View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by International Journal for Computational Biology (IJCB)

International Journal for Computational Biology (IJCB)
Vol.6, No.1, July 2017, pp. 01-06
ISSN: 2278-8115

Comparative Analysis of Inter residue Contact Energy Potentials with Surrounding Hydrophobicity Model

Konda Mani Saravanan[1], Samuel Selvaraj[2*]
[1] Centre of Advanced Study in Crystallography & Biophysics, University of Madras, Guindy Campus, Chennai – 600 025
[2] Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli – 620 024

Article Info

Article history:

Received Feb 28th, 2017
Revised Apr 20th, 2017
Accepted Apr 21st 2017

Keyword:

Statistical potentials
Inter residue interactions
Surrounding hydrophobicity
Protein folding
Protein stability

ABSTRACT

During the process of protein folding, the regular secondary structures are formed through backbone hydrogen bonding and the side chain interact each other as well as the surrounding medium to create the more complex tertiary structure. Covalent interactions between cysteine groups, non-covalent electrostatic interactions between polar groups and van-der waals interactions between non-polar groups are commonly observed in tertiary structures. To explore the role of various forces contributing to protein stability, models based on inter-residue interactions are an attractive choice. Hence, in the present work, inter residue contact energy statistical potentials are derived and related with the surrounding hydrophobicity model. Also, the statistical potentials derived by various leading research groups are also compared with the classical surrounding hydrophobicity model. Our analysis revealed the importance of hydrophobicity as a dominant force in the protein folding process.

Copyright © 2017 ***International Journal for Computational Biology****,* ***http:// www.ijcb.in****, All rights reserved.*

Corresponding Author:

Dr. Samuel Selvaraj,
Department of Bionformatics,
School of Life Sciences,
Bharathidasan University
Tiruchirappalli, India.
Email: selvarajsamuel@gmail.com

How to Cite:

Saravanan & Selvaraj. Comparative analysis of inter residue contact energy potentials with surrounding hydrophobicity model. IJCB. 2017; Volume 6 (Issue 1): Page 01-06.

1. INTRODUCTION

Inter residue interactions are one of the main focuses to understand protein folding and stability [1-3]. Theoretical models based on inter residue interactions are powerful tools for protein scientists to explore the contribution of different forces to protein stability [4-7]. As early as 1978, Manavalan and Ponnuswamy explored the hydrophobic environment of amino acid residues in globular proteins based on inter residue interactions which is termed as surrounding hydrophobicity model [8]. Later in 1990's, different statistical or empirical potentials based on inter residue interactions have been defined and their ability in discriminating the native structure from non-native ones has been explored for different proteins [9-14]. These statistical potentials use twenty amino acid alphabets and assume the energy contribution for a given residue pair regardless of its local environment. Also, it was observed that the statistical potentials derived by various research groups reflect mainly the hydrophobic interactions and these energy parameters depend upon the environment of the given residue [15]. Since the number of structures in the Protein Data Bank has increased substantially and also our main focus is on the inter residue interactions [16], we wanted to derive inter residue pairwise contact energies from a much larger dataset and to compare it with statistical potentials derived in the past. Since, statistical potentials and surrounding hydrophobicity models are computed based on inter residue contact propensities, it is of interest to relate the statistical potentials derived in the present work with surrounding hydrophobicity. Hence, in the present work, inter residue pairwise contact energy potentials are derived and the surrounding hydrophobicity and inter residue contact energy of proteins in the fold library are related by computing correlation coefficient. Also, the surrounding hydrophobicity was compared with other statistical potentials derived by various leading research groups such as Thomas and Dill [12], Berrera et al., [17] and Skolnick group [18] in different structural classes of proteins such as all-α, all-β, α+β and α/β.

2. RESEARCH METHOD

2.1. Dataset

A fold library has been developed by using a non-redundant dataset of 1334 high resolution x-ray structures by using PISCES server [19]. The proteins in the library share less than 25% sequence identity and within 3Å resolution. By writing suitable program, the C-alpha coordinates for all these proteins have been extracted from their PDB files. The structural class information of proteins in the fold library was obtained from SCOP database [20]. In the fold library, there are 302 all-α protein chains, 314 all-β protein chains, 308 α+β protein chains, 279 α/β protein chains and 131 other classes protein chains such as small proteins, designed proteins and coiled coil proteins etc.,

2.2. Deriving inter residue contact energy potentials

The inter residue contact preferences were computed by neglecting local contacts for 1334 proteins by using residue-residue distance cutoff as 6.5Å which is similar to Miyazawa-Jernigan Procedure [11]. The results were scored in the form of a 20 X 20 matrix. We consider the residue pairs that are separated by at least 4 amino acids in the sequence. By using the pair wise residue preferences, we have computed the energy function [21] by using the formula given below:

$$E_{pair} = -RT \ln \frac{M_{(i,j)}}{Me_{(i,j)}}$$

$$Me_{(i,j)} = \frac{M_{(i)} M_{(j)}}{M}$$

$$M_{(i,j)} = M_{(j,i)}$$

where M(i,j) is the number of pairs between residues of type i and j in the database, Me(i,j) is the estimated number of i-j pairs and M is The total number of pairs respectively.

2.3. Computation of surrounding hydrophobicity

The surrounding hydrophobicity for each residue in a protein molecule is computed following the procedure given by Manavalan and Ponnuswamy [8]. Each residue in a protein is represented by its C-α atom and fixing the first residue as the centre, the distances between this atom and the rest of the C-α atom in the protein molecules are computed. The residues appearing around the reference residue within a sphere of 8Å radius are found out. These residues are assigned with their respective hydrophobic indices [22] and the sum of the hydrophobic indices of those residues that are present within the above sphere is taken to be the 'Surrounding Hydrophobicity' of that particular residue. Thus, the surrounding hydrophobicity of the j^{th} residue of a given protein molecule in its folded form is given by,

$$H_j^f = \sum n_{ij} h_i$$

where n_{ij} is the sums of the surrounding residues of type i around the j^{th} residue of the protein, h_i is the hydrophobic index of the i^{th} residue and f refers to the folded state of the protein molecule respectively.

2.3. Computation of inter residue contact energy

By using the similar procedure used to compute inter residue contact preferences, the residues coming around a central residue are assigned with their interaction energy from the inter residue interaction energy table obtained and their sum represents the interaction energy of the central residue. The sum of interaction energy of all the residues in a protein is computed. Similarly, the sum of inter residue contact energy is calculated by using the statistical potentials derived by Thomas and Dill, Berrera et al., and Skolnick group. Correlation coefficient between surrounding hydrophobicity and inter residue interaction energy (our potential and other three potentials) has been calculated by using the equation below:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

Where X corresponds to surrounding hydrophobicity and Y corresponds to the inter residue contact energy

3. RESULTS AND ANALYSIS

3.1. Inter residue contact energy potentials

The number of inter residue contacts between all the twenty residues is computed and the inter residue contact energies estimated from the contact preference numbers is shown as an interactive grid diagram in figure 1. The residues with strong interactions is shown in blue color and those with weak interactions is shown as red color.The residues such as Ala, Ile, Leu and Val have more number of contacts and hence have minimum inter residue interaction energy. The residue-residue contact preferences of Asp, Gln, Glu and Met are low compared to other residues contact preferences and hence the positive values of inter residue interaction energy are observed. The contact preferences observed in the present work are similar to those obtained by Miyazawa and Jernigan [11]. The inter residue contact energy of hydrophobic residues such as Cys-Cys, Phe-Phe, Ile-Ile, Val-Val, Trp-Trp and Tyr-Tyr posses very minimum energy which is shown as blue color grid in the figure 1. The hydrophilic residues Arg-Arg, Asp-Asp, Glu-Glu and Lys-Lys have maximum energy values which are shown in dark red color in the figure 1. Other residues Asn-Asn, Gly-Gly, His-His, Leu-Leu and Met-Met have medium level of energy values which is shown as white grids in the figure 1.

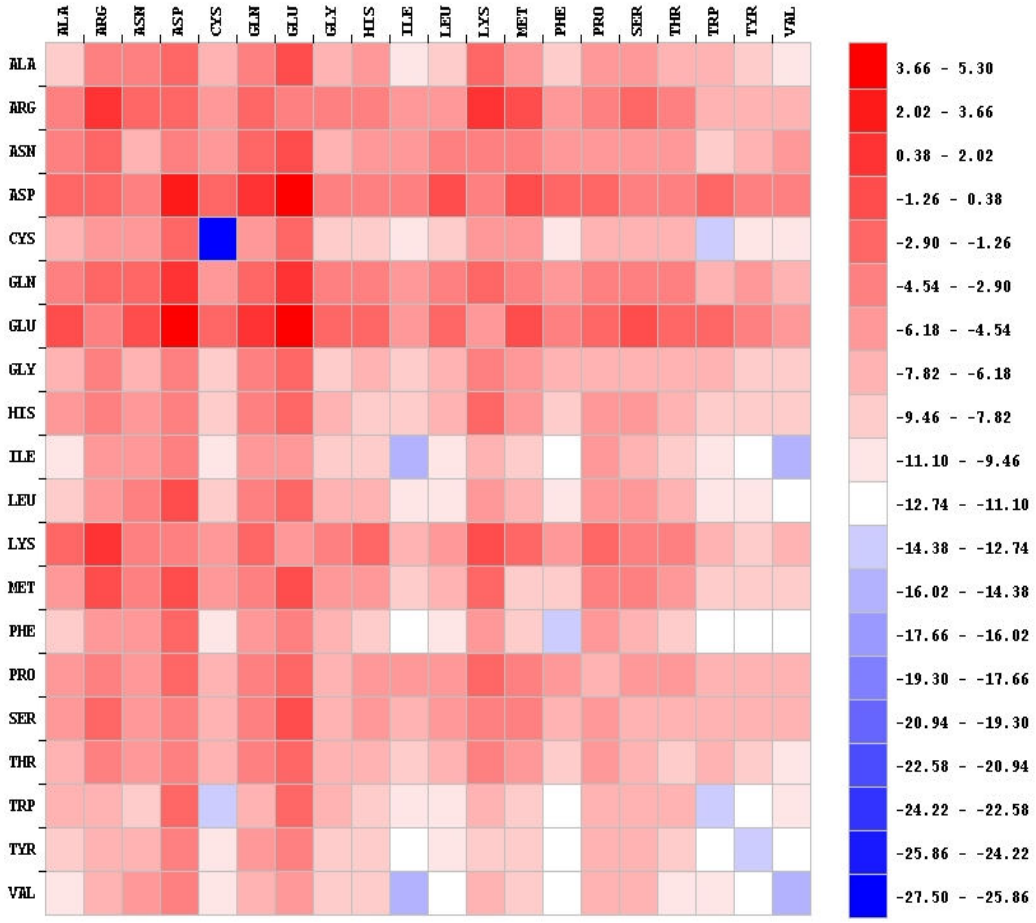

Figure 1. Grid diagram of inter residue contact energy (Kj/Mol) derived from the set of 1334 non-redundant proteins

A careful analysis of the statistical contact energy potentials, reveals that hydrophobicity is the strong determinant of the folding of a protein. It was also shown previously that the compatibility of a sequence for a given fold can be determined by aligning hydrophobic residues to buried positions in the structural motif [23]. However, not all proteins depend on hydrophobic interactions for achieving their stability; for example, Levitt and coworkers proposed a novel contact potential which do not take any information from known protein structures but follows the rule “hydrophobic residues inside and polar residues outside” [24]. Further, it is believed that the hydrophobic residues contribute to the stability of the protein depending on extent of burial in the structure and hence it is postulated that the contacts between the hydrophobic residues in the protein’s interior may be the most important link between sequence and structure which is clearly reflected in the inter residue pairwise contact energy potentials [25-27].

3.2. Comparison of surrounding hydrophobicity with statistical potentials

The relationship between surrounding hydrophobicity and inter residue contact energy of amino acid residues is shown in figure 2. A clear relationship between surrounding hydrophobicity and inter residue contact energy of amino acid residues could be observed. A correlation coefficient of -0.91 is observed between the surrounding hydrophobicity and pairwise interaction energy of amino acid residues of proteins in the fold library. This suggests that the interaction and placement of hydrophobic residues seems to be a more critical determinant of protein structure than the role of polar residues and local, sequence-dependent interactions. The hydrophobic residues are thought to be mostly responsible for the overall topological features of a protein [28]. It is shown that hydrophobic-hydrophobic residue pairs are the most preferred for long-range interactions and these long-range interactions play a key role in folding and stabilization of proteins [29].

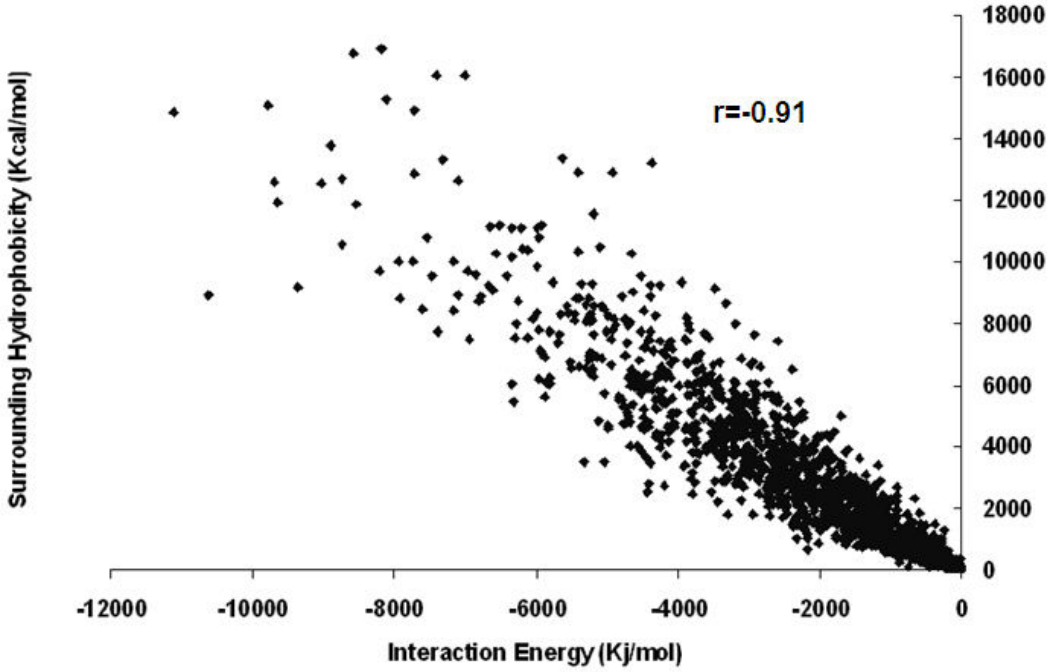

Figure 2. Relationship between surrounding hydrophobicity and inter residue interaction energy

It is noted that the hydrophobic residue-residue preferences and the inter residue contact energy derived here are mainly dominated by hydrophobic interactions. The correlation coefficient between surrounding hydrophobicity and inter residue contact energy implies that the relationship is due to the dominance of hydrophobic-hydrophobic residue preferences. Although, a good correlation coefficient is found between surrounding hydrophobicity and inter residue interaction energy, it is of interest to relate other statistical potentials developed by leading protein science groups with surrounding hydrophobicity. The correlation coefficient between surrounding hydrophobicity and statistical potential derived in the present work, Folgari group, Thomas and Dill and Skolnick group in different structural classes is shown in Table1. A good correlation coefficient between surrounding hydrophobicity and inter residue contact energy is observed in all structural classes (all-α, all-β, α+β and α/β respectively). A reasonably good correlation coefficient is observed between surrounding hydrophobicity and other three statistical potentials in all structural classes except all-β class of proteins in the fold library. The correlation coefficient of all-β class of proteins between surrounding hydrophobicity and interaction energy of Folgari group, Thomas and Dill and Skolnick group is -0.63, -0.48 and -0.36 respectively.

Table 2 Correlation coefficient between surrounding hydrophobicity and statistical potentials derived by us and other leading groups in different structural classes

Structural Class	Present work	Folgari Group	Thomas & Dill	Skolnick Group
All Alpha	-0.954	-0.781	-0.678	-0.644
All Beta	-0.878	-0.637	-0.487	-0.368
Alpha+Beta	-0.89	-0.789	-0.738	-0.657
Alpha/Beta	-0.85	-0.825	-0.764	-0.737
Others	-0.943	-0.802	-0.758	-0.744

In β-strands, hydrophobic residues are clustered together to form hydrophobic core [30] and these hydrophobic core residues contribute to two major interactions namely hydrophobic and packing interactions. Since most of the tertiary contacts formed by residues in β-strands are hydrophobic, there may be a fundamental difference between the structural and energetic principles governing β-strands. Sensible correlation coefficient is observed in the other classes of proteins such as small proteins, designed proteins and coiled-coiled proteins respectively. Interestingly, the unified statistical potentials describing protein-protein interactions derived by Skolnick and coworkers show good correlation (-0.744) with surrounding hydrophobicity. This may be due to the similar inter residue contacts between attractive (Hydrophobic-hydrophobic) and repulsive (Hydrophillic-hydrophillic) residue contacts in protein dimers and monomers.

4. CONCLUSION

The known three-dimensional structures of proteins contain a large amount of information on the forces stabilizing proteins. The inter residue contact preferences derived in the present work reveal that the hydrophobic-hydrophobic residue contacts are dominant and could be more responsible to form stable structure core than other contacts in a protein. Our analysis reveals that the interaction and placement of hydrophobic residues seems may be a more critical determinant of protein structure. Surrounding hydrophobicity of the proteins in the fold library are compared with the inter residue contact energy of proteins in the fold library. An inverse correlation coefficient of -0.91 was found between surrounding hydrophobicity and interaction energy derived in the present work. This implies that the energetics of the residues in the protein is highly dependent on its surrounding environment. It should be noted that 6.5Å distance cutoff is used to derive statistical potentials as used in Miyazawa-Jernigan procedure [11] whereas 8Å distance cutoff is used to compute pairwise interaction energy and surrounding hydrophobicity as defined as Manavalan and Ponnuswamy [8]. A good relationship has been found between surrounding hydrophobicity and interaction energy computed from three statistical potentials developed by Thomas and Dill, Berrera et al., and Skolnick group in all the structural classes except all-β class of proteins in the fold library. The correlation between the statistical potentials and surrounding hydrophobicity suggests that this property may be exploited further for fold recognition studies.

ACKNOWLEDGEMENTS

KMS is supported by a National Post-Doctoral Fellowship (File No: PDF/2015/000276) by Science Engineering Research Board (SERB), Government of India, under the mentorship of Dr. P. Karthe, Professor of Crystallography & Biophysics at the University of Madras.

REFERENCES

[1] Miyazawa S, Jernigan RL., Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation, Macromolecules, 1985; 18: 534-552.
[2] Bahar I, Jernigan RL., Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol, 1997; 266: 195-214.
[3] Keskin O, Bahar I, Jernigan RL, Badretdinov AY, Ptitsyn OB., Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions, Protein Sci, 1998; 7: 2578-2586.
[4] Gromiha MM., Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins, Biophys Chem, 2001; 91: 71-77.
[5] Fariselli P, Olmea O, Valencia A, Casadio R., Progress in predicting inter-residue contacts of proteins with neural networks and correlated mutations, Proteins, 2001; 45: 157-162.
[6] Gromiha MM, Selvaraj S., Inter-residue interactions in protein folding and stability, Prog Biophys Mol Biol, 2004; 86: 235-277.
[7] Mary RD, Saravanan MK, Selvaraj S., Conservation of inter-residue interactions and prediction of folding rates of domain repeats, J Biomol Struct Dyn, 2015; 33: 534-551.
[8] Manavalan P, Ponnuswamy PK., Hydrophobic character of amino acid residues in globular proteins, Nature, 1978; 275: 673-674.
[9] Maiorov VN, Grippen GM., Contact potential that recognizes the correct folding of globular proteins, J Mol Biol, 1992; 227: 876-888.
[10] Sippl MJ., Knowledge-based potentials for proteins, Curr Opin Struct Biol, 1995; 5: 229-235.
[11] Miyazawa S, Jernigan RL., Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading, J Mol Biol, 1996; 256: 623-644.
[12] Thomas PD, Dill KA., Statistical potentials extracted from protein structures: how accurate are they?. J Mol Biol, 1996; 257: 457-469.
[13] Melo F, Sánchez R, Sali A., Statistical potentials for fold assessment, Protein Sci, 2002; 11: 430-448.
[14] Saravanan KM, Suvaithenamudhan S, Parthasarathy S. Selvaraj S., Pairwise contact energy statistical potentials can help to find probability of point mutations, Proteins, 2017; 85: 54-64.
[15] Zhang C, Kim SH., Environment-dependent residue contact energies for proteins, Proc Natl Acad Sci, 2000; 97: 2550-2555.

[16] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE., The protein data bank, Nucleic Acids Res, 2000; 28: 235-242.
[17] Berrera M, Molinari H, Fogolari F., Amino acid empirical contact energy definitions for fold recognition in the space of contact maps, BMC Bioinformatics, 2003; 4: 8.
[18] Lu H, Lu L, Skolnick J., Development of unified statistical potentials describing protein-protein interactions, Biophys J, 2003; 84: 1895-1901.
[19] Wang G, Dunbrack RL., PISCES: a protein sequence culling server, Bioinformatics, 2003; 19: 1589-1591.
[20] Murzin AG, Brenner SE, Hubbard T, Chothia C., SCOP: a structural classification of proteins database for the investigation of sequences and structures, J Mol Biol, 1995; 247: 536-540.
[21] Xu Y, Xu D., Protein threading using PROSPECT: design and evaluation, Proteins, 2000; 40: 343–354.
[22]. Nozaki Y, Tanford C., The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale, J Biol Chem, 1971; 246: 2211-2217.
[23] Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT., Deciphering the message in protein sequences: tolerance to amino acid substitutions, Science, 1990; 247: 1306-1311.
[24] Huang ES, Subbiah S, Levitt M., Recognizing native folds by the arrangement of hydrophobic and polar residues, J Mol Biol, 1995; 252: 709-720.
[25] Gromiha MM, Pathak MC, Saraboji K, Ortlund EA, Gaucher EA., Hydrophobic environment is a key factor for the stability of thermophilic proteins, Proteins, 2013; 81: 715-721.
[26] Rishyakulya MC, Saravanan KM., Computational Structural Analysis of C-Terminal Residues of Proteins Containing Transmembrane Regions, Int J Comp Biol, 2015; 4: 44-54.
[27] Gromiha MM, Selvaraj S., Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction, J Mol Biol, 2001; 310: 27-32.
[28] Aftabuddin M, Kundu S., Hydrophobic, hydrophilic and charged amino acid networks within protein, Biophys J, 2007; 93: 225-231.
[29] Selvaraj S, Gromiha MM., Importance of long-range interactions in (α/β) 8 barrel fold, J Prot Chem, 1998; 17: 691-697.
[30] Gromiha MM, Suwa M., Variation of amino acid properties in all-β globular and outer membrane protein structures, Int J Biol Macromol, 2003; 32: 93-98.