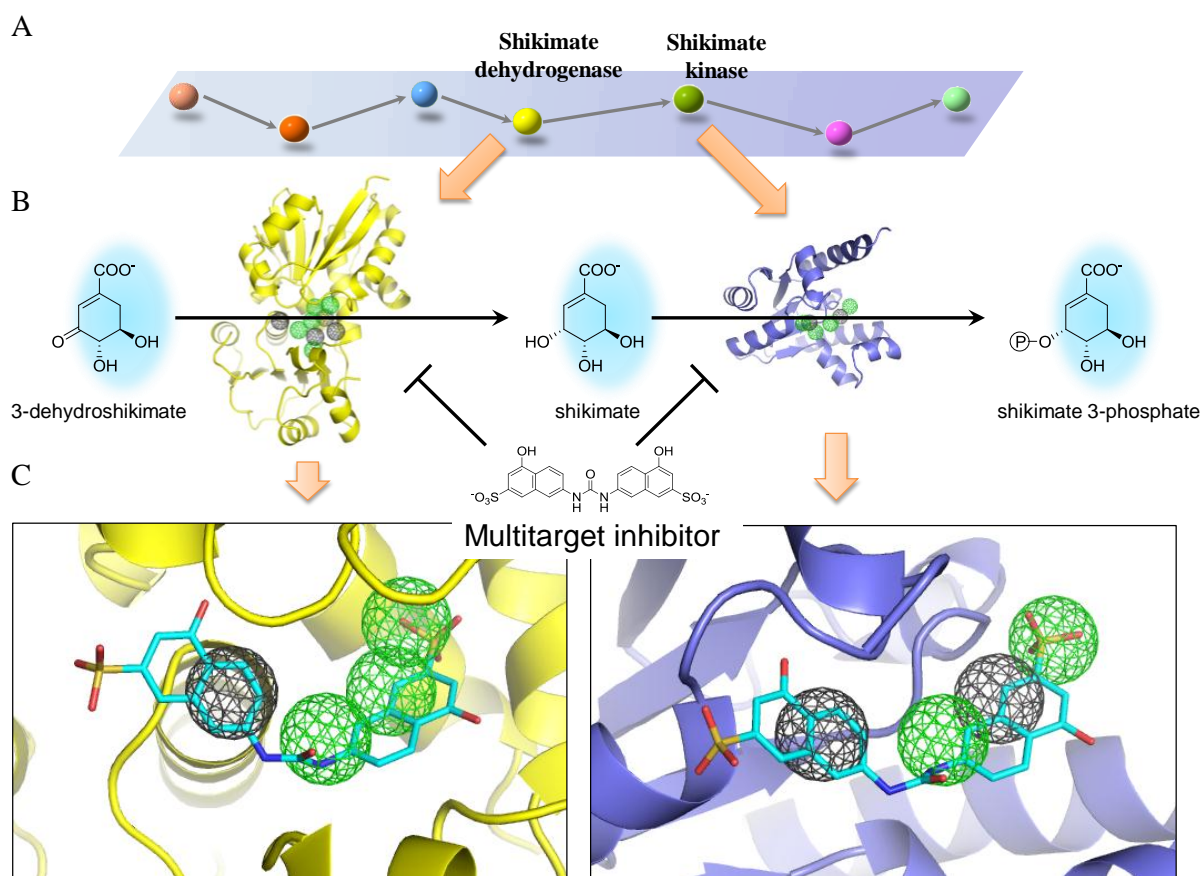


Figure 3.2. Discovery of multitarget inhibitors for inhibiting multiple proteins in a pathway. (A) Shikimate pathway that contains seven proteins. Shikimate dehydrogenase (SDH) and shikimate kinase (SK) are the fourth and fifth enzymes in the pathway. (B) Chemical reactions of SDH and SK. The two enzymes could share many similarities (core anchors) in their binding

environments because of catalyzing similar metabolites. (C) A multitarget inhibitor agrees with the core anchors and simultaneously inhibits SDH and SK.

3.2.2 Preparations of protein structures and screening

databases

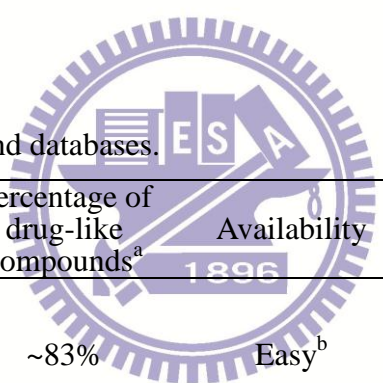
Apo-form structures of SDH and SK were selected for virtual screening because closed-form structures induced by bound ligands may limit the diversity of identified inhibitors. The apo-form structures of SDH were kindly provided by Dr. Wen-Chi Cheng (National Tsing Hua University, Hsinchu, Taiwan). The SK structures of *H. pylori* and *M. tuberculosis* were downloaded from Protein Data Bank (PDB), and their PDB codes are 1zuh⁷⁴ and 2iyt⁷⁵, respectively. Subsequently, to define binding sites, these three structures were aligned to a closed-form SK structure (PDB code 1zui⁷⁴) in complex with shikimate and PO₄ by using a structural alignment tool⁷⁶. Then, the bound ligands (*i.e.*, shikimate and PO₄) were used to determine the binding sites of the three structures. The binding sites of these structures were defined by considering the protein residues located ≤ 10 Å from the bound ligands.

Table 3.1 showed properties of five compound databases, including Maybridge, National Cancer Institute (NCI), Sigma, NCGC Pharmaceutical Collection (NPC)⁷⁷, and ZINC natural products^{78,79}. Most compounds of these databases obey Lipinski's rule of five⁸⁰ (Table 3.1), and these compounds could be regarded as drug-like compounds. Therefore, the drug-like compounds have a high probability of good absorption, distribution, metabolism, and excretion. Maybridge and Sigma contain ~83% and ~84% drug-like compounds, respectively. Compounds of the two databases are high quality and can be directly purchased from the suppliers. NCI contains > 200,000 compounds, and the most advantage of using this database

is that most compounds only charge a nominal fee. However, the supplier of NCI compounds does not guarantee the purity of most compounds. NPC provides a comprehensive collection of approved and investigational drugs, providing an opportunity to discover new uses for old drugs. The library of ZINC natural products contains a comprehensive collection of natural products. In history, natural products are the majority of developing new drugs⁸¹, and thereby natural products are promising sources for drug discovery.

In this study, we selected Maybridge and NCI to generate the site-moiety maps and discover multitarget inhibitors because of their rapid availability and low cost. The compound structures were obtained by CORINA⁸², and the compounds were filtered if their compound molecular weights are < 200 or > 650 dalton. In total, the number of the compounds selected for screening is 302,909.

Table 3.1. Properties of compound databases.



Database	Compound number	Percentage of drug-like compounds ^a	Availability	Description
Maybridge	56,515	~83%	Easy ^b	Provision of high-quality and drug-like compounds
National Cancer Institute	283,908	~85%	Based on inventory	Most compounds only charge a nominal fee
Sigma	115,507	~84%	Easy	Provision of high-purity and diverse compounds
NCGC				A comprehensive collection of approved and investigational drugs
Pharmaceutical Collection	14,814	~86%	Slightly difficult ^c	
ZINC natural products	89,425	~75%	Easy	A comprehensive collection of natural products

^a Drug-like compounds obey Lipinski's Rule of five.

^b Compounds can be directly purchased from the suppliers. ZINC natural products include the purchasing information (53 suppliers) of natural products

^c This database only includes suppliers of parts of compounds. For those compounds whose suppliers were unavailable, users have to rely on their own to find suppliers.

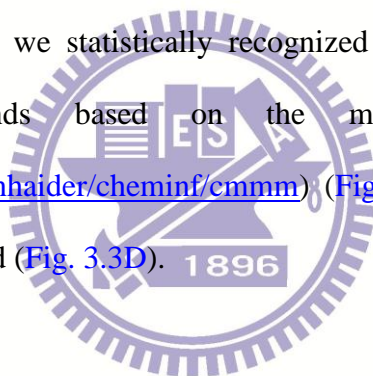
3.2.3 Computational screening and establishment of site-moiety maps

To establish site-moiety maps, docked poses of compounds in the target proteins were required. The 302,909 compounds were docked into the binding sites of the targets by using GEMDOCK³³ (Fig. 3.3A), which is an in-house developed docking program. GEMDOCK is compared to some docking methods (*e.g.*, DOCK³⁹, FlexX³⁴, and GOLD²⁵) in the binding pose prediction of the 100 protein-ligand complexes and the inhibitor identification of two virtual screening targets^{33,40}. Furthermore, we have successfully used GEMDOCK to identify novel inhibitors and binding sites for some targets^{51,56,57}. For each compound, the binding energy of the compound was predicted by the piecewise linear potential (PLP) of GEMDOCK. Subsequently, 1,000 top-ranked compounds and their respective binding sites were submitted to the SiMMap server for establishment of site-moiety maps¹⁶.

The SiMMap server inferred site-moiety maps consisting of anchors to recognize interaction preferences between protein pockets and moieties of the top-ranked compounds. First, the SiMMap server derived protein-compound interaction profiles of the compounds by applying the PLP of GEMDOCK for each target (Fig. 3.3B). The profiles described the interactions (*i.e.*, electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions) between the compounds and the protein residues. For the E or H profile, a profile entry between a compound and a protein residue was set to 1 if the compound forms electrostatic or hydrogen-bonding interactions with the residue (green regions in Fig. 3.3B); conversely the entry was set to 0 (black regions in Fig. 3.3B). For the V profile, a profile entry was set 1 if the V energy was less than -4 kcal/mol (green regions in Fig. 3.3B).

Consensus interactions of the profiles were recognized through Z-score, which is often

used to measure the statistical significance. For each profile, the Z-score value (Z_i) of the protein residue i was computed by $Z_i = \frac{A_i - \mu}{\sigma}$, where A_i is the interaction average of the residue i , and μ and σ are the mean and the standard deviation derived by 1,000 randomly shuffled profiles. Here, we regarded the interactions of the residue i as consensus interactions if its Z-score value was greater than 1.645 (95% confidence level). Consecutively, spatially neighbor residues with consensus interactions were regarded as anchor candidates, and the centers of these candidates were defined as the geometric centers of the interacting moieties of the compounds. Then, neighbor anchor candidates with the same interaction type were merged as an anchor if the distance between the centers of the anchor candidates was less than 3.5 Å. The interacting residues with the high Z-score values (≥ 1.645) of an anchor were regarded as key residues. For each anchor, we statistically recognized compositions of the interacting moieties of the compounds based on the moiety library of checkmol (<http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm>) (Fig. 3.3C). Finally, the site-moiety map of each target was generated (Fig. 3.3D).



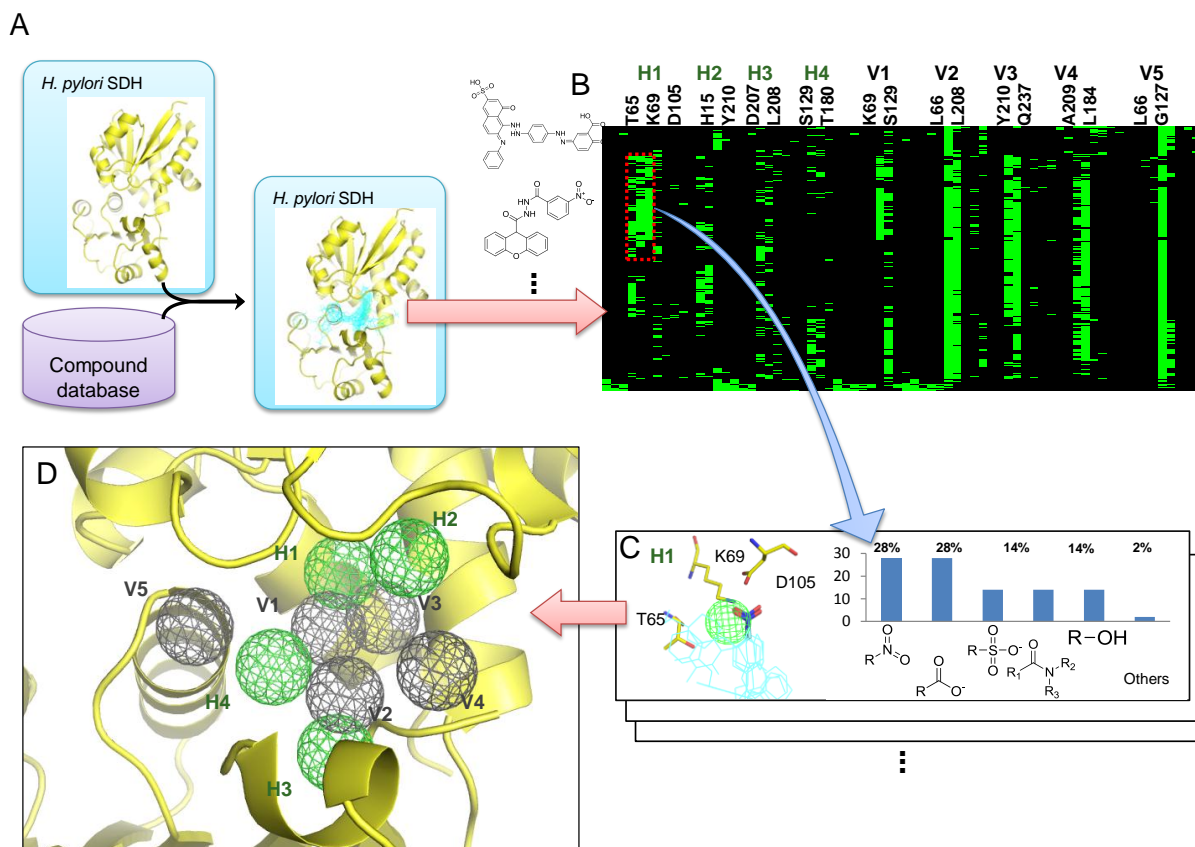


Figure 3.3. Establishment of a site-moiety map for a target binding site using shikimate dehydrogenase (SDH) of *Helicobacter pylori* (*H. pylori*) as the example. (A) Molecular docking for the screening target. (B) Merged protein-compound interaction profiles including electrostatic, hydrogen-bonding, and van der Waals profiles. A cell is colored by green if there is interaction between a compound moiety and a residue; conversely, the region is colored by black. (C) An anchor of the site-moiety map shown as the example. An anchor included conserved interacting residues, moiety preferences, and interaction type. (D) Site-moiety map of *H. pylori* SDH. The map consisted of four hydrogen-bonding anchors (H1-H4), and five van der Waals anchors (V1-V5).

3.2.4 Identification of core anchors, pharmapathlogs, and multitarget inhibitors

Core anchors, which are conserved anchors among the proteins, present key features including consensus interacting moieties of compounds, binding pockets of a binding site, and conserved interactions between the moieties and the pockets during the process of chemical reactions in a pathway. However, identifying the core anchors is a challenging task because of the low sequence identity (8.3%) and structure similarity (RMSD: 4.8Å) between HpSDH and HpSK, revealing that sequence or structure alignment methods are inapplicable to align the site-moiety maps of the two proteins for extracting the core anchors. Therefore, we aligned the site-moiety maps of the two proteins based on the spatial arrangements and the interaction types of the anchors (Figs. 3.4A and 3.4B). The alignment was achieved by maximizing the similarity score (S) between the site-moiety maps. The similarity score is defined as

$$S = \max(\sum_{i=1}^n AS(i))$$

, where $AS(i)$ is the alignment score of the aligned anchor pair i , and n is the number of the aligned anchors. Here, the alignment score is set to 1 if the distance between the centers of the aligned anchors is less than 2Å, and the aligned anchors share the same interaction types. For those with different interaction types, the alignment score is set to 0.5 if their center distances $\leq 2\text{Å}$. The alignment score is also set to 1 if an E anchor is aligned to an H anchor because the moieties of the E anchor are able to form hydrogen-bonding interactions with residues as the same as those of the H anchor. For the other aligned anchors, their alignment scores are set to 0. By exhaustively superimposing the anchors, the alignment of the three site-moiety maps with the highest similarity score was obtained. The aligned anchors with ≥ 0.5 alignment score were regarded as the core anchors, and these proteins sharing comparable core anchors were

considered pharmapathlogs (Fig. 3.4C).

Compounds that simultaneously matched the core anchors of HpSDH, HpSK, and MtSK were considered potential multitarget inhibitors for the shikimate pathway of *H. pylori* and *M. tuberculosis*. For compound *j*, the pharmapathlog score (*PS*), which is a measure of the capability of the multitarget inhibition, was calculated by

$$PS(j) = \sum_{t=1}^T \sum_{c=1}^C CAS_{ct}(j)$$

, where $CAS_{ct}(j)$ is the core anchor score of compound *j* in the core anchor *c* of the target *t*, *C* is the number of the core anchors, and *T* is the number of the targets. $CAS_{ct}(j)$ is set to 1 if compound *j* matches the core anchor *c* of the target *t*; conversely, it is set to 0. Here, *T* is 3, and *C* is 4 for all of the targets. Finally, the screening compounds were ranked based on their pharmapathlog scores, and top-ranked compounds that are commercially available were selected for bioassay. In addition, for each target, the top-ranked compounds derived from the anchor scores of the respective site-moiety maps were selected for bioassay¹⁶.

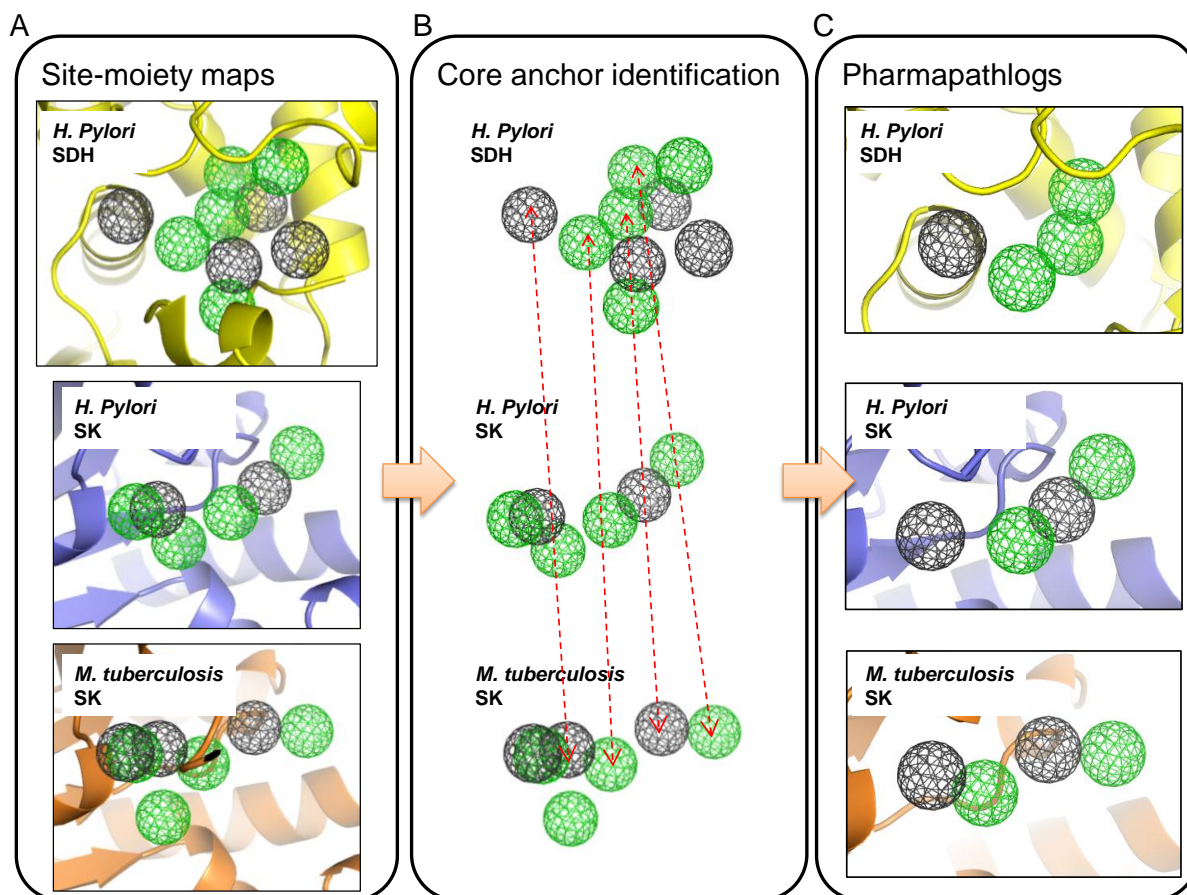


Figure 3.4. Identification of core anchors through alignment of site-moiety maps. (A) Site-moiety maps of the three screening targets. (B) Alignment process for identifying core anchors. (C) Pharma-pathlogs with core anchors.

3.3 Results and Discussion

3.3.1 Site-moiety map of shikimate dehydrogenase

The site-moiety map of HpSDH contained five H anchors (H1, H2, H3, H4, and H5) and four V anchors (V1, V2, V3, and V4) (Fig. 3.5). For each anchor, several residues comprising a binding pocket with specific chemical-physical properties, moiety compositions, and the

interaction type were identified from the top-ranked 1,000 screening compounds. The H1 anchor, consisting of residues T65, K69, and D105, preferred polar moieties (*e.g.*, carbonyl, amide, and nitro groups). The hydroxyl moieties of shikimate, one of which is involved the dehydrogenase reaction (Fig. 3.2B), consistently form hydrogen bonds with the three residues of the H1 anchor (Fig. 3.6A), indicating that the anchor is essential for catalysis and substrate binding. Furthermore, a previous study showed that the two residues of the anchor, LYS and ASP, are highly conserved and are responsible for transferring hydride ion between NADP(H) and shikimate in SDH of *Thermus thermophilus*⁸³. Three residues, H15, T65, and Y210, constituted the H2 anchor, and 258 compounds yielded hydrogen bonds with the residues by amide, carbonyl, sulfuric acid diester, amide, and carboxylic acid, which is the moiety of shikimate used to stable the binding. The H4 (S129, A179, and T180) and H5 (K69 and S129) anchors composed two polar binding pockets and may play important roles for nicotinic adenine dinucleotide phosphate (NADPH) binding (Fig. 3.6). The H3 anchor, which was far from the shikimate and NADPH binding sites, contained three residues T180, D207, and L208, presenting an additional binding pocket for designing inhibitors.

Ring moieties were the major moiety types of the V1, V2, V3, and V4 anchors. For example, 536, 233, 168, and 252 compounds yielded van der Waals (vdW) contact with the residues of the V1, V2, V3, and V4 anchors, respectively, by aromatic moieties. The number of the interacting compounds (800 of 1,000 compounds) of the V1 anchor was higher than those of the other V anchors because this anchor was composed by three hydrophobic residues, L66, L208, and A209 (Fig. 3.5). The ribose of NADPH is located in the V1 anchor when the substrate binds SDH (Fig. 3.6), suggesting the importance of the V1 anchor for maintaining the function of the protein. Although the V2 anchor only included one hydrophobic residue (L208), the long side chains of the other residues (Y210 and Q237) provides additional vdW contacts by the bulky moieties, such as aromatic and heterocyclic moieties, and this anchor had 496

interacting compounds. Furthermore, this anchor occupied as the same space as the pyridine ring of NADPH (Fig. 3.6). The V3 and V4 anchors possessed relatively small numbers of the interacting compounds because of polar or short side chains of the anchor residues. The V4 anchor was situated in the groove of the NADPH binding site, whereas the V3 anchor was positioned close to the entrance of the binding site. These V anchors can be combined with the H anchors to design inhibitors that block the binding of the shikimate or NADPH.



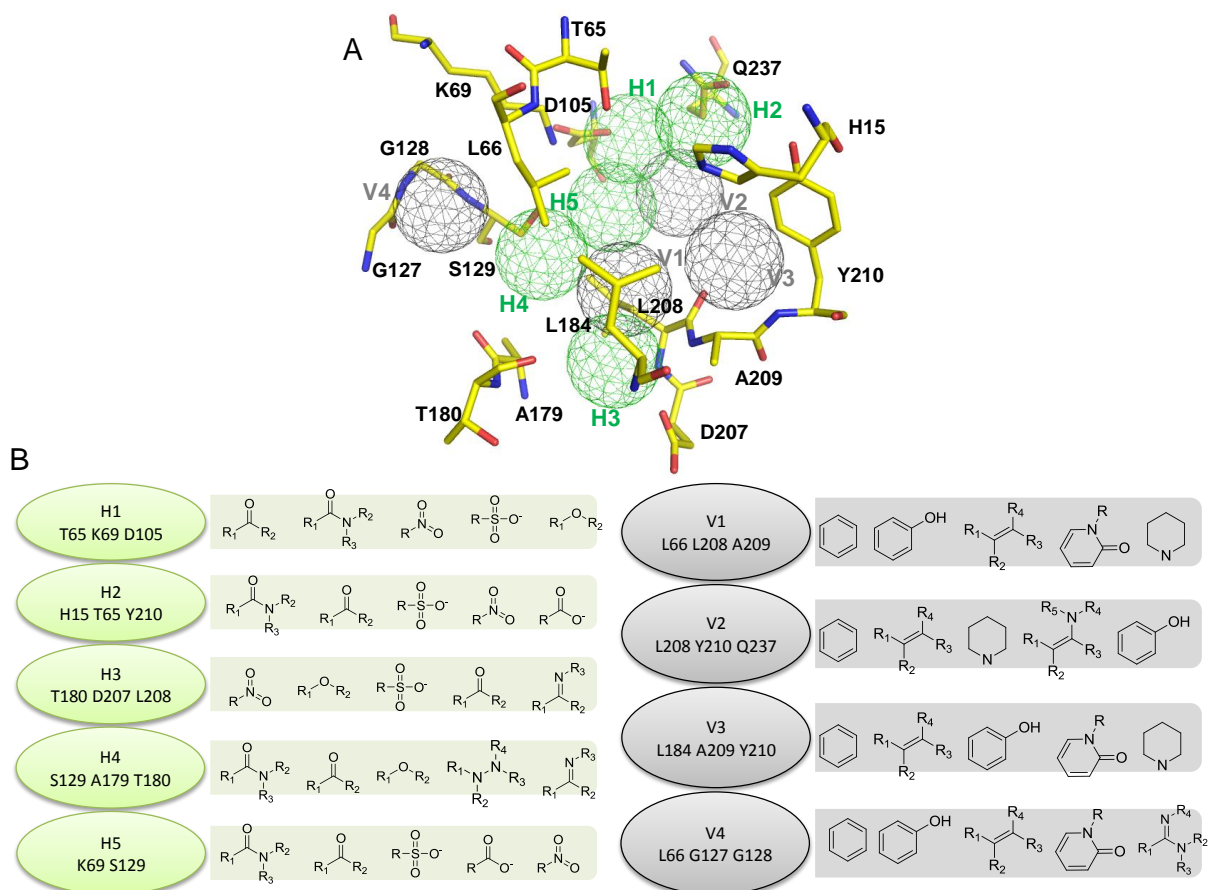


Figure 3.5. Site-moiety map of shikimate dehydrogenase. (A) Anchors with conserved interacting residues. Hydrogen-bonding and van der Waals anchors are colored green and grey, respectively. (B) Moiety preferences of anchors.

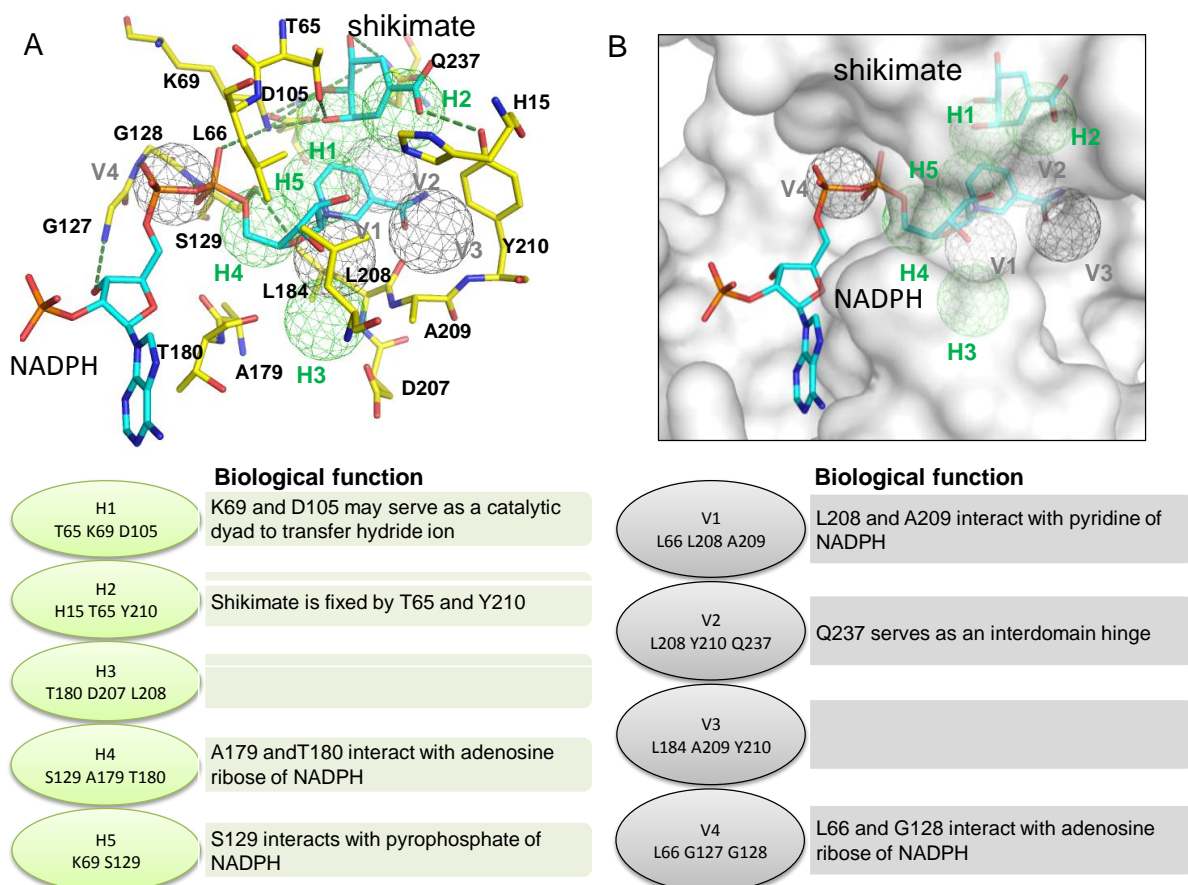


Figure 3.6. Importance of anchors in biological functions. (A) Binding modes and hydrogen bonds of the substrates, including NADPH and shikimate. (B) Binding modes of the substrates. The protein is present by surface. (C) Roles of anchors in biological functions.

3.3.2 Site-moiety maps of shikimate kinases

Shikimate kinases of HpSK and MtSK contained six (E1, H1–H3, V1, and V2) and seven anchors (E1, H1–H3, and V1–V3), respectively (Fig. 3.7). For each anchor, the relationships between residues of pockets and moieties of the docked compounds were analyzed, and moiety preferences of the anchors were then identified (Fig. 3.7B and 3.7D). The H1, H2, V1, and V3 anchors sat at the ATP site, while the H3, V2, and E1 anchors were situated at the shikimate site (Fig. 3.8A and 3.8B). Of the six consensus anchors (Fig. 3.7), the E1 anchor was a negatively

charged pocket containing R57 (R58 in MtSK), and R132 in HpSK (R136 in MtSK); these arginines are highly conserved in shikimate kinases and are critical for binding to shikimate⁷⁵. The moiety compositions of the E1 anchor consisted of carboxyl, sulfonate, and phosphate groups. The H1 anchor was enclosed with a tight turn (Walker A motif) that binds the β -phosphate of ATP⁷⁵ (Fig. 3.8). The preferred moieties of this anchor were carboxylic amide, sulfonate ester, carboxyl acid, and ketone. The H2 anchor was situated between the H1 anchor and the H3 anchor and possesses a hydrogen bonding environment from Walker A motif (K14 and S15 in HpSK; K15 and S16 in MtSK) and a DT/SD motif (D31 and D33 in HpSK; D32 and D34 in MtSK). Amide, ketone, sulfonate ester, and azine-contained compounds were the main moieties in this anchor. The H3 was situated above the central sheet including two conserved residues (D33, and G80 in HpSK; D34, G80 in MtSK), and preferred amide, sulfonate ester, and ester groups.

The V1 anchor was located at the ATP site, whereas the V2 anchor was in the shikimate site. V1 formed a vdW-binding environment for the phosphate groups and also contained residues from Walker A motif. The V2 anchor was situated at the border between shikimate and the nucleotide binding regions. In addition, The V2 anchor interacted with the ring of shikimate via vdW interactions (Fig. 3.8). The V1 anchor and V2 anchor, allowing the interactions with large chemical groups, preferred aromatic groups.

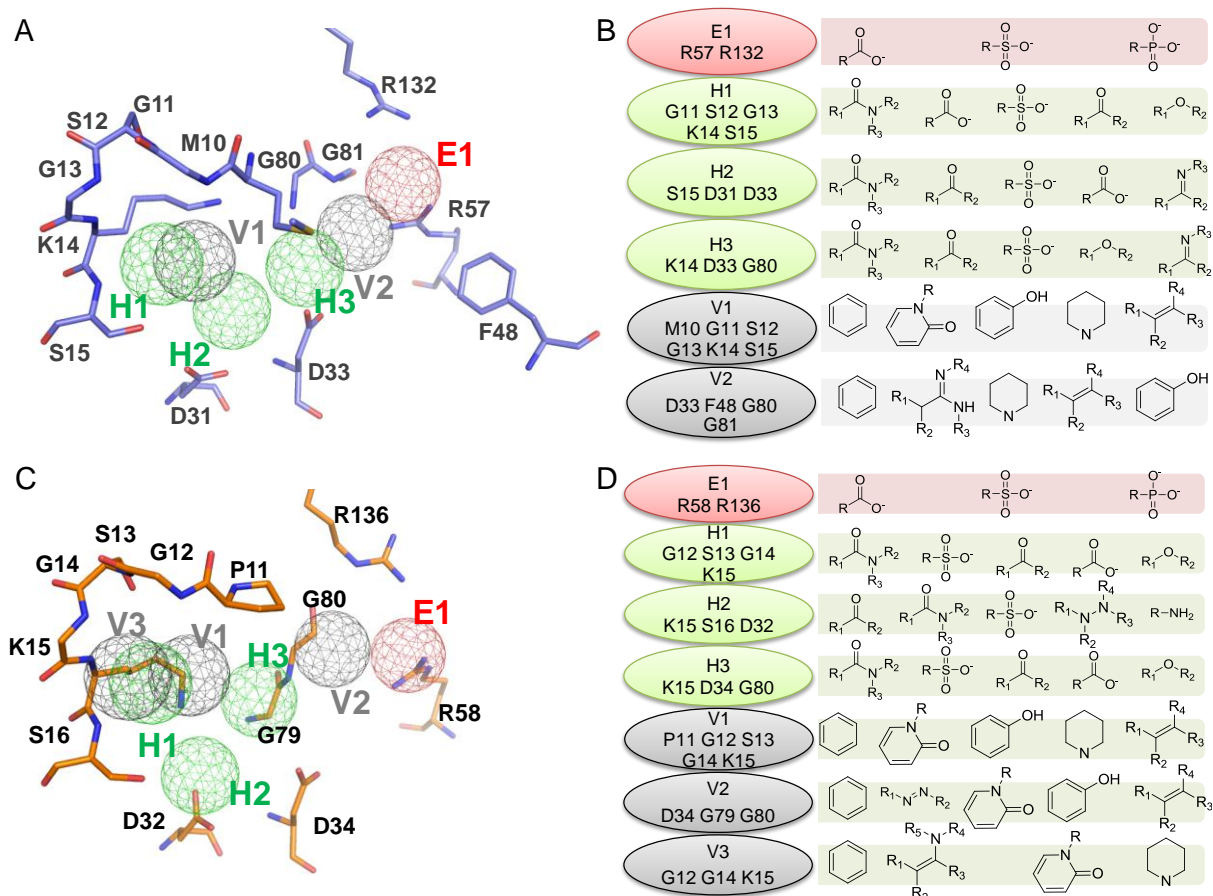


Figure 3.7. Site-moiety maps of shikimate kinases of *Helicobacter pylori* and *Mycobacterium tuberculosis*. Anchors with conserved interacting residues for *Helicobacter pylori* (A) and *Mycobacterium tuberculosis* (C). Negatively charged, hydrogen-bonding, and van der Waals anchors are colored red, green, and grey, respectively. Moiety preferences of anchors for *Helicobacter pylori* (B) and *Mycobacterium tuberculosis* (D).

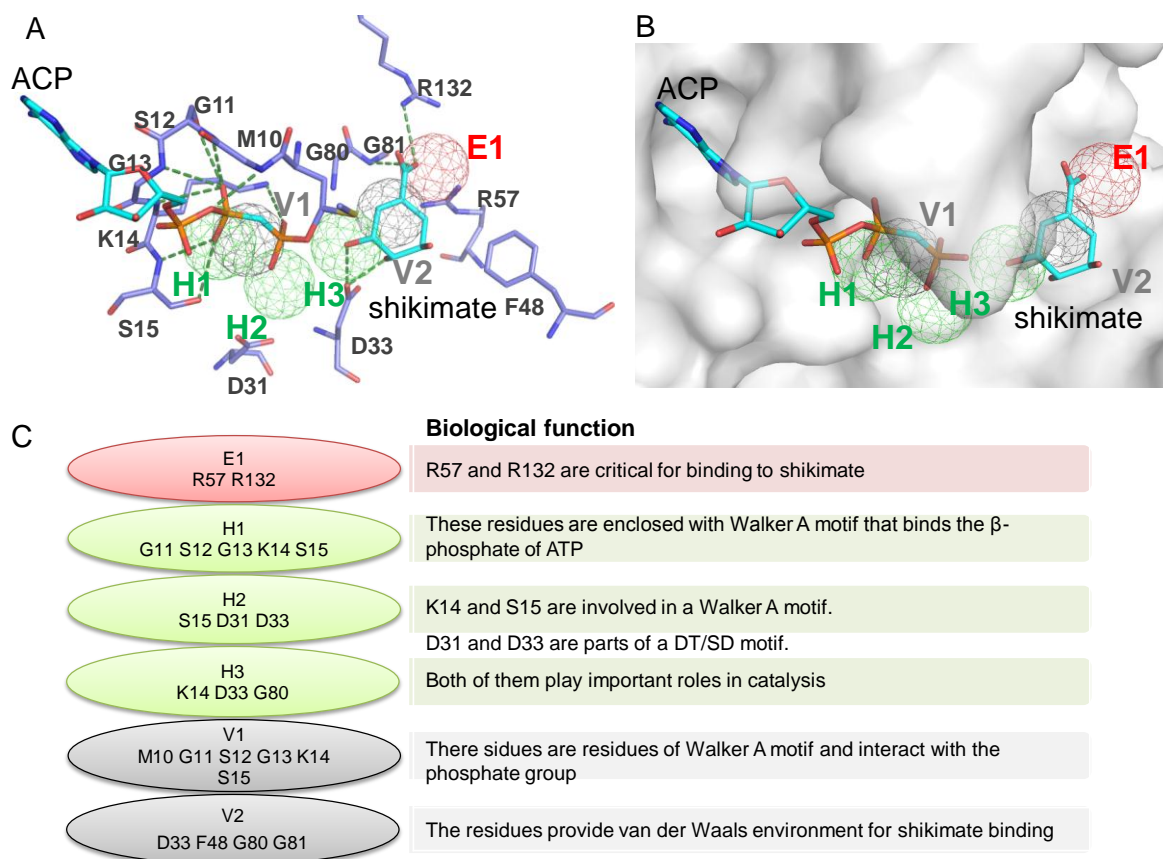


Figure 3.8. Importance of anchors in biological functions using shikimate kinase of *Helicobacter pylori* as the example. (A) Binding modes and hydrogen bonds of the phosphomethylphosphonic acid adenylate ester (ACP), an ATP analogue, and shikimate. ACP and shikimate obtained from shikimate kinase of *Mycobacterium tuberculosis* (PDB code 1zyu) are superimposed into the shikimate kinase of *Helicobacter pylori*. (B) Binding modes of the compounds. The protein is present by surface. (C) Roles of anchors in biological functions.

3.3.3 Core anchors of pharmapathlogs

HpSDH, HpSK, and MtSK shared four core anchors (CH1, CH2, CV1, and CV2) derived from the alignment of site-moiety maps despite of the low sequence and structure similarity (Fig. 3.9). Core hydrogen-bonding anchor 1 (CH1) contained H1 of HpSDH, E1 of HpSK, and E1 of

MtSk. The interaction type of the CH1 was assigned by hydrogen-bonding type because the preferred moieties of E1 were able to yield hydrogen bonds as the same as those of H1. Core hydrogen-bonding anchor 2 (CH2) consisted of H4 of HpSDH, H3 of HpSK, and H3 of MtSK. V4 of HpSDH, V1 of HpSK, and V1 of MtSK were consensus, and constituted core vdW anchor 1 (CV1). In addition, H5 of HpSDH, V2 of HpSK, and V2 of MtSK were spatially closed, and they were merged as the core vdW anchor (CV2) because two of three anchors were vdW types. The alignment of the site-moiety maps revealed that HpSDH, HpSK, and MtSK shared the four comparable core anchors (Fig. 3.4). As a result, the three targets were considered pharmapathlogs, and may be inhibited by the same compounds.

The CH1 anchor consisted of T65, K69, and D105 for HpSDH, R57 and R132 for HpSK, and R58 and R136 for MtSK (Fig. 3.9). The CH1 anchor preferred polar moieties (*e.g.*, carbonyl and sulfonate groups), and are involved in the dehydrogenase reaction for SDH and are critical for binding to shikimate⁷⁵ (Figs. 3.6 and 3.8). Interestingly, shikimates of SDH and SK are both occupied at the location of the CH1 anchor, indicating that the CH1 anchor is essential for catalysis and substrate binding in the shikimate pathway. For HpSDH, the residues (S129, A179, and T180) of the CH2 anchor interacted with nicotine adenine dinucleotide phosphate (NADPH); similarly, the residues (K14, D33, and G80 in HpSK; K14, D34, and G80 in MtSK) of CH2s are involved in Walker A motif and DT/SD motif, both of which are essential for shikimate and ATP binding in SK⁷⁴, suggesting that the CH2 anchor may play important roles for both substrate binding in SDH and SK.

In the meantime, the residues of the CV1 anchor (L66, G127, and G128 in HpSDH; M10, G11, S12, G13, K14, and S15 in HpSK; P11, G12, S13, G14, and K15 in MtSK) constituted hydrophobic binding pockets, and favored aromatic moieties, such as aromatic and heterocyclic moieties. The CV1 anchor accommodates phosphate groups of ATP and NADPH for all HpSDH, HpSK and MtSK (Figs. 3.7, 3.8, and 3.9), revealing the importance of this core

anchor for transferring phosphate groups during catalytic reactions. The final core anchor, CV2, shared the similar positions in HpSDH, HpSK, and MtSK. The long side chains of the residues (K69 and S129 for HpSDH; D33 and F48 for HpSK; D34 for MtSK) provided additional van der Waals contacts with the bulky moieties, such as pyridine ring of NADPH for SDH, and the cyclohexene group of shikimate for both HpSK and MtSK (Figs. 3.7, 3.8, and 3.9).

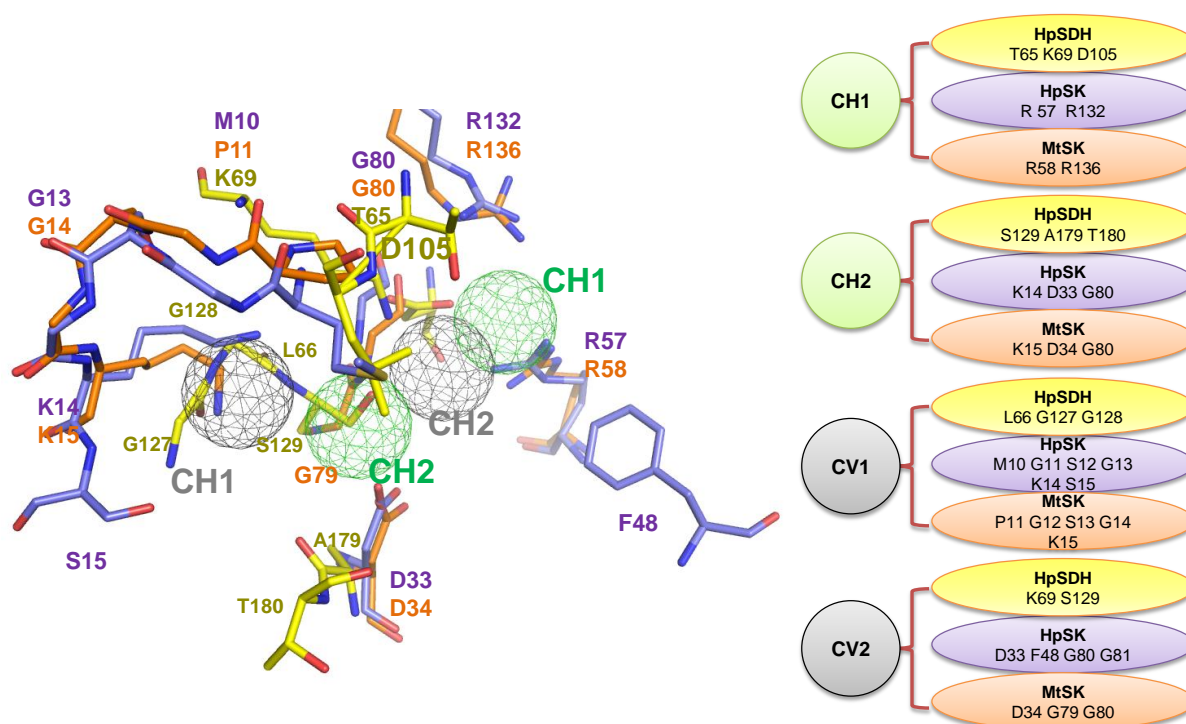


Figure 3.9. Core anchors of HpSDH, HpSK, and MtSK. HpSDH, HpSK, and MtSK are presented as yellow, slate, and orange sticks, respectively.

We further examined the core anchors through residue conservations. The residues of HpSDH and HpSK were divided into four groups: (1) core anchor residues; (2) anchor residues; (3) binding site residues; (4) other residues. The division process was described as the following (Fig. 3.10): (1) the residues of HpSDH and HpSK belonging to the core anchors

were selected as the "core anchor residues"; (2) the "anchor residues" were selected from the remaining residues; (3) the remaining residues were classified as "binding site residues" if they were the residues of the defined binding sites; (4) the remaining residues were classified as the "other residues". Each residue was only categorized once, and was assigned an evolutionary conservation score derived from the ConSurf server⁸⁴. Residues with high values (*e.g.*, 9) mean highly conserved positions. The statistics results revealed that the core anchor residues were higher conserved than anchor residues, binding site residues, and other residues (Fig. 3.11). For those residues with 9 conserved scores, they contained 81% core anchor residues, 63% anchor residues, 30% binding site residues, and 5% other residues. The core anchor residues were clearly more conserved than the rest residues of the proteins. For example, in the CH1 anchor, the conserved scores of the core anchor residues (T69, K69, and D105 for HpSDH; R57 and R132 for HpSK) were all 9. The structures (SDH-shikimate complex was provided by Dr. Wen-Ching Wang and Dr. Wen-Chi Cheng) in complex with shikimate reveals these residues play important roles for shikimate bindings⁸⁵. The results highlighted the strong selection pressures on the core anchor residues compared with the rest residues in the proteins, revealing the importance of the core anchors for maintaining biological functions during evolution.

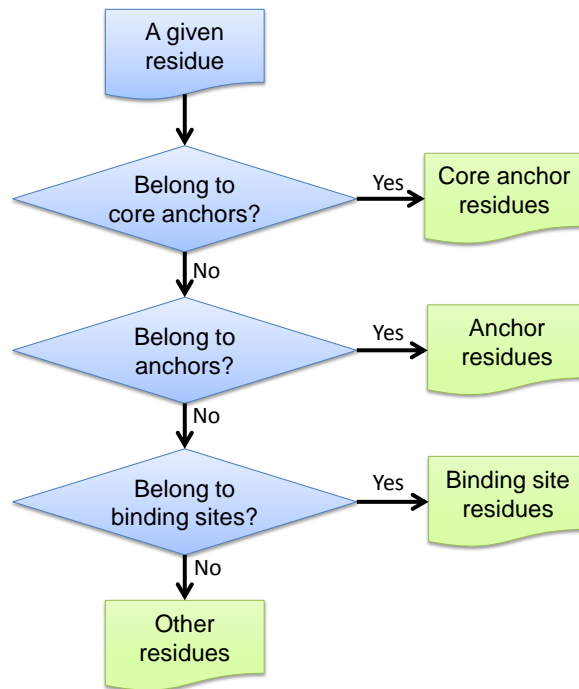


Figure 3.10. The division process for deciding core anchor residues, anchor residues, binding site residues, and other residues.

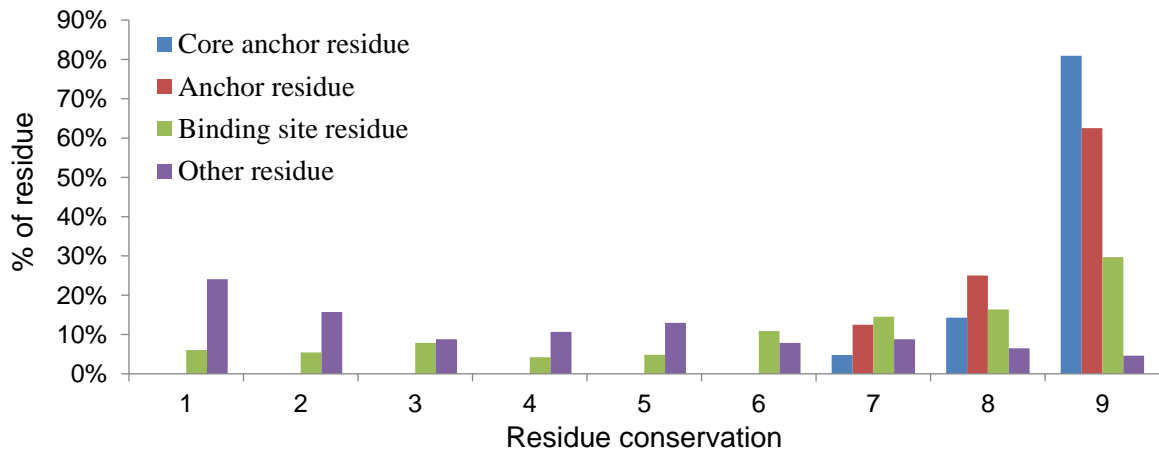


Figure 3.11. Conserved score distribution of core anchor residues, anchor residues, binding site residues, and other residues. The conserved scores are from 1 (least conserved) to 9 (most conserved).

3.3.4 New multitarget inhibitors

Following the pharmapathlog analysis, compounds were rescored using the rank-based consensus scoring (RCS⁵⁸). Our previous demonstrated that combining multiple scoring functions improves enrichment of true positives only if (a) each of the individual scoring functions has relatively high performance, and (b) the individual scoring functions are distinctive⁵⁸. As a result, we combined energy-based and pharmapathlog scores as the RCS. The top-ranked compounds that simultaneously agreed with the core anchors of HpSDH, HpSK, and MtSK, and interacted the three targets with low binding energies were thereby considered potential multitarget inhibitors. Then, we selected 30 compounds for subsequent bioassays.

Three multitarget inhibitors that simultaneously inhibited HpSDH and HpSK were identified. The three inhibitors were also able to inhibit MtSK. Two of them (NSC45174 and NSC45611) agreed with the four the core anchors in the three targets (Fig. 3.12), and their IC₅₀ values were consistently <10 μM. The other one (RH00037) did not contain a polar moiety nearby the CH1 anchor, resulting higher IC₅₀ values (24.8 μM in HpSDH; 23.8 μM in HpSK; <100 μM in MtSK) than those of NSC45174 and NSC45611 (Fig. 3.12). The sulfonate group of NSC45174 and the carboxyl group of NSC45611 formed hydrogen bonds with the residues of the CH1 anchor as like as the hydroxyl group of shikimate in HpSDH and the carboxyl groups of shikimates in HpSK and MtSK (Figs. 3.6, 3.8, and 3.12). The elimination of polar moieties in RH00037 caused about 10-fold reduction in inhibitory ability, revealing the importance of the CH1 anchor for design multitarget inhibitors.

Although the urea moiety of NSC45174 was different from the azo moieties of NSC45611 and RH00037, these moieties formed consistently hydrogen-bonding interactions with the pocket of the CH2 anchor. NSC45174, NSC45611, and RH00037 used the naphthalene, the

aromatic, and the aromatic moieties to interact the residues of the CV1 anchor via vdW interactions, respectively; similarly, NSC45174, NSC45611, and RH00037 use the aromatic, the aromatic, and the 9H-xanthene made vdW contact with the residues of the CV2 anchor. These ring moieties were able to consistently yield vdW interactions with residues of CV1 and CV2 despite of their different moieties. In these inhibitors, the different moieties with similar physico-chemical properties revealed the advantage of the pharmapathlog-based screening strategy in identifying multitarget inhibitors and an opportunity for further application in lead optimization.

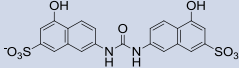
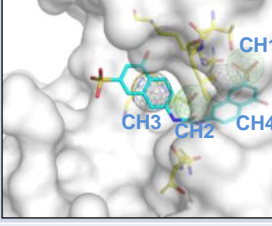
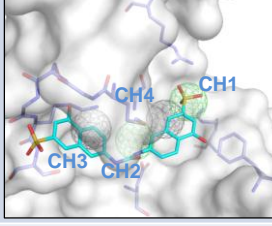
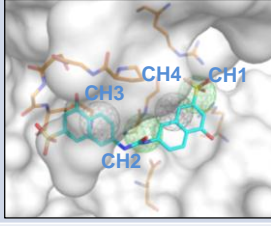
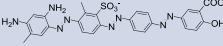
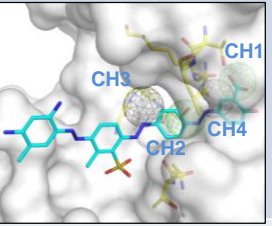
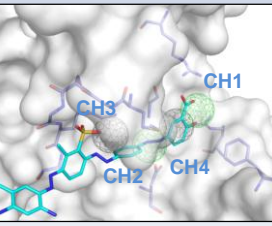
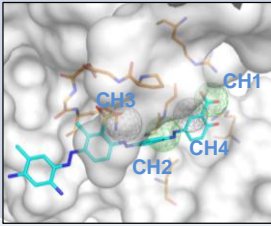
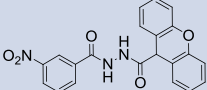
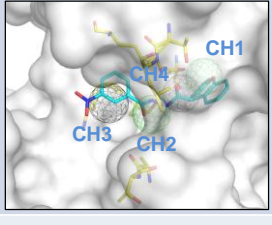
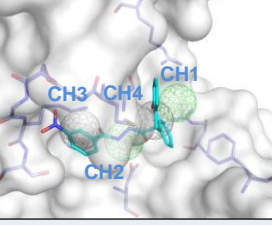
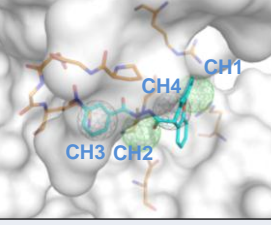
Compound		HpSDH	HpSK	MtSk
NSC45174 	Docking poses			
	IC ₅₀ (μM)	3.6	7.8	2.8
NSC45611 	Docking poses			
	IC ₅₀ (μM)	1.4	4.8	1.5
RH00037 	Docking poses			
	IC ₅₀ (μM)	24.8	23.8	<100

Figure 3.12. New multitarget inhibitors identified by the pharmapathlog-based screening strategy. In addition, the compound structures and their docking poses in HpSDH, HpSK, and MtSK are shown.

3.3.5 Specific site and inhibitors for shikimate

dehydrogenase

The alignment of the site-moiety maps of HpSDH, HpSK, and MtSK revealed a specific site for SDH despite of many similarities shared in the three targets (Fig. 3.13A). The specific site includes the H3, V1, and V3 anchors, and was not involved in the NADPH and shikimate binding sites (Fig. 3.6), suggesting an opportunity to discover specific inhibitors for HpSDH. Subsequently, we selected compounds that agreed with any of the three anchors located in the specific site. Two specific inhibitors were then identified (Fig. 3.13).

NRB03174 and HTS02873 inhibited HpSDH with IC_{50} values 9.7 μ M and 4.9 μ M, respectively, whereas the two compounds showed no inhibitory effect at 100 μ M for HpSK (Fig. 3.13D). NRB03174 used 4,5-dihydro-1H-tetrazole, chain, and bromobenzene moieties to interact with the residues of the H3, V1, and V3 anchors (Fig. 3.13B); on the other hand, HTS02873 made vdW contacts with the residues of the V1 anchor via the 4,5-dihydro-1H-tetrazole moiety (Fig. 3.13C). These results showed that the H3, V1, and V3 anchors are useful to design specific inhibitors for HpSDH.

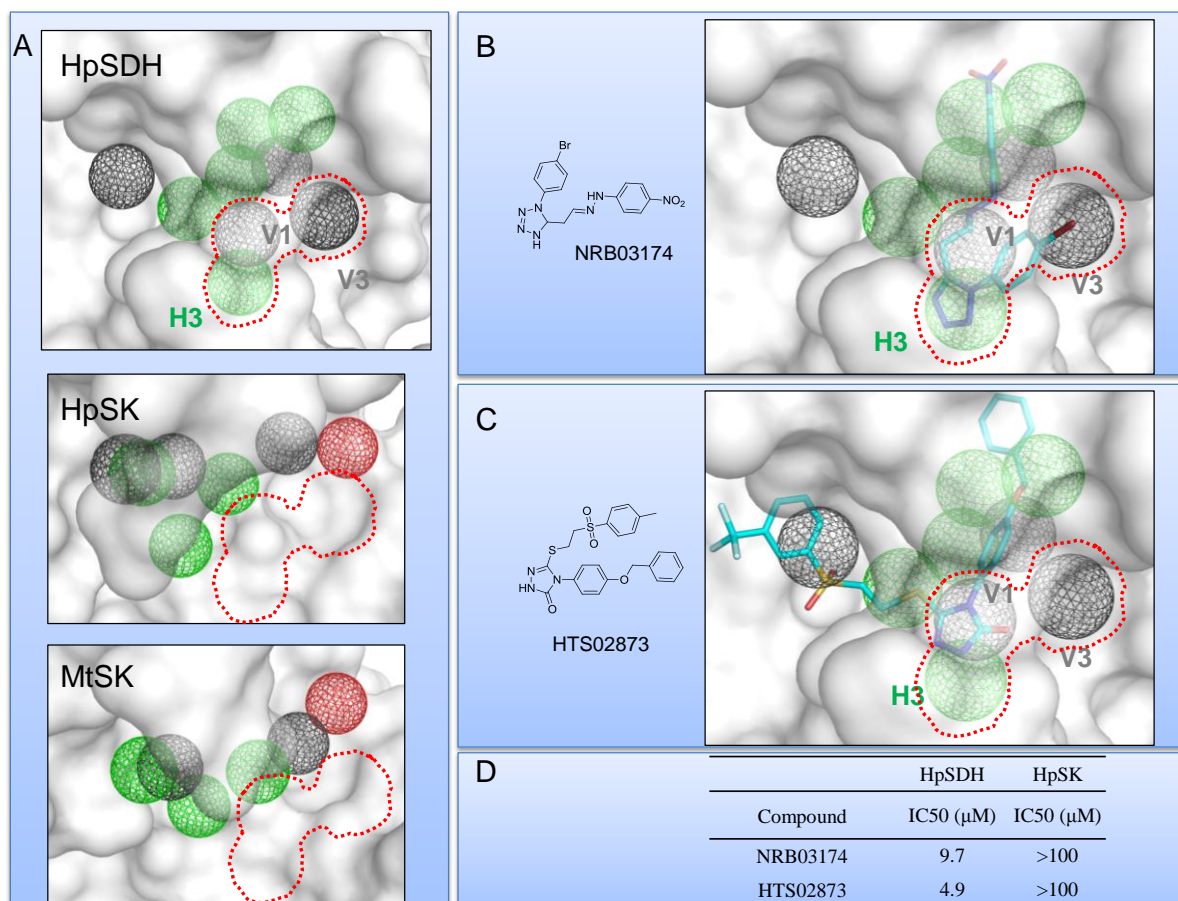


Figure 3.13. Specific site and inhibitors for shikimate dehydrogenase. (A) Specific site of shikimate dehydrogenase. The site includes the H3, V1, and V3 anchors. Docked poses of the specific inhibitors, including NRB03174 (B) and HTS02873 (C). (D) Bioassays of the compounds for HpSDH and HpSK.

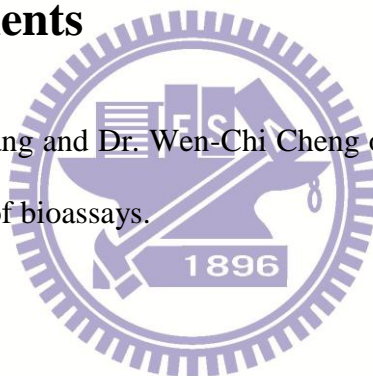
3.4 Conclusions

The study developed a new strategy, namely pharmapathlog-based screening strategy, to discover pharmapathlogs and multitarget inhibitors. We applied this strategy to identify multitarget inhibitors of shikimate pathways. By use of the strategy, the four core anchors consistently shared in HpSDH, HpSK, and MtSK were identified despite of the low sequence and structure similarity among these targets. The three proteins were thereby considered

pharmapathlogs, and three multitarget inhibitors were identified. Two of the multitarget inhibitors simultaneously inhibited HpSDH and HpSK with IC_{50} values $<10.0 \mu\text{M}$. The three inhibitors also inhibited MtSK with IC_{50} values $<10.0 \mu\text{M}$. In addition, our analyzed results showed that a core anchor often play important roles in catalysis or substrate binding, and is highly conserved. The above experimental results indicate that the pharmapathlog-based screening strategy is able to identify pharmapathlogs and then discovery multitarget inhibitors with novel scaffolds. We believe that the new strategy will be useful to discover new line drugs for the treatment of human diseases.

3.5 Acknowledgments

We thank to Dr. Wen-Ching Wang and Dr. Wen-Chi Cheng of National Tsing Hua University for kindly providing the results of bioassays.



Chapter 4. Pharmapathlogs for optimizing and identifying neuraminidase inhibitors

Influenza is an annual seasonal epidemic around the world and has continually drawing the public attentions because of the huge death toll and drug resistance. Neuraminidase, which is essential for the spread of influenza virus, has been regarded as a valid target for the treatment of influenza infection. In this chapter, we applied the pharmapathlog-based screening strategy to find and design new inhibitors for H1N1 and H5N1 neuraminidases by collaborating with Dr. John T.A. Hsu and Dr. Hui-Chen Hung of National Health Research Institutes, and Dr. Chun-Cheng Lin and Mr. Chien-Hung Lin of National Tsing Hua University. First, we applied the SiMMap server to infer site-moiety maps of H1N1, H5N1, and H3N2 neuraminidases. Subsequently, the core binding environments (core anchors) of the neuraminidases were extracted by comparing their maps. Based on the core anchors, we discovered three new inhibitors with IC_{50} values $<10 \mu\text{M}$. The experiment results revealed that the inhibitors could overcome the drug resistance introduced by H274Y and I222R for N1 neuraminidases without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains. In addition, five zanamivir derivatives were designed and their IC_{50} values were in the <10 nanomolar range. Our experimental results show that the pharmapathlog-based screening strategy is useful to identify and optimize novel inhibitors. We believe that the pharmapathlog-based screening strategy provides a great help for drug development and human health.

4.1 Introduction

Influenza virus causes severe respiratory illness and death each year. In recent years, outbreaks of avian influenza H5N1 virus have attracted the public attentions^{86,87}. In addition, a new strain of influenza H1N1 virus that originated in swine has rapidly spread to many countries⁸⁸. Although vaccination is the primary strategy to prevent influenza infection, vaccines are often inadequate because of the high mutation rate of influenza viral antigens⁸⁹. Currently, the viral enzyme neuraminidase (NA) has been regarded as a valid target for the treatment of influenza infection^{3,4}, and two successful drugs, zanamivir and oseltamivir, were reported⁹⁰. These two drugs were designed based on the transition state of sialic acid⁹¹⁻⁹³, and are generally used for the therapy of influenza virus infections⁹⁰. However, several drug-resistant mutations of NAs have been reported for oseltamivir and zanamivir during treatment^{6,94-96}. Therefore, it is apparent that developing new antiviral inhibitors to treat influenza virus infections is required.

NA cleaves the glycosidic linkage of sialic acid, and facilitates the release of newly formed virus from infected cells⁹⁷. The process is essential for the spread of virus in the respiratory tract. Therefore, NA is recognized as a suitable target protein for anti-influenza drug design. Subtypes of NAs can be classified into two group-1 (N1, N4, N5 and N8) and group-2 (N2, N3, N6, N7 and N9) according to their phylogenetic distances⁹¹. Recently, the crystal structures of group-1 NAs reveal that 150-loop of the group-1 NAs is able to maintain an open form and makes 150-cavity adjacent to the sialic acid binding site but not for that of group-2 NAs^{91,98}. However, 150-loop of group-2 NAs may be induced to be an open form by inhibitors because the two groups have high sequence similarity in 150-loops, and energy difference between open-form and closed-form conformations is not large⁹¹. The 150-cavity suggests new opportunities to design new type NA inhibitors, which extend to the 150-cavity. The new type inhibitors may be useful to inhibit drug-resistant influenza viruses introduced by the treatment of zanamivir or oseltamivir because drug-resistant mutations of these viruses often occur in the

sialic acid binding site^{6,94}. Although several inhibitors designed to be located at the 150-cavity have been reported^{49,99,100}, the potency of these inhibitors is low (micromolar range).

Core binding environments (core anchors) of pharmapathlogs often play important roles in biological functions because they are highly conserved during evolution to maintain structures or functions of proteins. In the previous chapter, we demonstrated that a site-moiety map, which contains several anchors, presents key binding environments of a protein binding site¹⁶. In addition, our experimental results show that an anchor is often a hot spot and the site-moiety map can help to assemble potential leads by optimal steric, hydrogen-bonding, and electronic moieties¹⁶. By uses of site-moiety maps, we are able to infer pharmapathlogs and their core binding environments presented as core anchors. Here, we applied the pharmapathlog-based screening strategy to identify and optimize inhibitors of H1N1, H5N1, and H3N2 NAs. Based on the core anchors of the site-moiety maps, we designed five zanamivir derivatives, which extend to the 150-cavity, with IC₅₀ values in the <10 nanomolar range. Moreover, the potent inhibitions of the zanamivir derivatives suggest that these inhibitors may interact with the residues of the 150-cavity of H1N1 and H3N2, revealing these derivatives were able to induce H3N2 NA to form the 150-cavity. Next, we found three novel inhibitors (IC₅₀ values <10 μM) predicted to be located in the 150-cavity. The three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 NAs without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains. The experimental results demonstrate that the pharmapathlog-based screening strategy is useful to identify and optimize novel inhibitors for drug-resistant strains. We believe that this strategy is helpful to design new drugs for diseases.

4.2 Methods and Materials

4.2.1 Overview of pharmapathlog-based screening strategy

Figure 4.1 presents the overview of the pharmapathlog-based screening strategy for optimizing lead compounds and identifying novel inhibitors. First, we performed large scale molecular docking to construct site-moiety maps (Fig. 4.2A). Approximately 420,000 compounds were docked to H1N1, H5N1, and H3N2 NAs by using GEMDOCK, which is an in-house developed docking program^{33,40,58,101} and has been successfully applied to identify inhibitors and binding sites for some targets^{51,56,57}. Based on the docking energies, top-ranked 1,000 compounds (~0.2%) were selected to the SiMMap server for constructing site-moiety maps of the NAs (Fig. 4.2A). The SiMMap server applied GEMDOCK program to generate protein-compound interaction profiles (Fig. 4.2B). Consensus interactions between compound moieties and binding pockets consisted of conserved interacting residues were identified based on these profiles. Moiety preferences of the binding pockets were also identified from the 1,000 compounds (Fig. 4.2C). The binding pockets, the moiety preferences of the pockets, and pocket-moiety interaction type [electrostatic (E), hydrogen-bonding (H), or van der Waals (V) interactions] comprised anchors of the site-moiety maps (Figs. 4.1B and 4.2D). The NAs sharing similar site-moiety maps were thereby considered pharmapathlogs, and their consensus anchors were treated as the core binding environments (core anchors).

The core anchors of the site-moiety maps can be used for optimization and identification of inhibitors (Figs. 4.1C and 4.1D). For lead optimization, we first found unmatched core anchors of a lead compound, and then modified the compound structure by adding moieties based on moiety preferences of the unmatched anchors (Fig. 4.1C). For inhibitor identification, the docked compounds were re-ranked by using the rank-based consensus scoring (RCS). RCS

obtained the new ranks of each compound through the pharmapathlog score ranks and energy-based ranks (Fig. 4.1D). Finally, the modified compounds and the predicated potential inhibitors were verified by bioassays (Figs. 4.1E and 4.1F)

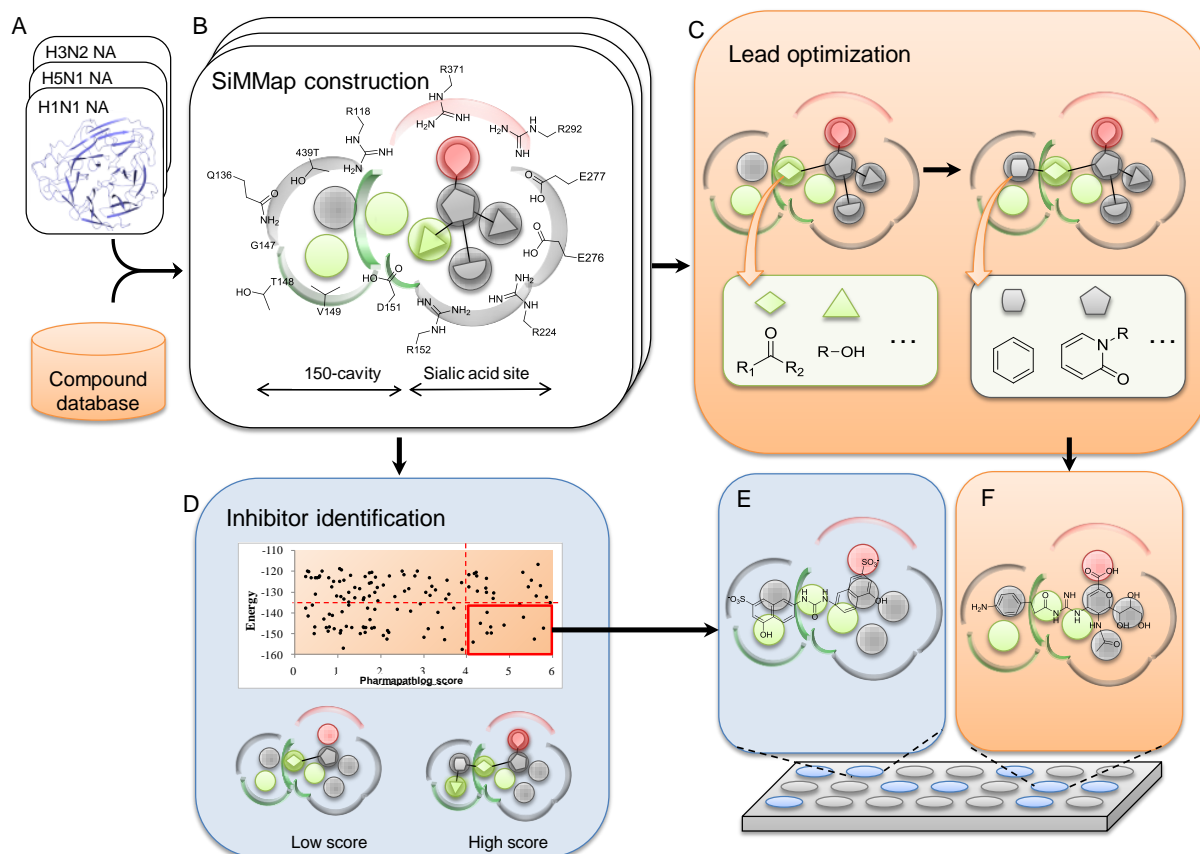


Figure 4.1. Overview of the pharmapathlog-based screening strategy for optimizing lead compounds and identifying novel inhibitors. The strategy includes five major steps, including (A) virtual screening for H1N1, H5N1, and H3N2 NAs, (B) construction of site-moiety maps for inferring core anchors of pharmapathlogs, (C) lead optimization, (D) inhibitor identification, and bioassays for identified inhibitors (E) and optimized inhibitors (F).

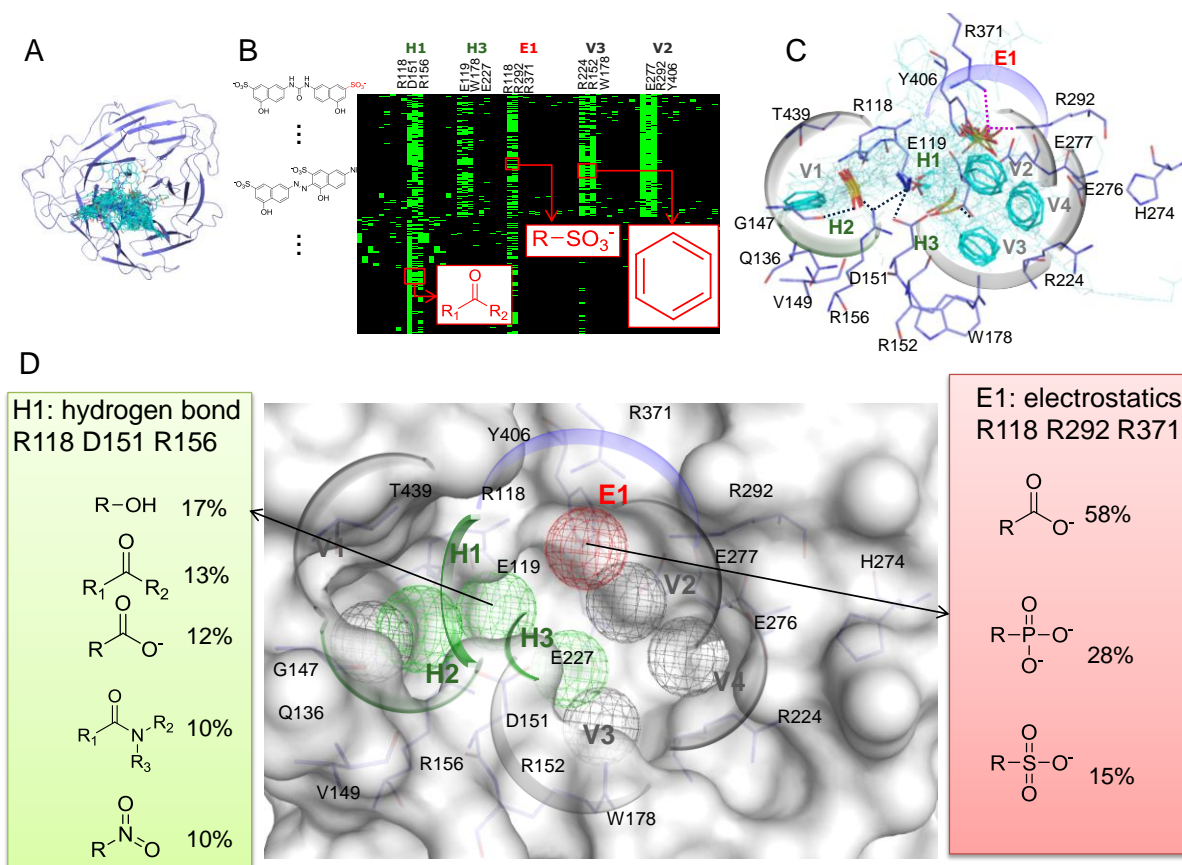


Figure 4.2. Main steps for constructing a site-moiety map. (A) Top-ranked compounds selected by GEMDOCK. (B) Protein-compound interaction profile for identifying consensus interactions between moieties of screening compounds and residues of pockets. (C) Examples of moieties that consistently interacted with residues of pockets. (D) Site-moiety map of NA using H1N1 NA as the example. The map consisted of eight anchors including conserved interacting residues, moiety preferences, and interaction types. Negatively charged, hydrogen-bonding, and van der Waals anchors are colored red, green, and grey, respectively.

4.2.2 Dataset preparation

To design zanamivir derivatives and to identify new type inhibitors that could bind the 150-cavity, we selected 150-open form structures of NAs including H1N1(PDB code 3beq⁹⁸) and

H5N1 (PDB code 2hty⁹⁰) for constructing site-moiety maps. In addition, a 150-closed form structure of H3N2 NA (PDB code 2aep¹⁰²) was selected. For the 150-open form of H1N1 NA, the binding site comprising the 150-cavity and the sialic acid binding site was defined to include the residues within a 10 Å radius sphere centered around the 150-loop (residues 147–152⁹¹) and zanamivir by superimposing a crystal structure of H1N1 NA (PDB code 3b7e⁹⁸). This binding site was then applied to derive binding sites of H5N1 and H3N2 NAs using a structural alignment tool⁷⁶.

H3N2 NA may maintain an open 150-open form because the 150-loop sequences of H1N1 and H3N2 NAs share a high sequence similarity (Fig. 4.3). Therefore, the 150-open form of H3N2 NA was modeled for further analysis. We selected a structure (PDB code 2hty⁹⁰) with the 150-open form as the template to model the 150-open form of H3N2 NA by using a homology modeling approach¹⁰³. Interestingly, the homology modeling results showed that the structure of H3N2 NA could be stable when its 150-loop was the open form (Fig. 4.4). The binding site of the modeled structure was prepared by the same steps as described above for the construction of site-moiety map.

The compounds used for virtual screening were collected from three compound libraries including Maybridge, NCI, and Sigma because of their rapid availability (Table 3.1). 3D structures of these compounds were obtained by CORINA⁸². The compounds were filtered if their compound molecular weights are < 200 or > 650 dalton. The number of the compounds selected for screening was 414,977.

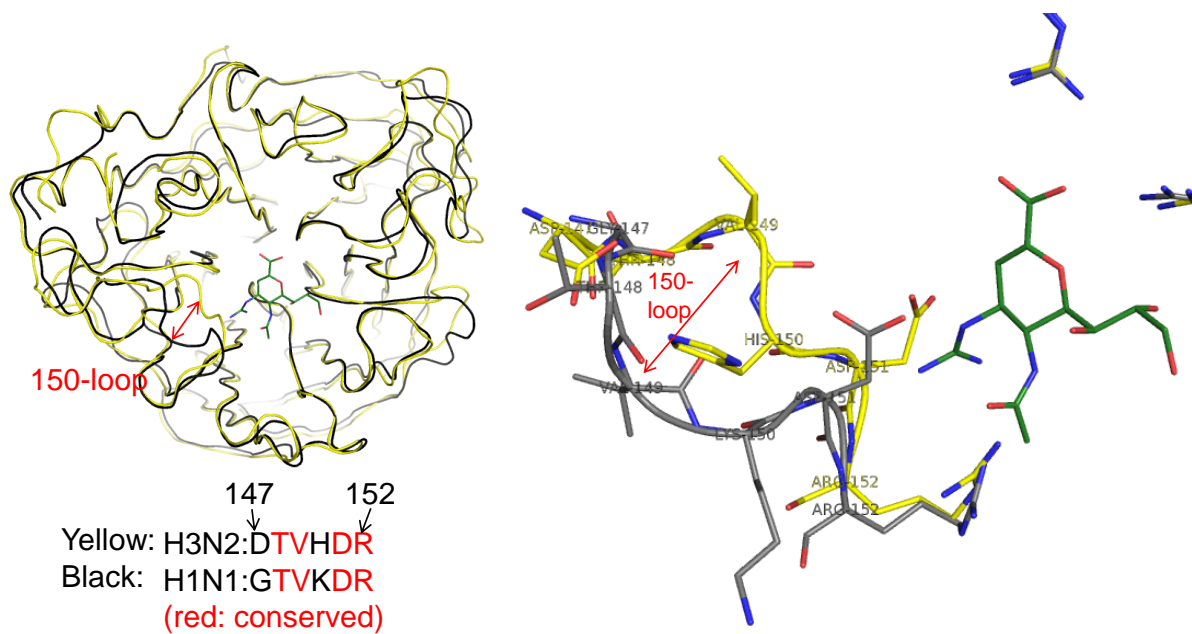
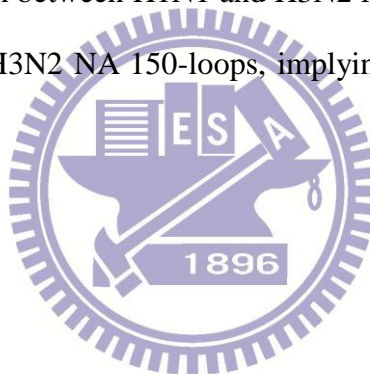


Figure 4.3. Sequence comparison between H1N1 and H3N2 NA 150-loops. Four residues were conserved between H1N1 and H3N2 NA 150-loops, implying H3N2 NA may contain a 150-open form.



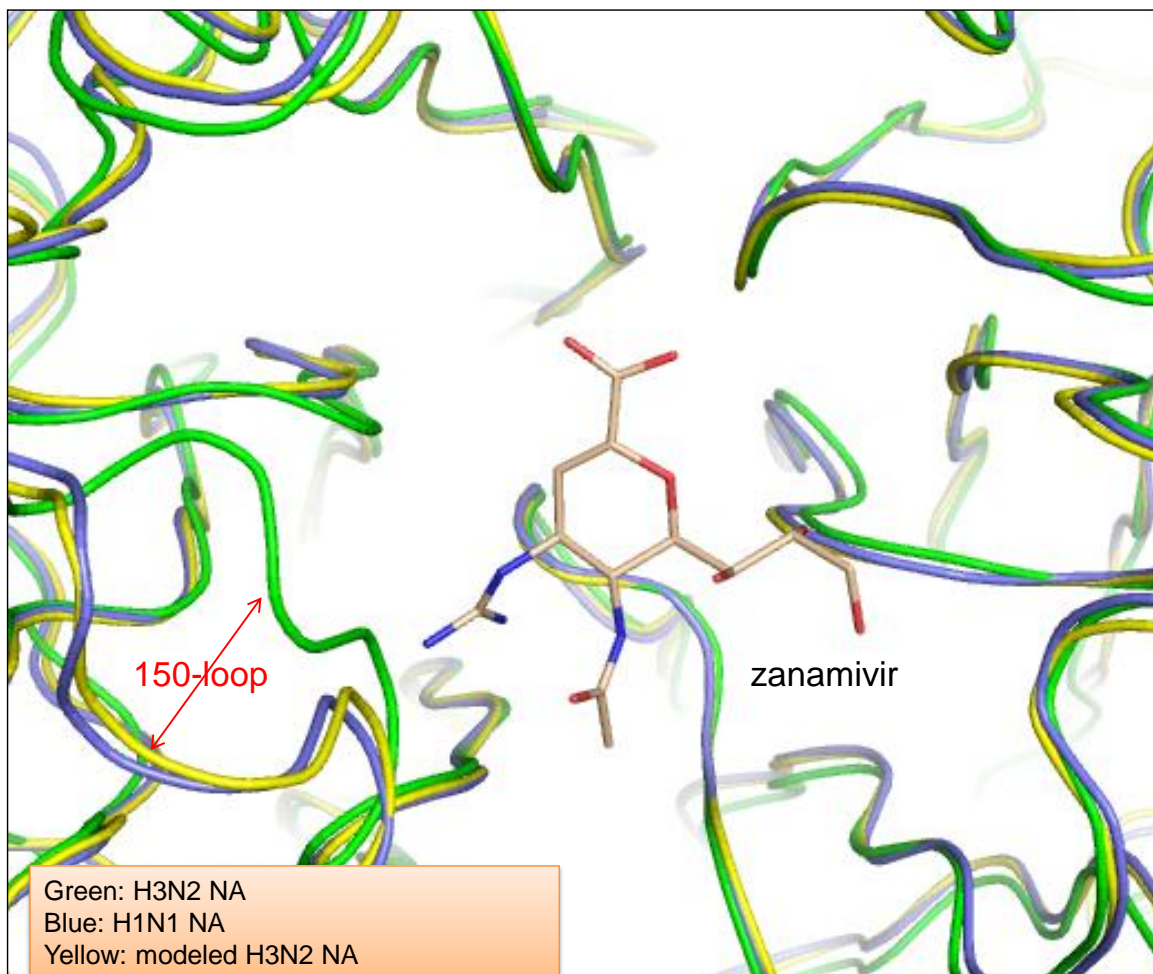


Figure 4.4. Modeled structure of H3N2 NA with a 150-open form.

4.2.3 Main procedure for identifying new type inhibitors and lead optimization

The main steps of the pharmapathlog-based screening strategy for identifying new type inhibitors and lead optimization were described as the following (Fig. 4.1):

- (1) Virtual screening of NAs. We docked the compounds of the databases into NAs of H1N1,

H5N1, and H3N2 by using in-house developed GEMDOCK program^{33,40,58,101} (Fig. 4.1A), which has been successfully applied to identify inhibitors and binding sites for some targets^{51,56,57}. Top-ranked 1,000 compounds ($\sim 0.2\%$) of each NA were selected for constructing respective site-moiety maps.

- (2) Site-moiety map construction of NAs. The top-ranked compounds and their corresponding binding sites were submitted to the SiMMap server (Fig. 4.2A). Subsequently, the server applied the GEMDOCK program to generate protein-compound interaction profiles (Fig. 4.2B). Based on these profiles, consensus interactions between compound moieties and binding pockets consisted of conserved interacting residues were identified. In addition, moiety preferences of the binding pockets were identified from the 1,000 compounds (Fig. 4.2C). The binding pockets, the moiety preferences of the pockets, and pocket-moiety interaction type [electrostatic (E), hydrogen-bonding (H), or van der Waals (V) interactions] comprised anchors of the site-moiety maps (Figs. 4.2B and 4.2D).
- (3) Identification of pharmapathlogs and core anchors. In this study, we used site-moiety map similarity to infer pharmapathlogs of NAs. Here, NAs that shared more 50% core anchors were regarded as pharmapathlogs. Two anchors between two site-moiety maps were defined as core anchors (core binding environments) if they had (1) similar positions in space (the distance of two anchor centers was less than 2.0\AA); and (2) the same interaction type. Core anchors were extracted through site-moiety map alignment that maximized the number of core anchors.
- (4) Lead optimization process. Moiety preferences of a core anchor are able to guide to assemble leads by optimal steric, hydrogen-bonding, and electronic moieties. For a lead, a near core anchor that was not matched by the lead can be identified on the site-moiety map. According to the moiety composition of the anchor, we are able to modify the lead structure by replacing or adding additional moieties (Fig. 4.1C). The moiety energy, which was used

to measure how well an added moiety sterically and physicochemically fit the anchor, was defined as

$$E_{moiety} = \sum_{i=1}^{moi} \sum_{j=1}^{pro} E_{ij}$$

,where E_{ij} is the interaction energy between atom i and j . The interaction energy was generated by GEMDOCK according to the interaction type of the anchor. The *moi* and *pro* indicate the numbers of the heavy atoms in the added moiety and the protein, respectively. Therefore, we can acquire ranks of all modified lead structures based on the moiety energy.

- (5) Identification of new type inhibitors. The core anchors of the site-moiety maps can be divided into 150-cavity group and sialic acid group based on their locations. The core anchors of the 150-cavity group were then used to identify new type inhibitors. For each compound, we obtained a pharmapathlog score, which was described in the previous chapter, by using the core anchors in the 150-cavity. Based on the scores, we can obtain ranks of all compounds in the 150-cavity.
- (6) Selection of top-ranked compounds. Combining different scoring methods is often useful to improve hit rate in identifying active compounds⁵⁸. Here, we used rank-based consensus scoring (RCS) to combine pharmapathlog-based and GEMDOCK scoring functions (Fig. 4.1D). For a compound x , we calculated its RCS by combining the ranks of pharmapathlog-based and GEMDOCK scoring functions as follows: $CS(x) = R_p(x) + R_G(x)$, where $R_p(x)$ is the rank of compound x based on its pharmapathlog score, and $R_G(x)$ is the rank of compound x using GEMDOCK scoring function. Finally, the RCS rank of the compound x was derived based on ascending order of $CS(x)$.
- (7) Bioassay. To discover new types of inhibitors, we selected top-ranked compounds occupied at the 150-cavity. In addition, we selected the modified zanamivir derivatives for bioassays

(Figs. 4.1E and 4.1F). The bioassays, including inhibition assay, cytopathic effect test, and cytotoxicity assay, were used to evaluate the utility for identification and optimization of lead compounds.



4.3 Results and Discussion

4.3.1 Site-moiety maps of H1N1, H5N1, and H3N2 NAs

The site-moiety maps of H1N1 and H5N1 NAs contained eight core anchors (E1, H1-H3, and V1-V4), whereas H3N2 NA had five core anchors (E1, H1, and V1-V3) in its site-moiety map because lacking the 150-cavity (Fig. 4.5). The E1, H1, V1, V2, and V3 were located in the sialic acid binding site and consistent in the site-moiety maps of the three NAs, implying the sites of the different NA subtypes share conserved physical-chemical features for the sialic acid binding. However, the site-moiety map of the modeled H3N2 NA was similar to those of H1N1 and H5N1, suggesting that the three NAs are pharmapathlogs and could be inhibited by new type inhibitors occupied in the 150-cavity (Fig. 4.5C).

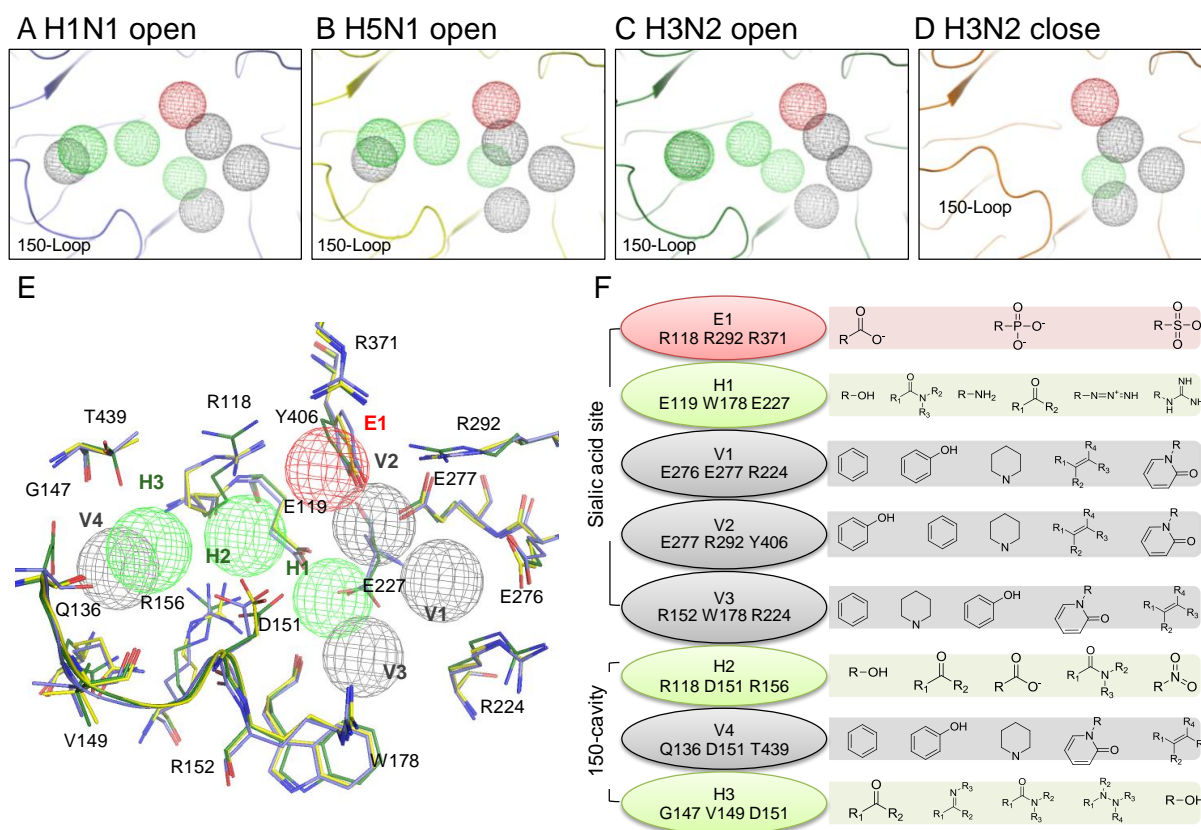


Figure 4.5. Site-moiety maps of NA pharmapathlogs (A-E). (F) The moiety preferences of the core anchors. Red, green, and gray anchors presents the electrostatic, hydrogen-bonding, and van der Waals interaction types, respectively. H1N1, H5N1, and H3N2 NAs are colored by slate, yellow, and green sticks.

The E1 core anchor was a positively charged pocket with three residues (R118, R292, and R371) which often formed electrostatic or hydrogen-bonding interactions with negatively charged or polar moieties among the docked compounds. These residues are highly conserved in all NAs and interact with the carboxylate moiety of the substrate sialic acid⁹¹. Except for the carboxylate moiety, sulfuric acid monoester and phosphoric acid monoester, which were physicochemically similar to the carboxylate moiety, were also identified from the docked compounds (Fig. 4.5F). The H1 core anchor possessed a polar binding pocket with residues, E119, W178, and E227, and the pockets preferred polar moieties (*e.g.*, hydroxyl group, carboxylic amide, ketone, and amine) (Fig. 4.5F). In addition, sialic acid, zanamivir, and oseltamivir consistently form hydrogen bonds with the pocket of the H1 core anchor^{90,98,104}. A study also indicates the E119G mutation reduces zanamivir susceptibility (1,400-fold)¹⁰⁵, suggesting this core anchor could play an important role for designing NA inhibitors in the sialic acid binding site.

The hydrophobic moieties of the docked compounds often formed van der Waals contacts with the long side chains of R224, E276, and E277 in the V1 core anchor. The van der Waals interactions between the hydrophobic pocket and oseltamivir are required for the binding process of oseltamivir¹⁰⁶. However, the H274Y mutation, which is the principal and frequent for the oseltamivir treatment^{49,106}, could alter the hydrophobic pocket. As a result, the mutation reduces the binding affinity of oseltamivir. Currently, a dual mutation (H274Y and I222R) causes ~20, ~12,000, and ~7,500 fold reduction in NA inhibition for zanamivir, oseltamivir, and peramivir, respectively⁹⁶. This suggests that the NA inhibitors located in the V1 core anchor may lose their potency when the H274 or I222R mutants. Furthermore, three residues (E277, R292, and Y406) were involved in the V2 core anchor (Fig. 4.5F). Y406 is supposed to be the catalytic residue for cleavage of the substrate¹⁰⁷, and the study of Ghate and Air also showed that the mutation of Y406 has a drastic effect on NA activity¹⁰⁸. Subsequently, the V3

core anchor, preferring hydrophobic moieties (*e.g.*, aromatic ring, heterocyclic group, alkenes, phenol, and oxohetarene) contained R152, W178, and R224 (Fig. 4.5F). The crystal structures (PDB codes 3b7e⁹⁸, 2hu4⁹⁰, and 1mwe¹⁰⁴) reveals that the acetamido moieties of sialic acid, zanamivir, and oseltamivir are located in this pocket, and interact with W178. Based on the above analysis and results, it may be concluded that the core anchors of NA pharmapathlogs often play important roles in catalysis or substrate binding, and the moiety preferences could be useful for drug design.



Compound ID	Subtype	IC50 (nM)	Compound Structure	Moiety composition				
				H3	E1	V2	V3	V4
GS4071	H1N1	1						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
Compound 3a	H1N1	0.3						
	H5N1	13.3		R-NH ₂	R-C(=O)O ⁻			
	N2							
Carbocyclic Analogue 53	H1N1	0.3						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
Carbocyclic Analogue 50	H1N1	1						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
C3-Aza Carbocyclic Analogue 3h	H1N1	12						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
Carbocyclic Analogue 12	H1N1	100						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
Carbocyclic Analogue 58	H1N1	100						
	H5N1				R-C(=O)O ⁻			
	N2							
Carbocyclic Analogue 31	H1N1	6300						
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2							
BANA 113	H1N1							
	H5N1				R-C(=O)O ⁻			
	N2	10000						
Benzoic Acid Inhibitor 8	H1N1							
	H5N1				R-C(=O)O ⁻			
	N2	15000						
BANA 108	H1N1							
	H5N1			R-NH ₂	R-C(=O)O ⁻			
	N2	>1.0E7						
Zanamivir	H1N1							
	H5N1				R-C(=O)O ⁻			
	N2							
ATA	H1N1	3300						
	H5N1	3300		R-C(=O)O ⁻	R-C(=O)O ⁻			
	N2	13800						

Figure 4.6. Relationship between core anchors and moieties of known inhibitors. The known inhibitors include GS4071, zanamivir, GS4071, and zanamivir analogues obtained from structure-activity relationship studies, and ATA. The cells are colored by yellow if the moieties of the inhibitors agree with the core anchors.

To validate the core anchors, we collected other known inhibitors including GS4071, zanamivir, GS4071, and zanamivir analogues obtained from structure-activity relationship studies, and ATA since these compounds contain various moieties in the core anchors (Fig. 4.6)¹⁰⁹⁻¹¹⁴. The moiety compositions of these compounds revealed that the moiety preferences of the core anchors may provide clues for lead optimization. For example, the compound 3a, which is an analogue of GS4071, has a phosphonate substitution at the E1 core anchor slightly enhance the IC₅₀ value from 1nM to 0.3nM. Another example is carbocyclic analogue 53, which contains an aromatic moiety to increase the inhibitory activity.

The open-form conformation of the 150-loop gave rise to three extra core anchors (H2, H3, and V4) (Fig. 4.5). The first core anchor (H2) was located in the passage between the sialic acid binding site and the 150-cavity, and was consisted of R118, D151, and R156. The major moiety types of the H2 core anchor were hydroxyl group, carboxylic amide, carboxylic acid, nitro group, and ketone (Fig. 4.5F). Most of these moieties yielded hydrogen bonds with this polar pocket, which suggested that compounds designed to occupy the 150-cavity may contain a polar moiety in this pocket to enhance the binding affinity. Moreover, 150-closed form NA structures collected from PDB showed that a conserved water atom forms a hydrogen-bonding network with E119, D151, and R156 (Fig. 4.7A), revealing the importance of the H2 core anchor.

The other extra two core anchors (H3 and V4) were in the 150-cavity (Fig. 4.5). The H2 core anchor, which contained G147, V149, and D151, was situated in the middle of the 150-loop (Fig. 4.5E). Furthermore, the 150-open form structures indicated that several water atoms often yield hydrogen-bonding interactions with G147, S145, and N146 in the H3 core anchor (Fig. 4.7B). The V4 core anchor was located in the center of the 150-cavity with Q136, D151, and T439, and preferred bulky moieties, such as aromatic ring, heterocyclic group, phenol, and oxohetarene. Based on the locations and moiety preferences of the H3 and V4 core anchors, it

revealed that an aromatic ring with polar moieties (*e.g.*, aniline and phenol moieties) may be sterically and physicochemically complementary to the 150-cavity.

Above results reveal that the core anchors of the sialic acid binding site are often involved in catalysis process or essential to the binding of the substrate and the inhibitors (*e.g.*, zanamivir and oseltamivir). However, there are already reports of drug-resistant mutations for these drugs^{94,115}, and most of the mutations are involved in the core anchors of the sialic acid binding site, such as R292K (E1 and V2) and E119V (H1)⁶. Therefore, developing drugs with a different action mechanism is required for the treatment of drug-resistant NAs.

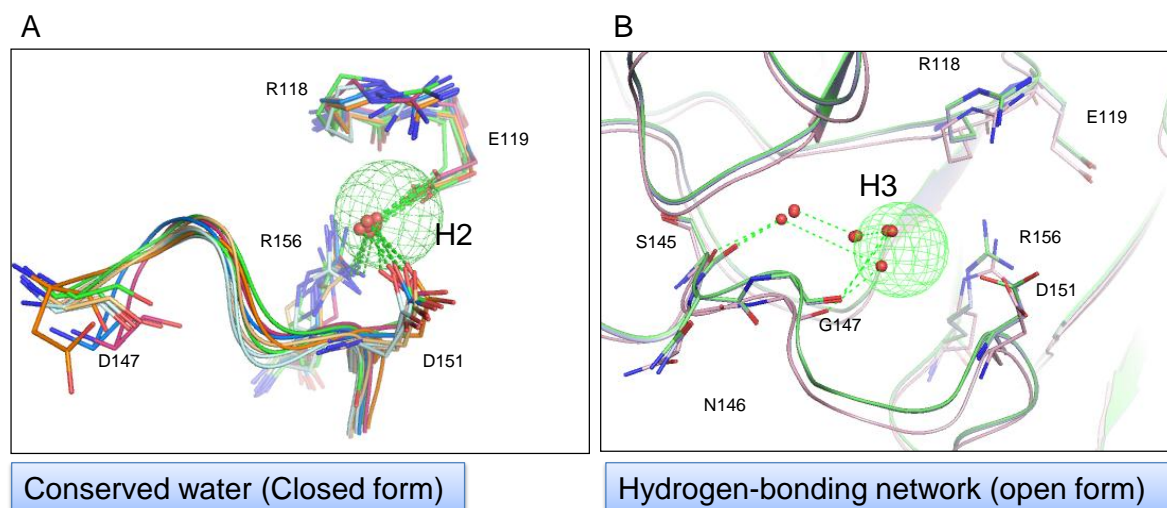


Figure 4.7. Conserved hydrogen-bonding interactions between water atoms and pockets of the H2 and H3 core anchors. (A) Conserved water atoms observed from the 150-closed form structures in the H2 core anchor. (B) Hydrogen-bonding networks observed from the 150-open form structures in the H3 core anchor. The red spheres indicate the water atoms. The hydrogen bonds are presented by green dash.

4.3.2 Optimization process of zanamivir derivatives

We first verified the pharmapathlog-based screening strategy on the lead optimization processes of NA drugs, including zanamivir, oseltamivir, and peramivir (Fig. 4.8). For instance, the initial inhibitor of zanamivir is DANA¹¹⁶, which inhibits NA with a K_i value $\sim 4.0 \mu\text{M}$ (Fig. 4.8A). Based on the binding mode of DANA and the site-moiety map of H1N1 NA, we were able to measure how well a compound moiety match a core anchor by using moiety energy, which is intermolecular energy between residues of a binding pocket and the moiety. Because the hydroxyl moiety of DANA had the highest moiety energy (-2.5 kcal/mol) in the H1 core anchor, we selected this anchor as the starting point to modify the compound structure of DANA to enhance binding affinity. The hydroxyl moiety was replaced by each preferred moiety (*i.e.*, amine, guanidine, ketone, amide, and azide moieties), and the replaced compounds were docked into the binding site of H1N1 NA to measure their moiety energies. We found that the two compounds modified by the amine moiety (4-amino-Neu5Ac2en, -3.5 kcal/mol) and the guanidine moiety (zanamivir, -14.2 kcal/mol) had lower moiety energies than that of DANA, suggesting the two moiety substitutions may increase the inhibitory ability of DANA. A previous study showed that the IC_{50} values of 4-amino-Neu5Ac2en and zanamivir were 150 and 1 nM, respectively, agreeing with our prediction for the lead optimization process of DANA. The same procedure was then applied to the optimization processes of oseltamivir and peramivir (Figs. 4.8B and 4.8C).

In addition, we collected zanamivir, oseltamivir, and peramivir analogues with experimental IC_{50} values from literatures for further analysis^{3-5,109,116-118} (Fig. 4.9). These analogues were classified to five groups: (1) Zanamivir_H1; (2) Oseltamivir_H1; (3) Oseltamivir_V1; (4) Oseltamivir_E1; (5) Peramivir_V1 according to their structures and locations of modified moieties. For example, DANA, 4-amino-Neu5Ac2en, and zanamivir

were classified to Zanamivir_H1 because their modified moieties (*e.g.*, hydroxyl moiety of DANA, amine moiety of 4-amino-Neu5Ac2en, and guanidine moiety of zanamivir) (Fig. 4.9). Then, we generated the moiety energies for each compound of the five groups. The Pearson's correlation coefficient between the moiety energies of the analogues and the IC₅₀ values was 0.76 (Fig. 4.10), showing the moiety preferences and moiety energies of core anchors are able to guide to assemble leads by optimal steric, hydrogen-bonding, and electronic moieties.

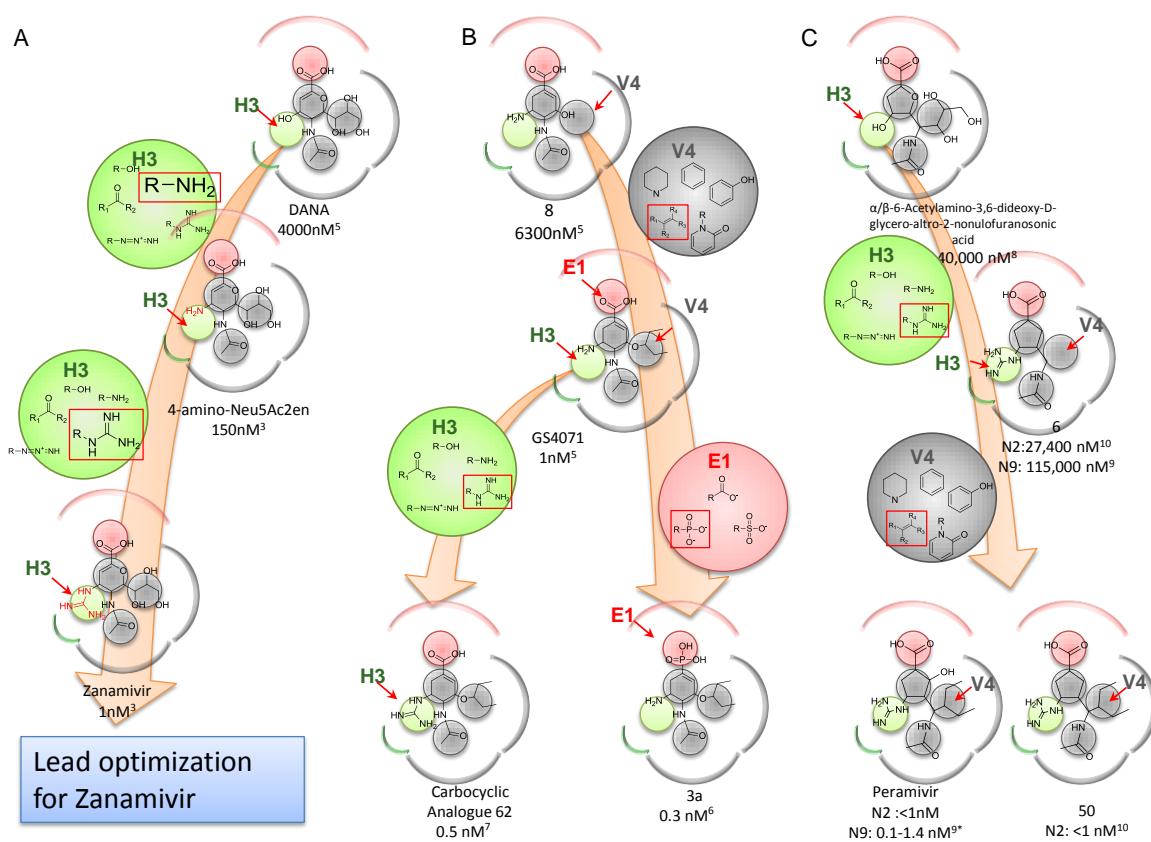


Figure 4.8. Core anchors of NA pharmacophores for lead optimization processes of (A) zanamivir, (B) oseltamivir, and (C) peramivir. The moieties in the circles present the preferred moieties of the core anchors. In this study, we selected zanamivir for further optimization process.

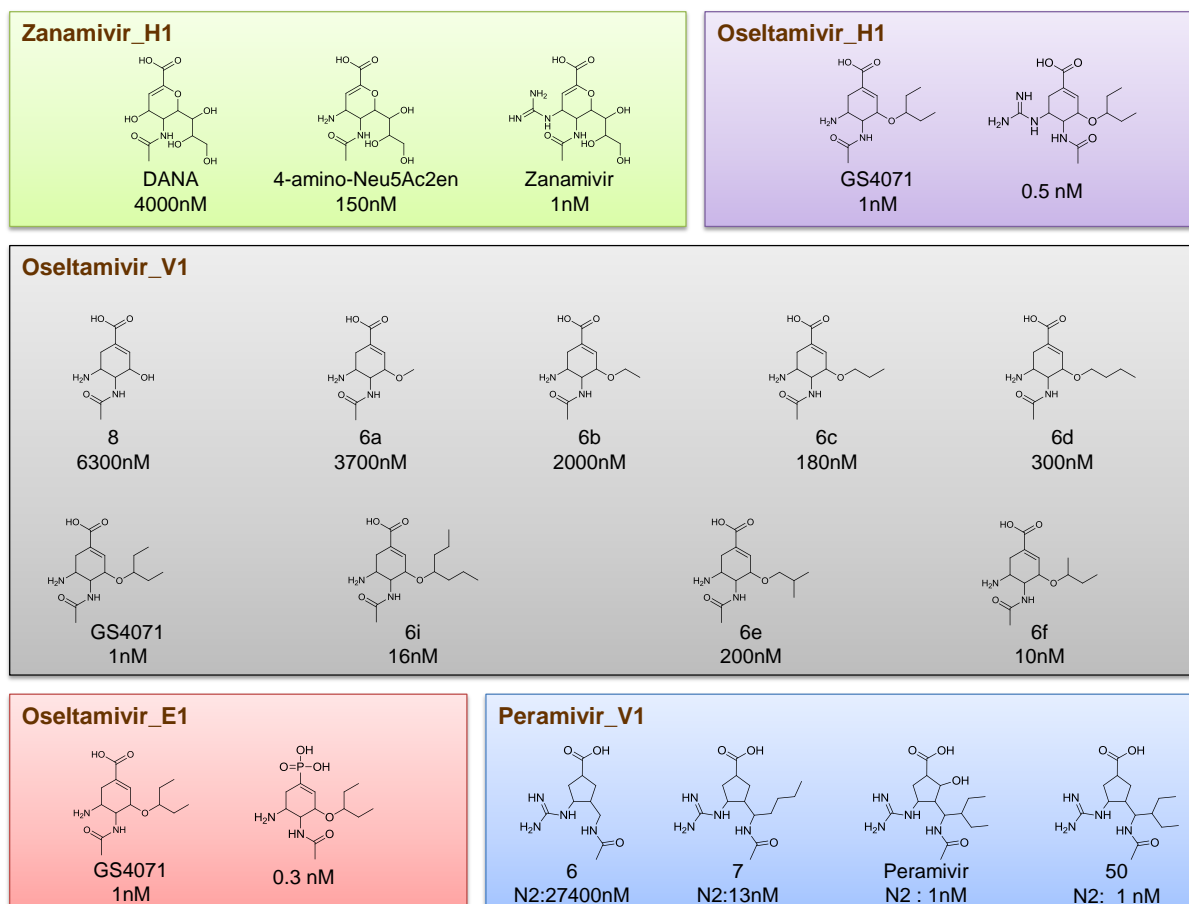


Figure 4.9. Zanamivir, oseltamivir, and peramivir analogues for verifying core anchors of pharmapathlogs on lead optimization processes. These analogues were classified to five groups: (1) Zanamivir_H1; (2) Oseltamivir_H1; (3) Oseltamivir_V1; (4) Oseltamivir_E1; (5) Peramivir_V1 for verification of moiety energy.

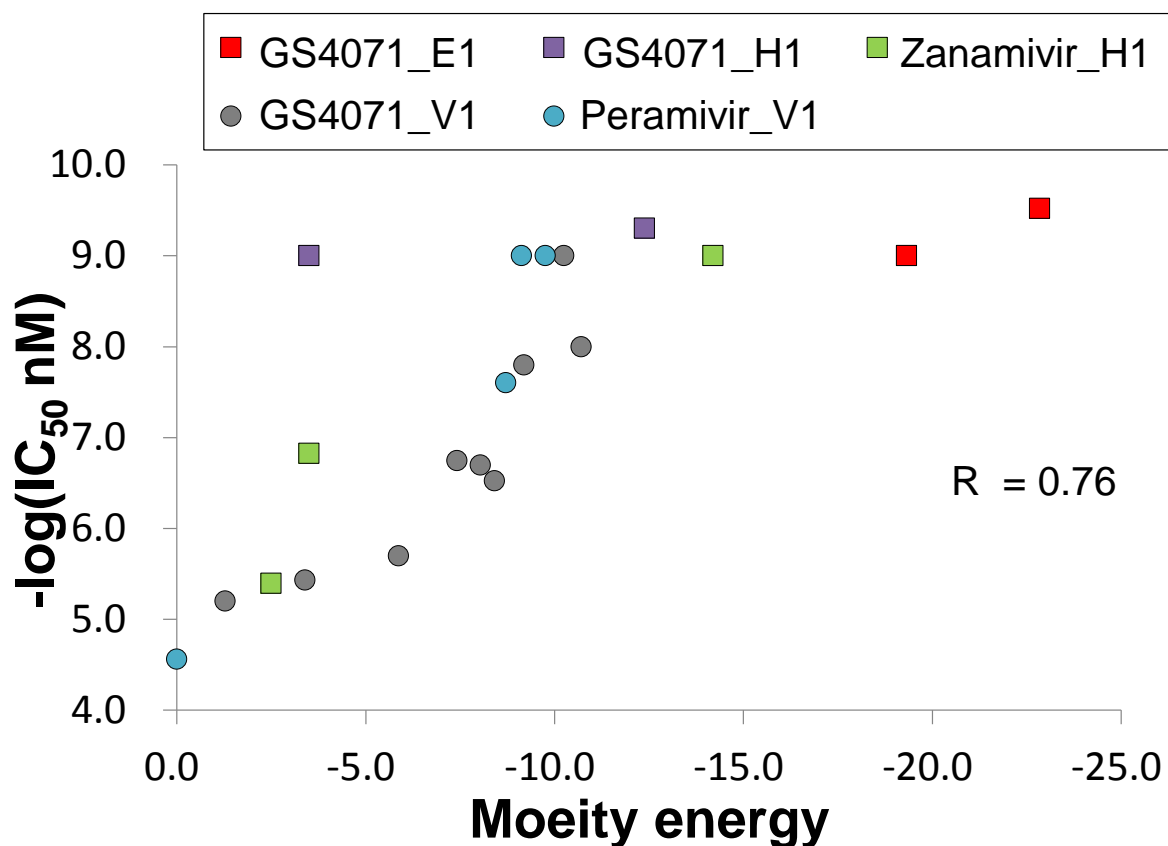
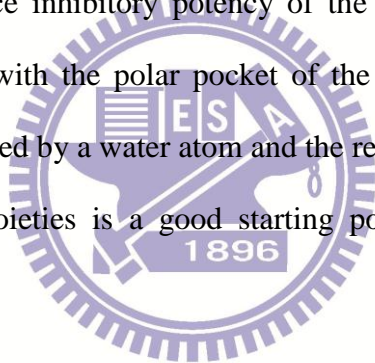


Figure 4.10. Pearson's correlation coefficient between moiety energies and IC_{50} values of compounds. The moiety energies were derived from the five groups of compounds. The high correlation (0.76) revealed that those compounds yielding strong interactions with pockets of anchors could have high antiviral potency.

The core anchors (*i.e.*, H2, V4, and H3) located in the 150-cavity were further applied to design zanamivir derivatives extending into the 150-cavity because zanamivir is less affected by a frequently drug-resistant mutation (H274Y) than oseltamivir^{6,106}. The site-moiety map showed that the closest mismatched anchor to zanamivir was the H2 core anchor (Fig. 4.11A), which was located at the midway between the 150-cavity and the sialic acid binding site, and preferred hydroxyl, carboxylic amide, carboxylic acid, nitro group, and ketone moieties (Fig. 4.5F). Then, 3-aminopropanal containing the ketone moiety was designed to be attached to the guanidine moiety of zanamivir because of the simple synthesis procedure (Fig. 4.11B). The

zanamivir derivative was synthesized by Dr. Chun-Cheng Lin and Mr. Chien-Hung Lin of NTHU. The inhibitory assay showed that the modified compound (compound 645) inhibited NA with a low IC_{50} value (25 nM). Furthermore, the docked conformation revealed that compound 645 was occupied at the 150-cavity and its ketone moiety yielded three hydrogen bonds with R118 and D151 of the H2 core anchor (Fig. 4.11G). To further study the importance of the H2 core anchor in designing the new type inhibitors, a 645 analogue (compound 636) without the ketone moiety was synthesized (Fig. 4.11C). The elimination of the ketone moiety caused at least 160-fold reduction in NA inhibitory ability. Moreover, both of the zanamivir derivatives reported by the previous studies do not contain the ketone moiety in the H2 core anchor and inhibit NA with less potency (IC_{50} or $K_i > 1,000$ nM)^{99,100}, revealing that polar moieties are required to enhance inhibitory potency of the new type inhibitors by yielding hydrogen-bonding interactions with the polar pocket of the H2 core anchor. The conserved hydrogen-bonding network formed by a water atom and the residues of the H2 core anchor also suggest that attaching polar moieties is a good starting point for design of the new type inhibitors (Fig. 4.7A).



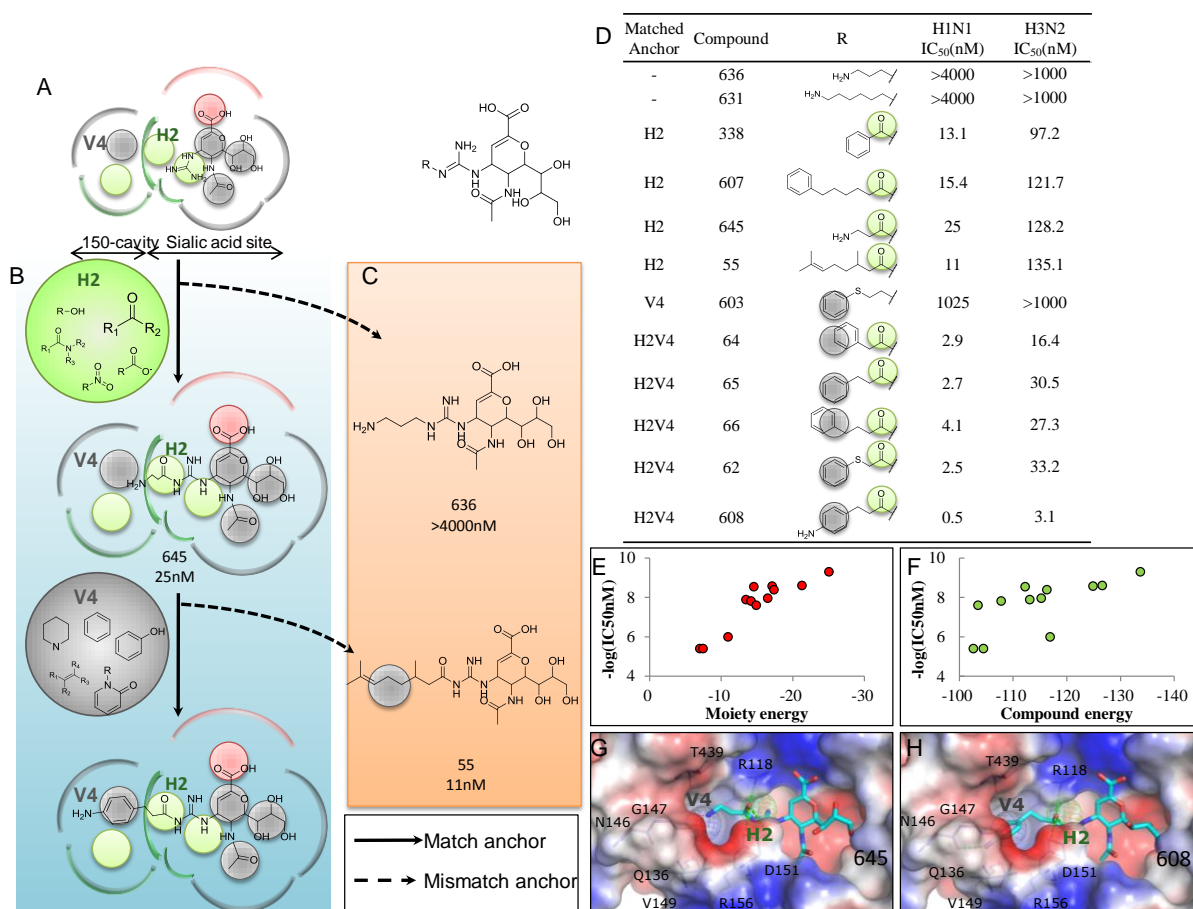


Figure 4.11. Lead optimization process of zanamivir derivatives for extending to the 150-cavity. (A) Zanamivir in the site-moiety map. (B) Design of zanamivir derivatives through matching the core anchors in the 150-cavity. Zanamivir derivatives mismatching the core anchors often show poor potency (C). (D) Enzyme-based assays of zanamivir derivatives. The core anchors matched by the derivatives are labeled. The zanamivir derivatives were synthesized by Dr. Chun-Cheng Lin and Mr. Chien-Hung Lin of NTHU. The purity and quality of the zanamivir derivatives will be further analyzed. Relationships between IC₅₀ values and docking energies including moiety energies (E) and compound energies (F). Docked poses of compound 645 (G) and compound 608 (H).

To further enhance the binding affinity of the compound 645, this compound was subsequently attached an aromatic moiety, which was the most preferred moiety in the V4 core anchor (a series of aromatic moieties with different chain length were attached to study structure-activity relationship (Fig. 4.11D), and these compounds were docked into the binding site of H1N1 NA. The moiety energies of benzene (compound 338), toluene (compound 64), ethylbenzene (compound 65), and propylbenzene (compound 66) were -11.0, -6.9, -12.8, and -11.6 kcal/mol, respectively. This suggests that adding the ethylbenzene moiety could increase the binding affinity of the compound 645 by providing additional van der Waals contacts with the residues of the 150-cavity. Next, the compounds with the aromatic moieties were synthesized and verified by NA inhibitory assays. In addition, a compound (compound 55) without aromatic moieties was synthesized to compare with the aromatic containing compounds (Fig. 4.11C). Among these compounds, compound 65 containing ethylbenzene moiety had the lowest IC_{50} value (2.7 nM) (Fig. 4.11D). The addition of the aromatic moiety increased the binding affinity of the compound 65 by approximately 10-fold, which reveals that the van der Waals interactions between Q136, D151, T439 of the V1 core anchor and the aromatic moieties could enhance the inhibitory ability of the new type inhibitors.

We tried to attach polar moieties to the ethylbenzene moiety of the compound 65 for increasing moiety energy in the H3 core anchor. Two 65 derivatives with benzene-1,3-diamine and aniline (compound 608) moieties were synthesized. However, the former compound was not tested because of instability. Although the docked conformation of compound 608 showed the amine of the aniline was not located in the H3 core anchor, compound 608 was able to form a hydrogen bond with N146 (Fig. 4.11H). The IC_{50} value of compound 608 was 0.53 nM, which was approximately 5-fold potency than that of compound 65 in NA inhibition. Furthermore, the Pearson's correlation coefficient between moiety energies and the IC_{50} values was -0.91, whereas using the compound energies showed a poor correlation coefficient (-0.69) (Figs. 4.11E and 4.11F). In the future, the purity and quality of the zanamivir derivatives will

be further analyzed. These experimental results revealed that the core anchors of the site-moiety map, and moiety energies are useful for guiding lead optimization process.

4.3.3 Identified novel inhibitors

Based on the RCS, which combined pharmapathlog-based ranks and energy-based ranks, we obtained ranks of each compound. Subsequently, 24 top-ranked compounds located in the 150-cavity were selected for bioassays. Of these 24 compounds, three new type inhibitors with low IC₅₀ values for H1N1 (<10.0 μM) and H5N1 (<20.0 μM) NAs were identified (Fig. 4.12).

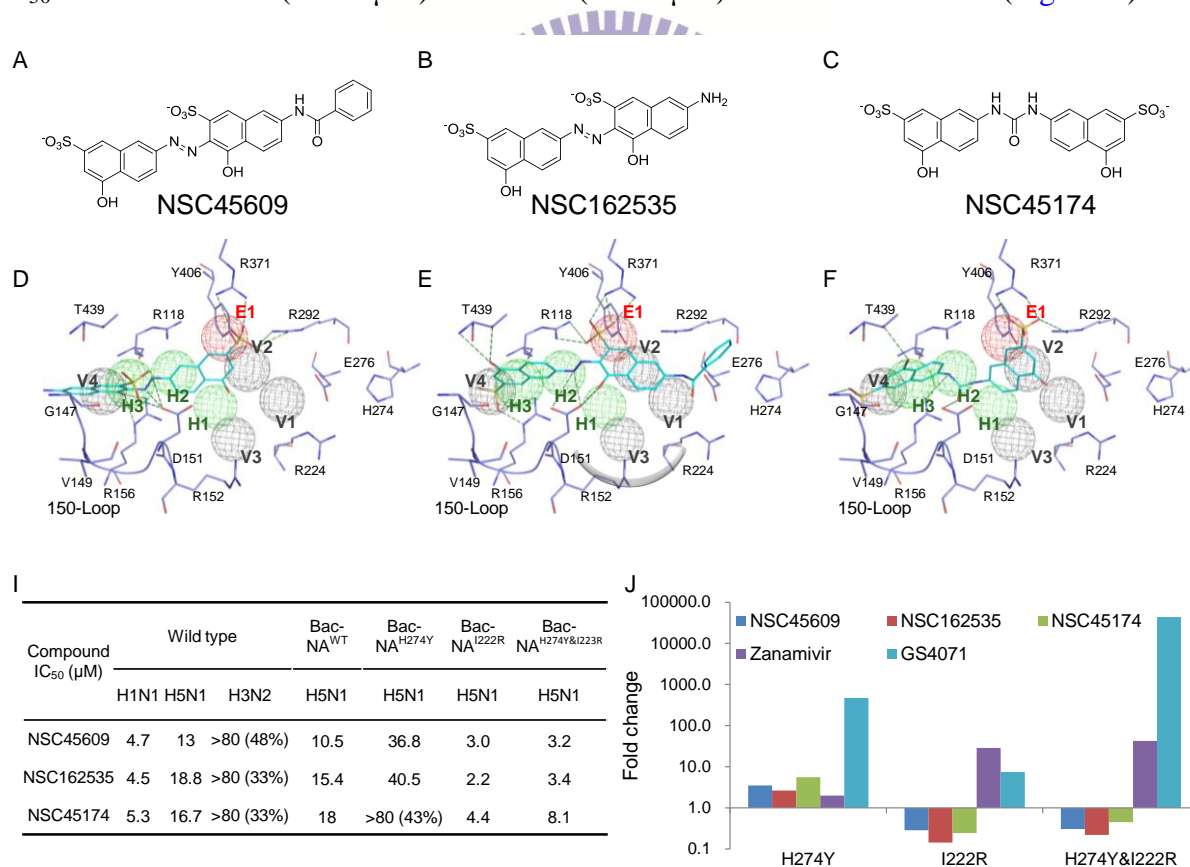


Figure 4.12. Discovered new type inhibitors. (A-C) Compound structures. (D-F) Docked conformations of the inhibitors. (I) Enzyme-based assays of the inhibitors. (J) The fold change in IC₅₀ values of the inhibitors for H274Y, I222R, and H274&I222R.

NSC45609, NSC45174, and NSC162535 agreed with 5, 4, 5 anchors, respectively. The three new type inhibitors consistently contained sulfuric acid monoester moieties interacting with R118, R292, and R371 of the E1 core anchor, and aromatic moieties interacting with Q136, D151, and T439 of the V4 core anchor (Fig. 4.12). The sulfuric acid monoester was a negative charged moiety, which was similar to the carboxyl groups of zanamivir, oseltamivir, and the substrate sialic acid. Although the urea moiety of NSC45174 was different from the azo moieties of the other two inhibitors, these moieties formed consistently hydrogen-bonding interactions with the pocket of the H2 core anchor. Similarly, the sulfuric acid monoester moiety of NSC162535, and the hydroxyl moieties of NSC45609 and NSC45174 yielded consistently hydrogen bonds with the pocket of the H3 core anchor despite of their different moieties. It should be noted that the three new type inhibitors were not located within the V1 core anchor, which was nearby I222 and H274. This suggests that these inhibitors may be affected by the two drug-resistant mutations.

We subsequently tested the inhibitory activities of these inhibitors on the H1N1 and H5N1 NAs with H274Y (NA^{H274Y}), I222R (NA^{I222R}), and H274Y&I222R (NA^{H274Y&I222R}). The NA inhibition assays showed that the new type inhibitors were not affect by these mutations (Fig. 4.12I), whereas zanamivir and oseltamivir highly reduce their potency (Fig. 4.12J). The experimental results supported our modeling predictions and indicated that designing new type inhibitors should avoid yielding van der Waals interactions with the residues of the V1 core anchor.

We further tested the inhibition ability of the three inhibitors on the cytopathic effect of cells infected by influenza viruses. They inhibited >50% of the cytopathic effect without causing apparent cytotoxicity for H1N1 NA (Fig. 4.13). The IC₅₀ values of NSC45609, NSC162535, and NSC45174 were 16.5, 61.5, and 36.5 μM, respectively, for H1N1 NA. These results indicate that the three inhibitors may be a good starting point for designing new

inhibitors to overcome the drug resistance introduced by H274Y and I222R for N1 NAs without causing apparent cytotoxicity.

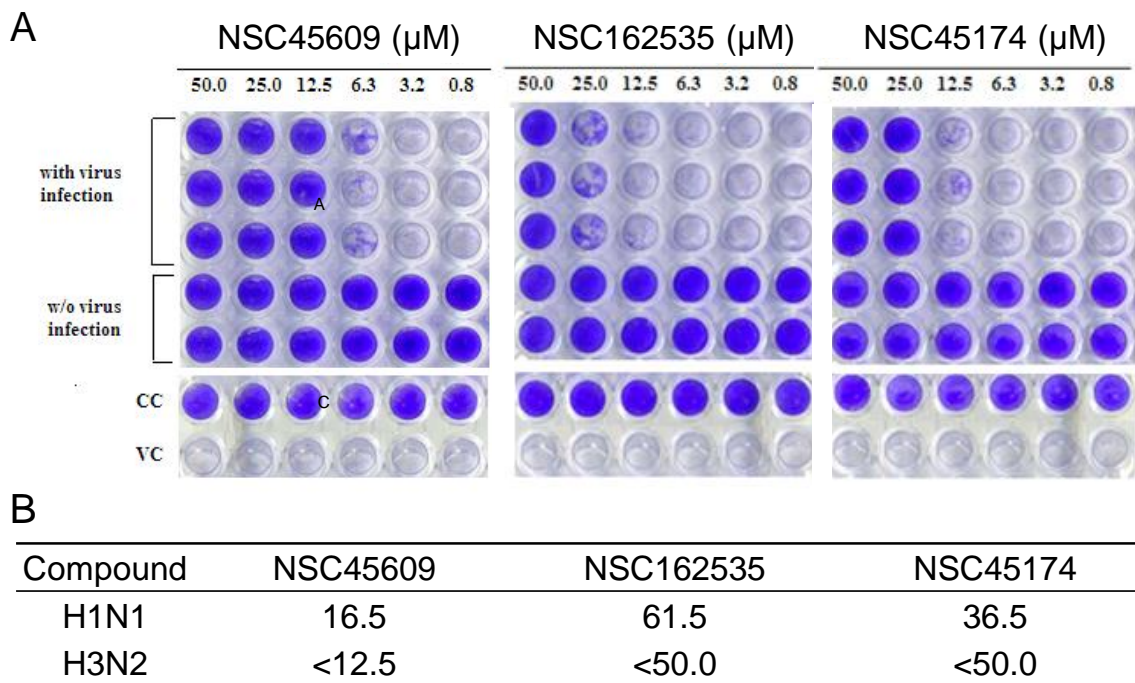


Figure 4.13. Effects of influenza virus in cytopathic effect inhibition assays of novel inhibitors. (A) Cytopathic effect inhibition assays for H1N1 virus. (B) The IC₅₀ values of the inhibitors for H1N1 and H3N2 viruses. This figure was obtained from the works of Dr. John T.A. Hsu and Dr. Hui-Chen Hung.

4.3.4 Proposed binding mechanism of novel inhibitors

The 150-loop of group-2 NAs might maintain an open form conformation as well as the 150-loop of group-1 NAs. Russell et al. indicated that the 150-loop sequences of the two group NAs share a high sequence similarity (Fig. 4.14), and the energy of the 150-loop open form and closed form conformations may be similar⁹¹. In addition, two H1N1 structures (PDB code 3o9k⁹⁹ and 3nss⁸⁵) were solved. One of them was locked in the open form conformation by 3-

(*p*-tolyl)allyl-Neu5Ac2en extending to the 150-cavity, and the other remained the closed form conformation without compound binding. Moreover, the 150-loop of the strain A/Paris/2590/2009 may be a closed form conformation because a structure (A/California/04/2009, PDB code 3o9k) with 100% sequence identify to A/Paris/2590/2009 has a closed form conformation without ligand binding (Fig. 4.14). However, the strain A/Paris/2590/2009 can be inhibited by 3-(*p*-tolyl)allyl-Neu5Ac2en (Fig. 4.14). The above analyses implied that the closed form conformation could be induced by a zanamivir derivative. Interestingly, our designed zanamivir derivatives extending to the 150-cavity inhibited the H3N2 NA with low IC₅₀ values (3.1nM ~97.2nM) (Fig. 4.11D), which also suggested the possibility that H3N2 NA may maintain a 150-loop open form. However, the IC₅₀ values obtained from H3N2 NA were about 10-folds higher than those obtained from H1N1 NA (Fig. 4.11D), implying that there was an energy barrier between the two conformations in H3N2 NA. On the basis of the intra-molecular interactions of the H3N2 NA, we found that the energy barrier may be formed by a salt bridge between D147 and H150.

Group	Subtype	150-form	147	149	151	431	Strain	Annotation		
Group 1	H5N1	unknown	G	T	V	K	D	PK	A/Hong	
Group 1	H5N1	open & closed	G	T	V	K	D	PK	A/Viet	
Group 1	H1N1	unknown	G	T	I	K	D	PK	A/Paris/2590/2009	nature commutation 2010; ki 1.7 uM
Group 1	H1N1	closed	G	T	I	K	D	PK	A/California/04/2009	nature structural & molecular biology 2010
Group 1	H1N1	unknown	G	T	F	K	D	PE	A/WSN/1933	NHRI
Group 1	H1N1	open & closed	G	T	V	K	D	PK	A/Brevig	
Group 1	H10N4	open & closed	G	T	V	K	D	PK	A/mink/Sweden/E1266	
Group 1	H3N8	open & closed	G	T	V	K	D	PE	A/duck/Ukraine/1/196	
Group 2	H3N2	unknown	D	T	I	H	D	EQ	A/udorn/1972	NHRI
Group 2	H2N2	closed	D	T	V	H	D	KQ	A/Tokyo/3/1967	
Group 2	H3N2	closed	D	T	V	H	D	KQ	A/Memphis/31/98	
Group 2	H11N9	closed	G	T	I	H	D	PK	A/tern/Australia/G70	
Group 2	H13N9	closed	G	T	I	H	D	PK	A/whale/Maine/1/1984	
Group 2	H11N6	closed	G	T	I	H	D	PK	A/duck/England/1/195	

Figure 4.14. Multiple sequence alignment of NAs. In 150-form column, “closed” or “open” indicate that 150-closed form or 150-open form structures have been solved for the strain, and

“unknown” indicates that no structures are available for the strain. The strains annotated by “NHRI” (National Health Research Institutes) are the strains used for bioassays in this study.

Based on the above observations, we thus proposed that the binding mechanism of the compound 608 for inhibiting H3N2 NA (Fig. 4.15). Firstly, the 150-loop was in the close form and remained slightly flexible (Figs. 4.15A and 4.15D), which was observed from the 150-loop B-factor of a H3N2 NA structure (PDB code 2aep¹⁰²). Subsequently, the compound 608 interacted with the NA, and the ketone moiety of the compound substituted the position of a water atom that formed a hydrogen-bonding network consisted of a water atom, E119, D151, and R156 in the H2 core anchor. Next, the ketone moiety formed a new hydrogen-bonding network with R118 and D151, providing the energy to break the salt bridge between D147 and H150 (Figs. 4.15B and 4.15E). Finally, the 150-loop became an open form conformation, and the compound 608 made additional interactions with the residues of the 150-cavity to stable the complex (Figs. 4.15C and 4.15F). In addition, the compound structure and binding mode of compound 348 are similar to those of 3-(*p*-tolyl)allyl-Neu5Ac2en obtained from a structure (PDB code 3o9k⁹⁹) (Figs. 4.16A and 4.16B), implying compound 348 may extend to the 150-cavity as same as 3-(*p*-tolyl)allyl-Neu5Ac2en. Based on the observation, compound 608 could also interact with the 150-cavity because the compound structures of compound 608 and compound 348 are similar (Fig. 4.16).

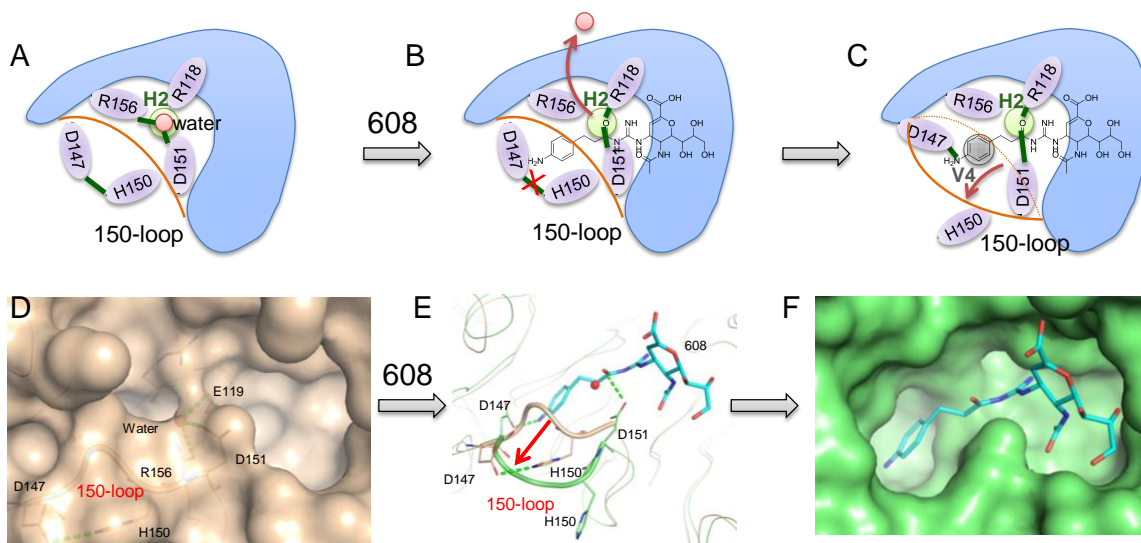
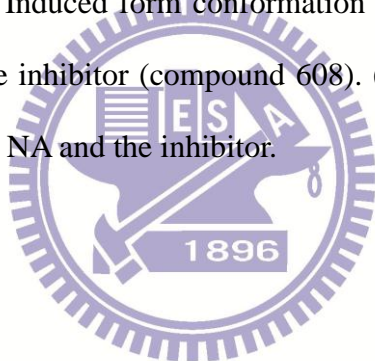


Figure 4.15. Proposed new mechanism of H3N2 NA. (A)(D) 150-closed form conformation before inhibitor binding. (B)(E) Induced form conformation of 150-loop. The 150-open form may be induced by the new type inhibitor (compound 608). (C)(F) The predicted interactions between the 150-cavity of H3N2 NA and the inhibitor.



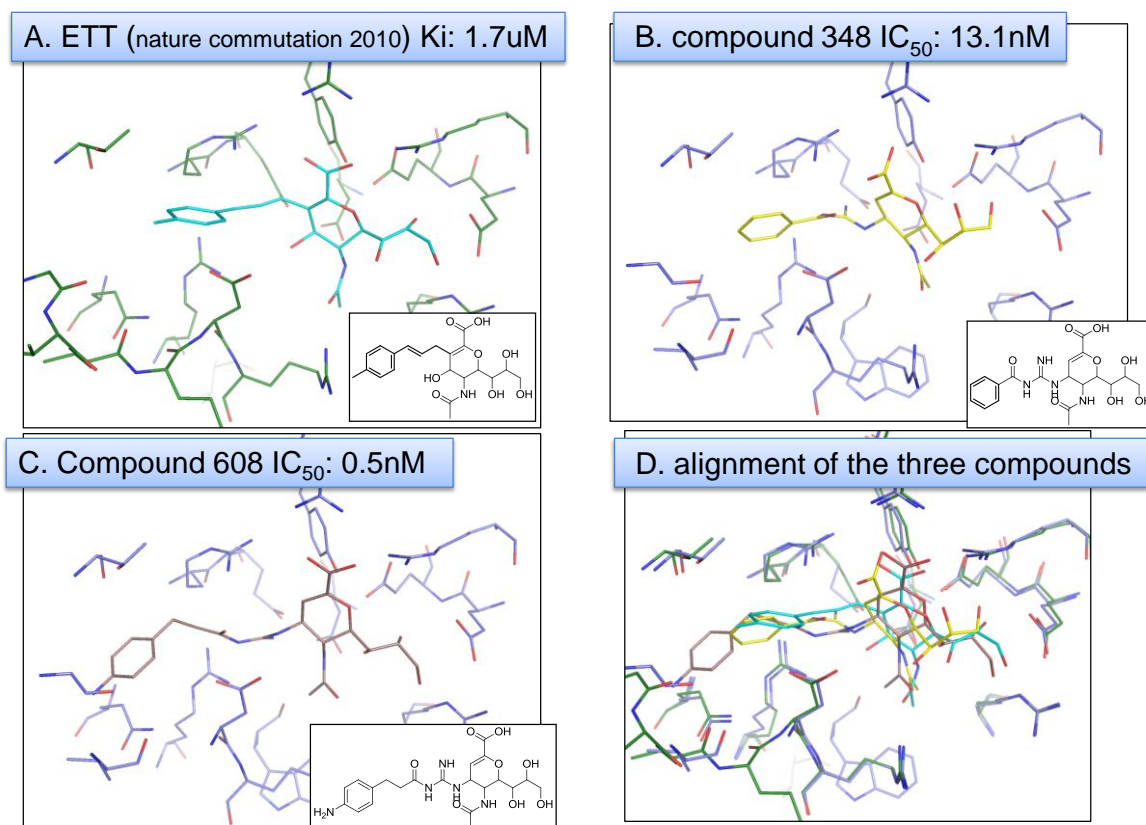


Figure 4.16. Binding mode comparison of new type inhibitors. (A) Binding mode of 3-(*p*-tolyl)allyl-Neu5Ac2en obtained from a structure (PDB code 3o9k). (B) Docked conformation of compound 348, which is structurally similar to 3-(*p*-tolyl)allyl-Neu5Ac2en. Compound 348 extends to the 150-cavity as same as 3-(*p*-tolyl)allyl-Neu5Ac2en. (C) Docked conformation of compound 608, which is similar to compound 343. Compound 608 interacts with residues of the 150-cavity. (D) Structure alignment of the three inhibitors.

4.3.5 Advantages of pharmapathlog-based screening

strategy

One of the great advantages of using the pharmapathlog-based screening strategy is that core anchors of site-moiety maps could guide to optimize potency of lead compounds by various key moieties. Generally, a group of lead compound derivatives are synthesized to find compounds with better inhibitory ability, which is time-consuming and high cost procedure. Although some computational approaches have been reported, these approaches remained problematic because of the inaccurate predictions, poor synthesizability of the suggested moieties, and low moiety diversity¹¹⁹⁻¹²¹. As a result, the lead optimization process takes a long time to complete. For example, it took about 20 years from the appearance of the first NA inhibitor (DANA) to the discovery of the first drug (zanamivir). The pharmapathlog-based screening strategy proposed here thus showed the value in the optimization process. The moiety compositions of anchors provide chemists numerous options to select a simple synthesis process for the lead optimization. Through this strategy, we successfully designed several potent zanamivir derivatives (<10 nM). Comparing the results with other reported zanamivir derivatives^{99,100}, compound 608 had a significantly improved inhibitory potency to NA. These results reveal that the pharmapathlog-based screening strategy is useful to guide the lead optimization process.

The pharmapathlog-based screening strategy could be used to discover novel compounds with different scaffolds. When this strategy was applied to identify NA inhibitors, we performed high-throughput screening on an in-house compound library of about 10,000 compounds in parallel. However, all of the compounds showed no inhibitory effect at 40 μ M on the NA assays, whereas three potent inhibitors with novel scaffolds were identified by the

pharmapathlog-based screening strategy. Moreover, core anchors of site-moiety maps showed the possibility of designing new type inhibitors for combating drug-resistant pathogens. For instance, our experimental results showed that the three inhibitors overcame the drug resistances introduced by H274Y and I222R for H1N1 NA, and the cell-based assays showed that the three inhibitors inhibited the cytopathic effect of cells infected with influenza viruses without causing apparent cytotoxicity. Therefore, we believe that the pharmapathlog-based screening strategy is useful to identify lead inhibitors efficiently and can be extend to other human diseases and drug-resistant pathogens.

4.4 Conclusions

The study demonstrated the utility and feasibility of the pharmapathlog-based screening strategy for identifying new type inhibitors and optimizing lead compounds. Three new inhibitors were identified, and these inhibitors showed the inhibitory abilities for an oseltamivir-resistant NA with H274Y and I222R mutations. The results reveal the advantages to of the pharmapathlog-based screening strategy to find new types of inhibitor for combating drug-resistant strains. Furthermore, five zanamivir derivatives were designed, and showed potent inhibitory effects for NAs. Based on the experimental results, we believe that the pharmapathlog-based screening strategy and these inhibitors are useful to design a novel class of inhibitors that may overcome the drug resistance.

4.5 Acknowledgments

We thank to Dr. John T.A. Hsu and Dr. Hui-Chen Hung of National Health Research Institutes for kindly providing the experimental results including enzyme inhibition assays and cytopathic effect inhibition assays. We also thank to Dr. Chun-Cheng Lin and Mr. Chien-Hung Lin of National Tsing Hua University for compound synthesis.



Chapter 5. Conclusions

5.1 Summary

The concept of "one-disease, one-target, one-drug" induces researchers to develop inhibitors with high specificity. Many drugs (*e.g.*, zanamivir and oseltamivir) have been reported by using this strategy. Currently, inhibitors designed by the single target concept often become less effective because of mutations during treatments. As a result, designing multitarget inhibitor is becoming a promising strategy for drug design. However, it is still a challenging task since many target proteins are not similar in their sequences or structures. Here, we proposed a new strategy, namely pharmapathlog-based screening strategy, to discover multitarget inhibitors by using site-moiety maps. The main concept of this strategy is to extract core binding environments (core anchors) among orthologs in the same pathways. For those orthologs that share comparable core anchors, these proteins can be considered pharmapathlogs. Then, we found compounds that agree with the core anchors of the pharmapathlogs for bioassays.

To verify the utility of the pharmapathlog-based screening strategy, we firstly applied this strategy to discover multitarget inhibitors of shikimate pathways in *Helicobacter pylori* and *Mycobacterium tuberculosis*. By use of this strategy, four core anchors consistently shared in shikimate dehydrogenase and shikimate kinase of *Helicobacter pylori*, and shikimate kinase of *Mycobacterium tuberculosis* were discovered. Three novel multitarget inhibitors that simultaneously inhibited shikimate dehydrogenase and shikimate kinase of *Helicobacter pylori* were also identified. The three inhibitors were also able to inhibit shikimate kinase of *Mycobacterium tuberculosis*. Subsequently, we applied this strategy to design and discover new type inhibitors of neuraminidases. Five zanamivir derivatives inhibiting H1N1 and H5N1 neuraminidases with IC_{50} values in the <10 nanomolar range were discovered. To our

knowledge, this was the first inhibitors that were predicted to be occupied in the 150-cavity with nanomolar inhibition. In addition, we found three novel inhibitors ($IC_{50} < 10 \mu M$) predicted to be located in the 150-cavity instead of the sialic acid binding site. Our experimental results revealed that the three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 NAs without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains. The experimental results demonstrate that the pharmapathlog-based screening strategy is useful to identify and optimize inhibitors. We believe that this strategy is helpful to design new drugs for enhancing public health.

5.2 Major contributions

In summary, the major contributions of this study were listed as the following:

- We proposed a new strategy, namely pharmapathlog-based screening strategy, to find pharmapathlogs, core anchors of pharmapathlogs, and multitarget inhibitors. A group of proteins are regarded as pharmapathlogs when they meet the following criteria: (1) they are orthologs in the same pathway; (2) they share similar binding environments in their binding sites; and (3) they can be bound or inhibited by the same compounds.
- We developed a web server, namely SiMMap, for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. A site-moiety map of a protein binding site can present key binding environments (anchors). Therefore, alignment of site-moiety maps of multiple proteins can extract core binding environments (core anchors) instead of using sequence or structure alignments. It is useful to discover pharmapathlogs sharing comparable core anchors

when proteins are dissimilar in sequences or structures.

- We discovered three multitarget inhibitors (IC_{50} values $<10 \mu\text{M}$) for shikimate dehydrogenase and shikimate kinase of *Helicobacter pylori*. The three inhibitors also showed inhibitory effects (IC_{50} values $<10 \mu\text{M}$) for shikimate kinase of *Mycobacterium tuberculosis*. In addition, two specific inhibitors (IC_{50} values $<10 \mu\text{M}$) for shikimate dehydrogenase of *Helicobacter pylori* were identified. These inhibitors are useful to develop new line antibiotics.
- We designed five zanamivir derivatives inhibiting neuraminidase with IC_{50} values in the <10 nanomolar range. In addition, three compounds with novel scaffolds were discovered with IC_{50} values $<10 \mu\text{M}$. These inhibitors were predicted to be located as the 150-cavity, which is a new cavity to design new type neuraminidase inhibitors. Our experimental results reveals the three inhibitors may overcome the drug resistances introduced by H274Y and I222R for H1N1 NAs without causing apparent cytotoxicity, suggesting a starting point to combat drug-resistant strains.

5.3 Future works

A large number of crystal structures reveals that proteins performing diverse functions may share similar binding sites¹³. Furthermore, some recent studies reveal that proteins sharing consistently key binding environments are able to be activated or inhibited by the same compounds despite of dissimilar whole protein structures^{45,122}. For example, an antiretroviral drug Cosalane is able to simultaneously inhibit multiple targets of HIV-1 proteins, including gp120, integrase, protease, and reverse transcriptase⁶⁵. Additional examples are isozymes, which catalyze identical chemical reactions but have different molecular structures or

sequences¹²³. The binding sites of these proteins are often similar in physico-chemical properties and geometric shapes for bindings of the same compounds. Therefore, it is possible to extend the pharmapathlog-based screening strategy to develop multitarget inhibitors for targets in multiple pathways. These multitarget inhibitors can enhance the efficiency of treatment by targeting several proteins in a disease for clinical use^{65,124}.

To address the issue, currently, we propose a new concept of "pharmacologs". Pharmacologs are a group of proteins sharing comparable binding environments at the protein-ligand interfaces, and can be bound the same compounds. Here, we described this new concept by using KDR (PDB code 3b8q¹²⁵), KIT (PDB code 1t46¹²⁶), and ABL1 (PDB code 2hyy¹²⁷), which are targets in acute myeloid leukemia, as the example (Fig. 5.1). Although KDR, KIT, and ABL1 are located at different pathways that are proliferation and antiapoptosis, these proteins share comparable core anchors derived by comparing their site-moiety maps (Fig. 5.1B). Therefore, the proteins are regarded as pharmacologs, and may be inhibited the same compounds. As expected, a cancer drug, imatinib, has been reported to inhibit KDR, KIT, and ABL1¹²⁵⁻¹²⁷ (Fig. 5.1C), suggesting the opportunity for designing multitarget inhibitors via pharmacologs.

The major difference between pharmapathlogs and pharmacologs is that pharmacologs could be constituted be targets of multiple disease-related pathways (Fig. 5.1A), whereas pharmapathlogs only contain targets in a pathway. The experience of studying pharmapathlogs can provide an opportunity to develop the new concept. In the future, we will identify pharmacologs according to the following steps: (1) we will perform large scale constructions of site-moiety maps of multiple disease-related targets; (2) pharmacologs with comparable core anchors will be discovered comparing the site-moiety maps of multiple targets (Fig. 5.1B); (3) we will screen compound databases to identify potential multiple targets that agree with the core anchors for bioassays (Fig. 5.1C). We believe that this new concept is useful for drug

development and understanding the mechanisms of drug actions.

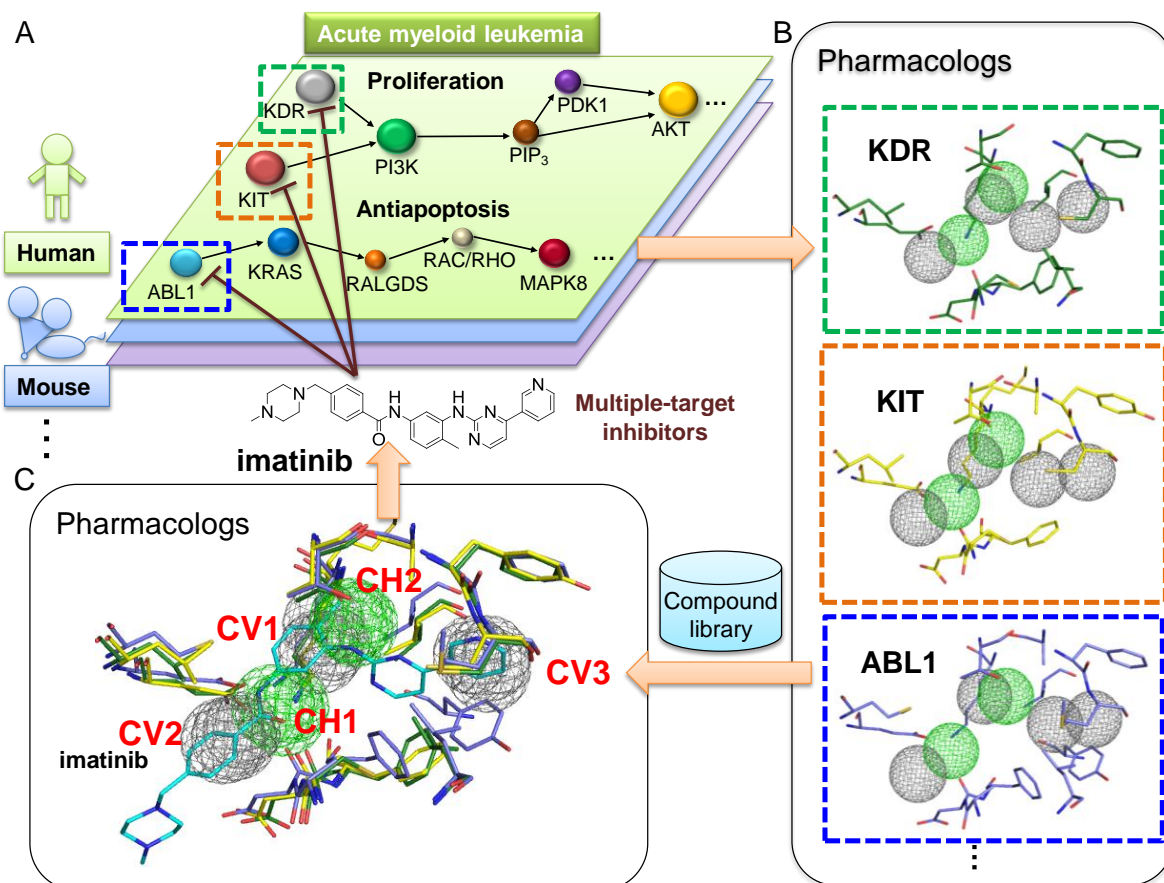


Figure 5.1. Pharmacologs using KDR, KIT, and ABL1 as the example. (A) Acute myeloid leukemia associated proteins. (B) Pharmacologs with the comparable core anchors derived from site-moiety maps. (C) Imatinib, a cancer drug, can inhibit the proteins of the pharmacologs.

List of publications

Journal papers

1. **K.-C. Hsu**, Y.-F. Chen, S.-R. Lin and J.-M. Yang*, "iGEMDOCK: A Graphical Environment of Enhancing GEMDOCK using Pharmacological Interactions and Post-screening Analysis," *BMC Bioinformatics*, 12(Suppl 1):S33, 2011
2. Y.-F. Chen[†], **K.-C. Hsu**[†], S.-R. Lin, W.-C. Wang, Y.-C. Huang and J.-M. Yang*, "SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties," *Nucleic Acids Research*, W424 - W430, 2010. [†] Contributed equally
3. D.-L. Clinciu, J.-M. Yang, **K.-C. Hsu**, C.-C. Lo, S. Wallace, and H.-C. Yu, "The Relevance of Protein-Ligand Interaction Profiles in Computer-Aided Novel Compound Design and Applications" *Current Bioinformatics*, in press.
4. C.-C. Tsai, H.-F. Liu, **K.-C. Hsu**, J.-M. Yang, C. Chen, K.-K. Liu, T.-S. Hsu, J.-I. Chao, "7-Chloro-6-piperidin-1-yl-quinoline-5,8-dione (PT-262), a novel ROCK inhibitor blocks cytoskeleton function and cell migration," *Biochemical Pharmacology*, vol. 62, pp. 799-808, 2011.
5. K.-P. Liu, **K.-C. Hsu**, J.-W. Huang, Y.-S. Chang, and J.-M. Yang*, "ATRIPPI: An atom-residue preference scoring function for protein-protein interactions," *International Journal on Artificial Intelligence Tools*, pp. 251-266, 2010.
6. **K.-C. Hsu**, Y.-F. Chen, and J.-M. Yang*, "GemAffinity: A scoring function for predicting binding affinity and virtual screening," *International Journal on Data Mining and Bioinformatics*, 2010, in press.

Conference Papers

1. **K.-C. Hsu**, Y.-F. Chen, and J.-M. Yang*. (2008) Binding Affinity Analysis of Protein-Ligand Complexes. *The 2nd International Conference on Bioinformatics and Biomedical Engineering*, pp. 167-171
2. Y.-F. Chen, **K.-C. Hsu**, P.-T. Lin, D.-F. Hsu, B.- S. Kristal, and J.-M. Yang *. (2011) LigSeeSVM: Ligand-Based Virtual Screening Using Support Vector Machines and Information Fusion. *International Journal of Computational Biology and Drug Design*, pp, in press.

References

- 1 Bleicher, K. H., Bohm, H. J., Muller, K. & Alanine, A. I. Hit and lead generation: Beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, 369-378 (2003).
- 2 Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **7**, 1047-1055 (2002).
- 3 von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418-423 (1993).
- 4 Maring, C. J. *et al.* Structure-based characterization and optimization of novel hydrophobic binding interactions in a series of pyrrolidine influenza neuraminidase inhibitors. *Journal of Medicinal Chemistry* **48**, 3980-3990 (2005).
- 5 Kim, C. U. *et al.* Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *Journal of the American Chemical Society* **119**, 681-690 (1997).
- 6 Moscona, A. Oseltamivir resistance--disabling our influenza defenses. *The New England Journal of Medicine* **353**, 2633-2636 (2005).
- 7 Gerrits, M. M., de Zoete, M. R., Arents, N. L. A., Kuipers, E. J. & Kusters, J. G. 16S rRNA mutation-mediated tetracycline resistance in *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy* **46**, 2996-3000 (2002).
- 8 Merlo, C. A. *et al.* Incidence and risk factors for multiple antibiotic-resistant *Pseudomonas aeruginosa* in cystic fibrosis. *Chest* **132**, 562-568 (2007).
- 9 Piddock, L. J. V. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clin Microbiol Rev* **19**, 382-+ (2006).
- 10 Navon-Venezia, S., Leavitt, A. & Carmeli, Y. High tigecycline resistance in multidrug-resistant *Acinetobacter baumannii* - authors' response. *J Antimicrob Chemoth* **60**, 178-179 (2007).
- 11 Dey, A., Tergaonkar, V. & Lane, D. P. Double-edged swords as cancer therapeutics: simultaneously targeting p53 and NF-kappa B pathways. *Nature Reviews Drug Discovery* **7**, 1031-1040 (2008).
- 12 Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* **10**, 130-137 (2010).
- 13 Jenwitheesuk, E., Horst, J. A., Rivas, K. L., Van Voorhis, W. C. & Samudrala, R. Novel paradigms for drug discovery: computational multitarget screening. *Trends in Pharmacological Sciences* **29**, 62-71 (2008).
- 14 Durrant, J. D. *et al.* A Multidimensional Strategy to Detect Polypharmacological Targets in the Absence of Structural and Sequence Homology. *PLoS Computational*

- Biology* **6**, - (2010).
- 15 Zhang, Y. *et al.* Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*. *Science* **325**, 1544-1549 (2009).
- 16 Chen, Y. F. *et al.* SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. *Nucleic Acids Research* **38**, W424-W430, doi:Doi 10.1093/Nar/Gkq480 (2010).
- 17 Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539-547 (1997).
- 18 Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176-180 (1999).
- 19 Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544 (1998).
- 20 Roberts, F. *et al.* Evidence for the shikimate pathway in apicomplexan parasites. *Nature* **393**, 801-805 (1998).
- 21 DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22**, 151-185 (2003).
- 22 Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews. Drug Discovery* **3**, 935-949 (2004).
- 23 Tanrikulu, Y. & Schneider, G. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nature Reviews. Drug Discovery* **7**, 667-677 (2008).
- 24 Holloway, M. K. *et al.* A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *Journal of Medicinal Chemistry* **38**, 305-317 (1995).
- 25 Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267**, 727-748 (1997).
- 26 Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-aided Molecular Design* **8**, 243-256 (1994).
- 27 Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design* **11**, 425-445 (1997).
- 28 Muegge, I. PMF scoring revisited. *Journal of Medicinal Chemistry* **49**, 5895-5902 (2006).
- 29 Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* **295**, 337-356 (2000).

- 30 Wang, R., Lai, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design* **16**, 11-26 (2002).
- 31 Puvanendrapillai, D. & Mitchell, J. B. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. *Bioinformatics* **19**, 1856-1857 (2003).
- 32 Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews. Drug Discovery* **6**, 211-219 (2007).
- 33 Yang, J.-M. & Chen, C.-C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics* **55**, 288-304 (2004).
- 34 Kramer, B., Rarey, M. & Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics* **37**, 228-241 (1999).
- 35 Morris, G. M., Goodsell, D. S., Huey, R. & Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design* **10**, 293-304 (1996).
- 36 Bajorath, J. Integration of virtual and high-throughput screening. *Nature Reviews. Drug discovery* **1**, 882-894 (2002).
- 37 Smith, G. M. *et al.* Positions of His-64 and a bound water in human carbonic anhydrase II upon binding three structurally related inhibitors. *Protein Science* **3**, 118-125 (1994).
- 38 Gutteridge, A. & Thornton, J. M. Understanding nature's catalytic toolkit. *Trends in Biochemical Sciences* **30**, 622-629 (2005).
- 39 Ewing, T. J., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-aided Molecular Design* **15**, 411-428 (2001).
- 40 Yang, J.-M. & Shen, T.-W. A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. *Proteins: Structure, Function, and Bioinformatics* **59**, 205-220 (2005).
- 41 Pan, Y., Huang, N., Cho, S. & MacKerell, A. D., Jr. Consideration of molecular weight during compound selection in virtual target-based database screening. *Journal of Chemical Information and Computer Sciences* **43**, 267-272 (2003).
- 42 Tafi, A. *et al.* Pharmacophore based receptor modeling: the case of adenosine A3 receptor antagonists. An approach to the optimization of protein models. *Journal of Medicinal Chemistry* **49**, 4085-4097 (2006).
- 43 Wolber, G., Seidel, T., Bendix, F. & Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discovery Today* **13**, 23-29 (2008).
- 44 Wiest, O. *et al.* Blueprint for antimicrobial hit discovery targeting metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **107**,

- 1082-1087 (2010).
- 45 Wei, D. *et al.* Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. *Journal of Medicinal Chemistry* **51**, 7882-7888 (2008).
- 46 Chen, J. & Lai, L. H. Pocket v.2: Further developments on receptor-based pharmacophore modeling. *Journal of Chemical Information and Modeling* **46**, 2684-2691 (2006).
- 47 Jensen, R. A. Enzyme recruitment in evolution of new function. *Annual Review of Microbiology* **30**, 409-425 (1976).
- 48 Schapira, M. *et al.* Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7354-7359 (2003).
- 49 An, J. *et al.* A novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. *Journal of Medicinal Chemistry* **52**, 2667-2672 (2009).
- 50 Powers, R. A., Morandi, F. & Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure* **10**, 1013-1023 (2002).
- 51 Yang, J. M., Chen, Y. F., Tu, Y. Y., Yen, K. R. & Yang, Y. L. Combinatorial computational approaches to identify tetracycline derivatives as flavivirus inhibitors. *PLoS ONE*, e428 (2007).
- 52 Champness, J. N. *et al.* Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins* **32**, 350-361 (1998).
- 53 Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927-937 (1998).
- 54 Warnmark, A. *et al.* Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. *Journal of Biological Chemistry* **277**, 21862-21868 (2002).
- 55 Bissantz, C., Folkers, G. & Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry* **43**, 4759-4767 (2000).
- 56 Chin, K. H. *et al.* The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas campestris*. *Journal of Molecular Biology* **396**, 646-662 (2010).
- 57 Yang, M.-C. *et al.* Rational design for crystallization of beta-lactoglobulin and vitamin D-3 complex: revealing a secondary binding site *Crystal Growth & Design* **8**, 4268-4276 (2008).
- 58 Yang, J.-M., Chen, Y.-F., Shen, T.-W., Kristal, B. S. & Hsu, D. F. Consensus scoring criteria for improving enrichment in virtual screening. *Journal of Chemical Information*

- and Modeling* **45**, 1134-1146 (2005).
- 59 Kussmann-Gerber, S., Kuonen, O., Folkers, G., Pilger, B. D. & Scapozza, L. Drug resistance of herpes simplex virus type 1 - structural considerations at the molecular level of the thymidine kinase. *European Journal of Biochemistry* **255**, 472-481 (1998).
- 60 Wild, K., Bohner, T., Folkers, G. & Schulz, G. E. The structures of thymidine kinase from herpes simplex virus type 1 in complex with substrates and a substrate analogue. *Protein Science* **10**, 2097-2106 (1997).
- 61 Zhou, H. B. *et al.* Structure-guided optimization of estrogen receptor binding affinity and antagonist potency of pyrazolopyrimidines with basic side chains. *Journal of Medicinal Chemistry* **50**, 399-403 (2007).
- 62 Maeda, M. The conserved residues of the ligand-binding domains of steroid receptors are located in the core of the molecules. *Journal of Molecular Graphics & Modelling* **19**, 543-551 (2001).
- 63 Evans, J. S. *et al.* Herpesviral thymidine kinases: laxity and resistance by design. *The Journal of General Virology* **79**, 2083-2092 (1998).
- 64 Pilger, B. D. *et al.* Substrate diversity of herpes simplex virus thymidine kinase. Impact Of the kinematics of the enzyme. *The Journal of Biological Chemistry* **274**, 31967-31973 (1999).
- 65 Santhosh, K. C. *et al.* Correlation of anti-HIV activity with anion spacing in a series of cosalane analogues with extended polycarboxylate pharmacophores. *Journal of Medicinal Chemistry* **44**, 703-714 (2001).
- 66 Kuchimanchi, K. R., Ahmed, M. S., Johnston, T. P. & Mitra, A. K. Binding of cosalane - A novel highly lipophilic anti-HIV agent - to albumin and glycoprotein. *J Pharm Sci* **90**, 659-666 (2001).
- 67 Cushman, M. *et al.* Design, Synthesis, and Biological Evaluation of Cosalane, a Novel Anti-Hiv Agent Which Inhibits Multiple Features of Virus Reproduction. *Journal of Medicinal Chemistry* **37**, 3040-3050 (1994).
- 68 Morphy, R., Kay, C. & Rankovic, Z. From magic bullets to designed multiple ligands. *Drug Discovery Today* **9**, 641-651 (2004).
- 69 Boshoff, H. I. M. & Barry, C. E. Tuberculosis - Metabolism and respiration in the absence of growth. *Nat Rev Microbiol* **3**, 70-80 (2005).
- 70 Blanchard, J. S., Hugonnet, J. E., Tremblay, L. W., Boshoff, H. I. & Barry, C. E. Meropenem-Clavulanate Is Effective Against Extensively Drug-Resistant Mycobacterium tuberculosis. *Science* **323**, 1215-1218 (2009).
- 71 Cameron, E. A. B. *et al.* Helicobacter pylori: antibiotic resistance and eradication rates in Suffolk, UK, 1991-2001. *J Med Microbiol* **53**, 535-538 (2004).
- 72 Brown, L. M. Helicobacter pylori: Epidemiology and routes of transmission. *Epidemiol Rev* **22**, 283-297 (2000).
- 73 Hu, L. H. *et al.* Biochemical characterization and inhibitor discovery of shikimate

- dehydrogenase from *Helicobacter pylori*. *Febs J* **273**, 4682-4692 (2006).
- 74 Cheng, W. C., Chang, Y. N. & Wang, W. C. Structural basis for shikimate-binding specificity of *Helicobacter pylori* shikimate kinase. *Journal of Bacteriology* **187**, 8156-8163 (2005).
- 75 Hartmann, M. D., Bourenkov, G. P., Oberschall, A., Strizhov, N. & Bartunik, H. D. Mechanism of phosphoryl transfer catalyzed by shikimate kinase from *Mycobacterium tuberculosis*. *Journal of Molecular Biology* **364**, 411-423 (2006).
- 76 Shindyalov, I. N. & Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* **11**, 739-747 (1998).
- 77 Austin, C. P. *et al.* The NCGC Pharmaceutical Collection: A Comprehensive Resource of Clinically Approved Drugs Enabling Repurposing and Chemical Genomics. *Sci Transl Med* **3** (2011).
- 78 Irwin, J. J. & Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* **45**, 177-182 (2005).
- 79 Medina-Franco, J. L. *et al.* Chemoinformatic Analysis of Combinatorial Libraries, Drugs, Natural Products, and Molecular Libraries Small Molecule Repository. *Journal of Chemical Information and Modeling* **49**, 1010-1024 (2009).
- 80 Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliver Rev* **46**, 3-26 (2001).
- 81 Vederas, J. C. & Li, J. W. H. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **325**, 161-165 (2009).
- 82 Sadowski, J., Gasteiger, J. & Klebe, G. Comparison of automatic three-dimensional model builders using 639 x-ray structures. *Journal of Chemical Information and Computer Sciences* **34**, 1000-1008 (1994).
- 83 Bagautdinov, B. & Kunishima, N. Crystal structures of shikimate dehydrogenase AroE from *Thermus thermophilus* HB8 and its cofactor and substrate complexes: Insights into the enzymatic mechanism. *Journal of Molecular Biology* **373**, 424-438 (2007).
- 84 Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research* **33**, W299-302 (2005).
- 85 Gao, G. F. *et al.* The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. *Nat Struct Mol Biol* **17**, 1266-1268 (2010).
- 86 Liu, Y., Zhang, J. & Xu, W. Recent progress in rational drug design of neuraminidase inhibitors. *Current Medicinal Chemistry* **14**, 2872-2891 (2007).
- 87 Chang, S. C., Cheng, Y. Y. & Shih, S. R. Avian influenza virus: the threat of a pandemic. *Chang Gung Medical Journal* **29**, 130-134 (2006).
- 88 Tanaka, T. *et al.* Safety of neuraminidase inhibitors against novel influenza A (H1N1) in

- pregnant and breastfeeding women. *Canadian Medical Association Journal* **181**, 55-58 (2009).
- 89 Wang, T. & Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *Journal of Medicinal Chemistry* **44**, 961-971 (2001).
- 90 Luo, M. Structural biology: antiviral drugs fit for a purpose. *Nature* **443**, 37-38 (2006).
- 91 Russell, R. J. *et al.* The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **443**, 45-49 (2006).
- 92 Burger, R. A. *et al.* Immunological effects of the orally administered neuraminidase inhibitor oseltamivir in influenza virus-infected and uninfected mice. *Immunopharmacology* **47**, 45-52 (2000).
- 93 Klumpp, K. & Graves, B. J. Optimization of small molecule drugs binding to highly polar target sites: lessons from the discovery and development of neuraminidase inhibitors. *Current Topics in Medicinal Chemistry* **6**, 423-434 (2006).
- 94 van der Vries, E., Stelma, F. F. & Boucher, C. A. B. Emergence of a Multidrug-Resistant Pandemic Influenza A (H1N1) Virus. *New Engl J Med* **363**, 1381-1382 (2010).
- 95 de Jong, M. D. *et al.* Oseltamivir resistance during treatment of influenza A (H5N1) infection. *New Engl J Med* **353**, 2667-2672 (2005).
- 96 Nguyen, H. T., Fry, A. M., Loveless, P. A., Klimov, A. I. & Gubareva, L. V. Recovery of a Multidrug-Resistant Strain of Pandemic Influenza A 2009 (H1N1) Virus Carrying a Dual H275Y/I223R Mutation from a Child after Prolonged Treatment with Oseltamivir. *Clin Infect Dis* **51**, 983-984 (2010).
- 97 Palese, P., Schulman, J. L., Bodo, G. & Meindl, P. Inhibition of influenza and parainfluenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA). *Virology* **59**, 490-498 (1974).
- 98 Xu, X., Zhu, X., Dwek, R. A., Stevens, J. & Wilson, I. A. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *Journal of Virology* **82**, 10493-10501 (2008).
- 99 Rudrawar, S. *et al.* Novel sialic acid derivatives lock open the 150-loop of an influenza A virus group-1 sialidase. *Nature Communications* **1**, 113 (2010).
- 100 Wen, W. H. *et al.* Analogs of zanamivir with modified C4-substituents as the inhibitors against the group-1 neuraminidases of influenza viruses. *Bioorgan Med Chem* **18**, 4074-4084 (2010).
- 101 Yang, J.-M. Development and evaluation of a generic evolutionary method for protein-ligand docking. *Journal of Computational Chemistry* **25**, 843-857 (2004).
- 102 Venkatramani, L. *et al.* An epidemiologically significant epitope of a 1998 human influenza virus neuraminidase forms a highly hydrated interface in the NA-antibody complex. *Journal of Molecular Biology* **356**, 651-663 (2006).
- 103 Chen, C. C., Hwang, J. K. & Yang, J. M. (PS)(2): protein structure prediction server.

- Nucleic Acids Research* **34**, W152-W157 (2006).
- 104 Varghese, J. N. *et al.* Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11808-11812 (1997).
- 105 Hurt, A. C., Holien, J. K. & Barr, I. G. In vitro generation of neuraminidase inhibitor resistance in A(H5N1) influenza viruses. *Antimicrobial Agents and Chemotherapy* **53**, 4433-4440 (2009).
- 106 Collins, P. J. *et al.* Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature* **453**, 1258-1261 (2008).
- 107 Varghese, J. N., Epa, V. C. & Colman, P. M. Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase. *Protein Science* **4**, 1081-1087 (1995).
- 108 Ghate, A. A. & Air, G. M. Site-directed mutagenesis of catalytic residues of influenza virus neuraminidase as an aid to drug design. *European Journal of Biochemistry* **258**, 320-331 (1998).
- 109 Shie, J. J. *et al.* Synthesis of tamiflu and its phosphonate congeners possessing potent anti-influenza activity. *Journal of the American Chemical Society* **129**, 11892-11893 (2007).
- 110 Kim, C. U. *et al.* Structure-activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *Journal of Medicinal Chemistry* **41**, 2451-2460 (1998).
- 111 Hung, H. C. *et al.* Aurintricarboxylic acid inhibits influenza virus neuraminidase. *Antiviral Research* **81**, 123-131 (2009).
- 112 Lew, W. *et al.* A new series of C3-aza carbocyclic influenza neuraminidase inhibitors: synthesis and inhibitory activity. *Bioorganic & Medicinal Chemistry Letters* **8**, 3321-3324 (1998).
- 113 Williams, M. A. *et al.* Structure-activity relationships of carbocyclic influenza neuraminidase inhibitors. *Bioorganic & Medicinal Chemistry Letters* **7**, 1837-1842 (1997).
- 114 Atigadda, V. R. *et al.* Hydrophobic benzoic acids as inhibitors of influenza neuraminidase. *Bioorgan Med Chem* **7**, 2487-2497 (1999).
- 115 Gubareva, L. V., Kaiser, L., Matrosovich, M. N., Soo-Hoo, Y. & Hayden, F. G. Selection of influenza virus mutants in experimentally infected volunteers treated with oseltamivir. *The Journal of Infectious Diseases* **183**, 523-531 (2001).
- 116 Meindl, P., Bodo, G., Palese, P., Schulman, J. & Tuppy, H. Inhibition of neuraminidase activity by derivatives of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid. *Virology* **58**, 457-463 (1974).
- 117 Babu, Y. S. *et al.* BCX-1812 (RWJ-270201): Discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *Journal of Medicinal Chemistry* **43**, 3482-3486 (2000).

- 118 Chand, P. *et al.* Systematic structure-based design and stereoselective synthesis of novel multisubstituted cyclopentane derivatives with potent antiinfluenza activity. *Journal of Medicinal Chemistry* **44**, 4379-4392 (2001).
- 119 Joseph-McCarthy, D., Baber, J. C., Feyfant, E., Thompson, D. C. & Humblet, C. Lead optimization via high-throughput molecular docking. *Curr Opin Drug Disc* **10**, 264-274 (2007).
- 120 Michel, J., Verdonk, M. L. & Essex, J. W. Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization? *Journal of Medicinal Chemistry* **49**, 7427-7439 (2006).
- 121 Liu, Q., Masek, B., Smith, K. & Smith, J. Tagged fragment method for evolutionary structure-based de novo lead generation and optimization. *Journal of Medicinal Chemistry* **50**, 5392-5402 (2007).
- 122 Woo, L. W. *et al.* Highly potent first examples of dual aromatase-steroid sulfatase inhibitors based on a biphenyl template. *Journal of Medicinal Chemistry* **53**, 2155-2170 (2010).
- 123 Hunter, R. L. & Markert, C. L. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* **125**, 1294-1295 (1957).
- 124 Bugatti, A. *et al.* Heparin-mimicking sulfonic acid polymers as multitarget inhibitors of human immunodeficiency virus type 1 Tat and gp120 proteins. *Antimicrobial Agents and Chemotherapy* **51**, 2337-2345 (2007).
- 125 Harmange, J. C. *et al.* Naphthamides as novel and potent vascular endothelial growth factor receptor tyrosine kinase inhibitors: Design, synthesis, and evaluations. *Journal of Medicinal Chemistry* **51**, 1649-1667 (2008).
- 126 Mol, C. D. *et al.* Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *Journal of Biological Chemistry* **279**, 31655-31663 (2004).
- 127 Cowan-Jacob, S. W. *et al.* Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr D* **63**, 80-93 (2007).