

Research Article

Transcriptional Protein-Protein Cooperativity in POU/HMG/DNA Complexes Revealed by Normal Mode Analysis

Debby D. Wang and Hong Yan

Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong

Correspondence should be addressed to Debby D. Wang; debby.d.wang@gmail.com

Received 7 August 2013; Accepted 22 September 2013

Academic Editor: Ao Yuan

Copyright © 2013 D. D. Wang and H. Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Biomolecular cooperativity is of great scientific interest due to its role in biological processes. Two transcription factors (TFs), Oct-4 and Sox-2, are crucial in transcriptional regulation of embryonic stem cells. In this paper, we analyze how Oct-1 (a similar POU factor) and Sox-2, interact cooperatively at their enhancer binding sites in collective motions. Normal mode analysis (NMA) is implemented to study the collective motions of two complexes with each involving these TFs and an enhancer. The special structure of Oct proteins is analyzed comprehensively, after which each Oct/Sox group is reassembled into two protein pairs. We subsequently propose a segmentation idea to extract the most correlated segments in each pair, using correlations of motion magnitude curves. The median analysis on these correlation values shows the intimacy of subunit POU (Oct-1) and Sox-2. Using those larger-than-median correlation values, we conduct statistical studies and propose several protein-protein cooperative modes (*S* and *D*) coupled with their subtypes. Additional filters are applied and similar results are obtained. A supplementary study on the rotation angle curves reaches an agreement with these modes. Overall, these proposed cooperative modes provide useful information for us to understand the complicated interaction mechanism in the POU/HMG/DNA complexes.

1. Introduction

Embryonic stem cells (ES cells) possess the pluripotency of differentiating into all the three germ layers (endoderm, mesoderm, and ectoderm), which correspond to hundreds of cell types. These pluripotent stem cells are transcriptionally regulated by a number of transcription factors (TFs) [1]. A specific TF called Oct-4, belonging to the POU class of homeodomain proteins, is regarded as a necessity for maintaining the undifferentiated state of embryonic ES cells. Generally, Oct-4 interacts with other TFs as a group to affect the gene expression of mouse ES cells in early embryo development [2], and Oct-4 coupled with its cofactor Sox-2 (HMG-box domain) is at the center of this group. Botquin and Nishimoto have both proven the cooperative effects of Oct-4 and Sox-2 on the expression of several genes in mouse embryonic ES cells [3, 4]. Dailey and Basilico further bring forward the idea that the interaction within the POU/HMG group, especially for groups composed of Oct and Sox proteins, at DNA binding sites is a fundamental mechanism for transcriptional regulation in early embryo development [5].

At the early stage of transcription, TFs bind to specific regulatory DNA regions to cooperatively affect the transcription sites. Enhancers, which act as activators or stimulators for transcription [6], are a major type of regulatory DNA regions. Unlike promoters, enhancers may be located kilobases away from their target genes, but geometrically they are most probably close to the genes due to the supercoil structure of DNA molecules, and thus there can be direct contacts between the enhancer-TF complexes and the transcription sites. Studies on the enhancer-TF complexes are very important for understanding the complicated mechanism of transcriptional regulation.

On the other hand, molecular dynamics are involved in many biological processes [7, 8], such as reproduction, regulation of gene expression, and protein interaction. As an indispensable component of gene expression, transcription must undergo a series of dynamical changes of biomolecules. Therefore, studies on dynamics of the aforementioned enhancer-TF complexes would provide a deep insight into their properties and functions in the transcriptional regulation. Specifically, deciphering the roles of Oct and Sox in

the interaction mechanism of their enhancer-bounded complexes in the collective dynamics is of great scientific interest. Further, the cooperativity of the two proteins is a major research topic in these studies.

In our work, the dynamics of the POU/HMG group at its enhancer binding sites, referred to as POU/HMG/DNA complexes, are surveyed. Two POU/HMG/DNA complexes, which are DNA-binding portions of a POU factor Oct-1 and an HMG factor Sox-2 bound to an enhancer, are specifically studied from a structural and molecular dynamic view. Normal mode analysis (NMA) is implemented to study the collective or cooperative motions of these POU/HMG/DNA ternary complexes, after which the interaction of the POU and HMG factors at their DNA binding sites in these collective motions is explored. We propose a segmentation idea for the proteins to construct an equal-length-chain comparison and measure the correlation of each protein segment pair using the linear correlation. A statistical analysis on the significantly correlated pairs provides useful information on how these TFs have a synergistic control on enhancer DNAs in transcriptional regulation.

2. Materials and Method

2.1. Normal Mode Analysis (NMA)

2.1.1. Introduction. NMA is an efficient method to detect the most cooperative or collective motions (essential modes) of large harmonic oscillating systems. With the constraint that the studied conformations are in the vicinity of the systematic equilibrium, which exists in most harmonic oscillating systems [9], NMA is useful for studying large structural deformations or motions of these systems. The idea is to use harmonic potentials to approximate a multidimensional energy landscape around an energy minimum for a system and to detect the most easily accessible modes on this energy landscape. NMA is broadly used to analyze the structural dynamics of biomolecules.

Specifically, if we describe an N -site-system with a position vector q , in which the three-dimensional coordinates of each site $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_N, y_N, z_N)$ are used, we can mathematically expand the potential energy V in a second-order Taylor series around the equilibrium conformation q^0 [9]. Finally, we obtain a quadratic approximation as follows:

$$V(q) = \frac{1}{2} \sum_{i,j} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0 (q_i - q_i^0) (q_j - q_j^0) = \frac{1}{2} \Delta q^T H \Delta q. \quad (1)$$

Here Δq stands for the systematic structural changes relative to q^0 , and H is a $3N \times 3N$ Hessian matrix, whose elements have the following form:

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j}. \quad (2)$$

Subsequently, the kinetic energy is brought in to slightly modify the Hessian to a mass-weighted one. These Hessian

matrices contain key structural information for our observed systems.

One broadly used construction method for the Hessian matrices is the elastic network models (ENMs) [9–12], which include the Gaussian network models (GNMs) [11] and anisotropic network models (ANMs) [12] as representatives. When ENM is applied, the equilibrium exploration can be skipped since the starting state is designed for this equilibrium. When constructing the ENM structure, the original system can be transformed into a network with nodes (CG-sites) and connecting springs, and a cutoff distance γ_c is used to define all the connecting springs γ_{ij} [9, 10]. Gaussian network model (GNM) selects representatives for substructures in the system, such as using C_α -atoms for amino acids [9, 11], to further lower the computational cost, leading to the potential form shown as (3) (R_i or R_j represents a CG-site):

$$V_{\text{GNM}} = \frac{1}{2} \sum_{i,j} \gamma_{ij} (\Delta R_i - \Delta R_j)^2. \quad (3)$$

Similarly, ANM proposes the potential form in (4) and ignores some influences caused by the distance vectors:

$$V_{\text{ANM}} = \frac{1}{2} \sum_{i,j} \gamma_{ij} (\|R_i^0 + \Delta R_i - R_j^0 - \Delta R_j\| - \|R_i^0 - R_j^0\|)^2. \quad (4)$$

Each eigenvalue of an above-constructed Hessian matrix denotes the associated systematic energy for the observed system, and its corresponding eigenvector represents the direction of a specific normal mode motion. Among the obtained $3N$ normal mode directions, the first six are trivial since they all correspond to zero eigenvalues, which means these structural changes have no effect on the systematic potential energy. For the remaining $3N - 6$ eigenvectors, we will select a small set that corresponds to small eigenvalues (essential modes) for analysis [9]. In previous research, the first 10~15 essential modes are chosen by many researchers for their work [13–15], and the first 10 are analyzed in our work.

2.1.2. Computational Platform. Several online tools are available for normal mode calculations. An online server called NOMAD-Ref at <http://lorenz.immstr.pasteur.fr/nomad-ref.php> [16] is utilized in our experiments. It is an ENM model-based method. The implementation of a rotation-translation block approach [16] and an ARPACK library for the sparse matrix data storage and decomposition [17] in the computations of Hessian matrices makes it possible to retain up to 100,000 atoms for each structure. In our work, when calculating the motions using NOMAD-Ref, all atoms in POU/HMG/DNA ternary complexes are used, while only motions of the POU and HMG proteins are analyzed since only protein-protein interactions in POU/HMG complexes at the DNA binding sites are of interest here.

2.2. Experimental Data and the Analysis on Their NMA Results

2.2.1. Experimental Data. Two POU/HMG/DNA ternary complexes, 1GT0 and 1O4X, are downloaded from the Protein Data Bank (PDB) [18] for analysis. Each structure is

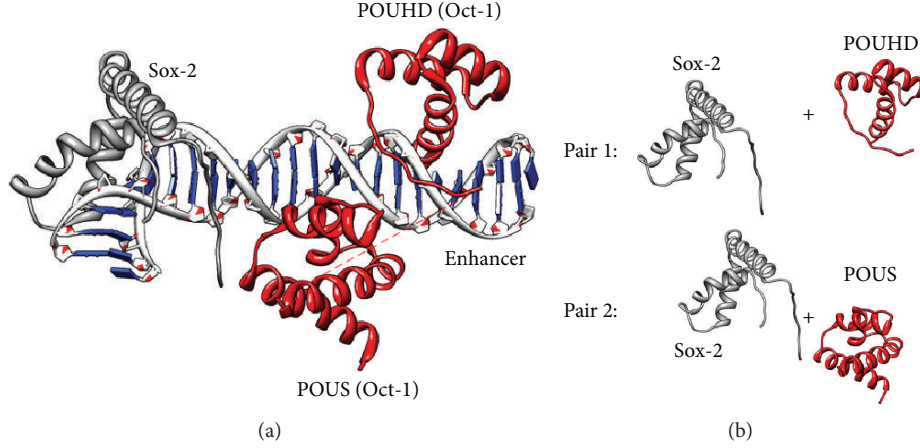


FIGURE 1: (a) The 3D structure of the POU/HMG/DNA ternary complex 1GT0. The gray protein represents an HMG factor Sox-2, and the red one is a POU factor Oct-1, which is composed of two subunits POUHD and POUS. (b) The two reassembled protein pairs, originated from (a), for our subsequent studies.

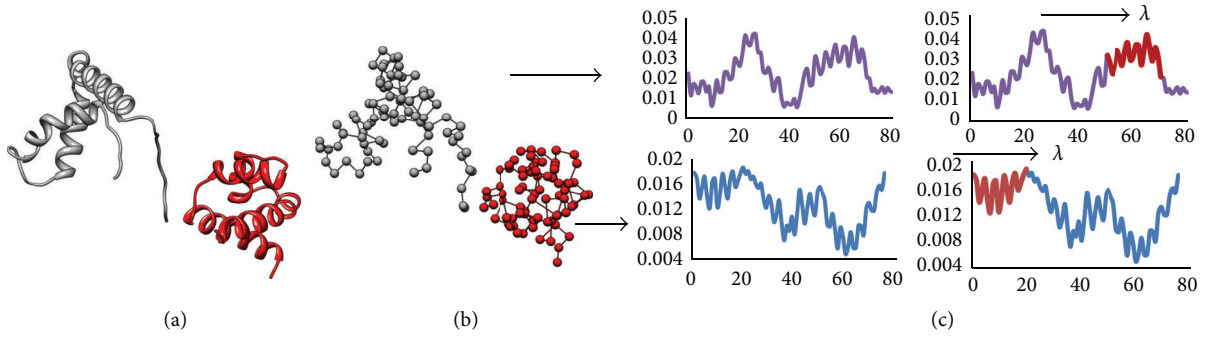


FIGURE 2: (a) The protein pair containing Sox-2 (gray) and POUS (red) in 1GT0. (b) The refined structure of the protein pair, where nodes represent residues. In each mode, a refined protein structure in the pair corresponds to a motion magnitude curve. (c) The searching process for the most cooperative segments of length λ in the protein pair in a specific mode.

composed of a POU factor Oct-1 (very similar to Oct-4), an HMG factor Sox-2 and an enhancer element. Figure 1(a) displays the 3D structure of complex 1GT0 and the diagram is produced using UCSF Chimera [19]. In 1GT0, the bounded DNA piece is a fibroblast growth factor 4 enhancer (FGF4) [20]; in 1O4X, the homeobox B1 (Hoxb1) enhancer is bounded by the two TFs [21].

Furthermore, each Oct protein contains two subunits (POUS and POUHD) that are connected by a flexible linker and control DNAs in a bipartite manner [21]. Based on the special structure of Oct proteins, we regard Oct-1 and Sox-2 in each complex as the two protein pairs for further investigation, namely, POUHD and Sox-2 as pair 1 and POUS and Sox-2 as pair 2, both of which are shown in Figure 1(b).

2.2.2. Analysis of Correlative Motions. After generating the motions of the two POU/HMG/DNA ternary complexes using NMA, we observe how the two protein pairs behave at the enhancer binding sites in these most collective or cooperative motions.

For each protein pair in each ternary complex, we analyze the first 10 obtained essential modes. In each mode, we firstly

refine an observed pair at the residue level from a view of motion magnitude. This can be achieved by calculating the motion magnitudes for all the atoms in each protein and subsequently computing the motion magnitude of each residue in this protein by averaging the motion magnitudes of all component atoms (see (5)):

$$\begin{aligned}
 MR_i &= \frac{1}{N} \sum_{j=1}^{N_i} MA_{ij} \\
 &= \frac{1}{N} \sum_{j=1}^{N_i} \sqrt{(x_{ij} - x_{ij}^0)^2 + (y_{ij} - y_{ij}^0)^2 + (z_{ij} - z_{ij}^0)^2}.
 \end{aligned} \tag{5}$$

Here atoms $j = 1 \sim N_i$ comprise the residue i ; $(x_{ij}^0, y_{ij}^0, z_{ij}^0)$ and (x_{ij}, y_{ij}, z_{ij}) represent the positions of atom j in its equilibrium position and in a specific mode, respectively. Therefore, for each mode, we will obtain a motion magnitude curve for each protein in an observed pair, and each curve point corresponds to a residue along the protein sequence (Figures 2(a) and 2(b)).

Next, in each protein pair we observe the potential protein-protein cooperativity in these motions based on the correlations of motion magnitude functions. An effective method to measure the dependence between two quantities is the Pearson product-moment correlation coefficient [22–24] also is usually called the correlation coefficient. This coefficient is calculated based on the expected values (μ) and standard deviations (σ) of the two variables (\mathbf{X} and \mathbf{Y}), as shown in (6):

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \quad (6)$$

We adopt this correlation coefficient in our studies. However, since each protein has a different length, we investigate the most cooperative/correlated segments among each protein pair in each mode. We introduce a segment length parameter λ here. For an observed pair of proteins that have different lengths of x and y , with a specific λ ($\lambda \leq x < y$) defined in a mode, we shift one motion magnitude function along the other to find the λ -length-segments which share the largest absolute correlation value (Figure 2(c)). We could further describe the process as follows:

$$\begin{aligned} \max_{ij} C(i, j) &= |\text{corr}(\text{Seg}_i, \text{Seg}_j)| \\ \text{So that } \begin{cases} \text{Seg}_i \in F_1 \\ \text{Seg}_j \in F_2 \\ \|\text{Seg}_i\| = \|\text{Seg}_j\| = \lambda. \end{cases} \end{aligned} \quad (7)$$

Here F_1 and F_2 represent motion magnitude functions of the two proteins in an observed pair, in a specific essential mode; Seg_i and Seg_j denote λ -length-segments of F_1 and F_2 , respectively.

In each modes with a list of λ values defined for each protein pair, we obtain a series of most cooperative segment pairs having correlation values c_{mn} , where m denotes different λ values and n (1~10) represents different modes. Here we replace λ by $p = \lambda/x$ for easier illustration and x is the shorter length in the observed pair. Since larger absolute value of correlation demonstrates more correlated segments (positively or negatively), we investigate how $|c_{mn}|$ distribute for the two protein pairs in each complex. For each p in an observed pair, the median value (8) is extracted and explored. Furthermore, the performances (based on \tilde{c}_m) of the two pairs in each complex are compared:

$$\begin{aligned} \tilde{c}_m &= \text{MED}_{n=1}^{10} \{|c_{mn}|\} = \text{MED} \{|c_{m1}|, |c_{m2}|, \dots, |c_{m10}|\}, \\ & \quad m = 1, \dots, 6. \end{aligned} \quad (8)$$

Now, we use medians in (8) as a filter and investigate how those $|c_{mn}|$ larger than \tilde{c}_m (supposed to be significant)

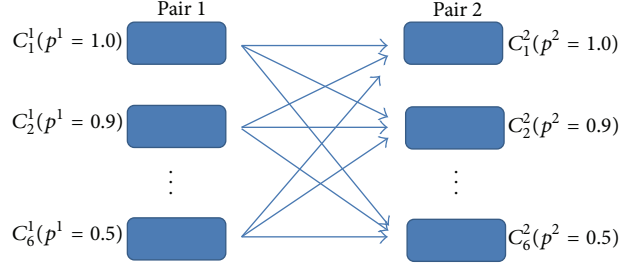


FIGURE 3: This diagram shows all the possibilities of the length pair (p^1, p^2). Specifically, for an observed complex, the significantly correlated segment pair (with a length parameter of p^1) in protein pair 1 and that (with a length parameter of p^2) in protein pair 2 are comparably investigated, with p^1 and p^2 taking all possible values ($m = 1 \sim 6$) as previously stated.

distribute. For each protein in an observed pair, we can obtain a logic matrix L that reflects this process:

$$\begin{aligned} L &= [l_{mn}] = ([|c_{mn}|] > [\tilde{c}_m]?) \\ &= ([C_m] > [\tilde{c}_m]?) = \left(\left(\begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_6 \end{pmatrix} \right) > \left(\begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \vdots \\ \tilde{c}_6 \end{pmatrix} \right) ? \right) \quad (9) \\ \text{So that } \begin{cases} C_m = (|c_{m1}|, |c_{m2}|, \dots, |c_{m10}|) \\ m = 1, \dots, 6 \\ n = 1, \dots, 10. \end{cases} \end{aligned}$$

Here $[x]$ gives the matrix that is composed of x .

We subsequently examine the relationship between the two protein pairs in each complex based on these logic matrices. The idea is to explore that in a single essential mode whether only one significantly correlated segment pair (either in protein pair 1 or pair 2) is involved or both pairs are involved. To balance the segment lengths (p) used by the two pairs, we take into consideration all the length pairs (p^1, p^2) between the two pairs, as presented in Figure 3. Here we use superscripts to distinguish pairs 1 and 2.

To fulfill the aforementioned operation, we conduct several iterations for all the p^j values and combine the results of these iterations. We now take rows in L^1 (denoting a specific p^1 value) and show how the whole procedure is accomplished. In each iteration, we firstly expand the involved row (identified with a subscript m) into a matrix in (10) and then carry out statistics on the cases where three situations occur: (a) S_1 (index s_1)—only the significantly correlated segment pair in pair 1 is detected in a single essential mode with a length pair (p^1, p^2), (b) S_1 (index s_2)—only the significantly correlated segment pair in pair 2 is detected, and (c) D (index d)—both pairs 1 and 2 are detected. The statistical analysis is based on logic operations, as shown in (11), which combines all the iterations to derive the final indexes for the two pairs:

$$L^{1,m} = \left(\frac{L_m^1, L_m^2, \dots, L_m^6}{6} \right)^T, \quad (10)$$

$$\begin{aligned}
s_1 &= \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times (\neg L^2)), \\
s_2 &= \sum_{m=1}^6 \text{sum} ((\neg L^{1,m}) \cdot \times L^2), \\
d &= \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times L^2).
\end{aligned} \tag{11}$$

Here “ $\cdot \times$ ” means a batch of multiplications of the corresponding elements in two matrices, and $\text{sum}(X)$ counts the number of ones (a logic “true” value) in a logic matrix X . Indexes s_1 , s_2 , and d separately show three cooperative modes (corresponding to the aforementioned three cases) between the two protein pairs in a POU/HMG/DNA complex. We exhibit some representative cooperative modes in Section 3, where we also list the above-mentioned indexes for the two complexes. Furthermore, to take the signs of the correlations c_{mm} into consideration, we introduce another logic matrix Z that describes the signs of c_{mm} , as stated in (12). Through combining logic operations of L^i and Z^i (13), we can divide the situations (S_1 , S_2 , and D) into subtypes (positive and negative), and all these subtypes are analyzed in Section 3:

$$Z = [z_{mn}] = ([c_{mn}] > [0]?) \quad \text{So that } \begin{cases} m = 1, \dots, 6 \\ n = 1, \dots, 10, \end{cases} \tag{12}$$

$$Z^{1,m} = \underbrace{(Z_m^1, Z_m^1, \dots, Z_m^1)^T}_6,$$

$$s_{1,\text{positive}} = \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times (\neg L^2) \cdot \times Z^{1,m})$$

$$s_{1,\text{negative}} = \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times (\neg L^2) \cdot \times (\neg Z^{1,m})),$$

$$s_{2,\text{positive}} = \sum_{m=1}^6 \text{sum} ((\neg L^{1,m}) \cdot \times L^2 \cdot \times Z^2),$$

$$s_{2,\text{negative}} = \sum_{m=1}^6 \text{sum} ((\neg L^{1,m}) \cdot \times L^2 \cdot \times (\neg Z^2)),$$

$$d_{\text{positive}}$$

$$= \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times L^2 \cdot \times [Z^{1,m} \cdot \times Z^2 + (\neg Z^{1,m})] \cdot \times (\neg Z^2)),$$

$$d_{\text{negative}}$$

$$= \sum_{m=1}^6 \text{sum} (L^{1,m} \cdot \times L^2 \cdot \times [Z^{1,m} \cdot \times (\neg Z^2) + (\neg Z^{1,m}) \cdot \times Z^2]). \tag{13}$$

To compare the scenarios where different filters are applied, we, respectively, apply the first tertile, the first quartile, and the mean value as filters to investigate the corresponding results. The mean filter can be described as (14), and the quantile filter as (15), where Pr represents probability.

Specifically, the tertile and quartile filters correspond to situations where $p = 1/3$ and $p = 1/4$, respectively. A series of operations are then carried out based on these filters, to reveal how the observed complexes behave in these situations:

$$\bar{c}_m = \frac{\sum_{n=1}^{10} \{|c_{mn}|\}}{10}, \quad m = 1, \dots, 6, \tag{14}$$

$$\tilde{c}_m(p) = \inf \{|c_{mn}| \mid F(|c_{mn}|) \geq p\},$$

$$F(|c_{mn}|) = \text{Pr}(C_m \leq |c_{mn}|), \tag{15}$$

$$m = 1, \dots, 6.$$

Finally, to gain a deep insight into the motions of these two complexes, we have also observed the rotation angles of the corresponding protein chains. In the above discussion, we regard residues as basic units in protein sequences, and here we consider the links between each consecutive two residues (Figure 2(b)). The angles between each pair of corresponding links in the original structure and in deformed structures (modes) are studied. We obtain a rotation angle function for each protein of a protein pair in each essential mode. Afterwards, we conduct a similar analysis as aforementioned on these rotation angle functions as a supplementary study. Principal component analysis (PCA) is implemented to reduce the effect of noisy rotation angles. We also investigate the suitability of the Fourier transform for data analysis.

3. Results and Discussion

3.1. Motion Magnitude Functions. For each protein in an observed pair of a ternary complex, we calculate the motion magnitude functions (5) for the first 10 essential modes. Figure 4 shows the motion magnitude curves for the two observed protein pairs in 1GT0 for the first essential mode.

After defining a list of p values, we calculate the most cooperative/correlated segment pairs among each protein pair in a complex for the 10 essential modes, using the mechanism discussed in Section 2.2.2. Since small p values correspond to shorter segment matching, whose results may be trivial due to the high correlation possibilities, we use a set of p values starting at 0.5 to 1.0 at a step of 0.1. Table 1 shows the results of correlations c_{mm} for the most cooperative segment pairs among protein pair 1 of 1GT0.

The larger the absolute value of correlation is, the more the two compared segments correlate with each other, either positively or negatively. Now we examine how the absolute correlation values $|c_{mn}|$ distribute, for the two protein pairs in each complex. The values are presented in Figure 5, where parts (a) and (b), respectively, show the values for the two pairs in 1GT0, and parts (d) and (e) show those for 1O4X. We can see from these diagrams that $|c_{mn}|$ becomes larger when p gets smaller, and this can also be detected from the median value \tilde{c}_m shown with a pink circle in each box (denoting a specific p). To give a comparison between the performances of the two pairs in each complex, we extract the above-mentioned median values \tilde{c}_m for each pair and present them in parts (c) (1GT0) and (f) (1O4X). In diagrams (c) and (f), especially (f), pair 2 presents a higher \tilde{c}_m than pair 1, which

TABLE 1: Motion correlations between POUHD and Sox-2 in protein pair 1 of IGT0.

P	Mode									
	Mode 7	Mode 8	Mode 9	Mode 10	Mode 11	Mode 12	Mode 13	Mode 14	Mode 15	Mode 16
1	-0.696	0.606	-0.477	0.324	-0.265	0.383	0.202	-0.326	-0.382	-0.520
0.9	-0.738	0.711	0.697	0.340	-0.419	0.772	0.454	-0.429	-0.420	0.739
0.8	-0.853	0.797	0.819	0.342	0.651	0.838	-0.607	-0.463	0.569	0.730
0.7	-0.859	0.836	0.820	-0.445	0.639	0.820	-0.621	-0.562	-0.716	0.769
0.6	-0.856	0.851	0.850	0.537	0.814	0.849	-0.699	-0.684	0.762	0.806
0.5	-0.865	-0.862	0.856	-0.757	0.858	-0.901	-0.733	-0.761	0.806	0.819

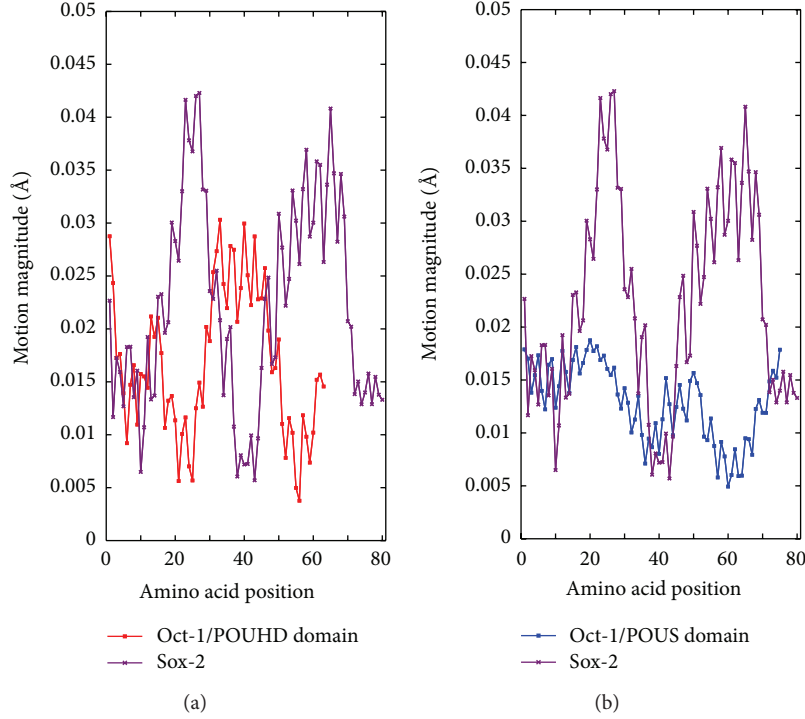


FIGURE 4: (a) The motion magnitude curves in mode 7 for proteins POUHD (red) and Sox-2 (purple) in pair 1 of IGT0. (b) The motion magnitude curves in mode 7 for proteins POUS (blue) and Sox-2 (purple) in pair 2 of IGT0.

to some extent implies that pair 2 may behave as a leading role in the Oct/Sox interactions.

Next, we use the above-mentioned medians as a filter and investigate how those $|c_{mm}|$ larger than \bar{c}_m (supposed to be significant) distribute. For each protein in an observed pair, we calculate its logic matrices L and Z (Section 2.2.2), which correspond to (9) and (12), respectively. We subsequently study the logic matrices of the two protein pairs (L^1 and Z^1 , L^2 and Z^2) in each complex, after which we propose several cooperative modes between the two pairs and conduct statistical analysis according to (10) and (11). In detail, these modes include (a) mode S_1 (index s_1)—only the significantly correlated segment pair in pair 1 is detected in a single essential mode with a length pair (p^1, p^2) , (b) mode S_2 (index s_2)—only the significantly correlated segment pair in pair 2 is detected, and (c) mode D (index d)—both pairs 1 and 2 are detected. To visually show the cooperative modes S and D , we select parts of the results of IGT0 for $p^1 = p^2 = 0.5$ as a display in Figure 6, in which S_1 , S_2 , and D modes are,

respectively, presented with the significantly correlated segment pairs colored.

Mode S denotes that only one protein pair, either pair 1 (S_1) or pair 2 (S_2), is significantly involved in a specific collective motion. This indicates that only one subunit, either POUHD or POUS, is significantly involved in the cooperativity with Sox-2 in an essential mode. Mode D implies that both subunits are involved in the interactions with Sox-2. Detailed statistical results are reported in Table 2. In this table, cooperative mode D occurs more frequently than modes S_1 and S_2 in the two complexes, which have the tuples of (82, 82, 98) and (85, 85, 95) for the indexes (s_1, s_2, d) , respectively. This implies that, compared with mode S_1 or S_2 , both subunits of Oct-1 frequently participate in the interactions with Sox-2 at the same time, as mode D .

Furthermore, we divide the modes S and D into subtypes, positive subtype and negative subtype, and their statistics are evaluated using (13) and listed in Table 2. In modes S_1 and S_2 , the positive subtype ($s_{1,positive}$ and $s_{2,positive}$) shows

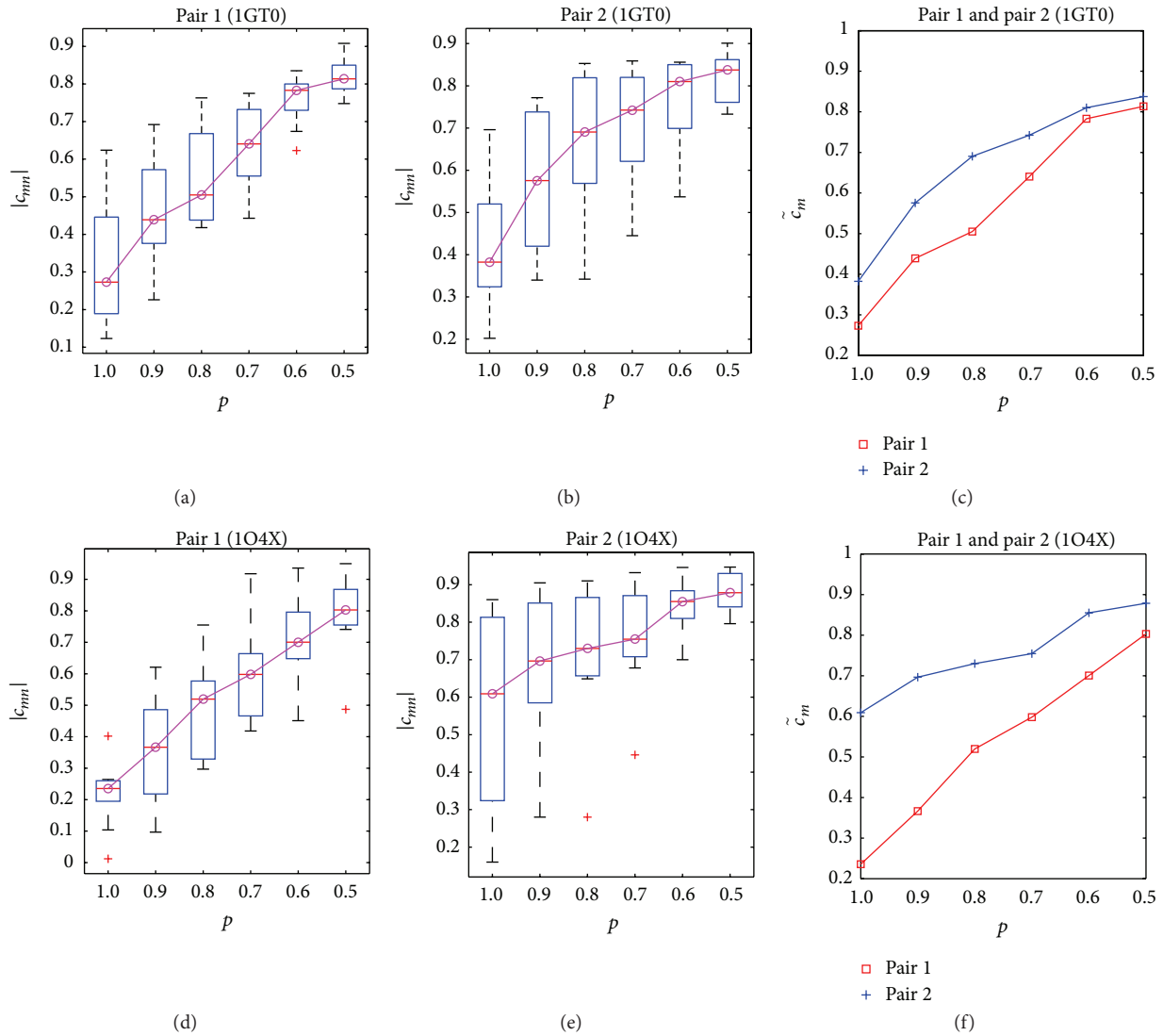


FIGURE 5: (a) and (b) show the distributions of absolute correlation values $|c_{nm}|$ for the two protein pairs in complex 1GT0, respectively. In (c), the median absolute correlation value \bar{c}_m for each p is extracted for the two pairs from (a) and (b). Similarly, (d), (e), and (f) are the plots for complex 1O4X.

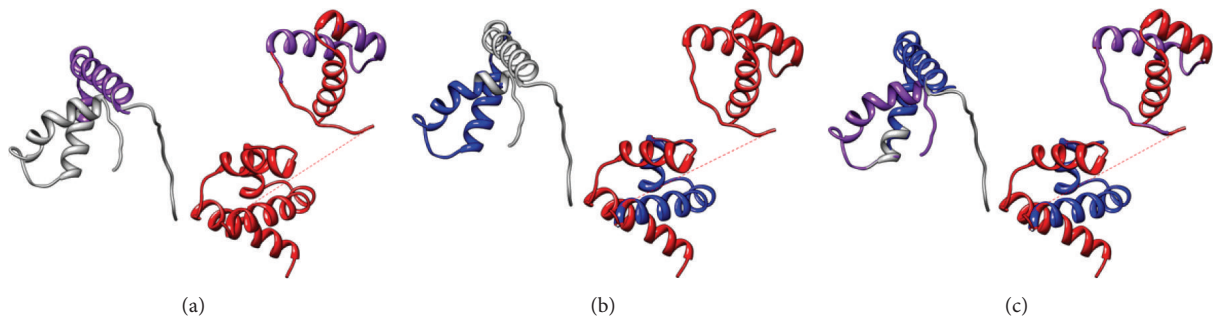


FIGURE 6: Parts of the results of cooperative modes S and D in complex 1GT0, with TFs Oct-1 and Sox-2 colored red and gray, for $p^1 = p^2 = 0.5$ in the first 10 essential modes. (a) displays mode S_1 in normal mode 13, with the significantly correlated segments in protein pair 1 colored purple; (b) displays mode S_2 in normal mode 12 and the correlated segments are colored blue in protein pair 2; (c) presents mode D in normal mode 16 with the correlated segment pairs colored purple and blue, respectively, in both protein pairs.

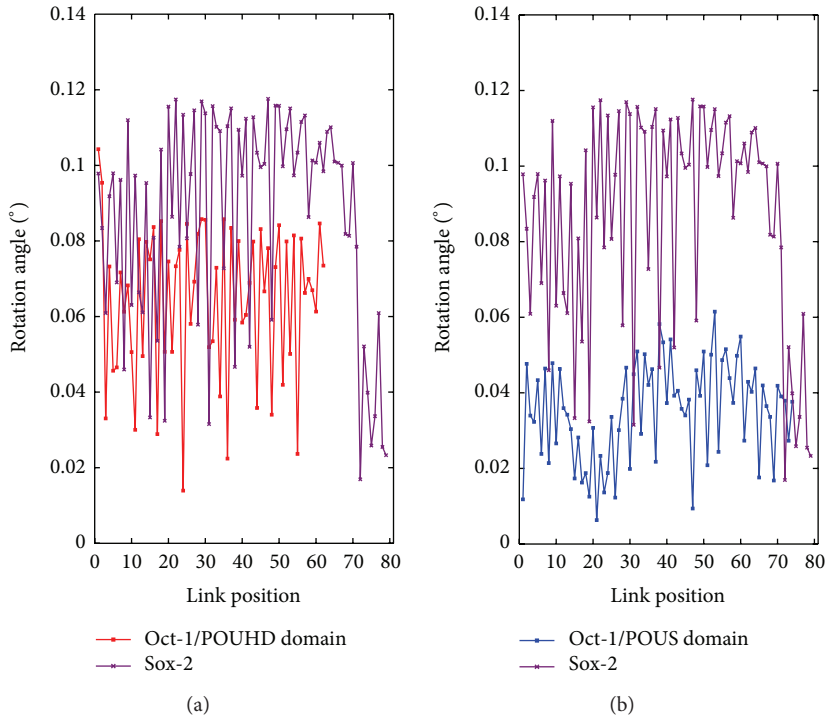


FIGURE 7: (a) displays the rotation angle curves for proteins POUHD (red) and Sox-2 (purple) in pair 1 of 1GT0 in mode 7; (b) shows the rotation angle curves for proteins POUS (blue) and Sox-2 (purple) in pair 2 of 1GT0 in mode 7.

a positive sign of c_{mn} for the significantly correlated segment pair in protein pair 1 or 2, and the negative one ($s_{1,negative}$ and $s_{2,negative}$) indicates a negative sign. In mode D , the positive subtype ($d_{positive}$) denotes a scenario where both significantly correlated segment pairs in the two protein pairs share the same sign of c_{mn} (+/+ or -/-), and the negative one ($d_{negative}$) represents two different signs (+/- or -/+). From Table 2 we notice that, for mode S_2 in both complexes 1GT0 and 1O4X, the positive subtype has a lead; for modes S_1 and D , the negative subtype is in the lead for 1GT0 while the positive one is in a dominant position for 1O4X.

We have also applied the first tertile, the first quartile and the mean value as filters and similarly conducted the statistical analysis as illustrated above. Tables 3, 4, and 5 present the results for these three scenarios, respectively. As shown in these tables, the gap between the occurrence frequencies of mode D and mode S_1 (or S_2) becomes larger, and the dominant occurrences of mode D are demonstrated. Besides, for modes S_1 and D , complexes 1GT0 and 1O4X have the opposite subtype distributions, while for mode S_2 , they present a similar distribution. Overall, these additional results are consistent with the previous one (the median filter).

3.2. Rotation Angle Functions. Subsequently, we calculate the rotation angle functions for each protein in each complex in the first 10 essential normal modes (described in Section 2.2.2). Figure 7 shows the rotation angle curves of proteins in the two protein pairs of 1GT0 in the first essential mode.

TABLE 2: Statistics on the occurrences of the cooperative modes S and D , and their subtypes, in the 10 essential modes for all the pairs (p^1, p^2), using the median filter for motion correlations.

1GT0								
s_1		s_2		d				
82	$s_{1,positive}$	$s_{1,negative}$	82	$s_{2,positive}$	$s_{2,negative}$	98	$d_{positive}$	$d_{negative}$
	36	46		49	33		35	63
1O4X								
s_1		s_2		d				
85	$s_{1,positive}$	$s_{1,negative}$	85	$s_{2,positive}$	$s_{2,negative}$	95	$d_{positive}$	$d_{negative}$
	62	23		75	10		68	27

TABLE 3: Statistics on the occurrences of the cooperative modes S and D , and their subtypes, in the 10 essential modes for all the pairs (p^1, p^2), using the first tertile as a filter for motion correlations.

1GT0								
s_1		s_2		d				
71	$s_{1,positive}$	$s_{1,negative}$	71	$s_{2,positive}$	$s_{2,negative}$	145	$d_{positive}$	$d_{negative}$
	28	43		47	24		70	75
1O4X								
s_1		s_2		d				
77	$s_{1,positive}$	$s_{1,negative}$	71	$s_{2,positive}$	$s_{2,negative}$	139	$d_{positive}$	$d_{negative}$
	57	20		62	9		98	41

Since the rotation angle functions contain a lot of noise, we apply the principal component analysis (PCA) to the 10

TABLE 4: Statistics on the occurrences of the cooperative modes S and D , and their subtypes, in the 10 essential modes for all the pairs (p^1, p^2), using the first quartile as a filter for motion correlations.

1GT0								
s_1		s_2		d				
62	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	62	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	190	d_{positive}	d_{negative}
	26	36		33	29		85	105
1O4X								
s_1		s_2		d				
70	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	64	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	182	d_{positive}	d_{negative}
	47	23		52	12		119	63

TABLE 5: Statistics on the occurrences of the cooperative modes S and D , and their subtypes, in the 10 essential modes for all the pairs (p^1, p^2), using the mean value as a filter for motion correlations.

1GT0								
s_1		s_2		d				
66	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	78	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	108	d_{positive}	d_{negative}
	33	33		43	35		44	64
1O4X								
s_1		s_2		d				
73	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	85	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	107	d_{positive}	d_{negative}
	52	21		72	13		77	30

rotation angle curves of each protein in the two complexes to obtain the first principal component (PC), leading the rotation angle curves ($n = 1 \sim 10$) of each protein to a single condensed PC curve. We similarly carry out the correlation analysis of the PC curves in each pair. Table 6 shows the statistical results for the two complexes. Intuitively, 1GT0 presents the distinct cooperative mode of S_1 , where pair 1 shows more significantly correlated segment pairs (with a positive subtype), while mode D is the dominant one in 1O4X, where many significantly correlated segment pairs occur in both pairs (with a positive subtype).

Now we apply the Fourier transform to analyze these noisy rotation angle values. Simply, the magnitudes of the transformed signals are regarded as our new data. The segmentation and correlation calculation are implemented, after which the statistical analysis is carried out. As an example, we use the first quartile as a filter for the correlations of rotation angle functions. The results are listed in Table 7, where we can see that the negative subtype of each cooperative mode is concealed after the transform. This implies that the Fourier transform may not be a suitable tool for handling these rotation angle values. More efficient strategies should be explored in the future to deal with these data.

4. Conclusions

In this paper, we performed NMA to study the collective motions of two TFs, Oct-1 and Sox-2, at their enhancer binding sites, aiming to gain an insight into the cooperative manner of these two TFs through the dynamics of their enhancer-bound complexes. Based on the special structure of Oct

TABLE 6: Correlations between PC curves of rotation angle functions for the two protein pairs in 1GT0 and 1O4X.

p	1GT0		1O4X	
	Pair 1	Pair 2	Pair 1	Pair 2
1.0	0.682	-0.278	-0.875	-0.310
0.9	0.836	-0.339	-0.884	-0.764
0.8	0.844	-0.354	-0.884	-0.821
0.7	0.854	-0.328	-0.884	-0.829
0.6	0.863	-0.498	-0.890	-0.856
0.5	0.876	0.703	-0.915	-0.859

TABLE 7: Statistics on the occurrences of the cooperative modes S and D , and their subtypes, in the 10 essential modes for all the pairs (p^1, p^2), using the first quartile as a filter for the correlations of rotation angle functions.

1GT0								
s_1		s_2		d				
108	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	108	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	144	d_{positive}	d_{negative}
	108	0		108	0		144	0
1O4X								
s_1		s_2		d				
48	$s_{1, \text{positive}}$	$s_{1, \text{negative}}$	48	$s_{2, \text{positive}}$	$s_{2, \text{negative}}$	204	d_{positive}	d_{negative}
	48	0		48	0		204	0

proteins, we treated an Oct/Sox group as two protein pairs and comparably investigated how these two pairs behave in the collective motions. A segmentation idea was introduced to explore the most correlated segments in each protein pair, according to the correlations of motion magnitude curves (or their segments). A median analysis on these correlations was conducted, which shows the leading role of subunit POU5 (pair 2). Furthermore, based on statistics of the correlated segment pairs having a correlation value above the corresponding median, we proposed several motion cooperative modes (S_1, S_2 , and D) and their subtypes (positive or negative). The first tertile, the first quartile, and the mean value provide consistent results. Moreover, the supplementary study on the rotation angle functions presents a consensus about these modes. These proposed modes provide a clue that when binding to different regulatory DNA regions or involved in different collective motions, Oct-1 has a synergistic relationship with Sox-2 either with one of the components, POU5 or POUHD, or both of them, POU5 and POUHD at the same time.

Cooperativity, in protein-DNA [25] and protein-protein [26] interactions, is an important feature in biomolecular interactions. In our work, we carried out a series of studies on the cooperative manner of Oct and Sox at their enhancer binding sites, which are important elements in the transcriptional regulation of embryonic stem cells. This work reveals how the two proteins work together physically and structurally at two specific DNA binding sites. The method developed here can be useful for the analysis of molecular interactions in other protein-protein and protein-DNA complexes.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by the City University of Hong Kong (Project 7002843).

References

- [1] L. A. Boyer, I. L. Tong, M. F. Cole et al., “Core transcriptional regulatory circuitry in human embryonic stem cells,” *Cell*, vol. 122, no. 6, pp. 947–956, 2005.
- [2] M. Pesce and H. R. Schöler, “Oct-4: gatekeeper in the beginnings of mammalian development,” *Stem Cells*, vol. 19, no. 4, pp. 271–278, 2001.
- [3] V. Botquin, H. Hess, G. Fuhrmann et al., “New POU dimer configuration mediates antagonistic control of an osteopontin pre-implantation enhancer by Oct-4 and Sox-2,” *Genes and Development*, vol. 12, no. 13, pp. 2073–2090, 1998.
- [4] M. Nishimoto, A. Fukushima, A. Okuda, and M. Muramatsu, “The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2,” *Molecular and Cellular Biology*, vol. 19, no. 8, pp. 5453–5465, 1999.
- [5] L. Dailey and C. Basilico, “Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes,” *Journal of Cellular Physiology*, vol. 186, pp. 315–328, 2001.
- [6] M. J. Borok, D. A. Tran, M. C. W. Ho, and R. A. Drewell, “Dissecting the regulatory switches of development: lessons from enhancer evolution in *Drosophila*,” *Development*, vol. 137, no. 1, pp. 5–13, 2010.
- [7] A. Warshel, “Molecular dynamics simulations of biological reactions,” *Accounts of Chemical Research*, vol. 35, pp. 385–395, 2002.
- [8] G. G. Dodson, D. P. Lane, and C. S. Verma, “Molecular simulations of protein dynamics: new windows on mechanisms in biology,” *EMBO Reports*, vol. 9, no. 2, pp. 144–150, 2008.
- [9] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava, “Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins,” *Chemical Reviews*, vol. 110, no. 3, pp. 1463–1497, 2010.
- [10] I. Bahar and A. J. Rader, “Coarse-grained normal mode analysis in structural biology,” *Current Opinion in Structural Biology*, vol. 15, no. 5, pp. 586–592, 2005.
- [11] B. Erman, “The Gaussian network model: precise prediction of residue fluctuations and application to binding problems,” *Biophysical Journal*, vol. 91, no. 10, pp. 3589–3599, 2006.
- [12] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model,” *Biophysical Journal*, vol. 80, no. 1, pp. 505–515, 2001.
- [13] S. Hayward, A. Kitao, and H. J. Berendsen, “Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme,” *Proteins*, vol. 27, pp. 425–437, 1997.
- [14] S. Hayward, A. Kitao, and N. Go, “Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis,” *Protein Science*, vol. 3, no. 6, pp. 936–943, 1994.
- [15] D. G. Teotico, M. L. Frazier, F. Ding, N. V. Dokholyan, B. R. S. Temple, and M. R. Redinbo, “Active nuclear receptors exhibit highly correlated AF-2 domain motions,” *PLoS Computational Biology*, vol. 4, no. 7, Article ID e1000111, 2008.
- [16] E. Lindahl, C. Azuara, P. Koehl, and M. Delarue, “NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis,” *Nucleic Acids Research*, vol. 34, pp. W52–W56, 2006.
- [17] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users’ Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, Pa, USA, 1998.
- [18] H. M. Berman, J. Westbrook, Z. Feng et al., “The protein data bank,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [19] E. F. Pettersen, T. D. Goddard, C. C. Huang et al., “UCSF Chimera—a visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [20] A. Reményi, K. Lins, L. J. Nissen, R. Reinbold, H. R. Schöler, and M. Wilmanns, “Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers,” *Genes and Development*, vol. 17, no. 16, pp. 2048–2059, 2003.
- [21] D. C. Williams Jr., M. Cai, and G. M. Clore, “Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex,” *Journal of Biological Chemistry*, vol. 279, no. 2, pp. 1449–1457, 2004.
- [22] B. G. Häne, K. Jäger, and H. G. Drexler, “The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms,” *Electrophoresis*, vol. 14, no. 10, pp. 967–972, 1993.
- [23] T. R. Derrick, B. T. Bates, and J. S. Dufek, “Evaluation of time-series data sets using the Pearson product-moment correlation coefficient,” *Medicine and Science in Sports and Exercise*, vol. 26, no. 7, pp. 919–928, 1994.
- [24] A. K. Gayen, “The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes,” *Biometrika*, vol. 38, no. 1-2, pp. 219–247, 1951.
- [25] D. D. Wang and H. Yan, “The relationship between periodic dinucleotides and the nucleosomal DNA deformation revealed by normal mode analysis,” *Physical Biology*, vol. 8, no. 6, Article ID 066004, 2011.
- [26] Y.-N. Zhang, X.-Y. Pan, Y. Huang, and H.-B. Shen, “Adaptive compressive learning for prediction of protein-protein interactions from primary sequence,” *Journal of Theoretical Biology*, vol. 283, no. 1, pp. 44–52, 2011.