Molecular Cell, Vol. 17, 761–771, March 18, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.molcel.2005.02.021

Structural Bases of Unphosphorylated STAT1 Association and Receptor Binding

Xiang Mao,[1,8] Zhiyong Ren,[1,4,8] Gregory N. Parker,[1] Holger Sondermann,[5] Michael A. Pastorello,[1] Wei Wang,[2] John S. McMurray,[3,4] Borries Demeler,[6] James E. Darnell, Jr.,[7] and Xiaomin Chen[1,4,*]

[1]Department of Biochemistry and Molecular Biology
[2]Department of Experimental Diagnostic Imaging
[3]Department of Neuro-Oncology
M.D. Anderson Cancer Center
The University of Texas
Houston, Texas 77030
[4]Graduate School of Biomedical Sciences
Houston, Texas 77030
[5]Department of Molecular and Cell Biology
University of California, Berkeley
Berkeley, California 94720
[6]Department of Biochemistry
University of Texas Health Science Center
San Antonio, Texas 78229
[7]Laboratory of Molecular Cell Biology
The Rockefeller University
New York, New York 10021

Summary

The crystal structure has been determined at 3.0 Å resolution for an unphosphorylated STAT1 (1–683) complexed with a phosphopeptide derived from the α chain of interferon γ (IFNγ) receptor. Two dimer interfaces are seen, one between the N domains (NDs) (amino acid residues 1–123) and the other between the core fragments (CFs) (residues 132–683). Analyses of the wild-type (wt) and mutant STAT1 proteins by static light scattering, analytical ultracentrifugation, and coimmunoprecipitation suggest that STAT1 is predominantly dimeric prior to activation, and the dimer is mediated by the ND interactions. The connecting region between the ND and the CF is flexible and allows two interconvertable orientations of the CFs, termed "antiparallel" or "parallel," as determined by SH2 domain orientations. Functional implications of these dimer conformations are discussed. Also revealed in this structure is the detailed interaction between STAT1 SH2 domain and its docking site on IFNγ receptor.

Introduction

The family of signal transducers and activators of transcription (STAT) plays an essential role in cytokine and growth factor signaling (Darnell, 1997; Leonard and O'Shea, 1998; Levy and Darnell, 2002; O'Shea et al., 2002; Stark et al., 1998). In untreated cells, the majority of these proteins reside in the cytoplasm in an unphosphorylated form, although recent evidence suggests they may shuttle in and out of the nucleus (Meyer et al., 2002). Upon cytokine and growth factor stimulation of cells, cytoplasmic STAT molecules become tyrosine phosphorylated (activated), dimerize through reciprocal phosphotyrosine (pTyr)-SH2 interactions, accumulate in the nucleus, bind to DNA, and activate gene transcription. Subsequent inactivation involves phospho-STATs coming off DNA, becoming dephosphorylated, and exported to the cytoplasm.

Seven STATs are known in mammals (STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6), and they range from 750–850 amino acids (aa) with similar domain structures. The functional domains of phospho-STAT1 and -STAT3 were identified from mutagenesis experiments and confirmed by the crystal structures (Becker et al., 1998; Chen et al., 1998; Darnell et al., 1994). These domains are called the ND, coiled-coil domain (CCD), DNA binding domain (DBD), linker domain, SH2 domain, and transcriptional activation domain (TAD) (Figure 1A). Aside from the TAD, the overall sequence conservation is high among all of the mammalian STAT molecules. Proteolytic digestion of STAT1 and STAT3 gives two structural regions, the ND (residues 1–130) and the CF (residues 130–700) (Vinkemeier et al., 1996).

The core structures of tyrosine-phosphorylated human STAT1 and mouse STAT3 dimers complexed with DNA have been solved (Becker et al., 1998; Chen et al., 1998) as have phosphorylated, dimeric *Dictyostelium* STATa without DNA (Soler-Lopez et al., 2004). The CF is a rigid structure with the four domains forming a contiguous hydrophobic core. C-terminal to SH2 is a flexible region containing the tyrosine phosphorylation site. In addition, the structure of the isolated dimeric ND of STAT4 was determined (Chen et al., 2003; Vinkemeier et al., 1998). However, no structural studies of an unphosphorylated STAT have been reported.

Unphosphorylated STAT1 was initially reported to be monomeric by glycerol gradient ultracentrifugation and native PAGE (Shuai et al., 1994). Further studies on the STAT proteins prior to tyrosine phosphorylation uncovered dimeric or higher order structures. Unphosphorylated STAT1 was reported to coimmunoprecipitate with STAT2 and STAT3 without cytokine stimulation (Haan et al., 2000; Stancato et al., 1996). Unphosphorylated STAT3 was found to form stable homodimers (Novak et al., 1998) or even higher-order complexes termed "statosomes" (Ndubuisi et al., 1999) by using size exclusion chromatography. STAT1 before activation was mainly dimeric as judged by surface plasmon resonance and other methods (Lackmann et al., 1998). Further, biophysical experiments using purified proteins and cell lysates showed compelling evidence that unphosphorylated STAT1 and STAT3 were mainly dimeric (Braunstein et al., 2003). Isolated STAT N-terminal domains were shown to exist as homodimers in solution (Baden et al., 1998; Byrd et al., 2002; Chen et al., 2003). Murphy and colleagues suggested that the presence of ND dimers in unphosphorylated, full-length proteins could explain dimerization/oligomerization of STAT4 and that for all of the mammalian STATs (except for

*Correspondence: xiaomin_chen@odin.mdacc.tmc.edu
[8]These authors contributed equally to this work.

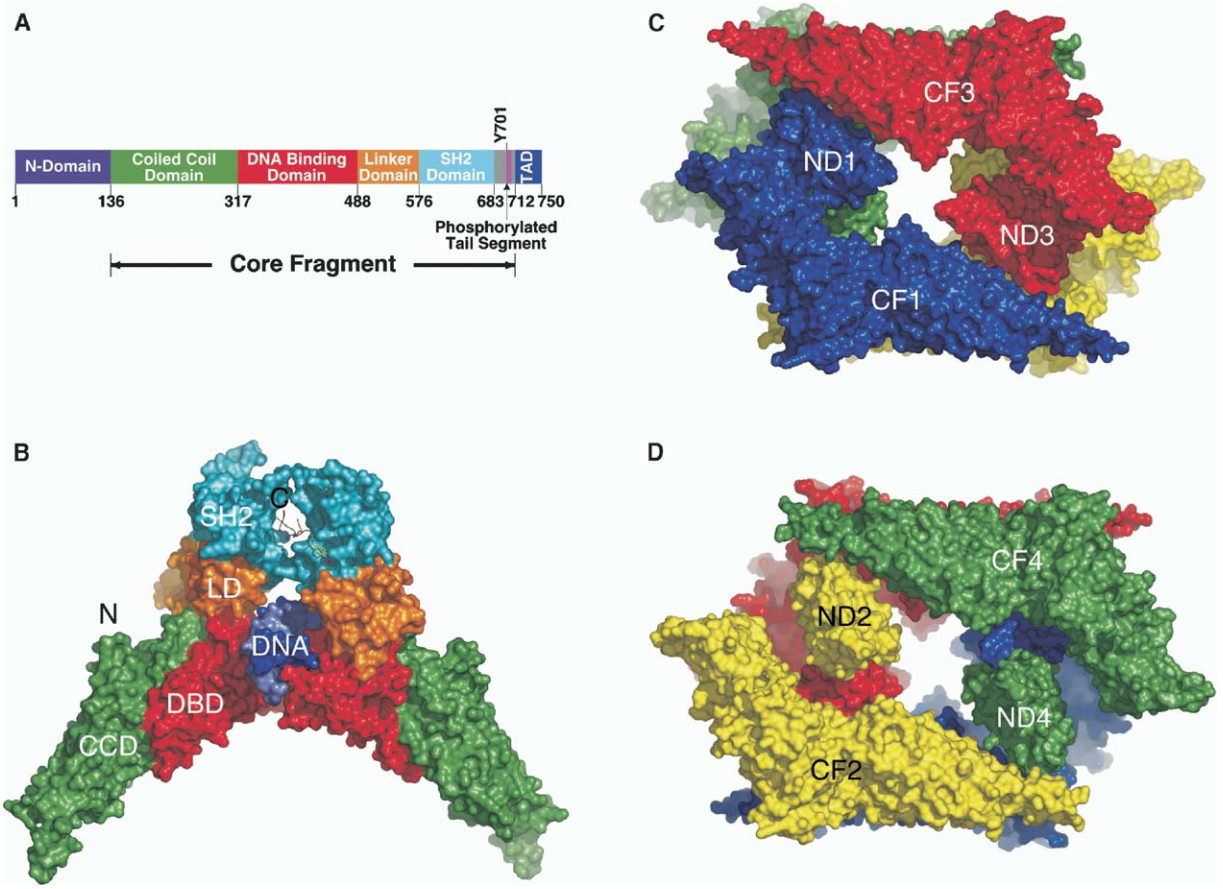

Figure 1. Structures of STAT1

(A) Linear representation of the domains of human STAT1.

(B) Structure of tyrosine-phosphorylated STAT1 core fragment (CF) bound to DNA (Chen et al., 1998).

(C) Structure of the STAT1 (1–683) (front view).

(D) View from the back of the same structure as in (C). All of the structural diagrams throughout the paper were made by using PyMOL (http://pymol.sourceforge.net).

STAT2) the ND-ND interactions were homotypic (Ota et al., 2004). To explore the nature of unphosphorylated STATs, we crystallized an unphosphorylated human STAT1 (1–683). The structure reported here reveals the structural basis of STAT association before tyrosine phosphorylation. The crystal structure also includes a phosphopeptide derived from the STAT1 docking site of human IFNγ receptor α chain and reveals the molecular interactions between STAT1 and the receptor.

Results and Discussion

Structure Determination and General Architecture

The crystal structure of human STAT1 (1–683) construct (molecular weight [MW] = 79.9 kDa) was determined by single anomalous dispersion (SAD) using a gold-derivatized crystal. The structure was refined by using data to 3.0 Å resolution. The $R_{conventional}$ value of this structure is 24.7% and R_{free} is 29.1% (Table 1). There are two molecules in each asymmetric unit. Residues 126–133, 181–188, 414–425, 544–550, and 621–625 in molecule one and residues 123–135, 182–191, 412–422, and 620–622 in molecule two are disordered and not included in the final model. Also, because the C-terminal 29 aa residues (684–712) were not visible in the crystal structure of the STAT1 CF (residues 132–712, X.C., J.E.D., and J. Kuriyan, unpublished data), they were not included in the construct in this study. In the previously determined structure of the phosphorylated dimer bound to DNA (Figure 1B), where the pY701 is stabilized by the SH2 domain and the dimer by reciprocal interaction of the tails, only 11 residues out of 29 of the tail were visible (Chen et al., 1998).

The structure reported here reveals a tetrameric arrangement of the molecules (Figures 1C and 1D). The four NDs exist in two pairs in the central region of the complex and these two pairs of the ND dimers do not interact with each other. The NDs are surrounded by the two pairs of the CFs (one in red and green and the other in blue and yellow). Each CF pair is in a boat-like arrangement. The orientation of these two boats is dictated by the NDs adopting the previously recognized dimer interface (Chen et al., 2003). As a consequence, the front view (arbitrarily defined) of the tetramer (Figure

Table 1. Summary of Crystallographic Analysis

Data Collection Statistics	SAD
Space group	$P6_122$
Unit cell	
A (Å)	102.55
B (Å)	102.55
C (Å)	646.48
α (°)	90
β (°)	90
γ (°)	120
Energy (wavelength)	12398.4 eV (1.0 Å)
Resolution range (Å)	30.0–3.0
Completeness (%)	98.9 (98.9)
I/σI	32.5 (5.0)
R_{sym} (%)	8.3 (50.8)
Figure of merit[a]	0.20 (30–3.0 Å)
Refinement statistics	30–3.0 Å
Reflections (|F|>2σ)	
Working set	65264
Test set	1542
Number of atoms	10863
Rmsd Bonds (Å)	0.009
Rmsd Angles (°)	1.4
$R_{conventional}$[b] (%)	24.7
R_{free}[c] (%)	29.1
Ramachandran plot	
Most favored	81.8%
Additionally allowed	17.3%
Disallowed	0.0%

[a] Figure of merit = $<|\Sigma P(\alpha)e^{i\alpha}/\Sigma|P(\alpha)|>$, where "α" is the phase, and P(α) is the phase probability distribution.
[b] $R_{conventional} = \Sigma\,|F(obs) - F(calc)|/\Sigma\,F(obs)$.
[c] R_{free} was calculated as $R_{conventional}$ by using 2.4% of the data not included in refinement.

1C) is different from the back view (Figure 1D) in that the tunnel going through the tetramer is narrower in the front than in the back.

In these molecules, the connecting region between an ND and a CF is disordered and not visible in the structure. The aa sequence of this region, ^{116}LENAQRFNQAQSGNQSTVMLDKQKEL^{142}, is primarily hydrophilic, consistent with its role as a flexible tether and its sensitivity to proteolysis (Vinkemeier et al., 1996).

The ND-ND Dimer Interface

There are two dimer interfaces seen in the structure, one between two NDs and the other between two CFs. The ND dimer interface is essentially the same as previously proposed for the isolated STAT4 ND dimer (Chen et al., 2003). Figure 2A shows the superimposition of the ND dimer structures from this study and of mouse STAT4. The sequence identity between the two is 63%, and structural alignment using program O (Jones et al., 1991) showed an rmsd of 1.4 Å (over 226 residues) between the Cα atoms of the two polypeptides. The buried surface area of this interface is about 1900 Å^2 (1.4 Å probe). The identity of this interface has been studied extensively by site-directed mutagenesis (Chen et al., 2003; Meyer et al., 2004; Ota et al., 2004), with individual mutations of F77A and L78A having been shown to affect the dimerization of the STAT1 ND (Chen et al., 2003). Similar mutations in STAT4 block dimerization of full-length STAT4α. The effects of these mutations on STAT1 biophysical and biological properties are presented in later sections.

The CF-CF Dimer Interface

Figures 2C and 2D illustrate the second dimer interface between two CFs. This interface has a nice shape complementarity, with the total buried surface area of approximately 2350 Å^2. Residues from the CCD of one molecule form reciprocal interactions with those from the DBD of the other. This interface is predominantly hydrophilic and/or polar, uncharacteristic of a typical protein-protein interface. There are only three hydrophobic residues from each molecule (F172, L383, and V389) with their side chains pointing toward the other molecule. Thus, this interface would likely afford a weak interaction and may only occur transiently unless reinforced by another interaction within the dimer, for example an ND-ND interaction. This dimer interface was independently seen in the crystal structure of an unphosphorylated STAT1 CF (132–712) (X.C., J.E.D., J. Kuriyan, unpublished data).

Residue F172 in the first helix of the CCD is one of the few hydrophobic residues at this interface (Figure 2E). It is inserted into a pocket created by residues Q340, L383, G384, T385, H406, and Q408 of the DBD of its partner molecule. L407, which has been implicated to mediate STAT1 nuclear localization (McBride et al., 2002), also contributes to the pocket, but its side chain points to the protein interior. As discussed later, residue F172 was mutated to a Trp to disrupt the interface.

The monomeric structures of the unphosphorylated and phosphorylated STAT1 CFs are very similar. Figure 2B shows the superimposition of two such fragments (132–683), which have a Cα RMSD of 1.4 Å (program O) over 492 residues. The structural similarity indicates that there is no intramolecular conformational change within each CF before and after tyrosine phosphorylation. This is consistent with the observation that the four domains in the CF form a contiguous hydrophobic core.

STAT Binding to the Receptor Docking Site

Another important feature revealed in the crystal structure is the binding of STAT1 SH2 to the phosphopeptide (^{440}pYDKPH^{444}) derived from the α chain of human IFNγ receptor. As shown in Figure 3A, pY440 of the peptide binds to R602 and K584 as well as the phosphate binding loop (shown in green) of the SH2 domain. In addition to these well-characterized pTyr-SH2 interactions, peptide residues D441 (pY + 1) and H444 (pY + 4), respectively, interact with H629 and Y634 of SH2 through hydrogen bonding. This observation is in perfect agreement with the previous peptide binding studies (Greenlund et al., 1995), which showed that phosphopeptide-SH2 interaction could be primarily attributed to residues pY, pY + 1, and pY + 4 of the peptide.

Figure 3B shows a superimposition of the SH2 domains and the phosphopeptides (^{700}GpYIKTEL^{706}) from the phospho-STAT1-DNA complex structure (Chen et al., 1998) and from the current work. The core secondary structural elements can be superimposed very well,

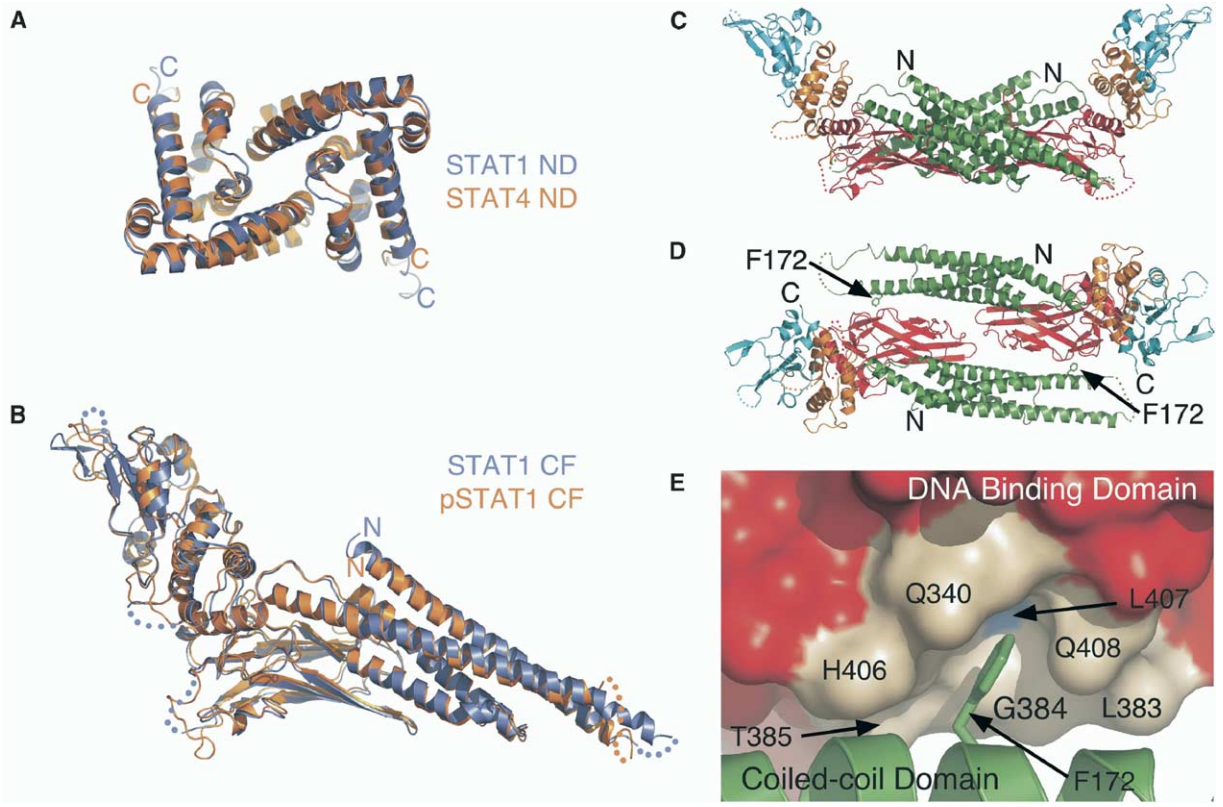

Figure 2. Illustration of the Two Dimer Interfaces

(A) Superimposition of human STAT1 (in slate) and mouse STAT4 (in orange) ND-ND dimers.

(B) Superimposition of monomeric CFs of phosphorylated STAT1 (in orange, Chen et al. [1998]) and unphosphorylated STAT1(1–683) (in slate).

(C) The side view of the CF-CF dimer interface.

(D) The top view of the same interface.

(E) Location of F172 at the CF dimer interface. The molecule where F172 belongs is shown as a ribbon in green, and its partner across the interface is shown as a molecular surface in red. F172 side chain is shown inserted into a pocket in the DNA binding domain of the other molecule. Residues in that area are labeled in wheat (except that L407 is in marine).

although some of the connecting loops are highly flexible and in different orientations. Superimposition based upon the three central β strands and helix αA resulted in the superimposition of the peptides in the orientation shown in the figure. As shown, the pTyr residues from the two peptides are superimposed very closely, but the rest of the peptides adopt very different orientations. Because of the interaction between peptide residue D441 and SH2, the peptide backbone is shifted toward H629 of SH2, causing a sharp bend, as compared to the tail peptide with an isoleucine at this position. The lysine residues at the pY + 2 positions in the two peptides point in opposite directions. The rest of the peptides do not have any sequence homology, and the interactions are different in nature.

Design for Mutational Analyses

As described in the Introduction, various studies have shown the existence in solution of unphosphorylated dimers. To further elucidate the unphosphorylated dimer interactions suggested from the crystal structure, we constructed a series of wild-type (wt) and mutant STAT1 proteins for examination: (1) full-length STAT1α (wt, residues 1–750); (2) F172W, a single mutant F172W in the CF interface of STAT1α; (3) F77A/L78A, a double mutant in the ND interface of STAT1α; and (4) F77A/L78A/F172W, a triple mutant in STAT1α. The ND (residues 1–124 with three extra residues GAH at the N terminus) was prepared separately as was the CF (residues 132–712). The mutants were designed to disrupt either the CF interface (F172W) or the ND dimer interfaces (F77A/L78A). The triple mutant is the combination of the two designed to disrupt both interfaces. As expected, all of the purified, full-length proteins (wt and mutants) showed very similar circular dichroism spectra from 200 to 250 nm (data not shown), indicating that the mutations did not change the overall folding of the protein. All of the constructs except the ND alone could also be tyrosine phosphorylated in vitro by using immunoprecipitated EGF receptor, and the phosphoproteins could bind to DNA with similar affinities (immunoblots and EMSA data not shown). These purified proteins were subjected to static light scattering and analytical ultracentrifugation analyses to determine their apparent MWs and consequently their oligomerization states. Similar constructs with differential affinity tags were

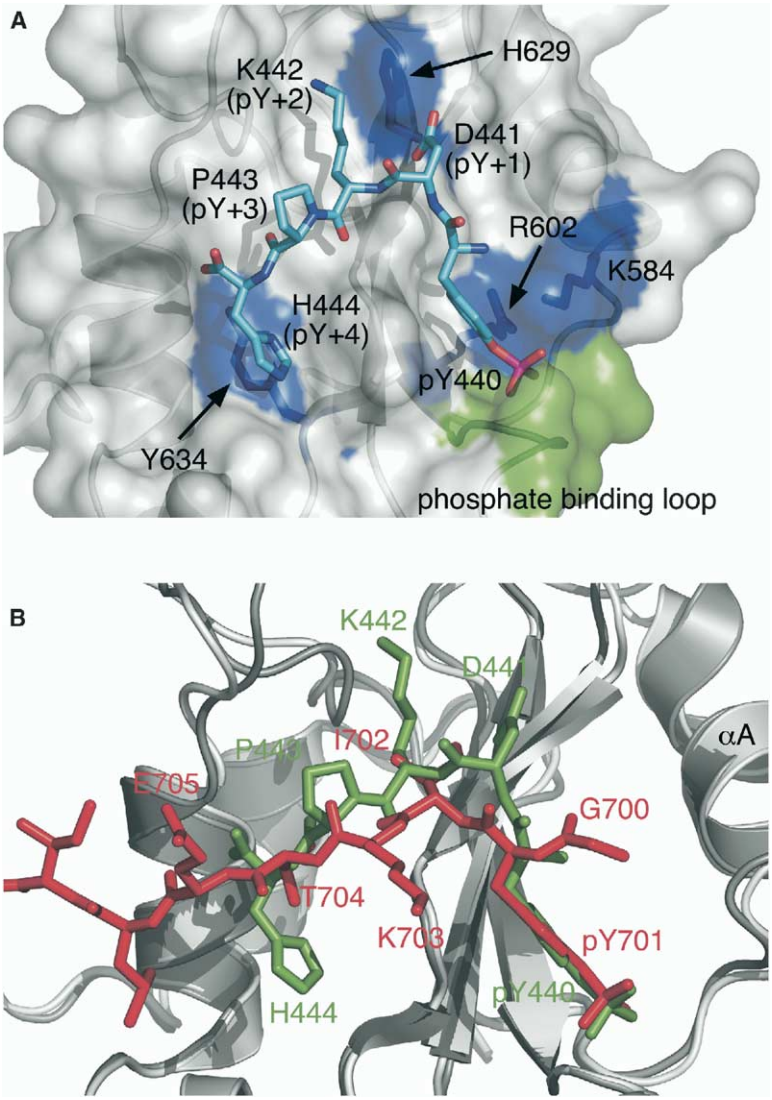

Figure 3. Binding between STAT1 SH2 and Phosphopeptides

(A) SH2 domain is shown as a white and partially transparent molecular surface revealing the secondary structure elements underneath. SH2 residues involved in binding are highlighted in blue (for R602, K584, H629, and Y634) and in green (for the phosphate binding loop). The phosphopeptide is shown "stick" (as defined by PyMOL).

(B) Superimposition of the SH2 domains and the phosphopeptides from previous (in white, Chen et al. [1998]) and current work (in black). The SH2 domains are shown as "cartoon" and the peptides as sticks. The docking site peptide (pYDKPH) is shown in green and the STAT1 tail peptide (GpYIKTEL) in red.

also generated and subjected to coimmunoprecipitation experiments to assess their ability to associate within cells.

Physico-Chemical Analyses

Purified protein samples were analyzed by static light scattering coupled with gel filtration chromatography. Although the concentrations used in these experiments were rather high, the effective concentrations at the point of measurement were significantly lower due to the dilution effect during gel filtration chromatography, with an estimated dilution factor of approximately five. For clarification, all of the concentrations reported in this section (light scattering) are those at sample injection and not at the point of measurement. Table 2 summarizes the results from these experiments.

STAT1α appeared to be almost exclusively dimeric with a measured MW of 174 kDa between 92 and 230 μM. The 67 kDa CF at 45 μM had a measured MW of 72 kDa. At 151 and 271 μM, it had a measured MW of 93 and 109 kDa, respectively, suggesting an equilibrium between monomeric and dimeric forms and a weaker tendency to dimerize than STAT1α. The single mutant (F172W) had an observed MW of 175 kDa at ~200 μM, meaning the molecules were still dimeric. At lower concentrations (23 and 69 μM), this mutant showed marginally lower MWs, indicating perhaps some monomers were present with dimers comprising the majority. The double mutant largely, but not completely, disrupted dimerization. This mutant had a MW ranging from 105 to 124 kDa between 23 and 172 μM. The triple mutant, aiming at disrupting both interfaces, showed only a monomeric MW of 84.4 kDa at ~114 μM.

We next performed analytical ultracentrifugation (sedimentation equilibrium) experiments, and the results are also summarized in Table 2. We found that STAT1α formed relatively tight dimers with a dissociation constant (K_d) of 0.68 μM. The F172W mutant had a higher K_d of 3.9 μM, whereas the isolated ND had a K_d of 6.4 μM, and the isolated CF had an even higher

Table 2. Summary of Static Light Scattering and Analytical Ultracentrifugation Results

		Static Light Scattering			Analytical Ultracentrifugation Monte Carlo Analysis			
	MW_{calc} (kDa)	Concentration at Injection (μM)	MW_{meas} (kDa)	Polydispersity (Mw/Mn)	Monomeric MW_{meas} (kDa)	95% Confidence Range	K_d[a] (μM)	95% Confidence Range
Full-length 1α (1–750)	87.39	91.5	173.3	1.003	87.31	+4.44/–4.02	0.68	+1.03/–0.36
		160	173.9	1.005				
		229	174.1	1.007				
Single mutant (1–750)	87.43	22.9	152.8	1.008	88.93	+12.73/–10.29	3.93	+7.41/–2.57
		68.6	160.9	1.003				
		206	175.7	1.007				
Double mutant (1–750)	87.27	22.9	104.5	1.005	85.17	+0.47/–0.50	NA[b]	NA
		68.8	108.1	1.014				
		172	124.3	1.008				
Triple mutant (1–750)	87.31	114	84.41	1.008	83.09	+0.28/–0.29	NA	NA
Core fragment (132–712)	66.31	45	72.4	1.007	66.99	–2.95/–3.66	21.82	+15.67/–9.12
		151	92.9	1.004				
		271	109.1	1.001				
N domain (1–124, with N-terminal GAH)	15.44	ND[c]	ND	ND	15.07	+0.36/–0.35	6.37	+1.56/–1.25

[a] Dissociation constant of the monomer-dimer equilibrium.
[b] Not applicable.
[c] Not determined.

K_d of 21.8 μM. The data for these four experiments were fitted with a monomer-dimer equilibrium model, and the fitted monomeric MWs are reported in Table 2. The double and triple mutants did not show any detectable dimerization and could be well fitted with a single ideal species suggesting only monomeric species.

These solution study results suggest the greater strength of ND-ND interactions compared to CF-CF interactions in forming dimers but also imply that both interfaces have a role in the dimer interactions of the full-length molecules. In order to further characterize the oligomerization state of STAT1α, we performed interference velocity experiments ranging in concentration between ~10 and ~230 μM (1–20 mg/ml). The results, shown in Figure S3 available with this article online, suggest that at concentrations higher than 10 mg/ml (115 μM), tetrameric forms appear in the sediment between 8 and 9 s. According to the velocity experiments, the predominant form of the STAT1α is dimeric, with some monomer component present at lower concentrations. The dimeric species sediments at ~6 s and the monomeric species between 3–4 s.

Because the IFNγ receptor-derived phosphopeptide was used in crystallization, we checked the effect of this peptide on the oligomerization state of the protein by static light scattering. Three protein samples were used in the analysis, STAT1α (1–750), STAT1β (1–712), and STAT1 (1–683) (the construct used for crystallization) were examined in the presence and absence of the peptide. The results clearly showed that addition of the peptide had no effect on protein dimerization (Table S1).

Immunochemical Analysis

To further explore whether the ND and the CF dimer interfaces play a role in the association of unphosphorylated STATs prior to tyrosine phosphorylation in vivo, we made two C-terminal differentially affinity-tagged (c-Myc and FLAG), full-length STAT1α molecules and expressed them in COS-7 cells. Whole-cell lysates were subjected to immunoprecipitation with anti-c-Myc antibody and the precipitates tested by Western blotting for the presence of the FLAG epitope (Figure 4A, top). When cells were cotransfected with both the c-Myc- and FLAG-tagged wt STAT1, the Myc antibody precipitated both epitopes, indicating an interaction between the FLAG- and Myc-tagged unphosphorylated molecules (lane 5). Controls showed that the Myc antibody could not precipitate the FLAG-tagged protein (lane 3) and that the tagged proteins had to be expressed in the same cells for detection. Further, the epitope-tagged proteins were found to be expressed at similar levels (Figure 4B).

We applied the same analysis to the three mutants F172W, F77A/L78A, and the F172W/F77A/L78A. All three mutations (lanes 6–8, top of Figure 4A) dramatically reduced the amounts of the FLAG-tagged STAT1 brought down by the Myc antibody. These experiments were repeated several times with similar results. Although the amounts of the total immunoprecipitate showed variation among the mutants (Figure 4A, bottom), it is clear that both the F172W and F77A/L78A mutations (either alone or in combination) drastically reduced unphosphorylated STAT1 dimer formation in cell extracts, suggesting the involvement of both interfaces in STAT preassociation.

Two Possible Conformations of the Dimer

From the aforementioned studies, it is clear that STAT1 exists predominantly in a dimeric state in solution that requires the ND-ND interaction. Because the connecting region between the ND and CF is not visible in our crystal structure we have yet to decide which ND and which CF are from the same molecule. Two pairs of the

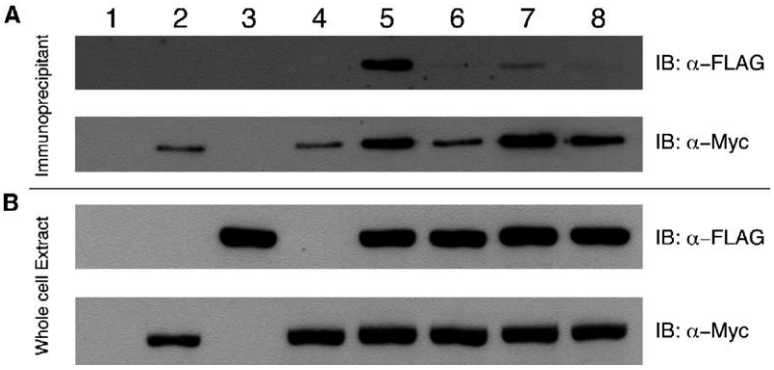

Figure 4. Coimmunoprecipitation of STAT1 Wt and Mutant Proteins with Different Affinity Tags without IFNγ Stimulation

Constructs of human STAT1α with a C-terminal FLAG or c-Myc tag were used to transiently transfect COS-7 cells. Cell lysates were analyzed by immunoprecipitation using anti-Myc antibody followed by Western blotting using anti-FLAG antibody (Figure 3A, top). The same membrane was stripped and reblotted with anti-c-Myc antibody (Figure 3A, bottom). Whole-cell lysates were also directly analyzed by Western blotting using anti-c-Myc antibody (Figure 3B, bottom). The same membrane was stripped and reblotted with anti-FLAG antibody (top). Lane 1, vector alone; lane 2, vector and wt-Myc; lane 3, vector and wt-FLAG; lane 4, wt-Myc and wt; lane 5, wt-Myc and wt-FLAG; lane 6, SM-Myc and SM-FLAG; lane 7, DM-Myc and DM-FLAG; and lane 8, TM-Myc and STAT(TM)-FLAG. Abbreviations: SM, F172 single mutant; DM, F77/L78A double mutant; and TM, F77A/L78A/F172W triple mutant.

ND dimers are symmetry related and we will focus on the one labeled ND2-ND3 in Figure 5A. Based on sidedness and distance considerations, this ND dimer has two ways to connect to the two CFs. The first one is illustrated in Figure 5A, in which ND2 (yellow) connects to CF2 (yellow) and ND3 (red) links to CF1 (blue). Further, these two CFs associate through their reciprocal CF-CF interactions. This particular orientation may be caused by tetramerization under crystallization conditions. Because the connecting regions between ND and CF are very flexible, we think that the ND dimer can adopt flexible orientations relative to the CF dimer. This is physically possible because the Cα distance between the two last anchor residues (L116) in the ND dimer is ~49 Å and that between the two first anchor residues (L142) in the CF is ~46 Å. We shall refer to this conformation as antiparallel because it puts the two pTyr binding SH2 domains on the opposite ends of the dimer (Figure 5C).

The second way to connect the NDs to the CFs is the parallel conformation, which puts the two SH2 domains on the same end of the dimer (Figures 5B and 5C). In this conformation, ND2 (yellow) goes to CF2 (yellow) and ND3 (red) to CF3 (red) without utilizing the CF interface. Although this conformation is weaker than the antiparallel structure due to the missing CF-CF interaction, it may be stabilized by binding to the cytokine receptor.

None of the physico-chemical data obtained can distinguish decisively between the two suggested dimer conformations. Most likely, an equilibrium exists between them (Figure 5C). Such an interconversion is possible because of the strength of the ND-ND interface, the weakness of the CF-CF interface, and the length of the flexible tether between ND and CF of STAT1. This tether allows the CF of the molecule to rotate reversibly from parallel to antiparallel orientations while still maintaining the ND dimers. We speculate that the antiparallel conformation represents the latent state of STAT1 in the cytoplasm before cytokine stimulation. It is the more stable dimer form because both interfaces are utilized. Our data support this notion because disruption of the CF-CF interface by the F172W mutation led to a weaker dimer, compared to wt, by light scattering and analytical ultracentrifugation (Table 2) and a reduced amount of association by coimmunoprecipitation (Figure 4A).

The parallel dimer conformation (Figure 5B) offers a scenario to explain STAT1 binding to IFNγ receptor and the subsequent parallel dimer formation after tyrosine phosphorylation (Figure 6). As revealed in the crystal structures (Randal and Kossiakoff, 2001; Thiel et al., 2000; Walter et al., 1995), two molecules of IFNγ bind to two molecules of the receptor α chain extracellular domain (for simplicity, two IFNγ receptor β chains and the Jak kinases are not shown in Figure 6A). The distance between the two receptor C termini extracellularly is ~90 Å, whereas the cytoplasmic region of the receptor is flexible. Now, imagine a parallel STAT dimer in the cytoplasm coming toward the cell membrane as seen from the receptor (Figure 6B). The two SH2 domains are separated so that there is no physical interaction between the two. The two R602 residues in the pTyr binding sites are both exposed and facing the receptor, and they are about 67 Å apart. The tail segment after the SH2 extends from its parent molecule (as the dotted curve in red or yellow) from the last residue (R683, in white) of SH2 and contains the unphosphorylated tyrosine residue Y701. The intermolecular distance between R683 and the phosphotyrosine binding site is about 40 Å. Given the flexible nature of the tail, the residues between R683 and Y701 can span a distance of 45 Å if fully extended, which is long enough for Y701 to be phosphorylated and in turn be recognized by R602 in the SH2 domain of the other molecule. Such a binding, because of its reciprocal nature, would likely displace the STATs from the receptors and bring the two SH2 domains closer to engage in further interactions (Figure 6C; Chen et al., 1998) that result in a tighter dimer than the one mediated by the NDs alone, breaking the ND-ND dimer interface in the process. Thus, the parallel conformation has the advantage of having the two SH2 domains and tails so positioned that once Y701 is phosphorylated on each tail, they can form the mutual handshake.

Figure 6B represents the unphosphorylated STAT in the cytoplasm, whereas Figure 6C illustrates the phosphorylated STAT in complex with DNA in the nucleus (Chen et al., 1998). The transition between the two states is most likely to be more complicated, and it may

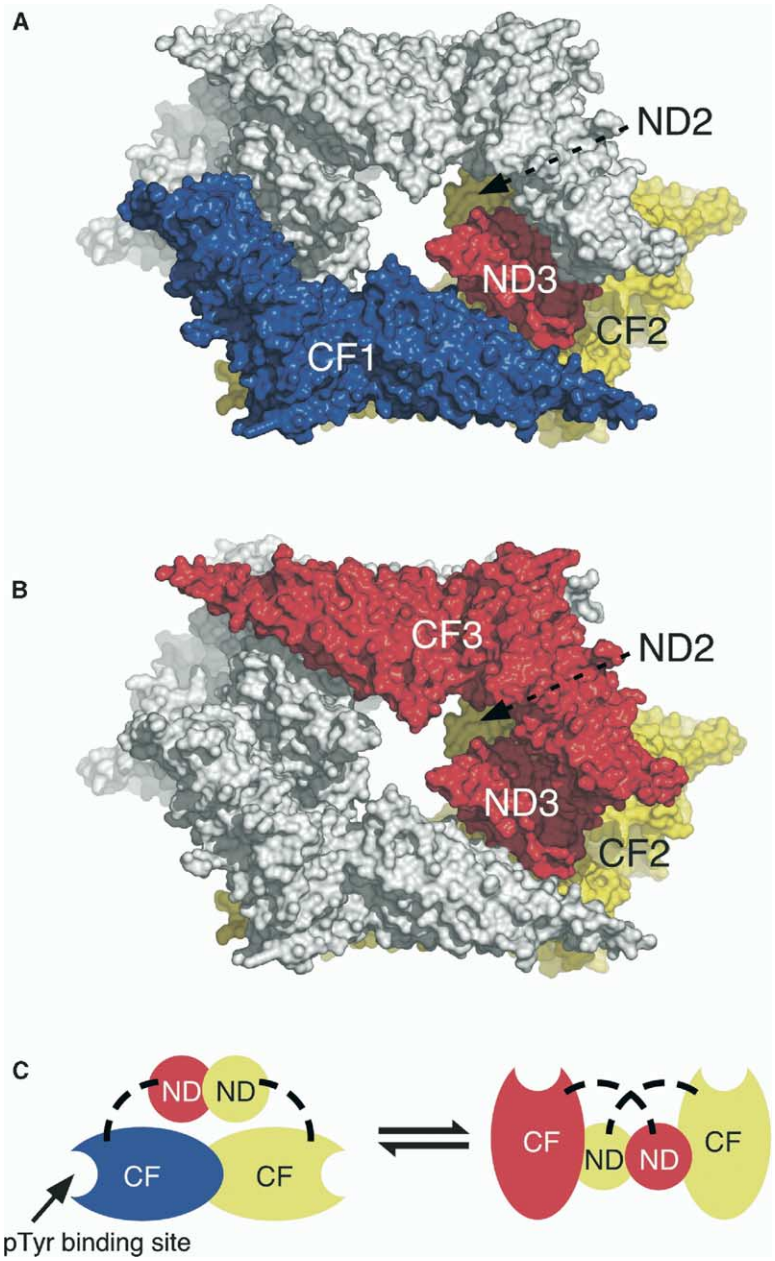

Figure 5. Two Possible Dimer Conformations as Seen in the Crystal Structure

(A and B) In both cases, the ND2 (in yellow) goes with the CF2 (in yellow). In the first case (A), ND3 (in red) connects to the CF1 (in blue), and in the second case (B), ND3 goes to the CF3 (in red).

(C) A cartoon illustrating the equilibrium between two possible dimer conformations: antiparallel and parallel. The location of the pTyr binding site is indicated by an arrow.

involve some intermediate states such as the completely antiparallel arrangement of the two molecules seen in the Dd-STATa structure (Soler-Lopez et al., 2004). One of these intermediates is likely to be recognized and stabilized by the component(s) of the STAT nuclear import machinery.

Conclusions

STATs are engaged in binding to various proteins (such as cytokine receptors, importins and exportins, general transcription machinery, and other transcription factors including other STATs) and DNA. An important feature of STAT structure is that there are two flexible loops: one connecting two rigid structural units (ND and CF), and the other linking CF to the pTyr tail (residues 684–699 in STAT1). These tethers enable the STAT proteins to adopt different conformations before and after tyrosine phosphorylation. The first loop allows the relative orientation of ND and CF to change, whereas the second loop allows the tail to be phosphorylated on the tyrosine and the resultant pTyr to be recognized by the SH2 by a mutual handshake. The ND and CF dimers are both relatively weak (compared to the phosphorylated STAT dimer) so that they can easily dissociate as needed. These structural features ensure maximal freedom of the molecules to achieve rapid and efficient transitions for the functional diversity of STATs. Our current work provides the structural bases of STAT associ-

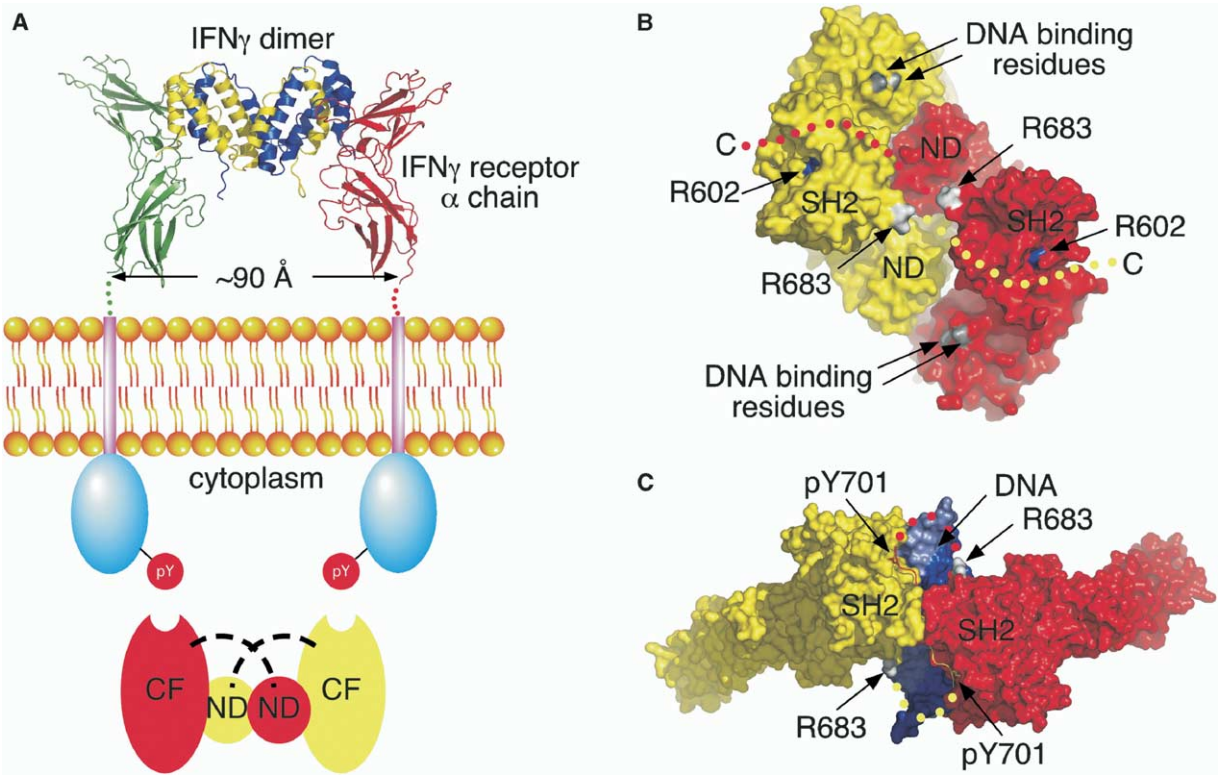

Figure 6. Proposed Model of STAT1 Binding to the IFNγ Receptor α Chain and STAT1 Activation

(A) Illustration of STAT1 binding to the ligand bound IFNγ receptor at cell surface (not to scale).

(B) Proposed view of the ND-mediated parallel dimer from the receptor, focusing on the SH2 domains. The R602 residues (responsible for phosphotyrosine binding) are in blue, the DNA binding residues (K336 and Q460) are in black, and the C-terminal residue R683 residues are in white. The tail segments C-terminal to the SH2 domains are represented as dotted curves in red and yellow.

(C) View of the phosphorylated STAT1 CF in complex with DNA (Chen et al., 1998) in the same orientation. Two strands of the DNA are shown in slate and blue.

ation and receptor binding before tyrosine phosphorylation.

Experimental Procedures

Molecular Cloning, Protein, and Peptide Preparation

Various STAT1 constructs (STAT1α, STAT1β, STAT1 [1–683], and CF) for bacterial overexpression were generated by inserting the corresponding DNA sequences into pET20b(+) (Novagen). The ND was expressed as a C-terminal fusion to GST (Chen et al., 2003). QuikChange site-directed mutagenesis was performed according to the manufacturer's specification (Stratagene). All of the STAT constructs and mutants were DNA sequenced over the entire open reading frame. All STAT1 proteins were purified from *E. coli* essentially as described (Bromberg and Chen, 2001). Purified STAT protein was concentrated, and aliquots were flash frozen in liquid nitrogen and stored at –80°C. Phosphopeptide (acetyl-pY-D-K-P-H-amide) was synthesized by using manual solid phase Fmoc methodology.

Crystallization

The complex of STAT1 (1–683) and the phosphopeptide were prepared by mixing the protein and the peptide at 1:1.5 molar ratio. The crystals were obtained at 4°C by mixing 1 μl of 0.5 mM protein-peptide complex and 1 μl of reservoir solution (100 mM HEPES [pH 7.0–7.1] and 10%–12% PEG 400). The best crystals were obtained in 1–2 weeks. Heavy atom derivatives were obtained by soaking the crystals in stabilization solution (100 mM HEPES [pH7.0–7.1] and 15% PEG 400) with 10 mM $KAu(CN)_2$ for 27 hr.

Data Collection, Structure Determination, and Refinement

The crystals were transferred through cryoprotection and dehydration solutions with increasing concentrations of PEG 4000 (100 mM HEPES, [pH 7.1], 10.5% PEG 400, and 10%–30% PEG 4000). The crystals then were flash frozen at 180 K in a stream of liquid nitrogen. Diffraction data were measured at Advanced Light Source (ALS) beamlines 8.3.1 and 8.2.2 by using ADSC CCD detector Quantum-315 and Cornell High Energy Synchrotron Source (CHESS) beamlines F1 and F2 by using a 2×2 ADSC CCD detector. Data processing and reduction were carried out by using DENZO and SCALEPACK programs (Otwinowski and Minor, 1997). The crystals are in space group $P6_122$ with cell dimensions of a = 102.6, b = 102.6, and c = 646.6 Å, with two molecules of STAT1 protein in an asymmetric unit.

Heavy-atom positions were located and refined by using SOLVE (Terwilliger and Berendzen, 1999), and three gold sites were found. The experimental electron density map calculated using phases derived from SOLVE with density modification by SOLOMON (Abrahams, 1996) was of sufficient quality to allow docking of the structural models of STAT4 ND (Vinkemeier et al., 1998) and of STAT1 CF (Chen et al., 1998) into the electron density. Molecular model was then modified in O (Jones et al., 1991) and refined with CNS (Brunger et al., 1998). The free R value of the model to 3.0 Å is 29.1%, with the conventional R value of 24.7%. The final model has 81.8% of the aa residues in the most favored regions of the

Ramachandran plot. Only nine residues were found in the generously allowed regions and none in the disallowed regions (Table 1).

Multiangle Light Scattering/Size Exclusion Chromatography

Purified protein was characterized by multiangle light scattering after size exclusion chromatography (MALS). Protein was injected onto a KW-803 size exclusion chromatography column (Shodex) equilibrated in MALS buffer (10 mM sodium phosphate [pH 7.5] and 100 mM KCl). The chromatography system was coupled to an 18-angle light scattering detector (DAWN EOS) and refractive index detector (Optilab DSP) (Wyatt Technology). Data were collected every 0.5 s at a flow rate of 0.4 ml/min. Data analysis was carried out by using the program ASTRA, yielding the molar mass and mass distribution (polydispersity) of the sample.

Analytical Ultracentrifugation and Monte Carlo Analysis

Sedimentation experiments were performed with a Beckman Optima XL-A at the Keck Biophysics Facility at Northwestern University and a Beckman Optima XL-I at the Center for Analytical Ultracentrifugation at the University of Texas Health Science Center at San Antonio. Data analyses and Monte Carlo analyses were performed with UltraScan version 6.2 (http://www.ultrascan.uthscsa.edu). Hydrodynamic corrections for buffer conditions were made according to data published (Laue et al., 1992) and as implemented in UltraScan. The partial specific volume of all samples was estimated from the peptide sequence as published (Durchschlag, 1986) and implemented in UltraScan and was found to be 0.73562 ccm/g for the STAT wt, 0.73557 ccm/g for the single mutant, 0.73536 ccm/g for the double mutant, 0.73531 for the triple mutant, 0.72299 ccm/g for the N-terminal domain, and 0.73998 ccm/g for the CF. Equilibrium data were fitted to multiple models. The most appropriate model was chosen based on visual inspection of the residual run patterns and on the best statistics. 95% confidence intervals were determined by Monte Carlo analysis. Sedimentation velocity experiments were analyzed with the enhanced van Holde-Weischet method as implemented in UltraScan (Demeler and van Holde, 2004). All samples were analyzed in a buffer containing 20 mM potassium phosphate, pH 7.0, and 100 mM KCl. Sedimentation equilibrium experiments were performed at 4°C and at multiple speeds corresponding to sigma values (defined as $\sigma = M(1 - \bar{\nu}\rho)\omega 2/2RT$) between one and four. Absorbance samples were spun in six-channel epon/charcoal centerpieces in the AN-60-TI or AN-50-TI rotor. Scans were collected at equilibrium at 230 nm and at 280 nm in radial step mode with 0.001 cm step-size setting and 20-point averages. Multiple loading concentrations ranging between 0.3 and 0.7 OD were measured at the given wavelength, data exceeding 0.9 OD were excluded from the fit. The concentration ranges examined were as follows: STAT1-wt, 0–12 μM; STAT1-F172W, 0–6 μM; STAT1-F77A/L78A, 0–6 μM; STAT1-F77A/L78A/F172W, 0–6 μM; ND, 0–31 μM; and CF, 0–8 μM. Velocity-interference experiments were performed for the wt in 0–250 μM concentration range. In order to compensate for the different absorption properties at different wavelengths, we measured wavelength scans between 220 and 350 nm with 1 nm intervals in triplicate, with 20 repetitions for each data point, by using the analytical ultracentrifuge. The wavelength scans were globally fitted to a sum of Gaussian terms whose width, amplitude, and offset were allowed to float but considered global for all scans. Individual concentrations were adjusted by floating the amplitude of the sum for each scan (data not shown). The extinction profile was normalized with the extinction coefficient at 280 nm, which was estimated from the protein sequence as published (Gill and von Hippel, 1989) and as implemented in UltraScan. The extinction values determined in this fashion were then used to convert optical densities to molar concentration units at the measured wavelengths. The resulting data were fitted to various reversible self-association models. To assure a good signal from both the monomer and dimer species, it is important to conduct the analyses over a large concentration range. This was accomplished by varying the loading concentration and the wavelength of the measurement, which exploit the various absorption properties of the protein. By globally fitting data observed under multiple conditions, such as multiple rotor speeds and multiple loading concentrations, it was possible to enhance the confidence in each fitted parameter value (Johnson et al., 1981). In such a fit, parameters such as monomer MW and association constants were considered global parameters and were forced to be the same for all included datasets. 95% confidence limits for each parameter were determined by Monte Carlo analysis, and the resulting data are listed in Table 2. Plots for the residuals and overlays from the fits to most appropriate models are shown in Figure S1. A plot of the relative dimer concentration distribution at different total concentrations is shown in Figure S2.

Coimmunoprecipitation

Wt and mutant STAT1α with C-terminal affinity tags in pRc/CMV (Invitrogen) were used to transfect COS-7 cells along with the proper controls using Fugene 6 (Roche). 48 hr after transfection, cells were harvested and lysed with IP buffer (100 mM Hepes, [pH 7.4], 150 mM KCl, 1 mM EDTA, and 1% Triton X-100). 2% of the whole-cell lysate was taken out for Western blotting analysis by using anti-FLAG and anti-c-Myc antibodies. The rest of the lysate was subjected to immunoprecipitation with anti-c-Myc antibody. The immunoprecipitates were eluted with a c-Myc peptide (Sigma) and then analyzed by immunoblotting using anti-FLAG antibody. The same membranes were stripped and reblotted with anti-c-Myc antibody.

Supplemental Data

Supplemental Data include three figures and one table and are available with this article online at http://www.molecule.org/cgi/content/full/17/6/761/DC1/.

Acknowledgments

We thank Drs. G. Zhang, J. Marcotrigiano, D. Jeruzalmi, Y. Jiang, and X. Mu for helpful discussions and/or critically reading the manuscript. Crystal structure of the unphosphorylated STAT1 CF was determined while X.C. was in the laboratory of Dr. John Kuriyan at The Rockefeller University. We are grateful to the staff at ALS beamlines 8.3.1 and 8.2.2 and CHESS beamlines F1 and F2 for export technical support. This work is supported by grants from the Robert A. Welch Foundation, National Institutes of Health (X.C.), and National Science Foundation (B.D.).

Received: December 2, 2004
Revised: January 13, 2005
Accepted: February 15, 2005
Published: March 17, 2005

References

Abrahams, J.P. (1996). Methods used in the structure determination of bovine mitochondrial F(1) ATPase. Acta Crystallogr. D Biol. Crystallogr. *52*, 30–42.

Baden, H.A., Sarma, S.P., Kapust, R.B., Byrd, R.A., and Waugh, D.S. (1998). The amino-terminal domain of human STAT4. Overproduction, purification, and biophysical characterization. J. Biol. Chem. *273*, 17109–17114.

Becker, S., Groner, B., and Muller, C.W. (1998). Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature *394*, 145–151.

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

Bromberg, J., and Chen, X. (2001). STAT proteins: signal tranducers and activators of transcription. Methods Enzymol. *333*, 138–151.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Byrd, R.A., Gaponenko, V., Sarma, S.P., Tarasova, N., Li, J.J., and Altieri, A.S. (2002). New insights into dimerization of STAT proteins:

a vehicle to modulate signaling for therapeutic purposes? (Meeting abstract). Proc. Amer. Assoc. Cancer Res. *43*, 139.

Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J.E., Jr., and Kuriyan, J. (1998). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell *93*, 827–839.

Chen, X., Bhandari, R., Vinkemeier, U., Van Den Akker, F., Darnell, J.E., Jr., and Kuriyan, J. (2003). A reinterpretation of the dimerization interface of the N-terminal domains of STATs. Protein Sci. *12*, 361–365.

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

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science *264*, 1415–1421.

Demeler, B., and van Holde, K.E. (2004). Sedimentation velocity analysis of highly heterogeneous systems. Anal. Biochem. *335*, 279–288.

Durchschlag, H. (1986). Specific volumes of biological macromolecules and some other molecules of biological interest. In Thermodynamic Data for Biochemistry and Biotechnology, H.-J. Hinz, ed. (New York: Springer-Verlag), pp. 45–128.

Gill, S.C., and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. *182*, 319–326.

Greenlund, A.C., Morales, M.O., Viviano, B.L., Yan, H., Krolewski, J., and Schreiber, R.D. (1995). Stat recruitment by tyrosine-phosphorylated cytokine receptors: an ordered reversible affinity-driven process. Immunity *2*, 677–687.

Haan, S., Kortylewski, M., Behrmann, I., Muller-Esterl, W., Heinrich, P.C., and Schaper, F. (2000). Cytoplasmic STAT proteins associate prior to activation. Biochem. J. *345*, 417–421.

Johnson, M.L., Correia, J.J., Yphantis, D.A., and Halvorson, H.R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. Biophys. J. *36*, 575–588.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47 (Pt 2)*, 110–119.

Lackmann, M., Harpur, A.G., Oates, A.C., Mann, R.J., Gabriel, A., Meutermans, W., Alewood, P.F., Kerr, I.M., Stark, G.R., and Wilks, A.F. (1998). Biomolecular interaction analysis of IFN gamma-induced signaling events in whole-cell lysates: prevalence of latent STAT1 in high-molecular weight complexes. Growth Factors *16*, 39–51.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J. C. Horton, eds. (Cambridge: Royal Society of Chemistry), pp. 90–125.

Leonard, W.J., and O'Shea, J.J. (1998). Jaks and STATs: biological implications. Annu. Rev. Immunol. *16*, 293–322.

Levy, D.E., and Darnell, J.E., Jr. (2002). Stats: transcriptional control and biological impact. Nat. Rev. Mol. Cell Biol. *3*, 651–662.

McBride, K.M., Banninger, G., McDonald, C., and Reich, N.C. (2002). Regulated nuclear import of the STAT1 transcription factor by direct binding of importin-alpha. EMBO J. *21*, 1754–1763.

Meyer, T., Begitt, A., Lodige, I., van Rossum, M., and Vinkemeier, U. (2002). Constitutive and IFN-gamma-induced nuclear import of STAT1 proceed through independent pathways. EMBO J. *21*, 344–354.

Meyer, T., Hendry, L., Begitt, A., John, S., and Vinkemeier, U. (2004). A single residue modulates tyrosine dephosphorylation, oligomerization, and nuclear accumulation of stat transcription factors. J. Biol. Chem. *279*, 18998–19007.

Ndubuisi, M.I., Guo, G.G., Fried, V.A., Etlinger, J.D., and Sehgal, P.B. (1999). Cellular physiology of STAT3: where's the cytoplasmic monomer? J. Biol. Chem. *274*, 25499–25509.

Novak, U., Ji, H., Kanagasundaram, V., Simpson, R., and Paradiso, L. (1998). STAT3 forms stable homodimers in the presence of divalent cations prior to activation. Biochem. Biophys. Res. Commun. *247*, 558–563.

O'Shea, J.J., Gadina, M., and Schreiber, R.D. (2002). Cytokine signaling in 2002: new surprises in the Jak/Stat pathway. Cell *109*, S121–S131.

Ota, N., Brett, T.J., Murphy, T.L., Fremont, D.H., and Murphy, K.M. (2004). N-domain-dependent nonphosphorylated STAT4 dimers required for cytokine-driven activation. Nat. Immunol. *5*, 208–215.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Randal, M., and Kossiakoff, A.A. (2001). The structure and activity of a monomeric interferon-gamma:alpha-chain receptor signaling complex. Structure (Camb) *9*, 155–163.

Shuai, K., Horvath, C.M., Huang, L.H., Qureshi, S.A., Cowburn, D., and Darnell, J.E., Jr. (1994). Interferon activation of the transcription factor Stat91 involves dimerization through SH2-phosphotyrosyl peptide interactions. Cell *76*, 821–828.

Soler-Lopez, M., Petosa, C., Fukuzawa, M., Ravelli, R., Williams, J.G., and Muller, C.W. (2004). Structure of an activated Dictyostelium STAT in its DNA-unbound form. Mol. Cell *13*, 791–804.

Stancato, L.F., David, M., Carter-Su, C., Larner, A.C., and Pratt, W.B. (1996). Preassociation of STAT1 with STAT2 and STAT3 in separate signalling complexes prior to cytokine stimulation. J. Biol. Chem. *271*, 4134–4137.

Stark, G.R., Kerr, I.M., Williams, B.R., Silverman, R.H., and Schreiber, R.D. (1998). How cells respond to interferons. Annu. Rev. Biochem. *67*, 227–264.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Thiel, D.J., le Du, M.H., Walter, R.L., D'Arcy, A., Chene, C., Fountoulakis, M., Garotta, G., Winkler, F.K., and Ealick, S.E. (2000). Observation of an unexpected third receptor molecule in the crystal structure of human interferon-gamma receptor complex. Structure Fold. Des. *8*, 927–936.

Vinkemeier, U., Cohen, S.L., Moarefi, I., Chait, B.T., Kuriyan, J., and Darnell, J.E., Jr. (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*, 5616–5626.

Vinkemeier, U., Moarefi, I., Darnell, J.E., Jr., and Kuriyan, J. (1998). Structure of the amino-terminal protein interaction domain of STAT-4. Science *279*, 1048–1052.

Walter, M.R., Windsor, W.T., Nagabhushan, T.L., Lundell, D.J., Lunn, C.A., Zauodny, P.J., and Narula, S.K. (1995). Crystal structure of a complex between interferon-gamma and its soluble high-affinity receptor. Nature *376*, 230–235.

Accession Numbers

The coordinates and structure factors of the crystal structure have been deposited in the Protein Data Bank under the accession code 1YVL.