

# SCIENTIFIC REPORTS



OPEN

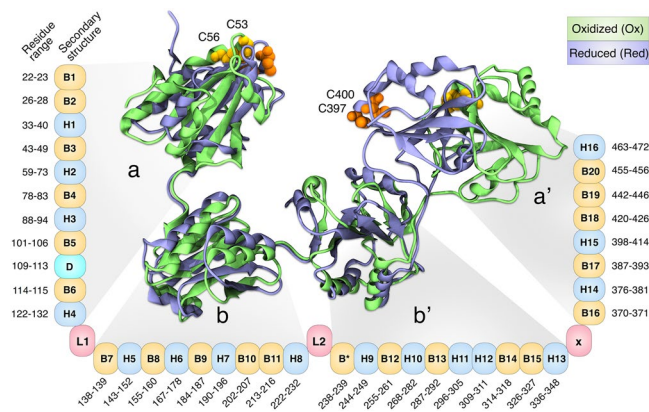
## Machine Learning and Network Analysis of Molecular Dynamics Trajectories Reveal Two Chains of Red/Ox-specific Residue Interactions in Human Protein Disulfide Isomerase

Razieh Karamzadeh<sup>1,2</sup>, Mohammad Hossein Karimi-Jafari<sup>3</sup>, Ali Sharifi-Zarchi<sup>4,5,6</sup>, Hamidreza Chitsaz<sup>6</sup>, Ghasem Hosseini Salekdeh<sup>2,7</sup> & Ali Akbar Moosavi-Movahedi<sup>1</sup>

The human protein disulfide isomerase (hPDI), is an essential four-domain multifunctional enzyme. As a result of disulfide shuffling in its terminal domains, hPDI exists in two oxidation states with different conformational preferences which are important for substrate binding and functional activities. Here, we address the redox-dependent conformational dynamics of hPDI through molecular dynamics (MD) simulations. Collective domain motions are identified by the principal component analysis of MD trajectories and redox-dependent opening-closing structure variations are highlighted on projected free energy landscapes. Then, important structural features that exhibit considerable differences in dynamics of redox states are extracted by statistical machine learning methods. Mapping the structural variations to time series of residue interaction networks also provides a holistic representation of the dynamical redox differences. With emphasizing on persistent long-lasting interactions, an approach is proposed that compiled these time series networks to a single dynamic residue interaction network (DRIN). Differential comparison of DRIN in oxidized and reduced states reveals chains of residue interactions that represent potential allosteric paths between catalytic and ligand binding sites of hPDI.

The human protein disulfide isomerase (hPDI), is one of the most abundant redox-regulated molecular chaperones accounting for the folding of almost one-third of proteins in cells<sup>1</sup>. As the first discovered protein-folding catalyst<sup>2</sup>, hPDI works as both an enzyme and a chaperone in various cellular processes including the oxidative stress, unfolded protein response, apoptosis and viral membrane fusion, using thiol disulfide exchange reactions<sup>3-6</sup>. A wide range of multifunctional features of hPDI are tightly associated with its unique molecular architecture. The horseshoe-like structure of hPDI is composed of four thioredoxin-like domains named a, b, b' and a' (Fig. 1). The N- and C-terminal domains – a and a' – contain conserved cysteine residues within CGHC motifs that are responsible for the formation, breakage and rearrangement of disulfide bonds on peptide/protein substrates; whereas the b and b' domains mostly contribute to substrate binding<sup>7</sup>. Targeted domain rearrangements during the redox-dependent activities of hPDI lead to the formation of two distinct “opened” and “closed”

<sup>1</sup>Department of Biophysics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. <sup>2</sup>Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran. <sup>3</sup>Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. <sup>4</sup>Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran. <sup>5</sup>Computer Engineering Department, Sharif University of Technology, Tehran, Iran. <sup>6</sup>Computer Science Department, Colorado State University, Fort Collins, Colorado, 80523, USA. <sup>7</sup>Department of Systems Biology, Agricultural Biotechnology Research Institute of Iran, Agricultural Research, Education, and Extension Organization, Karaj, Iran. Correspondence and requests for materials should be addressed to M.H.K.-J. (email: [mhkarimijafari@ut.ac.ir](mailto:mhkarimijafari@ut.ac.ir)) or A.A.M.-M. (email: [moosavi@ut.ac.ir](mailto:moosavi@ut.ac.ir))



**Figure 1.** Structural representation of hPDI in oxidized and reduced states superimposed from bb' domain along with secondary structure element of all residues. B: Sheet, H: Helix, D: Disordered region, L: Linker, X: X-linker. B\*: Exclusive sheet in oxidized form.

conformations in the oxidized (ox-hPDI) and reduced (red-hPDI) states, respectively<sup>7–10</sup>. Conformational transition of red/ox-hPDI is under the influence of a disulfide bond in the a' domain, which leads to the rearrangements of b' and a' domains, followed by switching its enzymatic activities<sup>7–11</sup>. Despite the available crystal structures of both states<sup>7</sup>, different functions of these conformations still conceal behind their dynamical complexities. Some studies have investigated the relations among the inter-domain flexibility and its effects on global domain motions of partial and full human and yeast PDI using limited proteolysis, SAXS, intrinsic fluorescence and NMR spectroscopies<sup>9,11–15</sup>. However, limited differences have been reported between these two redox states in terms of their distinct dynamical functions. From this perspective, identification of the structural determinants that control the dynamical behavior of each state is crucial for better understanding of their functions.

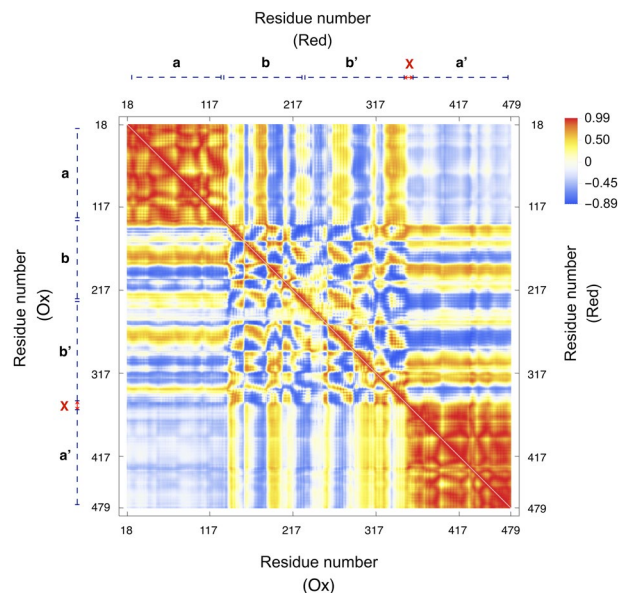
Differences in the dynamical behaviors of ox- and red-hPDI are the results of complex geometrical and physicochemical interplays of many residues as the structural units. An important question is which residues play more critical roles and through what mechanisms. In recent years, conversion of the 3D structure of proteins into a 2D network of interacting residues has found useful applications in dealing with the complexity of biomolecular structures<sup>16</sup>. Some tools such as RINerator and RING have been utilized to construct residue interaction networks (RIN) by considering geometrical, physicochemical, evolutionary and energetics of each residue<sup>17,18</sup>. Such networks could be used to compare and find hotspot players within the protein structure. However, the application of RIN is limited to the analysis of a single static snapshot of the protein structure obtained from experimental or computational resources. On the other hand, MD simulations are capable of producing a large number of conformations resulting from the time evolution of a protein. In these cases, each snapshot of the MD trajectory could be mapped to a corresponding RIN, and the dynamical analysis of such a huge number of networks would be a challenging task.

In the present work, we used MD simulations to reproduce the dynamical behaviors of hPDI in its oxidized and reduced states, independently. Regularities in large-scale domain motions of protein were then obtained through principal component analysis. Over the ensemble of the generated configurations, statistical machine learning methods were conducted to extract structural features that are related to the redox-dependent dynamics of proteins. By converting the dynamically sampled configurations to a sequence of residue interaction graphs, an approach was suggested, which provided a network based analysis of the dynamical differences of residue interactions between the two oxidation states.

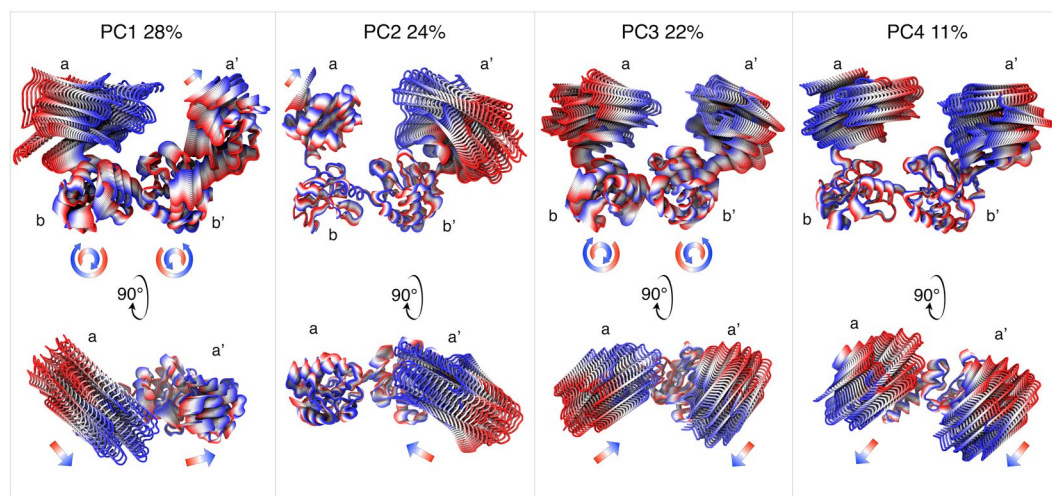
## Results and Discussion

**Pattern of domain motions.** Time series of backbone (root mean square deviation) RMSD values were monitored with respect to the initial structures to check whether the recorded atomic movements were caused by the collective domain motions or irregular random ones. Some trajectories showed RMSD values as large as 15 Å for whole protein backbone while the single domain values were mostly less than 2 Å (see Figure S1). Accordingly, most of the captured structural variations could be assigned to large scale domain motions and each of the single domains exhibited considerable structural stability within its domain boundaries. Moreover, starting from multiple initial configurations provides a more efficient sampling of the conformational space of protein. Collective atomic motions were characterized by cross-correlation matrix of atomic displacements around their average values (see Fig. 2). Clear domain boundaries and blocked structure of correlations were in agreement with patterns of RMSD values. Since all trajectories were superimposed to bb' domain, the central bb' blocks in Fig. 2 mainly reflect the intra-domain correlations, while other blocks reflect the inter-domain collective trends of atomic motions. The same overall patterns of correlations were obtained for both the oxidized and reduced trajectories. In both systems, terminal domains showed anti-correlated motions with respect to each other which is necessary for opening-closing rearrangements.

Essential modes of domain motions in hPDI were obtained from principal component analysis (PCA). The cumulative contribution of principal components (PCs) in total variance is shown in Figure S2. The first four PCs cover around 85% of total variance. As can be seen in Fig. 3, these essential components were different forms



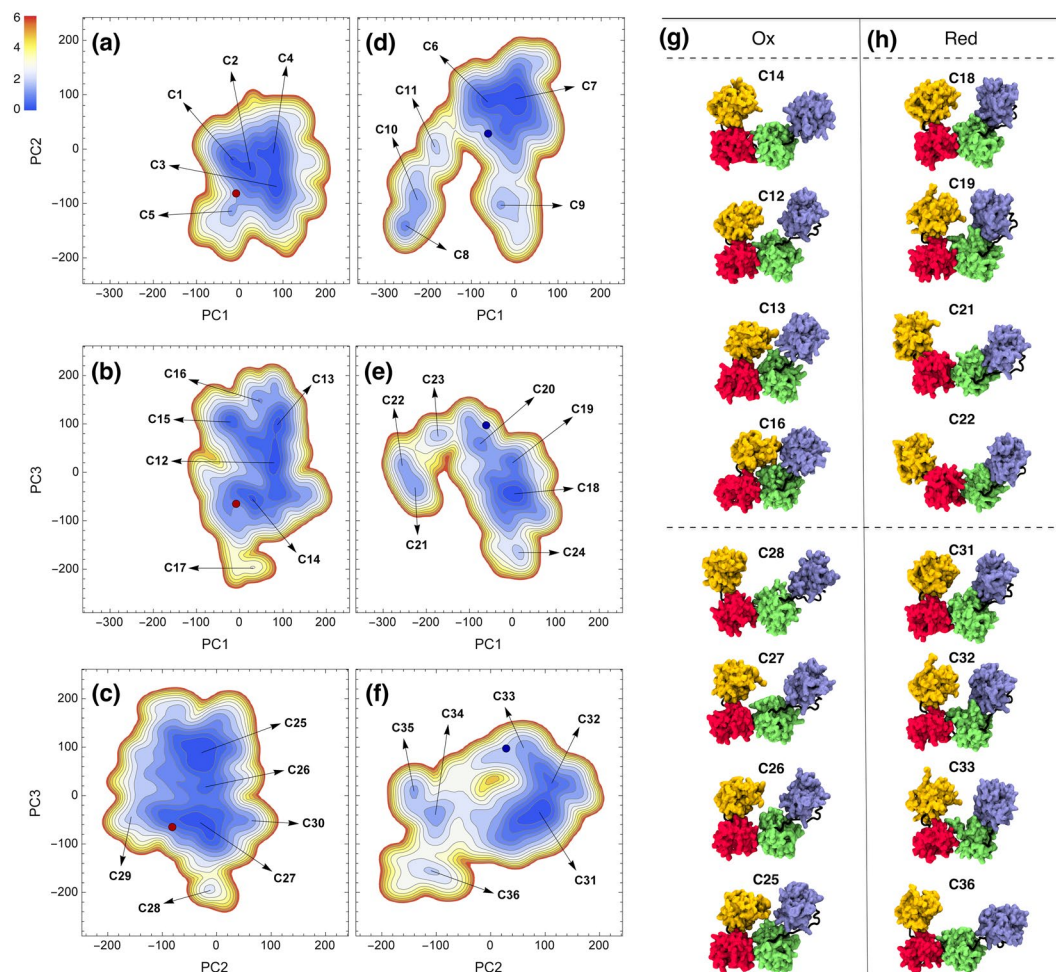
**Figure 2.** Cross-correlation matrix of atomic displacements around their average values for ox-hPDI (lower triangles) and red-hPDI (upper triangles). Values are ranged between 1 (red) and -1 (blue) for correlated and anti-correlated motions, respectively.



**Figure 3.** hPDI domain motion modes from PC1, PC2, PC3 and PC4 represented the dominant motions.

of hinge motions of domains a and a' with respect to the bb' base of protein. Single-domain hinge bending of terminal domains was the dominant motion in PC1 and PC2 while central domains showed a limited structural variation. In PC3 and PC4, both a and a' were involved in a concerted hinge motion with antiparallel and parallel directions, respectively. The antiparallel bending has a dominated contribution in total variance in agreement with the patterns in cross-correlation matrix as described previously. Moreover, PC3 imitated movements of the a and a' domains towards each other which could support previous studies showing the possibility of electron transfer between a and a' domains during the catalytic activities<sup>13, 19</sup>. The extent of the large amplitude domain motions identified in the current study was in agreement with a recent computational analysis of yeast PDI<sup>12</sup>. They reported a range of 15–80 Å for sampled values of inter-site distances. Nearly the same range of values was obtained here (Figure S3). The physical plausibility of fully closed conformations will be discussed in the next section.

It should be noted that while the hinge motions of a and a' domains were more highlighted in the first four PCs, there was a minor clockwise and counterclockwise twisting motion in b and b' domains which increased the amplitude of movements in the terminal domains. A quite different form of collective motion could be found in PC5 and PC6, which was a twisting of terminal domains around ab and a'b' inter-domain axes, respectively (Figure S4). Although a small fraction of structural variance was covered by these rotational movements, they are important for correct arrangement of catalytic sites towards substrates. The PCA approach was also used



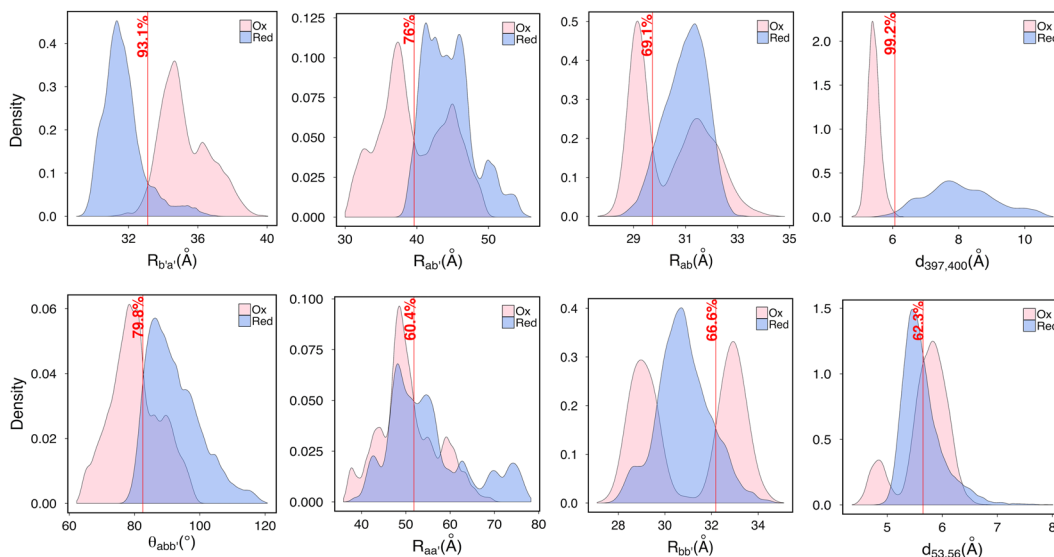
**Figure 4.** Projections of free energy landscape on two-dimensional spaces spanned by PC1-PC2, PC1-PC3 and PC2-PC3 in oxidized (a,b and c) and reduced (d,e and f) conformations. Conformers are numbered from C1 to C36. Representative conformers of FEL minima in oxidized (g) and reduced (h) states of hPDI. Yellow, red, green and blue stand for a, b, b' and a' domains, respectively.

to extract basic modes of motion in two-domain parts of protein including ab, bb' and b'a' (Figure S5). For all domain pairs, similar types of motion were obtained including two bending hinge-type motions in orthogonal directions and one twisting rotation along inter domain axis. Complicated four-domain motions were thus deduced from similar two-domain movements.

**Redox-dependent opening-closing routes.** Projections of free energy landscape (FEL) on two-dimensional spaces spanned by PC1-PC2, PC1-PC3 and PC2-PC3 were plotted in Fig. 4a–f. Some of the representative conformations presented in these figures, were depicted in Fig. 4g and h. The whole set of conformers could be found in Figures S6 and S7. Structural features of these conformers were reported in Table S1. All FEL projections of ox- and red-hPDI showed well-separated minima for open and closed conformations connected with low barriers. As could be seen in Fig. 4g, a path starting from open conformation of oxidized X-ray structure, passes through C14 and C12 conformations to reach C13 and C16 which were closed conformations. A similar closing route could be identified for ox-hPDI in PC2-PC3 subspace starting from C28, passing through C27 and C26 and reaching the C25 conformation. Comparison of these paths on FEL revealed that the ox-hPDI can easily switch between open and close states. In contrast, red-hPDI seems to be more restricted to intermediated semi-closed conformations similar to its X-ray structure. These semi-closed states were represented by a wide basin including C18, C19 and C20 conformations in PC1-PC3 subspace, or C31, C32 and C33 in PC2-PC3 subspace. In all of these states, the cleft between domains a and a' was not completely closed but was not as open as C34 and C21 conformations. Open states like C34, C35 and C36 in PC2-PC3 subspace or C21, C22 and C23 in PC1-PC3 subspace, were sampled in red-hPDI trajectories but FEL plots showed that they were less stable and less accessible from semi-closed states due to higher barriers. Moreover, fully closed conformations with a completely diminished cleft between a and a' domains were just seen on FEL of ox-hPDI (see C13, C16 and C25 conformations and their locations on FEL in Fig. 4). These opening-closing routes play a crucial role in regulation of substrate binding to hydrophobic pocket as reported in recent study about binding of  $\alpha$ -synuclein to hPDI<sup>20</sup>.

Features	$R_{ab}$	$R_{ab'}$	$R_{aa'}$	$R_{bb'}$	$R_{ba'}$	$R_{b'a'}$	$\theta_{abb'}$	$\theta_{bb'a'}$	$\Phi_{abb'a'}$
Linear discriminator	69%	76%	60%	67%	72%	93%	80%	60%	69%
Non-linear discriminator (SVM)	73%	76%	62%	80%	72%	93%	80%	61%	75%

**Table 1.** Accuracy of non-linear (SVM) and linear discriminator in classification of red/ox states based on different domain level features. R: inter-domain distance;  $\theta$ : angle;  $\Phi$ : dihedral torsion.

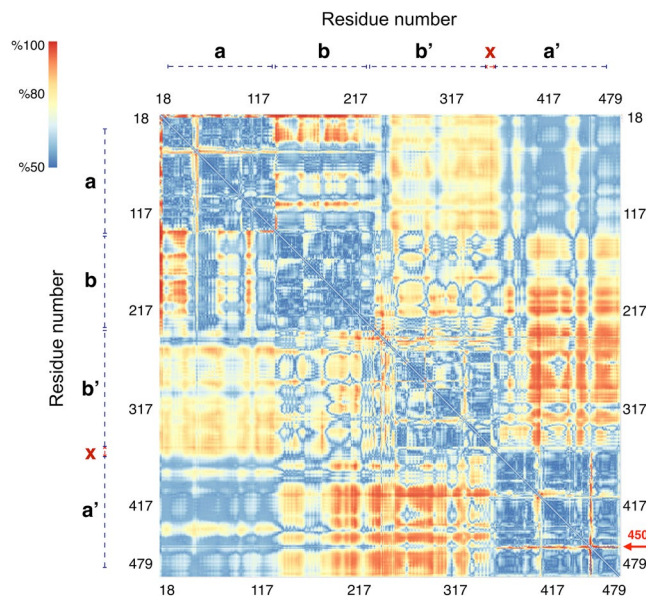


**Figure 5.** Domain and residue level distance distributions and their discrimination accuracies between the oxidized and reduced states of hPDI. Vertical red lines represent discrimination accuracy percentage. R: Domain distance; d: Residue distance;  $\theta$ : Inter-domain angle.

**Redox-dependency in dynamics of structural features.** To uncover which structural features of the hPDI could better discriminate between ox and red states, we employed both linear and non-linear classification methods<sup>21</sup>. With application of linear classifiers, we aimed to determine whether some structural features (e.g. inter-domain distances, angles, torsion or inter-residue distances) could discriminate ox and red snapshots of MD simulation by a simple threshold. We also used Support Vector Machine (SVM) with radial-basis kernel as a widely-used non-linear classifier since it could highlight more-complicated discriminative patterns in structural features such as bimodal distributions of domain level features (see Methods for more details).

Accuracy of linear and non-linear classifiers in discrimination of red/ox states, based on different domain level features, were reported in Table 1. As can be seen, the maximum discrimination was obtained from the inter-domain distance  $R_{b'a'}$  with an accuracy of 93%. The next most discriminative features were the inter-domain angle  $\theta_{abb'}$  and inter-domain distance  $R_{ab'}$  with classification accuracies of 80% and 76%, respectively. Different red/ox behavior of  $R_{b'a'}$  was also reflected in its separated distribution of values over ox- and red-hPDI trajectories in Fig. 5. Distributions of  $\theta_{abb'}$  and  $R_{ab'}$  had a considerable Red/Ox overlap but showed that the a and b' domains could be in closer contact with each other in ox-hPDI. Fully closed conformations as identified on ox-hPDI FEL were thus the result of the elongated distance between b' and a' domains, which permits domain a to be in close contact with b'. The wide range of sampled  $R_{b'a'}$  values in both ox- and red-hPDI was consistent with previous works that report the X-linker as the most flexible region of the protein<sup>9</sup>. However, it seems that the oxidized state was affected by the flexibility of the X-linker, to a higher extent. The underlying mechanism resulting in different values of  $R_{b'a'}$  in ox- and red-hPDI systems is discussed below. X-ray structures suggest different restricted values for  $R_{aa'}$  in ox- and red-hPDI but as could be seen in Fig. 5, they sampled a wide range of inter-domain distances between 35 to 80 Å, which indicates that both states have the ability of synergic cooperation of catalytic and chaperone activities upon various substrates.

Accuracy of the linear method in the case of residue-residue distances  $d_{ij}$  was presented as a residue-residue discrimination accuracy matrix (Fig. 6). Each element of matrix represents the accuracy of red/ox classification of conformations based on their  $d_{ij}$  values. Four domain features of protein were reflected in this matrix as indicated by the appearance of different intra- and inter-domain blocks. The most discriminative inter-domain  $d_{ij}$  values, were located between b' and a' domains, which was an expected result due to the different behaviors of  $R_{b'a'}$  in ox- and red-hPDI. However, in the case of intra-domain blocks, there were some hotspot residues that exhibited quite different distances with other residues in the same domain between ox- and red-hPDI. Moreover, A450 and D18 had different red/ox behavior which was highlighted as distinct red lines in corner aa and a'a' blocks, respectively (of Fig. 6). It should be noted that the A450 residue is in the middle of a loop connecting B19 and B20



**Figure 6.** Pairwise residue discrimination matrix between oxidized and reduced state of hPDI. Values range from 50% (blue) to 100% (red) discrimination accuracies.

sheets ( $L_{B19-B20}$ ). This loop had a conserved  $\omega$ -shaped structure over red-hPDI trajectories that became slightly distorted in ox-hPDI state.

Another important emerged from the linear classifier accuracies was the differential red/ox behavior of Cys-Cys distances in domains a and a'. Accuracy of the red/ox classifications based on  $d_{53,56}$  and  $d_{397,400}$  were 62% and 99%, respectively, as was evident from their distributions in Fig. 5. In the absence of the disulfide bond, C397-C400 distance took larger values in domain a' of red-hPDI as compared to that of ox-hPDI, while C53-C56 distance in domain a, showed the same range of values in both ox- and red-hPDI systems. This result was in agreement with previous studies showing the importance of redox sensitivity of the a' compared to the domain a<sup>10</sup>.

### A differential dynamic residue interaction network can highlight possible roadmap of redox-specific conformational changes.

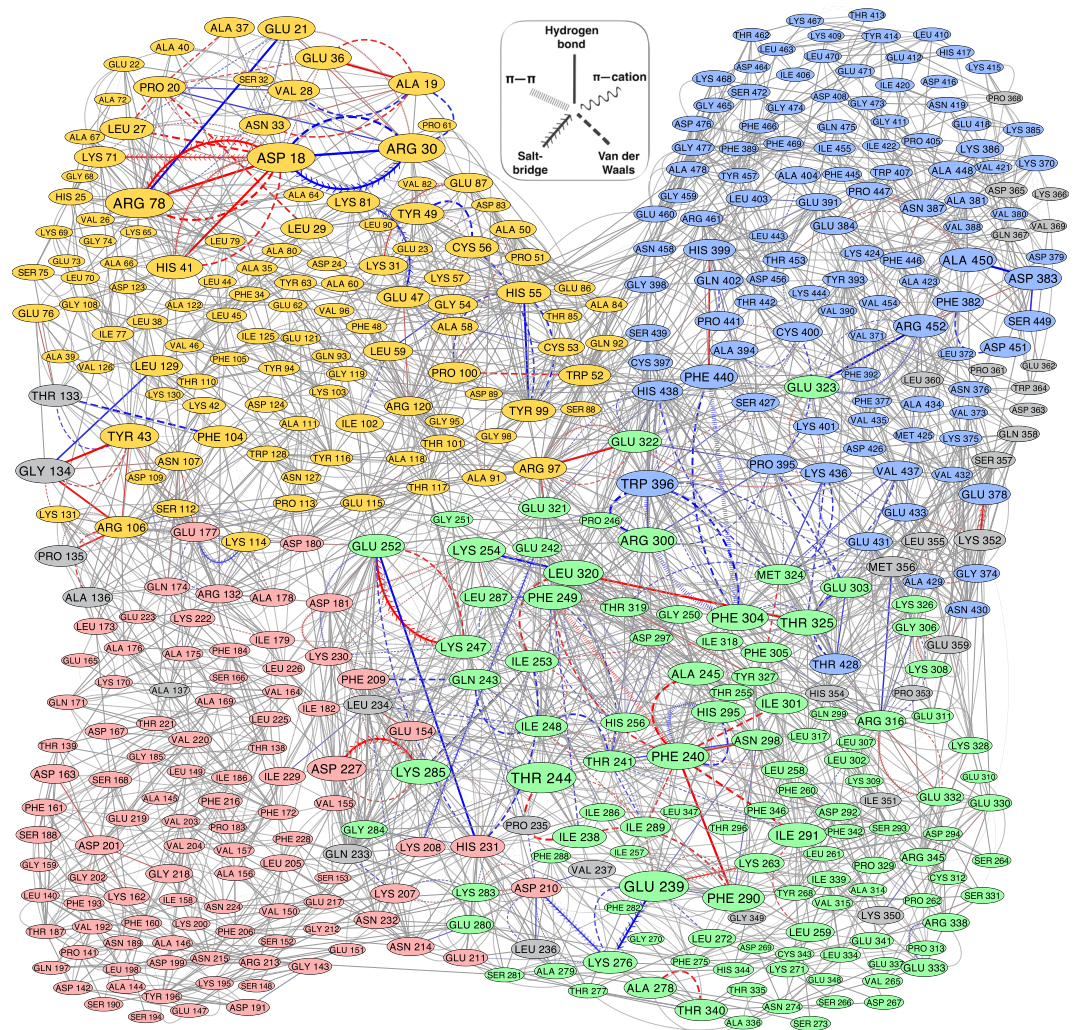
hPDI conformational changes are influenced by the allosteric redox regulation of disulfide bonds. Both causality and energetics of the allosteric pathway identify the mechanisms governing the regulation of this action<sup>22</sup>. There are some methods that examine the causality of the perturbations within the allosteric path<sup>23,24</sup>. Here, the time evolution of residue interaction networks were compiled to a dynamic residue interaction network named as DRIN. Since the DRIN graph reflected long lasting non-covalent interactions, a differential comparative DRIN graph could highlight central residues that are common or dominated in any of the ox- or red-hPDI systems. This representation was depicted in Fig. 7. As can be seen, the majority of red- or ox-dominated interactions occurred in a, b' and a' domains, which were responsible for substrate binding and enzymatic/chaperone activities of hPDI.

To test the importance of the identified hub residues of the DRIN graph on the functional activities of hPDI, we examined the situation of previously reported point mutations in differential DRIN graph (Table S2). Most of the point mutations that resulted in the alteration of hPDI activities were associated to the redox-dominated interacting residues. The majority of these mutations could affect the inter-domain coordination by changing the long-lasting interactions (Fig. 7 and Table S2). The network approach seems to be useful in identifying the critical residues in different oxidation states of hPDI.

The differential DRIN can explain the difference between ox- and red-hPDI in distribution of  $R_{bb'}$  by considering salt-bridges between D227, D210 and K208 in b domain with K285, K276 and E242 in the b' domain, respectively. These interactions which were mainly red-dominated locked the  $R_{bb'}$  around its unimodal distribution of values while their absence in ox-hPDI permitted two domains to rest in two different distances as reflected by bimodal form of  $R_{bb'}$  distribution (see Fig. 5).

Another example of the redox-dependent adjustment of inter-domain linkers can be found in region D of domain a in Fig. 1. This part of the sequence showed a sheet-coil oscillation during MD trajectories (Figure S8). The oscillations seemed to be biased to the sheet in red-hPDI as a result of the hydrogen bond of R106 to S112, while in ox-hPDI, R106 mainly interacted with G134 and P135 in L1 linker.

**Discrete lockstitch-like interactions in oxidized and reduced hPDI.** The machine learning approach used in this study revealed that b'a' inter-domain distance is a critical feature of different conformational dynamics of hPDI in reduced and oxidized states. We therefore attempted to identify residue-level interactions that regulate this feature. In Fig. 8 important red- and ox-dominated pairwise residue interactions were depicted on two representative conformations of red- and ox-hPDI, respectively. Two sequential interaction pathways in red-hPDI held b' and a' domains close to each other. The first one, denoted as *pathway I*, links H11 (e.g. R300, E303 and F304) and  $L_{B14-B15}$  (e.g. E322 and M324) of the b' domain to  $L_{B17-H15}$  (e.g. P395, W396) and  $L_{B18-B19}$  (e.g. K436, H438) of

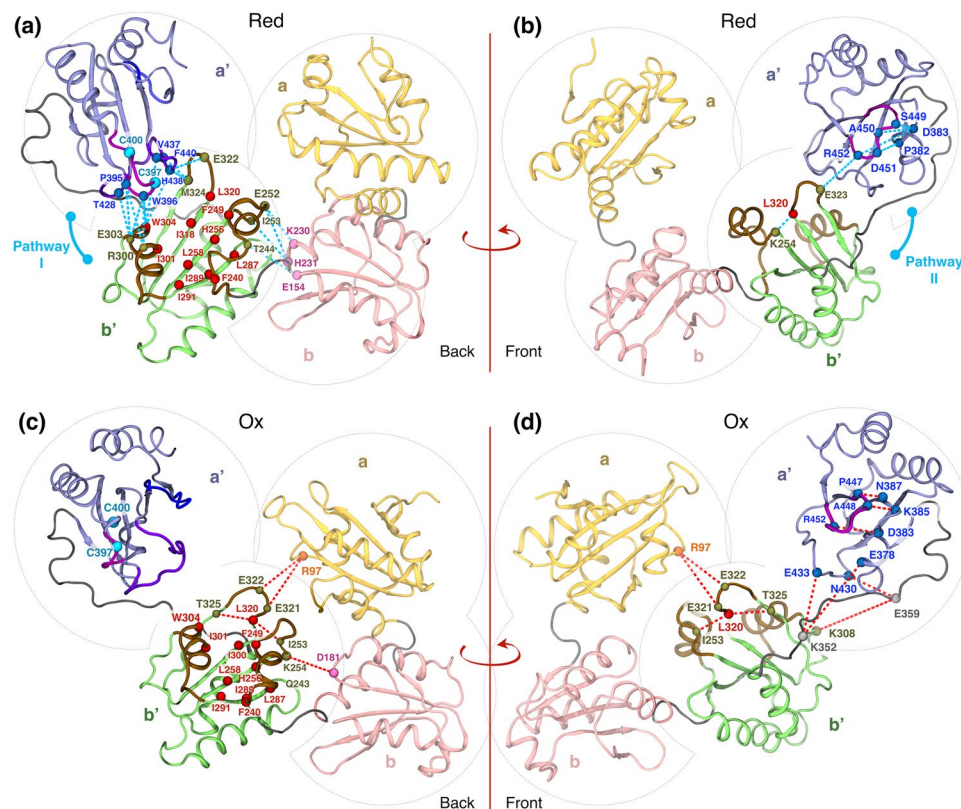


**Figure 7.** Dynamic residue interaction network (DRIN) in oxidized and reduced states of hPDI. The size of each vertex was assigned according to the maximum absolute weights of the edges arriving it. Sigmoid-like transformations from weights to the color or thickness of edges were used to show the interactions with unchanged weights between ox and red states as very thin grey edges, and the interactions with significantly higher persistence in one state as thick blue (red) or red (ox) edges. Nodes are colored by domains as yellow, red, green, blue and gray denote as domains a, b, b', a' and inter domain linkers, respectively.

the a' domain through hydrogen bonds, van der Waals and  $\pi$ - $\pi$  stacking interactions (Fig. 8a and b). Therefore, the catalytic site of a' (C397/C400) was connected to the core of hydrophobic pocket for further enzymatic/chaperone activities. The second sequence of interactions, denoted as *pathway II*, connected center of L<sub>B14-B15</sub> (E323) in the b' domain to L<sub>B19-B20</sub> (R452) in the a' domain via hydrogen bond and salt-bridge interactions (Fig. 8b). As noted before, these interactions keep the L<sub>B19-B20</sub> loop in a conserved  $\omega$ -shaped structure in red-hPDI. Two pathways (I and II) were connected to each other in the hydrophobic pocket of b' domain by involvement of L<sub>B14-B15</sub> (Fig. 8a and b). These interaction pathways act as a lockstitch that brings the b' and a' domains close to each other in red-hPDI.

In the absence of the red-dominated interaction of pathways I and II, different patterns of interactions occurred in ox-hPDI (Fig. 8c and d). The L<sub>B19-B20</sub> loop in domain a' became isolated with its own ox-dominated interactions, which distorted its  $\omega$ -shaped structure. New interactions were also created between the a' domain and the X-linker. In particular, interactions between K352 and E359 in X-linker with E378 and K308 in a' seem to adjust the b'a' distance indirectly based on the flexibility of X-linker (Fig. 8c and d).

Along with pathways I and II between the b'a' domains, additional chains of interactions seem to adjust the size of hydrophobic pocket in oxidized and reduced conformations. As can be seen in Fig. 8a, the rim of hydrophobic pocket mainly regulated by two helices (H9 and H11) and a loop between the B14-B15 (L<sub>B14-B15</sub>). As mentioned earlier, residues within the H11 and L<sub>B14-B15</sub> interacted with a' domain in red-hPDI through pathway I. In addition, residues around H9 also form red-dominated interactions with the b domain (K230, H231 and E154) based on hydrogen bonds, salt-bridge and van der Waals interactions. On the other hand, the H9 position was controlled by the formation of ox-dominated interactions in the oxidized state (e.g. D181—K254). These



**Figure 8.** Red- (a and b) and ox- (c and d) dominated residue-residue interaction networks on two representative conformations of red- and ox-hPDI (C3 and C6 of FEL) between the b' and a' domains. Residues ( $C_{\alpha}$ ) within the hydrophobic pocket in each state are colored in red. Red and blue dash-lines indicate the ox- and red-dominated interactions, respectively. Secondary structures involve in formation of the upper edge of hydrophobic pocket are colored in brown (H9, H11 and  $L_{B14-B15}$ ). Secondary structures of a' domain that contribute in  $ba'$  interactions are colored in violet ( $L_{B17-H15}$ ,  $L_{B18-B19}$  and  $L_{B19-B20}$ ).

results highlighted the importance of inter-domain long-lasting interactions in substrate selectivity of oxidized and reduced hPDI.

## Conclusion

hPDI plays diverse roles in various pathophysiological conditions such as the immune responses, coagulation, angiogenesis and even viral infection (reviewed in refs 25–27). Therefore, many attempts have been made to regulate its biological functions using small molecules for clinical applications<sup>28</sup>. Redox-associated structural changes are the key intrinsic feature to determine and control hPDI cross-functioning. For this propose, many studies have focused on finding the central residues that might play critical roles in manipulating different functions of hPDI<sup>28,29</sup>. In this study, we aimed to identify the redox-specific structural features of hPDI at residue and domain levels.

hPDI intrinsically controls its functions by balancing the structural features with long-lasting non-covalent interactions. Differential dominated interactions within the redox-regulated conformations could explain some of the inter- and intra-domain cooperative arrangements and suggest the mechanisms that potentially preserve the oxidized and reduced conformations. In addition, comparative analysis of the differences between the identified dynamic residue interaction network, and the earlier experimental point mutations elucidate the relative structural impacts of redox-dominated interactions on the biological function of hPDI and propose critical locations for future allosteric small molecule design. This approach can also be used for the extraction and discrimination of subtle structural dynamic alterations and hub residues between the redox-dependent conformations of other redox-dependent enzymes.

## Methods

**Molecular Dynamics Simulation.** X-ray crystal structures of ox-hPDI (PDB ID: 4EL1) and red-hPDI (PDB ID: 4EKZ) were used as the initial conformations for MD simulations<sup>7</sup> using the CHARMM 27 force field in the NAMD 2.10 package<sup>30</sup>. Short missing regions were recovered using Modeller 9.13<sup>31</sup>. Missing hydrogens were added by Reduce 3.23<sup>32</sup>. Each system was then solvated in a rectangular box of TIP3P water molecules and neutralized by sodium ions. Both systems were minimized and carefully equilibrated via NPT dynamics at 300 K and 1 atm. A Nosé-Hoover Langevin piston and a Langevin thermostat were used to control the pressure and temperature<sup>33,34</sup>, respectively. Periodic boundary conditions in conjunction with the particle mesh Ewald (PME)



method were applied<sup>35</sup>. To accelerate the conformational sampling, an initial 10 ns trajectory was clustered based on backbone RMSD in the VMD 1.9 package<sup>36</sup> and twelve representative structures (six for each system) from most populated clusters were used as starting points of 55 ns NPT dynamics. The whole procedure results in a sample of 156,000 structures from a total of 600 ns NPT simulation. More methodological details of MD simulations can be found in the supplemental text.

**Cross-Correlation and Principal Component Analysis.** The backbone RMSD for each domain and whole protein was calculated along each trajectory with respect to the first production frame. The secondary structure characters of all residues were calculated over trajectories using the STRIDE program<sup>37</sup>. The overall rotation and translation of the protein were eliminated by alignment from C $\alpha$  atoms of domains b and b', since the bb' pair of domains acts as a base for domain motions. To identify groups of residues with correlated motions, cross-correlation matrix of the atomic displacements was calculated<sup>38</sup> (see SI text for details). To quantitatively address domain motions, some inter-domain geometric variables were calculated over trajectories. Six inter-domain distances denoted as  $R_{ij}(i, j \in \{a, a', b, b'\})$ , two inter-domain angles  $\Theta_{abb'}$  and  $\Theta_{bb'a}$  and one torsion angle  $\Phi_{abb'a}$  were defined based on the domain centers (see Figure S9). Beside these mechanical “ball and spring” variable, a more comprehensive assessment of domain dynamics was performed via principal component analysis (PCA) of C $\alpha$  Cartesian coordinates (see Supplementary data for details). The standard PCA was performed for the whole protein and two-domain subsets of coordinates including ab, bb' and b'a' domain pairs to obtain basic two-domain motions that results in complicated four-domain dynamics. The most important collective modes of motions were selected based on their contribution in total variance. The kernel density estimation with a Gaussian kernel was used for the construction of the probability distribution functions (PDFs) in subspaces spanned by pairs of principal components. These PDFs were then used for the construction of projected FEL.

To check the performance of adopted multi-trajectory approach in sampling of conformational space of hPDI, the inner products between PCA eigenvectors obtained from all data and those of different halves of the data were compared in Figure S10. Comparison of diagonal and off diagonal elements in panel b of Figure S10 shows that the directions of important collective motions would be the same if one considers only the half-length of all trajectories while this was not true for full-length of all trajectories. Accordingly, starting from multiple configurations was crucial for efficient sampling and for avoiding from trapping regions of landscape.

**Statistical Machine Learning Methods.** Statistical machine learning methods were used to extract those structural features of protein which have a higher responsibility for the different dynamical behaviors of ox- and red-hPDI. Two types of structural variables were considered: *i*) Domain level features including  $R_{ij}$ ,  $\Theta_{abb'}$ ,  $\Theta_{bb'a}$  and  $\Phi_{abb'a}$  *ii*) Residue level features including pairwise residue-residue distances  $d_{ij}$ , defined as the Euclidean distance between C $\alpha$  atoms of residues *i* and *j* (see Figure S9). Values of the structural features in all trajectory snapshots were labeled as “ox” or “red” and a linear discrimination method was used for both residue and domain level features. In this regard, let *V* be any of the considered dynamical features, with observed ranges of values  $[V_{\min}^{\text{ox}}, V_{\max}^{\text{ox}}]$  and  $[V_{\min}^{\text{red}}, V_{\max}^{\text{red}}]$  in ox- and red-hPDI systems, respectively. The goal of the machine learning was to predict the “ox” or “red” label of each structure based on its *V* value. The linear method tried to find an optimum value denoted as  $V_{\text{sep}}$  from statistical distributions of *V* values in ox and red states. The features resulting in higher classification accuracy were assumed to be more important in description of dynamical differences between ox- and red-hPDI systems. Due to the bimodal distribution of some structural features (e.g.  $R_{bb'}$  distribution in ox states, see Fig. 5) we hypothesized that non-linear models might provide better classification accuracies than linear models, at least for a subset of features. To test this hypothesis, for each of 9 domain-level structural features (Table 1) we trained a separate Support Vector Machine (SVM) model to classify between ox and red states. The “svm” function of the “e1071” R package was used for this purpose, with default parameters and radial-basis kernel. To prevent overfitting, we assigned 156,000 available snapshots of both ox ( $N = 78,000$ ) and red ( $N = 78,000$ ) states into two disjoint training and test sets. Each snapshot was randomly assigned by equal probability to either training or test set, which resulted the sets to be of almost equal size. Each SVM model was trained based on one of the structural features to classify ox from red snapshots using training data, and then its performance was measured by test data. We observed identical classification accuracies between SVM and linear model for four structural features ( $R_{abb'}$ ,  $R_{ba'}$ ,  $R_{b'a'}$  and  $\theta_{abb'}$ ) and higher SVM accuracies for five other features ( $R_{ab'}$ ,  $R_{aa'}$ ,  $R_{bb'}$ ,  $\theta_{bb'a}$  and  $\Phi_{abb'a}$ ), which confirmed our hypothesis and also depicted that SVM models did not overfit to the training data (Table 1). As expected,  $R_{bb'}$  had the highest elevated accuracy between SVM and linear models (80% vs. 67%), which was due to the bimodal distribution of  $R_{bb'}$  in ox conformations (Fig. 5).

**Dynamic Residue Interaction Network.** The standalone version of RING 2.0 program was used for calculation of residue interaction network (RIN)<sup>17</sup>. Five types of non-covalent interactions, including hydrogen bond, van der Waals, salt bridge,  $\pi$ - $\pi$  and cation- $\pi$  interactions were considered. Each of the 156,000 structures was converted to a graph with residues as its vertices and the pairwise interactions of the residues as its edges. There could be different edges between a pair of residues due to different types of non-covalent interactions. For each type of interaction, a separate adjacency matrix of a RIN graph denoted as  $A_{ij}(t)$  was considered. From time series of  $A_{ij}(t)$ , the maximum interaction life time,  $\Gamma_{ij}$ , was then calculated according to equation (1) to build the DRIN matrix denoted as  $\Gamma$  for each type of non-covalent interactions.

$$\Gamma_{ij} = \arg \max_{\tau} \left\{ \exists \tau_0: \left( \prod_{t=\tau_0}^{\tau+\tau_0-1} A_{ij}(t) \right) = 1 \right\} \quad (1)$$

By this definition, for each type of interaction, 156,000 RIN graphs were compiled to two single DRIN graphs for each of the ox- and red-hPDI systems. To highlight the differences between the oxidized and reduced states, we computed the differential DRIN graph  $\Delta$  with adjusted weights that represented the fold change of  $\Gamma_{ij}$  values between ox- and red-hPDI systems. The elements of the differential DRIN matrix  $\Delta$  were computed for each type of interaction as equation (2):

$$\Delta_{ij} = \log_2 \left( \frac{\Gamma_{ij}^{\text{ox}} + \varepsilon}{\Gamma_{ij}^{\text{red}} + \varepsilon} \right) \quad (2)$$

where  $\varepsilon$  is a small positive value, which is considered to prevent the division by zero. By considering a cutoff on the absolute values of  $\Delta_{ij}$ , we could identify the pairs of residues that had a significant alternated pattern of interactions between the ox and red states. For visualization and analysis purposes Cytoscape 3.3.0<sup>39</sup> was used and differential DRIN graphs obtained for different types of interactions were merged together in a single graph with different edge styles.

## References

- Hatahet, F. & Ruddock, L. W. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid. Redox Signal.* **11**, 2807–50 (2009).
- Goldberger, R. F., Epstein, C. J. & Anfinsen, C. B. Acceleration of Reactivation of Reduced Bovine Pancreatic Ribonuclease by a Microsomal System from Rat Liver. *J. Biol. Chem.* **238**, 628–635 (1963).
- Hudson, D. A., Gannon, S. A. & Thorpe, C. Oxidative protein folding: From thiol-disulfide exchange reactions to the redox poise of the endoplasmic reticulum. *Free Radic. Biol. Med.* **80**, 171–182 (2015).
- Higa, A. & Chevet, E. Redox signaling loops in the unfolded protein response. *Cell. Signal.* **24**, 1548–1555 (2012).
- Grek, C. & Townsend, D. M. Protein Disulfide Isomerase Superfamily in Disease and the Regulation of Apoptosis. *Endoplasmic reticulum Stress Dis* **1**, 4–17 (2014).
- Walczak, C. P., Bernardi, K. M. & Tsai, B. Endoplasmic reticulum-dependent redox reactions control endoplasmic reticulum-associated degradation and pathogen entry. *Antioxid. Redox Signal.* **16**, 809–18 (2012).
- Wang, C. *et al.* Structural insights into the redox-regulated dynamic conformations of human protein disulfide isomerase. *Antioxid. Redox Signal.* **19**, 36–45 (2013).
- Serve, O. *et al.* Redox-dependent domain rearrangement of protein disulfide isomerase coupled with exposure of its substrate-binding hydrophobic surface. *J. Mol. Biol.* **396**, 361–374 (2010).
- Wang, C. *et al.* Plasticity of human protein disulfide isomerase: Evidence for mobility around the x-linker region and its functional significance. *J. Biol. Chem.* **285**, 26788–26797 (2010).
- Wang, C. *et al.* Human protein-disulfide isomerase is a redox-regulated chaperone activated by oxidation of domain a'. *J. Biol. Chem.* **287**, 1139–1149 (2012).
- Nguyen, V. D. *et al.* Alternative Conformations of the x Region of Human Protein Disulphide-Isomerase Modulate Exposure of the Substrate Binding b' Domain. *J. Mol. Biol.* **383**, 1144–1155 (2008).
- Römer, R. A. *et al.* The flexibility and dynamics of protein disulphide-isomerase. *Proteins Struct. Funct. Bioinforma.* **84**, 1776–1785 (2016).
- Yang, S. *et al.* Compact conformations of human protein disulfide isomerase. *PLoS One* **9**, e103472 (2014).
- Byrne, L. J. *et al.* Mapping of the ligand-binding site on the b' domain of human PDI: interaction with peptide ligands and the x-linker region. *Biochem. J.* **423**, 209–217 (2009).
- Li, S. J., Hong, X. G., Shi, Y. Y., Li, H. & Wang, C. C. Annular arrangement and collaborative actions of four domains of protein-disulfide isomerase: A small angle x-ray scattering study in solution. *J. Biol. Chem.* **281**, 6581–6588 (2006).
- Csermely, P. Creative elements: network-based predictions of active centres in proteins and cellular and social networks. *Trends Biochem. Sci.* **33**, 569–576 (2008).
- Piovesan, D., Minervini, G. & Tosatto, S. C. E. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* **44**, W367–W374 (2016).
- Doncheva, N. T., Klein, K., Domingues, F. S. & Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* **36**, 179–182 (2011).
- Wang, L., Wang, X. & Wang, C. C. Protein disulfide-isomerase, a folding catalyst and a redox-regulated chaperone. *Free Radic. Biol. Med.* **83**, 305–313 (2015).
- Yagi-Utsumi, M., Satoh, T. & Kato, K. Structural basis of redox-dependent substrate binding of protein disulfide isomerase. *Sci. Rep* **5**, 13909 (2015).
- Cai, Y. D., Liu, X. J., Xu, X. B. & Chou, K. C. Prediction of protein structural classes by support vector machines. *Comput. Chem.* **26**, 293–296 (2002).
- Guarnera, E. & Berezovsky, I. N. Allosteric sites: Remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.* **37**, 1–8 (2016).
- Allain, A. *et al.* Allosteric Pathway Identification through Network Analysis: from Molecular Dynamics Simulations to Interactive 2D and 3D Graphs. *Faraday Discuss.* **169**, 303–321 (2014).
- Guarnera, E. & Berezovsky, I. N. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Comput. Biol.* **12**, e1004678 (2016).
- Flaumenhaft, R., Furie, B. & Zwicker, J. I. Therapeutic implications of protein disulfide isomerase inhibition in thrombotic disease. *Arterioscler. Thromb. Vasc. Biol.* **35**, 16–23 (2015).
- Parakh, S. & Atkin, J. D. Novel roles for protein disulphide isomerase in disease states: a double edged sword? *Front. cell Dev. Biol* **3**, 30 (2015).
- Gallina, A. *et al.* Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *J. Biol. Chem.* **277**, 50579–50588 (2002).
- Xu, S., Sankar, S. & Neamati, N. Protein disulfide isomerase: A promising target for cancer therapy. *Drug Discov. Today* **19**, 222–240 (2014).
- Kaplan, A. *et al.* Small molecule-induced oxidation of protein disulfide isomerase is neuroprotective. *Proc. Natl. Acad. Sci.* **112**, E2245–E2252 (2015).
- Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
- Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science* **Chapter 2** (2007).
- Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).
- Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177 (1994).

34. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613 (1995).
35. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
36. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
37. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinforma.* **23**, 566–579 (1995).
38. Ichiye, T. & Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct. Funct. Bioinforma.* **11**, 205–217 (1991).
39. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

## Acknowledgements

Molecular dynamics simulations and data analysis were performed using the Computing Cluster Facility of the Computer Science Department, Colorado State University, Fort Collins, Colorado 80523, USA. The support of the University of Tehran, Royan Institute, and Iran National Science Foundation (INSF) is gratefully acknowledged.

## Author Contributions

M.K., A.M., A.S., G.H.S. and R.K. were involved in designing the project. M.K., A.S. suggested the analysis methods and mathematical framework. M.K., A.S. and R.K. developed the computational models and analyzed the data. R.K. reviewed the biological concepts and prepared figures. R.K., A.S. and M.K. wrote and/or reviewed the manuscript. A.M., G.H.S. and H.C. proofread the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-03966-5](https://doi.org/10.1038/s41598-017-03966-5)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017