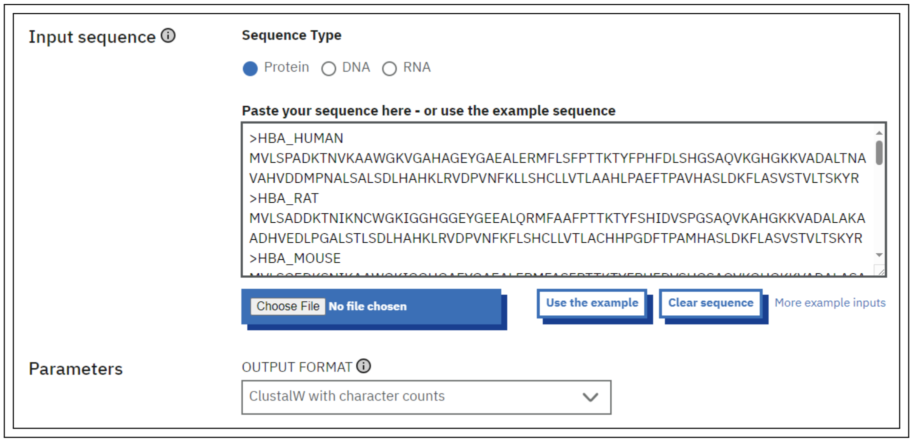
Practical 6

Question 1. Using AL2CO server (http://prodata.swmed.edu/al2co/al2co.php), obtain the positional conservation scores from multiple sequence alignment (**MSA**) of given set of protein sequences (set1 and set2) using the methods given below:

(i) Unweighted frequency and entropy-based measure
(ii) Unweighted frequency and variance-based measure
(iii) Unweighted frequency and sum of pairs measure
(iv) Weighted frequency and variance-based measure
(v) Normalize the scores obtained with (i)

Sequences:
Set 1: P69905, P01946, P01942, P01966, P01958, P01959, P01965, P06635, P60529, P80043, P01980
Set 2: TPIS_HUMAN, TPIS_YEAST, TPIS_GRAGA, TPIS_TRYCR, TPIS_MAIZE, TPIS_MOUSE, TPIS_DROME, TPIS_RABIT, TPIS_CAEEL

Solution. Firstly, I need to compute the Multiple Sequence Alignment of the given sets of sequences. It is because it is the aligned sequences itself that is passed as input to the **AL2CO** server, for it to compute conservation scores. So, below is an image of the input given to the **Clustal Omega** to compute the Multiple Sequence Alignment of set 1 and set 2. The outputs of the Multiple Sequence Alignment have been enclosed ahead of these input images.



Input sequence

Sequence Type

● Protein ○ DNA ○ RNA

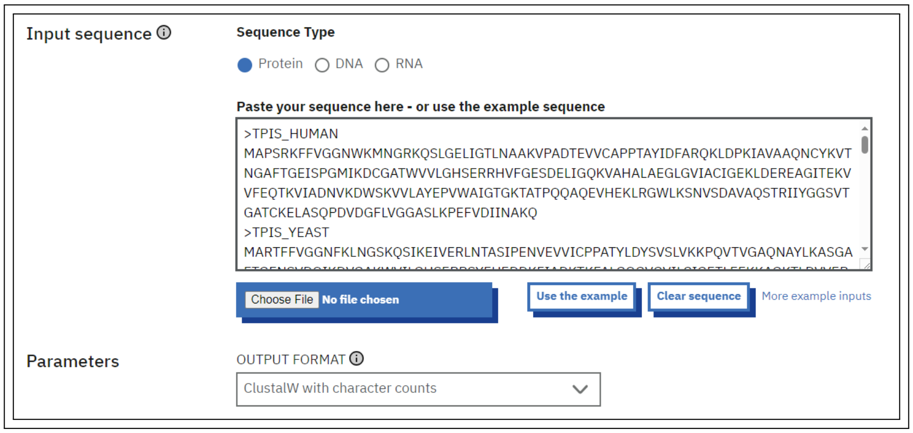
Paste your sequence here - or use the example sequence

```
>HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
>HBA_RAT
MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHIDVSPGSAQVKAHGKKVADALAKA
ADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVSTVLTSKYR
>HBA_MOUSE
```

Parameters

OUTPUT FORMAT: ClustalW with character counts

FIGURE 1. Input sequences to the CLUSTAL OMEGA from set 1



Input sequence

Sequence Type

● Protein ○ DNA ○ RNA

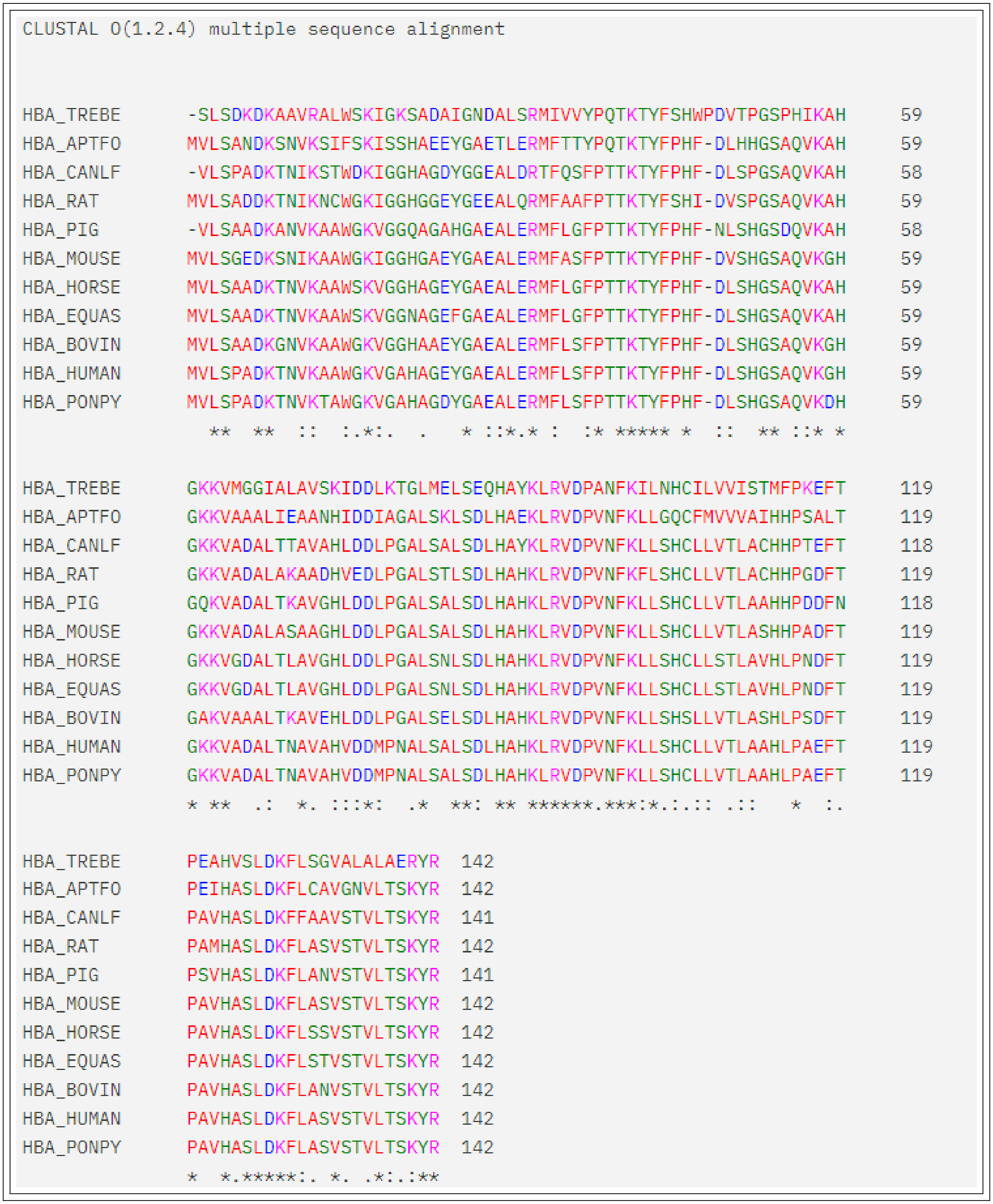
Paste your sequence here - or use the example sequence

```
>TPIS_HUMAN
MAPSRKFFVGGNWKMNGRKQSLGELIGTLNAAKVPADTEVVCAPPTAYIDFARQKLDPKIAVAAQNCYKVT
NGAFTGEISPGMIKDCGATWVVLGHSERRHVFGESDELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKV
VFEQTKVIADNVKDWSKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGWLKSNVSDAVAQSTRIIYGGSVT
GATCKELASQPDVDGFLVGGASLKPEFVDIINAKQ
>TPIS_YEAST
MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPENVEVVICPPATYLDYSVSLVKKPQVTVGAQNAYLKASGA
```

Parameters

OUTPUT FORMAT: ClustalW with character counts

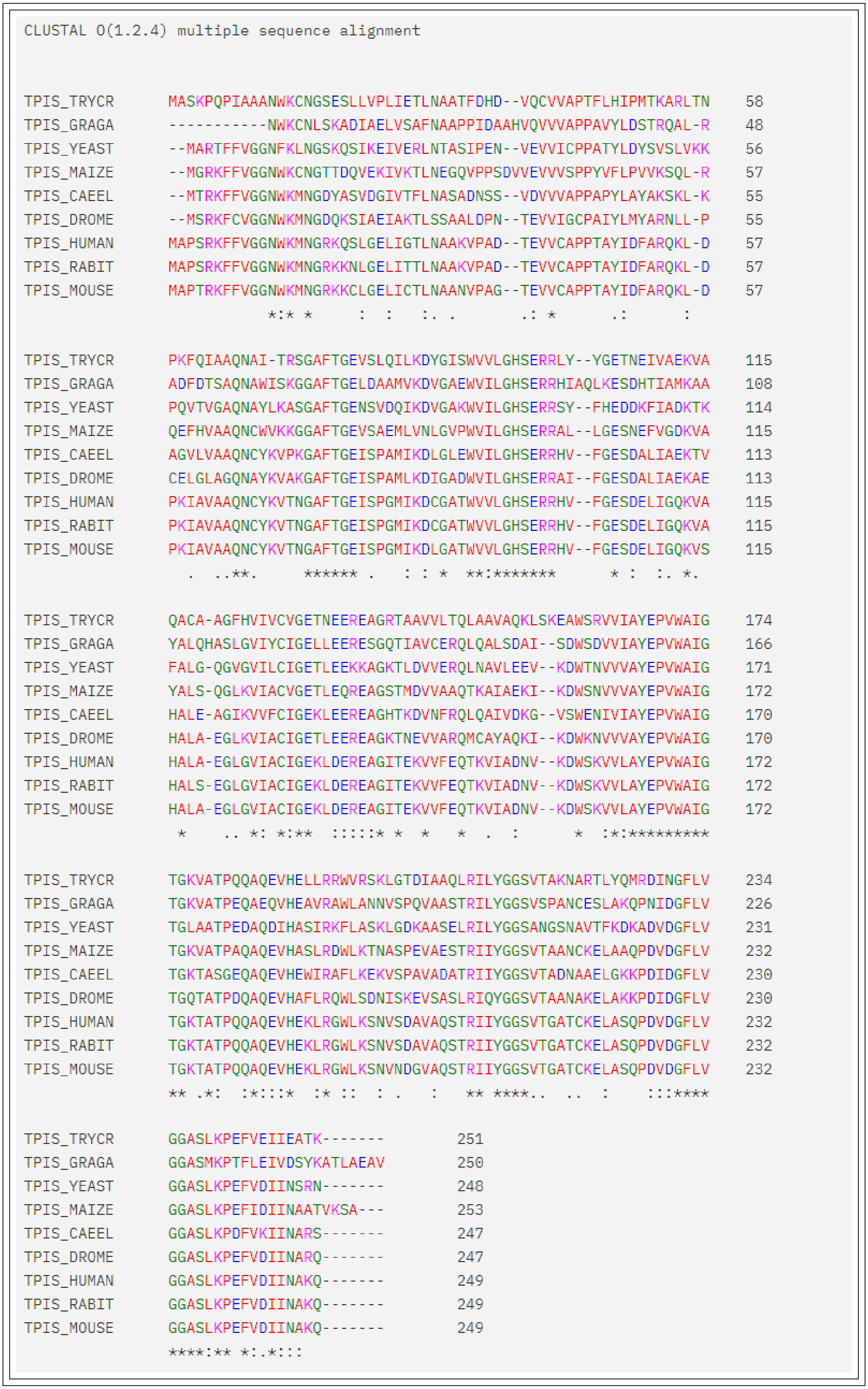
FIGURE 2. Input sequences to the CLUSTAL OMEGA from set 2



```
CLUSTAL O(1.2.4) multiple sequence alignment


HBA_TREBE      -SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAH     59
HBA_APTFO      MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-DLHHGSAQVKAH     59
HBA_CANLF      -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-DLSPGSAQVKAH     58
HBA_RAT        MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-DVSPGSAQVKAH     59
HBA_PIG        -VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-NLSHGSDQVKAH     58
HBA_MOUSE      MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-DVSHGSAQVKGH     59
HBA_HORSE      MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_EQUAS      MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_BOVIN      MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_HUMAN      MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_PONPY      MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKDH     59
                 **  **  ::  :.*:.  .   * ::*.* :  :* ***** *  ::  ** ::* *

HBA_TREBE      GKKVMGGIALAVSKIDDLKTGLMELSEQHAYKLRVDPANFKILNHCILVVISTMFPKEFT     119
HBA_APTFO      GKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALT     119
HBA_CANLF      GKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFT     118
HBA_RAT        GKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFT     119
HBA_PIG        GQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFN     118
HBA_MOUSE      GKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFT     119
HBA_HORSE      GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT     119
HBA_EQUAS      GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT     119
HBA_BOVIN      GAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT     119
HBA_HUMAN      GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT     119
HBA_PONPY      GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT     119
               * **  .:  *. :::*:  .*  **: ** ******.***:*.:.:: .::   *  :.

HBA_TREBE      PEAHVSLDKFLSGVALALAERYR  142
HBA_APTFO      PEIHASLDKFLCAVGNVLTSKYR  142
HBA_CANLF      PAVHASLDKFFAAVSTVLTSKYR  141
HBA_RAT        PAMHASLDKFLASVSTVLTSKYR  142
HBA_PIG        PSVHASLDKFLANVSTVLTSKYR  141
HBA_MOUSE      PAVHASLDKFLASVSTVLTSKYR  142
HBA_HORSE      PAVHASLDKFLSSVSTVLTSKYR  142
HBA_EQUAS      PAVHASLDKFLSTVSTVLTSKYR  142
HBA_BOVIN      PAVHASLDKFLANVSTVLTSKYR  142
HBA_HUMAN      PAVHASLDKFLASVSTVLTSKYR  142
HBA_PONPY      PAVHASLDKFLASVSTVLTSKYR  142
               *  *.*****:. *. .*:.:**
```

Figure 3. Output MSA of the CLUSTAL OMEGA from set 1

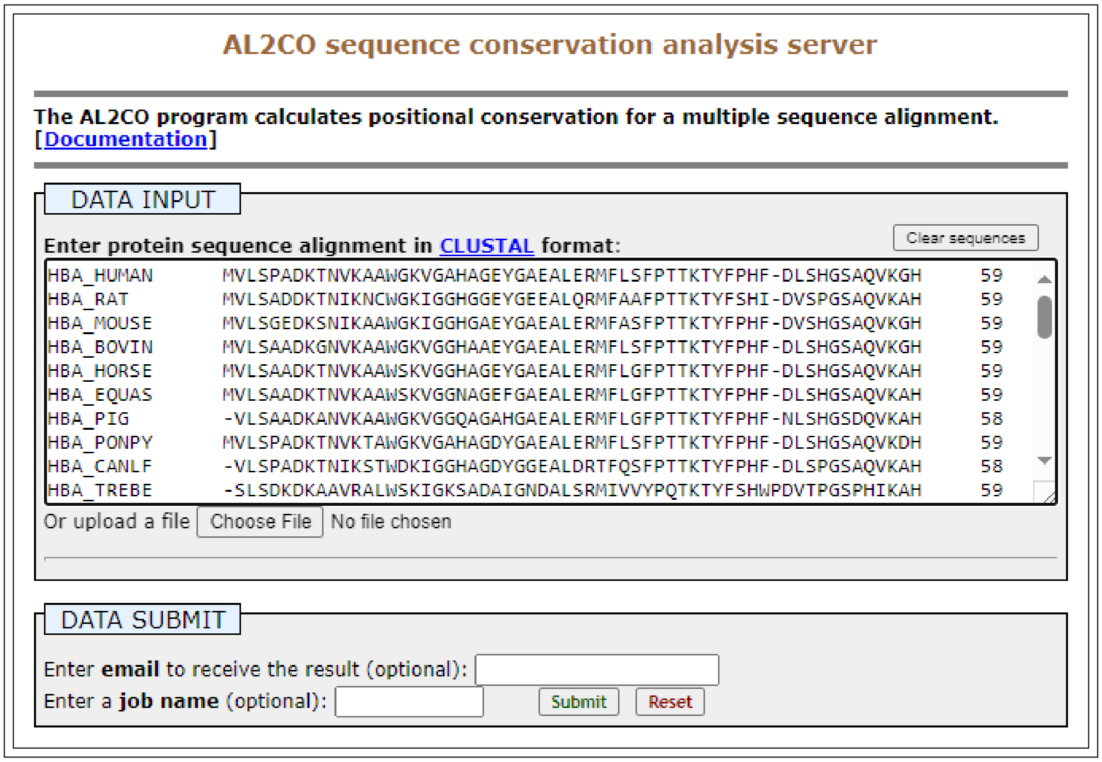
- The above image shows the multiple sequence alignment for the proteins enclosed in set 1. They are a total of 11 in number.
- The below image shows the multiple sequence alignment for the proteins enclosed in set 2. They are a total of 9 in number.
- These aligned sequences are those that will be passed to the AL2CO server tool to compute the conservation scores for these sets of proteins.



```
CLUSTAL O(1.2.4) multiple sequence alignment


TPIS_TRYCR      MASKPQPIAAANWKCNGSESLLVPLIETLNAATFDHD--VQCVVAPTFLHIPMTKARLTN    58
TPIS_GRAGA      -----------NWKCNLSKADIAELVSAFNAAPPIDAAHVQVVVAPPAVYLDSTRQAL-R    48
TPIS_YEAST      --MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPEN--VEVVICPPATYLDYSVSLVKK    56
TPIS_MAIZE      --MGRKFFVGGNWKCNGTTDQVEKIVKTLNEGQVPPSDVVEVVVSPPYVFLPVVKSQL-R    57
TPIS_CAEEL      --MTRKFFVGGNWKMNGDYASVDGIVTFLNASADNSS--VDVVVAPPAPYLAYAKSKL-K    55
TPIS_DROME      --MSRKFCVGGNWKMNGDQKSIAEIAKTLSSAALDPN--TEVVIGCPAIYLMYARNLL-P    55
TPIS_HUMAN      MAPSRKFFVGGNWKMNGRKQSLGELIGTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-D    57
TPIS_RABIT      MAPSRKFFVGGNWKMNGRKKNLGELITTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-D    57
TPIS_MOUSE      MAPTRKFFVGGNWKMNGRKKCLGELICTLNAANVPAG--TEVVCAPPTAYIDFARQKL-D    57
                           *:* *     :  :   :. .       .: *      .:      :

TPIS_TRYCR      PKFQIAAQNAI-TRSGAFTGEVSLQILKDYGISWVVLGHSERRLY--YGETNEIVAEKVA    115
TPIS_GRAGA      ADFDTSAQNAWISKGGAFTGELDAAMVKDVGAEWVILGHSERRHIAQLKESDHTIAMKAA    108
TPIS_YEAST      PQVTVGAQNAYLKASGAFTGENSVDQIKDVGAKWVILGHSERRSY--FHEDDKFIADKTK    114
TPIS_MAIZE      QEFHVAAQNCWVKKGGAFTGEVSAEMLVNLGVPWVILGHSERRAL--LGESNEFVGDKVA    115
TPIS_CAEEL      AGVLVAAQNCYKVPKGAFTGEISPAMIKDLGLEWVILGHSERRHV--FGESDALIAEKTV    113
TPIS_DROME      CELGLAGQNAYKVAKGAFTGEISPAMLKDIGADWVILGHSERRAI--FGESDALIAEKAE    113
TPIS_HUMAN      PKIAVAAQNCYKVTNGAFTGEISPGMIKDCGATWVVLGHSERRHV--FGESDELIGQKVA    115
TPIS_RABIT      PKIAVAAQNCYKVTNGAFTGEISPGMIKDCGATWVVLGHSERRHV--FGESDELIGQKVA    115
TPIS_MOUSE      PKIAVAAQNCYKVTNGAFTGEISPGMIKDLGATWVVLGHSERRHV--FGESDELIGQKVS    115
                  .  ..**.     ****** .   : : *  **:*******      * :  :. *.

TPIS_TRYCR      QACA-AGFHVIVCVGETNEEREAGRTAAVVLTQLAAVAQKLSKEAWSRVVIAYEPVWAIG    174
TPIS_GRAGA      YALQHASLGVIYCIGELLEERESGQTIAVCERQLQALSDAI--SDWSDVVIAYEPVWAIG    166
TPIS_YEAST      FALG-QGVGVILCIGETLEEKKAGKTLDVVERQLNAVLEEV--KDWTNVVVAYEPVWAIG    171
TPIS_MAIZE      YALS-QGLKVIACVGETLEQREAGSTMDVVAAQTKAIAEKI--KDWSNVVVAYEPVWAIG    172
TPIS_CAEEL      HALE-AGIKVVFCIGEKLEEREAGHTKDVNFRQLQAIVDKG--VSWENIVIAYEPVWAIG    170
TPIS_DROME      HALA-EGLKVIACIGETLEEREAGKTNEVVARQMCAYAQKI--KDWKNVVVAYEPVWAIG    170
TPIS_HUMAN      HALA-EGLGVIACIGEKLDEREAGITEKVVFEQTKVIADNV--KDWSKVVLAYEPVWAIG    172
TPIS_RABIT      HALS-EGLGVIACIGEKLDEREAGITEKVVFEQTKVIADNV--KDWSKVVLAYEPVWAIG    172
TPIS_MOUSE      HALA-EGLGVIACIGEKLDEREAGITEKVVFEQTKVIADNV--KDWSKVVLAYEPVWAIG    172
                 *    .. *: *:**  :::::* *  *   *  .  :      *  :*:*********

TPIS_TRYCR      TGKVATPQQAQEVHELLRRWVRSKLGTDIAAQLRILYGGSVTAKNARTLYQMRDINGFLV    234
TPIS_GRAGA      TGKVATPEQAEQVHEAVRAWLANNVSPQVAASTRILYGGSVSPANCESLAKQPNIDGFLV    226
TPIS_YEAST      TGLAATPEDAQDIHASIRKFLASKLGDKAASELRILYGGSANGSNAVTFKDKADVDGFLV    231
TPIS_MAIZE      TGKVATPAQAQEVHASLRDWLKTNASPEVAESTRIIYGGSVTAANCKELAAQPDVDGFLV    232
TPIS_CAEEL      TGKTASGEQAQEVHEWIRAFLKEKVSPAVADATRIIYGGSVTADNAAELGKKPDIDGFLV    230
TPIS_DROME      TGQTATPDQAQEVHAFLRQWLSDNISKEVSASLRIQYGGSVTAANAKELAKKPDIDGFLV    230
TPIS_HUMAN      TGKTATPQQAQEVHEKLRGWLKSNVSDAVAQSTRIIYGGSVTGATCKELASQPDVDGFLV    232
TPIS_RABIT      TGKTATPQQAQEVHEKLRGWLKSNVSDAVAQSTRIIYGGSVTGATCKELASQPDVDGFLV    232
TPIS_MOUSE      TGKTATPQQAQEVHEKLRGWLKSNVNDGVAQSTRIIYGGSVTGATCKELASQPDVDGFLV    232
                ** .*:  :*:::*  :* ::  : .   :   ** ****..  ..  :    :::****

TPIS_TRYCR      GGASLKPEFVEIIEATK-------        251
TPIS_GRAGA      GGASMKPTFLEIVDSYKATLAEAV        250
TPIS_YEAST      GGASLKPEFVDIINSRN-------        248
TPIS_MAIZE      GGASLKPEFIDIINAATVKSA---        253
TPIS_CAEEL      GGASLKPDFVKIINARS-------        247
TPIS_DROME      GGASLKPEFVDIINARQ-------        247
TPIS_HUMAN      GGASLKPEFVDIINAKQ-------        249
TPIS_RABIT      GGASLKPEFVDIINAKQ-------        249
TPIS_MOUSE      GGASLKPEFVDIINAKQ-------        249
                ****:** *:.*:::
```

Figure 4. Output MSA of the CLUSTAL OMEGA from set 2

The general format of providing the input to the **AL2CO** server is given below. It includes the syntax in which the input must be provided to the AL2CO. The aligned sequences should only be passed to the AL2Co server. The parameters involved in submitting the input have been enclosed later for each category mentioned above:



AL2CO sequence conservation analysis server

The AL2CO program calculates positional conservation for a multiple sequence alignment. [Documentation]

DATA INPUT

Enter protein sequence alignment in CLUSTAL format:

```
HBA_HUMAN      MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_RAT        MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-DVSPGSAQVKAH     59
HBA_MOUSE      MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-DVSHGSAQVKGH     59
HBA_BOVIN      MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_HORSE      MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_EQUAS      MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_PIG        -VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-NLSHGSDQVKAH     58
HBA_PONPY      MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKDH     59
HBA_CANLF      -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-DLSPGSAQVKAH     58
HBA_TREBE      -SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAH     59
```

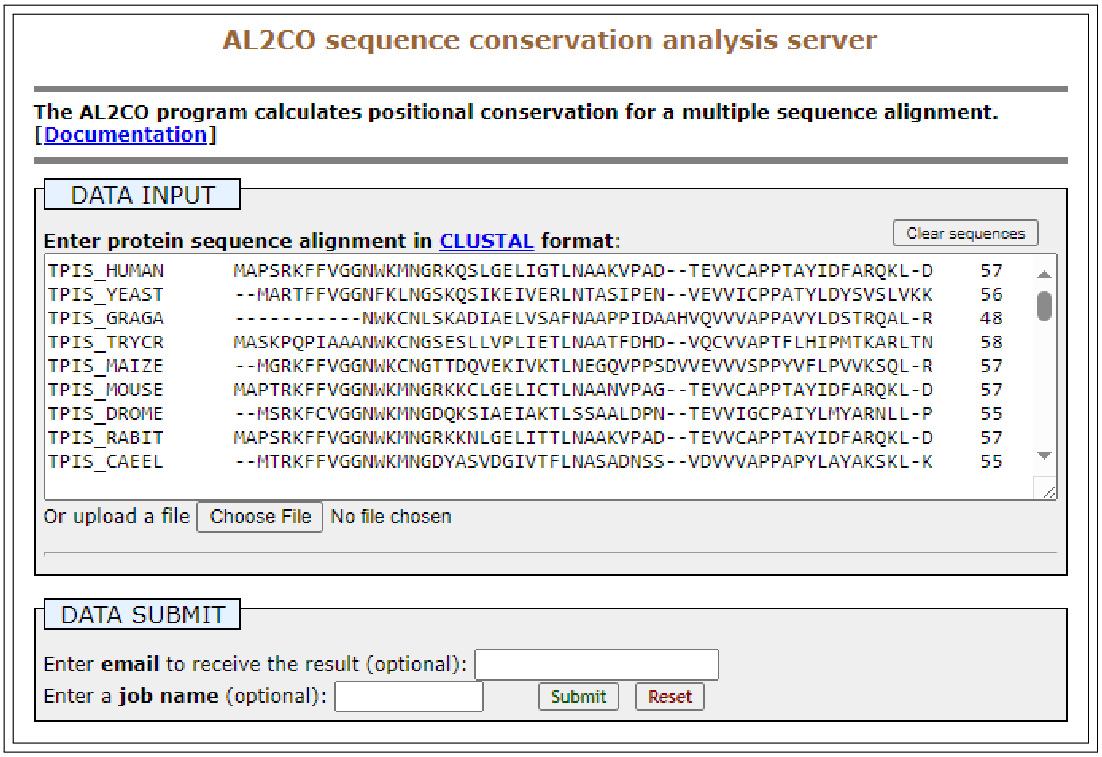
Or upload a file Choose File No file chosen

DATA SUBMIT

Enter **email** to receive the result (optional):

Enter a **job name** (optional):

FIGURE 5. Input sequences to the AL2CO server from set 1



AL2CO sequence conservation analysis server

The AL2CO program calculates positional conservation for a multiple sequence alignment. [Documentation]

DATA INPUT

Enter protein sequence alignment in CLUSTAL format:

```
TPIS_HUMAN      MAPSRKFFVGGNWKMNGRKQSLGELIGTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-D     57
TPIS_YEAST      --MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPEN--VEVVICPPATYLDYSVSLVKK     56
TPIS_GRAGA      -----------NWKCNLSKADIAELVSAFNAAPPIDAAHVQVVVAPPAVYLDSTRQAL-R     48
TPIS_TRYCR      MASKPQPIAAANWKCNGSESLLVPLIETLNAATFDHD--VQCVVAPTFLHIPMTKARLTN     58
TPIS_MAIZE      --MGRKFFVGGNWKCNGTTDQVEKIVKTLNEGQVPPSDVVEVVVSPPYVFLPVVKSQL-R     57
TPIS_MOUSE      MAPTRKFFVGGNWKMNGRKKCLGELICTLNAANVPAG--TEVVCAPPTAYIDFARQKL-D     57
TPIS_DROME      --MSRKFCVGGNWKMNGDQKSIAEIAKTLSSAALDPN--TEVVIGCPAIYLMYARNLL-P     55
TPIS_RABIT      MAPSRKFFVGGNWKMNGRKKNLGELITTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-D     57
TPIS_CAEEL      --MTRKFFVGGNWKMNGDYASVDGIVTFLNASADNSS--VDVVVAPPAPYLAYAKSKL-K     55
```

Or upload a file Choose File No file chosen

DATA SUBMIT

Enter **email** to receive the result (optional):

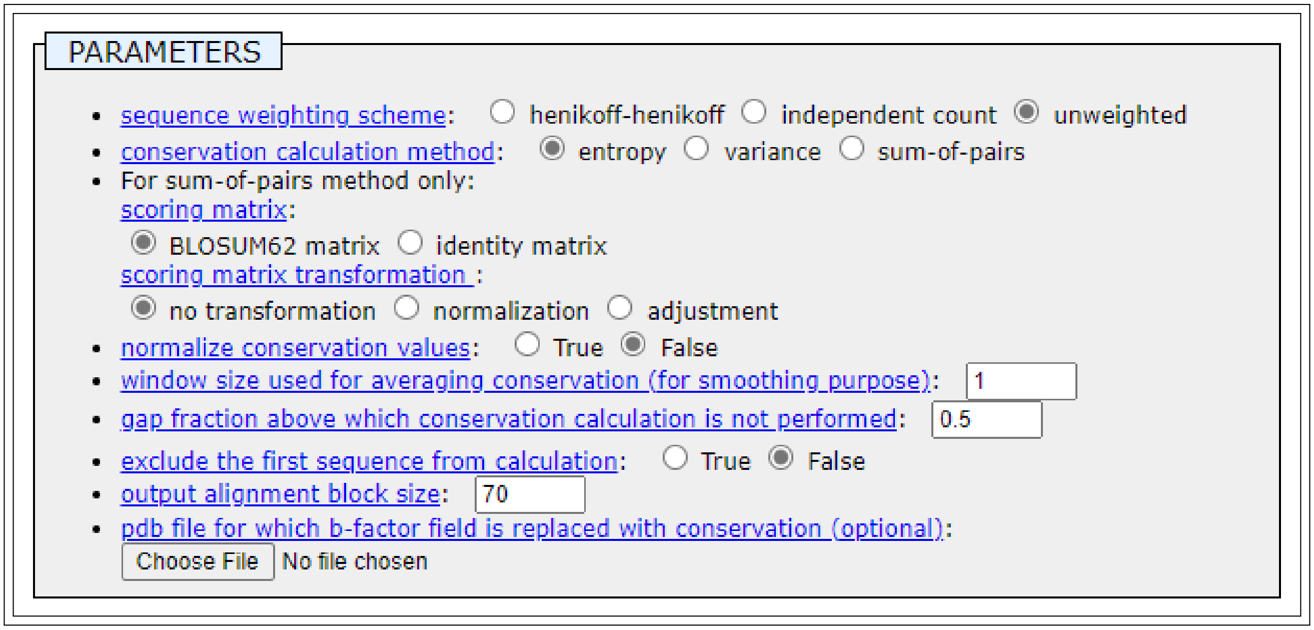
Enter a **job name** (optional):

FIGURE 6. Input sequences to the AL2CO server from set 2

(i) Unweighted frequency and entropy-based measure

The parameters for this scenario are enclosed in the image below. The important parameters are as follows:

- sequence weighting scheme
- conservation calculation method
- scoring matrix (for sum of pairs method only)
- scoring matrix transformation (for sum of pairs method only)
- normalize conservation values

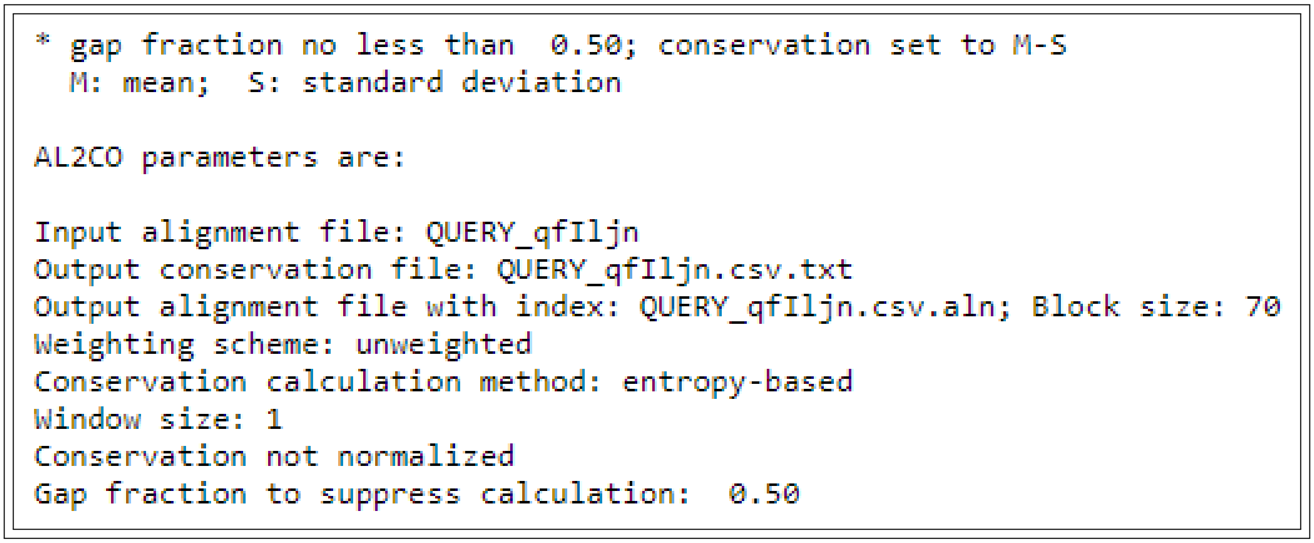


PARAMETERS

- sequence weighting scheme: ○ henikoff-henikoff ○ independent count ● unweighted
- conservation calculation method: ● entropy ○ variance ○ sum-of-pairs
- For sum-of-pairs method only:
  scoring matrix:
  ● BLOSUM62 matrix ○ identity matrix
  scoring matrix transformation :
  ● no transformation ○ normalization ○ adjustment
- normalize conservation values: ○ True ● False
- window size used for averaging conservation (for smoothing purpose): 1
- gap fraction above which conservation calculation is not performed: 0.5
- exclude the first sequence from calculation: ○ True ● False
- output alignment block size: 70
- pdb file for which b-factor field is replaced with conservation (optional):
  Choose File No file chosen

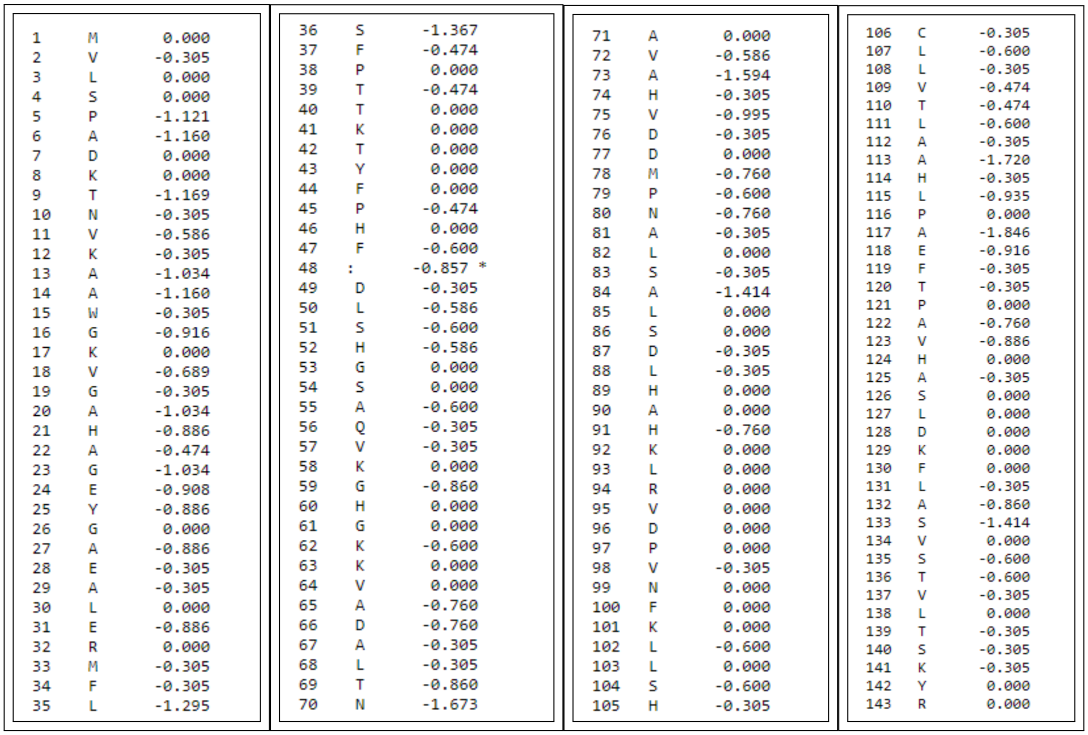
FIGURE 7. Parameters for the given scenario (i)

The AL2CO gives the list of positional conservation values and the alignment with integer conservation indices. The question asks to only calculate the positional conservation values.
The window of positional conservation values generates the following set of parameters which are the ones displayed in the above image.
It also displays some parameters taken into consideration to compute the desired positional conservation scores.



```
* gap fraction no less than  0.50; conservation set to M-S
  M: mean;  S: standard deviation

AL2CO parameters are:

Input alignment file: QUERY_qfIljn
Output conservation file: QUERY_qfIljn.csv.txt
Output alignment file with index: QUERY_qfIljn.csv.aln; Block size: 70
Weighting scheme: unweighted
Conservation calculation method: entropy-based
Window size: 1
Conservation not normalized
Gap fraction to suppress calculation:  0.50
```

FIGURE 8. Parameter output display in window



```
1     M     0.000
2     V    -0.305
3     L     0.000
4     S     0.000
5     P    -1.121
6     A    -1.160
7     D     0.000
8     K     0.000
9     T    -1.169
10    N    -0.305
11    V    -0.586
12    K    -0.305
13    A    -1.034
14    A    -1.160
15    W    -0.305
16    G    -0.916
17    K     0.000
18    V    -0.689
19    G    -0.305
20    A    -1.034
21    H    -0.886
22    A    -0.474
23    G    -1.034
24    E    -0.908
25    Y    -0.886
26    G     0.000
27    A    -0.886
28    E    -0.305
29    A    -0.305
30    L     0.000
31    E    -0.886
32    R     0.000
33    M    -0.305
34    F    -0.305
35    L    -1.295
```

(A) 1-35 pos

```
36    S    -1.367
37    F    -0.474
38    P     0.000
39    T    -0.474
40    T     0.000
41    K     0.000
42    T     0.000
43    Y     0.000
44    F     0.000
45    P    -0.474
46    H     0.000
47    F    -0.600
48   :    -0.857 *
49    D    -0.305
50    L    -0.586
51    S    -0.600
52    H    -0.586
53    G     0.000
54    S     0.000
55    A    -0.600
56    Q    -0.305
57    V    -0.305
58    K     0.000
59    G    -0.860
60    H     0.000
61    G     0.000
62    K    -0.600
63    K     0.000
64    V     0.000
65    A    -0.760
66    D    -0.760
67    A    -0.305
68    L    -0.305
69    T    -0.860
70    N    -1.673
```

(B) 36-70 pos

```
71    A     0.000
72    V    -0.586
73    A    -1.594
74    H    -0.305
75    V    -0.995
76    D    -0.305
77    D     0.000
78    M    -0.760
79    P    -0.600
80    N    -0.760
81    A    -0.305
82    L     0.000
83    S    -0.305
84    A    -1.414
85    L     0.000
86    S     0.000
87    D    -0.305
88    L    -0.305
89    H     0.000
90    A     0.000
91    H    -0.760
92    K     0.000
93    L     0.000
94    R     0.000
95    V     0.000
96    D     0.000
97    P     0.000
98    V    -0.305
99    N     0.000
100   F     0.000
101   K     0.000
102   L    -0.600
103   L     0.000
104   S    -0.600
105   H    -0.305
```

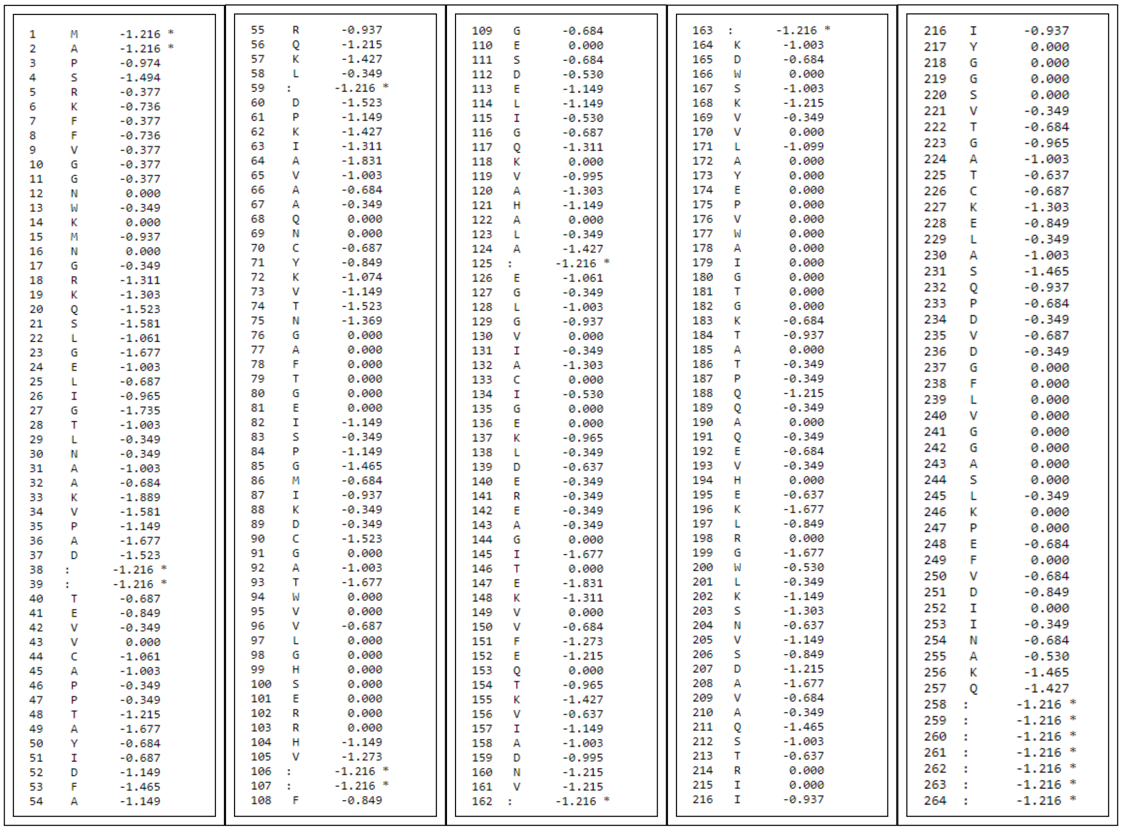
(C) 71-105 pos

```
106   C    -0.305
107   L    -0.600
108   L    -0.305
109   V    -0.474
110   T    -0.474
111   L    -0.600
112   A    -0.305
113   A    -1.720
114   H    -0.305
115   L    -0.935
116   P     0.000
117   A    -1.846
118   E    -0.916
119   F    -0.305
120   T    -0.305
121   P     0.000
122   A    -0.760
123   V    -0.886
124   H     0.000
125   A    -0.305
126   S     0.000
127   L     0.000
128   D     0.000
129   K     0.000
130   F     0.000
131   L    -0.305
132   A    -0.860
133   S    -1.414
134   V     0.000
135   S    -0.600
136   T    -0.600
137   V    -0.305
138   L     0.000
139   T    -0.305
140   S    -0.305
141   K    -0.305
142   Y     0.000
143   R     0.000
```

(D) 106-143 pos

FIGURE 9. Positional conservation scores for set 1

Above is the image of the positional conservation scores for the set 1 proteins. Below is the image of the positional conservation scores for the set 2 proteins.



```
1     M     -1.216 *
2     A     -1.216 *
3     P     -0.974
4     S     -1.494
5     R     -0.377
6     K     -0.736
7     F     -0.377
8     F     -0.736
9     V     -0.377
10    G     -0.377
11    G     -0.377
12    N      0.000
13    W     -0.349
14    K      0.000
15    M     -0.937
16    N      0.000
17    G     -0.349
18    R     -1.311
19    K     -1.303
20    Q     -1.523
21    S     -1.581
22    L     -1.061
23    G     -1.677
24    E     -1.003
25    L     -0.687
26    I     -0.965
27    G     -1.735
28    T     -1.003
29    L     -0.349
30    N     -0.349
31    A     -1.003
32    A     -0.684
33    K     -1.889
34    V     -1.581
35    P     -1.149
36    A     -1.677
37    D     -1.523
38   :     -1.216 *
39   :     -1.216 *
40    T     -0.687
41    E     -0.849
42    V     -0.349
43    V      0.000
44    C     -1.061
45    A     -1.003
46    P     -0.349
47    P     -0.349
48    T     -1.215
49    A     -1.677
50    Y     -0.684
51    I     -0.687
52    D     -1.149
53    F     -1.465
54    A     -1.149
```

(A) 1-54 pos

```
55    R     -0.937
56    Q     -1.215
57    K     -1.427
58    L     -0.349
59   :     -1.216 *
60    D     -1.523
61    P     -1.149
62    K     -1.427
63    I     -1.311
64    A     -1.831
65    V     -1.003
66    A     -0.684
67    A     -0.349
68    Q      0.000
69    N      0.000
70    C     -0.687
71    Y     -0.849
72    K     -1.074
73    V     -1.149
74    T     -1.523
75    N     -1.369
76    G      0.000
77    A      0.000
78    F      0.000
79    T      0.000
80    G      0.000
81    E      0.000
82    I     -1.149
83    S     -0.349
84    P     -1.149
85    G     -1.465
86    M     -0.684
87    I     -0.937
88    K     -0.349
89    D     -0.349
90    C     -1.523
91    G      0.000
92    A     -1.003
93    T     -1.677
94    W      0.000
95    V      0.000
96    V     -0.687
97    L      0.000
98    G      0.000
99    H      0.000
100   S      0.000
101   E      0.000
102   R      0.000
103   R      0.000
104   H     -1.149
105   V     -1.273
106  :     -1.216 *
107  :     -1.216 *
108   F     -0.849
```

(B) 55-108 pos

```
109   G     -0.684
110   E      0.000
111   S     -0.684
112   D     -0.530
113   E     -1.149
114   L     -1.149
115   I     -0.530
116   G     -0.687
117   Q     -1.311
118   K      0.000
119   V     -0.995
120   A     -1.303
121   H     -1.149
122   A      0.000
123   L     -0.349
124   A     -1.427
125  :     -1.216 *
126   E     -1.061
127   G     -0.349
128   L     -1.003
129   G     -0.937
130   V      0.000
131   I     -0.349
132   A     -1.303
133   C      0.000
134   I     -0.530
135   G      0.000
136   E      0.000
137   K     -0.965
138   L     -0.349
139   D     -0.637
140   E     -0.349
141   R     -0.349
142   E     -0.349
143   A     -0.349
144   G      0.000
145   I     -1.677
146   T      0.000
147   E     -1.831
148   K     -1.311
149   V      0.000
150   V     -0.684
151   F     -1.273
152   E     -1.215
153   Q      0.000
154   T     -0.965
155   K     -1.427
156   V     -0.637
157   I     -1.149
158   A     -1.003
159   D     -0.995
160   N     -1.215
161   V     -1.215
162  :     -1.216 *
```

(C) 109-162 pos

```
163  :     -1.216 *
164   K     -1.003
165   D     -0.684
166   W      0.000
167   S     -1.003
168   K     -1.215
169   V     -0.349
170   V      0.000
171   L     -1.099
172   A      0.000
173   Y      0.000
174   E      0.000
175   P      0.000
176   V      0.000
177   W      0.000
178   A      0.000
179   I      0.000
180   G      0.000
181   T      0.000
182   G      0.000
183   K     -0.684
184   T     -0.937
185   A      0.000
186   T     -0.349
187   P     -0.349
188   Q     -1.215
189   Q     -0.349
190   A      0.000
191   Q     -0.349
192   E     -0.684
193   V     -0.349
194   H      0.000
195   E     -0.637
196   K     -1.677
197   L     -0.849
198   R      0.000
199   G     -1.677
200   W     -0.530
201   L     -0.349
202   K     -1.149
203   S     -1.303
204   N     -0.637
205   V     -1.149
206   S     -0.849
207   D     -1.215
208   A     -1.677
209   V     -0.684
210   A     -0.349
211   Q     -1.465
212   S     -1.003
213   T     -0.637
214   R      0.000
215   I      0.000
216   I     -0.937
```

(D) 163-216 pos

```
216   I      -0.937
217   Y       0.000
218   G       0.000
219   G       0.000
220   S       0.000
221   V      -0.349
222   T      -0.684
223   G      -0.965
224   A      -1.003
225   T      -0.637
226   C      -0.687
227   K      -1.303
228   E      -0.849
229   L      -0.349
230   A      -1.003
231   S      -1.465
232   Q      -0.937
233   P      -0.684
234   D      -0.349
235   V      -0.687
236   D      -0.349
237   G       0.000
238   F       0.000
239   L       0.000
240   V       0.000
241   G       0.000
242   G       0.000
243   A       0.000
244   S       0.000
245   L      -0.349
246   K       0.000
247   P       0.000
248   E      -0.684
249   F       0.000
250   V      -0.684
251   D      -0.849
252   I       0.000
253   I      -0.349
254   N      -0.684
255   A      -0.530
256   K      -1.465
257   Q      -1.427
258   :     -1.216 *
259   :     -1.216 *
260   :     -1.216 *
261   :     -1.216 *
262   :     -1.216 *
263   :     -1.216 *
264   :     -1.216 *
```

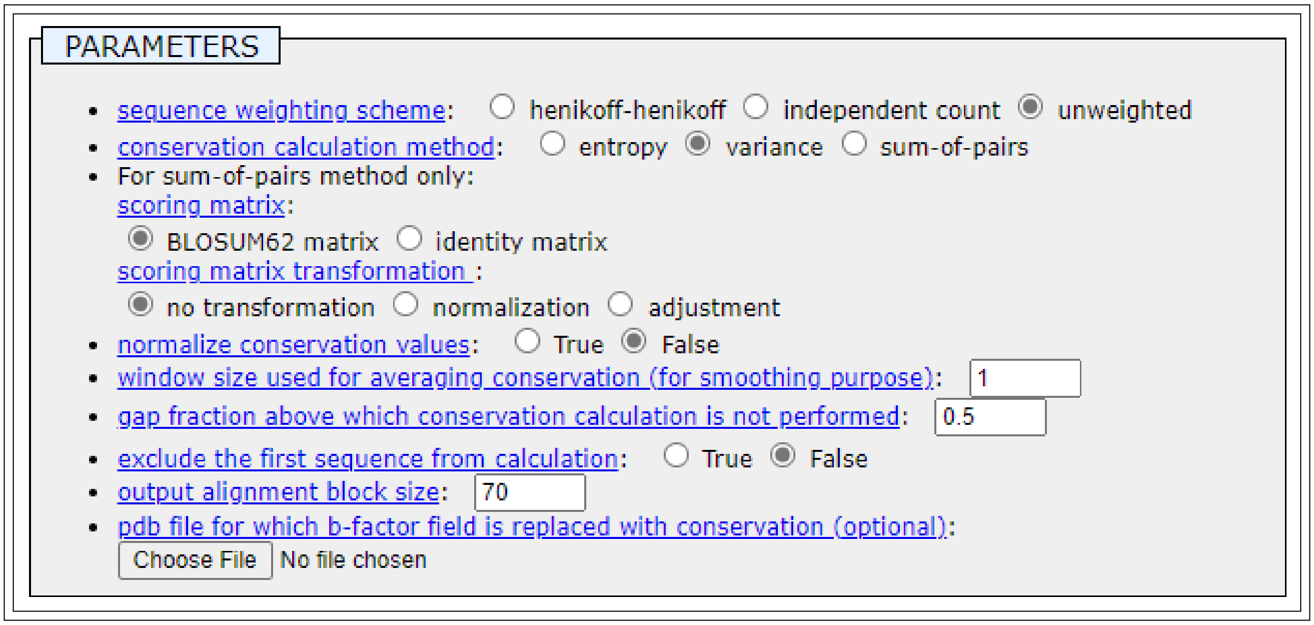
(E) 216-264 pos

FIGURE 10. Positional conservation scores for set 2

(ii) Unweighted frequency and variance-based measure

The parameters for this scenario are enclosed in the image below. The important parameters are as follows:

- sequence weighting scheme
- conservation calculation method
- scoring matrix (for sum of pairs method only)
- scoring matrix transformation (for sum of pairs method only)
- normalize conservation values



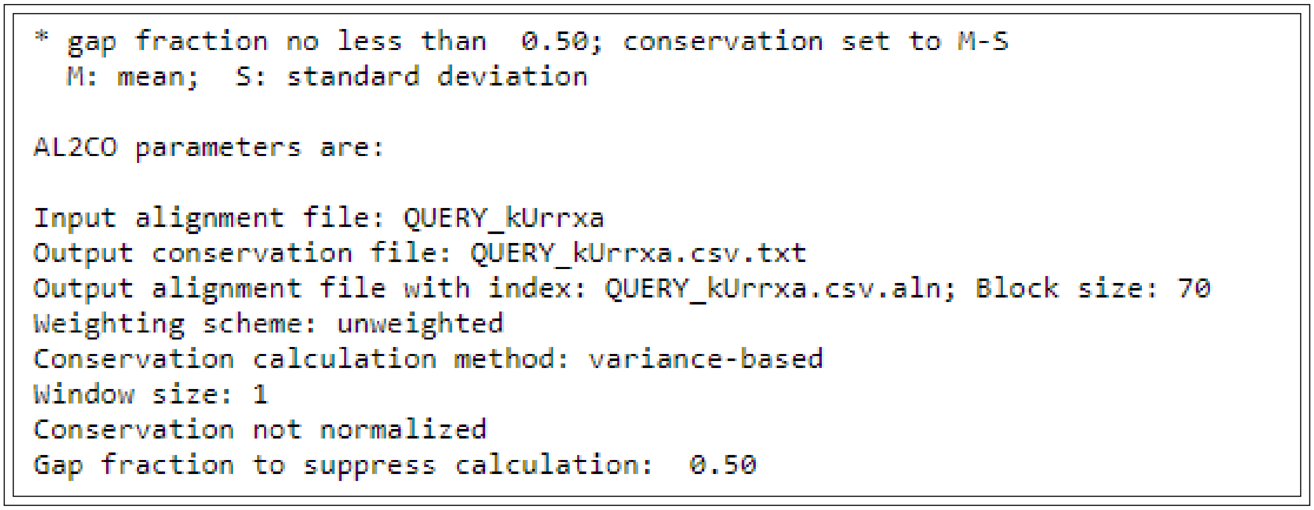
PARAMETERS

- sequence weighting scheme: ○ henikoff-henikoff ○ independent count ● unweighted
- conservation calculation method: ○ entropy ● variance ○ sum-of-pairs
- For sum-of-pairs method only:
  scoring matrix:
  ● BLOSUM62 matrix ○ identity matrix
  scoring matrix transformation :
  ● no transformation ○ normalization ○ adjustment
- normalize conservation values: ○ True ● False
- window size used for averaging conservation (for smoothing purpose): 1
- gap fraction above which conservation calculation is not performed: 0.5
- exclude the first sequence from calculation: ○ True ● False
- output alignment block size: 70
- pdb file for which b-factor field is replaced with conservation (optional):
  Choose File No file chosen

FIGURE 11. Parameters for the given scenario (ii)

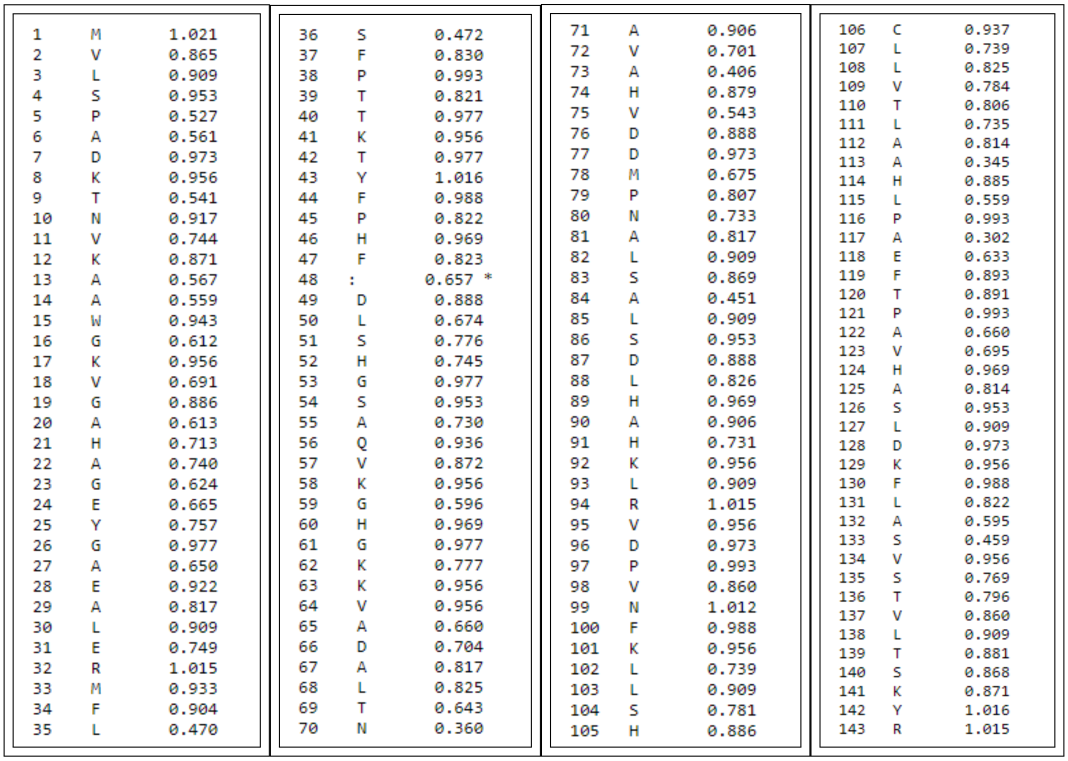
The AL2CO gives the list of positional conservation values and the alignment with integer conservation indices. The question asks to only calculate the positional conservation values. The window of positional conservation values generates the following set of parameters which are the ones displayed in the above image.

It also displays some parameters taken into consideration to compute the desired positional conservation scores.



```
* gap fraction no less than  0.50; conservation set to M-S
  M: mean;  S: standard deviation

AL2CO parameters are:

Input alignment file: QUERY_kUrrxa
Output conservation file: QUERY_kUrrxa.csv.txt
Output alignment file with index: QUERY_kUrrxa.csv.aln; Block size: 70
Weighting scheme: unweighted
Conservation calculation method: variance-based
Window size: 1
Conservation not normalized
Gap fraction to suppress calculation:  0.50
```

FIGURE 12. Parameter output display in window



```
1     M     1.021
2     V     0.865
3     L     0.909
4     S     0.953
5     P     0.527
6     A     0.561
7     D     0.973
8     K     0.956
9     T     0.541
10    N     0.917
11    V     0.744
12    K     0.871
13    A     0.567
14    A     0.559
15    W     0.943
16    G     0.612
17    K     0.956
18    V     0.691
19    G     0.886
20    A     0.613
21    H     0.713
22    A     0.740
23    G     0.624
24    E     0.665
25    Y     0.757
26    G     0.977
27    A     0.650
28    E     0.922
29    A     0.817
30    L     0.909
31    E     0.749
32    R     1.015
33    M     0.933
34    F     0.904
35    L     0.470
```

(A) 1-35 pos

```
36    S     0.472
37    F     0.830
38    P     0.993
39    T     0.821
40    T     0.977
41    K     0.956
42    T     0.977
43    Y     1.016
44    F     0.988
45    P     0.822
46    H     0.969
47    F     0.823
48   :     0.657 *
49    D     0.888
50    L     0.674
51    S     0.776
52    H     0.745
53    G     0.977
54    S     0.953
55    A     0.730
56    Q     0.936
57    V     0.872
58    K     0.956
59    G     0.596
60    H     0.969
61    G     0.977
62    K     0.777
63    K     0.956
64    V     0.956
65    A     0.660
66    D     0.704
67    A     0.817
68    L     0.825
69    T     0.643
70    N     0.360
```

(B) 36-70 pos

```
71    A     0.906
72    V     0.701
73    A     0.406
74    H     0.879
75    V     0.543
76    D     0.888
77    D     0.973
78    M     0.675
79    P     0.807
80    N     0.733
81    A     0.817
82    L     0.909
83    S     0.869
84    A     0.451
85    L     0.909
86    S     0.953
87    D     0.888
88    L     0.826
89    H     0.969
90    A     0.906
91    H     0.731
92    K     0.956
93    L     0.909
94    R     1.015
95    V     0.956
96    D     0.973
97    P     0.993
98    V     0.860
99    N     1.012
100   F     0.988
101   K     0.956
102   L     0.739
103   L     0.909
104   S     0.781
105   H     0.886
```

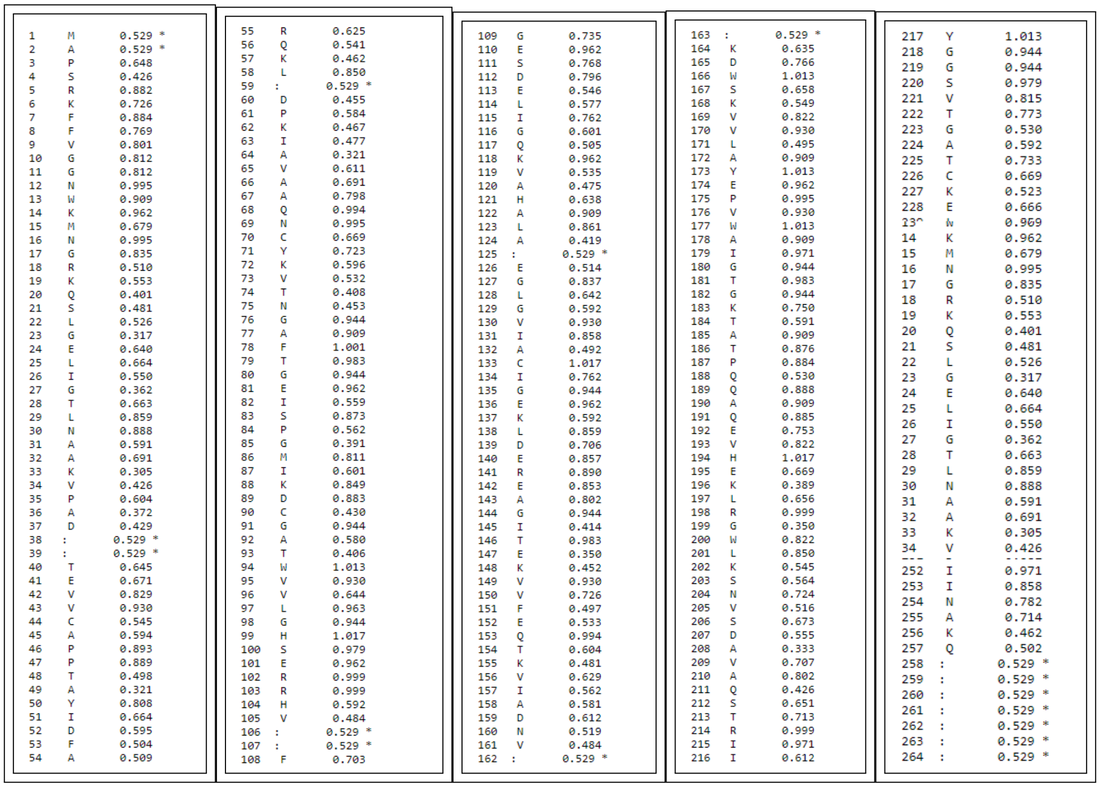
(C) 71-105 pos

```
106   C     0.937
107   L     0.739
108   L     0.825
109   V     0.784
110   T     0.806
111   L     0.735
112   A     0.814
113   A     0.345
114   H     0.885
115   L     0.559
116   P     0.993
117   A     0.302
118   E     0.633
119   F     0.893
120   T     0.891
121   P     0.993
122   A     0.660
123   V     0.695
124   H     0.969
125   A     0.814
126   S     0.953
127   L     0.909
128   D     0.973
129   K     0.956
130   F     0.988
131   L     0.822
132   A     0.595
133   S     0.459
134   V     0.956
135   S     0.769
136   T     0.796
137   V     0.860
138   L     0.909
139   T     0.881
140   S     0.868
141   K     0.871
142   Y     1.016
143   R     1.015
```

(D) 106-143 pos

Figure 13. Positional conservation scores for set 1

Above is the image of the positional conservation scores for the set 1 proteins. Below is the image of the positional conservation scores for the set 2 proteins.



```
1     M     0.529 *
2     A     0.529 *
3     P     0.648
4     S     0.426
5     R     0.882
6     K     0.726
7     F     0.884
8     F     0.769
9     V     0.801
10    G     0.812
11    G     0.812
12    N     0.995
13    W     0.909
14    K     0.962
15    M     0.679
16    N     0.995
17    G     0.835
18    R     0.510
19    K     0.553
20    Q     0.401
21    S     0.481
22    L     0.526
23    G     0.317
24    E     0.640
25    L     0.664
26    I     0.550
27    G     0.362
28    T     0.663
29    L     0.859
30    N     0.888
31    A     0.591
32    A     0.691
33    K     0.305
34    V     0.426
35    P     0.604
36    A     0.372
37    D     0.429
38   :     0.529 *
39   :     0.529 *
40    T     0.645
41    E     0.671
42    V     0.829
43    V     0.930
44    C     0.545
45    A     0.594
46    P     0.893
47    P     0.889
48    T     0.498
49    A     0.321
50    Y     0.808
51    I     0.664
52    D     0.595
53    F     0.504
54    A     0.509
```

(A) 1-54 pos

```
55    R     0.625
56    Q     0.541
57    K     0.462
58    L     0.850
59   :     0.529 *
60    D     0.455
61    P     0.584
62    K     0.467
63    I     0.477
64    A     0.321
65    V     0.611
66    A     0.691
67    A     0.798
68    Q     0.994
69    N     0.995
70    C     0.669
71    Y     0.723
72    K     0.596
73    V     0.532
74    T     0.408
75    N     0.453
76    G     0.944
77    A     0.909
78    F     1.001
79    T     0.983
80    G     0.944
81    E     0.962
82    I     0.559
83    S     0.873
84    P     0.562
85    G     0.391
86    M     0.811
87    I     0.601
88    K     0.849
89    D     0.883
90    C     0.430
91    G     0.944
92    A     0.580
93    T     0.406
94    W     1.013
95    V     0.930
96    V     0.644
97    L     0.963
98    G     0.944
99    H     1.017
100   S     0.979
101   E     0.962
102   R     0.999
103   R     0.999
104   H     0.592
105   V     0.484
106  :     0.529 *
107  :     0.529 *
108   F     0.703
```

(B) 55-108 pos

```
109   G     0.735
110   E     0.962
111   S     0.768
112   D     0.796
113   E     0.546
114   L     0.577
115   I     0.762
116   G     0.601
117   Q     0.505
118   K     0.962
119   V     0.535
120   A     0.475
121   H     0.638
122   A     0.909
123   L     0.861
124   A     0.419
125  :     0.529 *
126   E     0.514
127   G     0.837
128   L     0.642
129   G     0.592
130   V     0.930
131   I     0.858
132   A     0.492
133   C     1.017
134   I     0.762
135   G     0.944
136   E     0.962
137   K     0.592
138   L     0.859
139   D     0.706
140   E     0.857
141   R     0.890
142   E     0.853
143   A     0.802
144   G     0.944
145   I     0.414
146   T     0.983
147   E     0.350
148   K     0.452
149   V     0.930
150   V     0.726
151   F     0.497
152   E     0.533
153   Q     0.994
154   T     0.604
155   K     0.481
156   V     0.629
157   I     0.562
158   A     0.581
159   D     0.612
160   N     0.519
161   V     0.484
162  :     0.529 *
```

(C) 109-162 pos

```
163  :     0.529 *
164   K     0.635
165   D     0.766
166   W     1.013
167   S     0.658
168   K     0.549
169   V     0.822
170   V     0.930
171   L     0.495
172   A     0.909
173   Y     1.013
174   E     0.962
175   P     0.995
176   V     0.930
177   W     1.013
178   A     0.909
179   I     0.971
180   G     0.944
181   T     0.983
182   G     0.944
183   K     0.750
184   T     0.591
185   A     0.909
186   T     0.876
187   P     0.884
188   Q     0.530
189   Q     0.888
190   A     0.909
191   Q     0.885
192   E     0.753
193   V     0.822
194   H     1.017
195   E     0.669
196   K     0.389
197   L     0.656
198   R     0.999
199   G     0.350
200   W     0.822
201   L     0.850
202   K     0.545
203   S     0.564
204   N     0.724
205   V     0.516
206   S     0.673
207   D     0.555
208   A     0.333
209   V     0.707
210   A     0.802
211   Q     0.426
212   S     0.651
213   T     0.713
214   R     0.999
215   I     0.971
216   I     0.612
```

(D) 163-216 pos

```
217   Y     1.013
218   G     0.944
219   G     0.944
220   S     0.979
221   V     0.815
222   T     0.773
223   G     0.530
224   A     0.592
225   T     0.733
226   C     0.669
227   K     0.523
228   E     0.666
23^   W     0.969
14    K     0.962
15    M     0.679
16    N     0.995
17    G     0.835
18    R     0.510
19    K     0.553
20    Q     0.401
21    S     0.481
22    L     0.526
23    G     0.317
24    E     0.640
25    L     0.664
26    I     0.550
27    G     0.362
28    T     0.663
29    L     0.859
30    N     0.888
31    A     0.591
32    A     0.691
33    K     0.305
34    V     0.426
252   I     0.971
253   I     0.858
254   N     0.782
255   A     0.714
256   K     0.462
257   Q     0.502
258  :     0.529 *
259  :     0.529 *
260  :     0.529 *
261  :     0.529 *
262  :     0.529 *
263  :     0.529 *
264  :     0.529 *
```

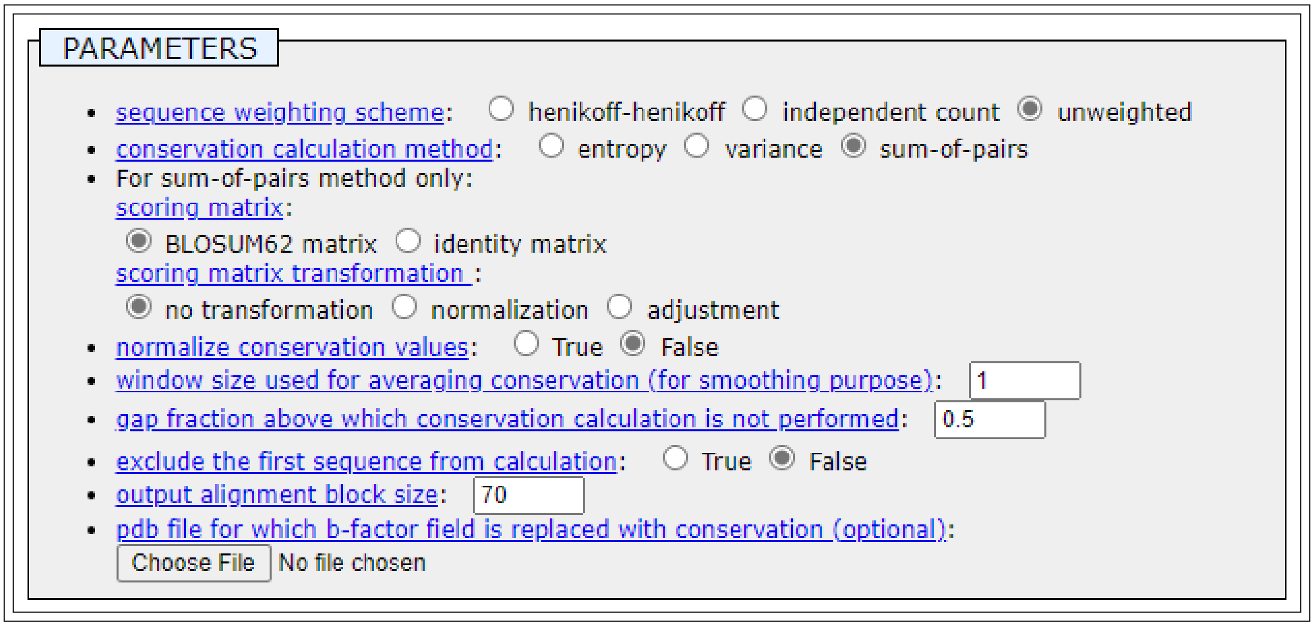
(E) 216-264 pos

Figure 14. Positional conservation scores for set 2

(iii) Unweighted frequency and sum of pairs measure

The parameters for this scenario are enclosed in the image below. The important parameters are as follows:

- sequence weighting scheme
- conservation calculation method
- scoring matrix (for sum of pairs method only)
- scoring matrix transformation (for sum of pairs method only)
- normalize conservation values



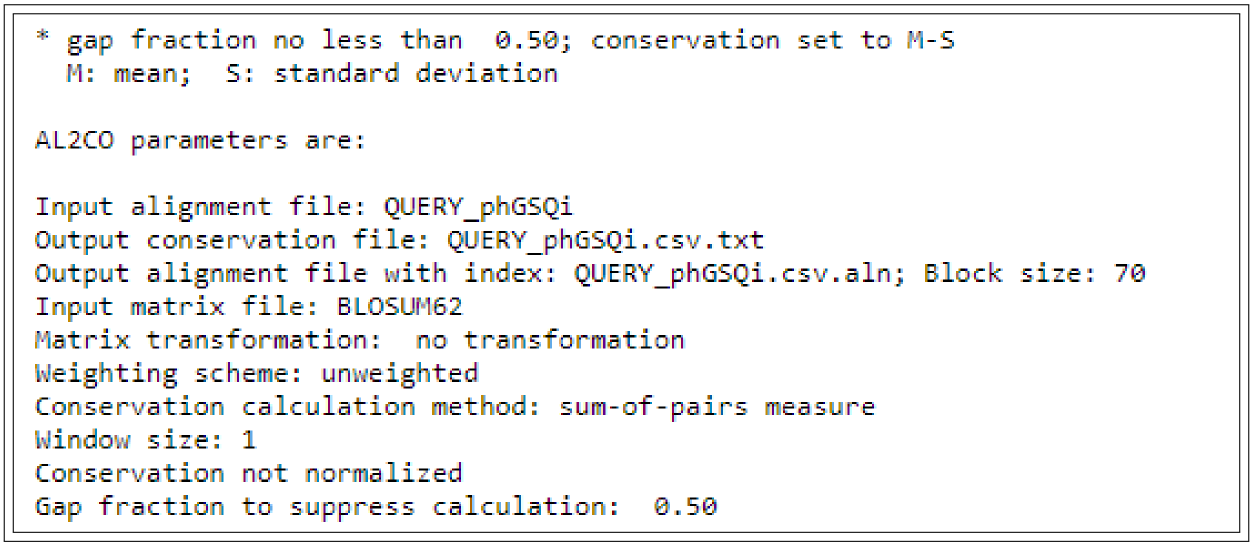
PARAMETERS

- sequence weighting scheme: ○ henikoff-henikoff ○ independent count ● unweighted
- conservation calculation method: ○ entropy ○ variance ● sum-of-pairs
- For sum-of-pairs method only:
  scoring matrix:
  ● BLOSUM62 matrix ○ identity matrix
  scoring matrix transformation :
  ● no transformation ○ normalization ○ adjustment
- normalize conservation values: ○ True ● False
- window size used for averaging conservation (for smoothing purpose): 1
- gap fraction above which conservation calculation is not performed: 0.5
- exclude the first sequence from calculation: ○ True ● False
- output alignment block size: 70
- pdb file for which b-factor field is replaced with conservation (optional):
  Choose File No file chosen

Figure 15. Parameters for the given scenario (iii)

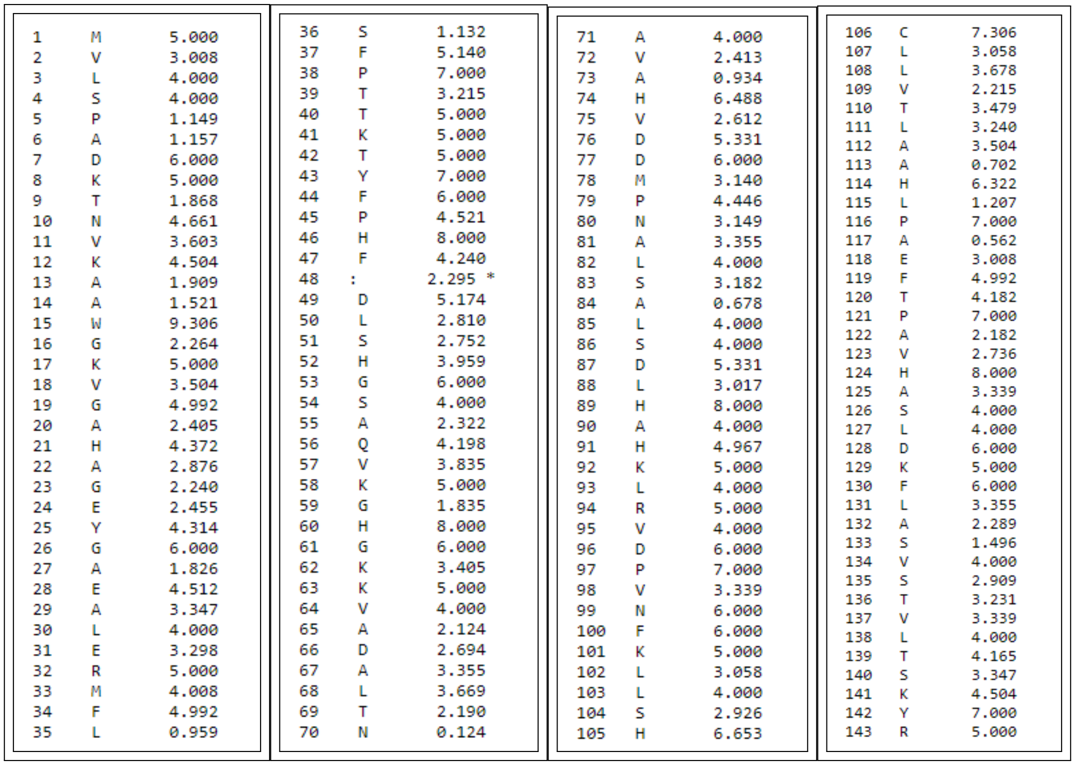
The AL2CO gives the list of positional conservation values and the alignment with integer conservation indices. The question asks to only calculate the positional conservation values. The window of positional conservation values generates the following set of parameters which are the ones displayed in the above image.

It also displays some parameters taken into consideration to compute the desired positional conservation scores.



```
* gap fraction no less than  0.50; conservation set to M-S
  M: mean;  S: standard deviation

AL2CO parameters are:

Input alignment file: QUERY_phGSQi
Output conservation file: QUERY_phGSQi.csv.txt
Output alignment file with index: QUERY_phGSQi.csv.aln; Block size: 70
Input matrix file: BLOSUM62
Matrix transformation:  no transformation
Weighting scheme: unweighted
Conservation calculation method: sum-of-pairs measure
Window size: 1
Conservation not normalized
Gap fraction to suppress calculation:  0.50
```

Figure 16. Parameter output display in window



```
1     M        5.000
2     V        3.008
3     L        4.000
4     S        4.000
5     P        1.149
6     A        1.157
7     D        6.000
8     K        5.000
9     T        1.868
10    N        4.661
11    V        3.603
12    K        4.504
13    A        1.909
14    A        1.521
15    W        9.306
16    G        2.264
17    K        5.000
18    V        3.504
19    G        4.992
20    A        2.405
21    H        4.372
22    A        2.876
23    G        2.240
24    E        2.455
25    Y        4.314
26    G        6.000
27    A        1.826
28    E        4.512
29    A        3.347
30    L        4.000
31    E        3.298
32    R        5.000
33    M        4.008
34    F        4.992
35    L        0.959
```

```
36    S        1.132
37    F        5.140
38    P        7.000
39    T        3.215
40    T        5.000
41    K        5.000
42    T        5.000
43    Y        7.000
44    F        6.000
45    P        4.521
46    H        8.000
47    F        4.240
48   :        2.295 *
49    D        5.174
50    L        2.810
51    S        2.752
52    H        3.959
53    G        6.000
54    S        4.000
55    A        2.322
56    Q        4.198
57    V        3.835
58    K        5.000
59    G        1.835
60    H        8.000
61    G        6.000
62    K        3.405
63    K        5.000
64    V        4.000
65    A        2.124
66    D        2.694
67    A        3.355
68    L        3.669
69    T        2.190
70    N        0.124
```

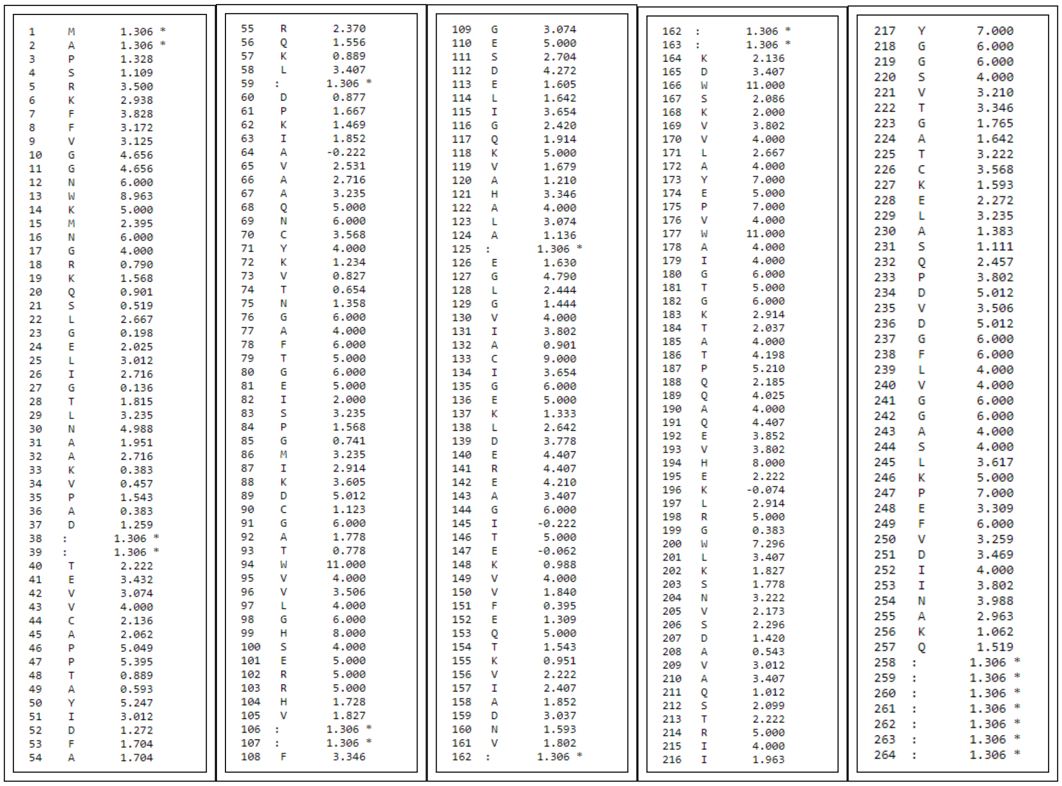
```
71    A        4.000
72    V        2.413
73    A        0.934
74    H        6.488
75    V        2.612
76    D        5.331
77    D        6.000
78    M        3.140
79    P        4.446
80    N        3.149
81    A        3.355
82    L        4.000
83    S        3.182
84    A        0.678
85    L        4.000
86    S        4.000
87    D        5.331
88    L        3.017
89    H        8.000
90    A        4.000
91    H        4.967
92    K        5.000
93    L        4.000
94    R        5.000
95    V        4.000
96    D        6.000
97    P        7.000
98    V        3.339
99    N        6.000
100   F        6.000
101   K        5.000
102   L        3.058
103   L        4.000
104   S        2.926
105   H        6.653
```

```
106   C        7.306
107   L        3.058
108   L        3.678
109   V        2.215
110   T        3.479
111   L        3.240
112   A        3.504
113   A        0.702
114   H        6.322
115   L        1.207
116   P        7.000
117   A        0.562
118   E        3.008
119   F        4.992
120   T        4.182
121   P        7.000
122   A        2.182
123   V        2.736
124   H        8.000
125   A        3.339
126   S        4.000
127   L        4.000
128   D        6.000
129   K        5.000
130   F        6.000
131   L        3.355
132   A        2.289
133   S        1.496
134   V        4.000
135   S        2.909
136   T        3.231
137   V        3.339
138   L        4.000
139   T        4.165
140   S        3.347
141   K        4.504
142   Y        7.000
143   R        5.000
```

(A) 1-35 pos (B) 36-70 pos (C) 71-105 pos (D) 106-143 pos

FIGURE 17. Positional conservation scores for set 1

Above is the image of the positional conservation scores for the set 1 proteins. Below is the image of the positional conservation scores for the set 2 proteins.



```
1     M        1.306 *
2     A        1.306 *
3     P        1.328
4     S        1.109
5     R        3.500
6     K        2.938
7     F        3.828
8     F        3.172
9     V        3.125
10    G        4.656
11    G        4.656
12    N        6.000
13    W        8.963
14    K        5.000
15    M        2.395
16    N        6.000
17    G        4.000
18    R        0.790
19    K        1.568
20    Q        0.901
21    S        0.519
22    L        2.667
23    G        0.198
24    E        2.025
25    L        3.012
26    I        2.716
27    G        0.136
28    T        1.815
29    L        3.235
30    N        4.988
31    A        1.951
32    A        2.716
33    K        0.383
34    V        0.457
35    P        1.543
36    A        0.383
37    D        1.259
38   :        1.306 *
39   :        1.306 *
40    T        2.222
41    E        3.432
42    V        3.074
43    V        4.000
44    C        2.136
45    A        2.062
46    P        5.049
47    P        5.395
48    T        0.889
49    A        0.593
50    Y        5.247
51    I        3.012
52    D        1.272
53    F        1.704
54    A        1.704
```

```
55    R        2.370
56    Q        1.556
57    K        0.889
58    L        3.407
59   :        1.306 *
60    D        0.877
61    P        1.667
62    K        1.469
63    I        1.852
64    A       -0.222
65    V        2.531
66    A        2.716
67    A        3.235
68    Q        5.000
69    N        6.000
70    C        3.568
71    Y        4.000
72    K        1.234
73    V        0.827
74    T        0.654
75    N        1.358
76    G        6.000
77    A        4.000
78    F        6.000
79    T        5.000
80    G        6.000
81    E        5.000
82    I        2.000
83    S        3.235
84    P        1.568
85    G        0.741
86    M        3.235
87    I        2.914
88    K        3.605
89    D        5.012
90    C        1.123
91    G        6.000
92    A        1.778
93    T        0.778
94    W       11.000
95    V        4.000
96    V        3.506
97    L        4.000
98    G        6.000
99    H        8.000
100   S        4.000
101   E        5.000
102   R        5.000
103   R        5.000
104   H        1.728
105   V        1.827
106  :        1.306 *
107  :        1.306 *
108   F        3.346
```

```
109   G        3.074
110   E        5.000
111   S        2.704
112   D        4.272
113   E        1.605
114   L        1.642
115   I        3.654
116   G        2.420
117   Q        1.914
118   K        5.000
119   V        1.679
120   A        1.210
121   H        3.346
122   A        4.000
123   L        3.074
124   A        1.136
125  :        1.306 *
126   E        1.630
127   G        4.790
128   L        2.444
129   G        1.444
130   V        4.000
131   I        3.802
132   A        0.901
133   C        9.000
134   I        3.654
135   G        6.000
136   E        5.000
137   K        1.333
138   L        2.642
139   D        3.778
140   E        4.407
141   R        4.407
142   E        4.210
143   A        3.407
144   G        6.000
145   I       -0.222
146   T        5.000
147   E       -0.062
148   K        0.988
149   V        4.000
150   V        1.840
151   F        0.395
152   E        1.309
153   Q        5.000
154   T        1.543
155   K        0.951
156   V        2.222
157   I        2.407
158   A        1.852
159   D        3.037
160   N        1.593
161   V        1.802
162  :        1.306 *
```

```
162  :        1.306 *
163  :        1.306 *
164   K        2.136
165   D        3.407
166   W       11.000
167   S        2.086
168   K        2.000
169   V        3.802
170   V        4.000
171   L        2.667
172   A        4.000
173   Y        7.000
174   E        5.000
175   P        7.000
176   V        4.000
177   W       11.000
178   A        4.000
179   I        4.000
180   G        6.000
181   T        5.000
182   G        6.000
183   K        2.914
184   T        2.037
185   A        4.000
186   T        4.198
187   P        5.210
188   Q        2.185
189   Q        4.025
190   A        4.000
191   Q        4.407
192   E        3.852
193   V        3.802
194   H        8.000
195   E        2.222
196   K       -0.074
197   L        2.914
198   R        5.000
199   G        0.383
200   W        7.296
201   L        3.407
202   K        1.827
203   S        1.778
204   N        3.222
205   V        2.173
206   S        2.296
207   D        1.420
208   A        0.543
209   V        3.012
210   A        3.407
211   Q        1.012
212   S        2.099
213   T        2.222
214   R        5.000
215   I        4.000
216   I        1.963
```

```
217   Y        7.000
218   G        6.000
219   G        6.000
220   S        4.000
221   V        3.210
222   T        3.346
223   G        1.765
224   A        1.642
225   T        3.222
226   C        3.568
227   K        1.593
228   E        2.272
229   L        3.235
230   A        1.383
231   S        1.111
232   Q        2.457
233   P        3.802
234   D        5.012
235   V        3.506
236   D        5.012
237   G        6.000
238   F        6.000
239   L        4.000
240   V        4.000
241   G        6.000
242   G        6.000
243   A        4.000
244   S        4.000
245   L        3.617
246   K        5.000
247   P        7.000
248   E        3.309
249   F        6.000
250   V        3.259
251   D        3.469
252   I        4.000
253   I        3.802
254   N        3.988
255   A        2.963
256   K        1.062
257   Q        1.519
258  :        1.306 *
259  :        1.306 *
260  :        1.306 *
261  :        1.306 *
262  :        1.306 *
263  :        1.306 *
264  :        1.306 *
```

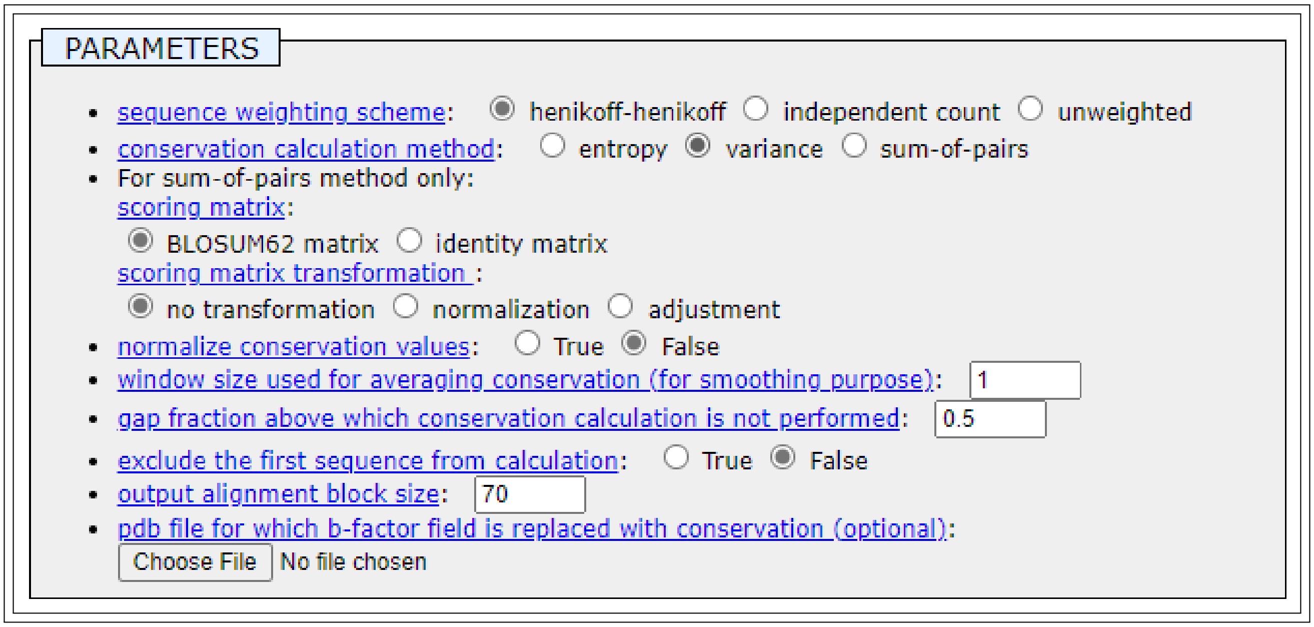
(A) 1-54 pos (B) 55-108 pos (C) 109-162 pos (D) 163-216 pos (E) 216-264 pos

FIGURE 18. Positional conservation scores for set 2

(iv) Weighted frequency and variance-based measure

The parameters for this scenario are enclosed in the image below. The important parameters are as follows:

- sequence weighting scheme
- conservation calculation method
- scoring matrix (for sum of pairs method only)
- scoring matrix transformation (for sum of pairs method only)
- normalize conservation values

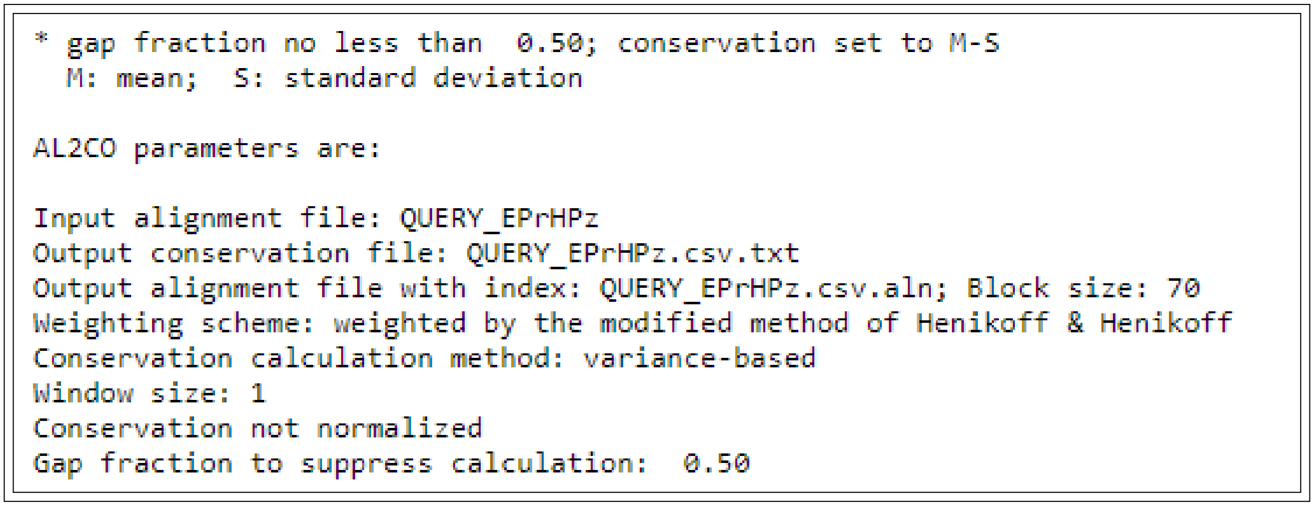


PARAMETERS

- sequence weighting scheme: ● henikoff-henikoff ○ independent count ○ unweighted
- conservation calculation method: ○ entropy ● variance ○ sum-of-pairs
- For sum-of-pairs method only:
  scoring matrix:
  ● BLOSUM62 matrix ○ identity matrix
  scoring matrix transformation :
  ● no transformation ○ normalization ○ adjustment
- normalize conservation values: ○ True ● False
- window size used for averaging conservation (for smoothing purpose): 1
- gap fraction above which conservation calculation is not performed: 0.5
- exclude the first sequence from calculation: ○ True ● False
- output alignment block size: 70
- pdb file for which b-factor field is replaced with conservation (optional):

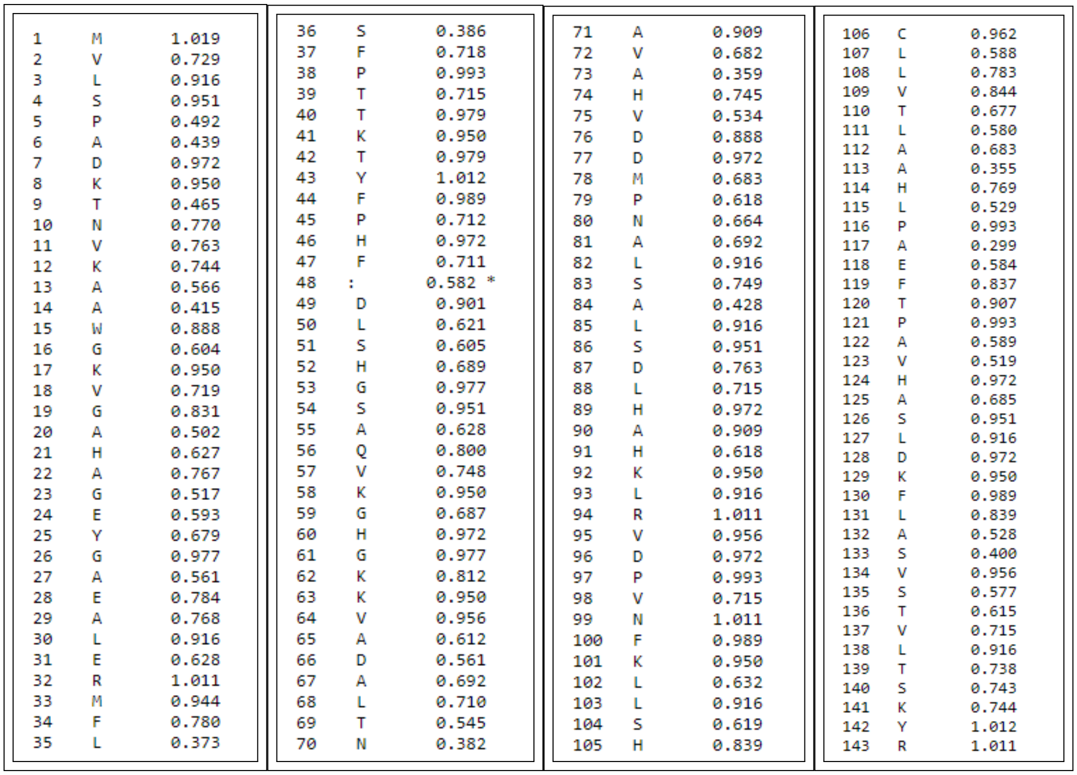
Figure 19. Parameters for the given scenario (iv)

The AL2CO gives the list of positional conservation values and the alignment with integer conservation indices. The question asks to only calculate the positional conservation values.
The window of positional conservation values generates the following set of parameters which are the ones displayed in the above image.
It also displays some parameters taken into consideration to compute the desired positional conservation scores.



```
* gap fraction no less than  0.50; conservation set to M-S
  M: mean;  S: standard deviation

AL2CO parameters are:

Input alignment file: QUERY_EPrHPz
Output conservation file: QUERY_EPrHPz.csv.txt
Output alignment file with index: QUERY_EPrHPz.csv.aln; Block size: 70
Weighting scheme: weighted by the modified method of Henikoff & Henikoff
Conservation calculation method: variance-based
Window size: 1
Conservation not normalized
Gap fraction to suppress calculation:  0.50
```

Figure 20. Parameter output display in window



```
1    M    1.019
2    V    0.729
3    L    0.916
4    S    0.951
5    P    0.492
6    A    0.439
7    D    0.972
8    K    0.950
9    T    0.465
10   N    0.770
11   V    0.763
12   K    0.744
13   A    0.566
14   A    0.415
15   W    0.888
16   G    0.604
17   K    0.950
18   V    0.719
19   G    0.831
20   A    0.502
21   H    0.627
22   A    0.767
23   G    0.517
24   E    0.593
25   Y    0.679
26   G    0.977
27   A    0.561
28   E    0.784
29   A    0.768
30   L    0.916
31   E    0.628
32   R    1.011
33   M    0.944
34   F    0.780
35   L    0.373
```

```
36   S    0.386
37   F    0.718
38   P    0.993
39   T    0.715
40   T    0.979
41   K    0.950
42   T    0.979
43   Y    1.012
44   F    0.989
45   P    0.712
46   H    0.972
47   F    0.711
48  :    0.582 *
49   D    0.901
50   L    0.621
51   S    0.605
52   H    0.689
53   G    0.977
54   S    0.951
55   A    0.628
56   Q    0.800
57   V    0.748
58   K    0.950
59   G    0.687
60   H    0.972
61   G    0.977
62   K    0.812
63   K    0.950
64   V    0.956
65   A    0.612
66   D    0.561
67   A    0.692
68   L    0.710
69   T    0.545
70   N    0.382
```

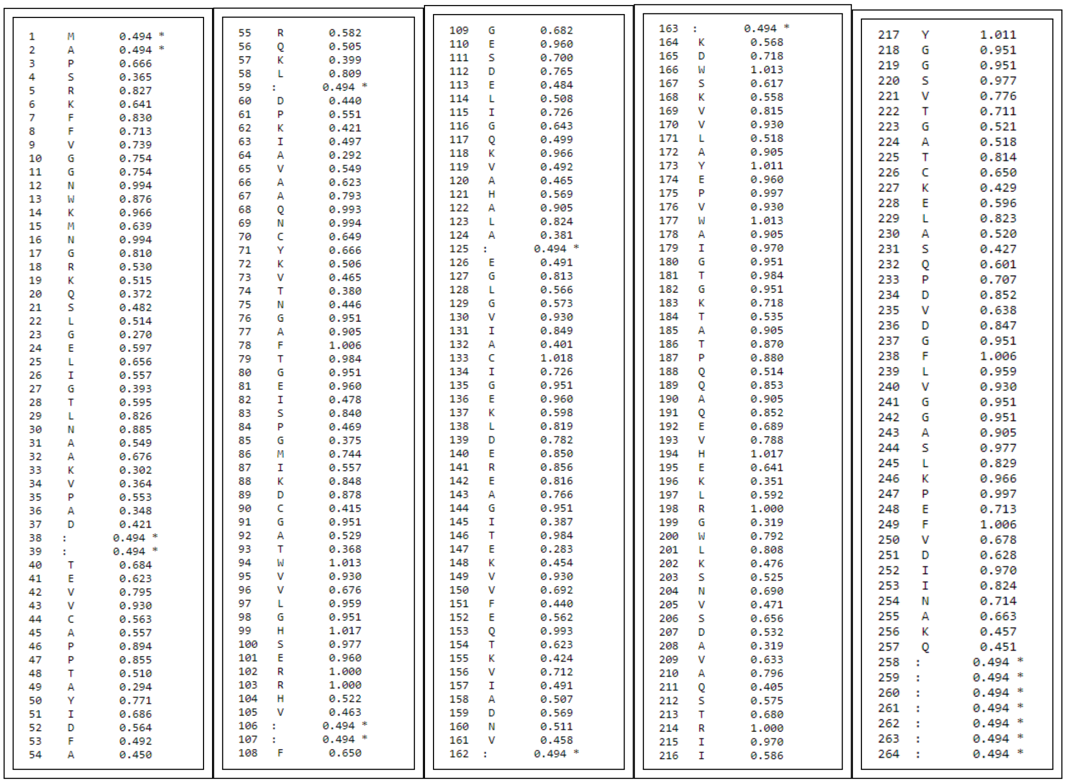
```
71   A    0.909
72   V    0.682
73   A    0.359
74   H    0.745
75   V    0.534
76   D    0.888
77   D    0.972
78   M    0.683
79   P    0.618
80   N    0.664
81   A    0.692
82   L    0.916
83   S    0.749
84   A    0.428
85   L    0.916
86   S    0.951
87   D    0.763
88   L    0.715
89   H    0.972
90   A    0.909
91   H    0.618
92   K    0.950
93   L    0.916
94   R    1.011
95   V    0.956
96   D    0.972
97   P    0.993
98   V    0.715
99   N    1.011
100  F    0.989
101  K    0.950
102  L    0.632
103  L    0.916
104  S    0.619
105  H    0.839
```

```
106  C    0.962
107  L    0.588
108  L    0.783
109  V    0.844
110  T    0.677
111  L    0.580
112  A    0.683
113  A    0.355
114  H    0.769
115  L    0.529
116  P    0.993
117  A    0.299
118  E    0.584
119  F    0.837
120  T    0.907
121  P    0.993
122  A    0.589
123  V    0.519
124  H    0.972
125  A    0.685
126  S    0.951
127  L    0.916
128  D    0.972
129  K    0.950
130  F    0.989
131  L    0.839
132  A    0.528
133  S    0.400
134  V    0.956
135  S    0.577
136  T    0.615
137  V    0.715
138  L    0.916
139  T    0.738
140  S    0.743
141  K    0.744
142  Y    1.012
143  R    1.011
```

(A) 1-35 pos (B) 36-70 pos (C) 71-105 pos (D) 106-143 pos

FIGURE 21. Positional conservation scores for set 1

Above is the image of the positional conservation scores for the set 1 proteins. Below is the image of the positional conservation scores for the set 2 proteins.



```
1    M    0.494 *
2    A    0.494 *
3    P    0.666
4    S    0.365
5    R    0.827
6    K    0.641
7    F    0.830
8    F    0.713
9    V    0.739
10   G    0.754
11   G    0.754
12   N    0.994
13   W    0.876
14   K    0.966
15   M    0.639
16   N    0.994
17   G    0.810
18   R    0.530
19   K    0.515
20   Q    0.372
21   S    0.482
22   L    0.514
23   G    0.270
24   E    0.597
25   L    0.656
26   I    0.557
27   G    0.393
28   T    0.595
29   L    0.826
30   N    0.885
31   A    0.549
32   A    0.676
33   K    0.302
34   V    0.364
35   P    0.553
36   A    0.348
37   D    0.421
38  :    0.494 *
39  :    0.494 *
40   T    0.684
41   E    0.623
42   V    0.795
43   V    0.930
44   C    0.563
45   A    0.557
46   P    0.894
47   P    0.855
48   T    0.510
49   A    0.294
50   Y    0.771
51   I    0.686
52   D    0.564
53   F    0.492
54   A    0.450
```

```
55   R    0.582
56   Q    0.505
57   K    0.399
58   L    0.809
59  :    0.494 *
60   D    0.440
61   P    0.551
62   K    0.421
63   I    0.497
64   A    0.292
65   V    0.549
66   A    0.623
67   A    0.793
68   Q    0.993
69   N    0.994
70   C    0.649
71   Y    0.666
72   K    0.506
73   V    0.465
74   T    0.380
75   N    0.446
76   G    0.951
77   A    0.905
78   F    1.006
79   T    0.984
80   G    0.951
81   E    0.960
82   I    0.478
83   S    0.840
84   P    0.469
85   G    0.375
86   M    0.744
87   I    0.557
88   K    0.848
89   D    0.878
90   C    0.415
91   G    0.951
92   A    0.529
93   T    0.368
94   W    1.013
95   V    0.930
96   V    0.676
97   L    0.959
98   G    0.951
99   H    1.017
100  S    0.977
101  E    0.960
102  R    1.000
103  R    1.000
104  H    0.522
105  V    0.463
106 :    0.494 *
107 :    0.494 *
108  F    0.650
```

```
109  G    0.682
110  E    0.960
111  S    0.700
112  D    0.765
113  E    0.484
114  L    0.508
115  I    0.726
116  G    0.643
117  Q    0.499
118  K    0.966
119  V    0.492
120  A    0.465
121  H    0.569
122  A    0.905
123  L    0.824
124  A    0.381
125 :    0.494 *
126  E    0.491
127  G    0.813
128  L    0.566
129  G    0.573
130  V    0.930
131  I    0.849
132  A    0.401
133  C    1.018
134  I    0.726
135  G    0.951
136  E    0.960
137  K    0.598
138  L    0.819
139  D    0.782
140  E    0.850
141  R    0.856
142  E    0.816
143  A    0.766
144  G    0.951
145  I    0.387
146  T    0.984
147  E    0.283
148  K    0.454
149  V    0.930
150  V    0.692
151  F    0.440
152  E    0.562
153  Q    0.993
154  T    0.623
155  K    0.424
156  V    0.712
157  I    0.491
158  A    0.507
159  D    0.569
160  N    0.511
161  V    0.458
162 :    0.494 *
```

```
163 :    0.494 *
164  K    0.568
165  D    0.718
166  W    1.013
167  S    0.617
168  K    0.558
169  V    0.815
170  V    0.930
171  L    0.518
172  A    0.905
173  Y    1.011
174  E    0.960
175  P    0.997
176  V    0.930
177  W    1.013
178  A    0.905
179  I    0.970
180  G    0.951
181  T    0.984
182  G    0.951
183  K    0.718
184  T    0.535
185  A    0.905
186  T    0.870
187  P    0.880
188  Q    0.514
189  Q    0.853
190  A    0.905
191  Q    0.852
192  E    0.689
193  V    0.788
194  H    1.017
195  E    0.641
196  K    0.351
197  L    0.592
198  R    1.000
199  G    0.319
200  W    0.792
201  L    0.808
202  K    0.476
203  S    0.525
204  N    0.690
205  V    0.471
206  S    0.656
207  D    0.532
208  A    0.319
209  V    0.633
210  A    0.796
211  Q    0.405
212  S    0.575
213  T    0.680
214  R    1.000
215  I    0.970
216  I    0.586
```

```
217  Y    1.011
218  G    0.951
219  G    0.951
220  S    0.977
221  V    0.776
222  T    0.711
223  G    0.521
224  A    0.518
225  T    0.814
226  C    0.650
227  K    0.429
228  E    0.596
229  L    0.823
230  A    0.520
231  S    0.427
232  Q    0.601
233  P    0.707
234  D    0.852
235  V    0.638
236  D    0.847
237  G    0.951
238  F    1.006
239  L    0.959
240  V    0.930
241  G    0.951
242  G    0.951
243  A    0.905
244  S    0.977
245  L    0.829
246  K    0.966
247  P    0.997
248  E    0.713
249  F    1.006
250  V    0.678
251  D    0.628
252  I    0.970
253  I    0.824
254  N    0.714
255  A    0.663
256  K    0.457
257  Q    0.451
258 :    0.494 *
259 :    0.494 *
260 :    0.494 *
261 :    0.494 *
262 :    0.494 *
263 :    0.494 *
264 :    0.494 *
```

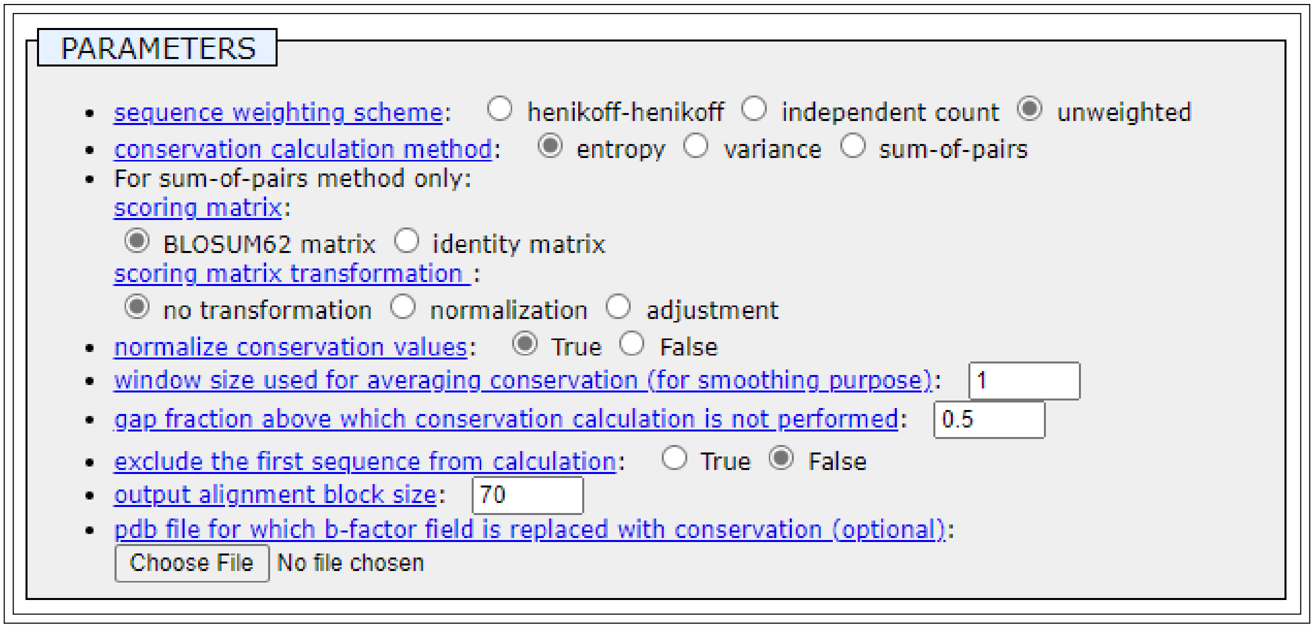
(A) 1-54 pos (B) 55-108 pos (C) 109-162 pos (D) 163-216 pos (E) 216-264 pos

FIGURE 22. Positional conservation scores for set 2

(v) Normalize the scores obtained with (i)

The parameters for this scenario are enclosed in the image below. The important parameters are as follows:

- sequence weighting scheme
- conservation calculation method
- scoring matrix (for sum of pairs method only)
- scoring matrix transformation (for sum of pairs method only)
- normalize conservation values

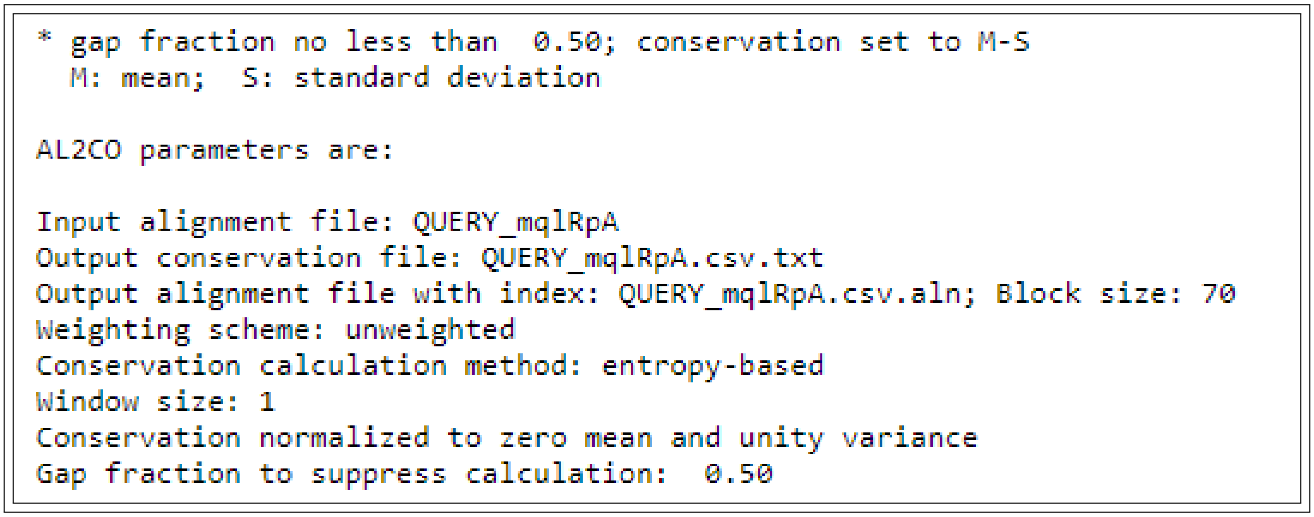


PARAMETERS

- sequence weighting scheme: ○ henikoff-henikoff ○ independent count ● unweighted
- conservation calculation method: ● entropy ○ variance ○ sum-of-pairs
- For sum-of-pairs method only:
  scoring matrix:
  ● BLOSUM62 matrix ○ identity matrix
  scoring matrix transformation :
  ● no transformation ○ normalization ○ adjustment
- normalize conservation values: ● True ○ False
- window size used for averaging conservation (for smoothing purpose): 1
- gap fraction above which conservation calculation is not performed: 0.5
- exclude the first sequence from calculation: ○ True ● False
- output alignment block size: 70
- pdb file for which b-factor field is replaced with conservation (optional):
  Choose File No file chosen

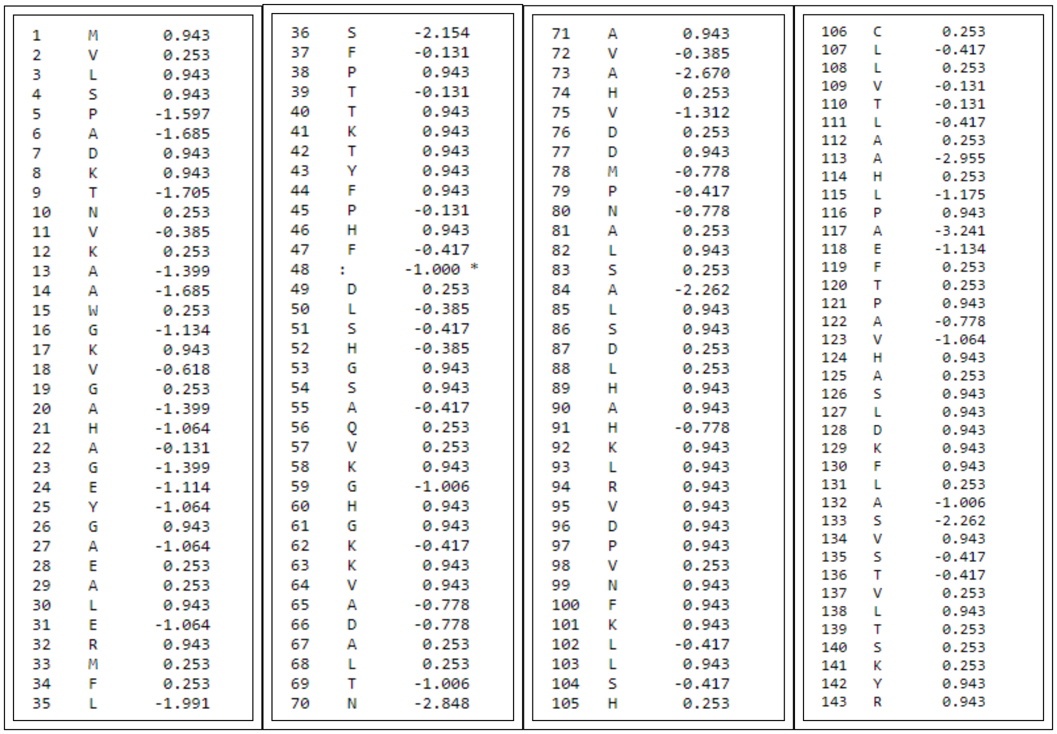
FIGURE 23. Parameters for the given scenario (v)

The AL2CO gives the list of positional conservation values and the alignment with integer conservation indices. The question asks to only calculate the positional conservation values.
The window of positional conservation values generates the following set of parameters which are the ones displayed in the above image.
It also displays some parameters taken into consideration to compute the desired positional conservation scores.



```
* gap fraction no less than  0.50; conservation set to M-S
  M: mean;  S: standard deviation

AL2CO parameters are:

Input alignment file: QUERY_mqlRpA
Output conservation file: QUERY_mqlRpA.csv.txt
Output alignment file with index: QUERY_mqlRpA.csv.aln; Block size: 70
Weighting scheme: unweighted
Conservation calculation method: entropy-based
Window size: 1
Conservation normalized to zero mean and unity variance
Gap fraction to suppress calculation:  0.50
```

FIGURE 24. Parameter output display in window



```
1    M    0.943
2    V    0.253
3    L    0.943
4    S    0.943
5    P   -1.597
6    A   -1.685
7    D    0.943
8    K    0.943
9    T   -1.705
10   N    0.253
11   V   -0.385
12   K    0.253
13   A   -1.399
14   A   -1.685
15   W    0.253
16   G   -1.134
17   K    0.943
18   V   -0.618
19   G    0.253
20   A   -1.399
21   H   -1.064
22   A   -0.131
23   G   -1.399
24   E   -1.114
25   Y   -1.064
26   G    0.943
27   A   -1.064
28   E    0.253
29   A    0.253
30   L    0.943
31   E   -1.064
32   R    0.943
33   M    0.253
34   F    0.253
35   L   -1.991
```

(A) 1-35 pos

```
36   S   -2.154
37   F   -0.131
38   P    0.943
39   T   -0.131
40   T    0.943
41   K    0.943
42   T    0.943
43   Y    0.943
44   F    0.943
45   P   -0.131
46   H    0.943
47   F   -0.417
48   :   -1.000 *
49   D    0.253
50   L   -0.385
51   S   -0.417
52   H   -0.385
53   G    0.943
54   S    0.943
55   A   -0.417
56   Q    0.253
57   V    0.253
58   K    0.943
59   G   -1.006
60   H    0.943
61   G    0.943
62   K   -0.417
63   K    0.943
64   V    0.943
65   A   -0.778
66   D   -0.778
67   A    0.253
68   L    0.253
69   T   -1.006
70   N   -2.848
```

(B) 36-70 pos

```
71   A    0.943
72   V   -0.385
73   A   -2.670
74   H    0.253
75   V   -1.312
76   D    0.253
77   D    0.943
78   M   -0.778
79   P   -0.417
80   N   -0.778
81   A    0.253
82   L    0.943
83   S    0.253
84   A   -2.262
85   L    0.943
86   S    0.943
87   D    0.253
88   L    0.253
89   H    0.943
90   A    0.943
91   H   -0.778
92   K    0.943
93   L    0.943
94   R    0.943
95   V    0.943
96   D    0.943
97   P    0.943
98   V    0.253
99   N    0.943
100  F    0.943
101  K    0.943
102  L   -0.417
103  L    0.943
104  S   -0.417
105  H    0.253
```

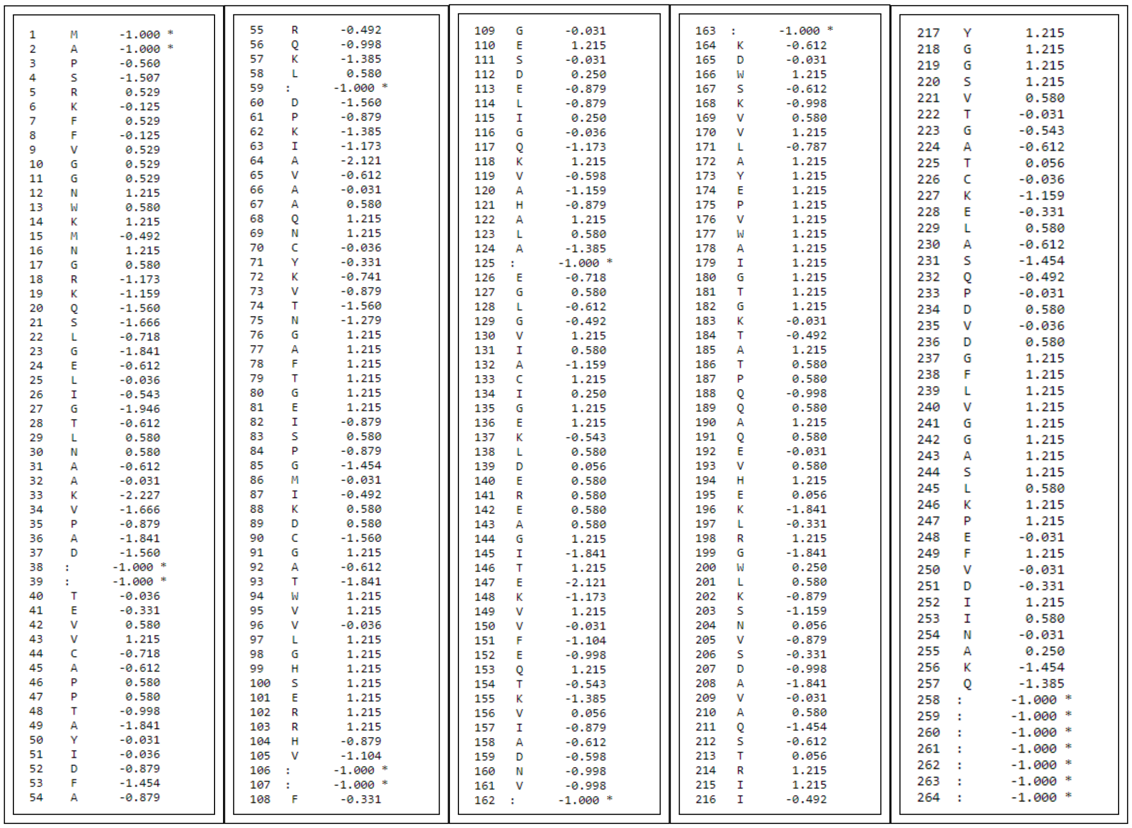
(C) 71-105 pos

```
106  C    0.253
107  L   -0.417
108  L    0.253
109  V   -0.131
110  T   -0.131
111  L   -0.417
112  A    0.253
113  A   -2.955
114  H    0.253
115  L   -1.175
116  P    0.943
117  A   -3.241
118  E   -1.134
119  F    0.253
120  T    0.253
121  P    0.943
122  A   -0.778
123  V   -1.064
124  H    0.943
125  A    0.253
126  S    0.943
127  L    0.943
128  D    0.943
129  K    0.943
130  F    0.943
131  L    0.253
132  A   -1.006
133  S   -2.262
134  V    0.943
135  S   -0.417
136  T   -0.417
137  V    0.253
138  L    0.943
139  T    0.253
140  S    0.253
141  K    0.253
142  Y    0.943
143  R    0.943
```

(D) 106-143 pos

Figure 25. Positional conservation scores for set 1

Above is the image of the positional conservation scores for the set 1 proteins. Below is the image of the positional conservation scores for the set 2 proteins.



```
1    M   -1.000 *
2    A   -1.000 *
3    P   -0.560
4    S   -1.507
5    R    0.529
6    K   -0.125
7    F    0.529
8    F   -0.125
9    V    0.529
10   G    0.529
11   G    0.529
12   N    1.215
13   W    0.580
14   K    1.215
15   M   -0.492
16   N    1.215
17   G    0.580
18   R   -1.173
19   K   -1.159
20   Q   -1.560
21   S   -1.666
22   L   -0.718
23   G   -1.841
24   E   -0.612
25   L   -0.036
26   I   -0.543
27   G   -1.946
28   T   -0.612
29   L    0.580
30   N    0.580
31   A   -0.612
32   A   -0.031
33   K   -2.227
34   V   -1.666
35   P   -0.879
36   A   -1.841
37   D   -1.560
38   :   -1.000 *
39   :   -1.000 *
40   T   -0.036
41   E   -0.331
42   V    0.580
43   V    1.215
44   C   -0.718
45   A   -0.612
46   P    0.580
47   P    0.580
48   T   -0.998
49   A   -1.841
50   Y   -0.031
51   I   -0.036
52   D   -0.879
53   F   -1.454
54   A   -0.879
```

(A) 1-54 pos

```
55   R   -0.492
56   Q   -0.998
57   K   -1.385
58   L    0.580
59   :   -1.000 *
60   D   -1.560
61   P   -0.879
62   K   -1.385
63   I   -1.173
64   A   -2.121
65   V   -0.612
66   A   -0.031
67   A    0.580
68   Q    1.215
69   N    1.215
70   C   -0.036
71   Y   -0.331
72   K   -0.741
73   V   -0.879
74   T   -1.560
75   N   -1.279
76   G    1.215
77   A    1.215
78   F    1.215
79   T    1.215
80   G    1.215
81   E    1.215
82   I   -0.879
83   S    0.580
84   P   -0.879
85   G   -1.454
86   M   -0.031
87   I   -0.492
88   K    0.580
89   D    0.580
90   C   -1.560
91   G    1.215
92   A   -0.612
93   T   -1.841
94   W    1.215
95   V    1.215
96   V   -0.036
97   L    1.215
98   G    1.215
99   H    1.215
100  S    1.215
101  E    1.215
102  R    1.215
103  R    1.215
104  H   -0.879
105  V   -1.104
106  :   -1.000 *
107  :   -1.000 *
108  F   -0.331
```

(B) 55-108 pos

```
109  G   -0.031
110  E    1.215
111  S   -0.031
112  D    0.250
113  E   -0.879
114  L   -0.879
115  I    0.250
116  G   -0.036
117  Q   -1.173
118  K    1.215
119  V   -0.598
120  A   -1.159
121  H   -0.879
122  A    1.215
123  L    0.580
124  A   -1.385
125  :   -1.000 *
126  E   -0.718
127  G    0.580
128  L   -0.612
129  G   -0.492
130  V    1.215
131  I    0.580
132  A   -1.159
133  C    1.215
134  I    0.250
135  G    1.215
136  E    1.215
137  K   -0.543
138  L    0.580
139  D    0.056
140  E    0.580
141  R    0.580
142  E    0.580
143  A    0.580
144  G    1.215
145  I   -1.841
146  T    1.215
147  E   -2.121
148  K   -1.173
149  V    1.215
150  V   -0.031
151  F   -1.104
152  E   -0.998
153  Q    1.215
154  T   -0.543
155  K   -1.385
156  V    0.056
157  I   -0.879
158  A   -0.612
159  D   -0.598
160  N   -0.998
161  V   -0.998
162  :   -1.000 *
```

(C) 109-162 pos

```
163  :   -1.000 *
164  K   -0.612
165  D   -0.031
166  W    1.215
167  S   -0.612
168  K   -0.998
169  V    0.580
170  V    1.215
171  L   -0.787
172  A    1.215
173  Y    1.215
174  E    1.215
175  P    1.215
176  V    1.215
177  W    1.215
178  A    1.215
179  I    1.215
180  G    1.215
181  T    1.215
182  G    1.215
183  K   -0.031
184  T   -0.492
185  A    1.215
186  T    0.580
187  P    0.580
188  Q   -0.998
189  Q    0.580
190  A    1.215
191  Q    0.580
192  E   -0.031
193  V    0.580
194  H    1.215
195  E    0.056
196  K   -1.841
197  L   -0.331
198  R    1.215
199  G   -1.841
200  W    0.250
201  L    0.580
202  K   -0.879
203  S   -1.159
204  N    0.056
205  V   -0.879
206  S   -0.331
207  D   -0.998
208  A   -1.841
209  V   -0.031
210  A    0.580
211  Q   -1.454
212  S   -0.612
213  T    0.056
214  R    1.215
215  I    1.215
216  I   -0.492
```

(D) 163-216 pos

```
217  Y    1.215
218  G    1.215
219  G    1.215
220  S    1.215
221  V    0.580
222  T   -0.031
223  G   -0.543
224  A   -0.612
225  T    0.056
226  C   -0.036
227  K   -1.159
228  E   -0.331
229  L    0.580
230  A   -0.612
231  S   -1.454
232  Q   -0.492
233  P   -0.031
234  D    0.580
235  V   -0.036
236  D    0.580
237  G    1.215
238  F    1.215
239  L    1.215
240  V    1.215
241  G    1.215
242  G    1.215
243  A    1.215
244  S    1.215
245  L    0.580
246  K    1.215
247  P    1.215
248  E   -0.031
249  F    1.215
250  V   -0.031
251  D   -0.331
252  I    1.215
253  I    0.580
254  N   -0.031
255  A    0.250
256  K   -1.454
257  Q   -1.385
258  :   -1.000 *
259  :   -1.000 *
260  :   -1.000 *
261  :   -1.000 *
262  :   -1.000 *
263  :   -1.000 *
264  :   -1.000 *
```

(E) 216-264 pos

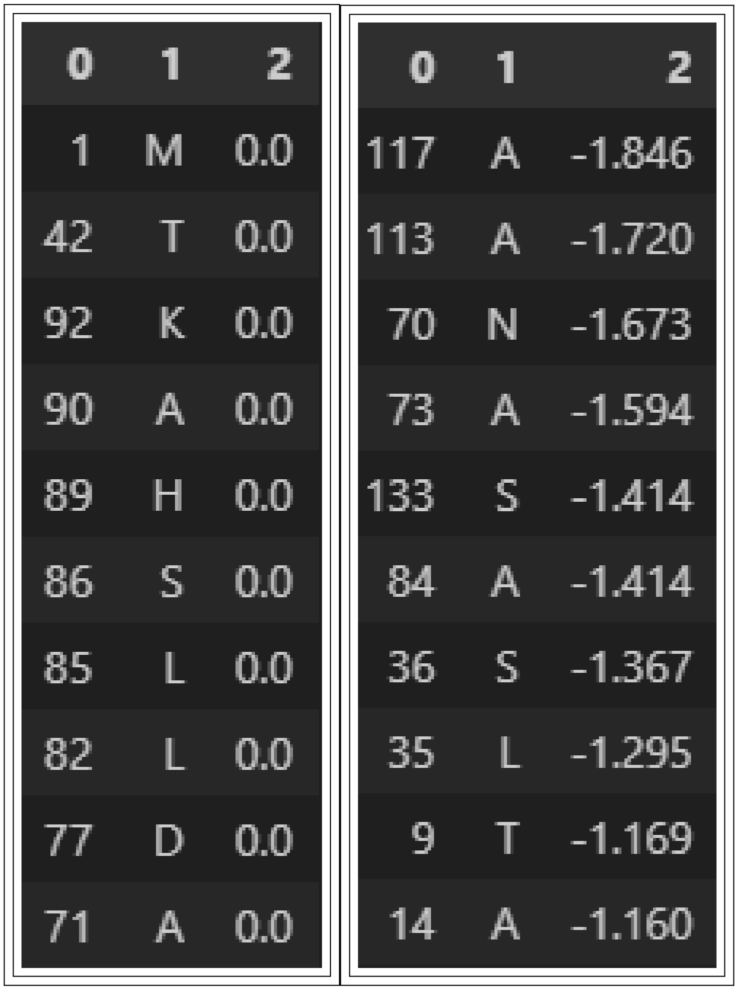
Figure 26. Positional conservation scores for set 2

Question 2. Tabulate the topmost 10 residues with highest and lowest conservation scores (in both Set1 and Set 2) obtained with method (i).

Solution. First the output obtained in the Q1 (i) is taken. The output has three columns as seen from its image. The position on sequence alignment, the residue, and its position conservation score.
Now, I have copied the above text from the AL2CO server's positional conservation score output to a text file. I have now written a code to read the file and then convert it into a dataframe, which has then been sorted to extract the top 10 and bottom 10 values as desired. The code is given below:

```python
import pandas as pd

new_line_list = []

# file.txt has the the positive conservation values the way it is shown in AL2CO server

with open("file.txt", ’r’) as f:

    # Read the text file and store in list

    line = f.read()
    line_list = line.split("\n")

    # Remove the additional spaces and positions with gap alignments

    for i in range(len(line_list)):
        line_i = line_list[i].split(" ")
        line_i = [str(value) for value in line_i if value != ""]
        if line_i[1] != ":":
            new_line_list.append(line_i)

# The above obtained list is then converted into a dataframe for easy readability and easier to perform sorting techniques and observe

df = pd.DataFrame(new_line_list)

# The first and third columns are converted to numeric data type because they represent the alignment position and position conservation scores respectively

df[2] = pd.to_numeric(df[2])
df[0] = pd.to_numeric(df[0])

# Sort the values in the descending order as per the last column, which is the column for positional conservation scores
final_df = df.sort_values(by=[2], ascending=False)

# Display the top 10 and bottom 10 residues
# Top 10 will be the highest scores
# Bottom 10 will be the lowest scores
final_df.head(10)
final_df.tail(10)
```

The last two lines will print the top 10 residues with highest and lowest positional conservation scores.



0	1	2
1	M	0.0
42	T	0.0
92	K	0.0
90	A	0.0
89	H	0.0
86	S	0.0
85	L	0.0
82	L	0.0
77	D	0.0
71	A	0.0

(A) Highest scores

0	1	2
117	A	-1.846
113	A	-1.720
70	N	-1.673
73	A	-1.594
133	S	-1.414
84	A	-1.414
36	S	-1.367
35	L	-1.295
9	T	-1.169
14	A	-1.160

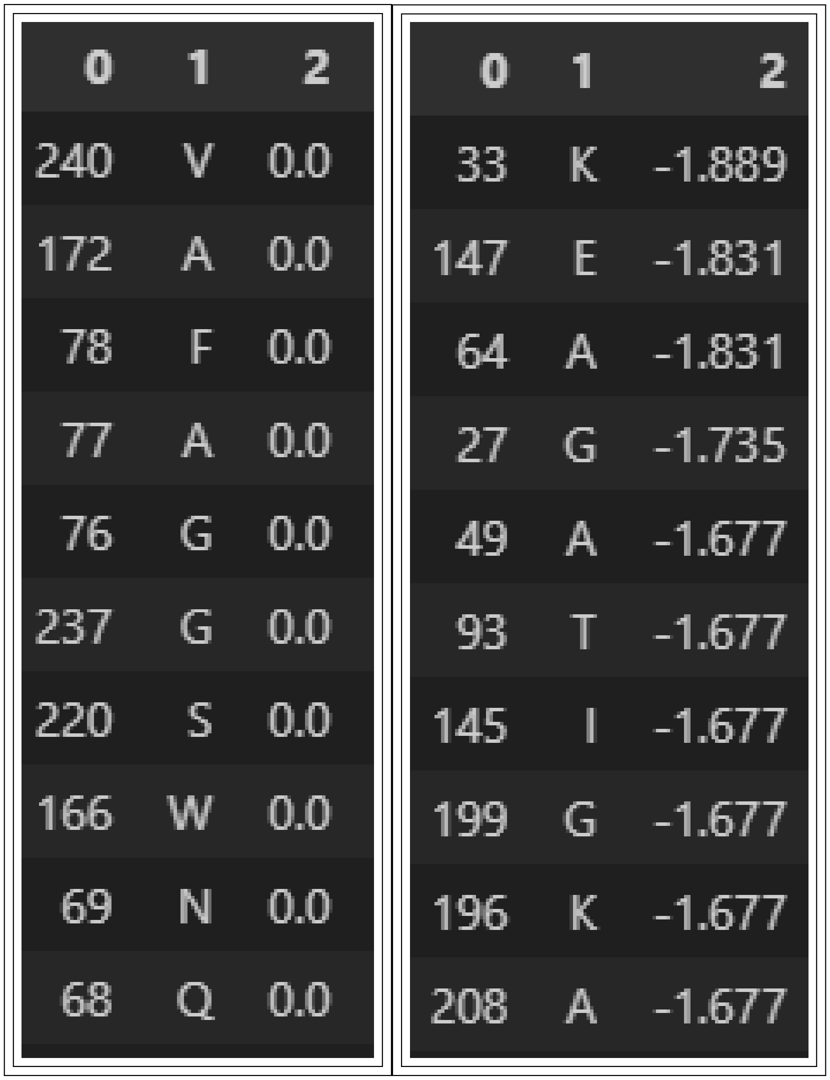
(B) Lowest scores

FIGURE 27. Top 10 residues with highest and lowest scores in set 1

The topmost 10 residues with highest and lowest conservation scores in set 1

In the following table, I have enclosed the readings from the above image with the top 10 residues with the highest and the lowest residues.
The results in this table are with respect to set 1 only.

Top 10 residues with the highest conservation scores			Top 10 residues with the lowest conservation scores		
Position	Residue	Score	Position	Residue	Score
1	Methionine	0.0	117	Alanine	-1.846
42	Threonine	0.0	113	Alanine	-1.720
92	Lysine	0.0	70	Asparagine	-1.673
90	Alanine	0.0	73	Alanine	-1.594
89	Histidine	0.0	133	Serine	-1.414
86	Serine	0.0	84	Alanine	-1.414
85	Leucine	0.0	36	Serine	-1.367
82	Leucine	0.0	35	Leucine	-1.295
77	Aspartate	0.0	9	Threonine	-1.169
71	Alanine	0.0	14	Alanine	-1.160



	0	1	2
	240	V	0.0
	172	A	0.0
	78	F	0.0
	77	A	0.0
	76	G	0.0
	237	G	0.0
	220	S	0.0
	166	W	0.0
	69	N	0.0
	68	Q	0.0

(A) Highest scores

	0	1	2
	33	K	-1.889
	147	E	-1.831
	64	A	-1.831
	27	G	-1.735
	49	A	-1.677
	93	T	-1.677
	145	I	-1.677
	199	G	-1.677
	196	K	-1.677
	208	A	-1.677

(B) Lowest scores

FIGURE 28. Top 10 residues with highest and lowest scores in set 2

The topmost 10 residues with highest and lowest conservation scores in set 2

In the following table, I have enclosed the readings from the above image with the top 10 residues with the highest and the lowest residues.
The results in this table are with respect to set 2 only.

Top 10 residues with the highest conservation scores			Top 10 residues with the lowest conservation scores		
Position	Residue	Score	Position	Residue	Score
240	Valine	0.0	33	Lysine	-1.889
172	Alanine	0.0	147	Glutamate	-1.831
78	Phenylalanine	0.0	64	Alanine	-1.831
77	Alanine	0.0	27	Glycine	-1.735
76	Glycine	0.0	49	Alanine	-1.677
237	Glycine	0.0	93	Threonine	-1.677
220	Serine	0.0	145	Isoleucine	-1.677
166	Tryptophan	0.0	199	Glycine	-1.677
69	Asparagine	0.0	196	Lysine	-1.677
68	Glutamine	0.0	208	Alanine	-1.677

Question 3. Write a program to compute the conservation score from MSA using unweighted frequency, and entropy, variance and sum of pairs-based measures.

Solution. The code for computing the conservation score from the MSA using the above mentioned techniques is given below:

```
# Creating a blosum62 matrix for sum of pairs measure
# The blosum matrix is read from a text file stored locally
# blosum_dict initializes the index for each amino acid in matrix
blosum_dict = {"A":0,"R":1,"N":2,"D":3,"C":4,"Q":5,"E":6,"G":7,"H":8,"I":9,
"L":10,"K":11,"M":12,"F":13,"P":14,"S":15,"T":16,"W":17,"Y":18,"V":19,"-":20}

blosum_matrix = []
with open("blosum62.txt", ’r’) as f:
    line = f.read()
    line_list = line.split("\n")

    for i in range(len(line_list)):
        line_i = line_list[i].split(" ")
        line_i = [int(value) for value in line_i if value != ""]
        blosum_matrix.append(line_i)
```

Listing 1. Blosum62 Matrix

```
import math

amino_acids_track =
{ "0": "G",  "1": "A",  "2": "V",  "3": "L",  "4": "I",
  "5": "T",  "6": "S",  "7": "M",  "8": "C",  "9": "P",
 "10": "F", "11": "Y", "12": "W", "13": "H", "14": "K",
 "15": "R", "16": "D", "17": "E", "18": "N", "19": "Q",
  "20": "-"}

amino_acids =
{"G":  0, "A":  1, "V":  2, "L":  3, "I":  4, "T":  5,
 "S":  6, "M":  7, "C":  8, "P":  9, "F": 10, "Y": 11,
 "W": 12, "H": 13, "K": 14, "R": 15, "D": 16, "E": 17,
 "N": 18, "Q": 19, "-": 20}

# Creating lists to store the conservation scores using the weighing scheme of unweighted amino acid frequencies with the three conservation calculation methods given in the question, entropy, variance and sum-of-pairs

entropy_score_list = []
variance_score_list = []
sum_pairs_score_list = []

# Iterating through entire aligned sequence
for i in range(len(seqs[0])):
    entropy_score = 0
    variance_score = 0
    sum_pairs_score = 0

    each_freq = [0 for k in range(21)]
    overall_freq = [0 for k in range(21)]
    total_non_aligns = 0
```

```python

    # Iterating through each column of sequences
    for j in range(len(seqs)):
        each_freq[amino_acids[seqs[j][i]]] += 1

    for j in range(len(each_freq)):
        each_freq[j] /= (len(seqs)-each_freq[20])

    for j in range(len(seqs)):
        for k in range(len(seqs[0])):
            overall_freq[amino_acids[seqs[j][k]]] += 1
            if seqs[j][k] == "-":
                total_non_aligns += 1

    # For the sake of variance based conservation score
    for j in range(len(overall_freq)):
        overall_freq[j] /= (len(seqs) * len(seqs[0]) - total_non_aligns)

    # Calculation of conservation score via the entropy method
    for j in range(len(each_freq)-1):
        if each_freq[j] != 0:
            entropy_score += each_freq[j] * math.log(each_freq[j])

    # Calculation of conservation score via the variance method
    for j in range(len(each_freq)-1):
        variance_score += abs(each_freq[j] - overall_freq[j])**2
    variance_score = variance_score**0.5

    # Calculation of conservation score via the sum-of-pairs method
    for j in range(len(each_freq)-1):
        for k in range(len(each_freq)-1):
            sum_pairs_score += (each_freq[j]*each_freq[k]*
            blosum_matrix[blosum_dict[amino_acids_track[str(j)]]]
                         [blosum_dict[amino_acids_track[str(k)]]])

    # Rounding off the values to 4 decimal places (just like AL2CO)
    entropy_score = round(entropy_score, 4)
    variance_score = round(variance_score, 4)
    sum_pairs_score = round(sum_pairs_score, 4)

    # Appending the scores for each column to the list
    entropy_score_list.append(entropy_score)
    variance_score_list.append(variance_score)
    sum_pairs_score_list.append(sum_pairs_score)

# Print the output lists
print(entropy_score_list)
print(variance_score_list)
print(sum_pairs_score_list)
```

Listing 2. Computing the different metric conservation scores

The above code generates the lists of entropy based, variance based, and sum-of-pairs based scores. The input sequences used were same as the one that is used in the Set 1 of the first question. The value of output generated by the code is compared against the ones generated by the AL2CO server. They match.

The input passed to the above code is the list of 11 sequences that were passed as input in the first question. They are: P69905, P01946, P01942, P01966, P01958, P01959, P01965, P06635, P60529, P80043 and P01980. These correspond to HBA_HUMAN, HBA_RAT, HBA_MOUSE, HBA_BOVIN, HBA_HORSE, HBA_EQUAS, HBA_PIG, HBA_PONPY, HBA_CANLF, HBA_TREBE, HBA_APTFO respectively. The inputs and their outputs are given below:

Input given to the above code

```
seqs[0] =
"MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP
VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR"
seqs[1] =
"MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-
DVSPGSAQVKAHGKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDP
VNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVSTVLTSKYR"
seqs[2] =
"MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-
DVSHGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDP
VNFKLLSHCLLVTLASHHPADFTPAVHASLDKFLASVSTVLTSKYR"
seqs[3] =
"MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDP
VNFKLLSHSLLVTLASHLPSDFTPAVHASLDKFLANVSTVLTSKYR"
seqs[4] =
"MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDP
VNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSKYR"
seqs[5] =
"MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDP
VNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSTVSTVLTSKYR"
seqs[6] =
"-VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-
NLSHGSDQVKAHGQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDP
VNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVSTVLTSKYR"
seqs[7] =
"MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKDHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP
VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR"
seqs[8] =
"-VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-
DLSPGSAQVKAHGKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDP
VNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVSTVLTSKYR"
seqs[9] =
"-SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWP
DVTPGSPHIKAHGKKVMGGIALAVSKIDDLKTGLMELSEQHAYKLRVDP
ANFKILNHCILVVISTMFPKEFTPEAHVSLDKFLSGVALALAERYR"
seqs[10] =
"MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-
DLHHGSAQVKAHGKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDP
VNFKLLGQCFMVVVAIHHPSALTPEIHASLDKFLCAVGNVLTSKYR"
```

Listing 3. Input given to the above code

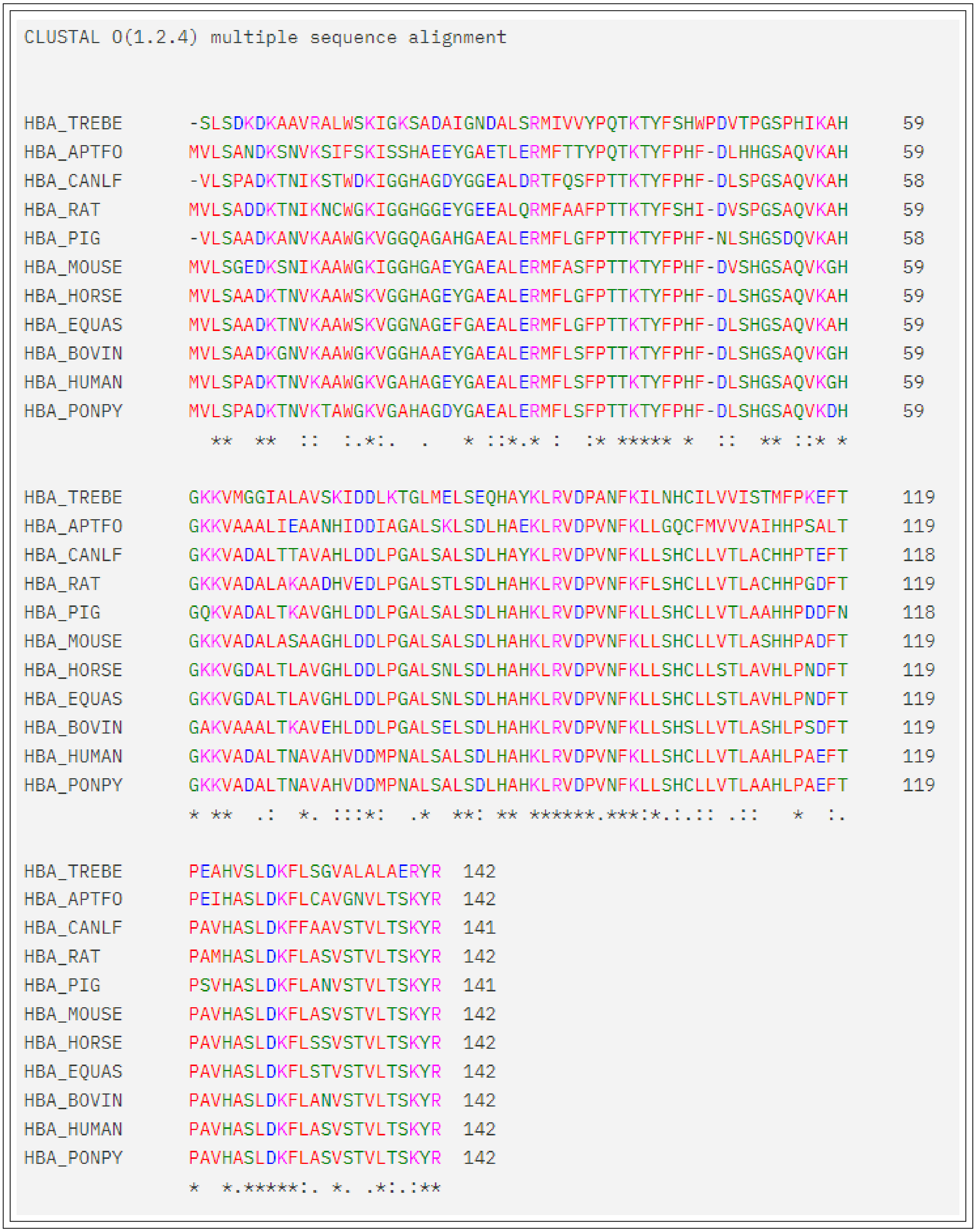
Output lists generated by the above code

```
# Entropy based conservation score
[0.0, -0.305, 0.0, 0.0, -1.121, -1.16, 0.0, 0.0, -1.169, -0.305,
-0.586, -0.305, -1.034, -1.16, -0.305, -0.916, 0.0, -0.689, -0.305,
-1.034, -0.886, -0.474, -1.034, -0.908, -0.886, 0.0, -0.886,
-0.305, -0.305, 0.0, -0.886, 0.0, -0.305, -0.305, -1.295, -1.367,
-0.474, 0.0, -0.474, 0.0, 0.0, 0.0, 0.0, 0.0, -0.474, 0.0, -0.6,
0.0, -0.305, -0.586, -0.6, -0.586, 0.0, 0.0, -0.6, -0.305, -0.305,
0.0, -0.86, 0.0, 0.0, -0.6, 0.0, 0.0, -0.76, -0.76, -0.305, -0.305,
-0.86, -1.673, 0.0, -0.586, -1.594, -0.305, -0.995, -0.305, 0.0,
-0.76, -0.6, -0.76, -0.305, 0.0, -0.305, -1.414, 0.0, 0.0, -0.305,
-0.305, 0.0, 0.0, -0.76, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -0.305, 0.0,
0.0, 0.0, -0.6, 0.0, -0.6, -0.305, -0.305, -0.6, -0.305, -0.474,
-0.474, -0.6, -0.305, -1.72, -0.305, -0.935, 0.0, -1.846, -0.916,
-0.305, -0.305, 0.0, -0.76, -0.886, 0.0, -0.305, 0.0, 0.0, 0.0,
0.0, 0.0, -0.305, -0.86, -1.414, 0.0, -0.6, -0.6, -0.305, 0.0,
-0.305, -0.305, -0.305, 0.0, 0.0]

# Variance based conservation score
[1.021, 0.865, 0.909, 0.953, 0.527, 0.561, 0.973, 0.956, 0.541,
0.917, 0.744, 0.871, 0.567, 0.56, 0.943, 0.612, 0.956, 0.691,
0.886, 0.613, 0.713, 0.74, 0.624, 0.665, 0.757, 0.977, 0.65, 0.922,
0.817, 0.909, 0.749, 1.015, 0.933, 0.904, 0.47, 0.473, 0.83, 0.993,
0.821, 0.977, 0.956, 0.977, 1.016, 0.988, 0.821, 0.969, 0.823,
0.993, 0.888, 0.674, 0.776, 0.745, 0.977, 0.953, 0.73, 0.936,
0.872, 0.956, 0.596, 0.969, 0.977, 0.777, 0.956, 0.956, 0.66,
0.704, 0.817, 0.825, 0.643, 0.36, 0.906, 0.701, 0.406, 0.879,
0.543, 0.888, 0.973, 0.675, 0.806, 0.733, 0.817, 0.909, 0.869,
0.451, 0.909, 0.953, 0.888, 0.826, 0.969, 0.906, 0.731, 0.956,
0.909, 1.015, 0.956, 0.973, 0.993, 0.86, 1.012, 0.988, 0.956,
0.739, 0.909, 0.781, 0.886, 0.937, 0.739, 0.825, 0.784, 0.806,
0.735, 0.814, 0.345, 0.885, 0.56, 0.993, 0.302, 0.633, 0.893,
0.891, 0.993, 0.66, 0.695, 0.969, 0.815, 0.953, 0.909, 0.973,
0.956, 0.988, 0.822, 0.595, 0.459, 0.956, 0.769, 0.797, 0.86,
0.909, 0.881, 0.868, 0.871, 1.016, 1.015]

# Sum-of-pairs based conservation score
[5.0, 3.008, 4.0, 4.0, 1.149, 1.157, 6.0, 5.0, 1.868, 4.661, 3.603,
4.504, 1.909, 1.521, 9.306, 2.264, 5.0, 3.504, 4.992, 2.405, 4.372,
2.876, 2.24, 2.455, 4.314, 6.0, 1.826, 4.512, 3.347, 4.0, 3.298,
5.0, 4.008, 4.992, 0.959, 1.132, 5.14, 7.0, 3.215, 5.0, 5.0, 5.0,
7.0, 6.0, 4.521, 8.0, 4.24, 7.0, 5.174, 2.81, 2.752, 3.959, 6.0,
4.0, 2.322, 4.198, 3.835, 5.0, 1.835, 8.0, 6.0, 3.405, 5.0, 4.0,
2.124, 2.694, 3.355, 3.669, 2.19, 0.124, 4.0, 2.413, 0.934, 6.488,
2.612, 5.331, 6.0, 3.14, 4.446, 3.149, 3.355, 4.0, 3.182, 0.678,
4.0, 4.0, 5.331, 3.017, 8.0, 4.0, 4.967, 5.0, 4.0, 5.0, 4.0, 6.0,
7.0, 3.339, 6.0, 6.0, 5.0, 3.058, 4.0, 2.926, 6.653, 7.306, 3.058,
3.678, 2.215, 3.479, 3.24, 3.504, 0.702, 6.322, 1.207, 7.0, 0.562,
3.008, 4.992, 4.182, 7.0, 2.182, 2.736, 8.0, 3.339, 4.0, 4.0, 6.0,
5.0, 6.0, 3.355, 2.289, 1.496, 4.0, 2.909, 3.231, 3.339, 4.0,
4.165, 3.347, 4.504, 7.0, 5.0]
```

Listing 4. The conversation scores output generated by the above code

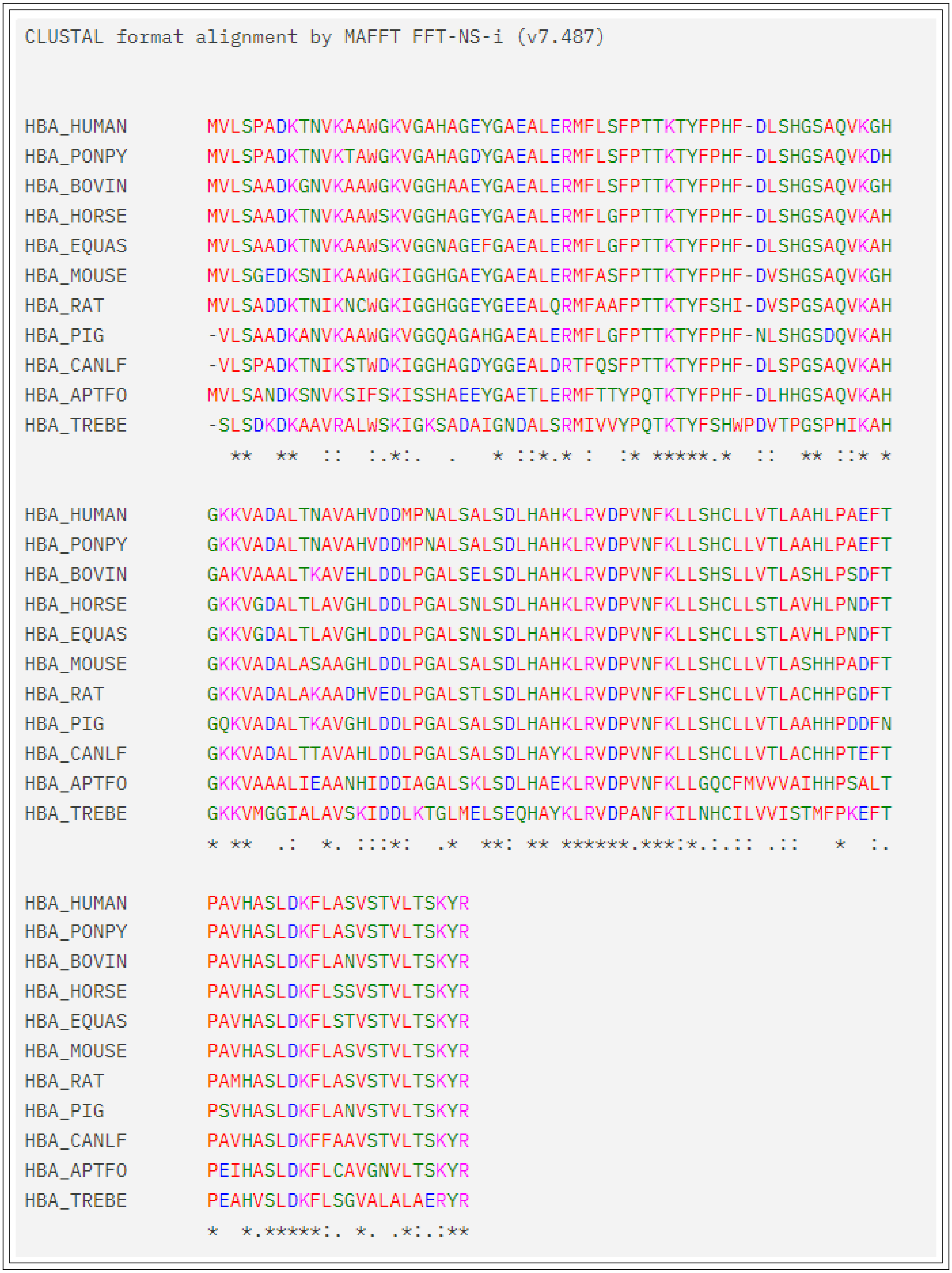
Question 4. Using the program written in Q3 (unweighted frequency and entropy-based measure), compare the MSA from Clustal Omega, MAFFT, and MUSCLE. Identify the residues with (i) similar and (ii) different conservation scores among the three alignment methods.

Solution. Below are the MSA obtained from the Clustal Omega, MAFFT, and MUSCLE for the protein sequences in set 1.



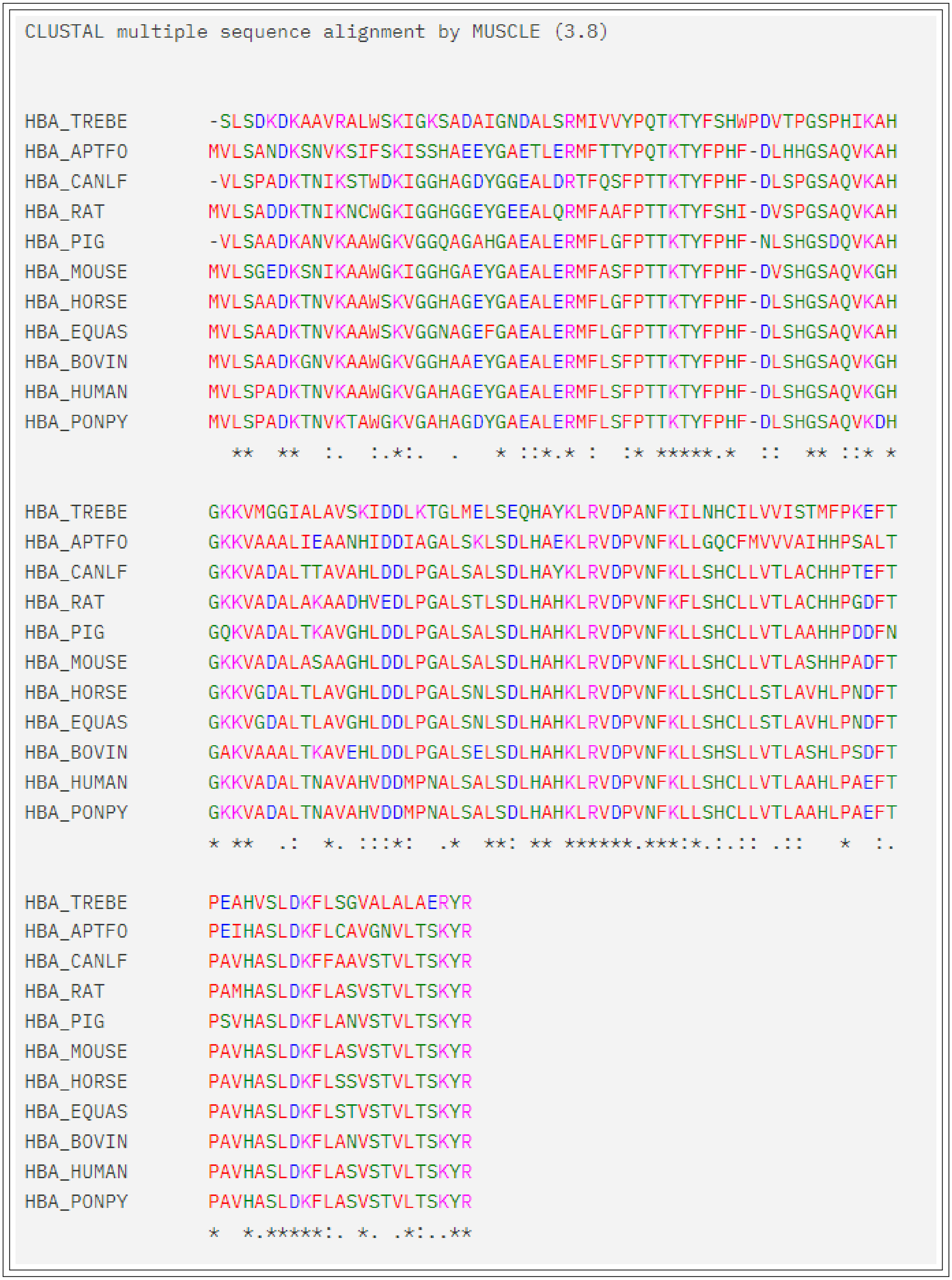
```
CLUSTAL O(1.2.4) multiple sequence alignment


HBA_TREBE      -SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAH     59
HBA_APTFO      MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-DLHHGSAQVKAH     59
HBA_CANLF      -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-DLSPGSAQVKAH     58
HBA_RAT        MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-DVSPGSAQVKAH     59
HBA_PIG        -VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-NLSHGSDQVKAH     58
HBA_MOUSE      MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-DVSHGSAQVKGH     59
HBA_HORSE      MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_EQUAS      MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH     59
HBA_BOVIN      MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_HUMAN      MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH     59
HBA_PONPY      MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKDH     59
                 **  **  ::  :.*:.  .   * ::*.* :  :* ***** *  ::  ** ::* *

HBA_TREBE      GKKVMGGIALAVSKIDDLKTGLMELSEQHAYKLRVDPANFKILNHCILVVISTMFPKEFT     119
HBA_APTFO      GKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALT     119
HBA_CANLF      GKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFT     118
HBA_RAT        GKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFT     119
HBA_PIG        GQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFN     118
HBA_MOUSE      GKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFT     119
HBA_HORSE      GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT     119
HBA_EQUAS      GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT     119
HBA_BOVIN      GAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT     119
HBA_HUMAN      GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT     119
HBA_PONPY      GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT     119
               * **  .:  *. :::*:  .*  **: ** ******.***:*.:.:: .::   *  :.

HBA_TREBE      PEAHVSLDKFLSGVALALAERYR  142
HBA_APTFO      PEIHASLDKFLCAVGNVLTSKYR  142
HBA_CANLF      PAVHASLDKFFAAVSTVLTSKYR  141
HBA_RAT        PAMHASLDKFLASVSTVLTSKYR  142
HBA_PIG        PSVHASLDKFLANVSTVLTSKYR  141
HBA_MOUSE      PAVHASLDKFLASVSTVLTSKYR  142
HBA_HORSE      PAVHASLDKFLSSVSTVLTSKYR  142
HBA_EQUAS      PAVHASLDKFLSTVSTVLTSKYR  142
HBA_BOVIN      PAVHASLDKFLANVSTVLTSKYR  142
HBA_HUMAN      PAVHASLDKFLASVSTVLTSKYR  142
HBA_PONPY      PAVHASLDKFLASVSTVLTSKYR  142
               *  *.*****:. *. .*:.:**
```

Figure 29. Output MSA by Clustal Omega for set 1



```
CLUSTAL format alignment by MAFFT FFT-NS-i (v7.487)


HBA_HUMAN       MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH
HBA_PONPY       MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKDH
HBA_BOVIN       MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH
HBA_HORSE       MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH
HBA_EQUAS       MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH
HBA_MOUSE       MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-DVSHGSAQVKGH
HBA_RAT         MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-DVSPGSAQVKAH
HBA_PIG         -VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-NLSHGSDQVKAH
HBA_CANLF       -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-DLSPGSAQVKAH
HBA_APTFO       MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-DLHHGSAQVKAH
HBA_TREBE       -SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAH
                  **  **  ::  :.*:.  .   * ::*.* :  :* *****.*  ::  ** ::* *

HBA_HUMAN       GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT
HBA_PONPY       GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT
HBA_BOVIN       GAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT
HBA_HORSE       GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
HBA_EQUAS       GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
HBA_MOUSE       GKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFT
HBA_RAT         GKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFT
HBA_PIG         GQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFN
HBA_CANLF       GKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFT
HBA_APTFO       GKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALT
HBA_TREBE       GKKVMGGIALAVSKIDDLKTGLMELSEQHAYKLRVDPANFKILNHCILVVISTMFPKEFT
                * **  .:  *. :::*:  .*  **: ** ******.***:*.:.:: .::   *  :.

HBA_HUMAN       PAVHASLDKFLASVSTVLTSKYR
HBA_PONPY       PAVHASLDKFLASVSTVLTSKYR
HBA_BOVIN       PAVHASLDKFLANVSTVLTSKYR
HBA_HORSE       PAVHASLDKFLSSVSTVLTSKYR
HBA_EQUAS       PAVHASLDKFLSTVSTVLTSKYR
HBA_MOUSE       PAVHASLDKFLASVSTVLTSKYR
HBA_RAT         PAMHASLDKFLASVSTVLTSKYR
HBA_PIG         PSVHASLDKFLANVSTVLTSKYR
HBA_CANLF       PAVHASLDKFFAAVSTVLTSKYR
HBA_APTFO       PEIHASLDKFLCAVGNVLTSKYR
HBA_TREBE       PEAHVSLDKFLSGVALALAERYR
                *  *.*****:. *. .*:.:**
```

Figure 30. Output MSA by MAFFT for set 1



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


HBA_TREBE       -SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAH
HBA_APTFO       MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-DLHHGSAQVKAH
HBA_CANLF       -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-DLSPGSAQVKAH
HBA_RAT         MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-DVSPGSAQVKAH
HBA_PIG         -VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-NLSHGSDQVKAH
HBA_MOUSE       MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-DVSHGSAQVKGH
HBA_HORSE       MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH
HBA_EQUAS       MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAH
HBA_BOVIN       MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH
HBA_HUMAN       MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGH
HBA_PONPY       MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKDH
                  **  **  :.  :.*:.  .   * ::*.* :  :* *****.*  ::  ** ::* *

HBA_TREBE       GKKVMGGIALAVSKIDDLKTGLMELSEQHAYKLRVDPANFKILNHCILVVISTMFPKEFT
HBA_APTFO       GKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALT
HBA_CANLF       GKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFT
HBA_RAT         GKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFT
HBA_PIG         GQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFN
HBA_MOUSE       GKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFT
HBA_HORSE       GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
HBA_EQUAS       GKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
HBA_BOVIN       GAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT
HBA_HUMAN       GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT
HBA_PONPY       GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT
                * **  .:  *. :::*:  .*  **: ** ******.***:*.:.:: .::   *  :.

HBA_TREBE       PEAHVSLDKFLSGVALALAERYR
HBA_APTFO       PEIHASLDKFLCAVGNVLTSKYR
HBA_CANLF       PAVHASLDKFFAAVSTVLTSKYR
HBA_RAT         PAMHASLDKFLASVSTVLTSKYR
HBA_PIG         PSVHASLDKFLANVSTVLTSKYR
HBA_MOUSE       PAVHASLDKFLASVSTVLTSKYR
HBA_HORSE       PAVHASLDKFLSSVSTVLTSKYR
HBA_EQUAS       PAVHASLDKFLSTVSTVLTSKYR
HBA_BOVIN       PAVHASLDKFLANVSTVLTSKYR
HBA_HUMAN       PAVHASLDKFLASVSTVLTSKYR
HBA_PONPY       PAVHASLDKFLASVSTVLTSKYR
                *  *.*****:. *. .*:..**
```

FIGURE 31. Output MSA by MUSCLE for set 1

Now, I will be putting these MSA aligned sequences (in the CLUSTAl format) to the code from question 3 to compute the given three score for each of these MSA alignments.
I used the following code to check for covert the MSA obtained by CLUSTAL OMEGA, MAFFT, and MUSCLE into their respective dataframes.

```
import pandas as pd

# Reading the CLUSTAL Omega MSA file
clustal = []
with open("clustal.txt", ’r’) as f:
    line = f.read()
    line_list = line.split("\n")

    for i in range(len(line_list)):
        line_i = line_list[i].split(" ")
        line_i = [str(value) for value in line_i if value != ""]
        clustal.append(line_i)

df1 = pd.DataFrame(clustal)
df1[2] = pd.to_numeric(df1[2])
df1[0] = pd.to_numeric(df1[0])

# Reading the MAFFT MSA file
mafft = []
with open("mafft.txt", ’r’) as f:
    line = f.read()
    line_list = line.split("\n")

    for i in range(len(line_list)):
        line_i = line_list[i].split(" ")
        line_i = [str(value) for value in line_i if value != ""]
        mafft.append(line_i)

df2 = pd.DataFrame(mafft)
df2[2] = pd.to_numeric(df2[2])
df2[0] = pd.to_numeric(df2[0])

# Reading the MUSCLE MSA file
muscle = []
with open("muscle.txt", ’r’) as f:
    line = f.read()
    line_list = line.split("\n")

    for i in range(len(line_list)):
        line_i = line_list[i].split(" ")
        line_i = [str(value) for value in line_i if value != ""]
        muscle.append(line_i)

df3 = pd.DataFrame(muscle)
df3[2] = pd.to_numeric(df3[2])
df3[0] = pd.to_numeric(df3[0])
```

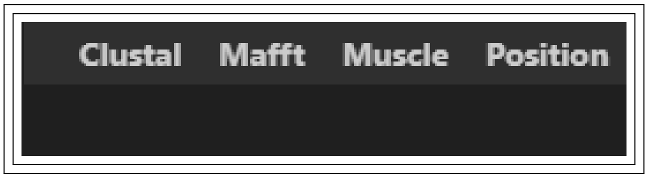
Listing 5. Code to create the dataframes for different MSAs

Now that I have the three dataframes for the MSA obtained by each of the alignment algorithms CLUSTAL OMEGA, MAFFT, and MUSCLE, I use the following code to list the similarities and differences.

```python
similarities_3 = [] # All three same residue same position
differences_2 = [] # Any two different residue same position
differences_3 = [] # All three different residue same position

for i in range(len(df1)):
  # Condition for checking any two sequences with different residues at same position
  if df1[2][i]!=df2[2][i] or df1[2][i]!=df3[2][i] or df2[2][i]!=df3[2][i]:
    print("Differences_2")
    print(df1[2][i],df1[3][i],df2[2][i],df2[3][i],df3[2][i],df3[3][i],i)
    differences_2.append([df1[2][i], df2[2][i], df3[2][i], i+1])

  # Condition for checking all three sequences with different residues at same position
  if df1[2][i]!=df2[2][i] and df1[2][i]!=df3[2][i] and df2[2][i]!=df3[2][i]:
    print("Differences_3")
    print(df1[2][i],df1[3][i],df2[2][i],df2[3][i],df3[2][i],df3[3][i],i)
    differences_3.append([df1[2][i], df2[2][i], df3[2][i], i+1])

  # Condtion for checking all three sequences with similar residues at same position
  if df1[2][i] == df2[2][i] and df1[2][i] == df3[2][i]:
    print("Similarities_3")
    print(df1[2][i],df1[3][i],df2[2][i],df2[3][i],df3[2][i],df3[3][i],i)
    similarities_3.append([df1[2][i], df2[2][i], df3[2][i], i+1])

# Converting all of these lists into dataframes for easy visualization
df1 = pd.DataFrame(similarities_3, columns=["Clustal", "Mafft", "Muscle", "Position"])
df2 = pd.DataFrame(differences_3, columns=["Clustal", "Mafft", "Muscle", "Position"])
df3 = pd.DataFrame(differences_2, columns=["Clustal", "Mafft", "Muscle", "Position"])
```

Listing 6. Code to find the similarities and differences in MSA

When I passed the above set 1 MSA, obtained by all three methods, to the code, unfortunately, all the differences dataframes were empty. This is indicative of the fact that all three MSA algorithms of CLUSTAL OMEGA, MAFFT, and MUSCLE aligned the given 11 sequences in set 1 in the exact same manner.



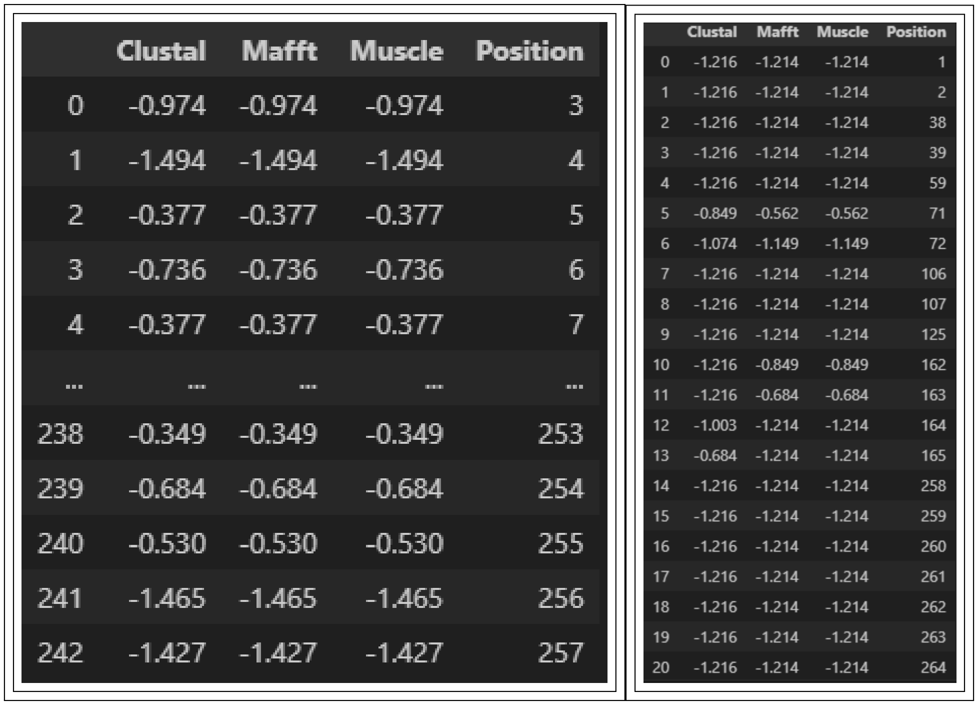
Clustal	Mafft	Muscle	Position

Figure 32. Output of differences_3 and differences_2

This is indicative of the fact that all the positions are same in all the three alignment techniques for the proteins in set 1. The similarities are:

Position	Clustal Residue	Mafft Residue	Muscle Residue	Score
3	Leucine	Leucine	Leucine	0.0
5	Proline	Proline	Proline	-1.121
143	Argenine	Argenine	Argenine	0.0

When I passed the above set 2 MSA, obtained by all three methods, to the code, unfortunately, the differences_3 dataframe was empty, however, the dataframes differences_2 and similarities_3 were not empty, indicating that differences and similarities exist upto some extent. This is indicative of the fact that all three MSA algorithms of CLUSTAL OMEGA, MAFFT, and MUSCLE do not align the given 9 sequences in set 2 in the exact same manner.



	Clustal	Mafft	Muscle	Position
0	-0.974	-0.974	-0.974	3
1	-1.494	-1.494	-1.494	4
2	-0.377	-0.377	-0.377	5
3	-0.736	-0.736	-0.736	6
4	-0.377	-0.377	-0.377	7
...	...	...	...	...
238	-0.349	-0.349	-0.349	253
239	-0.684	-0.684	-0.684	254
240	-0.530	-0.530	-0.530	255
241	-1.465	-1.465	-1.465	256
242	-1.427	-1.427	-1.427	257

(A) similarities_3

	Clustal	Mafft	Muscle	Position
0	-1.216	-1.214	-1.214	1
1	-1.216	-1.214	-1.214	2
2	-1.216	-1.214	-1.214	38
3	-1.216	-1.214	-1.214	39
4	-1.216	-1.214	-1.214	59
5	-0.849	-0.562	-0.562	71
6	-1.074	-1.149	-1.149	72
7	-1.216	-1.214	-1.214	106
8	-1.216	-1.214	-1.214	107
9	-1.216	-1.214	-1.214	125
10	-1.216	-0.849	-0.849	162
11	-1.216	-0.684	-0.684	163
12	-1.003	-1.214	-1.214	164
13	-0.684	-1.214	-1.214	165
14	-1.216	-1.214	-1.214	258
15	-1.216	-1.214	-1.214	259
16	-1.216	-1.214	-1.214	260
17	-1.216	-1.214	-1.214	261
18	-1.216	-1.214	-1.214	262
19	-1.216	-1.214	-1.214	263
20	-1.216	-1.214	-1.214	264

(B) differences_2

FIGURE 33. The two non empty dataframes in MSA comparison for set 2

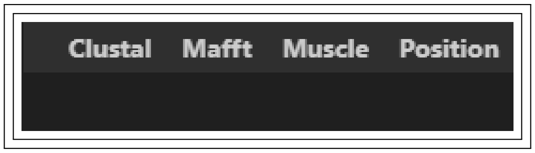
Another observation is that the differences observed in differences_2 is only the differences between either **mafft** or **muscle** and **clustal**. The mafft and muscle scores are exactly same throughout indicating their same alignment.
The similarities_3 are:

Position	Clustal Residue	Mafft Residue	Muscle Residue	Score
6	Lysine	Lysine	Lysine	-0.736
7	Phenylalanine	Phenylalanine	Phenylalanine	-0.377
255	Alanine	Alanine	Alanine	-0.530

The differences_2 are:

Position	Clustal	Score	Mafft	Score	Muscle	Score
71	Tyrosine	-0.849	Tyrosine	-0.562	Tyrosine	-0.562
72	Lysine	-1.074	Lysine	-1.149	Lysine	-1.149
162	Gap	*	Lysine	-0.849	Lysine	-0.849

The differences_2 are a total of 20 in number. The remaining are basically due to gap scores of clustal technique not aligning perfectly with the gap score of mafft and muscle techniques.
Also, there are no occurences of differences_3. The dataframe is empty.

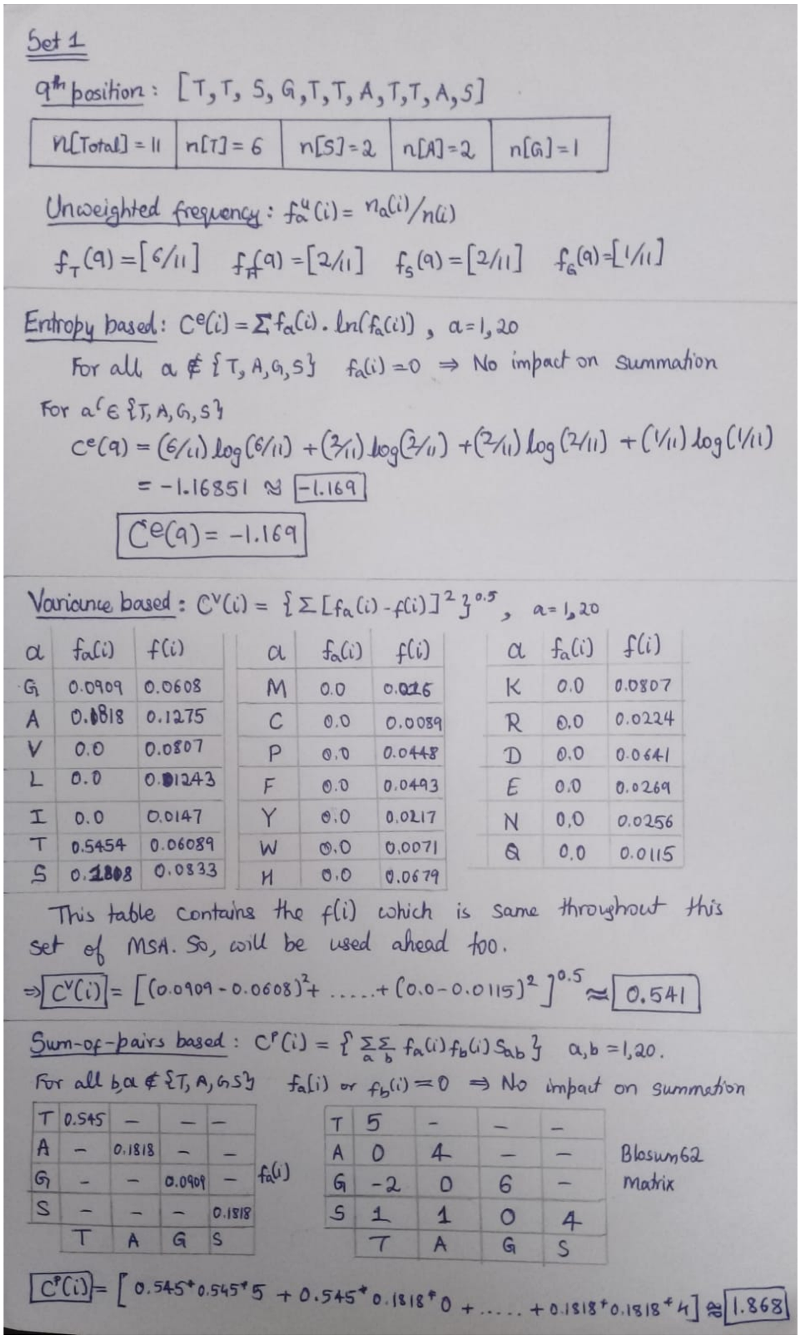


Clustal	Mafft	Muscle	Position

FIGURE 34. Output of differences_3 for seq 2

Question 5. Check the scores manually at positions 9, 11, 20, 22 and 30 (use MSA from Clustal Omega)

Solution. Below are the images of manual calculation of scores. The scores calculated are unweighted frequencies, with scoring computational methods, entropy based, variance based, and sum-of-pairs based measures.
Also, in the sum-of-pairs method, after the entire summation, it needs to be squared and then square root should to be taken. This measure is to ensure that the value is positive.
Also, I saw a research paper on AL2CO that also had omitted the squaring and square rooting component in the sum-of-pairs method. Along with that, none of the values I computed showed a negative sign, hence I have dropped the squaring and square rooting component and computed the values. Below are the images of the manual calculations.



Set 1

9th position : [T, T, S, G, T, T, A, T, T, A, S]

n[Total] = 11	n[T] = 6	n[S] = 2	n[A] = 2	n[G] = 1

Unweighted frequency : $f_a^u(i) = n_a(i)/n(i)$

$f_T(9) = [6/11]$ $f_A(9) = [2/11]$ $f_S(9) = [2/11]$ $f_G(9) = [1/11]$

Entropy based : $C^e(i) = \Sigma f_a(i) . ln(f_a(i))$, $a = 1, 20$

For all $a \notin \{T, A, G, S\}$ $f_a(i) = 0 \Rightarrow$ No impact on summation

For $a \in \{T, A, G, S\}$

$$C^e(9) = (6/11) \log (6/11) + (2/11) \log (2/11) + (2/11) \log (2/11) + (1/11) \log (1/11)$$

$$= -1.16851 \approx \boxed{-1.169}$$

$$\boxed{C^e(9) = -1.169}$$

Variance based : $C^v(i) = \{ \Sigma [f_a(i) - f(i)]^2 \}^{0.5}$, $a = 1, 20$

a	$f_a(i)$	$f(i)$	a	$f_a(i)$	$f(i)$	a	$f_a(i)$	$f(i)$
G	0.0909	0.0608	M	0.0	0.025	K	0.0	0.0807
A	0.1818	0.1275	C	0.0	0.0089	R	0.0	0.0224
V	0.0	0.0807	P	0.0	0.0448	D	0.0	0.0641
L	0.0	0.1243	F	0.0	0.0493	E	0.0	0.0269
I	0.0	0.0147	Y	0.0	0.0217	N	0.0	0.0256
T	0.5454	0.06089	W	0.0	0.0071	Q	0.0	0.0115
S	0.1808	0.0833	H	0.0	0.0679			

This table contains the f(i) which is same throughout this set of MSA. So, will be used ahead too.

$$\Rightarrow \boxed{C^v(i)} = [(0.0909 - 0.0608)^2 + \ldots\ldots + (0.0 - 0.0115)^2]^{0.5} \approx \boxed{0.541}$$

Sum-of-pairs based : $C^p(i) = \{ \sum_a \sum_b f_a(i) f_b(i) S_{ab} \}$ $a, b = 1, 20.$

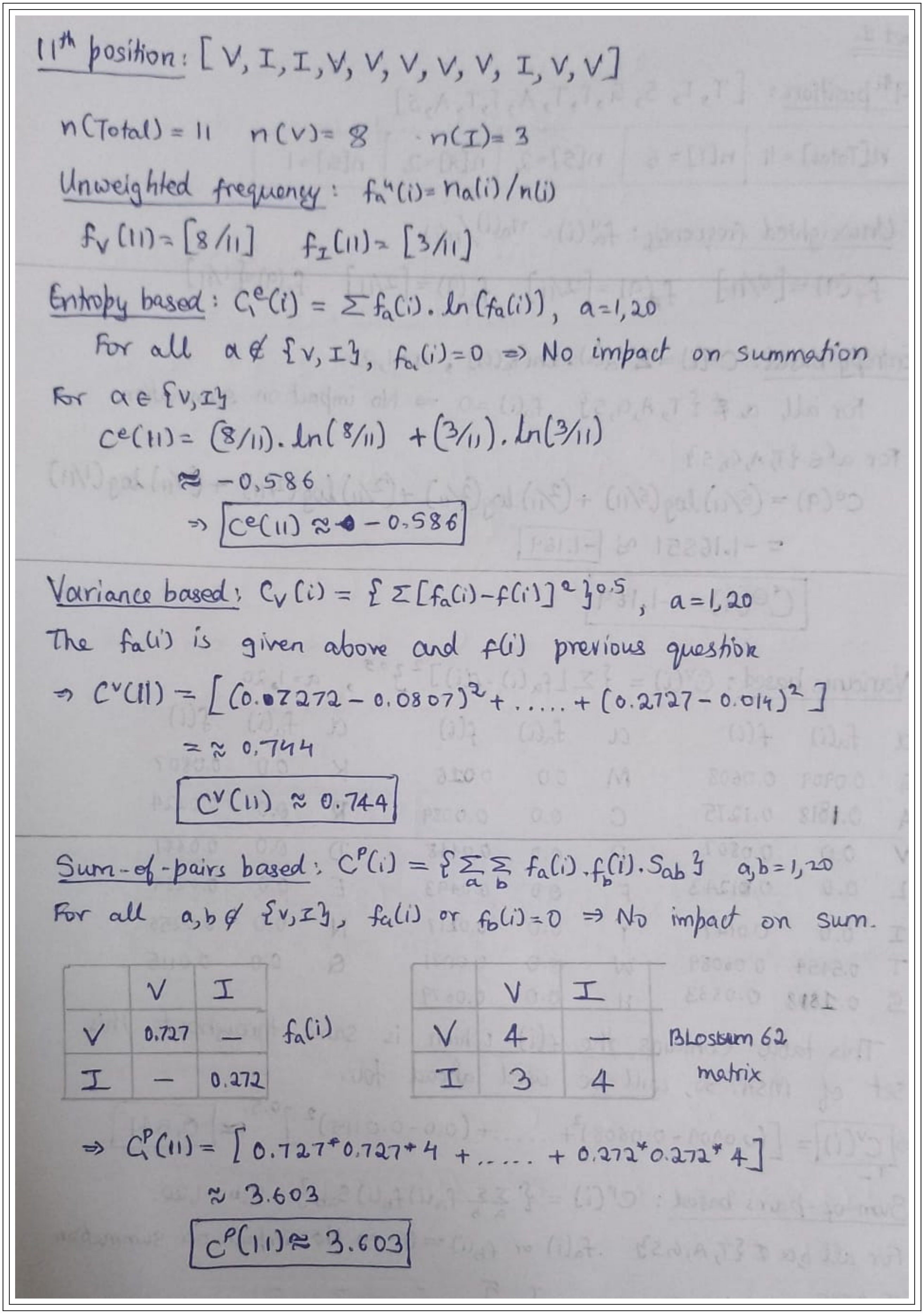
For all $b, a \notin \{T, A, G, S\}$ $f_a(i)$ or $f_b(i) = 0 \Rightarrow$ No impact on summation

$$\begin{array}{c|c|c|c|c} T & 0.545 & - & - & - \\ A & - & 0.1818 & - & - \\ G & - & - & 0.0909 & - \\ S & - & - & - & 0.1818 \\ \hline & T & A & G & S \end{array} \quad f_a(i) \qquad \begin{array}{c|c|c|c|c} T & 5 & - & - & - \\ A & 0 & 4 & - & - \\ G & -2 & 0 & 6 & - \\ S & 1 & 1 & 0 & 4 \\ \hline & T & A & G & S \end{array}$$

Blosum62 Matrix

$$\boxed{C^p(i)} = [0.545 * 0.545 * 5 + 0.545 * 0.1818 * 0 + \ldots\ldots + 0.1818 * 0.1818 * 4] \approx \boxed{1.868}$$

FIGURE 35. Calculations for 9th position of set 1



11th position: [V, I, I, V, V, V, V, V, I, V, V]

n(Total) = 11   n(V) = 8   n(I) = 3

Unweighted frequency: $f_a^u(i) = n_a(i)/n(i)$

$f_V(11) = [8/11]$ $f_I(11) = [3/11]$

Entropy based: $C_1^e(i) = \Sigma f_a(i) . \ln(f_a(i))$, $a = 1, 20$

For all $a \notin \{V, I\}$, $f_a(i) = 0$ ⇒ No impact on summation

For $a \in \{V, I\}$

$C^e(11) = (8/11) . \ln(8/11) + (3/11) . \ln(3/11)$

≈ -0.586

⇒ $\boxed{C^e(11) \approx -0.586}$

Variance based: $C_V(i) = \{\Sigma[f_a(i) - f(i)]^2\}^{0.5}$, $a = 1, 20$

The $f_a(i)$ is given above and $f(i)$ previous question

⇒ $C^v(11) = [(0.7272 - 0.0807)^2 + \dots + (0.2727 - 0.014)^2]$

$= \approx 0.744$

$\boxed{C^v(11) \approx 0.744}$

Sum-of-pairs based: $C^p(i) = \{\sum_a \sum_b f_a(i) . f_b(i) . S_{ab}\}$ $a, b = 1, 20$

For all $a, b \notin \{V, I\}$, $f_a(i)$ or $f_b(i) = 0$ ⇒ No impact on sum.

	V	I
V	0.727	—
I	—	0.272

$f_a(i)$

	V	I
V	4	—
I	3	4

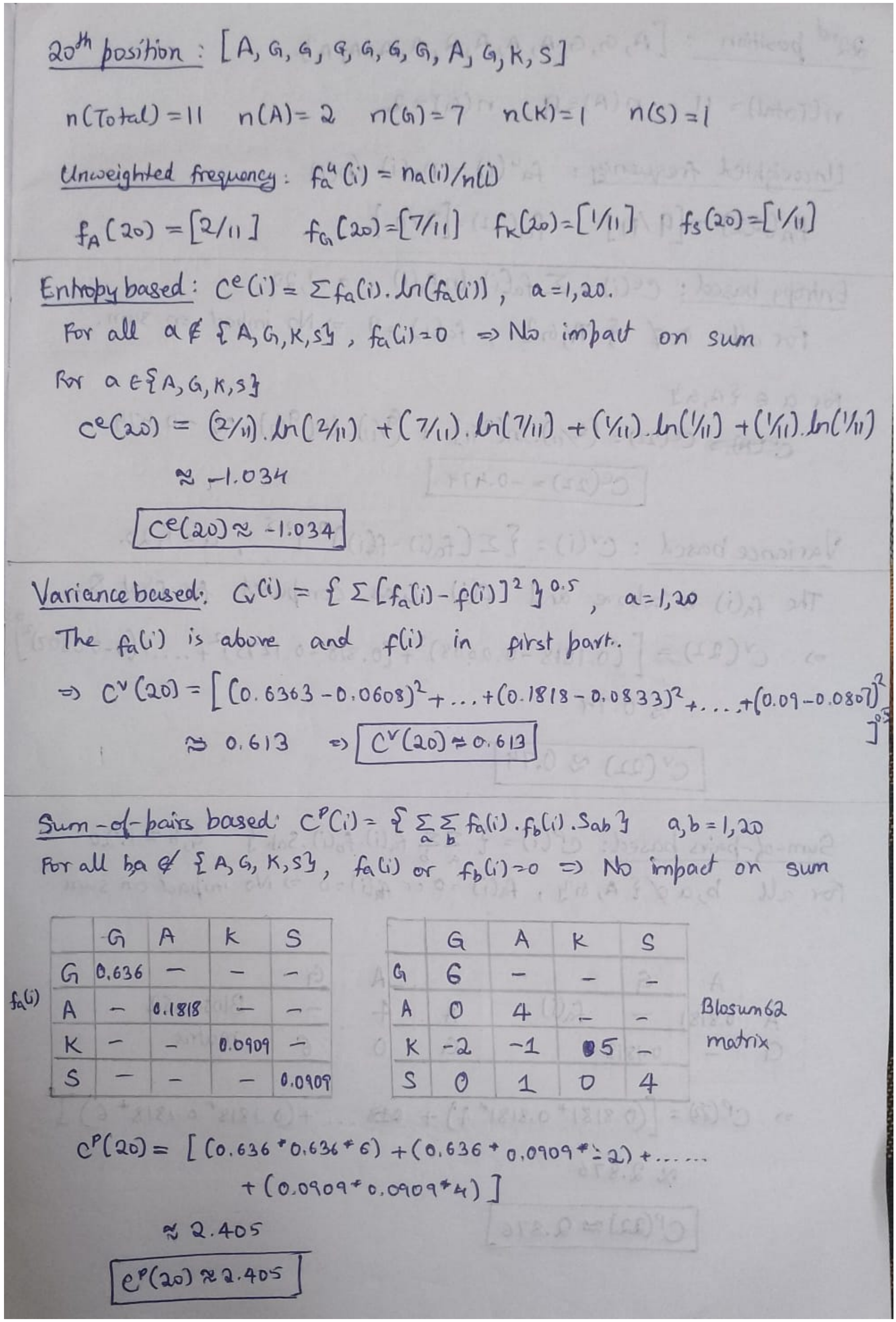
BLosum 62 matrix

⇒ $C_1^p(11) = [0.727 * 0.727 * 4 + \dots + 0.272 * 0.272 * 4]$

≈ 3.603

$\boxed{C^p(11) \approx 3.603}$

Figure 36. Calculations for 11th position of set 1



<u>20th position</u> : [A, G, G, G, G, G, G, A, G, K, S]

n(Total) = 11 n(A) = 2 n(G) = 7 n(K) = 1 n(S) = 1

<u>Unweighted frequency</u> : $f_a^u(i) = n_a(i)/n(i)$

$f_A(20) = [2/11]$ $f_G(20) = [7/11]$ $f_K(20) = [1/11]$ $f_S(20) = [1/11]$

<u>Entropy based</u> : $C^e(i) = \Sigma f_a(i) . ln(f_a(i))$, a = 1, 20.

For all $a \notin \{A, G, K, S\}$, $f_a(i) = 0$ ⇒ No impact on sum

For $a \in \{A, G, K, S\}$

$C^e(20) = (2/11).ln(2/11) + (7/11).ln(7/11) + (1/11).ln(1/11) + (1/11).ln(1/11)$

≈ -1.034

$\boxed{C^e(20) \approx -1.034}$

<u>Variance based</u>: $C_v(i) = \{ \Sigma [f_a(i) - f(i)]^2 \}^{0.5}$, a = 1, 20

The $f_a(i)$ is above and $f(i)$ in first part.

$\Rightarrow C^v(20) = [(0.6363 - 0.0608)^2 + \ldots + (0.1818 - 0.0833)^2 + \ldots + (0.09 - 0.0807)^2]^{0.5}$

≈ 0.613 ⇒ $\boxed{C^v(20) \approx 0.613}$

<u>Sum-of-pairs based</u>: $C^P(i) = \{ \sum_a \sum_b f_a(i) . f_b(i) . S_{ab} \}$ a, b = 1, 20

For all b,a $\notin$ {A, G, K, S}, $f_a(i)$ or $f_b(i) = 0$ ⇒ No impact on sum

$f_a(i)$

	G	A	K	S
G	0.636	–	–	–
A	–	0.1818	–	–
K	–	–	0.0909	–
S	–	–	–	0.0909

	G	A	K	S
G	6	–	–	–
A	0	4	–	–
K	-2	-1	5	–
S	0	1	0	4

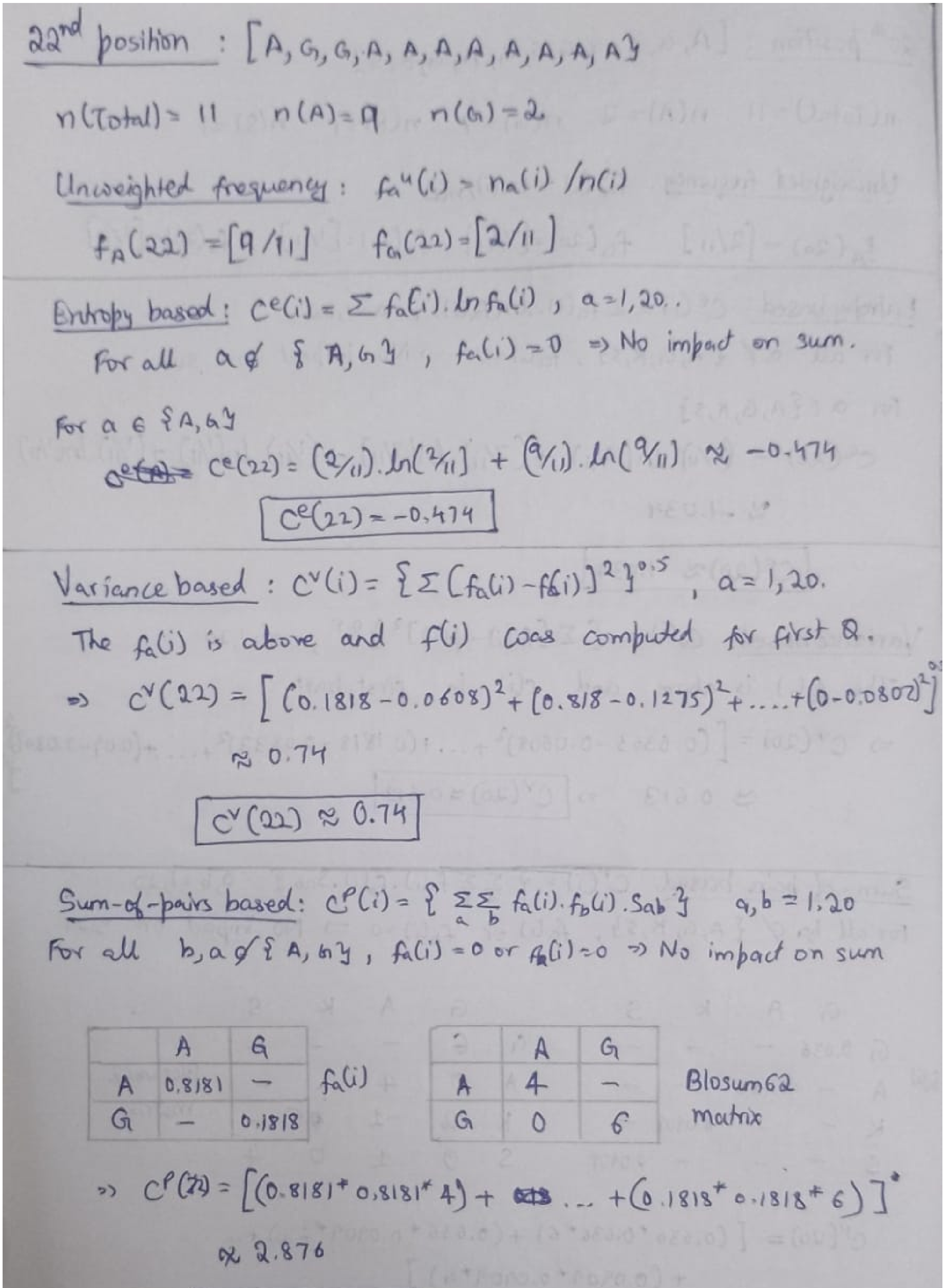
Blosum62 matrix

$C^P(20) = [(0.636 * 0.636 * 6) + (0.636 * 0.0909 * -2) + \ldots$
$+ (0.0909 * 0.0909 * 4)]$

≈ 2.405

$\boxed{C^P(20) \approx 2.405}$

FIGURE 37. Calculations for 20th position of set 1



22nd position : [A, G, G, A, A, A, A, A, A, A, A}

$n(\text{Total}) = 11 \quad n(A) = 9 \quad n(G) = 2$

Unweighted frequency : $f_a^u(i) = n_a(i) / n(i)$

$f_A(22) = [9/11] \quad f_G(22) = [2/11]$

Entropy based : $c^e(i) = \Sigma f_a(i) . \ln f_a(i)$, $a = 1, 20$.

For all $a \notin \{A, G\}$, $f_a(i) = 0 \Rightarrow$ No impact on sum.

For $a \in \{A, G\}$

$c^e(22) = (2/11) . \ln(2/11] + (9/11) . \ln(9/11) \approx -0.474$

$$\boxed{c^e(22) = -0.474}$$

Variance based : $c^v(i) = \{\Sigma [f_a(i) - f_a(i)]^2\}^{0.5}$, $a = 1, 20$.

The $f_a(i)$ is above and $f(i)$ was computed for first Q.

$\Rightarrow c^v(22) = [(0.1818 - 0.0608)^2 + (0.818 - 0.1275)^2 + \ldots + (0 - 0.0802)^2]^{0.5}$

≈ 0.74

$$\boxed{c^v(22) \approx 0.74}$$

Sum-of-pairs based : $c^p(i) = \{\sum_a \sum_b f_a(i) . f_b(i) . S_{ab}\}$ $a, b = 1, 20$

For all $b, a \notin \{A, G\}$, $f_a(i) = 0$ or $f_b(i) = 0 \Rightarrow$ No impact on sum

	A	G
A	0.8181	–
G	–	0.1818

$f_a(i)$

	A	G
A	4	–
G	0	6

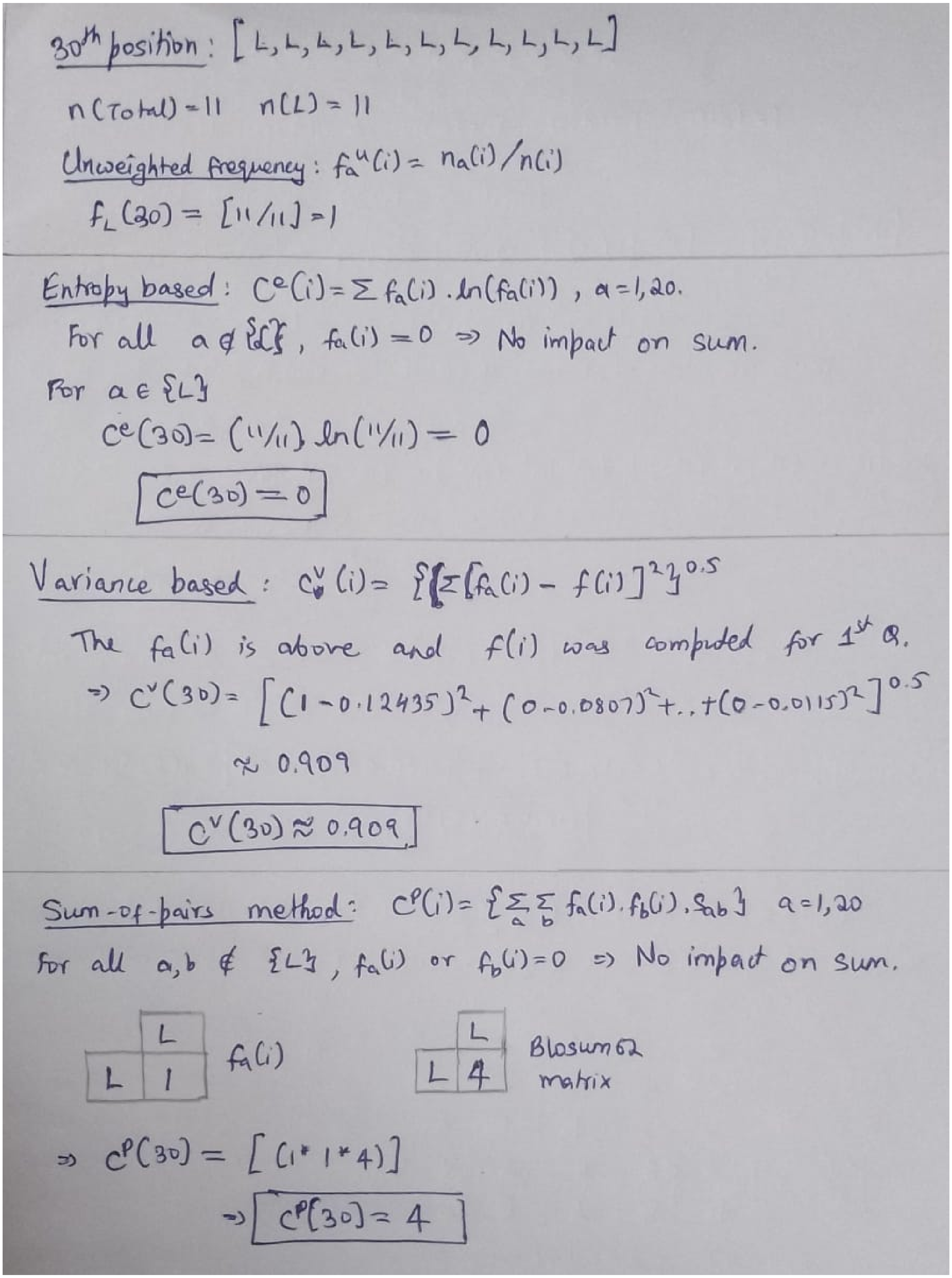
Blosum62 matrix

$\Rightarrow c^p(22) = [(0.8181 * 0.8181 * 4) + \ldots + (0.1818 * 0.1818 * 6)]$

≈ 2.876

$$\boxed{c^p(22) \approx 2.876}$$

Figure 38. Calculations for 22th position of set 1



30th position : [L, L, L, L, L, L, L, L, L, L, L]

n(Total) = 11   n(L) = 11

Unweighted frequency : $f_a^u(i) = n_a(i)/n(i)$

$f_L(30) = [11/11] = 1$

Entropy based : $C^e(i) = \Sigma f_a(i) . \ln(f_a(i))$, a = 1, 20.

For all $a \notin \{L\}$, $f_a(i) = 0 \Rightarrow$ No impact on sum.

For $a \in \{L\}$

$C^e(30) = (11/11) \ln(11/11) = 0$

$\boxed{C^e(30) = 0}$

Variance based : $C^v(i) = \{[\Sigma [f_a(i) - f(i)]^2\}^{0.5}$

The $f_a(i)$ is above and $f(i)$ was computed for 1st Q.

$\Rightarrow C^v(30) = [(1-0.12435)^2 + (0-0.0807)^2 + .. + (0-0.0115)^2]^{0.5}$

≈ 0.909

$\boxed{C^v(30) \approx 0.909}$

Sum-of-pairs method : $C^p(i) = \{\sum_a \sum_b f_a(i) . f_b(i) . S_{ab}\}$ a = 1, 20

For all $a, b \notin \{L\}$, $f_a(i)$ or $f_b(i) = 0 \Rightarrow$ No impact on sum.

	L
L	1

$f_a(i)$

	L
L	4

Blosum62 matrix

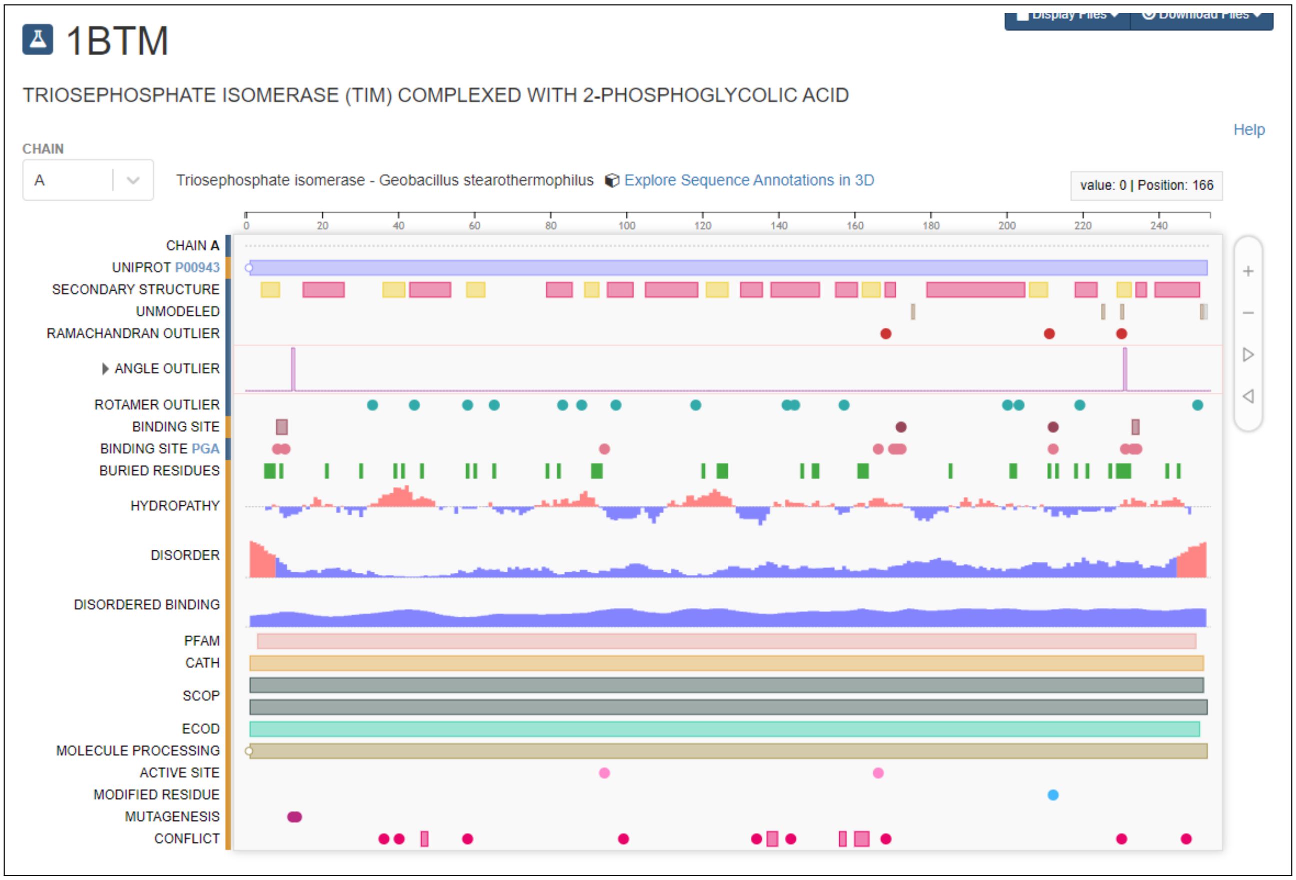
$\Rightarrow C^p(30) = [(1 * 1 * 4)]$

$\Rightarrow \boxed{C^p[30] = 4}$

Figure 39. Calculations for 30th position of set 1

Question 6. Obtain the conservation score of 1BTM, A-chain using Consurf server https://consurf.tau.ac.il/

Solution. First thing I need to find is the UniProt ID. I searched for the 1BTM, A chain. The below is the structure I found along with the accession number of the protein. The accession number of the protein is **P00943**. I searched for the same on the UniProtKB and found the similarity in structure, and other features and structural elements. Below is the structure I found initially.



1BTM

TRIOSEPHOSPHATE ISOMERASE (TIM) COMPLEXED WITH 2-PHOSPHOGLYCOLIC ACID

CHAIN A

Triosephosphate isomerase - Geobacillus stearothermophilus

value: 0 | Position: 166

FIGURE 40. Finding the protein sequence UniProt ID

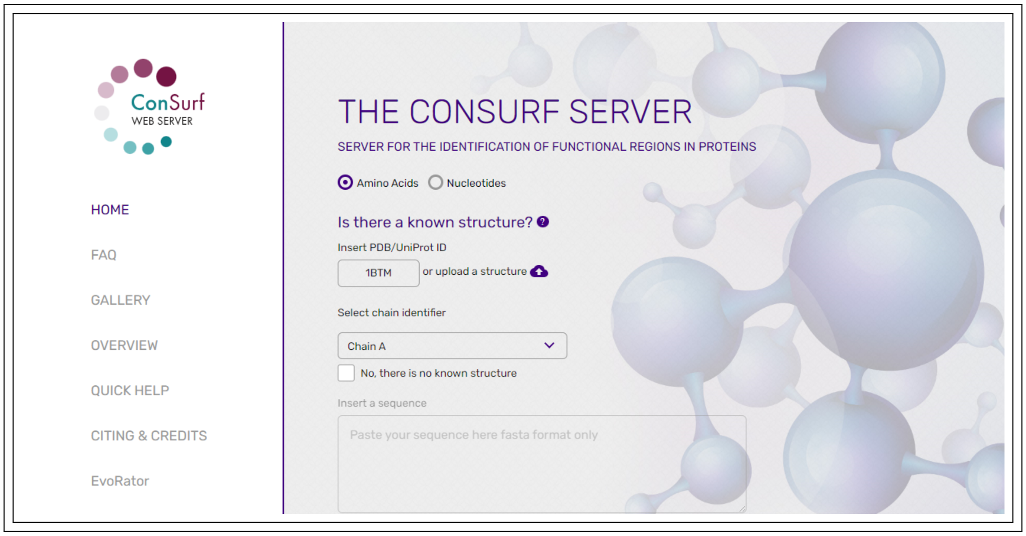
Protein sequence of 1BTM A-chain obtained from UniProtKB

```
>P00943
MRKPIIAGNWKMHKTLAEAVQFVEDVKGHVPPADEVISVVCAPFLFLDRLVQAAD
GTDLKIGAQTMHFADQGAYTGEVSPVMLKDLGVTYVILGHSERRQMFAETDETVN
KKVLAAFTRGLIPIICCGESLEEREAGQTNAVVASQVEKALAGLTPEQVKQAVIA
YEPIWAIGTGKSSTPEDANSVCGHIRSVVSRLFGPEAAEAIRIQYGGSVKPDNIR
DFLAQQQIDGPLVGGASLEPASFLQLVEAGRHE
```

LISTING 7. The protein sequence of the protein mentioned in the question

The above FASTA file is given as input to the **Consurf server**.
There also exists an alternative approach to execute the above given task. In the input page of the Consurf server, there is an option to enter the **PDB or UniProt ID**. Enter the term **1BTM** there. Then under the section of **Select the chain identifier**, choose **Chain A**. Now enter the job tite and email ID and **Run with default parameters**.

Below are two images, one for the input screen on the Consurf server, the other is the running parameters on Consurf server.


ConSurf WEB SERVER

HOME

FAQ

GALLERY

OVERVIEW

QUICK HELP

CITING & CREDITS

EvoRator

THE CONSURF SERVER

SERVER FOR THE IDENTIFICATION OF FUNCTIONAL REGIONS IN PROTEINS

● Amino Acids ○ Nucleotides

Is there a known structure?

Insert PDB/UniProt ID

1BTM or upload a structure

Select chain identifier

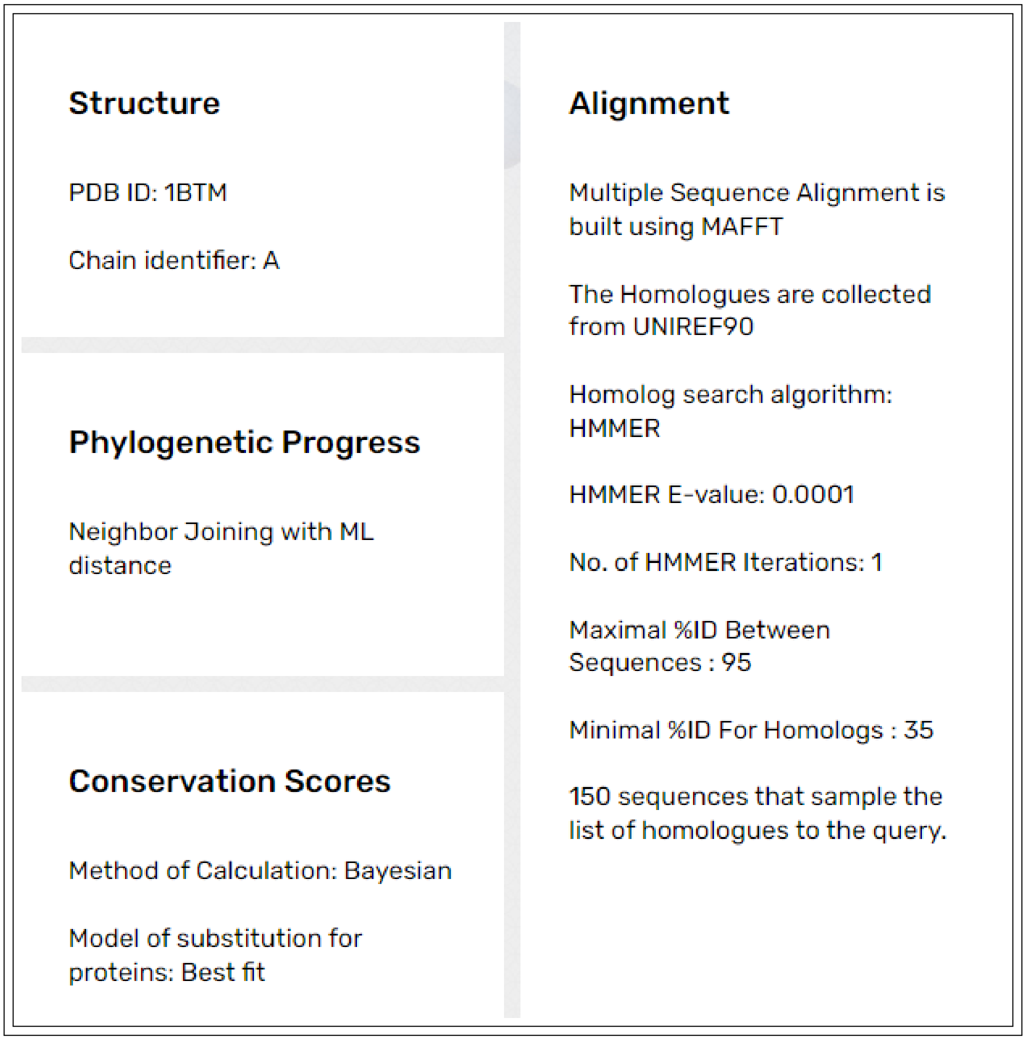
Chain A

□ No, there is no known structure

Insert a sequence

Paste your sequence here fasta format only

FIGURE 41. Input screen on the Consurf server


Structure

PDB ID: 1BTM

Chain identifier: A

Phylogenetic Progress

Neighbor Joining with ML distance

Conservation Scores

Method of Calculation: Bayesian

Model of substitution for proteins: Best fit

Alignment

Multiple Sequence Alignment is built using MAFFT

The Homologues are collected from UNIREF90

Homolog search algorithm: HMMER

HMMER E-value: 0.0001

No. of HMMER Iterations: 1

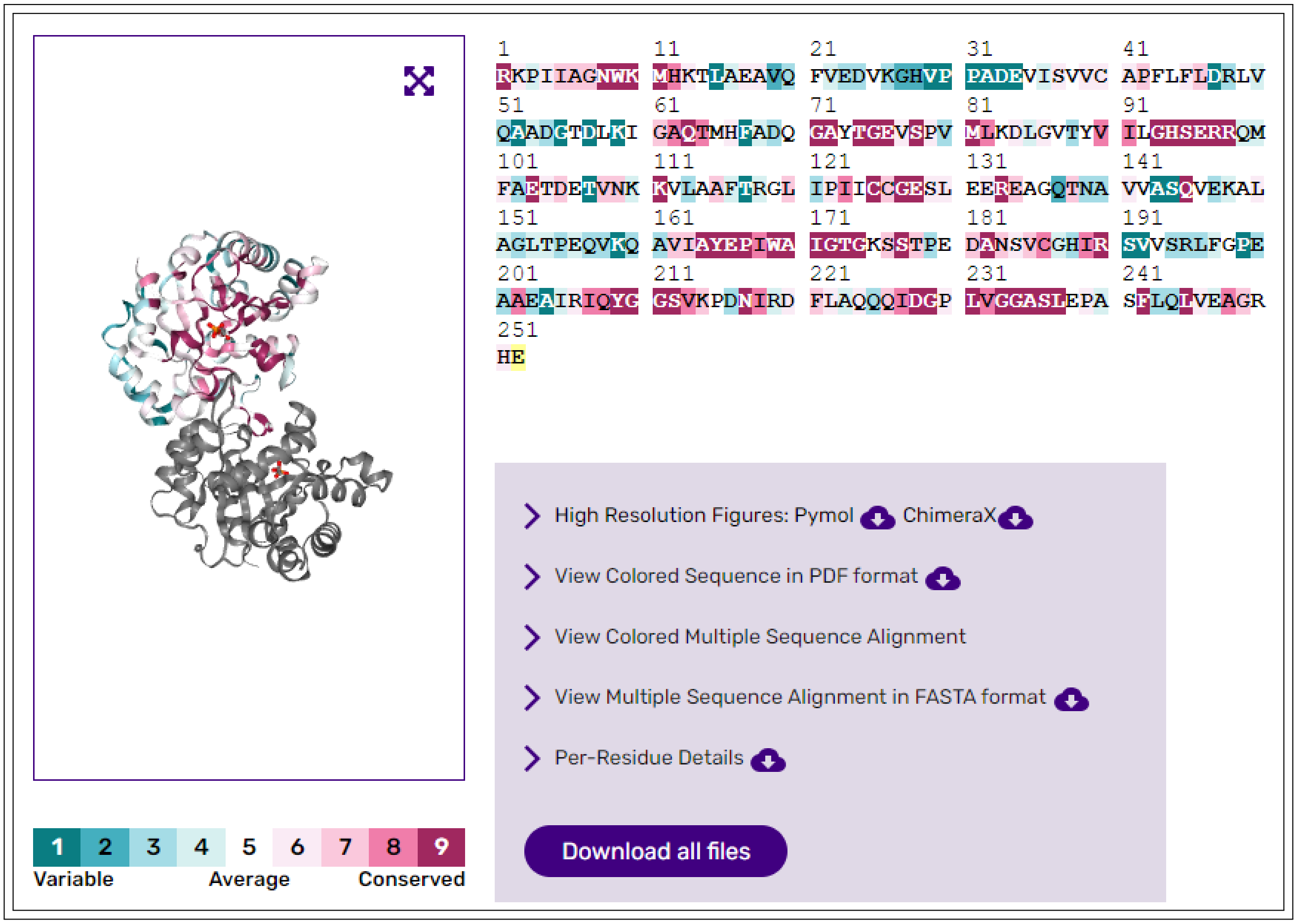
Maximal %ID Between Sequences : 95

Minimal %ID For Homologs : 35

150 sequences that sample the list of homologues to the query.

FIGURE 42. Running parameters of Consurf server

The results page for the above run on the Consurf server shows up as below. It consists of the structure of the protein sequence, along with its sequence and highlighted are highly and lowly conserved regions:

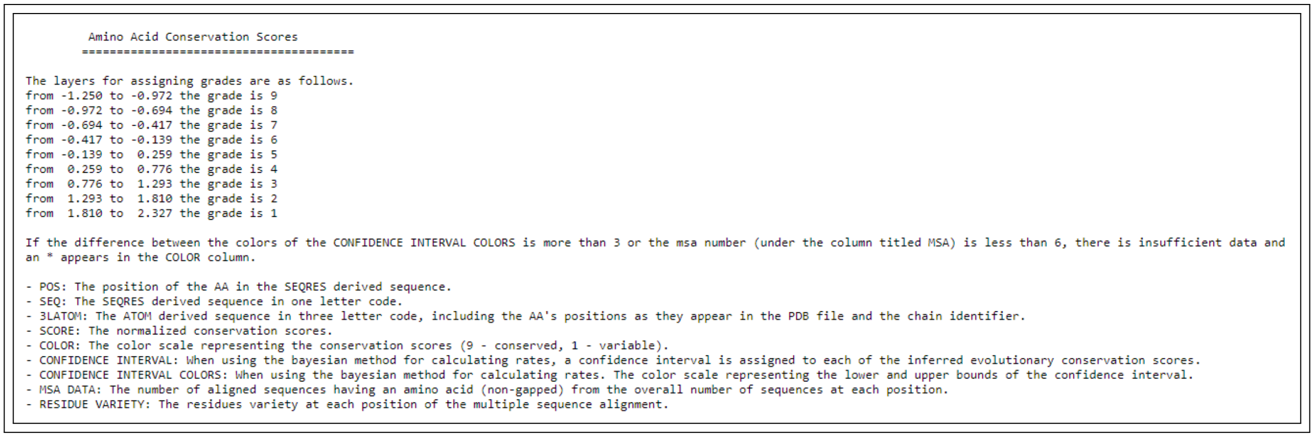


```
1          11         21         31         41
RKPIIAGNWK MHKTLAEAVQ FVEDVKGHVP PADEVISVVC APFLFLDRLV
51         61         71         81         91
QAADGTDLKI GAQTMHFADQ GAYTGEVSPV MLKDLGVTYV ILGHSERRQM
101        111        121        131        141
FAETDETVNK KVLAAFTRGL IPIICCGESL EEREAGQTNA VVASQVEKAL
151        161        171        181        191
AGLTPEQVKQ AVIAYEPIWA IGTGKSSTPE DANSVCGHIR SVVSRLFGPE
201        211        221        231        241
AAEAIRIQYG GSVKPDNIRD FLAQQQIDGP IVGGASLEPA SFLQLVEAGR
251
HE
```

High Resolution Figures: Pymol
ChimeraX
View Colored Sequence in PDF format
View Colored Multiple Sequence Alignment
View Multiple Sequence Alignment in FASTA format
Per-Residue Details
Download all files
1
2
3
4
5
6
7
8
9
Variable
Average
Conserved

Figure 43. Results page

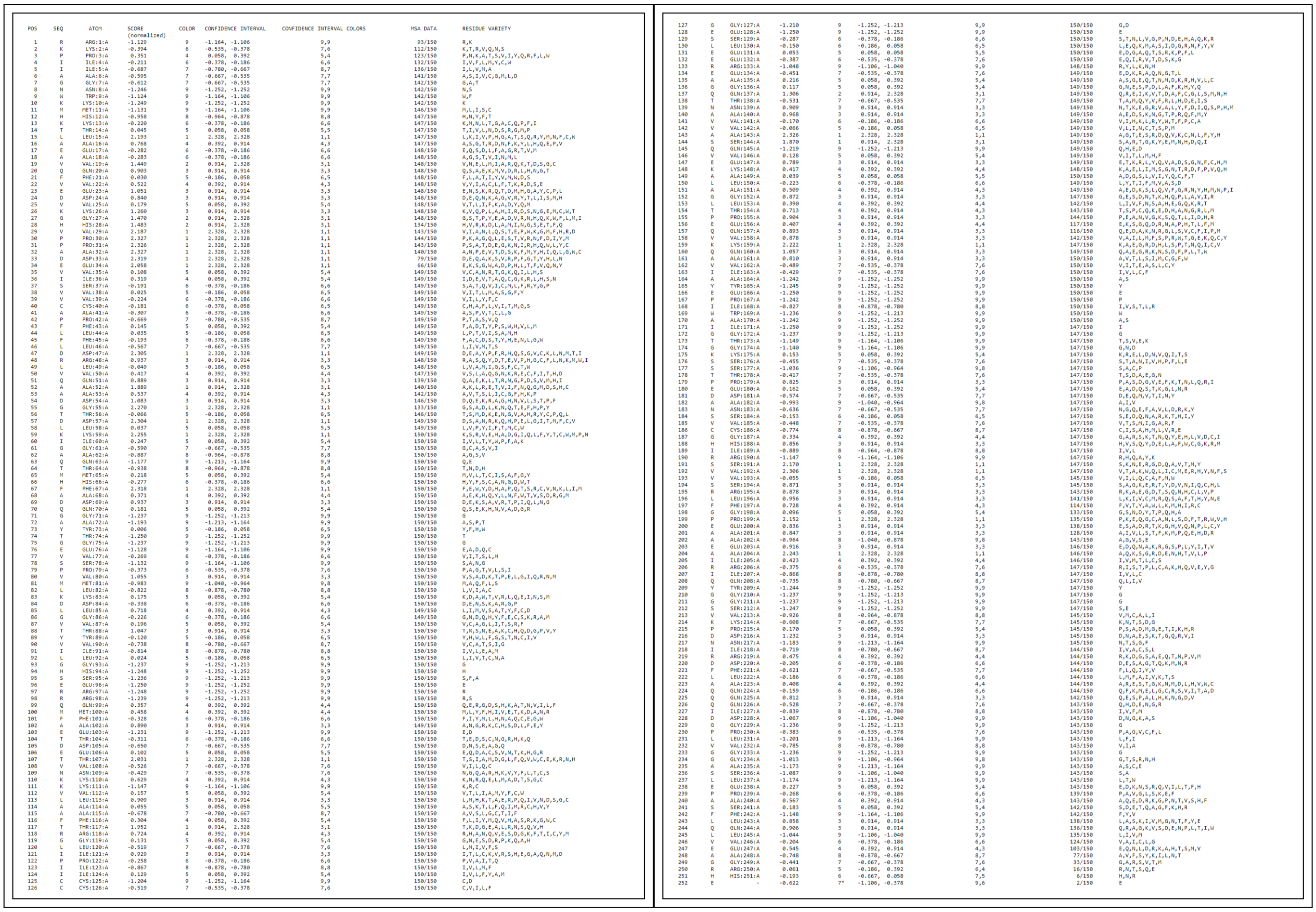
Given below are some of the parameters that play a role in assigning conservation scores, like, layers for assigning grades, confidence interval colors, residue variety.



```
         Amino Acid Conservation Scores
        =======================================

The layers for assigning grades are as follows.
from -1.250 to -0.972 the grade is 9
from -0.972 to -0.694 the grade is 8
from -0.694 to -0.417 the grade is 7
from -0.417 to -0.139 the grade is 6
from -0.139 to  0.259 the grade is 5
from  0.259 to  0.776 the grade is 4
from  0.776 to  1.293 the grade is 3
from  1.293 to  1.810 the grade is 2
from  1.810 to  2.327 the grade is 1

If the difference between the colors of the CONFIDENCE INTERVAL COLORS is more than 3 or the msa number (under the column titled MSA) is less than 6, there is insufficient data and
an * appears in the COLOR column.

- POS: The position of the AA in the SEQRES derived sequence.
- SEQ: The SEQRES derived sequence in one letter code.
- 3LATOM: The ATOM derived sequence in three letter code, including the AA's positions as they appear in the PDB file and the chain identifier.
- SCORE: The normalized conservation scores.
- COLOR: The color scale representing the conservation scores (9 - conserved, 1 - variable).
- CONFIDENCE INTERVAL: When using the bayesian method for calculating rates, a confidence interval is assigned to each of the inferred evolutionary conservation scores.
- CONFIDENCE INTERVAL COLORS: When using the bayesian method for calculating rates. The color scale representing the lower and upper bounds of the confidence interval.
- MSA DATA: The number of aligned sequences having an amino acid (non-gapped) from the overall number of sequences at each position.
- RESIDUE VARIETY: The residues variety at each position of the multiple sequence alignment.
```

Figure 44. Parameters to compute the conservation scores

Now, given below is the set of all 252 residues in the given protein and also all properties related to each residue position, like, SEQ (amino acid), COLOR (indicating conservation), CONFIDENCE INTERVAL (indicate the range of conservation values for COLOR), RESIDUE VARIETY (presence of multiple residues at same position), etc.



POS	SEQ	ATOM	SCORE (normalized)	COLOR	CONFIDENCE INTERVAL	CONFIDENCE INTERVAL COLORS	MSA DATA	RESIDUE VARIETY
1	R	ARG:1:A	-1.129	9	-1.164, -1.106	9,9	93/150	R,K
2	K	LYS:2:A	-0.394	6	-0.535, -0.378	7,6	112/150	K,T,R,V,Q,N,S
3	P	PRO:3:A	0.351	4	0.058, 0.392	5,4	123/150	P,N,K,A,T,S,V,I,Y,Q,R,F,L,W
4	I	ILE:4:A	-0.211	6	-0.378, -0.186	6,6	132/150	I,V,F,L,M,Y,C,W
5	I	ILE:5:A	-0.687	7	-0.780, -0.667	8,7	136/150	I,L,V,M,A
6	A	ALA:6:A	-0.595	7	-0.667, -0.535	7,7	141/150	A,S,I,V,C,G,M,L,D
7	G	GLY:7:A	-0.612	7	-0.667, -0.535	7,7	142/150	G,A,T
8	N	ASN:8:A	-1.246	9	-1.252, -1.252	9,9	142/150	N,S
9	W	TRP:9:A	-1.124	9	-1.164, -1.106	9,9	142/150	W,F
10	K	LYS:10:A	-1.249	9	-1.252, -1.252	9,9	142/150	K
11	M	MET:11:A	-1.131	9	-1.164, -1.106	9,9	146/150	M,L,I,S,C
12	H	HIS:12:A	-0.958	8	-0.964, -0.878	8,8	147/150	H,N,Y,F,T
13	K	LYS:13:A	-0.220	6	-0.378, -0.186	6,6	147/150	K,M,N,L,T,G,A,C,Q,P,F,I
14	T	THR:14:A	0.045	5	0.058, 0.058	5,5	147/150	T,I,V,L,N,D,S,R,G,M,P
15	L	LEU:15:A	2.193	1	2.328, 2.328	1,1	147/150	L,K,I,V,P,H,G,A,T,S,Q,R,Y,M,N,F,C,W
16	A	ALA:16:A	0.768	4	0.392, 0.914	4,3	147/150	A,S,G,T,R,D,N,F,K,Y,L,H,Q,E,P,V
17	E	GLU:17:A	-0.282	6	-0.378, -0.186	6,6	148/150	E,Q,S,D,L,F,A,G,R,T,V,M
18	A	ALA:18:A	-0.283	6	-0.378, -0.186	6,6	148/150	A,G,S,T,V,I,N,M,L
19	V	VAL:19:A	1.449	2	0.914, 2.328	3,1	148/150	V,N,E,L,M,I,A,R,Q,K,T,D,S,G,C
20	Q	GLN:20:A	0.903	3	0.914, 0.914	3,3	148/150	Q,S,A,E,K,M,V,D,R,L,H,N,G,T
21	F	PHE:21:A	0.030	5	-0.186, 0.058	6,5	148/150	F,L,A,T,I,Y,V,M,W,D,S
22	V	VAL:22:A	0.522	4	0.392, 0.914	4,3	148/150	V,Y,I,A,C,L,F,T,K,R,D,S,E
23	E	GLU:23:A	1.051	3	0.914, 0.914	3,3	148/150	E,N,S,K,R,Q,T,D,M,H,G,A,Y,C,P,L
24	D	ASP:24:A	0.840	3	0.914, 0.914	3,3	148/150	D,E,Q,N,K,A,G,V,R,Y,T,L,I,S,M,H
25	V	VAL:25:A	0.179	5	0.058, 0.392	5,4	148/150	V,T,L,I,F,K,A,D,Y,Q,M
26	K	LYS:26:A	1.260	3	0.914, 0.914	3,3	148/150	K,V,Q,P,L,A,H,I,R,D,S,N,G,E,M,C,W,T
27	G	GLY:27:A	1.470	2	0.914, 2.328	3,1	148/150	G,S,T,P,Y,E,A,D,V,R,N,H,Q,K,W,F,L,M,I
28	H	HIS:28:A	1.483	2	0.914, 2.328	3,1	134/150	H,V,R,K,D,L,A,M,I,N,G,S,E,T,F,Q
29	V	VAL:29:A	2.187	1	2.328, 2.328	1,1	143/150	V,I,A,N,L,Q,S,T,E,P,W,K,G,M,F,H,R,D
30	P	PRO:30:A	2.327	1	2.328, 2.328	1,1	144/150	P,K,A,G,Q,L,E,S,T,V,R,N,F,D,I,Y,M
31	P	PRO:31:A	2.326	1	2.328, 2.328	1,1	143/150	P,S,A,T,D,E,G,K,N,I,R,H,Q,W,L,Y,C
32	A	ALA:32:A	2.327	1	2.328, 2.328	1,1	140/150	A,N,P,E,V,T,D,S,K,F,M,Y,H,I,Q,L,G,W,C
33	D	ASP:33:A	2.319	1	2.328, 2.328	1,1	79/150	D,E,Q,A,K,S,V,R,P,F,G,T,Y,H,L,N
34	E	GLU:34:A	2.058	1	2.328, 2.328	1,1	66/150	E,K,S,G,W,A,D,P,H,L,T,F,V,Q,N,Y
35	V	VAL:35:A	0.108	5	0.058, 0.392	5,4	149/150	V,C,A,N,R,T,G,K,Q,I,L,H,S
36	I	ILE:36:A	0.319	4	0.058, 0.392	5,4	149/150	I,D,E,V,T,A,Q,C,G,K,R,L,H,S,N
37	S	SER:37:A	-0.191	6	-0.378, -0.186	6,6	149/150	S,A,T,Q,V,I,C,M,L,F,R,Y,G,P
38	V	VAL:38:A	0.025	5	-0.186, 0.058	6,5	149/150	V,I,T,L,M,A,S,G,F,Y
39	V	VAL:39:A	-0.224	6	-0.378, -0.186	6,6	149/150	V,I,L,Y,F,C
40	C	CYS:40:A	-0.181	6	-0.378, 0.058	6,5	149/150	C,H,A,F,L,V,I,T,M,G,S
41	A	ALA:41:A	-0.307	6	-0.378, -0.186	6,6	149/150	A,S,P,V,T,C,L,G
42	P	PRO:42:A	-0.669	7	-0.780, -0.535	8,7	149/150	P,T,A,S,V,Q
43	F	PHE:43:A	0.145	5	0.058, 0.392	5,4	149/150	F,A,D,T,Y,P,S,W,H,V,L,M
44	L	LEU:44:A	0.035	5	-0.186, 0.058	6,5	149/150	L,P,T,V,I,S,A,M,H
45	F	PHE:45:A	-0.193	6	-0.378, -0.186	6,6	149/150	F,A,C,D,S,T,Y,H,E,N,L,G,W
46	L	LEU:46:A	-0.567	7	-0.667, -0.535	7,7	149/150	L,I,V,M,T,S
47	D	ASP:47:A	2.305	1	2.328, 2.328	1,1	149/150	D,E,A,Y,P,F,R,H,Q,S,G,V,C,K,L,N,M,T,I
48	R	ARG:48:A	0.937	3	0.914, 0.914	3,3	148/150	R,A,S,Q,Y,D,T,E,V,P,H,G,C,F,L,N,K,M,W,I
49	L	LEU:49:A	-0.049	5	-0.186, 0.058	6,5	148/150	L,V,A,M,I,G,S,F,C,T,W
50	V	VAL:50:A	0.417	4	0.392, 0.392	4,4	147/150	V,S,L,A,Q,G,N,K,R,E,C,F,I,T,H,D
51	Q	GLN:51:A	0.889	3	0.914, 0.914	3,3	139/150	Q,A,E,K,L,T,R,N,G,P,D,S,V,M,H,I
52	A	ALA:52:A	1.889	1	0.914, 2.328	3,1	140/150	A,K,L,R,E,T,V,I,F,N,Q,G,M,D,S,H,C
53	A	ALA:53:A	0.537	4	0.392, 0.914	4,3	142/150	A,V,T,S,L,I,C,G,F,H,K,P
54	D	ASP:54:A	1.083	3	0.914, 0.914	3,3	146/150	D,Q,E,K,R,A,G,H,N,V,L,S,T,P,F
55	G	GLY:55:A	2.270	1	2.328, 2.328	1,1	133/150	G,S,A,D,L,K,N,Q,T,E,F,H,P,Y
56	T	THR:56:A	-0.066	5	-0.186, 0.058	6,5	146/150	T,S,M,D,K,E,N,G,V,A,H,R,Y,C,P,Q,L
57	D	ASP:57:A	2.304	1	2.328, 2.328	1,1	149/150	D,S,A,N,R,K,Q,H,P,E,L,G,I,T,M,F,C,V
58	L	LEU:58:A	0.037	5	0.058, 0.058	5,5	149/150	L,V,P,Y,I,F,T,M,C,W
59	K	LYS:59:A	2.255	1	2.328, 2.328	1,1	150/150	K,S,R,V,E,H,A,D,G,I,Q,L,F,Y,T,C,W,M,P,N
60	I	ILE:60:A	0.247	5	0.058, 0.392	5,4	150/150	I,V,L,T,Y,W,P,F,A,K
61	G	GLY:61:A	-0.590	7	-0.667, -0.535	7,7	150/150	G,C,A,S,V,I
62	A	ALA:62:A	-0.887	8	-0.964, -0.878	8,8	150/150	A,G,S,V
63	Q	GLN:63:A	-1.177	9	-1.213, -1.164	9,9	150/150	Q,E
64	T	THR:64:A	-0.938	8	-0.964, -0.878	8,8	150/150	T,N,D,H
65	M	MET:65:A	0.218	5	0.058, 0.392	5,4	150/150	M,V,L,T,C,I,S,A,F,G,Y
66	H	HIS:66:A	-0.277	6	-0.378, -0.186	6,6	150/150	H,Y,F,S,C,A,N,G,D,W,T
67	F	PHE:67:A	2.318	1	2.328, 2.328	1,1	150/150	F,E,W,Y,D,H,A,P,Q,T,S,R,C,V,N,K,L,I,M
68	A	ALA:68:A	0.371	4	0.392, 0.392	4,4	150/150	A,E,K,H,Q,Y,L,N,F,W,T,V,S,D,R,G,M
69	D	ASP:69:A	0.937	3	0.914, 0.914	3,3	150/150	D,E,K,S,A,V,R,T,P,I,Q,L,N,G
70	Q	GLN:70:A	0.181	5	0.058, 0.392	5,4	150/150	Q,S,E,K,H,N,V,A,D,G,R
71	G	GLY:71:A	-1.237	9	-1.252, -1.213	9,9	150/150	G
72	A	ALA:72:A	-1.193	9	-1.213, -1.164	9,9	150/150	A,S,P,T
73	Y	TYR:73:A	0.006	5	-0.186, 0.058	6,5	150/150	Y,F,H,W
74	T	THR:74:A	-1.250	9	-1.252, -1.252	9,9	150/150	T
75	G	GLY:75:A	-1.237	9	-1.252, -1.213	9,9	150/150	G
76	E	GLU:76:A	-1.128	9	-1.164, -1.106	9,9	150/150	E,A,D,Q,C
77	V	VAL:77:A	-0.269	6	-0.378, -0.186	6,6	150/150	V,I,T,S,L,H
78	S	SER:78:A	-1.132	9	-1.164, -1.106	9,9	150/150	S,A,N,G
79	P	PRO:79:A	-0.373	6	-0.535, -0.378	7,6	150/150	P,A,G,T,V,L,S,I
80	V	VAL:80:A	1.055	3	0.914, 0.914	3,3	150/150	V,S,A,D,K,T,P,E,L,G,I,Q,R,N,M
81	M	MET:81:A	-0.983	9	-1.040, -0.964	9,8	150/150	M,A,Q,F,L,S
82	L	LEU:82:A	-0.822	8	-0.878, -0.780	8,8	150/150	L,V,I,A,C
83	K	LYS:83:A	0.175	5	0.058, 0.392	5,4	150/150	K,D,A,W,T,V,R,L,Q,E,I,N,S,M
84	D	ASP:84:A	-0.338	6	-0.378, -0.186	6,6	150/150	D,E,N,S,K,A,R,G,P
85	L	LEU:85:A	0.718	4	0.392, 0.914	4,3	149/150	L,I,M,V,S,A,T,Y,F,C,D
86	G	GLY:86:A	-0.226	6	-0.378, -0.186	6,6	149/150	G,N,D,Q,H,Y,F,E,C,S,K,R,A,M
87	V	VAL:87:A	0.196	5	0.058, 0.392	5,4	150/150	V,C,A,G,L,I,T,S,R,F
88	T	THR:88:A	1.047	3	0.914, 0.914	3,3	150/150	T,R,S,N,E,A,K,C,H,Q,D,G,P,V,Y
89	Y	TYR:89:A	-0.120	5	-0.186, 0.058	6,5	150/150	Y,H,W,L,F,G,S,T,N,C,I,V
90	V	VAL:90:A	-0.738	8	-0.780, -0.667	8,7	150/150	V,C,A,T,S,I,G
91	I	ILE:91:A	-0.814	8	-0.878, -0.780	8,8	150/150	I,V,L,E,A,M
92	L	LEU:92:A	0.024	5	-0.186, 0.058	6,5	150/150	L,I,V,T,C,N,A
93	G	GLY:93:A	-1.237	9	-1.252, -1.213	9,9	150/150	G
94	H	HIS:94:A	-1.248	9	-1.252, -1.252	9,9	150/150	H
95	S	SER:95:A	-1.236	9	-1.252, -1.213	9,9	150/150	S,F,A
96	E	GLU:96:A	-1.250	9	-1.252, -1.252	9,9	150/150	E
97	R	ARG:97:A	-1.248	9	-1.252, -1.252	9,9	150/150	R
98	R	ARG:98:A	-1.239	9	-1.252, -1.213	9,9	150/150	R,S
99	Q	GLN:99:A	0.357	4	0.392, 0.392	4,4	150/150	Q,E,R,G,D,S,H,K,A,T,N,V,I,L,F
100	M	MET:100:A	0.458	4	0.392, 0.392	4,4	150/150	M,L,Y,F,H,I,V,E,T,K,D,A,N,R
101	F	PHE:101:A	-0.328	6	-0.378, -0.186	6,6	150/150	F,I,Y,M,L,H,N,A,Q,C,E,G,W
102	A	ALA:102:A	0.890	3	0.914, 0.914	3,3	149/150	A,N,G,R,K,C,H,S,D,L,F,E,Y
103	E	GLU:103:A	-1.231	9	-1.252, -1.213	9,9	150/150	E,D
104	T	THR:104:A	-0.311	6	-0.378, -0.186	6,6	150/150	T,E,D,S,C,N,G,R,H,K,Q
105	D	ASP:105:A	-0.650	7	-0.667, -0.535	7,7	150/150	D,N,S,E,A,G,Q
106	E	GLU:106:A	0.102	5	0.058, 0.058	5,5	150/150	E,Q,D,A,C,S,V,N,T,K,H,G,R
107	T	THR:107:A	2.031	1	2.328, 2.328	1,1	150/150	T,S,I,A,M,D,G,L,F,Q,V,W,C,E,K,R,N,H
108	V	VAL:108:A	-0.526	7	-0.667, -0.378	7,6	150/150	V,I,L,Q,C
109	N	ASN:109:A	-0.429	7	-0.535, -0.378	7,6	150/150	N,G,Q,A,R,H,K,V,Y,F,L,T,C,S
110	K	LYS:110:A	0.629	4	0.392, 0.914	4,3	150/150	K,N,R,Q,E,L,M,A,D,T,S,G,C
111	K	LYS:111:A	-1.147	9	-1.164, -1.106	9,9	150/150	K,R,C
112	V	VAL:112:A	0.157	5	0.058, 0.392	5,4	150/150	V,T,L,I,A,M,Y,F,C,W
113	L	LEU:113:A	0.909	3	0.914, 0.914	3,3	150/150	L,M,H,K,T,A,E,R,P,Q,I,V,N,D,S,G,C
114	A	ALA:114:A	0.055	5	0.058, 0.058	5,5	150/150	A,S,K,T,L,F,Q,I,M,R,C,H,V,Y
115	A	ALA:115:A	-0.678	7	-0.780, -0.667	8,7	150/150	A,V,S,L,G,C,T,I,F
116	F	PHE:116:A	0.304	4	0.058, 0.392	5,4	150/150	F,L,I,Y,M,Q,V,H,A,S,R,K,G,W,C
117	T	THR:117:A	1.952	1	0.914, 2.328	3,1	150/150	T,K,D,G,E,A,L,R,N,S,Q,V,H
118	R	ARG:118:A	0.724	4	0.392, 0.914	4,3	150/150	R,H,A,N,Q,V,E,S,D,G,K,F,T,I,C,Y,M
119	G	GLY:119:A	0.131	5	0.058, 0.392	5,4	150/150	G,N,E,S,D,R,P,K,Q,A,H
120	L	LEU:120:A	-0.519	7	-0.667, -0.378	7,6	150/150	L,M,I,V,F,S
121	I	ILE:121:A	0.929	3	0.914, 0.914	3,3	150/150	I,T,L,C,K,V,R,S,H,E,G,A,Q,N,M,D
122	P	PRO:122:A	-0.258	6	-0.378, -0.186	6,6	150/150	P,V,A,I,T,Q
123	I	ILE:123:A	-0.867	8	-0.878, -0.780	8,8	150/150	I,V,L,M,F
124	I	ILE:124:A	0.129	5	0.058, 0.392	5,4	150/150	I,V,L,F,Y,A,M
125	C	CYS:125:A	-1.204	9	-1.252, -1.164	9,9	150/150	C,D
126	C	CYS:126:A	-0.519	7	-0.535, -0.378	7,6	150/150	C,V,I,L,F
127	G	GLY:127:A	-1.210	9	-1.252, -1.213	9,9	150/150	G,D
128	E	GLU:128:A	-1.250	9	-1.252, -1.252	9,9	150/150	E
129	S	SER:129:A	-0.287	6	-0.378, -0.186	6,6	150/150	S,T,N,L,V,G,P,M,D,E,H,A,Q,K,R
130	L	LEU:130:A	-0.150	6	-0.186, 0.058	6,5	150/150	L,E,Q,K,M,A,S,I,D,G,R,N,F,Y,V
131	E	GLU:131:A	0.053	5	0.058, 0.058	5,5	150/150	E,D,G,A,Q,T,S,R,K,P,F,L
132	E	GLU:132:A	-0.387	6	-0.535, -0.378	7,6	150/150	E,Q,I,R,V,T,D,S,K,G
133	R	ARG:133:A	-1.048	9	-1.106, -1.040	9,9	148/150	R,Y,L,K,N,H
134	E	GLU:134:A	-0.451	7	-0.535, -0.378	7,6	149/150	E,D,K,R,A,Q,N,G,T,L
135	A	ALA:135:A	0.216	5	0.058, 0.392	5,4	149/150	A,S,G,E,Q,T,N,M,D,K,R,H,V,L,C
136	G	GLY:136:A	0.117	5	0.058, 0.392	5,4	149/150	G,N,E,S,P,D,L,A,F,K,H,Y,Q
137	Q	GLN:137:A	1.306	2	0.914, 2.328	3,1	149/150	Q,R,E,I,K,V,T,D,A,F,C,G,L,S,M,N,H
138	T	THR:138:A	-0.531	7	-0.667, -0.535	7,7	149/150	T,A,M,Q,Y,V,F,R,L,H,D,E,I,S
139	N	ASN:139:A	0.909	3	0.914, 0.914	3,3	149/150	N,T,K,E,G,R,V,A,L,Y,F,D,I,Q,S,P,H,M
140	A	ALA:140:A	0.968	3	0.914, 0.914	3,3	149/150	A,E,D,S,K,N,G,T,P,R,Q,F,M,Y
141	V	VAL:141:A	-0.170	6	-0.186, -0.186	6,6	149/150	V,I,H,K,L,R,Y,W,T,F,P,C,A
142	V	VAL:142:A	-0.066	5	-0.186, 0.058	6,5	149/150	V,L,I,N,C,T,S,P,M
143	A	ALA:143:A	2.326	1	2.328, 2.328	1,1	149/150	A,G,T,E,S,R,D,Q,V,K,C,N,L,F,Y,H
144	S	SER:144:A	1.870	1	0.914, 2.328	3,1	149/150	S,A,R,T,G,K,Y,E,M,N,H,D,Q,I
145	Q	GLN:145:A	-1.219	9	-1.252, -1.213	9,9	149/150	Q,H,E,D
146	V	VAL:146:A	0.128	5	0.058, 0.392	5,4	149/150	V,I,T,L,M,H,F
147	E	GLU:147:A	0.789	3	0.914, 0.914	3,3	149/150	E,T,K,R,L,Y,Q,V,A,D,S,G,N,F,C,H,M
148	K	LYS:148:A	0.417	4	0.392, 0.392	4,4	148/150	K,A,E,L,I,M,S,G,N,T,R,D,F,P,V,Q,H
149	A	ALA:149:A	0.039	5	0.058, 0.058	5,5	150/150	A,D,G,S,L,V,I,Y,Q,C,F,T
150	L	LEU:150:A	-0.223	6	-0.378, -0.186	6,6	149/150	L,Y,T,I,F,M,V,A,S,D
151	A	ALA:151:A	0.509	4	0.392, 0.914	4,3	149/150	A,E,D,K,S,L,Q,V,F,G,R,N,Y,H,M,W,P,I
152	G	GLY:152:A	0.872	3	0.914, 0.914	3,3	147/150	G,E,S,D,N,T,K,H,Q,P,L,A,V,I,R
153	L	LEU:153:A	0.390	4	0.392, 0.392	4,4	142/150	L,I,V,F,N,S,A,H,E,G,Q,K,R,T
154	T	THR:154:A	0.713	4	0.392, 0.914	4,3	143/150	T,S,P,C,Q,K,E,D,H,A,N,G,R,L,M
155	P	PRO:155:A	0.904	3	0.914, 0.914	3,3	144/150	P,E,A,N,V,G,K,S,Q,T,L,I,D,H,R
156	E	GLU:156:A	0.407	4	0.392, 0.392	4,4	117/150	E,K,S,G,Q,D,R,N,A,P,H,T,L,F,M
157	Q	GLN:157:A	0.893	3	0.914, 0.914	3,3	116/150	Q,E,D,A,K,N,R,G,L,S,V,C,F,I,P,M
158	V	VAL:158:A	0.878	3	0.914, 0.914	3,3	142/150	V,A,I,L,M,F,S,P,R,W,T,G,E,K,Q,C,Y
159	K	LYS:159:A	2.222	1	2.328, 2.328	1,1	147/150	K,A,E,G,R,D,H,L,S,P,T,N,Q,I,C,V
160	Q	GLN:160:A	1.057	3	0.914, 0.914	3,3	149/150	Q,A,E,G,R,K,N,S,D,F,P,L,T,W
161	A	ALA:161:A	0.810	3	0.914, 0.914	3,3	150/150	A,V,T,L,S,I,M,C,G,F,W
162	V	VAL:162:A	-0.489	7	-0.535, -0.378	7,6	150/150	V,I,T,E,A,S,L,C,Y
163	I	ILE:163:A	-0.429	7	-0.535, -0.378	7,6	150/150	I,V,L,C,F
164	A	ALA:164:A	-1.242	9	-1.252, -1.252	9,9	150/150	A,S
165	Y	TYR:165:A	-1.245	9	-1.252, -1.252	9,9	150/150	Y
166	E	GLU:166:A	-1.250	9	-1.252, -1.252	9,9	150/150	E
167	P	PRO:167:A	-1.242	9	-1.252, -1.252	9,9	150/150	P
168	I	ILE:168:A	-0.827	8	-0.878, -0.780	8,8	150/150	I,V,S,T,L,R
169	W	TRP:169:A	-1.236	9	-1.252, -1.213	9,9	150/150	W
170	A	ALA:170:A	-1.242	9	-1.252, -1.252	9,9	150/150	A,S
171	I	ILE:171:A	-1.250	9	-1.252, -1.252	9,9	147/150	I
172	G	GLY:172:A	-1.237	9	-1.252, -1.213	9,9	147/150	G
173	T	THR:173:A	-1.149	9	-1.164, -1.106	9,9	147/150	T,S,V,E,K
174	G	GLY:174:A	-1.140	9	-1.164, -1.106	9,9	147/150	G,N,D
175	K	LYS:175:A	0.153	5	0.058, 0.392	5,4	147/150	K,R,E,L,D,N,V,Q,I,T,S
176	S	SER:176:A	-0.455	7	-0.535, -0.378	7,6	147/150	S,T,A,N,I,V,H,P,F,L,E
177	S	SER:177:A	-1.036	9	-1.106, -0.964	9,8	147/150	S,A,C,P
178	T	THR:178:A	-0.417	7	-0.535, -0.378	7,6	147/150	T,S,D,A,E,G,N
179	P	PRO:179:A	0.825	3	0.914, 0.914	3,3	147/150	P,A,S,D,G,V,E,F,K,T,N,L,Q,R,I
180	E	GLU:180:A	0.162	5	0.058, 0.392	5,4	147/150	E,A,D,Q,S,T,K,G,L,N,R
181	D	ASP:181:A	-0.574	7	-0.667, -0.535	7,7	147/150	D,E,Q,M,V,T,I,N,Y
182	A	ALA:182:A	-0.993	9	-1.040, -0.964	9,8	147/150	A,I,V
183	N	ASN:183:A	-0.636	7	-0.667, -0.535	7,7	147/150	N,G,Q,E,F,A,V,L,D,R,K,Y
184	S	SER:184:A	-0.153	6	-0.186, 0.058	6,5	147/150	S,E,D,Q,N,A,R,K,T,H,I,Y
185	V	VAL:185:A	-0.448	7	-0.535, -0.378	7,6	147/150	V,T,S,M,I,G,A,R,F
186	C	CYS:186:A	-0.774	8	-0.878, -0.667	8,7	147/150	C,I,S,A,H,M,L,V,R,E
187	G	GLY:187:A	0.334	4	0.392, 0.392	4,4	147/150	G,A,R,S,K,T,N,Q,Y,E,H,L,V,D,C,I
188	H	HIS:188:A	0.856	3	0.914, 0.914	3,3	147/150	H,V,S,Q,Y,D,E,L,A,F,W,C,G,K,R,M
189	I	ILE:189:A	-0.889	8	-0.964, -0.878	8,8	147/150	I,V,L
190	R	ARG:190:A	-1.147	9	-1.164, -1.106	9,9	147/150	R,H,Q,A,Y,K
191	S	SER:191:A	2.170	1	2.328, 2.328	1,1	147/150	S,K,N,E,R,G,D,Q,A,V,T,M,Y
192	V	VAL:192:A	2.306	1	2.328, 2.328	1,1	147/150	V,T,A,K,W,Q,L,I,C,M,E,R,H,Y,N,F,S
193	V	VAL:193:A	-0.055	5	-0.186, 0.058	6,5	145/150	V,I,L,Q,C,A,F,M,W
194	S	SER:194:A	0.871	3	0.914, 0.914	3,3	145/150	S,A,G,K,E,R,T,Y,D,V,N,I,Q,C,H,L
195	R	ARG:195:A	0.878	3	0.914, 0.914	3,3	143/150	R,K,A,E,G,D,T,S,Q,N,H,C,L,V,P
196	L	LEU:196:A	0.956	3	0.914, 0.914	3,3	141/150	L,K,I,V,C,M,R,Q,S,A,F,T,H,Y,N,E
197	F	PHE:197:A	0.728	4	0.392, 0.914	4,3	114/150	F,V,T,Y,A,W,L,K,M,H,I,R,C
198	G	GLY:198:A	0.096	5	0.058, 0.392	5,4	133/150	G,S,N,D,Y,T,P,Q,H,A
199	P	PRO:199:A	2.152	1	2.328, 2.328	1,1	135/150	P,K,E,Q,G,C,A,N,L,S,D,F,T,R,W,V,H
200	E	GLU:200:A	0.836	3	0.914, 0.914	3,3	138/150	E,S,A,D,R,T,K,G,H,V,Q,N,P,L,C,Y
201	A	ALA:201:A	0.847	3	0.914, 0.914	3,3	128/150	A,I,V,L,S,T,F,K,M,P,Q,E,H,D,R
202	A	ALA:202:A	-0.964	8	-1.040, -0.878	9,8	143/150	A,G,V,S,E
203	E	GLU:203:A	0.916	3	0.914, 0.914	3,3	146/150	E,D,Q,N,A,K,R,G,S,P,L,Y,I,T,V
204	A	ALA:204:A	2.243	1	2.328, 2.328	1,1	146/150	A,Q,K,S,G,R,D,E,N,H,T,V,L,P
205	I	ILE:205:A	0.423	4	0.392, 0.392	4,4	146/150	I,V,M,T,L,C,S
206	R	ARG:206:A	-0.375	6	-0.535, -0.378	7,6	147/150	R,I,S,T,P,L,C,A,K,H,Q,V,E,Y,G
207	I	ILE:207:A	-0.868	8	-0.878, -0.780	8,8	147/150	I,V,L,C
208	Q	GLN:208:A	-0.735	8	-0.780, -0.667	8,7	147/150	Q,L,I,V
209	Y	TYR:209:A	-1.244	9	-1.252, -1.252	9,9	147/150	Y
210	G	GLY:210:A	-1.237	9	-1.252, -1.213	9,9	147/150	G
211	G	GLY:211:A	-1.237	9	-1.252, -1.213	9,9	147/150	G
212	S	SER:212:A	-1.247	9	-1.252, -1.252	9,9	147/150	S,E
213	V	VAL:213:A	-0.926	8	-0.964, -0.878	8,8	145/150	V,M,C,A,L,I
214	K	LYS:214:A	-0.608	7	-0.667, -0.535	7,7	145/150	K,N,T,S,D,G
215	P	PRO:215:A	0.170	5	0.058, 0.392	5,4	145/150	P,S,A,D,M,G,E,T,I,K,H,R
216	D	ASP:216:A	1.232	3	0.914, 0.914	3,3	145/150	D,N,A,E,S,K,T,G,Q,R,V,I
217	N	ASN:217:A	-1.183	9	-1.213, -1.164	9,9	145/150	N,T,S,G,F
218	I	ILE:218:A	-0.719	8	-0.780, -0.667	8,7	144/150	I,V,A,C,S,L
219	R	ARG:219:A	0.475	4	0.392, 0.392	4,4	144/150	R,K,D,G,S,A,E,Q,T,N,P,V,M
220	D	ASP:220:A	-0.205	6	-0.378, -0.186	6,6	144/150	D,E,S,A,G,T,Q,K,M,N,R
221	F	PHE:221:A	-0.621	7	-0.667, -0.535	7,7	144/150	F,L,Q,I,Y,V
222	L	LEU:222:A	-0.186	6	-0.378, -0.186	6,6	144/150	L,M,F,A,I,V,K,T,S
223	A	ALA:223:A	0.408	4	0.392, 0.392	4,4	144/150	A,R,E,S,T,G,K,N,M,D,L,H,V,W,C
224	Q	GLN:224:A	-0.159	6	-0.186, -0.186	6,6	144/150	Q,F,K,M,E,L,G,C,R,S,V,I,T,A,D
225	Q	GLN:225:A	0.812	3	0.914, 0.914	3,3	142/150	Q,E,S,P,A,L,H,K,N,G,D,V
226	Q	GLN:226:A	-0.528	7	-0.667, -0.378	7,6	143/150	Q,H,D,E,N,G,R
227	I	ILE:227:A	-0.839	8	-0.878, -0.780	8,8	143/150	I,V,F,M
228	D	ASP:228:A	-1.067	9	-1.106, -1.040	9,9	143/150	D,N,G,K,A,S
229	G	GLY:229:A	-1.236	9	-1.252, -1.213	9,9	143/150	G
230	P	PRO:230:A	-0.383	6	-0.535, -0.378	7,6	143/150	P,A,G,V,C,F,L
231	L	LEU:231:A	-1.201	9	-1.213, -1.164	9,9	143/150	L,F,I
232	V	VAL:232:A	-0.785	8	-0.878, -0.780	8,8	143/150	V,I,A
233	G	GLY:233:A	-1.236	9	-1.252, -1.213	9,9	143/150	G
234	G	GLY:234:A	-1.013	9	-1.106, -0.964	9,8	143/150	G,T,S,R,N,H
235	A	ALA:235:A	-1.173	9	-1.213, -1.164	9,9	143/150	A,S,C,E
236	S	SER:236:A	-1.087	9	-1.106, -1.040	9,9	143/150	S,A
237	L	LEU:237:A	-1.174	9	-1.213, -1.164	9,9	143/150	L,T,W
238	E	GLU:238:A	0.227	5	0.058, 0.392	5,4	143/150	E,D,K,N,S,R,Q,V,I,L,T,F,H
239	P	PRO:239:A	-0.268	6	-0.378, -0.186	6,6	139/150	P,A,V,G,L,S,K,E,F
240	A	ALA:240:A	0.567	4	0.392, 0.914	4,3	143/150	A,Q,E,D,R,K,G,P,N,T,V,S,H,F
241	S	SER:241:A	0.183	5	0.058, 0.392	5,4	142/150	S,D,E,T,Q,A,G,F,K,H,R
242	F	PHE:242:A	-1.148	9	-1.164, -1.106	9,9	142/150	F,Y,V
243	L	LEU:243:A	0.858	3	0.914, 0.914	3,3	138/150	L,A,S,K,I,V,M,G,N,T,F,Y,E
244	Q	GLN:244:A	0.906	3	0.914, 0.914	3,3	136/150	Q,R,A,G,K,V,S,D,E,N,P,L,T,I,W
245	L	LEU:245:A	-1.044	9	-1.106, -1.040	9,9	135/150	L,I,V,M
246	V	VAL:246:A	-0.204	6	-0.378, -0.186	6,6	124/150	V,A,I,C,L,G
247	E	GLU:247:A	0.545	4	0.392, 0.914	4,3	103/150	E,Q,N,L,D,R,K,A,H,T,S,M,V
248	A	ALA:248:A	-0.748	8	-0.878, -0.667	8,7	77/150	A,V,F,S,Y,K,I,L,N,T
249	G	GLY:249:A	-0.441	7	-0.667, -0.378	7,6	33/150	G,A,R,S,V,T,M
250	R	ARG:250:A	0.061	5	-0.186, 0.392	6,4	16/150	R,N,T,S,Q,E
251	H	HIS:251:A	-0.193	6	-0.667, 0.058	7,5	6/150	H,N,R
252	E	-	-0.622	7*	-1.106, -0.378	9,6	2/150	E

(A) 1-126 positions

(B) 126-252 positions

Figure 45. Each residue-wise properties for the given protein

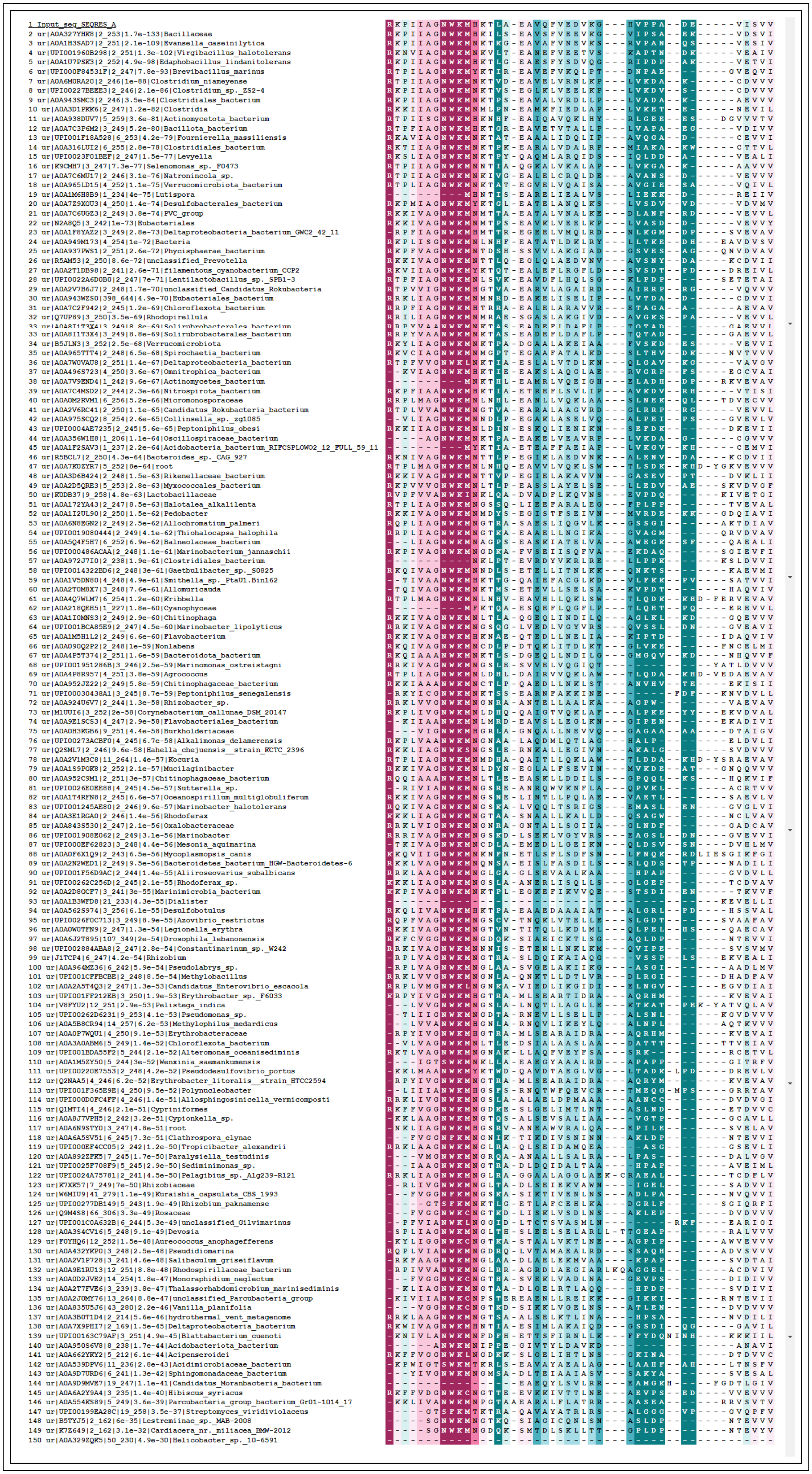
Names and Taxonomy related to protein from UniProtKB:

1. Recommended name is **Triosephosphate isomerase**
2. EC number is **EC:5.3.1.1**
3. Its optimal temperature is 60 degrees Celsius. It is thermostable
4. It plays a role in the following pathways: Carbohydrate biosynthesis; gluconeogenesis; Carbohydrate degradation; glycolysis; D-glyceraldehyde 3-phosphate from glycerone phosphate

The color scale in all of the above images are indicative of the conservation scores, with 9 indicating **conserved**, while 1 indicating **variable**. These numeric values have been assigned to different confidence interval ranges for the conservation scores. The split up is like this:

- from -1.250 to -0.972 the grade is 9
- from -0.972 to -0.694 the grade is 8
- from -0.694 to -0.417 the grade is 7
- from -0.417 to -0.139 the grade is 6
- from -0.139 to 0.259 the grade is 5
- from 0.259 to 0.776 the grade is 4
- from 0.776 to 1.293 the grade is 3
- from 1.293 to 1.810 the grade is 2
- from 1.810 to 2.327 the grade is 1

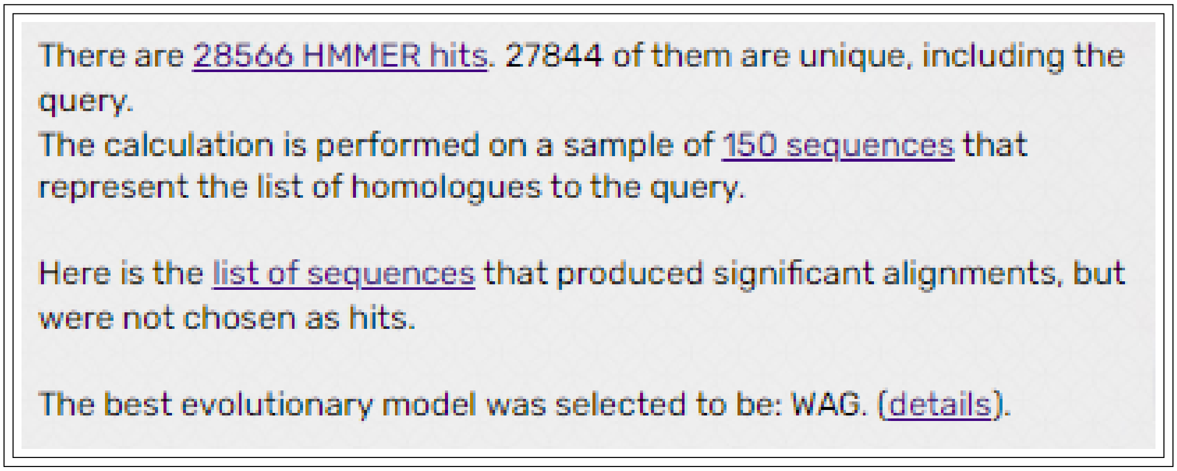
The Consurf can also perform multiple sequence alignment with some predefined sequences based on their closeness to the given sequence. Below image is an example of one such MSA.



```
1 Input_seq_SEQRES_A                                                  RKPIIAGNWKMHKTLA-EAVQFVEDVKG---HVPPA--DE-----VISVV
2 ur|A0A327YHK8|2_253|1.7e-133|Bacillaceae                            RKPIIAGNWKMHKILS-EAVSFVEEVKG---VIPSA--EK-----VDSVI
3 ur|A0A1H3SAD7|2_251|2.1e-109|Evansella_caseinilytica                RKPIIAGNWKMNKTKG-EAVAFVNEVKS---RVPAN--QS-----VESIV
4 ur|UPI001960B298|2_251|1.3e-102|Virgibacillus_halotolerans          RKNVIAGNWKMNKVLT-EANEFVEETVS---KAPTN--DK-----VEAIV
5 ur|A0A1U7PSK3|2_252|4.9e-98|Edaphobacillus_lindanitolerans         RKPIIAGNWKMHKTLG-EAESFYSDVQG---RIPDP--AK-----VETVI
6 ur|UPI000F84531F|2_247|7.8e-93|Brevibacillus_marinus                RTPILAGNWKMYKTIR-EAVEFVKQLPT---DNPAE--------GVEQVI
7 ur|A0A6M0RA20|2_246|1e-88|Clostridium_niameyense                   RTPIIAGNWKMNKTVD-EAVELVRNLKP---LVKDA--K------CDVVV
8 ur|UPI00227BEEE3|2_246|2.1e-86|Clostridium_sp._ZS2-4                RKPIIAGNWKMNKTVS-EGLKLVEELKP---LVKDV--S------CDVVV
9 ur|A0A943SMC3|2_246|3.5e-84|Clostridiales_bacterium                 RKPIIAGNWKMNKTPS-EAVALVRDLLP---LVADA--K------AEVVV
10 ur|A0A3D1PKK6|2_247|1.2e-82|Clostridia                              RKKIIAGNWKMNMLPN-EAMKFIEDLAP---LVKET--E------NEVIL
11 ur|A0A938DUV7|5_259|3.6e-81|Actinomycetota_bacterium                RTPIISGNWKMHKNHF-EAIQAVQKLHY---RLGEE--ES---DGVVVTV
12 ur|A0A7C3P6M2|3_249|5.2e-80|Bacillota_bacterium                     RTPFIAGNWKMHKTGR-EAVETVTALLP---LVAPA--EG-----VDIVV
13 ur|UPI001F18A528|6_253|4.2e-79|Fournierella_massiliensis            RKAVIAGNWKMNKTAT-EAAALIDQLIP---AVQGA--D------CEVVI
14 ur|A0A316LUI2|6_255|2.8e-78|Clostridiales_bacterium                 RKTIIAGNWKMNKTPK-EARALVDALRP---MIAKA--KW-----CTTVL
15 ur|UPI0023F01BEF|2_247|1.5e-77|Levyella                             RKSLIAGNWKMNKTPY-QAQMLARQIDS---IQLDD--A------VEALI
16 ur|K9CMH7|3_247|7.3e-77|Selenomonas_sp._F0473                       RTPIIAGNWKMNNTIA-QGKALVKALAP---LVKGA--K------AAVVV
17 ur|A0A7C6MU17|2_246|3.1e-76|Natronincola_sp.                        RTPIIAGNWKMNKIVG-EALELCRQLDE---AVANS--D------VEVVV
18 ur|A0A965LD15|4_252|1.1e-75|Verrucomicrobiota_bacterium             RTPLIAGNWKMNKTAT-EGVELVQAISA---AVGIE--SA-----VQIVL
19 ur|A0A1M6H8B9|1_234|4e-75|Lutispora                                 ----------MHNTIS-EARELIEALVS---LIEKK--D------REIIV
20 ur|A0A7Z9XGU3|4_250|1.4e-74|Desulfobacterales_bacterium             RKPMIAGNWKMYKTGL-EATELANQLKS---LVSDV--VD-----VDVMV
21 ur|A0A7C6UGZ3|2_249|3.8e-74|PVC_group                               RKKIVAGNWKMNMTTA-EATALVNALKE---DLANF--RD-----VEVVL
22 ur|N2A8Q5|3_242|1e-73|Eubacteriales                                 RKKIVAGNWKMNMTPS-EAVKLVEELKP---LVASD--D------VEVVY
23 ur|A0A1F8YAZ2|3_249|2.8e-73|Deltaproteobacteria_bacterium_GWC2_42_11 -RPFIAGNWKMHMTTR-EGEELVMQLRD---NLKGM--DP----SVEVAV
24 ur|A0A949M173|4_254|1e-72|Bacteria                                  RKPLISGNWKMNLNHF-EATATLDKLRY---LLTKE--DH---EAVDVSV
25 ur|A0A937PWS1|2_251|2.6e-72|Phycisphaerae_bacterium                 RKPFVAGNWKMNTDSH-SSVVLAKGIAD---GSVES--AG---QNVDVAV
26 ur|R5AM53|2_250|8.6e-72|unclassified_Prevotella                     RKKIVAGNWKMNTTLQ-EGLQLAEDVNV---AVSNY--DA----KCDVII
27 ur|A0A2T1DB98|2_241|2.6e-71|filamentous_cyanobacterium_CCP2         RKVIIAGNWKMYKTQT-EALEFLRGFLD---SVSDT--PD----DREIVL
28 ur|UPI0022A6D0B0|2_247|7e-71|Lentilactobacillus_sp._SPB1-3          RTPFIAGNWKMNLSVK-EAVDFLHQLSG---KLPDP-------SETETAI
29 ur|A0A2V7B6J7|2_248|1.7e-70|unclassified_Candidatus_Rokubacteria    RTPVVIGNWKMHGTVA-EARVLAGAIRD---AIRRP--RG-----VQVVV
30 ur|A0A943WZS0|398_644|4.9e-70|Eubacteriales_bacterium               RRKLIAGNWKMNMNRD-EAKELISELIP---LVTDA--D------CDVVI
31 ur|A0A7C2F942|2_245|1.2e-69|Chloroflexota_bacterium                 RRPFIAGNWKMHTRRA-EALELARAVVR---ETAGA--A------AEVAV
32 ur|Q7UP89|3_250|3.5e-69|Rhodopirellula                              RRILIAGNWKMNMRAE-SGASLAKGIVD---AVGKS--PA-----VEVVL
33 ur|A0A8I1T3X4|3_249|8.8e-69|Solirubrobacterales_bacterium           RRPYVAANWKMNKTAS-EADEFLDAFLP---TQTAD--------GAEVVL
34 ur|B5JLN3|3_252|2.5e-68|Verrucomicrobiota                           RKYLIAGNWKMNKTPA-DGAALAKEIAA---FVSKD--ES-----VEVVI
35 ur|A0A965TTT4|2_248|6.5e-68|Spirochaetia_bacterium                  RKVCIAGNWKMNMGPT-EGAAFATALKD---SLTHV--DK----NVTVVV
36 ur|A0A7W0VAU8|2_251|1.4e-67|Deltaproteobacteria_bacterium           RTPFVVGNWKLNKTIA-ESLALVTDLKN---QLGAV--KG-----VAVGV
37 ur|A0A496S723|4_250|3.6e-67|Omnitrophica_bacterium                  --KVIAGNWKMHKTIE-EAKSLAQGLAE---RVGRP--FS----EGCVAI
38 ur|A0A7V9END4|1_242|9.6e-67|Actinomycetes_bacterium                 ----------MNKTHL-EAMRLVQEIGH---ELADH--DP---RKVEVAV
39 ur|A0A7C4MSD2|2_244|2.3e-66|Nitrospirota_bacterium                  RKPFIAANWKMHKTIA-ETREFLSVLIP---AVKDV--RH-----VTISI
40 ur|A0A0M2RVM1|6_256|5.2e-66|Micromonosporaceae                      RRPLMAGNWKMNLNHL-EANLLVQKLAA---SLNEK--QL---TDVECVV
41 ur|A0A2V6RC41|2_250|1.1e-65|Candidatus_Rokubacteria_bacterium       RTPLVVANWKMNGTVA-EARALAAGVRD---GLRRP--RG-----VEVVL
42 ur|A0A975SCQ2|8_254|2.6e-65|Collinsella_sp._zg1085                  --VLIAGNWKMNNDLP-EGAKLASELVQ---ALPEI--PS----GVEVLV
43 ur|UPI0004AE7235|2_245|5.6e-65|Peptoniphilus_obesi                  RKKIIAGNWKMNLDIN-ESKQLIENIKN---SEFDK--------DKEVII
44 ur|A0A356W1H8|1_206|1.1e-64|Oscillospiraceae_bacterium              -----AGNWKMNKTPA-EAKELVEAVRP---LVKDA--G------CGVVV
45 ur|A0A1F2SAV3|1_237|2.2e-64|Acidobacteria_bacterium_RIFCSPLOWO2_12_FULL_59_11 ----------MYKTIA-ETEAFFAEIAP---LVKGV--KH-----CEMVI
46 ur|R5BCL7|2_250|4.3e-64|Bacteroides_sp._CAG_927                     RKNIVAGNWKMNTTLP-EGIKLAEDVNK---ALENV--DA----KCDVII
47 ur|A0A7K0ZYR7|5_252|8e-64|root                                      RTPLMAGNWKMNLNHQ-EAVVLVQKLSW---TLSDK--KHD-YGKVEVVV
48 ur|A0A3D6B424|2_248|1.5e-63|Rikenellaceae_bacterium                 RKKIVAGNWKMNTTVP-EGIELAKAVVN---GASEV--PT----DVKLII
49 ur|A0A2D5QRE3|5_253|2.8e-63|Myxococcales_bacterium                  RKPVVVGNWKMNLTLP-EASSLAYELSE---LLEDV--AE---SKVDVGF
50 ur|K0DB37|9_258|4.8e-63|Lactobacillaceae                            RVPFVVANWKINKLQA-DVADFLKQVNS---EVPDQ-------KIVETGI
51 ur|A0A172YA43|2_247|8.5e-63|Halotalea_alkalilenta                   RTPLIAGNWKMNGSSQ-LIEEFARALEG---FPLPP--------TVEVAL
52 ur|A0A1I2UL90|2_250|1.5e-62|Pedobacter                              RKKIVAGNWKMNMDYS-EGISTFSEIVN---MVRDE--KK---GDQIAVI
53 ur|A0A6N8EGN2|2_249|2.5e-62|Allochromatium_palmeri                  RQPLIAGNWKMNGTRA-SAESLIQGVLK---GSSGI-------AKTDIAV
54 ur|UPI0019080444|2_249|4.1e-62|Thiohalocapsa_halophila              RRPLIAGNWKMNGTKA-EAAELLNGIKA---GVAGM-------QRVDVAV
55 ur|A0A5Q4F5H7|6_252|6.9e-62|Balneolaceae_bacterium                  ---LIAGNWKMNAGPS-EASKIATELVA---AWEGK--SF----QAEALI
56 ur|UPI000486ACAA|2_248|1.1e-61|Marinobacterium_jannaschii           RKPIVAGNWKMNGDKA-FVSEIISQFVA---EKDAQ-------SGIEVFI
57 ur|A0A972J7I0|2_238|1.9e-61|Clostridiales_bacterium                 -----------NKLPT-EVRDYVKRLRE---LLPPK-------KSLDVVI
58 ur|UPI0014322BD6|2_248|3e-61|Gaetbulibacter_sp._S0825               RKQIVAGNWKMNNDLS-ETELLITNLKK---QNKTS--------KAEVMI
59 ur|A0A1V5DN80|4_248|4.9e-61|Smithella_sp._PtaU1.Bin162              --TIVAANWKMHKTTG-QAIEFACGLKD---VLFKK--PV----SATVVI
60 ur|A0A2T0M8X7|3_248|7.6e-61|Allomuricauda                           -TQIVAGNWKMNKTMS-ETEVLLSELSA---KVPDT--------HAQVIV
61 ur|A0A4Q7WLM7|6_254|1.2e-60|Kribbella                               RTPLMAGNWKMNLNHV-EAVHLLQKLSW---TLQDK--KHD-FERVEVAV
62 ur|A0A218QEH5|1_227|1.8e-60|Cyanophyceae                            ----------MFKTQA-ESQEFLQGFLP---TLQET--PQ----EREVVL
63 ur|A0A1I0MNS3|2_249|2.9e-60|Chitinophaga                            RKKIVAGNWKMNLTLA-QGEQLINDILQ---AGLKL--KD----GQEVVV
64 ur|UPI001BCA85E9|2_247|4.5e-60|Marinobacter_lipolyticus             RRKIVAGNWKMNGSQG-LVEDLVGYVRS---QVSSL--DN----GVEAVI
65 ur|A0A1M5H1L2|2_249|6.6e-60|Flavobacterium                          RKKIVAGNWKMHKNAE-QTEDLLNELIA---KIPTD-------IDAQVIV
66 ur|A0A090Q2P2|2_248|1e-59|Nonlabens                                 RKQIVAGNWKMNCDLP-DTQKLITDLKT---GLVKE-------FNCELMI
67 ur|A0A4P5T374|2_251|1.6e-59|Bacteroidota_bacterium                  RKQIAAANWKMNKTLS-DGEQLLNDILG---GMGQV--KD----NHQVVF
68 ur|UPI001951286B|3_246|2.5e-59|Marinomonas_ostreistagni             RRKIVAGNWKMNGSLE-SVVELVQGIQT--------------YATLDVVV
69 ur|A0A4P8R957|4_251|3.8e-59|Agrococcus                              RTPLIAGNWKMNLDHL-DAIRVVQKLAW---TLQDA--KHD-VEDAEVAV
70 ur|A0A952JZ22|2_249|5.8e-59|Chitinophagaceae_bacterium              RKKIAAANWKMNCTLP-QAEDLLNKLST---ANVHV--TE----EKISII
71 ur|UPI00030438A1|3_245|8.7e-59|Peptoniphilus_senegalensis          -RKYICGNWKMNKTSS-EARNFAKKINE---------FDF---KNVDVLL
72 ur|A0A924U6V7|2_244|1.3e-58|Rhizobacter_sp.                         RRKLVVGNWKMNGNRA-ANTELLAALKA---AGPW---------VAEVAV
73 ur|M1UUI6|3_252|2e-58|Corynebacterium_callunae_DSM_20147           RKPLIVGNWKMNLDHQ-QAIGTVQKLAF---ALPKE--YY---EKVDVAL
74 ur|A0A9E1SCS3|4_247|2.9e-58|Flavobacteriales_bacterium              --KIIAANWKMNLMRD-EAVSLLEGLKN---GIPEN-------EKADIVI
75 ur|A0A0H3KGB6|9_251|4.4e-58|Burkholderiaceae                        ----VIGNWKMHGRLA-GNQALLNEVVQ---GAGAA--AA----DTAIGV
76 ur|UPI00273ACBF0|4_245|6.7e-58|Alkalimonas_delamerensis             RKPLVAANWKMNGNAA-LAQDMLQTLAG---HALP---------ELDVLI
77 ur|Q2SML7|2_246|9.6e-58|Hahella_chejuensis__strain_KCTC_2396       RKKLIAGNWKSNGSLE-RNKALLEGIVN---AKALG--------SVDVVV
78 ur|A0A2V1M3C8|11_264|1.4e-57|Kocuria                                RTPLIAGNWKMNMDHA-QAITLLQKLAW---TLDDA--KHD-YSRAEVAV
79 ur|A0A1S9PGK8|2_252|2.1e-57|Mucilaginibacter                        RKKIVAGNWKMNLDYN-EGLALFSEVIN---MVKDE--AT---GNQQVVV
80 ur|A0A952C9M1|2_251|3e-57|Chitinophagaceae_bacterium                RQQIAAANWKMNLTLE-EASKLLDDILS---GPQQL--KS----HQKVIF
81 ur|UPI0026E0EE88|4_245|4.5e-57|Sutterella_sp.                       --RIVIANWKMNGSRE-ANRQWVKNFLA---QPVKL--------ACRTVV
82 ur|A0A1T4RFN8|2_245|6.6e-57|Oceanospirillum_multiglobuliferum      RKKIVAGNWKMNGSNE-LINTLLPQLAE---VAETL--------SAEVLV
83 ur|UPI001245AE80|2_246|9.6e-57|Marinobacter_halotolerans            RQKIVAGNWKMNGSKA-LVQQLTSRIGS---EMASL--EN----GVGLVI
84 ur|A0A3E1RGA0|2_246|1.4e-56|Rhodoferax                              KKKLIAGNWKMNGTAA-SNAALLKALLD---QSAGW--------NCLVAV
85 ur|A0A843S530|2_247|2.1e-56|Oxalobacteraceae                        RRKLVIGNWKMNGNRA-GNTALLSGIIA---GLNDF--------GADCAV
86 ur|UPI001908E062|2_249|3.1e-56|Marinobacter                         RRRIVAGNWKMNGSKD-LSEKLVGYVRS---EAGSL--DN----GVEVVI
87 ur|UPI000EF62823|3_248|4.4e-56|Mesonia_aquimarina                   -TKIVAGNWKMNCDLA-EMEDLLGEIKN---QLDSV--SN----DVHLMV
88 ur|A0A0F6X1Q9|2_243|6.5e-56|Mycoplasmopsis_canis                    KKQVIIGNWKMNKNFN-ETLSFLSDFSN---LFNQK--RDLIESGIKFGI
89 ur|A0A2N2WED1|2_249|9.5e-56|Bacteroidetes_bacterium_HGW-Bacteroidetes-6 RKKLVAGNWKMNQNSA-EISLFASDILS---RLQDS--TP----NADILI
90 ur|UPI001F56D9AC|2_244|1.4e-55|Aliiroseovarius_subalbicans          RRKLAAGNWKMNGLGA-GLSEVAALKAA----HPAP--------GVDVLL
91 ur|UPI00262C256D|2_245|2.1e-55|Rhodoferax_sp.                       KKKLIAGNWKMNGSLA-ANERLISQLLS---GLGEP--------TCDVAV
92 ur|A0A2D8GCF7|3_241|3e-55|Marinimicrobia_bacterium                  KKPIVAGNWKMNKTPL-EGKEFIKVVQE---STSDI--EN-----TKVVF
93 ur|A0A1B3WFD8|21_233|4.3e-55|Dialister                              -------------------------------------------KEVELLI
94 ur|A0A562S974|3_256|6.1e-55|Desulfobotulus                          RKQLIVANWKMHKTPA-EAAEDAAAIAT---ALGRL--PD----HSSVAL
95 ur|UPI0026F0C713|3_249|8.9e-55|Azovibrio_restrictus                 RQPFVAGNWKMNGSCV-TNQKLVTELLE---SLGDT-------SVAAFAV
96 ur|A0A0W0TFN9|2_247|1.3e-54|Legionella_erythra                      RKKIVAGNWKMNGTVN-TITQLLKDLML---QLPEL--HS----QAECAV
97 ur|A0A6J2T895|107_349|2e-54|Drosophila_lebanonensis                 RKFCVGGNWKMNGDQK-SIAEICKTLSG---AQLDP--------NTEVVI
98 ur|UPI002884ABA8|2_247|2.8e-54|Constantimarinum_sp._W242            RRKIVAGNWKMNNNIS-ETENLLNKLKM---QVIPE--------SVSVMV
99 ur|J1TCP4|6_247|4.2e-54|Rhizobium                                   -RPLVAGNWKMNGTRA-SLDQIKAIAQG----VSSP--LS---EKVEALI
100 ur|A0A964MZ36|6_242|5.9e-54|Pseudolabrys_sp.                       -RPLVAGNWKMNGLAA-SVGELAKIIKG----ASGI--------AADLMV
101 ur|UPI001CFFBCBE|2_248|8.5e-54|Methylobacillus                     RRKLVVGNWKMNGNLA-QNKTLLDTLVR---DLRGI-------DHADFAV
102 ur|A0A2A5T4Q3|2_247|1.3e-53|Candidatus_Enterovibrio_escacola       RRPLVMGNWKLNGNKA-VIEDLIKGIDI---ELNGV-------EGVDIAI
103 ur|UPI001FF212EB|3_250|1.9e-53|Erythrobacter_sp._F6033             KRPYIVGNWKMHGTRA-MLSEARTIDRA---AQRHM--------KVEVAI
104 ur|V8FYU2|12_251|2.9e-53|Pelistega_indica                          ---LVVGNWKMNGSLA-LNQTLLAGLLE---KTKAT--PEK-YATVQLAV
105 ur|UPI00262D6231|9_253|4.1e-53|Pseudomonas_sp.                     --TLIIGNWKMNGDTA-RNEVLLSELLP---ALSNL-------KGVDVVV
106 ur|A0A5B8CR94|14_257|6.2e-53|Methylophilus_medardicus              ---LVVANWKMHGNLA-RNQVLIKEYLQ---ALNPL-------AQTKVVV
107 ur|A0A0P7WQU1|4_250|9.1e-53|Erythrobacteraceae                     -RPYIVGNWKMHGTRA-MLSEARAIDRA---AQRHM--------KVEVAI
108 ur|A0A3A0ABM6|5_249|1.4e-52|Chloroflexota_bacterium                --VLVAGNWKMNLTLE-DSIALAASLAA---DATTT-------TTVEIAV
109 ur|UPI001BDA55F2|5_244|2.1e-52|Alteromonas_oceanisediminis         RKTLVAGNWKMNGNAK-LLAQFVEYFSA---SRK----------RCETVL
110 ur|A0A1M5ZY50|5_244|3e-52|Wenxinia_saemankumensis                  ----IGTSWKMNKLLA-EAEGYAAALRD---APAPP--------GITRFV
111 ur|UPI00220E7553|2_248|4.2e-52|Pseudodesulfovibrio_portus          -KKLMAANWKMYKTWD-QAVDTAEGLVG---LTADK-LPD----DREVLV
112 ur|Q2NAA5|4_246|6.2e-52|Erythrobacter_litoralis__strain_HTCC2594   -RPYIVGNWKMNGTRA-MLSEARAIDRA---AQRYM--------KVEVAV
113 ur|UPI001F365E9E|4_250|9.5e-52|Polynucleobacter                    --LIIIANWKMHGSFS-RNQTWFQEVCR---TMEQG-MPS----GRRYAV
114 ur|UPI000D0FC4FF|4_246|1.4e-51|Allosphingosinicella_vermicomposti  RRKLIVGNWKMNGSLA-ALAELDPMAAA---AANCC--------DVDVGI
115 ur|Q1MTI4|4_246|2.1e-51|Cypriniformes                              RKFFVGGNWKMNGDKE-SLGELIMTLNT---ASLND--------ETDVVC
116 ur|A0A8J7VPH5|2_242|3.2e-51|Cypionkella_sp.                        -KKLAAGNWKMNGTQS-ALSEVSALIAA----VGTP--------GCAVLL
117 ur|A0A6N9STY0|3_247|4.8e-51|root                                   -NKLIAGNWKMHGSRV-ANEAWVRALQA---EPILE--------SVELAV
118 ur|A0A6A5SV51|6_245|7.3e-51|Clathrospora_elynae                    ---FVGGNFKMNGNIK-TIKDIVSNINN---AKLDP--------NTEVVI
119 ur|UPI000EF4CC05|2_242|1.2e-50|Tropicibacter_alexandrii            RRKLAAGNWKMNGLRA-QLSEIDAMQEA-----ASG--------GSEVLI
120 ur|A0A892ZFK5|7_245|1.7e-50|Paralysiella_testudinis                ----VMGNWKMNGRQA-ANAALLSALRA---LPAS---------DALVGV
121 ur|UPI0025F708F9|5_245|2.9e-50|Sediminimonas_sp.                   ---IAAGNWKMNGTRA-DLDQIDALTAA----HPAP--------AVEIML
122 ur|UPI0024A75781|2_241|4.5e-50|Pelagibius_sp._Alg239-R121          RRKLIAGNWKMNGLRA-GGAALAGGLAEK-CRAEAL--------TCDFLV
123 ur|K7XK57|7_249|7e-50|Rhizobiaceae                                 --RLIVGNWKMNGLTA-DLSEIEKVAWN----IGEL--------SADVVI
124 ur|W6MIU9|41_279|1.1e-49|Kuraishia_capsulata_CBS_1993             ---FVGGNFKMNGSKA-SIKTIVENLNS---ADLPA--------NVQVVI
125 ur|UPI00277DB149|5_243|1.9e-49|Rhizobium_paknamense                -----GTSFKMNKTLG-ETLAFCEHLKA---DGRLF-------SGIQRFI
126 ur|Q9M4S8|66_306|3.3e-49|Rosaceae                                  ---FVGGNWKCNGTKD-LISKLVSDLNS---AKLEP--------DVDVVV
127 ur|UPI001C0A632B|6_244|5.3e-49|unclassified_Gilvimarinus           --PFVIANWKLNGGID-LTCTSVASMLN---------RKF----EARIGI
128 ur|A0A3S4CV16|5_248|9.1e-49|Devosia                                -SPLIAGNWKMNGLTH-SLEELSELARLL-TTGEAP--------RALVVV
129 ur|F0YHQ6|12_252|1.5e-48|Aureococcus_anophagefferens               -KYLIGGNWKCNGTKA-STAALVKTLNE---AGPIP-------AWVEVVV
130 ur|A0A432YKP0|3_248|2.5e-48|Pseudidiomarina                        RQPLVIANWKMNGDRQ-LVTAMAEALRD---SSAQH-------ADVSVVV
131 ur|A0A2V1P728|3_241|4.6e-48|Salibaculum_griseiflavum               -RKFAAGNWKMNGLAA-DLAELEKMVAA----KPAP--------SVDTAI
132 ur|A0A9E1RU13|12_251|8.8e-48|Rhodospirillaceae_bacterium           -RPIVVANWKMNGLRR-AGRDLAEGIARLKQAGGEL--------ACDVIV
133 ur|A0A0D2JVE2|14_254|1.8e-47|Monoraphidium_neglectum               ---FVGGNWKCNGTHA-SVEKLVADLNA---GEVPS--------DIDIVV
134 ur|A0A2T7FVE6|3_239|3.8e-47|Thalassorhabdomicrobium_marinisediminis --KLIAGNWKMNGTAA-DLGELRTLAQQ----HPDP--------SVDIVI
135 ur|A0A2J0MY76|13_264|8.8e-47|unclassified_Parcubacteria_group     -KIVIIANWKCNPSTEREAENLLREIKK---GIKKI-------RNTEVII
136 ur|A0A835U5J6|43_280|2.2e-46|Vanilla_planifolia                    ----VGGNWKCNGTKD-SIKKLVGELNS---ATLEN--------DVDVVV
137 ur|A0A3B0T1D4|2_214|5.6e-46|hydrothermal_vent_metagenome           RRKLAAGNWKMNGTSA-ALSELKTLINA----HPSA--------NVDILI
138 ur|A0A7X9PHI7|2_169|1.5e-45|Deltaproteobacteria_bacterium         RKWIVAGNWKMHNTIA-ESIMLAKAIQD---GSSDI--QG-----GAVIL
139 ur|UPI00163C79AF|3_251|4.9e-45|Blattabacterium_cuenoti             -KNIVLANWKMNFDFH-ETTSFIRNLLK---FFYDQNINH----KKKIIL
140 ur|A0A950S6V8|8_238|1.7e-44|Acidobacteriota_bacterium             ------ANWKMNIPPE-GIVTYLDAVKD-----------------ANAVI
141 ur|A0A662YKY2|5_212|6.1e-44|Acipenseroidei                        RKFFVGGNWKLNGDKK-SLGELIHTLNS---GKINA--------DTDVVC
142 ur|A0A539DPV6|11_236|2.8e-43|Acidimicrobiaceae_bacterium          -KPWIGTSWKMTKRLS-EAVAYAEALAG---LAAHF--AH----LTNSFV
143 ur|A0A9D7URD6|6_241|1.3e-42|Sphingomonadaceae_bacterium           ---YIVGNWKMNGMSA-DLTEIAAIASV---SAKYA--------SVESAL
144 ur|A0A9D9MVE7|19_247|1.1e-41|Candidatus_Moranbacteria_bacterium   -----------------EAEQYLSVLRR---EAMGKH-----FGDTLGIV
145 ur|A0A6A2Y9A4|3_235|1.4e-40|Hibiscus_syriacus                     RKFFVDGNWKCNGTTE-EVKKIVTTLNE---AEVPS--ED----VVEVVV
146 ur|A0A554KS89|5_249|3.6e-39|Parcubacteria_group_bacterium_Gr01-1014_17 -KKLIVANWKMNPGTAAEARALFLATRR---AASEA-------RRVSVTV
147 ur|UPI00199EA28C|19_258|3.5e-37|Streptomyces_viridiviolaceus      -----GTSFKMTKTRA-QAVAYATRLRS---ALGGG--VP----GVQPFV
148 ur|B5TYJ5|2_162|6e-35|Lestremiinae_sp._MAB-2008                   -----SGNWKMNGTKQ-SIAGICQNLSN---ASLDP--------NTEVVV
149 ur|K7Z649|2_162|3.1e-32|Cardiacera_nr._miliacea_BMW-2012          -----SGNWKMNGDQK-SMTDLSKLLTT---GPLDP--------NTEVYV
150 ur|A0A329ZQK5|50_230|4.9e-30|Helicobacter_sp._10-6591             --------------------------------------------------
```

Figure 46. MSA on Consurf Server

Some additional results obtained from the Consurf server are given below in the image:



There are 28566 HMMER hits. 27844 of them are unique, including the query.
The calculation is performed on a sample of 150 sequences that represent the list of homologues to the query.

Here is the list of sequences that produced significant alignments, but were not chosen as hits.

The best evolutionary model was selected to be: WAG. (details).

FIGURE 47. Additional Homology related results