行政院國家科學委員會專題研究計畫　成果報告

利用電腦模擬及新的胜肽合成法發展高效能及最佳的仿胜肽藥物

計畫類別：個別型計畫
計畫編號：NSC92-2323-B-002-013-
執行期間：92 年 08 月 01 日至 93 年 10 月 31 日
執行單位：國立臺灣大學資訊工程學系暨研究所

計畫主持人：歐陽彥正
共同主持人：阮雪芬，陳水田，張夢揚

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 4 月 30 日

行政院國家科學委員會補助專題研究計畫 ☑ 成 果 報 告 □期中進度報告

利用電腦模擬及新的胜肽合成法發展高效能及最佳的仿胜肽藥物

計畫類別：☑ 個別型計畫　　□ 整合型計畫
計畫編號：NSC 92－2323－B－002－013
執行期間：92 年 08 月 01 日至 93 年 07 月 31 日

計畫主持人：**歐陽彥正**
共同主持人：陳水田、阮雪芬
計畫參與人員：張夢揚

成果報告類型(依經費核定清單規定繳交)：☑精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立台灣大學資訊工程學系

中　華　民　國　94 年　4　月　30　日

生技製藥國家型科技計畫
92 年度成果報告

利用電腦模擬及新的胜肽合成法發展高效能及最佳的仿胜肽藥物

Development of highly potent and optimal RGD mimetic peptides drugs by computer modeling and new peptide synthesis methods.

計畫編號	NSC 92－2323－B－002－013		
執行期限	92 年 08 月 01 日至 93 年 07 月 31 日		
計畫類別	生技製藥國家型科技計畫 ☐天然藥物組　☑化學合成藥物組 ☐蛋白質晶片組　☐藥效評估組 ☐生技藥物組　☐臨床試驗		
研究型別	☑個別型計畫　☐整合型計畫		
計畫歸屬	☑國科會生物處　☐行政院衛生署(臨床試驗組)		
執行機關	國立台灣大學	執行系所	資訊工程學研究所
計畫主持人	**歐陽彥正**		
共同主持人	陳水田、阮雪芬		
計畫參與人員	張夢揚		
整合型總計畫名稱			
整合型總計畫主持人			
計畫計畫聯絡人	姓名：**歐陽彥正** 電話：(公) 02-23625336 ext 431　(宅)_____		
通訊地址	北市羅斯福路 4 段 1 號國立台灣大學資訊工程學研究所		
傳真號碼	02-23688675		
E-MAIL	yjoyang@csie.ntu.edu.tw		

計畫申請人(主持人)簽章：　　日期：

一、中文摘要

三胜肽序列 RGD(Arg-Gly-Asp)是一個常見的細胞辨識的要素(motif)；它是結合到 integrin 配位分子的一部份，這些配位分子有 fibronectin，fibrinogen 和 vitronectin。這條序列已經被利用來作為發展各種不同 intergrin 拮抗劑的先導化合物[1]。含 RGD 胜肽可以被當作 integrin-ligand 相互作用的有效抑制劑；也可以被用在研究細胞附著、遷移、生長和分化的研究上。最近，含 RGD 胜肽更進一步被發現其可藉由調節 caspase-3 的活化而能夠誘導細胞的自然死亡[2]。因此發展更有效的含 RGD 胜肽當作治癌藥物將是非常地有價值的。

在本計畫中我們所合成的藥物中以 cRGD 及 C-16 化合物抗癌的研究最為透徹，C-16 化合物能夠抑制很多種癌細胞的生長，包括腎臟癌 (A-498)、鼻咽癌 (NPC-tw01)、結腸癌 (HCT-116)、乳癌 (MCF-7)、肝癌(Hep-3B) 、胃癌 (MKN-45) 、子宮頸癌 (MES-SA) 、肺癌 (NCI-H226) 、腎小管癌 (RPTEC)。其中對乳癌的抗癌活性最好，對腎臟癌的生長抑制最差。C-16 化合物對乳癌細胞的 50%抑制性(GI_{50})已達到奈米莫耳量，顯示該藥物可以當做抗乳癌先導藥物。

本計畫另一項研究重點是研發先進的電腦輔助藥物設計(computer-aided drug design)方法及軟體。目前我們已經研發成功一套蛋白質序列分群(protein sequence clustering)軟體及一套 protein-ligand interaction 預測軟體。此兩套軟體已在本計畫中成功應用於協助生物學家尋找先導化合物。

關鍵詞：RGD 仿胜肽；固相合成法；電腦輔助藥物設計；蛋白質結構分析；藥物開發

二、英文摘要

The tripeptide sequence RGD (Arg-Gly-Asp) in integrin binding ligands like fibronectin, fibrinogen, and vitronectin is a common cell-recognition motif. This sequence has been used as a lead compound to develop different integrin antagonists. RGD-containing peptides have been found to be efficient inhibitors of integrin-ligand interactions in cells studies on adhesion, migration, growth and differentiation. Recently, the RGD-containing peptides have been further discovered to be able to induce cell apoptosis mediated by activation of caspase-3. Therefore, it is highly worthwhile to develop more potent RGD-containing peptides as drugs for cancer therapy.

In this project, the most understanding compounds we synthesized are cRGD and C-16. C-16 can inhibit the growth of many various tumor cells, including kidney caner (A-498 and RPTEC), nasopharyngeal carcinoma (NPC-tw01), colon cancer (HCT-116), breast cancer (MCF-7), liver cancer (HEp-3B), stomach cancer (MKN-45), Cervical Cancer (MES-SA), and lung cancer (NCI-H226). The best anticancer activity is anti-breast cancer, the worst one is anti-kidney cancer. Since the 50% inhibition (GI_{50}) of C-16 is about nano mole, this result indicated that this compound may be as a lead compound.

In this project, we also have developed two advanced software packages for computer-aided drug design in order to greatly enhance the efficiency of the drug discovery process. The first software package is a protein sequence clustering package. The second software package is for efficiently

identifying the possible protein-ligand interactions based on analysis of protein tertiary substructures.

三、研究方法與成果討論

RGD analogues are commercialized drugs that have been proved to be potent drugs. We have synthesized the more potent cyclic RGD mimetic peptide, c[3-mercaptopropionic acid-Arg-Gly-Asp-Trp-Pro-Cys] and assayed its biological activity (please see Fig. 1, Fig. 2, and Fig. 3). In the effect on inhibiting MCF-7 cell growth, cyclic RGD exerts 8-10 times potency more than that of linear RGD on the inhibiting proliferation and inducing clustering of MCF-7 cells. Besides, 48 μM cRGD can completely inhibit human platelet aggregation in platelet-rich plasma stimulated by ADP (20μM), and inhibit 92 % platelet aggregation in platelet-rich plasma stimulated by collagen (10μg).

Furthermore, we performed structure-similarity search in various chemical databases and manual survey for medical and commercial availability. Many RGD mimetic peptide drug candidates have been found. We will continue to test the bioactivity of these candidates.

Another major objective of this project is to develop advanced computer-aided drug design software packages in order to facilitate and expedite the process of drug discovery. In this project, we have designed a novel protein sequence clustering algorithm aimed at generating summarized dendrograms for analysis of protein databases. The proposed clustering algorithm employs a statistics-based model to summarize the distributions of the similarity scores among the proteins in the database and to control formation of clusters. Experimental results reveal that, due to the summarization mechanism incorporated, the proposed clustering algorithm offers the users highly concise dendrograms for analysis of protein clusters with biological significance. Fig. 4 compares the weighted average matching rates delivered by the single-link algorithm and the proposed clustering algorithm with the 4 benchmark datasets. In the results of the single-link algorithm in Fig. 4, no cutoff threshold is imposed to flatten the dendrogram. If a cutoff were imposed, then the weighted average matching rates delivered by the single-link algorithm would turn lower. In this experiment, the results corresponding to the proposed clustering algorithm are the averages of 5 independent runs with random order of input sequence. As Fig. 4 reveals, the proposed clustering algorithm and the single-link algorithm deliver comparable levels of weighted average matching rate. Table 1 and Table 2 present the most distinctive effects achieved with the proposed clustering algorithm. Table 1 compares the numbers of non-leaf nodes in the dendrograms generated by the proposed algorithm and by the single-link algorithm with the four benchmark datasets. Again, for the three smaller datasets, the results corresponding to the proposed clustering algorithm are the averages of 5 independent runs with random order of input sequence. Table 2 lists

the average depth that a user needs to traverse in each dendrogram in order to find a cluster that matches one family in InterPro best.

In addition to the protein sequence clustering algorithm described above, we have designed a software package for prediction of protein-ligand interaction based on analysis of protein tertiary substructures. Fig. 5 illustrates the application that the software package addresses. In this application, the biochemist is given the crystal structure of a protein bound with a specific ligand and wants to conduct a search in the PDB database (Berman, Westbrook, et al., 2000) for the other proteins that could interact with the specific ligand. In one experiment that we have conducted, the biochemists wanted to search the PDB database for all the other proteins that could interact with the ligand in the co-crystal structure reported in (Xiong, Stehle, et al. 2002), which consists of an integrin protein bound with a peptide ligand containing Arg-Gly-Asp. Table 3 presents the results outputted by the software package. In Table 3(b), only human proteins in the PDB database that has 18 residues successfully aligned with the 18 residues in the binding site of protein integrin αVβ3 are listed. In the experiment, the likelihood of residue substitution was also taken into account. If the entry in the PAM 250 matrix (Lesk, 2002; Altschul, 1991) corresponding to a pair of residues aligned by the geometric hashing algorithm is either zero or negative, then this pair of residues is excluded from the list of successfully aligned. Table 3(c) lists how the residues in protein fibrinogen, with PDB ID = 1fzg, are mapped to the residues in the binding site of the reference protein. The mapping detail of this pair is in conformity with our hypothesis that the ligand can also be a lead compound for anticoagulation. As Table 3(c) reveals, the substructure in protein fibrinogen shares a number of common chemical properties with the binding site of protein integrin αVβ3.

In summary, we found a RGD mimetic peptide, cyclic RGD, which can inhibit MCF-7 cell growth and inhibit platelet aggregation in platelet-rich plasma. We will use this cyclic RGD as a lead compound to find more potent anti-cancer and anti-platelet aggregation drugs. In addition, we have developed two computer-aided drug design packages. These two software packages have been successfully exploited in our project. The computer scientists and biochemists in our research team will continue to work together closely for developing more advanced computer-aided drug design packages. Eventually, we believe that we can make significant contributions in developing novel computer-aided drug design methodologies.

四、計畫成果列表

請依序列出與執行本計畫相關之已發表或已被接受發表著作(學術期刊論文、學會及研討會報告)、已申請或被接受之專利、專著、技術報告、或學生畢業論文等。

1. Chien-Yu Chen, Hsueh-Fen Juan, Po-Jen Hsiao, Shui-Tein Chen, Hsiang-Wen Tseng, and Yen-Jen Oyang , Design of an Incremental Clustering Package for Protein Function and Family Analysis, in Proceedings of IEEE 5th International Symposium on Multimedia Software Engineering, Taichung, Taiwan, 2003.
2. Yen-Jen Oyang, Darby Tien-Hau Chang, Chien-Yu Chen, and Shien-Ching Hwang, Expediting Protein Structural Analysis with an Efficient Kernel Density Estimation Algorithm , in Proceedings of IEEE 5th International Symposium on Multimedia Software Engineering, Taichung, Taiwan, 2003.

五、參考文獻

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Research, 28, 235-242.

Lesk, A.M. (2002) Introduction to bioinformatics, New York : Oxford University Press.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., and Arnaout, M.A. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science. 296, 151-5.

Figures and Tables

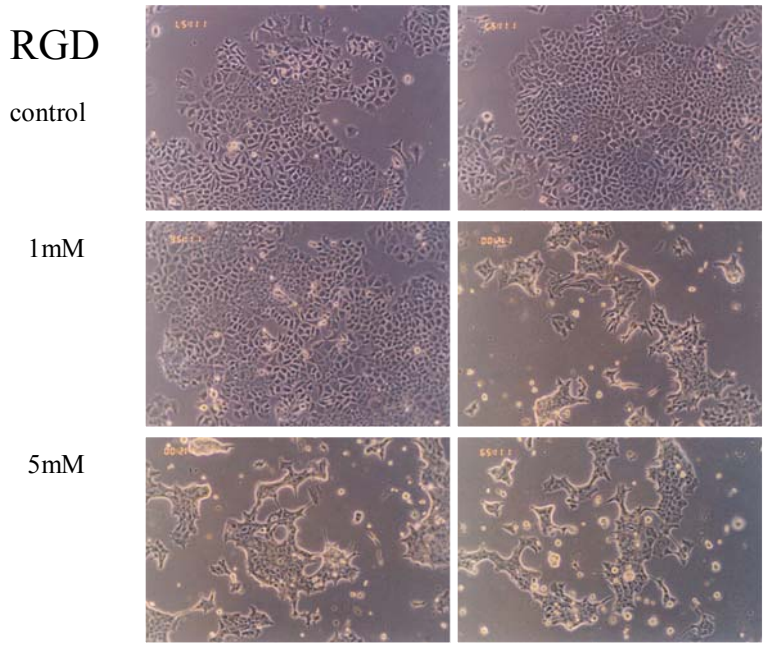


Fig. 1. Effect of cRGD on inhibiting MCF-7 cell growth. Cyclic RGD exerts more potency than that of liner RGD on the inhibiting cell growth. The cyclic RGD exerts 8-10 times potency more than that of liner RGD peptide in inhibiting proliferation and inducing clustering of MCF-7 cells.

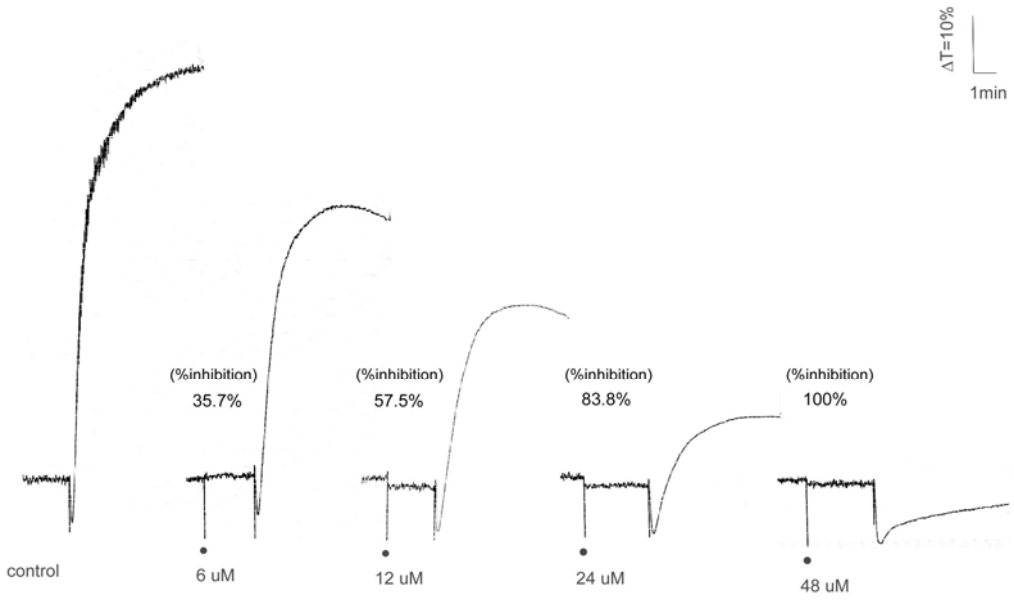


Fig. 2. Effect of cRGD on human platelet aggregation. Effect of various doses of cRGD or saline (control) on platelet aggregation in platelet-rich plasma stimulated by ADP(20μM).

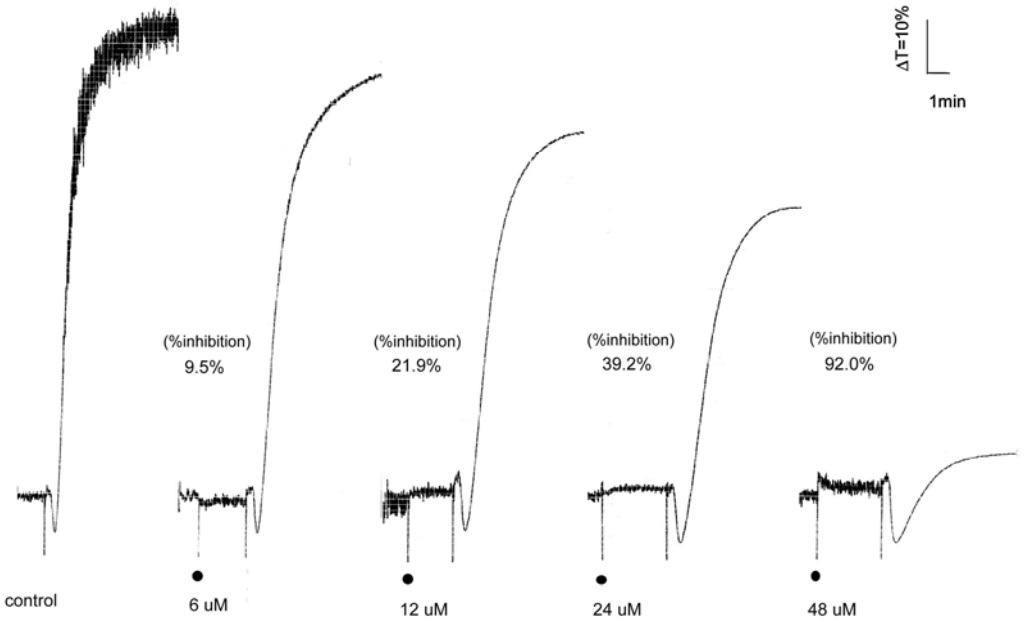
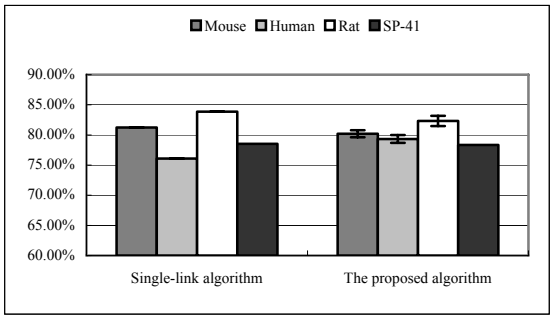
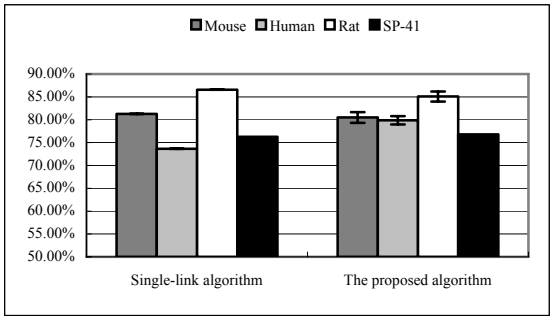


Fig. 3. Effect of cRGD on human platelet aggregation. Effect of various doses of cRGD or saline (control) on platelet aggregation in platelet-rich plasma stimulated by collagen(10μg).



(a) Comparison of the weighted average matching rate for protein families containing more than 10 proteins.



(b) Comparison of the weighted average matching rate for protein families containing more than 30 proteins.

Fig. 4. Comparison of the weighted average matching rates delivered by the protein sequence clustering algorithm developed in this project and the single-link algorithm.

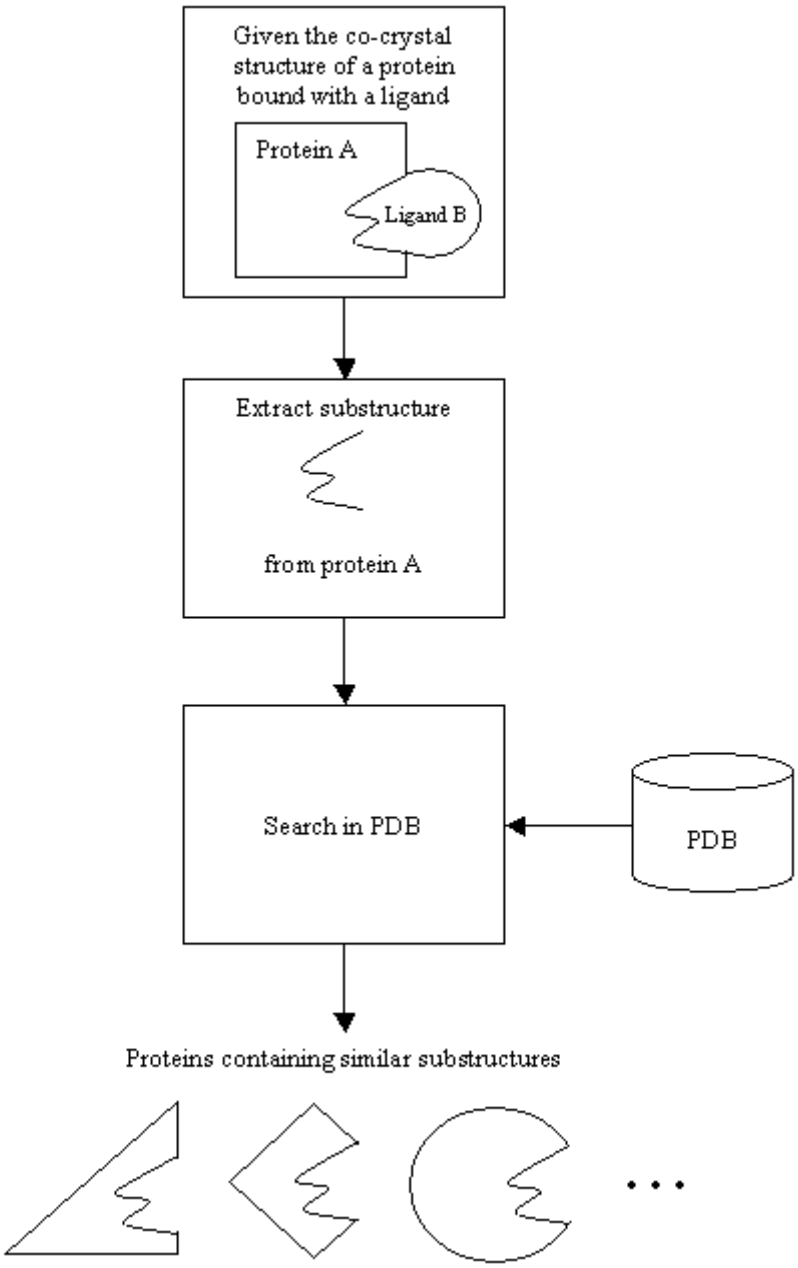


Fig. 5. The application that the software package developed in this project for prediction of protein-ligand interactions addresses.

Dataset	Number of non-leaf nodes	
	Single-link	The protein sequence clustering algorithm developed in this project
Mouse	4707	1333.8 ± 6.14
Human	7470	1986.8 ± 12.28
Rat	2915	806 ± 6.54
SP-41	122563	34352

Table 1. Comparison of the numbers of non-leaf nodes in the dendrograms generated by the protein sequence clustering algorithm developed in this project and the single-link algorithm.

Dataset	Depth	
	Single-link	The protein sequence clustering algorithm developed in this project
Mouse	798.33	1.85 ± 0.037
Human	1414.24	2.54 ± 0.050
Rat	520.67	1.78 ± 0.029
SP-41	15312.95	4.61

Table 2. Comparison of the average depths that a user needs to traverse in the dendrograms in order to find a cluster that matches one family in InterPro best.

Reference proteins			
PDB ID	# original residues	# of residues in the binding site	# of residues remaining with filtering applied
1l5g	1470	18	483

(a) Characteristics of the reference protein.

PDB ID of the Target Protein	# of residues	# of residues remaining with filtering applied	Geometric hashing without filtering		Geometric Hashing with filtering applied			Speedup
			Execution time of geometric hashing	# of residues in the cave that are successfully aligned	Execution time of filtering	Execution time of geometric hashing	# of residues in the cave that are successfully aligned.	
1bp5	1324	522	117.89	18	0.23	16.07	17	7.232515
1de8	594	150	39.35	18	0.08	3.45	17	11.14731
1dvm	1476	400	147.5	18	0.28	13.98	18	10.34362
1egc	1548	280	177.8	18	0.32	11.63	18	14.87866
1egd	1548	273	177.74	18	0.32	11.19	18	15.44222
1f4j	1918	670	259.99	18	0.44	32.88	18	7.802821
1foe	2152	719	280.48	18	0.55	33.12	18	8.330264
1fzb	1414	344	123.4	18	0.29	10.55	18	11.38376
1fzg	1313	344	113.45	18	0.24	10.25	18	10.81506
1h2i	4092	1398	760.03	18	1.52	94.56	18	7.910387
1h69	1092	322	87.84	18	0.17	8.75	18	9.847534
1haq	4852	247	551.02	18	2.37	12.16	17	37.92292
1ias	1644	450	174.76	18	0.31	16.22	18	10.57229
1iil	1361	180	136.94	18	0.26	6.31	18	20.84323
1itq	738	246	54.69	18	0.1	6.4	17	8.413846
1ivh	1548	364	178.46	18	0.31	15.08	18	11.59584
1jbq	2090	547	228.05	18	0.47	20.77	18	10.73682
1jv2	1466	446	154.88	18	0.27	16.87	18	9.036173
1kn0	2024	720	270.32	18	0.47	34.52	18	7.725636
1kv3	3906	1927	722.05	18	1.39	128.37	18	5.564504
1l5g	1470	483	156.49	18	0.26	18.33	18	8.417967
1l7x	1588	418	180.11	18	0.33	16.59	18	10.6448
1m1x	1466	485	154.33	18	0.27	18.3	18	8.310716
1n18	1530	380	153.79	18	0.31	13.29	18	11.30809
1nr1	2976	693	531.2	18	0.85	46.02	18	11.33348
1nzw	3952	1152	751.19	18	1.42	78.88	18	9.354795
1ogs	994	322	80.34	18	0.17	9.12	18	8.648009
1oqd	1734	153	189.05	18	0.38	6.06	18	29.35559
1os9	990	277	84.34	18	0.14	8.31	17	9.981065
1pq3	1836	349	209.66	18	0.39	13.83	18	14.74402
1qha	1802	466	214.46	18	0.41	19.88	18	10.56974
1qki	3912	1215	773.5	18	1.42	86.72	18	8.775811
1qmv	1950	568	219.82	18	0.4	22.1	18	9.769778
Average	-	-	-	-	-	-	-	11.78

(b) Effects of the prediction software package developed in this project.

Table 3. Experimental results of the software package developed in this project for prediction of protein-ligand interactions. (to be continued)

Protein integrin αVβ3 (reference protein)			Protein fibrinogen (PDB ID=1fzg)			PAM250 Score
Chain	Residue Index	Residue Type	Chain	Residue Index	Residue Type	
A	150	ASP	C	225	GLU	3
A	178	TYR	C	354	TYR	10
A	215	ALA	C	357	ALA	2
A	218	ASP	C	362	GLY	1
B	119	ASP	C	316	ASP	4
B	121	SER	C	351	GLY	1
B	122	TYR	C	167	TYR	10
B	123	SER	C	200	GLY	1
B	126	ASP	C	147	ASP	4
B	127	ASP	C	217	HIS	1
B	158	ASP	C	330	ASP	4
B	215	ASN	C	207	ASN	2
B	216	ARG	C	205	LYS	3
B	217	ASP	C	328	GLU	3
B	218	ALA	C	353	THR	1
B	219	PRO	C	327	ALA	1
B	220	GLU	C	251	GLU	4
B	251	ASP	C	150	GLY	1

(c) How the residues in protein fibrinogen, with PDB ID = 1fzg, are mapped to the residues in the binding site of the reference protein.

Table 3. Experimental results of the software package developed in this project for prediction of protein-ligand interactions. (continues)

計畫成果自評：本計劃已達成原計劃內容所規劃之第一年進度，所研發的電腦輔助藥物設計(computer-aided drug design)方法及軟體，包括一套蛋白質序列分群(protein sequence clustering)軟體及一套 protein-ligand interaction 預測軟體，成功地應用於協助生物學家尋找先導化合物。本計畫對 cRGD 及 C-16 化合物抗癌效果進行透徹的研究，對後續的相關藥物開發有極大的助益。執行期間已將階段性結果發表於國際會議。