Structure of the Small Ubiquitin-like Modifier (SUMO)-interacting Motif of MBD1-containing Chromatin-associated Factor 1 Bound to SUMO-3*†‡

Referred for publication, April 1, 2008, and in revised form, September 2, 2008 Published, JBC Papers in Press, October 7, 2008, DOI 10.1074/jbc.M802528200

Naotaka Sekiyma, Takahisa Ikegami, Tsutomu Yamane, Mitsunori Ikeguchi, Yasuhiro Uchimura, Daichi Baba, Mariko Ariyoshi, Hidehito Tochio, Hisato Saitoh, and Masahiro Shirakawa

From the 1Graduate School of Engineering, Kyoto University, Kyoto 615-8510, the 2Graduate School of Integrated Science, Yokohama City University, Yokohama 230-0045, the 3Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto 860-0811, the 4Graduate School of Science and Technology, Kumamoto University, Kumamoto 860-8555, 5CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, and the 6Genome Science Center, RIKEN, Tsurumi, Yokohama 230-0045, Japan

Post-translational modification by small ubiquitin-like modifier (SUMO) proteins has been implicated in the regulation of a variety of cellular events. The functions of sumoylation are often mediated by downstream effector proteins harboring SUMO-interacting motifs (SIMs) that are composed of a hydrophobic core and a stretch of acidic residues. MBD1-containing chromatin-associated factor 1 (MCAF1), a transcription repressor, interacts with SUMO-2/3 and SUMO-1, with a preference for SUMO-2/3. We used NMR spectroscopy to solve the solution structure of the SIM of MCAF1 bound to SUMO-3. The hydrophobic core of the SIM forms a parallel β-sheet pairing with strand β2 of SUMO-3, whereas its C-terminal acidic stretch seems to mediate electrostatic interactions with a surface area formed by basic residues of SUMO-3. The significance of these electrostatic interactions was shown by mutations of both SUMO-3 and MCAF1. The present structural and biochemical data suggest that the acidic stretch of the SIM of MCAF1 plays an important role in the binding to SUMO-3.

Small ubiquitin-like modifier (SUMO) proteins conjugate post-translationally with target proteins through a series of enzymatic reactions that resemble ubiquitination (1–5). In contrast to ubiquitination, which is largely involved in regulating the degradation of target proteins by proteasomes or lysosomes (6), sumoylation appears to regulate a wide variety of cellular events, such as nuclear transport, subnuclear localization, transcriptional regulation, DNA repair, and chromosome segregation (3). In particular, sumoylation of transcription factors represses transcription through a variety of different mechanisms, such as recruitment of histone deacetylases and regulation of nuclear body components (7). It has been suggested that sumoylation generally regulates the functions of target proteins by modulating their protein-protein or protein-DNA interactions.

In mammals, four SUMO paralogues, SUMO-1 through SUMO-4, have been identified, of which SUMO-1 to -3 can serve as protein modifiers (8). Whereas SUMO-2 and SUMO-3 share 97% amino acid identity, they have only 48 and 46% identity with SUMO-1, respectively. This indicates that SUMO-2 and SUMO-3 constitute a subgroup that is distinct from SUMO-1. Although these SUMO paralogues largely share common cellular functions, they also show paralogue-specific properties in regard to cellular localization and substrate specificity (9, 10). For example, SUMO-1 is preferentially found within nucleoli, nuclear envelopes, and cytoplasmic foci (8), whereas SUMO-2/3 accrue on chromosomes early in the nuclear reformation process (11). With respect to substrate specificity, RanGAP1 is preferentially modified by SUMO-1, whereas SUMO-2/3 show preference for topoisomerase 2 during mitosis. However, promyelocytic leukemia protein conjugates to all three SUMO paralogues (12). SUMO-1 is most abundant in the protein-conjugated form, whereas SUMO-2/3 isomers are more abundant in a free pool and are available to conjugate with target proteins upon various cellular stresses (13). However, at present, the mechanisms by which these different SUMO paralogues distinguish and conjugate to specific target proteins have yet to be elucidated. Additionally, it is unclear whether functional differences arise from modification by the various SUMO paralogues.

Previous studies have identified a SUMO-interacting motif (SIM) in downstream effector proteins that bind to SUMO (14). For example, the SIM sequence of the DNA helicase Srs2 is thought to mediate the interaction between Srs2 and sumoy-
lated proliferating cell nuclear antigen for suppression of illegitimate recombination events during DNA replication (15). The functional consequences of SUMO binding to SIM-containing proteins are beginning to emerge. The SUMO target protein thymine DNA glycosylase has an SIM-like sequence that interacts with an attached SUMO protein and is proposed to modulate SUMO-directed structural change (16).

The SIM consensus sequence consists of a hydrophobic core, V/I-V/I-X-V/I, and is often associated with a stretch of acidic residues flanking either its N or C terminus (17). An acidic amino acid is frequently present at the third position (X) of the hydrophobic core sequence (17). Structures of the complex formed between the SIM of PIASx and SUMO-1, and of the quaternary complex formed by SUMO-1-conjugated RanGAP1, Ubc9, and RanBP2 (Nup358) have shown that the hydrophobic core of SIM interacts with a hydrophobic groove formed between strand β2 and helix α1 of SUMO-1 (18–20). Thus, the interactions between canonical SIMs and SUMO-1 have been well characterized; however, because the residues that form the hydrophobic cleft between strand β2 and helix α1 are not fully conserved between SUMO-1 and SUMO-2/3, the molecular recognition of SUMO-3 by the canonical SIM has remained elusive, although the structure of SUMO-3-conjugated thymine DNA glycosylase (TDG) has revealed that the SIM-like sequence of TDG interacts with the hydrophobic cleft in a manner similar to that observed for the SUMO-1 SIM complex structures (16, 21). The SIM-like sequence of TDG, VEQQ, is partially similar to the SIM consensus and to the SIMs of PIASxα and RanBP2.

Although the importance of the acidic residues in and near SIM consensus sequences has been shown by mutagenesis (17), from a structural point of view, their roles in SUMO binding are not yet clear. The acidic residues are not well converged (18), or from a structural point of view, their roles in SUMO binding are not yet clear. The acidic residues are not well converged (18), or

To elucidate the mechanism of the molecular recognition of SUMO-2/3 by MCAF1, we used NMR spectroscopy to determine the solution structure of SUMO-3 in complex with the SIM of MCAF1. The structural data in combination with results of mutagenesis experiments provide insights into the molecular basis of binding of MCAF1 to SUMO-3 and also into the role of the acidic residues of SIM in SUMO binding.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—For the preparation of a peptid containing the SIM of MCAF1, a DNA fragment encoding the region of MCAF1, KGKATGSSSGVIDTMDEEASGQD, was amplified by PCR and subcloned into pGEX4T-1 (GE Healthcare). Recombinant SUMO-3 and SIMMCAF1 were each obtained by expressing a GST-tagged protein in *Escherichia coli* BL21(DE3) transformed with SUMO-3/pGEX4T-3 (22) and SIMMCAF1/pGEX4T-1, followed by affinity purification on a glutathione-Sepharose column (GE Healthcare). After GST tag cleavage with thrombin, recombinant SUMO-3 was purified by using Superdex 75 gel-filtration chromatography (GE Healthcare). After GST tag cleavage with thrombin, the recombinant peptide containing the SIM of MCAF1 was purified by reversed-phase high-performance liquid chromatography. The peptide, designated as SIMMCAF1, has an extra sequence, GSPEF, from the vector at its N terminus. Protein/peptide uniformly labeled with $^{13}$C, was obtained by growing the bacteria in minimal medium containing $^{12}$NH$_4$Cl and $^{13}$C$_6$-d$_2$-glucose as the sole nitrogen and carbon sources, respectively. SUMO-3-SIMMCAF1 complex, in which either SUMO-3 or SIMMCAF1 was labeled with $^{13}$C, was dissolved in a buffer containing 20 mM potassium phosphate (pH 6.8), and 9% D$_2$O and used for NMR experiments. These differently labeled samples were generated by titrating an excess amount of the unlabeled component into the $^{13}$C-$^{15}$N-labeled component. The molar ratios for preparation of these complexes were estimated by monitoring the chemical shift changes in the $^1$H-$^1$N HSQC spectra during titration. The final SUMO-3/SIMMCAF1 molar ratios of the samples were 1:1.5 ($^{13}$C-$^{15}$N-labeled SUMO-3 in complex with unlabeled SIMMCAF1) and 1:1.6 (unlabeled SUMO-3 in complex with $^{13}$C-$^{15}$N-labeled SIMMCAF1), respectively. The protein concentrations of these complexes were 0.7 mM ($^{13}$C-$^{15}$N-labeled SUMO-3 in complex with unlabeled SIMMCAF1) and 0.5 mM (unlabeled SUMO-3 in complex with $^{13}$C-$^{15}$N-labeled SIMMCAF1), respectively.

**NMR Spectroscopy**—All NMR experiments were performed at 298 K on a Bruker DRX-600, DRX-700, or DRX-800 spectrometer. Data processing and analysis were performed using NMRPipe (23) and Sparky (35), respectively. The sequential backbone chemical shift assignments of both SUMO-3 and SIMMCAF1 in the complex were obtained from the following triple-resonance experiments: three-dimensional HNCO, three-dimensional HN(CA)CO, three-dimensional HNCACB, three-dimensional HC(N)CO, three-dimensional CBCA(CO)NH, and three-dimensional HN(CA)CO. The side-chain chemical shifts of both proteins in the complex were assigned using five spectra: three-dimensional HBHA(CO)NH, three-dimensional CC(CO)NH, three-dimensional HCC(CO)NH,
three-dimensional HCCH-TOCSY, and two-dimensional 1H-13C HSQC. The intramolecular distance restraints were obtained from three-dimensional 15N-edited NOE and three-dimensional 13C-edited NOE spectra with a mixing time of 120 ms. The intermolecular distance restraints were obtained from 15N-edited (F2)/15N, 13C-filtered (F1) NOE and 13C-edited (F2)/15N, 13C-filtered (F1) NOE spectra of two isotopically labeled SUMO-3-SIMMCAF1 complexes. 15N T1 and T2, and 1H-15N steady-state NOE values were acquired at 298 K on a Bruker DRX-700 at a 15N frequency of 71.0 MHz, as described (24). The T1 relaxation delays were 10, 80, 180, 300, 450, 640, 800, and 1000 ms. The T2 relaxation delays were 16, 32, 48, 64, 80, 96, 128, and 160 ms. Data analysis and exponential curve fitting were performed with Sparky (35). In the experiments for 1H-15N steady-state NOEs, relaxation delays of 5 s before the 1H saturations of 3 s were applied.

Structure Calculations—All of the NOE cross-peaks were picked manually using Sparky (35). Intramolecular NOE peaks were assigned in an automated manner using the CANDID algorithm of CYANA (25). Intermolecular distance restraints derived from the 15N-edited (F2)/15N, 13C-filtered (F1) NOE and 13C-edited (F2)/15N, 13C-filtered (F1) NOE spectra were generated by manual assignment. Using the CNS program with the restraints defined above, the structures of SUMO-3-SIMMCAF1 were calculated in vacuo with no electrostatic interactions. The 20 CNS structures, which gave the lowest energies, were further refined by the simulated annealing method using the MARBLE molecular dynamics simulation program (26) in an explicit water environment (27) with full electrostatic interactions. The electrostatic interactions were calculated using the Particle Mesh Ewald method (33) without any cut-off operation. The systems contained about 11,000 water molecules and 8 counter ions. The force field used was CHARMM22/CMAP (34) and, in the calculation, two of the histidine residues in SUMO-3 were protonated. The restraint functions adopted in MARBLE were the same as those used in CNS with the default values. The simulated annealing processes applied here were the same as those in Ref. 25 except for the final energy minimization processes, which were performed by 1000 steps of the steepest descent method and 1000 steps of the conjugate gradient method with the increased NMR scaling factors and force constants of bond stretching. The total simulation time for generating 20 structures was ~40 ns, taking ~54 h in a 128-CPU parallel computer.

Mutational Binding Assays—Mutant constructs were prepared by site-directed mutagenesis (QuikChange, Stratagene). The binding affinities between SUMO-1/3 and SIMMCAF1 or its mutant and between SUMO-1/3 or their mutant and SIMMCAF1 were determined by isothermal titration calorimetry (ITC). ITC measurements were performed at 25 °C using a MicroCal VP-ITC calorimeter. The protein/peptide samples were dissolved in a buffer containing 20 mM phosphate (pH 6.8), 1 mM dithiothreitol, and 0.01, 0.15, 0.3, 0.6, or 1.0 mM NaCl, and thoroughly degassed before the titration experiments. The peptide concentrations were determined by a BCA protein assay (Pierce). Five microliters of ~0.7 mM SUMO-3, SUMO-1, or a SUMO mutant were injected at 5-min intervals into the 1.4-ml sample cell containing ~0.03 mM peptide. The baseline-corrected data were analyzed using software provided by the manufacturer (MicroCal). The first data point was excluded from the analysis. For each interaction, two independent titration experiments were performed.

RESULTS

Structure Determination of SUMO-3-SIMMCAF1—MCAF1 contains a SUMO-interacting motif (SIM), 965GVIDLTMDDE975, which is comprised of a hydrophobic core (underlined) and a C-terminal acidic stretch. We determined the binding affinity of the SIM-containing fragment of MCAF1 (938KTI-DASVSKKAADSTSQCGLGKATGS5DSSGVIDLTMDDDEE5GASQ981) with an extra sequence, GSPEF, at its N terminus (designated as SIMMCAF1) for SUMO-3 and SUMO-1 by ITC. SIMMCAF1 had a higher affinity for SUMO-3 (Kd = 1.3 ± 0.1 μM) than for SUMO-1 (Kd = 13.9 ± 2.1 μM) in a buffer containing 20 mM phosphate (pH 6.8), 1 mM dithiothreitol (see Fig. 5a and Table 2).

The molecular mechanism of the binding of SIM with SUMO-3 was investigated by determining the tertiary structure of SIMMCAF1 in complex with SUMO-3 using NMR spectroscopy. Resonance assignments and distance restraints for structure calculations were obtained from the heteronuclear multidimensional spectra of complexes in which either SUMO-3 or SIMMCAF1 was labeled uniformly with 15N and 13C. Intermolecular distance restraints were derived from 15N, 13C-filtered NOE experiments using the complex samples (Fig. 1).

The structure of the SUMO-3-SIMMCAF1 complex was determined from a total of 1109 NMR-derived distance restraints, including 85 intermolecular restraints (supplemental Table S1), and 112 dihedral angle restraints. After the structure calculations with CNS (28), the final structure refinements were performed by MARBLE (27). This refinement procedure significantly improved the geometry of the structures as assessed by the Ramachandran plot obtained by PROCHECK (29) (Table 1). The ensemble of the final 20 structures is well defined, except for the 32 N-terminal residues (five amino acids (GSPEF) from the vector and residues 938–964 of MCAF1) and the six C-terminal residues of SIMMCAF1 and the 15 N-terminal residues of SUMO-3 (Fig. 2a). Few long range NOEs were observed for the terminal regions of SUMO-3 and SIMMCAF1, which showed poor convergence, suggesting that these regions are flexible in solution. Intrinsic flexibility of the N-terminal
region of SUMO-3 has also been suggested by previous structural studies (21). The root-mean-square deviations (r.m.s.d.s) of backbone and all heavy atoms over residues 16–87 of SUMO-3 and residues 966–969 of SIMMCAF1 were 0.61 Å and 1.14 Å, respectively.

The acidic stretch of SIMMCAF1 (residues 972–975) was less converged than the hydrophobic core of MCAF1 (residues 966–969). The r.m.s.d.s of the backbone and heavy atoms over the acidic stretch in the 20 structures refined by the MARBLE calculations were 1.18 Å, respectively, whereas those in the 20 lowest energy structures generated by the CNS calculations were 0.91 Å and 2.39 Å, respectively. In contrast, the hydrophobic core of MCAF1 was well converged: the r.m.s.d.s of the backbone and heavy atoms over residues 966–969 in the MARBLE structures were 0.29 Å and 0.97 Å, respectively. The structure that had the smallest r.m.s.d. with respect to the mean structure was selected as a representative and used for further structural description.

Structure Description of the SIMMCAF1-SIM-SUMO-3 Complex—A ribbon diagram of the representative structure is shown in Fig. 2b. The overall fold of the structure excluding the N-terminal residues of SUMO-3 in the complex was nearly identical to the crystal structure of unliganded SUMO-3 (30). Residues 965–975 of SIMMCAF1 adopted an extended conformation, and were embedded in a hydrophobic cleft located between strand β2 and helix α1 of SUMO-3. Residues 966–969 of SIMMCAF1 formed a parallel β-sheet pairing with strand β2 of SUMO-3. The side chains of these SIMMCAF1 residues contributed to a major interface between SUMO-3 and SIMMCAF1, as evidenced by a large number of intermolecular NOEs observed for these regions (supplemental Table S1). The side-chain aliphatic groups of Ile-967 and Leu-969 (positions +2 and +4 of the SIM consensus, respectively; Fig. 2c) of MCAF1, which are two of the three conserved hydrophobic residues in SIMs, made extensive hydrophobic contacts with the side chains of Val-30, Phe-32, and Lys-42 and Ile-34, Thr-38, and Leu-43 of SUMO-3, respectively (Fig. 3a). The remaining conserved hydrophobic residue in the SIM, Val-966 (positions +1, Fig. 2c), interacted with the aliphatic part of the side chain of Val-30, Gln-31, and Arg-50. In addition, the negatively charged side chain of Asp-968 showed an electrostatic interaction with the side chain of Lys-33 of SUMO-3 (Fig. 3b). Uchimura et al. (22) used mutational analysis to demonstrate the importance of the residues Val-966, Ile-967, Asp-968, and Leu-969 of MCAF1 for SUMO-3 binding, and showed that alanine substitution of any of these residues markedly reduced binding affinity.

The complex structure also suggests that 972DDEE975, the acidic stretch of SIMMCAF1, interacted electrostatically with a positively charged surface patch formed by Lys-33, Lys-35, Arg-36, and His-37 residues from the loop connecting strand β2 and helix α1 of SUMO-3 (Fig. 3b), although the convergence of the acidic stretch was relatively poor in the final structures (Fig. 2a). In some of the final NMR structures, several potential salt bridges were observed between Lys-33, Lys-35, or His-37 of SUMO-3, and Asp-968, Asp-972, Asp-973, or Glu-975 of SIMMCAF1 (data not shown).

Despite its importance for SUMO-3 binding, the acidic stretch of SIMMCAF1 (residues 972–975) has higher conformational flexibility than the hydrophobic core, as suggested by the low convergence of NMR structures (Fig. 2a). The poor convergence of the acidic stretch is consistent with dynamic properties, as suggested by the low [1H]-15N steady-state NOE and larger 15N T2 values in this region (Fig. 4). The NOE values for the backbone amide groups of residues in the acidic stretch (<0.4) are significantly smaller than those for residues Ile-967 to Thr-970 (>0.7). The T2 values for the backbone amide 15N resonances of residues in the acidic stretch (>0.14 s) are also significantly larger than those for residues Val-966 to Thr-970 (<0.07 s). These results suggest that there is higher mobility in the main chain of the acidic stretch than in the hydrophobic core.

Mutational Assay—To further examine the significance of the electrostatic interactions between the SIMMCAF1 acidic stretch and the basic patch of SUMO-3, we analyzed the effect of amino acid substitution of the patch-forming residues, Lys-35, Arg-36, and His-37, in SUMO-3, on binding affinity to SIMMCAF1. ITC measurements showed that alanine substitution of Lys-35, Arg-36, or His-37 led to a moderate loss of affinity, giving a 4.4-, 4.3-, or 1.7-fold increase, respectively, in the dissociation constant (Kd, Table 2). These results suggest that the positive charges of Lys-35 and Arg-36 contributed equally to the binding, with lesser contribution from the side chain of His-37, although no intermolecular NOEs for Arg-36 or His-37 were identified in the NOESY spectra. Double (R36A,H37A) or triple (K35A,R36A,H37A) substitutions confirmed the significance of these residues for SIMMCAF1 binding and resulted in a 4.6- and 9.8-fold reduction in affinity, respectively.

We also analyzed the effects of substitutions of Asp-972 and Glu-975 in the acidic stretch of SIMMCAF1 on SUMO-3 or.

### Table 1

| NMR refinement statistics for the SUMO-3-SIMMCAF1 complex structures | CNS* | MARBLE* |
|---------------------------------------------------------------|------|--------|
| **NMR distance and dihedral constraints** | | |
| Distance constraints | | |
| Total NOE | 1109 | 1109 |
| Intramolecular | 1024 | 1024 |
| Sequential (i − j = 1) | 622 | 622 |
| Medium-range (i − j ≤ 4) | 166 | 166 |
| Long-range (i − j > 5) | 236 | 236 |
| Intermolecular | 85 | 85 |
| Hydrogen bonds | 56 | 56 |
| Total dihedral angle restraints | 112 | 112 |
| **Structure statistics** | | |
| Violations | None | None |
| Number of dihedral angle violation > 3° | None | None |
| Deviations from idealized geometry | | |
| Bond lengths (Å) | 0.0052 | 0.0033 |
| Bond angles (°) | 0.3918 | 2.8585 |
| Average pairwise r.m.s.d. (Å) | 1.05 ± 0.09 | 1.14 ± 0.23 |
| Heavy | 0.51 ± 0.10 | 0.61 ± 0.23 |
| Backbone | | |
| **Ramachandran plot statistics (%)** | | |
| Residues in most favored regions | 79.1 | 88.8 |
| Residues in additional allowed regions | 17.6 | 8.9 |
| Residues in generously allowed regions | 2.5 | 1.6 |
| Residues in disallowed regions | 0.8 | 0.7 |

* The NMR structures calculated by using the CNS program in vacuo and by using the MARBLE program in the explicit water environment.

** Structure of MCAF1 SIM-SUMO-3 Complex **

The structure of MCAF1 SIM-SUMO-3 Complex was determined through NMR spectroscopy. The overall fold of the complex includes the N-terminal residues of SUMO-3 in the SIMMCAF1. The hydrophobic core of MCAF1, consisting of residues 966–969, interacts with the side chains of Val-30, Phe-32, and Lys-42 and Ile-34, Thr-38, and Leu-43 of SUMO-3, respectively. The acidic stretch of SIMMCAF1 (residues 972–975), marked by 972DDEE975, interacts electrostatically with a positively charged surface patch formed by Lys-33, Lys-35, Arg-36, and His-37 of SUMO-3. The acidic stretch is less converged than the hydrophobic core, as indicated by the lower NOE values and larger T2 values in this region. The backbone amide 15N resonances of residues in the acidic stretch are significantly larger than those for residues Val-966 to Thr-970. These results suggest higher mobility in the main chain of the acidic stretch than in the hydrophobic core. Mutational analysis confirmed that the positive charges of Lys-35 and Arg-36 contribute equally to binding, with lesser contribution from His-37. Substitutions in SIMMCAF1, such as R36A,H37A or K35A,R36A,H37A, confirmed the importance of these residues for binding and resulted in 4.6- and 9.8-fold reductions in affinity, respectively.

### Supplemental Table S1

| Region | Value |
|--------|-------|
| 965–975 of SIMMCAF1 | 0.29 ± 0.14 |
| 966–969 of SIMMCAF1 | 0.97 ± 0.28 |
| 972–975 of SIMMCAF1 | 1.14 ± 0.23 |
| Backbone | 0.61 ± 0.23 |

These values are from NOE spectra and indicate the mobility and dynamic properties of the residues in the acidic stretch and hydrophobic core.
SUMO-1 binding. In the ensemble of structures, the positions of the side chains of these residues suggest that they have the potential to form salt bridges with basic residues of SUMO-3. Intermolecular NOE cross-peaks were observed for Asp-972 and Glu-975 in the NOESY spectra (supplemental Table S1). Simultaneous alanine substitutions of these residues (D972A,E975A) led to a drastic decrease in affinity for SUMO-3 and SUMO-1, as observed in an ITC experiment (Fig. 5b). The binding was so weak that we could not quantitate affinities by ITC measurements. These results suggest that the acidic stretch of SIM contributes significantly to binding for both SUMO paralogues.

The site of the positively charged patch in SUMO-3 differs from that in SUMO-1, which possesses the sequence 35KMT41 at the position of the 35KRH37 of SUMO-3 (Fig. 7a). Therefore, we questioned whether the preferential binding of SIMMCAF1 to SUMO-3 compared with SUMO-1 might be attributed to the interactions mediated by Arg-36 and His-37 of SUMO-3. The effect of simultaneous substitutions of these residues to the corresponding amino acids of SUMO-1, methionine and threonine, respectively, resulted in a 2.9-fold decrease in binding affinity, suggesting that the preferential binding to SUMO-3 over SUMO-1 may be attributed to the interactions mediated by Arg-36 and His-37 of SUMO-3. It is interesting to note that Smt3, the SUMO orthologue in Saccharomyces cerevisiae, exhibits a higher affinity to SIMMCAF1 (Kd = 5.49 ± 0.8 μM) than SUMO-1. Smt3 has a 40KKT42 sequence at the corresponding position to 35KRH37 of SUMO-3 (Fig. 7a). Thus, it can be assumed that the positive charge of Lys-41 may contribute to the relatively high affinity of Smt3 for SIMMCAF1.

To examine the significance of electrostatic interactions between SIMMCAF1 and the SUMO paralogues, we measured the binding affinities between SIMMCAF1 and SUMO-3 or SUMO-1 at various
NaCl concentrations by ITC. Although the binding affinity of the SIMMCAF1 interaction with SUMO-3 at a NaCl concentration of 0.15 M was found to be comparable to that in a solution without NaCl, the interaction at an NaCl concentration of 0.3 M was too weak to determine the binding affinity (Table 3 and supplemental Fig. S1). In contrast, SIMMCAF1 was found to retain its interaction with SUMO-1 at an NaCl concentration of 0.3 M. These results indicate that the SIMMCAF1-SUMO-3 interaction is more sensitive to ionic strength than is SIMMCAF1-SUMO-1, and suggest the significance of electrostatic interactions in the preferential binding of SIMMCAF1 to SUMO-3.

**DISCUSSION**

**Structure Comparison with Other SUMO-SIM Complexes**—Our understanding of the structural basis of the recognition of SUMO-1 or SUMO-3 by canonical SIM consensus sequence comes from determination of structures of SUMO-1 in complex with a SIM of PIASxα (18) and of the quaternary complex formed by SUMO-1- conjugated RanGAP1, Ubc9, and RanBP2 (Nup358) (19). In the structures, the hydrophobic core regions of the SIMs commonly undergo β-sheet pairing with strand β2 of SUMO-1, embedded in the clefts formed between strand β2 and helix α1 of SUMOs. However, the orientations of the SIMs with respect to SUMO-1 vary: the SIM of PIASxα pairs in a parallel fashion with the β2 strand, whereas that of the RanBP2 pair with strand β2 of SUMO-1 is in an antiparallel manner. Because the orientation of SIMMCAF1 to SUMO-3 is the same with that of SIMPIASxα to SUMO-1, we compare the structure of SIMMCAF1-SUMO-3 complex with that of SIMPIASxα-SUMO-1 complex below.

The consensus core sequences (positions +1 to +4) are identical between the SIMs of MCAF1 and PIASxα, whereas the positions of the acidic stretches are different (Fig. 6a). Comparison between the structures of SIMMCAF1-SUMO-3 and SIMPIASxα-SUMO-1 (18) reveals both conserved and non-conserved interactions of the SIM hydrophobic core sequences with SUMO-3 and SUMO-1 (Fig. 6b). The hydrophobic residues that are important for SIM binding in the clefts of SUMO-1 and SUMO-3 are homologous, but their amino acid identity is rather limited: in SUMO-1 these residues are Ile-34, Phe-36, and Val-38, whereas in SUMO-3 they are Val-30, Phe-32, and Ile-34; thus, only Phe-36 of SUMO-1 and Phe-32 of SUMO-3 are identical (Figs. 6b and 7). These conserved phenylalanines display similar interactions with SIMPIASxα and SIMMCAF1, respectively: Phe-32 of SUMO-3 makes hydrophobic contacts with Ile-967 (at position +2) and Leu-969 (+4) of SIMMCAF1, and similarly Phe-36 of SUMO-1 also contacts the aliphatic residues Ile-470 at position +2 and Leu-472 at +4 of PIASxα. Other contacts are conserved between Val-30 of SUMO-3 and Ile-967 (+2) of SIMMCAF1; the corresponding residue of SUMO-1, Ile-34, also makes a hydrophobic contact with aliphatic residues at position +2 of PIASxα. In contrast, the hydrophobic interactions of Val-966 (position +1) of SIMMCAF1 with Val-30, Gln-31, and Arg-50 of SUMO-3 seem to be unique, because Val-469 at position +1 of SIMPIASxα seems to contact only His-35 of SUMO-1. Another possible unique contact is made by Asp-968 at position +3 of SIMMCAF1; the side chain of this residue interacts with Lys-33 of SUMO-3, and intermolecular NOEs were observed between these residues. The importance of this electrostatic interaction is supported by a previous mutational study showing that the alanine substitution of Asp-968 of MCAF1 totally abolishes its interaction with SUMO-3 (22). In contrast, the side chain of the corresponding residue in SIMPIASxα, Asp-471, is not well defined in the SUMO-1 complex, and thus its interaction with Lys-37 (at the same position as Lys-33 of SUMO-3) is not clear. No inter-residue NOEs has been identified for Asp-471 (18). Collectively, the present SIMMCAF1-SUMO-3 structure reveals both conserved and non-conserved interactions of the hydrophobic core of canonical SIMs with SUMO-1 and SUMO-3.

In addition to these complex structures containing canonical SIMs, the structures of TDG conjugated to SUMO-1 and...
SUMO-3 provide the structural basis for the interaction of a SIM-like sequence of the substrate TDG with covalently linked SUMO-1 and SUMO-3 (16, 21). The SIM-like sequence of TDG has limited sequence similarity to the SIMs of MCAF1 and PIASx, and other canonical SIMs (Fig. 6a). Its orientation of binding to SUMO is opposite to that of MCAF1 and PIASx, but is the same as that of RanBP2. Therefore, the interactions with SUMO-3 are not well conserved between the SIM hydrophobic core of MCAF1 and TDG (Fig. 6b). For example, Arg-50 of SUMO-3 forms a salt bridge with the unconserved residue Glu-310 (position +2) of TDG, whereas Arg-50 makes no such a bond in the SIMMCAF1-SUMO-3 complex. On the other hand, Phe-32, Ile-34, and Leu-43 of SUMO-3 make hydrophobic contacts with both residues at position +4, Val-308 of TDG, and Leu-969 of MCAF1.

The Role of the Acidic Stretch of SIMMCAF1 in SUMO-3 Binding—In the reported structures of SUMO-SIM complexes, the residues at the acidic stretch are not well converged (18), or not included in the construct (19). Thus, it has been suggested that negative charges on the side chains of these residues might modulate SUMO-SIM interaction through long range electrostatic interactions with basic residues in the vicinity of the hydrophobic cleft of SUMO (17).

The importance of the acidic stretches of SIMs for SUMO-1 binding has been shown by yeast two-hybrid experiments (17). Hecker et al. reported that deletion of the acidic SIM sequence from PIASx considerably decreased its binding to SUMO-1 but had no effect on SUMO-2 binding (17). Based on chemical shift perturbation experiments, they proposed that the acidic stretch of PIASx interacts with Lys-78 of SUMO-1.

In contrast, for SIMMCAF1, there are evidences that suggest the acidic stretch contributes to binding to SUMO-3. First, simultaneous substitution of Asp-972 and Glu-975 in SIMMCAF1 led to a marked loss of binding to SUMO-3 and SUMO-1 (Fig. 5b). Consistent with this observation, GST pulldown analysis by Uchimura et al. found that alanine substitution of either Asp-973 or Glu-975 caused a large reduction in SUMO-3 binding (22). Second, in some of the final 20 structures of the SIMMCAF1-SUMO-3 complex, the acidic stretch is in a position that would allow electrostatic interactions with the side chains of either Lys-33, Lys-35, or His-37. Third, mutation of Lys-35, Arg-36, and His-37 of SUMO-3 indicated that these positively charged patch-forming residues

![Image](https://example.com/image.png)
Structure of MCAF1 SIM-SUMO-3 Complex

The NMR structure ensemble of the SIMMCAF1 sequence, TM (Fig. 6a). Due to this shorter spacer sequence, in the NMR structure ensemble of the SIMMCAF1-SUMO-3 complex, the acidic stretch of SIMMCAF1 does not appear to be in the position to interact with Ala-74 of SUMO-3, which is located at the position corresponding to Lys-78 of SUMO-1. Therefore, it is unlikely that the acidic stretch interacts with Lys-78 upon SIMMCAF1 binding to SUMO-1. On the other hand, the possible spatial distribution of the acidic stretch of the SIM of PIASxα, as deduced from the NMR structure ensemble of SUMO-1-SIMPIASxα (18), suggests that it could be close to the side chain of Lys-78, owing to the long spacer sequence.

Because the distributions of positive and negative charges near the clefts of SUMO-1 and SUMO-2/3 are not identical, the effect and mode of electrostatic interactions mediated by the acidic residues of SIM might be different in these SUMO paralogues (Fig. 7). The acidic stretch can be located at either the N- or C-terminal side of the SIM hydrophobic core with spacing from +1 to +6 residues (17). Therefore, the roles of the acidic patches of SIMs in binding to SUMO-1 or SUMO-3 may partly depend on their positions relative to the hydrophobic cores. Nevertheless, there are several observations that are inconsistent with this possible explanation. The deletion of the acidic stretch from the SIM of Sp100, whose spacer sequence is SS, was reported to decrease its affinity for SUMO-1, but had no effect on SUMO-2 binding, as shown by GST pulldown experiments using cell lysates (17). This observation, together with a previous yeast two-hybrid analysis of interactions between other SIM sequences and SUMO paralogues (17), suggests that factors other than spacer length might also affect the roles of the acidic stretch.

It is noteworthy that the surface-positive patch has been demonstrated to be important in SUMO-mediated transcriptional inhibition (31). Through a comprehensive function analysis, four basic residues (Lys-33, Lys-35, Lys-42, and Arg-50) of SUMO-3 are implicated in transcriptional repression through interactions with repressors (32).

**Preferential Binding to SUMO-2/3—SIMMCAF1 preferentially binds to SUMO-2/3 over SUMO-1. An electrostatic interaction mediated by the acidic patch of SIMMCAF1 may partly contribute to this preferential binding to SUMO-2/3. In some of the ensemble structures, the side chains of Asp-968, Asp-972, Asp-973, and Glu-975 of SIMMCAF1 are positioned where they may potentially interact with Lys-33, Lys-35, or His-37 of SUMO-3. Among these residues, only His-37 is not conserved in SUMO-1 (Fig. 7a). Alanine substitution of this residue causes 1.7-fold decrease in SIMMCAF1 binding (Table 2). The adjacent residue Arg-36, which is also not conserved in SUMO-1, has been shown by mutation analysis to contribute to binding. We therefore suggest that the surface-positive charges of these unconserved residues may be responsible, in part, for the preferential binding of SIMMCAF1 to SUMO-3. Consistent with this assumption, the binding of SIMMCAF1 to SUMO-3 is more sensitive to ionic strength as compared with the binding to SUMO-1 (Table 3 and supplemental Fig. S1); however, this salt sensitivity may also be attributable to the electrostatic interaction between Asp-968 at position 3 in the hydrophobic core sequence of SIMMCAF1 and Lys-33 of SUMO-3 (Fig. 6b).

The relatively low {1H}-{15N} steady-state NOE values and larger {15N} T2 values of the main-chain amide groups in the acidic patch of the SIM suggest that this region possesses some conformational flexibility and, thus, that the elec-
trostatic interactions are not as stable as the hydrophobic contacts made by the hydrophobic core sequence of the SIM (Fig. 4). It is possible that the acidic stretch has a relatively large spatial distribution, and thereby interacts with a large positively charged area formed by Lys-33, Lys-35, Arg-36, His-37, Lys-42, and Arg-50 (Fig. 7b). This possibility would be consistent with the observation that the simultaneous substitution of Asp-972 and Glu-975 of SIM$_{MCAF1}$ has a much larger effect on the SIM$_{MCAF1}$-SUMO-3 interaction than does the simultaneous substitution of Lys-35, Arg-36, and His-37 of SUMO-3 (Table 2). These results imply that, in SUMO-3, basic residues other than Lys-35, Arg-36, and His-37 also contribute to binding through long range electrostatic interactions.
Acknowledgments—We thank Akio Ojida and Itaru Hamachi for their help in recording the ITC data.

REFERENCES

1. Seeler, J., and Dejean, A. (2003) Nat. Rev. Mol. Cell Biol. 4, 690–699
2. Hay, R. (2005) Mol. Cell 18, 1–12
3. Gill, G. (2004) Genes & Dev. 18, 2046–2059
4. Johnson, E. (2004) Annu. Rev. Biochem. 73, 477–487
5. Gill, G. (2004) Genes Dev. 18, 2046–2059
6. Hochstrasser, M. (1996) Annu. Rev. Genet. 30, 405–439
7. Shin, J., Choi, E., Kim, H., Ho, J., Watts, F., Park, S., and Jang, Y. (2005) Mol. Cell 19, 817–828
8. Saitoh, H., and Hinchey, J. (2000) J. Biol. Chem. 275, 6252–6258
9. Azuma, Y., Arnaoutov, A., and Dasso, M. (2003) J. Cell Biol. 163, 477–487
10. Saitoh, H., and Hinchey, J. (2000) J. Biol. Chem. 275, 6252–6258
11. Ayaydin, F., and Dasso, M. (2004) Mol. Biol. Cell 15, 5208–5218
12. Sternsdorf, T., Puccetti, E., Jensen, K., Hoelzer, D., Will, H., Ottmann, O., and Ruthardt, M. (1999) Mol. Cell Biol. 19, 5170–5178
13. Kurepa, J., Walker, J., Smalle, J., Gosink, M., Davis, S., Durham, T., Sung, D., and Vierstra, R. (2003) J. Biol. Chem. 278, 6862–6872
14. Kerscher, O. (2007) EMBO Rep. 8, 550–555
15. Pfander, B., Moldovan, G., Sacher, M., Hoege, C., and Jentsch, S. (2005) Nature 436, 428–433
16. Baba, D., Maita, N., Jee, J., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) Nature 435, 979–982
17. Hecker, C., Rabiller, M., Haglund, K., Bayer, P., and Dikic, I. (2006) J. Biol. Chem. 281, 16117–16127
18. Song, J., Zhang, Z., Hu, W., and Chen, Y. (2005) J. Biol. Chem. 280, 40122–40129