JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

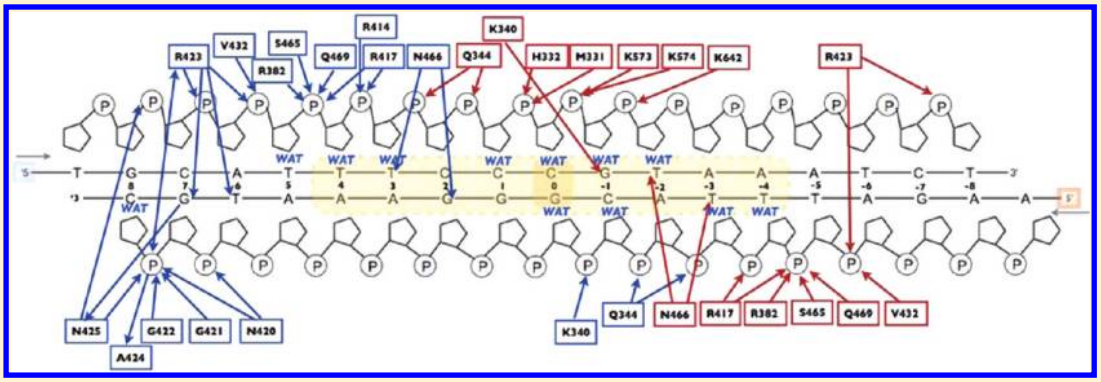
Molecular Dynamics Studies of the STAT3 Homodimer:DNA Complex: Relationships between STAT3 Mutations and Protein−DNA Recognition

Jarmila Husby,[†,‡] Alan K. Todd,[†] Shozeb M. Haider,[†,§] Giovanna Zinzalla,[‡] David E. Thurston,[‡,∥] and Stephen Neidle*[,†]

[†]CRUK Biomolecular Structure Group, UCL School of Pharmacy, University College London, WC1N 1AX London, U.K.

[‡]Department of Pharmacy, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, U.K.

Supporting Information



ABSTRACT: Signal Transducers and Activators of Transcription (STAT) proteins are a group of latent cytoplasmic transcription factors involved in cytokine signaling. STAT3 is a member of the STAT family and is expressed at elevated levels in a large number of diverse human cancers and is now a validated target for anticancer drug discovery.. Understanding the dynamics of the STAT3 dimer interface, accounting for both protein−DNA and protein−protein interactions, with respect to the dynamics of the latent unphosphorylated STAT3 monomer, is important for designing potential small-molecule inhibitors of the activated dimer. Molecular dynamics (MD) simulations have been used to study the activated STAT3 homodimer:DNA complex and the latent unphosphorylated STAT3 monomer in an explicit water environment. Analysis of the data obtained from MD simulations over a 50 ns time frame has suggested how the transcription factor interacts with DNA, the nature of the conformational changes, and ways in which function may be affected. Examination of the dimer interface, focusing on the protein−DNA interactions, including involvement of water molecules, has revealed the key residues contributing to the recognition events involved in STAT3 protein−DNA interactions. This has shown that the majority of mutations in the DNA-binding domain are found at the protein−DNA interface. These mutations have been mapped in detail and related to specific protein−DNA contacts. Their structural stability is described, together with an analysis of the model as a starting-point for the discovery of novel small-molecule STAT3 inhibitors.

INTRODUCTION

The Signal Transducer and Activator of Transcription protein STAT3, a member of the STAT family of cytosolic proteins, is a well-recognized and important mediator of tumor-induced immunosuppression at many levels. STAT3, as an oncogenic transcription factor, directs malignant progression via the aberration of key proteins, such as the cell survival (antiapoptotic) proteins Bcl-x_L and Mcl-1, the cell-cycle regulators cyclin D1/D2 and c-myc, and inducers of angiogenesis, such as vascular endothelial growth factor VEGF.[1,2] As a result, cancer cells with aberrantly active STAT3 show greater resistance to the initiation of apoptotic processes from the environment and apoptosis-initiating chemotherapeutics.[3]

Multiple distinct steps are involved within the STAT3 signaling pathway; upon extracellular receptor stimulation by growth factors or cytokines (e.g., IL-6), the receptor (gp130) becomes dimerized, and subsequently activated, which triggers the activation of the Janus protein tyrosine kinases JAK1 and JAK2, that are associated with the cytoplasmic tail of the receptor. This phosphorylation of the cytoplasmic tail provides a docking site for the recruitment of monomeric[1] or dimeric[4] unactivated STAT3 proteins via reciprocal interaction of their SH2 domains. Activated JAK tyrosine kinases phosphorylate recruited STAT3 proteins at a specific tyrosine within the C-

Received: December 27, 2011

   dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

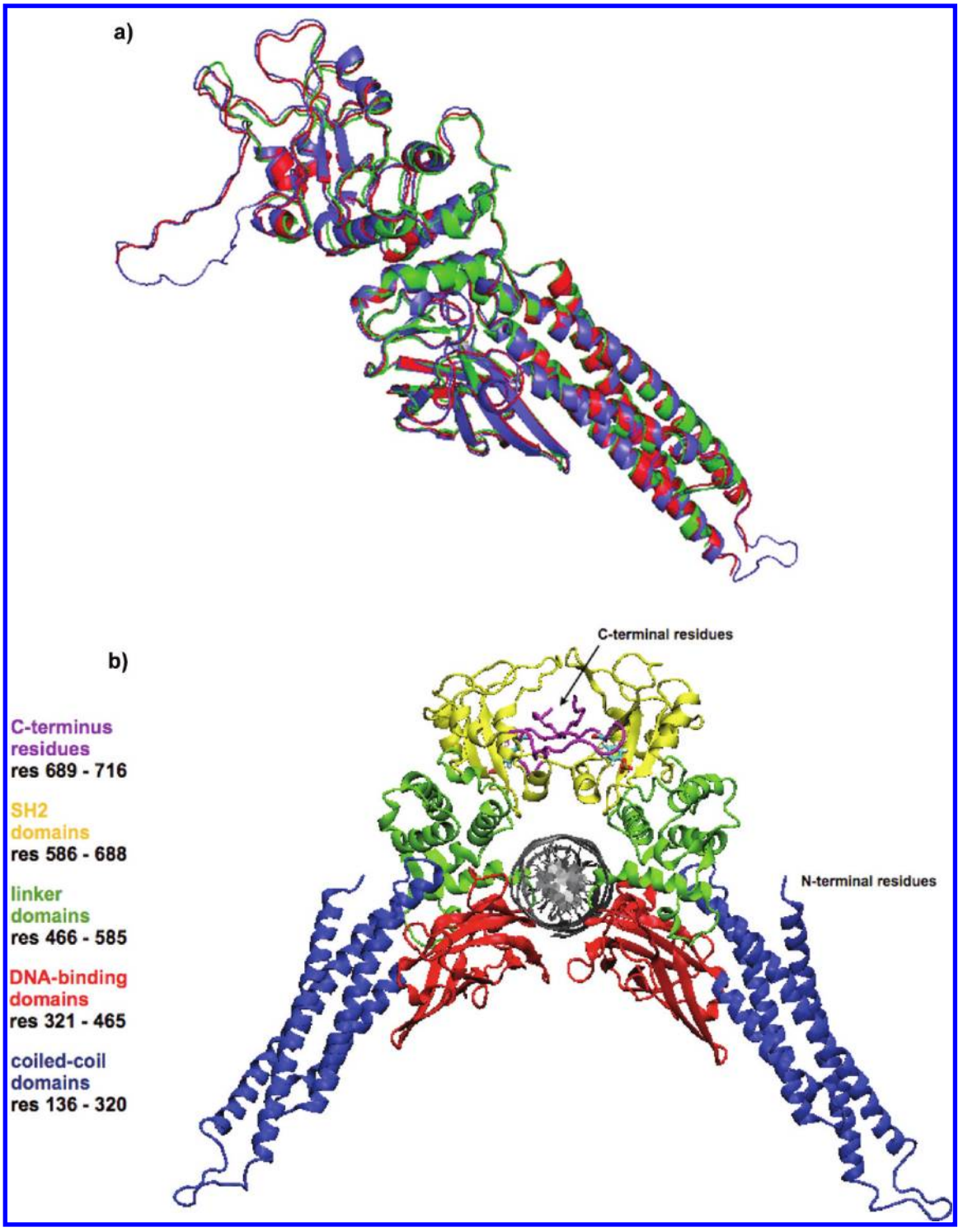


Figure 1. (a) Structural alignment of the STAT3β monomers in cartoon representation: unphosphorylated STAT3 core-fragment (PDB id 3CWG) *(green)*; phosphorylated STAT3β monomer used for model building (PDB id 1BG1) *(red)*; and phosphorylated STAT3β monomer of the model *(blue)*. (b) Model of the STAT3β-DNA complex in cartoon representation; Individual domains are color coded; Coiled-coil domains *(blue)*, DNA-binding domains *(red)*, linker domains *(green)*, SH2-domains *(yellow)*, and a stretch of ordered residues at the C-terminus *(magenta)*, with the pY705 highlighted in stick representation *(cyan)*. The 17-bp DNA with 5′ overhanging ends is located between the two monomers *(gray)*.

terminus (Tyr705), and the phosphorylated STAT3 monomers dissociate from the receptor to form STAT3-STAT3 homodimers that translocate from the cytoplasm to the nucleus, where they bind to target DNA motifs,[2] promoting the expression of proteins crucial for cell growth and survival. Normal STAT3 activation is relatively brief and regulated by a number of deactivation mechanisms but human cancer cells express constitutively active STAT3 at high concentrations.[1,5] There is a mounting body of evidence[6] suggesting that tumor formation can be caused by abnormally active STAT3. Thus inhibition of aberrant STAT3 activity induces growth arrest and apoptosis of tumor cells *in vitro* as well as *in vivo*, validating

dx.doi.org/10.1021/ci200625q | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

STAT3 as a suitable molecular target for anticancer drug discovery.[2,6,7] The transcriptionally active STAT3-STAT3 homodimer has been extensively targeted (both directly and indirectly) by many research groups with the ultimate goal of suppressing the aberrant STAT3 function in human cancer cells.[3,6] *In silico* screening techniques and structure-based drug design have been employed in a number of studies aimed at the discovery of small-molecule inhibitors of the STAT3-STAT3 dimerization. However, to date, only five categories of inhibitors have been reported.[3,8−12] This may be due to the inherent challenges of targeting protein−protein interactions.[13] There are currently two crystal structures of STAT3 deposited in the Protein Data Bank (PDB). Both of these structures, of the phosphorylated STAT3β-DNA complex (1BG1)[14] and the unphosphorylated STAT3 core fragment (3CWG),[15] have regions of tertiary structure missing in the STAT3 protein−protein interaction domains, which makes virtual screening even more challenging. Furthermore, X-ray crystal structures are spatially and temporally averaged, thus providing only limited information on dynamic properties.

Recent molecular dynamics studies[16] of the STAT3 complex and the STAT3 SH2 domain with docked ligands[17] have focused on protein−protein interactions. However, understanding the protein−DNA recognition process is of importance for a complete insight into the dynamic nature of the STAT3 homodimer:DNA complex which is in turn critical for structure-based design of STAT3-STAT3 inhibitors. Water molecules often play a role in molecular recognition and association[18,19] and especially in protein stability,[20,21] protein−protein interactions, protein−ligand recognition,[18] and protein−DNA recognition.[22,23] Protein−DNA contacts at the atomistic level can be explained in terms of hydrogen bonds (direct), water-mediated hydrogen bonds, van der Waals, electrostatic, and hydrophobic contacts.[23] Electrostatic charge interactions are considered to be a major determinant of protein−DNA interactions, contributing to indirect readout, depending on the distance between the charged groups. However, (weaker) hydrogen bonding contributes both to direct read-out (via contacts with base-pairs) as well as to indirect nonselective read-out (via phosphates and deoxyribose interactions). Therefore, if mutations occur at the protein−DNA interface, these interactions and their interaction networks become affected and altered. Dominant negative mutations in STAT3, mostly in the DNA-binding and SH2 domains, are associated with the Hyper-IgE Syndrome (HIES),[24] and these mutations have been mapped in STAT3 obtained from cells of clinical material.[24−27] We show here that the majority of mutations in the DNA-binding domain are located at the protein−DNA interface, and we have been able to map out these mutations in detail and relate them to specific protein−DNA contacts. The present study examines the dynamic features of (1) the Tyr705-phosphorylated STAT3β homodimer:DNA complex, (2) the unphosphorylated STAT3β monomer, and (3) the 17 base-pair DNA with 5′ overhanging ends, comprising the STAT3 binding site. A major focus is on the conformational changes at the protein−DNA interface with respect to STAT3-DNA complex formation and to the X-ray structural data of the STAT3-DNA complex (PDB code 1BG1 at 2.25 Å resolution).[14] These results will directly aid *in silico* approaches to the discovery of novel STAT3-STAT3 inhibitors for chemotherapeutic intervention.

MATERIALS AND METHODS

The three main stages of this study were (1) construction and preparation of the macromolecular systems, (2) calculation of classical MD full-atomic trajectories, and (3) data analysis.

Model Building. The STAT3 crystal structure (PDB id 1BG1 at 2.25 Å resolution) is deposited in the PDB as a monomer. The missing residues in the structure (185−193, 689−701, and 717−722) were identified as loops by the secondary structure prediction software JPRED (www.compbio.dundee.ac.uk/). These were modeled by the ModLoop program (www.modbase.compbio.ucsf.edu/modloop/). The loops were joined to the main structure and subjected to energy minimization procedures using GROMACS v3.3.3 (www.gromacs.org/) while keeping the position of the atoms of the core crystal structure fixed. The complete monomeric subunit was then used to construct the STAT3-DNA dimeric complex. The bases in the cDNA chain in the structure were modified using the Biopolymer software in the Insight II suite of programs (www.accelrys.com/). The full model of the STAT3-DNA complex was then subjected to a short cycle (1,000 steps) of molecular mechanics energy minimization to relieve any steric clashes from the structure.

System Setup and Molecular Dynamics Simulation. Each monomer of the generated STAT3 dimer comprised residues 136 to 716, while the DNA duplex, in a B-form conformation, comprised a 17-mer duplex with a nine base pair (9-bp) high-affinity binding site (M67) and overhanging 5′ ends. This high-affinity M67 site d(TTCCCGTAA) differs at the sixth base from the consensus 9-bp target site d(TTCCCGGAA) previously described.[28] A model of the unphosphorylated STAT3β monomer (U-STAT3) was generated by removing the coordinates of the phosphate group of the phospho-Tyr705 in the STAT3 dimer and saving the monomer as a new PDB file. The quality of the model was visually assessed by comparison and structural alignment with the crystal structure of the STAT3β-DNA complex (PDB code 1BG1) and also with the crystal structure of the unphosphorylated STAT3 core fragment (PDB code 3CWG). The calculated root mean-square deviation (rmsd) value of the corresponding protein residues of the unphosphorylated and phosphorylated STAT3 monomers was ∼0.87 Å upon structural alignment using the PyMol program (www.pymol.org/) with the structures visually well matching (Figure 1). All of the MD simulations were full-atom simulations and were performed with the GROMACS v 4.5.3 program (www.gromacs.org/),[29] employing the improved protein side-chain torsion potentials from the AMBER parm99sb-ILDN forcefield,[30] together with the parmbsc0 force field,[31] a refinement of the AMBER parm99 force field for nucleic acids. Further AMBER parameters for the phosphorylated-Tyr residues[32] were manually ported to GROMACS for the simulations.

The simulation protocols were consistent for all three systems, phosphorylated STAT3β homodimer:DNA complex (1), the unphosphorylated STAT3β monomer (2), and the 17-bp DNA STAT3 binding site (3). The simulation systems were constructed by immersing the macromolecules in a dodecahedron of explicit TIP3P water molecules, with a minimal clearance of 20 Å between periodic images for the starting configurations. The negative net charge of the STAT3-DNA complex and the DNA duplex was neutralized by addition of excess positively charged Na^+ counterions, with negatively charged Cl^- counterions, providing a final NaCl concentration

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

of ≈150 mM, which approximates physiological concentration. The counterions were automatically placed throughout the box by the *genion* utility of GROMACS 4.5.3 replacing solvent molecules, using the option to first position those atoms with the most favorable electrostatic potential. The total number of atoms in the resulting solvated systems were 559,839 (STAT3-DNA complex), 371,422 (U-STAT3 monomer), and 42,088 (17-bp DNA) respectively. Each of the systems was then allowed to adapt to the aqueous ionic environment (to remove any solvent−solvent and solvent−solute clashes created during the construction process) by applying 5,000 cycles of potential energy minimization, combining both steepest descent and conjugate gradient methods. This involved gradual relaxation of the initially used harmonic restraints on the macromolecular atoms. With the greatest strain dissipated from the systems, the solvent was allowed to adapt to the macromolecules by being able to move freely, while keeping the non-hydrogen atoms of the macromolecule harmonically restrained to their reference positions over the 150 ps period of molecular dynamics at 200 K. Subsequently the unconstrained systems were slowly heated to 300 K and equilibrated over 50 ps.

To compute MD trajectories, unrestrained production-level MD simulations were performed for 50 ns in the isobaric−isothermal ensemble (conserved NPT). Pressure and temperature were sustained at 1.0 bar and 300 K, with temperature modulated by a velocity rescaling thermostat with a stochastic term[33] and pressure was controlled via the Parrinello-Rahman barostat algorithm.[34,35] Nonbonded van der Waals interactions were calculated using the Lennard-Jones 12−6 potentials with a 9 Å cutoff. Long-range electrostatic effects were calculated using the Particle-Mesh-Ewald algorithm (PME),[36] with a cutoff for the real-space term of 9 Å. The corresponding cutoff values of 9 Å for both nonbonded and long-range electrostatic interactions were chosen to fulfill the force-field and PME algorithm requirements of a minimum cutoff value (<8 Å) and to obtain better performance with a minimum sacrifice in integration accuracy. The LINCS algorithm[37] was employed to constrain all bonds. The integration time step applied was 2.0 fs with the coordinates saved every 5.0 ps. All MD simulations were computed on in-house Linux 64-bit Intel Core-i7 workstations, with efficient parallel scaling and double-precision calculations to prevent any energy conservation and stability issues. Trajectories were analyzed with the programs in the GROMACS 4.5.3 suite package and visualized by means of the VMD program.[38] All graphs were plotted using the Xmgrace program (http://plasma-gate.weizmann.ac.il/Grace/).

Principal Component Analysis. Principal component analysis (PCA, i.e. Essential Dynamics) enables large-scale correlated motions of atoms in the molecule can be identified and extracted from MD trajectories, hence recurrent modes of structural changes can be eliminated from sets of structures, revealing the dominant modes, and structures in the motion.[39] The backbone atom coordinates of the protein part and backbone atoms of the DNA helix of (1) the complete STAT3-DNA complex, (2) monomer-A, (3) monomer-B, and (4) the U-STAT3 monomer were analyzed by PCA throughout the last 40 ns of the simulation time. Only residues 321 to 688 in the protein part of the complex were considered here in detail since the protein−DNA region is of main interest in this study and is stable throughout the MD simulation. The GROMACS program g_covar was used to calculate and diagonalize the mass-weighted covariance matrix. The generated eigenvectors (which provide a vectorial description of each component of the motion with their corresponding eigenvalues representing the energetic contribution of each component to the overall motion) were analyzed using the program g_anaeig in the GROMACS program suite. Protein and DNA backbone atoms only were used to construct the covariance matrix, since the size of the matrix varies with the square of the number of atoms for which the covariance is calculated and Cα atoms alone may adequately sample large-scale correlated motions in proteins.[40] Porcupine plots[41] were used to show the direction and magnitude of the extremes of fluctuations which were accounted for in the first eigenvector. The porcupine plots were generated using a porcupine script[42] and visualized using the VMD program.[38]

Cluster Analysis. By definition, cluster analysis is designed to detect hidden clusters in a set of objects which are described by numerical or structural data so that the members of each cluster behave similarly to each other and groups are well separated.[43] To identify clusters of structures in a trajectory, the RMSD was used to assign distances between cluster sets with respect to the distances between structures, reflecting the range of conformations and their relative populations. The gromos agglomerative clustering algorithm[44] was implemented via the GROMACS clustering utility (g_cluster). This was employed to extract the clusters of conformers in the STAT3β-DNA complex generated over the simulation time frame, with solely the duplex DNA, DNA binding domains, linker and SH2 domains being considered. The rmsd cutoff distance was 2.0 Å for two structures to be considered as neighbors.

Protein−DNA and Water Contact-Residues Analysis. Structures representing the conformations of the STAT3β-DNA complex generated over the course of the simulations obtained from the cluster analysis were subjected to contact analysis, focusing on the protein−DNA interface and water contacts. Interfacial interactions were defined on the basis of physicochemical and distance criteria between atoms. We used a 3.2 Å donor−acceptor distance for hydrogen bonds, and 5.0 Å for electrostatic interactions between the charged side-chains of residues and hydrophobic interactions. From the MD simulation generated trajectories, the interatomic distances between the hydrogen bond-forming residues were calculated using the GROMACS program g_mindist, and the overall time spent within the specified distance for each hydrogen bond was determined in terms of percentage of existence of the individual hydrogen bonds over the course of the 50 ns simulation. Also all the solvent molecules were analyzed to determine their interactions with the protein−DNA complex in terms of time spent within the hydrogen bonds formed with the protein, DNA, or both (bridging waters). Water molecules were only accepted if they formed hydrogen bonds for longer than 1 ns. Dedicated tools in the Chimera[45] and VMD[38] programs were used for hydrogen-bond detection and analysis.

Water Density Maps. Water density maps were generated from snapshots of the protein−DNA complex in explicit water that were generated throughout the simulation. For each area of interest, three consecutive residues on the same chain were chosen and used as a reference set to align all of the snapshots, using the program LSQMAN (http://xray.bmc.uu.se/usf/dejavu.html). The water coordinates were extracted from each of the aligned files and placed in a single file. These water files were then converted into density maps using the CCP4 programs SFALL and FFT (http://www.ccp4.ac.uk/). Maps were visualised with the PyMol program.

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

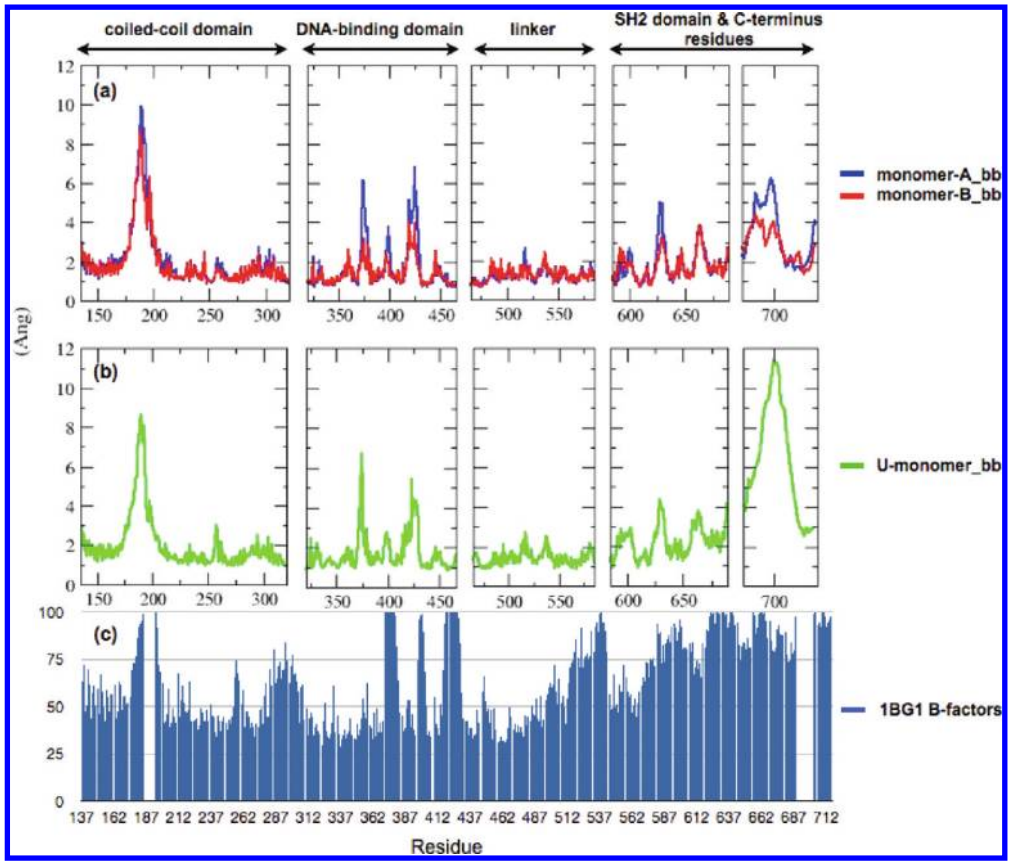


Figure 2. rms fluctuations averaged for each amino-acid residue for (a) STAT3β-DNA complex-forming monomer-A *(blue)* and monomer-B *(red)* and for (b) unphosphorylated STAT3β monomer *(green)* during the dynamics runs. (c) The rms fluctuations of the STAT3β models may be compared with the experimental B-factors for the STAT3β crystal structure (PDB id 1BG1), which was used to build the initial model.

RESULTS AND DISCUSSION

This study has focussed on 1. an analysis of the hydration at the protein:DNA interface and 2. detailed mapping of the mutations in the DNA-binding region and their location to specific contacts. The unphosphorylated unbound (U-STAT3) vs the phosphorylated complex-bound STAT3 monomers are also compared (Figure 1a), together with the effects of STAT3 dimerization followed by STAT3-DNA complex formation with the DNA duplex target sequence. These findings are then related to the X-ray crystal structure (PDB code 1BG1) that was used for the model building. In general the simulation results confirm earlier observations from the crystal structure but are able to extend them to provide a dynamic picture of this protein:DNA system.

A model of the STAT3β-DNA complex is shown in Figure 1b. Each of the two monomers forming the antiparallel dimer is composed of four domains: (1) a coiled-coil domain formed by an N-terminal four-helix bundle (residues 136 to 320), (2) a DNA-binding domain comprising an eight-stranded β-barrel (residues 321 to 465), (3) an α-helical linker domain (residues 466 to 585), and (4) a SH2 (Src homology domain (residues 586 to 688) with a stretch of ordered residues at the C-terminus (residues 689 to 716) containing the important phospho-Tyr705.[14]

Structural Stability and Conformational Variability. The overall 150 ns simulation for the STAT3 models resulted in energetically conserved and stable simulations for the ~560,000-atom STAT3-DNA complex, ~ 372,000-atom U-STAT3 monomer, and ~42,000-atom 17-bp DNA target sequence systems (Table S1 and Figure S1). RMSD values for the backbone atoms as a function of the simulation time were used as a measure of stabilization of the three models during the simulation, comparing both initial reference, and time-averaged structures of the models, with the latter providing superior insight into the structures reaching the plateau (Figure S2). The conformational stability of the higher-order structures was also examined over the course of the simulation.

The trajectory of the STAT3 complex was stabilized at around ~3.9 Å (and at around ~2.0 Å with respect to the time-averaged structure obtained from the MD), with the individual corresponding monomers-A and -B being stabilized with ~0.5 Å difference in their respective RMSDs. The complex-bound monomers show improved stability comparing to the unphosphorylated STAT3 monomer. The slightly larger fluctuations in the model suggested a stabilizing effect of the STAT3β-DNA complex formation towards both the STAT3 monomers and the DNA duplex (Table S1). The flexible regions in the protein part of the STAT3 complex, and the U-STAT3 monomer were also analyzed by examining their structural fluctuations in terms of root mean square fluctuations (RMSF) as a function of residue number (Figure 2). Large fluctuations indicated by sharp peaks correspond to the loops. The magnitude of fluctuations is most significant at both terminal regions of the models and in the DNA-binding domain, indicating significant interaction with the duplex DNA, in accord with the crystal structure. Differences in the fluctuations within individual domains were observed for

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

Table 1. Amino Acid Residues of the STAT3 Complex-Bound Monomers-A and -B Determined To Be within the 5 Å Cut-off Protein−DNA Interface of the First 5 Middle Structures Obtained from the Cluster Analysis[a]

monomer-A					monomer-B			
HB -solvent	Occu-pancy	HB-DNA	5Å interface		5Å interface	HB-DNA	Occu-pancy	HB-solvent
			--	M 329	✓			
			--	P 330	✓			
WAT			✓	M 331	✓	dC(0)	95%	
WAT			✓	H 332	✓	dC(0)	98%	
	60%	dG(0)	✓	K 340	✓	dG(-1)_O6	24%	3x WAT-bridge
WAT-bridge			✓	T 341	✓			2x WAT-bridge
			✓	G 342	✓			
			✓	V 343	✓			WAT-bridge
2x WAT-bridge	22%	dA(-2)	✓	Q 344	✓	dC(+2)	53%	
	73%	dC(-1)				dC(+1)	33%	
WAT-bridge, WAT	98%	dT(+4)	✓	R 382	✓	dT(-4)	97%	
			--	H 410	✓			
	2%	dT(+3)	✓	R 414	✓			
4x WAT-bridge, 5x WAT			✓	E 415	✓			5x WAT-bridge, 3x WAT
	69%	dT(+4)	✓	R 417	✓	dT(-4)	57%	
	83%	dT(+3)				dT(-3)	91%	
			✓	G 419	--			
	4%	dG(+7)	✓	N 420	--			
	7%	dC3'(+8)						
	8%	dC3'(+8)	✓	G 421	--			
	11%	dC3'(+8)	✓	G 422	--			
WAT	8%	dA(+6)	✓	R 423	✓	dT3'(-8)	5%	
	10%	dT(+5)				dT(-5)	42%	
	27%	dC3'(+8)						
	10%	dG(+7)_N3						
	35%	dT(+6)_O2						
	6%	dC3'(+8)	✓	A 424	--			
	7%	dG(+7)_N2	✓	N 425	--			
	4%	dC(+7)						
	3%	dC3'(+8)						
			✓	C 426	--			
			✓	L 430	✓			
			✓	I 431	✓			
	90%	dT(+5)	✓	V 432	✓	dT(-5)	77%	
WAT			✓	T 433	✓			3x WAT
			✓	I 464	✓			
	100%	dT(+4)	✓	S 465	✓	dT(-4)	100%	
	93%	dT(+3)_O4	✓	N 466	✓	dT(-2)_O4	95%	
	65%	dG(+2)_N7				dT(-3)_O4	90%	
	44%	dG(+2)_O6						
			✓	I 467	✓			WAT-bridge
			✓	C 468	✓			
WAT-bridge	92%	dT(+4)	✓	Q 469	✓	dT(-4)	95%	
			--	N 553	✓			
			✓	D 570	✓			2x WAT
			--	K 573	✓	dG(-1)	2%	
			✓	K 574	✓	dG(-1)	84%	
			--	E 616	✓			3x WAT
			--	K 642	✓	dT(-2)	78%	
			✓	Q 643	✓			

[a]There are 39 residues in the interface, 25 of which are common to the two monomers. Residues forming hydrogen bonds with the DNA are highlighted in grey, and unless stated otherwise, the hydrogen bonds involve the DNA backbone atoms. Hydrogen bonds comprising the atoms of the DNA bases are listed. Point mutations at the protein−DNA interface are shown in dark grey, and residues forming hydrogen bonds with the solvent molecules are marked (WAT), with water molecules bridging the protein−DNA interaction indicated as (WAT-bridge).

monomers A and B, so that despite the identity of their primary sequence, there is a real difference in their dynamic behavior. The complementary pattern of fluctuations can be observed for the U-STAT3 monomer with the largest difference at the SH2 domain and in particular in the elongated stretch of C-terminus residues. Whereas in the complex-bound form the "elongated arm-like" loop-forming residues 689-716 interact with the other STAT3 monomer (therefore the magnitude of fluctuations is partially limited), in the case of the U-STAT3 monomer the terminal loop was observed to be free to move and fold onto the SH2 domain which may explain the very large RMSF values for the U-STAT3 monomer C-terminal region. The pattern of experimental B-factor values in the crystal structure 1BG1, plotted as a function of residue number (Figure 2c), show certain similarities to the fluctuations observed in our STAT3 simulations. However, due to the extremely large values of the

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

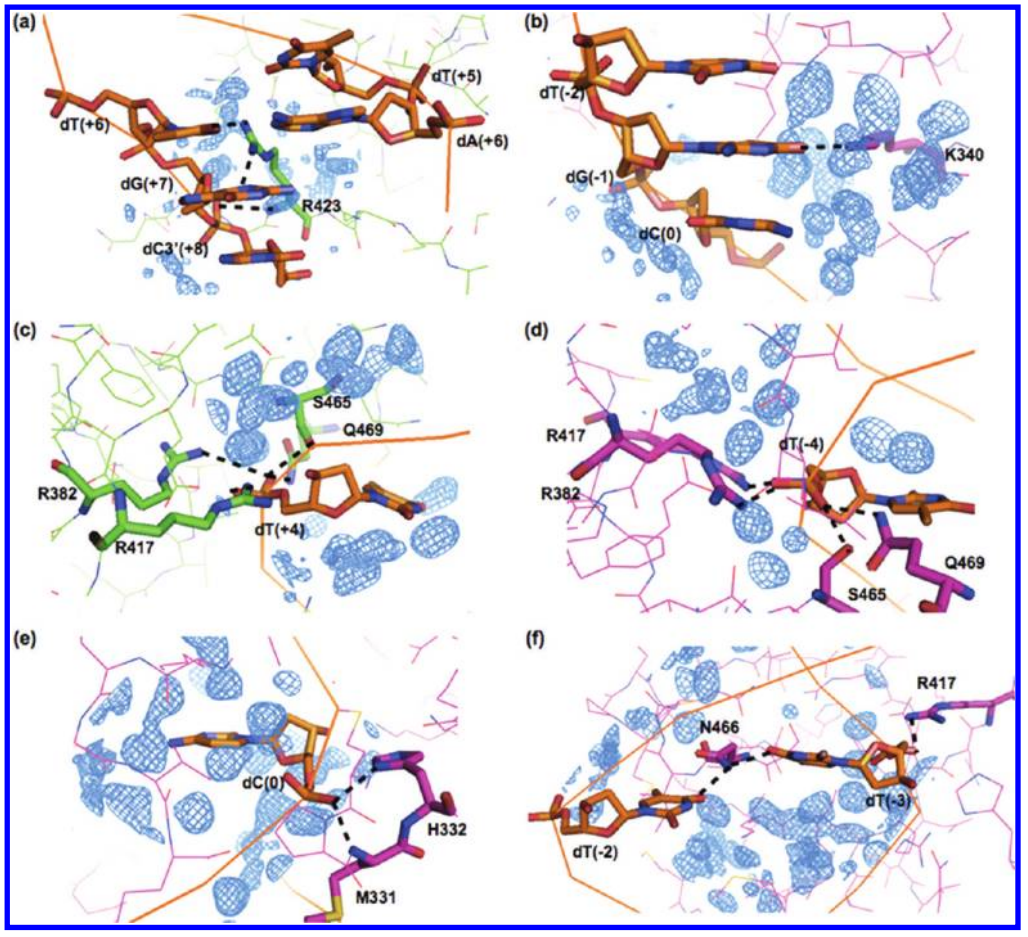


Figure 3. Water density maps showing the hydration at the STAT3 protein−DNA interface. Protein residues of monomer A are displayed in *green*, monomer B in *magenta,* and the DNA bases in *orange*. Regions with high water density, as observed throughout the MD run, are shown as *blue* "blobs" in mesh representation, where the size of these blobs corresponds to sizes of densely hydrated regions.

B factors for atoms in the SH2 domain and the uncertainties surrounding the stretch of C-terminal residues in particular (with a number of residues missing), the STAT3 conformations generated here during the MD run can provide greater insight into the arrangement of the protein−protein interaction than the crystal structure. A comparison has also not been made between the crystallographic water molecules and those in the MD simulations in view of the uncertainties surrounding the former, which also mostly have high B factors.

RMS fluctuation plots on a per residue basis were generated for the DNA-binding domain, linker, and SH2 domain (residues 321 to 688) of both complex-bound monomers A and B, as well as for the U-STAT3 monomer, in order to obtain more insight into the stability of this key region (Figure 2). These domains of monomer A and B were observed in the simulations to have comparable stabilities. Thus the mean RMSF per residue values are as follows: DNAbin domain of monoA: ∼ 2.8 Å, monoB: ∼ 3.3 Å, Umono: ∼ 3.0 Å; linker of monoA: ∼ 2.1 Å, monoB: ∼ 2.6 Å, Umono: ∼ 2.4 Å; SH2 domain of monoA: ∼ 3.5 Å, monoB: ∼ 4.3 Å, Umono: ∼ 4.1 Å for the U-STAT3 monomer domains.

Backbone-atom RMSD plots of the 17-bp duplex DNA (Figure S2d,e) unsurprisingly show that STAT3-DNA complex formation has an overall stabilizing effect on the DNA, in agreement with the RMSF plots (data not shown). Only the T5′ region shows amplified fluctuation in the complex-bound form. This observation can be explained by a strong interaction of the residues in monomer-A with the C(8) and G(7) nucleotides of the opposite DNA chain, hence unwinding the duplex DNA at that terminal. The protein−DNA interactions will be discussed in more detail in the following section.

Mapping the Protein−DNA−Solvent Interaction. The middle structures of the first 5 clusters obtained from the cluster analysis were used to describe the protein−DNA interactions, as they represent over 98% of the conformational space covered throughout the 50 ns MD run. Within a 5.0 Å distance cut-off, and a condition of at least one non-hydrogen atom present in that distance range, 25 corresponding residues were identified at the protein−DNA interface within each of the subunits (monomer-A and -B). A further 7 non-corresponding residues of monomer-A (G419, N420, G421, G422, A424, N425, and C426) and 7 noncorresponding residues of monomer-B (M329, P330, H410, N553, K573, E616, and K642) were observed at the interface (Table 1). Although the monomers are identical in terms of the reciprocal phospho-Tyr705 interaction at the SH2 domains and the high-affinity binding predominantly into the DNA major groove, the partial asymmetry of the monomeric protein−DNA interface and the specific dynamics of each monomer may be explicable by the asymmetric shape of the DNA duplex. A range of electrostatic, hydrophobic, and hydrogen bonds-forming contacts between the STAT3 protein and the DNA duplex occur predominantly in the DNA-binding domains, although they also occur in the linker domains of both phosphorylated

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

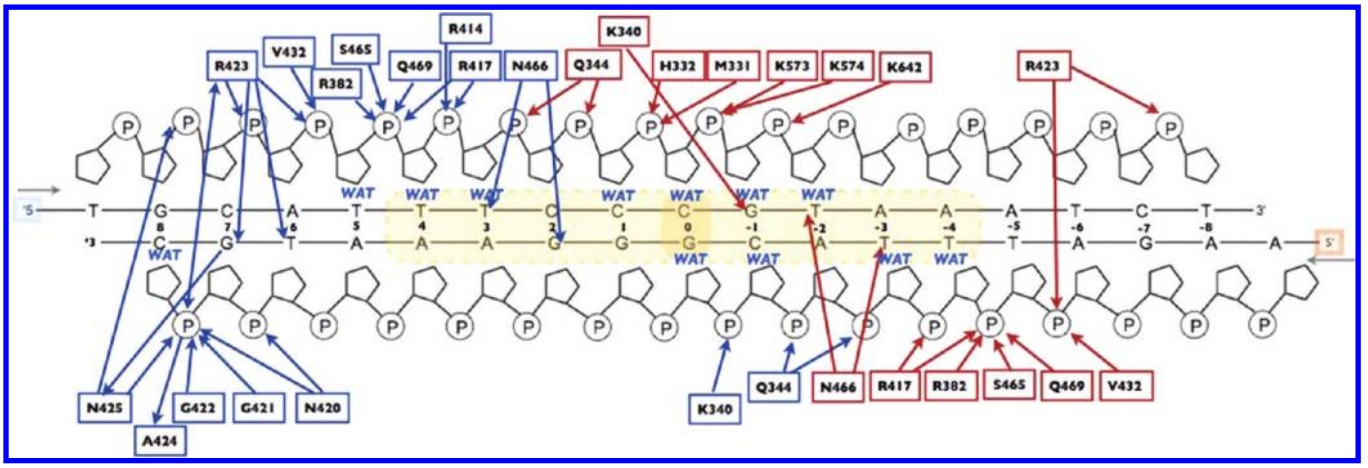


Figure 4. STAT3-DNA-water interactions map representing the residues of both STAT3 monomer-A *(blue)* and monomer-B *(red)* that form direct hydrogen bonds with the individual nucleotides of the DNA duplex over the course of the MD simulation. The 9-bp high affinity site is highlighted as a yellow bar. Arrows pointing toward/from the DNA backbone (represented by the phosphate group in the diagram) and the DNA bases represent the hydrogen bonds formed. Water molecules identified to form hydrogen bonds with the DNA with overall residency longer than 1 ns are shown (labeled as WAT).

STAT3β monomers, and at the SH2 domain of monomer-B (E616, K642, and Q643).

When analyzing the interfacial contacts, particular attention was paid to mapping the hydrogen bonds formed between the sugar-phosphate backbone of the DNA and protein residues and in particular to those hydrogen bonds that involve DNA bases and/or have been reported to be affected by point mutations. Since the hydration of the protein surface may well be crucial for the stability and flexibility of the protein itself, the folding process, and molecular recognition, both water−protein and water−DNA interactions were analyzed, as well as features of protein−DNA recognition. The analysis of the hydrogen bonds formed between the solvent and STAT3-DNA interface throughout the MD simulation (1920 frames with 25 ps time-steps) used the VMD program.[38] The water molecules were analyzed in terms of overall time spent within a distance required for a hydrogen bond formation, and only those water molecules with residence time exceeding 1 ns were considered. The hydration at the protein−DNA interface was also analyzed in terms of water density maps (Figure 3a-f).

Following these criteria, the STAT3-DNA-solvent hydrogen bonds were analyzed (Figures 4 and7 and Tables 1 and 2). Nine hydrogen bonds were identified involving the sugar-phosphate atoms of the DNA backbone, or the atoms from the bases, at both monomer interfaces, namely K340 (G(0) and G(-1)), Q344 (bp±1, ±2), R382 (bp±4), R417 (bp±3, ±4), R423 (bp±5, ±8), V432 (bp±5), S465 (bp±4), N466 (bp±2, ±3), and Q469 (bp±4). Monomer-A further formed stable hydrogen bonds between residues R414 − T(+3), N420 − C3'(+8) and G(+7), G421-C3'(+8), G422 − C3'(+8), R423 − A(+6), T(+6) and G(+7), N425 − G(+8), G(+7), and C(+7). Protein residues are always hydrogen bond-donors and the DNA bases are hydrogen bond-acceptors, with the exception of G(+7) and C(+8) that behaved both as hydrogen bond-donors and acceptors (Figure S). Monomer-B exclusively formed hydrogen bonds with the DNA sugar-phosphate backbone or the atoms from the bases; M331 − C(0), H332 − C(0), K573 − G(-1), K574 − G(-1), and K642 − T(-2). Despite fewer hydrogen bonds being formed between monomer-B and the DNA, the occupancy period of these interactions was in general longer comparing to the numerous interactions formed by monomer-A. This is due to the greater mobility of the DNA-binding domain of monomer A, allowing the hydrogen bond-interacting residues in loops ab, cx, ef, and gα5 to reach into the major groove while the interface with the monomer-B is less variable as the simulation time progresses (Figure S2). R382, S465, Q469, and N466 in both monomers formed hydrogen bonds with the DNA that lasted practically throughout the entire simulation time, with N466 specifically interacting with both DNA chains. Another newly described interaction arises from R423 of monomer-A, specifically recognizing bases T(+6) and G(+7) at the complementary DNA strand, and further forming hydrogen bond with the terminal base at the C3' end of the opposite DNA chain, and A(+6) and T(+5), acting both as a hydrogen bond donor and acceptor. Also R423 of monomer-B formed hydrogen bonds with both chains of the DNA, namely T(-5) and T(-8). N425 of monomer-A specifically recognized base G(+7) with the DNA base behaving like hydrogen bond donor.

The most frequently occurring residues in the protein−DNA interface are Arg, Asn, Lys, and Gln residues, with Arg residues located in the AT-rich minor groove, in agreement with experimental observations.[47] There are also a large number of

Table 2. Bridging Water Molecules at the Protein−DNA Interface That Mediate the Indirect Hydrogen Bonds Formed between the Residues of Monomer-A (*White*) and Monomer-B (*Grey*) and the DNA

STAT3 monomer	Bridging waters	duplex DNA
E415	WAT 1	dT(+4)
E415, R282, T341	WAT 2	dT(+4), dT(+3)
E415	WAT 3	dT(+4)
Q344	WAT 4	dC(-1)
E415	WAT 5	dT(+4)
Q469	WAT 6	dT(+5)
Q344	WAT 7	dC(-1)
K340, I467	WAT 8	dC(0), dG(-1), dT(-2)
E415, T341	WAT 9	dT(-3), dT(-4)
T341	WAT 10	dT(-3)
K340, V338, E415	WAT 11	dC(0)
E415	WAT 12	dT(-4)
E415	WAT 13	dT(-4)
E415	WAT 14	dT(-4)
E415	WAT 15	dT(-4)
K340, V343, L413	WAT 16	dC(0)

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Gly residues at the monomer-A -DNA interface, in particular forming hydrogen bonds with the terminal base at the C3' end of the DNA strand. There are a total of 25 contacts formed between 21 different water molecules and 12 individual nucleotides (+5, +4, +3, +1, 0, −1, and −2 in the 5′-T strand 1; −4, −3, −1, 0, and +8 in the 5′-A strand 2). Sixteen of these 21 water molecules (water molecules are numbered 1−16 for simplicity - see Table 2) bridge to the interfacial protein residues, forming indirect hydrogen bonds, for instance with E415 and T341 in both monomers, and additionally with Q344 and Q469 of monomer-A, and K340, V343, and I467 in monomer B. There is a pronounced pattern in the frequency of interactions of base pairs T±4, T±3 with both protein residues and water molecules, and there are also numerous hydrogen bonds involved with bp0 and bp±1, despite the high sequence variability at these positions in natural DNA target sites.[14]

We have shown here that the solvent molecules are indeed crucial for specific STAT3β-DNA recognition, expanding the contact area formed between the protein and duplex DNA. For comparison, the same water-DNA contact analysis performed with the 17-bp DNA alone suggest that only two water molecules interact with A(+5) and T(+4) of the opposite DNA strand and do not exceed 2% of the overall occupancy time spent within that hydrogen bond. In terms of the U-STAT3 monomer, only one (E616) out of many residues previously determined to be at the interface of the STAT3-DNA complex in interaction with a solvent molecule was found to actually form a corresponding hydrogen bond with a water molecule, again with a shorter occupancy time. Three more interfacial residues determined from the STAT3-DNA complex were found to form hydrogen bonds with water molecules in U-STAT3, namely A424, C426, and F384 (with occupancy times not exceeding 3% of the analyzed 48 ns). Hydration maps representing the water density at bp±4 are shown in Figure 3 (c, d), where numerous solvent molecules forming hydrogen bonds with the nucleotide were identified (as also shown in Table 2). Three hydrogen bonds were also found to be formed between C(0) and the solvent molecules (Table 2 and Figure 3e) or between K340 of monomer B with the solvent molecules (Figure 3b).

Mutations in the Interface. Thirty-one different dominant-negative STAT3 mutations have been identified to date, affecting predominantly the DNA-binding and SH2 domains of STAT3. The clinical importance of these mutations, accounting for the vast majority of HIES cases, has been reported and summarized.[26,27] Since the focus of this paper is on the protein−DNA recognition aspects of STAT3, those point mutations affecting the DNA-binding region (17 mutations) were examined for possible structural explanations of these mutations, which are summarized in Table 3. There is a level of symmetry between the monomers in terms of the residues affected by the point mutations. Thus R382 has the largest frequency of mutations observed in HIES patients, and it affects hydrogen-bonded DNA recognition with V463 and E435. Five different mutations of N466 have been reported, with low frequency of occurrence but with crucial effects on the direct readout involving bases at both strands of the DNA, and further interactions with C468 and Q469 (~65% and ~35% hydrogen bond occupancy respectively), causing a cascade of the subsequent alterations in STAT3-DNA recognition. rms fluctuations on a per residue basis, for the DNA-binding region, have been examined as a comparison between the phosphorylated complex-bound monomers-A and -B and unphosphorylated U-STAT3 monomer, to explore the dynamics of the interfacial residues, and in particular the residues for which mutations were observed. However there is not a significant correlation between the residues with larger values of rms and the occupancy time they spent forming hydrogen bonds with the DNA (Figure S3), and so are not discussed further.

In order to better map out the positions of the interfacial residues that may be mutated in terms of the phosphorylated complex-bound STAT3 and U-STAT3 monomer, the first eigenvectors for the backbone atoms were structurally aligned (duplex DNA was kept in place as a reference). The locations of the affected residues are in good agreement in the two structures, with the exception of R423 that is further away from the groove in the case of the U-STAT3 monomer (Figure 4). This may be explained by the high affinity of the residues within the *ef*-loop of the complex-bound STAT3 for DNA, hence having a large effect on the mobility of that particular region. Overall, the most significant structural difference between the two monomers is at the protein−protein interaction region, in agreement with the results discussed above.

Table 3. Point Mutations within the DNA-Binding Region and Their Interactions[a]

point mutation	patients reported	HB affected in monomer-A	HB affected in monomer-B
H332Y	3	M329, D334, R335, WAT	dC(0), M329, R335, K573
R335W	2	H332, D334, D566, 2x WAT	M329, H332, D334, D566, S574, 3x WAT
K340N	1	dG(0), M329, V343	T341, V343, WAT8, WAT11, WAT16
G342D	1	L413	L413
V343L	1	K340, T412	K340, T412, WAT16
R382L	2	dT(+4), E435, V463	dT(−4), E435, V463
R382W	35	WAT2, WAT	
R382Q	14		
F384L	2	2x WAT	D369, K383
F384S	1		
T389I	1	K409, H410	---
T412S	1	L387, V343	V343, L387, L411
R423Q	6	dA(+6), dT(+6), dT(+5)	dT(−5), dA(−6), dT3′(−8)
		dG(+7), dC3′(+8), G380, WAT	G380, A428, L430, E435
V432M	1	dT(+5), E435	dT(−5), E435
H437P	1	D369, E435, V461	D369, E435, V461
H437Y	1		
S465A	2	dT(+4), Q469	dT(−4), Q469
N466D	1	dG(+2), dT(+3), C468, Q469	dT(−2), dT(−3), C468, Q469
N466S	1		
N466T	1		
N466K	1		
N466H	1		
Q469H	1	dT(+4), S465, N466, A473, WAT6	dT(−4), S465, N466, A473
N472D	1	S476, 2x WAT	S476, 2x WAT
K642E	1	N567, D570, E616, L645, N646	dT(−2), D570, E616, L645, N646

[a]All hydrogen bonds formed between the mutated residue and any other protein residue, DNA base, or solvent molecule throughout the course of the simulation, are listed. Bridging water molecules involved in the interactions are specifically labeled (corresponding to those in Table 2).

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

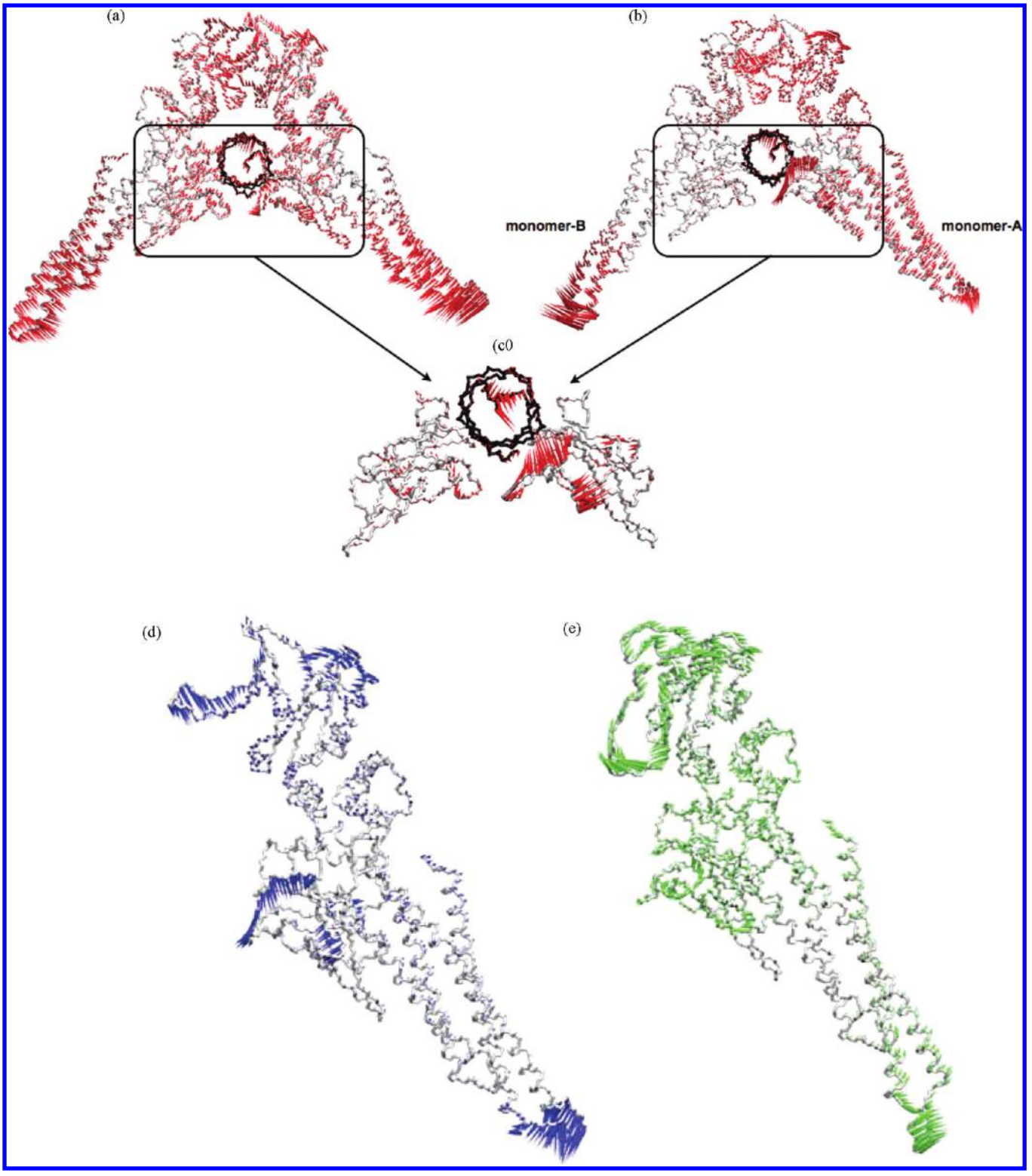


Figure 5. Porcupine plots of the first (a) and second (b) eigenvector for the STAT3β-DNA complex simulation with detailed focus on the characteristic movement of the monomer-A DNA-binding domain at the protein-DNA interface, which differs significantly from that of the monomer-B DNA-binding domain movement (c) and its subsequent interaction with the DNA. Porcupine plots of the first eigenvector for the STAT3 complex-bound monomer-A (d) and unphosphorylated STAT3 monomer (e). The models are shown as a bond trace, with the arrows attached to each backbone atom representing the scope and magnitude of the motions.

Cluster Analysis. Cluster analysis was applied to the large amount of data generated in order to provide a statistical description of the dynamics of the STAT3-DNA complex (only residues 321 to 688 of both monomers, corresponding to DNA-binding domains, linker domains and SH2 domain were used as the protein part of the complex for clustering, with the duplex DNA being kept intact). This analysis used the 1920 frames extracted from the MD trajectory at a time interval of 25

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

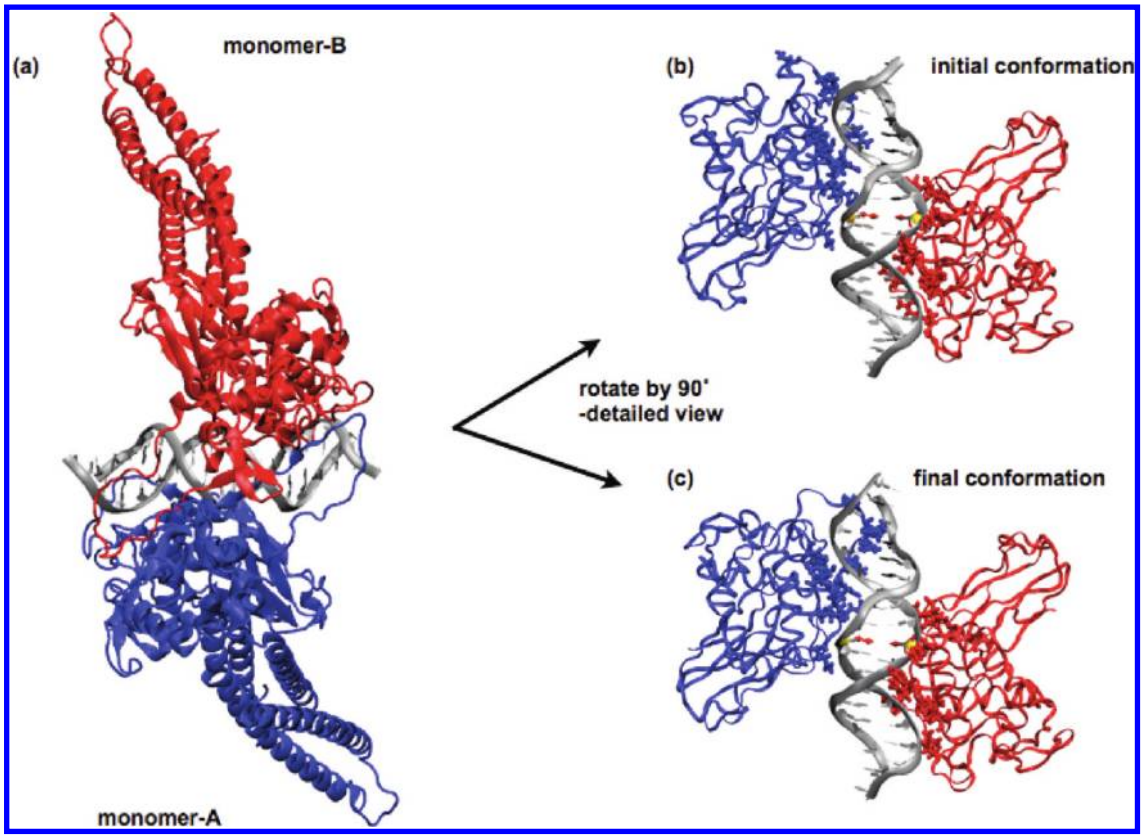


Figure 6. Structural model of STAT3β-DNA complex (top view) in cartoon representation, with the (a) monomer-A *(blue)*, monomer-B *(red)*, the 17-bp DNA duplex *(gray)* with bp(0) (see Figure 4) highlighted in stick representation *(yellow)*. Comparing the conformation of the initial (b) and final (c) structure of the MD simulation, tighter contact of the hydrogen bond-forming residues of the DNA-binding domain of monomer-A can be observed than in the corresponding region of monomer-B.

ps (~48 ns) for the matrix construction, with a 2.0 Å RMSD cut-off applied for the neighbour search. The first 5 of the 12 resulting clusters are represented by middle structures (Figure S3) – these are the conformations sampled during the simulation at 300 K. These were used for the analysis of the protein–DNA interface detailed above. The first five clusters covered approximately 98% of the total ensemble of sampled conformational space. Cluster 1 was the predominant conformer, being populated approximately 57% of the simulation time, whereas cluster 2 spanned over 10 ns of the simulation (22% of the simulation time).

Representative structures of the true conformational space at 300 K obtained from the cluster analysis of the MD trajectories subsequently provided multiple target conformations for small molecule virtual screenings (to be reported elsewhere). This approach is supported by recent studies demonstrating that state-of-the-art docking algorithms predict an incorrect binding pose for about 50–70% of ligands when only a single fixed receptor conformation is considered.[46]

Principal Component Analysis. The overall patterns of motions in the STAT3β homodimer:DNA complex (and its individual monomers), the U-STAT3 monomer, and the 17-bp duplex DNA were identified via principal component analysis (PCA) to give a quantitative assessment of the correspondence between the MD data sets. PCA was performed on the backbone atoms of the STAT3 models for residues 321 to 688 of the protein with the DNA duplex, employing a trajectory from the last 40 ns of each of the simulations (with a 25 ps time-step). The first 10 eigenvectors were considered for further analysis, which showed that the first three eigenvectors account for ~50% of all the motion in the simulated systems (Figure S4).

We have graphically represented the first two eigenvectors by means of porcupine plots[42] that show the scope and magnitude of the motion (given by the first eigenvector) for each of the backbone atoms. The immediate DNA-binding domain plus DNA show the same features as the complete protein and DNA backbone (Figure 5a). The complex-bound motion of STAT3 monomer-A is compared with the U-STAT3 monomer in Figure 5b. The first principal component for the STAT3-DNA complex confirms the observation made from the RMS plots that the most prominent motions are observed at the loops within the coiled-coil domain (the loop connecting helices α1 and α2), resulting in a characteristic scissor-like motion,[16] with the duplex DNA in the middle acting as a "hinge". Several observations can be made from the porcupine plots (corresponding to data from the RMS plots): (1) Monomer-A undergoes overall greater motion than monomer-B. (2) The dynamics/motion of the DNA-binding domain of monomer-A suggests that the DNA-interacting loops of monomer-A are inserted deeper into the DNA groove (Figures 6 and 7). This results in a greater span of interaction with the DNA duplex (further discussed below) and also indicates DNA unwinding at the T5′ end. The larger magnitude of movement at the T5′ of the DNA is also in accord with this observation. (3) The major flexibility at the SH2 domains region arises from the looplike stretch of C-terminal residues (residues 689–716) that forms an arm which reciprocally binds onto the other monomer.

dx.doi.org/10.1021/ci200625q | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

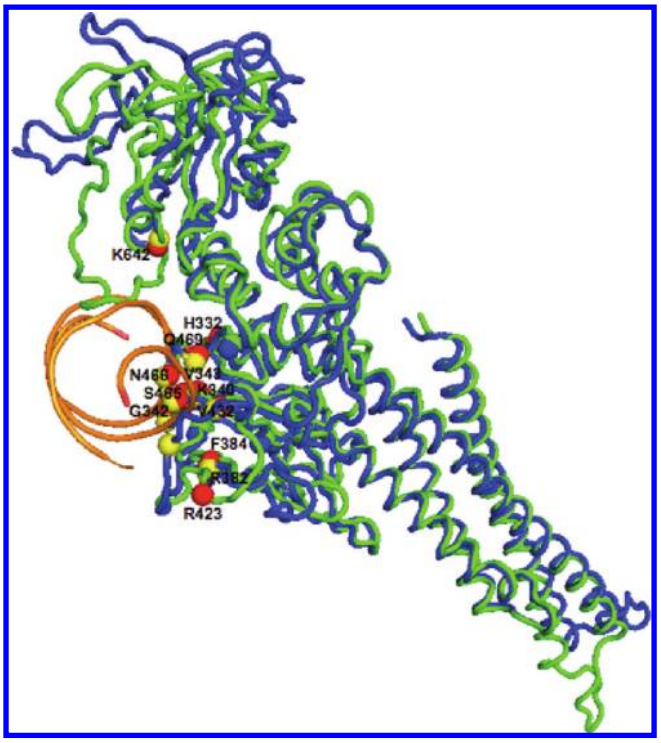


Figure 7. Structural alignment of the first eigenvectors for the backbone atoms of the phosphorylated STAT3 monomer-A *(blue)* and unphosphorylated STAT3 monomer *(green)* with the single point mutations at the DNA-binding region highlighted in sphere representation (*yellow* for phosphorylated monomer-A, *red* for U-monomer). The DNA duplex *(orange)* corresponding to the conformation of the first eigenvector (for the STAT3 complex) is displayed in order to better define the orientation of this view.

With respect to the predominant motion within the U-STAT3 monomer, the overall magnitude of fluctuations at the incremented regions is smaller comparing to those for monomer-A (Figure 5b), since there are no interaction partners. The most significant difference in the movement lies within the SH2 domain with the folded C-terminal stretch of residues, and the direction of the movement is opposite compared to that in complex-bound monomer-A.

SUMMARY AND CONCLUSIONS

1 Despite their sequence and secondary structure identity, distinct dynamic properties/behavior was observed for monomers A and B, and their characteristic dynamic properties contrast with that of the unphosphorylated STAT3 monomer. There are differences in their magnitude of motion, and in particular in the SH2 domain with the stretch of C-terminal residues, that clearly folds in the case of U-STAT3.

2 Protein−DNA-solvent interactions focused on those water molecules that remained at the interface with an occupancy greater than 1 ns of the simulation time. Water molecules were shown to be important mediators of protein−DNA recognition; there were a number directly involved in recognition, forming stable multiple hydrogen bonds with the protein and DNA residues. Bridging water molecules were identified not only in the protein−DNA interfacial interactions (residues within 5.0 Å of the DNA) but also between V338(B)− C(0) and L413(B)−C(0). While the water-DNA interactions in the STAT3 complex were stable over the simulation time, only two water molecules forming hydrogen bonds with the duplex DNA alone were identified which did exceed the overall 1 ns occupancy (hydrogen bonds with A(+5) and T(+4)). The simulation of the duplex DNA alone did not show unwinding at the terminal regions on the time scale of the simulations, despite the 5′ overhanging ends. The protein dimer does then have a stabilizing effect on the DNA, although DNA unwinding was observed for the complex-bound DNA in the 5′-T region as a consequence of the multiple interactions of monomer A with the opposite chain of the DNA in the region C3′(+8), G(+7).

3 Experimentally determined point mutations in the DNA-binding domain were located at the protein−DNA interface.

4 Molecular dynamics provides more detailed insight into the structural description of the protein−DNA recognition compared to the only currently available crystal structure of the STAT3β-DNA complex, which has been used to date for docking studies of STAT3 SH2 small-molecule ligands. These new data should therefore provide a more-robust platform for in silico approaches to the design of STAT3-STAT3 inhibitors.

ASSOCIATED CONTENT

Supporting Information

A summary of the simulations together with rmsd values (Table S1) and Figures S1−S4. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 44 207 753 5969. Fax: 44 207 753 5970. E-mail s.neidle@ucl.ac.uk. Corresponding author address: UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK.

Present Addresses

[§]School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, U.K.

[‖]Department of Pharmacy, King's College London, Franklin-Wilkins Building 150 Stamford Street, London SE1 9NH, U.K.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Cancer Research UK is thanked for providing financial support for these studies (A9327 to D.E.T. and C129/A4489 to S.N.).

REFERENCES

(1) Darnell, J. E. STATs and gene regulation. *Science* **1997**, *277*, 1630−1635.

(2) Fletcher, S.; Turkson, J.; Gunning, P. T. Molecular approaches towards the inhibition of the signal transducer and activator of transcription 3 (STAT3) protein. *ChemMedChem* **2008**, *3*, 1159−1168.

(3) Page, B. D. G.; Ball, D. P.; Gunning, P. T. Signal transducer and activator of transcription 3 inhibitors: a patent review. *Expert Opin. Ther. Pat.* **2011**, *21*, 65−83.

(4) Braunstein, J.; Brutsaert, S.; Olson, R.; Schindler, C STATs dimerize in the absence of phosphorylation. *J. Biol. Chem.* **2003**, *278*, 34133−34140.

(5) Yu, H.; Kortylewski, M.; Pardoll, D. Crosstalk between cancer and immune cells: role of STAT3 in the tumour microenvironment. *Nat. Rev. Immunol.* **2007**, *7*, 41−51.
(6) Yue, P.; Turkson, J. Targeting STAT3 in cancer: how successful are we? *Expert Opin. Invest. Drugs* **2009**, *18*, 45−56.
(7) Jing, N.; Tweardy, D. J. Targeting STAT3 in cancer therapy. *Anticancer Drugs* **2005**, *16*, 601−607.
(8) Song, H.; Wang, R.; Wang, S.; Lin, J. A low-molecular-weight compound discovered through virtual database screening inhibits STAT3 function in breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4700−4705.
(9) Siddiquee, K.; Zhang, S.; Guida, W. C.; Blaskovich, M. A.; Greedy, B.; Lawrence, H. R.; Yip, M. L. R.; Jove, R.; McLaughlin, M. M.; Lawrence, N. J.; Sebti, S. M.; Turkson, J. Selective chemical probe inhibitor of STAT3, identified through structure-based virtual screening, induces antitumor activity. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7391−7396.
(10) Siddiquee, K. A. Z.; Gunning, P. T.; Glenn, M.; Katt, W. P.; Zhang, S.; Schroeck, C.; Sebti, S. M.; Jove, R.; Hamilton, A. D. An oxazole-based small-molecule STAT3 inhibitor modulates STAT3 stability and processing and induces antitumor cell effects. *ACS Chem. Biol.* **2007**, *2*, 787−798.
(11) Xu, X.; Kasembeli, M. M.; Jiang, X.; Tweardy, B. J.; Tweardy, D. J. Chemical probes that competitively and selectively inhibit STAT3 activation. *PLoS One* **2009**, *4*, 1−12.
(12) Matsuno, K.; Masuda, Y.; Uehara, Y.; Sato, H.; Muroya, A.; Takahashi, O.; Yokotagawa, T.; Furuya, T.; Okawara, T.; Otsuka, M.; Ogo, N.; Ashizawa, T.; Oshita, C.; Sachiko, T.; Ishii, H.; Akiyama, Y.; Asai, A. Identification of a new series of STAT3 inhibitors by virtual screening. *ACS Med. Chem. Lett.* **2010**, *1*, 371−375.
(13) Zinzalla, G.; Thurston, D. E. Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Med. Chem.* **2009**, *1*, 65−93.
(14) Becker, S.; Groner, B.; Muller, C. W. Three-dimensional structure of the STAT3 [beta] homodimer bound to DNA. *Nature* **1998**, *394*, 145−151.
(15) Ren, Z.; Mao, X.; Mertens, C.; Krishnaraj, R.; Qin, J.; Mandal, P. K.; Romanwski, M. J.; McMurray, J. S.; Chen, X. Crystal structure of unphosphorylated STAT3 core fragment. *Biochem. Biophys. Res. Commun.* **2008**, *374*, 1−5.
(16) Lin, J.; Buettner, R.; Yuan, Y. C.; Yip, R.; Horne, G.; Jove, R.; Vaidehi, N. Molecular dynamics simulations of the conformational changes in signal transducers and activators of transcription, STAT1 and STAT3. *J. Mol. Graphics Modell.* **2009**, *28*, 347−356.
(17) Park, I.-H.; Li, C. Characterization of molecular recognition of STAT3 SH2 domain inhibitors through molecular simulation. *J. Mol. Recognit.* **2010**, *23*, 1−12.
(18) Baron, R.; Setny, P.; Andrew McCammon, J. Water in cavity−ligand recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091−12097.
(19) Henchman, R. H.; McCammon, J. A. Structural and dynamic properties of water around acetylcholinesterase. *Protein Sci.* **2002**, *11*, 2080−2090.
(20) De Simone, A.; Dodson, G. G.; Verma, C. S.; Zagari, A.; Fraternali, F. Prion and water: tight and dynamical hydration sites have a key role in structural stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7535−7540.
(21) Scorciapino, M. A.; Robertazzi, A.; Casu, M.; Ruggerone, P.; Ceccarelli, M. Heme proteins: the role of solvent in the dynamics of gates and portals. *J. Am. Chem. Soc.* **2010**, *132*, 5156−63.
(22) Reddy, C. K.; Das, A.; Jayaram, B. Do water molecules mediate protein-DNA recognition? *J. Mol. Biol.* **2001**, *314*, 619−632.
(23) Jayaram, B.; Jain, T. The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 343−361.
(24) Minegishi, Y.; Saito, M.; Tsuchiya, S.; Tsuge, I.; Takada, H.; Hara, T.; Kawamura, N.; Ariga, T.; Pasic, S.; Stojkovic, O.; Metin, A.; Karasuyama, H. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* **2007**, *448*, 1058−1062.
(25) Holland, S. M.; DeLeo, F. R.; Elloumi, H. Z.; Hsu, A. P.; Uzel, G.; Brodsky, N.; Freeman, A. F.; Demidowich, A.; Davis, J.; Turner, M. L.; Anderson, V. L.; Darnell, D. N.; Welch, P. A.; Kuhns, D. B.; Frucht, D. M.; Malech, H. L.; Gallin, J. I.; Kobayashi, S. D.; Whitney, A. R.; Voyich, J. M.; Musser, J. M.; Woellner, C.; Schäffer, A. A.; Puck, J. M.; Grimbacher, B. STAT3 mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* **2007**, *357*, 1608−1619.
(26) Woellner, C.; et al. Mutations in STAT3 and diagnostic guidelines for hyper-IgE syndrome. *J. Allergy Clin. Immunol.* **2010**, *125*, 424−432.
(27) Heimall, J.; Davis, J.; Shaw, P. A.; Hsu, A. P.; Gu, W.; Welch, P.; Holland, S. M.; Freeman, A. F. Paucity of genotype-phenotype correlations in STAT3 mutation positive Hyper IgE Syndrome (HIES). *Clin. Immunol.* **2011**, *139*, 75−84.
(28) Horvath, C. M.; Wen, Z.; Darnell, J.E. A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev.* **1995**, *9*, 984−994.
(29) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(30) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Margakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(31) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Chaetham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(32) Homeyer, N.; Horn, A.; Lanig, H.; Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.* **2006**, *12*, 281−289.
(33) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(34) Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7189.
(35) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(36) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(38) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(39) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412−425.
(40) Mesentean, S.; Fischer, S.; Smith, J. C. Analyzing large-scale structural change in proteins: comparison of principal component projection and Sammon mapping. *Proteins* **2006**, *64*, 210−218.
(41) Tai, K.; Shen, T.; Börjesson, U.; Philippopoulos, M.; McCammon, J. A. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J.* **2001**, *81*, 715−724.
(42) Haider, S.; Parkinson, G. N.; Neidle, S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J.* **2008**, *95*, 296−311.
(43) Bock, H. H. Probabilistic models in cluster analysis. *Comp. Stat. Data Anal.* **1996**, *23*, 5−28.
(44) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(45) Pettersen, E. F.; et al. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(46) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.

(47) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248−1253.