Bioinformatics analysis based on structure changes of RfaH

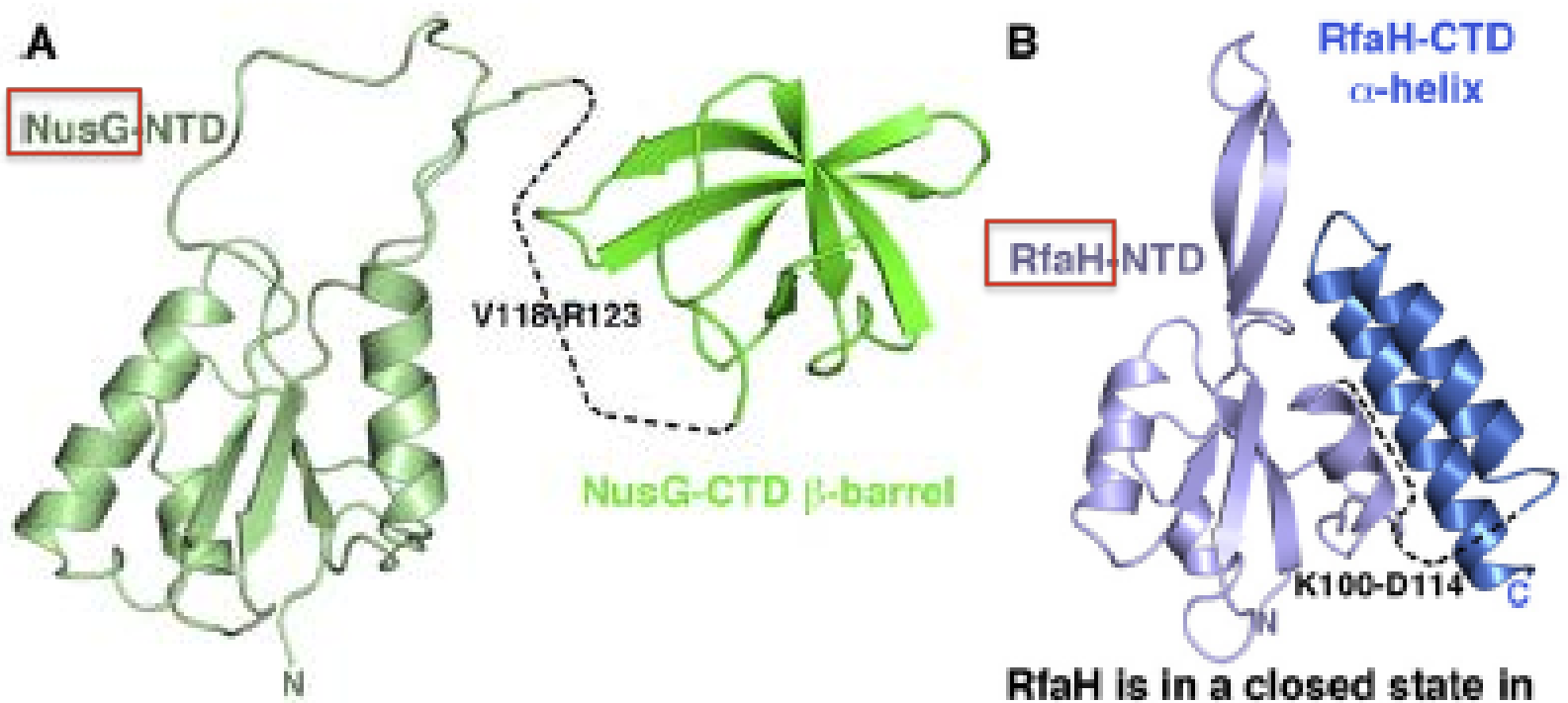
第8组 组员：黄骁骁，李鑫，邓琳娜，潘云龙

- The native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment.

----Christian B. Anfinsen, 1973

The Rules are made to be broken…

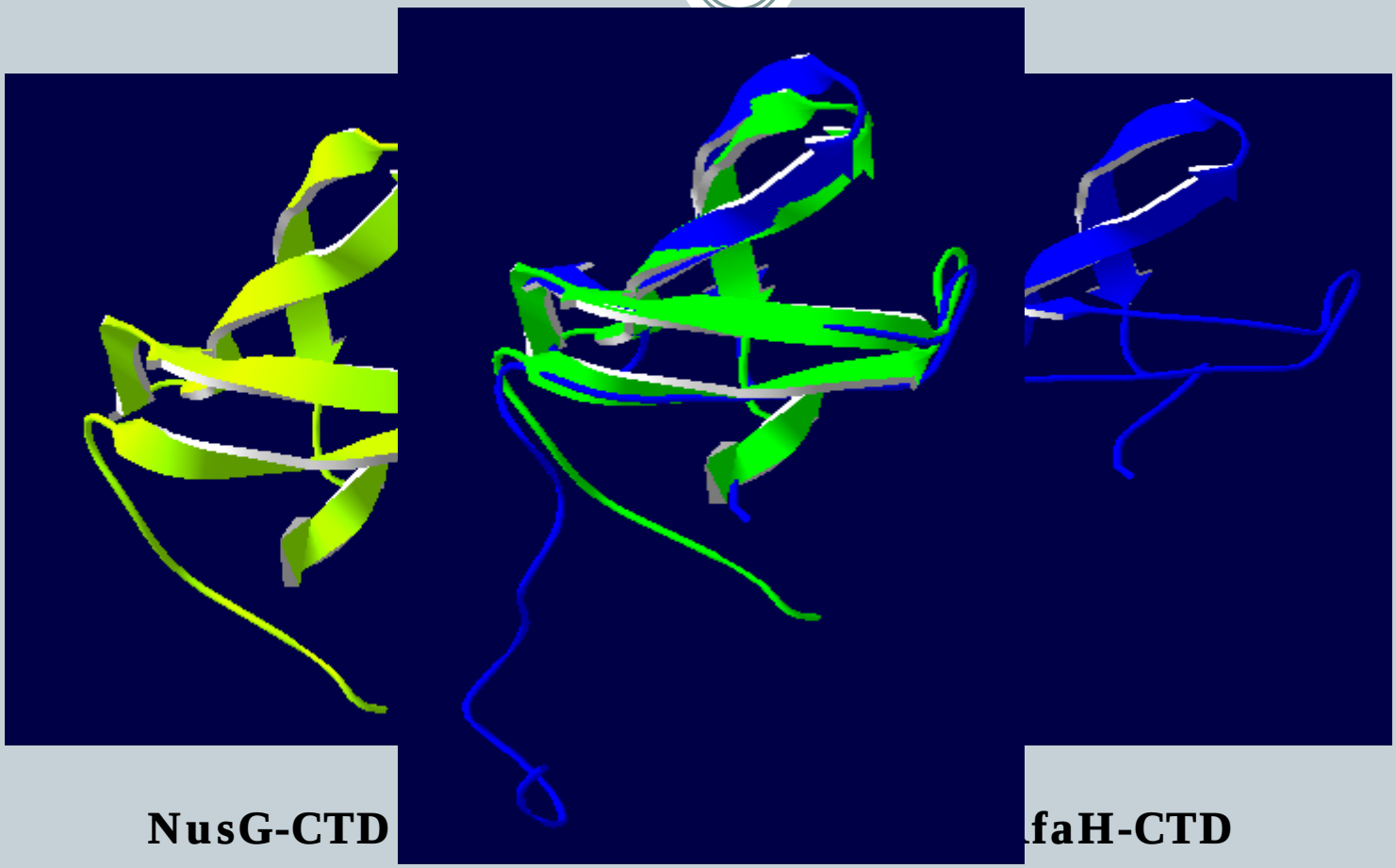
NusG homologs



NusG is in an open form in solution and has an all β-fold in its CTD

RfaH is in a closed state in the protein crystal and the CTD is in an all α-helical fold

NusG homolog

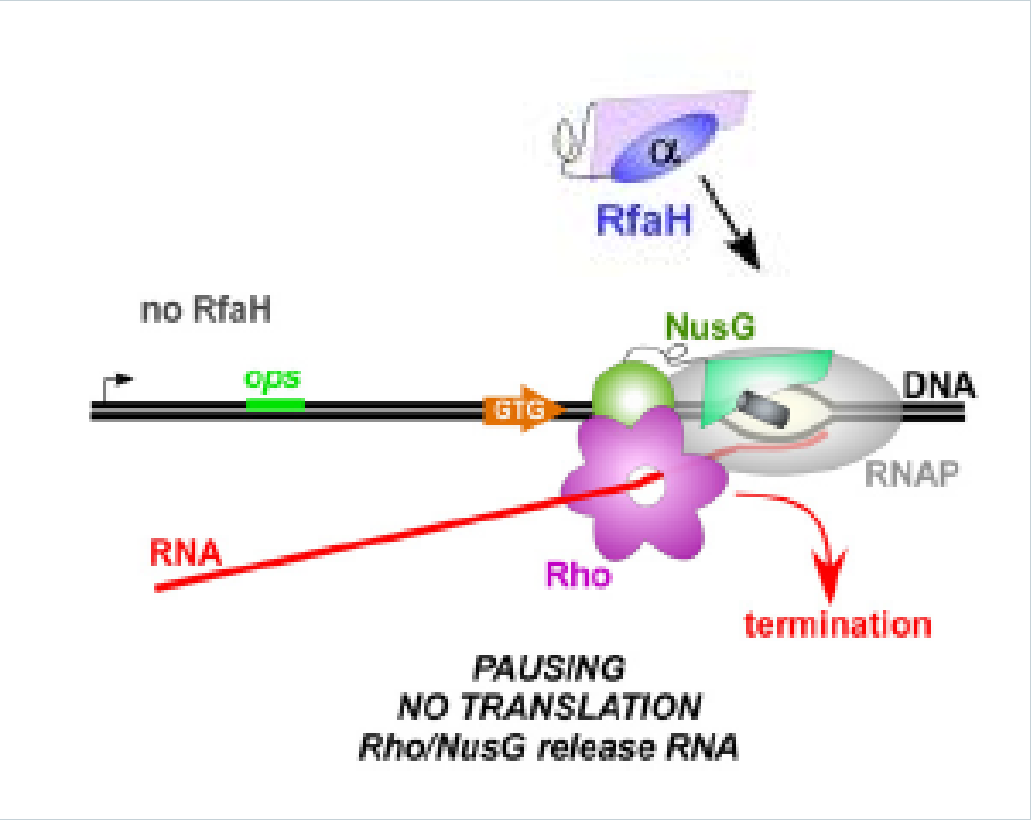


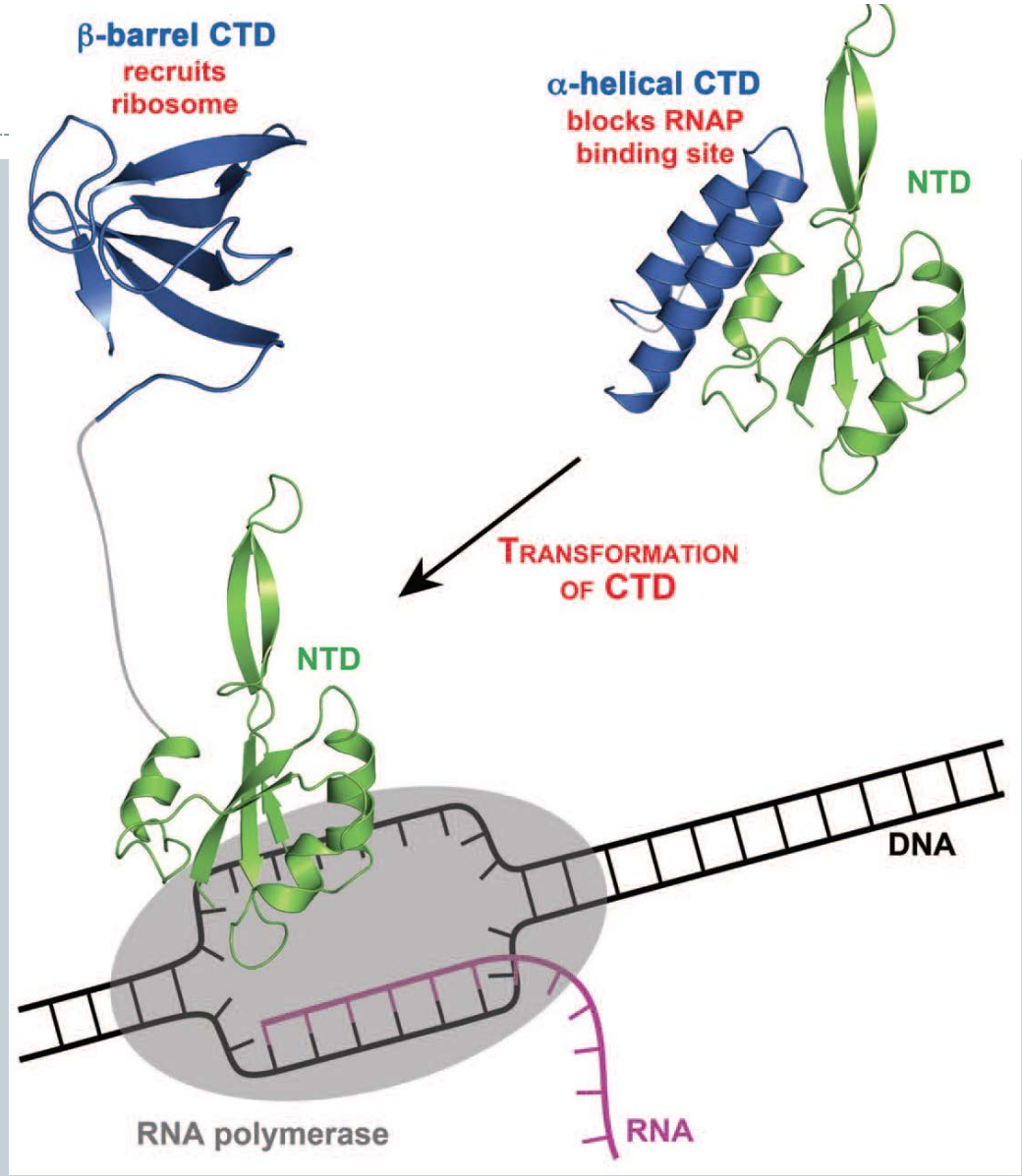
NusG-CTD

fa H-CTD

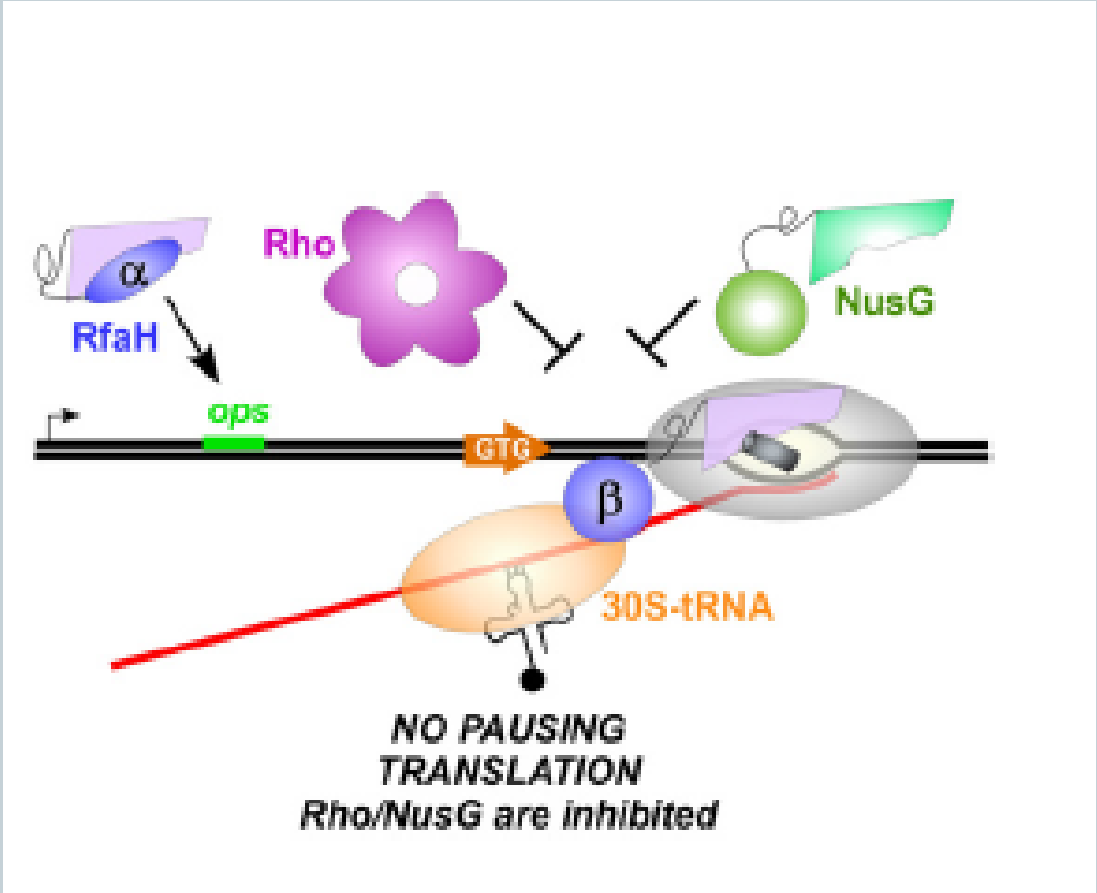
Backbone RMSD=0.65Å for P112-L162

No RfaH


α
RfaH
no RfaH
NusG
ops
GTG
DNA
RNAP
RNA
Rho
termination
PAUSING
NO TRANSLATION
Rho/NusG release RNA


β-barrel CTD
recruits
ribosome
α-helical CTD
blocks RNAP
binding site
NTD
TRANSFORMATION
OF CTD
NTD
DNA
RNA polymerase
RNA

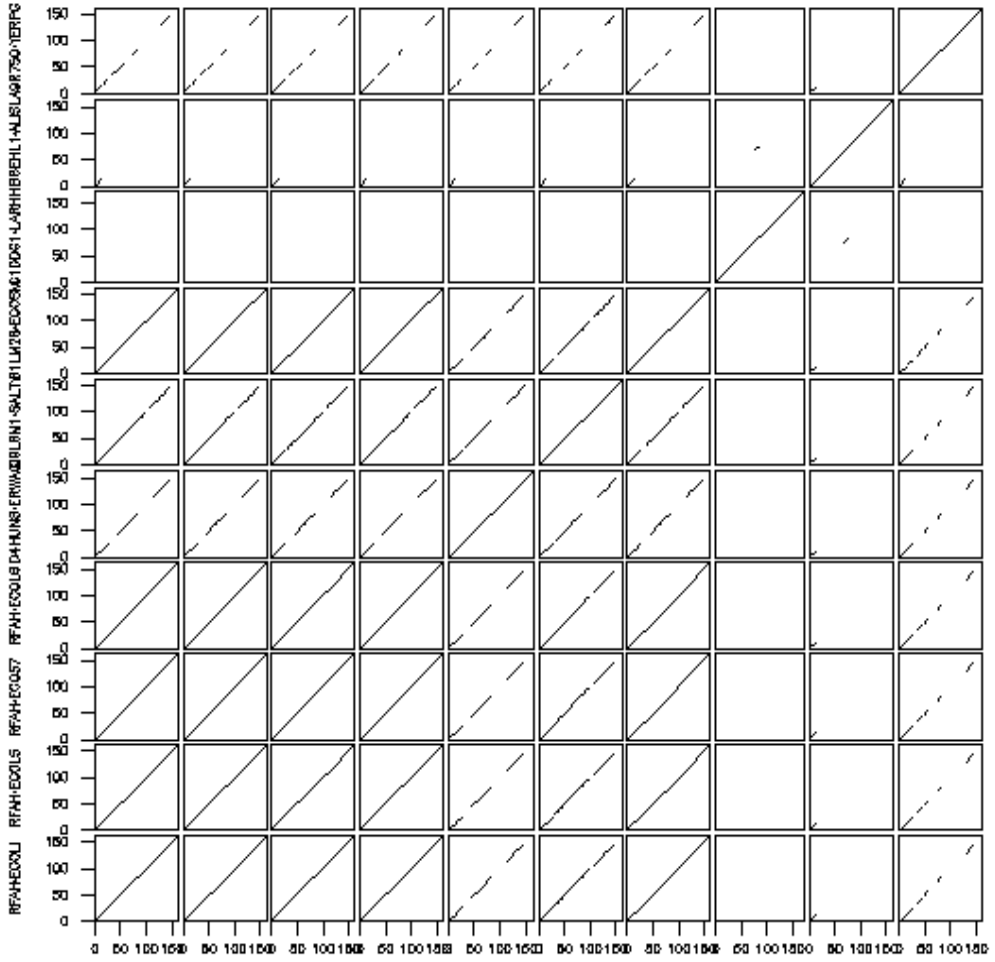
With RfaH


α
RfaH
Rho
NusG
ops
GTG
β
30S-tRNA
NO PAUSING
TRANSLATION
Rho/NusG are inhibited

Uniprot数据分析

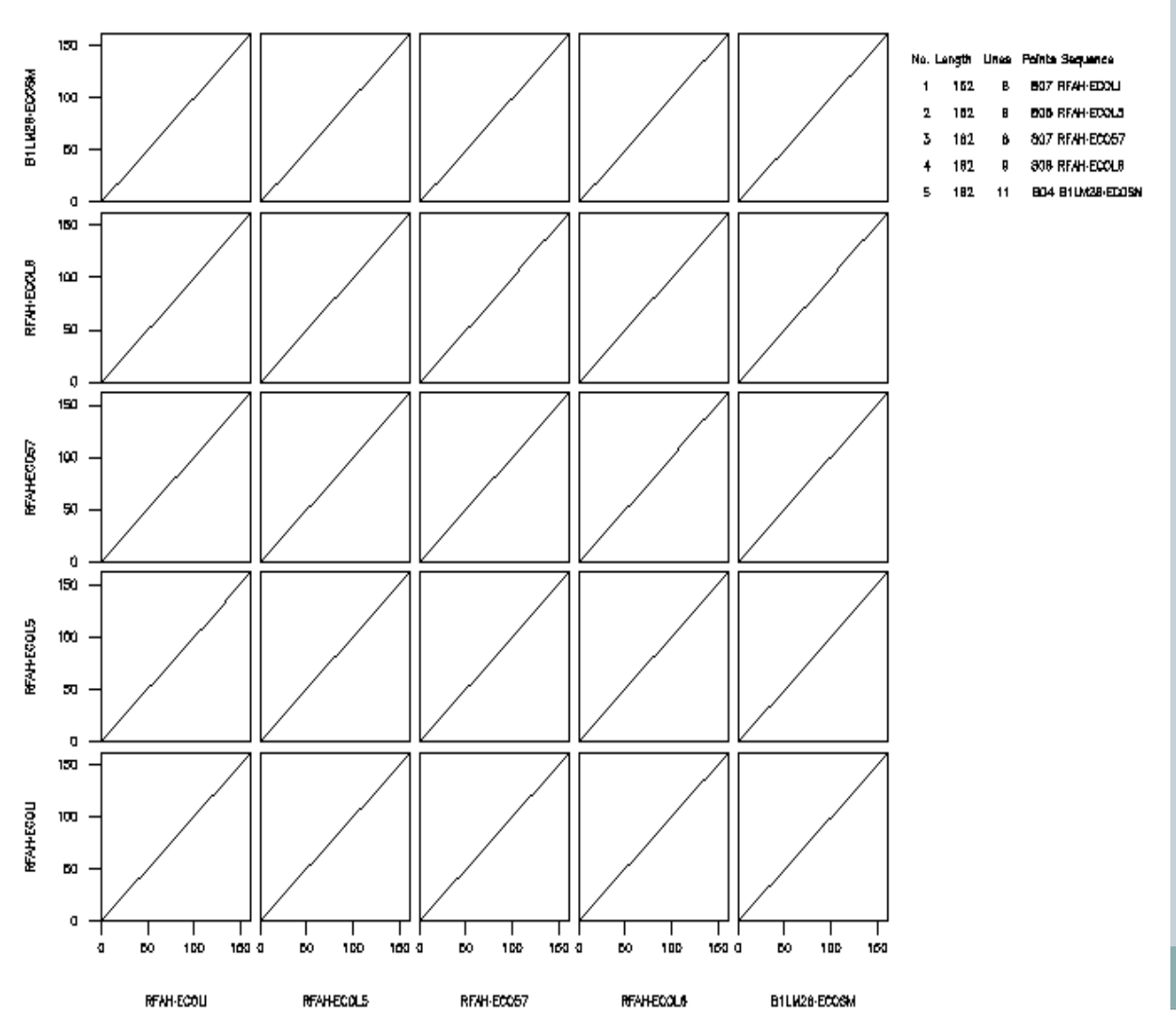
- >sp|P0AFW0|RFAH_**ECOLI** Transcription antitermination protein RfaH OS=Escherichia coli (strain K12) GN=rfaH PE=1 SV=1
- >sp|Q0TAL4|RFAH_**ECOL5** Transcription antitermination protein RfaH OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rfaH PE=1 SV=1
- >sp|P0AFW1|RFAH_**ECO57** Transcription antitermination protein RfaH OS=Escherichia coli O157:H7 GN=rfaH PE=3 SV=1
- >sp|Q8FBI4|RFAH_**ECOL6** Transcription antitermination protein RfaH OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rfaH PE=1 SV=1
- >tr|D4HUN8|D4HUN8_**ERWAC** Transcriptional activator RfaH OS=Erwinia amylovora (strain CFBP1430) GN=rfaH PE=4 SV=1
- >tr|Q9L6M1|Q9L6M1_**SALTY** 88% identity to E. coli transcriptional activator RFAH (SW:P26614) OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=rfaH PE=4 SV=1
- >tr|B1LM28|B1LM28_**ECOSM** Transcriptional activator RfaH OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=rfaH PE=4 SV=1
- >tr|C1DDG1|C1DDG1_**LARHH** RfaH OS=Laribacter hongkongensis (strain HLHK9) GN=rfaH PE=4 SV=1
- >tr|B6EHL1|B6EHL1_**ALISL** Transcriptional activator RfaH OS=Aliivibrio salmonicida (strain LFI1238) GN=rfaH PE=4 SV=1
- >tr|A9R750|A9R750_**YERPG** Transcriptional activator RfaH OS=Yersinia pestis bv. Antiqua (strain Angola) GN=rfaH PE=4 SV=1

Polydot初步分析



No.	Length	Lines	Points	Sequence
1	162	32	1077	RFAH-ECOLI
2	162	33	1076	RFAH-ECOL5
3	162	32	1077	RFAH-ECO57
4	162	33	1076	RFAH-ECOL6
5	188	50	748	D4HUN8-ERWAC
6	162	65	916	Q9L6M1-SALTY
7	162	34	1067	B1LM28-ECOSN
8	172	2	175	C1DDG1-LARHH
9	185	10	243	B6EHL1-ALISL
10	162	43	525	A8R750-YERPG

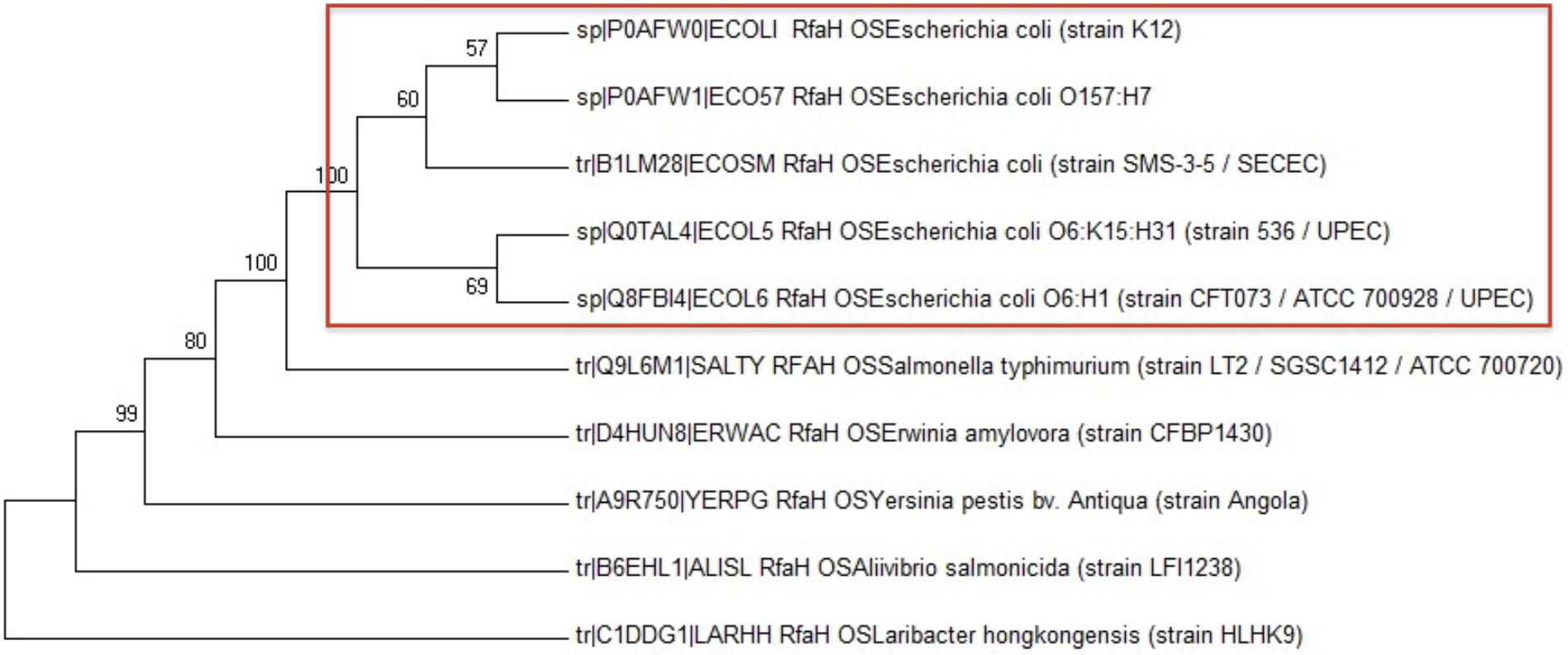
Polydot分析12347



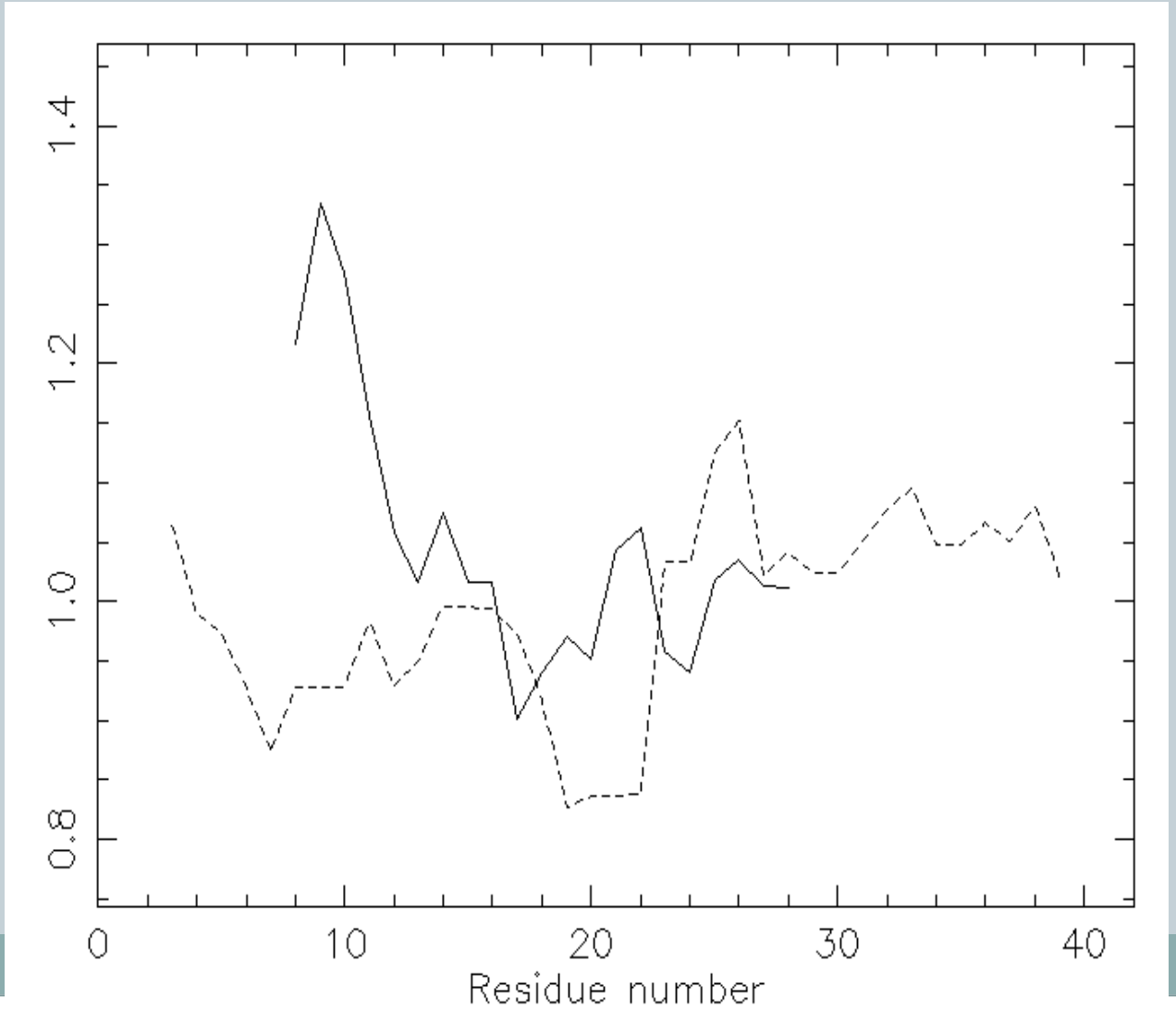
Muscle进行多序列比对

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


tr|B1LM28|B1LM28_ECOSM      MQSWYLLYCKRGQLQRAQEHLERQAVNCLAPMISLEKIVRGKRTAVSEPLFPNYLFVEFD
sp|Q0TAL4|RFAH_ECOL5        MQSWYLLYCKRGQLQRAQEHLERQAVNCLAPMITLEKIVRGKRTAVSEPLFPNYLFVEFD
sp|Q8FBI4|RFAH_ECOL6        MQSWYLLYCKRGQLQRAQEHLERQAVNCLAPMITLEKIVRGKRTAVSEPLFPNYLFVEFD
sp|P0AFW0|RFAH_ECOLI        MQSWYLLYCKRGQLQRAQEHLERQAVNCLAPMITLEKIVRGKRTAVSEPLFPNYLFVEFD
sp|P0AFW1|RFAH_ECO57        MQSWYLLYCKRGQLQRAQEHLERQAVNCLAPMITLEKIVRGKRTAVSEPLFPNYLFVEFD
                            ************************************:***********************

tr|B1LM28|B1LM28_ECOSM      PEVIHTTTINATRGVSHFVRFGASPAIVPSAVIHQLSVYKPKDIVDPATPYPGDKVIITE
sp|Q0TAL4|RFAH_ECOL5        PEVIHTTTINATRGVSHFVRFGASPAIVPSAVIHQLSVYKPKDIVDPSTPYPGDKVIITE
sp|Q8FBI4|RFAH_ECOL6        PEVIHTTTINATRGVSHFVRFGASPAIVPSAVIHQLSVYKPKDIVDPSTPYPGDKVIITE
sp|P0AFW0|RFAH_ECOLI        PEVIHTTTINATRGVSHFVRFGASPAIVPSAVIHQLSVYKPKDIVDPATPYPGDKVIITE
sp|P0AFW1|RFAH_ECO57        PEVIHTTTINATRGVSHFVRFGASPAIVPSAVIHQLSVYKPKDIVDPATPYPGDKVIITE
                            ************************************************:************

tr|B1LM28|B1LM28_ECOSM      GAFEGFQAIFTEPDGEARSMLLLNLINKEIKHSVKNTEFRKL
sp|Q0TAL4|RFAH_ECOL5        GAFEGFQAIFTEPDGEARSMLLLNLINKEIKHSVKNTEFRKL
sp|Q8FBI4|RFAH_ECOL6        GAFEGFQAIFTEPDGEARSMLLLNLINKEIKHSVKNTEFRKL
sp|P0AFW0|RFAH_ECOLI        GAFEGFQAIFTEPDGEARSMLLLNLINKEIKHSVKNTEFRKL
sp|P0AFW1|RFAH_ECO57        GAFEGFQAIFTEPDGEARSMLLLNLINKEIKHSVKNTEFRKL
                            *****************************************
```

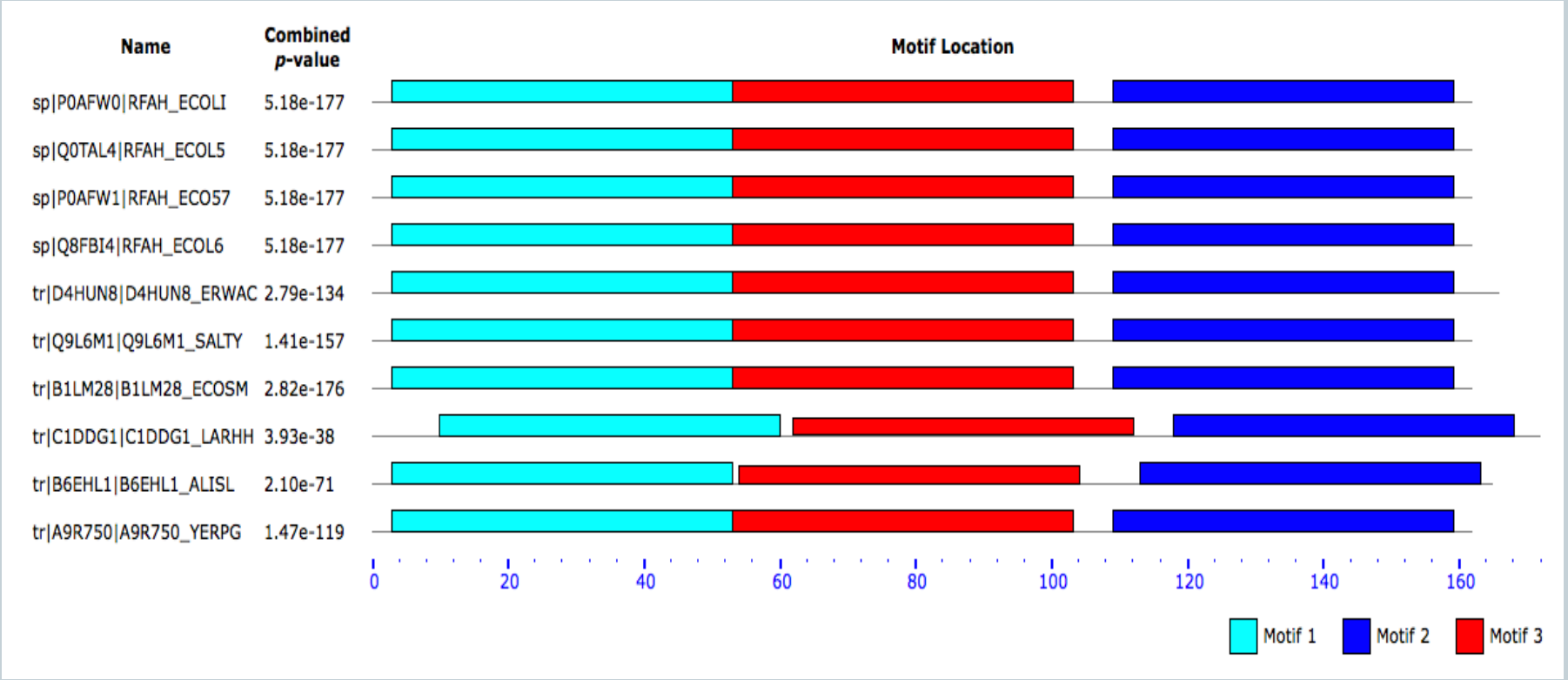
NJ法建树


sp|P0AFW0|ECOLI RfaH OSEscherichia coli (strain K12)
sp|P0AFW1|ECO57 RfaH OSEscherichia coli O157:H7
tr|B1LM28|ECOSM RfaH OSEscherichia coli (strain SMS-3-5 / SECEC)
sp|Q0TAL4|ECOL5 RfaH OSEscherichia coli O6:K15:H31 (strain 536 / UPEC)
sp|Q8FBI4|ECOL6 RfaH OSEscherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC)
tr|Q9L6M1|SALTY RFAH OSSalmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
tr|D4HUN8|ERWAC RfaH OSErwinia amylovora (strain CFBP1430)
tr|A9R750|YERPG RfaH OSYersinia pestis bv. Antiqua (strain Angola)
tr|B6EHL1|ALISL RfaH OSAliivibrio salmonicida (strain LFI1238)
tr|C1DDG1|LARHH RfaH OSLaribacter hongkongensis (strain HLHK9)
57
60
100
100
69
80
99

Tmap分析跨膜区域



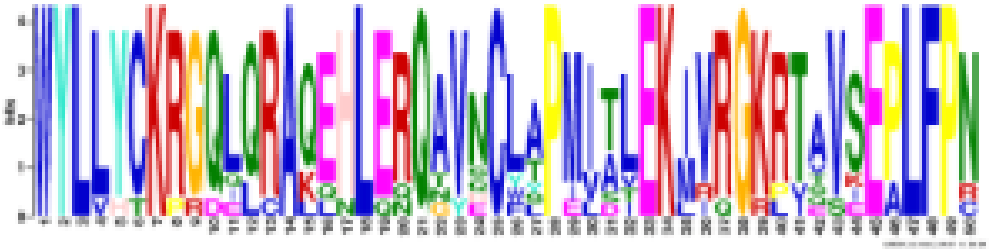
MEME 分析


Name
Combined p-value
Motif Location
sp|P0AFW0|RFAH_ECOLI 5.18e-177
sp|Q0TAL4|RFAH_ECOL5 5.18e-177
sp|P0AFW1|RFAH_ECO57 5.18e-177
sp|Q8FBI4|RFAH_ECOL6 5.18e-177
tr|D4HUN8|D4HUN8_ERWAC 2.79e-134
tr|Q9L6M1|Q9L6M1_SALTY 1.41e-157
tr|B1LM28|B1LM28_ECOSM 2.82e-176
tr|C1DDG1|C1DDG1_LARHH 3.93e-38
tr|B6EHL1|B6EHL1_ALISL 2.10e-71
tr|A9R750|A9R750_YERPG 1.47e-119
0
20
40
60
80
100
120
140
160
Motif 1
Motif 2
Motif 3

Sequence logo

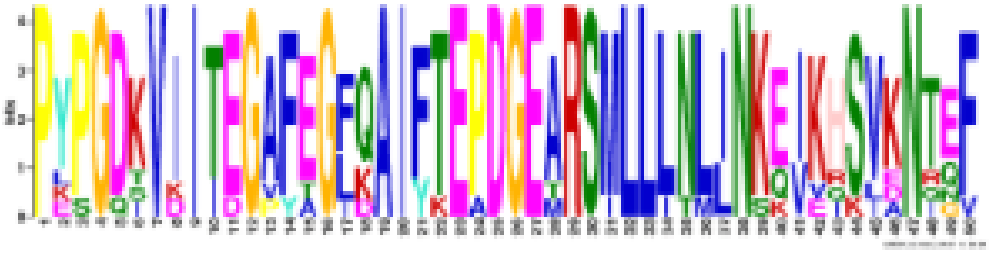
Motif 1

- 2.4e-291
- 10 sites



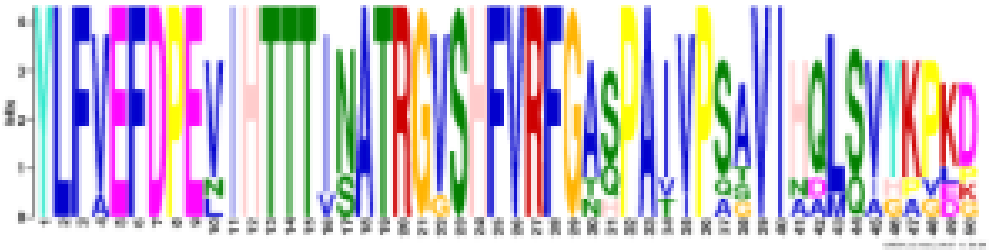
Motif 2

- 2.2e-240
- 9 sites

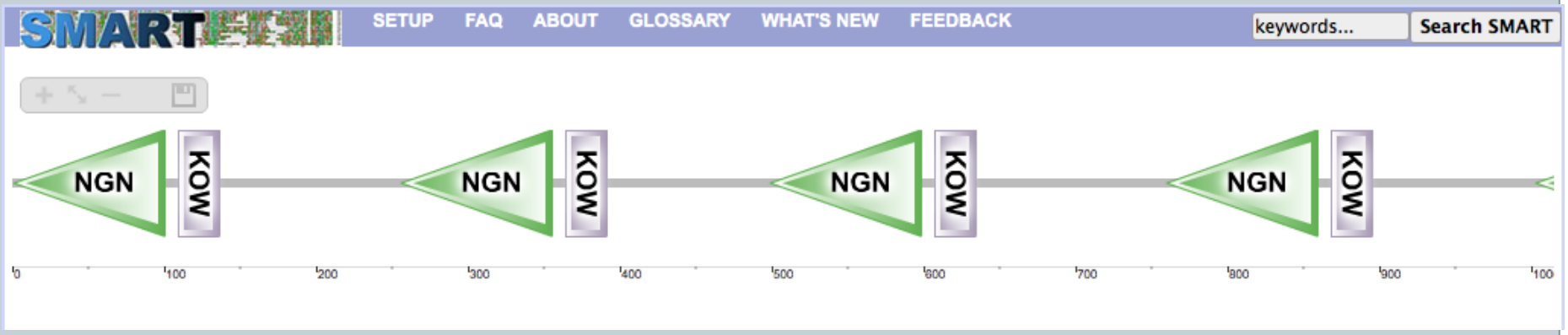


Motif 3

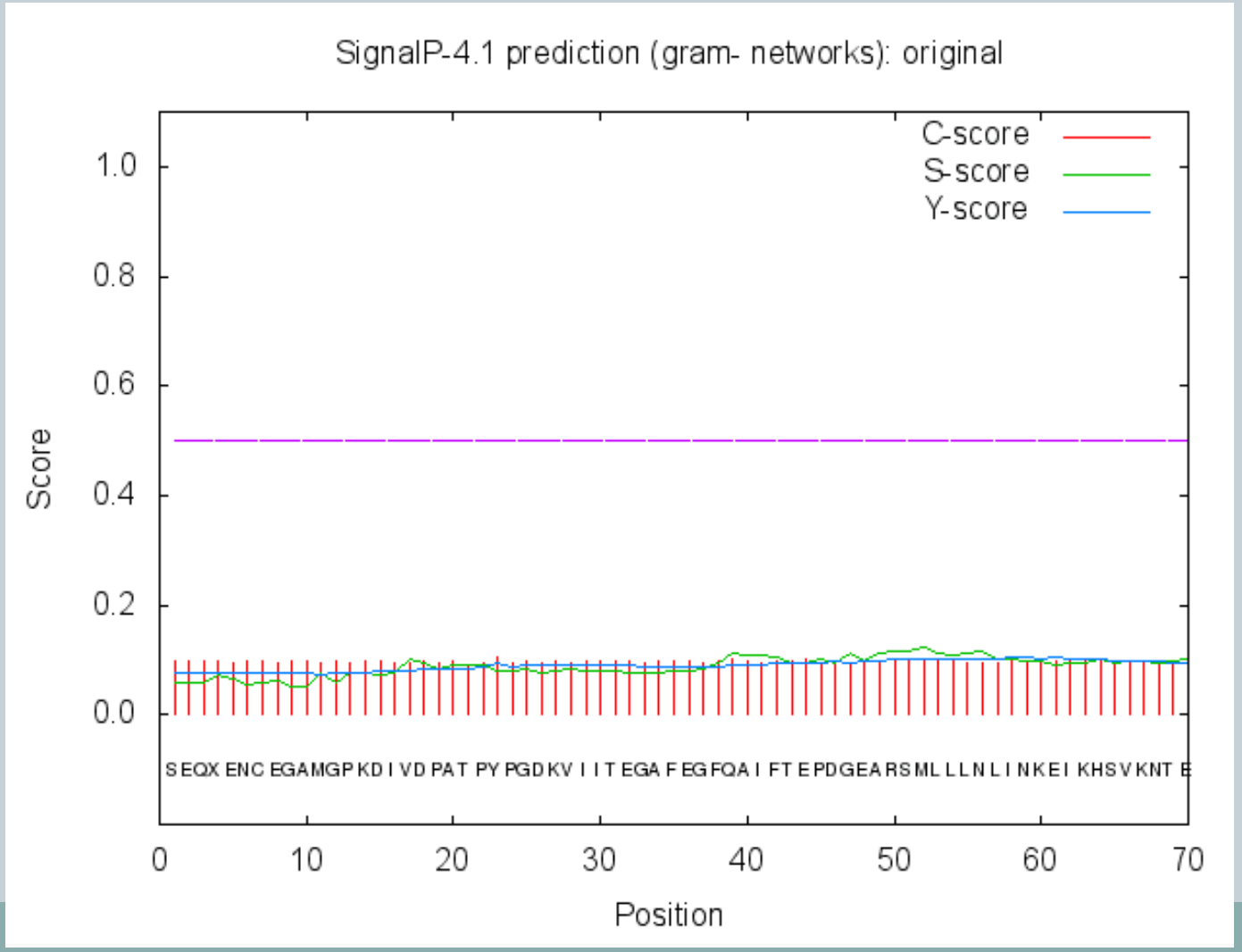
- 4.7e-226
- 8 sites



Smart预测


SMART
SETUP
FAQ
ABOUT
GLOSSARY
WHAT'S NEW
FEEDBACK
keywords...
Search SMART
NGN
KOW
NGN
KOW
NGN
KOW
NGN
KOW
0
100
200
300
400
500
600
700
800
900
100

Signalp分析


SignalP-4.1 prediction (gram- networks): original
C-score
S-score
Y-score
Score
1.0
0.8
0.6
0.4
0.2
0.0
SEQX ENC EGAMGP KD I VD PAT PY PGD KV I I T EGA F EG FQA I FT E PDGEA R S ML L L N L I N K E I KHS V K N T E
0
10
20
30
40
50
60
70
Position

targetP1.1预测亚细胞定位

```
### targetp v1.1 prediction results ###################################
Number of query sequences:  1
Cleavage site predictions not included.
Using NON-PLANT networks.

Name                    Len              mTP     SP  other  Loc   RC
----------------------------------------------------------------------
original                 74            0.060  0.072  0.940    _     1
----------------------------------------------------------------------
cutoff                                 0.000  0.000  0.000
```

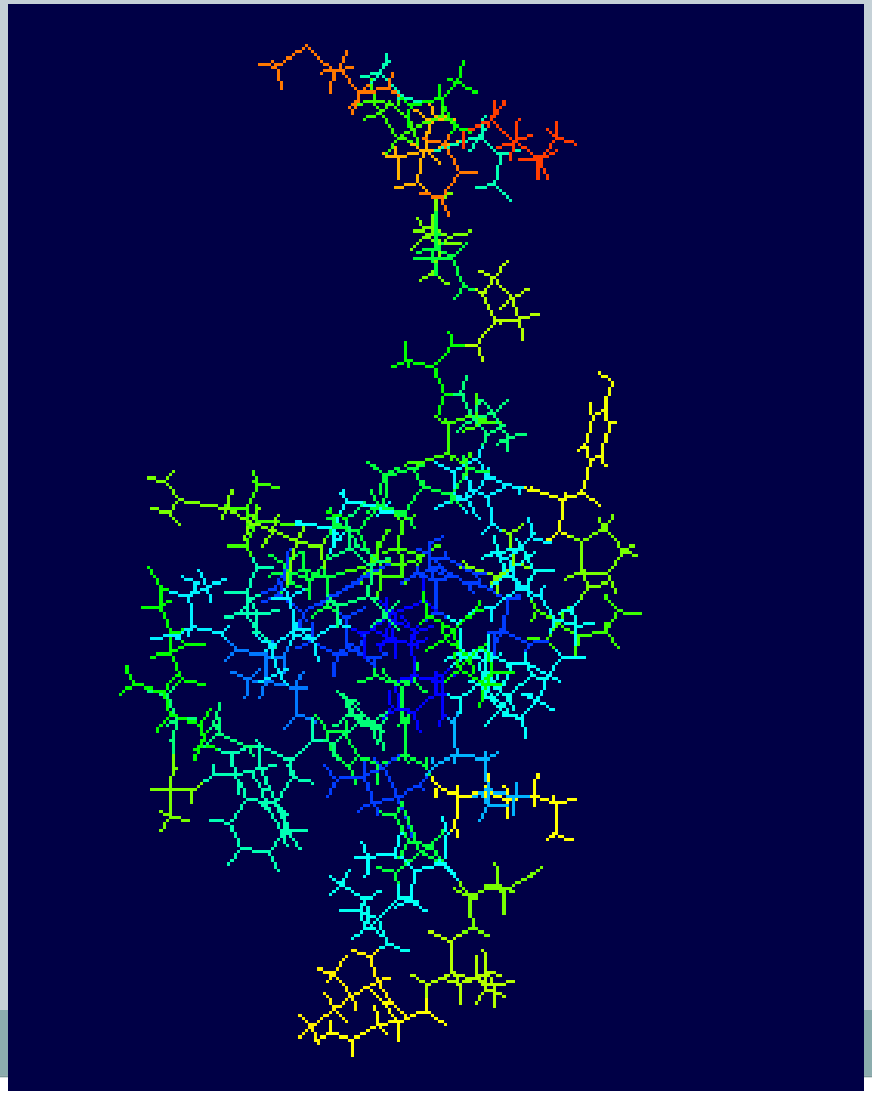
Pepstats分析氨基酸组成

PEPSTATS of SEQUENCE from 1 to 66

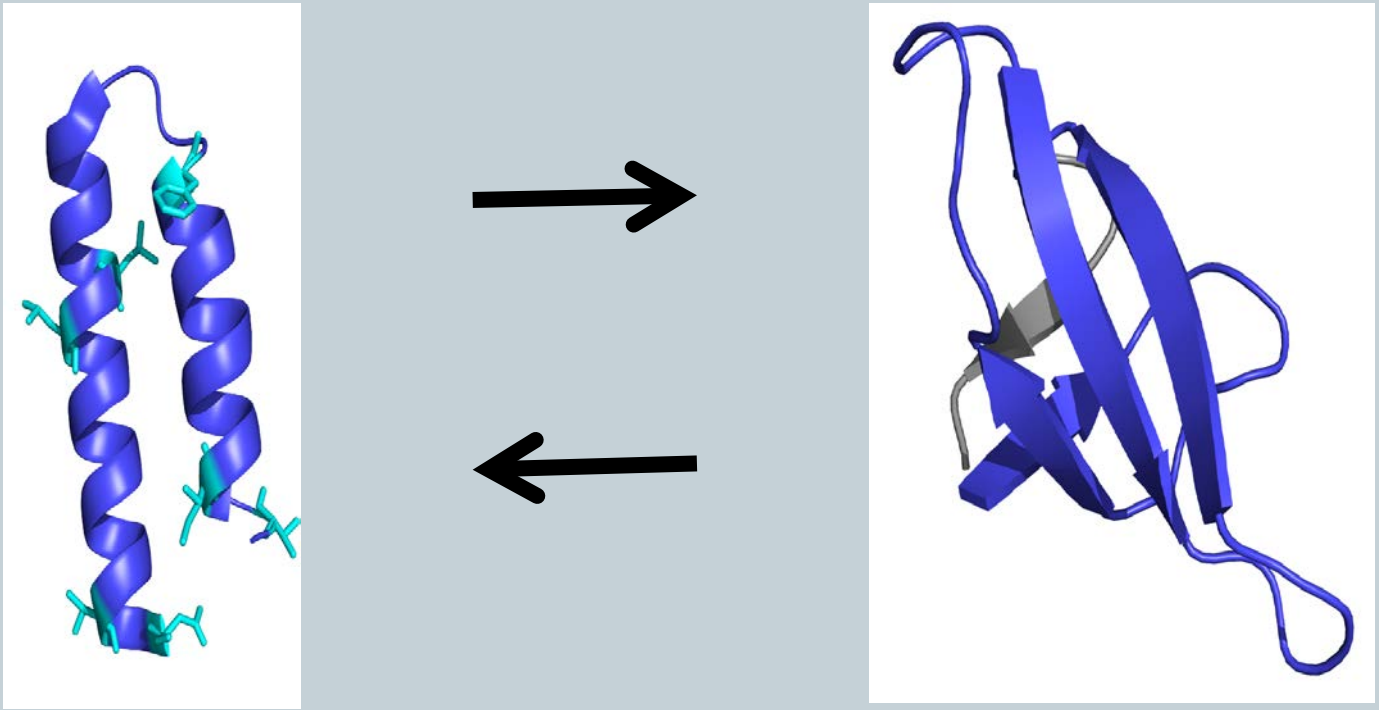
- Molecular weight = 7260.33 Residues = 66
- Average Residue Weight = 110.005 Charge = -1.5
- Isoelectric Point = 4.9874
- A280 Molar Extinction Coefficient = 1280
- A280 Extinction Coefficient 1mg/ml = 0.18
- Improbability of expression in inclusion bodies = 0.559

property	residues	number	Mol%
Tiny	A+C+S+T	17	25.758
Small	A+B+C+D+G+N+P+S+T+V	32	48.485
Aliphatic	A+I+L+Y	19	28.788
Aromatic	F+H+W+Y	6	9.091
Non-polar	A+C+F+G+I+L+M+P+V+W+Y	37	56.061
Polar	D+E+H+K+N+Q+R+S+T+Z	29	43.939
Charged	B+D+E+H+K+R+Z	19	28.788
Basic	H+K+R	9	13.636
Acidic	B+D+E+Z	10	15.152

Swiss-PDBViewer分析

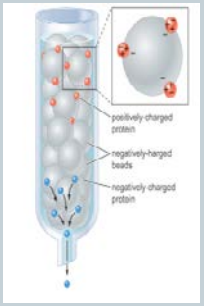


Point mutations that switch the two conformations?



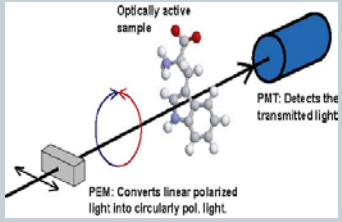
Our approach

Design mutations

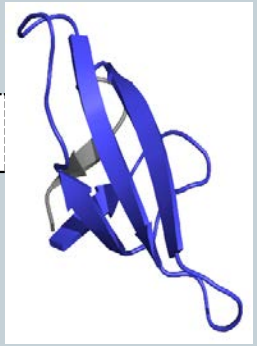


Purify proteins

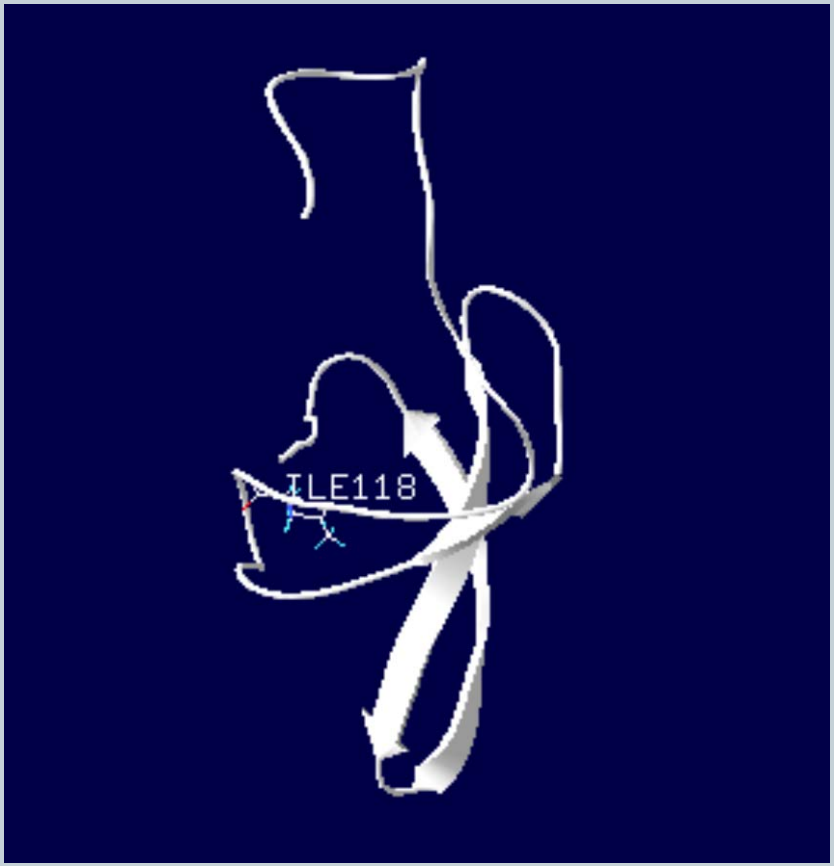
Circular Dichroism

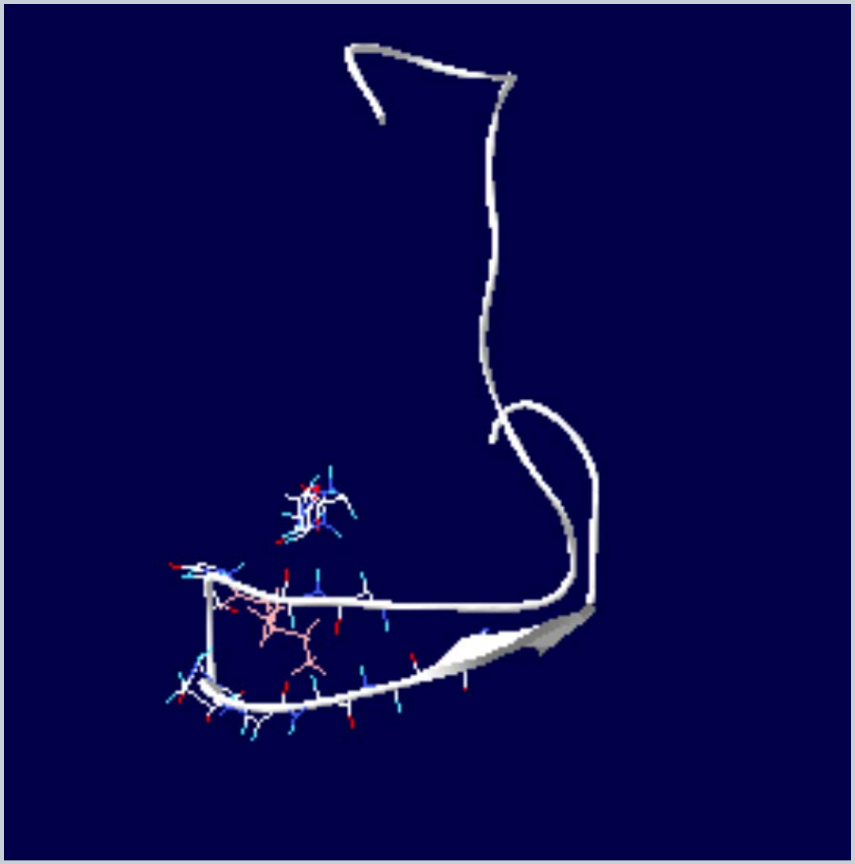


Structure solving



Example of one point mutation



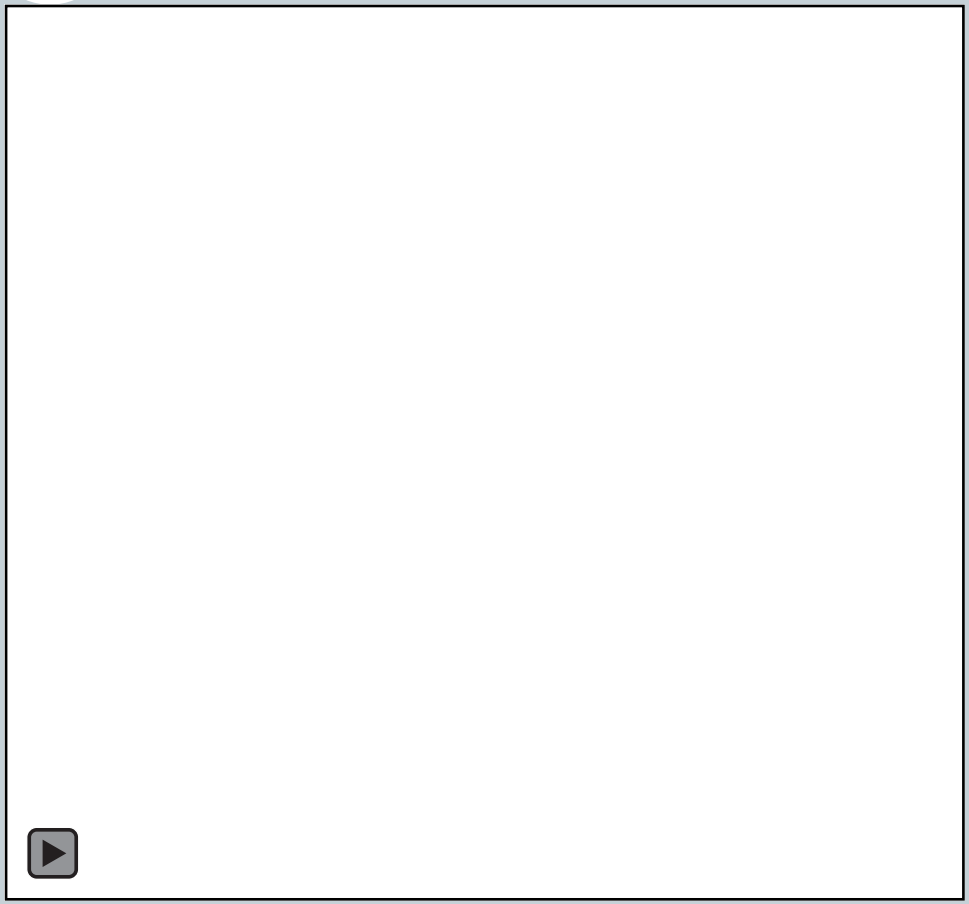


More point mutations to be explored

- P133G
- G121L
- G125L

Simulation with three mutations (G121L, G125L, P133A)

- Temperature: 300K;
- Pressure: 1atm;
- Box: cube(1nm^3);
- Solution model: SPC;
- Time: 20ns (2fs/step);
- Force field: OPLS/AA.



No obvious conformational change.

Thank you!