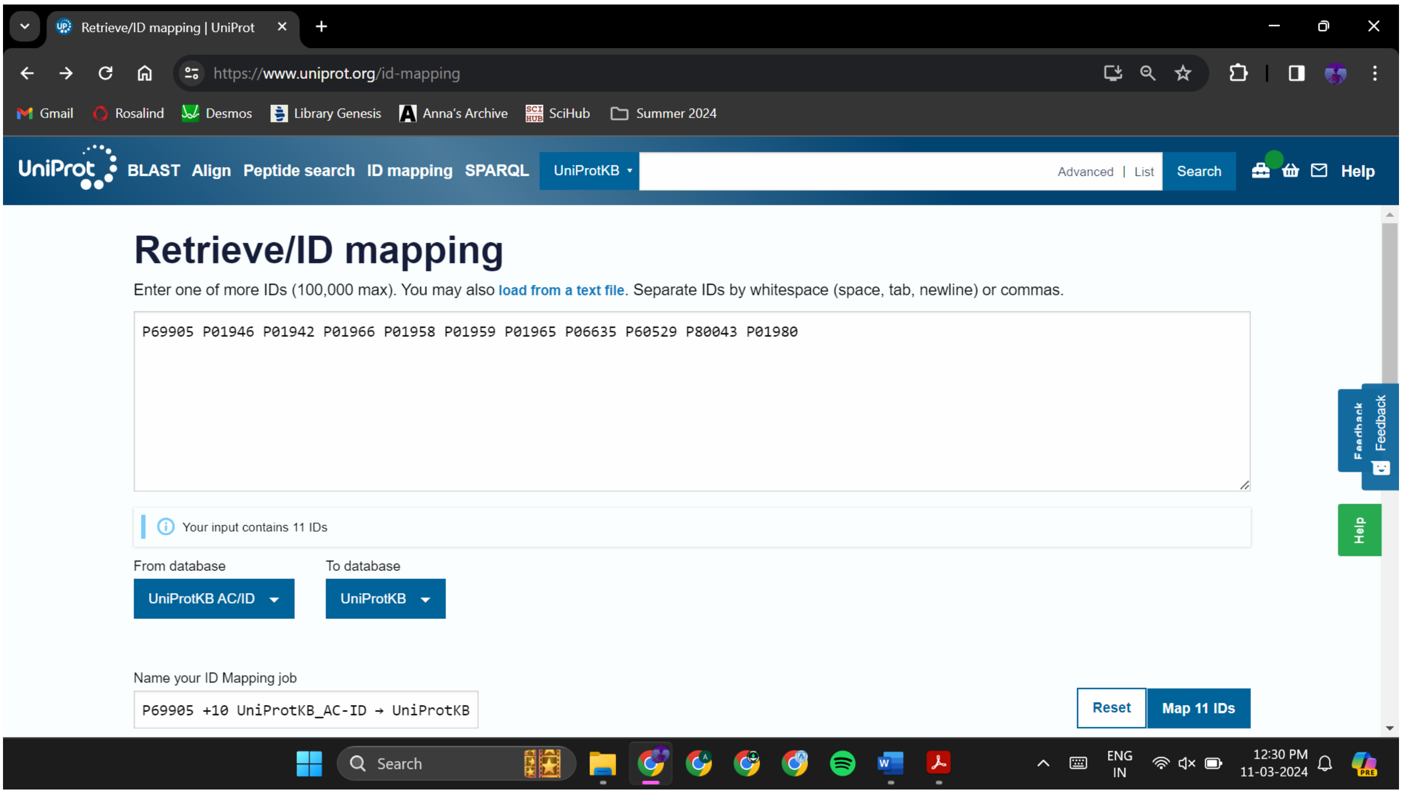
Anirudh Rao
BE21B004

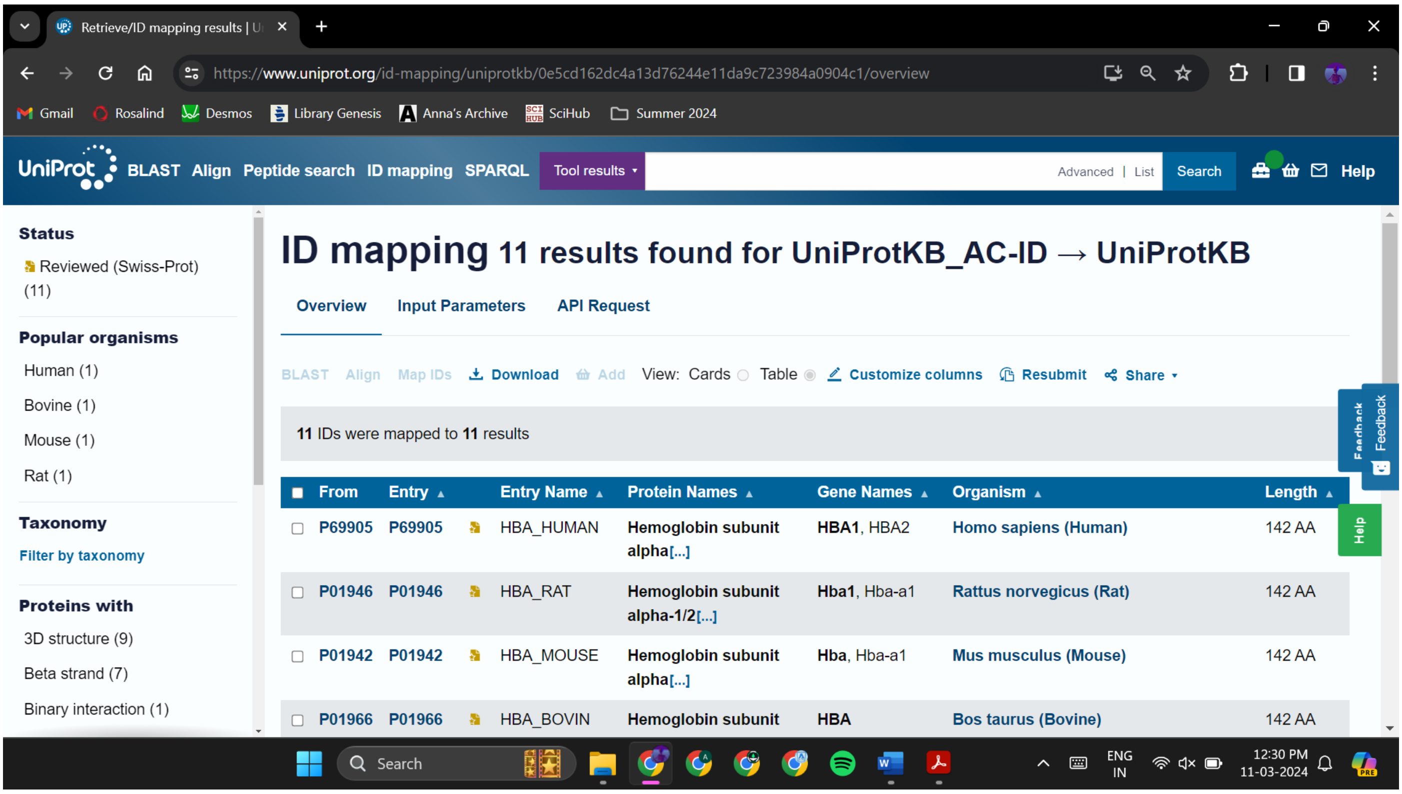
# BT3040 – Bioinformatics

## Practical 6

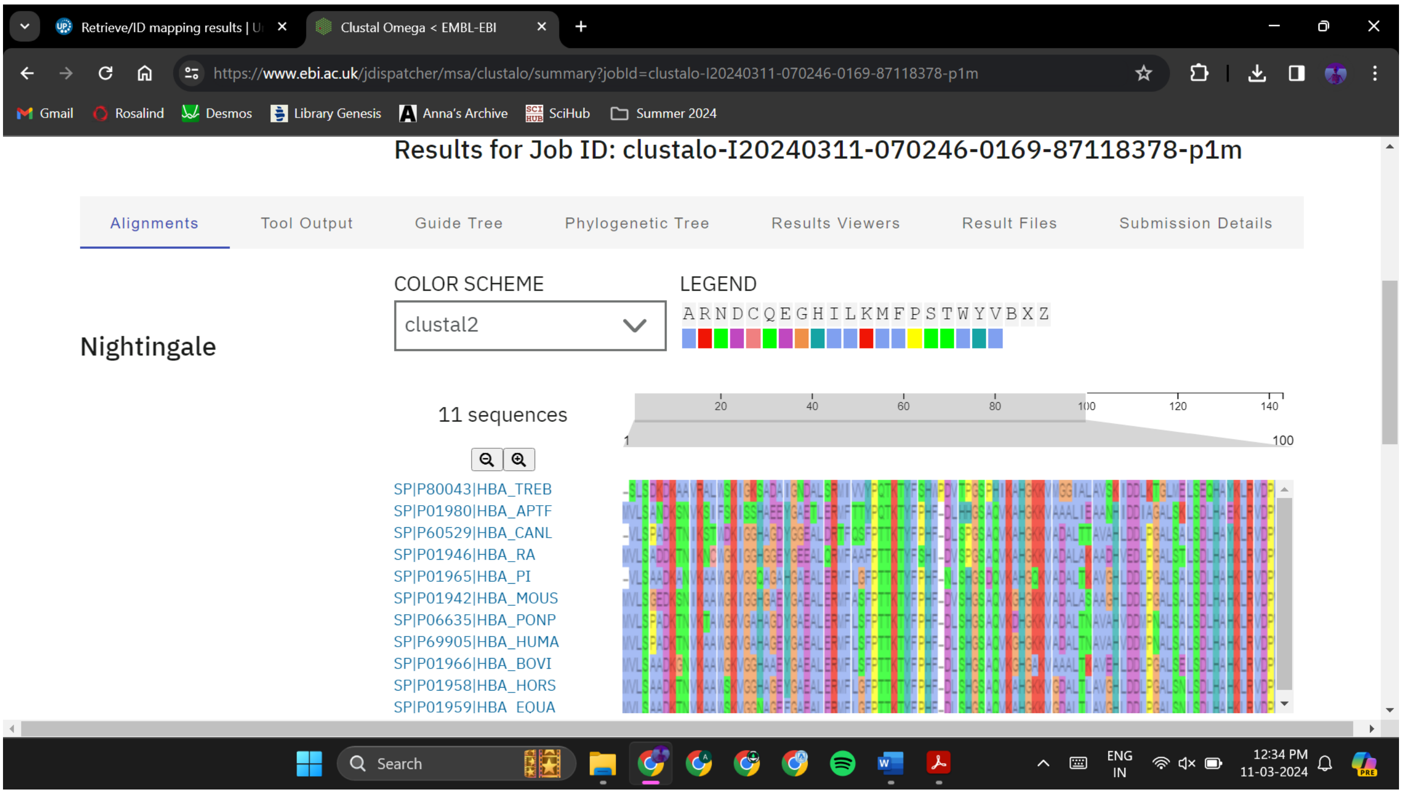
## 1

The given IDs in Set 1 were mapped using UniProt and the sequences were downloaded in FASTA format.

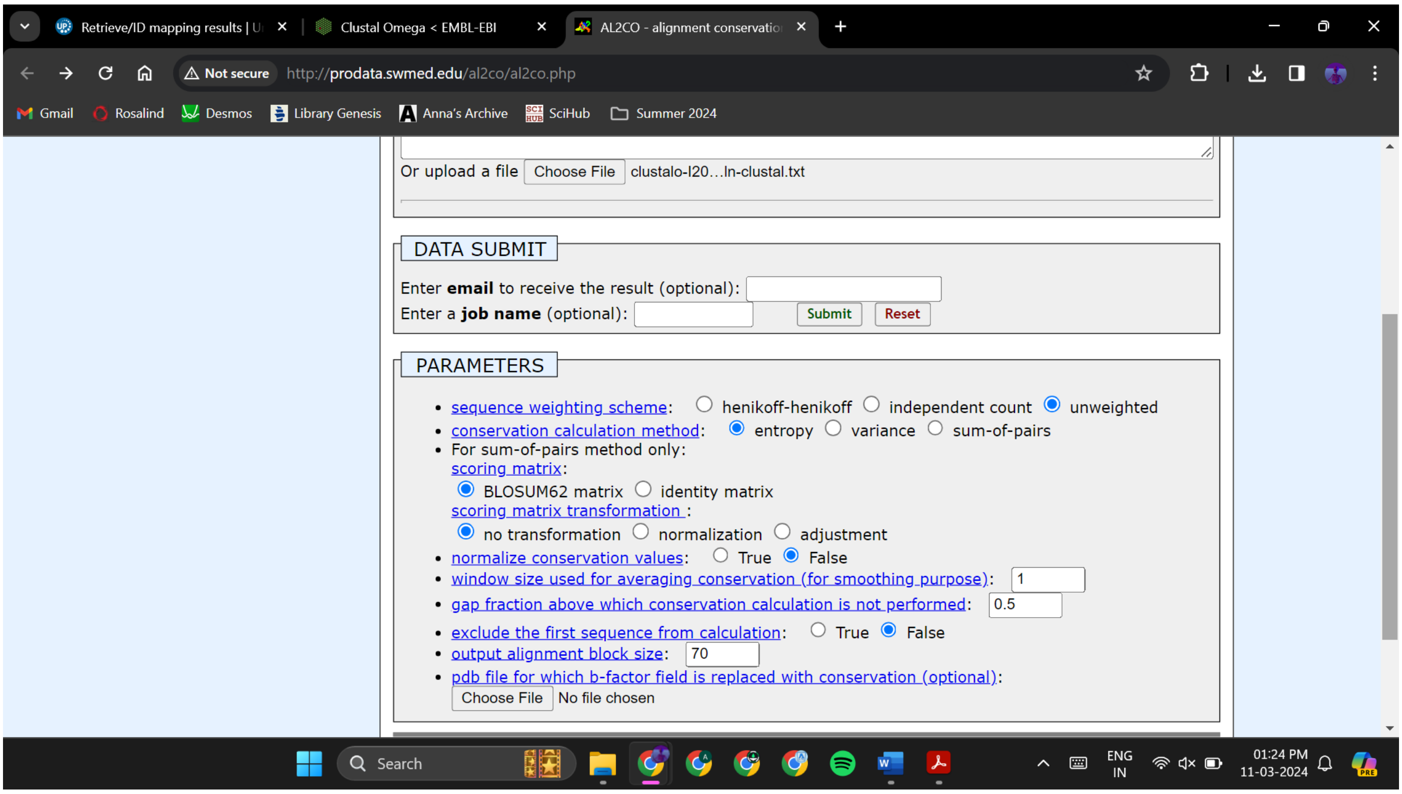


The FASTA file was fed into Clustal Omega to obtain the MSA. The result was downloaded in CLUSTAL format.




The file in CLUSTAL format was used as input for the AL2CO server along with the necessary scoring parameters.




The results are shown below:

| Set 1 | | | | | | |
|---|---|---|---|---|---|---|
| Position | Alignment | Score 1 | Score 2 | Score 3 | Score 4 | Score 5 |
| 1 | : | 0.000 | 1.021 | 5.000 | 1.019 | 0.943 |
| 2 | S | -0.305 | 0.865 | 3.008 | 0.729 | 0.253 |
| 3 | L | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 4 | S | 0.000 | 0.953 | 4.000 | 0.951 | 0.943 |
| 5 | D | -1.121 | 0.527 | 1.149 | 0.492 | -1.597 |
| 6 | K | -1.160 | 0.561 | 1.157 | 0.439 | -1.685 |
| 7 | D | 0.000 | 0.973 | 6.000 | 0.972 | 0.943 |
| 8 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 9 | A | -1.169 | 0.541 | 1.868 | 0.465 | -1.705 |
| 10 | A | -0.305 | 0.917 | 4.661 | 0.770 | 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 | -1.160 | 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 | 0.000 | 0.956 | 5.000 | 0.950 | 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 | 0.740 | 2.876 | 0.767 | -0.131 |
| 23 | D | -1.034 | 0.624 | 2.240 | 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 | 0.000 | 0.977 | 6.000 | 0.977 | 0.943 |
| 27 | N | -0.886 | 0.650 | 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 | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 31 | S | -0.886 | 0.749 | 3.298 | 0.628 | -1.064 |
| 32 | R | 0.000 | 1.015 | 5.000 | 1.011 | 0.943 |
| 33 | M | -0.305 | 0.933 | 4.008 | 0.944 | 0.253 |
| 34 | I | -0.305 | 0.904 | 4.992 | 0.780 | 0.253 |
| 35 | V | -1.295 | 0.470 | 0.959 | 0.373 | -1.991 |
| 36 | V | -1.367 | 0.472 | 1.132 | 0.386 | -2.154 |
| 37 | Y | -0.474 | 0.830 | 5.140 | 0.718 | -0.131 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 38 | P | 0.000 | 0.993 | 7.000 | 0.993 | 0.943 |
| 39 | Q | -0.474 | 0.821 | 3.215 | 0.715 | -0.131 |
| 40 | T | 0.000 | 0.977 | 5.000 | 0.979 | 0.943 |
| 41 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 42 | T | 0.000 | 0.977 | 5.000 | 0.979 | 0.943 |
| 43 | Y | 0.000 | 1.016 | 7.000 | 1.012 | 0.943 |
| 44 | F | 0.000 | 0.988 | 6.000 | 0.989 | 0.943 |
| 45 | S | -0.474 | 0.822 | 4.521 | 0.712 | -0.131 |
| 46 | H | 0.000 | 0.969 | 8.000 | 0.972 | 0.943 |
| 47 | W | -0.600 | 0.823 | 4.240 | 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 | 2.810 | 0.621 | -0.385 |
| 51 | T | -0.600 | 0.776 | 2.752 | 0.605 | -0.417 |
| 52 | P | -0.586 | 0.745 | 3.959 | 0.689 | -0.385 |
| 53 | G | 0.000 | 0.977 | 6.000 | 0.977 | 0.943 |
| 54 | S | 0.000 | 0.953 | 4.000 | 0.951 | 0.943 |
| 55 | P | -0.600 | 0.730 | 2.322 | 0.628 | -0.417 |
| 56 | H | -0.305 | 0.936 | 4.198 | 0.800 | 0.253 |
| 57 | I | -0.305 | 0.872 | 3.835 | 0.748 | 0.253 |
| 58 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 59 | A | -0.860 | 0.596 | 1.835 | 0.687 | -1.006 |
| 60 | H | 0.000 | 0.969 | 8.000 | 0.972 | 0.943 |
| 61 | G | 0.000 | 0.977 | 6.000 | 0.977 | 0.943 |
| 62 | K | -0.600 | 0.777 | 3.405 | 0.812 | -0.417 |
| 63 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 64 | V | 0.000 | 0.956 | 4.000 | 0.956 | 0.943 |
| 65 | M | -0.760 | 0.660 | 2.124 | 0.612 | -0.778 |
| 66 | G | -0.760 | 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 | 0.710 | 0.253 |
| 69 | A | -0.860 | 0.643 | 2.190 | 0.545 | -1.006 |
| 70 | L | -1.673 | 0.360 | 0.124 | 0.382 | -2.848 |
| 71 | A | 0.000 | 0.906 | 4.000 | 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 | -2.670 |
| 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 | 0.000 | 0.973 | 6.000 | 0.972 | 0.943 |
| 78 | L | -0.760 | 0.675 | 3.140 | 0.683 | -0.778 |
| 79 | K | -0.600 | 0.807 | 4.446 | 0.618 | -0.417 |
| 80 | T | -0.760 | 0.733 | 3.149 | 0.664 | -0.778 |
| 81 | G | -0.305 | 0.817 | 3.355 | 0.692 | 0.253 |
| 82 | L | 0.000 | 0.909 | 4.000 | 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 | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 86 | S | 0.000 | 0.953 | 4.000 | 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 | 0.000 | 0.969 | 8.000 | 0.972 | 0.943 |
| 90 | A | 0.000 | 0.906 | 4.000 | 0.909 | 0.943 |
| 91 | Y | -0.760 | 0.731 | 4.967 | 0.618 | -0.778 |
| 92 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 93 | L | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 94 | R | 0.000 | 1.015 | 5.000 | 1.011 | 0.943 |
| 95 | V | 0.000 | 0.956 | 4.000 | 0.956 | 0.943 |
| 96 | D | 0.000 | 0.973 | 6.000 | 0.972 | 0.943 |
| 97 | P | 0.000 | 0.993 | 7.000 | 0.993 | 0.943 |
| 98 | A | -0.305 | 0.860 | 3.339 | 0.715 | 0.253 |
| 99 | N | 0.000 | 1.012 | 6.000 | 1.011 | 0.943 |
| 100 | F | 0.000 | 0.988 | 6.000 | 0.989 | 0.943 |
| 101 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 102 | I | -0.600 | 0.739 | 3.058 | 0.632 | -0.417 |
| 103 | L | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 104 | N | -0.600 | 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 | -0.600 | 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 | -0.600 | 0.735 | 3.240 | 0.580 | -0.417 |
| 112 | S | -0.305 | 0.814 | 3.504 | 0.683 | 0.253 |
| 113 | T | -1.720 | 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 | 0.000 | 0.993 | 7.000 | 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 | 0.000 | 0.993 | 7.000 | 0.993 | 0.943 |
| 122 | E | -0.760 | 0.660 | 2.182 | 0.589 | -0.778 |
| 123 | A | -0.886 | 0.695 | 2.736 | 0.519 | -1.064 |
| 124 | H | 0.000 | 0.969 | 8.000 | 0.972 | 0.943 |
| 125 | V | -0.305 | 0.814 | 3.339 | 0.685 | 0.253 |
| 126 | S | 0.000 | 0.953 | 4.000 | 0.951 | 0.943 |
| 127 | L | 0.000 | 0.909 | 4.000 | 0.916 | 0.943 |
| 128 | D | 0.000 | 0.973 | 6.000 | 0.972 | 0.943 |
| 129 | K | 0.000 | 0.956 | 5.000 | 0.950 | 0.943 |
| 130 | F | 0.000 | 0.988 | 6.000 | 0.989 | 0.943 |
| 131 | L | -0.305 | 0.822 | 3.355 | 0.839 | 0.253 |
| 132 | S | -0.860 | 0.595 | 2.289 | 0.528 | -1.006 |
| 133 | G | -1.414 | 0.459 | 1.496 | 0.400 | -2.262 |
| 134 | V | 0.000 | 0.956 | 4.000 | 0.956 | 0.943 |
| 135 | A | -0.600 | 0.769 | 2.909 | 0.577 | -0.417 |
| 136 | L | -0.600 | 0.796 | 3.231 | 0.615 | -0.417 |
| 137 | A | -0.305 | 0.860 | 3.339 | 0.715 | 0.253 |
| 138 | L | 0.000 | 0.909 | 4.000 | 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 | 0.000 | 1.016 | 7.000 | 1.012 | 0.943 |
| 143 | R | 0.000 | 1.015 | 5.000 | 1.011 | 0.943 |

The same steps were repeated for Set 2. The results are shown below:

| Set 2 | | | | | | |
|---|---|---|---|---|---|---|
| **Position** | **Alignment** | **Score 1** | **Score 2** | **Score 3** | **Score 4** | **Score 5** |
| 1 | M | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 2 | A | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 3 | S | -0.974 | 0.648 | 1.328 | 0.666 | -0.56 |
| 4 | K | -1.494 | 0.426 | 1.109 | 0.365 | -1.507 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 | P | -0.377 | 0.882 | 3.5 | 0.827 | 0.529 |
| 6 | Q | -0.736 | 0.726 | 2.938 | 0.641 | -0.125 |
| 7 | P | -0.377 | 0.884 | 3.828 | 0.83 | 0.529 |
| 8 | I | -0.736 | 0.769 | 3.172 | 0.713 | -0.125 |
| 9 | A | -0.377 | 0.801 | 3.125 | 0.739 | 0.529 |
| 10 | A | -0.377 | 0.812 | 4.656 | 0.754 | 0.529 |
| 11 | A | -0.377 | 0.812 | 4.656 | 0.754 | 0.529 |
| 12 | N | 0 | 0.995 | 6 | 0.994 | 1.215 |
| 13 | W | -0.349 | 0.909 | 8.963 | 0.876 | 0.58 |
| 14 | K | 0 | 0.962 | 5 | 0.966 | 1.215 |
| 15 | C | -0.937 | 0.679 | 2.395 | 0.639 | -0.492 |
| 16 | N | 0 | 0.995 | 6 | 0.994 | 1.215 |
| 17 | G | -0.349 | 0.835 | 4 | 0.81 | 0.58 |
| 18 | S | -1.311 | 0.51 | 0.79 | 0.53 | -1.173 |
| 19 | E | -1.303 | 0.553 | 1.568 | 0.515 | -1.159 |
| 20 | S | -1.523 | 0.401 | 0.901 | 0.372 | -1.56 |
| 21 | L | -1.581 | 0.481 | 0.519 | 0.482 | -1.666 |
| 22 | L | -1.061 | 0.526 | 2.667 | 0.514 | -0.718 |
| 23 | V | -1.677 | 0.317 | 0.198 | 0.27 | -1.841 |
| 24 | P | -1.003 | 0.64 | 2.025 | 0.597 | -0.612 |
| 25 | L | -0.687 | 0.664 | 3.012 | 0.656 | -0.036 |
| 26 | I | -0.965 | 0.55 | 2.716 | 0.557 | -0.543 |
| 27 | E | -1.735 | 0.362 | 0.136 | 0.393 | -1.946 |
| 28 | T | -1.003 | 0.663 | 1.815 | 0.595 | -0.612 |
| 29 | L | -0.349 | 0.859 | 3.235 | 0.826 | 0.58 |
| 30 | N | -0.349 | 0.888 | 4.988 | 0.885 | 0.58 |
| 31 | A | -1.003 | 0.591 | 1.951 | 0.549 | -0.612 |
| 32 | A | -0.684 | 0.691 | 2.716 | 0.676 | -0.031 |
| 33 | T | -1.889 | 0.305 | 0.383 | 0.302 | -2.227 |
| 34 | F | -1.581 | 0.426 | 0.457 | 0.364 | -1.666 |
| 35 | D | -1.149 | 0.604 | 1.543 | 0.553 | -0.879 |
| 36 | H | -1.677 | 0.372 | 0.383 | 0.348 | -1.841 |
| 37 | D | -1.523 | 0.429 | 1.259 | 0.421 | -1.56 |
| 38 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 39 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 40 | V | -0.687 | 0.645 | 2.222 | 0.684 | -0.036 |
| 41 | Q | -0.849 | 0.671 | 3.432 | 0.623 | -0.331 |
| 42 | C | -0.349 | 0.829 | 3.074 | 0.795 | 0.58 |
| 43 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 44 | V | -1.061 | 0.545 | 2.136 | 0.563 | -0.718 |
| 45 | A | -1.003 | 0.594 | 2.062 | 0.557 | -0.612 |
| 46 | P | -0.349 | 0.893 | 5.049 | 0.894 | 0.58 |
| 47 | T | -0.349 | 0.889 | 5.395 | 0.855 | 0.58 |
| 48 | F | -1.215 | 0.498 | 0.889 | 0.51 | -0.998 |
| 49 | L | -1.677 | 0.321 | 0.593 | 0.294 | -1.841 |
| 50 | H | -0.684 | 0.808 | 5.247 | 0.771 | -0.031 |
| 51 | I | -0.687 | 0.664 | 3.012 | 0.686 | -0.036 |
| 52 | P | -1.149 | 0.595 | 1.272 | 0.564 | -0.879 |
| 53 | M | -1.465 | 0.504 | 1.704 | 0.492 | -1.454 |
| 54 | T | -1.149 | 0.509 | 1.704 | 0.45 | -0.879 |
| 55 | K | -0.937 | 0.625 | 2.37 | 0.582 | -0.492 |
| 56 | A | -1.215 | 0.541 | 1.556 | 0.505 | -0.998 |
| 57 | R | -1.427 | 0.462 | 0.889 | 0.399 | -1.385 |
| 58 | L | -0.349 | 0.85 | 3.407 | 0.809 | 0.58 |
| 59 | T | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 60 | N | -1.523 | 0.455 | 0.877 | 0.44 | -1.56 |
| 61 | P | -1.149 | 0.584 | 1.667 | 0.551 | -0.879 |
| 62 | K | -1.427 | 0.467 | 1.469 | 0.421 | -1.385 |
| 63 | F | -1.311 | 0.477 | 1.852 | 0.497 | -1.173 |
| 64 | Q | -1.831 | 0.321 | -0.222 | 0.292 | -2.121 |
| 65 | I | -1.003 | 0.611 | 2.531 | 0.549 | -0.612 |
| 66 | A | -0.684 | 0.691 | 2.716 | 0.623 | -0.031 |
| 67 | A | -0.349 | 0.798 | 3.235 | 0.793 | 0.58 |
| 68 | Q | 0 | 0.994 | 5 | 0.993 | 1.215 |
| 69 | N | 0 | 0.995 | 6 | 0.994 | 1.215 |
| 70 | A | -0.687 | 0.669 | 3.568 | 0.649 | -0.036 |
| 71 | I | -0.849 | 0.723 | 4 | 0.666 | -0.331 |
| 72 | : | -1.074 | 0.596 | 1.234 | 0.506 | -0.741 |
| 73 | T | -1.149 | 0.532 | 0.827 | 0.465 | -0.879 |
| 74 | R | -1.523 | 0.408 | 0.654 | 0.38 | -1.56 |
| 75 | S | -1.369 | 0.453 | 1.358 | 0.446 | -1.279 |
| 76 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 77 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 78 | F | 0 | 1.001 | 6 | 1.006 | 1.215 |
| 79 | T | 0 | 0.983 | 5 | 0.984 | 1.215 |
| 80 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 81 | E | 0 | 0.962 | 5 | 0.96 | 1.215 |
| 82 | V | -1.149 | 0.559 | 2 | 0.478 | -0.879 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 83 | S | -0.349 | 0.873 | 3.235 | 0.84 | 0.58 |
| 84 | L | -1.149 | 0.562 | 1.568 | 0.469 | -0.879 |
| 85 | Q | -1.465 | 0.391 | 0.741 | 0.375 | -1.454 |
| 86 | I | -0.684 | 0.811 | 3.235 | 0.744 | -0.031 |
| 87 | L | -0.937 | 0.601 | 2.914 | 0.557 | -0.492 |
| 88 | K | -0.349 | 0.849 | 3.605 | 0.848 | 0.58 |
| 89 | D | -0.349 | 0.883 | 5.012 | 0.878 | 0.58 |
| 90 | Y | -1.523 | 0.43 | 1.123 | 0.415 | -1.56 |
| 91 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 92 | I | -1.003 | 0.58 | 1.778 | 0.529 | -0.612 |
| 93 | S | -1.677 | 0.406 | 0.778 | 0.368 | -1.841 |
| 94 | W | 0 | 1.013 | 11 | 1.013 | 1.215 |
| 95 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 96 | V | -0.687 | 0.644 | 3.506 | 0.676 | -0.036 |
| 97 | L | 0 | 0.963 | 4 | 0.959 | 1.215 |
| 98 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 99 | H | 0 | 1.017 | 8 | 1.017 | 1.215 |
| 100 | S | 0 | 0.979 | 4 | 0.977 | 1.215 |
| 101 | E | 0 | 0.962 | 5 | 0.96 | 1.215 |
| 102 | R | 0 | 0.999 | 5 | 1 | 1.215 |
| 103 | R | 0 | 0.999 | 5 | 1 | 1.215 |
| 104 | L | -1.149 | 0.592 | 1.728 | 0.522 | -0.879 |
| 105 | Y | -1.273 | 0.484 | 1.827 | 0.463 | -1.104 |
| 106 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 107 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 108 | Y | -0.849 | 0.703 | 3.346 | 0.65 | -0.331 |
| 109 | G | -0.684 | 0.735 | 3.074 | 0.682 | -0.031 |
| 110 | E | 0 | 0.962 | 5 | 0.96 | 1.215 |
| 111 | T | -0.684 | 0.768 | 2.704 | 0.7 | -0.031 |
| 112 | N | -0.53 | 0.796 | 4.272 | 0.765 | 0.25 |
| 113 | E | -1.149 | 0.546 | 1.605 | 0.484 | -0.879 |
| 114 | I | -1.149 | 0.577 | 1.642 | 0.508 | -0.879 |
| 115 | V | -0.53 | 0.762 | 3.654 | 0.726 | 0.25 |
| 116 | A | -0.687 | 0.601 | 2.42 | 0.643 | -0.036 |
| 117 | E | -1.311 | 0.505 | 1.914 | 0.499 | -1.173 |
| 118 | K | 0 | 0.962 | 5 | 0.966 | 1.215 |
| 119 | V | -0.995 | 0.535 | 1.679 | 0.492 | -0.598 |
| 120 | A | -1.303 | 0.475 | 1.21 | 0.465 | -1.159 |
| 121 | Q | -1.149 | 0.638 | 3.346 | 0.569 | -0.879 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 122 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 123 | C | -0.349 | 0.861 | 3.074 | 0.824 | 0.58 |
| 124 | A | -1.427 | 0.419 | 1.136 | 0.381 | -1.385 |
| 125 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 126 | A | -1.061 | 0.514 | 1.63 | 0.491 | -0.718 |
| 127 | G | -0.349 | 0.837 | 4.79 | 0.813 | 0.58 |
| 128 | F | -1.003 | 0.642 | 2.444 | 0.566 | -0.612 |
| 129 | H | -0.937 | 0.592 | 1.444 | 0.573 | -0.492 |
| 130 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 131 | I | -0.349 | 0.858 | 3.802 | 0.849 | 0.58 |
| 132 | V | -1.303 | 0.492 | 0.901 | 0.401 | -1.159 |
| 133 | C | 0 | 1.017 | 9 | 1.018 | 1.215 |
| 134 | V | -0.53 | 0.762 | 3.654 | 0.726 | 0.25 |
| 135 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 136 | E | 0 | 0.962 | 5 | 0.96 | 1.215 |
| 137 | T | -0.965 | 0.592 | 1.333 | 0.598 | -0.543 |
| 138 | N | -0.349 | 0.859 | 2.642 | 0.819 | 0.58 |
| 139 | E | -0.637 | 0.706 | 3.778 | 0.782 | 0.056 |
| 140 | E | -0.349 | 0.857 | 4.407 | 0.85 | 0.58 |
| 141 | R | -0.349 | 0.89 | 4.407 | 0.856 | 0.58 |
| 142 | E | -0.349 | 0.853 | 4.21 | 0.816 | 0.58 |
| 143 | A | -0.349 | 0.802 | 3.407 | 0.766 | 0.58 |
| 144 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 145 | R | -1.677 | 0.414 | -0.222 | 0.387 | -1.841 |
| 146 | T | 0 | 0.983 | 5 | 0.984 | 1.215 |
| 147 | A | -1.831 | 0.35 | -0.062 | 0.283 | -2.121 |
| 148 | A | -1.311 | 0.452 | 0.988 | 0.454 | -1.173 |
| 149 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 150 | V | -0.684 | 0.726 | 1.84 | 0.692 | -0.031 |
| 151 | L | -1.273 | 0.497 | 0.395 | 0.44 | -1.104 |
| 152 | T | -1.215 | 0.533 | 1.309 | 0.562 | -0.998 |
| 153 | Q | 0 | 0.994 | 5 | 0.993 | 1.215 |
| 154 | L | -0.965 | 0.604 | 1.543 | 0.623 | -0.543 |
| 155 | A | -1.427 | 0.481 | 0.951 | 0.424 | -1.385 |
| 156 | A | -0.637 | 0.629 | 2.222 | 0.712 | 0.056 |
| 157 | V | -1.149 | 0.562 | 2.407 | 0.491 | -0.879 |
| 158 | A | -1.003 | 0.581 | 1.852 | 0.507 | -0.612 |
| 159 | Q | -0.995 | 0.612 | 3.037 | 0.569 | -0.598 |
| 160 | K | -1.215 | 0.519 | 1.593 | 0.511 | -0.998 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 161 | L | -1.215 | 0.484 | 1.802 | 0.458 | -0.998 |
| 162 | S | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 163 | K | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 164 | E | -1.003 | 0.635 | 2.136 | 0.568 | -0.612 |
| 165 | A | -0.684 | 0.766 | 3.407 | 0.718 | -0.031 |
| 166 | W | 0 | 1.013 | 11 | 1.013 | 1.215 |
| 167 | S | -1.003 | 0.658 | 2.086 | 0.617 | -0.612 |
| 168 | R | -1.215 | 0.549 | 2 | 0.558 | -0.998 |
| 169 | V | -0.349 | 0.822 | 3.802 | 0.815 | 0.58 |
| 170 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 171 | I | -1.099 | 0.495 | 2.667 | 0.518 | -0.787 |
| 172 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 173 | Y | 0 | 1.013 | 7 | 1.011 | 1.215 |
| 174 | E | 0 | 0.962 | 5 | 0.96 | 1.215 |
| 175 | P | 0 | 0.995 | 7 | 0.997 | 1.215 |
| 176 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 177 | W | 0 | 1.013 | 11 | 1.013 | 1.215 |
| 178 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 179 | I | 0 | 0.971 | 4 | 0.97 | 1.215 |
| 180 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 181 | T | 0 | 0.983 | 5 | 0.984 | 1.215 |
| 182 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 183 | K | -0.684 | 0.75 | 2.914 | 0.718 | -0.031 |
| 184 | V | -0.937 | 0.591 | 2.037 | 0.535 | -0.492 |
| 185 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 186 | T | -0.349 | 0.876 | 4.198 | 0.87 | 0.58 |
| 187 | P | -0.349 | 0.884 | 5.21 | 0.88 | 0.58 |
| 188 | Q | -1.215 | 0.53 | 2.185 | 0.514 | -0.998 |
| 189 | Q | -0.349 | 0.888 | 4.025 | 0.853 | 0.58 |
| 190 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 191 | Q | -0.349 | 0.885 | 4.407 | 0.852 | 0.58 |
| 192 | E | -0.684 | 0.753 | 3.852 | 0.689 | -0.031 |
| 193 | V | -0.349 | 0.822 | 3.802 | 0.788 | 0.58 |
| 194 | H | 0 | 1.017 | 8 | 1.017 | 1.215 |
| 195 | E | -0.637 | 0.669 | 2.222 | 0.641 | 0.056 |
| 196 | L | -1.677 | 0.389 | -0.074 | 0.351 | -1.841 |
| 197 | L | -0.849 | 0.656 | 2.914 | 0.592 | -0.331 |
| 198 | R | 0 | 0.999 | 5 | 1 | 1.215 |
| 199 | R | -1.677 | 0.35 | 0.383 | 0.319 | -1.841 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 200 | W | -0.53 | 0.822 | 7.296 | 0.792 | 0.25 |
| 201 | V | -0.349 | 0.85 | 3.407 | 0.808 | 0.58 |
| 202 | R | -1.149 | 0.545 | 1.827 | 0.476 | -0.879 |
| 203 | S | -1.303 | 0.564 | 1.778 | 0.525 | -1.159 |
| 204 | K | -0.637 | 0.724 | 3.222 | 0.69 | 0.056 |
| 205 | L | -1.149 | 0.516 | 2.173 | 0.471 | -0.879 |
| 206 | G | -0.849 | 0.673 | 2.296 | 0.656 | -0.331 |
| 207 | T | -1.215 | 0.555 | 1.42 | 0.532 | -0.998 |
| 208 | D | -1.677 | 0.333 | 0.543 | 0.319 | -1.841 |
| 209 | I | -0.684 | 0.707 | 3.012 | 0.633 | -0.031 |
| 210 | A | -0.349 | 0.802 | 3.407 | 0.796 | 0.58 |
| 211 | A | -1.465 | 0.426 | 1.012 | 0.405 | -1.454 |
| 212 | Q | -1.003 | 0.651 | 2.099 | 0.575 | -0.612 |
| 213 | L | -0.637 | 0.713 | 2.222 | 0.68 | 0.056 |
| 214 | R | 0 | 0.999 | 5 | 1 | 1.215 |
| 215 | I | 0 | 0.971 | 4 | 0.97 | 1.215 |
| 216 | L | -0.937 | 0.612 | 1.963 | 0.586 | -0.492 |
| 217 | Y | 0 | 1.013 | 7 | 1.011 | 1.215 |
| 218 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 219 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 220 | S | 0 | 0.979 | 4 | 0.977 | 1.215 |
| 221 | V | -0.349 | 0.815 | 3.21 | 0.776 | 0.58 |
| 222 | T | -0.684 | 0.773 | 3.346 | 0.711 | -0.031 |
| 223 | A | -0.965 | 0.53 | 1.765 | 0.521 | -0.543 |
| 224 | K | -1.003 | 0.592 | 1.642 | 0.518 | -0.612 |
| 225 | N | -0.637 | 0.733 | 3.222 | 0.814 | 0.056 |
| 226 | A | -0.687 | 0.669 | 3.568 | 0.65 | -0.036 |
| 227 | R | -1.303 | 0.523 | 1.593 | 0.429 | -1.159 |
| 228 | T | -0.849 | 0.666 | 2.272 | 0.596 | -0.331 |
| 229 | L | -0.349 | 0.859 | 3.235 | 0.823 | 0.58 |
| 230 | Y | -1.003 | 0.59 | 1.383 | 0.52 | -0.612 |
| 231 | Q | -1.465 | 0.444 | 1.111 | 0.427 | -1.454 |
| 232 | M | -0.937 | 0.637 | 2.457 | 0.601 | -0.492 |
| 233 | R | -0.684 | 0.777 | 3.802 | 0.707 | -0.031 |
| 234 | D | -0.349 | 0.883 | 5.012 | 0.852 | 0.58 |
| 235 | I | -0.687 | 0.637 | 3.506 | 0.638 | -0.036 |
| 236 | N | -0.349 | 0.883 | 5.012 | 0.847 | 0.58 |
| 237 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 238 | F | 0 | 1.001 | 6 | 1.006 | 1.215 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 239 | L | 0 | 0.963 | 4 | 0.959 | 1.215 |
| 240 | V | 0 | 0.93 | 4 | 0.93 | 1.215 |
| 241 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 242 | G | 0 | 0.944 | 6 | 0.951 | 1.215 |
| 243 | A | 0 | 0.909 | 4 | 0.905 | 1.215 |
| 244 | S | 0 | 0.979 | 4 | 0.977 | 1.215 |
| 245 | L | -0.349 | 0.862 | 3.617 | 0.829 | 0.58 |
| 246 | K | 0 | 0.962 | 5 | 0.966 | 1.215 |
| 247 | P | 0 | 0.995 | 7 | 0.997 | 1.215 |
| 248 | E | -0.684 | 0.752 | 3.309 | 0.713 | -0.031 |
| 249 | F | 0 | 1.001 | 6 | 1.006 | 1.215 |
| 250 | V | -0.684 | 0.715 | 3.259 | 0.678 | -0.031 |
| 251 | E | -0.849 | 0.682 | 3.469 | 0.628 | -0.331 |
| 252 | I | 0 | 0.971 | 4 | 0.97 | 1.215 |
| 253 | I | -0.349 | 0.858 | 3.802 | 0.824 | 0.58 |
| 254 | E | -0.684 | 0.782 | 3.988 | 0.714 | -0.031 |
| 255 | A | -0.53 | 0.714 | 2.963 | 0.663 | 0.25 |
| 256 | T | -1.465 | 0.462 | 1.062 | 0.457 | -1.454 |
| 257 | K | -1.427 | 0.502 | 1.519 | 0.451 | -1.385 |
| 258 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 259 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 260 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 261 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 262 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 263 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |
| 264 | : | -1.216 | 0.529 | 1.306 | 0.494 | -1 |

## 2

The top 10 residues with the highest scores in Set 1 and Set 2 obtained by unweighted frequency and entropy-based measure are:

| Set 1 | | | Set 2 | | |
|---|---|---|---|---|---|
| 1 | : | 0.00 | 12 | N | 0.00 |
| 3 | L | 0.00 | 14 | K | 0.00 |
| 4 | S | 0.00 | 16 | N | 0.00 |
| 7 | D | 0.00 | 43 | V | 0.00 |

| | | | | | |
|---|---|---|---|---|---|
| 8 | K | 0.00 | 68 | Q | 0.00 |
| 17 | K | 0.00 | 69 | N | 0.00 |
| 26 | G | 0.00 | 76 | G | 0.00 |
| 30 | L | 0.00 | 77 | A | 0.00 |
| 32 | R | 0.00 | 78 | F | 0.00 |
| 38 | P | 0.00 | 79 | T | 0.00 |

The top 10 residues with the lowest scores in Set 1 and Set 2 obtained by unweighted frequency and entropy-based measure are:

| Set 1 | | | Set 2 | | |
|---|---|---|---|---|---|
| 117 | K | -1.85 | 33 | T | -1.889 |
| 113 | T | -1.72 | 64 | Q | -1.831 |
| 70 | L | -1.67 | 147 | A | -1.831 |
| 73 | S | -1.59 | 27 | E | -1.735 |
| 84 | E | -1.41 | 23 | V | -1.677 |
| 133 | G | -1.41 | 36 | H | -1.677 |
| 36 | V | -1.37 | 49 | L | -1.677 |
| 35 | V | -1.30 | 93 | S | -1.677 |
| 9 | A | -1.17 | 145 | R | -1.677 |
| 6 | K | -1.16 | 196 | L | -1.677 |



The Python code to compute the conservation score from MSA using unweighted frequency, and entropy, variance and sum of pairs-based measures is given below:

```
import numpy as np
import pandas as pd

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
}


def compute_conservation_scores(msa):
    n = len(msa)
    residues = ["-", "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M",
"F", "P", "S", "T", "W", "Y", "V"]

    overall_freq = {residue: 0 for residue in residues}

    for i in range(n):
```

```
        for residue in residues:
            overall_freq[residue] += msa[i].count(residue)
    gap_count = overall_freq["-"]
    overall_freq = {residue: overall_freq[residue] / (n * len(msa[0]) - gap_count) for
residue in residues}

    entropy_scores = []
    variance_scores = []
    sum_of_pairs_scores = []

    for i in range(len(msa[0])):
        freq = {residue: 0 for residue in residues}
        count = 0
        for j in range(n):
            if msa[j][i] != "-":
                freq[msa[j][i]] += 1
                count += 1
        freq = {residue: freq[residue] / count for residue in residues}

        entropy_score = 0
        variance_score = 0
        sum_of_pairs_score = 0

        for residue in residues:
            if freq[residue] != 0:
                entropy_score += freq[residue] * np.log(freq[residue])
            variance_score += (freq[residue] - overall_freq[residue]) ** 2

        variance_score = np.sqrt(variance_score)

        for residue1 in residues:
            for residue2 in residues:
                if (residue1, residue2) in blosum62:
                    sum_of_pairs_score += freq[residue1] * freq[residue2] *
blosum62[(residue1, residue2)]
                elif (residue2, residue1) in blosum62:
                    sum_of_pairs_score += freq[residue1] * freq[residue2] *
blosum62[(residue2, residue1)]

        entropy_scores.append(round(entropy_score, 3))
        variance_scores.append(round(variance_score, 3))
        sum_of_pairs_scores.append(round(sum_of_pairs_score, 3))

    df = pd.DataFrame(
        data=([residue for residue in msa[0]], entropy_scores, variance_scores,
sum_of_pairs_scores)).transpose()
    df.columns = ["Residue", "Entropy", "Variance", "Sum of Pairs"]

    return df


msa = [
    "-
SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAHGKKVMGGIALAVSKIDDLKTGLMELSEQ
HAYKLRVDPANFKILNHCILVVISTMFPKEFTPEAHVSLDKFLSGVALALAERYR",
    "MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-
DLHHGSAQVKAHGKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALTPEIHASLDKFLCAVG
NVLTSKYR",
    "-VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-
DLSPGSAQVKAHGKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVS
TVLTSKYR",
    "MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-
DVSPGSAQVKAHGKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVS
TVLTSKYR",
    "-VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-
NLSHGSDQVKAHGQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVS
TVLTSKYR",
```

```
    "MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-
DVSHGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKDHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTPAVHASLDKFLANVS
TVLTSKYR",
    "MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVS
TVLTSKYR",
    "MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSTVS
TVLTSKYR"]

result = compute_conservation_scores(msa)
print(result)
```

The output is shown below:

```
Run - Anirudh-Python
Run:   workboard
C:\Users\anira\PycharmProjects\Anirudh-Python\ve
    Residue Entropy Variance Sum of Pairs
0         -     0.0    1.021          5.0
1         S  -0.305    0.865        3.008
2         L     0.0    0.909          4.0
3         S     0.0    0.953          4.0
4         D  -1.121    0.527        1.149
..      ...     ...      ...          ...
138       A  -0.305    0.881        4.165
139       E  -0.305    0.868        3.347
140       R  -0.305    0.871        4.504
141       Y     0.0    1.016          7.0
142       R     0.0    1.015          5.0

[143 rows x 4 columns]

Process finished with exit code 0
```

This matches the results obtained in Question 1 using AL2CO.

# 4

Multiple sequence alignments for Set 1 from Question 1 were obtained using Clustal Omega, MAFFT, and MUSCLE. These were used as input to the previous program to obtain the conservation score at each residue for each of these alignments.

The Python code used for this is shown below:

```
clustal = [
        "-
SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAHGKKVMGGIALAVSKIDDLKTGLMELSEQ
HAYKLRVDPANFKILNHCILVVISTMFPKEFTPEAHVSLDKFLSGVALALAERYR",
        "MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-
DLHHGSAQVKAHGKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALTPEIHASLDKFLCAVG
NVLTSKYR",
```

```
        "-VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-
DLSPGSAQVKAHGKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVS
TVLTSKYR",
        "MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-
DVSPGSAQVKAHGKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVS
TVLTSKYR",
        "-VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-
NLSHGSDQVKAHGQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVS
TVLTSKYR",
        "MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-
DVSHGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTPAVHASLDKFLASVS
TVLTSKYR",
        "MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKDHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
        "MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
        "MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTPAVHASLDKFLANVS
TVLTSKYR",
        "MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVS
TVLTSKYR",
        "MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSTVS
TVLTSKYR"]

mafft = [
         "MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-
DVSHGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTPAVHASLDKFLASVS
TVLTSKYR",
         "MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVS
TVLTSKYR",
         "MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSTVS
TVLTSKYR",
         "MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTPAVHASLDKFLANVS
TVLTSKYR",
         "MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKDHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
         "MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
         "-VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-
NLSHGSDQVKAHGQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVS
TVLTSKYR",
         "MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-
DVSPGSAQVKAHGKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVS
TVLTSKYR",
         "-VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-
DLSPGSAQVKAHGKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVS
TVLTSKYR",
         "MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-
DLHHGSAQVKAHGKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALTPEIHASLDKFLCAVG
NVLTSKYR",
         "-
SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAHGKKVMGGIALAVSKIDDLKTGLMELSEQ
HAYKLRVDPANFKILNHCILVVISTMFPKEFTPEAHVSLDKFLSGVALALAERYR"]

muscle = [
    "-
SLSDKDKAAVRALWSKIGKSADAIGNDALSRMIVVYPQTKTYFSHWPDVTPGSPHIKAHGKKVMGGIALAVSKIDDLKTGLMELSEQ
HAYKLRVDPANFKILNHCILVVISTMFPKEFTPEAHVSLDKFLSGVALALAERYR",
```

```
    "MVLSANDKSNVKSIFSKISSHAEEYGAETLERMFTTYPQTKTYFPHF-
DLHHGSAQVKAHGKKVAAALIEAANHIDDIAGALSKLSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALTPEIHASLDKFLCAVG
NVLTSKYR",
    "-VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-
DLSPGSAQVKAHGKKVADALTTAVAHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVS
TVLTSKYR",
    "MVLSADDKTNIKNCWGKIGGHGGEYGEEALQRMFAAFPTTKTYFSHI-
DVSPGSAQVKAHGKKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAMHASLDKFLASVS
TVLTSKYR",
    "-VLSAADKANVKAAWGKVGGQAGAHGAEALERMFLGFPTTKTYFPHF-
NLSHGSDQVKAHGQKVADALTKAVGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVS
TVLTSKYR",
    "MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHF-
DVSHGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSPADKTNVKTAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKDHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVS
TVLTSKYR",
    "MVLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-
DLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTPAVHASLDKFLANVS
TVLTSKYR",
    "MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVS
TVLTSKYR",
    "MVLSAADKTNVKAAWSKVGGNAGEFGAEALERMFLGFPTTKTYFPHF-
DLSHGSAQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSTVS
TVLTSKYR"]

df =
pd.DataFrame([compute_conservation_scores(clustal)["Entropy"],compute_conservation_scor
es(mafft)["Entropy"],compute_conservation_scores(muscle)["Entropy"]]).transpose()
df.columns = ["Clustal Omega","MAFFT","MUSCLE"]
print(df)

print("\nThe correlation between the conservation scores of the different alignments
are:\n")
print(df.corr())
```

The results obtained from this is shown below:

```
Run - Anirudh-Python
Run:   workboard
C:\Users\anira\PycharmProjects\Anirudh-Python\venv\Scripts\python.exe C:/Users/anira/F
     Clustal Omega  MAFFT  MUSCLE
0            0.000  0.000   0.000
1           -0.305 -0.305  -0.305
2            0.000  0.000   0.000
3            0.000  0.000   0.000
4           -1.121 -1.121  -1.121
..             ...    ...     ...
138         -0.305 -0.305  -0.305
139         -0.305 -0.305  -0.305
140         -0.305 -0.305  -0.305
141          0.000  0.000   0.000
142          0.000  0.000   0.000

[143 rows x 3 columns]

The correlation between the conservation scores of the different alignments are:

               Clustal Omega  MAFFT  MUSCLE
Clustal Omega            1.0    1.0     1.0
MAFFT                    1.0    1.0     1.0
MUSCLE                   1.0    1.0     1.0

Process finished with exit code 0
```

From this, we can say that the conservation scores obtained from the three alignment methods are identical.

## 5

The manual verification of the unweighted frequency entropy scores at the given positions is shown below:

$\boxed{\text{Pos. 9}}$ = A, S, T, T, A, S, T, T, G, T, T

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

Entropy score $= \sum_i f(i) \ln(f(i))$

$= 0.182(-1.704) + 0.182(-1.704) + 0.545(-0.607) + 0.091(-2.397)$

$= -1.169 \Rightarrow$ verified

$\boxed{\text{Pos. 11}}$ = V, V, I, I, V, I, V, V, V, V, V

$f(V) = \frac{8}{11} = 0.727$, $f(I) = \frac{3}{11} = 0.273$

Entropy score $= 0.727(-0.319) + 0.273(-1.298)$

$= -0.586 \Rightarrow$ verified

$\boxed{\text{Pos. 20}}$ = K, S, G, G, G, G, A, A, G, G, G

$f(K) = \frac{1}{11} = 0.091$, $f(S) = \frac{1}{11} = 0.091$, $f(G) = \frac{7}{11} = 0.636$, $f(A) = \frac{2}{11} = 0.182$

Entropy score $= 0.091(-2.397) + 0.091(-2.397) + 0.636(-0.453) + 0.182(-1.704)$

$= -1.034 \Rightarrow$ verified

$\boxed{\text{Pos. 22}}$ = A, A, A, G, A, G, A, A, A, A, A

$f(A) = \frac{9}{11} = 0.818$, $f(G) = \frac{2}{11} = 0.182$

Entropy score $= 0.818(-0.201) + 0.182(-1.704)$

$= -0.474 \Rightarrow$ verified

$\boxed{\text{Pos. 30}}$ = L, L, L, L, L, L, L, L, L, L, L

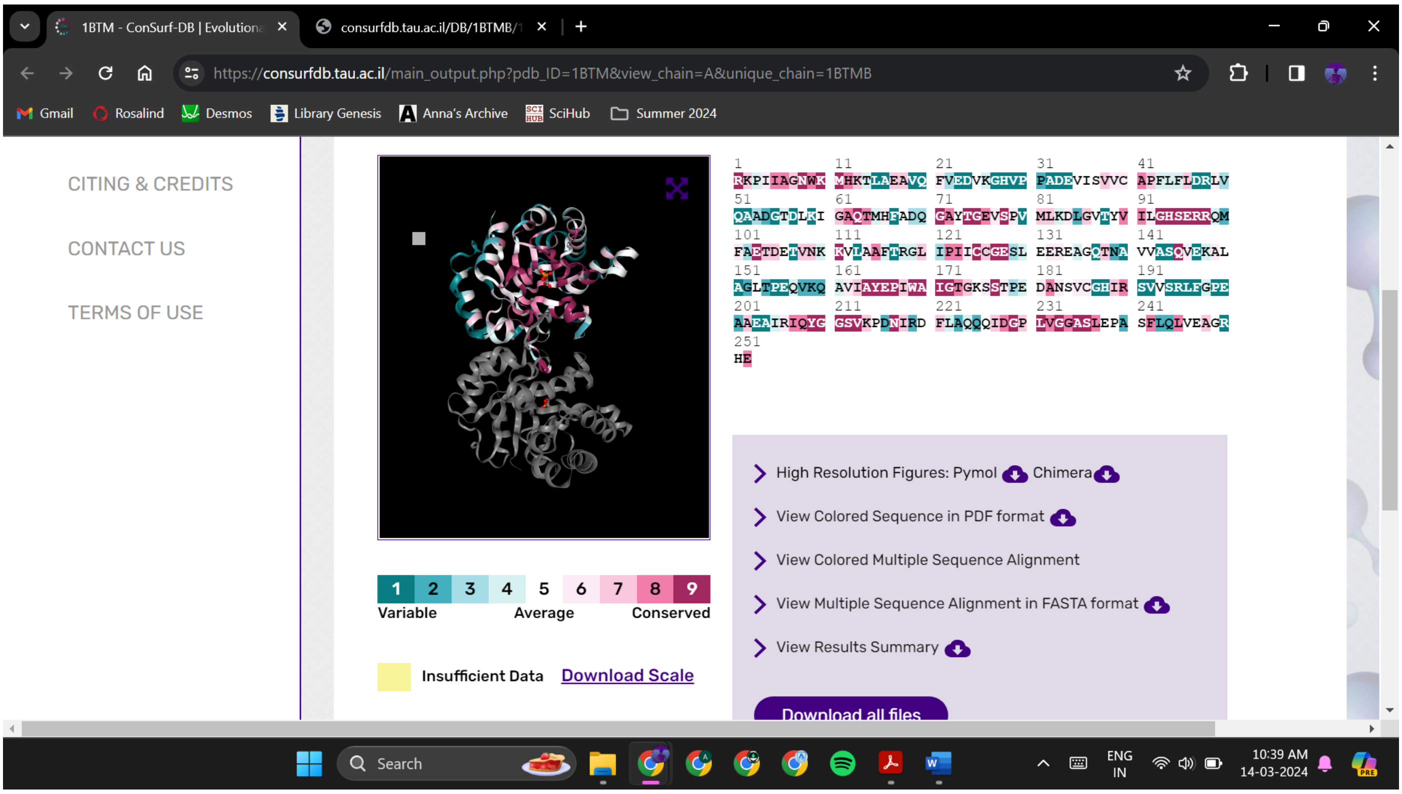
$f(L) = \frac{11}{11} = 1$

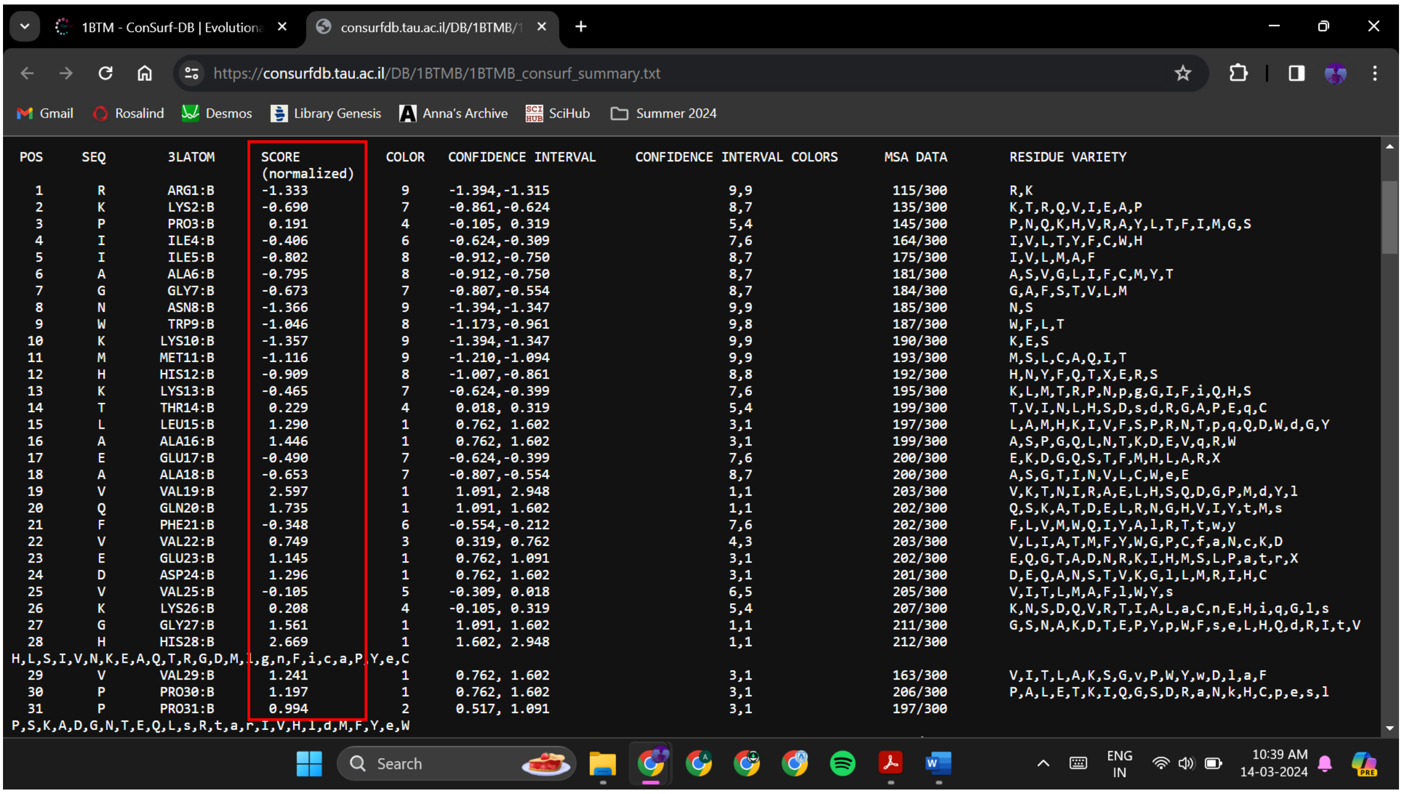
Entropy score $= 1(0)$

$= 0 \Rightarrow$ verified

## 6

The conservation scores for 1BTM Chain A were obtained used ConSurf. The results are shown below:




| POS | SEQ | 3LATOM | SCORE (normalized) | COLOR | CONFIDENCE INTERVAL | CONFIDENCE INTERVAL COLORS | MSA DATA | RESIDUE VARIETY |
|---|---|---|---|---|---|---|---|---|
| 1 | R | ARG1:B | -1.333 | 9 | -1.394,-1.315 | 9,9 | 115/300 | R,K |
| 2 | K | LYS2:B | -0.690 | 7 | -0.861,-0.624 | 8,7 | 135/300 | K,T,R,Q,V,I,E,A,P |
| 3 | P | PRO3:B | 0.191 | 4 | -0.105, 0.319 | 5,4 | 145/300 | P,N,Q,K,H,V,R,A,Y,L,T,F,I,M,G,S |
| 4 | I | ILE4:B | -0.406 | 6 | -0.624,-0.309 | 7,6 | 164/300 | I,V,L,T,Y,F,C,W,H |
| 5 | I | ILE5:B | -0.802 | 8 | -0.912,-0.750 | 8,7 | 175/300 | I,V,L,M,A,F |
| 6 | A | ALA6:B | -0.795 | 8 | -0.912,-0.750 | 8,7 | 181/300 | A,S,V,G,L,I,F,C,M,Y,T |
| 7 | G | GLY7:B | -0.673 | 7 | -0.807,-0.554 | 8,7 | 184/300 | G,A,F,S,T,V,L,M |
| 8 | N | ASN8:B | -1.366 | 9 | -1.394,-1.347 | 9,9 | 185/300 | N,S |
| 9 | W | TRP9:B | -1.046 | 8 | -1.173,-0.961 | 9,8 | 187/300 | W,F,L,T |
| 10 | K | LYS10:B | -1.357 | 9 | -1.394,-1.347 | 9,9 | 190/300 | K,E,S |
| 11 | M | MET11:B | -1.116 | 9 | -1.210,-1.094 | 9,9 | 193/300 | M,S,L,C,A,Q,I,T |
| 12 | H | HIS12:B | -0.909 | 8 | -1.007,-0.861 | 8,8 | 192/300 | H,N,Y,F,Q,T,X,E,R,S |
| 13 | K | LYS13:B | -0.465 | 7 | -0.624,-0.399 | 7,6 | 195/300 | K,L,M,T,R,P,N,p,g,G,I,F,i,Q,H,S |
| 14 | T | THR14:B | 0.229 | 4 | 0.018, 0.319 | 5,4 | 199/300 | T,V,I,N,L,H,S,D,s,d,R,G,A,P,E,q,C |
| 15 | L | LEU15:B | 1.290 | 1 | 0.762, 1.602 | 3,1 | 197/300 | L,A,M,H,K,I,V,F,S,P,R,N,T,p,q,Q,D,W,d,G,Y |
| 16 | A | ALA16:B | 1.446 | 1 | 0.762, 1.602 | 3,1 | 199/300 | A,S,P,G,Q,L,N,T,K,D,E,V,q,R,W |
| 17 | E | GLU17:B | -0.490 | 7 | -0.624,-0.399 | 7,6 | 200/300 | E,K,D,G,Q,S,T,F,M,H,L,A,R,X |
| 18 | A | ALA18:B | -0.653 | 7 | -0.807,-0.554 | 8,7 | 200/300 | A,S,G,T,I,N,V,L,C,W,e,E |
| 19 | V | VAL19:B | 2.597 | 1 | 1.091, 2.948 | 1,1 | 203/300 | V,K,T,N,I,R,A,E,L,H,S,Q,D,G,P,M,d,Y,l |
| 20 | Q | GLN20:B | 1.735 | 1 | 1.091, 1.602 | 1,1 | 202/300 | Q,S,K,A,T,D,E,L,R,N,G,H,V,I,Y,t,M,s |
| 21 | F | PHE21:B | -0.348 | 6 | -0.554,-0.212 | 7,6 | 202/300 | F,L,V,M,W,Q,I,Y,A,l,R,T,t,w,y |
| 22 | V | VAL22:B | 0.749 | 3 | 0.319, 0.762 | 4,3 | 203/300 | V,L,I,A,T,M,F,Y,W,G,P,C,f,a,N,c,K,D |
| 23 | E | GLU23:B | 1.145 | 1 | 0.762, 1.091 | 3,1 | 202/300 | E,Q,G,T,A,D,N,R,K,I,H,M,S,L,P,a,t,r,X |
| 24 | D | ASP24:B | 1.296 | 1 | 0.762, 1.602 | 3,1 | 201/300 | D,E,Q,A,N,S,T,V,K,G,l,L,M,R,I,H,C |
| 25 | V | VAL25:B | -0.105 | 5 | -0.309, 0.018 | 6,5 | 205/300 | V,I,T,L,M,A,F,l,W,Y,s |
| 26 | K | LYS26:B | 0.208 | 4 | -0.105, 0.319 | 5,4 | 207/300 | K,N,S,D,Q,V,R,T,I,A,L,a,C,n,E,H,i,q,G,l,s |
| 27 | G | GLY27:B | 1.561 | 1 | 1.091, 1.602 | 1,1 | 211/300 | G,S,N,A,K,D,T,E,P,Y,p,W,F,s,e,L,H,Q,d,R,I,t,V |
| 28 | H | HIS28:B | 2.669 | 1 | 1.602, 2.948 | 1,1 | 212/300 | H,L,S,I,V,N,K,E,A,Q,T,R,G,D,M,l,g,n,F,i,c,a,P,Y,e,C |
| 29 | V | VAL29:B | 1.241 | 1 | 0.762, 1.602 | 3,1 | 163/300 | V,I,T,L,A,K,S,G,v,P,W,Y,w,D,l,a,F |
| 30 | P | PRO30:B | 1.197 | 1 | 0.762, 1.602 | 3,1 | 206/300 | P,A,L,E,T,K,I,Q,G,S,D,R,a,N,k,H,C,p,e,s,l |
| 31 | P | PRO31:B | 0.994 | 2 | 0.517, 1.091 | 3,1 | 197/300 | P,S,K,A,D,G,N,T,E,Q,L,s,R,t,a,r,I,V,H,l,d,M,F,Y,e,W |