

# A Maximum Edge-weight Clique Finding Algorithmic Approach for Solving Protein Side-chain Positioning Problem

Dukka Bahadur K.C.<sup>1</sup>

dukka@kuicr.kyoto-u.ac.jp

Etsuji Tomita<sup>2</sup>

tomita@ice.uec.ac.jp

Suzuki Junichi<sup>2</sup>

jsuzuki@ice.uec.ac.jp

Tatsuya Akutsu<sup>1</sup>

takutsu@kuicr.kyoto-u.ac.jp

<sup>1</sup> Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto 611-0011, Japan

<sup>2</sup> Graduate School of Electro-communications, The University of Electro-Communications, Tokyo 182-8585, Japan

**Keywords:** protein side-chain packing, clique, combinatorial optimization, protein design

## 1 Introduction

We have developed a novel approach to solve the protein side-chain packing problem using the notion of a maximum edge-weight clique. Our approach is based on efficient reduction of protein side-chain packing problem to a graph and then solving the reduced graph to find the maximum clique. Since our approach is based on deterministic algorithms in contrast to the various existing algorithms based on heuristic approaches, our algorithm guarantees of finding an optimal solution. We have compared our results with the various existing approaches and have found that our results are comparable or considerably better than the existing methods.

## 2 Method and Results

We reduced the protein side-chain positioning problem into an weighted graph and then applied the maximum edge-weight clique finding algorithm developed by our co-authors [6].

The side-chain search space was sampled using discrete rotation angles defined by  $(2\pi k)/K$   $|k = 0, \dots, K - 1$ , where  $K$  is chosen to be 18. Let  $R = r_1, \dots, r_n$  be the set of residues of the given protein whose side-chain conformation has to be calculated and let  $r_{i,k}$  be the  $i$ -th residue whose side-chain atoms are rotated by  $(2\pi k)/K$  radian. This conformation is only considered as a node if the minimum distance between the atoms in  $r_{i,k}$  and the entire main-chain is greater than  $L_1 \text{Å}$ .

Similarly, for the generation of edges, let  $r_{i,k}$  be the  $i$ -th residue whose side chain atoms are rotated by  $(2\pi k)/K$  radian and  $r_{j,h}$  be the  $j$ -th residue whose side chains are rotated by  $(2\pi h)/K$  radian. Then an edge is drawn between conformation  $r_{i,k}$  and conformation  $r_{j,h}$  if the minimum distance between the atoms is greater than  $4.0 \text{Å}$ .

Finally, each edge is assigned weight based on the ‘Probability discriminatory function’ proposed by Samudrala et. al [4].

### 3 Computational Experiments and Results

The maximum edge-weight clique finding algorithm developed by Suzuki et. al [6](co-author) is utilized to solve the weighted graph to find the edge-weight maximum clique. The running environment of the program is ORIGIN 3600.

The approach is utilized to predict the side-chain positioning of a set of proteins and the results of the respective RMSD is given in the table 1. The H&S method refers to the method of Holm & Sander [2] and Lee &S refers to the method of Lee and Subbiah [3].

Table 1: Comparison of Weight-clique Algorithm other methods

PDB	SPWCQ	H&S	Lee&S
1crn	0.81	-	1.65
5pti	1.37	1.90	1.49
1ctf	0.86	1.70	1.86
7rsa	1.15	1.80	1.86
1lz1	0.99	1.60	1.62
3fxn	0.90	1.90	1.90
3app	1.43	1.40	1.22
2cro	0.96	2.30	2.39
3tln	1.56	1.70	-

### 4 Discussions

As per our initial motivation, we were able to design a branch and bound approach for the protein side-chain positioning problem. The results of the method are comparable or better than the existing methods. Moreover, our method can be applied to a protein of size of almost 400 residue.

### References

- [1] Akutsu, T., NP-Hardness results for protein side-chain packing, *Genome Informatics*, 8:180–186, 1997.
- [2] Holm, L. and Sander, C., Fast and simple Monte Carlo algorithm for side-chain optimization in proteins: application to model building by homology, *Proteins: Struct. Funct. Genet.*, 14, 213-23.
- [3] Lee, C. and Subbiah, S., Prediction of protein side-chain conformation by packing optimization, *J.Mol.Biol*, 217, 373-388.
- [4] Samudrala, R. and Moulton, J., An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction, *J. Molecular Biology*, 275:893-914, 1998.
- [5] Tomita, E. & Seki, T., An efficient branch-and-bound algorithm for finding a maximum clique, in 'DMTCS 2003', LNCS 2731, pp.278-289.
- [6] Suzuki, J., Tomita, E. and Seki, T., An algorithm for finding a maximum Clique with maximum edge-weight and computational experiments, in 'Technical Report' MPS, The Information Processing Society of Japan.