Structural basis for DNA recognition by STAT6

Jing Li[a,b,1], Jose Pindado Rodriguez[c,1], Fengfeng Niu[a], Mengchen Pu[a], Jinan Wang[d], Li-Wei Hung[e], Qiang Shao[d], Yanping Zhu[a], Wei Ding[a], Yanqing Liu[a], Yurong Da[f], Zhi Yao[f], Jie Yang[f], Yongfang Zhao[a], Gong-Hong Wei[g], Genhong Cheng[c], Zhi-Jie Liu[a,f,h,i], and Songying Ouyang[a,c,2]

[a]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; [b]University of Chinese Academy of Sciences, Beijing 100049, China; [c]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095; [d]Drug Discovery and Design Center, Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; [e]Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545; [f]Department of Immunology, Tianjin Medical University, Tianjin 300070, China; [g]Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu 90220, Finland; [h]iHuman Institute, Shanghai Tech University, Shanghai 201210, China; and [i]Institute of Molecular and Clinical Medicine, Kunming Medical University, Kunming 650500, China

Edited by Wei Yang, National Institutes of Health, Bethesda, MD, and approved October 10, 2016 (received for review July 9, 2016)

STAT6 participates in classical IL-4/IL-13 signaling and stimulator of interferon genes-mediated antiviral innate immune responses. Aberrations in STAT6-mediated signaling are linked to development of asthma and diseases of the immune system. In addition, STAT6 remains constitutively active in multiple types of cancer. Therefore, targeting STAT6 is an attractive proposition for treating related diseases. Although a lot is known about the role of STAT6 in transcriptional regulation, molecular details on how STAT6 recognizes and binds specific segments of DNA to exert its function are not clearly understood. Here, we report the crystal structures of a homodimer of phosphorylated STAT6 core fragment ($STAT6^{CF}$) alone and bound with the N3 and N4 DNA binding site. Analysis of the structures reveals that STAT6 undergoes a dramatic conformational change on DNA binding, which was further validated by performing molecular dynamics simulation studies and small angle X-ray scattering analysis. Our data show that a larger angle at the intersection where the two protomers of STAT meet and the presence of a unique residue, H415, in the DNA-binding domain play important roles in discrimination of the N4 site DNA from the N3 site by STAT6. H415N mutation of $STAT6^{CF}$ decreased affinity of the protein for the N4 site DNA, but increased its affinity for N3 site DNA, both in vitro and in vivo. Results of our structure–function studies on STAT6 shed light on mechanism of DNA recognition by STATs in general and explain the reasons underlying STAT6's preference for N4 site DNA over N3.

STAT6 | N4 site DNA recognition | JAK-STAT pathway | antiviral innate immunity | crystal structure

Proteins belonging to the STAT family mediate transmission of signals of numerous cytokines and growth factors from the cell membrane to the nucleus via the classical JAK-STAT pathway (1). Malfunctions in this pathway are known to result in immune system disorder and cancers. Therefore, the JAK-STAT pathway is considered to be of great importance in the development of therapeutic interventions (2). The mammalian STAT family is made up of seven structurally and functionally related proteins named STAT1, 2, 3, 4, 5a, 5b, and 6 (3). All of the STAT proteins share a conserved domain organization (Fig. 1*A*).

STAT6, an important member of the STAT family, plays a crucial role in the differentiation of Th2 cells and has been implicated in the development of asthma (4). This STAT is primarily stimulated by IL-4 and IL-13. A recent study reported that STAT6 plays a pivotal role in antiviral signaling initiated by host cells in response to viral infections (5). STAT6 could be activated by the stimulator of interferon genes/TBK1 cascade via phosphorylation of Y641. Intriguingly, residue S407 located in the DNA-binding domain (DBD) of STAT6 has been shown to be phosphorylated by TBK1. However, its implication for biological function of STAT6 is currently unknown (5). Thus, structural studies on STAT6 and its complex with DNA are essential to address several unanswered questions related to signals that morph STAT6 into a conformation, which is competent for binding DNA and transcription of genes.

STATs recognize DNA motifs with a consensus sequence of 5′-TTC$N_{3/4}$GAA-3′ in the regions of target genes that regulate expression of the protein, where $N_{3/4}$ denotes a spacer consisting of three (N3) or four (N4) nucleotides of any of the four types found in the DNA. STAT6 is reported as the only member that prefers N4 over N3 site DNA (6), which is in agreement with our observations during initial studies on DNA binding by STATs (Fig. S1) (7). Interestingly, STAT5 has been shown to bind N4 site DNA weakly (8). Thus far, only homodimeric structures of STAT1 and STAT3 in complex with N3 site DNA [phosphorylated STAT1 and STAT3, protein data bank (PDB) ID codes 1BF5 and 1BGl, respectively; unphosphorylated STAT3, PDB ID code 4E68] have been solved (9–11). In addition, structures of unphosphorylated monomeric STAT1 (PDB ID code 1YVL) and STAT3 (PDB ID code 3CWG), as well as unphosphorylated dimeric STAT5a (PDB ID code 1Y1U) in protein-alone form, have been published (12–14). These structures shed light on the domain organization, nature of dimerization, and mode of N3 site DNA binding by STATs. However, to the best of our knowledge, there was no instance where structures of a phosphorylated STAT protein alone and in complex with DNA were available for the same STAT protein. This gap has hampered our ability to comprehend the conformational changes induced on DNA binding by a phosphorylated STAT. Furthermore, there is no structural information available on STAT6. This situation has further limited

Significance

STAT6 is a transcription factor and plays a predominant role in IL-4/IL-13 and virus-mediated signaling pathways. Extensive studies have linked malfunctions of STAT6 to pathological features of asthma and cancer. Targeting the function of STAT6 has become an attractive therapy. Understanding the molecular mechanisms of STAT6 transcriptional regulation is still scarce. Here, we report the atomic-level structures of the phosphorylated STAT6 core fragment homodimer, both in DNA-free and complexed with N4 or N3 site DNA, uncovering both a larger dimer interface intersection angle and the unique residue H415 of STAT6 as important factors for discrimination of N4 from N3 site DNA. This study uncovers a dramatic conformational change in STAT6 dimer for recognizing and preferring N4 site DNA.

Author contributions: J.L., Y.L., Y. Zhao, and S.O. designed research; J.L., J.P.R., F.N., J.W., L.-W.H., Y. Zhu, Y.L., Y.D., Z.Y., and J.Y. performed research; J.L., M.P., J.W., L.-W.H., Q.S., Y. Zhu, W.D., Y. Zhao, G.-H.W., G.C., Z.-J.L., and S.O. analyzed data; and J.L. and S.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4Y5U, 4Y5W, and 5D39).

[1]J.L. and J.P.R. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: ouyangsy@ibp.ac.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1611228113/-/DCSupplemental.

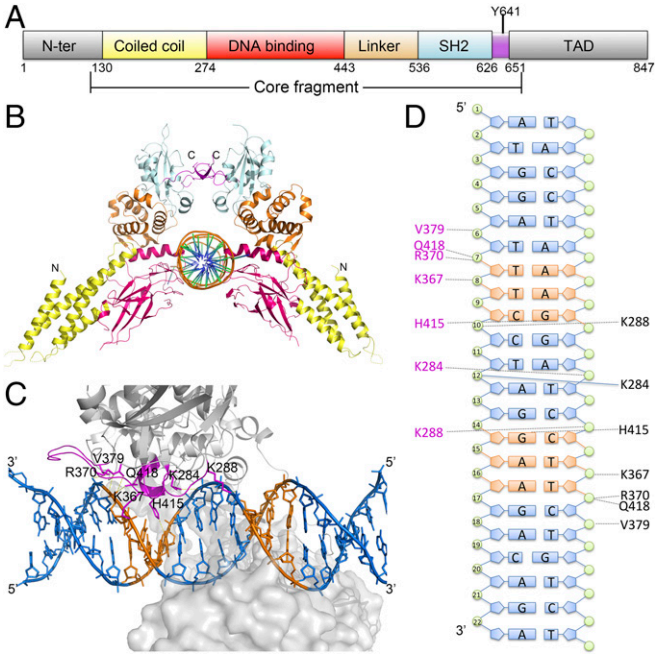

Fig. 1. Structure of STAT6^{CF} and N4 site DNA complex. (*A*) Schematic diagram showing the domain organization of human STAT6, including N-terminal domain (gray), coiled coil domain (yellow), DNA-binding domain (red), linker domain (orange), SH2 domain (cyan), and TAD domain (gray). The core fragment and phosphorylation site Y641 are indicated. (*B*) Cartoon diagram of the STAT6^{CF}-N4 complex (in front view). Colors of each domain are the same as in *A*. (*C*) Drawing depicting details of the STAT6^{CF}–DNA interface. The side chains of residues (A chain) donating hydrogen bonds are shown as magenta colored sticks, and the hydrogen bonds are shown as yellow dashed lines. The conserved palindromic bases (TTC/GAA) in both sides of the DNA molecule are shown in orange. (*D*) A schematic drawing highlighting STAT6^{CF}–DNA interactions. Residues forming hydrogen bonds are colored in magenta (D chain) and black (B chain).

our understanding of the mechanisms underlying DNA recognition by STAT6 and the molecular basis for how STAT6 discriminates N4 site DNA from N3 type of DNA.

Here, we report the crystal structures of phosphorylated dimeric STAT6 core fragment encompassing amino acids (aa) 123–658 (hereafter referred to as STAT6^{CF}) in unliganded form and its complexes with N3 and N4 site DNAs. Using these structures, we could analyze the conformational changes induced by DNA binding without having to resort to qualitative comparisons with phylogenetically distant homologs. Furthermore, we identified key residues in the DBD that are important for DNA recognition and verified their roles in the function of STAT6 by performing in vitro and in vivo experiments. Notably, we show that, a larger angle at the intersection where the two protomers of STAT meet and the unique residue H415 in the DBD of STAT6 are important for recognition of N4 site DNA.

Results

Phosphorylated STAT6^{CF} Forms a Homodimer and Binds DNA with High Affinity. Previous studies have shown that truncations of STAT, termed as the core fragment (STAT^{CF}), can bind DNA with affinities comparable to those of the full-length STATs (10, 15). Therefore, we designed a truncation encompassing human STAT6^{CF} (aa 123–658) based on alignment of primary sequence of STAT6 with STAT1 and STAT5a core fragments (Fig. 1*A*, Fig. S2, and Tables S1 and S2). The STAT6^{CF} purified to homogeneity existed as a monomer in solution. However, this form of STAT6^{CF} neither bound DNA nor did it crystallize after extensive screening. These features of STAT6 are unlike STAT1 and STAT3, which are known to bind DNA regardless of phosphorylation (11, 16). Because the unphosphorylated STAT6^{CF} did not bind DNA, we produced phosphorylated STAT6^{CF} as described previously (17, 18) (Fig. S3). The phosphorylated form of STAT6^{CF} eluted as a dimer (Fig. S3 *A* and *B*) and bound DNA with a high affinity (K_D = 79 nM for N4 site DNA; Table 1). Results of isothermal titration calorimetry (ITC), analytical ultracentrifugation (AUC) analysis, thermal shift assay (TSA), and gel filtration chromatography experiments were consistent with the ability of dimeric phosphorylated STAT6^{CF} to bind DNA (Fig. S3 *A*–*D*). Thus, phosphorylated STAT6^{CF} forms a homodimer that is capable of binding DNA in solution.

Overall Structure of STAT6^{CF} and Comparison with Other Unliganded STAT Structures. Crystal structure of the phosphorylated form of STAT6^{CF} homodimer was determined at 2.70-Å resolution (Table S3). Each asymmetric unit (ASU) contains two STAT6^{CF} protomers arranged in a dimer, consistent with the oligomeric state of the protein in solution. Two protomers within a dimer form a V-shaped structure. The architecture of STAT6^{CF} is similar to those of other STATs, and each protomer of STAT6^{CF} can be divided into five distinct modules: an N-terminal coiled coil domain, a DNA-binding domain, a linker domain, an SH2 domain, and a C-terminal phosphotyrosine tail segment (Fig. S4*A*).

Comparison of the structures indicates that all four α-helixes, especially the α1 and α2 helixes of the coiled coil domain of STAT6, are shorter than the corresponding regions of STAT1, STAT3, and STAT5a (Fig. S4*B*), which is consistent with the fact that STAT6 is 41 aa shorter than STAT1, 43 aa shorter than STAT3, and 49 aa shorter than STAT5a as indicated by primary sequence analysis (Fig. S2). In addition, this domain rotates ∼10° outward from the core of the dimer compared with unphosphorylated STAT1, 3, and 5 structures (Fig. S4*B*).

As expected, residue Y641 of STAT6^{CF} is phosphorylated, forming a phosphotyrosine tail segment, which mediates dimerization (Fig. S4*C*). The phosphorylated tail segment of each protomer passes through the gap between the two protomers to interact with the SH2 domain of the adjacent protomer and then returns, forming a gate with its own SH2 domain that allows the phosphorylated tail segment of the other molecule to pass through. These types of intermolecular interactions between the phosphorylated tail segments are also observed in STAT1 and STAT3 homodimers (9, 10). Two K647 residues located in the antiparallel β-sheet formed by the dimerization of STAT6 serve as hinge axis points around which the two molecules of STAT6^{CF} rotate when the dimer binds to DNA (Fig. S4*D*). In addition, our structure of STAT6^{CF} suggests that truncations ending before residue V650 lose the ability to bind to DNA (19) owing to the destruction of the intermolecular interactions of the β-strands of the phosphorylated tail segments, which affects dimerization.

A key structural difference between the dimers of STAT6 and STAT1/STAT3 is located at the C-terminal loop (aa 609–620) region of the SH2 domain. In STAT1 and STAT3, the C-terminal loop is 10 aa longer than in STAT6 (Fig. S4*E*). The loops from the two protomers are observed converging to form a closed tunnel just above the antiparallel β-sheet. In STAT6, this upper cover of the tunnel disappears completely because the two C-terminal loops are short. Another notable difference between the dimers of STAT6 and other STATs is observed at the intersection where the two protomers meet during dimerization. The angle formed at the intersection of dimerization between the protomers of STAT6 is larger than those observed for STAT1 and STAT3 dimers in complex with DNA, implying the possibility of functional differences between them (see below). Interestingly, like STAT6, STAT5 also contains a shorter C-terminal loop in its amino acid sequence (Fig. S2) so it would most likely assume a relatively more open dimeric structure after activation like STAT6. Thus, structural analysis seems to suggest that the dimeric assembly of STAT6

Table 1. Affinities of STAT6-WT, STAT1-WT, and their mutants for N3 and N4 site DNAs

Method	Protein	N4 site (CS4) K_D (M)	N4 site (IHG) K_D (M)	N3 site (M67) K_D (M)	N3 site (T1) K_D (M)
ITC	$STAT6^{CF}$-WT	$7.9 \times 10^{-8} \pm 7.8 \times 10^{-9}$	$1.2 \times 10^{-7} \pm 3.4 \times 10^{-8}$	$1.8 \times 10^{-6} \pm 2.0 \times 10^{-7}$	$2.8 \times 10^{-6} \pm 2.6 \times 10^{-7}$
	$STAT6^{CF}$-H415N	$2.2 \times 10^{-6} \pm 5.1 \times 10^{-7}$	$2.2 \times 10^{-6} \pm 8.7 \times 10^{-7}$	$2.4 \times 10^{-7} \pm 2.4 \times 10^{-8}$	$7.4 \times 10^{-7} \pm 2.5 \times 10^{-8}$
	$STAT6^{CF}$-K374E	$4.9 \times 10^{-7} \pm 2.0 \times 10^{-8}$	—	$5.5 \times 10^{-6} \pm 1.1 \times 10^{-7}$	—
SPR	$STAT6^{CF}$-WT	7.6×10^{-8}	—	6.5×10^{-7}	—
	$STAT6^{CF}$-H415N	1.0×10^{-6}	—	1.4×10^{-7}	—
	$STAT1^{CF}$-WT	3.2×10^{-7}	—	2.6×10^{-10}	—
	$STAT1^{CF}$-N460H	6.2×10^{-8}	—	2.2×10^{-9}	—

and STAT5 appears more flexible and accessible for DNA binding compared with other STATs.

Crystal Structures of $STAT6^{CF}$ in Complex with DNA. Crystal structures of $STAT6^{CF}$ in complex with N4 or N3 site DNA were both determined and refined to 3.10- and 3.20-Å resolutions, respectively (Table S3, Fig. 1, and Figs. S4 *F–H* and S5 *A–D*). B-form DNA molecules with lengths of 22 bp (N4 site) and 21 bp (N3 site, by deletion of an adenine in the N4 spacer) are modeled in our $STAT6^{CF}$-DNA complex structures (Fig. 1 and Fig. S5 *A–D*).

The two overall structures of $STAT6^{CF}$ bound with N4 and N3 DNA are similar, with an RMSD of 0.636 Å between the main chain C_α atoms of the protomer chains. The interfaces for DNA binding observed in the two complexes are highly conserved (Fig. S4*I*). One $STAT6^{CF}$ dimer binds one palindromic DNA duplex and each protomer binds only half of the palindromic sequence on average (Fig. 1*B*). Such interactions of $STAT6^{CF}$ with DNA are similar to those observed for $STAT1^{CF}$ and $STAT3^{CF}$ in their respective complexes with DNA (9, 10). Three loop regions (aa 284–288, aa 365–379, and aa 413–419) of the DBD participate in the recognition of DNA and binding of $STAT6^{CF}$ to the palindromic DNA (Fig. 1*C*). Several residues in the DBD, including K284, K288, K367, K370, K379, and Q418, form hydrogen bond interactions with the DNA backbone (Fig. 1*D*). Intriguingly, H415 is the only residue of $STAT6^{CF}$ that forms hydrogen bond with a base. Specifically, H415 forms a 2.91-Å hydrogen bond with the O^6 of guanine at position 14 in the N4 site DNA (position 13 in the N3 site DNA; Fig. 1*D* and Fig. S4*H*). This structural observation suggests an important role for residue H415 in recognition of DNA by $STAT6^{CF}$.

Residue H415 Is Essential for N4 Site DNA Recognition by STAT6. The architecture of the DBD of $STAT6^{CF}$ and the overall mode of DNA binding is similar to that of those reported for other STATs. However, STAT6 differs from other STATs in the ability to forge interactions with the bases of DNA. Although three residues of STAT1 form interactions with DNA bases, only one residue of STAT6, H415, forms a hydrogen bond with a DNA base. Furthermore, sequence alignment of all of the STATs indicates that H415 of STAT6 is replaced by an asparagine (N) in STAT1-4, whereas STAT5, which is reported to recognize a few N4 site DNAs with low affinity in addition to N3 site DNAs (8, 20), retains histidine (H471 in STAT5) at an equivalent position (Fig. 2*A*). Interestingly, in the unphosphorylated $STAT5^{CF}$ structure (PDB ID code 1Y1U), the side chain of H471 of STAT5 points toward the interior of the protein and does not form any interactions with its neighboring residues (Fig. S4*J*). However, in all of the structures of STAT6 being reported here, the side chain of residue H415 is oriented toward outside and poised for DNA binding. The relatively inaccessible nature of the position assumed by the side chain of H471 in STAT5 may partially explain its low affinity for N4 site DNA.

To investigate the role of residue H415 in recognition of DNA, we compared the affinities of a H415N mutant of $STAT6^{CF}$ for N3 and N4 site DNA with that of the WT $STAT6^{CF}$. Different types of DNA containing N3 or N4 site DNA motifs were selected as substrates for the studies. Results of ITC experiments revealed that the binding affinity of H415N mutant for an 18-bp-long stretch of the c-γ3 sterile transcript promoter (referred to as CS4; N4 site GAS motif) decreased dramatically from 79 nM to 2.2 μM (Table 1). Similarly, the affinity of H415N mutant for another N4 site DNA, IHG, decreased from 0.12 to 2.2 μM. In addition to N4 site DNAs, we also used two different types of N3 site DNA substrates, M67 and T1, for testing the affinity of the mutant for N3 site DNAs. Remarkably, the affinity of H415N mutant of $STAT6^{CF}$ for M67 increased 7.5 times (from 1.8 to 0.24 μM). Similarly, the affinity of the mutant for T1 increased 3.8 times (from 2.8 to 0.74 μM) (Table 1).

Next, we tested and confirmed the changes in DNA binding affinities of the H415N mutant independently by performing surface plasmon resonance (SPR) experiments. Similar to the ITC results, the affinity of mutant H415N for CS4 (N4 site DNA) decreased by almost 12-fold compared with that of $STAT6^{CF}$-WT, whereas the affinity for M67 (N3 site DNA) increased by 4.7-fold of that of $STAT6^{CF}$-WT (Table 1). Thus, a single mutation, H415N, switches the DNA binding preference of $STAT6^{CF}$ from the N4 site DNA to N3 in vitro.

To further corroborate our findings that residue H415 plays an important role in conferring specificity in recognition of DNA by STAT6, we mutated an equivalent residue of STAT1, N460, to histidine and tested the ability of N460H mutant of STAT1 to bind N3 and N4 site DNAs. The affinities of $STAT1^{CF}$-WT and the mutant $STAT1^{CF}$-N460H for the N3 and N4 site DNAs were tested using SPR method. Interestingly, $STAT1^{CF}$-N460H showed

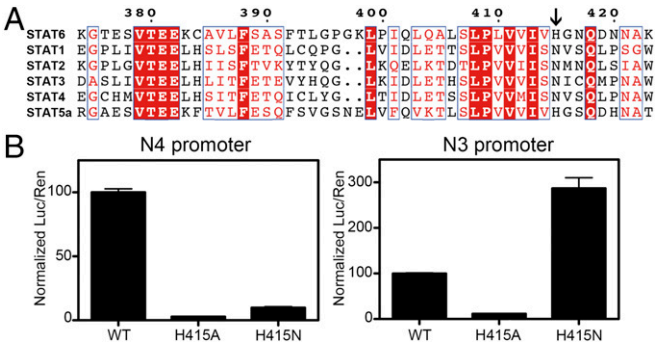

Fig. 2. Residue H415 is essential for N4 site DNA recognition by STAT6. (*A*) Multiple sequence alignment of STAT6 and other STAT proteins produced by ClusterW and ESpript (espript.ibcp.fr/ESPript/ESPript/). The location of residues (histidine in STAT6/STAT5 and asparagine in STAT1-4) used for distinguishing N4 and N3 site DNA are indicated by a black arrow. Strictly conserved residues are boxed in white on a red background, and highly conserved residues are boxed in red on a white background. (*B*) The plasmids containing $STAT6^{FL}$-WT, $STAT6^{FL}$-H415A, or $STAT6^{FL}$-H415N were transfected into HEK 293T cells together with renilla reporter and N4 site STAT6 luciferase reporter (*Left*) or N3 site STAT6 luciferase reporter genes (*Right*). After 24 h, cells were stimulated with IL-4 (10 ng/mL) for 2 h, and the results of dual-luciferase assay are shown for triplicate samples. Numbers are normalized with respect to the $STAT6^{FL}$-WT data and presented as percentages.

a remarkable decrease in affinity for N3 site DNA (M67) and an increase in affinity for N4 site DNA (CS4) compared with STAT1^{CF}-WT (Table 1). Thus, residue N460 of STAT1^{CF} probably plays an important role in conferring specificity for DNA binding. Taken together, the results of mutagenesis studies on H415 of STAT6 and N460 of STAT1 suggest that residue H415 plays an important role in conferring specificity on STAT6 for binding N4 site DNA.

To further verify the significance of residue H415 in conferring specificity for DNA binding, we performed in vivo luciferase reporter-based assays using full-length STAT6 (STAT6^{FL}). STAT6^{FL}-H415N and STAT6^{FL}-H415A mutants were generated and tested in vivo for their ability to regulate the reporter's expression via the N3 and N4 site DNAs on stimulation by IL-4 (Fig. 2*B*). As expected and consistent with the in vitro studies performed using STAT6^{CF}, STAT6^{FL}-H415N was deficient in its ability to activate an N4 site DNA. The same mutant exhibited a much stronger luciferase signal than that of STAT6^{FL}-WT when bound to the N3 site DNA (Fig. 2*B*). In stark contrast to the H415N mutation, STAT6^{FL}-H415A completely lost its ability to activate both N3 and N4 site DNAs, further confirming the essentiality of residue H415 in IL-4/IL-13 signaling (21). Thus, results of the in vitro and in vivo studies using mutants of STAT6 indicate that residue H415 of STAT6 plays an important role in the mechanism of discrimination of N4 site DNA from N3.

Structural Basis for Recognition of N4 Site DNA by STAT6. Superimposition of our N4 site DNA-bound STAT6^{CF} over N3 site DNA-bound STAT1 and STAT3 reveals that the mechanism of dimerization within STATs is similar (Fig. S4*K*). However, the angle of intersection where the two protomers of STAT6 meet in the DNA-bound and DNA-free STAT6^{CF} structures is larger than those observed for STAT1 or STAT3 in complex with DNA (Fig. S4*K*). Superimposition of our crystal structures of the DNA-free and N4 site DNA-bound STAT6^{CF} by aligning the SH2 domain in one protomer, as shown in Fig. 3*A*, reveals that the SH2 and linker domains overlapped well. However, the DNA-binding domain and coiled coil domain, when considered as a rigid body, have shifted slightly toward the DNA (Fig. 3*A*). The most remarkable difference between the two structures becomes evident when the dimers of unliganded and DNA-bound STAT6^{CF} are superimposed. The protomers within the dimer seem to have undergone a significant rotation during DNA binding. Consequently, the position of H415 shifts by 21.4 Å in the DNA-bound structure (Fig. 3*A* and Movie S1). Due to the flexibility between the four domains of STAT6^{CF}, the DBD of each STAT6^{CF} has rotated to an optimal position for DNA binding (Fig. 3*A*). The whole STAT6^{CF} dimer is squeezed into a more compact architecture after the phosphorylated STAT6^{CF} dimer encounters and binds N4 site DNA.

Similar to the STAT6^{CF}-N4 complex, the STAT6^{CF}-N3 complex also undergoes a rotational conformational change on DNA binding (Fig. S4*K*). We aligned the DBD of our STAT6^{CF}-N3 complex structure with the other three reported STAT protein structures in complex with N3 site DNA, i.e., STAT1 (PDB ID code 1BF5) and STAT3 (PDB ID codes 1BG1 and 4E68) (9–11). Although the proteins exhibit some flexibility, the position of the key residue, H415 in STAT6, N460 in STAT1, and N466 in STAT3, in both the protomers within the dimer of these STATs overlaps (Fig. S6*A*), suggesting that all members of the STAT family recognize N3 site DNA in a conserved fashion despite slight variations in the overall structures.

Superposition of the structure of the phosphorylated STAT6^{CF}-N4 site DNA over the structure of STAT1^{CF}-N3 DNA reveals that the overall topology of the members belonging to STAT family is highly conserved, further suggesting that they may share a similar mechanism of DNA recognition (Fig. S4*K*). We then compared the key residues of STAT6^{CF} and STAT1^{CF} participating in DNA binding. As shown in Fig. 3*B*, when DBDs are aligned, H415 of STAT6 and its counterpart N460 in STAT1, as well as

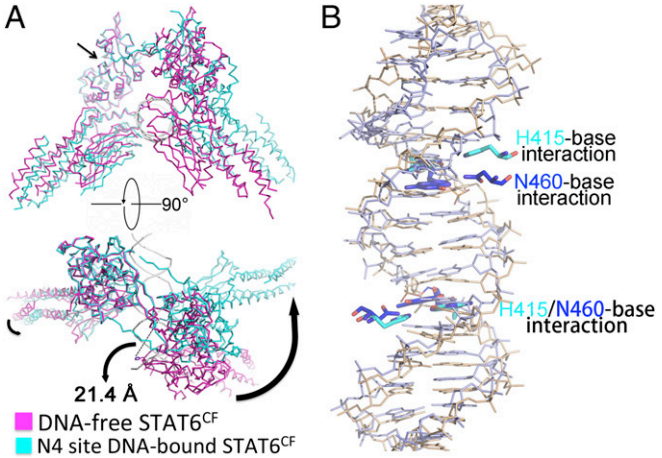

Fig. 3. Conformational change of STAT6^{CF} upon DNA binding. (*A*) Motion of STAT6^{CF} on DNA binding is shown by using SH2 domain as reference (indicated by an arrow in front view). The movements are indicated by bent black arrows in top view. Dash line indicates the movement of key residue H415 in DNA-free and N4 DNA-bound STAT6^{CF}. (*B*) Comparison of N4 DNA-bound STAT6^{CF} (residue and base in cyan) and N3 DNA-bound STAT1^{CF} (residue and base in blue) using the DNA-binding domain as the reference. The differences between N4 and N3 site DNA recognition by STAT proteins are shown.

their interacting bases in respective DNAs, overlap. Interestingly, in the adjacent protomer of the dimer, H415 of STAT6 has moved further inside by a distance of about 4.0 Å than the corresponding equivalent N460 residue in STAT1, which is very close to the regulation of ~3.4-Å rise/bp along the axis of a B-DNA double helix (Fig. 3*B*). A similar change in positions of H415 and its equivalent residue is observed between STAT6^{CF}-N4 and -N3 site DNA complexes (Fig. S5*E*). In addition, we noticed that the residue N417 of STAT6 would sterically clash with the fifth base of DNA, a thiamine, of the complex of STAT1 with the N3 site DNA, which further explains why STAT6 prefers the N4 site DNA but not N3 (Fig. S6*B*). Furthermore, our small angle X-ray scattering (SAXS) analysis of STAT6^{CF} indicates that the conformation of STAT6^{CF} could be stabilized by DNA (Fig. S7). In addition, molecular dynamics (MD) simulations results indicate the decreased structural flexibility of STAT6^{CF} after DNA binding (*SI Results*). Thus, the conformation of DNA-bound STAT6^{CF} is stable compared with the unliganded STAT6^{CF}.

Mutagenic Analysis of the STAT6 DNA-Binding Surface and Interpretation of Disease-Associated Mutations. Using the structures of STAT6^{CF}-DNA as a guide, we probed the role of residues located at the protein-DNA interface in binding DNA by mutagenesis. K284D, K288D, K367/369D, H415A, Q418A, K284A, K288A, and K367/369A mutations were introduced in both STAT6^{CF} and STAT6^{FL}. The in vitro affinity of the mutants for DNA was estimated using ITC (Fig. 4*A*). Most of the mutants of STAT6^{CF} did not bind the CS4 (N4 site DNA), which happens to be the core fragment of DNA used for formation of the STAT6^{CF}-N4 complex for crystallization (20). Only mutants K288A and K367/369A of STAT6^{CF} retained a low affinity of 2.4 and 8 μM, respectively, for the DNA. The WT STAT6^{CF} (STAT6^{CF}-WT) had an affinity of 79 nM for the same DNA when tested under identical conditions (Fig. 4*A*). We further tested the ability of the STAT6 mutants to recognize DNA in vivo using a luciferase reporter-based assay as described previously (5). Consistent with the results of the in vitro experiments performed using STAT6^{CF}, K288A and K367/369A mutants of STAT6^{FL} exhibited significantly reduced ability to activate the reporter gene (Fig. 4*B*) compared with WT STAT6^{FL}. As expected, K284D, K288D, K367/369D, H415A, Q418A, and K284A mutants of STAT6^{FL} almost completely lost their ability to activate the reporter

gene on stimulation by IL-4 (Fig. 4*B*). These mutagenesis results confirmed a role for the residues of the DBD of STAT6 in binding DNA and that the protein-DNA interface observed in the crystal structure is functionally relevant.

We next investigated whether our findings can be used to interpret the disease-associated STAT6 mutations with respect to the STAT6 protein–DNA interaction interface observed in our crystal structures. Recently, several mutations of STAT6 have been identified in follicular lymphoma cases, which include E372K, E377K, D419H, D419A, and D419G (22). Luciferase reporter-based assays and results of quantitative PCR (qPCR) studies have indicated that these mutations could result in an increase in the transactivation of STAT6 (22). Interestingly, mapping of these mutations on our structure of the STAT6^{CF}-N4 complex reveals that they locate on the DNA binding loops (Fig. 4*C*). Switching E or D to K, H, A, or G could decrease the electro-negativity of the DNA binding interface, enhancing the ability of the protein to bind DNA. In addition, we used the cBioPortal (www.cbioportal.org) web tool to analyze large-scale cancer genomics datasets (23). The results of the analysis identified 165 mutations of STAT6 that affected the pathophysiology of cancer. Of these, 52 unique mutations can be mapped on our STA6^{CF} structures. As shown in Fig. 4*D*, the majority of these mappable mutations are located on the STAT6 DBD, which is essential for DNA binding, and the dimerization domain (SH2 domain), which is critical for recognizing N4 site DNA due to its flexible nature. This analysis lends further support to previously published studies on STAT6 and other STAT proteins that recommend targeting of STATs with inhibitors for improving outcomes of therapeutic interventions (24, 25).

Discussion

STAT6 exhibits a preference for N4 site DNA during binding of regions of target genes that regulate expression of the protein. In this context, STAT6 is a unique member of STAT family because the other STATs show a preference for binding N3 site DNA. This study sheds light on the determinants of specificity underlying this preferential binding of STAT6 to N4 site DNA. The first crystal structures of phosphorylated STAT6^{CF}dimer and its two complexes with N3 and N4 site DNA unveil the mode of DNA binding by STAT6. We show that both residue H415 and dimer interface confer specificity on STAT6 for binding N4 site DNA. Thus far, except for STAT6, there are no crystal structures of DNA-free phosphorylated dimeric STATs belonging to the mammalian STAT family deposited in PDB. Results of MD and SAXS analysis indicate that STAT6 retains some degree of flexibility after dimerization. The crystal structure of STAT6^{CF}probably represents a snapshot of one of the dominant conformations assumed by STAT6. These results advance our understanding of the JAK-STAT pathway; in particular, they unveil the molecular mechanisms underlying the recognition and binding of DNAs by an activated phosphorylated STAT6 dimer.

Residue H415 was identified as a critical residue for DNA selection in STAT6. A single mutation, H415N, switched the DNA binding preference of STAT6 from N4 to N3 site DNA. Among the N3 site DNAs tested, the affinity of the H415N mutant for both M67 and T1 increased compared with that of STAT6^{CF}-WT. M67 and T1 have a mismatch (TTCN$_3$TAA) in the palindromic site. Because residue H415 interacts directly with the base G (TTCN$_{3/4}$GAA) of the palindromic site in the structures of our STAT6^{CF}-DNA complexes, STAT6^{CF}-WT probably has a low tolerance for the mismatch and therefore exhibited lower affinity for these kinds of DNAs. On mutation (H415N), the tolerance of STAT6 for the mismatch increased and therefore the affinities for the DNA increased accordingly. Thus, the H415N mutation decreased the affinity of STAT6 for N4 site DNA, but increased the affinity for N3 site DNA with a mismatch in the palindromic site. Interestingly, another reported functional site residue S407 (5), which is buried and not exposed

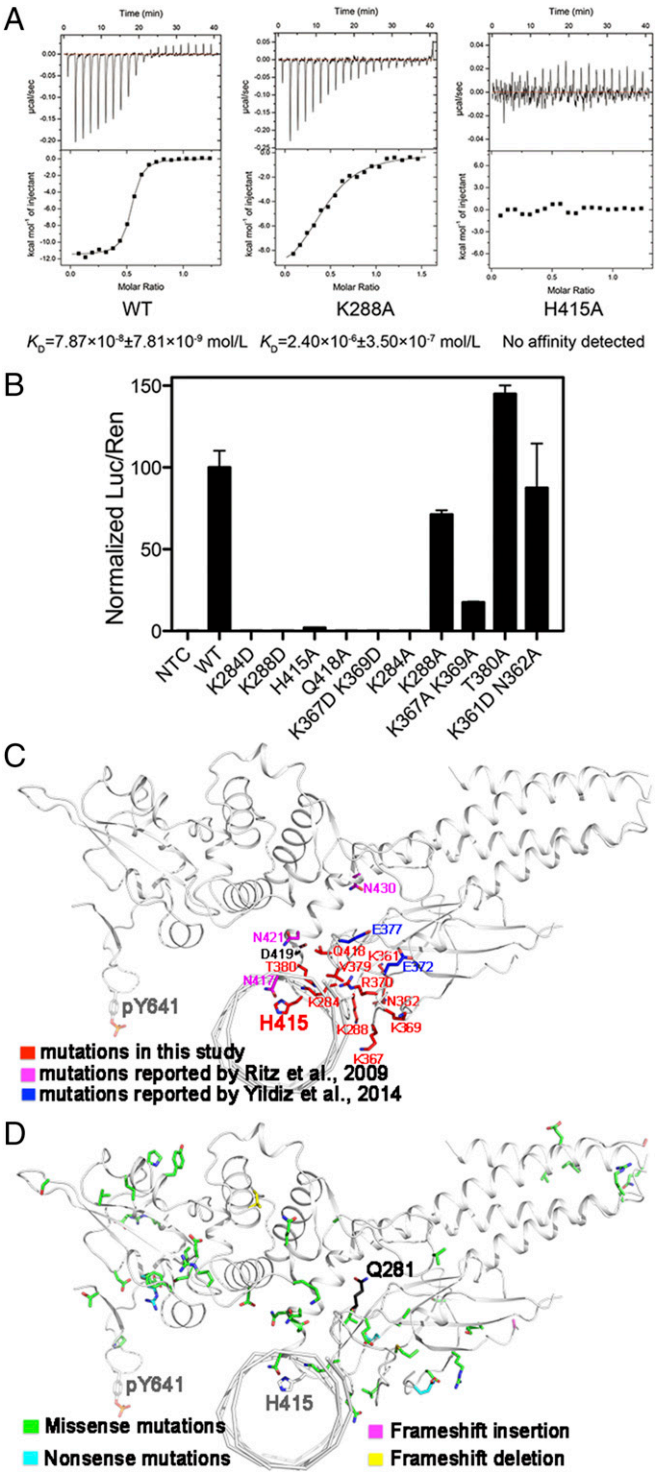

Fig. 4. Identification of key residues for DNA binding by mutagenesis. (*A*) ITC measurements of affinities of STAT6^{CF}-WT (*Left*), STAT6^{CF}-K288A (*Middle*), and STAT6^{CF}-H415A (*Right*) for N4 site DNA (CS4) are shown. (*B*) The ability of several mutants of STAT6^{FL} to activate gene transcription was assessed. Mean ratio luciferase/renilla light units activities are shown for triplicate samples. Normalized results are presented as percent activity relative to the activity in cells transfected with STAT6^{FL}-WT. (*C*) The residues (in red) mentioned above for mutagenesis, mutations reported by Ritz et al. (in magenta) (26) and Yildiz et al. (in blue) (22) were mapped on a protomer of STAT6^{CF}-N4 structure (in front view). The key residue Y641 is also shown. (*D*) Unique mutations from a large-scale cancer genomics dataset analysis in cBioPortal (www.cbioportal.org) web tool are mapped on a protomer of STAT6^{CF}-N4 structure.

on the surface of our structures, is not likely to be accessible for phosphorylation by any kinase (*SI Discussion*, Fig. S8).

In summary, crystal structures of phosphorylated $STAT6^{CF}$ dimer and its two complexes with N4 and N3 site DNAs clarify the differences in DNA binding and substrate specificities between STAT6 and other STATs. A remarkable conformational change is first observed on DNA binding. The dimer interface and residue H415 were identified as important factors for distinguishing N4 from N3 site DNA. These findings enhance our understanding of the STAT–DNA interactions that are crucial for the transcription of specific genes on activation of STATs via phosphorylation. The studies also open up avenues for targeting aberrant STAT6-mediated signaling for development of therapeutics against diseases like asthma and cancer.

Materials and Methods

Detailed discussion of materials and methods is given in *SI Materials and Methods*.

Phosphorylation of the protein was achieved by coexpressing $STAT6^{CF}$ with the tyrosine kinase receptor domain of Elk in the *Escherichia coli* BL21 (DE3) TKB1 strain (Agilent Technology) as described previously (17). Soluble recombinant protein was isolated and purified. The protein were pooled and concentrated for crystallization and other experiments. Unphosphorylated $STAT6^{CF}$ and $STAT1^{CF}$ (aa 132–713) were purified using the same procedures as described for phosphorylated $STAT6^{CF}$. All of the $STAT6^{CF}$-DNA complexes crystals were obtained by incubating the purified phosphorylated $STAT6^{CF}$ with annealed oligonucleotide duplexes at a molar ratio of 1:1.2 for 1 h in an ice bath. Phosphorylated $STAT6^{CF}$ in complex with the 22-bp N4 site duplex and 21-bp N3 site duplex formed crystals that were suitable for data collection. Crystals were frozen before data collection. Datasets were indexed, integrated, and scaled using HKL2000. All three structures were determined by molecular replacement (MR) method. The details of data collection and refinement statistics are listed in Table S3.

ACKNOWLEDGMENTS. We thank the staff at Synchrotron Beamlines (17U1 and 19U of SSRF, 5.0.1 and 12.3.1 of Advanced Light Source, and 23ID-C of the General Medical Sciences and Cancer Institutes Structural Biology Facility–Collaborative Access Team) for technical support and Y. Han, Y. Wang, P. Xue, and Y. Y. Chen (Protein Science Core Facility of Institute of Biophysics) for technical help with initial X-ray diffraction studies, automated protein crystallization, MS, and surface plasmon resonance assay experiments, respectively. This work was supported by National Natural Science Foundation of China Grants 31570875, 31330019, and 81590761; Ministry of Science and Technology of China Grants 2014CB910400 and 2013CB911103); Beijing Nova Program Grant Z141102001814020, and Youth Innovation Promotion Association Chinese Academy of Sciences Grant 2013065 (to S.O.).

1. Abroun S, et al. (2015) STATs: An old story, yet mesmerizing. *Cell J* 17(3):395–411.
2. Lai PS, et al. (2015) A STAT inhibitor patent review: Progress since 2011. *Expert Opin Ther Pat* 25(12):1397–1421.
3. Miklossy G, Hilliard TS, Turkson J (2013) Therapeutic modulators of STAT signalling for human diseases. *Nat Rev Drug Discov* 12(8):611–629.
4. Takeda K, et al. (1996) Essential role of Stat6 in IL-4 signalling. *Nature* 380(6575): 627–630.
5. Chen H, et al. (2011) Activation of STAT6 by STING is critical for antiviral innate immunity. *Cell* 147(2):436–446.
6. Seidel HM, et al. (1995) Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA binding and transcriptional activity. *Proc Natl Acad Sci USA* 92(7):3041–3045.
7. Weirauch MT, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158(6):1431–1443.
8. Moucadel V, Constantinescu SN (2005) Differential STAT5 signaling by ligand-dependent and constitutively active cytokine receptors. *J Biol Chem* 280(14): 13364–13373.
9. Becker S, Groner B, Müller CW (1998) Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* 394(6689):145–151.
10. Chen X, et al. (1998) Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* 93(5):827–839.
11. Nkansah E, et al. (2013) Observation of unphosphorylated STAT3 core protein binding to target dsDNA by PEMSA and X-ray crystallography. *FEBS Lett* 587(7):833–839.
12. Mao X, et al. (2005) Structural bases of unphosphorylated STAT1 association and receptor binding. *Mol Cell* 17(6):761–771.
13. Neculai D, et al. (2005) Structure of the unphosphorylated STAT5a dimer. *J Biol Chem* 280(49):40782–40787.
14. Ren Z, et al. (2008) Crystal structure of unphosphorylated STAT3 core fragment. *Biochem Biophys Res Commun* 374(1):1–5.
15. Vinkemeier U, et al. (1996) DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: Interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. *EMBO J* 15(20):5616–5626.
16. Chatterjee-Kishore M, Wright KL, Ting JP, Stark GR (2000) How Stat1 mediates constitutive gene expression: A complex of unphosphorylated Stat1 and IRF1 supports transcription of the LMP2 gene. *EMBO J* 19(15):4111–4122.
17. Becker S, Corthals GL, Aebersold R, Groner B, Müller CW (1998) Expression of a tyrosine phosphorylated, DNA binding Stat3beta dimer in bacteria. *FEBS Lett* 441(1): 141–147.
18. Baudin F, Müller CW (2013) Bacterial expression, purification, and crystallization of tyrosine phosphorylated STAT proteins. *Methods Mol Biol* 967:301–317.
19. Mikita T, Daniel C, Wu P, Schindler U (1998) Mutational analysis of the STAT6 SH2 domain. *J Biol Chem* 273(28):17634–17642.
20. Ehret GB, et al. (2001) DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. *J Biol Chem* 276(9):6675–6688.
21. Ohmori Y, Hamilton TA (2000) Interleukin-4/STAT6 represses STAT1 and NF-kappa B-dependent transcription through distinct mechanisms. *J Biol Chem* 275(48): 38095–38103.
22. Yildiz M, et al. (2015) Activating STAT6 mutations in follicular lymphoma. *Blood* 125(4):668–679.
23. Cerami E, et al. (2012) The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5):401–404.
24. Furqan M, et al. (2013) STAT inhibitors for cancer therapy. *J Hematol Oncol* 6:90.
25. Mandal PK, et al. (2015) Targeting the Src homology 2 (SH2) domain of signal transducer and activator of transcription 6 (STAT6) with cell-permeable, phosphatase-stable phosphopeptide mimics potently inhibits Tyr641 phosphorylation and transcriptional activity. *J Med Chem* 58(22):8970–8984.
26. Ritz O, et al. (2009) Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. *Blood* 114(6):1236–1242.
27. Stols L, et al. (2007) New vectors for co-expression of proteins: Structure of Bacillus subtilis ScoAB obtained by high-throughput protocols. *Protein Expr Purif* 53(2): 396–403.
28. Niu F, et al. (2013) Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. *Proc Natl Acad Sci USA* 110(22):9054–9059.
29. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromol Crystallogr Pt A* 276:307–326.
30. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
31. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
32. Ouyang S, et al. (2012) Structural analysis of the STING adaptor protein reveals a hydrophobic dimer interface and mode of cyclic di-GMP binding. *Immunity* 36(6): 1073–1086.
33. Hura GL, et al. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* 6(8):606–612.
34. Jiao L, et al. (2013) Structure of severe fever with thrombocytopenia syndrome virus nucleocapsid protein in complex with suramin reveals therapeutic potential. *J Virol* 87(12):6829–6839.
35. Liu T, et al. (2011) Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biol* 12(8):R83.
36. Svergun D, Barberato C, Koch M (1995) CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28(6): 768–773.
37. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst* 42(Pt 2):342–346.
38. Kozin MB, Svergun DI (2001) Automated matching of high-and low-resolution structural models. *J Appl Cryst* 34(1):33–41.