Contact Curve Based Simulation of Side Chains from Two Amino Acids in a Protein Molecule

Sanghun Jeong[1,2], Ku-Jin Kim[1,2]
[1]*School of Computer Science and Engineering, Kyungpook National University, Korea.*
[2]*Software Technology Research Center (SWRC), Kyungpook National University, Korea.*
kujinkim@gmail.com

***Abstract*—In this paper, an algorithm to compute the contact configuration between the rotating side chains from two amino acids in a protein molecule is proposed. The main chain and side chain parts in one amino acid are considered as two rigid-body parts combined with a revolute joint. While the atom positions in the main chain parts of two amino acids are fixed, the side chains possibly rotate, which can cause the collision between amino acids. On the plane of rotation angle parameters, we find the region for two side chains colliding each other. Then, by extracting the boundary of the region, the contact curve of the amino acids is found, where by using it, the amino acids with rotating side chains can be simulated.**

***Index Terms*—Protein Molecule; Side Chain Flexibility; Collision Detection**

I. Introduction

The protein molecule has flexibility both in the main and side chains of each amino acid [1,2,3] For visualizing the molecule or computing its geometric properties, the flexibility of the protein molecule is often simplified as side chain rotations only. For the case of computing the configuration of two rotating side chains from two different amino acids, it is essential for finding the atom positions not to have collisions, since two atoms with no bond do not exist within the van der Waals radius [4,5] in a natural state. When two amino acids are represented as two sets of van der Waals spheres, the spheres from each amino acid do not intersect each other. In this paper, a method to detect the collision and the contact configuration of rotating side chains in two amino acids based on the sphere representation of atoms is proposed. Then, the method to simulate the rotating side chains is given.

The remaining of this paper is organized as follows. In Section II, the geometric features of the side chain flexibility in the protein amino acids are explained. In Section III, we show the method for computing the contact curve between two rotating side chains. The experimental results are shown in Section IV, and this paper is concluded in Section V.

II. Side Chain Flexibility

A protein molecule is composed of a sequence of amino acids, where the amino acid is composed of atoms. In geometric approaches that deal with protein molecules, an atom is frequently represented as a sphere with van der Waals radius. When two atoms have a covalent bond, they are represented as two intersecting spheres. If there is no covalent bond, two spheres do not intersect each other.

Each amino acid in the protein molecule has two groups of atoms in a main chain and a side chain. For every amino acid, the main chain part is composed of four atoms: N, Cα, C, O. Each amino acid has a unique side chain, where an amino acid is classified by its side chain. The atom Cα in the main chain and the atom Cβ in the side chain have a covalent bond with each other, where there is no other covalent bond between the main chain atoms and the side chain atoms. When we draw a line L passes through the center of Cα and Cβ, it becomes a rotation axis of the side chain. Usually, there is no covalent bond between two atoms from different side chains.

Figure 1 shows an example of an amino acid with a rotating side chain.

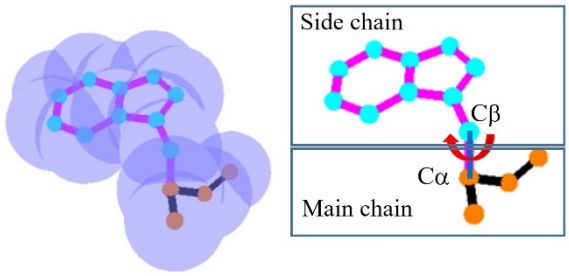


Figure 1: An example of the rotating side chain in an amino acid

III. Computation of Contact Curve for Two Rotating Side Chains

The side chain of an amino acid is represented as a set of spheres corresponding to the atoms. The relative position of atoms in a side chain is fixed and the side chain itself rotates as a rigid body. By analyzing and accumulating the contact information between the spheres from different side chains, the rotation of two side chains without collision can be simulated.

Kim et al. [6] presented equations for computing the contact between two rotating or translating simple surfaces. In this paper, rather than using the equation for computing the contact, we sample the parameter space, and then we check if two spheres collide, contact, or apart from each other.

Given two rotating spheres A and B, the sphere A is with the center position $\mathbf{c}_0=(x_0, y_0, z_0)$, radius r_0, and rotation axis (0, 0, 1). Then, the center trajectory of A is as follows:

$$A.c(s) = (R_0 \cos(s + s_0), R_0 \sin(s + s_0), z_0) \quad (1)$$

where s_0 = atan (x_0, y_0) and R_0 is the distance between $\mathbf{c}_0$ and the z-axis.

The sphere B is with the center position $\mathbf{c}_1 = (x_1, y_1, z_1)$, radius r_1, and rotation axis $\mathbf{p} - \mathbf{p}'$, where $\mathbf{p} = (p_x, p_y, p_z)$ and $\mathbf{p}' = (p_x', p_y', p_z')$. We define a plane P with a reference point $\mathbf{p}$ and normal vector $\mathbf{b}_3$, where $\mathbf{b}_3 = (\mathbf{p}' - \mathbf{p})/\|\mathbf{p}' - \mathbf{p}\|$. For two orthogonal basis vectors of P, $\mathbf{b}_1$ and $\mathbf{b}_2$, the followings are given:

$\mathbf{b}_1 = (b_{1x}, b_{1y}, b_{1z})$,
$\mathbf{b}_2 = (b_{2x}, b_{2y}, b_{2z})$,
$\mathbf{b}_3 = \mathbf{b}_1 \times \mathbf{b}_2 = (b_{3x}, b_{3y}, b_{3z})$,

and $\|\mathbf{b}_1\| = \|\mathbf{b}_2\| = \|\mathbf{b}_3\| = 1$. The sphere B is with the center position $\mathbf{c}_1=(x_1, y_1, z_1)$, radius r_1, and rotation axis is a line $\mathbf{p}$-$\mathbf{p}'$. Let c_1' be the projection of $\mathbf{c}_1$ to the plane P. Then, the center trajectory of B is as follows:

$$B.c(t) = \mathbf{p} + R_1 \cos(t + t_0)\,\mathbf{b}_1 + R_1 \sin(t + t_0)\,\mathbf{b}_2 + \|c_1 - c_1'\|\,\mathbf{b}_3, \quad (2)$$

where t_0 = atan(k_x, k_y) and k_x, k_y, and R_1 are the distances from $\mathbf{c}_1'$ to lines $\mathbf{p} + \alpha\mathbf{b}_1$, $\mathbf{p} + \beta\mathbf{b}_2$, and $\mathbf{p} + \gamma\mathbf{b}_3$, respectively.

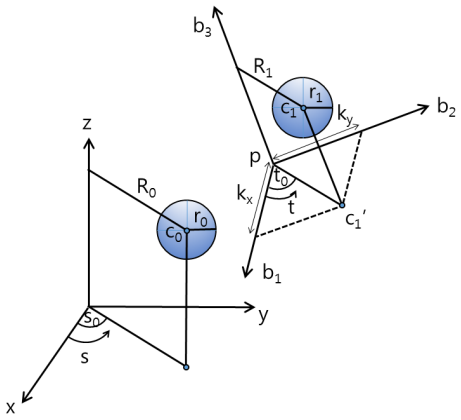


Figure 2: Two rotating spheres

For s*∈ s and t*∈ t, in st-plane, the status of two spheres A.c(s*) and B.c(t*) are determined by the following conditions:

1. Two spheres at (s*, t*) tangentially contact with each other $\|A.c(s^*) - B.c(t^*)\|^2 = (r_0 + r_1)^2$
2. Two spheres at (s*, t*) have an intersection with each other
   $\|A.c(s^*) - B.c(t^*)\|^2 < (r_0 + r_1)^2$
3. Two spheres at (s*, t*) are apart from each other
   $\|A.c(s^*) - B.c(t^*)\|^2 > (r_0 + r_1)^2$

Figure 3 shows an example for the collision area between two rotating spheres, where the point (s*,t*) is colored in black if two spheres at (s*, t*) have an intersection. We choose two spheres which correspond to NE1 and CD2 from two amino acids (Figure 3a), and then computed the collision area in st-plane, where their rotation angles are s and t, respectively (Figure 3b).

For two side chains A(s) and B(t), let us denote each atom in them as $A_i(s)$ and $B_j(t)$, respectively. Then, the center position and the radius of each atom in A(s) are represented as $A_i.c(s+s_i)$ and $A_i.r$, respectively. Those in B(t) are represented as $B_j.c(t+t_j)$ and $B_j.r$, respectively.

In st-plane, if a point (s*, t*) is given, we can decide if $A_i(s)$ and $B_j(t)$ collide or apart based on the following conditions:

1. If there exists any atom pair that satisfies the equation $\|A_i.c(s^*+s_i) - B_j.c(t^*+t_j)\|^2 < (A_i.r + B_j.r)^2$, then two side chains collide with each other.
2. If every atom pair satisfies the equation $\|A_i.c(s^*+s_i) - B_j.c(t^*+t_j)\|^2 > (A_i.r + B_j.r)^2$, then two side chains are apart from each other.

If two side chains are not either collide or apart, then we can decide it as tangential contact status. Algorithm 1 shows the process to construct the contact area for two side chains in the st-plane.

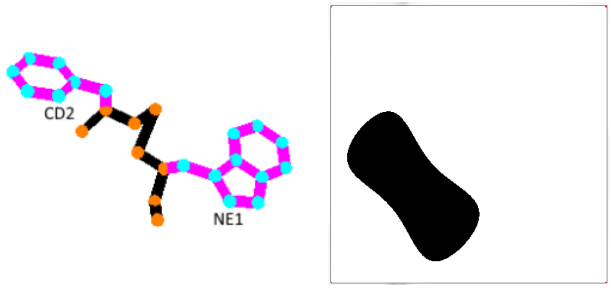


Figure 3: The collision area for two rotating atoms

Algorithm 1
Collision Area Computation

```
Input: δ // the step size
       A_i(s), 0≤ i < N_A // Atoms in side chain A
       B_j(t), 0≤ j < N_B // Atoms in side chain B
1   BEGIN
2       for ( s = 0 ; s < 2π ; s += δ ) BEGIN
3       for ( t = 0 ; t < 2π ; t += δ ) BEGIN
4           M[s][t] = FALSE;
5           for ( i = 0 ; i < N_A ; i ++) BEGIN
6           for ( j = 0 ; j < N_B ; j++ )
7               if (A_i(s) ∩ B_j(t) ≠ ∅) BEGIN
8                   Set TRUE to M[s][t];
9                   Break;
10                      END
11          END
12      END
13  END
14  END
```

When M[s*][t*] is TRUE, the point (s*, t*) represents that two side chains have collisions when their configuration is at A(s*) and B(t*). If M[s*][t*] is FALSE, two side chains A(s*) and B(t*) do not have collisions.

For the two side chains given in Figure 3a, Figure 4a represents the st-plane by representing the point at (s, t) as black in color when M[s][t] is TRUE, and as white in color when M[s][t] is FALSE. Generated matrix corresponds to a binary image, where the boundary of the black region corresponds to the curve that represents the tangential contact configuration between the two side chains. Figure 4b shows the contact curve.

In Figure 5, we represent the cases of two side chains without collisions (Figure 5a), with contact (Figure 5b), and with collisions (Figure 5c), where red dots on the left column shows the corresponding angles.

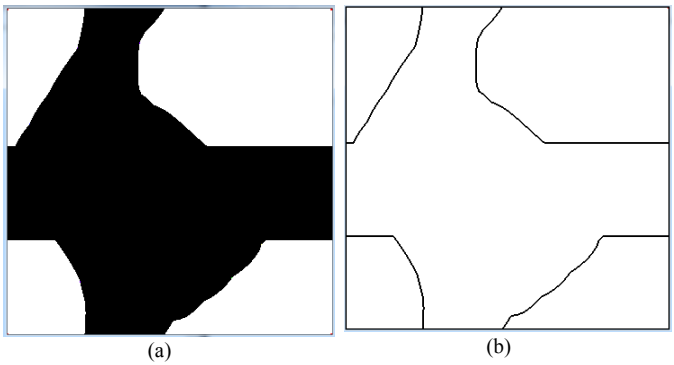
When we force amino acid A(s) to rotate around the axis, the other amino acid B(t) will rotate with respect to the movement of A(s). We simulate this situation, by following the contact curve with increasing s values.

IV. Experimental Results

We simulated the side chain rotation for several pair of amino acids. In Figure 6, we show one of the examples, where the simulation results are generated based on the (s,t) pairs in Figure 5. When two amino acids PHE and TRP are given as a PDB file format (http://www.pdb.org) such as in Table 1, the collision status in the (s,t)-plane is computed. Then, given a fixed value t=t*, we find a s* where (s*, t*) is on the contact curve. We increment the value of parameter s from s*. Each time the value of s is incremented, we compute the collision status of $A_i(s)$ and $B_j(t^*)$. While $A_i(s)$ and $B_j(t^*)$ are free, only the value of s is changed. If $A_i(s)$ and $B_j(t^*)$ collide or contact with each other, then we find the closest t* which is an apart status.

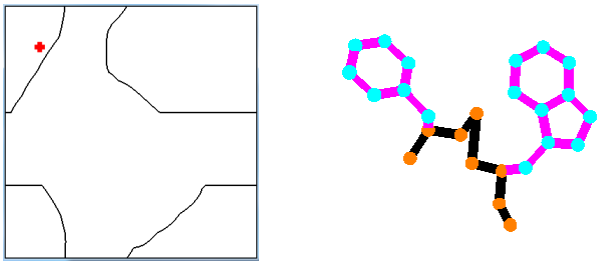
Table 1
Input PHE and TRP information

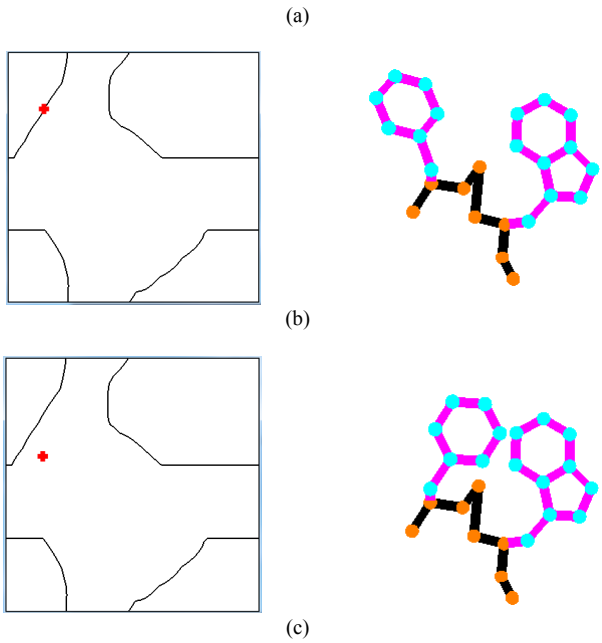
ATOM	1	N	PHE	A	1	21.320	22.197	64.569	1.00	10.94
ATOM	2	CA	PHE	A	1	19.900	22.372	64.205	1.00	11.49
ATOM	3	C	PHE	A	1	19.058	22.732	65.374	1.00	12.09
ATOM	4	O	PHE	A	1	17.939	23.205	65.206	1.00	12.13
ATOM	5	CB	PHE	A	1	19.317	21.042	63.629	1.00	12.11
ATOM	6	CG	PHE	A	1	20.330	20.438	62.718	1.00	8.33
ATOM	7	CD1	PHE	A	1	20.663	21.045	61.540	1.00	8.88
ATOM	8	CD2	PHE	A	1	21.011	19.291	63.124	1.00	7.80
ATOM	9	CE1	PHE	A	1	21.691	20.538	60.771	1.00	10.46
ATOM	10	CE2	PHE	A	1	22.046	18.812	62.377	1.00	9.63
ATOM	11	CZ	PHE	A	1	22.389	19.443	61.204	1.00	8.54
ATOM	12	N	TRP	A	2	19.533	22.544	66.617	1.00	10.96
ATOM	13	CA	TRP	A	2	18.829	22.870	67.838	1.00	10.24
ATOM	14	C	TRP	A	2	19.950	23.033	68.928	1.00	11.45
ATOM	15	O	TRP	A	2	20.112	22.121	69.741	1.00	11.40
ATOM	16	CB	TRP	A	2	17.870	21.716	68.159	1.00	9.36
ATOM	17	CG	TRP	A	2	16.866	22.129	69.200	1.00	8.64
ATOM	18	CD1	TRP	A	2	16.781	21.770	70.478	1.00	9.52
ATOM	19	CD2	TRP	A	2	15.749	23.012	68.983	1.00	13.34
ATOM	20	NE1	TRP	A	2	15.725	22.368	71.102	1.00	11.57
ATOM	21	CE2	TRP	A	2	15.072	23.163	70.204	1.00	14.95
ATOM	22	CE3	TRP	A	2	15.299	23.744	67.862	1.00	19.41
ATOM	23	CZ2	TRP	A	2	13.925	23.970	70.310	1.00	16.89
ATOM	24	CZ3	TRP	A	2	14.170	24.536	67.988	1.00	18.89
ATOM	25	CH2	TRP	A	2	13.524	24.639	69.209	1.00	15.29



(a) (b)

Figure 4: The collision configuration for two rotating side chains.





(a) (b) (c)

Figure 5: The contact curve and a pair of specific angles (s*, t*) in st-plane and the configuration of two amino acids with respect to (s*, t*).

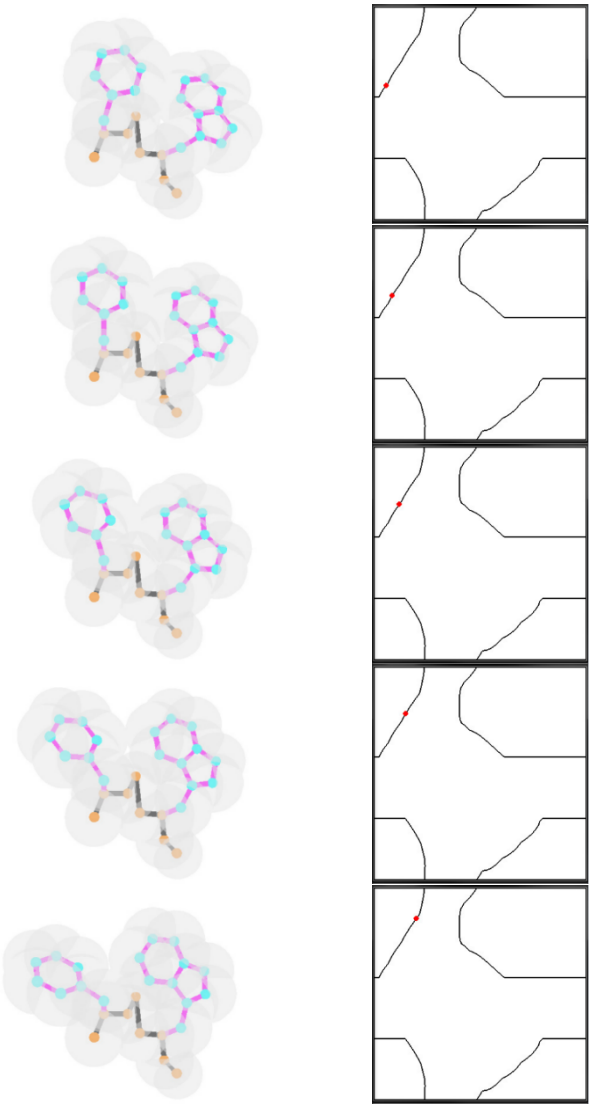


Figure 6: Simulation of the rotating side chains from two amino acids contacting each other using their contact curve.

V. CONCLUSION

In this paper, an algorithm for analyzing the contact between two rotating side chains in a protein amino acids was presented. Based on the curves representing the contact, the movement of two side chains was simulated. As a future work, we would like to extend our simulation to a set of side chains.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03930628) and the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

REFERENCES

[1] M. Diez, V. Petuya, L. A. Martinez-Cruz, A. Hernandez, Insights into mechanism kinematics for protein motion simulation, BMC Bioinformatics, 15(184) (2014).

[2] B. Gipson, D. Hsu, L. E. Kavraki, J.-C. Latombe, Computational Models of Protein Kinematics and Dynamics: Beyond Simulation, Annu. Rev. Anal. Chem., 5 (2012) 273-291.

[3] W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Computer simulation of biomolecular systems: theoretical and experimental applications, Vol. 3. Springer Science & Business Media (2013).

[4] A. Bondi, van der Waals volumes and radii, The Journal of physical chemistry, 68(3) (1964) 441-451.

[5] G. E. Schulz and R. H. Schirmer, Principles of protein structure, Springer Science & Business Media (2013).

[6] K. –J. Kim, E. Sacks, L. Joskowicz, Kinematic analysis of spatial fixed-axis higher pairs using configuration spaces, Computer-Aided Design 35(3) (2003) 279-291.