International Journal for Environmental Rehabilitation and Conservation
Volume I No. 1 2010 [22-27] [ISSN 0975 - 6272]

Deduction of Protein Structure with help of Tree Decomposition

Sudhir Prakash Srivastava[1] and Neha Srivastava[2]

Received: January 20, 2010 | **Accepted**: February 24, 2010 | **Online**: April 4, 2010

Abstract

Robertson and Seymour first time defined the notion of tree decomposition together with associated graph parameter tree width; this parameter played an important role in algorithm theory. In this paper we study a tree decomposition based algorithm for proteins structure. The tree based computational models are powerful, flexible and efficient way to modeling many type of proteins structure. We can developed algorithm that explore the fact that contain key parameter have complexity dependent on tree width of protein structure, when tree width is large, we can still use spectral methods to find protein sound solution in an efficient manner

Keywords: *Protein Structure | Tree - Decomposition | Computational models | Protein Side chain packing*

For Correspondence:

[1] IET, Dr. R.M.L. Avadh University, Faizabad, India
[2]Dept. of Botany, M.L.K.P.G. College, Balrampur, India

Introduction

The structure of protein play on instrumental role in determining in functional activity. The experimental method like NMR techniques and X-ray crystallography cannot generate protein structure in high through put way. Proteins structure has been used in many pharmaceutical companies to analyze the structure & functional characteristic of a protein. We can classify protein structure into two major steps. One is the predication of the backbone atom coordinates and other is the predication of side chain atom coordinate. A protein is a complex biological system consisting of dozens or hundreds of small molecules (i.e. amino acid) interact with specific shape. It may be possible that proteins also interact with each other to form and protein -protein interaction (PPI) network. A PPI network describes the interaction relationship among protein in cell; each vertex in the network corresponds to a protein and edge indicates a direction physical interaction between two proteins.

In this paper we study about protein structure with help of tree decomposition method. The tree decomposition based algorithm have fund a rich set of application in proteins structure.

Tree -Decomposition Concept

The notions of tree width and tree decomposition are introduced by Robertson and Seymour (1986) in their work on graph minor.

Graph minor is a branch of graph theory. In graph minor the 'decomposition theorem' describes the structural feature of all graphs excluding a given minor. In order way we can say that decomposition theorem say that sharse graph can be decomposed into a tree of component. Each component contains a small number of vertices from a graph. The width of tree decomposition is the maximum component size minus one. The tree width of a graph is the minimum width over all the tree decomposition. It any graph with tree width then computational problem related to the graph can be solve using dynamic programming with time complexity polynomial in graph size and tree width.

Definition

Let G = (V,E) be a graph. A tree decomposition of G pair (T,X) satisfying the following condition:

(i) T = (I, F) is a tree with a vertex set i.e. node set I and an edges set F,

(ii) $X \{ X_i / i \in I , X_i \in V\}$ and $U_{i\in I} X_i = V$

That is each node in the tree T represents a subset of V and Union of all the subset is V,

(iii) for every edge $e = (\upsilon,\omega) \in E$, there is at least one $i \in I$ such that both u and ω are in Xi, and

(iv) for all i, j, k ∈ I if j is anode of the path from i to k then $X_i \cap X_k \subseteq X_j$

The width of tree decomposition is $\max_{i\varepsilon I}\{|X_i|-1\}$

The tree width of a graph G, denoted by tw (G) is the minimum width overall the tree decomposition of G.

According to the above definition, the decomposition of a graph into bio-connected component corresponds to a vertex in T. and any two bio-connected component share one vertex of G. Then the width of bio-connected-component decomposition could be O (IVI). Then O (IVI) is bigger than tree width of G. if G is spares.

For example, when the graph is cycle, this graph has only on bio-connected component, itself and then tree width of cycle is only 2. Fig. 1, 2 & 3 shows on example of an interaction graph, its bio-connected component decomposition with width 6 and a tree decomposition with width 3. The width of the tree decomposition is a main factor determining the computational complexity of all the tree decomposition based algorithm. Smaller value of tree decomposition width algorithm is more efficient. Hence we try to optimize the tree decomposition of residue interaction graph.

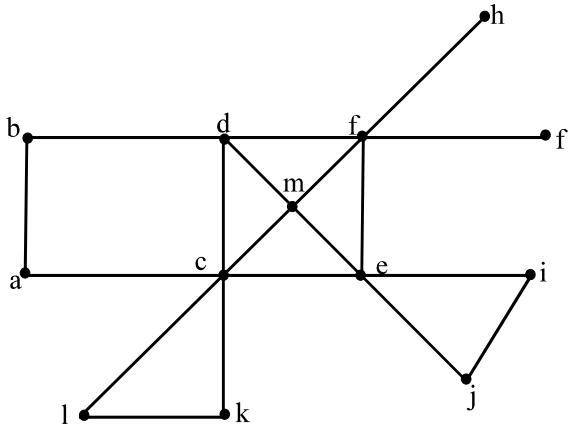


Fig. 1: Example of a residue interaction graph

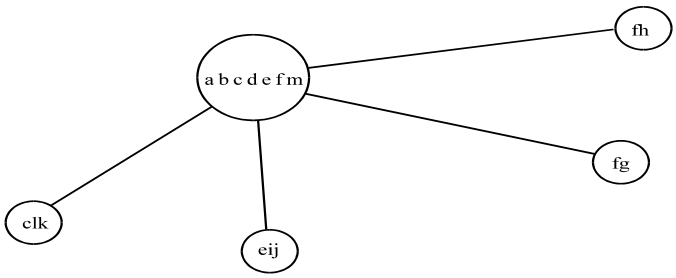


Fig.2: Example of the bio-connected component decomposition of a graph with width decomposition 6

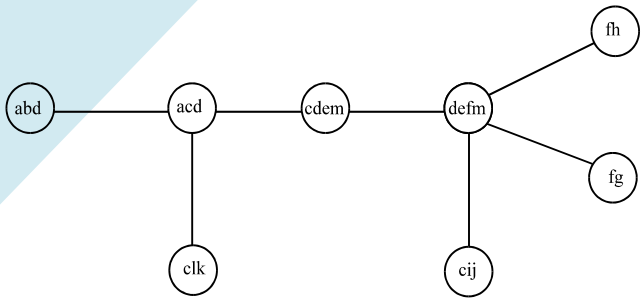


Fig.3: Example of tree decomposition of a graph with width 3

Tree Decomposition Based Algorithm

We can consider tree decomposition based algorithm for so many biological system. For example in the field of proteins structure, we can consider proteins side chain packing [Xu (2007) and Xu and Berger (2006)] and non sequential protein structure alignment (Xu *et al*., 2005, 2006).

Many biological problem including protein side chain packing and protein structure alignment can be formulated as a problem of assigning a label to each vertex in spares graph G= (V_1E) with bounded tree width ω. For any vertex ν in V.

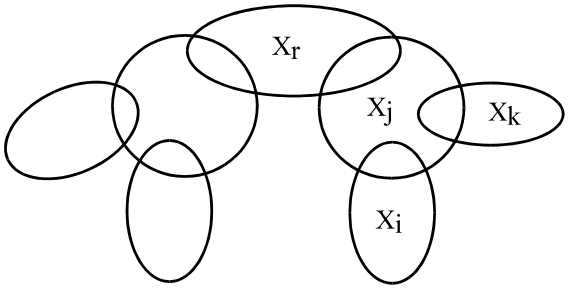
There is a set of candidate labels, denoted by D[v] for this vertex. There may be some restriction on the label assignment of two adjacent vertices that is given edges (v_1, v_2) in E. The feasible labels of v_1 and v_2 are restricted to a subset of D[v_1] x D [v_2]. We want to find a label assignment A(v) (A(v) D[v]) to vertex v ε V such that the following scoring function is optimized.

$$F(A) = \sum_{u\varepsilon v} S_v (A(v)) + \sum_{u\neq v,\,(u,v)\,\varepsilon\varepsilon} P_{u,v}(A(u)), A(v))$$

where S_v(A(v) denoted the preference of assigning a label A(v) to vertex u and P_{uv} (A(v), A(u)) measure how well two labels A(u) and A(v) are simultaneously assigned to u and v respectively.

The general graph labeling problem is computationally hard. However, if the graph has bounded tree width w, the graph labeling problem can be solved using tee decomposition with time complexity. Now we describe a recursive equation, that can be used to calculate F(A), based on tree decomposition of G.

Assume tree decomposition (T,X) of a sparse graph G. For simplicity, we assume that tree T has root X_r. Figure 4 shows an example of tree decomposition in which component X_r is the root.



Let X_{rij} denote the intersection between X_r and X_j. If we remove all the vertices in V_{rj}; then this tree decomposition become two disconnected sub tree. Let F[X_j I A($X_{r,j}$)] denote the optimal label assignment of sub-tree rooted at X_j, denoted by T(X_j), given that the label assignment to $X_{r,j}$ is fixed to A($X_{r,j}$). Then F[X_j ׀ A($X_{r,j}$)] is independent of T-T (X_j). Let C(j) denoted the set of child component of X_j and score (X_j ׀ A(X_j) denote the assignment score of component X_j with the label assignment being A(X_j). Let D[X] denote all the possible label assignment to the vertices in X. Therefore, we have the following recursion equation.

$$F(X_j \mid A(X_{r,j}) = \min \{F(X_i \mid B(X_{j,i})) + Score(X_j) \mid B(X_j - X_{i,j}) \cap A(X_{r,j})\}$$

According to this recursive equation we can calculate the optional label assignment using a divide strategy. First, we calculate the optional scoring function F(A) from bottom to top of the

tree decomposition and then we extract the optimal label assignment from top to bottom. A detailed account of the tree decomposition based algorithm is present in [Xu and Berger (2006), Xu *et al*., (2005)].

Application of Tree Decomposition:

We have much biological system which convertable in graph model. A Biological graph model have bounded tree wide can be study with tree decomposition algorithm. In this paper we take some example of tree decomposition based algorithm.

(1) Protein Side Chain Packing:

During the study protein side chain packing Xu, (2007) formulate problem with tree decomposition algorithm as

Let G denote the graph modeling the amino acid residue interaction relationship in a proteins, each vertex in G represent a residue in the proteins and there is one edge between any two rotamers of these two residues the protein side chain packing problem can be formulate to minimize the following scoring function.

$$E(G) = \sum_{i\varepsilon v, A(i)\ \varepsilon D(i)} S_i(A(i)) + \sum_{i\neq j,\ (i,j)\ \varepsilon E} P_{i,j}(A(i)), A(j))$$

Where D[i] is the set of candidate rotamer for position i and A[i] is rotamer assigned to position i. The score item $S_i(A(L))$ measure the preference of a rotamer occurring at the given position and Pij [A(i), A (j)] measures how well two rotamer can be assigned simultaneously to two interaction residue.

Using the geometrical feature of residue interact graph. We have proved that the residue interaction graph can be tree decomposed into many small components and thus there is an efficient tree decomposition based algorithm.

Protein contact map overlap.

Proteins structure alignment is a fundamental problem in structural bioinformatics. A proteins structure alignment algorithm align two proteins structure and calculates their similarly. In content graph based proteins structure alignment a proteins structure is modeled as contact graph and the similarity between two protein is measure by their maximum common sub graph [Bernstein et al., (1977) and Lathrop (1994)]. Some proteins structure alignment programs only generate sequential alignment [Holm and Sander (1993)] while other gerate non sequential alignment [Alexandrov (1996) and Yuan and Bystroff (2005)]

Let E(A) and E(B) denote the set of contacts in proteins A and B respectively. For any residue u in A. let M(u) denote its equivalent residue in B. If there is no equivalent residue for u. then M(u) =ϕ. The contant map overlap problem can be formulated as follows.

$$\max \sum_{(u,v)\ \varepsilon[A],v<u} f(u,vM(u), M(v))$$

For non sequential aligment

$$F\{u,vM(u), M(v)\} = \begin{cases} -\infty\ M(u) = M(u) \neq \emptyset \\ 1\ \{M(u)\} = M(u)\varepsilon E(B) \text{otherwise} \\ 0 \end{cases}$$

For sequence alignment f{u,v, M(u), N(v)} can be redefined as follows:

$$F\{u,vM(u), M(v)\} = \begin{cases} 0 \\ -\infty\ M(u) \text{ or } M(u) = \emptyset \\ 1\ \{M(u)\} \geq M(u) \\ 0\ \ M(u), M(u)\varepsilon E(B) \text{otherwise} \end{cases}$$

General Protein Threading

Protein threading is an important method for proteins structure predication. About new protein, templete protein date Bank more

concept please refer (Xu *et al*., 2005). Protein threading is computationally complicated, if the scoring function contain pair-wise constant potentials and gaps are allowed in the alignment [Akutsu and S. Miyano (1999)].

Many time many mathematician developed many algorithm for this problem. But none of this exact algorithm is guaranteed to terminate within reasonable theoretical time complexity.

So many publication and some preliminary studies on protein threading using tree decomposition are available we can used template contact graph to model a structural template in the Protein Data Bank. Then Protein threading can formulate as a problem of assigning some labels to a contact graph to minimize a scoring function.

Besides the above mentioned application, tree decomposition can also be applied to Protein complex threading protein complex. Threading problem can be formulated as a problem of aligning two sequence to a bipartite graph. The bipartite graph is geometric and also sparse. So it should have small tree width.

Some publication (Qu *et al*., 2004) protein. Threading with NMR data problem also formulated with help of tree decomposition algorithm.

Conclusion

This paper describe tree decomposition algorithm. Tree decomposition algorithm is very effective in solving problem in protein structure, because a protein structure usually can be modeled as spare geometric graph which treated as a tree width.

We are conducting a systematic study to identity proteins problem suitable for tree decomposition more non-trivial example, such as solution to other major protein problem should be given to demonstrate its usefulness. One challenge is to develop an empirically efficient algorithm for the protein problem with medium-sized tree width.

References

Alexandrov, N. (1996): SAR Fing the PDB. Protein Engineering Vol. 9: pp 727-732.

Akutsu, T. and Miyano, S. (1999): On the approximation of protein threading. Theoretical computer science Vol. 210: pp 3261-275.

Bernstein, F.C., Keetzle, T.F., and Williams, J. (1977): The protein data bank: A computer -based archived file for macromolecular structure. J. of Molecular biology Vol. 112(3): pp 535-542.

Holm, L. and Sander, S. (1993): Protein structure comparison by alignment of distance matrices, K. of Molecular biology Vol. 233: pp 123-138.

Lathrop, R. (1994): The protein threading problem with sequence amino acid interaction preference is NP-Complete. Protein Engineering Vol. 7: pp 1059-1068.

Qu, Y., Guo, J., Olman, V., and Xu, Y. (2004): Protein structure prediction using sparse dipolar coupling data. Nucleic acids Research, Vol 32(2): pp 551-561.

Robertson, N. and Seymour, P. (1986): Graph minor II algorithmic aspects of tree width, Journal of Algorithms Vol. 7: pp 309-322.

Xu, J. (2007): Solving the contact map overlap problem via tree decomposition and a dee-like pruning strategy. In proceeding the 46th IEEEE Conference on Decision and Control. (CDC), New Oreleans.

Xu, J. and Berger, B. (2006). Fast and accurate algorithm for protein side chain packing. Journal of ACM, Vol. 53: pp 533-557.

Xu, J., Jiao, F. and Berger, B. (2005): A tree-decomposition approach to protein structure prediction in proceeding computational system bio-informatics. pp 247-256.

Xu, J., Jiao, F. and Berger, B. (2006): A parameterized algorithm for protein structure alignment. In proceeding of the Tenth Annual International Conference on Research in Computational Molecular Biology, pp 488-499. Springer.

Yuan, X. and Bystroff, C. (2005): Non-sequential structure based alignments reveal topology-impendent core packing arrangements in proteins, Bioinformatics, Vol. 27: pp 1010-1019.