蛋白質的動態關連性

學生：何仙蕾　　指導教授：黃鎮剛　博士

國立交通大學生物科技學系（研究所）碩士班

摘　　要

有些蛋白質內部的氨基酸殘基並沒有直接和配體(ligand)有鍵結, 但是這些殘基卻能巨幅地影響蛋白質和配體間的鍵結親和力(十倍或更高). 這些氨基酸可以以丙氨酸掃描式突變在蛋白質的內部交界處來偵測得, 稱為熱點氨基酸(hot spot residues). 為了讓實驗進行地更有效率, 我們用我們發展出的方法發現蛋白質的鍵結位和熱點氨基酸間在幾何結構上具有高度的相關性, 可以辨識出這些熱點氨基酸和非熱點氨基酸. 分子模擬可以複雜的分子勢能來分析蛋白質內部的分子運動模式, 是一套非常有力的工具, 但是它非常地耗時. 因此, 我們使用簡化的物理分析模式, 如彈性網路模式(GNM)、中心模式（CM）、權重原子接觸模式(WCN)來探討蛋白質內部的動態關連性. 然而, GNM的交錯關連矩陣不能辨識出熱點氨基酸. 改良過的WCN和常態化後的CM可以辨識出這些熱點氨基酸. 我們發展出的這些方法可以顯示出熱點氨基酸和鍵結氨基酸間在幾何關係上具有高度地相關性. 最後, 也發現鍵結氨基酸和高度保留的氨基酸間也有高度的空間相關性.

Dynamical correlation in proteins

Student：Shian-lei,Ho　　Advisors：Dr. Jenn-Kang Hwang

Department of Biological Science and Technology
National Chiao Tung University

ABSTRACT

There are residues in proteins not directly bind to ligand, but they can dramatically affect the binding affinity(tenfold or higher)between protein and ligand. These residues called hot spot residues which can be detected by alanine scanning mutagenesis in protein interface. For the purpose to design experiment effectively, we develop some methods to find the geometric correlation between binding sites and hot spot residues.These methods can identify hot spot residues and non-hot spot residues. The molecular dynamics(MD) simulations with complex molecular potential energy can analize the motion of proteins. MD simulation is a powerful tool, but it is time-consuming. For this reason, we use simplified mechanical methods such as Gaussian network model(GNM), and our methods, centroid model(CM) and weighted contact-number model(WCN) to discuss the dynamical correlation in proteins. Howerer, the cross-correlation matrix in GNM can not identify the hot spot residues. The improved WCN and normalized centroid model can identify these hot spots. The methods we developed can show that there are high correlation between functional hot spot residues and binding sites in geometrical relation. In the end, we can also find the high spatial correlation between binding sites and highly sequence conserved residues.

誌謝

研究所兩年, 真的獲益良多, 當中最感謝的是我的指導老師黃鎮剛教授, 他讓我們知道最好的老師不是別人, 而是自己. 也很感謝實驗室的學長姐們的指導和協助, 特別感謝志鵬學長、少偉學長、小胖學長在程式上的協助, 還有小操學長在Mac電腦上的協助, 沒有大家的幫忙我也不能完成研究. 最後也最重要的是感謝家人的鼓勵與支持才能完成畢業論文, 也感謝老天爺的眷顧.

List of Tables

Table I : The total number residues of conservation > 7 and the percentage of conservation > 7 in iWCN and COS for ATP binding site in protein of PDB code 1A1V . 51

Table II : The percentage of conservation > 7 in iWCN and COS for DNA binding site of PDB code 1A1V. 51

Table III : The total number residues of conservation > 7 and the percentage of conservation > 7 in iWCN and COS for human hormone receptor binding site of PDB code 3HHR b chain . 52

Table IV : A. The comparison of experimental data and COS to identify hot spot residues in protein of PDB code 3HHR . B. The accuracy of COS . 53

Table V : A. The comparison of experimental data and iWCN to identify hot spot residues in protein of PDB code 3HHR . B. The accuracy of iWCN . 54

List of Figures

Figure 1 : The non-normalize correlation map of CM, COS, GNM, WCN, iWCN. 14

Figure 2 : The eigen value of CM, COS, GNM, WCN, iWCN, which are non-normalized. 17

Figure 3 : The eigen vector of CM, COS, GNM, WCN, iWCN, which are non-normalized. Red squares are hot spot residues. 20

Figure 4 : The direction of eigenvalue of CM, COS, GNM, WCN, iWCN, which are non-normalize . Red and yellow color represent contrarious direction of fluctuation . 23

Figure 5 : The correlation map of COS, GNM, WCN, iWCN, which are normalized. 26

Figure 6 : The eigenvalue of COS, GNM, WCN, iWCN, which are normalized. 28

Figure 7 : The eigenvector of COS, GNM, WCN, iWCN, which are normalized. Red squares are hot spot residues. 30

Figure 8 : The mode of eigen vector of COS, GNM, WCN, iWCN, which are normalized. 33

Figure 9 : Blue color denotes domain1 , green color denotes domain 2, red color denotes domain 3, magenta sphere denotes the center of CM, yellow sphere denotes the new center. 36

Figure 10 : The comparison of each method before improvement. 37

Figure 11 : The comparison of each method after improvement. 38

Figure 12 : The mode 1 comparison of each method after improvement. 39

Figure 13 : A. ATP binding site in protein of PDB code 1A1V . B. High conservation residues in 1A1V . C. The high correlation with ATP binding sites in COS . D. The high correlation with ATP binding sites in PCN. Here we only show α carbon spheres. Domain 1 : blue color. Domain 2 : green color. Domain 3 : red color. 40

Figure 14 : A. DNA binding site in protein of PDB code 1A1V . B. The high correlation with DNA binding sites in COS . C. The high correlation with DNA binding sites in iWCN. Here we only show α carbon spheres. Domain 1 : blue color. Domain 2 : green color. Domain 3 : red color. 44

Figure 15 : A. Human hormone receptor (PDB code 3HHR) binding site in chain b and its highly conserved residues . B. iWCN with high correlation . C. COS with high correlation . 47

Contents

摘要 i

ABSTRACT ii

誌謝 iii

List of Tables iv

List of Figures v

Content vii

Introduction 1

Method 1

Gaussian network model (GNM) 1

Centroid model (CM) 3

The normalization of centroid model(COS) 4

Weighted contact-number model(WCN) 4

The improvement of weighted contact-number model (iWCN) 5

Normalization 6

Prediction Accuracy 6

Data sets 6

Result 7

Correlation Map 7

Eigenvalue 7

Eigenvector 8

The improvement of eigenvector 9

Hot spot residues 9

Discussion 11

Reference 12