Chapter

Time Series Analysis on the Conformational Change of c-Src Tyrosine Kinase

Hyun Jung Yoon, Sungmin Lee, Suhyun Park and Sangwook Wu

Abstract

c-Src tyrosine kinase plays an important role in signal transduction pathways, where its activity is regulated by phosphorylation of the two tyrosine residues. We performed targeted molecular dynamics simulation to obtain trajectory of conformational change from inactive to active form. To investigate the conformational change of c-Src tyrosine kinase, we applied network analysis to time series of correlation among residues. The time series of correlation between residues during the conformational change generated by targeted molecular dynamic simulation. With centrality measures such as betweenness centrality, degree centrality, and closeness centrality, we observed a few important residues that significantly contribute to the conformational change of c-Src tyrosine kinase for the different time steps.

Keywords: c-Src tyrosine kinase, targeted molecular dynamics simulation, network analysis, time-series, clustering

1. Introduction

Tyrosine kinases play a critical role in various biological processes such as migration, angiogenesis, proliferation, differentiation, survival, and immune function [1–3]. c-Src (cellular Src), encoded by Src gene, is a non-receptor tyrosine kinase first isolated as the normal cellular homolog to the potent avian sarcoma viral transforming oncogene v-Src [4]. c-Src tyrosine kinase consists of the N-terminal unique region, the Src homology 3 (SH3), SH2, linker, kinase domain, and the regulatory C-terminal tail. Under normal circumstances, the SH3 domain of the c-Src tyrosine kinase binds to the proline-rich region in the linker domain. And the SH2 domain binds to Tyr527, which leads to a closed conformation [5, 6]. However, under certain conditions, the closed, inactive c-Src tyrosine kinase undergoes a transition to an open and active conformation. One of the important features of c-Src tyrosine kinase catalytic activation is to control its phosphorylation status. One of the major phosphorylation sites is Tyr527, which is located in the C-terminal tail. Dephosphorylation of pTyr527 (phosphorylated Tyr527) releases the closed conformation, which leads to the active state. Another major phosphorylation site is Tyr416, which is located in the activation loop of the tyrosine kinase domain [7, 8]. c-Src tyrosine kinase activity depends on the phosphorylation status of the two residues: Tyr527 and the Tyr416.

Using targeted molecular dynamics (TMD) simulation, we study the conformational change of c-Src tyrosine kinase by applying an external bias [9]. We generate sequential and continuous transition between two known conformations: the inactive and the active conformation for the case of Tyr527-pTyr416 (instead of pTyr527-Tyr416).

Network analysis has been successfully applied to biological problems [10, 11]. For instance, Goh et al. constructed a network of Mendelian gene-disease associations to identify unknown important genes causing disease [12]. In a biological network, a node corresponds to a gene and an edge is an interaction or correlation between genes. Centrality measures are very effective approaches to determine the ranking or importance of genes in the gene-disease network. Those centrality measures uncover important target genes relevant to disease. In this study, using network analysis, clustering method, and time-series analysis, we attempt to reveal the key residues, which influence the conformational transition of the c-Src tyrosine kinase between the inactive and active conformation.

2. Method

2.1 Targeted molecular dynamics (TMD) simulation

To perform the Targeted Molecular Dynamics simulation, an external potential, U_{TMD} [13], to be defined.

$$U_{TMD} = \frac{k}{2N}[RMSD(t) - RMSD^*(t)]^2 \quad (1)$$

In the above equation, RMSD(t) is defined as root-mean-square deviation (RMSD) of the simulated structure from the target structure at time t. RMSD*(t) is the RMSD value at time t assuming a linear decrease from the initial to the target structure. The inactive (PDB ID: 2SRC) and active form (PDB id: 1Y57) of c-Src tyrosine kinase are defined as initial and target conformation, respectively [14, 15]. Both the conformations were modified (to have the same number of atoms) for the TMD simulation. The spring constant '*k*' was set as 2500 kcal/mol ·Å^2 for 3,619 atoms (hydrogen atoms excluded). NAMD 2.9 [16] with the CHARMM 27 force field [17] was used to perform simulation and the protein parameters were incorporated with the CMAP corrections [18]. In TMD, the conformational change from inactive to the active state is guided by the external force within a reasonable time scale. In our calculation, we have used 10 ns time scale. To undergo the conformational change within 10 ns time scale and to avoid the bias effect, we employed the smallest spring constant 'k' (2500 kcal/mol ·Å^2). TIP3P water model [19] was used in the simulation. The particle mesh Ewald (PME) method was set as 12 Å direct space cut-off [20]. The damping coefficient was set as 5 ps^{-1} for the Langevin dynamics simulation. To maintain the constant pressure (1 atm), Nosé–Hoover method was used (1 atm) [21]. The NPT ensemble was carried out in 310K for the TMD simulation.

2.2 Network analysis

To analyze the correlation between residues in the trajectories of the TMD simulations, we have used the dynamical cross-correlation (DCCM) method [22–24].

$$C_{ij} = \frac{< (r_i(t) - < r_i(t) >)(r_j(t) - < r_j(t) >) >}{\sqrt{(< r_i^2(t) > - < r_i(t) >^2)(< r_j^2(t) > - < r_j(t) >^2)}} \quad (2)$$

Where, $r_i(t)$ and $r_j(t)$ = atomic positions of the i^{th} and j^{th} C_α atoms at time t. The quantity "$r_i(t) - <r_i(t)>$" corresponds to the fluctuation of the "i^{th}" atom and "$r_j(t) - <r_j(t)>$" corresponds to the fluctuation of the "j^{th}" atom. For all the Cα atoms, a 450 × 450 (residues 84-533) correlation map was obtained during the 10 ns TMD simulation. In Eq. (2), the quantity C_{ij} in the DCCM is an adjacency matrix. The weight w_{ij}, of the edge between the nodes, 'i and j', defined as [25, 26]

$$w_{ij} = -log|C_{ij}| \quad (3)$$

DCCM can be used to measure the weight, which is the probability of information transfer across the edge. In the constructed network, every node is a Cα atom and each edge is an information transfer probability, in the cross-correlation. To identify and quantify the nodes that occupy critical positions in a network, a few centrality measures have been proposed, including the degree, betweenness, and closeness centralities [27–29].

The degree centrality measures the number of edges incident on a node in a network, thus expressing the "popularity" of the node.

$$C_D(v_i) = d_i = \sum_j A_{ij} \quad (4)$$

where A_{ij} is the adjacency matrix: if $w_{ij} = 0$ then $A_{ij} = 0$, otherwise $A_{ij} = 1$. The closeness centrality is defined as the average length of the shortest paths between a node and all the other nodes in a network. It can be used to measure information spread from a given node to the other nodes. The closeness centrality is defined as

$$C_C(v_i) = \frac{n-1}{\sum_{j \neq i} g(v_i, v_j)} \quad (5)$$

where $g(v_i, v_j)$ is the shortest path with a weight between two nodes 'i' and 'j'. The betweenness centrality is to measure the number of information pathways that flow through a node in a network. The betweenness of node 'i' is the fraction of the shortest paths between pairs of nodes that pass through node 'i'. The betweenness centrality is defined as

$$b_i = \frac{\sum_{s<t} g_i^{st}/n_{st}}{\frac{1}{2}n(n-1)} \quad (6)$$

where g_i^{st} is the number of shortest paths from 's' to 't' with a weight that passes through node 'i'. n_{st} is the total number of shortest paths from to 's' to 't'. We obtained DCCM and three centralities using Bio3d [30–32].

2.3 Clustering

Clustering is an effective approach to discover interesting patterns of time series data. It is widely used in diverse fields of research from biology to economy:

functional clustering of time-series gene expression data [33], identification of functionally related genes [34–36], detecting brain activity [37, 38], identifying pathological cases from mass spectrometry (MS) clinical samples [39], discovering energy consumption pattern [40, 41], and pattern finding in stock time series [42, 43].

Using the targeted molecular simulation (TMD), we generated a dynamical correlation matrix between two residues, C_{ij}, for each time step during the conformational change from the inactive to active state. Then, the time-series of C_{ij} are grouped into a set of time series in such a way that their temporal profiles in the same group (cluster) are more similar to each other than those in other groups (clusters). We applied a hierarchical clustering algorithm and repeated clustering runs with a different number of clusters from 4 to 10 until getting clear clustering. The function 'tsclust [44]' of 'DTWclust [45]' in the R statistical package is used for clustering of time series data.

3. Results and discussion

3.1 Conformational change of c-Src tyrosine kinase

The electrostatic interaction between the Tyr527 and the positively charged Arg175/Lys203 becomes weaker in the inactive conformation at the early stage of the transition from the inactive to the active conformation by TMD simulation. The C-terminal tail, including Tyr527, is completely detached from the SH2 domain (the period of 0-3 ns) in the TMD simulation. The detachment of Tyr527 from the SH2 domain triggers the conformational change. The detached Tyr527 moves toward the kinase domain (4-6 ns period: **Figure 1b** and **c**). At this stage, the most prominent

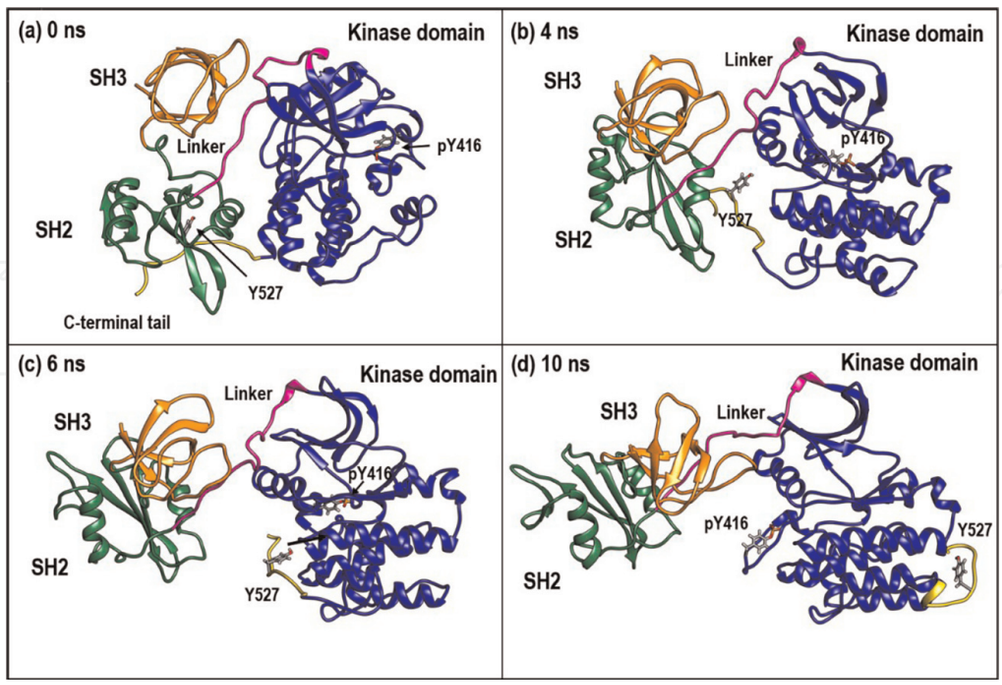

Figure 1.
Four snapshots of TMD simulation of c-Src tyrosine kinase between 0 and 10 ns. (a) 0 ns: inactive conformation (b) 4 ns: Detachment of Tyr527 in the C-terminal tail from the SH2 domain (c) 6 ns: Large-scale conformational change of c-Src tyrosine kinase (d) 10 ns: active conformation. Color code: SH3 (resid 84-142, Orange), SH2 (resid 143-245, Green), linker (resid 246-266, Magenta), and tyrosine kinase domains (resid 267-520, Blue), the regulatory C-terminal tail (resid 521-533, Yellow). The two residues, pTyr416 and Tyr527, are represented as licorice [9].

conformational change occurs in the kinase domain. The secondary structures, including the αC-helix in the kinase domain, rotate significantly. During this time, Tyr416 remains buried beneath the activation loop. At the final stage of transition, Tyr527 has reached the far side of the kinase domain relative to the SH2 domain. The activation loop has also moved from its original position. pTyr416 is now exposed to the surface.

The activation processes from the inactive state (0 ns: **Figure 1a**) to the active state (10 ns: **Figure 1d**) in the TMD simulation are shown in **Figure 1**. The conformational transition of c-Src tyrosine kinase from the inactive to the active state generated by TMD simulation is shown in the supplementary video material [9].

3.2 Centrality measures

The degree, closeness, and betweenness centralities are measured for the conformational change that occurred during the 10 ns TMD simulation (**Figure 2**). The residues with high values of the degree, closeness, and betweenness centralities are not located in the activation loop, linker, αC-helix. The residues with high centrality measures are mainly confined to the helix region adjacent to the αC-helix. The residues with the top 5 values of each centrality measure are listed in **Table 1**.

3.3 Clustering

We investigate correlation patterns between specific residues (Trp260, Tyr416, Tyr527, Lys321) and all the other residues using the clustering method. The clustering of residues provides us with information on not only the pattern of the time series but also the correlation values, C_{ij}, between residues at each time step.

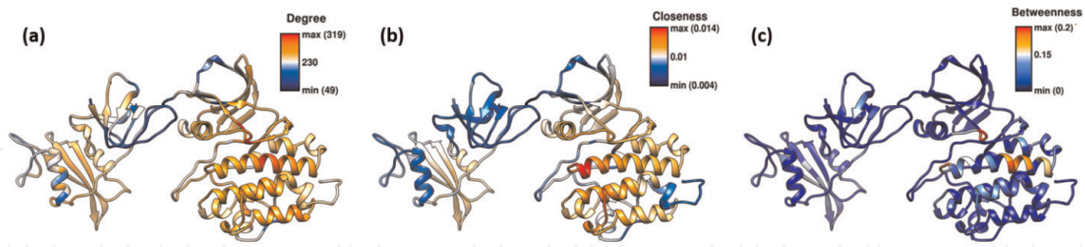

Figure 2.
Centrality measures during the 10 ns TMD simulation. (a) Degree centrality (b) Closeness centrality (c) Betweenness centrality. We mapped centrality measures onto the active conformation of c-Src tyrosine kinase [9].

	Betweenness	Closeness	Degree
1	Glu320	Val377	Val323
2	Lys321	Arg379	Val402
3	Ala368	Glu378	Lys321
4	Leu322	Met380	Ile370
5	Met380	Thr508	Thr301

Table 1.
Top 5 residues of each centrality measure [9].

3.3.1 Trp260

Trp260 has an important role in the conformational change of c-Src tyrosine kinase [46]. **Figure 3** shows a time series of correlation, C_{ij} (t), for Trp260. It shows distinctive patterns among clustering groups. Residues in both group 6 (**Figure 3a**) and group 11 (**Figure 3b**) are positively correlated to Trp260 for the entire time window (0-10 ns). The residues in clustering group 9 (**Figure 3c**), however, are negatively correlated to Trp260. **Figure 1a** shows no correlation in early time (0-4 ns). The correlation increases positively (4-7 ns), and reaches the highest correlation value (~7 ns). Eventually, the correlation decreases (8-10 ns). The location of residues in clustering group 6 (orchid), clustering group 9 (cyan), and clustering group 11 (pink) are shown (**Figure 3d**). The positively correlated (blue) and negatively correlated residues (red) are shown (**Figure 3e**). The positively correlated residues are mainly located near residue Trp260 in the inactive form of c-Src tyrosine kinase. The negatively correlated residues are spread around a relatively far from Trp260 in the inactive form. The time series analysis based on the clustering method and network analysis indicates that Trp260 is rapidly changing its position after the second half of the conformational transition.

3.3.2 pTyr416

Figure 4 shows the time series of correlation, C_{ij} (t), for pTyr416. The residues in the clustering group 12 are positively correlated to Tyr416 for the entire time window

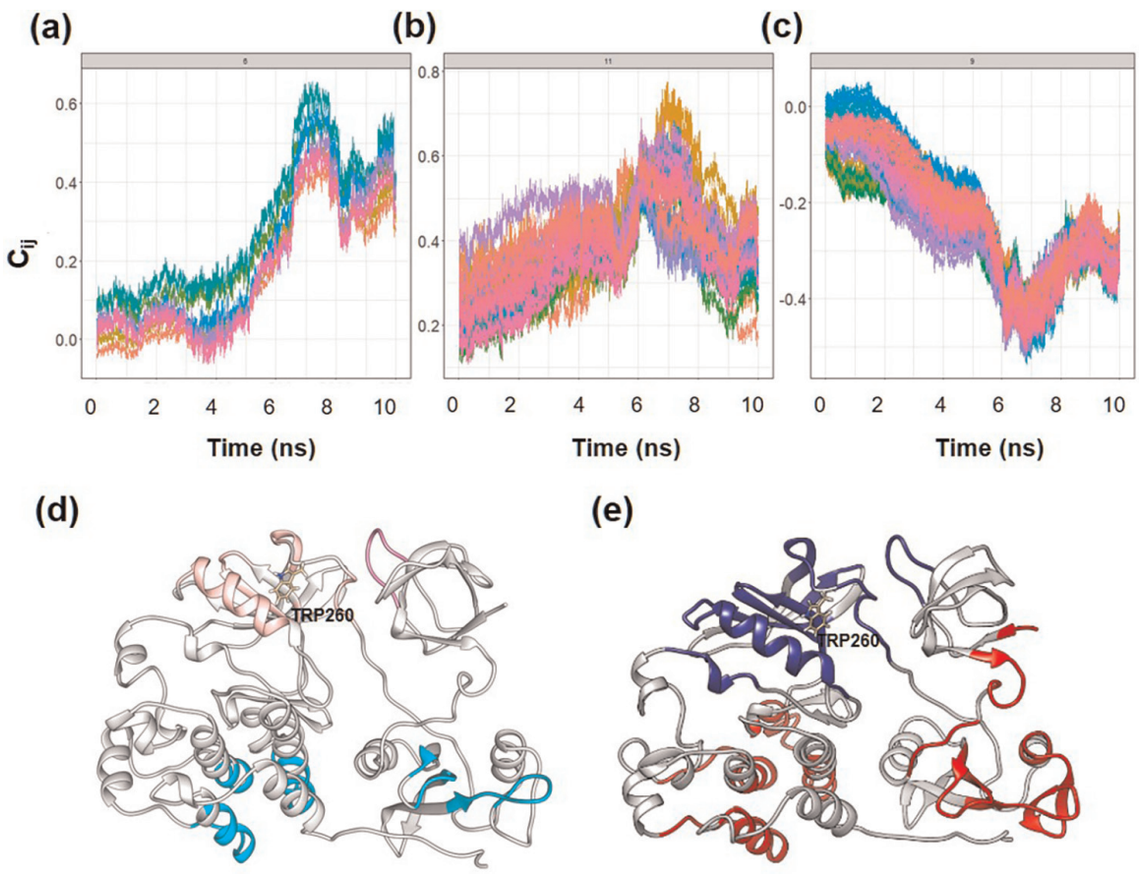

Figure 3.
Some distinctive clusters on the time series of correlation between Trp260 and other residues. (a) Clustering group 6 (b) Clustering group 11 (c) Clustering group 9 (d) Each cluster is mapped on the inactive conformation. Color code: orchid, pink and cyan for clustering 6, 11, and 9 respectively. (e) The positive correlation clusters (blue) and negative correlation clusters (red) on the inactive conformation.

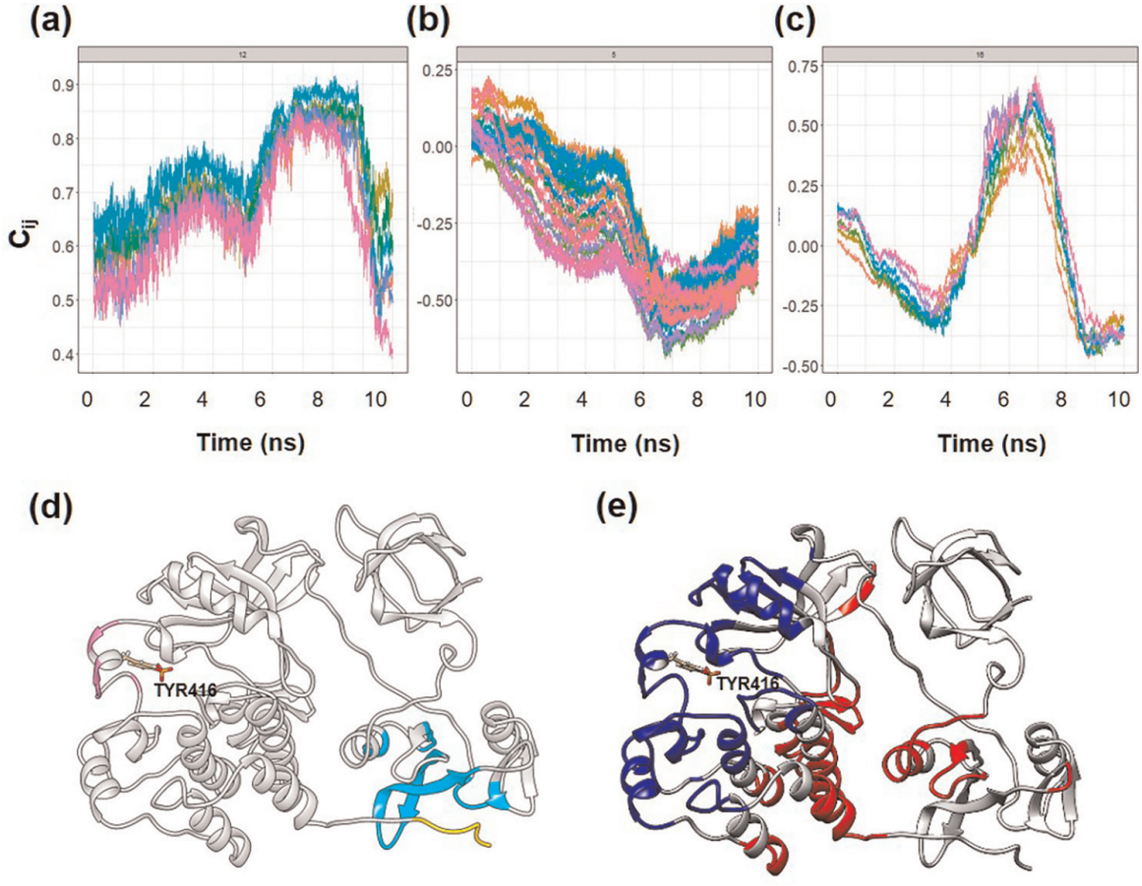

Figure 4.
Some distinctive clusters on the time series of correlation between pTyr416 and other residues. (a) Clustering group 12 (b) Clustering group 5 (c) Clustering group 16 (d) Each cluster is mapped on the inactive conformation. Color code: orchid, cyan, and yellow for clustering 12, 5, and 16 respectively. (e) The positive correlation clusters (blue) and negative correlation clusters (red) on the inactive conformation.

(0-10 ns) (**Figure 4a**). The residues in the clustering group 5, however, are negatively correlated to pTyr416 (**Figure 4b**). Interestingly, the residues in the clustering group 16 show correlation change in the two different time regimes (**Figure 4c**). As shown in **Figure 4a**, C_{ij} (t) has a large value (> 0.5) even in early time and increases to 0.9. During the 7-8 ns period, C_{ij} (t) reaches a plateau with the highest correlation value. After 8 ns, C_{ij} (t) returns to 0.5. The residues in the clustering group 12, shown in orchid in **Figure 4d**, are mainly located close to pTyr416. The positive correlation between the clustering group 12 and pTyr416 is due to "physical distance".

Most of the residues in the clustering group 5 are negatively correlated to pTyr416 during the conformational transition (**Figure 4b**). C_{ij} (t) decreases to a minimum value of -0.7 around 7 ns and returns to -0.3. The residues in group 5, shown as cyan in **Figure 4d**, are located far away from pTyr416. The negative value of C_{ij} (t) indicates that the movement of pTyr416 and the residues in the clustering group 5 are in a reverse direction.

The residues in the clustering group 16 show the correlation pattern change to pTyr416 during the conformational transition. During 0-4 ns period, residues in group 16 (shown as yellow in **Figure 4d**) shows negatively correlated to pTyr416 with a minimum value of -0.3 in C_{ij} (t). And after 4 ns, the correlation pattern shows positively correlated to pTyr416. C_{ij} (t) reaches 0.7 around 7 ns in the conformational transition process.

The positively correlated (blue) and negatively correlated residues (red) to pTyr416 are mapped into the inactive conformation (**Figure 4e**).

*3.3.3 T*yr527

Figure 5 shows the time series of correlation, C_{ij} (t), for Tyr527. The residues in the clustering group 14 are positively correlated to Tyr527 for the entire time window (0-10 ns) (**Figure 5a**). The residues in the clustering group 9, however, are negatively correlated to Tyr527 during the most time of conformational transition (**Figure 4b**). As with the case of pTyr416, the residues in the clustering group 2 show correlation pattern change in the two different time regimes (**Figure 5c**). As shown in **Figure 5a**, C_{ij} (t) increases to 0.6 during 0-4 ns period. After an abrupt decrease, C_{ij} (t) increases again to reach a plateau with the value of 0.6 during 6-10 ns. The residues in the clustering group 14 (shown as orchid in **Figure 5d**) are located close to Tyr527. Similarly, the positive correlation between the clustering group 14 and Tyr527 is due to a strong coupled through "physical distance".

Most of the residues in the clustering group 9 are negatively correlated to Tyr527 during the conformational transition (**Figure 5b**). The residues in the clustering group 9 (shown as cyan in **Figure 5d**) are located away from Tyr527 in the inactive conformation.

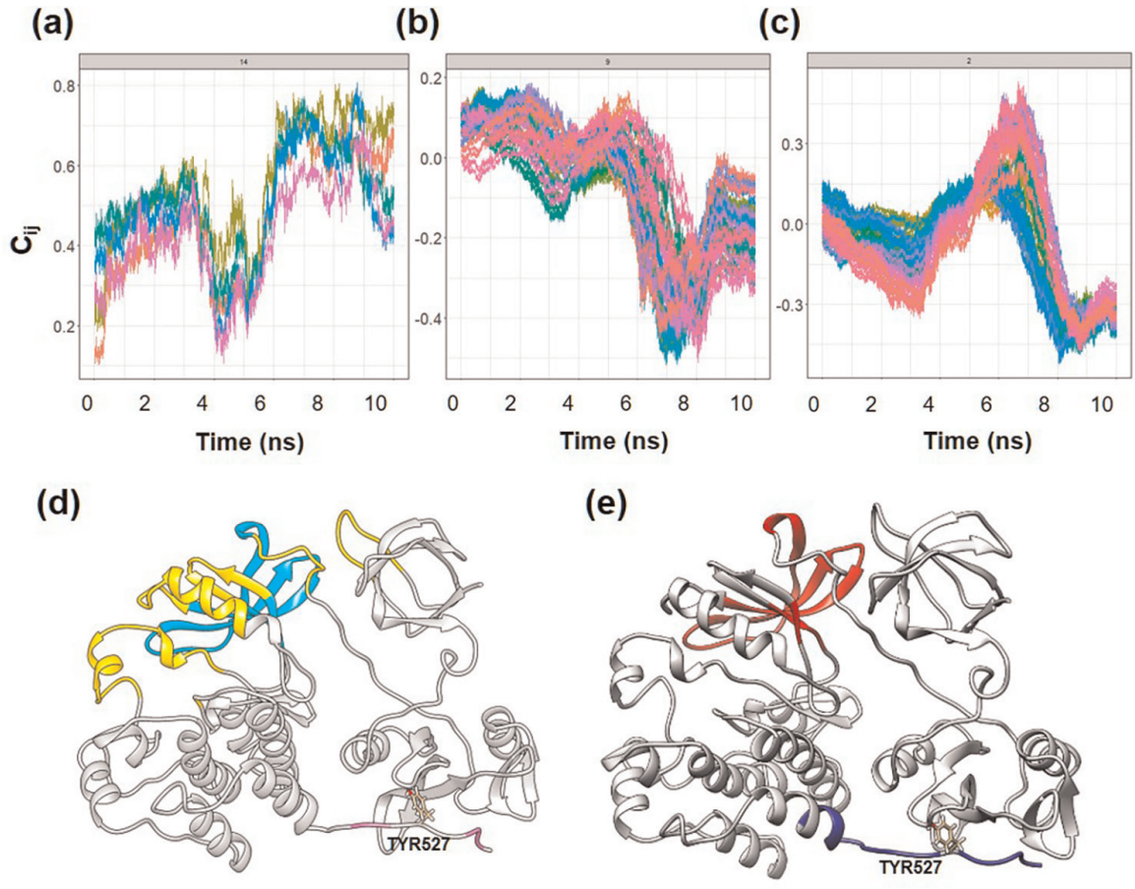

Figure 5.
Some distinctive clusters on the time series of correlation between pTyr416 and other residues. (a) Clustering group 14 (b) Clustering group 9 (c) Clustering group 2 (d) Each cluster is mapped on the inactive conformation. Color code: orchid, cyan, and yellow for clustering 14, 9, and 2 respectively. (e) The positive correlation clusters (blue) and negative correlation clusters (red) on the inactive conformation.

The residues in the clustering group 2 show the correlation pattern change to Tyr527 during the conformational transition, which is observed in the case of pTyr416. During 0-4 ns period, the residues in the clustering group 2 (shown as yellow in **Figure 5d**) shows negatively correlated to Tyr527 with a minimum value of -0.3 in C_{ij} (t). And after 4 ns, the correlation pattern shows positively correlated to Tyr527. In the last, C_{ij} (t) reaches 0.45 around 7 ns in the conformational transition process.

The positively correlated (blue) and negatively correlated residues (red) to pTyr416 are mapped into the inactive conformation (**Figure 5e**).

3.3.4 Lys321

According to network analysis, Lys321 has high centralities both in betweenness and degree centralities (**Table 1**). Within the framework of network theory, it implies that Lys321 would play an essential role in the conformational change of c-Src tyrosine kinase. The residues in the clustering group 13 are positively correlated to Lys321 for the entire time window (0-10 ns) (**Figure 6a**). The residues in the clustering group 1 are negatively correlated to Lys321 (**Figure 6b**). As with the cases of pTyr416 and Tyr527, the residues in the clustering group 16 show correlation pattern change in the two different time regimes (**Figure 6c**).

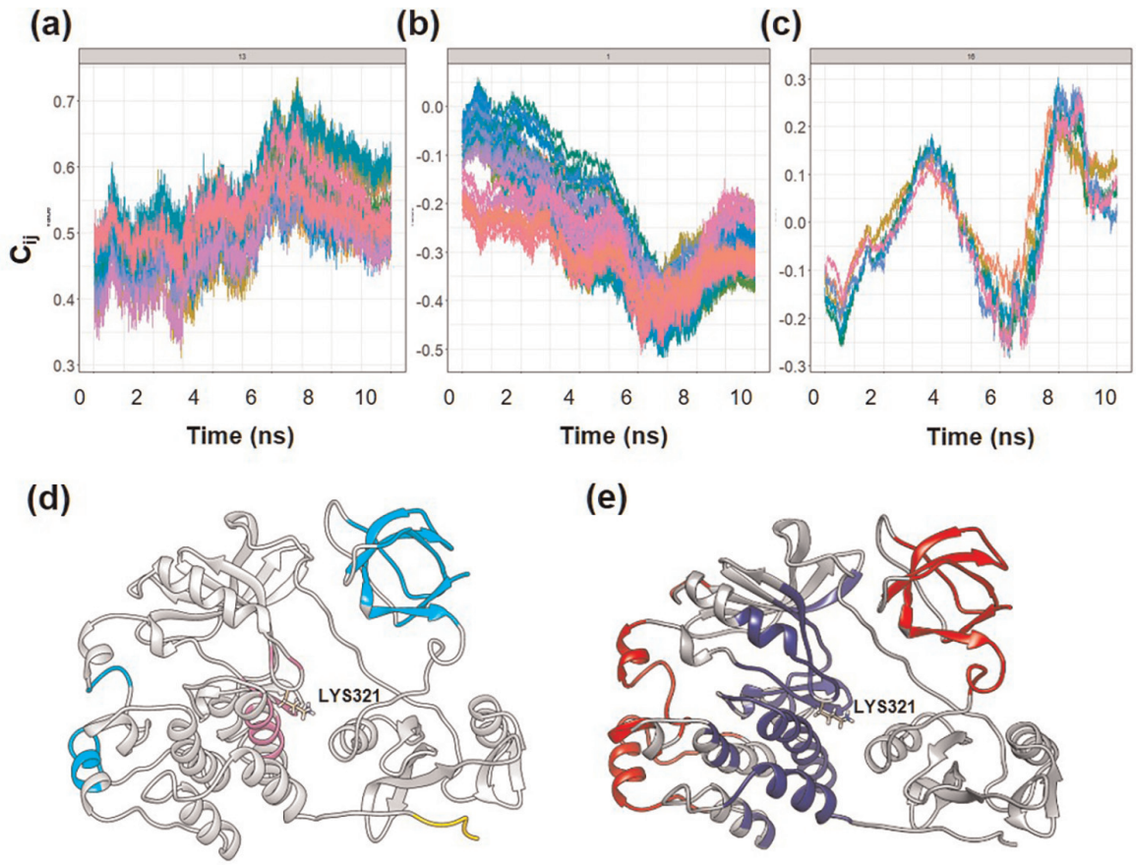

Figure 6.
Some distinctive clusters on the time series of correlation between Lys321and other residues. (a) Clustering group 13 (b) Clustering group 1 (c) Clustering group 16 (d) Each cluster is mapped on the inactive conformation. Color code: orchid, cyan, and yellow for clustering 13, 1, and 16 respectively. (e) The positive correlation clusters (blue) and negative correlation clusters (red) on the inactive conformation.

As shown in **Figure 6a**, C_{ij} (t) values fluctuate around 0.5 during the 0-5 ns period. After that, it increases to 0.7 around 7 ns. Finally, it decreases to around 0.55. The residues in the clustering group 13 (shown as an orchid in **Figure 6d**) are located close to Lys 321. Similarly, the positive correlation between the clustering group 14 and Lys321 is due to a strong coupled through "physical distance".

Most of the residues in the clustering group 1 are negatively correlated to Lys321 during the conformational transition (**Figure 6b**). The residues in the clustering group 1 (shown as cyan in **Figure 5d**) are located away from Lys321 in the inactive conformation. **Figure 6b** shows that C_{ij} decreases to -0.5 until 6 ns and increases and reaches -0.3 at the end of the conformational transition. The residues in the clustering group 1, shown in purple in **Figure 6d**, are located quite away from Lys321.

The residues in the clustering group 16 show the correlation pattern change to Lys321 during the conformational transition, which is observed in the cases of pTyr416 and Tyr527. During 0-2 ns period, the residues in the clustering group 16

TRP260	Positive correlation	Cluster 6	93, 94, 95, 96, 97, 98, 99
	Positive correlation	Cluster 11	253, 254, 255, 256, 257, 263, 264, 265, 266, 267, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 325, 326, 330, 331, 332, 333, 334, 335, 336, 337
	Negative correlation	Cluster 9	176, 177, 180, 181, 182, 183, 184, 185, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 452, 453, 454, 455, 456, 457, 458, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495
TYR416	Positive correlation	Cluster 12	413, 414, 418, 419, 420, 424
	Negative correlation	Cluster 5	162, 163, 169, 170, 171, 172, 173, 174, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 214, 215, 236, 237, 238, 239
	Switch	Cluster 16	527, 528, 529, 530, 531, 532, 533
TYR527	Positive correlation	Cluster 14	524, 525, 530, 531, 532
	Negative correlation	Cluster 9	263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 325, 326, 337, 338, 339, 340
	Switch	Cluster 2	93, 94, 95, 96, 97, 98, 99, 100, 101, 255, 256, 257, 258, 259, 260, 261, 262, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 386, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424
LYS321	Positive correlation	Cluster 13	319, 323, 368, 369, 370, 371, 372, 374, 375, 376, 399, 400, 402
	Negative correlation	Cluster 1	84, 85, 86, 87, 88, 89, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 136, 137, 138, 139, 140, 420, 421, 422, 423, 424, 466, 467, 468, 469, 470, 471, 472, 473, 474
	Switch	Cluster 16	528, 529, 530, 531, 532, 533

Table 2.
*The residues in the clustering group (**Figures 3d, 4d, 5d,** and **6d**).*

TRP260	Negative correlation	Cluster 1	84, 85, 86, 138, 139, 140, 141, 142, 143, 144, 145, 146, 149, 150, 151, 152, 153, 154, 174, 175, 178, 179, 186, 202, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 346, 347, 348, 349, 350, 351, 352, 353, 354, 389, 448, 449, 450, 451, 459, 460, 461, 462, 463, 464, 465, 479, 480, 481, 482, 496, 497, 498, 499
		Cluster 9	176, 177, 180, 181, 182, 183, 184, 185, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 452, 453, 454, 455, 456, 457, 458, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495
	Positive correlation	Cluster 6	93, 94, 95, 96, 97, 98, 99
		Cluster 10	:251, 252, 268, 284, 285, 286, 287, 288, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 316, 317, 318, 324, 338, 339, 406, 407, 408, 409, 410, 411, 412
		Cluster 11	253, 254, 255, 256, 257, 263, 264, 265, 266, 267, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 325, 326, 330, 331, 332, 333, 334, 335, 336, 337
		Cluster 12	258, 262, 327, 328, 329
		Cluster 13	259, 261
TYR416	Negative correlation	Cluster 4	149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 175, 176, 177, 178, 179, 180, 181, 182, 183, 208, 209, 210, 289, 290, 291, 320, 321, 339, 340, 341, 342, 343, 344, 346, 349, 350, 351, 352, 353, 367, 368, 369, 370, 372, 393, 394, 401, 452, 455, 456, 457, 458, 485, 486, 487, 488, 489, 490, 491, 492, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523
		Cluster 10	354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 395, 396, 397, 398, 399, 400
	Positive correlation	Cluster 6	261, 277, 278, 279, 296, 297, 298, 314, 328, 329, 330, 331, 332, 333, 334, 335, 336, 388, 404, 463, 466, 467
		Cluster 7	299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 385, 386, 406, 407, 427, 428, 429, 430, 431, 432, 438, 439, 440, 443, 468, 469, 470, 471, 472, 473
		Cluster 8	311, 312, 313, 377, 378, 380, 381, 382, 383, 384, 387, 405, 441, 442, 444, 445, 446, 447, 474, 475, 476, 477, 478, 479, 499, 500, 501, 502, 503, 504, 505, 506
		Cluster 11	408, 409, 410, 411, 412, 421, 422, 423, 425, 426, 433, 434, 435, 436, 437
		Cluster 12	413, 414, 418, 419, 420, 424
		Cluster 13	415, 417
TYR527	Negative correlation	Cluster 9	263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 325, 326, 337, 338, 339, 340
	Positive correlation	Cluster 13	516, 517, 518, 519, 520, 521, 522, 523, 533
		Cluster 14	524, 525, 530, 531, 532
		Cluster 15	526, 528, 529
LYS321	Negative correlation	Cluster 1	84, 85, 86, 87, 88, 89, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 136, 137, 138, 139, 140, 420, 421, 422, 423, 424, 466, 467, 468, 469, 470, 471, 472, 473, 474

	Cluster 3	141, 142, 143, 144, 145, 146, 147, 298, 299, 300, 301, 413, 414, 415, 416, 417, 418, 419, 425, 426, 427, 428, 429, 431, 434, 435, 436, 460, 461, 462, 463, 464, 465, 475, 476, 477, 478, 479, 480, 481, 482
Positive correlation	Cluster 9	290, 291, 292, 293, 311, 312, 313, 327, 337, 345, 349, 358, 359, 360, 383, 384, 387, 389, 391, 404, 405, 441, 444, 445, 448, 490, 491, 494, 507
	Cluster 10	314, 315, 325, 326, 338, 340, 343, 344, 346, 381, 382
	Cluster 11	316, 339, 341, 342, 361, 362, 363, 380, 396, 397, 508, 515, 516, 517, 518, 519, 520, 521, 522
	Cluster 12	317, 318, 324, 364, 365, 366, 367, 377, 378, 379, 392, 393, 394, 395, 398, 403, 509, 510, 511, 512, 513, 514
	Cluster 13	319, 323, 368, 369, 370, 371, 372, 374, 375, 376, 399, 400, 402
	Cluster 14	320, 322, 373, 401

Table 3.
*Positively/negatively correlated residues in the clustering group (**Figures 3e, 4e, 5e, 6e**).*

(shown as yellow in **Figure 6d**) shows negatively correlated to Lys321 with the minimum value of -0.25 in C_{ij} (t). And for 2-4 ns, the correlation pattern shows positively correlated to Lys321. After 4 ns, it sharply decreases to the minimum value of -0.25. Subsequently, it reaches the maximum positive value of 0.3 in C_{ij} (t). Eventually, it decreases to zero. The C_{ij} (t) for Lys321 shows more correlation pattern changes compared to the cases of pTyr416 and Tyr527.

The positively correlated (blue) and negatively correlated residues (red) to Lys321 are mapped into the inactive conformation (**Figure 6e**).

The residues in the clustering group (**Figures 3d**, **4d**, **5d**, and **6d**) are listed in **Table 2** positively and negatively correlated residues in the clustering group (**Figures 3e**, **4e**, **5e**, and **6e**) are listed in **Table 3**.

4. Conclusions

In this study, we investigated the conformational transition of c-Src tyrosine kinase from the inactive to the active state using TMD simulation and network theory. Tyr416 and Tyr527 are known to be significant residues in the conformational transition by controlling the phosphorylation status. They play as a *switch* for the conversion of the inactive/active conformation. The time-dependent correlation matrices, C_{ij} (t), between all the residues and pTyr416 and Tyr527 show very similar patterns (positively correlated, negatively correlated, negatively/positively correlated) during the conformational transition process. The time-series analysis supports that pTrp416 and Tyr527 act as a *switch* in a very concerted manner for completion of conformational transition of c-Src tyrosine kinase. Based on the analysis of the three centrality measures (betweenness, closeness, and degree), we observed that Lys321 plays an essential role in the conformational transition of c-Src tyrosine kinase. The time-series analysis shows that C_{ij} (t), between all the residues and Lys321 show positively correlated, negatively correlated, and pattern change from negatively to positively correlated one during the conformational transition process. Combining the network analysis with time-series analysis, Lys321 may be another candidate for a *switch* for the conformational transition of c-Src tyrosine kinase.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2018R1D1A1B07048762).

Conflict of interest

The authors declare no conflict of interest.

Author details

Hyun Jung Yoon[1,2†], Sungmin Lee[1†], Suhyun Park[2] and Sangwook Wu[1,2*]

1 R&D Center of PharmCADD, Busan, Republic of Korea

2 Department of Physics, Pukyong National University, Busan, Republic of Korea

*Address all correspondence to: sangwoow@pknu.ac.kr; s.wu@pharmcadd.com

† These authors contributed equally.

References

[1] Frame MC. Newest findings on the oldest oncogene; how activated src does it. Journal of cell science. 2004;117(7): 989–998.

[2] Summy JM, Gallick GE. Treatment for advanced tumors: SRC reclaims center stage. Clinical Cancer Research. 2006;12(5):1398–1401.

[3] Yeatman TJ. A renaissance for SRC. Nature Reviews Cancer. 2004;4(6): 470–480.

[4] Brown MT, Cooper JA. Regulation, substrates and functions of src. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 1996;1287(2):121–149.

[5] Roskoski R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. Pharmacological research. 2015;94:9–25.

[6] BOGGON TJ, ECK MJ. Structure and regulation of Src family kinases. Oncogene. 2004;23(48):7918–7927.

[7] Sun G, Sharma AK, Budde RJ. Autophosphorylation of Src and Yes blocks their inactivation by Csk phosphorylation. Oncogene. 1998;17(12):1587–1595.

[8] Meng Y, Roux B. Locking the active conformation of c-Src kinase through the phosphorylation of the activation loop. Journal of molecular biology. 2014;426 (2):423–435.

[9] Yoon HJ, Lee S, Park SJ, Wu S. Network approach of the conformational change of c-Src, a tyrosine kinase, by molecular dynamics simulation. Scientific Reports. 2018;8(1):5673.

[10] Ma'ayan A. Introduction to network analysis in systems biology. Science signaling. 2011;4(190):tr5–tr5.

[11] Almaas E. Biological impacts and context of network theory. Journal of Experimental Biology. 2007;210(9): 1548–1558.

[12] Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proceedings of the National Academy of Sciences. 2007; 104(21):8685–8690.

[13] Schlitter J, Engels M, Krüger P, Jacoby E, Wollmer A. Targeted molecular dynamics simulation of conformational change-application to the T↔R transition in insulin. Molecular Simulation. 1993;10(2-6):291–308.

[14] Xu W, Doshi A, Lei M, Eck MJ, Harrison SC. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Molecular cell.1999;3(5): 629–638.

[15] Cowan-Jacob SW, Fendrich G, Manley PW, Jahnke W, Fabbro D, Liebetanz J, et al. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. Structure. 2005;13(6):861–871.

[16] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry. 2005;26(16):1781–1802.

[17] MacKerell AD, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers. 2000;56(4): 257–265.

[18] MacKerell AD, Feig M, Brooks CL. Extending the treatment of back-bone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein

conformational distributions in molecular dynamics simulations. Journal of computational chemistry.2004;25(11): 1400–1415.

[19] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics. 1983;79 (2):926–935.

[20] Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993;98 (12):10089–10092.

[21] Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. The Journal of chemical physics. 1995;103(11): 4613–4621.

[22] Hünenberger P, Mark A, Van Gunsteren W. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. Journal of molecular biology. 1995;252(4):492–503.

[23] Karplus M, Ichiye T. Comment on a "fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". Elsevier; 1996.

[24] Kasahara K, Fukuda I, Nakamura H. A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer–DNA complex. PloS one. 2014;9(11):e112419.

[25] Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA: protein complexes. Proceedings of the National Academy of Sciences. 2009; 106(16):6620–6625.

[26] VanWart AT, Eargle J, Luthey-Schulten Z, Amaro RE. Exploring residue component contributions to dynamical network models of allostery. Journal of chemical theory and computation. 2012; 8(8):2949–2961.

[27] Freeman LC. Centrality in social networks conceptual clarification. Social networks. 1978;1(3):215–239.

[28] Mason O, Verwoerd M. Graph theory and networks in biology. IET systems biology. 2007;1(2):89–119.

[29] Newman ME. A measure of betweenness centrality based on random walks. Social networks. 2005;27 (1):39–54.

[30] Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006; 22(21):2695–2696.

[31] Skjærven L, Yao XQ, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC bioinformatics. 2014;15(1): 1–11.

[32] Skjærven L, Jariwala S, Yao XQ, Grant BJ. Online interactive analysis of protein structure ensembles with Bio3D-web. Bioinformatics. 2016;32(22): 3510–3512.

[33] Fujita A, Severino P, Kojima K, Sato JR, Patriota AG, Miyano S. Functional clustering of time series gene expression data by Granger causality. BMC systems biology. 2012;6(1):137.

[34] Möller-Levet CS, Klawonn F, Cho KH, Wolkenhauer O. Fuzzy clustering of short time-series and unevenly distributed sampling points. In: International Symposium on Intelligent Data Analysis. Springer; 2003. p. 330–340.

[35] Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. Bioinformatics. 2005;21 (suppl1):i159–i168.

[36] Pyatnitskiy M, Mazo I, Shkrob M, Schwartz E, Kotelnikova E. Clustering gene expression regulators: new approach to disease subtyping. PLoS One. 2014;9(1):e84955.

[37] Wang X, Nagarajan MB, Abidin AZ, DSouza A, Hobbs SK, Wismüller A. Investigating the use of mutual information and non-metric clustering for functional connectivity analysis on resting-state functional MRI. In: Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging. vol. 9417. International Society for Optics and Photonics; 2015. p. 94171N.

[38] Van Den Heuvel M, Mandl R, Pol HH. Normalized cut group clustering of resting-state FMRI data. PloS one. 2008;3(4):e2001.

[39] Gullo F, Ponti G, Tagarelli A, Tradigo G, Veltri P. A time series approach for clustering mass spectrometry data. Journal of Computational Science. 2012;3(5): 344–355.

[40] Košmelj K, Batagelj V. Cross-sectional approach for clustering time varying data. Journal of Classification. 1990;7(1):99–109.

[41] Iglesias F, Kastner W. Analysis of similarity measures in times series clustering for the discovery of building energy patterns. Energies.2013;6(2): 579–597.

[42] Fu Tc, Chung Fl, Ng V, Luk R. Pattern discovery from stock time series using self-organizing maps. In: Workshop Notes of KDD2001 Workshop on Temporal Data Mining; 2001. p. 26–29.

[43] Aghabozorgi S, Teh YW. Stock market co-movement assessment using a three-phase clustering method. Expert Systems with Applications.2014;41(4): 1301–1314.

[44] Montero P, Vilar JA, et al. TSclust: An R package for time series clustering. Journal of Statistical Software. 2014;62 (1):1–43.

[45] Sardá-Espinosa A. Comparing time-series clustering algorithms in r using the dtwclust package. R package vignette. 2017;12:41.

[46] Fajer M, Meng Y, Roux B. The activation of c-Src tyrosine kinase: conformational transition pathway and free energy landscape. The Journal of Physical Chemistry B. 2017;121(15): 3352–3363.