

SCIENTIFIC REPORTS



OPEN

Insight into the molecular recognition mechanism of the coactivator NCoA1 by STAT6

Luigi Russo^{1,3}, Karin Giller¹, Edith Pfitzner^{2,4}, Christian Griesinger¹ & Stefan Becker¹

Crucial for immune and anti-inflammatory cellular responses, signal transducer and activator of transcription 6 (STAT6) regulates transcriptional activation in response to interleukin-4 and -13-induced tyrosine phosphorylation by direct interaction with coactivators. The interaction of STAT6 with nuclear coactivator 1 (NCoA1) is mediated by a short region of the STAT6 transactivation domain that includes the motif LXXLL and interacts with the PAS-B domain of NCoA1. Despite the availability of an X-ray structure of the PAS-B domain/Leu⁷⁹⁴-Gly⁸¹⁴-STAT6 complex, the mechanistic details of this interaction are still poorly understood. Here, we determine the structure of the NCoA1^{257–385}/STAT6^{783–814} complex using Nuclear Magnetic Resonance (NMR) and X-ray crystallography. The STAT6^{783–814} peptide binds with additional N-terminal amino acids to NCoA1^{257–385}, compared to the STAT6^{794–814} peptide, explaining its higher affinity. Secondary and tertiary structures existing in the free peptide are more highly populated in the complex, suggesting binding by conformational selection.

STAT6 belongs to a family of transcription factors known as the signal transducers and activators of transcription (STAT). STAT family members share a similar protein structure, which is essential for their activation and function. They are composed of an N-terminal coiled-coil domain¹, a centrally located DNA-binding domain², a linker region, an SH2 domain for dimerization³ and a transactivation domain at the C-terminus⁴. STAT proteins mediate signaling from activated cytokine receptors to the nucleus⁵. After phosphorylation at a specific tyrosine by a receptor associated Janus kinase, STATs form homo- or heterodimers and translocate into the nucleus where they modulate transcription after binding to specific DNA sequence elements^{4,5}. STAT6 becomes activated in response to IL-4 and IL-13 and mediates most of the gene expression regulated by these cytokines⁶.

By direct interaction with specific parts of its transactivation domain, STAT6 recruits the co-activators p300/CBP and NCoA1 (also called steroid receptor coactivator-1, SRC-1), which are essential for transcriptional activation⁷. In particular, the interaction between STAT6 and NCoA1 is modulated by a short region of the transactivation domain that includes the motif LXXLL (where L is leucine and X is any amino acid)⁷. The crystal structure of a STAT6-derived peptide (Leu⁷⁹⁴-Gly⁸¹⁴) in complex with the NCoA1 PAS-B domain^{257–385} (PDB ID: 1OJ5) (Fig. 1) revealed that the leucine side-chains of the motif (Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶), are deeply embedded into a hydrophobic groove on the surface of NCoA1⁸. More recently, Robinson and coworkers⁹ demonstrated by fluorescence polarization (FP) binding assays that a peptide comprising STAT6 residues 783–814 (for sequence see Fig. 2) binds about 6.5 to 8 times stronger than Leu⁷⁹⁴-Gly⁸¹⁴ ($K_d = 0.04 \mu\text{M}$ vs. $0.32 \mu\text{M}$ from direct FP, $K_i = 0.04 \mu\text{M}$ vs. $0.26 \mu\text{M}$ from competitive FP). Obviously, more residues located N-terminally of the LXXLL motif in STAT6, play an important role in stabilizing the protein binding to NCoA1⁹. Yet, the molecular recognition mechanism of NCoA1 by the STAT6 transactivation domain is still poorly understood. Here, we report the structural characterization of the complex between a STAT6-derived peptide encompassing the region from Gly⁷⁸³ to Gly⁸¹⁴ and the NCoA1 PAS-B domain^{257–385} using NMR and X-ray crystallography. The structural characterization of the NCoA1^{257–385}/STAT6^{783–814} complex demonstrates that the STAT6^{783–814} peptide binds to the NCoA1 PAS-B domain^{257–385} by additional amino acids from its N-terminal region resulting in a more extended binding interface with NCoA1 compared to that identified before in the crystal structure with the STAT6^{794–814}

¹Department for NMR based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077, Göttingen, Germany. ²Friedrich-Schiller-University Jena, Institute of Biochemistry and Biophysics, Philosophenweg 12, 07743, Jena, Germany. ³Present address: Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania "Luigi Vanvitelli", 81100, Caserta, Italy. ⁴Present address: University of Kassel, Mönchebergstr. 19, 34109, Kassel, Germany. Correspondence and requests for materials should be addressed to S.B. (email: sabe@nmr.mpibpc.mpg.de)

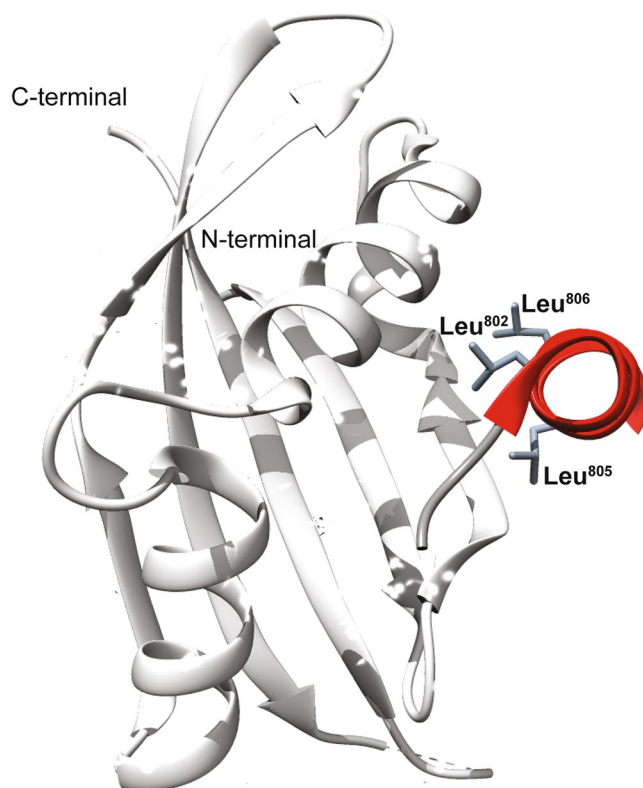


Figure 1. Overview of the NCoA1/STAT6^{794–814} X-ray structure (PDB ID: 1OJ5). Ribbon drawing representation of the NCoA1/STAT6^{794–814} complex. The helical region containing the LXXLL motif of the STAT6 derived peptide is depicted in red. The side chains of the STAT6 residues Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶ are shown as sticks.

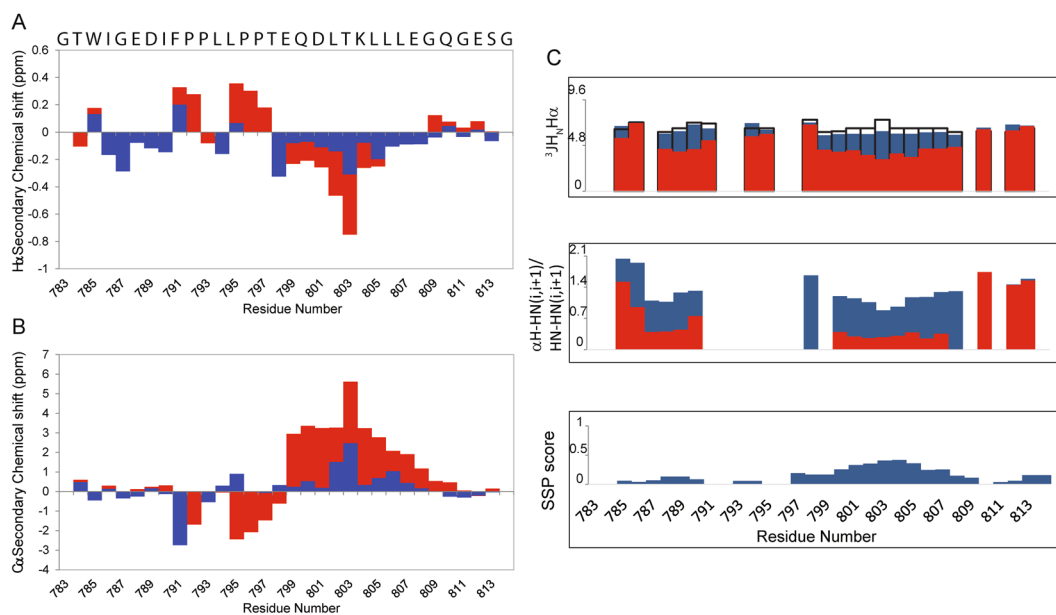


Figure 2. Secondary structure propensity analysis of STAT6^{783–814}. (**A,B**) Secondary chemical shifts of H α (**A**) and C α (**B**) for the STAT6^{783–814} peptide in the free (blue) and bound (red) state. (**C**) $^3J_{H_N H_\alpha}$ coupling constants (*upper*) and the intensity ratio of $\alpha H-HN(i,i+1)/HN-HN(i,i+1)$ NOEs (*middle*) for the STAT6^{783–814} peptide in the free (blue) and bound (red) state. The open bars indicate the reference $^3J_{H_N H_\alpha}$ values in the random coil conformation¹⁷ and the reference values of the intensity ratio of $\alpha H-HN(i,i+1)/HN-HN(i,i+1)$ NOEs are: 1.4 for random coil, 0.25 for α -helix and 55 for β -strand¹⁷. The SSP score (Secondary structural propensity) (*lower*) of the free STAT6^{783–814} peptide is also reported.

peptide⁸. Overall, the data indicate that the conformational propensity of free STAT6^{783–814} peptide in solution on the level of secondary and tertiary structure supports conformational selection as the key mechanism driving the molecular recognition of the coactivator by STAT6.

Results

Structural characterization of the STAT6^{783–814} peptide. To gain insight into the structural features of the protein-protein recognition mechanism between STAT6^{783–814} and the NCoA1 PAS-B domain we first investigated by NMR the STAT6^{783–814} peptide in the free form. The ¹H, ¹⁵N-HSQC spectrum of ¹⁵N-¹³C-labeled STAT6^{783–814} (SI Fig. 1) shows narrow dispersion of signals in both proton and nitrogen dimensions indicating that the peptide adopts an unstructured conformation in absence of the binding partner. This finding was further supported by the analysis of the 3D ¹H, ¹⁵N-NOESY-HSQC spectrum that does not show any medium and long range NOEs (SI Fig. 2). To better understand the structural details of the conformational ensemble sampled in solution by the STAT6^{783–814} peptide we also analyzed the backbone chemical shifts that are sensitive reporters of the secondary structure content¹⁰. Hence, we obtained the complete assignment of proton, C α and C β chemical shifts of STAT6^{783–814} in the free state (see Materials and Methods section) (Table SI 1a).

To obtain secondary structure propensities through the chemical shifts, we used the random coil values suitable for IDPs^{11,12}. Overall, the secondary C α and H α chemical shifts are relatively small and no patterns could be discerned, consistent with a lack of ordered secondary structure (Fig. 2A,B).

Nevertheless, the positive C α and negative H α secondary chemical shifts suggest the presence of a significant amount of transiently formed helix for the residues located in the region containing the LXXLL motif (Glu⁷⁹⁹-Glu⁸⁰⁸) (Fig. 2A,B). Additionally, ³JH_NH α coupling constants and the intensity ratio of α H-HN(i,i + 1)/HN-HN(i,i + 1) NOEs (Fig. 2C) were analyzed. Altogether the data for the residues located at the N-terminal part of the peptide in the region Ile⁷⁸⁶-Ile⁷⁹⁰ deviate from random coil values indicating helical propensity that is lower than the helical propensity for the helix between residues Glu⁷⁹⁹-Glu⁸⁰⁸.

To further investigate the conformational properties for STAT6^{783–814} in terms of helical populations, we utilized the secondary structure propensity (SSP) approach¹³. Our quantitative analysis (Fig. 2C) reveals more than 30% α -helical propensity in the region from Glu⁷⁹⁹ to Glu⁸⁰⁸ containing the LXXLL motif which is in quantitative agreement with the analysis of the coupling constants (Table SI 1b). In the region from Ile⁷⁸⁶ to Ile⁷⁹⁰ of the N-terminal part of the peptide 10% α -helical propensity is predicted from the SSP analysis as well as the ³JH_NH α coupling constants analysis (Table SI 1b). Taken together the data clearly indicate that the free STAT6^{783–814} peptide does not adopt a completely random coil conformation but contains two regions (Ile⁷⁸⁶-Ile⁷⁹⁰, Glu⁷⁹⁹-Glu⁸⁰⁸) with a significant α -helical propensity of around 10% for the first and 30% for the second sequence.

Structure of the NCoA1^{257–385}/STAT6^{783–814} complex by X-ray crystallography. In order to understand the increase of affinity of the STAT6^{783–814} peptide compared to the STAT6^{794–814} peptide in complex with NCoA1 by a factor of 10 we first solved the crystal structure of the NCoA-1 PAS-B domain in complex with the STAT6^{783–814} peptide (PDB ID: 5NWX) (SI Tables 2 and 3). In this crystal structure the residues located in the N-terminal part (Gly⁷⁸³ - Pro⁷⁹³) are not ordered indicating that this region is too dynamic to be resolved in the crystal (SI Fig. 3). Nevertheless, the new X-ray structure of the complex clarifies the structural role of Leu⁷⁹⁴ whose side chain fits into a deep pocket on the cofactor surface formed by Phe³¹⁴, Phe³⁰⁰ and Ala³¹⁰. Despite the fact that this amino acid was present in the construct crystallized previously⁸ it was not visible in the X-ray structure. Yet, there is no interaction visible between the residues N-terminal of Leu⁷⁹⁴ in the crystal structure. Therefore, we decided to perform a structural characterization of the NCoA1^{257–385}/STAT6^{783–814} complex by NMR.

Mapping of the NCoA1 PAS-B domain binding site on STAT6^{783–814}. The STAT6^{783–814}-NCoA1 PAS-B interaction was first described by investigating the structural changes of the STAT6^{783–814} peptide upon binding with the coactivator. Therefore, the interaction was monitored by acquiring the ¹H, ¹⁵N-HSQC spectrum of ¹⁵N-¹³C-labeled STAT6^{783–814} in complex with the unlabeled PAS-B domain of the coactivator. A subset of STAT6^{783–814} backbone amide resonances becomes well dispersed in the presence of the binding partner (Fig. 3A). These changes are indicative of a well-structured region within the bound STAT6^{783–814} peptide. In particular, the resonance peaks of the residues from Thr⁷⁹⁸ to Gly⁸¹¹ next to the LXXLL motif are markedly perturbed when STAT6^{783–814} is bound to the NCoA1 PAS-B domain, and among these, the three leucine residues (Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶) are strongly perturbed upon binding (Fig. 3B,C). In agreement with the X-ray data reported previously⁸, these findings indicate that the recognition mechanism of NCoA1 cofactor by STAT6 is principally mediated by the region containing the LXXLL motif with the three leucine residues playing an important role in the complex formation. In addition, in agreement with the X-ray structure (PDB ID: 5NWX), the residues Leu⁷⁹⁴ and Leu⁷⁹⁵ located at the N-terminal region of the peptide show small but significant chemical shift variations upon binding. Furthermore, Phe⁷⁹¹ also shows chemical shift changes suggesting that the complex may be further stabilized by additional residues flanking the LXXLL motif.

Chemical shift assignments and conformational analysis of the NCoA1 PAS-B domain in complex with STAT6^{783–814}. The PAS-B domain of NCoA1 adopts in the complex a stable folded structure (Fig. 4A). A nearly complete ¹H, ¹³C and ¹⁵N assignment of the NCoA1^{257–385} domain has been obtained using standard triple resonance experiments (see Materials and Methods). More than 91% of the backbone resonances (¹H_N, ¹⁵N, ¹³C α , and ¹³CO) and 89% of all side chain ¹³C and ¹H resonances were assigned. Secondary structure elements of NCoA1^{257–385} were identified by the analysis of the chemical shift index¹⁴, and then confirmed by ³JH_NH α coupling values and hydrogen exchange experiments (Fig. 4B). Specifically, the ³JH_NH α coupling values drop for the two helices identified before in the free peptide to values indicating close to 100% helix formation. Indeed, as shown in Fig. 4B, NCoA1^{257–385} in complex with the STAT6^{783–814} peptide preserves all secondary

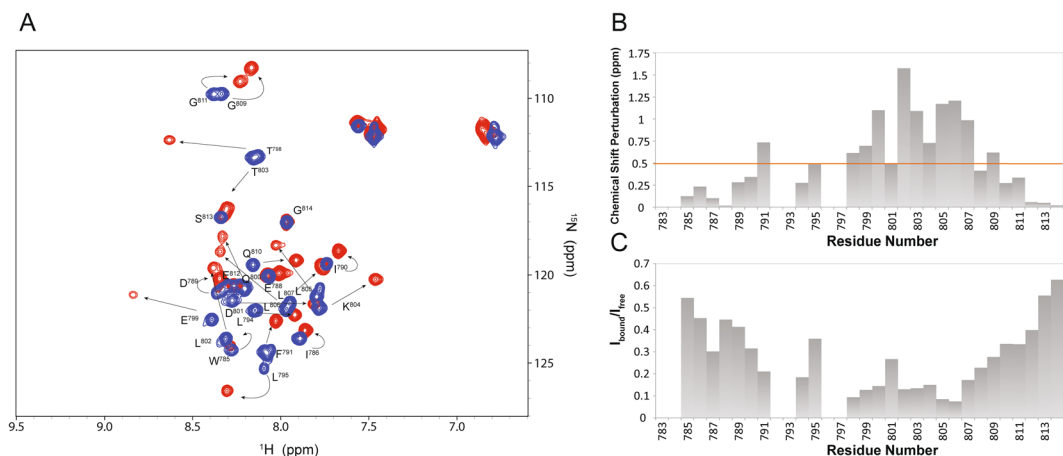


Figure 3. NMR analysis of STAT6 $^{783-814}$ binding to NCoA1 PAS-B domain. (A) Overlay of the ^1H , ^{15}N -HSQC spectra of STAT6 $^{783-814}$ in free form (blue) and in complex with the NCoA1 PAS-B domain (red). (B) Chemical shift perturbations (ppm) of STAT6 $^{783-814}$ upon binding plotted versus the primary sequence; the orange line indicates the mean value. (C) Signal intensity ratios of STAT6 $^{783-814}$ bound to NCoA1 PAS-B domain (I_{bound}) and free (I_{free}).

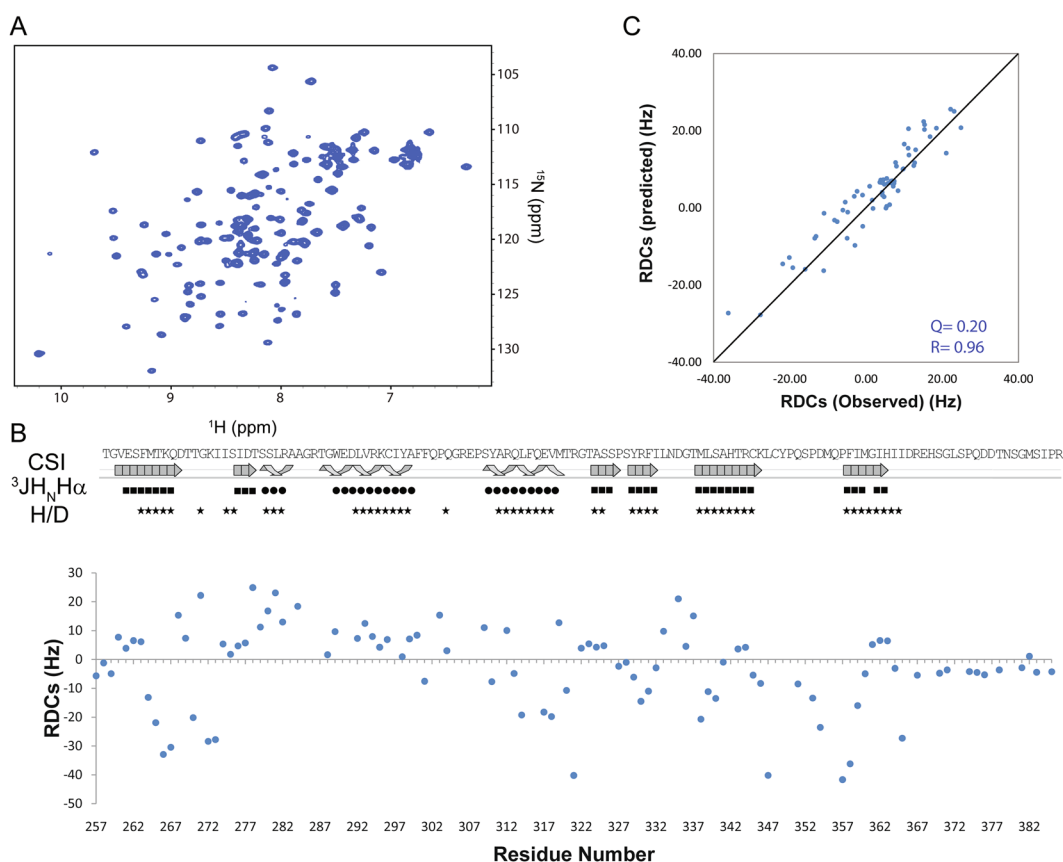


Figure 4. Conformational analysis of the NCoA1 PAS-B domain bound to STAT6 $^{783-814}$ by using NMR and X-ray (PDB ID: 5NWX) data. (A) ^1H , ^{15}N -HSQC spectrum of the ^{15}N - ^{13}C NCoA1 $^{257-385}$ in complex with unlabeled STAT6 $^{783-814}$. (B) RDC values of NCoA1 $^{257-385}$ in bound form measured at 900 MHz. A scheme of the secondary structure elements of NCoA1 $^{257-385}$ in dependence of the protein sequence shown on top as derived by the Chemical Shift Index (CSI) based on C_α and H_α resonance assignments. $^3\text{J}_{\text{H}_\text{N}\text{H}_\alpha}$ coupling constants are also reported and indicated by filled circles and filled squares for values of $^3\text{J}_{\text{H}_\text{N}\text{H}_\alpha} < 4.5$ Hz or > 8 Hz, respectively. The slowly exchanging amide protons are indicated by stars in the H/D line. (C) Plot of measured RDCs versus those calculated from the x-ray structure reported in this manuscript (PDB ID: 5NWX) using PALES 49 .

structure elements reported in the crystallographic structure of the PAS-B domain in complex with the shorter peptide STAT6^{794–814} (PDB ID: 1OJ5) as well as the structure with STAT6^{783–814} (PDB ID: 5NWX) with the exception of the helix Ile⁷⁸⁶ - Ile⁷⁹⁰ which is fully populated in solution, however, there is no electron density in the crystal structure. To check this result independently and get more insight into the secondary and especially of the tertiary structure adopted by the NCoA1 PAS-B domain in complex with the STAT6^{783–814} we analyzed RDCs that are a sensitive probe of local structure as well as of protein motions^{15,16}. In particular, we used the ¹D_{NH} RDCs to detect if the PAS-B domain of the coactivator undergoes local structural variations, in order to bind with high affinity the STAT6^{783–814} peptide. Weak alignment of the selectively ¹⁵N-labeled NCoA1^{257–385}/STAT6^{783–814} complex was achieved by the addition of filamentous bacteriophage Pf1¹⁷. Large ¹D_{NH} RDCs (40 Hz) (Fig. 4B) were obtained for the complex, which indicated substantial alignment and allowed for high sensitivity RDC measurement. In order to understand whether the PAS-B domain bound to STAT6^{783–814} adopts in solution a different conformation with respect to that observed in the crystal of the NCoA^{257–385}/STAT6^{794–814} complex we calculated theoretical RDC values from the X-ray structure reported in this manuscript (PDB ID: 5NWX). Alignment tensors were determined employing a linear fit procedure¹⁸ using the X-ray structure (PDB ID: 5NWX) and the measured RDCs, but considering only the residues located in the region having a secondary structure in accordance with the chemical shifts, ³J_{H_NH_α coupling and hydrogen exchange data. Using these alignment tensors together with the crystal coordinates, RDC values were predicted for the entire protein and then compared with the experimental RDC values measured on the NCoA^{257–385}/STAT6^{783–814} complex. All RDC values throughout the entire protein, apart from the additional C-terminal tail, which is missing in the crystal structure, are in good agreement with the crystal structure as reflected by the Q factor value (Q = 0.20) and by the Pearson's correlation coefficient (R = 0.96) (Fig. 4C). This demonstrates that the NCoA1 PAS-B domain in complex with the longer STAT6^{783–814} adopts a conformation similar to the crystal structure (PDB ID: 5NWX) and that any dynamic variations compared to the crystal structure coordinates must be small.}

Identification of the NCoA1 PAS-B/STAT6^{783–814} binding interface. The structural features of the binding mode of the coactivator NCoA1 with STAT6 was determined by intermolecular NOEs measured using an ¹³C-edited/¹²C-filter NOESY-HSQC experiment¹⁹, exploring the ¹⁵N-¹³C enriched STAT6^{783–814} peptide and the natural abundance NCoA1 PAS-B domain. Based on the intermolecular NOEs, the interface formed by NCoA1 with STAT6^{783–814} appears to be considerably larger compared to the shorter peptide STAT6^{794–814}. A large number of intermolecular NOEs is observed between the side chains of Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶ of STAT6^{794–814} and the NCoA1^{257–385} residues Ile²⁷², Ile²⁷³, Ser²⁷⁴, Thr²⁷⁷, Trp²⁸⁸, Val²⁹², Arg²⁹³ and Tyr²⁹⁷ (Fig. 5A). These findings, confirming the crystal structure of the NCoA1^{257–385}/STAT6^{794–814} complex, indicate that the three leucine residues of the LXXLL motif play indeed a crucial role in the molecular recognition process. Interestingly, the intermolecular NOE analysis indicates that the NCoA1^{257–385}/STAT6^{783–814} complex is stabilized by interactions of additional residues located in the N-terminal region of the peptide. In particular, in agreement with the X-ray structure of the NCoA1^{257–385}/STAT6^{783–814} complex the side chain of Leu⁷⁹⁴ shows intermolecular NOEs with the residues Gly²⁷⁰, Phe³⁰⁰, Ser³⁰⁸, Ala³¹⁰, Arg³¹¹ and Ile³⁵⁸.

The side chains of Ile⁷⁸⁶ and Ile⁷⁹⁰ that were not observed in the X-ray structure (PDB ID: 5NWX) interact with the NCoA1 residues Leu³⁴⁶, Tyr³⁴⁸, Gln³⁵⁵ and Pro³⁵⁶ (Fig. 5A). Notably, the NCoA1 residues showing intermolecular NOEs with the STAT6-derived peptide create a continuous patch on the surface of the PAS-B domain defining an extended binding interface (Fig. 5B). Residues Ile⁷⁸⁶ and Ile⁷⁹⁰ were not present in the peptide crystallized in the complex with the NCoA1 domain⁸ (PDB ID: 1OJ5) and are also not visible in the x-ray structure described above (PDB ID: 5NWX), such that their interaction is only observed in the solution structure and may be responsible for the enhanced binding.

Structure of the NCoA1^{257–385}/STAT6^{783–814} complex by NMR. The NMR structure of the NCoA1^{257–385}/STAT6^{783–814} complex (PDB ID: 5NWM), is of high quality (Table SI 4). In the complex the PAS-B domain adopts a well-defined globular fold (rmsd_{BackboneAtoms}^{260–367} = 0.485 Å) (Fig. 6A) ranging from Glu²⁶⁰ to Glu³⁶⁷ and a dynamically disordered tail at the C-terminus as confirmed by ¹⁵N-¹H heteronuclear NOE values (SI Fig. 4). The PAS-B domain shows all structural features known already from the two crystal structures (PDB IDs: 1OJ5, 5NWX) (SI Fig. 5) with a five-stranded anti-parallel β-sheet and three α-helices that connect the second and third β-strand (Fig. 6B).

The STAT6^{783–814} peptide in complex with the NCoA1 PAS-B domain presents a short flexible N-terminal tail with a one-turn α-helix (α1) in the region from Ile⁷⁸⁶-Ile⁷⁹⁰ (as confirmed by the low values of the ³J_{H_NH_α couplings and the NOEs (Fig. 2C) that is connected by a linker that adopts an extended conformation to the second α-helix (α2) formed by the residues from Glu⁷⁹⁹ to Glu⁸⁰⁸. The STAT6 peptide binds into a shallow groove at the surface of the NCoA1 PAS-B domain, (Fig. 6C,D) with the specific contacts already known from the crystal structures (SI Fig. 5). Shortly, Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶ form the major hydrophobic side chain contacts (Fig. 7A) while Pro⁷⁹⁶ and Pro⁷⁹⁷ make contact with the residues Ile²⁷² and Phe³⁰⁰ located in the C-terminal part of helix α3 of the PAS-B domain (Fig. 7B). These findings are also in line with the alanine scanning mutagenesis data of Robinson and coworkers⁹. They found that the L802A, L806A and P797A single mutants abolished the NCoA1^{257–385}/STAT6^{794–814} complex formation in their FP assay and that the L805A mutant reduced the affinity considerably.}

Importantly, the interaction between the NCoA1 PAS-B domain and STAT6^{783–814} is further stabilized by the interaction of STAT6 Leu⁷⁹⁴ with a shallow hydrophobic indentation constituted by the PAS-B domain residues Phe³¹⁴, Phe³⁰⁰ and Ala³¹⁰ (Fig. 7C), seen also in the X-ray structure and in agreement with the alanine scanning mutagenesis data where F794A reduced the affinity about 50-fold⁹. In addition, the two isoleucine residues (Ile⁷⁸⁶, Ile⁷⁹⁰) located in the N-terminal part of the STAT6 peptide bind into a shallow hydrophobic pocket on the surface of the PAS-B domain formed by Phe³¹⁴, Met³¹⁸, Leu³⁴⁶, Pro³⁵⁶ and Ile³⁵⁸ (Fig. 7D). Moreover, also STAT6 Trp⁷⁸⁵ slightly contributes to stabilize this interaction making contacts with Pro³⁵⁶ of the PAS-B domain. Altogether,

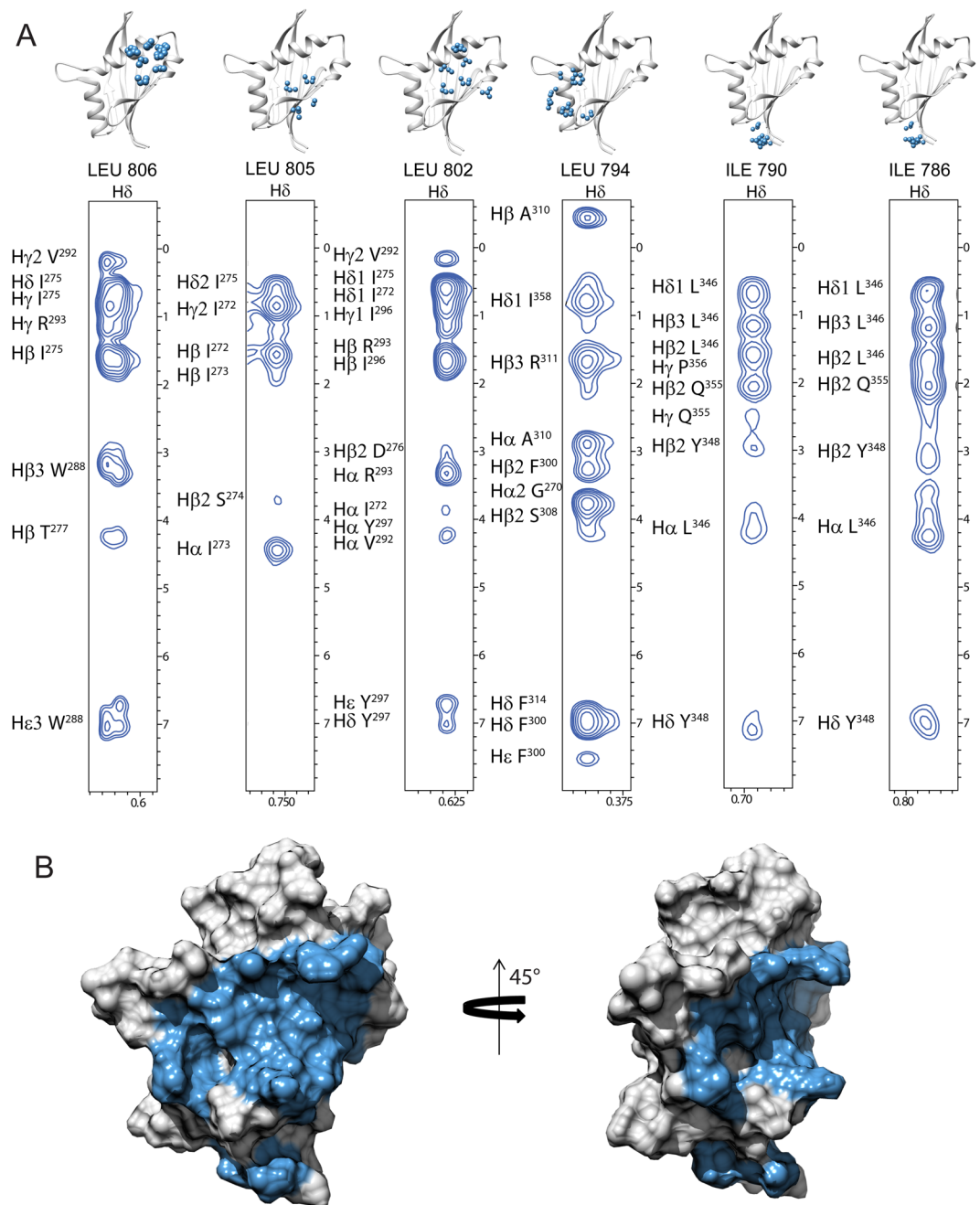


Figure 5. Mapping of the STAT6 binding surface on the NCoA1 PAS-B domain (PDB ID: 1OJ5). **(A)** Strips from the 3D ^{13}C -edited/ ^{12}C -filter NOESY experiment measured on the unlabeled NCoA1 PAS-B domain in complex with ^{15}N - ^{13}C STAT6 $^{783-814}$ showing inter-molecular NOEs. The strips are related to the residues STAT6 Leu 806 , Leu 805 , Leu 802 , Leu 794 , Ile 790 and Ile 786 directly involved in the interaction. **(B)** Mapping of the residues of NCoA1 involved in the interaction onto the x-ray structures (PDB ID: 1OJ5) in two orientations rotated by 45° around the z-axis.

the NMR structure indicates that the coactivator recognition mechanism by STAT6 occurs by the formation of a complex in which two folded regions are connected by a linker that adopts an extended conformation. This backbone conformation is further stabilized due to the side chain interaction of Leu 794 with NCoA1. Thus despite the invisibility of the peptide N-terminal to Leu 794 in the X-ray structure, the complex formation between the NCoA1 PAS-B domain and STAT6 $^{783-814}$ is well defined by specific interactions in solution. This is in contrast to a fuzzy complex 20,21 as it has been identified e.g. for the interaction between the transcription activator GCN4 and a subunit of the Mediator complex 22 .

Last, in both X-ray structures as well as in the NMR structure the C-terminal region of the STAT6 peptide downstream of Glu 808 is disordered, in line with the FP assay data with C-terminally truncated peptides that still bound with high affinity to the NCoA1 PAS-B domain 9 .

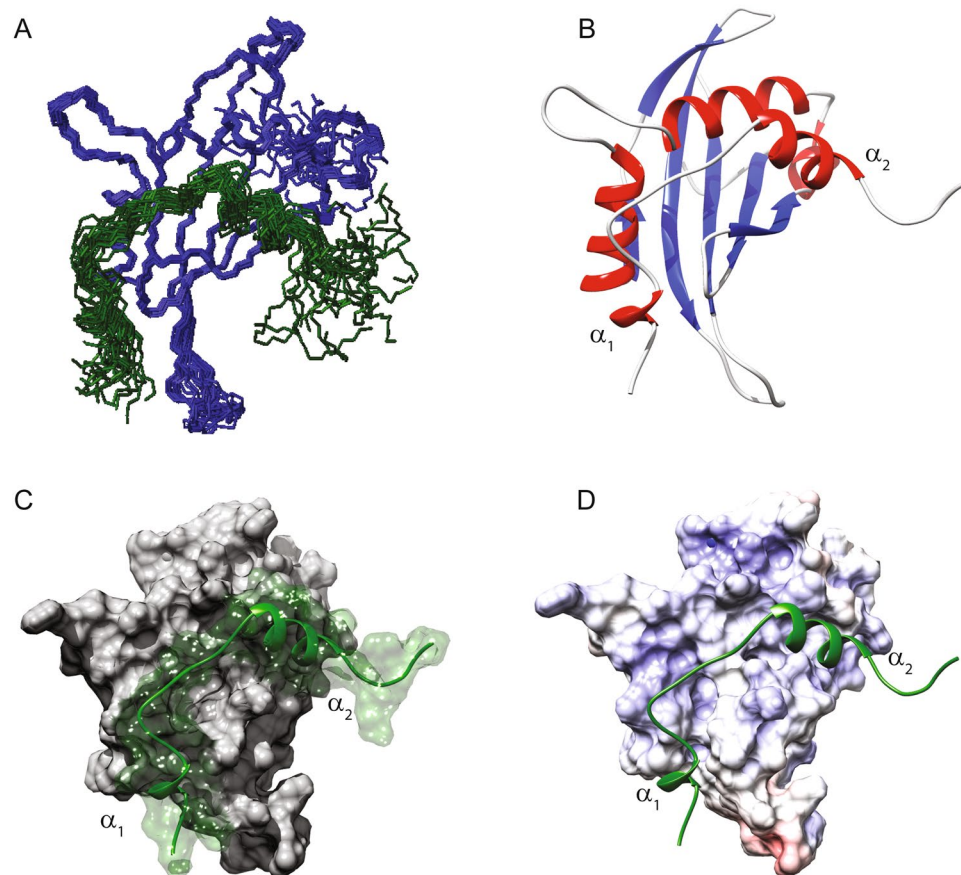


Figure 6. The NMR structure of the NCoA1²⁵⁷⁻³⁸⁵/STAT6⁷⁸³⁻⁸¹⁴ complex (PDB ID: 5NWM). **(A)** Overlay of the 20 lowest energy structures of the NCoA1²⁵⁷⁻³⁸⁵/STAT6⁷⁸³⁻⁸¹⁴ complex. **(B)** Ribbon drawing of one representative conformer of the NMR structure NCoA1²⁵⁷⁻³⁸⁵/STAT6⁷⁸³⁻⁸¹⁴ complex. **(C)** Solvent accessible surface of the NCoA1²⁵⁷⁻³⁸⁵/STAT6⁷⁸³⁻⁸¹⁴ complex. The NCoA1 PAS-B domain is depicted in light grey while the STAT6 peptide is reported in dark green. **(D)** Electrostatic surface of NCoA1²⁵⁷⁻³⁸⁵ bound to the STAT6⁷⁸³⁻⁸¹⁴ peptide. The positively charged residues are depicted in blue while the negatively charged residues are in red. STAT6⁷⁸³⁻⁸¹⁴ is reported in dark green.

Tertiary structural preorganization of free STAT6⁷⁸³⁻⁸¹⁴ before forming the NCoA1²⁵⁷⁻³⁸⁵/STAT6⁷⁸³⁻⁸¹⁴ complex.

We then set out to investigate whether in addition to the preformation of secondary structure elements, there is also a preformation of tertiary structure, i.e. a preorganization of the relative orientation of the two α -helices in the free STAT6 peptide, to facilitate binding to NCoA1²⁵⁷⁻³⁸⁵. Therefore, we evaluated the RDCs measured for STAT6⁷⁸³⁻⁸¹⁴ in the free (Table SI 5) and bound forms and compared them to the values predicted from the NMR structure of the complex (PDB ID: 5NWM). Remotely related are earlier studies where proteins were unfolded in urea or guanidinium hydrochloride and the RDCs were compared with the RDCs from the folded forms^{23,24} and the 3D topology was retained in part. Indeed, RDCs are faithful reporters on the relative orientation of structural segments to each other. For the bound form we find a Q factor of 0.17 which is slightly higher than for the PAS-B domain ($Q = 0.14$) but, as reported by the normal scalar product NSP, with a similar alignment tensor (Table SI 6). Accordingly, fitting of the experimental RDCs of STAT6⁷⁸³⁻⁸¹⁴ bound to PAS-B by using the tensor derived from the RDCs of the PAS-B domain resulted in a Q factor of 0.23 (Table SI 6). Taking now the experimental RDCs of the free form (ranging from 1 to 10 Hz) (Table SI 5), the Q factor increased only to 0.32. In addition, for several structural models (SI Fig. 5) in which the N-terminal helix was rotated by 30° about any axis perpendicular to the axis of the C-terminal helix, the Q factor increased to values between 0.36 and 0.54 (Table SI 7). Finally, in order to consider in our analysis the conformational heterogeneity of the STAT6⁷⁸³⁻⁸¹⁴ in the free form, we generated, as reported in the materials and methods and illustrated in the supplementary information (SI Fig. 7A–C), a pool of random conformers to properly describe the conformational space sampled by the peptide in solution. Interestingly, for all models of the conformational ensemble (SI Fig. 8A), we find Q factors higher than 0.32 for fitting the RDCs of the free form of STAT6⁷⁸³⁻⁸¹⁴ (SI Fig. 8B).

Even in this much more exhaustive ensemble, we find a Q factor close to 0.32 only for two structural models, namely SM20 and SM21. For these models the angle between the α -helices (interhelical angle θ) is $\theta_{SM20} = 51^\circ$, $\theta_{SM21} = 89^\circ$ compared to $\theta_{NMR} = 134^\circ$ (SI Fig. 8C) which will be further discussed below (see Discussion).

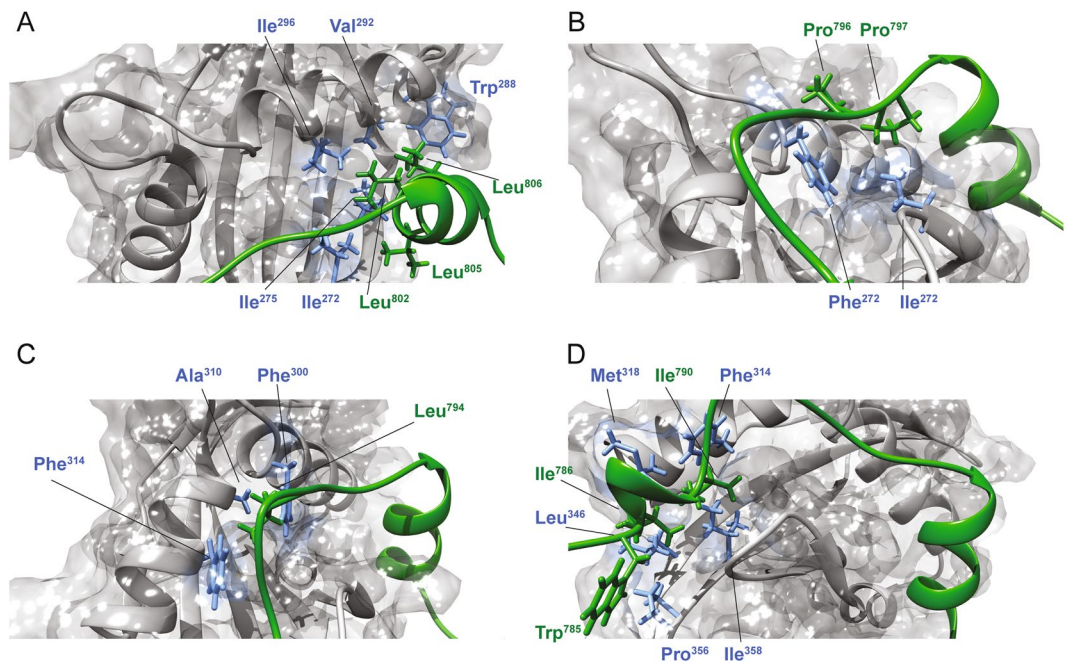


Figure 7. Structural details of the NCoA1^{257–385}/STAT6^{783–814} interface as reported by the NMR structure (PDB ID: 5NWM). **(A)** STAT6 Leu⁸⁰², Leu⁸⁰⁵ and Leu⁸⁰⁶ fit into the binding groove of the NCoA1 PAS-B domain. **(B)** STAT6 Pro⁷⁹⁶ and Pro⁷⁹⁷ interact with residues Ile²⁷², Ile²⁹⁶, Tyr²⁹⁷ and Phe³⁰⁰ located in the C-terminal part of helix α 3 of the NCoA1 PAS-B domain. **(C)** STAT6 Leu⁷⁹⁴ is embedded in a shallow hydrophobic depression constituted by the residues Phe³¹⁴, Phe³⁰⁰ and Ala³¹⁰ of the NCoA1 PAS-B domain. Ile⁷⁹⁰ is inside a less deep hydrophobic pocket formed by Phe³¹⁴, Met³¹⁸ and Ile³⁵⁸. **(D)** STAT6 Ile⁷⁸⁶ interacts with Met³¹⁸, Leu³⁴⁶, Pro³⁵⁶ and Ile³⁵⁸ of the NCoA1 PAS-B domain.

These findings indicate that indeed the tertiary structural elements are preorganized in the free form to facilitate binding. This finding differs from the mentioned studies^{23,24} in the sense that the “unfolding conditions” for STAT6^{783–814} are constituted by the absence of the binding partner PAS-B and don’t require chemical denaturants.

Discussion

How molecular recognition and binding occurs between highly flexible protein domains, is not yet well understood. The conformational selection theory provides a very elegant explanation for molecular recognition, especially in the context of partially structured protein regions^{25–28}.

A detailed understanding of the fundamental mechanisms of the molecular recognition of the coactivator NCoA1 by STAT6 is central to understanding biology at the molecular level of this interaction. The NMR structure of the NCoA1^{257–385}/STAT6^{783–814} complex indicates that the coactivator recognition mechanism by STAT6 occurs by the formation of a partially ordered complex in which two α -helical regions (Ile⁷⁸⁶-Ile⁷⁹⁰ and Glu⁷⁹⁹-Glu⁸⁰⁸) are connected by a linker that adopts an extended, albeit dynamic, conformation. This dynamics in the linker may act more strongly onto the N-terminal region of the STAT6^{783–814} peptide that has fewer interactions with the NCoA1 PAS-B domain than the C-terminal region, ultimately rendering residues 783–793 disordered in the X-ray structure.

As mentioned above, the conformational characterization of STAT6^{783–814} in the free form indicates that the peptide is prestructured in the two regions Ile⁷⁸⁶-Ile⁷⁹⁰ and Glu⁷⁹⁹-Glu⁸⁰⁸ with a significant propensity to occupy α -helical conformation, 10% and 30%, respectively. Regarding the STAT6^{783–814} peptide in the bound form, the analysis from chemical shifts, ³JH_NH α couplings and the relative weight of interresidual NOEs (Fig. 2C) suggests that the STAT6^{783–814} peptide presents a flexible N-terminal tail, followed by an α -helix in the region Ile⁷⁸⁶-Ile⁷⁹⁰ whose population increases from around 10% in the free form to close to 100% upon binding to NCoA1 PAS-B (Fig. 2C) (SI Fig. 9). Then a short extended linker follows and finally the second α -helix (Glu⁷⁹⁹-Glu⁸⁰⁸) (Fig. 2A,B) whose population increases from 30% to close to 100%. Given that existing prestructured motifs are enhanced in the bound form, the data suggest that the recognition of the coactivator by STAT6 occurs by a conformational selection²⁵ mechanism regarding the secondary structure. We then investigated whether there is also preorganization of the binding motif, i.e. the two helices on the tertiary structural level and fitted experimental RDCs of the free and bound forms to the bound structure and a distorted bound structure. We find indeed that the experimental RDCs of the free peptide fit well to the bound structure and that a rigid body rotation of the N-terminal helix away from this orientation as well as the use of additional conformers with larger conformational heterogeneity, deteriorates the fit. These findings indicate that not only on the secondary but also on the tertiary level, i.e. the arrangement of the two helices, there are prestructure motifs which are then also found in the bound form.

Overall the data supports conformational selection^{25–28} in this region as the key mechanism driving the molecular recognition of the NCoA1 PAS-B domain by STAT6 in which the coactivator strongly shifts the α -helical

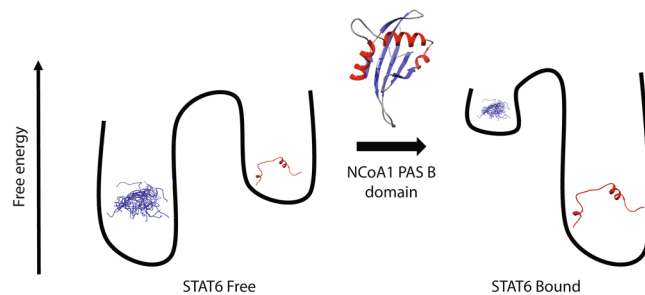


Figure 8. Molecular recognition mechanism. Cartoon representation of the conformational selection mechanism. The binding of the coactivator changes the free-energy landscape of STAT6^{783–814}.

propensity in the regions Ile⁷⁸⁶-Ile⁷⁹⁰ and Glu⁷⁹⁹-Glu⁸⁰⁸ to an even more populated helical state (Fig. 8). In particular, the two helices are present at 10% and 30% population in the free form and become fully populated in the bound form making conformational selection the probable binding mechanism. The 3D arrangement of the two helices found in the bound form is also very likely to prevail in the free form based on RDCs. This suggests a conformational selection not only for the secondary but also for the tertiary structure. This similarity of 3D arrangement of secondary structures in the free and bound form was further investigated by exhaustive sampling of the conformational space. In 77 clusters that deviated sufficiently from the bound structure we found only two that had a similar quality factor for the RDCs. This is still a very strong support for the similarity of the 3D arrangement in the free and bound structure for the following reason: It should be noted that the binding peptide comprises two helices in which the NH vectors point along the helix axis and indeed there is little variation of the respective RDCs (+6 Hz for the C-terminal helix and around -1 Hz for the N-terminal helix). In addition, the peptide contains an extended stretch (F⁷⁹¹, L⁷⁹⁴, and L⁷⁹⁵) where also little variation of the dipolar couplings is observed (around -9 Hz). For an accurate tensor determination one needs normally at least 5 independent orientations. Thus, it is not surprising that there is symmetry related degeneracies of possible orientations of the helices which fit to the RDCs. Even if these additional structures were populated, the conformational space would still be restricted to an ensemble with less members than a totally disordered peptide, supporting the mechanistic conclusion that a preformation of the bound 3D conformation happens in the free form.

Conformational selection has been observed for quite a number of interactions between structurally well ordered domains and disordered transactivation domains containing an amphipathic α -helical binding motif (Φ XX Φ , Φ being a bulky hydrophobic residue) of transcription factors, e.g. the interaction of the transactivation domain of c-Myb with the KIX domain of CBP/p300²⁹. In this complex as in other such complexes additional interactions outside the core binding motif that confer specificity have been structurally resolved. But to our knowledge the combination of a highly ordered amphipathic α -helical binding motif interaction with another N-terminally located second α -helical binding motif of equally high specificity in combination with a prearranged 3D arrangement of these helices is so far uniquely observed in the NCoA1 PAS-B/STAT6 transactivation domain complex described in this work and likely explains the higher affinity of STAT6^{783–814} compared to STAT6^{794–814} to the NcoA1 PAS-B domain.

Materials and Methods

Protein expression and purification. The recombinant NCoA-1 PAS-B domain comprising amino acids 257–385 of NCoA-1 was expressed and purified as previously published⁸. The fragment containing amino acids 783–814 of STAT6 was produced recombinantly as an N-terminal Z-tag fusion protein with a His₆-tag and a TEV cleavage site between Z-tag and the target sequence. After TEV cleavage and removal of the Z-tag by Ni-NTA resin (Qiagen) the STAT6 peptide was further purified by reversed phase HPLC. Expression of labeled protein was performed in Toronto minimal medium with ¹⁵N ammonium chloride as nitrogen source and U-¹³C₆-D-glucose as carbon source. The STAT6 peptide was added in 1.5 fold molar excess and the complex of the PAS-B domain with this peptide was purified by gel filtration on a Superdex 75 column (GE Healthcare).

NMR spectroscopy. NMR samples were made up in the following buffer: 50 mM HEPES, pH 7.0, 150 mM NaCl, 2 mM DTT and 10% ²H₂O. Complex samples were made up by adding STAT6 peptide in 1.5 fold molar excess. All spectra were recorded at a total protein concentration of 1 mM and were carried out at 309 K on Bruker 600, 700, 800, 900 NMR spectrometers equipped with a cryogenic probe and on Bruker 600, 700 NMR spectrometers equipped with a triple resonance probe head. The following experiments were recorded:

- on samples of ¹⁵N- or ¹⁵N-¹³C-labeled STAT6^{783–814} in the free form ¹H, ¹⁵N-HSQC, ¹H, ¹³C-HSQC, HNHA, HNCA, HNCACB, CBCACONH, ¹H, ¹⁵N NOESY-HSQC and ¹H, ¹⁵N TOCSY-HSQC³⁰.

- on sample of ¹⁵N-¹³C-labeled STAT6^{783–814} in complex with unlabeled NCoA1^{257–385} ¹H, ¹⁵N-HSQC, ¹H, ¹³C-HSQC aliphatic, ¹H, ¹³C-HSQC aromatic, HNHA, HNCA, HNCACB, CBCACONH, ¹H, ¹⁵N NOESY-HSQC, ¹H, ¹⁵N TOCSY-HSQC, HCCH-TOCSY, ¹H, ¹³C NOESY-HSQC aliphatic, ¹H, ¹³C NOESY-HSQC aromatic³⁰, ¹H, ¹³C-edited/¹²C-filter NOESY-HSQC¹⁹.

- on sample of ¹⁵N-¹³C-labeled NCoA1^{257–385} in complex with unlabeled STAT6^{783–814} ¹H, ¹⁵N-HSQC, ¹H, ¹³C-HSQC aliphatic, ¹H, ¹³C-HSQC aromatic, HNHA, HNCA, HNCACB, CBCACONH, ¹H, ¹⁵N NOESY-HSQC, ¹H, ¹⁵N TOCSY-HSQC. HCCH-TOCSY, ¹H, ¹³C NOESY-HSQC aliphatic, ¹H, ¹³C NOESY-HSQC aromatic.

The ^{15}N edited NOESY-HSQC and ^{13}C edited NOESY-HSQC experiments were acquired with a mixing time of 100 ms and 80 ms, respectively.

Slowly exchanging amide protons were identified in an ^1H , ^{15}N -HSQC spectrum reordered immediately after exchanging the proton into a buffer prepared with $^2\text{H}_2\text{O}$.

Vicinal (three-bond) HN-H α coupling constants ($^3J_{\text{HNH}\alpha}$) were evaluated from cross-peak intensities in quantitative J-correlation (HNHA) spectra³¹. Backbone torsion angles were estimated from C α , CO, C β , N, HN and H α chemical shift using the program TALOS+³².

Residual dipolar couplings $^1D_{\text{NH}}$ RDCs for the ^{15}N -labeled STAT6^{783–814}/U-NC α A1^{257–385} and U-STAT6^{783–814}/ ^{15}N -labeled NC α A1^{257–385} were measured by taking the difference in the one-bond ^1H - ^{15}N splittings ($^1J_{\text{NH}} + ^1D_{\text{NH}}$) in aligned (~ 20 mg/ml phage pf1¹⁷) and isotropic media using an in-phase/anti-phase (IPAP) HSQC experiment³³.

The observed chemical shift change ($\Delta\delta_{\text{obs}}$) for each backbone amide between the STAT6^{783–814} in the free and bound form was measured as the weighted average of the proton and nitrogen chemical shift changes by using equation (1)³⁴:

$$\Delta\delta_{\text{obs}} = [(\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}}^2/25)/2]^{1/2} \quad (1)$$

For the evaluation of the ^{15}N -[^1H] steady-state heteronuclear NOE two ^1H , ^{15}N -HSQC were acquired (in one the protons were unsaturated and in the other the protons were saturated for 3 s).

For the secondary structure analysis and the secondary structure propensity the random coil chemical shifts of Zhang *et al.*¹² were used. The SSP scores were calculated with the random coil chemical shifts and the average secondary shifts for fully formed secondary structure¹² as described previously¹³.

The evaluation of the alignment tensors was performed by using the normalized scalar product (NSP) defined by equation (2):

$$\text{NSP} = \frac{\langle S^{\text{sample1}} | S^{\text{sample2}} \rangle}{\sqrt{\langle S^{\text{sample1}} | S^{\text{sample1}} \rangle \langle S^{\text{sample2}} | S^{\text{sample2}} \rangle}} \quad (2)$$

where the scalar product between the two vectors formed from the Saupe matrix elements can be defined according to equation (3):

$$\langle S^{\text{sample1}} | S^{\text{sample2}} \rangle = \sum_{\substack{i=x,y,z \\ j=x,y,z}} S_{ij}^{\text{sample1}} S_{ij}^{\text{sample2}} \quad (3)$$

in which S_{ij} are the elements of the 3×3 Saupe matrices.

The values of the NSP close to 1.0 indicate that the two alignment tensors differ only by a scaling factor, whereas the values around 0.0 suggest that the alignment frames are orthogonal. The situation where NSP = -1 indicates that the two alignment tensors are antiparallel.

The quality factor (Q) is defined by equation (4)³⁵:

$$Q = \frac{(\sum_{i=1}^N (D_i^{\text{exp}} - D_i^{\text{calc}})^2 / N)^{1/2}}{(\sum_{i=1}^N (D_i^{\text{exp}})^2 / N)^{1/2}} \quad (4)$$

The spectra were processed using NMRpipe³⁶ and analyzed using SPARKY³⁷ and CARA³⁸. ^1H , ^{13}C and ^{15}N chemical shifts were calibrated indirectly by external DSS references.

Structure Calculation. NOE-derived distance constraints, coupling constants, TALOS dihedral angles, hydrogen bonds and residual dipolar couplings were used to calculate the structure of the NC α A1^{257–385}/STAT6^{783–814} complex with the program CYANA 3.0³⁹. The input data for the final structure are reported in Table SI 4. A total of 100 structures was calculated, and the 20 conformers with the lowest CYANA target function were selected. The small number of residual constraint violations indicates that the input data represent a self-consistent set and that the constraints are well satisfied in the calculated conformers. The structures were visualized and evaluated by using the programs, MOLMOL⁴⁰, CHIMERA⁴¹, PROCHECK-NMR⁴² and MOLPROBITY⁴³. The Adaptive Poisson-Boltzmann Solver (APBS)⁴⁴ was used to calculate spatial distributions of electrostatic potentials using the linearized Poisson-Boltzmann equation and parameters from the PQR files obtained using the PDB2PQR server⁴⁵. The electrostatic map was generated with CHIMERA⁴¹.

Generation of the conformational ensemble for RDCs evaluation. A conformational sampling approach was used to generate a conformational ensemble of the free STAT6 in order to describe the conformational space sampled by the peptide in the absence of the binding partner more comprehensively. The Normal Mode-based Simulation (NMSim)⁴⁶ approach has been shown to be a computationally efficient alternative to molecular dynamics simulation for conformational sampling of proteins. Therefore, starting from the NMR structure of the STAT6 peptide in complex with the co-activator, an ensemble of 2500 conformers was generated with the program NMSim⁴⁶ by using the default parameters for large scale motions. Then, the conformational ensemble was clustered by using the software NMRCLUST⁴⁷ implemented in the program CHIMERA⁴¹. A total of 153 clusters were found and for each of them the representative structure was considered as reference model. Successively, the representative models were further filtered by considering only the conformers in which the θ angle between the first and the second α -helix deviated from this angle in the NMR structure by more than

$\pm 30^\circ$, i.e. all structures were excluded where this angle was between 104° and 164° . This procedure resulted in the selection of 77 conformers which defined the final conformational ensemble for the evaluation of the RDCs. The structural models (SM1, SM2, SM3, SM4) of the STAT6 peptide obtained from the NMR structure of the complex by rotating the $\alpha 1$ helix by 30° with respect to the $\alpha 2$ helix were also included in the final ensemble.

Data availability. Coordinates and structural restraints for the NMR structure and the X-ray of NCoA1^{257–385}/STAT6^{783–814} complex have been deposited in the PDB under the accession number (PDB ID: 5NWM) and (PDB ID: 5NWX), respectively. The chemical shifts have been deposited in the BioMagResBank, accession number (BMRB ID: 34131). The conformational ensemble for RDCs evaluation generated during the current study is available from the corresponding author upon request.

References

- Xu, X., Sun, Y. L. & Hoey, T. Cooperative DNA binding and sequence-selective recognition conferred by the STAT amino-terminal domain. *Science* **273**, 794–797 (1996).
- Horvath, C. M., Wen, Z. & Darnell, J. E. A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev* **9**, 984–994 (1995).
- Stahl, N. *et al.* Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349–1353 (1995).
- Darnell, J. E. Jr. STATs and gene regulation. *Science* **277**, 1630–1635 (1997).
- Darnell, J. E., Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415–1421 (1994).
- Smerz-Bertling, C. & Duschl, A. Both interleukin 4 and interleukin 13 induce tyrosine phosphorylation of the 140-kDa subunit of the interleukin 4 receptor. *J Biol Chem* **270**, 966–970 (1995).
- Litterst, C. M. & Pfitzner, E. An LXXLL motif in the transactivation domain of STAT6 mediates recruitment of NCoA-1/SRC-1. *J Biol Chem* **277**, 36052–36060, <https://doi.org/10.1074/jbc.M203556200> (2002).
- Razeto, A. *et al.* Structure of the NCoA-1/SRC-1 PAS-B domain bound to the LXXLL motif of the STAT6 transactivation domain. *J Mol Biol* **336**, 319–329 (2004).
- Seitz, M., Maillard, L. T., Obrecht, D. & Robinson, J. A. Molecular characterization of the NCoA-1-STAT 6 interaction. *Chembiochem: a European journal of chemical biology* **9**, 1318–1322, <https://doi.org/10.1002/cbic.200700773> (2008).
- Wishart, D. S. & Sykes, B. D. The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**, 171–180 (1994).
- Kjaergaard, M., Brander, S. & Poulsen, F. M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* **49**, 139–149, <https://doi.org/10.1007/s10858-011-9472-x> (2011).
- Zhang, H., Neal, S. & Wishart, D. S. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* **25**, 173–195 (2003).
- Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* **15**, 2795–2804, <https://doi.org/10.1110/ps.062465306> (2006).
- Wishart, D. S., Sykes, B. D. & Richards, F. M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651 (1992).
- Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S. & Griesinger, C. Interdomain dynamics explored by paramagnetic NMR. *J Am Chem Soc* **135**, 17111–17120, <https://doi.org/10.1021/ja408143f> (2013).
- Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475, <https://doi.org/10.1126/science.1157092> (2008).
- Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**, 1065–1074, <https://doi.org/10.1038/4176> (1998).
- Sass, J. *et al.* Purple Membrane Induced Alignment of Biological Macromolecules in the Magnetic Field. *Journal of the American Chemical Society* **121**, 2047–2055, <https://doi.org/10.1021/ja983887w> (1999).
- Ikura, M. & Bax, A. Isotope-Filtered 2d NMR of a Protein Peptide Complex – Study of a Skeletal-Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin. *Journal of the American Chemical Society* **114**, 2433–2440 (1992).
- Fuxreiter, M. & Tompa, P. Fuzzy complexes: a more stochastic view of protein function. *Adv Exp Med Biol* **725**, 1–14, https://doi.org/10.1007/978-1-4614-0659-4_1 (2012).
- Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2–8, <https://doi.org/10.1016/j.tibs.2007.10.003> (2008).
- Brzovic, P. S. *et al.* The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol Cell* **44**, 942–953, <https://doi.org/10.1016/j.molcel.2011.11.008> (2011).
- Shortle, D. & Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487–489, <https://doi.org/10.1126/science.1060438> (2001).
- Ohnishi, S., Lee, A. L., Edgell, M. H. & Shortle, D. Direct demonstration of structural similarity between native and denatured eglin C. *Biochemistry* **43**, 4064–4070, <https://doi.org/10.1021/bi049879b> (2004).
- Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789–796, <https://doi.org/10.1038/nchembio.232> (2009).
- Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539–546, <https://doi.org/10.1016/j.tibs.2010.04.009> (2010).
- Ma, B., Kumar, S., Tsai, C. J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng* **12**, 713–720 (1999).
- Tsai, C. J., Kumar, S., Ma, B. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci* **8**, 1181–1190, <https://doi.org/10.1110/ps.8.6.1181> (1999).
- Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J Biol Chem* **291**, 6714–6722, <https://doi.org/10.1074/jbc.R115.692020> (2016).
- Sattler, M., Schleucher, J. R. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.
- Vuister, G. W. & A., B. Quantitative J correlation - a new approach for measuring homonuclear 3-bond J(H)(N)(H)(Alpha) coupling-constants in N-15-enriched proteins. *J Am Chem Soc* **115**, 7772–7777 (1993).
- Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213–223, <https://doi.org/10.1007/s10858-009-9333-z> (2009).
- Otting, N., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
- Grzesiek, S. *et al.* The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat Struct Biol* **3**, 340–345 (1996).

35. Cornilescu, G. & Ottiger, M. & A., B. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837 (1998).
36. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
37. Goddard, T. D. & Kneller, D. G. SPARKY 3, University of California, San Francisco. (1999).
38. Keller, R. The Computer Aided Resonance Assignment Tutorial. Cantina-Verlag, Goldau, Switzerland. (2004).
39. Güntert, P. Automated structure determination from NMR spectra. *Eur Biophys J* **38**, 129–143, <https://doi.org/10.1007/s00249-008-0367-z> (2009).
40. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51–55, 29–32, 0263785596000094 [pii] (1996).
41. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612, <https://doi.org/10.1002/jcc.20084> (2004).
42. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477–486 (1996).
43. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375–383, <https://doi.org/10.1093/nar/gkm216> (2007).
44. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041, <https://doi.org/10.1073/pnas.181342398> (2001).
45. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665–667, <https://doi.org/10.1093/nar/gkh381> (2004).
46. Krüger, D. M., Ahmed, A. & Gohlke, H. NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res* **40**, W310–316, <https://doi.org/10.1093/nar/gks478> (2012).
47. Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng* **9**, 1063–1065 (1996).

Acknowledgements

This work was supported by Max Planck Society.

Author Contributions

All authors designed the project, performed the experiments, analysed the data and commented the manuscript. L.R., C.G. and S.B. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-17088-5>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017