Cell, Vol. 93, 827–839, May 29, 1998, Copyright ©1998 by Cell Press

Crystal Structure of a Tyrosine Phosphorylated STAT-1 Dimer Bound to DNA

Xiaomin Chen,*§ Uwe Vinkemeier,*§
Yanxiang Zhao,* David Jeruzalmi,*
James E. Darnell, Jr.,* and John Kuriyan*†‡
*The Rockefeller University
†Howard Hughes Medical Institute
New York, New York 10021

Summary

The crystal structure of the DNA complex of a STAT-1 homodimer has been determined at 2.9 Å resolution. STAT-1 utilizes a DNA-binding domain with an immunoglobulin fold, similar to that of NFκB and the p53 tumor suppressor protein. The STAT-1 dimer forms a contiguous C-shaped clamp around DNA that is stabilized by reciprocal and highly specific interactions between the SH2 domain of one monomer and the C-terminal segment, phosphorylated on tyrosine, of the other. The phosphotyrosine-binding site of the SH2 domain in each monomer is coupled structurally to the DNA-binding domain, suggesting a potential role for the SH2-phosphotyrosine interaction in the stabilization of DNA interacting elements.

Introduction

It has been established for more than a decade that gene transcription can be initiated within minutes after the activation of cell surface receptors by polypeptide ligands (reviewed in Levy and Darnell, 1990). One of the most direct pathways of polypeptide stimulated gene activity is the so-called Jak-STAT pathway (Ihle et al., 1995; Briscoe et al., 1996; Leaman et al., 1996; Darnell, 1997a). STATs are so named because they serve both as **s**ignal **t**ransducers in the cytoplasm and **a**ctivators of **t**ranscription in the nucleus. Each STAT molecule contains a Src-homology 2 (SH2) domain, a modular unit that binds specifically to phosphotyrosine (Pawson, 1995; Kuriyan and Cowburn, 1997). The STAT SH2 domain acts as a phosphorylation-dependent switch that controls receptor recognition and DNA binding, thus allowing the STATs to couple the activation of cell surface receptors to gene regulation in a direct manner (Darnell, 1997a).

In animal cells, activation of the latent cytoplasmic STAT molecule is accomplished either through cell surface receptors for cytokines and their noncovalently associated Jak kinases, or by growth factor receptors with intrinsic tyrosine kinase activity (Ihle et al., 1995). Binding of the cognate ligand to the cell surface receptor causes the phosphorylation of tyrosines in the cytoplasmic regions of the receptor, thus creating docking sites for the STAT SH2 domain. The consequent recruitment of the STATs to the receptor leads, in turn, to their phosphorylation on tyrosine by the Jak or receptor kinases. The phosphorylated STATs form SH2-mediated dimers and are then translocated to the nucleus, where they bind to DNA and direct specific transcriptional initiation (Darnell, 1997a). STAT-1 and STAT-2 were originally discovered as transcription factors that are activated by interferons α and γ (Fu et al., 1990; Fu et al., 1992; Schindler et al., 1992; Veals et al., 1992). Seven mammalian STAT proteins have been discovered so far, and over 40 different polypeptides are now known to activate one or more STATs (reviewed in Darnell, 1997a).

Sustained efforts at dissecting the STATs into separable domains with distinct functions such as DNA binding have met with limited success. Molecular genetic experiments have, however, implicated specific regions of the protein in specific functions. A single phosphorylation site at Tyr-701 of STAT-1 was identified and proven to be necessary for STAT activity (Shuai et al., 1993). Just upstream from this residue is an SH2 domain, and biochemical experiments indicate that the SH2 domain and the phosphotyrosine in each of two STATs interact in a reciprocal manner to form a dimer (Shuai et al., 1994). The potential DNA-binding region of the STATs was shown to include residues in the 400–500 region (Horvath et al., 1995; Schindler et al., 1995). However, the architecture and mechanism of this DNA-binding region remain unknown.

Regions of STAT that are upstream from the DNA-binding region appear to be involved in protein–protein interactions. An IRF family member, p48, has been shown to interact with a region around Lys-161 in the ISGF3 protein complex (Horvath et al., 1996; Martinez-Moczygemba et al., 1997). Furthermore, CBP interacts with the N-terminal 150 residues (Zhang et al., 1996; J. J. Zhang and J. E. D., unpublished data). The amino-terminal 130 residues form a separable functional domain (N-domain) that strengthens interactions between STAT dimers on adjacent DNA-binding sites (Vinkemeier et al., 1996, 1998; Xu et al., 1996).

A deeper understanding of the mechanism of transcriptional activation by the STATs and the role of tyrosine phosphorylation in controlling this activity has been impeded greatly by the lack of three-dimensional structural information. We now present the crystal structure of a DNA complex of a 67 kDa core fragment of STAT-1, lacking only the N-domain and the C-terminal transcriptional activation domain. The structure lays bare the molecular architecture of the STAT proteins and reveals the mechanism by which the STAT SH2 domain controls dimer formation and DNA binding.

Results and Discussion

Structure Determination and General Architecture

The structure described here is that of STAT-1 core (residues 132–713, M_r = 67.3 kDa), crystallized with an 18-mer duplex DNA containing a binding site for one STAT-1 dimer. The structure determination was carried out by multiple isomorphous replacement (MIR), and the structure has been refined using data to 2.9 Å resolution,

‡To whom correspondence should be addressed.
§These authors contributed equally to this work.

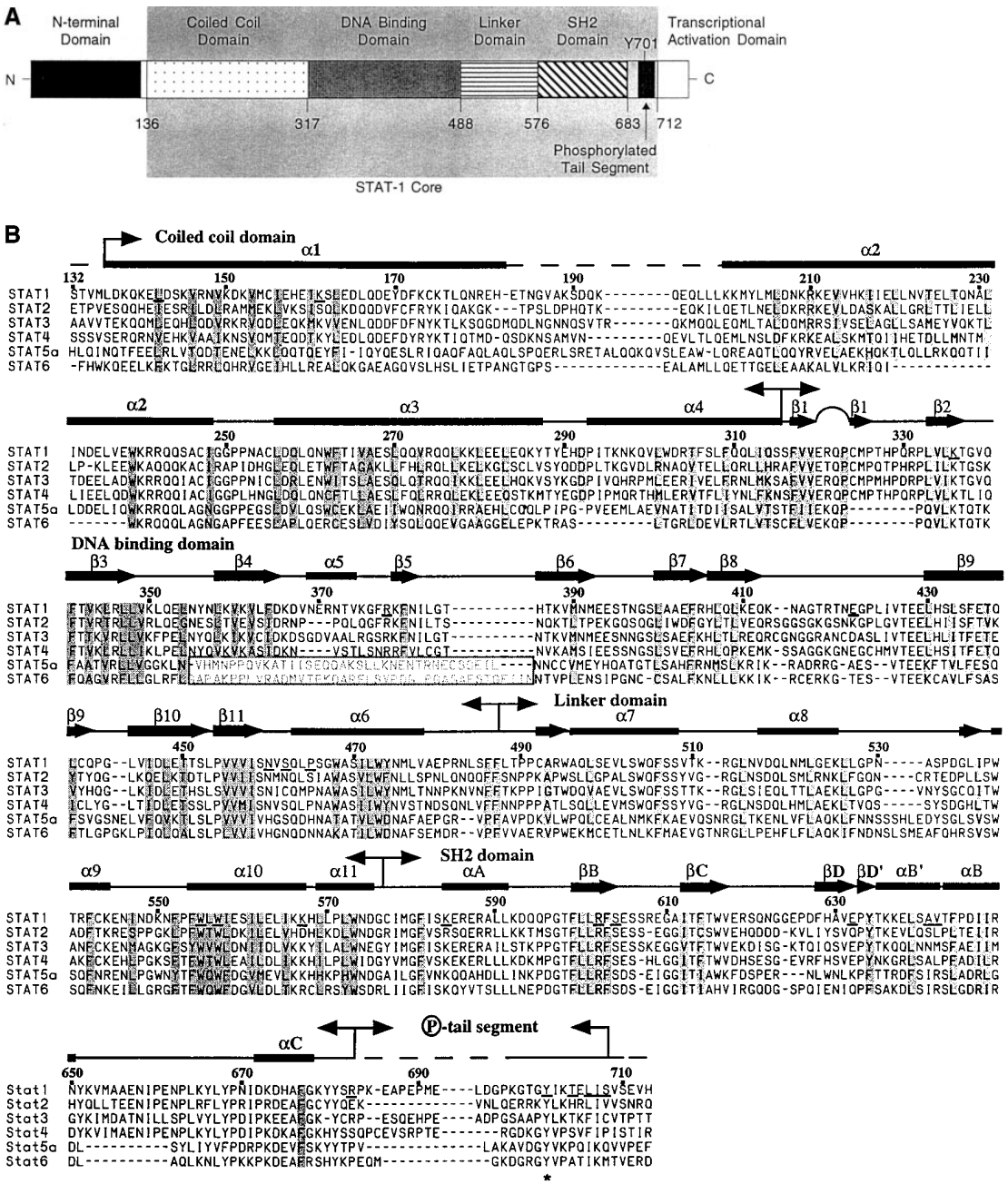


Figure 1. Domain Structure and Sequence Alignment of STATs

(A) Schematic diagram showing the domains of STAT-1.

(B) Sequence alignment of the core regions of human STATs. The secondary structure deduced from the crystal structure is indicated with arrows for β strands and rectangles for α helices. Buried residues that are in the hydrophobic core of the STAT-1 structure are highlighted in gray. A region in the DNA-binding domain of STAT-5 and STAT-6 that cannot be reliably aligned with STAT-1 is boxed. Residues mentioned in the text are underlined. Tyr-701 is marked with an asterisk and disordered loops are indicated by broken lines.

with a free R value of 29.4% and a conventional R value of 22.7%. The crystallographic model contains one STAT-1 molecule per asymmetric unit and includes residues 136–710 of STAT-1.

STAT-1 core contains four tandem structural domains (Figures 1A and 2). The first domain (residues 136–317) consists of four long helices (α1–4), and we shall refer to it as the coiled-coil domain. The DNA-binding domain follows next (residues 318–488) and contains an immunoglobulin-type fold. The next domain links the DNA-binding domain to the SH2 domain, and we shall refer to this as the linker domain (residues 488–576). This region had been predicted to contain an SH3 domain (Fu, 1992), but the all β-sheet architecture of the SH3

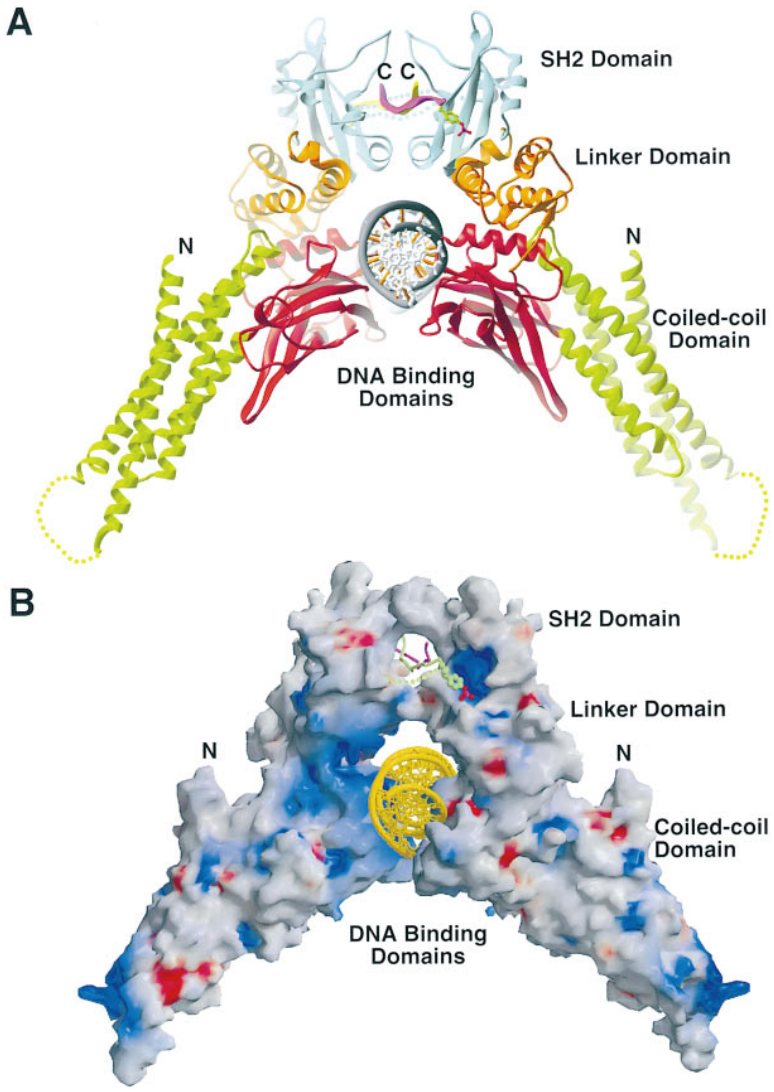


Figure 2. Structure of the STAT-1 DNA Complex

(A) Ribbon diagram of the STAT-1 core dimer on DNA. The component domains are colored green (coiled-coil domain), red (DNA-binding domain), orange (linker domain), cyan (SH2 domain). The tail segments are shown in magenta and yellow. Disordered loops (one in the coiled-coil domain and one connecting the SH2 domain to the tail segment) are shown as dotted lines. The phosphotyrosine residue is shown in a stick representation. The N and C termini of STAT-1 core are indicated by "N" and "C". The DNA backbone is shown in gray. This and other ribbon diagrams were rendered using RIBBONS (Carson, 1991).

(B) Molecular surface of the STAT-1 dimer in the same orientation as (A). The surface was calculated using GRASP (Nicholls et al., 1991) and rendered using RASTER3D (Merritt and Bacon, 1997). The tail segments, shown in green and magenta, were not included in the surface generation. The surface is colored according to the local electrostatic potential, with blue and red representing positive and negative potential, respectively.

domain is clearly missing. The SH2 domain (residues 577–683) is at the C-terminal end of the core structural unit. The C-terminal tail segment (residues 700–708) is phosphorylated on Tyr-701 and is connected to the SH2 domain by a flexible linker of 17 residues. Each of the four domains is fused to the adjacent ones by the formation of a contiguous hydrophobic core. The presence of extensive interdomain interfaces explains why previous efforts at constructing smaller units encompassing the distinct functions of STATs have met with failure.

Two STAT-1 molecules bind to DNA as a dimer, with each monomer in the dimer related to the other by a crystallographic 2-fold axis (Figure 2). The DNA oligonucleotide used in this work contains 18 base pairs that encompass two half-sites, and the spacing of half-sites on the DNA is such that the two DNA-binding domains are on opposite sides of the DNA and do not contact each other. The only protein–protein contacts between the monomers of the dimer occur between the SH2 domains, which exchange C-terminal segments in an intimate interaction. The C-terminal segments extend out of the SH2 domains of each monomer, bind to the SH2 domain of the other monomer, form an antiparallel β sheet arrangement with each other, and then return to make further interactions with the parent SH2 domain. This mutual handshake between SH2 domains seals the STAT dimer onto DNA in a closed embrace (Figure 2B).

The Coiled-Coil Domain

The two coiled domains in the dimer project outward from this C-shaped core in opposite directions and are not involved in interactions with the DNA or with the other monomer in the dimer (Figure 2). The coiled-coil domain has four α helices, two long ones (α1 and α2, 50 residues each) and two shorter ones (α3 and α4, 32 and 23 residues, respectively). The helices form a coiled-coil structure that presents a predominantly hydrophilic surface area for interaction with other proteins (Figure 3). A total of 11 aspartates, 16 glutamates, 7 arginines, 19 lysines, and 4 histidines are on the surface of the structurally defined part of this domain. This suggests

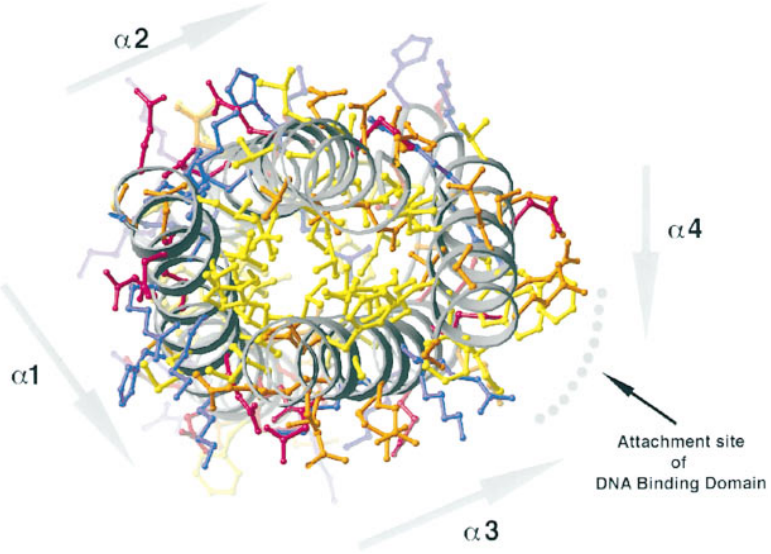


Figure 3. Structure of the Coiled-Coil Domain of STAT-1

The polypeptide backbone of the four helices is shown as a gray ribbon. The directions of the helices are indicated by arrows. All the side chains in the domain are depicted in the figure and are colored red (acidic), blue (basic), orange (polar), and yellow (hydrophobic). Note the clusters of acidic and basic residues on the surface. The only significant cluster of hydrophobic side chains on the surface corresponds to the site of attachment to the DNA-binding domain and is indicated by a dotted line.

that the helices of this domain could participate in interactions with other helical proteins with specificity arising from the interdigitation of complementary charges.

DNA-Binding Mechanism

The general architecture of the domain is that of an immunoglobulin fold (Bork et al., 1994) (Figure 4A). The β strands in the domain mainly run parallel to the major axis of the domain, and this axis is oriented perpendicular to the direction of the DNA axis (Figure 4C). As a consequence, all of the loops at one end of the β sheet arrangement face the DNA, and amino acids in four segments make contacts with DNA (see Figures 4A and 4B for the notation used in this discussion). DNA binding segment 1 includes two loops between β1 and β2, and between β2 and β3. Segment 1 positions Lys-336 in the major groove and makes additional contact with the phosphate backbone of DNA. Segment 2, connecting α5 to β5, is the most distant from the DNA, but the side chain of Arg-378 from this segment extends toward the DNA and makes contact with the phosphate backbone. Segment 3 is a long connector between strands β8 and β9, and it interacts with the minor groove and makes phosphate contacts in the major groove. The most important DNA recognition element is segment 4, the connector between β11 and helix α6 at the C-terminal end of the DNA-binding domain. Asn-460 is positioned deep into the major groove by segment 4, where it makes close contact with base pairs at positions 1 and 2 and can also interact, potentially via water molecules, with the A:T base pair at position 3. Segment 4 is coupled to the phosphotyrosine-binding site via the linker domain, as discussed below.

A crystallographic 2-fold axis of symmetry passes through the center of the oligonucleotide in the crystals of STAT-1 core complexed to DNA (Figure 4B). This 2-fold axis also relates one STAT-1 monomer in the dimer to the other monomer, and thus each STAT-1 monomer in the crystallographic unit is bound to a 2-fold averaged DNA. However, the 18 bp oligonucleotide that resulted in the best crystals of the STAT-1 DNA complex is not dyad symmetric (ACAGTTTCCCGTAAATGC; the core sequence element is underlined and the central C/G is numbered 0; see Figure 4B). This DNA corresponds to the so-called M67 variant of a region of the c-*fos* promoter (Wagner et al., 1990). The M67 site has been used widely in studies on STAT binding to DNA and binds to STAT-1 strongly (Vinkemeier et al., 1996). The lack of dyad symmetry in the M67 site complicates the interpretation of sequence-specific contacts between STAT-1 and DNA in this structure, since nonequivalent base pairs are superimposed at several positions in the crystallographic structure of the DNA.

Despite the asymmetry in the DNA sequence, the structure of the DNA-binding domain and of the DNA is in general very well resolved in the electron density maps (see Experimental Procedures). The temperature factors of atoms in in the DNA binding domain (average value of 37 Å^2) and the DNA (average of 33 Å^2) are among the lowest in the STAT-1 core structure (average of 46 Å^2 over the whole protein, excluding the DNA-binding domain). The heterogeneity in the DNA sequence does appear, however, to be correlated with localized regions of conformational disorder in the protein. For example, the region of segment 3 that contacts DNA in the minor groove is very poorly resolved in the electron density. The most likely explanation of this is that Glu-421, which can interact with the exocyclic amino group of guanine at position 7 in the minor groove, is expelled from the minor groove in the half-sites that contain thymine at this position instead (Figure 4B).

The crystal structure shows that the STAT dimer contacts DNA over a 15 bp region, consistent with studies on the sequence specificities of STAT-1 and -3 (Horvath et al., 1995). The selection experiments for STAT-1 suggest the following consensus sequence for optimal DNA binding: {G/A/C}A{A/C/T}TTCC{C/G}GGAA{G/A/T}TG (the core consensus element is underlined) (Horvath et al., 1995). Selection for C:G or G:C base pairs at the 0, 1, and 2 positions are likely to be mediated by Asn-460 and Lys-336, which make direct (in the case of Asn-460) or potentially water-mediated (in the case of both

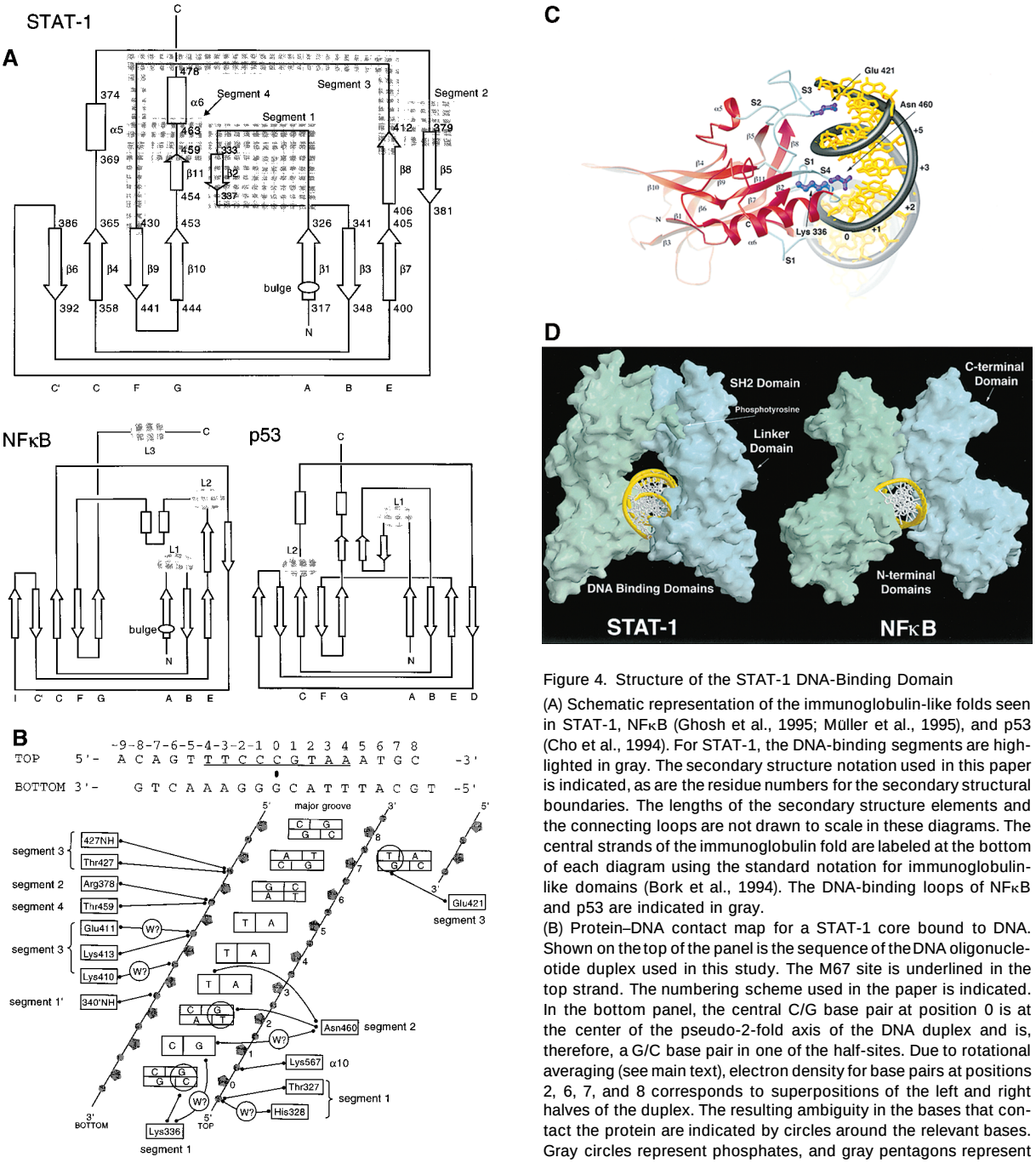


Figure 4. Structure of the STAT-1 DNA-Binding Domain

(A) Schematic representation of the immunoglobulin-like folds seen in STAT-1, NFκB (Ghosh et al., 1995; Müller et al., 1995), and p53 (Cho et al., 1994). For STAT-1, the DNA-binding segments are highlighted in gray. The secondary structure notation used in this paper is indicated, as are the residue numbers for the secondary structural boundaries. The lengths of the secondary structure elements and the connecting loops are not drawn to scale in these diagrams. The central strands of the immunoglobulin fold are labeled at the bottom of each diagram using the standard notation for immunoglobulin-like domains (Bork et al., 1994). The DNA-binding loops of NFκB and p53 are indicated in gray.

(B) Protein–DNA contact map for a STAT-1 core bound to DNA. Shown on the top of the panel is the sequence of the DNA oligonucleotide duplex used in this study. The M67 site is underlined in the top strand. The numbering scheme used in the paper is indicated. In the bottom panel, the central C/G base pair at position 0 is at the center of the pseudo-2-fold axis of the DNA duplex and is, therefore, a G/C base pair in one of the half-sites. Due to rotational averaging (see main text), electron density for base pairs at positions 2, 6, 7, and 8 corresponds to superpositions of the left and right halves of the duplex. The resulting ambiguity in the bases that contact the protein are indicated by circles around the relevant bases. Gray circles represent phosphates, and gray pentagons represent the ribose sugars. The DNA backbones are represented as straight lines connecting phosphates and sugars. Solid lines with black dots on both ends indicate potential hydrogen bonding interactions between protein residues and the DNA. Closed circles with "W?" inside represent possible water-mediated protein–DNA interactions. The indication of potential water-mediated interactions is not based on the direct observation of possible solvent sites in electron density maps, but simply on the distances between the interacting groups and their environment. Note that segment 3, shown to interact in the minor groove, is partially disordered. This may be correlated with heterogeneity at position 7, since guanine is necessary for stabilization of Glu-421 in the minor groove.

(C) A ribbon representation of the structure of the DNA-binding domain is shown in red, with the DNA interacting loops in cyan. The loops are denoted S1 to S4, corresponding to the segments 1 to 4 (see [A]). One of the two rotationally equivalent DNA duplexes is shown. The side chains of Lys-336, Glu-420, and Asn-460 are shown in blue.

(D) Surface representation of the STAT-1 dimer (left) and the NFκB dimer (right). The structure of NFκB shown here is that of Müller et al. (1995). The coiled-coil domains of STAT-1 are not shown.

residues) interactions with these base pairs. The rotamer of the Asn-460 side chain is defined by hydrogen bonding interactions between its terminal oxygen atom and the backbone amide group and the side chain hydroxyl of Ser-462. The nitrogen atom of the asparagine side chain is thus firmly positioned in the major groove, where

it can donate hydrogen bonds to the O6 and N7 atoms of a guanine base of a G:C base pair at position 2.

Selection for the two A:T base pairs at positions 3 and 4 is likely to involve interactions with Asn-460 and may also be an indirect consequence of DNA deformation at these positions. The minor groove is significantly narrowed at these positions (the phosphate–phosphate distances across the groove is ~8 Å, in contrast to ~12 Å in B-form DNA). This deformation may help select for A:T base pairs, and a similar minor groove narrowing and associated selection for A:T basepairs has been noted for the the p53-DNA interaction (Cho et al., 1994) and for NFκB (Müller et al., 1995). Finally, the structure suggests that selection for a G:C base pair at position 7 is likely to involve Glu-421 from segment 3, which can accept hydrogen bonds from guanine in the minor groove.

The general aspects of the interface between the STAT-1 DNA-binding domain and DNA suggest that relatively few direct contacts between STAT-1 side chains and the DNA bases are likely to occur. This is consistent with the pattern of sequences in natural STAT-binding sites, which do not show a sharply defined consensus sequence. Rather, specificity in DNA targeting is likely to arise from interactions between one STAT dimer on DNA and other proteins, particularly other STAT dimers bound to adjacent DNA sites (Vinkemeier et al., 1996; Xu et al., 1996). We discuss the implications of the STAT-1 core structure for STAT dimer–dimer interactions later.

Similarities in the DNA-Binding Domains of STAT, NFκB, and p53

The utilization of immunoglobulin folds for the recognition of DNA was first seen in the tumor suppressor p53 (Cho et al., 1994) and in proteins that contain Rel homology domains, such as the p50 subunit of NFκB (Ghosh et al., 1995; Müller et al., 1995). A search of the protein database using the DALI server (Holm and Sander, 1993) shows that the STAT-1 DNA-binding domain is most closely related in structure to the DNA-binding domains of p50-NFκB and p53. Structural alignments result in rms deviations of Cα positions of 3.0 Å over 106 residues and 3.4 Å over 113 residues for NFκB and p53, respectively.

NFκB and p53 are proteins that are unrelated except for their common immunoglobulin fold. NFκB is a member of the Rel family of transcription factors, and it plays an important role in cellular signal transduction in the immune system (Baeuerle and Henkel, 1994), while p53 is a tumor suppressor gene that is crucial for the control of DNA repair pathways (Friend, 1994). Analysis of the strand connectivity shows that the particular variations on the general immunoglobulin fold that are seen in the STAT-1 DNA-binding domain are similar to variations seen in either p53 or p50-NFκB (Figure 4A). All three proteins bind to DNA using the same face of the immunoglobulin fold, using a similar set of loops. However, there are differences in the lengths and detailed structures of the loops in the three proteins, and, consequently, the orientation of DNA with respect to the protein is different in each of the three cases and the specific DNA sequences that are recognized are unrelated.

Given the close structural similarity between immunoglobin folds in general (Bork et al., 1994), how significant is the structural correspondence between STAT-1, NFκB, and p53? The structural similarity between the three proteins is not reflected at the level of amino acid sequence, which makes it difficult to assign evolutionary significance to these relationships (sequence identity between STAT-1 and NFκB or p53 is 13% and 7%, respectively, for the structurally aligned regions). However, two aspects of the structural comparison are striking and suggest functional correspondences that go well beyond just the utilization of a common fold. These involve comparison of the DNA recognition mechanism of STAT-1 to that of p53 and the mechanism of dimerization of STAT-1 to that of p50-NFκB.

The structural segments that recognize DNA in STAT-1 are remarkably similar in detail to the corresponding elements of p53. A distinctive aspect of the STAT-1–DNA interaction is the positioning of Asn-460 of segment 4 in the major groove of DNA, which is brought about by strand β11 and the C-terminal helix α6 (Figure 4C). An analogous interaction occurs in p53, where a C-terminal α helix is important for positioning residues at the major groove (Cho et al., 1994). Likewise, the interaction of segment 3 of STAT-1 with the minor groove of DNA is mirrored in p53, which also interacts with the minor groove of DNA. Both these interactions are specific to the STAT-1-p53 comparison, since the p50 subunit of NFκB lacks the C-terminal α helix and does not interact directly with the minor groove of DNA.

Dimer formation in both NFκB and STAT-1 results from interactions made by domains other than the DNA-binding domain. The p50 subunit of NFκB contains two domains with immunoglobulin folds (Ghosh et al., 1995; Müller et al., 1995). The larger N-terminal domain makes sequence-specific contacts with the DNA, while the C-terminal domain mediates dimerization and makes contact with the DNA backbone. Superposition of the N-terminal domain of p50-NFκB with the DNA-binding domain of STAT-1 results in an overlay of the C-terminal dimerization domain of NFκB upon the linker and SH2 domains of STAT-1. These two STAT-1 domains are completely unrelated in structure to the C-terminal domain of p50-NFκB, but, like the C-terminal domain of NFκB, they are involved in forming the DNA-bound dimer (Figure 4D).

Comparison of the p50-NFκB and STAT-1 dimers on DNA also emphasizes a key difference in the DNA binding properties of the two molecules. The p50-NFκB homodimer binds DNA tightly, with a dissociation constant in the picomolar range (Baeuerle and Henkel, 1994). In contrast, STAT-1 binds to single DNA-binding sites much more weakly, with a short half-life and dissociation constants in the nanomolar range (Vinkemeier et al., 1996). One structural difference between the NFκB and STAT-1 dimers on DNA is likely to underlie the difference in interaction strengths. The dimerization domain of p50-NFκB makes extensive direct contacts with the phosphate backbone of DNA (Ghosh et al., 1995; Müller et al., 1995) (Figure 4D). In contrast, the structure of the STAT-1 dimer holds the linker and SH2 domains at a

greater distance from the DNA backbone and restricts direct contacts with the DNA to the STAT-1 DNA-binding domain (Figure 4D). The clear separation in STAT-1 of the dimerization region from the region of direct DNA contact might explain the much weaker binding of the STATs to DNA.

The STAT-1 SH2 Domain

The STAT SH2 domains are quite divergent in sequence from most other SH2 domains, perhaps reflecting their appearance early in the evolution of phosphotyrosine signaling in eukaryotic cells (Darnell, 1997b; Kawata et al., 1997). Nevertheless, the basic architecture of the STAT SH2 domain and the mechanism for recognizing the phosphotyrosyl polypeptide are both fundamentally the same as that elucidated for other SH2 domains (reviewed in Kuriyan and Cowburn, 1997). An antiparallel β sheet flanked by two α helices forms the core of the domain, and the phosphorylated tail segment, emanating from the other monomer in the dimer, binds in an extended conformation in a direction orthogonal to that of the strands of the sheet (Figure 5).

A defining aspect of the SH2-phosphotyrosine interaction is the recognition of the phosphate group of the phosphotyrosine by an arginine residue that rises up from the interior of the domain to engage the ligand. This arginine is strictly conserved in all known SH2 domains, and Arg-602 in strand βB of the STAT-1 SH2 domain plays this role (see Figure 5 for the notation used) (Shuai et al., 1993). Residues in the loop connecting strands βB and βC also coordinate the phosphate group, again with close similarity to the well-known mechanism of phosphotyrosine recognition. In most SH2 domains, helix αA provides another arginine side chain that interacts with the phosphate group, the tyrosine ring, and the polypeptide backbone of the ligand (Waksman et al., 1992). This residue is missing in the STAT-1 SH2 domain, and its place is taken by Lys-584, which coordinates only the phosphate group and is conserved in the STATs (Figure 1B).

The conservation in SH2 structure is particularly striking when one considers that only 16 residues are identical over the $\sim$100 residue span of the SH2 domains of STAT-1 and the prototypical one of the v-Src tyrosine kinase (Waksman et al., 1992). An alignment of the three-dimensional structures of the Src and STAT-1 SH2 domains shows that the two chain folds are in register from the N terminus of the Src SH2 domain through to the C-terminal boundary of the domain in v-Src (Figure 5A). The DALI (Holm and Sander, 1993) alignment of the two structures results in an rms deviation of 2.6 Å over 86 aligned residues, with quite limited insertions and deletions in the two sequences (Figure 5A).

Mechanism of SH2-Mediated Dimer Formation

The phosphotyrosine-binding sites are at the distal edges of the inter-SH2 interface, and the C-terminal segment emanating from one SH2 domain has to cross the length of the other one before arriving at the binding site. The linker connecting the last structured residue in the STAT-1 SH2 domain (Arg-683) to pTyr-701 is not visible in the electron density maps. This linker has the sequence ^{684}PKEAPEPMELDGPKGTK^{700}, and the preponderance of prolines, glycines, and hydrophilic residues in this sequence is consistent with its role as a flexible tether that allows the phosphotyrosine to span the 18 Å distance to the binding site on the other SH2 domain. An interdomain exchange of tail segments is enforced by the fact that the phosphotyrosine-binding site on the same SH2 domain is located on the other side of the domain from Arg-683 and is therefore not accessible to the tail segment (Shuai et al., 1994).

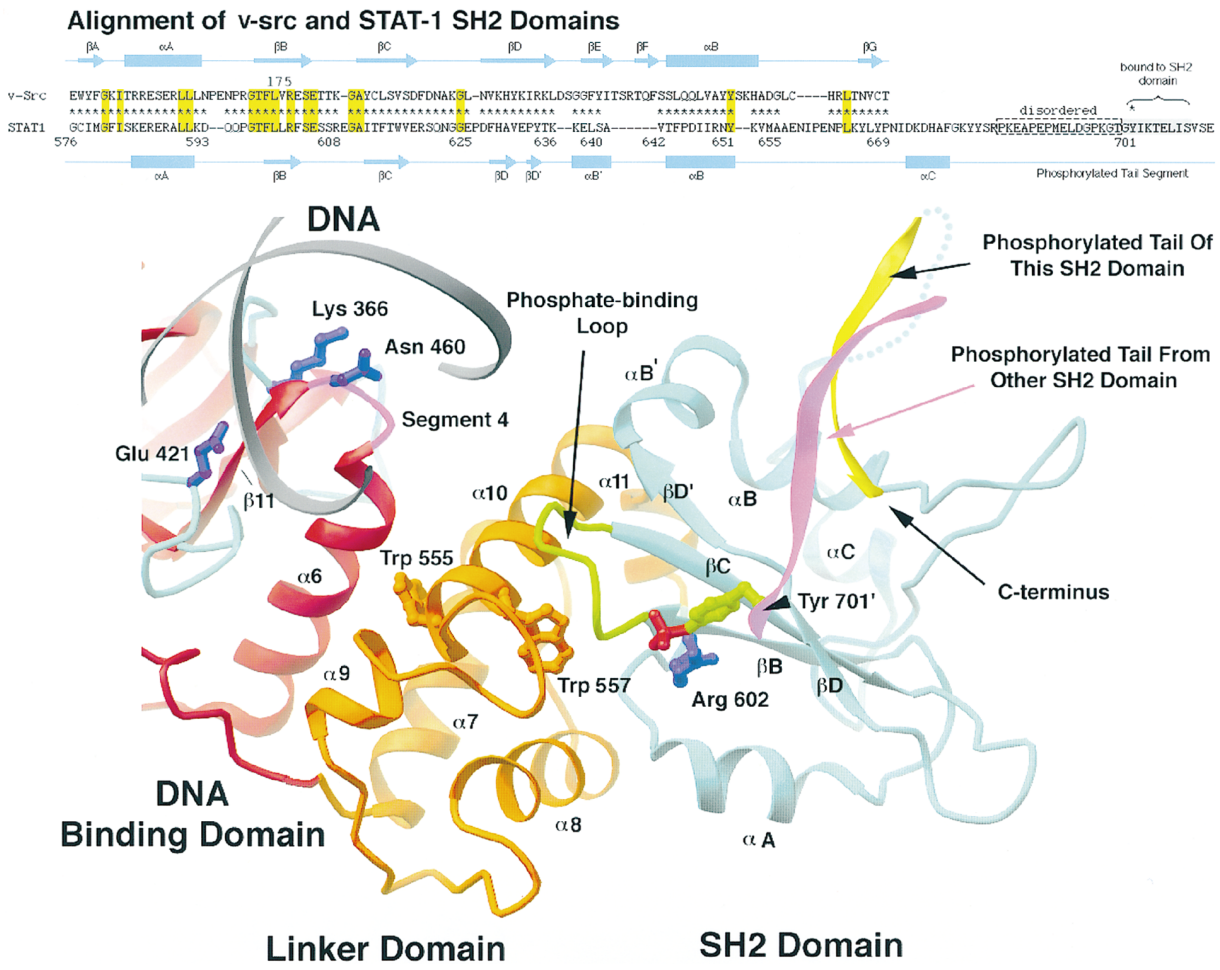
A characteristic aspect of SH2 domains that is preserved in STAT-1 is that interactions between the SH2 domain and its ligand are limited primarily to the residues that are C-terminal to the phosphotyrosine. This feature is a consequence of the geometry of phosphopeptide recognition, which occurs at one edge of the domain, and in STAT-1 it results in the formation of a pair of cross-over connections by the C-terminal segments of the two SH2 domains (Figure 5B). The two tail segments form a two-stranded antiparallel β sheet that passes through a tunnel formed by helices αB and αB′ and the C-terminal extension of the SH2 domain (Figures 2B and 5B). This structural arrangement results in the tail segment being recognized by the SH2 domain over a 7 residue length subsequent to the phosphotyrosine.

The interaction between the two SH2 domains of the STAT-1 dimer is mediated almost exclusively by the phosphorylated tail segment (residues 701–708). This explains the ability of a peptide corresponding to STAT-1 residues 693–707 that is phosphorylated on Tyr-701 to break apart the DNA complexes formed by phosphorylated STAT-1 (Shuai et al., 1994). The residues C-terminal to the phosphotyrosine are bound to the surface of the SH2 domain, with important interactions occuring at the +1, +3, and +5 positions of the tail segment, numbered relative to the phosphotyrosine. The residue at the +5 position is likely to be particularly crucial, because it is at this point that the tail segment enters the tunnel formed by helices αB and αB′ at the base and the C-terminal loop connecting helix αC to the C-terminal end of the SH2 domain at the top (see Figure 5B). Consequently, side chains at +5 (Leu-706), +6 (Ile-707), and +7 (Ser-708) are important mediators of the dimer interaction. In STAT-1, Leu-706 packs into a hydrophobic binding site that is formed by the close apposition of two symmetry related helices αB′ (see Figure 5B). The side chains of Ala-641 and Val-642 (in the SH2 domain) are also brought into close contact at this site in STAT-1. Considerable variation in size and chemical properties are seen at these positions between various STATs (Figure 5B). Subsequent to this point, the tail emerges onto the surface of the parent SH2 domain, where it interacts with the tail from the partner SH2 domain.

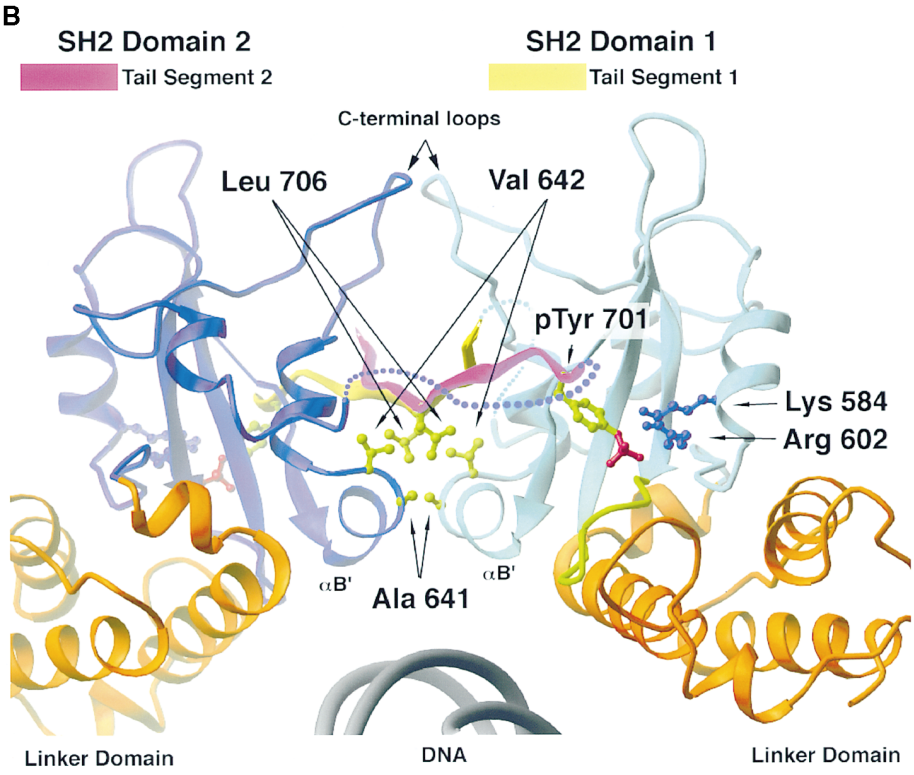
Structural Coupling between the Phosphotyrosine-Binding Site and the DNA-Binding Domain

A notable feature of the STAT-1 SH2 domain is that the phosphate-binding loop of the SH2 domain is buttressed by a number of interactions with elements of the linker domain, particularly with helix α10 (Figure 5A).

A



B



	641	642	706
STAT-1	Ala	Val	Leu
STAT-2	Ser	Leu	Leu
STAT-3	Asn	Met	Phe
STAT-4	Ala	Leu	Phe
STAT-5	Ile	Arg	Ile
STAT-6	Ile	Arg	Ile

A series of hydrophobic side chains presented by helix α10 pack into the hydrophobic core of the SH2 domain, right underneath the phosphate-binding loop. The conformation of one of these, Trp-557, is likely to be sensitive to phosphotyrosine ligation, since its side chain forms a hydrogen bond with the backbone carbonyl of Ser-604, the side chain of which is a phosphate ligand.

What is particularly intriguing about the interactions between the linker domain and the SH2 domain is that while one face of helix α10 interacts with the phosphate binding loop, the other face packs directly against helix α6 and segment 4 of the DNA-binding domain. Trp-555, located immediately before the tryptophan that packs under the phosphate-binding loop (see Figure 5B), is positioned so as to hydrogen bond with the carbonyl group of Pro-465 in helix α6. Changes in the conformation of Trp-555 are likely to be communicated directly to segment 4, which is at the base of this helix.

Segment 4 is the most crucial element of the DNA binding interface, and it is the one that is inserted most deeply into the major groove (Figure 4C). The coupled interactions seen here between segment 4 of the DNA-binding domain and the phosphate-binding loop of the SH2 domain raises the possibility that, in addition to the obvious effect of SH2 ligation upon dimerization, DNA binding in the STATs might be also modulated directly by the SH2 domain. This feature could be an important aspect of the disassembly of DNA-bound STAT complexes by phosphatases.

Implications for STAT-1 Dimer:Dimer Interaction on DNA

It is now clear that STAT proteins can achieve high affinity and specificity in their interactions with DNA by binding cooperatively to DNA sequences containing tandem arrays of multiple binding sites (Vinkemeier et al., 1996; Xu et al., 1996; Meyer et al., 1997). This synergistic recognition of DNA requires the presence of the N-terminal domain of the STATs, which is not required for the binding of STAT-1 core to a single DNA site (Vinkemeier et al., 1996; Xu et al., 1996). We have recently determined the structure of the N-domain of STAT-4, which is highly homologous to that of STAT-1, and have shown that it forms a dimer in the crystal (Vinkemeier et al., 1998). The determination of the structure of the STAT-1 dimer on DNA now allows us to construct a model for cooperative interactions between STATs on DNA.

An oligonucleotide containing two STAT-binding sites that have an 18 bp spacing between their centers (10 bp spacing between the ends of the core binding sites) exhibits cooperative binding by STAT-1 dimers (Vinkemeier et al., 1996). We generated a computer model for this oligonucleotide by taking one DNA duplex from the STAT-1:DNA crystal structure, adding a 2 bp B-form DNA extension to it, and then adding on a second DNA duplex from the crystal structure. A model for two STAT-1 core dimers bound to this oligonucleotide is shown in Figure 6.

Each STAT-1 dimer extends out the coiled-coil domains on either side of the DNA (Figure 6). The spacing of 18 bp between the centers of the two DNA sites results in a rotation of the two dimers with respect to each other in addition to the translation between the sites. The rotational offset between adjacent STAT-1 dimers results in the coiled-coil extensions fanning out around the DNA, much like the blades of a screw propeller. The N-terminal region of the coiled-coil domain is near the base of the domain, in close proximity to the DNA-binding domain. We have docked the N-domain dimer in between the coiled-coil domains of two adjacent STAT-1 dimers such that the C-terminal ends of each of the two monomers in the N-domain dimer are located at a minimal and equal distance (∼27 Å) from the N-terminal ends of two adjacent coiled domains.

Can a ∼27 Å distance be spanned by the linker between the N-domain and the coiled-coil domain? There are 24 residues separating the last hydrophobic anchor residue of the C-terminal α helix in the N-domain (Leu-116 in STAT4, corresponding to Leu-116 in STAT-1) and the first hydrophobic anchor residue in α1 of the coiled-coil domain (Leu-142). In a fully extended conformation, a 24 residue polypeptide can span ∼60 Å. The 24 residues in this region of STAT-1 are predominantly hydrophilic and are likely to be quite flexible in conformation (^{117}ENAQRFNQAQSGNQSTVMLDKQKE^{141}). While the ∼27 Å distance between the N-domains is not beyond the physical limit of extension of the polypeptide chain, in reality we expect that the distance will be reduced significantly by conformational flexibility in the DNA and the protein, which we have ignored in this simple model.

The model suggests that cooperativity in STAT binding to tandem sites on DNA does not result from direct

Figure 5. Structure of the STAT-1 SH2 Domain

(A) SH2 architecture and linkage to the DNA-binding domain. At the top of the panel is shown an alignment of the sequences of the SH2 domains of v-Src and STAT-1. This alignment was generated by the DALI program (Holm and Sander, 1993) based on the three-dimensional structures of the Src SH2 domain (Waksman et al., 1992) and of STAT-1. The asterisks indicate residues that are considered by DALI to be equivalent in three dimensions. The secondary structure elements are indicated using the standard SH2 notation. Identical residues are highlighted in yellow. The v-Src sequence shown spans the entire SH2 domain. Note that the structural conservation is maintained throughout this region of STAT-1. Ribbon diagrams for the SH2 domains, the linker region, and part of the DNA-binding domain are shown below. Two conserved tryptophan residues that pack tightly against the phosphate-binding loop of the SH2 domain (green) and helix α6 of the DNA-binding domain (red) are shown. Segment 4, which is crucial for DNA recognition, is shown in magenta. Also shown in magenta is the tail segment of the second SH2 domain (data not shown) that binds to this SH2 domain via pTyr-701. The tail segment of this SH2 domain, which interacts primarily with the other SH2 domain, is shown in yellow. The flexible connector to the tail is shown as a blue dotted line. The conserved arginine residue in the SH2 domain (Arg-602) and three residues that interact with DNA are shown in blue. The phosphate backbone of the DNA is shown as a gray spiral.

(B) The SH2-dimer interface, colored similarly as in (A). The second SH2 domain is shown, colored blue, while the DNA-binding domain is not shown. Hydrophobic side chains that pack at the dimer interface are shown in green, and residues found at three positions at the interface in STATs 1–6 are shown at the right.

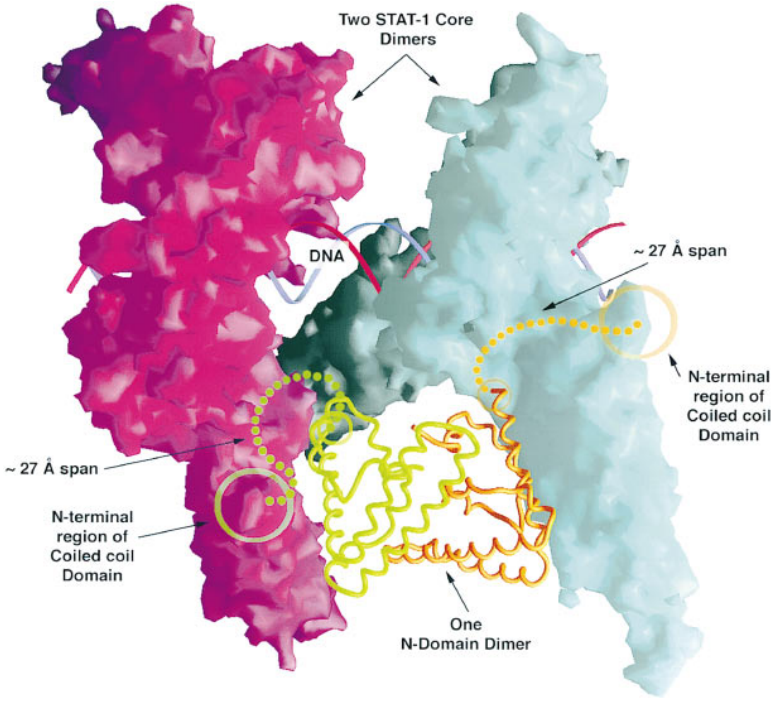


Figure 6. Model for Possible Interaction between Two STAT-1 Dimers on DNA

A DNA duplex containing two STAT-1-binding sites with center to center spacing of 18 bp was generated as described in the text. The DNA backbone is shown as blue and red ribbons. Two STAT-1 dimers (one blue, one purple) are shown bound to these two binding sites, based on the crystal structure of the STAT-1 DNA complex. One of the coiled-coil domains from each dimer is extended toward the viewer, and the location of the N-terminal region of this domain is indicated by orange and green circles. The structure of the STAT-4 N-domain dimer (Vinkemeier et al., 1998) is shown in a ribbon representation, and this dimer has been docked so as to place the last helical residue in each monomer at an equal distance from the N-terminal region of the two coiled-coil domains.

interactions between the core regions of the STAT dimer. The N-terminal region of the coiled-coil domain of each STAT-1 core dimer is positioned so that the loosely tethered N-domains can interact equally well with another STAT dimer that is on one side or the other of the parent dimer. This allows the formation of open-ended

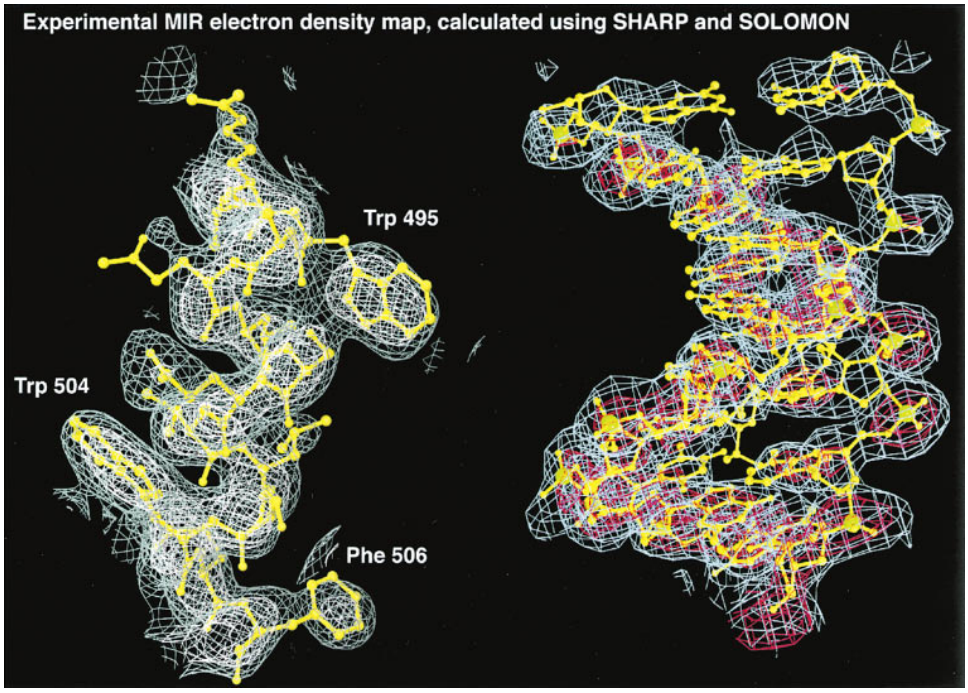


Figure 7. Experimental MIR Electron Density Maps at 3.0 Å Resolution

The maps were calculated using phases from SHARP (La Fortelle and Bricogne, 1997) after density modification by SOLOMON (Abrahams and Leslie, 1996). Left, electron density at the 0.7 σ and 2.0 σ levels are shown for a region in the linker domain. The strong electron density seen here for the backbone and for aromatic side chains is typical for the majority of the protein region and greatly facilitated accurate model building. Right, electron density at 0.7 σ and 2.5 σ levels for the DNA. One half-site is shown. This figure was composed in BOBSCRIPT (Esnouf, 1997) and rendered with RASTER3D (Merritt and Bacon, 1997).

Table 1. Summary of Crystallographic Analysis

Multiple Isomorphous Replacement		Native	Na_2OsCl_6	$KAu(CN)_2$	$Pb(OAc)_2$	$UO_2(NO_3)_2$
Resolution (Å)		30.0–2.9	30.0–3.0	30.0–3.0	30.0–3.0	30.0–3.0
Number of sites		-	3	2	3	3
Reflections	measured (unique)	261200 (23301)	174926 (21229)	324155 (21315)	209517 (21008)	158752 (19573)
R_{sym} (%)[a,f]	overall (outer shell)	7.5 (19.8)	7.7 (19.2)	9.4 (19.4)	7.9 (17.3)	9.9 (14.2)
Completeness ($I>1\sigma(I)$)	overall (outer shell)	89.8 (72.1)	79.9 (63.9)	77.4 (61.8)	73.6 (80.9)	77.0 (74.8)
$I/\sigma(I)$	overall (outer shell)	10.2 (4.9)	12.4 (5.0)	10.5 (4.9)	7.7 (3.4)	8.2 (5.1)
R_{iso} (%)[b]	overall (outer shell)	-	18.0 (22.1)	15.1 (24.8)	15.7 (18.8)	42.1 (35.9)
Phasing power[c]	centric/acentric	-	0.89/1.33	1.05/1.29	0.78/0.89[g]	0.63/0.69
Overall figure of merit[d]	centric/acentric			0.32/0.28		

Refinement	Resolution Range	Number of Reflection ($IF|>2\sigma$)	Total Number of Atoms	$R_{working}/R_{free}$[e] (%)	Rms Deviations		
					Bonds (Å)	Angles (Deg.)	B Values (Å^2)
	30.0–2.9	20526	5126	22.7/29.4	0.016	2.25	2.99

[a] $R_{sym}\% = 100 \times \Sigma|I-<I>|\Sigma I$, where I is the integrated intensity of a given reflection.
[b] $R_{iso}\% = 100 \times \Sigma|F_{PH}-F_P|/\Sigma F_P$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively.
[c] Phasing power $= \Sigma|F_{PH(calc)}|^2/\Sigma(|F_{PH\,(obs)}-F_{P(calc)}|^2)^{1/2}$.
[d] Figure of merit $= <|\Sigma\, P(\alpha)\, e^{i\alpha}/\Sigma|P(\alpha)|>$, where α is the phase and $P(\alpha)$ is the phase probability distribution.
[e] Free R factor was calculated with 10% of the data.
[f] The outershell for the native data is 3.0–2.9 Å, and that for all of the derivatives is 3.11–3.0 Å.
[g] The overall completion in this dataset is lower than that for the highest resolution shell because of the removal of data in resolution shells corresponding to ice rings.

complexes of STAT dimers on DNA, without particularly stringent requirements for site-to-site spacing.

Conclusions

The STATs utilize an immunoglobulin fold to bind DNA, much like NFκB and p53. It is striking that the STATs and NFκB, two of the limited number of families of latent cytoplasmic transcription factors that are translocated to the nucleus upon activation, both use similar DNA-binding motifs. There are, however, basic differences in their mechanism. Sequestration of NFκB in the cytoplasm is achieved by binding to an inhibitor that upon release reveals a nuclear localization signal, and nuclear translocation of NFκB follows. No nuclear localization signal for the STATs has been identified as yet, and the cytoplasmic unphosphorylated STATs are not bound to inhibitors. Rather, STAT activation requires tyrosine phosphorylation and dimerization, which somehow triggers nuclear translocation.

The unique feature of the STATs is the presence of the SH2 domain, which is fused into a contiguous structural element that includes the DNA-binding domain. A STAT homolog has been found in the slime mold *Dictyostelium discoideum* (Kawata et al., 1997), suggesting a very ancient evolutionary origin for the utilization of the immunoglobulin fold to bind DNA as well as for the interaction of SH2 domains with phosphotyrosines. The crystal structure of the STAT-1 DNA complex described here reveals that dimeric interactions between two SH2 domains are crucial to the formation of a DNA-binding clamp that wraps almost entirely around the duplex. By limiting the dimer interaction to the SH2 domain, the STATs ensure that dephosphorylation of the tail segment will result in the rapid dissociation of the STAT–DNA complex.

The SH2 domains first gained prominence because of their capacity to act as independently folded modular peptide binding units (Pawson, 1995). This concept holds true for the STATs in that the phosphorylated tail segment interacts only with the SH2 domain and does so in a conventional manner. However, the structure of STAT-1 makes clear that the SH2 module functions as a tightly integrated component of a complex signaling mechanism. This is reminiscent of the situation in the Src tyrosine kinases, which utilize the same conserved phosphopeptide-binding mechanism of the SH2 domain to coordinate an internal ligand, resulting in the inactivation of the enzyme via a subtle mechanism (Sicheri and Kuriyan, 1997). It would appear that as we learn more about how eukaryotic signaling switches operate at the molecular level, we shall have to expand the simpler modular models to include more integrative mechanisms of action.

Experimental Procedures

Protein and DNA Preparation and Crystallization

Human STAT-1 core protein (residues 132–713) was overexpressed in *E. coli* and purified essentially as described (Vinkemeier et al., 1996). Oligonucleotides were synthesized by standard phosphoramidite chemistry on an Expedite Nucleic Acid Synthesis System (PerSeptive) and purified by preparative, denaturing polyacrylamide gel electrophoresis. Purified oligonucleotides were extracted from gel slices using electroelution (Elutrap System, Schleicher and Schuell, Inc.) and desalted using a Resource RPC reverse phase column (Pharmacia) on HPLC at room temperature. Single-stranded DNA was quantified by UV spectrophotometry, mixed with an equimolar amount of a complementary strand, and annealed in the presence of 100 mM KCl and 10 mM $MgCl_2$.

Protein–DNA complex was prepared by mixing the protein and DNA samples with a molar ratio of 1:1.04 (protein dimer:DNA duplex). Crystals were obtained from a variety of oligonucleotide duplexes, but suitable diffraction was only obtained from crystals of the 18-mer duplex shown in Figure 4B. One large crystal grew over several months at 4°C from a hanging drop that had been set up by mixing 1 μl of 0.12 mM protein:DNA complex and 1 μl of the reservoir solution containing 100 mM Na acetate (pH 5.0), 100 mM KCl, 20 mM $MgCl_2$, 3% PEG400, and 0.01% NaN_3. Crystals grown by macroseeding, originally from this crystal, reach a size of at least 0.25 ×

$0.2 \times 0.1\ mm^3$ within 10 days. The crystals are in space group $C222_1$ with cell dimensions of a = 76.6, b = 148.2, and c = 181.1 Å, with one molecule of STAT-1 protein and a DNA half-site in the asymmetric unit.

Heavy atom derivatives were obtained by soaking crystals in stabilization solution with 1 mM Na_2OsCl_6 for 12 hr, with 10 mM $KAu(CN)_2$ for 12 hr, with 10 mM $Pb(OAc)_2$ for 1 hr, and with 10 mM $UO_2(NO_3)_2$ for 4 hr. The crystals were frozen in freshly thawed liquid propane (temperature $\sim -150°C$) after being serially transferred through the cryoprotection solutions with increasing concentrations of PEG400 (15%–45%). Diffraction data were measured at beamline A1 of Cornell High Energy Synchrotron Source using a CCD detector. Data processing and reduction was carried out using programs DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Electron density maps calculated using phases derived from MLPHARE (Collaborative Computational Project, 1994) with density modification by SOLOMON (Abrahams and Leslie, 1996) were of insufficient quality for model building. However, use of SHARP (La Fortelle and Bricogne, 1997) and SOLOMON gave a map at 3.0 Å of excellent quality (Figure 7; Table 1). Molecular models were built into this map using O (Jones et al., 1991) and refined with CNS, using a maximum likelihood residual (Brünger et al., 1998). A complete molecular model for the DNA was built on the basis of the experimental electron density map. Duplex DNA containing 17 base pairs and 1 overhang at each end was built to correspond to the sequence shown in Figure 4B. A crystallographic 2-fold axis passes through the central base pair of this duplex, as explained in the main text. The occupancy of the DNA was held fixed at 0.5 during the refinement to account for the conformational averaging inherent in the symmetry of the crystallographic system. Almost all of the molecular model for the protein could also be built into the original experimental map, and the map continued to provide valuable guidance until the very end of the model refinement.

The final model for the protein extends from residue 136 to residue 710 of STAT-1. Two loops in the structure are disordered, and these span residues 183–196 in the coiled-coil domain and residues 684–699 in the SH2 domain. The free R value of the model to 2.9 Å is 29.4%, with a working R value of 22.7%. The final model has 79% of the amino acid residues in the most favored regions of the Ramachandran plot. Only six residues are found in generously allowed regions, with none in the disallowed regions.

The coordinates will be deposited in the Protein Databank and are available immediately at the web site http://www.rockefeller.edu/kuriyan/.

Acknowledgments

We thank Gerard Bricogne for providing the SHARP program and for advice. We thank Stephen Burley, Jeff Bonanno, and Elena Conti for assistance and advice, Yossi Schlessinger for providing the antibody to the EGF receptor, and the staff at CHESS for support at beamline A1. The expert technical assistance provided by Stephen Jacques, Lore Leighton, Marc Uy, and Huguette Viguet is gratefully acknowledged. X. C. was supported by a Fellowship from the Arthritis Foundation. Macromolecular crystallography at CHESS is supported by the NSF and the NIH.

Received May 12, 1998; revised May 14, 1998.

References

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F^1 ATPase. Acta Crystollagr. D. *52*, 30–42.

Baeuerle, P.A., and Henkel, T. (1994). Function and activation of NF-κB in the immune system. Annu. Rev. Immunol. *12*, 141–79.

Bork, P., Holm, L., and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J. Mol. Biol. *242*, 309–320.

Briscoe, J., Guschin, D., Rogers, N.C., Watling, D., Muller, M., Horn, F., Heinrich, P., Stark, G.R., and Kerr, I.M. (1996). JAKs, STATs and signal transduction in response to the interferons and other cytokines. Phil. Trans. R. Soc. (London) *B351*, 167–171.

Brünger, A.T., Adams, P.D., Clore, G.M., Gros, P., Grosse-Kuntsleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., et al. (1998). Crystallography and NMR system: a new software system for macromolecular structure determination. Acta Crystallogr. D., in press.

Carson, M. (1991). Ribbons 2.0. J. Appl. Crystallogr. *24*, 958–961.

Cho, Y., Gorina, S., Jeffrey, P.D., and Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science *265*, 346–355.

Collaborative Computational Project, N. (1994). The CCP4 suite programs for protein crystallography. Acta Crytstallogr. D. *50*, 760–763.

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

Darnell, J.E., Jr. (1997b). Phosphotyrosine signaling and the single cell:metazoan boundary. Proc. Natl. Acad. Sci. USA *94*, 11767–11769.

Esnouf, R. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 133–138.

Friend, S. (1994). p53: a glimpse at the puppet behind the shadow play. Science *265*, 334–335.

Fu, X.-Y. (1992). A transcription factor with SH2 and SH3 domains is directly activated by an interferon α-induced cytoplasmic protein tyrosine kinase(s). Cell *70*, 323–335.

Fu, X.-Y., Kessler, D.S., Veals, S.A., Levy, D.E., and Darnell, J.E., Jr. (1990). ISGF3, the transcriptional activator induced by interferon alpha, consists of multiple interacting polypeptide chains. Proc. Natl. Acad. Sci. USA *87*, 8555–8559.

Fu, X.-Y., Schindler, C., Improta, T., Aebersold, R., and Darnell, J.E., Jr. (1992). The proteins of ISGF-3, the interferon α-induced transcriptional activator, define a gene family involved in signal transduction. Proc. Natl. Acad. Sci. USA *89*, 7840–7843.

Ghosh, G., Van Duyne, G., Ghosh, S., and Sigler, P.B. (1995). The structure of NFκB homodimer bound to a κB site. Nature *373*, 303–310.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

Horvath, C.M., Wen, Z., and Darnell, J.E., Jr. (1995). A STAT protein domain that determines DNA seqeunce recognition suggests a novel DNA-binding domain. Genes Dev. *9*, 984–994.

Horvath, C.M., Stark, G.R., Kerr, I.M., and Darnell, J.E., Jr. (1996). Interactions between STAT and non-STAT proteins in the ISGF3 complex. Mol. Cell. Biol. *16*, 6957–6964.

Ihle, J.N., Witthuhn, B.A., Quelle, F.W., Yamamoto, K., and Silvennoinen, O. (1995). Signaling through the hematopoietic cytokine receptors. Annu. Rev. Immunol. *13*, 369–398.

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

Kawata, T., Shevchenko, A., Fukuzawa, M., Jermyn, K.A., Totty, N.F., Zhukovskaya, N.V., Sterling, A.E., Mann, M., and Williams, J.G. (1997). SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. Cell *89*, 909–916.

Kuriyan, J., and Cowburn, D. (1997). Modular peptide binding domains. Annu. Rev. Biophys. Biomol. Struct. *26*, 259–288.

La Fortelle, E. d., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. Meth. Enzymol. *276*, 472–494.

Leaman, D.W., Leung, S., Li, X., and Stark, G.R. (1996). Regulation of STAT-dependent pathways by growth factors and cytokines. FASEB J. *10*, 1578–1588.

Levy, D.E., and Darnell, J.E. (1990). Interferon-dependent transcriptional activation: signal transduction without second messenger involvement. New Biologist *2*, 923–928.

Martinez-Moczygemba, M., Gutch, M.J., French, D.L., and Reich, N.C. (1997). Distinct STAT structure promotes interaction of STAT2

with the p48 subunit of the interfereon-alpha-stimulated transcription factor ISGF3. J. Biol. Chem. *272*, 20070–20076.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D photorealistic molecular graphics. Meth. Enzymol. *277*, 503–524.

Meyer, W.K., Reichenbach, P., Schindler, U., Soldaini, E., and Nabholz, M. (1997). Interaction of STAT5 dimers on two low affinity binding sites mediates interleukin 2 (IL-2) stimulation of IL-2 receptor alpha gene transcription. J. Biol. Chem. *272*, 31821–31828.

Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., and Harrison, S.C. (1995). Structure of the NF-κB homodimer bound to DNA. Nature *373*, 311–317.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct. Funct. Genet. *11*, 281–296.

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

Pawson, T. (1995). Protein modules and signaling networks. Nature *373*, 573–580.

Schindler, C., Fu, X.-Y., Improta, T., Aebersold, R., and Darnell, J.E., Jr. (1992). Proteins of transcription factor ISGF3: one gene encodes the 91 and 84 kDa ISGF-3 proteins that are activated by interferon-α. Proc. Natl. Acad. Sci. USA *89*, 7836–7839.

Schindler, U., Wu, P., Rothe, M., Brasseur, M., and McKnight, S.L. (1995). Components of a Stat recognition code: evidence for two layers of molecular selectivity. Immunity *2*, 689–697.

Shuai, K., Ziemiecki, A., Wilks, A.F., Harpur, A.G., Sadowski, H.B., Gilman, M.Z., and Darnell, J.E. (1993). Polypeptide signaling to the nucleus through tyrosine phosphorylation of Jak and Stat proteins. Nature *366*, 580–583.

Shuai, K., Horvath, C.M., Tsai-Huang, L.H., Qureshi, S., 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.

Sicheri, F., and Kuriyan, J. (1997). Structures of Src-family tyrosine kinases. Curr. Opin. Struct. Biol. *7*, 777–785.

Veals, S.A., Schindler, C., Leonard, D., Fu, X.-Y., Aebersold, R., Darnell, J.E., Jr., and Levy, D.E. (1992). Subunit of an alpha-interferon-responsive transcription factor is related to interferon regulatory factor and Myb families of DNA-binding proteins. Mol. Cell. Biol. *12*, 3315–3324.

Vinkemeier, U., Cohen, S.L., Moarefi, I., Chait, B.T., Kuriyan, J., and Darnell, J.E., Jr. (1996). DNA binding of in vitro activated STAT-1α, STAT-1β and truncated STAT-1: 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.

Wagner, B.J., Hayes, T.E., Hoban, C.J., and Cochran, B.H. (1990). The SIF binding element confers sis/PDGF inducibility onto the c-fos promoter. EMBO J. *9*, 4477–4484.

Waksman, G., Kominos, D., Robertson, S.R., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., et al. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. Nature *358*, 646–653.

Xu, X., Sun, Y.L., and Hoey, T. (1996). Cooperative DNA binding and sequence-selective recognition conferred by the STAT amino-terminal domain. Science *273*, 794–797.

Zhang, J.J., Vinkemeier, U., Gu, W., Chakravarti, D., Horvath, C.M., and Darnell, J.E., Jr. (1996). Two contact regions between STAT1 and CBP/p300 in interferon signaling. Proc. Natl. Acad. Sci. USA *93*, 15092–15096.