

Benchmarking and Extension of Protein3DFit

Nasir Mahmood

Benchmarking and Extension of Protein3DFit

- o **Introduction**

- o Why Align Structures?
- o Structure Alignment: Issues & problem

- o **Method**

- o Algorithm
- o Pitfalls
- o Extension

- o **Results**

- o **Outlook**

Why Align Structures?

Fundamental step in:

- **Analysis & Visualization**
- **Comparison**
- **Design**

- **Useful for:**
 - **Structure classification**
 - To predict protein folds.
 - Gold standard for sequence alignment.
 - **Structure prediction:**
 - Identifying "structural core".
 - **Function prediction**
 - Function of protein is determined by its structure.
 - **Drug discovery**
 - To pinpoint active sites accurately.

Structure Alignment: Issues

Theoretical Issues

- NP-hard geometric problem
- Heuristics needed
- No unique solution

Methodological Issues

Choices:

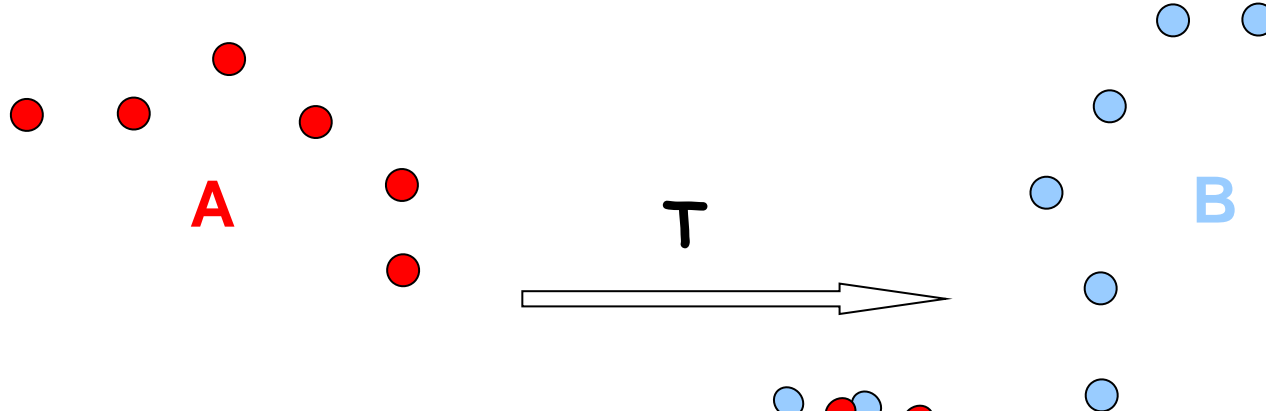
- Structure Representation
- Scoring function
- Search algorithm

Goals Desirable

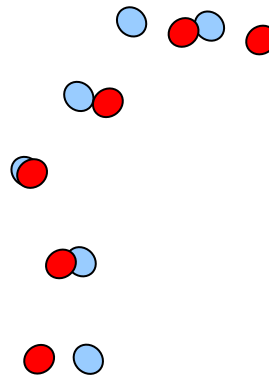
- Automatic
- Discriminating
- Fast

Structure alignment Problem

Simple case - two closely related proteins with the same number of amino acids.



Find a transformation to achieve the best superposition



Structure Alignment Problem

Two sets of Points in 3D space:

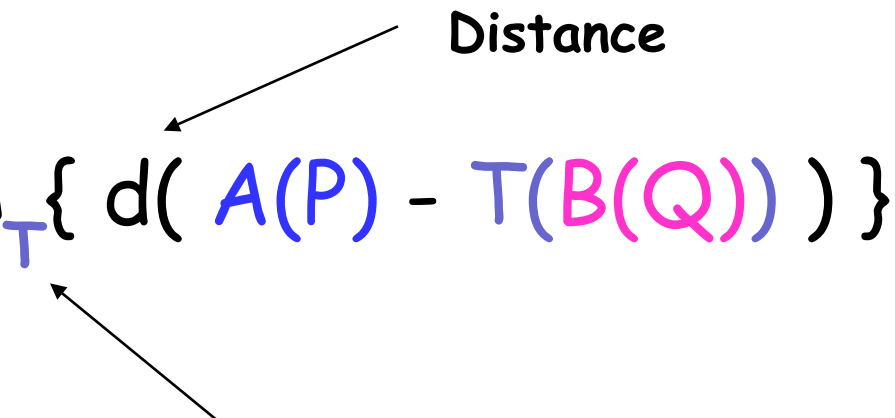
$$A = (a_1, a_2, \dots, a_n)$$

$$B = (b_1, b_2, \dots, b_m)$$

$A(P)$, $B(Q)$: Optimal subsets with $|A(P)| = |B(Q)|$,

$$\min_T \{ d(A(P) - T(B(Q))) \}$$

Distance

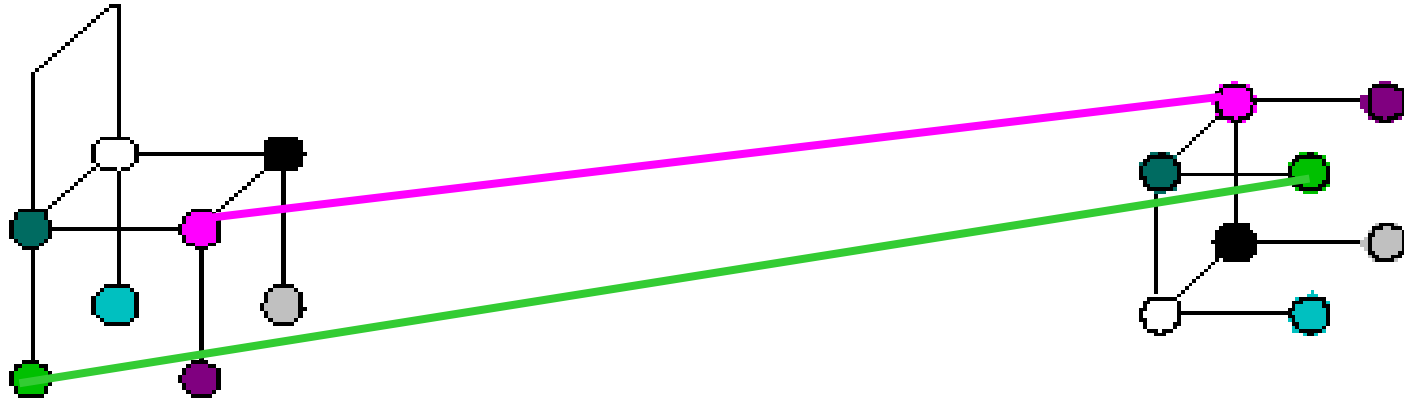


Optimal rigid body transformation
(calculated by Diamond (1998) method)

Two Subproblems

1. Find correspondence set
2. Find optimal transformation
(protein superposition problem)

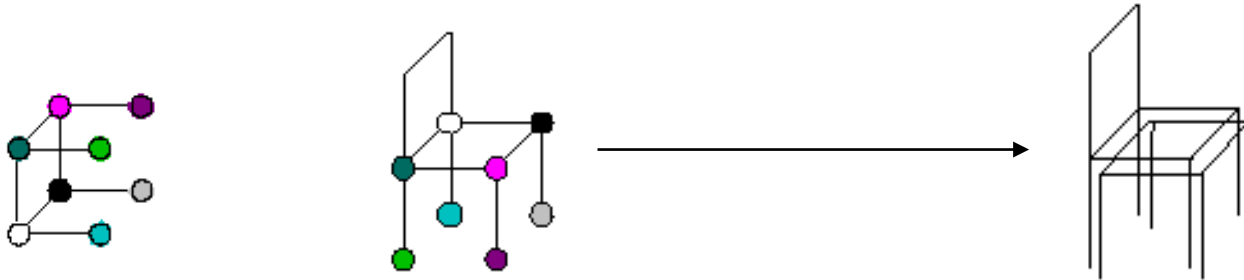
Correspondence set



- Correspondence is a key problem.
- Exhaustive search intractable.
- Heuristics are applied.

Superposition

- Two sets of points
 $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$



- Correspondence pairs (a_i, b_i)
- Find T

$$\min_T \{ \text{RMSD}(A, T(B)) \}$$

Aim of the project

- To make algorithm more efficient
 - debugging
 - optimization
 - extension
 - improve quality of alignments
 - organization of code

Benchmarking and Extension of Protein3DFit

- o **Introduction**

- o Why Align Structures?

- o Structure Alignment: Issues & problem

- o **Method**

- o Algorithm

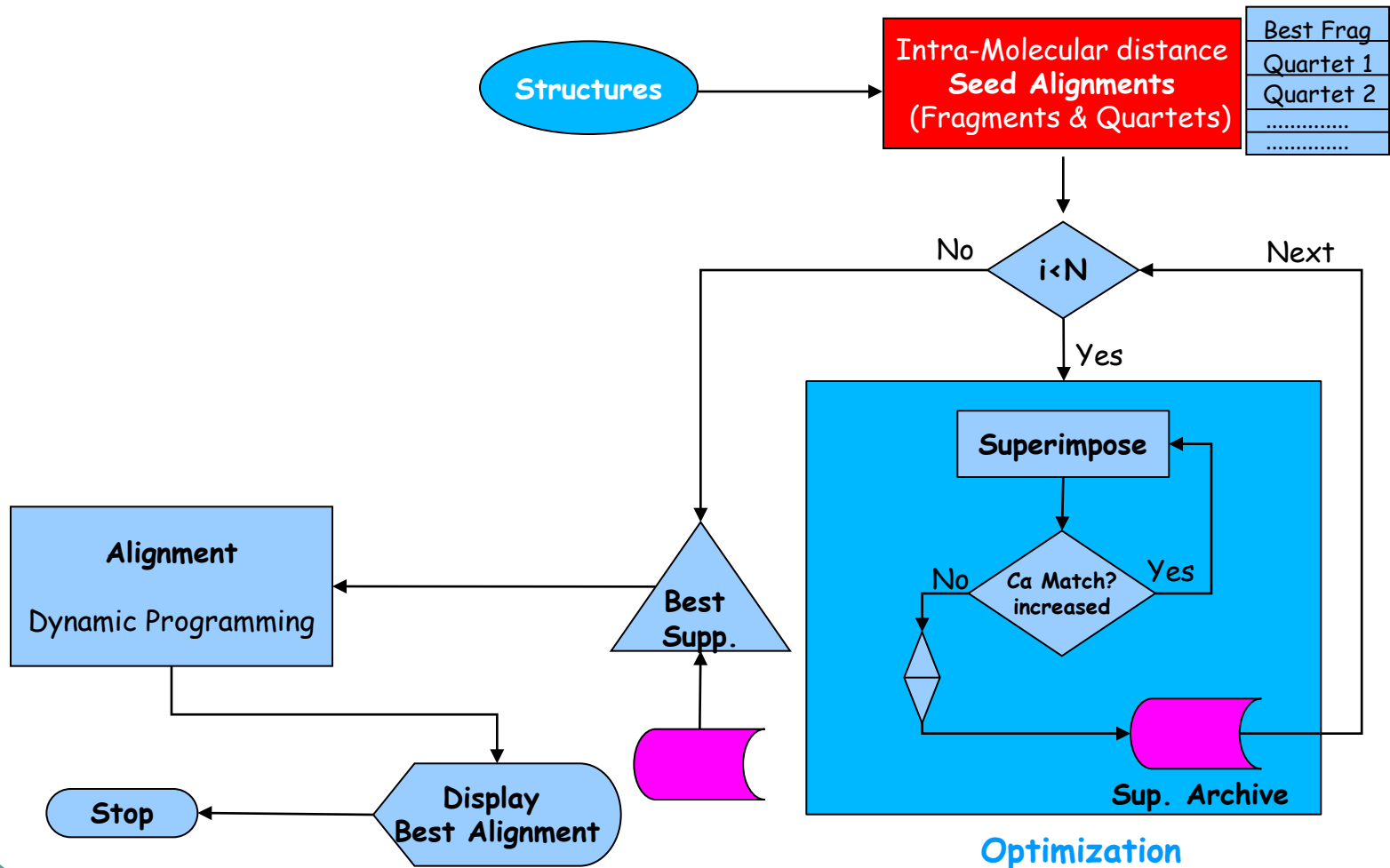
- o Pitfalls

- o Extension

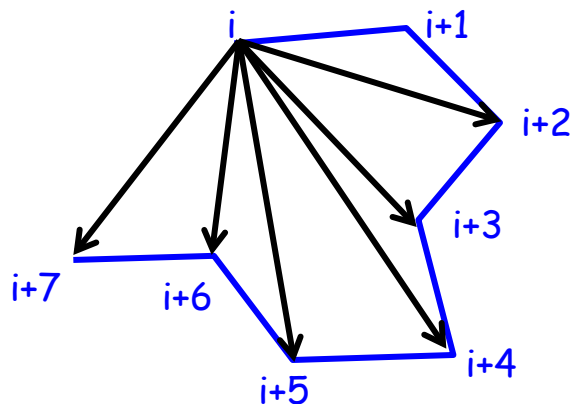
- o **Results**

- o **Outlook**

Algorithm

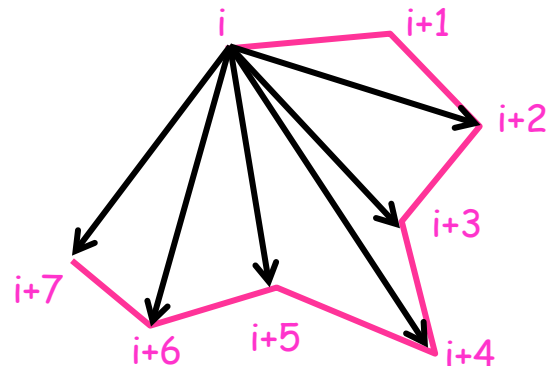


Intra-molecular Distance Matrices



| | 1 | 2 | 3 | 4 | 5 | | N |
|---|-----|-----|-----|-----|-----|-------|---|
| 1 | 0.0 | 1.4 | 1.4 | 0.0 | | | |
| 2 | 1.4 | 0.0 | 1.4 | 1.4 | 0.0 | | |
| 3 | 3.0 | 2.1 | 0.0 | 3.0 | 2.1 | 0.0 | |
| 4 | 0.7 | 0.3 | 0.0 | 0.0 | 0.3 | 0.0 | |
| 5 | 1.9 | 0.0 | 0.4 | 1.9 | 0.0 | 0.4 | |
| · | 0.0 | 3.1 | 0.0 | 0.0 | 3.1 | 0.0 | |
| · | 4.7 | 0.6 | 5.9 | 4.7 | 0.6 | 5.9 | |
| · | 0.0 | 2.9 | 0.0 | 0.0 | 2.9 | 0.0 | |
| N | 2.9 | 0.0 | 2.3 | 2.9 | 0.0 | 2.3 | |

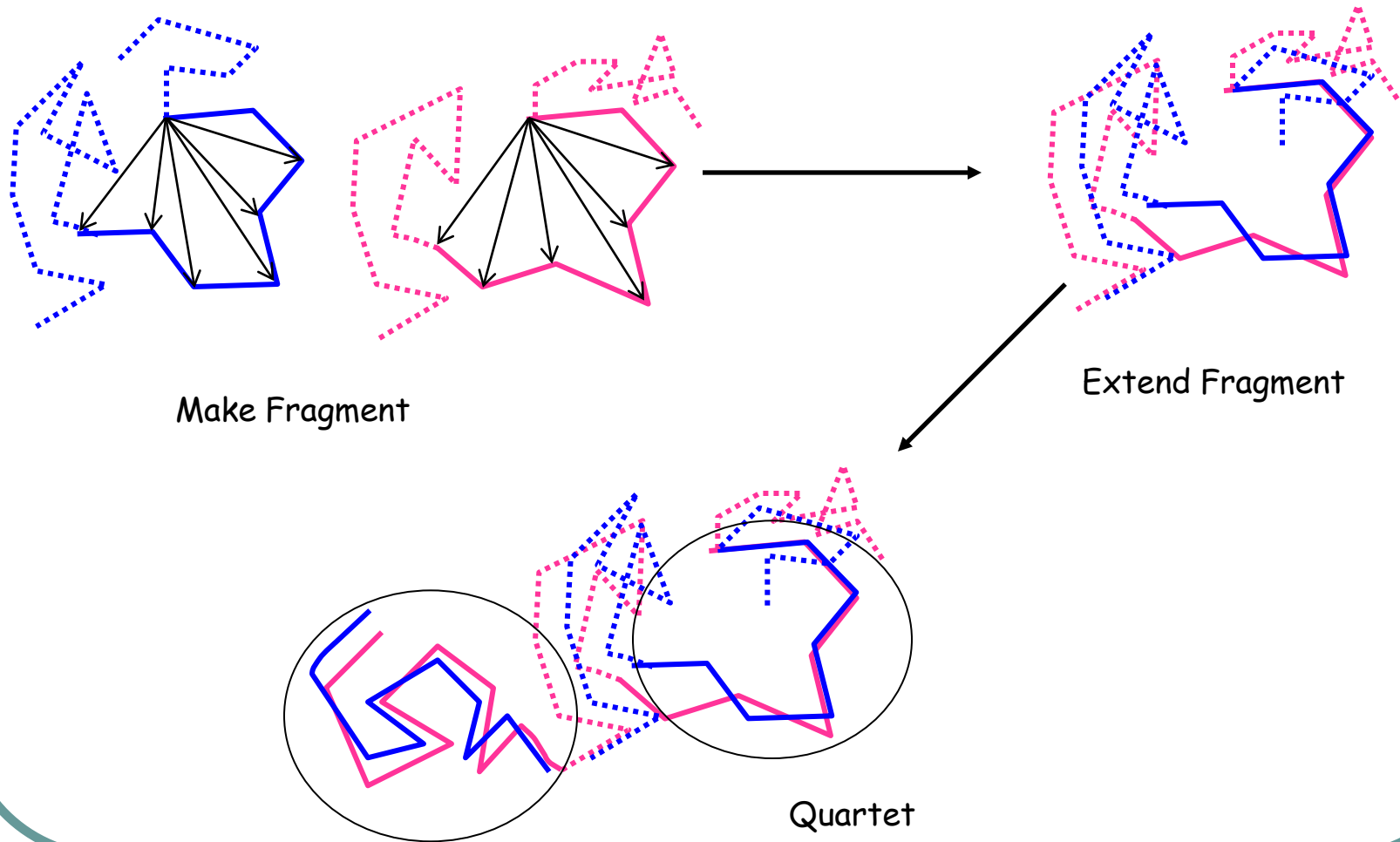
Structure A



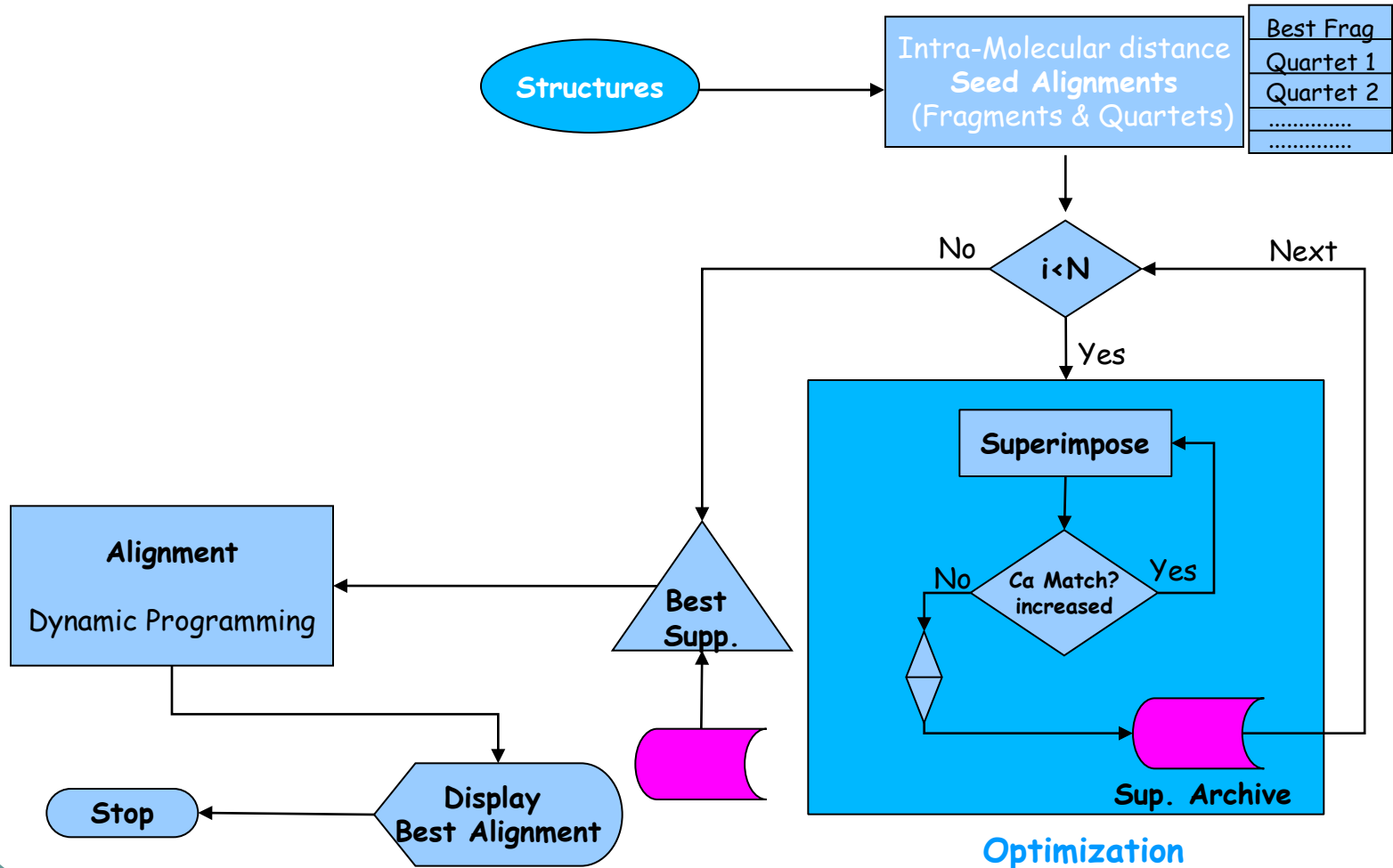
| | 1 | 2 | 3 | 4 | 5 | | M |
|---|-----|-----|-----|-----|-----|-------|---|
| 1 | 0.0 | | | | | | |
| 2 | 1.4 | 0.0 | | | | | |
| 3 | 3.0 | 2.1 | 0.0 | | | | |
| 4 | 0.7 | 0.3 | 0.0 | 0.0 | | | |
| 5 | 1.9 | 0.0 | 0.4 | 0.4 | 0.0 | | |
| · | 0.0 | 3.1 | 0.0 | 0.0 | 3.1 | | |
| · | 4.7 | 0.6 | 5.9 | 5.9 | 0.6 | 0.6 | |
| · | 0.0 | 2.9 | 0.0 | 0.0 | 2.9 | 2.9 | |
| M | 2.9 | 0.0 | 2.3 | 2.3 | 0.0 | 0.0 | |

Structure B

Seed Alignments

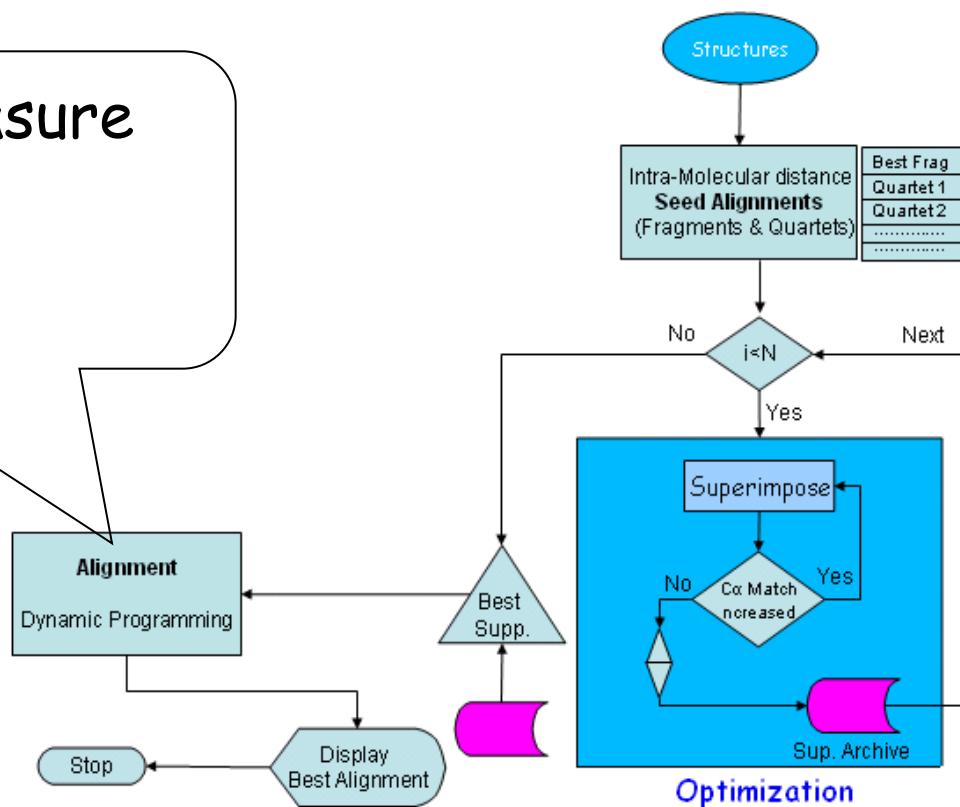


Algorithm



Algorithm- Pitfalls

- Simple Similarity measure
1: Match
0: No match



Algorithm- Pitfalls

```
1      11      21      31      41      51      61
APTATLANGDTITGLNAIINEAFLSIPF-AEPPVGNLRFKDPVPYSGSLDGQKFTSYGSPSCMQQNPEGTY
* * *
* * *
MINAIRT PDQRF SNLDQY PFS PNY--LDDL-----PGYPGLRAHYL-----
1      11      21      31

71      81      91      101      111      121      131
EENLPKAALDIVMQSKVFEAVSPSSDCLTINVVRPPGTKAGANLPVMLWIEGGGFVGGTSTFPPAQMI
A B C * * * * * * * * * * *
*** D * * * * * * * * * * *
-----DEGNSDAE-----DVFLCLHGEPTWS-----YLYRKMIP
41      51      61
```

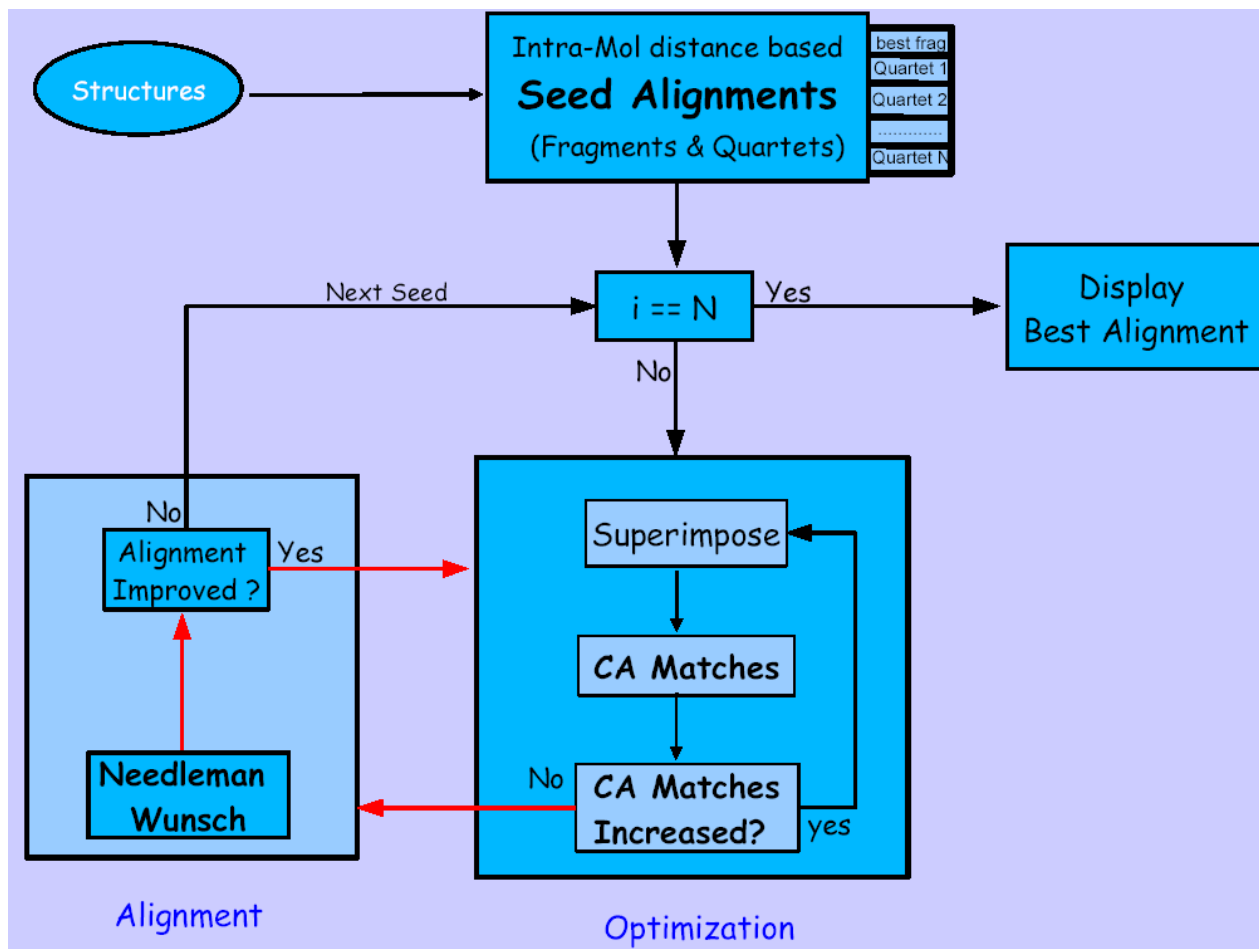
- Problems in aligning two structures low similarity.
- Alignments sometimes fragmented.
- Single matching residues not in sequential order (letters in the alignment).
- No optimization after dynamic programming.

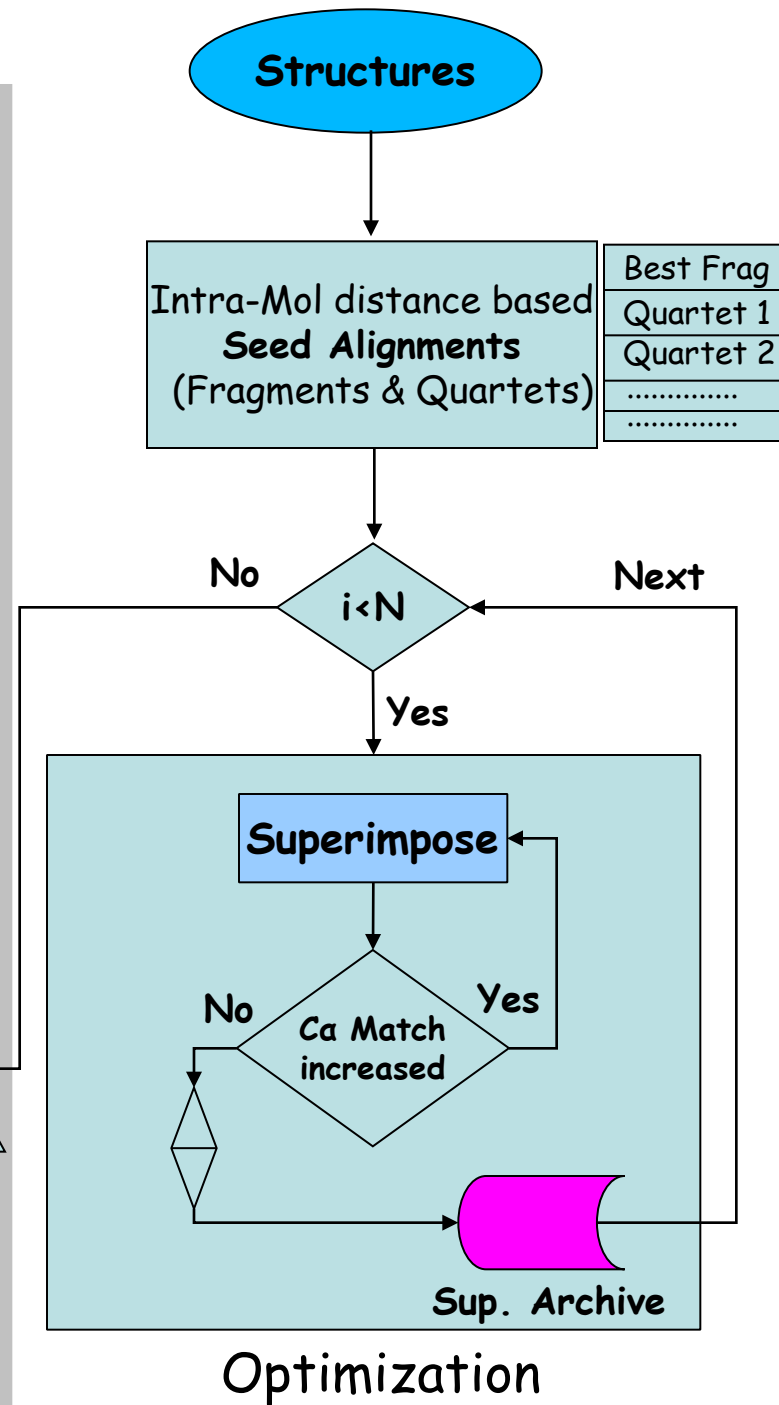
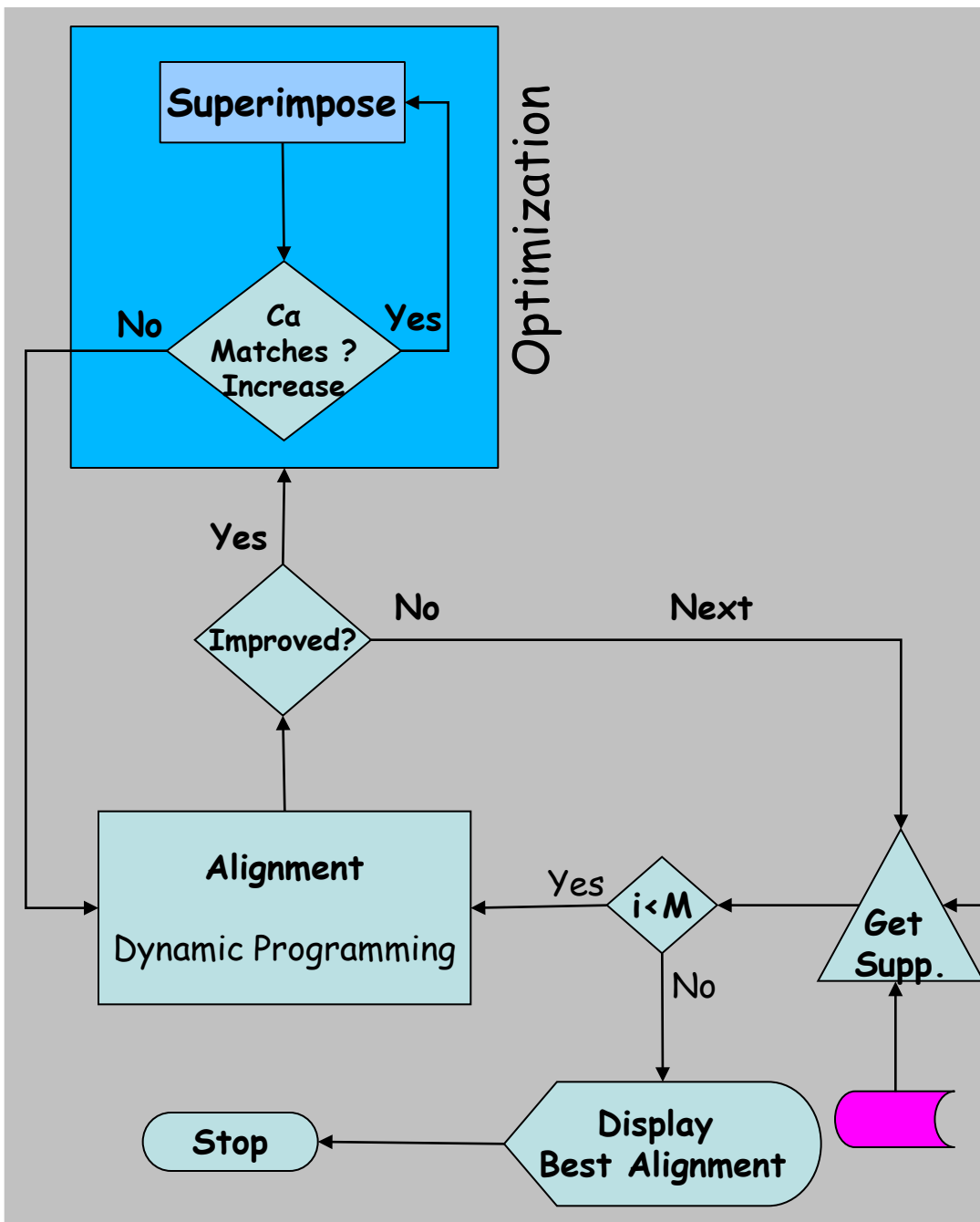
Algorithm- Pitfalls

| No | Chain 1 | Chain 2 | Protein3DFit $N^A / r.m.s.d.(A^\circ)$ | VAST $N^A / r.m.s.d.(A^\circ)$ | DALI $N^A / r.m.s.d.(A^\circ)$ | CE $N^A / r.m.s.d.(A^\circ)$ |
|----|------------|---------|---|-----------------------------------|-----------------------------------|---------------------------------|
| 1 | 1FXI:A | 1UBQ:_ | 39/1.0 | 48/2.1 | — | — |
| 2 | 1TEN:_ | 3HHR:B | 71/1.0 | 78/1.6 | 86/1.9 | 87/1.9 |
| 3 | 3HLA:B | 2RHE:_ | 42/1.1 | — | 63/2.5 | 85/3.5 |
| 4 | 2AZA: A | 1PAZ:_ | 52/1.2 | 74/2.2 | — | 85/2.9 |
| 5 | 1CEW:I | 1MOL:A | 54/1.2 | 71/1.9 | 81/2.3 | 69/1.9 |
| 6 | 1CID:_ | 2RHE:_ | 59/1.1 | 85/2.2 | 95/3.3 | 94/2.7 |
| 7 | 1CRL:_ | 1EDE:_ | 100/1.2 | | 211/3.4 | 187/3.2 |
| 8 | 2SIM:_ | 1NSB:A | 129/1.2 | 284/3.8 | 286/3.8 | 264/3.0 |
| 9 | 1BGE:B | 2GMF:A | 52/1.1 | 74/2.5 | 98/3.5 | 94/4.1 |
| 10 | 1TIE:_ | 4FGF:_ | 70/1.1 | 82/1.7 | 108/2.0 | 116/2.9 |

Reference: 10 'difficult' structures from (Fischer et al.) obtained by Dali (Holm & Sander, 1993) , VAST (Madej et al., 1995) & CE (Shindyalov & Bourne, 1998)

Algorithm - Extension





Similarity Score

$$S_{ij} = \frac{M}{1 + \left(\frac{d_{ij}}{d_0}\right)^2}$$

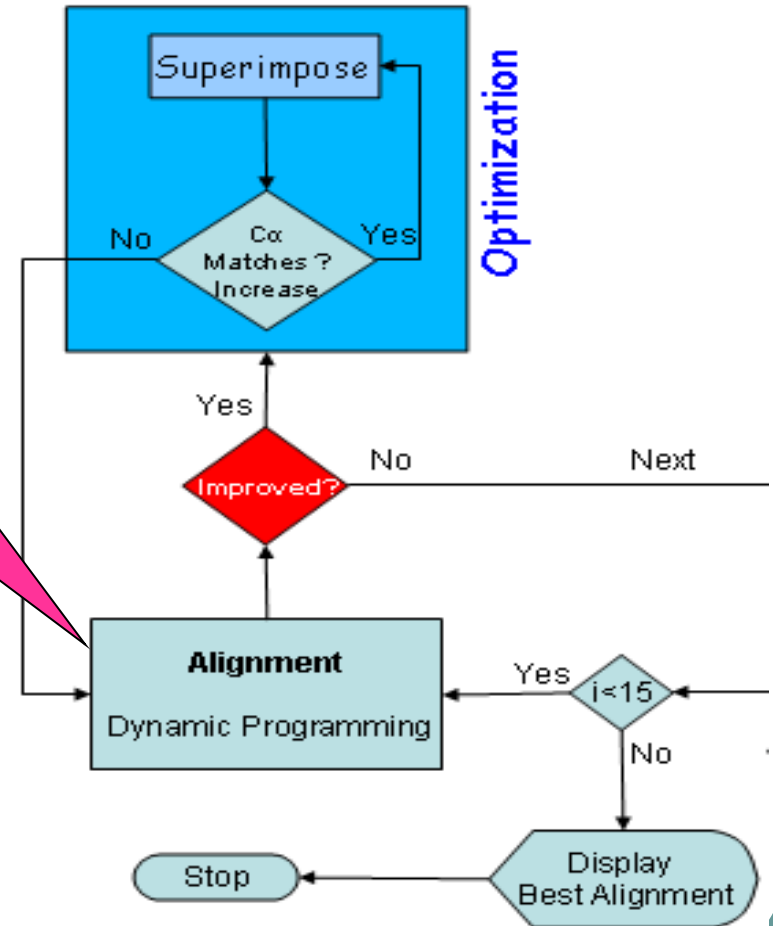
S_{ij} : similarity score used in DP

M : maximum score for a match

d_{ij} : distance

d_0 : distance resulting in half-maximal similarity

Source: Gerstein Levitt, 1998



Quality Function

Q score, ratio of no.
of aligned residues & RMSD

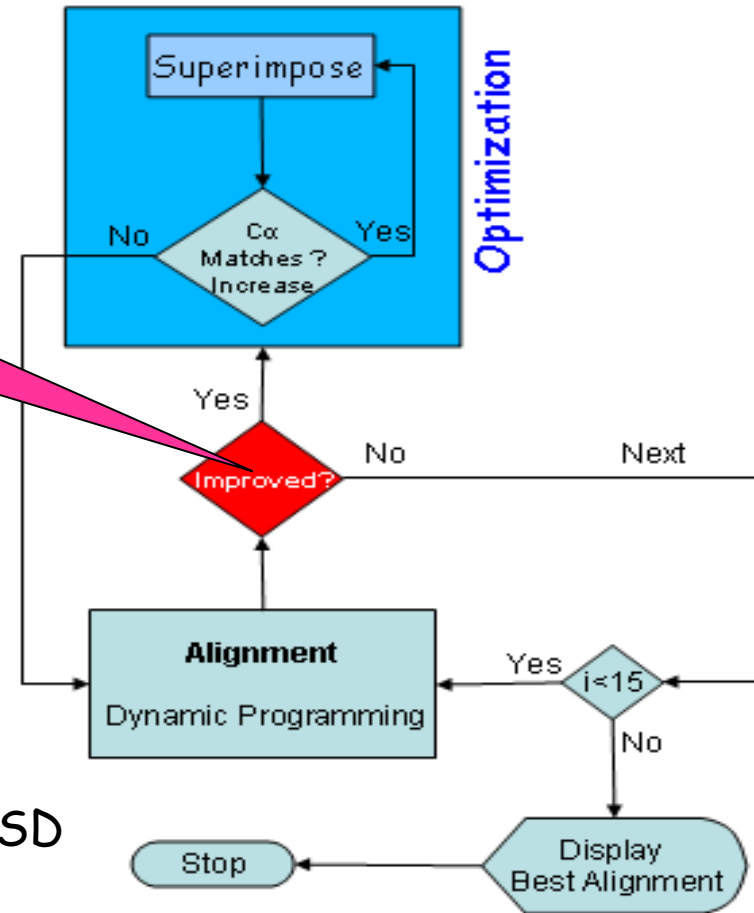
$$Q = \frac{N_{align}^2}{((1 + (RMSD/R_0)^2) N_1 N_2)}$$

N_{align} : number of matches (equivalent residues)

N_1, N_2 : length of structure 1 & 2 respectively

R_0 : empirical parameter measuring Relative significance of N_{align} & RMSD

Source: Krissinel & Henrich (2004)



Benchmarking and Extension of Protein3DFit

- o **Introduction**

- o Why Align Structures?
- o Structure Alignment: Issues & problem

- o **Method**

- o Algorithm
- o Pitfalls
- o Extension

- o **Results**

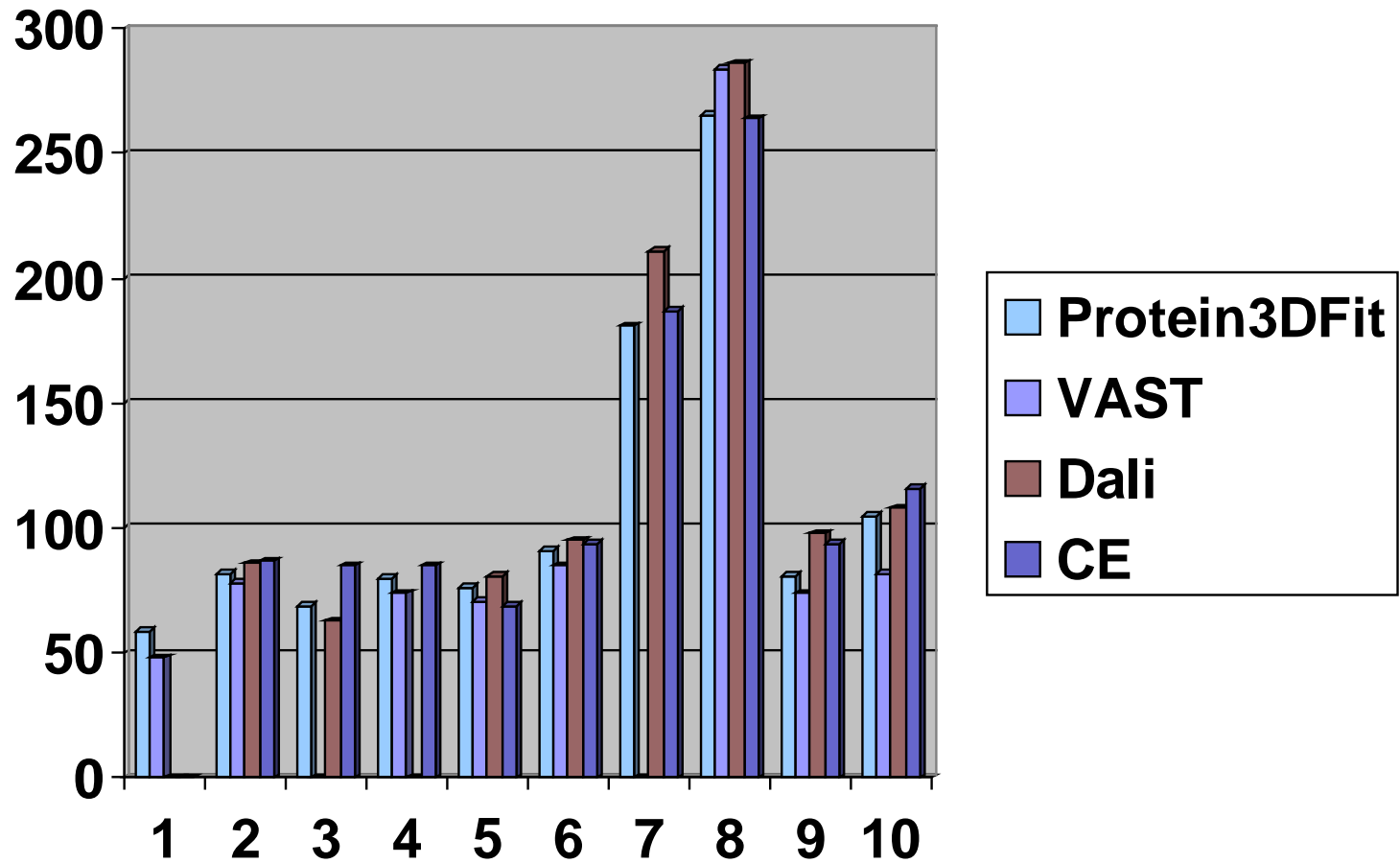
- o **Outlook**

Results

| No | Chain 1 | Chain 2 | Protein3DFit | | VAST | DALI | CE |
|----|---------|---------|--|--|--|--|---------|
| | | | N^A / <i>r.m.s.d.</i> (\AA°) | N^A / <i>r.m.s.d.</i> (\AA°) | N^A / <i>r.m.s.d.</i> (\AA°) | N^A / <i>r.m.s.d.</i> (\AA°) | |
| 1 | 1FXI:A | 1UBQ:_ | 59/2.9 | 39/1.0 | 48/2.1 | _ | _ |
| 2 | 1TEN:_ | 3HHR:B | 82/1.7 | 71/1.0 | 78/1.6 | 86/1.9 | 87/1.9 |
| 3 | 3HLA:B | 2RHE:_ | 69/3.4 | 42/1.1 | _ | 63/2.5 | 85/3.5 |
| 4 | 2AZA:A | 1PAZ:_ | 80/2.3 | 52/1.2 | 74/2.2 | _ | 85/2.9 |
| 5 | 1CEW:I | 1MOL:A | 76/2.1 | 54/1.2 | 71/1.9 | 81/2.3 | 69/1.9 |
| 6 | 1CID:_ | 2RHE:_ | 91/2.6 | 59/1.1 | 85/2.2 | 95/3.3 | 94/2.7 |
| 7 | 1CRL:_ | 1EDE:_ | 181/3.2 | 100/1.2 | | 211/3.4 | 187/3.2 |
| 8 | 2SIM:_ | 1NSB:A | 265/3.2 | 129/1.2 | 284/3.8 | 286/3.8 | 264/3.0 |
| 9 | 1BGE:B | 2GMF:A | 81/3.0 | 52/1.1 | 74/2.5 | 98/3.5 | 94/4.1 |
| 10 | 1TIE:_ | 4FGF:_ | 105/2.6 | 70/1.1 | 82/1.7 | 108/2.0 | 116/2.9 |

source: 10 'difficult' structures from (Fischer et al.) obtained by Dali (Holm & Sander, 1993) , VAST (Madej et al., 1995) & CE (Shindyalov & Bourne, 1998).

Results



Ca Matches: 59 (61.46%/77.63%), RMS-Deviation:2.95, Q-Value:0.84 Gaps: 44
 1FXI:A 96 residues -> 1UBQ:_ 76 residues

```

      11      21      31      41      51      61
YKVTLKT-PDGDNVITV-PD-DE--YILDVAEEEGEGL-DLPYSCRAGACSTCAGKLVSGPAPDEDQSFL
*****
*****
MQIFVKTLTGKTTITLEVEPSD-TIENVKAKIQDK-EGIPPD-----QQLRIF-----
1      11      21      31      41

      71      81      91
DDDQIQAGYIL-TCVAYPT-----GDCVIETHK-E
*****
*****
-----AGKQL-E-DGRTLSDYNIQKESTLHLV-LR
51      61      71

```

3Dfit-ext

CE

```

1FXI:A 5/6 VTLKTPD-GDNVITVPDDEY---ILDVAEEEGEGLDLPYSCRAGACSTCAGKLVSGPAPDED
1UBQ:_ 3/4 IFVKTLTGKTTITLEVEPSDTIENVKAKIQD-KEGIPPD-----QQLRIFA-----

1FXI:A 61/62 QSFLLDDQIQAGYILTCVAYPT-----GDCVIETHKKEAL
1UBQ:_ 47/48 -----GKQLEDGRTLSDYNIQKESTLHLVLRIRG

```

Rmsd = 3.8Å Z-Score = 3.1
 Sequence identity = 9.4%
 Aligned/gap positions = 64/36

3Dfit-old

pdb1fxi 96 residues -> pdb1ubq 76 residues
 matching Ca: 39 (40.62% / 51.32%)
 rms deviation: 1.027523 min. length: 6

```

1      11      21      31      41      51      61
ASYKVTLKTPD-GDNVITVPDDE---YILDVAEEEGEGL-DLPYSCRAGACSTCAGKLVSGPAPD-EDQSFL
* **** * * * A* *** B cc *
* **** * * * F E * * * A *** D *cc
M--QIFVKTLTGKTTITLEVEPSDTIENVKAKIQDK-EGIPPDQQR-----LIFAGKQLEDGRTLSDY
1      11      21      31      41      51

      71      81      91
DDDQIQAGYILTCVAYPTGDCVIETHKKEALY
D E F *****
B *****
NIQKES-----TLHLVLRIRGG
61      71

```

Dali

No 1: Query=mol1A Sbjct=mol2 Z-score=4.0

```

DSSP 11EEEEEEEL.LEEEEEEELLLL...LHHHHHHHLLLLLLLL11111111LLEEEe111
Query asYKVTLKTP.DGDNVITVPDDE...YILDVAEEEGEGLDLPYscragacsTCAGKlvsgpa 56
ident || | | |
Sbjct ..MQIFVKTLTGKTTITLEVEPSDtieNVKAKIQDKEGIPPD.....QQLRIF..... 45
DSSP ..LEEEEEEEL1LLEEEEEELLLL1hhHHHHHHHHHHLLLLHH.....HEEEEE.....

DSSP 1111111111111111LLEEEhHLLL....LLEEEEL111111
Query pdedqsfldddqiqAGYILTCVAYP.....TGDCVIETHKkealy 96
ident
Sbjct .....aGKQLE.DGRTLsdynIQKESTLHLVlrirgg 76
DSSP .....1LEELL.LLLLLhhhh1111111111111111

```

3Dfit-ext

```

CA Matches: 69 (69.70%/60.53%), RMS-Deviation:3.47, Q-Value:0.82 Gaps: 55
3HLA:B 99 residues -> 2RHE:_ 114 residues

      11      21      31      41
KIQVYSRHPAE-NG-KSNFLN-C-YVSGF--HPSDIEVDLLKN-----GERI-E-----K-----
**** *  ***  *  *****  ****  ***  *****  ****  *
**** *  ***  *  *****  ****  ***  *****  ****  *
VLTQ-PPSASGT-PGQRVTI-S-C-TGSATDIGS-NSVIWYQQVPGKAPKLLI-YYNDLLPSGVS
      11      21      31      41      51

      51      61      71      81      91
---VEHS-DLSFSKDWsfY-LLYYTEFT----PTEKDEYACRVNHVTLsQPKIV-KWD-R
***  *  *  *****  **  *****
***  *  *  *****  **  *****
DRFSAS-K-S-----GT-SASLAISGLESEDE--ADYYCAAWNDSLDEPGFGGGT-KL
61              71      81      91      101
  
```

CE

```

3HLA:B 4/5      TPKIQVYSRHPAENGKSNFLNCYVSGFHP-SDIEVDLLKN-----GERIE-----
2RHE:_ 3/4      VLTQPPSA----SGTPGQRVTISCTGSATDIGSNSVIWYQQVPGKAPKLLIYYNDLLPSGV

3HLA:B 48/49   ---KVEHSDLSFSKDWsfYLLYYTEFTPTEKDEYACRVNHV--TLsQPKIVKWDR
2RHE:_ 60/61   SDRFSASKSG-----TSASLAISGLESEDEADYYCAAWNDSLDEPGFGGGTKL
  
```

Rmsd = 3.4Å **Z-Score = 3.7**
Sequence identity = 2.4%
Aligned/gap positions = 84/31

```

1_pdb3hla 99 residues -> pdb2rhe 114 residues
matching Ca: 42 ( 42.42% / 36.84% )
rms deviation: 1.166318 min. length: 7
  
```

3Dfit-old

```

1              11
IQR-----TPKIQVYSRHPAE-----
      *  *****
      **  *****
F      E      D C B A      *  *****      ooo N
ESVLTQPPSASGTPGQRVTISCTGSATDIGSNSVIWYQQVPGKAPKLLIYYNDLLPSGVSDRFSASKSGT
1      11      21      31      41      51      61

      21      31      41      51      61      71
-----NGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWsfYLLY-YTEFTPTE
      *  *****      A B C D      E F *  G  **  *****hI
M L K J I *hh  ***** G      *  **  *****
SASLAISGLESEDE-ADYYCAAWNDS-----LDEPG---FGGGTKLTVLG
71      81      91              101      1

      81      91
KDEYACRVNHVTLsQPKIVKWDRDM
J K L M N      ooo

QPK
11
  
```

No 1: Query=mol1B Sbjct=mol2 Z-score=5.4

Dali

```

DSSP 11111EEEEEEe1LLLLLLLLLEEEEEEEEEEE111LEEEEEEEEL....LEELL.....
Query iqrtPKIQVYsrHPAENGKSNFLNCYVSGFhpsDIEVDLLKN....GERIE..... 47
ident | | | | |
Sbjct ...esVLTQPPsASGTPGQRVTISCTGSATdigSNSVIWYQQvpgkaPKLLIyyndllps 57
DSSP ..11LLLLLleEEELLLLLLEEEEEEEELLL1111LLEEEEEEL11111LEEEEE111111111

DSSP .....LLEELllee1111leEEEEEEEE.11LLLL..LEEEEEELLLLLLEEE.ELLL
Query .....KVEHSDlsfskdwsfYLLYYTEF.tpTEKD..EYACRVNHVTLsQPKIV.KWDR 97
ident | | | | |
Sbjct gvsdRFSASKS.....GTSASLAISglESEDeaDYYCAAWNDSLDEPGFGgGtKL 107
DSSP 1111leEEEE.E.....LLEEEEEEL111LHHHleEEEEEEEEELLLLLLEEEEL1LEEE
  
```

```

DSSP .....11
Query .....dm 99
ident
Sbjct tvlgqpk 114
DSSP ee11111
  
```

CA Matches: 265 (69.55%/67.95%), RMS-Deviation:3.20, Q-Value:0.80 Gaps: 182
2SIM: 381 residues -> INSB:A 390 residues

```
VEKSVVFK-AEGEFTDQKGNITVGS GSGGTTKYFRIPAMCTTSKGTIVVFD
***** ** ***
***** ** ***
KALLI-SPHRFG-EARGNSA-----PLIIREPFIACGPK-ECKHFAL
          96          106          116          126

          62          72          82          92          102
AR-H-N-----TASD-QSFIDTAAARSTDGKWTW----NKKIAI-YNDRVMSKLSRVMDPTCIVAN-
** * *          **** * * * *          * * * *          * * * *          * * * *
** * *          **** * * * *          * * * *          * * * *          * * * *
THYAAQPGGYNGTREDRNKLRHLISVK--LG--KIPTVENSIFH-M-A-----AWSGSACHD-G
          136          146          156          166          176

          112          122          132          142          152          162          172
IQGREILVMVGKWNNDKTWGA YRDKAPD TDWDLVLYKSTDDGVTF SKVETNI-HD IVTKNGTI--SAM
***** * * * *          * * * *          * * * *          * * * *          * * * *
***** * * * *          * * * *          * * * *          * * * *          * * * *
---REWYIGVDG--P---D---S-----N--ALIKIKYG-----EAYTD-TY---H---SYAMNII
          186          196          206          216

          182          192          202          212          222          232
LGGV-GSGLQ-LNDG-KLVFPVQMV-RTKNITVLTNSFIYSTDGIT-W-SLPSGY--CE-GFGSENN-I
*** * * * *          * * * *          * * * *          * * * *          * * * *
*** * * * *          * * * *          * * * *          * * * *          * * * *
RTQ-ESACNCI---GGCYLMTD-G---SASGISKRFLKIREGR-I-IKEI-FPTGRVEH-TEECT-
          226          236          246          256          266          276

          242          252          262          272
I-EFNASLVNNIRN---SGLRRSFETKD---FGKTWTEFPMD---KKVDNR--N-----
* * * * *          * * * *          * * * *          * * * *          * * * *
* * * * *          * * * *          * * * *          * * * *          * * * *
GFASNKTIECACRDNSTYAKRPFVKLVNVEDT--AEIR---LMCTETYLDTPRPDDGSITGPCE SNGDGK
          286          296          306          316          326          336

          282          292          302          312          322          332
-HGVQGSTITIPSGNKLVA AHSSAQNK-NNDYTRSDISLYAHN-LYSG-----EVKLIDDFY--PKVG
***** * * * *          * * * *          * * * *          * * * *          * * * *
***** * * * *          * * * *          * * * *          * * * *          * * * *
RGGIKGPFVHQRMASKIGRWY-SRTMSK-T--ERMGMELVYR-Y-DG-DPWTSDALAHSGVMVSMKE-P
          346          356          366          376          386          396

          342          352          362          372
NASGAGYSCLSYRKNVDKETL-YV-VYE-A-N-----GSIEFQDLSRHL P-V
***** * * * *          * * * *          * * * *          * * * *          * * * *
***** * * * *          * * * *          * * * *          * * * *          * * * *
---GWYSFGFEIKDKCDVPC-IG-IEM-VHDGKKTWHSAAATAIYCLM--GSG
```

3Dfit-ext

```
pdb2sim 381 residues -> pdb1nsb 390 residues
matching Ca: 129 ( 33.86% / 33.08% )
rms deviation: 1.276508 min. length: 10

2
TVEK-----

          gg H          JJJ KKK LLLLLLLLLL I
EPEWYPRLSQCGSTFQKALLISPHRFGEARGNSAPLIIREPFIACGPKCKHFALHYAAQPGGYNGT
76 86 96 106 116 126 136

          12 22 32 42
-----SVVFKAEGEHFTDQKGNITVGS GSGGTTKYFRIPAMCTTSKGTIVV
          **** * * * * * * * *
II A M * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
REDRNKLRHLISVKLGKIPTVENSIFHMAA-----WSGSACHDGR-EWTY
146 156 166 176 186

52 62 72 82 92 102 112
FADARHNTASDQSF-IDTAAARSTDGKWTWKKIAIYNDRVMSKLSRV-MDPTCI--VANIQGREILV
**** A * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
**** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
IGVDGP-----DSNALIKIKYGE-----AYTDTVHYSYA-----MNLIR-TQESACNIGG---DCYL
          196          206          216          226          23

          122 132 142 152 162 172 182
MVGKWNNDKTWGA YRDKAPD TDWDLVLYKSTDDGVTF SKVETNIHD IVTKNGTISAM LGVGSGLQLND
*** * * * *          * * * *          * * * *          * * * *          * * * *          * * * *
*** * * * *          * * * *          * * * *          * * * *          * * * *          * * * *
MITDGS---ASGISK-----CRFLKIREG---RIIKEIF-PT-GRVE----HTEECT--CGFAS-
6 246 256 266 276

          192 202 212 222
G-KLVFPVQMVRTKNITVLTNSFIYSTDG--ITWLSLPSGYC-E-GFGS-----
* * * *          C * * * *          * * * *          * * * *
* * * *          * * * *          * * * *          * * * *          C
NKTIECACRDNSTYAK-----RPFVKLVNVEDTAEIRLMCTETYLDTPRPDDGSITGPCE SNGDKRGGI
286 296 306 316 326 336 34

          232 242 252 262 272 282
ENNIIEFN---ASLVNNIRN---SGLRRSFETKDF-GKTW---TEFPMDKKVDNRNHGVQGSTITIP
* * * *          * * * *          * * * *          * * * *          * * * *          * * * *
* * * *          * * * *          * * * *          * * * *          * * * *          * * * *
KGGFVHQRMASKIGRWYSRTMSKTERMGEMELVYRYDGPWTSDALAHSGV---MVMKEPGWYSFGFEIK
6 356 366 376 386 396 406

          292 302 312 322 332 342 3
SGNKLVA A-HSSAQNKNDY--TRSDISLYAHNLYS---GEVKLIDDFYPKVGNASGAGYSCLSYRKNVD
* * * *          * * * *          * * * *          * * * *          * * * *          * * * *
* * * *          * * * *          * * * *          * * * *          * * * *          * * * *
DKKCDVPCIGIEMVHDGKKTWH-SAATAI-YCLMGSGQLLWDTVTGVDMAL
416 426 436 446 456

52 362 372 382
KETLVVYVEANGSIEFQDLSRHL PVIKSYN
LLLLLLLLLLLLMMMMMMMM
```

3Dfit-old

Outlook / Possible Extensions

- Use of secondary structure information in alignment step.
- Update Protein3DFit webserver
- Multiple structure superposition
- Alignment of a protein structure against PDB
-

Acknowledgments

- Prof. Dietmar Schomburg
- Pascal Benkert
- Structural Biology Group & CUBIC

Questions?

Thanks a lot for your patience!