Journal of Bioinformatics and Computational Biology
Vol. 11, No. 1 (2013) 1340012 (16 pages)

DOI: 10.1142/S021972001340012X

Imperial College Press
www.icpress.co.uk

PUTRACER: A NOVEL METHOD FOR IDENTIFICATION OF CONTINUOUS-DOMAINS IN MULTI-DOMAIN PROTEINS

SEYED SHAHRIAR ARAB[*,†,§,‡‡], MOHAMMADBAGHER PARSA GHARAMALEKI[†,‡,¶,‖], ZAIDDODINE PASHANDI[*,**] and REZVAN MOBASSERI[*,††]

[*]*Department of Biophysics, Faculty of biological Sciences*
Tarbiat Modares University (TMU), Tehran, Iran

[†]*Department of Bioinformatics*
Institute for Research in Fundamental Sciences (IPM)
Tehran, Iran

[‡]*Department of Mathematics*
University of Wisconsin Madison (UW-Madison) Madison, USA
[§]*sh.arab@modares.ac.ir*
[¶]*mbparsa@ipm.ir*
[‖]*gharamaleki@wisc.edu*
[**]*z.pashandi@modares.ac.ir*
[††]*rezvan.mobasseri@modares.ac.ir*

Received 1 October 2012
Revised 12 December 2012
Accepted 13 December 2012
Published 21 February 2013

Computer assisted assignment of protein domains is considered as an important issue in structural bioinformatics. The exponential increase in the number of known three dimensional protein structures and the significant role of proteins in biology, medicine and pharmacology illustrate the necessity of a reliable method to automatically detect structural domains as protein units. For this aim, we have developed a program based on the accessible surface area (ASA) and the hydrogen bonds energy in protein backbone (HBE). PUTracer (Protein Unit Tracer) is built on the features of a fast top-down approach to cut a chain into its domains (contiguous domains) with minimal change in ASA as well as HBE. Performance of the program was assessed by a comprehensive benchmark dataset of 124 protein chains, which is based on agreement among experts (e.g. CATH, SCOP) and was expanded to include structures with different types of domain combinations. Equal number of domains and at least 90% agreement in critical boundary accuracy were considered as correct assignment conditions. PUTracer assigned domains correctly in 81.45% of protein chains. Although low critical boundary accuracy in 18.55% of protein chains leads to the incorrect assignments, adjusting the scales causes to improve the performance up to 89.5%. We discuss here the success or failure of adjusting the scales with provided evidences. *Availability*: PUTracer is available at http://bioinf.modares.ac.ir/software/PUTracer/

‡‡ Corresponding author.

Keywords: Protein domain assignment; continuous domains; accessible surface area; hydrogen bond energies; top-down approach; recursive.

1. Introduction

Proteins as biologic machines in all organisms are worthy of being analyzed regarding their function, folding and so on. Achieving that typically requires parsing proteins into smaller substructures called domains. Although domains are mainly considered as basic components of proteins, an exact notion of a domain is still problematic.[1,2] In the case of concept, diverse definitions of protein domains can be classified into different groups.[2,3] Relying on structural information in addition to sequence information, domains are defined as compact units with hydrophobic cores,[1,4,5] which indicate independent functions. Moreover, structural domains are spatially distinct parts of protein[6] and may fold independently. Also in a different point of view, domains can be considered as independent parts of a protein molecule that display collective motions.[7] Many methods appeared to identify domains in protein due to the grave role as structural building blocks, units and evolutional elements.

In general, three distinct categories of approaches can be presumed. Visual inspection of a structure by human experts is the most straightforward approach.[2] Semi-automated methods are an alternative approach, which employs both the expert's proficiency and computer's fastness. However, the drastic increase in the number of known protein structures in the Protein Databank (PDB) favors fully automated ones.[3]

Majority of attempts on automatic methods have been made based on the geometric and topologic criteria.[2,3] Exclamation of the existence of stable units in proteins by Wetlaufer[8] followed by attending the local compactness in protein,[9] relationship among this compact unit[10] and focusing on inter-domain and intra-domain contact density.[6,11–15]

In the language of systems analysis, there are two groups comprising top-down and/or bottom-up approaches.[9] Starting from the whole chain of a protein and cutting into smaller units is the underlying idea in the first group and assembling smaller units to the bigger one is the second group strategy.[9] Some methods apply both strategies in their algorithm; nevertheless, most others consider only one. Both top-down and bottom-up algorithms are categorized into two generations: The first generation comprises methods published during the period 1974−1994, and the second generation includes methods published from 1995 until now.[16]

Herein we describe a top-down method for identifying continuous domains in protein chains based on simple principles derived from common definition of structural domains. Privileges, limitations, and possible development of our method in comparison to some of the best algorithms are discussed while the main goal is to present an accurate and assertive approach for identifying and evaluating structural domains. We have tested our method on a benchmark dataset of 124 proteins in the Balanced_Domain_Benchmak_2 of pDomains.

2. Methods

In this section, the algorithm of PUTracer and the description of using dataset are explained.

2.1. *The algorithms of PUTracer*

Conceptually, PUTracer as an automatic domain assignment method is established based on a number of assumptions. As a simple concept, we consider domain as a distinct unit in which the hydrogen bonds network is stronger than the hydrogen bonds network between separated units. However this criterion is not sufficient for domain determining. Due to lack of an exact definition of Hydrogen Bond Energy (HBE) there is an inherent error in results by applying Hydrogen Bond Energy alone in domain assignment. Hence we considered two factors as the logic of our algorithm: Hydrogen Bond Energy and accessible surface area.

Firstly, a structural domain should consist of at least 50 residues to be considered as a correct domain. Secondly, making a cut at domain boundary causes loosing hydrogen bound less than a cut at the middle of domains, hence noticing the difference of HBE level between the whole of a chain and its domain in this way is an inescapable fact.

DSSP output files included the information of accessible surface area and hydrogen bond energy for each residue.[17] First, the total HBE of the chain is calculated by summation of HBE (extracted from DSSP output) in all residues. The query is considered as a mother domain, which has the potential of comprising child domains. Second, to estimate the position of possible child domains, the mother domain (chain) is sliced into two children domains recursively in a top-down approach. The proportion of child domains energy in each step to energy of the mother chain is crucial to reaching the decision of accepting any child domain as the true domain. This proportion in each step is equal to summation of two child domains' hydrogen bond energy ($\sum E_i$) divided by the hydrogen bond energy of mother chain (E), multiplied by an empirical weighting factor. It gives the proportion of changes in energy for having two domains (P_{HBE}) (Eq. (1)). The smaller the P_{HBE}, the more likely it will be a two domain protein.

$$P_{HBE} = \left(\frac{E_1 + E_2}{E}\right)(W_{HBE}), \quad E > E_1 + E_2 \tag{1}$$

The correctness of domain assignment in the last condition of the algorithm is checked by accepting assignments with the minimum difference between summation of accessible surface area of two child domains ($\sum ASA_i$) and the accessible surface area of mother domain (ASA). Computationally, the proportion of ASA to $\sum ASA_i$ multiplied by an empirical weighting factor gives a crucial value, which benefits the algorithm to assign domain correctly (Eq. (2)).

$$P_{ASA} = \left(\frac{ASA}{ASA_1 + ASA_2}\right)(W_{ASA}), \quad ASA_1 + ASA_2 > ASA \tag{2}$$

$$\tau = P_{HBE} + P_{ASA} \tag{3}$$

If the summation of P_{HBE} and $P_{ASA}(\tau)$ becomes more than the main empirical threshold of algorithm, the algorithm will decide to merge child domains and get back to the mother domain; and if τ becomes less than the threshold, mother domain is divided to new child domains by cutting the ith residue. However, this process continues in a recursive approach until no more children with at least 50 can be found.

Figure 1 indicates recursive assignment algorithm in PUTracer. The process starts with a chain as mother domain, and then it is parsed into two new domains. The more domains are created when the child domains are processed in step two and three. PUTracer finishes the process when the length of domains in the last stage is less than 50 residues.

2.2. *Domain assignment criteria*

To calculate the accuracy of domain assignment, we use the equation generated by Islam *et al.*[19]

$$Accuracy = \left(\frac{N_{\text{correct}}}{N_{\text{total}}}\right) * 100 \tag{4}$$

In this equation N_{correct} is the number of residues in a domain assigned by PUTracer correctly in comparison to expert methods, and N_{total} denotes total number of residues. It should be noted that the accuracy will be calculated when the number of assigned domains by PUTracer is equal to expert methods. So we consider an assignment as correct if both criteria (equal number of domains and the accuracy above 90%) are achieved.

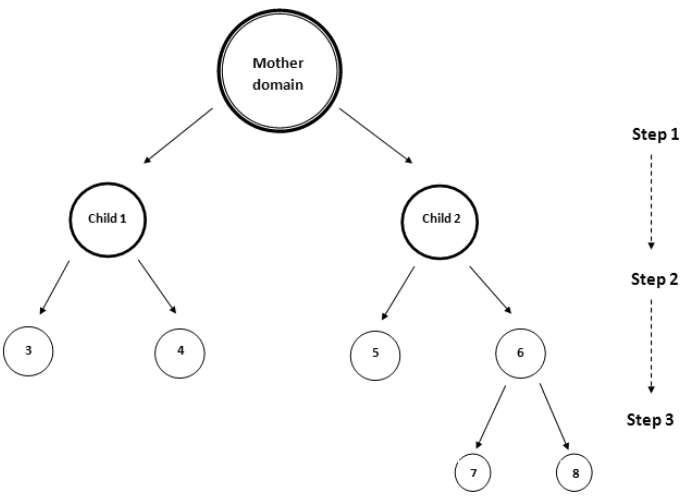

Fig. 1. Recursive assignment algorithm in PUTracer. The process starts with a chain with one mother domain, then it slides into two new domains recursively.

Table 1. Chains with different number of domains in selected dataset (124 chains).

Number of domain(s)*	Percent of chains in benchmark ()
1 Domain	43.5 %
2 Domain	39.5 %
3 Domain	11.3 %
4 Domain	2.4 %
5 Domain	2.4 %
6 Domain	0.8 %

*Number of chain domains indicated according to expert methods (SCOP, CATH, AUTHORS).

2.3. *The benchmark and dataset*

The performance of program is assessed by a comprehensive benchmark dataset of 124 protein chains, which is based on agreement among expert methods (e.g. CATH, SCOP) and is expanded to include structures with every type of domain combinations.

Previous studies indicate that the majority of problems that algorithms and experts encountered were in the multi-domain structures. To enable a comprehensive analysis of our automatic method, we used a much more improved benchmark dataset for multi-domain structures. The benchmark dataset was employed in the analysis of four automatic methods for domain decomposition: DomainParser, NCBI, PDP and PUU. This benchmark is available in pDomains site (http://pdomains.sdsc.edu/v2/index.php). It includes 124 protein structures in which over 66% of the structures are multi-domain proteins (see Table 1 for the detailed breakdown). Furthermore, the dataset does not have a tendency to each of 2, 3, …, 6-domains subsets and contains chains with smooth distributions in number of domains.

3. Results

3.1. *Dataset properties*

Since "55 chain" dataset[19] and the one introduced by Islam[18] seem to have bias to 1-domain proteins, Balanced_Domain_Benchmark_2 was selected.[20] This benchmark includes proteins in agreement with domain assignment by three expert methods: SCOP,[21] CATH[22] and AUTHORS.[18] Moreover, this benchmark was used by other automatic methods for domain partitioning such as PDP,[11] DomainParser,[6] NCBI,[23] DALI,[24] and PUU.[5] Therefore, the comparison of PUTracer with others can be feasible.

3.2. *Algorithm characteristics*

PUTracer is a top down approach with a recursive algorithm that selects a single cut per recursion according to the specified threshold. The cuts are at least 50 residues far from each other, the beginning and the end of a sequence.

Recursive algorithm provides the possibility to identify more than two domains. The accepting threshold for the best cut is based on differences between ASA and differences between HBE in the whole chain (mother domain) and new domains (child domains). The minimum length of domain to form 3-D structure presumed 50 residues. In both PDP and Domain Parser the minimum length of domain was considered 30 residues while here it was observed that changing the size limitation caused reduction in the number of correct assignments.

3.3. *Evaluation of the PUTracer*

Among 124 chains that were selected to evaluate the performance of the algorithm, 54 chains have single domain, 49 chains have two domains, 14 chains have three domains, 3 chains have four domains, 3 chains have five domains and 1 chain has six domains.

To accept an assignment as a correct one, the accuracy percentage is calculated for each assignment. If the assignment in case of domain number is in common with expert and the accuracy rate is more than 90%, the assignment will be considered correct. The PUTracer algorithm assigned 100 chains correctly (80.65%). The correct set contains 51 single domain proteins, 38 two-domain proteins, 9 three-domain proteins, 1 four-domain protein and 1 five-domain protein (Table 2).

Results in Table 2 show that PUTracer like other automatic methods performs approximately remarkable assignment at accuracy above 90% on single domain chains (94.44%). Although in almost all the automatic methods, the rate of correct assignment dropped as the number of domains increased, our results indicated good correct assignment rate in one, two and even three domain chains.

1qu6 (with accuracy 89.38%) has less accuracy than 90%. Figure 2 illustrates that a decrease in the accuracy of the mentioned assignment is because of cutting in the turn region of a protein structure and does not affect domain construction and truly can be waived.

Twenty-three chain assignments by PUTracer are challenging and could be categorized into two groups. The first group contains two domains (1civ and 1sky) with less than 90% accuracy (70.29% and 57.56% respectively) that are equal in the case of domain number but their domain boundaries are different. Our result for these two chains presented new domain decompositions especially in the case 1sky. In this query, one domain boundary was identified the same by both expert and PUTracer (Fig. 3).

To continue, PUTracer introduced a new domain boundary by cutting on GLU_{133} instead of ALA_{353}. According to our criteria (the minimum value for P_{ASA} and P_{HBE}) cutting ALA_{353} makes high value for τ_{353} in comparison to τ_{133} (i.e. cutting in GLU_{133} produces new domains that are far from each other). So the algorithm logically chose the other cut (GLU_{133}) with 6.5 as τ_{133} value, which are less than threshold (6.70) (Fig. 2 and 3). 1civ domain decomposition is not very different compared to expert. The difference between PUTracer and the expert result in this case could explain as is mentioned for the second cut in 1sky.

Table 2. Results of three automatic methods according to chains query in selected dataset.

		PUTracer				PD P				Domain Parser			
Dom.number	Total	Accuracy < 90 %	Undercut	Overcut	Correct %	Accuracy < 90 %	Undercut	Overcut	Correct %	Accuracy < 90 %	Undercut	Overcut	Correct %
1dom	54	3	0	3	94.44	2	0	2	96.30	1	0	1	98.15
2dom	49	10	5	4	79.59	9	1	5	81.63	11	8	3	77.55
3dom	14	5	2	2	64.29	6	1	2	57.14	3	1	0	78.57
4dom	3	2	1	1	33.33	1	0	0	66.67	2	2	0	33.33
5dom	3	2	1	1	33.33	2	1	0	33.33	3	3	0	0.00
6dom	1	1	1	0	0.00	1	1	0	0.00	1	1	0	0.00
total	124	23	10	11	81.45	21	4	9	83.06	21	15	4	83.06

Accuracy < 90% indicates all chains that have accuracy under 90% compared to expert method (Authors), including chains with under and over cut (zero accuracy), chains with correct number of domains but vary in boundary.

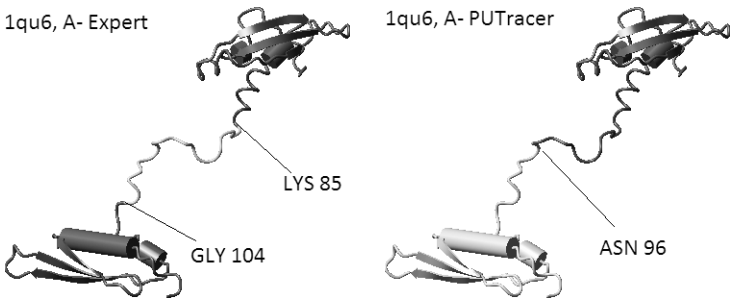

Fig. 2. Chain with accuracy less than 90% by PUTracer (right) but the same number of domain compared to expert (left). Obviously, the figure shows that PUTracer assignment is correct and differences are just due to site of turn\loop that was cut.

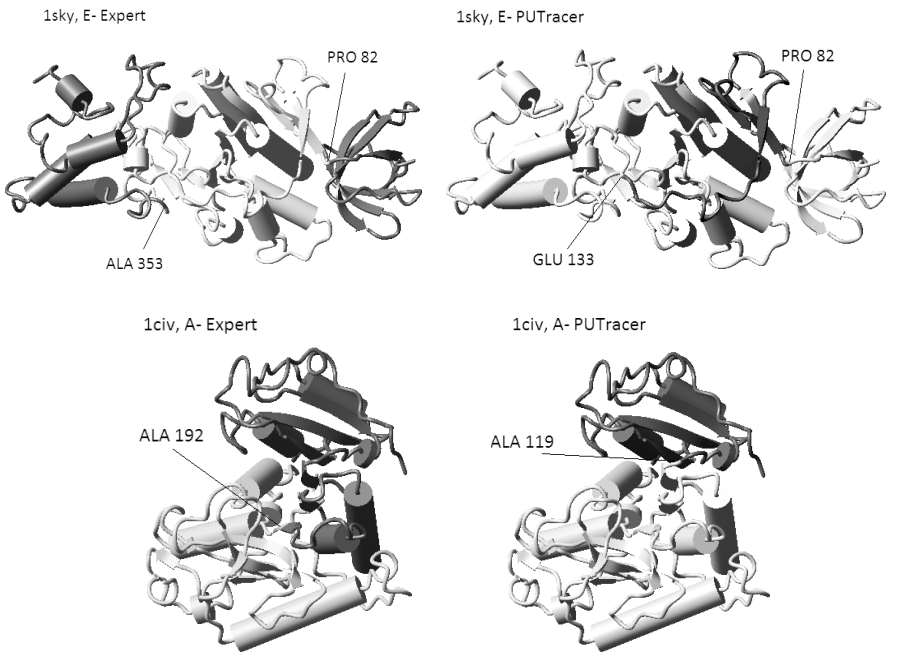

Fig. 3. Chains with accuracy less than 90% by PUTracer (right) but the same number of domain compared to expert (left). 1sky (57.56%) and 1civ (70.29%) take a new assignment by PUTracer.

Details of the result indicate that although cutting in ALA 192 made the least losses of HBE, the separation site in ALA 119 overcame ALA 192 due to the huge change in ASA.

The second group of those 23 chains discussed contains 18 chains with an accuracy less than 90%, which took new assignments by PUTracer and categorized as over-cut/under-cut groups compared with expert. Also comparison with other automatic methods reveals that they cannot assign all of these chains correctly and

Table 3. Eighteen chains with 0% accuracy i.e. new domain assignments according to PUTracer.

Expert assignment	pdb ID	PUTracer	PDP	DomainParser	DALI	NCBI
1-domain	4cp4, A	2*	2[fr]*	1	1	3[fr]*
	1xim, A	2*	2*	1	2*	1
	1wrp, A	2*	1	1	1	1
2-domain	1qnt, A	1*	2	2	2	2
	1ffu, A	1*	2	1*	2	1*
	1qni, A	3*	2	4[fr]*	2[fr]*	4[fr]*
	1whe, A	1*	1*	1*	1*	1*
	1crx, A	3*	3*	2	2	2
	1bpm, A	3*	2	2	2[fr]*	2
	1dg3, A	3*	5[fr]*	2	4[fr]*	2
	1yua, A	1*	2	1*	1*	2
	3sxl, A	1*	2	2	2	2<*
3-domain	1d0g, T	1*	2*	2*	1*	2*
	1igr, A	4*	4*	3	3	4*
	1bhg, A	4*	3[fr]*	3[fr]*	3[fr]*	3
	1hjp, A	2*	3	3	3	3
4-domain	1wgt, A	2*	4	4	4	4
	1qba, A	5*	4[fr]*	3[fr]*	4[fr]*	4[fr]*
5-domain	1cwv, A	4*	5	4*	5	5
	2gli, A	2*	1*	3*	2*	3*
6-domain	1bxr, C	9*	2[fr]*	5[fr]*	2[fr]*	8[fr]*

This Table compares different automatic method results in the case of chains that assigned as new domain partitioning by PUTracer against expert.
*Shows domain number of chains that were assigned incorrectly vs. expert. [fr]: contain fragmented domain.

even consider them fragmented domains in some cases (Table 3). Among these 21 chains, PUTracer assigned four chains (1wgt, 2gli, 1d0g, 1whe) differently from experts due to the size limitation considered for a domain.

It should also be noted that some gaps between domains (not recognized residues) are found in expert assignment. For instance, PUTracer parsed 1igr chain to four domains (1−183; 184−246; 247−298; 299−478) while other methods assigned it as three domains (1−184; 224−299; 300−478). The region between MET 184 to TYR 224 that is neglected by expert shows a new domain as seen in Fig. 4. It is clear that the difference between PUTracer and expert can be ignored in this case.

Another group of four chains (1xim, 1wrp, 1bpm, 1crx) can elicit among this 23 with a different assignment (Fig. 6). In these cases, DomainParser assigns all of them in the same way as expert but PDP agrees with PUTracer in the domain number of 1xim and 1crx. Being affected by structural configuration and accessible surface area and HBE may cause these differences.

As a further comparison, the chains assigned incorrectly are inspected by two automatic methods (PDP and DomainParser) while PUTracer does the assignments well. This argument demonstrates that PUTracer is able to identify all of the chains introduced in Table 4 the same as expert (with more than 90% accuracy).

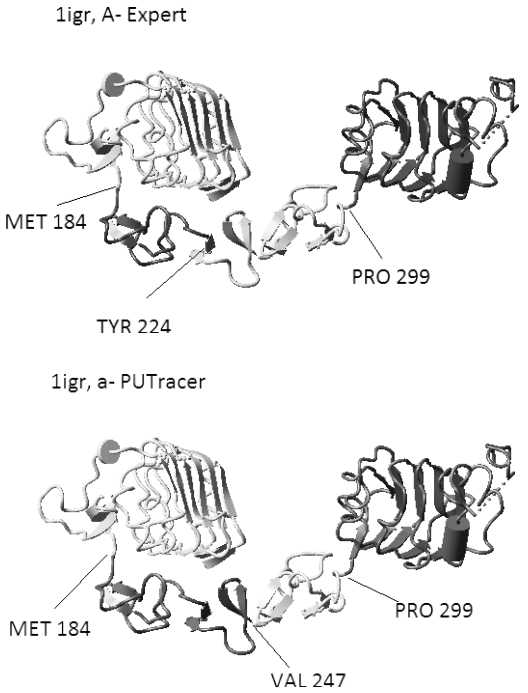

Fig. 4. Chain A of 1igr. Expert assignment (upper) and PUTracer assignment (bottom).

Table 4. Incorrectly assigned chains by automatic method (PDP and DomainParser) while correctly assigned by PUTracer.

Expert assignment &PUTracer	pdb ID, Chain	PDP	Domain parser
1dom	1myt, A	1	2*
2dom	1au7, A	2	1*
	1aua, A	2[fr]*	1*
	1bc5, A	3*	2
	1ffh, A	2[fr]*	1*
	1fbl, A	2	3*
	1bbw, A	3[fr]*	2
	1fmt, A	3*	3*
	1ega, A	2	1*
3dom	1djz, A	3[fr]*	3
	1ksi, A	4*	3
	2shp, A	3[fr]*	3
4dom	1cs6, A	4	1*
5dom	1fnm, A	5[fr]*	4*

*domain number of chains that were assigned incorrectly vs. expert. [fr]: contain fragmented domain.

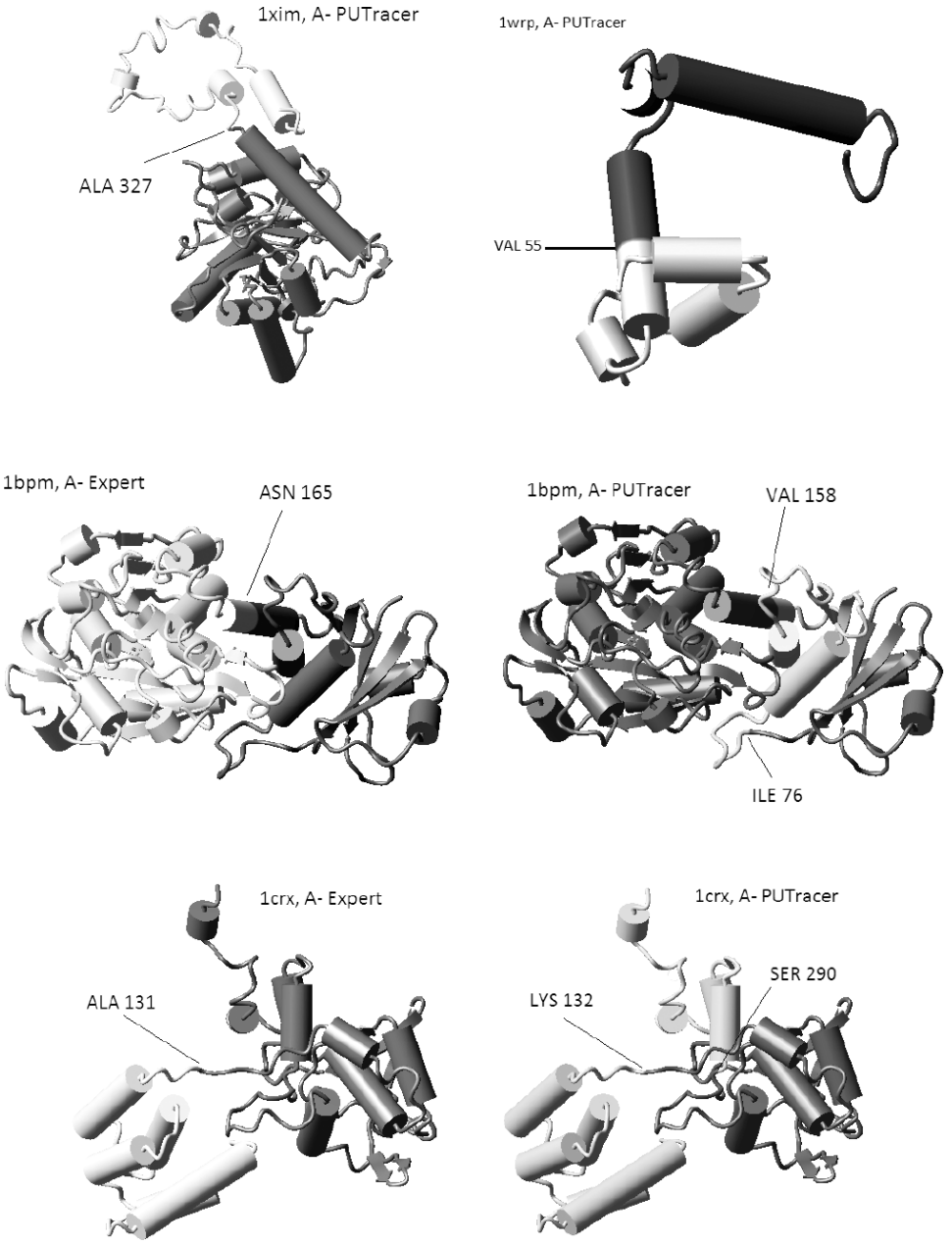

Fig. 5. New assignments by PUTracer (1xim, 1wrp as two domains, and 1bpm, 1crx as three domains), while they are assigned one and two domains respectively by expert.

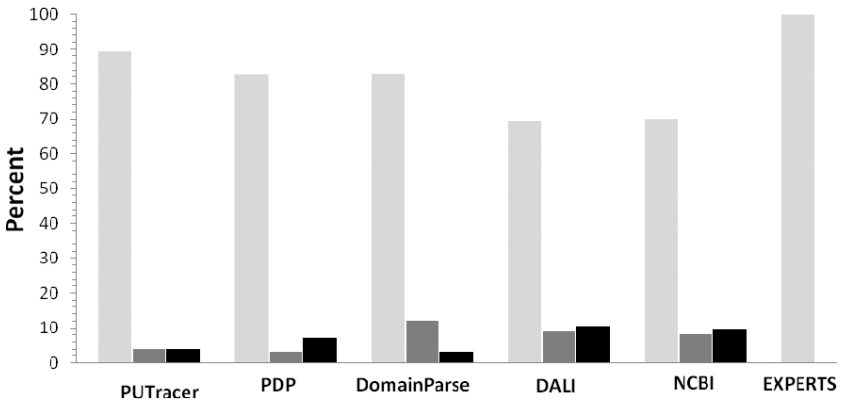

Fig. 6. The percentage of correct, under and over cut assignments. All three expert methods have consensus on this dataset except 2pf2. SCOP shows it as two domains with a cut in 65th residue (http://pdomains.sdsc.edu). Pale gray: percent of correct assignment. Bold gray: percent of under-cut. Black: percent of over-cut.

As previously mentioned, some chains from benchmark_2 are used that they are already assigned by expert methods as contiguous domains. Thus, as expected, there is no fragmentation in domain assignment by PUTracer while as indicated in Table 3 and 4, PDP and DomainParser show a tendency for fragmenting in some cases.

However, our primary results indicat a robust performance of PUTracer with 81.45% correct assignment with very small bias to over-cuts (8.06 and 8.87% for under and over-cuts respectively) while PDP and DomainParse show a tendency to over and under-cut, respectively. Investigating incorrectly assigned ones proves that 10 chains (1civ, 1xim, 1wrp, 1bpm, 1crx, 1wgt, 2gli, 1d0g, 1whe, 1igr) can be accepted as new correct parsing. With this regard four under-cuts and five over-cuts can be ignored, too (i.e. reduce to near 4%). Thus, the percentage of the correct assignment will be increased to 89.5%. Figure 6 illustrates the final results in comparison to other automatic methods.

4. Discussion

In this paper, we present PUTracer (Protein Unit Tracer) as a new and high performance algorithm for protein domain assignment according to inter and intra-domain interaction based on differences of accessible surface area and the hydrogen bonds energy in new child domains in comparison to their mother domain. Although nearly 80% of the cases could be assigned correctly by automatic approaches, drastic increase in the number of solved structures makes a necessity for rapid and automated methods. With this aim, PUTracer exhibits more than 90% efficiency with negligible bias in identifying contiguous domains in 124 chains of balanced domain benchmark_2. Moreover, our method has a potential in identifying two domain chains compared to other automatic methods (i.e. 81.63% after addition of 1civ) and

our mentioned results present new domain assignments for some chains such as 1xim, 1wrp, 1bpm and 1crx (Fig. 5).

Our algorithm is based on the simple concept of domain as a spatially distinct unit of proteins with independent folding, of which visual inspection of the results proved this notion. While almost all the previous methods that applied the concept of inter/intra domain interaction in domain assignment[5,6,11,12,23,25–28] emphasized Van der Waals interactions, considering the hydrogen bond energies in backbone atoms in domain recognition is a rarely used strategy.

Although the hydrogen bonds in the surface of protein may cause destabilization of protein structure due to change in environmental conditions such as pH,[29,30] in the backbone of protein (especially in protein core), the hydrogen bonds play a non-negligible role in secondary and tertiary structure formation and stabilization, especially when they are in a non-polar area inside the protein construction, where the dielectric coefficient is minimum and the electrostatic force increases. However, because expert methods (SCOP and CATH) turned out to be in agreement with the applied dataset, PUTracer relied on evolutionary and functional characteristics of domains.

Among 23 incorrectly assigned chains by PUTracer (Table 3), the latter ones can be discussed as acceptable decomposition in reference to PDP and DomainParser methods. PUTracer outputs for 1xim, 1igr and 1whe agree with PDP and DomainParser results respectively. However, we consider a domain with at least 50 residues in PUTracer; a simple comparison between these three methods indicates 89.5% correct assignments for PUTracer while PDP and DomainParser assigned 85.48% and 83.87% of chains correctly, respectively. In case of the under and over cuts, PUTracer by about 4% for each one, shows no bias to no one, while PDP and DomainParser show 6.45% and 11.29% for over-cuts and under-cuts respectively.

Both logics applied in DomainParser (graph based) and PDP (spatially contact based) are used in PUTracer. Accordingly, great role is hypothesized for changing accessible surface area and hydrogen bond energies in backbone atoms in parsing a protein into its domains. Obviously, by increasing chain length and structural complexity the correctness drops as seen in 1bxr. And as an end point, we are going to make improvements in our algorithm in the near future for a major defect due to the inability of non-contiguous domains assignment.

Acknowledgment

This research was in part supported by a grant from IPM (No. CS1385-102).

References

1. Taylor WR, Protein structural domain identification, *Protein Engineering* **12**(3): 203–216, 1999.
2. Emmert-Streib F, Mushegian A, A topological algorithm for identification of structural domains of proteins, *BMC Bioinformatics* **8**:237, 2007.

3. Tai CH, Sam V, Gibrat JF, Garnier J, Munson PJ, Lee B, Protein domain assignment from the recurrence of locally similar structures, *Proteins* **79**(3):853−866, 2011.
4. Gerstein M, Lesk AM, Chothia C, Structural mechanisms for domain movements in proteins, *Biochemistry* **33**(22):6739−6749, 1994.
5. Holm L, Sander C, Parser for protein folding units, *Most* **268**:256−268, 1994.
6. Xu Y, Xu D, Gabow HN, Protein domain decomposition using a graph-theoretic approach, *Bioinformatics* **16**(12):1091−1104, 2000.
7. Stepanova M, Dynamics of essential collective motions in proteins: Theory, *Phys Rev E Stat Nonlin Soft Matter Phys* **76**(5 Pt 1):051918, 2007.
8. Wetlaufer DB, Rapid folding, and globular intrachain, *Proc Natl Acad Sci USA* **70**(3):697−701, 1973.
9. Rose GD, Hierarchic organization of domains in globular proteins, *J Mol Biol* **134**(3):447−470, 1979.
10. Holm L, Sander C, Searching protein structure databases has come of age, *Proteins* **19**(3):165−173, 1994.
11. Alexandrov N, Shindyalov I, PDP: Protein domain parser, *Bioinformatics* **19**(3):429−430, 2003.
12. Wernisch L, Hunting M, Wodak SJ, Identification of structural domains in proteins by a graph heuristic, *Proteins* **35**(3):338−352, 1999.
13. Wernisch L, Wodak SJ, Identifying structural domains in proteins, *Methods Biochem Anal* **44**:365−385, 2003.
14. Holm L, Park J, DaliLite workbench for protein structure comparison, *Bioinformatics* **16**(6):566−567, 2000.
15. Jones S, Stewart M, Michie A, Swindells MB, Orengo C, Thornton JM, Domain assignment for protein structures using a consensus approach: Characterization and analysis, *Protein science: A Publication of the Protein Society* **7**(2):233−242, 1998.
16. Veretnik S, Gu J, Wodak SJ, Identifying structural domains in proteins, *Structural Bioinformatics* 487−516, 2009.
17. Kabsch W, Sander C, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* **22**(12):2577−2637, 1983.
18. Islam SA, Luo J, Sternberg MJ, Identification and analysis of domains in proteins, *Protein Eng* **8**(6):513−525, 1995.
19. Jones S, Stewart M, Michie A, Swindells MB, Orengo C, Thornton JM, Domain assignment for protein structures using a consensus approach: Characterization and analysis, *Protein Science* **7**:233−242, 1998.
20. Holland Ta, Veretnik S, Shindyalov IN, Bourne PE, Partitioning protein structures into domains: Why is it so difficult? *J Mol Biol* **361**(3):562−590, 2006.
21. 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* 536−540, 1995.
22. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM, CATH — a hierarchic classification of protein domain structures, *Structure* **5**(8):1093−1108, 1997.
23. Madej T, Gibrat J-F, Bryant SH, Threading a database of protein cores, *Library* **369**: 356−369, 1995.
24. Holm L, Sander C, Protein structure comparison by alignment of distance matrices, *J Mol Biol* 123−138, 1993.
25. Swindells MB, A procedure for detecting structural domains in proteins, *Protein Sci* **4**(1):103−112, 1995.
26. Wodak SJ, Janin J, Structural aspects of recognition and assembly in biological macromolecules, *Nature* **285**(5763):287−288, 1980.

27. Siddiqui AS, Barton GJ, Continuous and discontinuous domains: An algorithm for the automatic generation of reliable protein domain definitions, *Protein Sci* **4**(5):872−884, 1995.
28. Zhou H, Xue B, Zhou Y, DDOMAIN: Dividing structures into domains using a normalized domain-domain interaction profile, *Protein Sci* **16**(5):947−955, 2007.
29. Hendsch ZS, Tidor B, Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Science* 211−226, 1994.
30. Koczyk G, Berezovsky IN, Domain Hierarchy and closed Loops (DHcL):A server for exploring hierarchy of protein domain structure, *Nucleic Acids Res* **36**(Web Server issue): W239−245, 2008.

Seyed Shahriar Arab is an assistant professor in the department of Biophysics, Tarbiat Modares University. He completed his Ph.D. degree in Bioinformatics at the University of Tehran. He also holds a masters degree in Biophysics from Tarbiat Modares University. He is also a member of Bioinformatics group at Institute for Research in Fundamental Science (IPM). His research interests are Bioinformatics, especially protein structure prediction, classification and protein design.

Mohammadbagher Parsa Gharamaleki is a research fellow at University of Wisconsin Madison. Prior to joining the Department of Mathematics in University of Wisconsin in 2012, he was working at Institute for Fundamental Science (IPM) and Institute for Medical Plants (IMP), and he has two years of experience in bioinformatics and plant genomics. He graduated from Isfahan University of Technology with Master's degree in plant breeding. Systems biology and computational genomics are the central area of his research interests; his efforts led to a number of publications in these fields.

Zaiddodine Pashandi received his M.Sc. degree (September 2012) in Biophysics from Tarbiat Modares University (TMU), Iran, and is currently a Ph.D. Student in Biophysics at TMU. He has a B.E. degree in Applied Chemistry from Ferdowsi University of Mashhad, Iran (September 2010). His research interests include Bioinformatics, Proteomics and Bioelectronics.

Rezvan Mobasseri is a Ph.D. student in Nanobiotechnology at Tarbiat Modares University of Iran. Her research interest lies in Nanobiotechnolgy and Nanolithography as well as bioinformatics tools and drug design. She obtained her master's degree in Biophysics at Tarbiat Modares University (2012) and joined the TMU Bioinformatics group since 2011.