Analysis of Cancer Proteins: Pattern Identification on *Escherichia coli* Species

Abstract

A major challenge of protein sequence analysis is efficient and accurate detection of pattern and similarities that allow functional prediction. We analyzed the different species of protein sequences for getting a pattern. In this work, we have analyzed to find pattern for cancer disease from *Escherichia coli* protein sequences. The *Escherichia coli* protein sequences have been compared with different species to find the pattern. This work may be used to find and design the accurate drug for cancer disease on human species.

Introduction


Results and Discussion

Analysis of individual amino acid contents: (Figure 1)

We analyzed the amino acids of different species for Cancer disease (Table 1). From the 12 species, it is found that the *E. coli* has secured the alanine at highest level (7.87%). The lowest alanine has secured at (1.37%). The *E. coli* species has secured at 6.36% in human species. In cysteine amino acid, the highest score secured is 2.23 in rat species. The least score for cysteine is at (1.37%). The *E. coli* species has secured at 5.26% in human species. In other amino acids, *E. coli* species has secured the highest score.
The highest percentage of aspartic acid (6.59%). The least aspartic acid is secured at (5.43%) in rat species. The rat species has secured at high percentage of glutamic amino acid (7.36%). The E.coli secured lowest glutamic amino acid (5.49%). The highest percentage of phenylalanine amino acid secured in horse species (4.23%). The least percentage of phenylalanine amino acid is secured in rat species (3.58%). The highest percentage of glycine amino acid is found in horse species (7.60%). The lowest percentage of glycine amino acid is found in rat species (6.37%).

The highest percentage of histidine amino acid is found in E.coli species (3.73%). The lowest percentage of histidine amino acid is identified in cow species (2.46%). The highest percentage of isoleucine amino acid is detected in E.coli species (6.71%). The lowest percentage of isoleucine amino acid is found in rat species (4.41%). The highest percentage of lysine amino acid is identified in human species (5.37%). The lowest percentage of lysine amino acid is detected in E.coli species (3.97%).

The highest percentage of leucine amino acid is found in rat species (10.06%). The lowest percentage of leucine amino acid is found in horse species (9.12%). The highest percentage of methionine amino acid is detected in E.coli species (2.81%). The lowest percentage of methionine amino acid is found in rat species (1.6%).

The highest percentage of asparagine amino acid is identified in E.coli species (4.5%). The lowest percentage of asparagine amino acid is detected in whale species (3.63%). The highest percentage of proline amino acid is found in rat species (5.55%). The lowest percentage of proline amino acid is identified in whale species (4.78%). The highest percentage of glutamine amino acid is found in horse species (3.88%). The lowest percentage of glutamine amino acid is detected in E.coli and rat species (5.06%).

Name of the Amino Acid calculation for different species in Cancer Disease.

| Name of the Species | A   | C   | D   | E   | F   | G   | H   | I   | K   | L   | M   | N   | P   | Q   | R   | S   | T   | V   | W   | Y   |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| chimpanzee          | 6.95| 1.82| 5.63| 6.5 | 3.88| 7.07| 2.69| 5.11| 5.84| 9.26| 2.2 | 4.07| 5.07| 4.41| 5.08| 6.55| 5.68| 7   | 1.5 | 3.67|
| cow                 | 6.73| 1.67| 5.78| 6.82| 3.85| 7.08| 2.46| 5.37| 6.41| 9.16| 2.1 | 4.07| 5.06| 4.18| 4.83| 6.23| 5.82| 7.03| 1.6 | 3.77|
| donkey              | 6.98| 1.87| 5.68| 6.54| 3.88| 7.12| 2.74| 5.12| 5.9  | 9.24| 2.16| 3.95| 5.11| 4.25| 5.08| 6.63| 5.61| 7.02| 1.44| 3.67|
| duck                | 7.01| 1.78| 5.66| 6.59| 3.88| 7.05| 2.69| 5.36| 5.96| 9.35| 2.25| 4.02| 4.9 | 4.12| 5.08| 6.95| 5.68| 6.81| 1.41| 3.45|
| ecoli               | 7.87| 1.37| 6.59| 5.49| 3.65| 7.38| 3.73| 6.71| 3.97| 9.85| 2.81| 4.5 | 5.05| 5.06| 4.69| 4.22| 5.91| 6.67| 1.65| 2.84|
| elephant            | 7.03| 1.87| 5.66| 6.56| 3.89| 7.13| 2.74| 5.13| 5.93| 9.22| 2.17| 3.95| 5.1 | 4.24| 5.05| 6.61| 5.61| 7.02| 1.44| 3.66|
| fish                | 7.39| 1.71| 5.75| 6.47| 4.04| 7.18| 2.84| 5.32| 5.98| 9.22| 2.14| 4.22| 4.99| 4.09| 4.9 | 6.41| 5.47| 6.9 | 1.38| 3.61|
| horse               | 7.37| 2.04| 5.6 | 6.18| 4.23| 7.6 | 2.68| 5.08| 6.19| 9.12| 2.04| 3.98| 4.96| 3.88| 4.73| 6.75| 5.78| 7.32| 1.28| 3.2 |
| human               | 6.36| 2   | 5.49| 6.98| 4.02| 6.78| 2.88| 4.96| 5.93| 9.43| 2.14| 3.69| 5.25| 4.46| 5.37| 6.92| 5.29| 6.67| 1.56| 3.82|
| musculus            | 6.97| 1.87| 5.69| 6.55| 3.88| 7.12| 2.73| 5.12| 5.9  | 9.24| 2.17| 3.95| 5.11| 4.25| 5.09| 6.63| 5.62| 7.01| 1.44| 3.67|
| rat                 | 6.86| 2.23| 5.43| 7.36| 3.58| 6.37| 2.84| 4.41| 5.83|10.06| 1.6 | 3.78| 5.55| 5.06| 4.63| 6.9 | 4.43| 6.85| 1.88| 4.34|
| whale               | 7.53| 1.66| 5.48| 6.84| 3.9 | 7.16| 3.33| 5.21| 6.71| 9.46| 2.1 | 3.63| 4.78| 4.11| 4.76| 6.35| 5.34| 6.75| 1.46| 3.46|

Figure 1: Analysis individual amino acid contents in different species. (X-axis: Name of the amino acid & Y-axis: percentage value).
amino acid is found in duck species (6.95%). The lowest percentage of serine amino acid is identified in \textit{E.coli} species (4.22%). The highest percentage of threonine amino acid is detected in \textit{E.coli} species (7.32%). The lowest percentage of valine amino acid is found in rat species (4.43%). The highest percentage of tryptophan amino acid is found in rat species (1.88%). The lowest percentage of tryptophan amino acid is detected in horse species (1.28%). The highest percentage of tyrosine amino acid is found in \textit{E.coli} species (2.84%).

Analysis of group of amino acid contents (Figure 2 and Table 2)

The highest percentage of hydrophobic group amino acid species is identified in \textit{E.coli} (23.23%). The lowest percentage of hydrophobic group amino acid species is detected in human (21.06%). The highest percentage of hydrophobic group amino acid species is found in \textit{E.coli} (19.38%). The lowest percentage of hydrophobic group amino acid species is identified in rat (17.59%). The highest percentage of polar uncharged group amino acid species is detected in horse (27.99%). The lowest percentage of polar uncharged group amino acid species is found in rat (26.54%). The highest percentage of polar charged group amino acid species is identified in rat (11.29%). The lowest percentage of polar charged group amino acid species is detected in horse (9.2%). The highest percentage of charged positive group amino acid species is found in whale (11.47%). The lowest percentage of charged positive group amino acid species is identified in \textit{E.coli} (8.66%). The \textit{E.coli} species only has the least score of 8.66% when compared to other species. The other species has secured score greater than 10.46%. Therefore, the \textit{E.coli} has the different pattern when compared with the other species. The highest percentage of charged negative group amino acid species is identified in rat (12.79%). The lowest percentage of charged negative group amino acid species is detected in horse (11.78%).

Analysis of \textit{E.coli} pattern from individual amino acid contents

The \textit{E.coli} species maintains the alanine and aspartic amino acid at 7.87% & 6.59% respectively. The other species has secured alanine and aspartic acid amino acid less than 7.56% & 5.78% respectively. The \textit{E.coli} species maintains the glutamic amino acid at 5.49%. The other species has secured glutamic amino acid greater than 6.18%. The \textit{E.coli} other amino acid like histidine, isoleucine, serine and threonine also secures either high or low percentage score when compared with other species. Therefore, it is found that the \textit{E.coli} species is identified with the different pattern compared with the other species individual amino acid contents.

Analysis of \textit{E.coli} pattern from group of amino acid contents:

The \textit{E.coli} species maintains the highest hydrophobic strong and hydrophobic strong.
weak amino acid group at 23.23% & 19.38% respectively when compared with other species. The other species has secured highest hydrophobic strong and hydrophobic weak amino acid group at less than 21.56% & 18.56% respectively. The *E.coli* species also secured the least score of 8.66% when compared with other species for charged positive amino acid group. The other species has secured greater than 10.46% for charged positive amino acid group. Therefore, it is found that the *E.coli* species is identified with the different pattern when compared with the other species.

**Conclusion**

We have analyzed different species of protein sequences. We identified a pattern on *Escherichia coli* protein sequences when compared with 12 different species. The *Escherichia coli* protein sequence only is getting a different pattern on analysis of individual amino acid contents and group of amino acid contents also. These results may be used to design a new accurate drug for cancer disease on human species.
References


