A bioinformatical approach for a reliable determination of short motifs for SUMO and Atg8 interaction in *Saccharomyces cerevisiae*

Inaugural–Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Diplom-Chemiker

Benjamin Vogt

aus Bonn

April 2014

1. Berichterstatter: Professor Dr. Kay Hofmann

2. Berichterstatter: Professor Dr. R. Jürgen Dohmen

Tag der mündlichen Prüfung: 10. April 2014

meinem lieben Großvater

Zusammenfassung

Regulatorische Prozesse werden durch posttranslationale Proteinmodifikation eingeleitet, die seine Aktivität, seine Stabilität, seine Lokalisation innerhalb der Zelle oder seine Interaktion mit anderen Proteinen verändern. Neben vielen anderer solcher Prozesse stellen die Dekodierung des SUMOylierungssignals oder die Erkennung von lipidgebundenem Atg8 Ausgangspunkte für effektive Verarbeitungsrouten dar, die durch SUMO-Interaktionsmotive (SIMs) und Atg8-Interaktionsmotive (AIMs) in Proteinsequenzen eingeleitet werden. Ihr geringer Informationsgehalt verhindert jedoch ihre Unterscheidung von willkürlich auftretenden Sequenzen. Diese Dissertation handelt von einem Detektionsansatz mit bioinformatischen Methoden für bisher unbekannte SIMs und AIMs.

Der erste Teil dieser Arbeit beschreibt die bioinformatische SIM-Detektionsmethodik. Diese ist für alle drei SIM-Typen gleich, jedoch sind die einzelnen Suchläufe auf SIMa-, SIMb- und SIMr-Charakteristiken zugeschnitten. Als Maß für die Konservierung wurde eine informationstheoretische Metrik gewählt und auf jeweils zwei unterschiedlich weit gefasste phylogenetische Verwandtschaftsbereiche angewendet. Eine Kombination bioinformatischer Methoden dient zur Abschätzung, ob ein Proteinbereich unstrukturiert, nicht-globulär oder globulär ist. Der vorgestellte bioinformatische Ansatz verwendet Eigenschaften zu Konservierung und strukturellem Kontext für eine Wertungsmethodik zur Funktionalität neuer, noch unbekannter SIM-Instanzen.

Experimente zeigen SUMO-Interaktion für Dbp10, Drs1, Rfc1, Rad18 und Tdp1 aus den bioinformatischen SIM-Detektionsdurchläufen. Dbp10 und Drs1 treten in der ribosomalen Biogenese im Nukleolus auf. Ein SIM ist in diesem biologischen Kontext noch nicht bekannt, obwohl SUMOylierung hier im Zusammenhang mit dem Übergang präribosomaler Partikel ins Nukleoplasma auftritt. Tdp1, Rfc1 und Rad18 treten im Zusammenhang mit DNS-Replikation und in der DNS-Schadensreparatur auf, in denen SUMOylierung als wichtiger Aktivitätsfaktor für andere Proteine bekannt ist. Für das Motiv in Rfc1 konnte im Rahmen dieser Arbeit eine Verantwortlichkeit für SUMO-Bindung in Mutationsstudien gezeigt werden. Es verursacht auch einen beobachtbaren Wachstumsphänotyp unter chemisch-induziertem DNS-Schadensstress. Das Motiv in Rad18 wurde in der Zwischenzeit auch von Parker und Ulrich identifiziert.

Im zweiten Teil dieser Arbeit wird eine Anwendung analoger Methoden für einen bioinformatischen AIM-Detektionsansatz beschrieben. Die Konservierungscharaktika etablierter AIM-Instanzen sind ähnlich zu denen von etablierten SIMs, wobei der strukturelle Kontext für AIMs mit bioinformatischen Methoden schwerer abzubilden ist.

Abstract

Regulatory processes are initiated by posttranslational modifications of proteins which alter their activity, stability, localization or their interaction with other proteins. Among many other processes, decoding the SUMOylation signal or the recognition of lipidated Atg8 represent starting points to effective downstream signalling pathways, initiated by SUMO interacting motifs (SIMs) and Atg8 interacting motifs (AIMs) in protein sequences. The low information contents of SIMs and AIMs prevent their detection from spurious sequences. This thesis is about a detection approach for so far unknown SIMs and AIMs with bioinformatical methods.

The first part of this thesis describes the bioinformatical SIM detection method. The overall method is common for all SIM types, whereas the single SIM detection screens apply to the characteristics for SIMa, SIMb and SIMr. Two sets of phylogenetic distances from budding yeast in combination with information theoretical approaches and sliding averages serve as a conservation measure. A combination of bioinformatical tools is used for an estimation whether a protein segment is unstructured, non-globular or globular. The bioinformatical approach in this thesis uses these conservation and structural features for the evolution of a functionality scoring measure for unknown SIM instances.

Experimental interaction studies show so far unknown SUMO interaction for Dbp10, Drs1, Rfc1, Rad18 and Tdp1 from the bioinformatical SIM detection screens. Dbp10 and Drs1 are involved in ribosomal biogenesis in the nucleolus. A SIM in this biological context has not yet been reported, whereas SUMOylation is involved in the release of pre-ribosomal particles into the nucleoplasm. Tdp1, Rfc1 and Rad18 are involved in DNA replication and damage repair, where SUMOylation is a crucial activity factor for other proteins. The motif in Rfc1 identified from the bioinformatical detection screen is shown responsible for SUMO interaction in mutation studies. This motif also causes observable growth phenotype under chemically induced DNA damage stress. The motif in Rad18 was meanwhile identified by Parker and Ulrich.

The second part of this study describes the application of analogous methods for a bioinformatical AIM detection approach. The conservation characteristics of established AIM are found similar to those for SIMs, whereas the structural context is harder to be represented with bioinformatical methods.

Contents

Contents

Contents

List of Figures

List of Tables

List of Tables

1. Introduction

Cells constantly have to overcome diverse situations that endanger cellular homeostasis: Starvation, damage to cell compartments, or just any disturbance in the sensitive balance of protein levels have to be detected and responded to by the cell.

The ubiquitin proteasome system (UPS) and ubiquitin-like (Ubl) protein conjugation pathways are integral to regulated proteolysis in the cell, but they also have essential multiple nondegradative biological roles. The UPS regulates location and activity of cellular proteins. Their function and malfunction are important factors in various human diseases such as cancer or neurodegenerative disorders [Bedford et al., 2011, review]. Breakdown of bulk proteins, carbohydrates and lipids constantly replenishes the cell's supply for carbon and nitrogen under starvation conditions. The detection of nonfunctional substrates and subsequent increase in repair mechanisms may be sufficient for the adaptation to a stress situation. Several repair pathways exist. But depending on the kind and severity of damage to the cell more drastic measures like apoptosis or senescence can be triggered to protect the tissue or the whole organism.

1.1. Posttranslational protein modification

Posttranslational modifications of proteins (PTMs) drive a variety of cellular processes in eukaryotes like regulation of cell growth, cell division, or adaptive and developmental processes. Protein modifications are covalent attachments which alter its properties. Most modifications make major structural contributions to their target proteins, as it is the case for phosphorylation, acetylation or methylation [Sartorelli et al., 1999, Ubersax and Ferrell, 2007, Polevoda and Sherman, 2007, Webster et al., 2014]. These modifications affect the activity, stability, localization or the interaction of proteins with other proteins. There are also other types of PTMs such as modification with ubiquitin-like proteins (Ubls) such as ubiquitin (Ub), the small ubiquitin-like modifier (SUMO) or the neural precursor cell-expressed, developmentally downregulated gene 8 (NEDD8). Modifications by Ubls introduce a new interaction site for proteins. Ubls share a three-dimensional structure that resembles ubiquitin in its characteristic tightly

packed globular $\beta\beta\alpha\beta\beta$-fold, also termed as "$\beta$-grasp", with β-sheets wrapped around an α-helix, whereas both primary sequence and charge surface distributions are highly diverse [Hochstrasser, 2000, Burroughs et al., 2007]. The respective Ubl structures are virtually superimposable (figure 1.1) [Bayer et al., 1998]. The evolutionary relationship of different Ubls suggest functional similarities, as all Ubls described so far are conjugated via similar conjugation machineries. These machineries employ E1, E2 and E3 enzymes on their way onto their respective target.

| (a) Ubiquitin | (b) Atg8 | (c) Smt3 | (d) NEDD8 |
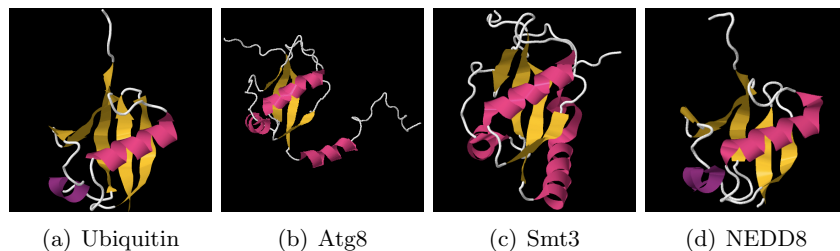
Figure 1.1. The ubiquitin fold as a structural element for ubiquitin-like modifiers. The β-grasp fold of ubiquitin (a) is a common structural core element for Ubls such as Atg8 [1UBI], Smt3 [2KQ7] and NEDD8 [2KO3]. Pictures were taken from the Protein Data Bank with the assigned identifiers. All structures were directly processed on the PDB web page using the "Jmol" application and the "screen shot" application. The structures were orientated in a similar conformation with the α-helix (bright purple) of the characteristic ubiquitin-like fold in front of the β-strands (yellow).

1.2. The ubiquitin conjugation machinery

Ubiquitin is a 76-residue protein, highly conserved within eukaryotes and found in all tissues of eukaryotic organisms, only with a sequence difference of three amino acids between plants, yeast and mammals [Hershko and Ciechanover, 1998]. It shares an almost identical conjugation process with its homologs in other species [Kerscher et al., 2006]. Ubiquitin is first expressed in an inactive form and needs processing at its carboxy terminus to expose a double glycine motif. After maturation of its C-terminus, ubiquitin is activated by an activating enzyme (E1) which catalyzes the formation of a high-energy thioester bond between its cysteine group and the carboxy terminus of ubiquitin [Haas et al., 1982]. The E1 acts as a catalyst and passes the modifier via transthiolation to the active cysteine side chain of a conjugating enzyme (E2) [Hershko et al., 1983, Hershko and Ciechanover, 1998]. Together with a substrate-specific E3 enzyme, the E2 catalyses ubiquitin transfer onto the target protein. The complexity of this posttranslational modification through ubiquitin is even increased by the fact that ubiquitin can form distinct polyubiquitin chains via seven lysine residues (K6, K11, K27, K29, K33, K48, K63)

or the N-terminus [Yeh, 2009]. This allows fine control of numerous cellular pathways: K48-polyubiquitinated proteins are generally degraded by the 26S proteasome, while K63-polyubiquitinated proteins can target membrane proteins to the lysosomal degradation pathway. Whereas the chain topology is determined by the E2 enzyme, substrate specificity comes with the E3 ligases [Yeh, 2009]. In this final step, the E3 ligase promotes the transfer of ubiquitin onto a lysine of its substrate. There are three classes of ubiquitin E3 ligases with regard to sequence and mechanism: E3 enzymes either contain a HECT-type (homologous to E6AP carboxyl terminus), a RING-type (really interesting new gene) or U-box domain [Barlow et al., 1994, Huibregtse, 1995, Borden et al., 1995, Hatakeyama et al., 2001, Jiang et al., 2001, Pringa et al., 2001]. The conserved HECT domain comprises ~ 350 amino acids with a conserved active site cysteine and is located at the E3 carboxy terminus. RING domains do not carry a covalently bound ubiquitin but rather serve as a scaffold to bring E2 and substrate in close proximity for transthiolation. A U-box domain forms a similar E2 binding surface as the RING domain. It lacks zinc ions, but is able to form stabilizing hydrogen bonds and salt bridges. The UPS pathway is initiated by the covalent attachment of ubiquitin to a target protein's amino group via the conjugation pathway [Pickart, 2001, Pickart and Eddins, 2004]. Ubiquitin is recognized by a number of pre-folded ubiquitin-binding domains (UBD) for substrate recognition. They are classified as UBA (ubiquitin-associated motif), UIM (ubiquitin interacting motif), UMI (UIM- and MIU-related UBD), DUIM (double-sided UIM), MIU (motif interacting with ubiquitin, or reversely orientated UIM), CUE (coupling of ubiquitin conjugation to endoplasmatic reticulum degradation), UEV (ubiquitin E2 variant), UBZ (ubiquitin-binding zinc finger) and UBM [Wilkinson et al., 2001, Mueller and Feigon, 2002, Hofmann and Falquet, 2001, Oldham et al., 2002, Pornillos et al., 2002, Swanson et al., 2003, Donaldson et al., 2003, Kang et al., 2003, Prag et al., 2003, Shih et al., 2003, Sundquist et al., 2004, Bienko et al., 2005, Penengo et al., 2006, Hirano et al., 2006, Pinato et al., 2011, Burschowsky et al., 2011]. They all have in common that they are composed of one or more α-helices binding to a hydrophobic patch on ubiquitin around Ile44. The general process of how ubiquitin and other Ubls get conjugated onto a substrate is well understood, however, the characteristics of their respective recognition sites in specific substrates and therefore the target specificity of a given modifier are still not known.

1.3. The Small Ubiquitin-like Modifier in Saccharomyces cerevisiae (SUMO)

SUMO proteins are expressed in all eukaryotes, but while *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* have one SUMO gene, organisms such as plants and vertebrates have several genes. It encodes a 101 amino acid polypeptide with 18% sequence identity to ubiquitin. The human genome encodes four SUMOs with SUMO1-3 expressed ubiquitously and SUMO4 mainly expressed in kidney, lymph node and spleen [Guo et al., 2004]. SUMO was shown to be a reversible modifier of proteins. Its attachment, called SUMOylation, can alter the protein localization by altering protein interactions. SUMOylation is a modification with important functions ranging from DNA damage control to regulation of mitochondrial dynamics [Livnat-Levanon and Glickman, 2011]. The *Saccharomyces cerevisiae* orthologue SMT3 of the mammalian SUMO was discovered in a genetic suppressor screen for the centromeric protein Mif2 [Meluh and Koshland, 1995, Mannen et al., 1996]. Additionally, SUMO was found to be covalently attached to the RanGTPase activating protein RanGAP1 and as binding partner for human Rad51 and Rad52 [Shen et al., 1996, Matunis et al., 1996, Mahajan et al., 1997].

1.3.1. SUMO conjugation

In yeast, the activation of SUMO is carried out by the heterodimeric protein complex Aos1-Uba2, after maturation of the otherwise unfunctional SUMO to expose the C-terminal GG motif [Johnson et al., 1997, Desterro et al., 1999, Gong et al., 1999, Okuma et al., 1999, Johnson, 2004]. Aos1 (activation of Smt3) resembles the N-terminus of Uba1, the ubiquitin activating enzyme (E1), whereas Uba2 shows structural homology to the C-terminus of Uba1. This close structure/function relationship shows how strongly related the conjugation machineries of ubiquitin and SUMO are. Similar to ubiquitin, the activation of SUMO is ATP-dependent and leads to a SUMO-adenylate conjugate which is an intermediate species in the thioester formation between SUMO and the active site cysteine in Uba2 within the Aos1–Uba2 complex [Dohmen et al., 1995, Johnson et al., 1997]. Upon activation, SUMO is transferred from the Aos1–Uba2 complex to a cysteine of the essential E2 conjugating enzyme Ubc9 [Johnson et al., 1997, Desterro et al., 1997, Schwarz et al., 1998]. In this step, Ubc9 binds both SUMO and the Aos1–Uba2 enzyme complex which places the two active cysteine residues into close proximity for SUMO transfer [Desterro et al., 1999, Gong et al., 1999, Lee et al., 1998, Saitoh et al., 1998, Wang et al., 2007]. Structural data suggest that the multi-protein binding surface

of Ubc9 establishes contact to the ΨKxE motif of a given substrate (with Ψ a large hydrophobic amino acid, K the SUMO target lysine residue, x a wild-card amino acid and E a glutamic acid residue) [Sampson et al., 2001, Bernier-Villamor et al., 2002].

1.3.2. SUMO ligases

Like ubiquitin, SUMO also engages E3 ligases to assist in the final step of conjugation and to promote the transfer of SUMO to a lysine residue in the substrate. The number of SUMO E3 enzymes is smaller than the number of E3s of ubiquitin. As a characteristic, the largest group of SUMO E3 ligases bears a SP–RING motif, which resembles the RING domain in ubiquitin E3 ligases [Hochstrasser, 2001]. SP–RING E3 ligases bind both their substrates and Ubc9 directly, and then position SUMO via a noncovalent SUMO interaction motif for a favorable SUMO-to-target orientation. The SP–RING ligases can be subdivided into different classes: The PIAS (protein inhibitor of activated STAT) protein family comprises a group of proteins which were initially described as negative regulators of cytokine signaling that inhibits STAT-transcription factors [Chung et al., 1997, Liu et al., 1998]. The PIAS-family with its yeast members Siz1 and Siz2 (SAP and miz-finger domain protein) and mammalian PIAS1, PIAS3 with splice variants PIASxα, PIASxβ and PIASy [Johnson and Gupta, 2001, Kahyo et al., 2001, Kotaja et al., 2002, Nakagawa and Yokosawa, 2002, Nishida and Yasuda, 2002, Sachdev et al., 2001, Schmidt and Müller, 2002]. Other SP–RING E3 ligases are Mms21 (methyl methanesulfonate sensitivity protein 21) and Zip3 [Roeder and Agarwal, 2000, Takahashi et al., 2001, Zhao and Blobel, 2005, Reindle et al., 2006]. Mms21 is part of an octametric Smc5–Smc6 complex essential for vegetative growth and DNA repair [Zhao and Blobel, 2005, Andrews et al., 2005, Potts and Yu, 2005]. Zip3 is part of the synapse-initiation complex [Cheng et al., 2006]. Human RanBP2 (Ran binding protein 2) is another SUMO E3 ligase, but without an SP–RING [Mahajan et al., 1997, Mahajan et al., 1998]. Its catalytic domain is located in a natively unfolded protein region assuming that the domain structure is established just upon binding to Ubc9 [Pichler et al., 2004, Reverter and Lima, 2005]. Further SUMO E3 ligases are the human Polycomb group member Pc2 (polycomb 2) [Kagey et al., 2003, Kagey et al., 2005], HDAC4 (Histone deacetylase 4) [Kirsh et al., 2002, David et al., 2002] and TOPORS (DNA topoisomerase I binding protein) [Weger et al., 2003, Weger et al., 2005].

A SUMO moiety attached to a target protein alters the protein shape and its surface charge distribution. The final effects of SUMOylation on a substrate may be numerous. SUMOylation may act as an ON-switch for interaction: only SUMOylated RanGAP1 interacts with RanBP2 and SUMOylated PCNA recruits yeast DNA helicase Srs2 to

replication forks [Matunis et al., 1996, Mahajan et al., 1997, Papouli et al., 2005, Pfander et al., 2005]. SUMOylation can also be an OFF-switch for interaction as it is the case for the transcription repressor ZNF76: Its SUMO-acceptor site interferes with its binding site for the TATA-binding protein and therefore, only one interaction is possible at a time [Zheng and Yang, 2004].

1.3.3. The SUMO interacting motif

SUMOylation proceeds via covalent modification of the target protein, whereas decoding of the SUMO signal depends on noncovalent interactions. Hecker *et al.* suggested a β-sheet of the SUMO-fold as SUMO/SIM interaction site [Hecker et al., 2006]. Minty and co-workers were the first to suggest a specific consensus pattern for SUMO interacting motifs (SIM) [Minty et al., 2000]. Minty *et al.* derived an 11-amino-acid stretch from sequence comparisons of the isolated proteins as an assumed SUMO-1 interaction site:

$$Pro - Ile - Ile - Leu - \underline{Ser} - Asp - \underline{Ser} - Glu - Glu - Glu - Glu$$
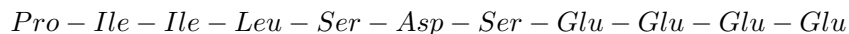
An "Ser–X–Ser" motif is preceded by hydrophobic residues and followed by acidic residues at its carboxy terminus. NMR studies by Song *et al.* drew the attention away from the serine residues in Minty's "Ser–X–Ser" motif as core element in SUMO-1 binding. They evolved a consensus pattern focused on a hydrophobic core [Song et al., 2004]:

$$\begin{pmatrix} Val \\ Ile \end{pmatrix} - X - \begin{pmatrix} Val \\ Ile \end{pmatrix} - \begin{pmatrix} Val \\ Ile \end{pmatrix}$$
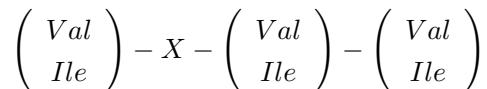
This motif has been found in several proteins like the SUMO ligases PIASX and RanBP2 and acts as binding site [Song et al., 2004, Song et al., 2005]. Solution structures of a SIM peptide in M–IR2 binding to SUMO1 (figure 1.2(a)) and a SIM peptide in MCAF1 binding to SUMO3 (figure 1.2(b)) indicate the same SUMO surface for SUMO/SIM interaction [Namanja et al., 2012, Sekiyama et al., 2008]. The same surface of the β-grasp fold is employed for binding in both structures.

Their structures and the works of Hannich *et al.* showed new SUMO interacting proteins like yeast Fir1, Nis1, Ris1 and Sap1 for a refined SIM consensus [Hannich et al., 2005]:

$$Lys - X_{3-5} - \begin{pmatrix} Val \\ Ile \end{pmatrix} - \begin{pmatrix} Leu \\ Ile \end{pmatrix} - \begin{pmatrix} Leu \\ Ile \end{pmatrix} - X_3 - \begin{pmatrix} Asp \\ Glu \\ Asn \\ Gln \end{pmatrix} - \begin{pmatrix} Asp \\ Glu \end{pmatrix} - \begin{pmatrix} Asp \\ Glu \end{pmatrix}$$

(a) SUMO1/M-IR2   (b) SUMO3/MCAF1
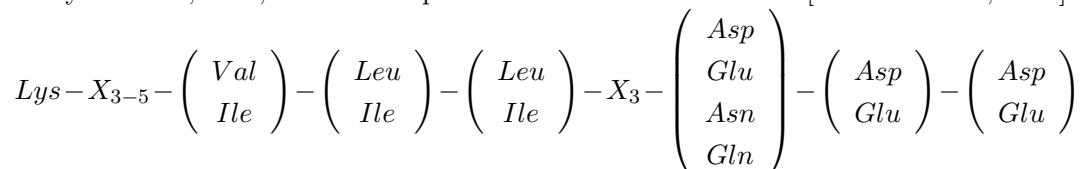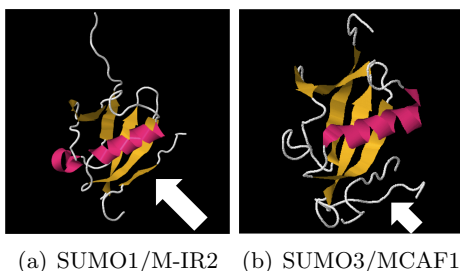
Figure 1.2. NMR solution structures show a functional binding site for noncovalent SUMO/SIM interaction. SIMs bind to a surface between α-helix (purple) and β-strand of the ubiquitin-like fold and extends to the β-sheet (both yellow). Depicted here are the following published structures: (a) A structural NMR study of SUMO1 in complex of a peptide comprising the SUMO1-specific M–IR2 SUMO interacting motif (SIM)[2LAS] [Namanja et al., 2012]. (b) A solution structure derived from NMR spectroscopy shows that the MCAF1 SIM employs the same surface of the SUMO fold [2RPQ] [Sekiyama et al., 2008]. The figures show the solution structures placed in a similar orientation to the observer to indicate common structural and binding characteristics. The SIM approaches the ubiquitin-like fold from a bottom right direction in a superimposable way. White arrows in the figures are pointed at the SIM. The arrow in figure (b) is drawn shorter not to cover the protein fold. All structures were taken from the Protein Database (PDB) web page using the "Jmol" application on that page. The images were taken by screen-shot from that application.

Additionally, it was reported that binding to SUMO can be performed in both sequence orientations, giving more variability in protein topology [Reverter and Lima, 2005, Song et al., 2005, Uzunova et al., 2007]. In 2007, Uzunova *et al.* performed yeast two-hybrid experiments with Smt3 as bait, in which they confirmed Siz1, Nis1, Fir1, Slx5 and Ris1 as SUMO interacting proteins (SIPs). They established a classification of SIMs into three types, 'SIMa', 'SIMb' and 'SIMr' [Uzunova et al., 2007]: A SIMa consensus sequence is composed of four hydrophobic amino acids followed by several acidic residues. The SIMr consensus sequence is like the one for SIMa, but with a reverse orientation of the sequence. The SIMb consensus sequence is composed of a shorter stretch of less variant hydrophobic residues. An acidic stretch at its carboxy terminus seems to be conserved, but less than for the other SIM consensus sequences. Experimental data helped to make a list of established SIMs (table 1.1). All motifs show close similarity to one of the three consensus types. Grouping these motifs in one of the three SIM types allows their information theoretical representation [Schneider et al., 1986, Schneider and Stephens, 1990]. These representations are based on information theory and display the residue variation in alignments of protein or nucleic sequences. A web-based application is provided by SeqLogo [Crooks et al., 2004]. The information theoretical SeqLogo approach can be applied to the list of established SIMs as another SIM representation aside from classical consensus pattern representations to each of the three SIM types SIMa, SIMb and SIMr (figure 1.2).

Table 1.1. Sequence characteristics of established SUMO interacting proteins.

Protein	Sequence motif	Reference(s)
Slx5	VILIDSDK	[Ii et al., 2007b, Xie et al., 2010]
	YVDLD	
Fir1	VILLDEDE	[Hannich et al., 2005, Uzunova et al., 2007]
Nis1	IIIPDSQD	[Hannich et al., 2005, Uzunova et al., 2007]
Uls1	IIILSDED	[Hannich et al., 2005, Uzunova et al., 2007]
	TIDLT	
Sap1	LIDLT	[Hannich et al., 2005]
Siz1	IIINLDSD	[Johnson and Gupta, 2001, Pichler et al., 2002, Uzunova et al., 2007]
Slx8	VLQISDD	[Uzunova et al., 2007, Sun et al., 2007]
Uba2	IVILDD	[Johnson et al., 1997]
Rad18	DDDLQIV	[Parker and Ulrich, 2012]
Elg1	QITIDD	[Parnas et al., 2010]
	DDLIVI	
	DDISII	
Srs2	IIVID	[Pfander et al., 2005, Kolesar et al., 2012], [Armstrong et al., 2012]
RanBP2	KKPEDSPSDDDVL	[Pichler et al., 2002, Hecker et al., 2006]
	DDVLIV	[Song et al., 2004]
PIASX	VIDLT	[Song et al., 2004]
PIAS1	VIDLT	[Hecker et al., 2006]
PIAS2	VIDLT	[Hecker et al., 2006]
PIAS3	VIDLT	[Hecker et al., 2006]
PIAS4	VVDLT	[Hecker et al., 2006]
TOPORS	VITIDS	[Weger et al., 2003]
Daxx	IIVLSD	[Lin et al., 2006]
MCAF1	VIDLT	[Sekiyama et al., 2008]
RNF4	IELVET	[Sun et al., 2007]
	VVDLT	
	VVIVDE	
	IVDLT	
Pc2	VILLSD	[Merrill et al., 2010, Yang and Sharrocks, 2010]
PML	VVVISS	[Minty et al., 2000]

Table 1.1. Sequence characteristics of established SUMO interacting proteins (continued).

Protein	Sequence motif	Reference(s)
Wss1	VVILDD	[Biggins et al., 2001, Uzunova et al.,
	VIDLT	2007, Mullen et al., 2010]
Rfp1	VIDLT	[Sun et al., 2007]
	IIDLD	
Rfp2	VIDLT	[Sun et al., 2007]
	LLDLT	

Table 1.2. SeqLogo representations (middle) and corresponding consensus patterns (right) of the three SUMO interacting motif types a, b and r.

SIMa		$\begin{pmatrix} Pro \\ Ile \\ Leu \\ Val \\ Met \end{pmatrix} - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \end{pmatrix} - X - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \end{pmatrix} - \begin{pmatrix} Asp \\ Glu \\ Ser \end{pmatrix}_3$
SIMb		$\begin{pmatrix} Pro \\ Ile \\ Leu \\ Val \\ Met \end{pmatrix} - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \end{pmatrix} - Asp - Leu - Thr$
SIMr		$\begin{pmatrix} Asp \\ Glu \\ Ser \end{pmatrix}_3 - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \end{pmatrix} - X - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \\ Pro \end{pmatrix} - \begin{pmatrix} Ile \\ Leu \\ Val \\ Met \\ Pro \end{pmatrix}$

1.3.4. SUMO-targeted ubiquitin ligases

Not only single SUMO moieties serve as targeting signals. Protein modification by SUMO chains represents recognition signals to a novel class of SUMO-targeted ubiquitin ligases (STUbLs) or ubiquitin ligases for SUMOylated proteins (ULS, E3-S). One function of this class of proteases is the STUbL-mediated ubiquitination of SUMOylated proteins as a signal for proteasomal degradation. For this purpose, STUbLs show two

characteristics: a RING domain for interaction with an E2 ubiquitin-conjugating enzyme and multiple SUMO interacting motifs for SUMO binding [Uzunova et al., 2007, Tatham et al., 2008].

The human Rnf4 protein (RING finger protein 4) is so far probably the best described STUbL [Prudden et al., 2007, Sun et al., 2007, Lallemand-Breitenbach et al., 2008, Tatham et al., 2008, Weisshaar et al., 2008]. Its architecture with at least three SUMO interacting motifs shows clear preference for binding of SUMO chains [Tatham et al., 2008]. A RING domain allows formation of a Rnf4 homodimer and binding of a single ubiquitin-charged E2 [Liew et al., 2010, Plechanovova et al., 2011, Plechanovova et al., 2012]. Several RNF4 substrates were found: Its ubiquitin ligase activity plays a crucial role in the ubiquitination of SUMOylated PML (promyelocytic leukemia protein) and in the disruption of PML nuclear bodies (PML–NB) in cells treated with arsenic trioxide (ATO) [Lallemand-Breitenbach et al., 2008, Tatham et al., 2008, Weisshaar et al., 2008]. ATO induces PML oligomerization and an increased affinity for Ubc9 [Zhang et al., 2010]. Poly-ubiquitination directs SUMOylated PML to the 26S proteasome for degradation [Tatham et al., 2008, Lallemand-Breitenbach et al., 2008].

In yeast, Slx5 is part of the DNA binding heterodimer Slx5–Slx8, also known as Uls2 [Wang et al., 2006, Yang et al., 2006, Ii et al., 2007a, Xie et al., 2007b, Uzunova et al., 2007]. This complex was also identified as STUbL from yeast two-hybrid data [Hannich et al., 2005, Uzunova et al., 2007, Xie et al., 2007b]. Both subunits contain a RING finger domain which was shown essential for dimer formation [Tatsuya Ii and Brill, 2007]. The Slx5 RING finger allows binding to Slx8, the Slx8 RING finger is the active ubiquitin ligase. Slx5 comprises several SIMs and is responsible for binding of SUMOylated proteins [Tatsuya Ii and Brill, 2007]. But Slx5–Slx8 can target substrates in a SUMO-independent manner [Xie et al., 2010]. The heterodimer is a ubiquitin ligase that links SUMOylation to recombinational DNA repair [Ii et al., 2007b, Xie et al., 2007b]. Slx5 and Slx8 are required for the viability of yeast cells lacking the Sgs1 DNA helicase [Kaliraman et al., 2001, Mullen et al., 2001]. *Slx5Δ* and *slx8Δ* null mutants show slow growth, sensitivity to hydroxy urea (HU) and increased rates of gross chromosomal rearrangements and mitotic recombination [Mullen et al., 2000, Ii et al., 2007a, Xie et al., 2007b]. Slx5 and Slx8 mutations are synthetically lethal when combined with mutations in the SUMO pathway [Wang et al., 2006]. SUMO interacting activity of Uls1 was found in a yeast two-hybrid screen [Hannich et al., 2005, Uzunova et al., 2007]. Uls1 is a ubiquitin ligase for SUMO conjugates. The Uls1 protein architecture comprises four SIMs in the amino terminal half, a Swi2/Snf2-like translocase motif and a RING domain in the carboxy terminal half [Dresser et al., 1997, Zhang and Buchman, 1997, Hannich et al., 2005, Uzunova

et al., 2007, Cal-Bakowska et al., 2011]. The combination of SIMs and RING domain suggests putative STUbL activity. Uls1 also binds to SUMO and SUMOylated proteins and shows interaction with Ubc4 ubiquitin ligase E2 in pulldown assays [Uzunova et al., 2007, Tan et al., 2013]. Uls1 mutant strains accumulate high molecular weight SUMO conjugates (HMW) and display synthetic growth effects [Uzunova et al., 2007]. These effects are increased in *uls1Δslx5Δ* or *uls1Δslx8Δ* double mutants [Pan et al., 2006, Uzunova et al., 2007]. Nevertheless, ubiquitin ligase activity of Uls1 has not yet been reported.

1.3.5. DeSUMOylation

For cell cycle progression and regulation of many processes, it is essential for a cell to be competent to remove the covalently attached SUMO from a given substrate at a given point of time. The de-modification is carried out by SUMO-specific proteases. This class of enzymes has two distinct functions: they C-terminally cleave the newly synthesized SUMO precursors to expose the di-glycine motif prior to conjugation, acting as so-called peptidases. They may also reverse SUMOylation by dissolving the SUMO isopeptide bond to its substrate and are therefore called isopeptidases. Both peptidases and isopeptidases are subclasses of hydrolases cleaving amide bonds by hydrolysis. Among these deSUMOylating enzymes are yeast Ulp1 and Ulp2 (Ubl specific protease) with a characteristic conserved approximate 200 amino acid C-terminal catalytic domain which is essential for the deSUMOylation activity [Li and Hochstrasser, 1999, Schwienhorst et al., 2000, Mossessova and Lima, 2000]. They have six human homologs SENP1–3, SENP5–7 (sentrin-specific proteases) [Mukhopadhyay and Dasso, 2007]. Ulp1 possesses cysteinyl proteinase activity. Additionally, Ulp1 deconjugates single SUMO moieties or SUMO chains from the ε-amino residue of the substrate. These functions are required for cell cycle progression in *Saccharomyces cerevisiae* [Li and Hochstrasser, 1999]. Ulp1 is SUMO specific. It lacks any sequence similarity to known ubiquitin deconjugating enzymes and is unable to deconjugate ubiquitin-targeted substrates. Ulp2, a second deSUMOylating proteinase, was found together with SUMO in the same screen for suppressors of a Mif2 mutation [Li and Hochstrasser, 2000]. Ulp2 was also found to have deSUMOylating activity and is located in the nucleus. It is not involved in SUMO maturation, but is required for chromosomal stability and for recovery from cell cycle checkpoint arrest [Li and Hochstrasser, 2000, Strunnikov et al., 2001, Bachant et al., 2002, Bylebyl et al., 2003, Felberbaum and Hochstrasser, 2008, Lee et al., 2011].

1.4. Autophagy

Autophagy plays a crucial role in the maintenance of a positive energy balance upon starvation stress [Kroemer et al., 2010, Ravikumar et al., 2010]. Dysfunctions in autophagic pathways are therefore associated with several diseases [Levine and Kroemer, 2008, Levine et al., 2011, Mizushima et al., 2008, Mizushima and Komatsu, 2011]. Autophagic pathways selectively remove aggregated proteins, surplus, damaged organelles and bacterial cells [Mizushima et al., 2008, Kirkin et al., 2009b, Noda and Yoshimori, 2009]. A crucial step is the formation of internal membranes in the cytoplasm which in turn form unique organelles, called autophagosomes [Baba et al., 1994, Kirisako et al., 1999, Suzuki et al., 2001, Mizushima and Komatsu, 2011]. Autophagosomes have been suggested to emerge from ER membranes [Hayashi-Nishino et al., 2009, Hayashi-Nishino et al., 2010]. More than 30 autophagy-related genes (Atgs) are involved in autophagy [Suzuki et al., 2007, Mizushima et al., 2011]. Most Atgs are involved in the following processes:

(i) The Atg1 kinase complex and its regulators [Matsuura et al., 1997, Cheong et al., 2008, Yeh et al., 2010, Kijanska and Peter, 2013]. Complex activity is enhanced upon starvation and plays a crucial role in autophagosome formation [Kamada et al., 2000]. It is composed of Atg1, Atg11, Atg13, Atg17, Atg20, Atg24, Atg29 and Atg31. Atg1 serves as a Ser/Thr kinase. It is directly connected to the Tor signaling pathway. Starvation induces dephosphorylation of Atg13. Atg13 now has a larger binding affinity to Atg1 increasing Atg1 kinase complex activity [Kamada et al., 2000, Funakoshi et al., 1997, Scott et al., 2000, Cheong et al., 2008].

(ii) A ternary Atg17–Atg29–Atg31 complex associates with Atg1. This also enhances Atg1 kinase complex activity. In a second step, the Atg17–Atg29–Atg31 complex recruits Atg proteins to the preautophagosomal structure (PAS) [Suzuki et al., 2001, Kawamata et al., 2008, Cheong et al., 2008]. Whereas Atg11 and Atg17 serve as a scaffold for PAS under nutrient-rich conditions, the Atg1–Atg13 complex associates with the Atg17–Atg29–Atg31 complex upon starvation [Shintani et al., 2002, Kawamata et al., 2008].

(iii) The phosphatidyl-inositol 3-kinase complex (PtdIns3) in *Saccharomyces cerevisiae* is Vps34 [Schu et al., 1993]. It forms a ternary Vps34–Atg6–Atg14 complex which is targeted to the PAS, whereas the Vps34–Atg11-Vps38 complex is required for vacuole protein sorting by the endosome [Kihara et al., 2001, Obara et al., 2006, Dove et al., 2004]. Atg14 recruits PtdIns3 binding proteins to the PAS, including Atg18. Atg18 in turn forms a Atg18–Atg2 complex which is responsible for Atg9 cycling between peripheral structures and the PAS [Noda et al., 2000, Reggiori et al., 2004, Suzuki et al., 2007].

(iv) Immunoelectron microscopy data suggest a central role for Atg8 in the autophagic pathway [Kirisako et al., 1999, Kirisako et al., 2000]. Atg8 is lipidated after ubiquitin-like conjugation to phosphatidylethanolamine into the autophagic membrane, a key process in autophagosome formation [Tanida et al., 2003, Kabeya et al., 2000, Kabeya et al., 2004].

1.4.1. The Atg8 conjugation cascade

Atg8 conjugation to phosphatidylethanolamine is dependent on two ubiquitin-like conjugation systems: Atg8 is attached to the lipid phosphatidylethanolamine (PE) via a ubiquitin-like conjugation mechanism. After proteolytic maturation cleavage by the protease Atg4, modified Atg8 is activated by the E1 enzyme Atg7 [Kirisako et al., 2000, Kim et al., 2001a]. Subsequently, activated Atg8 is transferred to the E2 enzyme Atg3 and finally conjugated to the amino group of the target lipid PE. This last step is catalyzed by an unusual E3 enzyme, which is a complex of Atg16 and Atg5, covalently modified by another ubiquitin-like modifier called Atg12 [Hanada et al., 2009]. Atg12 is activated by the E1 enzyme Atg7 and its specific E2 enzyme Atg10 for its covalent conjugation to Atg5. The Atg12-modified Atg5 associates with Atg16, thus forming the E3 for the transfer of Atg8 onto lipid–PE [Hanada and Ohsumi, 2005, Hanada et al., 2007, Kuma et al., 2002].

1.4.2. The Atg8 family proteins share a common fold

Atg8 with its central role in the autophagic system is the most prominent member of the Atg. Whereas only one ATG8 gene is known in yeast, there are six Atg8 homologs in mammals: LC3A, LC3B, LC3C, GABARAP, GABARAPL1, GATE-16/GABARAPL2 [Mann and Hammarback, 1994, Wang et al., 1999, Sagiv et al., 2000, Xin et al., 2001, He et al., 2003]. Their expressed proteins are here referred to as the Atg8 family proteins. The three-dimensional structures of Atg8 family proteins show a C-terminal ubiquitin fold and an N-terminal helical extension (figure 1.3). [Paz et al., 2000, Stangler et al., 2002, Sugawara et al., 2004, Noda et al., 2008, Duszenko et al., 2011]. The N-terminal region consists of two α-helices and is a unique feature of Atg8 that distinguishes it from other Ubls. All Atg8 family proteins have exposed β-strands, which are responsible for their interaction upon intermolecular β-sheet expansion. They all have two hydrophobic pockets, termed W-site (E17, I32, K48, L50, F104) and L-site (Y49, V51, P52, L55, F60, V63) [Noda et al., 2008]. The interaction between autophagic receptors and Atg8 contributes to specific cargo selection, whereas the AIM anchors the cargo with its

receptor to lipidated Atg8 [Ichimura et al., 2008, Kirkin et al., 2009a, Noda et al., 2008, Okamoto et al., 2009].

(a) Atg8　　　　(b) LC3　　　　(c) GABARAP
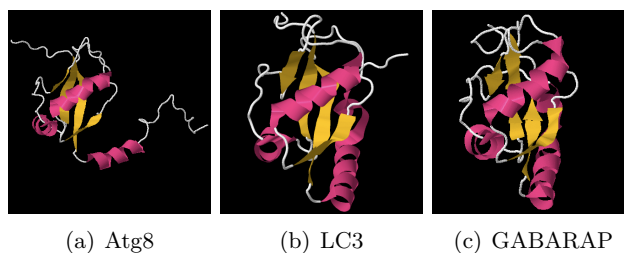
Figure 1.3. The solution structures of the Atg8 homologs show the characteristic β-grasp fold. The structures were taken from the Protein Database. Yeast Atg8 (a): 2KQ7 [Schwarten et al., 2010]. Mammalian LC3 (b): 1V49 [Kouno et al., 2005]. Mammalian GABARAP (c): 1KOT [Stangler et al., 2002]. All tertiary structures are placed in a similar orientation to display the common β-grasp fold with its β-strands (yellow) to the back and its characteristic α-helix (purple) to the front.

1.4.3. Substrate recognition and the Atg8 interacting motif

In autophagy, specific adaptor proteins link selective autophagy cargo to lipidated Atg8 family members and thus to the autophagy machinery. In yeast, such adaptors additionally bind Atg11 for both selective autophagy and the cytoplasm-to-vacuole targeting (CVT) pathway [Yorimitsu and Klionsky, 2005]. The CVT pathway is closely related to autophagy but serves as non-conventional targeting of Ams1 (α-mannosidase 1) and Ape1 (aminopeptidase 1) to the vacuole [Baba et al., 1997, Scott et al., 1997, Scott et al., 2001]. Atg19 and Atg34 mediate the incorporation of oligomerized structures of Ams1 and Ape1 in CVT vesicles [Watanabe et al., 2010]. Under both nutrient-rich and starvation conditions, Atg19 mediates the association of the CVT complex with the PAS via interaction with Atg8 and Atg11 [Hutchins and Klionsky, 2001, Shintani et al., 2002, Chang and Huang, 2007]. Under starvation conditions, Atg19 binding to Ams1 is no longer possible, which is then accomplished by the cargo receptor Atg34 [Watanabe et al., 2010, Suzuki et al., 2010]. The first Atg8 interacting motif (AIM) was described in the protein Atg19, which is the substrate recognition factor of the yeast CVT [Scott et al., 2001, Kim et al., 2001b]. The Atg19/Atg8 interaction could be attributed to a WEEL motif in Atg19. The Atg19 AIM adopts an extended β-conformation and an intermolecular parallel β-sheet with β2 of Atg8 (figure 1.4(a)) [Watanabe et al., 2010]. A similar site was found in Atg34 as an WEEI motif [Watanabe et al., 2010]. Atg11 binds cargo-receptors in an AIM-independent manner, implying that Atg8 and Atg11 use distinct binding sites in cargo receptors Atg19 and Atg34 [Okamoto et al., 2009].

Table 1.3. Characteristics of ubiquitin-like modifiers (Ubl) and respective enzymes.

Ubl	maturation by protease?	E1	E2	E3	conjugation product
ubiquitin	by E1 enzyme	Uba1[a], Uba6[b]	Ubc1–8, Ubc10, Ubc11	HECT[c]/U-Box[d]/RING[e] domain proteins	Ub-target
SUMO	Ulp1, Ulp2[f]	Aos1/Uba2[g]	Ubc9[h]	Siz1, Siz2[i]	SUMO-target
Atg8	Atg4[j]	Atg7[k]	Atg3[k]	Atg12–Atg5·Atg16[l]	Atg8–PE
Atg12	(none)	Atg7[m]	Atg10[n]	Atg5[o]	Atg12–Atg5[p]

[a] [McGrath et al., 1991]
[b] [Jin et al., 2007]
[c] [Huibregtse, 1995]
[d] [Hatakeyama et al., 2001, Jiang et al., 2001, Pringa et al., 2001, Murata et al., 2001]
[e] [Freemont et al., 1991, Barlow et al., 1994, Borden et al., 1995]
[f] [Mossessova and Lima, 2000]
[g] [Johnson et al., 1997, Gong et al., 1999]
[h] [Johnson and Blobel, 1997, Tatham et al., 2005]
[i] [Johnson and Gupta, 2001, Strunnikov et al., 2001]
[j] [Kirisako et al., 2000]
[k] [Ichimura et al., 2000]
[l] [Mizushima et al., 1999, Mizushima et al., 2003, Fujita et al., 2008]
[m] [Tanida et al., 1999]
[n] [Shintani et al., 1999]
[o] [Mizushima et al., 1998]
[p] [Mizushima et al., 1999, Kuma et al., 2002]

(a) Atg8/Atg19   (b) LC3/p62   (c) LC3B/OPTN   (d) GABARAP1/NBR1
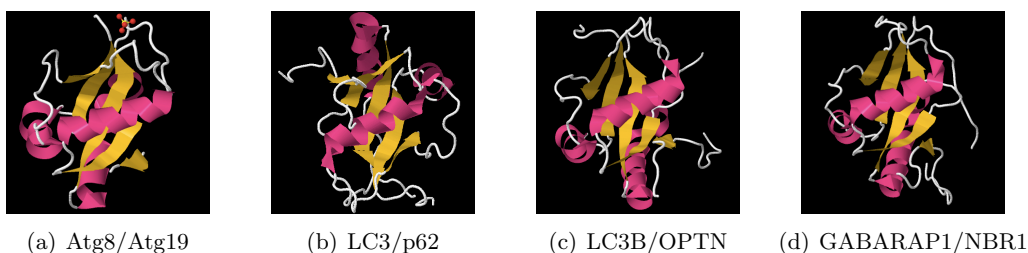
Figure 1.4. Solution structures of AIM (LIR in mammals) binding to different Atg8 homologs show a common binding site. The AIM (LIR) binding to a Atg8 homolog is structurally comparable to SIM binding to SUMO. It involves the same structural element with respect to the β-grasp fold, as can be seen from published structural data: Shown here are solution structures representing the interaction between Atg8/Atg19 (a, PDB 2ZPN), LC3/p62 (b, PDB 2K6Q), LC3B/OPTN (c, PDB 2LUE) and GABARAP1/NBR1 (d, PDB 2L8J). The respective structures are aligned to each other with respect to the ubiquitin-like structural core in a similar orientation with the β-strands to the back (yellow) and the characteristic α-helix to the front. The AIM approaches that fold at a cleft between that helix and the β-strands from a bottom right direction.

The Atg8 interaction of Atg19 is enhanced by concomitant binding to Atg11. There are other proteins bearing AIM. Atg32 is an outer-membrane, mitochondria-anchored receptor protein and bears an AIM for direct Atg8 binding [Kanki and Klionsky, 2008, Kanki et al., 2009, Okamoto et al., 2009, Aoki et al., 2011]. In mammals, a similar function is ascribed to Nix (mammalian Nip3-like protein, also known as BNIP3-like, BNIP3L) [Novak et al., 2010, Schwarten et al., 2009]. The pexophagy receptor Atg36 is shown to bind to Atg8 and the peroxisome via Pex3 [Motley et al., 2012, Farre et al., 2013]. Motley *et al.* suggested eight putative AIMs in Atg36 from sequence similarities to other AIMs, however the functional AIM could not be detected so far. A FDDI motif is better conserved than the other AIM candidates and thus might be functionally most important [Motley et al., 2012].

Other autophagic receptors responsible for the recognition of specific cargo have been described in mammals, including p62 [Bjørkøy et al., 2005, Komatsu et al., 2007, Pankiv et al., 2007, Ichimura et al., 2008] and NBR1 in mammals [Kirkin et al., 2009a, Waters et al., 2009]. p62 interacts with ubiquitinated proteins via its C-terminal UBA domain and oligomerizes with other "loaded" p62 via its N-terminal PB1 domain [Wilson et al., 2003, Moscat et al., 2007, Seibenhener et al., 2007, Saio et al., 2009, Nakamura et al., 2010, Isogai et al., 2011]. Docking to nucleating autophagosomes is mediated by the LC3 interacting motif (LIR in mammals, AIM in yeast) [Pankiv et al., 2007, Ichimura et al., 2008]. NDP52 was found as an receptor for ubiquitin-coated *Salmonella enterica serovar Typhimurium* [von Muhlinen et al., 2010, Thurston et al., 2009]. It has a carboxy-terminal zinc finger for ubiquitin binding and an LC3 interacting motif, char-

acteristics NDP52 shares with p62 and NRB1 [von Muhlinen et al., 2012]. Optineurin (OPTN) is an adaptor protein in the pathogen-induced autophagy pathway. OPTN has been shown to be an autophagy adaptor for ubiquitin-coated *Salmonella enterica* [Perrin et al., 2004, Wild et al., 2011]. It interacts with LC3/GABARAP via a LIR/AIM motif (figure 1.4(c)). Interestingly, Ser-170 within this LIR is phosphorylated by TBK1, making it more acidic and thus increasing its affinity for LC3/GABARAP [Kirkin et al., 2009a, Novak et al., 2010, Behrends et al., 2010]. A consensus pattern for AIM was evolved with experimental data (table 1.4, figure 1.5) [Alemu et al., 2012].

Table 1.4. Sequence characteristics of established Atg8 interacting proteins. The *Saccharomyces cerevisiae* protein sequences were used for bioinformatical pattern-based and profile-based screens.

Protein	Sequence	Reference(s)
Atg1	YVVV	[Yeh et al., 2010, Kraft et al., 2012]
Atg3	WEDL	[Hanada et al., 2009, Yamaguchi et al., 2010]
Atg4	YVDI	[Satoo et al., 2007, Fass et al., 2007]
Atg19	WEEL	[Shintani et al., 2002, Noda et al., 2008]
Atg32	WQAI	[Okamoto et al., 2009]
Atg34	WEEI	[Suzuki et al., 2010]
Nix	WVEL	[Schwarten et al., 2009, Novak et al., 2010]
NBR1	YIII	[Lamark et al., 2009, Rozenknop et al., 2011]
p62	WTHL	[Pankiv et al., 2007, Ichimura et al., 2008]

Figure 1.5. Information theoretical SeqLogo representation (left) and corresponding consensus pattern of the Atg8 family interaction motif.

$$\begin{pmatrix} Asp \\ Glu \end{pmatrix} - \begin{pmatrix} Asp \\ Glu \\ Ser \\ Thr \end{pmatrix} - \begin{pmatrix} Trp \\ Phe \\ Tyr \end{pmatrix} - \begin{pmatrix} Asp \\ Glu \\ Leu \\ Ile \\ Val \end{pmatrix} - X - \begin{pmatrix} Ile \\ Leu \\ Val \end{pmatrix}$$
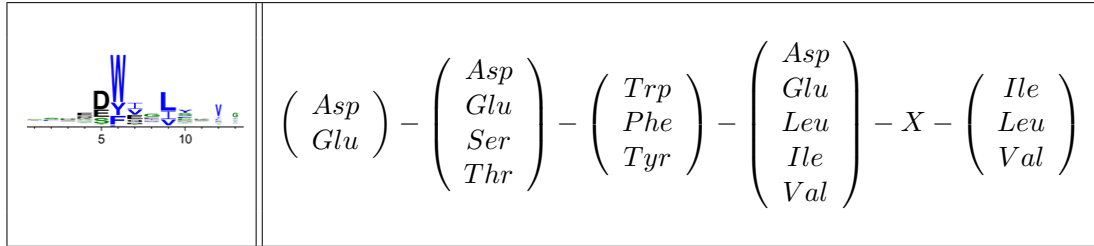
1.5. Comparison between Ubiquitin, SUMO and Atg8 interaction

The tightly packed globular $\beta\beta\alpha\beta\beta$ tertiary structure is a common characteristic for ubiquitin, SUMO and Atg8, despite their low degree of sequence similarity to each other [Vijay-Kumar et al., 1987, Bayer et al., 1998, Schwarten et al., 2010, Kumeta et al., 2010, Ganoth et al., 2013]. The differences in individual conjugation machineries are as follows: Whereas the basic principles of conjugation are the same with Ubls, the number of known representatives of each enzyme class E1, E2 and E3 varies. Ubiquitin employs numerous E2 enzymes and a large number of E3 ligases as a means of substrate specificity, SUMO has far less of each. Atg8 only involves one E1, E2 and E3 each. Substrate specificity comes from the cargo receptor. Ubiquitin, SUMO and Atg8 also differ in modifier recognition: Ubiquitin employs already pre-folded structural elements. Protein binding to SUMO and Atg8, on the contrary, proceed via small interaction motifs (SIM and AIM, respectively) forming β-strands upon Ubl binding (figures 1.2(a), 1.4(a)). These short binding motifs lie in structurally well solvent accessible, and therefore likely unstructured protein regions. These protein regions serve as a coupler between globular segments or lie close to sequence ending regions. In general, these sequence regions are not critical to function and show more sequence variation due to less evolutionary pressure. Functional short motifs of 3 to 10 amino acids in length in those protein segments show a distinct pattern of amino acid conservation. Multi-sequence alignment tools have difficulties in dealing with short conserved patches in otherwise more variant protein regions, as these tools are developed and trained on well-defined globular regions of more than 100 amino acids in length. A protein sequence database screen with a consensus motif pattern therefore gives a large number of potential motifs. An evaluation and distinction in true functional and incorrect occurrences due to the low information content of short linear motifs is a difficult computational problem.

1.6. Bioinformatics

Proteins usually fold natively into multiple sub-folding units, so called domains, rather than in one monolithic fold. These domains are able to act as intermolecular binding sites. A closer look into the three-dimensional structures and sequences of proteins shows: Hydrophobic amino acids are enriched within globular structures. Hydrophilic residues are found at sites of solvent contact — either on the outside of globular domains, or in unstructured linker regions.

The structure/function approach associates domains with particular functions such as

enzymatic activity or ligand recognition. This approach describes protein evolution as a number of single events with effects on domains. Autonomously folding domains evolve separately, as sequence changes in domains do not disturb the overall protein integrity. The comparative sequence analysis approach on proteomes from distinct species allows to derive functional information from already characterized sequences. Sequence homology uses the derived information as an anchor towards a transfer to novel sequences. In terms of evolution, homologs can be divided into orthologs and paralogs. Orthologous genes evolved from a single gene in the last common ancestor, paralogs arose from gene duplication [Ohno, 1970, Fitch, 1970]. Because inheritance represents the evolutionary relationship between orthologous genes, inherited functionality is therefore highly similar between those genes [Koonin, 2005, Schreiber and Sonnhammer, 2013].

Globular domains usually comprise more than 30 amino acids. Their information content is larger than for smaller protein segments. This allows distinct classification and storage in databases like Pfam [Bateman et al., 2004, Finn et al., 2014]. Protein regions shorter than about 20 amino acids show no natively folding substructure, as the internal stabilization is not sufficient. These regions are referred to as "motifs" or short linear motifs. Upon binding to a suitable binding partner, some of these motifs can adopt a tertiary structure. It often remains unclear, whether the folding or the binding step preceded. By contrast, short sequence modules such as motifs functioning independently from tertiary structure have been so far neglected.

The *Saccharomyces cerevisiae* genome has been completely sequenced [Goffeau et al., 1996]. The *Saccharomyces* Genome database (SGD)[1] stores more than 5,000 verified open reading frames (ORFs) as a center for different sequence data mining approaches [SGD, 2013]. That access to protein structures as well as sequence data sharpens the view on protein biology, as it allows both structure and sequence comparisons in combination with results from functional studies.

1.6.1. Short linear motifs

Traditionally, protein-protein interaction was believed to be restricted to large globular protein structures. Interaction between these distinct three-dimensional folded segments was thought to be the only way to fulfill a vast number of biological functions. In the last decades, scientific interest gets more and more focussed on highly flexible protein regions, especially highly conserved, but short sequence stretches within [Tompa et al., 2009, Gould et al., 2010]. These short linear motifs (SLiMs) are most abundant in

---

[1]data from October 16th, 2013, checked on October 17, 2013

natively disordered proteins, but are also found in solvent accessible parts of globular domains, such as exposed loops [Davey et al., 2012, Fuxreiter et al., 2007]. The majority of SLiMs catalogued so far, are between 3 and 10 amino acids in length and in most cases with one or more highly variable residues [Neduva and Russell, 2005, Fuxreiter et al., 2007, Diella et al., 2008]. Their information content is therefore too poor to be statistically significant in protein sequence searches. True-positive SLiM prediction from sequence and alignment data is difficult to distinguish from randomly occurring false-positive motif candidates, as they occur in unstructured regions with a high degree of amino acid variation. A well-curated database for storing data on SLiMs is the eukaryotic linear motif resource (ELM), a computational resource developed by the European Molecular Biology Laboratory (EMBL). The ELM datasets are useful for bioinformatical characterizations and predictions for established and new short linear motifs. ELM was developed on the basis of the PROSITE data [Bairoch, 1991].

2. Materials

2.1. Strains

Table 2.1. Yeast strains used in this study.

Name	Relevant genotype	Reference
AH109	MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3 GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2 URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ	[James et al., 1996]
JD24	URA3	Dohmen, collection
JD47–13C	MATα leu2-Δ1 trp1-Δ63 his3-Δ200 ura3-52 lys2-801 ade2-101	[Dohmen et al., 1995]
JD51	MATa/α leu2-Δ1/leu2-Δ1 trp1Δ63/trp1-Δ63 his3-Δ200/his3-Δ200 ura3-52/ura3-52 lys2-801/lys2-801 ade2-101/ade2-101	[Dohmen et al., 1995]
pJ69–4A	MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ GAL2-ADE2 LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ	[James et al., 1996]
YKU5	MATa NIS1-6HA::TRP1	Uzunova, PhD thesis

Table 2.2. Bacterial strains used in this study.

Strain	Genotype	Source
XL1 blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sub*E44 *rel*A1 lac[F' *pro*AB lacI^{q}ZΔM15 Tn10 (Tet^r)]	Stratgene, Cat No. 200249
XL10 gold	Tet^r Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 lac Hte [F' *pro*AB lacI^{q}ZΔM15 Tn10 (Tet^r) Amy Cam^r]	Stratgene, Cat No. 200314
BL21(DE3)	E. coli B F- dcam ompT hsdS(rB- mB-) gal *lambda*(DE3)(pLysS Camr)	Novagene
DH5α	F$^-$ ϕ80lacZΔM15 Δ(lacZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r$_k^-$, m$_k^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Invitrogen, Cat No. 18265-017
OneShot$^®$ *ccd*B Survival$^™$ 2 T1^R	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80lacZΔM15 ΔlacX74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *fhu*A::IS2	Invitrogen, Cat No. A10460

2.2. Plasmids

Table 2.3. Plasmids used in this study.

Plasmid	Description	Comment	Source/Reference
pBV1	P_{ADH1}-GAD-Sec27$^{930-2670}$-HA		this study
pBV2	P_{ADH1}-GAD-Rad5$^{1-1260}$-HA		this study
pBV3	P_{ADH1}-GAD-Rrb1$^{1-540}$-HA		this study
pBV4	P_{ADH1}-GAD-Rrb1$^{1-615}$-HA		this study
pBV5	P_{ADH1}-GAD-Dbp10$^{1-1529}$-HA		this study
pBV6	P_{ADH1}-GAD–Drs1$^{1-1731}$-HA		this study
pBV7	P_{ADH1}-GAD–Rfc1$^{1-705}$-HA		this study
pBV8	P_{ADH1}-GAD-Rad18$^{1-801}$-HA		this study
pBV9	Sec27-6HA		this study
pBV10	Rad5-6HA		this study
pBV11	Rrb1-6HA		this study
pBV12	Dbp10-6HA		this study
pBV13	Drs1-6HA		this study
pBV14	Rfc1-6HA		this study
pBV15	Rad18-6HA		this study
pGAD-C1	2μ / LEU2	two-hybrid vectors	[James et al., 1996]
pGBD-C1	2μ / TRP1		
pACT2™	2μ / LEU2		BD Biosciences Clontech
pD221™			BD Biosciences Clontech
	P_{ADH1}-GAD-Slx4		this study
	P_{ADH1}-GAD-Tdp1		this study
	P_{ADH1}-GAD-Top1		this study
pRS306	URA3		[Sikorski and Hieter, 1989]
pRS316	CEN/URA3		
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM1-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM2-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM3-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM1,2-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM1,3-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM2,3-HA			this study
P_{ADH1}-GAD-Rfc1$^{1-705}$ΔSIM1,2,3-HA			this study
	Rfc1ΔSIM:pRS306		this study
	Rfc1ΔSIM:pRS316		this study
	Rfc1ΔSIM:JD47-13C		this study

2.3. Oligonucleotides

Table 2.4. Oligonucleotides used in this study.

Name	Nucleotide sequence $5' \rightarrow 3'$	Description
MS3710	TCGACGCGGCCGCTTACCCATACGAT-GTTCCAGATTTACGCTTAA	GAD-HA (forward)
MS3711	GATCTTTAAGCGTAATCTGGACATCG-TATGGGTAAGCGGCCGCG	GAD-HA (reverse)
BV3712	CGCCCCGGGGTCGGTAAGCTAGTATG-GTCAGGT	Sec27 +930 (forward)
BV3713	CGCGCGGCCGCTTAACTCTCTTCTTC-CACAGGCTCCG	Sec27 $3'$-end (reverse)
BV3714	CGCCCCGGGATGAGTCATATTGAACA-GGAAGAA	Rad5 +1 (forward)
BV3715	CGCGCGGCCGCTAAAGATTCTGATGA-TTGTGCGGC	Rad5 +1260 (reverse)
BV3716	CGCCCCGGGATGTCGAAAGGTCTATC-GAGG	Rrb1 +1 (forward)
BV3717	CGCGCGGCCGCCAAAAGTGTTTTCGC-TAAATTAGA	Rrb1 +540 (reverse)
BV3718	CGCGCGGCCGCATTTTCATTCTCAAT-GACTGGATC	Rrb1 +615 (reverse)
BV3719	CGCCCCGGGGGGCTTCAAAAGCCGTA-GGTAA	Dbp10 +1 (forward)
BV3720	CGCGCGGCCGCTAATGTGTAATTAAT-AACATTGGC	Dbp10 +1529 (reverse)
BV3721	CGCCCCGGGATGGTGGTAGGAACTAA-AAAATAC	Drs1 +1 (forward)
BV3722	CGCGCGGCCGCACCTTCCCTACCAGC-TCTGGCGGT	Drs1 +1731 (reverse)
BV3723	CGCCCCGGGATGGTCAATATTTCTGA-TTTCTTTG	Rfc1 +1 (forward)
BV3724	CGCGCGGCCGCAGCAGGCATCCCAGC-AATTAATTG	Rfc1 +705 (reverse)
BV3727	CGCCCCGGGATGGACCACCAAATAAC-CACTGC	Rad18 +1 (forward)

Table 2.4. Oligonucleotides used in this study (continued).

Name	Nucleotide sequence $5' \rightarrow 3'$	Description
BV3728	CGCGCGGCCGCACTCATCATTGAGTT-TAAATACTT	Rad18 +801 (reverse)
BV3909	CCGGAGCAAGGAGAGGCAGTGCCGG-AGCCTGTGGAAGAAGAGAGTCGTACG-CTGCAGGTCGAC	Sec27 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3910	TTTCTCATTCGTTGATTAGTTCTTTAT-TTGTCTTTCTATGAGGTGGTTAATCG-ATGAATTCGAGCTCG	S2 + Sec27 $5'$-end (reverse) [Janke et al., 2004]
BV3911	TACGGTTTAGGAGATGATGCAGT	Sec27 +2523 (forward)
BV3912	GAGAGAAGAAAAAGGAGAATTGAAG-AAATCCAGATGCTGTTTGAACGTACG-CTGCAGGTCGAC	Rad5 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3913	GAAAATAATAATAAATAAAGTCTTTAT-ATATGAGTATGTGGTATGACTAATCG-ATGAATTCGAGCTCG	S2 + Rad5 $5'$-end (reverse) [Janke et al., 2004]
BV3914	TATAGCAGGCAAAAAATCCTATTACT	Rad5 +3199 (forward)
BV3915	AGTACCGGTACTGACGGTTTGAACGT-CTGGAAAACAATCAGTGTCCGTACGC-TGCAGGTCGAC	Rrb1 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3916	TTAATAATCTACTATGTGTAAATATAT-GATCCAAGTATTATACAGTTTAATCGA-TGAATTCGAGCTCG	S2 + Rrb1 $5'$-end (reverse) [Janke et al., 2004]
BV3917	TTGCTCAATATGACTTCCATAAGGG	Rrb1 +1229 (forward)
BV3918	AAAAAACGTGCGAAGAATGCTCGTCC-TTCCAAAAAGCGTAAATTTCGTACGC-TGCAGGTCGAC	Dbp10 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3919	ACATATGTAAACTATATATCTGTAGAT-CTAGATGGTTTTAAACTATCTAATCG-ATGAATTCGAGCTCG	S2 + Dbp10 $5'$-end (reverse) [Janke et al., 2004]
BV3920	ACCATCTAACTTGCTCGAAGACC	Dbp10 +2685 (forward)
BV3921	CAGTGATAAGTATGTTGATATTTTATG-ATAATGACGATGAAAATTATTAATCG-ATGAATTCGAGCTCG	S2 + Drs1 $5'$-end (reverse) [Janke et al., 2004]

Table 2.4. Oligonucleotides used in this study (continued).

Name	Nucleotide sequence $5' \rightarrow 3'$	Description
BV3922	AAGAGTACAAATTCAAATAAGAAGAA-GGGCTTCAAAAGCCGTAGGCGTACGC-TGCAGGTCGAC	Drs1 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3923	AATGCAATTAAGAAAGGGTGAAAATAT	Drs1 +1956 (forward)
BV3924	GCCACCAGTAAACCTGGTGGTAGCAA-AAAAAGGAAAACGAAAGCACGTACGC-TGCAGGTCGAC	S2 + Rfc1 $5'$-end (reverse) [Janke et al., 2004]
BV3925	ATCAATGAGAAGAAAGTGTAATTAT-AATCTTAGTGTATGAATAAATCAATCG-ATGAATTCGAGCTCG	Rfc1 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3926	GATACCAGCTACGGTTAAAAGTG	Rfc1 +2283 (forward)
BV3927	TTAATGGACTTGAATGAATATAGTAAA-GACCCACCCGGTAACAATCGTACGCT-GCAGGTCGAC	Rad18 $3'$-end + S3 (forward) [Janke et al., 2004]
BV3928	CAAATGTGCACAAGCTAACAAACAGG-CCTGATTACATATACACACCTTAATCG-ATGAATTCGAGCTCG	S2 + Rad18 $5'$-end (reverse) [Janke et al., 2004]
BV3929	GATGGTTATGTTTCGGAAGGATTTT	Rad18 +1161 (forward)
KH80	CGCGGTACCCTTCTATCCTTGATTCG-TATTCC	KpnI site at $5'$ Rfc1 promoter region[1]
KH81	CTAAACCAGAAGTTATCGACGCAGCT-ACTGAATCTGACCAAGAGTCGACCAA-TAAAACACCAAAG	Rfc1 SIMa mutagenese (forward)
KH82	CTTTGGTGTTTTATTGGTCGACTCTT-GGTCAGATTCAGTAGCTGCGTCGATA-ACTTCTGGTTTAG	Rfc1 SIMa mutagenese (reverse)
KH83	CGCGCGGCCGCAGCAGGCATCCCAGC-AATTAATTGCTTAAAG	NotI site at $3'$ Rfc1+705 (reverse)
KH84	TGGTCAGATTCAGTAGCTGCGTCGAT-AACTTCTGGTTTAGACGAA	GAD-Rfc1-HA SIMa mutagenese (reverse)
KH85	TTCGTCTAAACCAGAAGTTATCGACG-CAGCTACTGAATCTGACCA	GAD-Rfc1-HA SIMa mutagenese (forward)

---

[1] A sequence region of 1kb upstream of the coding sequence is regarded here "promoter region".

Table 2.4. Oligonucleotides used in this study (continued).

Name	Nucleotide sequence 5′ → 3′	Description
KH86	CGCGGTACCCTTCTATCCTTGATTCG-TATTCCAAAAAAACCTCT	KpnI at 5′ Rfc1 promoter region (forward)
KH87	CGATACCATCAGCCTCTGGATCTGTG-GACATTTTTGGAGGAA	Rfc1 +503 (forward)
KH88	AAGAAGGCCAAGAGTCATCCTCATGT-GGAGGACGTTGATGAAA	Rfc1 +610 (reverse)
KH89	CCCACTACAGACATGTTATCAAGAGC-GTTTTTAACACCGGCATTA	Rfc1 +2156 (reverse)
KH90	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGATTTCTGCTACTAATTG	Pwp1 with attB1 site (forward)
KH91	GGGGACCACTTTGTACAAGAAAGCTG-GGTTTATGCCATGTCATCATGCT	Pwp1 with attB2 site (reverse)
KH92	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGACTTCTGAAAATCCGGA	Prp6 with attB1 site (forward)
KH93	GGGGACCACTTTGTACAAGAAAGCTG-GGTTCACCTTTTAAATGACAGAT	Prp6 with attB2 site (reverse)
KH94	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGGATTCACAAAGAAGCCA	Sap1 with attB1 site (forward)
KH95	GGGGACCACTTTGTACAAGAAAGCTG-GGTTCATGAACCTGATGATCCGA	Sap1 with attB2 site (reverse)
KH96	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGGAACTTCAGAGGGCACA	Slx4 with attB1 site (forward)
KH97	GGGGACCACTTTGTACAAGAAAGCTG-GGTCTACTTTTTATCGTTTCTCA	Slx4 with attB2 site (reverse)
KH98	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGTCACGAGAAACAAATTT	Tdp1 with attB1 site (forward)
KH99	GGGGACCACTTTGTACAAGAAAGCTG-GGTTTAGTCGTTCTCATGACGAG	Tdp1 with attB2 site (reverse)
KH100	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGACTATTGCTGATGCTTC	Top1 with attB1 site (forward)
KH101	GGGGACCACTTTGTACAAGAAAGCTG-GGTTAAAACCTCCAATTTTCATC	Top1 with site attB2 (reverse)

Table 2.4. Oligonucleotides used in this study (continued).

Name	Nucleotide sequence $5' \rightarrow 3'$	Description
KH102	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGTCAGAGGGAAAAGTAGA	Ubx5 with attB1 site (forward)
KH103	GGGGACCACTTTGTACAAGAAAGCTG-GGTTTATTCTATTTCAGGGTCAA	Ubx5 with attB2 site (reverse)
KH104	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGCCAGTGTTGAAATCAGA	Glt1 with attB1 site (forward)
KH015	GGGGACCACTTTGTACAAGAAAGCTG-GGTTTAGACTTGACTAGCTAATT	Ifh1 with attB1 site (forward)
KH106	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGGCAGGCAAAAAAGTCC	Ifh1 with attB2 site (reverse)
KH107	GGGGACCACTTTGTACAAGAAAGCTG-GGTTCATTGTAAGACGTCATTGA	Prp14 with attB1 site (forward)
KH108	GGGGACAAGTTTGTACAAAAAAGCAG-GCTGGATGAGTAATTCACTTGAGGA	Prp14 with attB2 site (reverse)
KH109	GGGGACCACTTTGTACAAGAAAGCTG-GGTTTATTTAGGTCCTCCCTTCT	Rfc1 +705 NotI site (reverse)
KH110	GAGACGCGCGGCCGCAGCAGGCATCC-CAGCAATTAATTGCTTAAAG	KpnI at $5'$ promoter region of Rfc1 (forward)
KH111	GAGAGACGCGGTACCCTTCTATCCTT-GATTCGTATTCC	GAD-Rfc1-HA +1 mutagenesesis into MVNAAD (forward)
KH112	ACTAAGCTGAAGAAAATGGTCAATGC-TGCTGATTTCTTTGGTAAAAATAAGA-AATCCG	GAD-Rfc1-HA +1 mutagenesesis into MVNAAD (reverse)
KH113	CGGATTTCTTATTTTTACCAAAGAAA-TCAGCAGCATTGACCATTTTCTTCAG-CTTAG	GAD-Rfc1-HA +57 mutagenesis into VIDAAE (forward)
KH114	CCAAAGAAGATGCCTGTAAGTAATGT-AATTGATGCAGCAGAGACACCTGAAG-GAGAAAAAAG	GAD-Rfc1-HA +57 mutagenesis into VIDAAE (reverse)
KH115	CTTTTTTTCTCCTTCAGGTGTCTCTG-CTGCATCAATTACATTACTTACAGGCA-TCTTCTTTGG	GAD-Rfc1-HA +57 mutagenesis into VIDAAE (reverse)

Table 2.4. Oligonucleotides used in this study (continued).

Name	Nucleotide sequence $5' \rightarrow 3'$	Description
KH116	ATCGAATTCCCCGGGATGGTCAATGC-TGCTGATTTCTTTGGTAAAAATAAGA-AATCCG	GAD-Rfc1-HA +1 muta-genesesis into MVNAAD (forward)
KH117	CGGATTTCTTATTTTTACCAAAGAAA-TCAGCAGCATTGACCATCCCGGGGAA-TTCG	GAD-Rfc1-HA +1 muta-genesesis into MVNAAD (reverse)
KH206	AAAGCTGGGTACCCTTCTATC	Rfc1 Kpn1 at $5'$ promoter region (forward)
KH207	CAAAACGCTTGGTCTTGTAG	Rfc1 +1398 (reverse)
KH208	CTCTTCTGTACTTTCTTGCTC	Rfc1 +32 (forward)
KH209	CAGTATGACACATTATTGAGC	Rfc1 +251 (reverse)
KH210	CTTGTTCTTCATGAATGGTC	Rfc1 +767 (forward)
KH211	CTTGAAATTGATTTAGTAAC	Rfc1 +1479 (reverse)

2.4. Chemicals

Table 2.5. Chemicals used in this study.

Chemical	Company
Acetic acid	Merck, Darmstadt, Germany
Acrylamide	Serva, Heidelberg, Germany
Adenine	AppliChem, Darmstadt, Germany
Agar	Formedium, Hunstanton , UK
Agarose	Gibco, Paisley, Scotland
Ampicillin	Sigma-Aldrich, Munich, Germany
APS	Fluka, Deisenhofen, Germany
Arginine	Janssen, Beerse, Belgium
Bacto Tryptone	Formedium, Hunstanton, UK
Blotting grade blocker	BioRad, München, Germany
β-Mercaptoethanol	Roth, Karlsruhe, Germany
Bromophenol blue	Janssen, Beerse, Belgium
Disodium hydrogenphopsphate	Riedel de Haen, Seelze, Germany
Dimethyl sulfoxide	AppliChem, Darmstadt, Germany
Ebselen	Roth, Karlsruhe, Germany

Table 2.5. Chemicals used in this study (continued).

Chemical	Company
Ethanol	Baker, Deventer, The Netherlands
Ethidium bromide	Fluka, Deisenhofen, Germany
Ficoll	VWR, Darmstadt, Germany
FoA	Sigma-Aldrich, Munich, Germany
Formamide	Formedium, Hunstanton, UK
Galactose	Roth, Karlsruhe, Germany
Glutathion	Roth, Karlsruhe, Germany
Glutathion-Sepharose 4B	Amersham Bioscience, USA
Glucose	Caesar & Lortz, Hilden, Germany
Glycine	Serva, Heidelberg, Germany
Glycerol	Roth, Karlruhe, Germany
HEPES	Serva, Heidelberg, Germany
Histidine	Fluka, Deisenhofen, Germany
Hydroxy urea	Sigma-Aldrich, Munich, Germany
Isoleucine	Merk, Darmstadt, Germany
IPTG	Roth, Karlsruhe, Germany
Kanamycine	Sigma-Aldrich, Munich, Germany
Leucine	Roth, Karlsruhe, Germany
Lithium acetate	Sigma-Aldrich, Munich, Germany
Lysine	Roth, Karlsruhe, Germany
Lysogeny broth	Roth, Karlsruhe, Germany
Manganese(II) chloride	Roth, Karlsruhe, Germany
Manganese(II) sulfate	Roth, Karlsruhe, Germany
Methionine	Roth, Karlsruhe, Germany
Methanol	Baker, Deventer, The Netherlands
Methyl methanesulfonate	Sigma-Aldrich, Munich, Germany
PEG-3350	Roth, Karlsruhe, Germany
Peptone	Difco, Detroit, USA
Phenylalanine	Acros, Geel, Belgium
Potassium acetate	Acros, Geel, Belgium
Potassium chloride	Fluka, Deisenhofen, Germany
Potassium dihydrogenphosphate	Fluka, Deisenhofen, Germany
Protease inhibitor cocktail	Roche, Mannheim, Germany

Table 2.5. Chemicals used in this study (continued).

Chemical	Company
Sorbitol	Sigma-Aldrich, Munich, Germany
Sodium chloride	Baker, Deventer, The Netherlands
Sodium dodecyl sulfate (SDS)	Serva, Heidelberg, Germany
Sodium hydrogen diphosphate	Baker, Deventer, The Netherlands
TEMED	Sigma, Steinheim, Germany
Threonine	Roth, Karlsruhe, Germany
Tris	Roth, Karlsruhe, Germany
Triton-X-100	Serva, Heidelberg, Germany
Tryptophan	Roth, Karlsruhe, Germany
Tryptone	Roth, Karlsruhe, Germany
Tyrosine	Roth, Karlsruhe, Germany
Tween-20	VWR, Darmstadt, Germany
Uracil	Roth, Karlsruhe, Germany
Western blotting substrate	Roche, Mannheim, Germany
Yeast extract powder	Formedium, Hunstanton, UK
Yeast nitrogen base	Difco, Detroit, USA
X-Gal	Roth, Karlsruhe, Germany
Xylene cyanol	VWR, Darmstadt, Germany
1 kb DNA ladder	Fermentas, St. Leon-Rot, Germany
100 bp DNA ladder	Fermentas, St. Leon-Rot, Germany

2.5. Miscellaneous equipment

Table 2.6. Miscellaneous equipment used in this study.

Equipment	Company
glass beads (Ø 0.1–0.11 mm, for *Escherichia coli*)	Sartorius Stedim, Göttingen, Germany
glass beads (Ø 0.4–0.6 mm, for *Saccharomyces cerevisiae*)	Sartorius Stedim, Göttingen, Germany
calf tymus (CT-) DNA	Sigma-Aldrich, München, Germany
dNTP	Roche, Mannheim, Germany
PVDF membrane	Millipore, Zug, Switzerland
X-ray films	medical X-ray film SUPER RX, Japan

Table 2.6. Miscellaneous equipment used in this study (continued).

Equipment	Company
Chemiluminescence solution	Roche, Mannheim, Germany
protease inhibitor cocktail	Roche, Mannheim, Germany

2.6. Enzymes

Table 2.7. Enzymes used in this study.

Enzyme	Company
RedTaq PCR ready-mix	Sigma, St. Louis, USA
T4 DNA ligase	New England Biolabs, Ipswich, USA
Phusion High Fidelity Polymerase	New England Biolabs, Ipswich, USA
Antarctic Phosphatase	New England Biolabs, Ipswich, USA
SAP	New England Biolabs, Ipswich, USA
BP clonase	Sigma-Aldrich, Munich, Germany
LR clonase	Sigma-Aldrich, Munich, Germany

2.7. Antibodies

Table 2.8. Antibodies used in this study.

Antibody	Relevant characteristics	Derived from	Source
α-GST	polyclonal	rabbit	Santa Cruz Biotechnology, USA
α-GST-SUMO	polyclonal	rabbit	Santa Cruz Biotechnology, USA
α-HA 16B12	monoclonal	mouse	Covance, Berkeley, USA
α-HA 3F10	monoclonal	rat	Roche, Mannheim, Germany
α-Cdc11	polyclonal	rabbit	Santa Cruz Biotechnology, USA
α-Rat POD	horseraddish peroxidase coupled	sheep	Abacam Limited, Cambridge, UK
α-Mouse POD	horseraddish peroxidase coupled	goat	Dianova, Hamburg, Germany

Table 2.8. Antibodies used in this study (continued).

Antibody	Relevant characteristics	Derived from	Source
α-Rabbit POD	horseraddish peroxidase coupled	goat	Amersham Bioscience, USA

2.8. Media

2.8.1. Growth media for yeast

Media used for routine growth of Saccharomyces cerevisiae were "Yeast Peptone Dextrose" medium (YPD) medium, "Yeast Peptone Adenine Dextrose" (YPAD) medium, "Yeast Peptone Glycerol" (YPG) medium and "Yeast synthetic drop-out" (SD) medium.

Yeast complete medium "Yeast Peptone Dextrose" (YPD)

1 % yeast extract
2 % peptone
2 % glucose (dextrose)

Yeast complete medium "Yeast Peptone Adenine Dextrose" (YPAD)

1 % yeast extract
2 % peptone
0.004 % adenine hemisulfate
2 % glucose (dextrose)

Yeast complete medium "Yeast Peptone Glycerol" (YPG)

1 % yeast extract
2 % peptone
2 % glycerol

Yeast synthetic drop-out medium (SD) (complete)

6.7 $\frac{g}{L}$ yeast nitrogen base without amino acids	40 $\frac{mg}{L}$ lysine
20 $\frac{mg}{L}$ adenine	10 $\frac{mg}{L}$ methionine
20 $\frac{mg}{L}$ arginine	60 $\frac{mg}{L}$ phenylalanine
10 $\frac{mg}{L}$ histidine	50 $\frac{mg}{L}$ threonine
60 $\frac{mg}{L}$ isoleucine	40 $\frac{mg}{L}$ tryptophan
60 $\frac{mg}{L}$ leucine	2 % glucose

As stock solution for routine use were held:

50 % (w/v) glucose

250×	uracil
100×	histidine
100×	leucine
250×	tryptophan
500×	uracil

Solid growth media

For routine growth of Saccharomyces cerevisiae on agar plates, the respective medium is supplemented with 2 % agar.

2.8.2. Growth media for bacteria

Lysogeny Broth (LB) medium

0.5 %	Yeast extract
1 %	tryptone
1 %	sodium chloride

LB + Ampicillin medium

0.5 %	yeast extract
1 %	tryptone
1 %	sodium chloride
100 $\frac{\mu g}{mL}$	ampicillin, 100 $\frac{mg}{mL}$

LB + Kanamycin medium

0.5 %	yeast extract
1 %	tryptone
1 %	sodium chloride
50 $\frac{\mu g}{mL}$	kanamycin, 50 $\frac{mg}{mL}$

SOC medium

0.5 %	yeast extract
2 %	tryptone
10 mM	sodium chloride
2.5 mM	potassium chloride
10 mM	manganese(II) chloride
10 mM	manganese(II) sulfate
20 mM	glucose

2.9. Antibiotics

The following antibiotics were used when appropriate:

100 $\frac{mg}{mL}$ Ampicillin (1000× stock), used at 100 $\frac{\mu g}{mL}$

50 $\frac{mg}{mL}$ Kanamycin (1000× stock), used at 50 $\frac{\mu g}{mL}$

2.10. Protocol

Unless stated otherwise, *Saccharomyces cerevisiae* strains were grown at 30°C, *Escherichia coli* strains at 37°C.

3. Methods

3.1. Bioinformatical Methods

Bioinformatical analyses on *Saccharomyces cerevisiae* proteins were carried out using different tools provided by other groups in the internet (see table 3.1, 3.2 and 3.3). Data mining techniques and output data analyses were performed using shell and perl scripts generated and maintained by the author of this thesis.

3.1.1. Protein disorder prediction

Intrinsically unstructured or disordered proteins (IDPs or IUPs) are often referred to as to be natively unfolded and have a broad occurrence in living organisms. Though lacking specific three-dimensional structures — and instead, processing a large number of conformations under physiological conditions in the absence of a proper interaction partner — IUPs are involved in key biological processes such as cell cycle control, regulation, recognition, and signaling [Dunker et al., 2001]. These proteins are able to interact with and bind to a wide range of ligands, including partners such as proteins, membranes, and nucleic acids. Protein/protein interactions involve a combination of coupled binding and "folding-upon-binding" [Espinoza-Fonseca, 2009, Wright and Dyson, 2009]. Intrinsically unstructured regions have also been found to be important loci for alternative splicing [Romero et al., 2006] and for enzyme driven posttranslational modifications such as phosphorylation, methylation, or acetylation [Xie et al., 2007a, Gsponer et al., 2008]. While the opposite state of a protein being structured is rather simple to determine, the definition of being "disordered" or "unstructured" is still under discussion: A protein is "disordered" unless any structure is determined using methods such as X-ray crystallography, nuclear magnetic resonance (NMR), circular dichroism (CD), small-angle X-ray scattering. But absence of structural information is no good proof for an IUP. Therefore, there are different approaches to predict (and not verify) IUPs. Globular regions within *Saccharomyces cerevisiae* protein regions were predicted using locally installed versions of GlobPlot [Linding et al., 2003b] and IUPred [Dosztányi et al., 2005a, Dosztányi et al., 2005b]. Disorder prediction was performed using tools

35

such as DisEMBL [Linding et al., 2003a], RONN [Yang et al., 2005], FoldUnfold [Gar-buzynskiy et al., 2004], FoldIndex [Prilusky et al., 2005] and IUPred [Dosztányi et al., 2005a, Dosztányi et al., 2005b]. The latter features algorithms for both long-range and a short-range disorder prediction as well as one for globularity. Protein regions predicted to have no structural characteristic were referred to as unstructured. All tools were used according to the respective developers' manuals.

Table 3.1. Bioinformatical web based methods used in this study for the globularity and disorder prediction of proteins.

Name	uniform resource locator (URL)	Reference
GlobPlot	http://globplot.embl.de	[Linding et al., 2003b]
DisEMBL	http://dis.embl.de	[Linding et al., 2003a]
IUPred	http://iupred.enzim.hu	[Dosztányi et al., 2005a]
RONN	https://www.strubi.ox.ac.uk/RONN	[Yang et al., 2005]
FoldIndex	http://bip.weizmann.ac.il/fldbin/findex	[Prilusky et al., 2005]
FoldUnfold	http://www.bioinfo.protres.ru/ogu/	[Galzitskaya et al., 2006]

3.1.1.1. The GlobPlot method

IUPs generally have a biased amino acid composition, leading to the concept that secondary disorder prediction is based on propensities of each amino acid to occur in a distinct structural sequence context [Miyazawa and Jernigan, 1985, Sippl, 1990, Dunker et al., 2001, Linding et al., 2003b]. The idea behind IUPred is the hypothesis that the tendency for disorder can be described as

$$P = RC - SS$$

with the propensities for a given amino acid of beeing located rather in a 'random coil' (RC) or in any kind of 'secondary structured' (SS) region. These parameters were trained on protein data. These data sets comprised sequences of test bed proteins with determined secondary and random coil structure by Chou *et al.* [Chou and Fasman, 1974, Chou and Fasman, 1978], complemented by parameters [Deléage and Roux, 1987] and refined using data sets from the 'structural classification of proteins' (SCOP) database [Lo Conte et al., 2002a]. The frequencies RC and SS were calculated for each amino acid and published as Russel/Linding parameters. Additionally, Linding *et al.* developed a propensity scheme based on 'missing' or unresolved coordinates in proteins from the

protein data bank (PDB) [Westbrook et al., 2003]. Starting with the disorder propensity $P(a_i)$ of the ith amino acid a in a peptide of length L, a sum function Ω is introduced.

$$\Omega(a_i) = \sum_{j=1}^{i=1} \Omega(a_j) + \ln(i+1) \cdot P(a_i) \text{ with } i = 1, 2, ..., L$$

The next mathematic steps towards the "Russel-Linding algorithm" comprise low-pass Savitzky-Golay filter smoothing [Press and Teukolsky, 2002] over Ω and estimation of the first order derivative of the plot using the TISEAN 2.1 "Nonlinear Time Series Analysis" package[1] [Hegger et al., 1999]. The disordered and globular sequence position are selected from regions where the derivative shows positive or negative values, respectively, over a minimum length.

Whereas the globular domain (GlobDoms) prediction was benchmarked against data from the SMART server[2], the lack of a unique disorder definition led to different approaches for disorder prediction algorithms for GlobPlot, including:

- The **Russel/Linding** algorithm as stated above.

- **B-Factors**. These isotropic temperature factors were considered for the C_α atoms in the polypeptide chains [Wampler, 1997, Parthasarathy and Murthy, 1999, Parthasarathy and Murthy, 2000]. When benchmarking with the SCOP database[3] [Lo Conte et al., 2000, Lo Conte et al., 2002b], two different sets of B-factors were introduced as "2.0 standard deviation" and a more stringent variant "3.5 standard deviation".

- **REMARK465**. This algorithm is based on missing residues from electron density or NMR data of solution structures in the Protein Database[4] as stated above [Westbrook et al., 2003].

3.1.1.2. The DisEMBL method

The DisEMBL method [Linding et al., 2003a] is based on neural networks to allow the prediction of three different types of disorder:

- **loops/coils** as defined by the Dictionary of Secondary Structure of Proteins (DSSP) [Kabsch and Sander, 1983]. Loops are allowed to be part of structured protein regions (coils). The loops/coils algorithm is therefore not a unique requirement for disorder.

---

[1]http://www.mpipks-dresden.mpg.de/~tisean/
[2]http://smart.embl.de
[3]http://scop.mrc-lmb.cam.ac.uk/scop/
[4]http://www.pdb.org

- **Hotloops** are loops with an increased degree in mobility. This characteristic is derived from the C$_\alpha$ temperature factors (B-factors) approach.

- **REM465**. This disorder definition is based on missing atom coordinates in X-ray solution structures, such as stored in the Protein Database.

3.1.1.3. The IUPred method

This method claims to take the physical explanation for the occurrence for disordered/intrinsically unfolded and ordered/globular protein regions into account. The basic idea is that in globular proteins, amino acids are able to overcome the loss of entropy upon folding by the gain of stabilizing inter-residue interaction energy [Garbuzynskiy et al., 2004]. On the contrary, IUPs are less capable to form sufficient inter-residue interactions with their spatial neighbours. Therefore, it is possible to discriminate between the two borderline states, globular and intrinsically unfolded, by estimating the potential of a polypeptide to form stabilizing contacts [Thomas and Dill, 1996]. Dosztanyi *et al.* showed that the sum of interaction energies is proportional to a quadratic expression in the amino acid composition [Dosztányi et al., 2005b]. Additionally, the contribution to "globular" or "disorder" of a distinct amino acid depends not only on its chemical but also on its physical properties with respect to its potential interaction partner [Dosztányi et al., 2005b].

The IUPred algorithm estimates the inter-residue energy between each pair of amino acids based on their C$_\beta$ position [Dosztányi et al., 2005a, Dosztányi et al., 2005b]. The "heart of the method" is a 20×20 energy predictor matrix $\underline{P}$, whose elements P_{ij} had been trained on known globular structures. This position-specific method gives protein "likelihood of disorder" as probabilistic values in the range of 0 to 1. These values determine the borderline states, most residues correspond to values in between, complicating the determination of a complete sequence. The IUPred programme comes along with three different algorithms, "long (disorder)", "short (disorder)" and "structured (regions)", depending on the basic assumptions prior to parametrization.

3.1.1.4. The RONN method

This method is an evolution of a prediction method using a function neural network [Thomson and Esnouf, 2004] and is called "regional order neural network" (RONN) [Yang et al., 2005]. The Protein Data Bank and Molecular Structure Database (MSD) [Boutselakis et al., 2003] were taken as source for protein data. Trained on protein data sets comprising sequences with either mostly globular or mostly disordered or a mixture

of both folding states, multiple sequence alignments for each data set were generated. The observed differences in alignment scores are taken as a basis for the bio-based function neural network (BBFNN) method [Thomson et al., 2003, Yang and Thomson, 2005]. A sliding window technique gives ungapped sub-alignments training the BBFNN method to derive a receiver operator characteristic (ROC) for threshold determination. Query sequences are then evaluated against this threshold.

3.1.1.5. The FoldIndex method

This method was designed by Prilusky *et al.* and is based on the algorithm described by Uversky *et al.* [Uversky et al., 2000, Prilusky et al., 2005]. A formula for the net charge $\left|\langle R\rangle\right|$ of a given protein by its mean hydrophobicity $\langle H\rangle$ was introduced as

$$\left|\langle R\rangle\right| = 2.785\langle H\rangle - 1.151$$

with $\langle H\rangle$ calculated for each amino acid using the Kyte/Doolittle approximation and rescaled to values between 0 and 1 [Kyte and Doolittle, 1982]. "Natively unfolded" conformations are defined to have low mean hydrophobicity but a high net charge. Following this concept, interaction of natively unfolded protein regions with a binding partner affects $\left|\langle R\rangle\right|$ or $\langle H\rangle$ and therefore its folding state. The Fold Index$^{©}$ I_F^{KD} is then calculated as

$$I_F^{KD} = 2.785\langle H\rangle - \left|R\right| - 1.151$$

and trained on experimental data from the Protein Database.

3.1.1.6. The FoldUnfold method

A similar approach is FoldUnfold [Galzitskaya et al., 2006, Mamonova et al., 2010]. Globular structures from solution structures were taken as basis for the calculation of mean packing density. An average number of neighbouring residues in an 8 Å range for each amino acid was taken into account with additional use of a sliding window average technique. This method was trained on both long globular and long disordered protein segments to derive an ROC for the threshold determination [Galzitskaya et al., 2006].

3.1.2. Multiple sequence alignments

In terms of evolution, biological functionality of proteins is preserved within a phylogenetic family of species. Such biologically relevant protein segments are less residue

Table 3.2. Bioinformatical methods used in this study (miscellaneous).

Name	uniform resource location (URL)	Reference
SeqLogo	http://SeqLogo.threeplusone.com	[Schneider and Stephens, 1990] [Crooks et al., 2004]
belvu	http://sonnhammer.sbc.su.se/Belvu.html	[Sonnhammer and Hollich, 2005]
InParanoid	http://inparanoid.sbc.su.se	[Remm et al., 2001] [O'Brien et al., 2005]
gnuplot	http://www.gnuplot.info	

variant than others and can be detected by comparative sequence analysis. The detection of highly conserved subregions within sequences is a central issue to computational biology [Lander et al., 1991, Karp, 1993].

3.1.2.1. Multiple sequence alignment construction and substitution matrices

Sequence alignments allow a comparison of closely related sequences, sharing functional or sequential homology. For a biologically realistic alignment, input sequences are filled up with gap characters to give a rectangular array of rows with different levels of residue variation with the help of a substitution matrix. These matrices describe the variation rate over evolution for a specific amino acid. Matrix values are calculated from manually curated collections of sequences with known phylogenetic background. Dayhoff *et al.* developed the point accepted mutation (PAM) substitution matrix from alignments of protein sequences which share at least 85% sequence identity [Dayhoff et al., 1978]. Substitution matrices, in general, comprise not only residue-against-residue scores, but also scores for the possibility to get 'lost' or to get 'replaced' by a gap character. Depending on the substitution rate and divergence time, different matrix values were calculated, defining several PAM matrices based on different sequence variations between PAM1 and PAM250. As PAM matrices fail for the construction of alignments between evolutionarily more divergent sequences, Henikoff and Henikoff derived their block substitution matrix (BLOSUM) series from blocks of locally aligned sequences within an overall alignment of less identical sequences, such as the BLOSUM62 matrix which is based on alignment segments of at least 62% sequence identity [Henikoff and Henikoff, 1992]. The overall similarity between all input sequences may be less. The best-fit alignment of sequences is the one with the maximum sum-of-pairs score (SP score) [Bacon

and Anderson, 1986, Gotoh, 1986, Murata et al., 1985, Altschul et al., 1989]. Alignment construction with more than two input sequences is a complex problem which is computational (too) expensive. First improvements in computational performance were achieved using techniques based on the Needleman-Wunsch algorithm for global alignments or the Smith-Waterman algorithm for local alignments [Needleman and Wunsch, 1970, Smith and Waterman, 1981]. Altschul *et al.* evolved dynamic programming algorithms [Sankoff, 1975] based on the Carrillo-Lipman algorithm for less computational cost [Carrillo and Lipman, 1988, Altschul et al., 1989, Lipman et al., 1989].

3.1.2.2. Progressive alignment tools: MUSCLE

MUSCLE, multiple sequence comparison by log-expectation, was developed by Edgar *et al.* as a construction tool for multiple sequence alignments (MSA) with progressive refinement [Hogeweg and Hesper, 1984, Feng and Doolittle, 1987]. In a first step, a distant matrix is calculated from input sequences using either k-mer counting for unaligned sequences [Edgar, 2004c] or the Kimura approach for aligned sequences [Kimura and Takahata, 1983]. Distant matrix clustering by the UPGMA (unweighted pair group method with arithmetic mean) [Prelić et al., 2006] hierarchical clustering method gives an initial phylogenetic tree [Sneath and Sokal, 1973, Sneath, 1977]. Sequences are transformed into profiles which are pair-wisely aligned following the tree in 'root' direction. Refinement via alternate tree and alignment construction with internal improvement evaluation gives a best-fit solution for a biologically relevant MSA.

3.1.2.3. Progressive alignment tools: MAFFT

Katoh *et al.* developed MAFFT (multiple sequence alignment based on fast Fourier Transform) as another MSA software similar to Edgar's MUSCLE [Katoh et al., 2002]. MAFFT combines the construction of a rough initial MSA by the progressive method [Feng and Doolittle, 1987, Thompson et al., 1994, Katoh et al., 2002] with an iterative refinement method [Berger and Munson, 1991, Gotoh, 1993]. From the initial MSA an initial distant matrix is derived [Katoh et al., 2002] and a guide tree is constructed using an modified UPGMA method [Prelić et al., 2006]. Following that guide tree, the input sequences are aligned progressively [Feng and Doolittle, 1987, Thompson et al., 1994], which gives enough information for a new distant matrix [Edgar, 2004b, Katoh et al., 2002, Tateno et al., 1997]. The final MSA is obtained after alternate tree construction and immediate MSA refinement using an iterative method optimizing the weighted sum-of-pairs score (WSP) [Berger and Munson, 1991, Gotoh, 1993, Gotoh, 1995], an approx-

imate group-to-group alignment algorithm [Katoh et al., 2002] and the tree-dependent restricted partitioning technique [Plyusnin and Holm, 2012].

3.1.2.4. Dialign-TX

The assessment of alignment quality is a crucial step in MSA construction. Further improvements in alignment accuracy came with the development of pair-hidden Markov models comprising genetic algorithms (HMM) [Notredame and Higgins, 1996], simulated annealing [Kim et al., 1994], profile HMM [Krogh, 1994, Eddy, 1995] and multiple segment-to-segment comparisons [Morgenstern, 1996]. These hierarchical methods are not error-prone in early interim steps when initial alignments are constructed and refined stepwisely. Consistency-based schemes such as Dialign act differently. Whereas other methods sum up substitution matrix scores and gap penalties, the Dialign approach employs P-values [Morgenstern et al., 1998]. Only those sequence regions are aligned that share statistically relevant similarities. In this study, a more elaborated version, Dialign-TX, was used [Subramanian et al., 2005, Subramanian and Kaufmann, 2008]. A combination of segment-based and progressive alignment algorithms constructs pairwise sequence alignments for all possible combinations for a guide tree. The pairwise sub-alignments are taken to build the overall alignment using anchor points [Morgenstern et al., 2006] and a vortex-cover algorithm [Clarkson, 1983].

3.1.2.5. ProbCons

HMM represent a further approach to MSA generation via combination of progressive methods with probabilistic consistency [Rabiner, 1989, Baldi et al., 1994, Krogh, 1994, Hughey and Krogh, 1996]. Probabilistic calculations achieve a stronger conservation signal and may therefore give a higher accuracy compared to other methods. Probabilistic consistency is provided by ProbCons with its pair-hidden Markov model-based progressive alignment algorithm. In a first step, posterior-probability matrices are calculated from all possible pairs of input sequences to construct initial pairwise alignments. Dynamic programming allows calculation of the expected accuracies for all aligned pairs. A probabilistic consistency transformation re-estimates those quality scores. Hierarchical clustering constructs a guide tree, starting with progressive methods. Further iterative refinement gives the optimal final alignment.

Table 3.3. Web based multiple sequence alignment software used in this study.

Name	uniform resource locator (URL)	Reference
MUSCLE	http://www.ebi.ac.uk/Tools/msa/muscle/	[Edgar, 2004a] [Edgar, 2004b]
MAFFT	http://mafft.cbrc.jp/alignment/server/	[Katoh et al., 2002]
DIALIGN-TX	http://dialign-tx.gobics.de	[Subramanian and Kaufmann, 2008]
ProbCons	http://probcons.stanford.edu	[Do et al., 2005]

3.1.3. Computing sequence conservation from multiple alignments

A multiple sequence alignment comprises more or less phylogenetically related proteins to highlight sequence similarities as columns with higher degree of variation compared to others. Two different approaches have been tested in this study.

3.1.3.1. Computing sequence conservation from multiple alignments

The substitution matrix approach. Substitution matrices are mathematically derived from alignments of related sequences. They are constructed to display the evolutionary background as a basis for the residue variation within alignment columns. Therefore, substitution matrices such as BLOSUM [Henikoff and Henikoff, 1991, Henikoff and Henikoff, 1992] or PAM [Dayhoff et al., 1978] substitution matrices could be tested for alignment evaluation. Matrix values are regarded as amino acid/amino acid substitution scores. All alignment column elements are considered, whereas there are two approaches used in this study: First, all elements are regarded equal. Each residue-against-residue substitution is scored once and summed up to give the substitution score. Second, one sequence is taken as basis, in this case the *Saccharomyces cerevisiae* sequence. All possible substitutions are scored according to 'replacements' of the specific *Saccharomyces cerevisiae* amino acid. The pairwise substitution scores among the other sequences are ignored.

The information theoretical SeqLogo approach. This web-based tool graphically shows the information content of input multiple sequence alignments. The information theory behind is based on the works of Shannon *et al.* [Shannon, 1948, Schneider et al., 1986, Schneider and Stephens, 1990].

3. Methods

3.1.3.2. Information theory as mathematical background of the SeqLogo approach

Information theory based on the works of Shannon *et al.* and Pierce *et al.* give a mathematical basis for the description and evaluation of multiple sequence alignments at large scale [Shannon, 1948, Pierce et al., 1980].

For a defined ensemble with 4 nucleotides or 20 amino acids, respectively, the uncertainty H at an alignment position l is determined as $H(l) = -\sum_{b=a}^{t} f(b,l) \cdot \log_2 f(b,l)$ with $f(b,l)$ the frequency of residue b at position l. The total amount of information at position l, $R_{seq}(l)$, is

$$R_{seq}(l) = S_{max} - S_{obs}(l) = \log_2 X - (H(l) - e(n))$$

with S_{max} the maximum uncertainty of the ensemble, $S_{obs}(l)$ the observed uncertainty at position l, X the available number of residues in the ensemble and $e(n)$ as a correction factor for n-sized alignments of few sequences. Such a small-sample correction is approximated as

$$e(n) = \frac{1}{\ln 2} \cdot \frac{X-1}{2n}$$

For protein and nucleotide sequences the information content for a given position is

$$R_{seq}^{prot}(l) = \log_2 20 - (H(l) - e(n))$$

and

$$R_{seq}^{nucl}(l) = \log_2 4 - (H(l) - e(n)) = 2 - (H(l) - e(n))$$

Schneider *et al.* showed that the information approach is able to distinguish binding sites from the rest of the genome [Schneider et al., 1986] and evolved the graphical SeqLogo application from information theory [Schneider and Stephens, 1990]. SeqLogo was developed by Crooks *et al.* in 2004 as a web-based graphical application [Crooks et al., 2004].

3.1.3.3. The sliding window average technique

The sliding average techniques provides a mathematical calculation to average data points. It performs a series of subset averaging operations. The result is assigned to the subset's center position. The next averaging operation is performed with a subset 'shifted forwards'.

The sliding average technique is applied to determine a motif's 'absolute', 'relative' and 'background' conservation from conservation scores.

3.2. Genetic Techniques

All genetic-biological works comprised the amplification of oligonucleotides in polymerase chain reactions (PCR) on *Saccharomyces cerevisiae* DNA as a template [Kleppe et al., 1971, Saiki et al., 1985]. Multiplication of genetic material was preformed in bacterial strains of *Escherichia coli* prior to isolation from that strains. Verification and transformation of those constructs into *Saccharomyces cerevisiae* was performed for further investigations.

3.2.1. Preparation of plasmid DNA from Escherichia coli

Small-scale and large-scale bacterial plasmid preparations were performed using the plasmid isolation kits from Qiagen according to the manufacturer's instructions for high-copy number plasmids.

Escherichia coli cells were cultivated in lysogeny broth medium supplemented with the appropriate amount of antibiotic at 37°C for at least 14 hours while shaking at a constant rate of 220 rpm. The Qiagen "QIAprep" miniprep procedure uses the modified alkaline lysis method of Birnboim and Doly [Birnboim and Doly, 1979]. Bacterial cells are lysed in sodium hydroxide/SDS buffer in the presence of RNase A. One-step neutralization and adjustment to high-salt binding conditions of the lysate prepares the adsorption of plasmid DNA on silica membranes. After washing, the DNA is eluted from the membranes under low-salt conditions.

3.2.2. Quantitation of nucleic acids

DNA concentrations and purities of aqueous solutions were determined spectrophotometrically by measuring their absorbance at 260 nm. An absorbance (A_{260}) of 1 corresponds to 50 $\frac{\mu g}{\mu L}$ of double-stranded DNA and 40 $\frac{\mu g}{\mu L}$ of RNA or single-stranded DNA. The ratio $\frac{A_{260}}{A_{280}}$ corresponds to the sample purity. Ratios between 1.8 and 2.0 are regarded highly pure [Sambrook et al., 1989].

3.2.3. Generation and transformation of competent Escherichia coli cells

Chemically competent bacterial *Escherichia coli* cells were transformed according to the respective cell strain's manufacturer's protocol. With slight variations in the protocols, all used strains were processed as follows: Cells were kept on ice after removing from the -80°C storage. After adding 2–5 μL of miniprep isolated plasmid DNA to the cells, the cells were swirled and still kept on ice for 30 minutes. Cell tubes were heat shocked

at 42°C for 45 seconds, kept on ice for 2 minutes and the complete cell material streaked on lysogeny broth media plates supplemented with appropriate antibiotic. The plates were incubated overnight at 37°C.

3.2.4. DNA manipulations

Polymerase chain reaction were performed on BioRad Reaction Cyclers using DNA polymerases "Phusion high fidelity polymerase" from New England Biolabs, Cambridge, UK, and "Taq DNA polymerase" from Invitrogen with their respective reaction mixes. Restriction enzymes were purchased from New England Biolabs, Cambridge, UK. All enzymes and reagents were used according to the manufacturers' instructions. A common PCR programme set-up is given in table 3.4. A more detailed description of PCR steps is given in section 3.2.8. Denaturation temperature and duration depend on the polymerase used, whereas annealing values depend on primer characteristics. Elongation time is determined by the polymerase used and the desired amplicon size.

Typical reaction mixes contain:

component	final concentration
nuclease free water	-
buffer	$1\times$
dNTP mix, 10mM	200 μM
forward primer 10mM	0.5 μM
reverse primer 10mM	0.5 μM
template DNA	less than 250 ng
DNA polymerase	depending on the manufacturer's instructions

Table 3.4. Typical program set-up for a Polymerase Chain Reaction (PCR).

Step	Temperature	Duration	Description	Repetitions
1	92-96°C	30 sec to 2 min	initial denaturation	
2	92-96°C	1 sec	denaturation	
3	52-60°C	15 to 30 sec	annealing	25-30$\times$
4	72°C	30 - 60 $\frac{\text{sec}}{1\,\text{kb}}$	elongation	
5	72°C	10 min	final elongation	

3.2.5. DNA gel electrophoresis

Standard DNA gel electrophoresis was performed in agarose gels after PCR or restriction [Sambrook et al., 1989]. For resolution of DNA fragment sizes between 0.5 and 5 kb 1 % (w/v) agarose in 1× TAE buffer was used. DNA was visualized by adding ethidium bromide (UV light) or Sybr® Safe from Invitrogen (blue light) to the agarose gel, both according to the manufacturer's instructions. For gel extraction, the respective light sensitive band was cut out from the gel and the DNA was purified using the Qiagen Gel extraction kit.

1× TAE buffer		**6× DNA loading buffer**	
0.04 M	Tris-HCl, pH 8.0	0.25 % (w/v)	bromophenol blue
0.114 %	acetic acid	0.25 % (w/v)	xylene cyanol
1 mM	EDTA	15 %	Ficoll

3.2.6. Generation and transformation of competent Saccharomyces cerevisiae

A 50 mL culture was inoculated at an optical density of OD_{600}^{25} of 0.2 from a fresh overnight culture and incubated at 30°C while shaking until OD_{600}^{25} of 0.6 to 0.8. Cells were harvested at 4000 rpm for 5 minutes and washed once with 100 mM aqueous Lithium acetate solution. The cell pellet was resuspended in 500 μL of 100 mM aqueous Lithium acetate solution and aliquoted in 50 μL portions. After brief centrifuge spin the supernatant was removed, the pellet resuspended in "transformation mix" solution. The suspension was vortexted thoroughly and incubated at 30°C for 30 minutes. After heat shock at 42°C for 40 minutes, the cells were collected by centrifugation at 4000× g for 2 minutes, removing the supernatant and washing once with 1 mL of sterile water, resuspended in sterile water and plated on selective synthetic drop-out agar plates. Colonies of transformed cells should be isolated after incubation at 30°C for 2 – 4 days.

Transformation mix

240 μL	50 % (w/v) PEG-3350
36 μL	1M Lithium acetate, aqueous
5 μL	single stranded calf thymus DNA, 10 $\frac{\mu L}{mL}$
0.1 – 10 μg	plasmid DNA
70 mL	sterile water

3.2.7. Preparation of genomic DNA from Saccharomyces cerevisiae

Prior to the preparation step, *Saccharomyces cerevisiae* colonies were purified by dilution plating onto appropriate nutrient medium to obtain isolated colonies. A single colony is then inoculated shaking in liquid medium overnight at 30°C. Genomic DNA preparation is then processed using the Qiagen Puregene "Yeast/Bact. kit" according to the manufacturer's instructions.

3.2.8. Sanger DNA sequencing

Sequencing of double stranded DNA was done by the didesoxynucleotide method of Sanger using the "BigDye Terminator V.3.1" kit from Becton-Dickinson and the thermocycler ABI-Prism 3100 from Applied Biosystems [Sanger et al., 1977]. 500 ng of plasmid DNA were mixed with ready-to-use sequencing mix and 5 pmol of sequencing primers from table 2.4. PCR were performed according to the sequencer's instructions (see tables 3.5, 3.6(a)). The procedure begins with an pre-heating step to 96°C for 5 minutes (step 1). The cyclic steps comprise: Denaturation of the nucleic acid chain at 96°C for 30 seconds (step 2), annealing of primers to the target sequence at 50°C for 15 seconds (step 3) and extension of the complementary nucleic acid strand by the polymerase at 60°C for 4 minutes (step 4). A final extension step for 10 minutes ensures all remaining single-stranded DNA strands are fully extended (step 5). The processing of the samples was done after Exo-SAP PCR clean-up (see table 3.6(b)) in the central sequencing laboratory of the Cologne Center for Genomics[5] (CCG).

Table 3.5. Sanger sequencing PCR protocol using the "BigDye Terminator V.3.1" kit from Becton-Dickinson.

Step	Temperature	Duration	Repetitions
1	96°C	5 min	
2	96°C	10 sec	
3	55°C	5 sec	32×
4	60°C	4 min	
5	60°C	10 min	

3.2.9. The PCR epitope tagging technique

As the *Saccharomyces cerevisiae* genome is fully sequenced [Goffeau et al., 1996], PCR-techniques have come up which allow gene-targeting for a easy-to-proceed functional

---

[5]Cologne Center for Genomics, Weyertal 115b, 50931 Cologne, Germany.

Table 3.6. BigDye® Terminator reaction mix composition (a) and Exo/SAP clean-up protocol (b).

<div align="center">(a)</div>

	Volume
BigDye® Terminator v1.1/v3.1	0.25 μL
5× BigDye sequencing buffer	2.25 μL
Primer, 10 mM	0.25 μL
PCR template	2–10 ng per 100 bp
H$_2$O	1.95 μL

<div align="center">(b)</div>

	Volume
PCR product	7.0 μL
ExoI ($20\frac{U}{\mu L}$)	0.15 μL
SAP ($1\frac{U}{\mu L}$)	0.9 μL
H$_2$O	1.95 μL

ExoI: Exonuclease I, 15000 U from Neo Lab. SAP: Shrimps alkaline phosphatase, 500 U from Promega.

analysis of genes [Baudin et al., 1993, Wach et al., 1994]. With the PCR epitope tagging (PET) technique gene targeting for deletion, replacement or insertion of heterogenous DNA at a discrete locus on the genome is feasible [Knop et al., 1999, Longtine et al., 1998, Janke et al., 2004].

In the first step of the protocol described, the desired module is PCR amplified using S2- and S3-primers from table 2.4. These primers were designed to allow cassette amplification from the template plasmid. That cassettes comprise a selectable and an epitope tag. The cassette is flanked by homologous sequences with the gene of interest (GOI) in the target DNA (see table 2.4). Upon transformation into *Saccharomyces cerevisiae* strain JD47–13C after the Lithium acetate method (section 3.2.6), homologous recombination of the PCR fragment into the gene locus leads to in-frame fusion of the GOI and epitope tag [Schiestl and Gietz, 1989].

3.2.10. The site-directed mutagenesis technique

Site-directed mutagenesis of plasmid DNS was conducted using the "QuikChange® Lightning Site-Directed Mutagenesis Kit" from Stratgene, UK, following the manufacturer's instructions on primer design, reaction mix composition, reaction cycling conditions and reaction mix clean-up after cycling. Direct usage of the reaction mix for transformation into bacterial XL10-Gold® super-competent cells gave the desired transformants on lysogeny broth medium supplemented with appropriate antibiotic. Plasmids were isolated from transformants and digested with suitable enzymes to give a unique restriction pattern of bands on a agarose gel. Sanger sequencing confirmed the plasmid sequence.

3.2.11. The Gateway™ cloning technique

Some of the plasmid construction works were performed using the Gateway™ cloning technique. This technique is based on the site-specific recombinase system of bacteriophage lambda (λ). It facilitates the integration of λ genome into the *Escherichia coli* chromosome and the switch between lytic and lysogenic pathways [Ptashne, 1992]. Lambda-based recombination involves the DNA recombination sequences and proteins that mediate the recombination reaction. Recombination only occurs between the *att* sites, specific attachment sites on interacting DNA molecules. DNA segments flanking the *att* sites are interchanged in a way that the *att* sites are converted to hybrid sequences from parental vectors during the recombination step. The *att* sites have been engineered by the manufacturer to ensure specificity of the recombination reactions to maintain sequence orientation, to minimize secondary structure formations and to establish an incompatible sets of *att* site pairs: *att*B1/*att*P1, *att*B2/*att*P2, *att*P1/*att*L1, *att*P2/*att*L2. The actual cross-over reaction itself is catalyzed by a mixture of enzymes that bind to specific *att* sequences [Stark et al., 1989, Landy, 1989]: The "BP" reaction is mediated by bacteriophage λ integrase (Int) and *Escherichia coli* integration host factor (IHF) proteins. Reaction only occurs between bacterial *att*B and phage *att*P sites to specifically give the *att*L sites. These sites can be recombined in a second reaction to give the corresponding *att*R sites in an "LR" reaction. In this step, bacteriophage λ Int and excisionase (Xis) proteins together with *Escherichia coli* IHF are the mediators. The manufacturer provides prepared master mixes with the respective clonase enzyme/protein combination for both BP- and LR-reaction.

The Gateway™ cloning protocol

The first step for using this technique is the PCR amplification of the desired DNA fragment together with appropriate *att*B sites. The primers were designed according to the manufacturer's instructions (in $5' \rightarrow 3'$ direction):

*att*B1 forward oligo: GGGG ACAAGTTTGTACAAAAAAGCAGGCT TC X[6]
*att*B2 reverse oligo: GGGG ACCACTTTGTACAAGAAAGCTGGGT C X[6]

The manufacturer provides a number of entry vectors. In this study, pDONR™221 was exclusively used for BP reactions. Its features are: a chloramphenicol resistance gene located between the *att*P sites for counter-selection, a *ccdB* gene located between the

---

[6]18–25 gene-specific nucleotides at the $3'$ end.

two *att*P sites for negative selection, kanamycin resistance for selection in *Escherichia coli* and pUC origin for replication and maintenance of the plasmid in *Escherichia coli*. The BP reaction was set up as described in table 3.7 and run overnight at 25°C.

Table 3.7. The Gateway™ BP reaction mix used in this study.

Component	Volume
PCR with flanking *att*B sites	$1.0 - 7.0$ μL
pDONR™221 ($150 \frac{ng}{\mu L}$)	1.0 μL
BP clonase	0.25 μL
TE buffer	to 8.25 μL

BP clonase, Sigma-Aldrich

Transformation of the unpurified reaction solution into *Escherichia coli* strain XL1-blue gave the desired transformants on lysogeny broth media plates supplemented with kanamycin after incubation overnight at 37°C. The manufacturer offers a variety of destination vectors. In this study, pACT2™ with ampicillin selectable marker was used as destination vector together with the plasmid from the previous BP reaction. The LR reaction was set up as described in table 3.8 and run overnight at 25°C.

Table 3.8. The Gateway™ LR reaction mix used in this study.

Component	Volume
entry clone from previous BP reaction with *att*P sites	$1.0 - 10.0$ μL
destination vector ($150 \frac{ng}{\mu L}$)	2.0 μL
LR clonase reaction buffer, 5×	3 μL
LR clonase	0.25 μL
TE buffer	to 15.25 μL

LR clonase, Sigma-Aldrich

Transformation of the unpurified reaction solution into *Escherichia coli* strain XL1-blue gave the desired transformants on lysogeny broth media plates supplemented with ampicillin after incubation overnight at 37°C.

3.2.12. Preparation of crude extracts from Saccharomyces cerevisiae

Cells were routinely harvested from liquid cultures by centrifugation at 4000 rpm and washed once with water. Depending on further purpose, one of the following methods for protein extraction was used:

Boiling   For direct use in denaturing polyacrylamide gel electrophoresis (PAGE), cell pellets were simply resuspended in SDS-PAGE sample buffer and lysed by incubating at 100°C for 5 minutes. Unless stated otherwise insoluble cell debris was spun down and the supernatant was used for PAGE. Routinely, crude extracts from 0.5 to 1 OD_{600}^{25} units per sample were loaded onto the gel.

Glass bead extraction (I)   The following steps were all carried out at 4°C. After harvesting, cells were resuspended in lysis buffer including a protease inhibitor cocktail (complete protease inhibitors, Roche Diagnostics, Mannheim, Germany) and acid washed glass beads (Ø 0.45 mm) were added approximately in a 1:1 ratio. The sample was vortexted for 5 minutes at 4°C and then centrifuged at 16000×g for 10 minutes, the supernatant transferred to fresh tubes and used as crude extracts.

1× Lysis buffer		**Sample buffer**	
50 mM	HEPES, pH 7.5	62.5 mM	Tris
150 mM	sodium chloride	2%	SDS
1 % (v/v)	triton-X-100	10%	glycerol
5 mM	EDTA	0.002%	bromophenol blue
		adjust to pH 6.8	

3.2.13. Preparation of crude extracts from Escherichia coli

Glass bead extraction (II)   Cells were routinely harvested from liquid cultures by centrifugation at 10,000 rpm and washed once with water. The next steps were all carried out at 4°C. After harvesting, cells were resuspended in lysis buffer including a protease inhibitor cocktail (complete protease inhibitors, Roche Diagnostics, Mannheim, Germany) and acid washed glass beads (Ø 0.10 – 0.11 mm) were added approximately in a 1:1 ratio. The sample was vortexted for 5 minutes at 4°C and then centrifuged at 16000× g for 10 minutes, the supernatant transferred to fresh tubes and used as crude extracts.

3.2.14. SDS polyacrylamide gel electrophoresis

For protein separation, one-dimensional polyacrylamide gel electrophoresis (PAGE) under denaturing conditions was applied [Laemmli et al., 1970]. Gel compositions are given in tables 3.9 and 3.10. Usually, a acrylamide content of 10% was used. Samples were loaded after boiling in Laemmli loading buffer (LLB) for 5 minutes. Gel runs were

performed in Laemmli running buffer (LRB). Applied current and running time was applied according to gel size and expected sample masses to be resolved.

Reagents for SDS-PAGE

1 M	Tris-Cl, pH 6.8
1.5 M	Tris-Cl, pH 8.8
30 % (w/v)	acrylamide / 2 % bisacrylamide
10 % (w/v)	APS
10 % (w/v)	SDS
	TEMED

Laemmli running buffer (LRB)

25 M	Tris-Cl, pH 6.8
192 mM	glycine
0.1 % (w/v)	SDS

Laemmli loading buffer (LLB)

125 M	Tris-Cl, pH 6.8
10 %	β-mercaptoethanol
4 % (w/v)	SDS
20 % (w/v)	glycerol
0.04 % (w/v)	bromophenol blue

Table 3.9. Resolving gel compositions, with respect to different acrylamide contents.

	6 %	7 %	8 %	10 %	12 %
water	8.1 mL	7.6 mL	7.1 mL	6.1 mL	5.1 mL
1.5 M Tris-Cl, pH 8.8	3.75 mL	3.75 mL	3.75 mL	3.75 mL	3.75 mL
acrylamide / bisacrylamide	3 mL	3.5 mL	4 mL	5 mL	6 mL
10 % SDS	150 μL	150 μL	150 μL	150 μL	150 μL
10 % APS	50 μL	50 μL	50 μL	50 μL	50 μL
TEMED	10 μL	10 μL	10 μL	10 μL	10 μL

Table 3.10. Stacking gel, 3 % (v/v) acrylamide.

	3 %
water	3.6 mL
1 M Tris-Cl, pH 6.8	625 μL
acrylamide / bisacrylamide	650 μL
10 % SDS	50 μL
10 % APS	10 μL
TEMED	5 μL

3.2.15. Western blot analysis

After SDS gel electrophoresis proteins were transferred to a polyinylidene fluoride (PVDF) membrane from Millipore by semi-dry blotting at 240 mA for 1 – 3 hours depending on the size of proteins detected. After transferring, the proteins were fixed to the membrane by boiling in water for 30 minutes. The membrane was incubated in blocking solution (3 % (w/v) dry milk powder in PBST) for 45 minutes, followed by one hour incubation with primary antibody diluted in blocking solution. After two washing steps in PBST and SAIS for 5 minutes and another blocking for 10 minutes, the membrane was incubated with the secondary peroxidase coupled antibody. After several washing steps with PBST and SAIS, the membrane was washed in a final step with PBS and incubated in 'lumilight plus western blotting reagent' from Roche. X-ray films from Fuji were exposed for documentation.

Transfer buffer (TB)		**PBS, pH 7.4**	
25 mM	Tris-Cl, pH 8.3	137 mM	sodium chloride
19 mM	glycine	2.7 mM	potassium chloride
20 % (v/v)	Methanol	81 mM	disodium hydrogenphosphate
0.1 % (w/v)	SDS	1.5 mM	potassium dihydrogenphosphate

PBST, pH 7.4		**SAIS**	
137 mM	sodium chloride	1 M	sodium chloride
2.7 mM	potassium chloride	10 mM	disodium hydrogenphosphate
81 mM	disodium hydrogenphosphate	0.5 % (w/v) tween-20	
1.5 mM	potassium dihydrogenphosphate		
0.1 % (w/v) tween-20			

3.2.16. The Yeast Two-Hybrid assay

Protein-protein interactions (PPI) were detected *in vitro* using the yeast two-hybrid technique (Y2H) [Fields and Song, 1989, Chien et al., 1991]. Physical interaction between two proteins is detected linking interaction to an observable phenotype. The Gal4 transcription factor is divided into a activating (AD) and a binding (BD) domain and is only functional when those two domains are placed in close proximity. Fusing proteins either to AD and BD, transcription only occurs when interaction between the distinct proteins serves as linker for AD/BD interaction which then enables transcription of the reporter HIS and ADE genes and the observable phenotype. That transcriptional assay utilizes a genetically modified yeast strain like AH109 or PJD69–4A, in which the biosynthesis of

certain nutrients is lacking. In these strains, three reporter genes ADE2, HIS3 and MEL1 (or lacZ) under the control of distinct GAL4 upstream activating sequences (UAS) and TATA boxes are used. Genes, encoding the SUMO-interacting candidate proteins from the bioinformatical prediction screen, were HA-tagged and fused to the Gal4-activating domain, while SMT3 variants SMT3, SMT3Δ97, SMT3Δ97x3, SMT3Δ97x4, were fused to the Gal4-binding domain. Both Gal4 activating and Gal4 binding plasmids were simultaneously transformed into PJ69–4A and AH109, respectively, transformants selected on YP(A)D[7] plates lacking leucine and tryptophane. After replica streaking onto SD-LW plates, equal aqueous solutions from cell dilutions corresponding to 10 OD_{600}^{25} were spotted onto synthetic drop-out media plates additionally lacking histidine, histidine and variant amounts of 3-amino-1,2,4-1H-triazole (3-AT), and adenine, respectively [Fields, 1993, Joung et al., 2000, Fields, 1993]. Colony growth on these selective media plates is regarded as a measure of interaction between Gal4 activating and Gal4 binding domain and, therefore, between fused assumed interactors.

3.2.17. The GST-Pulldown assay

To verify the results from the yeast two-hybrid assays, GST pulldown assays were performed with new designed constructs as described in Kristina Uzunova's PhD thesis [Uzunova, 2006]. *Escherichia coli* strain BL21(DH3) was transformed with plasmids, expressing either GST or GST-SUMO. Proteins as a putative interaction partner were genetically 6×histidine tagged using the protocol developed by Knop *et al.* [Knop et al., 1999, Janke et al., 2004]. Protein expression was induced for 4 – 5 hours at 37°C upon isopropyl-1-thio-β-D-galactopyranoside (IPTG) addition. Cells corresponding to an optical density of 10 OD_{600}^{25} were harvested by centrifugation at 14000 rpm for 5 minutes and resuspended in 450 μL extraction buffer. Mechanical rupture after the glass bead lysis protocol and removal of cell debris by centrifugation at 14000 rpm at 4°C for 10 minutes gave the GST- and and GST-SUMO crude extracts. Epitope tagged proteins were obtained upon culturing respective yeast strains in YPD liquid media at 30°C for 3 hours. With an OD_{600}^{25} of 0.6 – 0.8, cells corresponding to an OD_{600}^{25} of 10 were harvested by centrifugation at 4000 rpm for 5 minutes, mechanically ruptured after the glass bead protocol (section 3.2.13). 80 μL of a 1:1 slurry of glutathion-sepharose CL-4B resin (Amersham Pharmacia Biotech, Piscataway, NJ, USA) was inverted with an equal amount of extraction buffer for 1 hour at 4°C, washed three times with 100 μL extraction buffer. Then 100 μL of GST containing crude extract were applied to the beads,

---

[7]AH109 is a adenine deficient yeast strain. Addition of adenine to growth media is recommended.

inverting the tubes for 2 – 4 hours at 4°C. After centrifugation for 3 seconds at 800 rpm at 4°C and washing once with 100 μL extraction buffer, 100 – 200 μL of the yeast crude extracts were applied to the Sepharose beads inverting the tubes for 2 – 4 hours at 4°C. The beads were washed at least once with extraction buffer and resuspended in sample buffer and incubated at 100°C for 5 minutes.

extraction buffer

150 mM	sodium acetate
2%	SDS
10%	glycerol
0.002%	bromophenol blue

adjust to pH 6.8

3.2.18. Two-step gene replacement

Replacement of a wild-type gene with a mutant allele at the endogenous locus was performed as using the two-step gene replacement technique [Widlund and Davis, 2005]. The mutant variant was PCR amplified and ligated into the shuttle vector pRS306 bearing an URA3 marker. Transformation into *Escherichia coli* strain DH5α onto synthetic dropout plate medium supplemented with the respective antibiotic and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) allowed blue/white screening for increased transformation efficiency. Plasmids were isolated as colorless transformants and digested with suitable enzymes to give a unique restriction pattern of bands on an agarose gel. Sanger sequencing confirmed the correct sequence of that integratable plasmid. Its linearization with a uniquely cutting restriction enzyme BamHI leads to homologous recombination to the endogenous gene locus as a tandem array of wild-typ and mutant allele flanking the URA3 selectable marker. Selection/counter-selection on YPD plate medium supplemented with 0.1% 5-fluoro-orotic acid (FoA) leads to reduction of that URA3 marker and flanking sequence repeats by mitotic recombination to give the mutant variant of the wild-type gene of interest. After purification of the mutant yeast strain by dilution streaking from single colonies onto YP/Glycerol plates and replica plating onto SD/FoA medium, the mutant yeast strain was verified by both PCR amplification using primers from table 2.4, restriction at the site of mutagenesis and sequencing of the amplicon.

4. Results

Short linear motifs (SLiMs) in general, or SUMO interacting motifs (SIMs) in specific terms, are hard to handle with bioinformatical methods (section 1.6). These methods are originally designed and trained on longer protein domains. Short motifs can be seen as special types of domains.

The following paragraphs describe the bioinformatical analysis of established SIMs and their flanking regions, with the aim to derive characteristics that can be used to discriminate functional SIMs from spurious SIM-like sequences. Subsequently, these characteristics are used in a proteome-wide screen for novel SIMs in the budding yeast *Saccharomyces cerevisiae*. In section 4.5, the same set of criteria is used for the detection of another short motif with SIM-like characteristics: the AIM motif that mediates binding to Atg8 (section 1.4.3).

4.1. General aspects for a bioinformatical prediction of SUMO interacting motifs

Protein sequence data was obtained from mining sources such as the Broad Institute of the Massachusetts Institute of Technology (MIT) and the Harvard University and the U.S. Department of Energy Joint Genome Institute (DOE JGI, or JGI). Retrieved data was processed for further use as database for bioinformatical applications. Unless stated otherwise, all applications used in this study were programmed by the author of this study.

4.1.1. SUMO interacting motifs belong to one of the three consensus patterns

SIM sequences belong to one of the three consensus patterns SIMa, SIMb and SIMr [Uzunova et al., 2007]. These three SIM consensus patterns form the basis for the discovery of novel SIM instances. A consensus search within a *Saccharomyces cerevisiae* sequence database yields a large number of retrieved motifs (table 4.2). Variations in the search pattern change the number of retrieved motifs. The exact definition of a sequence

pattern is important for the performance of a pattern-based database search. Too restrictive consensus patterns exclude biologically relevant motifs. With less stringent patterns, too many false-positive motif candidates are retrieved. For both borderline cases, the number of SIM-like protein sequences is much higher than the number believed to be relevant.

Table 4.1. SIM sequence characteristics as criterion in a bioinformatical prediction approach.

SIMa	Four hydrophobic residues, usually isoleucine, leucine, valine or methionine, but with a more variant third position compared to the others. At least two "acidic" termed amino acids such as aspartic acid or glutamatic acid follow that stretch. A possible consensus pattern is [ILVM] [ILVM] . [ILVM] [SDE] [SDE].
SIMb	Only few motif sequences are known for this consensus type. The consensus pattern is the most stringent one with a core comprising the amino acids valine, isoleucine, aspartic acid, leucine and threonine with more residue variability at the third position. A consensus pattern can be described as [PVILMT] [ILVM] DLT.
SIMr	This consensus pattern has only been observed in RanBP2, Elg1 and recently shown in Rad18 [Pichler et al., 2002, Hecker et al., 2006, Parnas et al., 2010, Parker and Ulrich, 2012]. The SIMr consensus pattern has close similarity to the one for SIMa, but in a reverse orientation. A possible consensus pattern is [DES] [DES] [ILVM] . [ILVM] [ILVM].

Table 4.2. Consensus patterns for SIMa, SIMb and SIMr for pattern-based *Saccharomyces cerevisiae* protein sequence screens.

consensus type	consensus pattern	number of sequence strings	number of different proteins
SIMa	[ILVM] [ILVM] . [ILVM] [SDE] [SDE]	1,555	1,264
	[PILVM] [ILVM] . [ILVM] [SDE] [SDE]	1,912	1,495
	[PILVM] [ILVM] . [ILVM] [SDE] [SDE] [SDE]	387	361
SIMb	[PVILMT] [ILVM] DLT	36	36
SIMr	[DES] [DES] [ILVM] . [ILVM] [ILVM]	8,502	1,159
	[DES][DES] [ILVM] . [ILVM] [PILVM]	1,760	1,387
	[DES] [DES][DES] [ILVM] . [ILVMF][ILVMF]	524	470

4.1.2. Four criteria for evaluation of database screen results for new SUMO interacting motifs

The number of motif candidates retrieved by the consensus pattern search is much higher than the expected number of biologically relevant SIM instances. There are more criteria to be defined as a measure for relevance from such a search. Multiple sequence alignments (MSA) of established SUMO interacting proteins show additional criteria (table 4.3):

Table 4.3. Four criteria for a reliable detection of relevant SUMO interacting motifs.

consensus adherence	Established SUMO-interacting motifs can be classified to one of the three consensus types SIMa, SIMb and SIMr (table 4.1). As such, any new SIM is expected to fulfill one of those consensus criteria.
absolute conservation	Functional motifs are under evolutionary constraints. Any variation in motif sequence only occurs in tight limits and in accordance with its interaction partner. Therefore, new SIMs are also expected to be conserved, at least over moderate evolutionary distances.
relative conservation	By contrast to functional SIMs, the flanking regions are typically not constrained. Therefore, new SIMs are expected to be better conserved than their flanking regions.
structural context	Additionally, SIMs occur outside globular structures. A possible explanation may be that SIMs have to be in flexible protein regions. Therefore, new SIMs are expected to reside in regions that are predicted to be unstructured, or at least not predicted to be globular.

4.1.3. Comparison of different disorder prediction tools for a bioinformatical SIM detection procedure

There are several tools for predicting unfolded ('disorder') regions in proteins, usually with different concepts of what constitutes an unfolded region. To investigate which of these tools is suited for enriching functional motifs, a number of web-based prediction tools was tested such as DisEMBL, GlobPlot, IUPred, FoldUnfold, FoldIndex and RONN (section 3.1.1) [Linding et al., 2003a, Linding et al., 2003b, Yang et al., 2005, Garbuzynskiy et al., 2004, Prilusky et al., 2005]. The test bed of proteins comprised the sequences of the human proteins Daxx and Prrg3 and *Saccharomyces cerevisiae* protein Wss1 [Biggins et al., 2001, Hitt and Wolf, 2004, Lin et al., 2006]. They contain motifs of different

classes (table 4.4). A disorder prediction method is considered as suitable if it places short functional motifs in disordered context while predicting folded domains as non-disordered. Globularity prediction methods are used as an additional measure where the prediction of non-globularity is interpreted as being potentially unstructured.

Table 4.4. Motifs in Daxx, Prrg3 and Wss1 proteins for a performance test of bioinformatical disorder and globularity prediction tools. A bioinformatical evaluation of protein sequences whether a distinct protein region is unfolded is a crucial step towards a reliable SIM detection. These proteins and motifs serve as test bed for several prediction tools.

protein	coordinates	motif	description	reference(s)
Daxx	7–15	IIVLDD	SIMa	[Lin et al., 2006]
	733–740	IIVLSDSD		
Prrg3	79–101	AMYVVVPLLGV-ALLIVIALFIIW	helical domain	[Kulman et al., 2001]
	191–194	PPSY	PY box	
	216–219	PPKY	PY box	
Wss1	209–219	RELAAFAAERR	VIM	[Hitt and Wolf, 2004]
	247–251	VVILDD	SIMa	[Biggins et al.,
	265–269	VIDLT	SIMb	2001]

IUPred was used with parameters favoring the detection of short (IUPred 'short') and longer unstructured regions (IUPred 'long'). These algorithms and DisEMBL with its parameters for 'LOOPS' and 'HOTLOOPS' gave the best results. RONN, FoldIndex and FoldUnfold predicted longer disordered regions. The criterion of a SIM's structural context is less suitable for short motifs with these methods. GlobPlot shows good results for the globularity prediction for the test motifs. This first overview addressed a range of different motifs, not all of which are guaranteed to reside within unstructured regions. The different test motifs are not all located in unstructured regions.

Therefore, a second round of evaluation was performed. The new test bed comprised the SIPs Daxx, RNF4, Fir1, Uba2, Slx5, Nis1, Siz1 and Uls1 (table 1.1) [Johnson and Gupta, 2001, Pichler et al., 2002, Hannich et al., 2005, Ii et al., 2007b, Uzunova et al., 2007, Sun et al., 2007]. The tested tools were as before.

FoldUnfold predicts the SIM hydrophobic part as folded. Therefore, this approach is not helpful for this study. FoldIndex gives slightly better results in this test. It yields larger segments of disorder covering the established SIMs. GlobPlot 'B-factors 2.0 std. dev.' and DisEMBL 'LOOPS' tend to give nearly identical results. The performances of the IUPred disorder tool are better with SIMs compared to the results from the previous test with shorter motifs only. The IUPred disorder server with both its 'IUPred short' and 'IUPred long' algorithms give again good results. Best disorder prediction results are

obtained with a combination of these three methods, GlobPlot 'B-factors 2.0 std. dev.', IUPred 'short' and 'long'. Both disorder and globular prediction were taken into account for a more suitable bioinformatical approach. SIM candidates located in unstructured regions — or at least outside globular regions — are considered as more likely to be functional.

4.2. Generation of multiple sequence alignments

4.2.1. Two phylogenetic ranges of species for the identification of orthologs

Orthology inference of *Saccharomyces cerevisiae* proteins was performed with the 'Inparanoid' approach [Remm et al., 2001, O'Brien et al., 2005, Berglund et al., 2008, Ostlund et al., 2010]. It allows the identification of one-to-one and one-to-many orthology and can also deal with many-to-many situations by including inparalogs into the cluster sets. The Inparanoid program was applied to all 5885 *Saccharomyces cerevisiae* protein sequences in the genome-wide collection. In order to address SIM candidates which are conserved over different evolutionary distances, two phylogenetic ranges were defined starting from *Saccharomyces cerevisiae*: A set of 12 species comprising the 'Saccharomycetales' orthologous sequences and a 'Saccharomycetaceae' set of additional 12 species (table 4.5). All *Saccharomyces cerevisiae* protein sequences were aligned with their orthologs to construct the two MSA sets. The performances of different MSA generation tools were tested.

Table 4.5. Two sets of MSA for a characterisation of SIM conservation.

'Saccharomycetaceae'	
'Saccharomycetales'	
Saccharomyces cerevisae	*Tetrapisispora phaffii*
Saccharomyces bayanus	*Torulaspora delbrueckii*
Saccharomyces mikatae	*Kluyveromyces africanus*
Saccharomyces castellii	*Saccharomyces arboricola*
Vanderwaltozyma polyspora	*Eremothecium cymbalariae*
Candida glabrata	*Naumovozyma dairenensis*
Zygosaccharomyces rouxii	*Debaryomyces hansenii*
Saccharomyces kluyveri	*Pichia guilliermondii*
Kluyveromyces thermotolerans	*Candida lusitaniae*
Kluyveromyces waltii	*Candida albicans*
Kluyveromyces lactis	*Candida dubliniensis*
Ashbya gossypii	*Lodderomyces elongisporus*

4.2.2. Choice of a suitable multiple sequence alignment tool

For assessing the conservation of SIM candidates, a proteome-wide set of aligned orthologs is required. Since SIMs typically reside within unstructured and poorly conserved sequence regions, the task of generating a meaningful multiple alignment is not trivial.

A number of different multiple alignment programmes are available and were tested for their ability to correctly align established SIM instances. ProbCons, Dialign-TX, MAFFT L-INS-I and MUSCLE were used (section 3.1.2) [Edgar, 2004b, Edgar, 2004c, Katoh et al., 2002, Subramanian and Kaufmann, 2008, Do et al., 2005]. ProbCons and MAFFT performed best, followed by MUSCLE and DIALIGN-TX. However, construction time increased in the order MUSCLE < MAFFT < DIALIGN-TX ≈ ProbCons. MAFFT L-INS-I gives the most accurate local alignments on SUMO interacting motifs among the test set alignment methods.

4.3. Conservation criteria from multiple sequence alignments

4.3.1. A suitable metric for assessing residue conservation

MSAs can be used to measure sequence conservation. Two approaches were used to read-out that MSA conservation content:

> **The substitution matrix approach**. Substitution matrices give a mathematical background for the evaluation of conservation within alignment columns (sections 3.1.2.1, 3.1.3). Alignment algorithms use these matrices for the maximization of sequence similarity. In this study, BLOSUM45 and BLOSUM62 matrices are used for the assessment of sequence conservation from MSA data (section 3.1.2.1).

> **The information theoretical SeqLogo approach**. The second approach is based on information theory and thus disregards similarity between the different amino acids (section 3.1.3.2) [Shannon, 1948, Schneider et al., 1986, Schneider and Stephens, 1990]. The information content of an alignment column is used as a conservation measure. This conservation metric is identical to that used by the popular SeqLogo approach for generating graphical alignment representations [Crooks et al., 2004].

There are two ways to describe conservation levels within aligned columns: An 'all-against-all scoring approach' ('N:N') treats all observed amino acids equally, whereas the 'one-against-all scoring approach' ('1:N') sets the *Saccharomyces cerevisiae* amino

acid as a reference. The 1:N scoring method measures the local similarity of the *Saccharomyces cerevisiae* sequence to all other sequences of the alignment, while the N:N scoring measures the overall conservation within an alignment column.

MSA data generated by MAFFT L-INS-I were used as a basis for bioinformatical procedures. Substitution and SeqLogo approach were used for 1:N and N:N conservation assessments.

In a first step, conservation assessment was performed for the established SIMs in Slx5 and Siz1 in a preliminary test (sections 3.1.2.1, 3.1.3.2 and table 4.6). High scores indicate high information content or high inter-residue similarity, respectively. The overall scores for SIMa in Slx5 and Siz1 are higher than for their flanking regions (not shown).

Table 4.6. The level of SIMa conservation can be calculated with substitution matrix and SeqLogo approaches. Conservation levels are derived from Saccharomycetales MSA. Both 'one-against-all' (1:N) and 'all-against-all' (N:N) approaches were performed.

Protein	SIMa	BLOSUM45		BLOSUM62		SeqLogo	
		1:N	N:N	1:N	N:N	1:N	N:N
Slx5	V	43	168	35	134	1.558	2.492
	I	68	490	55	401	3.516	3.750
	L	16	-12	6	-57	0.989	2.260
	I	67	475	554	386	3.516	3.750
	D	45	304	41	278	1.085	2.894
	S	53	370	53	371	3.516	3.750
Siz1	I	52	417	47	366	1.340	3.351
	I	70	525	56	420	4.322	4.322
	N	14	235	13	213	0.386	2.896
	L	70	525	56	420	4.322	4.322
	D	93	665	80	574	3.704	3.969
	S	56	420	56	420	4.322	4.322

For a better comparison of conservation features of different established SIMs, another preliminary test was performed. The aim was to decide, whether the substitution or the SeqLogo approach is suited best to assess the conservation criteria. These criteria are to be used for a refined bioinformatical sequence screen.

The conservation scores from 'Saccharomycetales' MSA data for Siz1 were compared to the Siz1 domain architecture. The domain architecture of Siz1 (YDR409W) with its established SIMa was retrieved from the Pfam database [Sonnhammer et al., 1997,

Bateman et al., 2000,Finn et al., 2010,Finn et al., 2014]: Siz1 has an SAP domain (34–68) in close proximity to the N-terminus, followed by a PINIT domain (170–312), a MIZ zinc finger (357–406) and a SIMa (`IINLDS`, 482–488) [Johnson and Gupta, 2001,Pichler et al., 2002, Shuai and Liu, 2005, Reindle et al., 2006, Takahashi and Kikuchi, 2005, Uzunova et al., 2007, Yunus and Lima, 2009]. This three-dimensional domain architecture can be directly mapped to the MSA (figure 4.1).
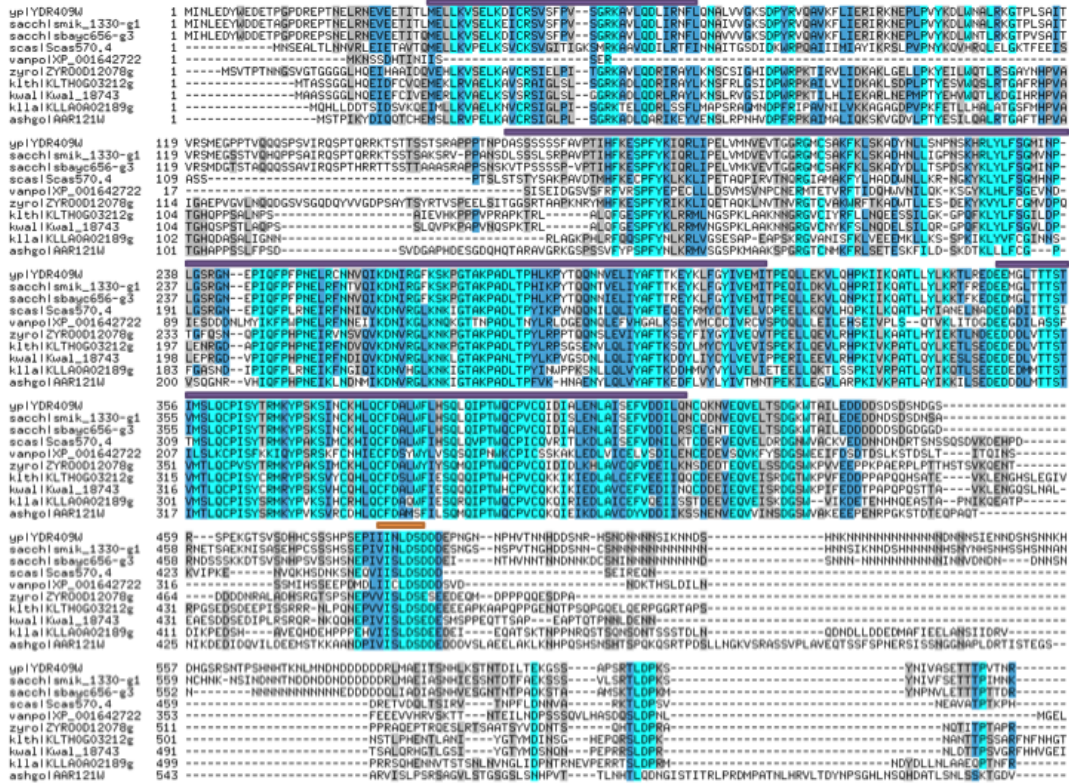
Figure 4.1. Siz1: MSA as a representation of the three-dimensional domain architecture of a protein. MAFFT L-INS-I MSA of *Saccharomyces cerevisiae* Siz1 with homologs within close phylogenetic range, here termed the 'Saccharomycetales'. Siz1 has an SAP domain (34–68) in close proximity to the N-terminus, followed by a PINIT domain (170–312), a MIZ zinc finger (357–406) and a SIMa (`IINLDS`, 482–488) [Johnson and Gupta, 2001, Pichler et al., 2002, Shuai and Liu, 2005, Reindle et al., 2006, Takahashi and Kikuchi, 2005, Uzunova et al., 2007, Yunus and Lima, 2009]. Violet boxes indicate conserved domains, whereas the orange box indicates the SIMa.

As shown in figure 4.1, globular protein domains and motifs usually correspond to regions with high sequence conservation. Furthermore, information theoretical approaches can mathematically address alignment files for conservation assessments (sections 3.1.2.1, 3.1.3.2, table 4.6). Gnuplot plots of the conservation scores against the *Saccharomyces*

cerevisiae protein sequence coordinates serve as a graphical representation for protein conservation (figure 4.2). The information theoretical SeqLogo approach gives scores in a narrow range between 0 and ~4.22 (section 3.1.3.2).

(a) 1:N BLOSUM45   (b) 1:N BLOSUM62   (c) 1:N SeqLogo

(d) N:N BLOSUM45   (e) N:N BLOSUM62   (f) N:N SeqLogo
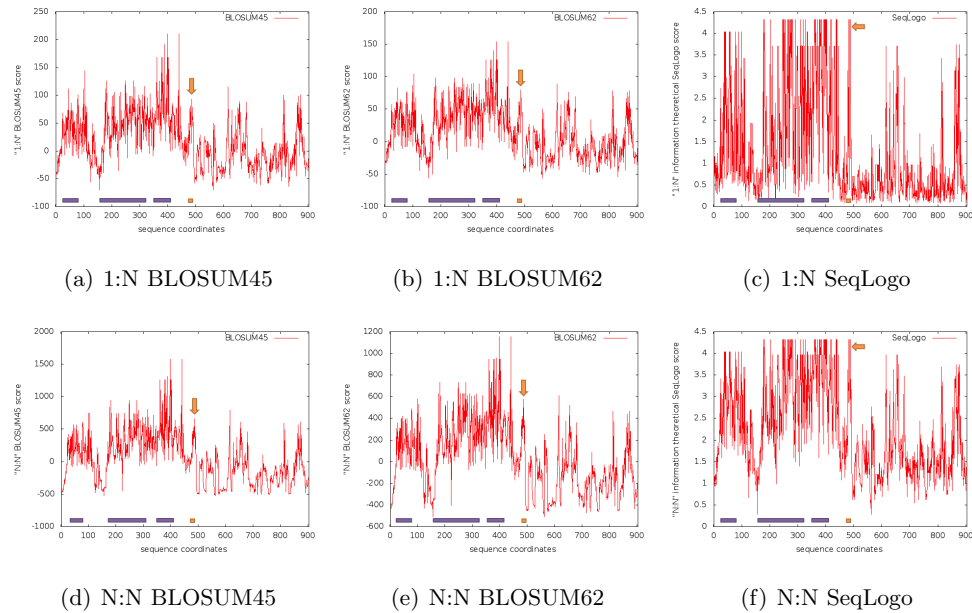
Figure 4.2. Gnuplot graphical representation of conservation scores from Siz1 'Saccharomycetales' MSA. The calculated conservation values are plotted against the *Saccharomyces cerevisiae* sequence coordinates. The scores are calculated with the substitution matrix approach (a,d: BLOSUM45; b,e: BLOSUM62) and with the information theoretical SeqLogo approach (c,f). Violet boxes indicate conserved domains as of a SAP domain (34–68), a PINIT domain (170–312) and a MIZ zinc finger (357–406). Orange box and arrow indicate the SIMa (482–488). It is shown highly conserved, whereas the level of background conservation is not.

4.3.2. Smoothing of information raw data as conservation measures

The overall conservation score of the SIM should be higher than it is for their flanking regions. The terms 'absolute conservation' and 'relative conservation' are measured with a sliding average technique on conservation scores. A sliding window with the same size as the motif returns the 'absolute conservation' of the motif ($w1$), whereas a sliding window of much larger size gives the 'background conservation' ($w2$). The difference between these sliding average values determines the 'relative' conservation (section 3.1.3.3):

$w1$: Absolute conservation of a putative motif.

$w1 - w2$: Relative conservation of a putative motif over its flanking regions.

$w2$: Background conservation of the putative motif.

Different sets of sliding average window sizes have been tested for characterisation and detection of established SIMa within the genome-wide protein sequence collection: Window sizes between 5 and 11 for absolute conservation were combined in a permutative way with sliding window sizes 25, 35, 51, 67 and 99 for the level of background conservation (table 4.7). These pairs of different window sizes define sets of three thresholds ($t1$, $t2$ and $t3$) for minimum absolute conservation, maximum background conservation and minimum relative conservation. These mathematical thresholds are specific for each $\begin{bmatrix} w1,w2 \end{bmatrix}$ pair and are defined as $w1(x)$ and $w2(x)$ with x the sequence position:

$$t1 < w1(x) \qquad t2 > w2(x) \qquad t3 < \Big(w1(x) - w2(x) \Big)$$

The large window ($w2$) is carefully chosen to include enough values to determine the motif's background conservation. Conserved flanking regions get a larger impact on the background conservation with larger sliding average windows.

The 1:N SeqLogo information theoretical approach was used for the final prediction procedure. The thresholds $t1$, $t2$ and $t3$ allow to scan a genome-wide protein sequence collection for putatively functional SIM candidates.

(a) 1:N SeqLogo

(b) 1:N SeqLogo (enlarged)
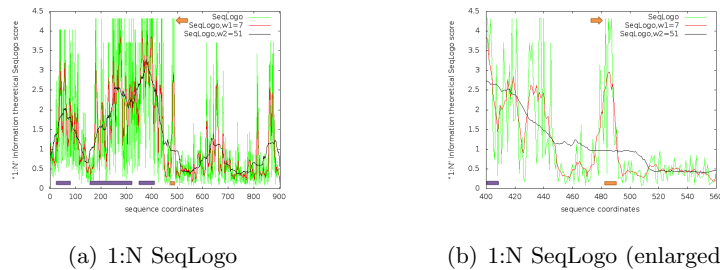
Figure 4.3. Conservation thresholds $t1$, $t2$ and $t3$ allow to describe an established SIMa. Gnuplot graphical representation of Siz1 conservation scores with employed sliding average technique. Conservation scores were calculated from 'Saccharomycetales' MSA with the information theoretical SeqLogo approach (green). Two different sliding averages $w1$=7 (red) and $w2$=51 (black) were applied. The difference between the plots $w1$ and $w2$ accounts for the relative sequence conservation. Plots for $w1$ and $w2$ show the absolute and background conservation, respectively. Violet boxes indicate conserved domains as of a SAP domain (34–68), a PINIT domain (170–312) and a MIZ zinc finger (357–406). Orange box and arrow indicate the SIMa (482–488). It is shown with high absolute (red) and moderate background conservation (black).

Table 4.7. The sliding average technique with different window sizes (w1,w2) determines a distinct conservation threshold triplet (t1,t2,t3) for established SIMa. Absolute conservation refers to t1 as a minimum information given by sliding average w1, whereas the background conservation is described by sliding average w2. Threshold t2 gives a maximum background limitation. The relative conservation of the motif is determined by the difference w1−w2 of these sliding averages. Threshold t3 gives a minimum amount for that.

protein	parameter	w1=5				w1=7				w1=9			
		w2				w2				w2			
		25	35	51	99	25	35	51	99	25	35	51	99
Fir1	t1	2.04	2.04	2.04	2.04	1.75	1.75	1.75	1.75	1.67	1.67	1.67	1.67
	t2	1.73	1.36	1.05	0.92	1.73	1.36	1.05	0.92	1.73	1.36	1.05	0.92
	t3	0.61	0.75	1.04	1.13	0.31	0.45	0.75	0.84	0.24	0.38	0.68	0.76
Uba2	t1	1.11	1.11	1.11	1.11	1.16	1.16	1.16	1.16	1.17	1.17	1.17	1.17
	t2	0.85	0.81	0.80	0.93	0.85	0.81	0.80	0.93	0.85	0.81	0.80	0.93
	t3	0.33	0.42	0.32	0.29	0.38	0.47	0.37	0.34	0.39	0.48	0.39	0.35
Slx5	t1	1.38	1.38	1.38	1.38	1.51	1.51	1.51	1.51	1.35	1.35	1.35	1.35
	t2	1.32	1.11	1.08	1.05	1.32	1.11	1.08	1.05	1.32	1.11	1.08	1.05
	t3	0.24	0.39	0.35	0.33	0.39	0.52	0.49	0.47	0.21	0.36	0.32	0.30
Slx8	t1	1.50	1.50	1.50	1.50	1.36	1.36	1.36	1.36	1.15	1.15	1.15	1.15
	t2	1.34	1.25	1.50	1.57	1.34	1.25	1.50	1.57	1.34	1.25	1.50	1.57
	t3	0.21	0.35	0.16	0.07	0.07	0.14	-0.14	-0.21	-0.14	-0.06	-0.34	-0.41
Uls1	t1	1.22	1.22	1.22	1.22	1.28	1.28	1.28	1.28	1.23	1.23	1.23	1.23
	t2	0.76	0.59	0.49	0.44	0.76	0.59	0.49	0.44	0.76	0.59	0.49	0.44
	t3	0.48	0.65	0.74	0.79	0.54	0.71	0.80	0.85	0.49	0.66	0.75	0.80
Wss1	t1	1.01	1.01	1.01	1.01	1.21	1.21	1.21	1.21	1.15	1.15	1.15	1.15
	t2	0.92	1.25	1.22	1.43	0.92	1.25	1.22	1.43	0.92	1.25	1.22	1.43
	t3	0.22	0.20	-0.16	-0.35	0.35	0.02	-0.01	-0.16	0.24	-0.09	0.02	-0.28
Siz1	t1	1.96	1.96	1.96	1.96	2.47	2.47	2.47	2.47	2.37	2.37	2.37	2.37
	t2	1.51	1.22	0.98	0.95	1.51	1.22	0.98	0.95	1.51	1.22	0.98	0.95
	t3	0.47	0.75	0.99	1.02	0.96	1.25	1.49	1.52	0.86	1.15	1.39	1.42

4.4. The bioinformatical SIM detection approach

The three SIM types have different requirements for a bioinformatical SIM detection procedure. The bioinformatical procedures are performed for each SIM type independently, whereas the basic assumptions for the approaches are the same.

4.4.1. Basic steps of the bioinformatical approach

The bioinformatical approach for a SIM detection within databases comprises the following steps:

- Established SIMs are used for SIM consensus profile construction.

- Conservation data is directly calculated from 'Saccharomycetales' and 'Saccharomycetaceae' MSA. Suitable thresholds ($t1$, $t2$ and $t3$) for the individual SIM types were chosen so that established SIMs pass them while spurious sequences do not.

- The structural contexts of established SIMs are tested with different bioinformatical tools. Such a bioinformatical tool is defined suitable when it predicts an established SIM in a non-disordered protein region.

The results from profile and consensus pattern searches are combined with conservation data and data for the structural context. The overall score is trained on established motifs. A scoring method gives lists of SIMs, in which functional motifs are enriched. The selection of SIM candidates for experimental validation is based on these overall scores together with the believed biological relevance for the respective motif.

4.4.2. SIMa

4.4.2.1. Consensus pattern and profile from established SIMa for a bioinformatical SIMa screen

Conservation criteria were optimized on established SIMa sequences in Fir1, Slx5, Slx8, Uls1, Siz1, Uba2 and Wss1 in the last section. These SIM sequences are used to derive both a SIM consensus pattern and a consensus profile from their 'Saccharomycetales' MSA was derived (figure 4.4).

For the SIMa consensus pattern, isoleucine, leucine, methionine and valine were accepted for the hydrophobic portion, while aspartate, glutamate and serine were accepted for the acid stretch. Searching a *Saccharomyces cerevisiae* protein sequence collection with a

$$
\begin{array}{ccccccccccc}
& & & \text{VILL DD} & & \text{Fir} & & \begin{pmatrix} V & I & L & L & D & D \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I & I & N & L & D & S \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ V & I & L & I & D & S \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ V & L & Q & I & S & D \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I & V & I & L & D & D \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ I & I & I & L & S & D \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ V & V & I & L & D & D \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}
\end{array}
$$

```
                                    VILL DD      ┌──────┐
                                    IINL DS      │ Fir  │
                                    VILI DS      │ Siz1 │
                                    VLQI SD      │ Slx5 │
[ILVM][ILVM].[ILVM][SDE][SDE]  ⟸    VLQI SD  ⟸   │ Slx8 │  ⟹
                                    IVIL DD      │ Uba2 │
                                    IIIL SD      │ Uls1 │
                                    VVIL DD      │ Wss1 │
                                                 └──────┘
```
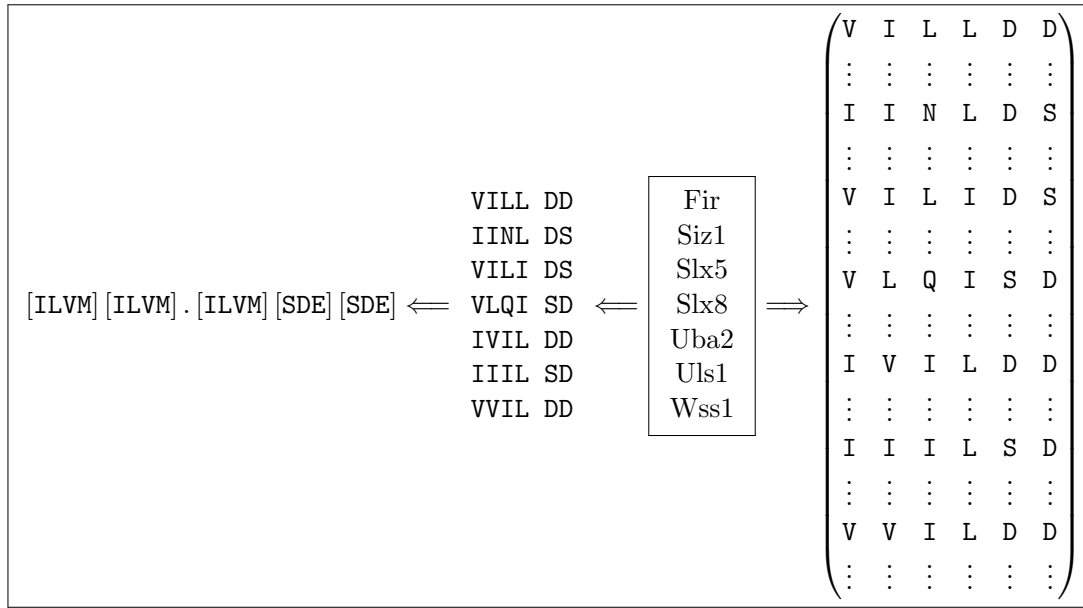
Figure 4.4. Schematic derivation of a consensus pattern (left) and a sequence profile (right) from established SIMa.

SIMa consensus pattern [ILVM][ILVM].[ILVM][SDE][SDE] returned 1955 consensus-like short sequence strings in 1264 different protein sequences, among them all established SIMa instances. (table 4.2). A preliminary profile screen of a corresponding SIMa sequence profile against a sequence collection of established SIPs gave a maximum sensitivity of $C = 4.0$ to return the established SIMs from a genome-wide sequence collection (table 4.8). Results from both the consensus pattern and the consensus profile search were taken as queries for conservation and disorder prediction data collections.

Table 4.8. Profile scores from a SIMa profile-based protein database screen returns established SIMa.

Protein	profile match	profile score
Uba2	IVILDDDE	6.961
Wss1	VVILDDDD	6.924
Fir1	VILLDEDE	6.331
Uls1	IIILSDED	5.887
Slx5	VILIDSDK	5.480
Siz1	IINLDSDD	5.258
Slx8	VLQISDDD	4.481
Nis1	IIIPDSQD	3.815

4.4.2.2. Conservation scores to discriminate between SIM-like sequences

The thresholds determine the borderline SeqLogo scores for a conservation threshold-based SIMa screen. Conservation thresholds were selected as $t1 = 1.16$, $t2 = 1.08$, $t3 = 0.37$ for 'Saccharomycetales' MSA. These thresholds were applied in the bioinformatical approach. Most established SIMs meet this conservation conditions and can be therefore selected from random motifs (table 4.7). Thresholds for Saccharomycetaceae MSA were not selected as an exclusive measure. Scores for absolute, relative and background conservation from the 'Saccharomycetaceae' MSA served as an additional measure for the quality of a putative SIM instance. Profile-based sequence scans for established and unknown SIMs were combined with a conservation threshold approach. This approach combines cut-off sensitivity from the profile-based screen and conservation thresholds. It is therefore more restrictive than the single approaches. After preliminary tests with different profile scan sensitivities and two sets of conservation thresholds, the effects of pattern-based, profile-based and conservation threshold-based search methods were investigated (table 4.9). For the final SIMa detection approach, a sensitivity of $C = 3.2$ was used for the profile-based screen in combination with a relaxed set of conservation thresholds $t1 = 1.0$, $t2 = 1.6$, $t3 = 0.2$.

Table 4.9. Effects of different combinations of sequence and conservation restrictions on the results of a genome-wide sequence search. Consensus pattern-based and profile-based search methods with and without use of conservation thresholds in a *Saccharomyces cerevisiae* protein sequence collection. [ILVM] [ILVM] . [ILVM] [SDE] [SDE] was used as a SIMa pattern.

consensus screen	$t1$	$t2$	$t3$	matching sequence positions	number of different protein sequences
-	-	-	-	2,915,898	5,885
	1.16	1.08	0.37	52,569	2,641
	1.0	1.6	0.2	227,383	3,982
pattern	-	-	-	9,330	1,264
	1.16	1.08	0.37	250	186
	1.0	1.6	0.2	913	234
profile ($C = 3.0$)	-	-	-	18,092	1,686
	1.16	1.08	0.37	431	102
	1.0	1.6	0.2	1,145	264
profile ($C = 3.2$)	-	-	-	12,498	1,247
	1.16	1.08	0.37	304	74
	1.0	1.6	0.2	1,145	264
profile ($C = 4.0$)	-	-	-	2,707	324
	1.16	1.08	0.37	123	27
	1.0	1.6	0.2	349	74

4.4.2.3. GlobPlot and IUPred disorder/globularity prediction

Disorder and globularity prediction assessments were performed for all *Saccharomyces cerevisiae* protein sequences. The use of a sliding average window was helpful using a small window size of 7 residues to gain the local disorder and globularity, respectively, of a putative motif. A sliding average of 51 residues was used to additionally indicate whether a putative SIM could lie in a disordered region and as an indicator for its distance to that disordered region. The sliding average approach is analogous to the method used to assess local conservation. A 'disorder score' was introduced as a mathematical combination of disorder assessments and their sliding averages. The results showed that established SIMs — except for Wss1 — were predicted as non-globular by both globularity predictors, GlobPlot and IUPred. The disorder prediction performed well.

4.4.2.4. Combination of sequence consensus, conservation thresholds and disorder/globularity prediction data

The previous paragraphs demonstrate that it is possible to define thresholds for consensus adherence, absolute and relative conservation, and disorder prediction, which are passed by most established SIMs but not by random sequences. Based on this finding, a universal screen for new SUMO interacting proteins was devised, which encompasses the following steps:

1) A profile scan with a reduced sensitivity of $C = 3.2$ as cut-off value was started. This cut-off value allows significant reduction of putative sequence positions (table 4.9). The returned profile scores were taken as a similarity measure to a SIMa consensus pattern.

2) Permissive thresholds for absolute, relative and background conservation were established both to a narrow ('Saccharomycetales') and wider phylogenetic range ('Saccharomycetaceae') (table 4.9). A scoring technique for the returned sequence positions was also applied to the obtained conservation scores $w1$, $w2$, $w3$ to result in a combined conservation score.

3) Disorder and globularity assignment data from GlobPlot and IUPred were combined with the results from the profile-based database screen.

Inclusion of disorder assessments into the SIM detection approach was even more restrictive. Therefore, conservation thresholds were reduced from $\begin{bmatrix} t1 = 1.16,\ t2 = 1.08, \\ t3 = 0.37 \end{bmatrix}$ to $\begin{bmatrix} t1 = 1.0,\ t2 = 1.6,\ t3 = 0.2 \end{bmatrix}$ to give more motif hits. This hit number

was further reduced by applying disorder assessment data. The final list contains all motif candidates that fulfill the thresholds for consensus similarity, conservation and disorder (tables 4.10, 4.11). The candidates are ranked by a combined score. Different predictions are shown for the narrow and wider set of MSAs, which differ mainly in their conservation scores.

The top part of the table shows the results of the scoring system applied to established SIMs (Fir1, Uls1, Slx5, Slx8, Siz1). They are predicted by GlobPlot and IUPred as expected: as non-globular and disordered. The combination with profile scores and conservation characteristics serves as a quantification method for a SIM candidate to be relevant. The results of so far unpublished SIMa-like sequences are given in the bottom part of the tables. Those putative motifs are collected by their high overall scores for different reasons:

Rad5 and Rfc1 are overall scored similar to established SIMs like the one in Fir1. Rrb1 shows too high background conservation for the given thresholds, but still good values for profile score, conservation characteristics and disorder assignments. Putative motifs in Dbp10 and Drs1 fail in the profile score cut-off values. The other values are still promising. Rad5 and Rfc1 are involved in the PCNA pathway, which is known to be regulated by SUMOylation [Torres-Ramos et al., 2002, Beckwith et al., 1998, Höge, 2002]. A SUMO interacting motif in Rad5 and Rfc1 may be biologically plausible. Rrb1, Dbp10 and Drs1 are involved in ribosomal biogenesis. SIMs in proteins with this biological background have not been reported yet. Further investigations on these proteins are carried out in this study and are described in later sections in this thesis.

Table 4.10. Final results from a genome–wide bioinformatical SIMa screen (part I). The results are compared to established SIMa in Fir1, Uls1, Slx5, Slx8 and Uls1. A profile scan with sensitivity $C = 3.2$ was run on a *Saccharomyces cerevisiae* protein sequence collection and the results were combined with conservation characteristics $t1 = 1.0$, $t2 = 1.6$, $t3 = 0.2$ and GlobPlot globularity/disorder assignments. Additional information on these established and putative SIMs are collected below (table 4.11).

| Protein | SIMa | profile score | 'Saccharomycetales' | | | 'Saccharomycetaceae' | | | GlobPlot | | | | | |
| | | | | | | | | | globular | | | disorder | | |
			raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51
Fir1	VILLDEDE	6.331	2.383	2.427	1.136	2.305	2.267	0.992	0.000	0.000	0.000	1.000	1.000	1.000
Uls1	IIILSDED	5.887	1.680	1.148	0.472	1.006	0.841	0.260	0.000	0.000	0.000	1.000	1.000	1.000
Slx5	VILIDSDK	5.480	2.249	2.103	1.054	1.692	1.497	0.589	0.000	0.000	0.000	1.000	1.000	1.000
Slx8	VLQISDDD	5.258	1.673	1.684	1.369	0.536	0.546	0.647	0.000	0.000	0.000	1.000	1.000	1.000
Siz1	IINLDSDD	4.481	2.733	2.632	0.961	2.019	1.833	0.505	0.000	0.000	0.000	1.000	1.000	0.946
Rad5	IIDLDND	5.295	1.841	1.715	0.892	0.796	0.733	0.351	0.000	0.000	0.000	1.000	1.000	1.000
Rrb1	IIEIDGDD	4.777	2.611	2.398	1.705	2.010	1.918	1.140	0.000	0.000	0.000	1.000	1.000	1.000
Dbp10	VIEYSSDE	3.296	2.985	2.814	1.638	2.300	2.199	0.958	0.000	0.000	0.000	1.000	1.000	1.000
Drs1	VPILDSSD	3.481	2.165	2.191	1.561	1.127	1.156	0.984	0.000	0.000	0.000	1.000	1.000	1.000
Rfc1	VIDLDTE	4.185	1.736	1.648	0.634	1.024	0.958	0.335	0.000	0.000	0.000	1.000	1.000	1.000

Table 4.11. Final results from a genome–wide bioinformatical SIMa screen (part II). The results are compared to established SIMa in Fir1, Uls1, Slx5, Slx8 and Uls1. Shown here is additional information on IUPred globularity/disorder prediction to table 4.10 above.

| Protein | SIMa | profile score | GlobPlot disorder | IUPred | | | | | | | | |
| | | | w2=51 | short disorder | | | long disorder | | | globular | | |
				raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51
Fir1	VILLDEDE	6.331	0.885	0.304	0.302	0.378	0.476	0.477	0.553	0.000	0.000	0.000
Uls1	IIILSDED	5.887	0.993	0.603	0.602	0.594	0.581	0.587	0.688	0.000	0.000	0.000
Slx5	VILIDSDK	5.480	0.726	0.590	0.585	0.596	0.738	0.736	0.754	0.000	0.000	0.000
Slx8	VLQISDDD	5.258	0.853	0.634	0.630	0.504	0.665	0.646	0.547	0.000	0.000	0.304
Siz1	IINLDSDD	4.481	0.843	0.757	0.749	0.747	0.880	0.870	0.854	0.000	0.000	0.000
Rad5	IIDLDND	5.295	0.887	0.380	0.381	0.375	0.461	0.456	0.496	0.125	0.179	0.441
Rrb1	IIEIDGDD	4.777	1.000	0.767	0.762	0.802	0.744	0.737	0.714	0.000	0.000	0.000
Dbp10	VIEYSSDE	3.296	1.000	0.610	0.618	0.588	0.603	0.589	0.536	0.000	0.000	0.000
Drs1	VPILDSSD	3.481	0.904	0.515	0.502	0.467	0.690	0.703	0.688	0.000	0.000	0.000
Rfc1	VIDLDTE	4.185	1.000	0.677	0.677	0.638	0.811	0.814	0.796	0.000	0.000	0.000

4.4.3. SIMb

4.4.3.1. Consensus pattern and profile from established motifs for a bioinformatical SIMb screen

The SIMb consensus pattern in SeqLogo representation (table 1.2) shows a well defined consensus pattern VIDLT. Alignment data of *Saccharomyces cerevisiae* Wss1 and Uls1 support such a profile close to that consensus pattern. An 'Saccharomycetales' MSA of Uls1 (figure 4.5) shows a conserved array of low variant alignment columns. They form the conserved SIMb in the amino terminus region of the MSA.

```
yp|YOR191W        |1|1619|-----------MAAV--PTIDLTLADSDNED-------------IFHSFSSST-----------SVDKIDIRKEN
sacch|sbayc583-g13|1|1620|-----------MAAV--PTIDLTLADSDNED-------------IFYSFSSST-----------SMDNTDVRKEN
vanpo|XP_001643729|1|1515|-----------MELI--PVIDLTSNESEEGHNNK-----------YSVFNASI-----------NHNDDTVKGTN
cagl|CAGL0G09493g |1|1408|-----------MPKV--TTIDLTLDDSEVEDYDA-----------------------------------------
zyro|ZYR00D15026g |1|1216|-----------MMGAV--PTIDLTGETLPSSS------------------------------------------
sklu|SAKL0A07546g |1|1574|-----------MNAL--PTIDLTKEDFDGEAPQVNNPMLSHKSIPVLSNFTMSRNTSTVNSTPDHTLSDVDATLDH
klth|KLTH0H06952g |1|1359|MEDQHAEVQTSVRPV--PSIDLTQDDSNPNE-------------------------------------------
kwal|Kwal_1868    |1|1357|-----------MSTI--PSIDLTGDSSDIEE-------------------------------------------
klla|KLLA0C05368g |1|1605|-------MTENGAEI--PVVNLTLDAEEEEENEVVS-----E------------LKRQETPKLFVKREIDEIEAEH
ashgo|AAR147W     |1|1580|-----------MASQV--PVVDLTSDEAEDNGRNGQQ-----DPPGALRRLYEDF----------GDGK-------
tetph|367005869  |1|1428|-----------MKI--PTIDLTTEDSDAENEQE-----------FYSMDSYP-------------------SE
torde|359750612  |1|1214|-----------MNAVSVPTIDLTLDDSDDGS-------------------------------------------
sacar|401623540  |1|1618|-----------MAAV--PTIDLTLADSDNED-------------KFYSFSSST-----------SMDNVDIRKEN
erecy|363750394  |1|1618|-----------MSGGI--PFVDLTSDDGS----NDVD-----DGLLPIGQTAESD----------GESGVL-----
naudal|365985598 |1|1484|-----------MTTAL--PTIDLTEDSDSDADENYD---------IFHSFHNTS-----------DPRNKTIETSH
```

Figure 4.5. SIMs are highly conserved and lie in more variant MSA regions. Extract from a multiple sequence alignment of *Saccharomyces cerevisiae* Uls1 (YOR191W) and orthologous sequences. The Uls1 SIMb near the protein's amino terminus outshines an area of low residue variation by its high conservation, highlighted by the graphic tool "belvu".

A preliminary SIMb profile and consensus pattern scan in the genome-wide *Saccharomyces cerevisiae* protein sequence collection with and without conservation thresholds $t1$, $t2$, $t3$ was run to fine-tune scan sensitivity and thresholds to use (tables 4.13, 4.12). These tests were set up in the same way as for the SIMa threshold determination. SIMb conservation characteristics are again presented by sliding window averages $w1 = 7$ and $w2 = 51$. Thresholds for this small test set were determined as $t1 = 1.75$, $t2 = 1.55$ and $t3 = 1.0$ (table 4.12). With these values, sequence collection scans with thresholds only or with additional use of either consensus pattern [PTVILM] [ILMV] DLT or a SIMb profile were performed (table 4.13). The results were combined with globularity and disorder assignments (table 4.14, 4.15).

Table 4.12. The sliding average technique with different window sizes ($w1$,$w2$) determines a distinct conservation threshold triplet ($t1$, $t2$, $t3$) for SIMb. Different sliding average window sizes for a shorter consensus sequences for SIMb compared to SIMa were tested on established SIMb in Uls1 and Wss1. Conservation data is based on 'Saccharomycetales' MSA. Different combinations of sliding averages of 3, 5 and 7 residues and sliding averages for background conservation of 25, 35, 51, 67 and 99 residues were investigated. Conservation characteristics were recorded as thresholds with respect to the given SIMb. A set of overall thresholds was finally derived as $t1 = 1.75$, $t2 = 1.55$, $t3 = 1.0$.

protein	parameter	w1=3 w2					w1=5 w2					w1=7 w2				
		25	35	51	67	99	25	35	51	67	99	25	35	51	67	99
Uls1	$t1$	2.35	2.35	2.35	2.35	2.35	2.21	2.21	2.21	2.21	2.21	2.23	2.23	2.23	2.23	2.23
	$t2$	1.65	1.44	1.19	1.04	0.80	1.65	1.44	1.20	1.04	0.80	1.65	1.44	1.20	1.04	0.80
	$t3$	0.89	1.03	1.24	1.38	1.60	0.74	0.88	1.10	1.24	1.46	0.75	0.87	1.10	1.24	1.47
Wss1	$t1$	2.87	2.87	2.87	2.87	2.87	2.84	2.84	2.84	2.84	2.84	2.66	2.66	2.66	2.66	2.66
	$t2$	1.90	1.68	1.55	1.27	1.33	1.90	1.68	1.55	1.27	1.33	1.90	1.68	1.55	1.27	1.33
	$t3$	1.27	1.27	1.51	1.69	1.55	1.23	1.24	1.46	1.65	1.52	1.06	1.07	1.29	1.48	1.34
overall	$t1$	2.35	2.35	2.35	2.35	2.35	2.21	2.21	2.21	2.21	2.21	2.23	2.23	2.23	2.23	2.23
	$t2$	1.90	1.68	1.55	1.27	1.33	1.90	1.68	1.55	1.27	1.33	1.90	1.68	1.55	1.27	1.33
	$t3$	0.89	1.03	1.24	1.38	1.55	0.74	0.88	1.10	1.24	1.46	0.75	0.87	1.10	1.24	1.34

Table 4.13. Comparison of results from SIMb pattern and profile scans in a *Saccharomyces cerevisiae* protein sequence collection. Several profile scan sensitivity cut-off values ($C = 3.0$, $C = 3.2$ and $C = 4.0$) were applied. The results are compared with a SIMb consensus pattern search ([PTVILM][ILMV]DLT) with and without conservation thresholds applied ($t1 = 1.75$, $t2 = 1.55$ and $t3 = 1.0$). The 7 different protein sequences after pattern scan are those of Ctf3, Dxo1, Psd1, Sap1, Ulp1, Uls1 and Wss1.

scan type	$t1$	$t2$	$t3$	matching sequence positions	different protein sequences
-	-	-	-	2,915,898	5,885
	1.75	1.55	1.0	39,629	2,962
pattern	-	-	-	450	90
	1.75	1.55	1.0	26	7
profile ($C = 3.0$)	-	-	-	29,103	2,009
	1.75	1.55	1.0	529	169
profile ($C = 3.2$)	-	-	-	20,520	1,551
	1.75	1.55	1.0	429	124
profile ($C = 4.0$)	-	-	-	3,735	349
	1.75	1.55	1.0	138	32

4.4.3.2. Combination of sequence consensus, conservation thresholds and disorder/globularity prediction data

The combination of conservation scores and disorder/globularity assessments to a ranking of putative SIMb candidates in *Saccharomyces cerevisiae* was applied to data for Wss1, Uls1 and Siz1. Additionally, a profile sensitivity of $C = 4.0$ was used for the *Saccharomyces cerevisiae* protein sequence collection scan. The filtering of the obtained motif matches by using the conservation thresholds ($t = 1.75$, $t2 = 1.55$ and $t3 = 1.0$) reduced the number of SIMb candidates to 32. Evaluation of these findings led to a selection of proteins for further investigations (tables 4.14, 4.15): As expected, the established SIMb of Uls1 and Wss1 were found in the screen with high profile scores. They show good absolute and relative conservation. However, GlobPlot gives no disorder or globularity propensities for the given SIMb. IUPred gives low propensities for disorder, but higher ones for globularity. A comparison with other putative SIMb from the collection gave mixed findings for both established SIMb and chosen candidates from a final selection:

Putative motifs in Rfc1, Sap1, Sec27, Slx4 and Tdp1 show similar tendencies in conservation characteristics as Wss1 and Uls1 SIMb. Motifs in Rfc1, Sap1 and Sec27 are similarly close to the SIMb consensus as are Wss1 and Uls1. They are predicted by IUPred to reside in putatively more likely disordered regions. Less confident disorder predictions were obtained for Slx4, Tdp1 and Top1. The Top1 motif shows poor relative conservation.

Rfc1 was already detected in a SIMa screen. Whereas Rfc1 could follow an intermediate consensus pattern, a SIMb-like sequence in Sec27 (`LIDLD`) appears atypical. Functionality of Sec27 SIMb was not expected, but tested along with the other chosen candidates from the final selection.

Table 4.14. Final results from a genome-wide bioinformatical SIMb screen (part I). The results are compared to established SIMb in Uls1 and Wss1. A profile-based scan with sensitivity $C = 4.0$ was performed in a *Saccharomyces cerevisiae* protein sequence collection. The results were combined with conservation thresholds $t1 = 1.75$, $t2 = 1.55$, $t3 = 1.0$ and GlobPlot globularity/disorder assignments. Additional information on these established and putative SIMs are collected in table 4.15 below.

Protein	SIMb	profile score	'Saccharomycetales'			'Saccharomycetaceae'			GlobPlot				
									globular			disorder	
			raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7
Uls1	TIDLT	6.559	2.896	2.490	1.147	1.144	0.966	0.429	0.000	0.000	0.000	0.000	0.000
Wss1	VIDLT	5.164	3.109	2.722	1.087	2.141	1.796	0.629	0.000	0.000	0.000	0.000	0.000
Sap1	LIDLT	6.889	2.559	1.968	0.719	1.954	1.420	0.396	0.000	0.000	0.000	0.800	0.714
Rfc1	VIDLDTE	4.530	1.563	1.532	0.635	0.916	0.911	0.338	0.000	0.000	0.000	1.000	1.000
Sec27	LIDLD	3.034	1.742	1.546	0.830	1.297	1.076	0.506	0.000	0.000	0.041	1.000	1.000
Slx4	IIDLT	5.443	2.635	2.544	1.566	1.682	1.616	1.009	1.000	1.000	1.000	0.000	0.000
Tdp1	IIDLT	6.179	1.374	1.250	0.691	0.674	0.604	0.362	0.000	0.000	0.147	0.800	0.786
Top1	IIDLT	4.023	0.758	0.886	0.880	1.797	1.956	1.722	0.000	0.000	0.000	1.000	1.000

Table 4.15. Final results from a genome-wide bioinformatical SIMa screen (part II). The results are compared to established SIMb in Uls1 and Wss1. Shown here is additional information on IUPred globularity/disorder prediction to table 4.14. The top two table entries comprise the characteristics of established SIMb in Uls1 and Wss1, the other entries show putative SIMb and their characteristics.

Protein	SIMb	profile score	GlobPlot disorder w2=51	IUPred									globular		
				short disorder			long disorder								
				raw	w1=7	w2=51	raw	w1=7	w2=51				raw	w1=7	w2=51
Uls1	TIDLT	6.559	0.393	0.455	0.442	0.381	0.434	0.421	0.421				1.000	0.971	0.909
Wss1	VIDLT	5.164	0.000	0.570	0.541	0.320	0.274	0.286	0.312				1.000	1.000	1.000
Sap1	LIDLT	6.889	0.824	0.436	0.416	0.464	0.543	0.537	0.587				0.000	0.000	0.000
Rfc1	VIDLDTE	4.530	1.000	0.679	0.678	0.639	0.814	0.815	0.796				0.000	0.000	0.000
Sec27	LIDLD	3.034	0.959	0.809	0.806	0.719	0.868	0.867	0.788				0.000	0.000	0.012
Slx4	IIDLT	5.443	0.267	0.298	0.294	0.384	0.330	0.329	0.455				1.000	1.000	0.735
Tdp1	IIDLT	6.179	0.666	0.397	0.404	0.402	0.412	0.418	0.403				1.000	1.000	1.000
Top1	IIDLT	4.032	0.892	0.277	0.272	0.232	0.389	0.386	0.320				1.000	1.000	1.000

4.4.4. SIMr

4.4.4.1. Consensus pattern from established SIMr for a bioinformatical SIMr screen

At the beginning of this study, only few SIMr data were known, including a "reverse SIMa" DDVLIV in human protein RanBP2 and two putative SIMr in *Saccharomyces cerevisiae* Elg1 [Song et al., 2004, Parnas et al., 2010]. During this study, a SIMr in Rad18 was published [Parker and Ulrich, 2012]. A *Saccharomyces cerevisiae* protein sequence collection scan was run with a reverse SIMa consensus pattern [SDE] [SDE] [ILVM] . [ILVM] [ILVM]. The results were combined with conservation thresholds from the previous SIMa and SIMb detection scans with and without employing a threshold (table 4.16). SIMa and SIMr consensus patterns show similar effects on a pattern search in a protein sequence collection. The restrictive SIMa thresholds $t1 = 1.16$, $t2 = 1.08$, $t3 = 0.37$ appeared to be too stringent for a SIMr screen with respect to a final collection of putative SUMO interacting proteins. Therefore, the most restrictive thresholds were applied and the results checked manually for biological relevance with SUMO interaction.

Table 4.16.   Genome-wide pattern-based sequence scan using three sets of conservation thresholds in synergy with SIM consensus patterns and profiles. *Saccharomyces cerevisiae* protein sequence collection scan was performed with a SIMr consensus pattern [SDE] [SDE] [ILVM] . [ILVM] [ILVM]. The findings are combined with conservation thresholds from both SIMa and SIMb screens.

consensus pattern	$t1$	$t2$	$t3$	matching sequence positions	number of different protein sequences
-	-	-	-	2,915,898	5885
	1.75	1.55	1.0	39,629	2,962
	1.16	1.08	0.37	52,569	2,641
	1.0	1.6	0.2	227,383	3,982
SIMa [ILVM] [ILVM] . [ILVM] [SDE] [SDE]	-	-	-	9330	1264
	1.75	1.55	1.0	211	68
	1.16	1.08	0.37	250	186
	1.0	1.6	0.2	913	234
SIMb [PTIVLM] [ILVM] DLT	-	-	-	450	90
	1.75	1.55	1.0	26	7
	1.16	1.08	0.37	21	5
	1.0	1.6	0.2	93	21
SIMr [SDE] [SDE] [ILVM] . [ILVM] [ILVM]	-	-	-	8502	1159
	1.75	1.55	1.0	162	51
	1.16	1.08	0.37	200	56
	1.0	1.6	0.2	744	196

The SIMr consensus pattern-based screen with [SDE] [SDE] [ILVM] . [ILVM] [ILVM] returns 51 putative SIM sequences within the *Saccharomyces cerevisiae* protein sequence database. Conservation thresholds $t1 = 1.75$, $t2 = 1.55$ and $t3 = 1.0$ were applied. Evaluation of these findings was performed analogous to the previous SIMa and SIMb screens. No profile scoring technique was used, as a suitable SIMr profile needed more data. The ranking of the SIMr pattern search was only conducted from conservation and disorder/globular assignment scores (tables 4.17, 4.18).

Selected for further investigations were a putative SIMr and a SIMa in Ulp2, which can be interpreted as SIMr with preceding serine residues, and a SIMr in Rad18 with its biological context of being involved in SUMO-dependent PCNA regulation. Conservation characteristics, disorder and globular assignments support this selection: Uls1 shows less local conservation ('raw' and 'w1') than the other candidates in the list, whereas its relative conservation ('w1-w2') is comparable. None of the candidates is predicted as being globular. Whereas one putative motif in Ulp2 (no. 2) and the one in Rad18 have less GlobPlot disorder scores, IUPred disorder works better for Ulp2 (no. 2).

Table 4.17. Final results of a genome-wide bioinformatical SIMr pattern-based screen in *Saccharomyces cerevisiae* (part I). The conservation characteristics of highly conserved SIMr instances and GlobPlot globularity/disorder assignments are shown from a final set up collection. Additional information on these putative SIMs are collected in table 4.18 below.

Protein	SIMr	'Saccharomycetales'			'Saccharomycetaceae'			GlobPlot					
								globular			disorder		
		raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51
Ulp2 (no. 1)	EEIQII	2.196	2.126	0.767	1.187	1.180	0.369	0.000	0.000	0.000	1.000	1.000	0.000
Ulp2 (no. 2)	SDVNLI	2.409	2.129	0.698	0.727	0.688	0.220	0.000	0.000	0.000	0.000	0.072	0.000
Uls1	SSIIIL	1.427	1.378	0.479	0.958	0.865	0.263	0.000	0.000	0.000	1.000	1.000	0.000
Rad18	DDLQIV	2.329	2.112	1.204	0.888	0.790	0.429	0.000	0.000	0.000	0.500	0.595	0.000

Table 4.18. Final results of a genome-wide bioinformatical SIMr screen in *Saccharomyces cerevisiae* (part II). Shown here is additional information on IUPred globularity/disorder prediction to table 4.17.

Protein	SIMr	GlobPlot disorder	IUPred								
			short disorder			long disorder			globular		
		w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51
Ulp2 (no. 1)	EEIQII	0.833	0.670	0.666	0.615	0.682	0.679	0.712	0.000	0.000	0.003
Ulp2 (no. 2)	SDVNLI	0.804	0.512	0.525	0.599	0.625	0.642	0.762	0.000	0.000	0.000
Uls1	SSIIIL	1.000	0.648	0.634	0.599	0.665	0.660	0.691	0.000	0.000	0.000
Rad18	DDLQIV	0.941	0.464	0.469	0.500	0.465	0.477	0.456	0.000	0.000	0.000

4.4.5. Selection of possible SUMO interacting proteins for validation

The following putative SUMO interacting motifs in *Saccharomyces cerevisiae* were selected from the final selection lists to show that the approach outlined above is able to enrich SIMs from a collection of protein sequences (table 4.19).

The selection is based on data from profile-based scans for new SIMs in a *Saccharomyces cerevisiae* protein sequence collection. The results were then combined with conservation data from 'Saccharomycetales' MSA and disorder/globular assignment data from GlobPlot and IUPred. Visual inspection of the final results was performed to further narrow the number of putative interacting motifs.

Rad5 is a known DNA helicase and ubiquitin ligase [Torres-Ramos et al., 2002, Parker and Ulrich, 2009, Carlile et al., 2009, Halas et al., 2011]. The proteins Rfc1 and Rad18 are known to be involved in the regulation of the SUMOylated protein PCNA [Bailly et al., 1997, Beckwith et al., 1998, Parker and Ulrich, 2012]. A SUMO interacting motif in those proteins is therefore biologically plausible. The putative SIM in Rad5 would be a SIMa type, in Rfc1 a SIMa/b intermediate and in Rad18 a SIMr type. Their high profile scores show a close sequence similarity to established SIMs. Combined with high values for absolute and relative conservation, these putative SIMs from the bioinformatical screen offer the possibility to investigate its performance on enriching a protein sequence collection. A putative motif sequence `LIDLD` in Sec27 would be similar to a SIMb consensus, but with a glutamate instead of a threonine. The sequence can be described as intermediate between the Uls1 SIMb `TIDLT` and a putative Rfc1 motif `VIDLDT`. Another group of SIM candidates in the proteins Rrb1, Dbp10 and Drs1 was selected due to their involvement in ribosomal biogenesis. So far, a regulation of this process by SUMO has not been described. Despite higher background (Rrb1) and low consensus pattern similarity (Dbp10, Drs1), SIM validation would open a new field of research for SUMO. In a second phase of this study, additional SIM candidates with lower relative conservation but a high profile score combined with low (Top1), or high (Slx4) absolute conservation were selected. Tdp1 shows good absolute conservation. Slx4 is predicted globular by both IUPred and GlobPlot. A SIM in Slx4 is not expected.

Table 4.19. Final selection of putative motifs for experimental validation from the bioinformatical genome-wide SIM screens. The final selection lists are set up for enrichment of SUMO interacting motifs. Further validation experiments on these proteins was performed.

SIM type	protein	sequence motif	comment
SIMa	Rad5	IIDL DN	-
	Rrb1	IIEI DG	higher background conservation
	Dbp10	VIEY SS	low profile score
	Drs1	VPIL DS	low profile score
	Rfc1	VIDL DT	-
SIMb	(Rfc1)	(VIDL DT)	-
	Sap1	LIDLT	-
	Sec27	LIDLD	putatively new consensus pattern
	Slx4	IIDLT	globular by GlobPlot
	Tdp1	IIDLT	low absolute conservation
	Top1	EIDLT	low relative conservation
SIMr	Rad18	(DD LQIV)	validated, [Parker and Ulrich, 2012]

4.5. Application of analogous methods for a genome-wide AIM screen in Saccharomyces cerevisiae

The characterized AIMs of the budding yeast autophagy regulators Atg1, Atg3, Atg4, Atg19, Atg32 and Atg34 and their homologs from other yeast species were used to derive an AIM-specific consensus pattern and sequence profile (table 1.4, figure 4.6). The consensus pattern has a low information content due to its short sequence and lack of additional sequences. An AIM consensus pattern is composed of a segment of four residues with sequence constraints at positions one and four. This stretch is usually preceded by acidic residues, neglected in figure 4.6. The two middle positions give a broad variation within the MSA. Therefore, these positions were left open completely in a putative consensus pattern [WYF] .. [ILV].

AIM are short linear motifs which adopt a β-strand structure upon binding to Atg8. But they are located in an otherwise more variant protein region. These features are also common with SIM. Therefore, the SIM prediction procedure may be adapted to an AIM prediction method, as well. Conservation data from 'Saccharomycetales' MSA were derived either by the substitution matrix approach or the information theoretical SeqLogo approach for established AIM (table 4.20). The information theoretical SeqLogo scores show the expected similarities to the SIM information scores: Conservation scores

Figure 4.6. Schematic derivation of a consensus pattern (left) and a sequence profile (right) from established AIM. The profile is represented as composed of a stack of single AIM. The pattern can be described as the overall composition of several sequences at once, whereas a profile is derived from aligned AIM instances and therefore has a much higher information content than its respective sequence pattern.
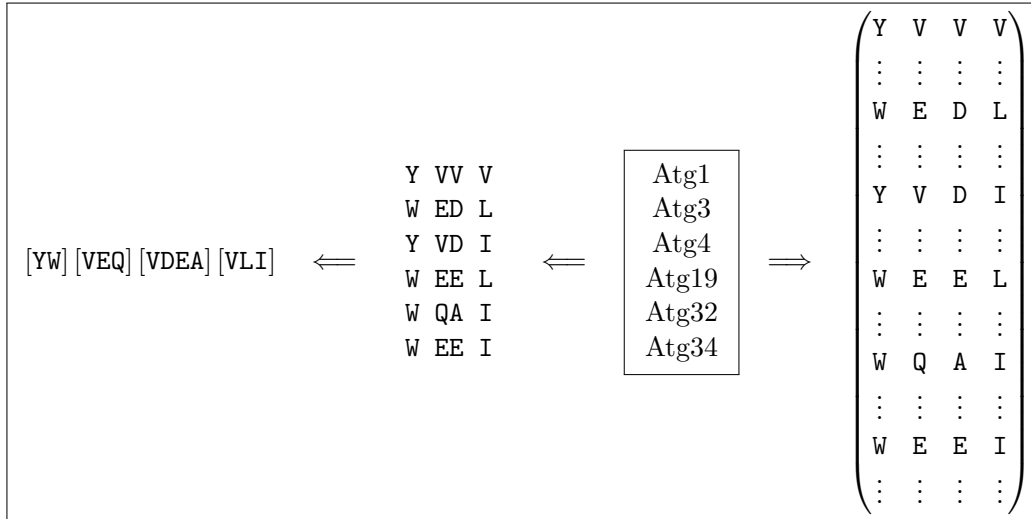
based on BLOSUM matrices are less suited for the detection of conserved motifs on a non-motif background than the information-theoretical scores of the SeqLogo approach. The SeqLogo scoring technique serves as descriptor of the information content of each AIM position within the 'Saccharomycetales' MSA.

First, a test for a set of conservation thresholds $t1$, $t2$, $t3$ most suitable for AIM detection was performed; the procedure was analogous to that employed for SIM binding as described above. Different permutative combinations of window sizes 3, 5 and 7 ($w1$) with 25, 35, 51, 67 and 99 ($w2$) were investigated (table 4.21): Window sizes 3 and 5 resulted in higher information contents and in higher absolute conservation values $t1$ than for a window of 7 residues. The differences between a 3– and a 5–residue window were marginal. Larger window sizes for background conservations returned larger information levels with increasing window size. However, the interplay of small and large window sizes for absolute, background and relative conservation limited the choice of suitable sliding average windows and conservation thresholds. Thresholds $t1$, $t2$, $t3$ for AIM detection were set for window sizes $w1 = 5$ and $w2 = 35$.

Table 4.20. The level of AIM conservation can be calculated the same way as for SIMs using substitution matrix and information theoretical approaches. 'One-against-all' (1:N) and 'all-against-all' (N:N) approaches were applied with these methods.

Protein	AIM	BLOSUM45		BLOSUM62		SeqLogo	
		1:N	N:N	1:N	N:N	1:N	N:N
Atg3	W	255	2295	187	1683	4.322	4.322
	E	102	918	85	765	4.322	4.322
	D	119	1071	102	918	4.322	4.322
	L	74	596	60	490	2.932	3.519
Atg19	W	66	94	46	48	1.271	2.178
	E	44	176	38	161	2.257	3.009
	E	44	176	38	161	2.257	3.009
	L	0	188	-9	146	0.752	3.009
Atg32	W	210	1575	154	1155	4.322	4.322
	Q	79	560	65	455	3.704	3.969
	A	27	151	21	118	0.920	2.299
	I	70	525	56	420	4.322	4.322
Atg34	W	81	175	57	105	1.398	2.272
	E	50	226	43	204	2.362	3.070
	E	50	226	43	204	2.362	3.070
	I	5	193	-5	141	0.915	2.974

The AIM in Atg19 is less well conserved than the other training AIMs, and require the use of rather permissive threshold values. Atg1 showed high values for both absolute and background conservation. High background conservation values affected the derivation of thresholds for less constraint values and therefore for more putative hits from a screen. Three different sets of conservation thresholds can be considered, depending on a tradeoff between sensitivity and selectivity (tables 4.22, 4.23). The established AIM in Atg1 has a low profile score of 1.552 due to its atypical AIM sequence YVVV. In its 'Saccharomycetales' MSA, this motif is at the C-terminal end of a highly conserved region.

Table 4.21. The sliding average technique with different window sizes ($w1$, $w2$) determines a distinct conservation threshold triplet ($t1$, $t2$, $t3$) for AIM. Sliding average window sizes of 3, 5 and 7 were combined with sliding averages for background conservation of 25, 35, 51, 67 and 99 residues. Conservation characteristics were recorded as thresholds with respect to the given AIM. A combination of sliding averages of 5 and 35 alignment positions was found useful. Two sets of overall thresholds were finally set as $t1 = 1.42$, $t2 = 2.12$, $t3 = 0.6$ (neglecting the Atg1 and Atg19 AIM) and $t1 = 1.42$, $t2 = 2.62$, $t3 = 0.6$ (only neglecting Atg19 AIM).

| protein | parameter | $w1 = 3$ | | | | | $w1 = 5$ | | | | | $w1 = 7$ | | | | |
| | | $w2$ | | | | | $w2$ | | | | | $w2$ | | | | |
		25	35	51	67	99	25	35	51	67	99	25	35	51	67	99
Atg1	$t1$	3.81	3.81	3.81	3.81	3.81	3.31	3.31	3.31	3.31	3.31	3.44	3.44	3.44	3.44	3.44
	$t2$	3.09	2.62	2.00	1.65	1.28	3.09	2.62	2.00	1.65	1.28	3.09	2.62	2.00	1.65	1.28
	$t3$	0.72	1.20	1.81	2.16	2.54	0.61	0.85	1.35	1.71	2.05	0.73	0.98	1.48	1.84	2.18
Atg3	$t1$	3.47	3.47	3.47	3.47	3.47	3.07	3.07	3.07	3.07	3.07	2.64	2.64	2.64	2.64	2.64
	$t2$	2.08	2.12	2.32	2.59	2.82	2.08	2.12	2.32	2.59	2.82	2.08	2.12	2.32	2.59	2.82
	$t3$	1.72	1.43	1.16	0.89	0.65	1.28	1.03	0.75	0.49	0.25	0.89	0.60	0.32	0.06	-0.18
Atg4	$t1$	2.38	2.38	2.38	2.38	2.38	2.34	2.34	2.34	2.34	2.34	2.37	2.37	2.37	2.37	2.37
	$t2$	1.12	0.92	1.05	1.12	1.17	1.12	0.92	1.05	1.12	1.17	1.12	0.92	1.05	1.12	1.17
	$t3$	1.29	1.47	1.34	1.27	1.26	1.24	1.43	1.30	1.22	1.21	1.28	1.46	1.33	1.25	1.25
Atg19	$t1$	1.23	1.23	1.23	1.23	1.23	1.29	1.29	1.29	1.29	1.29	1.07	1.07	1.07	1.07	1.07
	$t2$	0.73	0.69	0.86	0.84	0.66	0.73	0.69	0.86	0.84	0.66	0.73	0.69	0.86	0.84	0.66
	$t3$	0.58	0.54	0.33	0.43	0.48	0.64	0.60	0.38	0.49	0.53	0.42	0.39	0.17	0.27	0.16
Atg32	$t1$	2.07	2.07	2.07	2.07	2.07	2.00	2.00	2.00	2.00	2.00	2.11	2.11	2.11	2.11	2.11
	$t2$	1.59	1.41	1.35	1.56	1.64	1.59	1.41	1.35	1.56	1.64	1.59	1.41	1.35	1.56	1.64
	$t3$	0.48	0.67	0.72	0.52	0.44	0.41	0.60	0.66	0.45	0.37	0.52	0.71	0.76	0.56	0.48
Atg34	$t1$	1.36	1.36	1.36	1.36	1.36	1.42	1.42	1.42	1.42	1.42	1.18	1.18	1.18	1.18	1.18
	$t2$	0.87	0.73	0.77	0.83	0.78	0.87	0.73	0.77	0.83	0.78	0.87	0.73	0.77	0.83	0.78
	$t3$	0.75	0.66	0.59	0.54	0.60	0.63	0.72	0.65	0.60	0.66	0.39	0.48	0.42	0.37	0.42

Table 4.22. Derivation of conservation thresholds for AIM detection. Three sets of thresholds $t1$, $t2$, $t3$ could be determined from combinations of different sliding average windows. The test set comprised established AIM in Atg1, Atg3, Atg4, Atg19, Atg32 and Atg34 in *Saccharomyces cerevisiae*. With final sliding windows of 5 and 35 positions, the sets vary depending on the different conservation characteristics of the AIM included. Whereas the first row comprises the data from all AIM mentioned, the second and third lack Atg19 AIM and the AIM of Atg19 and Atg1, respectively.

included established AIM	$t1$	$t2$	$t3$
Atg1 , Atg3, Atg4, Atg19 , Atg32, Atg34	1.29	2.62	0.6
Atg1 , Atg3, Atg4, Atg32, Atg34	1.41	2.62	0.6
Atg3, Atg4, Atg32, Atg34	1.42	2.12	0.6

Table 4.23. Profile scores of established AIM. A profile was derived from established AIM in Atg1, Atg3, Atg4, Atg19, Atg32 and Atg34. A profile-based scan within a *Saccharomyces cerevisiae* protein sequence collection gave the presented profile scores.

Protein	AIM	profile score
Atg1	EYVVV	1.552
Atg3	DWEDL	5.521
Atg4	DYVID	3.980
Atg19	TWEEL	3.420
Atg32	SWQAI	2.953
Atg34	TWEEI	3.747

This region is large enough to give good absolute conservation scores for the AIM, but short enough to show good relative conservation with larger sliding windows $w2$. According to these scores, the adapted approach may be suitable for the detection of AIMs. The high level of background conservation interferes with the characteristics of the other AIM in terms of conservation and consensus pattern adherence. Therefore, the AIM in Atg1 is not considered any further to improve the conditions for parameter adjustment to the other established AIM (table 4.24). The used AIM profile was modified according to that. Improvements on profile scores compared to the previous profile screen (table 4.23) can be observed with the altered AIM profile (table 4.24). With these newly constructed profiles the genome-wide *Saccharomyces cerevisiae* protein sequence collection was searched for AIM consensus-like sequences (table 4.25) with the pattern [WYF]..[ILV] with and without application of conservation thresholds (table 4.25). For an AIM detection approach similar for the ones established with SIMs in this study, a profile scan was performed with a cut-off value of $C = 3.0$, which leaves the established AIM of Atg3, Atg4, Atg19, Atg32 and Atg34 within the collection of 1,312 putative AIM proteins (table 4.25). Additional use of conservation thresholds $t = 1.42$, $t2 = 2.62$ and $t3 = 0.60$ helped reduce the number of putative AIM candidates (tables 4.26, 4.27).

Table 4.24. Profile scores to established AIM. A profile was derived from established AIM in Atg3, Atg4, Atg19, Atg32 and Atg34. Atg1 was omitted. A profile scan within a *Saccharomyces cerevisiae* protein sequence collection gave the here presented profile scores.

Protein	AIM	profile score
Atg3	DWEDL	5.660
Atg4	DYVID	3.340
Atg19	TWEEL	3.862
Atg32	SWQAI	3.103
Atg34	TWEEI	4.102

Table 4.25. AIM consensus pattern search using [WYF]..[ILV] and a profile search based on MSA of established AIM was performed in a *Saccharomyces cerevisiae* protein sequence collection.

consensus screen	t1	t2	t3	matching sequence positions	number of different protein sequences
pattern	-	-	-	226,436	5,667
	1.29	2.62	0.60	35,804	4,025
	1.42	2.62	0.60	35,586	4,021
	1.42	2.12	0.60	8,030	1,731
profile ($C = 3.0$)	-	-	-	8,189	1,312
	1.29	2.62	0.60	1,303	445
	1.42	2.62	0.60	1,229	439
	1.42	2.12	0.60	869	304
profile ($C = 3.2$)	-	-	-	5,927	1,006
	1.29	2.62	0.60	947	339
	1.42	2.62	0.60	937	334
	1.42	2.12	0.60	613	227
profile ($C = 4.0$)	-	-	-	1,320	249
	1.29	2.62	0.60	279	91
	1.42	2.62	0.60	279	91
	1.42	2.12	0.60	208	68

Table 4.26. Final results from the bioinformatical AIM screen (part I). The results were compared to the values of established AIM. A profile scan with sensitivity $C = 3.0$ was run on a *Saccharomyces cerevisiae* protein sequence collection and the results were combined with conservation characteristics $t1 = 1.75$, $t2 = 1.55$, $t3 = 1.0$ (w1=5, w2=51) and GlobPlot globularity/disorder assignments. Shown here are results from a final set up collection. Additional information on these established and putative SIMs are collected in table 4.27 below.

Protein	AIM	profile score	'Saccharomycetales'			'Saccharomycetaceae'			GlobPlot				
									globular			disorder	
			raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7	w2=51	raw	w1=7
Atg3	DWEDL	5.660	3.535	3.253	2.036	1.629	1.413	1.189	0.000	0.000	0.000	1.000	1.000
Atg4	DYVDI	3.340	3.071	2.837	0.881	1.729	1.556	0.436	1.000	1.000	0.627	0.000	0.000
Atg19	TWEEL	3.862	1.331	1.359	0.687	1.333	1.356	0.654	1.000	1.000	1.000	0.000	0.000
Atg32	SWQAI	3.103	3.445	2.806	1.319	3.254	2.628	1.209	0.000	0.000	0.000	0.000	0.000
Atg34	TWEEI	4.102	1.464	1.484	0.701	1.445	1.456	0.656	0.000	0.000	0.000	0.000	0.000
Atg1	EYVVV	—	3.753	3.699	2.528	3.497	3.489	2.041	0.000	0.000	0.000	0.000	0.000

Table 4.27. Final results from the bioinformatical AIM screen (part II). The results were compared to the values of established AIM. Shown here is additional information on IUPred globularity/disorder prediction to table 4.26.

Protein	AIM	profile score	GlobPlot disorder w2=51	IUPred								
				short disorder			long disorder			globular		
				raw	w1=7	w2=51	raw	w1=7	w2=52	raw	w1=7	w2=52
Atg3	DWEDL	5.660	0.549	0.343	0.325	0.194	0.383	0.371	0.253	1.000	1.000	1.000
Atg4	DYVDI	3.340	0.471	0.299	0.308	0.347	0.274	0.283	0.366	1.000	1.000	1.000
Atg19	TWEEL	3.862	0.322	0.783	0.746	0.413	0.447	0.443	0.374	0.800	0.848	0.964
Atg32	SWQAI	3.102	0.141	0.354	0.359	0.427	0.465	0.476	0.553	0.000	0.000	0.000
Atg34	TWEEI	4.102	0.320	0.749	0.708	0.518	0.353	0.345	0.410	1.000	1.000	1.000
Atg1	EYVVV	—	0.549	0.289	0.289	0.415	0.411	0.416	0.560	0.000	0.000	0.000

4.6. Experimental validation of SUMO interacting motifs in Saccharomyces cerevisiae

The bioinformatical screens gave lists of SIM-like short sequences and their corresponding proteins. Among these protein sequences, putatively functional SIMs are believed to be enriched. After performance of bioinformatical prediction procedures for SIMa (section 4.4.2), SIMb (section 4.4.3) and SIMr (section 4.4.4), a selection was made for new putatively functional motifs for experimental interaction studies (tables 4.19, 4.28). Three experiments were performed:

First, truncated protein sequences of Sec27, Rad5, Rrb1, Dbp10, Drs1, Rfc1 and Rad18 from the bioinformatical screens were investigated in yeast two-hybrid experiments. Rfc1 and Rad18 show SUMO interaction. In a second experiment, these proteins were tested as full length-proteins using the GST-SUMO pulldown technique. SUMO interaction of Rfc1 and Rad18 was confirmed. Dbp10 and Drs1 show SUMO interaction, in contrast to negative interaction results from yeast two-hybrid experiments. In an additional yeast two-hybrid experiment, putative SIMs in Slx4, Tdp1 and Top1 from the bioinformatical SIMb detection screen were tested (table 4.28). SIMb instances in these proteins appeared less favourable for SUMO interaction than the instances in the previous experiments. SUMO interaction is shown for Tdp1.

Table 4.28. Selection of putative SIM sequences from the previous bioinformatical screens for experimental validation. SUMO interaction is shown for Rfc1, Rad18 and Tdp1 in yeast two-hybrid experiments. These interactions are further shown in GST-pulldown assays for Rfc1 and Rad18, and additionally for Dbp10 and Drs1. SUMO/SIM interaction for Rfc1 is shown in growth phenotype analysis upon chemically induced DNA replication stress.

SIM type	protein	sequence motif	coordinates	SUMO interaction?	SUMO/SIM interaction?
SIMa	Rad5	IIDL DN	$372 - 377$	☐	☐
	Rrb1	IIEI DG	$55 - 60$	☐	☐
	Dbp10	VIEY SS	$60 - 65$	$\sim$	☐
	Drs1	VPIL DS	$24 - 29$	$\sim$	☐
	Rfc1	VIDL DT	$33 - 38$	☒	☒
SIMb	(Rfc1)	(VIDLD)	$(33 - 37)$	(☒)	(☒)
	Sap1	LIDLT	$235 - 239$	☒	☐
	Sec27	LIDLD	$823 - 827$	☐	☐
	Slx4	IIDLT	$470 - 474$	☐	☐
	Tdp1	IIDLT	$57 - 61$	☒	☐
	Top1	EIDLT	$608 - 613$	☐	☐
SIMr	Rad18	DD LQIV	$137 - 142$	☒	☐

Rad18 and Rfc1 are involved in DNA binding and DNA damage repair [Bailly et al., 1997, Yao, 2003, Parker et al., 2008, Parker and Ulrich, 2012]. The biological relevance of SUMO/SIM interaction for these proteins was about to be tested in stress tests. However, SUMO interaction for Rad18 was meanwhile shown by Parker and Ulrich [Parker and Ulrich, 2012]. They also show a SIMr instance responsible for the interaction [Parker and Ulrich, 2012]. A SIM mutation in RFC1 was introduced into *Saccharomyces cerevisiae* strain JD47-13C. That *rfc1* SIM defective strain was exposed to thermal and chemical stress. It showed an observable phenotype compared to the wild-type strain.

4.6.1. Verification of SUMO/SIM interaction in Rfc1 using the yeast two-hybrid technique

The core component in this yeast two-hybrid approach is the *Saccharomyces cerevisiae* Gal4 transcriptional factor which regulates Gal4 expression resulting in the activation of a reporter gene (section 3.2.16) [Joung et al., 2000]. This transcriptional factor is composed of two physically separateable, functionally independent domains: Gal4 activating domain (GAD) and Gal4 binding domain (GBD). Cloning of gene fusions to these domains allows validation of interaction between the coding protein sequences. Interaction of the expressed proteins reassembles GAD and GBD to restore Gal4 transcriptor activity for *HIS3* and *ADE2* reporter gene expression.

Truncated protein versions of Sec27, Rad5, Rrb1, Dbp10, Drs1, Rfc1 and Rad18 comprising the respective SIM sequences were used in a first yeast two-hybrid experiment. In a second yeast two-hybrid experiment full-length proteins of Tdp1, Top1 and Slx4 were used. They were cloned and expressed as GAD fusions to their amino termini (table 4.29). The constructed plasmids were isolated and verified by their unique restriction pattern in agarose gel electrophoresis and by Sanger sequencing. GBD was cloned to nucleotide sequences coding for *Saccharomyces cerevisiae* SUMO paralog Smt3 and different Smt3 chain versions. These ΔG97 truncated Smt3 chains were used to investigate whether interaction differences occur with different Smt3 chain lengths. These GBD constructs were expressed as GBD fusions to the amino termini of Smt3. GBD-Smt3, GBD-Smt3Δ97, GBD-Smt3$\times$3 and GBD-Smt3$\times$4 fusions from different plasmids were taken from the Dohmen *et al.* lab collection. Expression of these GBD fusion proteins were shown in western blots in previous studies using α-GST-SUMO antibody (table 2.8). Plasmids only expressing GAD and GBD both with and without C-terminal HA tag were used as controls. They are referred to as 'GAD-C1' and 'GBD-C1' hereafter.

Table 4.29. GAD protein fusions with their different insert sizes used in this study. A plasmid expressing only the Gal4 binding domain was used with and without C-terminal HA tag as negative control in the yeast two-hybrid experiments. A C-terminal HA tag was introduced for expression level visualization using western blotting. The top part of the table comprises the GAD fusion proteins for a first yeast two-hybrid approach. After yeast two-hybrid experiments had only shown SUMO interaction for Rfc1 and Rad18, Tdp1, Top1 and Slx4 were investigated.

GAD protein fusion	inserted nucleotide sequence	
GAD	-	
GAD-HA	-	
GAD-Sec27-HA	Sec27,	930-1260
GAD-Rad5-HA	Rad5,	1-1260
GAD-Rrb1-HA	Rrb1,	1-615
GAD-Dbp10-HA	Dbp10,	1-1529
GAD-Drs1-HA	Drs1,	1-1731
GAD-Rfc1-HA	Rfc1,	1-705
GAD-Rad18-HA	Rad18,	1-801
GAD-Tdp1	Tdp1,	full length
GAD-Top1	Top1,	full length
GAD-Slx4	Slx4,	full length

Yeast two-hybrid experiments were performed (section 3.2.16). An HA tag was introduced for expression control in western blots (figure 4.7). The expression levels differ between fusion proteins. GAD-Sec27-HA, GAD-Dbp10-HA and GAD-Rrb1$^{1-540}$-HA are not sufficiently expressed. No SIM interaction results for GAD-Sec27-HA and GAD-Dbp10-HA are presented due to their low expression detected in the western blot analysis. The other proteins are expressed at higher levels.

SUMO interaction was detected as growth on synthetic drop-out (SD) media plates lacking leucine and tryptophan (SD-LW) and additionally either histidine (SD-LWH) or adenine (SD-LWA). 3-amino-1,2,4-1H-triazole (3-AT) is a competitive inhibitor to *HIS3* [Struhl and Davis, 1980, Brennan and Struhl, 1980, Joung et al., 2000]. Cells need higher expression levels of histidine for survival on SD-LWH selection plates supplemented with 3-AT. Experiments were repeated several times with equal results in both PJ69-4A and AH109 yeast strains (figure 4.8). From these yeast two-hybrid experiments, only GAD-Rfc1-HA and GAD-Rad18-HA show SUMO interaction (figure 4.8). SUMO interaction of the truncated versions of Sec27, Rad5, Rrb1, Dbp10 and Drs1 could not be observed (all not shown). GAD-Sap1 was chosen from the bioinformatical screen as an established SUMO interactor [Hannich et al., 2005]. Slx4, Tdp1 and Top1 were selected from the bioinformatical SIMb screens (tables 4.13, 4.19), despite their putative SIMb are either predicted to reside in a globular protein region (Slx4) or show low
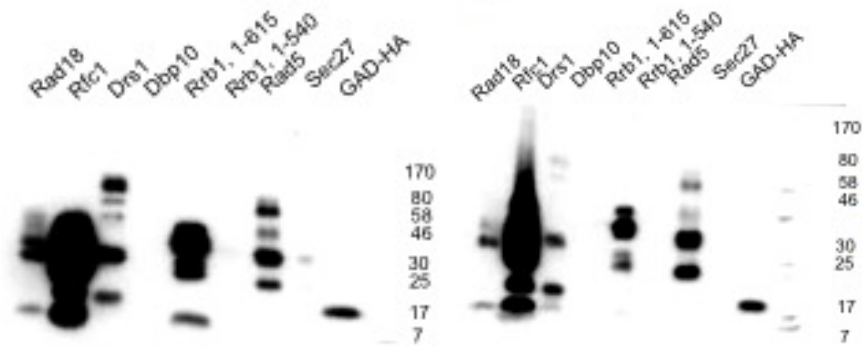
Figure 4.7. **α-HA signal intensity (3F10) as a measure for protein expression level.** Expression of GAD fusions of Rad18, Rfc1, Drs1, Dbp10, Rrb1, Rad5, Sec27 and unligated GAD-HA in *Saccharomyces cerevisiae* strain JD47-13C (table 4.29). Cell extracts were lysed with either the glass bead method (left) or direct boiling (right) (section 3.2.13). GAD-Sec27-HA, GAD-Rrb1[1-540]-HA and GAD-Dbp10-HA are not expressed.

relative conservation scores (Tdp1, Top1). Slx4, Tdp1 and Top1 were cloned using the Gateway™ protocol (section 3.2.11).

Yeast two-hybrid experiments are interpreted as follows: Growth on SD-LW serves as an indicator to correctly expressing both GAD and GBD fusion proteins. Growth on other plates was used as an indicator to SUMO interaction. Growth intensity and the plate type in the order SD-LWH < SD-LWH+3AT < SD-LWA are used as a measure for interaction intensity. Tdp1 shows interaction with all Smt3 versions on SD-LWH, whereas its interaction to Smt3Δ97 is suppressed upon 3-AT addition. The interaction of Tdp1 for conjugatable Smt3 is stronger than the Smt3Δ97 truncation and responsible for growth on SD-LWH+3AT and SD-LWA plates. GAD-Rad5-HA, GAD-Rrb1[1-615]-HA and GAD-Drs1-HA show no growth on other than SD-LW plates. Rfc1 has a preference for conjugatable Smt3, as strains expressing GAD-Rfc1-HA and GBD fusions of Smt3 and Smt3×3 chains grow on SD-LWH plates. That growth is prevented upon addition of 5 mM 3-AT to the SD-LWH plates and on SD-LWA plates. From these plates, it cannot not be concluded why Rfc1 is not interacting to Smt3×4 chains. GAD-Rad18-HA expressing strains grow on SD-LWH plates (and more stringent plates) when additionally expressed with GBD fusions of Smt3, Smt3×3 or Smt3×4. 3-AT reduces growth of these transformants to a similar extent as on SD-LWA plates for Rad18, Tdp1 and Sap1. These growth phenotypes indicate SUMO interaction (Smt3 and Smt3×3) with these fusions. Rad18 additionally interacts even better with the Smt3×4 GBD fusion. Rad18 therefore shows more preference for SUMO chain interaction than for the mono-Smt3 moiety (GBD-Smt3Δ97).
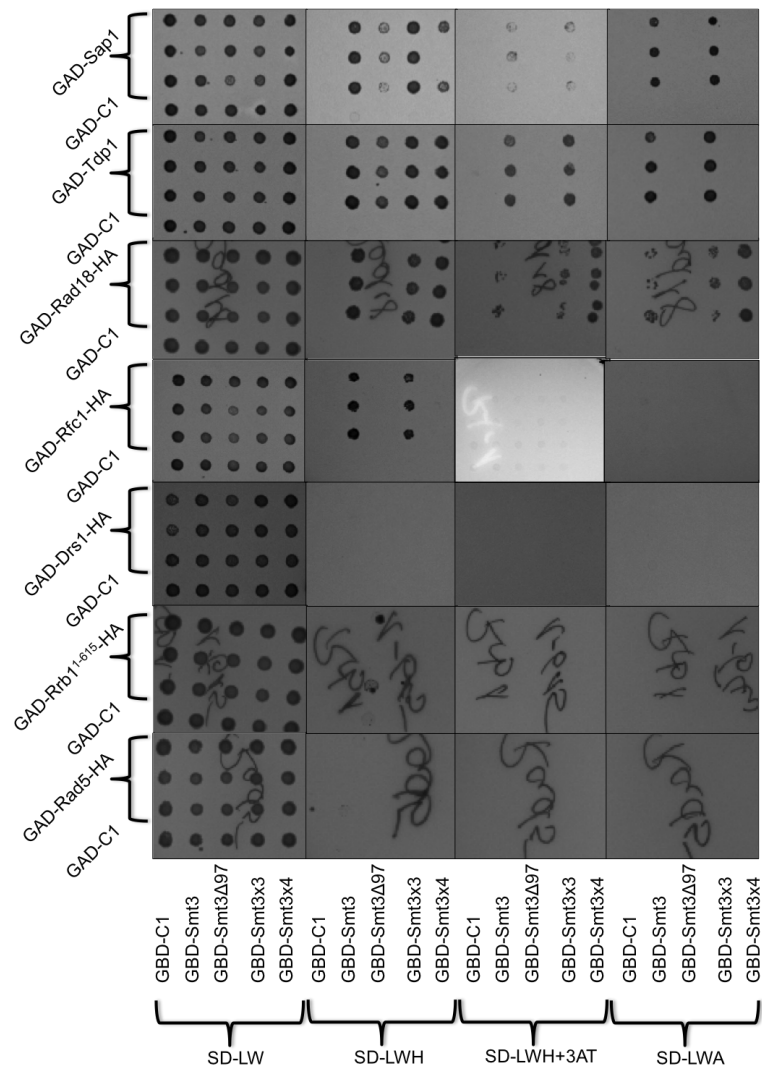
Figure 4.8. Rad18, Rfc1 and Tdp1 show SUMO interaction in yeast two-hybrid experiments. Results from yeast two-hybrid experiments are shown for Rad5, Rrb1, Drs1, Rfc1, Rad18, Tdp1, Sap1: Each spot on a plate corresponds to an independently grown transformant colony comprising one GAD and one GBD fusion protein expressing plasmid in *Saccharomyces cerevisiae* strain PJ69-4A. Equal results were obtained when yeast strain AH109 was used. Equal amounts from aqueous cell suspensions corresponding to an optical density of 10 OD_{600} were spotted onto SD-LW (first column), SD-LWH (second column), SD-LWH supplemented with 5 mM 3-AT (third column) and SD-LWA plates (last column) and incubated for 4 days at 30°C. Each plate was spotted with three independently grown transformants expressing the GAD fusion protein and one shown transformant expressing only GAD as a control, respectively, and expressing the GBD fusion protein and one shown transformant expressing only GBD as a control, respectively. SD plates are labeled in a way that each row corresponds to a distinct GAD fusion protein and each column to a distinct GBD fusion protein expressed in the respective strain.

4.6.2. Verification of SUMO/SIM interaction in GST-SUMO pulldown assays

For confirmation of SUMO interaction for Rfc1 and Rad18, GST pulldown assays were performed (section 3.2.17). Sec27, Rad5, Rrb1, Dbp10, Drs1, Rfc1 and Rad18 were used as full length proteins with C-terminal 6×HA-tag (table 4.30). Tdp1 was omitted from this test, as the SIM candidates in Tdp1, Top1 and Slx4 were investigated in yeast two-hybrid studies after GST-pulldown assays had been completed for the first set of proteins. Nis1 was used as a positive SUMO interaction control [Hannich et al., 2005, Uzunova et al., 2007]. The respective genes were genetically altered with a sequence encoding a C-terminal 6×HA tag using a PCR-based epitope tagging technique (section 3.2.17) [Longtine et al., 1998, Knop et al., 1999, Janke et al., 2004]. For this study, the pYM14 plasmid bearing the kanMX gene from Tn903 was used as a template. This amplified PCR product provided resistance to the aminoglycosid antibiotic G418 upon recombination into the yeast genome [Janke et al., 2004].

Table 4.30. Epitope tagged proteins in full-length comprising the SIM sequence from the bioinformatical screen used for GST pulldown assays. pYM14 served as template for the PCR amplification of a cassette comprising 6×HA tag and kanMX6 selectable marker [Longtine et al., 1998, Knop et al., 1999, Janke et al., 2004].

protein	size including tag	SUMO-SIM interaction?
Sec27-6HA	110 kDa	☐
Rad5-6HA	143 kDa	☐
Rrb1-6HA	66 kDa	☐
Dbp10-6HA	122 kDa	☒
Drs1-6HA	93 kDa	☒
Rfc1-6HA	104 kDa	☒
Rad18-6HA	64 kDa	☒
Nis1-6HA	52 kDa	☒

SUMO was expressed as GST protein fusion in *Escherichia coli*. An *E. coli* strain expressing GST alone was taken as a negative control. For the GST pulldown assay, *E. coli* strains expressing GST and GST-SUMO and the glass bead lysis protocol were used (section 3.2.13). GST and GST-SUMO were immobilized onto glutathion sepharose beads by applying the respective crude extracts to the glutathion sepharose matrix. Yeast strains expressing the 6×HA tagged proteins were incubated overnight at 30°C, harvested by centrifugation and lysed. Western blots indicate the HA signal intensity as a measure for protein expression levels (figure 4.9): Lanes correspond to different

crude extracts from lysis using the glass bead lysis (green) or the direct boiling method (red). All labels in the figures omit the 6×HA tag for simplicity. Using 3F10 as α-HA antibody, the HA signal intensity can be taken as an indicator for protein expression in *Saccharomyces cerevisiae*. All 6×HA tagged proteins are expressed but with very different levels. The glass bead lysis method gives more specific bands for the respective protein and was therefore used in the GST pulldown technique. Direct boiling of the *Saccharomyces cerevisiae* cell pellets leads to 'smearing' of the expected α-HA signal.

(a) (b)
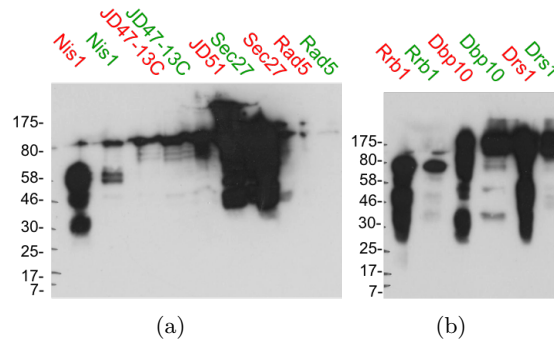
Figure 4.9. Glass bead lysis gives crude extracts with more specific α-HA signal in western blots than from the direct boiling protocol. α-HA western blots after glass bead lysis (green, soluble fraction) and boiled crude cell pellets (red) of the respective *Saccharomyces cerevisiae* JD47-13C strains.

4.6.2.1. Comparison of various lysis buffer compositions for best α-HA signal depletion in GST-SUMO pulldown assays

Different lysis buffer compositions are reported among others [Uzunova, 2006, Miteva, 2007, Kurian, 2009], as protein expression, folding and stability can be influenced by solvent conditions. Inter- and intramolecular salt-bridges in proteins are mainly affected when shifted to a solvent with different effective polarity. To find the best suitable lysis buffer composition for the GST-SUMO pulldown approach, different amounts of sodium chloride were used in a preliminary test with and without use of triton-X as detergent. Four different buffer compositions were applied to determine the optimal conditions for GST pulldown assay performance of Fir1-6HA and Nis1-6HA as established SIP, hereafter referred to as "Fir1" and "Nis1" (table 4.10). The GST pulldown assay was performed with GST and GST-SUMO crude extracts (section 3.2.11). Wild-type *Saccharomyces cerevisiae* strain and strains expressing "Fir1" and "Nis1" were used in the GST pulldown assay as bait for the immobilized GST fusion proteins. The *Saccharomyces cerevisiae* strains were incubated and lysed (section 3.2.13).

Table 4.10. Different lysis buffer compositions for best lysis performance in western blots.
The different conditions vary in their sodium chloride content ("low salt" and "high salt") and whether
Triton X-100 was used. All buffers were supplemented with protease inhibitor and used at 4°C.

NaCl	150 mM	50 mM	150 mM	50 mM
EDTA	5 mM	5 mM	5 mM	5 mM
HEPES	50 mM	50 mM	50 mM	50 mM
Triton	-	-	1 %	1 %

Samples were taken from these crude extracts before (input) and after applying to the
immobilized GST-SUMO fusion protein (flow-through, FT). The α-HA signal differences
between these samples were checked in α-HA western blots (figure 4.11). The α-HA
signal intensities were evaluated as FT/input ratio providing a measure for the efficiency
of the GST pulldown.
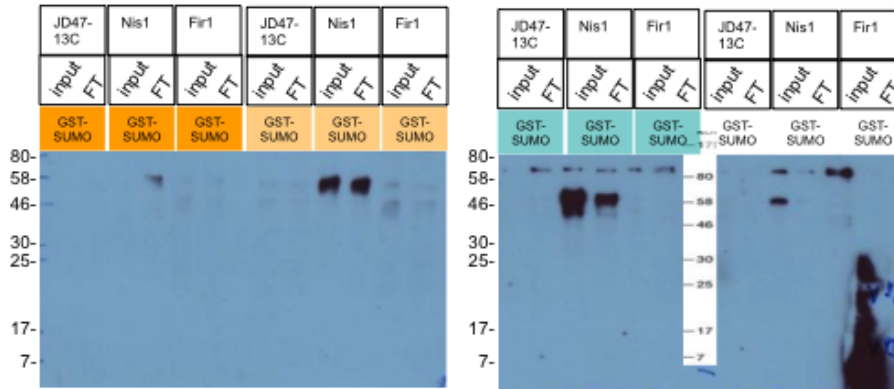
**Figure 4.11. Nis1 crude extract from glass bead lysis is a good positive control. Best
α-HA signal depletion is observed for buffer "green".** Wild-type yeast strain JD47-13C, Fir1-
6HA ("Fir1") and Nis1-6HA ("Nis1") were used in this GST pulldown assay. Different lysis buffer
compositions were applied (table 4.10). Shown here are the different levels of α-HA signal with crude
extracts before (input) and after applying (flow-through) to GST-SUMO immobilized on a sepharose
matrix.

α-HA signal depletion with GST-SUMO and $6\times$HA tagged Nis1 is best under "high salt"
lysing conditions and when triton is applied. Wild-type JD47-13C strain expressing no
HA-tagged protein gave a slight non-specific signal at high molecular weight region in
western blots and 3F10 antibody (figure 4.11).

4.6.2.2. GST-SUMO pulldown depletes α-HA signals of Dbp10 and Drs1

Western blots of Dbp10-6HA ("Dbp10") and Drs1-6HA ("Drs1") after GST pulldown show α-HA signal depletion upon SUMO interaction (figure 4.12). Depletion of HA signal intensity was defined as the difference between GST and GST-SUMO flow-through fractions. That depletion is taken as a measure for the amount of recombinant protein removed upon binding to SUMO (Smt3). Ponceau colorization and α-Cdc11 immunoblotting verified that similar amounts of extracts proteins were applied. DBP10 and DRS1 constructs show no interaction with SMT3 constructs in the two-hybrid assay, whereas α-HA signal depletion for these constructs in GST-SUMO pulldown assays is shown with shorter X-ray film exposure or with crude extract dilutions applied (figure 4.12): Elution of the GST- and GST-SUMO-coated glutathion sepharose beads with 10 mM Glutathion/lysis buffer give differences in α-HA signal intensities between the GST and GST-SUMO lanes. The blots for Dbp10 and Drs1, therefore, indicate weak interactions with SUMO.
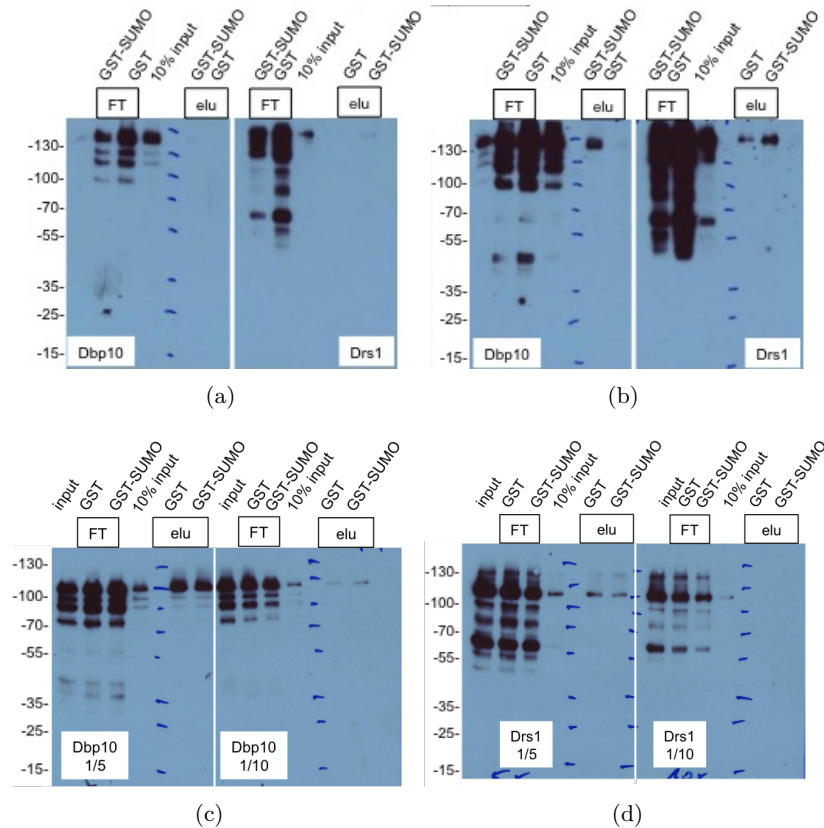
Figure 4.12. Dbp10 and Drs1 show α-HA signal depletion upon SUMO interaction. Dbp10 and Drs1 in their 6HA tagged versions were used in GST pulldown assays. Western blots were taken in two different x-ray film exposures (a/c, b/d) and different dilutions (a/b, c/d). α-HA signal depletion is regarded as a measure for recombinant protein removed from the applied crude extract upon SUMO binding.

97

4.6.2.3. GST-SUMO pulldown shows SUMO interaction for Rfc1 and Rad18

Western blots for the GST-SUMO pulldown assays applied to Rfc1-6HA ("Rfc1"), Rad18-6HA ("Rad18") and Nis1-6HA ("Nis1") show weak α-HA signal depletion from the flow-through (FT) samples (figure 4.13). Elution with 10 mM glutathion after washing returned more HA-signal for Dbp10, Drs1 and Rad18 from beads coated with GST-SUMO than from those with GST alone. Repeated GST-SUMO pulldown assays for Rfc1 and Rad18 did not give reproducible results. Even repetitions of the GST-SUMO pulldown assay with lysis buffer with different sodium chloride contents (25 mM, 50 mM, 100 mM and 150 mM) did not provide a clear α-HA signal depletion for Rad18 or Rfc1. Nis1 was used as a performance control in GST pulldown assays with the same buffer and GST/GST-SUMO exposures as Dbp10, Drs1, Rfc1 and Rad18 (figure 4.13(d)): Nis1 signal depletion is observed between the input lane and either of the FT lanes and between the FT lanes of GST and GST-SUMO.
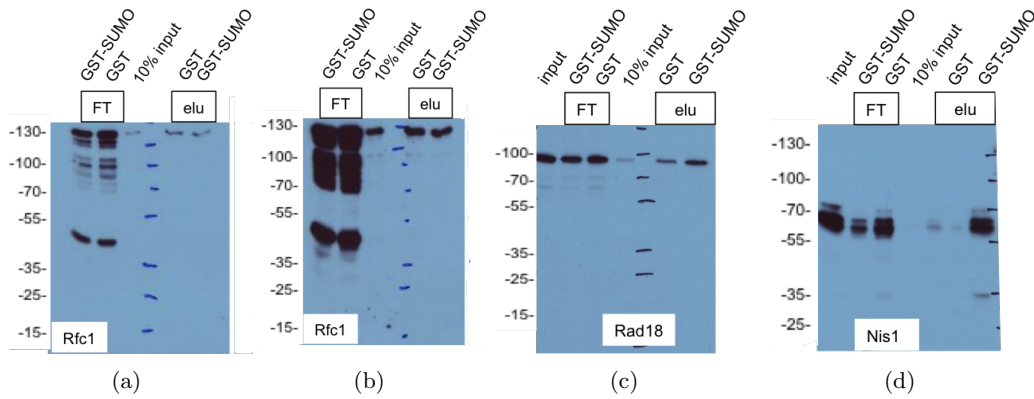
Figure 4.13. α-HA immumoblotting shows SUMO interaction for Rfc1 from α-HA signal depletion and for Rad18 from elution. GST-SUMO pulldown results for Rfc1 (in two different X-ray exposures 4.13(a), 4.13(a)), Rad18 (4.13(c)) and Nis1 as GST pulldown performance control (4.13(d)) show different levels of SUMO interaction.

The GST pulldown assay worked well under the used experimental conditions for the established SIP Nis1, as the conditions were optimized for Nis1/SUMO interaction (figures 4.11, 4.13(d)). For Rfc1 and Rad18 with SUMO interaction from the yeast two-hybrid results, however, the GST-SUMO pulldown assay did not give reproducible results. No results supporting SUMO interaction for Sec27-6HA, Rad5-6HA and Rrb1-6HA with SUMO were obtained in the GST-pulldown assays.

4.6.2.4. Strategy for SIMa mutation in Rfc1

Rfc1/SUMO interaction was shown in yeast two-hybrid experiments (section 4.6.1). For a better understanding of the biological relevance of the SIMa in Rfc1, mutation studies were performed. Protein sequence comparison within a 'narrow' phylogenetic range showed three putative SIMs in close distance to each other in the amino terminal region (figure 4.14): A highly conserved sequence MVNISD (green bar, 'SIM1'), the highly conserved VIDLDTE motif from the bioinformatical screen (red bar, 'SIM2') and a poorly conserved VIDVSE (blue bar, 'SIM3'). SIM1 and SIM3 were not included in the bioinformatical results, whereas they follow the SIMa consensus pattern. SIM1 shows good absolute and relative conservation but is not so close to the SIMa pattern as is SIM3. However, they may show supporting SUMO binding function for SIM2.
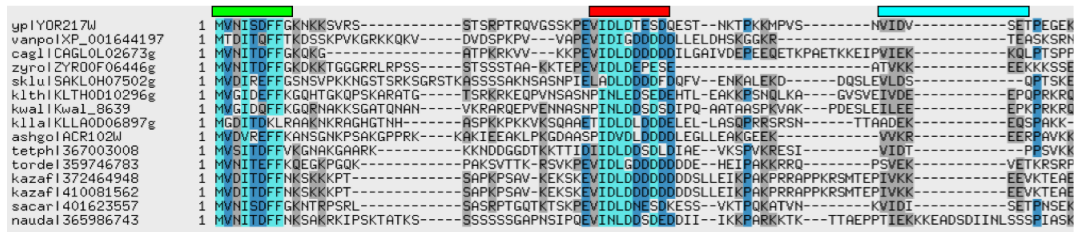
Figure 4.14. Three SIM-like sequences in the Rfc1 amino terminal region from the bioinformatical SIMa and SIMb detection screens. Carboxy terminal extract from a 'narrow' MSA for Rfc1 (YOR217W). A highly conserved SIM-like sequence at the very amino terminus (green bar), the well-conserved putative SIMa/b from the bioinformatical detection screens (red bar) and a poorly conserved SIMa-like sequence VIDVSE (blue bar) are indicated.

For the identification of SIM functionality for Rfc1, amino acid mutations were introduced at the interface between hydrophobic and acidic stretch (figure 4.15). For yeast two-hybrid experiments, these residue mutations were introduced by site-directed mutagenesis into the coding sequence of the previously used GAD-Rfc1-HA plasmid (table 4.29 and section 3.2.10). The three putative SIM1, SIM2 and SIM3 are referred to in the order of sequence occurrence from the protein's amino terminus. All seven possible SIM mutant permutations were used: ΔSIM1, ΔSIM2, ΔSIM3, ΔSIM1ΔSIM2, ΔSIM2ΔSIM3, ΔSIM1ΔSIM3, ΔSIM1ΔSIM2ΔSIM3. The Gal4 activating domain (GAD) was expressed alone as a negative control (GAD-C1) or as N-terminal fusions to Rfc1 (positive control), Rfc1ΔSIM1, Rfc1ΔSIM2, Rfc1ΔSIM3, Rfc1ΔSIM1ΔSIM2, Rfc1ΔSIM1ΔSIM3, Rfc1ΔSIM2ΔSIM3 and Rfc1ΔSIM1ΔSIM2ΔSIM3. The Gal4 binding domain (GBD) was expressed alone as a negative control (GBD-C1) or as fusions to Smt3, Smt3Δ97, Smt3×3 and Smt4×4.

```
ΔSIM1: MVNISD → MVNAAD
original: ACT AAG CTG AAG AAA ATG GTC AAT ATT TCT GAT TTC TTT GGT AAA AAT
           T   K   L   K   K  [M   V   N   I   S   D]  F   F   G   K   N
mutated:ACT AAG CTG AAG AAA ATG GTC AAT GCT GCT GAT TTC TTT GGT AAA AAT
           T   K   L   K   K  [M   V   N   A   A   D]  F   F   G   K   N


ΔSIM2: VIDLDT → VIDAA
original: TCG TCT AAA CCA GAA GTT ATC GAC TTA GAT ACT GAA TCT GAC CAA GAA
           S   S   K   P   E  [V   I   D   L   D   T]  E   S   D   Q   E
mutated:ATG CCT GTA AGT AAT GTA ATT GAT GCA GCA GAG ACA CCT GAA GGA GAA
           S   S   K   P   E  [V   I   D   A   A   T]  E   S   D   Q   E


ΔSIM3: VIDVISE → VIDAAE
original: ATG CCT GTA AGT AAT GTA ATT GAT GTA TCA GAG ACA CCT GAA GGA GAA
           M   P   V   S   N  [V   I   D   V   S   E]  T   P   E   G   E
mutated:ATG CCT GTA AGT AAT GTA ATT GAT GCA GCA GAG ACA CCT GAA GGA GAA
           M   P   V   S   N  [V   I   D   A   A   E]  T   P   E   G   E
```

Figure 4.15. Nucleotide exchanges for selective mutation of the Rfc1 SIM-like sequences in *Saccharomyces cerevisiae*.

All final carboxy termini of Smt3 were ΔG97 truncated. *Saccharomyces cerevisiae* strains AH109 and pJ69-4A were co-transformed with plasmids expressing a GAD and a GBD fusion protein using the lithium acetate method (section 3.2.6). Equal volumes from these cell dilutions were spotted onto synthetic drop-out (SD) media plates lacking leucine and tryptophane (SD-LW). SD-LW plates additionally lacking histidine (SD-LWH), or SD-LWH plates with 5 mM 3-AT (SD-LWH+3AT) and SD-LW plates lacking adenine (SD-LWA) were used.

The occurrence of growth phenotype is regarded as a measure for SUMO interaction. Cell growth phenotype on SD-LW plates for all strains used served as control (figure 4.16). Cells were present in the spotted suspensions at comparable concentrations. Growth on SD-LWH plates is expected from the previous yeast two-hybrid experiments (figure 4.8). Additionally, growth is observed in strains expressing any GBD-Smt3 fusion. Growth is less suppressed for two transformants expressing GBD-Smt3 and GBD-Smt3Δ97 on SD-LWH+3AT plates than expected from previous experiments. The results from strains expressing GAD fusions of Rfc1ΔSIM1, Rfc1ΔSIM3 and Rfc1ΔSIM1ΔSIM3 are comparable with previous two-hybrid findings for GAD-Rfc1-HA: Those strains show growth where GBD fusions of Smt3 and Smt3×3 are co-expressed with the respective GAD-Rfc1

mutant. All other strains show no growth on plates other than SD-LW.

In conclusion, these two hybrid assays show that SIM2, which was initially identified in the bioinformatical SIMa and SIMb detection screens, represents a functional SIM for SUMO interaction, whereas the other two sequences tested do not contribute detectably to SUMO binding.
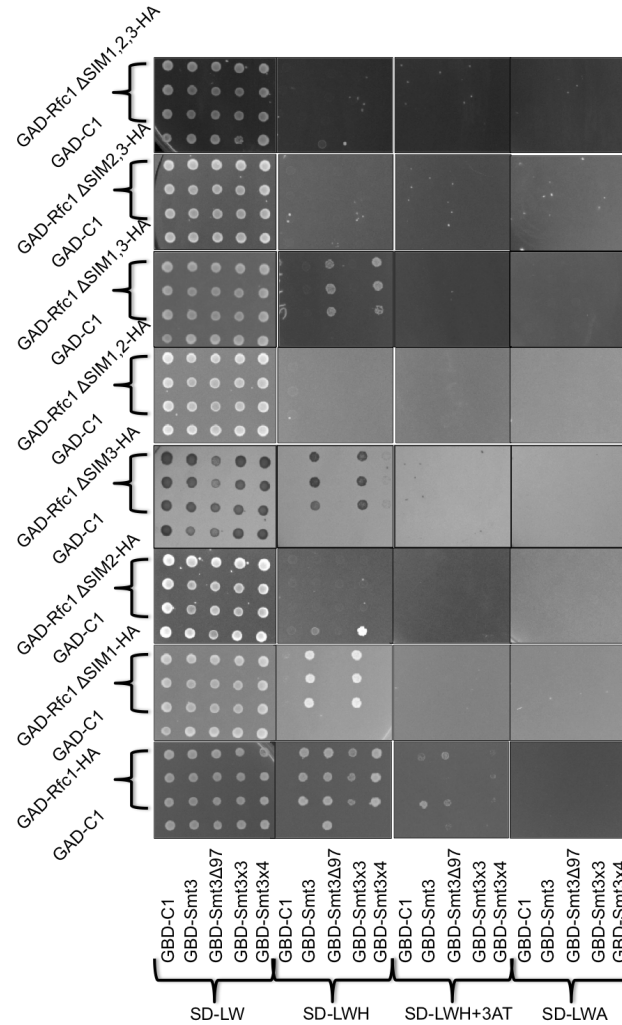
Figure 4.16. "SIM2" `VIDLDT` in Rfc1 from the bioinformatical SIMa and SIMb screens is responsible for SUMO interaction. Unmutated and mutated versions of GAD-Rfc1-HA were used together with different GBD constructs expressing GBD-C1 and fusions of Smt3, Smt3Δ97, Smt3×3 and Smt3×4. SD-LW plates served as internal transformant growth control.

4.6.3. Rfc1 SIM mutation may induce growth phenotype upon DNA damage

SUMO interaction has not yet been reported for Rfc1. With the SIM in Elg1 and the Elg1-RFC complex involved in SUMOylated PCNA binding, a functional SIM in Rfc1 may be relevant for RFC clamp loading activity [Parnas et al., 2010]. DNA replication and DNA damage repair may be negatively affected by an unfunctional SIM in Rfc1. Growth phenotype analysis of an *rfc1* SIM mutant strain is performed upon temperature stress and exposure to chemically induced DNA damage. Methyl methanesulfonate (MMS), Ebselen (2-phenyl-1,2-benzisoselenazol-3(2*H*)-one) and hydroxy urea (HU) are reported as such chemical agents [Azad et al., 2012].

4.6.4. Verification of Rfc1 SIM/Smt3 interaction

MMS is a reported alkylating agent [Prakash and Prakash, 1977, Fortini et al., 2000, Lundin, 2005]. It alkylates deoxyribonucleotide phosphates (dNTP). MMS modifies guanine and adenine to 7-methyl guanine and 3-methyl adenine, respectively [Beranek, 1990]. These DNA damages are repaired by the base excision repair pathway (BER) and DNA alkyltransferase [Lindahl, 1990, Jiricny, 2006]. Cells with dysfunctional homologous recombination pathway (HR) were found to be sensitive to MMS treatment [Krogh and Symington, 2004]. MMS causes double-strand breaks (DSBs). DSBs are usually corrected either by non-homologous end joining (NHEJ) or the HR pathway [Weterings and Chen, 2008, West, 2003, Li and Heyer, 2008, Caldecott, 2008]. HU is an inhibitor to ribonucleotide reductase (RNR) and inhibits DNA replication by stalling the replication fork [Krakoff et al., 1968, Slater, 1973]. The mechanisms leading to this replication stall remain unclear.

The underlying mechanisms how Ebselen exposure to yeast cells lead to DNA double-strand breaks are still unknown. It was found that Ebselen leads to RNR inhibition like HU treatment. MMS, HU and Ebselen are used in growth phenotype analysis when effects on DNA damage repair are expected [Azad et al., 2012].

This might be the case for a dysfunctional SIM in the *rfc1* mutant strain, if the SIM plays a crucial role in DNA damage repair pathways.

4.6.4.1. The SIMa mutation is integrated into RFC1 via two-step gene replacement

Gene replacement was performed with the two-step gene replacement method (section 3.2.18) [Sikorski and Hieter, 1989, Widlund and Davis, 2005]. A nucleotide sequence 1 kB upstream of the RFC1 gene and 705 nucleotides of its coding sequence was amplified by

PCR. Primers (table 2.4) were used to comprise a mutant version of SIMa ('SIM2', red), which was shown earlier in this thesis to be responsible for Rfc1/SUMO interaction in the two-hybrid assay. A silent SalI restriction site (blue) is introduced close to the SIMa mutation site to ease identification of correct clones.

VIDLTE → VIDAATE and insertion of a SalI restriction site $\begin{pmatrix} 5' \cdots \text{G}|\text{TC}\text{G}\text{A C} \cdots 3' \\ 3' \cdots \text{C AGCT}|\text{G} \cdots 5' \end{pmatrix}$

```
original: CCA GAA GTT ATC GAC TTA GAT ACT GAA TCT GAC CAA GAA TCA ACC AAT
           P   E   V   I   D   L   D   T   E   S   D   Q   E   S   T   N
mutated:  CCA GAA GTT ATC GAC GCA GCT ACT GAA TCT GAC CAA GAG TCG ACC AAT
           P   E   V   I   D   A   A   T   E   S   D   Q   E   S   T   N
```
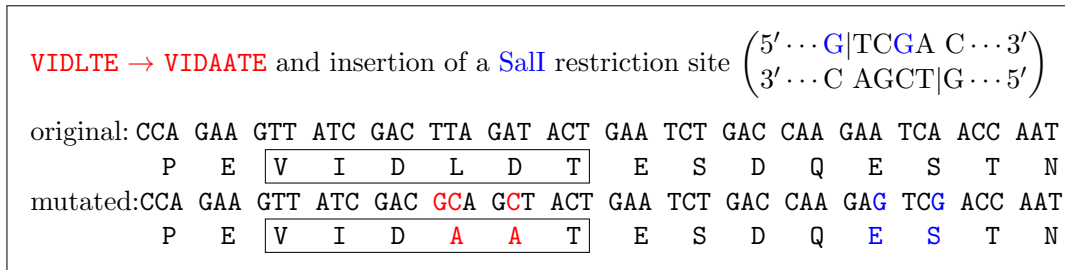
Figure 4.17. Strategy for a SIMa mutagenesis for RFC1 from the bioinformatical screens. The motif is changed from VIDLDT to VIDAAT. A SIM mutation VIDAAT, instead, lost that functionality in yeast two-hybrid experiments. Three nucleotides (red) were exchanged for a double-alanine translation. Silent mutagenesis introduced a unique SalI restriction site (blue) to allow mutant allele screening from many transformants by analytical PCR followed by SalI restriction (figure 4.20).

The amplicon was introduced into a pBluescript-based integrative vector pRS306 with *URA3* selectable marker (figures 4.18, 4.19). The multi-cloning site (MCS) is embedded in a sequence encoding a segment (alpha peptide) of the lacZ protein. The MCS is indicated in figure 4.18 and figure 4.19 by KpnI and NotI sites. They were used for the cloning step. Amplicon restriction using the unique SalI restriction site introduced by silent mutagenesis and Sanger sequencing confirmed the plasmid sequence.
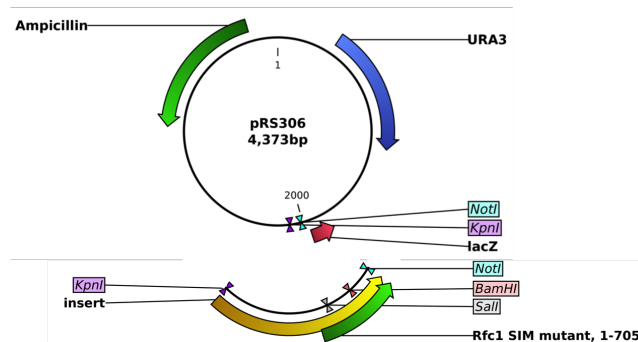
Figure 4.18. First step in the two-step gene replacement cloning strategy. Simplified pRS306 vector map (adapted from [Sikorski and Hieter, 1989]). The shuffle vector comprises ampicillin (green) and *URA3* (blue) selectable markers. A lacZ gene (red) is preceded by a multi-cloning site indicated by KpnI (violett) and NotI (turquoise) restriction sites. The insert depicted below (gold) was classically cloned and comprises 1 kb upstream sequence and 705 bases of the Rfc1 coding sequence (green). Compare figure 4.19 for further information.
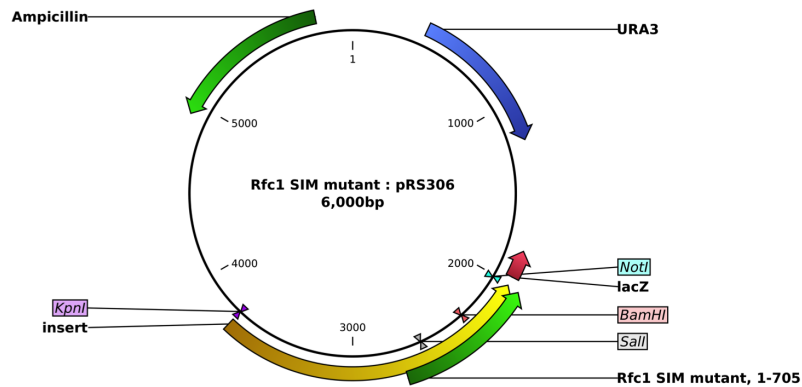
Figure 4.19. Vector map for the integratable shuffle vector *rfc1*:pRS306. The insert is cloned into a sequence encoding the so-called alpha-peptide of lacZ thereby preventing alpha-peptide expression and thus the generation of functional lacZ by alpha-complementation. This allows visual selection of integrants using the blue/white technique of transformed *Escherichia coli* cells on LB plates supplemented with ampicillin and X-Gal. A unique BamHI restriction site is introduced by the vector for linearization and a unique SalI restriction site for analytical PCR and restriction analysis of integrants.

Linearization of the vector using the unique BamHI site within the *RFC1* coding region directs integration at the endogenous *RFC1* gene locus upon transformation of wild-type yeast cells. Correct integration of the plasmid by homologous recombination results in a tandem array of a 5′ truncated wild-type allele of *RFC1* and a full-length mutant allele of *RFC1* with promoter flanked by the selectable *URA3* marker. Selection on SD plates lacking uracil, followed by counter-selection on SD plates with 0.1% 5-fluoro-orotic acid gives the desired strain with the wild-type *RFC1* gene replaced by the mutant allele (section 3.2.18) [Cregg and Russell, 1998, Widlund and Davis, 2005]. Replacement was checked by amplification and successive SalI restriction (figure 4.20, table 2.4). This was confirmed by Sanger sequencing.

The forward primer binds about 1 kB upstream the *RFC1* gene and the reverse primer to a nucleotide sequence downstream of the integrated allele. The SalI restriction site was introduced right after the mutagenized SIMa (figure 4.17). Amplification of that 2.8 kB nucleotide sequence with subsequent restriction with SalI verified the correct *rfc1* SIMa mutant strain (figure 4.20).
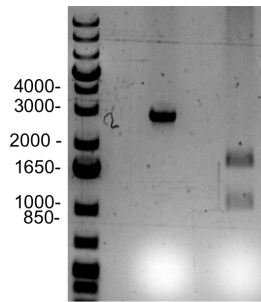
Figure 4.20. Amplification and subsequent SalI restriction helped identify a correct SIM mutant integrant. Amplification of a nucleotide sequence comprising 1 kb upstream the *RFC1* gene and about 1.8 kb of its coding sequences is performed with integrants as template. Primers are designed to bind in the front part of the integration and in the 'wild-type' coding sequence behind the integrated nucleotide sequence. An amplicon of 2.8 kB size is expected. SalI restriction verifies the presence of the introduced mutations by producing 1.7 kB and 1.1 kB fragments. This was confirmed by Sanger sequencing.

4.6.5. Relevance of Rfc1 SIMa/SUMO interaction

A biological relevance for a SIMa/SUMO interaction in *RFC1* can be investigated in thermal or chemical stress tests. The occurrence of growth phenotype for a *rfc1* SIMa mutant JD47-13C strain is regarded as a measure for dysfunctional or at least hindered stress response. Lack of growth, on the other hand, is no means for lack of interaction relevance. Stress of the mutant strain may also be compensated by other pathways.

4.6.5.1. No growth phenotype upon thermal stress

A mutant *Saccharomyces cerevisiae* strain with dysfunctional SIMa in the *RFC1* gene was constructed. Growth phenotype analyses are performed with different aqueous cell suspensions of the mutant strains as described by Azad *et al.* [Azad et al., 2012]. Growth characteristics of several independent isolates of the *rfc1* mutant strain were analyzed in comparison to the congenic wild-type strain (figures 4.21, 4.22, 4.23, 4.24). A preliminary growth test on YPD plates showed no difference whether cells were incubated at 30°C or 37°C. Plates were checked after 24h, 48h, 72h and 96h of incubation (figure 4.21). Thermal stress alone does not induce growth defects in the *rfc1* SIM mutant strain. Overall growth in the different strains is reduced by incubation at 37°C instead of 30°C.
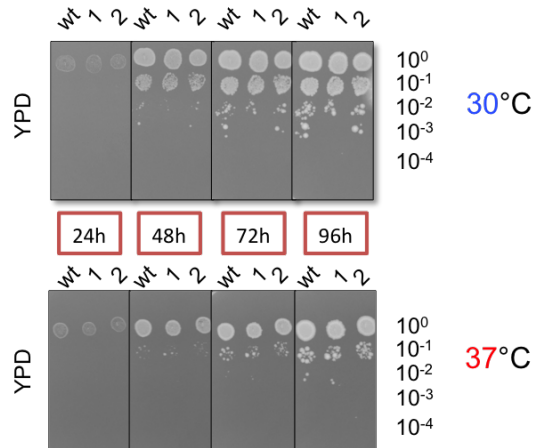
Figure 4.21. No observable growth phenotype of *rfc1* SIM mutant upon thermal stress.
Wild-type (wt) and SIM mutant strains (1, 2) were spotted onto YPD plates undiluted (10^0) and in
10^{-1}, 10^{-2} and 10^{-3} dilutions and incubated at 30°C (top row) and 37°C (bottom row). Shown here are
the respective plates after 24h, 48h, 72h and 96h. The left lanes correspond to an wild-type JD47-13C
strain and lanes 1 and 2 to independently grown mutant strains, all spotted in dilutions from the top
down.

4.6.5.2. No observable growth phenotype under chemical stress: methyl methanesulfonate

MMS was freshly dissolved in dimethyl sulfoxide (DMSO) and added to YPD plates at
a final concentration of 0.01%. Wild-type and *rfc1* SIM mutant strains were used in
the same dilutions as for the previous thermal stress test (section 4.6.5.1, figure 4.21).
Growth phenotype was checked after 24h, 48h, 72h and 96h incubation at 30°C and 37°C,
respectively (figure 4.22). No growth phenotype can be observed as difference between
wild-type and mutant strains (figure 4.22).
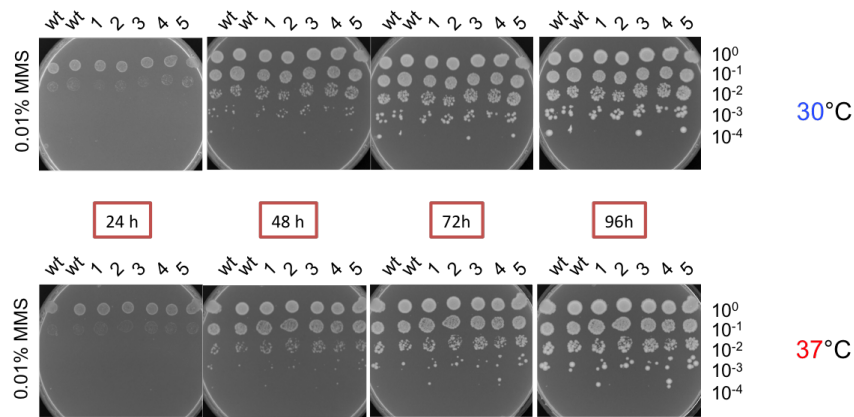
Figure 4.22. No observable growth phenotype of *rfc1* SIM mutant upon MMS exposure.
Wild-type (wt) and SIM mutant strains (1, 2) were spotted onto YPD plates supplemented with 0.01%
MMS in 10^0, 10^{-1}, 10^{-2} and 10^{-3} dilutions and incubated for 4 days at 30°C (top part) and 37°C (bottom
part).

4.6.5.3. Observable growth phenotype under chemical stress: hydroxy urea

Cells in dilutions of 10^0, 10^{-1}, 10^{-2} and 10^{-3} from the same culture were spotted onto
YPD plates supplemented to different final concentrations of hydroxy urea (HU) (10 mM,
40 mM, 70 mM and 100 mM) [Azad et al., 2012]. YPD/HU plates were incubated at
30°C and 37°C. Growth phenotype was checked after 24h, 48h, 72h and 96h (figure 4.23).
Strong growth phenotype can be observed on plates incubated at 37°C, especially with
40 mM and 70 mM HU concentration. On those plates, the wild-type strain shows in-
creased growth compared to the mutant strains. Similar growth phenotype is observable
on plates incubated at 30°C at longer incubation and with increased HU concentrations.
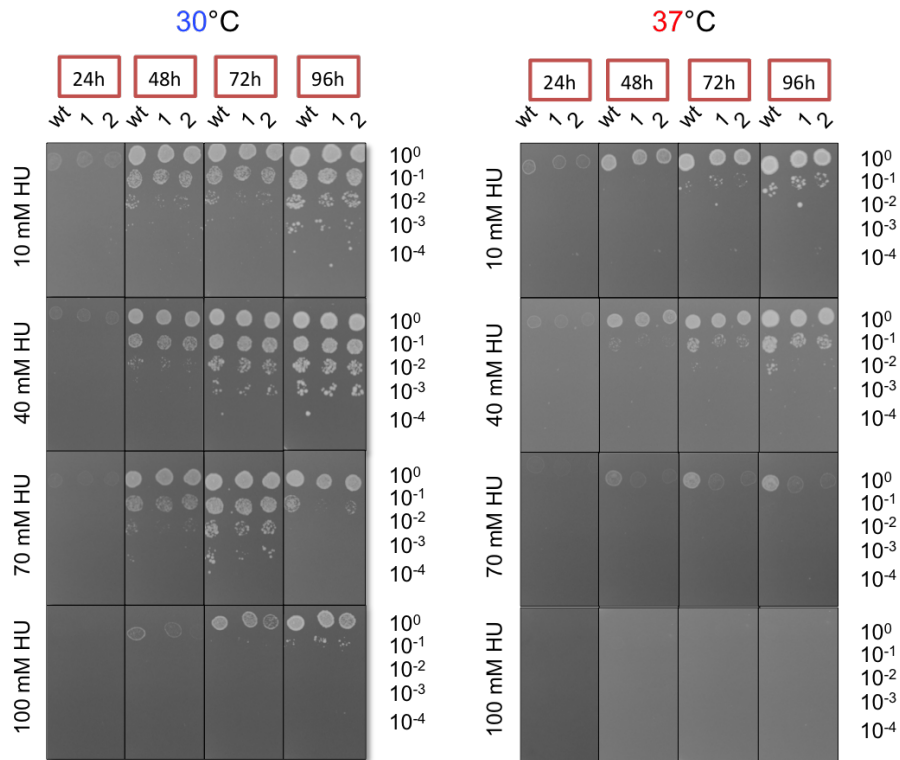
Figure 4.23. Observable growth phenotype of *rfc1* SIM mutant strain upon exposure to hydroxy urea. Wild-type (wt) and Rfc1 SIM mutant strains (1, 2) were spotted in 10^0, 10^{-1}, 10^{-2} and 10^{-3} dilutions. YPD plates supplemented with final concentrations of hydroxy urea (HU) of 10 mM, 40 mM, 70 mM and 100 mM were used. Incubation at 30°C (left part) and 37°C (right part) was recorded after 24h, 48h, 72h and 96h. The plates are arranged with increasing HU concentration from top down and with increasing incubation time from left to right. Growth phenotype is shown upon HU exposure and increased temperatures, but is already visible at 30°C with increased HU concentrations.

4.6.5.4. Observable growth phenotype under chemical stress: Ebselen

Different Ebselen (2-phenyl-1,2-benzisoselenazol-3(2H)-one) concentrations of 10 μM, 40 μM, 70 μM and 100 μM were supplemented to YPD plates. Cells of the wild-type and *rfc1* SIM mutant strain were spotted in dilutions of 10^0, 10^{-1}, 10^{-2} and 10^{-3} onto these YPD/Ebselen plates and incubated at 30°C and 37°C, respectively (figure 4.24). At 37°C, all strains spotted show decreased growth. The decrease is stronger with increasing Ebselen concentrations or increased temperature. It can be concluded from these experiments, that the *rfc1* SIM mutant exhibits a mild hypersensitivity to several genotoxic agents.
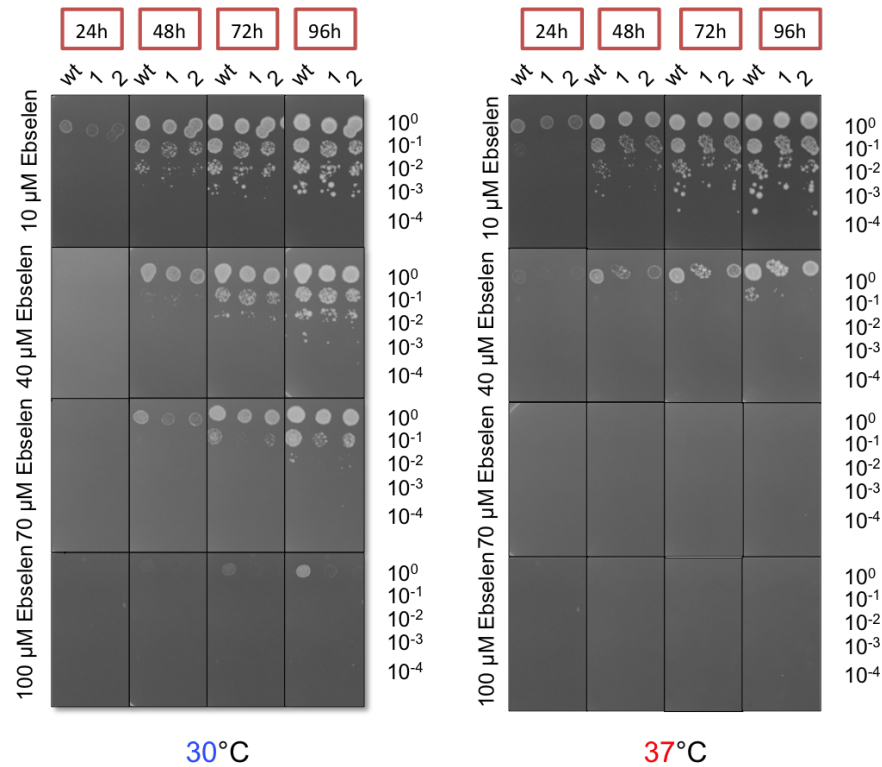
Figure 4.24. Observable growth phenotype of *rfc1* SIM mutant strain upon Ebselen exposure. Wild-type (wt) and SIM mutant strains (1, 2) were spotted in 10^0, 10^{-1}, 10^{-2} and 10^{-3} dilutions onto YPD plates supplemented with different concentrations of Ebselen (10 μM, 40 μM, 70 μM and 100 μM) and incubated for 4 days at 30°C (left part) and 37°C (right part). Growth phenotype is obvious at 37°C already with 10 μM and 40 μM Ebselen as growth difference between wild type and mutant strains.

5. Discussion

5.1. General aspects of bioinformatical motif discovery

Many regulatory processes are mediated by short linear motifs (SLiMs) which are usually found in unstructured linker regions. These highly conserved instances form short, mainly hydrophobic sequences of 3–10 amino acids which adopt a defined three-dimensional conformation upon interaction with their prefolded partner. Partner binding affinity is reduced compared to the binding between prefolded structures. SLiMs, therefore, allow regulation of biological processes. Such regulatory processes are involved in pathways connected to the ubiquitin proteasome system (UPS), the small ubiquitin-like modifier (SUMO) or Atg8. Mechanistical and structural similarities in conjugation cascades and recognition show the importance of these pathways for cellular survival. Whereas ubiquitin recognition involves different classes of prefolded structures, decoding of SUMOylated proteins depends on the noncovalent interaction to a SUMO interacting motif (SIM). The three different SIM types (SIMa, SIMb, SIMr) may differ in their consensus patterns, but also show common characteristics. These characteristics are used for a bioinformatical SIM detection approach described in this thesis.

AIMs show similar characteristics to SUMO interacting motifs. These short stretches are highly conserved and are located in unstructured, well solvent-accessible protein regions. One property of short motifs that complicates their bioinformatical detection is the narrow phylogenetic range of conservation. In more distantly related species, both SIMs and AIMs often disappear and/or reappear at a different position.

The basic assumptions for a bioinformatical detection approach are the same for both SIMs and AIMs. The number of established motifs is less for AIMs than for SIMs. This and the short sequence of AIMs reduces the sequence information in form of a consensus profile or pattern. The bioinformatical approach to detecting AIMs might therefore be less effective in enriching true-positive motifs than is the SIM approach. The descriptive part deals with the SIM detection approach. The same assumptions and methods were also applied in the AIM detection approach.

Pattern-based screens in sequence databases with only a consensus pattern return large numbers of false-positive sequences. Discrimination of functional motifs from spurious pattern matches is performed through additional criteria regarding their conservation

and structural context. Biological function is under strong evolutionary constraint. Therefore, the short SIM sequences are usually passed on to succeeding generations more or less as a whole. For that reason, most functional SIMs are highly conserved over several phylogenetically related species. Disordered regions in general lack the constraints associated with the conservation of folded structures. Thus, functional residues in SIM sequences are more conserved than their surrounding. Whereas folded structures show high contents of hydrophobic residues, acidic residues are enriched in unstructured regions. This is reflected in the SIM and AIM consensus sequences in which the conserved hydrophobic part is preceded or followed by acidic residues. Single mutations in the coding sequences allow SIMs to vanish, to emerge at a different position or to change its consensus type.

Whereas bioinformatics may be applied to a wide range of functional SIMs, these 'atypical' cases of poorly conserved, but still functional SIMs may not be identified by the bioinformatical approach described in this thesis. This approach will also neglect those SIMs, that are readily solvent-accessible despite being within or close to folded structures. Vice versa, this approach is not exclusive to false-positive motif-like sequences.

5.2. Experimental findings from the bioinformatical results

For the experimental interaction studies, it was important to not just use the SIM-containing peptide. One important criterion for the biological relevance of a putative SIM is its accessibility for SUMO interaction. For an isolated peptide, this condition would always be fulfilled - even if the SIM-like motif is biologically meaningless because it is found within the buried core of a folded domain. On the other hand, many of the SIM-containing proteins are unusually big and difficult to handle and express under experimental conditions. For that reason, an intermediate approach was used: Short proteins were used in their entirety, while big proteins were shortened by using truncations at defined domain boundaries. In all cases, the SIM-containing part of the protein was used within its usual sequence context. In the biologically interesting case of Rfc1, it was also tested if the observed SUMO binding by the protein was really due to the motif identified in the bioinformatical screen. A mutation within the predicted SIM region abrogated the SUMO interaction.

Putative SIM sequences in Rad5, Rrb1, Dbp10, Drs1, Rfc1 (all SIMa), Slx4, Top1, Tdp1 (all SIMb) and Rad18 (SIMr) were finally selected for experimental interaction studies. Yeast two-hybrid experiments show SUMO interaction for Tdp1, Rfc1 and Rad18. SUMO/SIM interaction could be confirmed for Rfc1 with SIM mutants in yeast

two-hybrid experiments. GST-SUMO pulldown experiments further suggested SUMO interaction for Dbp10 and Drs1. Rfc1/SUMO and Rad18/SUMO interaction were also shown in these experiments.

SUMO/SIM interaction of Rfc1 was investigated in phenotype analysis. A *rfc1* SIM mutant strain was exposed to thermal and thermal/chemical stress. The chemical compounds methyl methanesulfonate, hydroxy urea and Ebselen were used for their DNA damaging effects [Azad et al., 2012]. Hydroxy urea and Ebselen caused observable growth defects (sections 4.6.5.3, 4.6.5.4).

Rad18/SUMO interaction was meanwhile shown by Parker and Ulrich for the same SIMr as from the bioinformatical SIM detection screen [Parker and Ulrich, 2012].

5.3. Discussion of exemplary results

5.3.1. Dbp10/SUMO and Drs1/SUMO interaction

5.3.1.1. Biological context of Dbp10 and Drs1

The *Saccharomyces cerevisiae* ribosome complex is composed of two subunits, 40S and 60S [Mager et al., 1997]. These subunits in turn consist of different ribosomal RNAs (rRNA) and ribosomal proteins (RP) . Ribosomal biogenesis is a multi-step process. In *Saccharomyces cerevisiae*, there are 19 RNA helicases among which Dbp2, Dbp3, Dbp6, Dbp7, Dbp9, Dbp10 [Burger et al., 2000], Drs1 [Ripmaster et al., 1992, Ripmaster et al., 1993], Mak5, Mtr4 and Spb4 are involved in maturation of the pre-60S ribosomal particle in the nucleolus. They are recruited in a hierarchical manner to the pre-ribosomal particle at different stages of the biogenesis [Saveanu et al., 2003, Talkish et al., 2012]. Their molecular functions are unknown. During biogenesis, the pre-ribosomal particle is transfered from the nucleolus to the nucleoplasm and further to the cytoplasm for maturation. Yeast mutants with defects in their deconjugation pathways also develop defects in the pre-60S ribosomes [Panse et al., 2006]. Panse *et al.* also showed that many biogenesis factors are SUMOylated in this pathway [Panse et al., 2006]. SENP3-depleted cells accumulate pre-60S ribosomal particles in the nucleolus, indicating that the release into the nucleoplasm is dependent on deSUMOylation [Finkbeiner et al., 2011].

5.3.1.2. Experimental validation of SUMO interaction for Dbp10 and Drs1

The bioinformatical SIMa detection screen gave putative SIM instances with moderate profile scores in Dbp10 (`VIEYSS`, 3.296) and Drs1 (`VPILDS`, 3.481, tables 4.10, 4.11). These motif-like sequences show promising values for absolute, relative and background

conservation. They are additionally predicted to reside in unstructured regions. Yeast two hybrid experiments with nucleotide sequences comprising the first 1529 (Dbp10) and 1731 nucleotides (Drs1) were performed. There was no detectable SUMO interaction. GST-SUMO pulldown assays were performed with the C-terminal HA tagged full length versions. Observed α-HA signal depletion signal indicates SUMO interaction for both proteins (figure 4.12). Elution efficiency of bound material from immobilized GST-SUMO depended on the applied pulldown conditions. Pulldown results suggest SUMO interaction for both proteins.

Dbp10 and Drs1 belong to the DEAD box family of proteins. ATP-binding and helicase domains of Dbp10 and Drs1 indicate functional roles in early steps of the biogenesis pathway [Bernstein et al., 2006]. With SUMOylation as checkpoint control in ribosomal biogenesis, SUMO/SIM recognition of Dbp10 or Drs1 may be plausible.

A conclusion whether there is a biological relevant SUMO interaction for these proteins is not possible. The mechanisms that assemble the pre-60S ribosomal particle in the orchestration of ribosome assembly factors are not understood. Additionally, the mechanisms of SUMOylation-dependent checkpoint control or an involvement of Dbp10 or Drs1 are still unknown.

5.3.2. Tdp1/SUMO interaction

5.3.2.1. Biological context of Tdp1 and Top1

Topoisomerase I (Top1) is involved in DNA replication and transcription. It releases the topological stress upon DNA supercoiling in front of the replication fork by transient induction of single-strand breaks (SSBs). Top1 reversibly forms a thioester bond between its active tyrosine (Tyr-127) and the DNA 3'-end to form a covalent "cleavage complex" [Wang, 2002, Pommier et al., 2003]. This step allows DNA stress release through rotation of the DNA double helix around the Top1–DNA bond. In a second step, Top1 re-ligates the cleaved DNA ends to re-establish the original DNA double-strand. Cleavage complexes are short-living intermediates, as the DNA religation step is faster than the Top1-DNA bond formation step. Endogenous and exogenous factors provoke "trapping" of the cleavage complex leading to SSBs [Pourquier et al., 1999]. These cytotoxic DNA lesions can be further transformed into a double-strand break (DSB) upon collision with a replication fork [Tsao et al., 1993, Strumberg et al., 2000, Pommier et al., 2003]. Tyrosyl-DNA phosphodiesterase 1 (Tdp1) cleaves Top1 from the 3'-phosphotyrosyl-DNA intermediate [Yang et al., 1996, Pouliot et al., 1999, Pouliot et al., 2001, Champoux, 2001, Debethune et al., 2002]. The 3'-phosphate end is then further processed by an 3'-

phosphatase such as polynucleotide kinase phosphatase (PNKP) [Strumberg et al., 2000]. Top1 is ubiquitinated for its 26S proteasomal degradation as a response to transcription blocks and is independent from replication [D'Arpa and Liu, 1995, Fu et al., 1999, Desai et al., 2001, Desai et al., 2003]. Ubiquitination occurs at K117 and K153, whereas ubiquitination competes with SUMOylation at these residues [Mao et al., 2000, Horie et al., 2002].

5.3.2.2. Experimental validation of Tdp1/SUMO interaction

The bioinformatical SIMb screen were trained on SIMb such as the ones in Uls1 `TIDLT` and Wss1 `VIDLT`. Different putative SIMb were selected from the results (figures 4.14, 4.15). The sequences of putative SIMb in Top1 and Tdp1 (`IIDLT`) look the same. However, the different numbers of homologous sequences in the Saccharomycetales MSA give different profile scores: 6.179 for Tdp1 and 4.023 for Top1. The Top1 SIMb candidate shows low overall conservation. By contrast, the SIMb in Tdp1 is comparable in profile score, conservation and structural context to the established SIMb in Uls1 and Wss1. There is no clear preference to decide from the disorder prediction, whether Tdp1 or Top1 is more likely to reside in an unstructured protein region: GlobPlot gives better hints for Top1, whereas IUPred tends to assign Tdp1 more-likely as disordered. However, the IUPred algorithms have been found more trustworthy for SIMs throughout the study. This may be a borderline case for the structural context assignments for a bioinformatical SIMb detection.

Tdp1 and Top1 were investigated as putative SIPs in yeast two-hybrid experiments. Full-length coding sequences were used. Experimental results show a binding preference of Tdp1 for conjugatable Smt3, as interaction only occurs for GAD-Tdp1/GBD-Smt3 and GAD-Tdp1/GBD-Smt3×3 (figure 4.8). Whereas the Smt3 chain is Δ97 truncated, the single Smt3 moiety is not. According to the yeast two-hybrid results, Tdp1 does not appear to bind to isolated Smt3. One possible interpretation of this discrepancy would be a low-affinity SUMO binding by the SIM of Tdp1, which requires another contact between Tdp1 and the SUMOylated protein for robust interaction.

SIM mutations in Tdp1 were not checked for SUMO interaction. Whereas Tdp1 was found to interact with SUMO, Top1 was not.

A SIMb in Tdp1 might play a role in the recognition of SUMOylated Top1 prior to Tdp1 mediated Top1-DNA bond cleavage. The circumstances behind that interaction and its relevance for Top1 release from DNA or for DNA replication as a whole still remain unclear.

5.3.3. Rfc1/SUMO interaction

5.3.3.1. Biological context of Rfc1

The replication factor C (RFC) complex is an essential component of the DNA replication and DNA damage repair pathways. This complex is a heteropentameric ring complex and composed of Rfc1–5 subunits. It is responsible for loading a ring-shaped replication factor, the proliferating cell nuclear antigen (PCNA), onto DNA [Cullmann et al., 1995, Gomes and Burgers, 2001, Yao, 2003, Bowman et al., 2004, Kelch et al., 2011]. The carboxy terminal domains of the subunits hold each other for ring-complex formation. The N-terminal domains are responsible for DNA binding [Cullmann et al., 1995]. The Rfc1 subunit is the largest one and has amino and carboxy terminal extensions compared to the other subunits. The Rfc1 carboxy terminal domain extension packs against Rfc5 as a switch for ring completion of the RFC complex [Beckwith and McAlear, 2000, Jeruzalmi et al., 2001, Bowman et al., 2004]. Subunits Rfc2-5 were found to be involved in PCNA opening and unloading from DNA. Deletion of the Rfc1 amino terminal sequence has no observable replication phenotype, but shows increased PCNA unloading activity from DNA [Yao et al., 2006]. Other RFC-like complexes are known, in which the Rfc1 subunit is replaced by another protein: An Rfc1/Rad24 exchange in *Saccharomyces cerevisiae* forms a clamp loader complex required for checkpoint control during S-phase [Green et al., 2000]. That Rad24–RFC complex also acts as clamp loader of the 9–1–1 complex onto DNA damage sites [Lindsey-Boltz et al., 2001, Bermudez et al., 2003, Parrilla-Castellar et al., 2004, Delacroix et al., 2007, Lee et al., 2007]. A Ctf18–RFC complex is required for sister chromatid cohesion [Hanna et al., 2001, Mayer et al., 2001, Naiki et al., 2001]. An Elg1–RFC complex contributes to chromosome stability and suppresses chromosomal rearrangements [Kanellis et al., 2003, Bellaoui et al., 2003, Ben-Aroya et al., 2003, Parnas et al., 2009]. Elg1–RFC is involved in PCNA unloading after replication [Parnas et al., 2009, Kubota et al., 2013]. Elg1 is an SUMO interacting protein, bearing several SIMs and promoting a preferential association with SUMOylated PCNA [Parnas et al., 2010, Parnas et al., 2011].

5.3.3.2. Experimental validation of Rfc1/SUMO interaction

A SIMa `VIDLDT` was found in Rfc1 by its significant profile score (4.185 and 4.530) in the bioinformatical SIMa and SIMb detection screens (tables 4.10, 4.11, 4.14, 4.15). Its sequence can be interpreted as an intermediate between the consensus sequences in Fir1 (`VILLDE`) and the SIMb consensus sequence of the PIAS proteins (`VIDLT`) (table 1.1) [Song et al., 2004, Hannich et al., 2005, Hecker et al., 2006, Uzunova et al., 2007]. The

SIMa in Rfc1 has promising values for absolute, relative and background conservations in the bioinformatical screens. Additionally, it was predicted to be located most likely in an unstructured region.

Yeast two-hybrid experiments used a fragment comprising the first 705 nucleotides of the Rfc1 coding sequence. The results show Rfc1/SUMO interaction for this segment. Visual inspection of the Saccharomycetales MSA of Rfc1 showed the well-conserved SIM from the bioinformatical screen in a more variant protein region (table 4.14). Additionally, there are two SIM-like regions in close to the one from the screen — a highly conserved MVNISD sequence preceding and a nonconserved SIM consensus VIDVSE following that motif (figure 5.1). All three sequences match the SIMa consensus pattern.
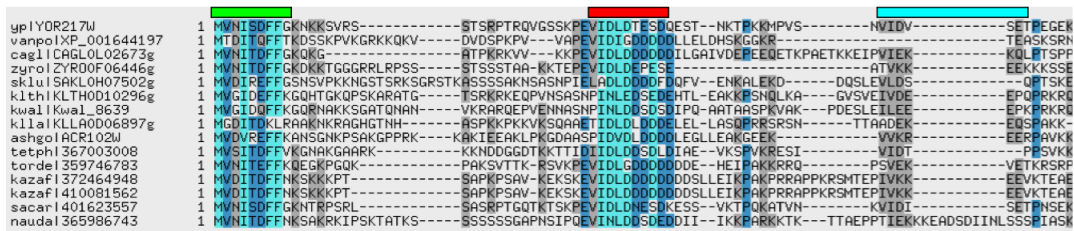
Figure 5.1. An excerpt from the Saccharomycetales MSA of Rfc1 shows three SIM-like sequences. A highly conserved MVINSD sequence (green), the SIMa VIDLDT from the bioinformatical screen (red) and a poorly conserved VIVDSE sequence (blue). All SIM-like sequences follow the SIMa consensus pattern.

The 'flanking' SIM sequences are not included in the final bioinformatical results. The former may give no good profile score, whereas the latter misses the applied conservation restrictions. However, a supporting role of these motifs for a Rfc1/SUMO interaction cannot be excluded. Therefore, mutations are introduced into the Rfc1 yeast two-hybrid construct (table 4.15). Rfc1/SUMO interaction is only observed for the SIMa candidate from the bioinformatical screen. Phenotype analysis of a SIM mutant *rfc1* strain was performed upon thermal stress and thermal/chemical stress. A growth phenotype is only observed for the SIM-mutant strain upon exposure to DNA damage-inducing agents.

On the one hand, these data suggest that a SIM in Rfc1 is biologically relevant in DNA damage response or DNA damage repair. On the other hand, it cannot be excluded that the mutant introduced to Rfc1 affects translation, folding or stability of the protein, which might also explain the observed phenotype. To exclude whether the protein sequence alteration alone reduces PCNA clamp loading activity, different experiments would have to be performed: The functional SIMa could be replaced by the sequence of a different SIM consensus type. SIM relevant Rfc1 function should be restored.

Parnas *et al.* used an experiment to show the biological relevance for a SIM in Elg1.

This SIM is shown to be involved in the process of unloading SUMOylated PCNA from DNA [Parnas et al., 2010].

Similar experiments for Rfc1 and SIM mutant *rfc1* strains could investigate whether the SIM in Rfc1 has a comparable impact on loading SUMOylated PCNA onto DNA as Elg1 has on unloading it. Parnas *et al.* performed Flag-pulldown assays of Elg1 constructs against SUMO-PCNA. They showed that Elg1 has a preference for binding SUMOylated PCNA [Parnas et al., 2010]. Similar to this experiment, Rfc1–SIM/SUMO–PCNA interaction should have an depleting effect on Rfc1 in the presence of SUMOylated PCNA. An *rfc1* SIM mutant construct should not show that effect. In another experiment, SUMOylated PCNA interaction could be restored in the *rfc1* SIM mutant by insertion of a SIM sequence of different consensus type to the original SIM position in the sequence.

5.3.4. Rad18/SUMO interaction

5.3.4.1. Biological context of Rad18

The heterodimeric Rad6–Rad18 protein complex monoubiquitinates the proliferating cell nuclear antigen (PCNA) at a highly conserved lysine (K164) in response to DNA damage [Hoege et al., 2002, Stelter and Ulrich, 2003, Kannouche et al., 2004, Bienko et al., 2005]. This step is important for translesion synthesis as an initial signal for DNA damage repair [Bailly et al., 1997, Hoege et al., 2002, Stelter and Ulrich, 2003, Kannouche et al., 2004, Bienko et al., 2005]. Extension of that monoubiquitin modification to a ubiquitin chain activates an error-free damage repair pathway [Hoege et al., 2002]. This active site K164 can also be SUMOylated by the SUMO E3 Siz1 [Hoege et al., 2002, Stelter and Ulrich, 2003]. SUMOylation at this position is observed during S phase and was shown to interact with Srs2 helicase which in turn inhibits unwanted recombination with Rad52 filaments [Hoege et al., 2002, Pfander et al., 2005, Papouli et al., 2005, Burkovics et al., 2013]. SUMOylation therefore acts as a competitor to ubiquitin to prevent unscheduled recombination events by enhancing Srs2 helicase binding during S phase [Papouli et al., 2005, Pfander et al., 2005, Parker et al., 2008]. Parker and Ulrich recently found that SUMOylated PCNA serves as a substrate for the *Saccharomyces cerevisiae* Rad18 ubiquitin ligase and that the SUMO binding step is mediated by a SIMr in Rad18 [Parker and Ulrich, 2012]. In this study, the same SIMr was found in the bioinformatical SIMr detection screen. SUMO/Rad18 interaction was also detected. Rad18 was proposed to mediate the cooperation between S phase-associated (mono-) SUMOylation and the DNA damage-induced ubiquitination of PCNA. Ubiquitination

of SUMOylated PCNA may be important during replication, when replication problems or DNA damage occur.

5.3.4.2. Experimental validation of Rad18/SUMO interaction

Rad18 was tested in yeast two-hybrid experiments in form of a nucleotide sequence comprising the first 801 and the putative SIMr coding sequence. The results show Rad18 interaction with Smt3 and with 3×Smt3 and 4×Smt3 chains. The interaction with 4×Smt3 chains is the strongest (figure 4.8). Full-length Rad18 was used in GST-SUMO pulldown experiments (section 4.6.2). Rad18/SUMO interaction is mainly detected as α-HA signal depletion after exposure of HA-taged Rad18 to immobilized GST and GST-SUMO as bait.

The α-HA signal difference after elution of immobilized material indicates weak SUMO interaction. Washing of immobilized material is the most sensitive step of the pulldown experiment. Therefore, α-HA signal intensities for Rad18 are significantly reduced in that experimental step. Putative differences in α-HA signal intensities are hard to detect. The GST-SUMO pulldown results are therefore a weak support to the clear yeast two-hybrid results. The Rad18/SUMO interaction via this SIM is very weak. Yeast two-hybrid experiments show a clear preference for Rad18 towards 4×SUMO chains. This is not covered by the GST-SUMO pulldown experiments with only a single SUMO moiety provided. SUMO binding may be therefore only sufficient when other binding factors are involved. Rad18 SIM/SUMO interaction is meanwhile confirmed by the works of Parker and Ulrich [Parker and Ulrich, 2012]. They showed that Rad18 interacts quite well with SUMOylated PCNA but hardly with isolated SUMO. This may explain, why GST-SUMO pulldown assays for Rad18 gave no clear result.

5.4. The application of analogous methods for a bioinformatical AIM detection approach

The number of established AIMs is low compared to the number of established SIMs. AIMs and SIMs are found in the same structural context. That is why a bioinformatical AIM approach was designed with analogous methods as for the SIM detection approach (tables 4.26 and 4.27).

AIM sequences seem to be not as strict to a consensus as are SIMs. The conservation characteristics of established AIM are found similar to those for SIMs, whereas the applied bioinformatical methods allow no clear judgement, whether an established AIM

is predicted as unstructured, non-globular or globular.

5.5. General aspects of the applied bioinformatical approach

The basic assumptions of what features constitute a good SUMO interacting motif are most relevant for the results of the bioinformatical SIM detection approach. This approach is suited for a detection of SIM which fit to these basic assumptions regarding conservation and structural context. Biologically relevant motifs in general are believed to be conserved among a certain phylogenetic range. SIMs which show less absolute conservation, which lie next to other conserved protein segments or which lie inside structured protein regions may not be detected. Biologically relevant SIMs like these exist, but are hard to distinguish from false-positive SIM candidates from the bioinformatical screen. The number of false-positive SIMs would be increased with a modification of the approach to yield more SIMs with 'atypical' conservation levels. While experimental methods are good at finding proteins and peptides that are physically able to interact with SUMO under the given experimental conditions, the bioinformatical approach is able to find SIMs that are selected for SUMO binding. While the latter set might be smaller than the former, it should be enriched for SIMs with biologically important functions.

5.6. General aspects of experimental validations

There are some critical aspects regarding the experimental validation of the results from bioinformatical SIM and AIM detection approaches. As already mentioned, the bioinformatical approach targets SIMs which have been evolved for efficient SUMO binding in the course of evolution. For experimental validation of SIMs and AIMs, however, the complete biological context has to be matched in the experimental conditions, with regard to all relevant binding partners, modifications etc. Experimental results from binding assays of a SIP against an isolated SUMO moiety or against an biologically irrelevant SUMOylated protein may give negative results. Proteins such as Rad18 bind distinct SUMOylated proteins via additional interaction between other parts of these proteins [Parker and Ulrich, 2012]. The SIP therefore is able to distinguish between SUMOylated proteins, with the SUMO binding being only one part of the overall interaction between proteins.
STUbLs belong to another class of SUMO interacting proteins. They are proteins with several SIMs for binding of all kinds of polySUMOylated proteins. The additive use

of SIMs allows SUMO chain recognition. The SUMOylated protein is not involved in the binding. The Slx5–Slx8 heterodimer in *Saccharomyces cerevisiae* is such a STUbL (section 1.3.4) [Wang et al., 2006]. This complex recognize high molecular weight SUMO conjugates via four SIMs (2 SIMa, 2 SIMb) [Uzunova et al., 2007, Mullen and Brill, 2008]. Experimental validations have therefore to be carefully designed. Their results have to be counter-checked by several methods.

A. Appendix: Motif list from the bioinformatical SIMa detection approach

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach. A SIMa profile search within a *Saccharomyces cerevisiae* protein sequence database with a sensitivity of $C = 3.2$ and further refinement by conservation criteria ($t1 = 1.0$, $t2 = 1.6$, $t3 = 0.2$) gave this list of putative SIMa instances. Evaluation of results was performed according to a combined 'bioinformatical motif score'. This score describes the profile adherence, conservation and structural context features of the instance.

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
1	SIZ1	YDR409W	482-489	IINLDSDD	5.258	2242.830
2	NFI1	YOR156C	472-479	IISLDSSD	4.444	1322.214
3	LEO1	YOR123C	388-395	DFLVDDDE	3.259	902.393
4	RIX1	YHR197W	749-756	AIELSDDE	3.963	818.023
5	DBP10	YDL031W	97-104	MLEMSDDE	3.370	628.850
6	DBP10	YDL031W	60-67	VIEYSSDE	3.296	465.979
7	SLX5	YDL013W	24-31	VILIDSDK	5.480	464.324
8	HSL1	YKL101W	1295-1302	TMLFDEEE	3.333	455.710
9	GAL83	YER027C	66-73	LIFNDDDD	3.741	400.741
10	UBA2	YDR390C	632-636	IVELD	3.222	394.650
11	TIF3	YPR163C	384-391	VLRTEDDD	3.370	382.034
12	NTR2	YKR022C	164-171	VKLLDSED	4.111	348.977
13	LEO1	YOR123C	454-461	VAVIEDDE	5.036	333.774
14	IST1	YNL265C	218-225	ILALDNDD	4.962	275.250
15	SGS1	YMR190C	324-331	IQVLDDED	5.110	274.918
16	ENT5	YDR153C	293-300	LIDLDDST	3.593	261.275
17	PNO1	YOR145C	70-77	TVVVDDQG	3.519	257.320
18	FIR1	YER032W	759-766	VILLDEDE	6.331	246.216
19	PSH1	YOL054W	398-405	VVLGDSDD	3.852	226.974
20	APL5	YPL195W	794-801	VQVLSDEP	3.370	214.335
21	TOM1	YDR457W	1940-1947	VVFSDEDD	3.815	196.191

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
22	ATG36	YJL185C	95-102	ILSISSDS	3.259	186.813
23	YPT35	YHR105W	13-20	IQLLDEDS	4.000	183.875
24	HPC2	YBR215W	524-531	VIELDDDE	6.257	179.245
25	DRS2	YAL026C	58-65	TIDLDADD	3.259	175.694
26	ULS1	YOR191W	371-378	IIILSDED	5.887	167.260
27	RFC1	YOR217W	33-40	VIDLDTES	4.185	163.977
28	UPF3	YGR072W	323-330	VVIIEEAG	3.259	142.679
29	KSP1	YHR082C	942-949	AIIFEDEE	4.111	134.064
30	MEF2	YJL102W	715-722	ILSIEDES	4.222	130.283
31	ATG33	YLR356W	130-137	IIDLGEDN	3.222	128.048
32	ZIP1	YDR285W	856-863	LLLVEDED	4.370	124.929
33	YGR016W	YGR016W	31-38	ILPLDTEE	3.889	119.976
34	UBA2	YDR390C	583-590	IVILDDDE	6.961	115.496
35	MMS22	YLR320W	9-16	IVISDSEA	3.407	95.976
36	DRS1	YLL008W	24-31	VPILDSSD	3.481	90.953
37	SNU66	YOR308C	305-312	VKLVDEDE	4.185	78.613
38	GCR2	YNL199C	412-419	VNTLDDEA	3.556	76.291
39	PTC5	YOR090C	462-469	VIDVSEDK	3.407	74.247
40	RTT106	YNL206C	435-442	PIEIDNDD	3.519	70.334
41	RAD5	YLR032W	372-379	IIDLDNDE	5.295	70.319
42	SFP1	YLR403W	535-542	IDDIDDDD	3.852	68.511
43	SPP381	YBR152W	171-178	CLLLDDND	3.593	66.674
44	RTC1	YOL138C	1008-1015	VIIEEDEH	3.593	64.145
45	UBP14	YBR058C	102-109	VIETSEDD	3.333	62.887
46	RCN1	YKL159C	204-211	IFDTDDDD	3.667	57.596
47	YDR186C	YDR186C	760-767	LLNLEEED	3.222	55.181
48	ARP8	YOR141C	18-25	IPLEDDDD	3.222	53.889
49	PNO1	YOR145C	41-48	DVLLDDSD	3.519	53.124
50	YHR131C	YHR131C	470-477	VIETEEDD	3.667	50.129
51	SAS10	YDL153C	63-70	VLAMDEDD	4.703	49.916
52	FYV8	YGR196C	437-444	ILLTSDEE	3.852	46.348
53	UBX2	YML013W	568-575	VEALDEED	3.704	43.593

A. Appendix: Motif list from the bioinformatical SIMa detection approach

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
54	SAP185	YJL098W	829-836	YVSFDEDE	3.296	43.085
55	CEF1	YMR213W	381-388	EIVLSEDE	3.444	42.893
56	SLX8	YER116C	179-186	VLQISDDD	4.481	42.725
57	PWP2	YCR057C	884-891	VVMESDDE	3.778	42.295
58	XRS2	YDR369C	768-775	IFITDEDD	4.148	41.512
59	STB3	YDR169C	201-208	VVYIDENA	3.852	39.926
60	SIR4	YDR227W	750-757	VYVSDSDD	3.333	37.203
61	YGP1	YNL160W	206-213	VAIVDDQD	4.592	36.276
62	VPS72	YDR485C	551-558	VAIIDTEE	4.444	34.115
63	NHA1	YLR138W	796-803	LLIIENED	4.148	28.527
64	ZRG8	YER033C	184-191	VIILNDPA	4.111	27.855
65	BNI4	YNL233W	614-621	LVMASDEE	3.444	26.784
66	MSS4	YDR208W	240-247	ILPMDDSD	3.815	24.196
67	YMR206W	YMR206W	62-69	VLNINDND	3.704	22.222
68	YML108W	YML108W	10-17	LVLLEDDT	4.481	22.015
69	CST6	YIL036W	526-533	LLMIDSDV	3.259	20.960
70	YFR016C	YFR016C	1091-1098	LKIVDDSE	3.333	20.535
71	MSH6	YDR097C	76-83	FVDVDEDN	3.407	19.388
72	MRE11	YMR224C	633-640	IIMVSTDE	4.222	18.525
73	RMR1	YGL250W	226-233	VIVMDIDD	4.333	18.500
74	BEM2	YER155C	1453-1460	TLILKDDN	3.333	18.411
75	ILM1	YJR118C	172-179	IEIINDEE	4.185	18.195
76	ESL2	YKR096W	1153-1160	VVLISDDD	5.665	17.911
77	SSL1	YLR005W	4-11	VVISESEE	3.593	17.723
78	BEM2	YER155C	215-222	VIHLNSEN	3.556	17.588
79	SYF2	YGR129W	152-159	ITIADDDK	3.926	17.283
80	MEC1	YBR136W	1579-1586	LIAISNED	3.704	16.027
81	YGL081W	YGL081W	8-15	VFAIEDTE	3.333	15.713
82	ECT1	YGR007W	261-268	AVIIDADA	3.333	15.121
83	LOS1	YKL205W	538-545	LLAIDNEQ	3.667	15.085
84	SVS1	YPL163C	120-127	IITLSTES	3.481	15.069
85	NUP188	YML103C	1375-1382	IVNLEDNT	3.519	14.784

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
86	RSE1	YML049C	560-567	ILQIDNES	4.148	13.922
87	VAN1	YML115C	111-118	TIDLEDEE	3.926	13.745
88	RAD2	YGR258C	304-311	AIVIDDKD	4.851	13.311
89	YDL176W	YDL176W	612-619	IVIINDID	4.037	13.086
90	BRR2	YER172C	207-214	VAILADDE	4.000	12.922
91	RUD3	YOR216C	261-268	MIILENNK	3.370	12.803
92	RSN1	YMR266W	889-896	VVEMNDEN	4.037	12.665
93	CEP3	YMR168C	547-554	IIELKNDE	4.074	12.478
94	FAR7	YFR008W	183-190	TIPLEDEE	3.741	12.041
95	MET12	YPL023C	338-345	DIVLDDSN	3.407	11.859
96	ULP2	YIL031W	900-907	IVISDTEQ	3.370	11.196
97	YMR279C	YMR279C	31-38	VVSTEDEE	3.926	10.948
98	ZIP1	YDR285W	689-696	VIVLKSEK	4.074	10.798
99	UGO1	YDR470C	160-167	PIILRDEE	3.407	10.079
100	DSE1	YER124C	404-411	SISLDSES	3.222	9.908
101	FKH1	YIL131C	250-257	TIMMEEDE	3.778	9.870
102	MLP1	YKR095W	586-593	IITLKSEK	3.407	9.809
103	TSR2	YLR435W	147-154	VVHIEGDD	3.519	9.676
104	AFI1	YOR129C	149-156	TIVLEDDE	5.110	9.526
105	OCA4	YCR095C	241-248	LIRVNEDD	3.444	9.360
106	YHL009W-B	YHL009W_B	1715-1722	IVMITDSK	3.556	9.137
107	RTT10	YPL183C	425-432	VTEYDDDS	3.370	8.238
108	URB1	YKL014C	1435-1442	VKMVKDDE	3.259	8.152
109	SHE10	YGL228W	237-244	TITLDQEE	3.926	7.969
110	AKR2	YOR034C	4-11	MSIIDDEN	4.148	7.069
111	DSN1	YIR010W	32-39	IPILESDS	3.222	6.887
112	CPR5	YDR304C	188-195	VIIVESGE	3.926	6.683
113	YJL113W	YJL113W	1461-1468	IVMITDSK	3.556	6.236
114	DLS1	YJL065C	148-155	IVEIDLDN	3.444	5.918
115	YHL008C	YHL008C	595-602	IVNLNKED	3.519	5.897
116	SPT21	YMR179W	541-548	TIALENED	3.444	5.551
117	YJL132W	YJL132W	428-435	ILTIKDET	3.481	5.376

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
118	MSC3	YLR219W	279-286	VMPLEEES	3.222	5.332
119	STE5	YDR103W	880-887	VLLLSDMD	3.444	5.332
120	IRA1	YBR140C	1033-1040	VILSDNDE	4.444	5.229
121	SAM37	YMR060C	62-69	VLILDNGT	3.519	4.364
122	BIR1	YJR089W	935-942	IETLEDDN	3.259	4.357
123	SNF2	YOR290C	440-447	VVDIDDPD	4.703	4.065
124	YPR117W	YPR117W	654-661	ILLADSQE	3.407	4.047
125	ARP10	YDR106W	202-209	IIIVNIEE	3.481	3.891
126	MCM22	YJR135C	161-168	VPNLDDND	3.333	3.732
127	RIC1	YLR039C	671-678	IILVTDTQ	3.444	3.732
128	NUM1	YDR150W	387-394	MIALPNDD	3.370	3.366
129	LIF1	YGL090W	299-306	VIKMEDDD	4.777	3.220
130	MGM1	YOR211C	145-152	ATSLDDDE	3.741	3.006
131	YGL176C	YGL176C	423-430	ILEISDDG	3.593	2.987
132	YER130C	YER130C	134-141	VLMVSDDE	4.814	2.834
133	NUM1	YDR150W	350-357	MILLSNDS	3.815	2.433
134	TLG1	YDR468C	59-66	IIVMKRDE	3.815	2.222
135	NTR2	YKR022C	265-272	VLFPDDDE	4.037	1.753
136	RTS3	YGR161C	118-125	SAILDDDD	4.185	1.708
137	SIZ1	YDR409W	442-449	AILEDDDD	3.889	1.586
138	OPY1	YBR129C	173-180	LIVVDEKA	3.556	1.499
139	MDM30	YLR368W	485-492	YVLLDDNN	4.185	1.364
140	INP51	YIL002C	856-863	IAVLSDDA	3.741	1.352
141	YDR239C	YDR239C	264-271	PLIVDNEE	3.333	1.264
142	GRC3	YLL035W	42-49	TVVLNSEE	3.815	1.095
143	RTC4	YNL254C	360-367	IIVADSDP	4.370	1.013
144	PRP31	YGR091W	141-148	ISILENEN	3.333	0.836
145	SCM3	YDL139C	161-168	IIDISDEE	4.888	0.807
146	ARE2	YNR019W	29-36	SITVDDEG	3.444	0.694
147	PTC5	YOR090C	124-131	VLILNDSK	3.926	0.635
148	YDR056C	YDR056C	93-100	IVDVDEDN	4.481	0.541
149	UFE1	YOR075W	13-20	VAVIDDAR	3.444	0.510

A. Appendix: Motif list from the bioinformatical SIMa detection approach

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
150	PTA1	YAL043C	615-622	SILLDDDD	5.332	0.492
151	SCC2	YDR180W	913-920	IVELNSDD	4.333	0.391
152	OPT2	YPR194C	93-100	IIYFKDDE	4.000	0.327
153	RAD9	YDR217C	905-912	KIILEDNE	3.667	0.246
154	VTH1	YIL173W	39-46	IVNFDDSN	3.481	0.212
155	VTH2	YJL222W	39-46	IVNFDDSN	3.481	0.210
156	PPM2	YOL141W	167-174	IIGLSEDK	3.259	0.194
157	SPP41	YDR464W	1382-1389	KIVIDDKE	4.185	0.172
158	MTW1	YAL034W_A	272-279	IVSIDIEE	3.630	0.148
159	VPS52	YDR484W	7-14	VLSLDQDK	3.778	0.077
160	NSA1	YGL111W	407-414	IVMLDDVE	4.777	0.033
161	PCL8	YPL219W	310-317	IPELSDDE	3.296	0.023
162	ECM16	YMR128W	680-687	VQVIDQDK	3.519	0.007
163	HAP1	YLR256W	496-503	IPILDEQN	3.296	0.000
164	ATG31	YDR022C	86-93	VIIVQLDE	3.444	0.000
165	GCR1	YPL075W	324-331	ILILDKNS	3.926	0.000
166	POL4	YCR014C	49-56	VILIEDSF	3.556	0.000
167	WHI3	YNL197C	189-196	VEIIDDTT	3.222	0.000
168	RAD61	YDR014W	302-309	LMIINDEE	4.740	0.000
169	YMR185W	YMR185W	539-546	LVLLNDEE	4.962	0.000
170	ATH1	YPR026W	393-400	VILTEDQP	3.407	0.000
171	WSS1	YHR134W	247-254	VVILDDDD	6.924	0.000
172	TFB6	YOR352W	88-95	VTMLDDVD	3.667	0.000
173	YBR071W	YBR071W	163-170	TIAIDDSK	3.519	0.000
174	YMR160W	YMR160W	102-109	PIVVDDNT	3.222	0.000
175	POP6	YGR030C	139-146	VTLVSDSE	3.333	0.000
176	SWP82	YFL049W	47-54	LQILDQDE	3.519	-0.006
177	CWC22	YGR278W	227-234	YIVSDEDE	3.593	-0.012
178	TFB6	YOR352W	269-276	VILTDDND	4.851	-0.067
179	PRP4	YPR178W	79-86	ILMVDEID	3.259	-0.127
180	TRF5	YNL299W	53-60	AIDVEDDD	3.778	-0.135
181	TSR1	YDL060W	427-434	MMEIDDEM	3.444	-0.176

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
182	YGL082W	YGL082W	306-313	VQQIEDDE	3.667	-0.223
183	PIG2	YIL045W	329-336	LLINDDDD	3.926	-0.265
184	ATG7	YHR171W	180-187	VCILDADD	3.630	-0.301
185	YDR262W	YDR262W	190-197	IVVYNDDK	4.185	-0.330
186	YDR476C	YDR476C	65-72	IYFIDNDT	3.519	-0.569
187	MSC6	YOR354C	446-453	LLMLDDEK	4.629	-0.641
188	KNS1	YLL019C	92-99	AIELDEEP	3.593	-0.717
189	MMS22	YLR320W	399-406	VISLDAAE	3.296	-0.728
190	BRR2	YER172C	1819-1826	FIEIDDTE	4.000	-0.746
191	YBL005W-B	YBL005W_B	1605-1612	LVVISDAS	3.259	-0.786
192	ATG14	YBR128C	33-40	LILLKDEN	4.148	-0.828
193	YBR012W-B	YBR012W_B	1606-1613	LVVISDAS	3.259	-0.836
194	FRE2	YKL220C	494-501	VIILKEKK	3.407	-0.875
195	YER160C	YER160C	1605-1612	LVVISDAS	3.259	-0.878
196	YMR050C	YMR050C	1605-1612	LVVISDAS	3.259	-0.878
197	YGR027W-B	YGR027W_B	1605-1612	LVVISDAS	3.259	-0.878
198	YGR161C-D	YGR161C_D	1605-1612	LVVISDAS	3.259	-0.878
199	YJR029W	YJR029W	1605-1612	LVVISDAS	3.259	-0.878
200	YMR045C	YMR045C	1605-1612	LVVISDAS	3.259	-0.878
201	YNL284C-B	YNL284C_B	1605-1612	LVVISDAS	3.259	-0.878
202	YOR142W-B	YOR142W_B	1605-1612	LVVISDAS	3.259	-0.878
203	YJL114W	YJL114W	182-189	IIEIDPRE	3.481	-1.158
204	YHL009W-A	YHL009W_A	182-189	IIEIDPRE	3.481	-1.168
205	TRS120	YDR407C	1213-1220	IVIFDSKT	3.407	-1.491
206	SPP41	YDR464W	92-99	VQNIDDEQ	3.630	-1.503
207	RIM101	YHL027W	578-585	MEDLDDEE	3.259	-1.541
208	RAX2	YLR084C	739-746	FIVLDNDY	3.889	-1.901
209	AFG2	YLR397C	178-185	VVITDASD	3.222	-2.027
210	YOR385W	YOR385W	195-202	LMFLDEDR	3.444	-2.384
211	VID27	YNL212W	368-375	ILHIDDRS	3.519	-2.671
212	RED1	YLR263W	460-467	YVHIDSED	3.593	-2.874
213	SIP4	YJL089W	539-546	LILTDDSN	3.519	-3.137

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
214	ESL2	YKR096W	761-768	VYLLDNQT	3.222	-3.170
215	MDN1	YLR106C	2348-2355	ILLLSDEE	5.073	-3.912
216	PEX6	YNL329C	170-177	LILVNDTE	3.778	-3.996
217	ULP2	YIL031W	198-213	VNLLDDVE	3.259	-4.428
218	TSC11	YER093C	624-631	VLILENSE	4.037	-5.163
219	YOS9	YDR057W	37-44	ISYIDEDD	4.111	-5.307
220	SPI1	YER150W	28-35	VIVVPSSD	3.556	-6.426
221	MUD1	YBR119W	53-60	LMLLDDQM	3.407	-6.663
222	ELP6	YMR312W	181-188	LIIVSNSD	3.296	-7.521
223	UFO1	YML088W	252-259	TLIMDDEK	3.963	-8.112
224	RAD2	YGR258C	432-439	ISVEDDDE	3.704	-8.723
225	YJR039W	YJR039W	561-568	VLVLNDEI	3.815	-9.034
226	UIP4	YPL186C	87-94	VATLNNEE	3.222	-9.305
227	CFT1	YDR301W	458-465	LMDINDDD	3.778	-10.049
228	NAT4	YMR069W	68-75	IIYIPEDD	4.259	-10.410
229	GPI15	YNL038W	20-27	TLVIDEDK	3.963	-11.301
230	DSL1	YNL258C	430-437	DVNIDDEE	3.556	-12.027
231	VPS16	YPL045W	513-520	LYNLDDDS	3.593	-12.620
232	MND2	YIR025W	232-239	VVDEDDNE	3.296	-12.678
233	YDL027C	YDL027C	334-341	VIILDESF	3.889	-13.276
234	CSF1	YLR087C	1636-1643	VLLVDKSE	3.778	-13.579
235	YIG1	YPL201C	446-453	LTILTDDN	3.222	-13.818
236	OST6	YML019W	33-40	ILQLKDDT	3.333	-14.189
237	PEP5	YMR231W	274-281	ILIVTEET	3.222	-14.887
238	LRG1	YDL240W	585-592	VIQTDDPS	3.259	-16.558
239	STV1	YMR054W	86-93	ILHIDDEG	4.185	-16.701
240	RIM15	YFL033C	108-115	MLLMNDDT	3.370	-17.102
241	SNG1	YGR197C	144-151	IVVLQDAP	3.556	-18.751
242	IRC6	YFR043C	10-17	ILVLSDHP	3.519	-20.514
243	RCY1	YJL204C	191-198	ILLISNEE	4.333	-20.872
244	RKR1	YMR247C	238-245	KVILSDES	3.296	-22.769
245	NUP133	YKR082W	1006-1013	YSMLDDEE	3.815	-24.786

A. Appendix: Motif list from the bioinformatical SIMa detection approach

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
246	VTH1	YIL173W	258-265	VIVLTQDD	4.518	-25.606
247	VTH2	YJL222W	258-265	VIVLTQDD	4.518	-25.606
248	YHL009W-B	YHL009W_B	182-189	IIEIDPRE	3.481	-25.808
249	NMD5	YJR132W	791-798	VLALDDSL	3.333	-27.049
250	SYT1	YPR095C	479-486	IILTEKDD	3.667	-27.062
251	IRA1	YBR140C	1081-1088	VSILDSNQ	3.407	-30.850
252	AIP1	YMR092C	421-428	AVLTNDDD	3.333	-35.000
253	DFG16	YOR030W	186-193	VLLLPSDN	3.519	-37.729
254	SPP382	YLR424W	255-262	ILKLSDDE	4.296	-38.094
255	KAP114	YGL241W	418-425	CILLNDDE	4.148	-39.429
256	SPO22	YIL073C	613-620	IVKVSEEE	3.704	-42.020
257	UBR2	YLR024C	269-276	AIQIEEEE	3.519	-42.717
258	OSH3	YHR073W	526-533	VIILPDTE	4.444	-44.358
259	SIR1	YKR101W	573-580	PILLDDQT	3.296	-44.566
260	PEX32	YBR168W	59-66	IMWLSDDK	3.222	-49.504
261	YOR342C	YOR342C	69-76	IVNIDNDS	4.629	-52.198
262	CSF1	YLR087C	1244-1251	VQVLDDIE	3.296	-52.209
263	YLL054C	YLL054C	719-726	IYLVDTDA	3.222	-52.702
264	UBA2	YDR390C	509-516	IILFSDEE	4.777	-61.970
265	RSC3	YDR303C	444-451	LIPLRDDE	3.556	-73.602
266	YKL100C	YKL100C	235-242	VTIADDNE	3.815	-74.568
267	TRS65	YGR166W	57-64	IILINEAQ	3.519	-74.835
268	VPS41	YDR080W	951-958	LVIINDET	4.407	-77.022
269	YDL183C	YDL183C	46-53	VIPITDKE	3.444	-78.445
270	CBP2	YHL038C	541-548	IVILDKKN	4.148	-79.621
271	EPS1	YIL005W	581-588	VIIIDKSN	4.444	-79.655
272	SWR1	YDR334W	1046-1053	VFTLNDKD	3.296	-82.663
273	ALF1	YNL148C	80-87	IVVVTDSN	3.704	-94.755
274	BLM10	YFL007W	1513-1520	IISLSDYE	3.481	-95.835
275	YJR030C	YJR030C	208-215	ILELDNGE	3.444	-95.883
276	TCA17	YEL048C	8-15	VSLIDESD	3.741	-96.184
277	DOC1	YGL240W	31-38	VLVLDDRI	3.593	-96.247

Table A.1. List of SIMa instances from the bioinformatical SIMa detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
278	VIK1	YPL253C	481-488	FVFLSDDE	4.000	-98.490
279	NUP82	YJL061W	142-149	IVVLKEDD	4.888	-109.273
280	VPS16	YPL045W	165-172	IILLDVDH	4.037	-124.996
281	GIS3	YLR094C	216-223	IININDLD	3.333	-136.543
282	GRX3	YDR098C	38-45	VIEINDQE	4.518	-137.793
283	SAS2	YMR127C	333-338	YLLIDD	3.444	-165.475
284	YLR049C	YLR049C	415-422	VILLEDLR	3.444	-173.160
285	OGG1	YML060W	52-59	VVILRQDE	4.222	-173.296
286	ELP6	YMR312W	145-152	IVIIEQPE	3.963	-244.552
287	PEX1	YKL197C	521-528	LIVLDNVE	3.778	-260.984
288	SHU1	YHL006C	26-33	VIVLGDTA	3.222	-274.184
289	SPO11	YHL022C	229-236	IVIVEKEA	3.593	-333.268
290	MSS1	YMR023C	390-397	IIVVNKSD	3.630	-498.190

B. Appendix: Motif list from the bioinformatical SIMb detection approach

Table B.1. List of SIMb instances from the bioinformatical SIMb detection approach. A SIMb profile search within a *Saccharomyces cerevisiae* protein sequence database with a sensitivity of $C = 4.0$ and further refinement by conservation criteria ($t1 = 1.2$, $t2 = 1.55$, $t3 = 1.0$) gave this list of putative SIMb instances. Evaluation of results was performed according to a combined 'bioinformatical motif score'. This score describes the profile adherence, the conservation and the structural context features of the instance.

For the SIMb bioinformatical detection approach, a more stringent criterion $t1 = 1.75$ was applied first, the weakened criterion $t1 = 1.2$ in a second screen. The results of both screens differ in one instance: The Tdp1 instance shows less conservation than the others in this list, affecting its overall 'bioinformatical score'.

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
1	CTF19	YPL018W	1-9	MDFTSDTTN	4.556	1828.752
2	SIZ1	YDR409W	482-491	IINLDSDDDE	5.139	1755.392
3	RIX1	YHR197W	749-758	AIELSDDEEE	4.835	1255.141
4	MPP10	YJR002W	273-282	KLDLFADEED	4.429	1059.138
5	DBP10	YDL031W	60-69	VIEYSSDEEE	4.505	768.928
6	SIP5	YMR140W	466-475	AIRLSLEDQD	4.099	735.609
7	ZRG8	YER033C	609-618	LFDLSDEDDN	4.251	706.659
8	DRS2	YAL026C	58-67	TIDLDADDDN	5.443	587.725
9	CST9	YLR394W	213-222	TVDLTVDDNS	6.534	573.074
10	MSB3	YNL293W	84-93	VIDLYGDEVE	5.672	486.758
11	DBR1	YKL149C	265-274	DLDLSSDEDE	4.936	331.376
12	YPT11	YNL304W	398-407	TVDITKPNDD	4.327	294.070
13	AKL1	YBR059C	904-913	SIDIDLDDAR	4.200	229.192
14	RFC1	YOR217W	33-42	VIDLDTESDQ	4.530	188.952
15	STU1	YBL034C	715-724	QIDLTDELSN	4.581	181.205
16	SGS1	YMR190C	511-520	HIDLLEDDLE	4.632	154.154
17	PTC5	YOR090C	462-471	VIDVSEDKEA	4.124	144.570

Table B.1. List of SIMb instances from the bioinformatical SIMb detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
18	SAP1	YER047C	235-244	LIDLTNDEDD	6.889	138.866
19	PSY4	YBL046W	192-201	SVNLGYDDED	4.226	135.789
20	IRC5	YFR038W	39-48	VNDLTADISD	4.987	102.798
21	IRE1	YHR079C	635-644	SLDLTTEKKK	4.429	71.083
22	KAR1	YNL188W	359-368	IIELLKDDTD	4.784	66.910
23	TDP1	YBR223C	57-66	IIDLTNQEQD	6.179	44.475
24	ULS1	YOR191W	6-15	TIDLTLADSD	6.559	37.880
25	PCH2	YBR186W	211-220	KIDLELDEDD	4.835	37.024
26	IWR1	YDL115C	294-303	FIDLDGQEGE	4.454	26.307
27	GPI16	YHR188C	361-370	CFDLSNDQNE	4.251	1.198
28	WSS1	YHR134W	265-269	VIDLT	5.164	0.000
29	IRC3	YDR332W	463-472	SVDLTLCSFD	4.251	0.000
30	CIN4	YMR138W	132-141	KIDLVEDKSE	4.911	0.000
31	YMD8	YML038C	418-427	SVDLTNQEYE	5.291	0.000
32	VPS8	YAL002W	761-770	VIDLLLDAMK	4.353	-136.843
33	ERG8	YMR220W	115-124	VIDIFSDDAY	4.479	-659.339
34	SLX4	YLR135W	470-479	IIDLTQESFK	5.443	-722.481

C. Appendix: Motif list from the bioinformatical SIMr detection approach

Table C.1. List of SIMr instances from the bioinformatical SIMr detection approach. A SIMr pattern search within a *Saccharomyces cerevisiae* protein sequence database and further refinement by conservation criteria ($t1 = 1.75$, $t2 = 1.55$, $t3 = 1.0$) gave this list of putative SIMr instances. Evaluation of results was performed according to a combined 'bioinformatical motif score'. This score describes the profile adherence, the conservation and the structural context features of the instance.

Rank	Gene	ORF	amino acid coordinates	sequence	bioinformatical motif score
1	VPS72	YDR485C	61-66	DEIGLL	268.104
2	FIP1	YJR093C	79-84	SDLEVI	163.276
3	YOL036W	YOL036W	125-130	EELALL	158.981
4	NAB6	YML117W	993-998	SDVEVI	118.589
5	RRP14	YKL082C	129-134	EDIKVI	115.133
6	ULP2	YIL031W	723-728	EEIQII	108.347
7	APL5	YPL195W	891-896	EEVIVI	65.484
8	EAR1	YMR171C	476-481	DSLPML	64.325
9	YJL070C	YJL070C	53-58	DDMDMI	55.846
10	TOP2	YNL088W	1356-1361	SDLEIL	51.268
11	SGS1	YMR190C	543-548	SDLELI	49.960
12	RAD2	YGR258C	446-451	EEIEMM	41.125
13	ULS1	YOR191W	369-374	SSIIIL	36.119
14	SLU7	YDR088C	214-219	EEIELM	34.873
15	RRP6	YOR001W	652-657	DDLVVL	33.595
16	FTH1	YBR207W	452-457	SSVPLI	31.348
17	YMR166C	YMR166C	8-13	SSIPII	19.854
18	RAD18	YCR066W	137-142	DDLQIV	18.922
19	PLM2	YDR501W	252-257	DELCLL	16.525
20	NTR2	YKR022C	96-101	DDLLIL	12.059
21	ATG36	YJL185C	91-96	DEIAIL	11.726

C. Appendix: Motif list from the bioinformatical SIMr detection approach

Table C.1. List of SIMr instances from the bioinformatical SIMr detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	bioinformatical motif score
22	PIF1	YML061C	64-69	EDLDLL	11.609
23	YML002W	YML002W	516-521	EEINMI	9.309
24	SET1	YHR119W	772-777	EDMLIL	8.719
25	YKL162C	YKL162C	394-399	SSLAVL	5.677
26	TAM41	YGR046W	64-69	SDLDLL	5.464
27	ULP2	YIL031W	929-934	SDVNLI	3.808
28	CST6	YIL036W	518-529	EELLMI	2.808
29	PSY4	YBL046W	162-167	EDVSLM	1.869
30	YIR024C	YIR024C	102-107	DELKLV	1.480
31	SET5	YHR207C	40-45	DDVVLL	0.946
32	NTG1	YAL015C	8-13	SSMAIL	0.923
33	PSO2	YMR137C	499-504	SSVHLV	0.000
34	CDC25	YLR310C	319-324	SELPLI	0.000
35	MSK1	YNL073W	429-434	SSLQIL	0.000
36	LSM8	YJR022W	59-64	SEIALV	0.000
37	CDC25	YLR310C	276-281	EEIEMI	0.000
38	ATG31	YDR022C	98-103	EDITLI	0.000
39	DNF3	YMR162C	621-626	SSIDVV	0.000
40	BUB3	YOR026W	15-20	SDIKII	-6.396
41	TEL1	YBL088C	1440-1445	DDVVLV	-7.832
42	IRA2	YOL081W	103-108	ESILII	-32.749
43	LAG2	YOL025W	498-503	SELSVL	-40.237
44	DCR2	YLR361C	187-192	SDINIL	-44.146
45	DOP1	YDR141C	830-835	DDLDML	-49.165
46	DOP1	YDR141C	693-698	ESMKLL	-49.882
47	SIR4	YDR227W	1031-1036	SDIIIL	-52.714
48	EAF5	YEL018W	6-11	SELVVL	-53.737
49	LAG2	YOL025W	171-176	ESLNLL	-58.756
50	YPR114W	YPR114W	100-105	SSVHLI	-65.019
51	DDI1	YER143W	68-73	DDLLLI	-73.246
52	SAM37	YMR060C	21-26	DSIALV	-77.422
53	NUP85	YJR042W	177-182	ESLTVL	-83.189

C. Appendix: Motif list from the bioinformatical SIMr detection approach

Table C.1. List of SIMr instances from the bioinformatical SIMr detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	bioinformatical motif score
54	BNR1	YIL159W	467-472	ESIKLL	-85.584
55	DYN3	YMR299C	234-239	DSIGLI	-138.595
56	YDR186C	YDR186C	32-37	DSLFLI	-216.419

D. Appendix: Motif list from the bioinformatical AIM detection approach

Table D.1. List of AIM instances from the bioinformatical AIM detection approach. A AIM profile search within a *Saccharomyces cerevisiae* protein sequence database with a sensitivity of $C = 3.0$ and further refinement by conservation criteria($t1 = 1.75$, $t2 = 1.55$, $t3 = 1.0$) gave this list of putative AIM instances. Evaluation of results was performed according to a combined 'bioinformatical motif score'. This score describes the profile adherence, the conservation and the structural context features of the instance.

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
1	DOS2	YDR068W	308-310	DWE	3.263	5704.649
2	NBP2	YDR162C	214-218	EWEDI	5.620	5516.317
3	SEC63	YOR254C	646-650	DYTDI	3.023	3357.059
4	FUN12	YAL035W	323-327	DWENL	4.701	3133.303
5	REI1	YBR267W	277-281	DWEDV	5.620	2836.780
6	CEG1	YGL130W	456-459	DWSD	3.982	1888.586
7	KAP114	YGL241W	923-927	DWEDV	5.620	1358.811
8	RPT1	YKL145W	6-10	DWEKY	3.462	1172.958
9	SCD5	YOR329C	4-8	DWLNV	3.982	1141.792
10	STB3	YDR169C	435-439	DWESI	4.701	935.862
11	SEM1	YDR363W_A	63-67	NWDDV	4.381	851.333
12	GCS1	YDL226C	348-352	KWDDF	3.103	775.180
13	MPE1	YKL059C	256-260	SWEDY	4.102	544.336
14	SPP41	YDR464W	129-133	EWAHI	3.303	503.329
15	SHY1	YGR112W	48-52	DWKPI	3.462	450.260
16	CDC123	YLR215C	83-87	DWEDD	3.343	422.515
17	MKS1	YNL076W	297-301	DWDSV	3.822	341.695
18	RAD2	YGR258C	371-375	EWEEV	4.701	332.595
19	YIL077C	YIL077C	246-250	SWEKI	3.662	331.688
20	YCR090C	YCR090C	157-161	EWYDY	3.742	330.139

D. Appendix: Motif list from the bioinformatical AIM detection approach

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
21	JJJ1	YNL227C	302-306	SWQTV	3.103	316.496
22	ATG3	YNR007C	269-273	DWEDL	5.660	314.174
23	EDE1	YBL047C	1287-1291	EWDEI	4.381	290.788
24	SDA1	YGR245C	574-578	EWVTM	3.103	285.343
25	NST1	YNL091W	102-106	HWESL	3.143	283.728
26	SPB4	YFL002C	581-585	DWKEI	4.581	263.204
27	SGT1	YOR057W	322-326	DWSKL	3.542	225.753
28	PRI1	YIR008C	273-277	KWNDI	3.862	224.631
29	HSV2	YGR223C	444-448	SWREL	3.423	223.755
30	SLP1	YOR154W	113-117	EWKKV	3.223	208.727
31	YFL042C	YFL042C	6-10	NWEPV	3.223	203.336
32	GPN2	YOR262W	329-333	EWENA	3.063	192.066
33	YIL077C	YIL077C	260-264	SWENI	4.142	186.175
34	FAA1	YOR317W	60-64	GWRDV	3.263	178.363
35	RAD18	YCR066W	119-123	SWIEL	3.782	174.454
36	CNE1	YAL058W	258-262	DWDDR	3.143	150.662
37	NOP4	YPL043W	228-232	RWEDY	3.462	144.039
38	REX4	YOL080C	6-10	NWQAL	3.023	140.640
39	GAT1	YFL021W	504-508	DWLNL	4.022	138.769
40	YBR197C	YBR197C	111-115	SWSEI	3.782	138.221
41	SEN54	YPL083C	31-35	DWSQL	3.742	130.479
42	EDC3	YEL015W	90-94	DWQDD	3.103	102.450
43	GCN4	YEL009C	119-123	EWTSL	3.423	91.987
44	REC107	YJR021C	301-305	PWEEL	3.502	85.403
45	GLT1	YDL171C	1663-1667	DWKEF	3.582	84.826
46	YML037C	YML037C	74-78	DWIDE	3.183	81.010
47	YGR015C	YGR015C	248-252	KWPDL	3.462	75.384
48	GLG1	YKR058W	592-596	DWEDS	3.822	66.036
49	RSC1	YGR056W	817-821	DYEDI	3.782	65.834
50	RAD34	YDR314C	53-57	DWEEV	4.981	64.745
51	CDC123	YLR215C	340-344	KWKEL	3.143	58.208
52	DSN1	YIR010W	361-365	DWEDE	3.622	57.479

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
53	PET8	YNL003C	130-134	SWQTL	3.143	57.268
54	UBC13	YDR092W	142-146	EWTKL	3.183	43.563
55	RNR1	YER070W	584-588	DWDTL	3.582	42.376
56	MRM1	YOR201C	48-52	QWETL	3.143	34.499
57	GAP1	YKR039W	51-55	KWQDF	3.462	31.362
58	FRK1	YPL141C	326-330	EWDKL	3.343	28.896
59	LCB5	YLR260W	558-562	DWERL	3.942	24.760
60	YPL107W	YPL107W	212-216	GWEDI	4.341	23.294
61	YOR1	YGR281W	538-542	EWEDY	4.621	22.608
62	RPH1	YER169W	100-104	QWKDL	3.942	20.621
63	PAA1	YDR071C	176-180	KWIDM	3.662	20.567
64	THR4	YCR053W	52-56	DWSKL	3.542	20.494
65	SGM1	YJR134C	360-364	NWDSI	3.462	20.450
66	MRD1	YPR112C	166-170	SWEKV	3.383	20.255
67	SGD1	YLR336C	566-570	SWEPL	3.103	20.042
68	AMS1	YGL156W	77-81	SWKSI	3.223	19.535
69	YAE1	YJR067C	4-8	TWDDV	3.862	18.282
70	REF2	YDR195W	440-444	DYMDI	3.263	18.103
71	GYP7	YDL234C	676-680	DWNDL	4.821	18.063
72	AIM9	YER080W	234-238	DWDPL	3.303	17.866
73	MPH2	YDL247W	31-35	SWIEM	3.423	17.209
74	MPH3	YJR160C	31-35	SWIEM	3.423	17.209
75	MSH6	YDR097C	560-564	DWPEV	3.982	16.896
76	BUD14	YAR014C	58-62	DYSDI	3.103	16.279
77	MAG2	YLR427W	489-493	DWRKI	3.662	14.192
78	COS9	YKL219W	333-337	GWDEI	3.103	11.765
79	NOC4	YPR144C	260-264	NWLSL	3.143	10.433
80	COS6	YGR295C	307-311	GWDEI	3.103	10.118
81	GSH1	YJL101C	579-583	DWKEL	4.341	9.688
82	COS8	YHL048W	307-311	GWDEI	3.103	9.552
83	SEY1	YOR165W	491-495	VWDDI	3.143	9.489
84	NPR3	YHL023C	164-168	DYLDI	3.103	8.427

D. Appendix: Motif list from the bioinformatical AIM detection approach

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
85	YCR016W	YCR016W	201-205	QWNDY	3.023	8.416
86	ARH1	YDR376W	458-462	DWERI	4.182	8.353
87	RLF2	YPR018W	591-595	NWENL	4.062	8.150
88	COS1	YNL336W	307-311	GWDEI	3.103	8.120
89	DSN1	YIR010W	473-477	EWQEL	4.501	7.875
90	ARP8	YOR141C	647-651	DWNSL	3.622	7.831
91	NCR1	YPL006W	67-71	EWKEV	4.022	7.155
92	COX19	YLL018C_A	73-77	EWSHL	3.143	6.919
93	TOP1	YOL006C	581-585	EWQKI	3.942	6.515
94	IOC2	YLR095C	27-31	DWKEY	3.582	6.406
95	GLG1	YKR058W	271-275	QWNEV	3.103	6.343
96	TOR2	YKL203C	1444-1448	EWEEL	4.741	6.227
97	MDY2	YOL111C	178-182	PWDDI	3.782	5.990
98	HXT16	YJR158W	555-559	SWKEV	3.502	5.635
99	HXT15	YDL245C	555-559	SWKEV	3.502	5.550
100	SLM5	YCR024C	319-323	RWEDL	4.222	5.431
101	SLA1	YBL007C	656-660	DWFEF	3.063	5.384
102	SFB3	YHR098C	847-851	NWQQV	3.502	5.329
103	PEP1	YBL017C	109-113	SWERI	3.383	5.284
104	YGR126W	YGR126W	13-17	EYEDI	3.502	5.278
105	NUP82	YJL061W	679-683	EWDEL	4.142	4.621
106	JJJ1	YNL227C	310-314	NWDEL	3.782	4.526
107	HOS1	YPR068C	90-94	KWSEL	3.143	4.006
108	FAR3	YMR052W	175-179	DWDRI	3.582	3.965
109	GIS1	YDR096W	49-53	EWLEL	4.062	3.789
110	SLX4	YLR135W	246-250	EWKDI	4.941	3.541
111	TRE2	YOR256C	693-697	GWENI	3.383	3.417
112	SWT1	YOR166C	398-402	EWKSL	3.502	2.565
113	UIP5	YKR044W	321-325	HWNDM	3.143	2.190
114	HUL5	YGL141W	450-454	EYEDL	3.263	1.989
115	YFL054C	YFL054C	203-207	DWKRL	3.263	1.731
116	HIR3	YJR140C	253-257	SWDEV	3.582	1.727

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
117	DRS1	YLL008W	617-621	DWVQI	4.222	1.635
118	GEX2	YKR106W	247-251	EWRSL	3.383	1.580
119	GEX1	YCL073C	247-251	EWRSL	3.383	1.557
120	MTC5	YDR128W	729-733	DWDDI	5.300	1.543
121	GLG2	YJL137C	361-365	DWEST	3.143	1.402
122	RAD4	YER162C	229-233	TWKEI	3.423	1.339
123	ETP1	YHL010C	48-52	DWQDW	3.742	1.138
124	UBP12	YJL197W	918-922	EWSEL	4.062	0.823
125	YGL262W	YGL262W	46-50	NWDNV	3.423	0.782
126	MVP1	YMR004W	495-499	EWEKL	3.942	0.781
127	IDH1	YNL037C	59-63	DWETI	4.421	0.770
128	VTH1	YIL173W	65-69	NWKTI	3.103	0.724
129	VTH2	YJL222W	65-69	NWKTI	3.103	0.724
130	TFC4	YGR047C	342-346	NWKKI	3.143	0.614
131	SPB4	YFL002C	6-10	EWDNL	3.822	0.600
132	PDR17	YNL264C	84-88	DWEKF	3.462	0.401
133	COS6	YGR295C	121-125	DWEVV	3.343	0.274
134	LST4	YKL176C	531-535	SWERL	3.143	0.266
135	COS4	YFL062W	121-125	DWEVV	3.343	0.260
136	YEL023C	YEL023C	664-668	SWTDI	4.341	0.225
137	GUP1	YGL084C	236-240	RWENL	3.263	0.109
138	UTP20	YBL004W	822-826	SWTEV	3.423	0.044
139	TIM54	YJL054W	207-211	DWRNV	3.862	0.019
140	YPR089W	YPR089W	48-52	SWVQL	3.183	0.009
141	LSM4	YER112W	34-38	NWMNL	3.542	0.000
142	ATG32	YIL146C	85-89	SWQAI	3.103	0.000
143	ASM4	YDL088C	524-528	GWNDL	3.263	0.000
144	YDL124W	YDL124W	114-118	DYVDL	3.103	0.000
145	ULS1	YOR191W	1302-1306	DWLRL	3.263	0.000
146	YOR338W	YOR338W	244-248	SWEKL	3.423	0.000
147	MSS11	YMR164C	136-140	EWWEI	3.462	0.000
148	YPR117W	YPR117W	1548-1552	DYIDL	3.103	0.000

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
149	CLD1	YGR110W	392-396	DWMDV	5.100	0.000
150	SIP1	YDR422C	556-560	NWFEV	3.143	0.000
151	YBR053C	YBR053C	40-44	LWVDI	3.223	0.000
152	HRB1	YNL004W	273-277	NWQAL	3.023	0.000
153	SSK2	YNR031C	388-392	EWQTM	3.303	0.000
154	ERG8	YMR220W	428-432	QWLDV	3.902	0.000
155	SLX4	YLR135W	734-738	EWADV	4.581	0.000
156	EAF3	YPR023C	235-239	DWEYV	3.502	0.000
157	YCS4	YLR272C	1143-1147	QWDDI	4.261	0.000
158	ADP1	YCR011C	689-693	EWAHL	3.063	0.000
159	YPR117W	YPR117W	1542-1546	DWYDY	4.022	0.000
160	YJR107W	YJR107W	308-312	DWLHV	3.383	0.000
161	ATG34	YOL083W	408-412	TWEEI	4.102	0.000
162	SHE3	YBR130C	247-251	SWLNL	3.223	0.000
163	NAT1	YDL040C	796-800	DWLNF	3.263	0.000
164	PHO87	YCR037C	396-400	TWKDM	3.462	0.000
165	KEX2	YNL238W	458-462	TWENV	3.502	0.000
166	PAF1	YBR279W	231-235	EWISM	3.383	0.000
167	YFH1	YDL120W	148-152	EWVSL	3.742	0.000
168	UGO1	YDR470C	99-103	EWAEL	3.982	0.000
169	TOM1	YDR457W	2230-2234	RWKDI	3.782	0.000
170	MTC4	YBR255W	195-199	HWEEL	3.702	0.000
171	TAF3	YPL011C	179-183	DWIKL	3.782	0.000
172	AIM41	YOR215C	157-161	DWKSL	3.782	0.000
173	DAD2	YKR083C	55-59	NWDSI	3.462	0.000
174	AUS1	YOR011W	1018-1022	DWSAL	3.223	0.000
175	ATG29	YPL166W	53-57	DWQNL	4.461	0.000
176	INO80	YGL150C	120-124	EWAEY	3.223	0.000
177	YPR022C	YPR022C	647-651	RWIDF	3.023	0.000
178	IFH1	YLR223C	815-819	DWYEV	4.102	0.000
179	UIP4	YPL186C	24-28	NWEGV	3.063	0.000
180	YPR022C	YPR022C	810-814	SWKDI	4.421	0.000

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
181	YAE1	YJR067C	123-127	KWTDI	3.942	0.000
182	SHE10	YGL228W	420-424	EWGDI	4.182	0.000
183	YPL245W	YPL245W	233-237	DWYTL	3.303	0.000
184	YPR097W	YPR097W	989-993	NWSDI	4.581	0.000
185	YBR053C	YBR053C	126-130	EWEYV	3.223	0.000
186	KRE28	YDR532C	227-231	NWQKL	3.343	0.000
187	GDB1	YPR184W	1311-1315	EWNQL	3.303	0.000
188	MRPL7	YDR237W	130-134	KWSNI	3.063	0.000
189	DYN1	YKR054C	3150-3154	NWRDI	4.461	0.000
190	HNT3	YOR258W	159-163	SWDDL	4.261	-0.008
191	FAT1	YBR041W	658-662	DWEAI	4.142	-0.061
192	PLB1	YMR008C	198-202	RWDDI	3.862	-0.064
193	GCD2	YGR083C	603-607	GWQEL	3.223	-0.399
194	YCS4	YLR272C	443-447	EWEEY	3.982	-0.521
195	FMP25	YLR077W	212-216	SWDDL	4.261	-0.948
196	LAS1	YKR063C	260-264	QWQEL	3.742	-1.271
197	GTO3	YMR251W	93-97	HWFDI	3.383	-1.468
198	ATG7	YHR171W	620-624	EWEDD	3.063	-1.537
199	NIP1	YMR309C	138-142	DWLTI	3.742	-1.627
200	GAC1	YOR178C	281-285	SWRDI	4.301	-2.063
201	SKI2	YLR398C	613-617	TWPEI	3.103	-2.126
202	CAT2	YML042W	302-306	QWREV	3.143	-2.375
203	RMP1	YLR145W	66-70	EWVKL	3.502	-2.543
204	FYV4	YHR059W	62-66	NWNNL	3.223	-2.638
205	YDR131C	YDR131C	249-253	EYIDI	3.063	-2.704
206	PSD2	YGR170W	78-82	EWLRI	3.223	-2.748
207	NPP1	YCR026C	429-433	IWEDL	3.303	-3.286
208	GSY1	YFR015C	62-66	DWEDE	3.622	-3.525
209	NAN1	YPL126W	192-196	EWHNV	3.662	-4.387
210	DAL2	YIR029W	296-300	TWVEL	3.423	-4.722
211	OST3	YOR085W	185-189	DWTPI	3.383	-4.866
212	SRP72	YPL210C	393-397	EWENA	3.063	-5.118

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
213	SYT1	YPR095C	1109-1113	KWEKL	3.023	-6.340
214	MAP2	YBL091C	101-105	HWNDV	3.462	-7.485
215	PMU1	YKL128C	36-40	NWKEL	3.702	-9.732
216	PET309	YLR067C	739-743	QWKDF	3.183	-9.744
217	NMD2	YHR077C	604-608	HWDDV	3.702	-9.761
218	DYN1	YKR054C	1444-1448	NWVEV	3.902	-10.552
219	PEP1	YBL017C	65-69	TWEKV	3.023	-10.765
220	YHR045W	YHR045W	223-227	TWKEL	3.183	-11.058
221	MTF1	YMR228W	223-227	EWDPI	3.263	-11.598
222	YMR209C	YMR209C	415-419	GWKDL	3.423	-13.083
223	MOT2	YER068W	465-469	SWDKI	3.063	-13.196
224	YMR262W	YMR262W	40-44	DWNNL	3.862	-13.862
225	NUD1	YOR373W	475-479	DWEKI	4.461	-15.381
226	DBR1	YKL149C	375-379	DWENY	3.942	-15.427
227	YAR023C	YAR023C	74-78	EWKTI	3.462	-16.590
228	IRR1	YIL026C	870-874	KWREI	3.263	-16.714
229	TOK1	YJL093C	611-615	DWSYI	3.103	-16.744
230	TRS120	YDR407C	1017-1021	DWIEY	3.822	-17.625
231	PPS1	YBR276C	342-346	DWHNY	3.223	-17.824
232	MYO3	YKL129C	878-882	DWMGV	3.183	-18.571
233	COS1	YNL336W	121-125	DWEVV	3.343	-18.933
234	COS5	YJR161C	121-125	DWEVV	3.343	-19.111
235	OST1	YJL002C	207-211	PWEDI	4.381	-19.505
236	TMA64	YDR117C	551-555	KWIDF	3.263	-20.214
237	MEI4	YER044C_A	355-359	EWQHL	3.582	-20.394
238	VPS41	YDR080W	581-585	EWADI	4.861	-20.476
239	COS8	YHL048W	121-125	DWEVV	3.343	-21.463
240	COS12	YGL263W	309-313	NWSQI	3.343	-21.677
241	PBA1	YLR199C	5-9	QWNDL	3.782	-21.934
242	SCC4	YER147C	155-159	NWASV	3.023	-22.380
243	COS2	YBR302C	121-125	DWEVV	3.343	-22.805
244	COS3	YML132W	121-125	DWEVV	3.343	-22.805

D. Appendix: Motif list from the bioinformatical AIM detection approach

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
245	HPR1	YDR138W	180-184	HWNDI	3.742	-22.936
246	PRM8	YGL053W	125-129	EWRTI	3.343	-23.146
247	UBP8	YMR223W	117-121	YWDDV	3.063	-23.161
248	SAD1	YFR005C	417-421	KWIEI	3.622	-23.308
249	SLN1	YIL147C	312-316	NWVAI	3.063	-23.469
250	ALG3	YBL082C	25-29	LWQDL	3.183	-23.693
251	FMP27	YLR454W	2467-2471	EWLSL	3.502	-23.726
252	PEP1	YBL017C	1192-1196	TWKDY	3.063	-23.791
253	ORC3	YLL004W	493-497	SWEQV	3.582	-24.144
254	SXM1	YDR395W	222-226	SWVQL	3.183	-24.568
255	COS9	YKL219W	149-153	DWETV	4.142	-24.677
256	CTF4	YPR135W	565-569	NWTKI	3.063	-25.942
257	SEC1	YDR164C	309-313	DWIDL	5.220	-26.028
258	PPE1	YHR075C	76-80	TWSDF	3.063	-26.475
259	AST1	YBL069W	402-406	DWKDH	3.103	-26.537
260	YMR114C	YMR114C	132-136	EWKTV	3.183	-26.588
261	ATX2	YOR079C	140-144	SWKDI	4.421	-26.998
262	SYS1	YJL004C	68-72	SWENI	4.142	-27.042
263	ATE1	YGL017W	208-212	SWEQL	3.622	-29.566
264	RSC30	YHR056C	610-614	EWKDT	3.383	-29.884
265	EXO84	YBR102C	353-357	NWMEL	3.862	-29.926
266	COG2	YGR120C	222-226	DYQDL	3.303	-30.303
267	YDR444W	YDR444W	419-423	DWRSI	3.902	-30.331
268	ATG7	YHR171W	170-174	KWFDV	3.223	-31.024
269	TRP1	YDR007W	156-160	DWNSI	3.862	-31.404
270	COS12	YGL263W	118-122	DWERV	3.902	-31.778
271	UTP20	YBL004W	1257-1261	SWSDI	4.421	-31.967
272	VTH1	YIL173W	1168-1172	TWKDY	3.063	-32.728
273	VTH2	YJL222W	1168-1172	TWKDY	3.063	-32.728
274	YNR065C	YNR065C	735-739	TWKDY	3.063	-32.728
275	CLF1	YLR117C	547-551	SWIEF	3.023	-33.547
276	YCL049C	YCL049C	96-100	DWEQY	3.662	-33.812

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
277	NDJ1	YOL104C	210-214	QWSEL	3.303	-33.892
278	YOR389W	YOR389W	553-557	GWIEL	3.023	-33.993
279	RKM1	YPL208W	147-151	EWFEL	3.542	-34.541
280	STE6	YKL209C	95-99	SWMHI	3.023	-35.585
281	TEL1	YBL088C	1918-1922	EWMDT	3.542	-35.911
282	YOL075C	YOL075C	599-603	NWITV	3.063	-35.914
283	MAK5	YBR142W	171-175	EWTNL	3.662	-36.533
284	YCR015C	YCR015C	303-307	SWDSL	3.063	-38.884
285	RIC1	YLR039C	482-486	LWEEI	3.023	-39.673
286	POL4	YCR014C	507-511	KWDEL	3.223	-40.416
287	RSC4	YKR008W	480-484	NWVEY	3.183	-40.829
288	IDS2	YJL146W	456-460	KWLDL	3.782	-42.426
289	ATG19	YOL082W	411-415	TWEEL	3.862	-43.803
290	KIP3	YGL216W	605-609	DWDET	3.103	-44.308
291	YGR026W	YGR026W	117-121	TWNDL	3.662	-44.788
292	AMD2	YDR242W	414-418	EWWDL	3.862	-46.659
293	TRM732	YMR259C	234-238	EWIQL	3.702	-46.759
294	AXL2	YIL140W	298-302	DWVAL	3.462	-46.861
295	ECM32	YER176W	584-588	SWNNL	3.063	-49.772
296	SEC16	YPL085W	1230-1234	NWKSI	3.383	-50.059
297	ALG14	YBR070C	218-222	QWQEL	3.742	-50.372
298	RIM21	YNL294C	76-80	DWQKF	3.223	-51.217
299	CWC22	YGR278W	17-21	NWEMI	3.023	-51.821
300	PGM3	YMR278W	342-346	EWRQL	3.343	-53.110
301	TRA1	YHR099W	1832-1836	KWLEL	3.143	-54.429
302	HPR1	YDR138W	47-51	EWEPL	3.622	-55.673
303	MYO5	YMR109W	878-882	DWVAI	3.702	-56.675
304	UBX7	YBR273C	244-248	KWVDV	3.982	-57.611
305	DYN1	YKR054C	1606-1610	EWLNI	3.982	-57.928
306	KTR3	YBR205W	159-163	SWIDT	3.103	-58.713
307	EMC3	YKL207W	49-53	EWQYL	3.023	-62.297
308	TDA9	YML081W	1046-1050	DWNSM	3.263	-64.111

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
309	FMP48	YGR052W	277-281	IWQDL	3.063	-64.176
310	NSI1	YDR026C	521-525	DWNSI	3.862	-64.184
311	THO2	YNL139C	572-576	KWIDY	3.263	-64.719
312	AVT3	YKL146W	683-687	SWQTI	3.383	-65.692
313	RTC1	YOL138C	535-539	TWRDL	3.702	-66.069
314	YIF1	YNL263C	132-136	NWQRI	3.303	-67.596
315	YGR109W-B	YGR109W_B	1168-1172	RWLDI	3.782	-68.112
316	YIL082W-A	YIL082W_A	1194-1198	RWLDI	3.782	-68.112
317	YSP1	YHR155W	535-539	NWNDL	4.182	-70.542
318	DCG1	YIR030C	134-138	EWIPI	3.423	-70.639
319	AGC1	YPR021C	218-222	DWNDF	4.062	-71.133
320	STT4	YLR305C	66-70	EWEVL	3.103	-71.320
321	CWC23	YGL128C	145-149	DWKHL	3.423	-72.386
322	IRC22	YEL001C	81-85	NWEDT	3.702	-74.191
323	TYR1	YBR166C	395-399	EWSSV	3.462	-76.009
324	TPP1	YMR156C	56-60	DWQFI	3.023	-78.015
325	PDR11	YIL013C	1028-1032	NWAEL	3.622	-79.654
326	NUP133	YKR082W	787-791	DWNHV	3.223	-80.052
327	CLB5	YPR120C	144-148	GWQDL	3.862	-81.854
328	COS10	YNR075W	114-118	DWDAV	3.263	-82.559
329	BNA2	YJR078W	51-55	KWEEI	4.062	-83.878
330	NUP170	YBL079W	1339-1343	KWDEL	3.223	-85.525
331	COS4	YFL062W	305-309	GWDEI	3.103	-87.953
332	CLF1	YLR117C	486-490	DWDRV	3.303	-88.049
333	VPS30	YPL120W	352-356	PWKEI	3.063	-88.292
334	COS2	YBR302C	305-309	GWDEI	3.103	-88.935
335	COS3	YML132W	305-309	GWDEI	3.103	-88.935
336	RAD30	YDR419W	414-418	SWLEV	3.502	-89.088
337	COS7	YDL248W	307-311	GWDEI	3.103	-89.905
338	CCE1	YKL011C	97-101	DWQKI	4.222	-91.635
339	TFC3	YAL001C	1152-1156	NWYSI	3.183	-92.191
340	YHR045W	YHR045W	198-202	DWYKL	3.343	-93.316

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
341	VRG4	YGL225W	237-241	DWSSV	3.742	-93.398
342	SLS1	YLR139C	339-343	DWARL	3.183	-94.700
343	ATG2	YNL242W	747-751	RWLEI	3.143	-95.308
344	COS5	YJR161C	307-311	GWDEI	3.103	-95.668
345	REV1	YOR346W	953-957	TWERI	3.023	-98.843
346	RKM2	YDR198C	201-205	DWETI	4.421	-100.354
347	KRE5	YOR336W	457-461	NWSEI	3.942	-100.693
348	IRR1	YIL026C	713-717	DWISI	4.261	-101.942
349	PPA2	YMR267W	262-266	SWKNL	3.223	-103.538
350	DYN1	YKR054C	1451-1455	YWLDL	3.023	-110.601
351	PAM17	YKR065C	50-54	TWSDF	3.063	-113.041
352	CUE3	YGL110C	213-217	NWIEI	4.182	-117.495
353	PRP28	YDR243C	173-177	NWEEL	4.381	-119.125
354	YPL247C	YPL247C	250-254	DWNTV	3.303	-120.634
355	ECM30	YLR436C	996-1000	TWANI	3.023	-121.061
356	YNL115C	YNL115C	262-266	EWVSI	3.982	-122.161
357	RRG7	YOR305W	226-230	EWLKL	3.263	-123.210
358	YLR460C	YLR460C	48-52	DWAHI	3.582	-129.541
359	YLR352W	YLR352W	450-454	SWQQI	3.622	-133.586
360	UBP5	YER144C	273-277	NWVKL	3.143	-135.314
361	SUE1	YPR151C	174-178	EWKNV	3.702	-136.638
362	YCR102C	YCR102C	40-44	DWAHI	3.582	-137.050
363	SKI3	YPR189W	290-294	EWTDY	3.862	-137.894
364	RRP1	YDR087C	180-184	EWERL	3.662	-139.173
365	YDL199C	YDL199C	482-486	SWDSI	3.303	-141.601
366	IZH2	YOL002C	41-45	SWDEI	3.862	-142.618
367	HBS1	YKR084C	347-351	NWVPI	3.063	-145.009
368	GLG2	YJL137C	215-219	EWIRL	3.223	-145.272
369	MSH5	YDL154W	505-509	QWEEI	4.222	-150.221
370	ARO80	YDR421W	353-357	KWSDY	3.023	-152.662
371	TEP1	YNL128W	406-410	SWQEL	3.982	-156.325
372	CHS6	YJL099W	373-377	SWYNL	3.023	-160.070

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
373	YJR061W	YJR061W	158-162	DWANF	3.183	-161.669
374	ECM5	YMR176W	1017-1021	EWLSV	3.462	-163.855
375	MSN5	YDR335W	138-142	SWVDM	4.062	-169.272
376	SLM3	YDL033C	79-83	DWRDV	4.821	-171.488
377	YLR063W	YLR063W	200-204	DWSTV	3.462	-172.996
378	FAD1	YDL045C	248-252	EWEII	3.143	-175.092
379	YGL082W	YGL082W	243-247	NWQSL	3.582	-175.441
380	ARP10	YDR106W	270-274	DWFDY	3.702	-177.873
381	SRB8	YCR081W	207-211	DWTDT	3.582	-179.172
382	YOR114W	YOR114W	71-75	EWDSM	3.223	-179.494
383	DUR1,2	YBR208C	19-23	DWIDF	4.461	-181.691
384	VPS16	YPL045W	7-11	DWERL	3.942	-187.800
385	CSF1	YLR087C	515-519	GWMDL	3.582	-190.712
386	FUI1	YBL042C	609-613	EWVEV	4.261	-191.255
387	SIA1	YOR137C	238-242	SWKEV	3.502	-194.500
388	ARC1	YGL105W	100-104	RWIDY	3.023	-194.611
389	YPR147C	YPR147C	58-62	DWEIL	3.183	-198.787
390	MNS1	YJR131W	49-53	SWRDY	3.303	-201.455
391	YPL071C	YPL071C	109-113	KWVDY	3.263	-202.141
392	OSH3	YHR073W	305-309	SWVDA	3.063	-204.340
393	DLD1	YDL174C	227-231	PWEDL	4.142	-212.062
394	YMR252C	YMR252C	43-47	KWHEL	3.103	-214.211
395	APC1	YNL172W	663-667	GWPDL	3.103	-214.705
396	CFT1	YDR301W	1339-1343	AWRDI	3.542	-221.190
397	FMP27	YLR454W	1795-1799	SWLDI	4.421	-221.747
398	BEM4	YPL161C	2-6	DYEEI	3.143	-221.996
399	REB1	YBR049C	752-756	DWDEL	4.421	-229.249
400	TPS3	YMR261C	558-562	EWRQL	3.343	-234.956
401	VPS8	YAL002W	422-426	SWSDI	4.421	-246.838
402	YIL151C	YIL151C	923-927	HWEKI	3.143	-249.698
403	IFM1	YOL023W	394-398	GWKDV	3.383	-263.129
404	IRC20	YLR247C	666-670	EWDNY	3.063	-279.489

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
405	IRC21	YMR073C	63-67	DWHSL	3.742	-281.803
406	YFR006W	YFR006W	233-237	DWYEI	4.381	-285.993
407	CTK3	YML112W	126-130	DWKSL	3.782	-287.449
408	HRT3	YLR097C	122-126	SWVNL	3.462	-291.531
409	ARC1	YGL105W	30-34	QWESV	3.383	-294.913
410	CDC123	YLR215C	280-284	SWNEI	3.622	-295.142
411	MTQ2	YDR140W	143-147	QWLDL	3.942	-298.684
412	YCL068C	YCL068C	108-112	DWYRL	3.063	-299.928
413	MID1	YNL291C	63-67	EWTPI	3.103	-300.081
414	LEU5	YHR002W	264-268	TWAEL	3.103	-300.735
415	YLR460C	YLR460C	226-230	DYHDI	3.063	-305.640
416	YDR333C	YDR333C	308-312	DWKDV	4.941	-308.301
417	YLR352W	YLR352W	302-306	EWLNV	3.702	-311.500
418	OST1	YJL002C	30-34	TWENV	3.502	-321.469
419	BUD5	YCR038C	142-146	DWYRL	3.063	-328.869
420	IRA1	YBR140C	2553-2557	SWSEL	3.542	-331.081
421	MDN1	YLR106C	3353-3357	EWEKY	3.183	-338.653
422	YCR102C	YCR102C	218-222	DYHDI	3.063	-340.267
423	YNL011C	YNL011C	89-93	EWNEI	4.142	-347.969
424	YJL016W	YJL016W	195-199	DWCEV	3.542	-349.187
425	YPL034W	YPL034W	139-143	EWFNV	3.183	-350.662
426	RRT6	YGL146C	198-202	SWSSI	3.223	-358.945
427	MCD4	YKL165C	751-755	QWIEI	3.782	-367.172
428	LST4	YKL176C	317-321	NWIEI	4.182	-370.937
429	NUP85	YJR042W	305-309	EWKNL	3.742	-401.862
430	RFC1	YOR217W	741-745	DWDSI	4.102	-433.948
431	DPB2	YPR175W	165-169	DWRDY	4.102	-444.270
432	DUO1	YGL061C	109-113	SWINI	3.702	-448.056
433	YBR225W	YBR225W	197-201	DWNTL	3.343	-453.700
434	IRA2	YOL081W	2541-2545	SWSEL	3.542	-479.613
435	SPT14	YPL175W	363-367	DWMDV	5.100	-481.155
436	PHO90	YJL198W	350-354	TWKDM	3.462	-484.820

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
437	LRO1	YNR008W	124-128	DWKDV	4.941	-491.295
438	ICT1	YLR099C	343-347	DWMDK	3.103	-506.525
439	VPS13	YLL040C	1912-1916	DWRSI	3.902	-513.062
440	PDR15	YDR406W	885-889	HWRDL	3.542	-517.839
441	CTS2	YDR371W	122-126	SWSDL	4.182	-518.804
442	ROT2	YBR229C	334-338	TWVDI	4.301	-528.063
443	BOP2	YLR267W	516-520	SWNEL	3.383	-538.468
444	YFR039C	YFR039C	393-397	EWGDV	3.902	-540.384
445	JID1	YPR061C	189-193	TWEDA	3.143	-542.945
446	YTA12	YMR089C	207-211	TWQDF	3.502	-554.255
447	PCH2	YBR186W	262-266	QWESL	3.423	-555.380
448	AUS1	YOR011W	752-756	SWKNI	3.462	-578.721
449	YLR118C	YLR118C	70-74	AWFDI	3.143	-594.046
450	PEX1	YKL197C	119-123	DWEII	3.423	-629.397
451	UTP9	YHR196W	312-316	SWLNV	3.183	-670.517
452	SLH1	YGR271W	66-70	DWDDI	5.300	-671.099
453	YNL134C	YNL134C	48-52	DWKHI	3.662	-676.269
454	PFS1	YHR185C	183-187	DWEET	3.702	-677.442
455	PDR11	YIL013C	752-756	SWKNI	3.462	-696.409
456	RAD1	YPL022W	470-474	KWEQL	3.223	-702.167
457	CAP2	YIL034C	155-159	NWDSI	3.462	-703.493
458	MET13	YGL125W	413-417	PWSDI	3.702	-760.413
459	MDM12	YOL009C	260-264	SWINL	3.462	-773.153
460	LTE1	YAL024C	1214-1218	DWKDL	4.981	-786.283
461	IML1	YJR138W	1053-1057	NWNQI	3.183	-801.179
462	LOS1	YKL205W	239-243	SWIDI	4.661	-848.876
463	HSL7	YBR133C	107-111	SWLEL	3.542	-854.713
464	SCP160	YJL080C	708-712	KWADI	3.942	-860.874
465	SKI2	YLR398C	301-305	EWAHV	3.023	-864.628
466	SAL1	YNL083W	180-184	QWRDF	3.063	-866.843
467	YHR202W	YHR202W	68-72	DWGDF	3.462	-913.337
468	KEX2	YNL238W	406-410	TWRDV	3.662	-959.171

Table D.1. List of AIM instances from the bioinformatical AIM detection approach (continued).

Rank	Gene	ORF	amino acid coordinates	sequence	profile score	bioinformatical motif score
469	SNT2	YGL131C	773-777	EWELV	3.023	-991.562
470	GPI18	YBR004C	60-64	SWDSV	3.023	-1029.188
471	RKM5	YLR137W	255-259	DWEKI	4.461	-1029.844
472	ATG4	YNL223W	423-427	DYVDI	3.343	-1072.129
473	TAF2	YCR042C	841-845	EWIRI	3.462	-1093.851
474	SHU2	YDR078C	213-217	EWLNL	3.742	-1116.865
475	YHI9	YHR029C	46-50	NWTNL	3.303	-1156.201
476	LPP1	YDR503C	236-240	HWYDV	3.423	-1258.567
477	NEW1	YPL226W	523-527	DWKRL	3.263	-1286.558
478	CAF130	YGR134W	706-710	DWRDV	4.821	-1290.481
479	NUP133	YKR082W	442-446	KWEDI	4.701	-1358.503
480	GCY1	YOR120W	104-108	DYVDL	3.103	-1383.343
481	YCK3	YER123W	331-335	DWMDL	5.140	-1438.413
482	UMP1	YBR173C	118-122	DWEDV	5.620	-1504.019
483	PTP3	YER075C	759-763	NWPDL	4.022	-1591.449
484	CHO2	YGR157W	689-693	DWIGL	3.303	-1676.704
485	CHS6	YJL099W	615-619	EWELL	3.063	-1738.280
486	NUP192	YJL039C	1305-1309	SWVQL	3.183	-1779.828
487	RAD30	YDR419W	232-236	DWDDV	5.021	-1796.379
488	BCH2	YKR027W	616-620	EWELL	3.063	-1893.714
489	YGR266W	YGR266W	598-602	NWSDI	4.581	-2017.565
490	CDC39	YCR093W	1578-1582	EWVKL	3.502	-2354.242

References

[SGD, 2013] (2013). The *Saccharomyces* Genome Database.

[Alemu et al., 2012] Alemu, E. A., Lamark, T., Torgersen, K. M., Birgisdottir, A. B., Larsen, K. B., Jain, A., Olsvik, H., Overvatn, A., Kirkin, V. and Johansen, T. (2012). ATG8 family proteins act as scaffolds for assembly of the ULK complex: sequence requirements for LC3-interacting region (LIR) motifs. J Biol Chem $\underline{287}$, 39275–39290.

[Altschul et al., 1989] Altschul, S., Carroll, R. and Lipman, D. (1989). Weights for data related by a tree. Journal of molecular biology $\underline{207}$, 647–653.

[Andrews et al., 2005] Andrews, E., Palecek, J., Sergeant, J., Taylor, E., Lehmann, A. and Watts, F. (2005). Nse2, a component of the Smc5-6 complex, is a SUMO ligase required for the response to DNA damage. Molecular and cellular biology $\underline{25}$, 185–196.

[Aoki et al., 2011] Aoki, Y., Kanki, T., Hirota, Y., Kurihara, Y., Saigusa, T., Uchiumi, T. and Kang, D. (2011). Phosphorylation of Serine 114 on Atg32 mediates mitophagy. Mol Biol Cell $\underline{22}$, 3206–3217.

[Armstrong et al., 2012] Armstrong, A. A., Mohideen, F. and Lima, C. D. (2012). Recognition of SUMO-modified PCNA requires tandem receptor motifs in Srs2. Nature $\underline{483}$, 59–63.

[Azad et al., 2012] Azad, G. K., Balkrishna, S. J., Sathish, N., Kumar, S. and Tomar, R. S. (2012). Multifunctional Ebselen drug functions through the activation of DNA damage response and alterations in nuclear proteins. Biochemical Pharmacology $\underline{83}$, 296–303.

[Baba et al., 1997] Baba, M., Osumi, M., Scott, S., Klionsky, D. and Ohsumi, Y. (1997). Two distinct pathways for targeting proteins from the cytoplasm to the vacuole/lysosome. The Journal of cell biology $\underline{139}$, 1687–1695.

[Baba et al., 1994] Baba, M., Takeshige, K., Baba, N. and Ohsumi, Y. (1994). Ultrastructural analysis of the autophagic process in yeast: detection of autophagosomes and their characterization. The Journal of cell biology $\underline{124}$, 903–913.

References

[Bachant et al., 2002] Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N. and Elledge, S. J. (2002). The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modification of DNA topoisomerase II. Molecular Cell 9, 1169–1182.

[Bacon and Anderson, 1986] Bacon, D. J. and Anderson, W. F. (1986). Multiple sequence alignment. Journal of Molecular Biology 191, 153–161.

[Bailly et al., 1997] Bailly, V., Lauder, S., Prakash, S. and Prakash, L. (1997). Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. The Journal of biological chemistry 272, 23360–23365.

[Bairoch, 1991] Bairoch, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. Nucleic acids research 19 Suppl, 2241–2245.

[Baldi et al., 1994] Baldi, P., Chauvin, Y., Hunkapiller, T. and McClure, M. A. (1994). Hidden Markov models of biological primary sequence information. Proceedings of the National Academy of Sciences 91, 1059–1063.

[Barlow et al., 1994] Barlow, P., Luisi, B., Milner, A., Elliott, M. and Everett, R. (1994). Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. Journal of molecular biology 237, 201–211.

[Bateman et al., 2000] Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. and Sonnhammer, E. L. L. (2000). The Pfam protein families database. Nucleic Acids Research 28, 263–266.

[Bateman et al., 2004] Bateman, A., Coin, L., Durbin, R., Finn, R., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D., Yeats, C. and Eddy, S. (2004). The Pfam protein families database. Nucleic acids research 32.

[Baudin et al., 1993] Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucleic acids research 21, 3329–3330.

[Bayer et al., 1998] Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. and Becker, J. (1998). Structure determination of the small ubiquitin-related modifier SUMO-1. Journal of molecular biology 280, 275–286.

References

[Beckwith and McAlear, 2000] Beckwith, W. and McAlear, M. A. (2000). Allele-specific interactions between the yeast RFC1 and RFC5 genes suggest a basis for RFC subunit-subunit interactions. Mol Gen Genet <u>264</u>, 378–391.

[Beckwith et al., 1998] Beckwith, W. H., Sun, Q., Bosso, R., Gerik, K. J., Burgers, P. M. and McAlear, M. A. (1998). Destabilized PCNA trimers suppress defective Rfc1 proteins in vivo and in vitro. Biochemistry <u>37</u>, 3711–3722.

[Bedford et al., 2011] Bedford, L., Lowe, J., Dick, L., Mayer, R. and Brownell, J. (2011). Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. Nature reviews. Drug discovery <u>10</u>, 29–46.

[Behrends et al., 2010] Behrends, C., Sowa, M., Gygi, S. and Harper, J. (2010). Network organization of the human autophagy system. Nature <u>466</u>, 68–76.

[Bellaoui et al., 2003] Bellaoui, M., Chang, M., Ou, J., Xu, H., Boone, C. and Brown, G. W. (2003). Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. EMBO J <u>22</u>, 4304–4313.

[Ben-Aroya et al., 2003] Ben-Aroya, S., Koren, A., Liefshitz, B., Steinlauf, R. and Kupiec, M. (2003). ELG1, a yeast gene required for genome stability, forms a complex related to replication factor C. Proc Natl Acad Sci U S A <u>100</u>, 9906–9911.

[Beranek, 1990] Beranek, D. T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat Res <u>231</u>, 11–30.

[Berger and Munson, 1991] Berger, M. and Munson, P. (1991). A novel randomized iterative strategy for aligning multiple protein sequences. Computer applications in the biosciences : CABIOS <u>7</u>, 479–484.

[Berglund et al., 2008] Berglund, A.-C., Sjölund, E., Ostlund, G. and Sonnhammer, E. L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. Nucleic acids research <u>36</u>.

[Bermudez et al., 2003] Bermudez, V. P., Lindsey-Boltz, L. A., Cesare, A. J., Maniwa, Y., Griffith, J. D., Hurwitz, J. and Sancar, A. (2003). Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex in vitro. Proc Natl Acad Sci U S A <u>100</u>, 1633–1638.

[Bernier-Villamor et al., 2002] Bernier-Villamor, V., Sampson, D., Matunis, M. and Lima, C. (2002). Structural basis for E2-mediated SUMO conjugation revealed by

a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. Cell 108, 345–356.

[Bernstein et al., 2006] Bernstein, K. A., Granneman, S., Lee, A. V., Manickam, S. and Baserga, S. J. (2006). Comprehensive mutational analysis of yeast DEXD/H box RNA helicases involved in large ribosomal subunit biogenesis. Mol Cell Biol 26, 1195–1208.

[Bienko et al., 2005] Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A. R., Hofmann, K. and Dikic, I. (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. Science 310, 1821–1824.

[Biggins et al., 2001] Biggins, S., Bhalla, N., Chang, A. and Smith, D. L. (2001). Genes involved in sister chromatid separation and segregation in the budding yeast Saccharomyces cerevisiae. Genetics 159, 453–470.

[Birnboim and Doly, 1979] Birnboim, H. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic acids research 7, 1513–1523.

[Bjørkøy et al., 2005] Bjørkøy, G., Lamark, T., Brech, A., Outzen, H., Perander, M., Overvatn, A., Stenmark, H. and Johansen, T. (2005). p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. The Journal of cell biology 171, 603–614.

[Borden et al., 1995] Borden, K., Boddy, M., Lally, J., O'Reilly, N., Martin, S., Howe, K., Solomon, E. and Freemont, P. (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. The EMBO journal 14, 1532–1541.

[Boutselakis et al., 2003] Boutselakis, H., Dimitropoulos, D., Fillon, J., Golovin, A., Henrick, K., Hussain, A., Ionides, J., John, M., Keller, P. A., Krissinel, E., McNeil, P., Naim, A., Newman, R., Oldfield, T., Pineda, J., Rachedi, A., Copeland, J., Sitnov, A., Sobhany, S., Suarez-Uruena, A., Swaminathan, J., Tagari, M., Tate, J., Tromm, S., Velankar, S. and Vranken, W. (2003). E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. Nucleic Acids Res 31, 458–462.

[Bowman et al., 2004] Bowman, G. D., O'Donnell, M. and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature 429, 724–730.

References

[Brennan and Struhl, 1980] Brennan, M. B. and Struhl, K. (1980). Mechanisms of increasing expression of a yeast gene in Escherichia coli. J Mol Biol <u>136</u>, 333–338.

[Burger et al., 2000] Burger, F., Daugeron, M. C. and Linder, P. (2000). Dbp10p, a putative RNA helicase from Saccharomyces cerevisiae, is required for ribosome biogenesis. Nucleic Acids Res <u>28</u>, 2315–2323.

[Burkovics et al., 2013] Burkovics, P., Sebesta, M., Sisakova, A., Plault, N., Szukacsov, V., Robert, T., Pinter, L., Marini, V., Kolesar, P., Haracska, L., Gangloff, S. and Krejci, L. (2013). Srs2 mediates PCNA-SUMO-dependent inhibition of DNA repair synthesis. The EMBO Journal <u>32</u>, 742–755.

[Burroughs et al., 2007] Burroughs, A. M., Balaji, S., Iyer, L. M. and Aravind, L. (2007). Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. Biol Direct <u>2</u>, 18.

[Burschowsky et al., 2011] Burschowsky, D., Rudolf, F., Rabut, G., Herrmann, T., Peter, M. and Wider, G. (2011). Structural analysis of the conserved ubiquitin-binding motifs (UBMs) of the translesion polymerase iota in complex with ubiquitin. J Biol Chem <u>286</u>, 1364–1373.

[Bylebyl et al., 2003] Bylebyl, G., Belichenko, I. and Johnson, E. (2003). The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast. The Journal of biological chemistry <u>278</u>, 44113–44120.

[Cal-Bakowska et al., 2011] Cal-Bakowska, M., Litwin, I., Bocer, T., Wysocki, R. and Dziadkowiec, D. (2011). The Swi2-Snf2-like protein Uls1 is involved in replication stress response. Nucleic Acids Res <u>39</u>, 8765–8777.

[Caldecott, 2008] Caldecott, K. (2008). Single-strand break repair and genetic disease. Nature reviews. Genetics <u>9</u>, 619–631.

[Carlile et al., 2009] Carlile, C. M., Pickart, C. M., Matunis, M. J. and Cohen, R. E. (2009). Synthesis of free and proliferating cell nuclear antigen-bound polyubiquitin chains by the RING E3 ubiquitin ligase Rad5. J Biol Chem <u>284</u>, 29326–29334.

[Carrillo and Lipman, 1988] Carrillo, H. and Lipman, D. (1988). The multiple sequence alignment problem in biology. SIAM Journal on Applied Mathematics <u>48</u>, 1073–1082.

[Champoux, 2001] Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem <u>70</u>, 369–413.

References

[Chang and Huang, 2007] Chang, C.-Y. and Huang, W.-P. (2007). Atg19 mediates a dual interaction cargo sorting mechanism in selective autophagy. Molecular biology of the cell **18**, 919–929.

[Cheng et al., 2006] 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 & development **20**, 2067–2081.

[Cheong et al., 2008] Cheong, H., Nair, U., Geng, J. and Klionsky, D. J. (2008). The Atg1 kinase complex is involved in the regulation of protein recruitment to initiate sequestering vesicle formation for nonspecific autophagy in Saccharomyces cerevisiae. Mol Biol Cell **19**, 668–681.

[Chien et al., 1991] Chien, C., Bartel, P., Sternglanz, R. and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proceedings of the National Academy of Sciences of the United States of America **88**, 9578–9582.

[Chou and Fasman, 1974] Chou, P. and Fasman, G. (1974). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry **13**, 211–222.

[Chou and Fasman, 1978] Chou, P. and Fasman, G. (1978). Empirical predictions of protein conformation. Annual review of biochemistry **47**, 251–276.

[Chung et al., 1997] Chung, C. D., Liao, J., Liu, B., Rao, X., Jay, P., Berta, P. and Shuai, K. (1997). Specific inhibition of Stat3 signal transduction by PIAS3. Science (New York, N.Y.) **278**, 1803–1805.

[Clarkson, 1983] Clarkson, K. L. (1983). A modification of the greedy algorithm for vertex cover. Information Processing Letters **16**, 23–25.

[Cregg and Russell, 1998] Cregg, J. M. and Russell, K. A. (1998). Transformation. Methods Mol Biol **103**, 27–39.

[Crooks et al., 2004] Crooks, G. E., Hon, G., Chandonia, J.-M. and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. Genome Research **14**, 1188–1190.

[Cullmann et al., 1995] Cullmann, G., Fien, K., Kobayashi, R. and Stillman, B. (1995). Characterization of the five replication factor C genes of Saccharomyces cerevisiae. Molecular and Cellular Biology **15**, 4661–4671.

References

[D'Arpa and Liu, 1995] D'Arpa, P. and Liu, L. F. (1995). Cell cycle-specific and transcription-related phosphorylation of mammalian topoisomerase I. Exp Cell Res 217, 125–131.

[Davey et al., 2012] Davey, N., Van Roey, K., Weatheritt, R., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H. and Gibson, T. (2012). Attributes of short linear motifs. Molecular bioSystems 8, 268–281.

[David et al., 2002] David, G., Neptune, M. and DePinho, R. (2002). SUMO-1 modification of histone deacetylase 1 (HDAC1) modulates its biological activities. The Journal of biological chemistry 277, 23658–23663.

[Dayhoff et al., 1978] Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978). A model of evolutionary change in proteins. Atlas of Protein Sequence and Structure 5, 345–352.

[Debethune et al., 2002] Debethune, L., Kohlhagen, G., Grandas, A. and Pommier, Y. (2002). Processing of nucleopeptides mimicking the topoisomerase I-DNA covalent complex by tyrosyl-DNA phosphodiesterase. Nucleic Acids Res 30, 1198–1204.

[Delacroix et al., 2007] Delacroix, S., Wagner, J. M., Kobayashi, M., Yamamoto, K.-i. and Karnitz, L. M. (2007). The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. Genes Dev 21, 1472–1477.

[Deléage and Roux, 1987] Deléage, G. and Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. Protein engineering 1, 289–294.

[Desai et al., 2001] Desai, S. D., Li, T. K., Rodriguez-Bauman, A., Rubin, E. H. and Liu, L. F. (2001). Ubiquitin/26S proteasome-mediated degradation of topoisomerase I as a resistance mechanism to camptothecin in tumor cells. Cancer Res 61, 5926–5932.

[Desai et al., 2003] Desai, S. D., Zhang, H., Rodriguez-Bauman, A., Yang, J.-M., Wu, X., Gounder, M. K., Rubin, E. H. and Liu, L. F. (2003). Transcription-dependent degradation of topoisomerase I-DNA covalent complexes. Mol Cell Biol 23, 2341–2350.

[Desterro et al., 1999] Desterro, J., Rodriguez, M., Kemp, G. and Hay, R. (1999). Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. The Journal of biological chemistry 274, 10618–10624.

References

[Desterro et al., 1997] Desterro, J., Thomson, J. and Hay, R. (1997). Ubch9 conjugates SUMO but not ubiquitin. FEBS letters <u>417</u>, 297–300.

[Diella et al., 2008] Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., Trave, G. and Gibson, T. J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front Biosci <u>13</u>, 6580–6603.

[Do et al., 2005] Do, C., Mahabhashyam, M., Brudno, M. and Batzoglou, S. (2005). ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome research <u>15</u>, 330–340.

[Dohmen et al., 1995] Dohmen, R., Stappen, R., McGrath, J., Forrová, H., Kolarov, J., Goffeau, A. and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. The Journal of biological chemistry <u>270</u>, 18099–18109.

[Donaldson et al., 2003] Donaldson, K. M., Yin, H., Gekakis, N., Supek, F. and Joazeiro, C. A. P. (2003). Ubiquitin signals protein trafficking via interaction with a novel ubiquitin binding domain in the membrane fusion regulator, Vps9p. Curr Biol <u>13</u>, 258–262.

[Dosztányi et al., 2005a] Dosztányi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics (Oxford, England) <u>21</u>, 3433–3434.

[Dosztányi et al., 2005b] Dosztányi, Z., Csizmók, V., Tompa, P. and Simon, I. (2005b). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. Journal of molecular biology <u>347</u>, 827–839.

[Dove et al., 2004] Dove, S. K., Piper, R. C., McEwen, R. K., Yu, J. W., King, M. C., Hughes, D. C., Thuring, J., Holmes, A. B., Cooke, F. T., Michell, R. H., Parker, P. J. and Lemmon, M. A. (2004). Svp1p defines a family of phosphatidylinositol 3,5-bisphosphate effectors. The EMBO Journal <u>23</u>, 1922–1933.

[Dresser et al., 1997] Dresser, M. E., Ewing, D. J., Conrad, M. N., Dominguez, A. M., Barstead, R., Jiang, H. and Kodadek, T. (1997). DMC1 functions in a Saccharomyces cerevisiae meiotic pathway that is largely independent of the RAD51 pathway. Genetics <u>147</u>, 533–544.

[Dunker et al., 2001] Dunker, A., Lawson, J., Brown, C., Williams, R., Romero, P., Oh, J., Oldfield, C., Campen, A., Ratliff, C., Hipps, K., Ausio, J., Nissen, M., Reeves,

References

R., Kang, C., Kissinger, C., Bailey, R., Griswold, M., Chiu, W., Garner, E. and Obradovic, Z. (2001). Intrinsically disordered protein. Journal of molecular graphics and modelling 19, 26–59.

[Duszenko et al., 2011] Duszenko, M., Ginger, M., Brennand, A., Gualdrón-López, M., Colombo, M., Coombs, G., Coppens, I., Jayabalasingham, B., Langsley, G., de Castro, S., Menna-Barreto, R., Mottram, J., Navarro, M., Rigden, D., Romano, P., Stoka, V., Turk, B. and Michels, P. (2011). Autophagy in protists. Autophagy 7, 127–158.

[Eddy, 1995] Eddy, S. R. (1995). Multiple alignment using hidden Markov models. In Proceedings of the Third International Conference on Intelligent Systems in Molecular Biology, (AAAI Press, Cambridge, U., ed.), pp. 114–120,.

[Edgar, 2004a] Edgar, R. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5, 113.

[Edgar, 2004b] Edgar, R. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792–1797.

[Edgar, 2004c] Edgar, R. C. (2004c). Local homology recognition and distance measures in linear time using compressed amino acid alphabets. Nucleic Acids Research 32, 380–385.

[Espinoza-Fonseca, 2009] Espinoza-Fonseca, L. (2009). Reconciling binding mechanisms of intrinsically disordered proteins. Biochemical and biophysical research communications 382, 479–482.

[Farre et al., 2013] Farre, J.-C., Burkenroad, A., Burnett, S. F. and Subramani, S. (2013). Phosphorylation of mitophagy and pexophagy receptors coordinates their interaction with Atg8 and Atg11. EMBO Rep 14, 441–449.

[Fass et al., 2007] Fass, E., Amar, N. and Elazar, Z. (2007). Identification of essential residues for the C-terminal cleavage of the mammalian LC3: a lesson from yeast Atg8. Autophagy 3, 48–50.

[Felberbaum and Hochstrasser, 2008] Felberbaum, R. and Hochstrasser, M. (2008). Ulp2 and the DNA damage response: desumoylation enables safe passage through mitosis. Cell Cycle 7, 52–56.

[Feng and Doolittle, 1987] Feng, D. and Doolittle, R. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of molecular evolution 25, 351–360.

References

[Fields, 1993] Fields, S. (1993). The two-hybrid system to detect protein-protein interactions. Methods 5, 116–124.

[Fields and Song, 1989] Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature 340, 245–246.

[Finkbeiner et al., 2011] Finkbeiner, E., Haindl, M., Raman, N. and Muller, S. (2011). SUMO routes ribosome maturation. Nucleus 2, 527–532.

[Finn et al., 2014] Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Research 42, D222–D230.

[Finn et al., 2010] Finn, R. D., Mistry, J., Tate, J., Coggill, P. and Heger, A. (2010). The Pfam protein families database. Nucleic Acids Research 38, D211–D222.

[Fitch, 1970] Fitch, W. (1970). Distinguishing homologous from analogous proteins. Systematic zoology 19, 99–113.

[Fortini et al., 2000] Fortini, P., Pascucci, B., Belisario, F. and Dogliotti, E. (2000). DNA polymerase beta is required for efficient DNA strand break repair induced by methyl methanesulfonate but not by hydrogen peroxide. Nucleic Acids Research 28, 3040–3046.

[Freemont et al., 1991] Freemont, P. S., Hanson, I. M. and Trowsdale, J. (1991). A novel gysteine-rich sequence motif. Cell 64, 483–484.

[Fu et al., 1999] Fu, Q., Kim, S. W., Chen, H. X., Grill, S. and Cheng, Y. C. (1999). Degradation of topoisomerase I induced by topoisomerase I inhibitors is dependent on inhibitor structure but independent of cell death. Mol Pharmacol 55, 677–683.

[Fujita et al., 2008] Fujita, N., Itoh, T., Omori, H., Fukuda, M., Noda, T. and Yoshimori, T. (2008). The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. Mol Biol Cell 19, 2092–2100.

[Funakoshi et al., 1997] Funakoshi, T., Matsuura, A., Noda, T. and Ohsumi, Y. (1997). Analyses of APG13 gene involved in autophagy in yeast, Saccharomyces cerevisiae. Gene 192, 207–213.

References

[Fuxreiter et al., 2007] Fuxreiter, M., Tompa, P. and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics (Oxford, England) <u>23</u>, 950–956.

[Galzitskaya et al., 2006] Galzitskaya, O. V., Garbuzynskiy, S. O. and Lobanov, M. Y. (2006). FoldUnfold: web server for the prediction of disordered regions in protein chain. Computer applications in the biosciences : CABIOS <u>22</u>, 2948–2949.

[Ganoth et al., 2013] Ganoth, A., Tsfadia, Y. and Wiener, R. (2013). Ubiquitin: molecular modeling and simulations. J Mol Graph Model <u>46</u>, 29–40.

[Garbuzynskiy et al., 2004] Garbuzynskiy, S., Lobanov, M. and Galzitskaya, O. (2004). To be folded or to be unfolded? Protein science : a publication of the Protein Society <u>13</u>, 2871–2877.

[Goffeau et al., 1996] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S. G. (1996). Life with 6000 genes. Science <u>274</u>, 546, 563–7.

[Gomes and Burgers, 2001] Gomes, X. and Burgers, P. (2001). ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. The Journal of biological chemistry <u>276</u>, 34768–34775.

[Gong et al., 1999] Gong, L., Li, B., Millas, S. and Yeh, E. (1999). Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. FEBS letters <u>448</u>, 185–189.

[Gotoh, 1986] Gotoh, O. (1986). Alignment of three biological sequences with an efficient traceback procedure. Journal of theoretical biology <u>121</u>, 327–337.

[Gotoh, 1993] Gotoh, O. (1993). Optimal alignment between groups of sequences and its application to multiple sequence alignment. Computer applications in the biosciences : CABIOS <u>9</u>, 361–370.

[Gotoh, 1995] Gotoh, O. (1995). A weighting system and algorithm for aligning many phylogenetically related sequences. Computer Applications in the Biosciences <u>11</u>, 543–551.

[Gould et al., 2010] Gould, C. M., Diella, F., Via, A., Puntervoll, P., Gemund, C., Chabanis-Davidson, S., Michael, S., Sayadi, A., Bryne, J. C., Chica, C., Seiler, M.,

References

Davey, N. E., Haslam, N., Weatheritt, R. J., Budd, A., Hughes, T., Pas, J., Rych-lewski, L., Trave, G., Aasland, R., Helmer-Citterich, M., Linding, R. and Gibson, T. J. (2010). ELM: the status of the 2010 eukaryotic linear motif resource. Nucleic Acids Res 38, D167–80.

[Green et al., 2000] Green, C. M., Erdjument-Bromage, H., Tempst, P. and Lowndes, N. F. (2000). A novel Rad24 checkpoint protein complex closely related to replication factor C. Curr Biol 10, 39–42.

[Gsponer et al., 2008] Gsponer, J., Futschik, M., Teichmann, S. and Babu, M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science (New York, N.Y.) 322, 1365–1368.

[Guo et al., 2004] Guo, D., Li, M., Zhang, Y., Yang, P., Eckenrode, S., Hopkins, D., Zheng, W., Purohit, S., Podolsky, R., Muir, A., Wang, J., Dong, Z., Brusko, T., Atkinson, M., Pozzilli, P., Zeidler, A., Raffel, L., Jacob, C., Park, Y., Serrano-Rios, M., Larrad, M., Zhang, Z., Garchon, H.-J., Bach, J.-F., Rotter, J., She, J.-X. and Wang, C.-Y. (2004). A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. Nature genetics 36, 837–841.

[Haas et al., 1982] Haas, A., Warms, J., Hershko, A. and Rose, I. (1982). Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. The Journal of biological chemistry 257, 2543–2548.

[Halas et al., 2011] Halas, A., Podlaska, A., Derkacz, J., McIntyre, J., Skoneczna, A. and Sledziewska-Gojska, E. (2011). The roles of PCNA SUMOylation, Mms2-Ubc13 and Rad5 in translesion DNA synthesis in Saccharomyces cerevisiae. Mol Microbiol 80, 786–797.

[Hanada et al., 2007] Hanada, T., Noda, N., Satomi, Y., Ichimura, Y., Fujioka, Y., Takao, T., Inagaki, F. and Ohsumi, Y. (2007). The Atg12-Atg5 conjugate has a novel E3-like activity for protein lipidation in autophagy. The Journal of biological chemistry 282, 37298–37302.

[Hanada and Ohsumi, 2005] Hanada, T. and Ohsumi, Y. (2005). Structure-function relationship of Atg12, a ubiquitin-like modifier essential for autophagy. Autophagy 1, 110–118.

[Hanada et al., 2009] Hanada, T., Satomi, Y., Takao, T. and Ohsumi, Y. (2009). The amino-terminal region of Atg3 is essential for association with phos-phatidylethanolamine in Atg8 lipidation. FEBS Lett 583, 1078–1083.

References

[Hanna et al., 2001] Hanna, J. S., Kroll, E. S., Lundblad, V. and Spencer, F. A. (2001). Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. Mol Cell Biol 21, 3144–3158.

[Hannich et al., 2005] Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S.-J., Heide, H., Emili, A. and Hochstrasser, M. (2005). Defining the SUMO-modified Proteome by Multiple Approaches in Saccharomyces cerevisiae. Journal of Biological Chemistry 280, 4102–4110.

[Hatakeyama et al., 2001] Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N. and Nakayama, K. (2001). U box proteins as a new family of ubiquitin-protein ligases. The Journal of biological chemistry 276, 33111–33120.

[Hayashi-Nishino et al., 2009] Hayashi-Nishino, M., Fujita, N., Noda, T., Yamaguchi, A., Yoshimori, T. and Yamamoto, A. (2009). A subdomain of the endoplasmic reticulum forms a cradle for autophagosome formation. Nature cell biology 11, 1433–1437.

[Hayashi-Nishino et al., 2010] Hayashi-Nishino, M., Fujita, N., Noda, T., Yamaguchi, A., Yoshimori, T. and Yamamoto, A. (2010). Electron tomography reveals the endoplasmic reticulum as a membrane source for autophagosome formation. Autophagy 6, 301–303.

[He et al., 2003] He, H., Dang, Y., Dai, F., Guo, Z., Wu, J., She, X., Pei, Y., Chen, Y., Ling, W., Wu, C., Zhao, S., Liu, J. O. and Yu, L. (2003). Post-translational modifications of three members of the human MAP1LC3 family and detection of a novel type of modification for MAP1LC3B. J Biol Chem 278, 29278–29287.

[Hecker et al., 2006] Hecker, C.-M., Rabiller, M., Haglund, K., Bayer, P. and Dikic, I. (2006). Specification of SUMO1- and SUMO2-interacting motifs. The Journal of biological chemistry 281, 16117–16127.

[Hegger et al., 1999] Hegger, R., Kantz, H. and Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. Chaos: An Interdisciplinary Journal of Nonlinear Science 9, 413–435.

[Henikoff and Henikoff, 1991] Henikoff, S. and Henikoff, J. (1991). Automated assembly of protein blocks for database searching. Nucleic acids research 19, 6565–6572.

[Henikoff and Henikoff, 1992] Henikoff, S. and Henikoff, J. (1992). Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America 89, 10915–10919.

164

References

[Hershko and Ciechanover, 1998] Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. Annual review of biochemistry 67, 425–479.

[Hershko et al., 1983] Hershko, A., Heller, H., Elias, S. and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. The Journal of biological chemistry 258, 8206–8214.

[Hirano et al., 2006] Hirano, S., Kawasaki, M., Ura, H., Kato, R., Raiborg, C., Stenmark, H. and Wakatsuki, S. (2006). Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. Nat Struct Mol Biol 13, 272–277.

[Hitt and Wolf, 2004] Hitt, R. and Wolf, D. H. (2004). Der1p, a protein required for degradation of malfolded soluble proteins of the endoplasmic reticulum: topology and Der1–like proteins. FEMS Yeast Research 4, 721–729.

[Hochstrasser, 2000] Hochstrasser, M. (2000). Evolution and function of ubiquitin-like protein-conjugation systems. Nature cell biology 2.

[Hochstrasser, 2001] Hochstrasser, M. (2001). SP-RING for SUMO: new functions bloom for a ubiquitin-like protein. Cell 107, 5–8.

[Hoege et al., 2002] Hoege, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G. and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 419, 135–141.

[Hofmann and Falquet, 2001] 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.

[Höge, 2002] Höge, C. (2002). RAD6-abhängige DNA-Reparatur wird durch Ubiquitin- und SUMO-Modifikation von PCNA reguliert. PhD thesis, Fakultät für Biologie der Ludwig-Maximilians-Universität München.

[Hogeweg and Hesper, 1984] Hogeweg, P. and Hesper, B. (1984). The alignment of sets of sequences and the construction of phyletic trees: An integrated method. Journal of molecular evolution 20, 175–186.

[Horie et al., 2002] Horie, K., Tomida, A., Sugimoto, Y., Yasugi, T., Yoshikawa, H., Taketani, Y. and Tsuruo, T. (2002). SUMO-1 conjugation to intact DNA topoisomerase I amplifies cleavable complex formation induced by camptothecin. Oncogene 21, 7913–7922.

References

[Hughey and Krogh, 1996] Hughey, R. and Krogh, A. (1996). Hidden Markov models for sequence analysis: extension and analysis of the basic method. Bioinformatics 12, 95–107.

[Huibregtse, 1995] Huibregtse, J. M. (1995). A Family of Proteins Structurally and Functionally Related to the E6-AP Ubiquitin-Protein Ligase. Proceedings of the National Academy of Sciences 92.

[Hutchins and Klionsky, 2001] Hutchins, M. U. and Klionsky, D. J. (2001). Vacuolar localization of oligomeric alpha-mannosidase requires the cytoplasm to vacuole targeting and autophagy pathway components in Saccharomyces cerevisiae. J Biol Chem 276, 20491–20498.

[Ichimura et al., 2000] Ichimura, Y., Kirisako, T., Takao, T., Satomi, Y., Shimonishi, Y., Ishihara, N., Mizushima, N., Tanida, I., Kominami, E., Ohsumi, M., Noda, T. and Ohsumi, Y. (2000). A ubiquitin-like system mediates protein lipidation. Nature 408, 488–492.

[Ichimura et al., 2008] Ichimura, Y., Kumanomidou, T., Sou, Y.-s., Mizushima, T., Ezaki, J., Ueno, T., Kominami, E., Yamane, T., Tanaka, K. and Komatsu, M. (2008). Structural basis for sorting mechanism of p62 in selective autophagy. The Journal of biological chemistry 283, 22847–22857.

[Ii et al., 2007a] Ii, T., Fung, J., Mullen, J. R. and Brill, S. J. (2007a). The yeast Slx5-Slx8 DNA integrity complex displays ubiquitin ligase activity. Cell cycle (Georgetown, Tex.) 6, 2800–2809.

[Ii et al., 2007b] Ii, T., Mullen, J. R., Slagle, C. E. and Brill, S. J. (2007b). Stimulation of in vitro sumoylation by Slx5–Slx8: evidence for a functional interaction with the SUMO pathway. DNA Repair 6, 1679–1691.

[Isogai et al., 2011] Isogai, S., Morimoto, D., Arita, K., Unzai, S., Tenno, T., Hasegawa, J., Sou, Y.-s., Komatsu, M., Tanaka, K., Shirakawa, M. and Tochio, H. (2011). Crystal structure of the ubiquitin-associated (UBA) domain of p62 and its interaction with ubiquitin. J Biol Chem 286, 31864–31874.

[James et al., 1996] James, P., Halladay, J. and Craig, E. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 144, 1425–1436.

References

[Janke et al., 2004] Janke, C., Magiera, M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E. and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast (Chichester, England) 21, 947–962.

[Jeruzalmi et al., 2001] Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M. and Kuriyan, J. (2001). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of E. coli DNA polymerase III. Cell 106, 417–428.

[Jiang et al., 2001] Jiang, J., Ballinger, C., Wu, Y., Dai, Q., Cyr, D., Höhfeld, J. and Patterson, C. (2001). CHIP is a U-box-dependent E3 ubiquitin ligase: identification of Hsc70 as a target for ubiquitylation. The Journal of biological chemistry 276, 42938–42944.

[Jin et al., 2007] Jin, J., Li, X., Gygi, S. and Harper, J. (2007). Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. Nature 447, 1135–1138.

[Jiricny, 2006] Jiricny, J. (2006). The multifaceted mismatch-repair system. Nature reviews. Molecular cell biology 7, 335–346.

[Johnson, 2004] Johnson, E. (2004). Protein modification by SUMO. Annual review of biochemistry 73, 355–382.

[Johnson and Blobel, 1997] Johnson, E. and Blobel, G. (1997). Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. The Journal of biological chemistry 272, 26799–26802.

[Johnson and Gupta, 2001] Johnson, E. and Gupta, A. (2001). An E3-like factor that promotes SUMO conjugation to the yeast septins. Cell 106, 735–744.

[Johnson et al., 1997] Johnson, E., Schwienhorst, I., Dohmen, R. and Blobel, G. (1997). The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. The EMBO journal 16, 5509–5519.

[Joung et al., 2000] Joung, J., Ramm, E. and Pabo, C. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America 97, 7382–7387.

References

[Kabeya et al., 2000] Kabeya, Y., Mizushima, N., Ueno, T., Yamamoto, A., Kirisako, T., Noda, T., Kominami, E., Ohsumi, Y. and Yoshimori, T. (2000). LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. The EMBO journal 19, 5720–5728.

[Kabeya et al., 2004] Kabeya, Y., Mizushima, N., Yamamoto, A., Oshitani-Okamoto, S., Ohsumi, Y. and Yoshimori, T. (2004). LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. Journal of cell science 117, 2805–2812.

[Kabsch and Sander, 1983] Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen‐bonded and geometrical features. Biopolymers 22, 2577–2637.

[Kagey et al., 2005] Kagey, M., Melhuish, T., Powers, S. and Wotton, D. (2005). Multiple activities contribute to Pc2 E3 function. The EMBO journal 24, 108–119.

[Kagey et al., 2003] Kagey, M., Melhuish, T. and Wotton, D. (2003). The polycomb protein Pc2 is a SUMO E3. Cell 113, 127–137.

[Kahyo et al., 2001] Kahyo, T., Nishida, T. and Yasuda, H. (2001). Involvement of PIAS1 in the sumoylation of tumor suppressor p53. Molecular cell 8, 713–718.

[Kaliraman et al., 2001] Kaliraman, V., Mullen, J. R., Fricke, W. M., Bastin-Shanower, S. A. and Brill, S. J. (2001). Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. Genes & Development 15, 2730–2740.

[Kamada et al., 2000] Kamada, Y., Funakoshi, T., Shintani, T., Nagano, K., Ohsumi, M. and Ohsumi, Y. (2000). Tor-mediated induction of autophagy via an Apg1 protein kinase complex. The Journal of cell biology 150, 1507–1513.

[Kanellis et al., 2003] Kanellis, P., Agyei, R. and Durocher, D. (2003). Elg1 forms an alternative PCNA-interacting RFC complex required to maintain genome stability. Curr Biol 13, 1583–1595.

[Kang et al., 2003] Kang, R. S., Daniels, C. M., Francis, S. A., Shih, S. C., Salerno, W. J., Hicke, L. and Radhakrishnan, I. (2003). Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. Cell 113, 621–630.

[Kanki and Klionsky, 2008] Kanki, T. and Klionsky, D. J. (2008). Mitophagy in yeast occurs through a selective mechanism. J Biol Chem 283, 32386–32393.

References

[Kanki et al., 2009] Kanki, T., Wang, K., Cao, Y., Baba, M. and Klionsky, D. (2009). Atg32 is a mitochondrial protein that confers selectivity during mitophagy. Developmental cell $\underline{17}$, 98–109.

[Kannouche et al., 2004] Kannouche, P. L., Wing, J. and Lehmann, A. R. (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Molecular Cell $\underline{14}$, 491–500.

[Karp, 1993] Karp, R. M. (1993). Mapping the genome: some combinatorial problems arising in molecular biology. some combinatorial problems arising in molecular biology, ACM, New York, New York, USA.

[Katoh et al., 2002] Katoh, K., Misawa, K., Kuma, K.-i. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research $\underline{30}$, 3059–3066.

[Kawamata et al., 2008] Kawamata, T., Kamada, Y., Kabeya, Y., Sekito, T. and Ohsumi, Y. (2008). Organization of the pre-autophagosomal structure responsible for autophagosome formation. Mol Biol Cell $\underline{19}$, 2039–2050.

[Kelch et al., 2011] Kelch, B., Makino, D., O'Donnell, M. and Kuriyan, J. (2011). How a DNA polymerase clamp loader opens a sliding clamp. Science (New York, N.Y.) $\underline{334}$, 1675–1680.

[Kihara et al., 2001] Kihara, A., Noda, T., Ishihara, N. and Ohsumi, Y. (2001). Two distinct Vps34 phosphatidylinositol 3-kinase complexes function in autophagy and carboxypeptidase Y sorting in Saccharomyces cerevisiae. The Journal of cell biology $\underline{152}$, 519–530.

[Kijanska and Peter, 2013] Kijanska, M. and Peter, M. (2013). Atg1 kinase regulates early and late steps during autophagy. Autophagy $\underline{9}$, 249–251.

[Kim et al., 2001a] Kim, J., Huang, W. and Klionsky, D. (2001a). Membrane recruitment of Aut7p in the autophagy and cytoplasm to vacuole targeting pathways requires Aut1p, Aut2p, and the autophagy conjugation complex. The Journal of cell biology $\underline{152}$, 51–64.

[Kim et al., 2001b] Kim, J., Kamada, Y., Stromhaug, P. E., Guan, J., Hefner-Gravink, A., Baba, M., Scott, S. V., Ohsumi, Y., Dunn, W. A. J. and Klionsky, D. J. (2001b).

169

Cvt9/Gsa9 functions in sequestering selective cytosolic cargo destined for the vacuole. J Cell Biol 153, 381–396.

[Kim et al., 1994] Kim, J., Pramanik, S. and Chung, M. J. (1994). Multiple sequence alignment using simulated annealing. Computer Applications in the Biosciences 10, 419–426.

[Kimura and Takahata, 1983] Kimura, M. and Takahata, N. (1983). Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method. Proceedings of the National Academy of Sciences 80, 1048–1052.

[Kirisako et al., 1999] Kirisako, T., Baba, M., Ishihara, N., Miyazawa, K., Ohsumi, M., Yoshimori, T., Noda, T. and Ohsumi, Y. (1999). Formation process of autophagosome is traced with Apg8/Aut7p in yeast. The Journal of cell biology 147, 435–446.

[Kirisako et al., 2000] Kirisako, T., Ichimura, Y., Okada, H., Kabeya, Y., Mizushima, N., Yoshimori, T., Ohsumi, M., Takao, T., Noda, T. and Ohsumi, Y. (2000). The reversible modification regulates the membrane-binding state of Apg8/Aut7 essential for autophagy and the cytoplasm to vacuole targeting pathway. The Journal of cell biology 151, 263–276.

[Kirkin et al., 2009a] Kirkin, V., Lamark, T., Sou, Y.-S., Bjørkøy, G., Nunn, J., Bruun, J.-A., Shvets, E., McEwan, D., Clausen, T., Wild, P., Bilusic, I., Theurillat, J.-P., Øvervatn, A., Ishii, T., Elazar, Z., Komatsu, M., Dikic, I. and Johansen, T. (2009a). A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. Molecular cell 33, 505–516.

[Kirkin et al., 2009b] Kirkin, V., McEwan, D., Novak, I. and Dikic, I. (2009b). A role for ubiquitin in selective autophagy. Molecular cell 34, 259–269.

[Kirsh et al., 2002] Kirsh, O., Seeler, J.-S., Pichler, A., Gast, A., Müller, S., Miska, E., Mathieu, M., Harel-Bellan, A., Kouzarides, T., Melchior, F. and Dejean, A. (2002). The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. The EMBO journal 21, 2682–2691.

[Kleppe et al., 1971] Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. and Khorana, H. (1971). Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. Journal of molecular biology 56, 341–361.

References

[Knop et al., 1999] Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K. and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. Yeast (Chichester, England) 15, 963–972.

[Kolesar et al., 2012] Kolesar, P., Sarangi, P., Altmannova, V., Zhao, X. and Krejci, L. (2012). Dual roles of the SUMO-interacting motif in the regulation of Srs2 sumoylation. Nucleic acids research 40, 7831–7843.

[Komatsu et al., 2007] Komatsu, M., Waguri, S., Koike, M., Sou, Y.-S., Ueno, T., Hara, T., Mizushima, N., Iwata, J.-I., Ezaki, J., Murata, S., Hamazaki, J., Nishito, Y., Iemura, S.-I., Natsume, T., Yanagawa, T., Uwayama, J., Warabi, E., Yoshida, H., Ishii, T., Kobayashi, A., Yamamoto, M., Yue, Z., Uchiyama, Y., Kominami, E. and Tanaka, K. (2007). Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. Cell 131, 1149–1163.

[Koonin, 2005] Koonin, E. (2005). Orthologs, paralogs, and evolutionary genomics. Annual review of genetics 39, 309–338.

[Kotaja et al., 2002] Kotaja, N., Karvonen, U., Jänne, O. and Palvimo, J. (2002). PIAS proteins modulate transcription factors by functioning as SUMO-1 ligases. Molecular and cellular biology 22, 5222–5234.

[Kouno et al., 2005] Kouno, T., Mizuguchi, M., Tanida, I., Ueno, T., Kanematsu, T., Mori, Y., Shinoda, H., Hirata, M., Kominami, E. and Kawano, K. (2005). Solution structure of microtubule-associated protein light chain 3 and identification of its functional subdomains. J Biol Chem 280, 24610–24617.

[Kraft et al., 2012] Kraft, C., Kijanska, M., Kalie, E., Siergiejuk, E., Lee, S. S., Semplicio, G., Stoffel, I., Brezovich, A., Verma, M., Hansmann, I., Ammerer, G., Hofmann, K., Tooze, S. and Peter, M. (2012). Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy. EMBO J 31, 3691–3703.

[Krakoff et al., 1968] Krakoff, I. H., Brown, N. C. and Reichard, P. (1968). Inhibition of ribonucleoside diphosphate reductase by hydroxyurea. Cancer Res 28, 1559–1565.

[Kroemer et al., 2010] Kroemer, G., Mariño, G. and Levine, B. (2010). Autophagy and the integrated stress response. Molecular cell 40, 280–293.

[Krogh, 1994] Krogh, A. (1994). Hidden Markov models for labeled sequences, vol. 2,. IEEE.

171

References

[Krogh and Symington, 2004] Krogh, B. O. and Symington, L. S. (2004). Recombination proteins in yeast. Annu Rev Genet <u>38</u>, 233–271.

[Kubota et al., 2013] Kubota, T., Myung, K. and Donaldson, A. D. (2013). Is PCNA unloading the central function of the Elg1/ATAD5 replication factor C-like complex? Cell Cycle <u>12</u>, 2570–2579.

[Kulman et al., 2001] Kulman, J. D., Harris, J. E., Xie, L. and Davie, E. W. (2001). Identification of two novel transmembrane gamma-carboxyglutamic acid proteins expressed broadly in fetal and adult tissues. Proc Natl Acad Sci U S A <u>98</u>, 1370–1375.

[Kuma et al., 2002] Kuma, A., Mizushima, N., Ishihara, N. and Ohsumi, Y. (2002). Formation of the approximately 350-kDa Apg12-Apg5.Apg16 multimeric complex, mediated by Apg16 oligomerization, is essential for autophagy in yeast. The Journal of biological chemistry <u>277</u>, 18619–18625.

[Kumeta et al., 2010] Kumeta, H., Watanabe, M., Nakatogawa, H., Yamaguchi, M., Ogura, K., Adachi, W., Fujioka, Y., Noda, N. N., Ohsumi, Y. and Inagaki, F. (2010). The NMR structure of the autophagy-related protein Atg8. J Biomol NMR <u>47</u>, 237–241.

[Kurian, 2009] Kurian, L. (2009). Mechanisms regulating ribosomal frameshifting in decoding Ornithine Decarboxylase Antizyme mRNA. PhD thesis, University of Cologne Zülpicher Straße 47a, 50674 Cologne.

[Kyte and Doolittle, 1982] Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology <u>157</u>, 105–132.

[Laemmli et al., 1970] Laemmli, U. K., Beguin, F. and Gujer-Kellenberger, G. (1970). A factor preventing the major head protein of bacteriophage T4 from random aggregation. Journal of Molecular Biology <u>47</u>, 69–85.

[Lallemand-Breitenbach et al., 2008] Lallemand-Breitenbach, V. e. r., Jeanne, M., Benhenda, S., Nasr, R., Lei, M., Peres, L., Zhou, J., Zhu, J., Raught, B. and de Th eacute, H. (2008). Arsenic degrades PML or PML—[ndash]—RAR—[alpha]— through a SUMO-triggered RNF4/ubiquitin-mediated pathway. Nature Cell Biology <u>10</u>, 547–555.

References

[Lamark et al., 2009] Lamark, T., Kirkin, V., Dikic, I. and Johansen, T. (2009). NBR1 and p62 as cargo receptors for selective autophagy of ubiquitinated targets. Cell cycle (Georgetown, Tex.) 8, 1986–1990.

[Lander et al., 1991] Lander, E. S., Langridge, R. and Saccocio, D. M. (1991). Mapping and interpreting biological information. Communications of the ACM 34, 32–39.

[Landy, 1989] Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annual review of biochemistry 58, 913–949.

[Lee et al., 1998] Lee, G., Melchior, F., Matunis, M., Mahajan, R., Tian, Q. and Anderson, P. (1998). Modification of Ran GTPase-activating protein by the small ubiquitin-related modifier SUMO-1 requires Ubc9, an E2-type ubiquitin-conjugating enzyme homologue. The Journal of biological chemistry 273, 6503–6507.

[Lee et al., 2007] Lee, J., Kumagai, A. and Dunphy, W. G. (2007). The Rad9-Hus1-Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. J Biol Chem 282, 28036–28044.

[Lee et al., 2011] Lee, M.-T., Bakir, A. A., Nguyen, K. N. and Bachant, J. (2011). The SUMO isopeptidase Ulp2p is required to prevent recombination-induced chromosome segregation lethality following DNA replication stress. PLoS Genet 7, e1001355.

[Levine and Kroemer, 2008] Levine, B. and Kroemer, G. (2008). Autophagy in the pathogenesis of disease. Cell 132, 27–42.

[Levine et al., 2011] Levine, B., Mizushima, N. and Virgin, H. (2011). Autophagy in immunity and inflammation. Nature 469, 323–335.

[Li and Hochstrasser, 1999] Li, S. and Hochstrasser, M. (1999). A new protease required for cell-cycle progression in yeast. Nature 398, 246–251.

[Li and Hochstrasser, 2000] Li, S. and Hochstrasser, M. (2000). The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. Molecular and cellular biology 20, 2367–2377.

[Li and Heyer, 2008] Li, X. and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell research 18, 99–113.

[Liew et al., 2010] Liew, C. W., Sun, H., Hunter, T. and Day, C. L. (2010). RING domain dimerization is essential for RNF4 function. Biochemical Journal 431, 23–29.

References

[Lin et al., 2006] Lin, D.-Y., Huang, Y.-S., Jeng, J.-C., Kuo, H.-Y., Chang, C.-C., Chao, T.-T., Ho, C.-C., Chen, Y.-C., Lin, T.-P., Fang, H.-I., Hung, C.-C., Suen, C.-S., Hwang, M.-J., Chang, K.-S., Maul, G. G. and Shih, H.-M. (2006). Role of SUMO-Interacting Motif in Daxx SUMO Modification, Subnuclear Localization, and Repression of Sumoylated Transcription Factors. Molecular Cell <u>24</u>, 341–354.

[Lindahl, 1990] Lindahl, T. (1990). Repair of intrinsic DNA lesions. Mutat Res <u>238</u>, 305–311.

[Linding et al., 2003a] Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. and Russell, R. B. (2003a). Protein Disorder Prediction. Structure <u>11</u>, 1453–1459.

[Linding et al., 2003b] Linding, R., Russell, R., Neduva, V. and Gibson, T. (2003b). GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic acids research <u>31</u>, 3701–3708.

[Lindsey-Boltz et al., 2001] Lindsey-Boltz, L. A., Bermudez, V. P., Hurwitz, J. and Sancar, A. (2001). Purification and characterization of human DNA damage checkpoint Rad complexes. Proc Natl Acad Sci U S A <u>98</u>, 11236–11241.

[Lipman et al., 1989] Lipman, D., Altschul, S. and Kececioglu, J. (1989). A tool for multiple sequence alignment. Proceedings of the National Academy of Sciences of the United States of America <u>86</u>, 4412–4415.

[Liu et al., 1998] Liu, N., Lamerdin, J. E., Tebbs, R. S., Schild, D., Tucker, J. D., Shen, M. R., Brookman, K. W., Siciliano, M. J., Walter, C. A., Fan, W., Narayana, L. S., Zhou, Z. Q., Adamson, A. W., Sorensen, K. J., Chen, D. J., Jones, N. J. and Thompson, L. H. (1998). XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. Mol Cell <u>1</u>, 783–793.

[Livnat-Levanon and Glickman, 2011] Livnat-Levanon, N. and Glickman, M. (2011). Ubiquitin-proteasome system and mitochondria - reciprocity. Biochimica et biophysica acta <u>1809</u>, 80–87.

[Lo Conte et al., 2000] Lo Conte, L., Ailey, B., Hubbard, T. J. P., Brenner, S. E., Murzin, A. G. and Chothia, C. (2000). SCOP: a Structural Classification of Proteins database. Nucleic Acids Research <u>28</u>, 257–259.

References

[Lo Conte et al., 2002a] Lo Conte, L., Brenner, S., Hubbard, T., Chothia, C. and Murzin, A. (2002a). SCOP database in 2002: refinements accommodate structural genomics. Nucleic acids research <u>30</u>, 264–267.

[Lo Conte et al., 2002b] Lo Conte, L., Brenner, S. E., Hubbard, T. J. P., Chothia, C. and Murzin, A. G. (2002b). SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Research <u>30</u>, 264–267.

[Longtine et al., 1998] Longtine, M. S., McKenzie, A. r., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P. and Pringle, J. R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast <u>14</u>, 953–961.

[Lundin, 2005] Lundin, C. (2005). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. Nucleic Acids Research <u>33</u>, 3799–3811.

[Mager et al., 1997] Mager, W. H., Planta, R. J., Ballesta, J. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R. and Woolford, J. (1997). A new nomenclature for the cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. Nucleic Acids Res <u>25</u>, 4872–4875.

[Mahajan et al., 1997] Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell <u>88</u>, 97–107.

[Mahajan et al., 1998] Mahajan, R., Gerace, L. and Melchior, F. (1998). Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. J Cell Biol <u>140</u>, 259–270.

[Mamonova et al., 2010] Mamonova, T. B., Glyakina, A. V., Kurnikova, M. G. and Galzitskaya, O. V. (2010). Flexibility and mobility in mesophilic and thermophilic homologous proteins from molecular dynamics and FoldUnfold method. J Bioinform Comput Biol <u>8</u>, 377–394.

[Mann and Hammarback, 1994] Mann, S. S. and Hammarback, J. A. (1994). Molecular characterization of light chain 3. A microtubule binding subunit of MAP1A and MAP1B. J Biol Chem <u>269</u>, 11492–11497.

[Mannen et al., 1996] Mannen, H., Tseng, H., Cho, C. and Li, S. (1996). Cloning and expression of human homolog HSMT3 to yeast SMT3 suppressor of MIF2 mutations

in a centromere protein gene. Biochemical and biophysical research communications 222, 178–180.

[Mao et al., 2000] Mao, Y., Sun, M., Desai, S. D. and Liu, L. F. (2000). SUMO-1 conjugation to topoisomerase I: A possible repair response to topoisomerase-mediated DNA damage. Proc Natl Acad Sci U S A 97, 4046–4051.

[Matsuura et al., 1997] Matsuura, A., Tsukada, M., Wada, Y. and Ohsumi, Y. (1997). Apg1p, a novel protein kinase required for the autophagic process in Saccharomyces cerevisiae. Gene 192, 245–250.

[Matunis et al., 1996] Matunis, M., Coutavas, E. and Blobel, G. (1996). A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. The Journal of cell biology 135, 1457–1470.

[Mayer et al., 2001] Mayer, M. L., Gygi, S. P., Aebersold, R. and Hieter, P. (2001). Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. Mol Cell 7, 959–970.

[McGrath et al., 1991] McGrath, J., Jentsch, S. and Varshavsky, A. (1991). UBA 1: an essential yeast gene encoding ubiquitin-activating enzyme. The EMBO journal 10, 227–236.

[Meluh and Koshland, 1995] Meluh, P. and Koshland, D. (1995). Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Molecular biology of the cell 6, 793–807.

[Merrill et al., 2010] Merrill, J., Melhuish, T., Kagey, M., Yang, S.-H., Sharrocks, A. and Wotton, D. (2010). A role for non-covalent SUMO interaction motifs in Pc2/CBX4 E3 activity. PloS one 5.

[Minty et al., 2000] Minty, A., Dumont, X., Kaghad, M. and Caput, D. (2000). Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. The Journal of biological chemistry 275, 36316–36323.

[Miteva, 2007] Miteva, M. (2007). Proteolytic Mechanisms Controlling SUMO Protein Modification. PhD thesis, University of Cologne Zülpicher Straße 47a, 50674 Cologne.

References

[Miyazawa and Jernigan, 1985] Miyazawa, S. and Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 18.

[Mizushima and Komatsu, 2011] Mizushima, N. and Komatsu, M. (2011). Autophagy: renovation of cells and tissues. Cell 147, 728–741.

[Mizushima et al., 2003] Mizushima, N., Kuma, A., Kobayashi, Y., Yamamoto, A., Matsubae, M., Takao, T., Natsume, T., Ohsumi, Y. and Yoshimori, T. (2003). Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. Journal of cell science 116, 1679–1688.

[Mizushima et al., 2008] Mizushima, N., Levine, B., Cuervo, A. and Klionsky, D. (2008). Autophagy fights disease through cellular self-digestion. Nature 451, 1069–1075.

[Mizushima et al., 1999] Mizushima, N., Noda, T. and Ohsumi, Y. (1999). Apg16p is required for the function of the Apg12p-Apg5p conjugate in the yeast autophagy pathway. The EMBO journal 18, 3888–3896.

[Mizushima et al., 1998] Mizushima, N., Noda, T., Yoshimori, T., Tanaka, Y., Ishii, T., George, M., Klionsky, D., Ohsumi, M. and Ohsumi, Y. (1998). A protein conjugation system essential for autophagy. Nature 395, 395–398.

[Mizushima et al., 2011] Mizushima, N., Yoshimori, T. and Ohsumi, Y. (2011). The Role of Atg Proteins in Autophagosome Formation. Annual Review of Cell and Developmental Biology 27, 107–132.

[Morgenstern, 1996] Morgenstern, B. (1996). Multiple DNA and protein sequence alignment based on segment-to-segment comparison. Proceedings of the National Academy of Sciences 93, 12098–12103.

[Morgenstern et al., 1998] Morgenstern, B., Frech, K., Dress, A. and Werner, T. (1998). DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics 14, 290–294.

[Morgenstern et al., 2006] Morgenstern, B., Prohaska, S. J., Pöhler, D. and F, S. P. (2006). Multiple sequence alignment with user-defined anchor points. Algorithms for Molecular Biology 1.

[Moscat et al., 2007] Moscat, J., Diaz-Meco, M. T. and Wooten, M. W. (2007). Signal integration and diversification through the p62 scaffold protein. Trends Biochem Sci 32, 95–100.

177

References

[Mossessova and Lima, 2000] Mossessova, E. and Lima, C. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Molecular cell 5, 865–876.

[Motley et al., 2012] Motley, A. M., Nuttall, J. M. and Hettema, E. H. (2012). Pex3-anchored Atg36 tags peroxisomes for degradation in Saccharomyces cerevisiae. EMBO J 31, 2852–2868.

[Mueller and Feigon, 2002] Mueller, T. D. and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. J Mol Biol 319, 1243–1255.

[Mukhopadhyay and Dasso, 2007] Mukhopadhyay, D. and Dasso, M. (2007). Modification in reverse: the SUMO proteases. Trends in biochemical sciences 32, 286–295.

[Mullen et al., 2010] Mullen, J., Chen, C.-F. and Brill, S. (2010). Wss1 is a SUMO-dependent isopeptidase that interacts genetically with the Slx5-Slx8 SUMO-targeted ubiquitin ligase. Molecular and cellular biology 30, 3737–3748.

[Mullen and Brill, 2008] Mullen, J. R. and Brill, S. J. (2008). Activation of the Slx5-Slx8 ubiquitin ligase by poly-small ubiquitin-like modifier conjugates. The Journal of biological chemistry 283, 19912–19921.

[Mullen et al., 2000] Mullen, J. R., Kaliraman, V. and Brill, S. J. (2000). Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae. Genetics 154, 1101–1114.

[Mullen et al., 2001] Mullen, J. R., Kaliraman, V., Ibrahim, S. S. and Brill, S. J. (2001). Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in Saccharomyces cerevisiae. Genetics 157, 103–118.

[Murata et al., 1985] Murata, M., Richardson, J. S. and Sussman, J. L. (1985). Simultaneous comparison of three protein sequences. Proceedings of the National Academy of Sciences of the United States of America 82, 3073–3077.

[Murata et al., 2001] Murata, S., Minami, Y., Minami, M., Chiba, T. and Tanaka, K. (2001). CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. EMBO reports 2, 1133–1138.

[Naiki et al., 2001] Naiki, T., Kondo, T., Nakada, D., Matsumoto, K. and Sugimoto, K. (2001). Chl12 (Ctf18) forms a novel replication factor C-related complex and functions

redundantly with Rad24 in the DNA replication checkpoint pathway. Mol Cell Biol 21, 5838–5845.

[Nakagawa and Yokosawa, 2002] Nakagawa, K. and Yokosawa, H. (2002). PIAS3 induces SUMO-1 modification and transcriptional repression of IRF-1. FEBS letters 530, 204–208.

[Nakamura et al., 2010] Nakamura, K., Kimple, A. J., Siderovski, D. P. and Johnson, G. L. (2010). PB1 domain interaction of p62/sequestosome 1 and MEKK3 regulates NF-kappaB activation. J Biol Chem 285, 2077–2089.

[Namanja et al., 2012] Namanja, A., Li, Y.-J., Su, Y., Wong, S., Lu, J., Colson, L., Wu, C., Li, S. and 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. The Journal of biological chemistry 287, 3231–3240.

[Neduva and Russell, 2005] Neduva, V. and Russell, R. (2005). Linear motifs: evolutionary interaction switches. FEBS letters 579, 3342–3345.

[Needleman and Wunsch, 1970] Needleman, S. and Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology 48, 443–453.

[Nishida and Yasuda, 2002] Nishida, T. and Yasuda, H. (2002). PIAS1 and PIASxalpha function as SUMO-E3 ligases toward androgen receptor and repress androgen receptor-dependent transcription. The Journal of biological chemistry 277, 41311–41317.

[Noda et al., 2008] Noda, N., Kumeta, H., Nakatogawa, H., Satoo, K., Adachi, W., Ishii, J., Fujioka, Y., Ohsumi, Y. and Inagaki, F. (2008). Structural basis of target recognition by Atg8/LC3 during selective autophagy. Genes to cells : devoted to molecular & cellular mechanisms 13, 1211–1218.

[Noda et al., 2000] Noda, T., Kim, J., Huang, W. P., Baba, M., Tokunaga, C., Ohsumi, Y. and Klionsky, D. J. (2000). Apg9p/Cvt7p is an integral membrane protein required for transport vesicle formation in the Cvt and autophagy pathways. The Journal of Cell Biology 148, 465–480.

[Noda and Yoshimori, 2009] Noda, T. and Yoshimori, T. (2009). Molecular basis of canonical and bactericidal autophagy. International immunology 21, 1199–1204.

References

[Notredame and Higgins, 1996] Notredame, C. and Higgins, D. G. (1996). SAGA: sequence alignment by genetic algorithm. Nucleic Acids Research <u>24</u>, 1515–1524.

[Novak et al., 2010] Novak, I., Kirkin, V., McEwan, D., Zhang, J., Wild, P., Rozenknop, A., Rogov, V., Löhr, F., Popovic, D., Occhipinti, A., Reichert, A., Terzic, J., Dötsch, V., Ney, P. and Dikic, I. (2010). Nix is a selective autophagy receptor for mitochondrial clearance. EMBO reports <u>11</u>, 45–51.

[Obara et al., 2006] Obara, K., Sekito, T. and Ohsumi, Y. (2006). Assortment of phosphatidylinositol 3-kinase complexes–Atg14p directs association of complex I to the pre-autophagosomal structure in Saccharomyces cerevisiae. Molecular biology of the cell <u>17</u>, 1527–1539.

[O'Brien et al., 2005] O'Brien, K., Remm, M. and Sonnhammer, E. L. (2005). Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic acids research <u>33</u>.

[Ohno, 1970] Ohno, S. (1970). Evolution by gene duplication. London: George Alien and Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.

[Okamoto et al., 2009] Okamoto, K., Kondo-Okamoto, N. and Ohsumi, Y. (2009). Mitochondria-anchored receptor Atg32 mediates degradation of mitochondria via selective autophagy. Developmental cell <u>17</u>, 87–97.

[Okuma et al., 1999] Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N. and Yasuda, H. (1999). In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. Biochemical and biophysical research communications <u>254</u>, 693–698.

[Oldham et al., 2002] Oldham, C. E., Mohney, R. P., Miller, S. L. H., Hanes, R. N. and O'Bryan, J. P. (2002). The Ubiquitin-Interacting Motifs Target the Endocytic Adaptor Protein Epsin for Ubiquitination. Current Biology <u>12</u>, 1112–1116.

[Ostlund et al., 2010] Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D., Roopra, S., Frings, O. and Sonnhammer, E. L. (2010). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic acids research <u>38</u>.

[Pan et al., 2006] Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S. and Boeke, J. D. (2006). A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell <u>124</u>, 1069–1081.

[Pankiv et al., 2007] Pankiv, S., Clausen, T., Lamark, T., Brech, A., Bruun, J.-A., Outzen, H., Øvervatn, A., Bjørkøy, G. and Johansen, T. (2007). p62/SQSTM1 binds

directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. The Journal of biological chemistry $\underline{282}$, 24131–24145.

[Panse et al., 2006] Panse, V., Kressler, D., Pauli, A., Petfalski, E., Gnädig, M., Tollervey, D. and Hurt, E. (2006). Formation and nuclear export of preribosomes are functionally linked to the small-ubiquitin-related modifier pathway. Traffic (Copenhagen, Denmark) $\underline{7}$, 1311–1321.

[Papouli et al., 2005] Papouli, E., Chen, S., Davies, A. A., Huttner, D., Krejci, L., Sung, P. and Ulrich, H. D. (2005). Crosstalk between SUMO and Ubiquitin on PCNA Is Mediated by Recruitment of the Helicase Srs2p. Molecular Cell $\underline{19}$, 123–133.

[Parker and Ulrich, 2012] Parker, J. and Ulrich, H. (2012). A SUMO-interacting motif activates budding yeast ubiquitin ligase Rad18 towards SUMO-modified PCNA. Nucleic acids research $\underline{40}$, 11380–11388.

[Parker et al., 2008] Parker, J. L., Bucceri, A., Davies, A. A., Heidrich, K., Windecker, H. and Ulrich, H. D. (2008). SUMO modification of PCNA is controlled by DNA. The EMBO Journal $\underline{27}$, 2422–2431.

[Parker and Ulrich, 2009] Parker, J. L. and Ulrich, H. D. (2009). Mechanistic analysis of PCNA poly-ubiquitylation by the ubiquitin protein ligases Rad18 and Rad5. EMBO J $\underline{28}$, 3657–3666.

[Parnas et al., 2011] Parnas, O., Amishay, R., Liefshitz, B., Zipin-Roitman, A. and Kupiec, M. (2011). Elg1, the major subunit of an alternative RFC complex, interacts with SUMO-processing proteins. Cell cycle (Georgetown, Tex.) $\underline{10}$, 2894–2903.

[Parnas et al., 2009] Parnas, O., Zipin-Roitman, A., Mazor, Y., Liefshitz, B., Ben-Aroya, S. and Kupiec, M. (2009). The ELG1 clamp loader plays a role in sister chromatid cohesion. PLoS One $\underline{4}$, e5497.

[Parnas et al., 2010] Parnas, O., Zipin-Roitman, A., Pfander, B., Liefshitz, B., Mazor, Y., Ben-Aroya, S., Jentsch, S. and Kupiec, M. (2010). Elg1, an alternative subunit of the RFC clamp loader, preferentially interacts with SUMOylated PCNA. EMBO J $\underline{29}$, 2611–2622.

[Parrilla-Castellar et al., 2004] Parrilla-Castellar, E. R., Arlander, S. J. H. and Karnitz, L. (2004). Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. DNA Repair (Amst) $\underline{3}$, 1009–1014.

References

[Parthasarathy and Murthy, 1999] Parthasarathy, S. and Murthy, M. (1999). On the correlation between the main-chain and side-chain atomic displacement parameters (B values) in high-resolution protein structures. Acta Crystallographica Section D: ... 55, 173–180.

[Parthasarathy and Murthy, 2000] Parthasarathy, S. and Murthy, M. (2000). Protein thermal stability: insights from atomic displacement parameters (B values). Protein engineering 13, 9–13.

[Paz et al., 2000] Paz, Y., Elazar, Z. and Fass, D. (2000). Structure of GATE-16, membrane transport modulator and mammalian ortholog of autophagocytosis factor Aut7p. The Journal of biological chemistry 275, 25445–25450.

[Penengo et al., 2006] Penengo, L., Mapelli, M., Murachelli, A. G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P. P., Polo, S. and Schneider, T. R. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. Cell 124, 1183–1195.

[Perrin et al., 2004] Perrin, A. J., Jiang, X., Birmingham, C. L., So, N. S. Y. and Brumell, J. H. (2004). Recognition of bacteria in the cytosol of Mammalian cells by the ubiquitin system. Curr Biol 14, 806–811.

[Pfander et al., 2005] Pfander, B., Moldovan, G.-L., Sacher, M., Hoege, C. and Jentsch, S. (2005). SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. Nature 436, 428–433.

[Pichler et al., 2002] Pichler, A., Gast, A., Seeler, J., Dejean, A. and Melchior, F. (2002). The nucleoporin RanBP2 has SUMO1 E3 ligase activity. Cell 108, 109–120.

[Pichler et al., 2004] Pichler, A., Knipscheer, P., Saitoh, H., Sixma, T. K. and Melchior, F. (2004). The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type. Nature Structural & Molecular Biology 11.

[Pickart and Eddins, 2004] Pickart, C. and Eddins, M. (2004). Ubiquitin: structures, functions, mechanisms. Biochimica et biophysica acta 1695, 55–72.

[Pickart, 2001] Pickart, C. M. (2001). Mechanisms underlying ubiquitination. Annual Review of Biochemistry 70, 503–533.

[Pierce et al., 1980] Pierce, J. R., Pierce, John Robinson, .-. S. s. and noise (1980). An introduction to information theory : symbols, signals & noise. 2nd, rev. ed edition,

References

New York : Dover Publications. Previously published as: Symbols, signals, and noise. 1961.

[Pinato et al., 2011] Pinato, S., Gatti, M., Scandiuzzi, C., Confalonieri, S. and Penengo, L. (2011). UMI, a novel RNF168 ubiquitin binding domain involved in the DNA damage signaling pathway. Mol Cell Biol 31, 118–126.

[Plechanovova et al., 2011] Plechanovova, A., Jaffray, E. G., McMahon, S. A., Johnson, K. A., Navrátilová, I., Naismith, J. H. and Hay, R. T. (2011). Mechanism of ubiquitylation by dimeric RING ligase RNF4. Nature Structural & Molecular Biology 18, 1052–1059.

[Plechanovova et al., 2012] 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.

[Plyusnin and Holm, 2012] Plyusnin, I. and Holm, L. (2012). Comprehensive comparison of graph based multiple protein sequence alignment strategies. BMC bioinformatics 13, 64.

[Polevoda and Sherman, 2007] Polevoda, B. and Sherman, F. (2007). Methylation of proteins involved in translation. Molecular microbiology 65, 590–606.

[Pommier et al., 2003] Pommier, Y., Redon, C., Rao, V. A., Seiler, J. A., Sordet, O., Takemura, H., Antony, S., Meng, L., Liao, Z., Kohlhagen, G., Zhang, H. and Kohn, K. W. (2003). Repair of and checkpoint response to topoisomerase I-mediated DNA damage. Mutat Res 532, 173–203.

[Pornillos et al., 2002] Pornillos, O., Alam, S. L., Rich, R. L., Myszka, D. G., Davis, D. R. and Sundquist, W. I. (2002). Structure and functional interactions of the Tsg101 UEV domain. EMBO J 21, 2397–2406.

[Potts and Yu, 2005] Potts, P. and Yu, H. (2005). Human MMS21/NSE2 is a SUMO ligase required for DNA repair. Molecular and cellular biology 25, 7021–7032.

[Pouliot et al., 2001] Pouliot, J. J., Robertson, C. A. and Nash, H. A. (2001). Pathways for repair of topoisomerase I covalent complexes in Saccharomyces cerevisiae. Genes Cells 6, 677–687.

[Pouliot et al., 1999] Pouliot, J. J., Yao, K. C., Robertson, C. A. and Nash, H. A. (1999). Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. Science 286, 552–555.

References

[Pourquier et al., 1999] Pourquier, P., Kohlhagen, G., Ueng, L. M. and Pommier, Y. (1999). Topoisomerase I and II Activity Assays. Methods Mol Med $\underline{28}$, 95–110.

[Prag et al., 2003] Prag, G., Misra, S., Jones, E. A., Ghirlando, R., Davies, B. A., Horazdovsky, B. F. and Hurley, J. H. (2003). Mechanism of ubiquitin recognition by the CUE domain of Vps9p. Cell $\underline{113}$, 609–620.

[Prakash and Prakash, 1977] Prakash, L. and Prakash, S. (1977). Isolation and characterization of MMS-sensitive mutants of Saccharomyces cerevisiae. Genetics $\underline{86}$, 33–55.

[Prelić et al., 2006] Prelić, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Hennig, L., Thiele, L. and Zitzler, E. (2006). A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics (Oxford, England) $\underline{22}$, 1122–1129.

[Press and Teukolsky, 2002] Press, W. H. and Teukolsky, S. A. (2002). Numerical Recipes in C++. The Art of Scientific Computing. 2nd ed. edition, Cambridge University Press, Cambridge.

[Prilusky et al., 2005] Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. and Sussman, J. L. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics (Oxford, England) $\underline{21}$, 3435–3438.

[Pringa et al., 2001] Pringa, E., Martinez-Noel, G., Muller, U. and Harbers, K. (2001). Interaction of the ring finger-related U-box motif of a nuclear dot protein with ubiquitin-conjugating enzymes. The Journal of biological chemistry $\underline{276}$, 19617–19623.

[Prudden et al., 2007] Prudden, J., Pebernard, S., Raffa, G., Slavin, D. A., Perry, J. J. P., Tainer, J. A., McGowan, C. H. and Boddy, M. N. (2007). SUMO-targeted ubiquitin ligases in genome stability. The EMBO Journal $\underline{26}$, 4089–4101.

[Ptashne, 1992] Ptashne, M. (1992). A Genetic Switch: Phage [lambda] and Higher Organisms. Cell Press.

[Rabiner, 1989] Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE $\underline{77}$, 257–286.

References

[Ravikumar et al., 2010] Ravikumar, B., Sarkar, S., Davies, J., Futter, M., Garcia-Arencibia, M., Green-Thompson, Z., Jimenez-Sanchez, M., Korolchuk, V., Lichtenberg, M., Luo, S., Massey, D., Menzies, F., Moreau, K., Narayanan, U., Renna, M., Siddiqi, F., Underwood, B., Winslow, A. and Rubinsztein, D. (2010). Regulation of mammalian autophagy in physiology and pathophysiology. Physiological reviews 90, 1383–1435.

[Reggiori et al., 2004] Reggiori, F., Tucker, K. A., Stromhaug, P. E. and Klionsky, D. J. (2004). The Atg1-Atg13 complex regulates Atg9 and Atg23 retrieval transport from the pre-autophagosomal structure. Developmental Cell 6, 79–90.

[Reindle et al., 2006] Reindle, A., Belichenko, I., Bylebyl, G., Chen, X., Gandhi, N. and Johnson, E. (2006). Multiple domains in Siz SUMO ligases contribute to substrate selectivity. Journal of cell science 119, 4749–4757.

[Remm et al., 2001] Remm, M., Storm, C. and Sonnhammer, E. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. Journal of molecular biology 314, 1041–1052.

[Reverter and Lima, 2005] Reverter, D. and Lima, C. (2005). Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. Nature 435, 687–692.

[Ripmaster et al., 1992] Ripmaster, T. L., Vaughn, G. P. and Woolford, J. L. J. (1992). A putative ATP-dependent RNA helicase involved in Saccharomyces cerevisiae ribosome assembly. Proc Natl Acad Sci U S A 89, 11131–11135.

[Ripmaster et al., 1993] Ripmaster, T. L., Vaughn, G. P. and Woolford, J. L. J. (1993). DRS1 to DRS7, novel genes required for ribosome assembly and function in Saccharomyces cerevisiae. Mol Cell Biol 13, 7901–7912.

[Roeder and Agarwal, 2000] Roeder, G. and Agarwal, S. (2000). Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. Cell 102, 245–255.

[Romero et al., 2006] Romero, P., Zaidi, S., Fang, Y., Uversky, V., Radivojac, P., Oldfield, C., Cortese, M., Sickmeier, M., LeGall, T., Obradovic, Z. and Dunker, A. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proceedings of the National Academy of Sciences of the United States of America 103, 8390–8395.

References

[Rozenknop et al., 2011] Rozenknop, A., Rogov, V. V., Rogova, N. Y., Lohr, F., Guntert, P., Dikic, I. and Dotsch, V. (2011). Characterization of the interaction of GABARAPL-1 with the LIR motif of NBR1. J Mol Biol <u>410</u>, 477–487.

[Sachdev et al., 2001] Sachdev, S., Bruhn, L., Sieber, H., Pichler, A., Melchior, F. and Grosschedl, R. (2001). PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. Genes & development <u>15</u>, 3088–3103.

[Sagiv et al., 2000] Sagiv, Y., Legesse-Miller, A., Porat, A. and Elazar, Z. (2000). GATE-16, a membrane transport modulator, interacts with NSF and the Golgi v-SNARE GOS-28. EMBO J <u>19</u>, 1494–1504.

[Saiki et al., 1985] Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science (New York, N.Y.) <u>230</u>, 1350–1354.

[Saio et al., 2009] Saio, T., Yokochi, M. and Inagaki, F. (2009). The NMR structure of the p62 PB1 domain, a key protein in autophagy and NF-kappaB signaling pathway. J Biomol NMR <u>45</u>, 335–341.

[Saitoh et al., 1998] Saitoh, H., Sparrow, D., Shiomi, T., Pu, R., Nishimoto, T., Mohun, T. and Dasso, M. (1998). Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. Current biology : CB <u>8</u>, 121–124.

[Sambrook et al., 1989] Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. 2nd ed. edition, New York: Cold Spring Harbor Laboratory Press.

[Sampson et al., 2001] Sampson, D. A., Wang, M. and Matunis, M. J. (2001). The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. Journal of Biological Chemistry <u>276</u>, 21664–21669.

[Sanger et al., 1977] Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America <u>74</u>, 5463–5467.

[Sankoff, 1975] Sankoff, D. (1975). Minimal mutation trees of sequences. SIAM Journal on Applied Mathematics <u>28</u>, 35–42.

References

[Sartorelli et al., 1999] Sartorelli, V., Puri, P., Hamamori, Y., Ogryzko, V., Chung, G., Nakatani, Y., Wang, J. and Kedes, L. (1999). Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. Molecular cell 4, 725–734.

[Satoo et al., 2007] Satoo, K., Suzuki, N. N., Fujioka, Y., Mizushima, N., Ohsumi, Y. and Inagaki, F. (2007). Crystallization and preliminary crystallographic analysis of human Atg4B-LC3 complex. Acta Crystallogr Sect F Struct Biol Cryst Commun 63, 99–102.

[Saveanu et al., 2003] Saveanu, C., Namane, A., Gleizes, P.-E., Lebreton, A., Rousselle, J.-C., Noaillac-Depeyre, J., Gas, N., Jacquier, A. and Fromont-Racine, M. (2003). Sequential protein association with nascent 60S ribosomal particles. Mol Cell Biol 23, 4449–4460.

[Schiestl and Gietz, 1989] Schiestl, R. and Gietz, R. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Current genetics 16, 339–346.

[Schmidt and Müller, 2002] Schmidt, D. and Müller, S. (2002). Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. Proceedings of the National Academy of Sciences of the United States of America 99, 2872–2877.

[Schneider and Stephens, 1990] Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Research 18, 6097–6100.

[Schneider et al., 1986] Schneider, T. D., Stormo, G. D., Gold, L. and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. Journal of Biological Chemistry 188, 415–431.

[Schreiber and Sonnhammer, 2013] Schreiber, F. and Sonnhammer, E. L. (2013). Hieranoid: hierarchical orthology inference. Journal of molecular biology 425, 2072–2081.

[Schu et al., 1993] Schu, P. V., Takegawa, K., Fry, M. J., Stack, J. H., Waterfield, M. D. and Emr, S. D. (1993). Phosphatidylinositol 3-kinase encoded by yeast VPS34 gene essential for protein sorting. Science (New York, N.Y.) 260, 88–91.

[Schwarten et al., 2009] Schwarten, M., Mohrlüder, J., Ma, P., Stoldt, M., Thielmann, Y., Stangler, T., Hersch, N., Hoffmann, B., Merkel, R. and Willbold, D. (2009). Nix directly binds to GABARAP: a possible crosstalk between apoptosis and autophagy. Autophagy 5, 690–698.

References

[Schwarten et al., 2010] Schwarten, M., Stoldt, M., Mohrluder, J. and Willbold, D. (2010). Solution structure of Atg8 reveals conformational polymorphism of the N-terminal domain. Biochem Biophys Res Commun <u>395</u>, 426–431.

[Schwarz et al., 1998] Schwarz, S., Matuschewski, K., Liakopoulos, D., Scheffner, M. and Jentsch, S. (1998). The ubiquitin-like proteins SMT3 and SUMO-1 are conjugated by the UBC9 E2 enzyme. Proceedings of the National Academy of Sciences of the United States of America <u>95</u>, 560–564.

[Schwienhorst et al., 2000] Schwienhorst, I., Johnson, E. and Dohmen, R. (2000). SUMO conjugation and deconjugation. Molecular & general genetics : MGG <u>263</u>, 771–786.

[Scott et al., 1997] Scott, S., Baba, M., Ohsumi, Y. and Klionsky, D. (1997). Aminopeptidase I is targeted to the vacuole by a nonclassical vesicular mechanism. The Journal of cell biology <u>138</u>, 37–44.

[Scott et al., 2001] Scott, S., Guan, J., Hutchins, M., Kim, J. and Klionsky, D. (2001). Cvt19 is a receptor for the cytoplasm-to-vacuole targeting pathway. Molecular cell <u>7</u>, 1131–1141.

[Scott et al., 2000] Scott, S. V., Nice, D. C., Nau, J. J., Weisman, L. S., Kamada, Y., Keizer-Gunnink, I., Funakoshi, T., Veenhuis, M., Ohsumi, Y. and Klionsky, D. J. (2000). Apg13p and Vac8p are part of a complex of phosphoproteins that are required for cytoplasm to vacuole targeting. The Journal of biological chemistry <u>275</u>, 25840–25849.

[Seibenhener et al., 2007] Seibenhener, M. L., Geetha, T. and Wooten, M. W. (2007). Sequestosome 1/p62–more than just a scaffold. FEBS Lett <u>581</u>, 175–179.

[Sekiyama et al., 2008] 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 <u>283</u>, 35966–35975.

[Shannon, 1948] Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal <u>27</u>, 379–423, 623–656.

[Shen et al., 1996] Shen, Z., Pardington-Purtymun, P., Comeaux, J., Moyzis, R. and Chen, D. (1996). UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins. Genomics <u>36</u>, 271–279.

References

[Shih et al., 2003] Shih, S. C., Prag, G., Francis, S. A., Sutanto, M. A., Hurley, J. H. and Hicke, L. (2003). A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. EMBO J $\underline{22}$, 1273–1281.

[Shintani et al., 2002] Shintani, T., Huang, W.-P., Stromhaug, P. and Klionsky, D. (2002). Mechanism of cargo selection in the cytoplasm to vacuole targeting pathway. Developmental cell $\underline{3}$, 825–837.

[Shintani et al., 1999] Shintani, T., Mizushima, N., Ogawa, Y., Matsuura, A., Noda, T. and Ohsumi, Y. (1999). Apg10p, a novel protein-conjugating enzyme essential for autophagy in yeast. The EMBO journal $\underline{18}$, 5234–5241.

[Shuai and Liu, 2005] Shuai, K. and Liu, B. (2005). Regulation of gene-activation pathways by PIAS proteins in the immune system. Nat Rev Immunol $\underline{5}$, 593–605.

[Sikorski and Hieter, 1989] Sikorski, R. and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics $\underline{122}$, 19–27.

[Sippl, 1990] Sippl, M. (1990). Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. Journal of molecular biology $\underline{213}$, 859–883.

[Slater, 1973] Slater, M. L. (1973). Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle. J Bacteriol $\underline{113}$, 263–270.

[Smith and Waterman, 1981] Smith, T. and Waterman, M. (1981). Identification of common molecular subsequences. Journal of molecular biology $\underline{147}$, 195–197.

[Sneath, 1977] Sneath, P. H. A. (1977). A method for testing the distinctness of clusters: A test of the disjunction of two clusters in Euclidean space as measured by their overlap. Journal of the International Association for Mathematical Geology $\underline{9}$, 123–143.

[Sneath and Sokal, 1973] Sneath, P. H. A. and Sokal, R. R. (1973). Numerical taxonomy: the principles and practice of numerical classification. Freeman, San Francisco, USA.

[Song et al., 2004] 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. Proceedings of the National Academy of Sciences of the United States of America $\underline{101}$, 14373–14378.

References

[Song et al., 2005] Song, J., Zhang, Z., Hu, W. and Chen, Y. (2005). Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. The Journal of biological chemistry 280, 40122–40129.

[Sonnhammer et al., 1997] Sonnhammer, E. L., Eddy, S. R. and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins: Structure, Function, and Bioinformatics 28, 405–420.

[Sonnhammer and Hollich, 2005] Sonnhammer, E. L. and Hollich, V. (2005). Scoredist: A simple and robust protein sequence distance estimator. BMC Bioinformatics 6, 108.

[Stangler et al., 2002] Stangler, T., Mayr, L. M. and Willbold, D. (2002). Solution structure of human GABA(A) receptor-associated protein GABARAP: implications for biolgoical funcrion and its regulation. J Biol Chem 277, 13363–13366.

[Stark et al., 1989] Stark, W., Sherratt, D. and Boocock, M. (1989). Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. Cell 58, 779–790.

[Stelter and Ulrich, 2003] Stelter, P. and Ulrich, H. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. Nature 425, 188–191.

[Struhl and Davis, 1980] Struhl, K. and Davis, R. W. (1980). A physical, genetic and transcriptional map of the cloned his3 gene region of Saccharomyces cerevisiae. J Mol Biol 136, 309–332.

[Strumberg et al., 2000] Strumberg, D., Pilon, A. A., Smith, M., Hickey, R., Malkas, L. and Pommier, Y. (2000). Conversion of topoisomerase I cleavage complexes on the leading strand of ribosomal DNA into 5'-phosphorylated DNA double-strand breaks by replication runoff. Mol Cell Biol 20, 3977–3987.

[Strunnikov et al., 2001] Strunnikov, A., Aravind, L. and Koonin, E. (2001). Saccharomyces cerevisiae SMT4 encodes an evolutionarily conserved protease with a role in chromosome condensation regulation. Genetics 158, 95–107.

[Subramanian et al., 2005] Subramanian, A., Weyer-Menkhoff, J., Kaufmann, M. and Morgenstern, B. (2005). DIALIGN-T: an improved algorithm for segment-based multiple sequence alignment. BMC bioinformatics 6, 66.

References

[Subramanian and Kaufmann, 2008] Subramanian, A. R. and Kaufmann, M. (2008). DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. Algorithms Mol ... 3.

[Sugawara et al., 2004] Sugawara, K., Suzuki, N., Fujioka, Y., Mizushima, N., Ohsumi, Y. and Inagaki, F. (2004). The crystal structure of microtubule-associated protein light chain 3, a mammalian homologue of Saccharomyces cerevisiae Atg8. Genes to cells : devoted to molecular & cellular mechanisms 9, 611–618.

[Sun et al., 2007] Sun, H., Leverson, J. D. and Hunter, T. (2007). Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. EMBO J 26, 4102–4112.

[Sundquist et al., 2004] Sundquist, W. I., Schubert, H. L., Kelly, B. N., Hill, G. C., Holton, J. M. and Hill, C. P. (2004). Ubiquitin recognition by the human TSG101 protein. Mol Cell 13, 783–789.

[Suzuki et al., 2001] Suzuki, K., Kirisako, T., Kamada, Y., Mizushima, N., Noda, T. and Ohsumi, Y. (2001). The pre-autophagosomal structure organized by concerted functions of APG genes is essential for autophagosome formation. The EMBO journal 20, 5971–5981.

[Suzuki et al., 2010] Suzuki, K., Kondo, C., Morimoto, M. and Ohsumi, Y. (2010). Selective transport of alpha-mannosidase by autophagic pathways: identification of a novel receptor, Atg34p. J Biol Chem 285, 30019–30025.

[Suzuki et al., 2007] Suzuki, K., Kubota, Y., Sekito, T. and Ohsumi, Y. (2007). Hierarchy of Atg proteins in pre-autophagosomal structure organization. Genes to cells : devoted to molecular & cellular mechanisms 12, 209–218.

[Swanson et al., 2003] Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L. and Radhakrishnan, I. (2003). Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. EMBO J 22, 4597–4606.

[Takahashi et al., 2001] Takahashi, Y., Kahyo, T., Toh-E, A., Yasuda, H. and Kikuchi, Y. (2001). Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. The Journal of biological chemistry 276, 48973–48977.

References

[Takahashi and Kikuchi, 2005] Takahashi, Y. and Kikuchi, Y. (2005). Yeast PIAS-type Ull1/Siz1 is composed of SUMO ligase and regulatory domains. J Biol Chem <u>280</u>, 35822–35828.

[Talkish et al., 2012] Talkish, J., Zhang, J., Jakovljevic, J., Horsey, E. W. and Woolford, J. L. J. (2012). Hierarchical recruitment into nascent ribosomes of assembly factors required for 27SB pre-rRNA processing in Saccharomyces cerevisiae. Nucleic Acids Res <u>40</u>, 8646–8661.

[Tan et al., 2013] Tan, W., Wang, Z. and Prelich, G. (2013). Physical and Genetic Interactions Between Uls1 and the Slx5-Slx8 SUMO-Targeted Ubiquitin Ligase. G3 (Bethesda) <u>3</u>, g3.113.005827v1.

[Tanida et al., 2003] Tanida, I., Komatsu, M., Ueno, T. and Kominami, E. (2003). GATE-16 and GABARAP are authentic modifiers mediated by Apg7 and Apg3. Biochem Biophys Res Commun <u>300</u>, 637–644.

[Tanida et al., 1999] Tanida, I., Mizushima, N., Kiyooka, M., Ohsumi, M., Ueno, T., Ohsumi, Y. and Kominami, E. (1999). Apg7p/Cvt2p: A novel protein-activating enzyme essential for autophagy. Molecular biology of the cell <u>10</u>, 1367–1379.

[Tateno et al., 1997] Tateno, Y., Ikeo, K., Imanishi, T., Watanabe, H., Endo, T., Yamaguchi, Y., Suzuki, Y., Takahashi, K., Tsunoyama, K., Kawai, M., Kawanishi, Y., Naitou, K. and Gojobori, T. (1997). Evolutionary motif and its biological and structural significance. Journal of molecular evolution <u>44 Suppl 1</u>.

[Tatham et al., 2008] Tatham, M., Geoffroy, M.-C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E., Palvimo, J. and Hay, R. (2008). RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nature cell biology <u>10</u>, 538–546.

[Tatham et al., 2005] Tatham, M., Kim, S., Jaffray, E., Song, J., Chen, Y. and Hay, R. (2005). Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. Nature structural & molecular biology <u>12</u>, 67–74.

[Tatsuya Ii and Brill, 2007] Tatsuya Ii, J. F. J. R. M. and Brill, S. J. (2007). The Yeast Slx5-Slx8 DNA Integrity Complex Displays Ubiquitin Ligase Activity. Cell cycle (Georgetown, Tex.) <u>6</u>, 2800–2809.

References

[Thomas and Dill, 1996] Thomas, P. and Dill, K. (1996). An iterative method for extracting energy-like quantities from protein structures. Proceedings of the National Academy of Sciences of the United States of America 93, 11628–11633.

[Thompson et al., 1994] Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22.

[Thomson and Esnouf, 2004] Thomson, R. and Esnouf, R. (2004). Prediction of natively disordered regions in proteins using a bio-basis function neural network. Lecture Notes in Computer Sciences 3177, 108–116.

[Thomson et al., 2003] Thomson, R., Hodgman, T. C., Yang, Z. R. and Doyle, A. K. (2003). Characterizing proteolytic cleavage site activity using bio-basis function neural networks. Bioinformatics 9, 1741–1747.

[Thurston et al., 2009] Thurston, T., Ryzhakov, G., Bloor, S., von Muhlinen, N. and Randow, F. (2009). The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. Nature immunology 10, 1215–1221.

[Tompa et al., 2009] Tompa, P., Fuxreiter, M., Oldfield, C. J., Simon, I., Dunker, A. K. and Uversky, V. N. (2009). Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays 31, 328–335.

[Torres-Ramos et al., 2002] Torres-Ramos, C. A., Prakash, S. and Prakash, L. (2002). Requirement of RAD5 and MMS2 for postreplication repair of UV-damaged DNA in Saccharomyces cerevisiae. Molecular and Cellular Biology 22, 2419–2426.

[Tsao et al., 1993] Tsao, Y. P., Russo, A., Nyamuswa, G., Silber, R. and Liu, L. F. (1993). Interaction between replication forks and topoisomerase I-DNA cleavable complexes: studies in a cell-free SV40 DNA replication system. Cancer Res 53, 5908–5914.

[Ubersax and Ferrell, 2007] Ubersax, J. and Ferrell, J. (2007). Mechanisms of specificity in protein phosphorylation. Nature reviews. Molecular cell biology 8, 530–541.

[Uversky et al., 2000] Uversky, V. N., Gillespie, J. R. and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: Structure, Function, and Bioinformatics 41, 415–427.

References

[Uzunova et al., 2007] Uzunova, K., Göttsche, K., Miteva, M., Weisshaar, S., Glanemann, C., Schnellhardt, M., Niessen, M., Scheel, H., Hofmann, K., Johnson, E., Praefcke, G. and Dohmen, R. (2007). Ubiquitin-dependent proteolytic control of SUMO conjugates. The Journal of biological chemistry 282, 34167–34175.

[Uzunova, 2006] Uzunova, K. M. (2006). Proteolytic control of SUMO conjugates. PhD thesis, University of Cologne, Institute for Genetics Zülpicher Straße 47a, 50674 Cologne.

[Vijay-Kumar et al., 1987] Vijay-Kumar, S., Bugg, C. and Cook, W. (1987). Structure of ubiquitin refined at 1.8 A resolution. Journal of molecular biology 194, 531–544.

[von Muhlinen et al., 2012] von Muhlinen, N., Akutsu, M., Ravenhill, B. J., Foeglein, A., Bloor, S., Rutherford, T. J., Freund, S. M. V., Komander, D. and Randow, F. (2012). LC3C, bound selectively by a noncanonical LIR motif in NDP52, is required for antibacterial autophagy. Mol Cell 48, 329–342.

[von Muhlinen et al., 2010] von Muhlinen, N., Thurston, T., Ryzhakov, G., Bloor, S. and Randow, F. (2010). NDP52, a novel autophagy receptor for ubiquitin-decorated cytosolic bacteria. Autophagy 6, 288–289.

[Wach et al., 1994] Wach, A., Brachat, A., Pöhlmann, R. and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast (Chichester, England) 10, 1793–1808.

[Wampler, 1997] Wampler, J. E. (1997). Distribution analysis of the variation of B-factors of X-ray crystal structures: Temperature and structural variations in lysozyme. Journal of Chemical Information and Modeling 37, 1171–1180.

[Wang et al., 1999] Wang, H., Bedford, F. K., Brandon, N. J., Moss, S. J. and Olsen, R. W. (1999). GABA(A)-receptor-associated protein links GABA(A) receptors and the cytoskeleton. Nature 397, 69–72.

[Wang et al., 2007] Wang, J., Hu, W., Cai, S., Lee, B., Song, J. and Chen, Y. (2007). The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. Molecular cell 27, 228–237.

[Wang, 2002] Wang, J. C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol 3, 430–440.

References

[Wang et al., 2006] Wang, Z., Jones, G. M. and Prelich, G. (2006). Genetic analysis connects SLX5 and SLX8 to the SUMO pathway in Saccharomyces cerevisiae. Genetics 172, 1499–1509.

[Watanabe et al., 2010] Watanabe, Y., Noda, N. N., Kumeta, H., Suzuki, K., Ohsumi, Y. and Inagaki, F. (2010). Selective transport of alpha-mannosidase by autophagic pathways: structural basis for cargo recognition by Atg19 and Atg34. J Biol Chem 285, 30026–30033.

[Waters et al., 2009] Waters, S., Marchbank, K., Solomon, E., Whitehouse, C. and Gautel, M. (2009). Interactions with LC3 and polyubiquitin chains link nbr1 to autophagic protein turnover. FEBS letters 583, 1846–1852.

[Webster et al., 2014] Webster, B. R., Scott, I., Traba, J., Han, K. and Sack, M. N. (2014). Regulation of Autophagy and Mitophagy by Nutrient Availability and Acetylation. Biochim Biophys Acta S1388-1981.

[Weger et al., 2003] Weger, S., Hammer, E. and Engstler, M. (2003). The DNA topoisomerase I binding protein topors as a novel cellular target for SUMO-1 modification: characterization of domains necessary for subcellular localization and sumolation. Experimental cell research 290, 13–27.

[Weger et al., 2005] Weger, S., Hammer, E. and Heilbronn, R. (2005). Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. FEBS letters 579, 5007–5012.

[Weisshaar et al., 2008] Weisshaar, S., Keusekotten, K., Krause, A., Horst, C., Springer, H., Göttsche, K., Dohmen, R. and Praefcke, G. (2008). Arsenic trioxide stimulates SUMO-2/3 modification leading to RNF4-dependent proteolytic targeting of PML. FEBS letters 582, 3174–3178.

[West, 2003] West, S. (2003). Molecular views of recombination proteins and their control. Nature reviews. Molecular cell biology 4, 435–445.

[Westbrook et al., 2003] Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H. (2003). The Protein Data Bank and structural genomics. Nucleic acids research 31, 489–491.

[Weterings and Chen, 2008] Weterings, E. and Chen, D. (2008). The endless tale of non-homologous end-joining. Cell research 18, 114–124.

References

[Widlund and Davis, 2005] Widlund, P. O. and Davis, T. N. (2005). A high-efficiency method to replace essential genes with mutant alleles in yeast. Yeast 22, 769–774.

[Wild et al., 2011] Wild, P., Farhan, H., McEwan, D., Wagner, S., Rogov, V., Brady, N., Richter, B., Korac, J., Waidmann, O., Choudhary, C., Dötsch, V., Bumann, D. and Dikic, I. (2011). Phosphorylation of the autophagy receptor optineurin restricts Salmonella growth. Science (New York, N.Y.) 333, 228–233.

[Wilkinson et al., 2001] Wilkinson, C. R., 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.

[Wilson et al., 2003] Wilson, M. I., Gill, D. J., Perisic, O., Quinn, M. T. and Williams, R. L. (2003). PB1 domain-mediated heterodimerization in NADPH oxidase and signaling complexes of atypical protein kinase C with Par6 and p62. Mol Cell 12, 39–50.

[Wright and Dyson, 2009] Wright, P. and Dyson, H. (2009). Linking folding and binding. Current opinion in structural biology 19, 31–38.

[Xie et al., 2007a] Xie, H., Vucetic, S., Iakoucheva, L., Oldfield, C., Dunker, A., Obradovic, Z. and Uversky, V. (2007a). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. Journal of proteome research 6, 1917–1932.

[Xie et al., 2007b] Xie, Y., Kerscher, O., Kroetz, M., McConchie, H., Sung, P. and Hochstrasser, M. (2007b). The yeast Hex3.Slx8 heterodimer is a ubiquitin ligase stimulated by substrate sumoylation. The Journal of biological chemistry 282, 34176—-34184.

[Xie et al., 2010] Xie, Y., Rubenstein, E. M., Matt, T. and Hochstrasser, M. (2010). SUMO-independent in vivo activity of a SUMO-targeted ubiquitin ligase toward a short-lived transcription factor. Genes & Development 24, 893–903.

[Xin et al., 2001] Xin, Y., Yu, L., Chen, Z., Zheng, L., Fu, Q., Jiang, J., Zhang, P., Gong, R. and Zhao, S. (2001). Cloning, Expression Patterns, and Chromosome Localization of Three Human and Two Mouse Homologues of GABAA Receptor-Associated Protein. Genomics 74, 408–413.

[Yamaguchi et al., 2010] Yamaguchi, M., Noda, N. N., Nakatogawa, H., Kumeta, H., Ohsumi, Y. and Inagaki, F. (2010). Autophagy-related protein 8 (Atg8) family in-

teracting motif in Atg3 mediates the Atg3-Atg8 interaction and is crucial for the cytoplasm-to-vacuole targeting pathway. J Biol Chem <u>285</u>, 29599–29607.

[Yang et al., 2006] Yang, L., Mullen, J. R. and Brill, S. J. (2006). Purification of the yeast Slx5-Slx8 protein complex and characterization of its DNA-binding activity. Nucleic Acids Research <u>34</u>, 5541–5551.

[Yang and Sharrocks, 2010] Yang, S.-H. and Sharrocks, A. D. (2010). The SUMO E3 Ligase Activity of Pc2 Is Coordinated through a SUMO Interaction Motif. Molecular and Cellular Biology <u>30</u>, 2193–2205.

[Yang et al., 1996] Yang, S. W., Burgin, A. B. J., Huizenga, B. N., Robertson, C. A., Yao, K. C. and Nash, H. A. (1996). A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc Natl Acad Sci U S A <u>93</u>, 11534–11539.

[Yang and Thomson, 2005] Yang, Z. R. and Thomson, R. (2005). Bio-basis function neural network for prediction of protease cleavage sites in proteins. IEEE Trans. Neural Networks <u>16</u>, 263–274.

[Yang et al., 2005] Yang, Z. R., Thomson, R., McNeil, P. and Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics (Oxford, England) <u>21</u>, 3369–3376.

[Yao, 2003] Yao, N. (2003). Replication Factor C Clamp Loader Subunit Arrangement within the Circular Pentamer and Its Attachment Points to Proliferating Cell Nuclear Antigen. Journal of Biological Chemistry <u>278</u>, 50744–50753.

[Yao et al., 2006] Yao, N. Y., Johnson, A., Bowman, G. D., Kuriyan, J. and O'Donnell, M. (2006). Mechanism of proliferating cell nuclear antigen clamp opening by replication factor C. J Biol Chem <u>281</u>, 17528–17539.

[Yeh, 2009] Yeh, E. (2009). SUMOylation and De-SUMOylation: wrestling with life's processes. The Journal of biological chemistry <u>284</u>, 8223–8227.

[Yeh et al., 2010] Yeh, Y.-Y., Wrasman, K. and Herman, P. K. (2010). Autophosphorylation within the Atg1 activation loop is required for both kinase activity and the induction of autophagy in Saccharomyces cerevisiae. Genetics <u>185</u>, 871–882.

References

[Yorimitsu and Klionsky, 2005] Yorimitsu, T. and Klionsky, D. J. (2005). Atg11 links cargo to the vesicle-forming machinery in the cytoplasm to vacuole targeting pathway. Mol Biol Cell <u>16</u>, 1593–1605.

[Yunus and Lima, 2009] Yunus, A. A. and Lima, C. D. (2009). Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. Mol Cell <u>35</u>, 669–682.

[Zhang et al., 2010] Zhang, X.-W., Yan, X.-J., Zhou, Z.-R., Yang, F.-F., Wu, Z.-Y., Sun, H.-B., Liang, W.-X., Song, A.-X., Lallemand-Breitenbach, V., Jeanne, M., Zhang, Q.-Y., Yang, H.-Y., Huang, Q.-H., Zhou, G.-B., Tong, J.-H., Zhang, Y., Wu, J.-H., Hu, H.-Y., De Thé, H., Chen, S.-J. and Chen, Z. (2010). Arsenic trioxide controls the fate of the PML-RARalpha oncoprotein by directly binding PML. Science <u>328</u>, 240–243.

[Zhang and Buchman, 1997] Zhang, Z. and Buchman, A. R. (1997). Identification of a member of a DNA-dependent ATPase family that causes interference with silencing. Molecular and Cellular Biology <u>17</u>, 5461–5472.

[Zhao and Blobel, 2005] Zhao, X. and Blobel, G. (2005). A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. Proceedings of the National Academy of Sciences of the United States of America <u>102</u>, 4777–4782.

[Zheng and Yang, 2004] Zheng, G. and Yang, Y.-C. (2004). ZNF76, a novel transcriptional repressor targeting TATA-binding protein, is modulated by sumoylation. The Journal of biological chemistry <u>279</u>, 42410–42421.

Acknowledgement

Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit — einschließlich Tabellen, Karten und Abbildungen —, die anderen Werken um Wortlaut oder Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch in keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie — abgesehen von unten angegebenen Teilpublikationen — noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Kay Hofmann betreut worden.

☐ Teilpublikationen liegen nicht vor
☒ Nachfolgend genannte Teilpublikationen liegen vor:

Benjamin Vogt and Kay Hofmann (2012). Bioinformatical detection of recognition factors for ubiquitin and SUMO. *Methods Mol. Biol.* <u>832</u>, 249–261.

Tania M. Roberts, Benjamin Vogt, Edyta Siergiuk, Yiming Chang, Caroline Wilson-Zbinden, Vikram Panse, Kay Hofmann and Matthias Peter. Bioinformatic analysis identified the 60S ribosomal maturation factor Rei1 as a cargo adaptor for ribophagy in *Saccharomyces cerevisiae*, *Nat. Cell Biol.*, submitted.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe, und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Angaben dem Dekanat unverzüglich mitzuteilen.

Köln, den 29. April 2014   Benjamin Vogt

<div align="right">

Benjamin Vogt
Im Kleinefeldchen 4 ◆ 53879 Euskirchen

</div>

Persönliche Daten

Geburtsdatum:	20. September 1981
Geburtsort:	Bonn-Duisdorf
Familienstand:	ledig

Promotionsstudium

07.2009 – 04.2014   Anfertigung der Dissertation im Fach Genetik,
Titel: „A bioinformatical approach for a reliable determination of short motifs for SUMO and Atg8 interaction in *Saccharomyces cerevisiae*",
Arbeitsgruppe Professor Dr. Kay Hofmann,
Institut für Genetik der Universität zu Köln

Hochschulstudium

10.2008 – 04.2009   Anfertigung der Diplomarbeit,
Titel: „Synthese und Charakterisierung eines Inhibitors der Dihydroceramid-Desaturase",
Arbeitsgruppe Professor Dr. Konrad Sandhoff,
Kekulé-Institut für organische Chemie und Biochemie der Rheinischen Friedrich Wilhelms-Universität zu Bonn
Diplom (Note: gut)

04.2002 – 04.2009   Diplom-Studiengang (Fachrichtung: Chemie),
Rheinische Friedrich Wilhelms-Universität zu Bonn
Diplom-Vorprüfung (Note: gut)

Schulbildung

08.1992 – 06.2001   Gymnasium Marienschule, Euskirchen
Abitur (Note: 2,2)

09.1989 – 07.1992   Grundschule Franziskusschule, Euskirchen

Zivildienst

08.2001 – 04.2002   Verein für Vormundschaften und Betreuungen,
Diakonisches Werk, Euskirchen

Köln,, den 29. April 2014   Benjamin Vogt