計 畫 編 號： NHRI-EX100-10009PI

國家衛生研究院整合性醫藥衛生科技研究計畫

分子間藥理作用介面家族應用在磷酸化酵素-藥物-疾病網絡與機制之研究

計畫名稱

100年度成果報告

執 行 機 構： 國立交通大學

計 畫 主 持 人： 楊進木 教授

本年度執行期間： 100 年 1 月 1 日 至 100 年 12 月 31 日

全文處理方式： 二年後對外提供參考

＊＊本研究報告僅供參考用，不代表本院意見＊＊

計畫編號：NHRI-EX100-10009PI

國家衛生研究院整合性醫藥衛生科技研究計畫

分子間藥理作用介面家族應用在磷酸化酵素−藥物−疾病網絡與機制之研究

100年度成果報告

執 行 機 構： 國立交通大學

計 畫 主 持 人： 楊進木 教授

本年度執行期間： 100 年 1 月 1 日 至 100 年 12 月 31 日

全文處理方式： 二年後對外提供參考

目錄

壹、100年度計畫研究成果摘要

計 畫 名 稱：分子間藥理作用介面家族應用在磷酸化酵素-藥物-疾病網
　　　　　　　絡與機制之研究

計 畫 編 號：NHRI-EX100-10009PI

執 行 機 構：國立交通大學

計 畫 主 持 人：楊進木

研 究 人 員：許凱程、林志達、林伸融、尤宣人

關 鍵 字：分子間藥理作用介面家族、舊藥新用、多標的藥物、磷酸
　　　　　　化酵素、癌症治療、副作用

成 果 分 類：　☑ 癌症基礎與臨床研究(可複選，最多三項)

　　　　　　　　□ 分子與基因醫學研究

　　　　　　　　□ 臨床研究

　　　　　　　　☑ 生物技術與藥物研究

　　　　　　　　☑ 生物統計與生物資訊研究

　　　　　　　　□ 醫療保健政策研究

　　　　　　　　□ 環境衛生與職業醫學研究

　　　　　　　　□ 醫學工程研究

　　　　　　　　□ 老年醫學研究

　　　　　　　　□ 精神醫學與藥物濫用研究

　　　　　　　　□ 疫苗研究

　　　　　　　　□ 幹細胞研究

　　　　　　　　□ 奈米醫學研究

　　　　　　　　□ 其他重要疾病或醫藥衛生問題研究

摘要

目前的藥物開發流程因其成本昂貴且需時極長，已逐漸不能滿足市場對新藥物的需求。再者，過去「單一藥物針對單一標的，治療單一疾病」的思維，在治療病理現象複雜的疾病(如癌症)上也遭遇諸多瓶頸。針對上述議題，本計劃的主要目標在於利用所提出之新概念「分子間藥理作用介面家族(molecular pharma-interface family)」以及新方法「區域官能基地圖(site-moiety map, SiMMap)」，希望達到三項目標：（一）為上市藥物與中草藥尋找新用途（舊藥新用）。（二）發展多標靶藥物(multi-target drugs)。（三）建立國人常見癌症之磷酸化酵素-藥物-疾病關聯性網絡資料庫，希望能藉此達到加速藥物發展及增進治療複雜疾病之效果。在本期（第一年）中我們已將相關成果發表於*Biochemical Pharmacology*（影響係數4.9）與*Nucleic Acids Research*（影響係數7.8），並即將再發表一篇期刊論文於*PLoS ONE*（影響係數4.4）。

本年度為五年期計畫之第一年，我們完成了以下事項：（一）針對147個具有結晶結構之磷酸化酵素、1,216個上市藥物、89,425個天然物及330,990個化合物建立「分子間藥理作用介面家族」。(二)利用此介面家族，我們發現數十個潛力新型抑制劑及多標靶藥物(包含現有上市藥物之舊藥新用及天然物)。(三)建立磷酸化酵素－藥物－疾病關聯性網絡資料庫，包含147種具

有結晶結構的磷酸化酵素、1,216個藥物與超過3萬個磷酸化酵素抑制劑、79種OMIM表型(Phenotype)。此外，我們挑選了64個潛力抑制劑及50種磷酸化酵素進行大規模磷酸化酵素－化合物之活性分析(3,200 spots)，以驗證及精進我們模型。

我們也應用分子間藥理作用介面家族於瞭解藥物作用機制、尋找新型抗生素及潛在標靶蛋白質。我們與交通大學趙瑞益教授合作，建立化合物PT-262對ROCK1的可能作用機制（已於2011年發表於期刊*Biochemical Pharmacology*)；與清華大學王雯靜教授合作，已找出三個μM等級之新型多標靶(multi-target)抗生素，可有效抑制幽門螺旋桿菌中的莽草酸磷酸化酵素(Shikimate Kinase)和莽草酸去氫化酵素(shikimate dehydrogenase) (審核於期刊*PLoS ONE*)。與義守大學袁行修醫師合作，替天然物-大金星蕨萃取物protoapigenone找出潛在標靶蛋白質-PI3K四種催化次單元（catalytic subunits）(準備投稿中)。此外，我們利用分子間藥理作用介面家族的概念建立以結構為基礎之蛋白質-胜肽抗原作用介面家族，能大規模搜尋病原體基因組數據庫找出潛在胜肽抗原(已於2011年發表於期刊*Nucleic Acids Research*)。

上述的研究成果已達到計畫今年之預期目標，初步的成果證實本計畫所提出的分子間藥理作用介面家族能有效地尋找新型抑制劑及了解藥物作用

機制。下年度的工作構想，主要分為兩大部分：(一)藉由大規模磷酸化酵素－化合物活性分析實驗資料，修正分子間藥理作用介面作用家族之計算模型，並探討普遍性 (general inhibitors)和專一性 (specific inhibitors)抑制劑之機制，進而設計具高專一性之藥物。(二) 擴增磷酸化酵素－藥物－疾病關聯性網絡資料庫，預期囊括全部 518 種已知之人類磷酸化酵素、與約一百萬化合物、與磷酸化酵素相關之 3,300 種 OMIM (Online Mendelian Inheritance in Man)基因遺傳疾病表型。我們相信當這套技術與成果，將對疾病治療、降低藥物副作用及藥物成本有所貢獻。

Abstract

The current framework of drug development is cost high and time-consuming and is unable to afford the requirements for new market drugs and emerging diseases. In addition, the "one drug for one target for one disease" drug-discovery strategy cannot fully solve complex diseases, such as cancer and diabetes. Therefore, developing a new concept to discover new uses for old drugs and new multi-target inhibitors would be useful for drug development. The new uses for old drugs can be fast evaluated by clinical trials and the multi-target inhibitors can an effective approach for complex diseases through inhibiting multiple targets. In this project, we propose a novel concept "molecular pharma-interface family" and Site-Moiety Map (SiMMap) to achieve the three goals: (1) finding new uses for old drugs and natural products, (2) discovering multi-target drugs, and (3) developing a kinase-drug-cancer (disease) network database. We believe that the new concept and SiMMap should accelerate drug development process and be useful for the treatment of complex diseases.

In the first year, we have achieved some preliminary results and published three papers in *Biochemical Pharmacology*, *Nucleic Acids Research*, and *PLoS ONE* (under revision on Impact factor: 4.4). We have achieved the following preliminary results. First, we have established molecular pharma-interface families by using 147 protein kinases with crystal structures, 1,216 marked drugs, 89,425 natural products, and 330,990 chemical compounds. Second, we have identified dozens of potential inhibitors and multi-target drugs from marked drugs and natural products by using the families. Third, we have created a preliminary kinase-drug-disease network database, which composes 147 protein kinases, 1,216 marked drugs, ~30,000 kinase inhibitors, and 79 OMIM

(Online Mendelian Inheritance in Man) phenotypes. Finally, we will perform large-scale kinase profiling on 64 selected potential inhibitors and 50 kinases to verify our model.

We have applied the molecular pharma-interface families to understand the mechanisms of drug action, identify multi-target antibiotics, and discover potential targets. We identified the binding mechanism of a ROCK1 inhibitor (PT-262) by cooperating with Professor Chao Jui-I, NCTU (published on *Biochemical Pharmacology*, 2011). Furthermore, we identified three novel multi-target antibiotics in μM level against shikimate kinase and shikimate dehydrogenase of *Helicobacter pylori* by collaborating with Professor Wang Wen-Ching, NTHU (revised in *PLoS ONE*). With the cooperation of Professor Yuan Shyng-Shiou, ISU, four catalytic subunits of PI3K were predicted as the potential target of protoapigenone. In addition, we applied molecular pharma-interface family to construct the structure-based "Peptide antigen family", which can identify the potential peptide antigens from the complete pathogen genome database (published on *Nucleic Acids Research*, 2011).

In summary, we have achieved the anticipated results of the first year, and the results reveal that the proposed molecular pharma-interface families are useful to discover inhibitors and reveal binding mechanisms. In the following year, we will verify and refine our computational models by analyzing results of large-scale kinase profiling assays. In addition, we will focus on the design of specific inhibitors, which can reduce side effects. Second, we will extend the current kinase-drug-disease network database by collecting 518 protein kinases, ~1,000,000 chemical compounds, and 3,300 OMIM disease phenotypes related to kinases. We believe that our kinase-drug-disease network database and the molecular pharma-interface families are useful for treatment of diseases, reducing side effects, and enhancing human health.

貳、100年度計畫著作一覽表

Journal

序號	計畫產出名稱	產出型式	Impact factor	致謝對象
1	K.-C. Hsu, W.-C. Cheng, Y.-F. Chen, H.-J. Wang, L.-T. Li, W.-C. Wang and J.-M. Yang Core site-moiety maps reveal inhibitors and binding mechanisms of orthologous proteins by screening compound libraries. PLoS One 2011; (SCI) Submitted	Foreign	4.411	NHRI
2	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 2011;856-865. (SCI) Published	Foreign	4.889	
3	I-H. Liu, Y.-S. Lo, and J.-M. Yang PAComplex: a web server to infer peptide antigen families and binding models from TCR-pMHC complexes. Nucleic Acids Research 2011;254-260. (SCI) Published	Foreign	7.836	NHRI

Patent

序號	計畫產出名稱
	無

Book

序號	計畫產出名稱
	無

Conference Paper

序號	計畫產出名稱
	無

Technical Report

序號	計畫產出名稱
	無

參、100年度計畫重要研究成果產出統計表

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

（係指執行100年度計畫之所有研究產出結果）

科技論文篇數			技術移轉			技術報告		0 項
發表地點 類　　型	國　內	國　外	類　　型	經　　費	項　　數	技術創新		0 項
期　刊 論　文	0篇	3篇	技　術 輸　入	0千元	0項	技術服務		0 項
研討會 論　文	0篇	0篇	技　術 輸　出	0千元	0項	專利權	國內	0 項
							國外	0 項
專　著	0篇	0篇	技　術 擴　散	0千元	0項	著作權	國內	0 項
							國外	0 項

〔註〕：

期刊論文：指在學術性期刊上刊登之文章，其本文部份一般包含引言、方法、結果、及討論，並且一定有參考文獻部份，未在學術性期刊上刊登之文章（研究報告等）與博士或碩士論文，則不包括在內。

研討會論文：指參加學術性會議所發表之論文，且尚未在學術性期刊上發表者。

專　　著：為對某項學術進行專門性探討之純學術性作品。

技術報告：指從事某項技術之創新、設計及製程等研究發展活動所獲致的技術性報告且未公開發表者。

技術移轉：指技術由某個單位被另一個單位所擁有的過程。我國目前之技術轉移包括下列三項：一、技術輸入。二、技術輸出。三、技術擴散。

技術輸入：籍僑外投資、與外國技術合作、投資國外高科技事業等方式取得先進之技術引進國內者。

技術輸出：指直接供應國外買主具生產能力之應用技術、設計、顧問服務及專利等。我國技術輸出方包括整廠輸出、對外投資、對外技術合作及顧問服務等四種。

技術擴散：指政府引導式的技術移轉方式，即由財團法人、國營事業或政府研究機構將其開發之技術擴散至民間企業之一種單向移轉（政府移轉民間）。

技術創新：指研究執行中產生的技術，且有詳實技術資料文件者。

技術服務：凡有關各項研究計畫之規劃與評審、技術督察與指導及專業技術服務事項等。

肆、100年度計畫重要研究成果

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

（一）計畫之新發現、新發明或對學術界、產業界具衝擊性(impact)之研究成果，請依性質勾選下列項目。

☑1.研發或改良國人重要疾病及癌症的早期診斷方式及治療技術

☐2.發展新的臨床治療方式

☑3.發展新生物製劑、篩檢試劑及新藥品

☐4.瞭解常見疾病及癌症之分子遺傳機轉

☑5.瞭解抗癌藥劑對癌細胞之作用機制

☐6.提供有效的疾病預防策略

☑7.利用生物統計與生物資訊研究，推動台灣生技醫藥研究，促進生物技術與基因體醫學之發展

☐8.醫療保健政策相關研究

☐9.瞭解環境毒理機制及重金屬對人體健康的影響

☐10.研發適合臨床使用的人造器官及生醫材料

☐11.縮短復健流程並增加復健效果的醫療輔助方式或器材之研究應用

☐12.改進現有醫療器材的功能或增加檢驗影像的解析能力

☐13.其他重要疾病或醫藥衛生問題研究

（二）計畫研究成果技術移轉情形

☐即將進行洽談或洽談中

☐已技轉

☑無

請概述 (1)技術或產品名稱(2)內容說明(3)產業別(4)應用範圍(5)技術移轉可行性及預期效益

註：1.如研究成果尚未申請專利，請勿揭露可申請專利之主要內容

2.如計畫研究成果沒有技術移轉，請寫「無」

無

肆、100 年度計畫重要研究成果

　　本研究團隊為全世界第一個提出且建立「分子間藥理作用介面家族」的團隊，並且完成建置第一個國人常見癌症的「磷酸化酵素－藥物－疾病關聯性網絡資料庫」。此兩項研究成果可篩選出高療效、低抗藥性的新型化合物，將有助於舊藥新用、開發新藥，了解藥物作用機制，及為國人健康提供重要福祉。下面將詳述我們在本年度計畫之研究成果、成效評估、未來一年的工作構想及展望。

（1）計畫之新發現、新發明或對學術界、產業界具衝擊性之研究成果

　　本團隊所提出的區域官能基地圖(SiMMap)[1]對於開發新藥及評估藥物副作用有重大貢獻。目前本技術使用人次已超過 2,500 人，遍及 46 個國家；從 2010 年發表至今，已有兩篇論文應用本技術且成果分別發表於 PLoS ONE[2] 和 Journal of Molecular Modeling[3] (圖一)。而另一套由本研究團隊發展出的高準確率分子嵌合工具 GEMDOCK[4-6]，下載人次則高達 3,477 人。

Fig. 5 The site-moiety map anchors in the binding pocket of EGFR-TK that interact with the eight candidate compounds
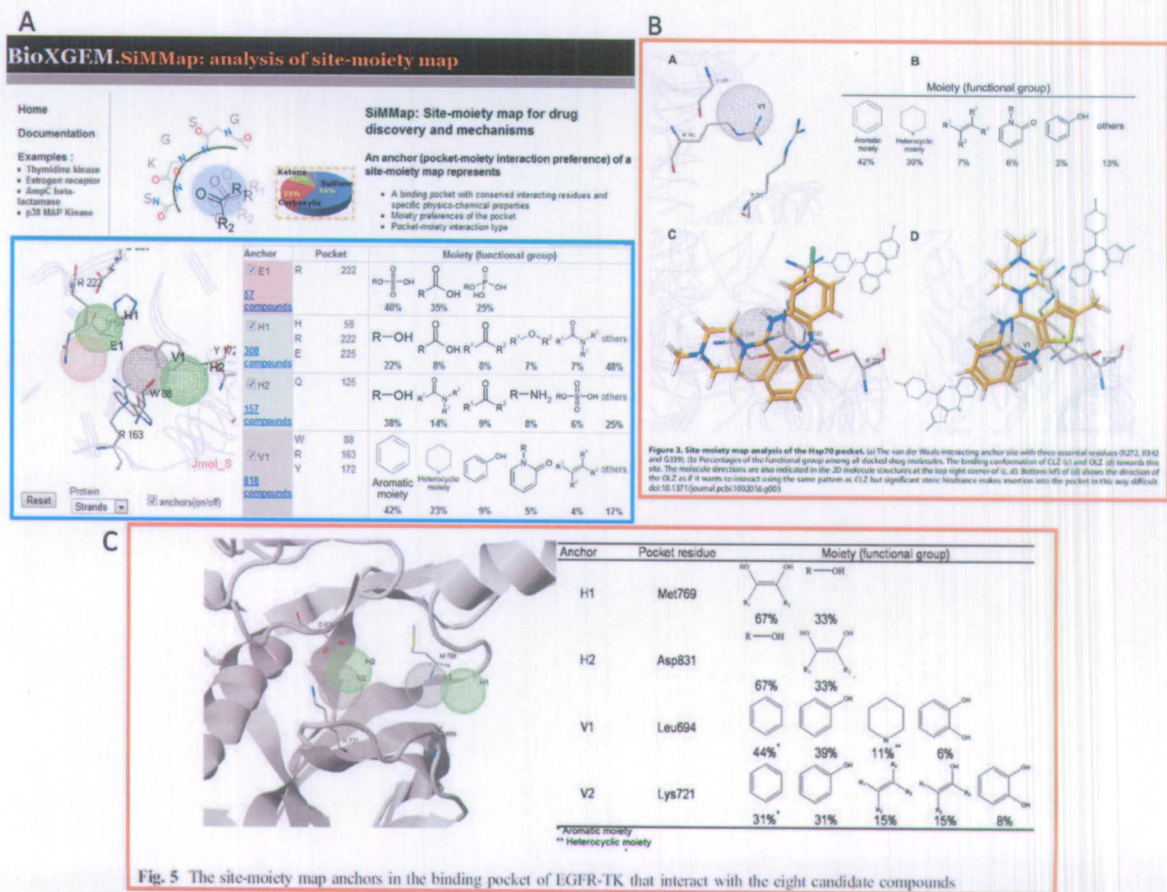
圖一、(A)本團隊提出的區域官能基地圖(SiMMap)。(B)發表於 *PLoS ONE* 的論文直接引用本團隊所提供之網頁服務圖片 [2]。(C)發表於 *Journal of Molecular Modeling* 的論文直接引用本團隊所提供之網頁服務圖片 [3]。

　　根據衛生署公佈 2010 年國人的十大死因,第一名為惡性腫瘤,已蟬聯長達 29 年。而目前治療癌症之普遍問題為療效不如預期、副作用的傷害,或抗藥性的產生,其主要原因為癌症通常牽涉複雜的生化代謝途徑,過去「單一藥物針對單一標的,治療單一疾病」的思維已逐漸不適合用於治療癌症。此外,單一標靶藥物容易因蛋白質突變產生抗藥性而造成藥效下降甚至喪失 [7,8]。若能設計針對多個標靶蛋白質的藥物將有效提高療效與降低

抗藥性，然而多標靶往往也可能伴隨著更多副作用的發生，如何克服上述問題將是當前藥物設計的難題。

　　癌症的發生主要原因之一為人體內重要生化途徑上的磷酸化酵素(kinases)失調，本團隊透過我們所發展的新方法「區域官能基地圖(site-moiety map, SiMMap)」成功建立分子間藥理作用介面家族(圖二 A 與 B)。透過家族成員會具有相同結合機制的特性，同一家族中的磷酸化酵素應具有相似分子間藥理作用(圖二 C 與 D)。利用這些磷酸化酵素的分子間藥理作用介面家族，將可以簡單且快速查詢哪些磷酸化酵素具有相似分子間藥理作用介面，因此，有助於針對特定的疾病設計出多標靶的藥物來治療，例如圖二 B 中 MAPK8，MAPK10 都和結腸癌(Colorectal cancer)有關，可以根據這兩個磷酸化酵素的分子間藥理作用介面設計專屬抑制劑。此外，由於是專一抑制這兩個磷酸化酵素，而對其他磷酸化酵素無抑制效果，不僅可以提高藥物療效，更可以降低副作用。此外，我們還可以快速從現有藥物資料庫及天然物資料庫篩選出其他潛在治療藥物(圖二 C 與 D)，不僅大幅降低了新藥開發的時間，也節省了大量的實驗成本。
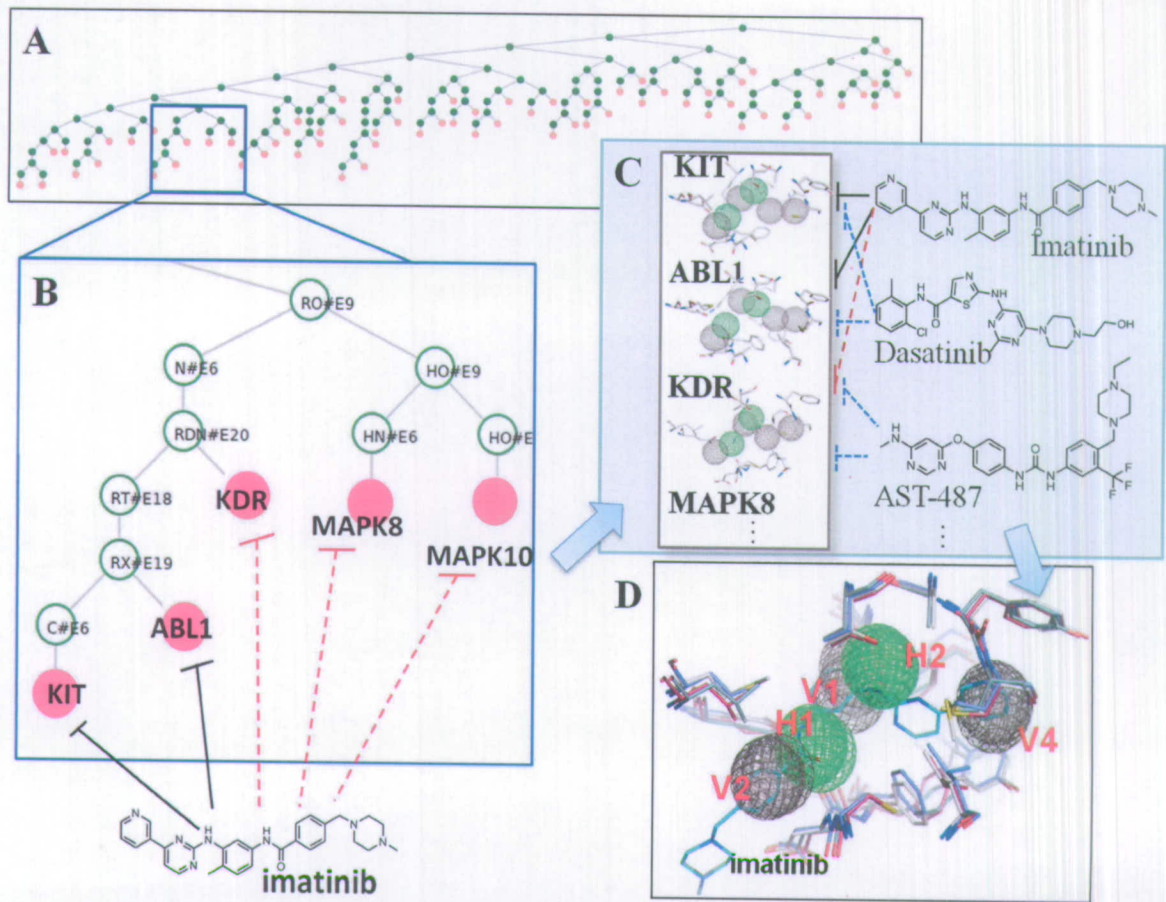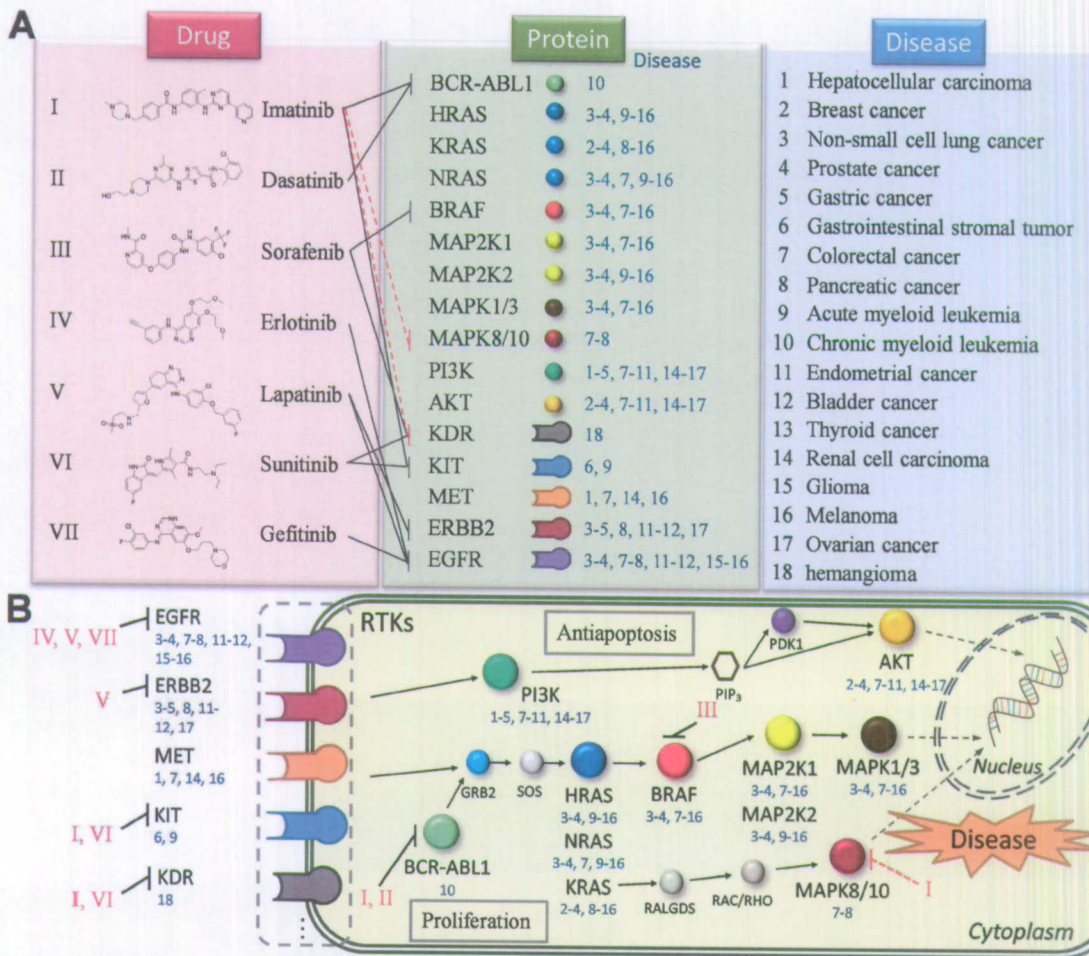
圖二、概述磷酸化酵素的分子間藥理作用介面家族。(A) 分子間藥理作用介面家族，紅色端點為人類磷酸化酵素(147 種)，綠色節點為分子間藥理作用介面之關鍵性位置(例如藥效基團位置)，同一家族之蛋白質具有相似之分子藥理作用介面。(B) 同一藥理作用介面家族之例子。已知 Imatinib 可抑制 KIT 與 ABL1(黑色線條)，而 KDR、MAPK8 及 MAPK10 與 KIT、ABL1 歸屬同一分子藥理作用介面家族，故能推論 Imatinib 亦能抑制 MAPK8、MAPK10 與 KDR (紅色線條)。(C) Imatinib 與 Dasatinib、AST-487 具有相似之分子藥理作用介面，則能預測 Dasatinib 及 AST-487 亦可抑制此一家族中的磷酸化酵素(藍色線條)。(D) 擁有相似分子間藥理作用介面的磷酸化酵素與多標靶藥物 Imatinib 之分子交互作用模型。

本團隊透過分子間藥理作用介面家族及結合 GEMDOCK、SiMMap、系

統生物等技術,成功建立第一個國人常見癌症的磷酸化酵素—藥物—疾病

關聯性網絡資料庫。本資料庫包含三大部分(磷酸化酵素、磷酸化酵素相關

藥物、磷酸化酵素相關疾病)及其關聯性網絡,收錄大量相關資料庫資料(例

如 DrugBank, BindingDB, KEGG, kinase.com, and OMIM)。當我們將這套觀

念運用在分析國內社會常見癌症(如肺癌、肝癌、腸癌及胰臟癌),初步成果

顯示上市藥物 Imatinib 除了能抑制 KIT 和 ABL1 磷酸化酵素而達到治療胃

腸道基質瘤(gastrointestinal stromal tumor)和慢性骨髓細胞白血病(chronic

myeloid leukemia)的效果之外,MAPK8(腸癌、胰臟癌)、MAPK10(腸癌、胰

臟癌)、KDR(血管增生,與肝癌及非小細胞肺癌有關)也具有可與 KIT 及

ABL1 歸屬於同一藥理作用介面家族的配體結合位置;換言之,Imatinib 也

可能抑制 MAPK8、MAPK10 及 KDR(圖三之紅色虛線),這一點在比對前人

發表的實驗論文後得到了證實[9-12]。此外,透過藥理作用介面家族與此資料

庫之分析,我們已從現有藥物及天然物中發現數十個潛力抑制劑,能抑制

與國人常見疾病相關之磷酸化酵素。我們相信此資料庫將對尋找出潛在傳

統單一標靶、新型多標靶抑制藥物與降低藥物副作用有極大貢獻。

A

Drug

I		Imatinib
II		Dasatinib
III		Sorafenib
IV		Erlotinib
V		Lapatinib
VI		Sunitinib
VII		Gefitinib

Protein — Disease

Protein	Disease
BCR-ABL1	10
HRAS	3-4, 9-16
KRAS	2-4, 8-16
NRAS	3-4, 7, 9-16
BRAF	3-4, 7-16
MAP2K1	3-4, 7-16
MAP2K2	3-4, 9-16
MAPK1/3	3-4, 7-16
MAPK8/10	7-8
PI3K	1-5, 7-11, 14-17
AKT	2-4, 7-11, 14-17
KDR	18
KIT	6, 9
MET	1, 7, 14, 16
ERBB2	3-5, 8, 11-12, 17
EGFR	3-4, 7-8, 11-12, 15-16

Disease

1 Hepatocellular carcinoma
2 Breast cancer
3 Non-small cell lung cancer
4 Prostate cancer
5 Gastric cancer
6 Gastrointestinal stromal tumor
7 Colorectal cancer
8 Pancreatic cancer
9 Acute myeloid leukemia
10 Chronic myeloid leukemia
11 Endometrial cancer
12 Bladder cancer
13 Thyroid cancer
14 Renal cell carcinoma
15 Glioma
16 Melanoma
17 Ovarian cancer
18 hemangioma

B

RTKs

IV, V, VII — EGFR 3-4, 7-8, 11-12, 15-16
V — ERBB2 3-5, 8, 11-12, 17
MET 1, 7, 14, 16
I, VI — KIT 6, 9
I, VI — KDR 18

Antiapoptosis — PDK1 — AKT 2-4, 7-11, 14-17 — Nucleus
PI3K 1-5, 7-11, 14-17 — PIP$_3$ — III
GRB2 — SOS — HRAS 3-4, 9-16 — BRAF 3-4, 7-16 — MAP2K1 3-4, 7-16 / MAP2K2 3-4, 9-16 — MAPK1/3 3-4, 7-16
NRAS 3-4, 7, 9-16
BCR-ABL1 10 — I, II
KRAS 2-4, 8-16 — RALGDS — RAC/RHO — MAPK8/10 7-8 — I
Proliferation — Disease — Cytoplasm

圖三、磷酸化酵素-藥物-疾病關聯性網絡資料庫部分資料。(A)磷酸化酵素─藥物─疾病三大子資料庫。圖中羅馬數字為7個獲得美國食品藥物管理局許可之磷酸化酵素藥物(紅色區塊部分),這些藥物所抑制的目標蛋白質參與數種癌症中(例如:肝癌及乳癌);阿拉伯數字為18種重要癌症(藍色區塊部分),並各自註解於相關的蛋白質(綠色區塊部分;藍色字為註解)。紅色虛線為本資料庫推測之Imatinib可能的抑制目標(例如:MAPK8、MAPK10及KDR)。(B)磷酸化酵素─藥物─疾病關聯性網絡。

(2)計畫對民眾具教育宣導之研究成果

　　生醫科技在後基因體時代是國家科技發展的重點項目之一,其中藥物開發更是最具潛力的主流。目前的藥物開發流程因其成本昂貴且需時極長,

已逐漸不能滿足市場對新藥物的需求和應付突發疫情。我們是國內少數有能力自行研發藥物設計技術的團隊，本計畫提出之「分子間藥理作用介面家族」，透過同家族的蛋白質－藥物具有相同結合特性，可應用於舊藥新用且快速找出潛在藥物並提出藥理機制。運用本概念探討國內常見癌症如肺癌、肝癌、腸癌及胰臟癌時，初步成果證實與先前研究相符。本研究成果顯示在舊藥新用、新藥探索、中草藥及生技藥品的研發具有潛力，我們相信值得後續研究與探討，將對疾病治療、降低藥物副作用與成本以及增進國民健康有所貢獻。

（3）簡述年度計畫成果之討論與結論，如有技術轉移、技術推廣或產業合作

今年底將進行「大規模蛋白質磷酸化酵素－化合物之活性分析」，期望應用「分子間藥理作用介面家族」能尋找出潛在傳統單一標靶與新型多標靶抑制劑，預計今年底將取得送測之結果(共 3,200 點實驗結果，50 種磷酸化酵素對 64 個化合物)。目前已收集 1,216 個 FDA 認證藥物和 89,425 個天然物及超過 33 萬個來自現有公開化合物資料庫的小分子化合物，經由本團隊所建立之分子間藥理作用介面家族篩選後，選出 64 個具有潛在抑制蛋白質磷酸化酵素能力的化合物，及 50 種分散在不同結構特性及疾病類型之蛋白質磷酸化酵素，例如與胃腸道基質瘤相關之酪胺酸激酶受體 (KIT)，與腸

癌和胰臟癌相關之有絲分裂活化蛋白質激酶 8 (MAPK8)。圖四為大規模蛋

白質磷酸化酵素—化合物之活性分析流程圖，根據預測結果現有藥物

Sidenafil 為單一標靶抑制劑，其僅抑制單一蛋白質磷酸化酵素；而現有藥物

Penetrex 與天然物 Quercetin 為多標靶抑制劑。此部分之實驗結果將於年底

取得。
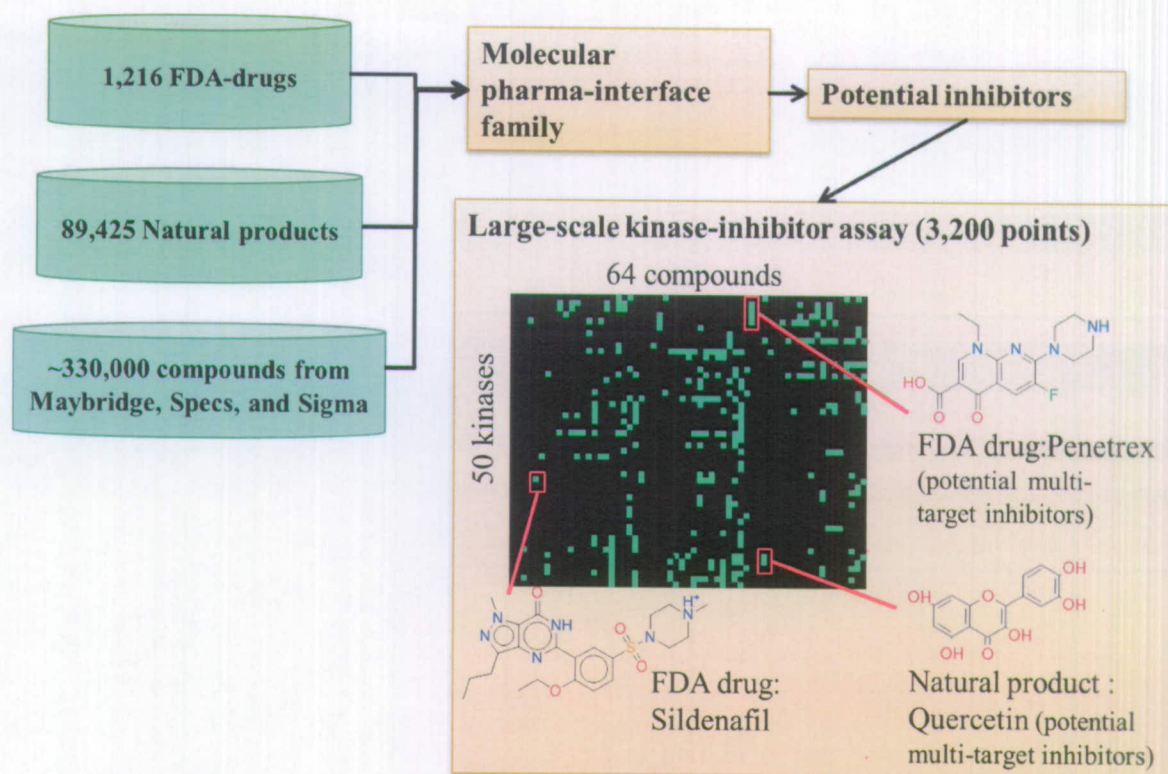
圖四、應用分子間藥理作用介面家族篩選潛在新型單一標靶與新型多標靶抑制劑之流程
圖。

我們也應用分子間藥理作用介面家族來尋找新型抑制劑及瞭解藥物作

用機制。我們與交通大學趙瑞益教授合作，建立了化合物 PT-262 對 ROCK1

的可能作用機制 [13]。ROCK1 為 Serine/Threonine protein kinase，在細胞移行、

增生和存活上扮演相當重要的角色。實驗結果顯示，PT-262 對 ROCK1 的抑制效果較已知 ROCK1 抑制劑 Y-27632 與 H-1152 有效，且可誘發肺癌細胞株 A549 細胞骨架的重塑（*cytoskeleton remodeling*）與移行的抑制（migration inhibition）(已於 2011 年發表於期刊 *Biochemical Pharmacology*)。

我們與清華大學王雯靜教授合作，已成功找出三個 μM 等級之新型多標靶(multi-target)抗生素，可有效抑制結核桿菌中的莽草酸磷酸化酵素(Shikimate Kinase)以及幽門螺旋桿菌中的莽草酸磷酸化酵素和莽草酸去氫化酵素。(已投稿於期刊 Plus One，目前正在審核階段)。

我們與義守大學袁行修醫師合作，發現萃取自常用於食用與民俗療法上之台灣原生植物-大金星蕨的新穎類黃酮 protoapigenone，有機會作用於 PI3K 四種催化次單元（catalytic subunits）。實驗結果亦與我們的預測一致：protoapigenone 可抑制子宮頸癌細胞株 HeLa 的生長（growth）、侵襲（invasion）與移行（migration）及誘發細胞凋亡（apoptosis），而這些現象皆與 PI3K 及其下游的調控有關(準備投稿中)。

此外，透過作用介面家族概念，我們亦建立以結構為基礎之蛋白質－胜肽抗原作用介面家族(Peptide antigen family)[14]，大規模搜尋病原體完整基因組數據庫後找出同源生態抗原家族，其為可能引發相同免疫反應之潛在胜

肽抗原。經由本方法提供之蛋白質-胜肽交互作用模型來解釋結合情形,進而探討同家族之免疫特異性(已於 2011 年發表於期刊 *Nucleic Acids Research*)。未來我們將延伸此蛋白質-胜肽家族理論,尋找磷酸化酵素與胜肽間的普遍性或專一性結合,以期剖析疾病發生時磷酸化酵素與蛋白質的關係,並運用於蛋白質藥物的開發。

(4)成效評估(技術面、經濟面、社會面、整合綜效)

技術面:

本研究提出一個新概念「分子間藥理作用界面作用家族 (molecular pharma-interface family)」,且結合本實驗室發展的分子嵌合工具 GEMDOCK 和「區域官能基地圖 (site-moiety map)」,用來分析和定位蛋白質-配體結合介面之藥效基團,瞭解蛋白質-配體之交互作用機制,與快速尋找藥物標的蛋白質。結合上述新概念和生物資訊相關技術,本研究可以增進治療國人重要疾病及癌症的新藥品之研發。

經濟面:

目前的藥物開發流程因其成本昂貴且需時極長,已逐漸不能滿足市場對新藥物的需求和應付緊急疫情。藉由本研究所發展之新概念「分子間藥理作用界面作用家族 (molecular pharma-interface family)」,可加快藥物開發速

度和增加虛擬藥物篩選的準確性，大幅降低尋找前導化合物(lead compound)所耗費的成本。

社會面：

目前市面上的抗癌藥物，因病患的個體差異會產生不同的副作用，甚至導致死亡。因此當療程中藥物對人體發生毒性時，則需重新評估療程，包括改變劑量、延遲治療或停藥。藉由磷酸化酵素-藥物-疾病關聯性網絡資料庫的建立，我們得以瞭解磷酸化酵素有關的藥物機制，如抗癌藥劑對於癌細胞之作用機制與其副作用。本研究不僅增進藥物開發速度，亦可協助評估臨床上藥物使用的副作用，因此對於減緩、治癒病患症狀，保持療程效果具有正面幫助。

整合綜效：

整體而言，本研究結合虛擬藥物篩選、分析藥效基團、及藥物影響路徑的整合性研究，透過建立磷酸化酵素—藥物—疾病關聯性網絡資料庫，對於瞭解藥物的治療機制、檢視藥物副作用、提供台灣生技醫藥研發技術、及降低藥物成本有所貢獻。

（５）下年度工作構想及重點之妥適性

目前，本研究藉由 147 個具有結晶結構的磷酸化酵素，初步建立分子間藥理作用界面作用家族 (molecular pharma-interface family)，並挑選 50 個磷酸化酵素與 64 個潛在抑制劑進行大規模磷酸化酵素－化合物活性分析 (Large-scale kinase profiling)。這部分工作，經由公開招標後，由默克-密理博藥廠（Merck Millipore）檢測。

下年度的工作構想，主要分為兩大部分，以進一步驗證分子間藥理作用界面作用家族和磷酸化酵素-藥物-疾病關聯性網絡資料庫。其一，改進且修正分子間藥理作用界面作用家族。藉由大規模磷酸化酵素－化合物活性分析資料，我們將繼續探討蛋白質與配體交互作用介面之分子機轉，進而修正分子間藥理作用界面作用家族之模型。此外，藉由大規模之實驗資料我們能夠探討普遍性 (general inhibitors)和專一性 (specific inhibitors)抑制劑之機制，進而設計具高專一性之藥物。

其二，擴增磷酸化酵素－藥物－疾病關聯性網絡資料庫。目前所建立之磷酸化酵素-藥物-疾病關聯性網絡資料庫，包含 147 種磷酸化酵素、1,216 個藥物、BindingDB 資料庫中超過 3 萬個已知之磷酸化酵素抑制劑、79 種 OMIM 基因遺傳疾病表型。下年度將陸續匯集擴充各項資料庫，預期囊括全部 518 種已知之人類磷酸化酵素、約一百萬個化合物、與磷酸化酵素相關之 3,300 種 OMIM 基因遺傳疾病表型。期望建構出包括國人各類疾病之

之磷酸化酵素-藥物-疾病關聯性網絡資料庫。

（6）檢討與展望

在第一年計畫中，我們將分子間藥理作用介面家族 (molecular pharma-interface family)的概念應用在 ROCK1 激酶和莽草酸激酶 (shikimate kinase)，成功的找出可能的作用機制與抑制劑，證明了此概念的可行性與潛力，並已有 2 篇 SCI 文獻發表和 1 篇投稿文章正在審核中。然而目前仍缺乏大規模的生物統計數據來證實分子間藥理作用介面家族的概念。為此，在年底時，我們已予默克-密理博藥廠（Merck Millipore）進行大規模（50 種磷酸化酵素對 64 個化合物）磷酸化酵素-化合物活性分析。

未來，將繼續利用分子間藥理作用介面家族概念為已上市藥物（與中草藥）找出其潛在標的蛋白與新的疾病應用。另外，我們將分析已送測之大規模磷酸化酵素－化合物之活性分析結果，以精進我們模型之準確性。最後，我們將擴大磷酸化酵素和藥物交互作用的資料範圍，並利用蛋白質在細胞網絡中的功能來協助疾病的註解，完成磷酸化酵素-藥物-疾病關聯性網絡的建構，以增進對疾病的瞭解，包含其可能的作用路徑與藥物作用方式，並利用此網絡找出具潛力之藥物設計標的，開發多標靶藥物與進行藥物副作用之評估。

參考文獻

1　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 (2010).

2　Yang, L. *et al.* Exploring Off-Targets and Off-Systems for Adverse Drug Reactions via Chemical-Protein Interactome - Clozapine-Induced Agranulocytosis as a Case Study. *PLoS Computational Biology* **7** (2011).

3　Sawatdichaikul, O., Hannongbua, S., Sangma, C., Wolschann, P. & Choowongkomon, K. In silico screening of epidermal growth factor receptor (EGFR) in the tyrosine kinase domain through a medicinal plant compound database. *Journal of Molecular Modeling* (2011).

4　Hsu, K. C., Chen, Y. F., Lin, S. R. & Yang, J. M. iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC Bioinformatics* **12 Suppl 1**, S33 (2011).

5　Yang, J.-M. & Chen, C.-C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics* **55**, 288-304 (2004).

6　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).

7　Moscona, A. Oseltamivir resistance--disabling our influenza defenses. *The New England Journal of Medicine* **353**, 2633-2636 (2005).

8　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).

9　Fabian, M. A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* **23**, 329-336, doi:nbt1068 [pii] 10.1038/nbt1068 (2005).

10　Karaman, M. W. *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132, doi:nbt1358 [pii] 10.1038/nbt1358 (2008).

11　Buchdunger, E., Matter, A. & Druker, B. J. Bcr-Abl inhibition as a modality of CML therapeutics. *Biochim Biophys Acta: Reviews on Cancer* **1551**, M11-M18 (2001).

12　Halper, J. Growth factors as active participants in carcinogenesis: a perspective. *Veterinary Pathology* **47**, 77-97 (2010).

13　Tsai, C. C. *et al.* 7-Chloro-6-piperidin-1-yl-quinoline-5,8-dione (PT-262), a novel ROCK inhibitor blocks cytoskeleton function and cell migration. *Biochem Pharmacol* **81**, 856-865 (2011).

14   Liu, I. H., Lo, Y. S. & Yang, J. M. PAComplex: a web server to infer peptide antigen families and binding models from TCR-pMHC complexes. *Nucleic Acids Research* **39**, W254-W260 (2011).

伍、100年度計畫所培訓之研究人員

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

種類				人數	備註
專任人員	1.	博士後研究人員	訓練中	1	許凱程
			已結訓	0	
	2.	碩士級研究人員	訓練中	0	
			已結訓	0	
	3.	學士級研究人員	訓練中	0	
			已結訓	0	
	4.	其他	訓練中	0	
			已結訓	0	
兼任人員	1.	博士班研究生	訓練中	2	林志達、林伸融
			已結訓	0	
	2.	碩士班研究生	訓練中	1	尤宣人
			已結訓	0	
醫師			訓練中	0	
			已結訓	0	

特殊訓練課程（請於備註欄說明所訓練課程名稱）

陸、參與100年度計畫所有人力之職級分析

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

職級	所 含 職 級 類 別	參與人次
第一級	研究員、教授、主治醫師	1 人
第二級	副研究員、副教授、總醫師、助教授	2 人
第三級	助理研究員、講師、住院醫師	0 人
第四級	研究助理、助教、實習醫師	2 人
第五級	技術人員	0 人
第六級	支援人員	0 人
合計		5 人

〔註〕：

第一級：研究員、教授、主治醫師、簡任技正，若非以上職稱則相當於博士滿三年、碩士滿六年、或學士滿九年之研究經驗者。

第二級：副研究員、副教授、助研究員、助教授、總醫師、薦任技正，若非以上職稱則相當於博士、碩士滿三年、學士滿六年以上之研究經驗者。

第三級：助理研究員、講師、住院醫師、技士，若非以上職稱則相當於碩士、或學士滿三年以上之研究經驗者。

第四級：研究助理、助教、實習醫師，若非以上職稱則相當於學士、或專科滿三年以上之研究經驗者。

第五級：指目前在研究人員之監督下從事與研究發展有關之技術性工作，且具備下列資格之一者屬之：具初（國）中、高中（職）、大專以上畢業者，或專科畢業目前從事研究發展，經驗未滿三年者。

第六級：指在研究發展執行部門參與研究發展有關之事務性及雜項工作者，如人事、會計、秘書、事務人員及維修、機電人員等。

柒、參與100年度計畫所有人力之學歷分析

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

類別	學　歷　別	參與人次
1	博士	2 人
2	碩士	0 人
3	學士	0 人
4	專科	0 人
5	博士班研究生	2 人
6	碩士班研究生	1 人
7	其他	0 人
合計		5 人

捌、參與100年度計畫所有協同合作之研究室

註：群體/中心計畫者，不論是否提出各子計畫資料，都必須提出總計畫整合之資料

機構	研究室名稱	研究室負責人
國立交通大學	分子抗癌研究室	趙瑞益
國立清華大學	生物醫學科技研發中心主任	王雯靜
義守大學	生物科技學系	袁行修

玖、100 年度計畫執行情形

（1）簡述原計畫書中 100 年預計達成之研究內容

本計畫之第一年預計工作目標如下：

1. 依據蛋白質-配體結合介面之性質建立分子間藥理作用介面家族

2. 為已上市藥找出新用途與開發多標的藥物

3. 建立磷酸化酵素-藥物-疾病關聯性網絡資料庫

（2）100 年度計畫執行達成情形

1. 建立分子間藥理作用介面家族

   透過我們所提出「分子間藥理作用介面家族」的概念，我們透過 147 種磷酸化酵素的結晶結構，依照其蛋白質－配體結合介面之立體形狀、物理化學特性及藥效基團分群，成功的建立了一個磷酸化酵素的分子間藥理作用介面家族，屬於同一家族的蛋白質會與同一小分子化合物結合，目前已有 5 萬個化合物可供使用者查詢，將持續增加至 30 萬個化合物以上。

2. 舊藥新用及開發多標的藥物

   我們挑選 50 種磷酸化酵素及 64 個小分子化合物進行大規模蛋白質磷酸化酵素－化合物之活性分析，計畫於今年年底完成 3200 點(50 種磷酸化酵素對 64 個小分子化合物)活性測試實驗。選取磷酸化酵素包含與國人

常見癌症相關之磷酸化酵素如胃腸道基質腫瘤（GIST：Gastrointestinal stromal tumor）(KIT)、腸癌(MAPK8)等，而化合物部分是從已知 1216 個 FDA 藥物，約 9 萬種天然化合物以及超過 33 萬種來自現有公開化合物資料庫(Maybridge、Sigma 及 Specs)中，利用分子間藥理作用介面家族並結合分子嵌合工具(GEMDOCK)、區域官能基地圖(SiMMap)挑選而出。實驗結果將有助於開發針對國人常見癌症相關之磷酸化酵素的藥物、多標的藥物設計以及舊藥新用。

3. 建立磷酸化酵素-藥物-疾病關聯性網絡資料庫

根據上述研究成果，我們已建立了磷酸化酵素-藥物-疾病關聯性網絡資料庫之雛形，包含 147 種具有結晶結構的磷酸化酵素、1216 個藥物與超過 3 萬個抑制劑、79 種 OMIM 表型(Phenotype)，可用於開發抑制特定疾病之生化途徑相關磷酸化酵素的化合物（藥物、天然物）、降低藥物副作用。預期將擴展至包含 518 種磷酸化酵素、1216 個藥物及超過 3 萬個抑制劑、3300 種 OMIM 表型(Phenotype)之國人常見癌症相關磷酸化酵素-藥物-疾病關聯性網絡資料庫。

拾、附錄

無

拾壹、本年度之著作抽印本或手稿

Core site-moiety maps reveal inhibitors and binding mechanisms of orthologous proteins by screening compound libraries

Kai-Cheng Hsu[1†], Wen-Chi Cheng[2†], Yen-Fu Chen[1], Hung-Jung Wang[2], Ling-Ting Li[1], Wen-Ching Wang[2]* and Jinn-Moon Yang[1,3]*

[1]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, 30050, Taiwan

[2]Institute of Molecular and Cellular Biology & Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan

[3]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, 30050, Taiwan

[†]Joint first authorship

*Corresponding authors

ABSTRACT

Members of protein families often share conserved structural subsites for interaction with chemically similar moieties despite low sequence identity. We propose a core site-moiety map of multiple proteins (called CoreSiMMap) to discover inhibitors and mechanisms by profiling subsite-moiety interactions of immense screening compounds. The consensus anchor, the subsite-moiety interactions with statistical significance, of a CoreSiMMap can be regarded as a "hot spot" that represents the conserved binding environments involved in biological functions. Here, we derive the CoreSiMMap with six consensus anchors and identify six inhibitors (IC_{50} <8.0 μM) of shikimate kinases (SKs) of *Mycobacterium tuberculosis* and *Helicobacter pylori* from the NCI database (236,962 compounds). Studies of site-directed mutagenesis and analogues reveal that these conserved interacting residues and moieties contribute to pocket-moiety interaction spots and biological functions. These results reveal that our multi-target screening strategy and the CoreSiMMap can increase the accuracy of screening in the identification of novel inhibitors and subsite-moiety environments for elucidating the binding mechanisms of targets.

1

INTRODUCTION

The expanding number of protein structures and advances in bioinformatics tools have offered an exciting opportunity for structure-based virtual screening in drug discovery [1]. Although there are some successful agents in the antibiotic development, few agents act at novel molecular binding sites to target multiple antibiotic–resistant pathogenic bacteria [2,3]. However, screening tools are often designed for one-target paradigm and the scoring methods are highly target-dependent and energy-based. As a result, they cannot consistently and persuasively identify true leads, leading to a low success rate [4-6].

Orthologous proteins often perform similar functions, despite low sequence identity. Importantly, they frequently share conserved binding environments for interacting with partners. These proteins and their interacting partners (inhibitors or substrates) can be regarded as a pharmacophore family, which is a group of protein-compound complexes that share similar physical-chemical features and interaction patterns between the proteins and their partners. Such a family is analogous to a protein sequence family [7,8] and a protein structure family [9]. However, the establishment of pharmacophores often requires a set of known active ligands that were acquired experimentally [10-12]. Developing an efficient method for identifying new adaptive inhibitors against multiple targets from public compound libraries is therefore becoming an important task [13-15].

To address the above issues, we propose a core site-moiety map to discover inhibitors and mechanisms of multiple targets from large-scale docked compounds. The consensus anchors, which are subsite-moiety interactions with statistical significance in site-moiety maps of these proteins, represent the conserved binding environments that are involved in biological functions. The new method (called CoreSiMMap-based screening method) was heavily modified and improved from that "SiMMap" in our earlier work [16], which constructed a site-moiety map comprising of anchors from a target protein and thousands of docked compounds. An anchor contains three crucial elements, which are conserved interacting residues that constitute a binding pocket (part of the binding site), the preference of moieties, and a pocket-moiety interaction type.

The major enhancements of the CoreSiMMap for multi-target inhibitors from SiMMap are as follows: 1) we developed the robust theoretical model for the SiMMap; 2) the CoreSiMMap is designed for multiple target proteins by modifying the SiMMap on a single target protein ; 3) we added an anchor alignment method to identify core binding environments (anchors) among multiple targets to reveal binding mechanisms; 4) we added a rank-based consensus score (RCS) of multiple targets to improve the enrichment of true positives. Based on these enhancements and modifications, the CoreSiMMap-based screening method is useful to infer core pharmacophores both to identify adaptive inhibitors of multiple targets and to improve screening accuracy.

2

Here, we have applied the CoreSiMMap strategy to discover core pharmacophores and adaptive inhibitors of shikimate kinase (SK) of *Mycobacterium tuberculosis* and *Helicobacter pylori* (MtSK and HpSK) by screening large compound libraries. Mt causes tuberculosis and killed 1.7 million people in 2006[17]. Therefore, it is becoming a major public health threat[18]. We first derived core site-moiety maps that often represent the conserved binding environment elements or "hot spots" among orthologous targets based on virtual screening. In using core site-moiety maps, six potent adaptive inhibitors of MtSK and HpSK with low IC_{50} values (<8.0 μM) were identified. Site-directed mutagenesis revealed that the core pharmacophores often contribute to specific pocket-moiety interaction anchors. These results reveal the CoreSiMMap is useful to identify adaptive SK inhibitors and provide insight into the binding mechanisms of compounds.

RESULTS

Overview of CoreSiMMap

A CoreSiMMap is the consensus site-moiety maps, which consist of several consensus anchors derived from multiple targets, to represent essential features that are involved in the common biological functions of these targets (Figs. 1 and 2). The following criteria are considered for a CoreSiMMap: (1) the binding sites of the screening targets share conserved physical-chemical features; (2) pocket-moiety interaction profiles of these targets and well-docked compounds are similar; and (3) the site-moiety maps of these targets share comparable anchors (pharmacophores) with respect to their sites and crucial protein-ligand interactions.

Figure 1 presents the major steps of the CoreSiMMap-based method for orthologous targets (HpSK and MtSK). For each target, we used the top 2% (~6000) of compounds obtained by screening compound libraries to analyze target-compound interaction profiles (Fig. 1B) to establish the site-moiety maps and "pharmacophore spots (anchors) (Fig. 1C)" with statistically significant Z-scores. The superimposed pharmacophore models with anchors among orthologous proteins revealed overlapping regions, which are regarded as "hot spots". A set of hot spots from orthologous targets therefore form a core site-moiety map that represents the conserved binding elements (Fig. 1E). After enzyme inhibition assay, active and inactive compounds were used to refine the CoreSiMMap model and core anchors.

CoreSiMMaps of HpSK and MtSK

The CoreSiMMap that has comparable spots or hot spots is useful in providing biological insights and guiding the process of drug discovery including hit search and lead optimization. Here, we utilized the CoreSiMMap strategy to identify adaptive inhibitors of MtSK and HpSK.

3

The SK is the fifth enzyme in the shikimate pathway and converts shikimate to shikimate 3-phosphate[19]. The SK is an attractive target for the development of new antimicrobial agents, herbicides, and antiparasitic agents since it is essential to bacteria, fungi, and plants, but not animals [20]. Because there are distinct binding pockets with or without ligands for shikimate kinase, we sought to establish apo-form and closed-form CoreSiMMaps of SKs (Fig. 2). We first generated the apo-form SiMMaps of HpSK (6 anchors) and MtSK (7 anchors), allowing us to derive the CoreSiMMap with 6 consensus anchors (Figs. 2A and 2C). In parallel, the closed-form CoreSiMMap with 6 consensus anchors was also derived (Figs. 2B and 2D).

The apo-form and closed-form CoreSiMMaps have six comparable anchors: E1, H1–H3, V1, and V2 (Figs. 2C and 2D). H1, H2, and V1 sit at the ATP site, while H3, V2, and E1 are situated at the shikimate site (Figs. S1 and S2 in supporting material). The protein-ligand relationship was analyzed for each hot spot; a set of chemically related entities that contribute to intermolecular interactions were then identified (Figs. 2E, S1, and S2). Our results support the notion that a hot spot has a conserved binding environment with a specific chemical-physical property, which can be used as a guide in further lead optimization. The compounds moieties, anchors, SiMMaps and CoreSiMMaps are available at http://simmap.life.nctu.edu.tw/orthsimmap/.

Of the six consensus anchors (Fig. 2E), E1 is a negatively charged pocket that interacts with R57 (R58 in MtSK), R116 (R117 in MtSK), and R132 in HpSK (R136 in MtSK); these arginines are highly conserved in SKs and are critical for binding to shikimate [21] (Fig 2, S1, and S2 in supporting material). The chemical entities on E1 consisted of carboxyl, sulfonate, and phosphate groups. H1 is enclosed with a tight turn (Walker A motif) that binds the β-phosphate of ATP [21]. The identified moieties were carboxylic amide, sulfonate ester, carboxyl acid, and ketone (Fig. 2E). H2 is located between H1 and H3 and possesses a hydrogen-bonding environment from a 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-containing compounds fit into this pocket. H3 is situated above the central sheet in which two conserved residues (D33, and G80 in HpSK; D34, G80 in MtSK) contribute to H3. Amide, sulfonate ester, and ester groups were frequently identified.

V1, which is adjacent to H1, bears a vdW-binding environment and also contains residues from the Walker A motif. V2, which is near H3, is situated at the border between shikimate and the nucleotide binding regions. V1 and V2, allowing the interactions with large chemical groups, prefer aromatic groups. Analysis of the closed-form SiMMaps indicates that E114 and R116 (T115 and R117 in MtSK) in LID are conserved interacting residues (Fig. 2E). Analysis of the closed-form model shows that the four hot spots (E1, H1, H2, and V1) occupy a similar region in the closed-form pharmacophores. Given the closure of the LID region, E1, H2, and V1, interact with a residue from LID [19,22]. On the other hand, H3 and V2 were absent in the

closed form, perhaps because of its tight binding pocket.

Inhibitors and inhibition assay

Following the analyses of the SiMMap and the compound-anchor-residue profiles (Fig. 3), we used the CoreSiMMap in post-screening analysis to rescore docked compounds by the rank-based consensus scoring (RCS [23]), which combines energy-based and anchor-based scoring (Table 1). Since a compound that simultaneously docks into apo and closed-form binding sites of HpSK and MtSK is regarded as a potentially useful hit, we selected common top-ranked compounds from the closed-form and apo-form CoreSiMMap analysis for subsequent bioassay. After compounds were ranked by using RCS in both the Maybridge and the NCI databases, 48 available compounds (either requested or purchased) were subjected to MtSK and HpSK inhibitory assays. Among those, 10 compounds had $IC_{50} \leqq 100$ μM for both HpSK and MtSK, of which six [NSC45611 (**1**), NSC162535 (**2**), NSC45612 (**3**), NSC45174 (**4**), NSC45547 (**6**), and NSC45609 (**7**)] had $IC_{50} \leqq 10$ μM. In parallel, 65 existing kinase inhibitors were tested to evaluate their inhibitory effectiveness against shikimate kinase. Of the two compounds [AG538 (**5**) and GW5074 (**12**)] that showed inhibitory effects, AG538 (**5**) had a low IC_{50} value. Enzymatic kinetic analysis revealed that NSC45611 (**1**), NSC162535 (**2**), NSC45612 (**3**) and AG538 (**5**) were competitive inhibitors of ATP, in agreement with the docked poses (Table 2). Of these, NSC45611 (**1**), NSC162535 (**2**) and NSC45612 (**3**) competed with shikimate and had low IC_{50} and αK_i values, showing potent inhibition.

We then further evaluated the binding and pharmacophore modes of these inhibitors. Figure 4 shows that these three compounds have lower IC_{50} (≤ 10 μM) and fit well into five hot spots (H1, V1, H3, V2, and E1). The compounds with $IC_{50} \geq 20$ μM lack the negatively charged groups that are required to form electrostatic interactions with arginines (R57 and R136 in HpSK) on E1. On the other hand, kinase inhibitors AG538 (**5**) and GW5074 (**12**) did not occupy the shikimate site. Moieties with 1–3 rings were present at V1 and V2, yielding a number of vdW contacts. The binding groups of active inhibitors matched well with the identified moieties that were found from the consensus anchors. For example, the sulfonate groups of NSC162535 (**2**), NSC45611 (**1**), and NSC45612 (**3**) were found to occupy H1. The moieties of NSC162535 (**2**) (SO_3^- group), NSC45611 (**1**) (CO_2^- group), and NSC45612 (**3**) (CO_2^- group) occupied E1.

Consensus residues of CoreSiMMap and site-directed mutagenesis

A consensus anchor of orthologous targets that is identified from the conserved binding pockets has conserved interacting residues and a specific physico-chemical property, which often engage in specific enzymatic functions. We sought to investigate the roles of identified

consensus anchor residues of the CoreSiMMaps in catalysis (Fig. 2). Each of the selected residues was replaced with alanine and expressed in *Escherichia coli*. After purification using affinity chromatography, all mutants migrated to a major band of an apparent molecular mass of ~18 kDa in SDS-PAGE, as expected. We first investigated mutants of E1 residues (R57, R116, and R132) that contact with shikimate [19]. Enzymatic analysis revealed that these arginines had extremely low activity (Fig. 4G), suggesting their importance in catalysis. Indeed, R117 of MtSK, which corresponds to R116 of HpSK, has thus been suggested as a primary candidate to stabilize the transition-state intermediate [24].

For the H3 (D33) and V2 (F48) residues, D33A completely lost the enzymatic activity while F48A exhibited hardly any detectable activity (1%, Fig. 4G). D33 and F48 are in direct contact with shikimate. More importantly, D33 forms a hydrogen bond with the 3-OH group of shikimate, which may increase the nucleophilicity of the O atom or accept the proton from the 3-OH group of shikimate, facilitating catalysis. E114A, a LID residue whose side chain faces the solvent, retained 82% relative activity. On the other hand, the F48 side chain contacts with those from several residues nearby (V44, E53, F56, R57 and P117), potentially forming a stable platform that interacts with the ligand for a subsequent catalytic reaction.

We then evaluated residues from H1, H2, and V1 at the nucleotide site. H1 residues are primarily from the Walker A motif (P loop; residues 11–16, GSGKSS) that surrounds the phosphate groups of the nucleotides. Of the two alanine mutants (S12A and S15A), S12A retained 59% of the relative activity, while S15A had extremely low activity (1%). The S15 side chain resides near the β-phosphate of ADP. Furthermore, the adjacent lysine (K14) corresponding to K15 of MtSK has been identified as a critical catalytic residue in MtSK since its side chain points toward the γ-phosphate [24]. The other H2 mutant, D31A, retained 62% of the relative activity, possibly because of its remoteness from the phosphate group. For V1 that is just next to H1, several H1 residues are also shared by V1. Enzymatic analysis indicated that M10A retained 38% relative activity. These site-directed mutagenesis studies revealed that the critical conserved interacting residues of a pocket-moiety interaction spot participated in biological functions.

Analogues assay and CoreSiMMap

To verify the moiety preferences of consensus anchors, we identified four analogues of NSC162535 (2) [NSC45547 (6), NSC45609 (7), NSC37215, and NSC45208] for inhibitory assays (Fig. 3H). NSC45547 (6) and NSC45609 (7) that occupy E1 (SO$_3^-$ group) and H1 (SO$_3^-$ and NO$_2$ groups) retained good IC$_{50}$ values (7.8 and 7.0 μM for HpSK; 3.4 and 2.0 μM for MtSK). Conversely, NSC37215 and NSC45208, which cannot anchor at E1, lost the inhibitory.

To evaluate the significance of the pocket-moiety interaction preferences of consensus

anchors in the CoreSiMMaps, we performed clustering analysis on 27 inhibitory assay compounds (Fig. 3). These compounds can be roughly clustered into three groups. The potent inhibitors of Group I [NSC162535 (2), NSC45609 (7), NSC45547 (6), NSC45174 (4), NSC45611 (1) and NSC45612 (3)] match more than 5 consensus anchors (Fig. 3B). Each of the Group II compounds [RH00037 (8), RH00016 (9), GK01385 (10), and SPB01099 (11)] matches four of six anchors; Group III comprises kinase inhibitors [AG538 (5) and GW5074 (12)] and these compounds match anchors at the ATP site. The inactive compounds often match few consensus anchors in the HpSK/MtSK (usually 4), in particular, E1 is the least seen anchor. While the inhibitors of Group I and II match anchors at the ATP site and the shikimate site, kinetic assay indicates competitive inhibitions for ATP and shikimate acid (Table 2). The kinase inhibitors of Group III occupied the anchors of the ATP site, and only showed the competitive inhibitions for ATP. The pocket environment of ATP is generally conserved in kinase family, and the inhibitors of Group III also have the broadband inhibition for multiple kinases, including AG538 (5), which is observed on insulin-like growth factor-1 receptor (IGF-1R) [25], IR, EGFR [26], and Src kinases [27].

We sought to evaluate the binding-site features of the apo and closed forms that show a significant structural change because of LID closure and domain movement [19,21,28]. Indeed, the apo-form HpSK structure shows higher deviations in the LID region and in the adenine binding loop region (residues 143–150) (Fig. S3 in supporting material). Superimposition of the apo and closed HpSK structures revealed a tight binding pocket, particularly for the shikimate binding site. While the apo and closed forms had the same number of consensus anchors (E1, H1, H2, H3, V1 and V2), the spatial arrangements of these anchors were more packed in the closed form (Figs. 2C and 2D). Residues (D31 and D33) that contribute to H2 of the apo form were closer to each other in the closed conformation, resulting in a lower volume at this pocket. Likewise, the corresponding pocket at V2 surrounded by F48, G80, and G81 in HpSK had less space in the closed form, hindering the accommodation of large moieties in this pocket. The above evidences demonstrate that the induced LID conformation of shikimate kinases was sensitive in the structure-based drug discovery strategy (Table 1). The CoreSiMMap, considering both apo-form and close-form structures, can reduce the ill effects.

Accuracy of the CoreSiMMap-based screening method

We next evaluated the accuracy of energy-based and CoreSiMMap-based scoring methods. The energy-based score (docking energy) of a compound was generated using a docking program, GEMDOCK [23,29]. The scoring function in GEMDOCK is piecewise linear potential (PLP), which is a simple scoring function and is comparable to some energy-based scoring functions in estimating binding affinities [30]. Here, we used the hit rate and enrichment to assess the overall accuracy. The hit rate is defined as A_h/T_h (%) and the

enrichment is $(A_h/T_h)/(A/T)$, where A_h is the number of active compounds among the T_h highest ranking compounds (hit list), A is the total number of active compounds in the database, and T is the total number of compounds in the database. Here, A_h is 8 (hits, compounds with $IC_{50} <$ 100 µM) based on the bioassay results, and T is 6000 based on the 6000 top-ranked compounds from screening databases. We computed the average of enrichments, defined as $\sum_{i=1}^{A}[(i/T_h^i)/(A/T)]/A$, where T_h^i is the number of compounds in a hit list containing i active compounds.

As shown in Figure 5A, the CoreSiMMap-based scoring method (solid lines) significantly outperformed the energy-based scoring method (dashed lines), which is often used in docking tools, when applied to apo-form HpSK and MtSK (Fig. S5A). The average enrichments of 3.73 (HpSK), 1.59 (MtSK), and 2.74 (fusion of HpSK and MtSK) were obtained using energy-based scoring method, as compared to 11.18 (HpSK), 35.51 (MtSK), and 93.69 (fusion of HpSK and MtSK), obtained using the CoreSiMMap scoring method. The average hit rates were 0.92% (HpSK), 0.21% (MtSK), and 0.37% (fusion of HpSK and MtSK) using energy-based scoring methods as compared to 34.13% (HpSK), 17.57% (MtSK), and 67.02% (fusion of HpSK and MtSK) using the CoreSiMMap score. Additionally, the CoreSiMMap-based RCS strategy was more accurate than the SiMMap-based strategy for a single target (HpSK or MtSK).

The CoreSiMMap scoring method can reduce the deleterious effects of screening ligand structures that are rich in charged or polar atoms. Generally, energy-based scoring functions favor the selection of high-molecular-weight compounds that yield high vdW potentials, as well as polar compounds that produce hydrogen-bonding and/or electrostatic potentials [23]. The average molecular weights of the 100 top-ranked compounds of the CoreSiMMap-based and energy-based scoring methods were 459.9 and 532.6, respectively; the average numbers of polar atoms were 11.3 (CoreSiMMap-based method) and 14.1 (energy-based method) (Figs. 5B and 5C). The ranks of the 10 active compounds were much higher in the CoreSiMMap-based scoring analysis than in the energy-based analysis. Notably, NSC162535 (2) was ranked as 1 and 1821 using the apo-form CoreSiMMap-based and energy-based scoring methods, respectively (Table 1).

DISCUSSION

By far the largest obstacle in structure-based drug discovery is the relatively low hit rate of scoring methods owing to the lack of adequate quantities of binding partners for a given target. The accuracy of a given individual scoring function is generally unknown and/or cannot be evaluated. The emphasis of the CoreSiMMap-based screening method developed here is therefore to provide a useful index to improve screening accuracy for the identification of

adaptive inhibitors when the target proteins share conserved binding sites. Through the employment of this developed method, we successfully found six new potent inhibitors (<8.0 μM) of HpSK and MtSK. Two of the 65 kinase inhibitors were also found to inhibit both HpSK and MtSK activity. The finding that NSC45611 (1), NSC162535 (2), and NSC45612 (3) were competitive inhibitors of ATP and shikimate suggests that they belong to a novel class of shikimate kinase inhibitors. These results illustrate a robust CoreSiMMap-based screening approach for identifying selective kinase inhibitors.

The combined apo/closed CoreSiMMap analysis utilized here is considered to be useful for induced-fit P-loop kinases. Of six potent inhibitors, all except NSC45609 (7) have a higher rank in the apo-form than in the closed-form CoreSiMMap-based scoring analysis. Additionally, the top-ranked inhibitors according to the apo-form CoreSiMMap-based scoring analysis often have larger moieties (such as naphthalene or nitrobenzene) on both sides as opposed to those with a relatively small moiety (e.g., amide or aliphatic chain). The closed-form CoreSiMMap-based scoring analysis has, nonetheless, yielded useful hits, including NSC45609 (7) and SPB01099 (11). Target proteins with dynamic induced-fit forms, such as the P-loop SKs, represent a major limitation for the structure-based screening approach.

The CoreSiMMap that shares consensus binding environments among orthologous targets will be valuable for the development of effective common inhibitors. Indeed, the inhibitors identified herein had comparable IC_{50} values, docked poses, and competition properties for both HpSK and MtSK (Tables 1 and 2 and Fig. 4). More importantly, this CoreSiMMap, which includes both nucleotide and shikimate sites, specifically identified inhibitors that competed with both substrates, whereas the two general kinase inhibitors (AG538 (5) and GW5074 (12)) that did not compete well with the shikimate site. Notably, the potent competitive shikimate inhibitors have a common -N=N- moiety. A further lead optimization will be required to verify the importance of this moiety.

The developed CoreSiMMap-based screening method is database-independent. Comparable anchors were identified in compounds from the Maybridge and NCI databases. Each of the anchors also included similar chemical moieties. Nonetheless, the derived proportion of these moieties was different because the Maybridge and NCI databases contain heterogeneous distribution of compounds. For example, the proportions of carboxyl, sulfonate, and phosphate were significantly higher in compounds from the NCI database than in those from the Maybridge database. On the other hand, the derived model was sensitive to binding-site properties, as revealed by the difference between the apo- and closed-form models (Fig. 2). In summary, we anticipate that the CoreSiMMap-based screening method will be useful for discovering new inhibitors, investigating binding mechanisms, and guiding lead optimization for orthologous targets.

CONCLUSIONS

We have developed a CoreSiMMap-based screening method to derive conserved pocket-moiety environments (e.g., hot spots) between orthologous targets and inhibitors from screening compound libraries. Studies of site-directed mutagenesis and analogues revealed that critical conserved interacting residues of pocket-moiety interaction anchors participate in biological functions. Experimental results reveal that the CoreSiMMap-based screening method is database-independent and the CoreSiMMap scoring method significantly outperformed the energy-based scoring method, which is often used in docking tools. Furthermore, we have successfully obtained several potent inhibitors and revealed the binding mechanisms for MtSK and HpSK. The CoreSiMMap is useful in providing biological insights and guiding the process of drug discovery, including hit search and lead optimization.

MATERIALS AND METHODS

Overview of CoreSiMMap-based screening method

Figure 1 shows the main steps of the CoreSiMMap-based screening method. For each orthologous target, we first docked selected compounds into the binding site using GEMDOCK, which is an in-house molecular docking program that uses PLP to measure intermolecular potential energy between proteins and compounds [29]. Our previous works showed that GEMDOCK yields very comparable results to some other docking tools (e.g., DOCK, FlexX, and GOLD) when applied to 100 protein-ligand complexes and some virtual screening targets [29,30]. Additionally, GEMDOCK has been successfully applied to identify novel inhibitors and binding sites for some targets [31-34]. After the docking procedure, we used the top 2% (~6,000) of compounds based on docking energy to analyze target-compound interaction profiles (Fig. 1B and Fig 3) to establish the site-moiety map (Figs. 2A and 2B). Each anchor represented a local binding environment with a specific physico-chemical property or pharmacophore spot, which was derived by identifying statistically significant interacting residues and compound moieties. The CoreSiMMap (Figs. 2C and 2D) that consists of the aligned anchors of orthologous proteins is generated by extracting the consensus anchors from the orthologous SiMMaps.

Preparations of SK structures of four target proteins and screening databases

Four structures of the target proteins were selected from Protein Data Bank (PDB) for virtual screening. They were the apo-form structure (PDB code: 1ZUH[19]) and the closed-form structure (PDB code: 1ZUI[19]) of HpSK, and the apo-form structure (PDB code:

2IYT[21]) and the closed-form structure (PDB code: 1ZYU[24]) of MtSK. The residues of the binding sites of these four structures were prepared by the following steps. Firstly, the apo-form and closed-form structures of MtSK were aligned to the closed-form structure of HpSK using the structural alignment tool [35]. The bound ligands (SKM and ACP) of HpSK were used to determine the binding sites of these four structures. For each structure, the residues of the binding site were obtained by considering the protein atoms that were ≤10 Å from these two bound ligands.

We selected compounds from the Maybridge and NCI databases for both HpSK and MtSK to establish the site-moiety maps and to identify novel candidates. In total, 65,947 (Maybridge) and 236,962 (NCI) compounds with molecular weights between 200 and 650 dalton were prepared for screening. In addition, we collected a dataset of 37 orthologous target pairs with biological function annotations (*e.g.*, substrate binding, metal binding, and catalytic residues) summarized from UniProt [36], Consurf server [37], and Catalytic Site Atlas [38] (Table. S1) to verify the consensus anchors residues that were derived using the CoreSiMMap-based screening method. Experimental results reveal that the consensus anchors often represent the conserved binding environments that are involved in biological functions, such as substrate binding, metal binding, or catalytic functions. Residues of consensus anchors of these 37 orthologous pairs are often the key residues (*i.e.*, substrate binding residues, metal binding residues, catalytic residues, or high conserved residues with Consurf conservation score 8 or 9) or key anchors (*i.e.*, anchors that contain key residues) (Fig. S4). For example, D81, D84, and D201 of inositol-1-monophosphatase of *Methanocaldococcus jannaschii* (D82, D85, and D200 in *Archaeoglobus fulgidus*) are the consensus anchor residues and coordinate two metal ions for catalysis [39]. Another example is the androgen receptor of *Homo sapiens*, which is a target in the treatment of prostate cancer [40]. Three residues (N705, R752, and T877) of the consensus anchors are essential for the ligand binding [41]. These results reveal that the consensus anchor residues often play important roles in biological functions.

CoreSiMMap-based screening method

The main steps of the CoreSiMMap-based method for producing SiMMaps and a CoreSiMMap from orthologous targets are described as follows (Fig. 1):

(1) Virtual screening of orthologous targets. We used in-house GEMDOCK program [23,29] to screen Maybridge (65,947 compounds) and NCI (236,962 compounds) databases for four targets, including apo and closed forms of HpSK and MtSK. GEMDOCK first assigned the formal charge and atom type (*i.e.,* donor, acceptor, both, or nonpolar) based on physico-chemical property of each atom of both compounds and binding sites. GEMDOCK then utilizes PLP to measure the intermolecular potential energies between the binding sites

(rigid) and the compounds (flexible). Finally, these docked compounds were ranked in order of energy. The top 2% (~6,000) of compounds of each target were selected from the screening results for subsequent protein-compound profiling. A personal computer cluster (80 nodes, each with an Intel Woodcrest 2.66 GHz processor and 4 GB of RAM) was used to implement the docking procedure. On average, a docking run (in which a compound was docked into a binding site of a protein) took 63s; therefore, docking 302,909 compounds into these four proteins took about 11 days (~1,200,000 docking runs).

(2) Selection of a diverse set of top-ranked compounds. To maintain a wide range of potential functional groups in the sampling of anchors, a cluster method was utilized to cluster the top-ranked compounds. A hierarchical clustering method that exploited the topological features of the compounds was used. The topological features were generated by the atom pair (AP) approach [42,43]. An AP was a substructure of a compound and was presented as "atom type I – bond distance – atom type J", where atom types I and J are the atom types of atoms I and J, respectively, and bond distance is the number of bonds measured along the shortest path between atoms I and J. Atoms were classified into 10 types based on their chemical properties (Table. S2), and there are 55 different combinations of these 10 atom types in total. The value of the AP was set to 1 (ON) if the compound contained the substructure; otherwise, the value was set to 0. Here, the maximum number of the bonds was set to 15. Therefore, the number of APs was 825 (55 × 15), and the topology of a compound was represented as a string of 825 binary bits.

Subsequently, the AP binary strings of the top-ranked compounds were used to cluster the compounds in a hierarchical clustering procedure[44]. In the procedure, we firstly used the Tanimoto coefficient (Tc) [43] to quantify the similarity between two compounds (A and B). Tc was defined as

$$Tc(A,B) = \frac{|N_A \cap N_B|}{|N_A \cup N_B|} \tag{1}$$

, where $|N_A \cap N_B|$ is the number of ON bits that are common to both A and B, and $|N_A \cup N_B|$ is the number of ON bits in either A or B. Based on the Equation (1), the similarity matrix of the compounds was generated, and applied to construct a dendrogram by using complete linkage hierarchical clustering. The Tc threshold that was used to group compounds was set to 0.9. The compound with the lowest energy in each group was selected as the representative compound for the group. Accordingly, we could select a diverse set of compounds to ensure that many functional groups would form the anchors.

(3) Profiling analysis of target-compound interactions (Fig. 1B and Fig. 3). The profiles described the interactions (*i.e.*, electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions) between the diverse top-ranked compounds and the protein residues, from

which the anchors could be derived. The target-compound interactions were identified by applying the PLP of GEMDOCK program [29]. In the E or H profiles, a profile entry was set to 1 if the compound forms electrostatic or hydrogen-bonding interactions with the residue (green regions); otherwise, the entry was set to 0 (black regions). I the V profile, a profile entry was set 1 if the V energy was less than -4 kcal/mol.

(4) Identification of anchors (Fig. 1C). We identified consensus interactions between residues and compound moieties from the profiles. For an interacting residue, we used the Z-score to measure the interacting conservation between the residue and moieties. The interacting conservation is treated as a binomial distribution, which is approximated as a normal distribution when either $p \leq 0.5$ and $np > 5$ or $p > 0.5$ and $n(1-p) > 5$, where n is the number of selected compounds and p is the probability of forming an interaction. Theoretically, at least 500 compounds should be selected to construct a target-compound interaction profile. Spatially neighboring interacting residues and moieties with statistically significant Z-score ≥ 1.645 are referred as an anchor. A set of anchors derived from the target-compound interaction profile can be used to establish a site-moiety map for each orthologous target.

(5) Establishment of CoreSiMMap of orthologous targets (Figs. 1D, 1E and Fig. 2). Firstly, the apo-form and closed-form structures of orthologous proteins (HpSK and MtSK) were aligned using a structural alignment tool [35]. The superimposed SiMMaps (anchors) of the orthologous proteins revealed an overlapping region of matched anchors which form the CoreSiMMap (Figs. 2C and 2D). For compound x, the CoreSiMMap-based score, combining anchor (CoreSiMMap) and PLP (GEMDOCK) scores, is defined as

$$CAS(x) = \sum_{i=1}^{a} w_i AS_i(x) - 0.001E(x),$$ where w_i is the conservation of the anchor i on

orthologous targets; $AS_i(x)$ is the anchor score of compound x in anchor i; a is the number of anchors, and $E(x)$ is the docked energy of the compound x. The CoreSiMMap-based rank of compound x was obtained by arranging the compounds in order of descending $CAS(x)$. Table S3 presents the parameters that are used to construct the CoreSiMMap.

(6) Inhibition assay. We selected top-ranked compounds using rank-based consensus scoring (RCS) for subsequent inhibitory assay. The RCS of a compound x was calculated by combining the ranks of m (apo/closed) forms of n orthologous targets as follows:

$$S_R(x) = \sum_{i=1}^{n} \sum_{k=1}^{m} R_{C_{ki}}(x)/2mn,$$ where $R_{C_{ki}}(x)$ is the CoreSiMMap-based rank of the

compound x on the k (apo/closed) form of target i. Here, m and n are 2. We obtained the RCS rank of the compound x by arranging the compounds in order of ascending $S_R(x)$.

(7) Refinement of CoreSiMMaps. Active and inactive compounds from the enzyme inhibition assay were used to evaluate and refine the CoreSiMMaps.

Cloning, expression, and purification of *M. tuberculosis* shikimate kinase (MtSK) and *H. pylori* shikimate kinase (HpSK)

AroK, which encodes MtSK, was amplified from the chromosomal DNA of *Mycobacterium tuberculosis* str. Haarlem with PCR using pfu DNA polymerase. Primers (MtSK-F (forward), 5'-CGCGGATCCATGGCACCCAAAGCGGTTCTCGTCG-3'; MtSK-R (reverse), 5'-AAACTGCAGTCATGTGGCCGCCTCGCTGGGGCTG-3') that contain sequences for the *Bam*HI and *Pst*I sites, respectively, were designed based on the nucleotide sequence of the reported MtSK gene of *M. tuberculosis* (accession number AASN01000059; Genome Project: 17353). The amplified fragment was inserted into the pET28a expression vector to generate pET28a-MtSK containing an N-terminal six-histidine tag (His$_6$) for purification purposes. The recombinant MtSK in *E. coli* BL-21 (DE3) cells was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 16°C. Bacterial pellets were fractionated, and soluble proteins in cytosolic fractions were collected. The expressed MtSK protein with a His$_6$ tag was purified by immobilized-nickel ion chromatography, followed by Superdex-75 gel filtration chromatography (Pharmacia), and then analyzed using SDS-PAGE to verify its purity. The cloning and purification of HpSK were based on previous methods [19].

Preparation of mutant HpSKs

Site-directed mutagenesis was performed using the overlap extension PCR [45] method with the plasmid pQE30-HpSK as the template. All the mutations were confirmed by sequencing of the whole ligated PCR fragment. Mutant proteins were expressed and purified by the same procedures as described for the wild-type HpSK.

Enzymatic activity assay and analysis of inhibitor kinetics

Shikimate kinase activity was determined by coupling the release of ADP from the SK-catalyzed reaction to the oxidation of NADH using pyruvate kinase (EC 2.7.1.40) and lactate dehydrogenase (EC 1.1.1.27) as coupling enzymes [46]. Shikimate-dependent oxidation of NADH was measured by monitoring the decrease in $A340$ ($\varepsilon = 6,200$ M^{-1} cm^{-1}). The assay was carried out at 25°C in a mixture of 100 mM Tris-HCl buffer, pH 7.5, 50 mM KCl, 5 mM MgCl$_2$, 1.6 mM shikimic acid, 2.5 mM ATP, 1 mM phosphoenolpyruvate (PEP), 0.1 mM NADH, 2.5 units of pyruvate kinase per ml, and 2.7 units of lactate dehydrogenase per ml. All assays were conducted in a 96-well microplate and analyzed with a spectrophotometer (FLUOstar OPTIMA, BMG LABTECH).

Stock solutions of compounds were prepared in dimethylsulfoxide (DMSO). Each set of measurements included 10% DMSO as a negative control. Approximately 80 nM enzyme was added to a reaction of 200 μl containing compounds. The initial velocities of enzyme activity were determined in the presence of various concentrations of test compounds to investigate the dose-dependent inhibition effects. The IC_{50} values of these compounds were obtained by fitting the data to a sigmoid dose-response equation in GraphPad Prism 4.

After the preliminary screening, the inhibitor modality was determined by measuring the effects of inhibitor concentrations on the enzymatic activity as a function of substrate concentration. In the inhibition experiment in which the ATP concentration was fixed at 2.5 mM, shikimate was a varied substrate (0.06, 0.12, 0.24, 0.48, and 0.96 mM) when the concentration of inhibitor was varied from 0 to 50 μM. In parallel, in the inhibition experiment in which the shikimate concentration was fixed at 1.6 mM, ATP was a varied substrate (0.06, 0.12, 0.24, 0.48, and 0.96 mM) when the concentration of inhibitor was varied from 0 to 50 μM. A nonlinear least square-fitting algorithm was applied to the absorbance data to determine the kinetic mechanism data.

ACKNOWLEDGEMENTS

REFERENCES

1. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1: 882-894.
2. Ginsberg AM, Spigelman M (2007) Challenges in tuberculosis drug research and development. Nat Med 13: 290-294.
3. Lock RL, Harry EJ (2008) Cell-division inhibitors: new insights for future antibiotics. Nat Rev Drug Discov 7: 324-338.
4. Lyne PD (2002) Structure-based virtual screening: an overview. Drug Discov Today 7: 1047-1055.
5. Shoichet BK (2004) Virtual screening of chemical libraries. Nature 432: 862-865.

6. Shoichet BK, McGovern SL, Wei B, Irwin JJ (2002) Lead discovery using molecular docking. Curr Opin Chem Biol 6: 439-446.

7. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res 38: D211-222.

8. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37: D211-215.

9. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, et al. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res 32: D226-229.

10. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1: 882-894.

11. Wei D, Jiang X, Zhou L, Chen J, Chen Z, et al. (2008) Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. J Med Chem 51: 7882-7888.

12. Wolber G, Seidel T, Bendix F, Langer T (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discov Today 13: 23-29.

13. Freire E (2002) Designing drugs against heterogeneous targets. Nat Biotechnol 20: 15-16.

14. Zimmermann GR, Lehar J, Keith CT (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. Drug Discov Today 12: 34-42.

15. Wei D, Jiang X, Zhou L, Chen J, Chen Z, et al. (2008) Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. J Med Chem 51: 7882-7888.

16. Chen YF, Hsu KC, Lin SR, Wang WC, Huang YC, et al. (2010) SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. Nucleic Acids Res 38 Suppl: W424-430.

17. WHO (2008) Global tuberculosis control - surveillance, planning, financing (WHO/HTM/TB/2008.393).

18. Dye C (2006) Global epidemiology of tuberculosis. Lancet 367: 938-940.

19. Cheng WC, Chang YN, Wang WC (2005) Structural basis for shikimate-binding specificity of *Helicobacter pylori* shikimate kinase. J Bacteriol 187: 8156-8163.

20. Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, et al. (1998) Evidence for the shikimate pathway in apicomplexan parasites. Nature 393: 801-805.

21. Hartmann MD, Bourenkov GP, Oberschall A, Strizhov N, Bartunik HD (2006) Mechanism of phosphoryl transfer catalyzed by shikimate kinase from *Mycobacterium tuberculosis*. J Mol Biol 364: 411-423.

22. Gu Y, Reshetnikova L, Li Y, Wu Y, Yan H, et al. (2002) Crystal structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the dynamic role of the LID domain in catalysis. J Mol Biol 319: 779-789.

23. Yang JM, Chen YF, Shen TW, Kristal BS, Hsu DF (2005) Consensus scoring criteria for

improving enrichment in virtual screening. J Chem Inf Model 45: 1134-1146.

24. Gan J, Gu Y, Li Y, Yan H, Ji X (2006) Crystal structure of *Mycobacterium tuberculosis* shikimate kinase in complex with shikimic acid and an ATP analogue. Biochemistry 45: 8539-8545.

25. Blum G, Gazit A, Levitzki A (2003) Development of new insulin-like growth factor-1 receptor kinase inhibitors using catechol mimics. J Biol Chem 278: 40442-40454.

26. Hallak H, Moehren G, Tang J, Kaou M, Addas M, et al. (2002) Epidermal growth factor-induced activation of the insulin-like growth factor I receptor in rat hepatocytes. Hepatology 36: 1509-1518.

27. Zahradka P, Litchie B, Storie B, Helwer G (2004) Transactivation of the insulin-like growth factor-I receptor by angiotensin II mediates downstream signaling from the angiotensin II type 1 receptor to phosphatidylinositol 3-kinase. Endocrinology 145: 2978-2987.

28. Gan J, Gu Y, Li Y, Yan H, Ji X (2006) Crystal structure of *Mycobacterium tuberculosis* shikimate kinase in complex with shikimic acid and an ATP analogue. Biochemistry 45: 8539-8545.

29. Yang JM, Chen CC (2004) GEMDOCK: a generic evolutionary method for molecular docking. Proteins: Structure, Function, and Bioinformatics 55: 288-304.

30. Yang JM, Shen TW (2005) A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. Proteins: Structure, Function, and Bioinformatics 59: 205-220.

31. Yang JM, Chen YF, Tu YY, Yen KR, Yang YL (2007) Combinatorial computational approaches to identify tetracycline derivatives as flavivirus inhibitors. PLoS One 2: e428.

32. Chin KH, Lee YC, Tu ZL, Chen CH, Tseng YH, et al. (2010) The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas campestris*. J Mol Biol 396: 646-662.

33. Hung HC, Tseng CP, Yang JM, Ju YW, Tseng SN, et al. (2009) Aurintricarboxylic acid inhibits influenza virus neuraminidase. Antiviral Res 81: 123-131.

34. Yang M-C, Guan H-H, Yang J-M, Ko C-N, Liu M-Y, et al. (2008) Rational design for crystallization of beta-lactoglobulin and vitamin D-3 complex: revealing a secondary binding site Cryst Growth Des 8: 4268-4276.

35. Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11: 739-747.

36. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, et al. (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34: D187-D191.

37. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, et al. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 33: W299-302.

38. Porter CT, Bartlett GJ, Thornton JM (2004) The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 32: D129-D133.

39. Johnson KA, Chen LJ, Yang HY, Roberts MF, Stec B (2001) Crystal structure and catalytic mechanism of the MJ0109 gene product: A bifunctional enzyme with inositol monophosphatase and fructose 1,6-bisphosphatase activities. Biochemistry 40: 618-630.

40. Heinlein CA, Chang CS (2004) Androgen receptor in prostate cancer. Endocr Rev 25: 276-308.

41. Bisson WH, Cheltsov AV, Bruey-Sedano N, Lin B, Chen J, et al. (2007) Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. Proc Natl Acad Sci U S A 104: 11927-11932.

42. Carhart RE, Smith DH, Venkataraghavan R (1985) Atom pairs as molecular features in structure-activity studies: definition and applications. J Chem Inf Comp Sci 25: 64-73.

43. Willet P (1998) Chemical similarity searching. J Chem Inf Comp Sci 38: 983-996.

44. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: A review. Acm Comput Surv 31: 264-323.

45. Chiu WC, You JY, Liu JS, Hsu SK, Hsu WH, et al. (2006) Structure-stability-activity relationship in covalently cross-linked N-carbamoyl D-amino acid amidohydrolase and N-acylamino acid racemase. J Mol Biol 359: 741-753.

46. Millar G, Lewendon A, Hunter MG, Coggins JR (1986) The cloning and expression of the aroL gene from Escherichia coli K12. Purification and complete amino acid sequence of shikimate kinase II, the aroL-gene product. Biochem J 237: 427-437.
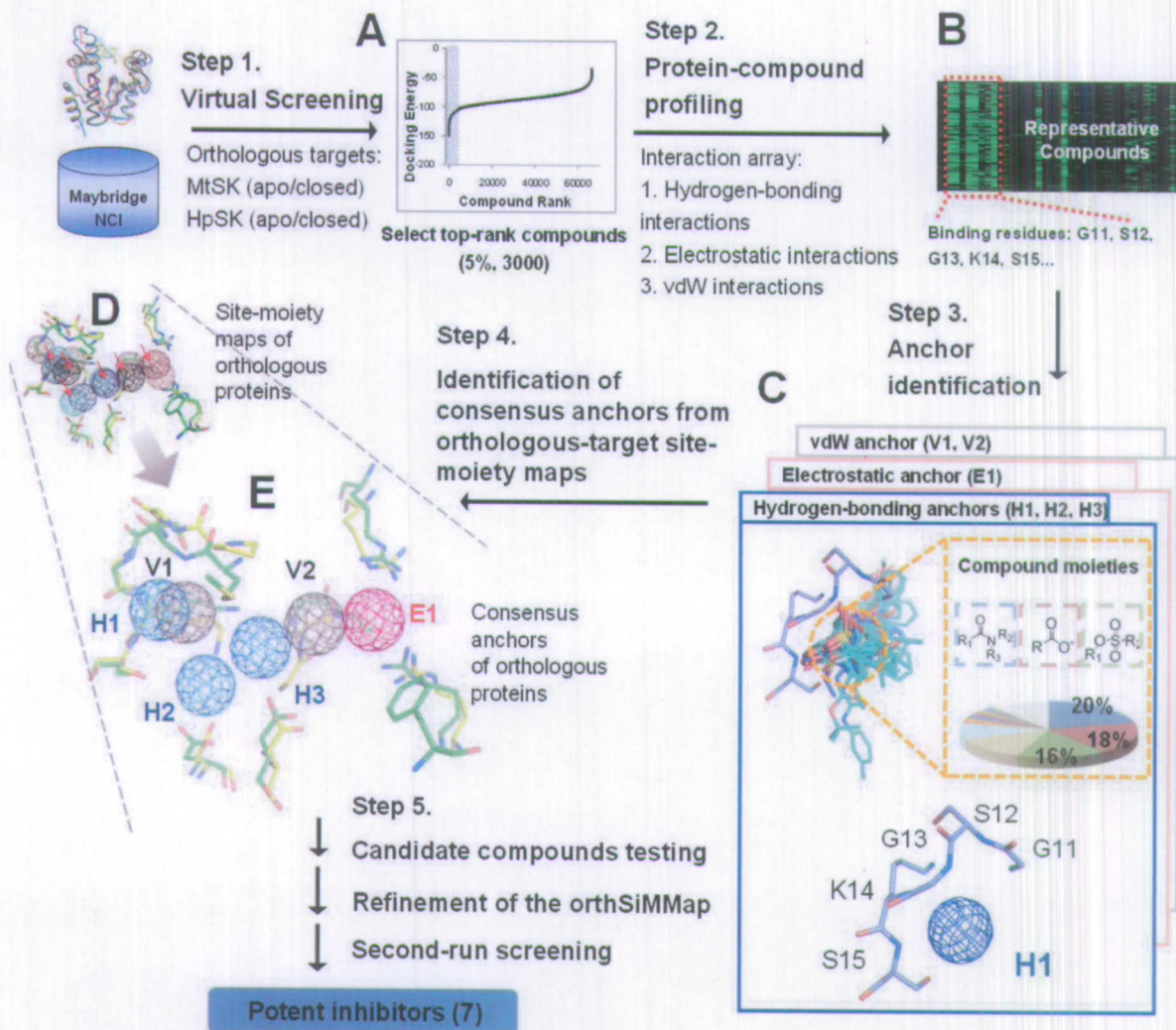
Figure 1. Framework of CoreSiMMap-based screening method. In Step 1, GEMDOCK was used to generate docked poses for HpSK and MtSK by screening compound libraries (Maybridge and NCI). For each target (HpSK or MtSK), the protein-compound interacting profile was derived by the top 2% (~6,000) of compounds ranked by docking energy. In Step 3, conserved interactions of the target protein and chemical moieties of ligands are identified to deduce the anchors of HpSK and MtSK. The CoreSiMMap is constructed based on the features that are conserved between orthologous target site-moiety maps, which will be used to select candidate compounds for the enzymatic assay. Finally, the model is refined based on the bioassay of candidate compounds.
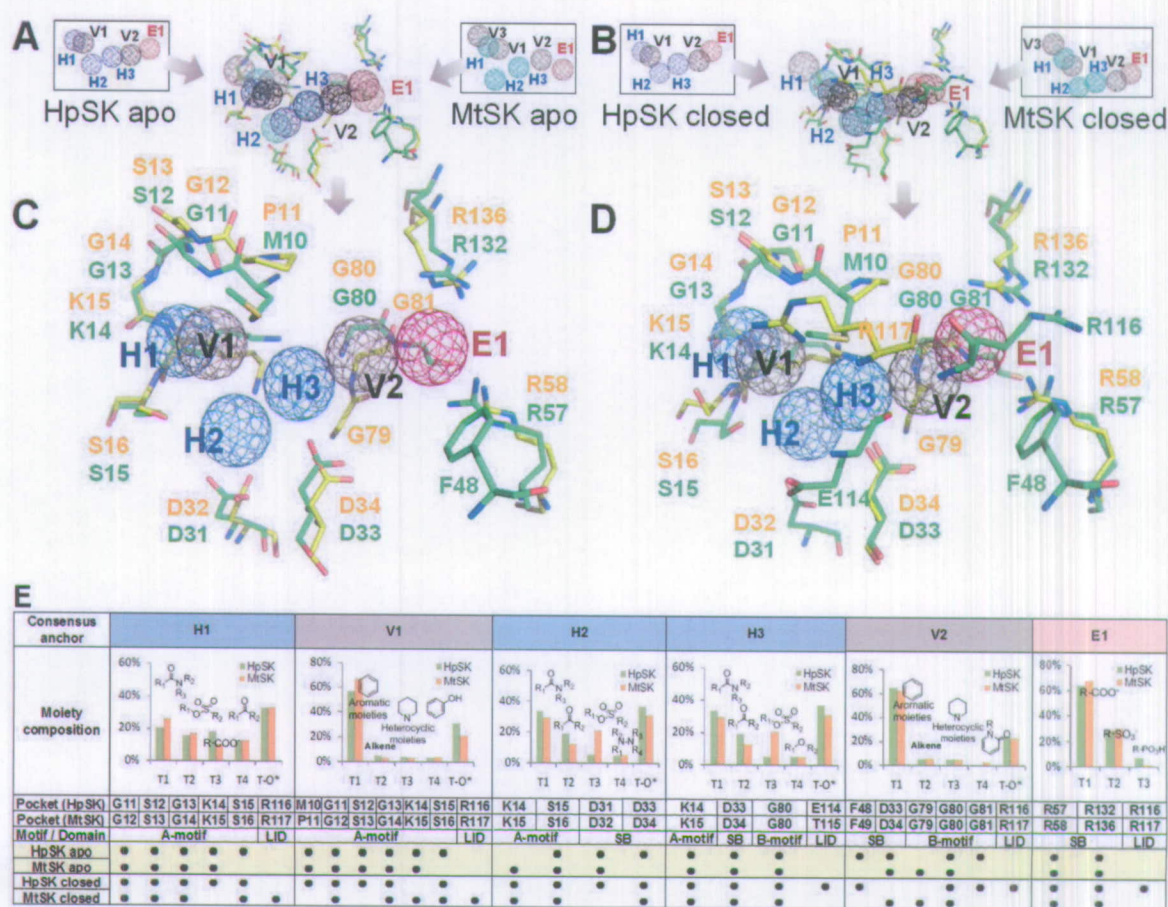
Figure 2. Shikimate kinase CoreSiMMaps. (A) Superimposed apo-form anchors of HpSK and MtSK. (B) Superimposed closed-form anchors of HpSK and MtSK. (C) The apo-form CoreSiMMap and (D) the closed-form CoreSiMMap include six consensus anchors derived from consensus anchors (A) and (B), respectively. Each consensus anchor shares conserved residues between HpSK and MtSK and has the same interaction type of binding environment. (E) Features of the six consensus anchors of the apo-form CoreSiMMap. Each of the T groups (T1–T4) represents a given chemical moiety and T-O* indicates other chemical groups. H1, V1, and H2 are situated at the ATP-binding site, while H3, V2, and E1 are at the shikimate-binding site. Each consensus anchor includes conserved interacting residues (●) and the major chemical moieties of the compound candidates.
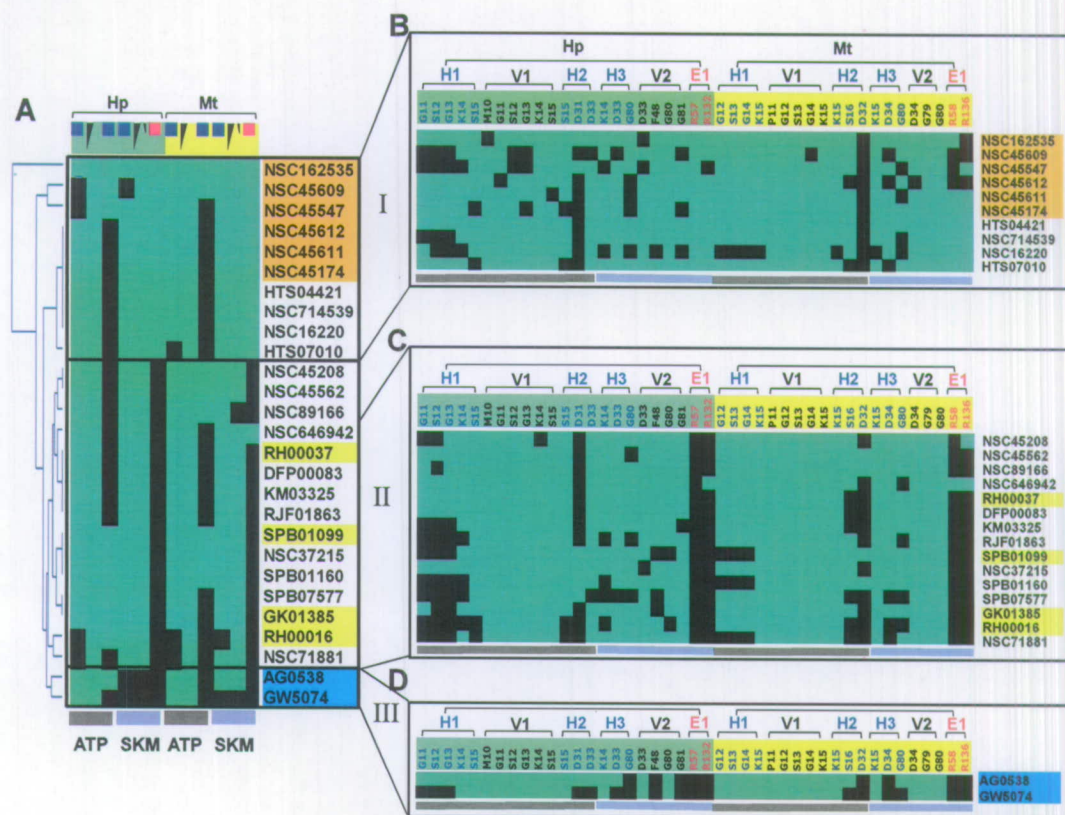
Figure 3. **Interaction profiles between selected anchor residues and 27 tested compounds.** (A) Anchor profile of tested compounds on shikimate kinases. (B) Group I: NCI compounds (orange). (C) Group II: Maybridge compounds (yellow). (D) Group III: kinase inhibitors (cyan). The NCI compounds consistently occupy anchors E1 and V2 at both ATP and shikimate sites. The NCI compounds except NSC45174 are competitive inhibitors with both ATP and shikimate. For the Maybridge compounds, none form electrostatic interactions with R57 and R132 on the consensus anchor E1. The two kinase compounds are located at the ATP site, which fact is consistent with the kinetic results that reveal that these compounds exhibited competitive inhibition with ATP and noncompetitive inhibition with shikimate.
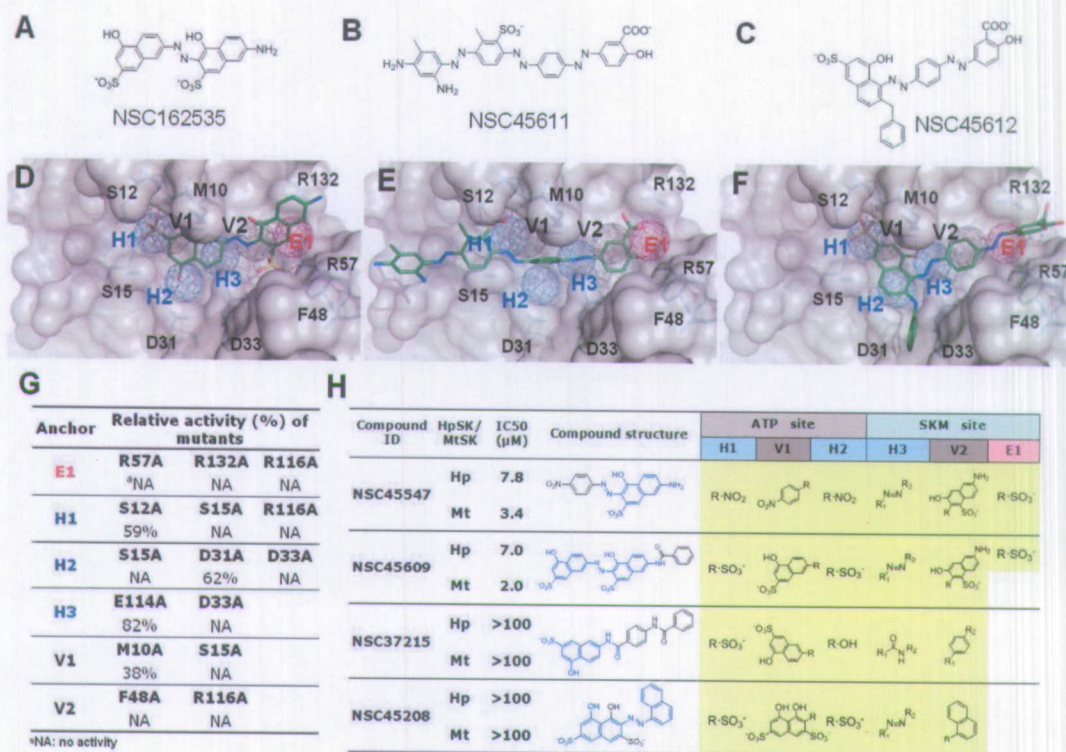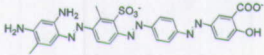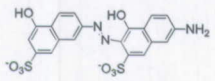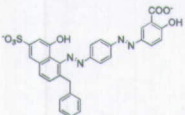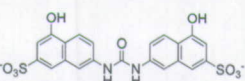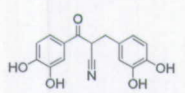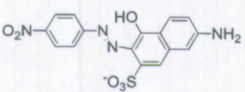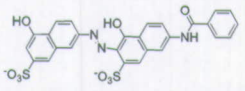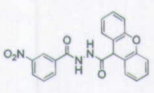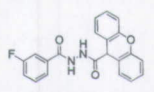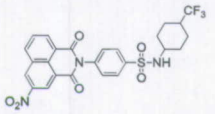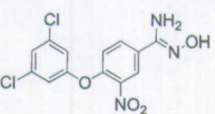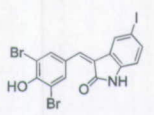
Figure 4. Characterization of shikimate kinase inhibitors by enzyme assay, CoreSiMMaps, site-mutagenesis studies, and analogues. (A-C) Structures of three inhibitors, NSC162535 (**2**), NSC45611 (**1**), and NSC45612 (**3**). (D-F) Relationships between anchors and docked mode of each inhibitor for HpSK. These compounds consistently include two negative charge moieties (SO_3^- or CO_2^-) that form hydrogen bonds with conserved interacting residues of anchors E1 and H1. (G) Comparison of relative activities of HpSK mutants. The conserved interacting residues for each anchor were mutated. R57, R132, R116, and F48 located in the shikimate site were critical for the enzymatic functions. (H) Potency of NSC162535 analogues. The substitution moieties of analogues are indicated in black. Those that lack the E1 moiety greatly lost the inhibitory effects ($IC_{50} > 100$ µM).

G

Anchor	Relative activity (%) of mutants		
E1	R57A [a]NA	R132A NA	R116A NA
H1	S12A 59%	S15A NA	R116A NA
H2	S15A NA	D31A 62%	D33A NA
H3	E114A 82%	D33A NA	
V1	M10A 38%	S15A NA	
V2	F48A NA	R116A NA	

[a]NA: no activity

H

Compound ID	HpSK/MtSK	IC50 (µM)	Compound structure	ATP site H1	V1	H2	SKM site H3	V2	E1
NSC45547	Hp / Mt	7.8 / 3.4		R-NO2		R-NO2			R-SO3
NSC45609	Hp / Mt	7.0 / 2.0		R-SO3		R-SO3			R-SO3
NSC37215	Hp / Mt	>100 / >100		R-SO3		R-OH			
NSC45208	Hp / Mt	>100 / >100		R-SO3		R-SO3			

Figure 5. Performance of CoreSiMMap method on apo-form HpSK and MtSK. (A) True-hit rates of energy-based and CoreSiMMap scoring approaches. The CoreSiMMap scores (solid line) of adaptive inhibitors are significantly better than the energy-based scores (dashed line) for the 6000 top-ranked compounds by combining the Maybridge and NCI databases. (B) Distribution of number of polar atoms. (C) Molecular weights of top 100 compounds from CoreSiMMap scores and energy-based scores.
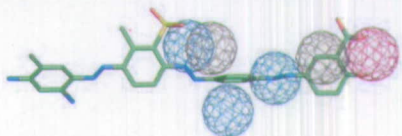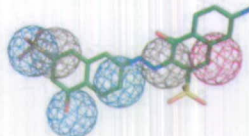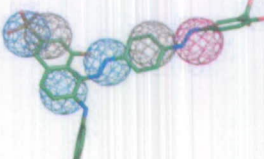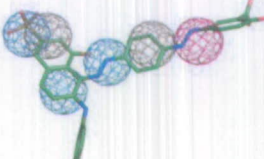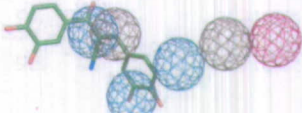
Table 1. Ranks of active compounds using CoreSiMMap, energy-based, and combined scoring methods for apo and closed forms of HpSK and MtSK

Compound ID	SK species	IC$_{50}$ (μM)	Compound structure	Apo form		Closed form		RCS [a]
				CoreSiMMap	Energy	CoreSiMMap	Energy	
NSC45611 (1)	Hp	4.8		48	435	515	827	96
	Mt	1.5						
NSC162535 (2)	Hp	4.9		1	1821	242	253	25
	Mt	1.6						
NSC45612 (3)	Hp	6.1		32	106	238	229	31
	Mt	2.8						
NSC45174 (4)	Hp	7.8		38	162	737	110	147
	Mt	2.8						
AG538 (5) [b]	Hp	2.3				- [b]		
	Mt	0.4						
NSC45547 (6)	Hp	7.8		130	5999	308	1540	67
	Mt	3.4						
NSC45609 (7)	Hp	7.0		18	4017	5	17	3
	Mt	2.0						
RH00037 (8)	Hp	23.8		786	876	891	1549	371
	Mt	<100						
RH00016 (9)	Hp	40.2		3765	6000	1219	5824	1837
	Mt	<100						
GK01385 (10)	Hp	<100		286	1774	730	49	199
	Mt	<100						
SPB01099 (11)	Hp	<100		117	5940	68	2871	19
	Mt	<100						
GW5074 (12) [b]	Hp	31.4				- [b]		
	Mt	29.6						

[a] The rank is the rank combination of CoreSiMMap and energy.

[b] The compound was derived from the 65 existing kinase inhibitors.

Table 2. Properties of some potent inhibitors of HpSK and MtSK

Compound ID	SK species	Inhibition mode[a]		αKi, ATP (μM)	αKi, SKM (μM)	Docked pose	
		ATP	SKM			ATP site	SKM site
						H1 V1 H2	H3 V2 E1
NSC45611 (**1**)	Hp	■	■	1.1	1.7		
	Mt	■	■	0.3	0.7		
NSC162535 (**2**)	Hp	■	■	1.9	1.8		
	Mt	■	■	0.2	0.6		
NSC45612 (**3**)	Hp	■	■	2.0	2.4		
	Mt	■	■	0.7	1.0		
AG538 (**5**)	Hp	■	□	3.1	5.4		
	Mt	■	□	0.04	0.4		

[a]■: Competitive inhibition; □: Non-competitive inhibition

7-Chloro-6-piperidin-1-yl-quinoline-5,8-dione (PT-262), a novel ROCK inhibitor blocks cytoskeleton function and cell migration

Chih-Chien Tsai [a,b], Huei-Fang Liu [a], Kai-Cheng Hsu [a,c], Jinn-Moon Yang [a,c], Chinpiao Chen [d], Kuang-Kai Liu [a], Tzu-Sheng Hsu [b], Jui-I. Chao [a,e,*]

[a] Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 30068, Taiwan
[b] Institute of Pharmacology and Toxicology, Tzu Chi University, Hualien 970, Taiwan
[c] Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu 30068, Taiwan
[d] Department of Chemistry, National Dong Hwa University, Hualien 970, Taiwan
[e] Institute of Molecular Medicine and Bioengineering, National Chiao Tung University, Hsinchu 30068, Taiwan

ARTICLE INFO

ABSTRACT

The 5,8-quinolinediones are precursors for producing multiple types of bioactive products. In this study, we investigated a new compound derived from 5,8-quinolinediones, 7-chloro-6-piperidin-1-yl-quinoline-5,8-dione (designated as PT-262), which markedly induced cytoskeleton remodeling and migration inhibition in lung carcinoma cells. Comparison with various cytoskeleton inhibitors, including paclitaxel, colchicine and phallacidin, the cell morphology following treatment with PT-262 was similar to phallacidin on the cell elongation and abnormal actin polymerization. However, PT-262 did not directly bind to actin filaments. ROCK (Rho-associated coiled-coil forming protein kinase) is a downstream effector of RhoA to mediate the phosphorylation of myosin light chain (MLC) and cytoskeleton reorganization. The RhoA–ROCK–MLC pathway has been shown to promote cancer cell migration and metastasis. Interestingly, PT-262 was more effective on inhibiting ROCK kinase activities than specific ROCK inhibitors Y-27632 and H-1152. PT-262 induced cytoskeleton remodeling and migration inhibition in A549 lung carcinoma cells. The total MLC and phosphorylated MLC proteins and stress fibers were blocked after treatment with PT-262. Nonetheless, the RhoA protein and GTPase activity were not altered by PT-262. A computational model suggests that PT-262 interacts with the ATP-binding site of ROCK protein. Together, these findings demonstrate that PT-262 is a new ROCK inhibitor.

1. Introduction

The reorganization of cytoskeleton plays important roles in the regulation of cancer cell migration and metastasis. The RhoGTPases are key regulators of cytoskeleton dynamics [1,2]. The RhoGTPase family members include RhoA, Rac1, and Cdc42 in mammalian cells. RhoA is involved in the regulation of cytoskeleton reorganization and focal adhesion, whereas Rac1 and Cdc42 work together at the regulation of cell leading edges to form

lamellipodia and filopodia [3,4]. RhoA is a key family of RhoGTPases that participates in cancer migration and metastasis [5,6]. Overexpression of RhoA has been found in a variety of cancers including, lung, bladder, testicular, ovarian, colon, and breast [7]. ROCK (Rho-associated coiled-coil forming protein kinase) is a RhoA downstream protein [5]. There are two isoforms of ROCK, known as ROCK1 and ROCK2 [8]. Both ROCK1 and ROCK2 can regulate the activity of myosin light chain (MLC) proteins by direct MLC phosphorylation [9,10]. Cell migration of actomyosin contractility is mediated by the phosphorylation of MLC for cell movement [1]. Moreover, ROCK has been shown to induce the stress fiber formation and cancer cell migration and metastasis [5,11,12]. Thus, the development of strategies or drugs to block the RhoA–ROCK pathway is highly desirable for cancer therapy.

Y-27632, [(+)-(R)-trans-4-(1-aminoethyl)-N-(4-pyridyl)cyclohexane carboxamide dihydrochloride], is a specific ROCK inhibitor [10,13]. The inhibiting ROCK mechanism of Y-27632 is through binding to the catalytic site of ROCK by competing with the ATP binding site [14]. Inhibition of ROCK by Y-27632 blocks cancer cell

migration, invasion and metastasis [15–17]. Y-27632 inhibits ROCK kinase activity to disrupt networks and remodeling of actin filaments in cancer cells [15,17]. Furthermore, Y-27632 reduces cancer invasion by reducing the levels of matrix metalloproteinases (MMPs) such as MMP-2 and MMP-9 [17]. In addition, H-1152, (S)-(+)-2-methyl-1-[(4-methyl-5-isoquinoline)sulfonyl]-homopiperazine, is another specific ROCK inhibitor derived from isoquinoline–sulfonamide analogues similar to Y-27632 by binding to catalytic site of ROCK [10,18].

The 5,8-quinolinediones are useful precursors for producing multiple types of bioactive products [19]. The derivatives of quinolinediones have been shown to possess many biological activities including anti-tumor and anti-microbial actions. 6-Anilino-5,8-quinolinedione (LY83583) is an inhibitor of guanylyl cyclase that can reduce cancer cell proliferation [20]. 6-Chloro-7-(2-morpholin-4-ylethylamino)quinoline-5,8-dione (NSC 663284), an inhibitor of CDC25 protein phosphatases, induced cell cycle arrest by reducing CDC2 activation in cancer cells [21]. Lavendamycin is a bacterially derived quinolinedione that displays significant anticancer activities [22]. 7-Chloro-6-piperidin-1-yl-quinoline-5,8-dione (designated as PT-262) is a new derivative of 5,8-quinolinedione, which induced apoptosis associated with inhibiting CDC2 and ERK phosphorylation [23].

In this study, we investigated a new ROCK inhibitor, PT-262, which inhibited cytoskeleton function, stress fiber formation, and migration in human lung carcinoma cells. In addition, PT-262 was more effective in inhibiting ROCK kinase activity and cancer cell migration than Y-27632 and H-1152. PT-262 can be developed as a novel ROCK inhibitor.

2. Materials and methods

2.1. Materials

The chemical synthesis of PT-262 was according to a previous study [23]. PT-262 was dissolved in 100% dimethyl sulfoxide (DMSO). The concentration of DMSO was <0.2% in the control and drug-containing media in each experiment. Hoechst 33258, propidium iodide (PI), 3-(4,5-dimethyl-thiazol-2-yl)2,5-diphenyl tetrazolium bromide (MTT), colchicine, paclitaxel, and the Cy3-labeled mouse anti-β-tubulin, gelatin, fibronectin and Y-27632 were purchased from Sigma Chemical Co. (St. Louis, MO). Anti-ERK-2 (C-14) and anti-RhoA (sc-179) were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). BODIPY FL phallacidin (B-607) and Lipofectamine™ 2000 were purchased from Invitrogen (Carlsbad, CA). The Cy5-labeled goat anti-rabbit IgG was purchased from Amersham Pharmacia Biotech (Little Chalfont Buckinghamshire, UK). Anti-phospho-MLC (serine-19), anti-MLC, and anti-ROCK1 antibodies were purchased from Cell Signaling Technology, Inc. (Beverly, MA). H-1152 was purchased from Merck Calbiochem (San Diego, CA).

2.2. Cell culture

The A549 cell line (ATCC, #CCL-185) was derived from a non-small cell lung adenocarcinoma tumor. A549 cells were cultured in RPMI-1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS), 100 units/ml penicillin, 100 μg/ml streptomycin, and L-glutamine (0.03%, w/v). The cells were maintained at 37 °C and 5% CO_2 in a humidified incubator (310/Thermo, Forma Scientific, Inc., Marietta, OH).

2.3. Cytoskeleton staining and confocal microscopy

The cells were cultured on coverslips, which were kept in a 60-mm Petri dish for 16–20 h. At the end of treatment, the cells were washed with isotonic PBS (pH 7.4), and then the cells were fixed in 4% paraformaldehyde solution in PBS for 1 h at 37 °C. Subsequently, the coverslips were washed three times with PBS, and non-specific binding sites were blocked in PBS containing 10% FBS and 0.3% Triton X-100 for 1 h. Actin filament (F-actin) and β-tubulin were stained with 20 U/ml BODIPY FL phallacidin and anti-β-tubulin Cy3 (1:50) for 30 min at 37 °C, respectively. Finally, the nuclei were stained with 2.5 μg/ml Hoechst 33258 for 30 min. The samples were examined under a Leica confocal laser scanning microscope (Mannheim, Germany) that equipped with an UV laser (351/364 nm), an Ar laser (457/488/514 nm), and a HeNe laser (543 nm/633 nm). The fluorescence images were displayed through the frame store in the computer, and the cell length was calculated using a Leica confocal software (Ver. Lite).

2.4. Bio-atomic force microscopy (Bio-AFM)

To observe the cytoskeleton alteration and cell elongation by PT-262, the cells were analyzed by a BD CARV II confocal microscope combined a NanoWizard Bio-AFM (JPK Instruments, Berlin). Bio-AFM was mounted on an inverted microscope, TE-2000-U (Nikon, Japan). The silicon nitride non-sharpened cantilever was used by a nominal force constant of 0.06 N/m (DNP-20, Veeco). The images were scanner by using contact mode. Line scan rates are varied from 0.5 to 2 Hz.

2.5. Actin polymerization assay

Actin polymerization assays in vitro (BK003) were purchased from Cytoskeleton, Inc. (Denver, CO). After treatment with or without PT-262, the actin polymerization reaction was analyzed according to the manufacturer's recommendations. Phallacidin was used as a positive control in actin polymerization. Briefly, the actin polymerization assays were based on the enhanced fluorescence of pyrene-conjugated actin that occurs during polymerization. The enhanced fluorescence were measured by pyrene monomer globular-actin (G-actin) formed polymer pyrene F-actin in a fluorometer at excitation wavelength 365 nm and emission wavelength 407 nm.

2.6. Boyden chamber analysis

The cell migration was examined by the Boyden chamber analysis. Boyden's chamber system was used two chambers separated by a collagen coated 8 μm pore size polycarbonate membrane. Briefly, the polycarbonate filter of 8 μm pore size was soaked in 0.5 M acetic acid overnight. The filter was washed with distilled water and then incubated for 16 h in a 100 μg/ml gelatin solution and air dried. The gelatin coating filters were incubated again in a 10 μg/ml fibronectin solution for 2 h. The tested drugs were added to the bottom well of a Bodyen chamber. Cells on the upper surface of the filters were carefully removed with a cotton swab. Cells on the lower surface were counted after stained with hematoxylin under a light microscope.

2.7. Wound healing assay

The cells were cultured in six-well plate for 24 h overnight, resulting in a monolayer that was more than 90% confluent. The cell monolayer was scraped with a 200 μl pipette tip to generate the wound of 6–7 mm. Thereafter, the cells were treated with or without PT-262, washed with PBS, and then re-cultured in fresh medium for various periods. Photographs were taken at 8 and 24 h on the same position of the wound.

2.8. ROCK kinase assay

A549 cells were treated with or without PT-262, Y-27632 or H-1152. After drug treatment, the cell lysates were diluted in kinase buffer and allowed to phosphorylate the bound substrate myosin binding subunit (MBS). The amount of MBS was calculated in a binding reaction with a horseradish peroxidase conjugate of AF20, an anti-phospho-MBS threonine-696 specific antibody, which catalyzes the conversion of the chromogenic substrate tetra-methylbenzidine from a clear to a blue-colored solution. The color reaction product was quantified by measuring the absorbance at

450 nm using a VERSAmax microtiter plate reader (Molecular Devices Corp., Sunnyvale, CA).

2.9. Western blot analysis

Western analyses of RhoA, ERK-2, MLC, phospho-MLC, and ROCK1 were performed using specific antibodies. At the end of the treatment, the cells were lysed in ice-cold whole cell extract buffer containing protease inhibitors as described [24]. Briefly, equal amounts of proteins in samples were subjected to electrophoresis using 10–12% sodium dodecyl sulfate-polyacrylamide gels. After
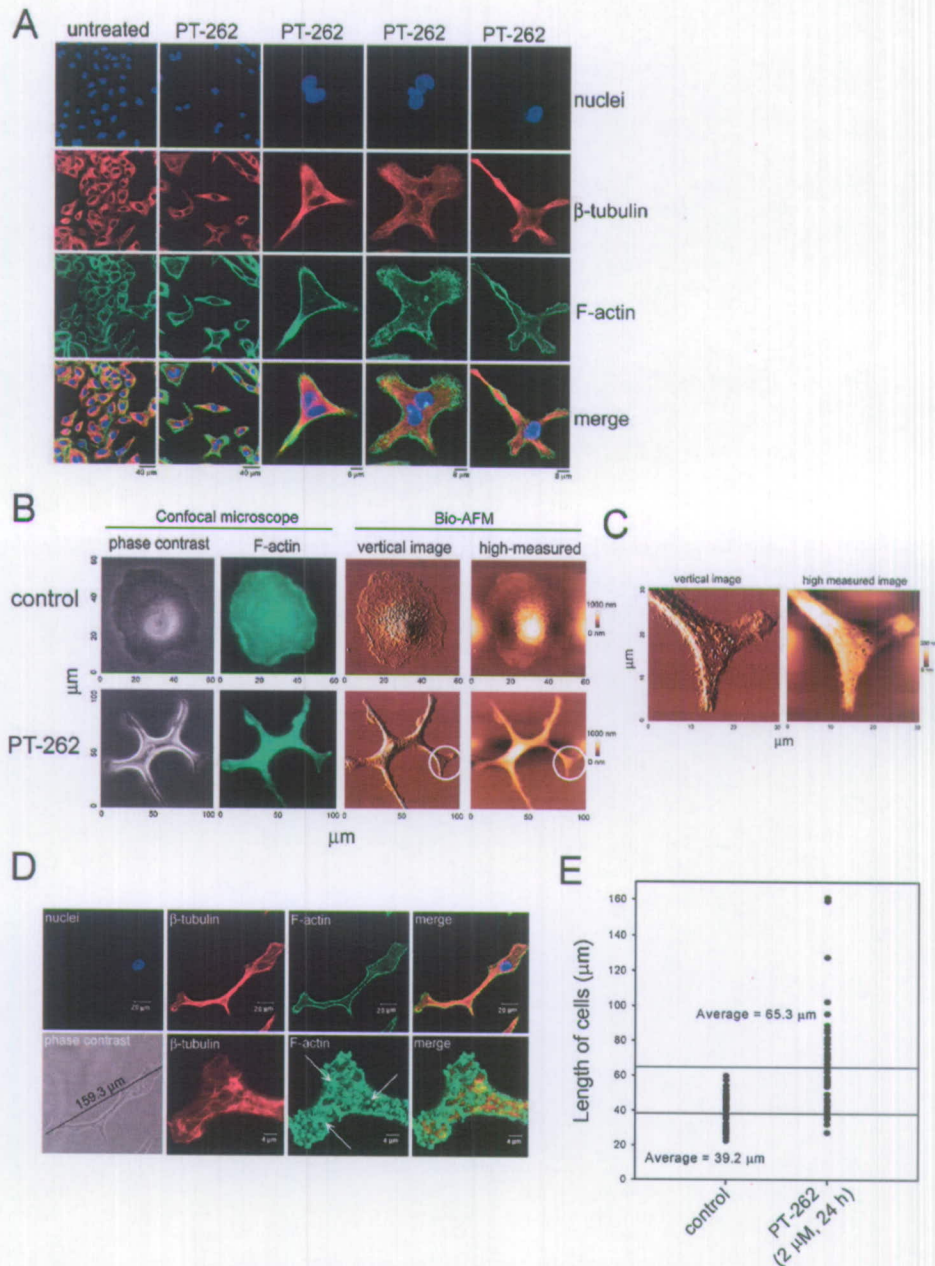
Fig. 1. Effect of PT-262 on the cytoskeleton remodeling and cell elongation in lung carcinoma cells. (A) A549 cells were treated with or without 2 μM PT-262 for 24 h. At the end of treatment, β-tubulin, F-actin, and nuclei were stained with Cy3-labeled mouse anti-β-tubulin, BODIPY FL phallacidin, Hoechst 33258, respectively. The β-tubulin displayed a red color. The blue color indicated the location of nuclei or chromosomes. The green color indicated the location of F-actin. (B) The untreated or PT-262-treated cells were analyzed by confocal microscope-combined Bio-AFM. The green color indicated the location of F-actin proteins. (C) The amplified pictures were obtained from circle marks of (B). (D) A549 cells were treated with or without 2 μM PT-262 for 24 h. The β-tubulin, F-actin, and nuclei were stained with the Cy3-labeled mouse anti-β-tubulin, BODIPY FL phallacidin, Hoechst 33258, respectively. The treated or untreated cells were subject to confocal microscopy analysis. The arrows indicate the F-actin spikes. (E) The cell length was measured using Leica confocal software. The average cell length (long diameter) was calculated from three separate experiments. (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)
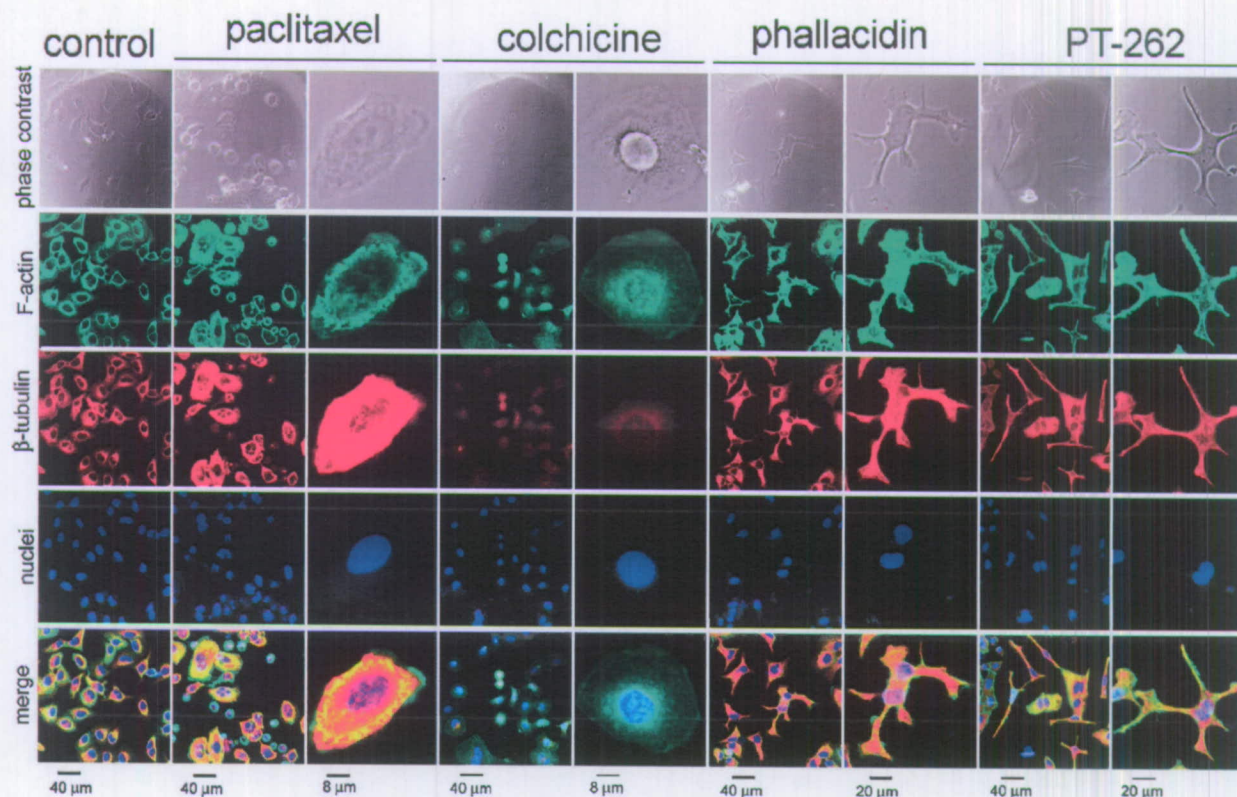
Fig. 2. Comparison of various cytoskeleton blockers on the cytoskeleton alteration and cell morphology in lung carcinoma cells. A549 cells were treated with or without cytoskeleton blockers. β-tubulin, F-actin, and nuclei were stained with Cy3-labeled mouse anti-β-tubulin, BODIPY FL phallacidin, Hoechst 33258, respectively. The treated or untreated cells were subject to confocal microscopy analysis. The β-tubulin displayed a red color. The blue color indicated the location of nuclei or chromosomes. The green color indicated the location of F-actin. (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

electrophoretic transfer of proteins onto polyvinylidene difluoride membranes, they were sequentially hybridized with primary antibody and followed with a horseradish peroxidase-conjugated second antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). Finally, the protein bands were visualized using the enhanced chemiluminescence detection system (NEN, Boston, MA).

2.10. RhoA GTPase activity assay

Activated GTP-RhoA proteins were detected by a pull-down assay kit (Upstate Biotechnology, Lake Placid, NY) for Rho–GTP bound to the Rho-binding domain (RBD) of Rhotekin, a downstream effector of Rho. Rho activation kit used the Rhotekin-RBD fused with GST to affinity precipitate cellular Rho–GTP. Briefly, the cells were treated with or without PT-262. At the end of treatments, the cell lysates were collected and then added equal volumes to incubate with GST–Rhotekin–RBD and glutathione–sepharose beads according to the manufacturer's instructions. Activated GTP–RhoA proteins bound to beads in cell lysates were precipitated and then subjected to Western blot using specific anti-RhoA antibody. The levels of active RhoA were determined by normalizing with the total RhoA proteins present in the cell lysates.

2.11. Transfection

Control siRNA (5′-UUCUCCGAACGUGUCACGU-3′) and RhoA siRNA (5′-CGGAAUGAUGAGCACACAA-3′) were purchased from Qiagen Inc. (Valencia, CA) that were used for transfection by using Lipofectamine™ 2000 (Invitrogen) according to the manufacturer's recommendations.

Table 1
The actions of various cytoskeleton blockers.

Compounds	Effects			
	Microtubule polymerization	Microtubule depolymerization	Actin polymerization	Actin depolymerization
Paclitaxel	+[a]	−	−	−
Colchicine	−	+[b]	−	−
Cytochalasin B[c]	−	−	−	+[d]
Phallacidin	−	−	+[e]	−
PT–262	−	−	+[f]	−

[a] Induction of mitotic arrest by inhibiting spindle function but not inducing cell elongation.

[b] Induction of mitotic arrest by inhibiting spindle function but not inducing cell elongation.

[c] A cytokinesis inhibitor induces actin depolymerization.

[d] Cytochalasin B inhibits actin polymerization and cytokinesis.

[e] Cell elongation by inducing direct binding to F-actin.

[f] Cell elongation by inducing abnormal actin polymerization but did not directly bind to F-actin.

2.12. Statistical analysis

All results were obtained from at least three separate experiments. Data were analyzed by one-way or two-way analysis of variance (ANOVA), and further post hoc tests using the statistic software of GraphPad Prism 4 (GraphPad software, Inc. San Diego, CA). A p value of <0.05 was considered as statistically significant in each experiment.

3. Results

3.1. PT-262 induces the cytoskeleton alteration and cell elongation in lung carcinoma cells

As shown in Fig. 1A, treatment with 2 μM PT-262 dramatically induced cytoskeleton alteration in A549 lung cancer cells. The cells were elongated and formed an abnormal cytoskeleton following
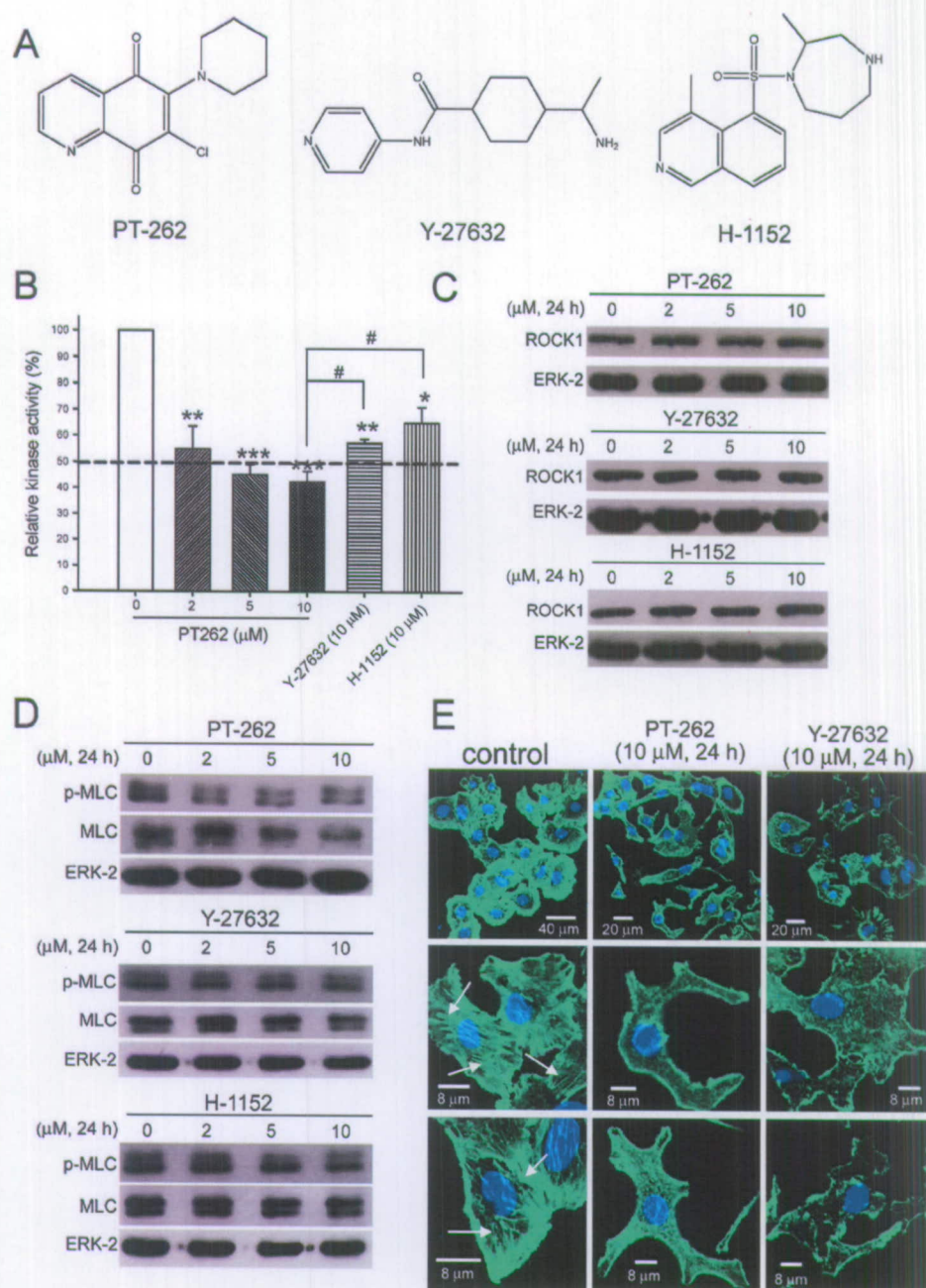

Fig. 3. Effect of PT-262 on the ROCK kinase activity, phospho-MLC and total-MLC expression, and stress fibers in lung carcinoma cells. (A) Chemical structures of PT-262, Y-27632, and H-1152. (B) A549 cells were treated with or without PT-262, Y-27632 and H-1152. At the end of the treatments, the cell lysates were subjected to ROCK kinase activity assays. The amount of phosphorylated MBS was calculated in a binding reaction with a horseradish peroxidase conjugated an anti-phospho-MBS threonine-696 specific antibody, which catalyzes the chromogenic substrate tetra-methylbenzidine to blue-color. Results were obtained from 3 separate experiments. The bar represents the mean ± S.E. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$ indicate significant difference between control and inhibitor treated samples. #$p < 0.05$ indicates a significant difference between PT-262 and ROCK inhibitor (Y-27632 or H-1152) treated samples. (C) and (D), A549 cells were treated with or without 2–10 μM PT-262, Y-27632, and H-1152. Representative Western blot data of the protein levels of ROCK1, phospho-MLC, and total MLC were shown from one of three separate experiments with similar findings. (E) F-actin proteins were stained with BODIPY FL phallacidin, which displayed a green color. The nuclei were stained with Hoechst 33258, which displayed a blue color. The arrows indicate the location of stress fibers. (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

exposure to PT-262. To view the elaborate structure of cell morphology, the cells were observed by confocal microscope combined Bio-AFM. The cytoskeleton alteration and cell elongation were clearly induced by PT-262 by comparison with the untreated sample (Fig. 1B). The green color exhibited by F-actin formed an elongated shape after treatment with 2 μM PT-262 for 24 h. The magnified pictures of the circle marks in Fig. 1B show the fork shape in the elongated cell (Fig. 1C). Cell length elongated to ~160 μm following treatment with 2 μM PT-262 (Fig. 1D). Moreover, PT-262 increased the formation of F-actin concentrated spikes (Fig. 1D, arrows). The average cell length was extended from 39.15 to 65.30 μm following treatment with PT-262 by calculating the long diameter (Fig. 1E).

3.2. Various cytoskeleton inhibitors induce cytoskeleton remodeling and morphological alteration

To investigate the cytoskeleton alteration by PT-262, A549 cells were compared with various cytoskeleton blockers, including paclitaxel, colchicine, and phallacidin. β-tubulin and F-actin were displayed by staining with Cy3-labeled mouse anti-β-tubulin and BODIPY FL phallacidin, respectively. Treatment with 50 nM paclitaxel for 24 h increased the red fluorescence intensity of β-tubulin by inducing the microtubule polymerization (Fig. 2). In contrast, colchicine (50 nM, 24 h) reduced the red intensity of β-tubulin by the inhibition of microtubulin polymerization (Fig. 2). Phallacidin (0.5 U/ml, 24 h) increased the green fluorescence
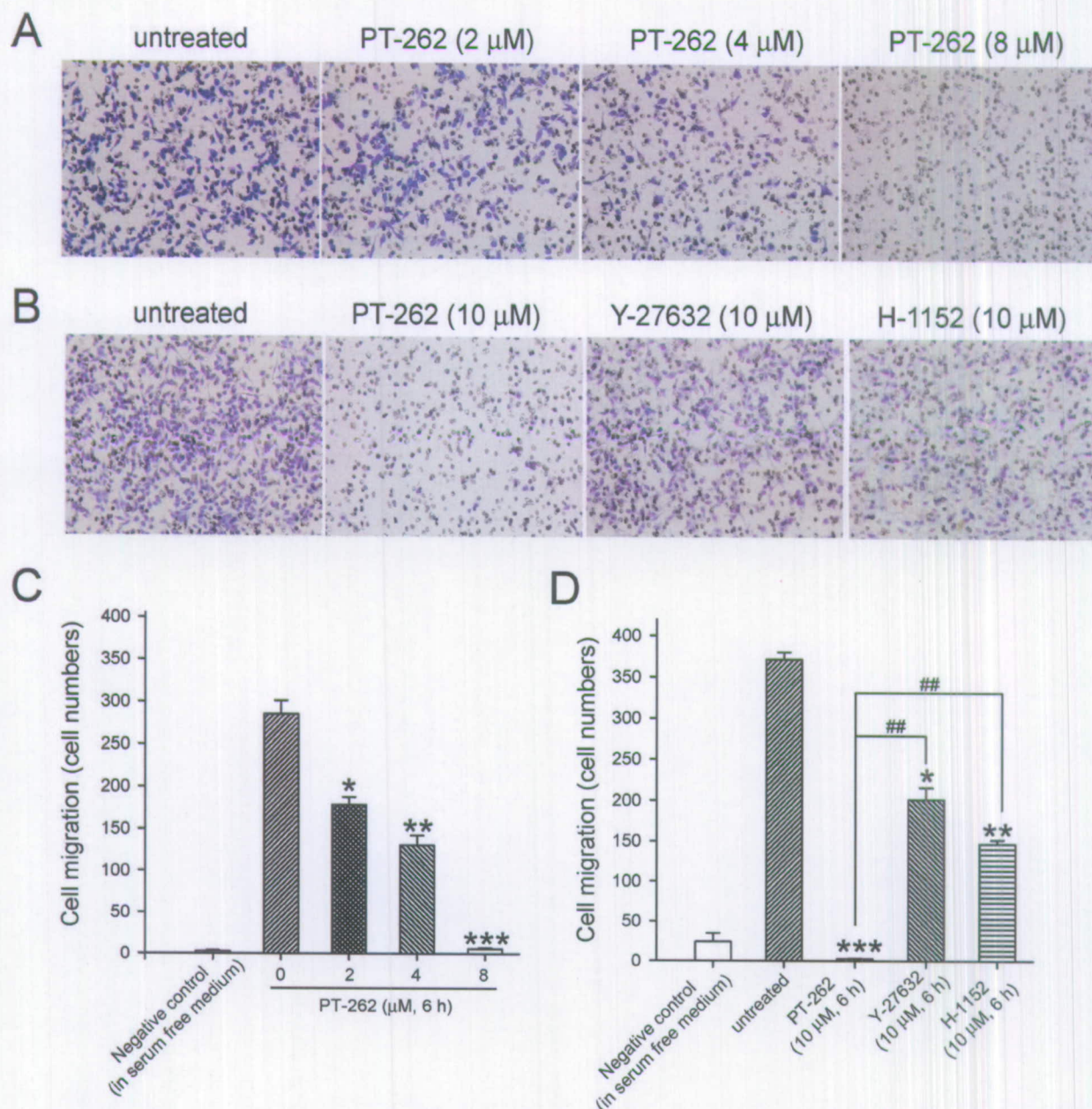
Fig. 4. Effect of PT-262 on cell migration by Boyden chamber analysis in lung carcinoma cells. (A) A549 cells were treated with or without 2–8 μM PT-262 for 6 h. (B) A549 cells were treated with or without 10 μM PT-262, Y-27632 or H-1152 for 6 h. At the end of treatment, the migrated cells on the lower surface of membranes were stained with hematoxylin and observed under a light microscope. Gray color spots indicate 8 μm pores on the polycarbonate membranes. The migrated cells indicate purple color. (C) and (D), the migrated cells on the lower surface were counted after stained with hematoxylin under a light microscope. Results were obtained from 3 separate experiments. The bar represents the mean ± S.E. $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$ indicate a significant difference between control and inhibitor treated samples. $##p < 0.05$ indicates a significant difference between PT-262 and ROCK inhibitor (Y-27632 or H-1152) treated samples. (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

intensity of F-actin by promoting actin polymerization and caused cell elongation in A549 cells. PT-262 was similar to phallacidin on the cell elongation (Fig. 2). Phallacidin (0.5 U/ml) increased 88.4% of actin polymerization than control using actin polymerization assays in vitro; nonetheless, treatment with 2–10 μM PT-262 did not alter actin polymerization. Cytochalasin B inhibited actin polymerization and cytokinesis but without inducing cell elongation in A549 cells [25]. The actions of these cytoskeleton inhibitors are summarized in Table 1.

3.3. PT-262 inhibits ROCK kinase activity and stress fiber formation

The chemical structures of PT-262, Y-27632, and H-1152 are shown in Fig. 3A. Treatment with 2–10 μM PT-262 inhibited ROCK kinase activities via a concentration-dependent manner in A549 cells (Fig. 3B). Furthermore, PT-262 was more effective on inhibiting ROCK kinase activities than Y-27632 and H-1152 (Fig. 3B, $p < 0.05$). The IC_{50} value (the concentration of 50% kinase activity inhibition) was around 5 μM by PT-262; the IC_{50} values of Y-27632 and H-1152 were >10 μM. Nonetheless, PT-262, Y-27632, and H-1152 did not reduce ROCK protein expression (Fig. 3C). ERK-2 protein has been used as an internal control [24,26]. The protein level of ERK-2 was not altered by PT-262 (Fig. 3C). In addition, treatment with 5–10 μM PT-262 decreased the levels of phospho-MLC and total MLC proteins (Fig. 3D). Y-27632 and H-1152 also decreased the MLC protein expression. The formation of stress fibers was found in the PT-262-untreated cells (Fig. 3E, arrows); in contrast, PT-262 or Y-27632 can block the stress fiber formation completely (Fig. 3E).

3.4. PT-262 inhibits cancer cell migration

To examine the effect of PT-262 on cancer cell migration, the cells were examined by Boyden chamber analysis. Treatment with 2–10 μM PT-262 for 6 h significantly blocked the cell migration in a concentration-dependent manner (Fig. 4A and C). The negative control of cell migration was used by incubating in serum free medium. Moreover, PT-262 was more effective in inhibiting cell migration than Y-27632 and H-1152 (Fig. 4B and Fig. 4D). The migration inhibition of PT-262 was verified by the wound-healing assay. As shown in Fig. 5A Fig. 5, the wound width by the cell migration was observed under a phase contrast microscope following treatment with 2 μM PT-262 for 8 h or 24 h. PT-262 significantly increased the wound width than untreated samples (Fig. 5B).

3.5. Blockage of RhoA expression inhibits cancer cell migration

Transfection with 20–80 nM RhoA siRNA for 48 h decreased RhoA gene expression and protein levels in A549 cells (Fig. 6A and C). The blockage of RhoA expression inhibited cell migration via a concentration-dependent manner (Fig. 6B). Furthermore, the phospho-MLC and total MLC proteins also were reduced by transfection with RhoA siRNA (Fig. 6B). The ROCK protein expression was not altered by RhoA siRNA (Fig. 6C). The effect of PT-262 on RhoA GTPase activity was further analyzed by RhoA pull-down assays. However, treatment with PT-262, Y-27632 and H-1152 did not significantly alter the RhoA GTPase activity and protein expression (Fig. 6D and E).

4. Discussion

The RhoA–ROCK pathway has been shown to promote cancer cell migration, invasion and metastasis. The RhoA–ROCK pathway is a potential target for cancer chemotherapy. In this study, we provide a newly synthesized compound, PT-262, that markedly

Magnification: 40X
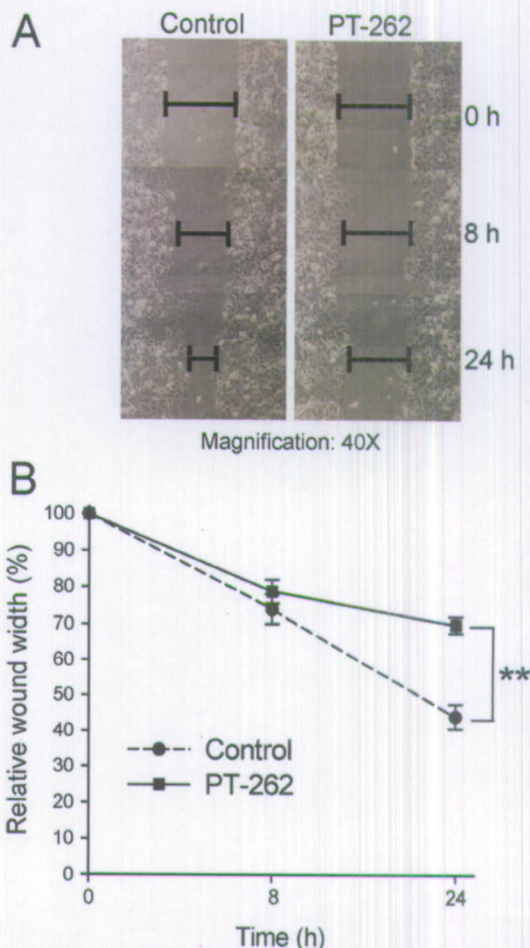
Fig. 5. Effect of PT-262 on cell migration by wound healing assay in lung carcinoma cells. (A) The confluent cell monolayer scraped by a pipette tip to generate wounds of 6–7 mm. Then the cells were treated with or without 8 μM PT-262 and then re-cultured for various periods. Photographs were taken at the same position of the wound. (B) Results were obtained from 3 separate experiments. The bar represents the mean ± S.E. **$p < 0.01$ indicates a significant difference between the control and PT-262 treated samples.

induces cytoskeleton remodeling, cell elongation, and migration inhibition in human lung carcinoma cells. The concentration range of PT-262 at 2–10 μM blocked ROCK kinase activity and cancer cell migration in a concentration-dependent manner. Interestingly, PT-262 is more effective in inhibiting the ROCK kinase activity and migration than ROCK inhibitors Y-27632 and H-1152. The findings provide that PT-262 is a novel and potent ROCK inhibitor.

Small RhoGTPases are key regulators of cytoskeleton dynamics [1,2]. RhoA is a key family of RhoGTPases that is involved in the regulation of cytoskeleton reorganization [5,6]. Overexpression of RhoA has been shown in various cancers [7]. ROCK is a downstream effector protein of RhoA [5]. The Rho–ROCK signaling pathway participates in cancer cell migration and transformation [27]. Blockage of RhoA expression can prevent cell migration in A549 lung cancer cells. However, PT-262 inhibited ROCK kinase activity but did not alter RhoA GTPase activity and protein expression. ROCK can regulate the activity of MLC proteins by direct MLC phosphorylation [9,10]. Cell migration of actomyosin contractility is mediated by the phosphorylation of MLC [1]. The inhibition of ROCK function by PT-262 reduced the protein levels of MLC and phosphorylated MLC. Also, PT-262 prevented stress fiber formation in lung cancer cells. Thus, these results show that PT-262 can prevent cancer cell migration by blocking the ROCK-MLC signaling pathway.
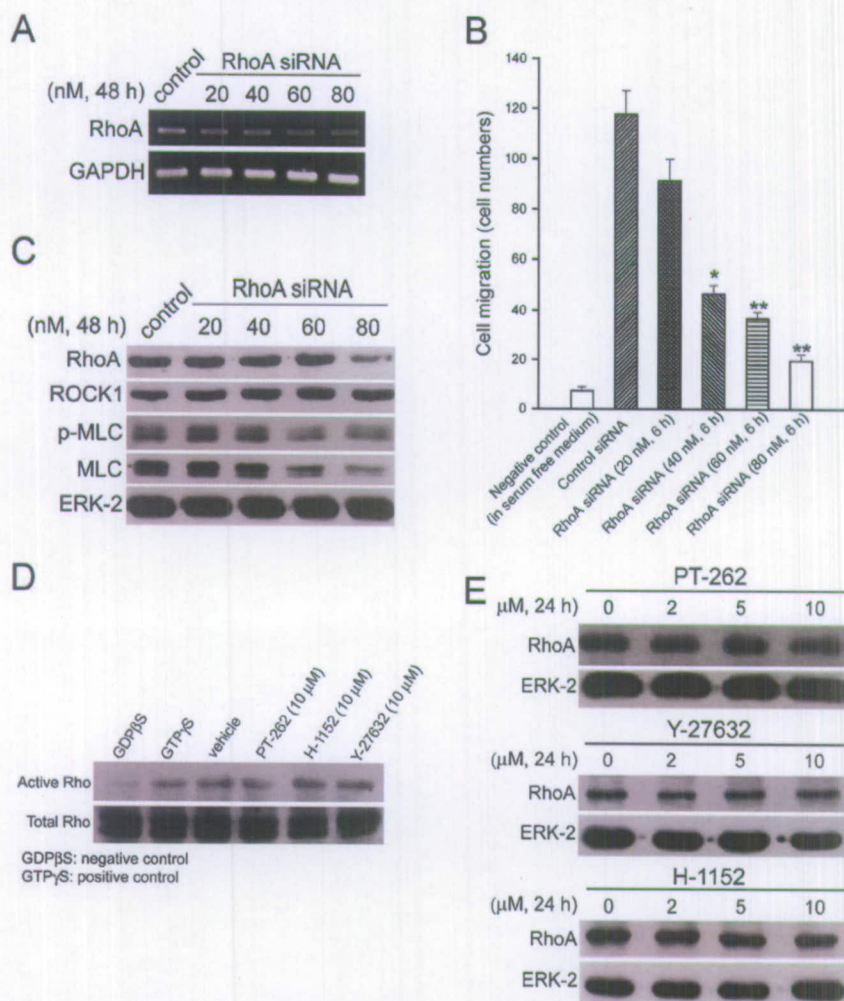
Fig. 6. Effect of PT-262 on RhoA GTPase activity and protein expression in lung carcinoma cells. (A) A549 cells were transfected with 20–80 nM control or RhoA siRNA for 48 h. The RhoA gene expression was analyzed by RT-PCR. GAPDH was an internal control gene. (B) The cell migration was examined by Boyden chamber analysis following transfection with control or RhoA siRNA. Results were obtained from 3 separate experiments. The bar represents the mean ± S.E. $^*p < 0.05$ and $^{**}p < 0.01$ indicate a significant difference between the control and RhoA siRNA treated samples. (C) The protein levels of RhoA, ROCK1, phospho-MCL, MLC, and ERK-2 were analyzed by Western blot. (D) RhoA GTPase activity was detected by a RhoA-GTP pull-down assay kit according to the manufacturer's instructions. A549 cells were treated with or without 10 μM PT-262, Y-27632 and H-1152. Activated RhoA-GTP and total RhoA in cell lysates were detected by Western blot using anti-RhoA. GDPβS and GTPγS were used as a negative control and positive control, respectively. (E) A549 cells were treated with or without 2–10 μM PT-262, Y-27632 and H-1152. At the end of treatment, the protein levels of RhoA and ERK-2 were analyzed by Western blot. The data are shown from one of three separate experiments with similar findings.

The cytoskeleton proteins of microtubules and F-actin are potential targets for cancer chemotherapy [2,28]. The disruption of F-actin has been found in malignant transformed cells [2]. Actin polymerization and remodeling play a critical role in the morphologic and phenotypic events in cancer cells [2,28]. Chemotherapeutic drugs that target cytoskeleton remodeling can prevent malignant transformation and metastasis processes. Paclitaxel is an effective anticancer drug that acts by stabilizing microtubules and inducing formation of microtubule bundles to block mitosis [29,30]. The vinca alkaloids and colchicine induce mitotic arrest by inhibiting microtubule polymerization and disassembling mitotic spindle [28,31]. Cytochalasins bind to the plus end of F-actin, reduce F-actin mass, and prevent actin polymerization [28,32]. Malignant cells are reported to be more sensitive to cytochalasin B than normal cells [28,33]. Phalloidins can bind to F-actin to inhibit actin depolymerization [32]. Among these cytoskeleton blockers, we found paclitaxel and colchicine induced microtubule polymerization and depolymerization, respectively. However, phallacidin promoted actin polymerization by stabilizing F-actin to induce the cell elongation in lung cancer cells. Interestingly, PT-262 increased cell elongation in a manner comparable to phallacidin treatment, indicating that PT-262 may

act in a similar fashion, promoting F-actin stability. Nonetheless, PT-262 did not bind to F-actin directly using actin polymerization *in vitro* assays. Therefore, we suggest that PT-262 blocks cytoskeleton functions indirectly by inhibiting ROCK signaling.

The ROCK1 protein crystal structure has been reported [10]. It has been shown that four crystal structures contain ATP-competitive inhibitors of ROCK1 in the third pocket site [10]. Fig. 7A shows a computational structure model of the interaction of PT-262 and ROCK1 protein. The 3D structures of ROCK inhibitors on the binding site of ROCK1 were utilized for docking analysis [10]. The coordinates of binding-pocket atoms were taken from the PDB for the docking tools such as GEMDOCK [34,35]. Furthermore, the potential binding sites were predicted by using Q-SiteFinder, which was an energy-based method for the prediction of protein–ligand binding sites [36,37]. We have further compared the binding of PT-262, H-1152, and Y-27632 to ROCK1 (Supplementary Fig. 1). The nitrogen atoms on the quinoline group of PT-262, the pyridine group of H-1152, and the isoquinoline group of Y-27632 all form hydrogen bonds with the nitrogen on the main chain of M156. Three functional groups occupy the same space as the adenine of ATP occupies (Supplementary Fig. 1). In addition, the chlorine and quinoline moieties of PT-262 contact with the MET153 amino
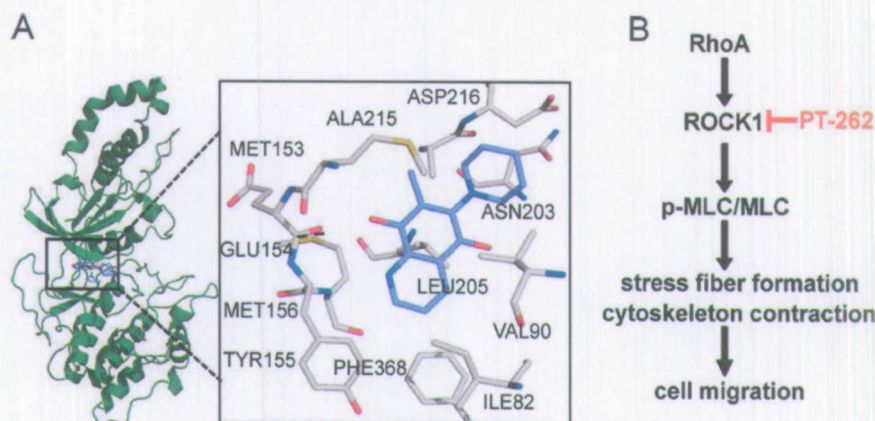
Fig. 7. A proposed model of PT-262 on the inhibition of ROCK. (A) A computational model of the interaction of PT-262 (blue color) and ROCK protein (green color). (B) The inhibition of ROCK signaling pathway by PT-262. (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.)

residue, and the chlorine and piperidin moieties of PT-262 contact with ALA215 amino residue, which may determine the selectivity at the ROCK inhibition. The chlorine of PT-262 may be the reason more effective on inhibition of ROCK activity than Y-27632 and H-1152. Accordingly, the computational model supports the findings that PT-262 blocks the ROCK kinase activities by binding on ATP-binding site.

The derivatives of quinolinediones have been shown to possess many biological activities including anti-tumor actions. For example, 6-anilino-5,8-quinolinedione (LY83583) is an inhibitor of guanylyl cyclase that can prevent cancer cell proliferation [20]. In addition, 6-chloro-7-(2-morpholin-4-ylethylamino) quinoline-5,8-dione (NSC 663284) is an inhibitor of CDC25 protein phosphatases that mediates cell cycle arrest by reducing CDC2 activation [21]. In this report, we provide a novel function of quinolinedione derivatives that is a ROCK inhibitor.

In conclusion, we presented a model of the inhibition of ROCK and its downstream regulation of cell migration by PT-262 (Fig. 7B). The model provides a novel mechanism of cell migration inhibition by the 5,8-quinolinedione derivative through the irreversible effects of cytoskeleton alteration and cell elongation in blocking ROCK pathway.

Acknowledgment

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bcp.2011.01.009.

References

[1] Hall A. Rho GTPases and the actin cytoskeleton. Science 1998;279:509–14.
[2] Rao J, Li N. Microfilament actin remodeling as a potential target for cancer drug development. Curr Cancer Drug Targets 2004;4:345–54.
[3] Hall A. Signal transduction through small GTPases—a tale of two GAPs. Cell 1992;69:389–91.
[4] Ridley AJ, Hall A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. Cell 1992;70:389–99.
[5] Wyckoff JB, Pinner SE, Gschmeissner S, Condeelis JS, Sahai E. ROCK- and myosin-dependent matrix deformation enables protease-independent tumor-cell invasion in vivo. Curr Biol 2006;16:1515–23.
[6] Narumiya S, Tanji M, Ishizaki T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. Cancer Metastasis Rev 2009;28:65–76.
[7] Sahai E, Marshall CJ. RHO-GTPases and cancer. Nat Rev Cancer 2002;2:133–42.
[8] Leung T, Chen XQ, Manser E, Lim L. The p160 RhoA-binding kinase ROK alpha is a member of a kinase family and is involved in the reorganization of the cytoskeleton. Mol Cell Biol 1996;16:5313–27.
[9] Kawano Y, Fukata Y, Oshiro N, Amano M, Nakamura T, Ito M, et al. Phosphorylation of myosin-binding subunit (MBS) of myosin phosphatase by Rho-kinase in vivo. J Cell Biol 1999;147:1023–38.
[10] Jacobs M, Hayakawa K, Swenson L, Bellon S, Fleming M, Taslimi P, et al. The structure of dimeric ROCK I reveals the mechanism for ligand selectivity. J Biol Chem 2006;281:260–8.
[11] Amano M, Chihara K, Kimura K, Fukata Y, Nakamura N, Matsuura Y, et al. Formation of actin stress fibers and focal adhesions enhanced by Rho-kinase. Science 1997;275:1308–11.
[12] Liu S, Goldstein RH, Scepansky EM, Rosenblatt M. Inhibition of rho-associated kinase signaling prevents breast cancer metastasis to human bone. Cancer Res 2009;69:8742–51.
[13] Uehata M, Ishizaki T, Satoh H, Ono T, Kawahara T, Morishita T, et al. Calcium sensitization of smooth muscle mediated by a Rho-associated protein kinase in hypertension. Nature 1997;389:990–4.
[14] Ishizaki T, Uehata M, Tamechika I, Keel J, Nonomura K, Maekawa M, et al. Pharmacological properties of Y-27632, a specific inhibitor of rho-associated kinases. Mol Pharmacol 2000;57:976–83.
[15] Takamura M, Sakamoto M, Genda T, Ichida T, Asakura H, Hirohashi S. Inhibition of intrahepatic metastasis of human hepatocellular carcinoma by Rho-associated protein kinase inhibitor Y-27632. Hepatology 2001;33:577–81.
[16] Somlyo AV, Bradshaw D, Ramos S, Murphy C, Myers CE, Somlyo AP. Rho-kinase inhibitor retards migration and in vivo dissemination of human prostate cancer cells. Biochem Biophys Res Commun 2000;269:652–9.
[17] Chang HR, Huang HP, Kao YL, Chen SL, Wu SW, Hung TW, et al. The suppressive effect of Rho kinase inhibitor, Y-27632, on oncogenic Ras/RhoA induced invasion/migration of human bladder cancer TSGH cells. Chem Biol Interact 2010;183:172–80.
[18] Sasaki Y, Suzuki M, Hidaka H. The novel and specific Rho-kinase inhibitor (S)-(+)-2-methyl-1-[(4-methyl-5-isoquinoline)sulfonyl]-homopiperazine as a probing molecule for Rho-kinase-involved pathway. Pharmacol Ther 2002;93:225–32.
[19] Yoon E, Choi HY, Shin KJ, Yoo KH, Chi DI, Kim DJ. The regioselectivity in the reaction of 6,7-dihaloquinoline-5,8-diones with amine nucleophiles in various solvents. Tetrahedron Lett 2000;41:7475–80.
[20] Lodygin D, Menssen A, Hermeking H. Induction of the Cdk inhibitor p21 by LY83583 inhibits tumor cell proliferation in a p53-independent manner. J Clin Invest 2002;110:1717–27.
[21] Lazo JS, Aslan DC, Southwick EC, Cooley KA, Ducruet AP, Joo B, et al. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. J Med Chem 2001;44:4042–9.
[22] Fang Y, Linardic CM, Richardson DA, Cai W, Behforouz M, Abraham RT. Characterization of the cytotoxic activities of novel analogues of the antitumor agent, lavendamycin. Mol Cancer Ther 2003;2:517–26.
[23] Hsu TS, Chen C, Lee PT, Chiu SJ, Liu HF, Tsai CC, et al. 7-Chloro-6-piperidin-1-yl-quinoline-5,8-dione (PT-262), a novel synthetic compound induces lung carcinoma cell death associated with inhibiting ERK and CDC2 phosphorylation via a p53-independent pathway. Cancer Chemother Pharmacol 2008;62:799–808.
[24] Chao JI, Kuo PC, Hsu TS. Down-regulation of survivin in nitric oxide-induced cell growth inhibition and apoptosis of the human lung carcinoma cells. J Biol Chem 2004;279:20267–76.
[25] Chao JI, Liu HF. The blockage of survivin and securin expression increases the cytochalasin B-induced cell death and growth inhibition in human cancer cells. Mol Pharmacol 2006;69:154–64.
[26] Kuo PC, Liu HF, Chao JI. Survivin and p53 modulate quercetin-induced cell growth inhibition and apoptosis in human lung carcinoma cells. J Biol Chem 2004;279:55875–8.
[27] Li B, Zhao WD, Tan ZM, Fang WG, Zhu L, Chen YH. Involvement of Rho/ROCK signalling in small cell lung cancer migration through human brain microvascular endothelial cells. FEBS Lett 2006;580:4252–60.

[28] Jordan MA, Wilson L. Microtubules and actin filaments: dynamic targets for cancer chemotherapy. Curr Opin Cell Biol 1998;10:123–30.

[29] Schiff PB, Horwitz SB. Taxol stabilizes microtubules in mouse fibroblast cells. Proc Natl Acad Sci USA 1980;77:1561–5.

[30] Jordan MA, Toso RJ, Thrower D, Wilson L. Mechanism of mitotic block and inhibition of cell proliferation by taxol at low concentrations. Proc Natl Acad Sci USA 1993;90:9552–6.

[31] Dhamodharan R, Jordan MA, Thrower D, Wilson L, Wadsworth P. Vinblastine suppresses dynamics of individual microtubules in living interphase cells. Mol Biol Cell 1995;6:1215–29.

[32] Cooper JA. Effects of cytochalasin and phalloidin on actin. J Cell Biol 1987;105:1473–8.

[33] Stournaras C, Stiakaki E, Koukouritaki SB, Theodoropoulos PA, Kalmanti M, Fostinis Y, et al. Altered actin polymerization dynamics in various malignant cell types: evidence for differential sensitivity to cytochalasin B. Biochem Pharmacol 1996;52:1339–46.

[34] Yang JM, Chen CC. GEMDOCK: a generic evolutionary method for molecular docking. Proteins 2004;55:288–304.

[35] Yang JM. Development and evaluation of a generic evolutionary method for protein–ligand docking. J Comput Chem 2004;25:843–57.

[36] Laurie AT, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. Bioinformatics 2005;21:1908–16.

[37] Huang B, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. BMC Struct Biol 2006;6:19.

PAComplex: a web server to infer peptide antigen families and binding models from TCR–pMHC complexes

I-Hsin Liu[1], Yu-Shu Lo[1] and Jinn-Moon Yang[1,2,]*

[1]Institute of Bioinformatics and Systems Biology and [2]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan

ABSTRACT

One of the most adaptive immune responses is triggered by specific T-cell receptors (TCR) binding to peptide-major histocompatibility complexes (pMHC). Despite the availability of many prediction servers to identify peptides binding to MHC, these servers are often lacking in peptide–TCR interactions and detailed atomic interacting models. PAComplex is the first web server investigating both pMHC and peptide-TCR interfaces to infer peptide antigens and homologous peptide antigens of a query. This server first identifies significantly similar TCR–pMHC templates (joint Z-value ≥ 4.0) of the query by using antibody–antigen and protein–protein interacting scoring matrices for peptide-TCR and pMHC interfaces, respectively. PAComplex then identifies the homologous peptide antigens of these hit templates from complete pathogen genome databases ($\geq 10^8$ peptide candidates from 864 628 protein sequences of 389 pathogens) and experimental peptide databases (80 057 peptides in 2287 species). Finally, the server outputs peptide antigens and homologous peptide antigens of the query and displays detailed interacting models (e.g. hydrogen bonds and steric interactions in two interfaces) of hitTCR-pMHC templates. Experimental results demonstrate that the proposed server can achieve high prediction accuracy and offer potential peptide antigens across pathogens. We believe that the server is able to provide valuable insights for the peptide vaccine and MHC restriction. The PAComplex sever is available at http://PAcomplex.life.nctu.edu.tw.

INTRODUCTION

An immune system protects an organism from diseases by identifying and killing pathogens (1). One of the most adaptive immune responses is triggered by specific T-cell receptors (TCRs) binding to peptide-major histocompatibility complexes (pMHC) molecules. An increasing number of available binding peptide antigens that are reliable (2–4) and high-throughput experiments that provide systematic identification of pMHC interactions explain the growing requirement for fast and accurate computational methods for discovering homologous peptide antigens of a new peptide antigen and developing peptide-based vaccines for pathogens.

Many methods have been proposed for predicting pMHC interactions. These methods can be roughly divided into the sequence-based methods such as motif matching (5,6), matrix methods [e.g. SYFPEITHI (7), MAPPP (8), IEDB (9)] and machine learning approaches [e.g. SVMHC (10)]; and structure-based approaches [e.g. PREDEP (11) and MODPROPEP (12)]. However, these methods are often lack of the TCR and pMHC binding, which is critical to trigger adaptive immune responses. Since the increasing number of TCR–pMHC crystal structures to investigate both pMHC and peptide-TCR interfaces provides further insights for understanding TCR–pMHC interactions and binding mechanisms. Additionally, discovering homologous peptide antigens (called peptide antigen family) to a known peptide antigen often provides a valuable reference for efforts to elucidate the functions of a new peptide antigen.

To address these issues, we propose the PAComplex server for predicting TCR–pMHC interactions and inferring antigen families across organisms of a query protein or a set of peptides. To our best knowledge, PAComplex is the first web server investigating both pMHC and peptide-TCR interfaces to infer peptide antigens and homologous peptide antigens of a query. Additionally, peptide antigen families are derived from a complete pathogen genome database ($\geq 10^8$ peptide candidates from 389 pathogens) and experimental peptide databases to demonstrate the feasibility of the PAComplex server and increase the number of potential antigens. Moreover, for a peptide antigen family, the amino acid composition and conservation are

*To whom correspondence should be addressed. Tel: 886 3 5712121 (ext. 56942); Fax: 886 3 5729288; Email: moon@faculty.nctu.edu.tw

evaluated at each position. Experimental results demonstrate that the server can improve the peptide antigen prediction accuracy and is useful for identifying peptide antigen families by using two interfaces of TCR–pMHC structures. Furthermore, the proposed server provides a valuable reference for efforts to develop peptide vaccines and elucidate MHC restriction and T-cell activation.

METHOD AND IMPLEMENTATION

Homologous peptide antigen

The concept of homologous peptide antigen is the core of this server. We define the homologous peptide antigen (p') of the peptide (p) in template complex as follows: (i) p and p' can be bound by the same MHC forming pMHC and p'MHC, respectively, with the significant interface similarity ($Z_{MHC} \geq 1.645$); (ii) pMHC and p'MHC can be recognized by the same TCR with significant peptide-TCR interface similarity ($Z_{TCR} \geq 1.645$); and (iii) TCR-pMHC and TCR-p'MHC share significant complex similarity (joint $Z \geq 4.0$). The joint Z-value (J_z) is defined as

$$J_z = \sqrt{Z_{MHC} \times Z_{TCR}} \qquad (1)$$

The Z_{MHC} and Z_{TCR} of a TCR-p'MHC candidate with interaction score (E) can be calculated by $(E-\mu)/\sigma$, where μ is the mean and σ is the standard deviation from 10 000 random interfaces (Supplementary Figure S1). For a TCR–pMHC template collected from Protein Data Bank (PDB), these 10 000 random interfaces are generated by substituting with another amino acid according to the amino acid composition derived from UniProt (13). Here, $J_Z \geq 4.0$ is considered a significant similarity according to the statistical analysis of 41 TCR–pMHC structure complexes; 80 057 experimental peptide antigens; and $\geq 10^8$ peptide candidates derived from 864 628 protein sequences in 389 pathogens.

Template-based scoring function

We have recently proposed a template-based scoring function to determine the reliability of protein–protein interactions derived from a 3D-dimer structure (14). For measuring the pMHC interaction score, the scoring function is defined as

$$E_{tot} = E_{VDW} + E_{SP} + E_{sim} \qquad (2)$$

Where E_{VDW} and E_{SP} denote steric force and special energy (i.e. hydrogen bond energy and electrostatic energy), respectively, according to four knowledge-based scoring matrices (14) which have a good achievement between pMHC and protein–protein interactions. E_{sim} refers to the peptide similarity score between p and p'.

To model E_{VDW} and E_{SP} of the peptide-TCR interactions, we developed a new residue-based matrix (Supplementary Figure S2) because the peptide-TCR interface resembles antigen–antibody interactions and differs from protein–protein interfaces (15,16). The matrix is derived from a non-redundant set which consists of 62 structural antigen–antibody complexes (including 131 interfaces) constructed by Ponomarenko *et al.* (17).

According to this matrix, the peptide-TCR (antigen–antibody) interface prefers aromatic residues (i.e. Phe, Trp and Tyr), which interact with aliphatic residues (i.e. Ala, Val, Leu, Ile and Met) or long side-chain polar residues (i.e. Gln, His, Arg, Lys and Glu), to form strong van der Waals (VDW) forces (yellow boxes). Additionally, the scores are high if basic residues (i.e. Arg and Lys) interact to acidic residues (i.e. Asp and Glu). Conversely, the scores are low (purple box) when non-polar residues interact with polar residues.
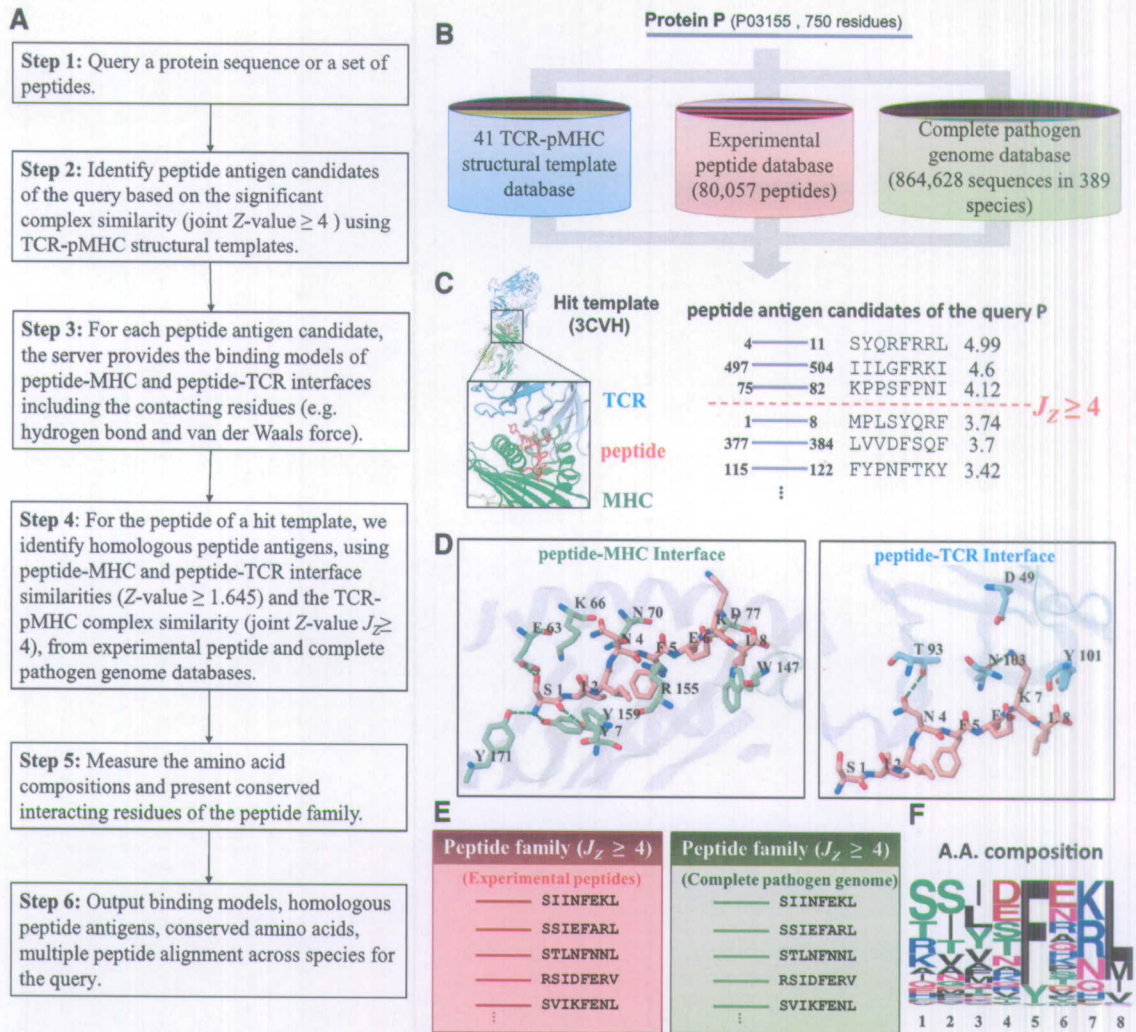
Overview

Figure 1 shows the details of the PAComplex server to predict peptide antigens and search the template-based homologous peptide antigens of a query protein sequence (or a set of peptides) by the following steps (Figure 1A). The server initially divides the query protein sequence into fix length (ranging from 8 to 13) peptides based on selected MHC class I allele and templates. Each peptide (p') is then aligned to the bound peptide (p) of TCR–pMHC templates collected from PDB. Next, the peptide antigen is examined by utilizing the template-based scoring function to statistically evaluate the complex similarity ($J_z \geq 4.0$) between TCR–pMHC and TCR-p'MHC (Figure 1B and C). For each peptide antigen, the server introduces the potential TCR–pMHC binding models and the detailed residues interactions (e.g. hydrogen bonds and VDW forces) of pMHC and peptide-TCR interfaces (Figure 1D). For the hit templates, the server identifies the homologous peptide antigens with $J_z \geq 4.0$ from an experimental peptide database (80 057 peptides in 2287 species) and a complete pathogen genome database ($\geq 10^8$ peptide antigen candidates with $J_z \geq 1.645$ derived from 864 628 protein sequences of 389 pathogens) (Figure 1B and E). For a peptide antigen family, we measure the amino acid composition and conservation at each position (Figure 1F) by WebLogo program (18). Finally, this server provides peptide antigens, visualization of the TCR–pMHC interaction models, and peptide antigen families with conserved amino acids.

INPUT, OUTPUT AND OPTIONS

The PAComplex server is easy to use (Figure 2). Users input a protein sequence in FASTA format (or a set of fix-length peptides) and select the parameters (e.g. MHC class I allele and templates) (Figure 2A). The PAComplex server typically infers peptide antigens and homologous peptide antigens of the query within 4 s if the sequence length is ≤300. For a query, PAComplex shows the detailed atomic interactions and binding models using Jmol and amino acid profiles (Figure 2C) of homologous peptide antigens from experimental peptide (Figure 2E) and complete pathogen genome databases (Figure 2D). For each peptide antigen, PAComplex also presents the source proteins, organisms and experimental data. In addition, users can download summarized results of query peptides or protein sequences, the modeling TCR–pMHC complex, template structure of TCR–pMHC and peptide family of the template.

Figure 1. Overview of the PAComplex server for peptide antigens and homologous peptide antigens search using protein P of HBV as the query. (**A**) Main procedure. (**B**) Template-based scoring function to infer the peptide antigens and homologous peptide antigens through structural templates, experimental peptides and complete pathogen genome databases. (**C**) Peptide antigen candidates of the query using hit TCR–pMHC complex templates. (**D**) Atomic binding models with hydrogen bonds (green dash lines) of both pMHC and peptide-TCR interfaces. (**E**) Peptide antigen families of the query from the experimental peptide and complete pathogen genome databases. (**F**) Amino acid compositions (profiles) of the homologous peptide antigens.

Example analysis

Protein P of hepatitis B virus. While affecting over 350 million people worldwide, *hepatitis B virus* (HBV) infection is a leading cause of liver diseases and hepatocellular carcinoma (19,20). Figure 1 shows the PAComplex derived results using protein P [UniProt (13) accession number: P03155, 750 residues divided into 743 8-mer peptides] of HBV *genotype D* as the query. Protein P, a multifunctional enzyme, converts the viral RNA genome into dsDNA in viral cytoplasmic capsids. This enzyme displays a DNA polymerase activity that can replicate either DNA or RNA templates, and a ribonuclease H (RNase H) activity that cleaves the RNA strand of RNA–DNA heteroduplexes in a partially processive 3′- to 5′-endonucleasic mode (21,22). For this query, the PAComplex server found three hit peptide antigen candidates ($J_z \geq 4.0$; Figure 1C) and 73 homologous peptide antigens in 21 organisms (Supplementary Figure S3A) by using H-2K^b-peptide-TCR template [PDB entry 3 CVH (23)] and the experimental peptide database. Among these three hit peptides, the peptide 497–504 (IILGFRKI) recorded in IEDB (4) is the epitope of protein P and PAComplex presents its binding models and detailed residue interactions of peptide-TCR and pMHC interfaces (Figure 1D). Position 1of the homologous peptide antigens prefers the polar residues (e.g. Ser, Thr, Arg and Lys; Figure 1F) and the first position of this peptide (pink) is polar residue Ser forming five hydrogen bonds with residues Tyr7, Glu63 and Tyr171 on MHC molecule (green) (Figure 1D). Additionally, position 7 (Lys) of the homologous peptide antigens prefers the positive residues (Arg and Lys, Figure 1F) and Lys7 of
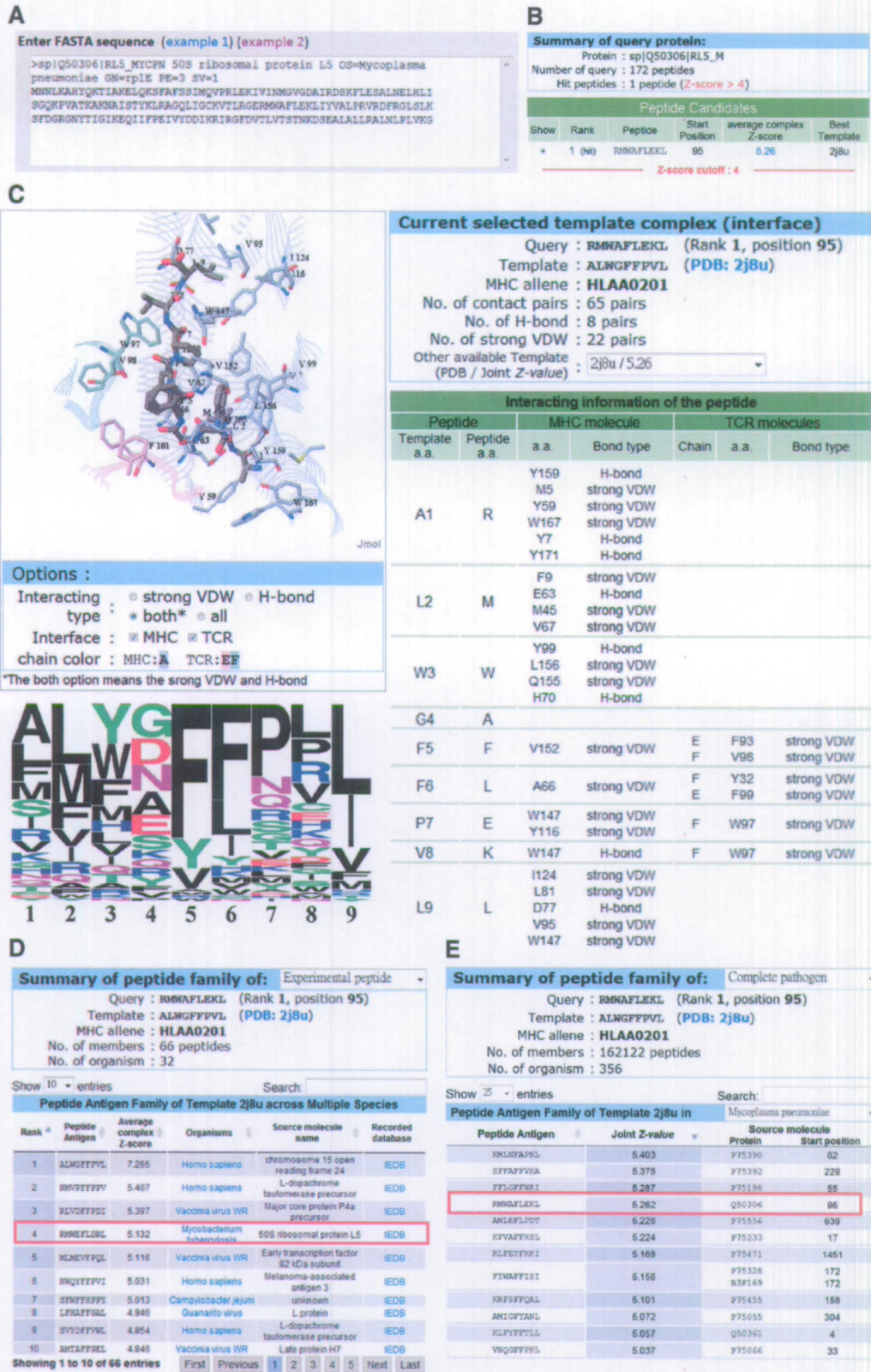
Figure 2. PAComplex server search results using 50S ribosomal protein L5 (rplE) of *M. pneumonia* as the query. (**A**) User interface for inputting the query protein sequence, MHC class I allele, and templates. (**B**) The peptide antigen candidates ($J_z \geq 4.0$) of the query. (**C**) Detailed atomic interactions and binding models with hydrogen bond residues and strong VDW forces. Peptide antigen families from (**D**) experimental peptide database and (**E**) complete pathogen genome database, respectively.

this hit peptide forms electrostatic interactions with Asp49 in TCR.

Two other hit peptide antigens 4–11 (SYQRFRRL) and 75–82 (KPPSFPNI) correlate well two homologous peptide antigens (SYQHFRKL and KTPSFPNI), which are epitopes of HBV *alpha 1* recorded in IEDB, respectively (orange box in Supplementary Figure S3A). According to the amino acid composition (profile) of this peptide antigen family (Figure 1F), position 7 prefers positive charged residues and positions 5 and 8 prefer non-polar residues. Conversely, the compositions of Positions 2 and 4 are diverse. These two hit antigens match the profile of the antigen family on Positions 1 (polar residues), 5 (conserved residue Phe), 7 (positive or polar residues) and 8 (non-polar residues). For instance, the two hit antigens are conserved on the Position 5 with residue Phe forming strong VDW interactions with MHC [Phe74, Val97, Tyr22, Val9 and Tyr116) and TCR (Phe104) molecules (Supplementary Figure S3B)]. These residue–residue interactions (i.e. Phe-Phe, Phe-Val and Phe-Try) are high scores according to pMHC (14) and peptide-TCR (Supplementary Figure S2) scoring matrices. For peptides SYQRFRRL and SYQHFRKL, they have the positively charged residue type (e.g. Arg and Lys) on Position 7 and different residue types on Position 4. For peptides KPPSFPNI and KTPSFPNI, the only different residue type is located on Position 2. Therefore, these two hit peptides are potential peptide antigens. These results suggest that investigating multiple TCR–pMHC interfaces and the peptide antigen family are useful for predicting peptide antigens and providing valuable insight into MHC restriction and T-cell activation.

50S ribosomal protein L5 (rplE) of Mycoplasma pneumonia. 50S ribosomal protein L5 (rplE), interacting with 5S rRNA and tRNA, is an essential protein of *M. pneumonia,* which is the cause of human walking pneumonia (24). Based on use of the *M. pneumonia* rplE (Q50306, 180 residues are divided into 172 9-mer peptides) as the query (Figure 2A), the PAComplex server infers one hit candidate ($J_z \geq 4.0$; Figure 2B), 95–103 (RMWAFLEKL) and its 66 homologous peptide antigens in 32 organisms, based on the HLA-A0201-peptide-TCR template [PDB entry 2J8U (25); Figure 2D]. This server provides the binding model (Figure 2C) and homologous peptide antigens via the experimental peptide database and complete pathogen genome database (Figure 2D and E).

The hit candidate is similar to the Rank 4 peptide (RMWEFLDRL, red box) in the peptide family (Figure 2D). But they have three different amino acid types on Positions 4, 7 and 8 whose amino acid compositions of this family are diverse (Figure 2C). Based on binding models and interactions, Position 4 lacks any hydrogen bonds and strong VDW contacts. On the other hand, the hit peptide correlates well with the amino acid profile on the conserved positions (i.e. 1, 2, 3, 5, 6, 9) forming strong VDW forces (the right-side table of Figure 2C) based on the interactions of both peptide-TCR and pMHC interfaces in HLA-A0201-peptide-TCR template (2J8U). Above results imply that

the hit peptide of rplE is a potential antigen-activating immune response. Furthermore, PAComplex provides the potential peptide antigens derived from all proteins of the query *M. pneumoniae* (Figure 2E) and the other 388 pathogens using a complete pathogen genome database. These potential peptide antigens across pathogens can be useful in identifying specific peptides for the target pathogen for vaccine design.

RESULTS

To evaluate the performance of PAComplex for identifying the peptide antigens and peptide antigen families, we selected two peptide sets, termed BothMT (Figure 3A) and CPD (Figure 3B). BothMT consists of 86 positive and 67 negative octamers with experimental data of both H-2K^b and TCR sides collected from IEDB (4). PAComplex aligned these 153 peptides to six H-2K^b-peptide-TCR complex templates extracted from PDB released on 25 December 2010 to evaluate the accuracies of scoring functions on variant conditions (e.g. single template, multiple templates, single side and both sides). The CPD set, which comprises $\geq 10^8$ peptide candidates ($J_Z \geq 1.645$) derived from 864 628 protein sequences of 389 pathogens, was used to evaluate the reliability of homologous peptide antigens and it was collected by the following steps: (i) extract 389 pathogens (e.g. bacteria, archaea and virus) recorded in both IEDB and UniProt (13) databases and their respective complete genomes collected from UniProt database (13); (ii) derive the positive and negative data sets from IEDB for these pathogens; and (iii) extract 41 TCR–pMHC complexes from PDB.

Figure 3A illustrates the receiver operating characteristic (ROC) curves (i.e. true positive and false positive rates) of our scoring functions on single and multiple templates using one interface (i.e. pMHC and peptide-TCR interfaces) and two interfaces (i.e. TCR–pMHC complex). We observed several interesting results: (i) the scoring function using pMHC interface (blue lines) yields a higher accuracy than using peptide-TCR interface (green lines); (ii) using multiple templates (solid lines) is better than using single template (dot lines); and (iii) using two interfaces with multiple templates (red) is the best among these six combinations.

Next, the J_Z threshold for reliable homologous peptide antigens is determined by evaluating the PAComplex server on the large-scale CPD data set (Figure 3B). This server was tested on $>10^{10}$ peptides derived from 864 628 protein sequences of 389 pathogens. Among these peptides, over 10^8 peptide candidates with $J_Z \geq 1.645$ were selected for analyzing the relationships between J_Z values with both the numbers of positive homologous peptide antigens (blue, recorded in IEDB) and precision (red). When J_Z is higher than 4.0, the precision >0.8 and the number of positive antigens exceeds 1600 according to the positive and negative data sets. If the J_Z threshold is set to 4.0, the total number of inferring possible peptide antigens surpasses 4 000 000 (Supplementary Figure S4) statistically derived from 41 TCR–pMHC complexes. The amino acid compositions (profiles) of these 1600 positive peptide
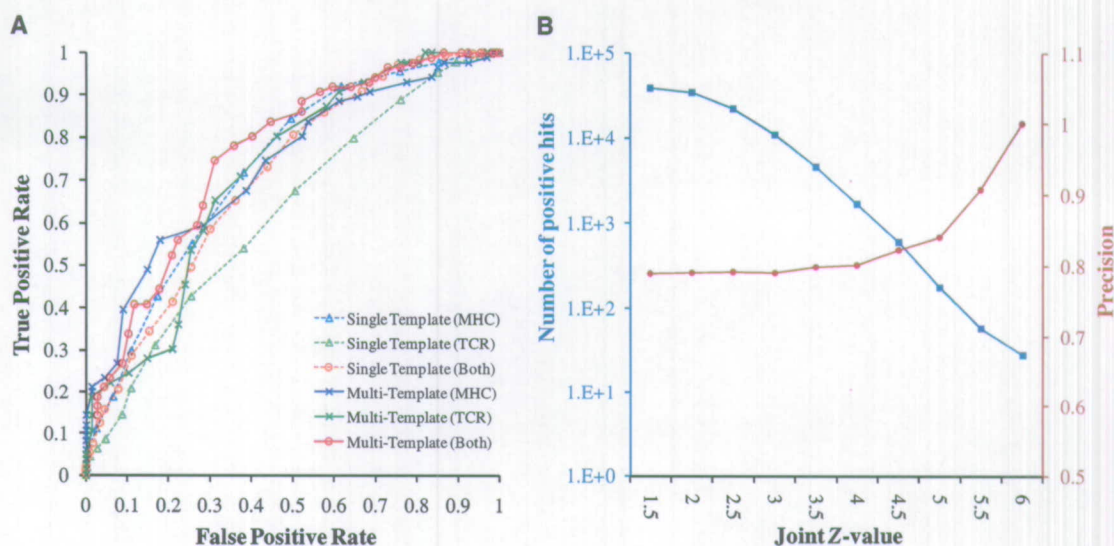
Figure 3. Evaluations of the PAComplex server on BothMT and CPD sets. (**A**) ROC curves of PAComplex using single and multiple templates with one interface (pMHC and peptide-TCR) and two interfaces of TCR–pMHC complex using BothMT set. (**B**) Relationship between the distribution of positive hits (blue line) and precision values (red line) with different joint Z-value thresholds using CPD set.

antigens closely correspond to the ones obtained from peptide antigen families (4 000 000 antigens). These experimental results demonstrate that this server achieves high accuracy and is able to provide potential peptide antigens across pathogens.

CONCLUSIONS

This work demonstrates the feasibility of using the PAComplex server to identify peptide antigens and homologous peptide antigens. The proposed server provides detailed atomic interactions, binding models, amino acid compositions of peptide families, source proteins and organisms and experimental data. PAComplex server is the first to infer peptide antigens and homologous peptide antigens by considering two TCR–pMHC interfaces from complete pathogen genome and experimental peptide databases. Experimental results demonstrate that the server is highly accurate and capable of providing potential peptide antigens across pathogens. We believe that PAComplex is a fast homologous peptide antigens search server and is able to provide valuable insights into the peptide vaccine, MHC restriction and T-cell activation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Authors are grateful to both the hardware and software supports of the Structural Bioinformatics Core Facility at National Chiao Tung University.

REFERENCES

1. Marrack,P., Scott-Browne,J.P., Dai,S., Gapin,L. and Kappler,J.W. (2008) Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.*, **26**, 171–203.
2. Bhasin,M., Singh,H. and Raghava,G.P. (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics*, **19**, 665–666.
3. Brusic,V., Rudy,G. and Harrison,L.C. (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.*, **26**, 368–371.
4. Vita,R., Zarebski,L., Greenbaum,J.A., Emami,H., Hoof,I., Salimi,N., Damle,R., Sette,A. and Peters,B. (2010) The immune epitope database 2.0. *Nucleic Acids Res.*, **38**, D854–D862.
5. Rudensky,A., Preston-Hurlburt,P., Hong,S.C., Barlow,A. and Janeway,C.A. Jr (1991) Sequence analysis of peptides bound to MHC class II molecules. *Nature*, **353**, 622–627.
6. Hammer,J., Valsasnini,P., Tolba,K., Bolin,D., Higelin,J., Takacs,B. and Sinigaglia,F. (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. *Cell*, **74**, 197–203.
7. Schuler,M.M., Nastke,M.D. and Stevanovikc,S. (2007) SYFPEITHI: database for searching and T-cell epitope prediction. *Methods Mol. Biol.*, **409**, 75–93.
8. Hakenberg,J., Nussbaum,A.K., Schild,H., Rammensee,H.G., Kuttler,C., Holzhutter,H.G., Kloetzel,P.M., Kaufmann,S.H. and Mollenkopf,H.J. (2003) MAPPP: MHC class I antigenic peptide processing prediction. *Appl. Bioinformatics*, **2**, 155–158.
9. Peters,B. and Sette,A. (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*, **6**, 132.

10. Donnes,P. and Kohlbacher,O. (2006) SVMHC: a server for prediction of MHC-binding peptides. *Nucleic Acids Res.*, **34**, W194–W197.

11. Altuvia,Y., Sette,A., Sidney,J., Southwood,S. and Margalit,H. (1997) A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. *Hum. Immunol.*, **58**, 1–11.

12. Kumar,N. and Mohanty,D. (2007) MODPROPEP: a program for knowledge-based modeling of protein-peptide complexes. *Nucleic Acids Res.*, **35**, W549–W555.

13. Apweiler,R., Martin,M.J., O'Donovan,C., Magrane,M., Alam-Faruque,Y., Antunes,R., Barrell,D., Bely,B., Bingley,M., Binns,D. *et al.* (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.

14. Chen,Y.C., Lo,Y.S., Hsu,W.C. and Yang,J.M. (2007) 3D-partner: a web server to infer interacting partners and binding models. *Nucleic Acids Res.*, **35**, W561–W567.

15. Rudolph,M.G., Luz,J.G. and Wilson,I.A. (2002) Structural and thermodynamic correlates of T cell signaling. *Annu. Rev. Biophys. Biomol. Struct.*, **31**, 121–149.

16. Rudolph,M.G. and Wilson,I.A. (2002) The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.*, **14**, 52–65.

17. Ponomarenko,J.V. and Bourne,P.E. (2007) Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct. Biol.*, **7**, 64.

18. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.

19. Bhattacharya,D. and Thio,C.L. (2010) Review of hepatitis B therapeutics. *Clin. Infect. Dis.*, **51**, 1201–1208.

20. Dienstag,J.L. (2008) Hepatitis B virus infection. *N. Engl. J. Med.*, **359**, 1486–1500.

21. Bartenschlager,R. and Schaller,H. (1988) The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO J.*, **7**, 4185–4192.

22. Bartenschlager,R. and Schaller,H. (1992) Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.*, **11**, 3413–3420.

23. Mareeva,T., Martinez-Hackert,E. and Sykulev,Y. (2008) How a T cell receptor-like antibody recognizes major histocompatibility complex-bound peptide. *J. Biol. Chem.*, **283**, 29053–29059.

24. Miyata,M. (2010) Unique centipede mechanism of Mycoplasma gliding. *Annu. Rev. Microbiol.*, **64**, 519–537.

25. Miller,P.J., Pazy,Y., Conti,B., Riddle,D., Appella,E. and Collins,E.J. (2007) Single MHC mutation eliminates enthalpy associated with T cell receptor binding. *J. Mol. Biol.*, **373**, 315–327.