# Table of training set

I made a training set with 9 high resolution files (resolution < 1.2 Å). The high resolution files that we got from Protein Date Bank website by these standards:

1. The resolution is as high as possible.
2. The file should have at least two different helices.

I checked all helix records of these files and picked helices by our standards. Difference standards will get difference helices. Our standards are:

1. The length of the helix is longer than 6 residues.
2. The helix should consistent with protein theory( like Ramachandran plot).
3. The helix should look like “a real helix” in protein crystallography software (like PyMol).
4. Pay attention to the helices which include Proline. Because Proline will influence helix structure.

I checked all helix-records by PyMol cartoon, Ramachandran plot and DSS method that is a method from PyMol to define secondary structure based on backbone geometry. The red modified records mean doubt.

Finally, we got this table

|  |  |  |  |
| --- | --- | --- | --- |
| file name | original records | modified records | notes |
| 2vb1  (0.65 Å) | 4-15 | 5-14 |  |
| 24-37 | 26-35 |  |
| 80-85 | delete | poor helix |
| 88-101 | 89-100 |  |
| 104-107 | delete | poor helix |
| 108-115 | delete | poor helix |
| 119-124 | delete | poor helix |
| 2WFI  (0.75 Å) | 33-46 | 33-46 | PRO:34 |
| 131-135 | delete | Length < 6 CA |
| 147-156 | 147-156 | poor helix |
| 4REK  (0.74 Å) | 19-32 | 20-33 | poor helix |
| 62-66 | delete | Length < 6 CA |
| 78-83 | delete | poor helix |
| 84-86 | delete | Length < 6 CA |
| 113-118 | delete | poor helix |
| 127-135 | 127-135 | PRO: 136 |
| 139-146 | delete | poor helix |
| 146-155 | 146-155 | PRO: 149 |
| 161-168 | 161-168 |  |
| 171-184 | 171-183 |  |
| 196-205 | 196-205 |  |
| 211-215 | delete | Length < 6 CA |
| 231-239 | 231-239 |  |
| 289-304 | 292-304 |  |
| 402-405 | delete | Length < 6 CA |
| 406-425 | 406-424 |  |
| 474-478 | delete | Length < 6 CA |
| 486-506 | 486-503 |  |
| 1us0  (0.66 Å) | 23-38 | 26-37 | PRO:23 |
| 45-49 | delete | Length < 6 CA |
| 50-64 | 51-64 |  |
| 68-72 | delete | Length < 6 CA |
| 79-83 | delete | Length < 6 CA |
| 87-100 | 87-100 |  |
| 136-150 | 137-150 |  |
| 162-171 | 162-171 |  |
| 192-202 | 192-202 |  |
| 230-241 | 230-241 |  |
| 243-255 | 243-252 | PRO:252 |
| 265-274 | 265-274 |  |
| 281-290 | 283-290 |  |
| 300-304 | delete | Length < 6 CA |
| 5l87  (0.87Å) | 5-20 | 6-20 |  |
| 27-38 | 28-38 |  |
| 41-52 | 42-52 |  |
| 57-66 | 58-65 |  |
| 5gv8  （0.78 Å） | 1096-1105 | 1097-1103 |  |
| 1154-1167 | 1156-1167 |  |
| 1183-1187 | delete | Length < 6 CA |
| 1189-1201 | 1191-1199 |  |
| 1225-1233 | delete | poor helix |
| 1235-1239 | delete | Length < 6 CA |
| 1247-1254 | delete | poor helix |
| 1255-1263 | 1255-1262 |  |
| 1265-1269 | delete | Length < 6 CA |
| 5r4h  (1.18 Å) | 2794-2800 | delete | poor helix |
| 2803-2819 | 2804-2819 |  |
| 2823-2827 | delete | Length < 6 CA |
| 2837-2842 | delete |  |
| 2847-2857 | delete | poor helix |
| 2862-2881 | 2863-2880 |  |
| 2885-2911 | 2891-2910 |  |
| 5yce  (0.77 Å) | 3-18 | 3-17 |  |
| 20-36 | 22-36 |  |
| 36-42 | delete | poor helix |
| 51-57 | delete | poor helix |
| 58-77 | 59-77 |  |
| 82-96 | 86-96 |  |
| 100-119 | 103-119 |  |
| 124-149 | 126-146 |  |
| 5dj7  （0.87 Å） | 8-12 | delete |  |
| 40-45 | delete |  |
| 48-51 | delete |  |
| 82-86 | delete |  |
| 110-114 | delete |  |
| 146-154 | 146-154 |  |
| 214-223 | 214-222 |  |