Structure and functional characterization of the conserved JAK interaction region in the intrinsically disordered N-terminus of SOCS5

Indu R. Chandrashekaran[1*], Biswaranjan Mohanty[1], Edmond M. Linossi[2,3], Laura F. Dagley[2,3], Eleanor W. W. Leung[1], James M. Murphy[2,3], Jeffrey J. Babon[2,3], Sandra E. Nicholson[2,3] and Raymond S. Norton[1*]

[1] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

[2] The Walter & Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia

[3] The Department of Medical Biology, University of Melbourne, Parkville, Victoria 3052, Australia

Corresponding Authors:

* E-mail: ray.norton@monash.edu

Telephone: +613 99039167

* E-mail: indu.chandrashekaran@monash.edu

Telephone: +613 99039101

Funding

This work was supported in part by the National Health and Medical Research Council (NHMRC), Australia (Program grant 1016647), as well as an NHMRC IRIISS grant 361646, a Cancer Council Victoria grant 1065180 and a Victorian State Government Operational Infrastructure Scheme grant. E.M.L was supported by an Australian Postgraduate Award. R.S.N. acknowledges fellowship support from the NHMRC. J.M.M. and J.J.B acknowledge fellowship support from the Australian Research Council.

Abbreviations: SOCS, suppressor of cytokine signaling; JAK, Janus kinase; STAT, signal transducer and activator of transcription; EGF, epidermal growth factor; JIR, JAK interacting region; NMR, nuclear magnetic resonance; $mSOCS5_{175-244}$, N-terminal conserved region in mouse SOCS5; SH2, Src-homology 2; miR, microRNA; SPR, surface plasmon resonance; TEV, tobacco etch virus; IDP, intrinsically disordered protein; HSQC, heteronuclear single quantum coherence; NOE, nuclear overhauser effect; TCEP, tris(2-carboxyethyl)phosphine; RMSD, root-mean-square deviation

Abstract

SOCS5 can negatively regulate both JAK/STAT and EGF-receptor pathways and has therefore been implicated in regulating both the immune response and tumorigenesis. Understanding the molecular basis for SOCS5 activity may reveal novel ways to target key components of these signaling pathways. The N-terminal region of SOCS5 coordinates critical protein interactions involved in inhibition of JAK/STAT signaling, and a conserved region within the N-terminus of SOCS5 mediates direct binding to the JAK kinase domain. Here we have characterized the solution conformation of this conserved JAK interaction region (JIR) within the largely disordered N-terminus of SOCS5. Using nuclear magnetic resonance (NMR) chemical shift analysis, relaxation measurements and NOE analysis, we demonstrate the presence of pre-formed structural elements in the JIR of mouse SOCS5 ($mSOCS5_{175-244}$). The pre-formed structure consists of an α-helix encompassing residues 224–233, which is preceded by a turn and an extended structure. We have identified a phosphorylation site (Ser211) within the JIR of mSOCS5 and have investigated the role of phosphorylation in modulating JAK binding using site-directed mutagenesis.

The Suppressor of Cytokine Signaling (SOCS) proteins are small intracellular proteins that negatively regulate cytokine responses by binding to receptor complexes and inhibiting the activity of the associated JAK tyrosine kinases, or by targeting signaling components for ubiquitination and proteasomal degradation.[1-3] The mammalian SOCS family consists of eight proteins, SOCS1-7 and cytokine-inducible SH2-containing protein (CIS), that share a common domain organization, with a central SH2 domain, a highly conserved C-terminal SOCS box and an N-terminal region of low sequence conservation.[4, 5] The N-terminal domains of SOCS proteins vary in sequence and length, with SOCS4-7 being distinguished from the other SOCS proteins by a long N-terminal region, which is predicted to be disordered.[6] In general, SOCS proteins are thought to bind to target proteins via their SH2 domains or N-terminal regions, leading to ubiquitination of the target protein by the SOCS box-associated E3 ubiquitin ligase complex and its subsequent proteasomal degradation.[7, 8]

While the physiological functions of SOCS1-3 and CIS are well characterised,[2, 5, 9] comparatively little is known regarding the biological roles of SOCS4 and SOCS5. Although SOCS4 and SOCS5 share high sequence homology across their SH2 domain (~90%) and SOCS box, it is unclear whether they have overlapping or distinct biological roles.[5, 8, 10, 11] However, SOCS5 can interact independently of the canonical SH2-phosphotyrosine interactions via its N-terminus, which mediates interactions with the EGF receptor[8, 10] and IL-4 receptor.[12] Recently, SOCS36E, the SOCS5 homologue in *Drosophila*, was shown to negatively regulate JAK/STAT signaling via a SOCS box-independent interaction mediated by the N-terminus.[13] Studies in *Drosophila* suggest that the EGF receptor and JAK/STAT pathways cooperate in oncogenic transformation and are both able to induce the expression of SOCS36E, which, in turn, serves to limit the activity of both pathways.[14] There is also increasing evidence that mammalian SOCS5

may function as a tumor suppressor.[15, 16] Recently, SOCS5 was identified as a target of microRNA (miR) miR-9, and down-regulation of SOCS5 expression by miR-9 resulted in aberrant activation of JAK/STAT signaling in tumor-associated endothelial cells.[17] These findings highlight the potential role of SOCS5 in the regulation of oncogenic signaling, although the molecular mechanisms by which it acts in these pathways have not been well defined. Understanding these events may provide a rationale for targeting these pathways through the development of novel cancer therapeutics that mimic or block the activity of SOCS5.

In contrast to the relative wealth of structural and functional data available for the SH2 domain and the SOCS box,[7, 18-21] little is known about the structure and function of the N-terminal region of the SOCS proteins. In most studies the N-terminal regions of SOCS proteins were removed to avoid interference with structure determination. However, we recently identified a highly conserved 70-residue region in the disordered N-terminus of SOCS4 and SOCS5 that is predicted to be more ordered than the surrounding sequence.[6] The high degree of sequence conservation of this region across species and extending back to lower vertebrates implies that it has an important functional role.[6] We also demonstrated that SOCS5 could bind all four JAK family members, although it was found to selectively inhibit only JAK1 and JAK2 autophosphorylation and this function was dependent on the SOCS5 N-terminal region.[22] The conserved N-terminal region in SOCS5 ($SOCS5_{175-244}$) facilitates a direct interaction with the kinase domain of JAK and has therefore been designated the JAK interaction region (JIR).[22] The equilibrium dissociation constant (K_D) for the interaction of SOCS5 JIR with the kinase domain of JAK1 was measured to be 0.5 μM.[22] Deletion of the JIR impaired the ability of SOCS5 to inhibit IL-4 induced STAT6 activity, suggesting that this region was functionally important.[22]

In this study, we characterize the conformation and dynamics of the JIR of mouse SOCS5 ($mSOCS5_{175-244}$) by solution NMR spectroscopy. We show that the distal C-terminal region of mSOCS5 JIR adopts a helical structure and exhibits partially restricted dynamics, while the proximal N-terminal region is disordered and flexible. We identify a phosphorylation site within the conserved JIR of mSOCS5 and use site-directed mutagenesis and surface plasmon resonance (SPR) to probe the role of phosphorylation in modulating JAK binding.

EXPERIMENTAL PROCEDURES

Protein expression and purification. The expression and purification of $mSOCS5_{175-244}$ were as described previously.[22] The protein purity was determined to be > 95% by analytical HPLC and SDS-PAGE and the molecular mass confirmed by mass spectrometry (8101 Da). [U-^{13}C, ^{15}N]-$mSOCS5_{175-244}$ was expressed in 1 L of M9 medium supplemented with 1 g/L of $^{15}NH_4Cl$ and 2 g/L of ^{13}C-glucose. Expression and purification conditions were as described for unlabeled $mSOCS5_{175-244}$.[22] The protein yields per liter of culture were generally 1–2 mg for [U-^{13}C, ^{15}N]-$mSOCS5_{175-244}$ and 3 mg for [U-^{15}N]-$mSOCS5_{175-244}$. Ser211Glu and Ser211Ala mutants of $mSOCS5_{175-244}$ were generated using site-directed mutagenesis. Expression and purification conditions for these mutants were as described for $mSOCS5_{175-244}$.[22]

Codon-optimized DNA corresponding to residues 110-313 in the N-terminus of mouse SOCS5 was custom synthesized (Genscript). EcoR1 and TEV sites were engineered upstream of $mSOCS5_{110-313}$ gene and a BamH1 site was incorporated downstream. This gene was ligated into the pGEX-2T vector (GE Healthcare) via EcoRI and BamH1 sites and transformed into *E. coli* BL21-(DE3) cells. $mSOCS5_{110-313}$ was expressed as a glutathione S-transferase (GST)-fusion protein. The cells were grown at 28 °C in 1 L of Luria-Bertani medium to an OD_{600} 0.5, cooled

to 18 °C, and protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 20 h at 18 °C. The soluble GST-fusion protein was purified using glutathione-Sepharose™ 4B (GE healthcare). One unit of TEV protease per 30 mg of fusion protein was used to cleave at 4 °C for 20 h on a rotating mixer. $mSOCS5_{110-313}$ was further purified from the cleavage mixture using cation-exchange chromatography. Recombinant JAK kinase domains were expressed in insect cells and purified as described previously.[19]

NMR spectroscopy and resonance assignments. Lyophilized [U-^{13}C, ^{15}N]-$mSOCS5_{175-244}$ was dissolved in a buffer containing 20 mM sodium citrate (pH 4.5), 5 mM TCEP, 0.02% (w/v) sodium azide (NaN_3), 94% H_2O, and 6% ^2H_2O, at a final protein concentration of 0.3 mM. NMR data were acquired on Bruker Avance spectrometers operating at ^1H frequencies of 600 and 800 MHz, both equipped with cryogenically cooled triple-resonance probes. Sequence-specific resonance assignments (H^N, N, C^α and C^β) were obtained at 30 °C using the following three-dimensional experiments, employing non-uniform sampling in the ^{15}N and ^{13}C dimensions: HNCA, CBCA(CO)NH, HNCACB. The datasets were processed with the multidimensional decomposition algorithm[23] using TOPSPIN (version 3.2) from Bruker BioSpin or NMRPipe.[24] Spectra were referenced to dioxane at 3.75 ppm. Data analysis and peak-picking were done using CARA and NMRView[25] and automated backbone assignments were made using the software MATCH in the UNIO software package.[26] The backbone assignments were then manually complemented and completed interactively using 3D ^{15}N-resolved [^1H,^1H]-NOESY, and further extended to the side-chains using 3D HBHA(CO)NH, 3D ^{13}C(aliphatic)-resolved, and 3D ^{13}C(aromatic)-resolved [^1H,^1H]-NOESY experiments. NOESY spectra were recorded with a mixing time of 225 ms at 800 MHz. C^α, C^β, H^α, and N chemical shifts were used in the program TALOS [27] to obtain backbone torsion angles.

Structure Calculation. The software UNIO-ATNOS/CANDID [28, 29] was used in combination with the torsion angle dynamics algorithm CYANA3.0 [30] in a standard seven-cycle protocol for NOESY peak picking, NOE assignment and structure calculation.[31] The intensities from the NOEs were converted into distance restraints and combined with the backbone dihedral angles derived from TALOS,[27] of which 18 φ and ψ dihedral angle constraints predicted with high confidence were used to supplement the NOE distance restraints in the final input for structure calculations. The 40 conformers with the lowest CYANA target function values obtained from the final UNIO-ATNOS/CANDID/CYANA3.0 cycle were subjected to restrained simulated annealing and energy minimization in a water shell with the program OPALp.[32] Finally, a family of 20 conformers of lowest target function values was chosen for further analysis based on stereochemistry and energy considerations using the programs PROCHECK[33] and MOLMOL[34] as previously described.[31] The structures were displayed and analyzed using PyMOL and MOLMOL.[34]

^{15}N Relaxation Studies. Backbone $\{^1H\}$-^{15}N heteronuclear NOE and ^{15}N longitudinal (T_1) and transverse (T_2) relaxation experiments were performed using standard methods[35] at 30 °C on Bruker 600 MHz and 800 MHz spectrometers, respectively. T_1 values were obtained from a series of ^{1}H-^{15}N correlation spectra with 0, 100, 400, 500, 800, 1000, 1500, and 2500 ms relaxation delays. T_2 values were acquired with 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, and 210 ms delays. Recycle delay was set to 4 s for both T_1 and T_2 measurements. The relaxation rates were calculated in SPARKY by least-squares fitting of peak intensities *versus* relaxation delay times to an equation for single-order exponential decay. Steady-state $\{^1H\}$-^{15}N-NOE values were determined from the ratio of peak intensities for spectra collected with and without 5 s proton pre-saturation.

Surface Plasmon Resonance (SPR). Surface plasmon resonance measurements were performed on a BIAcore T200 instrument at 25 °C as described previously.[22] Recombinant $SOCS5_{110-313}$ was diluted to 10 μg/mL in 10 mM sodium acetate, pH 5.0, and immobilized to a CM5 Biosensor chip (GE Healthcare) using amine coupling. A reference flow cell was prepared by a similar procedure, albeit in the absence of protein. Recombinant JAK1 and JAK2 kinase domains were injected over the immobilized surface at a rate of 30 μL/min for 2 min in running buffer (10 mM HEPES, pH 8.0, 200 mM NaCl, 3.0 mM EDTA, 0.05% v/v Tween20) at concentrations ranging from 31.25 nM to 1 μM. Regeneration of the surface was achieved with 20 mM NaOH injected at a rate of 20 μL/min for 20s. The ligand flow cell responses were double reference subtracted for all analyses. The binding responses at the steady-state region of the sensorgrams were plotted as a function of the analyte concentration to obtain the binding curves. These curves were fitted to a steady-state model to derive the apparent equilibrium dissociation constant (K_D). Similarly, the binding affinities of $mSOCS5_{175-244}$ and Ser211 mutants for JAK1 kinase domain were analyzed by SPR by immobilizing JAK1 kinase domain to a CM5 chip using amine coupling.

Cell transfections, lysis and tryptic digest. Flag-tagged SOCS5 was transiently transfected into 293T cells using FuGENE6 (Promega), and cells were lysed 48 h later for 30 min in 1% NP-4 buffer (1% v/v NP-4 , 50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM NaF, 1 mM Na_3VO_4). Lysates were pre-cleared with protein-A-Sepharose prior to immunoprecipitation of Flag-tagged SOCS5 with anti-Flag M2 affinity gel (Sigma). Proteins were eluted from the M2 Sepharose with 0.16% phosphoric acid (pH 1.8) on ice. Proteins were prepared for mass spectrometry analysis using the FASP protein digestion kit (Protein Discovery, Knoxville, TN) as described previously[36] with the following modifications. Eluted material was neutralized with

urea/Tris-HCl, pH 8.5 and reduced with TCEP (5 mM final concentration). 2 μg of sequence-grade modified trypsin (Promega) was added in 50 mM NH_4HCO_3 and incubated overnight at 37 °C. Peptides were eluted by two rounds of washing with 40 μL of 50 mM NH_4HCO_3.

Phosphopeptide enrichment. Tryptic peptides were subjected to phosphopeptide enrichment using the immobilized metal ion affinity chromatography (IMAC) Phosphopeptide Isolation Kit (gallium/IDA, Pierce) as per manufacturer's instructions with the following described modifications: an equal amount of binding buffer (5% acetic acid) and peptide sample was added to the resin and incubated at room temperature for 30 min. Samples were centrifuged at 1000 x *g* for 1 min and subsequently washed with 50 μL of 0.1% acetic acid followed by 50 μL of 0.1% acetic acid, 20% acetonitrile, pH 3.0. Phosphopeptides were eluted with 30 μL of 0.4 M NH_4HCO_3, 20% acetonitrile, pH 8.7 by centrifugation after incubating at room temperature for 7 min. This step was repeated twice. Phosphopeptides were acidified with formic acid (1% final concentration), subjected to vacuum centrifugation and resuspended in 25 μL 1% formic acid/2% acetonitrile.

Mass spectrometry. Tryptic peptides were separated by nanoflow reversed-phase LC on a Waters nanoAcquity C18 150 mm x 75 μm I.D. column with a 60 min gradient set at a flow rate of 400 nL/min from 100% buffer A (0.1% formic acid, 3% acetonitrile, 97% water) to 55% B (80% acetonitrile, 0.1% formic acid, 20% water). The nano HPLC was coupled on-line to a Q-Exactive mass spectrometer equipped with a nanoelectrospray ion source (Thermo Fisher Scientific™) for automated MS/MS. The Q-Exactive was run in a data-dependent acquisition mode with resolution set at 70K and the 10 most abundant precursor ions were dynamically chosen from the survey scan (350–1850 Th) for HCD fragmentation with the resolution set at 17.5K. Dynamic exclusion was enabled for 90 s.

Database search and identification of SOCS5 post-translational modifications. Mass spectra were processed using the MaxQuant software (version 1.4.1.2) [37] and searched against a concatenated database consisting of the Ludwig Institute non-redundant database (NR, September 2013) supplemented with common contaminants (including keratins, trypsin, BSA) and the reversed-sequence version of the same database.[38] The search parameters were: minimum peptide length 7, peptide tolerance 4.5 ppm, mass tolerance 0.5 Da, cleavage enzyme trypsin/P, and a total of 3 missed cleavages were allowed. Carbamidomethyl (C) was set as a fixed modification and oxidation (M), acetylation (Protein N-term), pyro-Glu/Gln (N-term) Phospho (ST), and Phospho (Y) were set as variable modifications. The peptide and protein false discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 0.01. The ProteinProspector MS-Product program (http://prospector.ucsf.edu/) was used to calculate the theoretical masses of fragments of identified peptides for manual validation including a phosphopeptide corresponding to Ser211 of mouse SOCS5.

RESULTS

N-terminus of SOCS5. The N-terminal region of SOCS5 accounts for a large proportion of the protein and is 368 amino acid residues in length (Figure 1A). This region is enriched in polar and charged amino acids and relatively depleted in hydrophobic amino acids, a feature typical of intrinsically disordered proteins (IDPs).[39] The disorder prediction for SOCS5 N-terminus was made using the metaPrDOS web server, which integrates the results of eight different disorder prediction methods.[40] The prediction output is consistent with the SOCS5 N-terminus being largely unstructured, with the exception of a few short regions with ordered structure, one of

which encompasses the conserved JIR (Figure S1 of the Supporting Information). Prediction of protein binding regions within the N-terminus of SOCS5 using the web server ANCHOR[41] identified several segments that have the propensity to undergo disorder-to-order transition on binding a globular partner, including residues 183–192 and 210–220 in the JIR (Figure S1 of the Supporting Information).

NMR assignments of mSOCS5 $_{175\text{-}244}$. Given that the JIR of SOCS5 may contain some ordered structure, we expressed this region of mouse SOCS5 (residues 175–244) in *E.coli* in order to characterize its solution structure by NMR spectroscopy. The ^{1}H-^{15}N HSQC spectrum of mSOCS5$_{175\text{-}244}$ is characteristic of a largely disordered protein, with the amide proton shifts confined to 6.9–8.9 ppm (Figure 1B). While IDPs usually have uniformly sharp resonances, the mSOCS5$_{175\text{-}244}$ spectra displayed a diversity of peak intensities and peak shapes, indicating conformational heterogeneity of the molecular ensemble. Unambiguous sequence-specific backbone assignments were obtained for all residues at 30 °C. Side chain assignments were obtained for most residues using 3D HBHA(CO)NH and 3D NOESY experiments. A representative strip plot from the 3D ^{15}N- resolved [^{1}H, ^{1}H] NOESY spectrum for residues Leu215-Phe220 is shown in Figure S2A of the Supporting Information. The $^{13}C^{\beta}$ chemical shifts of the two cysteine residues, Cys188 and Cys218, at 28.2 and 28.0 ppm respectively, suggested that these residues existed in the reduced state. All of the assigned proline residues were found to be in the *trans* conformation based on the presence of strong NOEs between H^{δ} and the preceding H^{α} (Figure S2B of the Supporting Information). A few minor peaks with intensities < 10% of the major peaks were observed in the ^{1}H-^{15}N HSQC, which might arise from minor conformers of the protein as a result of *cis–trans* isomerization of the prolines, but none of these resonances were assigned owing to their low intensity and the absence of cross peaks in the 3D

spectra. The NMR assignments for mSOCS5$_{175-244}$ have been deposited in the BioMagResBank database (accession no.19966).

The JIR of SOCS5 contains secondary structural elements. The resonance assignments of mSOCS5$_{175-244}$ were used to investigate the presence of secondary structure by evaluating the chemical shift deviations from random coil.[42, 43] The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts of mSOCS5$_{175-244}$ (Figure 2A and B) show that the C-terminal region of the polypeptide, encompassing residues Asp225–Ile233, is helical, while most of the residues in the N-terminal region show essentially random coil $^{13}C^{\alpha}$ chemical shifts. The mean helix percentage for the residues Asp225–Ile233, calculated on the basis of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts is 37%.[44] Secondary $^{1}H^{\alpha}$ chemical shifts (Figure 2C) are consistent with helical structure in the C-terminal segment Asp225–Ile233. A chemical shift deviation of ~ -2 ppm is observed for H^{α} of Cys218, which could be attributed to the ring current effects induced by the indole ring of Trp230 and the aromatic ring of Phe220. $^{13}C^{\alpha}$ chemical shift deviations ranging from 0.2 to 1.4 ppm were observed in the region Leu210–Leu213 (Figure 2A), indicating the presence of transient structure in this region.

Backbone Dynamics of mSOCS5$_{175-244}$. Backbone $\{^{1}H\}$-^{15}N-NOEs and ^{15}N -T_1 and T_2 relaxation times for mSOCS5$_{175-244}$ are shown in Figure 2 (D-F). The observation of positive heteronuclear NOE values ranging from 0.2 to 0.6 for residues Ile208–His236 is consistent with limited mobility and restrained conformation in this region (Figure 2F). Other segments, including residues Ser176–Ser203 in the N-terminal region and Ala238–Ser244 at the C-terminus, show negative heteronuclear NOE values, consistent with a high degree of flexibility and disorder (Figure 2F). Consistent with the heteronuclear NOE values, the T_2 values for residues His209–His236 are significantly lower than those for the rest of mSOCS5$_{175-244}$ (Figure

2E). It is evident from the average heteronuclear NOE, T_1 and T_2 values (Table S1 of the Supporting Information) that mSOCS5$_{175\text{-}244}$ is largely disordered. Nevertheless, the most structured region of mSOCS5$_{175\text{-}244}$ encompassing residues Cys218–Ile233 exhibits restricted backbone mobility, with heteronuclear NOE values well above the average (Table S1 of the Supporting Information).

Solution structure of mSOCS5$_{175\text{-}244}$. The C-terminal region of mSOCS5$_{175\text{-}244}$ displayed several NOEs indicative of helix and turn-like structures (Figure S3 of the Supporting Information). Residues Ser224–Ile233 show $d_{\alpha N}(i,i+3)$,$d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ NOEs typical for regular α-helices, indicating the presence of a fully formed α-helix in this region.[45] Furthermore, weak $d_{\alpha N}(i,i+2)$ NOEs indicative of turn-like structures were observed for Ile208–Ser211 and Pro221–Gly223, although the local structure could not be classified as any specific type of turn.[45, 46] A continuous stretch of sequential NOEs supports an extended structure in the region Cys218–Phe220 preceding the turn at Pro221–Gly223. Residues Arg175–Lys207 and Lys234–Ser244 were disordered as assessed by both NOE and chemical shift analysis. Chemical shift deviations from random coil values of the C^{α} and H^{α} resonances, as well as the backbone relaxation parameters (Figure 2), are in agreement with the NOE data in identifying the most prominent secondary structure element in mSOCS5$_{175\text{-}244}$ as an α-helix located at the C-terminus in the region Ser224–Ile233. Although there are a few long-range NOE connectivities among Cys218, Pro219, Phe220, Leu226, Ala227 and Trp230, mSOCS5$_{175\text{-}244}$ lacks a stable tertiary structure. This is also supported by hydrogen-deuterium exchange NMR experiments, in which all backbone amide resonances underwent rapid exchange (data not shown), indicating high solvent accessibility.

Table 1 summarizes the long-range, medium-range, sequential, and intra-residual NOEs from the final cycle of UNIO-ATNOS/CANDID and CYANA 3.0.[30] The solution structure of mSOCS5$_{175-244}$ is best described as a conformational ensemble of inter-converting states. The global structure of the polypeptide was not well defined, with the average root-mean-square deviation (RMSD) over the backbone heavy atoms (N, C^{α}, C′) of the entire peptide being 11.2 Å. However, ordered local structure was apparent in the C-terminal region of the molecule (Figure S4 of the Supporting Information). The ensemble of 20 energy-minimized conformers that represent the most-populated state of mSOCS5$_{175-244}$ is shown in Figure 3A. The ensemble of 20 structures showed excellent agreement with the experimental data, with RMSD values to the mean coordinates of 0.49 Å for the backbone and 1.12 Å for all heavy atoms in the structured region of residues Ser224–Ile233 (Figure 3B). Angular order parameters (S) were used to determine the precision of dihedral angles within the ensemble; the φ and ψ angles were generally well ordered ($S \geq 0.8$) for residues Ile208–Ser211 and Ser224–Ile233 (Figure S5 of the Supporting Information). The final conformers had no distance violations > 0.2 Å or dihedral violations > 5 °. The NMR ensemble for mSOCS5$_{175-244}$ has been deposited in the Protein Data Bank (PDB ID 2N34).

The transient local structure in mSOCS5$_{175-244}$ could be stabilized by hydrophobic interactions involving Cys218, Pro219, Phe220, Leu226 and Trp230 (Figure 3C). The exceptional upfield shift of the Phe220 backbone amide resonance observed in the ^{1}H-^{15}N HSQC NMR spectrum could be due to the flanking proline residues, Pro219 and Pro221. As a consequence of the spatial proximity of Cys218, Pro219 and L226 to the aromatic rings of Phe220 and Trp230 (Figure 3C), some of their side chain resonances are perturbed by ring current effects. The differences between the observed and calculated average ring current shifts (Table S2 of the

Supporting Information) for these resonances arise from the variability in the calculated shifts across the ensemble owing to sparse NOE restraints and less well-defined side chains of Phe220 and Trp230 (Figure S4D of the Supporting Information).

Interaction of SOCS5 N-terminal region with the kinase domain of JAK. In an earlier study, we showed that $mSOCS5_{175-244}$ binds specifically to the kinase domain of the JAKs with an equilibrium dissociation constant of 0.5 μM.[22] Although $mSOCS5_{175-244}$ was sufficient for binding to the JAK kinase domain, additional residues of the N-terminus were required for inhibition of JAK activity. Deletion analysis of the SOCS5 N-terminus identified a larger segment (residues 110-313) as being critical for inhibition of JAK activity.[22] To determine whether this larger segment made additional contacts with the JAK kinase domain we generated recombinant protein corresponding to $mSOCS5_{110-313}$ and measured its binding affinity for JAK1 and JAK2 kinase domains by SPR. $mSOCS5_{110-313}$ bound the JAK1 and JAK2 kinase domains with equilibrium dissociation constants of 0.23 and 1.5 μM, respectively (Figure 4), demonstrating that high affinity binding of $mSOCS5_{175-244}$ to the JAK kinase domain is recapitulated by $mSOCS5_{110-313}$ and further validating $mSOCS5_{175-244}$ as the primary JAK interaction region in the SOCS5 N-terminus.

Identification of a phosphorylation site in SOCS5 JIR and its role in JAK binding. Post-translational modifications such as phosphorylation often regulate interactions between proteins and are known to modulate the function of intrinsically disordered proteins.[47, 48] Using mass spectrometry, we identified Ser211 within the JAK interaction region (JIR) of SOCS5, as a candidate phosphorylation site (Figure 5A). 293T cells were transiently transfected with full-length, Flag-tagged SOCS5 and the protein was enriched from cell lysates, digested with trypsin and analysed for the presence of phosphorylation modifications by mass spectrometry. A high

confidence peptide corresponding to phosphorylated Ser211 was identified with a MaxQuant score of 143.25. Ser211 is located within a turn-like structure adjacent to the structured region (Figure 5B) in the JIR of SOCS5 and had the potential to impact on JAK binding to this region.

In order to investigate the effect of Ser211 phosphorylation on the structure and function of $mSOCS5_{175-244,}$ Ser211 was mutated either to the phosphomimetic glutamic acid or to alanine. These mutations had no deleterious effects on either the protein expression or yield. The effect of these mutations on the structure of $mSOCS5_{175-244}$ was analyzed by NMR spectroscopy (Figure S6 of the Supporting Information). Chemical shift changes in the 1D NMR spectra of the Ser211Glu and Ser211Ala mutants appear to be limited to the mutated residue and its adjacent residues. No significant change in chemical shift dispersion was observed in the 1D NMR spectra of the mutants, suggesting that phosphorylation of Ser211 does not induce folding or disrupt the structured region of $mSOCS5_{175-244}$.

Surface plasmon resonance was used to probe the role of Ser211 phosphorylation on JAK binding (Figure 6). The phosphomimetic mutant (Ser211Glu) of $mSOCS5_{175-244}$ bound the JAK1 kinase domain with an equilibrium dissociation constant of 0.3 μM (Figure 6B). The kinetics of Ser211Glu mutant binding to JAK1 show fast association and slow dissociation phases similar to that of the wild-type. The kinetic data could not be fit to a 1:1 binding model, thus precluding accurate quantification of the association and dissociation rates. Replacement of Ser211 with alanine resulted in a slight decrease in affinity for the JAK1 kinase domain ($K_D \sim 0.85$ μM) (Figure 6C). Thus, our SPR data show that $mSOCS5_{175-244}$ and its phosphomimetic Ser211Glu mutant bind JAK1 with similar affinities, suggesting that phosphorylation of Ser211 only had a moderate effect on modulating JAK1 binding. It is also possible that phosphorylation of Ser211

modulates other interactions of the JIR or plays a more important role in the context of the full-length protein.

DISCUSSION

Intrinsically disordered regions of proteins play specific roles in protein-protein recognition, [49, 50] but their modes of interaction with partner proteins are distinct from those used by their folded counterparts. IDPs utilize a multitude of binding processes, including folding-coupled to-binding, conformational selection and 'fly casting', and the increased flexibility in IDPs is inherent to these processes.[51-53] The mechanism of interaction often depends on the structure of the IDP in its unbound form.[54] Pre-formed structural elements in IDPs can play a crucial role in molecular recognition, so their identification and characterization is an area of active investigation.[54, 55]

In this study, we define the solution structure of the conserved JAK interaction region of mouse SOCS5 ($mSOCS5_{175-244}$) and characterize its binding to the kinase domain of JAK1. This is the first structural study on the apparently disordered N-terminal region of SOCS proteins and provides insights into the structure and dynamics of the JAK interaction region, which is unique to SOCS4 and SOCS5. NMR studies identified pre-formed structural elements in $mSOCS5_{175-244}$, including a helical segment spanning residues Ser224–Ile233. NMR relaxation measurements support the presence of a transient population of secondary structural elements with restricted mobility for residues Ile208–His236. The presence of such motionally restricted structured regions in the SOCS5 N-terminus may provide a thermodynamic advantage by reducing the entropic penalty associated with coupled folding and binding.[56] Pre-formed structural elements, such as those present in $mSOCS5_{175-244}$, are likely to mediate interactions that play important

roles in regulation of signaling pathways.[54] The plasticity of the disordered SOCS5 N-terminus would enable it to function as a scaffold that can bind multiple components of the signalling pathway. The pre-formed hairpin structure in the SOCS5 JIR is likely to contribute to the scaffolding functionality of the N-terminus by spatially constraining the flanking disordered regions and the proteins bound to them, thus bringing together various signalling components in close proximity to JAK, facilitating efficient interactions.

Phosphorylation is often overrepresented in disordered regions of proteins and is implicated in modulating the activity of these regions.[47, 48, 57] The identification of Ser211 as a candidate phosphorylation site on SOCS5 adds weight to the argument that the JIR is important for SOCS5 function, particularly given its conservation across species for SOCS5.[6] As the primary JAK1 binding site on SOCS5 includes Ser211 and the surrounding residues, it is conceivable that phosphorylation of this residue in cells would modulate JAK1 binding. However, our SPR data on the phosphomimetic Ser211Glu mutant of SOCS5 showed only a moderate increase in affinity for the JAK1 kinase domain. It is possible that the Ser-Glu mutation has a weaker effect than phosphorylation or that the phosphate group is crucial for the functional effects of phosphorylation at Ser211. It appears that the interaction of SOCS5 with the kinase domain of JAK1 is not critically regulated by a Ser211 phosphorylation switch; rather phosphorylation of this residue may facilitate interaction with additional proteins or facilitate binding to another region of JAK outside its kinase domain.

In conclusion, we have confirmed the presence of pre-formed structural elements in the largely disordered N-terminus of SOCS5. We propose that the pre-formed helix in the SOCS5 JIR has mostly a structural role and serves as a “staple” that connects and spatially restricts the flanking flexible ligand binding regions in the unstructured N-terminus of SOCS5, thus favoring

the binding process. We predict that the other SOCS family members with extended and disordered N-terminal regions, such as SOCS4, may exhibit similar modes of interaction. Further experiments will be required to validate this hypothesis and define the binding mechanism. Given the role of SOCS5 in the inhibition of JAK1 activity and tumor suppression, molecular insights into the specific interactions that mediate inhibition of JAK are expected to lay the groundwork for the design of novel JAK inhibitors.

Acknowledgements

We thank Dr. Ann Kwan at the NMR facility, University of Sydney for her assistance in acquiring the NMR spectra.

Supporting information

Disorder prediction for the N-terminus of mouse SOCS5 made using metaPrDOS (Figure S1), Representative strip plots from 3D ^{15}N-resolved and ^{13}C (aliphatic)-resolved [^{1}H,^{1}H]-NOESY spectra of mSOCS5$_{175\text{-}244}$ (Figure S2), Schematic of characteristic NOEs identified for mSOCS5$_{175\text{-}244}$ (Figure S3), Structured regions of mSOCS5$_{175\text{-}244}$ (Figure S4), Angular order parameters of the backbone dihedral angles of mSOCS5$_{175\text{-}244}$ (Figure S5), Comparison of amide region of 1D ^{1}H NMR spectra of mSOCS5$_{175\text{-}244,}$ Ser211Ala and Ser211Glu mutants (Figure S6), Average ^{15}N relaxation parameters for mSOCS5 JIR (Table S1), Comparison of the observed shift differences and the calculated ring current shifts for side-chain protons of Cys218, Pro219 and Leu226 (Table S2). This material is available free of charge via the Internet at http:// pubs.acs.org.

References

1. Starr, R., Willson, T. A., Viney, E. M., Murray, L. J., Rayner, J. R., Jenkins, B. J., Gonda, T. J., Alexander, W. S., Metcalf, D., Nicola, N. A., and Hilton, D. J. (1997) A family of cytokine-inducible inhibitors of signalling. *Nature 387*, 917-921.
2. Alexander, W. S., and Hilton, D. J. (2004) The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response. *Annu Rev Immunol 22*, 503-529.
3. Endo, T. A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor, N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997) A new protein containing an SH2 domain that inhibits JAK kinases. *Nature 387*, 921-924.
4. Hilton, D. J., Richardson, R. T., Alexander, W. S., Viney, E. M., Willson, T. A., Sprigg, N. S., Starr, R., Nicholson, S. E., Metcalf, D., and Nicola, N. A. (1998) Twenty proteins containing a C-terminal SOCS box form five structural classes. *Proc Natl Acad Sci U S A 95*, 114-119.
5. Linossi, E. M., Babon, J. J., Hilton, D. J., and Nicholson, S. E. (2013) Suppression of cytokine signaling: the SOCS perspective. *Cytokine Growth Factor Rev 24*, 241-248.
6. Feng, Z. P., Chandrashekaran, I. R., Low, A., Speed, T. P., Nicholson, S. E., and Norton, R. S. (2012) The N-terminal domains of SOCS proteins: a conserved region in the disordered N-termini of SOCS4 and 5. *Proteins 80*, 946-957.
7. Bullock, A. N., Rodriguez, M. C., Debreczeni, J. E., Songyang, Z., and Knapp, S. (2007) Structure of the SOCS4-ElonginB/C complex reveals a distinct SOCS box interface and the molecular basis for SOCS-dependent EGFR degradation. *Structure 15*, 1493-1504.
8. Nicholson, S. E., Metcalf, D., Sprigg, N. S., Columbus, R., Walker, F., Silva, A., Cary, D., Willson, T. A., Zhang, J. G., Hilton, D. J., Alexander, W. S., and Nicola, N. A. (2005) Suppressor of cytokine signaling (SOCS)-5 is a potential negative regulator of epidermal growth factor signaling. *Proc Natl Acad Sci U S A 102*, 2328-2333.
9. Alexander, W. S. (2002) Suppressors of cytokine signalling (SOCS) in the immune system. *Nat Rev Immunol 2*, 410-416.
10. Kario, E., Marmor, M. D., Adamsky, K., Citri, A., Amit, I., Amariglio, N., Rechavi, G., and Yarden, Y. (2005) Suppressors of cytokine signaling 4 and 5 regulate epidermal growth factor receptor signaling. *J Biol Chem 280*, 7038-7048.
11. Sutherland, J. M., Keightley, R. A., Nixon, B., Roman, S. D., Robker, R. L., Russell, D. L., and McLaughlin, E. A. (2012) Suppressor of cytokine signaling 4 (SOCS4): moderator of ovarian primordial follicle activation. *J Cell Physiol 227*, 1188-1198.
12. Seki, Y., Hayashi, K., Matsumoto, A., Seki, N., Tsukada, J., Ransom, J., Naka, T., Kishimoto, T., Yoshimura, A., and Kubo, M. (2002) Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proc Natl Acad Sci U S A 99*, 13003-13008.
13. Stec, W., Vidal, O., and Zeidler, M. P. (2013) *Drosophila* SOCS36E negatively regulates JAK/STAT pathway signaling via two separable mechanisms. *Mol Biol Cell 24*, 3000-3009.
14. Herranz, H., Hong, X., Hung, N. T., Voorhoeve, P. M., and Cohen, S. M. (2012) Oncogenic cooperation between SOCS family proteins and EGFR identified using a *Drosophila* epithelial transformation model. *Genes Dev 26*, 1602-1611.

15. Francipane, M. G., Eterno, V., Spina, V., Bini, M., Scerrino, G., Buscemi, G., Gulotta, G., Todaro, M., Dieli, F., De Maria, R., and Stassi, G. (2009) Suppressor of cytokine signaling 3 sensitizes anaplastic thyroid cancer to standard chemotherapy. *Cancer Res 69*, 6141-6148.
16. Calvisi, D. F., Ladu, S., Gorden, A., Farina, M., Lee, J. S., Conner, E. A., Schroeder, I., Factor, V. M., and Thorgeirsson, S. S. (2007) Mechanistic and prognostic significance of aberrant methylation in the molecular pathogenesis of human hepatocellular carcinoma. *J Clin Invest 117*, 2713-2722.
17. Zhuang, G., Wu, X., Jiang, Z., Kasman, I., Yao, J., Guan, Y., Oeh, J., Modrusan, Z., Bais, C., Sampath, D., and Ferrara, N. (2012) Tumour-secreted miR-9 promotes endothelial cell migration and angiogenesis by activating the JAK-STAT pathway. *EMBO J 31*, 3513-3523.
18. Babon, J. J., McManus, E. J., Yao, S., DeSouza, D. P., Mielke, L. A., Sprigg, N. S., Willson, T. A., Hilton, D. J., Nicola, N. A., Baca, M., Nicholson, S. E., and Norton, R. S. (2006) The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. *Mol Cell 22*, 205-216.
19. Babon, J. J., Kershaw, N. J., Murphy, J. M., Varghese, L. N., Laktyushin, A., Young, S. N., Lucet, I. S., Norton, R. S., and Nicola, N. A. (2012) Suppression of cytokine signaling by SOCS3: characterization of the mode of inhibition and the basis of its specificity. *Immunity 36*, 239-250.
20. Babon, J. J., Sabo, J. K., Zhang, J. G., Nicola, N. A., and Norton, R. S. (2009) The SOCS box encodes a hierarchy of affinities for Cullin5: implications for ubiquitin ligase formation and cytokine signalling suppression. *J Mol Biol 387*, 162-174.
21. Kershaw, N. J., Murphy, J. M., Liau, N. P., Varghese, L. N., Laktyushin, A., Whitlock, E. L., Lucet, I. S., Nicola, N. A., and Babon, J. J. (2013) SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. *Nat Struct Mol Biol 20*, 469-476.
22. Linossi, E. M., Chandrashekaran, I. R., Kolesnik, T. B., Murphy, J. M., Webb, A. I., Willson, T. A., Kedzierski, L., Bullock, A. N., Babon, J. J., Norton, R. S., Nicola, N. A., and Nicholson, S. E. (2013) Suppressor of Cytokine Signaling (SOCS) 5 utilises distinct domains for regulation of JAK1 and interaction with the adaptor protein Shc-1. *PLoS One 8*, e70536.
23. Orekhov, V. Y., and Jaravine, V. A. (2011) Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog Nucl Magn Reson Spectrosc 59*, 271-292.
24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR 6*, 277-293.
25. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol 278*, 313-352.
26. Volk, J., Herrmann, T., and Wüthrich, K. (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR 41*, 127-138.
27. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR 13*, 289-302.

28. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol 319*, 209-227.
29. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR 24*, 171-189.
30. Güntert, P. (2003) Automated NMR protein structure calculation. *Prog Nucl Magn Reson Spectrosc 43*, 105-125.
31. Serrano, P., Pedrini, B., Mohanty, B., Geralt, M., Herrmann, T., and Wüthrich, K. (2012) The J-UNIO protocol for automated protein structure determination by NMR in solution. *J Biomol NMR 53*, 341-354.
32. Luginbuhl, P., Güntert, P., Billeter, M., and Wüthrich, K. (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J Biomol NMR 8*, 136-146.
33. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR 8*, 477-486.
34. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph 14*, 51-55, 29-32.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984-6003.
36. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Methods 6*, 359-362.
37. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol 26*, 1367-1372.
38. Moritz, R. L., Ji, H., Schutz, F., Connolly, L. M., Kapp, E. A., Speed, T. P., and Simpson, R. J. (2004) A proteome strategy for fractionating proteins and peptides using continuous free-flow electrophoresis coupled off-line to reversed-phase high-performance liquid chromatography. *Anal Chem 76*, 4811-4824.
39. Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, 89-100.
40. Ishida, T., and Kinoshita, K. (2008) Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics 24*, 1344-1348.
41. Dosztanyi, Z., Meszaros, B., and Simon, I. (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics 25*, 2745-2746.
42. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*, 1647-1651.
43. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR 5*, 67-81.

44. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci 15*, 2795-2804.
45. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley, New York.
46. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) Turns in peptides and proteins. *Adv Protein Chem 37*, 1-109.
47. Wright, P. E., and Dyson, H. J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol 16*, 18-29.
48. Bah, A., Vernon, R. M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L. E., and Forman-Kay, J. D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature 519*, 106-109.
49. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol 293*, 321-331.
50. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit 18*, 343-384.
51. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol 12*, 54-60.
52. Fuxreiter, M., and Tompa, P. (2012) Fuzzy complexes: a more stochastic view of protein function. *Adv Exp Med Biol 725*, 1-14.
53. Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr Opin Struct Biol 19*, 31-38.
54. Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol 338*, 1015-1026.
55. Lee, S. H., Kim, D. H., Han, J. J., Cha, E. J., Lim, J. E., Cho, Y. J., Lee, C., and Han, K. H. (2012) Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. *Curr Protein Pept Sci 13*, 34-54.
56. Sivakolundu, S. G., Bashford, D., and Kriwacki, R. W. (2005) Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J Mol Biol 353*, 1118-1128.
57. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res 32*, 1037-1049.

Table 1. Structure Statistics for mSOCS5$_{175-244}$

Number of Distance Constraints	450
Intra-residue ($i = j$)	189
Sequential ($\|i - j\| = 1$)	198
Medium-range ($1<\|i - j\|<5$)	45
Long-range ($\|i - j\| \geq 5$)	18
Dihedral constraints (Φ, Ψ)	18
Final CYANA target function value (Å^2)	0.70 $\pm$ 0.24
AMBER energies (kcal/mol)	
Total energy	-1835 $\pm$ 115.5
Vander Waals energy	-42 $\pm$ 13.2
Electrostatic energy	-2426 $\pm$109.3
Deviations from ideal geometry[a]	
Bond (Å)	0.007 $\pm$ 0.0002
Angles (°)	1.88 $\pm$ 0.061
Average RMSD to mean coordinates (Å)	
Residues 218-233	
Backbone heavy atoms (N, C$^\alpha$, C′)	0.83 $\pm$ 0.27
All heavy atoms	1.46 $\pm$ 0.28
Residues 224-233	
Backbone heavy atoms (N, C$^\alpha$, C′)	0.49 $\pm$ 0.17
All heavy atoms	1.12 $\pm$ 0.23
Ramachandran plot[b]	
Residues 175-244	
Most favored (%)	50.5
Additionally allowed (%)	41.6
Generously allowed (%)	4.6
Disallowed (%)	3.3
Residues 224-233	
Most favored (%)	100
Additionally allowed (%)	0
Generously allowed (%)	0
Disallowed (%)	0

[a]The values for the bond and angle show the deviations from ideal values based on perfect stereochemistry.
[b]As determined by PROCHECK-NMR for all residues except Gly and Pro.

Figure legends

Figure 1. NMR assignments of mSOCS5$_{175-244}$. (A) Schematic showing domain organization in SOCS5. (B) ^{1}H-^{15}N HSQC NMR spectrum of mSOCS5$_{175-244}$ with backbone resonance assignments labelled. The spectrum was recorded in 20 mM citrate buffer, pH 4.5, containing 5 mM TCEP at 30 °C on a 600 MHz spectrometer.

Figure 2. Secondary structure and backbone amide relaxation parameters measured for mSOCS5$_{175-244}$. Chemical shift deviations of (A) C^{α} (B) C^{α}-C^{β} (average over three residues) and (C) H^{α} chemical shifts of mSOCS5$_{175-244}$ from random-coil values are plotted vs residue number. (D-F) ^{15}N *T*1, *T*2 relaxation parameters were measured at 800 MHz and steady-state {^{1}H}-^{15}N-NOE was measured at 600 MHz at 30 °C. Missing histogram bars correspond to overlapped residues and prolines.

Figure 3. Solution structure of mSOCS5$_{175-244}$. (A) Ensemble of 20 conformers of mSOCS5$_{175-244}$ with backbone heavy atoms superimposed over residues 218–233 (RMSD 0.83 Å) and (B) with all heavy atoms superimposed over residues 224–233 (RMSD 1.12 Å). (C) Closest-to-average conformer of mSOCS5$_{175-244}$ over residues 218–233. The side chain protons of Cys218, Pro219 and Leu226 that display ring current shifts due to their proximity to the aromatic rings of Trp230 and Phe220 are highlighted in magenta. The poorly defined disordered regions of the polypeptide are not shown in the figure.

Figure 4. mSOCS5$_{110-313}$ binds JAK kinase domain with high affinity. SPR analysis of mSOCS5$_{110-313}$ binding to (A) JAK1 and (B) JAK2 JH1 domain. Serially diluted JAK JH1 domains (31.25 nM – 1 μM) were flowed over immobilised mSOCS5$_{110-313}$. Binding affinities were determined using steady-state analysis.

Figure 5. SOCS5 JIR contains a serine phosphorylation site. (A) MS/MS spectrum identifying SOCS5 phosphorylation site at Ser211. The spectrum displayed prominent product ions at 944.5 m/z and 831.4 m/z that corresponded to loss of H_3PO_4 from the y8 and y7 ions, respectively. Furthermore, the fragmentation pattern shows long uninterrupted b- and y-ion series that define the phosphopeptide sequence and site of phosphorylation. (B) Surface representation of the closest-to-average NMR conformer of mSOCS5$_{175-244}$ with the phosphorylation site (Ser211) highlighted in purple. The surface is shown at 60 % transparency and the regions exhibiting restricted backbone mobility is shaded blue.

Figure 6. Effect of Ser211 phosphorylation on JAK1 binding. SPR analysis of (A) mSOCS5$_{175-244}$ (B) Ser211Glu mutant and (C) Ser211Ala mutant binding to JAK1 kinase domain. Serially diluted mSOCS5$_{175-244}$ and mutants (31.2 – 500 nM) were flowed over immobilised JAK1 JH1 domains. Left panels represent sensorgrams showing the kinetics of binding and right panels show steady-state analysis. K_D values are representative of two experiments.

Figure 1.

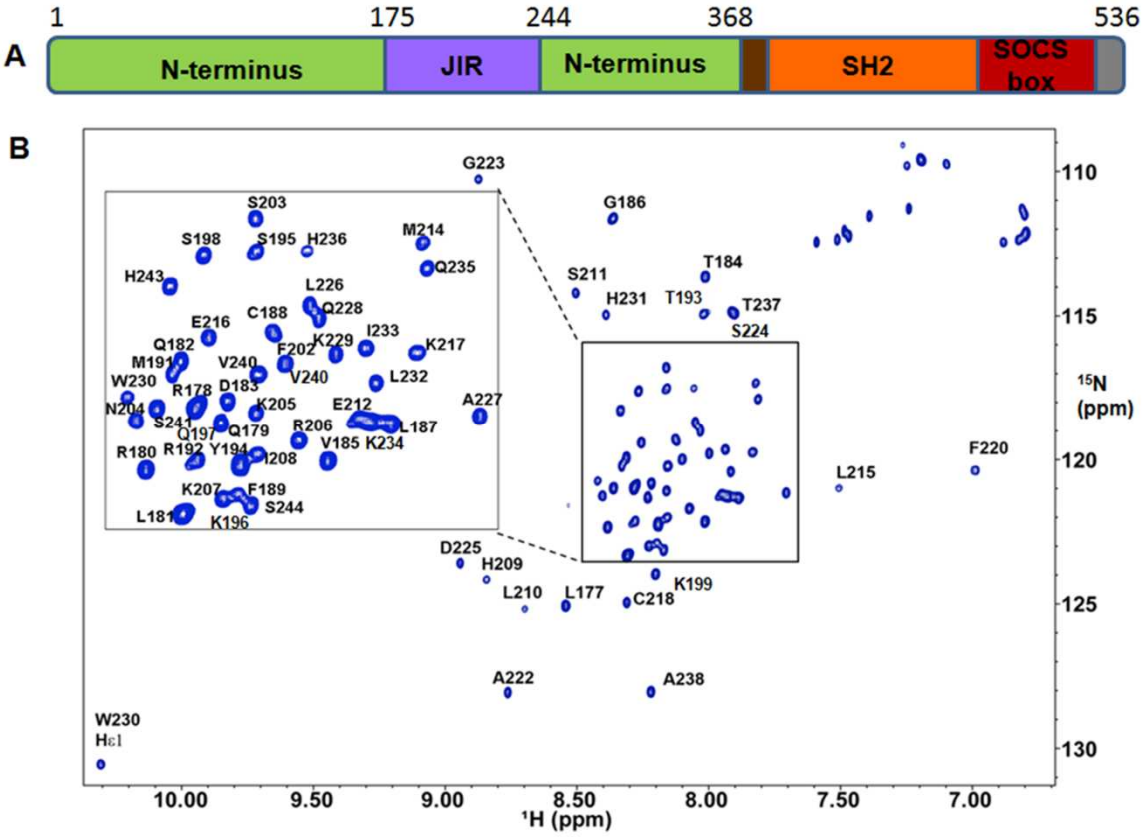

A
1
175
244
368
536
N-terminus
JIR
N-terminus
SH2
SOCS box
B
G223
S203
G186
S198
S195
H236
M214
Q235
S211
T184
H243
L226
H231
T193
T237
Q228
S224
E216
C188
I233
Q182
K229
K217
F202
M191
V240
V240
L232
W230
R178
D183
A227
N204
K205
E212
S241
Q197
Q179
R206
L187
V185
K234
R180
R192
Y194
I208
F220
L215
K207
F189
S244
L181
K196
D225
H209
K199
L210
L177
C218
A222
A238
W230
Hε1
110
115
120
125
130
15N (ppm)
10.00
9.50
9.00
8.50
8.00
7.50
7.00
1H (ppm)

Figure 2.

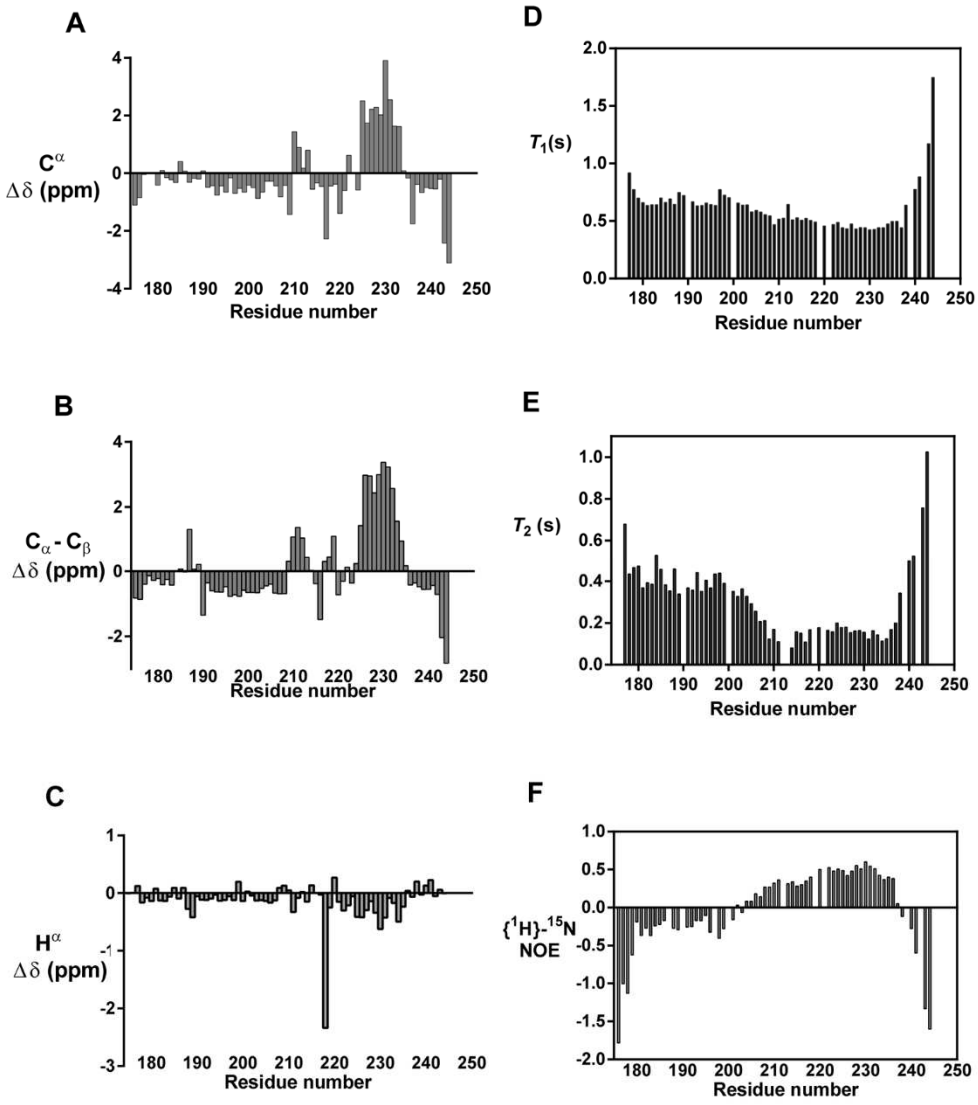

A
C^{α} $\Delta\delta$ (ppm)
Residue number
D
T_1(s)
Residue number
B
C_{α} - C_{β} $\Delta\delta$ (ppm)
Residue number
E
T_2 (s)
Residue number
C
H^{α} $\Delta\delta$ (ppm)
Residue number
F
{^{1}H}-^{15}N NOE
Residue number

Figure 3.

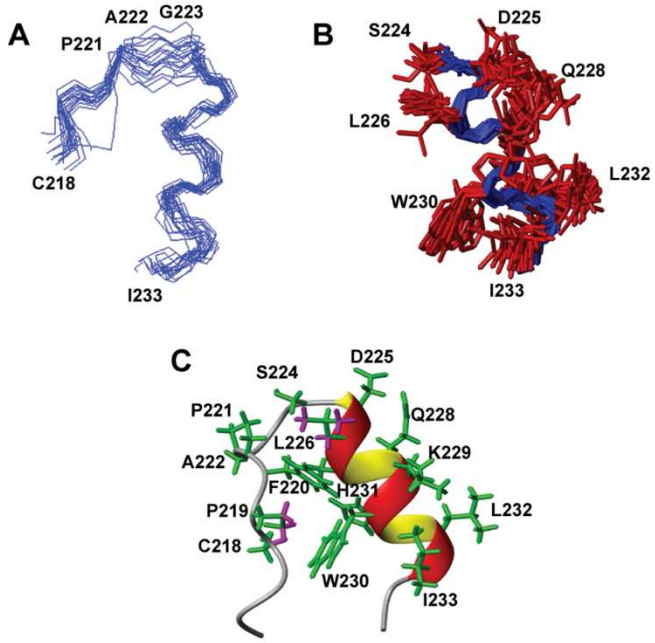


Figure 4.

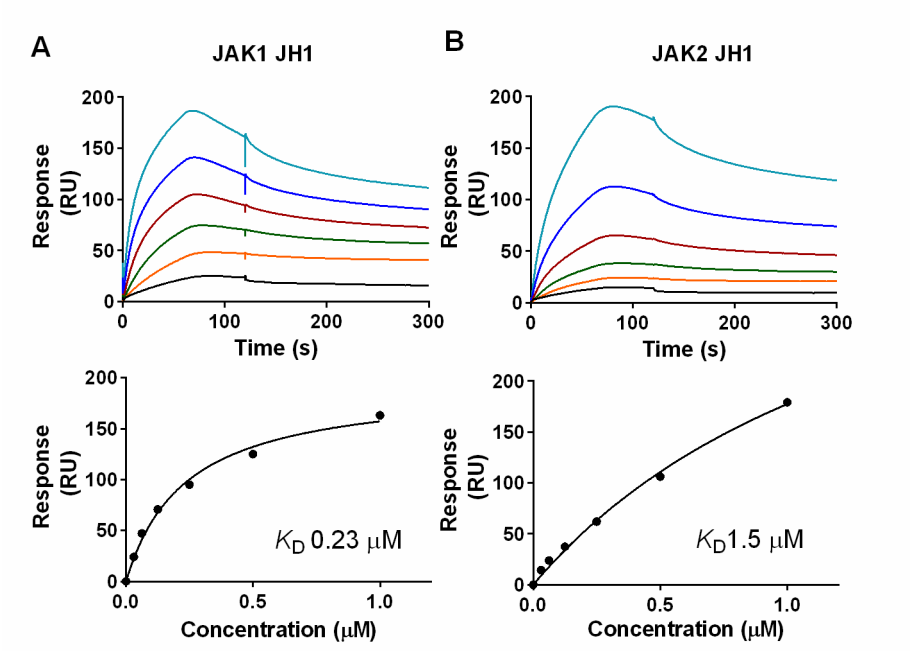


Figure 5.

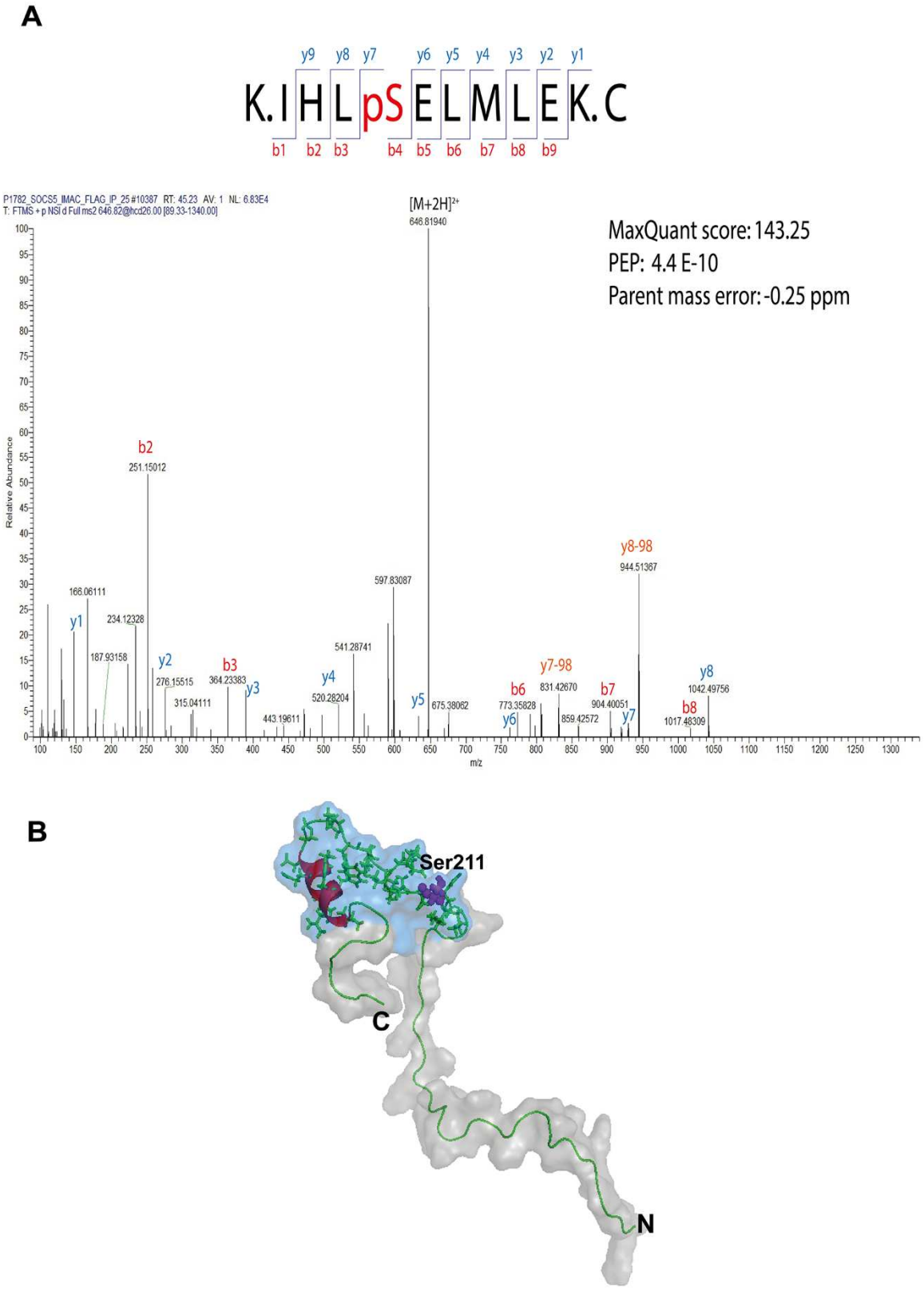
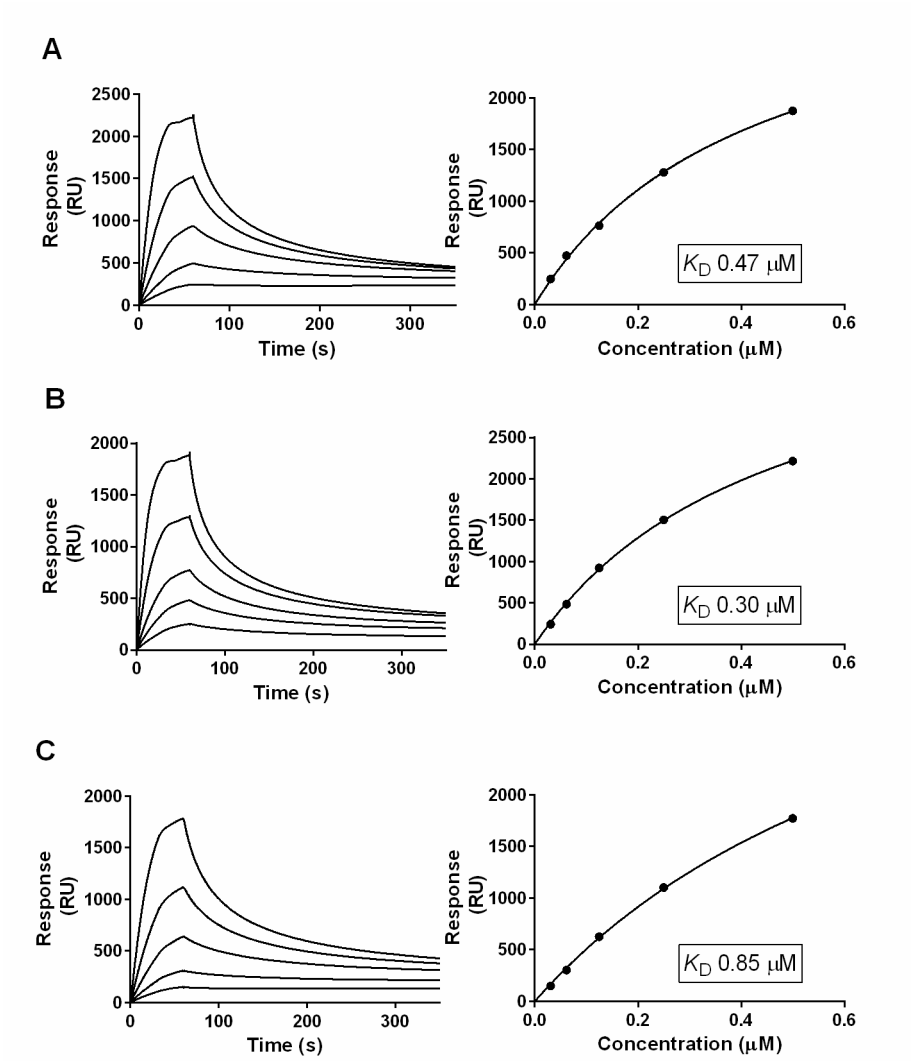
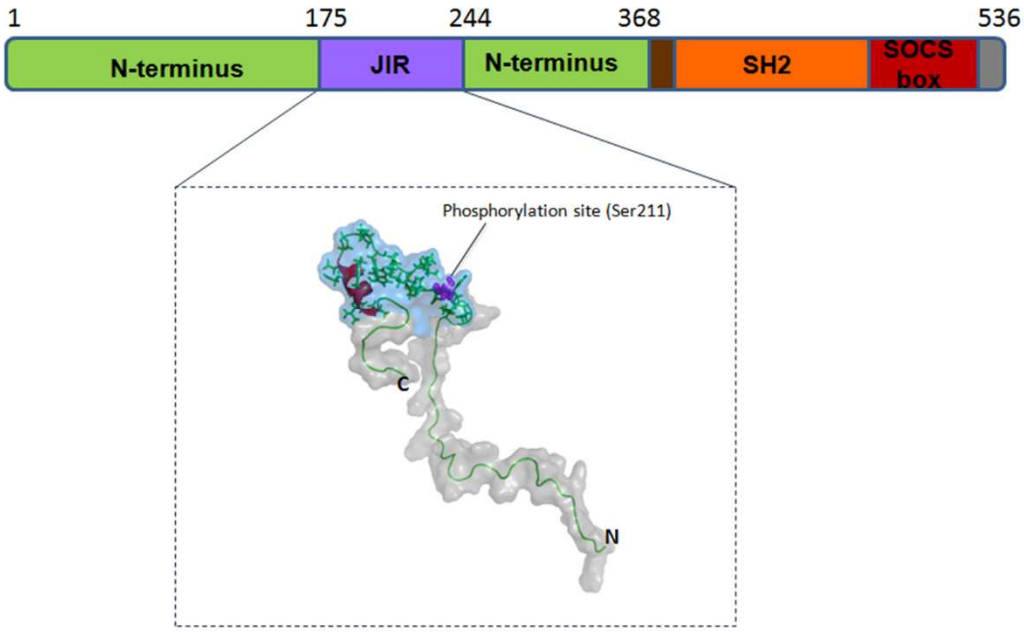

A
y9 y8 y7 y6 y5 y4 y3 y2 y1
K.IHLpSELMLEK.C
b1 b2 b3 b4 b5 b6 b7 b8 b9
P1782_SOCS5_IMAC_FLAG_IP_25 #10387 RT: 45.23 AV: 1 NL: 6.83E4
T: FTMS + p NSI d Full ms2 646.82@hcd26.00 [89.33-1340.00]
[M+2H]2+
646.81940
MaxQuant score: 143.25
PEP: 4.4 E-10
Parent mass error: -0.25 ppm
Relative Abundance
m/z
b2 251.15012
y8-98 944.51367
597.83087
166.06111
y1
234.12328
187.93158
y2
276.15515
b3
364.23383
y3
315.04111
541.28741
y4
520.28204
443.19611
y5
675.38062
y7-98
831.42670
b6
773.35828
y6
859.42572
b7
904.40051
y7
y8
1042.49756
b8
1017.48309
B
Ser211
C
N

Figure 6.


A
Response (RU)
Time (s)
Concentration (μM)
K_D 0.47 μM
B
Response (RU)
Time (s)
Concentration (μM)
K_D 0.30 μM
C
Response (RU)
Time (s)
Concentration (μM)
K_D 0.85 μM

Graphic for the Table of Contents



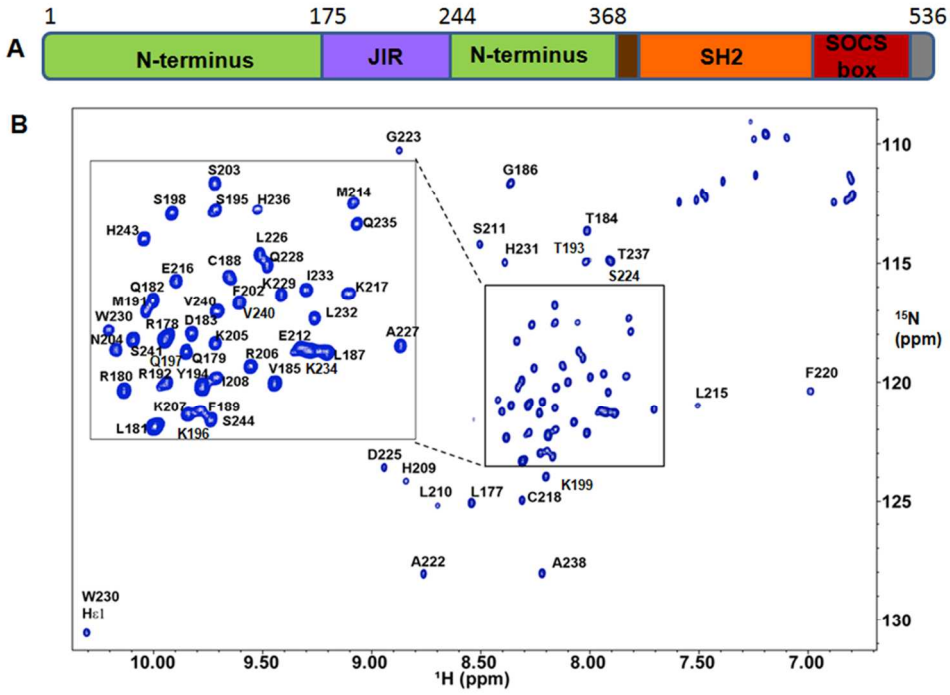


Figure 1. NMR assignments of mSOCS5$_{175-244}$. (A) Schematic showing domain organization in SOCS5. (B) ^{1}H-^{15}N HSQC NMR spectrum of mSOCS5$_{175-244}$ with backbone resonance assignments labelled. The spectrum was recorded in 20 mM citrate buffer, pH 4.5, containing 5 mM TCEP at 30 °C on a 600 MHz spectrometer. 160x120mm (300 x 300 DPI)

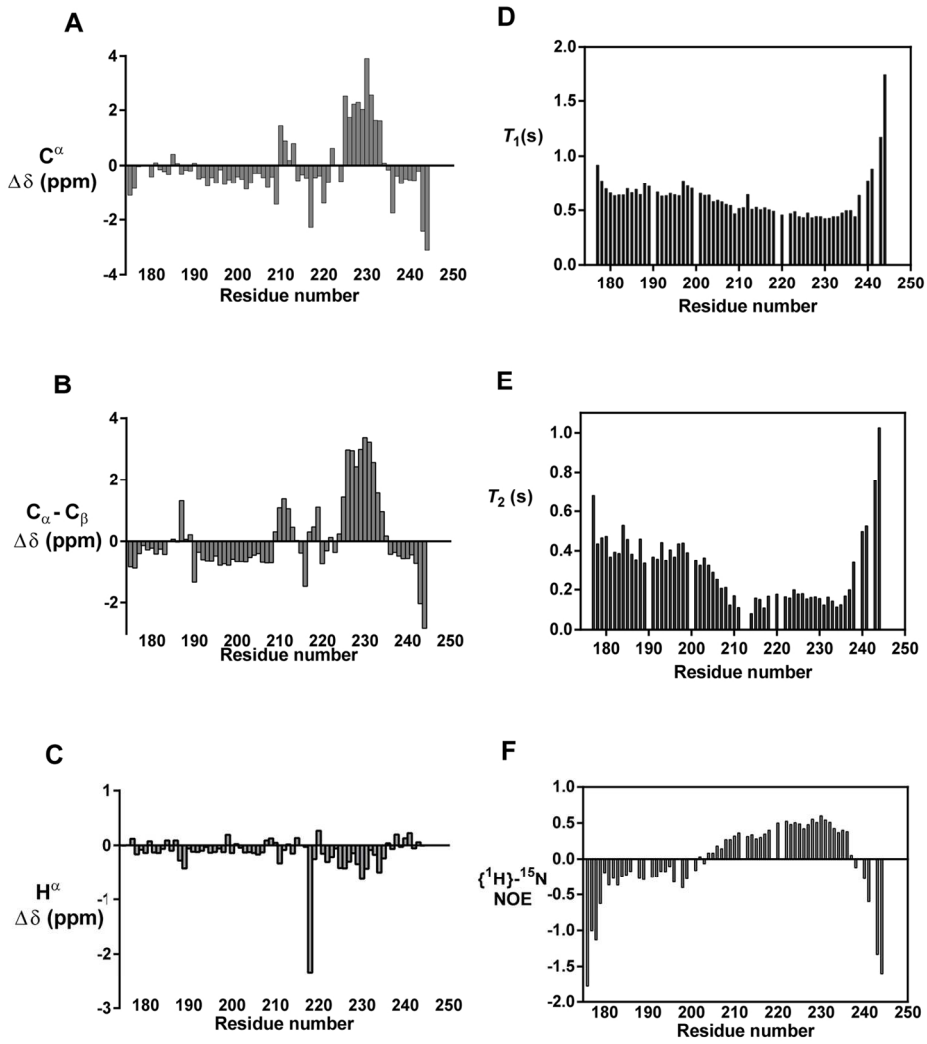


Figure 2. Secondary structure and backbone amide relaxation parameters measured for mSOCS5$_{175-244}$. Chemical shift deviations of (A) C^{α} (B) C^{α}-C^{β} (average over three residues) and (C) H^{α} chemical shifts of mSOCS5$_{175-244}$ from random-coil values are plotted vs residue number. (D-F) ^{15}N *T*1, *T*2 relaxation parameters were measured at 800 MHz and steady-state {^{1}H}-^{15}N-NOE was measured at 600 MHz at 30 °C. Missing histogram bars correspond to overlapped residues and prolines.
140x152mm (300 x 300 DPI)

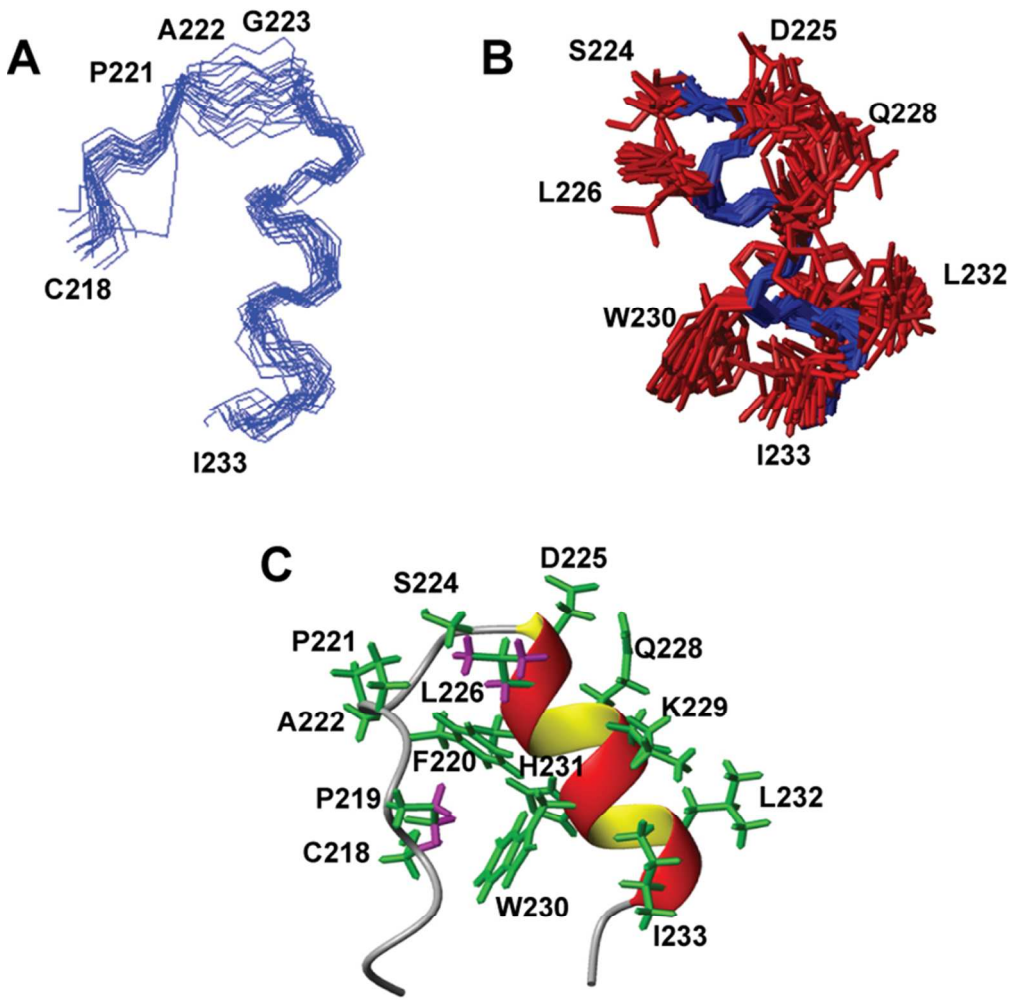
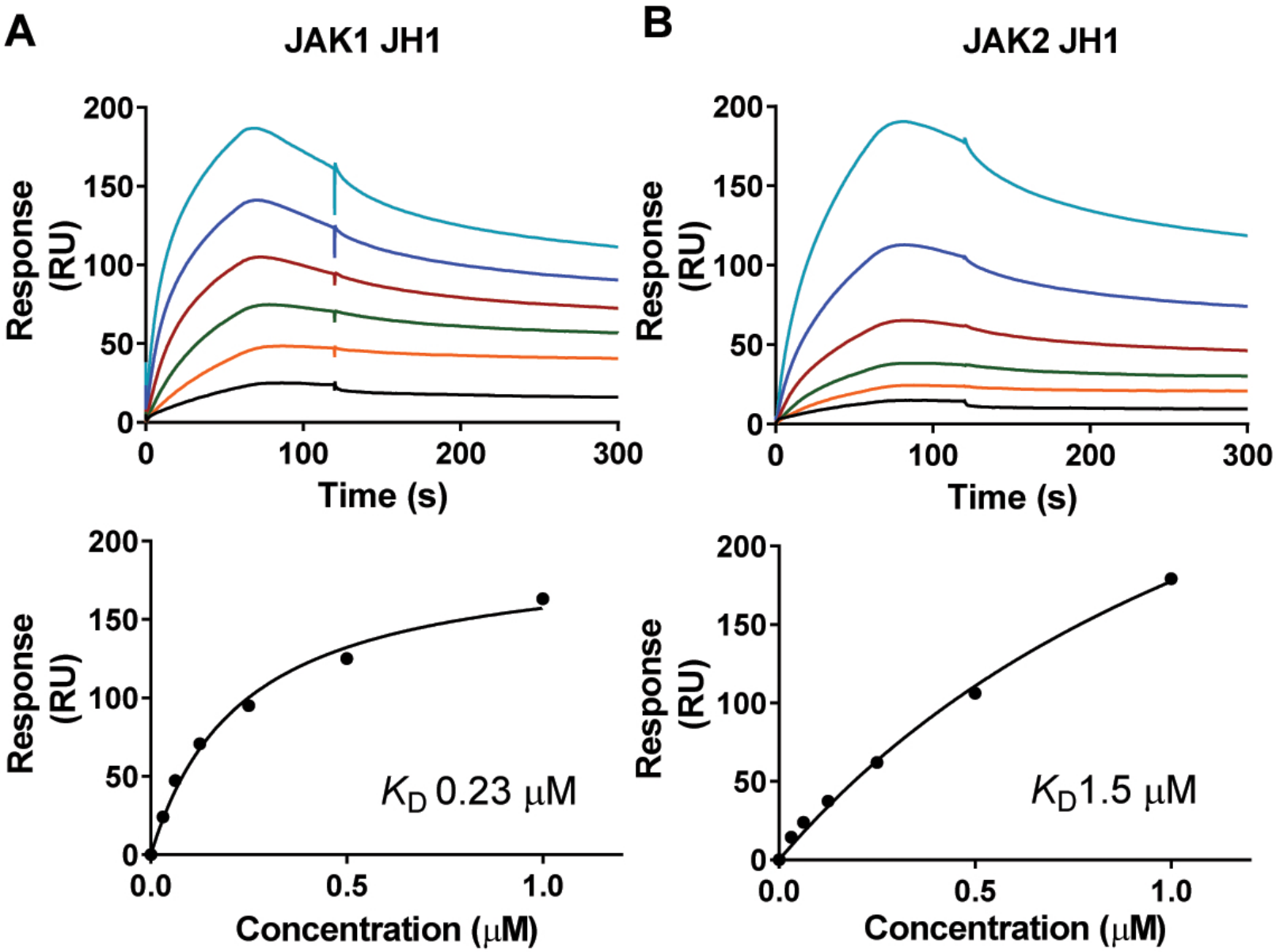


Figure 3. Solution structure of mSOCS5$_{175-244}$. (A) Ensemble of 20 conformers of mSOCS5$_{175-244}$ with backbone heavy atoms superimposed over residues 218–233 (RMSD 0.83 Å) and (B) with all heavy atoms superimposed over residues 224–233 (RMSD 1.12 Å). (C) Closest-to-average conformer of mSOCS5$_{175-244}$ over residues 218–233. The side chain protons of Cys218, Pro219 and Leu226 that display ring current shifts due to their proximity to the aromatic rings of Trp230 and Phe220 are highlighted in magenta. The poorly defined disordered regions of the polypeptide are not shown in the figure.
85x94mm (300 x 300 DPI)


A
JAK1 JH1
Response (RU)
200
150
100
50
0
0
100
200
300
Time (s)
Response (RU)
200
150
100
50
0
0.0
0.5
1.0
Concentration (μM)
K_D 0.23 μM
B
JAK2 JH1
Response (RU)
200
150
100
50
0
0
100
200
300
Time (s)
Response (RU)
200
150
100
50
0
0.0
0.5
1.0
Concentration (μM)
K_D1.5 μM

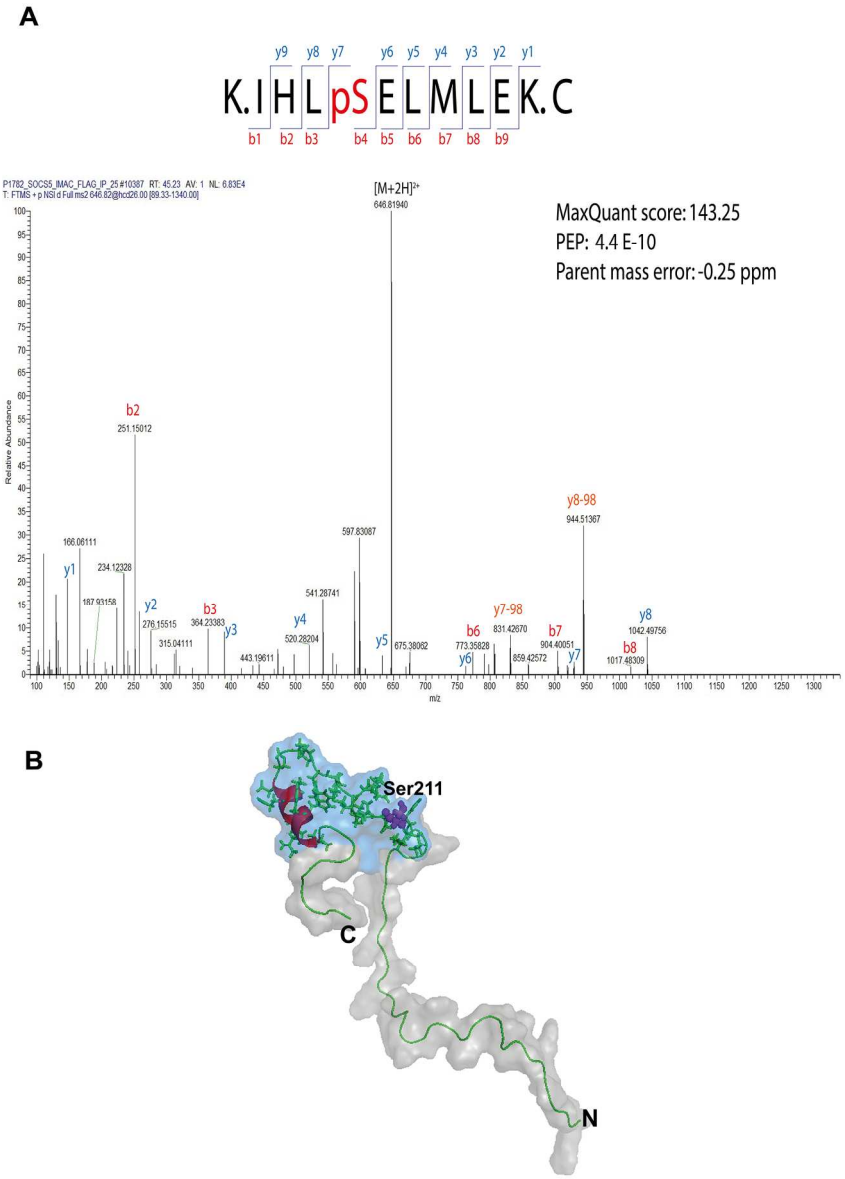
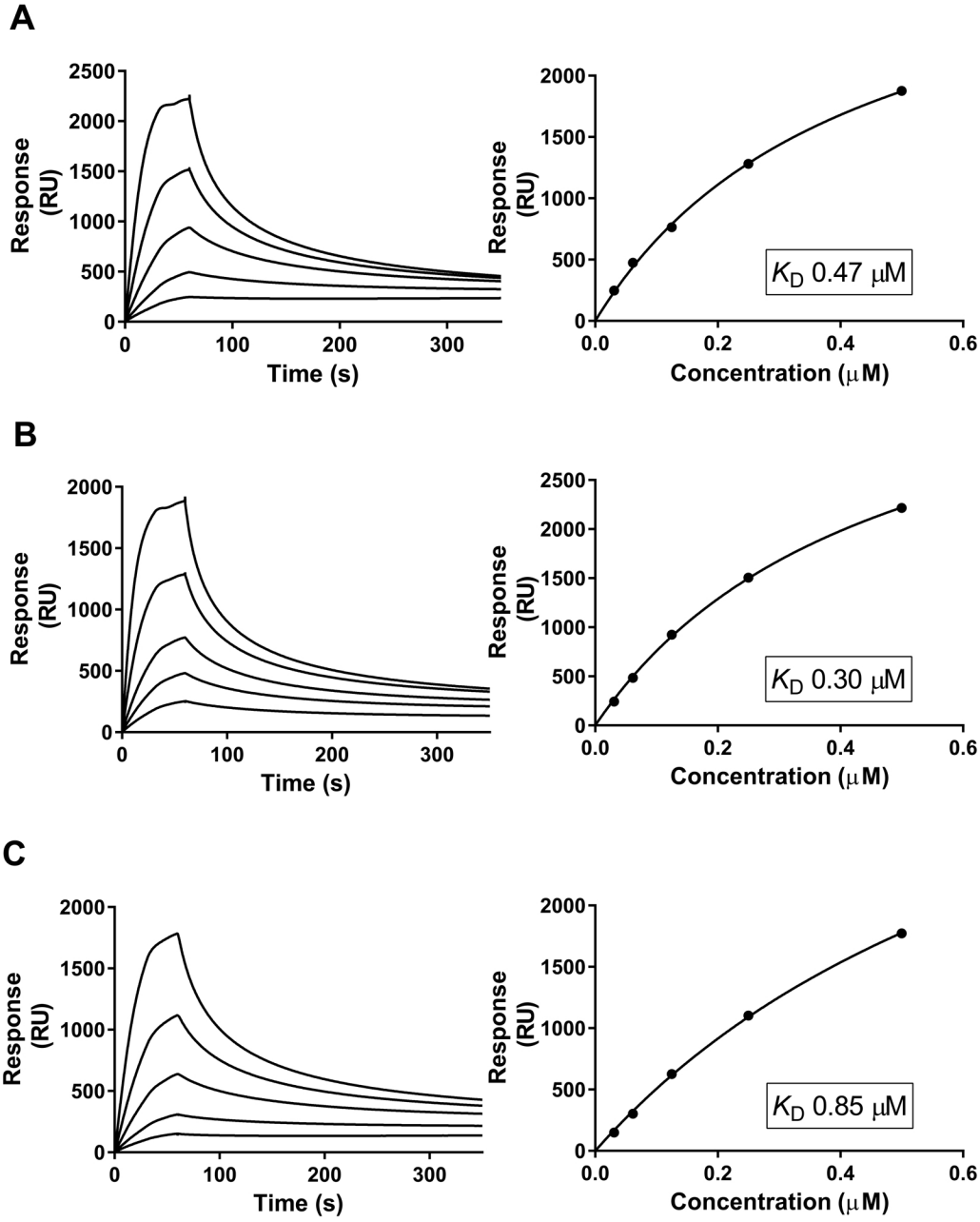
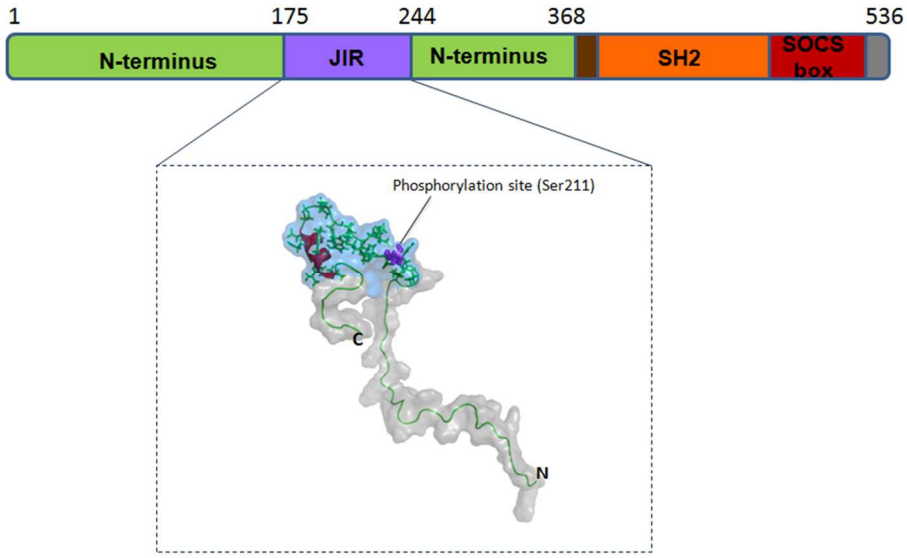


Figure 5. SOCS5 JIR contains a serine phosphorylation site. (A) MS/MS spectrum identifying SOCS5 phosphorylation site at Ser211. The spectrum displayed prominent product ions at 944.5 m/z and 831.4 m/z that corresponded to loss of H_3PO_4 from the y8 and y7 ions, respectively. Furthermore, the fragmentation pattern shows long uninterrupted b- and y-ion series that define the phosphopeptide sequence and site of phosphorylation. (B) Surface representation of the closest-to-average NMR conformer of $mSOCS5_{175-244}$ with the phosphorylation site (Ser211) highlighted in purple. The surface is shown at 60 % transparency and the regions exhibiting restricted backbone mobility is shaded blue.
160x226mm (300 x 300 DPI)


A
Response (RU)
Time (s)
Response (RU)
Concentration (μM)
K_D 0.47 μM
B
Response (RU)
Time (s)
Response (RU)
Concentration (μM)
K_D 0.30 μM
C
Response (RU)
Time (s)
Response (RU)
Concentration (μM)
K_D 0.85 μM


1
175
244
368
536
N-terminus
JIR
N-terminus
SH2
SOCS box
Phosphorylation site (Ser211)
C
N