# CHAPTER 4

# RESULTS AND DISCUSSIONS

This research study aims to; (1) predict the rational CD4 – DARPin 23.2 complex structure, (2) create an algorithm template histogram-based analysis for identifying the key residues of CD4 interacting with DARPin 23.2, and (3) validate the predicted results by checking the physicochemical properties and comparing with other programs. In this chapter, the results and discussions are divided into three main parts: (1) computational simulation, (2) designing algorithm, and (3) validating the results.

## Computational Simulation

### Searching and preparing for 3D structures

There are 39 crystal structures of human CD4 (Table 17) that has been reported in protein data bank (updating on 2013, June 6$^{th}$). The proposed structure template of CD4 in this study is structure that CD4 did not bind with other proteins and CD4 consisted of at least domain 1 and 2 of CD4 without mutation. Among four possible structures, the PDB code: 3CD4 [26] was selected (Fig. 16) with the lowest resolution, 2.2 Å. For 3D structure of DARPin 23.2, this structure was constructed by homology modeling as described in the previous study [109].

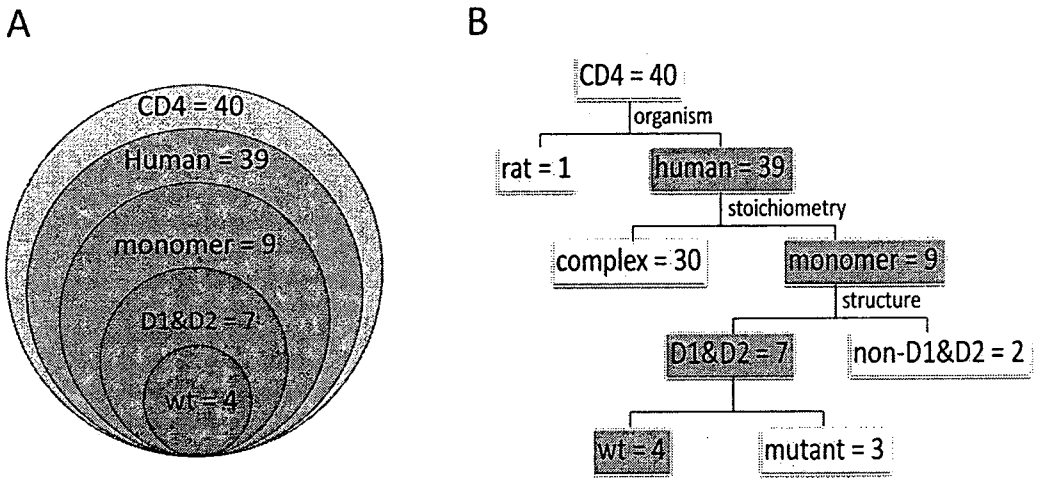


**Figure 16** Diagram data flow of selecting proposed CD4 showing in Venn diagram (A) and tree diagram (B). There are 40 PDB files reported as CD4 structures (updating on 2013, May 1st). The 40 structures consist of a rat CD4 and 39 human CD4. For 39 human CD4, the structures are solved in formation of complex structure (30) and single structure (9). In monomer structures, the structures divide into two groups; (1) structures present full domain 1 and 2 of CD4 and (2) structures are not part of domain 1 and 2. The full structures of domain 1 and 2 of CD4 consist of wide type (wt) forms and mutant form.

The structure of human CD4 and DARPin 23.2 were optimized and alternate conformation of Lys72 and Asn73 were deleted. Moreover, the incomplete residues, Asp105 and Ala178, were corrected. For DARPin 23.2, the atom names of Gln121 were standardized and the atoms in 110 residues were reordered. Then, the cleaned structures could be further typed force field and minimized energy. The aim of energy minimization of both proteins was preparation of structure before docked process by relaxing the structures and eliminating the steric overlapping. The outputs of minimization were presented in term of potential energy and RMS gradient. The words of "Initial" and "Final" as shown in Table 5 were defined as "before

performing minimization" and "after performing minimization", respectively. To estimate the correctness of structure, the initial and final energy were compared and also RMS gradient. The results showed that the final potential energy of both structures is more negative valuable than the initial potential energy as shown in Table 5. For the RMS gradient value, the structure which was very close zero indicated the stable state. The final RMS gradient of both CD4 and DARPin 23.2 was closer to zero than the initial RMS gradient.

**Table 5** The potential energy and RMS gradient of CD4 and DARPin 23.2

| name | initial potential energy (kcal/mol) | final potential energy (kcal/mol) | initial RMS gradient ($kcal \cdot mol^{-1} \cdot Å^{-1}$) | final RMS gradient ($kcal \cdot mol^{-1} \cdot Å^{-1}$) |
|---|---|---|---|---|
| CD4 | -106.873 | -11804.9279 | 362.91254 | 0.09795 |
| DARPin 23.2 | -5821.73201 | -7675.11933 | 39.6614 | 0.09848 |

Moreover, the stereochemistry quality of minimized proteins was validated using the Ramachandran plot as shown results in Table 6 and Fig. 17. The minimized CD4 structures showed a large number of residues in the core and allowed region (96.8%). There was an amino acid residue (0.9%) which was Ser124 in disallowed region. However, it did not belong to the target region which located in domain 2 of CD4. Same as CD4, the DARPin 23.2 was found to have a large number of residues in core and allowed region (97.1%). Although there were 2 residues (1.9%), Leu2 and His57, in disallowed region, these residues were included in the constant part of ankyrin which did not be the binding part. Therefore, both of minimizing CD4 and DARPin 23.2 were accepted for further protein-protein docking simulation.

**Table 6** The Ramachandran plot of minimizing structures of CD4 and DARPin 23.2

| name | region of Ramachandran plot (% residue) | | | |
|---|---|---|---|---|
| | cored | allowed | generously allowed | disallowed |
| CD4 | 66.90% | 29.90% | 2.50% | 0.60% |
| DARPin 23.2 | 67.90% | 29.20% | 0.90% | 1.90% |

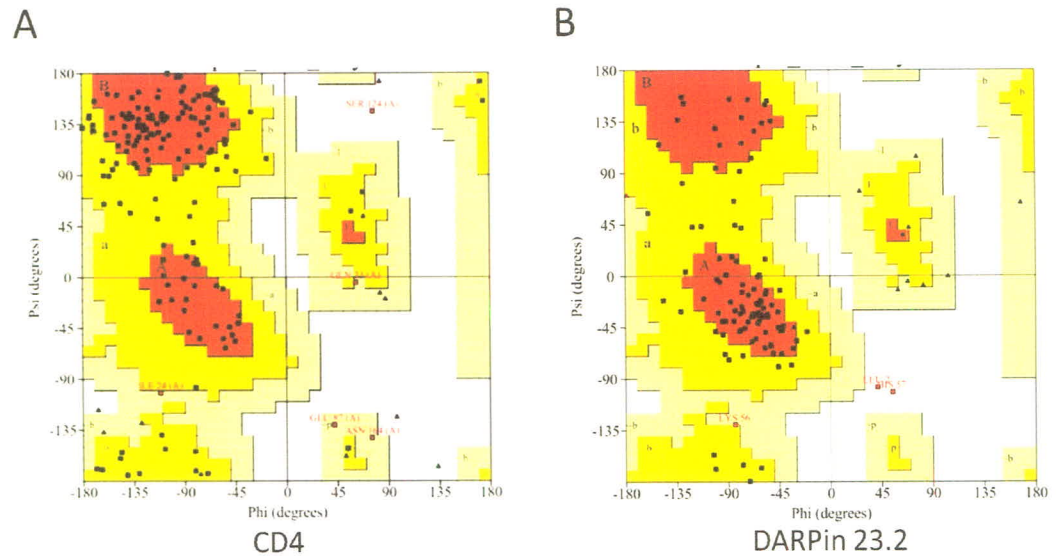


**Figure 17** The Ramachandran plot of minimized CD4 (A) and minimized DARPin 23.2 (B). The red, strong yellow, light yellow, and whit color are core region, allowed region, generously allowed region, and disallowed region respectively. The black points are the amino acid residues which the number of residues related with Table 6.

**Predicting the heterodimer complex**

The 2,000 docked poses with the highest positive ZDock scoring from ZDOCK protocol were reported. The top 20 ZDock scores were selected to put into RDOCK protocol for refining the complex structures. The results showed that there are 11 out of 20 poses that DARPin 23.2 binds to CD4 on domain 1. Therefore, nine poses,

DARPin 23.2 bound to CD4 on domain 2 were excluded because Schweizer's experiment. It demonstrated that human-CD4-spcific DARPins bind to domain 1 of CD4 for interfering interaction between CD4 and gp120 as well as CD4 and MHCII [14]. The 11 poses were divided, based on orientation of complex (cluster), into three groups as shown in Figs. 18 C, D, E; seven poses in the first group (pose (p) 16, p21, p26, p642, p1302, and p1513); three poses in the second group (p1128, p1266, and p1454); and one pose (p85) in the last group. Those of three groups were filtered to get the rational complex by x-ray information concerning binding site of CD4 bound to gp120 and MHCII as shown in Figs.18A, B). The result showed that the binding area of CD4 bound to DAPRin 23.2 in all three groups are overlapped with the binding site of CD4 binding to gp120 and also MHCII as shown in Table 7 and Fig. 19. Althought the binding site of CD4 to both gp120 and MHCII had not been used for eliminating anygroup from three groups, those of them were rational to describe binding area of CD4 interacting DARPin 23.2.

**Table 7** The inteface residue of CD4 bound to gp120, MHCII, and DARPin 23.2 complex at 5.0 Å

| | interface residues on CD4 | | | | |
|---|---|---|---|---|---|
| | **2NXY** | **1JL4** | **group 1; p1513** | **group 2; p1128** | **group 3; p85** |
| **PPI** | S23, Q25, H27, K29, S31, N32, Q33, I34, K35, Q40, G41, S42, F43, L44, W45, K46, G47, P48, N52, R59, S60, L61, W62, D63, Q64, E85 | K35, Q40, S42, F43, L44, T45, K46, G47, P48, S60, D63 | I24, Q25, F26, H27, K29, N32, Q33, I34, K35, N39, Q40, G41, S42, T45, K46, G47, P48, S49, K50, N52, E85, E87, D88, K90 | K1, Q25, F26, H27, K29, N30, S31, N32, Q33, I34, K35, T81, I83, E85, V86, E87, D88, Q89, K90, E91, E92, Q94 | G9, D10, T11, F43, L44, T45, K46, G47, P48, S49, K50, L51, N52, D53, R54, A55, D56, S57, R58, R59, S60, I70, K72, N73, K75 |
| **OA[a]** | ND | K35, Q40, S42, F43, L44, T45, K46, G47, P48, S60, D63 | Q25, H27, K29, N32, Q33, I34, K35, Q40, G41, S42, T45, K46, G47, P48, E85 | Q25, H27, K29, S31, N32, Q33, I34, K35, E85 | F43, L44, T45, K46, G47, P48, R59, S60 |
| **OA[b]** | ND | ND | K35, Q40, S42, T45, K46, G47, P48 | K35 | F43, L44, T45, K46, G47, P48, S60 |

OA[a] is overlapping area between CD4-gp120 and CD4-DARPin 23.2; OA[a] is overlapping area between CD4-MHCII and CD4-DARPin 23.2; 2NXY is crystallography structure of CD4-gp120 complex; 1JL4 is crystallography structure of CD4-MHCII complex; ND is not done.

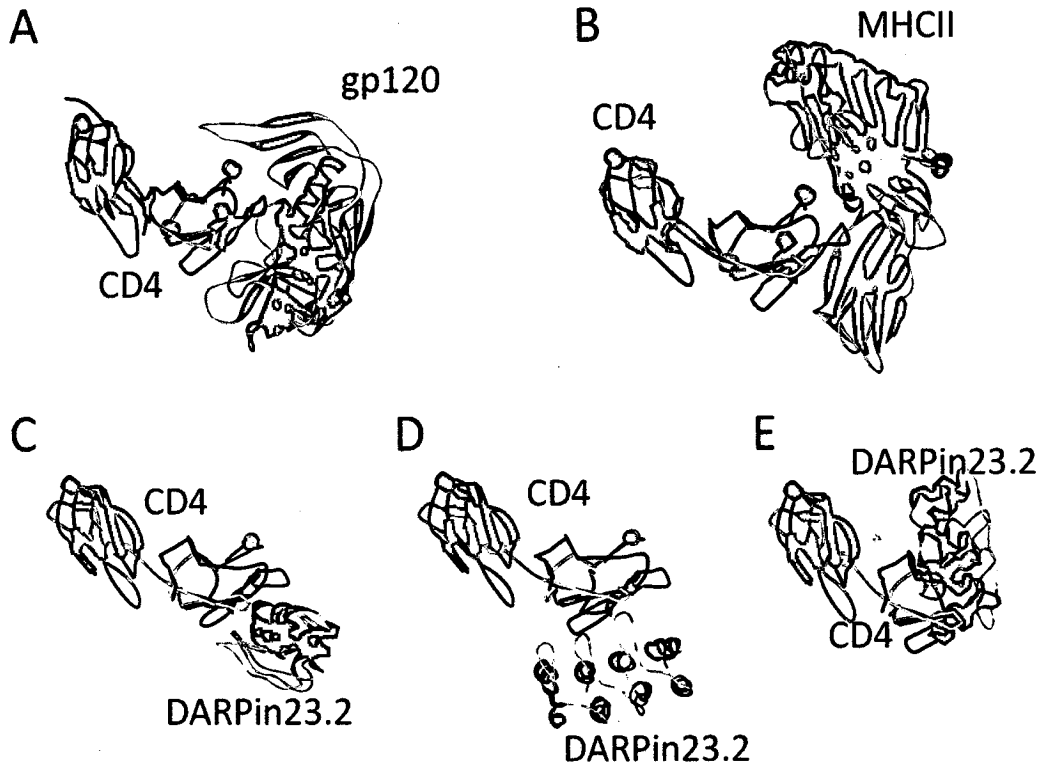


**Figure 18** The structures of CD4-gp120 complex, CD4-MHCII complex and CD4-DARPin 23.2 complexes. The X-ray structure of CD4-gp120 complex (PDB code: 2NXY) (A) [34] and CD4-MHCII complex (PDB code: 1JL4) (B) [28] shows that gp120 binds CD4 on domain 1 and MHCII binds domain 1 of CD4 as well. The docked structures of CD4-DARPin 23.2 in group 1, p1513 (C); group 2, p1128 (D); and group 3, p85 (E) demonstrate that DARPin 23.2 binds domain 1 of CD4.

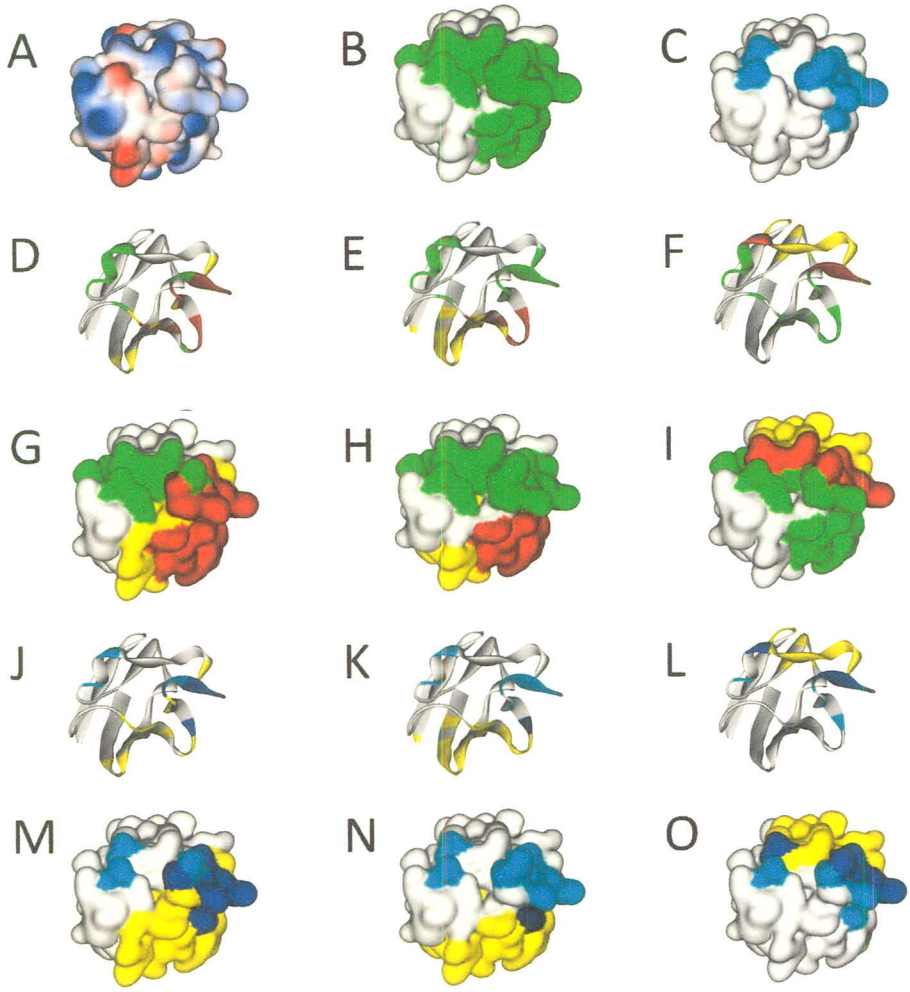


**Figure 19** The binding area on the tip of domain 1 of CD4. (A), Electrostatic potential surface. Red and blue referred to negative charges and positive charges, respectively. (B), Binding region of CD4 bound to gp120 (PDB code: 2NXY) representing in green. (C), Interface area of CD4 bound to MHCII (PDB code: 1JL4) representing in sky. The overlapping region on CD4 between gp120 and DARPin 23.2 show in ribbon style (D – F) and surface style (G – I); first group (D and G), second group (E and H), and third group (F and I). The green region is CD4 bound to gp120, the yellow region is bound by DARPin 23.2, and the red area is bound by both gp120 and DARPin 23.2. Characteristics of the overlapping interface region on domain 1 of CD4 between MHCII and DARPin 23.2 show in ribbon style (J – L) and surface style (M – O); first group (J and M), second group (K and N), and third group (L and O). The sky region is CD4 bound to MHCII, the yellow region is bound by DARPin 23.2, and the blue area is bound by both MHCII and DARPin 23.2.

Moreover, the docking score values were applied to decide the rational complex. Since DS program provides scoring more than one value to decide the good complex, in this case, the ZDock, Cluster, ZRank, and E_RDock were used to decide together. The ZDock score is based on PSC, DE as well as ELEC and recommended that the high value is good complex. In all three groups, the ZDock score had high value when comparing with 2000 poses (Table 8 and Fig. 20A). This value could not eliminate any pose in 11 complex, however, all 11 poses had a good reason to be candidate complex. For cluster parameter, the small value represents the good complex. The group 1 had the lowest value of cluster (the value was 8), as shown in Fig. 20B, and also the most member (seven poses). In ZRank and E_RDock score which based on energy calculation are recommended that the high negative value is good complex. The results showed that ZRank score of group 2 and 3 have less negative value when comparing with group 1. Moreover, there were no negative E_RDock value in group 2 and 3. So, selection the group 1 to further analyze was rational when diciding with cluster, ZRank and E_RDock parameters. Within group 1, the cluster parameter did not be used to identify the good complex. The interesting poses were pose number (p) 16, p26, and p1513. The highest negative ZRank score is p16 but it had less negative E_RDock value when comparing with p16 and p1513 (Table 8 and Figs. 20C, D). For p1513, although it had low negative ZRank value, it was the highest E_RDock score. The remained pose, p26, it did not have the best ZRank or E_RDock score, however, it showed high ZRank and E_RDock scores. Nevertheless, those of 3 complexes had the same orentation between CD4 and DARPin 23.2 (Fig. 21).

**Table 8** Docking scores of 11 complexes

| group | PoseNum | ZDock | Cluster | ZRank | E_RDock |
|---|---|---|---|---|---|
| 1 | 16 | 25.24 | 8 | -74.047 | -2.46328 |
| | 21 | 28.65 | 8 | -72.89 | -2.40144 |
| | 26 | 31.66 | 8 | -70.935 | -3.27997 |
| | 241 | 28.09 | 8 | -52.891 | 0.083766 |
| | 642 | 27 | 8 | -42.02 | -0.74093 |
| | 1302 | 26.79 | 8 | -29.252 | -1.45436 |
| | 1513 | 29.25 | 8 | -24.937 | -6.93017 |
| 2 | 1128 | 26.76 | 11 | -32.704 | 4.88564 |
| | 1266 | 26.71 | 11 | -29.992 | 6.2419 |
| | 1454 | 26.41 | 11 | -26.192 | 8.64862 |
| 3 | 85 | 26.93 | 87 | -61.523 | 2.29653 |

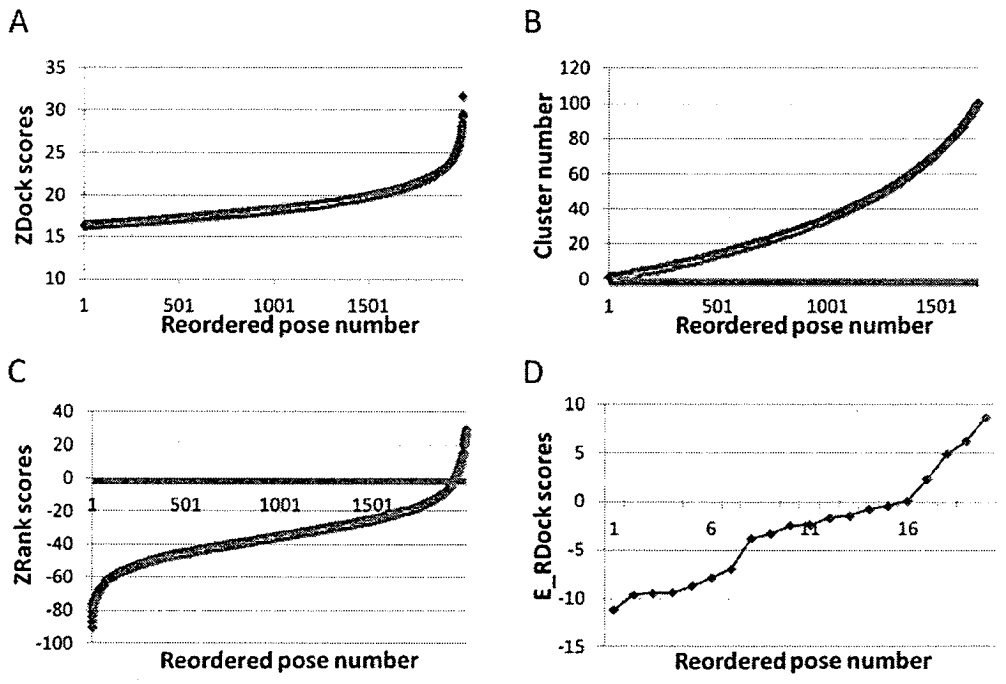


**Figure 20** The docked scoring of docked poses. The pose number of docked poses is reordered depended on each docked score. (A), Sorting pose number from small to large positive ZDock score, 2000 poses. (B), Sorting pose number from small to large cluster, 1616 poses. (C), Sorting pose number from more negative value to more positive value of ZRank score, 2000 poses. (D), Sorting pose number of E_RDock score from more negative value to more positive value, 20 poses.

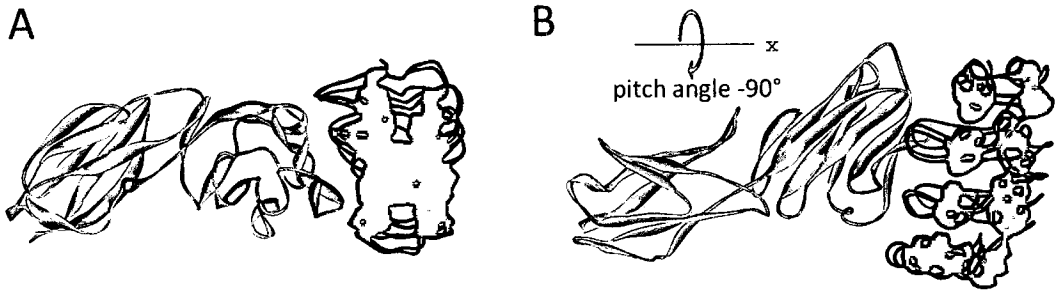


**Figure 21** The superimposition of three rational CD4 – DARPin 23.2 complexes. The rotated form along X axis -90° of (A) is presented in (B). The gray, pink, blue, and green colors are CD4, DARPin in p16, DARPin in p26, and DARPin inp1153 respectively.

**Finding the intermolecular/interface neighbor**

Although three rational complexes showing good docked scores could be used to describe mechanism, identifying hot spots should be solved in all 11 poses. The reason was that increasing the chance to find the hot spots from different orientation of complex. The hot spots were analyzed from $1^{st}$-$3^{rd}$ key residue of all 11 complexes. For the key binding residues, they were identified from intermolecular/ interface neighbor. All 11 docked complexes were carried out to find the intermolecular neighbors, using DS 2.5, with distance threshold of 5.0 Å. The characteristic of intermolecular neighbor was protein:residue:atom - protein:residue:atom as shown in Fig. 22. Moreover, the hydrogen bonds in each pose were identified as shown in Table 9. Since, the number of intermolecular neighbors in each pose was huge, averaged 1,000 pairs, it was hard to manually manage these data. So, the programming language was available to create algorithm to speedily and precisely identify key binding residues.

**Table 9** The hydrogen bond in 11 poses with the distance cutoff and angle as 2.5 Å and 120-180°.

| Donor - Acceptor (protein:residue:atom) | Donor - Acceptor (protein:residue:atom) |
|---|---|
| **p16** | **p1266** |
| A:GLN40:HE21 - D:ALA43:O | A:LYS29:HZ3 - D:GLU65:OE2 |
| A:GLN40:HE22 - D:LEU74:O | A:LYS90:HZ1 - D:LEU34:O |
| A:LYS90:HZ3 - D:THR33:O | A:LYS90:HZ3 - D:LEU34:O |
| D:ARG36:HH12 - A:GLN25:OE1 | **p1302** |
| **p21** | A:LYS35:HZ2 - D:GLU65:OE2 |
| A:LYS35:HZ2 - D:GLU65:OE2 | D:LYS99:HZ2 - A:ASN32:O |
| A:GLN40:HE21 - D:ALA44:O | **p1454** |
| A:GLN40:HE22 - D:LEU74:O | A:LYS29:HZ3 - D:GLU65:OE2 |
| A:LYS90:HZ3 - D:THR33:OG1 | A:GLN33:HE22 - D:HIS40:NE2 |
| D:ARG36:HH12 - A:GLN25:OE1 | A:GLN33:HE22 - D:MET69:O |
| **p26** | A:LYS90:HZ1 - D:LEU34:O |
| A:LYS35:HZ2 - D:GLU65:OE2 | A:LYS90:HZ3 - D:LEU34:O |
| A:GLN40:HE21 - D:ALA44:O | D:LYS99:HZ2 - A:GLU92:OE1 |
| A:GLN40:HE22 - D:LEU74:O | **p1513** |
| D:ARG36:HH12 - A:GLN25:OE1 | A:LYS29:HZ1 - D:GLU66:OE2 |
| **p241** | A:LYS29:HZ2 - D:GLU66:OE2 |
| A:GLN40:HE22 - D:ALA43:O | A:LYS35:HZ1 - D:HIS40:NE2 |
| D:ARG36:HH12 - A:GLN25:OE1 | A:LYS35:HZ2 - D:GLU65:OE2 |
| **p642** | A:LYS90:HZ3 - D:THR33:O |
| A:LYS35:HZ2 - D:GLU65:OE2 | D:ARG36:HH12 - A:GLN25:OE1 |
| A:GLN40:HE22 - D:LEU74:O | D:LYS99:HZ2 - A:ASN32:O |
| A:LYS90:HZ3 - D:THR33:OG1 | **p85** |
| D:ARG36:HH12 - A:GLN25:OE1 | A:ARG59:HH22 - D:LEU34:O |
| **p1128** | D:TRP45:HE1 - A:GLY9:O |
| A:LYS29:HZ3 - D:GLU65:OE2 | |
| A:GLN33:HE22 - D:HIS40:NE2 | |
| A:LYS90:HZ1 - D:LEU34:O | |
| D:LYS99:HZ2 - A:GLU92:OE1 | |
| D:TYR111:HH - A:ASN32:OD1 | |

```
A:LYS22:HZ2 - D:ARG11:HH11
A:LYS22:HZ2 - D:ARG11:HH12
A:SER23:C   - D:TRP45:CZ2
A:SER23:O   - D:TRP45:CE2
A:SER23:O   - D:TRP45:NE1
A:SER23:O   - D:TRP45:HE1
A:ILE24:N   - D:TRP45:CZ2
A:ILE24:CA  - D:TRP45:CZ2
A:ILE24:CA  - D:TRP45:CH2
A:ILE24:CG1 - D:ARG11:CB
A:ILE24:CG1 - D:ARG11:CG
A:ILE24:CG1 - D:ARG11:CD
A:ILE24:CG1 - D:ARG11:NE
```

**Figure 22** The characteristic of intermolecular neighbor. The distance cutoff is 5.0 Å.

## Designing algorithm for finding key residue

### Exacting considered CD4 residues

The intermolecular neighbor data in each pose were changed into histogram value relied on five criteria excepting criterion of hydrogen bond as shown in Table 18-21 and Fig. 23. Note that first criterion was defined as the number of DARPin's amino acid positions those were bound to each CD4's amino acid. The second criterion was the number of interactions in each CD4's amino acids bound to DARPin. The third criterion was the number of CD4's atom types in each CD4's amino acids those were bound to the DARPin's residues. The fourth criterion was defined as the percentage of CD4's atom types in each CD4's amino acids those were bound to the DARPin's residues. Then the histogram values in four criteria were carried out to determine the considered CD4 residues and the results were shown in Table 10. In group 1, seven poses, there were 10 considered CD4residues that were the same 100%; these amino

acids were Gln25, His27, Gln33, Ile34, Lys35, Asn39, Gln40, Gly41, Phe48, Glu85, and Lys90. In group 2, there were 10 considered CD4 residues that were the same 100%; these residues were His27, Lys29, Ser31, Asn32, Gln33, Glu85, Glu87, Asp88, Lys90, and Glu92. The identically considered CD4 residues between cluster 1 and 2 were His27, Gln33, Glu85, and Lys90. These amino acids may coincide with the key amino acids of CD binding to DARPin 23.2.

**Table 10** The considered CD4 residues that were a combination of top 10 in criteria 1-4.

| group | pose | positions of considered CD4 residues |
|---|---|---|
| 1 | **16** | 24 25 27 33 34 35 39 40 41 48 85 87 88 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 29 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 33 34 35 39 40 41 48 85 90 |
| | **1513** | 25 27 29 33 34 35 39 40 41 48 85 87 90 |
| 2 | **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 |
| 3 | **85** | 44 45 46 52 53 54 56 59 60 72 73 |

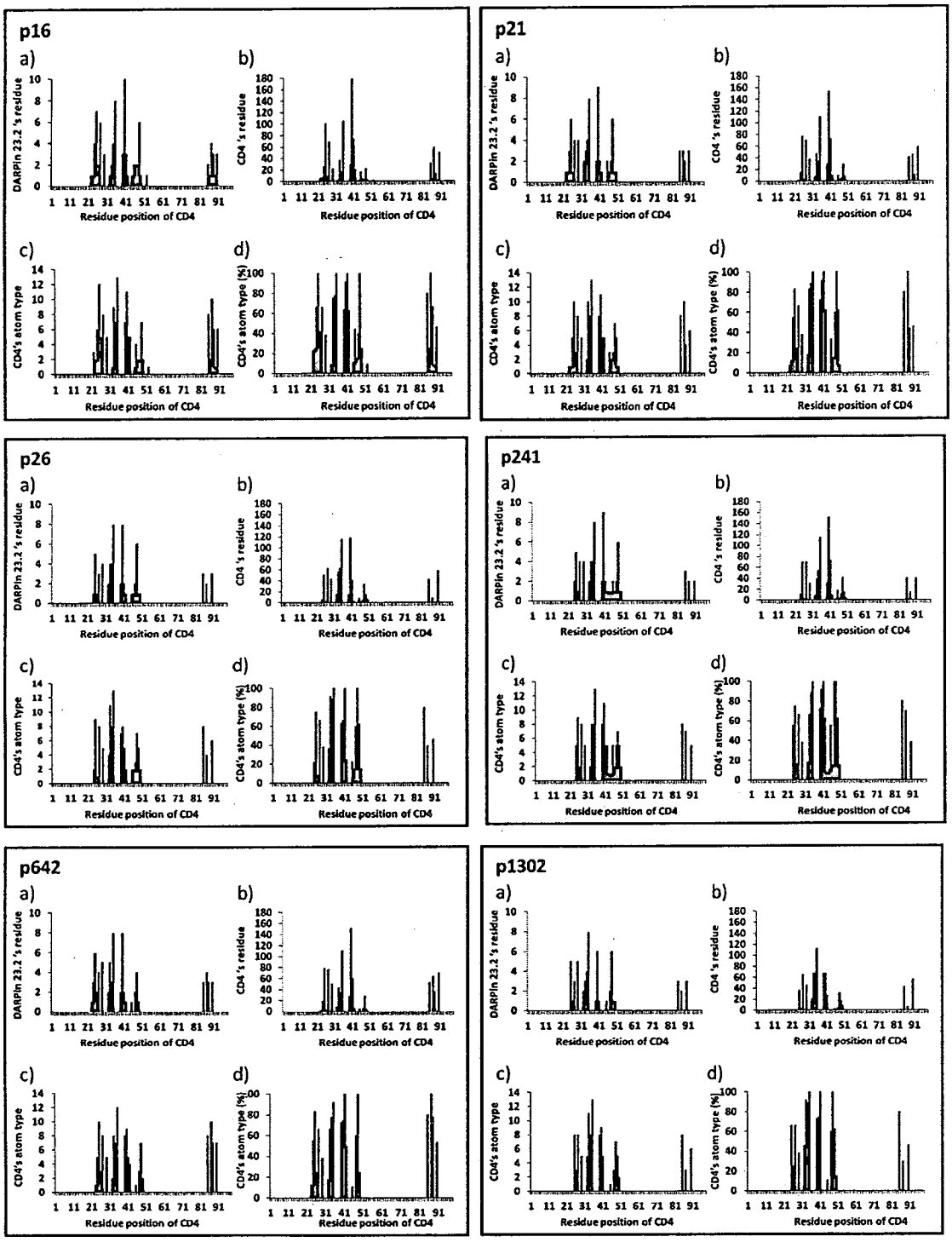


**Figure 23** The intermolecular neighbors in all 11 poses are shown in histogram values in four criteria; the first (a), the second (b), the third (c), and the fourth (d) criterion.

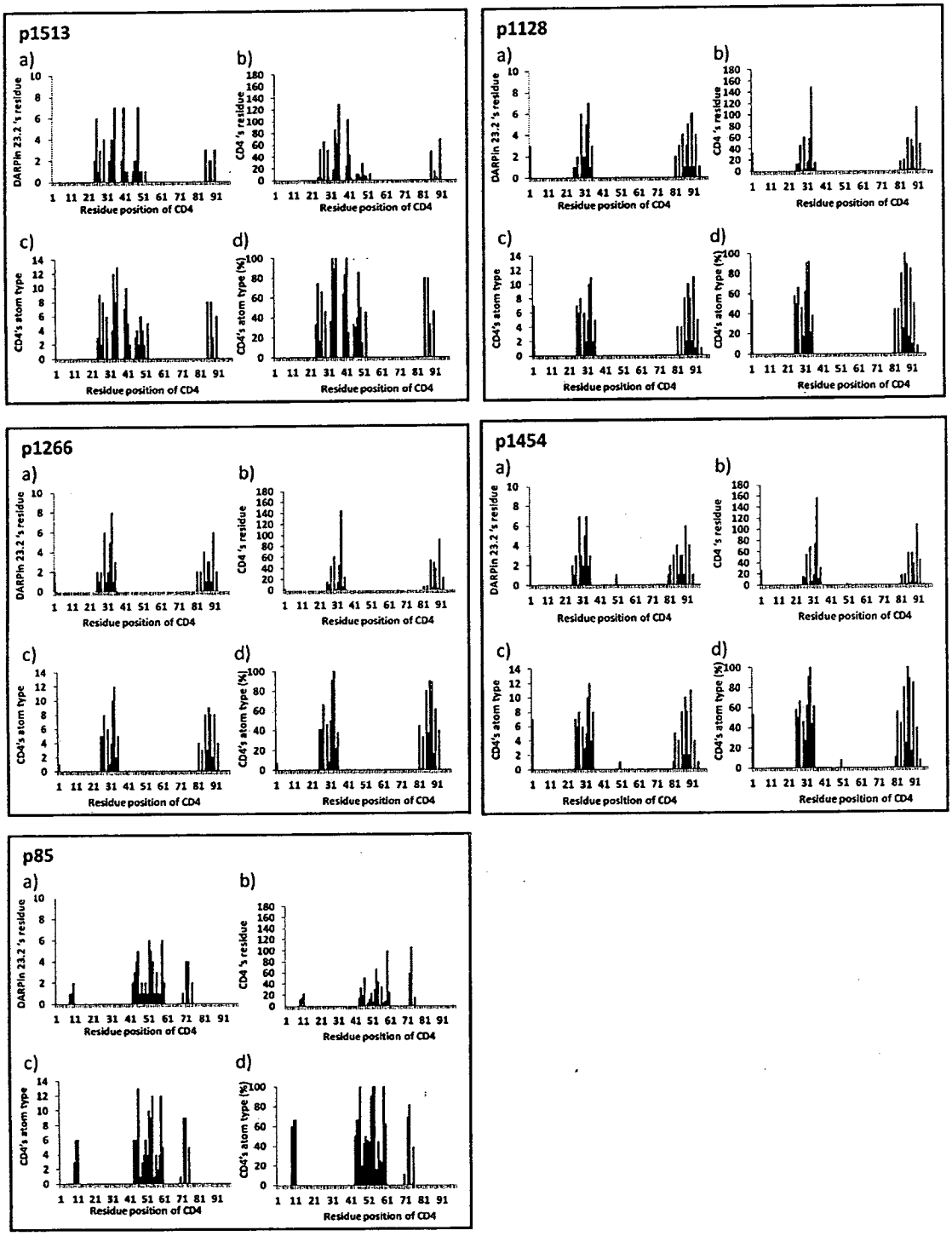


**Figure 23 (continue)** The intermolecular neighbors in all 11 poses are shown in histogram values in four criteria; the first (a), the second (b), the third (c), and the fourth (d) criterion.

**Identifying key binding residues**

The criteria combination was create in six patterns. Then the histogram values of considered CD4 residues in each pattern were combined and normalized. The final normalized histogram values in six patterns were carried out to detect the top 3 key CD4 residues by using maximum detection. The normalized value and $1^{st}$-$3^{rd}$ key residue were shown in Table 11, 12, 22, and 23. In Table 12, cluster 1, there were four residues for $1^{st}$-$3^{rd}$ key residue viz. Lys35, Gln40, Gln25, and Gln33 with probability of 100%, 100%, 57.1% and 42.9% respectively. Considering only $1^{st}$ key CD4 residue, most of this residue is Lys35with probability of 57.1% (four in seven) and following by Gln40 (42.4%; three in seven). The results implied that Lys35 is the interesting residue for focusing on hot spot of CD4 to DARPin 23.2. Gln40 and Gln25 was the secondary and tertiary important residue in the group 1. In group 2, Asn32, Gln33, and Lys90 have 100 % probability as the top 3 key residues. Gln33 was the attentive residue because only Gln33 was the $1^{st}$ key residue and three out of seven residues of Gln33 were found in group 1 (Table 12). Furthermore, Gln33 was part of identically considered CD4 residues between group 1 and 2 (refer to above discussion). Hence, in group 2, Gln33 was the most probable key residue, followed by Lys90 and Gln32, respectively. In group 3, Arg59, Asp53, and Lys46 were the first, second, and third key residues for binding with DARPin. Therefore, the key binding residurs in group 1 were Lys35, Gln40, and Gln25; in group 2 were Gln33, Lys90, and Asn32; and in group 3 were Arg59, Asp53, and Lys46.

**Table 11** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

A. Pose number 16 (p16)

| considered CD4 residues of p16 | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **24** | **25** | **27** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **87** | **88** | **90** |
| **name** | | **I** | **Q** | **H** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **E** | **D** | **K** |
| **pattern** | **A** | -0.79 | 1.43 | -0.14 | -0.52 | -0.58 | 1.26 | -0.82 | 2.37 | -0.26 | -0.03 | -0.58 | 0.27 | -0.97 | -0.64 |
| | **B** | -0.68 | 1.35 | 0.07 | -0.5 | -0.62 | 1.14 | -0.66 | 2.57 | -0.62 | -0.35 | -0.65 | 0 | -0.89 | -0.17 |
| | **C** | -0.7 | 1.3 | -0.15 | -0.74 | -0.57 | 0.95 | -0.86 | 2.58 | 0.08 | 0.12 | -0.69 | 0.11 | -0.92 | -0.52 |
| | **D** | -0.8 | 1.35 | -0.03 | -0.47 | -0.55 | 1.63 | -0.83 | 1.97 | -0.17 | 0.1 | -0.55 | 0.46 | -1.01 | -1.11 |
| | **E** | -0.69 | 1.28 | 0.26 | -0.46 | -0.61 | 1.62 | -0.66 | 2.22 | -0.61 | -0.27 | -0.66 | 0.18 | -0.95 | -0.65 |
| | **F** | -0.7 | 1.17 | 0 | -0.75 | -0.53 | 1.38 | -0.9 | 2.15 | 0.29 | 0.33 | -0.69 | 0.32 | -0.97 | -1.11 |

B. Pose number 21 (p21)

| considered CD4 residues of p21 | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **87** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **E** | **K** |
| **pattern** | **A** | 0.74 | -0.36 | -1.27 | -0.21 | -0.34 | 1.75 | -0.8 | 2.23 | -0.48 | -0.1 | -0.51 | -0.01 | -0.64 |
| | **B** | 0.8 | -0.2 | -0.83 | -0.28 | -0.49 | 1.7 | -0.79 | 2.36 | -0.81 | -0.39 | -0.56 | -0.29 | -0.23 |
| | **C** | 0.7 | -0.39 | -0.83 | -0.48 | -0.37 | 1.55 | -0.94 | 2.41 | -0.14 | 0.06 | -0.58 | -0.23 | -0.47 |
| | **D** | 0.62 | -0.25 | -1.38 | -0.08 | -0.23 | 1.86 | -0.8 | 1.97 | -0.4 | 0.06 | -0.44 | 0.17 | -1.09 |
| | **E** | 0.7 | -0.04 | -0.87 | -0.14 | -0.43 | 1.89 | -0.82 | 2.18 | -0.85 | -0.29 | -0.52 | -0.15 | -0.66 |
| | **F** | 0.54 | -0.27 | -1.24 | -0.39 | -0.24 | 1.66 | -1 | 2.17 | 0.07 | 0.32 | -0.52 | -0.06 | -1.03 |

**Table 11 (Continued)** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

C. Pose number 26 (p26)

| considered CD4 residues of p26 | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **K** |
| **pattern** | **A** | 0.39 | -0.38 | -1.12 | 0.35 | 0.03 | 2.11 | -1.07 | 1.55 | -0.62 | 0.06 | -0.39 | -0.9 |
| | **B** | 0.45 | -0.28 | -0.71 | 0.17 | -0.14 | 2.01 | -1.01 | 1.85 | -1.01 | -0.26 | -0.48 | -0.58 |
| | **C** | 0.33 | -0.48 | -0.97 | -0.01 | 0.02 | 1.89 | -1.19 | 1.9 | -0.35 | 0.21 | -0.49 | -0.84 |
| | **D** | 0.15 | -0.3 | -1.21 | 0.59 | 0.2 | 2.26 | -1.15 | 1.08 | -0.59 | 0.24 | -0.32 | -0.94 |
| | **E** | 0.2 | -0.18 | -0.74 | 0.42 | 0.01 | 2.24 | -1.13 | 1.45 | -1.13 | -0.15 | -0.44 | -0.56 |
| | **F** | -0.01 | -0.42 | -1.08 | 0.21 | 0.25 | 2.08 | -1.37 | 1.43 | -0.25 | 0.5 | -0.44 | -0.9 |

D. Pose number 241 (p241)

| considered CD4 residues of p241 | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **47** | **48** | **85** | **87** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **G** | **P** | **E** | **E** | **K** |
| **pattern** | **A** | 0.92 | -0.11 | -1.05 | -0.32 | 0.05 | 1.59 | -0.52 | 2.34 | -0.24 | -0.65 | 0.22 | -0.26 | -0.77 | -1.21 |
| | **B** | 1.08 | 0.03 | -0.64 | -0.21 | -0.09 | 1.5 | -0.52 | 2.46 | -0.58 | -1.04 | -0.05 | -0.33 | -0.77 | -0.83 |
| | **C** | 0.97 | -0.19 | -0.95 | -0.45 | 0.02 | 1.27 | -0.7 | 2.49 | 0.06 | -0.45 | 0.37 | -0.38 | -0.89 | -1.16 |
| | **D** | 0.37 | -0.01 | -1.12 | -0.26 | 0.18 | 2.01 | -0.49 | 2.06 | -0.16 | -0.65 | 0.39 | -0.19 | -0.8 | -1.32 |
| | **E** | 0.49 | 0.17 | -0.65 | -0.12 | 0.02 | 2 | -0.5 | 2.2 | -0.58 | -1.15 | 0.08 | -0.27 | -0.81 | -0.89 |
| | **F** | 0.26 | -0.08 | -1.04 | -0.41 | 0.18 | 1.76 | -0.72 | 2.17 | 0.24 | -0.41 | 0.63 | -0.32 | -0.96 | -1.3 |

**Table 11 (Continued)** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

E. Pose number 642 (p642)

| considered CD4 residues of p642 | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **87** | **88** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **E** | **D** | **K** |
| **pattern** | **A** | 1.17 | -0.25 | -1.18 | -0.37 | -0.71 | 2.12 | -0.83 | 1.75 | -0.59 | -0.28 | -0.38 | 0.44 | -0.68 | -0.21 |
| | **B** | 1.15 | -0.09 | -0.62 | -0.22 | -0.76 | 2.03 | -0.81 | 1.92 | -1 | -0.66 | -0.45 | 0.1 | -0.74 | 0.16 |
| | **C** | 1.1 | -0.33 | -0.97 | -0.49 | -0.73 | 1.96 | -1.07 | 2.03 | -0.22 | -0.18 | -0.5 | 0.18 | -0.7 | -0.09 |
| | **D** | 0.94 | -0.11 | -1.26 | -0.26 | -0.68 | 2.13 | -0.83 | 1.67 | -0.53 | -0.14 | -0.27 | 0.76 | -0.65 | -0.78 |
| | **E** | 0.93 | 0.11 | -0.58 | -0.06 | -0.77 | 2.07 | -0.83 | 1.93 | -1.07 | -0.64 | -0.35 | 0.37 | -0.73 | -0.37 |
| | **F** | 0.78 | -0.17 | -1.02 | -0.37 | -0.7 | 1.92 | -1.15 | 2.03 | -0.01 | 0.04 | -0.38 | 0.53 | -0.66 | -0.84 |

F. Pose number 1302 (p1302)

| considered CD4 residues of p1302 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **K** |
| **pattern** | **A** | -0.32 | -0.25 | -0.92 | 0.43 | 0.19 | 2.82 | -0.39 | 0.38 | -0.85 | 0.06 | -0.32 | -0.83 |
| | **B** | -0.21 | -0.13 | -0.48 | 0.28 | 0.05 | 2.85 | -0.38 | 0.46 | -1.29 | -0.26 | -0.4 | -0.49 |
| | **C** | -0.41 | -0.33 | -0.75 | 0.12 | 0.25 | 2.91 | -0.51 | 0.34 | -0.66 | 0.23 | -0.41 | -0.78 |
| | **D** | -0.31 | -0.23 | -1.09 | 0.65 | 0.34 | 2.56 | -0.41 | 0.58 | -1 | 0.18 | -0.31 | -0.97 |
| | **E** | -0.16 | -0.06 | -0.53 | 0.49 | 0.18 | 2.57 | -0.39 | 0.73 | -1.61 | -0.23 | -0.42 | -0.55 |
| | **F** | -0.45 | -0.32 | -0.92 | 0.3 | 0.49 | 2.61 | -0.58 | 0.62 | -0.79 | 0.46 | -0.45 | -0.97 |

**Table 11 (Continued)** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

G. Pose number 1513 (p1513)

| considered CD4 residues of p1513 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **33** | **34** | **35** | **39** | **40** | **41** | **48** | **85** | **87** | **90** |
| **name** | | **Q** | **H** | **K** | **Q** | **I** | **K** | **N** | **Q** | **G** | **P** | **E** | **E** | **K** |
| **pattern** | **A** | 0.46 | -0.47 | -0.28 | 0.87 | -0.01 | 2.55 | -1.16 | 0.94 | -0.8 | -0.19 | -0.41 | -0.86 | -0.64 |
| | **B** | 0.58 | -0.32 | 0.25 | 0.59 | -0.2 | 2.47 | -1.04 | 0.96 | -1.28 | -0.35 | -0.49 | -1 | -0.16 |
| | **C** | 0.47 | -0.58 | -0.03 | 0.52 | 0.01 | 2.55 | -1.33 | 0.93 | -0.54 | 0.09 | -0.51 | -1.1 | -0.5 |
| | **D** | 0.29 | -0.33 | -0.95 | 1.16 | 0.18 | 2.19 | -1.09 | 1.23 | -0.69 | -0.02 | -0.26 | -0.76 | -0.94 |
| | **E** | 0.42 | -0.15 | -0.47 | 0.93 | -0.01 | 2.19 | -1.02 | 1.38 | -1.3 | -0.19 | -0.36 | -0.96 | -0.46 |
| | **F** | 0.24 | -0.42 | -0.91 | 0.87 | 0.28 | 2.11 | -1.29 | 1.35 | -0.36 | 0.37 | -0.33 | -1.02 | -0.89 |

H. Pose number 1128 (p1128)

| considered CD4 residues of p1128 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **1** | **25** | **27** | **29** | **31** | **32** | **33** | **85** | **87** | **88** | **90** | **92** |
| **name** | | **K** | **Q** | **H** | **K** | **S** | **N** | **Q** | **E** | **E** | **D** | **K** | **E** |
| **pattern** | **A** | -0.88 | -1.22 | -0.63 | -0.03 | -1.23 | 0.91 | 1.91 | -0.11 | 0.25 | 0.05 | 1.44 | -0.44 |
| | **B** | -0.73 | -1.21 | -0.64 | 0.39 | -1.29 | 0.78 | 1.94 | -0.24 | -0.14 | -0.2 | 1.5 | -0.16 |
| | **C** | -0.95 | -1.38 | -0.79 | 0.26 | -1.1 | 0.83 | 1.94 | -0.13 | 0.01 | 0.06 | 1.35 | -0.11 |
| | **D** | -0.79 | -1.2 | -0.5 | -0.37 | -1.21 | 0.73 | 1.9 | 0.12 | 0.53 | 0.29 | 1.35 | -0.86 |
| | **E** | -0.62 | -1.23 | -0.51 | 0.07 | -1.33 | 0.56 | 2.04 | 0 | 0.12 | 0.05 | 1.48 | -0.63 |
| | **F** | -0.88 | -1.42 | -0.67 | -0.12 | -1.07 | 0.6 | 2 | 0.16 | 0.34 | 0.39 | 1.26 | -0.59 |

**Table 11 (Continued)** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

I. Pose number 1266 (p1266)

| considered CD4 residues of p1266 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **27** | **29** | **31** | **32** | **33** | **35** | **81** | **85** | **87** | **88** | **90** | **92** |
| **name** | | **H** | **K** | **S** | **N** | **Q** | **K** | **T** | **E** | **E** | **D** | **K** | **E** |
| **pattern** | **A** | -0.3 | 0.32 | -1.08 | 0.58 | 1.95 | -0.92 | -1.21 | 0.18 | 0.24 | 0.05 | 1.32 | -1.14 |
| | **B** | -0.36 | 0.66 | -1.07 | 0.36 | 1.88 | -0.72 | -1.14 | 0.04 | -0.03 | -0.24 | 1.65 | -1.01 |
| | **C** | -0.49 | 0.56 | -0.98 | 0.38 | 1.88 | -0.91 | -1.14 | 0.13 | 0.07 | -0.04 | 1.59 | -1.06 |
| | **D** | -0.22 | 0.01 | -1.07 | 0.75 | 2.24 | -0.89 | -1.2 | 0.31 | 0.38 | 0.17 | 0.65 | -1.13 |
| | **E** | -0.28 | 0.34 | -1.1 | 0.57 | 2.33 | -0.7 | -1.19 | 0.19 | 0.11 | -0.14 | 0.91 | -1.03 |
| | **F** | -0.42 | 0.19 | -0.98 | 0.6 | 2.33 | -0.89 | -1.17 | 0.3 | 0.24 | 0.11 | 0.76 | -1.07 |

J. Pose number 1454 (p1454)

| considered CD4 residues of p1454 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **25** | **27** | **29** | **31** | **32** | **33** | **35** | **85** | **87** | **88** | **90** | **92** |
| **name** | | **Q** | **H** | **K** | **S** | **N** | **Q** | **K** | **E** | **E** | **D** | **K** | **E** |
| **pattern** | **A** | -1.02 | -0.43 | 0.06 | -1.17 | 0.48 | 2.21 | -0.64 | -0.11 | 0.2 | -0.26 | 1.47 | -0.79 |
| | **B** | -0.98 | -0.41 | 0.46 | -1.21 | 0.3 | 2.2 | -0.58 | -0.23 | -0.15 | -0.53 | 1.55 | -0.44 |
| | **C** | -1.13 | -0.53 | 0.35 | -1.05 | 0.34 | 2.24 | -0.79 | -0.13 | 0 | -0.31 | 1.45 | -0.45 |
| | **D** | -1.05 | -0.33 | -0.11 | -1.22 | 0.76 | 2.1 | -0.58 | 0.05 | 0.43 | -0.13 | 1.21 | -1.14 |
| | **E** | -1.04 | -0.3 | 0.37 | -1.34 | 0.63 | 2.15 | -0.52 | -0.06 | 0.05 | -0.45 | 1.32 | -0.81 |
| | **F** | -1.23 | -0.44 | 0.2 | -1.13 | 0.7 | 2.18 | -0.78 | 0.09 | 0.26 | -0.16 | 1.14 | -0.85 |

**Table 11 (Continued)** The renormalized values of considered CD4 residues in six pattern of 11 complexes. The white in black box, white in gray box, and black in light gray box are 1st, 2nd, and 3rd key residue respectively.

K. Pose number 85 (p85)

| considered CD4 residues of p85 | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **position** | | **44** | **45** | **46** | **52** | **53** | **54** | **56** | **59** | **60** | **72** | **73** |
| **name** | | **L** | **T** | **K** | **N** | **D** | **R** | **D** | **R** | **S** | **K** | **N** |
| **pattern** | **A** | -0.86 | -0.77 | 0.87 | 0.48 | 0.66 | 0.03 | -1.36 | 1.85 | -1.3 | -0.12 | 0.52 |
| | **B** | -0.86 | -0.76 | 0.56 | 0.27 | 0.31 | 0.11 | -1.05 | 2.27 | -1.3 | -0.04 | 0.49 |
| | **C** | -0.81 | -0.7 | 0.63 | 0.46 | 0.8 | -0.34 | -1.23 | 1.99 | -1.25 | -0.19 | 0.63 |
| | **D** | -0.9 | -0.8 | 1.1 | 0.65 | 0.86 | 0.13 | -1.49 | 1.21 | -1.41 | -0.04 | 0.7 |
| | **E** | -0.97 | -0.84 | 0.83 | 0.46 | 0.51 | 0.26 | -1.2 | 1.67 | -1.53 | 0.07 | 0.74 |
| | **F** | -0.86 | -0.73 | 0.9 | 0.69 | 1.11 | -0.29 | -1.38 | 1.19 | -1.41 | -0.11 | 0.9 |

**Table 12** The top 3 key CD4 residues bound to DARPin 23.2

| pose | key CD4 residue | | |
|---|---|---|---|
| | 1st | 2nd | 3rd |
| 16 | Q40 | K35 | Q25 |
| 21 | Q40 | K35 | Q25 |
| 26 | K35 | Q40 | Q33 |
| 241 | Q40 | K35 | Q25 |
| 642 | K35 | Q40 | Q25 |
| 1302 | K35 | Q40 | Q33 |
| 1513 | K35 | Q40 | Q33 |
| 1128 | Q33 | K90 | N32 |
| 1266 | Q33 | K90 | N32 |
| 1454 | Q33 | K90 | N32 |
| 85 | R59 | D53 | K46 |

To increase the reliable of these key binding residues, they were analyzed with the bio-information of CD4-gp120 and CD4-MHC complex solving by crystallization and biochemical mutagenesis. All 1st-3rd key CD4 residues in 11 poses were located on the binding site of gp120 specific CD4 within a 5.0 Å radius, except for Lys90 and Asp53 (refer to Table 7 and 12). According to the key amino acids, Lys35 and Arg59 were found to be parts of the critical residues on CD4 that gp120 recognition, which are Lys29, Lys35, Phe43, Leu44, Lys46, Gly47 and Arg59; these are studied by biochemical mutagenesis (Phe43 and Gly47) [32] and compiled from mutagenesis studies by Ryu et.al [33]. Focusing on interface on CD4 to MHCII at 5 Å, Lys35, Gln40, and Lys46 were part on this area (refer to Table 7 and 12). Whereas critical CD4 residues to MHCII suggested by Moebius et al. [27], Lys35, Phe43, Lys46, and Arg59, had overlapping residues with our prediction only 2 residues (Lys35 and Arg59). Altogether, Lys35 was the most important hot spots and Gln33, Gln40, as well as Arg59 were second order of hot spots. Specific mutation experiments to

confirm the theoretical investigation of key CD4 amino acids will be undertaken and observed for the interactions between CD4 and DARPin.

## Validating the predicted key residues

The 1st- 3rd key CD4 residues, in case of validation called hot spots, consisted of Gln25, Asn32, Gln33, Lys35, Gln40, Lys46, Asp53, and Arg59, were validated with physical properties and other programs. Interestingly, all three groups, the physiology of these hot spots were polar group which Gln and Asn were polar uncharged; Lys and Arg were basic polar; Asp was acidic polar. From several studies [63-69] indicated that hydrophobic residues play a dominant role in the protein-protein interface, however, hydrophilic residues are preferred in interface area [63-67]. Therefore, our algorithm for predicting hot spots should not be immediately rejected.

Considering the order of hot spot propensity, it was analyzed by Bogan & Thorn [58] showed as Trp > Arg > Tyr > Ile > Asp > His > Pro > Lys > Asn > Glu > Gly > Gln > Phe > Met > Thr > Ser > Leu > Cys = Val (refer to table 13). Here, we clustered this propensity into four classes: high, moderate, low, and rare propensity. According suggesting of Bogan & Thron, the high and rare propensities were defined. The high propensity composed of Trp, Arg, and Tyr, whereas, Met, Thr, Ser, Leu, Cys, and Val were rare propensity. Since the range of frequency excepting high and rare propensity were 3.01%-9.62%, the cut-off diving data into two classes was average of these data, 6.10%. Therefore, the moderate propensity which percentage

of hot spot had frequency value more 6.10% consisted of Ile, Asp, His, Pro, and Lys. The low propensity composed of Gln, Asp, Gly, Gln, and Phe.

**Table 13** The hot spot propensity analyzed by Bogan & Thorn was divided into 4 classes. Our hot spots were predicted at high, moderate, and low propensity.

| residue | in database | | contribute ≥ 2 kcal/mol | | class of propen-sity | our hot spot |
|---|---|---|---|---|---|---|
| | (number) | (percent) | (number) | (percent) | | |
| Trp | 19 | 0.82 | 4 | 21.05 | high | Arg59 |
| Arg | 218 | 9.38 | 29 | 13.3 | | |
| Tyr | 122 | 5.25 | 15 | 12.3 | | |
| Ile | 104 | 4.47 | 10 | 9.62 | moderate | Lys35 Lys46 Asp53 |
| Asp | 177 | 7.61 | 16 | 9.04 | | |
| His | 50 | 2.15 | 4 | 8 | | |
| Pro | 89 | 3.83 | 6 | 6.74 | | |
| Lys | 143 | 6.15 | 9 | 6.29 | | |
| Asn | 99 | 4.26 | 5 | 5.05 | low | Gln25 Gln33 Gln40 Asn32 |
| Glu | 220 | 9.46 | 8 | 3.64 | | |
| Gly | 28 | 1.2 | 1 | 3.57 | | |
| Gln | 160 | 6.88 | 5 | 3.13 | | |
| Phe | 166 | 7.14 | 5 | 3.01 | | |
| Met | 69 | 2.97 | 2 | 2.9 | rare | - |
| Thr | 131 | 5.63 | 2 | 1.53 | | |
| Ser | 178 | 7.66 | 2 | 1.12 | | |
| Leu | 242 | 10.41 | 2 | 0.83 | | |
| Cys | 3 | 0.13 | 0 | 0 | | |
| Val | 107 | 4.6 | 0 | 0 | | |

In group 1 of docked poses, Lys and Gln, which were predicted to be hot spot, were estimated at moderate and rare propensity as same as hot spots in group 2 (Lys, Gln, and Asn). Although our prediction suggested the hot spots at low (Gln and Asn) to moderate (Lys and Asp) propensity, these amino acids were hydrophilic residues which were able to make hydrogen bond to one another. Moreover, considered

CD4's residues (refer to Table 10) in group 1 and 2 did not have residue in high propensity; it was no way to predict hot spot at high propensity. Considering group 3, one out of hot spots was Arg which inhabited at high propensity. Probably, our procedure could predict hot spot at high propensity when residues at high propensity appeared in considered CD4's residues. Altogether our algorithm was not too bad because most of predicted hot spot could be found in the interface which was moderate class of hot spot propensity as shown in Table 13. Moreover, these residues which were polar residues were not bad result to be hot spot, at least these residues helped to stability the protein-protein interaction.

All 11 complexes were performed to find hot spots by HotPOINT and HSPred web server. The HotPOINT bases on solvent accessibility and pair potentials of residues, whereas the HSPred depends on energy calculation. The results showed that there are 4-8 hot spots, average at 6.8, (Table 14 and 24) and 2-6 residues, average at 4.4, (Table 15 and 25) predicted by HotPOINT and HSPred respectively. Notably, in both servers, the hot spots were somewhat identified to be hydrophilic residues more than hydrophobic residues as shown in Table 14-16. So, the hot spots of CD4 bound to DARPin 23.2 were possibly identified as hydrophilic residues. Since the number of hot spots predicted from three different methods was different, validating hot spots held algorithm having smallest number of hot spot. So the identity residues between our algorithm and two programs were performed with the maximum length residues of our prediction, which value was 3. The compared results showed in percentage of identity prediction ($PID^{predict}$) as shown in last column of Table 14-16. The $PID^{predict}$ showed the percent identity residue between our prediction and other software. Note that the red hot spots in these tables were our predicted hot spots.

Among all 11 complexes, at least one residue in each pose was identified agreeable to two programs. The average $PID^{predict}$ in both HotPOINT and HSPred was high value (69.7%). Especially in group 1, there were four out of seven poses that our prediction accorded 100% with two programs and the $PID^{predict}$ had high value (85.71% for HotPOINT, 76.19% for HSPred). Interestingly, Lys35 and Gln33 were found in both server in all poses of group 1 and 2 respectively. So, this result supported the reliable of Lys35 and Gln33. Unlike Arg59 of group 3, it did not be predicted from both programs. Therefore, our algorithm was reliable for predicting hot spot. Although our algorithm did not have the threshold to distinguish the hot spot and non-hot spot, our hot spot prediction was part of server hot spot prediction. Possibly, the loss hot spot from our algorithm could identify by increasing the number of key residue. The number of hot spots intersecting between HotPOINT and HSPred estimated three residues, therefore, 3 residues for hot spot maybe were appropriate with our algorithm.

**Table 14** The hot spots predicted by HotPOINT. The red residues are the same hot spots as our prediction.

| group | pose | positions of hot spot | | | | | | | | R | $PID^{predict}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 16 | I24 | Q25 | H27 | K35 | N39 | Q40 | T45 | E85 | 1 : 7 | 100 |
| | 21 | I24 | Q25 | H27 | K35 | N39 | Q40 | E85 | | 1 : 6 | 100 |
| | 26 | Q25 | H27 | K29 | K35 | Q40 | E85 | | | 0 : 6 | 66.67 |
| | 241 | I24 | Q25 | H27 | I34 | K35 | Q40 | T45 | E85 | 2 : 6 | 100 |
| | 642 | I24 | Q25 | H27 | K35 | N39 | Q40 | E85 | E87 | 1 : 7 | 100 |
| | 1302 | Q25 | H27 | K29 | I34 | K35 | Q40 | E85 | | 1 : 6 | 66.67 |
| | 1513 | Q25 | I34 | K35 | N39 | Q40 | E85 | | | 1 : 5 | 66.67 |
| 2 | 1128 | F26 | K29 | Q33 | E85 | | | | | 1 : 3 | 33.33 |
| | 1266 | F26 | K29 | Q33 | I83 | E85 | E92 | | | 2 : 4 | 33.33 |
| | 1454 | F26 | K29 | S31 | Q33 | I83 | E85 | E92 | | 1 : 6 | 33.33 |
| 3 | 85 | L44 | K46 | N52 | D53 | R54 | D56 | S57 | I70 | 2 : 6 | 66.67 |

R is ratio of hydrophobic to hydrophilic residue

**Table 15** The hot spots predicted by HSPred. The red residues are the same hot spots as our prediction.

| group | pose | positions of hot spot | | | | | | R | $PID^{predict}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 16 | Q25 | F26 | H27 | K35 | Q40 | | 1 : 4 | 100 |
| | 21 | Q25 | F26 | H27 | K29 | K35 | Q40 | 1 : 5 | 100 |
| | 26 | H27 | K29 | K35 | Q40 | | | 0 : 4 | 66.67 |
| | 241 | Q25 | H27 | K35 | Q40 | | | 0 : 4 | 100 |
| | 642 | Q25 | H27 | K29 | K35 | Q40 | | 0 : 5 | 100 |
| | 1302 | H27 | K29 | K35 | | | | 0 : 3 | 33.33 |
| | 1513 | H27 | K29 | K35 | E85 | | | 0 : 4 | 33.33 |
| 2 | 1128 | H27 | K29 | Q33 | I83 | E85 | K90 | 1 : 5 | 66.67 |
| | 1266 | K29 | Q33 | K90 | | | | 0 : 3 | 66.67 |
| | 1454 | H27 | K29 | Q33 | I83 | E85 | K90 | 1 : 5 | 66.67 |
| 3 | 85 | D53 | R54 | | | | | 0 : 2 | 33.33 |

R is ratio of hydrophobic to hydrophilic residue

**Table 16** The intersection of hot spots between HotPOINT and HSPred. The red residues are the same hot spots as our prediction.

| group | pose | positions of hot spot | R | $PID^{predict}$ |
|---|---|---|---|---|
| 1 | 16 | Q25 H27 K35 Q40 | 0 : 4 | 100 |
| | 21 | Q25 H27 K35 Q40 | 0 : 4 | 100 |
| | 26 | H27 K29 K35 Q40 | 0 : 4 | 66.67 |
| | 241 | Q25 H27 K35 Q40 | 0 : 4 | 100 |
| | 642 | Q25 H27 K35 Q40 | 0 : 4 | 100 |
| | 1302 | H27 K29 K35 | 0 : 3 | 33.33 |
| | 1513 | K35 E85 | 0 : 2 | 33.33 |
| 2 | 1128 | K29 Q33 I83 E85 | 1 : 3 | 33.33 |
| | 1266 | K29 Q33 | 0 : 2 | 33.33 |
| | 1454 | K29 Q33 I83 E85 | 1 : 3 | 33.33 |
| 3 | 85 | D53 R54 | 0 : 2 | 33.33 |

R is ratio of hydrophobic to hydrophilic residue