J Mol Model (2015) 21:50
DOI 10.1007/s00894-015-2597-1

ORIGINAL PAPER

Binding properties of SUMO-interacting motifs (SIMs) in yeast

Christophe Jardin · Anselm H. C. Horn · Heinrich Sticht

Received: 29 October 2014 / Accepted: 26 January 2015

Abstract Small ubiquitin-like modifier (SUMO) conjugation and interaction play an essential role in many cellular processes. A large number of yeast proteins is known to interact non-covalently with SUMO via short SUMO-interacting motifs (SIMs), but the structural details of this interaction are yet poorly characterized. In the present work, sequence analysis of a large dataset of 148 yeast SIMs revealed the existence of a hydrophobic core binding motif and a preference for acidic residues either within or adjacent to the core motif. Thus the sequence properties of yeast SIMs are highly similar to those described for human. Molecular dynamics simulations were performed to investigate the binding preferences for four representative SIM peptides differing in the number and distribution of acidic residues. Furthermore, the relative stability of two previously observed alternative binding orientations (parallel, antiparallel) was assessed. For all SIMs investigated, the antiparallel binding mode remained stable in the simulations and the SIMs were tightly bound via their hydrophobic core residues supplemented by polar interactions of the acidic residues. In contrary, the stability of the parallel binding mode is more dependent on the sequence features of the SIM motif like the number and position of acidic residues or the presence of additional adjacent interaction motifs. This information should be helpful to enhance the prediction of SIMs and their binding properties in different organisms to facilitate the reconstruction of the SUMO interactome.

Keywords Linear motif · Molecular dynamics (MD) simulations · Protein-protein interaction · SUMO-binding motif (SBM) · SUMO-interacting motif (SIM) · SUMO-SIM complexes

Introduction

Many proteins can achieve their cellular function only when they are part of larger assemblies of two or more proteins. As a matter of fact, protein-protein interactions (PPIs) play a key role in the regulation of many biological processes. Their elucidation is thus crucial to understand processes such as metabolic control, signal transduction, and gene regulation [1–6]. The list of PPIs provided by large-scale studies using for example yeast two-hybrid assays or mass spectrometry is increasing continuously [7–11].

Based on the large number of identified PPIs, it has become obvious that PPI networks are often controlled by posttranslational modification (PTM) of amino acids. One type of modification is the attachment of a functional group, such as methyl, acetyl, or phosphate. Another type of PTM is the attachment of an entire protein, like ubiquitin or SUMO (small ubiquitin-like modifier) to a lysine residue of the target protein. These SUMO moieties mediate PPIs by recognizing short peptide stretches that were termed SUMO-interacting motifs (SIMs) or SUMO-binding motifs (SBMs). SUMO conjugation and interaction have been shown to be involved in various cellular processes, such as transcriptional regulation, cell cycle progression, DNA damage response, and signal transduction [12]. In eukaryotic cells, SIMs are also utilized for host-pathogen interactions [13], e.g., for anti-viral response [14]. In human cells, at least three SUMO paralogs (hSUMO1, 2, and 3) are expressed [15, 16]. The hSUMO2 and hSUMO3

Electronic supplementary material The online version of this article (doi:10.1007/s00894-015-2597-1) contains supplementary material, which is available to authorized users.

C. Jardin · A. H. C. Horn · H. Sticht (✉)
Bioinformatik, Institut für Biochemie,
Friedrich-Alexander-Universität Erlangen-Nürnberg, Fahrstr. 17,
91054 Erlangen, Germany
e-mail: h.sticht@biochem.uni-erlangen.de

Published online: 19 February 2015

isoforms are almost identical and share ~50 % identity with hSUMO1 (Fig. 1).

Structural analyses of SUMO-SIM complexes have shown that all SIMs bind to a surface patch between the α-helix and a β-sheet of the SUMO protein and extend the β-sheet of SUMO by one additional strand. However, two different binding orientations are observed, in which the SIM either attaches as a parallel or an antiparallel strand to the SUMO β-sheet. Structurally characterized examples of SIMs binding in an antiparallel orientation are the M-IR2 domain of RanBP2 [17], and the IR1 domain of RanBP2 [18]. The parallel binding mode is observed for the SIMs from PIASx [19], MCAF1 [20], and Daxx [21]. A very recent publication revealed that the ubiquitin ligase RNF2 recognizes a SUMO-dimer via two SIM motifs. One of the motifs interacts in a parallel fashion with SUMO, while the second SIM concomitantly binds to the second SUMO moiety in an antiparallel fashion [22]. Regardless of the orientation, binding is primarily mediated by a stretch of four residues containing 3–4 hydrophobic amino acids (I, V, or L). This core interaction motif is a common property of all SIMs, whereas the remaining sequence features exhibit a considerable variability. For example the stretch of hydrophobic residues may contain an insertion of a variable residue at either the second or the third position, like exemplified for the V-I-D-L motif present in the SIMs of the PIASx and MCAF1 proteins. Acidic residues were not only observed within the core motif but also in the flanking regions of the core motif [23] and were shown to enhance binding affinity due to interactions with basic residues of hSUMO on SIM interaction interface [21, 24, 25]. In addition, SIM binding may also be regulated by the phosphorylation of serine residues which allows a fine-tuning of the affinity of binding to hSUMO [21, 23, 26].

Several studies have also addressed the issue of binding specificity of SIMs for the different human SUMO paralogs. Most SIMs investigated do not display a pronounced specificity for either hSUMO1 or hSUMO2 [23]. However, more recent work revealed several examples of hSUMO2/3-specific SIMs [20, 27–29]. For the SIM sequence of the human Daxx protein, it has been additionally shown that phosphorylation promotes binding affinity toward hSUMO1 over hSUMO2/3 thus favoring an interaction with hSUMO1-modified proteins [21]. The studies above indicate that despite striking common structural features, human SUMOs exhibit paralog-specific properties.

The observations above also raise the question about the binding properties of SUMO in those organisms that exhibit only one single SUMO protein and no paralogs. This situation is present in the eukaryotic model organism baker's yeast (*Saccharomyces cerevisiae*) for which a large number of SUMO-mediated interactions have been recently identified from genomic interaction studies [30–32]. Yeast SUMO (ySUMO) is encoded by the Smt3 gene and exhibits approximately 50 % sequence identity to the human SUMO proteins (Fig. 1). Sequence analyses based on small datasets of yeast SIMs revealed the presence of a hydrophobic core and of acidic flanking residues [33] suggesting that the overall mode of interaction is similar to the human SUMO-SIM system, but structural information on the yeast SUMO-SIM interaction is yet rather limited and to date there exists only one complex structure of a physiological interaction. This structure is a complex of SUMOylated PCNA with Srs-2 [34]. Srs-2 interacts with ySUMO in a parallel orientation via a short sequence stretch of six residues including the hydrophobic core motif I-I-V-I (Fig. 2a, b). Binding affinity of Srs-2 is significantly enhanced by the presence of a PIP motif that is connected with the SIM motif via a flexible linker (Fig. 2a). Both PIP and SIM motifs are required to recognize SUMO-PCNA specifically [34, 35]. In addition, there exists a complex structure that comprises a so-called monobody, which targets the SIM-binding site of ySUMO [36] (Fig. 2c, d). This monobody consists of a fibronectin type III scaffold with a long loop, which has been optimized by phage display for high-affinity SUMO-binding [36]. The resulting sequence D-L-Y-Y-S-Y is rather different from that of physiological SIMs but nevertheless interacts with SUMO in a highly similar fashion than observed previously for the antiparallel binding mode of human SIMs. These structures suggest that the two different binding orientations observed in human are also feasible in yeast. However, since the latter complex represents a phage-display optimized monobody rather than a physiological SIM, it gives only little information about the sequence features

```
ySUMO     MSDSEVNQEA KP-----EVK PEVKPETHIN LKVS-DGSSE IFFKIKKTTP LRRLMEAFAK RQGKEMDSLR 64
hSUMO1    MSD----QEA KPSTEDLGDK KEG---EYIK LKVIGQDSSE IHFKVKMTTH LKKLKESYCQ RQGVPMNSLR 63
hSUMO2    MAD----EKP KE-----GVK TENN--DHIN LKVAGQDGSV VQFKIKRHTP LSKLMKAYCE RQGLSMRQIR 59
hSUMO3    MSE----EKP KE-----GVK TEN---DHIN LKVAGQDGSV VQFKIKRHTP LSKLMKAYCE RQGLSMRQIR 58
Consensus *::    ::. *       *  *     :*: ***  :..*  : **:*  *  * :* ::::.: ***  * .:*

ySUMO     FLYDGIRIQA DQTPEDLDME DNDIIEAHRE QIGGATY--- ----- 101
hSUMO1    FLFEGQRIAD NHTPKELGME EEDVIEVYQE QTGGHSTV-- ----- 101
hSUMO2    FRFDGQPINE TDTPAQLEME DEDTIDVFQQ QTGGVY---- ----- 95
hSUMO3    FRFDGQPINE TDTPAQLEME DEDTIDVFQQ QTGGVPESSL AGHSF 103
Consensus * ::*  *    .** :* ** ::* *:..:: * **
```

Fig. 1 Sequence alignment of the yeast SUMO (ySUMO; SMT3_YEAST) and three human SUMO isoforms. The key interacting residues in the β-sheet and α-helix regions of SUMO for SIM binding are highlighted in bold

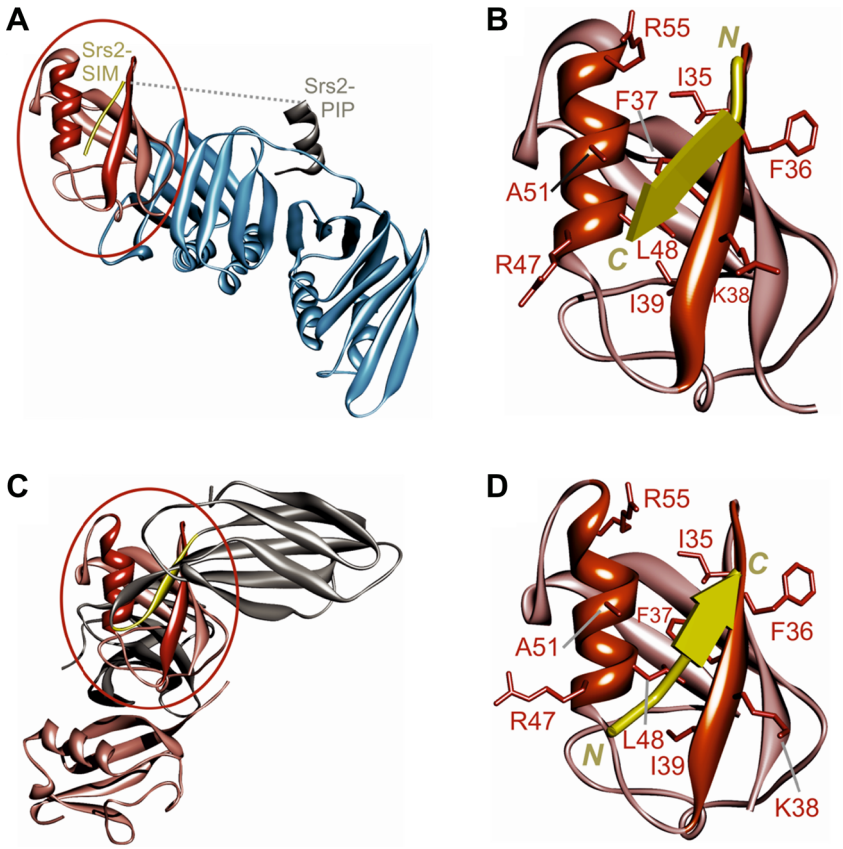

Fig. 2 Structure of ySUMO in complex with different ligands. (**a**) Interaction of SUMOylated PCNA with Srs-2 [34]. Srs-2 interacts with SUMOylated PCNA via a PIP (gray) and a SIM (yellow) motif that target the PCNA (blue) and SUMO (red) moieties, respectively. The dashed line depicts the Srs2 amino acids connecting the PIP and SIM motifs, which are not resolved in the crystal structure. (**b**) Detailed view of the Srs-2 SIM (yellow) bound to ySUMO (red) in a parallel orientation. (**c**) Structure of a fibronectin III-derived monobody (gray and yellow) in complex with ySUMO (red) [36]. (**d**) Detailed view of the monobody SIM (yellow) bound to ySUMO (red) in an antiparallel orientation. In B and D, the SUMO structural elements involved in SIM binding are highlighted in darker color

favoring parallel or antiparallel binding and about the role of flanking residues for the interaction.

To shed light on these issues, we analyzed a large dataset of 148 yeast SIMs and classified the physiologically observed binding motifs. Molecular dynamics (MD) simulations were then performed to investigate the binding preferences for four representative SIM peptides that differ in the number and distribution of acidic residues. Furthermore, the relative stability of two previously observed alternative binding orientations (parallel, antiparallel), which were observed in previous complex structures, was assessed.

Methods

Dataset of SIMs in saccharomyces cerevisiae

SIMs were retrieved from a large-scale study by Srikumar et al., who used an affinity-purification mass spectrometry (AP-MS) approach to analyze the global ySUMO interactome [32]. They identified 452 protein-protein interactions, encompassing 321 different proteins for *Saccharomyces cerevisiae*. For 106 of these proteins a total of 148 high-confidence SIMs were deduced by Srikumar et al. [32], which provided the basis of the present SIM classification analysis. The sequence conservation within these SIMs was computed using the online tool Weblogo [37].

Molecular modeling of the yeast SUMO-SIM peptide complexes

Basically, two sets of ySUMO-SIM complexes were generated. The first set was based on the ySUMO complex crystal structures 3QHT [36] and 3V62 [34] for the antiparallel and parallel binding mode, respectively. In these complexes, the original ligand sequence was replaced by the D-V-V-V-L-D, E-V-V-V-L-E, A-V-V-V-L-A, and D-V-I-D-L-T sequences investigated in the present work using SwissPDBviewer [38]. The second set, which was used for the control simulations, was based on the structure of unliganded ySUMO (PDB 1EUV [39]), where the ligand binding mode of the SIM ligand was modeled according to that observed in homologous human SUMO-SIM complexes. We used PDB-templates 2ASQ [19] and 2RPQ [20] for the parallel ligand orientation and template 2LAS [17] for the antiparallel ligand orientation.

Molecular dynamics (MD) simulations

Separate molecular dynamics simulations were performed for each ySUMO-SIM peptide system with the AMBER12 package [40] and the parm99SB force field [41] following a

protocol well established in our group [42–44]. Acetyl and N-methyl groups were added at the N- and C-termini of the SIM peptides to avoid terminal charge effects during the simulations. The systems were neutralized with an appropriate number of counterions, and solvated with TIP3P water molecules in a box with a distance of at least 10 Å between the solute and the borders. The systems were minimized in two steps, and heated up to 298 K in 1 ns using SHAKE [45] constraints on hydrogens and small positional restraints on the Cα atoms. For each unrestrained system, 500 ns of production run were then carried out. Salt bridges and hydrogen bonds were identified using VMD [46]. Discovery Studio Visualizer [47] and VMD were used for creation of graphical representations.

Results and discussion

Determination of consensus sequences for yeast SUMO binding SIMs

Our dataset comprises 148 SIMs deduced from an affinity-purification mass spectrometry approach [32]. All motifs exhibit a characteristic stretch of 3–4 hydrophobic aliphatic residues (I, V, or L) that constitutes the core motif for SUMO binding. To allow for a finer classification of the core motifs, we defined four sub-patterns that differ in the number and spacing of the hydrophobic residues ϕ (where ϕ represents I, V, or L): the pattern ϕ-ϕ-ϕ-ϕ corresponding to four hydrophobic residues core is present in 38 SIMs. The patterns ϕ-X-ϕ-ϕ and ϕ-ϕ-X-ϕ, in which one of the hydrophobic residues is replaced by another type of amino acid, are detected in 19 and 75 sequences, respectively. The remaining 16 instances of the dataset exhibit the pattern ϕ-ϕ-ϕ-X corresponding to a hydrophobic core motif of three residues.

A representation of the sequence preferences for each of the four subclasses is shown in Fig. 3 as sequence logos. In this presentation, the height of a letter indicates how well the corresponding amino acid is conserved among all instances in the dataset at this site. Inspection of the four patterns reveals that each ϕ-position (positions 1 to 4 in Fig. 3) tolerates either I, V, or L. There are minor differences with respect to the relative frequencies of these amino acids at the individual positions, e.g., ϕ-ϕ-ϕ-ϕ prefers valine at the first position, whereas ϕ-X-ϕ-ϕ prefers isoleucine at the first position. However, due to the rather small size of the present dataset we refrain from a quantitative interpretation of these differences, but rather focus on the sequence preferences observed at the X-positions. For the ϕ-X-ϕ-ϕ pattern, no amino acid preference is detected at the X-position (Fig. 3b), while in the ϕ-ϕ-X-ϕ pattern, the X-position is preferentially occupied by a negatively charged residue (D or E). In the ϕ-ϕ-ϕ-X motif, the predominant residues at the fourth position are M and F (Fig. 3d), which are both hydrophobic in nature. Thus, the ϕ-ϕ-ϕ-X motif is rather similar to the ϕ-ϕ-ϕ-ϕ motif mainly differing in the identity of the hydrophobic residue at the fourth position.

The flanking regions are less conserved than the core motif, but nevertheless exhibit distinct sequence properties. This applies in particular to those positions, which are N- and C-terminal immediately adjacent to the core (positions 0 and 5 in Fig. 3): these positions are predominantly occupied by negatively charged amino acids. The only exception is the ϕ-ϕ-X-ϕ motif, which does not display a preference for acidic residues at the flanking positions. This larger tolerance of more residue types might be explained by the fact that this motif type frequently exhibits a negatively charged residue at the X-position within the core motif itself.

In addition to acidic residues, the polar amino acids serine and threonine are enriched at the flanking regions. This

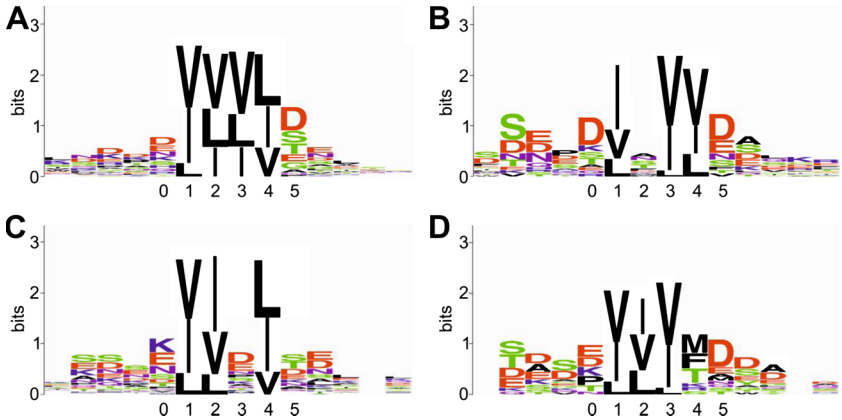

Fig. 3 Sequence logos calculated for 148 SIM instances from yeast. The SIMs were classified into four sub-patterns depending on the sequence of the core motif. (**a**) ϕ-ϕ-ϕ-ϕ, (**b**) ϕ-X-ϕ-ϕ, (**c**) ϕ-ϕ-X-ϕ, and (**d**) ϕ-ϕ-ϕ-X. The core residues are denoted as sequence positions 1–4 in the diagram. The degree of sequence conservation is given in 'bits' with high values indicating a stronger conservation of the respective sequence position. In addition, the sequence conservation is also shown for the five sequence positions N- and C-terminal of the core motif. The positions immediately adjacent to the core motif, which were also investigated in this study, are denoted as positions 0 and 5

enrichment was noted previously for human SIMs and it was also shown that these sites become phosphorylated thus giving rise to additional negative charges in the vicinity of the core motif [21, 23, 26]. For example, the SIM in human Daxx protein exhibits the sequence I-I-V-L-S-D-S-D, in which an S-x-S motif is immediately adjacent to the hydrophobic core motif. Both serines are phosphorylated by casein kinase 2 thereby increasing SUMO affinity [21]. A preference for negatively charged and phosphorylatable amino acids is not only detected for the 0 and 5 positions, but also for the residues more distant from the core — although to lesser extent (Fig. 3). This is in line with the properties of human SIMs curated in the ELM databank [48] showing that the flanking stretches frequently comprise more than one acidic or phosphorylatable residue.

In summary, our analyses show that yeast SIMs contain a four residue core motif that comprises 3–4 hydrophobic aliphatic residues (L, I, V). Flanking regions display a preference for negatively charged amino acids (D, E) or phosphorylatable residues (S, T). Thus, the sequence preferences deduced for yeast SIMs are highly similar to those reported for human SIMs.

Choice of complexes for molecular dynamics simulations

The analysis above gives insight into the sequence preferences of SIMs, but does not allow to answer the following key questions related to the SIM binding properties in yeast:

1) Is there a preferred binding orientation (parallel, antiparallel) for SIMs to ySUMO?
2) Is the binding orientation critically affected by the sequence properties of the SIM (i.e., do the different patterns detected above favor different binding orientations)?
3) What is the role of negatively charged amino acids for SUMO binding?

We addressed these issues by molecular modeling and molecular dynamics simulations. Since MD simulations are computationally expensive, we selected several model systems based on the following considerations:

a) SIM core sequences: The most prevalent sub-patterns deduced from our analysis are ϕ-ϕ-ϕ-ϕ and ϕ-ϕ-X-ϕ. For these two sub-patterns, we used the sequences V-V-V-L and V-I-D-L, which represent the consensus core sequence derived from the logo in Fig. 3a and c, respectively.
b) Flanking residues: To assess the role of acidic flanking residues the V-V-V-L core was extended at the N- and C-terminus either by aspartate, glutamate, or alanine resulting in the peptides D-V-V-V-L-D, E-V-V-V-L-E, and A-V-V-V-L-A. We refrained from modeling additional flanking residues further apart from the core, since these positions are generally less well conserved compared to those immediately adjacent to the core. For the V-I-D-L core, the sequence logo revealed no clear sequence preferences for the flanking residues. Therefore, we investigated only one single peptide, for which the identity of the flanking residues (D,T) was chosen according to the SIM present in the PIASx protein.
c) Binding orientation: All four peptides D-V-V-V-L-D, E-V-V-V-L-E, and A-V-V-V-L-A, and D-V-I-D-L-T were modeled and simulated both in the parallel and in the antiparallel binding orientation. For each of these eight complexes a 500-ns MD-simulation was performed to assess conformational stability and to identify the key interactions for binding.
d) Choice of modeling templates: In order to avoid that the results of the simulations might become biased by the choice of the modeling template, we performed eight independent control simulations, in which the respective complexes were modeled using alternative template structures (see Methods for details) resulting in a total of 16 simulations. To avoid confusion, the results for the control simulations are presented in the supplement.

Binding properties of SIMs in the antiparallel orientation

For the simulations starting from the antiparallel orientation, all four peptides remained bound tightly to ySUMO in the initial position over the entire simulation time, and exhibited only small fluctuations from the starting conformation (Fig. 4).

For a more detailed analysis of the antiparallel binding mode, we first analyzed the contribution of the hydrophobic core residues to binding by measuring the van der Waals interaction energy (Fig. 5). For all four peptides the first position in the core displays the strongest interaction energy (V1; black curves in Fig. 5). Consequently, the hydrophobic residue at position 1 of the core has the most stabilizing contribution to the hydrophobic interaction of SIM with ySUMO. This can be rationalized by its tight packing into a hydrophobic pocket formed by the SUMO residues F37, I39, L48, A51, and the hydrophobic part of the R47 sidechain (Fig. 6a, f). The second position of the core (V2; red curves in Fig. 5) generally has the smallest contribution to van der Waals energy, and is thus less important with regard to SUMO binding. In the MD simulations, this residue is rather oriented toward the solvent, only forming few contacts with the methylene groups of the K38 sidechain (Fig. 6b, g). The third and fourth residues of the core (V3, L4; green and blue curves in Fig. 5), also contribute significantly to the van der Waals energy, albeit less than the first position. This is also in line with the smaller number of interacting hydrophobic residues (Fig. 6c, d).

When V at position 3 is replaced by an aspartate in the V-I-D-L peptide, D3 establishes a salt-bridge with R55 of ySUMO

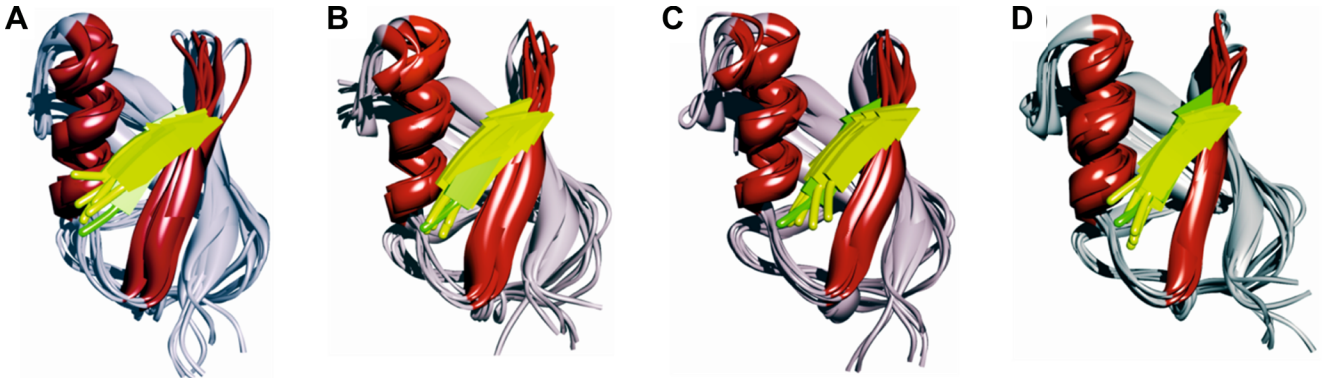

Fig. 4 Conformational stability of the SIM peptides bound to ySUMO in antiparallel orientation. The SIM peptides (**a**) AVVVLA, (**b**) DVVVLD, (**c**) EVVVLE, and (**d**) DVIDLT are depicted as yellow strands (green for the initial structure). The ySUMO protein is shown as ribbons with the secondary elements (α-helix and β-sheet) forming the binding interface in red. Snapshots were recorded for the initial structure and at every 100 ns of the simulations

(Fig. 6h) during the simulation. This favorable interaction offers an explanation, why V3, which is involved in numerous hydrophobic interactions (Fig. 3c), can be replaced by aspartate without affecting the overall conformational stability of the complex (Fig. 4a, d). Thus, depending on the polarity of the residue at this core position, interactions are either formed with the methylene groups (Fig. 6c) or the guanidino group (Fig. 6h) of the R55 sidechain. The dynamic properties and energetic interaction profiles detected above were confirmed by our control simulations starting from template structures obtained by an alternative modeling approach (Figs. S1, S2).

These results obtained for the ϕ-ϕ-ϕ-ϕ and ϕ-ϕ-X-ϕ core are also of relevance for deducing the binding properties of the two other subclasses ϕ-X-ϕ-ϕ and ϕ-ϕ-ϕ-X that were not explicitly investigated by MD simulations. The finding that the first position of the core has the strongest contribution to the interaction is in line with the observation that this position is exclusively occupied by I, V, or L in all four subclasses shown in Fig. 3. The second position of the core is rather oriented toward the solvent and has the smallest contribution to the van der Waals energy. This result is consistent with the sequence logo generated for the sequences matching the ϕ-X-ϕ-ϕ pattern (Fig. 3b). In these sequences, the X-position exhibits no sequence preferences indicating that at this position V/I/L can be replaced by almost every amino acid type without losing binding capability. A correlation between the interaction energy and the conservation of the respective position is also detected for position 4 of the core. MD simulations indicate that the leucine at this position is important for binding (Fig. 5). Consequently, this sequence position appears not to be completely variable in the ϕ-ϕ-ϕ-X core (Fig. 3d), but instead is preferentially occupied by the hydrophobic residues F and M. Thus, the binding properties observed in the MD simulation of the ϕ-ϕ-ϕ-ϕ core also offer an explanation for the sequence variability observed at the X-position of the ϕ-X-ϕ-ϕ and ϕ-ϕ-ϕ-X cores. This also suggests that the sequences exhibiting a ϕ-X-ϕ-ϕ or ϕ-ϕ-ϕ-X core motif can also bind to ySUMO in an antiparallel fashion similar to the ϕ-ϕ-ϕ-ϕ core.

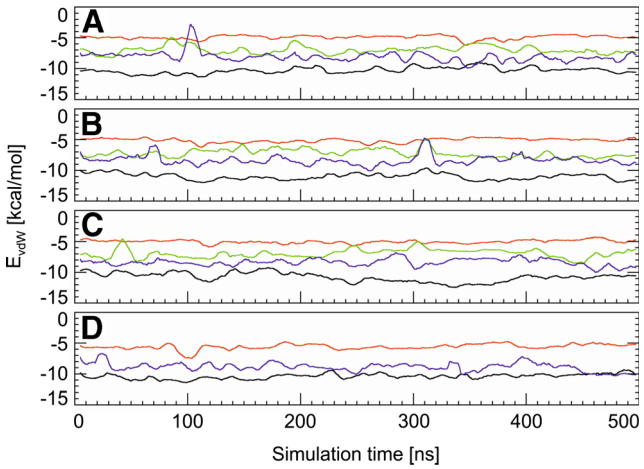

Fig. 5 Interaction energies of the SIM peptides bound to ySUMO in antiparallel orientation. (**a**) AVVVLA, (**b**) DVVVLD, (**c**) EVVVLE, and (**d**) DVIDLT. The van der Waals (vdW) energies over simulation time were calculated as a 1-ns average over a sliding window. The black, red, green, and blue curves represent the individual vdW energies for the hydrophobic residues 1, 2, 3, and 4 of the SIM core motif, respectively

The simulations above have demonstrated that D3 within the core sequence can form favorable polar interactions explaining the frequent observation of this residue within the core motif. However, acidic residues are not only observed within the core motif but also at the flanking sequence positions. We assessed the role of these residues from a comparison of the simulations of the D-V-V-V-L-D, E-V-V-V-L-E, and A-V-V-V-L-A peptides. For all three peptides the overall dynamic properties (Fig. 4a-c) and the interaction energies of the individual core positions (Fig. 5a-c) are very similar. This shows that flanking residues do not enhance the strength of the interaction of the hydrophobic core motif itself. Nevertheless, the enrichment of acidic residues at the flanking positions in the present dataset suggests that these residues might enhance binding affinity by forming additional polar interactions with

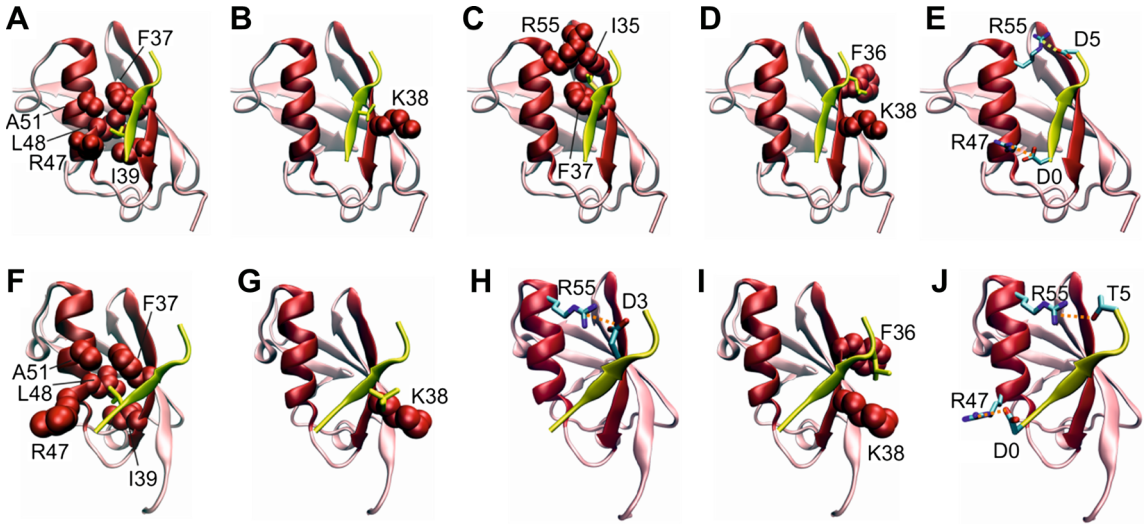

Fig. 6 Contacts formed between the SIM peptides and ySUMO in the antiparallel orientation. Top row (**a**-**e**): contacts for DVVVLD as a representative for SIM peptides with a four hydrophobic residues core. For clarity, the interactions of individual SIM residues are shown as separate panels: (**a**) V1, (**b**) V2, (**c**) V3, (**d**) L4, (**e**) D0 and D5. Hydrophobic SIM residues are shown as yellow sticks and their interacting SUMO residues are indicated in space-filled presentation (red) and labeled according to their sequence position. Side chains of residues forming salt bridges or polar interactions are shown as sticks (cpk coloring). Bottom row: interactions of the DVIDLT peptide. The interactions of individual SIM residues are shown as separate panels: (**f**) V1, (**g**) I2, (**h**) D3, (**i**) L4, (**j**) D0 and T5

ySUMO. This is supported by the observation that both the N-terminal and C-terminal acidic residues can form polar interactions with basic residues of SUMO. In the simulation of the DVVVLD peptide, D0 interacts with R47 whereas D5 interacts with R55 of SUMO (Fig. 6e). The same interactions with residues R47 and R55 of SUMO are observed for E0 and E6 in the simulation of the EVVVLE peptide (data not shown). Thus, the presence of multiple interaction sites on the ySUMO surface also offers an explanation for the observation that acidic residues might be present either at the N- or C-terminal side of the core motif (Fig. 3a).

Binding properties of SIMs in the parallel orientation

As next aspect, we investigated whether the SIM-peptides can also interact with ySUMO in a parallel fashion. When modeled in a parallel orientation all peptides underwent large fluctuations during the simulations and at least partially dissociated from ySUMO (Fig. 7). These motions are shown in detail for representative snapshots from the D-V-V-V-L-D simulation in Fig. 8. Instability of the parallel SIM-ySUMO complexes is also detected in the control simulations and even a complete dissociation is observed for the A-V-V-V-L-A and E-V-V-V-L-E peptides (Fig. S3). Interestingly, the E-V-V-V-L-E peptide re-bound to SUMO after ~350 ns between the α-helix and the β2-sheet in a conformation similar to that observed for the antiparallel SIM peptides. This association event was probably favored by the close spatial proximity between SIM and SUMO and could therefore be detected on the timescale of the simulations. Representative snapshots from this simulation showing the dissociation and rebinding are shown in Fig. S4. This observation suggests that the antiparallel orientation is significantly more stable for this peptide.

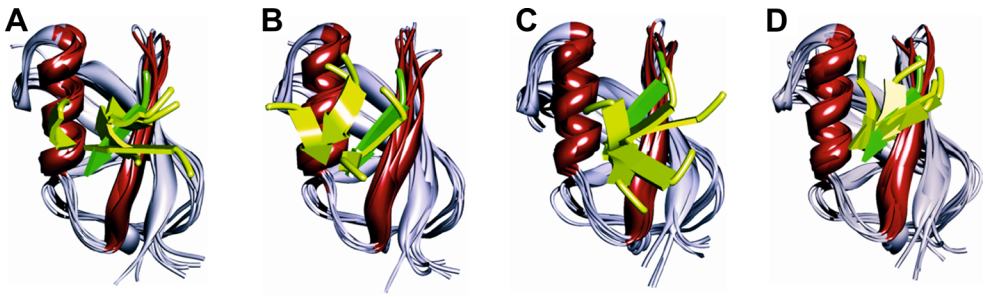

Fig. 7 Conformational stability of the SIM peptides bound to ySUMO in the parallel orientation. The SIM peptides (**a**) AVVVLA, (**b**) DVVVLD, (**c**) EVVVLE, and (**d**) DVIDLT are depicted as yellow strands (green for the initial structure). The ySUMO protein is shown as ribbons with the secondary elements (α-helix and β-sheet) forming the binding interface in red. Snapshots were recorded for the initial structure and at every 100 ns of the simulations

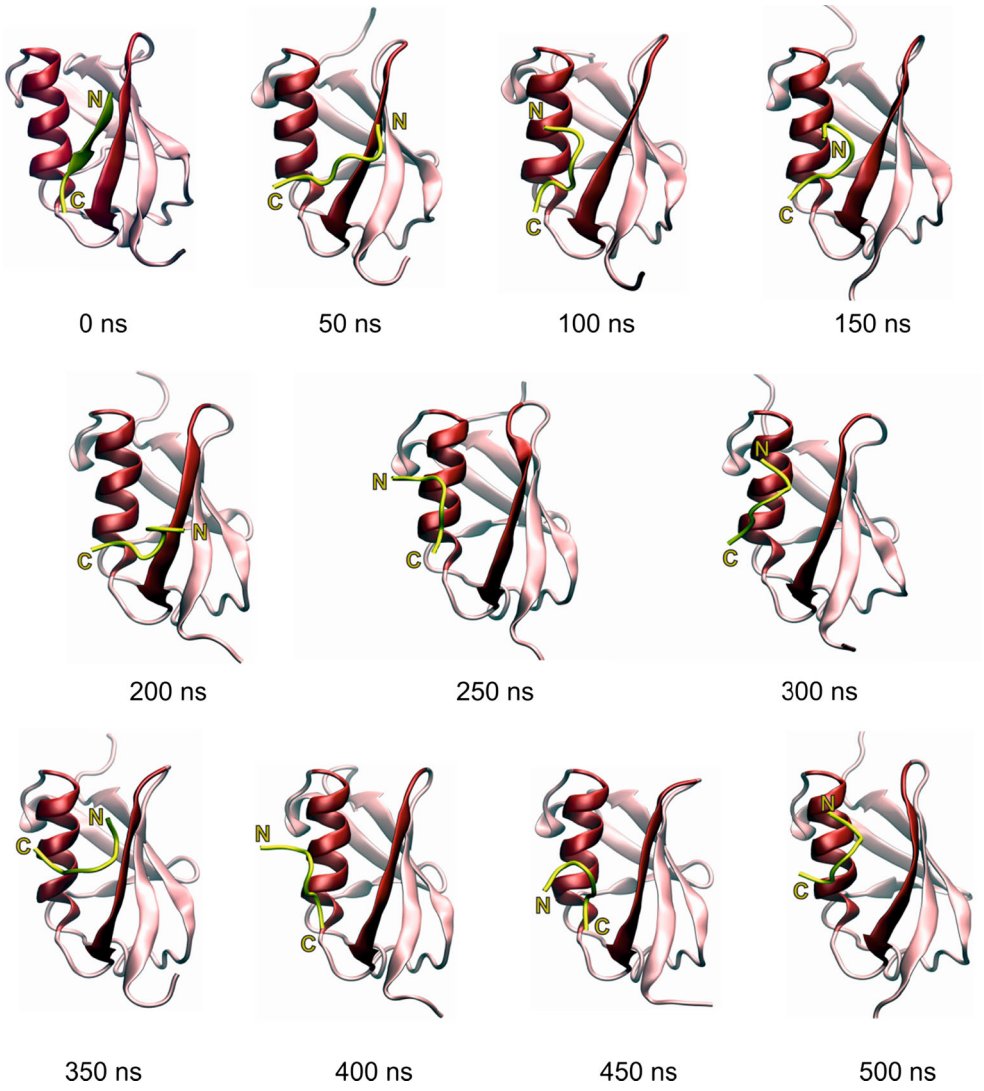

Fig. 8 Snapshots from the MD simulation of the DVVVLD SIM peptide (yellow) bound in the parallel orientation to ySUMO (pink with red ribbons for the secondary elements forming the binding interface)

To assess the energetics of the parallel SIM-SUMO interaction mode in more detail, the van der Waals interaction energies of the hydrophobic SIM residues were monitored (Fig. 9). As expected from the observed flexibility of the complexes (Fig. 7), there are significant fluctuations of the interaction energies over the simulation time (Fig. 9). As a common feature of all simulations, L4 is the hydrophobic residue that forms the tightest interactions with ySUMO (Fig. 9) by packing into a pocket formed by residues F37, I39, L48, A51, and the hydrophobic part of the R47 sidechain (Fig 10c). The same pocket is occupied by V1 for the SIMs bound in the antiparallel orientation (Fig. 6a, f) and V1 exhibits the strongest interactions with ySUMO in the antiparallel interaction (Fig. 5). Therefore, one can conclude that binding to the hydrophobic pocket formed by residues residues F37, I39, L48, and A51 represents a key interaction regardless of the SIM binding orientation.

Comparison of the simulations for the A-V-V-V-L-A, D-V-V-V-L-D, and E-V-V-V-L-E peptides (Fig. 7a-c) indicates that

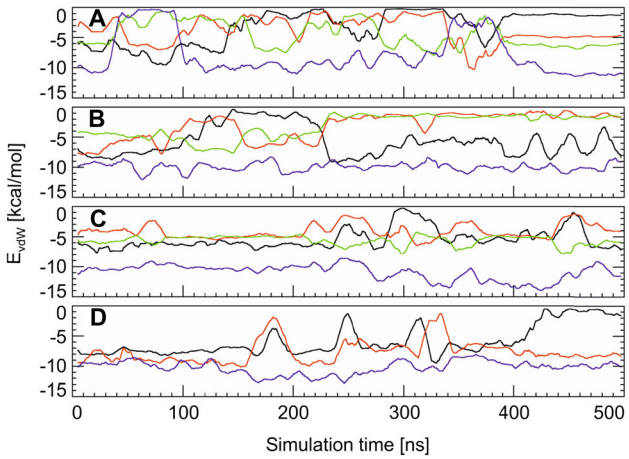

Fig. 9 Interaction energies of the SIM peptides bound to ySUMO in the parallel orientation. (**a**) AVVVLA, (**b**) DVVVLD, (**c**) EVVVLE, and (**d**) DVIDLT. The van der Waals (vdW) energies over simulation time were calculated as a 1-ns average over a sliding window. The black, red, green, and blue curves represent the individual vdW energies for the hydrophobic residues 1, 2, 3, and 4 of the SIM core motif

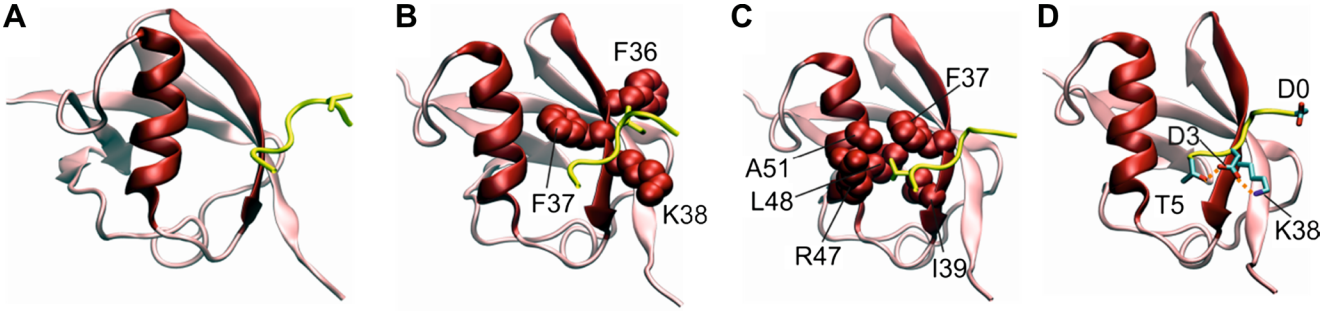

Fig. 10 Contacts formed between the DVIDLT SIM peptide and ySUMO in the parallel orientation. For clarity, the interactions of individual SIM residues are shown as separate panels: (**a**) V1, (**b**) I2, (**c**) L4, (**d**) D0, D3, and T5. Hydrophobic SIM residues are shown as yellow sticks and their interacting SUMO residues are indicated in space-filled presentation (red) and labeled according to their sequence position. Side chains of residues forming salt bridges or polar interactions are shown as sticks in cpk coloring)

the flanking acidic residues do not significantly stabilize the parallel binding mode. In contrary, a central aspartate at position 3 confers stabilization (Fig. 7d) by forming a salt-bridge with K38 of ySUMO (Fig. 10d). In addition, the two hydrophobic residues (I2, L4) flanking the aspartate form favorable hydrophobic interactions with SUMO (Figs. 9d and 10b, c) thus providing a strong anchor for the C-terminal part of the D-V-I-L-T peptide. The N-terminal part of this peptide displays significantly larger motions and neither D0 nor V1 form stable interactions with ySUMO (Figs. 7d, 9d, and 10a, d).

The data above shows that a central aspartate at position 3 does not only stabilize the antiparallel but also the parallel binding orientations. This is due to the fact that ySUMO exhibits several basic residues close to the binding cleft that can form salt-bridges to D3 of SIM. These residues are R55 in the antiparallel orientation (Fig. 6h) and K38 in the parallel orientation (Fig. 10d). Alternatively to K38, K40 may also be used for salt-bridge formation as evidenced from our control simulation and from human SUMO-SIM complexes [22]. This observation suggests that there is a certain degree of plasticity in the parallel binding mode. In this context, it is interesting to note that the parallel orientation frequently relies on additional factors for stabilization.

From the analysis of human SIM-SUMO complexes we noted that several human SIM motifs that bind in a parallel orientation display a distinct enrichment of acidic residues in the peptide stretch C-terminally adjacent to the core motif. One example is the ^{961}SDSSGVIDLTMDDEES^{976} SIM of human MCAF1 [20]. In this motif, the acid stretch D972-E975 is positioned over a basic patch of hSUMO2 and interacts with K33, K35, R36, and H37 (equivalent to residues K38, K40, K41, and T42 of ySUMO; Fig. 1). Simultaneous replacement of SUMO residues K35, R36, and H37 by alanine decreases binding MCAF1 affinity by one order of magnitude [20]. A double mutation D972A, E975A in the MCAF1 SIM even reduces binding affinity beyond the limit of experimental detection [20] thus underscoring the importance of flanking acidic residues in the SIM-SUMO interaction.

A similar asymmetric enrichment of acidic residues with respect to the core motif is also detected for the ^{466}KVDVIDLTIESSSDEEED SIM from human PIASx [19] and the first SIM ^{1}MATANSIIVLDDDDEDE^{17} of human Daxx [49], which both bind in a parallel orientation. These examples suggest that a high negative charge in the C-terminal proximity of the core motif might generally favor parallel binding. The relevance of C-terminal acidic residues is also detected for the second ^{721}KTSVATQCDPEEIIVLSDSD^{740} SIM from human Daxx [21] in which D738/D740 interact with K17 and K39 of hSUMO1 in the parallel binding mode. In the respective SIM, the distribution of acidic residues is less asymmetric compared to the other examples discussed above. This might also explain, why the respective SIM from human Daxx interconverts between parallel and antiparallel binding modes on a ms to μs timescale [49]. Additional support for the idea that an asymmetric distribution of negative charges affects the binding orientation comes from the ^{2705}DNEIEVIIVWEKK^{2717} M-IR2 SIM [17] and RanBP2 SIM ^{2625}DSPSDDDVLIVY^{2636} [18]. These SIMs bind in an antiparallel orientation, which allows the N-terminally charged residues to interact with the same surface patch of hSUMO that is contacted by the C-terminally acidic residues in the parallel binding mode.

Structural analysis of human SIMs bound in a parallel orientation suggests that in particular large clusters of acidic residues, which are located C-terminally of the core motif, might play a crucial role for enhancing binding affinity and binding orientation. Interestingly, only a subset of these acidic residues forms defined polar interactions with SUMO in the complex structures. This is also in line with the observation that these acidic clusters can be separated from the core motif by spacers of variable lengths [23]. This suggests that acidic clusters most probably enhance the on-rate of complex formation by increasing the electrostatic attraction between SIM and SUMO. A correlation between electrostatic complementarity and rate of complex formation has been previously reported for other protein-protein complexes like TEM1-BLIP [50, 51] or RalGDS-Ras [51]. Since a quantification of the effect of such

additional residues on the initial steps of recognition is not possible from MD simulations of the bound complex, we refrained from investigating the role of these acidic clusters in our MD simulations.

Analysis of protein complexes, in which the SUMOylated protein is recognized via multiple interaction motifs, revealed an additional factor that is likely to affect binding orientation. One example is yeast Srs2 [34] that targets SUMOylated PCNA via the synergistic action of a SIM and a PIP sequence motif (Fig. 1a). The length of the linker and the relative orientation between the two interactions motifs only permit parallel binding of the SIM to allow a concomitant interaction. The situation is similar for the human RNF4 [22] that contains two SIM motifs connected by a short 8-residue linker. A simultaneous interaction of both SIMs with a SUMO-dimer is therefore only feasible, if one of the motifs is bound in a parallel orientation.

Conclusions

In the present study, we have analyzed a large dataset of yeast SIMs and modeled their interaction with ySUMO. Generally, the overall sequence properties of ySIMs proved to be highly similar to those previously described in human.

For each peptide investigated, molecular modeling was performed using two alternative modeling templates either based on hSUMO or ySUMO complex geometries. The structural properties of the peptides detected in the MD simulations were not critically affected by the choice of the modeling template. There is, e.g., a large similarity between the dynamics and interactions detected for the antiparallel binding mode, regardless if the yeast monobody or a human SUMO-SIM complex was used as template (see Figs. 4 and 5 vs. S1, S2). This underscores that the monobody complex, despite its unusual D-L-Y-Y-S-Y sequence, represents a suitable template for modeling the interaction of physiological ySIM sequences in this binding orientation.

For all peptides investigated by MD simulations the antiparallel binding mode proved to be more stable than the parallel one. However, we noted that the presence of an acidic residue at position 3 of the core motif has a particularly large stabilizing effect on the parallel orientation (Fig. 7). The idea that a central aspartate is beneficial for parallel binding is also supported by the fact that a significant portion of the human SIMs that bind in a parallel orientation also contains a central aspartate: examples are the V-I-D-L core sequences present in the PIASx and MCAF1 SIMs as well as the I-V-D-L SIM in RNF4.

In the antiparallel orientation, flanking acidic residues immediately adjacent to the core motif can form favorable electrostatic interactions on each side of the core motif (Fig. 6e, j) and are thus expected to enhance binding affinity. In contrary, the parallel orientation did not experience a significant stabilization by the presence of acidic residues on each side of the core motif in our simulations (Figs. 7, 8, and 9), suggesting that this mode of interaction relies on additional sequence requirements. Additional factors that might critically affect the binding orientation are the presence of a cluster of acidic or phosphorylatable residues, the presence of additional interaction motifs as well as the linker length between them and the core motif. Thus, there may exist additional mechanisms for fine-tuning the ySUMO-SIM binding orientation, which exploit structural principles not coded in the core motif itself. Despite these differences in the fine-tuning, our study indicates that the overall sequence and structure properties of SIM-SUMO interactions are evolutionary conserved. This also suggests that the SUMO-SIM interaction data derived from different organisms may be combined to create a larger database for the development of more powerful SIM prediction tools, which will facilitate the reconstruction of the SUMO interactome in different organisms.

Acknowledgments The project was funded by the Deutsche Forschungsgemeinschaft (SFB796, project A2) to HS. The authors thank Melanie Schneider and Jakob Bader for fruitful discussions, as well as Thomas Zeiser from the High Performance Computing group of the Regionales Rechenzentrum Erlangen (RRZE) for providing optimized AMBER executables.

References

1. Jones S, Thornton JM (1995) Protein-protein interactions: a review of protein dimer structures. Prog Biophys Mol Biol 63:31–65
2. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285:2177–2198
3. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci U S A 93:13–20
4. Nooren IM, Thornton JM (2003) Diversity of protein-protein interactions. EMBO J 22:3486–3492
5. Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. Science 300:445–452
6. Aloy P, Russell RB (2004) Ten thousand interactions for the molecular biologist. Nat Biotechnol 22:1317–1321
7. Gietz RD, Triggs-Raine B, Robbins A, Graham KC, Woods RA (1997) Identification of proteins that interact with a protein of interest: applications of the yeast two-hybrid system. Mol Cell Biochem 172:67–79
8. Young KH (1998) Yeast two-hybrid: so many interactions, (in) so little time. Biol Reprod 58:302–311
9. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207
10. Gavin AC, Superti-Furga G (2003) Protein complexes and proteome organization from yeast to man. Curr Opin Chem Biol 7:21–27
11. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M (2008) High-quality

binary protein interaction map of the yeast interactome network. Science 322:104–110
12. Gareau JR, Lima CD (2010) The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat Rev Mol Cell Biol 11:861–871
13. Wimmer P, Schreiner S, Dobner T (2012) Human pathogens and the host cell SUMOylation system. J Virol 86:642–654
14. Cuchet D, Sykes A, Nicolas A, Orr A, Murray J, Sirma H, Heeren J, Bartelt A, Everett RD (2010) PML isoforms I and II participate in PML-dependent restriction of HSV-1 replication. J Cell Sci 124:280–291
15. Hay RT (2001) Protein modification by SUMO. Trends Biochem Sci 26:332–333
16. Müller S, Hoege C, Pyrowolakis G, Jentsch S (2001) SUMO, ubiquitin's mysterious cousin. Nat Rev Mol Cell Biol 2:202–210
17. Namanja AT, Li YJ, Su Y, Wong S, Lu J, Colson LT, Wu C, Li SS, Chen Y (2012) Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis. J Biol Chem 287:3231–3240
18. Reverter D, Lima CD (2005) Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. Nature 435:687–692
19. Song J, Zhang Z, Hu W, Chen Y (2005) Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. J Biol Chem 280:40122–40129
20. Sekiyama N, Ikegami T, Yamane T, Ikeguchi M, Uchimura Y, Baba D, Ariyoshi M, Tochio H, Saitoh H, 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
21. Chang CC, Naik MT, Huang YS, Jeng JC, Liao PH, Kuo HY, Ho CC, Hsieh YL, Lin CH, Huang NJ, Naik NM, Kung CC, Lin SY, Chen RH, Chang KS, Huang TH, Shih HM (2011) Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. Mol Cell 42:62–74
22. Xu Y, Plechanovova A, Simpson P, Marchant J, Leidecker O, Kraatz S, Hay RT, Matthews SJ (2014) Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. Nat Commun 5:4217
23. Hecker CM, Rabiller M, Haglund K, Bayer P, Dikic I (2006) Specification of SUMO1- and SUMO2-interacting motifs. J Biol Chem 281:16117–16127
24. Kerscher O (2007) SUMO junction-what's your function? New insights through SUMO-interacting motifs. EMBO Rep 8:550–555
25. Lin DY, Huang YS, Jeng JC, Kuo HY, Chang CC, Chao TT, Ho CC, Chen YC, Lin TP, Fang HI, Hung CC, Suen CS, Hwang MJ, Chang KS, Maul GG, Shih HM (2006) Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. Mol Cell 24:341–354
26. Stehmeier P, Muller S (2009) Phospho-regulated SUMO interaction modules connect the SUMO system to CK2 signaling. Mol Cell 33:400–409
27. Chang PC, Izumiya Y, Wu CY, Fitzgerald LD, Campbell M, Ellison TJ, Lam KS, Luciw PA, Kung HJ (2010) Kaposi's sarcoma-associated herpesvirus (KSHV) encodes a SUMO E3 ligase that is SIM-dependent and SUMO-2/3-specific. J Biol Chem 285:5266–5273
28. Meulmeester E, Kunze M, Hsiao HH, Urlaub H, Melchior F (2008) Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. Mol Cell 30:610–619
29. Cai Q, Cai S, Zhu, C, Verma, SC, Choi JY, Robertson ES (2013) A Unique SUMO-2-Interacting Motif within LANA Is Essential for KSHV Latency. PLoS Pathog 9:e1003750
30. Makhnevych T, Sydorskyy Y, Xin X, Srikumar T, Vizeacoumar FJ, Jeram SM, Li Z, Bahr S, Andrews BJ, Boone C, Raught B (2009) Global map of SUMO function revealed by protein-protein interaction and genetic networks. Mol Cell 33:124–135
31. Sung MK, Lim G, Yi DG, Chang YJ, Yang EB, Lee K, Huh WK (2013) Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. Genome Res 23:736–746
32. Srikumar T, Lewicki MC, Raught B (2013) A global S. cerevisiae small ubiquitin-related modifier (SUMO) system interactome. Mol Syst Biol 9:668
33. Hannich JT, Lewis A, Kroetz MB, Li SJ, Heide H, Emili A, Hochstrasser M (2005) Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. J Biol Chem 280:4102–4110
34. Armstrong AA, Mohideen F, Lima CD (2012) Recognition of SUMO-modified PCNA requires tandem receptor motifs in Srs2. Nature 483:59–63
35. Kolesar P, Sarangi P, Altmannova V, Zhao X, Krejci L (2012) Dual roles of the SUMO-interacting motif in the regulation of Srs2 sumoylation. Nucleic Acids Res 40:7831–7843
36. Gilbreth RN, Truong K, Madu I, Koide A, Wojcik JB, Li NS, Piccirilli JA, Chen Y, Koide S (2011) Isoform-specific monobody inhibitors of small ubiquitin-related modifiers engineered using structure-guided library design. Proc Natl Acad Sci U S A 108:7751–7756
37. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190
38. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
39. Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell 5:865–876
40. Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2012) AMBER 12. University of California, San Francisco
41. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling CL (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
42. Jardin C, Horn AHC, Schürer G, Sticht H (2008) Insight into the phosphoryl transfer of the Escherichia coli glucose phosphotransferase system from QM/MM simulations. J Phys Chem B 112:13391–13400
43. Mazumder ED, Jardin C, Vogel B, Heck E, Scholz B, Lengenfelder D, Sticht H, Ensser A (2012) A molecular model for the differential activation of STAT3 and STAT6 by the herpesviral oncoprotein tip. PLoS One 7:e34306
44. Jardin C, Sticht H (2012) Identification of the structural features that mediate binding specificity in the recognition of STAT proteins by dual-specificity phosphatases. J Biomol Struct Dyn 29:777–792
45. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Physiol 23:327–341
46. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33–38
47. Accelrys Software Inc (2004) Discovery Studio Modeling Environment, Release 2.1. Accelrys Inc, San Diego
48. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Kruger D, Grebnev G, Kuban M, Strumillo M, Uyar B, Budd A, Altenberg B, Seiler M, Chemes LB, Glavina J, Sanchez IE, Diella F, Gibson TJ (2014) The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res 42:D259–266
49. Escobar-Cabrera E, Okon M, Lau DK, Dart CF, Bonvin AM, McIntosh LP (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
50. Selzer T, Albeck S, Schreiber G (2000) Rational design of faster associating and tighter binding protein complexes. Nat Struct Biol 7:537–541
51. Kiel C, Selzer T, Shaul Y, Schreiber G, Herrmann C (2004) Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proc Natl Acad Sci U S A 101:9223–9228