Expression of a tyrosine phosphorylated, DNA binding Stat3β dimer in bacteria

Stefan Becker[a], Garry L. Corthals[1,b], Ruedi Aebersold[b], Bernd Groner[2,c], Christoph W. Müller[a,*]

[a]*European Molecular Biology Laboratory (EMBL), Grenoble Outstation, c/o ILL, P.O. Box 156, 38042 Grenoble Cedex 9, France*
[b]*Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195-7730, USA*
[c]*Institute for Experimental Cancer Research, Tumor Biology Center, Breisacher Strasse 117, D-79106 Freiburg i. Br., Germany*

Received 27 October 1998

Abstract The signal transducer and activator of transcription (STAT) proteins deliver signals from the cell membrane to the nucleus. An N-terminally truncated fragment of murine Stat3β, Stat3βtc (127–722), was produced in bacteria. STAT proteins must be specifically phosphorylated at a single tyrosine residue for dimerization and DNA binding. Therefore, Stat3βtc was coexpressed with the catalytic domain of the Elk receptor tyrosine kinase. Stat3βtc was quantitatively phosphorylated by this kinase domain. Gel filtration chromatography revealed a Stat3βtc dimer. Y705 was identified as the major phosphorylated residue of Stat3βtc. This corresponds to the tyrosine residue which is phosphorylated by the Janus kinase in vivo. The phosphorylated Stat3βtc specifically bound to DNA binding sites. The described protocol allows the production of large amounts of activated protein for biochemical and pharmaceutical studies.

© 1998 Federation of European Biochemical Societies.

Key words: Stat3β; Bacterial expression; In vivo activation; Tyrosine phosphorylation

1. Introduction

Signal transducers and activators of transcription (STATs) constitute a family of eukaryotic transcription factors [1]. Their latent, cytoplasmic form is activated by phosphorylation of a single tyrosine residue [2]. This activation step is triggered by the binding of growth factors, hormones or cytokines to their receptors [3]. The tyrosine kinases involved in activation are either intrinsic activities of ligand-activated transmembrane receptors or Janus kinases which are associated with cytokine receptors [4,5]. After tyrosine phosphorylation the STATs homo- or heterodimerize, translocate to the nucleus and regulate gene expression by binding to specific response elements [2,6].

So far seven mammalian STATs have been identified. In most cases the targeted disruption of these STAT genes in mice led to phenotypes in various organs, including the mammary gland and the immune system [1,7]. Initial studies in cultured cells suggested that STATs can be activated by a large number of polypeptides. More recent studies, however, indicate that the STATs can be differentially activated depending on the subset of cell surface receptors in specific tissues [8]. Therefore the STATs combine activation by a large array of different signals with tissue dependent specificity. The STAT signalling pathway from the cell membrane to the nucleus is also a comparably direct signal transduction pathway. This has turned the STATs into important targets for drug development with the aim to modulate a large number of cellular functions.

Biochemical and molecular genetic investigations established a common structural organization of the STAT proteins. The N-terminal part of the STATs, named N-domain, is responsible for dimer dimerization on clustered DNA binding sites and interaction with other proteins [9,10]. The DNA recognition domain was located between residues 300 and 500 [11]. The highly conserved residues 600–700 form a Src homology 2 (SH2) domain [12]. This SH2 domain mediates the interaction with the cytoplasmic domain of the receptor through binding to phosphotyrosine residues as well as the dimerization of the STATs through reciprocal binding of the phosphotyrosine residue following the SH2 domain [1]. The C-terminus of the STATs contains long stretches of acidic residues which constitute the transactivation domain [9,10]. For several STATs splice variants have been identified within this domain [13–15]. Recently these studies have been matched by X-ray structures (Fig. 1). The N-domain of human Stat4 shows a hook-like structure and forms stable dimers [16]. The complementary crystal structures of N-terminally truncated fragments of the murine Stat1 and Stat3β dimers in complex with DNA have also been reported [17,18].

The availability of large amounts of active Stat protein has been a prerequisite for structural studies. In this paper we describe the expression of mg amounts of N-terminally truncated active Stat3β in *Escherichia coli* and a simple three-step purification protocol. We achieved specific phosphorylation by the coexpression of the protein-tyrosine kinase domain of the Elk receptor in bacteria[19]. The purified protein was used for co-crystallization with DNA oligonucleotides and led to the structure determination of the Stat3β:DNA complex [17]. Our procedure should be applicable to other Stat proteins and

*Corresponding author. Fax: (33) 476 207199.
E-mail: mueller@embl-grenoble.fr

[1]Present address: Garvan Institute of Medical Research, St. Vincent's Hospital, Sydney NSW 2010, Australia.

[2]Present address: Chemotherapeutisches Forschungsinstitut, Georg-Speyer-Haus, Paul-Ehrlich-Strasse 42–44, D-60596 Frankfurt a. M., Germany.

Abbreviations: STAT, signal transducer and activator of transcription; SH2, Src homology 2 domain; PCR, polymerase chain reaction; APRE, acute phase response element; DTT, dithiothreitol; Stat3βtc, a fragment of Stat3β consisting of amino acid residues 127–722; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid); CAPS, 3-(cyclohexyl-amino)-1-propanesulfonic acid; PMSF, phenylmethylsulfonyl fluoride; MS, mass spectrometry; MS/MS, tandem mass spectrometry; EGF receptor, epidermal growth factor receptor; AspN, endoproteinase Asp-N

0014-5793/98/$19.00 © 1998 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(98)01543-9

allows the large-scale expression of active Stats for structural and biochemical studies as well as large-scale drug screening.

2. Materials and methods

2.1. Expression and purification of Stat3βtc

A C-terminal fragment of the murine Stat3β cDNA, starting at residue G127, was amplified by polymerase chain reaction (PCR) using the primers 5′-dGGGATCTACTTCCATATGGGCCAGGCCAACCACC and 5′-dGGAATTCATCATTTCCAAACTGCATCAATGAATGGTGTCACACAGATGAACTTGGT (restriction sites underlined). The second primer was designed to start 20 bases upstream of the splice site between Stat3 and Stat3β [13,14], because the template for the PCR reaction was a cDNA encoding Stat3. Therefore the primer had to encode the differing seven amino acids after the splice site at P713.

The PCR fragment was digested with *Nde*I/*Eco*RI and cloned into the expression vector pET32a (Novagen). The resultant expression plasmid pET32aStat3βtc was transformed into the *E. coli* strain BL21(DE3)TBK1 (Stratagene). For expression 10 ml of an overnight culture in 2×YT broth [20], supplemented with 100 μg/ml ampicillin and 12 μg/ml tetracycline, was diluted 1/50 into 0.5 l of LB broth [20] supplemented with the same antibiotics. The culture was grown at 37°C to A_{600} = 0.3. Then the temperature was switched to 21°C. At mid-log phase (A_{600} = 0.6–0.7) the expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside. The cells were harvested 5 h after induction by centrifugation and redissolved in 1 l of kinasing medium (1×M9 [20], 1 mM $MgSO_4$, 11mM D(+)-glucose, 0.1% (w/v) casamino acids (Difco), 1.5 μM thiamine-HCl, 53 μM 3β-indoleacrylic acid), supplemented with 50 μg/ml ampicillin and 12.5 μg/ml tetracycline. The 3β-indoleacrylic acid induces the expression of the Elk receptor protein-tyrosine kinase domain under the control of the *trp* promoter [21]. This culture was grown for another 2.5 h at 37°C. Finally the cells were harvested by centrifugation and the pellet was stored at −80°C.

The cells were resuspended in ice-cold extraction buffer (20 mM HEPES-HCl, pH 7.6, 0.1 M KCl, 10% glycerol, 1 mM EDTA, 10 mM $MnCl_2$, 20 mM DTT, 0.5 mM PMSF; 10 ml/g cells). The lysis was performed by ultrasonication and the lysate was centrifuged at 27 000×*g* for 45 min at 4°C. To remove nucleic acids polyethyleneimine (0.1% final) was added to the ice-cooled, stirred supernatant. The resulting suspension was stirred on ice for another 15 min and then centrifuged for 20 min at 27 000×*g*. Ammonium sulfate was slowly added to the supernatant up to 35% saturation while the solution was stirred on ice. The precipitated protein was collected by centrifugation, redissolved in 0.3 ml of buffer D (20 mM HEPES-HCl, pH 7.0, 200 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT, 0.5 mM PMSF) and dialyzed over night at 4°C against 0.5 l of buffer D.

After dialysis the protein was loaded on a Pharmacia Superose 12 HR 10/30 column, which had been equilibrated with buffer D. The protein was eluted at a flow rate of 0.5 ml/min. The peak fractions were analyzed by a 10% SDS-polyacrylamide gel. Fractions of highest purity were combined and stored on ice for further use. Typical yields were 5 mg of >95% pure (as judged by Coomassie stain) Stat3βtc protein from 1 l of starting culture.

To obtain ^{32}P-labelled Stat3βtc, the protein was expressed in a 0.15 l culture as described above. After harvesting the cells were redissolved in 0.3 l of a modified kinasing medium in which sodium phosphate was replaced by 50 mM Tris-HCl, pH 7.0. In addition 1 mCi of [^{32}P]sodium phosphate (Amersham) was added to the culture. Otherwise the expression conditions for the Elk receptor tyrosine kinase domain were as described above. The radiolabelled Stat3βtc protein was purified by preparative 10% SDS-polyacrylamide gel electrophoresis. 30 μg of protein was loaded per well on the gel (1.5 mm). After electrophoretic separation the protein was blotted onto nitrocellulose (Schleicher and Schuell, 60 V for 1 h) in blotting buffer (10 mM CAPS, pH 11, 10% methanol). The nitrocellulose was stained with Ponceau S dye (Merck) and the bands corresponding to Stat3βtc were cut out. The nitrocellulose pieces were completely destained by washing with double distilled water and dried. The specific activity of the purified ^{32}P-labelled Stat3βtc was 0.53 μCi/pmol protein.

2.2. Enzymatic digestion of ^{32}P-labelled Stat3βtc

The radioactive filter fragments containing Stat3βtc were washed three times with H_2O, blocked with 0.5% w/v polyvinylpyrrolidone-40 (Sigma) in 100 mM acetic acid and washed again 10 times with H_2O. The filter fragments were sliced into small pieces and placed in a microcentrifuge tube. Two proteolytic digestion methods were used: trypsin was used exclusively or AspN and trypsin were added sequentially. The trypsin digestion was performed as previously described [22]. The sequential digestion with AspN/trypsin was performed in 10 mM *N*-ethylmorpholine, 1% (w/v) Zwittergent 3-16 (Calbiochem) digestion buffer, pH 7.9. The digestion buffer was added until the membrane pieces were covered, subsequently 0.5 μg AspN was added. The AspN digestion was performed overnight at 37°C. Next, 2 μg trypsin was added to the same solution and left to digest at 37°C for 3 h. After the sequential digestion the microcentrifuge tube was sonicated for 15 s. and centrifuged on a bench-top centrifuge at 14 000 rpm for 1 min. Supernatant was collected and the membrane pieces were washed once with digestion buffer (without zwittergent), sonicated and centrifuged. The supernatant was collected, pooled with the first supernatant and vacuum dried.

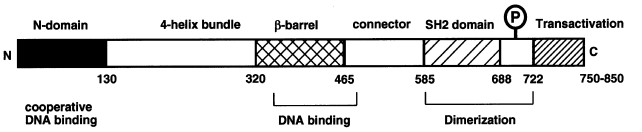


Fig. 1. Schematic diagram showing the domains of Stat3βtc as derived from the crystal structure of the Stat3βtc homodimer bound to DNA [17].

2.3. Liquid chromatography tandem mass spectrometric (LC-MS/MS) analysis of radiolabelled peptides

Peptides were separated on a Michrom Ultrafast Microprotein analyzer (Michrom Bioresources, Auburn, CA) equipped with a Michrom Reliasil, 0.5 mm diameter, 300 Å, 5 μm, C18 column. A linear gradient was developed for 45 min, at 50 μl/min from buffer A (2% CH_3CN, 98% H_2O, 0.4% CH_3COOH, 0.005% $C_4HF_7O_2$) to buffer B (80% CH_3CN, 20% H_2O, 0.4% CH_3COOH, 0.005% $C_4HF_7O_2$). The column eluent was fractionated through a splitting Tee (Valco Instruments, Houston, TX). 95% of the flow was collected for Cerenkov counting and 5% was used for MS/MS. Cerenkov counting was performed on a Beckman LS 6500. Flow splitting was essentially performed as previously described [23,24]. Mass spectrometry was conducted on a Finnigan MAT TSQ 7000 (San Jose, CA) with a standard ionization source [25]. The system was configured so that after flow splitting the remaining eluent was sprayed directly into the ionization source. Under control of the Instrument Control Language (Finnigan) software, MS/MS was performed on selected peptides above a specified threshold in an automated fashion as described previously [25]. Peptide fragmentation spectra were analyzed using SEQUEST [26], where the experimental data was correlated with theoretical spectra from the OWL database (http://www.biochem.ucl.ac.uk/bsm/

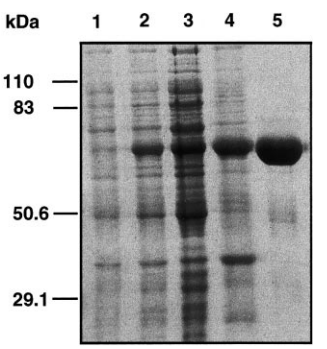


Fig. 2. Purification of the recombinant, phosphorylated Stat3βtc. An SDS-polyacrylamide (10%) gel is shown. Lane 1, uninduced Stat3βtc transformed cells; lane 2, induced Stat3βtc transformed cells; lane 3, supernatant of sonication of induced cells; lane 4, pellet of sonication of induced cells; lane 5, protein precipitated with 35% saturated ammonium sulfate at 0°C.

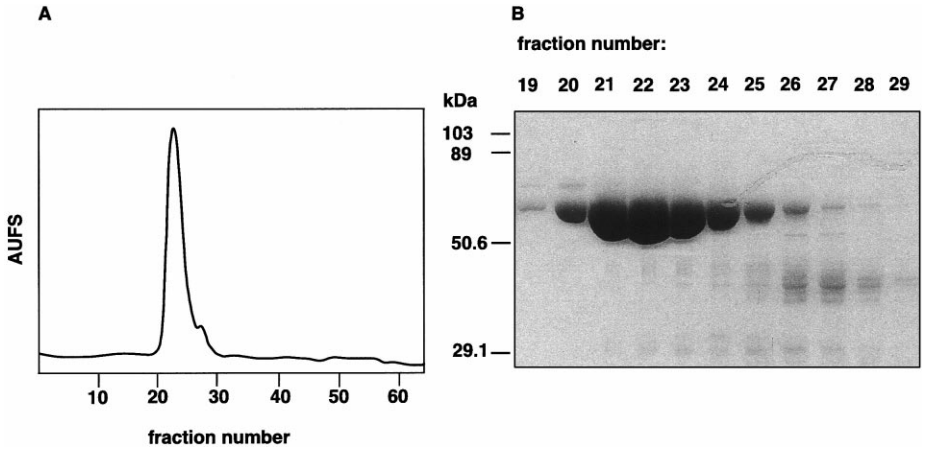


Fig. 3. Purification of Stat3βtc by gel filtration chromatography. A: After precipitation with 35% saturated ammonium sulfate and dialysis overnight against buffer D, Stat3βtc was loaded on a Pharmacia Superose 12 HR 10/30 column in buffer D. Flow rate, 0.5 ml/min; fraction size, 0.5 ml. B: A 10% SDS-polyacrylamide gel shows fractions 19–29 of the gel chromatography run (A).

dbbrowser/OWL/OWL.html) and a FASTA formatted file containing only the STAT3βtc amino acid sequence.

2.4. Determination of protein concentration

The purified protein was quantitated by its absorbance at 280 nm. The extinction coefficient ε_{280} was calculated according to [27]: $\varepsilon_{280} = (5700 \times W + 1300 \times Y)$/molecular weight with W = number of tryptophans, Y = number of tyrosines. For Stat3βtc the value of ε_{280} was 0.74 assuming a molecular weight of 136 kDa for the dimeric protein.

2.5. Immunoblot analysis

A Western blot of the supernatant of sonication and of purified protein was first developed with the monoclonal anti-phosphotyrosine antibody pY20 (Santa Cruz). Subsequently bound antibodies were removed by washing the membrane for 1 h in stripping buffer (62.5 mM Tris-HCl, pH 6.7, 2% SDS, 100 mM 2-mercaptoethanol). Next the blot was developed with the polyclonal anti-Stat3 antibody K15 (Transduction Laboratories). The detection was performed by enhanced chemoluminescence (ECL, Amersham). The secondary horseradish peroxidase-conjugated antibodies were obtained from Promega.

2.6. Electrophoretic mobility shift assay

20 μl of reaction volume contained binding buffer (10 mM Tris-HCl, pH 7.5, 1 mM DTT, 0.2 mM PMSF, 0.1 mM EDTA, 5% glycerol, 50 mM NaCl, 0.1% NP-40, 5 mM $MgCl_2$), 0.5 fmol radiolabelled DNA (see below), 2.0 fmol purified Stat3βtc protein, 20 ng bovine serum albumin (BSA) and the indicated amounts of unlabelled DNA. The reaction mix was incubated for 30 min at room temperature. Then the reaction products were loaded on a 4% polyacrylamide gel (1.5 mm thick) containing 0.25×Tris-borate/EDTA. Before loading the gel had been prerun at 15 V/cm for 30 min at 4°C. After loading the gels were run for another 3.5 h at 4°C, dried and exposed to X-ray film overnight.

2.7. Binding site oligonucleotides

Oligonucleotides with the trityl group cleaved off were obtained in quantities of 1 μmol from the DNA synthesis facility, EMBL Heidelberg. The oligonucleotides were purified by anion exchange chromatography. Oligonucleotide concentrations were determined assuming that $A_{260} = 1.0$ corresponds to 40 μg/ml of single stranded oligonucleotide. 1 mM solutions of complementary oligonucleotides were hybridized for 12 h in 5 mM Tris-HCl, pH 8.0, 50 mM KCl, 10 mM $MgCl_2$, after denaturation at 95°C. 5 pmol of duplex DNA was labelled with [α-^{32}P]dATP to a specific activity of about 2.25 μCi/fmol using polynucleotide kinase [28]. Unincorporated nucleotides were removed by gel chromatography (G25 spin columns; Pharmacia). The following fragments were used as probes (core binding sites are underlined): M67, TGCATTTCCCGTAAATCT [29]; acute phase response element (APRE), AGCTTCCTTCTGGGATTCCT [30].

3. Results

3.1. Expression and purification of tyrosine phosphorylated Stat3βtc

We expressed Stat3βtc protein in the *E. coli* strain BL21(DE3)TKB1 (Fig. 2, lanes 1, 2) at high levels. To obtain soluble protein it was essential to induce the expression at room temperature. The expressed protein had the expected molecular weight of about 70 kDa and about 30% of the protein was obtained in soluble form after sonication (Fig. 2, lanes 3, 4). After DNA removal and ammonium sulfate precipitation the protein was about 90% pure (Fig. 2, lane 5). A gel chromatography step removed smaller proteins with molecular weights between 25 and 50 kDa (Fig. 3A,B). Fractions of the major peak contained 98% pure protein. N-terminal sequencing confirmed the identity of Stat3βtc. Immunoblots with an anti-phosphotyrosine antibody (pY20) and subsequently with an anti-Stat3 antibody (K15) showed that

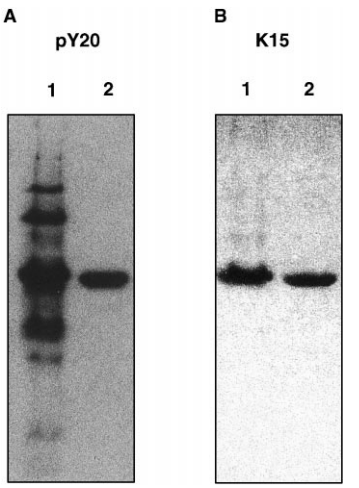


Fig. 4. Immunoblot analysis of Stat3βtc. Supernatant of the sonication and purified Stat3βtc protein were run on a 10% SDS-polyacrylamide gel and transferred to nitrocellulose membrane. A: Development with anti-phosphotyrosine monoclonal antibody pY20 (1:1000-fold dilution) using enhanced chemoluminescence. B: Redevelopment with the anti-Stat3 polyclonal antibody K15 (1:1000-fold dilution) using enhanced chemoluminescence.

a number of proteins of different molecular weights had been phosphorylated by the Elk receptor tyrosine kinase domain (Fig. 4A, lane 1). One of these phosphorylated proteins was purified and clearly identified as Stat3βtc (Fig. 4A,B, lanes 2).

3.2. Phosphorylated Stat3βtc forms functional homodimers

The retention on a gel filtration column of phosphorylated Stat3βtc was compared with that of purified Stat3βtc, which had been expressed without the tyrosine kinase domain of the Elk receptor. The gel filtration column had been calibrated with BSA (dimer 132 kDa, monomer 66 kDa) and cytochrome *c* (12.4 kDa). Phosphorylated Stat3βtc was running as a 140 kDa protein while unphosphorylated Stat3βtc had a retention profile equivalent to about 60 kDa (Fig. 5), suggesting that Stat3βtc can form a stable homodimer.

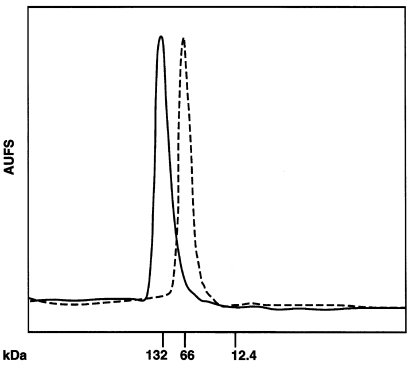
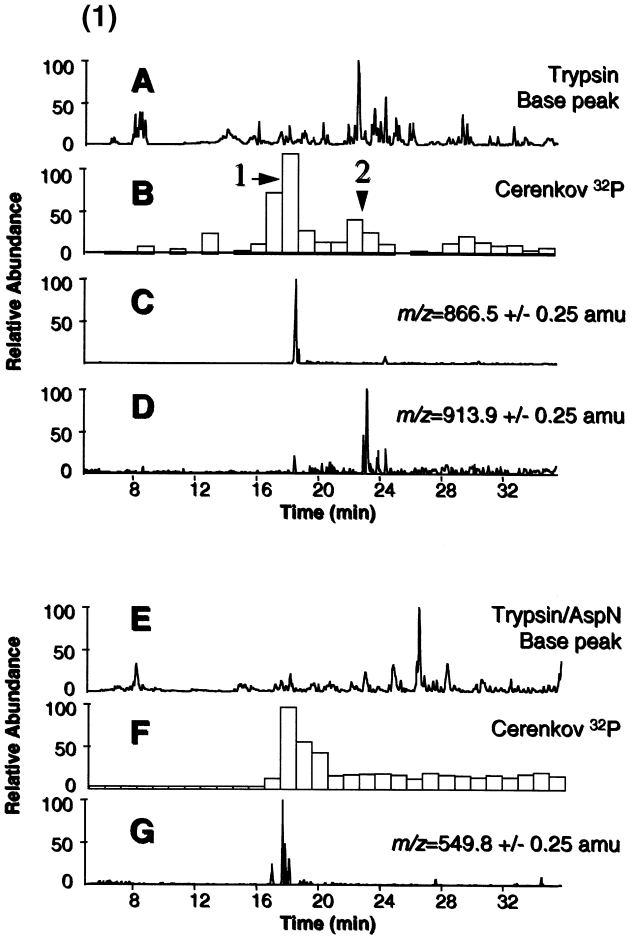


Fig. 5. Size exclusion chromatography using a Pharmacia Superose 12 HR 10/30 column. Protein standards were BSA (132 kDa, 66 kDa) and cytochrome *c* (12.4 kDa). Solid line, phosphorylated Stat3βtc; dotted line, non-phosphorylated Stat3βtc.

3.3. Determination of Stat3βtc phosphorylation sites

γ-^{32}P-Labelled Stat3βtc was used to identify the exact sites



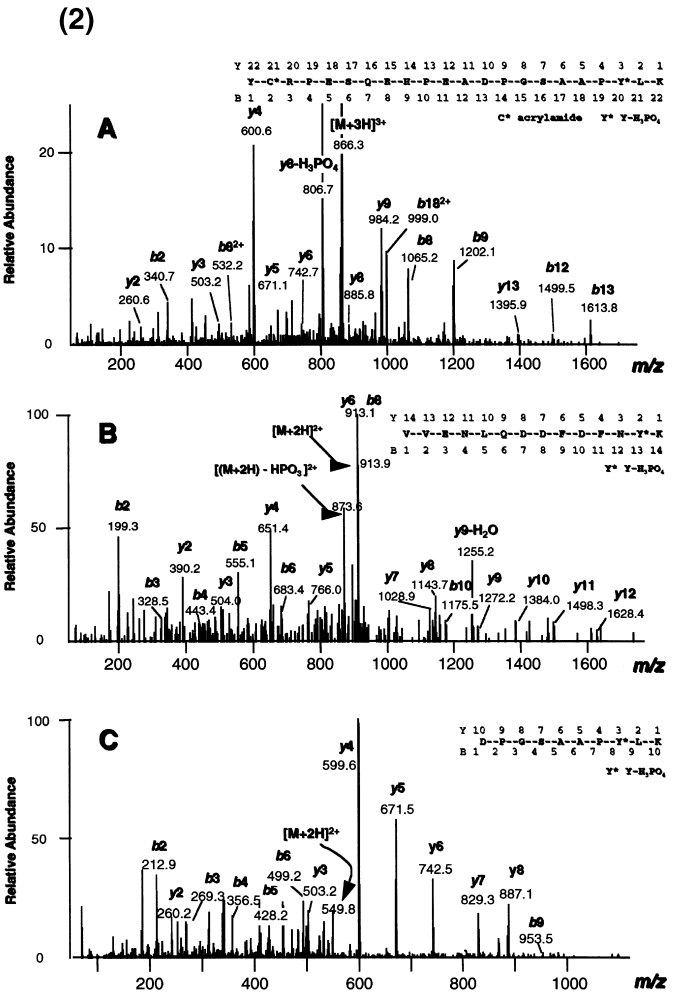


Fig. 6. 1: The mass spectrometric base peak of trypsin (A) and AspN/trypsin (E) digested peptides separated by microbore HPLC. B and F show the Cerenkov counts in the fractions that were collected post-HPLC separation by flow spliting the column eluate, indicating the precence of ^{32}P-containing peptides. Panel C shows the trace of all occurences where ions with a *m/z* ratio of 866.5 ± 0.25 amu eluted. 866.5 amu is the triply charged mass of the peptide YC^{acryl}RPESQEHPEADPGSAAPY^{phos}LK (see 6.2A). Likewise, panel D shows the trace where ions with 913.9 ± 0.25 amu eluted. 913.9 amu is the doubly charged mass of the peptide VVENLQDDFDFNY^{phos}K (see 6.2B). G shows where ions with 549.8 ± 0.25 amu eluted; 549.8 amu is the doubly charged mass of the peptide DPGSAAPY^{phos}LK (see 6.2C). C^{acryl} = cysteine-acrylamide adduct, Δmass 71; Y^{phos} = phosphotyrosine Δmass 80. 2: Tandem mass spectra of identified phosphopeptides. Fragmentation pattern of the peptides (A) YC^{acryl}RPESQEHPEADPGSAAPY^{phos}LK, *m/z* 866.5^{3+}, (B) VVENLQDDFDFNY^{phos}K, 913.9^{2+}, and (C) DPGSAAPY^{phos}LK, 549.8^{2+}. In each MS/MS spectrum the doubly or triply charged parent ion is indicated. Panel B shows the parent ion the loss of a phosphate (−40 *m/z*) at *m/z* 873.6.

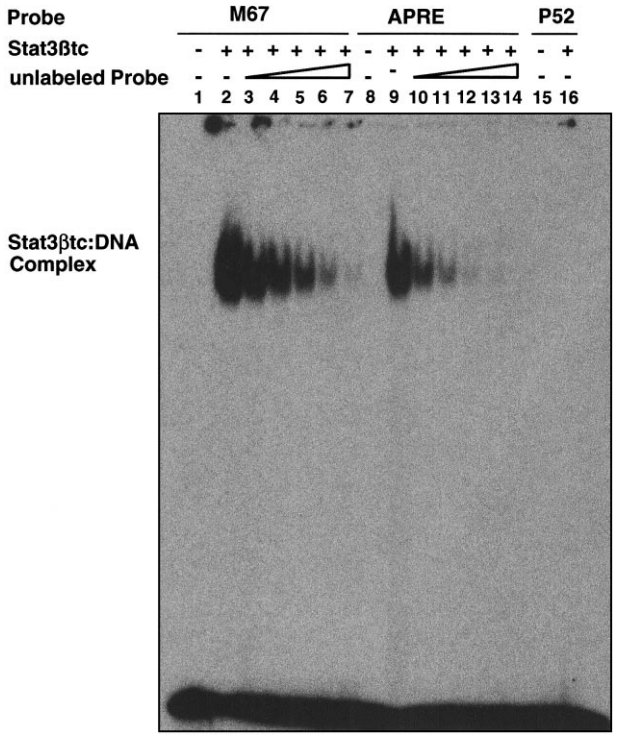


Fig. 7. Stat3βtc binds to mutant and wild type STAT DNA binding sites. The labelled probes only (lanes 1, 8, 15) or reaction mixes of Stat3βtc with the probes (lanes 2–7, 9–14, 16) were analyzed by gel retardation. Unlabelled probe was added to the reaction mixes in excess: 5-fold (lanes 3, 10), 10-fold (lanes 4, 11), 20-fold (lanes 5, 12), 50-fold (lanes 6, 13) and 100-fold (lanes 7, 14).

of phosphorylation (cf. Section 2). After the purified ^{32}P-labelled protein was first digested with trypsin two major peaks were detected by Cerenkov counting (Fig. 6.1B). Two phosphopeptides were identified with the following sequences: (K)YC^{acryl}RPESQEHPEADPGSAAPY^{phos}LK (C^{acrylamide}-aa686; Y^{phosphate}- aa705) and (K)VVENLQDDFDFNY^{phos}K (Y^{phosphate}-aa176), see Fig. 6.2A,B respectively. Both peptide masses, a triply charged 866.5 (*m*/*z*) and doubly charged 913.9 (*m*/*z*) respectively (Fig. 6.1C,D), aligned with the Cerenkov peaks (Fig. 6.1B) associating the presence of ^{32}P with these peptides. To unambiguously assign the phosphorylation site of the peptide in Cerenkov peak #1 to tyrosine residue 705, a second digestion was performed with AspN and trypsin. This digestion was expected to only reveal the peptide DPGSAAPY^{phos}LK during an LC-MS/MS run, since the small hydrophilic peptide DFNY^{phos}K (aa176) was not retained by the C18 column under the conditions used. Indeed, the subsequent digestion of a Stat3βtc sample with AspN and trypsin revealed one major peak containing ^{32}P (Fig. 6.1F, doubly charged 549.8 *m*/*z*). MS/MS sequencing of the ^{32}P-containing peptide (*m*/*z* 549.8) yielded the sequence DPGSAAPY^{phos}LK (Fig. 6.2C), thus confirming the phosphorylation site of residue Y 705, identified in the first analysis.

3.4. DNA binding

The phosphorylated Stat3βtc was able to bind to Stat3 DNA binding sites in a bandshift assay that was performed with ^{32}P-labelled, double stranded DNA fragments containing a very strong, mutant (M67) and a wild type (APRE) binding site [29,30]. As expected Stat3βtc showed the strongest interaction with M67, while APRE was bound weaker. (Fig. 7, lanes 1, 8). In both cases the binding was specific, as the unlabelled DNA fragments could competitively bind against the respective ^{32}P-labelled fragments to Stat3βtc (Fig. 7, lanes 2–7, 9–14). M67 binding could still be observed when unlabelled probe was used in a 100-fold excess while APRE binding could not be observed any more in this case (Fig. 7, lanes 7, 14). Stat3βtc could not bind to a NF-κB P52 response element (Fig. 7, lane 16).

4. Discussion

Bacterial expression of a soluble fragment of Stat3β critically depends on the fragment size. Initially we tried to express full length clones of murine Stat3, Stat4 and Stat5 in *E. coli* using a T7-RNA polymerase based expression system (pET, Novagen) [31]. Although we could express large amounts of protein, all expression conditions, including a variety of media, resulted in the presence of 100% of the protein in inclusion bodies. Vinkemeier et al. [32] achieved the bacterial expression of a soluble fragment of Stat1β lacking the N-domain. Following their approach we selected N-terminally truncated fragments of murine Stat3β, Stat4 and Stat5 and tested their expression in *E. coli*. The largest amount of soluble protein was obtained for Stat3βtc, starting at Gly127. This residue is part of a flexible linker connecting the core of the Stat molecule to the N-domain [16–18].

Compared to Stat3 the natural splice variant Stat3β has a shortened carboxyl terminal transactivation domain [14]. The transactivation domain of Stat3 is 48 residues longer, highly acidic and probably poorly structured. Furthermore it had been shown for Stat1, Stat3 and Stat5 that removal of the transactivation domain results in a decreased dephosphorylation rate resulting in increased DNA binding [10,14,33]. Stable phosphorylation was essential for our aim to produce functional STAT protein for crystallization studies and therefore the short variant of Stat3β was chosen.

Stat1 and Stat3 can be activated by ligands binding to the epidermal growth factor (EGF) receptor [34]. Correspondingly the N-terminally truncated Stat1β fragment has been activated in vitro using the tyrosine kinase activity of immunoprecipitated EGF receptor from A431 cells [32,35]. This approach resulted in yields of up to 75% of phosphorylated protein. The activation of Stat3βtc with immunoprecipitated EGF receptor was not successful in our hands. Cao et al. [36] have demonstrated that Src might be involved in the activation of Stat3 through the EGF receptor. Our attempts to activate Stat3βtc with Src were successful, but the yields were very low (not shown).

An alternative to these in vitro activations is the in vivo activation through the tyrosine kinase activity of the Elk receptor, a member of the Eph receptor family [37,38]. The strain BL21(DE3)TBK1 (Stratagene) carries an expression plasmid for this tyrosine kinase domain, which is induced following the expression of the Stat protein. Our simple purification protocol allowed us to prepare phosphorylated Stat3β that was running on a gel filtration column as a single peak corresponding to the size of a dimer, while no monomer peak could be detected (Figs. 3 and 5). This shows that all Stat3βtc protein coexpressed with the Elk receptor kinase domain dimerizes and is probably nearly quantitatively phosphorylated.

This is an important advantage of the in vivo activation compared to the less efficient in vitro activation [32], where a chromatographic step to separate phosphorylated from non-phosphorylated protein is necessary. Our purification procedure includes gel filtration chromatography, but crystals of the protein:DNA complex [17] were also obtained without the final gel filtration step. The specific binding to natural DNA target sites confirmed that the bacterially expressed protein not only dimerized but also retained its DNA binding ability.

In Stat3β Y705 is the only physiological tyrosine phosphorylation site [39]. It is responsible for the reciprocal binding of the monomers. Mass spectrometry identified this residue as the major phosphorylation site of Stat3βtc (Fig. 6). Y176 was identified as the only other phosphorylated residue of Stat3βtc (Fig. 6). This residue is located at the end of helix α1 of the four helix bundle and points into the solvent [17]. Cerenkov counting estimated its phosphorylation site has a low occupancy, which is consistent with our finding that no electron density of a phosphate group at Y176 was observed in the electron density map of the Stat3βtc:DNA cocrystals [17]. Nevertheless the specificity of the kinase activity of the Elk receptor tyrosine kinase domain seemed to be lower in the case of Stat3βtc than that of the immunoprecipitated EGF receptor from A431 cells in the case of the N-terminally truncated Stat1β. Here only the expected Y701 had been phosphorylated [32].

Stat1 and Stat3 probably bind through their SH2 domains to phosphotyrosines outside the catalytic center of the EGF receptor [40,41], similar to their specific interactions with phosphotyrosines on the cytoplasmic domains of cytokine receptors [42]. In our bacterial expression system the receptor component with its docking site as an element of substrate specificity is missing. The kinase and its substrate are therefore not held in a rigid configuration by the receptor, which might render Y176 of Stat3βtc a substrate of the kinase domain. Alternatively it is possible that the difference in the phosphorylation patterns between the EGF receptor and the Elk receptor tyrosine kinase domain reflects a real difference in specificity of the kinase activities of these two protein-tyrosine kinase receptors.

Our in vivo activation procedure allows the easy preparation of large amounts of active Stat3 protein. Using the same procedure we also observed the phosphorylation of a N-terminally truncated fragment of Stat5 (not shown) and indeed, all activated STATs might be obtained by this approach. The easy accessibility to large quantities of active STAT proteins should prove beneficial for further investigations such as structural studies and large-scale drug screens. Furthermore protein-tyrosine kinase receptors activate many other effector proteins [43,44]. It is therefore intriguing to investigate if any of these proteins can also be activated in vivo by the Elk receptor tyrosine kinase domain.

Acknowledgements: R.A. was supported by National Science Foundation (USA) Center for Molecular Biotechnology (NIHNCRR 11823) and by the National Institutes of Health (USA) Center for Comprehensive Biology (BIR 9214821 AM). S.B. was supported by an institutional fellowship of the European Union and a short-term EMBO fellowship.

References

[1] Darnell, J.E. (1997) Science 277, 1630–1635.
[2] Darnell, J.E., Kerr, I.M. and Stark, G.R. (1994) Science 264, 1415–1421.
[3] Schindler, C. and Darnell, J.E. (1995) Annu. Rev. Biochem. 64, 621–651.
[4] Leaman, D.W., Leung, S., Li, X. and Stark, G.R. (1996) FASEB J. 10, 1578–1588.
[5] Taniguchi, T. (1996) Science 268, 251–255.
[6] Shuai, K., Horvath, C.M., Tsai-Huang, L.H., Qureshi, S., Cowburn, D. and Darnell, J.E. (1994) Cell 76, 821–828.
[7] Leonard, W.J. and O'Shea, J. (1998) Annu. Rev. Immunol. 16, 293–332.
[8] Vaisse, C., Halaas, J.L., Horvath, C.M., Darnell, J.E., Stoffel, M. and Friedman, J.M. (1996) Nature Genet. 14, 95–97.
[9] Qureshi, S.A., Leung, S., Kerr, I.M., Stark, G.R. and Darnell, J.E. (1996) Mol. Cell. Biol. 16, 288–293.
[10] Moriggl, R.M., Gouilleux-Gruart, V., Jähne, R., Berchtold, S., Gartmann, C., Liu, X., Hennighausen, L., Sotiropoulos, A., Groner, B. and Gouilleux, F. (1996) Mol. Cell. Biol. 16, 5611–5700.
[11] Horvath, C.M., Wen, Z. and Darnell, J.E. (1995) Genes Dev. 9, 984–994.
[12] Fu, X.Y. (1992) Cell 70, 323–335.
[13] Schaefer, T.S., Sanders, L.K. and Nathans, D. (1995) Proc. Natl. Acad. Sci. USA 92, 9097–9101.
[14] Caldenhoven, E., van Dijk, T.B., Solari, R., Armstrong, J., Raaijmakers, J.A.M., Lammers, J.W.J., Kowenderman, L. and de Groot, R.P. (1996) J. Biol. Chem. 271, 13221–13227.
[15] Shuai, K., Ziemecki, A., Wilks, A.F., Harpur, A.G., Sadowski, H.B., Gilman, M.Z. and Darnell, J.E. (1993) Nature 366, 580–583.
[16] Vinkemeier, U., Moarefi, I., Darnell, J.E. and Kuriyan, J. (1998) Science 279, 1048–1052.
[17] Becker, S., Groner, B. and Müller, C.W. (1998) Nature 394, 145–151.
[18] Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J.E. and Kuriyan, J. (1998) Cell 93, 827–839.
[19] Letwin, K., Yee, S.-P. and Pawson, T. (1988) Oncogene 3, 621–627.
[20] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[21] Wilson, M.L. and MacNab, R.M. (1988) J. Bacteriol. 170, 588–597.
[22] Oostveen, V.-I., Ducret, A. and Aebersold, R. (1997) Anal. Biochem. 247, 310–318.
[23] Hess, D., Covey, T.C., Winz, R., Brownsey, R.W. and Aebersold, R. (1993) Protein Sci. 2, 1342–1351.
[24] Ducret, A., Bruun, C.F., Bures, E.J., Marhaug, G., Husby, G. and Aebersold, R. (1996) Electrophoresis 17, 866–876.
[25] Ducret, A., VanOostveen, I., Eng, J.K., Yates, J.R. and Aebersold, R. (1998) Protein Sci. 7, 706–719.
[26] Eng, J.K., McCormack, A.L. and Yates III, J.R. (1994) J. Am. Soc. Mass Spectrom. 5, 976–989.
[27] Cantor, C.R. and Schimmel, P.R. (1980) in: Biophysical Chemistry, Part II, Techniques for the Study of Biological Structure and Function, pp. 380–381, W.H. Freeman and Company, San Francisco, CA.
[28] Berchtold, S., Moriggl, R.M., Gouilleux, F., Silvennoinen, O., Beisenherz, C., Pfitzner, E., Wissler, M., Stöcklin, E. and Groner, B. (1997) J. Biol. Chem. 272, 30237–30243.
[29] Wagner, B.J., Hayes, T.E., Hoban, C.J. and Cochran, B.H. (1990) EMBO J. 9, 4477–4484.
[30] Ito, T., Tanahashi, H., Misumi, Y. and Sakaki, Y. (1989) Nucleic Acids Res. 17, 9425–9435.
[31] Studier, F.W. and Moffat, B.A. (1986) J. Mol. Biol. 189, 113–130.
[32] Shuai, K., Liao, J. and Song, M.M. (1996) Mol. Cell. Biol. 16, 4932–4941.
[33] Fu, X.Y. and Zhang, J.J. (1993) Cell 74, 1135–1145.
[34] Vinkemeier, U., Cohen, S.L., Moarefi, I., Chait, B.T., Kuriyan, J. and Darnell, J.E. (1996) EMBO J. 15, 5616–5626.

[35] Yarden, Y., Harari, I. and Schlessinger, J. (1985) J. Biol. Chem. 260, 315–319.
[36] Cao, X., Tay, A., Guy, G.R. and Tan, Y.H. (1996) Mol. Cell. Biol. 16, 1595–1603.
[37] Lemmon, M.A. and Schlessinger, J. (1993) Trends Biochem. Sci. 19, 459–463.
[38] Heldin, C.-H. (1995) Cell 80, 212–223.
[39] Wen, Z. and Darnell, J.E. (1996) Nucleic Acids Res. 25, 2062–2067.
[40] Leaman, D.W., Pisharody, S., Flickinger, T.W., Commane, M.A., Schlessinger, J., Kerr, I.M., Levy, D.E. and Stark, G.R. (1996) Mol. Cell. Biol. 16, 369–375.
[41] Coffer, P.J. and Kruijer, W. (1995) Biochem. Biophys. Res. Commun. 210, 74–81.
[42] Stahl, N., Farrugella, T.J., Boulton, T.G., Zhong, Z., Darnell, J.E. and Yancopoulos, G.D. (1995) Science 267, 1349–1353.
[43] Goldschmidt-Clermont, P.J., Kim, J.W., Machesky, L.M., Rhee, S.G. and Pollard, T.D. (1991) Science 251, 1231–1233.
[44] Feng, G.S., Hui, C.C. and Pawson, T. (1993) Science 259, 1607–1611.