Chiang Mai J. Sci. 2012; 39(X) : 1-8
http://it.science.cmu.ac.th/ejournal/
Contributed Paper

MATLAB Process for Validating Amino Acids on CD4 Involving in DARPin Binding Site from ZDOCK Molecular Docking Database†

Tanchanok Wisitponchai [a], Piyarat Nimmanpipug [b,c], Vannajan Sanghiran Lee [c,d] and Chatchai Tayapiwatana*[a,f,g]

[a] Biomedical Engineering Center, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand.

[b] Department of Chemistry, Center of Excellence for Innovation in Chemistry and Materials Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand.

[c] Thailand Center of Excellence in Physics, Commission on Higher Education, Bangkok 10400, Thailand.

[d] Department of Chemistry, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, Malaysia.

[f] Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand.

[g] Biomedical Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand.

*Author for correspondence; e-mail: asimi002@hotmail.com

† *Presented at the Pure and Applied Chemistry International Conference 2012 (PACCON 2012), 11th-13th January 2012, Chiang Mai, Thailand.*

Received: 15 January 2012
Accepted: 7 April 2012

ABSTRACT

Human immunodeficiency virus (HIV) inherits the active mutation, which promotes the viral survival in human and is the main hurdle of vaccine development. The mutations on HIV surface i.e. CD4 binding site on HIV gp120, causes the escaping from the immune-surveillance but still retains the infectivity for CD4+ T cells. Formerly, Designed Ankyrin Repeat Protein (DARPin) technology has been innovated to inhibit the HIV infection. The CD4-specific DARPin specifically excludes the adhesion step of HIV to CD4 molecule and efficiently prevents HIV infection *in vitro* by competing with gp120, however, not *in vivo*. Therefore, insight into the interaction between DARPin and CD4 molecules will provide the information to improve the interaction affinity of DARPin for *in vivo* purpose. The binding activity can be modified by directly defining the key amino acid residues of CD4-specific DARPin and replacing them with other possible residues. To discover the remarkable CD4's residues, a molecular docking software i.e. ZDOCK was used to predict the complex structures between CD4 and DARPin. In this study, the candidate residues of CD4-specific DARPin were identified by the residues of CD4 molecule involving in gp120 interaction. The defined DARPin poses that interrelated with them were then extracted. The MATLAB software was implemented to generate the assigned screening criteria

for recruiting the most relevant binding residues in eleven complex-structure-candidates derived from ZDOCK. The invented data processing methodology will assist the molecular simulation researchers to rapidly and precisely define important residues.

Keywords: DARPins, CD4, HIV, Protein Docking, MATLAB.

1. INTRODUCTION

The evaluation and mutation of Human immunodeficiency virus (HIV), effect of high rates of viral replication in human host cells, are able to evade recognition by cytotoxic T lymphocytes (CTLs) [1, 2]. The entry of HIV into host cell is initiated by a binding between the viral envelope protein, gp120, and its primary receptor, CD4, on surface of T-helper lymphocytes or macrophages. The mutations on HIV surface i.e. gp120, CD4 binding site on HIV, cause the evasion of effective host-neutralizing antibodies, however, still retain the infectivity for CD4+ T cells [3]. Antiretroviral (ARV) drugs, currently available agents, are more classified by the phase of the retrovirus life cycle than direct binding with HIV particles. Several HIV drugs have been developed to hinder entry of HIV, such as maraviroc, enfuvirtide [4], and ibalizumab [5]. Drug resistance problems from virus mutations and serious side effects are the importance problem for treatments, so the discoveries for new agents are needed. One of microbicide agents for inhibiting entry of HIV has been developed, which is Designed Ankyrin Repeat Protein (DARPin) technology [6]. DARPin consists of repeat motifs and flanked constant regions. Each repeat motif, containing 33 amino acids, comprises one beta-turn and two antiparalled alpha helices [7]. Advantages of using ankyrin repeat proteins are due to their tight structure, specific binding at nanomolar range, and each repeat can contribute to target binding [8]. DARPins specific for human CD4 were reported by Schwizer *et al.* [9], *in vitro*, they are potent and highly specific to CD4 to block HIV entry. In addition, they are able to act against a wide range of virus strains and cause no effect on basis T cell function. *In vivo*, CD4-specific DARPin 57.2 efficiently binds to CD4+ cells (T helper cells, DCs, and monocytes) but has no effect on SHIV-infected rhesus macaques [10]. Therefore, insight into the interaction between DARPins and CD4 molecules may provide a clue for improving the binding activity of DARPins. To improve the binding activity, the key amino acids of CD4-specific DARPins were defined directly [11] and replaced with other possible residues. In the other hand, the replaceable amino acids on DARPin were discovered indirectly by analysis of key CD4's amino acid.

Our study aims to clarify the roles of crucial CD4's amino acid in the interaction of DARPin-CD4 complex, one idea to find replaced amino acids on DARPin, by using ZDOCK [12, 13] and MATLAB [14] processing. Protein-protein docking tool is used to predict protein complex structure. ZDOCK, an initial stage of docking, is an algorithm that optimizes desolvation, grid-based shape complementarity (GSC) and electrostatics by using fast Fourier transform (FFT) algorithm. The refinement stage, RDOCK, a small number (tens to thousands) of structures obtained in the initial stage is refined and re-ranked. The scoring from ZDOCK or RDOCK was carried out to gain the possible complex

structures. The intermolecular neighbors, binding residues, between complex structures were identified from the possible complexes. Numerous data of the binding residues were rapidly and precisely analyzed by MATLAB processing. These data are divided into five groups based on the generated criteria for the decision determining the key CD4's amino acid. Our model of key CD4's residues decision is using pair matching histogram analysis.

2. MATERIALS AND METHODS

2.1 Protein Docking

The ankyrin-CD4 docking simulations were performed by using a ZDOCK [12] and RDOCK [15] protocol in the Discovery studio (DS) 2.5. The 3D structure of CD4 specific ankyrin was generated by homology modeling [11] and the CD4 was downloaded from protein data bank (PDB code: 3CD4) [16]. The initial step, ZDOCK protocol, the docked poses were generated by 54,000 predictions around the CD4 that are possible binding regions within a small cluster radius of 6.0 Å for the RMSD cutoff, and a smaller interface cutoff of 9.0 Å. To filter the possible poses, the binding site [8] and the blocked residues of ankyrin were specified. The 2,000 docked poses with the highest scores were reported. The next step of refinement stage, the CD4 and ankyrin structures were typed by the CHARMm polarH force field. The selected docked poses with top 20 ZDOCK values were inputted into an RDOCK protocol to get complex structure.

2.2 Finding Intermolecular Neighbors

Each ankyrin-CD4 complex structure from RDOCK process, involving the gp120-CD4 complex (PDB code: 2NXY) [17], was carried out to find the binding atom neighbors. The set of neighbor atom was investigated within a distance threshold of 5.0 Å. The default parameter for the distance between the hydrogen bond donor and the acceptor was 2.5 Å and the donor proton-acceptor angles between 120-180°, were selected.

2.3 Data Transformation

Because the data from DS are in a string form, they were changed into numbering data to suit MATLAB. The data were classified into five criteria, first criterion; the number of DARPin's amino acid positions was bound to each CD4's amino acid. The second, the number of CD4's atom types in each CD4's amino acids was bound to DARPin's residues. Third criterion, the percentage of CD4's atom types in each CD4's amino acids was bound to DARPin's residues. The forth, the number of interactions in each CD4's amino acids was bound to DARPin. The last, each hydrogen-bonded CD4's amino acids were bound to DARPin's residues.

2.4 Data Selection

For each CD4's amino acid, we counted a number of times for a CD4's amino acid that was matched with the DARPin's amino acid. Then we created a histogram of the matching. In each criterion, the CD4's amino acids with the top 10 highest values in the histogram were selected to be candidates for considering key CD4's amino acids. Next the histogram values in each considered CD4's amino acids were combined and normalized by using:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

, where x is histogram value of CD4's amino acid, μ is CD4's criterion mean, and σ is CD4's amino acid standard deviation.

2.5 Decision Making

We created 6 patterns of combination, i.e., patterns A, B, C, D, E and F. For each pattern, the normalized histogram values at each considered CD4's amino acid were combined. However, the criteria 1, 2, 3, 4, and 5 were used in pattern A, whereas, 1, 3, 4, and 5 were used in pattern B. For pattern C, D, E, and F, the used criteria were; 1, 2, 4, and 5; 1, 2, 3, and 4; 1, 2, and 4; and 1, 3, and 4, respectively. The combined CD4's amino acids were normalized again using equation (1). The position of the maximum normalized combined CD4's amino acid was selected for each pattern. Finally, the key CD4's residue decision, the maximum one from all patterns was selected. Because we desired only the top 3 of considered CD4's residues, the maximum was performed as 3 times.

3. RESULTS AND DISCUSSION

3.1 Characteristics of Ankyrin-CD4 Complex Structure

In 11 out of the 20 poses, the ankyrin was found to bind to the CD4 molecule on the domain 1, where the region was bound by gp120 [9]. We divided them, with the same binding region on CD4, into 3 clusters; 7 poses in the first cluster (pose 16, 21, 26, 642, 1302, and 1513); 3 poses in the second cluster (pose 1128, 1266, and 1454); and 1 pose (pose 85) in the last cluster. All clusters, the binding CD4's residues were overlapped with the binding CD4's residues that bind to gp120, as shown in figure 1.

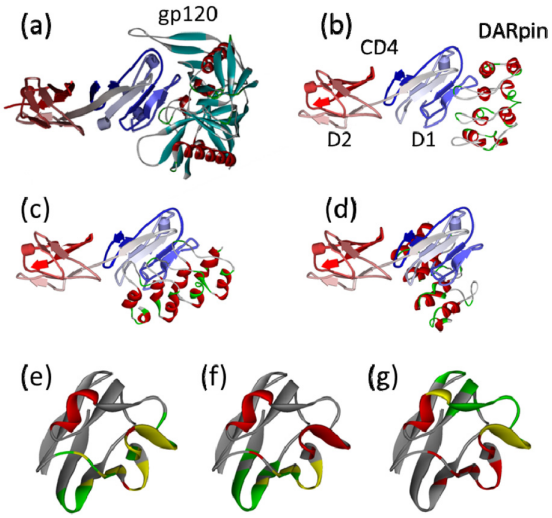

Figure 1. Complex structures of CD4-gp120 (a) and CD4-DARPin; first cluster (b), second cluster (c), and third cluster (d). Characteristics of the binding region on domain 1 of CD4; first cluster (e), second cluster (f), and third cluster (g). The red region is bound by gp120, the green region is bound by DARPin, and the yellow region is bound by both DARPin and gp120.

3.2 Data Analysis

In every pose, the values in each criterion, excluding the fifth criterion, were counted to be considered as CD4's amino acids, as shown in figure 2. The considered CD4's amino acids to be key residues in each poses were shown in table 1. In cluster 1, 7 poses, there were 11 considered CD4's amino acids that were the same 100%; these amino acids were Q25, H27, Q33, I34, K35, N39, Q40, G41, P48, E85, E87, and K90. In cluster 2, 3 poses, there were 10 considered CD4's amino acids that were the same 100%. These amino acids were H27, K29, S31, N32, Q33, E85, E87, D88, K90,

Table 1. The considered CD4's amino acids that were a combination of top ten in each criterion.

Pose	position of considered CD4's amino acids
16	25, 27, 33, 34, 35, 39, 40, 41, 48, 85, 87, 90
21	25, 27, 29, 33, 34, 35, 39, 40, 41, 48, 85, 87, 90
26	25, 27, 29, 33, 34, 35, 39, 40, 41, 48, 85, 90
241	25, 27, 33, 34, 35, 39, 40, 41, 47, 48, 85, 87, 90
642	25, 27, 29, 33, 34, 35, 39, 40, 41, 48, 85, 87, 88, 90
1302	25, 27, 29, 32, 33, 34, 35, 39, 40, 41, 48, 85, 90
1513	25, 27, 29, 32, 33, 34, 35, 39, 40, 41, 48, 85, 87, 90
1128	1, 25, 27, 29, 31, 32, 33, 85, 87, 88, 90, 92
1266	27, 29, 31, 32, 33, 35, 81, 85, 87, 88, 90, 92
1454	25, 27, 29, 31, 32, 33, 35, 85, 87, 88, 90, 92
85	9, 44, 45, 46, 52, 53, 54, 56, 59, 60, 72, 73

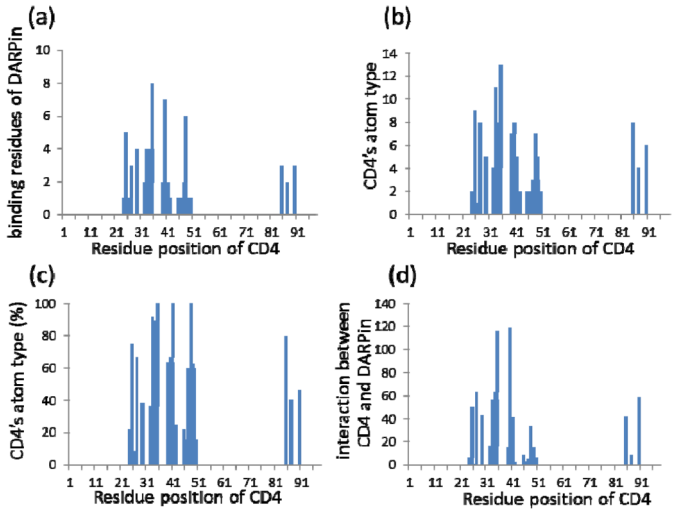

Figure 2. Example of pose 26, one of cluster 1, the binding data are shown in four criteria; the first (a), the second (b), the third (c), and fourth (d) criterion.

and E92. The identical CD4's residues considered between cluster 1 and 2 were H27, Q33, E85, E87, and K90. These results may coincide with the key amino acids of CD4.

In each pose, the normalized histogram values were carried out to detect the top 3 of the key CD4's residues by using maximum detection. The top 3 residues were shown in table 2. In cluster 1, K35 was found the most key CD4's residues, with 100% probability, followed by Q40, Q25, Q33, and P48 with probability of 85.7%, 57.1%, 42.9%, and 14.3 respectively. The results implied that K35 is the most probable CD4's amino acid to be considered as the key residue of CD4-DARPin interaction because of the above reason. Similarly, it had 63.6% probability when we considered throughout 11 poses. Moreover, the possibility of the first residue was 71.4%. Q40 and Q25 was the secondary and tertiary important residue in the cluster 1. In cluster 2, N32, Q33, and K90 had 100 % probability as the top 3 key residues. When we considered possibility of the first key residues, Q33 and K90 had

probability of 66.7% and 33.3%, respectively. Furthermore, Q33 was found in cluster 1 with a probability of 42.9%, and at 54.5% when considered throughout the 11 poses. Hence, in cluster 2, Q33 was the most probable key residue, followed by K90 and N32, respectively. In cluster 3, R59, S46, and N52 were the first, second, and third key residues for binding with DARPin. Therefore, the key amino acids in cluster 1 were K35 (as shown in figure 3) Q40, and Q25; cluster 2 were Q33, N32, and K90; cluster 3 were R59, K46, and N52. All key CD4's amino acids were located on the binding site of gp120 specific CD4 within a 5 Å radius, except for K90 and D93 (data not shown). According to the key amino acids, K35 and R59, were found to be parts of the critical residues on CD4 that gp120 recognition, which were Lys29, Lys35, Phe43, Leu44, Lys46, Gly47 and Arg59; these were studied by biochemical mutagenesis (Phe43 and Gly47) [18] and compiled from mutagenesis studies by Ryu et.al [19]. Specific mutation experiments to confirm the theoretical investigation of key CD4's amino acids will be undertaken and observed for the interactions between mutated CD4 and DARPin.

Table 2. The top 3 key CD4's amino acids bound to DARPin

Pose	Key CD4 amino acid		
	1st	2nd	3rd
16	35	25	40
21	40	35	25
26	35	40	33
241	40	35	25
642	35	40	25
1302	35	48	33
1513	35	33	40
1128	33	90	32
1266	33	90	32
1454	90	32	33
85	59	46	52

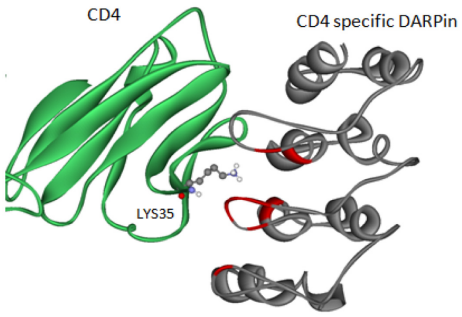

Figure 3. Complex structure of CD4 bound to DARPin. The CD4:LYS35 is one of validating amino acid on CD4 related with DARPin binding site. The relative binding residues of CD4-specific DARPin are marked in red color.

4. CONCLUSIONS

The binding site pattern from our simulation of CD4-ankyrin docking was related with Schweizer *et al.* [9] report. Although we obtained different binding patterns, the CD4-DARPin binding regions of all three patterns overlap with CD4-gp120 interacting region. Moreover, some key amino acids from our model are part of the crucial CD4 that bind to gp120 [18, 19]. They are the candidates for finding the relevant amino acids of ankyrin, which will be mutated further to enhance the binding affinities. Also, from this study, we have shown that effective data analysis for a very large dataset can be done successfully in MATLAB.

ACKNOWLEDGEMENTS

This work was supported by the Thailand Research Fund (TRF), the National Centre

for Genetic Engineering and Biotechnology (BIOTEC) of the National Science and Technology Development Agency (NSTDA), the National Electronics and Computer Technology (NECTEC) for Discovery software. We would like to thank Associate Professor Nipon Theera-Umpon and Assistant Professor Sansanee Auephanwiriyakul for consultations on the data analysis.

REFERENCES

[1] Lieberman J., Defying Death-HIV Mutation to Evade Cytotoxic T Lymphocytes*, N. Engl. J. Med.*, 2002; **347:** 1203-1204.

[2] Troyer R.M., McNevin J., Liu Y., Zhang S.C., Krizan R.W., Abraha A., and et al., Variable Fitness Impact of HIV-1 Escape Mutations to Cytotoxic T Lymphocyte (CTL) Response, *PLoS Pathog.,* 2009; **5**: e1000365.

[3] Bolmstedt A., Hemming A., Flodby P., Berntsson P., Travis B., Lin J.P., and et al., Effects of Mutations in Glycosylation Sites and Disulphide Bonds on Processing, CD4-Binding and Fusion Activity of Human Immunodeficiency Virus Envelope Glycoproteins, *J. Gen. Virol.,* 1991; **72 (Pt 6):** 1269-1277.

[4] Pugach P., Ketas T.J., Michael E., and Moore J.P., Neutralizing Antibody and Anti-Retroviral Drug Sensitivities of HIV-1 Isolates Resistant to Small Molecule CCR5 Inhibitors, *Virology*, 2008; **377:** 401-407.

[5] Bruno C.J., and Jacobson J.M., Ibalizumab: an Anti-CD4 Monoclonal Antibody for the Treatment of HIV-1 Infection, *J. Antimicrob. Chemother.,* 2010; **65:** 1839-1841.

[6] Binz H.K., Stumpp M.T., Forrer P., Amstutz P., and Pluckthun A., Designing Repeat Proteins: Well-Expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins, *J. Mol. Biol.*, 2003; **332:** 489-503.

[7] Sedgwick S.G., and Smerdon S.J., The Ankyrin Repeat: a Diversity of Interactions on a Common Structural Framework, *Trends Biochem. Sci.,* 1999; **24:** 311-316.

[8] Zahnd C., Wyler E., Schwenk J.M., Steiner D., Lawrence M.C., McKern N.M., and et al., A Designed Ankyrin Repeat Protein Evolved to Picomolar Affinity to Her2, *J. Mol. Biol.,* 2007; **369:** 1015-1028.

[9] Schweizer A., Rusert P., Berlinger L., Ruprecht C.R., Mann A., Corthesy S., and et al., CD4-Specific Designed Ankyrin Repeat Proteins are Novel Potent HIV Entry Inhibitors with Unique Characteristics, *PLoS Pathog.,* 2008; **4:** e1000109.

[10] Pugach P., Krarup A., Gettie A., Kuroda M., Blanchard J., Piatak M. and et al., *In Vivo* Binding and Retention of CD4-Specific DARPin 57.2 in Macaques, *PLoS One*, 2010; **5:** e12455.

[11] Nimmanpipug P., Khampa C., Lee V.S., Nangola S., and Tayapiwatana C., Identification of Amino Acid Residues of a Designed Ankyrin Repeat Protein Potentially Involved in Intermolecular Interactions with CD4: Analysis by Molecular Dynamics Simulations, *J. Mol. Graph. Model.,* 2011; **31:** 65-75.

[12] Chen R. and Weng Z., A Novel Shape Complementarity Scoring Function for Protein-Protein Docking, *Proteins,* 2003; **51:** 397-408.

[13] Chen R., Li L., and Weng Z., ZDOCK: an Initial-Stage Protein-Docking Algorithm, *Proteins,* 2003; **52:** 80-87.

[14] Sobie E.A., An Introduction to MATLAB, *Sci. Signal.,* 2011; **4:** tr7.

[15] Li L., Chen R., and Weng Z., RDOCK: Refinement of Rigid-Body Protein Docking Predictions, *Proteins,* 2003; **53:** 693-707.

[16] Garrett T.P., Wang J., Yan Y., Liu J., and Harrison S.C., Refinement and Analysis of the Structure of the First Two Domains of Human CD4, *J. Mol. Biol.,* 1993; **234:** 763-778.

[17] Zhou T., Xu L., Dey B., Hessell A.J., Van R.D., Xiang S.H., and et al., Structural Definition of a Conserved Neutralization Epitope on HIV-1 gp120*, Nature,* 2007; **445:** 732-737.

[18] Wu H, Myszka DG, Tendian SW, Brouillette CG, Sweet RW, Chaiken IM, and et al., Kinetic and Structural Analysis of Mutant CD4 Receptors that are Defective in HIV gp120 Binding, *Proc. Natl. Acad. Sci. U. S. A.,* 1996; **93:** 15030-15035.

[19] Ryu S.E., Truneh A., Sweet R.W., and Hendrickson W.A., Structures of an HIV and MHC Binding Fragment from Human CD4 as Refined in Two Crystal Lattices, *Structure,* 1994; **2:** 59-74.