Molecular evolution of Sde2

Bhargesh Patel

A dissertation submitted for the partial fulfilment

of BS-MS dual degree in Science.

Indian Institute of Science Education and Research Mohali

May 2020

Certificate of Examination

This is to certify that the dissertation titled '**Molecular evolution of Sde2**' submitted by Mr. Bhargesh Patel (Reg. No. MS15115) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Shashi Bhushan Pandit Dr. Purnananda Guptasarma Dr. Shravan Kumar Mishra

Date: May 4, 2020

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Shravan Kumar Mishra at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Bhargesh Patel
MS15115
May 4, 2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Shravan Kumar Mishra

Acknowledgements

At the outset, I would like to thank Dr Shravan Kumar Mishra for giving me the great opportunity to work in his lab and gain deeper experience in the fields of cell and molecular biology.

I would also like to thank my lab members – Rakesh, Anupa, Poulami, Kiran, Sugata, Balashanker and Bhavya for their incredible support and fruitful scientific discussions.

I would also like to thank Mr Baidnath Mandal (Vidhya) for his efforts and support in the lab work.

In the end, I would like to thank KVPY for fellowship and Welcome DBT for lab funding.

List of Figures

List of tables

Notations

Sde2-UBL	N-terminal domain of Sde2
Sde2-C	C-terminal domain of Sde2
DUB	Deubiquitinating enzyme
Sp	*Schizosaccharomyces pombe*
Sj	*Schizosaccharomyces japonicus*
At	*Arabidopsis thaliana*
Ce	*Caenorhabditis elegans*
Dm	*Drosophila melanogaster*
Dr	*Danio rerio*
Hs	*homo saipens*
Sc	*Saccharomyces cerevisiae*

Contents

Abstract

Ubiquitination is one of the most common post-translational modification in eukaryotes. Ubiquitin usually regulates protein function by serving as a tag for degradation. A special group of enzymes called deubiquitinating enzymes (DUBs) perform the task of removing ubiquitin moieties from proteins. Recently, DUBs have also been reported to cleave the ubiquitin fold (Sde2-UBL) from the splicing regulator Sde2 in *Schizosaccharomyces pombe* (fission yeast). Such a resemblance in the structural architecture and processing machinery of ubiquitin and Sde2-UBL tempted us to trace the evolutionary origins of the two. Our study suggests the likely possibility that Sde2-UBL might have evolved from ubiquitin by diverging from it as a consequence of minimal selection pressure. In the latter parts of our study, we also attempted to analyse our results in the context of the factors that would determine the specificity of the DUBs towards Sde2-UBL. We infer that the ubiquitin fold of Sde2-UBL itself may not play an important role in determining the specificity of the DUBs, but a meagre 8 amino acid sequence at the C-terminus of Sde2-UBL might be the key player in dictating specificity to the DUBs.

Chapter 1: Introduction

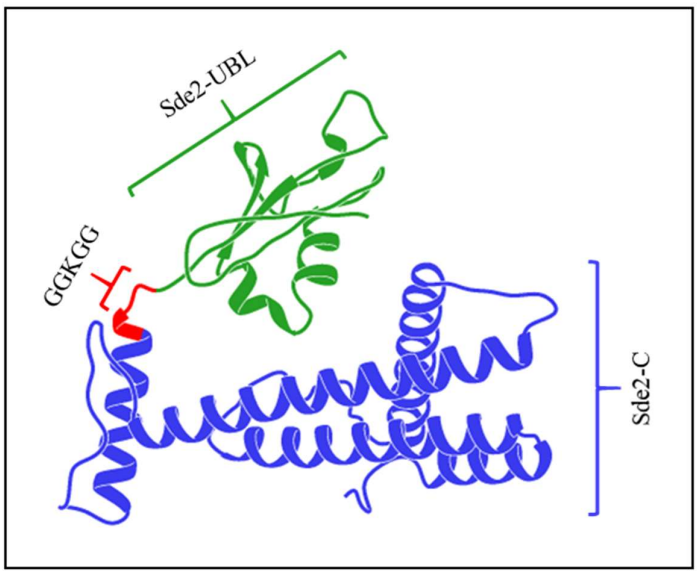


Figure 1.1 Predicted structure of *S. pombe* Sde2

Silencing defective (Sde2) is an important splicing regulator in *Schizosaccharomyces pombe* (fission yeast). It has a β-grasp fold at the N-terminus called Sde2-UBL (Ubiquitin-like) and a loosely folded helical domain at the C-terminus called Sde2-C, connected by a highly conserved GGKGG motif (Fig. 1.1). Sde2-C is the actual catalytic domain of the protein which carries out splicing of a particular subset of pre-mRNAs. The exact molecular function of Sde2-UBL in *S. pombe* is not known. Although, it has been reported that the presence of Sde2-UBL is inhibitory for the association of Sde2-C with the spliceosome, and hence its removal from the precursor protein is essential for activating Sde2-C.[1,2] This is achieved by the processing of Sde2 at the GG~KGG site to release Sde2-C with an N-terminal lysine(K), which can then freely associate with the spliceosome and mediate intron specific pre-RNA splicing of particular genes (Fig. 1.2).[1,3]

Putative orthologs of Sde2 are present in all eukaryotes with important exceptions of *Saccharomyces cerevisiae*, *Candida albicans* and *Pichia pistoris*. A multiple sequence alignment of the orthologs reveals that Sde2-UBL is poorly conserved, Sde2-C is moderately conserved, whereas the GGKGG sequence separating the two, is entirely conserved among eukaryotes.

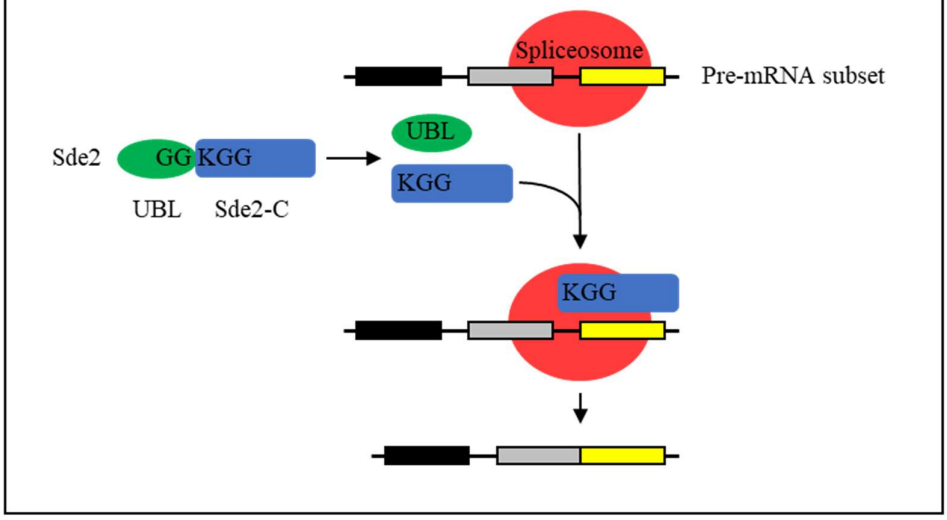


Figure 1.2 Simplified mechanism of Sde2 activation and functioning

The β-grasp fold of Sde2-UBL closely resembles a small, 76 amino acid-long protein called ubiquitin. On comparing Sde2-UBL and ubiquitin, we found that the two have less than 30% protein sequence similarity. But despite this, the structural similarity between ubiquitin and Sde2-UBL exceeds more than 60% (Fig. 1.3). Ubiquitin is present in all eukaryotes, but unlike Sde2-UBL, it is highly conserved among eukaryotes. It is involved in the post-translational modification of proteins by getting conjugated to various proteins.[3,4] The complexity and the extent of ubiquitin conjugation is responsible for regulating the function of the targeted protein and determining its fate.[5]

Ubiquitin is synthesised as a precursor fused with either a ribosomal protein or itself as polyubiquitin. Like Sde2, processing of the ubiquitin precursor is also necessary for the activation of the fused ribosomal protein as well as ubiquitin. A specific group of enzymes called the deubiquitinating enzymes (DUBs) carry out the processing of the ubiquitin precursor.[6] Additionally, as the name suggests, DUBs are also responsible for the removal of ubiquitin moieties from substrate proteins.[7] In *S. pombe*, a total of 20 DUBs have been reported to carry out the process of deubiquitination.[8]

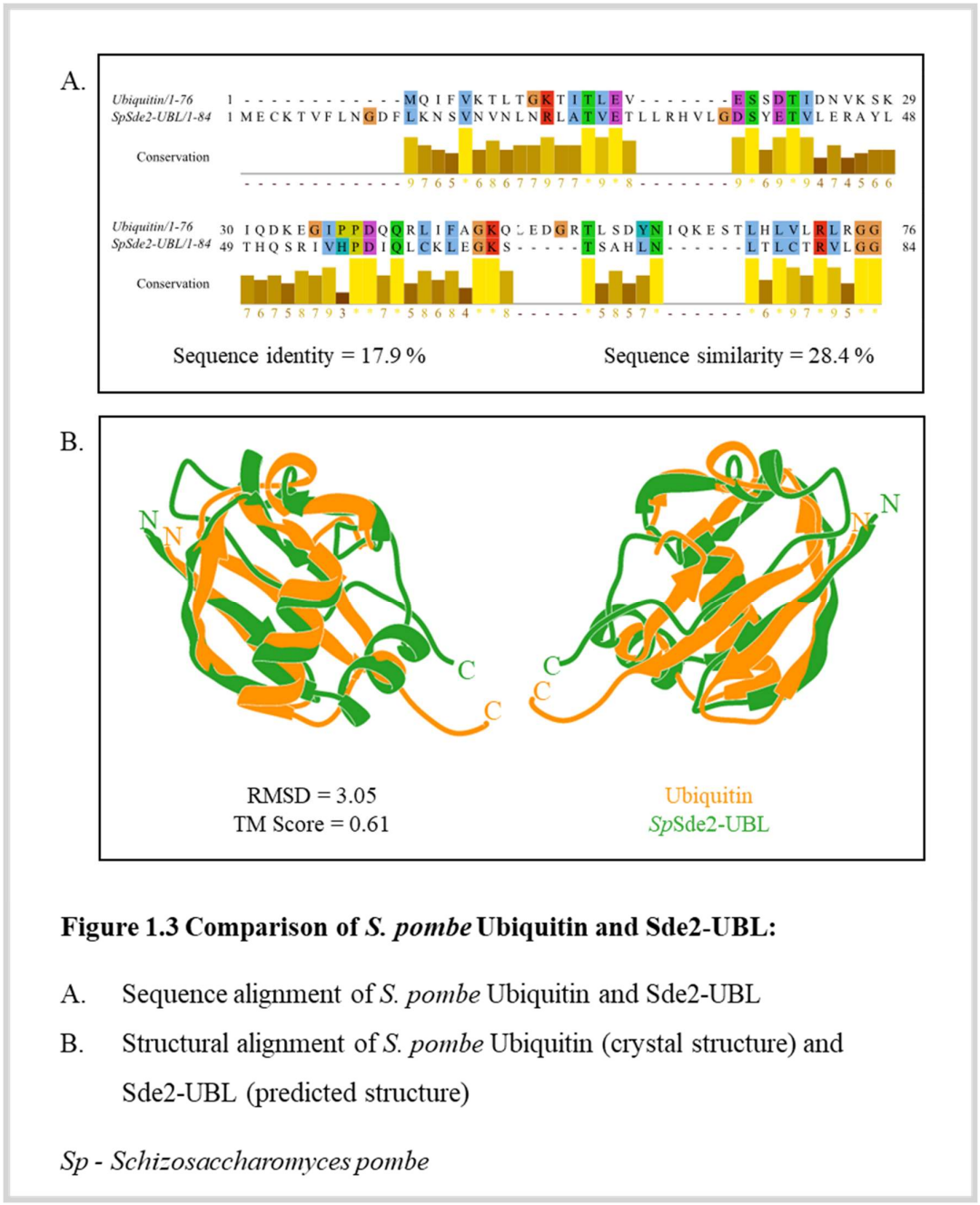


Figure 1.3 Comparison of *S. pombe* Ubiquitin and Sde2-UBL:

A. Sequence alignment of *S. pombe* Ubiquitin and Sde2-UBL

B. Structural alignment of *S. pombe* Ubiquitin (crystal structure) and Sde2-UBL (predicted structure)

Sp - Schizosaccharomyces pombe

In one of our previous studies, we have shown that two paralogous DUBs- Ubp5 and Ubp15 are required for the processing of Sde2 in *S. pombe*.[1] Such a resemblance in the structural architecture and processing machinery of Sde2-UBL and ubiquitin despite poor sequence conservation suggests a common evolutionary relationship between the two. In this study, we aim to gain deeper insights into this evolutionary relationship between ubiquitin and Sde2-UBL as we try to determine the region in the ubiquitin fold that might be responsible for providing specificity to the DUBs towards ubiquitin or Sde2-UBL.

Hypothesis

Given the facts that the same machinery of DUBs can process both ubiquitin as well as Sde2-UBL, and the ubiquitin sequence is highly conserved in eukaryotes as opposed to Sde2-UBL which is poorly conserved, we are hypothesising that Sde2-UBL sequence emerged from, and was the same as ubiquitin at some point of time. This proto-Sde2-UBL was able to get processed by many DUBs, just like ubiquitin. But due to the sheer redundancy in the number of DUBs and ubiquitin genes, the proto-Sde2-UBL diverged from ubiquitin and evolved freely as long as at least one of the DUBs was able to process it (Fig. 1.4).

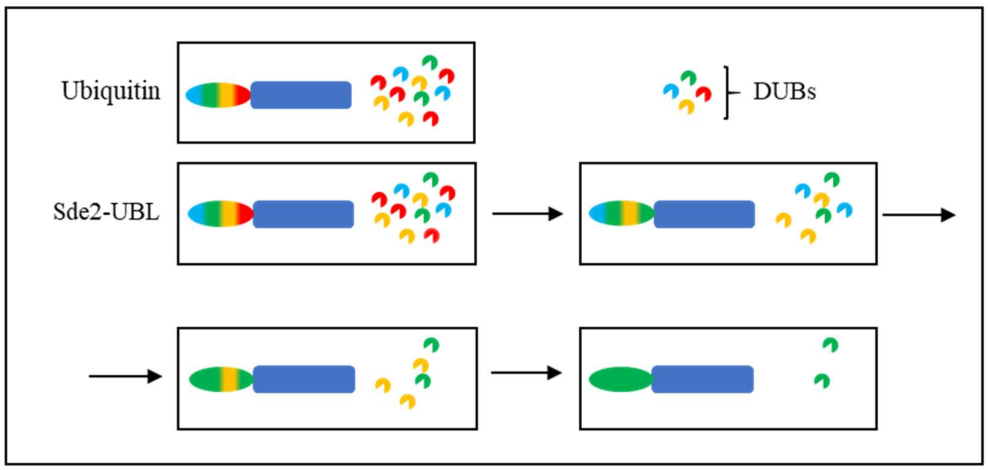


Figure 1.4 Our hypothesis on the molecular evolution of Sde2-UBL: Sde2-UBL sequence was identical to ubiquitin in the beginning and could be processed by many DUBs. With time, the proto-Sde2-UBL diverged from ubiquitin and evolved freely as long as it was able to be processed by at least one DUB.

If our hypothesis is correct, then we would expect very different Sde2-UBL sequences to get processed by the same set of DUBs. Following that, we would also expect Sde2-UBL sequences having higher similarity with ubiquitin to get processed by a larger number of DUBs. Lastly, DUBs other than Ubp5 and Ubp15 should also emerge as new candidates to process the different Sde2-UBL sequences.

With these ideas, we sought to refine the understanding of the evolutionary relationship between ubiquitin and Sde2-UBL.

Chapter 2: Results

Processing of distinct Sde2-UBL sequences in *S. pombe*

In order to obtain distinct Sde2-UBL sequences, we identified six different organisms – *Schizosaccharomyces japonicus (Sj)*, *Arabidopsis thaliana (At)*, *Caenorhabditis elegans (Ce)*, *Drosophila melanogaster (Dm)*, *Danio rerio (Dr)*, and *Homo sapiens (Hs)* whose Sde2-UBL sequences showed low homology to ubiquitin as well as *S. pombe* Sde2-UBL (Fig. 2.1). There are to paralogs of Sde2 in *A. thaliana* – AT3G06455 and ATG01000. We have used Sde2-UBL from the latter for our analysis. We fused these six Sde2-UBL sequences with Sde2-C from *S. pombe (Sp)* to negate any artefacts from variable C-termini.

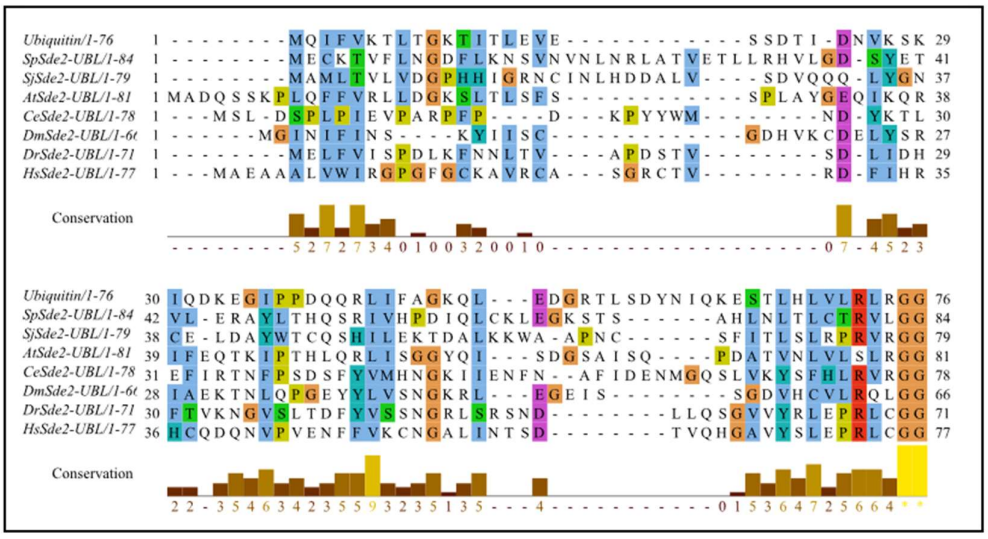


Figure 2.1 Alignment of the different Sde2-UBL sequences used for analysis

Sp - Schizosaccharomyces pombe, Sj - Schizosaccharomyces japonicus, At -Arabidopsis thaliana, Ce - Caenorhabditis elegans, Dm - Drosophila melanogaster, Dr - Danio rerio, Hs - Homo sapiens

Following this, a PhD student in the lab, Rakesh performed complementation of the *S. pombe Δsde2* strain using these chimeric Sde2s. We found that only the *At*Sde2-UBL-*Sp*Sde2-C chimera was able to fully complement the phenotype of *Δsde2*. Moreover, we also observed that the *Sj*Sde2-UBL– and *Ce*-Sde2-UBL–*Sp*Sde2-C chimeras complemented the phenotype only partially whereas the *Dr*Sde2-UBL– and *Hs*Sde2-UBL–*Sp*Sde2-C chimeras almost completely did not complement the phenotype of *Δsde2* (Fig. 2.2A).

We tried to check the processing of the chimeras by western blotting. Since the constructs were tagged with a C-terminal FLAG-tag, a simple anti-FLAG blot revealed the processing. We found that there was a fair amount of correspondence in the processivity of the Sde2 chimeras to the *Δsde2* complementation results. The Sde2-UBLs from *S. japonicus*, *A. thaliana*, and *C. elegans* got strongly processed, whereas the Sde2-UBLs from *D. melanogaster, D. rerio*, and *H. sapiens* got very weakly processed (such that the processed form was detectable only at high exposures) in *S. pombe* (Fig. 2.2B).

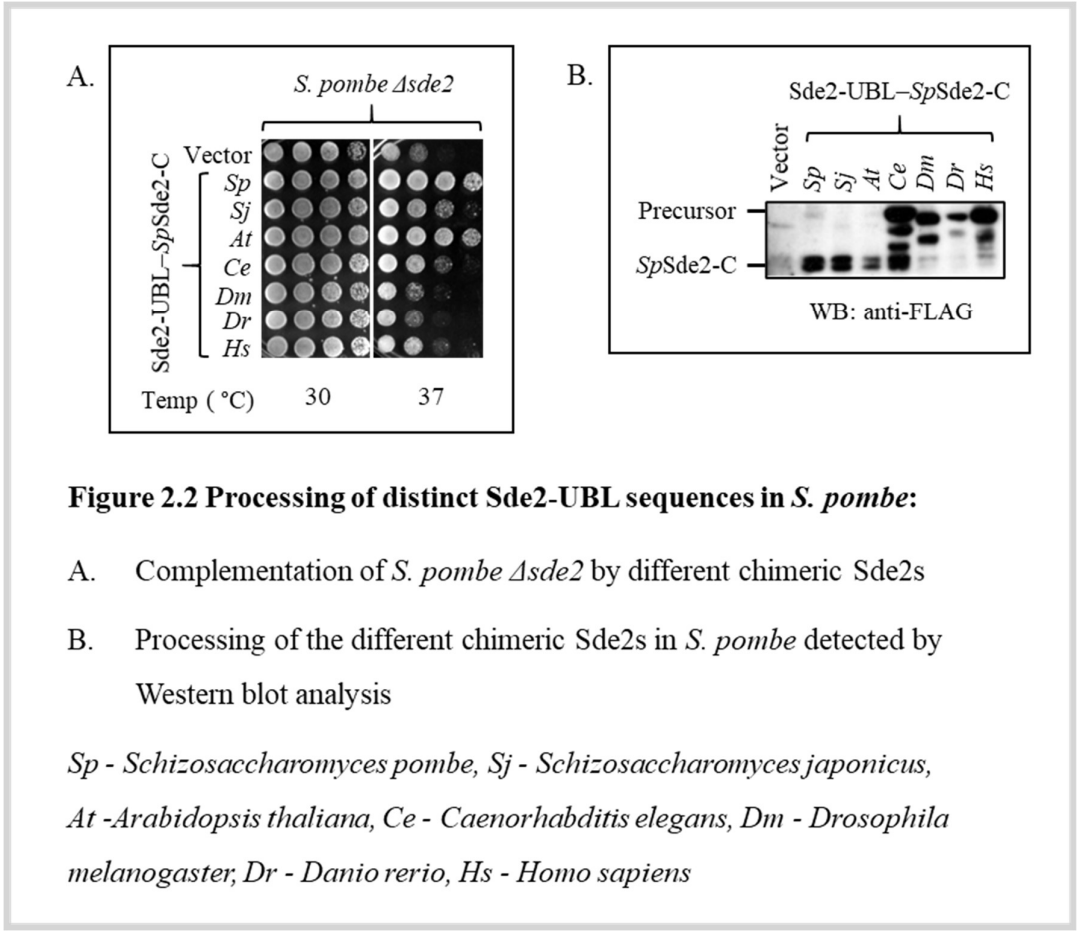


Figure 2.2 Processing of distinct Sde2-UBL sequences in *S. pombe*:

A. Complementation of *S. pombe Δsde2* by different chimeric Sde2s

B. Processing of the different chimeric Sde2s in *S. pombe* detected by Western blot analysis

Sp - Schizosaccharomyces pombe, Sj - Schizosaccharomyces japonicus, At -Arabidopsis thaliana, Ce - Caenorhabditis elegans, Dm - Drosophila melanogaster, Dr - Danio rerio, Hs - Homo sapiens

Identifying the *S. pombe* DUBs that process each of the Sde2 chimeras

We next wanted to find out the individual *S. pombe* DUBs which process each of the Sde2 chimeras. Since there are a large number of DUBs in *S. pombe* (20), it becomes challenging to use an *in vivo* system for this purpose. So, we used an *in vitro* expression system to overcome this obstacle. I created the *S. pombe* DUB library in a bacterial expression vector(pCDFDuet-1) and simultaneously, also cloned the Sde2 chimeras in a separate, compatible bacterial expression vector(pET28a). Following this, I co-expressed each of the chimeras with every DUB in a one-to-one fashion, in *E. coli*. Processing of the chimeras was checked by western blot (Fig. 2.3). One DUB, Ubp14 was excluded from this screen since there were some difficulties in cloning it and the clone could not be obtained in time. Two additional DUBs - USP7 from *Homo sapiens* (referred hereafter as *Hs*USP7) and Ubp15 from *Saccharomyces cerevisiae* (referred hereafter as *Sc*Ubp15) were also included

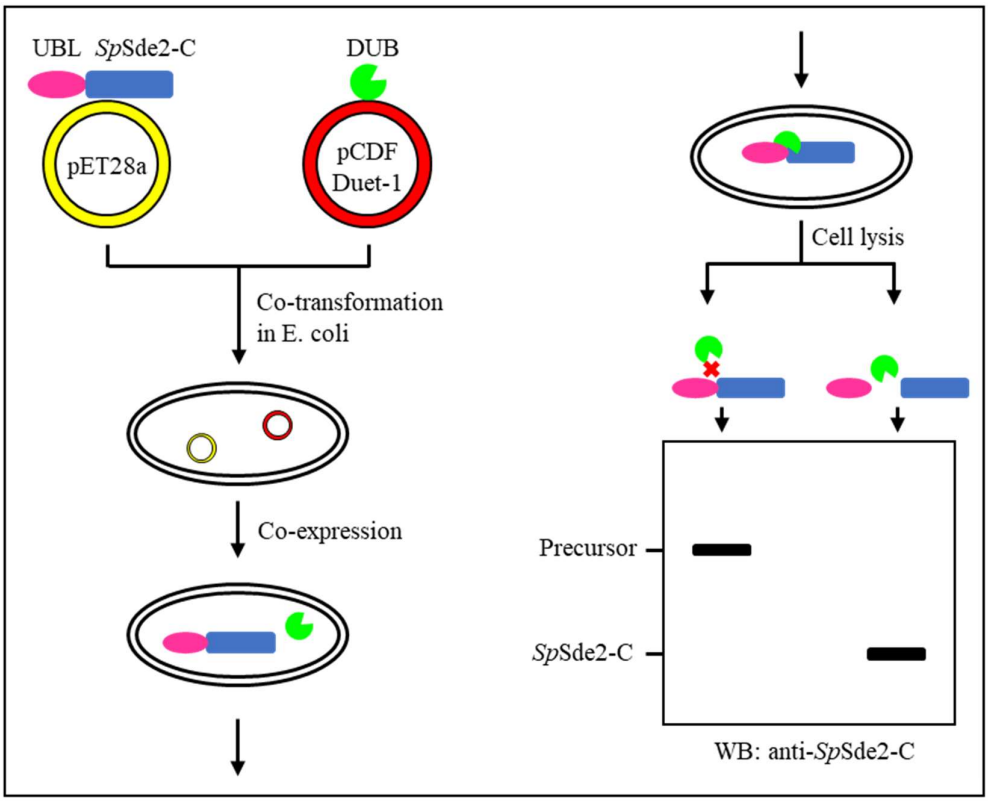
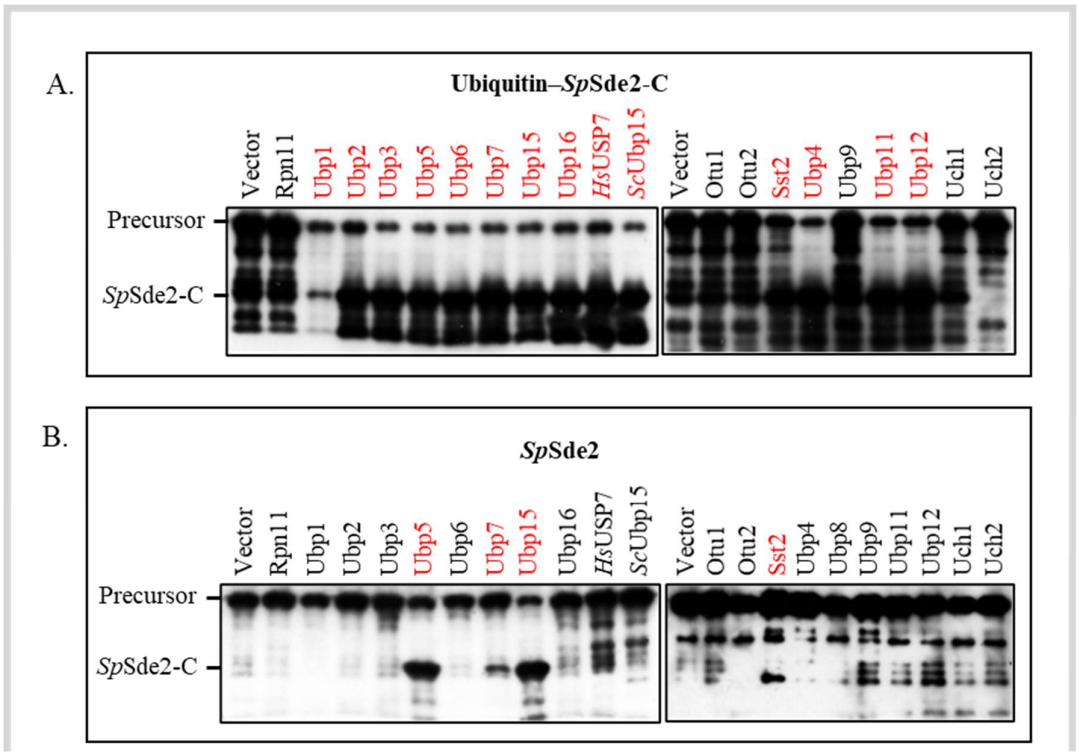


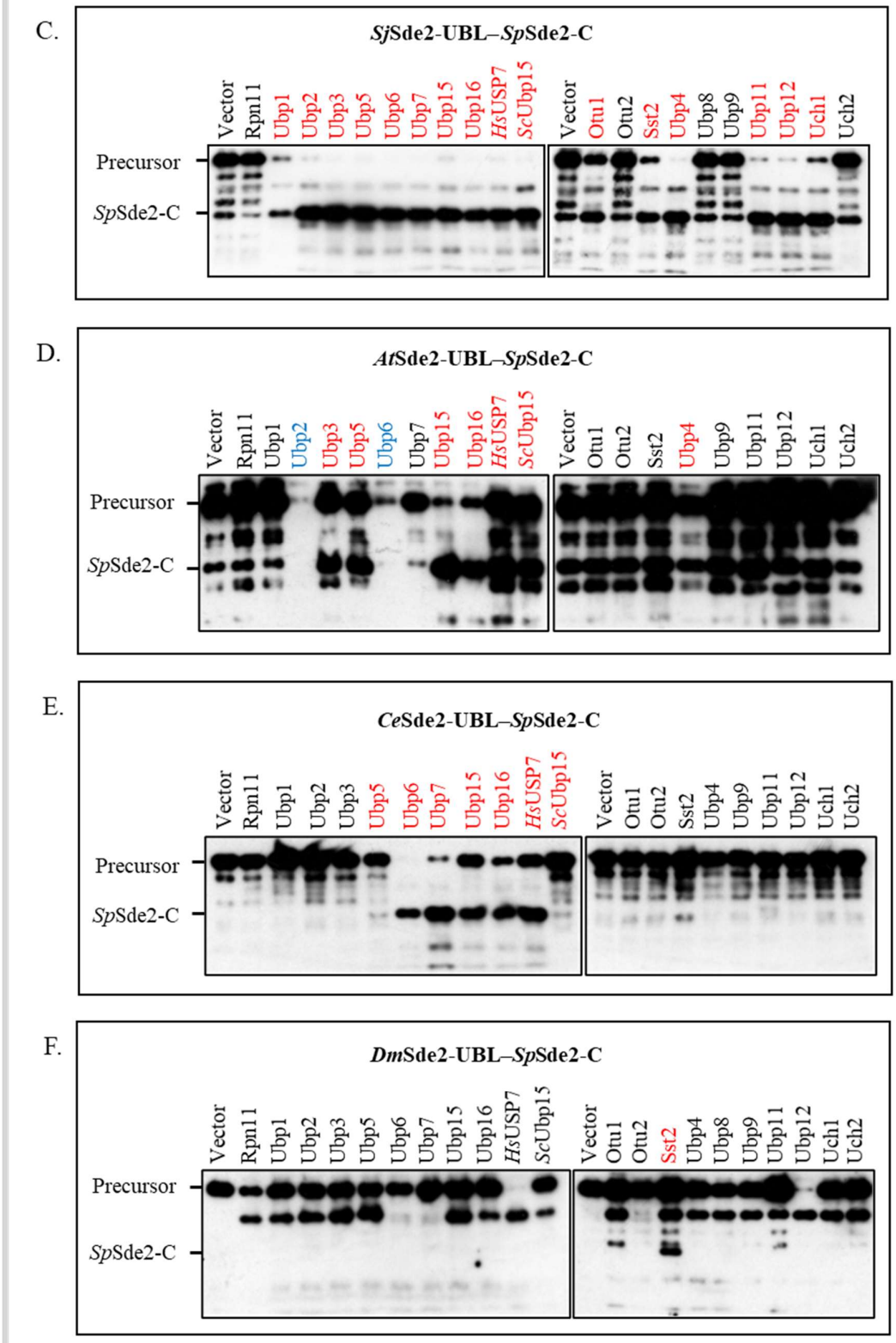
Figure 2.3 Work flow of the experiment to identify the *S. pombe* DUBs that process a particular Sde2 chimera

Each chimera is co-expressed with every DUB separately

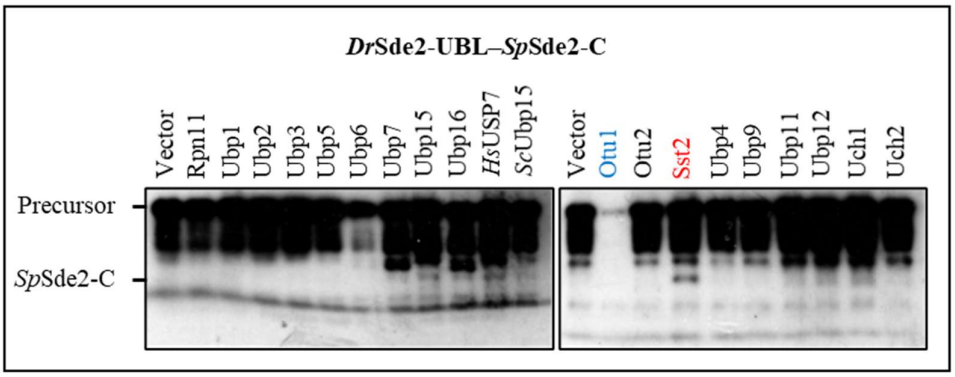
for our analysis since they are close paralogs of the *S. pombe* DUBs Ubp5 and Ubp15, which process *S. pombe* Sde2.

We found that all the chimeras got processed by at least one *S. pombe* DUB *in vitro*. The Sde2-UBL from *S. japonicus* got processed by the greatest number of *S. pombe* DUBs (14) whereas Sde2-UBLs from *D. melanogaster*, *D. rerio*, and *H. sapiens* got processed by the least number of DUBs (1 each). Five DUBs - Rpn11, Otu2, Ubp8, Ubp9, and Uch2 did not show processivity towards any of the chimeras, whereas all the other DUBs process some of the other chimeras (Fig. 2.4). The DUBs' processivity results have been summarised in Table 2.1. Most interestingly, we did not observe any correlation between the sequence similarity of the Sde2-UBLs with ubiquitin and the number of DUBs that are able to process them, as we were expecting from the predictions of our hypothesis (Table 2.2).

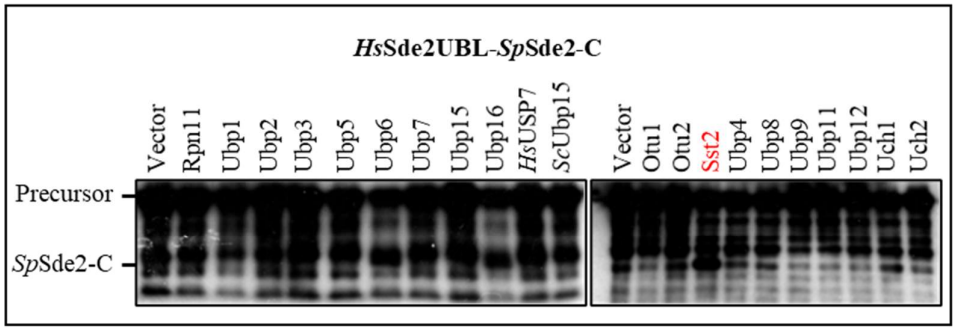



C.
SjSde2-UBL–SpSde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
HsUSP7
ScUbp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp8
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
SpSde2-C
D.
AtSde2-UBL–SpSde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
HsUSP7
ScUbp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
SpSde2-C
E.
CeSde2-UBL–SpSde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
HsUSP7
ScUbp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
SpSde2-C
F.
DmSde2-UBL–SpSde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
HsUSP7
ScUbp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp8
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
SpSde2-C

G.


*Dr*Sde2-UBL–*Sp*Sde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
*Hs*USP7
*Sc*Ubp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
*Sp*Sde2-C

H.


*Hs*Sde2UBL-*Sp*Sde2-C
Vector
Rpn11
Ubp1
Ubp2
Ubp3
Ubp5
Ubp6
Ubp7
Ubp15
Ubp16
*Hs*USP7
*Sc*Ubp15
Vector
Otu1
Otu2
Sst2
Ubp4
Ubp8
Ubp9
Ubp11
Ubp12
Uch1
Uch2
Precursor
*Sp*Sde2-C

I.

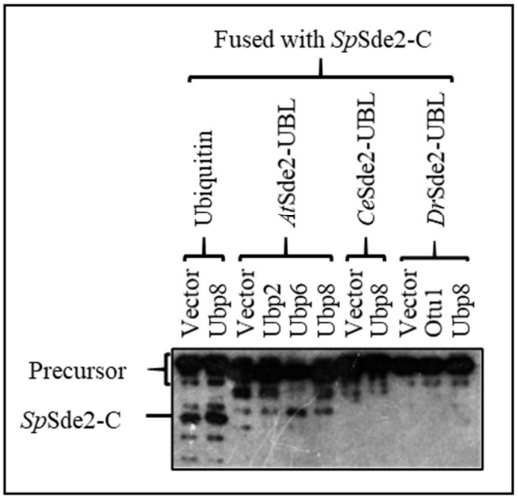

Fused with *Sp*Sde2-C
Ubiquitin
*At*Sde2-UBL
*Ce*Sde2-UBL
*Dr*Sde2-UBL
Vector
Ubp8
Vector
Ubp2
Ubp6
Ubp8
Vector
Ubp8
Vector
Otu1
Ubp8
Precursor
*Sp*Sde2-C

Figure 2.4 Processing of the different chimeric Sde2s by *S. pombe* DUBS

Western blot results for screening of *S. pombe* DUBs to check processivity towards:

A. Ubiquitin–*Sp*Sde2-C

B. *Sp*Sde2

C. *Sj*Sde2-UBL–*Sp*Sde2-C

D. *At*Sde2-UBL–*Sp*Sde2-C

E. *Ce*Sde2-UBL–*Sp*Sde2-C

F. *Dm*Sde2-UBL–*Sp*Sde2-C

G. *Dr*Sde2-UBL–*Sp*Sde2-C

H. *Hs*Sde2-UBL–*Sp*Sde2-C

I. Repeated samples (samples which had inconclusive results or were missed previously)

as detected by anti-*Sp*Sde2-C.

The DUBs in red indicate positive processivity, and the DUBs in blue indicate inconclusive result from the sample.

Sp - Schizosaccharomyces pombe, Sj - Schizosaccharomyces japonicus, At - Arabidopsis thaliana, Ce - Caenorhabditis elegans, Dm - Drosophila melanogaster, Dr - Danio rerio, Hs - Homo sapiens, Sc - Saccharomyces cerevisiae

Sde2-UBL from:	Processing by S. pombe DUBs (*Hs - H. sapiens, Sc - S. cerevisiae*)		
	Strong	**Weak**	**Very weak**
S. pombe	Ubp15	Ubp5	Ubp7, Sst2
S. japonicus	Ubp2, Ubp3, Ubp5, Ubp6, Ubp7, Ubp15, Ubp16, *Hs*USP7, *Sc*Ubp15, Ubp4, Ubp11, Ubp12, Uch1	Ubp1, Otu1, Sst2	
A. thaliana	Ubp15, Ubp16	Ubp3, Ubp5, *Hs*USP7, *Sc*Ubp15	Ubp4
C. elegans	Ubp6, Ubp7, Ubp15, Ubp16, *Hs*USP7		Ubp5, *Sc*Ubp15
D. melanogaster			Sst2
D. rerio			Sst2
H. sapiens			Sst2
S. pombe Ubiquitin	Ubp2, Ubp3, Ubp5, Ubp6, Ubp7, Ubp15, Ubp16, *Hs*USP7, *Sc*Ubp15, Ubp4, Ubp11, Ubp12		Ubp1, Sst2

Table 2.1 Summary of Sde2-UBL processing by *S. pombe* DUBs

Strong processing - Appearance of processed form and diminishing of precursor

Weak processing - Appearance of processed form but no diminishing of precursor

Very weak processing - Faint appearance of processed form

Sde2-UBL from:	Similarity of the Sde2-UBL with Ubiquitin (in %)	No. of S. pombe DUBs that process the Sde2-UBL
A. thaliana	54.1	5
D. melanogaster	45	1
S. japonicus	34.8	14
D. rerio	33.3	1
C. elegans	31.9	5
S. pombe	28.4	4
H. sapiens	14.2	1
S. pombe Ubiquitin	100	12

Table 2.2 Comparison of Sde2-UBL sequence similarity with Ubiquitin and the number of DUBs that process it.

Chapter 3: Discussion

Sde2 processing in the correct subcellular compartment may be essential for the activation of Sde2-C

We found that the six Sde2 chimeras which we used for our analysis, get processed by at least one DUB. Despite this, only a subset of these chimeras complements the phenotype of *Δsde2* in *S. pombe*. This can be explained by considering the subcellular localisation of the specific DUBs that process each of these chimeras. In one of our previous studies, we have shown that Sde2 gets processed inside the nucleus.[1] Since the released active Sde2-C also functions by associating with the spliceosome within the nucleus, this suggests that not just the processing of Sde2, but also the processing of Sde2 in the right subcellular compartment (the nucleus) might be important for its functioning. So, if the DUB(s) that processes a particular chimera is localised at any organelle other than the nucleus, a complete complementation of the *S. pombe Δsde2* phenotype may not happen by that chimera, since it will not get processed within the nucleus.

Both of the DUBs which process the Sde2-UBL from *A. thaliana* strongly (Ubp15 and Ubp16) have been reported to localise in the nucleus[8–10] and as expected, the chimera complements the *S. pombe Δsde2* phenotype almost as good as native *S. pombe* Sde2.

On the other hand, three (Ubp2, Ubp7, and Ubp4) of the eleven *S. pombe* DUBs that strongly process the *S. japonicus* Sde2-UBL have been reported to be localised either in the cytoplasm (Ubp2 and Ubp7)[8,9] or the endosomes (Ubp4)[8]. This would cause a fraction of the precursor pool to be processed in the nucleus, whereas the other fraction will be processed at a different subcellular location. This could explain the poor complementation of the *S. pombe Δsde2* phenotype by the *Sj*Sde2-UBL–*Sp*Sde2-C chimera. The similar result of complementation by the *Ce*Sde2-UBL–*Sp*Sde2-C chimera can also be explained by the same reasoning since one (Ubp7) of the four S. pombe DUBs that strongly process the Sde2-UBL from *C. elegans* also gets localised in the cytoplasm[8,9].

Lastly, none of the S. pombe DUBs strongly process the Sde2-UBLs from either of *D. melanogaster*, *D. rerio*, or *H. sapiens*. The only DUB, which very weakly process them (Sst2), itself gets localised in the endosomes[8], hence explaining an almost complete lack of complementation by the corresponding chimeras.

The type of ubiquitin branching controls the set of DUBs that will process it

We used ubiquitin–SpSde2-C fusion protein in our *in vitro* experiment to check the processivity of *S. pombe* DUBs to serve as a positive control. But we found that 7 DUBs – Rpn11, Otu1, Otu2, Ubp8, Ubp9, Uch1, and Uch2 do not process the linear fusion protein. One reason for this could be that these DUBs might be involved in the cleavage of ubiquitin by recognising the ubiquitin-ubiquitin isopeptide bond (which is formed during polyubiquitination) instead of the peptide bond present in ubiquitin–SpSde2-C. Apart from that, the UCH domain proteins(Uch1 and Uch2 in *S. pombe*) have been reported to have an active site crossover loop which prevents large substrates from passing through.[11] So, it could also be possible that in case of Uch1 and Uch2, the presence of a fused C-terminal domain denies ubiquitin of the chance to bind the catalytic cleft of the DUB and hence, leaving it unprocessed.

Checking the expression of the DUBs by western blotting was also attempted. But since the Sde2 chimeras also had the same 6xHis tag the DUBs were tagged with, the blot appeared too complicated to interpret.

The ubiquitin fold of Sde2 might be playing a little role in determining the specificity of the DUBs towards Sde2-UBL

The observations that six very different Sde2-UBL sequences are able to get processed by the same set of *S. pombe* DUBs, and DUBs other than Ubp5 and Ubp15 are also able to process the different Sde2-UBL sequences support our initial hypothesis. But the absence of correlation between the sequence similarity of Sde2-UBLs with ubiquitin and the number of DUBs that can process them suggests that there is more to it than that. A crucial inference of this observation could be that the high degree of ubiquitin sequence conservation that we observe in the case of eukaryotes, may not be a result of its interactions with DUBs. Ubiquitin also interacts with other complexes like the E1-E2-E3 enzyme

ubiquitin thioester cascade and the proteasomal machinery. It could be the interactions with these complexes that might not have allowed the sequence of ubiquitin to change rather than its interactions with the DUBs. On expanding this idea to Sde2, it is possible that even though Sde2-UBL attains a ubiquitin fold, the fold itself may not be critical in determining the specificity of the DUBs towards Sde2-UBL.

Also, interaction of Sde2-UBL with DUBs is important for the activation of Sde2-C. But following that, the free UBL might not be participating in any of the pathways redundant with ubiquitin. Hence, it would be under no pressure to keep its sequence conserved. This could be the reason for the divergence of Sde2-UBL from ubiquitin rather than the redundancy in the number of DUBs that we originally hypothesised.

A major burden of determining the specificity of DUBs towards Sde2-UBL might lie on the UBL-tail

Through crystal structure studies, it has previously been reported that ubiquitin binds to the DUBs via two hydrophobic patches (Ile44 and Ile36) to offer its C-terminal "tail" to the catalytic cleft of the DUB in the correct orientation.[12,13] In independent studies, many of the C-terminal ubiquitin residues have been reported to play an important role in maintaining proper orientation of the ubiquitin tail inside the catalytic cleft by making essential interactions with residues on the DUBs (e.g., Leu71 and Leu73 of ubiquitin interact with Phe409, Tyr411, Lys420 and Tyr514 of the human DUB USP7[14]; Arg74 forms an extensive network of hydrogen bonds and electrostatic interactions with 11 residues of the human DUB UCH-L3, Arg72 forms salt bridge with Asp33 of UCH-L3[15]). Since hydrophobic interactions are non-specific, they can be offered by any hydrophobic amino acids. In such a scenario, we would expect that the amino acids that comprise the ubiquitin fold might not be crucial in determining the specificity of the DUBs as long as the fold is able to present two hydrophobic patches to the DUB in the correct position. On the other hand, the electrostatic, hydrogen-bonding and other Vander Waals interactions offered by the ubiquitin/Sde2-UBL tail are highly specific and might carry a majority of the burden of determining the specificity of the DUBs towards ubiquitin/Sde2-UBL.

In the future, this proposition can be tested by checking the processivity of *S. pombe* DUBs towards "tail-only" (Sde2-UBL tail–*Sp*Sde2-C) chimeras. If the tail is indeed the major determinant of DUB specificity, then such a construct should get processed by some DUBs at least *in vitro*.

On a closer analysis of the Sde2-UBL sequences from our experiment, we realised that the last eight amino acids (Sde2-UBL tails) showed relatively lesser variation as compared to the amino acids comprising the ubiquitin fold. Further, we noticed that among the non-hydrophobic residues present in the Sde2-UBL tails, arginine was the only amino acid that was conserved. The observation that *S. japonicus* Sde2-UBL gets processed by the greatest number of DUBs (11) suggests that *S. japonicus'* Sde2-UBL tail might be ideal for processing by DUBs. This ideal tail has three arginines, each alternating with a hydrophobic residue, in the sequence. Next, the *C. elegans* Sde2-UBL has 2 arginines in its tail, separated by a hydrophobic residue, and it gets strongly processed by 4 *S. pombe* DUBs. Finally, Sde2-UBLs from *S. pombe*, *A. thaliana*, *D. melanogaster*, *D. rerio*, and *H. sapiens* have only one arginine in their tails and none of these Sde2-UBLs gets processed strongly by more than 2 DUBs. Such a trend suggests that presence of alternating arginines in the Sde2-UBL tails might be largely responsible for determining the specificity of the DUBs towards the Sde2-UBL. Presence of more arginines in the Sde2-UBL tails might increase the specificity of the DUBs towards the Sde2-UBL and presence of a smaller number of arginines in the Sde2-UBL tails might permit only the DUBs with a promiscuous catalytic cleft to be active over the Sde2-UBL (Table 3.1).

Sde2-UBL from:	Sde2-UBL tail sequence	No. of DUBs that strongly process the Sde2-UBL
S. japonicus	L R P R V R G G	11
C. elegans	F H L R V R G G	4
A. thaliana	L V L S L R G G	2
S. pombe	L C T R V L G G	1
D. melanogaster	C V L R Q L G G	0
D. rerio	L E P R L C G G	0
H. sapiens	L E P R L C G G	0
S. pombe Ubiquitin	L V L R L R G G	10

Table 3.1 Possible relation between Sde2-UBL tail sequence and processivity of DUBs

The Sde2-UBLs that have a greater number of arginnies in their tails are strongly processed by a larger number of DUBs

This conjecture can also explain the processivity results of *Hs*USP7 and *Sc*Ubp15. As DUBs, *Hs*USP7 and *Sc*Ubp15 could easily accommodate an Sde2-UBL tail with two or three arginines since it resembles either the ubiquitin tail or the ideal Sde2-UBL tail, respectively. But they may lack the promiscuity in their catalytic clefts to accommodate an Sde2-UBL tail with only one arginine. Hence, these two DUBs show strong processivity towards the Sde2-UBLs from *S. japonicus* and *C. elegans* but either very weak or no processivity towards the Sde2-UBLs from *S. pombe*, *A. thaliana*, *D. melanogaster*, *D. rerio*, or *H. sapiens*.

In the future, this proposition can be tested by transplanting the ideal Sde2-UBL tail from *S. japonicus* to a poorly processed Sde2-UBL like the one from *H. sapiens*. If the arginines in the Sde2-UBL tails are indeed important for determining the specificity of DUBs then such a chimeric Sde2-UBL should get processed by a greater number of DUBs *in vitro*.

Chapter 4: Materials and methods

Complementation assay

Fivefold serial dilutions of cells were spotted on SC-Leu plates. The plates were kept at 30 °C and 37 °C and the cells were allowed to grow. The constructs were expressed under the *nmt81* promoter, which is a thiamine repressible promoter. 5 µg/ml thiamine was already added in the media to repress the promoter and prevent overexpression of the protein.

Protein expression in *S. pombe*

Freshly transformed cells were inoculated in SC-Leu and allowed to grow for 24 hours at 30 °C. A secondary culture was inoculated with 2% primary culture and incubated at 30 °C. 1 OD_{600} of logarithmically growing cells were harvested and stored at -80 °C after flash freezing in liquid nitrogen.

Co-transformation in *E. coli*

The plasmid with Sde2 chimera was transformed in chemical competent *E. coli* BL21 cells. New chemical competent cells were prepared using a colony from the transformed cells. These Sde2 chimera-containing competent cells were then transformed with the plasmid with DUB.

Protein expression in *E. coli*

Freshly co-transformed cells were inoculated in LB with appropriate antibiotics and allowed to grow overnight at 37 °C. A secondary culture was inoculated with 2% primary culture and allowed to grow at 37 °C for 2.5 hours. IPTG was added to the culture at a final concentration of 100 µM and the cells was allowed to incubated at 18 °C for 16 hours. 0.5 OD_{600} cells were harvested and stored at -80 °C after flash freezing in liquid nitrogen.

Western Blot

S. pombe cells were lysed using TCA lysis method and E. coli BL21 cells were lysed using B-PER lysis method. The cell lysates were loaded on SDS-PAGE. Proteins were separated on the gel, and transferred to PVDF membrane using semi-dry transfer assembly. The PVDF membrane was then probed for Sde2-C using either anti-FLAG for yeast lysate, or anti-Sde2-C for bacterial lysate.

Bioinformatics

trRosetta[16] was used for Sde2 structure prediction. TM-align[17] was used to perform structural alignment of ubiquitin and *Sp*Sde2-UBL. EMBOSS Needle and Clustal Omega[18] were used for pair-wise and multiple sequence alignment, respectively. JalView[19] was used to view the protein sequence alignment. UCSF Chimera[20] was used to visualize protein 3D structures.

References

1. Thakran, P. *et al.* Sde2 is an intron-specific pre-mRNA splicing regulator activated by ubiquitin-like processing. *EMBO J.* **37**, 89–101 (2018).

2. Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science (80-.).* **363**, 710 LP – 714 (2019).

3. Chanarat, S. & Mishra, S. K. Emerging Roles of Ubiquitin-like Proteins in Pre-mRNA Splicing. *Trends Biochem. Sci.* **43**, 896–907 (2018).

4. Hershko, A. & Ciechanover, A. THE UBIQUITIN SYSTEM. *Annu. Rev. Biochem.* **67**, 425–479 (1998).

5. Wilkinson, K. D. Ubiquitin-Dependent Signaling: The Role of Ubiquitination in the Response of Cells to Their Environment. *J. Nutr.* **129**, 1933–1936 (1999).

6. Reyes-Turcu, F. E., Ventii, K. H. & Wilkinson, K. D. Regulation and Cellular Roles of Ubiquitin-Specific Deubiquitinating Enzymes. *Annu. Rev. Biochem.* **78**, 363–397 (2009).

7. Wilkinson, K. D. Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* **11**, 1245–1256 (1997).

8. Kouranti, I. *et al.* A Global Census of Fission Yeast Deubiquitinating Enzyme Localization and Interaction Networks Reveals Distinct Compartmentalization Profiles and Overlapping Functions in Endocytosis and Polarity. *PLOS Biol.* **8**, e1000471 (2010).

9. Matsuyama, A. *et al.* ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.* **24**, 841–847 (2006).

10. Richert, K., Schmidt, H., Gross, T. & Käufer, N. The deubiquitinating enzyme Ubp21p of fission yeast stabilizes a mutant form of protein kinase Prp4p. *Mol. Genet. Genomics* **267**, 88–95 (2002).

11. Johnston, S. C., Riddle, S. M., Cohen, R. E. & Hill, C. P. Structural basis for the specificity of ubiquitin C-terminal hydrolases. *EMBO J.* **18**, 3877–3887 (1999).

12. Komander, D. & Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

13. Mevissen, T. E. T. & Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu. Rev. Biochem.* **86**, 159–192 (2017).

14. Hu, M. *et al.* Crystal Structure of a UBP-Family Deubiquitinating Enzyme in Isolation and in Complex with Ubiquitin Aldehyde. *Cell* **111**, 1041–1054 (2002).

15. Misaghi, S. *et al.* Structure of the Ubiquitin Hydrolase UCH-L3 Complexed with a Suicide Substrate. *J. Biol. Chem.* **280**, 1512–1520 (2005).

16. Yang, J. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci.* **117**, 1496 LP – 1503 (2020).

17. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

18. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz268.

19. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

20. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Appendix

Reviewers' comments

Reviewer: Dr. Purnanda Guptasarma

Comment 1:

Structure, and specific residues at specific locations, are conserved more than the overall sequence, in any protein that performs a function (this is well-known). So, if Sde2-UBL's structure is not important at all, and only the C-terminal eight amino acids are important, and it has diverged away greatly from ubiquitin, and Sde2-UBL sequences in different organisms also are quite different, then why is the structure of Sde2-UBL conserved at all? If it is just a piece of protein, like a piece of flesh, why is it there in Sde2 and not something else with the same eight amino acids at the C-terminus. Surely, Sde2-UBL must provide some structural context for the presentation of the eight amino acids-long sequence to the DUBs, or otherwise, it would retain its structural similarity with ubiquitin?

Response:

It has been previously reported that two hydrophobic patches, formed by Ile44 and Ile36 on ubiquitin are important for maintaining the correct orientation of the C-terminal "tail" of ubiquitin in the catalytic cleft of DUBs. Since ubiquitin and Sde2-UBL both use the same machinery for their processing, the Sde2-UBL must at least possess these hydrophobic patches if not any of the other features of ubiquitin. The hydrophobic patches may be offered in the correct conformation only by the ubiquitin fold and not any other fold. This could be a possible reason why the ubiquitin fold of Sde2 did not get lost over evolution, even though it may not be the prime DUB specificity determining factor.

Comment 2:

What happens to the released Sde2-UBL after its removal (which allows Sde2-C to do its work)? Does it do something else? Does it work like ubiquitin? Does it get coupled to ubiquitin and degraded through a ubiquitin-dependent pathway? If so, is it through an isopeptide bond? If not through a ubiquitin-dependent pathway, is it degraded through proteasomes?

Response:

The exact function and degradation mechanism for the released Sde2-UBL are not known. From preliminary experiments in the lab, we have observed that Sde2-UBL does get conjugated to proteins like ubiquitin. But it is very speculative at this stage to comment on its significance.

Comment 3:

Ultimately, what is the connection between RNA processing through splicing, and protein degradation through ubiquitin-dependent pathways? Why is a ubiquitin-like fold used? Clearly, it is not an incidental quirk of nature, and there is suggestive evidence of evolution from ubiquitin, so it is just the use of some domain for a stuffer-role like function? By running-down the role of the ubiquitin fold itself in the cleavage of Sde2-UBL, and thinking of the possibility that just those eight amino acids will work (in a different structural context), one is running down the connection of the ubiquitin-like fold with the RNA processing function. Is this sufficiently established? I think that if the fold is being conserved, there is a purpose (not yet identified).

Response:

One reason for using the same ubiquitin fold in both RNA splicing and protein degradation machineries could be to regulate stress response. Due to an accumulation of misfolded proteins, the cell may downregulate the expression of DUBs to allow for an increased degradation of the misfolded proteins via ubiquitin-dependent pathways. As an additional effect, this could also downregulate the processing of Sde2 and slow down the splicing of Sde2-dependent mRNA transcripts, thereby delaying the production of new proteins.

Regarding the role of Sde2-UBL, we are suggesting that the ubiquitin fold plays a little role in determining the specificity of DUBs towards Sde2-UBL. By that, we do not mean to imply that the ubiquitin fold has no function at all. The free Sde2-UBL itself may likely have some role in the cell, and that could be one of the reasons why it maintained the ubiquitin fold. But it is very speculative at this stage to comment on it.

Strain used

SP20 h+ *JY741 Δsde2::Nat-NT2*

Plasmid list

Stock ID.	Plasmid
D052	pREP81x 3MYC-*Spsde2*-3FLAG
RP01	pREP81x 3MYC-*Sjsde2*-3FLAG
RP02	pREP81x 3MYC-*Atsde2*-3FLAG
RP03	pREP81x 3MYC-*Cesde2*-3FLAG
RP04	pREP81x 3MYC-*Dmsde2*-3FLAG
RP05	pREP81x 3MYC-*Drsde2*-3FLAG
RP06	pREP81x 3MYC-*Hssde2*-3FLAG
D441	pPROEX 6HIS-*ubi1GG-KGGsde2*-C
D067	pPROEX 6HIS-*Spsde2*
BP01	pET28a 6HIS-*Sjsde2*-6HIS
BP02	pET28a *Atsde2-6HIS*
BP03	pET28a 6HIS-*Cesde2*-6HIS
BP04	pET28a 6HIS-*Dmsde2*-6HIS
BP05	pET28a 6HIS-*Drsde2*-6HIS
BP06	pET28a 6HIS-*Hssde2*-6HIS
BP07	pCDFDuet-1 6His-*Sprpn11*
BP08	pCDFDuet-1 6His-*Spubp1*
BP09	pCDFDuet-1 6His-*Spubp2*
BP10	pCDFDuet-1 6His-*Spubp3*
D437	pCDFDuet-1 6His-*Spubp5*
BP11	pCDFDuet-1 6His-*Spubp6*
BP12	pCDFDuet-1 6His-*Spubp7*
D439	pCDFDuet-1 6His-*Spubp15*
D442	pCDFDuet-1 6His-*Spubp16*
D446	pCDFDuet-1 6His-*HsUSP7*
BP13	pCDFDuet-1 6His-*ScUBP15*
BP14	pCDFDuet-1 6His-*Spotu1*
BP15	pCDFDuet-1 6His-*Spotu2*
BP16	pCDFDuet-1 6His-*Spsst2*
BP17	pCDFDuet-1 6His-*Spubp4*
BP18	pCDFDuet-1 6His-*Spubp8*
BP19	pCDFDuet-1 6His-*Spubp9*
BP20	pCDFDuet-1 6His-*Spubp11*
BP21	pCDFDuet-1 6His-*Spubp12*
BP22	pCDFDuet-1 6His-*Spuch1*
BP23	pCDFDuet-1 6His-*Spuch2*