CHAPTER 5

CONCLUSIONS

DARPins 23.2 had been proved that it potentially and specifically binds to CD4 leading to block CD4-gp120 interaction *in vitro* [14], however, not *in vivo* [15]. In animal model, maybe due to the weak interaction between CD4 and DARPin 23.2, so, insight the interaction between them may provide a clue for improving the binding activity of DARPin. To clarify the CD4-DARPin 23.2 interaction, in this study, the hot spots (binding amino acids) on CD4 bound to DARPin are identified by incorporation of computational simulation, bio-information, and computer programming. The advantage of this study is not only getting hot spots of CD4 to DARPin but also exhibiting the procedure to filter the rational complex and creating a new algorithm for identifying hot spots.

The computational model is used to generate protein-protein complexes, called protein-protein docking method, for further identifying key binding residues. The computational method starts with searching 3D structure of CD4 and DARPin 23.2 molecule. The CD4 X-ray structure gets from protein data bank, PDB code: 3CD4 [26]; and the DARPin 23.2 is constructed by homology modeling [109]. Both of proteins are prepared before docking by cleaning, typing force field, and minimizing energy. The minimized proteins can be further carried out to dock because these structures show (1) high negative potential energy (Table 5), (2) RMS gradient closer

zero (Table 5), and (3) a large number of residues in the core and allowed region of Ramachandran plot (Table 6 and Fig. 17). The 2,000 docked poses generating from ZDOCK protocol [49] are filtered to find the rational complex. In 2,000 poses, first filter, the docked complexes with top 20 ZDock score are kept. Then, the Schweizer's experiment [14] which demonstrates DARPin 23.2 binds domain 1 of CD4 is available to filter the 20 complexes. There are 11 out of 20 poses related to above experiment. The knowledge of interaction of CD4-gp120 [34] and CD4-MHCII [28] are available to make ensure that all of 11 poses are rational to describe interaction of CD4-DARPin 23.2. The reason is that there are overlapping area on CD4 between DARPin 23.2 as well as gp120 and also MHCII (Table 7 and Figs. 18, 19). Next filter, the other docking scores which is not ZDock such as Cluster, ZRank, and E_RDock are used to eliminate unreliable complex. Those of 11 poses are divided into three groups based on the same orientation between CD4 and DARPin 23.2. The group 2 and 3, totally four poses, are eliminated because these groups present (1) low cluster parameter, (2) low negative ZRank score when comparing with group 1, and (3) no negative E_RDock scoring (Table 8 and Fig. 20). The seven remained poses in group 1, ZRank and E_RDock scores are used to decide the reliable complex which is the last filter for docking score. There are three complexes, p16, p26 and p1513, showing the good scoring. The p16 shows highest negative ZRank but low negative E_RDock, whereas, the p1153 shows the inverse values. The p26 does not present the highest ZRank and E_RDock score, however, both scores of p26 show high value. Nevertheless, those of 3 complexes had the same orentation between CD4 and DARPin 23.2 (Fig. 21).

To increase the chance of hot spot identification, not only three poses but 11 poses are performed to find intermolecular neighbors. Since the intermolecular neighbor data are complicated, the computer programming is available for management these data. The designing algorithm (Figs. 12-15) to identify key CD4 residues starts with converting the neighbor data into histogram values relied on four criteria. Next, the top 10 histogram value in each criterion is merged together to be considered CD4 residues. Then the combination of four criteria and hydrogen bond property is created into six patterns. Last step, the histogram value of considered CD4 residues in six patterns is performed to identify the key CD4 residues. The results show that 1^{st}-3^{rd} key CD4 residues in group 1 are Lys35, Gln40, and Gln25; in group 2 are Gln33, Lys90, and Asn32; and in group 3 are Arg59, Asp53, and Lys46 (Table 12). These 1^{st}-3^{rd} key CD4 residues and bio-information are analyzed to identify hot spots of CD4 to DARPin 23.2. All of these residues excepting Asp53 and Lys90 located on the binding site CD4 bound to gp120 within 5 Å (refer to Table 7 and 12) [34]. Interestingly, Lys35 and Arg59 are overlapping residues with critical residues on CD4 recognizing gp120 are studied by biochemical mutagenesis [32]. Focusing on interface area on CD4 to MHCII at 5 Å [28], Lys35, Gln40, and Lys46 were part on this area (refer to Table 7 and 12). Incidentally, binding CD4 residues to MHCII studied by site-directed mutagenesis [27] are overlapping residues with our prediction only 2 residues (Lys35 and Arg59). Altogether, Lys35 is the most important hot spots and Gln33, Gln40, as well as Arg59 were second order of hot spots of CD4 to DARPin 23.2.

To validate the hot spots predicted from our algorithm, they are checked with physical properties of hot spot [58, 63-69] and also compared with HotPOINT [78]

and HSPred [85, 86] programs. In case of validation, the hot spot predicted by our algorithm are defined as 1^{st}-3^{rd} key CD4 residues. Interestingly, our prediction accords to both software, which hot spots of CD4 bound to DARPin are somewhat hydrophilic residues (Tables 14-17). Although several studies suggested that hydrophobic residues are dominate in protein-protein interface [63-69], hydrophilic have been reported to be preferred in interface area [63-67]. Moreover, our predicted hot spots show the low (Gln and Asn), moderate (Lys and Asp), and high (Arg) propensity of hot spot based on Bogan & Thorn (Table 13) [58]. In case of finding the identical hot spot between our algorithm and two programs, the results show that our prediction has high identity with both methods (Table 17) by calculating the $PID^{predict}$. So, we can conclude that our algorithm is reliable for predicting hot spots.

A benchmark of our algorithm finding hot spot describes in terms of strength and benefit. The first strength is that our algorithm can identify the order of hot spots i.e. most crucial hot spots etc. Whereas, HotPOINT and HSPred software are able to discriminate between hot spot and non-hot spot, however, not ordering the hot spot. The benefit of this strength is that it is easy for user to choose the hot spot for *in vitro* mutation by following the order of hot spot. In other words, both servers provide a large number of hot spots, especially in HotPOINT giving 6-8 hot spots. It is increasing the payment if all un-ordering hot spots are performed to *in vitro* mutate in the same time. However, the threshold to distinguish hot spot and non-hot spot does not mention in our algorithm. The second good point is that the scoring function of our algorithm uses information of pair potential interface between CD4's amino acid and DARPin's amino acid, CD4's atom and DARPin's atom, CD4's atom and DARPin's amino acid, as well as atom making hydrogen bond. Whereas, both

softwares use information of pair potential interface between CD4's atom and DARPin's atom which is applied to calculate scoring with other functions such as ASA and energy terms. The benefit of this point suggests that our algorithm increases the chance to get the hot spot. Moreover, the six patterns which are the various criteria combination help to enhance the accuracy of hot spot.