Supporting Information

Unexpected implications of STAT3 acetylation revealed by genetic encoding of acetyl-lysine

Yael Belo[a], Zack Mielko[b], Hila Nudelman[a], Ariel Afek[b], Oshrit Ben-David[a,c], Anat Shahar[c], Raz Zarivach[d,c], Raluca Gordan[b,e], and Eyal Arbely *[a,d]

[a]Department of Chemistry, Ben Gurion University of the Negev, P.O. Box 653, Beer-Sheva, 8410501, Israel

[b]Center for Genomic and Computational Biology, Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27708, USA

[c]The National Institute for Biotechnology in the Negev, Ben Gurion University of the Negev, P.O. Box 653, Beer-Sheva, 8410501, Israel

[d]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, 8410501, Israel

[e]Department of Computer Science, Department of Molecular Genetics and Microbiology, Duke University, Durham, NC 27708, USA

*Corresponding Author: arbely@bgu.ac.il

Supplementary Tables

Table S1. Data collection and refinement statistics.

PDB code	6QHD
Protein	AcK685+pY705 STAT3
Beamline	ID29–ESRF
Wavelength (Å)	0.97625
Space group	$P4_1$
Cell Dimensions:	
a, b, c (Å)	175.49, 175.49, 79.07
α, β, γ (°)	90, 90, 90
Resolution (Å)	50–2.85
Unique reflections	47795
Completeness (%)	99.7 (98.4)
Mean $I/\sigma(I)$	10.22 (0.97)
Redundancy	7.62 (7.81)
Refinement:	
R_{work}	0.294
R_{free}	0.343
$CC_{1/2}$	0.999 (0.653)
Number of atoms (non hydrogens):	
Protein	9365
Water	89
Average B-factor (Å^2)	97.54
Protein	101.6
Solvent	81.59
R.M.S deviations:	
Bond lengths (Å)	0.019
Bond angles (°)	2.232
Ramachandran statistics:	
Allowed	349 (94.29%)
Partially allowed	14 (4.07%)
Disallowed	15 (1.65%)

The values in parentheses refer to the data of the corresponding upper resolution shell. One crystal was used per data set. Data was collected at 100°K.
R_{free} calculated using 5% of the reflection data chosen randomly and omitted from refinement.
R.M.S deviations for bonds and angles are the respective root-mean-square deviations from ideal values.

Supplementary Figures

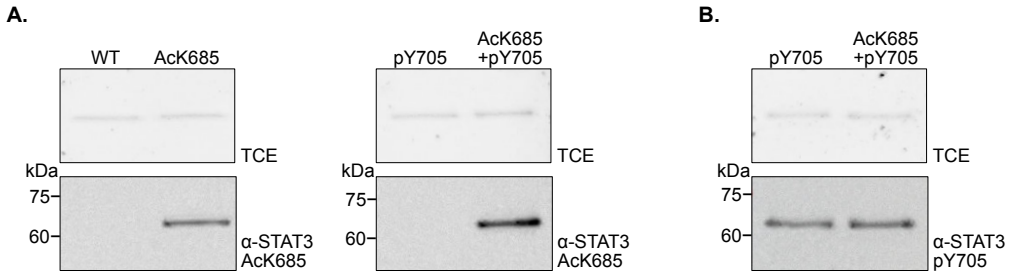

Figure S1. Expression of STAT3 in bacteria. **A.** Indicated purified STAT3 proteins were analyzed by SDS-PAGE and visualized by Western blotting using specific antibodies against Lys685-acetylated STAT3. As loading control, proteins were also visualized by in-gel fluorescence following UV-induced reaction between tryptophan residues and 2,2,2-trichloroethanol (TCE). **B.** Indicated purified STAT3 proteins were analyzed as described in panel A, using specific antibodies against Tyr705-phosphorylated STAT3.

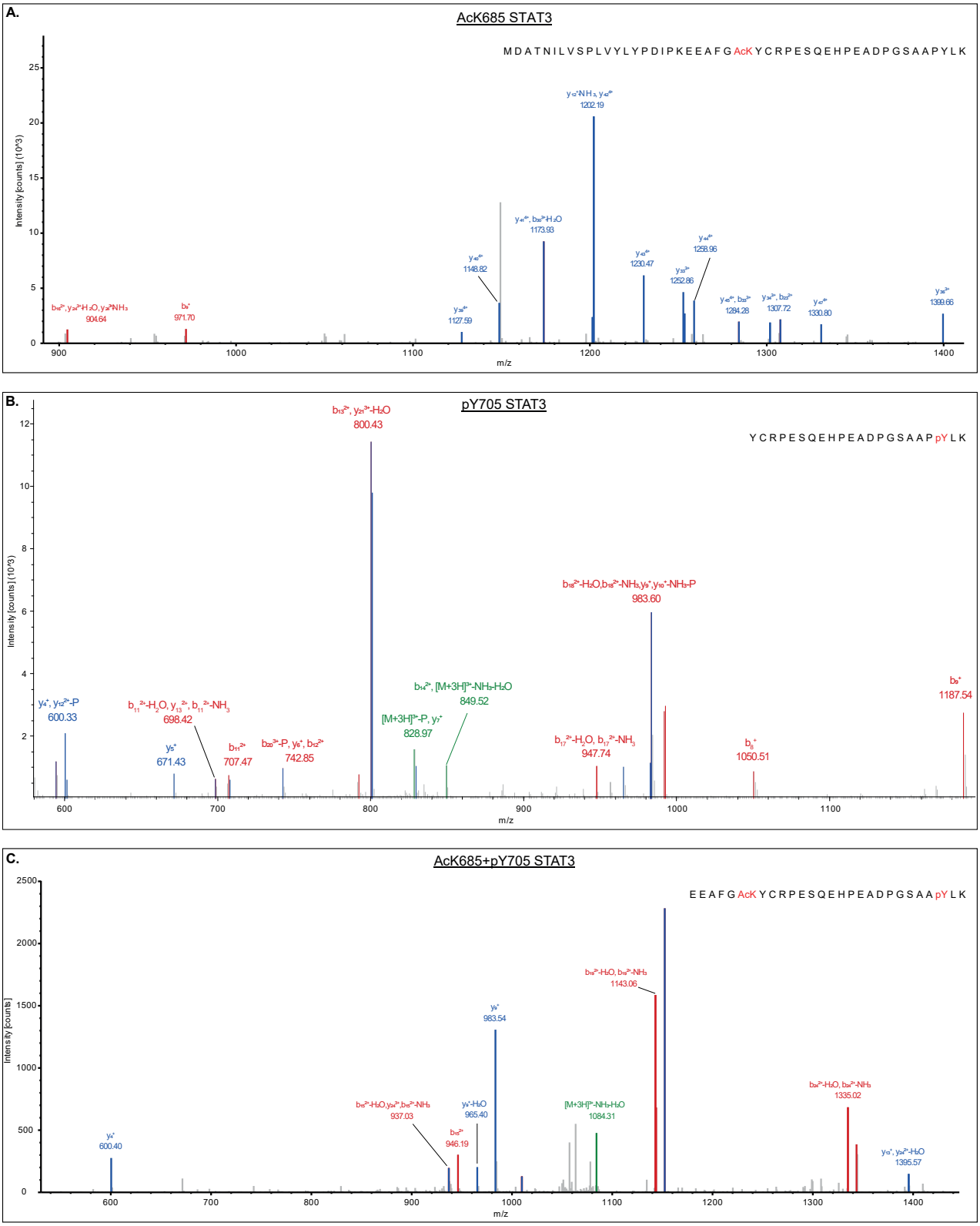

Figure S2. LC-MS/MS identification of acetyl lysine and phosphotyrosine. MS/MS data confirm the incorporation of acetyl lysine at position 685 of AcK685 STAT3 (**A**), phosphotyrosine at position 705 of pY705 STAT3 (**B**), and both acetyl-lysine and phosphotyrosine at positions 685 and 705, respectively, of AcK685+pY705 STAT3 (**C**).

Supplementary Methods

General

General chemicals and DNA oligomers for molecular cloning were ordered from Sigma Aldrich (Darmstadt, Germany). DNA sequencing was performed by the sequencing facility at Ben-Gurion University. The STAT3 gene was amplified from pDONR221 plasmid carrying STAT3 cDNA (DNASU plasmid ID HsCD00295594). pBK vector for expression of evolved acetyl-lysine synthetase was kindly provided by Dr. Jason W. Chin (MRC-LMB, Cambridge, UK).[1] Enzymes for molecular cloning were purchased from NEB (Ipswich, MA) and used according to the manufacturer's instructions. DNA was purified using spin columns from Macherey Nagel (Düren, Germany). Acetyl lysine was purchased from Chem-Impex International Inc. (Wood Dale, IL) and used without further purification. DH10B *E. coli* strain (Life technologies, Carlsbad, CA) was used for molecular cloning and plasmid propagation. BL21(DE3) *E. coli* strain (NEB, Ipswich, MA) was used for protein expression. Bacteria were incubated in liquid LB media or on LB/agar plates supplemented with antibiotics (50 μg/mL kanamycin, spectinomycin, or chloramphenicol). Primary antibodies: anti-6×His (#G020) was purchased from abm (Richmond, ON); anti Y705-phosphorylated STAT3 (#ab76315) was purchased from Abcam (Cambridge, UK); anti K685-acetylated STAT3 (#PA5-17429) was purchased from Thermo Fisher Scientific (Waltham, MA). Secondary antibodies: anti-mouse IgG (#ab7068) and anti-rabbit IgG (#ab92080) were purchased from Abcam.

Molecular cloning

Non-phosphorylated STAT3 (residues 128–715, accession number NP_644805.1) was expressed as a fusion protein with C-terminal tobacco etch virus (TEV) cleavage site, followed by the lipoyl domain and 6×His-tag. The gene was cloned into the first open reading frame of a pCDF-Duet vector using Gibson Assembly Kit (NEB, Ipswich, MA).[2,3] This vector also contained a U25C mutant of PylT under constitutive expression.[4] To enable the co-translational incorporation of acetylated lysine, an in-frame TAG mutation was introduced at position Lys685 by site-directed mutagenesis. To express phosphorylated STAT3 proteins, STAT3 and K685-TAG STAT3 (residues 128–715) with C-terminal 6×His-tag were cloned into the first open reading frame of the above mentioned pCDF-Duet vector. Phosphorylated STAT3 variants were expressed without C-terminal TEV cleavage site and lipoyl domain, since TEV protease was incapable of digesting the Tyr705-phosphorylated variants of STAT3. Next, the gene coding for the

kinse domain of the Elk receptor was amplified from TKB1 cells (Agilent Technologies, Santa Clara, CA) and cloned with a C-terminal HA-tag between NdeI and EcoRV restriction sites within the second open reading frame of the pCDF-Duet vector. Thus, this plasmid enabled the IPTG-controlled co-expression of STAT3 (either acetylated or non-acetylated) and Elk.

Protein Purification

Frozen bacterial pellet (∼20 gr) was resuspended in 100 mL of buffer A (50 mM Tris pH 8.0, 100 mM NaCl, 15 mM β-mercaptoethanol, 20 mM imidazole pH 8.0) supplemented with 1.2 μg/mL leupeptin, 1 μM pepstatin A, 100 μM PMSF, 1 μg/mL aprotinin, 0.4 mg/mL lysozyme, 20 μg/mL DNAse, 10 mM $MgCl_2$, and 10 mM nicotinamide. For the purification of Tyr705-phosphorylated STAT3, 100 μM of sodium orthovanadate were add to the buffer. Cells were incubated on ice with stirring for 30 min, lysed by sonication and the lysate was centrifuged at 20,000 g for 30 min at 4°C. Clear lysate was loaded on a 5 mL HisTrap HP column (GE Healthcare, Chicago, IL) pre-equilibrated with buffer A. Column was washed with at least 10 column volumes of buffer A and protein was eluted with a linear gradient (0–100% over 20 column volumes) of buffer B (50 mM Tris pH 8.0, 100 mM NaCl, 15 mM β-mercaptoethanol, 500 mM imidazole pH 8.0). Fractions were analyzed by SDS-polyacrylamide gel electrophoresis, and fractions of highest purity were collected. The combined fractions of non-phosphorylated STAT3 variants were diluted 1:3 with dialysis buffer (25 mM Tris pH 7.6, 300 mM NaCl, 10% Glycerol, 15 mM β-mercaptoethanol) and filtered. TEV protease was added to the sample and dialysis was performed against 2×4 L of dialysis buffer at 4°C. Dialyzed protein solution was supplemented with imidazole (25 mM) and the protein was further purified by a second Ni^{2+} affinity chromatography. Flow-through containing the cleaved protein was collected and diluted 1:10 in ice-cold buffer C (25 mM Tris pH 7.6, 10% Glycerol, 15 mM β-mercaptoethanol). When phosphorylated STAT3 variants were purified, dialysis was performed without TEV protease, and dialyzed sample was diluted 1:10 into ice-cold buffer C, without a second Ni^{2+} affinity chromatography. Diluted protein samples in buffer C were loaded on a heparin HP 5 mL column (GE Healthcare, Chicago, IL) and protein was eluted with buffer D (25mM Tris pH 7.6, 10% Glycerol, 15 mM β-mercaptoethanol, 1 M NaCl) using reverse flow. Eluted protein was concentrated using Amicon Ultra-15 centrifugal filter unit with nominal molecular weight limit of 10 kDa (Merck Millipore, Burlington, MA), loaded on a HiLoad 26/600 Superdex 200 column (GE Healthcare, Chicago, IL) pre-equilibrated with buffer E (HEPES pH 7.0, 200 mM NaCl, 5 mM DTT, 10 mM $MgCl_2$ and 0.5 mM PMSF) and protein was eluted at a flow rate

of 2.6 ml/min with fractionation. Fractions were analyzed by 12% SDS-polyacrylamide gel electrophoresis and fractions of highest purity were combined, concentrated to ∼0.5 mg/mL (based on UV absorption, ε=89840) using Amicon Ultra-15 centrifugal filter unit and stored at -80°C.

Crystallization

Single-stranded DNA sequences (5'-TGCATTTCCCGTAAATCT-3' and 5'-AAGATTTAC GGGAAATGC-3', IDT, Coralville, IA) were dissolve in annealing buffer (1 M Tris-HCl pH 8, 1 M NaCl and 0.5 M EDTA pH 8), heated to 90°C, and annealed by a slow temperature gradient (-1°C/min). Double-stranded molecules were dialyzed extensively against DDW (18.2 MΩ·cm) at 4°C, filtered through a 0.2 μM filter, lyophilized and redissolved in ultra pure water at concentration of 0.175 mM. Purified AcK685+pY705 STAT3 and double-stranded DNA were mixed at 1:2 molar ratio (STAT3 dimer:double stranded DNA), and the sample was incubated on ice for 1 h. Crystals were grown at 20°C using the sitting drop vapor diffusion technique. Drops composed of 2 μl protein/DNA complex and 0.5 μl of crystallization solution were equilibrated above a reservoir of 80 μl. Initial crystallization experiments were performed using the Hampton Research index screen (Aliso Viejo, CA). Final crystallization conditions were 0.2 M lithium sulfate, 0.1 M Bis-Tris pH 7.0 and 20% w/v polyethylene glycol 3350. Before data collection, crystals were transferred into a cryo-protectant solution consisting of 60% mother liquor, 25% ethylene glycol, and 15% DDW. The protected crystals were mounted on Hampton Research CryoCapHT nylon loops and flash-frozen in liquid nitrogen.

Data collection, structure determination, and data analysis

Diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France), beamline ID-29. Data were indexed and integrated with XDS. Initial phase determination was performed by molecular replacement with Phaser from the CCP4 package, using a previously solved STAT3 structure (PDB: 1BG1)[5] as the search model. The structure was further refined using CCP4 Phenix.[6] Successive rounds of model building and manual corrections were performed with COOT.[7] Figures were prepared using PyMol.

Western blot

Protein-expressing bacteria from 1 mL of overnight culture normalized to O.D600=1 were precipitated (15,000 rpm, 10 min, 4°C) and cell pellet was resuspended in 500 μL of 1×Laemmli

sample buffer. Cells were lysed by heating to 95°C with agitation (400 rpm) for 7 min and cell debris were precipitated by centrifugation at 4°C (15,000 rpm). Proteins in equal volumes of cleared lysates were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a 0.2 μm nitrocellulose membrane using a semi-dry transfer apparatus (Trans-Blot Turbo, BioRad, Hercules, CA). Membranes were blocked with Tris-buffered saline containing 0.05% (v/v) Tween-20 (TBST) and 5% (w/v) non-fat dry milk and incubated over night with primary antibody diluted in 5% (w/v) bovine serum albumin in TBST, at 4°C. On the following day, membranes were washed with 1×TBST, incubated with secondary antibody for 1 h at room temperature, and washed again. Finally, proteins were visualized using ECL reagent (GE Healthcare, Chicago, IL) and immunoblot intensities were quantified with ImageJ.[8].

MS/MS

Purified protein samples were separated by SDS polyacrylamide gel electrophoresis, and the band corresponding to ~64 kDa was incised and in-gel digested by trypsin according to the manufacturer's protocol (Promega). Peptides were then extracted from the gel and analyzed by LC-MS using an Eksigent nano-HPLC (model nanoLC-2D, Netherlands) connected to an LTQ Orbitrap XL ETD mass spectrometer (Thermo Fisher Scientific, Germany & USA). Reverse-phase chromatography of peptides was performed using a C-18 column (IntegraFrit, 360 µm OD × 75 µm ID; New Objective USA). Peptides were separated by a 70 min linear gradient, starting with 100% buffer A (5% acetonitrile, 0.1% formic acid) and ending with 80% buffer B (80% acetonitrile, 0.1% formic acid), at a flow rate of 300 nl/min. A full scan, acquired at 60,000 resolution, was followed by CID MS/MS analysis performed for the five most abundant peaks, in the data-dependent mode. Fragmentation (with minimum signal trigger threshold set at 500) and detection of fragments were carried out in the linear ion trap. Maximum ion fill time settings were 500 ms for the high-resolution full scan in the Orbitrap analyzer and 200 ms for MS/MS analysis in the ion trap. The AGC settings were 5×10^5 and 1×10^4 (MS/MS) for Orbitrap and linear ion trap analyzers, respectively. Proteins were identified and validated using the SEQUEST and Mascot search engines operated under the Proteome Discoverer 1.4 software (Thermo Fisher Scientific). Mass tolerance for precursors and fragmentations was set to 10 ppm and 0.8 'Da, respectively. Only proteins containing at least two peptides of high confidence (Xcore 2 or 2.5 or more for doubly or triply charged species, respectively) were chosen.

Supplementary References

[1] Neumann, H. *et al.* A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Molecular Cell* **36**, 153–163 (2009).

[2] Vainer, R., Cohen, S., Shahar, A., Zarivach, R. & Arbely, E. Structural Basis for p53 Lys120-Acetylation-Dependent DNA-Binding Mode. *Journal of Molecular Biology* **428**, 3013–3025 (2016).

[3] Arbely, E. *et al.* Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. *Proceedings of the National Academy of Sciences* **108**, 8251–8256 (2011).

[4] Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52**, 1828–1837 (2013).

[5] Becker, S., Corthals, G. L., Aebersold, R., Groner, B. & Müller, C. W. Expression of a tyrosine phosphorylated, DNA binding Stat3β dimer in bacteria. *FEBS Letters* **441**, 141–147 (1998).

[6] Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213–221 (2010).

[7] Emsley, P. & Cowtan, K. Coot : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **60**, 2126–2132 (2004).

[8] Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).