Sidechain prediction validation set 1: 4BVG

Residues included in the consideration (for sidechain prediction and Prime minimization) are 144-180.195,199,204,207,210,227-234,248,251,291,294,324 plus intermediate in minimization.

Two starting structures were used.

1. One used the 4BVG prepared using Protein Preparation Wizard and OPLS minimization on hydrogen only (without relaxation of heavy atoms).

|  |  |
| --- | --- |
| **Structures** | **Prime Energy** |
| Struct 1: 4BVG as prepared w/ h-opt | -9614.9 |
| Struct 2: Struct 1 w/ sidechain opt w/ backbone within 1 residue | -9871.5 |
| Struct 3: Struct 1 w/ sidechain opt w/ default option | -9896.1 |
| Struct 4: Struct 1 w/ sidechain opt w/ Monte Carlo, rank2 | -10174.6 |
| Struct 5: Struct 1 w/ sidechain opt w/ Monte Carlo, rank1 | -10175.5 |
| Struct 6: Struct 1 w/ sidechain opt w/ CA-CB vector sampling | -10179.1 |
| Struct1 with prime minimization on selected residues | -10301.2 |
| Struct2 with prime minimization on selected residues | -10316.1 |
| Struct3 with prime minimization on selected residues | -10316.4 |
| Struct4 with prime minimization on selected residues | -10321.0 |
| Struct5 with prime minimization on selected residues | -10321.1 |
| The above structure with prime re-minimization on selected residues | -10320.9 |
| Struct6 with prime minimization on selected residues | -10320.7 |
| Struct 7: Struct 1 w/ sidechain opt w/ backbone within 3 residue | -10268.2 |
| Struct7 with prime minimization on selected residues | -10463.6 |
| The above structure with prime re-minimization on selected residues | -10467.1 |
| The above structure with prime re-minimization on selected residues | -10467.5 |

There are four options available in Prime Sidechain Prediction: Default, Monte Carlo, w/ CA-CB vector sampling, and w/ backbone sampling (by default with 3 residues each time, leading to extra residues included in optimization.) Reducing backbone sampling to 1 actually remove the backbone sampling, and has to be manually edited in the input file to carry out the calculation.

The results also show extra steps of minimization in testing the convergence.

The RMSDs for each residue are included in EXCEL file 4BVG\_h\_minimized\_as\_Starting\_for\_sidechain\_prediction\_1.xlsx.

Some convergence on structures can be identified. Two minimums identified at around -10316 and -10321 with only the selected residues included. And the default sidechain prediction didn’t locate the better minimum. Residues with significant change from crystal structure were highlight in EXCEL file.

1. The structure prepared using Protein Preparation Wizard and OPLS minimization on hydrogen on 4BVG were further minimized using Prime and serve as a starting structure for further sidechain prediction investigation.

|  |  |
| --- | --- |
| **Structures** | **Prime Energy** |
| Struct 1a: 4BVG as prepared w/ h-opt followed by Prime minimization of selected residues | -10301.2 |
| Struct 2a: Struct 1a w/ sidechain opt w/ default option | -10261.6 |
| Struct 3a: Struct 1a w/ sidechain opt w/ CA-CB vector sampling | -10301.9 |
| Struct 4a: Struct 1a w/ sidechain opt w/ backbone within 1 residue | -10304.7 |
| Struct 5a: Struct 1a w/ sidechain opt w/ Monte Carlo, rank1 | -10322.7 |
| Struct 6a: Struct 1a w/ loop refine with ultra extended sampling (res. 156-169) rank 1 | -10334.8 |
| Struct2a with prime minimization on selected residues | -10289.7 |
| Struct3a with prime minimization on selected residues | -10304.3 |
| Struct4a with prime minimization on selected residues | -10320.4 |
| Struct5a with prime minimization on selected residues | -10324.8 |
| Struct6a with prime minimization on selected residues | -10374.4 |
| Struct 7a: Struct 1a w/ sidechain opt w/ backbone within 3 residue | -10464.5 |
| Struct7a with prime minimization on selected residues | -10470.2 |

The results above show that sampling error does exist as least for the sidechain prediction with default option, as it located a structure with higher energy. And various sidechain predictions also gave different minimums. The fact that loop refinement on only a subset of selected residues (but it includes extra residues within 7.5 A in optimization) reduced structure significantly suggest there may be potential energy errors as well.

The above two sidechain prediction run results point to a limited set of residues that contribute to the change of energy in sidechain prediction (as seem from RMSD data), which can be used in the future in making a small set of residues for optimization.

We carried out something similar for 4FVT using the carba-NAD+ as in the crystal structure, and almost the same choice of residues except carbaNAD+ and ac-LYS in the place of intermediate in 4BVG.

|  |  |
| --- | --- |
| **Structures** | **Prime Energy** |
| Struct 1: 4FVT prepared w/ Protein Preparation Wizard w/ h-opt | -10055.6 |
| Struct 2: Struct 1 w/ sidechain opt w/ backbone sampling of 1 residue | -10339.6 |
| Struct 3: Struct 1 w/ sidechain opt w/ default option | -10342.2 |
| Struct 4: Struct 1 w/ sidechain opt w/ CA-CB vector sampling | -10601.9 |
| Struct 5: Struct 1 w/ sidechain opt w/ Monte Carlo, rank1 | -10605.9 |
| Struct 6: Struct 1 w/ sidechain opt w/ Monte Carlo, rank2 | -10605.1 |
| Struct 7: Struct 1 w/ loop refine with ultra extended sampling (res. 156-169) rank 1 | -10382.2 |
| Struct1 with prime minimization on selected residues | -10667.1 |
| Struct2 with prime minimization on selected residues | -10714.7 |
| Struct3 with prime minimization on selected residues | -10715.9 |
| The above structure with prime re-minimization on selected residues | -10716.2 |
| Struct4 with prime minimization on selected residues | -10725.0 |
| Struct5 with prime minimization on selected residues | -10728.2 |
| Struct6 with prime minimization on selected residues | -10726.2 |
| Struct7 with prime minimization on selected residues | -10742.8 |
| Struct 8: Struct 1 w/ sidechain opt w/ backbone sampling of 3 residue | -10690.8 |
| Struct8 with prime minimization on selected residues | -10894.0 |

Better sidechain refinement was found for Monte Carlo method when backbone flexibility was included in the refinement. Among the four options in the sidechain prediction, the predicted energy quite often correlated with the degrees of freedom considered, which is increasing from default option, to w/ CA-CB vector sampling, to w/ Monte Carlo method, to w/ backbone sampling of 3 residues. The following Prime minimizations do not change the rank order most of the time.

The loop refinement of residue 156-169 identified loop structures dramatically different with lower energies, suggesting energy error in this case.

The corresponding RMSDs are included in EXCEL file 4FVT\_carbaNAD\_h-opt\_as\_start\_for\_sidechain\_prediction\_1.xlsx.

We also run Prime minimization of the selected residues and used it as the starting point for sidechain prediction and other structural refinement.

|  |  |
| --- | --- |
| **Structures** | **Prime Energy** |
| Struct 1a: 4FVT as prepared w/ h-opt followed by Prime minimization of selected residues | -10667.1 |
| Struct 2a: Struct 1a w/ sidechain opt w/ default option | -10680.9 |
| Struct 3a: Struct 1a w/ sidechain opt w/ backbone within 1 residue | -10690.7 |
| Struct 4a: Struct 1a w/ sidechain opt w/ Monte Carlo, rank1 | -10716.3 |
| Struct 5a: Struct 1a w/ sidechain opt w/ Monte Carlo, rank2 | -10715.2 |
| Struct 6a: Struct 1a w/ sidechain opt w/ Monte Carlo, rank3 | -10714.5 |
| Struct 7a: Struct 1a w/ sidechain opt w/ CA-CB vector sampling | -10735.6 |
| Struct 8a: Struct 1a w/ sidechain opt w/ backbone within 3 residue | -10855.8 |
| Struct 9a: Struct 1a w/ loop refine with ultra extended sampling (res. 156-169) rank 1 | -10736.6 |
| Struct 10a: Struct 1a w/ loop refine with ultra extended sampling (res. 156-169) rank 2 | -10736.3 |
| Struct2a with prime minimization on selected residues | -10724.3 |
| Struct3a with prime minimization on selected residues | -10726.5 |
| Struct4a with prime minimization on selected residues | -10720.9 |
| Struct5a with prime minimization on selected residues | -10717.8 |
| Struct6a with prime minimization on selected residues | -10717.9 |
| Struct7a with prime minimization on selected residues | -10741.0 |
| Struct8a with prime minimization on selected residues | -10888.8 |
| Struct9a with prime minimization on selected residues | -10764.0 |
| Struct10a with prime minimization on selected residues | -10762.1 |

With the new starting structure, the Monte Carlo method isn’t the better method vs w/ CA-CB vector sampling we saw when we started with 4FVT w/ h-opt preparation. The loop refinement of residue 156-169 again identified lower energy structures with dramatically different loop conformations.