STRUCTURE BASED PROPERTIES - Physiochemical Properties

1. Center of mass

$$x_{COM} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i}$$

where x_i is the X- Cartesian coordinate of the atom i

m_i = atomic mass

Similarly, the y and z coordinates of the center of mass can be calculated.

Ref: Costantini, S., Paladino, A., & Facchiano, A. M. (2008). CALCOM: A software for calculating the center of mass of proteins. *Bioinformation*, **2** , 271-272.

2. Disulphide interactions :

Calculate distance : $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2)}$

Pairs of cysteines within 2.2 Å are considered as disulphide bridges.

Ref: Tina, K. G., Bhadra R, and Srinivasan N.(2007) PIC: protein interactions calculator. *Nucleic Acids Research* **35**, 473-476.

3. Ionic Interactions:

Calculate distance : $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2)}$

Ionic residue pairs(R,K,H) : (D,E) falling within a distance of 6Å contribute to ionic interactions

Ref: Tina, K. G., Bhadra R, and Srinivasan N.(2007) PIC: protein interactions calculator. *Nucleic Acids Research* **35**, 473-476.

4. Hydrophobic interactions

Calculate distance : $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2)}$

CB residues of A,V,L,I,M,F,W,P,Y show hydrophobic interactions when they fall within 5Å range.

Ref: Namdeo, R., Thakur, P., Sachan, R., Karatti, A., Wariar, M., Sharma, M., ... & Kusmakar, S. (2011). PIPE: Protein Interaction and Properties Explorer.*Journal of Natural Science, Biology and Medicine*, *2*(3), 88.

5. Hydrogen bond interactions

It is classified into

i. Main chain –Main chain interactions
ii. Main chain – side chain interactions
iii. Side chain – side chain interactions

Ref: McDonald, I. K., Naylor, D. N., Jones, D. T., & Thornton, J. M. (1993). HBPLUS computer program. *Department of Biochemistry and Molecular Biology, University College, London, UK*.

6. Aromatic-Aromatic interactions

Pairs of phenyl ring centroids that are seperated by a distance of 4.5 to 7 Å contribute to aromatic – aromatic interactions.

Ref: Tina, K. G., Bhadra R, and Srinivasan N.(2007) PIC: protein interactions calculator. *Nucleic Acids Research* **35**, 473-476.

7. Aromatic-sulphur interactions

Sulphur atoms of C/M and the aromatic rings of F,Y,W falling within a distance of 5.3 Å accounts for aromatic - sulphur interactions.

Ref: Tina, K. G., Bhadra R, and Srinivasan N.(2007) PIC: protein interactions calculator. *Nucleic Acids Research* **35**, 473-476.

8. Cation- π interactions

When a cationic side chain (K,R) is near an aromatic side chain (F,Y,W) within 10 Å separation they account for cation-pi interactions

Ref: Tina, K. G., Bhadra R, and Srinivasan N.(2007) PIC: protein interactions calculator. *Nucleic acids research* **35**, 473-476.

9. Radius of gyration ROG

$$(\mathrm{ROG}) = \sqrt{\frac{\sum_i \boldsymbol{m_i}\,|\boldsymbol{x_i} - \boldsymbol{CM}|^2}{\sum_i \boldsymbol{m_i}}}$$

CM = center of mass of protein

m_i = mass of each atom; x_i = atomic coordinate

Ref : Banerji, A., & Ghosh, I. (2009). Revisiting the myths of protein interior: studying proteins with mass-fractal hydrophobicity-fractal and polarizability-fractal dimensions. *PloS one*, **4**, 7361.

10. Accessible surface area for the native protein

ASA for each residue is obtained from DSSP output.

Total ASA = ∑ ASA of each residue

Ref : Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

11. Free energy due to Disulphide interactions

$G_{ss}= 2.3\ N_{ss}$

N_{ss} is the no. of disulphide bonds in the protein

Ref : Ponnuswamy, P. K., & Gromiha, M. M. (1994). On the conformational stability of folded proteins. *Journal of theoretical biology*, *166*(1), 63-74.

12. Surrounding hydrophobicity

$H_p\ (i) = \sum_{j=1}^{20} n_{ij} * h_j$

where n_{ij} is the total number of surrounding residues of type j around i^{th} residue of the protein ;

h_j = hydrophobicity index (kcal/mol) given by Tanford and Jones(1971)

ALA	ASP	CYS	GLU	PHE	GLY	HIS	ILE	LYS	LEU
0.87	0.66	1.52	0.67	2.87	0.1	0.87	3.15	1.64	2.17
MET	ASN	PRO	GLN	ARG	SER	THR	VAL	TRP	TYR
1.67	0.09	2.77	0	0.85	0.07	0.07	1.87	3.77	2.67

The same formula can be used to calculate surrounding hydrophobicity in α-helix/β-sheet/turn regions.

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

13. Surrounding hydrophobicity in unfolded state of jth residue is given by

$$H_j^u = \sum_{\substack{k=j-2 \\ k \neq j}}^{k=j+2} h_k$$

14. Average gain in surrounding hydrophobicity or hydrophobic enrichment

$$H_j = H_j^f - H_j^u$$

15. Average gain ratio in surrounding hydrophobicity

$$G_j = H_j^f / H_j^u$$

16. B_r Buriedness

$$\text{Buriedness} = \frac{\text{no of residues (i) in the interior of protein}}{\text{total number of residues (i)}}$$

$$\text{Example: Buriedness of Alanine} = \frac{\text{no of Ala in the interior of protein}}{\text{total number of Ala}}$$

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

17. Ra Solvent accessible reduction ratio or Mean fractional area loss

$<R_A> = (A^0 - <A>)/ A^0$

A^0 = accessible area in unfolded state of protein

<A> = accessible area in folded state of protein

ASA(Å^2) of residues in the unfolded state

ALA	ASP	CYS	GLU	PHE	GLY	HIS	ILE	LYS	LEU
110.2	144.1	140.4	174.7	200.7	78.7	181.9	185	205.7	183.1
MET	ASN	PRO	GLN	ARG	SER	THR	VAL	TRP	TYR
200.1	146.4	141.9	178.6	229	117.2	138.7	153.7	240.5	213.7

Ref: Gromiha, M.M. *Protein bioinformatics: from sequence to function*. Academic Press, 2010.
Ahmad, S., Gromiha, M., Fawareh, H., & Sarai, A. (2004). ASAView: database and tool for solvent accessibility representation in proteins. *BMC bioinformatics*,*5*, 51.

18. Mean area buried on transfer

Mean area buried on transfer = A^0 - <A>

A^0 = accessible area in unfolded state of protein

<A> = accessible area in folded state of protein

Ref : Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science*, **229**, 834-838.

19. Ns Average number of surrounding residues

Distance between C_α carbon atoms of two residues ≤ 8Å
Count the number of such residues.

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

20. Normalized flexibility parameters (B-values), average

$B_{norm} = \frac{B - B_{mean}}{B_{\sigma}}$

$B_{norm} < 1.0 \rightarrow$ rigid

Ref : Vihinen, M., Torkkila, E., & Riikonen, P. (1994). Accuracy of protein flexibility predictions. *Proteins: Structure, Function, and Bioinformatics*, *19*, 141-149.

21. Normalized flexibility parameters (B-values) for each residue surrounded by none rigid neighbours (Vihinen et al., 1994)

For i^{th} residue, if B_{norm} of i±1 residue > 1.0 then it is has no rigid neighbors.

22. Normalized flexibility parameters (B-values) for each residue surrounded by one rigid neighbours (Vihinen et al., 1994)

For i^{th} residue, if B_{norm} of (i+ 1) or (i-1) residue < 1.0, then it has one rigid neighbor.

23. Normalized flexibility parameters (B-values) for each residue surrounded by two rigid neighbours (Vihinen et al., 1994)

For i^{th} residue, if B_{norm} of any 2 residues around i^{th} residue < 1.0, then it has two rigid neighbors.

24. Surface hydrophobicity

$\Phi_{surface} = \sum \frac{s_{i} * \varphi_{i}}{s_{p}}$

$\Phi_{surface}$ = surface hydrophobicity

s_i = solvent accessible area occupied by amino acid 'i'

φ_i = hydrophobicity value assigned to amino acid

s_p = total solvent accessible area of protein

Normalized census hydrophobicity values :

ALA	ASP	CYS	GLU	PHE	GLY	HIS	ILE	LYS	LEU
0.81	0.88	0.72	0.46	0.95	0.77	0.54	1	0.26	0.92
MET	ASN	PRO	GLN	ARG	SER	THR	VAL	TRP	TYR
0.81	0.45	0.68	0.43	0	0.6	0.63	0.92	0.85	0.71

Ref : Mahn, A., Lienqueo, M. E., & Asenjo, J. A. (2004). Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *Journal of Chromatography A*, **1043**, 47-55.

Gromiha, M. M. (2005). A statistical model for predicting protein folding rates from amino acid sequence with structural class information. *Journal of chemical information and modeling*, *45*, 494-501.

25. ASA of hydrophobic residues

The solvent accessible area of the hydrophobic residues are obtained from DSSP.
Hydrophobic residues : A, V, L, I, M,F,P

Ref : Mahn, A., Lienqueo, M. E., & Asenjo, J. A. (2004). Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *Journal of Chromatography A*, **1043**, 47-55.

26. Hydrophobic free energy

$$G\text{hy}=\sum \Delta\sigma_i\,[A_i\,(\text{folded}) - A_i\,(\text{unfolded})]$$

A_i (folded) = ASA of each atom in folded state
ASA calculated using NACCESS
σ_i atomic salvation parameter

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