A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain

Curt M. Horvath, Zilong Wen, and James E. Darnell Jr.

Laboratory of Molecular Cell Biology, The Rockefeller University, New York, New York 10021

Stat1 and Stat3 are two members of the ligand-activated transcription factor family that serve the dual functions of signal transducers and activators of transcription. Whereas the two proteins select very similar (not identical) optimum binding sites from random oligonucleotides, differences in their binding affinity were readily apparent with natural STAT-binding sites. To take advantage of these different affinities, chimeric Stat1:Stat3 molecules were used to locate the amino acids that could discriminate a general binding site from a specific binding site. The amino acids between residues ~400 and ~500 of these ~750-amino-acid-long proteins determine the DNA-binding site specificity. Mutations within this region result in Stat proteins that are activated normally by tyrosine phosphorylation and that dimerize but have greatly reduced DNA-binding affinities.

The STAT (signal transducers and activators if transcription) proteins have the dual purpose of, first, signal transduction from ligand-activated receptor kinase complexes, followed by nuclear translocation and DNA binding to activate transcription (Darnell et al. 1994). To function as specific transcriptional activators, STAT proteins by themselves or in combination with other proteins must have the ability to recognize specific DNA sequence elements in the promoters of their target genes. The binding of the STATs to DNA occurs only after tyrosine phosphorylation when the proteins form either homodimers (Shuai et al. 1994) or heterodimers (Schindler et al. 1992b; Zhong et al. 1994a,b) that bind DNA either alone or in combination with other proteins (Fu et al. 1990; Schindler et al. 1992b). Because a number of mutations in the STAT proteins block phosphorylation and thus dimerization (Shuai et al. 1993a; Improta et al. 1994), and none of the STAT sequences resembles previously well defined DNA-binding domains in other proteins, it has not been possible to define the DNA-binding domains of the STATs quickly and easily.

This paper reports an alternative approach to locating the DNA-binding domain. First, we sought DNA-binding sites that would distinguish Stat1 from Stat3 binding. Optimum binding sites (Pollock and Triesman 1990) were defined for Stat1 and Stat3 by determining the sequences each protein selected from a random deoxyoligonucleotide pool. Only minor differences in base preference were found for the two proteins outside a common identical core sequence TTCC[C or G]GGAA.

Whereas oligonucleotides representing these selected sequences exhibited slight binding preferences, the consensus sites overlapped sufficiently to be recognized by both factors. However, by screening different natural sites for affinity to the two proteins, Stat3 was shown not to bind to some sequences that occur in γ interferon (IFN-γ)-activated genes that were recognized well by Stat1, which is preferentially activated by IFN-γ (Shuai et al. 1992). Because both activated Stat1 and Stat3 bound strongly to synthetic sites we could score both general and specific binding. We then substituted parts of Stat1 into Stat3 or Stat3 into Stat1 and located a segment of amino acids between residues 400 and 500 that controlled the specificity of DNA binding. Mutations introduced into Stat3 within this region allowed phosphorylation and dimerization of protein but prevented binding to the general DNA site. Comparison of the sequences of STAT family members in this region revealed a potential helical region and some highly conserved residues in five different family members but did not reveal any obvious similarities with other DNA-binding domains, suggesting the STAT DNA-binding domains may be a newly recognized class of contact regions.

Results

In vitro binding site selection for Stat1 and Stat3

To determine whether Stat1 and Stat3 homodimers preferred different high-affinity oligonucleotide-binding

sites we carried out synthesis of a set of deoxyolignucle-otides 76 bases long: a random stretch of 26 bases was sandwiched between two constant 25 oligonucleotide regions that could be used as PCR primers. Stat1 optimum binding sites were determined first. Stat1 activation was carried out by IFN-γ treatment of Bud-8 fibroblast cells, and total cell extracts were exposed to the random deoxyoligonucleotide mixture. Stat1 carboxy-terminal antiserum (Schindler et al. 1992b) was used to immunoprecipitate the protein/DNA complexes followed by PCR amplification of the DNA in the precipitate (Pollock and Triesman 1990). Five such cycles were carried out, and

individual DNA segments were cloned after the final amplification. Sequencing of 55 individual clones demonstrated a clear consensus binding site with strong similarity to the earlier identified GAS elements (Fig. 1A; Decker et al. 1991, 1994; Lew et al. 1991). The most prominent feature of the selected sequence was a 9-bp inverted repeat with TTCCC/G as the half-site consensus, a feature consistent with the fact that Stat1 binds DNA as a dimer (Shuai et al. 1994). The symmetry around the central C or G [designated position 0 (zero)] is also reflected in the flanking sequence by a strong preference for A at position −6 and T at +6. There was also
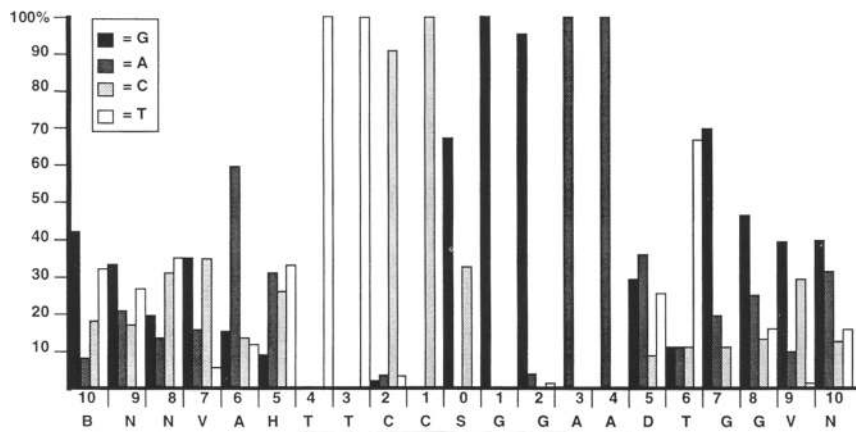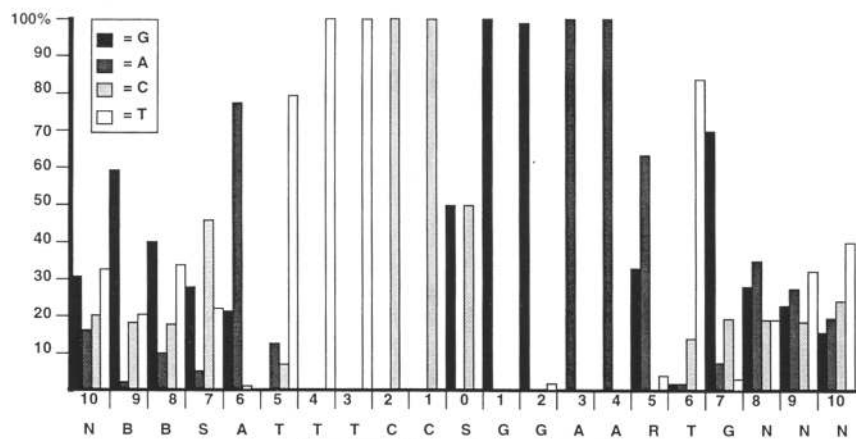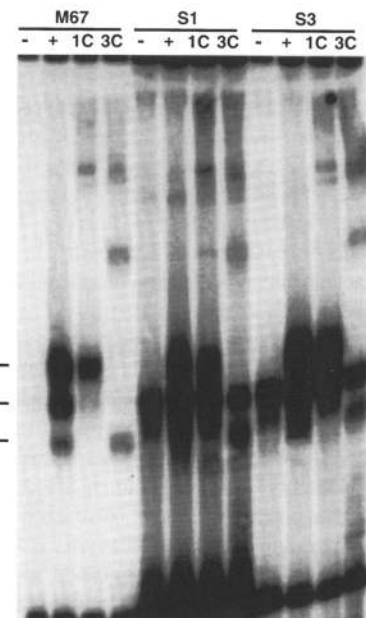

Figure 1. (A) Binding site selection for Stat1 and Stat3. Graphic representation of the nucleotide frequency in 55 independent binding sites selected by Stat1 (top) and Stat3 (bottom) in vitro from a pool of random oligodeoxynucleotides. Sequences were aligned to fit the TTNNNNNAA consensus previously recognized to be present in natural GAS elements (Table 1). The common core consensus is underlined, with the central nucleotide assigned position zero(0). The optimum consensus sequence and base preference in the flanking region is written beneath the graphs in International Union of Biochemists (I.U.B.) code. (N) G, C, A, T; (D) G, A, T; (H) A, C, T; (S) G, C; (K) G, T; (B) G, C, T; (V) G, A, C; (R) G, A. (B) EMSA with labeled Stat1 and Stat3 consensus site oligonucleotides. A radiolabeled probe that corresponds either to the Stat1 (S1) or Stat3 (S3) consensus sites was incubated with HepG2 nuclear extracts of cells that were untreated (−) or treated (+) with IL-6. Positions of SIF A, SIF B, and SIF C complexes are marked. Supershifting of the IL-6-induced complexes with Stat1 (1C)-or Stat3 (3C)-specific antisera is indicated above the lanes. (Top) Probes are identified. (*) The position of the constitutive comigrating band described in the text.

a preference at position +7 for a G, but position −7 did not show a preference, suggesting that the flanking sequences surrounding the core sequence may contribute to optimum binding.

A double-stranded deoxyoligonucleotide of 22 bp, containing in its center the consensus core sequence (TTC-CCGGAA), was synthesized and used as probe in the electrophoretic mobility shift assay (EMSA) (Fig. 1B; Fried and Crothers 1981; Levy et al. 1989). Extracts were used from both IFN-γ-treated HepG2 cells and HepG2 cells treated with a high dose of interleukin-6 (IL-6), which induces three well-recognized bands (Sadowski et al. 1993) described as SIF A, SIF B, and SIF C because there are three DNA-binding complexes inducible by medium from cells expressing the *sis* oncogene [(SIE) *sis*-inducible element; (SIF) *sis*-inducible factor] (Wagner et al., EMBO, 1990). The SIF C complex is identical in mobility and protein content to the IFN-γ-induced complex (Sadowski et al. 1993) and is therefore a Stat1 homodimer. This complex reacts with Stat1-specific antiserum. The SIF A complex, which migrates more slowly (most likely because of a greater number of positively charged amino acids in addition to a slightly longer polypeptide chain) reacts with the Stat3 antiserum (Zhong et al. 1994b) and is considered to contain a Stat3 homodimer. The SIF B complex that migrates between complex A and C and reacts with both Stat1 and Stat3 antisera is considered a Stat1:3 heterodimer. [These earlier conclusions are supported by results in Fig. 1B, lanes 1–4, with the synthetic oligonucleotide M67 (Wagner et al. 1990) as the labeled DNA probe.] The Stat1-selected consensus oligonucleotide bound weakly to some protein in untreated cells (Fig. 1B, lane 5) but also bound strongly to the induced STAT proteins that form SIF A, B, and C. Thus, it seemed possible there would be overlap of the Stat1 optimum binding site and any Stat3 response element.

To determine the optimum binding site for Stat3, extracts were used that contained high levels of activated Stat3 with much less Stat1. This was achieved by preparing extracts of epidermal growth factor (EGF)-treated, Stat3-transfected COS cells as the source of binding activity (Zhong et al. 1994b); the activated Stat3 homodimer bound to the random 76-bp probe (corresponding to the SIF A band) was identified by electrophoretic separation. The position of SIF A was marked using one of the Stat1-selected 76-nucleotide high-affinity sites that binds to Stat3 as shown in Figure 1B. The gel electrophoretic band was excised and DNA amplified, and five cycles of gel shifts and amplification were carried out before cloning of individual examples of DNA from the SIF A complex. Sequencing of 55 individual clones with Stat3-selected sequences also revealed a clear consensus sequence that was identical in the core sequence TTCC[C or G]GGAA to that selected by the Stat1 (Fig. 1A). Just as did the Stat1 site, the Stat3-selected site contained an A or T at positions +6 and −6, respectively, but in addition the Stat3 site also showed a strong preference for A and T at positions +5 and −5 making a 13-nucleotide palindrome the favored Stat3 site. As with

Stat1, a preference for G at position +7 was not matched by a C at position −7. Also, position −9 was G in ~60% of cases. As with Stat1, these flanking sequence preferences may contribute to the optimum site.

An oligonucleotide probe was synthesized to represent the Stat3 optimal site (position −9 to +9) and used in a gel shift experiment (Fig. 1B, lanes 9–13). Because the Stat1 optimum site core is contained within the Stat3 probe, it was not surprising that, like the selected Stat1 probe, the Stat3 probe bound well to all of the SIF complexes. Unfortunately, the Stat3 consensus probe used also bound even more strongly to a constitutively active protein (marked by the asterisk in Fig. 1B) that comigrates closely with SIF B, obscuring the center section of the gel shift pattern. It was noted that the Stat3 consensus probe bound somewhat better in the SIF A complex from which it had been selected than did the Stat1 optimum probe, but this was estimated by competition experiments to be only a three- to fivefold difference. Although it is clear that such relatively minor differences might be important at individual sites in genomic DNA, we could not use these "consensus" probes to easily distinguish the binding affinities of Stat1 from Stat3.

Stat protein binding to natural sites

We then examined previously identified Stat protein-binding elements to determine whether any sites gave sufficient specificity to distinguish easily Stat1 from Stat3 binding. Oligonucleotide probes representing GAS [IFN-γ-activated sites (Decker et al. 1991; Lew et al. 1991)] from the murine surface antigen Ly6E (Khan et al. 1993), IFN-γ response region (the GRR) of the FcγR1 gene (Pearse et al. 1993), the c-*fos* SIE and its high-affinity mutated form, M67 (Wagner et al. 1990; Sadowski et al. 1993), and the optimum Stat1 core sequence were surveyed by EMSA for their effectiveness as Stat1- or Stat3-binding sites (Fig. 2). Using extracts from HepG2 cells treated with IL-6 that contain SIF A-, SIF B-, and SIF C-binding activity, differences were clearly observed among these probes. The M67 SIE probe bound to form in near equimolar amounts the SIFA, SIFB, and SIFC complexes, whereas the natural c-*fos* site gave a very weak signal with STAT proteins. The Stat1 optimum core sequence was also bound by all of the SIF species but with overall lower affinity as judged by the intensity of the binding signal. Thus, the M67 probe binds well to both Stat1 or Stat3 but cannot distinguish between them. In contrast, the GRR and Ly6e probes were both bound by the SIF C protein (Stat1 homodimer), with the GRR probe giving two- to threefold more binding than the Ly6E probe. Both probes were bound poorly by the SIF B complex, the heterodimer of Stat3 and Stat1. Most significantly, the SIF A complex that represents Stat3 homodimer binding, was not observed with the GRR or Ly6E probes unless the autoradiographs were overexposed. Thus, the two closely related proteins Stat3 and Stat1 differ in their ability to recognize these two natural GAS elements. Other GAS elements tested (from the IRF1 gene, the α-2 macroglobulin gene, the guanylate-
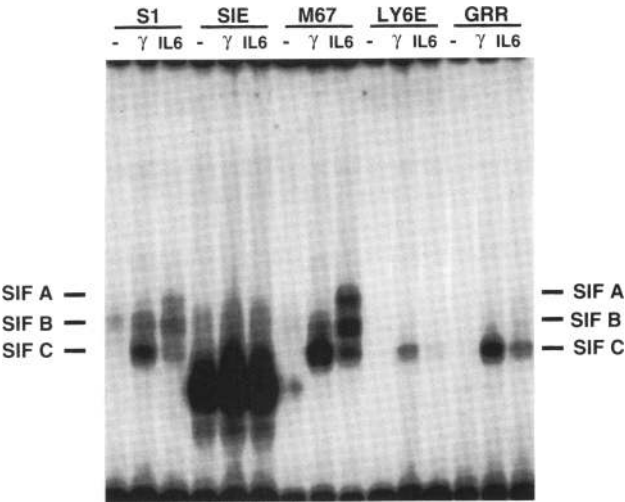
Figure 2. Binding of Stat1 and Stat3 to known GAS elements reveals differential binding patterns. Nuclear extracts from untreated (–), IFN-γ-treated (γ), and IL-6-treated HepG2 cells were incubated with the indicated probes and DNA protein complexes detected by EMSA. Positions of SIF A, SIF B, and SIF C are marked. (S1) Stat1-selected consensus sequence; (SIE) c-*fos* promoter *sis*-inducible element; (M67) hyperactive mutated form of SIE; (Ly6E) GAS element from the Ly6E gene promoter; (GRR) FcγR1 promoter IFN-γ response element.

binding protein gene, and the B-casein gene) displayed intermediate binding properties with respect to Stat1 and Stat3 binding and were not useful for this analysis (data not shown).

Localization of specific DNA-binding region of Stat proteins

We proceeded to use the differential binding affinities of Stat1 and Stat3 to the GRR compared with uniform binding to the M67 SIE probe in determining the STAT protein region that discriminates between the probes. The Stat1 SH2 group lies between amino acids 573 and 700 (residues ~600–700) (Fu 1992; Schindler et al. 1992a,b), and the Y that becomes phosphorylated is at residue 701. Mutations at the Y701 and in R602 in the pocket of Stat1 SH2 have proved the necessity of these regions in STAT tyrosine phosphorylation and subsequent activation as a DNA-binding protein (Shuai et al. 1993a,b; 1994). Moreover the SH2 region of Stat1 has been shown to confer IFN-γ inducibility on Stat2 (Heim et al. 1994). Thus, a chimeric protein with the Stat1 carboxyl terminus can be activated by IFN-γ. Stat3 also contains an SH2 region from ~600–700 and a Y in a position comparable to Stat1 at residue 705, but Stat3 is not activated by IFN-γ (Zhong et al. 1994a). [Mutations of the Stat3 Y residue at 705 to phenylalanine likewise blocks phosphorylation of Stat3 (Z. Wen and J.E. Darnell, unpubl.)]

As the segment of STAT proteins from ~600 to ~750 appears to function in activation and dimerization, we focused on the amino-terminal regions as a possible source of DNA-binding specificity. Gene fusions were constructed that code for chimeric Stat proteins containing regions of Stat1 fused to Stat3 or vice versa (Fig. 3). The chimeras are named to specify the source of the fused Stat protein from the amino to the carboxyl terminus with the amino acid number of the joint in subscript.
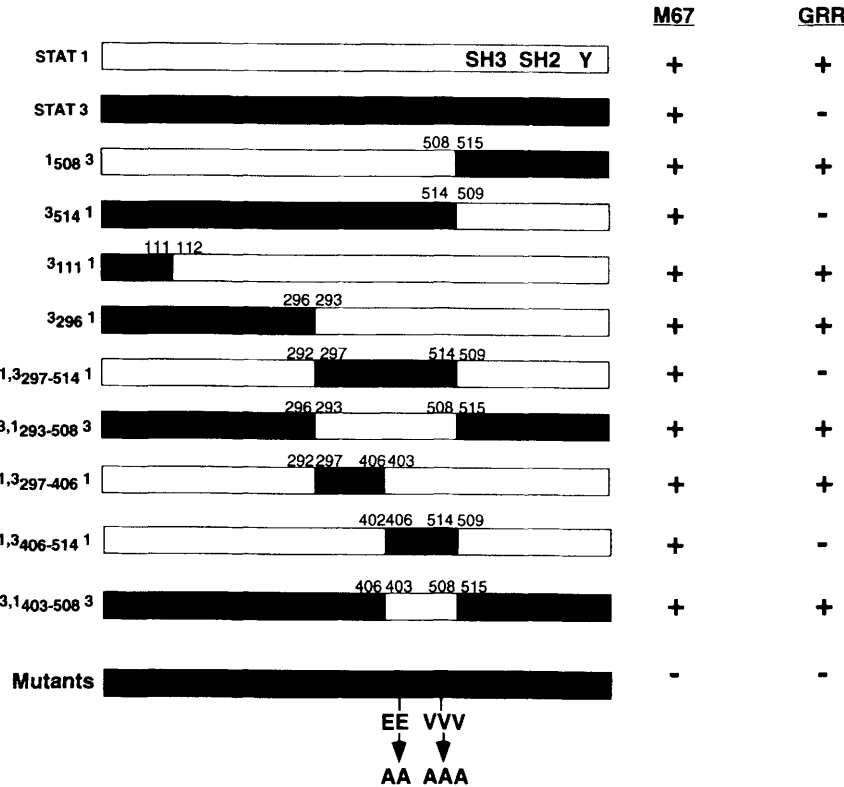

Figure 3. Diagrammatic representation of the Stat1/Stat3 chimeras used in this study. (Open box) The Stat1 molecule; (solid box) Stat3. The numbers above the boxes refer to the amino acid residues of Stat1 or Stat3 before and after the chimeric junction. Positions of the *src* homology domains (SH3, SH2) and activating tyrosine (Y) are indicated for Stat1. Binding properties for the M67 and GRR oligodeoxynucleotides, as determined in this study (see Fig. 4), are indicated at *right*. (*Bottom*) The box depicts the postions of the two mutations made in Stat3 (see Fig. 5), drawn to approximate scale.

For example, $1_{500}3$ means Stat1 amino acids 1–500 joined to Stat3 at amino acid 500. The cDNAs were transfected into U3A cells, and permanent cell lines expressing the recombinant proteins were selected. U3A cells lack expression of Stat1 protein but contain active receptors for IFN-γ or IFN-α (Pellegrini et al. 1989; Muller et al. 1993). Stat1 (and chimeric proteins containing the Stat1 carboxy-terminal activation regions) introduced into this cell line can be activated by IFN-γ or

IFN-α (Fig. 4; Muller et al. 1993; Improta et al. 1994). Stat3 can be activated by IFN-α in the U3A precursor cell line, 2FTGH (I. Kerr, pers. comm.; C.M. Horvath, Z. Zhong, and J.E. Darnell Jr., unpubl.), but we found that the U3A cells derived from 2FTGH by extensive mutagenesis (Pellegrini et al. 1989) did not respond by activating the endogenous Stat3. However, the wild-type Stat3 permanently introduced into U3A cells was activated by IFN-α (Fig. 4A, last lane) (C.M. Horvath and J.E.
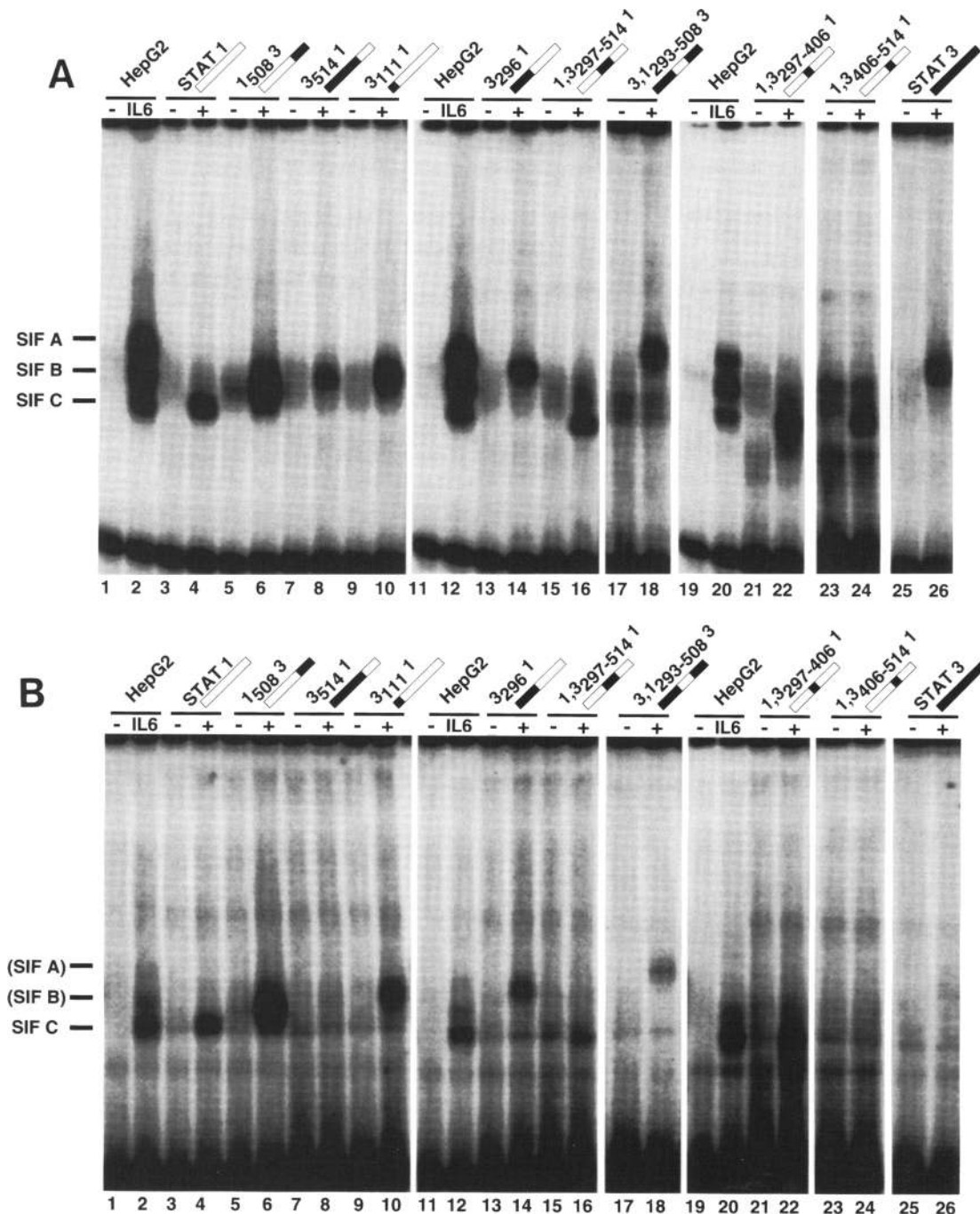

Figure 4. (See facing page for C and legend.)

Darnell Jr., unpubl.). Therefore, we used IFN-α to activate in U3A-derived cells lines the chimeric proteins containing the Stat3 carboxy-terminal activation regions.

Consistent with the the results using IL-6-treated HepG2 extracts (Fig. 1B), extracts of U3A cells permanently transfected with either Stat1 and treated with IFN-γ or transfected with Stat3 and treated with IFN-α, displayed the same differential DNA-binding properties as did the same proteins activated in HepG2 cells (Fig. 4). Activated Stat1 binds well to both M67 and GRR probes, whereas activated Stat3 binds to M67 but not (or very poorly) to the GRR probe (Fig. 4A, B, lanes 4 and 26). Chimeric junctions in the first ~500 amino acids were chosen based on regions of amino acid sequence identity between Stat1 and Stat3 so as not to disrupt potentially important domains of the resulting hybrid proteins. As mentioned earlier, a greater number of charged amino acid residues plus a slightly greater length in Stat3 compared with Stat1 is the cause for the slower migration of Stat3 homodimers compared with Stat1 homodimers. In chimeric proteins, these differences were reflected in protein/DNA complexes that migrated at intermediate rates. A chimeric Stat protein containing the first 508 amino acids of Stat1 and the carboxyl terminus of Stat3 exhibited the general binding property of Stat1 in that the chimeric protein, designated $1_{508}3$, bound well to both test probes and migrated just slightly slower than Stat1 (Fig. 4A, B, lane 6). The complementary chimera, $3_{514}1$, with the amino-terminal 514 amino acids of Stat3 fused to the carboxyl terminus of Stat1, had the recognition property of Stat3, that is, it bound well to M67 probe, but not to GRR (Fig. 4A, B, lane 8). Thus, the STAT DNA recognition capacity was localized to the amino-terminal 508 amino acids of Stat1 or 514 amino acids of Stat3 and was not influenced by the putative SH3 domain (~500–600), the SH2 domain (~600–700), or other sequences in the carboxy-terminal third of the molecule, which itself can utilize different ligand–receptor complexes for activation (IFN-γ for Stat1 and IFN-α for Stat3).

To further dissect the STAT DNA recognition region, additional chimeras were constructed containing the amino-terminal 111 or 296 amino acids of Stat3 substituted into Stat1. Both recombinant molecules, $3_{111}1$ or $3_{296}1$, retained the binding characteristic of Stat1 (Fig. 4A, B, lanes 10 and 14), recognizing both M67 and GRR probes. These results suggest that the amino-terminal 296 amino acids do not determine the specificity of DNA sequence recognition. It seemed reasonable to infer from this set of chimeras that the region from amino acid 297 to 514 of Stat3 (or 508 of Stat1) imparted the ability to discriminate between DNA elements. To test this suggestion directly, the region of Stat1 between amino acids 292 and 509 was replaced with the Stat3 amino acids 297–514 (chimera $1,3_{297-514,}1$) and a corresponding Stat3 with a Stat1 insertion, chimera $3,1_{293-508,}3$ was made. The differential binding to the test sequences of the $1,3_{297-514,}1$ molecule showed that although the amino acid sequence was primarily Stat1, the recombinant molecule now bound M67 but failed to bind the GRR showing that recognition capacity of Stat3 was transferred to Stat1. Reciprocally, when chimera $3,1_{293-508,}3$ was tested, the recombinant, largely Stat3 sequence could now bind well to both the M67 and GRR probes, transferring the DNA-binding property of Stat1 (cf. Fig. 4 A and B, lanes 18 and 26). We conclude that the portion of the STAT protein that recognizes the DNA response el-
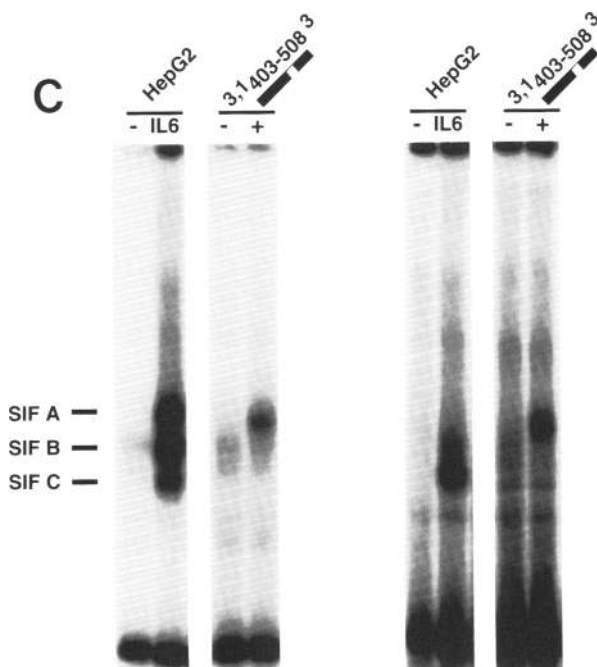
Figure 4. Differential binding of the chimeric STAT proteins. (A) Nuclear extracts from untreated (−) and IFN-treated (+) U3A cells expressing the chimeric STAT proteins were incubated with M67 probe to reveal all DNA-binding complexes. Positions of SIF A, SIF B, and SIF C are marked as determined from IL-6-treated HepG2 cell nuclear extracts. (B) The same extracts incubated with GRR probe. The position of SIF C from IL-6-treated HepG2 cell nuclear extracts is marked, and the positions where SIF A and SIF B would migrate (in parentheses) are marked. (C) Nuclear extracts from U3A cells expressing chimera $3,1_{403-508}3$ were incubated with M67 (left) or GRR (right) probes.

ement lies between amino acids 297 and 514 of Stat3 and between amino acids 293 and 508 of Stat1. A final set of chimeric molecules that positioned the Stat3 recognition capacity more accurately were then constructed. The 200-amino-acid region was divided into two ~100-amino-acid insertions of Stat3 into Stat1. These chimeras showed that amino acids 297–406 left Stat1 recognition intact, whereas insertions of amino acids 406–514 resulted in the transfer of Stat3 recognition (Fig. 4A, B, lanes 22, 24). The complementary substitution, amino acids 403–508 of Stat1 inserted into Stat3, binds well to both the M67 and GRR probes, restoring the site-specific recognition property of Stat1 (Fig. 4C). We conclude that the amino acids that detemine DNA-binding specificity lie in this ~108-amino-acid segment between residues 406 and 514 of Stat3 and 403 and 508 of Stat1.

Point mutations alter DNA-binding affinity

The proposed DNA recognition domain (~400–500) encompasses one of the most highly conserved regions of the STAT protein family, although no function had been assigned previously to this region either from experiment or from sequence comparison with other proteins in the data banks. To ascertain whether specific amino acids within the conserved amino acid stretches were important for binding to DNA, mutations were made in two of the highly conserved regions of Stat3 in the region of amino acids ~400–500. The sequence VTEEL (residues 432–436) was changed to VTAAL (mutant EE → AA), or the conserved sequence SLPVVVISN (residues 458–466) was changed to SLPAAAISN (mutant VVV → AAA). Each mutant protein was expressed transiently in COS-1 cells [which have low endogenous Stat3 protein levels (Zhong et al. 1994b)] and nuclear extracts prepared following activation with EGF. Neither of the two mutants produced STAT proteins capable of binding the M67 element to the same extent as wild-type STAT 3, suggesting that both mutations influenced DNA recognition. Mutant EE → AA had a more severe effect on DNA binding (nearly undetectable) than mutant VVV → AAA, which exhibited a distinctly reduced but still detectable binding (Fig. 5A). To determine whether these mutations blocked activation of the protein, Stat3 antiserum was used to precipitate proteins from the same COS cell extracts, and the precipitates were tested by immunoblotting with antiphosphotyrosine antibody. Both mutant proteins were phosphorylated as well as the wild-type protein (Fig. 5B). To determine whether the mutant STAT proteins were capable of dimerization, the mutant EE → AA or mutant VVV → AAA were tagged with a FLAG epitope (Hopp et al. 1988) so that they could be distinguished from endogenous Stat 3 and transfected into COS cells along with nontagged Stat1 cDNA. Extracts of the COS cells treated with EGF were then precipitated with monoclonal antibody to the FLAG epitope (M2). If dimerization occurred, the FLAG-tagged protein should carry along both endogenous and transfected activated Stat1 protein in heterodimers into the precipitate. Figure 5C shows clearly that this was the
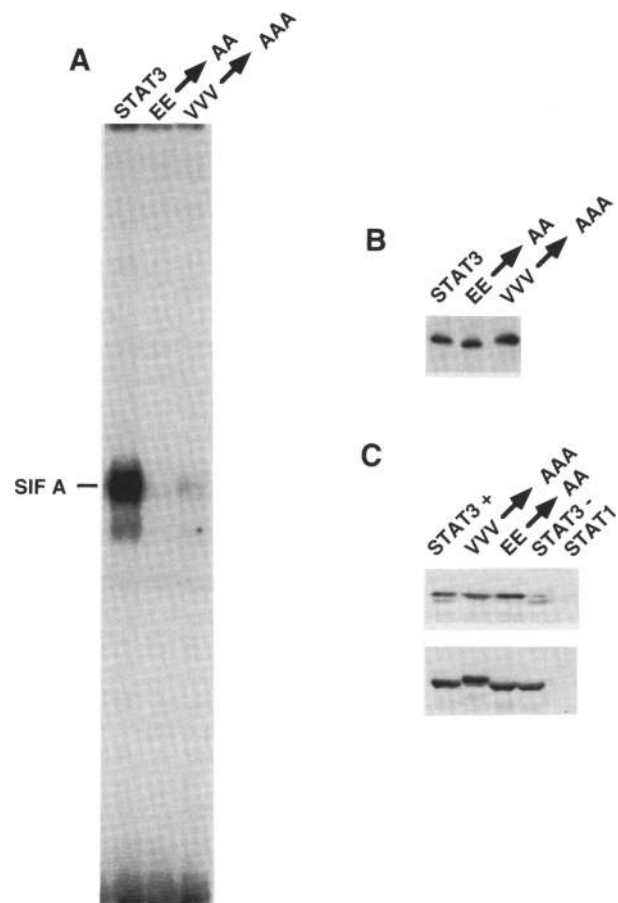

Figure 5. Mutations in Stat3 influence DNA-binding affinity. (A) EMSA analysis of DNA–protein complexes. Nuclear extracts from EGF-treated COS cells transfected with Stat3, mutant EE → AA or mutant VVV → AAA (see Materials and methods) were incubated with labeled M67 probe to reveal DNA-binding complexes. The position of SIF A is marked. (B) Phosphotyrosine immunoblotting. Extracts from the cells in A were immunoprecipitated with Stat3-specific antiserum, separated by SDS-PAGE, transferred to nitrocellulose, and probed with monoclonal antibody PY20. (C) Coimmunoprecipitation of Stat1 and Stat3 mutants. COS cells were transfected with FLAG-tagged Stat3 or mutants along with untagged Stat1 and treated (+) or not treated (−) with EGF. FLAG immunoprecipitates were separated by SDS-PAGE, transferred to nitrocellulose, and probed with Stat1-specific antiserum (top). STAT1 refers to transfection with Stat1 alone. (Bottom) An immunoblot with FLAG-specific monoclonal antibody to demonstrate similar expression levels in the transfected cells.

case; Stat1 was detected in all FLAG-containing extracts but not in control cells transfected with Stat1 alone. A small amount of Stat1 coprecipitated with FLAG–Stat3 from untreated COS cells, reflecting a low basal level of Stat3 activation. The amount of Stat1 from the treated cells was about fivefold greater than from the untreated cells, indicating a ligand-induced heterodimerization. These data support the conclusion that the mutant EE → AA and VVV → AAA proteins become phosphorylated in response to ligand and dimerize but cannot bind

DNA as well as wild-type Stat3. These results greatly strengthen the conclusion that this highly conserved region of the STAT proteins between 406 and 514 participate in recognition of and binding to GAS-like DNA response elements.

Discussion

In the past 2 years a large number of reports have indicated that sequences of the general motif TTNCNN-NAA, the originally defined GAS consensus, can be used to detect activated STAT DNA binding (Khan et al. 1993; Lew et al. 1989; Pearse et al. 1993; Wegenka et al. 1993). We sought to determine first whether two specific STAT members that are activated by different ligands would select individual binding sites. However, optimum site selection experiments showed that both Stat1 and Stat3 preferred very similar 9-bp core elements and only minor differences in flanking sequences. The selection of highly similar optimum sites is characteristic of other DNA-binding protein families such as homeo box protein (Wilson et al. 1993), yet it is clear that specific biologic events are controlled by different family members. It is generally believed therefore that optimum binding sites may be used less commonly in evolution but that chromosomal binding sites evolved that are differentially distinguished by particular members of protein families. In line with this conjecture we found that two sites from genes known to be activated by IFN-γ, the GRR of the FcγR1 gene, and the GAS site in the promoter of the Ly6E gene are bound by Stat1 homodimers but not by Stat3 homodimers. The high-affinity synthetic derivative of the cfos promoter M67, in contrast, is bound by both proteins and served to monitor the binding of either protein. It is interesting to note that the GRR sequence differs from the selected core sequence only at position +1, where A replaces G. Similarly, the Ly6E sequence differs from the M67 probe at only one position within the core (T replaces C at the zero position). Thus, these central nucleotides within the 9 bp are important for Stat3 binding, whereas Stat1 binding is less demanding at these sites.

Most of the genomic DNA sites (Table 1) that presumably function to bind STAT proteins do not contain the perfect 9-base palindrome selected by the optimum site selection techniques. Considerable additional work will be required to determine the in vivo binding specificity of chromosomal GAS sites for particular STAT proteins, especially as few experiments have yet been reported on the influence of adjacent binding sites for additional transcription factors that may bind coordinately with STAT proteins.

The high amino acid sequence identity between Stat1 and Stat3, coupled with the inherent ability of Stat3 to distinguish between M67 and GRR elements, made it possible to define the DNA-binding domain of the STAT proteins by exchanging regions between two proteins and assaying the substituted proteins for DNA site-binding preference. This technique resulted in the identification of residues 406–514 as capable of the transfer of

Table 1. *Comparison of GAS-like promoter elements*

Source	Core element	Reference
S3	TTCCGGGAA	this study
S1	TTCCGGGAA	this study
M67 SIE	TTCCCGTAA	Wagner et al. (1990)
cFOS SIE	TTCCCGTCA	Wagner et al. (1990)
Ly6E/A	TTCCTGTAA	Khan et al. (1993)
FcγR1	TTCCCAGAA	Pearse et al. (1993)
GBP	TTACTCTAA	Decker et al. (1989)
MIG	TTACTATAA	Wong et al. (1994)
IFP53	TTCTCAGAA	Strehlow et al. (1993)
ICAM-1	TTCCCGGAA	Yuan et al. (1994)
IRF1	TTCCCCGAA	Yuan et al. (1994)
ICSBP	TTCTCGGAA	Kanno et al. (1993)
α2-macroglobulin	TTCCCGTAA	Wegenka et al. (1993)
Acid glycoprotein	TTCCCAGAA	Wegenka et al. (1993)

binding specificity, as an activated Stat1 molecule containing residues 406–514 of Stat3 could bind only to the M67 probe and not the GRR probe while activated Stat1 itself and a Stat3 with amino acids 403-508 of Stat1 binds to both probes. Within these 108 amino acids, Stat1 and Stat3 have only 43 amino acid differences. Counting conservative amino acid changes the sequences are even more similar. Consistent with the observations of basic residues within DNA-binding domains there are three positively charged amino acids conserved in the first 20 amino acids of this segment in Stat1, Stat3, Stat4, Stat5, and Stat6. Mutations targeted to the most conserved sequences in this domain have no effect on phosphorylation or dimerization of the STAT proteins but reduce DNA binding. We conclude that this region of the Stat1 and Stat3 proteins between 406 and 514 controls DNA-binding specificity and is likely to be part of the DNA-binding domain. Because the region between 400 and 500 is highly conserved in all of the other reported STATs it seems likely that this region will function for all family members.

No clear homologies to known DNA-binding proteins have been found at the level of primary sequence comparison. Recently, the structure of NF-κB was solved. As opposed to short recognition helices, the 38 DNA contacts of this dimeric transcription factor are distributed over a 300-amino-acid region with multiple loops making phosphate contacts (Ghosh et al. 1995; Muller et al. 1995). It is not unreasonable to think the STAT DNA-binding domain might also be composed of multiple contact regions and that we have identified one important specificity-determining region.

To suggest any possible folding motifs in the putative DNA-binding regions, amino acids in the 293–467 region of all the presently cloned STATs (1–6) were analyzed by computer comparison, predicting secondary structure motifs by the algorithm of Chou and Fasman (Fig. 6; Genetics Computer Group 1991). The consensus prediction suggests a helical domain surrounding the VTEEL sequence, which extends to the SLPVVV sequence; this sequence is at the beginning of a predicted β-sheet. Com-
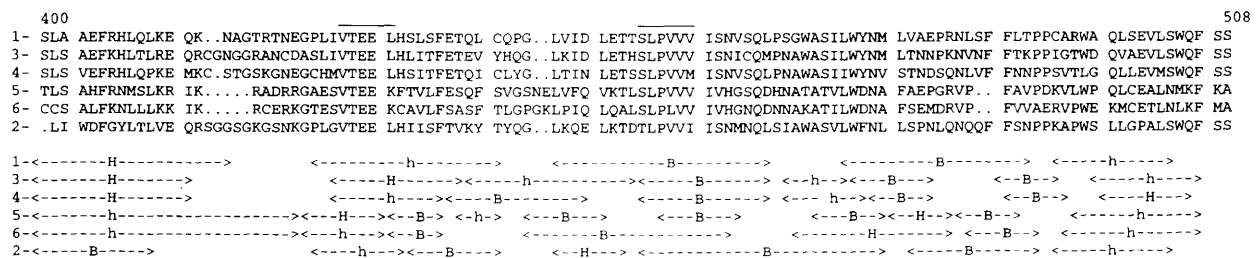
```
     400                                                                                                                    508
1- SLA AEFRHLQLKE QK..NAGTRTNEGPLIVTEE LHSLSFETQL CQPG..LVID LETTSLPVVV ISNVSQLPSGWASILWYNM LVAEPRNLSF FLTPPCARWA QLSEVLSWQF SS
3- SLS AEFKHLTLRE QRCGNGGRANCDASLIVTEE LHLITFETEV YHQG..LKID LETHSLPVVV ISNICQMPNAWASILWYNM LTNNPKNVNF FTKPPIGTWD QVAEVLSWQF SS
4- SLS VEFRHLQPKE MKC.STGSKGNEGCHMVTEE LHSITFETQI CLYG..LTIN LETSSLPVVM ISNVSQLPNAWASIIWYNV STNDSQNLVF FNNPPSVTLG QLLEVMSWQF SS
5- TLS AHFRNMSLKR IK.....RADRRGAESVTEE KFTVLFESQF SVGSNELVFQ VKTLSLPVVV IVHGSQDHNATATVLWDNA FAEPGRVP.. FAVPDKVLWP QLCEALNMKF KA
6- CCS ALFKNLLLKK IK.....RCERKGTESVTEE KCAVLFSASF TLGPGKLPIQ LQALSLPLVV IVHGNQDNNAKATILWDNA FSEMDRVP.. FVVAERVPWE KMCETLNLKF MA
2- .LI WDFGYLTLVE QRSGGSGKGSNKGPLGVTEE LHIISFTVKY TYQG..LKQE LKTDTLPVVI ISNMNQLSIAWASVLWFNL LSPNLQNQQF FSNPPKAPWS LLGPALSWQF SS

1-<-------H----------->       <--------h-------->       <----------B--------->      <---------B-------->  <-----h----->
3-<-------H------->                <-----H------><-----h---------><-----B---->  <--h--><---B--->     <--B-->   <-----h---->
4-<-------H------->                  <-----h----><----B-------->  <-----B------> <--- h--><--B-->     <--B-->   <----H--->
5-<-------h------------------><---H---><---B-> <-h-> <---B--->    <---B--->      <---B--><--H--><--B-->      <------h----->
6-<-------h------------------><---h---><<--B-> <------B---------->    <------H-------> <---B--> <------h----->
2-<-----B----->        <----h---><---B---->     <--H---> <------------B-----------> <-----B-------> <-----h----->
```

Figure 6. Alignment of STAT family members in the putative DNA-binding region. Broken lines indicate boundaries of putative helices (H,h) and β-sheets (B,b) predicted by the algorithms of Chou and Fasman for each of the family members. Numbers above the alignment refer to the Stat1 sequence. The conserved amino acids mutated in this study are overlined. Sequences were aligned using the GCG PILEUP program, and secondary structure was predicted using the GCG peptide structure program (Genetics Computer Group 1991).

parison of the possible DNA-binding region that we define here to known DNA-binding domains does not reveal any similarity. Perhaps the STAT protein DNA-binding domain will represent an unusual class of DNA-binding domain. It is interesting also that this domain lies between the SH3 homology, which binds proline-rich sequences (Cicchetti et al. 1992), and the conserved STAT sequence PCMPXXPXXP. If these two sequences interacted within a STAT molecule prior to phosphory-lation of the protein, the DNA-binding domain might be shielded in the nonphosphorylated protein or, conversely, such an interaction after phosphorylation might present the putative DNA-binding domain.

The exchange of this 108-amino-acid domain can substitute the DNA recognition properties of these two STAT proteins. A more direct demonstration that this region is the DNA contact domain would be to tranfer this domain to another class of dimeric transcription factors. We have attempted to reconstitute specific DNA recognition by grafting these sequences onto an unrelated dimerization domain from the heterologous basic leucine zipper (bZIP) or helix–loop–helix (HLH) families. STAT amino acids ~300–500 were joined to the CCAAT/enhancer-binding protein (c/EBP) leucine zipper and the E47 HLH domains, but demonstration of specific DNA binding by these fusion proteins has been unsuccessful so far. One reason might be that specific structural properties inherent in the STAT family of transcription factors are not provided simply by the dimerization motifs of these other factors. For example, the primary dimerization of the STAT proteins is mediated by intermolecular SH2/phosphotyrosyl interactions (~600–710), which predicts an antiparallel interaction of the two chains in this dimeric region (Shuai et al. 1994). Perhaps this orientation requires compensation as the chains emerge from the dimer to present the residues of the 400–500 region to DNA. bZIP and HLH dimerization domains are parallel with a short hinge region that allows the short DNA contact helices of those proteins to rotate correctly to form "induced sites" on the DNA (Burley 1994). Because the potential STAT DNA contact region has only a limited helical content, it is possible that the domain must make a protein fold that has not yet been described in other DNA-binding proteins.

Materials and methods

Cell culture, cytokines, and antisera

Human U3A cells [generously provided by Drs. George Stark (Cleveland Clinic Foundation Research Institute, OH) and Ian Kerr (Imperial Cancer Research Foundation)], HepG2 cells, and COS-1 cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% bovine calf serum. Transfection of cells and selection of stable cell lines were carried out by standard procedures (Shuai et al. 1993a). Treatment of cells with cytokines was for 15 min unless otherwise noted. IFN-γ (a gift from Amgen) was used at a concentration of 5 ng/ml; IFN-α (a gift from Hoffman LaRoche) was used at a concentration of 500 IU/ml. IL-6 (UBI) was used at a concentration of 30 ng/ml. EGF (a generous gift of Dr. S. Cohen; Vanderbilt University School of Medicine, Nashville, TN) was used at 50 ng/ml. Cytoplasmic and nuclear extracts were prepared as described (Sadowski and Gilman 1993). For immunoprecipitation of cell extracts, Stat1 or Stat3 carboxy-terminal antiserum was used at a 1:200 dilution. Immobilized FLAG-specific monoclonal anibody was used for precipitiation according to the manufacturer's instructions (Kodak). Phosphotyrosine-specific monoclonal antbody PY20 was used at a dilution of 1:2000 according to the manufacturer's instructions (Transduction Laboratories).

Plasmid Construction

Expression plasmid pRcCMV (Invitrogen) carrying Stat1 or Stat3 cDNA (Improta et al. 1994; Zhong et al. 1994b) was used for all cell lines. All of the recombinant STAT proteins were constructed by polymerase chain reaction (PCR) amplification using Vent polymerase (NEB) and verified by DNA sequencing. The chimeric Stat1 and Stat3 cDNAs included the FLAG epitope [Kodak IBI (Hopp et al. 1988)] to easily identify the recombinant proteins.

EMSA

Gel moblity shift assays were carried out as described (Levy et al. 1989). Double-stranded oligonucleotide probes were synthesized for use as the probe with 5'-GATC protruding ends. Probe sequences used in this study are SIE, 5'-CAGTTCCCGTCAAT-CAT-3'; M67, 5'-CATTTCCCGTAAATCAT-3'; Ly6E, 5'-ATAT-TCCTGTAAGTGAT-3'; GRR, 5'-GTATTTCCCAGAAAAGG-3'; S1, 5'-GTTGTTCCGGGAAAATT-3'; and S3, 5'-TATTTCC-GGGAAATCCC-3'.

Binding site selection

In vitro binding site selection for Stat1 was carried out essen-

tially according to the method of Pollock and Triesman (1990). IFN-γ-treated BUD 8 fibroblast nuclear extracts were mixed with a double-stranded random 76-base oligomer and immunoprecipitated with antiserum specific for Stat1 and protein A–agarose. The copurifying DNA was isolated, amplified by PCR, and analyzed for binding by EMSA. Following five rounds of selection, a Stat-specific complex was observed, eluted from the gel, and subcloned. To obtain the Stat3 optimum site, nuclear extracts from EGF-treated COS-1 cells transfected with Stat3 expression vector were bound to the random oligomer and applied to an EMSA gel. The region corresponding to the mobility of the Stat3 gel shift on one of the 76-bp Stat1-selected sites was excised, and the DNA was amplified by PCR. Following five rounds of selection from the gel, the resulting complex was supershifted by Stat3-specific antiserum and the DNA isolated from the supershifted complex was eluted from the gel, amplified, and subcloned.

Acknowledgments

References

Burley, S.K. 1994. DNA-binding motifs from eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **4:** 3–11.

Cicchetti, P., B.J. Mayer, G. Thiel, and D. Baltimore. 1992. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science* **257:** 803–806.

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

Decker, T., D.J. Lew, Y.-S. Cheng, D.E. Levy, and J.E. Darnell Jr. 1989. Interactions of alpha- and gamma-interferon in the transcriptional regulation of the gene encoding a guanylate-binding protein. *EMBO J* **8:** 2009–2014.

Decker, T., D.J. Lew, J. Mirkovitch, and J.E. Darnell Jr. 1991. Cytoplasmic activation of GAF, an IFN-γ regulated DNA binding factor. *EMBO J* **10:** 927–932.

Fried, M. and D.M. Crothers. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9:** 6505–6525.

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., D.S. Kessler, S.A. Veals, D.E. Levy, and J.E. Darnell Jr. 1990. ISGF3, the transcriptional activator induced by interferon α, consists of multiple interacting polypeptide chains. *Proc. Natl. Acad. Sci.* **87:** 8555–8559.

Genetics Computer Group. 1991. Program manual for the GCG package, version 7. Madison, WI.

Ghosh, G., G.V. Duyne, S. Ghosh, and P.B. Sigler. 1995. Struc-

ture of NF-κB p50 homodimer bound to a κB site. *Nature* **373:** 303–310.

Heim, M.H., I.M. Kerr, G.R. Stark, and J.E. Darnell Jr. 1994. STAT SH2 groups contribute to specific interferon signalling by the Jak-STAT pathway. *Science* **267:** 1347–1349.

Hopp, T.P., K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D.P. Cerretti, D.L. Urdal, and P.J. Conlon. 1988. A short polypeptide marker sequence useful for recombinant protein identification and purification. *BioTechnology* **6:** 1204–1210.

Improta, T., C. Schindler, C.M. Horvath, I.M. Kerr, G.R. Stark, and J.E. Darnell Jr. 1994. ISGF-3 formation requires phosphorylated Stat91 but Stat113 is phosphorylated independently of Stat91. *Proc. Natl. Acad. Sci.* **91:** 4776–4780.

Kanno, Y., C.A. Kozak, C. Schindler, P.H. Driggers, D.L. Ennist, S.L. Gleason, J.E. Darnell, and K. Ozato. 1993. The genomic structure of the murine ICSBP gene reveals the presence of the gamma interferon responsive element to which an ISGF3a subunit (or similar) molecule binds. *Mol. Cell. Biol.* **13:** 3951–3963.

Khan, K.D., K. Shuai, G. Lindwall, S.E. Maher, J.E. Darnell Jr., and A.L.M. Bothwell. 1993. Induction of the Ly-6A/E gene by both interferon α/β and γ requires a DNA element to which a tyrosine phosphorylated 91 kDa protein binds. *Proc. Natl. Acad. Sci.* **90:** 6806–6810.

Levy, D.E., D.S. Kessler, R.I. Pine, and J.E. Darnell Jr. 1989. Cytoplasmic activation of ISGF3, the positive regulator of interferon-α stimulated transcription, reconstituted in vitro. *Genes & Dev.* **3:** 1362–1372.

Lew, D.J., T. Decker, and J.E. Darnell Jr. 1989. Alpha interferon and gamma interferon stimulate transcription of a single gene through different signal transduction pathways. *Mol. Cell. Biol.* **9:** 5404–5411.

Lew, D., T. Decker, I. Strehlow, and J.E. Darnell Jr. 1991. Overlapping elements in the GBP gene promoter mediate transcriptional induction by alpha and gamma interferon. *Mol. Cell. Biol.* **11:** 182–191.

Muller, M., C. Laxton, J. Briscoe, C. Schindler, T. Improta, J.E. Darnell Jr., G.R. Stark, and I.M. Kerr. 1993. Complementation of a mutant cell line: Central role of the 91-kDa polypeptide of ISGF3 in the interferon-α and -γ signal transduction pathway. *EMBO J.* **12:** 4221–4228.

Muller, C.W., F.A. Rey, M. Sodeoka, G.L. Verdine, and S.C. Harrison. 1995. Structure of the NF-κB p50 homoder bound to DNA. *Nature* **373:** 311–317.

Pearse, R.N., R. Feinman, K. Shuai, J.E. Darnell, and J.V. Ravetch. 1993. Interferon gamma-induced transcription of the high affinity Fc receptor for IgG requires assembly of a complex that includes the 91 kDa subunit of transcription factor ISGF-3. *Proc. Natl. Acad. Sci.* **90:** 4314–4318.

Pellegrini, S., J. John, M. Shearer, I.M. Kerr, and G.R. Stark. 1989. Use of a selectable marker regulated by alpha interferon to obtain mutations in the signaling pathway. *Mol. Cell. Biol.* **9:** 4605–4612.

Pollock, R. and R. Triesman. 1990. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* **18:** 6197–6204.

Sadowski, H.B. and M.Z. Gilman. 1993. Cell-free activation of a DNA-binding protein by epidermal growth factor. *Nature* **362:** 79–83.

Sadowski, H.B., K. Shuai, J.E. Darnell Jr., and M.Z. Gilman. 1993. A common nuclear signal transduction pathway activated by growth factor and cytokine receptors. *Science* **261:** 1739–1744.

Schindler, C., X.-Y. Fu, T. Improta, R. Aebersold, and J.E. Darnell Jr. 1992a. Proteins of transcription factor ISGF-3: One

gene encodes the 91 and 84 kDa ISGF-3 proteins that are activated by inteferon-α. *Proc. Natl. Acad. Sci.* **89:** 7836–7839.

Schindler, C., K. Shuai, V.R. Prezioso, and J.E. Darnell Jr. 1992b. Interferon-dependent tyrosine phosphorylation of a latent cytoplasmic transcription factor. *Science* **257:** 809–815.

Shuai, K., C. Schindler, V.R. Prezioso, and J.E. Darnell Jr. 1992. Activation of transcription by IFN-γ: Tyrosine phosporylation of a 91 kD DNA binding protien. *Science* **259:** 1808–1812.

Shuai, K., G.R. Stark, I.M. Kerr, and J.E. Darnell Jr. 1993a. A single phosphotyrosine residue of Stat91 required for gene activation by interferon-γ *Science* **261:** 1744–1746.

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

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

Strehlow, I., D. Seegert, C. Frick, F.-C. Bange, C. Schindler, E.C. Boettger, and T. Decker. 1993. The gene encoding IFP 53/ tryptophanyl tRNA synthetase is regulated by the gamma interferon activation factor. *J Biol. Chem.* **260:** 15280–15285.

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

Wegenka, U.M., J. Buschmann, C. Lutticken, P.C. Heinrich, and F. Horn. 1993. Acute-phase response factor, a nuclear factor binding to acute-phase response elements, is rapidly activated by interleukin-6 at the posttranslational level. *Mol. Cell. Biol.* **13:** 276–288.

Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan. 1993. Cooperative dimerization of paired class homeo domains on DNA. *Genes & Dev.* **7:** 2120–2134.

Wong, P., C.W. Severns, N.B. Guyer, and T.M. Wright. 1994. A unique palindromic element mediates gamma interferon induction of mig gene expression. *Mol. Cell. Biol.* **14:** 914–922.

Yuan, J., U.M. Wegenka, C. Hutticken, J. Buschmann, T. Decker, C. Schindler, P. C. Heinrich, and F. Horn. 1994. The signalling pathways of interleukin-6 and gamma interferon converge by the activation of different transcription factors which bind to common responsive DNA elements. *Mol. Cell. Biol.* **14:** 1657–1668.

Zhong, Z., Z. Wen, and J.E. Darnell Jr. 1994a. Stat3 and Stat4: Members of the family of signal transducers and activators of transcription. *Proc. Natl. Acad. Sci.* **91:** 4806–4810.

———. 1994b. Stat3: A STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science* **264:** 95–98.

A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain.

C M Horvath, Z Wen and J E Darnell

References	This article cites 38 articles, 21 of which can be accessed free at: **http://genesdev.cshlp.org/content/9/8/984.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**