BT3040 – BIOINFORMATICS – Assignment 5

Submitted by Sahana (BE17B038)

Question 1

Algorithm –

1. Go to Uniprot.org -> Retrieve ID Mapping.
2. Input all the identifiers at once and map it from UniProt AC/ID to UniProtKB.
3. Download the Fasta file for all the sequences in both Sets.
4. Run them through a Multiple Sequence Alignment tool, such as CLUSTAL Omega/ MUSCLE/ MAFFT. Download the alignments in the canonical format (ClustalW)
5. Paste the MSA results in AL2CO server and select required criteria for analyses.
6. Make sure that you don't normalize for options (i) -> (iv) in the question.

Set 1 - P69905, P01946, P01942, P01966, P01958, P01959, P01965, P06635, P60529, P80043, P01980

Position	Residue	Set 1 Condition 1	Set 1 Condition 2	Set 1 Condition 3	Set 1 Condition 4	Set 1 Condition 5
'1'	':'	0	1.021	5	1.019	0.943
'2'	'S'	-0.305	0.865	3.008	0.729	0.253
'3'	'L'	0	0.909	4	0.916	0.943
'4'	'S'	0	0.953	4	0.951	0.943
'5'	'D'	-1.121	0.527	1.149	0.492	-1.597
'6'	'K'	-1.16	0.561	1.157	0.439	-1.685
'7'	'D'	0	0.973	6	0.972	0.943
'8'	'K'	0	0.956	5	0.95	0.943
'9'	'A'	-1.169	0.541	1.868	0.465	-1.705
'10'	'A'	-0.305	0.917	4.661	0.77	0.253
'11'	'V'	-0.586	0.744	3.603	0.763	-0.385
'12'	'R'	-0.305	0.871	4.504	0.744	0.253
'13'	'A'	-1.034	0.567	1.909	0.566	-1.399
'14'	'L'	-1.16	0.559	1.521	0.415	-1.685
'15'	'W'	-0.305	0.943	9.306	0.888	0.253
'16'	'S'	-0.916	0.612	2.264	0.604	-1.134
'17'	'K'	0	0.956	5	0.95	0.943
'18'	'I'	-0.689	0.691	3.504	0.719	-0.618
'19'	'G'	-0.305	0.886	4.992	0.831	0.253
'20'	'K'	-1.034	0.613	2.405	0.502	-1.399
'21'	'S'	-0.886	0.713	4.372	0.627	-1.064
'22'	'A'	-0.474	0.74	2.876	0.767	-0.131
'23'	'D'	-1.034	0.624	2.24	0.517	-1.399
'24'	'A'	-0.908	0.665	2.455	0.593	-1.114
'25'	'I'	-0.886	0.757	4.314	0.679	-1.064

'26'	'G'	0	0.977	6	0.977	0.943
'27'	'N'	-0.886	0.65	1.826	0.561	-1.064
'28'	'D'	-0.305	0.922	4.512	0.784	0.253
'29'	'A'	-0.305	0.817	3.347	0.768	0.253
'30'	'L'	0	0.909	4	0.916	0.943
'31'	'S'	-0.886	0.749	3.298	0.628	-1.064
'32'	'R'	0	1.015	5	1.011	0.943
'33'	'M'	-0.305	0.933	4.008	0.944	0.253
'34'	'I'	-0.305	0.904	4.992	0.78	0.253
'35'	'V'	-1.295	0.47	0.959	0.373	-1.991
'36'	'V'	-1.367	0.472	1.132	0.386	-2.154
'37'	'Y'	-0.474	0.83	5.14	0.718	-0.131
'38'	'P'	0	0.993	7	0.993	0.943
'39'	'Q'	-0.474	0.821	3.215	0.715	-0.131
'40'	'T'	0	0.977	5	0.979	0.943
'41'	'K'	0	0.956	5	0.95	0.943
'42'	'T'	0	0.977	5	0.979	0.943
'43'	'Y'	0	1.016	7	1.012	0.943
'44'	'F'	0	0.988	6	0.989	0.943
'45'	'S'	-0.474	0.822	4.521	0.712	-0.131
'46'	'H'	0	0.969	8	0.972	0.943
'47'	'W'	-0.6	0.823	4.24	0.711	-0.417
'48'	'P'	-0.857	0.657	2.295	0.582	-1
'49'	'D'	-0.305	0.888	5.174	0.901	0.253
'50'	'V'	-0.586	0.674	2.81	0.621	-0.385
'51'	'T'	-0.6	0.776	2.752	0.605	-0.417
'52'	'P'	-0.586	0.745	3.959	0.689	-0.385
'53'	'G'	0	0.977	6	0.977	0.943
'54'	'S'	0	0.953	4	0.951	0.943
'55'	'P'	-0.6	0.73	2.322	0.628	-0.417
'56'	'H'	-0.305	0.936	4.198	0.8	0.253
'57'	'I'	-0.305	0.872	3.835	0.748	0.253
'58'	'K'	0	0.956	5	0.95	0.943
'59'	'A'	-0.86	0.596	1.835	0.687	-1.006
'60'	'H'	0	0.969	8	0.972	0.943
'61'	'G'	0	0.977	6	0.977	0.943
'62'	'K'	-0.6	0.777	3.405	0.812	-0.417
'63'	'K'	0	0.956	5	0.95	0.943
'64'	'V'	0	0.956	4	0.956	0.943
'65'	'M'	-0.76	0.66	2.124	0.612	-0.778
'66'	'G'	-0.76	0.704	2.694	0.561	-0.778
'67'	'G'	-0.305	0.817	3.355	0.692	0.253
'68'	'I'	-0.305	0.825	3.669	0.71	0.253
'69'	'A'	-0.86	0.643	2.19	0.545	-1.006
'70'	'L'	-1.673	0.36	0.124	0.382	-2.848

'71'	'A'	0	0.906	4	0.909	0.943
'72'	'V'	-0.586	0.701	2.413	0.682	-0.385
'73'	'S'	-1.594	0.406	0.934	0.359	-2.67
'74'	'K'	-0.305	0.879	6.488	0.745	0.253
'75'	'I'	-0.995	0.543	2.612	0.534	-1.312
'76'	'D'	-0.305	0.888	5.331	0.888	0.253
'77'	'D'	0	0.973	6	0.972	0.943
'78'	'L'	-0.76	0.675	3.14	0.683	-0.778
'79'	'K'	-0.6	0.807	4.446	0.618	-0.417
'80'	'T'	-0.76	0.733	3.149	0.664	-0.778
'81'	'G'	-0.305	0.817	3.355	0.692	0.253
'82'	'L'	0	0.909	4	0.916	0.943
'83'	'M'	-0.305	0.869	3.182	0.749	0.253
'84'	'E'	-1.414	0.451	0.678	0.428	-2.262
'85'	'L'	0	0.909	4	0.916	0.943
'86'	'S'	0	0.953	4	0.951	0.943
'87'	'E'	-0.305	0.888	5.331	0.763	0.253
'88'	'Q'	-0.305	0.826	3.017	0.715	0.253
'89'	'H'	0	0.969	8	0.972	0.943
'90'	'A'	0	0.906	4	0.909	0.943
'91'	'Y'	-0.76	0.731	4.967	0.618	-0.778
'92'	'K'	0	0.956	5	0.95	0.943
'93'	'L'	0	0.909	4	0.916	0.943
'94'	'R'	0	1.015	5	1.011	0.943
'95'	'V'	0	0.956	4	0.956	0.943
'96'	'D'	0	0.973	6	0.972	0.943
'97'	'P'	0	0.993	7	0.993	0.943
'98'	'A'	-0.305	0.86	3.339	0.715	0.253
'99'	'N'	0	1.012	6	1.011	0.943
'100'	'F'	0	0.988	6	0.989	0.943
'101'	'K'	0	0.956	5	0.95	0.943
'102'	'I'	-0.6	0.739	3.058	0.632	-0.417
'103'	'L'	0	0.909	4	0.916	0.943
'104'	'N'	-0.6	0.781	2.926	0.619	-0.417
'105'	'H'	-0.305	0.886	6.653	0.839	0.253
'106'	'C'	-0.305	0.937	7.306	0.962	0.253
'107'	'I'	-0.6	0.739	3.058	0.588	-0.417
'108'	'L'	-0.305	0.825	3.678	0.783	0.253
'109'	'V'	-0.474	0.784	2.215	0.844	-0.131
'110'	'V'	-0.474	0.806	3.479	0.677	-0.131
'111'	'I'	-0.6	0.735	3.24	0.58	-0.417
'112'	'S'	-0.305	0.814	3.504	0.683	0.253
'113'	'T'	-1.72	0.345	0.702	0.355	-2.955
'114'	'M'	-0.305	0.885	6.322	0.769	0.253
'115'	'F'	-0.935	0.559	1.207	0.529	-1.175

'116'	'P'	0	0.993	7	0.993	0.943
'117'	'K'	-1.846	0.302	0.562	0.299	-3.241
'118'	'E'	-0.916	0.633	3.008	0.584	-1.134
'119'	'F'	-0.305	0.893	4.992	0.837	0.253
'120'	'T'	-0.305	0.891	4.182	0.907	0.253
'121'	'P'	0	0.993	7	0.993	0.943
'122'	'E'	-0.76	0.66	2.182	0.589	-0.778
'123'	'A'	-0.886	0.695	2.736	0.519	-1.064
'124'	'H'	0	0.969	8	0.972	0.943
'125'	'V'	-0.305	0.814	3.339	0.685	0.253
'126'	'S'	0	0.953	4	0.951	0.943
'127'	'L'	0	0.909	4	0.916	0.943
'128'	'D'	0	0.973	6	0.972	0.943
'129'	'K'	0	0.956	5	0.95	0.943
'130'	'F'	0	0.988	6	0.989	0.943
'131'	'L'	-0.305	0.822	3.355	0.839	0.253
'132'	'S'	-0.86	0.595	2.289	0.528	-1.006
'133'	'G'	-1.414	0.459	1.496	0.4	-2.262
'134'	'V'	0	0.956	4	0.956	0.943
'135'	'A'	-0.6	0.769	2.909	0.577	-0.417
'136'	'L'	-0.6	0.796	3.231	0.615	-0.417
'137'	'A'	-0.305	0.86	3.339	0.715	0.253
'138'	'L'	0	0.909	4	0.916	0.943
'139'	'A'	-0.305	0.881	4.165	0.738	0.253
'140'	'E'	-0.305	0.868	3.347	0.743	0.253
'141'	'R'	-0.305	0.871	4.504	0.744	0.253
'142'	'Y'	0	1.016	7	1.012	0.943
'143'	'R'	0	1.015	5	1.011	0.943

Set 2 - TPIS_HUMAN, TPIS_YEAST, TPIS_GRAVE, TPIS_TRYCR, TPIS_MAIZE, TPIS_MOUSE, TPIS_DROME, TPIS_RABIT, TPIS_CAEEL

Position	Residue	Set 1 Condition 1	Set 2 Condition 2	Set 2 Condition 3	Set 2 Condition 4	Set 2 Condition 5
1	-	-1.216	0.529	1.306	0.494	-1
2	-	-1.216	0.529	1.306	0.494	-1
3	-	-1.216	0.529	1.306	0.494	-1
4	-	-1.216	0.529	1.306	0.494	-1
5	-	-1.216	0.529	1.306	0.494	-1
6	-	-1.216	0.529	1.306	0.494	-1
7	-	-1.216	0.529	1.306	0.494	-1
8	-	-1.216	0.529	1.306	0.494	-1
9	-	-1.216	0.529	1.306	0.494	-1
10	-	-1.216	0.529	1.306	0.494	-1

11	-	-1.216	0.529	1.306	0.494	-1
12	-	-1.216	0.529	1.306	0.494	-1
13	-	-1.216	0.529	1.306	0.494	-1
14	-	-1.216	0.529	1.306	0.494	-1
15	-	-1.216	0.529	1.306	0.494	-1
16	-	-1.216	0.529	1.306	0.494	-1
17	-	-1.216	0.529	1.306	0.494	-1
18	-	-1.216	0.529	1.306	0.494	-1
19	-	-1.216	0.529	1.306	0.494	-1
20	-	-1.216	0.529	1.306	0.494	-1
21	-	-1.216	0.529	1.306	0.494	-1
22	-	-1.216	0.529	1.306	0.494	-1
23	-	-1.216	0.529	1.306	0.494	-1
24	-	-1.216	0.529	1.306	0.494	-1
25	-	-1.216	0.529	1.306	0.494	-1
26	-	-1.216	0.529	1.306	0.494	-1
27	-	-1.216	0.529	1.306	0.494	-1
28	-	-1.216	0.529	1.306	0.494	-1
29	-	-1.216	0.529	1.306	0.494	-1
30	-	-1.216	0.529	1.306	0.494	-1
31	-	-1.216	0.529	1.306	0.494	-1
32	-	-1.216	0.529	1.306	0.494	-1
33	-	-1.216	0.529	1.306	0.494	-1
34	-	-1.216	0.529	1.306	0.494	-1
35	-	-1.216	0.529	1.306	0.494	-1
36	-	-1.216	0.529	1.306	0.494	-1
37	-	-1.216	0.529	1.306	0.494	-1
38	-	-1.216	0.529	1.306	0.494	-1
39	-	-1.216	0.529	1.306	0.494	-1
40	-	-1.216	0.529	1.306	0.494	-1
41	-	-1.216	0.529	1.306	0.494	-1
42	-	-1.216	0.529	1.306	0.494	-1
43	-	-1.216	0.529	1.306	0.494	-1
44	-	-1.216	0.529	1.306	0.494	-1
45	-	-1.216	0.529	1.306	0.494	-1
46	-	-1.216	0.529	1.306	0.494	-1
47	-	-1.216	0.529	1.306	0.494	-1
48	-	-1.216	0.529	1.306	0.494	-1
49	-	-1.216	0.529	1.306	0.494	-1
50	-	-1.216	0.529	1.306	0.494	-1
51	-	-1.216	0.529	1.306	0.494	-1
52	-	-1.216	0.529	1.306	0.494	-1
53	M	-1.216	0.529	1.306	0.494	-1
54	A	-1.216	0.529	1.306	0.494	-1
55	S	-0.974	0.648	1.328	0.666	-0.56

56	K	-1.494	0.426	1.109	0.365	-1.507
57	P	-0.377	0.882	3.5	0.827	0.529
58	Q	-0.736	0.726	2.938	0.641	-0.125
59	P	-0.377	0.884	3.828	0.83	0.529
60	I	-0.736	0.769	3.172	0.713	-0.125
61	A	-0.377	0.801	3.125	0.739	0.529
62	A	-0.377	0.812	4.656	0.754	0.529
63	A	-0.377	0.812	4.656	0.754	0.529
64	N	0	0.995	6	0.994	1.215
65	W	-0.349	0.909	8.963	0.876	0.58
66	K	0	0.962	5	0.966	1.215
67	C	-0.937	0.679	2.395	0.639	-0.492
68	N	0	0.995	6	0.994	1.215
69	G	-0.349	0.835	4	0.81	0.58
70	S	-1.311	0.51	0.79	0.53	-1.173
71	E	-1.303	0.553	1.568	0.515	-1.159
72	S	-1.523	0.401	0.901	0.372	-1.56
73	L	-1.581	0.481	0.519	0.482	-1.666
74	L	-1.061	0.526	2.667	0.514	-0.718
75	V	-1.677	0.317	0.198	0.27	-1.841
76	P	-1.003	0.64	2.025	0.597	-0.612
77	L	-0.687	0.664	3.012	0.656	-0.036
78	I	-0.965	0.55	2.716	0.557	-0.543
79	E	-1.735	0.362	0.136	0.393	-1.946
80	T	-1.003	0.663	1.815	0.595	-0.612
81	L	-0.349	0.859	3.235	0.826	0.58
82	N	-0.349	0.888	4.988	0.885	0.58
83	A	-1.003	0.591	1.951	0.549	-0.612
84	A	-0.684	0.691	2.716	0.676	-0.031
85	T	-1.889	0.305	0.383	0.302	-2.227
86	F	-1.581	0.426	0.457	0.364	-1.666
87	D	-1.149	0.604	1.543	0.553	-0.879
88	H	-1.677	0.372	0.383	0.348	-1.841
89	D	-1.523	0.429	1.259	0.421	-1.56
90	:	-1.216	0.529	1.306	0.494	-1
91	:	-1.216	0.529	1.306	0.494	-1
92	V	-0.687	0.645	2.222	0.684	-0.036
93	Q	-0.849	0.671	3.432	0.623	-0.331
94	C	-0.349	0.829	3.074	0.795	0.58
95	V	0	0.93	4	0.93	1.215
96	V	-1.061	0.545	2.136	0.563	-0.718
97	A	-1.003	0.594	2.062	0.557	-0.612
98	P	-0.349	0.893	5.049	0.894	0.58
99	T	-0.349	0.889	5.395	0.855	0.58
100	F	-1.215	0.498	0.889	0.51	-0.998

101	L	-1.677	0.321	0.593	0.294	-1.841
102	H	-0.684	0.808	5.247	0.771	-0.031
103	I	-0.687	0.664	3.012	0.686	-0.036
104	P	-1.149	0.595	1.272	0.564	-0.879
105	M	-1.465	0.504	1.704	0.492	-1.454
106	T	-1.149	0.509	1.704	0.45	-0.879
107	K	-0.937	0.625	2.37	0.582	-0.492
108	A	-1.215	0.541	1.556	0.505	-0.998
109	R	-1.427	0.462	0.889	0.399	-1.385
110	L	-0.349	0.85	3.407	0.809	0.58
111	T	-1.216	0.529	1.306	0.494	-1
112	N	-1.523	0.455	0.877	0.44	-1.56
113	P	-1.149	0.584	1.667	0.551	-0.879
114	K	-1.427	0.467	1.469	0.421	-1.385
115	F	-1.311	0.477	1.852	0.497	-1.173
116	Q	-1.831	0.321	-0.222	0.292	-2.121
117	I	-1.003	0.611	2.531	0.549	-0.612
118	A	-0.684	0.691	2.716	0.623	-0.031
119	A	-0.349	0.798	3.235	0.793	0.58
120	Q	0	0.994	5	0.993	1.215
121	N	0	0.995	6	0.994	1.215
122	A	-0.687	0.669	3.568	0.649	-0.036
123	I	-0.849	0.723	4	0.666	-0.331
124	:	-1.074	0.596	1.234	0.506	-0.741
125	T	-1.149	0.532	0.827	0.465	-0.879
126	R	-1.523	0.408	0.654	0.38	-1.56
127	S	-1.369	0.453	1.358	0.446	-1.279
128	G	0	0.944	6	0.951	1.215
129	A	0	0.909	4	0.905	1.215
130	F	0	1.001	6	1.006	1.215
131	T	0	0.983	5	0.984	1.215
132	G	0	0.944	6	0.951	1.215
133	E	0	0.962	5	0.96	1.215
134	V	-1.149	0.559	2	0.478	-0.879
135	S	-0.349	0.873	3.235	0.84	0.58
136	L	-1.149	0.562	1.568	0.469	-0.879
137	Q	-1.465	0.391	0.741	0.375	-1.454
138	I	-0.684	0.811	3.235	0.744	-0.031
139	L	-0.937	0.601	2.914	0.557	-0.492
140	K	-0.349	0.849	3.605	0.848	0.58
141	D	-0.349	0.883	5.012	0.878	0.58
142	Y	-1.523	0.43	1.123	0.415	-1.56
143	G	0	0.944	6	0.951	1.215
144	I	-1.003	0.58	1.778	0.529	-0.612
145	S	-1.677	0.406	0.778	0.368	-1.841

146	W	0	1.013	11	1.013	1.215
147	V	0	0.93	4	0.93	1.215
148	V	-0.687	0.644	3.506	0.676	-0.036
149	L	0	0.963	4	0.959	1.215
150	G	0	0.944	6	0.951	1.215
151	H	0	1.017	8	1.017	1.215
152	S	0	0.979	4	0.977	1.215
153	E	0	0.962	5	0.96	1.215
154	R	0	0.999	5	1	1.215
155	R	0	0.999	5	1	1.215
156	L	-1.149	0.592	1.728	0.522	-0.879
157	Y	-1.273	0.484	1.827	0.463	-1.104
158	:	-1.216	0.529	1.306	0.494	-1
159	:	-1.216	0.529	1.306	0.494	-1
160	Y	-0.849	0.703	3.346	0.65	-0.331
161	G	-0.684	0.735	3.074	0.682	-0.031
162	E	0	0.962	5	0.96	1.215
163	T	-0.684	0.768	2.704	0.7	-0.031
164	N	-0.53	0.796	4.272	0.765	0.25
165	E	-1.149	0.546	1.605	0.484	-0.879
166	I	-1.149	0.577	1.642	0.508	-0.879
167	V	-0.53	0.762	3.654	0.726	0.25
168	A	-0.687	0.601	2.42	0.643	-0.036
169	E	-1.311	0.505	1.914	0.499	-1.173
170	K	0	0.962	5	0.966	1.215
171	V	-0.995	0.535	1.679	0.492	-0.598
172	A	-1.303	0.475	1.21	0.465	-1.159
173	Q	-1.149	0.638	3.346	0.569	-0.879
174	A	0	0.909	4	0.905	1.215
175	C	-0.349	0.861	3.074	0.824	0.58
176	A	-1.427	0.419	1.136	0.381	-1.385
177	:	-1.216	0.529	1.306	0.494	-1
178	A	-1.061	0.514	1.63	0.491	-0.718
179	G	-0.349	0.837	4.79	0.813	0.58
180	F	-1.003	0.642	2.444	0.566	-0.612
181	H	-0.937	0.592	1.444	0.573	-0.492
182	V	0	0.93	4	0.93	1.215
183	I	-0.349	0.858	3.802	0.849	0.58
184	V	-1.303	0.492	0.901	0.401	-1.159
185	C	0	1.017	9	1.018	1.215
186	V	-0.53	0.762	3.654	0.726	0.25
187	G	0	0.944	6	0.951	1.215
188	E	0	0.962	5	0.96	1.215
189	T	-0.965	0.592	1.333	0.598	-0.543
190	N	-0.349	0.859	2.642	0.819	0.58

191	E	-0.637	0.706	3.778	0.782	0.056
192	E	-0.349	0.857	4.407	0.85	0.58
193	R	-0.349	0.89	4.407	0.856	0.58
194	E	-0.349	0.853	4.21	0.816	0.58
195	A	-0.349	0.802	3.407	0.766	0.58
196	G	0	0.944	6	0.951	1.215
197	R	-1.677	0.414	-0.222	0.387	-1.841
198	T	0	0.983	5	0.984	1.215
199	A	-1.831	0.35	-0.062	0.283	-2.121
200	A	-1.311	0.452	0.988	0.454	-1.173
201	V	0	0.93	4	0.93	1.215
202	V	-0.684	0.726	1.84	0.692	-0.031
203	L	-1.273	0.497	0.395	0.44	-1.104
204	T	-1.215	0.533	1.309	0.562	-0.998
205	Q	0	0.994	5	0.993	1.215
206	L	-0.965	0.604	1.543	0.623	-0.543
207	A	-1.427	0.481	0.951	0.424	-1.385
208	A	-0.637	0.629	2.222	0.712	0.056
209	V	-1.149	0.562	2.407	0.491	-0.879
210	A	-1.003	0.581	1.852	0.507	-0.612
211	Q	-0.995	0.612	3.037	0.569	-0.598
212	K	-1.215	0.519	1.593	0.511	-0.998
213	L	-1.215	0.484	1.802	0.458	-0.998
214	S	-1.216	0.529	1.306	0.494	-1
215	K	-1.216	0.529	1.306	0.494	-1
216	E	-1.003	0.635	2.136	0.568	-0.612
217	A	-0.684	0.766	3.407	0.718	-0.031
218	W	0	1.013	11	1.013	1.215
219	S	-1.003	0.658	2.086	0.617	-0.612
220	R	-1.215	0.549	2	0.558	-0.998
221	V	-0.349	0.822	3.802	0.815	0.58
222	V	0	0.93	4	0.93	1.215
223	I	-1.099	0.495	2.667	0.518	-0.787
224	A	0	0.909	4	0.905	1.215
225	Y	0	1.013	7	1.011	1.215
226	E	0	0.962	5	0.96	1.215
227	P	0	0.995	7	0.997	1.215
228	V	0	0.93	4	0.93	1.215
229	W	0	1.013	11	1.013	1.215
230	A	0	0.909	4	0.905	1.215
231	I	0	0.971	4	0.97	1.215
232	G	0	0.944	6	0.951	1.215
233	T	0	0.983	5	0.984	1.215
234	G	0	0.944	6	0.951	1.215
235	K	-0.684	0.75	2.914	0.718	-0.031

236	V	-0.937	0.591	2.037	0.535	-0.492
237	A	0	0.909	4	0.905	1.215
238	T	-0.349	0.876	4.198	0.87	0.58
239	P	-0.349	0.884	5.21	0.88	0.58
240	Q	-1.215	0.53	2.185	0.514	-0.998
241	Q	-0.349	0.888	4.025	0.853	0.58
242	A	0	0.909	4	0.905	1.215
243	Q	-0.349	0.885	4.407	0.852	0.58
244	E	-0.684	0.753	3.852	0.689	-0.031
245	V	-0.349	0.822	3.802	0.788	0.58
246	H	0	1.017	8	1.017	1.215
247	E	-0.637	0.669	2.222	0.641	0.056
248	L	-1.677	0.389	-0.074	0.351	-1.841
249	L	-0.849	0.656	2.914	0.592	-0.331
250	R	0	0.999	5	1	1.215
251	R	-1.677	0.35	0.383	0.319	-1.841
252	W	-0.53	0.822	7.296	0.792	0.25
253	V	-0.349	0.85	3.407	0.808	0.58
254	R	-1.149	0.545	1.827	0.476	-0.879
255	S	-1.303	0.564	1.778	0.525	-1.159
256	K	-0.637	0.724	3.222	0.69	0.056
257	L	-1.149	0.516	2.173	0.471	-0.879
258	G	-0.849	0.673	2.296	0.656	-0.331
259	T	-1.215	0.555	1.42	0.532	-0.998
260	D	-1.677	0.333	0.543	0.319	-1.841
261	I	-0.684	0.707	3.012	0.633	-0.031
262	A	-0.349	0.802	3.407	0.796	0.58
263	A	-1.465	0.426	1.012	0.405	-1.454
264	Q	-1.003	0.651	2.099	0.575	-0.612
265	L	-0.637	0.713	2.222	0.68	0.056
266	R	0	0.999	5	1	1.215
267	I	0	0.971	4	0.97	1.215
268	L	-0.937	0.612	1.963	0.586	-0.492
269	Y	0	1.013	7	1.011	1.215
270	G	0	0.944	6	0.951	1.215
271	G	0	0.944	6	0.951	1.215
272	S	0	0.979	4	0.977	1.215
273	V	-0.349	0.815	3.21	0.776	0.58
274	T	-0.684	0.773	3.346	0.711	-0.031
275	A	-0.965	0.53	1.765	0.521	-0.543
276	K	-1.003	0.592	1.642	0.518	-0.612
277	N	-0.637	0.733	3.222	0.814	0.056
278	A	-0.687	0.669	3.568	0.65	-0.036
279	R	-1.303	0.523	1.593	0.429	-1.159
280	T	-0.849	0.666	2.272	0.596	-0.331

281	L	-0.349	0.859	3.235	0.823	0.58
282	Y	-1.003	0.59	1.383	0.52	-0.612
283	Q	-1.465	0.444	1.111	0.427	-1.454
284	M	-0.937	0.637	2.457	0.601	-0.492
285	R	-0.684	0.777	3.802	0.707	-0.031
286	D	-0.349	0.883	5.012	0.852	0.58
287	I	-0.687	0.637	3.506	0.638	-0.036
288	N	-0.349	0.883	5.012	0.847	0.58
289	G	0	0.944	6	0.951	1.215
290	F	0	1.001	6	1.006	1.215
291	L	0	0.963	4	0.959	1.215
292	V	0	0.93	4	0.93	1.215
293	G	0	0.944	6	0.951	1.215
294	G	0	0.944	6	0.951	1.215
295	A	0	0.909	4	0.905	1.215
296	S	0	0.979	4	0.977	1.215
297	L	-0.349	0.862	3.617	0.829	0.58
298	K	0	0.962	5	0.966	1.215
299	P	0	0.995	7	0.997	1.215
300	E	-0.684	0.752	3.309	0.713	-0.031
301	F	0	1.001	6	1.006	1.215
302	V	-0.684	0.715	3.259	0.678	-0.031
303	E	-0.849	0.682	3.469	0.628	-0.331
304	I	0	0.971	4	0.97	1.215
305	I	-0.349	0.858	3.802	0.824	0.58
306	E	-0.684	0.782	3.988	0.714	-0.031
307	A	-0.53	0.714	2.963	0.663	0.25
308	T	-1.465	0.462	1.062	0.457	-1.454
309	K	-1.427	0.502	1.519	0.451	-1.385
310	:	-1.216	0.529	1.306	0.494	-1
311	:	-1.216	0.529	1.306	0.494	-1
312	:	-1.216	0.529	1.306	0.494	-1
313	:	-1.216	0.529	1.306	0.494	-1
314	:	-1.216	0.529	1.306	0.494	-1
315	:	-1.216	0.529	1.306	0.494	-1
316	:	-1.216	0.529	1.306	0.494	-1

Question 2

Lowest conservation scores

Position	Residue	Set 1 Condition 1
'117'	'K'	-1.846
'113'	'T'	-1.72
'70'	'L'	-1.673
'73'	'S'	-1.594
'84'	'E'	-1.414
'133'	'G'	-1.414
'36'	'V'	-1.367
'35'	'V'	-1.295
'9'	'A'	-1.169
'6'	'K'	-1.16

Position	Residue	Set 2 Condition 1
85	T	-1.889
116	Q	-1.831
199	A	-1.831
79	E	-1.735
75	V	-1.677
88	H	-1.677
101	L	-1.677
145	S	-1.677
197	R	-1.677
248	L	-1.677

Highest conservation scores

Position	Residue	Set 1 Condition 1
'1'	':'	0
'3'	'L'	0
'4'	'S'	0
'7'	'D'	0
'8'	'K'	0
'17'	'K'	0
'26'	'G'	0
'30'	'L'	0
'32'	'R'	0
'38'	'P'	0

Position	Residue	Set 2 Condition 1
64	N	0
66	K	0
68	N	0
95	V	0
120	Q	0
121	N	0
128	G	0
129	A	0
130	F	0
131	T	0

Question 3

Algorithm –

1. The program takes a list of strings as input. This list is basically the output after MSA.
2. Next step is to calculate the unweighted frequency of all amino acids at each position from the given list.
3. Conservation score through Entropy, Variance and Sum of pairs are then calculated.

CODE-

```
import numpy as np

def conservation_MSA(s):
    total = len(s) %total number of sequences aligned
    length = len(s[1]) %length of each sequence aligned
    unweightedf = [[0 for i in range(20)] for i in range(length)]
%frequency of each AA
```

```
    AA_all
=['A','C','D','E','F','G','H','I','K','L','M','N','P','Q','R','S','T','V','
W','Y']
    AA_freq=[0]*20

    for i in range(length):
        for j in range(total):
            if s[j][i].isalpha()==1:
                a = AA_all.index(str(s[j][i]))
                unweightedf[i][a]+=1
    unweightedf = np.divide(unweightedf,total)

    for i in range(length):
        for j in range(total):
            if s[j][i].isalpha()==1:
                a = AA_all.index(str(s[j][i]))
                AA_freq[a]+=1
    AA_freq = np.divide(AA_freq,(total*length))

    %scoring matrix is taken as blosum62
    blosum62 = {
    ('W', 'F'): 1, ('L', 'R'): -2, ('S', 'P'): -1, ('V', 'T'): 0,
    ('Q', 'Q'): 5, ('N', 'A'): -2, ('Z', 'Y'): -2, ('W', 'R'): -3,
    ('Q', 'A'): -1, ('S', 'D'): 0, ('H', 'H'): 8, ('S', 'H'): -1,
    ('H', 'D'): -1, ('L', 'N'): -3, ('W', 'A'): -3, ('Y', 'M'): -1,
    ('G', 'R'): -2, ('Y', 'I'): -1, ('Y', 'E'): -2, ('B', 'Y'): -3,
    ('Y', 'A'): -2, ('V', 'D'): -3, ('B', 'S'): 0, ('Y', 'Y'): 7,
    ('G', 'N'): 0, ('E', 'C'): -4, ('Y', 'Q'): -1, ('Z', 'Z'): 4,
    ('V', 'A'): 0, ('C', 'C'): 9, ('M', 'R'): -1, ('V', 'E'): -2,
    ('T', 'N'): 0, ('P', 'P'): 7, ('V', 'I'): 3, ('V', 'S'): -2,
    ('Z', 'P'): -1, ('V', 'M'): 1, ('T', 'F'): -2, ('V', 'Q'): -2,
    ('K', 'K'): 5, ('P', 'D'): -1, ('I', 'H'): -3, ('I', 'D'): -3,
    ('T', 'R'): -1, ('P', 'L'): -3, ('K', 'G'): -2, ('M', 'N'): -2,
    ('P', 'H'): -2, ('F', 'Q'): -3, ('Z', 'G'): -2, ('X', 'L'): -1,
    ('T', 'M'): -1, ('Z', 'C'): -3, ('X', 'H'): -1, ('D', 'R'): -2,
    ('B', 'W'): -4, ('X', 'D'): -1, ('Z', 'K'): 1, ('F', 'A'): -2,
    ('Z', 'W'): -3, ('F', 'E'): -3, ('D', 'N'): 1, ('B', 'K'): 0,
    ('X', 'X'): -1, ('F', 'I'): 0, ('B', 'G'): -1, ('X', 'T'): 0,
    ('F', 'M'): 0, ('B', 'C'): -3, ('Z', 'I'): -3, ('Z', 'V'): -2,
    ('S', 'S'): 4, ('L', 'Q'): -2, ('W', 'E'): -3, ('Q', 'R'): 1,
    ('N', 'N'): 6, ('W', 'M'): -1, ('Q', 'C'): -3, ('W', 'I'): -3,
    ('S', 'C'): -1, ('L', 'A'): -1, ('S', 'G'): 0, ('L', 'E'): -3,
    ('W', 'Q'): -2, ('H', 'G'): -2, ('S', 'K'): 0, ('Q', 'N'): 0,
    ('N', 'R'): 0, ('H', 'C'): -3, ('Y', 'N'): -2, ('G', 'Q'): -2,
    ('Y', 'F'): 3, ('C', 'A'): 0, ('V', 'L'): 1, ('G', 'E'): -2,
    ('G', 'A'): 0, ('K', 'R'): 2, ('E', 'D'): 2, ('Y', 'R'): -2,
    ('M', 'Q'): 0, ('T', 'I'): -1, ('C', 'D'): -3, ('V', 'F'): -1,
    ('T', 'A'): 0, ('T', 'P'): -1, ('B', 'P'): -2, ('T', 'E'): -1,
    ('V', 'N'): -3, ('P', 'G'): -2, ('M', 'A'): -1, ('K', 'H'): -1,
    ('V', 'R'): -3, ('P', 'C'): -3, ('M', 'E'): -2, ('K', 'L'): -2,
    ('V', 'V'): 4, ('M', 'I'): 1, ('T', 'Q'): -1, ('I', 'G'): -4,
    ('P', 'K'): -1, ('M', 'M'): 5, ('K', 'D'): -1, ('I', 'C'): -1,
    ('Z', 'D'): 1, ('F', 'R'): -3, ('X', 'K'): -1, ('Q', 'D'): 0,
    ('X', 'G'): -1, ('Z', 'L'): -3, ('X', 'C'): -2, ('Z', 'H'): 0,
    ('B', 'L'): -4, ('B', 'H'): 0, ('F', 'F'): 6, ('X', 'W'): -2,
    ('B', 'D'): 4, ('D', 'A'): -2, ('S', 'L'): -2, ('X', 'S'): 0,
    ('F', 'N'): -3, ('S', 'R'): -1, ('W', 'D'): -4, ('V', 'Y'): -1,
    ('W', 'L'): -2, ('H', 'R'): 0, ('W', 'H'): -2, ('H', 'N'): 1,
    ('W', 'T'): -2, ('T', 'T'): 5, ('S', 'F'): -2, ('W', 'P'): -4,
    ('L', 'D'): -4, ('B', 'I'): -3, ('L', 'H'): -3, ('S', 'N'): 1,
    ('B', 'T'): -1, ('L', 'L'): 4, ('Y', 'K'): -2, ('E', 'Q'): 2,
```

```
    ('Y', 'G'): -3, ('Z', 'S'): 0, ('Y', 'C'): -2, ('G', 'D'): -1,
    ('B', 'V'): -3, ('E', 'A'): -1, ('Y', 'W'): 2, ('E', 'E'): 5,
    ('Y', 'S'): -2, ('C', 'N'): -3, ('V', 'C'): -1, ('T', 'H'): -2,
    ('P', 'R'): -2, ('V', 'G'): -3, ('T', 'L'): -1, ('V', 'K'): -2,
    ('K', 'Q'): 1, ('R', 'A'): -1, ('I', 'R'): -3, ('T', 'D'): -1,
    ('P', 'F'): -4, ('I', 'N'): -3, ('K', 'I'): -3, ('M', 'D'): -3,
    ('V', 'W'): -3, ('W', 'W'): 11, ('M', 'H'): -2, ('P', 'N'): -2,
    ('K', 'A'): -1, ('M', 'L'): 2, ('K', 'E'): 1, ('Z', 'E'): 4,
    ('X', 'N'): -1, ('Z', 'A'): -1, ('Z', 'M'): -1, ('X', 'F'): -1,
    ('K', 'C'): -3, ('B', 'Q'): 0, ('X', 'B'): -1, ('B', 'M'): -3,
    ('F', 'C'): -2, ('Z', 'Q'): 3, ('X', 'Z'): -1, ('F', 'G'): -3,
    ('B', 'E'): 1, ('X', 'V'): -1, ('F', 'K'): -3, ('B', 'A'): -2,
    ('X', 'R'): -1, ('D', 'D'): 6, ('W', 'G'): -2, ('Z', 'F'): -3,
    ('S', 'Q'): 0, ('W', 'C'): -2, ('W', 'K'): -3, ('H', 'Q'): 0,
    ('L', 'C'): -1, ('W', 'N'): -4, ('S', 'A'): 1, ('L', 'G'): -4,
    ('W', 'S'): -3, ('S', 'E'): 0, ('H', 'E'): 0, ('S', 'I'): -2,
    ('H', 'A'): -2, ('S', 'M'): -1, ('Y', 'L'): -1, ('Y', 'H'): 2,
    ('Y', 'D'): -3, ('E', 'R'): 0, ('X', 'P'): -2, ('G', 'G'): 6,
    ('G', 'C'): -3, ('E', 'N'): 0, ('Y', 'T'): -2, ('Y', 'P'): -3,
    ('T', 'K'): -1, ('A', 'A'): 4, ('P', 'Q'): -1, ('T', 'C'): -1,
    ('V', 'H'): -3, ('T', 'G'): -2, ('I', 'Q'): -3, ('Z', 'T'): -1,
    ('C', 'R'): -3, ('V', 'P'): -2, ('P', 'E'): -1, ('M', 'C'): -1,
    ('K', 'N'): 0, ('I', 'I'): 4, ('P', 'A'): -1, ('M', 'G'): -3,
    ('T', 'S'): 1, ('I', 'E'): -3, ('P', 'M'): -2, ('M', 'K'): -1,
    ('I', 'A'): -1, ('P', 'I'): -3, ('R', 'R'): 5, ('X', 'M'): -1,
    ('L', 'I'): 2, ('X', 'I'): -1, ('Z', 'B'): 1, ('X', 'E'): -1,
    ('Z', 'N'): 0, ('X', 'A'): 0, ('B', 'R'): -1, ('B', 'N'): 3,
    ('F', 'D'): -3, ('X', 'Y'): -1, ('Z', 'R'): 0, ('F', 'H'): -1,
    ('B', 'F'): -3, ('F', 'L'): 0, ('X', 'Q'): -1, ('B', 'B'): 4
    }
    %Entropy, Variance and Sum_of_Pairs based conservation score is
calculated based on the formula mentioned in class
    entropy = [0]*length
    variance = [0]*length
    sum_of_pairs = [0]*length
    for i in range(length):
        for j in range(20):
            if unweightedf[i][j]==0:
                entropy[i]+=0
            else:
                entropy[i]+=unweightedf[i][j]*(np.log(unweightedf[i][j]))
            variance[i]+=(unweightedf[i][j]-AA_freq[j])**2
            for k in range(20):
                pair = (AA_all[j],AA_all[k])
                if pair in blosum62:
                    pair = (AA_all[j],AA_all[k])
                else:
                    pair = (AA_all[k],AA_all[j])

                  sum_of_pairs[i]+=unweightedf[i][j]*unweightedf[i][k]*blos
                  um62[pair]
    for i in range(length):
        variance[i]=variance[i]**0.5
        sum_of_pairs[i]=sum_of_pairs[i]**0.5
%The conservation scores are printed
    print('\nEntropy of each position in alignment = ')
    print(entropy)
    print('\nVariance of each position in alignment = ')
    print(variance)
    print('\nSum of all pairs for each position in alignment = ')
    print(sum_of_pairs)
```

```
    return [entropy, variance, sum_of_pairs]
```

RUNNING THE CODE –

```
s = ['AAANWKCNGSESLLVPLIETLNAATFDHD--VQCVVAPTFLHIPMTKARLTNPKFQIAAQ','---
NWKCNLSKADIAELVSAFNAAPPIDAAHVQVVVAPPAVYLDSTRQAL-
RADFDTSAQ','VGGNFKLNGSKQSIKEIVERLNTASIPEN--
VEVVICPPATYLDYSVSLVKKPQVTVGAQ','VGGNWKCNGTTDQVEKIVKTLNEGQVPPSDVVEVVVSPPYVFL
PVVKSQL-RQEFHVAAQ','VGGNWKMNGDYASVDGIVTFLNASADNSS--VDVVVAPPAPYLAYAKSKL-
KAGVLVAAQ','VGGNWKMNGDQKSIAEIAKTLSSAALDPN--TEVVIGCPAIYLMYARNLL-
PCELGLAGQ','VGGNWKMNGRKKCLGELICTLNAANVPAG--TEVVCAPPTAYIDFARQKL-
DPKIAVAAQ','VGGNWKMNGRKQSLGELIGTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-
DPKIAVAAQ','VGGNWKMNGRKKNLGELITTLNAAKVPAD--TEVVCAPPTAYIDFARQKL-DPKIAVAAQ']

conservation_MSA(s)
```

OUTPUT –

```
Entropy of each position in alignment =
[-0.43960284170028463, -0.43960284170028463, 0.0, -0.34883209584303193, 0.0, -0.9368883075390159, 0.0, -
0.34883209584303193, -1.310783678099714, -1.3030924037617193, -1.5229550675313184, -1.5810937501718236, -1.0608569471580214, -
1.6769877743224173, -1.0027182645175161, -0.6869615765973234, -0.9649629230074277, -1.7351264569629226, -1.0027182645175161, -
0.34883209584303193, -0.34883209584303193, -1.0027182645175161, -0.6837389058487535, -1.8891591637540217, -1.5810937501718236,
-1.1490596969706202, -1.6769877743224173, -1.5229550675313182, -0.48827212829693767, -0.48827212829693767, -0.6869615765973234,
-0.8486855577264172, -0.34883209584303193, 0.0, -1.0608569471580214, -1.0027182645175161, -0.34883209584303193, -0.348832095843
03193, -1.2148896539491203, -1.6769877743224173, -0.6837389058487535, -0.6869615765973234, -1.1490596969706202, -1.464816384890
813, -1.1490596969706202, -0.936888307539016, -1.2148896539491205, -1.4270610433807247, -0.34883209584303193, -0.48827212829693
767, -1.5229550675313184, -1.1490596969706202, -1.4270610433807247, -1.310783678099714, -1.8310204811135165, -1.002718264517516
1, -0.6837389058487535, -0.34883209584303193, 0.0]

Variance of each position in alignment =
[0.6951181916850706, 0.7451213064440395, 0.7451213064440395, 0.9707978984684618, 0.9142560896876017, 0.963137448579364, 0.67163
2738168282, 0.9707978984684618, 0.8595026794414767, 0.524365185473469, 0.5593021813305866, 0.3960452649027025, 0.48097551763023
32, 0.5278849793660824, 0.3425575557659554, 0.6731627946850213, 0.6700991880391287, 0.5537563458191512, 0.3913414725634738, 0.6
657864559588694, 0.8547013532762078, 0.8654668808283189, 0.5872969397785511, 0.6849799949242456, 0.28120112744097253, 0.41382984
913237203, 0.5763336653598856, 0.36069776913426593, 0.42511194362079907, 0.22511999543727568, 0.2518654680415113, 0.63768808247
84006, 0.7013068342174947, 0.8172900115399446, 0.9198653506761115, 0.5394517869584592, 0.5872969397785511, 0.8642773331760907,
0.8609378559277941, 0.4907162817105241, 0.29686288420154666, 0.8081758151220242, 0.6700991880391287, 0.5809563564208965, 0.5076
168802545298, 0.4982066604765636, 0.6341288051837624, 0.5363916939911691, 0.45953546055356037, 0.8433106423532035, 0.2518654680
415113, 0.44866055799044363, 0.5504020229171362, 0.4873502724589816, 0.48903617310054176, 0.31762363523917964, 0.60285671099505
02, 0.684979949242456, 0.7891105583193198, 0.9878167304738846]

Sum of all pairs for each position in alignment =
[1.5713484026367723, 1.9180751668480078, 1.9180751668480078, 2.449489742783178, 2.9938207967349952, 2.23606797749979, 1.5475986
974649023, 2.449489742783178, 2.0, 0.8888888888888888, 1.2521586299538492, 0.9493337494797257, 0.7200822998230956, 1.6329931618
55452, 0.4444444444444444, 1.4229164972072996, 1.7356110390903676, 1.6480441082434805, 0.3685138655950444, 1.3471506281091268,
1.7984904506931827, 2.23330569358242, 1.3966450099973928, 1.6480441082434805, 0.6186404847588913, 0.6758625033664688, 1.2422599
874998832, 0.6186404847588913, 1.1221672153735642, 0.2721655269759087, 0.2721655269759087, 1.4907119849998598, 1.85259244450367
4, 1.7533037597843888, 2.0, 1.4614384931073228, 1.4358719981466759, 2.2470831573507426, 2.322727217818541, 0.9428090415820635,
0.769800358919501, 2.290614236454256, 1.7356110390903676, 1.12765461834358, 1.3052600138300812, 1.3052600138300812, 1.539600717
8390023, 1.247219128924647, 0.9428090415820635, 1.8459164139817945, 0.31426968052735443, 0.936238863686262, 1.2909944487358058,
1.2120791238484128, 1.3608276348795434, nan, 1.5908690070307059, 1.6480441082434805, 1.7984904506931827, 2.23606797749979]
```

Question 4

FUNCTION/PROGRAM TO COMPARE THE CONSERVATION SCORES

```
def compare_MSA(l1,l2,l3):
    similar = []
    dissimilar = []
    for i in range(len(l1)):
        if l1[i]==l2[i]==l3[i]: %If all scores from clustal/mafft/muscle
are same, then it is added to the list - Similar
            similar.append((i,l1[i]))
        elif l1[i]!=l2[i]!=l3[i]: %If all scores from clustal/mafft/muscle
don’t match, then it is added to the list - Dissimilar

            dissimilar.append((i,l1[i],l2[i],l3[i]))
```

```
    print('\nTotal number of similar scores = %d. \nSimilar residues
through Clustal Omega, MAFFT and MUSCLE are = (position, common
conservation score)' %len(similar))
    print(similar)

    print('\nTotal number of different scores = %d. \nDifferent residues
through Clustal Omega, MAFFT and MUSCLE are = (position,conservation scores
in the same order)' %len(dissimilar))
    print(dissimilar)
```

FUNCTION/PROGRAM THAT TAKES MSA FROM CLUSTAL, MAFFT, MUSCLE, COMPUTES CONSERVATION SCORE, AND COMPARES THEM. THIS FUNCTION INVOKES BOTH THE ABOVE FUNCTIONS AS WELL.

```
def result_compare():
    #The following are MSAs of Set1 sequence from Clustal Omega, MAFFT and
MUSCLE.
    cl = ['-----------MAAQIPESDQIKQFK------EFLGTYNKLTETCFLDCVKD-
FTTREVKPEETTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALAAKAGLLGQPR-----','-----------
MAAQIPESDQIKQFK------EFLGTYNKLTETCFLDCVKD-
FTTREVKPEEVTCSEHCLQKYLKMTQRISVRFQEYHIQQNEALAAKAGLLGQPR-----','-------------
--MTSEQNIQTFR------DFLTQYNLVAEQCFNSCVNE-
FGSRTVSGKEESCANNCLDKFLKMTQRVSQRFQEHQLLNAQANGAAIKVENGGKINKIQ','---
MSFLGFGGGQPQLSSQQKIQAAE--
AELDLVTDMFNKLVNNCYKKCINTSYSEGELNKNESSCLDRCVAKYFETNVQVGENMQKMGQSFNAAG----KF-
---------','MDSYSSPPMGGSGSSVSPEVMMESVKTQLAQAYAEELIETLRTKCFDKCVTKP--
GSSLGGSESSCISRCVERYMEATAIISRSLFTQR------------------------']
    ma = ['M---SFLGFGGGQPQLSSQQKIQAAEAELDL--
VTDMFNKLVNNCYKKCINTSYSEGELNKNESSCLDRCVAKYFETNVQVGENM--QKM----GQSFNA------
AGKF----','MDSYSSPPMGGSGSSVSPEVMMESVKTQLAQAYAEELIETLRTKCFDKCVTKPGS--
SLGGSESSCISRCVERYMEATAIISRSLFTQR-------------------------','M-----------
AAQIPESDQIKQFKEFLGT------YNKLTETCFLDCVKDFTTR-EVKPEETTCSEHCLQKYLKMTQRISMRF--
QEYHIQQNEALAA-----KAGLLGQPR','M-----------AAQIPESDQIKQFKEFLGT------
YNKLTETCFLDCVKDFTTR-EVKPEEVTCSEHCLQKYLKMTQRISVRF--QEYHIQQNEALAA-----
KAGLLGQPR','M---------------TSEQNIQTFRDFLTQ------YNLVAEQCFNSCVNEFGSR-
TVSGKEESCANNCLDKFLKMTQRVSQRF--QEHQLLNAQANGAAIKVENGGKINKIQ']
    mu = ['---MSFLGFGGGQPQLSSQQKIQAAEAELDL--
VTDMFNKLVNNCYKKCINTSYSEGELNKNESSCLDRCVAKYFETNVQVGENMQKMGQSFNAAGKF----------
----','MDSYSSPPMGGSGSSVSPEVMMESVKTQLAQAYAEELIETLRTKCFDKCVT-KPGS-
SLGGSESSCISRCVERYMEATAIISRSLFTQR----------------------','-----------
MAAQIPESDQIKQFKEFLGT------YNKLTETCFLDCVK-
DFTTREVKPEETTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALAA-----KAGLLGQPR','-----------
MAAQIPESDQIKQFKEFLGT------YNKLTETCFLDCVK-
DFTTREVKPEEVTCSEHCLQKYLKMTQRISVRFQEYHIQQNEALAA-----KAGLLGQPR','------------
---MTSEQNIQTFRDFLTQ------YNLVAEQCFNSCVN-
EFGSRTVSGKEESCANNCLDKFLKMTQRVSQRFQEHQLLNAQANGAAIKVENGGKINKIQ']

    cl_e, cl_v, cl_s = conservation_MSA(cl)
    ma_e, ma_v, ma_s = conservation_MSA(ma)
    mu_e, mu_v, mu_s = conservation_MSA(mu)

    print('\nComparing entropy-based conservation scores from all three
alignments: ')
    compare_MSA(cl_e,ma_e,mu_e)
    print('\nComparing variance-based conservation scores from all three
alignments: ')
    compare_MSA(cl_v,ma_v,mu_v)
    print('\nComparing sum of all pairs-based conservation scores from all
three alignments: ')
    compare_MSA(cl_s,ma_s,mu_s)
```

```
result_compare()
```

OUTPUT

```
Comparing entropy-based conservation scores from all three alignments:

Total number of similar scores = 65.
Similar residues through Clustal Omega, MAFFT and MUSCLE are =
(position, common conservation score)
[(1, -0.3218875824868201), (2, -0.3218875824868201), (4, -
0.366516292749662), (5, -0.6437751649736402), (6, -0.6437751649736402),
(7, -0.6437751649736402), (8, -0.6437751649736402), (9, -
0.366516292749662), (10, -0.366516292749662), (12, -
1.0102914577233022), (13, -1.0102914577233022), (14, -
0.6283829567464145), (16, -1.0549201679861442), (17, -
1.0549201679861442), (18, -1.0549201679861442), (19, -
1.0549201679861442), (20, -1.3321790402101223), (21, -
0.5004024235381879), (22, -1.0549201679861442), (23, -
1.3321790402101223), (24, -0.9502705392332347), (25, -
0.9502705392332347), (37, -0.9502705392332347), (38, -
0.5004024235381879), (39, -0.9502705392332347), (40, -
0.5004024235381879), (41, -1.3321790402101223), (42, -
0.9502705392332347), (43, -1.3321790402101223), (44, 0.0), (45, -
0.5004024235381879), (46, -1.3321790402101223), (47, -
1.0549201679861442), (48, 0.0), (49, -0.5004024235381879), (50, -
1.0549201679861442), (57, -0.9502705392332347), (58, -
0.6730116670092565), (59, -1.3321790402101223), (60, -
1.0549201679861442), (61, -1.3321790402101223), (62, 0.0), (63, -
1.3321790402101223), (64, -0.6730116670092565), (65, 0.0), (66, -
1.3321790402101223), (67, -1.3321790402101223), (68, -
1.0549201679861442), (69, 0.0), (70, -0.6730116670092565), (71, -
1.3321790402101223), (72, -0.5004024235381879), (73, -
0.5004024235381879), (74, -0.9502705392332347), (75, -
0.6730116670092565), (76, -0.9502705392332347), (77, -
0.5004024235381879), (78, -0.9502705392332347), (79, -
0.9502705392332347), (80, -0.6730116670092565), (81, -
0.5004024235381879), (82, -1.6094379124341005), (83, -
0.9502705392332347), (84, -0.9502705392332347), (89, -
1.0102914577233022)]

Total number of different scores = 29.
Different residues through Clustal Omega, MAFFT and MUSCLE are =
(position,conservation scores in the same order)
[(0, -0.3218875824868201, 0.0, -0.3218875824868201), (3, -
0.6437751649736402, -0.3218875824868201, -0.6437751649736402), (11, -
1.0102914577233022, -0.6437751649736402, -1.0102914577233022), (15, -
1.3321790402101223, -1.0102914577233022, -1.3321790402101223), (52, -
0.6437751649736402, -0.9502705392332347, -1.3321790402101223), (53, -
0.6283829567464145, -1.0549201679861442, -0.9502705392332347), (54, -
1.0102914577233022, -0.6730116670092565, -1.0549201679861442), (55, -
1.3321790402101223, -0.6283829567464145, -1.0549201679861442), (56, -
0.9502705392332347, -0.3218875824868201, -0.6283829567464145), (85, -
0.5004024235381879, -0.3218875824868201, -0.5004024235381879), (86, -
0.9502705392332347, -0.3218875824868201, -0.9502705392332347), (87, -
1.3321790402101223, 0.0, -1.3321790402101223), (88, -
1.3321790402101223, -0.9502705392332347, -1.3321790402101223), (90, -
1.0102914577233022, -0.688403875236482, -1.0102914577233022), (91, -
1.0102914577233022, -0.688403875236482, -1.0102914577233022), (92, -
```

0.6283829567464145, -0.688403875236482, -0.6283829567464145), (93, -1.0102914577233022, -0.688403875236482, -1.0102914577233022), (94, -0.17851484105136778, -1.0102914577233022, -0.17851484105136778), (95, -1.0102914577233022, -0.733032585499324, -1.0102914577233022), (96, -0.688403875236482, -0.6283829567464145, -1.0102914577233022), (97, -0.30649537425959444, -1.0102914577233022, -0.6283829567464145), (98, -0.688403875236482, -1.0102914577233022, -0.3218875824868201), (99, -0.688403875236482, -0.17851484105136778, -0.3218875824868201), (103, -0.688403875236482, -0.3218875824868201, -0.688403875236482), (104, -0.688403875236482, -0.3218875824868201, -0.688403875236482), (106, -0.688403875236482, -0.6283829567464145, -0.688403875236482), (107, -0.3218875824868201, -0.17851484105136778, -0.688403875236482), (108, -0.3218875824868201, -0.733032585499324, -0.688403875236482), (109, -0.3218875824868201, -1.0102914577233022, -0.688403875236482)]

Comparing **variance-based conservation scores** from all three alignments:

Total number of similar scores = 0.
Similar residues through Clustal Omega, MAFFT and MUSCLE are =
(position, common conservation score)
[]

Total number of different scores = 112.
Different residues through Clustal Omega, MAFFT and MUSCLE are =
(position,conservation scores in the same order)
[(0, 0.26079988809938753, 0.9909252759070775, 0.26079988809938753), (1, 0.2648762706916386, 0.26258416991731715, 0.2648762706916386), (2, 0.23179919617405662, 0.22980447213658295, 0.23179919617405662), (3, 0.3142146835317197, 0.2625841699173172, 0.3142146835317197), (4, 0.38472923158067035, 0.3841183899779295, 0.38472923158067035), (5, 0.27674538876751253, 0.2756229129640996, 0.27674538876751253), (6, 0.29546961928152166, 0.29409914304582613, 0.29546961928152166), (7, 0.30615497276206177, 0.3046474936938978, 0.30615497276206177), (8, 0.30145269030114147, 0.3000051297396661, 0.30145269030114147), (9, 0.4046367103303144, 0.4037146690553426, 0.4046367103303144), (10, 0.4046367103303144, 0.4037146690553426, 0.4046367103303144), (11, 0.46153719420286493, 0.2743469259256438, 0.46153719420286493), (12, 0.43852285677985 75, 0.4387017417574167, 0.43852285677985 75), (13, 0.45215611580650383, 0.4520940640866618, 0.45215611580650383), (14, 0.5745950215373522, 0.5750694151631425, 0.5745950215373522), (15, 0.4929671202348621, 0.4628321555228239, 0.4929671202348621), (16, 0.5582006898993244, 0.5583524548482134, 0.5582006898993244), (17, 0.5319930278045504, 0.5326230265360575, 0.5319930278045504), (18, 0.5128514225705659, 0.5138464416432328, 0.5128514225705659), (19, 0.5569195981248193, 0.5570942066321426, 0.5569195981248193), (20, 0.4599869674285149, 0.4605522131398444, 0.4599869674285149), (21, 0.8103011848714538, 0.8101350920332968, 0.8103011848714538), (22, 0.5163216150850957, 0.517249395861645, 0.5163216150850957), (23, 0.4433824649262549, 0.44426512233433935, 0.4433824649262549), (24, 0.6263403993982565, 0.626465656058647, 0.6263403993982565), (25, 0.6042842367425367, 0.6048078633370473, 0.6042842367425367), (26, 0.24234452200728362, 0.46055221313984446, 0.4599869674285149), (27, 0.22712994123206573, 0.61000683686989, 0.6095801917747014), (28, 0.27284639096243235, 0.9585275205724015, 0.9581318185054983), (29, 0.2648762706916386, 0.482867448278465, 0.48271791103361084), (30, 0.25942686044117974, 0.5279914844906832, 0.5272727772535324), (31, 0.299073825436504, 0.24025481496887402, 0.24234452200728362), (32, 0.46844973376150456, 0.2625841699173172, 0.2648762706916386), (33,

0.6263403993982565, 0.28807209863238187, 0.2893628644918673), (34,
0.597747685832716, 0.26391703370023584, 0.2648762706916386), (35,
0.47450968248868236, 0.284394568957584, 0.28563615003021176), (36,
0.4630822314246407, 0.28807209863238187, 0.2893628644918673), (37,
0.6492980904053866, 0.6490232261218026, 0.6492980904053866), (38,
0.7906702465475165, 0.7908487051018939, 0.7906702465475165), (39,
0.6013218856852639, 0.6019001409075262, 0.6013218856852639), (40,
0.7792960258581892, 0.7796780198980983, 0.7792960258581892), (41,
0.4790043351173604, 0.4792204092869287, 0.4790043351173604), (42,
0.5971499059734345, 0.5978055694979308, 0.5971499059734345), (43,
0.457651781446576, 0.4582609277148925, 0.457651781446576), (44,
0.9838783368042276, 0.9838179784453115, 0.9838783368042276), (45,
0.8009918914721184, 0.8009881554920367, 0.8009918914721184), (46,
0.46997204649829544, 0.47035215322729473, 0.46997204649829544), (47,
0.553703901199984, 0.5539360821730895, 0.553703901199984), (48,
0.9838783368042276, 0.9838179784453115, 0.9838783368042276), (49,
0.8041069644046629, 0.8040486962493562, 0.8041069644046629), (50,
0.5452543405772827, 0.545639419503144, 0.5452543405772827), (51,
0.47676230547136095, 0.4770188020605024, 0.24234452200728362), (52,
0.29304218892472794, 0.62814367892934, 0.47148944412188837), (53,
0.615411136840193, 0.5596078739679742, 0.6409943248499131), (54,
0.44095935533926994, 0.6645163844771135, 0.5445989443655594), (55,
0.4537330982965602, 0.61000683686989, 0.5272727772535324), (56,
0.6303192016100789, 0.24602722137923588, 0.6188833344279462), (57,
0.5893235918562632, 0.590125996635635, 0.5893235918562632), (58,
0.6825494300706696, 0.6827469727119314, 0.6825494300706696), (59,
0.45920989154176106, 0.4597897200264962, 0.45920989154176106), (60,
0.568973520778374, 0.5689352233747087, 0.568973520778374), (61,
0.4425762357942369, 0.44347462671136056, 0.4425762357942369), (62,
0.9468833139326523, 0.9474821244307083, 0.9468833139326523), (63,
0.4513655584412027, 0.45209406408666175, 0.4513655584412027), (64,
0.6639832260605654, 0.6645163844771135, 0.6639832260605654), (65,
0.9838783368042276, 0.9838179784453115, 0.9838783368042276), (66,
0.4584314984547967, 0.45902596032912873, 0.4584314984547967), (67,
0.45920989154176106, 0.4597897200264962, 0.45920989154176106), (68,
0.5917427133269241, 0.5913139611721776, 0.5917427133269241), (69,
0.9838783368042276, 0.9838179784453115, 0.9838783368042276), (70,
0.6757129643599282, 0.6760330034323495, 0.6757129643599282), (71,
0.4529452933583814, 0.45364363938484187, 0.4529452933583814), (72,
0.7806696824273534, 0.7810269379926112, 0.7806696824273534), (73,
0.8155730738416456, 0.8153158406258136, 0.8155730738416456), (74,
0.6148305308234563, 0.6151618734135698, 0.6148305308234563), (75,
0.6553206490423025, 0.6560136851187606, 0.6553206490423025), (76,
0.6353981958952503, 0.635363902339777, 0.6353981958952503), (77,
0.787502024616769, 0.7877368809822145, 0.787502024616769), (78,
0.6001328509384487, 0.6007331111671556, 0.6001328509384487), (79,
0.6325815668951398, 0.6325966459495379, 0.6325815668951398), (80,
0.700623810964055, 0.7005032706962037, 0.700623810964055), (81,
0.7737770139691014, 0.7742588469501916, 0.7737770139691014), (82,
0.379118539967302, 0.3795236034681007, 0.379118539967302), (83,
0.6314513974542607, 0.6314863479875675, 0.6314513974542607), (84,
0.6263403993982564, 0.626465656058647, 0.6263403993982564), (85,
0.768683110579253, 0.24744928378468276, 0.768683110579253), (86,
0.5905343913329171, 0.2402548149688740 8, 0.5905343913329171), (87,
0.5079533262344613, 0.9493319482907718, 0.5079533262344613), (88,
0.5037170508796965, 0.5978055694979308, 0.5037170508796965), (89,
0.4489855345154024, 0.49506645732503374, 0.4489855345154024), (90,
0.41681034937599476, 0.4513172812960188, 0.41681034937599476), (91,

0.43607274482075836, 0.43468430646984096, 0.43607274482075836), (92,
0.6078200005439311, 0.399345417002174, 0.6078200005439311), (93,
0.41681034937599476, 0.4088961485832977, 0.41681034937599476), (94,
0.7723910992522388, 0.45595812824449045, 0.7723910992522388), (95,
0.4433824649262549, 0.49718815384746884, 0.4433824649262549), (96,
0.42445192752007876, 0.5930914464355611, 0.4425762357942368), (97,
0.5807773204235573, 0.44426512233433935, 0.6001328509384486), (98,
0.41250386516432813, 0.4513172812960188, 0.24234452200728362), (99,
0.43030156724733204, 0.7724440073796811, 0.25804652720794685), (100,
0.5245564209649046, 0.24025481496887402, 0.234860466608389), (101,
0.4457924038685931, 0.25581570389900216, 0.2538604992140851), (102,
0.4028675910499222, 0.23283815019429985, 0.2255520438608259), (103,
0.4327843526727455, 0.25166726287589974, 0.41766631784383157), (104,
0.40727598074253774, 0.2236136799708682, 0.42445192752007876), (105,
0.45373309829656017, 0.41738903218209195, 0.588110299595064), (106,
0.43525297593609225, 0.5995638098489936, 0.40815194499607815), (107,
0.25804652720794685, 0.7796780198980984, 0.42192011285627645), (108,
0.2510310371899321, 0.5124789330474788, 0.4327843526727455), (109,
0.234860466608389, 0.448196687685853, 0.399305839002167), (110,
0.25804652720794685, 0.43225592319050044, 0.45920989154176106), (111,
0.22712994123206573, 0.399345417002174, 0.43113075103707915)]

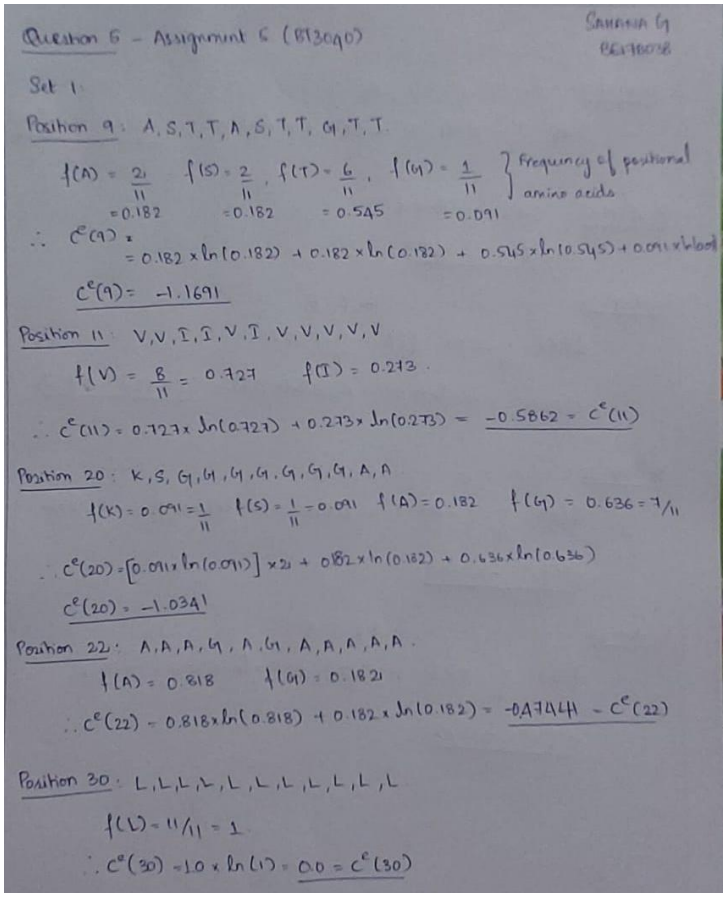
Comparing **sum of all pairs-based conservation scores** from all three alignments:

Total number of similar scores = 63.
Similar residues through Clustal Omega, MAFFT and MUSCLE are =
(position, common conservation score)
[(1, 0.48989794855663565), (2, 0.4), (4, 0.8), (5,
0.48989794855663565), (6, 0.4472135954999579), (7, 0.6000000000000001),
(8, 0.66332495807108), (9, 0.9797958971132713), (10,
0.9797958971132713), (12, 0.8717797887081347), (13, 1.0), (14, 1.4),
(16, 1.2806248474865698), (17, 1.1832159566199232), (18,
1.4000000000000001), (19, 1.0583005244258363), (20,
1.1661903789690602), (21, 1.7549928774784247), (22,
1.6124515496597103), (23, 1.077032961426901), (24, 1.32664991614216),
(25, 1.7088007490635062), (37, 1.8439088914585775), (38,
2.0099751242241783), (39, 1.16619037896906), (40, 1.7435595774162698),
(41, 0.9165151389911681), (42, 1.414213562373095), (43,
1.0954451150103324), (44, 3.0), (45, 2.2538855339169293), (47,
1.2649110640673352), (48, 3.0), (49, 1.9183326093250879), (50,
1.3416407864998738), (57, 1.414213562373095), (58, 1.6), (59,
1.0954451150103324), (60, 1.077032961426901), (61, 1.2806248474865698),
(62, 2.23606797749979), (63, 0.8944271909999161), (64,
1.6492422502470643), (65, 3.0), (66, 0.8), (67, 1.385640646055102),
(68, 1.5748015748023623), (69, 3.0), (70, 1.6), (71,
1.216552506059644), (72, 2.0099751242241783), (73, 2.383275057562597),
(74, 1.5362291495737217), (75, 1.7549928774784245), (76,
1.296148139681572), (77, 1.854723699099141), (78, 1.1832159566199232),
(79, 1.2), (80, 1.876166303929372), (81, 1.6733200530681513), (82,
0.692820323027551), (83, 1.42828568570857), (84, 1.6370705543744901)]

Total number of different scores = 33.
Different residues through Clustal Omega, MAFFT and MUSCLE are =
(position,conservation scores in the same order)
[(0, 0.447213595499958, 2.23606797749979, 0.447213595499958), (3,
0.6324555320336759, 0.5291502622129182, 0.6324555320336759), (11,
0.7483314773547883, 0.6324555320336759, 0.7483314773547883), (15,

```
1.562049935181331, 1.3564659966250538, 1.562049935181331), (46, nan,
nan, nan), (52, 0.6000000000000001, 1.0392304845413265,
1.32664991614216), (53, 1.7776388834631178, 0.7745966692414834,
1.4966629547095767), (54, 1.019803902718557, 1.6492422502470643,
1.2000000000000002), (55, 0.9591663046625439, 1.414213562373095,
1.3416407864998738), (56, 1.2165525060596438, 0.48989794855663565,
1.2489995996796797), (85, 1.5748015748023623, 0.48989794855663565,
1.5748015748023623), (86, 1.4560219778561037, 0.447213595499958,
1.4560219778561037), (87, 1.2961481396815722, 2.23606797749979,
1.2961481396815722), (88, 1.1661903789690602, 1.61245154965971,
1.1661903789690602), (89, 0.8246211251235321, 1.2806248474865698,
0.8246211251235321), (90, 0.8, 1.216552506059644, 0.8), (91,
0.7483314773547883, 1.0583005244258363, 0.7483314773547883), (92,
1.3564659966250538, 0.8, 1.3564659966250538), (93, 1.1135528725660044,
1.019803902718557, 1.1135528725660044), (94, 1.6, 1.019803902718557,
1.6), (95, 0.0, 1.4966629547095767, 0.0), (96, 0.938083151964686,
1.3564659966250538, 0.8717797887081347), (97, 1.2, 0.6324555320336759,
1.0954451150103321), (98, 0.894427190999916, 0.8944271909999161, 0.4),
(99, 0.8, 1.6, 0.4), (100, 1.0583005244258363, 0.4, 0.447213595499958),
(103, 1.0954451150103324, 0.4, 1.019803902718557), (104,
0.8485281374238571, 0.447213595499958, 0.938083151964686), (106,
1.148912529307606, 1.296148139681572, 0.7211102550927979), (107, 0.4,
1.9595917942265426, 1.0583005244258363), (108, 0.48989794855663565,
0.894427190999916, 1.0954451150103324), (109, 0.447213595499958,
1.1661903789690602, 1.077032961426901), (110, 0.4, 1.0954451150103324,
0.8944271909999159)]
```

Question 5 – Manual calculation for Set 1 and 2 with condition (i) as given in question



Question 5 – Assignment 6 (813040)

Set 1:

Position 9: A, S, T, T, A, S, T, T, G, T, T

$f(A) = \frac{2}{11} = 0.182$, $f(S) = \frac{2}{11} = 0.182$, $f(T) = \frac{6}{11} = 0.545$, $f(G) = \frac{1}{11} = 0.091$ } Frequency of positional amino acids

$\therefore C^e(9) = 0.182 \times \ln(0.182) + 0.182 \times \ln(0.182) + 0.545 \times \ln(0.545) + 0.091 \times$ blood

$C^e(9) = -1.1691$

Position 11: V, V, I, I, V, I, V, V, V, V, V

$f(V) = \frac{8}{11} = 0.727$ $f(I) = 0.273$

$\therefore C^e(11) = 0.727 \times \ln(0.727) + 0.273 \times \ln(0.273) = -0.5862 = C^e(11)$

Position 20: K, S, G, G, G, G, G, G, G, A, A

$f(K) = 0.091 = \frac{1}{11}$ $f(S) = \frac{1}{11} = 0.091$ $f(A) = 0.182$ $f(G) = 0.636 = 7/11$

$\therefore C^e(20) = [0.091 \times \ln(0.091)] \times 2 + 0.182 \times \ln(0.182) + 0.636 \times \ln(0.636)$

$C^e(20) = -1.0341$

Position 22: A, A, A, G, A, G, A, A, A, A, A

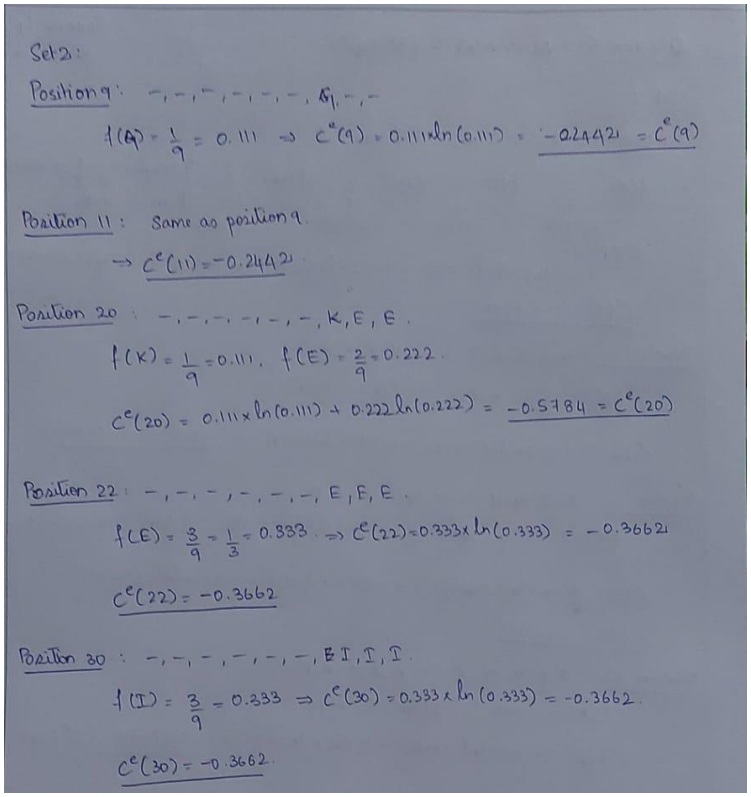
$f(A) = 0.818$ $f(G) = 0.182$

$\therefore C^e(22) = 0.818 \times \ln(0.818) + 0.182 \times \ln(0.182) = -0.47444 = C^e(22)$

Position 30: L, L, L, L, L, L, L, L, L, L, L

$f(L) = 11/11 = 1$

$\therefore C^e(30) = 1.0 \times \ln(1) = 0.0 = C^e(30)$



Question 6

Algorithm –

1. Go to ConSurf server
2. Select analyse Amino acids.
3. Select yes for protein structure. PDB ID = 1BTM
4. There is no MSA available yet. So, click No.
5. Allow the server to select homologs for analyses automatically (suggested).
6. Change parameters if needed.
7. Select Submit. The server might take a while, but a detailed report will be returned.

Result - https://consurf.tau.ac.il/results/1583001613/output.php

The below steps will be followed in order for the analysis –

1. Extract sequence from PDB file
2. Find sequence homologs
3. Align sequences
4. Select best evolutionary model
5. Calculate conservation scores
6. Project conservation scores onto the molecule

Alignment details

1. The average number of replacements between any two sequences in the alignment;

2. A distance of 0.01 means that on average, the expected replacement for every 100 positions is 1.
3. Average pairwise distance: 0.974575
4. Lower bound: 0.175208
5. Upper bound: 1.27945

The phylogenetic tree after alignment looks like this –