STRUCTURE BASED PROPERTIES (Features) – Propensities

1. Amino acid compositions in high B-value regions

Normalize B-values: $\qquad B_{norm} = \dfrac{B - B_{mean}}{B_{\sigma}}$

Amino acids with B-values greater than $B_{mean} + 0.5\ B_{\sigma}$ are high B-value residues.

Output: amino acid % $= \dfrac{Number\ of\ particular\ amino\ acid\ residue * 100}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

Ex: amino acid % $= \dfrac{No.of\ ALA\ in\ high\ B-value\ regions\ * 100}{Total\ No\ of\ ALA\ in\ the\ protein}$

Ref: Parthasarathy, S., & Murthy, M. R. N. (2000). Protein thermal stability: insights from atomic displacement parameters (B values). *Protein engineering*, **13**, 9-13.

2. Frequency of occurrence in β-bends

β-bends are identified from DSSP output.

Amino acid frequency $= \dfrac{Number\ of\ particular\ amino\ acid\ residue}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

Ref : Lewis, P. N., Momany, F. A., & Scheraga, H. A. (1971). Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proceedings of the National Academy of Sciences*, **68**, 2293-2297.

3. Normalized frequency of turn

Turns are identified from DSSP output.

Amino acid frequency $= \dfrac{Number\ of\ particular\ amino\ acid\ residue}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

Ref : Crawford, J. L., Lipscomb, W. N., & Schellman, C. G. (1973). The reverse turn as a polypeptide conformation in globular proteins. *Proceedings of the National Academy of Sciences*, **70**, 538-542.

4. Normalized frequency of N-Helix (Crawford et al., 1973)

Frequency of N-helix $= \dfrac{Number\ of\ particular\ amino\ acid\ residue\ in\ the\ N-helix\ region\ * 100}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

5. Normalized frequency of C-Helix (Crawford et al., 1973)

Frequency of C-helix $= \dfrac{Number\ of\ particular\ amino\ acid\ residue\ in\ the\ C-helix\ region\ * 100}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

6.   Normalized frequency of middle helix (Crawford et al., 1973)

Frequency of middle-helix =

$$\frac{Number\ of\ particular\ amino\ acid\ residue\ in\ the\ middle-helix\ region * 100}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$$

7.   Normalized frequency of β-sheet (Crawford et al., 1973)

Frequency of β sheet $= \dfrac{Number\ of\ particular\ amino\ acid\ residue\ in\ the\ beta\ sheet\ region * 100}{total\ number\ of\ particular\ amino\ acid\ residue\ in\ the\ protein}$

8. Propensity to form MCI for two state proteins

$P_{mc}(i) = f_{mc}(i) / f_t(i)$

$f_{mc}(i)$ = frequency of occurrence of amino acids that form multiple contacts

$f_t(i)$ = frequency of residues in the whole protein

Ref: Gromiha, M.M. *Protein bioinformatics: from sequence to function*. Academic Press, 2010.

9. Propensity to form MCI for three state proteins

$P_{mc}(i) = f_{mc}(i) / f_t(i)$

$f_{mc}(i)$ = frequency of occurrence of amino acids that form multiple contacts

$f_t(i)$ = frequency of residues in the whole protein

Ref: Gromiha, M.M. *Protein bioinformatics: from sequence to function*. Academic Press, 2010.

10. P_α, α-helical tendency;

% of residue in α-*helix* $= \dfrac{no.of\ particular\ residue\ in\ \alpha-helix}{total\ no.of\ particular\ residue\ in\ whole\ protein}$

    Ref: Gromiha, M.M. *Protein bioinformatics: from sequence to function*. Academic Press, 2010.

11. P_β β -sheet tendency

% of residue in β sheet $= \dfrac{no.of\ particular\ residue\ in\ \beta-sheet}{total\ no.of\ particular\ residue\ in\ whole\ protein}$

    Ref: Gromiha, M.M. *Protein bioinformatics: from sequence to function*. Academic Press, 2010.