10.48047/jocaaa.2024.33.08.352

Computational Methods in Studies of Protein Folding and Analysis

Kirti Patel*
Department of Chemistry, N. B. Mehta (V) Science College, Bordi, Dahanu, Dist. Palghar, Maharashtra, India – 401701

Abstract

In this study, we present the account of various computational methods used in polypeptide modeling and protein folding analysis. The software used for computational analysis is Gromacs accompanied with GROMOS forcefield. The polypeptide folds thus obtained are then clustered using standard procedures. The methods for studying thermodynamic properties of polypeptide folds are reported.

Keywords: Protein Folding, Gromacs, Clustering, Peptide fold energetics, simulated annealing and docking

1. Introduction

Although proteins are astronomical in possibilities of interactions along their sequences, they are restricted to only about 1000 topological possibilities.[1] Main chain define folds due to solvent mediated screening of electrostatics of inter-peptide interactions in conflict with the inter-peptide hydrogen bonding due to homochiral structure.[2,3] Due to the effect of homochirality, α−helix and β−sheet motifs are the building blocks of protein structure, being sequentially selectable options of secondary structure. This limits the scope of morphological possibility in protein tertiary structure to just about 1000 topological forms.[1] Geometric rules of interpeptide interaction under constraint of polypeptide stereochemistry are the defining consideration of protein structure.

Due to poly−L structure, proteins access mainly left hand space of the φ coordinate of Ramachandran diagram in correspondence of α−helix and β−sheet secondary structure. Due to stereospecificity of conformation, residue stereochemistry can be a powerful tool for design of proteins with L and D structure as the alphabet. The principle is observed in Nature in the form of Gramicidin-A, a microbial peptide alternately L,D in structure is a β−helical fold. In the synthetic variants, peptides with even number of alternately L,D residues are ring shaped molecules self stackable as nanotubes.[4]

10.48047/jocaaa.2024.33.08.352

D-amino acids have recently found place in *de novo* protein design. The most common usage of D amino acids is for design of β−turn and inter-helical linkers. Nanda and DeGrado reported putative helices in polypeptide sequences of mixed L,D amino acids.[5] In possibly the first demonstration of design in L and D alphabet, our lab reported a hexapeptide in a complex stereochemical fold.[6] More recently, a family of shape specific heterochiral mini-proteins for chosen sequence plans in L and D structure were reported as bracelet[7], boat[8], canoe[9] and cup[10] shaped molecules.

2. Computational Methods

Molecular dynamics was performed with gromos-96 43A1 force field[11] in GROMACS package in a box of explicit solvent with periodic boundary conditions under NVT (constant number of particles, volume, and temperature), as detailed below. Non-bonded list cutoff was 1.4 nm with shift at 0.8 nm; integration step was 2 fs; initial velocities were drawn from Maxwellian distribution; temperature was coupled to external bath with relaxation time constant 0.1 ps. Bond lengths were constrained with SHAKE[12] to the geometric accuracy 0.0001. For electrically charged system, counter ions were added by replacing the solvent molecule to achieve the electrical neutrality. For each case, first the model peptides energy minimized were placed in a periodic cubic box of appropriate edge length and soaked in either of solvent model like SPC water[13], to the density in correspondence of 1 atm at 298 K. The system was energy minimized first in solvent restraining the solute and then in both solvent and the solute relived of the restraint. Molecular dynamics was initialized and the trajectory was sampled at intervals reported in the individual cases after allowing initial 3 ns for equilibration.

3. Computational Hardware

Molecular dynamics (MD) simulations are carried out on Intel Pentium PIV Linux server, Intel Xeon dual CPU 2.4 GHz computational servers. Certain studies require high computational power and higher memory which necessitates the use of PARAM Padma super computer equipped with 248 1GHz processors and an aggregate memory of 512 GB. MD was carried out using GROMACS package (versions 3.1.4, 3.2.1, 3.3.1, 3.3.2 and 3.3.3). The cluster is configured with two 1.76 GHz AMD® Opteron® processors and 4 GB of memory per node.

4. Investigative Methods

a. Microstate density by conformational clustering

10.48047/jocaaa.2024.33.08.352

Conformational clustering was performed in Cartesian space using van Gunsteren and co-workers[14] procedure. Briefly, a least squares translational-rotational fit on backbone atoms (N, Cβ, Cα, C) between every pair of conformers gave the number of neighbors for every conformer, with root-mean-square deviation (RMSD) ≤ 0.15 nm defining neighbors. The conformer with the largest number of neighbors was defined as central member of the first cluster or the most populous microstate. Removing all members of this microstate from the ensemble, the procedure was iterated until all members in the residual ensemble could be assigned to specific microstates, diminishing in population.

5. Thermodynamic properties of canonical ensemble of microstates

a. Helmholtz Free Energy

Difference Helmholtz free energy (ΔF_{A-B}) between specific microstates was calculated from relative probabilities pA and pB of finding the system in microstate A and B as: $\Delta F_{A-B} = -RT \ln pB/pA$, with R as gas constant, T as temperature, and pA and pB the number of members in microstates A and B.

b. Lennard-Jones Potential

Lennard-Jones potential was calculated in gromacs software with parameters for 12-6 equation drawn from gromos force field. Coulomb calculations were performed with atomic partial charge assignments drawn from gromos-96 and under dielectric constant (ε) one.

c. Coulomb Energy

The Coulomb energy over all peptide groups in the chain gave C_{Tot}, over the peptide groups within a residue and its immediate sequence neighbors, as defined in Figure 1, summed over the entire chain, gave C_{Lo}. Coulomb energy over the peptide groups satisfying the cut-off criteria of distance ≤ 0.35 nm and angle of ≤ 30° for hydrogen bond, summed over the entire chain, gave the Coulomb hydrogen-bond energy (C_{Hb}). C_{Nl} was calculated from the relation $C_{Nl} = C_{Tot} – (C_{Lo} + C_{Hb})$.

d. Hydrophobic Burial Energy

Solvent accessible surface area (ASA) was calculated with NACCESS[15] using probe radius 1.4 Å. Energy of hydrophobic burial was calculated in C_α and C_β atoms as $E_{Bu} = kh (ASA_{PPII} - ASA_{Fold})$[16], with ASA_{PPII} and ASA_{Fold} being solvent accessible surface area in standard-PPII conformer and in the fold of interest. The value for proportionality coefficient k_h was taken as 13.0 kJ mol^{-1} nm^{-2}.

e. Electrostatic Solvation Free Energy

10.48047/jocaaa.2024.33.08.352

Electrostatic solvation free energy ESF, i.e., the solvation energy of transfer from homogeneous medium of dielectric constant of protein to water, was calculated for the central member of microstates using finite-difference solution of Poisson-Boltzmann equation as implemented in DelPhi software[17]. Hydrogens were added using GROMACS software[18]. Internal and external dielectric values were respectively 4.0 and 80.0, and probe radius was 1.4 Å. Grid size of 65 with scale 1.0 was used. The convergence threshold value was set to maxc = 0.0001. The energy values obtained in kT were converted to kJ/mol.

6. Conformational Descriptors

Hydrogen bonds were enumerated to 0.35 nm cutoff in donor-acceptor (N—O) distance and 30^o cutoff in hydrogen-donor-acceptor (H—N—O) angle.

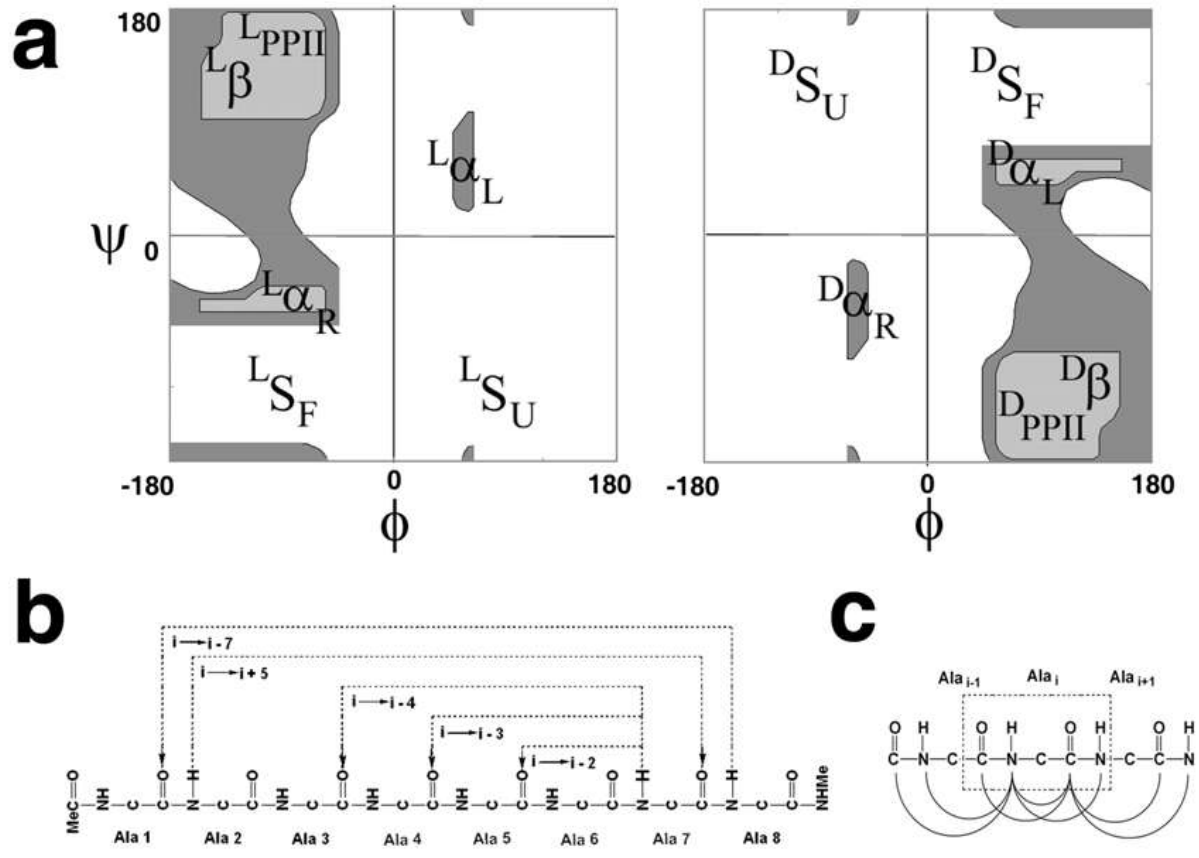

Figure 1 (Panel a) Ramachandran diagrams stereospecific for α-amino acid residues of L- and D- structure, (Panel b) schematic definition of interpeptide hydrogen bonds, and (Panel c) peptide group interactions defining Coulomb-local energy C_{Lo}. Ramchandran diagrams in Panel a are labeled in their α_L and α_R basins, in their β basins inter-nested with PPII sub-basins, and in their sterically favoured (S_F) and unfavoured (S_U) halves in φ.

10.48047/jocaaa.2024.33.08.352

Specific φ,ψ basins were defined in Figure 1 as L/D$_{\beta}$: φL/D = -/+ 90 to -/+ 170, ψL/D = +/- 80 to +/- 180); L/D$_{PPII}$: φL/D = -/+ 30 to -/+ 90, ψL/D = +/- 80 to +/- 170, L/Dα: φL/D = -/+ 20 to -/+ 100, ψL/D = -/+ 20 to -/+ 80; SF: Lφ = 0 to -180, Dφ = 0 to 180, and SU: Lφ = 0 to 180, Dφ = 0 to -180.

7. Simulated Annealing

Mixed-L, D structures in Ac-(Leu)$_{12}$-NHMe were generated randomly using Ribosome. The structures set to +/- 120° in φ and +/- 120° in ψ, as appropriate for a residue based on stereochemistry, were energy minimized in 2000 steps of steepest descent and then vacuum annealed with molecular dynamics by heating at 800 K for 300 ps, cooling in steps to 300 K at the rate of 1 K/ps, and energy minimized in 2000 steps of steepest descent. The annealing was simulated with the force field customized specifically for the purpose[19], while the annealed folds were energy minimized with the standard gromos-96 force field.

8. Ligand Docking

The flexible docking was performed with AutoDock-4.0. The central members top 10 clusters obtained by clustering the three aromatic residues over the molecular dynamics trajectory were chosen for modelling the receptor structure in ligand-peptide complex. Genetic algorithm was used for docking. Using an RMSD tolerance of 2 Å structurally distinct conformational clusters of the ligand were ranked in terms of increasing energy.

9. Sequence Optimization using IDeAS software

The sequence optimization on the heterochiral folds was accomplished using in-house software Inverse Design Algorithm for Sequence program.[20] The program implements LJ, solvation, Coulomb, orientation-dependent hydrogen-bond energies, and an entropy term. The software tests side-chain rotamers drawn from Richardson[21] or Dunbrack[22] libraries for L-residue positions, and suitable symmetry transforms for D-residue positions, using DEE[23] and Monte Carlo search algorithms, towards sequence solutions compatible with an input fold. The sequence optimization was performed at all the positions except at the positions with glycine or proline residues. The residues to be optimized were initially mutated to alanine and the software was allowed to optimize a position with any of the amino acid side chains except glycine, proline, cysteine and methionine amino acid side chains.

10. Conclusion

10.48047/jocaaa.2024.33.08.352

Methods of computational modeling of polypeptide chain are described in the study. The study also describes various methods to analyze the polypeptide folds based on energetics and docking methods. Stereochemical aspects in designing polypeptide fold using simulated annealing are described to facilitate designing of polypeptide folds variable in stereochemistry on a large scale.

References

1. Chothia, C. *Nature* **1992**, *357*, 543.
2. Ramakrishnan, V.; Ranbhor, R.; Kumar, A.; Durani, S. *J. Phys. Chem. B* **2006**, *110*, 9314
3. Ranbhor, R.; Ramakrishnan, V.; Kumar, A.; Durani, S. *Biopolymers* **2006**, *83*, 537.
4. Horne, W. S.; Stout, C. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2003**, *125*, 9372.
5. Nanda, V.; DeGrado, W. F. *J. Am. Chem. Soc.* **2006**, *128*, 809.
6. Fabiola, F.; Pattabhi, V.; Rawale, S.; Raju, E. B.; Durani, S. *Chem. Comm.* **1997**, *15*, 1379.
7. Rana, S., Kundu, B., and Durani, S. *Chem. Commun.* **2004**, 2462.
8. Rana, S., Kundu, B., and Durani, S. *Chem. Commun.* **2005**, 207.
9. Rana, S., Kundu, B., and Durani, S. *Biopolymers,* **2007**, *87(4),* 231.
10. Rana, S., Kundu, B., and Durani, S. *Bioorganic and Medicinal Chemistry* **2006**, *15(11)*, 3874.
11. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Hochschulverlag AG and the ETH Zurich: Zurich, Switzerland, 1996.
12. Ryckaert, J. P.; Ciccoti, G.; Berendsen, H. J. C. *J. Comp Phys* **1997**, *23*, 327.
13. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydration. Intermolecular Forces*, Pullman, B., Reidel Publishing Company, The Netherlands, *331,* 1981.
14. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem. Int. Ed.* **1999**, *38*, 236.
15. Baldwin, R. L. *J. Mol. Biol.* **2007**, *371*, 283.
16. Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biomol. Str.* **1997**, *26*, 425.

10.48047/jocaaa.2024.33.08.352

17. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Compt. Chem.* **2002**, *23(1)*, 128.

18. Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.

19. Ramakrishnan, V.; Ranbhor, R.; Durani, S. *Biopolymers* **2005**, *78*, 96.

20. Ranbhor, R. Computational design and engineering of heterochiral protein structures, Indian Institute of Technology Bombay, **2006**.

21. Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins* **2000**, *40*, 389

22. Dunbrack, R. L.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543.

23. Goldstein, R. F. *Biophys. J.* **1994**, *66*, 1335.