Interacting amino acid preferences of 3D pattern pairs at the binding sites of transient and obligate protein complexes

Suryani Lukman, Kelvin Sim

Institute for Infocomm Research,
21 Heng Mui Keng Terrace, Singapore 119613
E-mail: slukman, shsim@i2r.a-star.edu.sg

Jinyan Li

School of Computer Engineering,
Nanyang Technological University, Singapore 639798
E-mail: JYLi@ntu.edu.sg

Yi-Ping Phoebe Chen

Faculty of Science and Technology,
Deakin University, Australia
E-mail: phoebe@deakin.edu.au

To assess the physico-chemical characteristics of protein-protein interactions, protein sequences and overall structural folds have been analyzed previously. To highlight this, discovery and examination of amino acid patterns at the binding sites defined by structural proximity in 3-dimensional (3D) space are essential. In this paper, we investigate the interacting preferences of *3D pattern pairs* discovered separately in transient and obligate protein complexes. These 3D pattern pairs are not necessarily sequence-consecutive, but each residue in two groups of amino acids from two proteins in a complex is within certain Å threshold to most residues in the other group. We develop an algorithm called *AA-pairs* by which every pair of interacting proteins is represented as a bipartite graph, and it discovers all *maximal quasi-bicliques* from every bipartite graph to form our 3D pattern pairs. From 112 and 2533 highly conserved 3D pattern pairs discovered in the transient and obligate complexes respectively, we observe that Ala and Leu is the highest occuring amino acid in interacting 3D patterns of transient (20.91%) and obligate (33.82%) complexes respectively. From the study on the dipeptide composition on each side of interacting 3D pattern pairs, dipeptides Ala-Ala and Ala-Leu are popular in 3D patterns of both transient and obligate complexes. The interactions between amino acids with large hydrophobicity difference are present more in the transient than in the obligate complexes. On contrary, in obligate complexes, interactions between hydrophobic residues account for the top 5 most occuring amino acid pairings.

Keywords: bipartite graph; amino acid preferences; pattern pairs; transient complexes; obligate complexes

1. Introduction

Amino acid interactions are fundamental for protein-protein interactions. These protein-protein interactions are important in all facets of life, from metabolism to disease fighting. Since different amino acids possess distinct functional groups, a preliminary step towards

2

understanding protein-protein interactions is to analyze the pairing preferences of amino acids at the binding sites of distinct protein complexes. Previous studies to determine interacting amino acid preferences are of various conclusions. Some studies report that the amino acid composition of interfaces of distinct protein complexes are similar,[7,9] whereas others report significant differences.[5,10,11] Furthermore, some groups discovered that polar and charged residues are the major contributors in protein-protein interactions,[14,18] whereas other reported that hydrophobic interactions are favoured.[19] Therefore, we introduce the concept of interacting 3D pattern pairs, defined by spatial proximity, to understand the interacting preferences of conserved amino acids involved in distinct protein complexes. We focus our study on two types of protein complexes: transient and obligate.

Individual proteins are capable of adopting their native fold as monomers and they may interact transiently upon a molecular stimulus, to fulfill a particular function and dissociate after that. This type of protein complexes is termed as transient complex. Another type of protein complexes is obligate complex, in which the protein chains remain bounded to each other throughout their functional lifetime. These proteins may not fold and function properly when they are unbound. A classical example of obligate complex is the interaction between the β- and γ- subunits of heterotrimeric G proteins, whereas the α-subunit forms a transient interaction with the β- and γ-subunits. Since the transient and obligate interactions are characterized by distinct physico-chemical characteristics,[16] it is crucial to distinguish between these two kinds of interactions when analyzing the interacting preferences of amino acid pattern pairs.

We propose a graph-based approach to discover interacting 3D pattern pairs efficiently. We represent a pair of interacting protein chains using a *bipartite* graph[6] based on their residues' 3D-coordinate distance information. We discover *maximal quasi-biclique* subgraphs[15] from every bipartite graph. We then mine across the maximal quasi-bicliques to obtain quasi-bicliques that are significantly frequent and large, and each of them corresponds to a 3D pattern pair. We choose to discover maximal quasi-bicliques, instead of the classical *maximal bicliques*[6] because maximal quasi-bicliques are more tolerant to missing data.

2. *AAPAIRS*: Our algorithm to discover 3D pattern pairs

We present an algorithm to discover **A**mino **A**cid 3D pattern **pairs**, called *AApairs* (Algorithm 2.1). A preprocessing step of AApairs is to classify a pair of interacting polypeptide chains into one of three different classes: crystal packing, transient or obligate complex. Crystal packing is excluded from our further consideration. The first step of our *AApairs* algorithm finds a special type of subgraph—maximal quasi-biclique[15]—from a pair of interacting polypeptide chains which we represent as a bipartite graph in our implementation. At the second step, *closed patterns*[8,13] across the maximal quasi-biclique subgraphs are detected. If a pair of closed patterns can form a quasi-biclique subgraph, and the pair occurs frequently in many pairs of interacting proteins, then we call such a pair a 3D pattern pair, which is of our interest. Our experiments were conducted in Windows XP environment, using Intel Xeon CPU 3.4GHz with 2GB RAM. *AApairs* was implemented in C++.

2.1. *Classifying pairs of interacting proteins into obligate or transient complexes*

This preprocessing step deals with all X-ray crystallographic protein structures having resolution better than 2.5Å from the Protein Data Bank (PDB, http://www.rcsb.org). We do not consider Nuclear Magnetic Resonance (NMR)-determined protein structures and any of the nucleic acids in PDB. We consider only PDB entries with two or more polypeptide chains. Given such an entry, we use the NOXclass method,[20] a support-vector machine prediction method, to remove biologically-irrelevant crystal packing interaction between any two polypeptides, and then to classify those remaining biologically-relevant interactions into either transient or obligate interactions. Using interface properties such as interface area, ratio of interface area to protein surface area, and amino acid composition of the interface as input, NOXclass was reported to be highly accurate, achieving an accuracy of 91.8% for the classification of those interactions.[20]

2.2. *AApair (step one): Discovering maximal quasi-biclique subgraphs from transient and obligate complexes*

We represent a pair of transient or obligate polypeptide chains as an undirected bipartite graph. An undirected bipartite graph G is a graph consisting of two disjoint vertex sets V_1 and V_2 satisfying the condition that there is no edge between any two vertices within V_1 or within V_2. Such a graph is usually denoted as $G = (V_1, V_2, E)$ where E is the set of edges of G. A polypetide chain can be mathematically represented as a set of amino acid residues (with location information). Thus, to transform a pair of polypeptide chains into a bipartite graph, we just represent every residue as a vertex, and we assign an edge between a residue x_1 in one chain and a residue x_2 in the other chain if and only if there exists at least a pair of atoms between x_1 and x_2 whose distance is less than a threshold. In this study, we use a threshold of 5.0Å.[17]

After constructing a bipartite graph $G = (V_1, V_2, E)$ representing a pair of interacting polypetide chains, we discover the complete set of maximal quasi-biclique subgraphs from G. A *quasi-biclique* H of G is a subgraph consisting of two sets of vertices $X_1 \subseteq V_1$ and $X_2 \subseteq V_2$ such that every vertex in X_i, $i = 1, 2$, is adjacent to at least $|X_j| - \epsilon$, $j \neq i$, vertices in X_j. The tolerance rate ϵ is a small integer number, e.g. 1 or 2, defined by users. A quasi-biclique subgraph H is maximal in G if there is no other quasi-biclique in G that contains H. We use our $CompleteQB$ algorithm[15] to discover maximal quasi-bicliques. The $CompleteQB$ algorithm has another user input parameter ms—the minimum number of vertices in each side of a maximal quasi-biclique. That is, only those maximal quasi-bicliques, whose vertex set size is at least ms, are enumerated. Therefore, by mining maximal quasi-biclique subgraphs, we can discover pairs of closely interacting residues from a pair of interacting polypeptide chains. We note that the residues in one side of a maximal quasi-biclique are not necessarily consecutive in one chain.

The above procedure is performed for all possible pairs of interacting (transient or obligate) polypetide chains within a PDB entry. Thus, after going through all PDB entries, we can obtain many maximal quasi-bicliques representing pairs of closely interacting residues.

4

2.3. *AApair (step two): Identifying significant 3D pattern pairs*

As some maximal quasi-bicliques may occur in interacting polypetide chains by chance, we identify those quasi-bicliques that occur in PDB entries with a high frequency. Let m be the number of all pairs of interacting polypetide chains from all PDB entries used in this study, and let CHAINPAIRS represents all these pairs. We then denote CHAINPAIRS as $\{chainPair^{(i)} \mid i = 1, 2, \cdots, m\}$, where $chainPair^{(i)} = (C_1^{(i)}, C_2^{(i)})$, $C_1^{(i)}$ or $C_2^{(i)}$ represents the set of amino acid residues in one of the two chains. Let n be the number of all maximal quasi-bicliques discovered from $\{chainPair^{(i)} \mid i = 1, 2, \cdots, m\}$, after transforming every $chainPair^{(i)} = (C_1^{(i)}, C_2^{(i)})$ into a bipartite graph $G^{(i)} = (V_1^{(i)}, V_2^{(i)}, E^{(i)})$. A maximal quasi-biclique is denoted as $H^{(j)} = (X_1^{(j)}, X_2^{(j)})$, $j = 1, 2, \cdots, n$. We then discover frequent closed patterns from $segmentDB$ to construct our desired patterns, where

$$segmentDB = \{X_1^{(1)}, X_2^{(1)}, X_1^{(2)}, X_2^{(2)}, \cdots, X_1^{(n)}, X_2^{(n)}\}$$

A closed pattern is the maximal pattern among patterns that occur in the same set of objects. For example, abc is the closed pattern occurring in $abcde$ and $abcfg$, but ab is not. Suppose there are k number of closed patterns of $segmentDB$, denoted as $\{P_1, P_2, \cdots, P_k\}$. We then pairwise $P_1, P_2, \cdots$, and P_k, and for every pair (P_u, P_v), we go through $\{chainPair^{(i)} \mid i = 1, 2, \cdots, m\}$ to count the number of them containing (P_u, P_v). If the number exceeds a pre-defined threshold sup, then (P_u, P_v) is a significant interacting 3D pattern pair.

Formally,

Definition 2.1. (Interacting 3D pattern pair). A pair of closed patterns P and Q forms a 3D pattern pair (P, Q) if and only if

- $|P| \geq ms$ and $|Q| \geq ms$, as specified in Step one;
- The occurrence number in CHAINPAIRS exceeds sup, as specified in Step two;
- P and Q can form a quasi-biclique.

3. Results

From 17,630 X-ray crystallographic protein complexes data in the Protein Data Bank (PDB, http://www.rcsb.org), we collect 4,661 transient and 7,985 obligate interactions. Only polypeptide chains containing >30 amino acids are considered in our analysis. Similar sequences at 90% identity were removed. To ensure that our results are supported with experimental evidence, we consider only interactions between two polypeptide chains found in a single PDB entry.

Our $AApair$ algorithm allows users to set three parameter values prior to discovering interacting 3D pattern pairs. The three parameters are: the minimum number of residues in one side of pattern pairs (ms), the minimum occurence in the pairs of interacting proteins (sup), and the error tolerance rate (ϵ) of the maximal quasi-bicliques.

By varying the three parameters, we obtain different numbers of 3D pattern pairs. We observe that ms set to 3 or 4 is ideal because the average number of residues in one side

---

Algorithm 2.1 Algorithm $AApairs$

Input:

Set of pairs of interacting polypeptide chains, $ppDB$

ms is the minimum size threshold; ϵ is the error tolerance rate;

sup is the minimum occurrence;

Description:

1: use $NOXclass$ to classify all pairs of interacting polypeptide chains in $ppDB$. Let the set of qualified pairs of interacting polypeptide chains be CHAINPAIRS $= \{chainPair^{(i)}, i = 1, 2, \ldots, m\}$;

2: convert $\{chainPair^{(i)}, i = 1, 2, \ldots, m\}$ into a set of bipartite graphs $\{G^{(i)}, i = 1, 2, \ldots, m\}$;

3: use $CompleteQB$ to mine maximal quasi-biclique subgraphs from every $G^{(i)}$. Let the set of maximal quasi-biclique subgraphs be $\{H^{(j)}, j = 1, 2, \ldots, n\}$, where $H^{(j)} = (X_1^{(j)}, X_2^{(j)})$;

4: $segmentDB = \{X_1^{(1)}, X_2^{(1)}, X_1^{(2)}, X_2^{(2)}, \cdots, X_1^{(n)}, X_2^{(n)}\}$

5: use $MineLMBC$ on $segmentDB$ to mine a set of closed patterns $\{P_1, P_2, \cdots, P_k\}$;

6: **for all** $P_i, P_j \in \{P_1, P_2, \cdots, P_k\}$ **do**

7:    $count = 0$;

8:    **for all** $chainPair^{(x)} \in$ CHAINPAIRS **do**

9:        **if** P_i and $P_j \in chainPair^{(x)}$ **then** $count$++;

10:    **if** $count \geq sup$ **then** (P_i, P_j) forms a 3D pattern pair;

11: output all 3D pattern pairs;

---

of the pattern pairs mined by maximal quasi-bicliques are 3.60 and 5.14 for transient and obligate complexes respectively. When ϵ is set to 1, ms set to 3, sup set to 100, we discover 112 and 2,533 3D pattern pairs from the transient and obligate interaction datasets respectively. When ϵ is set to 0, while the two other parameters remain the same, only one 3D pattern pair is discovered each from the transient and obligate datasets.

The introduction of error tolerance rate in maximal-quasi bicliques, allows reasonable numbers of 3D pattern pairs to be found across different sup settings, up to sup=100. This also solves the challenges faced by maximal biclique technique in which very few or even zero 3D pattern pairs can be found when sup is high. Mining maximal quasi-bicliques is appropriate as not all existing structural data are complete. Even in a complete data of protein complexes, we cannot expect that the inter-protein residual interactions have a perfect all-versus-all relationship as represented by maximal biclique subgraphs. In addition, the ϵ parameter also accommodates the rapidly growing structural data.

The statistic of the numbers of 3D pattern pairs discovered by varying the parameters is reported in `http://research.i2r.a-star.edu.sg/AApairs`. A note of caution for setting the parameters is essential as parameters set too low (e.g. sup) may result in very big numbers of 3D pattern pairs.

3.1. *Amino acid distributions on each side of 3D pattern pairs*

We consider average amino acid composition in percentage for the complete database in release 53.0 of UniProtKB/Swiss-Prot (`http://cn.expasy.org/sprot/relnotes/`). If the percentage of a particular amino acid in our 3D pattern pairs is much greater than its percentage in Swiss-Prot database, the amino acid is likely to play important role in protein-protein interactions. Amino acids such as Leu, Ala, and Gly have high compositions in protein sequences, whereas amino acids such as Cys and Trp have

6

low compositions. We study the amino acid distributions in 3D patterns found within both transient and obligate complexes and compare them with the amino acid distributions in protein sequences (Figure 1). Leu is the highest occuring amino acid in interacting 3D patterns of obligate complexes, accounting for 33.82% as compared to 9.66% in protein sequences. Ala, a less hydrophobic amino acid than Leu, is the highest and second highest occuring amino acid in interacting 3D patterns of transient (20.91%) and obligate complexes (16.78%) respectively. Ala was reported to have high α helix-forming tendency and two α helices can wrap around one another to form a coiled-coil as a mechanism for protein-protein interaction.[1] Though Ala side chain is very non-reactive, it can play a role in substrate recognition or specificity.[3]

Except for Ala whose presence in 3D patterns of transient complexes shows a significant increase, the frequencies of hydrophobic amino acids (Ile, Val, Leu, Phe, and Met) in 3D patterns of transient complexes are generally lower than those in overall protein sequences. On the other hand, more polar amino acids, especially the charged ones (His, Glu, Asp, Lys, Arg) occur less frequently in interacting 3D patterns of obligate complexes. Only 0.06% of amino acids in 3D patterns of obligate complexes are His. For the other two positively-charged residues, Arg occurs more often than Lys does in the 3D patttern pairs of both transient (6.06% versus 0.89%) and obligate (2.16% versus 0.33%) complexes. This is in agreement with previous studies.[12]
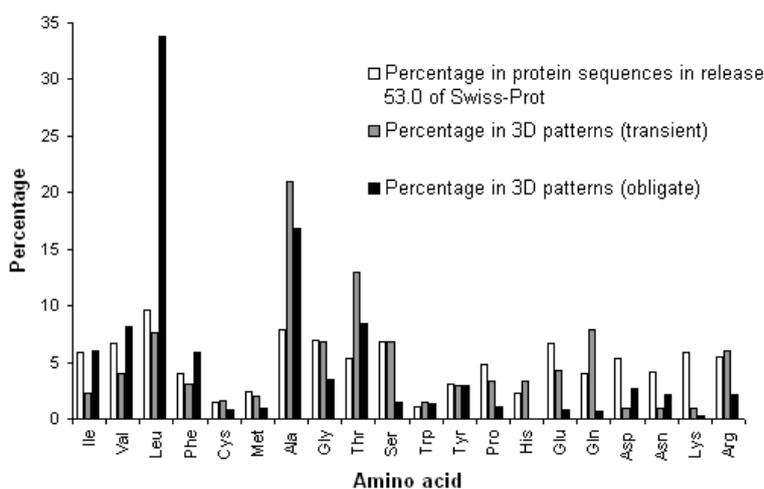

Fig. 1.   The comparison of amino acid distributions among protein sequences, 3D patterns in transient and those in obligate protein complexes. The amino acid residues are ordered according to their hydrophobicity, with Ile as the most hydrophobic and Arg as the least hydrophobic.

We also calculate the frequencies of all 210 possible dipeptides in each side of interacting 3D pattern pairs of both transient and obligate complexes. As we exclude sequence-consecutiveness, we consider dipeptide AT and TA as the same, and any occurence of TA is added to the number of occurence of AT. We present the top 10 dipeptides to compare the

3D patterns on the interfaces of the two distinct protein-protein interactions (Table 1). Only AA and AL are within the top 10 dipeptides of 3D patterns of both transient and obligate complexes. The transient dipeptides contain more combinations of hydrophobic and polar amino acids (AT, AQ, GT, AS, GS, and AE) whereas such combinations are less observed in the obligate dipeptides, in which only LT and DL are highly present.

Table 1.   The list of the top 10 highest occuring dipeptides in 3D patterns of transient and obligate protein-protein interactions.

Dipeptide	Transient (%)	Dipeptide	Obligate (%)
AT	6.63	LL	9.59
AQ	6.51	AL	8.84
AA	4.25	LT	7.73
TT	4.19	LV	6.34
GT	4.04	FL	6.26
AL	3.71	IL	5.74
AS	2.89	AA	4.20
GS	2.20	AV	3.91
AE	2.13	DL	3.16
AG	1.90	GL	2.38

Table 2.   The distributions of dipeptides comprising charged residues in 3D patterns of transient and obligate complexes.

Dipeptide containing residues of same charges	Transient (%)	Obligate (%)	Dipeptide containing residues of opposite charges	Transient (%)	Obligate (%)
DD	0	0	DH	0	1.37
DE	0	1.58	DK	0	1.38
EE	18.47	0	DR	15.36	23.31
HH	0	0.69	EH	3.50	4.19
HK	13.45	0	EK	11.05	0.70
HR	28.09	1.97	ER	6.51	11.52
KK	0	0			
KR	3.56	2.76			
RR	0	50.53			

Note: The percentage of each dipeptide containing charged side chains is calculated from the 3D patterns found separately in the transient and obligate complexes.

The clustering of neighboring polar amino acid side chains can alter their reactivity.[1] For example, if a number of negatively charged side chains are forced together against their mutual repulsion in a particular site by the way of protein folding, the affinity of this site for a positively charged ion is greatly increases. We investigate if such phenomenon is present in our 3D pattern pairs (Table 2). We observe that 50.53% of contributing charged dipeptides in 3D patterns of obligate complexes is RR, which seems to support the hypothesis of clustering of same charged amino acids. However, caution has to be exercised while interpreting this result as Arg is one of the three abundant hotspot residues in binding energy

8

for protein interfaces.[4] To support this, DR and ER dipeptides are also highly present in the 3D patterns of obligate complexes.

In 3D patterns of transient complexes, dipeptides HR and EE are highly present. The high percentage of EE in 3D patterns of transient complexes compared to those in obligate ones, suggests the role of EE as characteristic dipeptide of 3D patterns in transient complexes.

3.2. *Amino acid pairing preferences of 3D pattern pairs*

To reveal residue pairing preferences across various protein-protein interactions, we study the pairing preferences of amino acids in our 3D pattern pairs (Figure 2). In transient complexes, the interactions between residues of large hydrophobic difference are observed more often than those in obligate complexes. The pairings between Gly and Thr, Ala and Ser, and Ala and Glu (4.85%, 4.47%, and 3.90% respectively) are among the top 10 pairing preferences of interacting 3D pattern pairs of transient complexes. On contrary, no pairings between Gly and Thr, Ala and Glu, and only 1.14% of pairing is observed for Ala and Ser.

In obligate complexes, there is a high occurrence of interactions between identical amino acids. The interactions between all identical amino acids, especially the hydrophobic residues, account for 29.48% of all possible residual interactions. Interactions between hydrophobic residues, such as Ile, Val, Leu, and Ala, occur much more often than those between polar residues. In particular, interactions involving Leu are highly common. The pairings between Leu and Leu, Ala and Leu, Leu and Thr, Leu and Val, which are 14.6%, 6.92%, 6.56% and 4.99% respectively, are among the top 5 pairing preferences of interacting 3D pattern pairs of obligate complexes.

Covalent interactions such as the Cys-Cys disulphide bonds are also observed, though they are uncommon. Only six Cys-Cys pairings are present in 3D pattern pairs of obligate complexes. The sulphur atoms of the two Cys residues from the interacting proteins form a disulphide bond if they are at most 2.0Å apart. Since disulphide bonds are more rigid and stable than ionic and van der Waals interactions, it is not surprising to detect such interactions only in 3D pattern pairs of obligate complexes, but not in those of transient ones.

4. Discussion

To draw veritable observations and knowledge from the available structural data, it is essential to analyze as many protein-protein interactions as possible. Our study addresses this requirement by analyzing 12,646 interactions, which is further classified into either transient or obligate interactions. Inclusion of large dataset allows for distinguishing transient and obligate complexes, which cannot be achieved by most studies using small datasets.[12]

3D pattern pairs can be used as building blocks for the model building of protein complexes in crystallography. They are also applicable to automated structure prediction and determination of protein complexes based on crystallography data. Furthermore, 3D pattern pairs can facilitate the incremental acquisition and indexing of structural data of protein complexes into knowledge bases, which can be organized based on substructural similar-
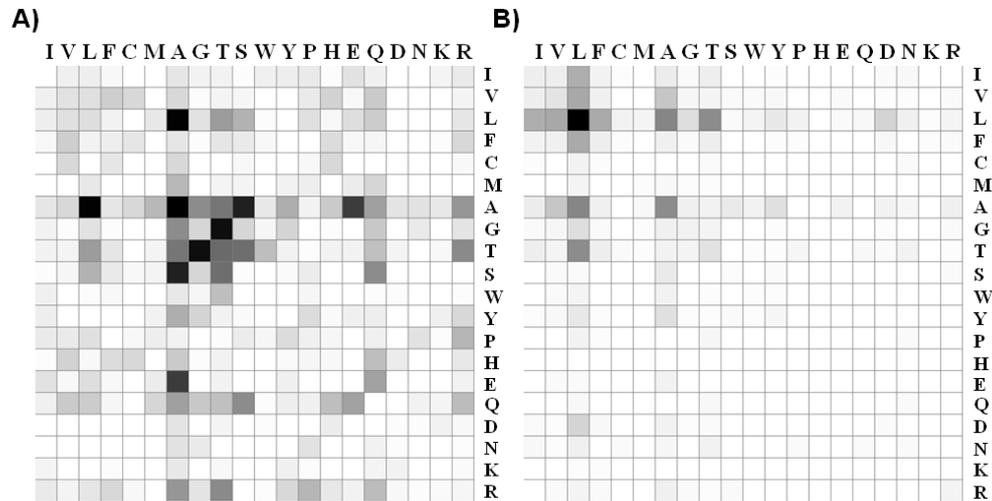
A)   **B)**

Fig. 2.   The pairing preferences of amino acids of the 3D pattern pairs discovered in A) transient, and B) obligate complexes. The darkness of each cell corresponds to the normalized frequency of amino acid interactions: the darker it is, the more frequently the interaction occurs. The amino acid residues are ordered according to their hydrophobicity, with I (Ile) as the most hydrophobic and R (Arg) as the least hydrophobic.

ity. From this study, we observe the high occurence of interactions between hydrophobic amino acids in obligate complexes. These hydrophobic interface resemble domain interfaces or the protein core. As obligate complexes may not fold and function properly when their proteins are unbound, it is sensible to consider obligate complexes as an extension of protein folding.[2] On the other hand, the high-occuring pairings between amino acids of large hydrophobicity difference in transient complexes suggest a kind of interactions that are less permanent than obligate ones. This supports that transient complexes share some similarities with active sites of enzymes, but they are more conservative than active sites of enzymes.[2]

Although 3D pattern pairs are not sufficient to predict a complete structure of protein complexes, there are differences between the interacting amino acid preferences in 3D pattern pairs of transient and those of obligate complexes, which are useful in understanding the difference between transient and obligate complexes. This will be useful in large-scale structural proteomics initiatives, especially for assemblies of protein complexes, in which the physico-chemical characterization is incomplete. In brief, using maximal quasi-bicliques allows us the flexibility of having a wider range of parameter settings to obtain 3D pattern pairs.

Although there are only 20 possible types of amino acid in most binding sites of protein-protein interactions, there are many more variations that can occur through subsequent modification. This necessistates the inclusion of post-translational modification information in the future analysis of binding sites of distinct protein-protein interactions.

10

References

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York and London, 2002.
2. A. I. Archakov, V. M. Govorun, A. V. Dubanov, Y. D. Ivanov, A. V. Veselovsky, P. Lewi, and P. Janssen. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*, 3:380–391, 2003.
3. M. J. Betts and R. B. Russell. Amino acid properties and consequences of substitutions. *Bioinformatics for Geneticists*, pages 289–316, 2003.
4. A. A. Bogan and K. S. Thorn. Anatomy of hot spots in protein interfaces. *J Mol Biol*, 280:1–9, 1998.
5. S. De, O. Krishnadev, N. Srinivasan, and N. Rekha. Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Structural Biology*, 5:15, 2005.
6. D. Eppstein. Arboricity and bipartite subgraph listing algorithms. *Information Processing Letters*, 51:207–211, 1994.
7. F. Glaser, D. M. Steinberg, I. A. Vakser, and N. Ben-Tal. Residue frequencies and pairing preferences at protein-protein interfaces. *Nucleic Acids Res*, 43:89–102, 2001.
8. G. Grahne and J. Zhu. Fast algorithms for frequent itemset mining using fp-trees. *IEEE Transactions on Knowledge and Data Engineering*, 17(10):1347–1362, October 2005.
9. S. Jones, A. Marin, and J. M. Thornton. Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Engineering*, 13:77–82, 2000.
10. S. Jones and J. Thornton. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol*, 272:121–32, 1997.
11. L. Lo Conte, C. Chothia, and J. Janin. The atomic structure of protein-protein recognition sites. *J Mol Biol*, 285(2):177–198, 1999.
12. Y. Ofran and B. Rost. Analysing six types of protein-protein interfaces. *J Mol Biol*, 325(2):377–387, 2003.
13. N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. *ICDT*, pages 398–416, 1999.
14. F. B. Sheinerman, R. Norel, and B. Honig. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol*, 10:153–159, 2000.
15. K. Sim, J. Li, V. Gopalkrishnan, and G. Liu. Mining maximal quasi-bicliques to co-cluster stocks and financial ratios for value investor. *ICDM*, pages 1059–1063, 2006.
16. E. Sprinzak, Y. Altuvia, and H. Margalit. Characterization and prediction of protein-protein interactions within and between complexes. *PNAS*, 103:14718–14723, 2006.
17. C. J. Tsai, S. L. Lin, H. J. Wolfson, and R. Nussinov. A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *J Mol Biol*, 260:604–620, 1996.
18. D. Xu, S. Lin, and R. Nussinov. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J Mol Biol*, 265:68–84, 1997.
19. H.-X. Zhou and Y. Shan. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins Structure Function and Genetics*, 44:336 – 343, 2001.
20. H. Zhu, F. S. Domingues, I. Sommer, and T. Lengauer. Noxclass: prediction of protein-protein interaction types. *BMC Bioinformatics*, 7:27, 2006.