Journal of Molecular Graphics and Modelling 28 (2009) 347–356

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Molecular dynamics simulations of the conformational changes in signal transducers and activators of transcription, Stat1 and Stat3

Jianping Lin, Ralf Buettner, Yate-Ching Yuan, Richard Yip, David Horne, Richard Jove, Nagarajan Vaidehi *

Departments of Immunology and Molecular Medicine, Beckman Research Institute, City of Hope, 1500 E Duarte Road, Duarte, CA 91010, United States

ARTICLE INFO

Article history:
Received 22 April 2009
Received in revised form 28 August 2009
Accepted 31 August 2009
Available online 6 September 2009

Keywords:
Molecular dynamics
Signal transducers and activators of transcription
Clustering
Correlation map
Domain motion
Druggable pocket

ABSTRACT

All signal transducers and activators of transcription (STAT) factors are a family of cytoplasmic transcription factors that mediate the signal response to cytokines, growth factors, and hormonal factors. The phosphorylation and subsequent activation of Stat3, a member of the STAT family, has been found to be elevated in a large number of diverse human cancers. Understanding of the dynamics of the Stat3 dimer interface is pertinent to designing small molecule inhibitors to activated Stat3 dimer. To this end, we have performed molecular dynamics simulations in explicit water of the activated Stat3 homodimer, and also its closely related member of the STAT family, activated Stat1 homodimer. We observed a large-scale domain motion in the Stat3 dimer while the structure of the monomer remains intact. The driving force for this conformational change is enhanced binding of the Stat3 dimer to the DNA, thereby regulating gene expression. Our model shows that the carboxy terminus of one monomer wraps around the core of the SH2 domain of the other monomer, and this region that makes up the dimer interface remains intact during the dynamics. Water diffuses into a cavity under this dimer interface, thus expanding a pre-existing cavity that gets gated and closed by the loops in the SH2 domain. This cavity could serve as a potential binding pocket for inhibitor design.

1. Introduction

Signal transducers and activators of transcription (STAT) factors are a family of cytoplasmic transcription factors that mediate signal responses to cytokines, growth factors, and hormonal factors [1–3]. The STAT family comprises of seven members in mammalian cells namely, Stat1, Stat2, Stat3, Stat4, Stat5a, Stat5b, and Stat6 [1]. Upon recruitment to cell surface receptors, STAT proteins become phosphorylated at a specific carboxy terminal tyrosine by Janus kinases (JAKs) or receptor-intrinsic kinases. Phosphorylation of this tyrosine induces activation and nuclear translocation of dimerized STAT proteins. Once in the nucleus, activated STAT dimers bind to specific target genes and regulate gene expression [4–7]. Activation of Stat1, Stat3, and Stat5 have been found to be elevated in a large number of diverse human cancers including blood malignancies (multiple myeloma, leukaemias, lymphomas) and solid tumors [8,9]. The constitutive activation of Stat3 leads to tumor formation, cell proliferation and tumor growth [10], whereas elevated Stat1 activation has been associated with growth suppression rather than malignant transformation, and thus can be considered a potential tumor suppressor [11]. In contrast to Stat1 function, Stat3 has been identified as an oncogene and therefore, selective targeting of disrupting the activated Stat3 dimer is an attractive strategy for cancer therapy. However disrupting protein–protein interaction with small molecules is an emerging therapeutic strategy and a challenging task.

Structurally, Stat1 and Stat3 proteins are homodimers with the possibility of formation of heterodimers [12]. The monomer of both Stat1 and Stat3 is a multi-domain protein with 750–850 amino acids, and share 50% pairwise sequence identity (Fig. S1 in the supporting material shows the sequence alignment between Stat1 and Stat3). Stat3 and Stat1 protein structures consist of a coiled-coil domain (residues 138–320 for Stat3, residues 136–317 for Stat1), a DNA-binding domain (residues 321–465 for Stat3, residues 318–488 for Stat1), a connector domain (residues 466–585 for Stat3, residues 489–576 for Stat1), a Src homology 2 (SH2) domain (residues 586–688 for Stat3, residues 577–683 for Stat1), and a carboxy terminus that is called the transcriptional activation domain (residues 702- for Stat3, residues 700- for Stat1), as shown in Fig. 1A and B [13,14]. STAT proteins could exist in dimeric form even in the absence of phosphorylation of the conserved Tyr in the SH2 domain [15], but it is well understood that homodimers or heterodimers form via tight interaction of the phosphotyrosine in the SH2 domain from one monomer with the other monomer following tyrosine phosphorylation [13,14]. It has been speculated that the conformational change that accompanies tyrosine

* Corresponding author at: Department of Immunology, Beckman Research Institute, City of Hope, 1500 E Duarte Road, Duarte, CA 91010, United States.
Tel.: +1 626 301 8408; fax: +1 626 301 8186.
E-mail address: NVaidehi@coh.org (N. Vaidehi).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.08.013

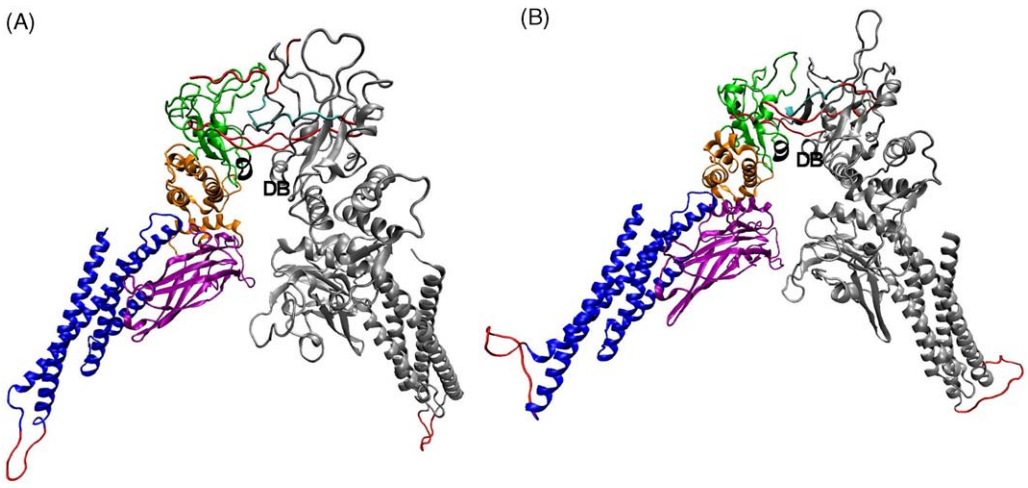

Fig. 1. (A) Crystal structure of the Stat3 dimer (pdb code: 1BG1) with the added missing loops (B). Crystal structure of the Stat1 dimer (pdb code: 1BF5) with the missing loops added; coiled-coil domain is in blue; DNA-binding domain is in purple; the connector domain is in orange; the SH2 domain is in green; the carboxy terminus is in cyan; the missing loops added with MODELLER are in red, DB represents DB loop region.

phosphorylation strengthens the dimer as well as improve their ability to bind specific target DNA [13].

Our end goal is to design Stat3 selective small molecules that would disrupt the preformed, activated dimers of Stat3. Many known Stat3 dimerization inhibitors target the well conserved phosphorylated tyrosine site in the SH2 domain that may not confer Stat3 selectivity. We took a different approach to use molecular dynamics (MD) simulations to examine the role of water in the dynamics of Stat3 dimer interface and Stat3/DNA interface possibly leading to alternate, less conserved and potentially druggable binding pockets in the Stat3 dimer structure. MD simulations starting with the crystal structure offer insight into localized motion and low frequency motion that lead to conformational changes in proteins. In this work, we perform MD simulations on the carboxy terminus truncated Stat3β to understand the comparative conformational flexibility of Stat3 versus Stat1 dimers, and provide insight into the protein–protein interactions in the dimer. Concisely, the results of the MD simulations show that the Stat3 dimer undergoes a significant conformational change resembling "scissor-like" motion when bound to the DNA. This conformational change leads to the reorientation of the DNA-binding domain with respect to the dimer interface, resulting in tighter binding of Stat3 to the DNA. These findings are in agreement with the experimental evidence that the Stat3β shows both increased DNA-binding activity and dimer stability [16]. *k*-means clustering algorithm and principal component analysis show five distinct conformational clusters for Stat3, and eight conformational clusters for Stat1 during the MD simulations. However the five conformational clusters for Stat3 show distinct and large conformational changes, while such large-scale motions were not observed for the Stat1 dimer. We observe that the large conformational change in the Stat3 dimer leads to stronger interaction energy of the Stat3 protein to the bound DNA. Water penetrating into the dimer interface of Stat3 leads to formation of a potential binding pocket in the dimer interface that can be targeted for inhibitor design. We have further used the conformations from the MD simulations to successfully design inhibitors for Stat3 dimerization and these results will be published elsewhere.

2. Computational methods

2.1. System setup and molecular dynamics simulations

The crystal structures of the Stat3 and Stat1 monomers were downloaded from the pdb entries 1BG1 [13] and 1BF5 [14] respectively. These are the monomer structures with one strand of DNA bound to Stat3, and double-stranded DNA bound to Stat1. We generated the dimer structures of Stat3 and Stat1 from the pdb files and added the missing residues in Stat3 (residues 185–193, 689–701, and 717–722) using Modeller [17–20] (Fig. 1A and Fig. S2A show the added loop of Stat3). The complementary chain of the DNA in the Stat3 structure was added using VMD [21]. The missing residues 183–196 in a loop region of Stat1 were also added using Modeller, and the residues 684–699, near the carboxy terminus, were added in a conformation similar to that of Stat3 to keep both structures consistent. The missing atoms in the Stat1 structure were added using VMD. Sodium and chloride counter-ions were added using VMD to neutralize the system. The protein was solvated using the water box in "solvate plugin" in VMD. The total number of atoms in the solvated Stat3, Stat1 systems are 234,009, 292,759, respectively. Fig. S3 of the Supplementary material shows the starting conformations of Stat3 and Stat1 and the MD simulations were performed using NAMD [22] program. We used the CHARMM 27 force field [23], NPT ensemble with periodic boundary conditions, and particle-mesh Ewald (PME) full electrostatics calculations [24]. TIP3P model was used for water [25]. Initially, the added loops were minimized using conjugate gradient method with the rest of the protein fixed, and minimizing the potential energy of the solvent at 0 K temperature, for 1000 steps. Subsequently the whole system was slowly heated to 310 K (increased the temperature by 1 K per 5 steps) within 100 ps. The production MD was run for the whole system for 50 ns for both Stat3 and Stat1 system (Fig. S7) and the snapshots from the MD runs were saved every 5.0 ps.

2.2. Principal component analysis

The conformational changes in Stat3 and Stat1 have been analyzed using the principal component analysis. The positional covariance matrix from the MD trajectory is given by

$$\Gamma_{ij} = \langle (x_i(t) - \langle x_i(t) \rangle) \times (x_j(t) - \langle x_j(t) \rangle) \rangle \quad (1)$$

where Γ_{ij} is the element of covariance matrix, $x_i(t)$ and $x_j(t)$ are the Cartesian coordinates of atom i and atom j, respectively, at time t, $\langle\rangle$ represents time average over the trajectory. The principal modes were obtained by diagonalizing the covariance matrix, and the eigenvectors associated with the largest, and second largest eigenvalues were selected as the two significant principal

components that form the axes of the essential space. The MD snapshots at every 50 ps were then projected to the selected principal components. The principal component analysis was performed using *ptraj* of AMBER 8 [26]. Due to the matrix size limit, we use the coordinates of C_α atoms of the protein, and the backbone atoms of the DNA to generate the covariance matrix.

2.3. Cluster analysis

k-mean clustering algorithm has been used to identify the structural representations of conformation families that are most populated during MD simulations [27,28]. This analysis would offer insight into the nature of the large-scale conformational changes, if any in Stat3 and Stat1 dimers. With the projections of MD snapshots onto the essential space [29], we calculated the distribution of the conformer density in the essential space, ρ, and the corresponding molecular potential given by [28]

$$E = -RT\ln(\rho) \quad (2)$$

where R is the gas constant and T is temperature. The adaptation of *k*-means clustering algorithm available in "*Weka*" [30] was applied to projections of MD snapshots in the essential space to identify the conformation clusters. The *k* values were refined from 10 clusters to the final value based on mapping back each cluster to the contour map of molecular potential energies.

2.4. Cross-correlation analysis of MD trajectories

The correlated motion between pairs of residues in proteins can be used to probe the inter-domain coupling of proteins during the MD simulations [31]. Correlated motion between pairs of residues can be represented by cross-correlation coefficient, C_{ij}, between atom i and atom j, which is given by [31]

$$C_{ij} = \frac{\langle \Delta r_i \times \Delta r_j \rangle}{(\langle \Delta r_i \times \Delta r_i \rangle \langle \Delta r_j \times \Delta r_j \rangle)^{1/2}} \quad (3)$$

where Δr_i and Δr_j are the displacement vectors for atom i and atom j, respectively. $\langle\rangle$ represents average over the entire trajectory. The value of C ranges from -1 to 1, where negative sign signifies uncorrelated motion of the residue pair in opposite directions, and a positive sign shows a correlated motion of the residue pair in the same direction. In our calculations, the correlated motion of a residue pair is represented by the motion of the C_α atoms of those residues. The cross-correlation coefficient matrix was calculated using GROMACS [32] and the correlation maps were generated using MATLAB [33].

2.5. Calculation of the interaction energy

To understand the strength of interaction between various domains of Stat3 and Stat1, we have calculated the interaction energy, which is the sum of van der Waals and electrostatic energies between the stipulated residues in domains, or between the DNA and the protein. For each trajectory from our MD simulations, we used NAMD energy plugin in VMD [21] to calculate the interaction energies for each snapshot of the MD trajectory, using a cutoff of 12 Å for the non-bond energy calculations.

3. Results and discussion

3.1. Root mean square deviation (RMSD) in coordinates

Fig. 2 shows the calculated root mean square deviation (RMSD) in coordinates (labeled as CRMSD henceforth) of the C_α atoms for the two monomers, and the dimer of both Stat3 and Stat1 from

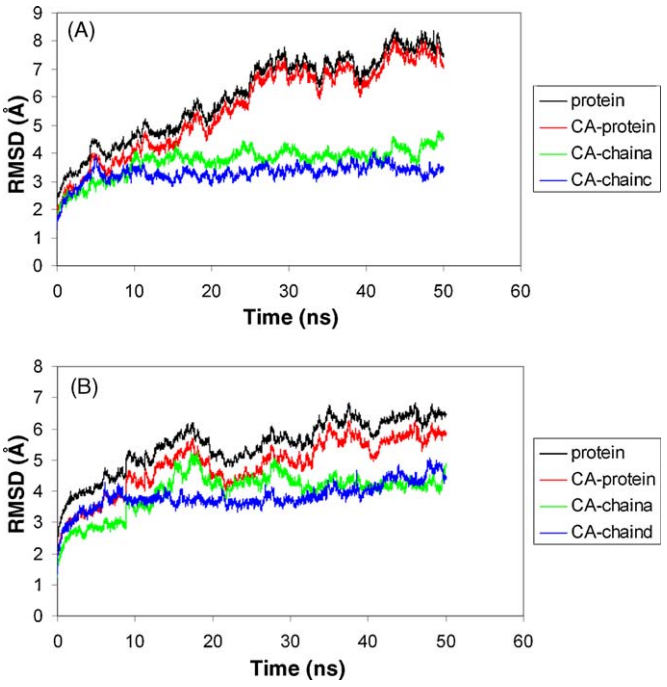

Fig. 2. Root mean square deviation (RMSD) in coordinates as a function of the simulation time (A) Stat3 and (B) Stat1. Protein represents RMSD of entire STAT protein with all atoms. CA-protein represents RMSD of C_α atoms (CRMSD) of STAT protein. CA-chaina represents CRMSD of the C_α atoms of monomer A. CA-chainc represents CRMSD of the C_α atoms of monomer C. CA-chaind represents CRMSD of C_α atoms of the monomer D.

their respective crystal structures. It is seen in Fig. 2A, that the CRMSD of both the Stat3 monomers stabilize at 3.5 Å, while there is a 2.0 Å increase in the CRMSD of the Stat3 dimer between 20 and 30 ns. This jump in CRMSD for the dimer is stabilized further

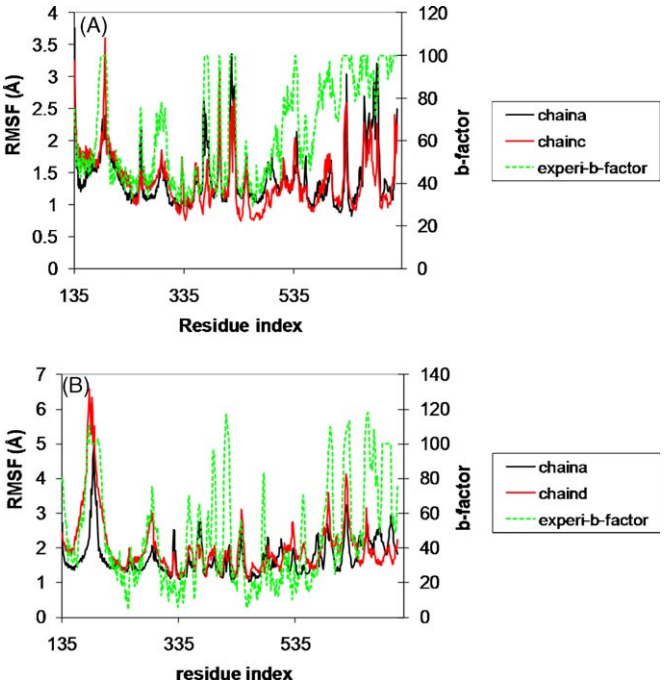

Fig. 3. RMSF of C_α atoms in coordinates for each residue averaged over the last 5 ns of the MD trajectory compared to the experimental B-factor from the crystal structures (A) Stat3, chaina and chainc represent monomers A and C respectively. (B) Stat1, chaina and chaind represent monomers A and D respectively. Experi-b-factor represent experimental b-factor [from references 13,14].

between 30 and 50 ns of the MD simulation. This indicates that the dimer undergoes substantial conformational change while the internal domain structure of the monomer remains less dynamic during the MD simulations. As seen in Fig. 2B, the CRMSD of Stat1 monomers stay intact within 3.5 Å, and the CRMSD for the dimer show less change than Stat3 dimer. These results indicate that the dimer of Stat3 undergoes significant conformational changes compared to the Stat1 dimer.

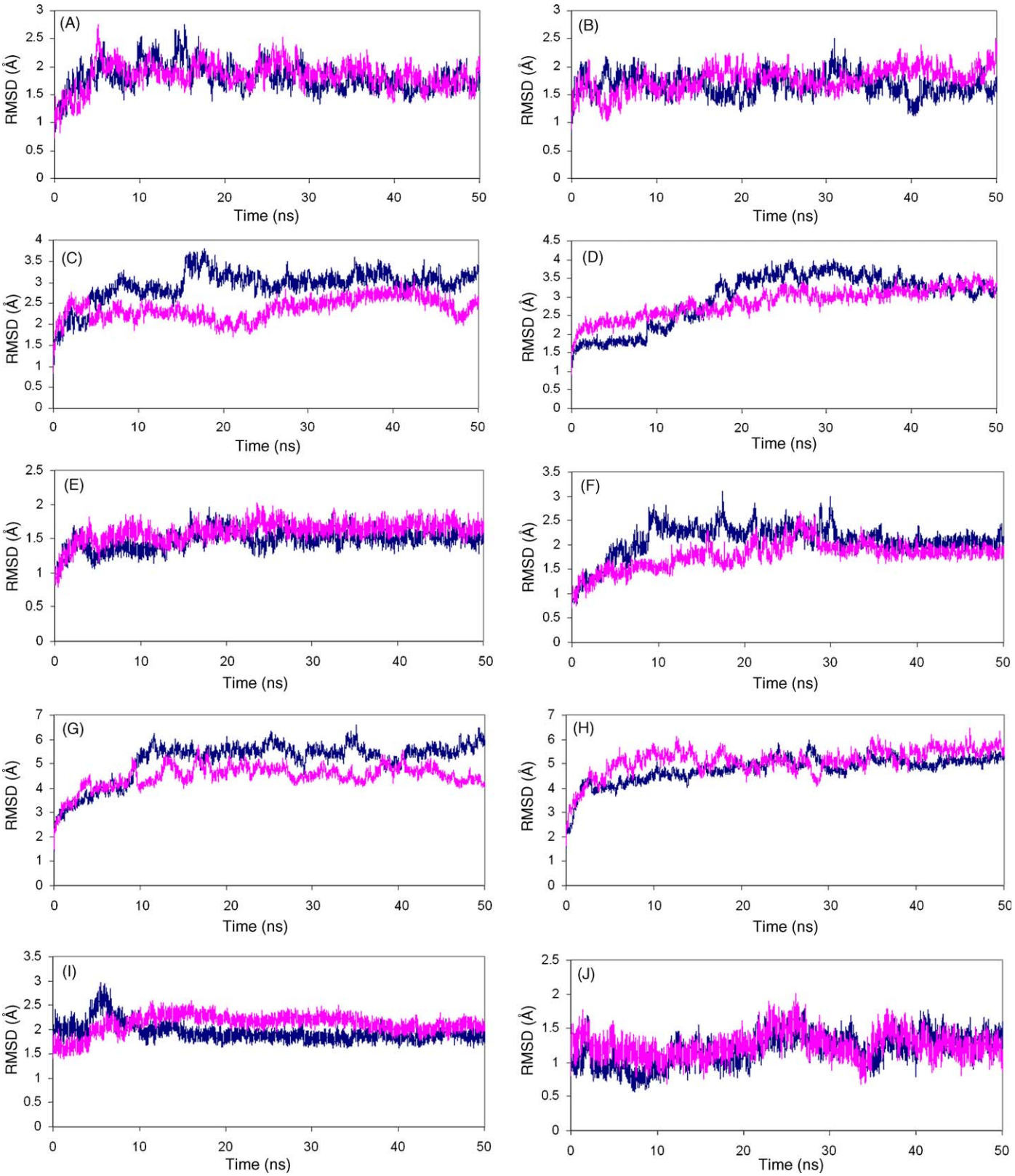

Fig. 4. RMSD in coordinates of the C_α atoms for different domains. (A) Stat3, coiled-coil domain without added loop; (B) Stat1, coiled-coil domain without added loop; (C) Stat3, DNA-binding domain; (D) Stat1, DNA-binding domain; (E) Stat3, connector domain; (F) Stat1, connector domain; (G) Stat3, SH2 domain; (H) Stat1, SH2 domain; (I) Stat3, carboxy terminus (residues 708–716); (J) Stat1, carboxy terminus (residues 703–710).

3.2. Root mean square fluctuation (RMSF) of C_α atoms by residues and CRMSD by domains

To understand the dynamic flexibility in each domain, we have calculated the RMSF of C_α atoms for each residue in Stat1 and Stat3 monomer, averaged over the last 5 ns of the MD simulations. Fig. 3 shows the comparison of the RMSF for each residue to the experimental B-factors taken from the corresponding pdb files [13,14]. The B-factors of the missing residues in the crystal structures have been arbitrarily assigned a maximum value of 100. Fig. 3 shows that the calculated RMSF by residue, matches well with the trends of the fluctuations in the B-factors, except for the region between residues 535 to the carboxy terminus where the B-factors are high. Fig. 4 shows the variation in CRMSD with time for various domains of Stat3 and Stat1. For both Stat1 and Stat3, the loop connecting the helices (residues ranging from 183 to 196) in the coiled-coil domain is very flexible and shows a RMSF by residue in the range of 4–6 Å. The rest of the coiled-coil domain (residues range from 136 to 182 and 197 to 320) is quite rigid with a CRMSD less than 3 Å as seen in Fig. 4A and B. Fig. 4C and D shows that the residues in the DNA-binding domain are flexible and the CRMSD goes up to 3.5–4.0 Å for Stat3 and Stat1, which is in corroboration with the variation of the RMSF by residue of the DNA-binding domain (residues 321–465 for Stat3 and residues 318–488 for Stat1) in Fig. 3. For Stat3, the residues, H332, K340, Q344, R382, R417, V432, S465, N466, and Q469 that make contact with the DNA [13] are rigid with their RMSF varying from 0.8 to 1.7 Å. For Stat1, the important residues T327, H328, Q340, R378, K410, E411, K413, E421, T427, T459, D460, and K567 that make contact with the DNA [14] are rigid with a RMSF that varies from 1.0 to 2.5 Å. In Fig. 3, it is seen that the connector domain ranging from residues 466 to 585 for Stat3 and from residues 489 to 576 in Stat1, is rigid with a RMSF by residue less than 2 Å. This is in agreement with the instantaneous CRMSD of the connector domain that is also less than 3 Å, as seen in Fig. 4E and F.

The variations in CRMSD and RMSF show the highest fluctuations for the SH2 domains in both Stat3 and Stat1 as shown in Figs. 3 and 4G and H with a RMSF of more than 2.5 Å, and a CRMSD for the entire SH2 domain of 5–6 Å. Residues K591, R609, S611, and S613 show direct polar interactions with phosphotyrosine 705 in Stat3 and these residues are quite rigid during the MD simulations. Residues K584, R602, S604 and S606 that show direct polar interactions with phosphotyrosine 701 in Stat1, are also less flexible. Summarizing, the flexibilities observed in the DNA-binding domain and SH2 domain agree with the hypothesis that the high thermal mobility of SH2 domain would allow the STAT proteins to reorient their DNA-binding domains with respect to the dimerization surface [13].

The residues in the carboxy terminus (residues 708–716 for Stat3, residues 703–710 for Stat1) that form most of the dimer interface (as shown in Fig. S2B) do not show much flexibility in the MD simulations for both Stat1 and Stat3. From Fig. 3, the RMSF for residues 708–716 for Stat3 varies from 1.0 to 1.4 Å indicating that these residues are very rigid. The RMSF for residues 703–710 for Stat1 varies from 1.6 to 2.2 Å that indicates these residues are not as rigid as Stat3.

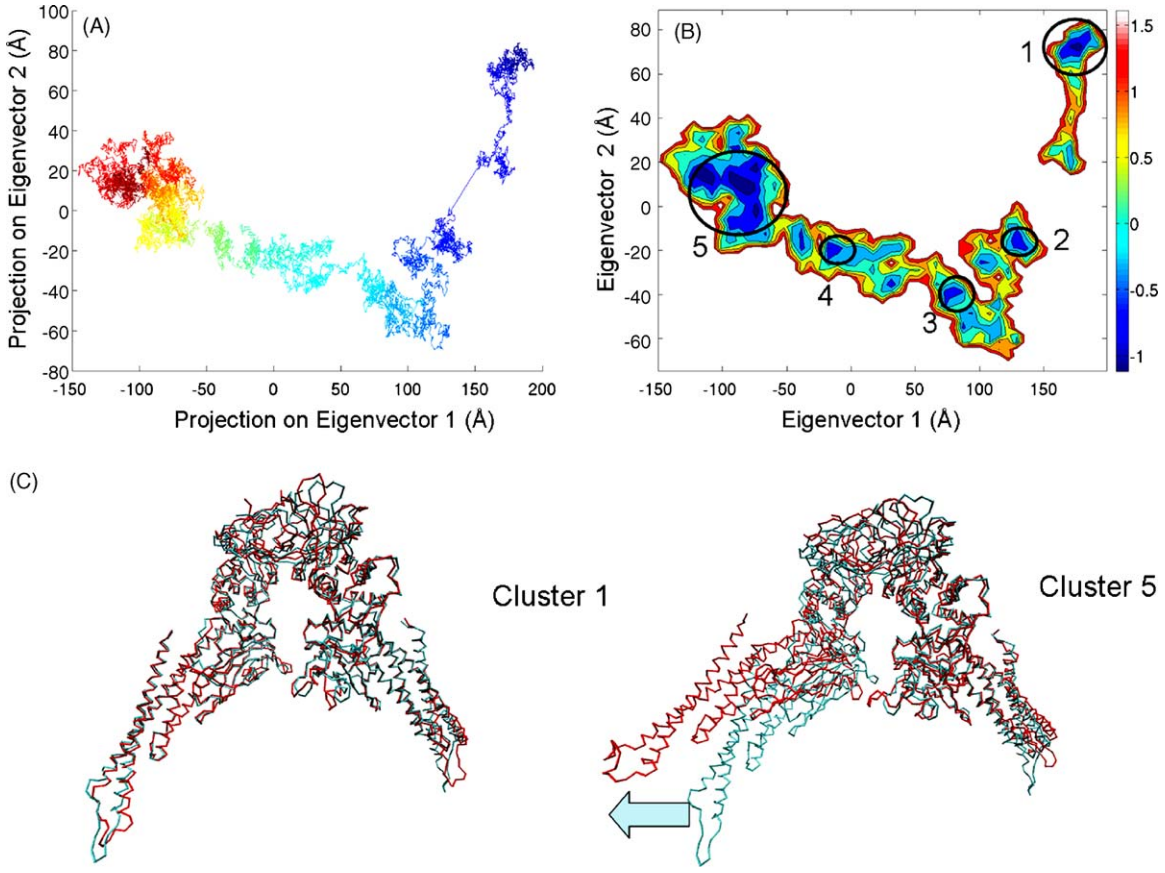

Fig. 5. Projection of MD snapshots every 5 ps of the 50 ns Stat3 trajectory onto the first two principal components. (A) Color gradient from blue at the starting conformer to the darkest red at the 50 ns snapshot. (B) A contour plot of distribution conformer density, ρ, represented by the molecular potential, $-RT\ln(\rho)$ in units of kcal/mol. (C) Averaged structures of clusters 1 and 5. The reference structure is cyan, averaged structure is red.

3.3. Conformational changes in Stat1 and Stat3

We have used principal component analysis and *k*-means cluster analysis to analyze the conformational changes in Stat3 and Stat1 dimers bound to DNA during the MD simulations. Fig. 5A shows the projection of the 50 ns MD snapshots of Stat3 dimer on its two most significant principal components, calculated as described in the Methods section. Fig. 5B shows the contour plot of the distribution of the conformer density represented by the molecular potential (Eq. (2)) generated from Fig. 5A. There are five distinct clusters of conformations seen in Fig. 5B. Fig. 5C shows the averaged structures for cluster 1 and cluster 5 (averaged structures for clusters 2–4 are shown in Fig. S4 of the supporting material). During the MD simulation, the coiled-coil domain of Stat3 undergoes big scissor-like motion shown in the averaged structure of cluster 5 (Fig. 5C). There is also a reorientation of the DNA-binding domain observed for Stat3 as shown in Fig. S5 of supporting material. These results are in agreement with the changes observed in the CRMSD for the DNA-binding domain. Fig. 6 shows distinct conformation clusters for Stat1, and the coiled-coil domain of the Stat1 dimer only undergoes slight scissor-like motion as shown in the averaged conformation of cluster 8 in Fig. 6C. Averaged structures for clusters 2–7 are shown in Fig. S4 of the supporting material. This is also consistent with the CRMSD calculation that there is no big domain motion for Stat1. Also no significant reorientation of DNA-binding domain has been observed for Stat1. This analysis combined with the CRMSD calculations show that the conformational changes observed in Stat3 dimer are different from that of the Stat1 dimer.

3.4. Analysis of domain motion

To understand the nature of the domain motion in Stat1 and Stat3 dimers, we have calculated the correlated fluctuations of the C_α coordinates for various residue pairs in Stat1 and Stat3 dimers. Fig. 7A and B shows the correlation maps for both Stat3 and Stat1, respectively. There are four distinct domains marked as regions 1–4 in Fig. 7A. Residues in regions 1 and 3 show higher positive correlation than residues in region 2. Region 1 is the coiled-coil domain, region 2 is the DNA-binding domain, the connector and the SH2 domain is region 3 and the carboxy terminus is numbered as region 4. For Stat3, regions 5 and 6 both show uncorrelated motion, the coiled-coil domain moving in the opposite direction with respect to the connector and SH2 domains in both the monomers and vice versa. This corresponds to the coiled-coil domain of the Stat3 protein undergoing a conformational change deciphered as the "scissor-like" motion in Fig. 5C. This is also consistent with the fact that the DNA-binding domain is reoriented with respect to the dimerization surface [13]. For Stat1, the coiled-coil domain and DNA-binding domain exhibit less motion with respect to the connector region and the SH2 domain of the same monomer, indicating that there is no big domain motion for Stat1.

3.5. Flexibility of the Stat3 dimer Interface and solvent effect

To study the flexibility of the residues that make up the dimer interface, we examined the correlated motion of the residues in the carboxy terminus of one monomer looping around the other monomer, forming the dimer interface as shown in Fig. 1. Fig. 7C and D shows the correlation map for this region for Stat3 and Stat1 respectively. Region 8 in Fig. 7C and D shows highly correlated motion revealing that the carboxy terminus of one monomer still stays buried and moves in the same direction as the SH2 domain of the other monomer, keeping the dimer interface intact during the dynamics. From region 9 in both Fig. 7C and D, it is seen that the carboxy termini of both the monomers move in the opposite direction

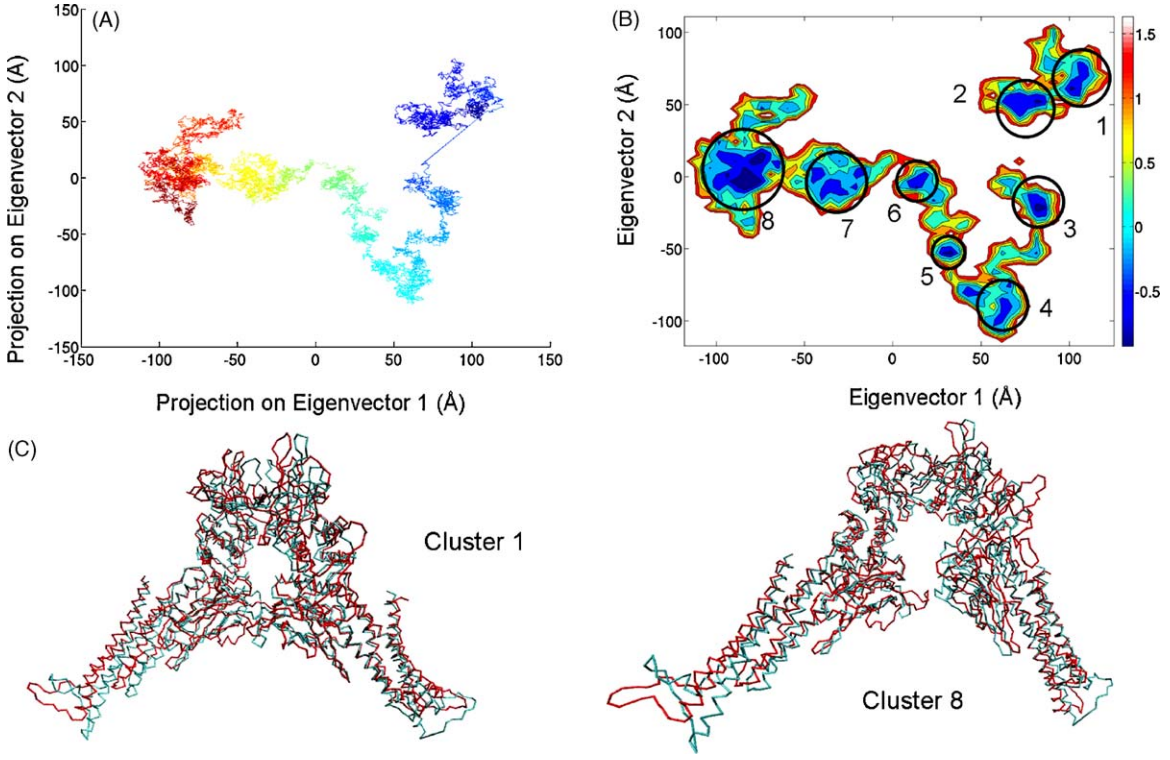

Fig. 6. Projection of MD snapshots at every 5 ps of the 50 ns Stat1 trajectory onto the first two principal components. (A) Color gradient from blue at the starting conformer to the darkest red at the 50 ns snapshot. (B) A contour plot of distribution conformer density, ρ, represented by the molecular potential, $-RT\ln(\rho)$ in units of kcal/mol. (C) Averaged structures of clusters 1 and 8. The reference structure is cyan, averaged structure is red.

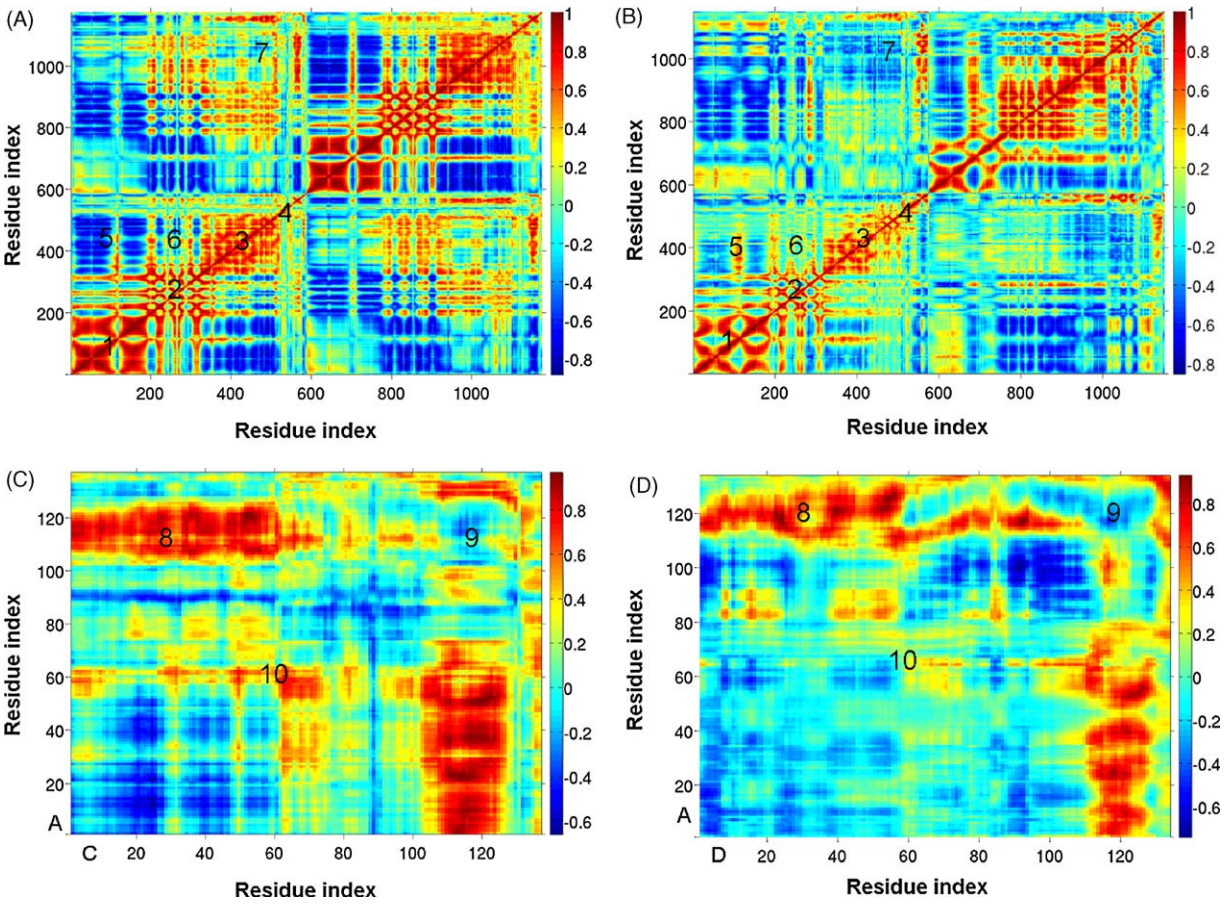

Fig. 7. Cross-correlation coefficient of the C_α atoms. (A) Stat3, (B) Stat1, (C) correlation between SH2 domain and carboxy terminus of one monomer to the other monomer, in Stat3 and (D) correlation between SH2 domain and the carboxy terminus of one monomer to the other monomer, in Stat1. Positive correlations in descending order are shown in red, orange, and yellow. Anti-correlated motions are shown in cyan and blue. Light green shows region with no correlation. In (A), residues 1–587 correspond to residues 136–722 of monomer A, residues 588–1174 correspond to residues 136–722 of monomer C. In (B), residues 1–575 correspond to residues 136–710 of monomer A, residues 576–1150 correspond to residues 136–710 of monomer D. In (C), residues 1–135 correspond to residues 588–722, and in (D), residues 1–137 correspond to residues 574–710. In (C), the *x*-axis is for monomer C, and the *y*-axis is for monomer A. In (D), the *x*-axis is for monomer D, the *y*-axis is for monomer A.

with respect to each other to keep the dimer interface intact. Residues 642–647 of Stat3 (residues 636–641 in Stat1) also form a part of the dimer interface. This region is called the DB loop region in the crystal structure [13], since it connects the β-strands D and B in the SH2 domain. This region is marked as DB in Fig. 1 and in Fig. S2 of supporting material. The DB loops of both the monomers act as a gate to the cavity under the dimer interface, shown in Fig. 8A. The DB loop of Stat3 is quite rigid with RMSF by residue of 0.9–1.2 Å, while the DB loop of Stat1 is not as rigid as Stat3 with a RMSF by residue of 1.6–2.3 Å. Examination of the correlated motion of these residues in region 10 of Fig. 7C shows that the DB loops move in the same direction for Stat3 dimer, thereby tightening the dimer interface after the dynamics. On the contrary these residues move away in the opposite direction in Stat1 dimer, as shown in region 10 of Fig. 7D.

Fig. 9 shows the time evolution of the distance between the geometric center of the DB loops in Stat3 and Stat1 dimers. The two DB loops in Stat3 draw close during simulation and they move apart in Stat1. This distance stabilizes for both Stat3 and Stat1 after 20 ns of MD simulations. The calculated solvent accessible surface area (SASA) decreases for Stat3, whereas it increases in Stat1 as shown in Fig. S6 of the supporting material. After 50 ns of MD, the distance between the two DB loops in Stat3 is about 6 Å closer and about 250 $Å^2$ less solvent accessible compared to Stat1. The SASA values were calculated using VMD. During the MD simulations we observed opening of the cavity under the dimer interface of Stat3 as shown in Fig. 8A–C, allowing water molecules to enter this cavity. The entrance to this cavity is gated by the two DB loops on either monomer. When the DB loops draw closer together as the MD simulations proceed, this water-filled cavity is closed by the DB loops acting as gates as seen in Fig. 8C. Similar to Stat3, there is a water-filled cavity formed under the dimer interface in Stat1 but this cavity is not closed by the DB loops, since they move apart during the dynamics. This cavity under the dimer interface makes it a potential druggable pocket for disrupting the protein–protein interaction in the Stat3 dimer.

3.6. Modulation of dimer and protein-DNA interaction energies upon large conformational changes

To understand the driving force for these conformational changes, we calculated the interaction energies between the SH2 domain including the carboxy terminus (residues from 586 to 722 for Stat3, residues from 577 to 710 for Stat1) of one monomer with the other monomer. Fig. 10A and B shows the variation of the interaction energy with time for Stat3 and Stat1 respectively. The interaction energy between the SH2 domain including the carboxy terminus between the two monomers becomes more favorable by about 300 kcal/mol in Stat3 after the conformational changes, and

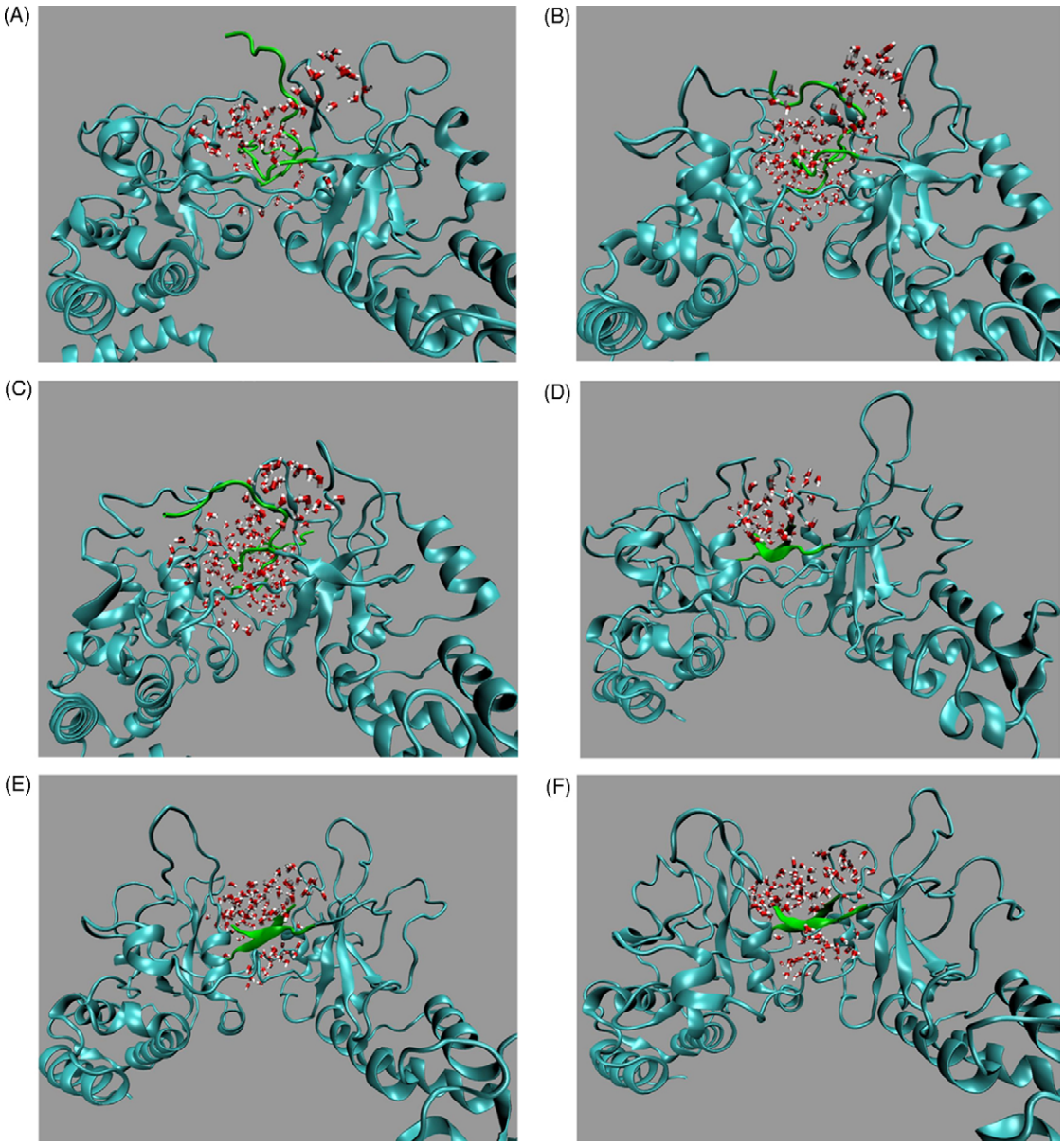

Fig. 8. Water diffusing into the dimer interface. (A) Initial conformation of Stat3; (B) snapshot at 25 ns of Stat3; (C) snapshot at 50 ns of Stat3; (D) initial conformation of Stat1; (E) snapshot at 25 ns of Stat1; (F) snapshot at 50 ns of Stat1.

there is no significant change in the interaction energy for Stat1. We also calculated the interaction energies between DNA and protein (defined by residues within 14 Å around the DNA) using NAMD energy plugin in VMD, for each snapshot of the MD trajectories. Fig. 10C and D show the time progression of the DNA protein interaction energies for both Stat3 and Stat1 respectively. These interaction energies become more favorable in Stat3 by about 100 kcal/mol, whereas there is no substantial change for the interaction energies in Stat1. These calculations in conjunction with the conformational change observed in Stat3 show that the domain motion in Stat3 is driven by enhanced dimer stability and better binding to the DNA. DNA-binding measurements with Stat3β show both increased DNA-binding activity and dimer stability [16] compared to Stat3α, which is in agreement with the increased DNA interaction of Stat3β from our calculations. Furthermore, both Stat1 and Stat3 bind to the same as well as different DNA target sequences in the nuclei of cells and it is not known how the STATs distinguish between these target sequences. Based on our MD simulations, we speculate that the conformational changes that occur in the DNA-binding domain might play a role in differentiating interaction with different target gene sequence.

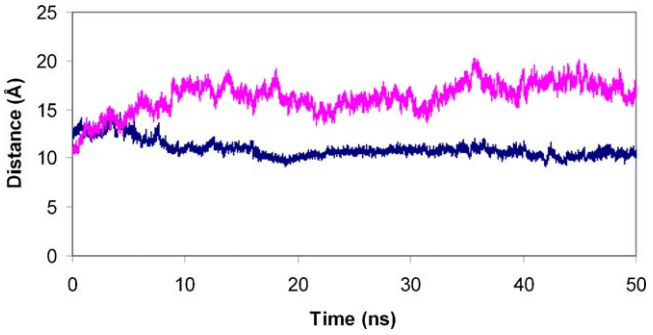

Fig. 9. The distance between the geometric centers of the DB loop of the two monomers (residues 642–647 for Stat3, residues 636–641 for Stat1) versus time. Purple line is for Stat1, dark blue line is for Stat3.

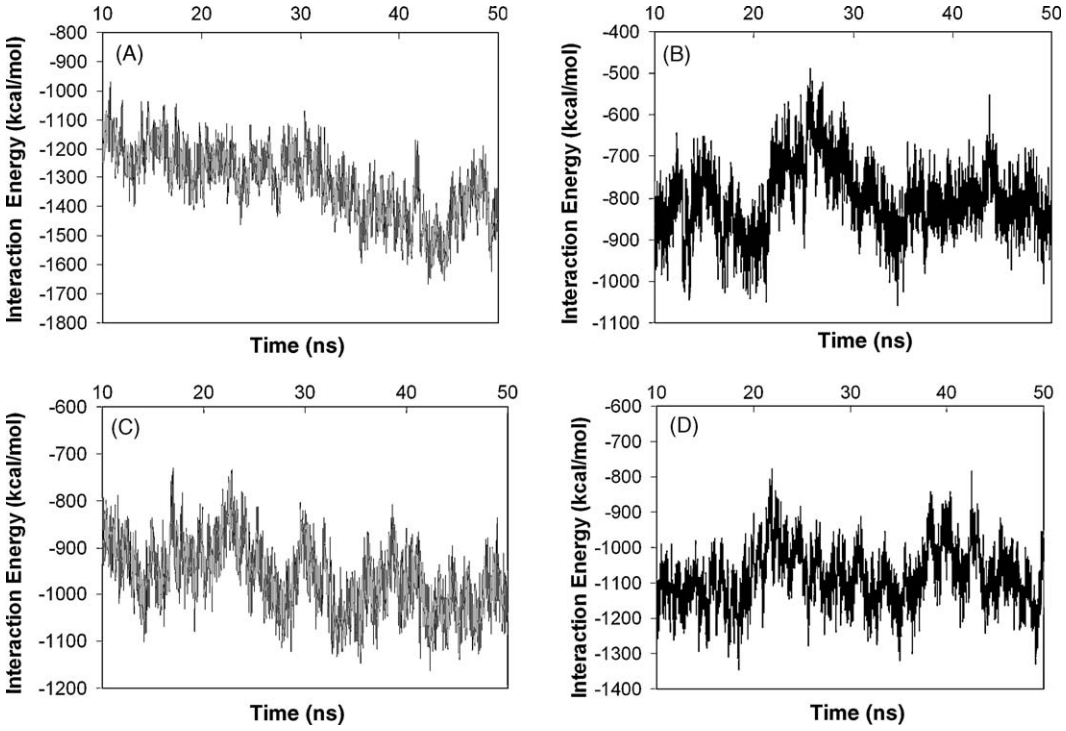

Fig. 10. The interaction energies versus simulation time. (A) Interaction energy between SH2 domain including the carboxy terminus of one monomer to the other monomer for Stat3. (B) Interaction energy between SH2 domain including the carboxy terminus of one monomer to the other monomer for Stat1. (C) Interaction energy between DNA and protein (residues within 14 Å around the DNA) for Stat3. (D) Interaction energy between DNA and protein (residues within 14 Å around the DNA) for Stat1.

4. Conclusions

Using MD simulations in explicit water, we have observed a large conformational change in the Stat3 dimer, while such a large change has not been observed for the Stat1 dimer. The large domain motion is driven by more favorable interaction energies between the DNA and the protein, leading to tightening of the SH2 domains. However, during the dynamics, the domain structure of the individual monomers remain fairly intact. The flexibility of the DNA-binding domain, and the SH2 domain leading to the conformational change is in agreement with the hypothesis that the high thermal mobility of the SH2 domain would allow the STAT proteins to reorient their DNA-binding domains with respect to the dimerization interface. The conformational change leads to better binding of the Stat3 dimer to the DNA than the crystal structure, to regulate transcription. We also observed that the carboxy terminus of one monomer that is not resolved in the crystal structure, wraps around the core of the SH2 domain of the other monomer and this region makes up the dimer interface, which remains intact during the dynamics. Water diffuses into a cavity under this interface thus expanding this cavity that is gated by the DB loops. This cavity is a potential binding pocket to target for Stat3 dimer inhibitor design. Cluster analysis of the MD trajectories showed five distinct conformations for the Stat3 dimer as compared to eight conformations for the Stat1 dimer. These conformations can be further exploited for the on-going structure based dimer inhibitor design for the development of more potent and selective inhibitors. Using the Stat3 receptor conformations from the MD simulations we now have successfully designed and tested inhibitors that show selectivity for Stat3 over Stat1, the results of which will be published elsewhere.

Acknowledgments

We thank Dr. S. Bhattacharya and Mr. A. Mao at the Beckman Research Institute of the City of Hope for help in MD trajectory analysis. We thank Prof. C.W. Müller at the European Molecular Biology Laboratory, France, for sharing the Stat3β dimer structure. We thank Mr. P.L. Freddolino from UIUC for helpful discussion in MD simulations. We also thank Dr. X. Cheng from University of California, San Diego, and Dr. J.M. Bui from the University of Cambridge for helpful discussions on the clustering methods and correlation maps. We thank the San Diego Super computer Center resources for granting us supercomputer time for performing the MD simulations.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.08.013.

References

[1] J.E. Darnell Jr., STATs and gene regulation, Science 277 (1997) 1630–1635.
[2] J.N. Ihle, STATs: signal transducers and activators of transcription, Cell 84 (1996) 331–334.
[3] S. Pellegrini, I. Dusanter-Fourt, The structure, regulation and function of the Janus kinases (JAKs) and the signal transducers and activators of transcription (STATs), Eur. J. Biochem. 248 (1997) 615–633.
[4] J. Bromberg, J.E. Darnell Jr., The role of STATs in transcriptional control and their impact on cellular function, Oncogene 19 (2000) 2468–2473.
[5] T. Decker, P. Kovarik, Transcription factor activity of STAT proteins: structural requirements and regulation by phosphorylation and interacting proteins, Cell Mol. Life Sci. 55 (1999) 1535–1546.
[6] J.E. Darnell Jr., I.M. Kerr, G.R. Stark, Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins, Science 264 (1994) 1415–1421.
[7] J.E. Darnell Jr., Studies of IFN-induced transcriptional activation uncover the Jak-Stat pathway, J. Interferon Cytokine Res. 18 (1998) 549–554.
[8] H. Yu, R. Jove, The STATs of cancer—new molecular targets come of age, Nature Rev. Cancer 4 (2004) 97–105.
[9] R. Buettner, L.B. Mora, R. Jove, Activated STAT signaling in human tumors provides novel molecular targets for therapeutic intervention, Clin. Cancer Res. 8 (2002) 945–954.
[10] T. Bowman, R. Garcia, J. Turkson, R. Jove, STATs in oncogenesis, Oncogene 19 (2000) 2474–2488.

[11] J.F. Bromberg, C.M. Horvath, Z. Wen, R.D. Schreiber, J.E. Darnell Jr., Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 7673–7678.
[12] N.C. Reich, STAT dynamics, Cytokine Growth Factor Rev. 18 (2007) 511–518.
[13] S. Becker, B. Groner, C.W. Müller, Three-dimensional structure of the Stat3beta homodimer bound to DNA, Nature 394 (1998) 145–151.
[14] X. Chen, U. Vinkemeier, Y. Zhao, D. Jeruzalmi, J.E. Darnell Jr., J. Kuriyan, Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA, Cell 93 (1998) 827–839.
[15] X. Mao, Z. Ren, G.N. Parker, H. Sondermann, M.A. Pastorello, W. Wang, J.S. McMurray, B. Demeler, J.E. Darnell Jr., X. Chen, Structural bases of unphosphorylated STAT1 association and receptor binding, Mol. Cell 17 (2005) 761–771.
[16] O.K. Park, L.K. Schaefer, W. Wang, T.S. Schaefer, Dimer stability as a determinant of differential DNA binding activity of Stat3 isoforms, J. Biol. Chem. 275 (2000) 32244–32249.
[17] M.A. Marti-Renom, A. Stuart, A. Fisher, R. Sánchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.
[18] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[19] A. Fisher, R.K.G.A. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.
[20] N. Eswar, M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, Current Protocols in Bioinformatics, Suppl. 15, John Wiley & Sons, Inc., Hoboke, NJ, 2000, pp. 5.6.1–5.6.30.
[21] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[22] J.C. Phillips, R. Braun, W. Wang, J. Gumbat, E. Tajkhorshid, E. Villa, C. Chipot, D. Skeel, R.L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[23] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B. 102 (1998) 3586–3616.
[24] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N–log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[25] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[26] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, P.A. Kollman, AMBER, 8, University of California, San Francisco, 2004.
[27] B.L. de Groot, X. Daura, A.E. Mark, H. Grubmuller, Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds, J. Mol. Biol. 309 (2001) 299–313.
[28] J.M. Bui, Z. Radic, P. Taylor, J.A. McCammon, Conformational transitions in protein–protein association: binding of fasciculin-2 to acetylcholinesterase, Biophys. J. 90 (2006) 3280–3287.
[29] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins Struct. Funct. Genet. 17 (1993) 412–425.
[30] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools And Techniques, second ed., Morgan Kaufmann, San Francisco, 2005.
[31] X. Cheng, I. Ivanov, H. Wang, S.M. Sine, J.A. McCammon, Nanosecond-timescale conformational dynamics of the human alpha7 nicotinic acetylcholine receptor, Biophys. J. 93 (2007) 2622–2634.
[32] D. van der Spoel, E. Lindahl, B. Hees, A.R. Van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feestra, R. Van Drunen, H.J.C Berendsen, Gromacs User Manual Version 3.3, University of Groningen, Groningen, the Netherlands, 2005, www.gromacs.org.
[33] MATLAB, R2008a, 2008, The Mathworks, Natick, Massachusetts, http://www.mathworks.com.