Hydrophobic-Hydrophilic Forces and their Effects on Protein Structural Similarity

Trent Higgs, Bela Stantic, Md Tamjidul Hoque, and Abdul Sattar

Institute for Integrated and Intelligent Systems
Griffith University, Queensland, Australia
{T.Higgs, B.Stantic, T.Hoque, A.Sattar}@griffith.edu.au

Abstract. Hydrophobic-hydrophilic interactions have a strong impact on the three-dimensional structure a protein will adopt. Because structure, not amino acid sequence order, carry out certain functions it is important to understand how these forces affect the protein folding process. In recent years, a lot of focus has been dedicated towards *ab initio* protein folding prediction, which tries to predict a proteins native conformation from its sequence alone. To aid this type of prediction sub-conformations from already known proteins are used to limit the free energy conformational search space. In this paper we looked into the sub-conformations' hydrophobic-hydrophilic nature by incorporating a HP approach and proposed a way of evaluating how these type of forces affect the protein folding process. By doing this, we can gain insight into how hydrophobic-hydrophilic interactions affect protein structural similarity, and thus aid us in picking more suitable sub-conformations based off their HP shape for use in protein structure prediction.

1 Introduction

The idea of being able to successfully predict a proteins three-dimensional structure is a mystery that has baffled scientists for many years. The reason why a solution to the protein folding problem has been heavily sought after is due to their importance. Proteins carry out all of the main functionality within an organism on a cellular level. For example, red blood cells contain a protein known as the hemoglobin. This protein carries out the functionality of carrying oxygen to the blood stream. To understand proteins and their functionality better an investigation into their three-dimensional structure is required. This is because structure, not the sequence of amino acids, carry out certain biological functions [6] [7].

To elicit a proteins three-dimensional structure numerous computational methods have been developed [13] [8] [15] [5]. Each of these methods use a variation of one or more of the following prediction techniques: comparative modelling, threading, and *ab initio*. For our work we have primarily focused on the *ab initio* method, which predicts a proteins three-dimensional structure from its sequence alone. In this type of prediction sub-conformations can be used to limit the free energy conformational search. In previous work [4] the length of these sub-conformations or fragments and how they affected the protein folding prediction process was investigated. To extend on this work we are interested in how hydrophobic-hydrophilic interactions will affect

M. Chetty, S. Ahmad, A. Ngom, S. W. Teng (Eds.): PRIB2008,
Supplementary Conf. Proc., pp.1-12, 2008.

these sub-conformations and structural similarity within the prediction process, due to these forces being highly desirable within protein folding [14].

In this paper by incorporating a HP model approach we have investigated the hydrophobic and hydrophilic nature of sub-conformations. A HP pattern model can add an extra layer of sensitivity to the fragment selection process within fragment-based protein prediction software. By presuming a certain HP shape for a particular sub-conformation it can allow the best sub-conformation to be used at a particular segment of the protein chain in less time, therefore quicken the accuracy and computation time of the prediction process. To investigate this we have ran experiments to find out how hydrophobic or hydrophilic a k-sized (where, k is a non-zero positive integer number) conformation was, and what effects this had on identical sub-conformations structural similarity. We also applied a HP model concept to our database of already folded proteins, to discover how viable it is to use HP shapes within the sub-conformation selection/generation process.

The reminder of this paper is organized as follows: in Section 2 we present background information, Section 3 discusses the proposed methodology, Section 4 presents preliminary experimental results, and in Section 5 we conclude our findings and mention future work.

2 Background

A protein is made up of a collection of amino acids, which are molecules that have both carboxyl and amino groups. An amino acid contains a carbon atom (Cα), and has four different connections, these include an amino group, carboxyl group, a hydrogen atom, and a side chain (this differs depending on the amino acid). The Cα atom is the central atom of the amino acid and all of the other connectors are attached to it [2]. There are 20 of these standard amino acids, and they are the building blocks for the numerous conformations a protein can adopt.

To predict a protein's structure by means of conformational search can take an enumerable amount of conformations to be computed. Even for the simplified assumption [1] that if each amino acid can have 3 degrees of rotation, a protein chain that has 200 residues could at the very minimum have 3^{200} possible conformations, which is an astronomical number. Hence, it is very hard to predict a proteins three-dimensional structure by searching all possible structures available. To alleviate this problem 3 main computational methods to predict a proteins structure have been developed, these are: comparative modelling, threading, and *ab initio*.

As our work is focused primarily on the *ab initio* technique we will only talk about this method. *Ab initio*, unlike the other two techniques, will predict a proteins structure from the proteins primary sequence alone. It does this by applying Afinsen's theory that a proteins native three-dimensional structure is at its lowest free energy minimum [9]. The problem with this technique is that the free energy conformational space is quite large. Therefore, a lot of protein folding prediction methods use sub-conformations to limit the amount of conformations considered for a particular segment of the protein chain. A sub-conformation is several secondary structures joined together to form a partially conformed segment of the protein chain.

Due to sub-conformations being partially conformed segments of a protein, they are therefore composed of amino acids. Amino acids have numerous properties, however the most important in concerns to protein folding is if there side chains are hydrophobic (H) or hydrophilic (P). hydrophobic-hydrophilic (HP) interactions happen when a protein is folding into its tertiary structure. The amino acids with hydrophobic side chains move to the core of the protein to be away from water, and the hydrophilic side chains move towards the outside of the protein because they have an affinity with water. The shape of the protein is then defined by the Van Der Waals attractions that strengthen these hydrophobic interactions. Even though these interactions are weak the sheer number of them cause a protein to take on a specific shape [2].

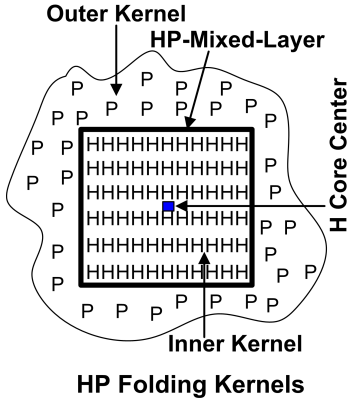

Fig. 1. Metaphoric HP Folding Kernels [5]

State-of-the-art fragment/sub-conformation based protein structure prediction software like: Rosetta [11] [12] and Tasser [15] [16] do not take into consideration the polarity of the amino acids when they are randomly inserting sub-conformations. Instead they apply the HP interactions after a rough model has been generated from random sub-conformation insertions. By investigating the effects of HP interactions in regards to protein structural similarity within sub-conformations/fragments the protein folding prediction process could be significantly improved (i.e. more likely sub-conformations could be chosen before the prediction process begins and by defining a HP pattern particular HP shapes could be more accurately picked during the process also). This sort of approach was used in [5] by packing likely sub-conformations (mapped from short sub-sequences) that were mainly hydrophobic (H) within the proteins core (H-Core), while placing mostly hydrophilic (P) amino acids on the outside kernel (see Figure 1).

3 Testing the Polarity of Sub-Conformations

To test hydrophobicity within sub-conformations we have used Kyte's and Doolittle's 'Hydropathy Index' [10] as a model. This works by assigning each amino acid a value

to represents its hydrophobic (H) or hydrophilic (P) properties. The larger the value is the more hydrophobic the amino acid is (see Table 1).

Table 1. Kyte's and Doolittle's Hydropathy Index [10]

Amino Acid	Three-Letter	Hydropathy Index
Glycine	GLY	-0.4
Alanine	ALA	1.8
Proline	PRO	1.6
Valine	VAL	4.2
Leucine	LEU	3.8
Isoleucine	ILE	4.5
Methionine	MET	1.9
Phenylalanine	PHE	2.8
Tyrosine	TYR	-1.3
Tryptophan	TRP	-0.9
Serine	SER	-0.8
Threonine	THR	-0.7
Cysteine	CYS	2.5
Asparagine	ASN	-3.5
Glutamine	GLN	-3.5
Lysine	LYS	-3.9
Histidine	HIS	-3.2
Arginine	ARG	-4.5
Aspartate	ASP	-3.5
Glutamate	GLU	-3.5

From Table 1 you can see that the most hydrophobic amino acids are Isoleucine (4.5) and Valine (4.2), and the most hydrophilic amino acids are Arginine (-4.5) and Lysine (-3.9). This is very important in concerns to protein structure due to hydrophobic amino acids tend to be internal and hydrophilic amino acids tend to be external within a proteins final three-dimensional structure.

To evaluate hydrophobicity within sub-conformations using Kyte's 'Hydropathy Index' we have extended upon the work of [4]. In this work already folded proteins from the PDB (Protein Data Bank) were stored within a Relational Database Management System (RDBMS) and a sub-conformation similarity algorithm was used to check how often a particular motif/fragment of a set size within each protein in the protein database matched with every other protein sequence contained within the database.

For example, if one of the motifs contained the amino acids alanine, threonine, and glycine (ATG) right after one another. ATG would be searched for throughout every protein sequence in the database and every occurrence of it that appeared in the exact same order (i.e. A as the first amino acid in the motif, T as the second amino acid in the motif, and G as the third amino acid in the motif) would be recorded and analysed. Other than recording matches the distance between the amino acids (i.e. between the

Cα atoms) of the motif/fragment were also stored to determine how structurally similar identical sub-conformations are within different proteins.

$$HPmeasure = \sum hp_i \tag{1}$$

To extend this work into a HP approach we have altered the current algorithm and created a new algorithm. Instead of just looking at the length of the sub-conformation, we now introduce a 'Hydropathy Index Measure'(HIM) which is two fold. First of all for each sub-conformation a sum of each residues hydropathy index is carried out (see equation 1 - for every residue *i* add the hydropathy index value for *i* to the total hydropathy index value for the sub-conformation *i* belongs to) (extension of current algorithm). This is to gauge how hydrophobic or hydrophilic a particular sub-conformation is. The second part of this 'Hydropathy Index Measure' is to evaluate, using the X, Y, Z coordinates of the sub conformation, if a residue/s within that sub-conformation is hydrophobic then is the shape of the sub-conformation altering due to it - moving more towards the inside of the protein (new algorithm).

To determine structural similarity between identical sub-conformations/fragments (the same sequence of amino acids), just like in [4], we have used the RMSD equation [3]. This works by measuring the distances between the Cα atoms of each amino acid within the fragment (see equation 2). We used the root mean squared distance equation rather than other statistical structural measures (e.g. root mean squared deviation) due to it being less computation intensive, and due to it producing close-enough structural similarity values.

$$rmsd = \sqrt{\sum d_i^2 \div (n(n-1) \div 2)} \tag{2}$$

3.1 HP Sub-Conformation Similarity Algorithm

The algorithm that we used for the HP protein sub-conformation search can be found in Algorithm 1. This algorithm works by first grabbing all of the proteins within the database (`protAll`). It will then iterate throughout all of `protAll` so that all proteins within the database are searched. The main body introduces two sub-conformations `A` and `B`. `A` is a protein sub-conformation of size k (where k $\geq$ 4) that starts from the amino acid of a particular protein that `protAll` is currently on (curr) and ends k-1 amino acids past curr (i.e. A = protAll.currentAcidID to protAll.currentAcidID + k-1). `B` on the other hand holds many different sub-conformations depending on the first amino acid in sub-conformation A.

`B` is assigned by finding the amino acid sequence positions (within any protein, even the same one) in the database that have the same amino acid as the first acid in A (`protAcid`). `protAcid` is then iterated through and each time B is assigned the fragment, k in length (where k $\geq$ 4), generated by protAcids current amino acid for a particular protein (curr) up to k-1 amino acids past curr (i.e. B = protAcid.currentAcidID to protAcid.currentAcidID + k-1). After `A` and `B` are both found then checking if the amino

acid of A at position two matches B's amino acid at position two and then same for position three. If this is the case then the root mean squared distance for B is calculated. Then the hp measure for that sub-conformation is calculated.

Algorithm 1: HP Sub-Conformation Similarity Prediction Algorithm

```
protAll = get all proteins in the database;
WHILE (protAll NOT NULL)
    IF (protAll.prot_name != previous.protAll.prot_name)
        Mark previous.protAll.prot_name AS DONE;
    END IF
    A() = sub-conformation of protein from protAll.acidID
    to protAll.acidID + k - 1;
    dist = RMSD for A(1) to A(3);
    Add A(1) ... A(3) & dist to match, report;
    calculate HP measure for A(1) to A(3) and add to report;
    protAcid = All amino acids within all proteins
                that contain A(1).acid_name;
    WHILE (protAcid is NOT NULL)
        B() = sub-conformation of protein from protAcid.acidID
        to protAcid.acidID + k - 1;
        IF (A(2).acid_name = B(2).acid_name}
            IF (A(3).acid_name = B(3).acid_name)
                dist = RMSD for B(1) to B(3);
                Add B(1) ... B(3) & dist to match, report;
                calculate HP for B(1) to B(3) and add to report
                (if not recorded already);
                IF (A(4).acid_name = B(4).acid_name)
                    dist = RMSD for A(1) to A(4);
                    Add A(1) ... A(4) & dist to match, report;
                    calculate HP for A(1) to A(4) and add to
                    report (if not recorded already);
                    dist = RMSD for B(1) to B(4);
                    Add B(1) ... B(4) & dist to match, report;
                    calculate HP for B(1) to B(4) and add to report
                    (if not recorded already);
                    ...
                    IF(A(k).acid_name = B(k).acid_name)
                            ...
                    END IF;
                END IF;
            END IF;
        END IF;
    END WHILE;
END WHILE;
Mark last protein AS DONE;
```

Once all of these values have been calculated the RMSD, and HP measure are all recorded within the database. The same is done for 4, 5, 6, 7 ... k fragments in length, but there is one slight difference. Due to the algorithm automatically adding `A`, 3 sized fragments to the database as a default, there is a need to calculate the root mean squared distance, and HP measure for A 4, 5, 6, 7...k sized fragment matches and add them to the match and report tables before calculating the RMSD value, and HP measure for `B`.

3.2 HP Shape Analysis Algorithm

To observe HP interactions within sub-conformations we have designed idealistic HP shape patterns and analysed the pdb structures within our database against these patterns. For an example of these HP patterns please see Figure 2.

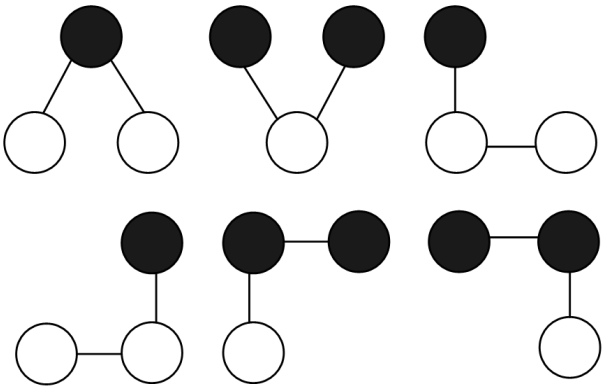

Fig. 2. 3-Residue HP Patterns

Figure 2 contains a set of patterns we applied for 3 residue sub-conformations. The black residues are hydrophobic (which can be represented as H), and the white residues are hydrophilic (which can be represented as P). For each pattern the hydrophobic residues face more towards the center of the proteins core, therefore they are idealistic depictions of how a protein is meant to fold based on desirable hydrophobic interactions [14].

The algorithm to do this shape analysis can be found in Algorithm 2. This algorithm works by going through ever k-sized sub-conformation within the protein database, and comparing those sub-conformations against a HP pattern database (for example of HP patterns contained within the database see Figure 2). If a sub-conformation matches a HP pattern in the database, it is then checked to see if it is valid. To be valid the hydrophobic residues must be closer to the core of the protein than the hydrophilic residues. If it is then the HP pattern and sub-conformation are recorded, and it is marked as valid, otherwise it is marked as invalid.

Algorithm 2: HP Shape Analysis Algorithm

```
protAll = get all proteins in the database;
WHILE (protAll NOT NULL)
    frag() = sub-conformation of protein from protAll.acidID
    to protAll.acidID + k - 1;
    hp_pattern = compare frag(1), frag(2)... frag(k)
    to HP pattern database;
    IF (hp_pattern exists)
        IF (frag matches hp_pattern)
            record frag details, hp_pattern and mark as true;
        ELSE
            record frag details, hp_pattern and mark as false;
        END IF;
    END IF;
END WHILE;
```

4 Experimental Results

For our preliminary results we have carried out the HP sub-conformation similarity algorithm and HP shape analysis algorithm on our current database of 24 000 pdb structures, that contain just one chain and the Cα atom of each amino acid. In Table 2 we present the average HP measure, average minimum RMSD, and average maximum RMSD for a particular sub-conformation length. This is accompanied by a graph (Figure 3) of the average HP measure for a particular sub-conformation length (in our case 3-12). The x axis is the length of the sub-conformation and the y axis is the average HP measure. Figure 4 shows the average structural difference for every matching sub-conformation for a particular length. This is calculated by subtracting the average maximum RMSD from the average minimum RMSD. The x axis is the length of the sub-conformation, and the y axis is the average structural difference.

Table 2. Average HP Measure for Sub-Conformations 3-12 in length

Length	Avg HP Measure	Avg Min RMSD	Avg Max RMSD
3	-1.0360245	4.131	39.089
4	-0.9014449	4.639	10.775
5	-1.0022609	5.690	7.223
6	-1.1843423	6.669	7.325
7	-1.3879915	7.383	7.975
8	-1.5934906	8.034	8.636
9	-1.8164757	8.654	9.279
10	-2.049479	9.248	9.897
11	-2.293114	9.813	10.483
12	-2.5264208	10.346	11.033

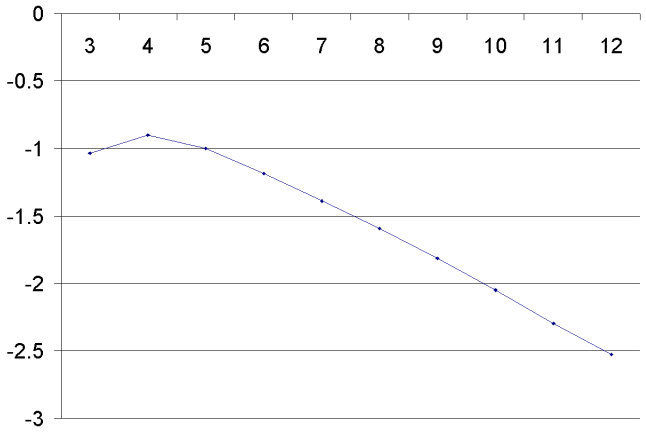

Fig. 3. Average HP measure for Sub-Conformations 3-12 in Length

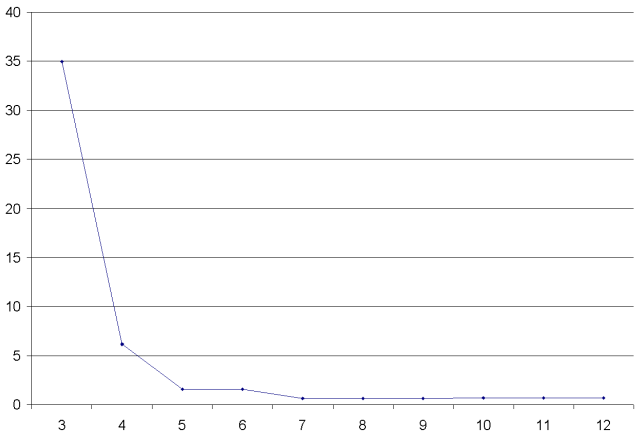

Fig. 4. Average Structural Difference between Identical Matching Sub-Conformations (i.e. same amino acid order) for Lengths 3-12

For the HP shape analysis experiment we only used patterns that were 3-residues in length for our preliminary results. Each of these patterns can be found in Figure 2. In Table 3 we show how many valid and invalid sub-conformations there were for a particular HP pattern. Each HP pattern is represented by H for hydrophobic, and P for hydrophilic.

Table 3. HP Patterns and their Validity

HP Pattern	No of Valid Sub-Conformations	No of Invalid Sub-Conformations
PHP	259526	516588
HPH	210112	534814
HPP	424232	369426
PPH	402392	378949
PHH	0	0
HHP	823661	686055

4.1 Analysis and Discussion

From our results our main intentions were to observe how hydrophobic or hydrophilic a k-sized sub-conformation was on average (and how this affected structural similarity between identical fragments). We were also interested in if HP patterns were suitable to be used to improve the sub-conformation selection process within fragment-based protein structure prediction software. First of all we will discuss the HP measure results, which looked into how hydrophobic or hydrophilic a certain k-sized sub-conformation was. These results can be found in Table 2.

The first observation that can be made from this is that on average the longer the sequence gets the more hydrophilic the overall sub-conformation becomes (see Figure 3). In Figure 3 you can see that it slightly becomes more hydrophobic on 4-sized sub-conformations, but after that it steadily goes down, meaning the overall sub-conformation is becoming more hydrophilic. This could be attributed to there being more hydrophilic or neutral (e.g. could be hydrophobic but its 'Hydropathy Index' is lower than other hydrophobic amino acids) contained within the sub-conformation. Take for example a 12-sized sub-conformation that contains LYS, ASP, GLU, ALA, GLU, LYS, LEU, PHE, ASN, GLN, ASP, and VAL, four of these amino acids are highly hydrophobic, and 8 are highly hydrophilic, according to [10].

In regards to structural similarity, as depicted in [4], the longer the sequence becomes the less the structural difference will be between identical sub-conformations found within different proteins (see Table 2 and Figure 4). In Figure 4 you can see that on average once the length of the sub-conformation becomes 7, the structural difference is minimal, which means the longer the sequence becomes the more hydrophilic it will be, and the structural difference between identical fragments becomes significantly less (which could be due to the increase in hydrophilic amino acids within the sub-conformation).

Apart from the HP measure, we considered HP patterns 3-residues in length to see if a sub-conformation's shape would change based on the polarity of the amino acids that it was composed of. These results can be found in Table 3. In this experiment patterns PHP and HPH had more invalid sub-conformations, than valid sub-conformations. From this we can conclude that PHP, and HPH patterns occur a lot within our pdb database (approximately 760520 appearances on average), however the hydrophobic residues appear to be more towards the outside of the protein's core, rather than the idealistic picture of them moving towards the inside of the protein (i.e. due to the huge dif-

ference between valid and invalid sub-conformations). Despite this, due to there being some valid conformations, there must be some other force/s involved that determines whether or not the protein shape should change due to the polarity of its amino acids.

As for the pattern PHH, it appears that this pattern does not occur at all. There are no valid or invalid sub-conformations for this HP pattern. If this is the case then this sub-conformation pattern PHH could be easily discarded when picking sub-conformations for a particular insertion within a fragment based protein folding prediction software, due to the knowledge that they do not exist in real proteins. It is important to note that there is a possibility for this being incorrect. The definition of the center of the protein's core that we have used for our experiments may not being 100% accurate, due to there being numerous ways to do this. If the definition we have used is slightly off, it may cause some of our results to be inaccurate, which may be the case here.

The HP patterns that had more valid than invalid conformations were HPP, PPH, and HHP, which is exactly half of the HP patterns used. However, even though these HP patterns performed better than PHP, HPH, and PHH there was still a large number of invalid conformations within these HP patterns (approximately 71952 on average). This further proves that there must be one or more other force/s involved that defines the shape a particular sub-conformation will take on, rather than hydrophobicity alone.

5 Conclusions and Future Work

In this paper we have conducted a investigation into hydrophobic-hydrophilic interactions within k-sized sub-conformations to improve the fragment generation process for protein folding prediction software. To do this we came up with our 'Hydropathy Index Measure' that looked into the hydrophobic-hydrophilic nature of a k-sized sub-conformation. This involved looking at how hydrophobic or hydrophilic a k-sized sub-conformation was overall with our HP measure, and how likely that HP patterns could be used to determine a sub-conformations rough shape (i.e. hydrophobic amino acids being located more towards the protein's core, and the hydrophilic amino acids being more towards the outer surface of the protein).

The most significant observations we came across in our study was that only a few of the amino acids contained within a k-sized sub-conformation would be hydrophobic, and the rest would most likely be hydrophilic or neutral amino acids. This was noted due to the fact that as the sub-conformation grew in size the more hydrophilic it became overall, and the structural difference gap also lessened a fair bit, which could be attributed to its hydrophilic nature. As for HP patterns, it was quite obvious that our results could be used to rank more likely HP patterns higher over less likely patterns in a fragment-based protein folding prediction software, rather than random insertions (i.e use a HHP HP pattern sub-conformation, instead of PHP or PHH). However, due to their being so many invalid sub-conformations for each HP pattern, there must be more forces involved than hydrophobicity alone that determines a sub-conformation's shape.

In regards to future work we plan to conduct more experiments that include $>$3-sized HP patterns to determine if the size increases does the valid:invlaid ratio move to a more acceptable level. Also, it would be interesting to add different forces (e.g. positively and negatively charges) to the HP patterns to determine how other forces

affect the shape for a particular sub-conformation, and if they can be applied to the fragment selection process.

Acknowledgements

This research is partly sponsored by ARC (Australian Research Council) grant no DP0557303.

References

1. Baker, D.: Proteins by Design. The Scientist. 26–32 (2006)
2. Campbell, N.A., Reece, J.B., Mitchell, L.G.: Biology: Fifth Edition. Benjamin Cummings, USA (1999)
3. Carugo, O.: Statistical Validation of the Root-Mean-Square-Distance, a Measure of Protein Structural Proximity. Protein Engineering, Design and Selection. 20, 33–38 (2007)
4. Higgs, T., Stantic, B., Hoque, T.: Optimal Length of Fragments for Use in Protein Structure Prediction. In: IASTED Advances in Computer Science and Technology, pp. 156–161. Langkawi, Malaysia (2008)
5. Hoque, T., Madhu, C., Laurence, D.: A New Guided Genetic Algorithm for 2D Hydrophobic-Hydrophilic Model to Predict Protein Folding. In: IEEE Congress on Evolutionary Computation, pp. 259–266. Edinburgh, UK (2005)
6. Huang, Z.H., Zhou, X.: High Dimensional Indexing for Protein Structure Matching using Bowties. In: 3rd Asia-Pacific Bioinformatics Conference, pp. 21–30. Singapore (2005)
7. Huang, Z.H., Zhou, X., Song, D., Bruza, P.: Dimensionality Reduction in Patch-Signature Based Protein Structure Matching. In: 17th Australiasian Database Conference, pp. 89–97. Hobart, Australia (2006)
8. Hung, L-H., Samudrala, R.: PROTINFO: Secondary and Tertiary Protein Structure Prediction. Nucleic Acids Research. 31, 3296–3299 (2003)
9. Jordan, I., Kondrashov, F., Adzhubei, Y., Wolf, E.K., Kondrashov, A., Sunyaev, S.: A Universal Trend of Amino Acid Gain and Loss in Protein Evolution. Nucleic Letter to Nature. 433 (2005)
10. Kyte, J., Doolittle, R.F.: A Simple Method for Displaying the Hydropathic Character of a Protein. Journal of Molecular Biology. 157, 105–132 (1982)
11. Simons, K.T., Kooperberg, C., Huang, E., Baker D.: Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. Journal of Molecular Biology. 268, 209–225 (1997)
12. Simons, K.T., Bonneau, R., Ruczinski, I., Baker D.: Ab Initio Protein Structure Prediction of CASP III Targets using ROSETTA. Proteins. 37, 171–176 (1999)
13. Simons, K.T., Strauss, C., Baker D.: Prospects for Ab Initio Protein Structural Genomics. Journal of Moleculer Biology. 306, 1191–1199 (2001)
14. Yue, K., Dill, K.A.: Forces of Tertiary Structure Organization in Globular Proteins. Biophysics. 92, 146–150 (1995)
15. Zhang, Y., Skolnick, J.: Automated Structure Prediction of Weakly Homologous Proteins on a Genomic Scale. PNAS. 101, 7594–7599 (2004)
16. Zhang, Y., Skolnick, J.: Tertiary Structure Predictions on a Comprehensive Benchmark of Medium to Large Size Proteins. Biophysical Journal. 87, 2647-2655 (2004)