Biochem. J. (2014) **462**, 53–65 (Printed in Great Britain) doi:10.1042/BJ20140521

Structural analysis of poly-SUMO chain recognition by the RNF4-SIMs domain

Camy C.-H. KUNG*†‡, Mandar T. NAIK*[1], Szu-Huan WANG§, Hsiu-Ming SHIH*, Che-Chang CHANG‖, Li-Ying LIN¶, Chia-Lin CHEN†¶, Che MA¶, Chi-Fon CHANG¶ and Tai-Huang HUANG*¶**[1]

*Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan
†Chemical Biology and Molecular Biophysics, Taiwan International Graduate Program, Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan
‡Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan
§Scientific Instrument Center, Academia Sinica, Taipei 11529, Taiwan
‖Graduate Institute of Translational Medicine, Taipei Medical University, Taipei 110, Taiwan
¶Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan
**Department of Physics, National Taiwan Normal University, Taipei 11677, Taiwan

The E3 ubiquitin ligase RNF4 (RING finger protein 4) contains four tandem SIM [SUMO (small ubiquitin-like modifier)-interaction motif] repeats for selective interaction with poly-SUMO-modified proteins, which it targets for degradation. We employed a multi-faceted approach to characterize the structure of the RNF4-SIMs domain and the tetra-SUMO2 chain to elucidate the interaction between them. In solution, the SIM domain was intrinsically disordered and the linkers of the tetra-SUMO2 were highly flexible. Individual SIMs of the RNF4-SIMs domains bind to SUMO2 in the groove between the β2-strand and the α1-helix parallel to the β2-strand. SIM2 and SIM3 bound to SUMO with a high affinity and together constituted the recognition module necessary for SUMO binding. SIM4 alone bound to SUMO with low affinity; however, its contribution to tetra-SUMO2 binding avidity is comparable with that of SIM3 when in the RNF4-SIMs domain. The SAXS data of the tetra-SUMO2–RNF4-SIMs domain complex indicate that it exists as an ordered structure. The HADDOCK model showed that the tandem RNF4-SIMs domain bound antiparallel to the tetra-SUMO2 chain orientation and wrapped around the SUMO protamers in a superhelical turn without imposing steric hindrance on either molecule.

Key words: nuclear magnetic resonance (NMR), RING finger 4 (RNF4), small-angle X-ray scattering (SAXS), small ubiquitin-like modifier (SUMO), SUMO-interaction motif (SIM).

INTRODUCTION

SUMOylation, enzyme-mediated post-translational modification by a SUMO (small ubiquitin-like modifier), is a crucial signalling event that regulates diverse cellular processes [1,2]. At least four paralogues of SUMO exist in vertebrates. Among these, the mature forms of SUMO2 and SUMO3 are highly conserved and are collectively referred to as SUMO2/3. SUMO2/3 shares approximately 50 % of its sequence identity with SUMO1 and contains a consensus SUMO modification motif, V^{10}KTE, in its N-terminal region. The motif allows these paralogues to act as a Ubc9 (ubiquitin carrier protein 9) substrate in the SUMOylation cascade, leading to the formation of polymeric SUMO chains [3]. Thus the SUMOylation machinery can mark target proteins by using three different processes: single SUMO modification, polymeric SUMO chain modification and multiple SUMO modifications at various sites. Each of these modifications can regulate cellular pathways differently by enhancing or blocking downstream interactions. Poly-SUMO chain formation plays a crucial role in chromosome segregation, DNA-damage response and meiosis [4,5]. Recognition of SUMOylated proteins is mostly mediated through SIMs (SUMO-interaction motifs) present on effectors. SIMs bind to the groove between β2-strands and α1-helices of a SUMO [6–9]. This interaction can be regulated by post-translational modifications, such as phosphorylation of the SIM [7] and acetylation of the SUMO [10].

The discovery of STUbLs (SUMO-targeted ubiquitin ligases) directly links the SUMOylation process to ubiquitination pathways. By means of tandem SIMs, STUbLs recognize poly-SUMOylated proteins and target them for Lys^{48}-linked polyubiquitylation and degradation through their E3 ubiquitin ligase activities. Only a few STUbLs have been identified thus far, including Slx5–Slx8 (where Slx is synthetic lethal of unknown function) in *Saccharomyces cerevisiae*, Rfp1 (RING finger protein 1)/Rfp2–Slx8 in *Schizosaccharomyces pombe*, human RNF111/Arkadia [11] and RING finger 4 (RNF4) in mammalian cells [12]. RNF4 is a dimeric STUbL. Its N-terminal half contains four tandem SIM repeats (SIM1–SIM4), referred to collectively in the present study as the SIMs domain, that recognize poly-SUMOylated substrates. The RING domain at the C-terminal half acts together with the SIMs domain to facilitate ubiquitination of proteins modified with poly-SUMO chains [12,13]. Bruderer et al. [14] identified more than 900 putative endogenous poly-SUMOylated proteins by using the RNF4-SIMs domain as bait. A computational string search identified additional novel SIM clusters in many other proteins [15]. More recently, the human RNF111/Arkadia was identified as a new STUbL, using three adjacent SIMs for specific recognition of poly-SUMO2/3 chains

Abbreviations: CSP, chemical shift perturbation; GST, glutathione transferase; HSQC, heteronuclear single-quantum coherence; ITC, isothermal titration calorimetry; MTSL, *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate; PRE, paramagnetic relaxation enhancement; RNF/Rfp, RING finger protein; SIM, SUMO-interaction motif; Slx, synthetic lethal of unknown function; STUbL, SUMO-targeted ubiquitin ligase; SUMO, small ubiquitin-like modifier; TEV, tobacco etch virus; Ubc, ubiquitin carrier protein; UIM, ubiquitin-interacting motif.

[1] Correspondence may be addressed to either of these authors (email mandarn@ibms.sinica.edu.tw or bmthh@ibms.sinica.edu.tw).

Co-ordinates for the tetra-SUMO2 structure have been deposited in the PDB under accession code 4NPN.

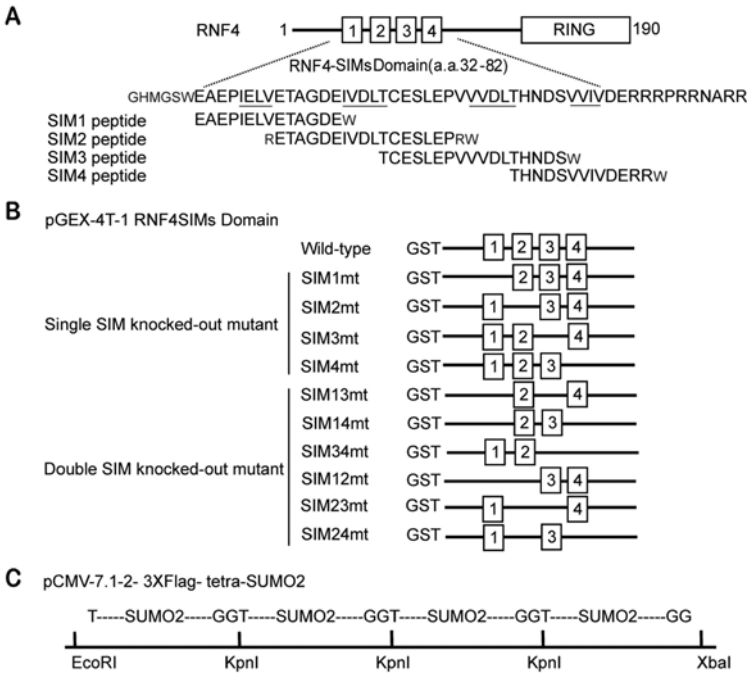

Figure 1 Constructs used in the present study

(**A**) Organization of the RNF4-SIMs domain in the pET-30a system and primary sequences of peptides used. Small capital letters represent the amino acids that are not part of the original RNF4-SIMs domain. (**B**) Wild-type and mutant constructs of RNF4-SIMs domain in pGEX-4T-1 employed for GST pull-down experiments. (**C**) The restricted cutting sites used to clone four repeats of SUMO2 into p3XFLAG-CMV-7.12 vector. By using KpnI as the restricting enzyme, each SUMO2 subunit was linked from Gly^{93} to the Thr^{12} of the next unit. No artefact residues were present in the linker.

and using Ubc13–Mms2 as a cognate E2 enzyme to promote the Lys^{63}-linked ubiquitination of SUMOylated target proteins [11]. RNF4 relies on the SIMs domain for selective binding of poly-SUMO chains over monomer SUMO. For instance, only poly-SUMOylated PML (promyelocytic leukaemia) proteins can be recognized by RNF4 [16]; however, the structural basis of this recognition is unclear. Recently, Keusekotten et al. [17] showed that the sequence and spacing of the RNF4-SIMs domain regulate the avidity-driven recognition of poly-SUMO chains. They also concluded that the SIM2 and SIM3 regions are necessary and sufficient for binding to a SUMO chain, whereas the SIM4 region is only needed for recognition of longer chains. We provide a structural analysis of the RNF4-SIMs domain and tetra-SUMO chains using a combination of NMR spectroscopy, X-ray crystallography and SAXS. We also report a previously unexplored weak binding between isolated SIM peptides and SUMO, and interpreted it to delineate interactions between these fundamental units. On the basis of these experimental results we built a HADDOCK model of the RNF4-SIMs domain–tetra-SUMO2 complex that provides insights on poly-SUMO recognition.

EXPERIMENTAL

Cloning and protein purification

A human RNF4-SIMs domain (residues Glu^{32} to Arg^{82}) was subcloned in a custom pET-30a vector with an N-terminal His tag followed by a TEV (tobacco etch virus) protease cutting site (Figure 1A). GST fusion proteins for pull-down experiments were prepared by subcloning the respective RNF4-SIMs fragments in a pGEX-4T-1 vector. SIM-knockout mutants were prepared using site-directed mutagenesis by replacing the hydrophobic residues in targeted SIM(s) with alanine as follows: $SIM1_{mt}$, $I^{35}ELV$ to AEAA; $SIM2_{mt}$, $I^{45}VDLT$ to AADAT; $SIM3_{mt}$, $V^{56}VVDLT$ to AAADAT; and $SIM4_{mt}$, $V^{66}VIV$ to AAAA. Double SIM mutants of $SIM13_{mt}$, $SIM14_{mt}$, $SIM12_{mt}$, $SIM23_{mt}$, $SIM24_{mt}$ and $SIM34_{mt}$ were also prepared accordingly (Figure 1B).

The linear poly-SUMO2 chains, with the N-terminal 11 residues of each constituent SUMO2 deleted (ΔN11-poly-SUMO2), were generated using a previously described strategy [12]. Briefly, ΔN11-poly-SUMO2 (residues Thr^{12} to Gly^{93}) repeats were subcloned in the p3XFLAG-CMV-7.1-2 vector by the stepwise linking of each subunit with a combination of restriction enzymes, including EcoRI, KpnI and XbaI (Figure 1C). Thus, unlike previous studies [17], our construct did not include any cloning artefact residues in the subunit linker because the DNA sequence recognized by KpnI translated as -GlyThr-. Poly-SUMO2 chains from the p3XFLAG-CMV-7.1-2 vector were then transferred to pMAL-c2X using EcoRI and XbaI. Finally, SUMO2 chains were subcloned into pET28a from pMAL-c2X using EcoRI and HindIII as restriction enzymes. Monomeric SUMO1 and SUMO2 were subcloned in pET-15b and were used in the binding study with four individual RNF4 SIM peptides. The commercially synthesized SIM peptides that were used are presented in Figure 1(A). All proteins were expressed in the Rosetta(DE3) strain of *Escherichia coli*. Isotope-enriched samples for NMR experiments were prepared using an M9 medium supplemented with $^{15}NH_4Cl$ and/or [^{13}C]glucose. The samples were first purified using affinity chromatography, followed by proteolysis to remove the purification tag. They were finally purified using FPLC.

NMR spectroscopy

All NMR experiments were conducted at 17 °C using either a Bruker Avance 600- or a 500-MHz NMR spectrometer equipped with a 5-mm triple-resonance cryoprobe. Data were acquired

and processed using Topspin 2.1 and further analysed using SPARKY (http://www.cgl.ucsf.edu/home/sparky/) or CcpNMR analysis [18]. Protein backbone resonance assignments were achieved using standard triple-resonance experiments, including HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO. ^{15}N-edited TOCSY-HSQC (heteronuclear single-quantum coherence) acquired with a 60-ms mixing time was used to derive side-chain assignments. The steady-state heteronuclear ^{1}H-^{15}N-NOE experiment was conducted in triplicate in an interleaved manner, with and without proton saturation. SIM peptide binding was studied using systematic titration against 0.12 mM [^{15}N]SUMO1 or [^{15}N]SUMO2 in 10 mM potassium phosphate (KPi) buffer, pH 7.5. The microscopic dissociation constant of each perturbing amino acid residue in SUMO1/2 was calculated using a non-linear regression module provided by CcpNMR software. Dissociation constants of the residues in the binding groove of SUMO1/2 were averaged and recorded as equilibrium dissociation constants, EC_{50}, between SUMO and isolated SIM peptides. CSP (chemical shift perturbation) was calculated using the formula:

$$\Delta\delta_{comp} = \sqrt[2]{(\Delta\delta_H)^2 + (0.154 \times \Delta\delta_N)^2}$$

where $\Delta\delta_H$ is the perturbation of the proton and $\Delta\delta_N$ is the perturbation in the ^{15}N chemical shift. The binding specificity of the SUMO2 chain was studied by titrating di-, tri- and tetra-SUMO chains against recombinant [^{15}N]RNF4-SIMs domain peptide dissolved in 10 mM KPi, 100 mM NaCl and 2 mM DTT (pH 7.5). The resonance broadening or disappearance of the [^{15}N]RNF4-SIMs domain that resulted from the titration was analysed using the peak intensity ratio (I/I_o). The intensity of each resonance at a different titration ratio (I) was compared with its intensity in a free form state (I_o). The averaged intensity ratio of the hydrophobic (core) residues in each SIM was taken as a measure of the relative amount of bound form for each SIM.

Paramagnetic relaxation enhancement

PRE (paramagnetic relaxation enhancement) experiments were conducted at 290 K on a 600-MHz NMR spectrometer. An MTSL [*S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate] spin label was coupled to the Cys^{51} of the SIM3 peptide to determine its binding orientation on SUMO2 following the method described in [7]. Labelled peptide and free MTSL were separated using Superdex Peptide HR10/30 (catalogue number 9648074; Pharmacia Biotech) and the identity of RNF4-SIM3^{CMTSL} was verified using MS. SIM3^{CMTSL} peptide was added to a 1:1 molar ratio with [^{15}N]SUMO2 at a final concentration of 0.25 mM in 10 mM KPi buffer (pH 7.5). Ascorbic acid was added to quench the paramagnetic state. ^{15}N-HSQC spectra were acquired before and after the addition of ascorbic acid to determine the peak intensity ratios at paramagnetic and diamagnetic states (I_{para}/I_{dia}). For PRE by Mn^{2+}, [^{13}C,^{15}N]SUMO2 with a His tag at the C-terminus for chelating the paramagnetic Mn^{2+} ion, was mixed at 1:1 molar ratio with SIM peptides to a volume of 400 μl at 0.62 mM in 10 mM deuterated Tris buffer (pH 7.5). The ^{13}C,^{15}N-filtered TOCSY spectra of two samples, one containing 0.5 mM Zn^{2+} and the other containing 0.5 mM Mn^{2+}, were acquired to determine the peak intensity ratios of each residue (I_{Mn2+}/I_{Zn2+}) of SIM in the fingerprint region at a diamagnetic state (I_{Zn2+}) and at a paramagnetic state (I_{Mn2+}) [19,20].

GST pull-down experiment

Purified GST (glutathione transferase; 500 μl of 0.5 mM) that was fused with wild-type or mutant RNF4-SIMs domain (Figure 1B) was bound to 10 ml of GST–agarose resin and washed thoroughly with a binding buffer (50 mM KPi) and 0.5 M NaCl (pH 8.0). A mixture representing 0.5 mM di-, tri- and tetra-SUMO2 was incubated with 10 ml of RNF4-SIMs domain bounded GST–resin for 1 h at 25 °C. Resins were then washed with 150 ml of 10 mM potassium phosphate and 300 mM NaCl (pH 7.5) buffer to remove excess proteins. Subsequently, 10 μl of resin was taken for SDS/PAGE.

X-ray crystallography

Tetra-SUMO2 was crystallized using the hanging-drop method at 16 °C in 0.1 M Tris/HCl, 0.1 M CHES and 30 % PEG-600 (pH 9.4). Diffraction data were collected using the 0.9 Å X-ray at SPring-8, Japan, on beam line BL44XU and processed using an HKL2000 program suite. The tetra-SUMO structure was determined by using the molecular replacement method with a SUMO monomer as a model molecule (PDB code 1WM3) [21,22].

SAXS

SAXS data were collected using X-rays generated at 15 keV at the 23A SWAXS beamline at the National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan [23]. Sample solutions were buffered in 10 mM KPi, 100 mM NaCl and 2 mM TCEP [tris-(2-carboxyethyl)phosphine] (pH 7.5), and loaded into thermostated cells, 45 μl each, for a 2.5-mm X-ray pathlength at 20 °C. The concentration ranges used for each sample are listed in Supplementary Table S1 (http://www.biochemj.org/bj/462/bj4620053add.htm). Solution cells sealed with thin 30-μm quartz (or 8-μm Kapton) windows 4-mm in diameter were slowly rocked within an area of 1.5 mm × 1.5 mm to minimize prolonged spot exposure (approximately 0.5 mm in beam diameter) and reduce radiation effects on the samples. Data were collected using a Pilatus 1M-F area detector. Identical SAXS profiles extracted from the exposures (implying a small radiation effect) were selectively combined for improved data statistics. Sample solution scattering was subtracted by scattering from a corresponding buffer solution collected under identical experimental conditions and scaled to the absolute intensity scales (i.e. scattering cross-section per unit volume of units of cm^{-1}) through the scattering from water (further counter-checked with the scattering from a solution of cytochrome *c*), following the standard data reduction procedures previously detailed [23]. Data were processed and analysed using the PRIMUS v3.2, GNOM program: SAS DATA analysis and ATSAS online software were provided by EMBL, as described in [24–30].

Experimentally guided modelling

We used HADDOCK software, version 2.1 [31], to interpret the sparse NMR data and build models of individual SIM–SUMO2 complexes. Docking was performed using the high-resolution crystal structure of SUMO2 (PDB code 1WM3) and 100 random conformations of the respective SIM peptides as templates. Ambiguous iterative restraints were defined between strongly perturbed and solvent-exposed residues on SUMO2 and six core residues from each SIM, defined as SIM1 (I^{36}ELVET), SIM2 (I^{46}VDLTC), SIM3 (V^{58}VDLTH) or SIM4 (V^{67}VIVDE), while treating the entire peptide as fully flexible. The dihedral angle and hydrogen-bond restraints were predicted from chemical shifts of each bound form of SUMO2 [32]. Structural restraints

from these runs were implemented in a five-component simulation that included 100 random conformers of the intact RNF4-SIMs domain and four templates of SUMO2 with the intersubunit polypeptide bonds mimicked by unambiguous distance restraints. This approach was selected to provide maximal freedom for the docking of each SUMO2 subunit [33]. These structures were then externally ranked using analysis of the SAXS data acquired on the tetra-SUMO2–RNF4-SIMs domain complex [34]. Finally, the intersubunit linkers were externally rebuilt and the complex structure was refined using HADDOCK.

RESULTS

The RNF4-SIMs domain was intrinsically disordered in solution

The N-terminal region of RNF4 protein spanning residues Glu^{32} to Arg^{82} contained numerous hydrophobic residues that were mostly congregated in four segments identified as SIM1–SIM4 or collectively as the SIMs domain (Figure 1A). Secondary structure predictions made by using various algorithms suggested that the SIMs domain may contain β-strands and α-helices (Supplementary Figure S1 at http://www.biochemj.org/bj/462/bj4620053add.htm) [15]. Clustered SIMs were also predicted to form a folded β-structure or exist in dynamic balance between a folded, standalone β-sheet and an open less ordered conformation with multiple extended SIMs [15]. The presence of ordered structures affected the interaction of the SIMs domain with poly-SUMOs, thus more detailed structural and functional characterization of the SIMs domain was essential for understanding how poly-SIM-containing proteins interpret the SUMO signal. This prompted us to further probe the conformation and dynamics of the RNF4-SIMs domain in solution by using NMR. The amide resonances in the ^{15}N-HSQC NMR spectrum of SIMs domain were well-resolved and assigned using standard triple-resonance experiments (Figure 2A). The clustering of the amide resonances in the 8.0–8.7 p.p.m. region at the proton dimension is characteristic of an unstructured peptide, suggesting that the SIMs domain exists in a disordered state [35]. We then applied sequence-specific NMR chemical shifts to predict secondary structure propensities [36]. The results indicated that the SIMs domain is largely unstructured with varying small propensities for β- and α-conformation between residues Ile^{36}–Val^{70} and Glu^{72}–Asn^{79} respectively (Figure 2B). Specifically, the SIM1 region exhibited a 20 % propensity to form β structures, whereas the poly-arginine segment that follows SIM4 exhibited an approximately 20 % propensity to form α structures. The middle SIM2 and SIM3 segments, which were identified as the most critical for poly-SUMO recognition, exhibited little secondary structure propensity [15,17].

We then applied SAXS, which is well-suited for identifying and modelling flexibility as an ensemble set of biological macromolecules in solution, to probe the solution conformation of the RNF4-SIMs domain (Figure 2C, red) [37,38]. Rambo and Tainer [38,39] have proposed several criteria as a necessary and sufficient condition of the presence of flexibility that warrants the modelling of SAXS data as an ensemble set: (i) lack of convergence in a Kratky plot; (ii) invalidation of the Porod–Debye law; (iii) Flory's inequality, that is, $R_g > R_g^{RC}$ where R_g is the radius of gyration derived from SAXS data and R_g^{RC} is related to the number of residues, N, as $R_g^{RC} = (2.54 \pm 0.01)N^{(0.522 \pm 0.01)}$; (iv) low particle density; and (v) inability to model data with a single model. An analysis of SAXS data indicated that the RNF4-SIMs domain existed as a disordered peptide. Specifically, (i) the Kratky plot was a hyperbolic shape (Figure 2D, red) and the Porod–Debye plot (Figure 2E, red) did not converge to a plateau; (ii) SAXS data fitted to the Guinier equation predicted a radius of gyration of 25.8 Å for the SIMs domain, which is larger than its theoretical R_g^{RC}, calculated to be 20.961 ± 0.167 Å; (iii) the particle density of the RNF4-SIMs domain calculated from SAXS data was 1.108 g·cm^{-3}, which is considerably smaller than the canonical density of approximately 1.37 g·cm^{-3} for a compact folded protein; and (iv) we could not fit the SAXS data of the RNF4-SIMs domain to a single model.

The dynamics of the SIMs domain were further characterized using NMR relaxation (Figure 2F). The uniformly small ^{1}H-^{15}N-NOE values, which approached zero for all residues up to the end of SIM4 (Val^{70}), indicated that this region behaved like a random coil and was highly dynamic at the picosecond timescale. The C-terminal residues were apparently less flexible. In particular, residues between Glu^{72}–Pro^{76} exhibited considerably high NOEs, up to 0.4, and large R2 values, suggesting that this region may exist as an area of ordered structure. This is consistent with the prediction of approximately 20 % helical propensity. The observation of negative values of the N-terminal residues indicated that these residues move with correlation times that are expected for small molecules (Figure 2F).

Binding of an individual SIM peptide to SUMO1 and SUMO2

We employed NMR CSP to probe the low-affinity binding of an individual SIM to SUMO that could not be measured using ITC (isothermal titration calorimetry). Because such interactions are known to be weakened by a high-salt concentrations [8], we used a potassium phosphate buffer without salt to detect this weak binding. A systematic titration was conducted using the stepwise addition of concentrated SIM peptide solution to ^{15}N-labelled SUMO1 or SUMO2. The change at each step was monitored using the ^{15}N-HSQC spectrum (Figure 3A). A parallel set of experiments was conducted on SUMO1 (Supplementary Figure S2 at http://www.biochemj.org/bj/462/bj4620053add.htm). The binding equilibrium appeared on the fast-exchange NMR timescale, as characterized by concentration-dependent shifts of specific residues (Figure 3B). The average CSP for each SUMO residue was calculated from two titration end point spectra described previously [7]. Figure 3(C) summarizes the perturbation of SUMO2 residues on binding isolated SIM peptides. The adjoining figure shows the degree of perturbation for each residue as mapped on to the structure of SUMO2 (Figure 3D). As expected, the most perturbed residues were located primarily in the β2-strand and α1-helix, which comprised the SIM-binding groove. The adjoining β1-strand was perturbed to a lesser extent. The binding equilibria were analysed using non-linear regression to extract microscopic dissociation constants for each residue on the binding interface. The apparent dissociation constants, EC_{50}, of the SIM–SUMO complexes are summarized in Table 1. SIM2 had the highest affinity with an EC_{50} of 0.04 ± 0.01 mM for SUMO2, whereas those for SIM1, SIM3 and SIM4 were lower by factors of 8, 2 and 23 respectively. Consistent with previously reported ITC data [17], the middle two SIMs, SIM2 and SIM3, were strong contenders for binding with SUMO and both exhibited a significant preference by factors of 4.25 and 1.55 for SUMO2 compared with SUMO1. Because only SUMO2/3 exhibited a genuine SUMO modification consensus sequence, the SUMO chains were mostly composed of SUMO2/3 with the possibility of SUMO1 acting as an end cap to terminate further elongation [3]. The findings explain why the SIM2 and SIM3 regions of the RNF4-SIMs domain are necessary for SUMO chain recognition [12,17]. The remaining two outer SIM regions were also capable of binding to the SUMO and may have contributed

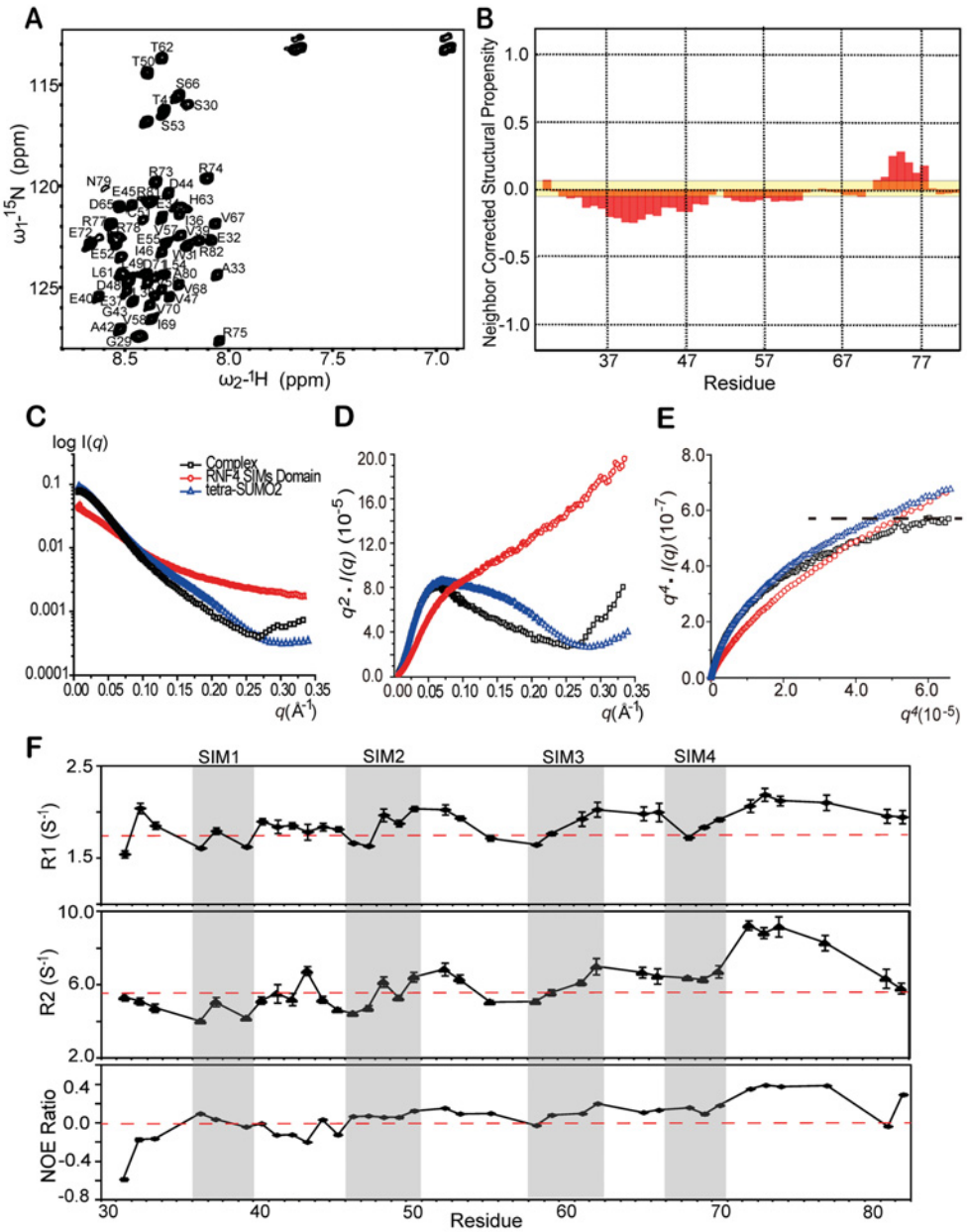

Figure 2 Characterization of the structure and dynamics of the RNF4-SIMs domain

(**A**) Assigned ^{15}N-HSQC spectrum of [^{15}N]RNF4-SIMs domain purified from the His tag fusion pET-30a system. The His tag was removed by using the TEV enzyme. (**B**) Neighbour-corrected structural propensity analysis for the RNF4-SIMs domain by using chemical shifts assigned from standard triple-resonance experiments. The merged SAXS scattering profile, Kratky plot and Porod–Debye plot of the RNF4-SIMs domain (red), tetra-SUMO2 (blue) and RNF4-SIMs domain–tetra-SUMO2 complex (black) are shown in (**C**), (**D**) and (**E**) respectively. (**F**) The amide ^{15}N longitudinal relaxation rate (R1), transverse relaxation rate (R2) and the ^{1}H-^{15}N-NOE of the [^{15}N]RNF4-SIMs domain are present on the top, middle and bottom panel of (**F**) respectively. The red broken lines are a visual guide only.

to the binding avidity through the multivalency effect. Between these two regions, SIM1 exhibited a 1.84-fold higher affinity for SUMO2 than SUMO1, whereas SIM4 bound both paralogues with similar affinity.

We attempted to determine NMR solution structures of all four SIM complexes, but, because of weak binding affinities, we failed to observe intermolecular NOE cross-peaks. We then used the experimental-guided modelling approach HADDOCK to interpret our NMR observations, as was performed previously (Figure 3D) [31]. SIMs are typically composed of a cluster of hydrophobic residues, I/V/L-X-I/V/L-I/V/L, followed by acidic or polar residues [6]. The most favourable HADDOCK score structures indicated tentative interaction between SUMO2 and the core residues of the respective SIM module, including SIM1 (I^{36}ELVET), SIM2 (I^{46}VDLTC), SIM3 (V^{58}VDLTH) and SIM4 (V^{67}VIVDE). All SIM regions of RNF4 contained a short acidic/polar residue patch; therefore, because electrostatic contacts were lacking, these exhibited weak binding affinity for the SUMO. The first three repeats, SIM1–SIM3, also contained an atypical acidic residue in the hydrophobic cluster. An acidic residue in this position can interact with Lys^{33} of SUMO2/3 and is known to impart paralogue preference for SUMO2/3 [6,9].

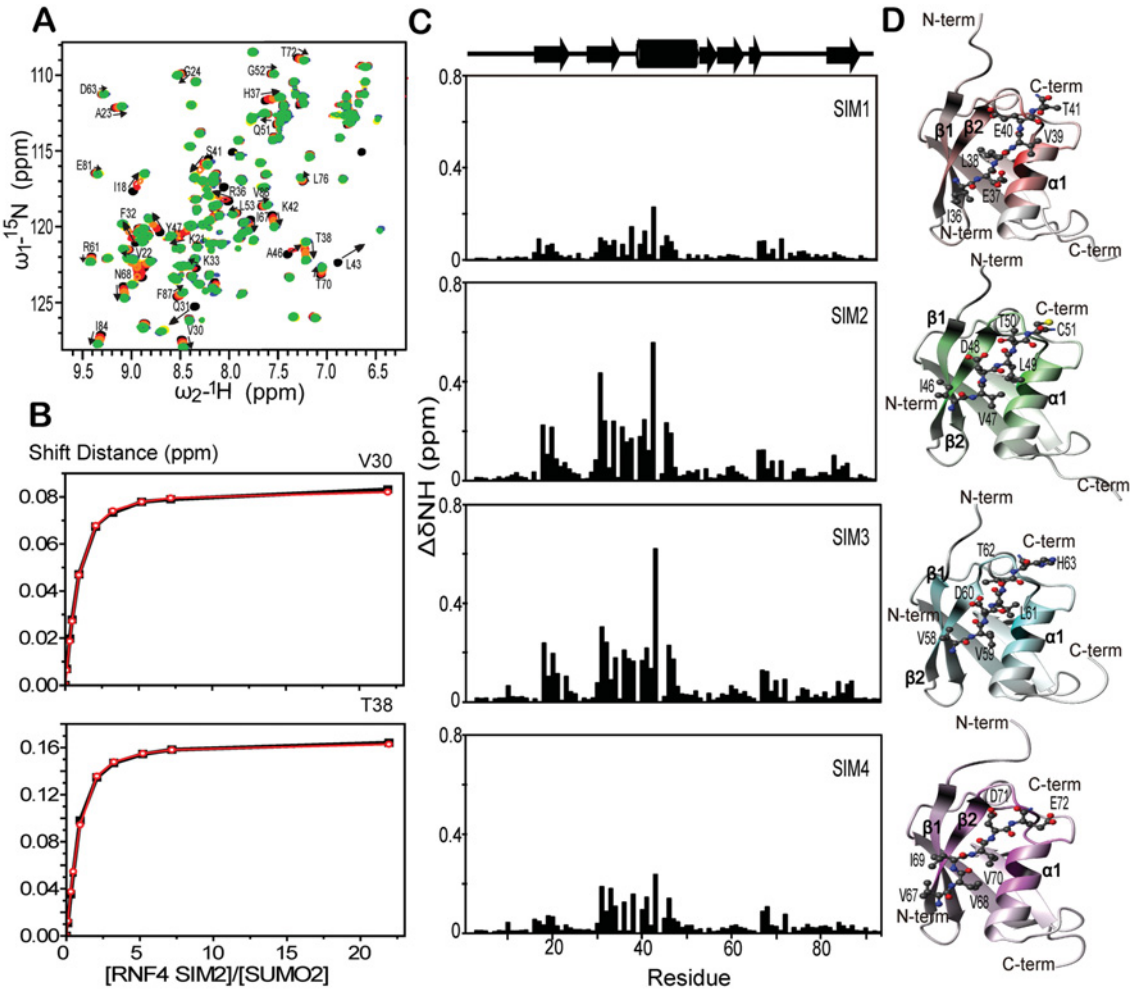

Figure 3 Mapping the SIM peptide binding site on SUMO2 by NMR

(**A**) ^{15}N-HSQC spectra of [^{15}N]SUMO2 in the absence (black) and presence of 2:1 (red), 1:1 (orange), 1:3 (yellow), 1:7 (green) and 1:21 (blue) molar ratio of SUMO2–SIM-2 respectively. (**B**) The chemical shift titration curves of Val^{30} and Thr^{38} of SUMO2. The composite CSP is defined as $\Delta\delta_{comp} = \sqrt[2]{(\Delta\delta_H)^2 + (0.154 \times \Delta\delta_N)^2}$, where $\Delta\delta_H$ is the proton chemical shift change and $\delta\Delta_N$ is ^{15}N chemical shift change. The titration curves were fit to a linear regression equation provided in the CcpNMR program to determine the binding constants. The binding constants of the SIM domain residues were averaged and the results are listed in Table 1. (**C**) End point composite CSP of SUMO2 induced by >20:1 molar excess of SIM1, SIM2, SIM3 and SIM4 (top to bottom). (**D**) Respective HADDOCK models of individual SIM–SUMO complexes. The extent of chemical perturbation was coded as a colour gradient with darker colours representing more perturbed residues. All SIM peptides were docked parallel to the β2-strand of SUMO2.

Table 1 Equilibrium dissociation constants between SUMO paralogues and isolated SIM peptides

The reported EC_{50} values are means ± S.D.from a minimum of five residues.

	EC_{50} (mM)		
SIM	SUMO1	SUMO2	Ratio SUMO1/SUMO2
SIM1	0.61 ± 0.29	0.33 ± 0.24	1.84
SIM2	0.17 ± 0.03	0.04 ± 0.01	4.25
SIM3	0.14 ± 0.04	0.09 ± 0.04	1.56
SIM4	0.86 ± 0.49	0.91 ± 0.29	0.95

Thus the RNF4-SIMs domain adorned an optimized sequence to identify central SUMO2/3 subunits of a SUMO chain and was fine-tuned to a weaker affinity to avoid binding to the monomer SUMO.

Binding orientation of an individual SIM peptide to SUMO2

We employed PRE to probe the binding orientation of an individual SIM on SUMO2. Paramagnetic line broadening is proportional to the inverse six power of the distance between nuclear spin and the paramagnetic ion; therefore, it is highly sensitive to distance variations. We conducted the PRE experiments in two ways. First, we attached the spin label, MTSL, at Cys^{51} of the SIM3 peptide (SIM3^{CMTSL}) and monitored the broadening of SUMO2 amide resonances. ^{15}N-HSQC spectra at a 1:1 molar ratio of the [^{15}N]SUMO2–SIM3^{CMTSL} complex were acquired before (paramagnetic state) and after the addition of excess ascorbic acid to reduce MTSL in the diamagnetic state. The intensity ratio of each residue in SUMO2 before (I_{para}) and after (I_{dia}) the addition of ascorbic acid, I_{para}/I_{dia}, was measured. As shown in Figure 4(A), the residues that experienced a substantial drop in intensity were located at four regions: Val^{22}–Val^{30}, Leu^{40}–Arg^{61}, Asn^{68}–Asn^{71} and Val^{86}–Gln^{88}. Mapping of these residues on to the structure of SUMO2 indicated that they were positioned

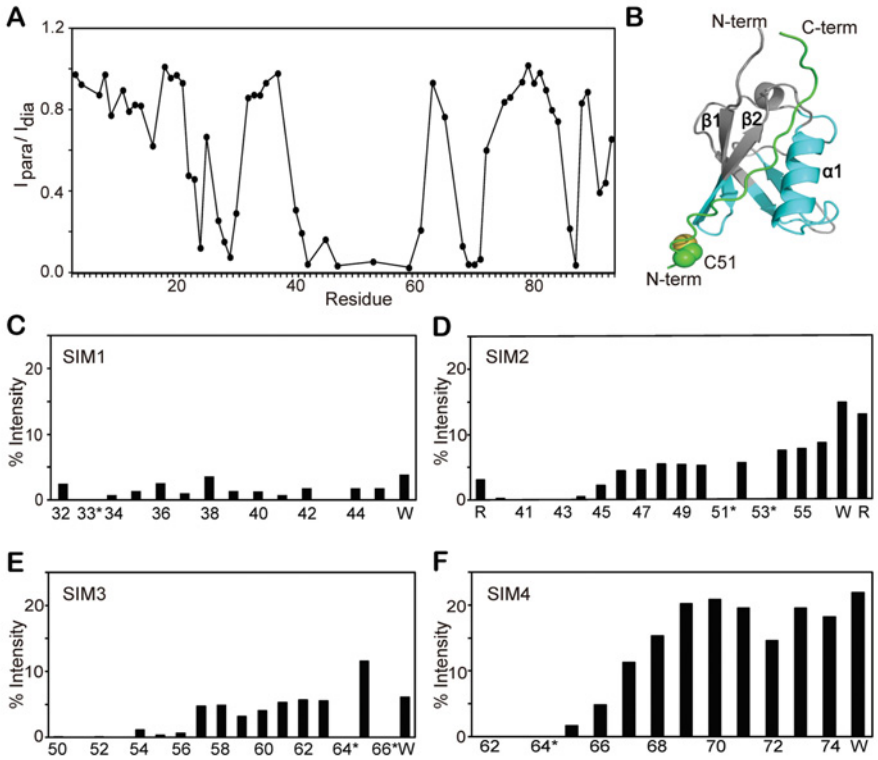

Figure 4 Binding orientation of RNF4-SIMs to SUMO2

(**A**) Intensity ratio of the SUMO2 amide resonances in paramagnetic (I_{para}) and diamagnetic (I_{dia}) states at 1:1 molar ratio of SIM3–SUMO2. The MTSL spin label was coupled with Cys^{51} on the SIM3 peptide. (**B**) SIM3–SUMO2 model showing the spatial locations of the SIM3 peptide (green) and Cys^{51} (spheres). SUMO2 is presented in grey. Residues affected by PRE [I_{para}/I_{dia}] (cyan) were mostly congregated in the N-terminal side of the β2-strand of SUMO2, indicating that SIM3 is bound in a parallel orientation. (**C**–**F**) Variation with sequence of the percentage intensity of SIM1–SIM4 peptide respectively, upon binding to metal chelated SUMO2. The percentage intensity was measured as the percentage of peak height in the presence of Mn^{2+} (I_{Mn2+}) or Zn^{2+} (I_{Zn2+}). The asterisk (*) represents residues that could not be properly assigned in the $^{13}C,^{15}N$-filtered TOCSY spectra. Residues shown as a single letter were artificially added in the peptide sequence for the purpose of increasing solubility.

at the N-terminal side of the binding cleft near the spin label at Cys^{51} the N-terminus of the β2-strand (Figure 4B), which was consistent with the peptide binding parallel to the β2-strand. Labelling MTSL at Cys^{51} in SIM2 was not appropriate because it lay too close to the binding site and could have affected the binding of SIM2 to SUMO2.

To further corroborate the study, we prepared a SUMO2 construct containing a His tag at the C-terminus to serve as a metal ion-chelating site. $^{13}C,^{15}N$-filtered TOCSY spectra of two samples, each containing the [$^{13}C,^{15}N$]SUMO2–SIM complex at a 1:1 molar ratio, were taken to measure the Mn^{2+}-induced perturbation on the SIMs. One of the samples contained a 0.8 molar ratio of diamagnetic Zn^{2+} ion and the other contained the paramagnetic Mn^{2+} ion. The $^{13}C,^{15}N$-filtered TOCSY experiment effectively filtered out all proton signals from the $^{13}C,^{15}N$-labelled SUMO2 and allowed for observation of only the unlabelled SIMs in the TOCSY spectra whose ^{1}H resonances had been previously assigned. Ideally, the paramagnetic Mn^{2+} ion would chelate to the C-terminal His tag of SUMO2 that resided near the N-terminus of the β2-strand on one end of the SIM-binding pocket. Amide resonances from residues bound near the N-terminus of the β2-strand would be preferentially broadened with a loss of intensity. Figures 4(C)–4(F) show the sequence variations of the intensity ratio with Mg^{2+} or Zn^{2+}, I_{Mn2+}/I_{Zn2+} of the four SIM/$^{13}C,^{15}N$-labelled SUMO2 complexes. SIM2, SIM3 and SIM4 indicated strong preferential intensity loss for residues on the N-terminus, suggesting that these three SIMs bound to SUMO2 parallel to the β2-strand. SIM1 lost approximately 95 % of the intensity of all resonances without an orientation preference. Because all three high-affinity SIMs bound to the SUMO2 in parallel orientation and SIM1 was the weakest binder, we concluded that the RNF4-SIMs domain bound to the tetra-SUMO2 in parallel orientation. We attributed the loss of intensity in SIM1 to its high-acidic-residue content (six residues), which rendered it a high affinity for chelating the positively charged Mn^{2+} ions. Although SIM2 contained six acidic residues, the high affinity of SIM2 for SUMO2 may have prevented SUMO2 from binding to metal ions non-specifically. Viewed together with the MTSL experiment involving the SIM3 peptide, we concluded that the RNF4-SIMs domain bound to the poly-SUMO2 in a parallel orientation.

Avidity of interaction between the SIMs domain and the poly-SUMO2

To investigate preferential interaction with individual SIM regions in the SIMs domain, unlabelled poly-SUMO2 chains of various lengths were titrated against the ^{15}N-labelled SIMs domain. This resulted in selective resonance disappearance/broadening, characteristics of a slow to intermediate exchange (Figure 5A). Furthermore, most resonance disappeared when a 1:1 molar ratio of poly-SUMO2 was added. In comparison, the binding of mono-SUMO–mono-SIM was in the fast exchange limit and a high ($\times$ 20) molar ratio of SUMOs needed to be added for

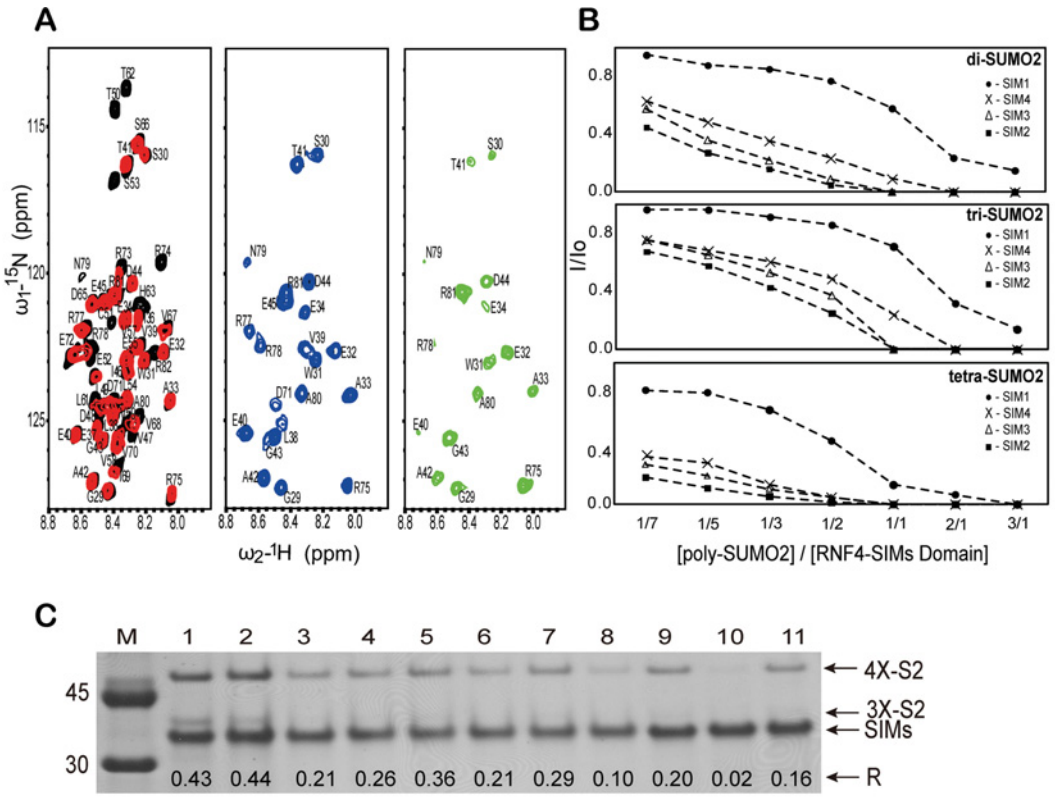

Figure 5 Binding avidity of RNF4-SIMs domain to poly-SUMO2

(**A**) ^{15}N-HSQC spectra of the free [^{15}N]RNF4-SIMs domain (black) and its complexes with tetra-SUMO2 at a SIMs domain/tetra-SUMO2 molar ratio of 3:1 (red), 1:1 (blue) or 1:3 (green). Peaks in the SIM regions shift and/or broaden as the molar ratio of tetra-SUMO2 increases. (**B**) Variation in average peak intensity ratios of the SIM1 (●), SIM2 (■), SIM3 (Δ) and SIM4 (×) regions upon binding to a di-, tri- and tetra-SUMO2 chain. Each point is the average intensity ratio of the hydrophobic residues in the SIMs motifs. (**C**) Gel pull-down assay. Equal amount of wild-type RNF4-SIMs domain (lane 1) or its mutant was added to a mixture containing 0.5 mM each of di-, tri- and tetra-SUMO2. RNF4-SIMs domain mutants added to the poly-SUMO2 mixture were: $SIM1_{mt}$ (lane 2), $SIM2_{mt}$ (lane 3), $SIM3_{mt}$ (lane 4), $SIM4_{mt}$ (lane 5), $SIM13_{mt}$ (lane 6), $SIM14_{mt}$ (lane 7), $SIM34_{mt}$ (lane 8), $SIM12_{mt}$ (lane 8), $SIM23_{mt}$ (lane 10) and $SIM24_{mt}$ (lane 11). The bands corresponding to tetra-SUMO2, tri-SUMO2 and the GST–RNF4-SIMs domain mutants are labelled as 4X-S2, 3X-S2 and SIMs respectively. The ratios of the band intensities of tetra-SUMO2 and the RNF4-SIMs domain mutants, $R = I_{\text{tetra-SUMO2}}/I_{\text{SIMs mutants}}$, are given below each lane.

the perturbation to reach a plateau. Thus linking multiple SIMs together in the SIMs domain greatly enhanced its binding to poly-SUMO2 synergistically. Determining the binding affinity was impossible because the correlation between resonance line broadening and the amount of bound form is not simple in the slow-exchange regime.

We analysed the site-specific interaction between specific SIMs in the SIMs domain and the poly-SUMO2 chain by monitoring the resonance line intensity of each residue on the SIMs domain (Supplementary Figure S3 at http://www.biochemj.org/bj/462/bj4620053add.htm). Consistent with SUMO binding directly to the SIMs sites, residues in the SIM sites were affected more than those in the spacer regions under all titration conditions. The average intensity ratios of residues in each SIM were plotted against the titration ratios for monitoring the preferential SUMO2 binding of each SIM in the SIMs domain (Figure 5B). In all cases, the decrease in intensity for RNF4-SIMs followed the order SIM2>SIM3>SIM4>>SIM1. Consistent with the binding of individual SIMs to mono-SUMO and with previous findings [12,15,17], SIM2 and SIM3 both exhibited the highest affinity towards poly-SUMO2. Whereas SIM1 was not considerably perturbed at low SUMO2 ratios, the binding affinity of SIM4 was greatly enhanced to that of SIM2/3 when linked together in the RNF4-SIMs domain. A tentative explanation of this effect is discussed below. Notably, tetra-SUMO bound to the SIMs domain at much higher affinity compared with di- or tri-SUMO2. The contribution of SIM1 to binding avidity seems minimal, but SIM4 evidently contributes substantially to the binding avidity.

Effect of mutant SIMs on binding to poly-SUMO2

To further dissect the contribution of individual SIMs in poly-SUMO2 recognition, GST pull-down assays was performed using wild-type and SIM-knockout mutants (SIM_{mt}) (Figure 1B). The ability of a RNF4-SIMs domain mutant to pull down tetra-SUMO can be expressed as the ratio of the intensity of the tetra-SUMO2 band ($I_{\text{tetra-SUMO2}}$) and that of the RNF4-SIMs domain mutant band ($I_{\text{SIMs mutant}}$) in the polyacrylamide gel (Figure 5C), $R = I_{\text{tetra-SUMO2}}/I_{\text{SIM mutants}}$. The results are listed below each lane in Figure 5(C). Wild-type SIMs domain predominantly pulled down tetra-SUMO2 ($R = 0.43$) from a mixture of di-, tri- and tetra-SUMO2 chains, whereas only a trace amount of tri-SUMO2 and none of di-SUMO2 were visible on gel (Figure 5C, lane 1). This once more demonstrated the avidity interaction between SIMs domain and poly-SUMO chain. We then tested the contribution of single SIM by replacing the hydrophobic residues in selective SIMs with alanines (Figure 5C, lanes 2–5). The effect of $SIM1_{mt}$ (Figure 5C, lane 2) on binding with tetra-SUMO2 was negligible ($R = 0.44$). Single knockout of SIM2 or SIM3 (Figure 5C, lane 3 or 4) significantly reduced the intensity of both tri- and

tetra-SUMO2 bands with $R = 0.21$ and 0.26 for $SIM2_{mt}$ and $SIM3_{mt}$ respectively. Surprisingly, $SIM4_{mt}$ (Figure 5C, lane 5) also caused significant reduction in the intensity of tri- and tetra-SUMO2 band ($R = 0.36$), although to a lesser extent than that of $SIM2_{mt}$ or $SIM3_{mt}$. This result was congruent with previous experiments and again emphasized the crucial contribution by SIM2 and SIM3 and the peripheral role of SIM1. Consistent with the results observed for the NMR titration experiment SIM4 apparently played a significant role in binding avidity of RNF4-SIMs domain to tetra-SUMO2. To further corroborate the single SIM mutation experiments we also examined the binding of tetra-SUMO2 to RNF4-SIMs domain with double-SIM mutants. In general, double-SIM mutants containing a mutation at SIM1 have tetra-SUMO-binding affinities similar to those of the corresponding single SIM mutants with a wild-type SIM1, indicating that SIM1 does not have much of a role in binding. Significantly, double-SIM mutants containing a combination of SIM2, SIM3 and SIM4, including $SIM34_{mt}$ ($R = 0.10$) (Figure 5C, lane 8), $SIM23_{mt}$ ($R = 0.02$) (Figure 5C, lane 10) and $SIM24_{mt}$ ($R = 0.16$) (Figure 5C, lane 11) all have substantially reduced binding affinity towards tetra-SUMO2. As expected, $SIM23_{mt}$ cannot bind to any of the poly-SUMO2s. To our surprise, $SIM34_{mt}$ (Figure 5C, lane 8) has the second lowest ability to pull down tetra-SUMO2.

The gel pull-down experiments were in general agreement with the NMR results and those of Keusekotten et al. [17]. The major surprise when compared with the results of individual SIMs was the observation of a much enhanced contribution of SIM4 towards poly-SUMO2 binding in RNF4-SIMs domain. In isolation, SIM4 peptide is the weakest binder to SUMO2; however, in the RNF4-SIMs domain the contribution of SIM4 is comparable with those of SIM2 and SIM3. We speculate that the enhancement is due to the inclusion of the poly-arginine patch immediately following the SIM4 site. In our HADDOCK model of the RNF4-SIMs domain–tetra-SUMO2 complex this poly-arginine patch is located in the vicinity of a large acidic surface on SUMO2 (discussed below). The Coulomb interaction between the two charged patches could play a pivotal role in the avidity of RNF4-SIMs domain–poly-SUMO2 interaction.

Structure of tetra-SUMO2 and its complex with the RNF4-SIMs domain

Previous studies have demonstrated that a polypeptide-linked ΔN11-SUMO chain can bind to a RNF4-SIMs domain with similar affinities as a native isopeptide-linked SUMO chain [12,17]. Because of the ease of preparation of homogeneous samples at a high concentration, we employed NMR and X-ray crystallography to assess the tertiary structure of such a polypeptide-linked tetra-SUMO2 chain. The ^{15}N-HSQC spectrum of tetra-SUMO2 was practically identical with that of SUMO2, except that resonance in tetra-SUMO2 was broader because of its longer correlation time (Figure 6A). The lack of CSP from monomer SUMO2 indicated a beads-on-a-string scenario whereby all four SUMO2 subunits in the tetra-SUMO2 chain moved freely with little interaction among themselves. We also solved the crystal structure of the tetra-SUMO2 chain (PDB code 1WM3) (Supplementary Figure S4 and Table S2 at http://www.biochemj.org/bj/462/bj4620053add.htm). Similar to the previously reported crystal structure of the di-SUMO2 [17], the structure of the individual SUMO2 in the tetra-SUMO2 crystal was indistinguishable from that of the free SUMO2. The subunit linkers could not be observed and no unique or large discernible contacts between the SUMO2 subunits were observed. At present we are unsuccessful in obtaining crystals of the RNF4-SIMs domain–tetra-SUMO2 complex.

SAXS was employed to characterize the solution conformation of tetra-SUMO2 and its complex with the RNF4-SIMs domain (Figure 2C). The SAXS data of the tetra-SUMO2–RNF4-SIMs domain complex (black) presents a parabolic shape in the Kratky plot (Figure 2D) and converges to a plateau at the high q^4 value in the Porod plot (Figure 2E), indicating that it exists as an ordered structure. The SAXS plots for tetra-SUMO2 (blue) are between the two extremes. Whereas the Kratky plot of tetra-SUMO2 is a parabolic shape, the Porod–Debye plot did not converge to the Porod plateau (Figure 2E), suggesting that the tetra-SUMO alone is partially ordered. This is expected because it contains four structured SUMO2 molecules that are linked together in a beads-on-a-string fashion. Using the merged SAXS scattering profile of tetra-SUMO2 for data acquired at two different concentrations, we built linkers in the crystal structure using the ensemble optimization method [28] (Supplementary Figure S4). The average structure of tetra-SUMO2 thus calculated presented a radius of gyration of 36.7 Å and was similar to that of a probable structure (with distance between individual SUMO2s comparable with that expected for the linker lengths) traced out from the crystal packing.

To further explore the structure of the RNF4-SIMs domain–tetra-SUMO2 complex we performed a knowledge-based HADDOCK modelling, as detailed in the Experimental section. This model was based on the free SUMO2 structure, sparse NMR data acquired from the individual peptide-binding experiment, binding orientation as determined by PRE and fitting the experimental SAXS profile with a reliable χ value of 0.92 (Figures 6B and 6C). The spacing between tandem SIMs and the orientation of SIM-binding sites on SUMO2 dictated the arrangement of the complex (Figures 6D and 6E). Several notable features can be discerned from the model. First, in contrast with the previous belief that the spacing of SIM3 and SIM4 is too close to allow interaction with two adjacent SUMOs [17], we found that, with the proper orientation, both SIM3 and SIM4 can simultaneously bind to two adjacent SUMOs. Secondly, the SUMO2 protomers of the tetra-SUMO were observed to bind the four SIM regions of the RNF4-SIMs domain in a zigzag arrangement, forming a compact superhelical turn. Thirdly, the SIM domain bound antiparallel to the tetra-SUMO2 chain such that SIM4 bound to the first SUMO2 (SUMO2a). Fourthly, the poly-arginine segment immediately following SIM4 was positioned beside a large negatively charged surface on SUMO2a, composed of Glu^{13}, Asp^{16}, Asp^{63}, Glu^{77}, Glu^{79}, Asp^{80}, Glu^{81}, Asp^{82} and Asp^{85} (Figures 6F and 6G). Residues Glu^{79}-Asp^{82} and Asp^{85} were conserved among all SUMOs, hence the negatively charged surface feature. Although the precise structure of the poly-arginine segment could not be modelled here, the poly-arginine segment was shown to possess substantial order and rigidity, as demonstrated by the secondary structure prediction and NMR relaxation. In this arrangement the arginine residues, particularly Arg^{81} and Arg^{82}, could readily interact with the negatively charged groups in SUMO2, thereby increasing the binding affinity of SIM4 towards the poly-SUMO (Figure 5B). Experiments to test the validity of the HADDOCK model are ongoing in our laboratory.

DISCUSSION

Intrinsic disorder confers flexibility in RNF4-SIMs domain–poly-SUMO interaction

The recognition of ubiquitylated proteins was mediated by conserved modules which includes the UIM (ubiquitin-interacting

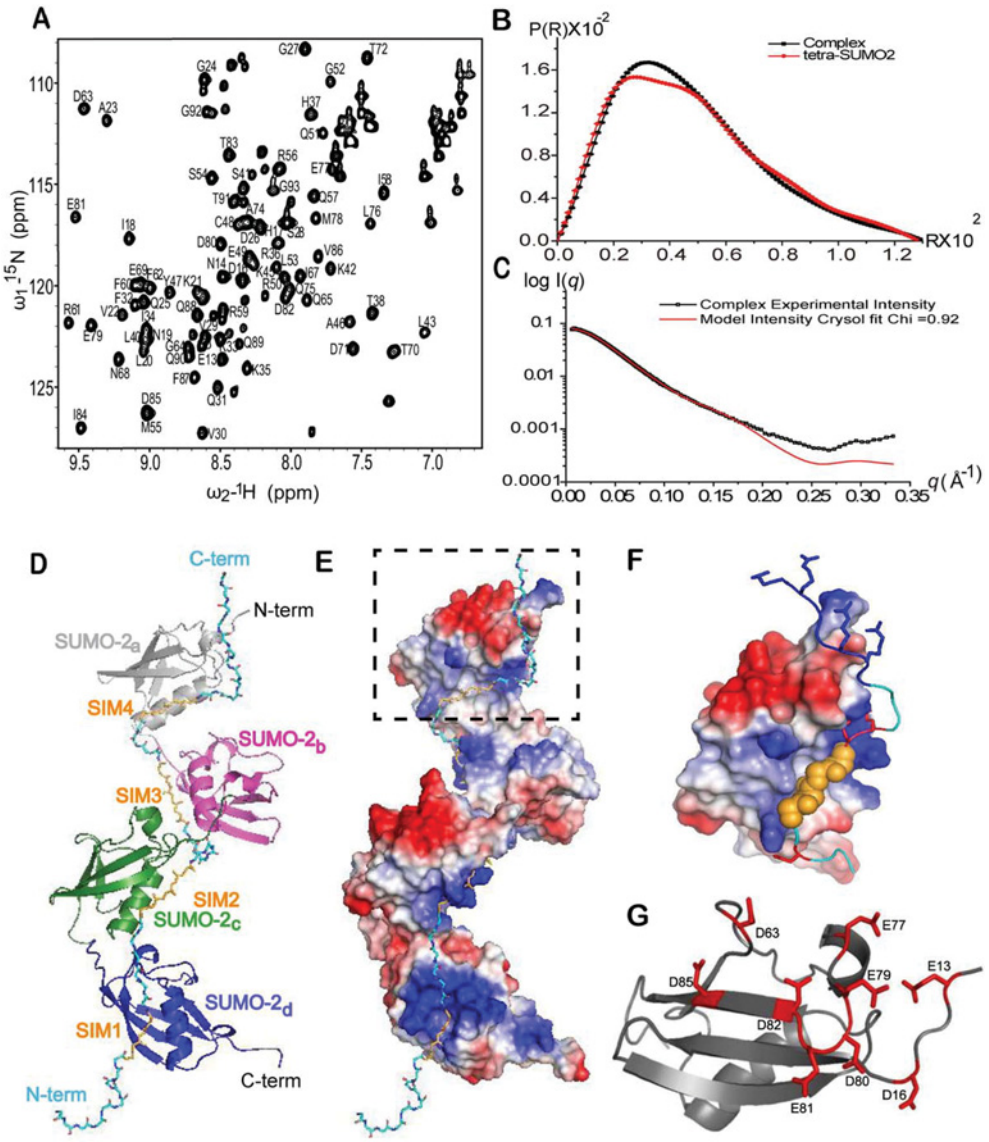

Figure 6 Structural analysis of tetra-SUMO2 and its complex with the RNF4-SIMs domain

(**A**) ^{15}N-HSQC spectrum of the free tetra-SUMO2. The spectrum is practically identical with that of the free SUMO2 and all resonances have been assigned. (**B**) Distance distribution function, *P*(r), of tetra-SUMO2 (red) and its complex with the RNF4-SIMs domain (black). (**C**) SAXS scattering profile of RNF4-SIMs domain–tetra-SUMO2 complex (black). The red curve represents the profile calculated by CRYSOL using the HADDOCK model of the complex, yielding a Chi value of 0.92. (**D**) Knowledge-based HADDOCK model of the RNF4-SIMs domain–tetra-SUMO2 complex, showing the antiparallel orientation of the two molecular chains. The four SUMO2 protomers are colour-coded and labelled as SUMO2$_a$–SUMO2$_d$ for the first to fourth protomer. The RNF4-SIMs domain peptide is coloured cyan and the hydrophobic residues in the SIM motifs are coloured orange. (**E**) Surface charge representation of the RNF4-SIMs domain–tetra-SUMO2 complex. The RNF4-SIMs domain is presented in stick render in cyan and the hydrophobic residues in the SIMs motifs are coloured orange. (**F**) Magnified view of the SUMO2$_a$ region, broken box region in (**E**), showing the interaction of SUMO2$_a$ with a SIM4 peptide. The hydrophobic residues in the SIM4 motif are shown as orange spheres. The three acidic residues (red sticks) in the SIM4 interact with the basic residues in SUMO2$_a$. The C-terminal poly-arginine residues (blue sticks) are located near the large acidic patch (red) of SUMO2$_a$. (**G**) The acidic residues (red) that form the large patch of acidic surface on SUMO2.

motif) [40,41], UBA (ubiquitin-associated domain) [42,43], UEV (ubiquitin E2 variant domain) [44,45], NZF (np14 zinc finger domain) [46] and a CUE (coupling of ubiquitin conjugation to endoplasmic reticulum degradation) domain [47]. Each of these motifs can bind to a variety of ubiquitinated proteins resulting in the modulation of diverse biological functions. UIMs consist of 20 residue sequences that typically adopt a single α-helix structure in solution and form a helical bundle in crystals [40,48]. UIMs also exist in tandem repeats that facilitate binding towards polyubiquiylated proteins. The binding of tandem UIMs resulted in substantial structural changes. For example, the two UIMs of RAP80 (receptor-associated protein 80) merged into a long helix when bound to Lys^{63}-linked diubiquitin [49], whereas in its unbound form the structure is considerably more compact with hydrophobic contacts between the two helices [50]. Thus a tandem UIM is designed to recognize specific polyubiquitin-linked proteins through various lysine residues [51,52]. In comparison, our extensive analysis of NMR and SAXS data clearly demonstrated that the RNF4-SIMs domain is flexible and disordered with little secondary structure propensity. Numerous proteins have been discovered to be intrinsically disordered, or to possess long stretches of disordered segments [53,54]. Intrinsic

disorder plays a crucial role in many biological functions. Among them, biological regulation is particularly noteworthy because its inclusion increases binding affinity and enhances binding allostery, thereby permitting more precise regulation of biological activity [55,56]. The results suggested that intrinsic disorder may also play a specialized functional role in the RNF4-SIMs domain–poly-SUMO interaction. The extended conformation and intrinsic disorder exposed all four SIMs and increased the collision radius with poly-SUMO molecules, thus enhancing their interaction. This was particularly beneficial for RNF4 because it functions as a dimer and the flexibility conferred by the SIMs domain facilitates the binding of either chain to the poly-SUMO [12,13,57–60]. The flexibility conferred by disorder also facilitates conformational optimization, leading to the formation of the final tight complex structure. We speculated that flexibility in the SIMs domain also facilitates a 'zipping-up' or sliding process during the recognition between clustered SIMs and multiple SUMO moieties before adopting a stable configuration, as proposed by Sun and Hunter [15]. Apart from poly-SUMO interaction, clustered SIMs may target other forms of SUMOylation, such as a protein or protein complex that is monoSUMOylated at multiple sites. The avidity effect from multiple SIM–SUMO contacts may thus provide sufficient affinity and specificity for the formation of SUMOylation-regulated protein complexes, which would be distinct from those involving a singular SIM and secondary SUMO-independent physical contact [15]. In this context, flexibility in the SIMs cluster is essential for recognizing various structures in various combinations of multiple-site SUMOylation. Human RNF111/Arkadia is a new STUbL that uses three adjacent SIMs (VVVI, VEIV and VVDL) for specific recognition of poly-SUMO2/3 chains and Ubc13–Mms2 as a cognate E2 enzyme to promote the non-proteolytic Lys^{63}-linked ubiquitylation of SUMOylated target proteins [11]. The spacing between adjacent SIMs in RNF111/Arkadia is considerably longer: 22 and 52 residues between SIM1/2 and SIM 2/3 respectively. In comparison, the spacings between SIM1/2, SIM2/3 and SIM3/4 are 6, 7 and 4 respectively in RNF4. The free RNF111/Arkadia SIMs domain is also likely to be disordered and a similar poly-SUMO binding mechanism is employed.

SUMO paralogue selectivity

Several previous studies have addressed the question of SUMO paralogue selectivity [7,12,15,17]. SIM1, SIM2 and SIM3 indicated a significant binding affinity for SUMO2 over SUMO1, by a factor of 1.85, 4.25 and 1.56 respectively. By contrast, SIM4 indicated no paralogue preference. This is consistent with its likely binding to SUMO1, which could serve as the end cap in a poly-SUMO chain [3]. Viewed together, and assuming that the four SIMs bound to tetra-SUMO synergistically, the RNF4-SIMs domain had a paralogue preference of SUMO2 over SUMO1 by a factor of 11. Thus the RNF4-SIMs domain was tailored to bind to poly-SUMO2 preferentially, similar to that of Arkadia/RNF111 *in vivo* [11,61,62]. However, no paralogue binding preference was observed in the gel pull-down assay in Arkadia/RNF111. This preferential binding is apparently unimportant in the poly-SUMO-mediated proteasome degradation in RNF4 because SUMO1 only serves as the end cap and the RNF4-SIMs domain can still bind to poly-SUMO2 tightly enough for efficient ubiquitylation [12]. Although whether poly-SIMs can bind to substrates modified with SUMOs at multiple distinct sites is unclear, we envisioned that in such a case the SUMO paralogue selectivity would be more crucial because SUMO1 is expected to be more equally populated than SUMO2/3.

Spacing between consecutive SIMs does not impose steric hindrance

An approximately 20-amino-acid spacing between two tandem SIMs, such as that from SIM1 to SIM2 in Rfp1, Rfp2 and C5orf25 (chromosome 5 ORF 25), as well as SIM1 to SIM3 and SIM2 to SIM4 in RNF4, may contribute to the optimal recognition of poly-SUMO chains, similar to the tandem UIM for polyubiquitin chains [15,50,63]. Two adjacent SUMO units were speculated to form physical contacts with either the SIM1/3 or the SIM2/4 pair in a stable RNF4–poly-SUMO chain complex [15]. To test this, we used NMR to monitor the poly-SUMO2 binding of each SIM in the RNF4-SIMs domain (Figure 5). If the hypothesis is correct, then resonance of residues from SIM1 and SIM3, and that of SIM2 and SIM4, would be affected simultaneously as pairs. Inconsistent with the hypothesis, we determined that binding the RNF4–SIMs domain to di-, tri- and tetra-SUMO2 consistently followed a trend of SIM2$\geqslant$SIM3$>$SIM4$\gg$SIM1, with SIM2–SIM4 forming a group and SIM1 being a considerably weak binder (Figure 5B). This observation appears consistent with a model wherein all four SIMs bind to the poly-SUMO2 as a whole.

To further corroborate the observation we performed a knowledge-based HADDOCK modelling of the RNF4-SIMs domain–poly-SUMO2 complex. The model showed that the spacing between tandem SIMs and the orientation of SIM-binding sites on SUMO2 dictated the arrangement of the complex (Figures 6D and 6E). Importantly, in agreement with the NMR data we found that both SIM3 and SIM4 can simultaneously bind to two adjacent SUMOs. Thus the linker spacing between adjacent SIMs did not necessarily prohibit the simultaneous binding of four SIMs on to the tetra-SUMO chain. However, this did not indicate that the spacing in the model was optimal for interaction. Moreover, we could not exclude the possible existence of alternative complex structures. In particular, because RNF4 exists as a dimer *in vivo* and the two N-tails of RNF4 are considerably close in spacing, a poly-SUMO2 chain can conceivably be bound by different SIMs forming two neighbouring chains, as previously proposed [15,17,58,60]. The functional structure of the RNF4-SIMs domain–poly-SUMO complex is likely to be more complex than that shown on the HADDOCK model (Figure 6D).

Conclusion

Deciphering their complex structure is a major step forward in unravelling the molecular basis of RNF4-SIMs domain interaction with poly-SUMO. By using NMR, SAXS and X-ray crystallography, we demonstrated that, contrary to bioinformatics predictions, the SIMs domain is largely unstructured. Likewise, linkers between the adjacent SUMO molecules in the poly-SUMO chain are also disordered and the chain behaves similar to 'beads on a string'. The isolated SIM peptides show SUMO-binding affinities in increasing order, as SIM2$\geqslant$ SIM3$>>$SIM1$>$SIM4, with SIM2–SIM4 exhibit preferential binding in an orientation parallel to the β2-strand. The binding affinity preference was also preserved in the long construct containing all four SIMs in the RNF4-SIMs domain, except that the binding affinity of SIM4 was substantially enhanced to a level comparable with that of SIM2 and SIM3. The results are consistent with the notion that SIM2, SIM3 and SIM4 are necessary and sufficient for interaction with the poly-SUMO chain [12,15,17]. Individual SIMs of the RNF4-SIMs domain recognize hydrophobic and polar residues of SUMO2 in the groove between β2-strands and α1-helices parallel to the β2-strands. HADDOCK modelling demonstrated that the spacing between adjacent SIMs did not necessarily prohibit the simultaneous binding of four SIMs on to the tetra-SUMO

chain. The RNF4-SIMs domain binds in antiparallel orientation to the poly-SUMO chain, suggesting a potential role of the poly-arginine track immediately following the SIM4 sequence.

AUTHOR CONTRIBUTION

Camy Kung, Che-Chang Chang and Szu-Huan Wang prepared the protein samples. Camy Kung and Mandar Naik performed the NMR experiments and analysed the data under the supervision of Tai-Huang Huang and Li-Ying Lin. Chia-Lin Chen crystallized poly-SUMO and solved its structure under the guidance of Che Ma. Camy Kung and Chi-Fon Chang. performed the SAXS experiments and analysed the results data. Mandar Naik, Hsiu-Ming Shih and Tai-Huang Huang conceived the project and planned the experiments. Camy Kung, Mandar Naik and Tai-Huang Huang wrote the paper.

ACKNOWLEDGEMENTS

The NMR experiments were conducted on NMR spectrometers at the High-Field Nuclear Magnetic Resonance Center (HFNMRC), supported by the National Research Program for Biopharmaceuticals, the National Science Council of the Republic of China. We acknowledge the use of Academia Sinica Biophysics Core Facility and the National Synchrotron Radiation Research Center, Hsinchu, Taiwan and the SPring-8 in Japan. We thank Dr U-Ser Jeng and Dr Shu-Chuan Jao for their help in data collection and experimental setup.

FUNDING

This work was supported by the Ministry of Science and Technology, Republic of China [grant numbers NSC101-2311-B-001-025 and NSC102-2113-M-001-010 (to T.-H.H.)].

REFERENCES

1 Gareau, J. R. and Lima, C. D. (2010) The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat. Rev. Mol. Cell Biol. **11**, 861–871 CrossRef PubMed
2 Hay, R. T. (2013) Decoding the SUMO signal. Biochem. Soc. Trans. **41**, 463–473 CrossRef PubMed
3 Matic, I., van Hagen, M., Schimmel, J., Macek, B., Ogg, S. C., Tatham, M. H., Hay, R. T., Lamond, A. I., Mann, M. and Vertegaal, A. C. O. (2007) *In vivo* identification of human SUMO polymerization sites by high accuracy mass spectrometry and an *in vitro* to *in vivo* strategy. Mol. Cell. Proteomics **7**, 132–144 CrossRef PubMed
4 Cheng, C. H., Lo, Y. H., Liang, S. S., Ti, S. C., Lin, F. M., Yeh, C. H., Huang, H. Y. and Wang, T. F. (2006) SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of *Saccharomyces cerevisiae*. Genes Dev. **20**, 2067–2081 CrossRef PubMed
5 Schwartz, D. C., Felberbaum, R. and Hochstrasser, M. (2007) The Ulp2 SUMO protease is required for cell division following termination of the DNA damage checkpoint. Mol. Cell. Biol. **27**, 6948–6961 CrossRef PubMed
6 Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G. and Chen, Y. (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proc. Natl. Acad. Sci. U.S.A. **101**, 14373–14378 CrossRef PubMed
7 Chang, C. C., Naik, M. T., Huang, Y. S., Jeng, J. C., Liao, P. H., Kuo, H. Y., Ho, C. C., Hsieh, Y. L., Lin, C. H., Huang, N. J. et al. (2011) Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. Mol. Cell **42**, 62–74 CrossRef PubMed
8 Escobar-Cabrera, E., Okon, M., Lau, D. K., Dart, C. F., Bonvin, A. M. and McIntosh, L. P. (2011) Characterizing the N- and C-terminal small ubiquitin-like modifier (SUMO)-interacting motifs of the scaffold protein DAXX. J. Biol. Chem. **286**, 19816–19829 CrossRef PubMed
9 Sekiyama, N., Ikegami, T., Yamane, T., Ikeguchi, M., Uchimura, Y., Baba, D., Ariyoshi, M., Tochio, H., Saitoh, H. and Shirakawa, M. (2008) Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. J. Biol. Chem. **283**, 35966–35975 CrossRef PubMed
10 Ullmann, R., Chien, Christopher D., Avantaggiati, Maria L. and Muller, S. (2012) An acetylation switch regulates SUMO-dependent protein interaction networks. Mol. Cell **46**, 759–770 CrossRef PubMed
11 Poulsen, S. L., Hansen, R. K., Wagner, S. A., van Cuijk, L., van Belle, G. J., Streicher, W., Wikstrom, M., Choudhary, C., Houtsmuller, A. B., Marteijn, J. A. et al. (2013) RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. J. Cell Biol. **201**, 797–807 CrossRef PubMed
12 Tatham, M. H., Geoffroy, M. C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E. G., Palvimo, J. J. and Hay, R. T. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nat. Cell Biol. **10**, 538–546 CrossRef PubMed
13 Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. and Hay, R. T. (2012) Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. Nature **489**, 115–120 CrossRef PubMed
14 Bruderer, R., Tatham, M. H., Plechanovova, A., Matic, I., Garg, A. K. and Hay, R. T. (2011) Purification and identification of endogenous polySUMO conjugates. EMBO Rep. **12**, 142–148 CrossRef PubMed
15 Sun, H. Y. and Hunter, T. (2012) Poly-small ubiquitin-like modifier (polySUMO)-binding proteins identified through a string search. J. Biol. Chem. **287**, 42071–42083 CrossRef PubMed
16 Lallemand-Breitenbach, V., Jeanne, M., Benhenda, S., Nasr, R., Lei, M., Peres, L., Zhou, J., Zhu, J., Raught, B. and de The, H. (2008) Arsenic degrades PML or PML-RARα through a SUMO-triggered RNF4/ubiquitin-mediated pathway. Nat. Cell Biol. **10**, 547–555 CrossRef PubMed
17 Keusekotten, K., Bade, V. N., Meyer-Teschendorf, K., Sriramachandran, A. M., Fischer-Schrader, K., Krause, A., Horst, C., Schwarz, G., Hofmann, K., Dohmen, R. J. and Praefcke, G. J. K. (2013) Multivalent interactions of the SUMO-interaction motifs in the RING-finger protein 4 (RNF4) determine the specificity for chains of the small ubiquitin-related modifier (SUMO). Biochem. J. **457**, 207–214 CrossRef
18 Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins **59**, 687–696 CrossRef PubMed
19 Van Horn, W. D., Beel, A. J., Kang, C. and Sanders, C. R. (2010) The impact of window functions on NMR-based paramagnetic relaxation enhancement measurements in membrane proteins. Biochim. Biophys. Acta **1798**, 140–149 CrossRef PubMed
20 Iwahara, J., Schwieters, C. D. and Clore, G. M. (2004) Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. J. Am. Chem. Soc. **126**, 5879–5896 CrossRef PubMed
21 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Cryst. **40**, 658–674 CrossRef
22 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**, 2126–2132 CrossRef
23 Jeng, U., Su, C. H., Su, C. J., Liao, K. F., Chuang, W. T., Lai, Y. H., Chang, J. W., Chen, Y. J., Huang, Y.-S., Lee, M.-T. et al. (2010) A small/wide-angle X-ray scattering instrument for structural characterization of air-liquid interfaces, thin films and bulk specimens. J. Appl. Crystallogr. **43**, 110–121 CrossRef
24 Svergun, D. I., Barberato, C. and Koch, M. H. J. (1995) CRYSOL: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates J. Appl. Crystallogr. **28**, 768–773 CrossRef
25 Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. and Svergun, D. I. (2003) PRIMUS: a Windows-PC based system for small-angle scattering data analysis. J. Appl. Crystallogr. **36**, 1277–1282 CrossRef
26 Svergun, D. I., Petoukhov, M. V. and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. **80**, 2946–2953 CrossRef PubMed
27 Volkov, V. V. and Svergun, D. I. (2003) Uniqueness of *ab-initio* shape determination in small-angle scattering. J. Appl. Crystallogr. **36**, 860–864 CrossRef
28 Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. **129**, 5656–5664 CrossRef PubMed
29 Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. **25**, 495–503 CrossRef
30 Konarev, P. V., Petoukhov, M. V., Volkov, V. V. and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. J. Appl. Crystallogr. **39**, 277–286 CrossRef
31 Dominguez, C., Boelens, R. and Bonvin, A. M. (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. **125**, 1731–1737 CrossRef PubMed
32 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 CrossRef PubMed
33 van Dijk, A. D. J., Fushman, D. and Bonvin, A. M. J. J. (2005) Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked Di-ubiquitin and validation against N-15-relaxation data. Proteins **60**, 367–381 CrossRef PubMed
34 Karaca, E. and Bonvin, A. M. J. J. (2013) On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. Acta Crystallogr. D. Biol. Crystallogr. **69**, 683–694 CrossRef PubMed

35 Wishart, D. S. and Case, D. A. (2001) Use of chemical shifts in macromolecular structure determination. Methods Enzymol. **338**, 3–34 CrossRef PubMed

36 Tamiola, K. and Mulder, F. A. A. (2012) Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem. Soc. Trans. **40**, 1014–1020 CrossRef PubMed

37 Lipfert, J. and Doniach, S. (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu. Rev. Biophys. Biomol. Struct. **36**, 307–327 CrossRef PubMed

38 Rambo, R. P. and Tainer, J. A. (2013) Super-resolution in solution X-Ray scattering and its applications to structural systems biology. Annu. Rev. Biophys. **42**, 415–441 CrossRef PubMed

39 Rambo, R. P. and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod–Debye law. Biopolymers **95**, 559–571 CrossRef PubMed

40 Hofmann, K. and Falquet, L. (2001) A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. Trends Biochem. Sci. **26**, 347–350 CrossRef PubMed

41 Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M. and Rechsteiner, M. (1998) Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. J. Biol. Chem. **273**, 5461–5467 CrossRef PubMed

42 Bertolaet, B. L., Clarke, D. J., Wolff, M., Watson, M. H., Henze, M., Divita, G. and Reed, S. I. (2001) UBA domains of DNA damage-inducible proteins interact with ubiquitin. Nat. Struct. Biol. **8**, 417–422 CrossRef PubMed

43 Wilkinson, C. R. M., Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C. and Gordon, C. (2001) Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. Nat. Cell Biol. **3**, 939–943 CrossRef PubMed

44 VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M. and Wolberger, C. (2001) Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. Cell **105**, 711–720 CrossRef PubMed

45 Koonin, E. V. and Abagyan, R. A. (1997) TSG101 may be the prototype of a class of dominant negative ubiquitin regulators. Nat. Genet. **16**, 330–331 CrossRef PubMed

46 Meyer, H. H., Wang, Y. and Warren, G. (2002) Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. EMBO J. **21**, 5645–5652 CrossRef PubMed

47 Ponting, C. P. (2000) Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. Biochem. J. **351**, 527–535 CrossRef PubMed

48 Fisher, R. D., Wang, B., Alam, S. L., Higginson, D. S., Robinson, H., Sundquist, W. I. and Hill, C. P. (2003) Structure and ubiquitin binding of the ubiquitin-interacting motif. J. Biol. Chem. **278**, 28976–28984 CrossRef PubMed

49 Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A. and Fukai, S. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. EMBO J. **28**, 2461–2468 CrossRef PubMed

50 Song, A. X., Zhou, C. J., Peng, Y., Gao, X. C., Zhou, Z. R., Fu, Q. S., Hong, J., Lin, D. H. and Hu, H. Y. (2010) Structural transformation of the tandem ubiquitin-interacting motifs in ataxin-3 and their cooperative interactions with ubiquitin chains. PLoS ONE **5**, e13202 CrossRef PubMed

51 Kulathu, Y. and Komander, D. (2012) Atypical ubiquitylation: the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. Nat. Rev. Mol. Cell Biol. **13**, 508–523 CrossRef PubMed

52 Keusekotten, K., Elliott, P. R., Glockner, L., Fiil, B. K., Damgaard, R. B., Kulathu, Y., Wauer, T., Hospenthal, M. K., Gyrd-Hansen, M., Krappmann, D. et al. (2013) OTULIN Antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. Cell **153**, 1312–1326 CrossRef PubMed

53 Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. **293**, 321–331 CrossRef PubMed

54 Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C. and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac. Symp. Biocomput. 473–484

55 Dyson, H. J. (2011) Expanding the proteome: disordered and alternatively folded proteins. Q. Rev. Biophys. **44**, 467–518 CrossRef PubMed

56 Hilser, V. J. and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc. Natl. Acad. Sci. U.S.A. **104**, 8311–8315 CrossRef PubMed

57 Rojas-Fernandez, A., Plechanovova, A., Hattersley, N., Jaffray, E., Tatham, M. H. and Hay, R. T. (2014) SUMO chain-induced dimerization activates RNF4. Mol. Cell **53**, 880–892 CrossRef PubMed

58 Liew, C. W., Sun, H. Y., Hunter, T. and Day, C. L. (2010) RING domain dimerization is essential for RNF4 function. Biochem. J. **431**, 23–29 CrossRef PubMed

59 Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. and Huang, D. T. (2012) BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. Nat. Struct. Mol. Biol. **19**, 876–883 CrossRef PubMed

60 Plechanovova, A., Jaffray, E. G., McMahon, S. A., Johnson, K. A., Navratilova, I., Naismith, J. H. and Hay, R. T. (2011) Mechanism of ubiquitylation by dimeric RING ligase RNF4. Nat. Struct. Mol. Biol. **18**, 1052–1059 CrossRef PubMed

61 Matic, I., van Hagen, M., Schimmel, J., Macek, B., Ogg, S. C., Tatham, M. H., Hay, R. T., Lamond, A. I., Mann, M. and Vertegaal, A. C. O. (2008) *In vivo* identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an *in vitro* to *in vivo* strategy. Mol. Cell. Proteomics **7**, 132–144 CrossRef

62 Tatham, M. H., Jaffray, E., Vaughan, O. A., Desterro, J. M. P., Botting, C. H., Naismith, J. H. and Hay, R. T. (2001) Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. J. Biol. Chem. **276**, 35368–35374 CrossRef PubMed

63 Sims, J. J. and Cohen, R. E. (2009) Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. Mol. Cell **33**, 775–783 CrossRef PubMed

Received 22 April 2014/14 May 2014; accepted 20 May 2014
Published as BJ Immediate Publication 20 May 2014, doi:10.1042/BJ20140521

Biochem. J. (2014) **462**, 53–65 (Printed in Great Britain) doi:10.1042/BJ20140521

online data

SUPPLEMENTARY ONLINE DATA

Structural analysis of poly-SUMO chain recognition by the RNF4-SIMs domain

Camy C.-H. KUNG*†‡, Mandar T. NAIK*[1], Szu-Huan WANG§, Hsiu-Ming SHIH*, Che-Chang CHANG‖, Li-Ying LIN¶, Chia-Lin CHEN†¶, Che MA¶, Chi-Fon CHANG¶ and Tai-Huang HUANG*¶**[1]

*Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan
†Chemical Biology and Molecular Biophysics, Taiwan International Graduate Program, Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan
‡Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan
§Scientific Instrument Center, Academia Sinica, Taipei 11529, Taiwan
‖Graduate Institute of Translational Medicine, Taipei Medical University, Taipei 110, Taiwan
¶Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan
**Department of Physics, National Taiwan Normal University, Taipei 11677, Taiwan

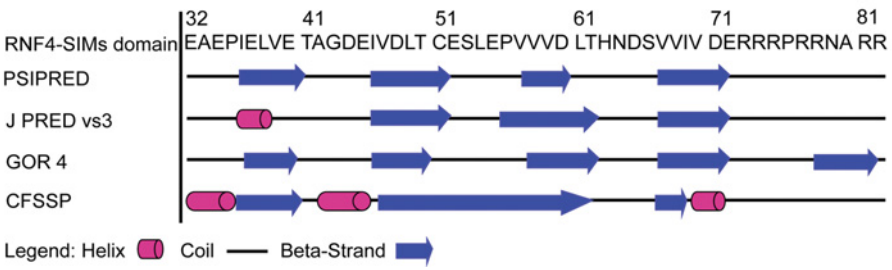

Figure S1 Secondary structure prediction of RNF4-SIMs domain

Residues spanning from Glu[32] to Arg[82] were predicted by software programs PSIPRED, Jpred, GOR 4 and CFSSP [1–5].

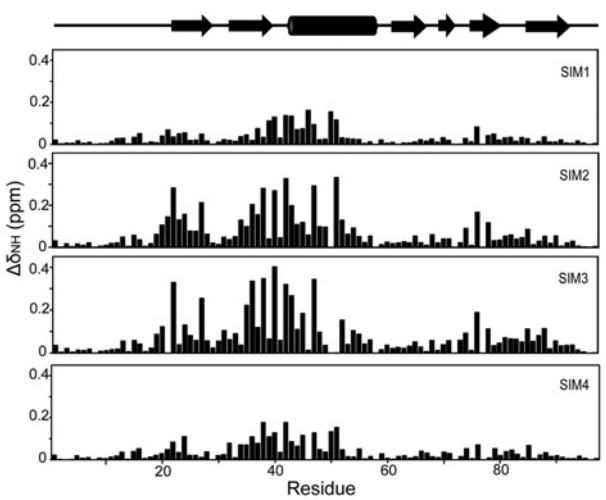

Figure S2 Sequence variation of the composite amide resonance chemical shift of SUMO-1 perturbed by the addition of individual RNF4-SIMs

[1] Correspondence may be addressed to either of these authors (email mandarn@ibms.sinica.edu.tw or bmthh@ibms.sinica.edu.tw).
Co-ordinates for the tetra-SUMO2 structure have been deposited in the PDB under accession code 4NPN.

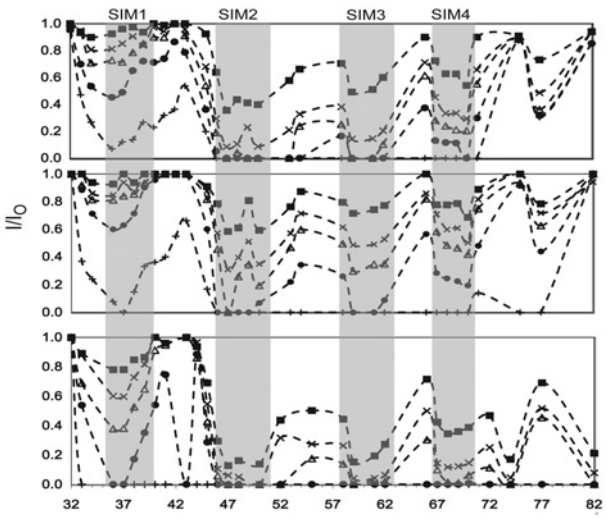

Figure S3 Sequence variation of the intensity ratios (*I*/*Io*) for amide resonances in ^{15}N-labelled RNF4-SIMs domain with (*I*) and without (Io) the addition of di- (top), tri- (middle) or tetra- (bottom) SUMO2

poly-SUMO2/RNF4-SIMs domain ratios were 1:7 (■), 1:3 (×), 1:2 (△), 1:1 (●) and 3:1 (+).

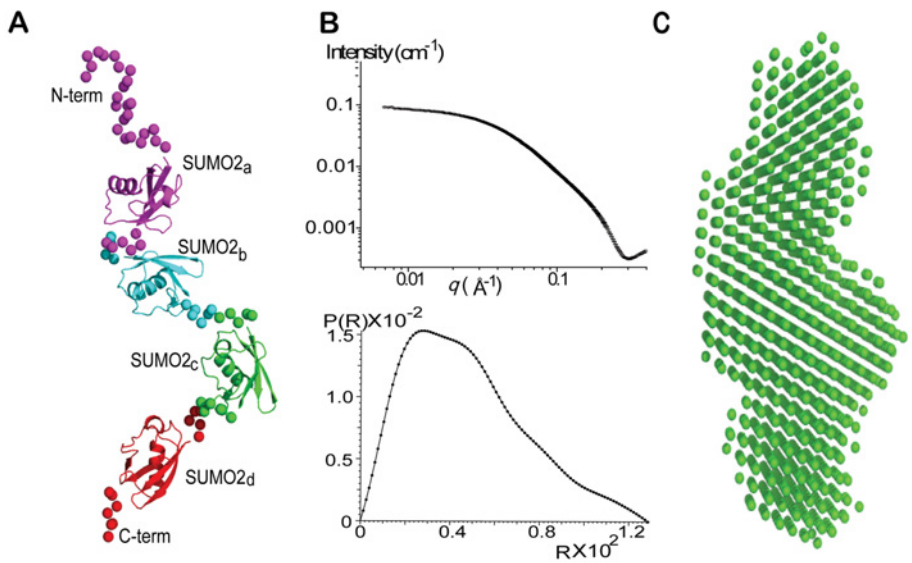

Figure S4 Crystal structure and SAXS analyses of the free tetra-SUMO2 chain

(**A**) Ribbon representation of the crystal structure of tetra-SUMO2 with linkers modelled using SAXS data (spheres). (**B**) Merged SAXS scattering profile (3.0 and 28 ng/ml) (top) and distance distribution function (bottom) of tetra-SUMO2 in solution. (**C**) CRYSOL model of tetra-SUMO2 deduced from SAXS scattering profile.

Table S1 SAXS analysis of the RNF4-SIMs domain, tetra-SUMO2 and a complex of the two

EMBL, European Molecular Biology Laboratory; N/A, not appliable; NSRRC, National Synchrotron Radiation Research Center.

Parameter	RNF4-SIMs domain	Tetra-SUMO2	Tetra-SUMO2–RNF4-SIMs domain complex
Data collection			
Instrument	23A SWAXS end-station of NSRRC		
Beam geometry	0.5 mm pinhole		
Wavelength (Å)	0.8267		
q-range ($Å^{-1}$)	0.007–0.39		
Exposure time (s)	300 s with 60 s per exposure for five successive exposures		
Temperature (K)	293		
Concentration range (mg/ml)	8.0–11.5	3.0–28.0	2.3
Structural parameters at the above concentration ranges			
I(0) (cm^{-1})*	0.035 ± 0.000	0.085 ± 0.000	0.082 ± 0.000
R_g (Å)*	25.50 ± 0.06	38.21 ± 0.07	37.54 ± 0.11
I(0) (cm^{-1})†	0.036 ± 0.001	0.084 ± 0.001	0.081 ± 0.001
R_g (Å)†	25.82 ± 3.916	36.74 ± 2.827	36.42 ± 1.13
D_{max} (Å)	84.55 ± 8	128.57 ± 12	130.46 ± 13
Porod volume estimate ($Å^3$)	9072 ± 1814	59040 ± 11808	63604 ± 12721
Dry volume calculated from sequence ($Å^3$)‡	7826	47889	55693
Molecular mass determination			
Partial specific volume (cm^3/g)	0.73	0.73	0.73
Contrast ($\Delta\rho$ $X10^{-10}$ cm^{-2})	2.90	2.94	2.94
M_r§	6056 ± 937	36807 ± 3117	46293 ± 2539
Calculated monomeric M_r from sequence‡	6468	39578	46027
Software employed			
Primary data reduction	NSRRC 23A SWAXS package		
Data processing	PRIMUS v3.2 and SAS DATA analysis, EMBL		
Ab initio analysis	GASBOR	GASBOR	N/A
Validation and averaging	DAMAVER	DAMAVER	N/A
Rigid-body modelling	N/A	N/A	HADDOCK
Atomic modelling	N/A	EOM	N/A
Computation of model intensities	N/A	N/A	CRYSOL
3D graphics representation	PyMOL		

*Calculated using the distance distribution function, *P*(r).

†Calculated using the Guinier equation.

‡Data reported from Peptide Property Calculator Version 1.00 (Chazan), Center for Biotechnology, Northwestern University Evanston, IL 60201. U.S.A.

§Calculated using the *I*(0).

Table S2 Data collection and refinement statistics for X-ray crystallography of tetra-SUMO2

Values in parentheses are for highest-resolution shell.

Parameter	Tetra-SUMO2
Data collection	
Space group	*R*3
Cell dimensions	
a, *b*, *c* (Å)	75.123, 75.123, 33.116
α, β, γ (°)	90, 90, 120
Resolution (Å)	30.0–1.63 (1.69–1.63)
R_{sym} or R_{merge}	4.8 (35.7)
R_{pim}	1.9 (13.8)
$I/\sigma I$	25.52 (6.22)
Completeness (%)	99.23 (97.32)
Redundancy	7.5 (7.6)
$CC_{1/2}$	0.999 (0.958)
Refinement	
Resolution (Å)	23.21–1.63
Number of reflections	8595
R_{work}/R_{free} (%)	19.81/21.83
Number of atoms	
Protein	575
Water	29
B-factors	
Protein	25.80
Water	28.20
RMSD	
Bond lengths (Å)	0.006
Bond angles (°)	1.087

REFERENCES

1 Garnier, J., Gibrat, J. F. and Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. **266**, 540–553 CrossRef PubMed

2 Cole, C., Barber, J. D. and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res. **36**, W197–W201 CrossRef PubMed

3 McGuffin, L. J., Bryson, K. and Jones, D. T. (2000) The PSIPRED protein structure prediction server. Bioinformatics **16**, 404–405 CrossRef PubMed

4 Chou, P. Y. and Fasman, G. D. (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry **13**, 211–222 CrossRef PubMed

5 Chou, P. Y. and Fasman, G. D. (1974) Prediction of protein conformation. Biochmistry **13**, 222–245 CrossRef

Received 22 April 2014/14 May 2014; accepted 20 May 2014
Published as BJ Immediate Publication 20 May 2014, doi:10.1042/BJ20140521