

# The amino-terminal domain of human signal transducers and activators of transcription 1: Overexpression, purification and characterization

ARATI PRABHU, EVANS COUTINHO\* and SUDHA SRIVASTAVA<sup>†</sup>

Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Mumbai 400 098, India

<sup>†</sup>Tata Institute of Fundamental Research, Homi Bhabha Road, Navy Nagar, Colaba, Mumbai 400 005, India

\*Corresponding author (Fax, 91-22-26670816; Email, evans@bcplindia.org)

The dual functional signal transducers and activators of transcription (STAT) proteins are latent cytoplasmic transcription factors that play crucial roles in host defense. Animals that lack these proteins are highly susceptible to microbial and viral infections and chemically induced primary tumours. We have over expressed the amino-terminal domain of human STAT1 (hSTAT1) in *Escherichia coli* and purified it by affinity chromatography and gel filtration chromatography. The entire process has been monitored by gel electrophoresis. The pure protein has been characterized by mass spectrometry and 2-dimensional nuclear magnetic resonance (2D-NMR) spectroscopy. Our results indicate that the N-terminus of hSTAT1 exists as a dimer in solution.

[Prabhu A, Coutinho E and Srivastava S 2005 The amino-terminal domain of human signal transducers and activators of transcription 1: Overexpression, purification and characterization; *J. Biosci.* **30** 611–618]

## 1. Introduction

Interferons (IFNs) play key roles in a variety of cellular functions in the human body such as mediating antiviral responses, regulation of cell growth and tumour surveillance (Bach *et al* 1997; Boehm *et al* 1997; Darnell *et al* 1994; Stark *et al* 1998). They are also involved in modulating immune responses (Durbin *et al* 1996; Meraz *et al* 1996). IFNs are essential for the survival of higher vertebrates because they provide an early line of defense against viral infections – hours to days before immune responses commence. The main IFN signalling pathways are rapid and direct and involve the activation of signal transducers and activators of transcription (STAT) fac-

tors by Janus tyrosine kinases (JAKs) at the cell membrane. This is followed by the release of the STAT proteins and their migration to the nucleus where they accumulate to drive transcription.

At least seven STATs and four JAKs have already been identified (Darnell 1997). The JAKs activate the STAT proteins through phosphorylation of a specific tyrosine residue (Darnell *et al* 1994; Schindler and Darnell 1995). Thereafter, the STAT proteins form homo- and heterodimers through mutual phosphotyrosine Src homology region 2 (SH2) interactions (Qureshi *et al* 1995; Shuai *et al* 1994). This dimerization of STAT is an essential step as it provides a means of disengagement from the receptor or kinase, following which the STAT proteins migrate

**Keywords.** Affinity chromatography; amino-terminal domain; SDS-PAGE electrophoresis; STAT proteins

Abbreviations used: CBP, CREB binding protein; CREB protein, cyclic-AMP response element binding protein; DTT, dithiothreitol; 2D-NMR, 2-dimensional nuclear magnetic resonance; EDTA, ethylene diamine tetraacetate; FPLC, fast performance liquid chromatography; GST, glutathione S-transferase; HMM, hidden Markov model; HSQC, heteronuclear single quantum coherence; IFN, interferons; IPTG, isopropyl *b*-D-thiogalactopyranoside; JAK, Janus tyrosine kinases; MALDI-TOF, matrix assisted laser desorption/ionization-time of flight; N-terminal, amino-terminal; PBS, phosphate buffered saline; pI, isoelectric point; SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis; SH2, Src homology region 2; STAT, signal transducers and activators of transcription; human STAT1 (hSTAT1).

to the nucleus. Here they induce the expression of many gene products involved in immune responses, cell growth and differentiation, cell survival and apoptosis. STAT dimers interact with each other and with several different transcription factors, primarily through the N-terminal domain (Kumar *et al* 1997; Xu *et al* 1996). Thus, development of small molecules that bind to the amino-terminal domain of the STAT protein so as to prevent its dimerization might be a useful strategy in the development of immunosuppressant drugs.

All immunosuppressants currently used for organ transplant and autoimmune diseases have ubiquitously expressed targets. This results in major complications during therapy, and serious side effects such as nephrotoxicity, neurotoxicity, bone loss, cataracts, new-onset post-transplant diabetes mellitus, hyperlipidaemia and hypertension (Hong and Kahan 2000). Targeting the JAK-STAT proteins which are restricted in their expression to immune cells would provide effective immunosuppression without the toxicity that is associated with present therapies.

The physiological relevance of the JAK-STAT pathway in IFN $\gamma$  signalling has been established unequivocally through the generation and characterization of mice with a targeted disruption of the *STAT1* gene (Durbin *et al* 1996; Meraz *et al* 1996). STAT1-knockout mice show normal tissue and organ development, produce normal numbers and distributions of immune cell populations, and are able to reproduce. However, cells from these mice are incapable of manifesting any biologic responses to either IFN $\gamma$  or IFN $\alpha$  and the mice display severe defects in their ability to resist microbial and viral infections.

Recent publications have established important connections between STAT1 and the CBP/p300 transcription factors (CREB binding protein). Specifically, two interaction regions have been identified: the amino terminal (N-terminal) region which interacts with the (cyclic-AMP response element binding protein) CREB-binding domain of CBP/p300 and the C-terminal region of STAT1 which interacts with the domain of CBP/p300 which is also known to be involved in the binding to the adenovirus protein E1A (Zhang *et al* 1996). As CBP/p300 binding is required for the adenovirus E1A protein to regulate transcription of many genes during viral replication and cellular transformation, it is possible that the antiviral effect of IFN $\gamma$  is based at least in part on direct competition between nuclear STAT1 and E1A for CBP/p300 binding.

Thus the N-terminal domain of STAT1 seems to be crucial to its intrinsic activity. As a first step towards understanding the various STAT1 interactions, we have over expressed the amino-terminal domain of STAT1 in *Escherichia coli* as a glutathione S-transferase (GST)-tagged fusion protein. The strategy of affinity chromatography to selectively extract the fusion protein, cleavage

of the tag and further purification on a gel filtration column yielded a protein with a high degree of purity. This protein has been characterized by mass spectrometry and 2-dimensional nuclear magnetic resonance (2D-NMR) spectroscopy and was found to have a well ordered conformation in solution.

## 2. Materials and methods

The N-terminal domain of human STAT1 (hSTAT1) comprises 128 amino acid residues and has a molecular weight of 15.56 kDa. Its theoretical isoelectric point (pI) is 4.98 and it has an extinction coefficient of 23470 ( $M^{-1} cm^{-1}$ ) at 280 nm (<http://expasy.org/cgi-bin/protparam>).

### 2.1 Expression and purification of N-terminal hSTAT1

**2.1a Transformation:** The cDNA encoding the N-terminal 128 amino acids of hSTAT1 fused to a glutathione S-transferase (GST) moiety was cloned into the expression vector pGEX-2T (Amersham Biosciences, USA) and used to transform the *E. coli* strain BL21 (DE3, Novagen) using the method of Hanahan (1983).

**2.1b Protein expression:** The transformants were checked for protein expression by growing them at 37°C in 10 ml of M9 minimal medium (Sambrook and Russell 2001) containing 100  $\mu$ g/ml of ampicillin to mid log phase as indicated by the absorbance of 0.6 in the visible range at 600 nm ( $A_{600}$ ). Following this, the inducer isopropyl *b*-D-thiogalactopyranoside (IPTG) (Sigma-Aldrich Co., USA) was added to a final concentration of 1 mM and the cells were grown for an additional 5 h. Two such transformants were screened for protein expression on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions as described by Laemmli (1970). Reducing PAGE analysis was performed in a 12% polyacrylamide gel in the presence of SDS after reduction of protein by dithiothreitol (DTT) and heating at 95°C.

Cells from single, ampicillin-resistant transformed colonies of *E. coli* were grown to saturation at 37°C in M9 minimal media containing 0.68%  $Na_2HPO_4$  (Sigma-Aldrich Co. USA); 0.3%  $KH_2PO_4$  (SD Fine-Chem Ltd., India); 0.06% NaCl (Qualigens Fine Chemicals, India); 0.1%  $NH_4Cl$  (SRL, India) and supplemented with 0.4% glucose (Qualigens Fine Chemicals, India); 2 mM  $MgSO_4$  (E Merck, India); 0.1 mM  $CaCl_2$  (SRL, India); 0.04% CAS amino acids (Sigma Chemical Co., USA); and 0.01% ampicillin (Sigma Chemical Co., USA). The saturated cultures were diluted 100-fold in the same medium and grown in shake-flasks to mid-log phase ( $A_{600} = 0.6$ ) at which time isopropyl *b*-D-thiogalactopyranoside (IPTG)

was added to a final concentration of 1 mM. After 5 h, the cells were harvested by centrifugation at 4500 rpm for 1 h using a Beckman Sorvall (GSA rotor) centrifuge.  $^{15}\text{N}$ -labelled N-terminal hSTAT1 was expressed in M9 minimal medium containing 0.1% of  $^{15}\text{NH}_4\text{Cl}$  (Spectra Stable Isotopes, USA).

**2.1c Preparation of the cell extract:** *E. coli* cells containing N-terminal hSTAT1 (1–128) were diluted 100-fold in phosphate buffered saline (PBS) (Sambrook and Russell 2001) containing 0.1% lysozyme (Sigma Chemical Co., USA) and incubated at 4°C for 30 min. Triton X-100 (Roche Diagnostics Corporation, USA) at 0.001% concentration was forcibly injected into the cell lysate. DNase (Amersham Biosciences, USA) and RNase (Amersham Biosciences, USA) were added to the solution to a final concentration of 1 µg/ml and incubation was continued for a further 30 min. DTT (Sigma-Aldrich Co., USA) was added to a final concentration of 1 mM and the cells were ruptured by sonication while cooling on ice with seven pulses at full power with a gap of at least 2 min between pulses. A soluble extract was prepared by centrifuging the disrupted cell suspension at 8500 rpm for 1 h at 4°C. The supernatant was filtered through a sterile 0.45 µm syringe filter.

**2.1d Purification of the fusion protein:** The clear cell lysate was combined with glutathione-agarose beads (50% slurry in PBS, Amersham Biosciences, USA) in the proportion of one part of beads to thousand parts of cell culture volume and shaken gently at 4°C for 2 h. The unbound proteins were washed from the beads with PBS buffer till the UV absorption of the eluate at 280 nm showed no presence of protein.

The protein profile of each step (cell culture, cell extract and washes) was analysed by SDS-PAGE (Davis 1964; Ornstein 1964).

**2.1e Proteolytic cleavage of the target protein from the bound GST moiety:** Thrombin protease (Amersham Biosciences, USA) in the proportion of 50 units per ml of glutathione agarose resin was added to the beads in 1 ml of PBS. The mixture was agitated gently and incubated for 24 h at room temperature. The cleaved protein was eluted from the beads using PBS and concentrated to 2 ml by ultrafiltration through a 3 kDa membrane filter. The processes of thrombin cleavage and sample concentration were monitored by SDS-PAGE.

**2.1f Gel filtration chromatography:** A 2 ml aliquot of the concentrated protein solution (concentration 0.25 mM) was injected onto a Superdex Hi load-16/60 gel filtration FPLC column (Amersham Biosciences, 60 cm, 120 ml) and developed under a flow rate of 1 ml/min of column

buffer [PBS, 1 mM EDTA (SRL, India), 1 mM DTT]. N-terminal hSTAT1 eluted as a dimer appearing in the mobile phase fractions collected between 60 and 80 ml after the start of the chromatogram. The molecular weight of the eluted protein was confirmed by injecting a standard sample (carbonic anhydrase; MW 29 kDa) and developing it under similar conditions as the sample protein.

The fractions containing the N-terminal STAT dimer were concentrated to a final volume of 500 µl using ultrafiltration (concentration 0.5 mM). The purity of the protein sample was confirmed by SDS-PAGE. The concentration of the pure protein in the final sample was determined from its absorbance at 280 nm.

## 2.2 Characterization

The purified protein was characterized by matrix assisted laser desorption/ionization (MALDI) and NMR spectroscopy.

**2.2a MALDI:** The MALDI spectrum of the purified protein was recorded on a Micromass MALDI-TOF (linear) mass spectrometer. A 10 µM solution mixed with a Sinnapinnic acid matrix was used for acquiring the mass profile.

**2.2b NMR spectroscopy:** NMR experiments were recorded on a Varian Unity Plus 600 MHz NMR spectrometer equipped with pulsed field gradients. A [ $^1\text{H}$ - $^{15}\text{N}$ ] HSQC (Kay *et al* 1992; Palmer *et al* 1991; Schleucher *et al* 1994) spectrum was recorded at 25°C.  $^1\text{H}$  shifts were referenced with respect to the water signal at 4.702 ppm and  $^{15}\text{N}$  chemical shifts were referenced indirectly (Wishart *et al* 1995). The protein had all its nitrogens uniformly labelled with  $^{15}\text{N}$  as  $^{15}\text{NH}_4\text{Cl}$  was used as the sole source of nitrogen during cell growth and protein expression. The NMR spectrum was recorded at a concentration of 0.5 mM at a pH of 6.5.

**2.2c Secondary structure prediction:** Several algorithms were used for the prediction of the secondary structure of N-terminal hSTAT1. Protein databases of functional motifs (PROSITE) (Bairoch *et al* 1997) and putative protein domains (PRODOM) (Sonnhammer and Kahn 1994) were searched for motifs and domains which correspond to the N-terminal STAT1 protein. Chemically conserved amino acid residues were identified and the secondary structural elements were predicted using different algorithms like hidden Markov model (HMM) (Karplus *et al* 1998), neural networks (Baldi *et al* 1999; Pollastri *et al* 2002) and multiple sequence alignments (Sander and Schneider 1991).

### 3. Results

*E. coli* BL21/DE3 cells containing the GST-fused N-terminal STAT1 expression vector (pGEX-2T) were grown to mid log phase in minimal M9 media at 37°C and induced for several hours with IPTG after which the total, soluble protein was extracted in the lysis buffer by lysozyme treatment and sonication. Small aliquots were analysed by SDS-PAGE at various stages along with samples from the uninduced and induced BL21/DE3 cells. The results are shown in figure 1. These indicate that upon induction with IPTG, GST-tagged N-terminal STAT1 (42 kDa) accumulates to high concentrations (figure 1, lane 2) and virtually all of this material is soluble in the crude cell extract (figure 1, lane 3).

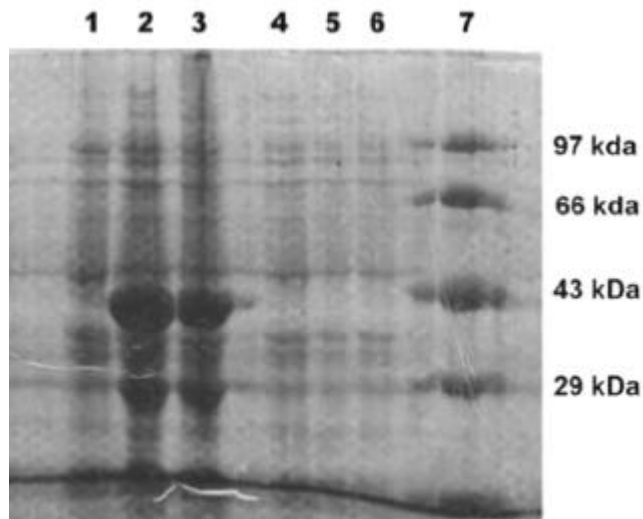
The cell extract was sonicated and the clear sonicate was allowed to react with glutathione agarose beads for affinity binding of the GST-tagged protein. Fusion proteins expressed from pGEX vectors containing a GST moiety can be purified to near homogeneity by affinity chromatography on glutathione-agarose (Smith and Johnson 1988). As the affinity of GST for its substrate is in the submillimolar range, immobilization of glutathione on an agarose matrix makes a highly efficient affinity chromatography resin. The eluent from the previous operation which contains the unbound fraction shows complete absence of any GST-tagged protein (figure 1, lane 4), which proves the efficiency of the affinity binding process. The glutathione agarose beads were washed with the

binding buffer (PBS) till no residual material appeared in the effluent. This was confirmed by checking the absorbance of the eluate at 280 nm.

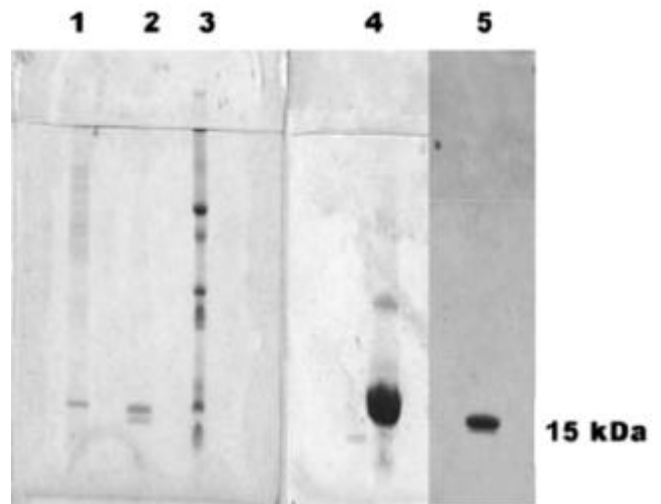
The GST fragment was then removed from the tagged protein and the untagged protein was eluted out from the beads in a single step by treatment with thrombin protease for 24 h. Fusion proteins containing a thrombin recognition site can be easily cleaved from the fused tag with thrombin protease (Gearing *et al* 1989; Smith and Johnson 1988). Small aliquots of the cleaved protein were tested at intervals on SDS-PAGE shown in figure 2 (lanes 1 and 2). Completion of thrombin cleavage was confirmed by eluting the fraction containing GST tag from the beads with elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) and analysing by SDS-PAGE. Complete cleavage of GST-tagged protein is indicated by absence of a band at 42 kDa (GST-tagged STAT1).

The cleaved protein was concentrated by ultrafiltration to a volume of 2 ml (0.25 mM) and then subjected to gel filtration chromatography. A small aliquot was checked by SDS-PAGE (figure 2, lane 4) which shows that the sample had a high degree of purity.

The results of gel filtration chromatography are shown in figure 3. The FPLC chromatogram shows that the N-terminal STAT1 exists as a dimer with a molecular weight slightly more than 29 kDa. The fractions from the FPLC eluate, which contained the dimer, were concentrated by ultrafiltration to a volume of 500 µl (concentration



**Figure 1.** Reducing SDS-PAGE analysis during protein purification. Lane 1, noninduced cell culture; lane 2, cell culture after induction with IPTG; lane 3, cell lysate after lysozyme treatment and sonication; lane 4, unbound fraction after affinity chromatography; lanes 5 and 6, eluate after washing with 5 ml and 10 ml PBS respectively; lane 7, molecular weight marker.



**Figure 2.** Reducing SDS-PAGE analysis during the processes of thrombin cleavage and sample concentration (before and after gel filtration chromatography). Lane 1, eluate after 8 h of treatment with thrombin protease; lane 2, eluate after 24 h of treatment with thrombin protease; lane 3, molecular weight marker; lane 4, protein concentrate after affinity chromatography; lane 5, protein concentrate after gel filtration chromatography (FPLC).

0.5 mM) and its purity was checked by SDS-PAGE (figure 2, lane 5). This final sample shows a single band at 15 kDa. Thus, the dimeric FPLC fraction of N-terminal STAT1 appears as a monomer on the denaturing SDS-PAGE gel. By the above method, 2 mg of the purified protein could be recovered per liter of the starting culture.

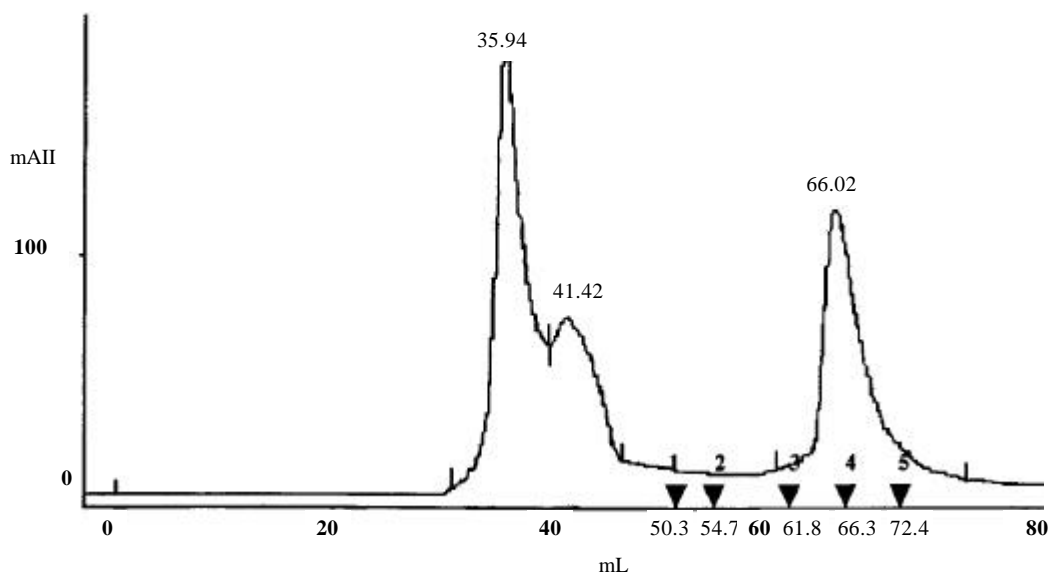
The pure protein was characterized by mass spectrometry (MALDI) and 2-dimensional heteronuclear NMR spectroscopy. The MALDI spectrum in figure 4 shows a predominant peak at about 15 kDa. This leads to the inference that in dilute solutions such as that required for the mass measurement, the protein could exist as a monomer. However this monomeric state could also be an artifact of ionization during the mass spectrometric measurement.

#### 4. Discussion

The [ $^1\text{H}$ - $^{15}\text{N}$ ] HSQC NMR spectrum in figure 5 shows a good dispersion of the  $\text{H}^{\text{N}}$ -N resonances including the backbone amide and the side chain amino protons. This indicates that N-terminal STAT1 has a well-structured conformation in solution. However, even after making allowances for some spectral overlap, the number of peaks in the HSQC spectrum corresponds more closely to the monomer than a dimeric molecule. This could also

arise from a symmetric dimer for this protein. A similar result was observed for the N-terminal domain of human STAT4 (Baden *et al* 1998). An asymmetrical arrangement of the secondary structural elements in the molecule would have resulted in the presence of almost twice as many  $\text{H}^{\text{N}}$ -N resonances. Recently the crystal structure of unphosphorylated STAT1 (1-683) complexed with a phosphopeptide derived from the alpha chain of interferon gamma (IFN $\gamma$ ) receptor, has been determined by X-ray crystallography (Mao *et al* 2005). Analyses of both the wild-type and mutant STAT1 proteins have shown that STAT1 is predominantly dimeric prior to activation, and the dimer is mediated by interactions of the N-terminal domain. Our FPLC and NMR show that the dimeric state is indeed a reality for the amino terminal domain of STAT1.

All the methods used for structure prediction indicate the presence of six helices spread through the length of the protein. In some of the methods, the last three helices towards the C-terminal end show discontinuous stretches of alpha features with intervening loops. This results in an increase in the number of helices to seven or eight. The beginning and end positions of each helix also varies among the different methods by a few residues. A consensus prediction (figure 6) over all the methods returned a structure which could be classified as 'all-alpha' with about 73% of the protein being made up of alpha helices and the rest being characterized as loop regions.



**Figure 3.** Fast performance liquid chromatogram (FPLC) of N-terminal STAT1 purified by affinity chromatography. Arrows at bottom of plot shows the elution profile of protein molecular weight standards. (Gel filtration molecular weight markers, Product code MW-GF-200, Sigma.) (1) *b*-Amylase (mw 200 kDa) elutes at 50.3 ml. (2) Alcohol dehydrogenase (mw 150 kDa) elutes at 54.7 ml. (3) Bovine serum albumin (mw 66 kDa) elutes at 61.8 ml. (4) Carbonic anhydrase (mw 29 kDa) elutes at 66.3 ml. (5) Cytochrome C (mw 12.4 kDa) elutes at 72.4 ml.

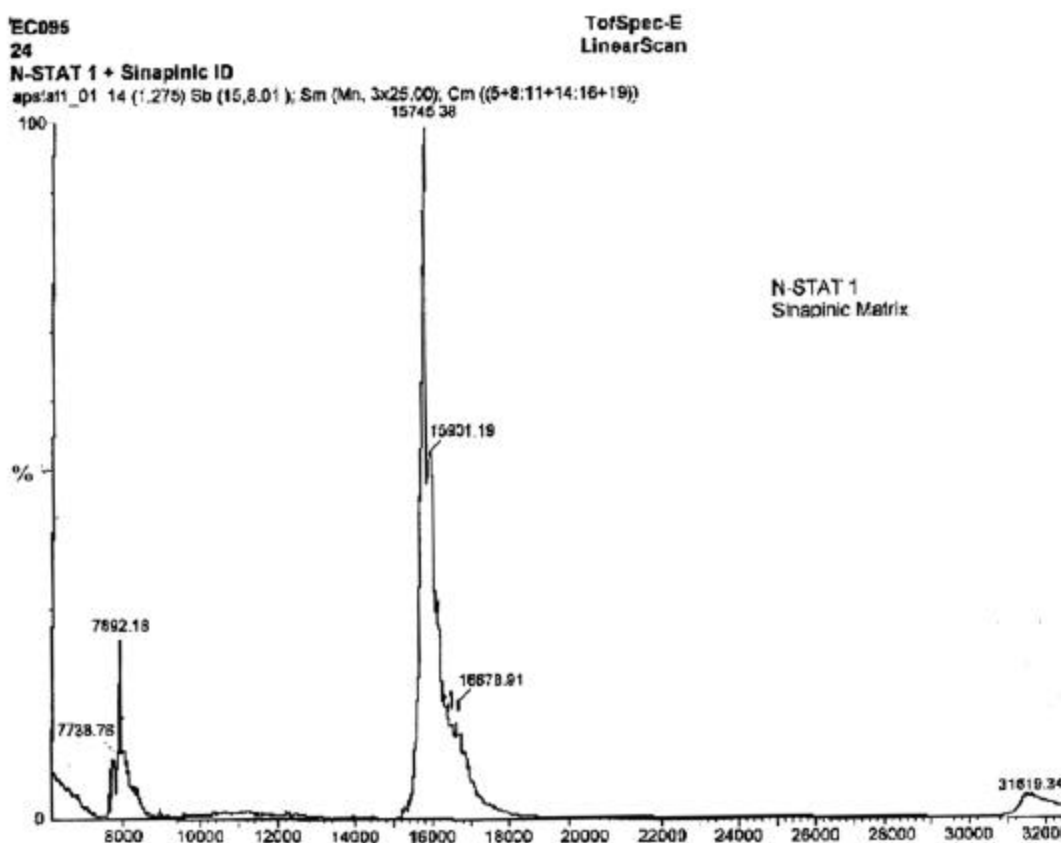


Figure 4. Mass spectrum (MALDI) of N-terminal STAT1.

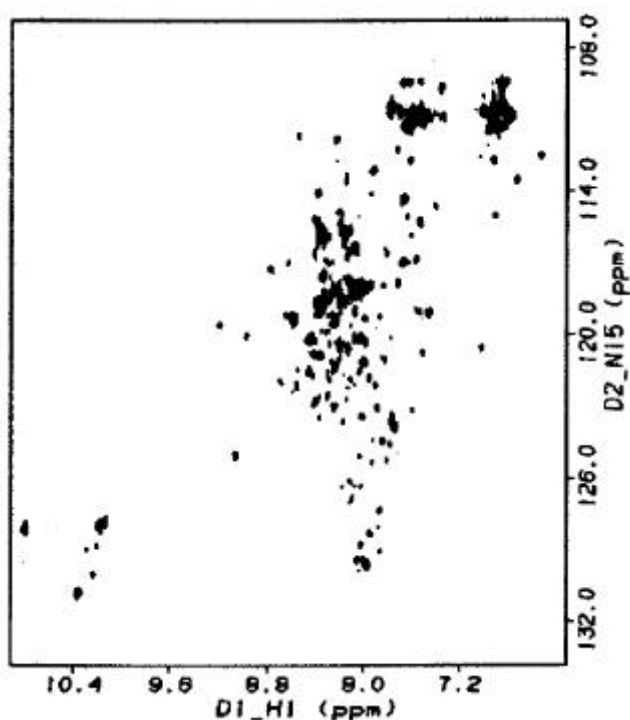
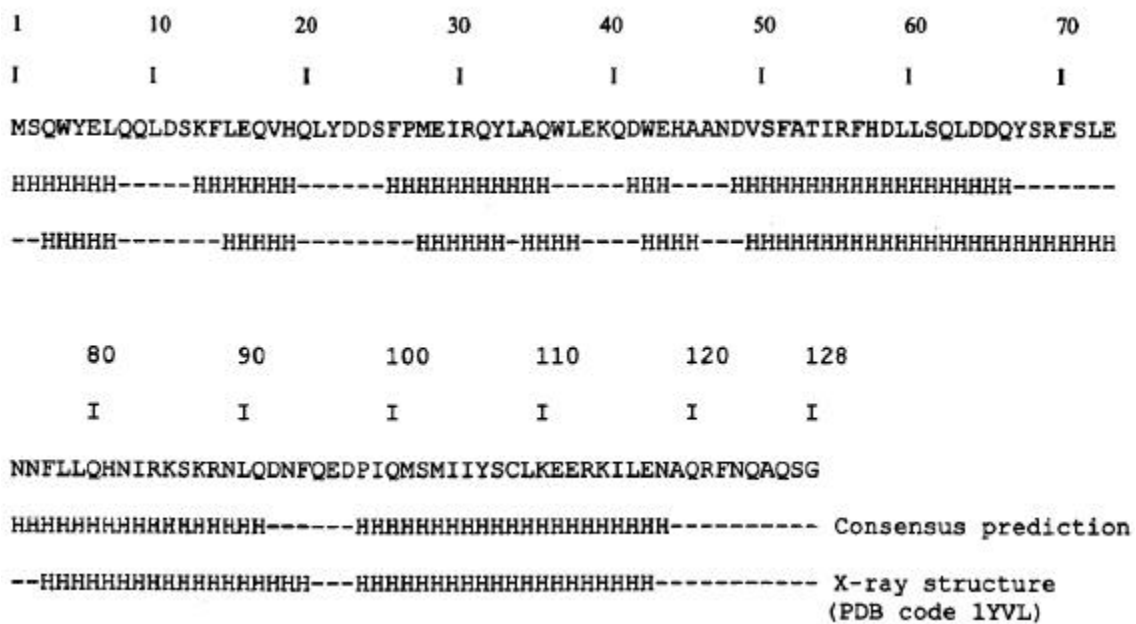


Figure 5.  $[^{15}\text{N}-^1\text{H}]$  HSQC spectrum of N-terminal STAT1.

The various domains of the signal transducers and transcription activator factors have generated a lot of interest in the last few years for their individual contribution towards biological function (Becker *et al* 1998; Chen *et al* 1998; Darnell *et al* 1994). The N-terminal domain of STAT1 has been shown to be stable to proteolytic treatments (Vinkemeier *et al* 1996). This coupled with the fact that the N-termini of the various STATs share high sequence homology indicates a significant role for the N-terminal amino acids in STAT signalling (Becker *et al* 1998; Chen *et al* 1998; Murphy *et al* 2000).

The crystal structure of the amino-terminus of STAT4 (residues 1–130) has been solved independently of the core structure and shows a series of eight short helices (Vinkemeier *et al* 1998). The X-ray crystal structure of unphosphorylated STAT1 (pdb code 1YVL) as well as several secondary structure prediction algorithms, point to a similar conformation for the amino-terminal domain of STAT1. We have shown that the N-terminal domain of STAT1 can be purified by affinity chromatography. Moreover, our results indicate that the protein has an inherent tendency to dimerize. This molecule is a promising candidate for detailed NMR characterization for its 3-dimensional structure as evidenced by our  $[^1\text{H}-^{15}\text{N}]$  HSQC NMR results.



**Figure 6.** Comparison of consensus prediction of the secondary structure of the N-terminal domain of STAT1 with the X-ray crystal structure of the amino terminus of unphosphorylated STAT1 (pdb id. 1YVL).

### References

- Bach E A, Aguet M and Schreiber R D 1997 The IFN gamma receptor: a paradigm for cytokine receptor signaling; *Annu. Rev. Immunol.* **15** 563–591
- Baden H A, Sarma S P, Kapust R B, Byrd R A and Waugh D S 1998 The Amino-terminal domain of human STAT4. Overproduction, purification and biophysical characterization; *J. Biol. Chem.* **273** 17109–17114
- Bairoch A, Bucher P and Hofmann K 1997 The PROSITE database, its status in 1997; *Nucleic Acids Res.* **25** 217–221
- Baldi P, Brunak S, Frasconi P, Pollastri G and Soda G 1999 Exploiting the Past and the Future in Protein Secondary Structure Prediction; *Bioinformatics* **15** 937–946
- Becker S, Groner B and Muller C W 1998 Three-dimensional structure of STAT3b homodimer bound to DNA; *Nature (London)* **394** 145–151
- Boehm U, Klamp T, Groot M and Howard J C 1997 Cellular responses to interferon gamma; *Annu. Rev. Immunol.* **15** 749–795
- Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell J E Jr and Kuriyan J 1998 Crystal structure of a tyrosine phosphorylated STAT1 dimer bound to DNA; *Cell* **93** 827–839
- Darnell J E Jr 1997 STATS and gene regulation; *Science* **277** 1630–1635
- Darnell J E Jr, Kerr I and Stark G M 1994 JAK-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins; *Science* **264** 1415–1421
- Davis B J 1964 Disc electrophoresis. II. Method and application to human serum proteins; *Ann. N.Y. Acad. Sci.* **121** 404–427
- Durbin J E, Hackenmiller R, Simon M C and Levy D E 1996 Targeted disruption of the mouse STAT1 gene results in compromised innate immunity to viral disease; *Cell* **84** 443–450
- Gearing D P, Nicola N A, Metcalf D, Foote S, Willson T A, Gough N M and Williams R L 1989 Production of leukemia factor in *Escherichia coli* by a novel procedure and its use in maintaining embryonic stem cells in culture; *Bio/Technology* **7** 1157–1161
- Hanahan D 1983 Studies on transformation of *Escherichia coli* with plasmids; *J. Mol. Biol.* **166** 557–580
- Hong J C and Kahan B D 2000 Immunosuppressive agents in organ transplantation: past, present and future; *Semin. Nephrol.* **20** 108–125
- Karplus K, Barrett C and Hughey R 1998 Hidden Markov Models for Detecting Remote Protein Homologies; *Bioinformatics* **14** 846–856
- Kay L E, Keifer P and Saarinen T 1992 Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity; *J. Am. Chem. Soc.* **114** 10663–10665
- Kumar A, Commane M, Flickinger T W, Horvath C M and Stark G R 1997 Defective TNF-alpha-induced apoptosis in STAT1 null-cells due to low constitutive levels of caspases; *Science* **278** 1630–1632
- Laemmli U K 1970 Cleavage of structural proteins during the assembly of the head of bacteriophage T4; *Nature (London)* **227** 680–685
- Mao X, Ren Z, Parker G N, Sondermann H, Pastorello M A, Wang W, McMurray J S, Demeler B, Darnell J E Jr and Chen X 2005 Structural bases of unphosphorylated STAT1 association and receptor binding; *Mol. Cell.* **17** 761–771
- Meraz M A, White J M, Sheehan K C, Bach E A, Rodig S J, Dighe A S, Kaplan D, Riley J K, Greenlund A C, Campbell D, Carver-Moore K, Du Bois R N, Clark R, Aguet M and Schreiber R D 1996 Targeted disruption of the STAT1 gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway; *Cell* **84** 431–442
- Murphy T L, Geisal E D, Farrar D and Murphy K M 2000 Role

- of the STAT 4 N-domain in receptor proximal tyrosine phosphorylation; *Mol. Cell. Biol.* **20** 7121–7131
- Ornstein L 1964 Disc Electrophoresis. I. Background and theory; *Ann. N.Y. Acad. Sci.* **121** 321–349
- Pollastri G, Przybylski D, Rost B and Baldi P 2002 Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles; *Proteins* **47** 228–235
- Palmer III A G, Cavanagh J, Wright P E and Rance M 1991 Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy; *J. Magn. Reson.* **93** 151–170
- Qureshi S A, Salditt-Georgieff M and Darnell J E Jr 1995 Tyrosine phosphorylated STAT1 and STAT2 plus a 48 kDa protein all contact DNA in forming the interferon stimulated gene factor 3 (ISGF3); *Proc. Natl. Acad. Sci. USA* **92** 3829–3833
- Sambrook J and Russell D W 2001 *Molecular cloning: A laboratory manual* Third Edition (New York: Cold Spring Harbor Laboratory Press)
- Sander C and Schneider R 1991 Database of homology-derived protein structures and the structural meaning of sequence alignment; *Proteins* **9** 56–68
- Schindler C and Darnell J E Jr 1995 Transcriptional responses to polypeptide ligands: The JAK-STAT pathway; *Annu. Rev. Biochem.* **64** 621–651
- Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser S J, Sorensen O W and Griesinger C 1994 A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients; *J. Biomol. NMR* **4** 301–306
- 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 STAT1 involves dimerisation through SH2-phosphotyrosyl peptide interactions; *Cell* **76** 821–828
- Smith D B and Johnson K S 1988 Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase; *Gene* **67** 31–40
- Sonnhammer E L L and Kahn D 1994 Modular arrangement of proteins as inferred from analysis of homology; *Protein Sci.* **3** 482–492
- Stark G R, Kerr I M, Williams B R, Silverman R H and Schreiber R D 1998 How cells respond to interferons; *Annu. Rev. Immunol.* **67** 227–264
- Vinkemeier U, Cohen S L, Moarefi I, Chait B T, Kuriyan J and Darnell J E Jr 1996 DNA binding of in-vitro activated STAT1 alpha, STAT1 beta and truncated STAT1: interaction between amino-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 STAT4; *Science* **279** 1048–1052
- Wishart D S, Bigam C G, Yao J, Abilgaard F, Dyson H J, Oldfield E, Markley J L and Sykes B D 1995 <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR; *J. Biomol. NMR* **6** 135–140
- Xu X A, 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 gamma signaling; *Proc. Natl. Acad. Sci. USA* **93** 15092–15096

MS received 9 May 2005; accepted 21 October 2005

ePublication: 9 November 2005

Corresponding editor: SHAHID JAMEEL