Results assignment 2:

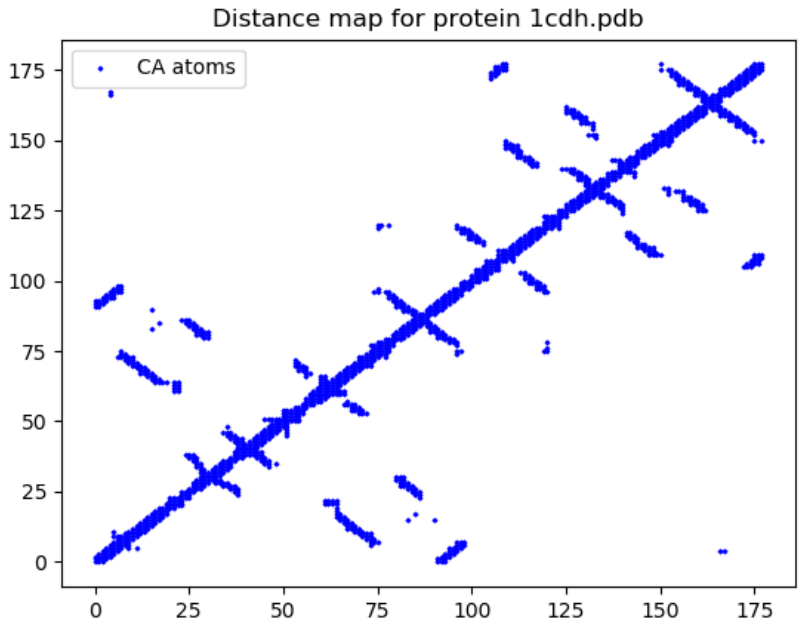
Emma Rydholm

All tests are run with threshold = 7

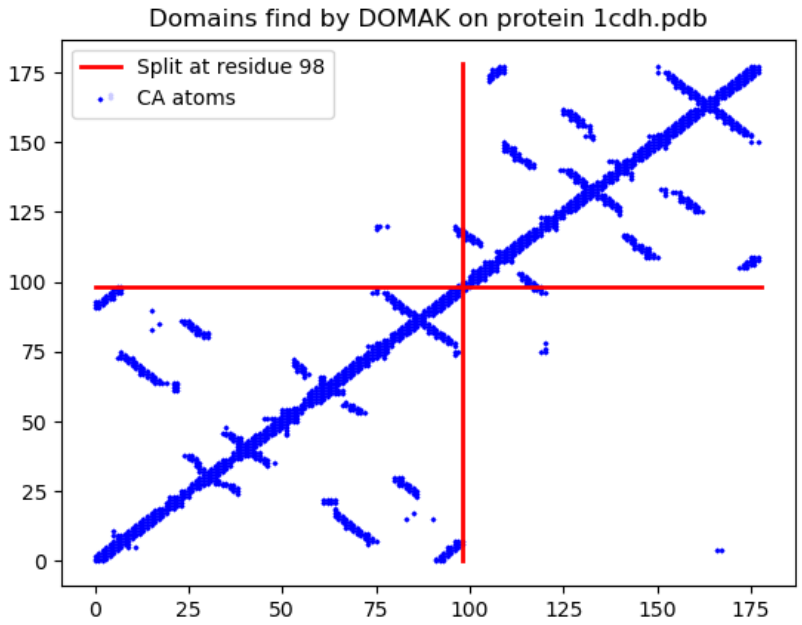
To run the script just run: **python3 find_domains.py**

Protein = 1cdh.pdb

Distance map:



The best partition index using DOMAK is: [98]



The best partition index using STRUDL is:

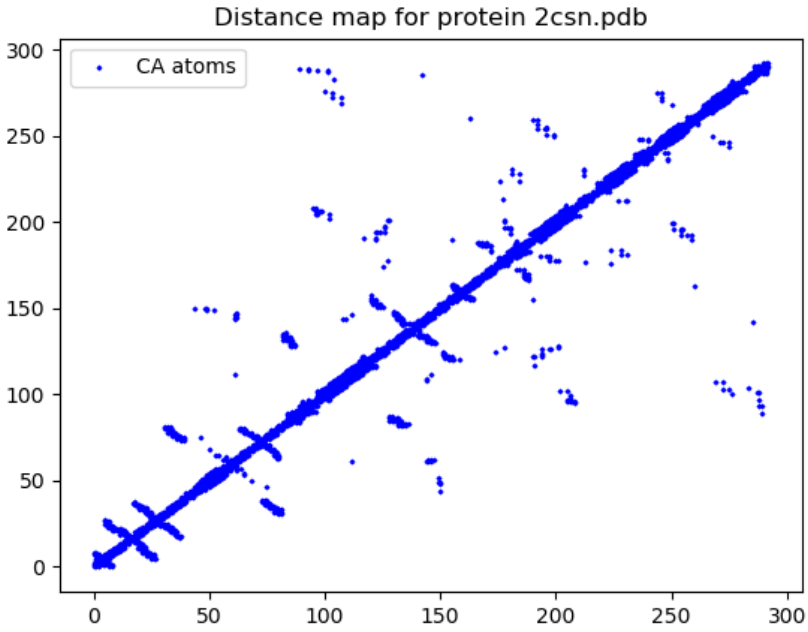
U = [0 1 2 3 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
65 66 67 68 69 70 71 72 79 80 81 82 83 84 85 86 87 88 89 90
91 92 93 121 122 123 124 163 164 165]

and

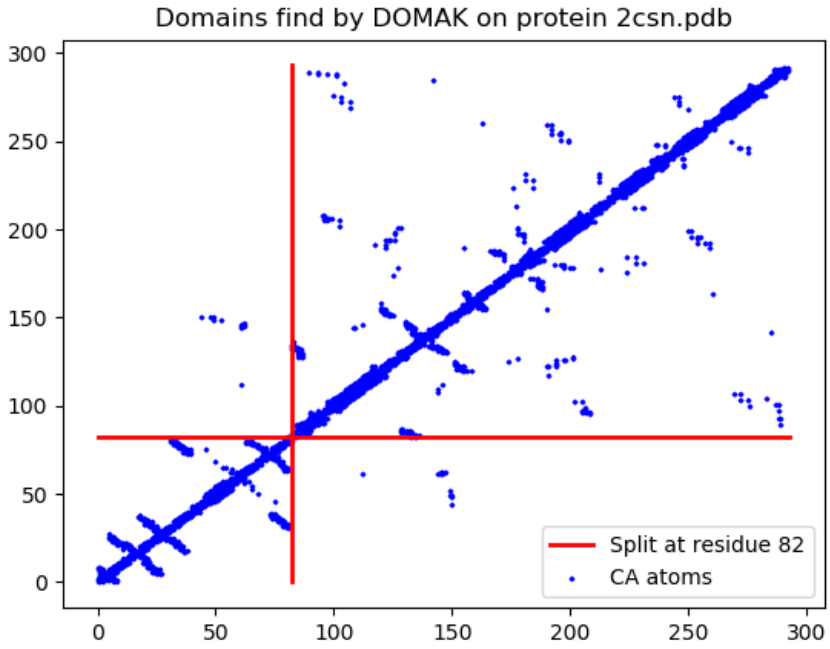
V = [4 5 6 7 8 9 73 74 75 76 77 78 94 95 96 97 98 99 100 101
102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
117 118 119 120 125 126 127 128 129 130 131 132 133 134 135
136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
151 152 153 154 155 156 157 158 159 160 161 162 166 167 168
169 170 171 172 173 174 175 176 177]

The difference from using STRUDL and DOMAK is that DOMAK results in 2 domains that where all residues are aligned, compared to STRUDL that can have gaps in the domain. Even though there are some gaps in the domains the breaking point seem to be around residue 94, which is close to what was found using DOMAK.

Protein = 2csn.pdb
Distance map:



The best partition index using DOMAK is: [82]



The best partition index using STRUDL is:

U = [0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 139 148
149 150 156 157 158 159 160 161 162 163 164 165 166 167 168
169 170 171 172 173 174 175 176 177 186 209 210 211 212 213
214 215 216 217 218 219 220 258 259 260 261 262 263 264 265
266 277 278 279 280 281 282 283 284 285 286 287 288 289 290
291 292]

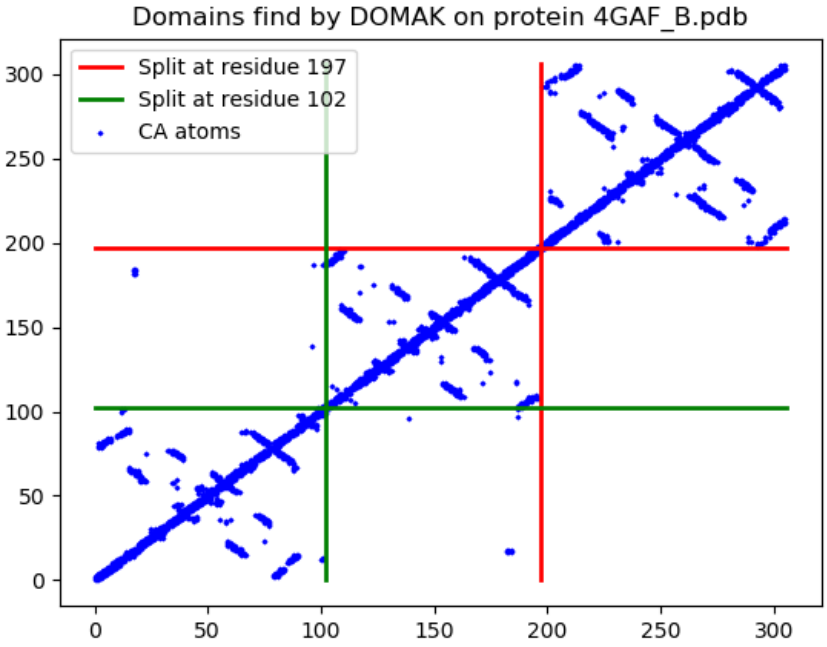
and

V = [82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115
116 117 118 119 120 121 122 123 124 125 126 127 128 129 130
131 132 133 134 135 136 137 138 140 141 142 143 144 145 146
147 151 152 153 154 155 178 179 180 181 182 183 184 185 187
188 189 190 191 192 193 194 195 196 197 198 199 200 201 202
203 204 205 206 207 208 221 222 223 224 225 226 227 228 229
230 231 232 233 234 235 236 237 238 239 240 241 242 243 244
245 246 247 248 249 250 251 252 253 254 255 256 257 267 268
269 270 271 272 273 274 275 276]

The difference from using STRUDL and DOMAK is that DOMAK results in 2 domains that where all residues are aligned, compared to STRUDL that can have gaps in the domain. Even though there are some gaps in the domains the breaking point seem to be around residue 82, which is close to what was found using DOMAK. There seems to be a larger gap for the last residues and more overlaps between V and U for the residues towards the end in general. This could potentially depend on that there are more residues with connections towards the end than the beginning.

Protein = 4GAF_B.pdb

The best partition index using DOMAK is: [197, 102]
and there is 3 domains



Protein = 1HZH_H.pdb

The best partition index using DOMAK is: [236, 127]
and there is 3 domains