

# 1 Rotation matrices

In three dimensions, we can rotate any set of points with respect to the  $x$ ,  $y$ , or  $z$  axes. Thus, given a set of  $n$  points  $d = \{p_1, \dots, p_n\}$ , we can obtain any orientation by performing a rotation along each of the three axes. We consider each point to be in the form of a three-dimensional vector  $p = [x, y, z]^T$ . Rotating by some angle  $\theta$  along an arbitrary axis (here, the  $x$ -,  $y$ -, or  $z$ -axis) can be performed using a 3x3 rotation matrix  $R(\theta)$ , where the new point  $P$  is given by  $P = R(\theta) \cdot p$ . For the  $x$ -,  $y$ -, and  $z$ -axis, we have

$$R_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \quad (1)$$

$$R_y(\theta_y) = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \quad (2)$$

$$R_z(\theta_z) = \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Combining the three rotation matrices, such that  $R(\theta_x, \theta_y, \theta_z) = R_z(\theta_z) \cdot R_y(\theta_y) \cdot R_x(\theta_x)$ , we have

$$\begin{aligned} R &= R_z(\theta_z) \cdot R_y(\theta_y) \cdot R_x(\theta_x) \\ &= \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \\ &= \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_y & \sin \theta_x \sin \theta_y & \cos \theta_x \sin \theta_y \\ 0 & \cos \theta_x & -\sin \theta_x \\ -\sin \theta_y & \sin \theta_x \cos \theta_y & \cos \theta_x \cos \theta_y \end{bmatrix} \\ &= \begin{bmatrix} \cos \theta_y \cos \theta_z & \sin \theta_x \sin \theta_y \cos \theta_z - \cos \theta_x \sin \theta_z & \cos \theta_x \sin \theta_y \cos \theta_z + \sin \theta_x \sin \theta_z \\ \cos \theta_y \sin \theta_z & \sin \theta_x \sin \theta_y \sin \theta_z + \cos \theta_x \cos \theta_z & \cos \theta_x \sin \theta_y \sin \theta_z - \sin \theta_x \cos \theta_z \\ -\sin \theta_y & \sin \theta_x \cos \theta_y & \cos \theta_x \cos \theta_y \end{bmatrix} \end{aligned} \quad (4)$$

Below, we define  $r_x$ ,  $r_y$ , and  $r_z$  to be the three row vectors of  $R$ , such that

$$R = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \quad (5)$$

## 2 Structural alignments

Our goal is to determine, for two similar macromolecule molecules, the orientation in 3-dimensional space at which the structures align in an optimal way. The approach is to minimize the sum of the squares between the corresponding atoms. We are given two sets of points  $d = \{p_1, \dots, p_n\}$  and  $D = \{P_1, \dots, P_n\}$ , where each  $p_i$  roughly corresponds to  $P_i$ . Each point can be thought of as a three-dimensional vector where  $p_i = [x_i, y_i, z_i]^T$  and  $P_i = [X_i, Y_i, Z_i]^T$ . We can define an isometry  $f(P) = [f_x(P), f_y(P), f_z(P)]^T$  that outputs the  $x$ ,  $y$ , and  $z$  coordinates of a new point after translating and rotating point  $P$  by some amount. Thus, we have that

$$f_x(P) = r_x \cdot P + c_x \quad (6)$$

$$f_y(P) = r_y \cdot P + c_y \quad (7)$$

$$f_z(P) = r_z \cdot P + c_z \quad (8)$$

where  $c_x$ ,  $c_y$ , and  $c_z$  are scalar parameters, and  $r \cdot P$  is the dot product of 3 dimensional vectors  $r$  and  $P$ . Our goal, then, is to determine the isometry  $f$  that minimizes the sum of the squares

$$S = \sum_{i=1}^n \sqrt{(x_i - f_x(P_i))^2 + (y_i - f_y(P_i))^2 + (z_i - f_z(P_i))^2} \quad (9)$$

Differentiating  $S$  with respect to  $\theta_j$ , where  $\theta_j \in \{\theta_x, \theta_y, \theta_z, c_x, c_y, c_z\}$ , gives

$$\frac{\partial S}{\partial \theta_j} = - \sum_{i=1}^n \frac{\frac{\partial f_x(P_i)}{\partial \theta_j}(x_i - f_x(P_i)) + \frac{\partial f_y(P_i)}{\partial \theta_j}(y_i - f_y(P_i)) + \frac{\partial f_z(P_i)}{\partial \theta_j}(z_i - f_z(P_i))}{\sqrt{(x_i - f_x(P_i))^2 + (y_i - f_y(P_i))^2 + (z_i - f_z(P_i))^2}} \quad (10)$$

and differentiating with respect to both  $\theta_j$  and  $\theta_k$  gives

$$\begin{aligned} \frac{\partial^2 S}{\partial \theta_j \partial \theta_k} = & -2 \sum_{i=1}^n \frac{\partial^2 f_x(P_i)}{\partial \theta_j \partial \theta_k}(x_i - f_x(P_i)) + \frac{\partial^2 f_y(P_i)}{\partial \theta_j \partial \theta_k}(y_i - f_y(P_i)) + \frac{\partial^2 f_z(P_i)}{\partial \theta_j \partial \theta_k}(z_i - f_z(P_i)) \\ & + 2 \sum_{i=1}^n \frac{\partial f_x(P_i)}{\partial \theta_j} \frac{\partial f_x(P_i)}{\partial \theta_k} + \frac{\partial f_y(P_i)}{\partial \theta_j} \frac{\partial f_y(P_i)}{\partial \theta_k} + \frac{\partial f_z(P_i)}{\partial \theta_j} \frac{\partial f_z(P_i)}{\partial \theta_k} \end{aligned} \quad (11)$$

Here, instead of the second derivative  $\partial^2 S / \partial \theta_j \partial \theta_k$  we use its expected value, which is only the second summation in Eq 11, giving

$$\frac{\partial^2 S}{\partial \theta_j \partial \theta_k} = 2 \sum_{i=1}^n \frac{\partial f_x(P_i)}{\partial \theta_j} \frac{\partial f_x(P_i)}{\partial \theta_k} + \frac{\partial f_y(P_i)}{\partial \theta_j} \frac{\partial f_y(P_i)}{\partial \theta_k} + \frac{\partial f_z(P_i)}{\partial \theta_j} \frac{\partial f_z(P_i)}{\partial \theta_k} \quad (12)$$

Taking the derivative of  $f_x$ ,  $f_y$ , and  $f_z$  is straight forward. For  $a, b \in \{x, y, z\}$ , we have

$$\frac{\partial f_a(P)}{\partial c_b} = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a \end{cases} \quad (13)$$

For the rotation angle parameters...

$$\frac{\partial f_x(P)}{\partial \theta_x} = (\cos \theta_x \sin \theta_y \cos \theta_z + \sin \theta_x \sin \theta_z) Y + (-\sin \theta_x \sin \theta_y \cos \theta_z + \cos \theta_x \sin \theta_z) Z \quad (14)$$

$$\frac{\partial f_x(P)}{\partial \theta_y} = (-\sin \theta_y \cos \theta_z) X + (\cos \theta_y \sin \theta_x \cos \theta_z) Y + (\cos \theta_y \cos \theta_x \cos \theta_z) Z \quad (15)$$

$$\frac{\partial f_x(P)}{\partial \theta_z} = (-\sin \theta_z \cos \theta_y) X + (-\sin \theta_z \sin \theta_x \sin \theta_y - \cos \theta_z \cos \theta_x) Y + (-\sin \theta_z \cos \theta_x \sin \theta_y + \cos \theta_z \sin \theta_x) Z \quad (16)$$

$$\frac{\partial f_y(P)}{\partial \theta_x} = (\cos \theta_x \sin \theta_y \sin \theta_z - \sin \theta_x \cos \theta_z) Y + (-\sin \theta_x \sin \theta_y \sin \theta_z - \cos \theta_x \cos \theta_z) Z \quad (17)$$

$$\frac{\partial f_y(P)}{\partial \theta_y} = (-\sin \theta_y \sin \theta_z) X + (\cos \theta_y \sin \theta_x \sin \theta_z) Y + (\cos \theta_y \cos \theta_x \sin \theta_z) Z \quad (18)$$

$$\frac{\partial f_y(P)}{\partial \theta_z} = (\cos \theta_z \cos \theta_y) X + (\cos \theta_z \sin \theta_x \sin \theta_y - \sin \theta_z \cos \theta_x) Y + (\cos \theta_z \cos \theta_x \sin \theta_y + \sin \theta_z \sin \theta_x) Z \quad (19)$$

$$\frac{\partial f_z(P)}{\partial \theta_x} = (\cos \theta_x \cos \theta_y) Y + (-\sin \theta_x \cos \theta_y) Z \quad (20)$$

$$\frac{\partial f_z(P)}{\partial \theta_y} = (-\cos \theta_y) X + (-\sin \theta_y \sin \theta_x) Y + (-\sin \theta_y \cos \theta_x) Z \quad (21)$$

$$\frac{\partial f_z(P)}{\partial \theta_z} = 0 \quad (22)$$

### 3 Molecular free energy

To determine the structure of any given protein molecule, we minimize the function  $\Delta G$  where

$$\Delta G = W_{atr}A + W_{rep}R + W_z(d - d_0)^2 \quad (23)$$

where  $A$  and  $R$  measure the inter-residue attractive and repulsive atomic forces, respectively, and  $Z$  represents the ionic bond energies of the zinc ion in the protein structure.  $W$ -values represent the weights attributed to components  $A$  and  $R$ , which are set to  $W_{atr} = 0.777$  and  $W_{rep} = 0.627$ , as in Havernek et al (2004).

The functional form and parameter values of the attractive and repulsive forces  $A$  and  $R$  were derived in Havernek et al (2004). These functions partition atoms into 25 categories, where each category is associated with a single value for the atomic radii ( $r$ ) and well depth ( $e$ ). Given the distance ( $d_{ij}$ ) between two atoms  $i$  and  $j$  and the sum ( $r_{ij}$ ) of the two atomic radii ( $r_{ij} = r_i + r_j$ ), the values for  $A$  and  $R$  for a given structure is then calculated as

$$A = \begin{cases} \sum_i \sum_{j>i} \sqrt{e_i e_j} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] & \text{if } \frac{r_{ij}}{d_{ij}} < 1.12 \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

and

$$R = \begin{cases} \sum_i \sum_{j>i} \sqrt{e_i e_j} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] & \text{if } 1.33 > \frac{r_{ij}}{d_{ij}} > 1.12 \\ \sum_i \sum_{j>i} \sqrt{e_i e_j} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] + \sum_i \sum_{j>i} [y_{ij} - b_{ij} d_{ij}] & \text{if } \frac{r_{ij}}{d_{ij}} > 1.33 \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

where

$$b_{ij} = -12 \frac{\sqrt{e_i e_j}}{r_{ij}} \quad (26)$$

$$y_{ij} = -b_{ij} \left( \frac{r_{ij}}{1.33} \right) + \sqrt{e_i e_j} (1.33^{12} - 2 \cdot 1.33^6) \quad (27)$$

The values for the atomic radii and well depths are given in Table 1.

The value  $P$  is used to promote higher-probability torsions of the  $\phi$  and  $\psi$  angles of each amino acid within the polymer.  $P$  has the functional form

$$P = \sum_k \left[ \frac{V_k[1]}{2} [1 + \cos(\phi_k - \gamma)] + \frac{V_k[2]}{2} [1 + \cos(2\phi_k - \gamma)] + \frac{V_k[3]}{2} [1 + \cos(3\phi_k - \gamma)] \right] \quad (28)$$

$$+ \sum_k \left[ \frac{U_k[1]}{2} [1 + \cos(\phi_k - \gamma)] + \frac{U_k[2]}{2} [1 + \cos(2\phi_k - \gamma)] + \frac{U_k[3]}{2} [1 + \cos(3\phi_k - \gamma)] \right] \quad (29)$$

where the sum is over all amino acids  $k$  within the polymer. The values for  $V_k$  and  $U_k$  are parameters unique to each type of amino acid, and are optimized according to empirical data.

The component  $Z$  incorporates bond interactions between the zinc ion and the 2 cysteine and 2 histidine amino acids.

Atom types				
Atom type	Description	Atom	Radius ( $r$ )	Well depth ( $e$ )
1	Side chain carbonyl C	C	2.00	0.1200
2	Side chain carboxyl C	C	2.00	0.1200
3	Aliphatic C with 1 H	C	2.00	0.0486
4	Aliphatic C with 2 H	C	2.00	0.1142
5	Aliphatic C with 3 H	C	2.00	0.1811
6	Aromatic ring C	C	2.00	0.1200
7	Protonated N	N	1.75 <sup>1</sup>	0.2384
8	Deprotonated N	N	1.75 <sup>1</sup>	0.2384
9	Neutral NH <sub>2</sub> Nitrogen	N	1.75 <sup>1</sup>	0.2384
10	Lysine N	N	1.75 <sup>1</sup>	0.2384
11	Arginine N	N	1.75 <sup>1</sup>	0.2384
12	Proline N	N	1.75 <sup>1</sup>	0.2384
13	Hydroxyl O	O	1.55 <sup>1,2</sup>	0.1591
14	Side chain carbonyl O	O	1.55 <sup>2</sup>	0.1591
15	Side chain carboxyl O	O	1.55 <sup>2</sup>	0.2100
16	Sulfur	S	1.90	0.1600
17	Backbone N	N	1.75	0.2384
18	Backbone C $\alpha$	C	2.00	0.0486
19	Backbone carbonyl C	C	2.00	0.1400
20	Backbone O	O	1.55	0.1591
21	Zinc	Zn	—	—
22	Polar H	H	0.10	0.0500
23	Nonpolar H	H	0.70	0.0500
24	Aromatic H	H	0.70	0.0500
25	Backbone HN	H	0.10	0.0500

Table 1: Atom types and their corresponding atomic radii ( $r$ ) and well depths ( $e$ ). All values were determined as in Havernek et al 2004. Atomic radii and well depths for zinc were not calculated by Havernek et al. <sup>1</sup> Hydrogen bond acceptors. <sup>2</sup> Hydrogen bond donors. When hydrogen bond acceptors are paired with hydrogen bond donors, the value  $r_{ij} = r_i + r_j$  is set to 2.95.