



HAL
open science

Synthèse totale de dimères SUMO-2/3

Jennifer Bouchenna

► **To cite this version:**

Jennifer Bouchenna. Synthèse totale de dimères SUMO-2/3. Médecine humaine et pathologie. Université de Lille, 2020. Français. NNT : 2020LILUS038 . tel-03383471

HAL Id: tel-03383471

<https://theses.hal.science/tel-03383471>

Submitted on 18 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Lille – Droit et Santé
École Doctorale Biologie – Santé de Lille (EDBSL)

THÈSE

Pour obtenir le titre de

Docteur de l'Université de Lille

Spécialité : Chimie organique

Présentée par

Jennifer BOUCHENNA

SYNTHÈSE TOTALE DE DIMÈRES SUMO-2/3

Soutenue le 10 juin 2020

Devant le jury composé de :

Rapporteurs : Dr. Jean-Jacques LACAPERE

Dr. Fabienne BURLINA

Examineur : Dr. Florine CAVELIER

Directeur de thèse : Dr. Oleg MELNYK

Institut de Biologie de Lille, CNRS UMR 8204

Institut Pasteur de Lille

1 Rue du Professeur Calmette 59019 Lille Cedex France

REMERCIEMENTS

Je tiens à remercier l'ensemble des membres de mon jury qui ont accepté de juger mon travail. Je remercie le Dr. Jean-Jacques Lacapère et le Dr. Fabienne Burlina d'avoir accepté d'être mes rapporteurs. Merci au Dr. Florine Cavelier d'avoir accepté de présider mon jury et d'être également mon examinateur.

Je remercie le Pr. Camille Locht pour m'avoir accueillie au sein du CIIL dans l'UMR8204.

A l'issue de la rédaction de ce manuscrit, je suis convaincue que la thèse est loin d'être un travail solitaire. En effet, je n'aurais pas pu réaliser ce travail de thèse sans le soutien d'un bon nombre de personnes dont la générosité, la bonne humeur et l'intérêt manifesté à l'égard de mes recherches m'ont permis d'avancer et de progresser dans cette phase délicate du « futur chercheur ».

En premier lieu, je tiens à remercier tout particulièrement le Dr. Oleg Melnyk pour m'avoir encadré durant ma thèse et donné l'opportunité de réaliser ce travail fascinant mêlant chimie et biologie. Merci de m'avoir guidé durant ces trois années et de m'avoir toujours poussé dans mes retranchements pour pouvoir donner le meilleur de moi-même. Je vous remercie également pour votre disponibilité, vos conseils et le partage de vos connaissances.

J'aimerais aussi remercier le Dr. Jérôme Vicogne pour toute l'aide qu'il m'a apporté en biologie du début jusqu'au dernier point final de ce manuscrit. Merci pour tes conseils, ta patience, ta bonne humeur et ton soutien précieux.

Je remercie également toutes les personnes qui m'ont épaulé dans mes travaux de recherche. Hervé, le dompteur des machines analytiques et le « super-sonic » des mots fléchés, Julien et Reda pour leurs précieux conseils scientifiques.

Je souhaite aussi remercier mes deux acolytes Benoît et Florent pour les moments de fou rire, leur soutien, leur aide et leur écoute dans mes moments de « down ». Une chose est sûre c'est que le Golden Wave se souviendra de nous (Free from desire, 90's).

Je remercie ensuite Nathalie, Magalie (on s'en souviendra du CD6 !), Vangélis, Rémi, Vincent et Olga pour leurs conseils, leur aide et leur écoute. Mais également la team biologie Alex, Stéphanie, Lucile pour leur aide et leur disponibilité.

Enfin, je terminerai par remercier ceux qui me sont chers : mes parents, ma petite-sœur et son fiancé (Et oui !), ma famille dans sa globalité et mon amie Muz qui est devenu un membre à part entière de ma famille. Merci pour votre soutien au quotidien, ce travail de thèse est aussi le votre.

TABLE DES MATIÈRES

REMERCIEMENTS	3
TABLE DES MATIÈRES	4
RÉSUMÉ.....	6
ABSTRACT	7
ÉTUDE BIBLIOGRAPHIQUE	9
I. Les modifications post-traductionnelles	11
A. Les dessous de la grande diversité des modifications post-traductionnelles.....	11
B. L'ubiquitine et sa découverte	14
C. Les modifications post-traductionnelles apparentées à l'ubiquitine	17
II. La SUMOylation	19
A. La découverte des protéines SUMO	19
B. Les différents homologues SUMO	21
1. La protéine SUMO-1.....	23
2. Les protéines SUMO-2 et SUMO-3	24
3. La protéine SUMO-4.....	26
C. La polySUMOylation.....	27
D. Le mécanisme enzymatique de la SUMOylation.....	28
1. La maturation des protéines SUMO.....	29
2. Etape d'activation par l'enzyme E1.....	30
3. Etape de conjugaison par l'enzyme E2.....	31
4. Etape de ligation par les enzymes E3 ligases.....	32
5. Les enzymes impliquées dans la déSUMOylation des protéines	36
6. Motifs régulateurs de la SUMOylation	37
E. Interactions entre la SUMOylation et les autres modifications post-traductionnelles	42
1. SUMOylation versus ubiquitination	42
2. SUMOylation versus phosphorylation	43
3. La SUMOylation versus l'acétylation	44
III. Rôle de la SUMOylation dans la physiologie cellulaire	46
A. SUMO et Cancer	46
1. Leucémie aigüe promyélocytaire	49
2. Cancer de la prostate	49
B. SUMO et le stress cellulaire.....	51
C. SUMO : expression et intégrité génomique	53
D. SUMO et l'inflammation.....	56
E. Bilan/Conclusion.....	57
IV. Objectifs généraux.....	57
V. Synthèse des protéines par voie chimique	58

A. Approches modernes de la synthèse chimique des protéines.....	58
1. La synthèse peptidique sur phase solide	60
2. Le principe de la ligation chimique.....	62
3. Assemblage de grands peptides et de protéines	71
4. Les procédés de synthèse « one pot » de N vers C	75
B. Méthodes pour la synthèse de grandes protéines.....	78
1. Aperçu statistique de la synthèse chimique de protéines.....	79
2. Fréquence des acides aminés protéinogéniques.....	80
3. Les réactions de déchalcogénéation	82
C. Bilan/Conclusion.....	85
RÉSULTATS & DISCUSSIONS	87
I. Synthèse par voie chimique des protéines SUMO-2 et SUMO-3.....	89
A. Formation d'aspartimide	91
1. Formation d'aspartimide pendant la synthèse peptidique en phase solide.....	91
2. Formation d'aspartimide pendant l'étape de ligation chimique.....	94
3. Synthèse des domaines SUMO-2, SUMO-3.....	95
4. Synthèses des domaines SUMO-2, SUMO-3	97
B. Analyses biophysiques et biochimiques des protéines SUMO-2, SUMO-3	99
1. Analyses en dichroïsme circulaire des domaines SUMO-2 et SUMO-3	99
2. Analyses biochimiques : La SUMOylation	103
II. Synthèse totale de conjugués SUMO-2 et SUMO-3.....	108
A. La stratégie convergente « one pot ».....	108
B. Problématique de la désulfuration	110
C. Cinétique de déconjugaison avec SENP1 et SENP2.....	112
III. Synthèse par voie chimique du domaine SUMO-2/3 core.....	114
IV. Synthèse totale des dimères SUMO-2/3	116
A. La stratégie de synthèse générale	117
B. L'assemblage « one pot » des dimères SUMO-2/3.....	120
C. L'étape de désulfuration sélective des dimères SUMO-2/3	122
D. Caractérisations biophysiques et biochimiques du dimère SUMO-2/SUMO-3	124
1. Analyse structurale par dichroïsme circulaire	124
2. Test de déconjugaison par les enzymes SENP1 et SENP2.....	124
CONCLUSION & PERSPECTIVES	127
RÉFÉRENCES BIBLIOGRAPHIQUES	131
ANNEXES	152

RÉSUMÉ

Les progrès du génie génétique ont permis l'expression d'un nombre considérable de protéines dans des cellules pour étudier les mécanismes du vivant, développer des approches thérapeutiques et de biotechnologie. Ces protéines, dites recombinantes, présentent certaines limitations associées à leur mode de production, notamment dans le choix des acides aminés de la séquence qui est limité aux vingt acides aminés naturellement codés par le génome. Une autre limitation de ces méthodes est la difficulté d'incorporer des acides aminés porteurs de modifications post-traductionnelles (MPT). Ces limitations ont motivé les chimistes organiciens à développer de nouvelles méthodologies de synthèse afin d'accéder à ces macromolécules d'intérêt biologique. Si la synthèse chimique s'est clairement imposée depuis quelques années comme une alternative crédible à l'utilisation de méthodes biologiques pour la synthèse de protéines, la synthèse chimique de protéines de grande taille (au-delà de 100 acides aminés) reste un véritable défi.

SUMO est une MPT apparentée à l'ubiquitine constituée d'environ 90 acides aminés. Cinq isoformes différentes de SUMO ont été identifiées à ce jour chez l'Homme. Parmi celles-ci, SUMO-2 et SUMO-3 sont exprimées par tous les types cellulaires et similaires à 97% au niveau de leur séquence. L'objectif de ce projet de thèse a été d'étudier la conformation et les propriétés biochimiques de conjugués SUMO-2 et SUMO-3, ainsi que de réaliser la synthèse totale des différentes combinaisons de dimères SUMO-2 / SUMO-3. La taille de ces constructions moléculaires se situe à la limite haute des protéines pouvant être produite par synthèse chimique. Pour atteindre ces objectifs, nous avons développé des stratégies de synthèse performantes en one-pot et des réactions de désulfuration sélectives permettant d'accéder aux protéines cibles à l'échelle du milligramme.

Nous montrons au cours de ses recherches que les protéines SUMO-2 et SUMO-3 ont des propriétés structurales et biochimiques spécifiques, et ne doivent pas être considérées comme redondantes. Nous décrivons également pour la première fois la synthèse totale de dimères SUMO-2/3. Ce travail devrait aider à mieux comprendre le rôle spécifique des modifications SUMO-2 et/ou SUMO-3 sur la structure et la fonction des protéines qui sont la cible de ces MPT.

ABSTRACT

Advances in genetic engineering enabled the expression of a huge number of proteins in cells to study biological mechanisms and to develop novel therapeutic and biotechnological approaches. Nevertheless, recombinant strategy of protein production has inherent limitations. In particular, the sequence composition is usually limited to the twenty naturally encoded amino acids.

Moreover, the incorporation of amino acids carrying post-translational modifications (PTM) appears difficult. This second limitation motivated organic chemists to develop new synthetic methodologies to access biologically relevant macromolecules. Although chemical synthesis clearly emerged in recent years as a useful alternative to biological methods, the chemical synthesis of large proteins (beyond 100 amino acids) remains a real challenge.

SUMO is an ubiquitin-like PTM consisting of approximately 90 amino acids. Five different SUMO isoforms have been identified in humans so far. Among them SUMO-2 and SUMO-3 are expressed by all cell types and are 97% similar in sequence. The aim of this thesis was to study the conformation and the biochemical properties of SUMO-2 and SUMO-3 conjugates. The total synthesis of the different combinations of SUMO-2 / SUMO-3 dimers was also carried out. The size of these molecules coincides with the upper limit of the proteins that can be produced by chemical synthesis. Considering this, we developed powerful one-pot synthetic strategies and selective desulfurization reactions to access the target proteins on a milligram scale.

During this research, we show that the SUMO-2 and SUMO-3 proteins have specific structural and biochemical properties and should not be considered redundant. We also described the first total synthesis of SUMO-2/3 dimers. This work will help to better understand the specific role of SUMO-2 and/or SUMO-3 modifications on the structure and function of the SUMOylated proteins.

ÉTUDE BIBLIOGRAPHIQUE

I. Les modifications post-traductionnelles

A. Les dessous de la grande diversité des modifications post-traductionnelles

Au cours de l'évolution, les organismes vivants ont développé des mécanismes leur permettant de générer un protéome (ensemble des protéines exprimées dans une cellule) d'une diversité quasiment infinie à partir d'un nombre restreint de gènes. Chez l'Homme, il existe environ 26 500 gènes, ce qui est en de ça du nombre et de la grande diversité de son protéome. En effet, ce protéome semble être deux à trois plus complexe (supérieur à 1 000 000 protéines) que le nombre de protéines prédites par le génome. C'est seulement vingt ans après la détermination de la structure en double hélice de l'ADN que les scientifiques ont découvert que les gènes des organismes eucaryotes étaient « morcelés », c'est-à-dire constitués d'exons (partie codante, une dizaine en moyenne par gène) séparés par des introns (partie non codante). Ainsi, les exons et les introns sont transcrits à partir de l'ADN en un ARN (dit pré-messager, noté pré-ARN). Ensuite, certains introns et/ou exons sont éliminés, puis les exons sont rassemblés pour former l'ARN messager final (ARNm). Ce mécanisme, nommé épissage, a lieu dans le noyau des cellules avant l'export des ARNm « épissés » dans le cytoplasme, où ils y seront finalement traduits en protéines par les ribosomes. L'épissage est donc une source de diversité des protéines produites (Figure 1). Cependant, il existe bien d'autres mécanismes permettant d'augmenter encore les combinatoires du protéome, notamment après la phase de traduction. On parle alors de modifications post-traductionnelles (MPT).

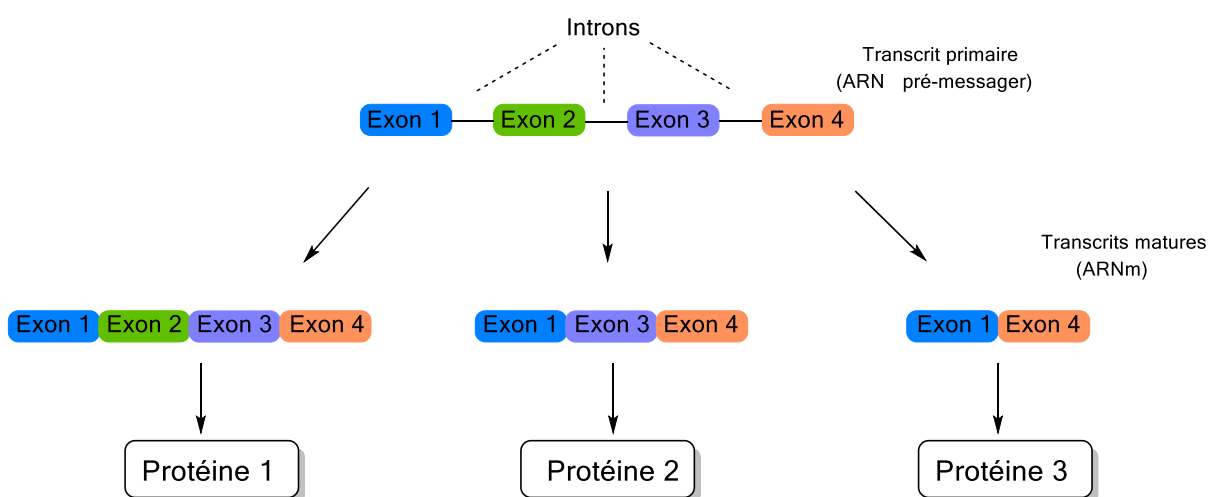


Figure 1. Mécanisme d'épissage des ARN pré-messager en ARN messager mature.

Même si elle existe chez les procaryotes, les MPT sont majoritairement rencontrées chez les eucaryotes. Environ 5% du génome est d'ailleurs dédié à la synthèse des enzymes réalisant ce type de modifications (pour revue voir¹). Cette régulation post-traductionnelle du protéome est un mécanisme cellulaire à faible coût énergétique car il permet une adaptation rapide de la cellule eucaryote à des changements de son environnement sans requérir de nouveaux gènes. Ces modifications covalentes peuvent modifier fortement les propriétés physico-chimiques de la protéine : il peut s'agir de clivages protéolytiques, d'ajout de groupements chimiques ou de polypeptides sur un ou plusieurs acides aminés. Loin d'être un simple changement d'architecture de la protéine, la présence de ces MPT peut impacter sur son activité biologique, sa localisation, sa stabilité, ainsi que ses capacités à interagir avec d'autres partenaires. Ceci impacte directement la transduction du signal, la multiplication cellulaire, l'apoptose ou encore l'autophagie. A l'heure actuelle, la phosphorylation est la MPT la plus étudiée et beaucoup de ses rôles sont quasiment établis. Dans le cas de la signalisation cellulaire, les protéines kinases peuvent être activées ou inactivées par l'ajout de manière réversible de groupements phosphates². Cependant, il existe bien d'autres MPT telles que la méthylation, l'acétylation, l'ubiquitination, la SUMOylation, la O-GlcNAcylation, l'ADP-ribosylation, la nitration, l'ajout d'ancrage lipidique (myristoylation, prénylation, ...), la formation sélective de ponts disulfure et la glycosylation³. Parmi toutes les MPT répertoriées chez les eucaryotes supérieurs, une trentaine d'entre elles sont plus communément rencontrées⁴ (Tableau 1).

Modification post-traductionnelle	Résidu modifié	Caractéristiques de la MPT
SUMOylation	Lys	Ajout d'un peptide d'environ 100 acides aminés par établissement d'une liaison isopeptidique. Possibilité de multi- ou polySUMOylation. Il existe 5 isoformes SUMO.
Ubiquitylation	Lys	Ajout d'un peptide d'environ 76 acides aminés par établissement d'une liaison isopeptidique ou par une liaison peptidique sur l'acide aminé N-terminal de la protéine cible. Possibilité de multi- ou polyubiquitination.
Neddylation	Lys	Addition d'un peptide ubiquitin-like NEDD8.
Méthylation	Lys, His, Arg	Ajout d'un groupement méthyle (-CH ₃) sur des résidus basiques. Augmentation de l'hydrophobicité de la protéine modifiée.
Acétylation	Lys	Ajout d'un groupement acétyle sur l'amine ε. Neutralisation de la charge positive.
Hydroxylation	Lys, Pro	Ajout d'un groupement hydroxyle (-OH). Modification essentielle pour la maturation des fibres de collagène ⁵ .
Sulfatation	Tyr	Transfert d'un groupement sulfate activé (SO ₃ ⁻) sur une protéine acceptrice par une sulfotransférase.
Nitration	Tyr	Addition de groupement nitro (-NO ₂) sur les résidus tyrosine. Permet la dégradation protéasomale des protéines modifiées.
Phosphorylation	Ser, Thr	Ajout d'un groupement phosphate (-PO ₃ ²⁻) par une liaison ester sur des résidus polaires. Apport de charges négatives
Carboxylation	Glu	Modification du glutamate en γ-carboxyglutamate par une γ-glutamylcarboxylase. Protéines impliquées dans la coagulation du sang ⁶ .
C-mannosylation	Trp	Premier résidu Trp de la séquence est modifié par un α-mannopyranose par une liaison C-C. Site consensus : WxxW
Myristoylation	Gly	Ajout d'un groupement myristoyl (acide gras saturé de 14 carbones) sur l'extrémité N-terminale d'un résidu Gly par une liaison amide stable. Réaction catalysée par la myristoyl-CoA: protéine N-myristoyltransférase (NMT).
Prénylation	Cys	Ajout d'un groupement farnésyl pyrophosphate (C15) ou géranylgeranyl pyrophosphate (C20). Réaction catalysée par la farnésyl transférase ou la géranylgeranyl transférase, respectivement.
S-acylation	Cys	On parle aussi de palmitoylation. Ajout d'acides gras sur des résidus cystéine par une famille d'acyltransférases. La réaction inverse est catalysée par les acylthioestérases.
ADP-ribosylation	Arg, Glu, Asp	Ajout d'un ou plusieurs motifs d'ADP-ribose provenant du NAD par l'ADP ribosyltransférase via une liaison ester.
Citrullinisation	Arg	Conversion d'une arginine en citrulline au sein d'une séquence consensus par l'enzyme peptidylarginine deiminase (PAD).

Tableau 1. Principales modifications post-traductionnelles rencontrées et les acides aminés impliqués¹. Il n'y a pas de modification connue à ce jour pour les acides aminés Leu, Ile, Val et Phe.

Certaines de ces MPT sont susceptibles d'être retrouvées systématiquement sur une protéine ou d'y apparaître de manière transitoire. Dans certains cas, la présence d'une MPT peut devenir un signal pour l'ajout d'une autre MPT. C'est l'exemple même de la phosphorylation qui précède bien souvent l'ubiquitination ou la SUMOylation. Cela signifie que les MPT peuvent être greffées de façon unique à la surface de la protéine cible, de façon concomitante sur des résidus différents ou être en compétition pour la modification d'un résidu précis. Par conséquent, étudier les interactions entre les différentes MPT est essentiel pour comprendre leurs régulations et leurs fonctions au sein de la cellule. Il est établi que la dérégulation de certaines MPT est à l'origine de l'apparition ou de la progression de certaines pathologies d'où l'intérêt de mieux comprendre leur rôle au sein de la cellule. Etant donné leur grande diversité et de leurs combinaisons, les MPT sont semblables à des « codes-barres moléculaires » qu'on se doit de décrypter pour mieux comprendre les mécanismes du vivant et découvrir de nouveaux traitements thérapeutiques. Mon travail de thèse s'est attaché à l'étude d'une MPT spécifique, semblable à l'ubiquitination ou « Ubiquitin-like » c'est-à-dire la SUMOylation.

B. L'ubiquitine et sa découverte

L'ubiquitination est la première MPT peptidique ayant été décrite dans la littérature scientifique. Cette MPT consiste à l'ajout d'un polypeptide de 8,5 kDa, l'ubiquitine (Ub), sur une protéine cible. Initialement découverte par Goldstein *et al.* en 1975⁷, ce n'est qu'au début des années 1980 que les travaux d'Aaron Hershko, Avram Ciechanover et Irwin Rose ont permis la caractérisation de la cascade enzymatique permettant cette modification. Les études ayant menées à la découverte de l'ubiquitine visaient à étudier une voie de dégradation non lysosomale ATP-dépendante sur un modèle de réticulocytes⁸. Le fait que cette voie utilise de l'énergie pour dégrader les protéines était en soi intrigant, puisque la rupture d'une liaison peptidique classique est une réaction thermodynamiquement favorisée. Partant de ce postulat, cela suggérait que l'énergie utilisée devait servir à autre chose qu'à l'hydrolyse proprement dite des protéines. C'est donc en essayant de comprendre ce paradoxe qu'ils ont découvert l'ATP-dependent proteolysis factor 1 (APF-1), qui de par son caractère « ubiquitaire » (c'est-à-dire hautement conservé chez tous les eucaryotes) sera rapidement renommé ubiquitine. Ces découvreurs ont également mis en évidence une corrélation entre la polyubiquitination d'une protéine et sa dégradation protéasomale. L'ensemble de ces découvertes a été récompensé par un prix Nobel de Chimie en 2004. Toutefois, en plus de son rôle avéré dans le processus de dégradation des protéines, l'ubiquitine est également associée à d'autres mécanismes cellulaires tels que la réparation de l'ADN, la régulation des histones, l'activation

de protéines kinases etc. Cela fait maintenant plus de 30 ans que l'ubiquitine fût découverte et demeure depuis une cible thérapeutique de première importance.

L'ubiquitine (Ub) est une petite protéine de 76 acides aminés avec un poids moléculaire de 8,5 kDa. Elle est formée d'une hélice α et de quatre feuillets β (Figure 2). C'est une protéine globulaire très stable thermiquement.

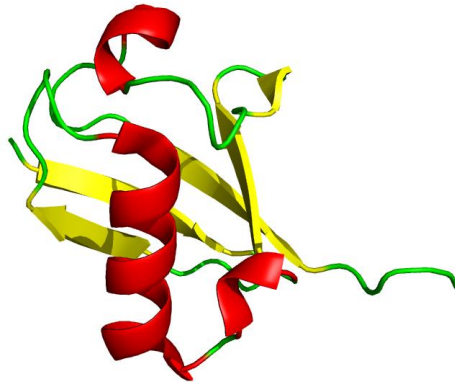


Figure 2. Structure tridimensionnelle de l'ubiquitine (code pdb: 2MWS). Les hélices α en rouge, les feuillets β en jaune, les boucles en vert.

L'ubiquitine est conservée et exprimée par toutes les cellules eucaryotes. Elle est codée par 4 gènes différents (*Uba52*, *Rps27a*, *Ubb*, *Ubc*), ce qui souligne son rôle central pour la cellule. Chez la souris et l'Homme, les gènes *Uba52* et *Rps27a* codent des protéines de fusions entre l'Ub et des protéines ribosomales, et sont exprimés de façon constitutive au sein de la cellule. Les gènes *Ubb* et *Ubc* quant à eux codent la polyubiquitination souvent induite lors d'un stress cellulaire. Des recherches de biologie moléculaire ont récemment démontré l'importance de ces gènes puisqu'une délétion du gène *Ubc* est létale au niveau embryonnaire et celle du gène *Ubb* est responsable de l'infertilité pour le mâle et la femelle ainsi qu'une dégénérescence des cellules du système nerveux^{9,10}.

Au sein de la cellule, l'ubiquitine est présente sous plusieurs formes : libre ou conjuguée à des substrats nucléaires comme cytosoliques voire même membranaires. L'ubiquitination comprend plusieurs étapes enzymatiques successives permettant *in fine* le couplage de l'Ub sur la lysine cible (K) du substrat (Figure 3). Tout d'abord, l'ubiquitine est produite sous la forme d'un précurseur soit polyubiquitine, soit en fusion avec une protéine ribosomale, ainsi elle doit passer par une étape de maturation avant d'être prise en charge par la machinerie d'ubiquitination. Ensuite une enzyme d'activation E1 assure l'adénylation du résidu glycine en position 76 de l'Ub en présence d'adénosine triphosphate (ATP) et d'un ion magnésium (Mg^{2+}) pour former une liaison thioester. Puis l'Ub est transférée à une enzyme E2 en charge de sa

conjugaison sur le groupe ϵ -NH₂ d'une lysine au sein de la protéine. L'enzyme E2 est fréquemment retrouvée en étroite association avec la protéine E3, qui contrôle la spécificité de ce mécanisme en reconnaissant le substrat cible pour conduire à la formation d'un lien isopeptidique. Dans la cellule eucaryote, il existe plusieurs enzymes E2 et la famille des ligases E3 ne cesse de s'agrandir au fur et à mesure des recherches réalisées sur l'ubiquitination⁸.

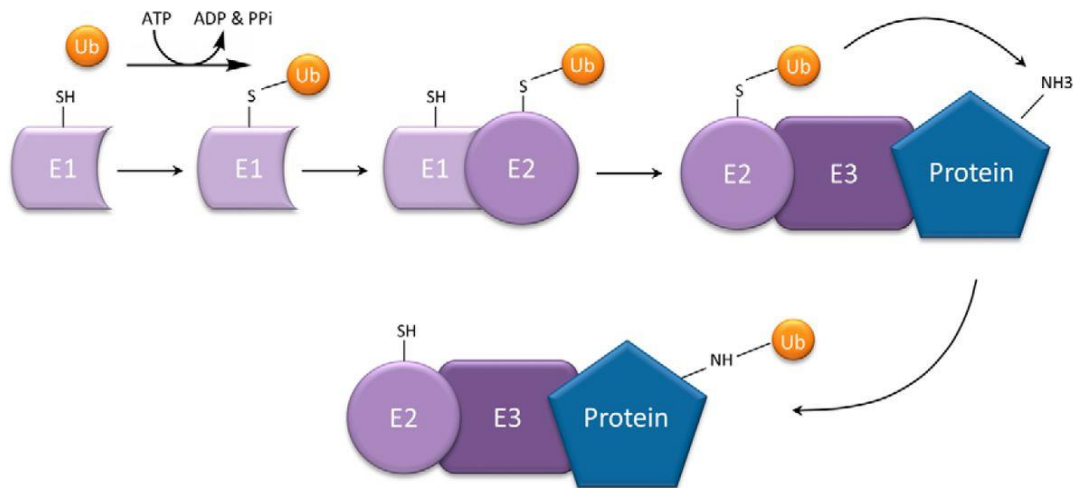


Figure 3. Le processus enzymatique de l'Ubiquitination (D'après Dwayne et al. 2017)¹¹.

Au cours des années 1990, l'étude du protéome a connu un essor spectaculaire avec l'avènement de la spectrométrie de masse. Cette technique a permis l'identification au sein de l'ubiquitine de plusieurs résidus lysines impliqués dans la formation de liens isopeptidiques^{12, 13}. En effet, l'ubiquitine possède sept lysines internes (K6, K11, K27, K29, K33, K48, K63) pouvant être modifiées, et conduire à l'élaboration de chaînes polyUb d'architecture variées. Cette diversité dans l'architecture des chaînes polyUb permet d'orienter les substrats vers des voies de signalisation cellulaire différentes (Figure 4). En effet, des travaux montrent que seuls les substrats modifiés par une chaîne de polyUb K48 d'au moins 4 molécules d'ubiquitine sont dirigés vers le protéasome en vue d'une dégradation. La diversité des signaux cellulaire induits par les chaînes polyUb est due à leur architecture qui diffère de manière significative d'un type à l'autre, permettant ainsi leur reconnaissance spécifique par des effecteurs dédiés^{14, 15}.

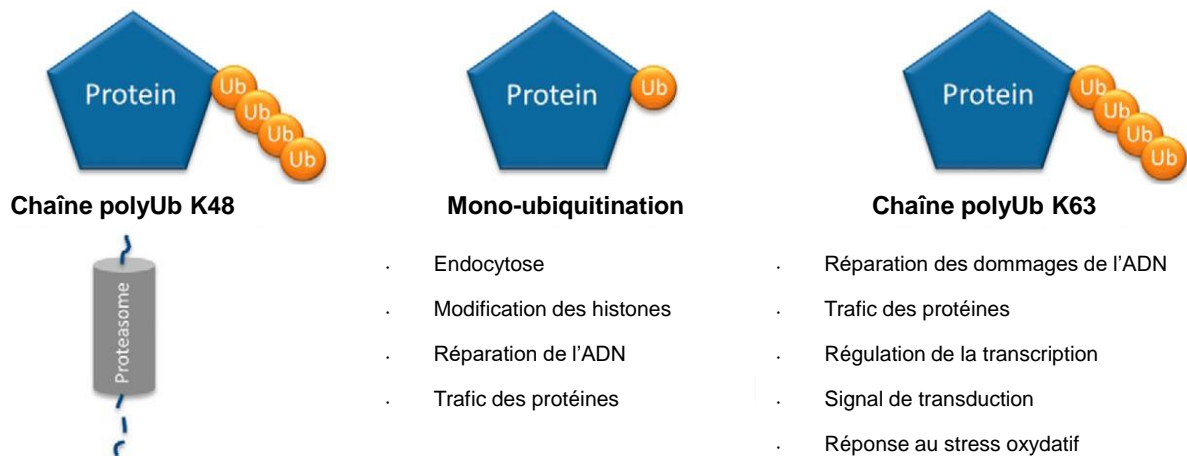


Figure 4. Conséquences fonctionnelles des différentes topologies d'ubiquitination (adapté de Dwayne et al. 2017)¹¹.

Pour finir, l'ubiquitination est une réaction rendue réversible par l'action d'enzymes appelées déubiquitinases (DUBs), capables d'hydrolyser les liaisons isopeptidiques des différents conjugués ubiquitines rencontrés (chaînes polyUb ou mono-ubiquitination ou multi-ubiquitination).

C. Les modifications post-traductionnelles apparentées à l'ubiquitine

Au cours des années 2000, la modification réversible des protéines par l'ajout de polypeptide s'est diversifiée avec la mise en évidence de plusieurs protéines apparentées à l'ubiquitine, appelées Ubiquitin-like, UBL. Bien que ces UBL présentent une faible homologie de séquence avec l'ubiquitine, elles partagent malgré tout une importante similarité dans leur structure tertiaire (structure globulaire à feuillets β). De même, leur mécanisme de conjugaison s'effectue suivant une cascade enzymatique à la fois spécifique mais similaire à celle de l'Ub. Cependant, la présence d'une enzyme E3 n'a pas toujours été démontrée pour chacune de ces voies enzymatiques. Actuellement, il existe une vingtaine d'UBL, répertoriées dans le tableau ci-dessous (Tableau 2).

Protéines apparentées à l'Ub ^a	% d'identité avec l'Ub	E1 ^a	E2 ^a	Caractéristiques
Ubiquitine	100	Uba1& 6	Plusieurs	Codé par plusieurs gènes
ISG15	32/37 ^b	Ube1L	UbcH8	Induit par l'interféron I
Rub1/NEDD8	55	Uba3-Ula1	Ubc12	Substrats : cullins, p53
Smt3/SUMO1-5	18	Uba2-Aos1	Ubc9	Vertébrés : 5 gènes <i>sumo</i>
FAT10	32/40 ^b	Uba6	/	Substrats inconnus, repliement de type β -grasp
FUBI/MNSF β	38	/	/	Dérivé à partir d'un précurseur ribosomal
Ubl-1	40	/	/	Précurseur protéique ribosomal chez le Nématode
SF3a120	30	/	/	Ubl à l'ext. C-terminale, pas de données trouvées sur la conjugaison
Oligo(A) synthetase	30/42 ^b	/	/	Ubl à l'ext. C-terminale, pas de données trouvées sur la conjugaison
UFM1	ND	Uba5	Ufc1	Repliement de type β -grasp

Tableau 2. Tableau résumant quelques protéines apparentées à l'Ub ou susceptibles de l'être¹⁶. ^a Les noms des protéines de levures sont listés pour E1 et E2 à l'exception de ISG15, FAT10 et UFM1, qui ne sont pas trouvés dans *S. cerevisiae* (Uba6 a une distribution phylogénétique beaucoup plus limitée); pour les Ubl, les noms de *S. cerevisiae* sont donnés s'ils sont présents dans cet organisme et sont listés en premier si un orthologue de vertébré est connu et porte un nom différent, ^b Deux domaines liés à l'Ub, ^c Non détectable par les recherches BLAST standard.

Historiquement, ISG15 (Interferon-stimulated gene 15) est la première protéine UBL à avoir été découverte. L'expression du gène *Isg15* est régulée positivement après un stress cellulaire, en particulier ceux induits par des infections bactériennes et virales. Le stress subit par la cellule induit l'expression des facteurs pro-inflammatoires tels que les interférons de type I (INF de type I) qui déclenchent l'expression d'ISG15 et permettent d'assurer une réponse antivirale complète qui cible toutes les étapes de réplication des différents virus. A titre d'exemple, les souris knowout (KO) pour le gène *Isg15*, c'est-à-dire dont le gène *Isg15* a été éteint, présentent une sensibilité exacerbée aux virus Influenza A, Influenza B, HSV-1 et Sindbis¹⁷. Ainsi, l'ISGylation joue un rôle central dans les mécanismes de défense de la cellule

(réponse immunitaire innée) face à une infection virale. Enfin, le mécanisme permettant cette modification suit le même schéma enzymatique que l'Ub (E1, E2, et E3).

Parmi les UBL, on retrouve également la protéine NEDD8 qui possède la plus forte homologie de séquence avec l'Ub (55%). La protéine NEDD8 est essentielle pour les eucaryotes car une délétion de son gène *Nedd8* est létale chez tous les organismes étudiés excepté chez *Saccharomyces cerevisiae*. Les substrats de la neddylation (modification par NEDD8) les mieux caractérisés sont les membres de la famille des Cullines, protéines adaptatrices charnières à la formation de complexes ubiquitine ligases E3 de la famille RING (CRL, Cullin-RING-ligase). De ce fait, la modification des CRLs par NEDD8 stimule l'ubiquitination de son substrat par le recrutement d'ubiquitine ligases E3^{18, 19}. C'est le cas de la protéine Parkine, une ubiquitine ligase E3 dont le rôle est de favoriser la dégradation des mitochondries endommagées par un processus connu sous le nom de mitophagie. Lorsque la protéine parkine est endommagée elle est aussitôt NEDDylée ce qui augmente son activité d'ubiquitination *in vivo*. En outre, l'interférence entre la NEDDylation et l'ubiquitination lors de différents stress cellulaires et/ou de conditions pathologiques devrait être prise en compte dans l'évaluation de la NEDDylation en tant que cible thérapeutique prometteuse.

Aujourd'hui, la modification des substrats par l'Ub et les UBL apparaît comme un système de régulation essentiel pour l'homéostasie cellulaire, et dont l'importance est capitale dans le fonctionnement normal ou pathologique de la cellule.

Cette thèse a pour objectif de participer à la compréhension de l'une de ces UBL, la protéine SUMO.

II. La SUMOylation

A. La découverte des protéines SUMO

La protéine SUMO (Small Ubiquitin-Like Modifier) a été découverte et caractérisée dans la deuxième partie des années 90, soit 20 ans après la découverte de l'ubiquitine. L'étude de cette MPT est un challenge en partie dû à son caractère hautement dynamique, ou seule une petite partie du protéome (estimé à moins de 1%) est SUMOylé *in vivo*²⁰. On retrouve la SUMOylation (modification par SUMO) chez tous les eucaryotes étudiés, de *Saccharomyces cerevisiae* à *Homo sapiens*. En premier lieu, elle a été découverte chez la levure après un criblage génomique pour des répresseurs du phénotype de la mutation *Mif2* (Mitotic fidelity of chromosome transmission protein 2). C'est pourquoi, la protéine SUMO a été nommée initialement Smt3p (Suppressor of mif two 3 protein). L'année suivante, le gène humain *Sumo1* a été cloné, et nommé à cette occasion *hSmt3*^{21, 22}. Ainsi, son clonage a permis de réaliser des expériences de criblages en double hybride mettant en avant l'implication de la

SUMOylation dans plusieurs processus cellulaires essentiels tels que la réparation de l'ADN via RAD51 et RAD52, l'apoptose induite par Fas et TNF. En 1997, l'acronyme SUMO-1 est introduit pour la première fois lors de l'étude de la SUMOylation ATP-dépendante de RanGAP1 (Ran GTPase Activating Protein 1) par Stm3p^{23, 24}.

Les eucaryotes unicellulaires et les invertébrés ne possèdent qu'un gène codant la protéine SUMO, et il s'avère être essentiel à la viabilité cellulaire. Chez les eucaryotes supérieurs tels que les plantes et les vertébrés, ils expriment plusieurs variants du gène. En effet, le génome d'*Arabidopsis thaliana* contient huit gènes codant des paralogues SUMO²⁵, le génome humain cinq dont un pseudo-gène pour la protéine SUMO-5 et dans le génome de la souris on en dénombre trois. Dans ce manuscrit seules les protéines SUMO humaines seront décrites. D'un point de vue de l'homologie de séquence, la protéine SUMO-1 partage environ 20% d'identité de séquence avec l'ubiquitine, les protéines SUMO-2 et -3 sont identiques à 97% entre elles mais elles ne partagent que 50% d'identité avec SUMO-1²⁶ (Figure 5). Leur forte homologie de séquence fait qu'elles sont couramment appelées par la contraction SUMO-2/3.

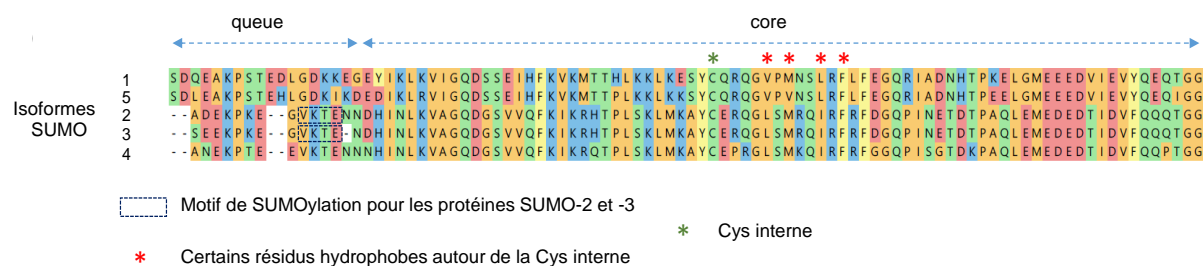


Figure 5. Les séquences protéiques des isoformes SUMO chez l'Homme (*Homo sapiens*), hSUMO-1 (P63165), hSUMO-2 (P61956), hSUMO-3 (P55854), hSUMO-4 (Q6EEV6) et hSUMO-5 (G2XKK0) ont été alignées en utilisant MegAlign Pro™ (Lasergen © DNASTar®) avec l'algorithme MAFFT (matrice de score BLOSUM30). Les séquences ont été associées en fonction de leurs proximités phylogéniques. La couleur de fond des acides aminés correspond à la nature chimique de la chaîne latérale (aromatique - jaune, acide - rouge, basique - bleu, non polaire - orange, polaire - verte).

B. Les différents homologues SUMO

Bien que les protéines SUMO ne partagent qu'une faible identité de séquence avec l'ubiquitine, leurs structures tridimensionnelles en sont très proches. En effet, les protéines SUMO adoptent une forme globulaire compacte de type $\beta\beta\alpha\beta\beta\alpha\beta$, avec un motif di-glycine en C-terminal permettant l'entrée dans le cycle de la SUMOylation.

Les protéines SUMO se distinguent de l'ubiquitine par la présence d'une extension au niveau de leur extrémité N-terminale dans une direction opposée à celle du côté C-terminal. Par ailleurs, la fonction exacte de cette extension N-terminale reste à l'heure d'aujourd'hui inconnue, mais ses caractéristiques riches en acides aminés chargés, en glycine et en proline, en font une excellente région pour des interactions protéine-protéine. En ce qui concerne la distribution des charges à la surface de la protéine, les protéines SUMO sont bien différentes de l'Ub bien qu'elles partagent un domaine structuré identique. A la différence de l'ubiquitination, la SUMOylation des protéines s'effectue à un endroit bien particulier de la séquence qui correspond à un motif consensus de type $\Psi KxE/D$, où Ψ est un acide aminé hydrophobe, K la lysine acceptrice, x n'importe quel acide aminé et E/D un acide aminé glutamate ou aspartate.

Au delà des différences au sein de leur séquence, les isoformes SUMO se différencient également par une certaine spécificité de substrat. En effet, une analyse protéomique réalisée par Vertegaal et ses collaborateurs a permis d'identifier 53 substrats de la SUMOylation²⁷. Parmi ces substrats, 25 présentaient une préférence pour SUMO-1, 19 pour SUMO-2, et 9 étaient modifiés par SUMO-1 ou -2. Cette analyse est intéressante car elle révèle que tous les substrats de la SUMOylation ne sont pas exclusivement nucléaires. Une autre étude effectuée sur les cellules HeLa a mis en évidence cette fois-ci 232 substrats SUMOylables, sans surexpression des paralogues SUMO. De la même manière, on observe une spécificité envers l'un ou l'autre paralogue²⁸ (Figure 6).

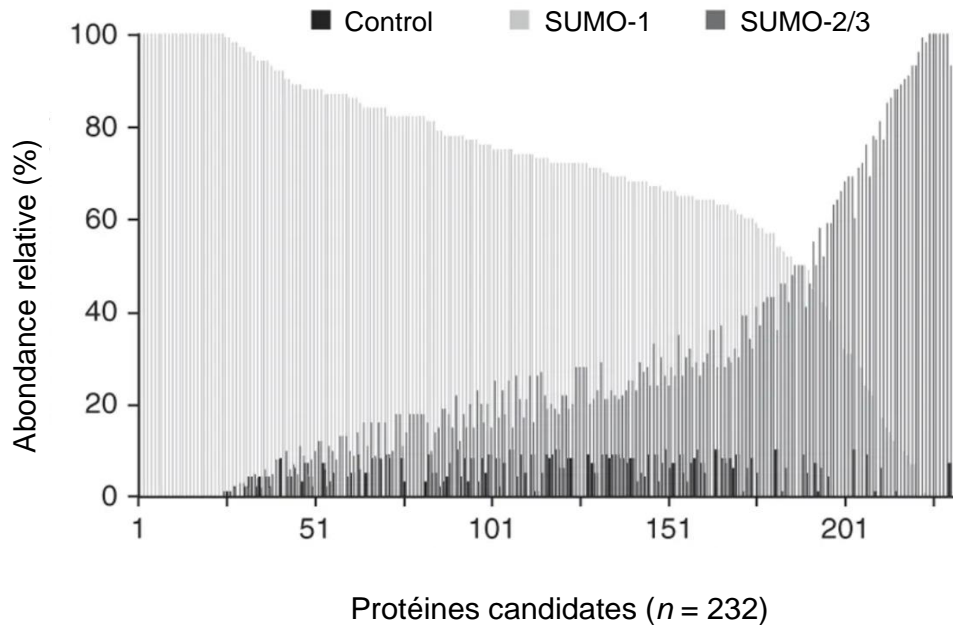


Figure 6. Analyse protéomique des substrats de SUMO démontrant une spécificité relative de substrats selon le paralogue SUMO. Les barres gris clair représentent l'abondance relative de peptides modifiés par SUMO-1. Les barres gris foncé représentent l'abondance relative des peptides modifiés par SUMO-2/3, et les barres noires correspondent au bruit de fond.

Malgré cette spécificité de substrat, il est intéressant de noter que deux équipes indépendantes ont montré que des souris KO pour le gène *Sumo1* étaient parfaitement viables et présentaient un développement embryonnaire normal^{29, 30}. Ces résultats indiquent que les protéines SUMO-2 et -3 sont capables de compenser la perte de la protéine SUMO-1. Cependant, la réciproque n'a pas été testée mais il est probable que l'invalidation des gènes *Sumo2* et *Sumo3* ait un impact important sur le fonctionnement de la cellule. En effet, la protéine SUMO-1 est incapable de former des chaînes polySUMO. S'ajoute à cela, la dynamique de conjugaison qui est différente entre les protéines SUMO-1, SUMO-2 et -3. En effet, la protéine SUMO-1 est souvent liée de manière stable à ses substrats, avec un faible pool de protéine SUMO-1 libre. A contrario, les formes libres des protéines SUMO-2 et -3 sont beaucoup plus abondantes et se conjuguent rapidement aux substrats suite à un stress cellulaire^{31, 32}. Ainsi, l'une des fonctions des protéines SUMO-2 et -3 serait de fournir à la cellule un stock de protéines SUMO prêtes à répondre à un stress cellulaire important (choc thermique, osmotique, oxydatif etc.). Chez l'Homme, le rôle de la protéine SUMO-4 semble être assez controversé. En effet, son profil d'expression au sein de certains organes tout comme son implication dans le diabète de type I sont des sujets de discordes entre différents groupes de recherche^{33, 34}.

Enfin, très récemment une protéine SUMO-5 a été identifiée dans certains tissus. La protéine SUMO-5 est hautement conservée parmi les espèces primates³⁵. Comme pour les

protéines SUMO-2 et -3, SUMO-5 est capable de former des chaînes polySUMO. Cependant, ses rôles biologiques sont encore inexplorés³⁶.

1. La protéine SUMO-1

La protéine SUMO-1 est identique à 46% à SUMO-2 et à 47% à SUMO-3³⁷. Comme mentionné précédemment, la protéine SUMO se différencie de l'ubiquitine par la présence d'une région flexible d'une quinzaine d'acides aminés située du côté N-terminal (Figure 7).

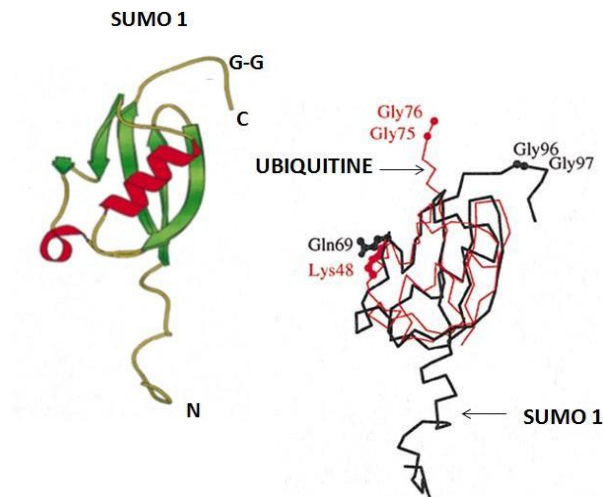


Figure 7. La structure tridimensionnelle de SUMO-1 est similaire à celle de l'ubiquitine. A) Représentation en rubans de la structure 3D de SUMO-1 en utilisant le logiciel MOLSCRIPT. Les protéines SUMO appartenant aux membres de la famille des Ubiquitine-like (UBL) se caractérisent par un repliement dit en « en β -grasp » qui se définit par une structure secondaire organisée selon une succession de feuillets beta-beta-alpha-beta-beta. B) Superposition des structures de SUMO-1 (noir) et de l'ubiquitine (rouge). La superposition des carbones Ca met en évidence la similitude de structure des deux protéines³⁷.

Du point de vue de la répartition des charges, le domaine SUMO-1 expose une surface étendue concentrant des charges négatives y compris au niveau de la région N-terminale, non présente chez l'ubiquitine. On trouve des charges positives sur la face concave de la molécule (Figure 8). Cette répartition particulière des charges peut expliquer les interactions que SUMO-1 réalise avec ses différents partenaires comme l'enzyme Ubc9 et la SUMO protéase Ulp1^{37, 38}.

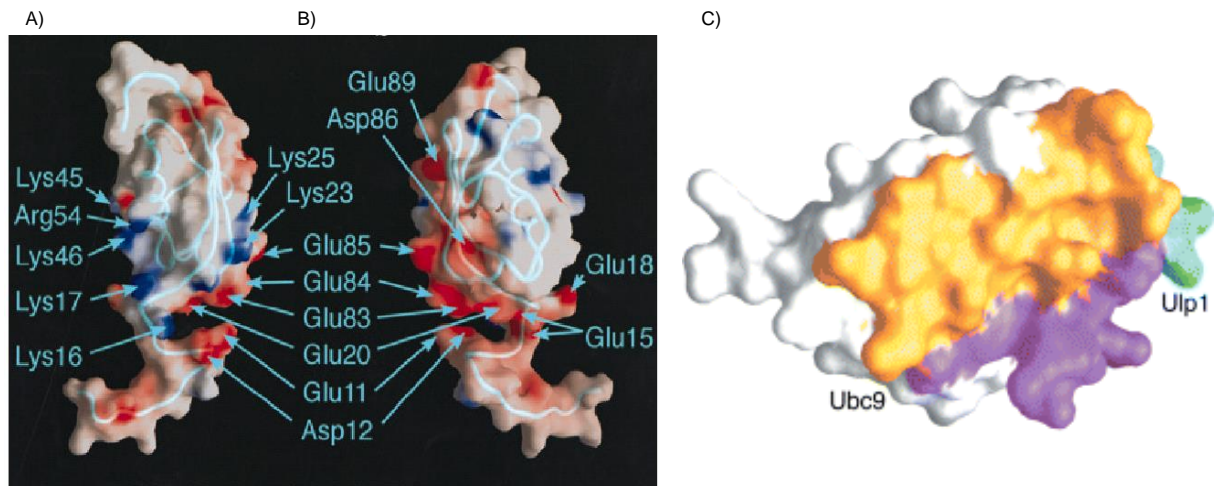


Figure 8. Répartition des charges à la surface de SUMO-1 et modélisation des régions d'interaction. A-B) Répartition des charges électrostatiques sur les deux faces opposées de la molécule. Les orientations sont indiquées en filigrane et correspondent au squelette. Les charges positives sont représentées en bleu et les charges négatives en rouge en utilisant le programme GRASP³⁹. La première vue de SUMO-1 (A) montre les épitopes chargés positivement (Arg54, Lys46, Lys25, Lys23, Lys17 and Lys16). La seconde vue (B) révèle une large zone composée d'acides aminés chargés négativement (Glu89, Asp86, Glu85 et Glu84) ainsi que la poche d'acides aminés acides³⁷ (Glu83, Glu18, Glu15, Asp12, Glu11 et Glu20). C) Modélisation des zones d'interaction de SUMO-1 avec l'enzyme de conjugaison Ubc9 et la SUMO protéase Ulp1 colorées en orange, magenta et bleu cyan³⁸.

SUMO-1 a une localisation principalement nucléaire⁴⁰. Elle se retrouve essentiellement sous forme conjuguée et son ajout covalent sur une protéine ne semble pas être lié à des conditions physiologiques spécifiques³². SUMO-1 régule de nombreux acteurs du transport noyau-cytoplasme, de la transcription, de la réparation de l'ADN, de la régulation du cycle cellulaire et de l'apoptose^{41, 42}.

2. Les protéines SUMO-2 et SUMO-3

Les protéines SUMO-2 et -3 sont homologues à 97% et ne partagent que 50% d'identité avec SUMO-1²⁶ (Figure 9). Dans la littérature elles sont souvent désignées par la contraction SUMO-2/3. Actuellement, aucun anticorps permettant de discriminer SUMO-2 de SUMO-3 n'a pu être mis au point, pourtant cela serait un outil indispensable pour les différentes recherches menées sur la compréhension du rôle de ces MPT au sein d'une cellule.

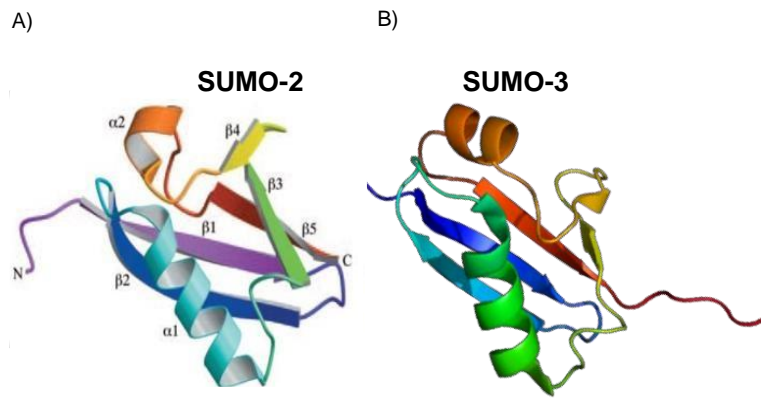


Figure 9. Structures tridimensionnelles de SUMO-2 (code pdb 1WM2) et SUMO-3 (code pdb 2MP2). A) Représentation en ruban de la structure 3D de SUMO-2 tronquée en N-terminal (structure correspondant aux aa 12-89) obtenue par diffraction aux rayons X. B) Représentation en ruban de la structure 3D de SUMO-3 (structure correspondant aux aa 12-92) obtenue par résonance magnétique nucléaire. Le repliement dit en « en β -grasp » est retrouvé comme dans l'ubiquitine et SUMO-1^{38, 43}.

Le site de SUMOylation (Ψ KxE/D) est retrouvé au sein de la queue N-terminale des protéines SUMO-2 et SUMO-3, et permet ainsi la formation de chaînes polySUMO⁴⁴. Ce site de SUMOylation est fortement conservé au sein des eucaryotes supérieurs (levures, souris et Homme). Récemment, une étude de protéomique a mis en évidence la présence d'autres sites de branchements au sein des protéines SUMO-2 et -3 tels que les lysines 5 et 7 (K5, K7), en plus du site de branchement majeur qui est la lysine 11 (K11). Ainsi, ces données suggèrent la possibilité pour les protéines SUMO-2 et -3 de former des chaînes d'architecture différente, chose que l'on retrouve dans la polyubiquitination. Il serait donc intéressant de pouvoir accéder à ces différentes chaînes polySUMO afin de mieux comprendre les conséquences biologiques de ces modifications, à l'instar de ce qui est fait pour les chaînes d'ubiquitine.

Pour les protéines SUMO-2 et SUMO-3, la répartition des charges de surface est semblable. La région la plus caractéristique des protéines SUMO-2 et -3 est la zone qui correspond à la partie concave de la molécule. Dans le domaine SUMO-1 cette partie est chargée positivement alors que dans les domaines SUMO-2 et -3 elle est neutre. Ceci est dû au remplacement de la Lys en position 48 de SUMO-1 par une Met en position 44 dans les domaines SUMO-2 et -3 (Figure 10, flèches rouge). La proximité de cette surface concave avec l'extrémité C-terminale pourrait servir de site potentiel pour discriminer les isoformes (SUMO-1 vs SUMO-2/3) dans les cellules humaines.

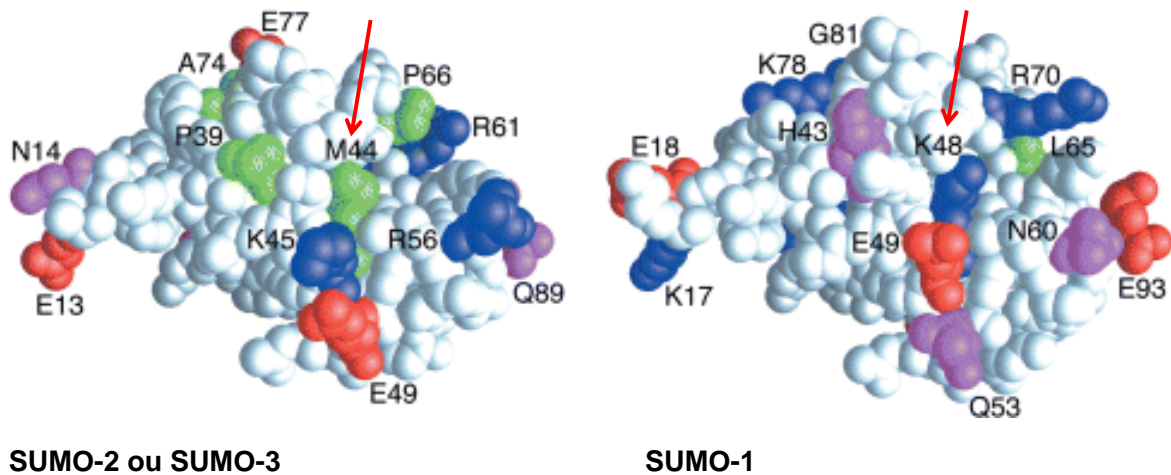


Figure 10. Identification des acides aminés qui diffèrent par leur charge entre la protéine SUMO-2 ou -3 et la protéine SUMO-1. Les acides aminés chargés positivement, négativement et les plaies neutres sont colorés en bleu, rouge et violet respectivement. Les résidus non polaires sont en vert. Les flèches rouges pointent vers le résidu Met44 pour SUMO-2, et Lys48 pour SUMO-1³⁸.

Pour finir sur la localisation subcellulaire, la protéine SUMO-2 est retrouvée associée aux corps nucléaires alors que la protéine SUMO-3 est cytosolique⁴⁰. Comme précédemment mentionné, les protéines SUMO-2 et -3 sont majoritairement libres et sous cette forme constituent un réservoir de protéines SUMO disponible à tout moment pour la cellule.

3. La protéine SUMO-4

La protéine SUMO-4 a été découverte plus récemment, elle partage 86% d'identité avec la protéine SUMO-2, et 76% avec la protéine SUMO-3 (Figure 5). *In vivo*, cette isoforme ne participe pas à la formation de conjugués dû à son incapacité à être maturée et donc à dévoiler ses deux résidus glycine C-terminales³⁴. En effet, le résidu de proline en position 90 dans la séquence de la protéine SUMO-4 provoque une contrainte conformationnelle ne permettant pas l'hydrolyse du précurseur sous sa forme native. A la différence des autres isoformes, son expression se retrouve limitée à certains tissus^{33, 45}. La protéine SUMO-4 est principalement exprimée dans le rein, les ganglions lymphatiques, la rate et les îlots pancréatiques. Elle modifie essentiellement des facteurs de transcription ayant un rôle dans la réponse immunitaire par des interactions non covalentes⁴⁶. Récemment, des recherches ont montré que la protéine SUMO-4 inhibe indirectement l'activité transcriptionnelle du facteur NFκB (Nuclear factor-kappa B) en SUMOylant IκBα (Inhibitor of kappa B α). Cette modification IκBα par SUMO-4 favorise ainsi la séquestration de NFκB dans le cytoplasme et inhibe la transcription d'une série de gènes anti-apoptotiques. En 2008, Wei et ses collaborateurs ont montré que la protéine SUMO-4 joue un rôle clé lors d'un stress cellulaire⁴⁷. En effet, la

protéine SUMO-4 modifie plusieurs facteurs de transcriptions importants pour la régulation du stress intracellulaire tels que AP-1, AP-2, STAT-1, STAT-3 et le récepteur des glucocorticoïdes⁴⁸. Par conséquent, les substrats de SUMO-4 appartiennent notamment à un groupe de protéines anti-stress incluant des enzymes antioxydantes comme par exemple Cu/ZnSOD, la catalase, les peroxyredoxines, des protéines chaperonnes comme HSP70 et Grp58 (protéine disulfure isomérase retrouvée dans le lumen du Réticulum Endoplasmique).

C. La polySUMOylation

Bien que SUMO-2 et SUMO-3 partagent 97% d'identité, SUMO-2 semble être un meilleur substrat pour la polymérisation⁴⁴. *In vivo*, les chaînes de polySUMO-2/3 sont souvent terminées par un domaine SUMO-1 puisqu'elle contient un site interne de SUMOylation, suggérant un rôle de « terminateur » d'élongation^{36, 44}. Des études récentes réalisées dans des conditions favorisant la SUMOylation montrent que les résidus de lysine K5, K7 et K42 des protéines SUMO-2 et -3 et les résidus de lysine K7, K16, K17, K25, K37, K39 et K46 de SUMO-1 seraient SUMOylables^{49, 50}. Cependant, la polySUMOylation demeure encore largement inexplorée.

Par ailleurs, des chaînes mixtes de polyubiquitine/SUMO ont été décrites^{51, 52} (Figure 11) ce qui soulève la question du rôle de ces chaînes dans le processus de dégradation des protéines via le protéasome. C'est en 2012, qu'une équipe de recherche démontre qu'I κ B α , la protéine inhibitrice de NF κ B, est modifiée par des chaînes mixtes de SUMO-2/3-ubiquitine et que la modification d'I κ B α par ces chaînes mixtes après une stimulation de la cellule au TNF α engendre une dégradation protéasomale plus efficace que lorsqu'elle est modifiée seulement soit par SUMO-2 et -3 soit par l'ubiquitine seule⁵².

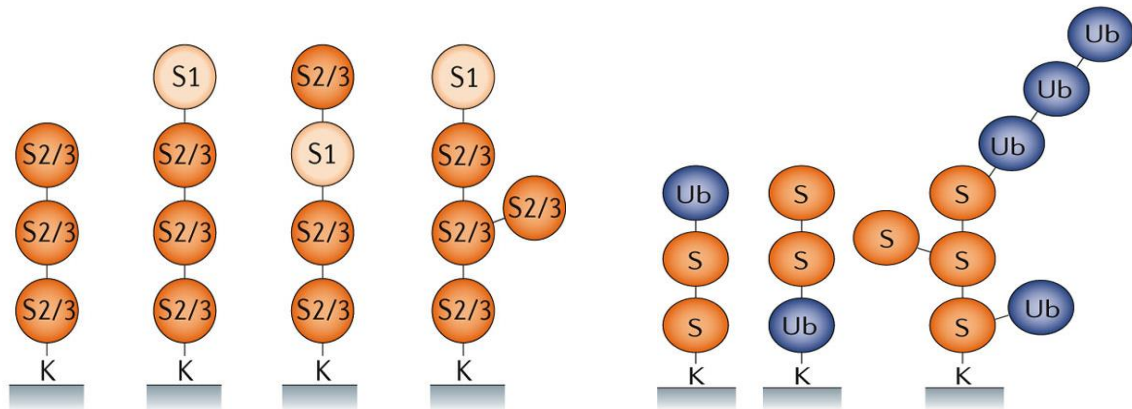


Figure 11. Les protéines SUMO-2 et SUMO-3 (S2/3) forment des chaînes via le site de SUMOylation internes pouvant inclure SUMO-1 (S1), ce dernier étant situé généralement à l'extrémité de la chaîne polySUMO. La protéine SUMO-1 contient un site de SUMOylation inversé, qui peut être utilisé pour la formation de chaînes. De plus, les chaînes SUMO peuvent être ramifiées. Les protéines SUMO forment également des chaînes mixtes avec l'ubiquitine.

D. Le mécanisme enzymatique de la SUMOylation

La SUMOylation est un mécanisme dynamique et réversible semblable à celui de l'ubiquitination⁵³. Ainsi, on retrouve dans ce processus des enzymes de maturation, d'activation, de conjugaison, de ligation et de déconjugaison (Figure 12). Cependant, les enzymes impliquées sont spécifiques de ces réactions. En effet, l'absence de redondance entre les enzymes de la SUMOylation et celles des autres UBL ont permis de les étudier spécifiquement.

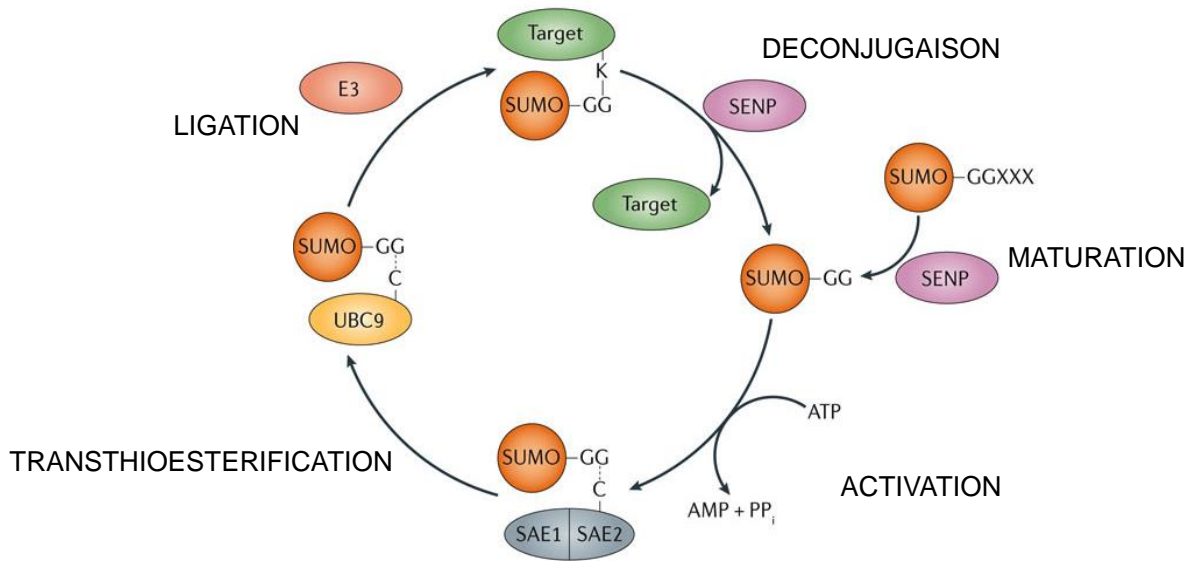


Figure 12. Le cycle de conjugaison des protéines SUMO. Dans un premier temps, le précurseur de SUMO est clivé en C-terminal pour exposer un motif di-glycine. Ensuite, la protéine SUMO mature est activée par l'enzyme E1 composée de deux sous unités SAE1 et SAE2, puis transférée à l'enzyme E2 de conjugaison Ubc9. Enfin, l'enzyme E3 permet le rapprochement d'Ubc9 avec la protéine cible facilitant ainsi sa modification covalente par SUMO. La réaction est dite réversible par l'intervention de SUMO protéases appelées aussi deSUMOylases²⁶.

1. La maturation des protéines SUMO

La traduction des ARNm des protéines SUMO aboutit à la production de pro-protéines ou pro-peptides appelés précurseurs. Ces précurseurs possèdent une séquence de 2 à 11 acides aminés permettant de masquer le motif di-glycine en C-terminal, signal d'entrée dans le cycle de la SUMOylation. Cette maturation est réalisée par des enzymes présentant une activité endopeptidase, et qui sont spécifiques des protéines SUMO. Par ailleurs, une autre source de protéines SUMO peut provenir du recyclage des substrats SUMOylés après action des enzymes à activité isopeptidase.

Chez la levure, deux SUMO protéases ont été découvertes, Ulp1 et Ulp2 (Ubiquitin-like protein 1 et 2). Elles sont capables de maturer la protéine SUMO, Smt3p et de cliver la liaison isopeptidique avec le substrat. Chez les mammifères, trois groupes de SUMO protéases (ou déSUMOylases) ont été décrites et classées en fonction de leur structure : les SENP (Sentrin protease), DESI1 et 2 (Desumoylating isopeptidase 1 et 2), et USPL1 (Ubiquitin-specific peptidase-like protein 1). A ce jour, les déSUMOylases les mieux caractérisées sont les SENP. Leur découverte au début des années 2000 a été réalisée d'après une homologie de séquence avec Ulp1⁵⁴. On dénombre six SENP : SENP1, -2, -3, -5, -6, et -7. Leur spécificité de substrat découle en grande partie de leur localisation subcellulaire. Elles sont essentiellement retrouvées dans des structures bien différenciées du

noyau⁵⁵ (Tableau 3). Par ailleurs, certaines SENPs possèdent une affinité plus importante pour l'un ou l'autre paralogue SUMO (SUMO-1, -2, -3). En particulier, SENP1 and SENP2 sont spécifiques pour SUMO-1, SUMO-2 et -3 alors que SENP3 et SENP5 sont plus spécifiques pour les domaines SUMO-2 et -3⁵⁵. De la même manière, seules SENP1, 2 et 5 sont capables de maturer les précurseurs SUMO alors que SENP6 et 7 sont uniquement aptes à dépolymériser les chaînes polySUMO. Ce qui signifie que SENP6 et 7 sont capables de rompre les liens isopeptiques entre deux monomères SUMO, sans ôter le dernier monomère lié à la protéine cible^{56, 57}.

DéSUMOylase	Localisation	Spécificité	Act. endopeptidase	Act. isopeptidase	Edition des chaînes
SENP1	Pore nucléaire, foci nucléaires	SUMO-1 et SUMO-2/3	Oui	Oui	Non
SENP2	Pore nucléaire, foci nucléaires, cytoplasme	SUMO-2/3 > SUMO-1	Oui	Oui	Non
SENP3	Nucléole	SUMO-2/3	Inconnu	Oui	Non
SENP5	Nucléole et mitochondrie	SUMO-2/3	Oui	Oui	Non
SENP6	Nucléoplasme	SUMO-2/3	Non	Oui	Oui
SENP7	Nucléoplasme	SUMO-2/3	Non	Oui	Oui

Tableau 3. Caractéristiques des déSUMOylases SENP ; (Act. pour activité).

2. Etape d'activation par l'enzyme E1

Contrairement à l'ubiquitine, il n'existe qu'une seule et unique enzyme E1 pour la SUMOylation. Cette étape d'activation de la protéine SUMO mature fait appel à l'enzyme E1, présentée sous la forme d'un hétérodimère SAE1:SAE2 (SUMO Activating Enzyme 1 et 2). L'unité SAE1 porte le domaine important pour l'adénylation ATP-dépendante de la glycine en position C-terminale de SUMO, alors que SAE2, lui porte la cystéine catalytique. Cette enzyme est capable d'activer les trois protéines SUMO (SUMO-1, -2 et -3). En effet, les résidus nécessaires à la reconnaissance par l'enzyme d'activation SAE1:SAE2 sont conservés dans ces trois isoformes^{44, 58}.

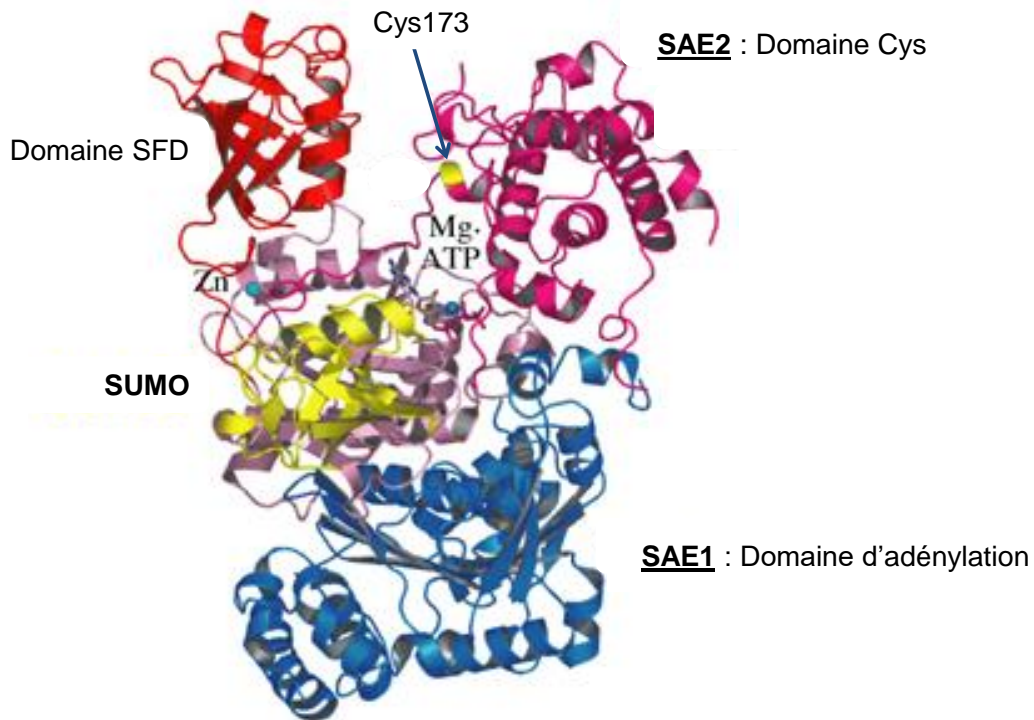


Figure 13. Structure aux rayons X du complexe entre SUMO-1 et l'enzyme d'activation E1 SAE1:SAE2. SUMO-1 (jaune) est complexé au domaine d'adénylation formé de SAE1 (bleu):SAE2 (rose). La cystéine catalytique (C173) du domaine SAE2 est colorée en jaune (flèche bleue). Le domaine SFD (rouge) de SAE2 sert au recrutement de l'enzyme de conjugaison E2 Ubc9. SFD : SUMO-Fold Domain (code pdb 3KYC).

La réaction chimique d'activation semble en tout point être identique à celle de l'ubiquitination. Après l'étape de maturation, la protéine SUMO est adénylée en présence d'ATP et d'ion magnésium (Mg^{2+}). Suite à cela, une liaison thioester de haute énergie est formée entre SAE2 et SUMO-AMP (SAE2-SUMO) grâce à l'attaque nucléophile de la cystéine catalytique (SAE2-C173) sur le groupe carbonyle C-terminal de la protéine SUMO adénylée. De cette manière, le processus d'activation s'accompagne de changements de conformation de la sous-unité SAE2 permettant le recrutement de l'enzyme de conjugaison Ubc9.

3. Etape de conjugaison par l'enzyme E2

La seconde étape de la SUMOylation implique le transfert de la protéine SUMO de la cystéine catalytique de SAE2 à celle de l'enzyme E2 (Ubc9) par une étape de transthiostérfication, sans consommation d'une molécule d'ATP. L'enzyme Ubc9 est conservée chez tous les organismes eucaryotes. Il n'existe qu'une seule enzyme E2 ce qui explique pourquoi elle est essentielle au sein d'un organisme. En effet, son invalidation

entraîne une létalité à un stade précoce du développement chez la levure *S. cerevisiae* et chez la souris^{59, 60}.

Ubc9 est recrutée par l'enzyme d'activation E1 via une interaction avec le domaine SFD (SUMO-Fold Domain) de SAE2^{61, 62} (Figure 13). Cette interaction permet un changement de conformation de SAE2 qui permet le rapprochement des sites actifs de SAE2 et Ubc9. Dans cette configuration, le transfert de SUMO de la cystéine catalytique de SAE2 (C173) à la cystéine catalytique d'Ubc9 (C93) est facilité⁶³ (Figure 13). Ubc9 semble être suffisante pour la reconnaissance du substrat ainsi que de la lysine du site consensus^{64, 65}. Or, dans certains cas Ubc9 interagit avec les différentes enzymes de ligation E3 situées en aval de la voie de SUMOylation. Cependant, la nature de cette interaction reste mal connue car peu de SUMO ligases E3 ont été caractérisées à ce jour.

Enfin, la régulation de cette cascade enzymatique passe par la SUMOylation de la sous-unité SAE2 mais aussi d'Ubc9. En effet, l'auto-SUMOylation de SAE2 n'altère pas le processus d'activation mais la formation du lien thioester est inhibée⁶⁶. L'auto-SUMOylation d'Ubc9 cible principalement le résidu lysine 14 (K14) qui se situe au sein de l'hélice α 1 correspondant à la région d'interaction avec l'enzyme E1, SAE1:SAE2. Cette SUMOylation n'a pas d'impact sur l'interaction E1:Ubc9 mais sur la capacité d'Ubc9 à modifier ses protéines cibles. Ainsi, la SUMOylation des enzymes conduit à une auto-inhibition de la SUMOylation. Par ailleurs, la phosphorylation d'Ubc9 par la kinase CDK1 régule de manière positive la formation du lien thioester, ainsi que l'activité de catalytique d'Ubc9^{67, 68}.

4. Etape de ligation par les enzymes E3 ligases

L'étape finale de la SUMOylation consiste à former une liaison isopeptidique entre le dernier résidu glycine en position C-terminale de SUMO et la lysine acceptrice de la protéine à modifier. Comme mentionné précédemment, les SUMO ligases E3 sont caractérisées par leurs capacités à lier l'enzyme Ubc9 et ainsi d'augmenter l'efficacité de la SUMOylation *in vitro* et *in vivo*. En effet, elles recrutent le complexe Ubc9-SUMO facilitant ainsi la modification^{69, 70} (Figure 14).

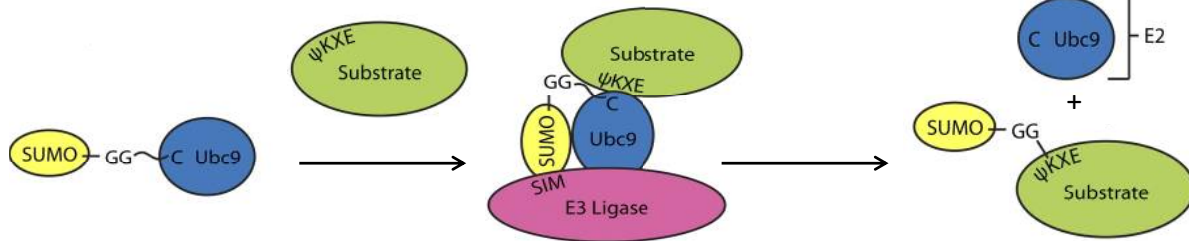


Figure 14. L'étape de ligation assistée par une SUMO ligase E3. Transfert de la protéine SUMO de la cystéine catalytique de l'enzyme de conjugaison Ubc9 à la protéine cible possédant le site consensus via l'intervention d'une SUMO ligase⁷¹.

Dans ce contexte, la SUMOylation se distingue de l'ubiquitination car les enzymes E3 ne forment pas d'intermédiaire covalent avec le peptide SUMO. De plus, les enzymes E3 sont capables de s'auto-SUMOyler. Si à ce jour, plusieurs centaines d'E3 sont caractérisées pour l'ubiquitine, on en dénombre très peu pour les protéines SUMO. La majorité des SUMO ligases présentent peu de spécificité au niveau de leurs substrats, mais elles possèdent des localisations subcellulaires différentes. Par exemple, la protéine RanBP2 associée au complexe du pore nucléaire est retrouvée du côté cytoplasmique alors que PIAS est localisée dans le nucléoplasme et les corps nucléaires. Topors se situe dans le nucléoplasme tandis que Pc2 est localisée dans des structures subnucléaires appelées complexes Polycomb. Ainsi, leurs différentes localisations contribuent à la régulation spatiale de la SUMOylation, mais également à la spécificité fonctionnelle de ces enzymes *in vivo*.

Malgré leur diversité structurale, les SUMO ligases E3 présentent plusieurs dénominateurs communs. En effet, elles recrutent le complexe Ubc9-SUMO, elles favorisent la SUMOylation du substrat, elles interagissent toutes avec les isoformes SUMO via le motif SIM, que nous discuterons un peu plus tard dans ce manuscrit (SUMO interacting motif, Figure 14). Ainsi, il est possible de les classer en deux groupes en fonction de la présence (PIAS et Topors) ou non (Pc2 et RanBP2) d'un domaine de type RING, similaire au domaine RING (Really interesting new gene) de certaines ubiquitines E3 ligases.

Les E3 ligases à domaine SP-RING (Siz/PIAS-RING) sont les plus représentées avec leur domaine RING essentiel à leur fonction ligase. Elles sont capables de lier la protéine, Ubc9 ainsi que la protéine SUMO via le motif SIM ce qui facilite le processus de SUMOylation. Au sein de cette famille, on retrouve notamment les protéines PIAS (Protein Inhibitor of Activated STAT) qui ont été initialement nommées comme cela en raison de leurs capacités à inhiber les protéines STAT (Signal transducer and activator of transcription), des protéines impliquées dans la régulation de la réponse immunitaire. Fait intéressant, ces protéines sont homologues aux E3 ligases chez *S. cerevisiae*^{72, 73} (Siz1 et Siz2). Chez l'Homme et la souris, quatre gènes codent pour les protéines PIAS (*Pias1*, -2, -3, -4 ou plus connue sous le nom *Piasy*) ce qui

donne lieu à différentes isoformes. D'un point de vue de leur structure, elles contiennent toutes quatre domaines conservés : un domaine avec le motif SAP (Scaffold attachment factor A/B/acinus/PIAS) du côté N-terminal qui permet de se lier à l'ADN, un avec le motif «PINIT » qui contribue à la sélectivité du substrat et un avec le motif SIM, permettant l'interaction avec la protéine SUMO⁷⁴. Toutefois, elles se différencient par leur domaine C-terminal qui varie beaucoup en fonction de la protéine PIAS étudiée (Figure 15).

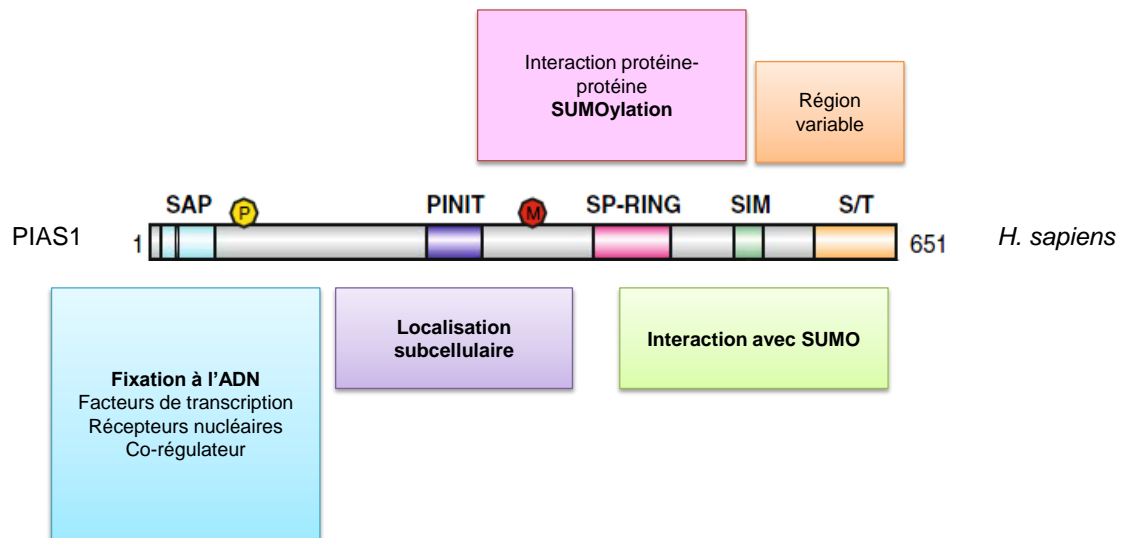


Figure 15. Représentation schématique de la SUMO ligase PIAS1. Le domaine SAP (cyan), PINIT (mauve), SP-RING (magenta) ainsi que le motif SIM (vert) et la région variable riche en résidus sérine et thréonine S/T (orange) sont positionnés. Les positions des modifications post-traductionnelles de PIAS1 sont représentées⁷⁴ (P, phosphorylation ; M, méthylation).

La seconde famille de SUMO ligases identifiées est illustrée par la protéine RanBP2 (Ran Binding Protein 2 ou Nup358 pour Nucleoporin 358) bien qu'elle ne possède pas de domaine RING. On la retrouve uniquement chez les vertébrés avec une localisation plutôt au niveau des pores nucléaires dans les cellules en interphase, plus précisément proche des kinétochores du fuseau mitotique⁷⁵. C'est une protéine essentielle au transport nucléocytoplasmique, et son activité est nécessaire pour la ségrégation chromosomique lors de la mitose. RanBP2 appartient au complexe du pore nucléaire composé notamment de RanGAP1. Ainsi, l'association RanGAP1-SUMO1-Ubc9 avec RanBP2 forme un complexe multi-protéique localisé sur la face cytoplasmique des pores nucléaires permettant la SUMOylation des protéines à destinée nucléaire. Ceci permet d'expliquer pourquoi la plupart des protéines connues pour être SUMOylées possèdent un NLS (Nuclear localization signal) et sont retrouvées dans le noyau des cellules^{76,77}. Aussi surprenant soit-il, RanBP2 n'est pas la SUMO ligase de RanGAP1, mais sa présence empêche l'action des désSUMOylases. Ainsi,

l'encombrement stérique imposé par RanBP2 réduit l'action endopeptidase des SENP pouvant expliquer la spécificité observée pour l'un ou l'autre des paralogues au niveau de RanGAP1. En effet, *in vitro* RanGAP1 peut-être SUMOylé par SUMO-1, SUMO-2 ou -3, or seule la forme RanGAP1-SUMO1 est retrouvée *in vivo*. Dans ce contexte, cela peut s'expliquer par le fait que les protéines SUMO-2 et -3 interagissent moins fortement avec RanBP2 ce qui laisse la possibilité aux déSUMOylases d'agir. Par conséquent, c'est le complexe tout entier RanBP2/RanGAP1-SUMO-1-Ubc9 qui servirait de SUMO ligase E3 pour les protéines dont le devenir est d'être transférées dans le noyau de la cellule.

Pc2 (Polycomb 2) ou CBX4 (Chromobox protein homolog 4) présentent aussi une activité SUMO ligase. C'est une protéine nucléaire associée au complexe PRC1 (Polycomb repressive complex 1), inhibant la transcription de nombreux gènes durant le développement des vertébrés au moyen de la méthylation des histones de la chromatine⁷⁸. Du point de vue de la structure, Pc2 est différente de PIAS et RanBP2, avec une localisation restreinte à la structure du noyau appelé corps PcG (Polycomb group body). D'une manière générale, elle est associée à l'hétérochromatine péricentromérique⁷⁹. Pc2 présente deux domaines d'interaction avec Ubc9 et SUMO qui sont importants pour sa fonction de ligase^{70, 79, 80}. Bien qu'*in vitro* cette enzyme ne semble pas être nécessaire pour la modification du co-répresseur CtBP1 (C-terminal-binding protein 1), *in vivo* son expression accentue fortement la SUMOylation de ce dernier⁸¹. Ainsi la SUMOylation de CtBP1 reste malgré tout dépendante de Pc2 et ces protéines sont localisées dans des structures subnucléaires appelées PcG avec la présence d'Ubc9.

Enfin, les HDAC (Histone Deacetylases) possèdent une activité SUMO ligase leur permettant de SUMOyler des substrats tels que MEF2 et HIC1 (Hypermethylated in cancer 1), ou encore PML (Promyelocytic leukemia protein)^{82, 83}. La protéine PML est la protéine la plus SUMOylée de la cellule. Par ailleurs, les protéines de la famille TRIM possédant un domaine RING avec un motif TRIM/RBCC ont permis la SUMOylation de la protéine p53 (« gardien du génome ») et de sa protéine associée MDM2. Les protéines TRIM jouent un rôle dans la réponse antivirale puisque la plupart de ces protéines sont induites par les interférons de type I et II. De plus, elles montrent également un rôle dans la suppression antivirale au moyen de différents mécanismes passant par l'ubiquitination de nombreux effecteurs. Ainsi, la famille des protéines TRIM constitue une nouvelle classe de ligases.

A ce jour, la seule ligase E3 identifiée comme ayant une activité pour le monomère ubiquitine et celui de SUMO est la ligase Topors (Topoisomerase I-binding arginine/serine-rich protein). On la retrouve essentiellement en relation avec des protéines nucléaires, notamment celles associées à la chromatine^{84, 85}.

5. Les enzymes impliquées dans la désSUMOylation des protéines

Un aspect important de la SUMOylation est qu'il s'agit d'un processus hautement dynamique où la déconjugaison (désSUMOylation) est catalysée par une famille de cystéine protéases appelée SUMO protéases ou désSUMOylases. Les enzymes de déconjugaison de cette famille fonctionnent à la fois pour les conjugués SUMO et pour la dépolymérisation des chaînes polySUMO-2/3. Ainsi, la protéine SUMO libre peut être générée de deux façons : soit par synthèse d'une nouvelle protéine SUMO, soit par la désSUMOylation d'un substrat cible. Les enzymes de la désSUMOylation ont d'abord été caractérisées chez la levure, elles contiennent un domaine C-terminal d'environ 200 acides aminés, nommé domaine Ulp⁸⁶. Il possède l'activité protéase et son domaine N-terminal est variable permettant ainsi la régulation de la localisation des enzymes au sein des différents compartiments de la cellule eucaryote^{87, 88}. Chez les mammifères, ce sont les protéines de la famille des SENP qui sont les plus représentées et caractérisées⁸⁹. Elles partagent un mécanisme catalytique similaire mais appartiennent à différentes super-familles qui se distinguent par le repliement de leurs domaines catalytiques. En effet, les protéines Ulp/ SENP appartiennent au même groupe de cystéine protéases dont le membre fondateur est Ulp1, découverte pour la première fois dans la levure *S. cerevisiae*. Par la suite, Ulp2 a été identifiée comme une seconde désSUMOylase de la levure. Par conséquent, toutes les SENP contiennent un domaine C-terminal commun d'environ 250 acides aminés (domaine Ulp) responsable de l'activité catalytique. Cependant, elles diffèrent au niveau de leur domaine N-terminal qui est responsable de leur localisation au sein de la cellule ainsi que de la spécificité de substrat^{86, 88}.

Le génome humain code six SUMO protéases de la famille Ulp/SENP : SENP1, -2, -3, -5, -6 et -7. Des expériences biochimiques ont été réalisées dans le but d'établir le rôle de SENP1 dans le processus de maturation et/ou de désSUMOylation (déconjugaison). Les résultats montrent que la protéine SUMO-1 est maturée plus rapidement que les protéines SUMO-2 et -3 par cette enzyme. Récemment, des études de génomique sur le modèle murin montrent que SENP1 a un rôle restreint dans l'activité de déconjugaison de conjugués SUMO-2 et -3, mais s'avère être essentiel pour la désSUMOylation des protéines modifiées par SUMO-1. Ainsi, bien que certains membres de la famille SENP (SENP1 et -2) déconjuguent aussi bien SUMO-1 que SUMO-2 ou -3 de leur substrats, d'autres (SENP3 et -5) déconjugent préférentiellement les protéines modifiées par SUMO-2 ou -3^{54, 90, 91}. Enfin, chez les levures comme chez l'Homme différents travaux ont mis en évidence, une autre classe de protéines impliquée dans la dégradation spécifique des protéines polySUMOylées⁹². Ces protéines appartiennent à la famille des ubiquitine-ligases E3 à domaine RING et ont notamment été nommées STUbL (SUMO Targeted Ubiquitin Ligases)^{93, 94}.

6. Motifs régulateurs de la SUMOylation

6.1. Les sites consensus de la SUMOylation

6.1.1 La séquence consensus

La SUMOylation a pour cible un résidu lysine (K) retrouvé dans la majorité des cas au sein d'un site consensus correspondant à la séquence canonique $\Psi KxE/D$ où Ψ est un acide aminé hydrophobe, x correspond à n'importe quel résidu et E/D est un résidu acide, généralement un acide glutamique^{95, 96} (Tableau 4).

La découverte du motif consensus a permis d'imaginer plusieurs algorithmes de prédiction des sites de SUMOylation potentiels pour une protéine donnée (SUMOplot et SUMOp). Cependant, l'incorporation de ce motif dans une protéine n'est pas suffisante pour que la SUMOylation soit initiée *in vitro*. *In vivo* un signal d'importation dans le noyau (séquence NLS) est aussi nécessaire^{97, 98}. Tenant compte de ces observations, il est désormais établi que la présence du site consensus doit être associée à des critères structuraux. La résolution du complexe protéique Ubc9-RanGAP1-SUMO-1 (lien isopeptidique entre SUMO-1 et la lysine K524 de RanGAP1) par diffraction des rayons X a permis d'établir les principales caractéristiques structurales du site consensus. Cette structure donne des informations importantes sur les critères requis pour la formation du lien isopeptidique, éléments qui ont été confortés par d'autres études⁶⁵ (Figure 16).

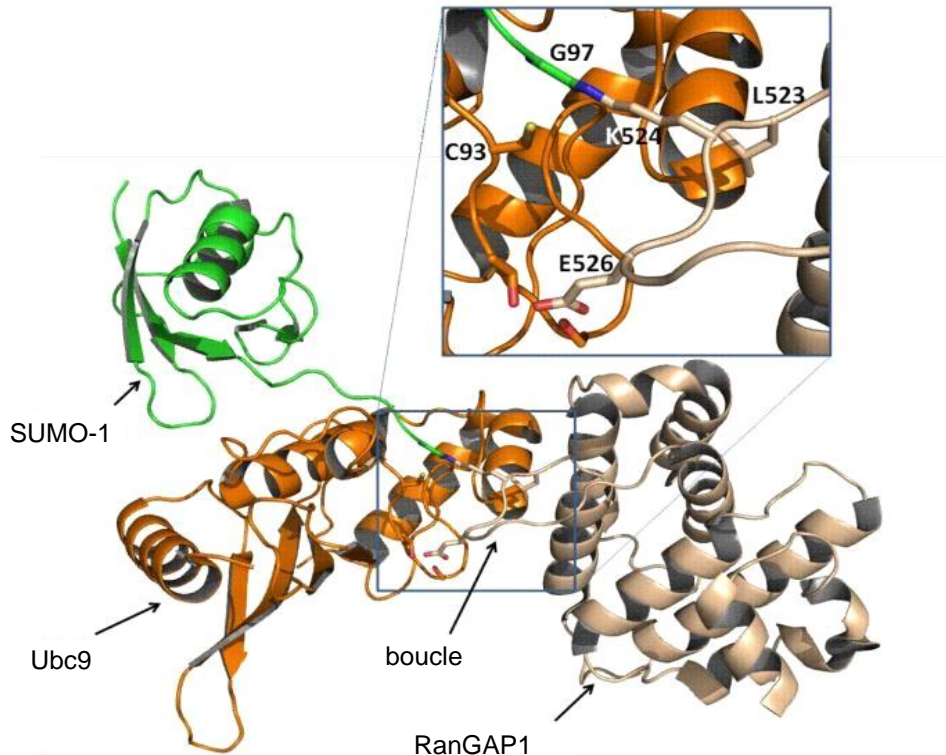


Figure 16. Structure cristallographique du complexe Ubc9:RanGAP1-SUMO-1. Dans l'encadré, on observe le lien isopeptidique entre le résidu lysine (K524) de RanGAP1 (beige) et le résidu glycine (G97) de SUMO-1 (vert). SUMO-1 est en interaction avec Ubc9 (marron). La cystéine catalytique (C93) d'Ubc9 ainsi que les résidus leucine (L523) et acide glutamique (E526) du site de SUMOylation (LKSE) de RanGAP1 sont représentés. (code pdb1Z5S)⁶⁵.

Cette structure montre que la SUMOylation demande que le site consensus de la protéine soit localisé dans une région non-structurée ou au niveau d'une boucle⁷¹. Par ailleurs, l'hypothèse reposant sur une éventuelle reconnaissance du site de SUMOylation par Ubc9 a conduit à des études d'interactions. Ces recherches montrent que les résidus qui entourent le site actif d'Ubc9 interagissent avec les acides aminés du motif consensus et ainsi participent à la reconnaissance^{65, 99}.

6.1.2. Les variants de la séquence consensus

En plus de la séquence canonique, des séquences consensus alternatives ont été décrites au moyen d'analyses protéomiques. En effet, des approches par spectrométrie de masse ont mis en évidence deux types de sites¹⁰⁰. D'une part, on a le site ICM (Inverted consensus motif) qui correspond au site consensus canonique, mais dans le sens inverse (E/DxKΨ) ce qui signifie son orientation n'est pas nécessaire à sa reconnaissance par Ubc9. D'autre part, on a le site HCSM (Hydrophobic cluster SUMO motif) qui contient une séquence riche en acides aminés hydrophobes après le résidu lysine^{97, 100, 101}.

Certaines études se sont intéressées au rôle des régions adjacentes au site consensus. On a pu constater que celles-ci pouvaient influencer de manière positive le processus de SUMOylation. En effet, lorsque le site de SUMOylation est suivi de résidus chargés négativement cela favorise la modification de la lysine cible. Dans un autre cas, la phosphorylation de résidus sérine et/ou thréonine apportant des charges négatives à proximité du site consensus permet une interaction plus étroite avec la surface basique (chargée positivement) d'Ubc9. Cette interaction électrostatique entre ces deux acteurs de la SUMOylation facilite le processus de modification^{102, 103}. Ce motif alternatif est nommé PDSM (Phosphorylation-dependent SUMO motif). Plus récemment, un autre site a été identifié, appelé pSuM (phosphorylated SUMOylation motif). D'un point de vue de la séquence, il est quasiment identique au site consensus canonique avec la particularité de présenter un résidu sérine phosphorylable à l'emplacement du résidu acide (Ψ KxS)¹⁰⁴. Ainsi, l'ajout d'une charge négative au moyen de la phosphorylation de ce résidu sérine accentue l'interaction avec Ubc9 et favorise la réaction de SUMOylation. En outre, la SUMOylation dépendante de la phosphorylation est bien souvent impliquée dans la régulation des facteurs de transcriptions (activateurs, répresseurs) en modifiant leur activité, c'est notamment le cas de PPAR γ (Peroxisome proliferator-activated receptor γ) ou MEF2 (Myocyte enhancer factor 2)¹⁰⁵. Par conséquent, ces études montrent l'importance potentielle de la phosphorylation dans la régulation de SUMOylation.

Motif consensus	Sigle	Séquence*
Consensus Motif	CM	Ψ KxE/D
Inverted Consensus Motif	ICM	E/DxK Ψ
Negatively charged amino-acid-Dependent SUMO Motif	NDSM	Ψ KxE _{xx} EEEE
Hydrophobic Cluster SUMO Motif	HCSM	$\Psi\Psi\Psi$ KxE
Phosphorylation-Dependent SUMO Motif	PDSM	Ψ KxE/D _{xx} SP
Phosphorylated SuMOylation Motif	pSuM	Ψ KxS

* Les résidus du motif consensus sont colorés en fonction de leur nature. Résidu hydrophobe (Ψ , bleu), lysine (K), acide (E/D, rouge) et sérine phosphorylable (S, magenta). x représente n'importe quel résidu (Adapté de Da Silva-Ferrada *et al.* 2012)¹⁰⁶.

Tableau 4. Les différents motifs de SUMOylation.

Au travers d'une analyse protéomique, il a été démontré que la majeure partie des protéines cibles de SUMO-2 étaient SUMOylés au niveau d'une lysine située dans le motif consensus Ψ KxE/D. Un quart des autres lysines cibles a été trouvé dans un site consensus légèrement modifié tel que PDSM ou même dans un site consensus inversé (ICM). Pour finir, 10% seulement des substrats sont SUMOylés dans un site très différent du site consensus.

6.2. Les sites non consensus de la SUMOylation

Dans le but de mieux comprendre les dessous de la SUMOylation, il était essentiel de découvrir de manière perpétuelle de nouvelles protéines cibles. Ainsi, les avancées de la bioinformatique ont permis de concevoir des algorithmes de prédictions de sites consensus (SUMOplot <http://www.abgent.com/tools/sumoplot>, SUMOp <http://SUMOp.biocuckoo.org/>). Cependant, l'idée que la seule présence du site consensus est suffisante pour confirmer un site de SUMOylation est trop restreinte. En effet, de nombreuses protéines sont retrouvées SUMOylées alors qu'elles ne présentaient pas de site consensus.

La SUMOylation de protéines en dehors des lysines du site consensus conventionnel est mal comprise. En effet, ces modifications peuvent être induites par la structure secondaire de la protéine, par un encombrement stérique imposées par d'autres MPT (compétition) ou bien par la nature des SUMO ligases E3. L'enzyme Ubc9 qui est SUMOylée sur un résidu de lysine n'appartenant pas au site consensus classique (K14). En effet, cette lysine 14 est logée sur l'hélice α H1 d'Ubc9¹⁰⁷. Autre exemple, la protéine PCNA (Proliferating cell nuclear antigen) est SUMOylée sur un résidu de lysine localisé au sein d'une structure β -turn, qui peut être aussi ubiquitinée^{108, 109}.

6.3. La SUMOylation SIM (SUMO interacting motif) - dépendante

D'autres interactions favorisant la modification d'une protéine par SUMO ont été identifiées. Notamment des interactions non-covalente de nature hydrophobe se sont révélées être bénéfiques pour la SUMOylation. Grâce à un alignement de séquence d'un ensemble de protéines SUMOylées, un motif SIM (SUMO Interacting Motif) de type $\Psi\Psi x\Psi$ ou $\Psi x\Psi\Psi$ (où Ψ correspond à un résidu hydrophobe, le plus souvent Val, Ile ou Leu et x correspond à n'importe quel résidu d'acides aminés) a été identifié comme étant favorable à la SUMOylation^{69, 110, 111}. Par ailleurs, l'étude des acides aminés adjacents à cette séquence consensus particulière a permis d'en savoir plus sur l'intérêt de sa présence au sein des différents effecteurs de la SUMOylation. En effet, au sein d'un large panel de protéines SUMOylées, ce motif est bien souvent précédé et/ou suivi de résidus acides (aspartate et/ou glutamate), de résidus sérine et/ou thréonine phosphorylables¹¹². Par conséquent, les interactions électrostatiques produites par ces résidus jouent un rôle important sur l'affinité, l'orientation et la fonctionnalité du complexe à l'origine de la formation de la liaison isopeptidique^{110, 113}. Fait intéressant, une variabilité dans la composition de cette région hydrophobe ainsi que dans l'emplacement des acides aminés chargés a été observée. Cela permet au motif SIM présent au sein de la protéine cible d'interagir avec le feuillet β 2 de la protéine SUMO selon deux orientations, soit de manière parallèle soit de manière antiparallèle^{110, 113}. Désormais, la prédiction de motifs

SIM peut être réalisée en utilisant le serveur : GPS-SBM 1.0-prediction of SUMO-binding motifs¹¹⁴

La découverte de ce motif a permis d'effectuer des recherches au sein des différents éléments de la SUMOylation, notamment au niveau des enzymes impliquées dans ce processus. En effet, des motifs SIM ont été retrouvés dans la séquence de l'enzyme d'activation E1 et les SUMO ligases E3^{62, 111}. Par ailleurs, d'autres enzymes telles que les STUbL, impliquées dans la dégradation des chaînes polySUMO, présentent des motifs SIM répétés en séries au sein de leur structure, et il faut à minima quatre unités de répétitions du motif SIM au sein de STUbL pour faciliter la reconnaissance de la chaîne polySUMO¹¹⁵. Certains motifs SIM peuvent être spécifiques soit de SUMO-1 soit de SUMO-2 et -3 expliquant dans une moindre mesure pourquoi certaines protéines seraient préférentiellement modifiées par SUMO-1 versus SUMO-2 et -3^{30, 116}.

Enfin, l'ajout de MPT sur le domaine SUMO (phosphorylation de Ser/Thr, acétylation de Lys) entraîne une régulation importante des interactions non covalentes entre SUMO et ses substrats porteurs de motifs SIM. En effet, il a été récemment démontré que l'acétylation des domaines SUMO pouvait altérer les éventuelles interactions avec certains motifs SIM¹¹⁷.

6.4. La SUMOylation dépendante d'une E3 ligase

Un autre niveau de complexité dans la régulation de la SUMOylation découle de l'ajout de MPT sur l'enzyme de conjugaison Ubc9 telles que la SUMOylation, la phosphorylation ou l'acétylation. En effet, *in vitro* l'affinité des protéines substrats est augmentée par l'autoSUMOylation de Ubc9 (K14) permettant l'interaction avec le motif SIM proche du site consensus, c'est notamment le cas pour la SUMOylation de la protéine Sp100. En effet, la SUMOylation d'Ubc9 permet de créer une interface supplémentaire Ubc9-SUMO différente de la liaison thioester améliorant la modification du régulateur de transcription SP100¹⁰⁷. Dans le cas de la phosphorylation d'Ubc9, un mécanisme similaire a été suggéré, bien qu'une analyse plus approfondie doit être menée¹¹⁸. En ce qui concerne l'acétylation d'Ubc9, cela permet de réguler de manière sélective l'interaction avec les substrats SUMO contenant le motif NDSM¹¹⁹.

En conclusion, on s'aperçoit rapidement qu'il existe des interactions de nature différente entre les différents acteurs de la SUMOylation permettant de favoriser et/ou d'inhiber la réaction de SUMOylation.

E. Interactions entre la SUMOylation et les autres modifications post-traductionnelles

Les interconnexions entre SUMOylation, ubiquitination, acétylation sont établies et la dynamique qui existe entre ces MPT se retrouve de plus en plus étudiée dans la littérature. A cette complexité s'ajoute la régulation par phosphorylation mentionnée précédemment.

1. SUMOylation versus ubiquitination

L'ubiquitination et la SUMOylation ciblent toutes les deux un résidu de lysine exposé à la surface de la protéine. Fait intéressant, une analyse protéomique révèle que 37,8% des sites de SUMOylation caractérisés sont potentiellement modifiés par l'ubiquitine également, ce qui suggère que la compétition entre l'Ub et la protéine SUMO pour la modification de certaines lysines est probablement un processus fréquent. D'une manière générale, la SUMOylation de facteurs de transcription est très souvent associée à une répression transcriptionnelle alors que leur ubiquitination, est quant à elle, associée à leur activation¹²⁰. Alors que la SUMOylation semble aider à la stabilisation des protéines, l'ubiquitination les dirige fréquemment vers une dégradation protéasomale.

La relation compétitive entre l'ubiquitination et la SUMOylation a été décrite pour la première fois lors de l'étude de la voie de signalisation du facteur de transcription NFκB. En effet, la modification d'IκBα (Inhibitor kappa-B alpha) par SUMO-1 sur sa lysine 21 (K21) empêche son ubiquitination et donc sa dégradation. La SUMOylation d'IκBα conduit donc à sa stabilisation et à promouvoir l'inhibition du facteur de transcription NFκB. De plus, il a été établi que dans ce cas l'ubiquitination d'IκBα est sous le contrôle de la phosphorylation. En effet, pour être ubiquitinée la protéine IκBα doit être préalablement phosphorylée. Cette observation a été faite au moyen de mutants phosphomimétiques d'IκBα qui ne pouvaient pas être SUMOylés¹²¹. D'autres exemples de modification d'un seul et même résidu de lysine par la SUMOylation et ubiquitination sont mis en avant. C'est notamment le cas de ERβ (estrogen receptor β) sur la lysine K4¹⁰⁴, de Smad4 sur les lysines K113 et K159¹²², de la huntingtine sur K6, K9 et K15¹²³ et de la protéine NEMO (NF-κB essential modulator ou IKKγ) sur les résidus de lysine K 277 et K309¹²⁴. Enfin en 2014, Bellail et ses collaborateurs ont démontré que la SUMOylation de CDK6 sur sa lysine K216 permet d'inhiber son ubiquitination sur un autre résidu de lysine, la lysine K147. Cette SUMOylation permet une meilleure stabilisation de CDK6 en empêchant sa dégradation, permettant ainsi la progression des cellules cancéreuses dans le cycle cellulaire¹²⁵.

La SUMOylation et l'ubiquitination n'induisent pas nécessairement de conséquences antagonistes. C'est notamment le cas de la protéine PCNA. La protéine PCNA est semblable à une pince coulissante qui entoure l'ADN servant de plateforme pour guider l'ADN polymérase et les différents facteurs liés à la réplication. Elle subit une régulation très fine par la balance

SUMOylation/ubiquitination, sur le même résidu lysine Ce mécanisme permet d'orienter le choix du type de réparation à effectuer sur l'ADN durant la phase répliative. En réponse aux dommages de l'ADN, le complexe PCNA est ubiquitiné. Cependant, si la protéine PCNA est monoubiquitinée sur la lysine K164, la réplication se poursuit malgré les cassures de l'ADN, ce qui peut entraîner des erreurs de réplication. Ce phénomène s'appelle TLS (translesion synthesis) pour synthèse transliésionnelle. Toutefois, si le processus de TLS échoue, la protéine PCNA se retrouve poly-ubiquitinée (chaînes polyUb K63) puis s'ensuit la mise en place d'une recombinaison homologue pour une réplication sans erreur. La protéine PCNA peut également être SUMOylée sur ce résidu de lysine K164 au cours de la réplication. Une fois SUMOylée, PCNA empêche la recombinaison homologue car elle recrute Srs2, une hélicase anti-recombinogène qui empêche la formation de la recombinaison Rad51^{71, 108, 126, 127}.

Cependant, la relation entre la SUMOylation et l'ubiquitination ne doit pas être exclusivement restreinte à une compétition. En effet, une interaction coopérative existe entre elles, notamment avec la découverte des ubiquitines ligases dépendantes de la SUMOylation (STUbl). C'est le cas de l'ubiquitine ligase RFN4 qui est dépendante de la polySUMOylation des protéines. En effet, la polySUMOylation de la protéine PML est reconnue par RFN4 au moyen de ses motifs SIM permettant l'ubiquitination et la dégradation de la protéine PML^{93, 128}. Ainsi, la SUMOylation peut être un point de départ pour l'ubiquitination et la dégradation.

2. SUMOylation versus phosphorylation

La phosphorylation semble également fortement impliquée dans la régulation des mécanismes de SUMOylation et/ou d'ubiquitination. Cette combinatoire apporte un degré supplémentaire de complexité à l'étude des protéines substrats, et aux mécanismes cellulaires induits par ces MPT. Aujourd'hui, il est clairement reconnu qu'il n'y a quasiment pas d'ubiquitination ou de SUMOylation qui ne soit contrôlée par une étape de phosphorylation. Comme les voies SUMOylation/phosphorylation, ubiquitination/déubiquitination ciblent souvent les mêmes résidus de lysine, le système phosphorylation/dé-phosphorylation apporte une régulation qui permet à la cellule d'orienter le type de modification à greffer sur une lysine donnée.

La SUMOylation dépendante de la phosphorylation a été traitée lors de l'identification des différentes séquences consensus. En effet, il existe une séquence de SUMOylation phospho-dépendante (motif PDSM, ψ KxExxSP)¹²⁹. Pour rappel, ce motif a été identifié au sein de la séquence de plusieurs facteurs de transcription et de protéines kinases (HIPK2/3, MAPKK2, MAK) SUMOylés^{130, 131}. De même, le motif SIM peut être régulé par phosphorylation comme mentionné pour la SUMO ligases E3 PIAS α ¹¹³. D'autre part, la phosphorylation peut influencer la SUMOylation de protéines substrats de manière plus éloignée des séquences

déjà pré-définies. En effet, la protéine MEF2D phosphorylée sur la serine S444 logée au coeur d'une séquence PDSM permet de réguler positivement la SUMOylation sur le résidu de lysine K439, alors que la phosphorylation du résidu S179 inhibe la SUMOylation⁸². D'une autre manière, la phosphorylation peut parfois entrer en concurrence avec la SUMOylation. C'est notamment le cas pour la protéine SATB1 (Special AT rich sequence binding protein 1) dont la phosphorylation altère l'interaction et le recrutement de la SUMO ligase PIAS1 et donc sa SUMOylation. Dans un dernier exemple, la phosphorylation de la protéine GCM1 (Glial cell missing 1) favorise le recrutement de SENP1 et a pour conséquence de faciliter sa déSUMOylation¹³².

3. La SUMOylation versus l'acétylation

L'acétylation interne des protéines sur les résidus de lysine est un processus distinct de l'acétylation en position N-terminale qui est un processus co- traductionnel et irréversible. L'acétylation N-terminale concerne environ 50% des protéines de la levure et plus de 80% des protéines des eucaryotes supérieurs.

La protéine acétylée la plus étudiée est la protéine p53. Elle est connue pour son rôle de suppresseur de tumeur qui déclenche une réponse lorsque l'ADN est endommagé au sein de la cellule. L'acétylation de p53 en réponse aux dommages à l'ADN¹³³ conduit au recrutement de nombreux co-régulateurs transcriptionnels et à sa relocalisation intracellulaire¹³⁴. La phosphorylation de l'extrémité N-terminale de p53 est due aux protéines kinases ATM (Ataxia Telangiectasia Mutated) et/ou ATR (Ataxia Telangiectasia and Rad3-related protein) qui ciblent la sérine S15, et Chk2 et/ou Chk1 (Checkpoint homolog) qui ciblent la sérine S20. Cette phosphorylation permet de stabiliser la protéine p53 en la dissociant de son partenaire Mdm2, une ubiquitine ligase qui promeut sa dégradation protéasomale. Au-delà de cette stabilisation, l'acétylation permet également le recrutement des acétyltransférases CBP/p300 et PCAF¹³⁵ (P300/CBP-Associated Factor) permettant l'acétylation de p53 à son extrémité C-terminale sur les résidus de lysine K373 et K382¹³⁶. L'association de ces MPT permet de stimuler l'activité transcriptionnelle de p53 avec, pour conséquence, l'induction de l'expression de nombreux gènes impliqués dans l'arrêt du cycle cellulaire et de l'apoptose. Cependant, la SUMOylation de la lysine K386 de p53 empêche son acétylation ce qui inhibe son interaction avec l'ADN et donc affecte négativement son activité transcriptionnelle¹³⁷.

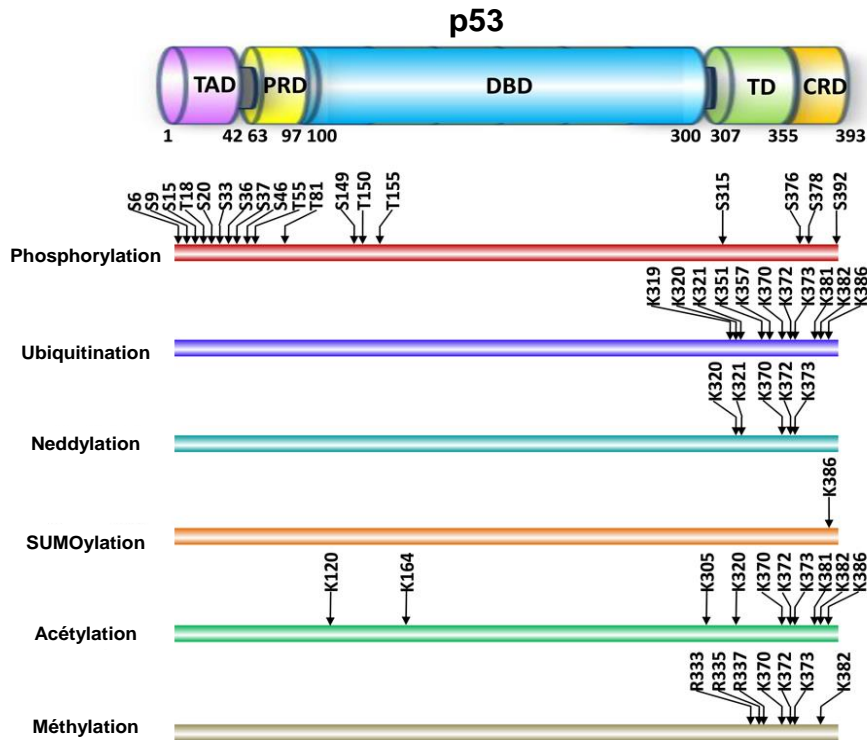


Figure 17. Ensemble des modifications post-traductionnelles de la protéine p53 identifiées à ce jour (adapté de Dai et al. 2010)¹³⁸.

La SUMOylation et l'acétylation ciblent dans de nombreux cas le même résidu de lysine. Prenons l'exemple de la protéine HIC-1 (Hypermethylated in cancer 1), dont l'activité transcriptionnelle est modulée par ces deux types de modifications. Le domaine central présente deux courts motifs conservés phylogénétiquement : le motif GLDLSKK qui permet le recrutement de son co-répresseur CtBP (C-terminal binding protein) et le motif M/VK314XEP ou «SUMO-acetyl switch», permettant un dialogue entre la SUMOylation et l'acétylation^{139, 140}. Lors de dommages à l'ADN (cassures double brin), la protéine HIC-1 est retrouvée SUMOylée suite la mise en place du complexe SIRT1-SUMO-1/HDAC4/Ubc9 responsable de la désacétylation au profit de la SUMOylation. L'acétylation de la protéine HIC-1 par p300/CBP sur sa lysine K314 permet un contrôle positif de son activité transcriptionnelle avec pour conséquence une diminution de son interaction avec son co-répresseur CtBP^{139, 141}. La complexité de l'interaction entre la SUMOylation et l'acétylation est particulièrement exacerbée dans le cas des HDAC (Histones deacetylases). Elles fonctionnent également comme des SUMO ligases E3, bien que leur fonction première soit d'hydrolyser les groupements acétylés présents sur des protéines ayant un effet répressif sur la transcription. Par exemple, HDAC4 grâce à son domaine N-terminal à caractère SUMO ligases peut induire la SUMOylation de la lysine K314 de la protéine HIC-1.

En définitive, les inter-relations entre les différentes MPT permettent une régulation très fine des processus cellulaires. Il est clairement établi que la phosphorylation, l'acétylation, l'ubiquitination et la SUMOylation sont interconnectés les uns aux autres, comme en témoigne la protéine p53 (Figure 17). Par conséquent, se focaliser sur une seule MPT ne permet pas de rendre compte de la dynamique des événements se déroulant *in cellulo*. En effet, au-delà de la compréhension de la MPT elle-même, il faut décrypter le dialogue subtil entre ces MPT, dont la dérégulation peut conduire à certaines pathologies.

III. Rôle de la SUMOylation dans la physiologie cellulaire

La SUMOylation suscite de plus en plus d'intérêt car c'est une MPT qui régule un grand nombre de processus biologiques, tels que la réparation des dommages à l'ADN, les réponses immunitaires, la cancérogenèse, la progression du cycle cellulaire et l'apoptose. La SUMOylation joue donc un rôle central au sein des réseaux complexes destinés qui régulent l'homéostasie cellulaire. Une dérégulation de la SUMOylation peut ainsi conduire au développement de pathologies telles que le cancer. Par conséquent, les protéines SUMO peuvent devenir des cibles thérapeutiques potentielles.

A. SUMO et Cancer

Le rôle important de SUMOylation dans la tumorigenèse a progressivement émergé au cours des vingt dernières années. Des altérations de l'expression ou de l'activité de différents acteurs de la SUMOylation peuvent complètement modifier la physiologie de la cellule. La dérégulation de la SUMOylation peut induire une prolifération cellulaire anarchique, une résistance à l'apoptose et accroître le potentiel métastatique en dérégulant des protéines impliquées dans la cancérogenèse^{142, 143} (Figure 18). Bien que l'association entre l'expression de divers composants dans la voie de la SUMOylation et la progression ou les métastases du cancer ne soit pas encore entièrement comprise, un nombre croissant d'études a montré que la SUMOylation joue un rôle déterminant dans le cancer¹⁴⁴.

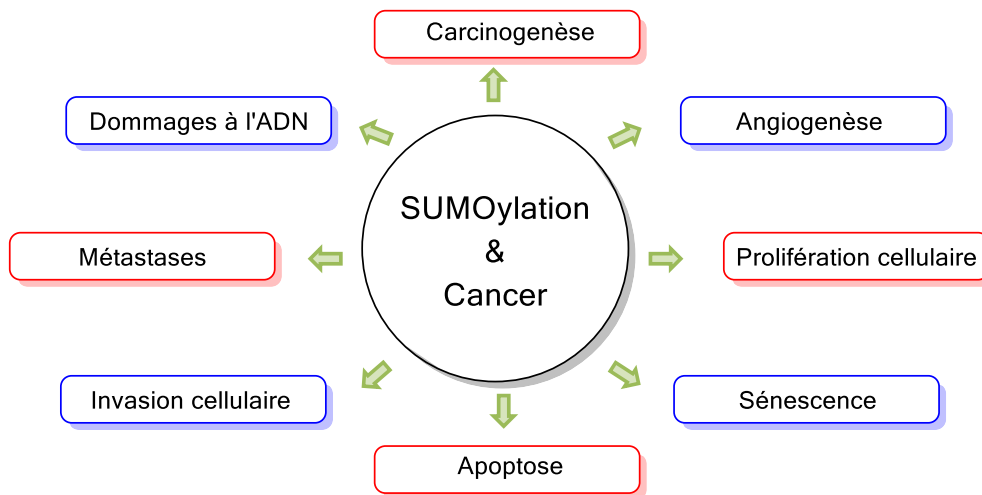


Figure 18. Liens supposés entre SUMOylation et Cancer. La SUMOylation a un impact sur la signalisation intracellulaire des cellules cancéreuses et sur toute une série de gènes régulant l'inflammation, l'immunité et les dommages à l'ADN, ce qui montre un lien entre cancérogenèse, prolifération cellulaire, métastases et l'apoptose¹⁴⁵ (D'après Han et al. 2018).

A ce propos, de nombreuses études basées sur des analyses transcriptomiques ou immunohistochimiques révèlent qu'une dérégulation de la machinerie de la SUMOylation peut être observée dans divers types de cancer (Tableau 5).

Protéine dérégulée	Sens de la dérégulation	Type de cancer	Références
SUMO-1	↑	Lymphome anaplasique à grandes cellules, côlon	¹⁴⁶ Villalva, C. et al. (2002). ¹⁴⁷ Zhang et al. (2013).
SUMO-2, SAE2	↑	Carcinome hépatocellulaire	¹⁴⁸ Lee et al. (2004).
Ubc9	↑	Cancer de l'ovaire, Leucémie myéloïde Aigüe (LMA), Mélanome, Tumeurs primaires du côlon et de la prostate, Adénocarcinomes pulmonaires	¹⁴⁹ Mo et al. (2005) ¹⁵⁰ McDoniels-Silvers et al. (2002) ¹⁵¹ Geletu et al. (2007) ¹⁵² Moschos et al. (2007) ¹⁵³ Moschos et al. (2010)
Ubc9	↓	Métastases de cancer du sein, de la prostate et du poumon	¹⁵³ Moschos et al. (2010)
PIAS3	↑	Cancers des poumons, du sein, de la prostate, du côlon, du rectum et du cerveau	⁴⁵ Wang et al. (2004)
PIASy	↓	Syndrome myélodysplasique	¹⁵⁴ Ueda et al. (2003)
SENP1	↑	Adénome oncocytaire thyroïdien, Cancer de la prostate	¹⁵⁵ Jacques et al. (2005) ¹⁵⁶ Cheng et al. (2006)

Tableau 5. Dérégulation des composants de la SUMOylation dans différents types de cancers (adapté et mis à jour d'après Lee et al. 2009)¹⁵⁷. Flèche vers le haut signifie transcription augmentée, flèche vers le bas signifie transcription diminuée.

Actuellement, les deux exemples pour lesquels la SUMOylation s'avère jouer un rôle majeur dans la cancérogenèse concernent la leucémie aigüe promyélocytaire (LAP) et le cancer de la prostate. Ainsi, nous nous focaliserons sur ces exemples qui témoignent de l'importance d'une meilleure compréhension de la SUMOylation à des fins thérapeutiques.

1. Leucémie aigüe promyélocytaire

La LAP est une forme agressive de leucémie qui provoque des hémorragies soudaines mettant rapidement en jeu le pronostic vital. C'est une maladie rare car elle représente environ 100 nouveaux cas par an en France. Cette leucémie est la conséquence d'une translocation chromosomique donnant naissance à des protéines de fusion dans les cellules myéloïdes et, dans la majorité des cas, elle est causée par une translocation des chromosomes 15 et 17 (t15,17) provoquant la fusion de la protéine PML au récepteur de l'acide rétinoïque RAR α ¹⁵⁸. L'oncoprotéine PML/RAR α est responsable de l'arrêt de la différenciation et de la prolifération accrue des promyélocytes^{159, 160}. La thérapie utilisée pour les LAP consiste en l'association de l'acide trans-rétinoïque (ATRA) et du trioxyde d'arsenic (ATO). Il s'agit du premier exemple de thérapie ciblée. Dans une grande majorité des cas, les patients sont guéris grâce à ce traitement combiné qui induit de manière spécifique la dégradation de l'oncoprotéine PML/RAR α et l'élimination des cellules souches leucémiques. En quelque sorte, il s'agit de tuer spécifiquement les cellules souches cancéreuses en détruisant la protéine qui les fait proliférer et indifférenciées¹⁶¹. D'un point de vue du mécanisme, le traitement à l'arsenic provoque le maintien de la polySUMOylation de la protéine PML sur trois lysines, K380, K400 et K497, ce qui permet l'interaction avec l'enzyme StUbl, RNF4 via son domaine SIM. La protéine RNF4 provoque la polyubiquitination de la PML et de l'oncoprotéine PML/RAR α , entraînant leur dégradation par le protéasome^{93, 162}. Et la perte soudaine de PML/RAR α induit l'apoptose des cellules.

2. Cancer de la prostate

Chez l'homme, le cancer de la prostate est le cancer le plus répandu dans les pays développés. La SUMOylation joue un rôle important dans la survenue de ce type de cancer. En effet, dans la plupart des cancers de la prostate, la dérégulation de la SUMOylation ou de la désSUMOylation a des effets sur l'expression des oncogènes, l'expression des suppresseurs de tumeurs et sur certains intermédiaires du cycle cellulaire. Par exemple, la désSUMOylase SENP1 se retrouve fortement induite dans la majorité des cas du cancer de la prostate^{156, 163}. L'augmentation de son activité est le résultat de la surexpression de l'enzyme elle-même stimulée par les androgènes (AR) ainsi que les cytokines pro-inflammatoires telles que l'interleukine 6 (IL-6). La surexpression de SENP1 a pour conséquence l'augmentation de

l'activité transcriptionnelle de différents facteurs tels que le récepteur des androgènes, de l'oncogène c-jun et de la cycline D1^{156, 164}. Une étude chez la souris montre que la surexpression de SENP1 induite par les androgènes conduit à une néoplasie intra-épithéliale de la prostate de haut grade (HGPIIN) en modulant des voies régulant l'angiogenèse, un mécanisme important du développement des néoplasies.

Un autre exemple est particulièrement bien illustré : à propos de l'enzyme de conjugaison Ubc9 et de la SUMO ligase PIAS1 qui se retrouve impliqués dans le cancer de la prostate. Ubc9 est impliquée dans la transcription des gènes cibles des AR, et son inactivation provoque ainsi des défauts dans l'organisation nucléaire conduisant à une létalité embryonnaire^{60, 165}. Dans le cancer de la prostate primaire une sur-expression de l'enzyme Ubc9 est observée, à la différence dans les états métastatiques où il s'agit d'une diminution de l'expression d'Ubc9¹⁵³. Par ailleurs un déficit en SUMO ligase PIAS1 chez la souris conduit à de petits testicules avec un nombre réduit de spermatozoïdes, ce qui est logique dans la mesure où cette SUMO ligase est fortement exprimée dans les testicules¹⁶⁶. PIAS1 (protein inhibitor of activated STAT-1), agit aussi en tant que co-régulateur des AR, et est fortement augmenté dans les tumeurs primitives et métastatiques. En particulier, son inactivation entraîne une transcription accrue du suppresseur de tumeur p21 et une expression réduite de la protéine anti-apoptotique Mcl1. Par conséquent, PIAS1 joue donc un rôle d'interrupteur oncogénique dans le cancer de la prostate¹⁶⁷.

La SUMOylation régule également des fonctions essentielles jouant un rôle dans l'homéostasie cellulaire telles que p53, NFκB et PTEN. La SUMOylation du gène *p53* permet d'augmenter sa transcription associée à l'expression de la SUMO ligase PIAS1, cela conduit à une induction accrue de la sénescence qui agit comme un mécanisme anti-cancéreux. L'activité du facteur NFκB, est régulée positivement dans le cancer de la prostate, et se trouve réduit une fois SUMOylé par SUMO-2. La protéine PTEN (Phosphatase and TENSin homolog) joue également un rôle essentiel dans le cancer de la prostate, car la perte de ce suppresseur de tumeurs entraîne une hyper-activation de la voie PI3K (Phosphoinositide 3 Kinase)/AKT/mTOR, le cancer de la prostate¹⁶⁸.

B. SUMO et le stress cellulaire

La SUMOylation peut être également régulée lors de différents stress environnementaux que peut subir la cellule (stress osmotique, stress oxydatif, choc thermique). En effet, un état de stress cellulaire induit généralement une augmentation de la SUMOylation principalement en affectant la SUMOylation via les isoformes SUMO-2 et -3. Il existe un pool libre plus important que pour l'isoforme SUMO-1 celui-ci étant majoritairement conjugué à une protéine substrat³². De même, il a été démontré que les stress oxydatifs et liés à l'éthanol augmentent la SUMOylation globale dans la levure¹⁶⁹. Des travaux supplémentaires ont révélé que la situation n'est pas aussi simple au niveau de toutes les cibles individuelles, et qu'elle dépend de l'intensité ou de la durée du stress (Figure 18).

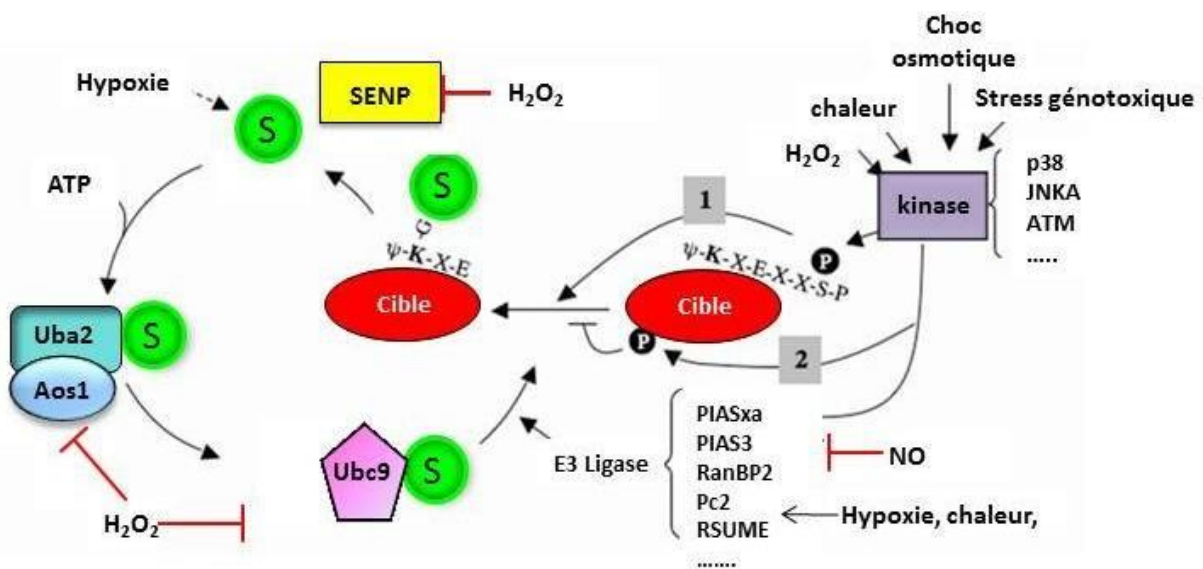


Figure 19. Régulation de la SUMOylation en réponse à différents stress environnementaux. La régulation affecte l'activité des enzymes de la voie de la SUMOylation (activation, conjugaison, ligation, S pour SUMO et SENP). Le stress active différentes kinases qui augmentent la SUMOylation en particulier dans le cas des motifs PDSM (1) ou inhibent la SUMOylation (2) (Schéma adapté de Tempé et al., 2008)¹⁷⁰

Après un choc thermique, une augmentation importante de la SUMOylation est observée, bien que l'état de SUMOylation de certains substrats reste inchangé et que dans d'autres cas, diminue. La SUMOylation du facteur de transcription HSF1 (Heat- Shock Factor 1) est fortement augmentée lors d'un choc thermique et conduit à une augmentation de sa fixation à l'ADN et de son activité¹⁷¹. Cette SUMOylation induite par sur la K298 nécessite une phosphorylation de la S303, située dans un motif PDSM¹²⁹. De même, la SUMOylation de c- Myb est fortement accentuée après un choc thermique. Toutefois, les mécanismes

impliqués n'ont pas encore été clarifiés car cette augmentation de la SUMOylation n'est pas due à l'activation de la voie p38 et JNK (c-Jun N-terminal kinase), et la phosphorylation au sein du motif PDSM de c-Myb n'a que peu d'effet. A l'opposé, la SUMOylation de la topoisomerase 1¹⁴⁹ et de la protéine PML¹⁷² se retrouve fortement réduite suite à un choc thermique. Outre ces dernières, il y a également la SUMOylation de c-Fos qui disparaît de manière concomitante à la phosphorylation d'un résidu de thréonine proche de la lysine SUMOylable¹⁷³. Ainsi, ces différents travaux ont permis d'identifier une nouvelle SUMO ligase, RSUME (RWD-containing SUMOylation enhancer), dont la transcription est activée en réponse à un stress cellulaire d'origine thermique¹⁷⁴.

Après un traitement des cellules au peroxyde d'hydrogène (H₂O₂) pour mimer un état de stress oxydatif, on observe une hyperSUMOylation, alors qu'initialement on s'attendait à ce que le mécanisme soit similaire à celui qui se produit lors d'un choc thermique. Lors d'une exposition à des concentrations élevées en H₂O₂, dites supra-physiologiques, on observe une inactivation de nombreuses déSUMOylases. A l'inverse, à faible concentration de H₂O₂, équivalente à celle observable en condition physiologique, on mesure une déSUMOylation globale rapide. Ce phénomène de déconjugaison massif est provoqué par la formation d'un point disulfure entre les cystéines catalytiques des enzymes E1 et Ubc9 de la SUMOylation, inhibant leur fonction alors que l'activité d'isopeptidase et d'endopeptidase des SENPs n'est pas affectée¹⁷⁵.

Autre point, en situation d'hypoxie cellulaire, la modification par SUMO-1 est considérablement augmentée en favorisant la transcription de nombreux gènes spécifique à l'état d'hypoxie de la cellule. Ceci suggère que la protéine SUMO-1 est importante à la cellule pour faire face à cet état de stress et d'hypoxie¹⁷⁶. L'ensemble de ces stress peuvent causer également des stress génotoxiques.

Enfin, le stress du réticulum endoplasmique (RE) dû à une accumulation de protéines non ou mal repliées sollicite également la SUMOylation pour rétablir l'homéostasie cellulaire en développant un mécanisme appelé : « Unfolded Protein Response » (UPR). Lors de ce stress, l'ARNm du facteur de transcription XBP1 est phosphorylé par une protéine kinase transmembranaire du RE, appelée Ire1, donnant naissance à des facteurs XBP1 actifs. Les facteurs XBP1 sont essentiels et responsables de l'activation de nombreux gènes associés à la restauration de la fonction du RE, tels que ceux de la voie ERAD (Regulation of Endoplasmic Reticulum-Associated Protein Degradation), ceux du repliement et de l'importation de protéines¹⁷⁷. Cependant, des recherches menées à l'aide de mutants XBP1 suggèrent que la SUMOylation a un effet négatif sur la réponse au stress du RE dans les moments d'urgence. En effet, la SUMOylation de XBP1 au moyen d'Ubc9 et de PIAS2 montre une inhibition de l'activité de transcriptionnelle des facteurs XBP1^{178, 179}. A l'inverse, une déSUMOylation de XBP1 par SENP1 renforce son activité transcriptionnelle. Ainsi, lors de faibles stress du RE,

l'UPR favorise la viabilité cellulaire, à contrario lors de stress plus important, il peut induire l'apoptose de la cellule.

C. SUMO : expression et intégrité génomique

D'un point de vue structural, le génome est organisé sous forme de chromatines composées d'ADN et de protéines appelées histones organisées en forme d'octaèdres autour desquels l'ADN s'enroule pour former le nucléosome. Cette organisation permet de compacter l'ADN. Les histones sont fortement régulées par les modifications post-traductionnelles telles que l'acétylation, la méthylation, la phosphorylation, l'ubiquitination et la SUMOylation¹⁸⁰.

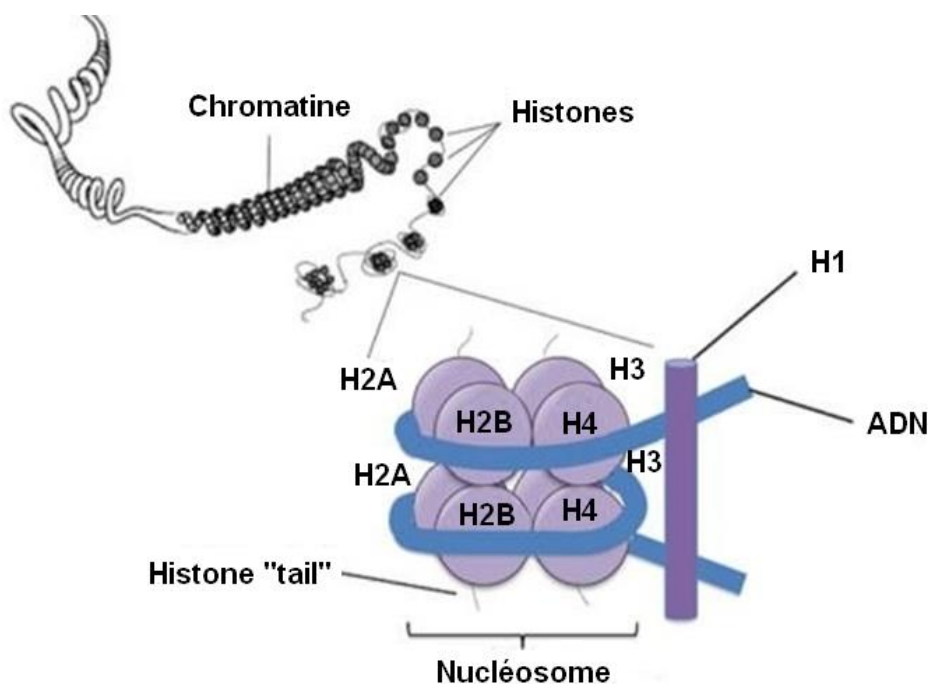


Figure 20. Schéma représentant la chromatine condensée formée d'ADN enroulé autour du nucléosome. Le nucléosome est composé de huit protéines histones. Les modifications post-traductionnelles des histones permettent de moduler l'activité transcriptionnelle.

Chez l'Homme, la SUMOylation de l'histone H4 permet le recrutement d'histones déacétylases (HDAC) et d'autres acteurs transcriptionnels de la famille HIP (Heterochromatin Protein 1) tels que HP1, γ conduisant à la répression de la transcription¹⁸¹. Dans la levure, les quatre histones principales (H2A, H2B, H3 et H4) peuvent être SUMOylées et leur SUMOylation est principalement associée à la répression transcriptionnelle¹⁸². Dans la plupart des cas, la SUMOylation est donc associée à un effet répresseur. Cependant, un nombre

grandissant d'études lie également la SUMOylation à l'activation de la transcription¹⁸³. En effet, des expériences de ChIP (ImmunoPrecipitation de la Chromatine) ont permis de démontrer pour la première fois, chez la levure *S. cerevisiae*, que de la SUMOylation est observée dans les régions promotrices des gènes actifs, qu'ils soient constitutifs ou induits, mais également absents des gènes réprimés ou non induits. Cette observation nuance l'idée que la SUMOylation soit systématiquement associée à une répression transcriptionnelle. Ces mêmes expériences ont été réalisées chez l'Homme, associées à du séquençage à haut débit ce qui a permis de collecter des informations sur l'impact de la SUMOylation sur la répression transcriptionnelle. En effet, des travaux de recherches menés par le laboratoire du Dr. Anne Dejean à Paris ont montré l'existence d'une corrélation étroite entre des régions modifiées par SUMO et des promoteurs actifs, phénomène également observé par un autre groupe de recherche^{184, 185}. Cependant, les conclusions qui en découlent sont controversées quant à l'influence de la SUMOylation sur les promoteurs des gènes actifs¹⁸⁵. En effet, si un groupe met en avant une diminution de l'expression des gènes concernés lors d'une réduction de la SUMOylation, à l'inverse, les résultats du laboratoire parisien révèlent une augmentation de l'expression des gènes les plus fortement modifiés par SUMO, codant les histones, les ARN ribosomiaux, les ARN de transferts et ceux impliqués dans la synthèse protéique. Ces résultats divergents proviennent de la méthode de modulation de la SUMOylation utilisée : le groupe de Liu a effectué une déplétion de SUMO-1 tandis que Neyret-Kahn et ses collaborateurs ont eux réduit l'expression d'Ubc9, affectant la SUMOylation globale, c'est-à-dire pour toutes les isoformes.

Au niveau chromosomique, les télomères sont des structures de nucléoprotéines essentielles qui permettent de maintenir l'intégrité du génome en protégeant les extrémités du chromosome. Ces extrémités peuvent être reconnues et traitées comme un ADN endommagé. En outre, les télomères limitent physiquement la capacité de prolifération des cellules au cours de phases de division cellulaire deviennent inévitablement limitante. Lorsque les télomères deviennent trop courts, la cellule déclenche une réaction aux dommages de l'ADN entraînant une sénescence, une apoptose ou une instabilité génomique. Ces mécanismes ont un impact sur le vieillissement et la tumorigenèse¹⁸⁶. La majorité des cellules cancéreuses surmontent cette limitation cellulaire télomériques en sur-exprimant la télomérase, une enzyme qui prolonge les télomères en ajoutant des petites répétitions d'ADN¹⁸⁷. Ces dernières années, le rôle des MPT dans les processus liés aux télomères, notamment la maintenance, la réplication et le dysfonctionnement des télomères, a considérablement progressé. Des études ont d'ailleurs montré que la SUMOylation régulait négativement l'élongation des télomères en modulant l'activité d'une protéine de liaison à ces derniers, Cdc13. Cdc13 agit comme régulateur pour le recrutement de la télomérase. Plus précisément, sa SUMOylation favorise l'association de Cdc13 avec Stn1, un inhibiteur de la télomérase, conduisant à des télomères

plus courts¹⁸⁸. Par ailleurs, des expériences réalisées chez la levure exprimant un mutant de Cdc13 ne pouvant pas être SUMOylé, des télomères de plus grande taille ont été mis en évidence¹⁸⁹. De plus, les «shelterines» qui s'assemblent en complexe aux extrémités des chromosomes afin de les protéger de l'érosion des télomères, sont également retrouvées SUMOylées^{189, 190}.

Des agents chimiques, des radiations ionisantes ou une exposition aux rayons UV induisent la formation de cassures simples et doubles brins de l'ADN. Ce stress dit génotoxique chez la levure *S. pombe* conduit à la SUMOylation de Rad52, impliquée dans les phénomènes de recombinaison homologue¹⁹¹. De même, au sein de cellules de mammifères exposées aux rayons UV, une surexpression de la SUMOylation est retrouvée sur le régulateur transcriptionnel DJ1¹⁹² (aussi connue sous le nom de PARK7 pour Parkinsonism associated deglycase 7), protéine qui régule la transcription, sur la protéine XPC (Xeroderma Pigmentosum, complementation group C), impliquée de la voie de réparation par excision de nucléotides ou NER¹⁹³ (Nucleotide Excision Repair) et sur l'histone acétyltransferase Tip60¹⁹⁴. Les agents anti-cancéreux comme la doxorubicine qui inhibe les hélicases et induisent des cassures à l'ADN sont beaucoup utilisés en chimiothérapie. Ils affectent la SUMOylation de certains facteurs comme le co-répresseur KAP-1^{195, 196} (Kinesin- Associated Protein-1) ou p53¹⁹⁷.

L'exposition répétée à des agents mutagènes oblige la cellule à disposer de mécanismes fidèles et efficaces pour la réparation de son ADN¹⁹⁸. Récemment, la SUMOylation a été associée à différents processus de réparation de l'ADN. A ce sujet, la protéine la plus étudiée est la protéine PCNA car c'est un acteur essentiel de la réplication de l'ADN. La SUMOylation de PCNA permet le recrutement de l'hélicase Srs2 qui inhibe le mécanisme de recombinaison homologue. Par ailleurs, chez l'Homme, la lysine 164 (K164) de PCNA peut être SUMOylée. En effet, la protéine PCNA peut être modifiée par SUMO-1, ce qui permet le recrutement de l'hélicase PARI (PCNA associated recombination inhibitor) empêchant la recombinaison homologue en cas de blocage de la fourche de réplication au niveau de lésions de l'ADN. Ainsi, ce mécanisme évite la formation de cassures double-brins de l'ADN¹⁹⁹ (DSB). Cependant, des expériences *in vivo* remettent actuellement en question la nécessité et le rôle de la SUMOylation²⁰⁰.

L'ensemble de ces études réalisées sur la chromatine, les télomères et les cassures à l'ADN démontrent clairement un lien entre la SUMOylation et le maintien de l'intégrité du génome, et ce, quels que soient les stress à l'origine des éventuels dommages à l'ADN.

D. SUMO et l'inflammation

L'inflammation fait partie de la réponse physiologique de l'organisme lors de conditions potentiellement nocives, telles qu'une infection ou lors de lésions tissulaires. Dans le domaine de l'oncologie, il est bien connu que l'inflammation chronique offre aux cellules tumorales un environnement favorable à leur prolifération et à leur survie^{201, 202}. En effet, de nombreuses formes de cancer sont associées à des maladies inflammatoires en lien avec une hyper activation des facteurs de transcription clés impliqués dans les processus inflammatoires, tels que NFκB ou les STAT (Signal transducers and activators of transcription)²⁰¹. D'autre part, la SUMOylation semble réguler les fonctions cellulaires de ces facteurs de transcription, soulignant un rôle de la SUMOylation dans l'inflammation physiologique et dans celle associée aux cancers.

En cas de dommages à l'ADN, NEMO est modifié par SUMO-1 induisant l'activation de la voie du NFκB²⁰³. Dans ce cas, cette SUMOylation favorise l'inflammation et agit en faveur de l'expansion de la zone tumorale. Cependant, SUMO cible également l'inhibiteur de NFκB, IκBβ empêchant son ubiquitination et par conséquent bloque sa dégradation¹²¹. SUMO agit donc comme un suppresseur de tumeur. A ce propos, des études ont été mené pour analyser le rôle de la SUMOylation dans l'inflammation, notamment l'effet du déficit en enzyme Ubc9 sur l'expression des gènes pro-inflammatoires. En effet, l'invalidation du gène *ubc9* augmente la production de cytokines pro-inflammatoires dépendantes des facteur INFβ (interferon β), TNF (Tumor Necrosis Factor), l'IL6 (interleukine-6) et le CCL4²⁰⁴ (Chemokine à ligand 4 et à motif C-C) après une stimulation modérée des cellules dendritiques (DC) par le lipopolysaccharide bactérien (LPS). Pour démontrer la pertinence de ces résultats, des études sur des modèles murins ont été réalisées. Après une stimulation par le LPS, les souris immunodéprimées qui ont reçu une greffe de moelle osseuse des souris *ubc9* *-/-* ont montré une hypersensibilité au choc endotoxinique induit par le LPS accompagné de concentrations plus élevées de cytokines pro- et anti-inflammatoires dans leur sérum comparativement aux souris *ubc9* *+/+*. Ces résultats démontrent un rôle critique de la SUMOylation dans la régulation de l'inflammation *in vivo*^{205, 206}. Ainsi, après stimulation par le LPS, c'est plus de 529 gènes qui ont été induits au moins deux fois plus que dans les cellules dendritiques *ubc9* *+/+*, par conséquent une altération importante de la SUMOylation globale affecte de manière considérable la réponse des gènes inflammatoires.

Tenant compte de ces recherches, on constate que la SUMOylation régule également l'inflammation au sein d'un organisme.

E. Bilan/Conclusion

Alors que l'ubiquitination est principalement décrite comme un signal de dégradation des protéines par le protéasome, la modification par SUMO peut modifier la localisation, la stabilité ou l'activité des protéines substrat en modulant les interactions protéine-protéine et par compétition ou interaction avec d'autres MPT. Effectivement, au travers de cette partie du manuscrit nous avons pu constater que la SUMOylation jouait un rôle important dans de nombreux processus cellulaires, ce qui place la SUMOylation au cœur de nombreuses recherches dans le domaine thérapeutique. En effet, les machineries de la SUMOylation et la déSUMOylation sont désormais considérées comme des cibles prometteuses pour le développement de médicaments, notamment contre le cancer²⁰⁷.

Contrairement à SUMO-1, les isoformes SUMO-2 et -3 contiennent un site de SUMOylation à leur extrémité N-terminale située au niveau de leur lysine 11. Cette caractéristique permet ainsi la production de chaînes polySUMO-2/3 *in vivo*. Un effort considérable est fait pour comprendre comment la SUMOylation et, en particulier, comment la modification par des chaînes polySUMO-2/3 module la structure et la fonction de la protéine substrat. Par conséquent, de nombreuses méthodologies sont développées pour accéder à la production de conjugués SUMO-2/3 homogènes et natifs, comme cela est actuellement le cas avec les conjugués ubiquitines synthétisés par voie chimique. L'accès à ces conjugués Ub ou polyUb ont permis de comprendre le rôle de l'ubiquitination, mais surtout celui de la polyubiquitination en fonction de la topologie de la chaîne polyUb présente sur la protéine à modifier (polyUb K68, polyUb K63, ...).

IV. Objectifs généraux

Actuellement, nous ne savons pas si les protéines SUMO-2 et -3 seules fonctionnent de manière distincte par rapport aux chaînes polySUMO-2/3 (de la même manière que l'Ub seule et les chaînes polyUb sont structurellement et fonctionnellement distinctes l'une de l'autre). Bien que les concepts de polySUMOylation et de polyubiquitination soient similaires, la morphologie de la chaîne générée se révèle sans doute bien différente. Une chaîne polySUMO-2/3 peut être perçue comme étant moins complexe qu'une chaîne polyUb dans la mesure où il existe au sein de l'ubiquitine 7 résidus lysine pouvant être ubiquitinylés contre une seule pour les protéines SUMO-2/3^{36, 44}. Cependant, SUMO-2 et SUMO-3 ne sont pas identiques. Etudier la complexité induite par l'ordre dans lequel les domaines SUMO-2 et -3 sont enchaînés, et pour laquelle aucune donnée n'est disponible, est un défi formidable. L'impact de ces modifications sur la structure et les fonctions des protéines cibles est un domaine encore largement inexploré.

Le but de ce projet de thèse est d'abord de clarifier s'il existe des différences structurales et fonctionnelles entre les protéines SUMO-2 et SUMO-3, puis de produire par synthèse totale les différentes combinaisons de dimères SUMO-2/3 avec un contrôle précis de leur structure de branchement afin de caractériser l'action des SENP. A terme, une meilleure compréhension des relations structure-fonction impliquant cette modification pourrait permettre de mettre au point des stratégies de criblage dans le but de découvrir des composés d'intérêt thérapeutique ciblant l'activité enzymatique des SENP.

V. Synthèse des protéines par voie chimique

A. Approches modernes de la synthèse chimique des protéines

Les peptides et les protéines jouent un rôle central dans de nombreux processus biochimiques intervenant au sein des organismes vivants. Comprendre les mécanismes et les principes moléculaires impliqués dans les propriétés structurales et fonctionnelles des protéines est une thématique importante dans la recherche scientifique, notamment dans le domaine biomédical.

Pendant de nombreuses années, la production des protéines a été assurée pour l'essentiel par les techniques de l'ADN recombinant, qui permettent de programmer l'expression de protéines au sein d'organismes vivants tels que les bactéries (la plus connue étant *Escherichia coli*). Ces méthodologies recombinantes ont donc permis de grandes avancées dans la compréhension des activités biologiques des protéines, en particulier au travers des études de relations structure/fonction permises par la facilité d'introduire des mutations dans les protéines exprimées. Bien que ce type de production se soit démocratisé au sein des laboratoires de recherche, son utilisation souffre encore de certaines limitations majeures. Une limitation majeure est la difficulté d'introduire à l'aide de ces méthodes des modifications post-traductionnelles sur des résidus précis au sein de la séquence de la protéine cible (phosphorylation, glycosylation, ubiquitination, SUMOylation, ...). Par ailleurs, l'introduction d'acides aminés non codés par le code génétique reste complexe. De nombreux développements ont rendu possible l'ajout de nombreux acides aminés non canoniques au code génétique des organismes. Néanmoins et malgré les optimisations apportées à ce jour, ces méthodologies sont réservées à des spécialistes du domaine et ne sont pas facilement utilisables au quotidien. Ces limitations ont stimulé de nouveaux développements facilitant l'incorporation d'une plus grande diversité chimique au sein des protéines et ce de manière site-spécifique.

Aujourd'hui, la synthèse des peptides et des protéines par voie chimique est perçue comme une alternative remarquable à l'utilisation des systèmes recombinants. La synthèse

chimique permet la production de protéines avec un contrôle total de sa composition. L'introduction de modifications non naturelles n'a pour seule limite que la créativité des chimistes. De plus, les protéines modifiées présentant une architecture atypique (protéines SUMOylées, protéines cycliques, protéines ramifiées, ...) difficilement produites par les systèmes recombinants peuvent être accessibles par voie chimique. Au vu de son intérêt, la synthèse chimique des protéines s'est enrichie de nombreuses avancées méthodologiques et continue son expansion de façon remarquable, comme en témoigne la Protein Chemical Synthesis database (PCS-db) (Figure 21). Cette base de données collecte les protéines d'intérêt biologique produites par synthèse chimique et comprend aujourd'hui plus de 1000 entrées. La chimie apparaît donc clairement comme une science complémentaire aux biotechnologies pour accéder à des protéines fonctionnelles²⁰⁸.

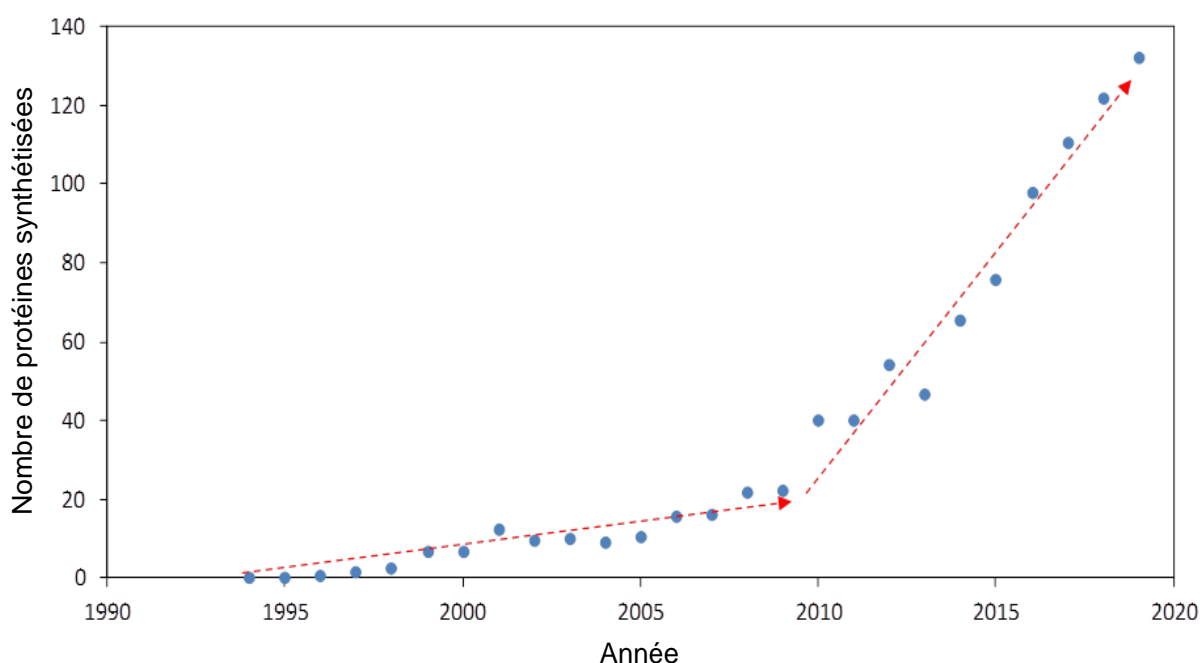


Figure 21. Nombre de protéines préparées par synthèse chimique (ligations chimiosélectives) par année sur une période de 1990 à 2020.

La synthèse des protéines par voie chimique, lorsqu'elles excèdent 150 acides aminés, constituent un véritable défi pour les chimistes. C'est pourquoi faciliter l'accès à ces molécules attire encore aujourd'hui l'attention de nombreux groupes de recherche. De nos jours, les protéines sont produites par un assemblage séquentiel de segments peptidiques non protégés grâce à des méthodes dites de ligation chimique native. Nous allons maintenant détailler les principaux concepts employés pour accéder à des protéines natives par synthèse chimique.

1. La synthèse peptidique sur phase solide

En 1963, Merrifield (prix Nobel de chimie en 1984) a décrit pour la première fois le principe de la synthèse peptidique sur phase solide ou SPPS²⁰⁹. Celle-ci permet la production de segments peptidiques par couplages successifs d'acides aminés sur un support polymérique insoluble et inerte (Figure 22). L'assemblage du peptide est réalisé par répétition du cycle déprotection/couplage (étape 2 et 3, Figure 22) du C-terminal vers le N-terminal. Une étape dite de *capping* permet de masquer les fonctions amines encore présentes résultant de couplages incomplets (réaction d'acylation, étape 4, Figure 22). Après la phase d'élongation, un traitement chimique permet d'éliminer le groupement protecteur présent sur la fonction amine du dernier acide aminé couplé au support solide. Enfin, le processus se termine par un traitement de la résine par un acide fort afin de rompre le lien acido-labile situé entre le peptide et le support solide et d'éliminer les groupements protecteurs des chaînes latérales (étape 5, Figure 22). Quelques années après son introduction, ce procédé de synthèse a rapidement été automatisé pour permettre la production de polypeptides de manière plus fiable. Par son efficacité, la SPPS a révolutionné la synthèse chimique des peptides.

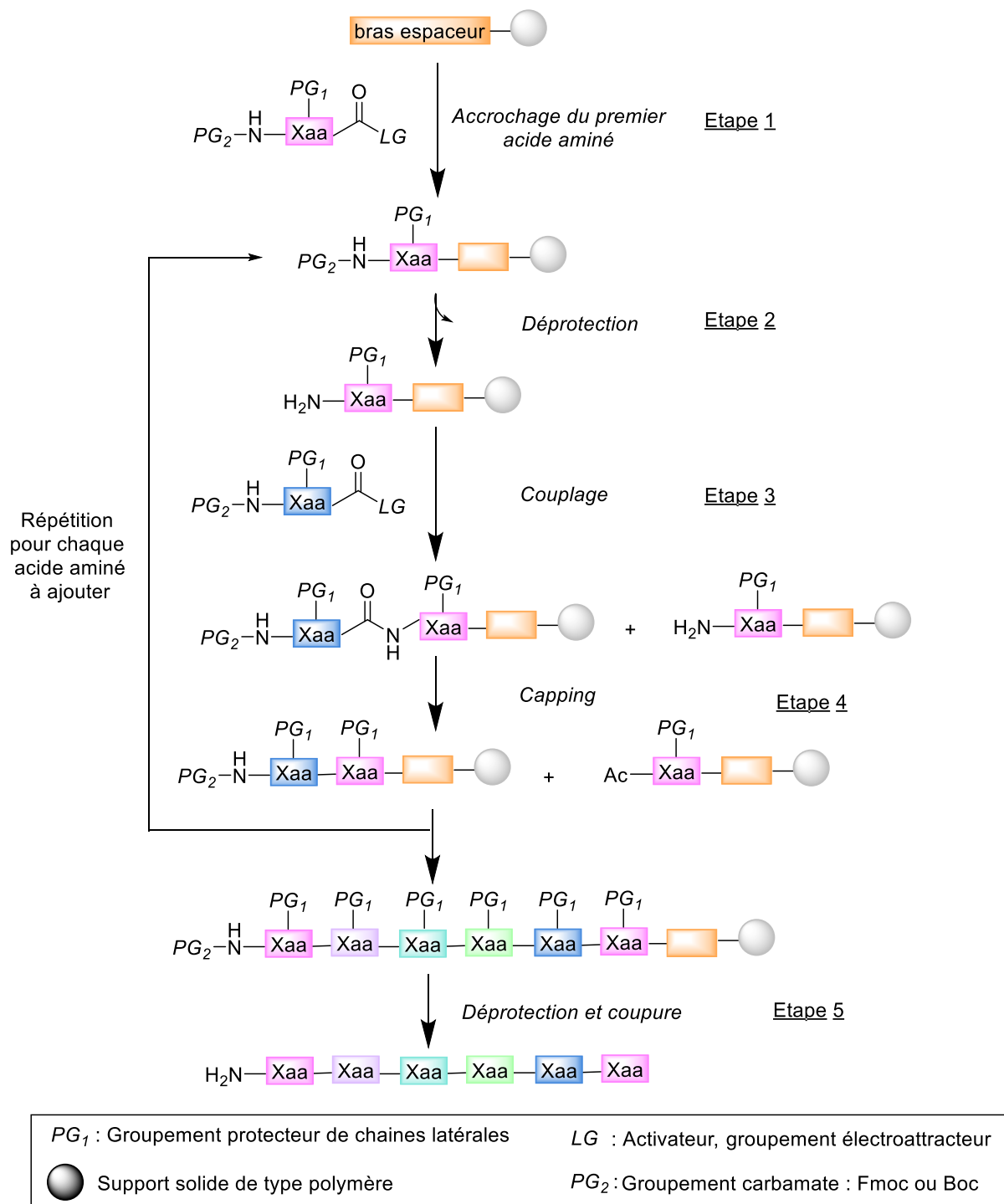


Figure 22 : Principe de la synthèse peptidique sur support solide (SPPS).

La méthodologie Boc-Bzl développée par Merrifield s'est rapidement démocratisée au sein des groupes de recherche. Elle consiste à utiliser un groupement protecteur acido-labile, *i.e.* le groupement *tert*-butoxycarbone (Boc), pour masquer les fonctions amines N-terminales, et le groupement benzyle (Bzl) comme base pour la protection des chaînes latérales. Dans ce contexte, l'amine N-terminale est déprotégée à chaque cycle après un

traitement au TFA. A la fin, le peptide est libéré et démuné de tous groupements protecteurs après un traitement à l'acide fluorhydrique anhydre (HF). Cependant, l'utilisation de l'HF au quotidien nécessite des précautions importantes lors de son utilisation puisque c'est un composé volatil (Point d'ébullition ~ 20 °C), très toxique et corrosif. D'autre part, l'utilisation de conditions acides (TFA à chaque cycle et le traitement final à l'HF) pour cette stratégie est incompatible avec la synthèse de protéines arborées de modifications post-traductionnelles (MPT) acido-sensibles telles que la glycosylation.

En 1978, Sheppard a introduit une nouvelle stratégie de synthèse, la stratégie Fmoc/*t*Bu²¹⁰. Elle utilise le groupement baso-labile 9-fluorénylméthoxycarboxyle (Fmoc) pour protéger les fonctions amines alpha des acides aminés. En général, le groupement Fmoc est éliminé après un traitement à la pipéridine²¹¹. L'intérêt du groupement Fmoc réside dans la possibilité de quantifier l'avancement le rendement d'élongation en dosant par spectrophotométrie l'adduit fluorénylméthyle-pipéridine. Cet avantage combiné à des conditions plus douces pour l'étape finale de déprotection et coupure explique la popularité de cette méthode de synthèse.

Actuellement, il existe une grande variété de polymères constituant les résines²¹², d'agents de couplage²¹³ et de groupements protecteurs de chaînes latérales²¹⁴ permettant une amélioration certaine des procédés de synthèse. Tous ces progrès permettent l'accès à des peptides composés jusqu'à 30-40 acides aminés sans trop de difficulté. A noter qu'il existe des exceptions notables dans la littérature montrant que le procédé en phase solide permet parfois la synthèse de polypeptides de plus de 100 acides aminés²¹⁵. Au-delà de ces cas particuliers, la taille des peptides produits par SPPS n'excède pas en général 40-50 acides aminés. En effet, les synthèses longues mènent inévitablement à des couplages incomplets lors de l'élongation et à des réactions secondaires, responsables de l'accumulation de sous-produits, de séparations complexes et de rendements faibles. Ces difficultés ont stimulé les chimistes à mettre au point de nouvelles méthodologies de synthèse pour obtenir des peptides de plus grande taille, à savoir les techniques dites de ligation chimique.

2. Le principe de la ligation chimique

Le terme de ligation chimique a été introduit en 1992 par M. Schnölzer et S. Kent pour décrire une réaction régio- et chimiosélective entre deux fragments peptidiques non-protégés porteurs de fonctions mutuellement réactives (fonctions X et Y, Figure 23)²¹⁶. Le caractère chimiosélectif de la réaction permet d'utiliser des fragments peptidiques totalement déprotégés. L'absence de groupements protecteurs à caractère hydrophobe minimise les problèmes de solubilité des segments peptidiques, ce qui facilite le bon déroulement de la

réaction. Le terme « ligation » vient des ligases, une classe d'enzymes capables de catalyser la formation de liaisons covalentes entre deux biomolécules.

Depuis les premiers concepts, de nombreuses réactions de ligation chimiosélectives ont été conçues. Les réactions amenant à un lien non natif ne seront pas discutées dans le cadre de cette thèse (on pourra consulter d'excellentes revues à ce sujet^{217, 218}). Bien que ces méthodes de ligations chimiques non natives ont permis d'accéder à de nombreuses macromolécules, la nécessité de produire par voie chimique des protéines conservant une structure native s'est révélée être cruciale pour l'étude, et surtout la compréhension de leurs fonctions biologiques au sein du vivant. La suite sera ainsi consacrée aux ligations dites « natives » c'est-à-dire à l'origine de la formation d'une liaison peptidique classique (Figure 23).

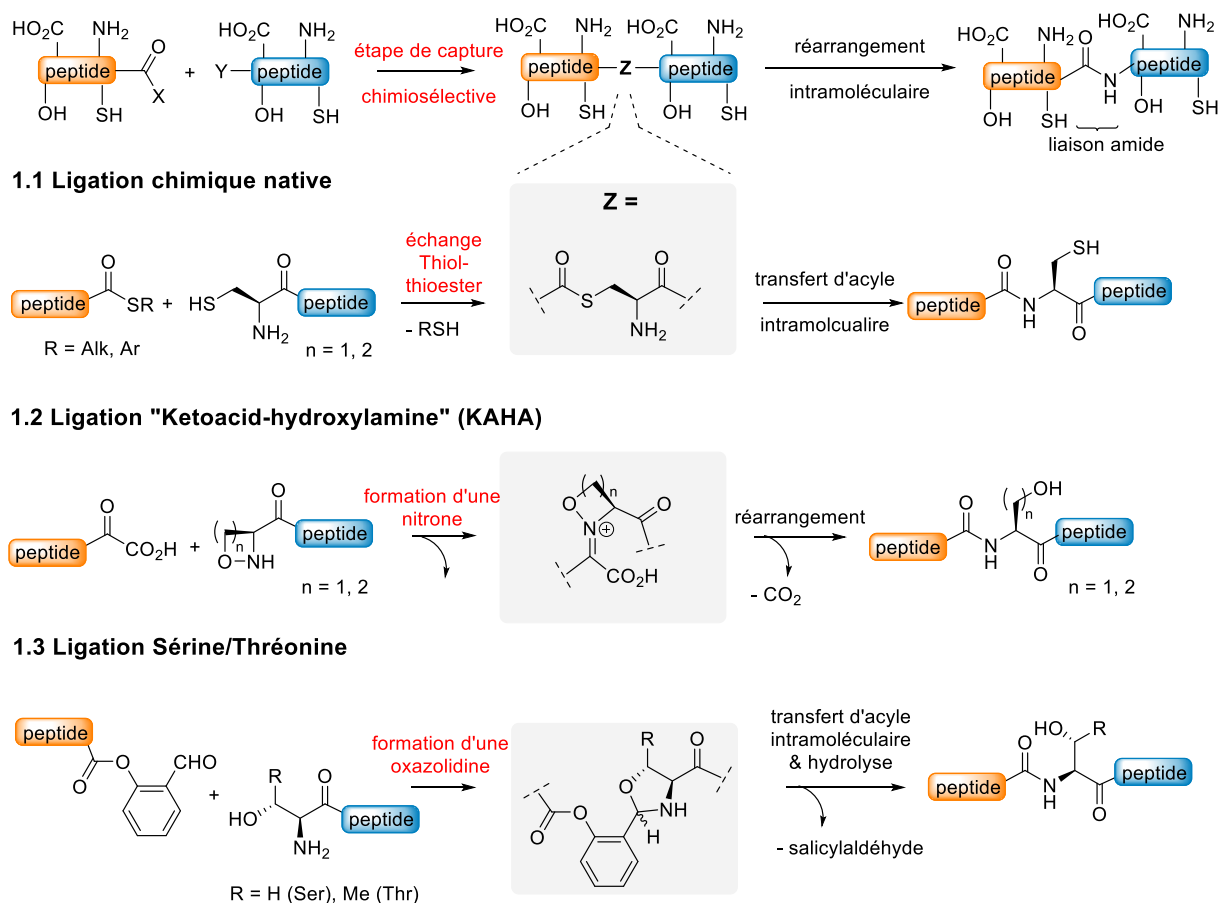


Figure 23. Différentes méthodes de formation de liaisons peptidiques basées sur le concept de capture chimiosélective suivie par un transfert d'acyle intramoléculaire. 1.1 La ligation chimique native, 1.2 La ligation « ketoacid-hydroxylamine » (KAHA) 1.3 La ligation Ser/Thr.

2.1. La ligation chimique native

En 1994, Kent et ses collaborateurs ont décrit une réaction chimique clé dans le domaine de la synthèse totale des protéines appelée « ligation chimique native » (abrégée par NCL pour *Native Chemical Ligation*)²¹⁹. Elle se réalise dans des conditions aqueuses à un pH proche de la neutralité (pH = 7), et met en jeu un peptide thioester (peptide₁, Figure 24) et un peptide comportant un résidu cystéine libre à son extrémité N-terminale (ou cystéinyl peptide, peptide₂, Figure 24). Le lien produit dans cette réaction est un lien natif de type Xaa-Cys. Cette stratégie a été illustré pour la première fois avec la synthèse de l'interleukine 8 humaine²²⁰ (IL- 8, 72 acides aminés).

D'un point de vue mécanistique, le caractère chimiosélectif de cette réaction repose sur l'étape de capture du peptide thioester (peptide₁) par la fonction thiol du résidu cystéine (peptide₂) permettant ainsi le rapprochement de la fonction amine et de la fonction thioester. L'intermédiaire transitoire formé (thioester transitoire, Figure 24) à une structure favorable pour un réarrangement intramoléculaire. Celui-ci implique la migration du groupement acyle de l'atome de soufre vers l'atome d'azote pour former le lien amide natif entre les deux segments peptidique mis en jeu (Transfert d'acyle, Figure 24). Par ailleurs, cette réaction nécessite la présence d'une étape de catalyse due à la réactivité modérée des peptides alkylthioester (peptide₁, Figure 24) et des concentrations habituellement utilisées dans cette réaction pour les segments peptidiques. Par conséquent, l'ajout en large excès d'un arylthiol tel que l'acide 4-mercaptophénylacétique (MPAA, Figure 24) permet la formation d'un peptide aryl-thioester plus réactif que le peptide alkyl-thioester de départ par des réactions équilibrées d'échange thiol-thioester (étape de catalyse, Figure 24).

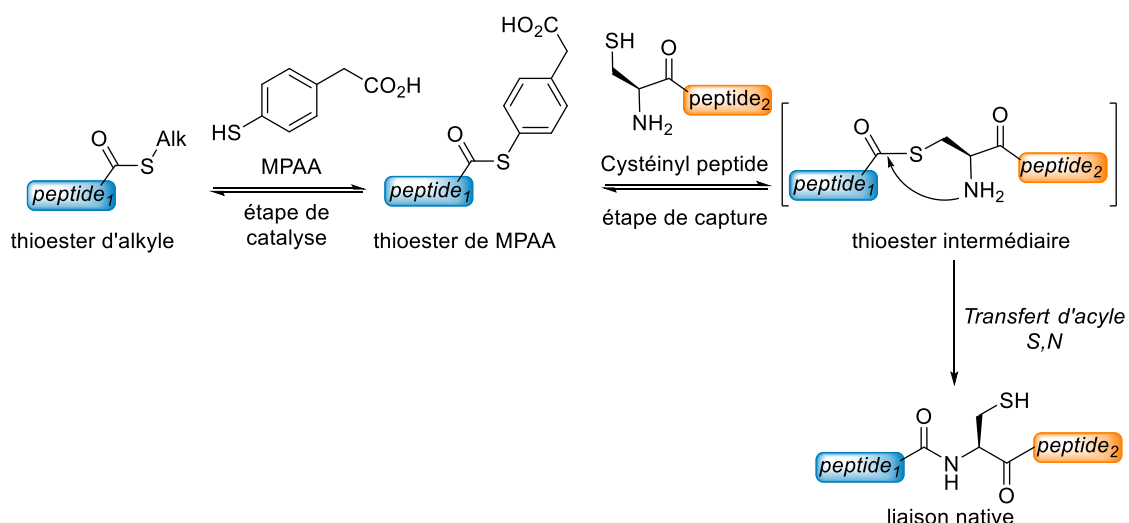


Figure 24 Mécanisme de la ligation chimique native (NCL).

Depuis 1994, la NCL a rendu possible de nombreuses applications synthétiques en chimie biologique et en chimie médicinale. Plus spécifiquement, la combinaison de la NCL et de la SPPS a largement contribué à orienter son utilisation vers la synthèse chimique de longs peptides et de protéines^{221, 222}. Des progrès continus et significatifs dans le domaine ont récemment abouti à l'assemblage de protéines de tailles exceptionnelles, comme des analogues de polymérases bactériennes^{223, 224} (~ 350 acides aminés). Au-delà de leur taille, la diversité des protéines produites traduit la polyvalence des approches de synthèse par voie chimique basées sur le principe de la NCL. Les données statistiques de la PCS démontrent qu'à l'heure actuelle la NCL est la méthode de ligation chimiosélective la plus utilisée pour la synthèse totale de protéines.

2.2. La ligation « Ketoacid-hydroxylamine » (KAHA)

Bode et ses collaborateurs ont décrit en 2006, une nouvelle méthode de ligation native basée sur la réaction d'un peptide comportant une fonction hydroxylamine en position N-terminale avec un peptide porteur d'un résidu α -cétoacide en C-terminal²²⁵ (Figure 25). Cette réaction ne nécessite pas de réactifs additionnels (ex : catalyseur). De plus, l'eau et le dioxyde de carbone sont les seuls sous-produits formés. Le mécanisme de cette ligation n'est pas entièrement élucidé et passe probablement par la formation d'une nitroène. La liaison amide est formée après un réarrangement intramoléculaire responsable d'une décarboxylation. A l'origine, cette réaction était limitée aux milieux organiques (DMF, DMSO, ou MeOH) et nécessitait des concentrations importantes de peptides (20 mM à 100 mM). Plus tard, des améliorations ont été apportées avec notamment l'utilisation d'hydroxylamines O-substituées²²⁶. Ces dernières sont plus appropriées pour la synthèse de protéines puisqu'elles autorisent l'utilisation de conditions aqueuses pour l'étape de ligation. En particulier, le résidu 5-oxaproline conduit à la formation d'une homosérine après réarrangement intramoléculaire de l'ester intermédiaire à pH basique (pH 9.5)²²⁷ (Figure 25A). Il a notamment permis la synthèse totale de petites protéines comme l'analogue de l'ubiquitine chez les procaryotes²²⁸ (protéine Pup de 66 acides aminés) et les protéines SUMO-2 et -3 humaines²²⁹ (~90 acides aminés).

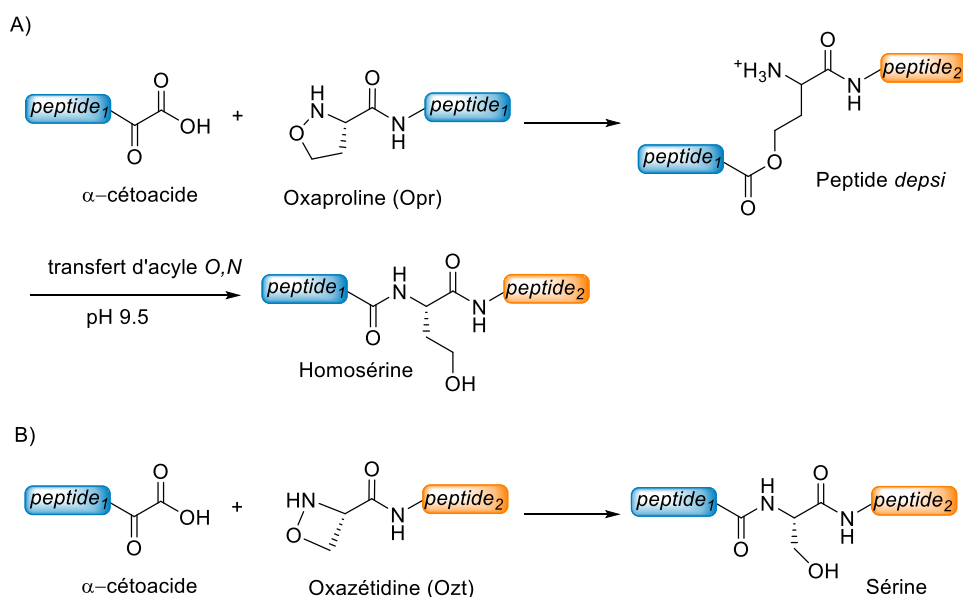


Figure 25. Optimisation de la ligation KAHA en utilisant A).un motif oxaproline. B) un motif oxazétidine.

Plus récemment, l'utilisation du résidu oxazétidine a permis la formation de lien natifs Xaa-Ser et l'utilisation de plus faibles concentrations de segments peptidiques tout en gardant des cinétiques de ligation convenables. Ceci s'explique par l'importance de la tension de cycle imposée par la présence du motif oxazétidine, d'où une plus grande réactivité. Cette variante de la ligation KAHA a rendu possible la synthèse totale de la protéine de liaison au calcium S104²³⁰ (101 acides aminés).

2.3 La ligation Ser/Thr

En 2010, Li et ses collaborateurs se sont inspirés des travaux pionniers de J. P. Tam pour mettre au point une ligation chimiosélective faisant intervenir un résidu sérine ou une thréonine en position N-terminal et un ester de salicyaldéhyde en position C-terminale^{231, 232}. Cette « ligation sérine/thréonine » (STL) est habituellement réalisée dans un mélange pyridine/acide acétique. Après évaporation du solvant, le produit brut est traité avec un mélange TFA/H₂O/TIIS pour permettre l'élimination du groupement acétal, générant ainsi une liaison peptidique native.

D'un point de vue mécanistique, la fonction amine du résidu sérine ou thréonine (peptide₂, Figure 25) se lie réversiblement au groupement aldéhyde présent du côté C-terminal de l'autre fragment peptidique (peptide₁, Figure 25). Le lien formé est de type oxazolidine. Cet intermédiaire subit un réarrangement intramoléculaire et irréversible par le biais d'un transfert

d'acyle de l'atome d'oxygène vers l'atome d'azote pour former le lien amide. Un point important de cette approche est qu'aucune épimérisation n'a lieu au niveau du site de jonction.

Comme nous avons pu le constater, la formation de peptide thioester est le point d'ancrage de tout bio-procédés conduisant à la synthèse de macromolécules par NCL ou ses méthodes apparentées (Figure 25A). Bien que les cystéinyl peptides sont faciles à préparer par la Fmoc-SPPS conventionnelle, les premiers développements des réactions de NCL ont été compliqués par la rareté des méthodes chimiques permettant l'accès à des peptides thioesters. En effet, la stratégie Fmoc implique l'utilisation de la pipéridine comme agent de déprotection, qui de par son caractère nucléophile et basique, est en général incompatible avec les fonctions thioesters (Figure 25B).

Les chimistes du peptide ont contourné cette difficulté et ont développé de nombreuses méthodologies pour accéder à des peptides thioesters en stratégie Fmoc-SPPS. Dans la suite de ce manuscrit, je ne vais décrire que les méthodes exploitant le transfert d'acyle intramoléculaire *N,S* (*N,S* acyl shift en anglais) pour produire des peptides thioesters, puisque ce sont les méthodes que j'ai principalement utilisé durant mon travail de thèse.

2.4. Le système de transfert d'acyle de X, S comme source de thioesters

De nombreuses méthodes s'appuyant sur l'utilisation du transfert d'acyle intramoléculaire *X,S* ($X = O, N$) au moyen d'une étape de transthioestérification intramoléculaire ont été développées^{233, 234}. Ces systèmes sont soit de types β -mercapto ester (transfert d'acyle $O \rightarrow S$) soit de type β -mercapto amide *N,N*-disubstitué (transfert d'acyle $N \rightarrow S$) sauf à de rares exceptions. De manière générale, ces dispositifs servent de lien entre le support solide et le peptide au moyen d'un bras acido-labile. Seulement, après l'étape de déprotection et de coupure, le groupement thiol du dispositif peut exercer son rôle de nucléophile pour conduire à la formation du peptide thioester à la suite d'un réarrangement intramoléculaire. L'utilisation du système amide plutôt qu'ester rend le réarrangement moins favorable compte tenu de la faible électrophilie de la fonction amide due à la délocalisation du doublet non-liant de l'atome d'azote vers le groupe carbonyle. Le transfert d'acyle *O, S* des peptides β -mercapto ester s'effectue toujours à pH neutre, le transfert d'acyle *N,S* y est quant à lui réalisé en général en milieu acide.

Le mécanisme réactionnel de ces systèmes de transfert d'acyle *N,S* est décrit à la Figure 26. La première étape est une réaction de migration d'acyle intramoléculaire de l'azote vers le soufre, qui est équilibrée. D'un point de vue thermodynamique, la fonction thioester n'est en général pas favorisée à pH neutre car l'amine libérée lors du réarrangement n'est pas protonée de manière significative. Ainsi, de nombreux systèmes de transfert d'acyle *N,S* ne sont déplacés par l'ajout de thiol exogène qu'en milieu acide, conditions dans lesquelles

l'amine libérée est masquée par protonation. Dans ce cas, le peptide thioester est produit et isolé avant la réaction de NCL (Figure 26).

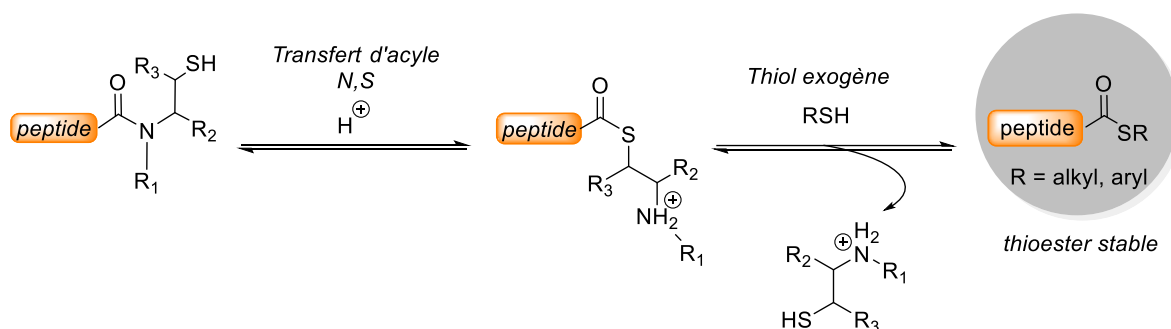


Figure 26. Principe général des systèmes de transfert d'acyle N,S basés sur le motif N-(2-sulfanyléthyl)amide.

Dans la suite de ce manuscrit, je ne décrirai que les systèmes de type β -mercapto amide N,N-disubstitués qui constituent la classe la plus importante dans le domaine. De nombreux systèmes de ce type ont été conçus dont voici un échantillon non exhaustif (Figure 27) tel que le système N-alkyl cystéine²³⁵, l'ester de cystéinyl proline^{236, 237} (CPE), N-sulfanyléthylanilide²³⁸ (SEAlide), α -méthylcystéine²³⁹, N-hydroxynitrobenzyle cystéine²⁴⁰ et bis(2-sulfanyléthyl)amido²⁴¹ (SEA).

Système de transfert d'acyle intramolécule N vers S

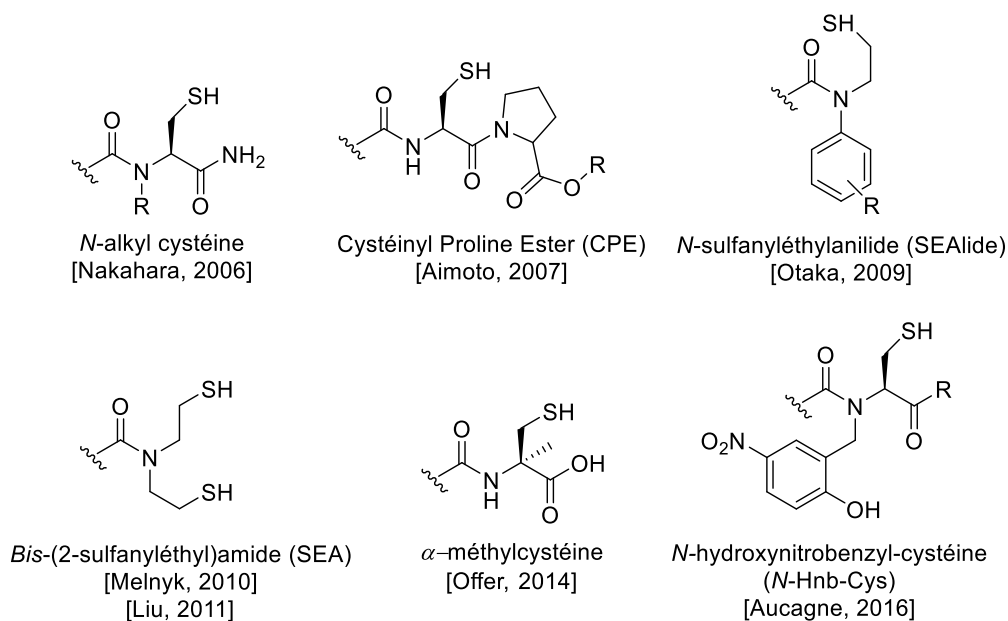


Figure 27. Système de transfert d'acyle de type N,S.

Ce principe d'échange amide/thiol est particulièrement performant et a été fréquemment utilisé dans ce travail de thèse avec le système de transfert d'acyle *N,S* bis(2-sulfanyléthyl)amido (SEA) et l'acide 3-mercaptopropionique (MPA, RSH sur la Figure 28) utilisé comme thiol.

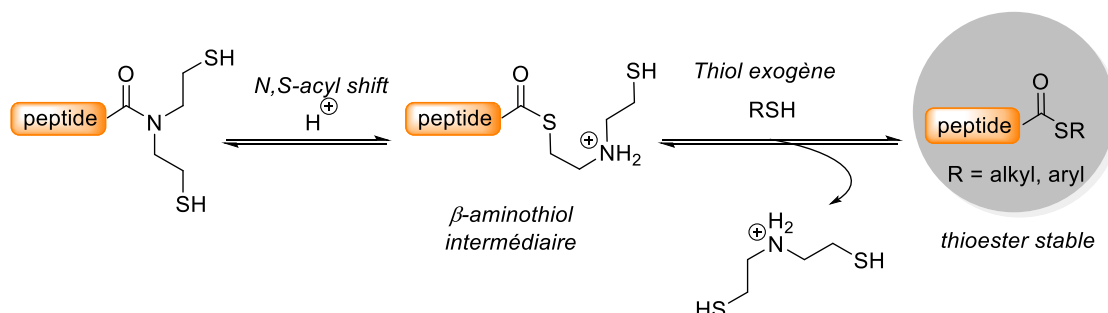


Figure 28. Formation d'un peptide thioester à partir d'un peptide bis(2-sulfanyléthyl)amido (SEA).

La transthioestérification du peptide SEA à pH acide (pH 4) assistée par un excès de MPA comme thiol exogène permet l'échange SEA-thiol pour conduire *in fine* à la formation du peptide thioester (Figure 28) qui après purification pourra être engagé dans une NCL²⁴².

2.5. Le système de transfert d'acyle de *N, S* comme source de thioesters *in situ*

De façon intéressante, une catégorie restreinte de ces systèmes de transfert d'acyle *N,S* ont la capacité de se réarranger *in situ* à pH neutre et ainsi de réagir directement avec le cystéinyl peptide (Figure 29). On parle souvent de peptides « crypto-thioesters » ou de thioesters masqués. Cette approche est particulièrement intéressante lorsque la réactivité des peptides porteurs de ce dispositif peut être contrôlée à la demande.

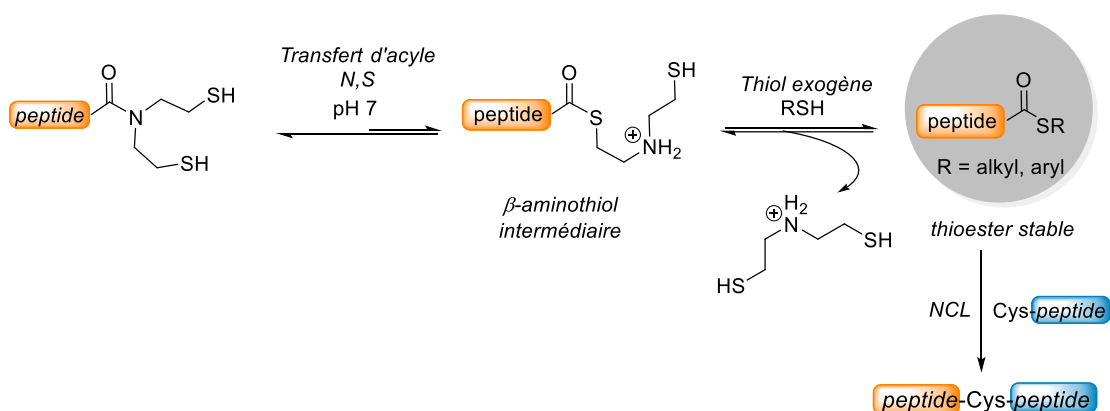


Figure 29. Ligation à l'aide de systèmes de transfert d'acyle *N,S*, illustrée avec le système *N, N*-bis(2-sulfanyléthyl)amide (SEA).

Il est à noter que l'utilisation de tels systèmes pour la formation de liens peptidiques natifs conduit à des cinétiques de réactions plus lentes que celles observées pour la NCL, l'étape de migration d'acyle N,S étant l'étape limitante. Néanmoins, les conditions requises pour que ces dispositifs agissent en tant que donneurs d'acyles sont très douces, sans risque de déclencher un réarrangement similaire avec la cystéine. Cela en fait des outils chimiques précieux comme en témoigne leur capacité à accéder à des protéines complexes.

En 2010 et 2011, les groupes de Melnyk²⁴¹ puis de Liu²⁴³ ont décrit de manière indépendante le système *N,N-bis(2-sulfanyléthyl)amido*. Dans le procédé de synthèse développé par Liu, le motif est lié à la résine par un de ses groupements thiols, à la différence du groupe Melnyk où les deux groupements thiols sont engagés avec la résine polymérique (Figure 30A). Les peptides SEA sont obtenus après un simple clivage au TFA. Les peptides sont libérés de la résine sous leur forme dithiol reactive, appelée SEA^{on}. Celle-ci peut être sujette à un transfert d'acyle *N,S* intramoléculaire, notamment durant l'étape de purification. Toutefois, le groupe Melnyk a mis en évidence la possibilité de former un disulfure cyclique après une étape d'oxydation, permettant de cette manière l'inactivation du groupement SEA sous une forme dite SEA^{off} (Figure 30A). Les peptides SEA^{off} sont incapables de se réarranger en thioester et sont donc totalement stables durant les étapes de purification par HPLC et leur stockage.

La ligation SEA sert principalement à coupler un peptide équipé d'un groupement SEA à l'extrémité C-terminale à un cystéinyl peptide. De la même manière que la NCL, cette réaction conduit à la formation d'une liaison amide native. Contrairement à la NCL qui demande de travailler à un pH proche de la neutralité, le système SEA permet de travailler dans une large gamme de pH (pH 3 - 7,5), et donc d'adapter les conditions réactionnelles en fonction de la solubilité des segments peptidiques utilisés.

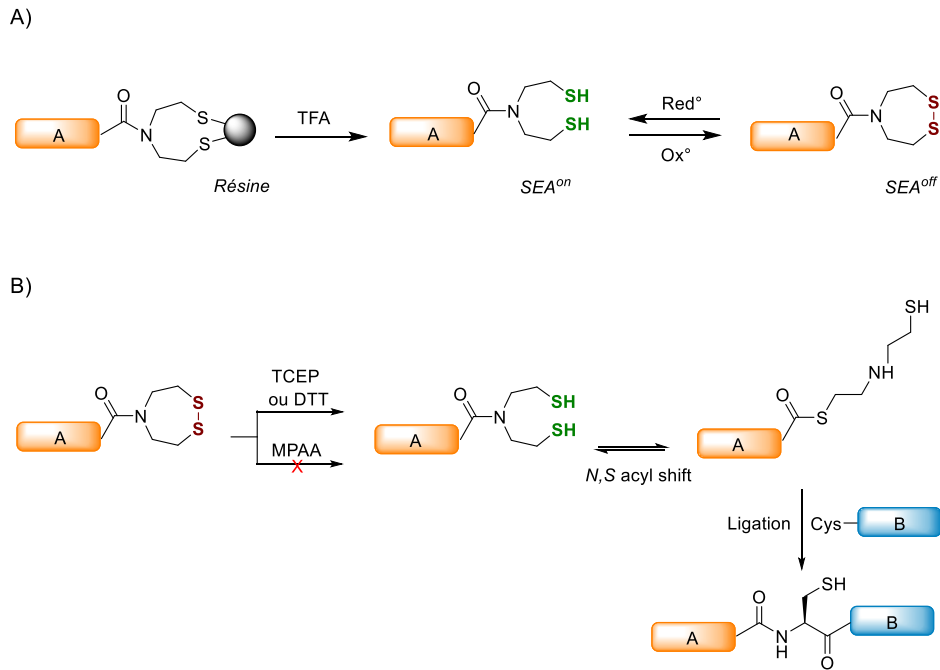


Figure 30. A) Synthèse de peptides SEA. Une étape d'oxydation post-SPPS est réalisée pour convertir le groupement SEA^{on} en SEA^{off}. B) Principe de la ligation SEA. Le groupement SEA est un thioester latent activé par réduction à l'aide de réducteurs tels que la TCEP ou le DTT.

L'une des propriétés intéressantes du groupement SEA réside dans la possibilité de passer de la forme SEA^{off} à la forme SEA^{on} en fonction du potentiel d'oxydo-réduction du milieu réactionnel. En effet, la forme SEA^{off} peut être activée très facilement à l'aide d'un agent réducteur tel que la TCEP ou le DTT, simplifiant ainsi la conception de procédés « one pot » ou le groupement SEA^{off} réagit dans une deuxième étape après augmentation du pouvoir réducteur du milieu réactionnel (Figure 30B). Par exemple, le MPAA catalysant la réaction de NCL est un arylthiol faiblement réducteur qui ne permet pas l'ouverture et donc l'activation du groupement SEA. Une réaction de NCL peut donc s'opérer en présence d'un groupement SEA^{off} sans que ce dernier vienne interférer. Puis, l'ajout de TCEP ou de DTT déclenche la formation du groupement SEA^{on} par réduction du groupement SEA^{off} et ainsi une deuxième réaction de ligation avec le cystéinyl peptide nouvellement introduit avec le réducteur.

3. Assemblage de grands peptides et de protéines

Précédemment, nous avons vu que la SPPS est utilisée pour accéder à des fragments peptidiques d'une cinquantaine d'acides aminés qui auront vocation à être engagés dans des réactions de ligations chimiques. Par conséquent, la taille du produit de ligation de deux fragments peptidiques n'excède pas une centaine d'acides aminés habituellement. Toutefois, les propriétés de la NCL et de ses méthodes apparentées telles que la ligation SEA (vu ci-dessus) permettent d'assembler plus de deux segments peptidiques. Deux approches sont

possibles en fonction du sens d'élongation choisi pour l'assemblage de la protéine, soit de l'extrémité N-terminale vers l'extrémité C-terminale (N vers C) ou l'inverse (C vers N).

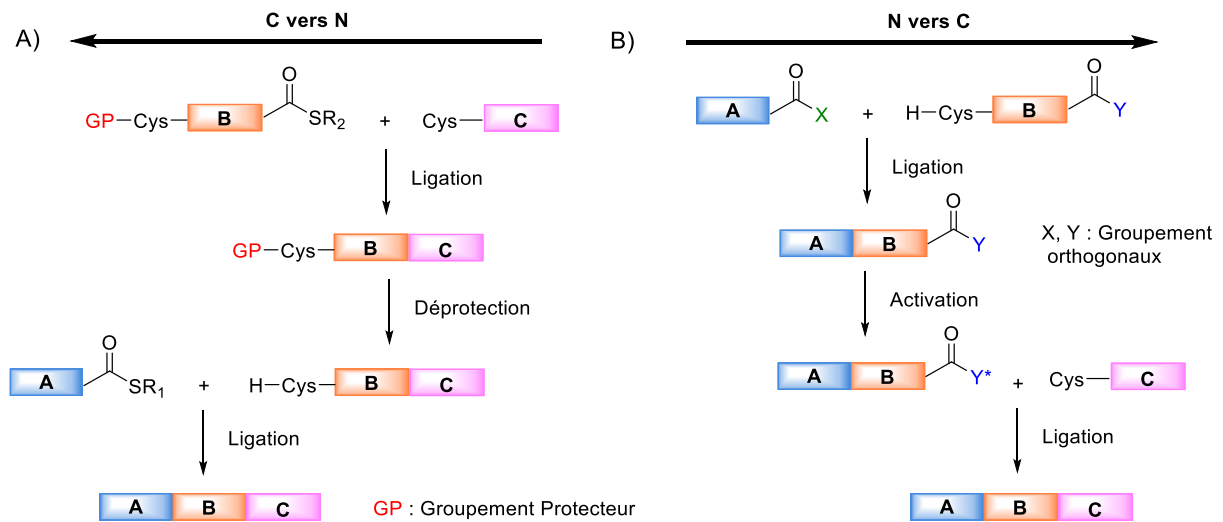


Figure 31. Stratégie d'assemblage de A) C vers N, B) N vers C.

L'assemblage de peptides de l'extrémité C vers N nécessite la protection temporaire du résidu cystéine à l'extrémité N-terminale du fragment à incrémenter (segment B, Figure 31A). Cette protection de la cystéine est essentielle pour éviter des phénomènes d'oligomérisation ou de cyclisation. Ce groupement doit être stable dans les conditions de ligation, et son élimination doit pouvoir s'effectuer dans des conditions douces en solution aqueuse. Parmi les groupements protecteurs de la cystéine, nous pouvons souligner l'utilisation du groupement protecteur de la fonction amine (Figure 32A), le 2-méthylsulfonyléthoxycarbonyl^{244, 245} (Msc), ou de la fonction thiol tels que, l'acétoacétamidométhyle^{246, 247} (Acm), ou thiazolidine^{248, 249} (Thz), ce dernier étant très utilisé. Très récemment, l'intérêt du groupement N-acétoacétyl²⁵⁰ (ACA) a été décrit par l'équipe. Ce groupement a l'avantage d'être introduit en solution par la mise en œuvre d'une ligation chimique native (NCL) (Figure 32B)^{251, 252}.

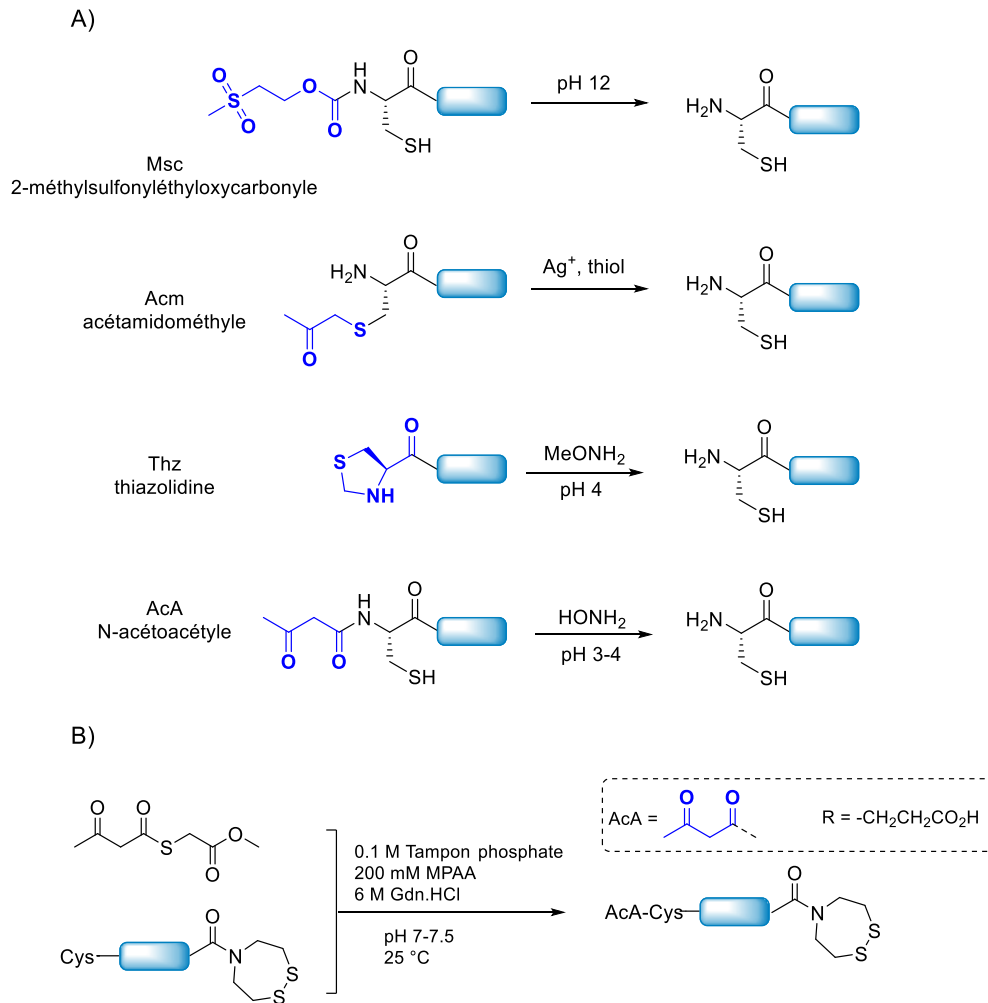


Figure 32. A) Exemples de groupements protecteurs pour l'assemblage dans le sens C vers N. B) Réaction de protection du résidu de cystéine N-terminal à l'aide du groupement N-acétoacétyle.

La stratégie d'assemblage de C vers N a permis la synthèse de nombreuses protéines, telle la sous-unité F-ATPase d'*Escherichia Coli*, une protéine de 286 acides aminés qui a été assemblée à partir de 6 segments peptidiques²⁴⁸. Cependant, l'accumulation des étapes chimiques et de purifications intermédiaires a conduit à des rendements faibles (Figure 33). Cette stratégie de synthèse illustre ainsi l'importance de trouver des méthodes d'assemblage plus simples, tout en réduisant le nombre d'étapes chimiques, y compris les étapes de purifications intermédiaires.

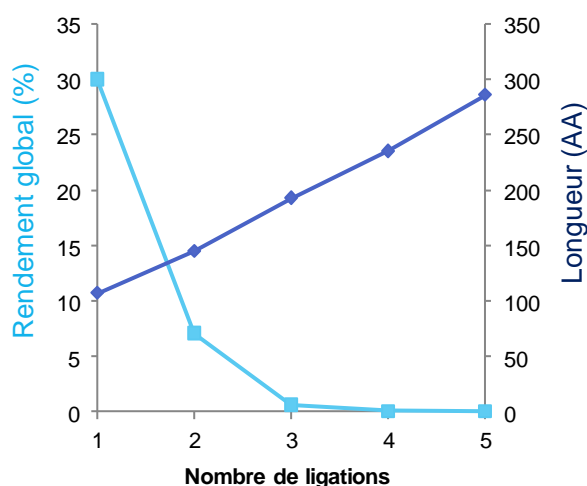
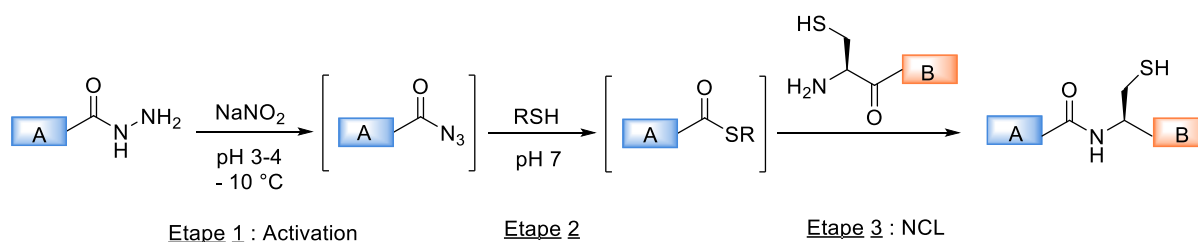


Figure 33. Évolution du rendement en fonction du nombre d'étapes de ligation pour la synthèse de la sous unité F-ATPase d'*Escherichia coli*. Les données présentées sont issues de la référence²⁴⁸.

L'assemblage des protéines dans le sens N vers C bénéficie de l'auto purification des segments peptidiques porteurs de la cystéine en leur extrémité N-terminale. C'est la stratégie qui a notamment été choisie dans ce travail de thèse. Cette approche nécessite l'utilisation de fonctions thioesters latentes à l'extrémité C-terminal du segment peptidique à incrémenter (segment B, Figure 31). Bien évidemment, les conditions pour activer ou générer la fonction thioester doivent être douces et compatibles avec l'utilisation de peptides complètement déprotégés. Le développement de tels précurseurs de peptides thioesters est un domaine encore en pleine expansion. Je ne détaillerai pas ici que les méthodologies essentiellement utilisées au cours de ma thèse.

J'ai déjà discuté les principales propriétés des systèmes de transfert d'acyle N,S comme peptides thioester masqués. De façon complémentaire, Liu et ses collaborateurs²⁵³ ont décrit l'utilisation de peptides hydrazides en tant que précurseurs de peptides thioesters. Les peptides hydrazides sont inertes dans les conditions de la NCL. Cependant, ils peuvent être convertis en azoture d'acyle par oxydation (nitrosation), puis en thioester après l'ajout d'un thiol exogène comme le MPAA. Le thioester ainsi formé peut être isolé ou mis à réagir directement avec le cystéinyl peptide dans une NCL (étape 1-3, Figure 34). Les peptides hydrazides sont donc classés dans la catégorie des peptides thioester masqués, tout comme les peptides SEA^{off}. Malgré l'utilisation de conditions oxydantes, les peptides hydrazides sont les précurseurs de thioesters les plus utilisées de nos jours. Leur préparation par Fmoc-SPPS est aisée permettant ainsi l'accès à de nombreuses cibles protéiques²⁵⁴.



R : groupement alkyle ou aryle

Figure 34. Les peptides hydrazides, précurseurs de thioesters pour la NCL.

Il est intéressant de combiner différents donneurs d'acyle, par exemple des peptides thioester, SEA et hydrazide, pour développer des procédés de synthèse « one pot » de N vers C. De tels développements ont constitué une part importante de mes travaux de thèse.

4. Les procédés de synthèse « one pot » de N vers C

Comme déjà mentionné, une des caractéristiques uniques du groupe SEA est la présence de deux fonctions thiols en étroite relation permettant une réversibilité du processus qui bloque le transfert d'acyle *N,S* par la formation d'une liaison disulfure intramoléculaire. Les propriétés d'oxydo-réduction uniques du système chimique SEA^{off} / SEA^{on} sont au cœur des méthodes d'assemblage dans le sens N vers C développées à Lille. Par exemple, il est possible d'établir une stratégie d'assemblage de trois segments peptidiques basée sur un processus séquentiel combinant la NCL et la ligation SEA (Figure 35). Ainsi, un premier peptide équipé d'un thioester d'alkyle (peptide₁) peut être engagé dans une réaction de NCL avec un peptide bifonctionnel (peptide₂), équipé d'un groupement SEA^{off} à son extrémité C-terminale et d'un résidu cystéine à son extrémité N-terminale. En absence de TCEP ou de DTT, le groupement SEA reste sous sa forme disulfure cyclique empêchant toutes réactions parasites de cyclisation ou d'oligomérisation. La réduction *in situ* du nouveau polypeptide C formé par l'ajout d'un réducteur fort dans le milieu (TCEP, DTT) active le groupement SEA en position C-terminale pour conduire à une ligation SEA avec un troisième segment peptidique (peptide₃, Figure 35).

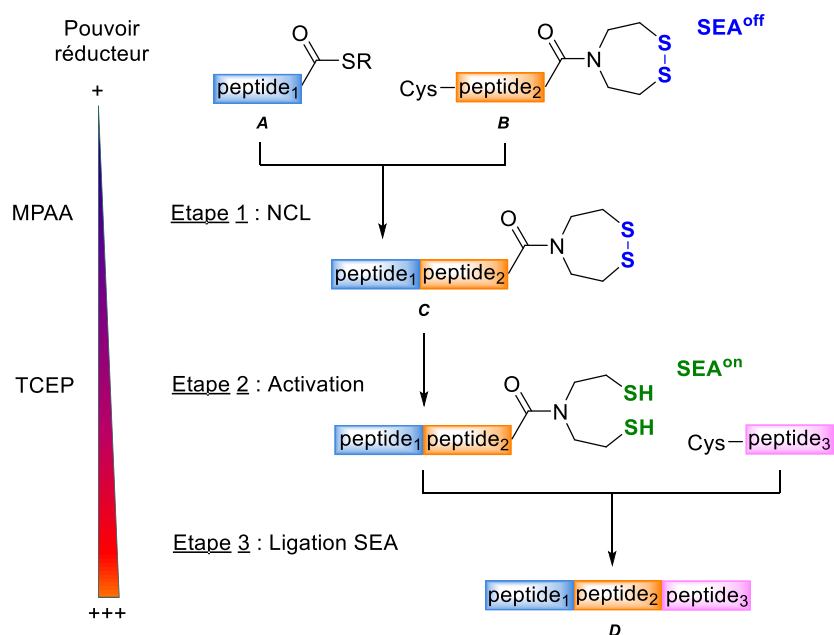


Figure 35. Assemblage « one pot » de trois segments peptidiques basé sur l'utilisation du système SEA^{off}/SEA^{on} .

Fréquemment, la formation de disulfures mixtes entre les fonctions thiol des résidus cystéine et le MPAA sont observés lors de la première l'étape de NCL en raison du faible pouvoir réducteur du MPAA. Néanmoins, la formation de ces disulfures mixtes est réversible et ne peut donc nuire à l'efficacité de la NCL. D'autre part, la particularité de la ligation SEA est la possibilité de pouvoir travailler dans une gamme de pH relativement large, notamment comprise entre 5.5 à 7.5. Une baisse du pH à 5.5 peut être avantageuse car cela favorise le transfert d'acyle N, S du motif SEA et donc augmente la vitesse de la ligation SEA. Le pH de la deuxième étape peut également tenir compte de contraintes de solubilité des peptides mis en jeu. En dessous de 5.5, le MPAA tend à précipiter du fait du pKa de la fonction thiol qui est de 6,6.

Pour illustrer l'utilité du procédé one-pot décrit ci-dessus pour accéder à une cible protéique complexe en lien avec la MPT SUMO, je discuterai les travaux récents réalisés au sein du laboratoire et portant sur la synthèse de conjugués p53-SUMO-1^{255, 256} (Figure 36).

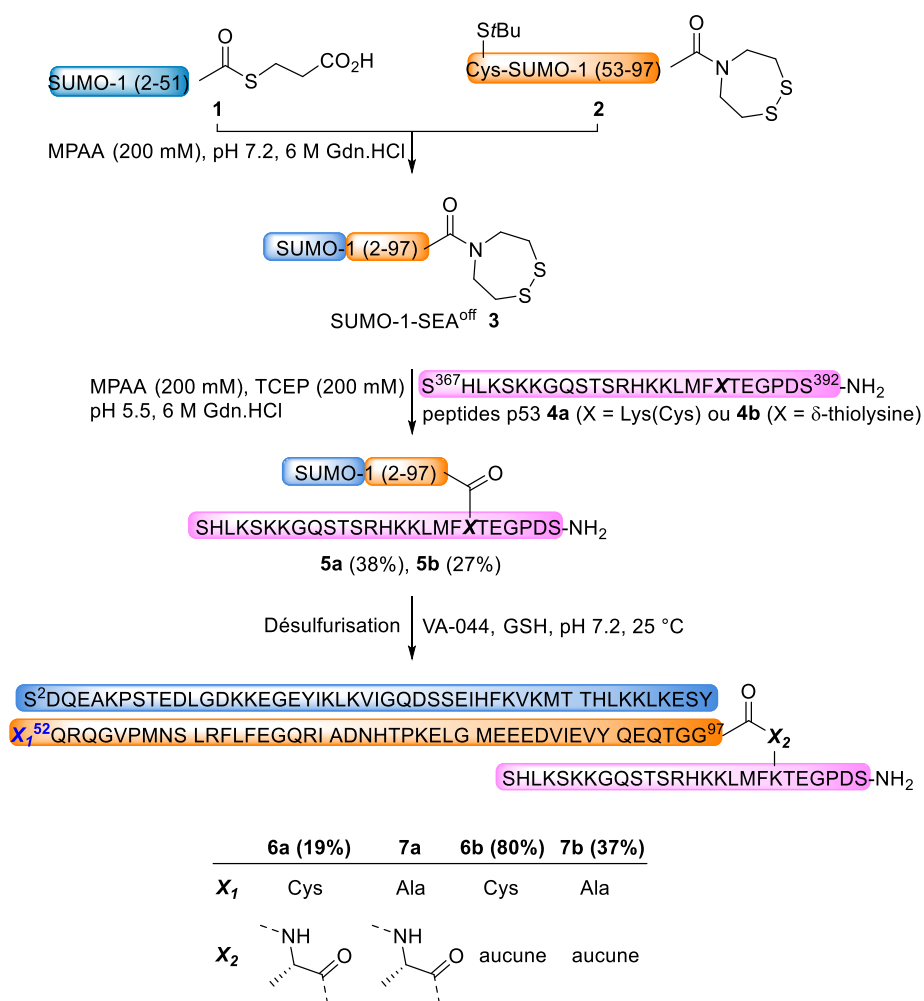


Figure 36. Stratégie de synthèse « one pot » du conjugué p53-SUMO-1.

Dans cette stratégie d'assemblage, le domaine SUMO-1-SEA^{off} 3 a été assemblé par NCL avec la seule présence de l'acide 4-mercaptophénylacétique (MPAA). Une ligation ultérieure avec le peptide cible p53 4a ou 4b en présence de MPAA et de tris (2-carboxyéthyl) phosphine (TCEP) donne lieu à la synthèse des conjugués p53 SUMO-1 5a, b sous une forme homogène avec un rendement de 27 à 38% (Figure 36). Cependant, l'accès à des conjugués p53 SUMO-1 natifs nécessite une étape de désulfuration de l'acide aminé β-amino thiol (Lys(Cys) ou δ-thiolysine) utilisé pour construire la liaison isopeptidique. La désulfuration se doit d'être sélective car la transformation de la cystéine interne du domaine SUMO-1 en alanine conduit à une perte partielle de la structure et de la fonctionnalité biochimique du conjugué.

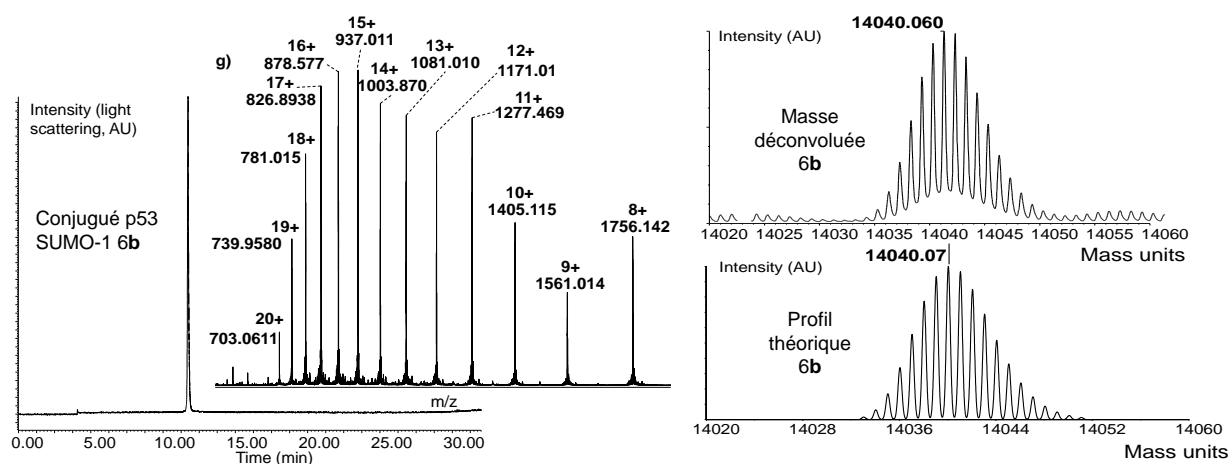


Figure 37. Analyses LC-MS et HRMS du conjugué natif p53 SUMO-1 6b purifié.

La NCL et les méthodes apparentées sont devenues des outils courants et indispensables à la synthèse totale des protéines par voie chimique. Ces méthodologies sont à l'origine de l'apparition de nombreuses revues illustrant autant les aspects synthétiques que les applications biologiques. Il n'y a pas de stratégie de synthèse universelle puisque chaque protéine possède des caractéristiques de séquence et de structure qui lui sont propres. Il est important de connaître les forces et les limitations des différentes méthodes chimiques employées afin de sélectionner la stratégie la mieux adaptée à la synthèse d'une protéine donnée. Dans la suite de ce manuscrit, je vais présenter rapidement les limitations des méthodes existantes et les développements qui ont permis de les contourner et d'étendre la gamme des protéines accessibles par synthèse totale.

B. Méthodes pour la synthèse de grandes protéines

Aujourd'hui, de nombreuses méthodes d'assemblage séquentielles à trois segments peptidiques sont disponibles dans la littérature^{249, 257, 258}. Ces méthodes permettent notamment d'économiser une étape d'isolement intermédiaire, et de simplifier l'accès à des protéines composées de moins de 150 acides aminés. Cependant, la synthèse de grandes protéines (> 150 acides aminés) nécessite souvent l'assemblage de plus de trois segments peptidiques. Par conséquent, la conception d'approches dites « one pot » pour assembler quatre segments peptidiques est d'un réel intérêt. Cependant, les approches « one-pot » nécessitent de réaliser toutes les étapes chimiques dans un seul réacteur ce qui limite considérablement la nature des groupements protecteurs et des méthodes de ligations chimiques pouvant être combinées. Sans compter que la grande difficulté de ces approches réside aussi dans l'accumulation inévitable de segments peptidiques n'ayant pas réagi ou de polypeptides délétés, source de complication à l'étape de purification finale. Pour éviter ces difficultés, il est

primordial d'avoir un contrôle de la stœchiométrie des segments peptidiques mis en jeu et des réactifs utilisés.

1. Aperçu statistique de la synthèse chimique de protéines

Bien que la taille des peptides accessibles par synthèse ait longtemps été limitée à moins de 100 acides aminés du fait des limitations inhérentes de la SPPS, la possibilité d'assembler chimiquement et ce de manière sélective deux segments peptidiques non protégés à l'aide de la ligation chimique native (NCL) ou de ses méthodes associées ont propulsé la synthèse totale de protéines dans une nouvelle ère^{219, 259, 260}. En effet, ces méthodologies ont permis la préparation de protéines de plus en plus complexes et constituent aujourd'hui des outils robustes et polyvalents pour étudier le repliement et la fonction biologique des protéines. Globalement, nous estimons que l'utilisation de la NCL et des méthodes apparentées ont abouti à la synthèse chimique d'environ 800 cibles d'importance biologique sur la période 1994 à 2018. La majorité des protéines sont composées d'environ 50 acides aminés en moyenne (452, 60%) et ont été produits à partir de deux segments peptidiques (Figure 38). Les 40% restants sont des protéines synthétisées par voie chimique avec une taille supérieure à 80 acides aminés, et ont été assemblés en effectuant deux à cinq ligations consécutives.

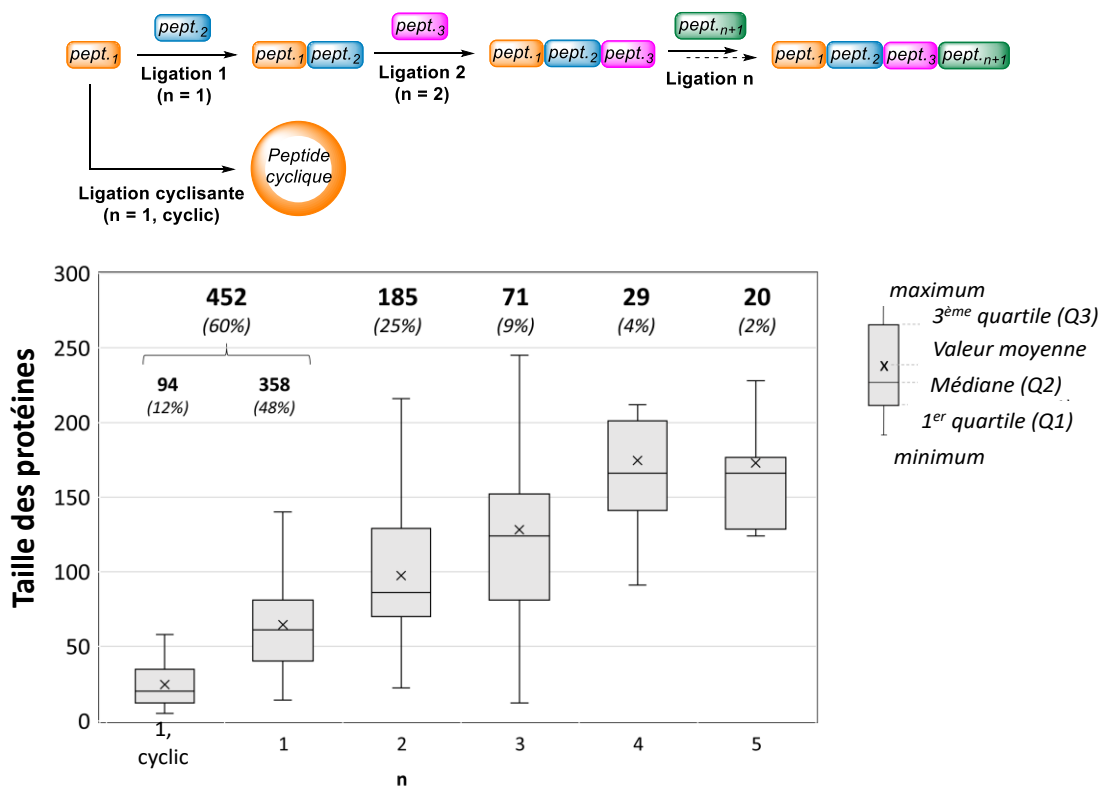


Figure 38. La taille des protéines produites par NCL et les méthodes apparentées sur la période 1994-2018 est comparée au nombre de ligations nécessaire à leur synthèse. Le nombre total de protéines synthétisées dans chaque catégorie est indiqué en caractères gras au-dessus de chaque colonne. Les ligations cyclisantes à 1 seul segment sont désignées par 1c dans le graphique. Quatre protéines ont été obtenues après 6 ligations ou plus et ne sont pas représentées sur cette figure pour plus de clarté^{224, 261, 262} (Weinstock et al. 2015 ; Xu et al. 2017 ; Tang et al. 2017). Les données ont été extraites de la Protein Chemical Synthesis (PCS) (<http://pcs-db.fr>). Il convient de noter que la base de données de la PCS comprend uniquement les peptides synthétiques et les protéines d'importance biologique et ne prend pas en compte les réactions de ligation utilisées pour la synthèse de peptides modèles, de polymères ou de matériaux hybrides.

Une limitation majeure de ces approches est la nécessité d'avoir un résidu cystéine à la jonction entre les segments peptidiques. Cette condition peut ainsi restreindre considérablement le nombre de cibles protéiques potentielles.

2. Fréquence des acides aminés protéinogéniques

La fréquence des acides aminés protéinogéniques peut être estimée à partir de la base de données UniRef50 qui regroupe un vaste échantillon de protéines avec leurs caractéristiques (Figure 39). Cette base de données nous informe sur la fréquence des différents acides aminés au sein de ce panel de protéine. Nous pouvons constater que la fréquence du résidu cystéine est faible (1,48 %) alors que celle de l'alanine est 5,7 fois supérieure (8,42 %).

En outre, Carugo a montré que la fréquence du résidu cystéine diminue avec la taille des protéines, à l'inverse celle de l'alanine augmente²⁶³ (Figure 39B). L'observation faite que les cystéines sont plus fréquemment retrouvées au sein de petites protéines peut être expliquée par le fait que ces molécules sont trop petites pour former un noyau hydrophobe et nécessitent ainsi des ponts disulfures pour stabiliser leur repliement naturel²⁶⁴. Pour illustrer ce point de manière plus concrète, prenons l'exemple d'une protéine de 50 acides aminés pour laquelle la fréquence des cystéines est de 2,25 % contre 6,68 % pour l'alanine. Comparativement, ces valeurs passent à 1,45 % pour la cystéine et 8,75 % pour l'alanine pour une protéine de 200 acides aminés. Ce constat a stimulé le développement de nouvelles méthodologies de synthèse chimique pour étendre la ligation chimique native à d'autres types de jonctions, notamment à la jonctions Xaa-Ala dont la formation est rendue possible par des réactions de désulfuration de la cystéine appliquées après l'étape de ligation^{265, 266}. J'ai choisi d'exposer brièvement, sans être exhaustive, la méthode de désulfuration la plus utilisée dans le domaine, et celle aussi que j'ai appliqué pour les conjugués SUMO synthétiques durant mon travail de thèse.

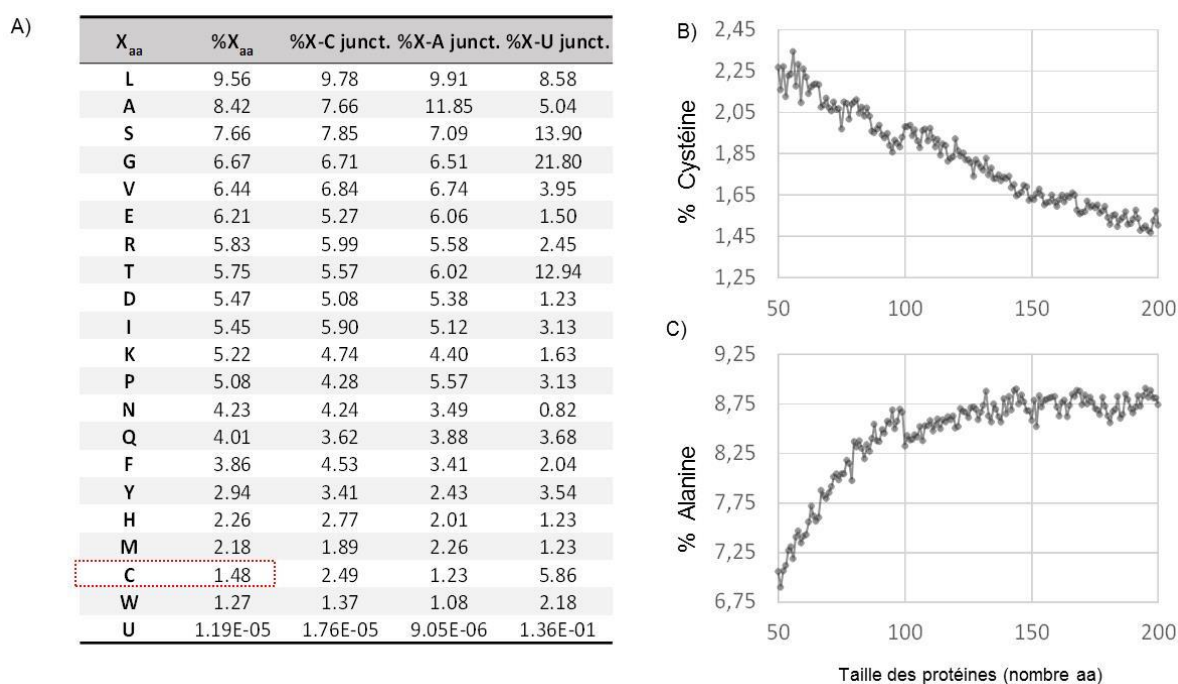


Figure 39. A) Statistiques issues de la base de données UniRef50 : fréquence des acides aminés dans les protéines. B-C) : Evolution de la fréquence des résidus Cys (a) ou Ala (b) en fonction de la taille des protéines (données reproduites à partir de Carugo et al. 2008)²⁶³.

3. Les réactions de déchalcogénation

Dawson et ses collègues ont mis au point des approches novatrices axées sur la désulfuration du résidu Cys afin de produire des liaisons peptidiques Xaa-Ala. Ces approches sont facilitées par l'introduction d'un résidu Cys dans la séquence des peptides assemblés par SPSS et par la fréquence significative des résidus Ala contenu dans les protéines. Depuis, le principe de la désulfuration du résidu Cys en Ala a été étendue à presque tous les acides aminés codés génétiquement (Figure 40).

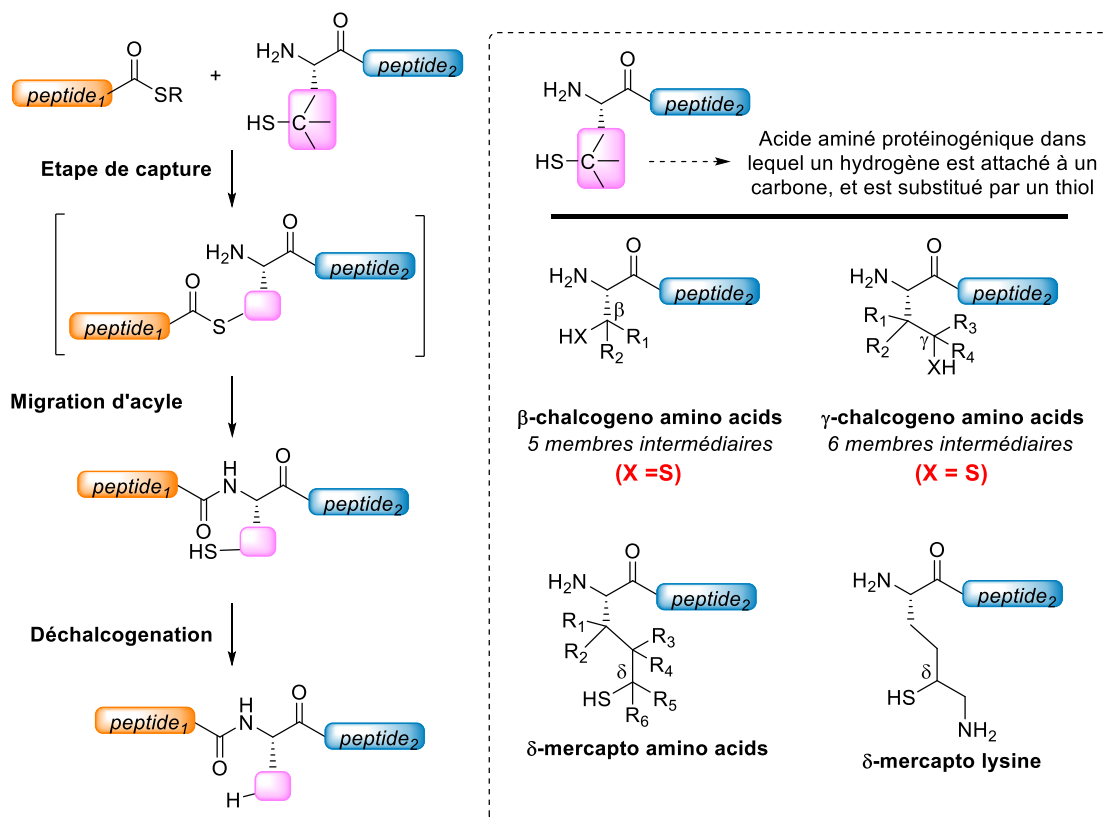


Figure 40. Les dérivés d'acides aminés porteurs d'un motif thiol permettent la formation de pratiquement toutes les jonctions en combinant la NCL avec une étape de déchalcogénation.

Ces acides aminés sont obtenus en substituant un atome d'hydrogène de la chaîne latérale par un groupe hydrogénosulfényle (HS). Le résidu est par la suite converti en un résidu protéinogénique au moyen d'une réaction de désulfuration. Le même principe donne accès à des liaisons isopeptidiques au moyen de dérivés β-, γ- ou δ-mercapto-lysine. La δ-mercaptolysine (ou δ-thiolysine) a largement été utilisée pour accéder à des conjugués ubiquitines natifs, ou comme vu précédemment à un conjugué p53 SUMO1 natif^{255, 267}. Toutefois, la synthèse de la δ (ou de la γ-mercaptolysine) qui serait utile à mon projet nécessite

de nombreuses étapes chimiques (Figure 41). Par conséquent, nous avons décidé d'utiliser à la place le motif Lys₁₁(Cys), plus facile à produire en grande quantité, pour identifier la meilleure stratégie de synthèse et réaliser les premières preuves de concept. Nous sommes ainsi venus modifier chimiquement le groupement ϵ -amine (ϵ -NH₂) d'une lysine avec un résidu cystéine.

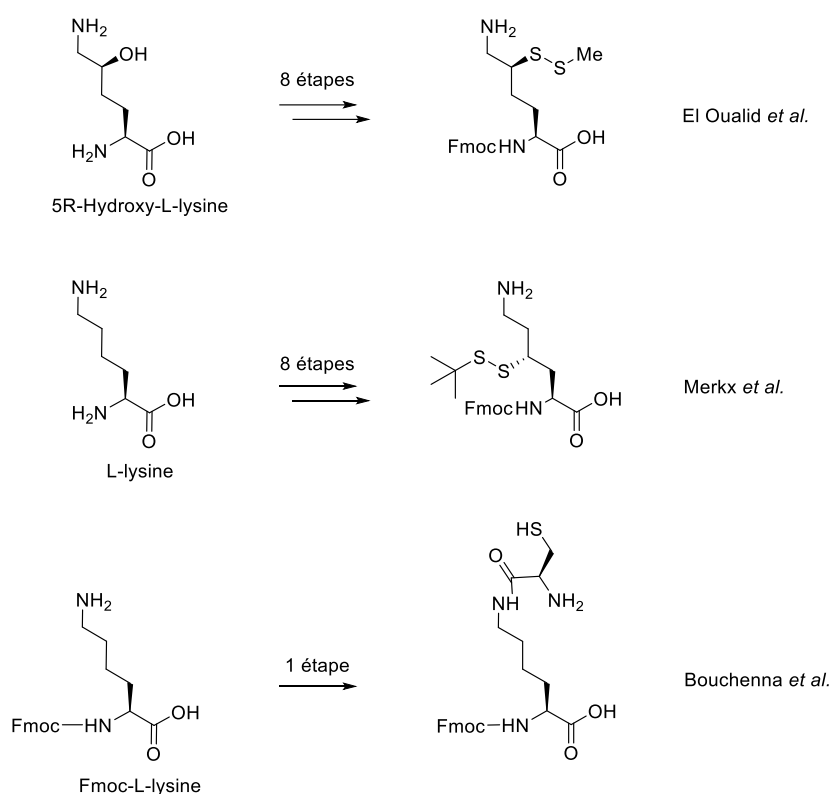


Figure 41. La synthèse de δ/γ -thiolysine à incorporer au peptide pendant la SPPS demande de nombreuses étapes chimiques.

La désulfuration d'alkylthiols par des phosphites a été observée pour la première fois par le groupe de Hoffman en 1956²⁶⁸. Ils ont notamment démontré que la réaction était catalysée par des températures élevées ou par irradiation UV à température ambiante. Un an plus tard, le groupe de Walling a décrit l'utilisation d'un initiateur de radicaux, l'AIBN (azobisisobutyronitrile), pour accélérer cette réaction de déchalcoxylation²⁶⁹. Un schéma mécanistique a été proposé par lequel une combinaison entre des radicaux thiyles et l'atome de phosphore conduirait à la formation d'un radical phosphoryle. Ce radical phosphoryle se décompose par une coupure homolytique de la liaison C-S en radical alkyle et en sulfure de phosphine. Suite à cela, le radical d'alkyle capte un proton du solvant ou d'un résidu thiol présent dans le milieu réactionnel pour former un groupement alkyle. Donc ce mécanisme

s'applique particulièrement à la désulfuration de peptides ou protéines par la TCEP (X= Cys, Figure 42).

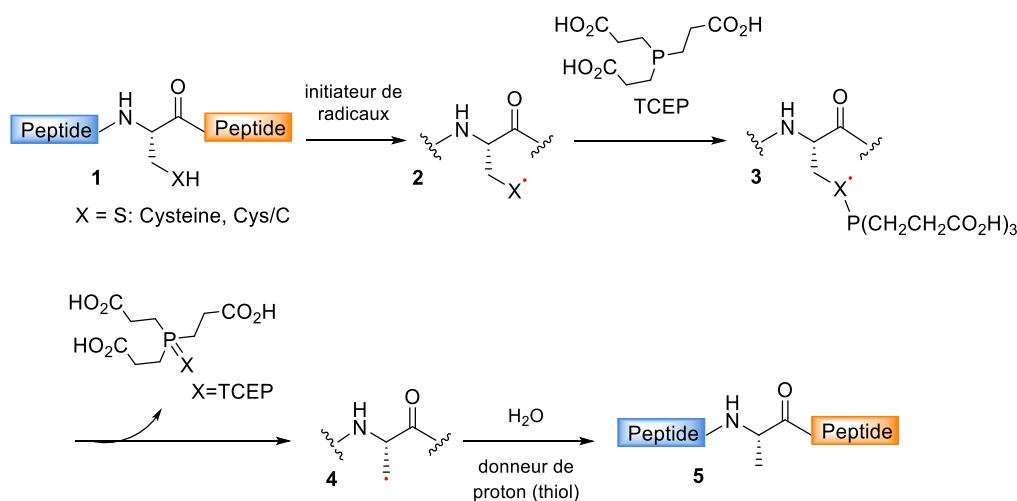


Figure 42. Schéma mécanistique de la déchalcogénation par voie radicalaire.

En 2007, le groupe de Danishefsky a décrit une méthode de désulfuration par voie radicalaire en utilisant un analogue hydrosoluble de l'AIBN, le VA-044²⁶⁵. Depuis, c'est devenu une méthode utilisée de manière courante et étendue à la déchalcogénation de nombreux dérivés d'acides aminés thiolés. Cette méthode permet d'envisager des ligations à tout type de jonction à condition de disposer des acides aminés chalcogénés adéquats. Par exemple, le groupe de Danishefsky a réalisé la synthèse totale d'une protéine ne contenant aucune cystéine, l'hormone parathyroïde humaine Hpth, en utilisant des dérivés de valine et de la leucine thiolés²⁷⁰.

Pour finir, la mise en place de différents procédés de désulfuration ont permis d'une part de surpasser la faible abondance du résidu cystéine dans les protéines naturelles mais également le positionnement inadéquat de celle-ci dans la séquence primaire des protéines cibles. C'est pourquoi l'utilisation de méthodes de ligation combinées à des processus de désulfuration est en constante évolution depuis les années 2010, comme en témoigne la Figure 43.

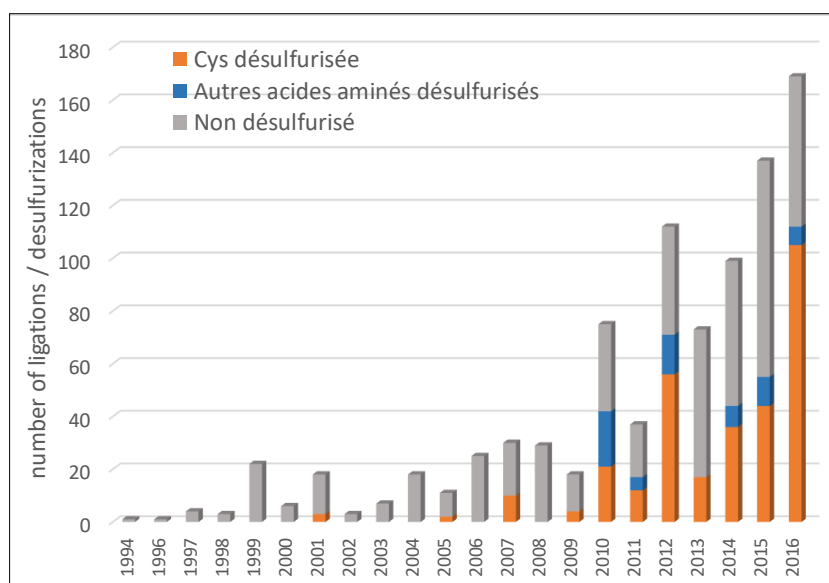


Figure 43. Histogramme du nombre d'acides aminés de type thiol (Cys ou dérivés thiols d'acides aminés) désulfurisés après ligation avec le nombre de ligations à la cystéine sans désulfuration.

C. Bilan/Conclusion

Les systèmes recombinants rencontrent de nombreuses difficultés à produire certaines protéines. Cela comprend particulièrement les protéines qui présentent des architectures inhabituelles comme les protéines cycliques ou ramifiées. Les protéines modifiées par SUMO possèdent une structure branchée et constituent un exemple typique de protéines difficiles à obtenir au moyen des techniques recombinantes. Par contre, ce type de protéine est accessible par voie chimique, une méthode qui s'est aujourd'hui clairement imposée comme une alternative intéressante aux méthodes recombinantes pour accéder à des protéines complexes. L'émergence de quelques ligations chimiques donne accès à la synthèse de polypeptides natifs. Notamment, la NCL et les méthodes apparentées Ces dernières permettent d'atteindre des objets moléculaires de taille significative (> 300 AA). J'ai dans mon travail de thèse développé des procédés performants basés sur cette chimie des thioesters pour accéder aux protéines SUMO-2 et SUMO-3 ainsi qu'à des dimères SUMO-2/3 de 21 kDa.

RÉSULTATS & DISCUSSIONS

I. Synthèse par voie chimique des protéines SUMO-2 et SUMO-3

La synthèse chimique a été utilisée de façon intensive pour produire l'ubiquitine (Ub), des conjugués ubiquitine, mais également des chaînes polyubiquitine,^{267, 271, 272} donnant ainsi de nouveaux outils pour étudier cette MPT. Comparativement, les publications décrivant la synthèse de conjugués SUMO sont plus rares. Cette situation est due en partie à la différence de longueur entre les domaines Ub (76 acides aminés) et SUMO (92-96 acides aminés). En effet, le domaine SUMO dépasse les limites de la synthèse peptidique en phase solide (SPPS). Une autre différence majeure entre les protéines Ub et SUMO réside également dans la présence d'un résidu cystéine (Cys) interne et central au cœur de toutes les isoformes de SUMO (Figure 5, * Cys interne), rendant les stratégies de synthèse plus complexes.

De manière générale, les conjugués Ub peuvent-être synthétisés en faisant réagir un domaine Ub équipé d'une fonction thioester à l'extrémité C-terminale avec un peptide cible contenant un résidu γ ou δ -mercaptolysine selon les conditions de la NCL (Figure 44). Toutefois, l'ajout d'un résidu Cys sur la chaîne latérale de la lysine cible constitue également une alternative intéressante à l'utilisation des mercaptolysines, qui ont un processus de synthèse relativement onéreux comme discuté auparavant. Par la suite, le groupe thiol de la mercaptolysine ou du motif Lys(Cys) doit être éliminé par désulfuration. Etant donné que l'Ub contient un résidu alanine central dans sa séquence (Ala46), l'analogue Ub Ala46Cys peut être assemblé facilement à partir de deux segments peptidiques par NCL et ensuite converti en un conjugué Ub natif par désulfuration globale (conversion Cys→Ala). Pour les domaines SUMO, le résidu Cys est idéalement situé en milieu de séquence facilitant son assemblage par NCL. Or si un domaine SUMO natif doit être obtenu, l'étape de désulfuration nécessitera des adaptations des procédés de synthèse. En effet, les conditions de désulfuration de l'acide aminé thiol utilisé pour construire la liaison isopeptidique, c'est-à-dire la mercaptolysine ou le motif Lys(Cys), peuvent potentiellement entraîner la désulfuration de la cystéine interne du domaine SUMO.

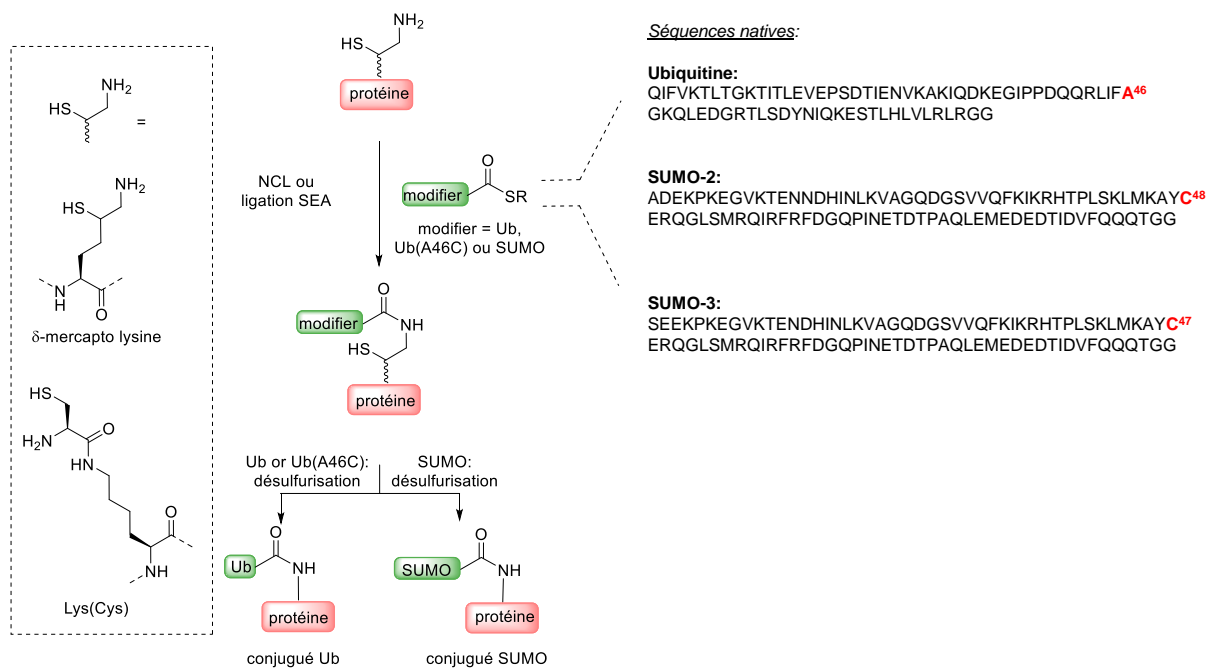


Figure 44. Synthèse chimique de conjugués Ub et/ou SUMO en combinant la NCL et la désulfuration. Modifier correspond à l'Ub ou la protéine SUMO.

Dans ce contexte, il est essentiel de connaître la contribution de la cystéine interne sur la structure et la fonctionnalité du domaine SUMO. Cette connaissance peut significativement influencer la conception de la stratégie de synthèse de conjugués SUMO. En effet, si cette cystéine peut être remplacée par une alanine, une étape de désulfuration globale pourra être appliquée aux conjugués, simplifiant considérablement l'accès à ces molécules. Dans le cas contraire, une désulfuration sélective sera nécessaire. Nous avons ainsi voulu investiguer l'importance de cette cystéine interne sur la structure et les propriétés des domaines SUMO-2 et SUMO-3, sujets de l'étude.

A cette fin, il a été nécessaire de réaliser la synthèse des monomères SUMO-2 et -3 ainsi que leurs variants desulfurés pour déterminer la contribution de la Cys interne sur la structure secondaire, la stabilité thermique et la capacité de ces derniers à être conjugué sur un substrat cible. De la même manière, la synthèse de conjugués SUMO-2 et -3 et de leur analogues desulfurés seront utiles pour investiguer l'impact de la modification Cys → Ala sur la cinétique de déconjugaison, qui dépend de la reconnaissance des domaines SUMO par les SENPs. Au final, les informations acquises grâce à ces expériences permettront de mettre en place la meilleure stratégie de synthèse pour accéder à des conjugués SUMO-2/3 natifs.

A. Formation d'aspartimide

1. Formation d'aspartimide pendant la synthèse peptidique en phase solide

La chimie des peptides a fait de grands progrès au cours des dernières décennies, mais la formation d'aspartimide (Asu) lors de la synthèse de peptides reste une difficulté fréquente. Jusqu'à présent, aucune méthode simple et générale n'a été développée pour supprimer cette réaction secondaire lors de la SPPS. Cette réaction secondaire peut conduire à de faibles rendements et rend certaines séquences peptidiques inaccessibles, gênant à la fois la recherche et les applications industrielles. D'un point de vue mécanistique, c'est le résultat de la fermeture d'un cycle α -amino-succinimide entre l'azote de la liaison amide et la chaîne latérale de l'acide aspartique, entraînant ainsi la perte de son groupe protecteur ester (OtBu)^{273, 274}. C'est une problématique majeure dans la stratégie Fmoc SPPS car la cyclisation est favorisée par la présence de base telle que la pipéridine utilisée pour éliminer le groupement Fmoc^{275, 276}. D'autre part, sa sensibilité aux bases peut mener à de l'épimérisation²⁷⁷, qui combiné à l'ouverture du cycle imide par des nucléophiles conduit à la formation de divers sous-produits (Figure 45). Une attaque par une molécule d'eau donne le peptide β -aspartyle et le peptide α -aspartyle dans un rapport de 3:1²⁷⁸. L'ouverture du cycle par la pipéridine donne un mélange d' α et de β -pipéridides. Bien que les aspartimides et les α/β -pipéridides générés par cette réaction secondaire puissent être facilement séparés du peptide cible par HPLC, les peptides β -aspartyle et le peptide α -aspartyle épimérisés sont quant à eux difficiles à éliminer car ils ont souvent le même temps de rétention que le peptide cible. De plus, ce sont des produits isobares, par conséquent la présence de ces deux produits est difficile à détecter.

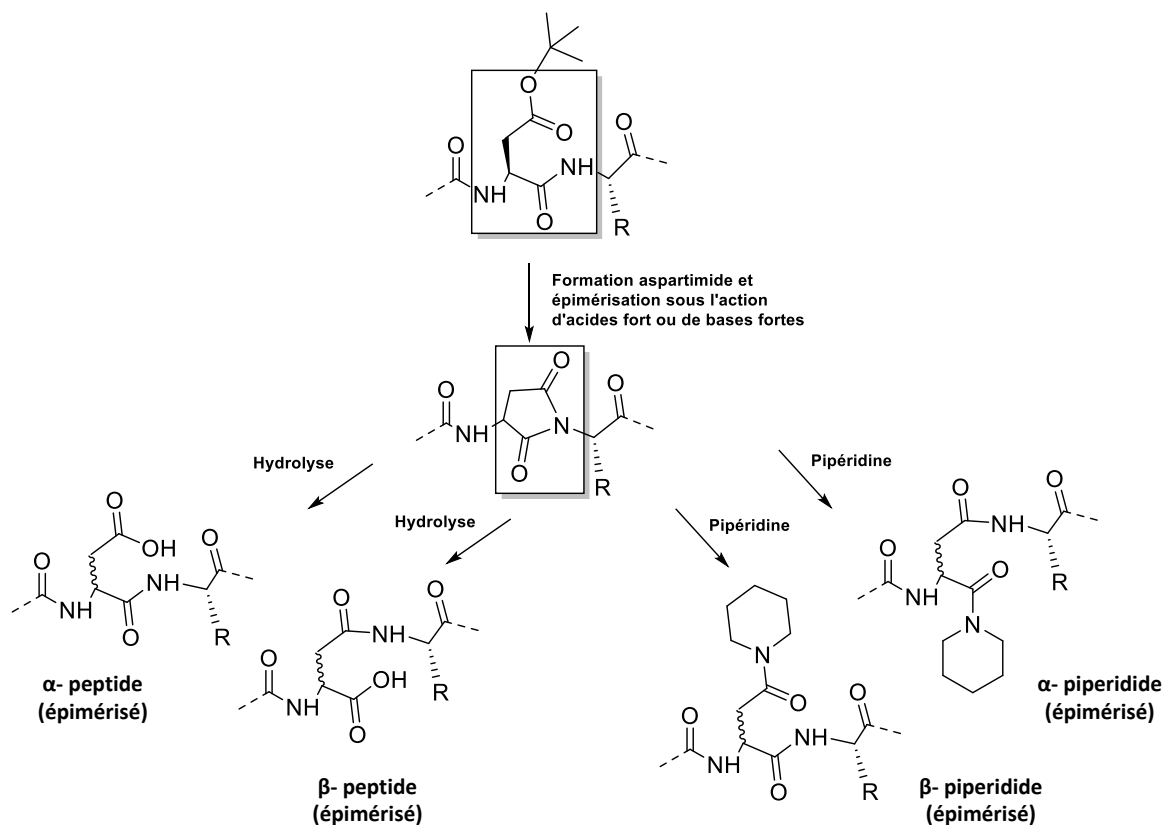


Figure 45. Réactions secondaires résultant de la formation d'aspartimide.

Un paramètre important déterminant la facilité de formation d'aspartimide pendant la SPSS est la nature de l'acide aminé suivant le résidu d'acide aspartique (Tableau 6). La conformation et donc la séquence du peptide joue également un rôle. D'autre part, la réaction d'aspartimide est catalysée par les bases, les acides et est favorisée par une augmentation de la température.

Nature de l'acide aminé en position C-terminale	Tendance à la formation d'aspartimide
Ala	+
Arg(Pbf)	++
Asp(OtBu)	+++
Asn(Trt)	+++
Cys(Acm)	++
Cys(Trt)	+
Gly	++++

Tableau 6. Les séquences Asp-Xaa sujettes à la formation d'aspartimide sous traitement basique. Faible +, élevée ++++^{273, 275, 276}.

Les peptides contenant le motif Asp(D)-Gly(G) dans leur séquence sont particulièrement sensibles à la formation d'aspartimide, dont on estime la quantité formée lors de la SPPS à environ 0.5 % par cycle de déprotection lors de la SPPS²⁷⁹. De toute évidence, le problème est plus sévère dans les segments peptidiques contenant plus d'un site sensible à la formation d'aspartimide, ainsi que les longs segments qui inévitablement sont exposés plus longtemps à pipéridine. Néanmoins, l'ajout d'acide organique à la solution de pipéridine utilisée pour ôter le groupement Fmoc s'est avéré être efficace pour minimiser la formation d'aspartimide. Les réactifs les plus utilisés sont l'hydroxybenzotriazole (HOBt) ou l'éthyl cyano(hydroxyimino)acétate (Oxyma pure), utilisés habituellement à 0,1 M dans un mélange 80 :20 de diméthylformamide et de pipéridine. Toutefois, il est généralement admis que dans les peptides contenant le motif DG l'ajout d'acide ne réduit pas suffisamment la formation d'aspartimide. De même, diminuer la basicité de la solution de pipéridine c'est prendre le risque d'avoir des déprotections incomplètes. Par conséquent, des alternatives ont été développées notamment des dipeptides protégés Asp-Xaa qui peuvent être incorporés pendant la SPPS comme des acides aminés classiques. Dans ces systèmes, la nucléophilie de l'azote de la liaison amide est bloquée à l'aide d'un groupement protecteur de type benzyle.

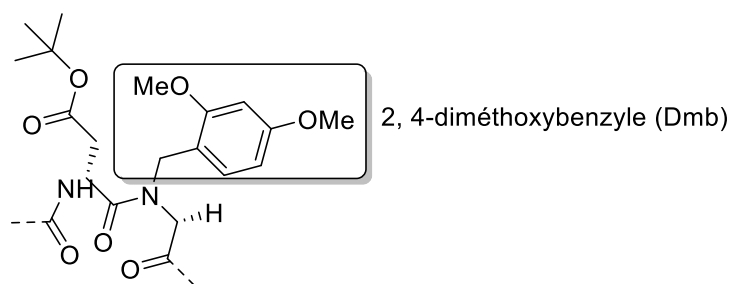


Figure 46. Dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH commercial utilisé dans mon travail de thèse.

En ce qui concerne, la synthèse par voie chimique des protéines SUMO, nous avons été confrontés à cette problématique de formation d'aspartimide pendant la SPPS des segments peptidiques. En effet, deux motifs Asp-Gly sont présents dans chacun des segments peptidiques nécessaires à l'assemblage des protéines SUMO-2 et de SUMO-3 (Figure 47). C'est pourquoi, durant ma thèse j'ai utilisé le dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH où l'azote α du résidu glycine est protégé par un groupement protecteur 2, 4-diméthoxybenzyle (Dmb). Lors de la SPPS, le dipeptide Fmoc-Asp(OtBu)-(Dmb)Gly-OH a été couplé manuellement.

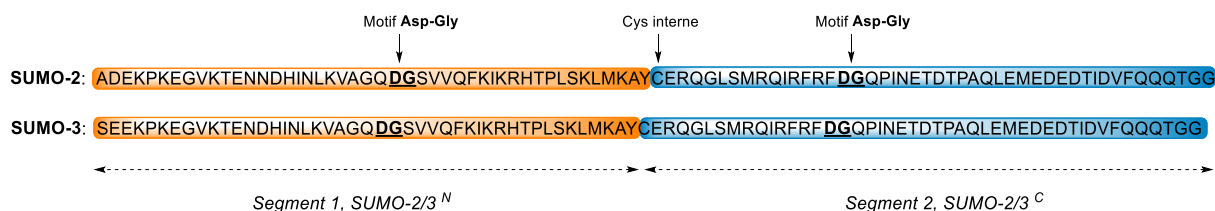


Figure 47. Séquence des segments peptidiques des domaines SUMO-2 et -3. Les motifs Asp-Gly, sensibles à la formation d'aspartimide, sont soulignés.

2. Formation d'aspartimide pendant l'étape de ligation chimique

La formation d'aspartimide à partir des résidus acides aspartique est un sujet de préoccupation dans de nombreux domaines. Au-delà des modifications post-traductionnelles non enzymatiques des protéines, les modifications des résidus conventionnels de L- α -Asp en résidus L- β -Asp, D-Asp et D- β -Asp ont récemment attiré l'attention en raison de leur pertinence dans le vieillissement cellulaire, et notamment dans certaines pathologies telles que la cataracte ou encore la maladie d'Alzheimer. La formation d'Asu a été considérée comme une substitution nucléophile au niveau du carbone carbonyle comprenant deux étapes, un mécanisme de cyclisation et de déshydratation (Figure 48). D'après les calculs de chimie quantique effectués sur la base de la densité électronique des éléments (DFT pour Density Functional Theory), il semblerait que les deux étapes nécessitent une catalyse car les

processus non catalysés ont des barrières énergétiques élevées incompatibles avec les vitesses de réactions observées en pratique^{280, 281}. L'eau ou les tampons aqueux peuvent potentiellement jouer le rôle de catalyseur^{282, 283}.

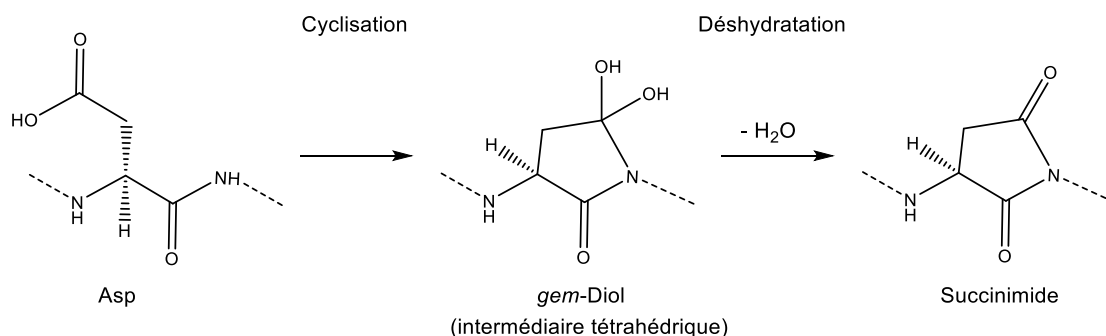


Figure 48. Mécanisme en deux étapes conduisant à la formation du succinimide à partir de l'acide aspartique (Asp)

Des expériences réalisées sur le lysozyme ont permis de clarifier cet aspect. L'incubation du lysozyme à pH 4 (tampon acétate, 40 °C) conduit à la formation d'Asu au niveau du résidu Asp101²⁸⁴. La quantité d'Asu formée augmente avec la concentration du tampon et la température, en accord avec une catalyse de la formation d'aspartimide par le tampon. D'autre part, il a été constaté que l'intermédiaire succinimide est plus stable (ou que sa formation est plus rapide) à pH légèrement acide qu'à pH élevé^{285, 286}. Il ressort de ces différentes expériences que les acides carboxyliques catalysent la formation d'aspartimide. Très récemment, la désamidation de résidus d'asparagine (Asn) (qui passe également par l'intermédiaire succinimide) s'est révélée être elle aussi catalysée par divers acides carboxyliques²⁸⁷.

3. Synthèse des domaines SUMO-2, SUMO-3

La formation d'Asu pendant les étapes de ligation est un aspect qui reste relativement peu discuté dans la littérature malgré la fréquence des motifs Asp-Gly dans les protéines. Cette problématique n'a pas été rencontrée au laboratoire durant la synthèse du domaine SUMO-1 ainsi que du conjugué p53-SUMO-1 car le domaine SUMO-1 ne contient pas de motif Asp-Gly. Comme explicité précédemment, la situation diffère pour SUMO-2 et -3 avec la présence de deux motifs Asp-Gly de part et d'autre du site de ligation.

La ligation SEA nécessite la présence de divers réactifs porteurs d'une ou plusieurs fonctions acides carboxylique tels que le MPAA et la TCEP, qui peuvent potentiellement promouvoir la formation d'aspartimide. Dans ces conditions, il nous semblait logique de

regarder l'impact des conditions réactionnelles de l'étape de ligation sur la formation d'aspartimide. La synthèse de SUMO-2 par ligation SEA à 37 °C (pH 5,5, 200 mM TCEP, 200 mM MPAA) s'accompagne effectivement de la formation d'aspartimide (environ 30%). Nous avons pu constater qu'une réduction de la température entraînait une diminution significative de la formation d'aspartimide (environ 10%), avec certes l'inconvénient d'induire également une baisse de la cinétique de réaction, qui reste toutefois utile d'un point de vue synthétique. Ainsi, l'assemblage des différentes protéines SUMO (SUMO-2, SUMO-3) a été effectué selon les conditions classiques de la ligation SEA mais à 25 °C, sans que l'on ait eu besoin d'adapter la concentration du MPAA et de la TCEP. Pour conclure, en plus des précautions prises durant la synthèse des segments par SPPS, les conditions réactionnelles menant à la synthèse des domaines SUMO-2 et -3 ont dues être également optimisées.

Les nouvelles conditions de ligation SEA ont permis la synthèse totale des protéines SUMO-3 et SUMO-2 et du domaine core comme indiqué dans la Figure 49.

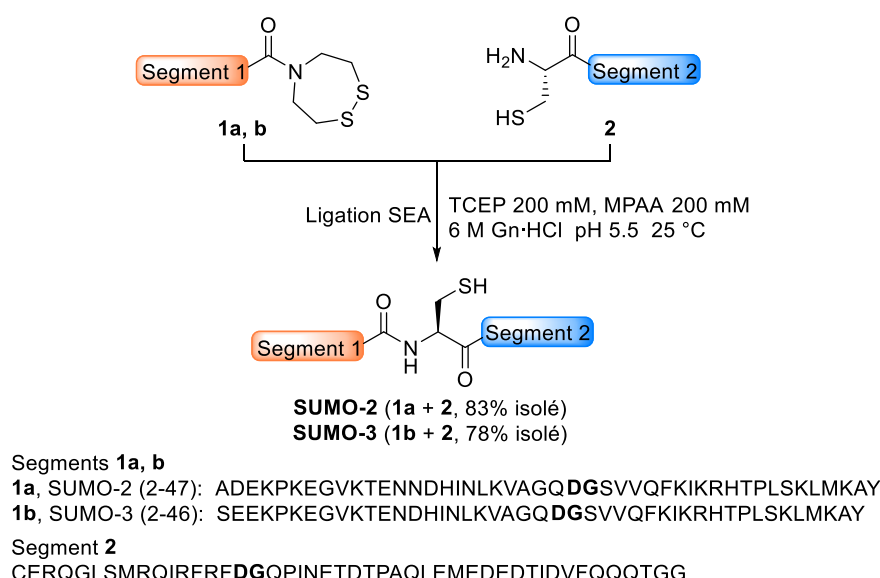


Figure 49. Synthèse en solution des protéines SUMO-2, -3 et du domaine core SUMO-2/3.

A noter que la ligation s'effectue en présence de chlorure de guanidinium (Gdn.HCl)²⁸⁸, utilisé pour favoriser la solubilisation des segments peptidiques et empêcher toute structuration qui pourrait masquer les groupements fonctionnels. En utilisant ces conditions, les protéines SUMO-2 et SUMO-3 ont été isolés avec un bon rendement (60-83% en produit isolé) et une bonne pureté comme en témoignent les analyses UPLC-MS montrées à la Figure 50.

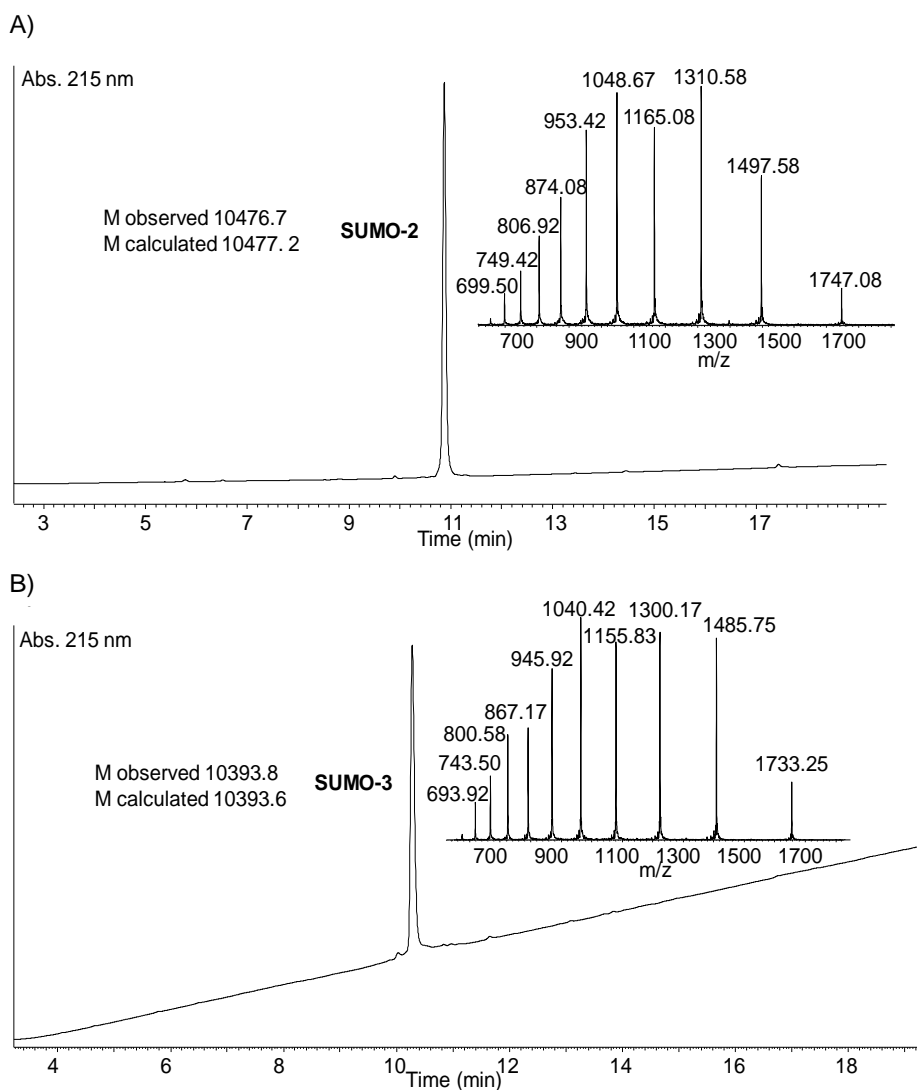


Figure 50. Analyse des protéines synthétiques SUMO-2, SUMO-3. A) UPLC-MS du domaine SUMO-2 synthétique ; B) UPLC-MS du domaine SUMO-3 synthétique.

Nous avons pu isoler les protéines SUMO-2 et -3 à l'échelle de plusieurs mg et poursuivre nos études.

4. Synthèses des domaines SUMO-2, SUMO-3 désulfurisés

Les protéines SUMO-2, SUMO-3 et SUMO-2/3 core ont été désulfurisées avec succès comme décrit à la Figure 51. De la méthionine a été ajoutée au milieu de désulfuration pour minimiser l'oxydation des méthionines. Nous avons procédé à leurs caractérisations, notamment par spectrométrie de masse (voir annexe Figure S 12, Figure S 13, Figure S 14), et par SDS-PAGE, dont les résultats sont montrés à la Figure 52. Ces données démontrent la qualité des produits obtenus permettant de concevoir d'autres analyses afin de répondre à l'objectif.

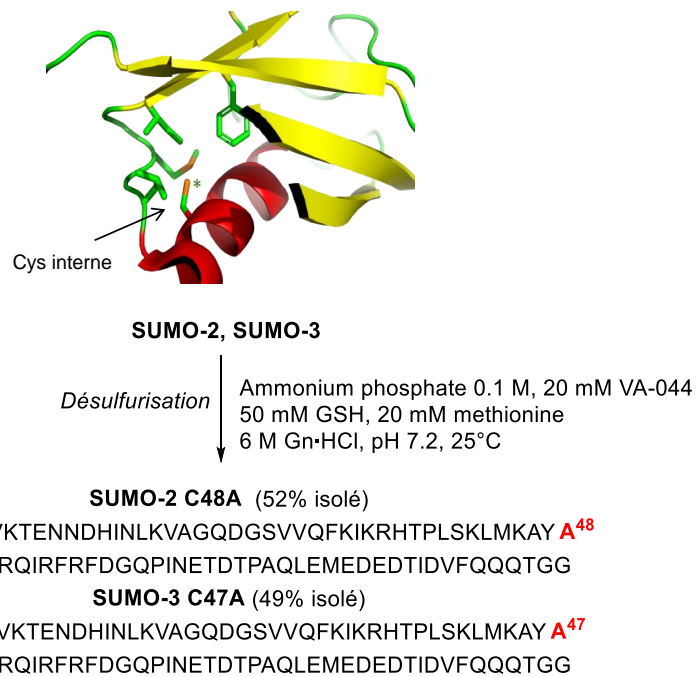


Figure 51. Préparation des analogues SUMO-2 C48A et SUMO3 C47A

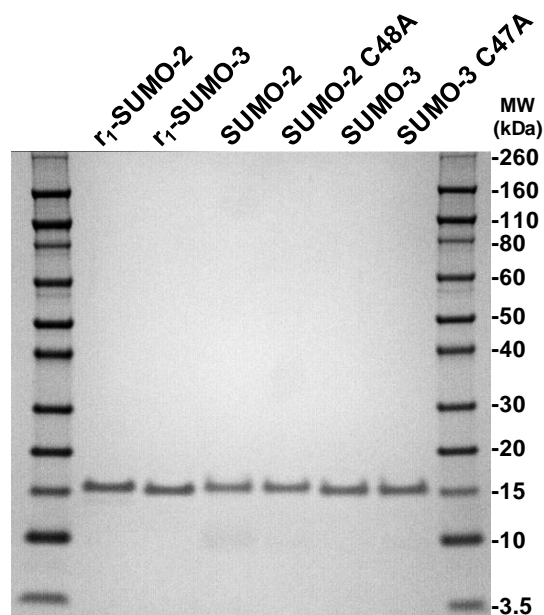


Figure 52. SDS-PAGE des protéines SUMO-2, SUMO-3 et de leur variant désulfurisés (SUMO-2 C48A, SUMO-3 C47A) et des protéines SUMO-2,3 recombinantes (*r1*-SUMO-2 et *r1*-SUMO-3). Le gel a été chargé avec 1 µg de chaque protéine et coloré au bleu de Coomassie R520 (MW= marqueur de poids moléculaire).

L'ensemble de ces produits a permis d'évaluer l'impact de la désulfuration de la cystéine centrale sur leurs structures secondaires et leurs propriétés biochimiques. Ces interrogations font notamment suite aux travaux réalisés sur le domaine SUMO-1, pour lequel l'équipe a pu démontrer le rôle important de la cystéine centrale sur la stabilité thermique du domaine SUMO-1 ainsi que sur ses propriétés biochimiques²⁵⁵.

B. Analyses biophysiques et biochimiques des protéines SUMO-2, SUMO-3

Pour l'étude par dichroïsme circulaire, nous avons rencontré des difficultés lors de la détermination de la concentration des différentes protéines. En effet, les protéines SUMO-2 et -3 ne possèdent qu'un seul résidu tyrosine pouvant absorber à 280 nm, par conséquent les techniques d'absorption en UV pour déterminer la concentration des protéines sont à priori peu adaptées lorsque les mesures sont faites avec un spectrophotomètre classique. De ce fait, nous avons tenté d'estimer la concentration à l'aide du dosage BCA (acide bicinchonique). Cependant, nous avons observé un facteur 2 entre la concentration théorique calculée à partir de la masse pesée et celle obtenue par cette technique de dosage. Cette différence importante peut être due à la composition en acide aminé des protéines mesurées, puisque la présence de thiols libres affins pour les ions Cu(II) peut fausser la mesure mais également à la quantité d'eau absorbée par la poudre lyophilisée lors de la pesée. Les mesures UV à 280 nm effectuées avec un nanodrop ont été plus concluantes en étant proches des concentrations estimées par la pesée, tout en ayant l'avantage d'utiliser peu de produit (3 µL suffisent). Ainsi, la concentration des différents variant des domaines SUMO a été mesurée au nanodrop.

Nous avons analysé les différentes protéines SUMO par CD afin de déterminer l'impact de la modification Cys→Ala sur la structure secondaire et la stabilité thermique des domaines SUMO-2 et SUMO-3.

1. Analyses en dichroïsme circulaire des domaines SUMO-2 et SUMO-3

Les analyses CD des protéines synthétiques comparées aux protéines recombinantes montrent un chevauchement des spectres prouvant la bonne conformation des domaines SUMO obtenu par synthèse chimique (Figure 53). De même, les spectres CD des domaines SUMO-2 et -3 synthétiques sont identiques, avec un taux d'alpha hélicité d'environ 15 % similaire à celui trouvé pour les protéines recombinantes²⁸⁹. Ces résultats sont en accord avec le taux d'alpha hélicité décrit pour 20 domaines SUMO-2/3 de la *Protein Data Base* (pdb) (voir annexe Table 1).

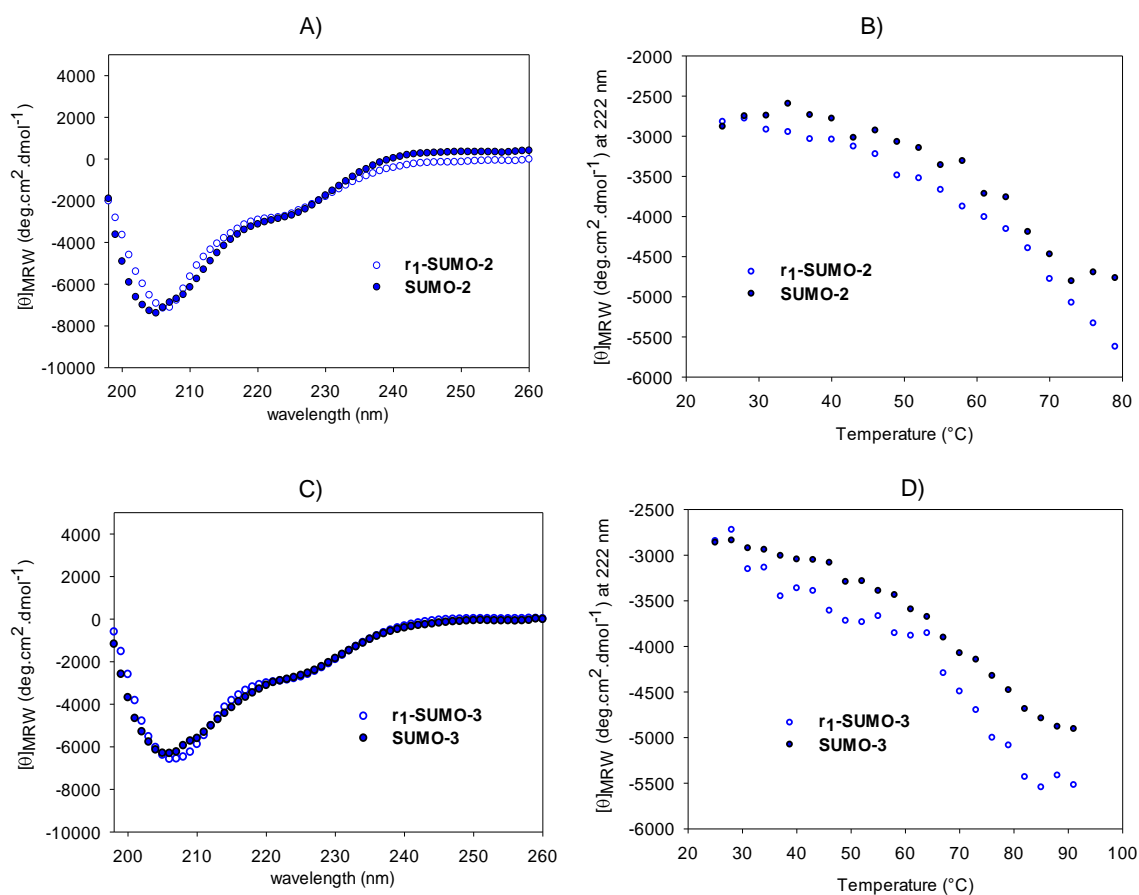


Figure 53. Analyses CD des protéines SUMO-2 et -3 synthétiques et recombinantes (r1-SUMO-2, r1-SUMO-3). (Tampon phosphate 10 mM pH 7.2). A) Analyse CD à 25 °C des protéines SUMO-2 synthétiques et recombinantes. B) Dénaturation thermique des protéines SUMO-2 synthétiques et recombinantes. C) Analyse CD à 25 °C des protéines SUMO-3 synthétiques et recombinantes. D) Dénaturation thermique des protéines SUMO-3 synthétiques et recombinantes. Les échantillons ont été soigneusement dégazés pendant 30 min avec de l'argon et immédiatement transférés sous argon dans la cuve à CD.

Nous avons ensuite comparé les protéines SUMO-2, SUMO-3 natives et les analogues désulfurisés pour mesurer l'impact de la modification Cys→Ala sur la structure secondaire des domaines SUMO (Figure 54). Les données montrent que la modification Cys→Ala induit de légers changements dans le spectre CD du domaine SUMO-2 avec une diminution du taux d'alpha hélicité d'environ 1%, $15.2 \pm 0.12\%$ ($n = 10$) pour SUMO-2 contre $14.1 \pm 0.23\%$ ($n = 12$) pour SUMO-2 C48A (Figure 54A). Cette baisse d'alpha hélicité est faible mais significative. A l'inverse pour le domaine SUMO-3, l'impact de la modification Cys→Ala est plus prononcé avec une diminution du taux d'alpha hélicité d'environ 2%, $15.1 \pm 0.21\%$ ($n = 8$) pour SUMO-3 vs $13.3 \pm 0.29\%$ ($n = 10$) pour SUMO-3 C47A (Figure 54B). Nous pouvons conclure que les protéines SUMO-2 et -3 ont la même structure secondaire mais répondent différemment à la modification Cys→Ala. Pourtant, les protéines SUMO-2 et 3 diffèrent seulement de quelques acides aminés dans leur partie N-terminale (Tail).

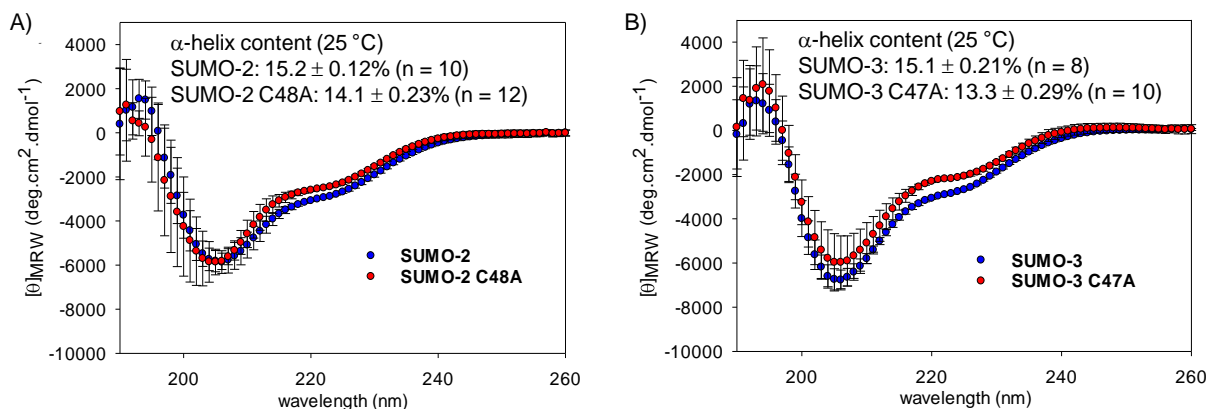


Figure 54. Analyses CD des protéines synthétiques SUMO-2, SUMO-2 C48A, SUMO-3 et SUMO-3 C47A. A) Analyses des protéines SUMO-2 ($n = 10$) et SUMO-2 C48A ($n = 12$) à 25 °C. B) Analyses des protéines SUMO-3 ($n = 8$) et SUMO-3 C47A ($n = 10$) à 25 °C. Conditions : 10 mM tampon phosphate pH 7.2. Les échantillons sont dégazés au préalable pendant 30 min à l'argon puis transférés immédiatement dans la cuve CD sous argon. Concentrations pour les analyses à 25 °C : SUMO-2 13.8-25.3 μM , SUMO-2 C48A 9.1-31.6 μM , SUMO-3 14.4-24.0 μM , SUMO-3 C47A 25.1-34.7 μM . Les données correspondent aux moyennes et aux écart-types.

Dans un second temps, nous avons regardé l'impact de la modification Cys \rightarrow Ala sur la stabilité thermique des protéines SUMO-2 et -3. En ce qui concerne les dénaturations thermiques, nous avons regardé l'évolution de l'ellipticité molaire par résidu à 222 nm ($[\theta]_{\text{MRW}} 222 \text{ nm}$), longueur d'onde représentative des hélices α . Les données montrent que les courbes de dénaturation des différents domaines SUMO-3 ne se chevauchent pas contrairement aux domaines SUMO-2, ce qui veut dire que la modification Cys \rightarrow Ala a un impact plus significatif sur la stabilité thermique du domaine SUMO-3 plutôt que sur celle de SUMO-2 (Figure 55).

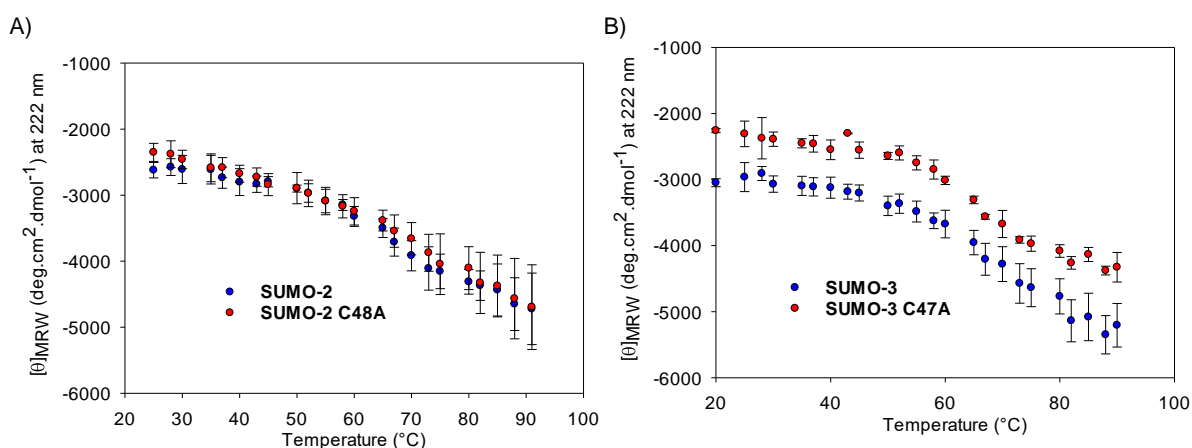


Figure 55. Analyses CD des protéines synthétiques SUMO-2, SUMO-2 C48A, SUMO-3 et SUMO-3 C47A. A) Dénaturation thermique des protéines SUMO-2 ($n = 5$) et SUMO-2 C48A ($n=7$) ($[\theta]_{MRW}$ à 222 nm). B) Dénaturation thermique des protéines SUMO-3 ($n = 9$) et SUMO-3 C47A ($n = 5$) ($[\theta]_{MRW}$ at 222 nm). Conditions : 10 mM tampon phosphate pH 7.2. Les échantillons sont dégazés au préalable pendant 30 min à l'argon puis transférés immédiatement dans la cuve CD sous argon. Concentrations des protéines pour les dénaturations thermiques : SUMO-2 12.9-28.6 μM , SUMO-2 C48A 12.0-32.5 μM , SUMO-3 12.5-40.4 μM , SUMO-3 C47A 13.5-34.7 μM

D'un point de vue structural, la Cys interne (Cys48 pour SUMO-2, Cys47 pour SUMO-3) est logée au sein d'une hélice α appartenant au domaine core (domaine « Ubiquitin-like »). La chaîne latérale de cette Cys pointe vers l'intérieur du domaine structuré dont certains résidus hydrophobes entourent sa chaîne latérale (Figure 56). Récemment, l'importance de la phénylalanine 60 (Phe60) dans le repliement de SUMO-2 a été discuté par Bhattacharya et Ainaravaru²⁸⁹. La Phénylalanine en position 60, Ph60, interagit avec de nombreux résidus du domaine core, et est localisée à moins de 5 Å de la Cys 48 de SUMO-2. Elle agirait comme un « Hub »²⁹⁰ en établissant de multiples interactions au sein de la structure tertiaire de SUMO-2. Logiquement, on peut s'attendre à ce qu'une modification du noyau hydrophobe de SUMO-2 et -3 par une étape de désulfuration du résidu Cys interne vienne altérer l'environnement de Phe60 et par là leur structure et leur stabilité.

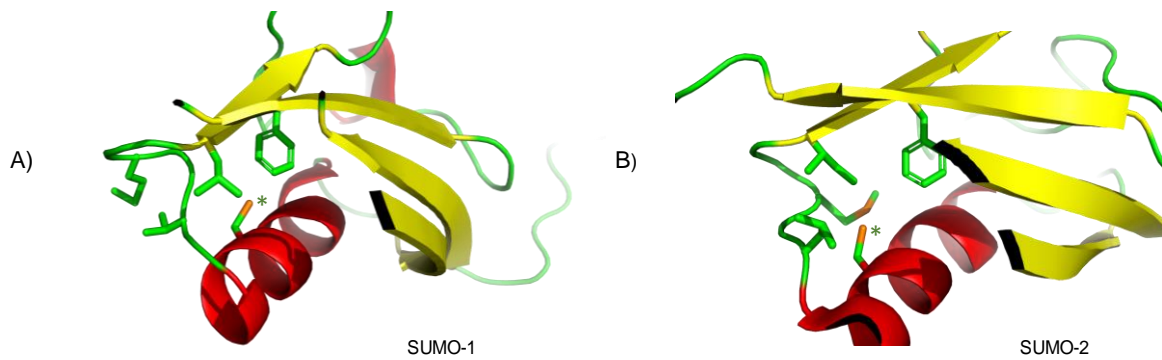


Figure 56. A) Structure RMN de SUMO-1, (code pdb : 1A5R) (Bayer et al. 1998)³⁷. B) Structure RMN SUMO-2, code pdb 2N1W. Pour A) et B) Certaines chaînes latérales de résidus hydrophobes entourant la Cys interne sont mises en évidence (Tyr47, Phe32, Phe60, Phe62).

Puisque la désulfuration de la Cys interne entraîne des perturbations dans la structure secondaire et réduit la stabilité thermique des domaines SUMO-2 et -3, il est pertinent de s'interroger sur l'influence potentielle de cette modification sur la capacité des domaines SUMO-2 et -3 à être pris en charge par la machinerie enzymatique de la SUMOylation pour être conjugués à une protéine cible.

2. Analyses biochimiques : La SUMOylation

Pour la SUMOylation *in vitro*, les protéines SUMO-2, SUMO-2 C48A, SUMO-3 et SUMO-3 C47A ont été incubées avec les différents réactifs d'un kit commercial de SUMOylation : ATP/Mg⁺, UbC9, E1, RanGAP (protéine cible), tampon. Les réactions ont été analysées par SDS-PAGE à l'aide d'un anticorps polyclonal anti-SUMO 2/3. Les différents analogues, c'est-à-dire les protéines recombinantes du kit commercial ainsi que les différents variant des protéines de synthèse, ont été efficacement transférés sur la protéine cible RanGAP1, sans que l'on puisse mettre en évidence une différence significative entre les protéines SUMO étudiées (Figure 57).

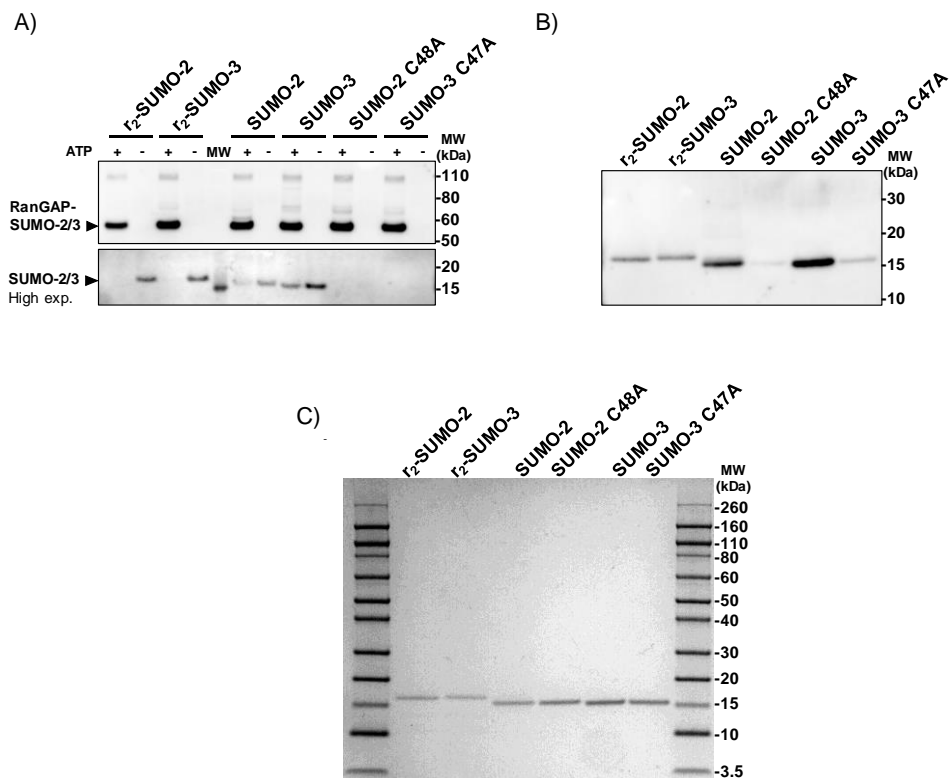


Figure 57. A) Essai de conjugaison utilisant RanGAP comme protéine cible. Le test a été effectué selon le protocole du fabricant en utilisant 100 ng des différentes protéines SUMO, y compris les protéines SUMO-2 et 3 recombinantes du kit marqué par un His-tag (notées r2-SUMO-2 et r2-SUMO-3). La présence des protéines SUMO a été testée par Western-blot en utilisant un anticorps commercial anti-SUMO-2/3. B) Analyse du Western blot de r2-SUMO-2 et r2-SUMO-3, de SUMO-2 et 3 synthétiques wild-type et de SUMO-2 synthétisé désulfuré (3) C48A (C47A) (100 ng de chaque protéine) en utilisant de l'anticorps anti-SUMO -2/3. C) Coloration au Coomassie des différentes protéines (1 μ g chacune).

De manière inattendue, nous avons constaté que les protéines désulfurisées SUMO-2 C48A et SUMO-3 C47A étaient mal reconnues par l'anticorps polyclonal anti-SUMO-2/3 fourni par le kit commercial (Figure 57A), tandis que les conjugués RanGAP correspondants présentaient quant à eux une forte bande dans la même analyse. Cette observation a notamment été confirmée au moyen d'une autre analyse par western blot des différents analogues de SUMO-2 et -3 synthétiques, toujours en utilisant le même anticorps anti-SUMO-2/3 (Figure 57B). Cette différence de reconnaissance ne peut être due à un différentiel en termes de quantité de matériel protéique puisque le chargement du gel (quantité en masse de protéine) a été contrôlé par la réalisation d'une électrophorèse en utilisant la coloration au bleu de Coomassie, montrée à la Figure 57C.

Dans la mesure où l'essai de conjugaison décrit ci-dessus pouvait être biaisé par la reconnaissance différentielle des analogues SUMO par l'anticorps anti-SUMO-2/3, nous avons effectué des essais de bioconjugaison quantitatifs basés sur la technique du FRET en

collaboration avec le laboratoire du Dr. Frauke Melchior en Allemagne. Le principe de cette technique est notamment décrit à la Figure 58. En quelques mots, dès lors où le substrat cible RanGAPtail marqué par la protéine fluorescente cyan (CFP) est conjugué enzymatiquement à la protéine SUMO-3 marquée par la protéine fluorescente jaune (YFP), un signal est émis. Cette émission de signal est rendue possible car la protéine CFP absorbe la lumière à 430 nm et émet un signal de fluorescence à 485 nm, longueur d'onde d'excitation du chromophore YFP. De ce fait, lorsque les protéines fluorescentes CFP et YFP sont à proximité (c'est-à-dire lorsque RanGAPtail est SUMOylé), le transfert d'énergie de CFP à YFP donne lieu à un signal d'émission à 527 nm. En conséquence, le rapport d'intensité de l'émission à 527 nm et de l'émission à 485 nm est représentatif de la quantité de RanGAPtail SUMOylé. De cette façon le test permet de suivre la réaction de conjugaison des domaines SUMO en temps réel et en parallèle.

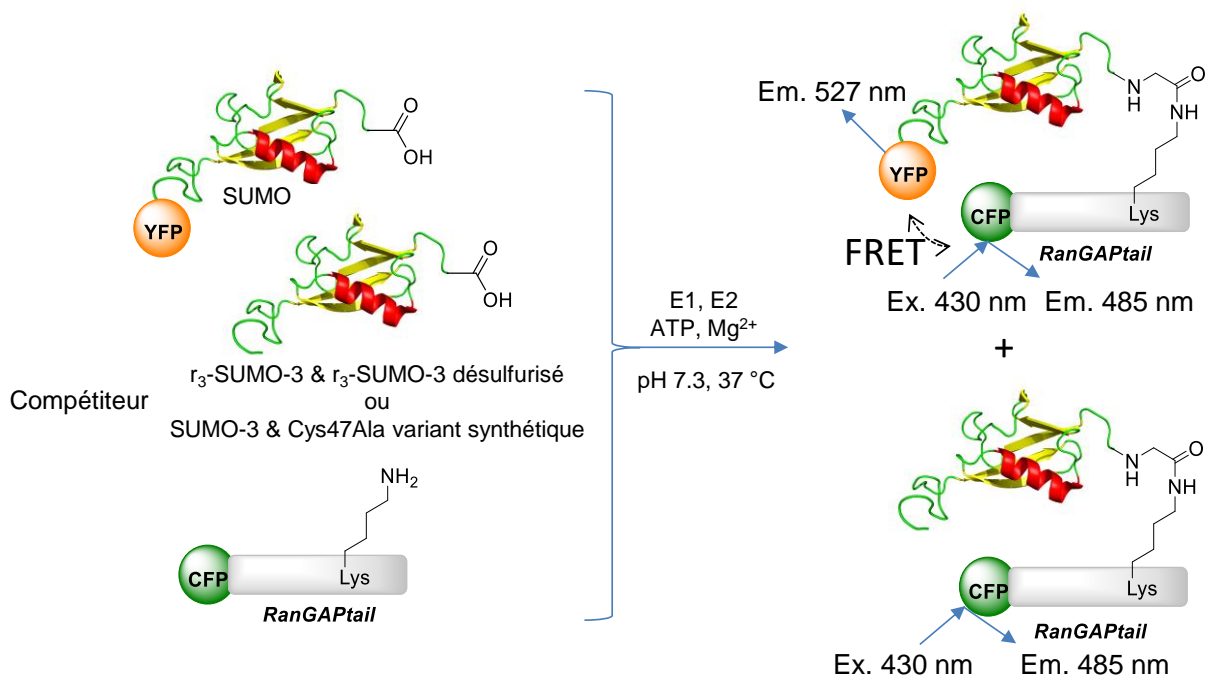
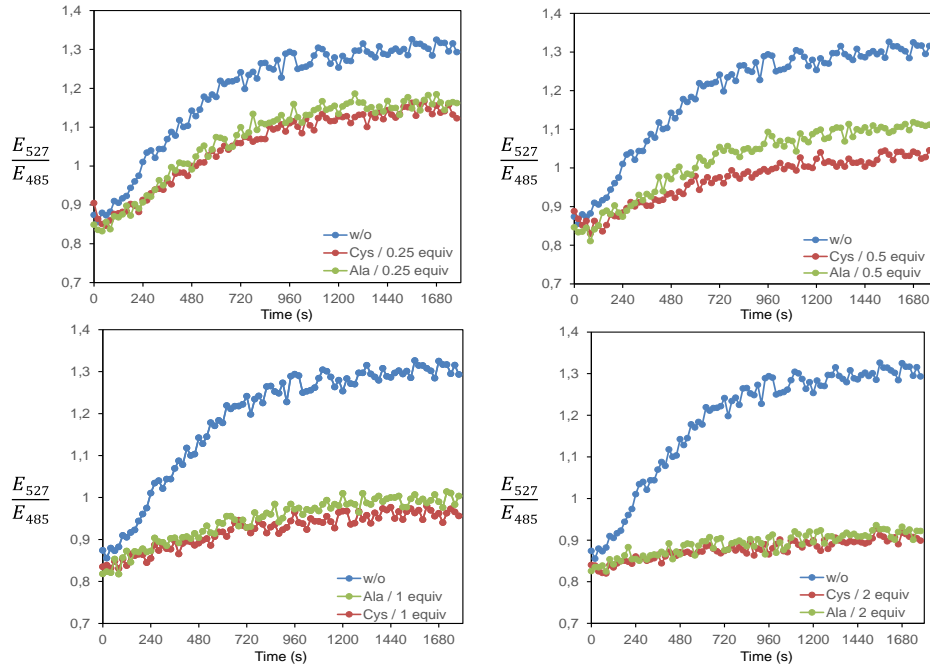


Figure 58. Principe du test de conjugaison par FRET en utilisant RanGAP tail comme substrat cible.

Avec cet essai, nous avons analysé une protéine SUMO-3 recombinante commerciale et son analogue désulfuré chimiquement avec les protéines synthétiques SUMO-3 et SUMO-3 C47A. Le but était d'utiliser les différents analogues comme concurrents du domaine SUMO-3 marqué YFP (YFP-SUMO-3). Les résultats présentés à la Figure 59 montrent que les protéines entrent en compétition avec le conjugué YFP-SUMO-3 de façon similaire. Les données

montrent clairement que la modification Cys→Ala n'a pas d'effet significatif sur la capacité des domaines SUMO-3 à se conjuguer à RanGAPtail, que ce soit avec la protéine recombinante ou synthétique, et ce à différents ratio SUMO-3/YFP-SUMO-3 (Figure 59).

A) Domaines SUMO-3 wild-type et désulfuré recombinants



B) Domaines SUMO-3 wild-type et désulfuré synthétiques

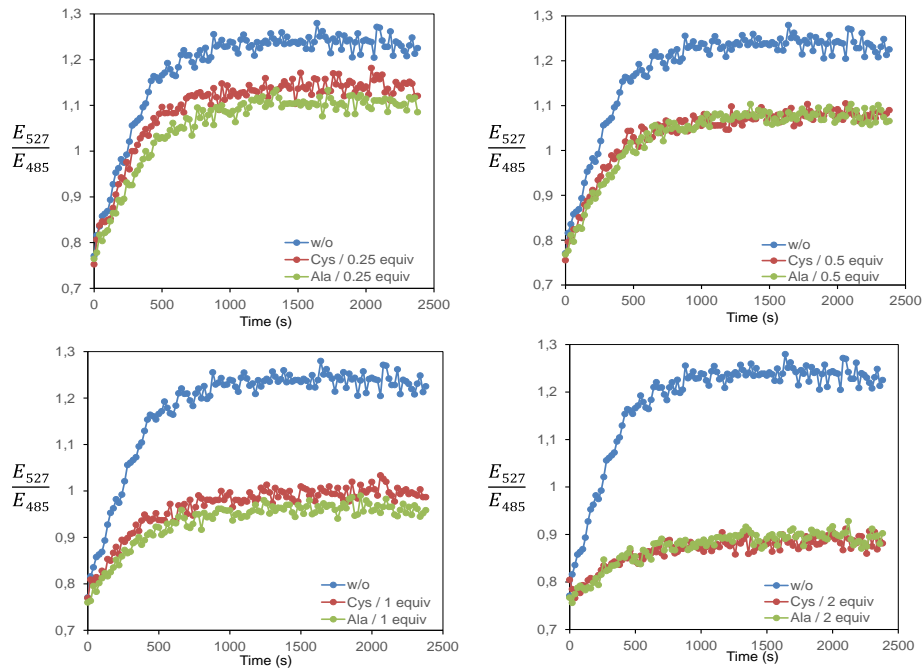


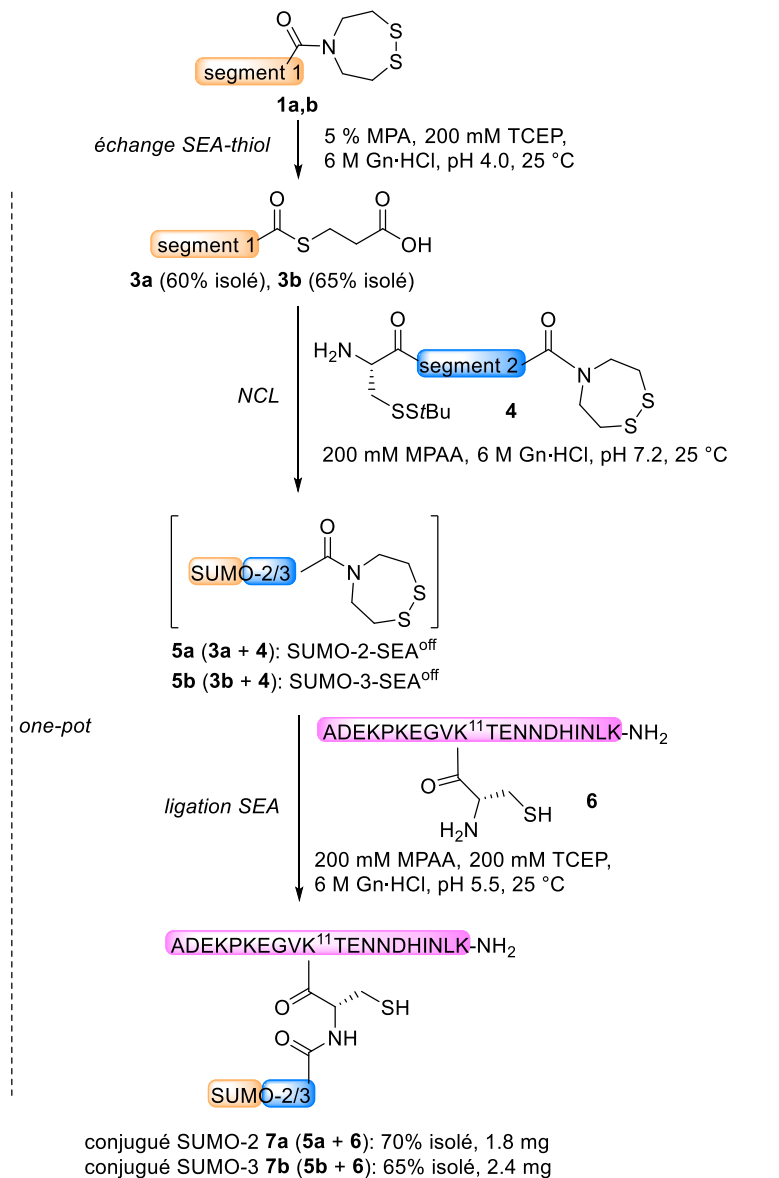
Figure 59. Conjugaison basée sur la technique du FRET en utilisant RanGAPtail comme substrat cible. A) Compétition avec la protéine SUMO-3 recombinante et son analogue désulfuré chimiquement. B) Compétition avec les protéines synthétiques SUMO-3 et SUMO-3 C47A. Conditions : 15 nM E1, 20 nM E2, 300 nM YFP-SUMO-3, 300 nM CFP-RanGAPtail and compétiteurs (75 nM-600 nM) ont été incubés à 37°C pour 10 min dans un milieu de conjugaison (110 mM KOAc, 20 mM HEPES pH 7.3, 2 mM Mg(OAc)₂, 1 mM EGTA, supplémenté avec 1 µg/mL de chaque, leupeptine, pepstatine et aprotinine, 1 mM DTT, 0.05% Tween®-20 and 0.2 mg/mL ovalbumine). La réaction de SUMOylation *in vitro* a commencé avec l'addition de 1 mM ATP. Tous les 20 s, le ratio E527/E485 a été déterminé après excitation à 430 nm.

Nous pouvons conclure que les résultats obtenus par la technique FRET sont en accord avec ceux du test de SUMOylation *in vitro*. Dans le but de compléter ces études, nous avons réalisé la synthèse totale de divers conjugués SUMO-2 et -3 pour étudier l'impact de la désulfuration sur la réaction de déconjugaison au moyen de deux déSUMOylases commerciales humaines, SENP1 et SENP2.

II. Synthèse totale de conjugués SUMO-2 et -3

A. La stratégie convergente « one pot »

La stratégie de synthèse des conjugués SUMO-2 et -3 utilisé est décrite à la Figure 60. La première étape débute avec la préparation des peptides SEA **1a, b** qui par une étape de transthioestérification à l'aide de MPA conduit à l'échange SEA-thiol en condition acide (pH 4) permettant de produire le peptide **3a, b** (voir annexe Figure S 9, Figure S 10). La deuxième étape correspond à l'assemblage « one pot ». Cette étape repose essentiellement sur la production de l'intermédiaire SUMO-SEA^{off} **5a, b** issu de la NCL entre les peptides **3a, b** et **4**. Puisque la NCL est réalisée en absence de TCEP ceci a permis de maintenir le groupement SEA sous sa forme disulfure cyclique inactive (SEA^{off}) ou comme mentionné précédemment dans ce manuscrit sous sa forme thioester latent (Figure 60). Ainsi, une fois l'intermédiaire formé, on vient activer le groupement SEA du domaine **5a, b** en additionnant de la TCEP au milieu réactionnel afin de débiter la ligation SEA avec le peptide cible **6**. Le peptide **6** correspond à la partie N-terminale de SUMO-2 qui inclut le site de SUMOylation VKTE (Figure 5).



1,3a, SUMO-2 (2-47): ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY
1,3b, SUMO-3 (2-46): SEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY
4, SUMO-2 (48-93) or SUMO-3 (47-92): CERQGLSMRQIRFRFDGQPINETDTPAQLME-
 DEDTIDVFQQQTGG

Figure 60. Synthèse totale des conjugués SUMO-2 et -3.

Les analyses UPLC-MS présentées à la Figure 61 nous ont permis de confirmer l'identité et la qualité des produits obtenus.

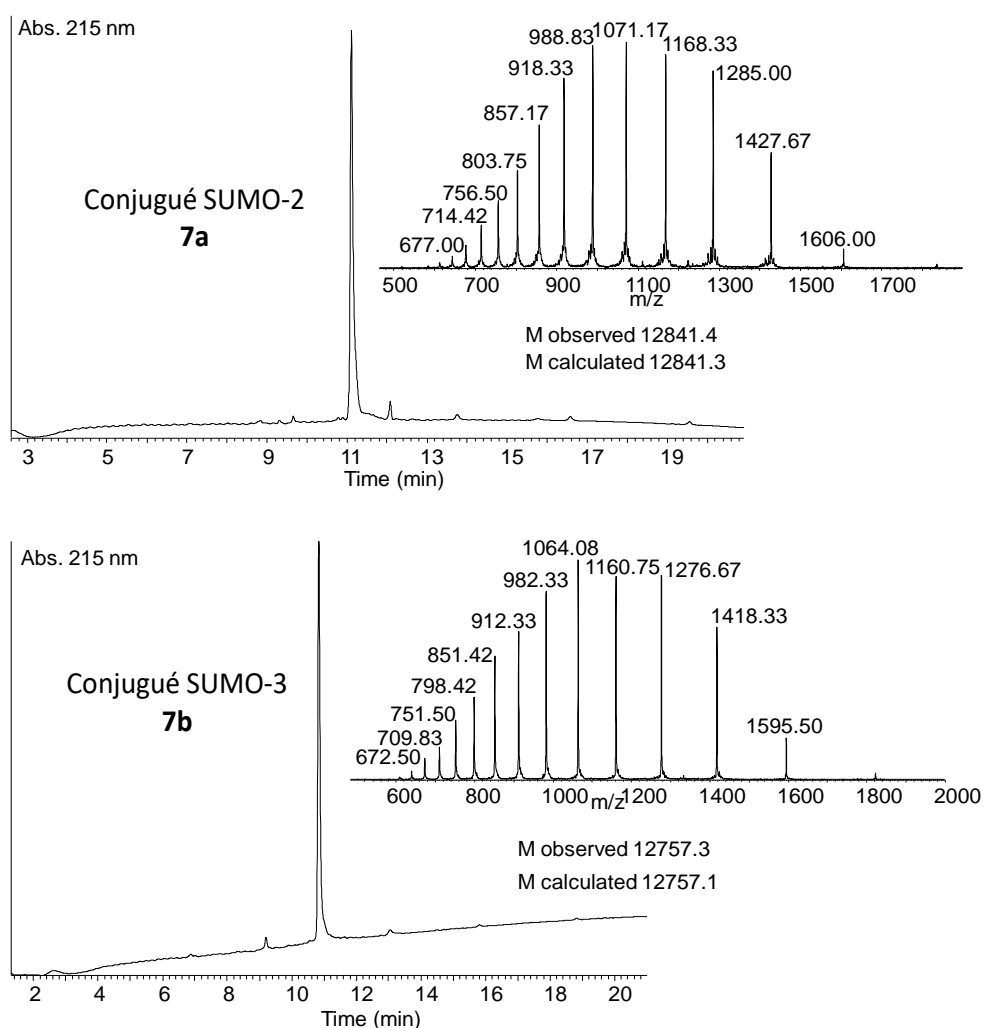


Figure 61. Analyses UPLC-MS des conjugués SUMO-2 **7a** et SUMO-3 **7b**.

B. Problématique de la désulfuration

La synthèse des conjugués fait intervenir une étape finale de désulfuration du résidu thiol impliqué dans la liaison isopeptidique afin d'obtenir un conjugué SUMO natif. Tout bien considéré, le but de ces assemblages étant de comprendre le rôle de la Cys interne sur la réaction de déconjugaison il est donc nécessaire de réaliser deux types de désulfuration permettant de moduler l'exposition des résidus thiols présents au sein des conjugués. Une désulfuration sélective (sans Gn.HCl, Figure 62, partie gauche) conduira à des conjugués présentant un domaine SUMO natif, alors qu'une désulfuration globale donne accès à des conjugués présentant la modification Cys→Ala (6M Gn.HCl, Figure 62, partie droite).

La désulfuration en conditions natives conduit à une perte de 32 uma (uma, unité de masse atomique), due à une désulfuration de l'acide aminé amino-thiol impliqué dans le lien isopeptidique. Cette sélectivité a été clairement établie par une analyse protéomique des produits obtenus (voir annexe Figure S 28, Figure S 29, Figure S 30, Figure S 31). La désulfuration en conditions dénaturantes conduit comme attendu à une perte de 64 uma.

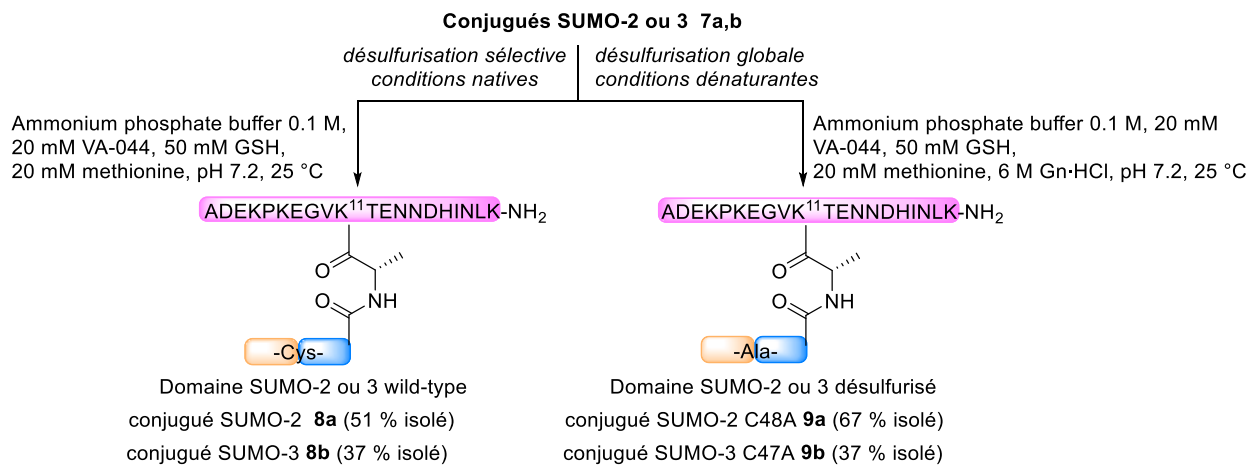


Figure 62. Désulfuration sélective ou globale des conjugués SUMO-2/3 8a, b conjugués SUMO-2/3 sélectivement désulfurés, 9a, b conjugués SUMO-2/3 totalement désulfurés.

Les conjugués SUMO-2/3 issus des différentes désulfurations (sélective ou globale) ont été analysés en spectrométrie de masse (voir annexe Figure S 15, Figure S 16) et en SDS-PAGE. Ils migrent à un poids moléculaire apparent d'environ 20 kDa alors que les monomères SUMO-2 et 3 migrent quant à eux à un poids moléculaire apparent d'environ 15 kDa (Figure 63. SDS-PAGE des protéines synthétiques SUMO-2/3 et des différents conjugués SUMO-2/3.). L'utilisation de ces conditions ont permis d'obtenir les conjugués SUMO-2 et -3 désulfurés sélectivement ou totalement avec de bon rendement et de bonnes puretés comme en témoigne le SDS-PAGE. La qualité de ces conjugués a permis la réalisation de cinétiques de déconjugaison avec 2 déSUMOylases commerciales humaines, SENP1 & SENP2.

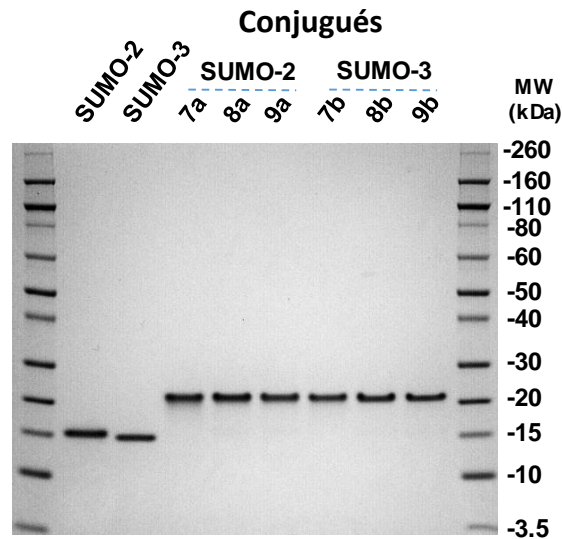


Figure 63. SDS-PAGE des protéines synthétiques SUMO-2/3 et des différents conjugués SUMO-2/3.

C. Cinétique de déconjugaison avec SENP1 et SENP2

L'effet de la modification Cys→Ala sur le clivage des conjugués SUMO-2 et -3 synthétiques par les enzymes SENP1 et SENP2 a été examiné. Le clivage enzymatique donne le peptide **10** et les domaines SUMO-2 (-3) ou SUMO-2 C48A (-3 C47A) (Figure 64). La vitesse de la réaction de déconjugaison a été facilement déterminée par LC-MS.

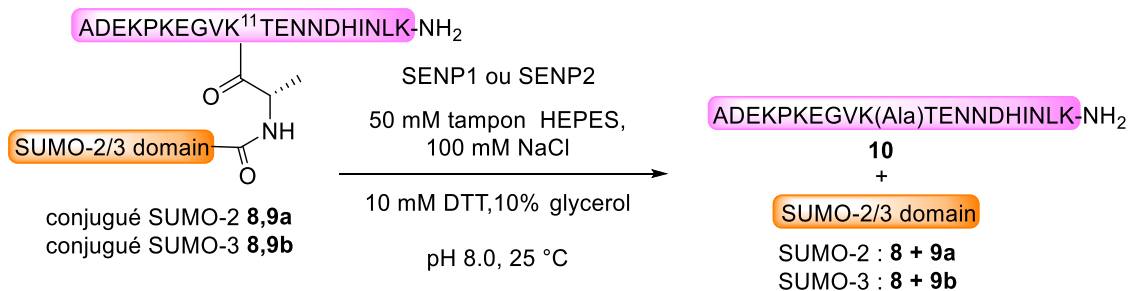


Figure 64. Déconjugaison catalysée par les deSUMOylases SENP1 et SENP2 sur les différents conjugués SUMO-2/3.

Les données issues de ces expérimentations montrent que les conjugués SUMO-2 et SUMO-2 C48A sont clivés à la même vitesse, quel que soit l'enzyme utilisée (Figure 65). Donc le domaine SUMO-2 est reconnu par SENP1 et SENP2 que la cystéine interne soit présente ou substituée par une alanine. A contrario, le conjugué SUMO-3 est clivé plus rapidement par

les deux enzymes que le conjugué SUMO-3 C47A. Ces données confirment les résultats obtenus lors des analyses en dichroïsme circulaire.

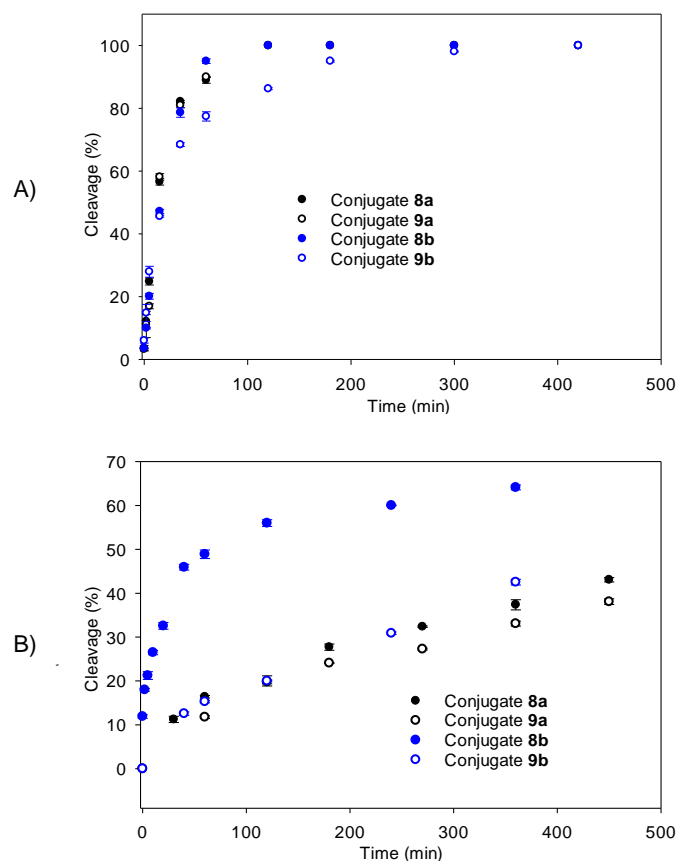


Figure 65. Cinétique de déconjugaison des conjugués SUMO-2 et SUMO-3 par SENP1 et SENP2, A) Déconjugaison par SENP1, B) Déconjugaison par SENP2. Conditions : 50 mM tampon HEPES pH 8.0, 100 mM NaCl, 10 mM DTT, 10% glycérol, 25 °C, conjugué 0.5 mg/mL, volume final 100 μ L, SENP1 5.0 ng par réaction de coupure, SENP2 0.4 units par réaction de coupure. La réaction de déconjugaison a été réalisée sous atmosphère inerte, stoppée avec 10 % acide acétique et analysée en UPLC-MS. L'expérience a été effectuée en triplicat. Les données correspondent aux moyennes et écart-types.

En conclusion, ces expériences démontrent que les protéines SUMO-2 et -3 répondent différemment à la modification Cys \rightarrow Ala autant sur le plan conformationnel que fonctionnel. Une explication possible de cette différence de comportement pourrait être l'existence d'une interaction stabilisante entre la queue N-terminale et le domaine core de SUMO-2, qui serait moins forte ou absente dans la protéine SUMO-3. Pour approfondir ce point, nous avons souhaité mesurer l'effet de la désulfuration sur la structure secondaire du domaine core SUMO-2/3 (sans l'extension N-terminale).

III. Synthèse par voie chimique du domaine SUMO-2/3 core

Les domaines SUMO-2 et -3 partagent 97 % d'homologie de séquence, elles ne diffèrent seulement que de quelques acides aminés localisés dans leur queue N-terminale relativement flexible, on parle également de « Tail » (Figure 66).

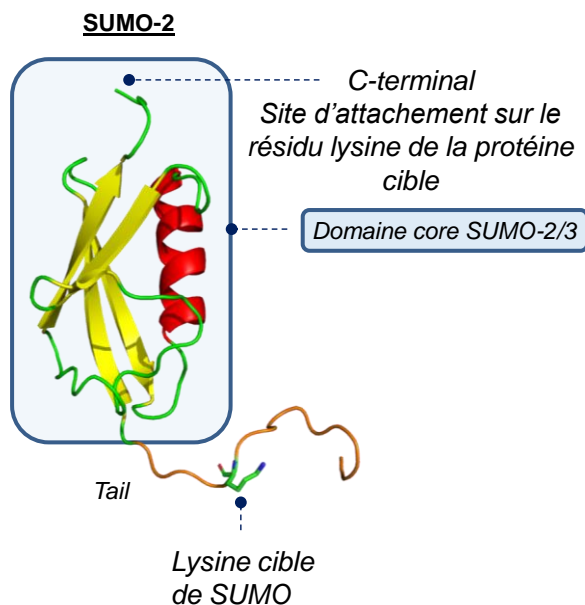


Figure 66. Structure tertiaire du domaine SUMO-2 (structure RMN, code pdb : 2N1W).

Pour comprendre les différences observées à la Figure 54 entre les protéines SUMO-2 et SUMO-3, nous avons synthétisé et désulfuré le domaine core des protéines SUMO-2/3 afin de déterminer l'influence de la partie N-terminale sur la stabilité de celui-ci. Les variant du domaine core SUMO-2/3 ont été préparé comme décrit à la Figure 67. Le domaine core SUMO-2/3 est obtenu après la réaction de ligation entre le peptide SEA **1c** et le peptide **2**. Concernant l'étape de désulfuration, les conditions utilisées pour les protéines SUMO-2 et SUMO-3 ont été utilisées pour la préparation du domaine core SUMO-2/3 C48A.

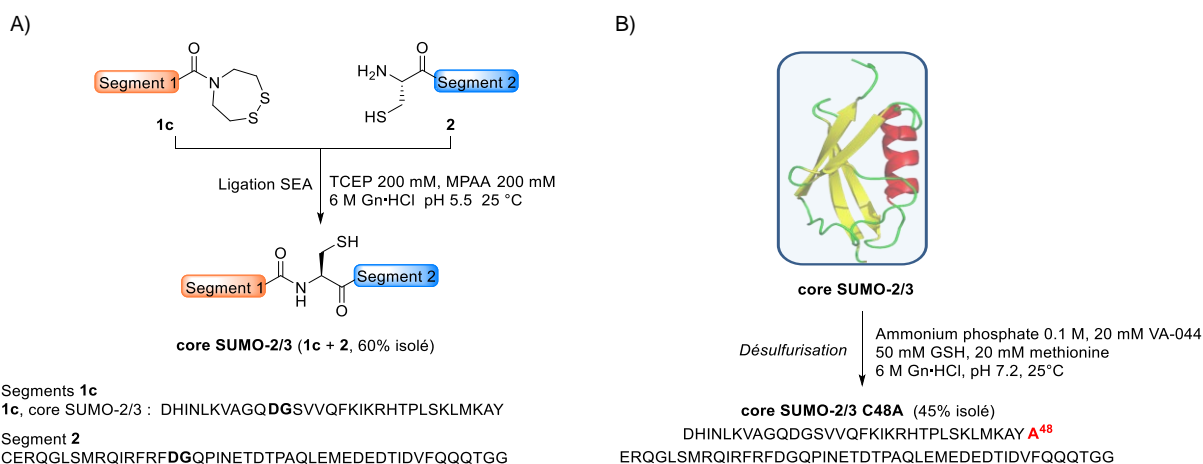


Figure 67. A) Synthèse chimique du domaine core SUMO-2/3. B) Préparation de l'analogue désulfuré du domaine core SUMO-2/3 C48A.

L'analyse des analogues core SUMO-2/3 par CD montre que le domaine core est également sensible à la modification Cys→Ala, avec une perte de contenu en hélice α due à la désulfuration de la Cys interne similaire à celle observée pour le domaine SUMO-3 (Figure 68).

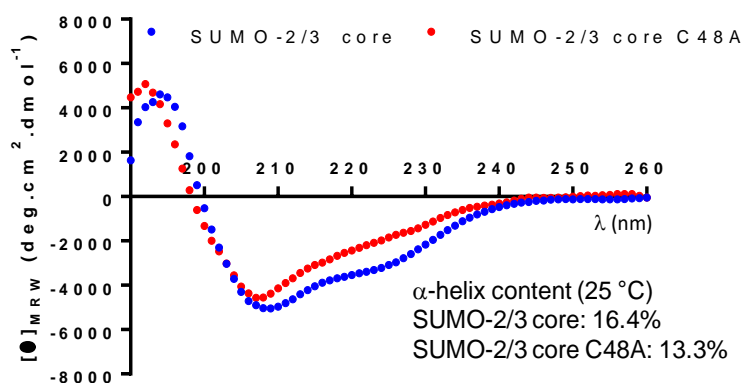


Figure 68. Analyse CD des domaines core SUMO-2/3 et SUMO-2/3 C48A. Conditions : tampon phosphate 10 mM pH 7.2. Les échantillons ont été soigneusement dégazés pendant 30 min avec de l'argon et immédiatement transférées sous argon dans la cuve à CD. Concentration en protéines : core SUMO-2/3 à 21.2 μ M, core SUMO-2/3 C48A à 22.4 μ M.

Cette dernière expérience suggère l'existence d'une interaction stabilisante entre la partie N-terminale qui était considérée jusqu'à là comme une région peu structurée et le domaine core SUMO-2/3, principalement dans la protéine SUMO-2. La modification Cys→Ala apparaît comme un outil intéressant pour découvrir les interactions faibles se produisant entre la queue et le domaine core des protéines SUMO. Cette conclusion est importante et constitue

la première démonstration que les protéines SUMO-2 et SUMO-3 ne sont pas identiques en terme de propriétés conformationnelles.

Pour résumer, nous avons préparé avec succès par synthèse totale les protéines SUMO-2 et SUMO-3 ainsi que leurs analogues désulfurisés. La production de ces analogues a permis d'étudier l'impact de la modification Cys→Ala sur la structure secondaire, la stabilité thermique et la capacité des domaines SUMO-2 et -3 à être conjugués sur un substrat cible, RanGAP. Nous avons également regardé l'effet de la désulfuration sur la vitesse de déconjugaison des conjugués SUMO-2 et SUMO-3 avec les deSUMOylases SENP1 et SENP2. Contrairement au domaine SUMO-2, la structure secondaire du domaine SUMO-3 a été significativement affectée par la modification Cys→Ala. De même, la vitesse de déconjugaison du domaine SUMO-3 par SENP1 et SENP2 est réduite lorsque la Cys interne est convertie en Ala. Les analyses CD des analogues SUMO-2, -3 et du domaine core SUMO-2/3 suggèrent l'existence d'une interaction stabilisante entre le Tail et le domaine core dans la protéine SUMO-2 qui se retrouve moins prononcée voire absente dans la protéine SUMO-3, rendant cette dernière plus sensible à la modification Cys→Ala. Ces résultats mettent en avant une différence subtile entre les protéines SUMO-2 et -3 qui ne doivent plus être considérées comme identiques malgré leur forte homologie de séquence. Par conséquent, la modification de la Cys interne dans le domaine SUMO-3 doit être prise avec précaution. Ce travail permet de clarifier la stratégie de synthèse à adopter pour accéder aux dimères SUMO-2/3.

IV. Synthèse totale des dimères SUMO-2/3

Alors que la synthèse totale des conjugués Ub ou de chaînes polyUb fait l'objet d'une attention particulière durant les dernières décennies, les publications décrivant la synthèse de conjugués SUMO se font plus rares. Cependant, déterminer le rôle de la PolySUMOylation (polySUMO-2/3) nécessite la production de tels objets de façon homogène et native, pour en étudier les propriétés physico-chimiques et les fonctionnalités biochimiques.

Puisque les protéines SUMO-2 et -3 ne peuvent plus être considérées comme redondantes, cela implique que les chaînes polySUMO-2/3 peuvent être modulées en fonction de la proportion de chaque isoforme présentes dans ces dernières. D'autre part, bien que la présence d'une interaction entre la queue N-terminale et le domaine core de SUMO-2 reste encore hypothétique, son éventuelle existence ouvre la voie à un questionnement quant à la modification de la Lys₁₁ (appartenance au site de SUMOylation interne, Figure 5) sur la fonction et la structure de la chaîne polySUMO-2/3. Ces différences entre les deux monomères impactent potentiellement la structure des chaînes polySUMO-2/3 et leur agencement dans l'espace. Pour initier l'étude de tels objets et tenter de répondre à ces questions, il est

nécessaire d'accéder à toutes les combinaisons de dimères SUMO-2/3. Cependant, la synthèse de tels objets de poids moléculaire proche de 21 kDa (188-187 acides aminés) s'avèrent être un réel défi avec de nombreuses difficultés à surmonter.

A. La stratégie de synthèse générale

Tenant compte des éléments précédemment décrits, nous avons conçu une stratégie de synthèse facile et robuste donnant accès à toutes les combinaisons des dimères SUMO-2/3. La stratégie est décrite à la Figure 69, en prenant comme exemple l'assemblage du dimère SUMO-2-SUMO-3. La première étape de cette synthèse ambitieuse permet l'assemblage des segments SUMO- X^N ($X= 2$ ou 3) et SUMO- X^C ($X= 2$ ou 3) pour former les protéines SUMO-2 ou SUMO-3 équipées en C-terminal d'un groupement réactif SEA, sous sa forme thioester latent SEA^{off}. Parallèlement, la synthèse du domaine SUMO-2 Lys₁₁(Cys) ou SUMO-3 Lys₁₁(Cys) est réalisée en « one pot ». Enfin, les deux domaines équipés de fonctionnalités complémentaires peuvent être reliés entre eux par le biais d'un pseudo lien isopeptidique grâce à une ligation SEA.

peptide ajouté par la suite en one-pot. Les segments **14a** ou **14b** ne sont pas isolés et sont directement mis en ligation avec les segments **2** ou **13** en ajoutant un excès de MPAA et de TCEP. Une fois la seconde étape de ligation terminée, nous avons réalisé la déprotection du groupement AcA in situ en additionnant 2 équivalents d'hydroxylamine dans le milieu réactionnel. L'utilisation d'un détergent, le N-octylglucoside, et de 6 M Gn.HCl comme additifs dans le mélange réactionnel s'est avérée obligatoire pour assurer la solubilisation des segments peptidiques de départ et des produits formés tout au long du procédé de synthèse. Enfin, les domaines SUMO-2/3 Lys₁₁(Cys) **16a, b** ont été isolés avec de bons rendements globaux (58-60% après purification par HPLC, voir annexe Figure S 18, Figure S 19) et de bonnes puretés. La stratégie a également été appliquée avec succès à la synthèse du domaine SUMO-3 Lys₁₁(Cys) contenant un groupe hydrazide C-terminal, c'est-à-dire SUMO-3 Lys₁₁(Cys) **17**, permettant d'envisager dans le futur la synthèse de chaînes polySUMO-2/3 de plus grande complexité.

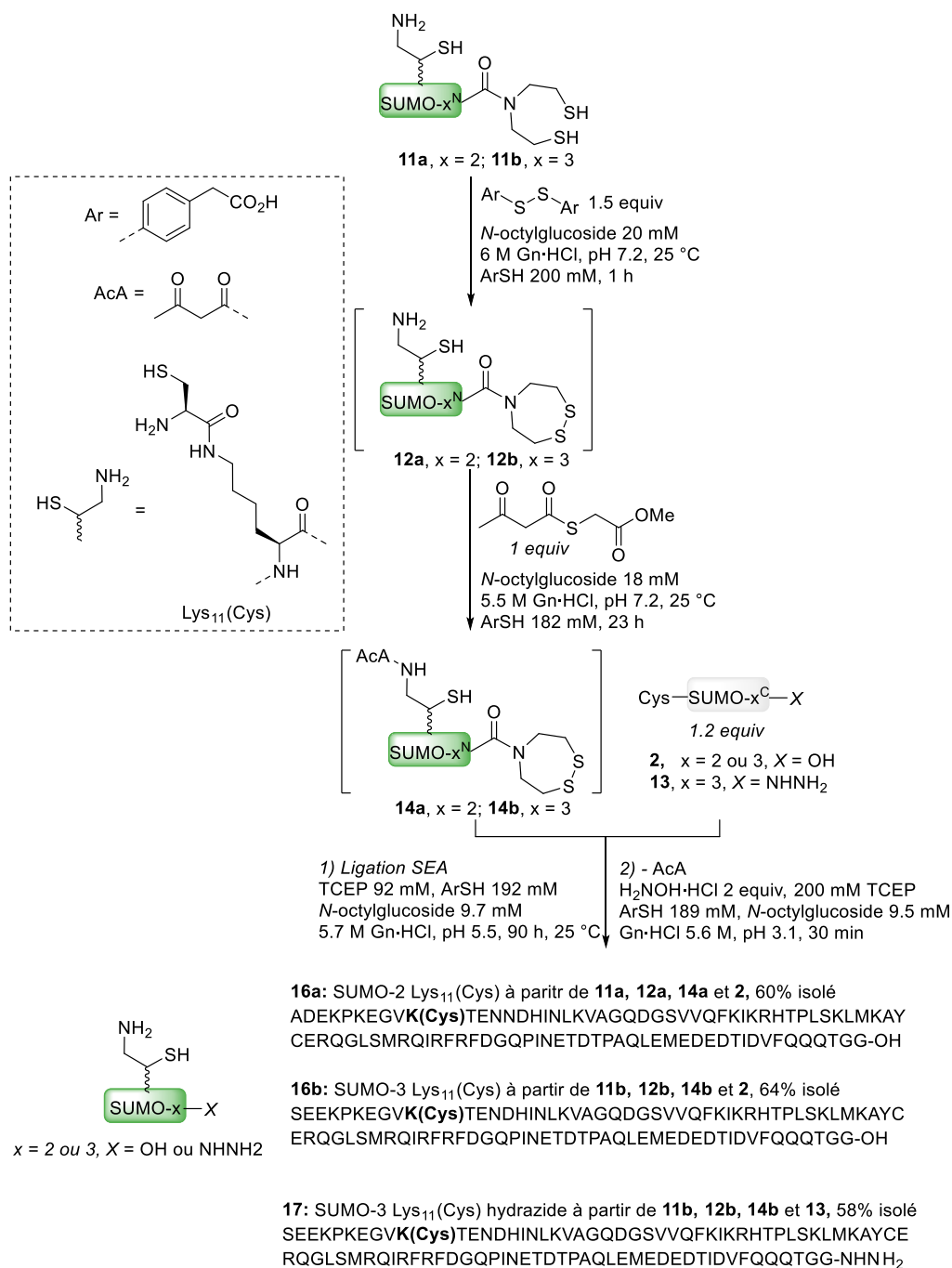


Figure 70. Synthèse « one pot » des domaines SUMO-2 et -3 équipés du motif Lys₁₁(Cys).

B. L'assemblage « one pot » des dimères SUMO-2/3

La réalisation de la synthèse des domaines SUMO **16a, b** et **17** a ouvert la voie à l'assemblage des dimères SUMO-2/3 comme décrit à la Figure 71. Les différents dimères SUMO-2/3 ont été produits avec succès et isolés par HPLC avec une excellente pureté (voir annexe Figure S 20).

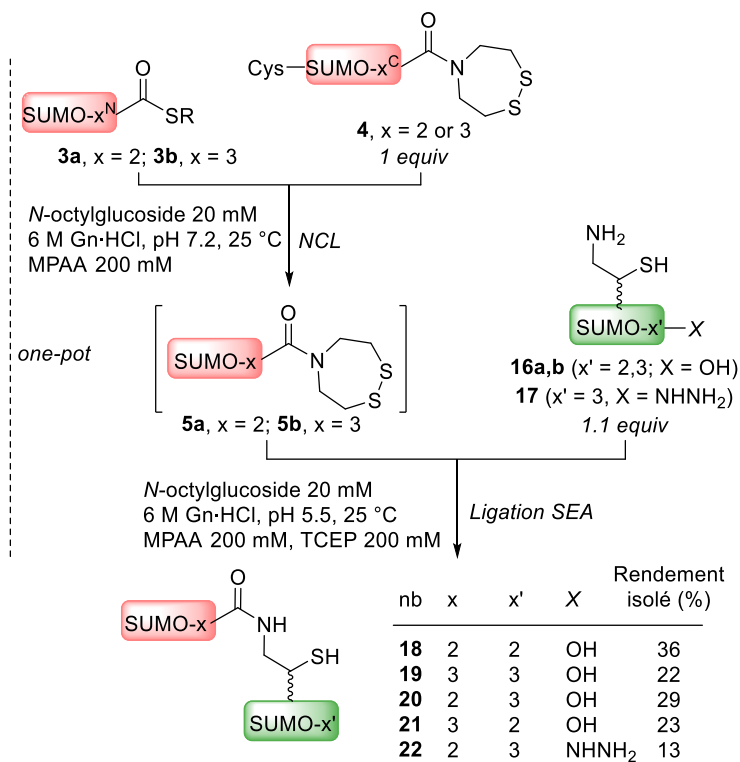


Figure 71. A) Synthèse chimique des dimères SUMO-2/3. SDS-PAGE des dimères SUMO-2/3.

Tous les dimères ont été testés par SDS-PAGE et migrent tous selon un poids moléculaire apparent de 30 kDa (Figure 72). Ces molécules composées de 183 à 185 résidus d'acides aminés sont une nouvelle démonstration de la puissance des systèmes de *N,S*-acyl shift pour accéder à des protéines complexes

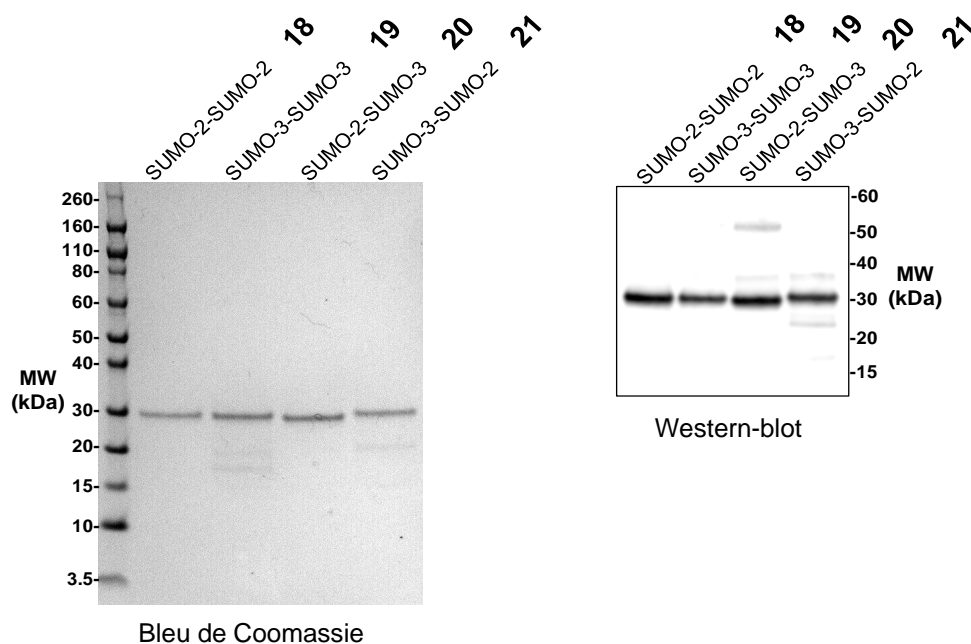


Figure 72. Bleu de Coomassie des dimères SUMO-2/3, 1 μ g de chaque dimère SUMO-2/3. Western-blot des dimères SUMO-2/3 en utilisant l'anticorps polyclonal SUMO-2/3, 100 ng pour chaque dimère SUMO-2/3, MW= marqueur de poids moléculaire.

La synthèse réussie des différents dimères SUMO-2/3 a ouvert la voie aux expériences de désulfuration et par la suite aux analyses biophysiques et biochimiques.

C. L'étape de désulfuration sélective des dimères SUMO-2/3

Outre les étapes d'assemblage, la stratégie fait intervenir une étape de désulfuration du résidu β -aminothiols impliqué dans la formation du lien isopeptidique entre les deux domaines SUMO. Cette désulfuration peut être non sélective, et ainsi conduire également à la désulfuration des cystéines internes des domaines SUMO-2/3, ou au contraire être sélective. Le choix de l'une ou l'autre option reposait en grande partie sur le rôle que pouvait jouer cette cystéine centrale sur la stabilité et la conformation adoptée par les domaines SUMO-2 et -3. Les résultats préliminaires ont montré l'importance de la cystéine interne pour la structure et la fonction des domaines SUMO-2 et SUMO-3. Ainsi, nous avons envisagé une désulfuration sélective afin de préserver les Cys internes des domaines SUMO. De manière générale, la sélectivité de la désulfuration repose sur une différence d'accessibilité entre le résidu thiol à éliminer, qui est exposé aux solvants, et le thiol de la Cys interne des domaines SUMO, qui est enfouie au cœur du noyau hydrophobe du domaine SUMO structuré. Cependant, l'extension du procédé de désulfuration sélective aux dimères pour lesquels la liaison isopeptidique est située entre deux domaines structurés devait être établie. A noter que

Liu et ses collaborateurs ont pu observer un masquage des liaisons isopeptidiques dans certaines chaînes de polyubiquitine²⁹¹.

La désulfuration des quatre dimères induits par la TCEP et le VA-044 dans des conditions natives a montré sur une sélectivité totale pour le résidu thiol de la cystéine présente entre les deux domaines SUMO (Figure 73), comme le démontre une analyse protéomique approfondie des produits désulfurés (voir annexe Figure S 32 à Figure S 38). Dans une expérience témoin, nous avons effectué la désulfuration du dimère SUMO-2/3 **20** en conditions dénaturantes (6 M Gn.HCl). Dans ce cas, tous les résidus Cys ont été convertis en Ala (voir annexe Figure S 39). Ainsi, dans des conditions natives, le thiol du motif Lys₁₁(Cys) utilisé pour lier les domaines SUMO reste accessible au solvant, et ce quelque soit le dimère considéré. La protection efficace des résidus Cys internes dans des conditions natives montre que les deux domaines SUMO sont correctement structurés.

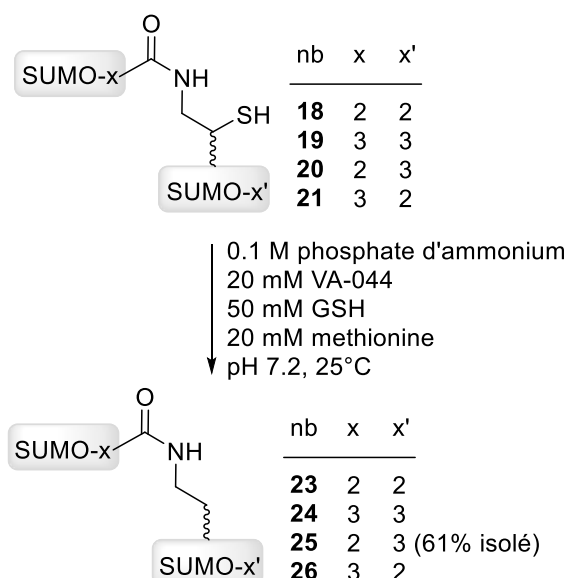


Figure 73. Procédé de désulfuration sélective des dimères SUMO-2/3 **18-21**

Une fois le procédé de désulfuration pour les dimères SUMO-2/3 validé, nous avons réalisé la désulfuration sélective du dimère SUMO-2/3 **20** sur une échelle préparative pour produire l'analogue désulfuré **25** avec un rendement isolé de 61% (Figure 73). L'obtention à l'échelle du milligramme de ces dimères **20** et **25** nous a permis d'effectuer différentes analyses biochimiques ainsi qu'une étude par dichroïsme circulaire.

D. Caractérisations biophysiques et biochimiques du dimère SUMO-2/SUMO-3

1. Analyse structurale par dichroïsme circulaire

Dans un premier temps, les dimères **20** et **25** ont été analysés par dichroïsme circulaire (CD) comme le montre la Figure 74 afin de confirmer la bonne conformation des dimères. L'analyse CD indique que la structure secondaire des dimères **20** et **25**, en particulier leur contenu en hélice α (18%) sont très similaires. De plus, le contenu en hélice alpha trouvé pour les dimères est cohérent avec celui mesuré pour les monomères SUMO-2 et SUMO-3 (15%). De plus, les spectres CD des dimères **20** et **25** sont similaires au spectre CD reconstitué à partir des spectres CD expérimentaux obtenus pour les domaines SUMO-2 et SUMO-3 individuels

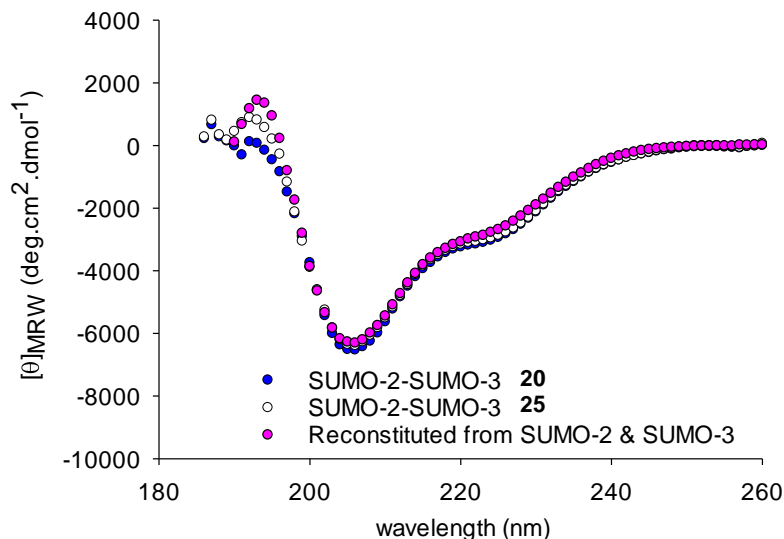


Figure 74. Analyses CD des dimères SUMO-2/3 **20** et **25**, comparaison avec la reconstitution calculée à partir des spectres CD des monomères.

La confirmation de la bonne conformation des dimères nous a permis de poursuivre les analyses en effectuant un test de déconjugaison *in vitro*. Ceci nous a permis d'explorer l'activité biochimique du dimère **25**, celui dont la structure se rapproche le plus de la structure native.

2. Test de déconjugaison par les enzymes SENP1 et SENP2

Nous avons regardé le clivage du dimère **25** a été examiné par les déSUMOylases 1 (SENP1)²⁹² et 2 (SENP2)^{293, 294}, qui sont des enzymes de déconjugaison humaines spécifiques de SUMO-1, -2 et -3^{55, 295}. Les réactions enzymatiques ont été réalisées à 37 °C

et suivies par SDS-PAGE. Les résultats démontrent que le dimère **25** a été clivé par les deux enzymes SENP (Figure 75). Ce clivage a conduit à la production des protéines SUMO-2 et SUMO-3 Lys₁₁(Ala) avec un poids moléculaire apparent de ~ 15 kDa qui ont été révélées par coloration au bleu de Coomassie ou par Western blot.

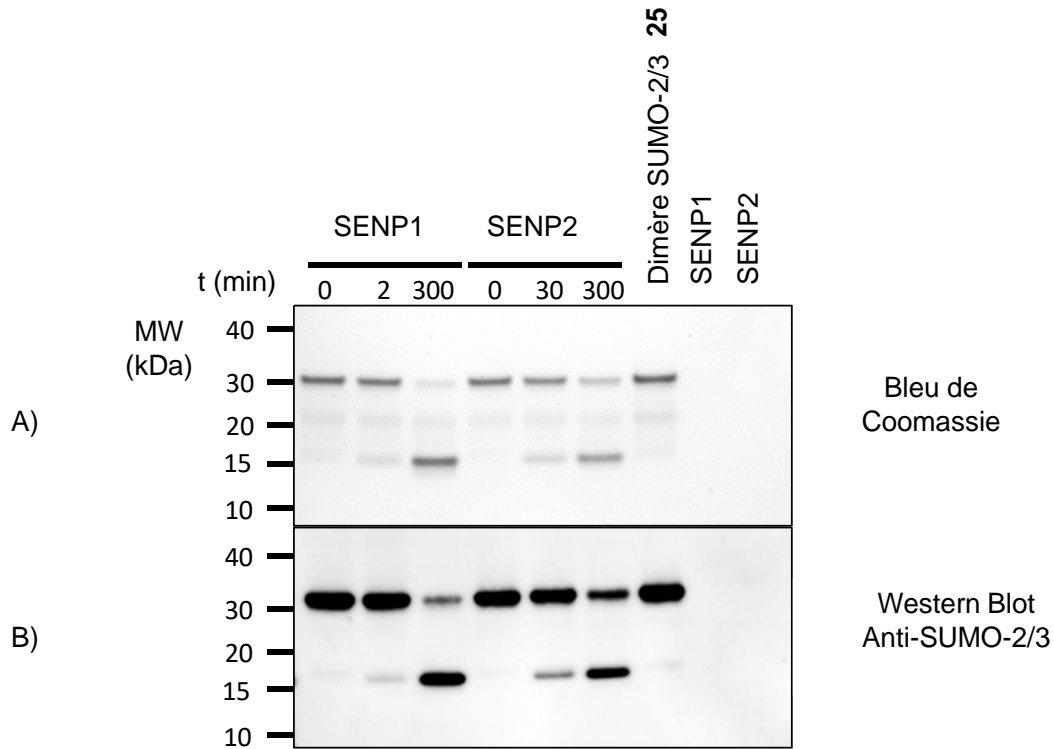


Figure 75. Clivage du dimère SUMO-2/3 **25** par les enzymes de déconjugaison SENP1 et SENP2 (temps en minutes). A) Coloration de Coomassie, 2 µg du dimère **25**. B) Analyse par Western blot en utilisant un anticorps polyclonal anti-SUMO-2/3, 100 ng du dimère **25**. MW: marqueurs de poids moléculaire apparent en kDa.

Pris ensemble, les données des analyses CD du dimère SUMO-2/3 **25** et son clivage par les déSUMOylases SENP1 et SENP2 montrent que le dimère est correctement plié et fonctionnel. Ainsi, cela nous permet de confirmer la pertinence de la stratégie synthèse mise en place pour accéder à des objets moléculaires complexes de part leur taille et leur structure (ramifié).

CONCLUSION & PERSPECTIVES

Dans un premier temps, nous avons réalisé l'assemblage d'une collection de monomères SUMO-2 et -3 dans le but d'étudier l'impact de la désulfuration de la Cys interne sur la conformation et la fonctionnalité des domaines SUMO-2 et -3. Les analyses par dichroïsme circulaire montrent que la désulfuration altère la structure secondaire et la stabilité thermique des protéines SUMO-2 et -3, SUMO-3 étant plus sensible à la modification que SUMO-2. De même les tests biochimiques montrent un effet significatif de la modification Cys→Ala sur la vitesse sur la déconjugaison des conjugués SUMO-3 par les enzymes SENP1 et SENP2. Les analyses CD de SUMO-2, SUMO-3 et core SUMO-2/3 suggèrent l'existence d'une interaction stabilisante entre la queue du domaine SUMO et le domaine principal dans la protéine SUMO-2, interaction qui serait moins marquée ou absente chez la protéine SUMO-3. Pris ensemble, nos résultats montrent que la modification de la cystéine interne de la protéine SUMO-3 doit être évitée impérativement. Ils révèlent également des différences subtiles entre les protéines SUMO-2 et -3 qui ne peuvent plus être considérées comme identiques, bien que leurs séquences soient très similaires (97% d'homologie). Nous montrons ici pour la première fois que SUMO-2 et SUMO-3 sont des protéines bien distinctes et ne doivent plus être considérées comme redondantes.

Dans un second temps, les connaissances acquises sur le rôle de la Cys interne sur les propriétés des protéines SUMO-2 et -3 nous ont permis de concevoir une stratégie simple et efficace pour accéder aux dimères SUMO-2/3. Au-delà de leur architecture particulière ne permettant pas aux systèmes vivants de les produire, la taille de ces objets se situe à la limite haute des protéines pouvant être produites par synthèse chimique. Ces synthèses ont nécessité de nombreuses optimisations pour minimiser la formation d'aspartimide, le nombre de purifications intermédiaires et atteindre une grande sélectivité dans la réaction de désulfuration. Pour valider les protéines ainsi produites, nous avons procédé à différentes analyses. D'une part, les analyses par dichroïsme circulaire nous ont permis de nous rendre compte de la bonne conformation adoptée par le dimère SUMO-2/3. Nous avons également déterminé la cinétique de coupure du dimère SUMO-2/3 **25** par les déSUMOylases SENP1 et 2. La production de dimères SUMO-2/3 à l'échelle mg et fonctionnels est une étape importante vers de nouvelles recherches visant à déterminer l'effet de la composition des dimères SUMO-2/3 et de l'ordre de concaténation sur leurs propriétés biophysiques et biochimiques. Enfin, la stratégie de synthèse est compatible avec la production de dimères SUMO-2/3 dotés d'une fonctionnalité hydrazide en position C-terminale. L'accès à des dimères SUMO-2/3 hydrazide ouvre la voie à la production de chaînes polySUMO-2/3 plus élaborées en exploitant les propriétés de thioester masqué du groupement hydrazide.

Mes travaux de thèse ouvrent la voie à une meilleure compréhension du rôle de la SUMOylation et du dialogue avec d'autres modifications post-traductionnelles. Par exemple, il a été démontré que la protéine SUMO-1 peut être phosphorylé sur son résidu de sérine 2.

Cette sérine est conservée dans la protéine SUMO-3 et est absente dans SUMO-2, la position étant occupée par une alanine. Aucune donnée n'a été publiée à ce jour concernant la phosphorylation du domaine SUMO-3 *in vivo*. Mais si cette phosphorylation a effectivement lieu, il serait du plus grand intérêt d'étudier l'effet de la phosphorylation de la sérine 2 sur les propriétés de SUMO-3 et naturellement des dimères SUMO-2/SUMO-3, SUMO-3/SUMO-2, SUMO-3/SUMO-3.

En conclusion, la SUMOylation est une PTM au cœur de nombreux processus cellulaires pouvant contribuer à l'initiation, le développement et la progression de certaines pathologies humaines. De ce fait, il est essentiel de comprendre tous les dessous de cette PTM dans le but de prédire l'impact des différentes dérégulations de la balance SUMOylation/polySUMOylation/déSUMOylation et ainsi identifier de nouvelles stratégies thérapeutiques.

RÉFÉRENCES BIBLIOGRAPHIQUES

- (1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed Engl.* **2005**, *44* (45), 7342–7372. <https://doi.org/10.1002/anie.200501023>.
- (2) Cohen, P. The Regulation of Protein Function by Multisite Phosphorylation – a 25 Year Update. *Trends Biochem. Sci.* **2000**, *25* (12), 596–601. [https://doi.org/10.1016/S0968-0004\(00\)01712-6](https://doi.org/10.1016/S0968-0004(00)01712-6).
- (3) Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y. Chemical Synthesis of an Erythropoietin Glycoform Containing a Complex-Type Disialyloligosaccharide. *Angew Chem Int Ed* **2012**, *51*, 3567–3572. <https://doi.org/10.1002/anie.201109034>.
- (4) Seo, J.; Lee, K.-J. Post-Translational Modifications and Their Biological Functions: Proteomic Analysis and Systematic Approaches. *J. Biochem. Mol. Biol.* **2004**, *37* (1), 35–44. <https://doi.org/10.5483/BMBRep.2004.37.1.035>.
- (5) Pekkala, M.; Hieta, R.; Bergmann, U.; Kivirikko, K. I.; Wierenga, R. K.; Myllyharju, J. The Peptide-Substrate-Binding Domain of Collagen Prolyl 4-Hydroxylases Is a Tetratricopeptide Repeat Domain with Functional Aromatic Residues. *J. Biol. Chem.* **2004**, *279* (50), 52255–52261. <https://doi.org/10.1074/jbc.M410007200>.
- (6) Bandyopadhyay, P. K. Vitamin K-Dependent Gamma-Glutamylcarboxylation: An Ancient Posttranslational Modification. *Vitam. Horm.* **2008**, *78*, 157–184. [https://doi.org/10.1016/S0083-6729\(07\)00008-8](https://doi.org/10.1016/S0083-6729(07)00008-8).
- (7) Goldstein, G.; Scheid, M.; Hammerling, U.; Schlesinger, D. H.; Niall, H. D.; Boyse, E. A. Isolation of a Polypeptide That Has Lymphocyte-Differentiating Properties and Is Probably Represented Universally in Living Cells. *Proc. Natl. Acad. Sci.* **1975**, *72* (1), 11–15. <https://doi.org/10.1073/pnas.72.1.11>.
- (8) Ciechanover, A. The Ubiquitin Proteolytic System: From an Idea to the Patient Bed. *Proc. Am. Thorac. Soc.* **2006**, *3* (1), 21–31. <https://doi.org/10.1513/pats.200510-106JH>.
- (9) Wiborg, O.; Pedersen, M. S.; Wind, A.; Berglund, L. E.; Marcker, K. A.; Vuust, J. The Human Ubiquitin Multigene Family: Some Genes Contain Multiple Directly Repeated Ubiquitin Coding Sequences. *EMBO J.* **1985**, *4* (3), 755–759.
- (10) Baker, R. T.; Board, P. G. The Human Ubiquitin-52 Amino Acid Fusion Protein Gene Shares Several Structural Features with Mammalian Ribosomal Protein Genes. *Nucleic Acids Res.* **1991**, *19* (5), 1035–1040.
- (11) Dwane, L.; Gallagher, W. M.; Chonghaile, T. N.; O'Connor, D. P. The Emerging Role of Non-Traditional Ubiquitination in Oncogenic Pathways. *J. Biol. Chem.* **2017**, *292* (9), 3543–3551. <https://doi.org/10.1074/jbc.R116.755694>.
- (12) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The Absolute Quantification Strategy: A General Procedure for the Quantification of Proteins and Post-Translational Modifications. *Methods* **2005**, *35* (3), 265–273. <https://doi.org/10.1016/j.ymeth.2004.08.018>.
- (13) Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. A Proteomics Approach to Understanding Protein Ubiquitination. *Nat. Biotechnol.* **2003**, *21* (8), 921–926. <https://doi.org/10.1038/nbt849>.
- (14) Varadan, R.; Assfalg, M.; Raasi, S.; Pickart, C.; Fushman, D. Structural Determinants for Selective Recognition of a Lys48-Linked Polyubiquitin Chain by a UBA Domain. *Mol. Cell* **2005**, *18* (6), 687–698. <https://doi.org/10.1016/j.molcel.2005.05.013>.
- (15) Raasi, S.; Orlov, I.; Fleming, K. G.; Pickart, C. M. Binding of Polyubiquitin Chains to Ubiquitin-Associated (UBA) Domains of HHR23A. *J. Mol. Biol.* **2004**, *341* (5), 1367–1379. <https://doi.org/10.1016/j.jmb.2004.06.057>.
- (16) Hochstrasser, M. Origin and Function of Ubiquitin-like Proteins. *Nature* **2009**, *458* (7237), 422–429. <https://doi.org/10.1038/nature07958>.
- (17) Speer, S. D.; Li, Z.; Buta, S.; Payelle-Brogard, B.; Qian, L.; Vigant, F.; Rubino, E.; Gardner, T. J.; Wedeking, T.; Hermann, M.; Duehr, J.; Sanal, O.; Tezcan, I.; Mansouri, N.; Tabarsi, P.; Mansouri, D.; Francois-Newton, V.; Daussy, C. F.;

- Rodriguez, M. R.; Lenschow, D. J.; Freiberg, A. N.; Tortorella, D.; Piehler, J.; Lee, B.; García-Sastre, A.; Pellegrini, S.; Bogunovic, D. ISG15 Deficiency and Increased Viral Resistance in Humans but Not Mice. *Nat. Commun.* **2016**, *7* (1), 1–10. <https://doi.org/10.1038/ncomms11496>.
- (18) Amir, R. E.; Iwai, K.; Ciechanover, A. The NEDD8 Pathway Is Essential for SCF(Beta-TrCP)-Mediated Ubiquitination and Processing of the NF-Kappa B Precursor P105. *J. Biol. Chem.* **2002**, *277* (26), 23253–23259. <https://doi.org/10.1074/jbc.M200967200>.
- (19) Wu, K.; Fuchs, S. Y.; Chen, A.; Tan, P.; Gomez, C.; Ronai, Z.; Pan, Z.-Q. The SCFHOS/ β -TRCP-ROC1 E3 Ubiquitin Ligase Utilizes Two Distinct Domains within CUL1 for Substrate Targeting and Ubiquitin Ligation. *Mol. Cell. Biol.* **2000**, *20* (4), 1382–1393.
- (20) Johnson, E. S. Protein Modification by SUMO. *Annu. Rev. Biochem.* **2004**, *73*, 355–382. <https://doi.org/10.1146/annurev.biochem.73.011303.074118>.
- (21) Lapenta, V.; Chiurazzi, P.; van der Spek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C. SMT3A, a Human Homologue of TheS. Cerevisiae SMT3 Gene, Maps to Chromosome 21qter and Defines a Novel Gene Family. *Genomics* **1997**, *40* (2), 362–366. <https://doi.org/10.1006/geno.1996.4556>.
- (22) Mahajan, R.; Gerace, L.; Melchior, F. Molecular Characterization of the SUMO-1 Modification of RanGAP1 and Its Role in Nuclear Envelope Association. *J. Cell Biol.* **1998**, *140* (2), 259–270.
- (23) Johnson, P. R.; Hochstrasser, M. SUMO-1: Ubiquitin Gains Weight. *Trends Cell Biol.* **1997**, *7* (10), 408–413. [https://doi.org/10.1016/S0962-8924\(97\)01132-X](https://doi.org/10.1016/S0962-8924(97)01132-X).
- (24) Meulmeester, E.; Kunze, M.; Hsiao, H. H.; Urlaub, H.; Melchior, F. Mechanism and Consequences for Paralog-Specific Sumoylation of Ubiquitin-Specific Protease 25. *Mol. Cell* **2008**, *30* (5), 610–619. <https://doi.org/10.1016/j.molcel.2008.03.021>.
- (25) Kurepa, J.; Walker, J. M.; Smalle, J.; Gosink, M. M.; Davis, S. J.; Durham, T. L.; Sung, D.-Y.; Vierstra, R. D. The Small Ubiquitin-like Modifier (SUMO) Protein Modification System in Arabidopsis. Accumulation of SUMO1 and -2 Conjugates Is Increased by Stress. *J. Biol. Chem.* **2003**, *278* (9), 6862–6872. <https://doi.org/10.1074/jbc.M209694200>.
- (26) Hay, R. T. SUMO: A History of Modification. *Mol. Cell* **2005**, *18* (1), 1–12. <https://doi.org/10.1016/j.molcel.2005.03.012>.
- (27) Vertegaal, A. C. O. SUMO Chains: Polymeric Signals. *Biochem. Soc. Trans.* **2010**, *38* (Pt 1), 46–49. <https://doi.org/10.1042/BST0380046>.
- (28) Becker, J.; Barysch, S. V.; Karaca, S.; Dittner, C.; Hsiao, H.-H.; Diaz, M. B.; Herzig, S.; Urlaub, H.; Melchior, F. Detecting Endogenous SUMO Targets in Mammalian Cells and Tissues. *Nat. Struct. Mol. Biol.* **2013**, *20* (4), 525–531. <https://doi.org/10.1038/nsmb.2526>.
- (29) Evdokimov, E.; Sharma, P.; Lockett, S. J.; Lualdi, M.; Kuehn, M. R. Loss of SUMO1 in Mice Affects RanGAP1 Localization and Formation of PML Nuclear Bodies, but Is Not Lethal as It Can Be Compensated by SUMO2 or SUMO3. *J. Cell Sci.* **2008**, *121* (Pt 24), 4106–4113. <https://doi.org/10.1242/jcs.038570>.
- (30) Zhang, X.; Lu, X.-W.; Liu, C.-F. Solid-Phase Synthesis of Peptide Thioacids through Hydrothiolysis of Resin-Bound Peptide Thioesters. *Tetrahedron Lett* **2008**, *49*, 6122–6125. <https://doi.org/10.1016/j.tetlet.2008.08.018>.
- (31) Matunis, M. J.; Coutavas, E.; Blobel, G. A Novel Ubiquitin-like Modification Modulates the Partitioning of the Ran-GTPase-Activating Protein RanGAP1 between the Cytosol and the Nuclear Pore Complex. *J. Cell Biol.* **1996**, *135* (6 Pt 1), 1457–1470. <https://doi.org/10.1083/jcb.135.6.1457>.
- (32) Saitoh, H.; Hinchev, J. Functional Heterogeneity of Small Ubiquitin-Related Protein Modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **2000**, *275* (9), 6252–6258. <https://doi.org/10.1074/jbc.275.9.6252>.
- (33) Bohren, K. M.; Nadkarni, V.; Song, J. H.; Gabbay, K. H.; Owerbach, D. A M55V Polymorphism in a Novel SUMO Gene (SUMO-4) Differentially Activates Heat Shock

- Transcription Factors and Is Associated with Susceptibility to Type I Diabetes Mellitus. *J. Biol. Chem.* **2004**, 279 (26), 27233–27238. <https://doi.org/10.1074/jbc.M402273200>.
- (34) Owerbach, D.; McKay, E. M.; Yeh, E. T. H.; Gabbay, K. H.; Bohren, K. M. A Proline-90 Residue Unique to SUMO-4 Prevents Maturation and Sumoylation. *Biochem. Biophys. Res. Commun.* **2005**, 337 (2), 517–520. <https://doi.org/10.1016/j.bbrc.2005.09.090>.
- (35) Liang, Y.-C.; Lee, C.-C.; Yao, Y.-L.; Lai, C.-C.; Schmitz, M. L.; Yang, W.-M. SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci. Rep.* **2016**, 6, 26509. <https://doi.org/10.1038/srep26509>.
- (36) Matic, I.; Hagen, M. van; Schimmel, J.; Macek, B.; Ogg, S. C.; Tatham, M. H.; Hay, R. T.; Lamond, A. I.; Mann, M.; Vertegaal, A. C. O. In Vivo Identification of Human Small Ubiquitin-like Modifier Polymerization Sites by High Accuracy Mass Spectrometry and an in Vitro to in Vivo Strategy. *Mol. Cell. Proteomics* **2008**, 7 (1), 132–144. <https://doi.org/10.1074/mcp.M700173-MCP200>.
- (37) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenicke, R.; Becker, J. Structure Determination of the Small Ubiquitin-Related Modifier SUMO-1. *J Mol Biol* **1998**, 280, 275–286. <https://doi.org/10.1006/jmbi.1998.1839>.
- (38) Huang, W.-C.; Ko, T.-P.; Li, S. S.-L.; Wang, A. H.-J. Crystal Structures of the Human SUMO-2 Protein at 1.6 Å and 1.2 Å Resolution: Implication on the Functional Differences of SUMO Proteins. *Eur. J. Biochem.* **2004**, 271 (20), 4114–4122. <https://doi.org/10.1111/j.1432-1033.2004.04349.x>.
- (39) Nicholls, A.; Sharp, K. A.; Honig, B. Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins* **1991**, 11 (4), 281–296. <https://doi.org/10.1002/prot.340110407>.
- (40) Woo, C.-H.; Abe, J. SUMO — A Post-Translational Modification with Therapeutic Potential? *Curr. Opin. Pharmacol.* **2010**, 10 (2), 146–155. <https://doi.org/10.1016/j.coph.2009.12.001>.
- (41) Meluh, P. B.; Koshland, D. Evidence That the MIF2 Gene of *Saccharomyces Cerevisiae* Encodes a Centromere Protein with Homology to the Mammalian Centromere Protein CENP-C. *Mol. Biol. Cell* **1995**, 6 (7), 793–807.
- (42) Okura, T.; Gong, L.; Kamitani, T.; Wada, T.; Okura, I.; Wei, C. F.; Chang, H. M.; Yeh, E. T. Protection against Fas/APO-1- and Tumor Necrosis Factor-Mediated Cell Death by a Novel Protein, Sentrin. *J. Immunol. Baltim. Md 1950* **1996**, 157 (10), 4277–4281.
- (43) Ding, H.; Xu, Y.; Chen, Q.; Dai, H.; Tang, Y.; Wu, J.; Shi, Y. Solution Structure of Human SUMO-3 C47S and Its Binding Surface for Ubc9. *Biochemistry* **2005**, 44 (8), 2790–2799. <https://doi.org/10.1021/bi0477586>.
- (44) Tatham, M. H.; Jaffray, E.; Vaughan, O. A.; Desterro, J. M.; Botting, C. H.; Naismith, J. H.; Hay, R. T. Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **2001**, 276 (38), 35368–35374. <https://doi.org/10.1074/jbc.M104214200>.
- (45) Guo, D.; Li, M.; Zhang, Y.; Yang, P.; Eckenrode, S.; Hopkins, D.; Zheng, W.; Purohit, S.; Podolsky, R. H.; Muir, A.; Wang, J.; Dong, Z.; Brusko, T.; Atkinson, M.; Pozzilli, P.; Zeidler, A.; Raffel, L. J.; Jacob, C. O.; Park, Y.; Serrano-Rios, M.; Larrad, M. T. M.; Zhang, Z.; Garchon, H.-J.; Bach, J.-F.; Rotter, J. I.; She, J.-X.; Wang, C.-Y. A Functional Variant of SUMO4, a New I Kappa B Alpha Modifier, Is Associated with Type 1 Diabetes. *Nat. Genet.* **2004**, 36 (8), 837–841. <https://doi.org/10.1038/ng1391>.
- (46) Wang, C.-Y.; She, J.-X. SUMO4 and Its Role in Type 1 Diabetes Pathogenesis. *Diabetes Metab. Res. Rev.* **2008**, 24 (2), 93–102. <https://doi.org/10.1002/dmrr.797>.
- (47) Wei, W.; Yang, P.; Pang, J.; Zhang, S.; Wang, Y.; Wang, M.-H.; Dong, Z.; She, J.-X.; Wang, C.-Y. A Stress-Dependent SUMO4 Sumoylation of Its Substrate Proteins. *Biochem. Biophys. Res. Commun.* **2008**, 375 (3), 454–459. <https://doi.org/10.1016/j.bbrc.2008.08.028>.
- (48) Guo, D.; Han, J.; Adam, B.-L.; Colburn, N. H.; Wang, M.-H.; Dong, Z.; Eizirik, D. L.; She, J.-X.; Wang, C.-Y. Proteomic Analysis of SUMO4 Substrates in HEK293 Cells

- under Serum Starvation-Induced Stress. *Biochem. Biophys. Res. Commun.* **2005**, 337 (4), 1308–1318. <https://doi.org/10.1016/j.bbrc.2005.09.191>.
- (49) Blomster, H. A.; Imanishi, S. Y.; Siimes, J.; Kastu, J.; Morrice, N. A.; Eriksson, J. E.; Sistonen, L. In Vivo Identification of Sumoylation Sites by a Signature Tag and Cysteine-Targeted Affinity Purification. *J. Biol. Chem.* **2010**, 285 (25), 19324–19329. <https://doi.org/10.1074/jbc.M110.106955>.
- (50) Cooper, H. J.; Tatham, M. H.; Jaffray, E.; Heath, J. K.; Lam, T. T.; Marshall, A. G.; Hay, R. T. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Analysis of Small Ubiquitin-like Modifier (SUMO) Modification: Identification of Lysines in RanBP2 and SUMO Targeted for Modification during the E3 AutoSUMOylation Reaction. *Anal. Chem.* **2005**, 77 (19), 6310–6319. <https://doi.org/10.1021/ac058019d>.
- (51) Schimmel, J.; Larsen, K. M.; Matic, I.; Hagen, M. van; Cox, J.; Mann, M.; Andersen, J. S.; Vertegaal, A. C. O. The Ubiquitin-Proteasome System Is a Key Component of the SUMO-2/3 Cycle. *Mol. Cell. Proteomics* **2008**, 7 (11), 2107–2122. <https://doi.org/10.1074/mcp.M800025-MCP200>.
- (52) Aillet, F.; Lopitz-Otsoa, F.; Egaña, I.; Hjerpe, R.; Fraser, P.; Hay, R. T.; Rodriguez, M. S.; Lang, V. Heterologous SUMO-2/3-Ubiquitin Chains Optimize IκBα Degradation and NF-κB Activity. *PLoS ONE* **2012**, 7 (12). <https://doi.org/10.1371/journal.pone.0051672>.
- (53) Müller, S.; Hoegel, C.; Pyrowolakis, G.; Jentsch, S. SUMO, Ubiquitin's Mysterious Cousin. *Nat. Rev. Mol. Cell Biol.* **2001**, 2 (3), 202–210. <https://doi.org/10.1038/35056591>.
- (54) Gong, L.; Yeh, E. T. H. Characterization of a Family of Nucleolar SUMO-Specific Proteases with Preference for SUMO-2 or SUMO-3. *J. Biol. Chem.* **2006**, 281 (23), 15869–15877. <https://doi.org/10.1074/jbc.M511658200>.
- (55) Hickey, C. M.; Wilson, N. R.; Hochstrasser, M. Function and Regulation of SUMO Proteases. *Nat. Rev. Mol. Cell Biol.* **2012**, 13 (12), 755–766. <https://doi.org/10.1038/nrm3478>.
- (56) Alegre, K. O.; Reverter, D. Swapping Small Ubiquitin-like Modifier (SUMO) Isoform Specificity of SUMO Proteases SENP6 and SENP7. *J. Biol. Chem.* **2011**, 286 (41), 36142–36151. <https://doi.org/10.1074/jbc.M111.268847>.
- (57) Békés, M.; Prudden, J.; Srikumar, T.; Raught, B.; Boddy, M. N.; Salvesen, G. S. The Dynamics and Mechanism of SUMO Chain Deconjugation by SUMO-Specific Proteases. *J. Biol. Chem.* **2011**, 286 (12), 10238–10247. <https://doi.org/10.1074/jbc.M110.205153>.
- (58) Desterro, J. M.; Rodriguez, M. S.; Kemp, G. D.; Hay, R. T. Identification of the Enzyme Required for Activation of the Small Ubiquitin-like Protein SUMO-1. *J. Biol. Chem.* **1999**, 274 (15), 10618–10624. <https://doi.org/10.1074/jbc.274.15.10618>.
- (59) Betting, J.; Seufert, W. A Yeast Ubc9 Mutant Protein with Temperature-Sensitive in Vivo Function Is Subject to Conditional Proteolysis by a Ubiquitin- and Proteasome-Dependent Pathway. *J. Biol. Chem.* **1996**, 271 (42), 25790–25796. <https://doi.org/10.1074/jbc.271.42.25790>.
- (60) Nacerddine, K.; Lehembre, F.; Bhaumik, M.; Artus, J.; Cohen-Tannoudji, M.; Babinet, C.; Pandolfi, P. P.; Dejean, A. The SUMO Pathway Is Essential for Nuclear Integrity and Chromosome Segregation in Mice. *Dev. Cell* **2005**, 9 (6), 769–779. <https://doi.org/10.1016/j.devcel.2005.10.007>.
- (61) Desterro, J. M. P.; Thomson, J.; Hay, R. T. Ubch9 Conjugates SUMO but Not Ubiquitin. *FEBS Lett.* **1997**, 417 (3), 297–300. [https://doi.org/10.1016/S0014-5793\(97\)01305-7](https://doi.org/10.1016/S0014-5793(97)01305-7).
- (62) Lois, L. M.; Lima, C. D. Structures of the SUMO E1 Provide Mechanistic Insights into SUMO Activation and E2 Recruitment to E1. *EMBO J.* **2005**, 24 (3), 439–451. <https://doi.org/10.1038/sj.emboj.7600552>.

- (63) Olsen, S. K.; Capili, A. D.; Lu, X.; Tan, D. S.; Lima, C. D. Active Site Remodelling Accompanies Thioester Bond Formation in the SUMO E1. *Nature* **2010**, *463* (7283), 906–912. <https://doi.org/10.1038/nature08765>.
- (64) Giraud, M. F.; Desterro, J. M.; Naismith, J. H. Structure of Ubiquitin-Conjugating Enzyme 9 Displays Significant Differences with Other Ubiquitin-Conjugating Enzymes Which May Reflect Its Specificity for Sumo Rather than Ubiquitin. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54* (Pt 5), 891–898. <https://doi.org/10.1107/s0907444998002480>.
- (65) Bernier-Villamor, V.; Sampson, D. A.; Matunis, M. J.; Lima, C. D. Structural Basis for E2-Mediated SUMO Conjugation Revealed by a Complex between Ubiquitin-Conjugating Enzyme Ubc9 and RanGAP1. *Cell* **2002**, *108* (3), 345–356. [https://doi.org/10.1016/s0092-8674\(02\)00630-x](https://doi.org/10.1016/s0092-8674(02)00630-x).
- (66) Truong, K.; Lee, T. D.; Li, B.; Chen, Y. Sumoylation of SAE2 C Terminus Regulates SAE Nuclear Localization. *J. Biol. Chem.* **2012**, *287* (51), 42611–42619. <https://doi.org/10.1074/jbc.M112.420877>.
- (67) Tomasi, M. L.; Tomasi, I.; Ramani, K.; Pascale, R. M.; Xu, J.; Giordano, P.; Mato, J. M.; Lu, S. C. S-Adenosyl Methionine Regulates Ubiquitin-Conjugating Enzyme 9 Protein Expression and Sumoylation in Murine Liver and Human Cancers. *Hepatology* **2012**, *56* (3), 982–993. <https://doi.org/10.1002/hep.25701>.
- (68) Su, Y.-F.; Yang, T.; Huang, H.; Liu, L. F.; Hwang, J. Phosphorylation of Ubc9 by Cdk1 Enhances SUMOylation Activity. *PLoS One* **2012**, *7* (4), e34250. <https://doi.org/10.1371/journal.pone.0034250>.
- (69) Reverter, D.; Lima, C. D. Insights into E3 Ligase Activity Revealed by a SUMO-RanGAP1-Ubc9-Nup358 Complex. *Nature* **2005**, *435* (7042), 687–692. <https://doi.org/10.1038/nature03588>.
- (70) Yang, S.; Sharrocks, A. D. The SUMO E3 Ligase Activity of Pc2 Is Coordinated through a SUMO Interaction Motif. *Mol. Cell Biol.* **2010**, *30* (9), 2193–2205. <https://doi.org/10.1128/MCB.01510-09>.
- (71) Gareau, J. R.; Lima, C. D. The SUMO Pathway: Emerging Mechanisms That Shape Specificity, Conjugation and Recognition. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (12), 861–871. <https://doi.org/10.1038/nrm3011>.
- (72) Johnson, E. S.; Gupta, A. A. An E3-like Factor That Promotes SUMO Conjugation to the Yeast Septins. *Cell* **2001**, *106* (6), 735–744. [https://doi.org/10.1016/s0092-8674\(01\)00491-3](https://doi.org/10.1016/s0092-8674(01)00491-3).
- (73) Yeast Ull1/Siz1 Is a Novel SUMO1/Smt3 Ligase for Septin Components and Functions as an Adaptor between Conjugating Enzyme and Substrates <http://www.jbc.org/content/276/52/48973> (accessed Jan 2, 2020).
- (74) Rytinki, M. M.; Kaikkonen, S.; Pehkonen, P.; Jääskeläinen, T.; Palvimo, J. J. PIAS Proteins: Pleiotropic Interactors Associated with SUMO. *Cell. Mol. Life Sci. CMLS* **2009**, *66* (18), 3029–3041. <https://doi.org/10.1007/s00018-009-0061-z>.
- (75) Görlich, D.; Kutay, U. Transport between the Cell Nucleus and the Cytoplasm. *Annu. Rev. Cell Dev. Biol.* **1999**, *15*, 607–660. <https://doi.org/10.1146/annurev.cellbio.15.1.607>.
- (76) Wilken, J.; Hoover, D.; Thompson, D. A.; Barlow, P. N.; McSparron, H.; Picard, L.; Wlodawer, A.; Lubkowski, J.; Kent, S. B. Total Chemical Synthesis and High-Resolution Crystal Structure of the Potent Anti-HIV Protein AOP-RANTES. *Chem. Biol.* **1999**, *6* (1), 43–51. [https://doi.org/10.1016/S1074-5521\(99\)80019-2](https://doi.org/10.1016/S1074-5521(99)80019-2).
- (77) Saitoh, H.; Pu, R. T.; Dasso, M. SUMO-1: Wrestling with a New Ubiquitin-Related Modifier. *Trends Biochem. Sci.* **1997**, *22* (10), 374–376. [https://doi.org/10.1016/s0968-0004\(97\)01102-x](https://doi.org/10.1016/s0968-0004(97)01102-x).
- (78) Wotton, D.; Merrill, J. C. Pc2 and SUMOylation. *Biochem. Soc. Trans.* **2007**, *35* (Pt 6), 1401–1404. <https://doi.org/10.1042/BST0351401>.
- (79) Kagey, M. H.; Melhuish, T. A.; Wotton, D. The Polycomb Protein Pc2 Is a SUMO E3. *Cell* **2003**, *113* (1), 127–137. [https://doi.org/10.1016/s0092-8674\(03\)00159-4](https://doi.org/10.1016/s0092-8674(03)00159-4).

- (80) Merrill, J. C.; Melhuish, T. A.; Kagey, M. H.; Yang, S.-H.; Sharrocks, A. D.; Wotton, D. A Role for Non-Covalent SUMO Interaction Motifs in Pc2/CBX4 E3 Activity. *PLoS One* **2010**, *5* (1), e8794. <https://doi.org/10.1371/journal.pone.0008794>.
- (81) Zhao, J. Sumoylation Regulates Diverse Biological Processes. *Cell. Mol. Life Sci. CMLS* **2007**, *64* (23), 3017–3033. <https://doi.org/10.1007/s00018-007-7137-4>.
- (82) Grégoire, S.; Yang, X.-J. Association with Class IIa Histone Deacetylases Upregulates the Sumoylation of MEF2 Transcription Factors. *Mol. Cell. Biol.* **2005**, *25* (6), 2273. <https://doi.org/10.1128/MCB.25.6.2273-2287.2005>.
- (83) Gao, C.; Ho, C.-C.; Reineke, E.; Lam, M.; Cheng, X.; Stanya, K. J.; Liu, Y.; Chakraborty, S.; Shih, H.-M.; Kao, H.-Y. Histone Deacetylase 7 Promotes PML Sumoylation and Is Essential for PML Nuclear Body Formation. *Mol. Cell. Biol.* **2008**, *28* (18), 5658–5667. <https://doi.org/10.1128/MCB.00874-08>.
- (84) Weger, S.; Hammer, E.; Heilbronn, R. Topors Acts as a SUMO-1 E3 Ligase for P53 in Vitro and in Vivo. *FEBS Lett.* **2005**, *579* (22), 5007–5012. <https://doi.org/10.1016/j.febslet.2005.07.088>.
- (85) Pungaliya, P.; Kulkarni, D.; Park, H.-J.; Marshall, H.; Zheng, H.; Lackland, H.; Saleem, A.; Rubin, E. H. TOPORS Functions As A SUMO-1 E3 Ligase for Chromatin-Modifying Proteins. *J. Proteome Res.* **2007**, *6* (10), 3918–3923. <https://doi.org/10.1021/pr0703674>.
- (86) Mossessova, E.; Lima, C. D. Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Element Essential for Cell Growth in Yeast. *Mol Cell* **2000**, *5*, 865–876.
- (87) Hang, J.; Dasso, M. Association of the human SUMO-1 protease SENP2 with the nuclear pore. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/11896061> (accessed Jan 16, 2020).
- (88) Li, S.-J.; Hochstrasser, M. The Ulp1 SUMO Isopeptidase: Distinct Domains Required for Viability, Nuclear Envelope Localization, and Substrate Specificity. *J. Cell Biol.* **2003**, *160* (7), 1069–1081. <https://doi.org/10.1083/jcb.200212052>.
- (89) Mukhopadhyay, D.; Dasso, M. Modification in Reverse: The SUMO Proteases. *Trends Biochem. Sci.* **2007**, *32* (6), 286–295. <https://doi.org/10.1016/j.tibs.2007.05.002>.
- (90) Reverter, D.; Lima, C. D. A Basis for SUMO Protease Specificity Provided by Analysis of Human Senp2 and a Senp2-SUMO Complex. *Struct. Lond. Engl.* **1993** **2004**, *12* (8), 1519–1531. <https://doi.org/10.1016/j.str.2004.05.023>.
- (91) Xu, Z.; Au, S. W. N. Mapping Residues of SUMO Precursors Essential in Differential Maturation by SUMO-Specific Protease, SENP1. *Biochem. J.* **2005**, *386* (Pt 2), 325–330. <https://doi.org/10.1042/BJ20041210>.
- (92) Perry, J. J. P.; Tainer, J. A.; Boddy, M. N. A SIM-Ultaneous Role for SUMO and Ubiquitin. *Trends Biochem. Sci.* **2008**, *33* (5), 201–208. <https://doi.org/10.1016/j.tibs.2008.02.001>.
- (93) Tatham, M. H.; Geoffroy, M.-C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E. G.; Palvimo, J. J.; Hay, R. T. RNF4 Is a Poly-SUMO-Specific E3 Ubiquitin Ligase Required for Arsenic-Induced PML Degradation. *Nat. Cell Biol.* **2008**, *10* (5), 538–546. <https://doi.org/10.1038/ncb1716>.
- (94) Xie, Y.; Rubenstein, E. M.; Matt, T.; Hochstrasser, M. SUMO-Independent in Vivo Activity of a SUMO-Targeted Ubiquitin Ligase toward a Short-Lived Transcription Factor. *Genes Dev.* **2010**, *24* (9), 893–903. <https://doi.org/10.1101/gad.1906510>.
- (95) Sternsdorf, T.; Jensen, K.; Reich, B.; Will, H. The Nuclear Dot Protein Sp100, Characterization of Domains Necessary for Dimerization, Subcellular Localization, and Modification by Small Ubiquitin-like Modifiers. *J. Biol. Chem.* **1999**, *274* (18), 12555–12566. <https://doi.org/10.1074/jbc.274.18.12555>.
- (96) Ling, Y.; Sankpal, U. T.; Robertson, A. K.; McNally, J. G.; Karpova, T.; Robertson, K. D. Modification of de Novo DNA Methyltransferase 3a (Dnmt3a) by SUMO-1 Modulates Its Interaction with Histone Deacetylases (HDACs) and Its Capacity to

- Repress Transcription. *Nucleic Acids Res.* **2004**, *32* (2), 598–610. <https://doi.org/10.1093/nar/gkh195>.
- (97) Rodriguez, M. S.; Dargemont, C.; Hay, R. T. SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting. *J. Biol. Chem.* **2001**, *276* (16), 12654–12659.
- (98) Pichler, A.; Knipscheer, P.; Oberhofer, E.; van Dijk, W. J.; Körner, R.; Olsen, J. V.; Jentsch, S.; Melchior, F.; Sixma, T. K. SUMO Modification of the Ubiquitin-Conjugating Enzyme E2-25K. *Nat. Struct. Mol. Biol.* **2005**, *12* (3), 264–269. <https://doi.org/10.1038/nsmb903>.
- (99) Sampson, D. A.; Wang, M.; Matunis, M. J. The Small Ubiquitin-like Modifier-1 (SUMO-1) Consensus Sequence Mediates Ubc9 Binding and Is Essential for SUMO-1 Modification. *J. Biol. Chem.* **2001**, *276* (24), 21664–21669. <https://doi.org/10.1074/jbc.M100006200>.
- (100) Matic, I.; Schimmel, J.; Hendriks, I. A.; van Santen, M. A.; van de Rijke, F.; van Dam, H.; Gnad, F.; Mann, M.; Vertegaal, A. C. O. Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif. *Mol. Cell* **2010**, *39* (4), 641–652. <https://doi.org/10.1016/j.molcel.2010.07.026>.
- (101) Spratt, D. E.; Shaw, G. S. Association of the Disordered C-Terminus of CDC34 with a Catalytically Bound Ubiquitin. *J. Mol. Biol.* **2011**, *407* (3), 425–438. <https://doi.org/10.1016/j.jmb.2011.01.047>.
- (102) Grégoire, S.; Tremblay, A. M.; Xiao, L.; Yang, Q.; Ma, K.; Nie, J.; Mao, Z.; Wu, Z.; Giguère, V.; Yang, X.-J. Control of MEF2 Transcriptional Activity by Coordinated Phosphorylation and Sumoylation. *J. Biol. Chem.* **2006**, *281* (7), 4423–4433. <https://doi.org/10.1074/jbc.M509471200>.
- (103) Mohideen, F.; Capili, A. D.; Bilimoria, P. M.; Yamada, T.; Bonni, A.; Lima, C. D. A Molecular Basis for Phosphorylation-Dependent SUMO Conjugation by the E2 Ubc9. *Nat. Struct. Mol. Biol.* **2009**, *16* (9), 945–952. <https://doi.org/10.1038/nsmb.1648>.
- (104) Picard, N.; Caron, V.; Bilodeau, S.; Sanchez, M.; Mascle, X.; Aubry, M.; Tremblay, A. Identification of Estrogen Receptor β as a SUMO-1 Target Reveals a Novel Phosphorylated Sumoylation Motif and Regulation by Glycogen Synthase Kinase 3 β . *Mol. Cell Biol.* **2012**, *32* (14), 2709–2721. <https://doi.org/10.1128/MCB.06624-11>.
- (105) Hietakangas, V.; Anckar, J.; Blomster, H. A.; Fujimoto, M.; Palvimo, J. J.; Nakai, A.; Sistonen, L. PDSM, a Motif for Phosphorylation-Dependent SUMO Modification. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (1), 45–50. <https://doi.org/10.1073/pnas.0503698102>.
- (106) Da Silva-Ferrada, E.; Lopitz-Otsoa, F.; Lang, V.; Rodríguez, M. S.; Matthiesen, R. Strategies to Identify Recognition Signals and Targets of SUMOylation <https://www.hindawi.com/journals/bri/2012/875148/> (accessed Nov 25, 2019). <https://doi.org/10.1155/2012/875148>.
- (107) Knipscheer, P.; Flotho, A.; Klug, H.; Olsen, J. V.; van Dijk, W. J.; Fish, A.; Johnson, E. S.; Mann, M.; Sixma, T. K.; Pichler, A. Ubc9 Sumoylation Regulates SUMO Target Discrimination. *Mol. Cell* **2008**, *31* (3), 371–382. <https://doi.org/10.1016/j.molcel.2008.05.022>.
- (108) Hoege, C.; Pfander, B.; Moldovan, G.-L.; Pyrowolakis, G.; Jentsch, S. RAD6-Dependent DNA Repair Is Linked to Modification of PCNA by Ubiquitin and SUMO. *Nature* **2002**, *419* (6903), 135–141. <https://doi.org/10.1038/nature00991>.
- (109) Yunus, A. A.; Lima, C. D. Structure of the Siz/PIAS SUMO E3 Ligase Siz1 and Determinants Required for SUMO Modification of PCNA. *Mol. Cell* **2009**, *35* (5), 669–682. <https://doi.org/10.1016/j.molcel.2009.07.013>.
- (110) Song, J.; Zhang, Z.; Hu, W.; Chen, Y. Small Ubiquitin-like Modifier (SUMO) Recognition of a SUMO Binding Motif: A Reversal of the Bound Orientation. *J. Biol. Chem.* **2005**, *280* (48), 40122–40129. <https://doi.org/10.1074/jbc.M507059200>.
- (111) Song, J.; Durrin, L. K.; Wilkinson, T. A.; Krontiris, T. G.; Chen, Y. Identification of a SUMO-Binding Motif That Recognizes SUMO-Modified Proteins. *Proc. Natl. Acad.*

- Sci. U. S. A.* **2004**, *101* (40), 14373–14378.
<https://doi.org/10.1073/pnas.0403498101>.
- (112) Deshaies, R. J.; Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annu. Rev. Biochem.* **2009**, *78*, 399–434.
<https://doi.org/10.1146/annurev.biochem.78.101807.093809>.
- (113) Hecker, C.-M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. Specification of SUMO1- and SUMO2-Interacting Motifs. *J. Biol. Chem.* **2006**, *281* (23), 16117–16127.
<https://doi.org/10.1074/jbc.M512757200>.
- (114) Liu, Z.; Gao, X.; Cao, J.; Ma, Q. GPS-CCD: A Novel Computational Program for the Prediction of Calpain Cleavage Sites
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019001> (accessed Jan 2, 2020).
- (115) Sun, H.; Levenson, J. D.; Hunter, T. Conserved Function of RNF4 Family Proteins in Eukaryotes: Targeting a Ubiquitin Ligase to SUMOylated Proteins. *EMBO J.* **2007**, *26* (18), 4102–4112. <https://doi.org/10.1038/sj.emboj.7601839>.
- (116) Vertegaal, A. C. O.; Andersen, J. S.; Ogg, S. C.; Hay, R. T.; Mann, M.; Lamond, A. I. Distinct and Overlapping Sets of SUMO-1 and SUMO-2 Target Proteins Revealed by Quantitative Proteomics. *Mol. Cell. Proteomics* **2006**, *5* (12), 2298–2310.
<https://doi.org/10.1074/mcp.M600212-MCP200>.
- (117) Ullmann, R.; Chien, C. D.; Avantaggiati, M. L.; Muller, S. An Acetylation Switch Regulates SUMO-Dependent Protein Interaction Networks. *Mol. Cell* **2012**, *46* (6), 759–770. <https://doi.org/10.1016/j.molcel.2012.04.006>.
- (118) Tomasi, M. L.; Ramani, K.; Ryoo, M. Ubiquitin-Conjugating Enzyme 9 Phosphorylation as a Novel Mechanism for Potentiation of the Inflammatory Response. *Am. J. Pathol.* **2016**, *186* (9), 2326–2336.
<https://doi.org/10.1016/j.ajpath.2016.05.007>.
- (119) Hsieh, Y.-L.; Kuo, H.-Y.; Chang, C.-C.; Naik, M. T.; Liao, P.-H.; Ho, C.-C.; Huang, T.-C.; Jeng, J.-C.; Hsu, P.-H.; Tsai, M.-D.; Huang, T.-H.; Shih, H.-M. Ubc9 Acetylation Modulates Distinct SUMO Target Modification and Hypoxia Response. *EMBO J.* **2013**, *32* (6), 791–804. <https://doi.org/10.1038/emboj.2013.5>.
- (120) Gill, G. SUMO and Ubiquitin in the Nucleus: Different Functions, Similar Mechanisms? *Genes Dev.* **2004**, *18* (17), 2046–2059.
<https://doi.org/10.1101/gad.1214604>.
- (121) Desterro, J. M.; Rodriguez, M. S.; Hay, R. T. SUMO-1 Modification of I κ B α Inhibits NF- κ B Activation. *Mol. Cell* **1998**, *2* (2), 233–239.
[https://doi.org/10.1016/s1097-2765\(00\)80133-1](https://doi.org/10.1016/s1097-2765(00)80133-1).
- (122) Lee, P. S. W.; Chang, C.; Liu, D.; Derynck, R. Sumoylation of Smad4, the Common Smad Mediator of Transforming Growth Factor- β Family Signaling. *J. Biol. Chem.* **2003**, *278* (30), 27853–27863. <https://doi.org/10.1074/jbc.M301755200>.
- (123) Steffan, J. S.; Agrawal, N.; Pallos, J.; Rockabrand, E.; Trotman, L. C.; Slepko, N.; Illes, K.; Lukacsovich, T.; Zhu, Y.-Z.; Cattaneo, E.; Pandolfi, P. P.; Thompson, L. M.; Marsh, J. L. SUMO Modification of Huntingtin and Huntington's Disease Pathology. *Science* **2004**, *304* (5667), 100–104. <https://doi.org/10.1126/science.1092194>.
- (124) Perkins, N. D. Post-Translational Modifications Regulating the Activity and Function of the Nuclear Factor Kappa B Pathway. *Oncogene* **2006**, *25* (51), 6717–6730.
<https://doi.org/10.1038/sj.onc.1209937>.
- (125) Bellail, A. C.; Olson, J. J.; Hao, C. SUMO1 Modification Stabilizes CDK6 Protein and Drives the Cell Cycle and Glioblastoma Progression. *Nat. Commun.* **2014**, *5*, 4234.
<https://doi.org/10.1038/ncomms5234>.
- (126) Prakash, S.; Sung, P.; Prakash, L. DNA Repair Genes and Proteins of *Saccharomyces Cerevisiae*. *Annu. Rev. Genet.* **1993**, *27*, 33–70.
<https://doi.org/10.1146/annurev.ge.27.120193.000341>.
- (127) Watts, F. Z. Sumoylation of PCNA: Wrestling with Recombination at Stalled Replication Forks. *DNA Repair* **2006**, *5* (3), 399–403.
<https://doi.org/10.1016/j.dnarep.2005.11.002>.

- (128) Geoffroy, M.-C.; Hay, R. T. An Additional Role for SUMO in Ubiquitin-Mediated Proteolysis. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (8), 564–568. <https://doi.org/10.1038/nrm2707>.
- (129) Hietakangas, V.; Ahlskog, J. K.; Jakobsson, A. M.; Hellesuo, M.; Sahlberg, N. M.; Holmberg, C. I.; Mikhailov, A.; Palvimo, J. J.; Pirkkala, L.; Sistonen, L. Phosphorylation of Serine 303 Is a Prerequisite for the Stress-Inducible SUMO Modification of Heat Shock Factor 1. *Mol. Cell. Biol.* **2003**, *23* (8), 2953–2968. <https://doi.org/10.1128/MCB.23.8.2953-2968.2003>.
- (130) Gresko, E.; Möller, A.; Roscic, A.; Schmitz, M. L. Covalent Modification of Human Homeodomain Interacting Protein Kinase 2 by SUMO-1 at Lysine 25 Affects Its Stability. *Biochem. Biophys. Res. Commun.* **2005**, *329* (4), 1293–1299. <https://doi.org/10.1016/j.bbrc.2005.02.113>.
- (131) Hofmann, T. G.; Jaffray, E.; Stollberg, N.; Hay, R. T.; Will, H. Regulation of Homeodomain-Interacting Protein Kinase 2 (HIPK2) Effector Function through Dynamic Small Ubiquitin-Related Modifier-1 (SUMO-1) Modification. *J. Biol. Chem.* **2005**, *280* (32), 29224–29232. <https://doi.org/10.1074/jbc.M503921200>.
- (132) Chang, C.-C.; Naik, M. T.; Huang, Y.-S.; Jeng, J.-C.; Liao, P.-H.; Kuo, H.-Y.; Ho, C.-C.; Hsieh, Y.-L.; Lin, C.-H.; Huang, N.-J.; Naik, N. M.; Kung, C. C.-H.; Lin, S.-Y.; Chen, R.-H.; Chang, K.-S.; Huang, T.-H.; Shih, H.-M. Structural and Functional Roles of Daxx SIM Phosphorylation in SUMO Paralog-Selective Binding and Apoptosis Modulation. *Mol. Cell* **2011**, *42* (1), 62–74. <https://doi.org/10.1016/j.molcel.2011.02.022>.
- (133) Barlev, N. A.; Liu, L.; Chehab, N. H.; Mansfield, K.; Harris, K. G.; Halazonetis, T. D.; Berger, S. L. Acetylation of P53 Activates Transcription through Recruitment of Coactivators/Histone Acetyltransferases. *Mol. Cell* **2001**, *8* (6), 1243–1254. [https://doi.org/10.1016/s1097-2765\(01\)00414-2](https://doi.org/10.1016/s1097-2765(01)00414-2).
- (134) Liang, S. H.; Clarke, M. F. Regulation of P53 Localization. *Eur. J. Biochem.* **2001**, *268* (10), 2779–2783. <https://doi.org/10.1046/j.1432-1327.2001.02227.x>.
- (135) Appella, E.; Anderson, C. W. Post-Translational Modifications and Activation of P53 by Genotoxic Stresses. *Eur. J. Biochem.* **2001**, *268* (10), 2764–2772. <https://doi.org/10.1046/j.1432-1327.2001.02225.x>.
- (136) Gu, W.; Roeder, R. G. Activation of P53 Sequence-Specific DNA Binding by Acetylation of the P53 C-Terminal Domain. *Cell* **1997**, *90* (4), 595–606. [https://doi.org/10.1016/S0092-8674\(00\)80521-8](https://doi.org/10.1016/S0092-8674(00)80521-8).
- (137) Wu, S.-Y.; Chiang, C.-M. P53 Sumoylation. *Epigenetics* **2009**, *4* (7), 445–451.
- (138) Dai, C.; Gu, W. P53 Post-Translational Modification: Deregulated in Tumorigenesis. *Trends Mol. Med.* **2010**, *16* (11), 528–536. <https://doi.org/10.1016/j.molmed.2010.09.002>.
- (139) Stankovic-Valentin, N.; Deltour, S.; Seeler, J.; Pinte, S.; Vergoten, G.; Guérardel, C.; Dejean, A.; Leprince, D. An Acetylation/Deacetylation-SUMOylation Switch through a Phylogenetically Conserved PsiKXEP Motif in the Tumor Suppressor HIC1 Regulates Transcriptional Repression Activity. *Mol. Cell. Biol.* **2007**, *27* (7), 2661–2675. <https://doi.org/10.1128/MCB.01098-06>.
- (140) Dehennaut, V.; Loison, I.; Dubuissez, M.; Nassour, J.; Abbadie, C.; Leprince, D. DNA Double-Strand Breaks Lead to Activation of Hypermethylated in Cancer 1 (HIC1) by SUMOylation to Regulate DNA Repair. *J. Biol. Chem.* **2013**, *288* (15), 10254–10264. <https://doi.org/10.1074/jbc.M112.421610>.
- (141) Van Rechem, C.; Boulay, G.; Pinte, S.; Stankovic-Valentin, N.; Guérardel, C.; Leprince, D. Differential Regulation of HIC1 Target Genes by CtBP and NuRD, via an Acetylation/SUMOylation Switch, in Quiescent versus Proliferating Cells. *Mol. Cell. Biol.* **2010**, *30* (16), 4045–4059. <https://doi.org/10.1128/MCB.00582-09>.
- (142) Bergink, S.; Jentsch, S. Principles of Ubiquitin and SUMO Modifications in DNA Repair. *Nature* **2009**, *458*, 461–467. <https://doi.org/10.1038/nature07963>.

- (143) Sarangi, P.; Zhao, X. SUMO-Mediated Regulation of DNA Damage Repair and Responses. *Trends Biochem. Sci.* **2015**, *40* (4), 233–242. <https://doi.org/10.1016/j.tibs.2015.02.006>.
- (144) Seeler, J.-S.; Dejean, A. SUMO and the Robustness of Cancer. *Nat. Rev. Cancer* **2017**, *17* (3), 184–197. <https://doi.org/10.1038/nrc.2016.143>.
- (145) Han, Z.-J.; Feng, Y.-H.; Gu, B.-H.; Li, Y.-M.; Chen, H. The Post-Translational Modification, SUMOylation, and Cancer (Review). *Int. J. Oncol.* **2018**, *52* (4), 1081–1094. <https://doi.org/10.3892/ijo.2018.4280>.
- (146) Villalva, C.; Trempat, P.; Greenland, C.; Thomas, C.; Girard, J. P.; Moebius, F.; Delsol, G.; Brousset, P. Isolation of Differentially Expressed Genes in NPM-ALK-Positive Anaplastic Large Cell Lymphoma. *Br. J. Haematol.* **2002**, *118* (3), 791–798. <https://doi.org/10.1046/j.1365-2141.2002.03671.x>.
- (147) Zhang, H.; Kuai, X.; Ji, Z.; Li, Z.; Shi, R. Over-Expression of Small Ubiquitin-Related Modifier-1 and Sumoylated P53 in Colon Cancer. *Cell Biochem. Biophys.* **2013**, *67* (3), 1081–1087. <https://doi.org/10.1007/s12013-013-9612-x>.
- (148) Lee, J.-S.; Thorgeirsson, S. S. Genome-Scale Profiling of Gene Expression in Hepatocellular Carcinoma: Classification, Survival Prediction, and Identification of Therapeutic Targets. *Gastroenterology* **2004**, *127* (5 Suppl 1), S51-55. <https://doi.org/10.1053/j.gastro.2004.09.015>.
- (149) Mo, Y.; Yu, Y.; Theodosiou, E. A role for Ubc9 in tumorigenesis. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/15735760> (accessed Jan 3, 2020).
- (150) McDoniels-Silvers, A. L. Differential Gene Expression in Human Lung Adenocarcinomas and Squamous Cell Carcinomas | Clinical Cancer Research <https://clincancerres.aacrjournals.org/content/8/4/1127> (accessed Jan 3, 2020).
- (151) Geletu, M. Target proteins of C/EBPalphap30 in AML: C/EBPalphap30 enhances sumoylation of C/EBPalphap42 via up-regulation of Ubc9. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/17671234> (accessed Jan 3, 2020).
- (152) Moschos, S. J.; Smith, A. P.; Mandic, M.; Athanassiou, C.; Watson-Hurst, K.; Jukic, D. M.; Edington, H. D.; Kirkwood, J. M.; Becker, D. SAGE and Antibody Array Analysis of Melanoma-Infiltrated Lymph Nodes: Identification of Ubc9 as an Important Molecule in Advanced-Stage Melanomas. *Oncogene* **2007**, *26* (29), 4216–4225. <https://doi.org/10.1038/sj.onc.1210216>.
- (153) Moschos, S. J.; Jukic, D. M.; Athanassiou, C.; Bhargava, R.; Dacic, S.; Wang, X.; Kuan, S.-F.; Fayewicz, S. L.; Galambos, C.; Acquafondata, M.; Dhir, R.; Becker, D. Expression Analysis of Ubc9, the Single Small Ubiquitin-like Modifier (SUMO) E2 Conjugating Enzyme, in Normal and Malignant Tissues. *Hum. Pathol.* **2010**, *41* (9), 1286–1298. <https://doi.org/10.1016/j.humpath.2010.02.007>.
- (154) Ueda, M. DNA microarray analysis of stage progression mechanism in myelodysplastic syndrome. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/14531911> (accessed Jan 3, 2020).
- (155) Jacques, C. Two-step differential expression analysis reveals a new set of genes involved in thyroid oncocytic tumors. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/15623817> (accessed Jan 3, 2020).
- (156) Cheng, J.; Bawa, T.; Lee, P.; Gong, L.; Yeh, E. T. H. Role of Desumoylation in the Development of Prostate Cancer. *Neoplasia N. Y. N* **2006**, *8* (8), 667–676. <https://doi.org/10.1593/neo.06445>.
- (157) *SUMO Regulation of Cellular Processes*, 2nd ed.; Wilson, V., Ed.; Advances in Experimental Medicine and Biology; Springer International Publishing, 2017. <https://doi.org/10.1007/978-3-319-50044-7>.
- (158) de Thé, H.; Lavau, C.; Marchio, A.; Chomienne, C.; Degos, L.; Dejean, A. The PML-RAR Alpha Fusion MRNA Generated by the t(15;17) Translocation in Acute Promyelocytic Leukemia Encodes a Functionally Altered RAR. *Cell* **1991**, *66* (4), 675–684. [https://doi.org/10.1016/0092-8674\(91\)90113-d](https://doi.org/10.1016/0092-8674(91)90113-d).
- (159) Brown, R. S.; Aman, A. Intramolecular Catalysis of Thiol Ester Hydrolysis by a Tertiary Amine and a Carboxylate. *J Org Chem* **1997**, *62*, 4816–4820.

- (160) Grisolano, J. L.; Wesselschmidt, R. L.; Pelicci, P. G.; Ley, T. J. Altered Myeloid Development and Acute Leukemia in Transgenic Mice Expressing PML-RAR Alpha under Control of Cathepsin G Regulatory Sequences. *Blood* **1997**, *89* (2), 376–387.
- (161) Dos Santos, G. A.; Kats, L.; Pandolfi, P. P. Synergy against PML-RAR α : Targeting Transcription, Proteolysis, Differentiation, and Self-Renewal in Acute Promyelocytic Leukemia. *J. Exp. Med.* **2013**, *210* (13), 2793–2802. <https://doi.org/10.1084/jem.20131121>.
- (162) Lallemand-Breitenbach, V.; Jeanne, M.; Benhenda, S.; Nasr, R.; Lei, M.; Peres, L.; Zhou, J.; Zhu, J.; Raught, B.; de Thé, H. Arsenic Degrades PML or PML-RAR α through a SUMO-Triggered RNF4/Ubiquitin-Mediated Pathway. *Nat. Cell Biol.* **2008**, *10* (5), 547–555. <https://doi.org/10.1038/ncb1717>.
- (163) Wang, Q.; Xia, N.; Li, T.; Xu, Y.; Zou, Y.; Zuo, Y.; Fan, Q.; Bawa-Khalife, T.; Yeh, E. T. H.; Cheng, J. SUMO-Specific Protease 1 Promotes Prostate Cancer Progression and Metastasis. *Oncogene* **2013**, *32* (19), 2493–2498. <https://doi.org/10.1038/onc.2012.250>.
- (164) Cheng, J. SENP1 enhances androgen receptor-dependent transcription through desumoylation of histone deacetylase 1. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/15199155> (accessed Jan 3, 2020).
- (165) Kim, K. I.; Baek, S. H. SUMOylation Code in Cancer Development and Metastasis. *Mol. Cells* **2006**, *22* (3), 247–253.
- (166) Santti, H.; Mikkonen, L.; Anand, A.; Hirvonen-Santti, S.; Toppari, J.; Panhuysen, M.; Vauti, F.; Perera, M.; Corte, G.; Wurst, W.; Jänne, O. A.; Palvimo, J. J. Disruption of the Murine PIASx Gene Results in Reduced Testis Weight. *J. Mol. Endocrinol.* **2005**, *34* (3), 645–654. <https://doi.org/10.1677/jme.1.01666>.
- (167) Puhr, M.; Hoefler, J.; Neuwirt, H.; Eder, I. E.; Kern, J.; Schäfer, G.; Geley, S.; Heidegger, I.; Klocker, H.; Culig, Z. PIAS1 Is a Crucial Factor for Prostate Cancer Cell Survival and a Valid Target in Docetaxel Resistant Cells. *Oncotarget* **2014**, *5* (23), 12043–12056.
- (168) Song, M.; Salmena, L.; Pandolfi, P. P. The functions and regulation of the PTEN tumour suppressor. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/22473468> (accessed Jan 3, 2020).
- (169) Zhou, W.; Ryan, J. J.; Zhou, H. Global Analyses of Sumoylated Proteins in *Saccharomyces Cerevisiae*: Induction of Protein Sumoylation by Cellular Stresses. *J. Biol. Chem.* **2004**, *279* (31), 32262–32268. <https://doi.org/10.1074/jbc.M404173200>.
- (170) Tempé, D.; Piechaczyk, M.; Bossis, G. SUMO under Stress. *Biochem. Soc. Trans.* **2008**, *36* (Pt 5), 874–878. <https://doi.org/10.1042/BST0360874>.
- (171) Hong, Y.; Rogers, R.; Matunis, M. J.; Mayhew, C. N.; Goodson, M. L.; Park-Sarge, O. K.; Sarge, K. D.; Goodson, M. Regulation of Heat Shock Transcription Factor 1 by Stress-Induced SUMO-1 Modification. *J. Biol. Chem.* **2001**, *276* (43), 40263–40267. <https://doi.org/10.1074/jbc.M104714200>.
- (172) Nefkens, I.; Negorev, D. G.; Ishov, A. M.; Michaelson, J. S.; Yeh, E. T. H.; Tanguay, R. M.; Müller, W. E. G.; Maul, G. G. Heat Shock and Cd²⁺ Exposure Regulate PML and Daxx Release from ND10 by Independent Mechanisms That Modify the Induction of Heat-Shock Proteins 70 and 25 Differently. *J. Cell Sci.* **2003**, *116* (Pt 3), 513–524. <https://doi.org/10.1242/jcs.00253>.
- (173) Bossis, G.; Malnou, C. E.; Farras, R.; Andermarcher, E.; Hipskind, R.; Rodriguez, M.; Schmidt, D.; Muller, S.; Jariel-Encontre, I.; Piechaczyk, M. Down-Regulation of c-Fos/c-Jun AP-1 Dimer Activity by Sumoylation. *Mol. Cell. Biol.* **2005**, *25* (16), 6964–6979. <https://doi.org/10.1128/MCB.25.16.6964-6979.2005>.
- (174) Carbia-Nagashima, A.; Gerez, J.; Perez-Castro, C.; Paez-Pereda, M.; Silberstein, S.; Stalla, G. K.; Holsboer, F.; Arzt, E. RSUME, a Small RWD-Containing Protein, Enhances SUMO Conjugation and Stabilizes HIF-1 α during Hypoxia. *Cell* **2007**, *131* (2), 309–323. <https://doi.org/10.1016/j.cell.2007.07.044>.

- (175) Bossis, G.; Melchior, F. Regulation of SUMOylation by Reversible Oxidation of SUMO Conjugating Enzymes. *Mol. Cell* **2006**, *21* (3), 349–357. <https://doi.org/10.1016/j.molcel.2005.12.019>.
- (176) Shao, R.; Zhang, F.-P.; Tian, F.; Anders Friberg, P.; Wang, X.; Sjöland, H.; Billig, H. Increase of SUMO-1 Expression in Response to Hypoxia: Direct Interaction with HIF-1alpha in Adult Mouse Brain and Heart in Vivo. *FEBS Lett.* **2004**, *569* (1–3), 293–300. <https://doi.org/10.1016/j.febslet.2004.05.079>.
- (177) Hetz, C. The Unfolded Protein Response: Controlling Cell Fate Decisions under ER Stress and Beyond. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (2), 89–102. <https://doi.org/10.1038/nrm3270>.
- (178) Chen, H.; Qi, L. SUMO Modification Regulates the Transcriptional Activity of XBP1. *Biochem. J.* **2010**, *429* (1), 95–102. <https://doi.org/10.1042/BJ20100193>.
- (179) Uemura, A.; Taniguchi, M.; Matsuo, Y.; Oku, M.; Wakabayashi, S.; Yoshida, H. UBC9 Regulates the Stability of XBP1, a Key Transcription Factor Controlling the ER Stress Response. *Cell Struct. Funct.* **2013**, *38* (1), 67–79. <https://doi.org/10.1247/csf.12026>.
- (180) Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007**, *128* (4), 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>.
- (181) Shiio, Y.; Eisenman, R. N. Histone Sumoylation Is Associated with Transcriptional Repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230. <https://doi.org/10.1073/pnas.1735528100>.
- (182) Nathan, D.; Ingvarsdottir, K.; Sterner, D. E.; Bylebyl, G. R.; Dokmanovic, M.; Dorsey, J. A.; Whelan, K. A.; Krsmanovic, M.; Lane, W. S.; Meluh, P. B.; Johnson, E. S.; Berger, S. L. Histone Sumoylation Is a Negative Regulator in *Saccharomyces Cerevisiae* and Shows Dynamic Interplay with Positive-Acting Histone Modifications. *Genes Dev.* **2006**, *20* (8), 966–976. <https://doi.org/10.1101/gad.1404206>.
- (183) Lyst, M. J.; Stancheva, I. A Role for SUMO Modification in Transcriptional Repression and Activation. *Biochem. Soc. Trans.* **2007**, *35* (Pt 6), 1389–1392. <https://doi.org/10.1042/BST0351389>.
- (184) Neyret-Kahn, H.; Benhamed, M.; Ye, T.; Le Gras, S.; Cossec, J.-C.; Lapaquette, P.; Bischof, O.; Ouspenskaia, M.; Dasso, M.; Seeler, J.; Davidson, I.; Dejean, A. Sumoylation at Chromatin Governs Coordinated Repression of a Transcriptional Program Essential for Cell Growth and Proliferation. *Genome Res.* **2013**, *23* (10), 1563–1579. <https://doi.org/10.1101/gr.154872.113>.
- (185) Liu, H.; Zhang, J.; Heine, G. F.; Arora, M.; Gulcin Ozer, H.; Onti-Srinivasan, R.; Huang, K.; Parvin, J. D. Chromatin Modification by SUMO-1 Stimulates the Promoters of Translation Machinery Genes. *Nucleic Acids Res.* **2012**, *40* (20), 10172–10186. <https://doi.org/10.1093/nar/gks819>.
- (186) Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E. Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science* **1998**, *279* (5349), 349–352. <https://doi.org/10.1126/science.279.5349.349>.
- (187) Shay, J. W.; Bacchetti, S. A Survey of Telomerase Activity in Human Cancer. *Eur. J. Cancer Oxf. Engl. 1990* **1997**, *33* (5), 787–791. [https://doi.org/10.1016/S0959-8049\(97\)00062-2](https://doi.org/10.1016/S0959-8049(97)00062-2).
- (188) Xhemalce, B.; Riising, E. M.; Baumann, P.; Dejean, A.; Arcangioli, B.; Seeler, J.-S. Role of SUMO in the Dynamics of Telomere Maintenance in Fission Yeast. *Proc. Natl. Acad. Sci.* **2007**, *104* (3), 893–898. <https://doi.org/10.1073/pnas.0605442104>.
- (189) Hang, L. SUMOylation regulates telomere length homeostasis by targeting Cdc13. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/21743457> (accessed Jan 3, 2020).
- (190) Pebernard, S. Localization of Smc5/6 to centromeres and telomeres requires heterochromatin and SUMO, respectively <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2585169/> (accessed Nov 26, 2019).

- (191) Sacher, M.; Pfander, B.; Hoege, C.; Jentsch, S. Control of Rad52 Recombination Activity by Double-Strand Break-Induced SUMO Modification. *Nat. Cell Biol.* **2006**, *8* (11), 1284–1290. <https://doi.org/10.1038/ncb1488>.
- (192) Shinbo, Y.; Niki, T.; Taira, T.; Ooe, H.; Takahashi-Niki, K.; Maita, C.; Seino, C.; Iguchi-Arigo, S. M. M.; Ariga, H. Proper SUMO-1 Conjugation Is Essential to DJ-1 to Exert Its Full Activities. *Cell Death Differ.* **2006**, *13* (1), 96–108. <https://doi.org/10.1038/sj.cdd.4401704>.
- (193) Wang, Q.-E.; Zhu, Q.; Wani, G.; El-Mahdy, M. A.; Li, J.; Wani, A. A. DNA Repair Factor XPC Is Modified by SUMO-1 and Ubiquitin Following UV Irradiation. *Nucleic Acids Res.* **2005**, *33* (13), 4023–4034. <https://doi.org/10.1093/nar/gki684>.
- (194) Cheng, Z.; Ke, Y.; Ding, X.; Wang, F.; Wang, H.; Ahmed, K.; Liu, Z.; Xu, Y.; Aikhionbare, F.; Yan, H.; Liu, J.; Xue, Y.; Powell, M.; Liang, S.; Reddy, S. E.; Hu, R.; Huang, H.; Jin, C.; Yao, X. Functional Characterization of TIP60 Sumoylation in UV-Irradiated DNA Damage Response. *Oncogene* **2008**, *27* (7), 931–941. <https://doi.org/10.1038/sj.onc.1210710>.
- (195) Ivanov, A. V.; Peng, H.; Yurchenko, V.; Yap, K. L.; Negorev, D. G.; Schultz, D. C.; Psulkowski, E.; Fredericks, W. J.; White, D. E.; Maul, G. G.; Sadofsky, M. J.; Zhou, M.-M.; Rauscher, F. J. PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. *Mol. Cell* **2007**, *28* (5), 823–837. <https://doi.org/10.1016/j.molcel.2007.11.012>.
- (196) Li, X.; Lee, Y.-K.; Jeng, J.-C.; Yen, Y.; Schultz, D. C.; Shih, H.-M.; Ann, D. K. Role for KAP1 Serine 824 Phosphorylation and Sumoylation/Desumoylation Switch in Regulating KAP1-Mediated Transcriptional Repression. *J. Biol. Chem.* **2007**, *282* (50), 36177–36189. <https://doi.org/10.1074/jbc.M706912200>.
- (197) Lin, J.-Y.; Ohshima, T.; Shimotohno, K. Association of Ubc9, an E2 Ligase for SUMO Conjugation, with P53 Is Regulated by Phosphorylation of P53. *FEBS Lett.* **2004**, *573* (1), 15–18. <https://doi.org/10.1016/j.febslet.2004.07.059>.
- (198) Stratton, M. R.; Campbell, P. J.; Futreal, P. A. The Cancer Genome. *Nature* **2009**, *458* (7239), 719–724. <https://doi.org/10.1038/nature07943>.
- (199) Gali, H.; Juhasz, S.; Morocz, M.; Hajdu, I.; Fatyol, K.; Szukacsov, V.; Burkovics, P.; Haracska, L. Role of SUMO Modification of Human PCNA at Stalled Replication Fork. *Nucleic Acids Res.* **2012**, *40* (13), 6049–6059. <https://doi.org/10.1093/nar/gks256>.
- (200) McLaughlin, D.; Coey, C. T.; Yang, W.-C.; Drohat, A. C.; Matunis, M. J. Characterizing Requirements for SUMO Modification and Binding on Base Excision Repair Activity of Thymine DNA Glycosylase in Vivo. *J. Biol. Chem.* **2016**, *jbc.M115.706325*. <https://doi.org/10.1074/jbc.M115.706325>.
- (201) Mantovani, A.; Allavena, P.; Sica, A.; Balkwill, F. Cancer-Related Inflammation. *Nature* **2008**, *454* (7203), 436–444. <https://doi.org/10.1038/nature07205>.
- (202) Grivennikov, S.; Karin, E.; Terzic, J.; Mucida, D.; Yu, G.-Y.; Vallabhapurapu, S.; Scheller, J.; Rose-John, S.; Cheroutre, H.; Eckmann, L.; Karin, M. IL-6 and Stat3 Are Required for Survival of Intestinal Epithelial Cells and Development of Colitis-Associated Cancer. *Cancer Cell* **2009**, *15* (2), 103–113. <https://doi.org/10.1016/j.ccr.2009.01.001>.
- (203) Huang, T. T.; Wuerzberger-Davis, S. M.; Wu, Z.-H.; Miyamoto, S. Sequential Modification of NEMO/IKKgamma by SUMO-1 and Ubiquitin Mediates NF-KappaB Activation by Genotoxic Stress. *Cell* **2003**, *115* (5), 565–576. [https://doi.org/10.1016/s0092-8674\(03\)00895-x](https://doi.org/10.1016/s0092-8674(03)00895-x).
- (204) Decque, A.; Joffre, O.; Magalhaes, J. G.; Cossec, J.-C.; Blecher-Gonen, R.; Lapaquette, P.; Silvin, A.; Manel, N.; Joubert, P.-E.; Seeler, J.-S.; Albert, M. L.; Amit, I.; Amigorena, S.; Dejean, A. Sumoylation Coordinates the Repression of Inflammatory and Anti-Viral Gene-Expression Programs during Innate Sensing. *Nat. Immunol.* **2016**, *17* (2), 140–149. <https://doi.org/10.1038/ni.3342>.
- (205) Pascual, G.; Fong, A. L.; Ogawa, S.; Gamliel, A.; Li, A. C.; Perissi, V.; Rose, D. W.; Willson, T. M.; Rosenfeld, M. G.; Glass, C. K. A SUMOylation-Dependent Pathway

- Mediates Transrepression of Inflammatory Response Genes by PPAR-Gamma. *Nature* **2005**, *437* (7059), 759–763. <https://doi.org/10.1038/nature03988>.
- (206) Ghisletti, S.; Barozzi, I.; Mietton, F.; Polletti, S.; De Santa, F.; Venturini, E.; Gregory, L.; Lonie, L.; Chew, A.; Wei, C.-L.; Ragoussis, J.; Natoli, G. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. *Immunity* **2010**, *32* (3), 317–328. <https://doi.org/10.1016/j.immuni.2010.02.008>.
- (207) Bawa-Khalfe, T.; Yeh, E. T. H. SUMO Losing Balance: SUMO Proteases Disrupt SUMO Homeostasis to Facilitate Cancer Development and Progression. *Genes Cancer* **2010**, *1* (7), 748–752. <https://doi.org/10.1177/1947601910382555>.
- (208) Agouridas, V.; El Mahdi, O.; Cargoët, M.; Melyk, O. A Statistical View of Protein Chemical Synthesis Using NCL and Extended Methodologies. *Bioorg. Med. Chem.* **2017**, *25* (18), 4938–4945. <https://doi.org/10.1016/j.bmc.2017.05.050>.
- (209) Merrifield, R. B. Automated Synthesis of Peptides. *Science* **1965**, *150* (3693), 178–185. <https://doi.org/10.1126/science.150.3693.178>.
- (210) Sheppard, R. The Fluorenylmethoxycarbonyl Group in Solid Phase Synthesis. *J. Pept. Sci.* **2003**, *9* (9), 545–552. <https://doi.org/10.1002/psc.479>.
- (211) Fields, G. B. Methods for Removing the Fmoc Group. *Methods Mol Biol* **1994**, *35*, 17–27.
- (212) Meldal, M. [6] Properties of Solid Supports. In *Methods in Enzymology*; Solid-Phase Peptide Synthesis; Academic Press, 1997; Vol. 289, pp 83–104. [https://doi.org/10.1016/S0076-6879\(97\)89045-3](https://doi.org/10.1016/S0076-6879(97)89045-3).
- (213) Han, S.-Y.; Kim, Y.-A. Recent Development of Peptide Coupling Reagents in Organic Synthesis. *Tetrahedron* **2004**, *60* (11), 2447–2467. <https://doi.org/10.1016/j.tet.2004.01.020>.
- (214) Albericio, F. Orthogonal Protecting Groups for N(Alpha)-Amino and C-Terminal Carboxyl Functions in Solid-Phase Peptide Synthesis. *Biopolymers* **2000**, *55* (2), 123–139. [https://doi.org/10.1002/1097-0282\(2000\)55:2<123::AID-BIP30>3.0.CO;2-F](https://doi.org/10.1002/1097-0282(2000)55:2<123::AID-BIP30>3.0.CO;2-F).
- (215) Shimko, J. C.; North, J. A.; Bruns, A. N.; Poirier, M. G.; Ottesen, J. J. Preparation of Fully Synthetic Histone H3 Reveals That Acetyl-Lysine 56 Facilitates Protein Binding within Nucleosomes. *J Mol Biol* **2011**, *408*, 187–204.
- (216) Schnolzer, M.; Kent, S. B. H. Constructing Proteins by Dovetailing Unprotected Synthetic Peptides: Backbone-Engineered HIV Protease. *Science* **1992**, *256*, 221–225.
- (217) Tam, J. P.; Xu, J.; Eom, K. D. Methods and Strategies of Peptide Ligation. *Biopolymers* **2001**, *60*, 194–205.
- (218) Thapa, P.; Zhang, R.-Y.; Menon, V.; Bingham, J.-P. Native Chemical Ligation: A Boon to Peptide Chemistry. *Molecules* **2014**, *19* (9), 14461–14483. <https://doi.org/10.3390/molecules190914461>.
- (219) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776–779.
- (220) Dawson, P. E. Synthesis of Chemokines by Native Chemical Ligation. In *Methods in Enzymology*; Chemokines; Academic Press, 1997; Vol. 287, pp 34–45. [https://doi.org/10.1016/S0076-6879\(97\)87005-X](https://doi.org/10.1016/S0076-6879(97)87005-X).
- (221) D'Andrea, L. D.; Romanelli, A. *Chemical Ligation. Tools for Biomolecule Synthesis and Modification*; John Wiley & Sons, Inc.: New York, U. S. A., 2017.
- (222) Liu, S.; Pentelute, B. L.; Kent, S. B. H. Convergent Chemical Synthesis of [Lysine(24,38,83)] Human Erythropoietin. *Angew Chem Int Ed* **2015**, *51*, 993–999.
- (223) Pech, A.; Achenbach, J.; Jahnz, M.; Schulzchen, S.; Jarosch, F.; Bordusa, F.; Klussmann, S. A Thermostable D-Polymerase for Mirror-Image PCR. *Nucleic Acids Res* **2017**, *45*, 3997–4005. <https://doi.org/10.1093/nar/gkx079>.
- (224) Xu, W.; Jiang, W.; Wang, J.; Yu, L.; Chen, J.; Liu, X.; Liu, L.; Zhu, T. F. Total Chemical Synthesis of a Thermostable Enzyme Capable of Polymerase Chain Reaction. *Cell Discov* **2017**, *3*, 17008. <https://doi.org/10.1038/celldisc.2017.8>.

- (225) Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and α -Ketoacids. *Angew Chem Int Ed* **2006**, *45*, 1248–1252.
- (226) Pusterla, I.; Bode, J. W. The Mechanism of the α -Ketoacid-Hydroxylamine Amide-Forming Ligation. *Angew. Chem. Int. Ed Engl.* **2012**, *51* (2), 513–516. <https://doi.org/10.1002/anie.201107198>.
- (227) Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. Chemical Protein Synthesis by Chemoselective α -Ketoacid-Hydroxylamine (KAHA) Ligations with 5-Oxaproline. *Angew. Chem.* **2012**, *124* (21), 5204–5208. <https://doi.org/10.1002/ange.201200907>.
- (228) Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. Chemical Protein Synthesis by Chemoselective Alpha-Ketoacid-Hydroxylamine (KAHA) Ligations with 5-Oxaproline. *Angew Chem Int Ed* **2012**, *51*, 5114–5118. <https://doi.org/10.1002/anie.201202064>.
- (229) Wucherpfennig, T. G.; Rohrbacher, F.; Pattabiraman, V. R.; Bode, J. W. Formation and Rearrangement of Homoserine Depsipeptides and Depsiproteins in the Alpha-Ketoacid-Hydroxylamine Ligation with 5-Oxaproline. *Angew Chem Int Ed* **2014**, *53*, 12244–12247.
- (230) Pusterla, I.; Bode, J. W. An Oxazetidine Amino Acid for Chemical Protein Synthesis by Rapid, Serine-Forming Ligations. *Nat Chem* **2015**, *7*, 668–672.
- (231) Liu, C.-F.; Tam, J. P. Chemical Ligation Approach To Form a Peptide Bond between Unprotected Peptide Segments. Concept and Model Study. *J. Am. Chem. Soc.* **1994**, *116* (10), 4149–4153. <https://doi.org/10.1021/ja00089a001>.
- (232) Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. Salicylaldehyde Ester-Induced Chemoselective Peptide Ligations: Enabling Generation of Natural Peptidic Linkages at the Serine/Threonine Sites. *Org. Lett.* **2010**, *12* (8), 1724–1727. <https://doi.org/10.1021/ol1003109>.
- (233) Mende, F.; Seitz, O. 9-Fluorenylmethoxycarbonyl-Based Solid-Phase Synthesis of Peptide α -Thioesters. *Angew Chem Int Ed* **2011**, *50*, 1232–1240.
- (234) Kawakami, T. Peptide Thioester Formation via an Intramolecular N to S Acyl Shift for Peptide Ligation. *Top. Curr. Chem.* **2015**, *362*, 107–135. https://doi.org/10.1007/128_2014_575.
- (235) Hojo, H.; Onuma, Y.; Akimoto, Y.; Nakahara, Y.; Nakahara, Y. Corrigendum to “N-Alkyl Cysteine-Assisted Thioesterification of Peptides” • : [Tetrahedron Lett. 48 (2007) 25]. *Tetrahedron Lett* **2007**, *48*, 1299.
- (236) Kawakami, T.; Aimoto, S. Peptide Ligation via the In-Situ Transformation of an Amide into a Thioester at a Cysteine Residue. *Adv Exp Med Biol* **2009**, *611*, 117–118.
- (237) Kawakami, T.; Hasegawa, K.; Teruya, K.; Akaji, K.; Horiuchi, M.; Inagaki, F.; Kurihara, Y.; Uesugi, S.; Aimoto, S. Polypeptide Synthesis Using an Expressed Peptide as a Building Block for Condensation with a Peptide Thioester: Application to the Synthesis of Phosphorylated P21Max Protein(1-101). *J Pept Sci* **2001**, *7*, 474–487. <https://doi.org/10.1002/psc.341>.
- (238) Sato, K.; Shigenaga, A.; Tsuji, K.; Tsuda, S.; Sumikawa, Y.; Sakamoto, K.; Otaka, A. N-Sulfanylethylanilide Peptide as a Crypto-Thioester Peptide. *ChemBioChem* **2011**, *12* (12), 1840–1844. <https://doi.org/10.1002/cbic.201100241>.
- (239) Burlina, F.; Papageorgiou, G.; Morris, C.; White, P. D.; Offer, J. In Situ Thioester Formation for Protein Ligation Using A-Methylcysteine. *Chem Sci* **2014**, *5*, 766–770. <https://doi.org/10.1039/c3sc52140k>.
- (240) Terrier, V. P.; Adihou, H.; Arnould, M.; Delmas, A. F.; Aucagne, V. A Straightforward Method for Automated Fmoc-Based Synthesis of Bio-Inspired Peptide Crypto-Thioesters. *Chem Sci* **2016**, *7*, 339–345.
- (241) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org Lett* **2010**, *12*, 5238–5241.
- (242) Dheur, J.; Ollivier, N.; Vallin, A.; Melnyk, O. Synthesis of Peptide Alkylthioesters Using the Intramolecular N,S-Acyl Shift Properties of Bis(2-Sulfanylethyl)Amido Peptides. *J Org Chem* **2011**, *76*, 3194–3202.

- (243) Hou, W.; Zhang, X.; Li, F.; Liu, C. F. Peptidyl N,N-Bis(2-Mercaptoethyl)-Amides as Thioester Precursors for Native Chemical Ligation. *Org Lett* **2011**, *13*, 386–389.
- (244) Tesser, G. I.; Balvert-Geers, I. C. The Methylsulfonylethoxycarbonyl Group, a New and Versatile Amino Protective Function. *Int J Pept Protein Res* **1975**, *7*, 295–305.
- (245) Camarero, J. A.; Pavel, J.; Muir, T. W. Chemical Synthesis of a Circular Protein Domain: Evidence for Folding-Assisted Cyclization. *Angew Chem Int Ed* **1998**, *37*, 347–349.
- (246) Veber, D.; Milkowski, J.; Varga, S.; Denkwalter, R.; Hirschmann, R. Acetamidomethyl. A Novel Thiol Protecting Group for Cysteine. *J Am Chem Soc* **1972**, *94*, 5456–5461.
- (247) Bang, D.; Chopra, N.; Kent, S. B. H. Total Chemical Synthesis of Crambin. *J Am Chem Soc* **2004**, *126*, 1377–1383.
- (248) Wintermann, F.; Engelbrecht, S. Reconstitution of the Catalytic Core of F-ATPase (Ab)3g from Escherichia Coli Using Chemically Synthesized Subunit g. *Angew Chem Int Ed* **2013**, *52*, 1309–1313. <https://doi.org/10.1002/anie.201206744>.
- (249) Bang, D.; Kent, S. B. H. A One-Pot Total Synthesis of Crambin. *Angew. Chem. Int. Ed.* **2004**, *43* (19), 2534–2538. <https://doi.org/10.1002/anie.200353540>.
- (250) Di Bello, C.; Filira, F.; Giormani, V.; D'Angeli, F. An Investigation of Racemisation during the Use of Acetoacetyl-L-Valine in Peptide Synthesis. *J Chem Soc C* **1969**, 350–352. <https://doi.org/10.1039/j39690000350>.
- (251) Boll, E.; Ebran, J. P.; Drobecq, H.; El-Mahdi, O.; Raibaut, L.; Ollivier, N.; Melnyk, O. Access to Large Cyclic Peptides by a One-Pot Two-Peptide Segment Ligation/Cyclization Process. *Org Lett* **2015**, *17*, 130–133. <https://doi.org/10.1021/ol503359w>.
- (252) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L. One-Pot Native Chemical Ligation of Peptide Hydrazides Enables Total Synthesis of Modified Histones. *Org Biomol Chem* **2014**, *12*, 5435–5441. <https://doi.org/10.1039/c4ob00715h>.
- (253) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew Chem Int Ed* **2011**, *50*, 7645–7649.
- (254) Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. Chemical Ligation of Unprotected Peptides Directly from a Solid Support. *J Pept Res* **1998**, *51*, 303–316.
- (255) Drobecq, H.; Boll, E.; Sénéchal, M.; Desmet, R.; Saliou, J.-M.; Lacapère, J.-J.; Mougel, A.; Vicogne, J.; Melnyk, O. A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide–Protein Conjugates. *Bioconjug. Chem.* **2016**, *27* (6), 1540–1546. <https://doi.org/10.1021/acs.bioconjchem.6b00211>.
- (256) Boll, E.; Drobecq, H.; Ollivier, N.; Blanpain, A.; Raibaut, L.; Desmet, R.; Vicogne, J.; Melnyk, O. One-Pot Chemical Synthesis of Small Ubiquitin-like Modifier (SUMO) Protein-Peptide Conjugates Using Bis(2-Sulfanylethyl)Amido Peptide Latent Thioester Surrogates. *Nat Protoc* **2015**, *10*, 269–292. <https://doi.org/10.1038/nprot.2015.013>.
- (257) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El-Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew Chem Int Ed* **2012**, *51*, 209–213.
- (258) Bang, D.; Pentelute, B. L.; Kent, S. B. H. Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew Chem Int Ed* **2006**, *45*, 3985–3988.
- (259) Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept - Nat. Rev. Chem. - X-MOL <https://www.x-mol.com/paper/5416047> (accessed Jan 4, 2020).
- (260) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**. <https://doi.org/10.1021/acs.chemrev.8b00712>.

- (261) Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S. Synthesis and Folding of a Mirror-Image Enzyme Reveals Ambidextrous Chaperone Activity. *Proc Natl Acad Sci U A* **2015**, *111*, 11679–11684.
- (262) Tang, S.; Liang, L.-J.; Si, Y.-Y.; Gao, S.; Wang, J.-X.; Liang, J.; Mei, Z.; Zheng, J.-S.; Liu, L. Practical Chemical Synthesis of Atypical Ubiquitin Chains by Using an Isopeptide-Linked Ub Isomer. *Angew Chem Int Ed* **2017**, *56*, 13333–13337. <https://doi.org/10.1002/anie.201708067>.
- (263) Carugo, O. Amino Acid Composition and Protein Dimension. *Protein Sci* **2008**, *17*, 2187–2191.
- (264) Carugo, O.; Pongor, S. A Normalized Root-Mean-Square Distance for Comparing Protein Three-Dimensional Structures. *Protein Sci. Publ. Protein Soc.* **2001**, *10* (7), 1470–1473.
- (265) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed Engl.* **2007**, *46* (48), 9248–9252. <https://doi.org/10.1002/anie.200704195>.
- (266) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533. <https://doi.org/10.1021/ja003265m>.
- (267) El Oualid, F.; Merckx, R.; Ekkebus, R.; Hameed, D. S.; Smit, J. J.; de Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. Chemical Synthesis of Ubiquitin, Ubiquitin-Based Probes, and Diubiquitin. *Angew Chem Int Ed* **2010**, *49*, 10149–10153. <https://doi.org/10.1002/anie.201005995>.
- (268) Hoffmann, F. W.; Ess, R. J.; Simmons, T. C.; Hanzel, R. S. THE DESULFURIZATION OF MERCAPTANS WITH TRIALKYL PHOSPHITES. *J. Am. Chem. Soc.* **1956**, *78* (24), 6414–6414. <https://doi.org/10.1021/ja01605a034>.
- (269) Walling, C.; Rabinowitz, R. The Reaction of Thiyl Radicals with Trialkyl Phosphites. *J Am Chem Soc* **2014**, *79*, 5326–5326. <https://doi.org/10.1021/ja01576a077>.
- (270) Shang, S.; Tan, Z.; Danishefsky, S. J. Application of the Logic of Cysteine-Free Native Chemical Ligation to the Synthesis of Human Parathyroid Hormone (HPTH). *Proc Natl Acad Sci U A* **2011**, *108*, 5986–5989.
- (271) Kumar, K. S.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. Total Chemical Synthesis of Di-Ubiquitin Chains. *Angew Chem Int Ed* **2011**, *49*, 9126–9131.
- (272) Pan, M.; Zheng, Q.; Ding, S.; Zhang, L.; Qu, Q.; Wang, T.; Hong, D.; Ren, Y.; Liang, L.; Chen, C.; Mei, Z.; Liu, L. Chemical Protein Synthesis Enabled Mechanistic Studies on the Molecular Recognition of K27-Linked Ubiquitin Chains. *Angew. Chem.* **2019**, *131* (9), 2653–2657. <https://doi.org/10.1002/ange.201810814>.
- (273) Bodanszky, M.; Deshmane, S. S.; Martinez, J. Side Reactions in Peptide Synthesis. 11. Possible Removal of the 9-Fluorenylmethoxycarbonyl Group by the Amino Components during Coupling. *J. Org. Chem.* **1979**, *44* (10), 1622–1625. <https://doi.org/10.1021/jo01324a008>.
- (274) Barany, G.; Kneib-Cordonier, N.; Mullen, D. G. Solid-Phase Peptide Synthesis: A Silver Anniversary Report*. *Int. J. Pept. Protein Res.* **1987**, *30* (6), 705–739. <https://doi.org/10.1111/j.1399-3011.1987.tb03385.x>.
- (275) Dölling, R.; Beyermann, M.; Haenel, J.; Kernchen, F.; Krause, E.; Franke, P.; Brudel, M.; Bienert, M. Piperidine-Mediated Side Product Formation for Asp(OBut)-Containing Peptides. *J. Chem. Soc. Chem. Commun.* **1994**, No. 7, 853–854. <https://doi.org/10.1039/C39940000853>.
- (276) Lauer, J. L.; Fields, C. G.; Fields, G. B. Sequence Dependence of Aspartimide Formation during 9-Fluorenylmethoxycarbonyl Solid-Phase Peptide Synthesis. *Let. Pept. Sci.* **1995**, *1* (4), 197–205. <https://doi.org/10.1007/BF00117955>.
- (277) Schön, I.; Nyéki, O. Unprecedented Transformation of Aspartyl Peptides by Conjugative Degradation. *J. Chem. Soc. Chem. Commun.* **1994**, No. 4, 393–394. <https://doi.org/10.1039/C39940000393>.

- (278) Orpiszewski, J.; Schormann, N.; Kluge-Beckerman, B.; Liepnieks, J. J.; Benson, M. D. Protein Aging Hypothesis of Alzheimer Disease. *FASEB J.* **2000**, *14* (9), 1255–1263. <https://doi.org/10.1096/fasebj.14.9.1255>.
- (279) Quibell, M.; Owen, D.; Packman, L. C.; Johnson, T. Suppression of Piperidine-Mediated Side Product Formation for Asp(OBu)-Containing Peptides by the Use of N-(2-Hydroxy-4-Methoxybenzyl)(Hmb) Backbone Amide Protection. *J. Chem. Soc. Chem. Commun.* **1994**, No. 20, 2343–2344. <https://doi.org/10.1039/C39940002343>.
- (280) Aylin, F.; Konuklar, S.; Aviyente, V. Modelling the Hydrolysis of Succinimide: Formation of Aspartate and Reversible Isomerization of Aspartic Acid via Succinimide. *Org. Biomol. Chem.* **2003**, *1* (13), 2290–2297. <https://doi.org/10.1039/b211936f>.
- (281) Catak, S. Deamidation of asparagine residues: direct hydrolysis versus succinimide-mediated deamidation mechanisms. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/19152321> (accessed Jan 4, 2020).
- (282) Takahashi, O.; Oda, A.; Oda, A. M. G. R.; Oda, E. Amide-Iminol Tautomerization of the C-Terminal Peptide Groups of Aspartic Acid Residues: Two-Water-Assisted Mechanism, Cyclization from the Iminol Tautomer Leading to the Tetrahedral Intermediate of Succinimide Formation, and Implication to Peptide Group Hydrogen Exchange. **2012**.
- (283) Takahashi, O.; Kirikoshi, R.; Manabe, N. Racemization of the Succinimide Intermediate Formed in Proteins and Peptides: A Computational Study of the Mechanism Catalyzed by Dihydrogen Phosphate Ion. *Int. J. Mol. Sci.* **2016**, *17* (10). <https://doi.org/10.3390/ijms17101698>.
- (284) Tomizawa, H.; Yamada, H.; Ueda, T.; Imoto, T. Isolation and Characterization of 101-Succinimide Lysozyme That Possesses the Cyclic Imide at Asp101-Gly102. *Biochemistry* **1994**, *33* (29), 8770–8774. <https://doi.org/10.1021/bi00195a019>.
- (285) Capasso, S. Thermodynamic Parameters of the Reversible Isomerization of Aspartic Residues via a Succinimide Derivative. *Thermochim. Acta* **1996**, *286* (1), 41–50. [https://doi.org/10.1016/0040-6031\(96\)02937-1](https://doi.org/10.1016/0040-6031(96)02937-1).
- (286) Violand, B. N.; Schlittler, M. R.; Kolodziej, E. W.; Toren, P. C.; Cabonce, M. A.; Siegel, N. R.; Duffin, K. L.; Zobel, J. F.; Smith, C. E.; Tou, J. S. Isolation and Characterization of Porcine Somatotropin Containing a Succinimide Residue in Place of Aspartate129. *Protein Sci. Publ. Protein Soc.* **1992**, *1* (12), 1634–1641. <https://doi.org/10.1002/pro.5560011211>.
- (287) Connolly, B. D.; Tran, B.; Moore, J. M. R.; Sharma, V. K.; Kosky, A. Specific Catalysis of Asparaginyl Deamidation by Carboxylic Acids: Kinetic, Thermodynamic, and Quantitative Structure–Property Relationship Analyses. *Mol. Pharm.* **2014**, *11* (4), 1345–1358. <https://doi.org/10.1021/mp500011z>.
- (288) Dooley, K. H.; Castellino, F. J. Solubility of Amino Acids in Aqueous Guanidinium Thiocyanate Solutions. *Biochemistry* **1972**, *11* (10), 1870–1874. <https://doi.org/10.1021/bi00760a022>.
- (289) Bhattacharya, S.; Ainarapu, S. R. K. Mechanical Softening of a Small Ubiquitin-Related Modifier Protein Due to Temperature Induced Flexibility at the Core. *J. Phys. Chem. B* **2018**, *122* (39), 9128–9136. <https://doi.org/10.1021/acs.jpcc.8b06844>.
- (290) Brinda, K. V.; Vishveshwara, S. A Network Representation of Protein Structures: Implications for Protein Stability. *Biophys. J.* **2005**, *89* (6), 4159–4170. <https://doi.org/10.1529/biophysj.105.064485>.
- (291) Pan, M.; Gao, S.; Zheng, Y.; Tan, X.; Lan, H.; Tan, X.; Sun, D.; Lu, L.; Wang, T.; Zheng, Q.; Huang, Y.; Wang, J.; Liu, L. Quasi-Racemic X-Ray Structures of K27-Linked Ubiquitin Chains Prepared by Total Chemical Synthesis. *J Am Chem Soc* **2016**, *138*, 7429–7435. <https://doi.org/10.1021/jacs.6b04031>.
- (292) Gong, L.; Millas, S.; Maul, G. G.; Yeh, E. T. H. Differential Regulation of Sentrinized Proteins by a Novel Sentrin-Specific Protease. *J. Biol. Chem.* **2000**, *275* (5), 3355–3359. <https://doi.org/10.1074/jbc.275.5.3355>.

- (293) Yeh, E. Ubiquitin-like proteins: new wines in new bottles. - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/10806345> (accessed Jan 4, 2020).
- (294) Nishida, T. Characterization of a novel mammalian SUMO-1/Smt3-specific isopeptidase, a homologue of rat axam, which is an axin-binding protein promoting beta-c... - PubMed - NCBI <https://www.ncbi.nlm.nih.gov/pubmed/11489887> (accessed Jan 4, 2020).
- (295) Sommer, S.; Weikart, N. D.; Linne, U.; Mootz, H. D. Covalent Inhibition of SUMO and Ubiquitin-Specific Cysteine Proteases by an in Situ Thiol-Alkyne Addition. *Bioorg. Med. Chem.* **2013**, 21 (9), 2511–2517. <https://doi.org/10.1016/j.bmc.2013.02.039>.

ANNEXES

TABLE OF CONTENTS

ANNEX A: EXPERIMENTAL PART	156
I. General Methods	158
A. Reagents and solvents	158
B. HPLC analysis & purification.....	158
C. MALDI-TOF analysis	159
II. Peptide synthesis	159
A. Synthesis of peptides 1a-e, 2, 4 and 5.....	159
1. Characterization of peptide 1a	164
2. Characterization of peptide 1b	165
3. Characterization of peptide 1c	166
4. Characterization of peptide 11a	167
5. Characterization of peptide 11b	168
6. Characterization of peptide 2	169
7. Characterization of peptide 4	170
8. Characterization of peptide 13	171
B. Synthesis of peptide thioesters by thiol-SEA exchange	172
1. Synthesis and characterization of peptide 3a	172
2. Synthesis and characterization of peptide 3b	174
III. Proteins synthesis	175
A. Chemical synthesis of SUMO-2, SUMO-3 and SUMO-2/3 core proteins.....	175
1. Synthesis and characterization of SUMO-2	175
2. Synthesis and characterization of SUMO-2/3 core	176
B. Desulfurization of SUMO proteins.....	177
1. Synthesis and characterization of SUMO-2 C48A	177
2. Synthesis and characterization of SUMO-3 C47A	178
3. Synthesis and characterization of SUMO-2/3 core C48A.....	179
4. Desulfurization of SUMO-2/3 conjugates	180
C. Chemical synthesis of SUMO-2/3 Lys(Cys) proteins	182
1. One-pot synthesis and characterization of SUMO-2 Lys ₁₁ (Cys) 16a	182
2. Characterization of SUMO-2/3 Lys ₁₁ (Cys) 16a-b	184
D. Chemical synthesis of SUMO-2/3 dimers	185
1. Synthesis of SUMO dimer 11	185
2. Desulfurization of SUMO-2/3 dimers	188
IV. Proteomics analysis	190
A. Proteomic analysis of r1-SUMO-2 & 3 Boston Biochem (ref. UL-752 & UL-762).....	191
B. Proteomic analysis of SUMO-2/3 conjugates 7-9a/b.....	195
1. Characterization of SUMO-3 conjugates 7-9a by proteomics analysis.....	195

2. Characterization of SUMO-3 conjugates 7-9b by proteomics analysis.....	197
C. Characterization of SUMO-2/3 dimers by proteomics analysis	199
1. Proteomic analysis of SUMO-2/3 dimer 23	199
2. Proteomic analysis of SUMO-2/3 dimer 24	201
3. Proteomic analysis of SUMO-2/3 dimer 25	203
4. Proteomic analysis of SUMO-2/3 dimer 26	204
5. Proteomic analysis of SUMO-2/3 dimer 27	206
V. Biophysical characterization	207
ANNEX B: ARTICLES	208
BIOCONJUGATE CHEMISTRY	210
The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis.....	212
Jennifer Bouchenna, ^a Magalie Sénéchal, ^a Hervé Drobecq, ^a Nicolas Stankovic-Valentin, ^b Jérôme Vicogne, ^{*a} Oleg Melnyk ^{*a}	212
Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations	237
Jennifer Bouchenna, Magalie Sénéchal, Hervé Drobecq, Jérôme Vicogne,* Oleg Melnyk*	237
METHODS IN MOLECULAR BIOLOGY	255
The problem of aspartimide formation during protein chemical synthesis using SEA-mediated ligation	257

ANNEX A: EXPERIMENTAL PART

I. General Methods

A. Reagents and solvents

N-[(dimethylamino)-1*H*-1,2,3-triazolo-[4,5-*b*]pyridin-1-ylmethylene]-*N* methylmethanaminium hexafluorophosphate *N*-oxide (HATU) and *N*α-Fmoc protected amino acids were obtained from Iris Biotech GmbH. Side-chain protecting groups used for the amino acids were Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Asp(OtBu)[Dmb-Gly]-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(OtBu)-OH, Fmoc-Thr(OtBu)-OH, Fmoc-Tyr(OtBu)-OH, Fmoc-Val-OH, Fmoc-Cys(StBu)-OH or Fmoc-Cys(Trt)-OH.

Synthesis of *bis*(2-sulfanylethyl)aminotriyl polystyrene (SEA PS) resin was carried out as described elsewhere.^{1, 2} 4-mercaptophenylacetic acid (MPAA) and *tris*(2-carboxyethyl)phosphine hydrochloride (TCEP) were purchased from Sigma-Aldrich. All other reagents were purchased from Acros Organics or Merck and were of the purest grade available. Peptide synthesis grade *N,N*-dimethylformamide (DMF), dichloromethane (CH₂Cl₂), diethylether (Et₂O), acetonitrile (CH₃CN), heptane, LC–MS-grade acetonitrile (CH₃CN, 0.1% TFA and CH₃CN, 0.1% formic acid), LC–MS-grade water (H₂O, 0.1% TFA and H₂O, 0.1% formic acid), *N,N*-diisopropylethylamine (DIEA), acetic anhydride (Ac₂O) were purchased from Biosolve and Fisher-Chemical. Trifluoroacetic acid (TFA) was obtained from Biosolve. Water was purified with a Milli-Q Ultra-Pure Water Purification System.

B. HPLC analysis & purification

The reactions were monitored by analytical UPLC–MS (Dionex UltiMate 3000 LC/ LCQ Fleet Ion Trap) on a reverse phase column. The column, eluent system and gradient used are indicated in the figure legends. The column eluate was monitored with a Dionex DA detector (215 nm, 254 nm) and Corona Veo charged aerosol detector. The peptide masses were measured by on-line UPLC–MS: Ionization mode: ES+, *m/z* range 300–2000, capillary voltage 3.5 kV, cone voltage 10 V, tube lens 75 V, capillary voltage temperature 350 °C.

Ligations were analyzed by quenching aliquots (1.5-2 µL) of the reaction mixtures with 100 µL of 10% aqueous acetic acid. The mixture was extracted with Et₂O to remove MPAA before analysis.

The peptides were purified by semi-preparative HPLC (Waters 600 controller, UV 2487 Detector, 215 nm, TL 105 HPLC column heater) on a reverse phase column (Waters XBridge BEH300 C18, 20 x 100 mm; pore size 300 Å; particle size : 5 µm). The eluent system and gradient used are indicated in the figure legends.

C. MALDI-TOF analysis

MALDI-TOF mass spectra were recorded with a BrukerAutoflex Speed mass spectrometer.

The matrix used for the analysis is indicated in the figure legends.

II. Peptide synthesis

A. Synthesis of peptides 1a-c, 2, 4, 11a-b and 13

Peptide **1a**: ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY-SEA^{off}

Peptide **1b**: SEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY-SEA^{off}

Peptide **1c**: DHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY-SEA^{off}

Peptide **11a**: ADEKPKEGVK(C)TENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY-SEA^{on}

Peptide **11b**: SEEKPKEGVK(C)TENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY-SEA^{on}

Peptide **2**: CERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

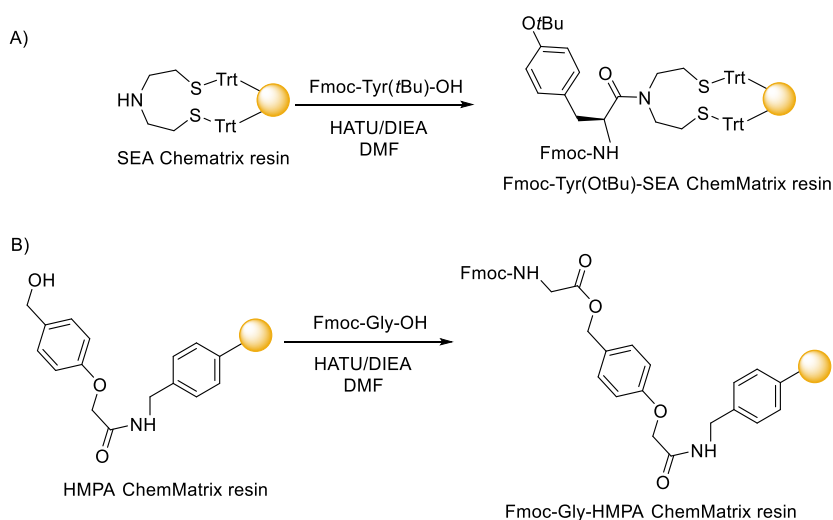
Peptide **4**: C(S^tBu)ERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-SEA^{off}

Peptide **13**: CERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-NHNH₂

Coupling of the first amino acid residue to the SEA ChemMatrix resin and HMPA ChemMatrix resin

The first amino acid was coupled manually to the SEA ChemMatrix[®] resin for SEA peptides (Scheme S 1A) or to the HMPA ChemMatrix[®] resin for peptide acids (Scheme S 1B) using HATU/DIEA activation in DMF.

Scheme S 1. Loading of the first amino acid on the resin. A) Loading on SEA ChemMatrix[®] resin. B) Loading on HMPA ChemMatrix[®] resin



SEA ChemMatrix resin (0.220 mmol/g, 454 mg, 100 μ mol) was conditioned in CH₂Cl₂ (3 \times 2 min, 3 mL) and then in DMF (3 \times 2 min, 3 mL) in a manual SPPS glass reactor.

Fmoc-Tyr(O^tBu)-OH (460 mg, 1.00 mmol) and HATU (376 mg, 0.950 mmol) was dissolved in the minimal volume of DMF (2 mL). DIEA (348 μ L, 2.00 mmol) was added to the above solution to start the activation of the amino acid. This solution was agitated for 1 min and then added to the resin which was shaken during 1 h 30 min at room temperature. The resin was subsequently washed with DMF (5 \times 2 min, 3 mL). The chloranil assay was negative. The resin was then acetylated with a mixture of acetic anhydride/DIEA/DMF: 10/5/85 by vol (2 \times 3 mL, 2 min and then 20 min) before being washed successively with DMF (3 \times 2 min, 3 mL), CH₂Cl₂ (3 \times 2 min, 3 mL) and diethyl ether (3 \times 2 min, 3 mL). The resin was finally dried in vacuo for ~2 h.

HMPA ChemMatrix resin (500 mg, Lot 08k11-20-03-043) was conditioned in CH₂Cl₂ (3 \times 2 min, 3 mL), in DMF (3 \times 2 min, 3 mL) and then in CH₂Cl₂ (3 \times 2 min, 2 mL) in a manual SPPS glass reactor.

Fmoc-Gly-OH (297 mg, 1.00 mmol) was dissolved in CH₂Cl₂/DMF (15/1 by vol, 7 mL) to which was added *N,N'*-diisopropylcarbodiimide (DIC, 78.3 μ L, 0.500 mmol) to start the activation. After 15 min, 4-dimethylaminopyridine (DMAP, 12.2 mg, 0.100 mmol) was added to the above solution, which was agitated for 1 min and then added to the resin. The resin was shaken for 1 h 30 min at room temperature, and subsequently washed with DMF (5 \times 2 min, 3 mL). The resin was then acetylated with a mixture of acetic anhydride/DIEA/DMF: 10/5/85 by vol (2 \times 3 mL, 2 min and then 20 min) before being washed successively with DMF (3 \times 2 min, 3 mL), CH₂Cl₂ (3 \times 2 min, 3 mL) and diethyl ether (3 \times 2 min, 3 mL). The resin was finally dried in vacuo for ~2 h.

Preparation of hydrazine PS resin

Commercial 2-chloro-4'-polystyryl triphenylmethyl chloride resin (0.600 mmol/g, 667 mg, 400 μ mol) was swollen for 15 min in DMF (150 mL) and cooled at 0 °C. A mixture of trimethylamine (6 eq, 240 mmole, 334.5 μ L) and hydrazine hydrate (4 eq, 160 mmole, 102.4 μ L) in DMF (4 mL) was added dropwise and the suspension was stirred for 60 min at room temperature. Methanol (107 μ L) was then added and stirring continued for 10 min in order to quench the excess of reactive Cl. The resin was filtered, washed with DMF (2 \times 5 mL), water (2 \times 5 mL), DMF (2 \times 5 mL), methanol (2 \times 5 mL) and ether and dried in vacuo for 2 h at room temperature (protocol adapted from ref).

Coupling of the first amino acid residue to the hydrazide PS resin

Fmoc-Gly-OH (297 mg, 1.00 mmol) and HATU (376 mg, 0.950 mmol) were dissolved in the minimal volume of DMF (2 mL). DIEA (348 μ L, 2.00 mmol) was added to the above solution to start the activation of the amino acid. This solution was agitated for 1 min and then added to the hydrazine resin prepared above (1.00 mmole) which was shaken during 1 h 30 min at room temperature. The resin was subsequently washed with DMF (5 \times 2 min, 3 mL). The picrylsulfonic acid assay was negative. The resin was then acetylated with a mixture of acetic anhydride/DIEA/DMF: 10/5/85 by vol (2 \times 3 mL, 2 min and then 20 min) before being washed successively with DMF (3 \times 2 min, 3 mL), CH₂Cl₂ (3 \times 2 min, 3 mL) and diethylether (3 \times 2 min, 3 mL). The resin was finally dried in vacuo for ~2 h.

Determination of resin loading

The loading of the resins was determined by UV quantification at 290 nm of the dibenzofulvene-piperidine adduct formed by treating aliquots of the resin with piperidine (20% by vol in DMF). We found 0.18 mmol/g for Fmoc-Tyr(OtBu)-SEA ChemMatrix resin, 0.58 mmol/g for Fmoc-Gly-HMPA ChemMatrix resin and 0.18 mmol/g for Fmoc-Gly-hydrazine PS resin.

Automated solid phase peptide synthesis

The peptide elongation step was performed using an automated column peptide synthesizer and standard Fmoc-SPPS protocols (0.1 mmol scale). The amino acids (10 equiv) were activated using HATU (9.9 equiv)/DIEA (20 equiv) in DMF. The peptidyl resin was acetylated with Ac₂O/DIEA/DMF 10/5/85 by vol after each coupling step.

Coupling of the dipeptide unit Fmoc-Asp(OtBu)-(Dmb)Gly-OH

Fmoc-Asp(OtBu)-(Dmb)Gly-OH (186 mg, 0.300 mmol, 3 equiv) and HATU (108 mg, 0.284 mmol, 2.85 equiv) was dissolved in the minimal volume of DMF (2 mL). DIEA (104 μ L, 0.600 mmol, 6 equiv) was added to the above solution to start the activation of the amino acid. This solution was agitated for 1 min and then added to the resin which was shaken for 1 h 30 min at room temperature. The resin was subsequently washed with DMF (5 \times 2 min, 3 mL).

The chloranil assay was negative. An aliquot of the peptidyl resin was also treated with piperidine in DMF (20 % by vol) and cleaved for 1 h 30 min in a mixture of TFA/triisopropylsilane (TIS)//thioanisole/H₂O/thiophenol: 87.5/5/2.5/2.5/2.5 by vol (1 mL) for SEA ChemMatrix[®] resin, or TFA/TIS/H₂O/ethanedithiol (EDT): 90/5/2.5/2.5 by vol (1 mL) for HMPA-ChemMatrix[®] resin. 1 μ L of the cleavage mixture was diluted with water (50 μ L) and analyzed by MALDI-TOF to confirm the coupling of the dipeptide unit.

Deprotection and cleavage step

The peptidyl resins (0.1 mmol scale) were deprotected and cleaved in a mixture of TFA/TIS/thioanisole/H₂O/thiophenol: 87.5/5/2.5/2.5/2.5 by vol (10 mL) for 1 h 30 min (twice) for peptides **1a-e** and **4**, and in a mixture of TFA/TIS/EDT/H₂O : 90/5/2.5/2.5/2.5 by vol (10 mL) for 1 h 30 min (twice) for peptide **2** and **5**. The crude peptides were precipitated in an ice-cold mixture of Et₂O/*n*-heptane : 1/1 by vol (200 mL) to give 380.2 mg (63% crude) of peptide **1a**, 355.1 mg (54% crude) of peptide **1b**, 366.8 mg (42% crude) of peptide **1c**, 280.5 mg (53% crude) of peptide **1d**, 276.8 mg (51% crude) of peptide **1e**, 246.6 mg (42% crude) of peptide **2**, 297.1 mg (51% crude) peptide **4** 197.1 mg (37% crude) peptide **5**.

Purification of SEA^{off} peptide segments

The crude SEA^{on} peptide segments **1a-c**, **4** were dissolved in phosphate buffer/6 M guanidine hydrochloride (final peptide concentration ~0.25 mM). Iodine solution (200 mM in DMSO) was added dropwise until the appearance of a yellow color. After 30 s, dithiothreitol (DTT, 65 mM in water) was added dropwise until the disappearance of the yellow color to quench the excess of iodine. The mixture was then immediately filtered and purified by reversed-phase HPLC. The crude peptide segments **1d-e**, **2** and **5** were dissolved in phosphate buffer/6 M guanidine hydrochloride (final peptide concentration ~0.25 mM). The mixture was then immediately filtered and purified by reversed-phase HPLC.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

Yield for peptide **1a**: 34.4 mg of crude product furnished (11.2 mg, 20.5% global yield) of peptide **1a**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

Yield for peptide **1b**: 30.8 mg of crude product furnished (11.2 mg, 19.6% global yield) of peptide **1b**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

Yield for peptide **1c**: 30.5 mg of crude product furnished (13.7 mg, 18.9% global yield) of peptide **1c**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-45% B in 50 min, flow rate 6 mL/min, 65 °C, UV detection at 215 nm.

Yield for peptide **11a**: 34.4 mg of crude product furnished (11.2 mg, 17.2% global yield) of peptide **11a**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-45% B in 50 min, flow rate 6 mL/min, 65 °C, UV detection at 215 nm.

Yield for peptide **11b**: 30.8 mg of crude product furnished (11.2 mg, 18.5% global yield) of peptide **11b**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-20% B in 5 min, then 20-40% B in 40 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm.

Yield for peptide **2**: 31.3 mg of crude product furnished (10.6 mg, 14.2% global yield) of peptide **2**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-45% B in 50 min, flow rate 6 mL/min, 65°C, UV detection at 215 nm.

Yield for peptide **4**: 29.2 mg of crude product furnished (9.6 mg, 16.8% global yield) of peptide **4**.

Gradient used for the HPLC purification: eluent C water containing 0.1% of FA, eluent D CH₃CN/water : 4/1 by vol containing 0.1% of FA, gradient: 0-10% D in 10 min, then 10-30% D in 50 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm, ZORBAX 300SB-C3 (5 μm, 9.4 x 250 mm).

Yield for peptide **13**: 36.3 mg of crude product furnished (6.2 mg, 6.3% global yield) of peptide **13**.

1. Characterization of peptide **1a**

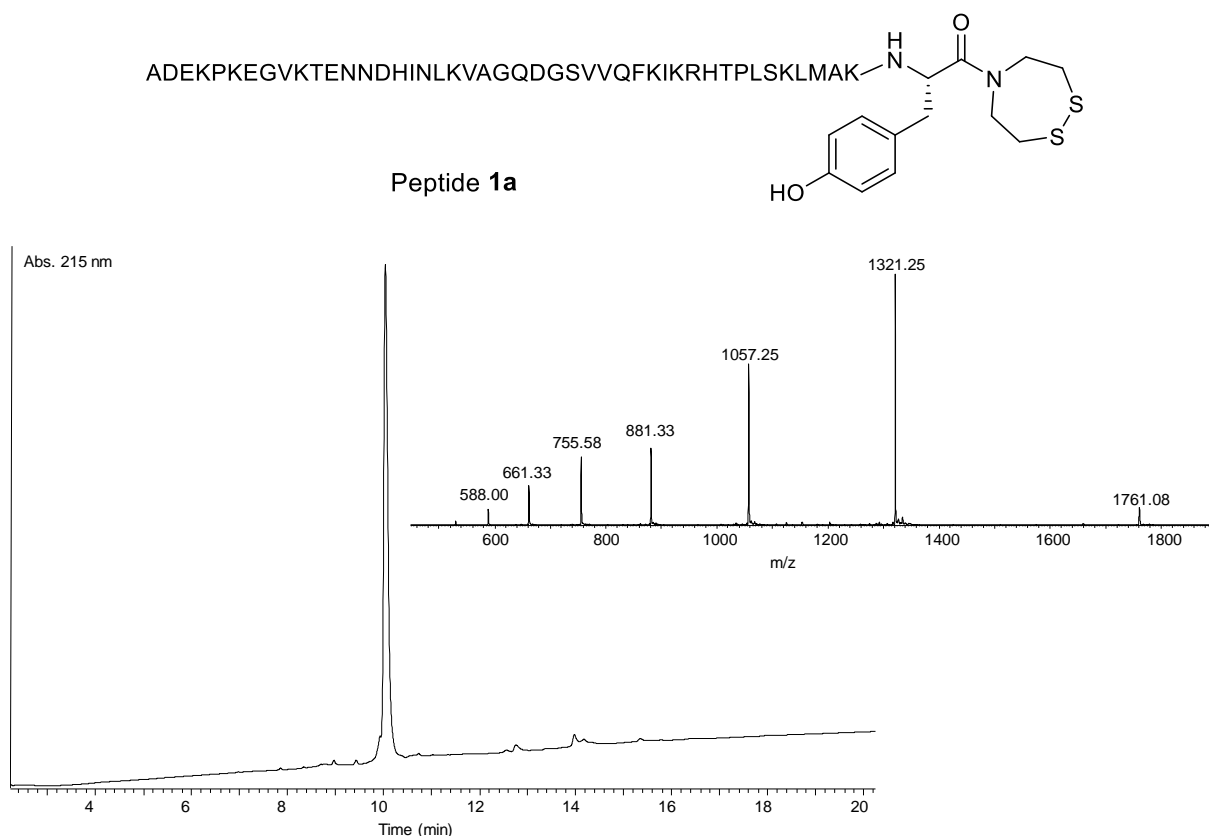


Figure S 1. Analysis of peptide **1a**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5281.17, observed 5281.13 after deconvolution

2. Characterization of peptide **1b**

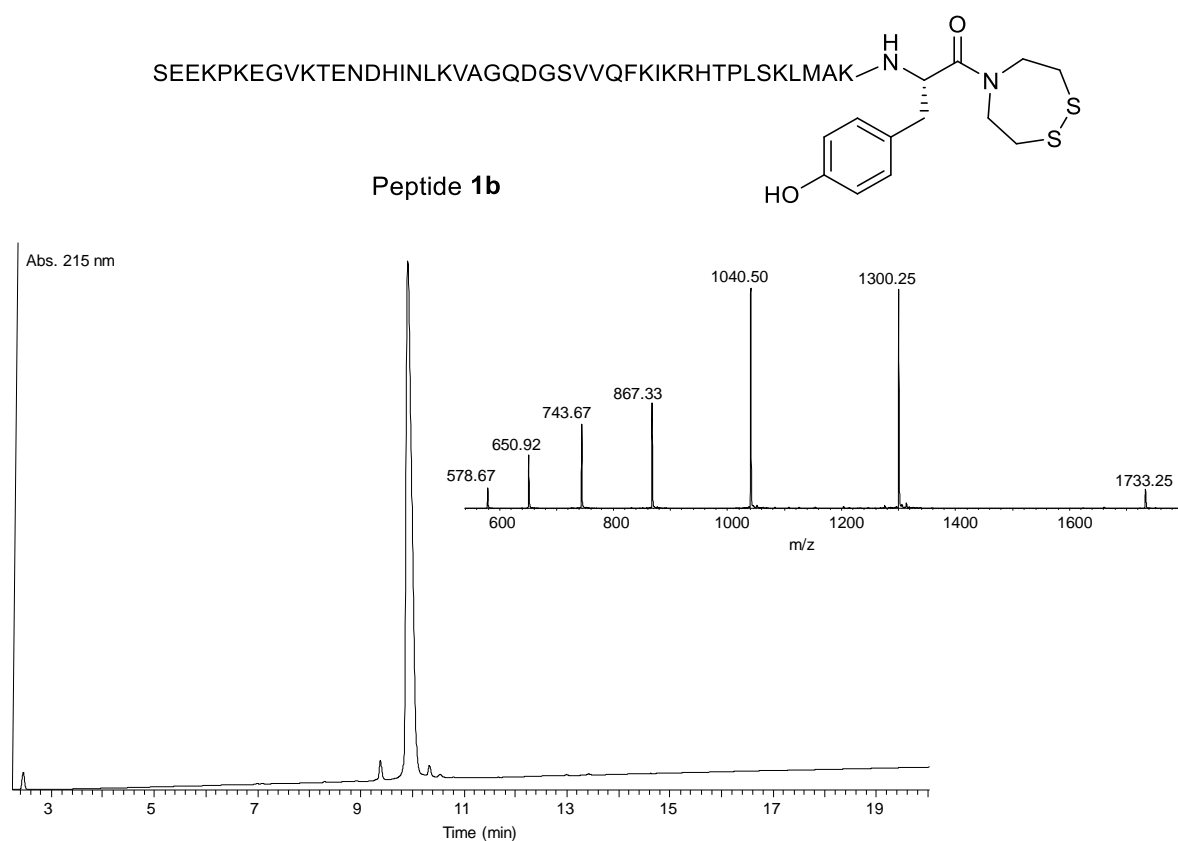


Figure S 2. Analysis of peptide **1b**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5197.09, observed 5197.17 after deconvolution.

3. Characterization of peptide 1c

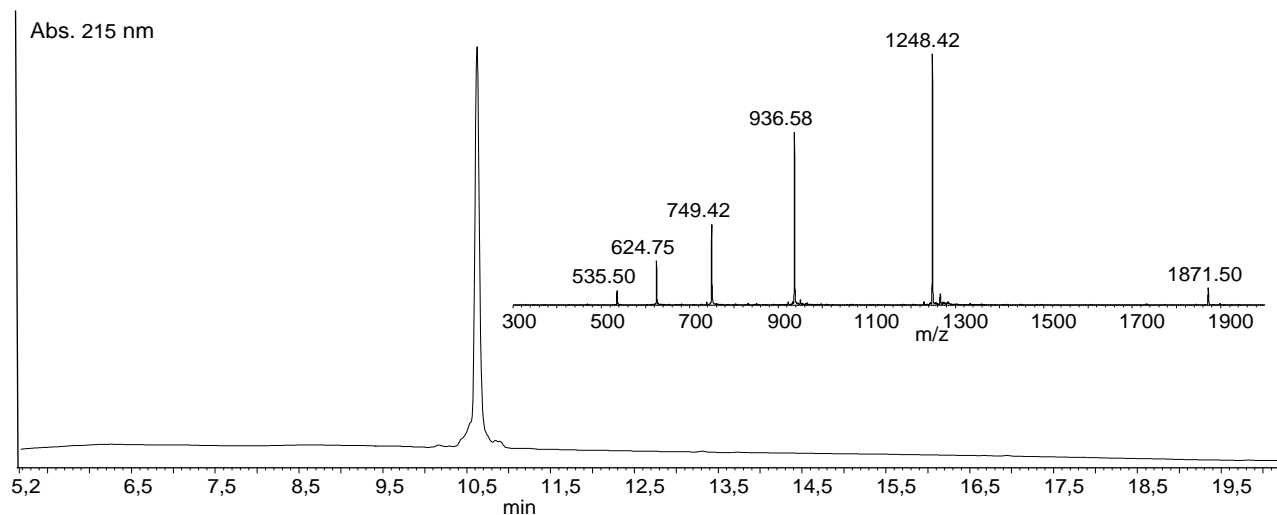
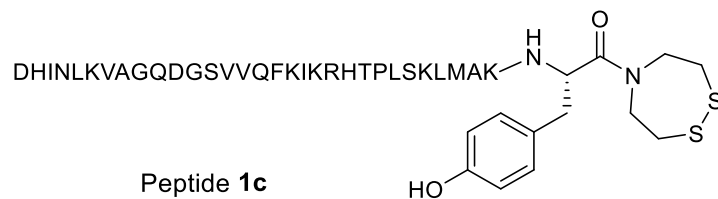


Figure S 3. Analysis of peptide 1c. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 3741.53, observed 3742.29 after deconvolution.

4. Characterization of peptide **11a**

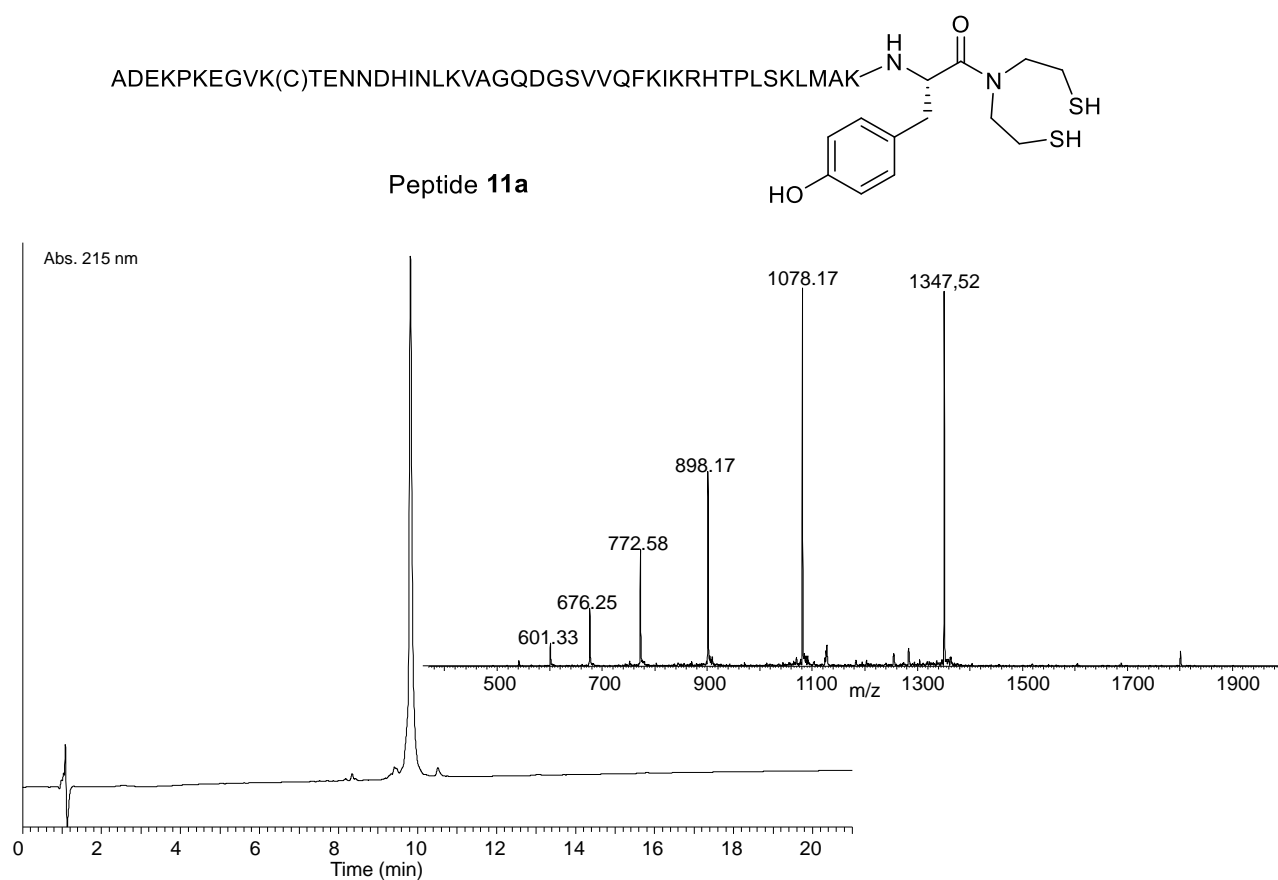


Figure S 4. Analysis of peptide **11a**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5386.33, observed 5385.97 after deconvolution.

5. Characterization of peptide **11b**

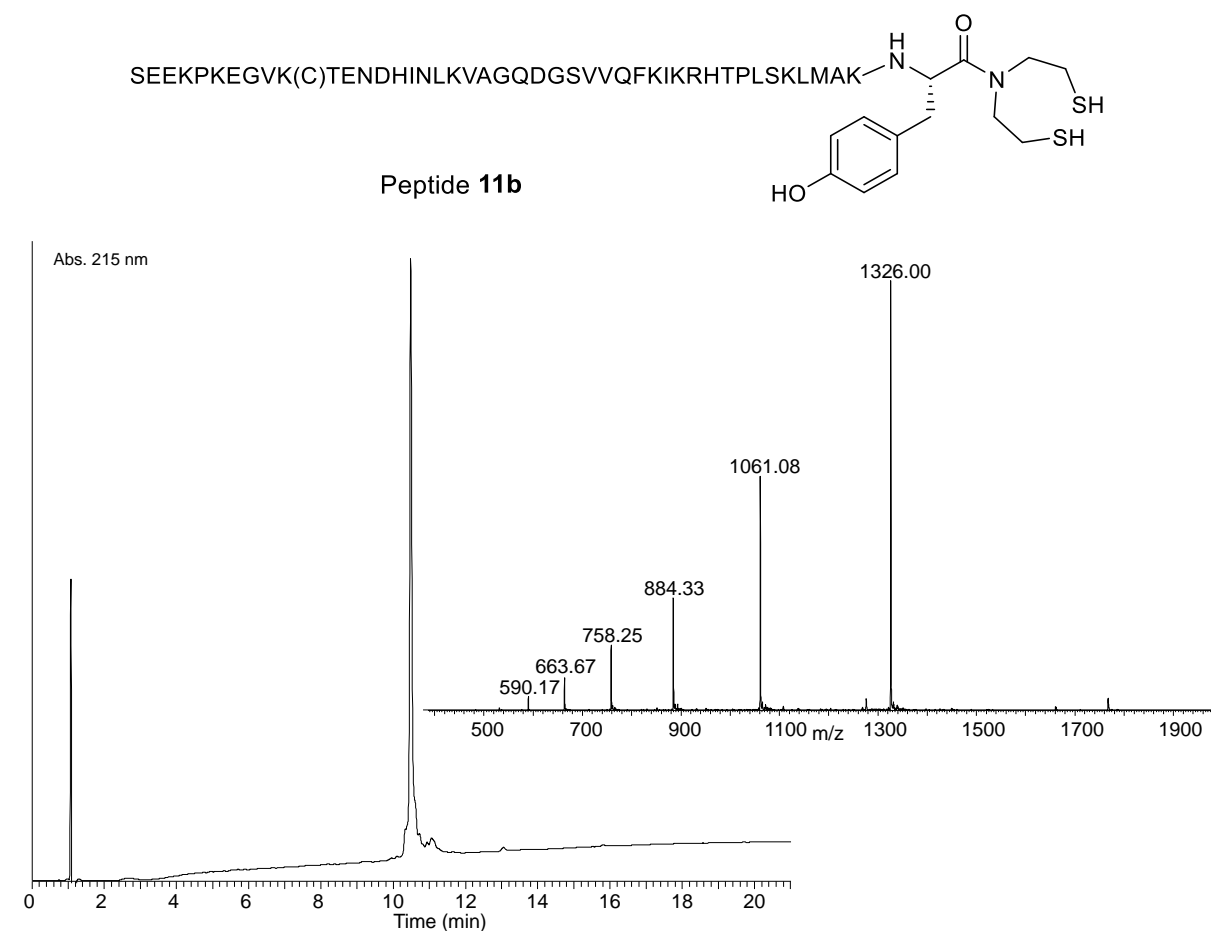


Figure S 5. Analysis of peptide **11b**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5302.25, observed 5302.20 after deconvolution.

6. Characterization of peptide 2

CERQGLSMRQIRFRFDGQPINETDTPAGLEMEDEDTIDVFQQQTGG

Peptide 2

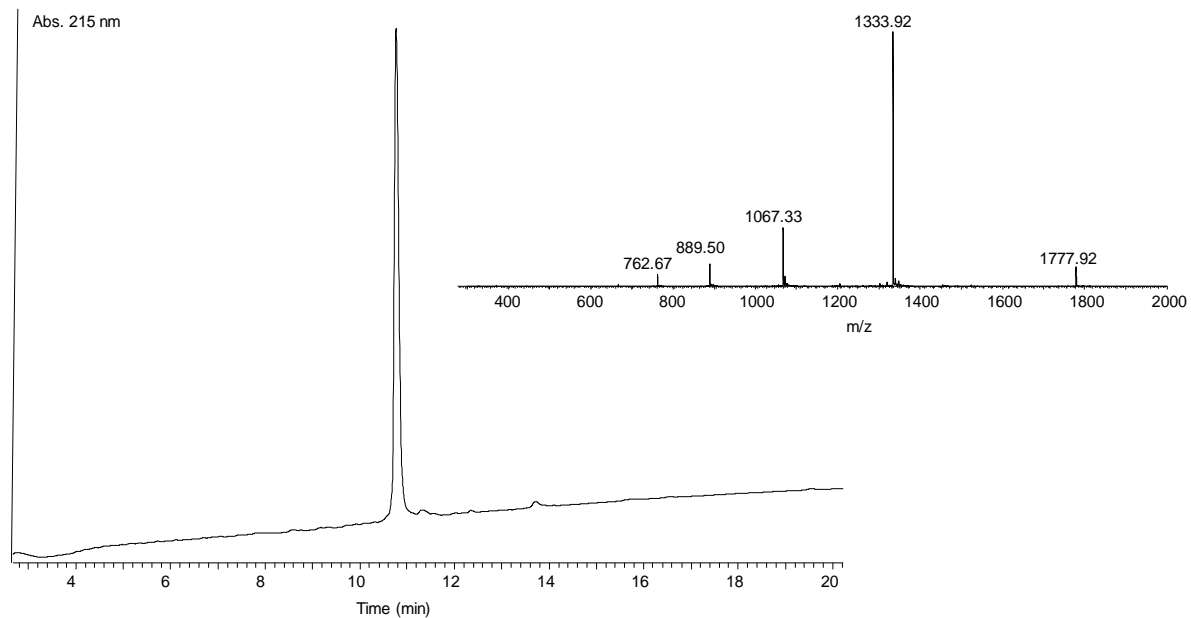


Figure S 6. Analysis of peptide 2. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5331.82, observed 5331.95 after deconvolution.

7. Characterization of peptide 4

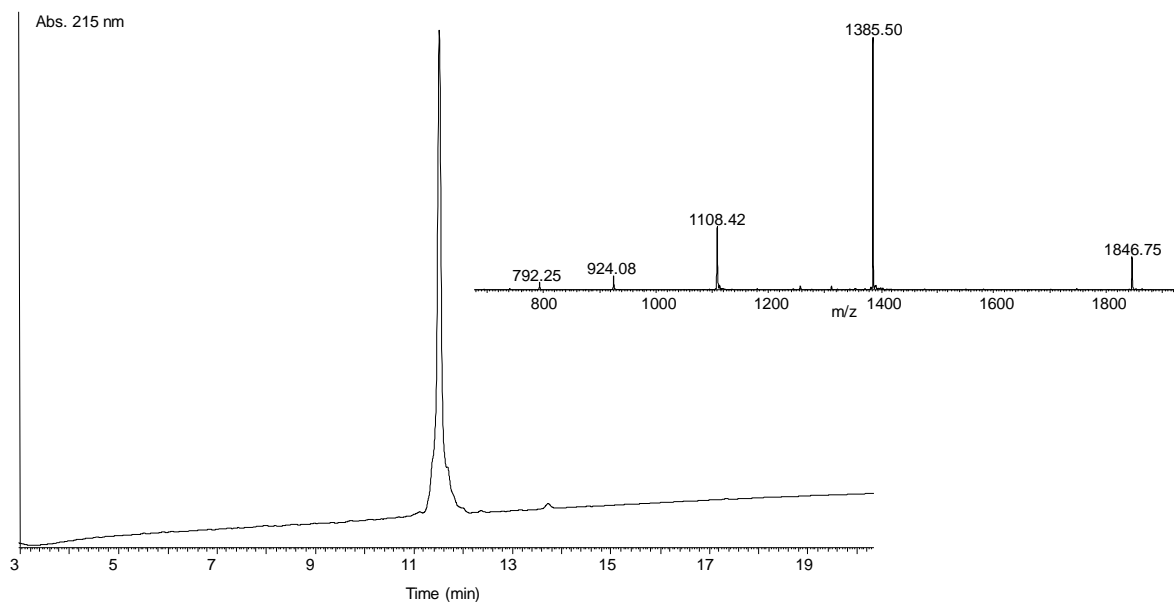
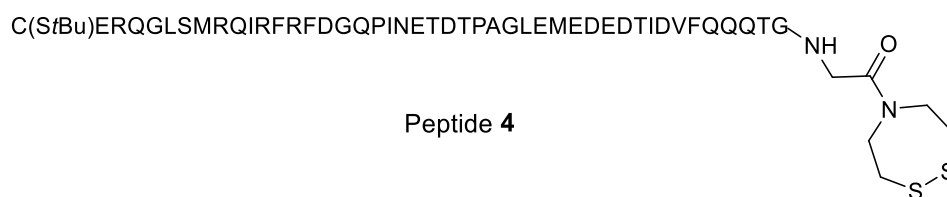


Figure S 7. Analysis of peptide 4. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace Calculated for M (average mass) 5537.21, observed 5537.59 after deconvolution.

8. Characterization of peptide **13**

CERQGLSMRQIRFRFDGQPINETDTPAGLEMEDEDTIDVFQQQTGG-NH₂

Peptide **13**

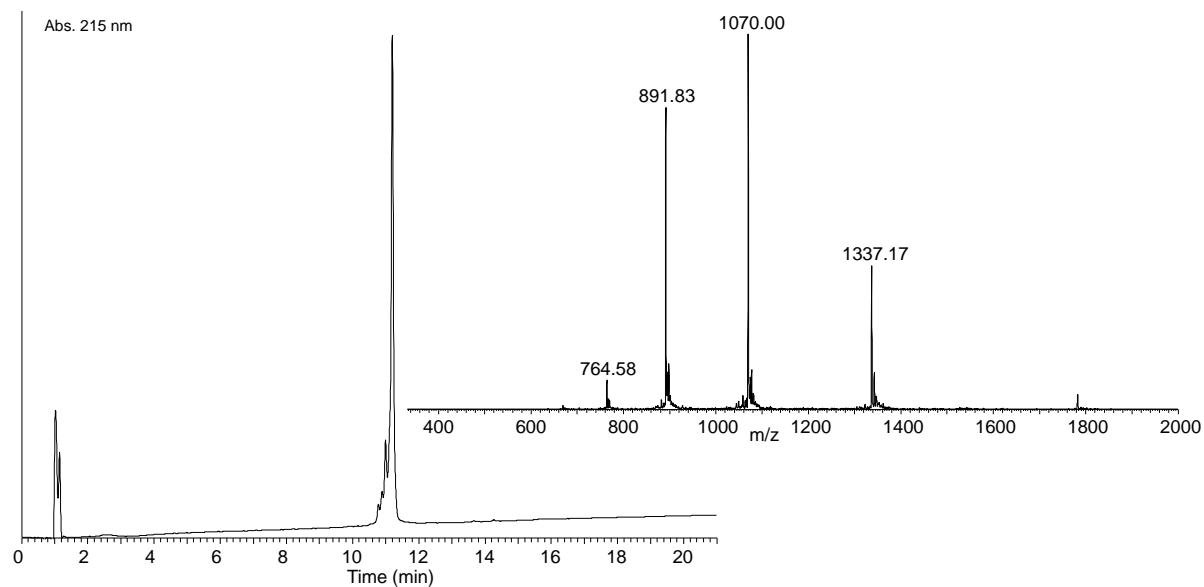
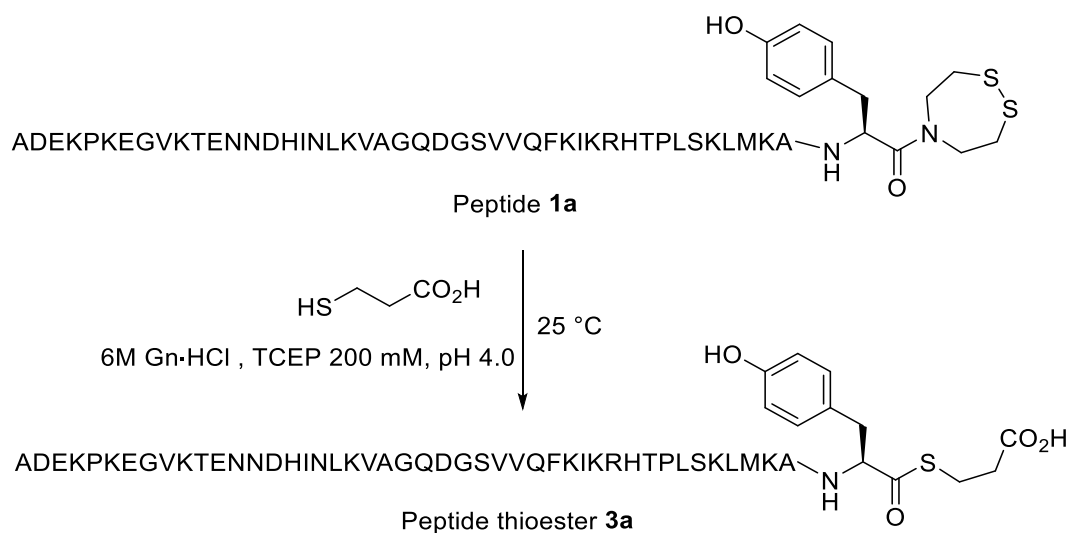


Figure S 8. Analysis of peptide **13**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5345.85, observed 5344.85 after deconvolution.

B. Synthesis of peptide thioesters by thiol-SEA exchange

1. Synthesis and characterization of peptide **3a**

Scheme S 2. SEA-thiol exchange reaction for the preparation of peptide thioester



The reaction was carried out under nitrogen atmosphere. TCEP (29.7 mg, 0.103 μmol) was dissolved in 0.1 M pH 7.2 sodium phosphate buffer/6 M Gn·HCl (1 mL). Peptide **1a** (18.04 mg, 2.752 μmol) was dissolved in the above solution (0.680 mL). Then, 3-mercaptopropionic acid (MPA, 25 μL , 0.29 μmol) was added and the pH of the reaction mixture was adjusted to 4.0 by addition of 6 N NaOH. The reaction was agitated at 25 °C for 26 h and then acidified with acetic acid (60 μL). The mixture was extracted with diethyl ether (3 \times 2 mL) to remove the excess of MPA and immediately purified by RP-HPLC using a C18 Xbridge BEH300 prep column (300 Å, 10 \times 250 mm, 5 μm) to give 10.8 mg (60%) of peptide thioester **3a**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

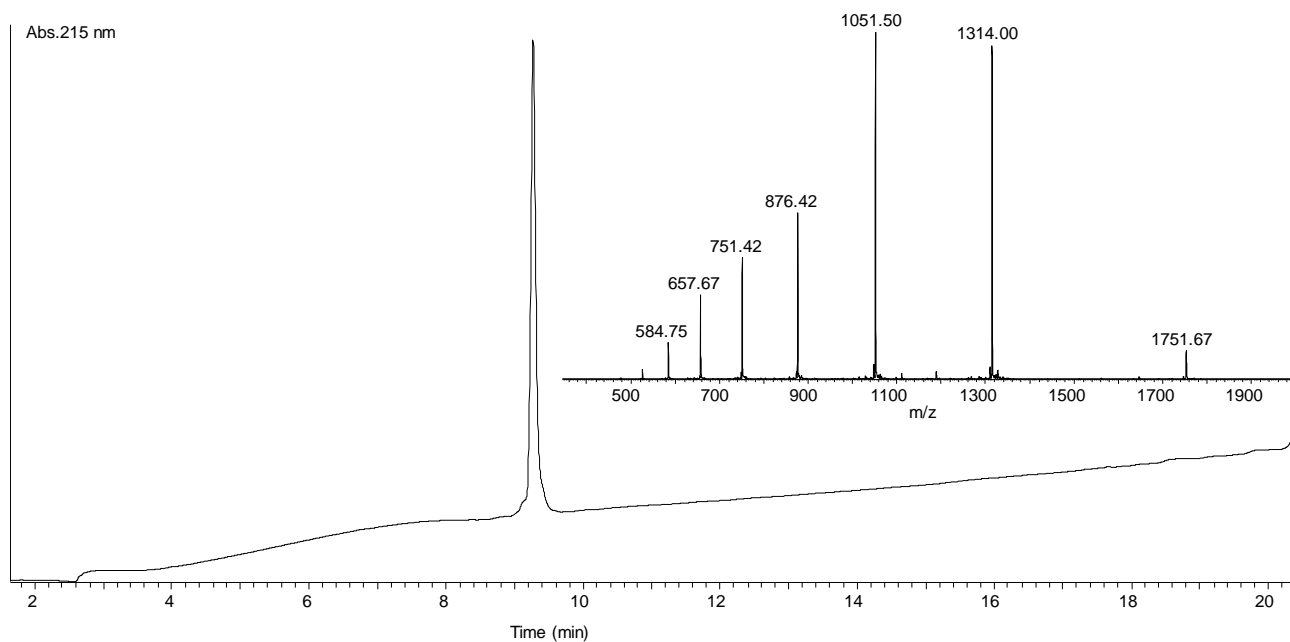
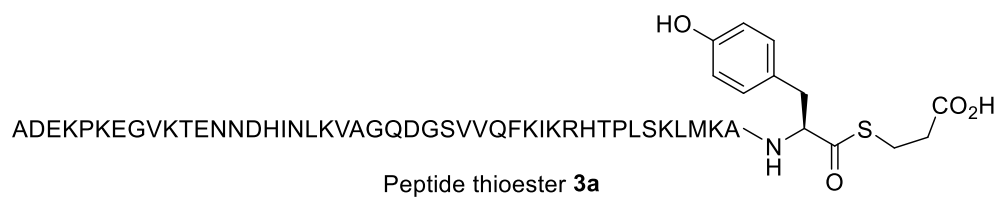
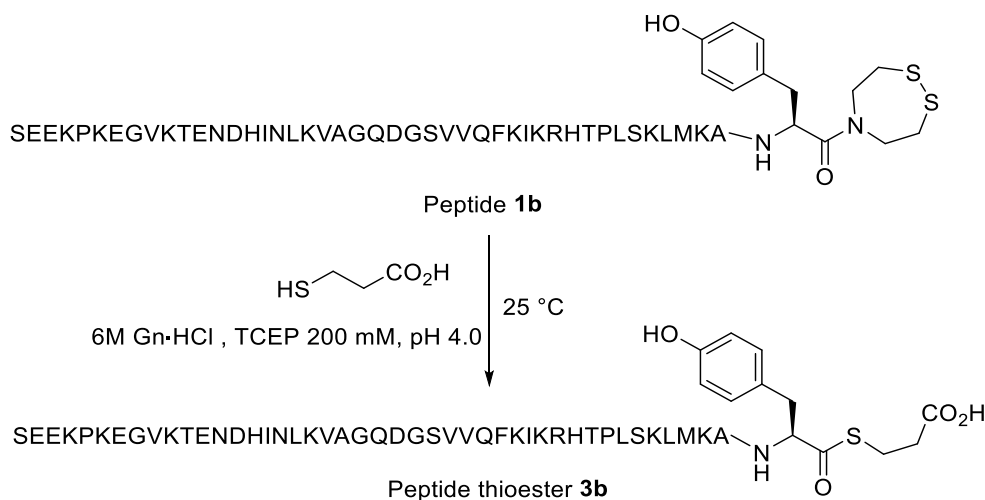


Figure S 9. Analysis of peptide **3a**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm x 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5252.07, observed 5252.25 after deconvolution.

2. Synthesis and characterization of peptide **3b**

*Scheme S 3. SEA-thiol exchange reaction for the preparation of peptide thioester **3b**.*



The reaction was carried out under nitrogen atmosphere. TCEP (29.7 mg, 0.103 μmol) was dissolved in 0.1 M pH 7.2 sodium phosphate buffer/6 M Gn·HCl. Peptide **1b** (11.09 mg, 1.692 μmol) was dissolved in the above solution (0.420 mL). Then, 3-mercaptopropionic acid (MPA, 42 μL , 0.48 μmol) was added and the pH of the reaction mixture was adjusted to 4.0 by addition of 6 M NaOH. The reaction was agitated at 25 °C for 26 h and then acidified with acetic acid (42 μL). The mixture was extracted with diethyl ether (3 \times 2 mL) to remove the excess of MPA and immediately purified by RP-HPLC using a C18 Xbridge BEH300 prep column (300 \AA , 10 \times 250 mm, 5 μm) to give 7.10 mg (65 %) of peptide thioester **3b**.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B $\text{CH}_3\text{CN}/\text{water}$: 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

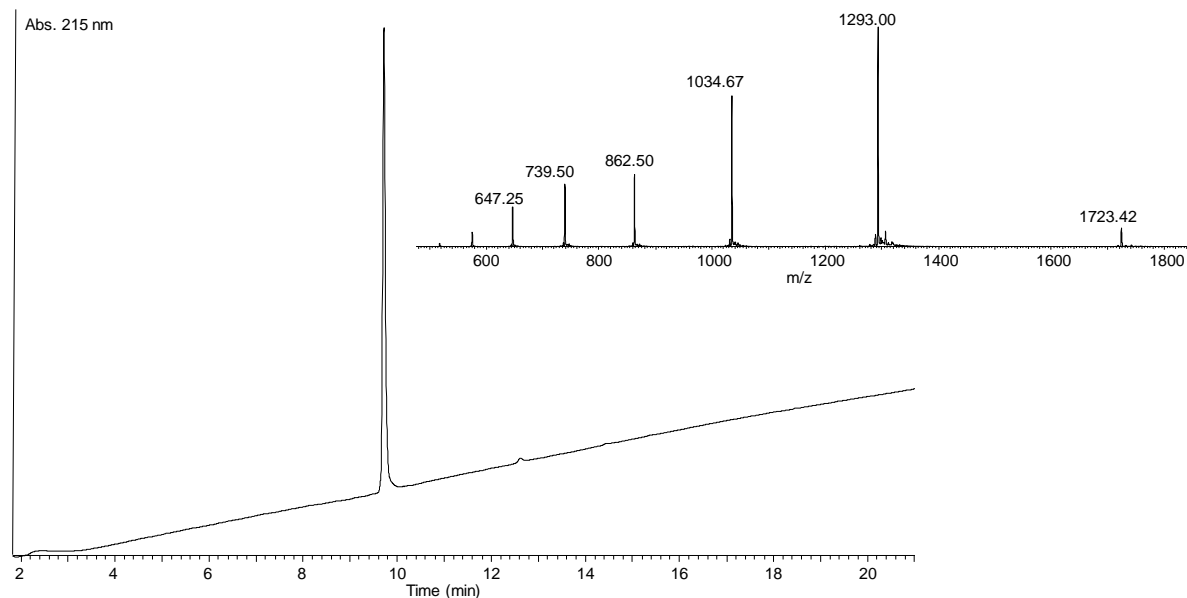
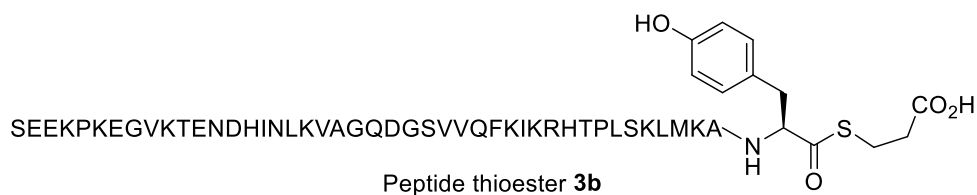


Figure S 10. Analysis of peptide **3b**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 5167.99, observed 5168.18 after deconvolution.

III. Proteins synthesis

A. Chemical synthesis of SUMO-2, SUMO-3 and SUMO-2/3 core proteins.

A typical procedure is illustrated with the preparation of SUMO-2 protein.

1. Synthesis and characterization of SUMO-2

To a solution of TCEP·HCl (31.64 mg, 110.0 µmol, 200 mM final concentration) in pH 7.2 sodium phosphate buffer/6 M Gn·HCl (200 µL) was added MPAA (13.42 mg, 79.77 µmol, 200 mM final concentration) dis-solved in the same buffer (500 µL). The pH of the mixture was then adjusted to 5.5 with 6 N NaOH.

SEA peptide 1a (2.23 mg, 0.335 µmol, 5 mM) and Cys peptide 2 (1.98 mg, 0.335 µmol, 5 mM, 1 equiv) were dissolved in the above solution (67 µL) and the reaction mixture was stirred at 25 °C under nitrogen atmosphere. The reaction was agitated for 33 h.

The progress of the reaction was monitored by HPLC. For each point, a 2 µL aliquot was taken from the reaction mixture and quenched by adding 100 µL of 10 % acetic acid. The sample was then extracted with Et₂O to remove MPAA prior to HPLC analysis.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-20% B in 5 min, then 20-40% B in 60 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.

2. Synthesis and characterization of SUMO-2/3 core

To a solution of TCEP·HCl (31.6 mg, 110 μmol, 200 mM final concentration) in pH 7.2 sodium phosphate buffer/6 M Gn·HCl (200 μL) was added MPAA (13.4 mg, 79.8 μmol, 200 mM final concentration) dissolved in the same buffer (500 μL). The pH of the mixture was then adjusted to 7.2 with 6 N NaOH.

The SEA peptide **1c** (5.95 mg, 1.25 μmol, 4 mM) and Cys peptide **2** (7.23 mg, 1.25 μmol, 4 mM, 1.0 equiv) were dissolved in the above solution (313 μL) and the reaction mixture was stirred at 25 °C under nitrogen atmosphere. The reaction was agitated for 35 h.

The progress of the reaction was monitored by HPLC. For each point, a 2 μL aliquot was taken from the reaction mixture and quenched by adding 100 μL of 10 % acetic acid. The sample was then extracted with Et₂O to remove MPAA prior to HPLC analysis.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-20% B in 5 min, then 20-40% B in 60 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm.

The purification yielded 3.57 mg (60% yield) of SUMO-2/3 core.

**SUMO-2/3 core: DHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFR
FDGQPINETDTPAQLEMEDEDITIDVFQQQTGG**

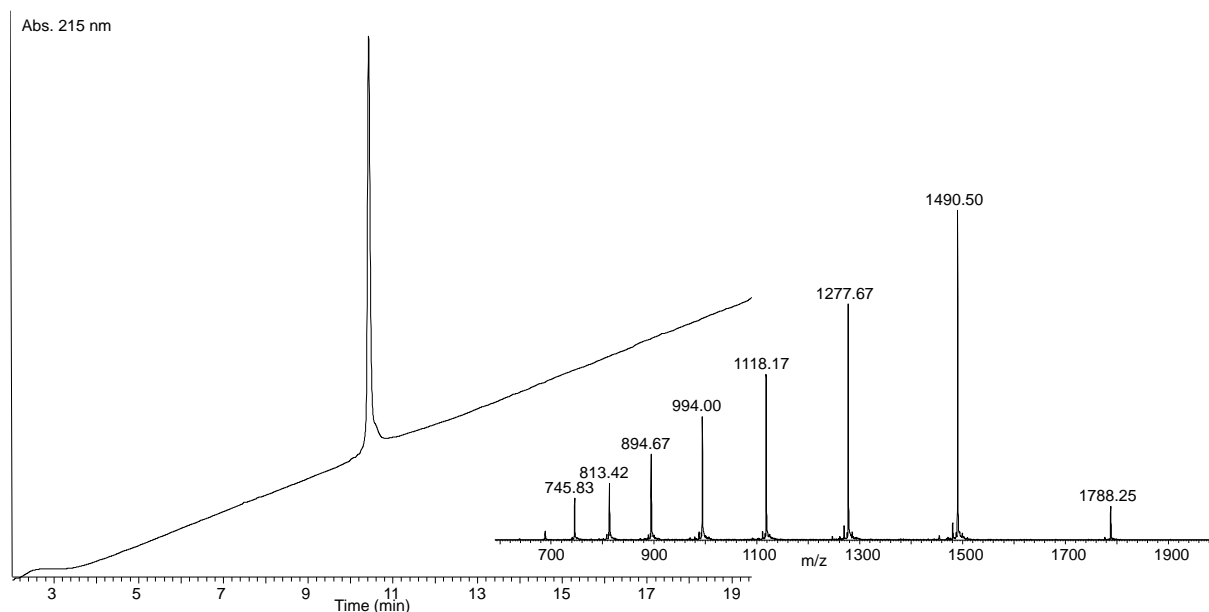


Figure S 11. Analysis of SUMO-2/3 core. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 8936.85, observed 8937.03 after deconvolution.

B. Desulfurization of SUMO proteins

1. Synthesis and characterization of SUMO-2 C48A

SUMO-2 (2.25 mg, 0.214 µmol, 0.100 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.5 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.4 mg/mL, 50.0 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer/6 M Gn·HCl (4 mL). The desulfurization was carried out at 25 °C and monitored by MALDI-TOF mass spectrometry. After 25 h, the reaction product was purified by HPLC.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-10% B in 5 min, then 10-100% B in 15 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm.

The purification yielded 1.17 mg (52% yield) of SUMO-2 C48A.

SUMO-2 C48A:
ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYAERQGLSMRQIRFR
FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

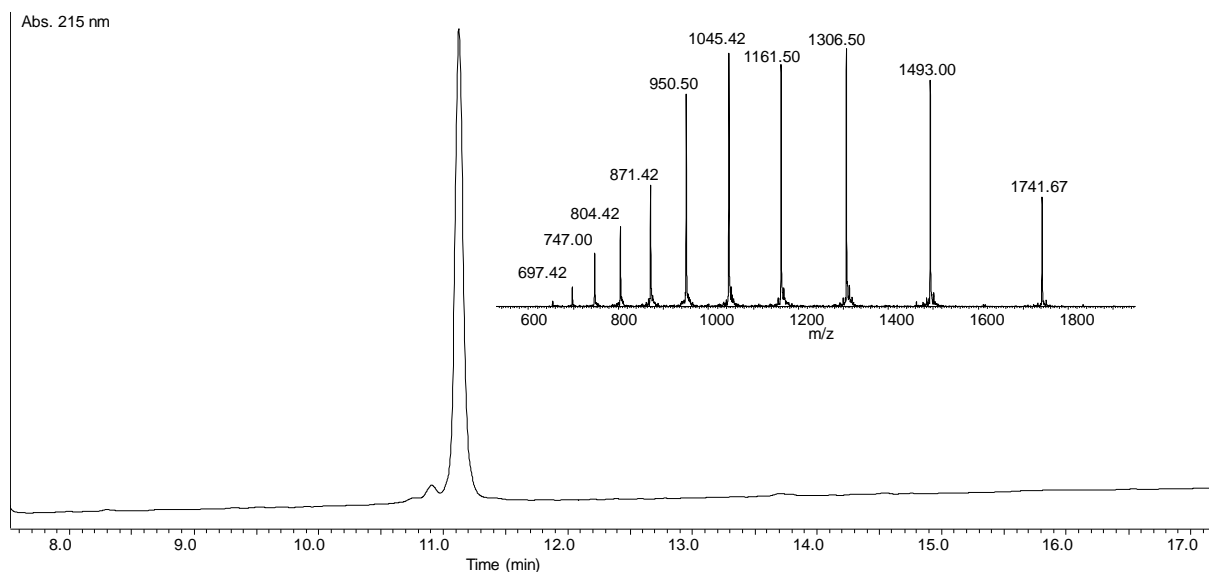


Figure S 12. Analysis of SUMO-2 C48A protein. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 10445.68, observed 10444.25 after deconvolution.

2. Synthesis and characterization of SUMO-3 C47A

SUMO-3 (3.00 mg, 0.300 µmol, 0.100 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.5 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.4 mg/mL, 50.0 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer/6 M Gn-HCl (4 mL). The desulfurization was carried out at 25 °C and monitored by MALDI-TOF mass spectrometry. After 24 h, the reaction product was purified by HPLC.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-10% B in 5 min, then 10-100% B in 15 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm

The purification yielded 1.47 mg (49% yield) of SUMO-3 C47A.

SUMO-3 C47A:
SEKPKKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYAERQGLSMRQIRFR
FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

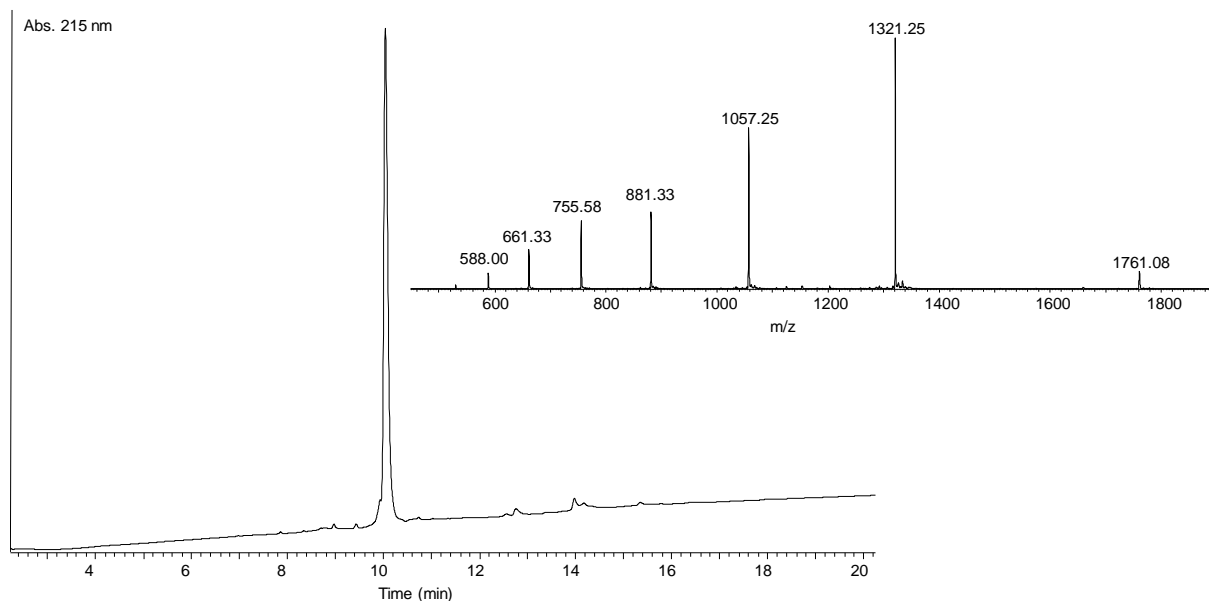


Figure S 13. Analysis of SUMO-3 C47A protein. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 10361.60, observed 10360.67 after deconvolution.

3. Synthesis and characterization of SUMO-2/3 core C48A

SUMO-2/3 core (3.00 mg, 0.336 µmol, 0.100 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.50 mg/mL, 20 mM final concentration) and reduced glutathione (15.4 mg/mL, 50.0 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer/6 M Gn-HCl (4 mL). The desulfurization was carried out at 25 °C and monitored by MALDI-TOF mass spectrometry. After 24 h, the reaction product was purified by HPLC.

Gradient used for the HPLC purification: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-10% B in 5 min, then 10-100% B in 15 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm.

The purification yielded 1.35 mg (45% yield) of SUMO-2/3 C48A.

SUMO-2/3 core C48A:
DHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYAERQGLSMRQIRFR
FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

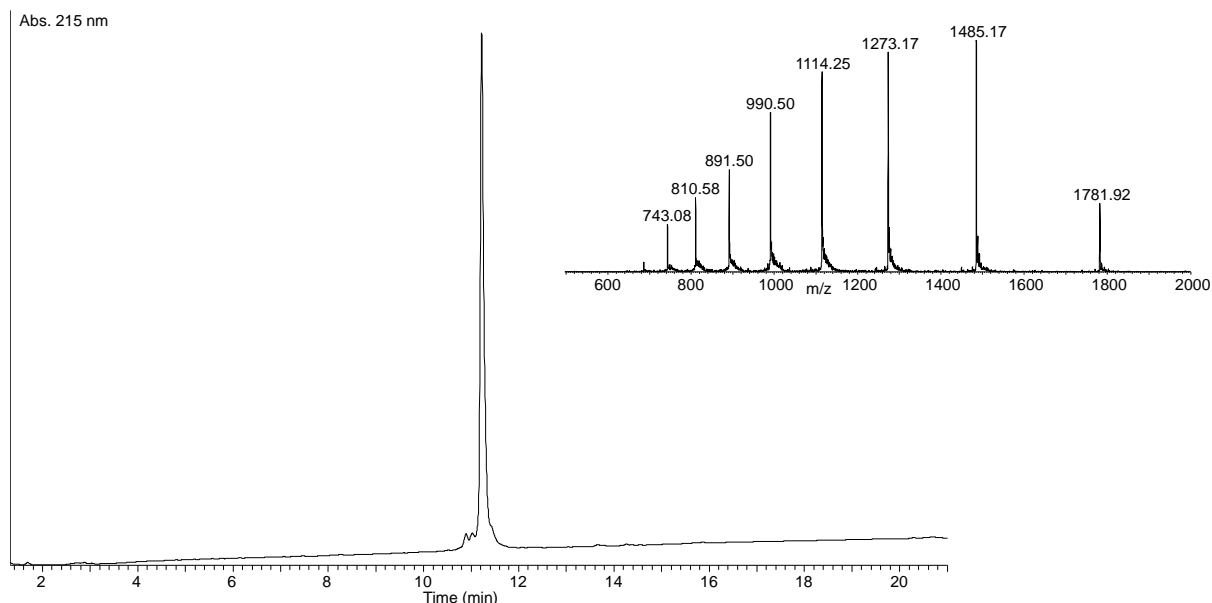


Figure S 14. Analysis of SUMO-2/3 core C48A. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 8905.11, observed 8905.07 after deconvolution.

4. Desulfurization of SUMO-2/3 conjugates

4.1. Desulfurization of conjugate **7a** in native conditions

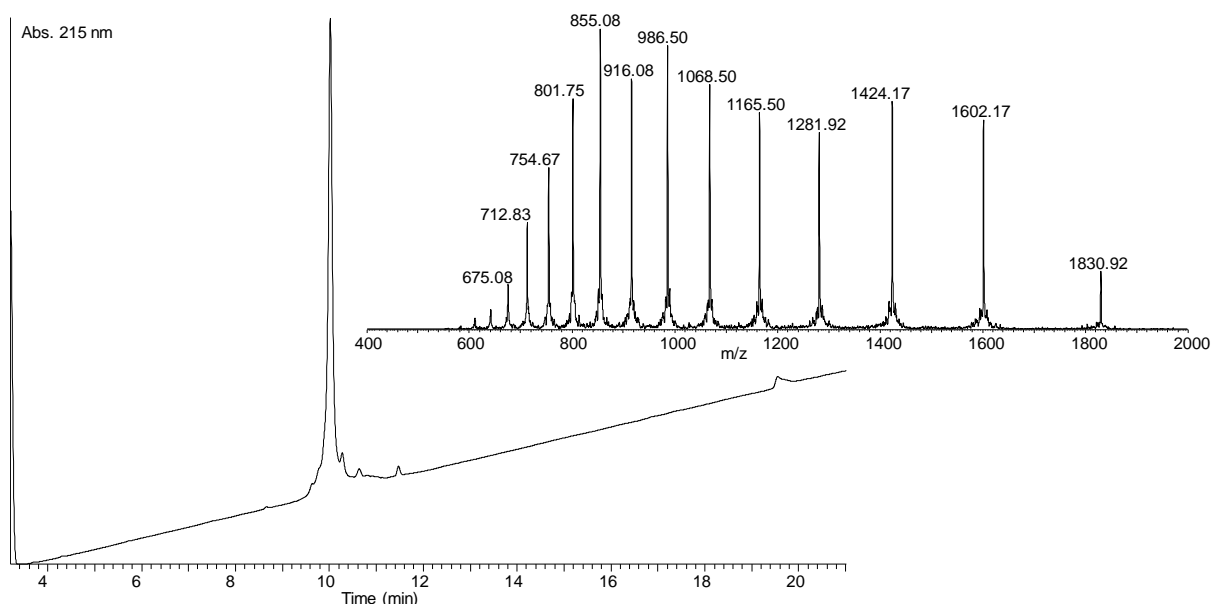
Conjugate **7a** (0.61 mg, 0.46 µmol, 0.10 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.5 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.4 mg/mL, 50 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer (3492 µL). The desulfurization was carried out at 25 °C and monitored by MALDI-TOF mass spectrometry.

After 25 h, the reaction product was purified by HPLC to yield 311 µg (51 %) of conjugate **8a**. Semi-preparative HPLC conditions: XBridge BEH300 C18 (5 µm, 300 Å, 10 × 250 mm) column, eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-10 % B in 5 min, 10-100 % B in 15 min flow rate 6 mL/min, UV detection at 215 nm, 50 °C.

SUMO-2 conjugate **8a**:

ADEKPK(ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCER

QGLSMRQIRFRFDGQPINETDTPAQLMEDEDTIDVFQQQTGGA)EGVKTENNDHINK-NH₂



*Figure S 15 Analysis of conjugate **8a**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 12809.32, observed 12809.75 after deconvolution.*

4.2. Desulfurization of conjugate **7b** in native conditions

Conjugate **7b** (0.9 mg, 0.7 µmol, 0.1 mM final concentration) was dissolved in a solution of TCEP (57.4 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.5 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.4 mg/mL, 50 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer (935 µL). The desulfurization was carried out at 25 °C and monitored by MALDI-TOF mass spectrometry.

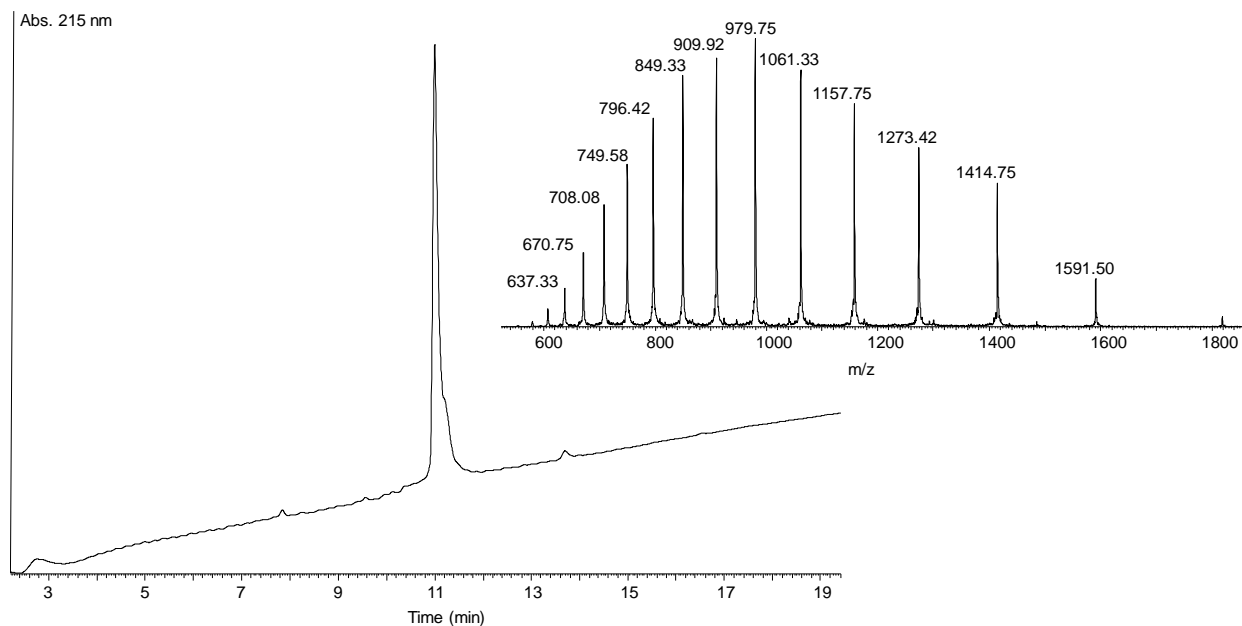
After 25 h, the reaction product was purified by HPLC to yield 371 µg (37 %) of pure conjugate **8b**.

Semi-preparative HPLC conditions: XBridge BEH300 C18 (5 µm, 300 Å, 10 × 250 mm) column, eluent A water containing 0.1% of TFA, eluent B CH₃CN/water: 4/1 by vol containing 0.1% of TFA, gradient: 0-10 % B in 5 min, 10-100 % B in 15 min flow rate 6 mL/min, UV detection at 215 nm, 50 °C.

SUMO-3 conjugate **8b**:

ADEKPK(SEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCER

QGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGA)EGVKTENNDHINK-NH₂



*Figure S 16. Analysis of conjugate **8b**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace Calculated for M (average mass) 12725.18, observed 12723.80 after deconvolution.*

C. Chemical synthesis of SUMO-2/3 Lys(Cys) proteins

A typical procedure is illustrated with the preparation of peptide **16a**.

1. One-pot synthesis and characterization of SUMO-2 Lys₁₁(Cys) **16a**

1. Oxidation of SEA disulfide into SEA cyclic sulfide

A solution of *N*-octylglucoside (6.90 mg, 20 mM) in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer (1.180 mL) was prepared (solution A). Solution A (501 µL) was then used to dissolve 4-mercaptophenylacetic acid (MPAA, 16.85 mg, 200 mM) to give solution B. Then 4-mercaptophenylacetic acid disulfide (MPAA ox, 0.183 mg, 5 mM) was dissolved in solution B (109 µL) and the pH was adjusted to 7.2 by addition of aqueous NaOH 6 N to give solution C. Peptide **11a** (2.45 mg, 0.364 µmol, 5 mM) was solubilized in solution C (72 µL) and the mixture was stirred for 60 min to give peptide **12a**.

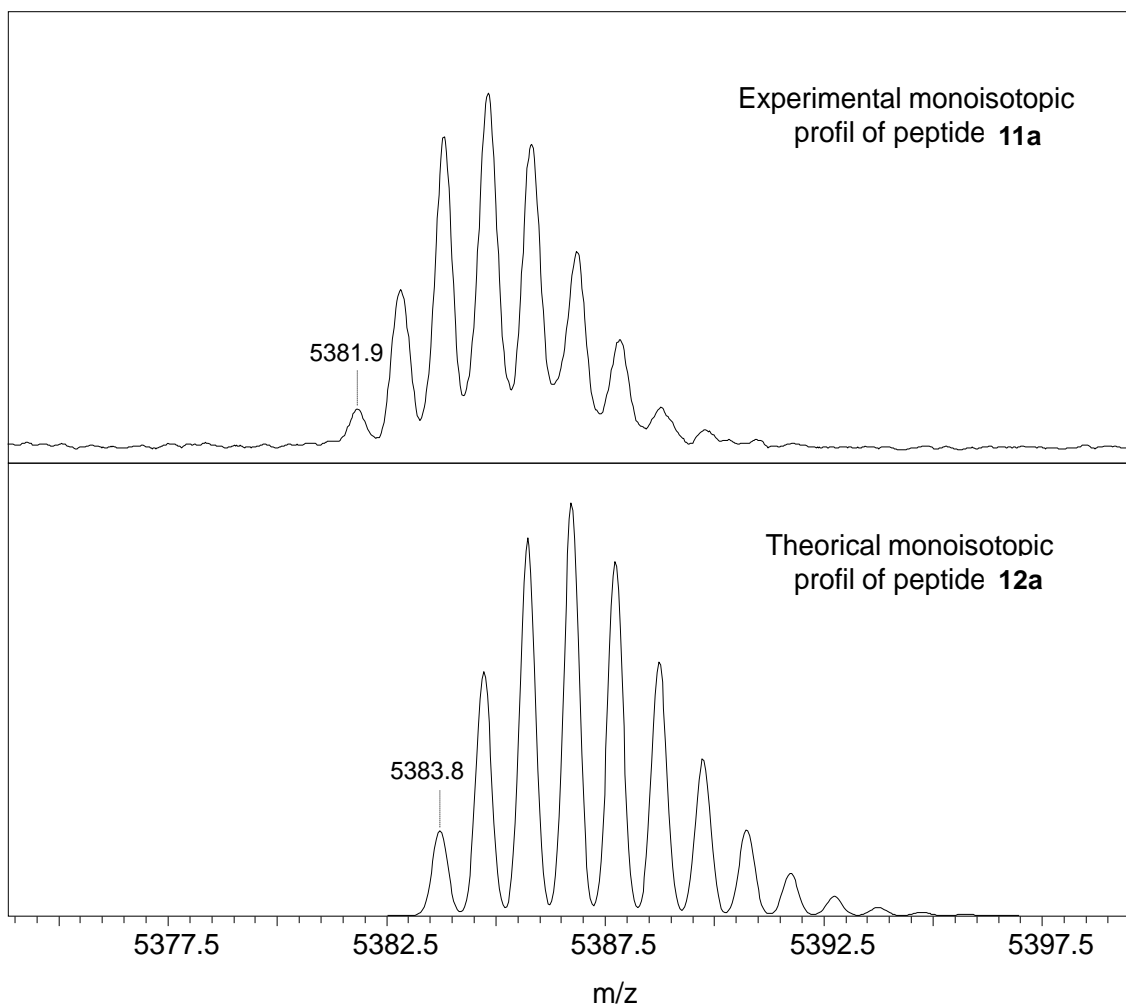


Figure S 17. MALDI-TOF analysis peptide **12a**. Matrix 2,5-dihydroxybenzoic acid (DHB), positive detection mode, $[M+H]^+$ calcd. (monoisotopic) 5381.9, found 5380.7.

2. Protection of Lys₁₁(Cys) with AcA

Protection of Lys₁₁(Cys) residue was performed by adding acetoacetyl-methylthioglycolate (AcA-MTG, 1.00 eq, 0.364 μ mol, 7 μ L of a 10 mg/mL solution in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer) to the above peptide solution. The reaction was shaken for 23 h to give solution D.

3. Ligation

TCEP (29.02 mg, 101.2 μ mol) and MPAA (16.43 mg, 97.68 μ mol) were dissolved in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer (506 μ L). Peptide **4** (2.14 mg, 0.364 μ mol, 20 mM) was dissolved in the TCEP/MPAA solution (68 μ L) and the pH of was adjusted to 5.5. Then, this solution was added to solution D. The pH of the resulting mixture was adjusted again to 5.5 by addition of aqueous NaOH 6 N (3 μ L). The final peptide concentration was 2.1 mM. The reaction was shaken for 90 h to give solution E.

4. AcA removal

To remove AcA protecting group, a 10 mg/mL hydroxylamine hydrochloride (NH₂OH.HCl) solution in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer was prepared. The

solution is acidic due to the presence of $\text{NH}_2\text{OH}\cdot\text{HCl}$. This solution (0.73 μmol , 2.0 eq, 5.0 μL) was added to solution E. The deprotection reaction was achieved in 30 min.

5. Purification

The crude mixture was purified by semi-preparative HPLC using XBridge BEH300 C18 column (5 mm, 300 Å, 10 × 250 mm), eluent A water containing 0.1% of TFA, eluent B CH_3CN /water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25 % B in 5 min, 25-45 % B in 50 min, flow rate 6 mL/min, UV detection at 215 nm, 65 °C to give **6d** (1.58 mg, 60% overall isolated).

2. Characterization of SUMO-2/3 $\text{Lys}_{11}(\text{Cys})$ **16a-b**

Peptide **16a**:

ADEKPKEGVK(C)TENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

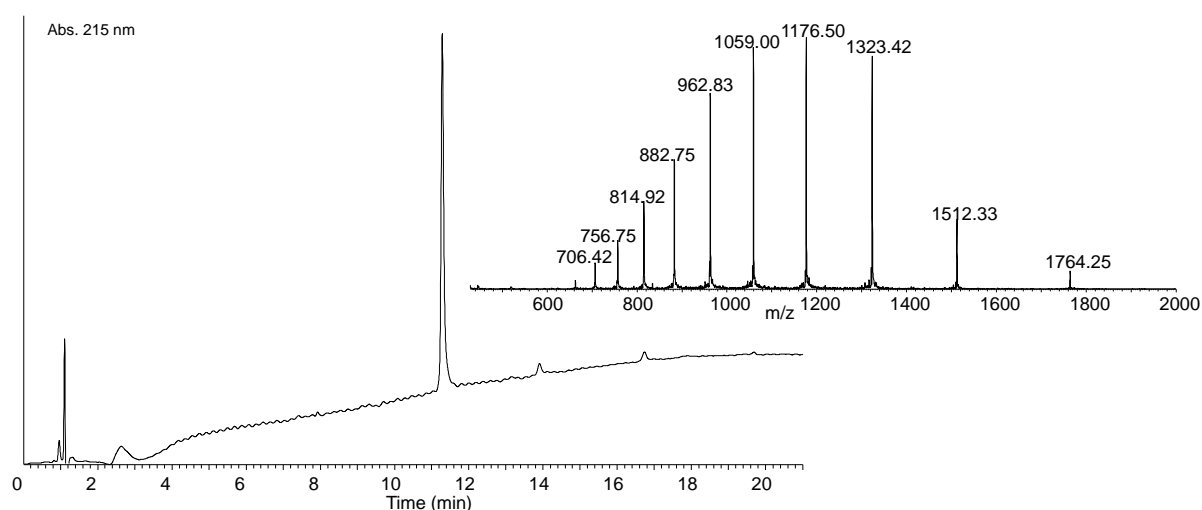


Figure S 18. Analysis of SUMO-2 $\text{Lys}_{11}(\text{Cys})$ **16a**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 μm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 10580.88, observed 10579.75 after deconvolution.

Peptide 16b:
ADEKPKEGVK(C)TENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIR
FRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

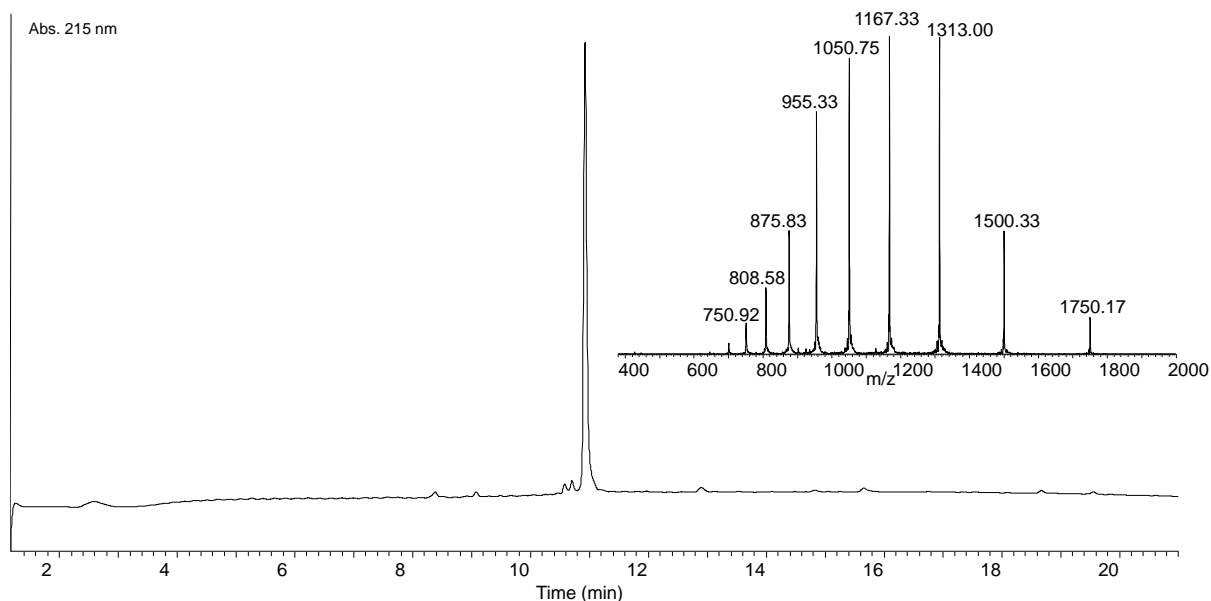


Figure S 19. Analysis of SUMO-3 Lys₁₁(Cys) **16b**. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 2.1 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (UV detection) and MS trace. Calculated for M (average mass) 10496.80, observed 10496.49 after deconvolution.

D. Chemical synthesis of SUMO-2/3 dimers

A typical procedure is illustrated with the synthesis of SUMO-2/3 dimer **20**.

1. Synthesis of SUMO dimer **11**

Step 1. Synthesis of SUMO-2 SEA^{off} **5a**

A solution of *N*-octylglucoside (6.90 mg, 20 mM) in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer (1.180 mL) was prepared (solution A). Solution A (501 µL) was used to dissolve 4-mercaptophenylacetic acid (MPAA, 16.85 mg, 200 mM) and the pH of the mixture was then adjusted to 7.2 with 6 N NaOH to give solution B.

Thioester peptide **3a** (2.04 mg, 0.308 µmol, 5 mM) and peptide **4** (1.85 mg, 0.308 µmol, 5 mM, 1 equiv) were dissolved in solution B (62 µL) and the reaction mixture was stirred at 25 °C for 25 h under nitrogen atmosphere to give the solution C.

The progress of the reaction was monitored by HPLC. For each analysis, a 1.5 µL aliquot was taken from the reaction mixture and quenched by adding 100 µL of 10 % acetic acid. The sample was then extracted with Et₂O to remove MPAA prior to HPLC analysis.

Step 2. Assembly of the dimer

A fresh solution of *N*-octylglucoside (6.90 mg, 20 mM) in 6 M guanidinium chloride/0.1 M pH 7.0 sodium phosphate buffer (1.180 mL) was prepared again (solution D). Solution D (421 μ L) was used to dissolve TCEP (24.14 mg, 84.20 μ mol, 200 mM) to give the solution E. Then MPAA (14.16 mg, 84.20 μ mol, 200 mM) was dissolved in solution E (421 μ L) and the pH of the solution was adjusted to 5.5 with 6 N NaOH to give solution F. Peptide **16a** (3.8 mg, 0.307 μ mol, 5 mM) was dissolved in solution F (60 μ L) to give solution G.

Solutions C and G were mixed and the pH was adjusted to 5.5 by addition of 6 N NaOH. The final peptide concentration was 2.5 mM. The reaction was agitated for 96 h and then acidified with glacial acetic acid (10 μ L) to quench the reaction. The solution was extracted with diethyl ether (3 \times 2 mL) to remove the excess of MPAA. The mixture was immediately purified by HPLC to yield 1.85 mg (36% overall isolated) of SUMO dimer **18**, 0.836 mg (22% overall isolated) of SUMO dimer **19**, 1.102 mg (29% overall isolated) of SUMO dimer **20** and 0.874 mg (23% overall isolated) of SUMO dimer **21**.

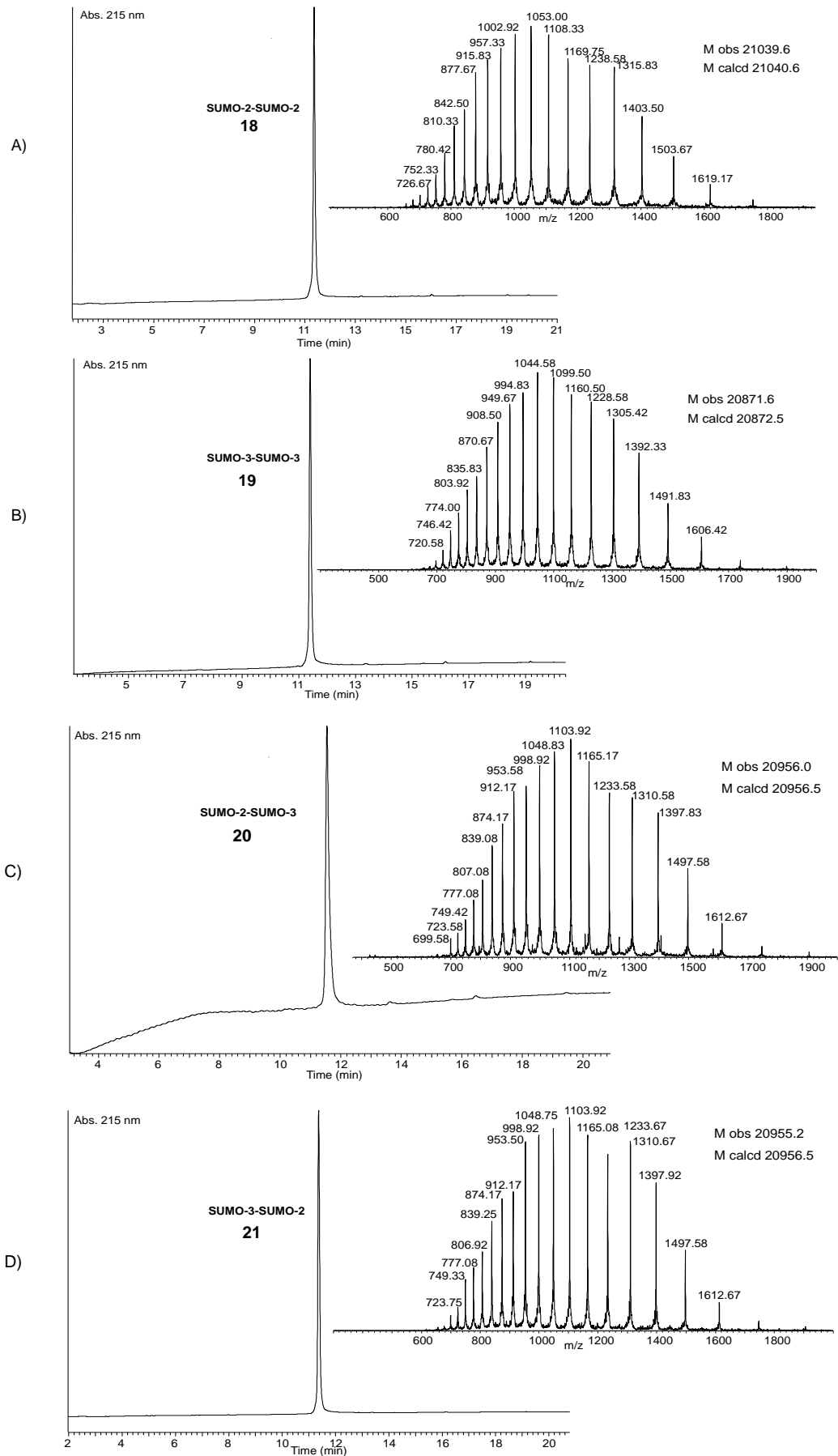


Figure S 20. LC-MS analysis of isolated SUMO-2/3 dimers 18-21.

2. Desulfurization of SUMO-2/3 dimers

The desulfurization protocols in native or denaturing conditions were optimized and validated using SUMO-2/3 dimer **20**.

2.1. Desulfurization of SUMO-2/3 dimer **20** in denaturing conditions

SUMO-2/3 dimer **20** (0.130 mg, 6.20 μ mol, 0.048 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.50 mg/mL, 20 mM final concentration), methionine (Met, 2.98 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.40 mg/mL, 50 mM final concentration) in 6 M guanidinium chloride/ 0.1 M pH 7.2 ammonium phosphate buffer (3492 μ L). The desulfurization was carried out at 25 °C and monitored by UPLC-MS. The reaction product was SUMO-2/3 dimer **27**.

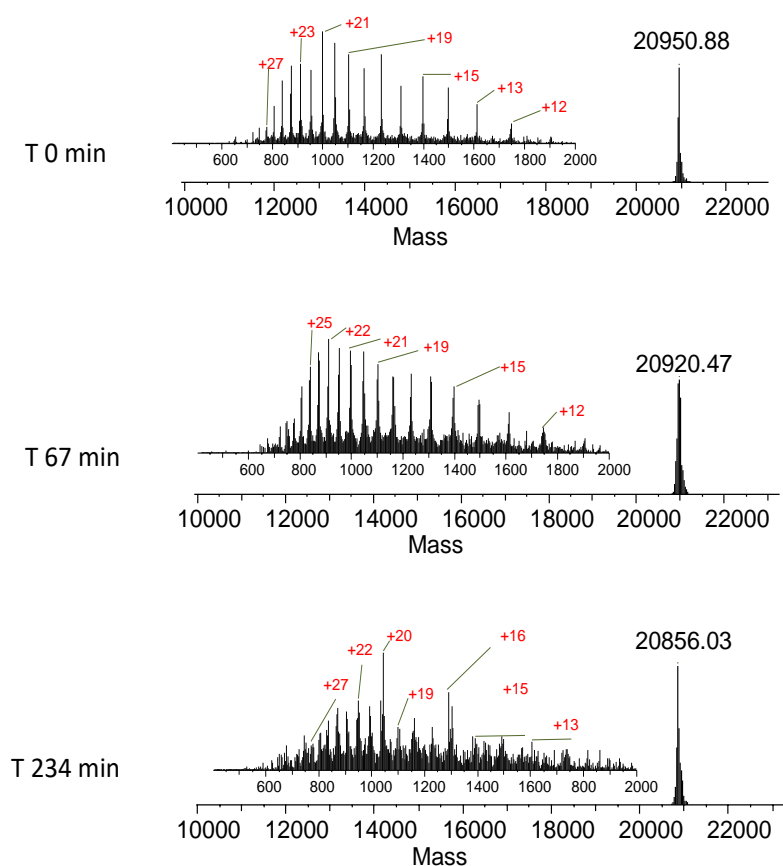


Figure S 21. Analysis of the desulfurization reaction in denaturing condition and leading to the production of SUMO-2/3 dimer **20**. UPLC-MS analysis using ATLANTIS waters T3column (3 μ m 2.1 mm \times 50 mm). 50 °C, flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% aqueous acetonitrile. Gradient from 0% eluent B to 70% eluent B in 4 min. MS trace.

2.2. Desulfurization of SUMO-2/3 dimer **20** in native conditions

SUMO-2/3 dimer **13** (0.136 mg, 6.49 μ mol, 0.048 mM final concentration) was dissolved in a solution of TCEP (57.3 mg/mL, 200 mM final concentration), 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044, 6.50 mg/mL, 20 mM final concentration), methionine (Met, 2.98 mg/mL, 20 mM final concentration) and reduced glutathione (GSH, 15.40 mg/mL, 50 mM final concentration) in 0.1 M pH 7.2 ammonium phosphate buffer (3492 μ L). The desulfurization was carried out at 25 °C and monitored by UPLC-MS. The reaction yielded selectively SUMO-2/3 dimer **25**.

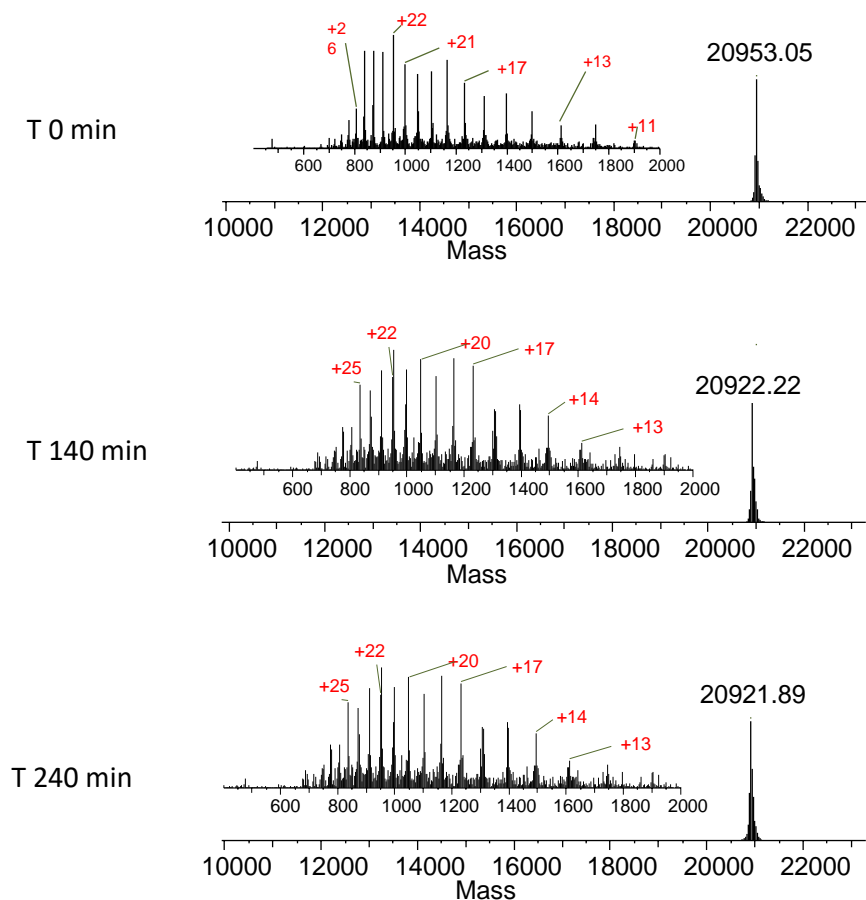


Figure S 22. Analysis of the desulfurization reaction in native conditions. UPLC-MS analysis using ATLANTIS waters T3column (3 μ m, 2.1 mm \times 50 mm), 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% aqueous acetonitrile. Gradient from 0% eluent B to 70% eluent B in 4 min. MS trace.

This protocol was applied to the selective desulfurization of SUMO-2/3 dimers **18-21**.

IV. Proteomics analysis

Tryptic digestion

The proteins were dissolved in 25 mM ammonium bicarbonate pH 7.8 (1 mg/mL). The protein solution (5 μ L, 5 μ g) was mixed with DTT (5 mg/mL, 1 μ L, 5 μ g) and iodoacetamide (10 mg/mL, 5 μ L, 50 μ g). After 30 min, trypsin (1 μ L, 500 ng) solution in 25 mM ammonium bicarbonate (0.5 mg/mL) was added to the protein which was digested at 37 °C overnight.

Endopeptidase GluC digestion

The proteins were dissolved in 25 mM ammonium phosphate pH 7.5 (1 mg/mL). The protein solution (5 μ L, 5 μ g) was mixed with DTT (5 mg/mL, 1 μ L, 5 μ g) and iodoacetamide (10 mg/mL, 5 μ L, 50 μ g). After 30 min, endopeptidase GluC (1 μ L, 500 ng) solution in 25 mM ammonium phosphate (0.5 mg/mL) was added to the above solution and was kept at room temperature overnight.

The digested protein solutions were directly spotted on a MALDI plate and mixed with the matrix (α -cyano-4-hydroxycinnamic acid or 2, 5-dihydroxybenzoic acid) before analysis.

A. Proteomic analysis of r1-SUMO-2 & 3 Boston Biochem (ref. UL-752 & UL-762)

r₁-SUMO-2:

**ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMAKYCERQGLSMRQIRFR
FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG**

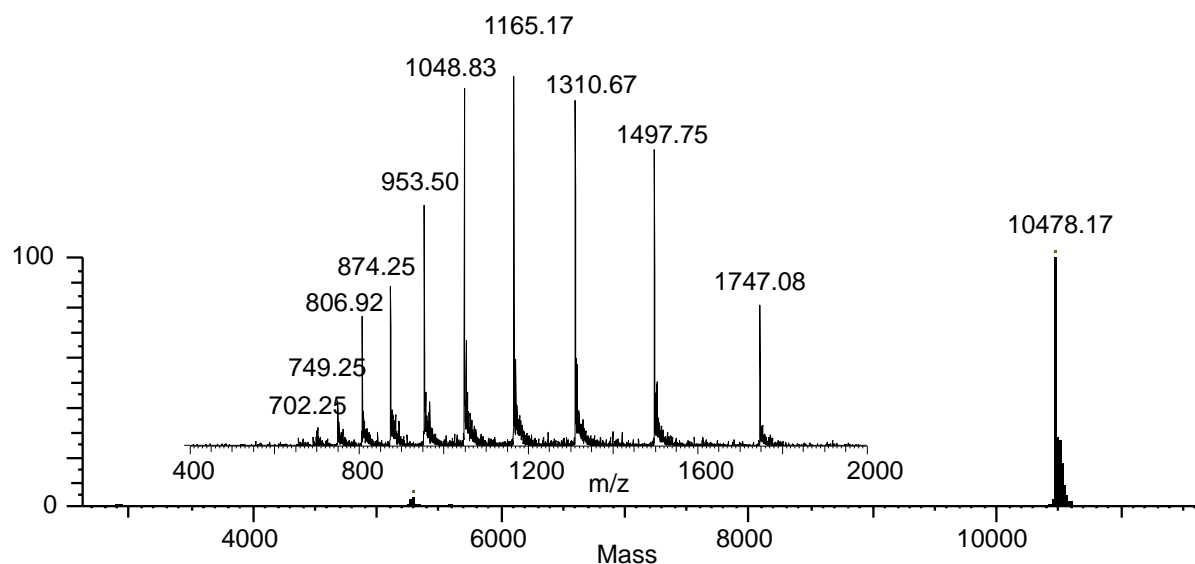


Figure S 23. Analysis of r1-SUMO-2 protein. UPLC-MS analysis ATLANTIS waters column T3 3 μ m 2.1 mm \times 50 mm, 50 $^{\circ}$ C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 4 min. MS trace. Calculated for M (average mass) 10477.19 observed 10478.17 after deconvolution.

Search Parameter: Tit...							
✓ peak 7	687.34	687.37	-41.35	1 - 6	0	ADEKPK	
✓ peak 24	1197.65	1197.59	49.91	11 - 20	0	TENNDHINLK	
✓ peak 26	1234.71	1234.64	50.61	21 - 32	0	VAGQDGSVVQFK	
✓ peak 20	838.52	838.49	31.77	35 - 41	1	RHTPLSK	
✓ peak 6	682.36	682.39	-46.45	36 - 41	0	HTPLSK	
✓ peak 1	641.24	641.27	-55.99	45 - 49	0	AYCER	
✓ peak 8	691.34	691.36	-21.17	50 - 55	0	QGLSMR	
✓ peak 29	3871.77	3871.72	13.11	59 - 92	1	FRFDGQPINETDTPAQLEMEDEDITIDVFQQQTGG	

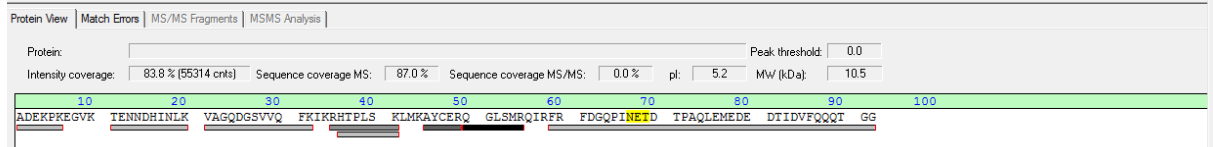
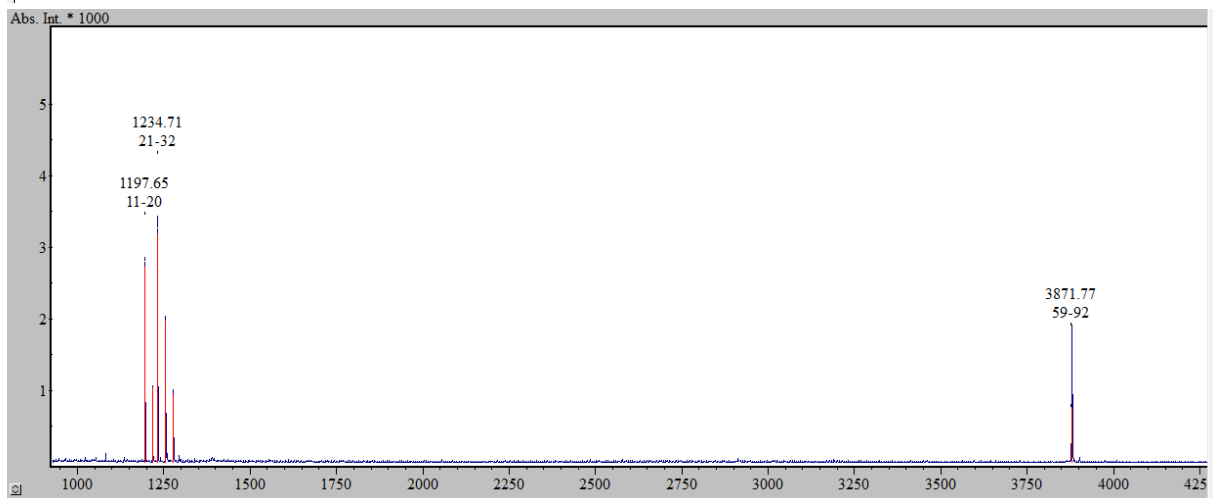
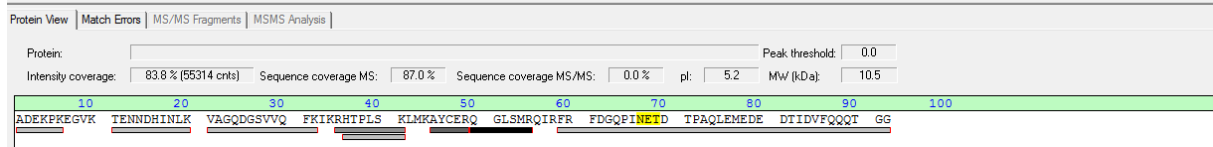
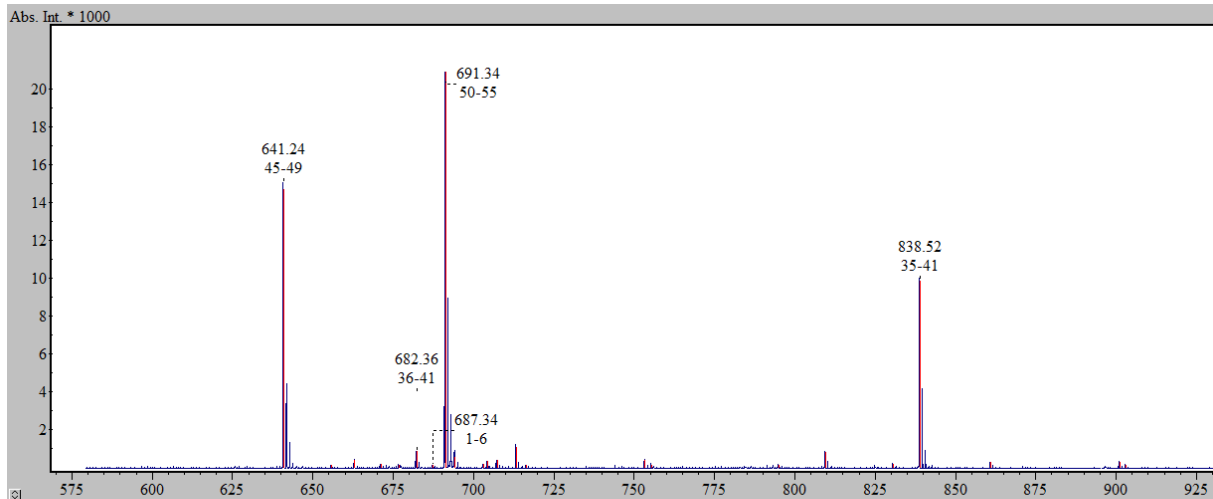


Figure S 24. MALDI-TOF spectra of the fragments generated after reduction, alkylation and trypsin cleavage of r1-SUMO-2 (ref. UL-752).

r₁-SUMO-3:

**SEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMAKYCERQGLSMRQIRFRF
DGQPINETDTPAQLEMEDEDTIDVFQQQTGG**

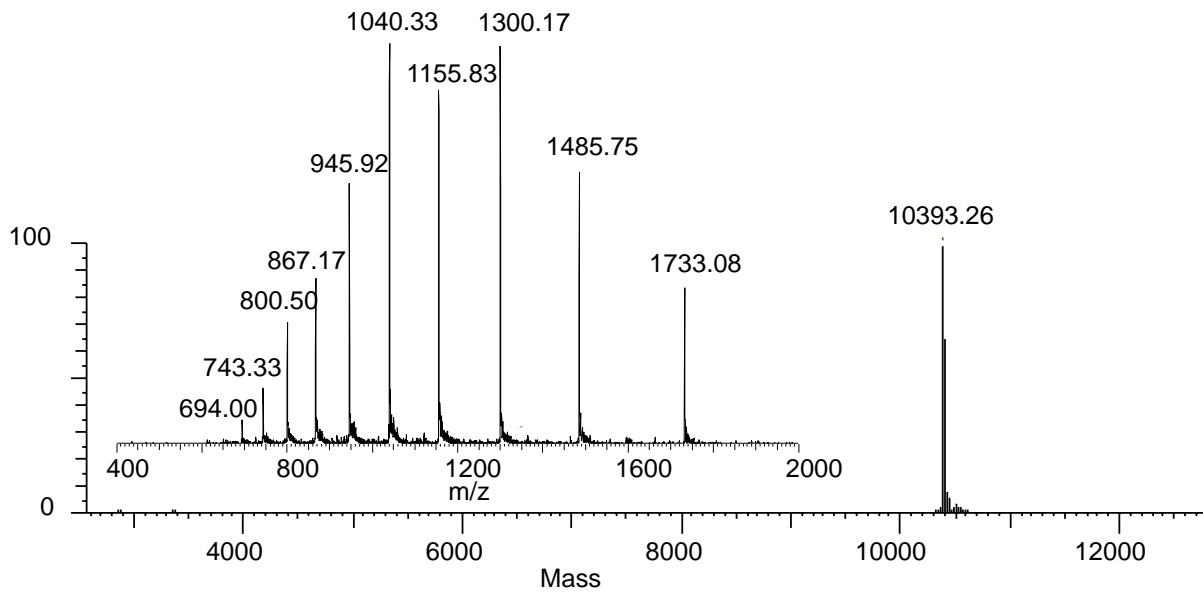
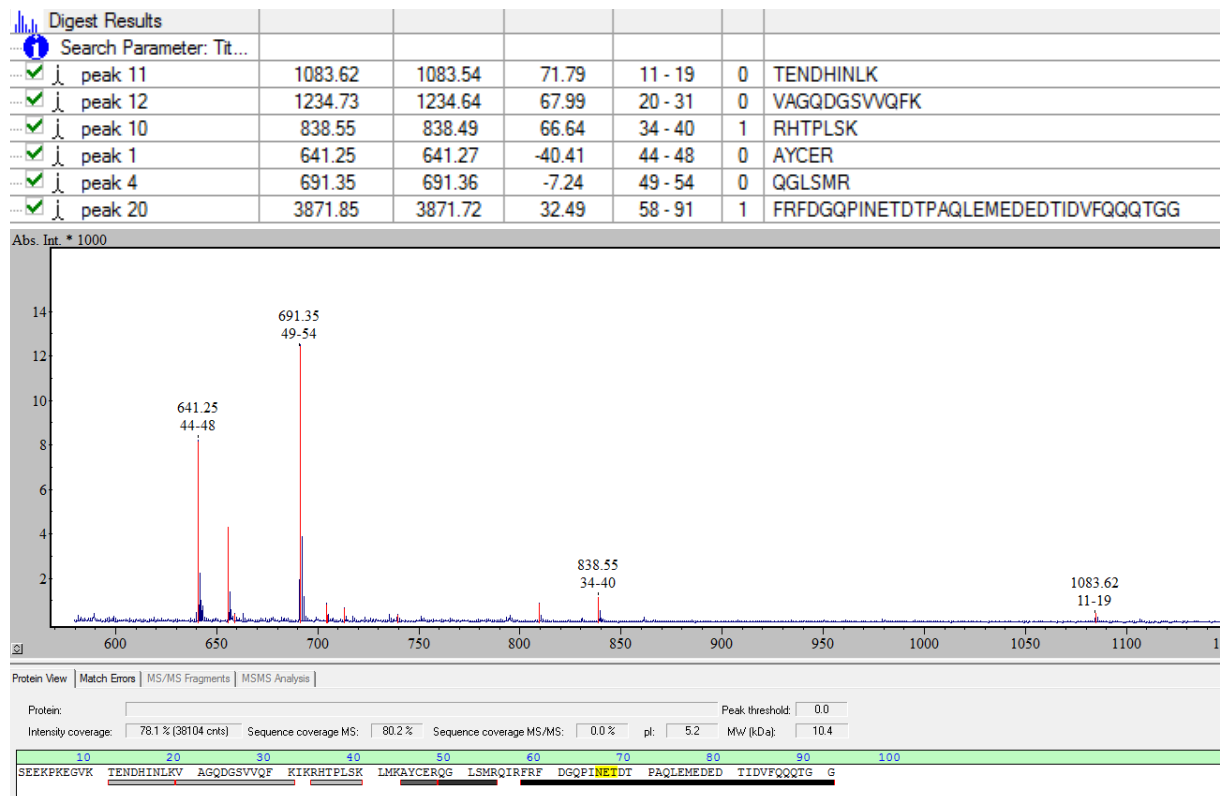


Figure S 25. Analysis of r₁-SUMO-3 protein. UPLC-MS analysis ATLANTIS waters column T3 3 μm 2.1 mm × 50 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 4 min. MS trace. Calculated for M (average mass) 10393.66, observed 10393.26 after deconvolution.



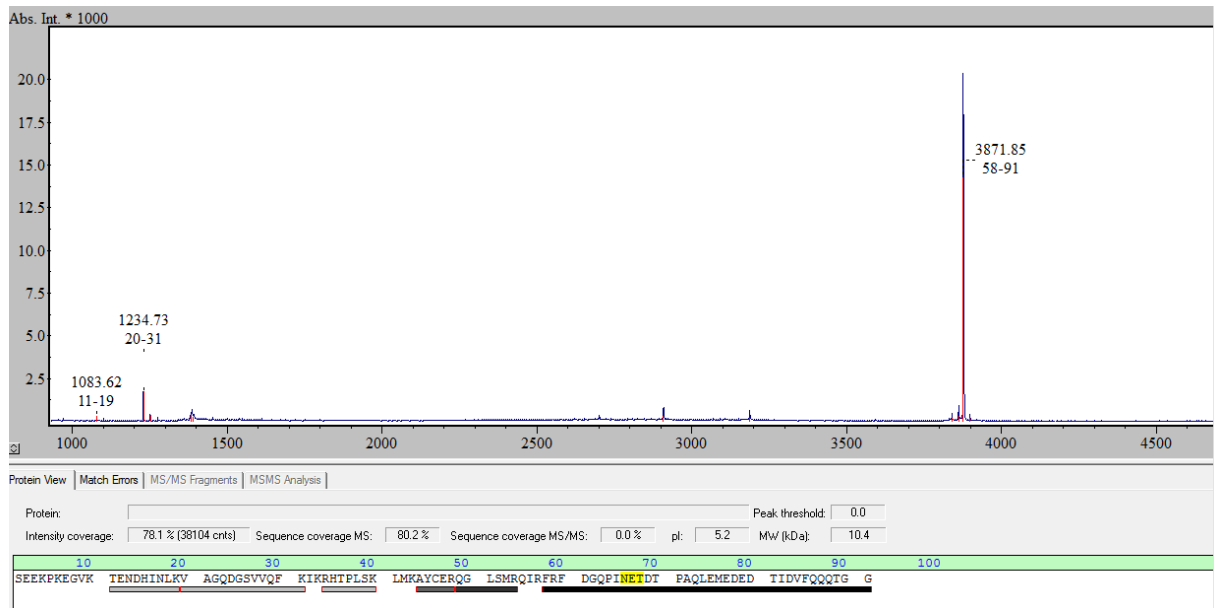


Figure S 26. MALDI-TOF spectra of the fragments generated after reduction, alkylation and trypsin cleavage of r1-SUMO-3 (ref. UL-762).

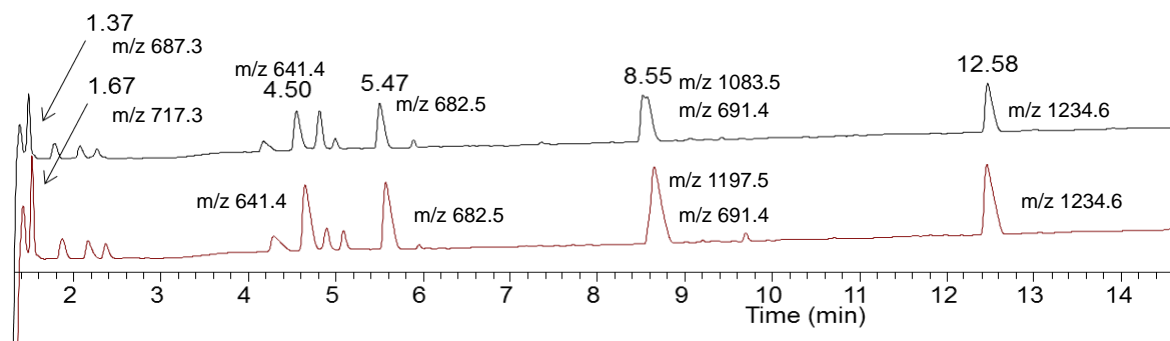


Figure S 27. Analysis of trypsin digests. UPLC-MS analysis ACQUITY UPLC peptide BEH C18 300 Å 1.7 µm 4.6 mm × 150 mm, 50 °C. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% tri-fluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min. HPLC trace (light scattering detection) and MS data, r1-SUMO-2 in red, r1-SUMO-3 in black.

B. Proteomic analysis of SUMO-2/3 conjugates 7-9a/b

1. Characterization of SUMO-3 conjugates 7-9a by proteomics analysis

5622.54 Da	Mo.			Amino Acids	interchain links: ☰
Cys:	Carbamidomethyl (C)	Left:	Hydrogen	Right:	Free Acid
New interchain set0					
FRFDG	10 QPINE	20 TDTPA	30 QLEME	DEDTI	DUFQQ QTGGC
EGUKT	10 ENNDH	INLK			

5533.55 Da	Mo.			Amino Acids	interchain links: ☰
Cys:	Carbamidomethyl (C)	Left:	Hydrogen	Right:	Free Acid
New interchain set0					
FRFDG	10 QPINE	20 TDTPA	30 QLEME	DEDTI	DUFQQ QTGGA
EGUKT	10 ENNDH	INLK			

m/z 2720.47, GSVVQFKIKRHTPLSKLMKAYC*E (*Carbamidomethyl(C))

m/z 2631.48, GSVVQFKIKRHTPLSKLMKAYAE

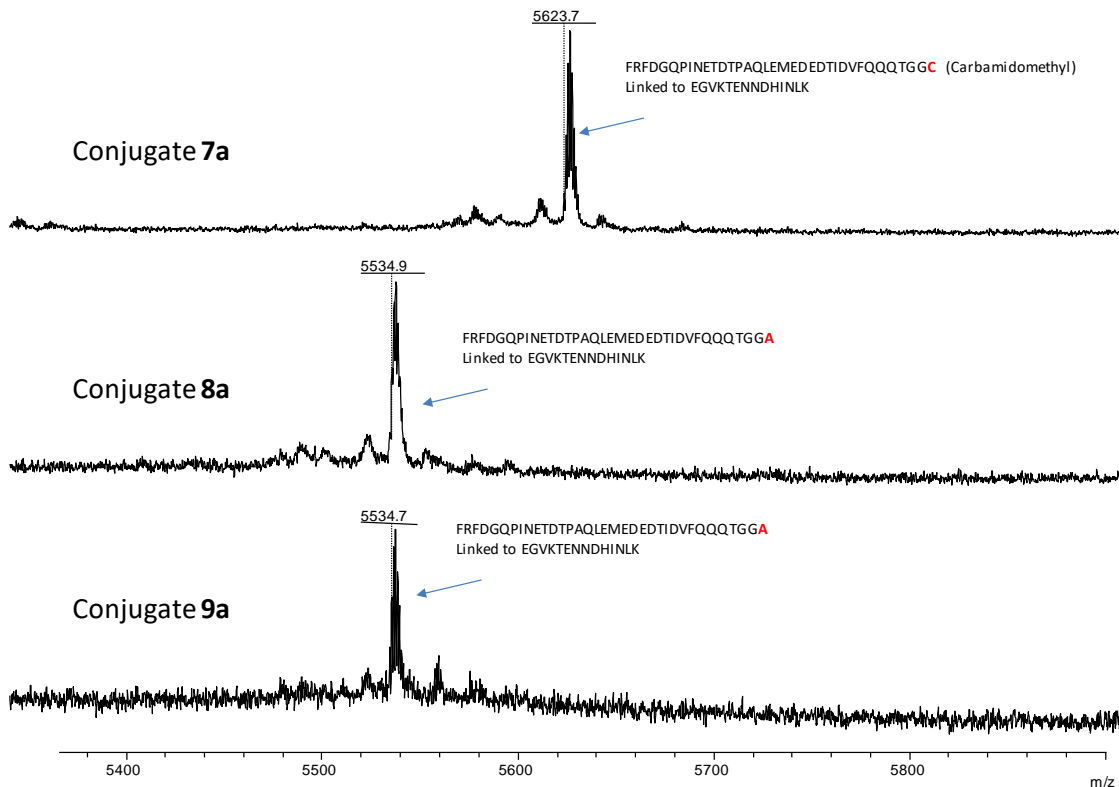


Figure S 28. MALDI-TOF spectra of the fragments generated after reduction, alkylation and trypsin cleavage of SUMO-2 conjugates 7-9a.

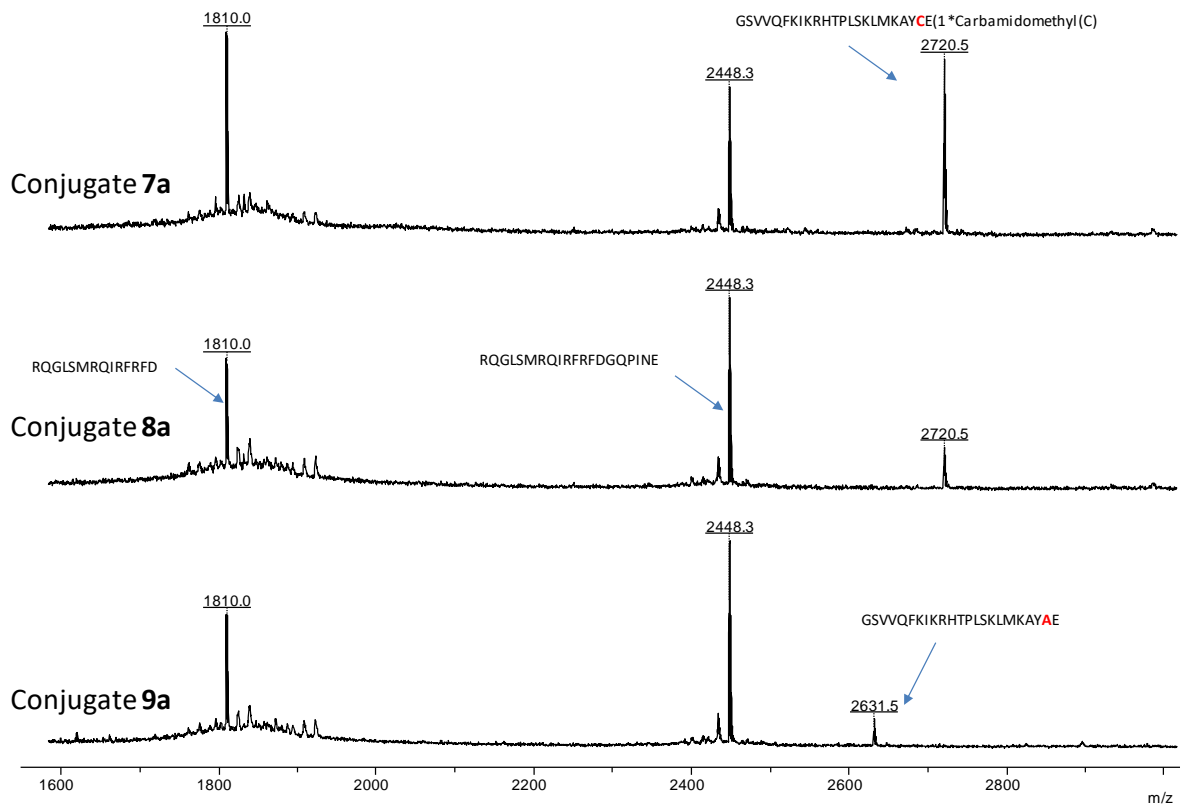


Figure S 29. MALDI-TOF spectra of the fragments generated after reduction, alkylation and Endo GluC cleavage of SUMO-2 conjugates 7-9a.

2. Characterization of SUMO-3 conjugates 7-9b by proteomics analysis

5622.54 Da	Mo.				Amino Acids	interchain links:
Cys:	Carbamidomethyl (C)	Left:	Hydrogen	Right:	Free Acid	
New interchain set0						
<input type="checkbox"/>	10	20	30			
FRFDG	QPINE	TDTPA	QLEME	DEDTI	DVFQQ	QTGGC
<input type="checkbox"/>	10					
EGUKT	ENNDH	INLK				

5533.55 Da	Mo.				Amino Acids	interchain links:
Cys:	Carbamidomethyl (C)	Left:	Hydrogen	Right:	Free Acid	
New interchain set0						
<input type="checkbox"/>	10	20	30			
FRFDG	QPINE	TDTPA	QLEME	DEDTI	DVFQQ	QTGGA
<input type="checkbox"/>	10					
EGUKT	ENNDH	INLK				

m/z 2720.47, GSVVQFKIKRHTPLSKLMKAYC*E (*Carbamidomethyl(C))

m/z 2631.48, GSVVQFKIKRHTPLSKLMKAYAE

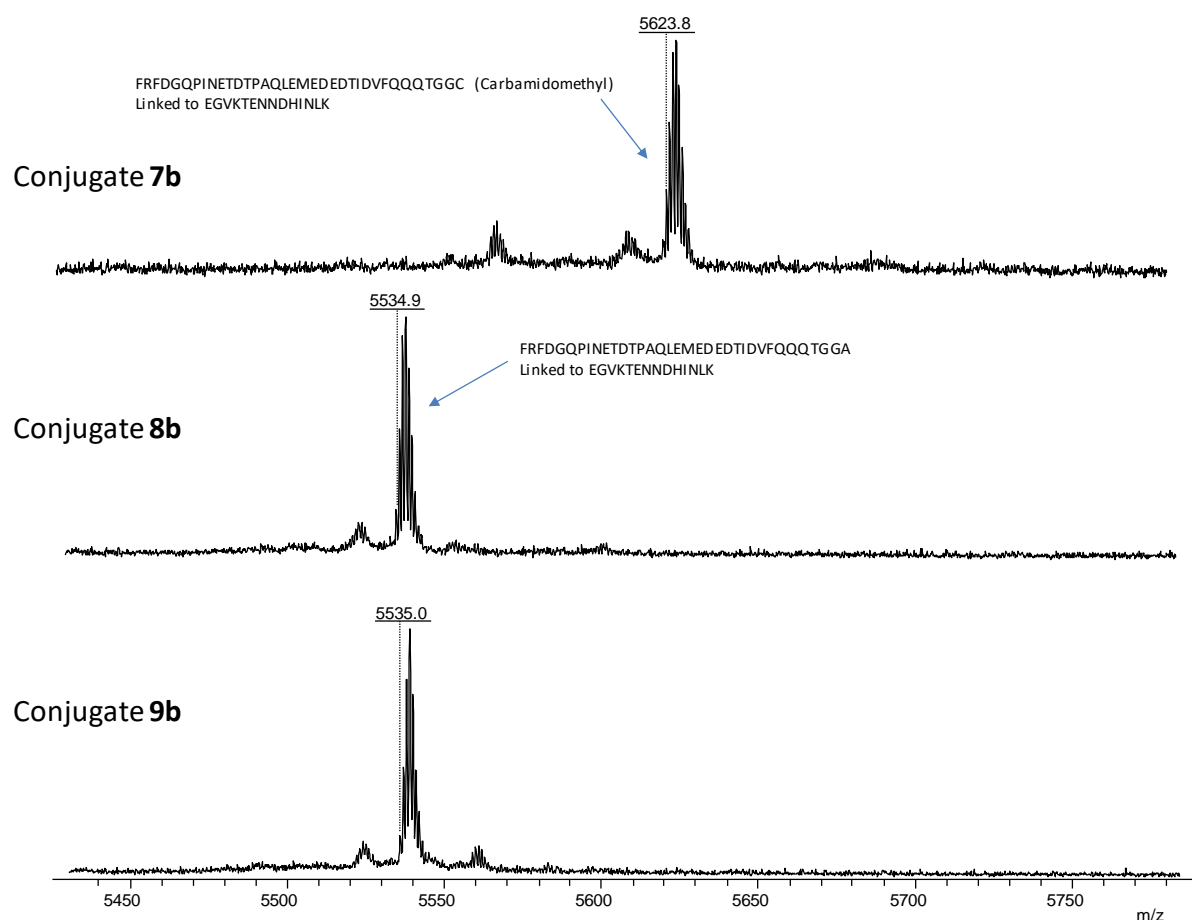


Figure S 30. MALDI-TOF spectra of the fragments generated after reduction, alkylation and trypsin cleavage of SUMO-3 conjugates 7-9b.

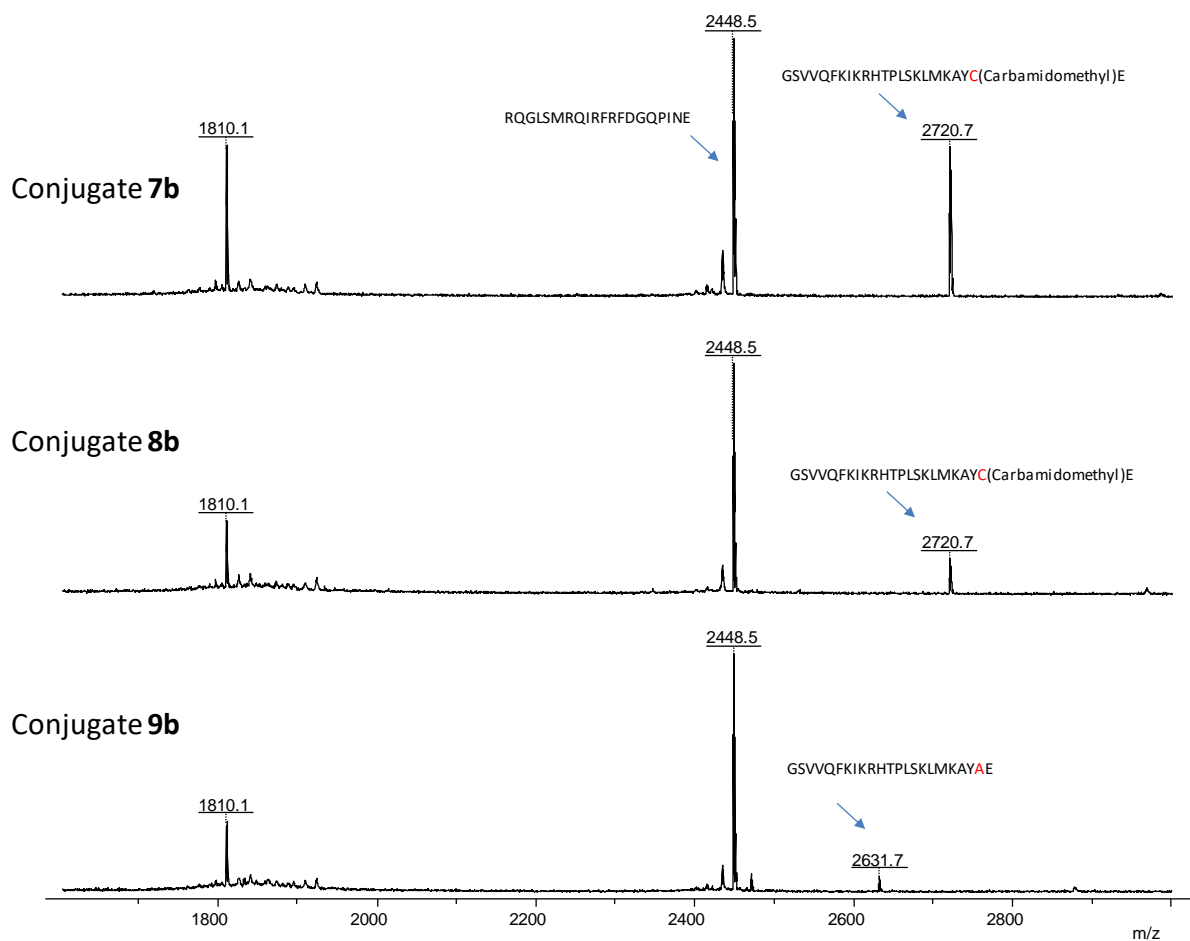


Figure S 31. MALDI-TOF spectra of the fragments generated after reduction, alkylation and Endo GluC cleavage of SUMO-3 conjugates **7-9b**.

C. Characterization of SUMO-2/3 dimers by proteomics analysis

1. Proteomic analysis of SUMO-2/3 dimer **23**

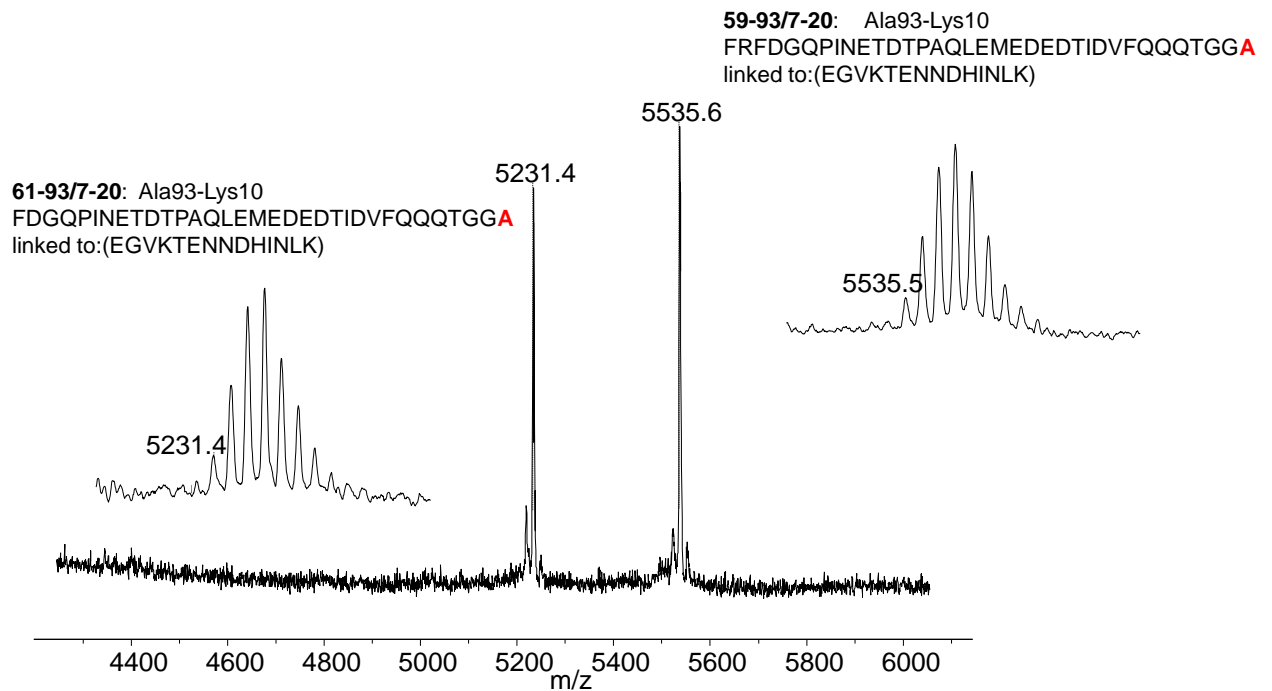


Figure S 32. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and trypsin cleavage of SUMO-2/3 dimer **23**.

Internal cysteine state

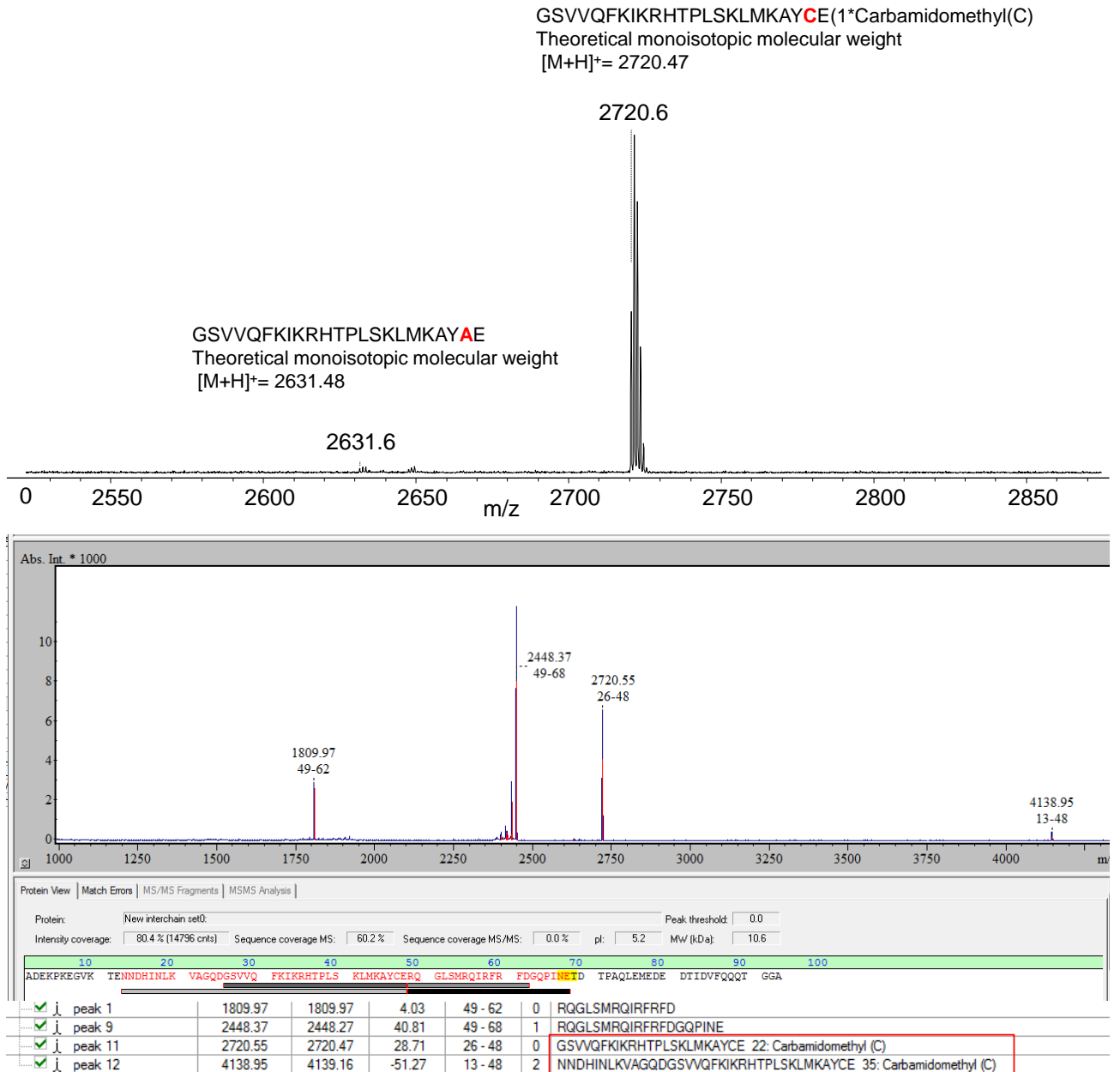


Figure S 33. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and endopeptidase GluC cleavage of SUMO-2/3 dimer 23.

2. Proteomic analysis of SUMO-2/3 dimer **24**

60-92/7-19: Ala92-Lys10

FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG**A**

linked to:(EGVKTENDHINLK)

58-92/7-19: Ala92-Lys10

FRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG**A**

linked to:(EGVKTENDHINLK)

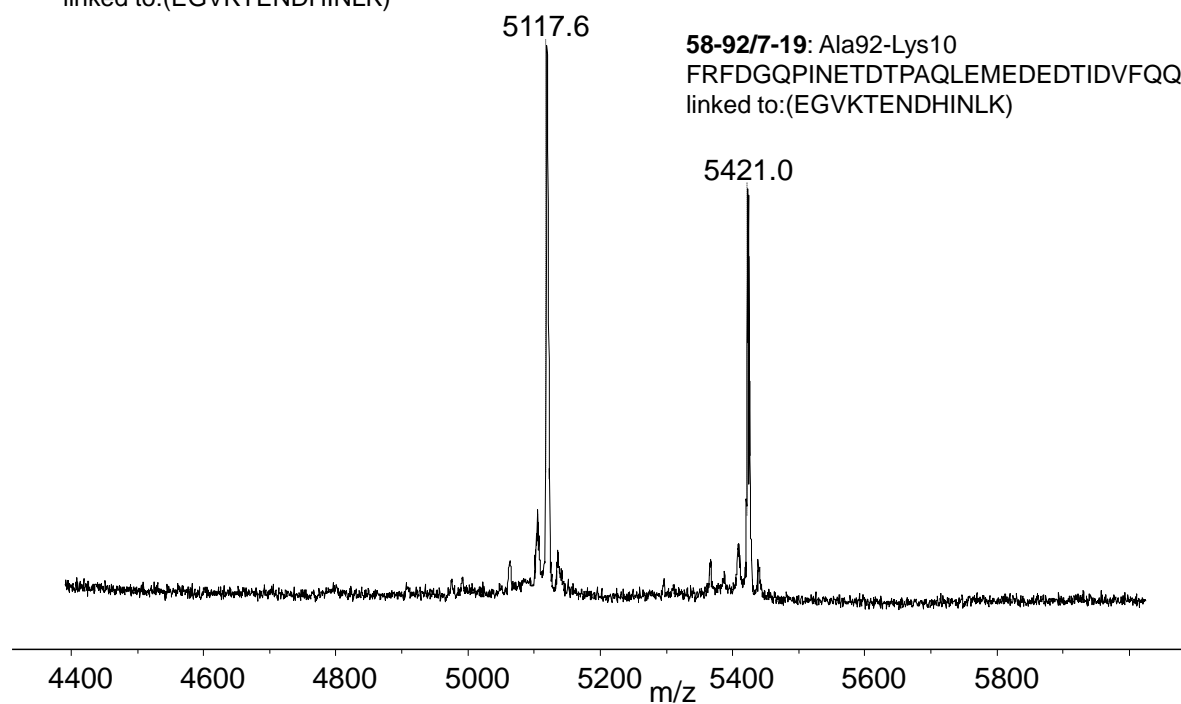
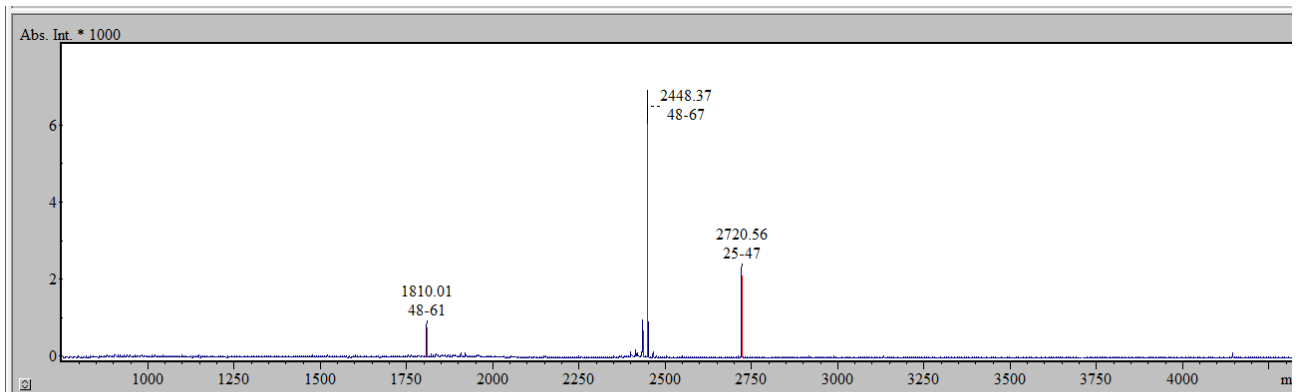
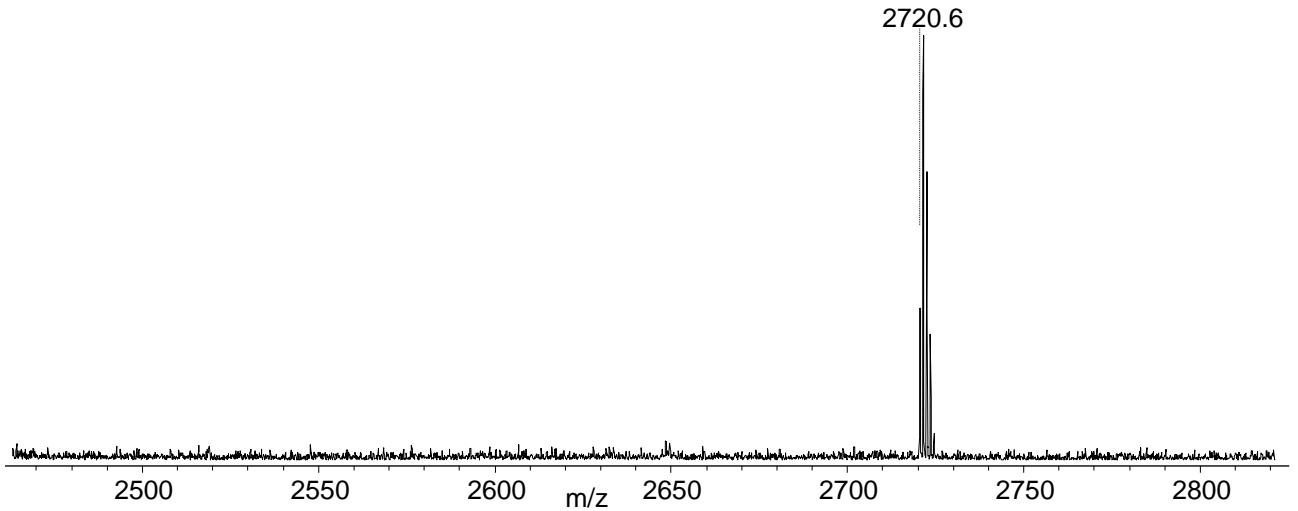


Figure S 34. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and trypsin cleavage of SUMO-2/3 dimer **24**.

Internal cysteine state

GSVVQFKIKRHTPLSKLMKAYCE(1*Carbamidomethyl)(C)
 Theoretical monoisotopic molecular weight
 [M+H]⁺= 2720.47



Protein View | Match Errors | MS/MS Fragments | MSMS Analysis

Protein: New interchain set0
 Intensity coverage: 91.4% (8858 cnts) | Sequence coverage MS: 46.7% | Sequence coverage MS/MS: 0.0% | pI: 5.2 | MW (kDa): 10.5 | Peak threshold: 0.0

10 20 30 40 50 60 70 80 90 100
 SEEKPEGVK TENDHINKV AGQDGSVVQF KIKRHTPLSK LMKAYCERQG LSMRQIRFRF DGQPINE²²DT PAQLEMEDED TIDVFQQQTG GA

Search Parameter: Tit...						
✓ j peak 4	1810.01	1809.97	25.83	48 - 61	0	RQGLSMRQIRFRFD
✓ j peak 1	2448.37	2448.27	42.44	48 - 67	1	RQGLSMRQIRFRFDGQPINE
✓ j peak 2	2720.56	2720.47	31.00	25 - 47	0	GSVVQFKIKRHTPLSKLMKAYCE 22: Carbamidomethyl (C)

Figure S 35. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and endopeptidase GluC cleavage of SUMO-2/3 dimer **24**.

3. Proteomic analysis of SUMO-2/3 dimer 25

Internal cysteine state

GSVVQFKIKRHTPLSKLMKAYCE(1*Carbamidomethyl(C)
 Theoretical monoisotopic molecular weight
 [M+H]⁺= 2720.47

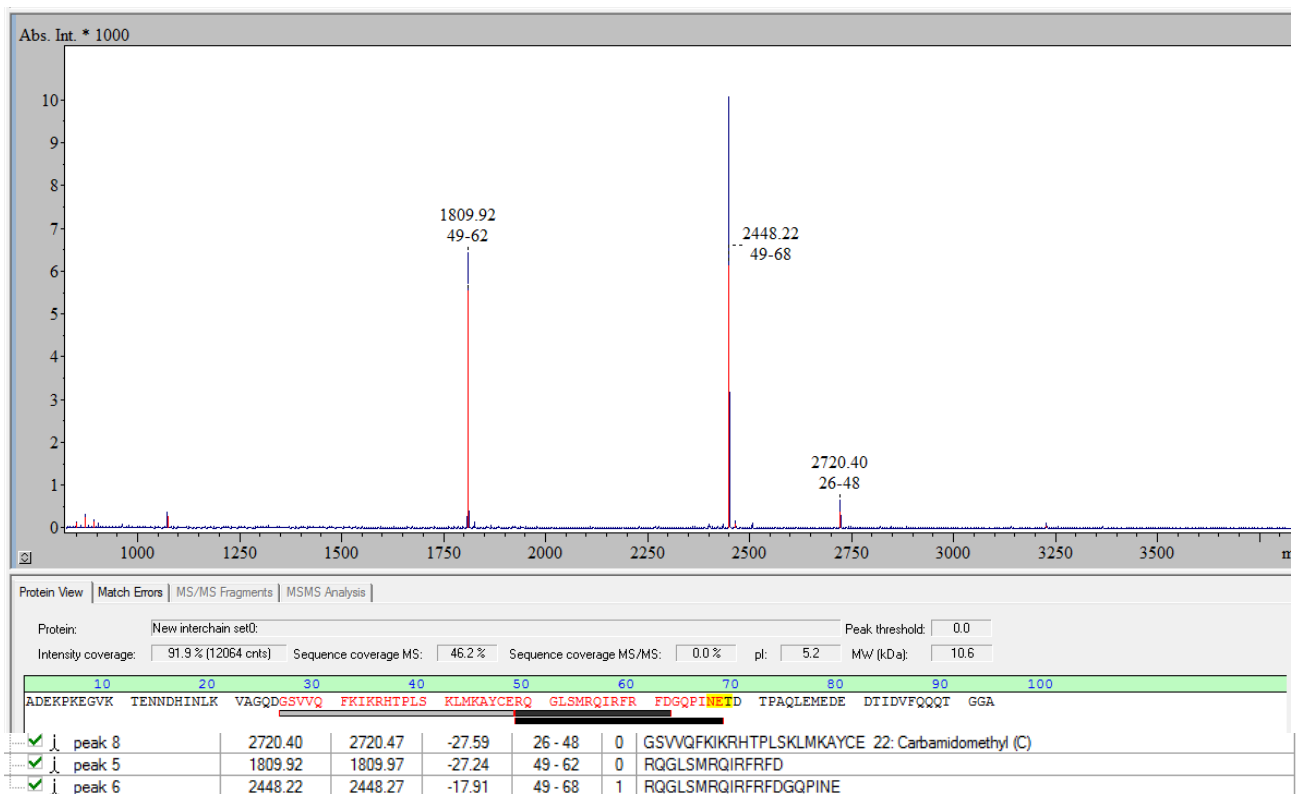
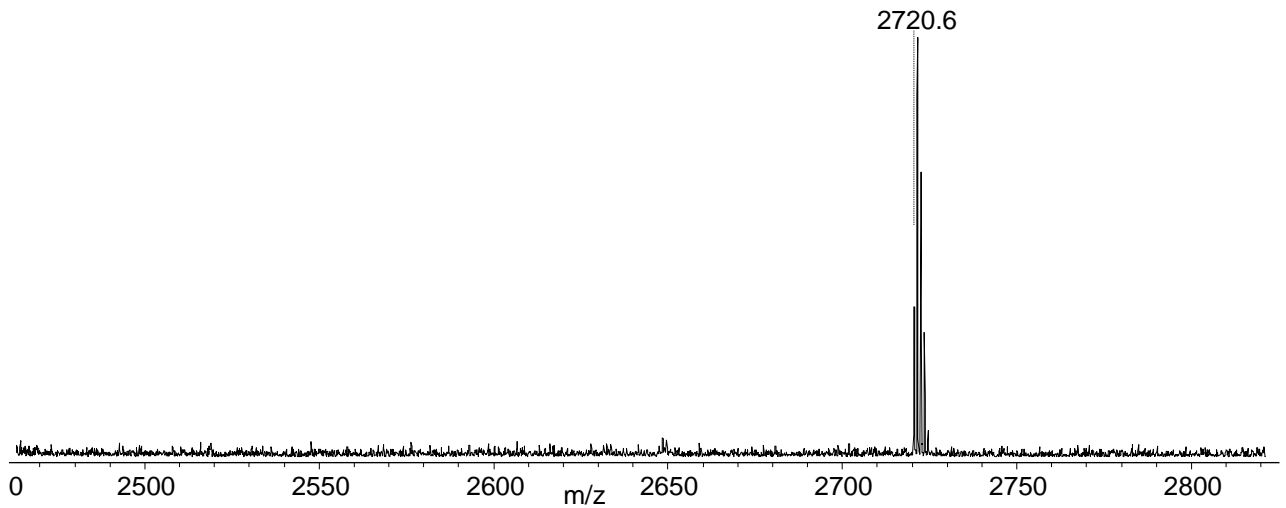


Figure S 36. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and endopeptidase GluC cleavage of SUMO-2/3 dimer 25.

4. Proteomic analysis of SUMO-2/3 dimer 26

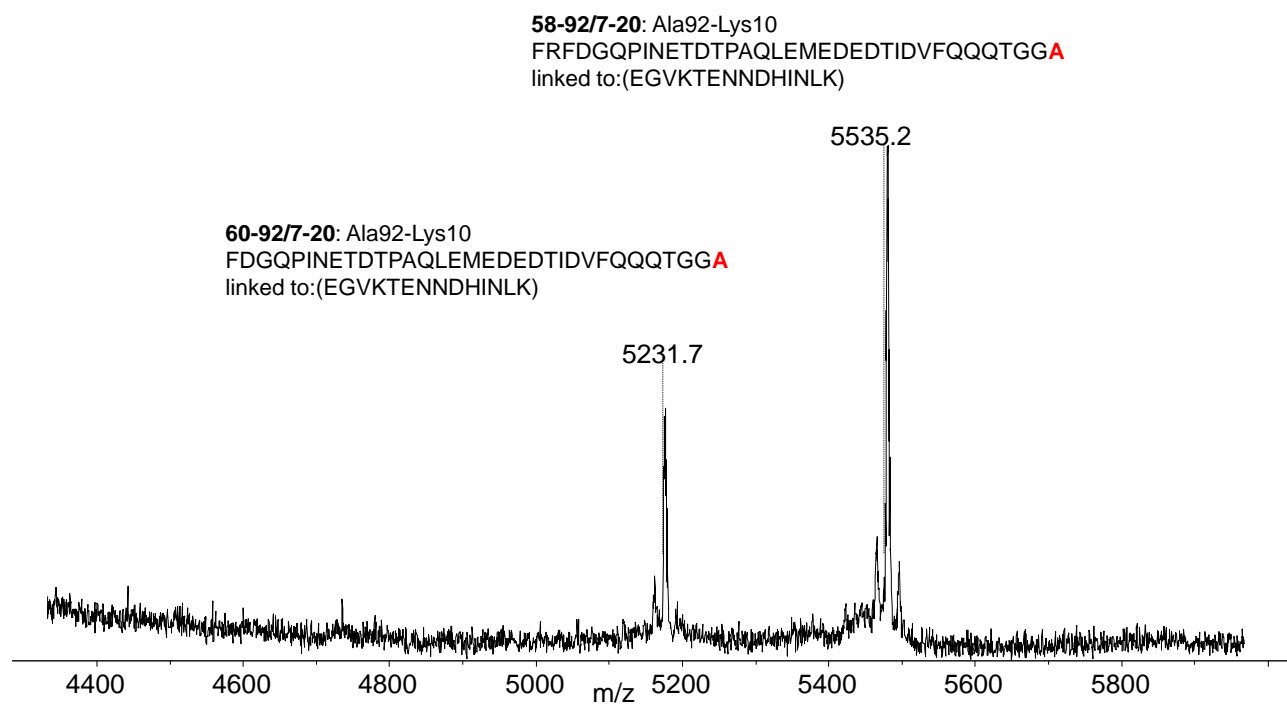


Figure S 37. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and trypsin cleavage of SUMO-2/3 dimer 26.

Internal cysteine state

GSVVQFKIKRHTPLSKLMKAYCE(1*Carbamidomethyl(C)
 Theoretical monoisotopic molecular weight
 [M+H]⁺= 2720.47

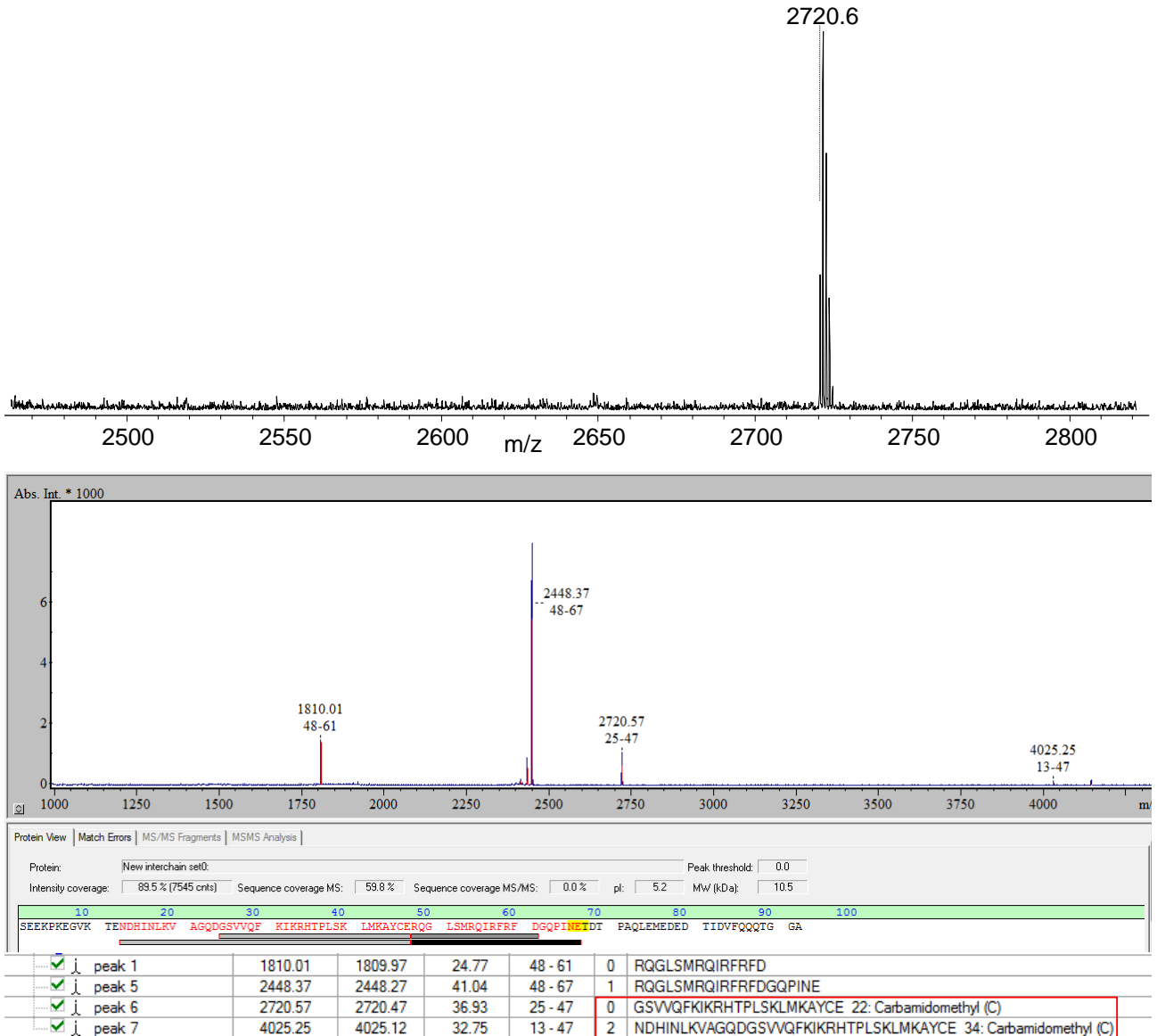


Figure S 38. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and endopeptidase GluC cleavage of SUMO-2/3 dimer 21.

5. Proteomic analysis of SUMO-2/3 dimer 27

SUMO-2/3 is the product of the desulfurization of SUMO-2/3 dimer **20** in denaturing conditions.

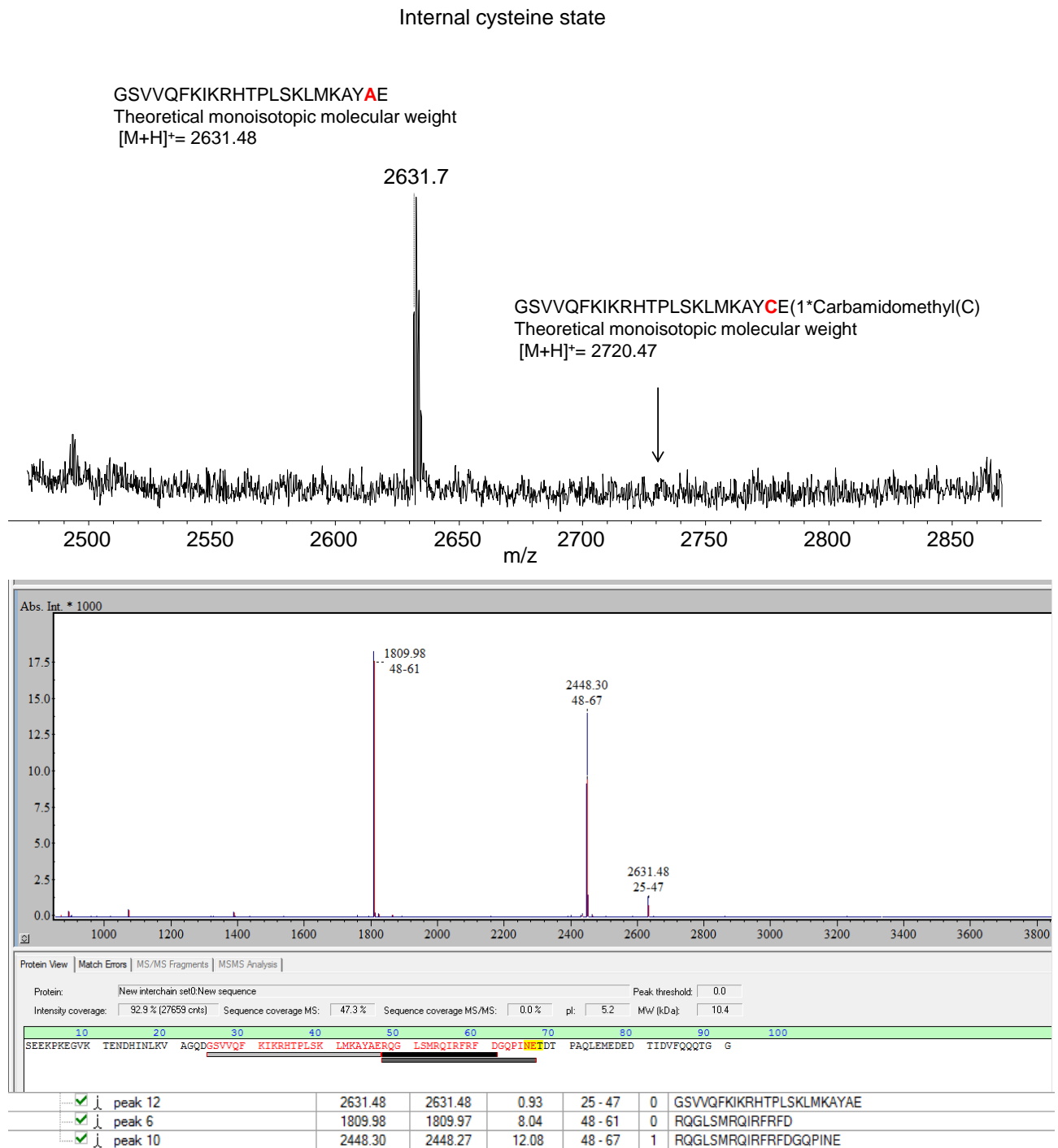


Figure S 39. MALDI-TOF spectrum of the fragments generated after reduction, alkylation and endopeptidase GluC cleavage of SUMO-2/3 dimer **27**.

V. Biophysical characterization

Table 1. Table of 20 different pdb entries describing the structure of SUMO-2 or -3 proteins with determination of % α helix.

Pdb name	SUMO	Sequence	Modifications	Technique	% α helix
2RPQ	SUMO-2/3	15-93		NMR	12%
2D07	SUMO-2/3	15-89		X-Ray	16%
2MP2	SUMO-3	12-92		NMR	15%
2N1W	pro-SUMO-2	1-93		NMR	10%
5GHB	SUMO-2	1-93	Lys42 acetylated	NMR	14%
5GHC	SUMO-2	1-93	Lys33 acetylated	NMR	14%
2IO1	SUMO-2/3	15-88		X-Ray	14%
2IO3	SUMO-2/3	20-93		X-Ray	14%
2IYD	SUMO-2/3	14 -92		X-Ray	14%
2CKH	SUMO-2/3	14 -92		X-Ray	16%
1WZ0	SUMO-2	8 -100		NMR	11%
1WM2	SUMO-2	12-89		X-Ray	17%
1WM3	SUMO-2/3	17-88		X-Ray	15%
2N9E	SUMO-2	1-95		NMR	15%
2AWT	SUMO-2	12-89		X-Ray	11%
3ZO5	SUMO-2/3	16-95		X-Ray	16%
4BKG	SUMO-2	17-89		X-Ray	8%
2MP2	SUMO-3	1-90		NMR	15%
2IO0	SUMO-2/3	16-89		X-Ray	10%
5D2M	SUMO-2/3	18-93		X-Ray	13%
				average % α helix	14%

ANNEX B: ARTICLES

BIOCONJUGATE CHEMISTRY

Article 1: The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis

Publication Date: September 18, 2019

DOI: [10.1021/acs.bioconjchem.9b00598](https://doi.org/10.1021/acs.bioconjchem.9b00598)

Article 2: Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations

Publication Date: November 8, 2019

DOI: [10.1021/acs.bioconjchem.9b00661](https://doi.org/10.1021/acs.bioconjchem.9b00661)

The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis

Jennifer Bouchenna,^a Magalie Sénéchal,^a Hervé Drobecq,^a Nicolas Stankovic-Valentin,^b Jérôme Vicogne,^{*a} Oleg Melnyk^{*a}

^a University of Lille, CNRS, Institut Pasteur de Lille, INSERM, UMR CNRS 8204, Centre d'Immunité et d'Infection de Lille, F-59000 Lille, France.

^b Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), DKFZ - ZMBH Alliance, 69120, Heidelberg, Germany.

emails of corresponding authors:

oleg.melnyk@ibl.cnrs.fr

Jerome.vicogne@ibl.cnrs.fr

ABSTRACT

While the semi or total synthesis of ubiquitin or polyubiquitin conjugates have attracted a lot of attention the last decade, the preparation of small-ubiquitin like modifier (SUMO) conjugates is much less developed. We describe hereinafter some important molecular features to consider when preparing SUMO-2/3 conjugates by chemical synthesis using the native chemical ligation and extended methods. In particular, we clarify the role of the conserved cysteine residue on SUMO-2/3 domain stability and properties. Our data reveal that SUMO-2 and 3 proteins behave differently to the Cys→Ala modification with SUMO-2 being less impacted than SUMO-3 likely due to a stabilizing interaction occurring in SUMO-2 between its tail and the SUMO core domain. While the Cys→Ala modification has no effect on the enzyme-catalyzed conjugation, it shows a deleterious effect on the enzyme-catalyzed deconjugation process, especially with the SUMO-3 conjugate. Whereas it is often stated that SUMO-2 and SUMO-3 are structurally and functionally indistinguishable, here we show that these proteins have specific structural and biochemical properties. This information is important to consider when designing and preparing SUMO-2/3 conjugates, and should help making progress in the understanding of the specific role of SUMO-2 and/or SUMO-3 modifications on protein structure and function.

KEY-WORDS

Protein chemical synthesis, SUMO conjugates, SUMO-2, SUMO-3, desulfurization, circular dichroism

INTRODUCTION

Protein synthesis¹⁻³ is a powerful mean for accessing proteins harboring a large variety of small post-translational modifications such as methylation,⁴ acetylation,^{5,6} phosphorylation^{5,7} or sulfation.^{8,9} Its increasing capacity to address the production of proteins decorated by post-translational modifications of much higher complexity stimulates the creativity of protein chemists worldwide.¹⁰⁻²⁴ In particular, chemical synthesis has been used to produce ubiquitin (Ub),^{12,25-27} polyubiquitin,^{12-14,22,24,28-30} small ubiquitin-like modifiers (SUMO)^{19,20,31} and conjugates thereof^{10,11,21,23,29,32-35} (for reviews see ^{36,37}) and as such has permitted to investigate in a new way how these modifications modulate protein structure and function.

Among the five human SUMO isoforms identified so far, SUMO-1-3 are the most studied since they are constitutively expressed in all eukaryotic cells (Fig. 1A).^{38,39} SUMO-2 and 3 are highly homologous (~95%) but are only 45% identical to SUMO-1. SUMO proteins are linked to the side-chain of exposed lysine residues (Lys) through an isopeptidic bond involving the C-terminus of the modifier. While ubiquitination is principally a signal for protein degradation by the proteasome, SUMO modification can change the localization, the stability, or the activity of

the target proteins by modulating protein-protein interactions and by competing or intersecting with other lysine modifications such as ubiquitin.^{40,41} Many proteins are the target of SUMO modification.⁴² The SUMOylation and deSUMOylation machinery, by being implicated in the regulation of many cellular mechanisms, is now considered as a promising target for the development of drugs against diseases such as cancer.⁴³⁻⁴⁵

Contrary to SUMO-1, SUMO-2 and 3 isoforms contain an internal SUMOylation site within their N-terminal tail situated at lysine 11. This feature makes possible the production of polySUMO-2/3 chains *in vivo*. Therefore, understanding how SUMOylation and, in particular, how polySUMO-2/3 modification modulates protein structure and function can strongly benefit from the production of homogeneous and branched SUMO-2/3 conjugates, as this is currently pursued with synthetic ubiquitin scaffolds. Such scaffolds can nicely complement studies performed with linear and recombinant polySUMO constructs.^{46,47}

While methods enabling the chemical synthesis of polyubiquitin chains have been extensively developed during the last decade, the reports describing the synthesis of SUMO conjugates are few and far between.^{10,11,21,23,32-35} This situation arises probably in part from the difference in length between Ub (76 AA) and SUMO domains (92-96 AA), the latter being beyond the limits of conventional SPPS: the SPPS of Ub has been reported by Ramage and co-workers about 30 years ago,²⁶ when an efficient SPPS approach to SUMO has been disclosed only recently.⁴⁸ A second and important difference between Ub and SUMO proteins is also the presence of a central cysteine (Cys) residue in the core of all SUMO isoforms (Fig. 1A).⁴⁹

Ub conjugates can be produced by reacting an Ub thioester domain with a target peptide featuring a γ ^{50,51} or δ -mercaptolysine^{13,14,52} residue according to the native chemical ligation (NCL⁵³, for a recent review see ²). Appending a Cys residue to the side-chain of the target lysine is an interesting alternative to the use of mercaptolysines.^{54,55} The thiol group from mercaptolysine or Cys is subsequently removed by desulfurization (Scheme 1).^{56,57} Since Ub features an central alanine residue in its sequence (Ala46), Ub (Ala46Cys) analogue can also be assembled from two peptide segments by NCL and converted afterwards into native Ub by global desulfurization.

The central Cys residue in SUMO domains is ideally located and facilitates their assembly by NCL. However, the production of SUMO conjugates requires an adaptation of desulfurization protocols if a native SUMO domain has to be obtained.^{19,20,23} Indeed, the desulfurization of the thiol amino acid used to produce the isopeptidic bond, *i.e.* a mercaptolysine or Lys(Cys), can potentially result in the desulfurization of the SUMO domain. This is especially the case when desulfurization is conducted in denaturing conditions, *e.g.*, in concentrated guanidine hydrochloride (Gn-HCl), since all thiols are equally exposed. Interestingly, the possibility of a selective desulfurization of SUMO-1-3 domains in native conditions was reported by Geurink *et al.*³³ and confirmed afterwards by Drobecq *et al.* with the preparative synthesis of SUMO-1

conjugates (Scheme 1).²³ The basis of this selectivity is the masking of the internal Cys thiol provided by the hydrophobic core of the folded SUMO domain, whereas the thiol from the residue used to produce the isopeptidic bond is solvent exposed and accessible to the desulfurization reagents.

Nevertheless, the selective desulfurization protocol discussed above might not preclude the desulfurization of the Cys residues present in the SUMOylated polypeptide that are accessible to the solvent. Moreover, the popularity of 6 M Gn·HCl as a medium for protein synthesis originates from its capacity to significantly improve the solubility of polypeptides. Therefore, another pitfall of performing a desulfurization in native conditions can be the limited solubility of the reactants. In this context, knowing the relative contribution of the internal Cys thiol to SUMO-2/3 domain structure and functionality can significantly influence the design of SUMO conjugates and the way they are produced.

While the importance of the internal Cys residue on SUMO-1 thermal stability and biochemical properties has been investigated,²³ this information is not yet available for SUMO-2 or SUMO-3. The differences between SUMO-1 and SUMO-2/3 in amino acid composition and in the Cys thiol environment within the protein hydrophobic core of the SUMO domain hampers any generalization and stimulated the investigations reported hereinafter (Fig. 1B, C). In particular, we used circular dichroism (CD) to compare SUMO-2/3 and desulfurized SUMO-2/3 secondary structure and thermal stability. These proteins were also compared using SUMO biochemical conjugation assays, including FRET-based conjugation experiments. We also determined the kinetics for the SENP1 and SENP2-catalyzed cleavage of synthetic conjugates featuring a wild type or desulfurized SUMO-2/3 domain. In addition of showing the importance of the internal Cys residue on SUMO-2/3 secondary structure and recognition by deconjugation enzymes, this work establishes that SUMO-2 and 3 behave differently to the Cys→Ala perturbation.

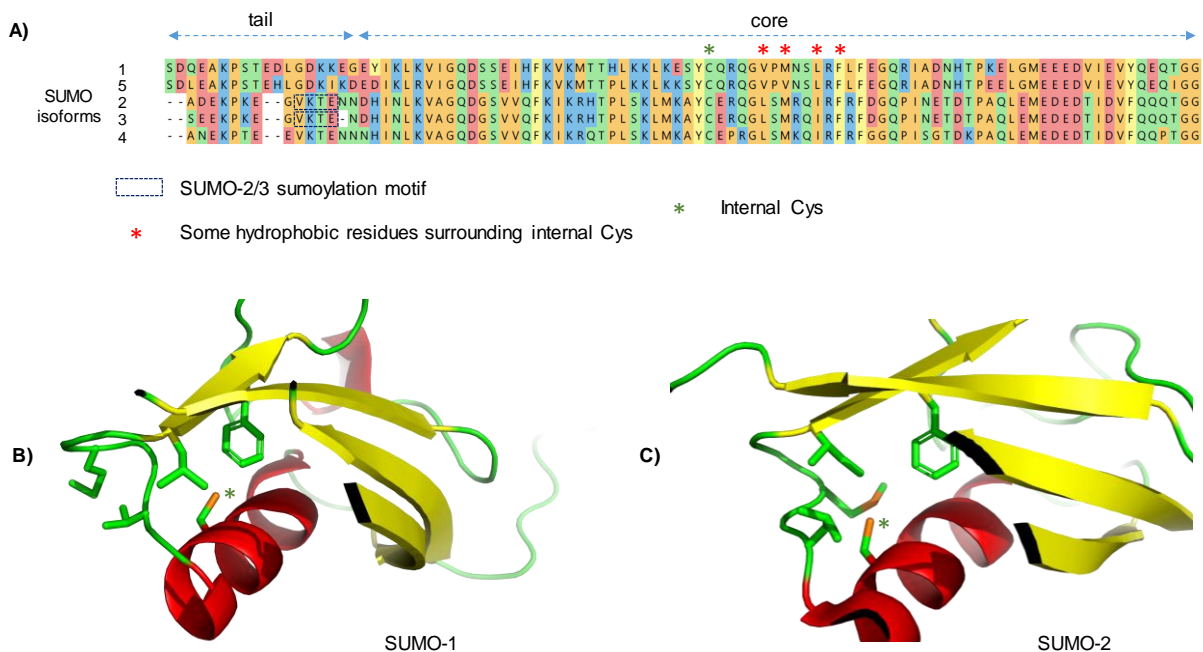
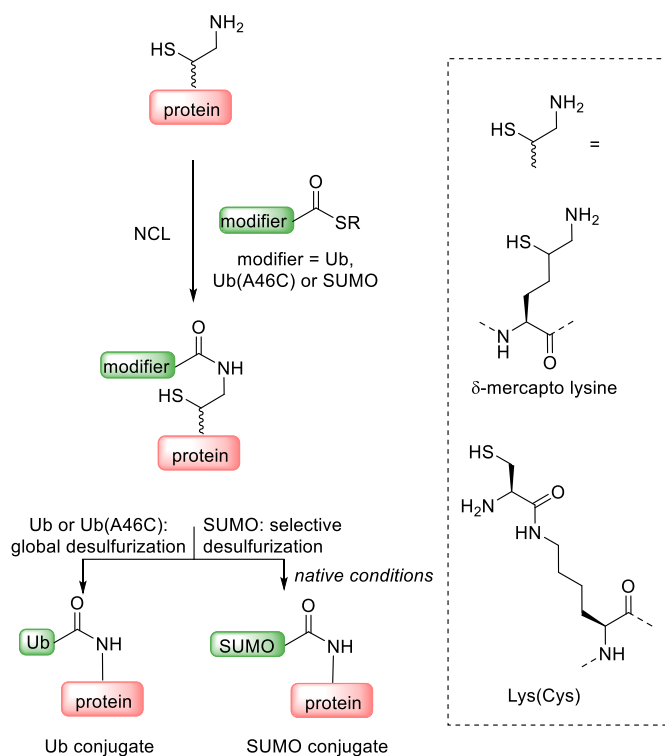


Figure 1. A) Protein sequences from Homo sapiens (human) SUMO isoforms hSUMO-1 (Accession number P63165), hSUMO-2 (P61956), hSUMO-3 (P55854), hSUMO-4 (Q6EEV6) and hSUMO-5 (G2XKQ0) have been aligned using MegAlign Pro™ (Lasergen© DNASTar®) with MAFFT algorithm (BLOSUM30 scoring matrix). Sequences have been associated according to their phylogenic proximities. Color background of amino acids corresponds to the chemical nature of the side chain (aromatic – yellow, acidic – red, basic – blue, nonpolar – orange, polar – green). B) NMR structure of SUMO-1, Protein Data Bank (pdb) entry 1A5R.⁵⁸ C) NMR structure of SUMO-2 (pdb entry 2N1W). For B,C some side-chains of hydrophobic residues surrounding the internal Cys are highlighted.

Scheme 1. Chemical synthesis of Ub and SUMO-conjugates by the combined use of NCL and desulfurization



RESULTS AND DISCUSSION

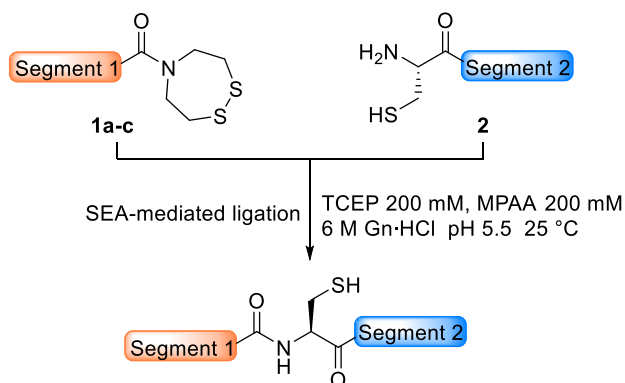
Chemical synthesis of SUMO-2/3 proteins

Wild-type and desulfurized SUMO-2 and 3 proteins (SUMO-2 C48A and SUMO-3 C47A) were required for performing comparative studies by circular dichroism and by SUMO biochemical conjugation assays. The synthesis of wild-type and desulfurized SUMO-2/3 core domain was also undertaken to further investigate the role of the SUMO tail in protein folding.

The chemical synthesis of the different SUMO proteins was undertaken as described in Scheme 2. The presence in SUMO proteins of a Cys residue in central position enabled their assembly in two pieces. The segments corresponding to the sequence of SUMO-2, SUMO-3 and SUMO-2/3 core on both sides of the central cysteine residue were produced by 9-fluorenylmethyloxycarbonyl (Fmoc) SPPS using bis(2-sulfanylethyl)amino (SEA) ChemMatrix® resin^{19,20} for segments **1a-c** and 4-hydroxymethylphenoxyacetyl (HMPA) ChemMatrix® resin for segment **2**. The use of Fmoc-Asp(*Or*Bu)-(Dmb)Gly-OH dipeptide unit for introducing the Asp and Gly residues indicated in bold in Scheme 2 was found to be mandatory for suppressing aspartimide formation during SPPS, as already noticed by other groups.^{31,48} SEA-mediated ligation⁵⁹ of segments **1a-c** and **2** proceeded efficiently at pH 5.5 in the presence of 4-mercaptophenylacetic acid (MPAA) to produce the target SUMO proteins. While SEA-mediated ligations are usually carried out at 37°C, we performed the assembly of the different SUMO proteins at a lower temperature of 25°C, since higher temperatures resulted in the appearance of aspartimide side-products. Using these conditions, SUMO-2, SUMO-3 and SUMO-2/3 core proteins were isolated in good yield (60-83% after HPLC

purification). Typical HPLC-MS and SDS-PAGE analyses of synthetic SUMO-2 and 3 proteins are shown in Fig. 2. For the latter analysis, the synthetic proteins are compared to SUMO-2 and SUMO-3 recombinant products (see Notes i, ii).

Scheme 2. Chemical synthesis of SUMO-2, SUMO-3 and SUMO-2/3 core proteins



SUMO-2 (from **1a** and **2**, 83% isolated)
SUMO-3 (from **1b** and **2**, 78% isolated)
SUMO-2/3 core (from **1c** and **2**, 60% isolated)

Segments **1a-c**

1a, SUMO-2 (2-47): ADEKPKEGVKTENNDHINLKVAGQ**DGS**VVQFKIKRHTPLSKLMKAY

1b, SUMO-3 (2-46): SEEKPKEGVKTENDHINLKVAGQ**DGS**VVQFKIKRHTPLSKLMKAY

1c, SUMO-2/3 core: DHINLKVAGQ**DGS**VVQFKIKRHTPLSKLMKAY

Segment **2**

CERQGLSMRQIRFR**FDG**QPINETDTPAQLEMEDEDTIDVFQQQTGG

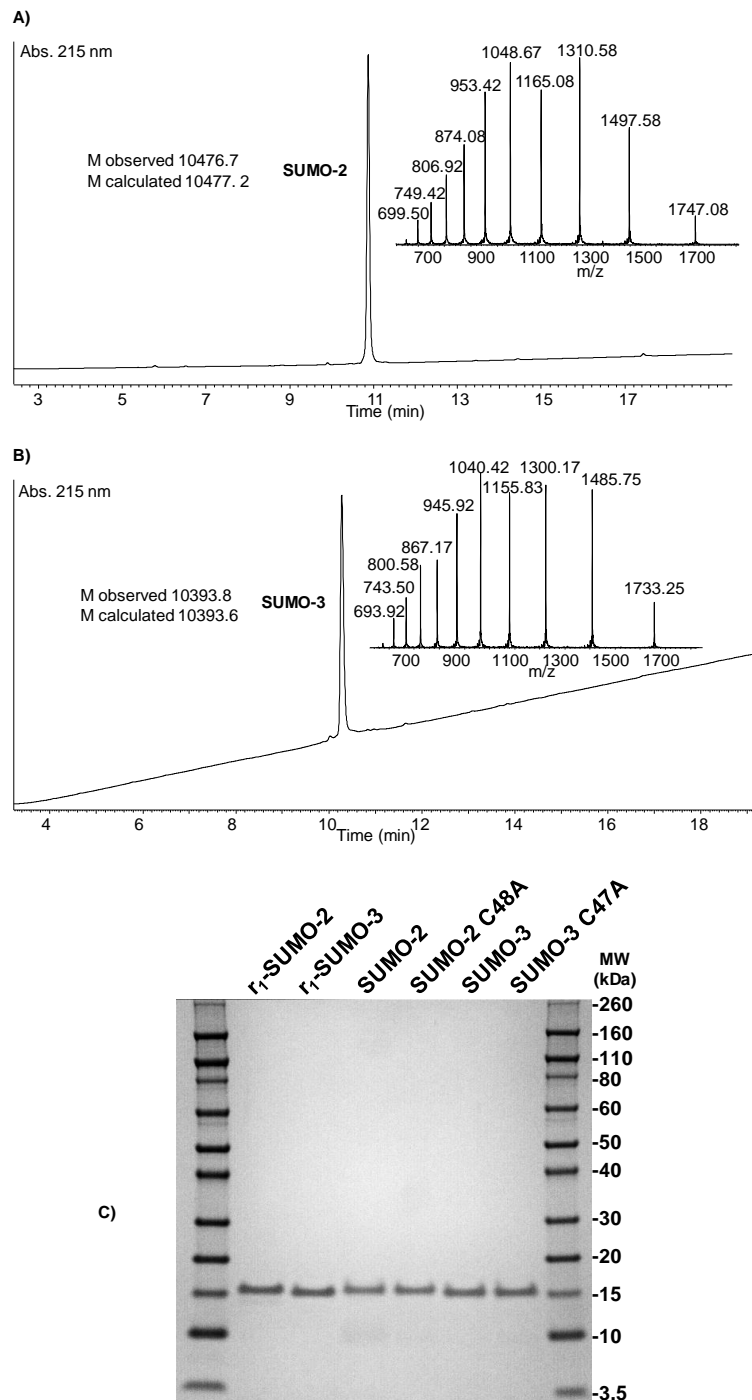
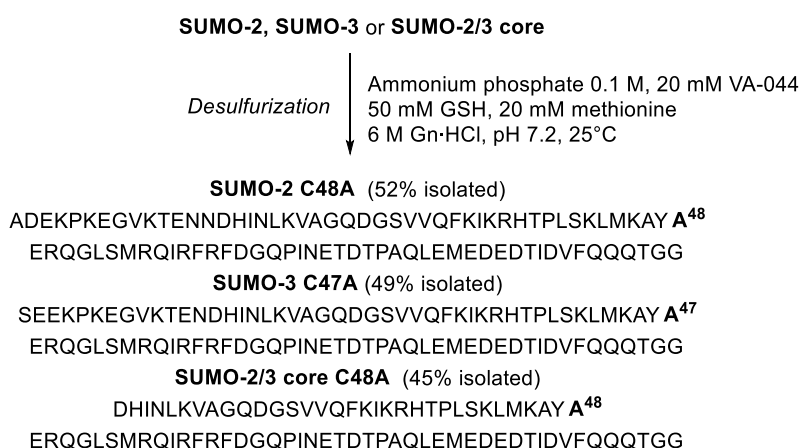


Figure 2. Analysis of synthetic SUMO-2 and SUMO-3 proteins. A) UPLC-MS analysis of synthetic SUMO-2; B) UPLC-MS analysis of synthetic SUMO-3; C) SDS-PAGE analysis of SUMO-2, SUMO-3, desulfurized SUMO-2 and 3 proteins (i.e. SUMO-2 C48A and SUMO-3 C47A respectively), and recombinant SUMO-2 and 3 proteins (i.e. r₁-SUMO-2 and r₁-SUMO-3 respectively, see Notes i, ii). The gel was loaded with 1 μ g of each protein and stained using Coomassie blue R250, MW = molecular weight markers).

The successful preparation of the different SUMO proteins enabled the preparation of SUMO-2 C48A, SUMO-3 C47A and SUMO-2/3 core C48A by *tris*(2-carboxyethyl)phosphine (TCEP)-induced radical desulfurization of the internal Cys residue in denaturing conditions (Scheme 3, see Note iii).⁵⁷ The presence of methionine in the desulfurization mixture proved to be important for minimizing the oxidation of the methionine residues during the reaction. The SDS-PAGE analysis of SUMO-2 and 3 Cys→Ala variants is shown in Fig. 2C in comparison with wild-type synthetic and recombinant SUMO-2 and 3 proteins (for other analyses, see Supporting Information).

Scheme 3. Preparation of SUMO-2 C48A, SUMO-3 C47A and SUMO-2/3 core C48A analogues by desulfurization



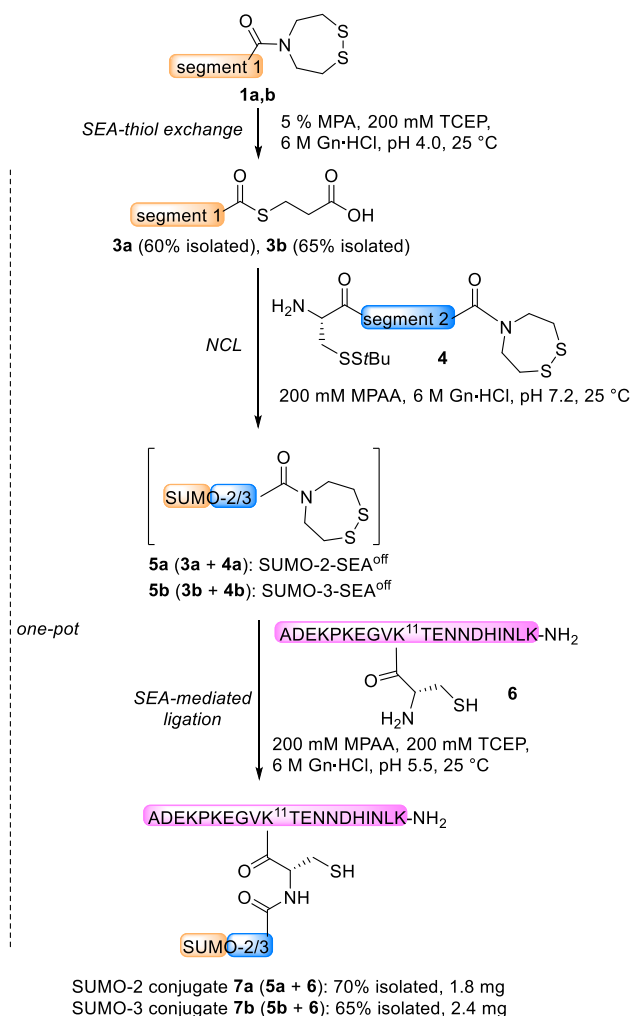
Chemical synthesis of SUMO-2 and 3 conjugates

The chemical synthesis of SUMO-2, SUMO-3 and desulfurized conjugates derived thereof was undertaken for studying the impact of the Cys→Ala modification on the deconjugation reaction catalyzed by Sentrin-specific protease 1 (SEN1⁶⁰) or 2 (SEN2^{61,62}), which are human SUMO-1, 2 and 3 specific deconjugating enzymes.^{43,63}

The conjugates were prepared starting from SEA peptides **1a,b** as described in Scheme 4. Briefly, peptide thioesters **3** derived from 3-mercaptopropionic acid (MPA) were first prepared from SEA peptides **1** by SEA-thiol exchange in mildly acidic conditions (pH 4.0).⁶⁴ The one-pot assembly of the conjugates started with the production of SUMO-SEA intermediates **5** by coupling peptide segments **3** and **4** according to the NCL reaction. This step was performed in the absence of TCEP to keep the SEA group in its latent form as a cyclic disulfide.^{20,65} Once formed, the SUMO-SEA domain **5** was activated by adding TCEP to the ligation mixture and ligated in one-pot with the target peptide **6**. The latter corresponds to the N-terminal tail of SUMO-2, which includes the SUMO-2 SUMOylation site VKTE (see Figure 1A). Typical HPLC-MS analyses of SUMO-2 and 3 conjugates **7a,b** are shown in Fig. 3A,B.

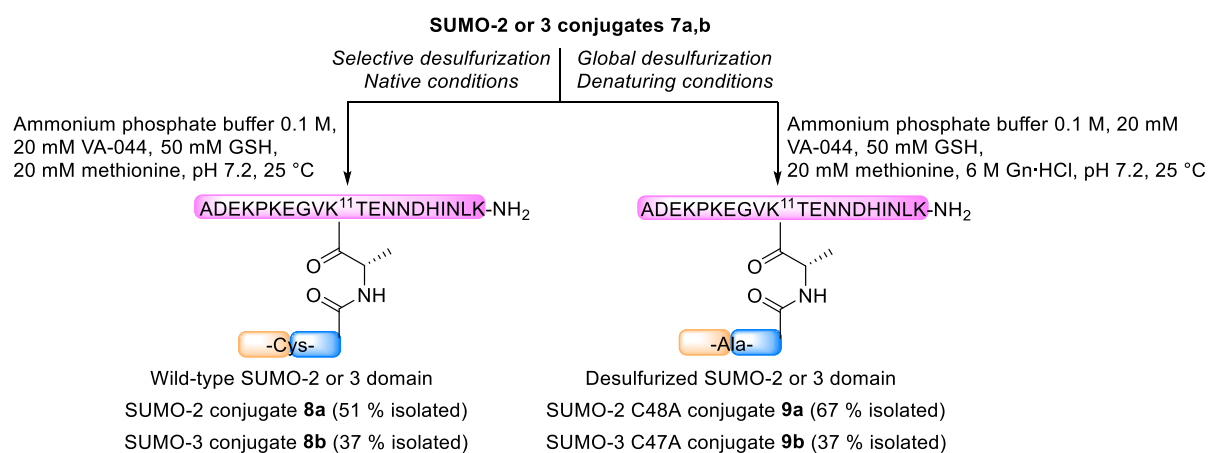
The successful synthesis of SUMO 2 and 3 conjugates **7a,b** set the stage for the desulfurization reactions, which were performed under native conditions, i.e., in the absence of Gn·HCl (Scheme 5, left), or in denaturing conditions that is in 6 M Gn·HCl (Scheme 5, right). Importantly, the native conditions resulted in the selective desulfurization of the Cys thiol linking the SUMO domains to the target peptide and hence in the production of conjugates **8a,b**, while denaturing conditions furnished the fully desulfurized products **9a,b**. The selectivity of the desulfurization reaction in native conditions for the Cys residue between SUMO domain and the target lysine was confirmed by proteomic analysis of conjugates **8a,b** (see Supporting Information). The SDS-PAGE analysis of the different conjugates is shown in Fig. 3C. All the conjugates migrate to an apparent molecular weight of ~20 kDa, while wild-type SUMO-2 and 3 proteins migrate at around 15 kDa.

Scheme 4. Synthesis of SUMO-2 and 3 conjugates **7a,b**



1,3a, SUMO-2 (2-47): ADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY
1,3b, SUMO-3 (2-46): SEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY
4, SUMO-2 (48-93) or SUMO-3 (47-92): CERQGLSMRQIRFRFDGQPINETDTPAQLEME-
 DETDIDVFQQQTGG

Scheme 5. Selective or global desulfurization of SUMO conjugates 7a,b. Production of selectively desulfurized SUMO conjugates 8a,b and of fully desulfurized conjugates 9a,b



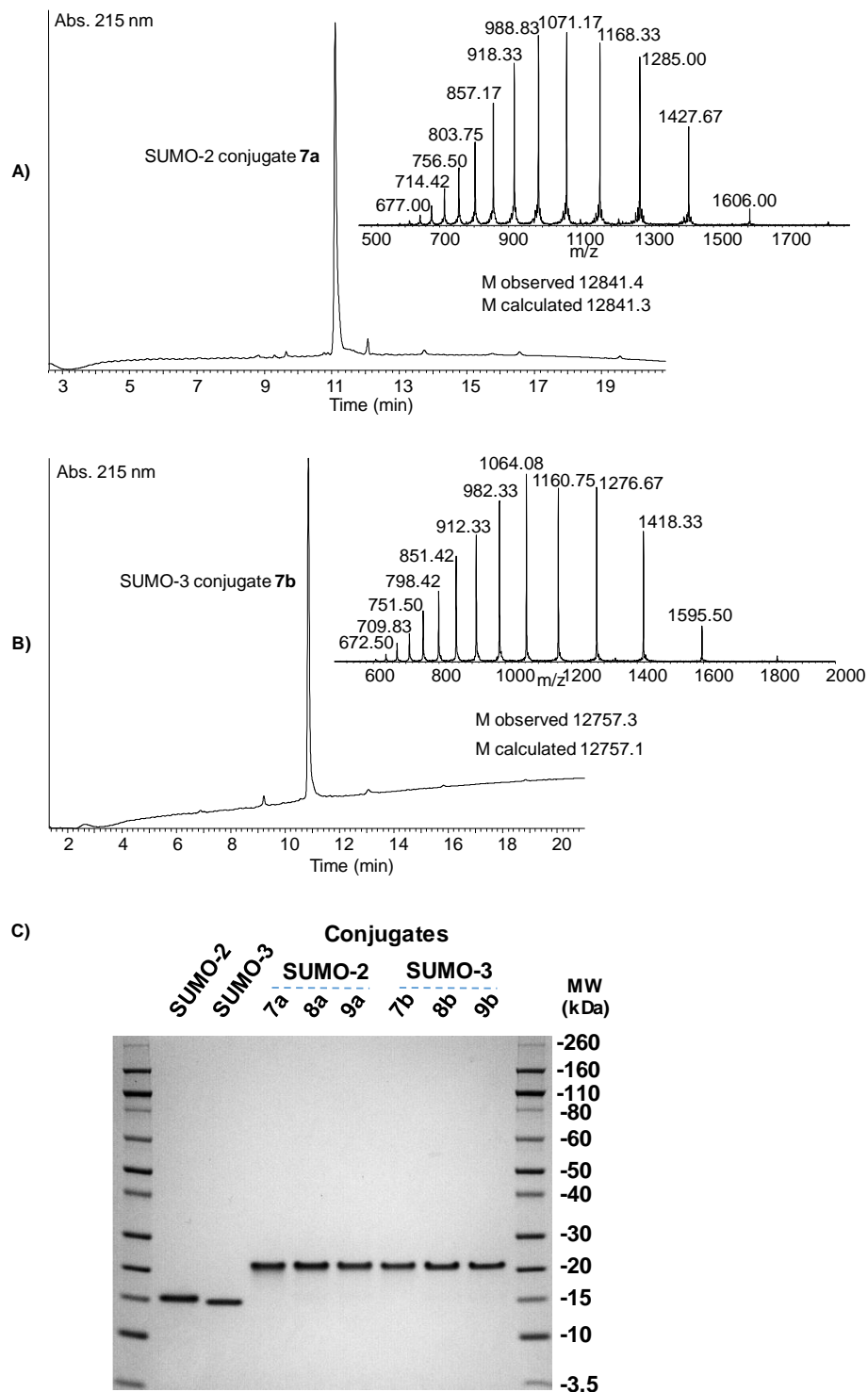


Figure 3. Analysis of synthetic SUMO-2 and SUMO-3 conjugates. A) HPLC-MS analysis of synthetic SUMO-2 conjugate **7a**; B) HPLC-MS analysis of synthetic SUMO-3 conjugate **7b**; C) SDS-PAGE analysis of synthetic SUMO-2 and SUMO-3 proteins and of the different conjugates derived thereof, i.e. **7-9a** and **7-9b** respectively (1 μ g of each protein, Coomassie blue R250 staining, MW= molecular weight markers).

Role of the internal Cys residue on SUMO-2 and 3 secondary structure and stability

The successful synthesis of the different SUMO 2 and 3 variants set the stage for the analysis by CD of the impact of the Cys→Ala modification on the secondary structure and thermal stability of the SUMO-2 and SUMO-3 domains. The CD spectra of synthetic and recombinant SUMO-2 and 3 proteins are highly similar and the same for the thermal denaturation curves (see Supporting Information, Fig. S28 and S29). The CD spectra of wild-type synthetic SUMO-2 and 3 are highly similar. The α -helical content of ~15% found for synthetic and recombinant SUMO-2 and 3 proteins is identical to the CD α -helical content recently reported for recombinant SUMO-2.⁶⁶ Determined α -helical content is also consistent with that extracted from 20 different pdb entries describing the structure of SUMO-2 or SUMO-3 (see Supporting Information, Table S1).

Figure 4 shows the far-UV CD spectra at 25 °C, the deduced α -helical content according to the method of Greenfield & Fasman⁶⁷ and the evolution of the mean residue ellipticity at 222 nm, $[\theta]_{MRW222}$ nm, as a function of the temperature for synthetic SUMO-2 and 3 proteins and their Cys→Ala variants. The data show that the Cys→Ala modification induced only subtle changes in the far-UV CD spectrum of the SUMO-2 protein with a decrease of protein α -helical content of ~1% ($15.2 \pm 0.12\%$ ($n = 10$) for SUMO-2 vs $14.1 \pm 0.23\%$ ($n = 12$) for SUMO-2 C48A). The thermal denaturation curves for SUMO-2 and SUMO-2 C48A were found to be similar (SUMO-2 $T_m = 70.7 \pm 2.4$ °C, Fig. 4B).

The impact of the Cys→Ala modification was more pronounced for SUMO-3 with a decrease of the protein α -helical content of ~2% ($15.1 \pm 0.21\%$ ($n = 8$) for SUMO-3 vs $13.3 \pm 0.29\%$ ($n = 10$) for SUMO-3 C47A). The T_m measured for SUMO-3 C47A (65.5 ± 0.7 °C) is 4.4 °C lower than those found for the wild type protein SUMO-3 (69.9 ± 1.5 °C). Therefore and contrary to SUMO-2, the effect of the Cys→Ala modification on SUMO-3 is significant.

The internal Cys (Cys48 for SUMO-2, Cys47 for SUMO-3) is located in the α -helix of the SUMO-2/3 core domain with the Cys side-chain pointing toward the interior of the hydrophobic core. Some hydrophobic residues surrounding the Cys side-chain are highlighted in Fig. 1B. The case of Phe60 (SUMO-2) which is within 5 Å of Cys48 is particularly interesting to discuss in the light of a recent report by Bhattacharya and Ainaravaru.⁶⁶ The authors have proposed that Phe60 plays a central role in the reduction of the tertiary structure and mechanical stiffness of SUMO-2, as measured by CD and single-molecule force spectroscopy (SMFS) respectively, which is observed when SUMO-2 is heated from 5 °C to 45 °C. Obviously, the thermal denaturation of SUMO-2 is complex and involves some intermediates between the native and unfolded states. Therefore, the modification of the hydrophobic core of SUMO-2 and 3 by desulfurization of the central Cys residue can potentially alter their structure, stability and/or the distribution of intermediates along the protein energy landscape. Interestingly, the data presented in Fig. 4 show that although wild-type SUMO-2 and SUMO-3 proteins adopt the

same secondary structure, they respond differently to the Cys→Ala modification. SUMO-2 and 3 differ only by some amino acid residues in their tail (Figure 1A). In other words, SUMO-2 and SUMO-3 share the same core domain and the close environment of the central Cys is the same for both proteins. A likely explanation for the observed difference is the existence of a cross-talk between the tail and the SUMO-2/3 core domain which is perceived by CD when the latter is perturbed by desulfurization. In order to further investigate this point, we analyzed the synthetic SUMO-2/3 core and SUMO-2/3 core C48A proteins by CD (Figure 5). The data show that the SUMO-2/3 core domain is also sensitive to the Cys→Ala modification, with a loss of α -helical content upon Cys desulfurization which is similar to what is observed for SUMO-3. Taken together, these experiments strongly support the existence of stabilizing interactions between the N-terminal tail of SUMO-2 and the core domain that are absent or much less pronounced in SUMO-3. Therefore, the Cys→Ala modification is an interesting tool to uncover weak interactions occurring between the tail and the core domain of SUMO proteins.

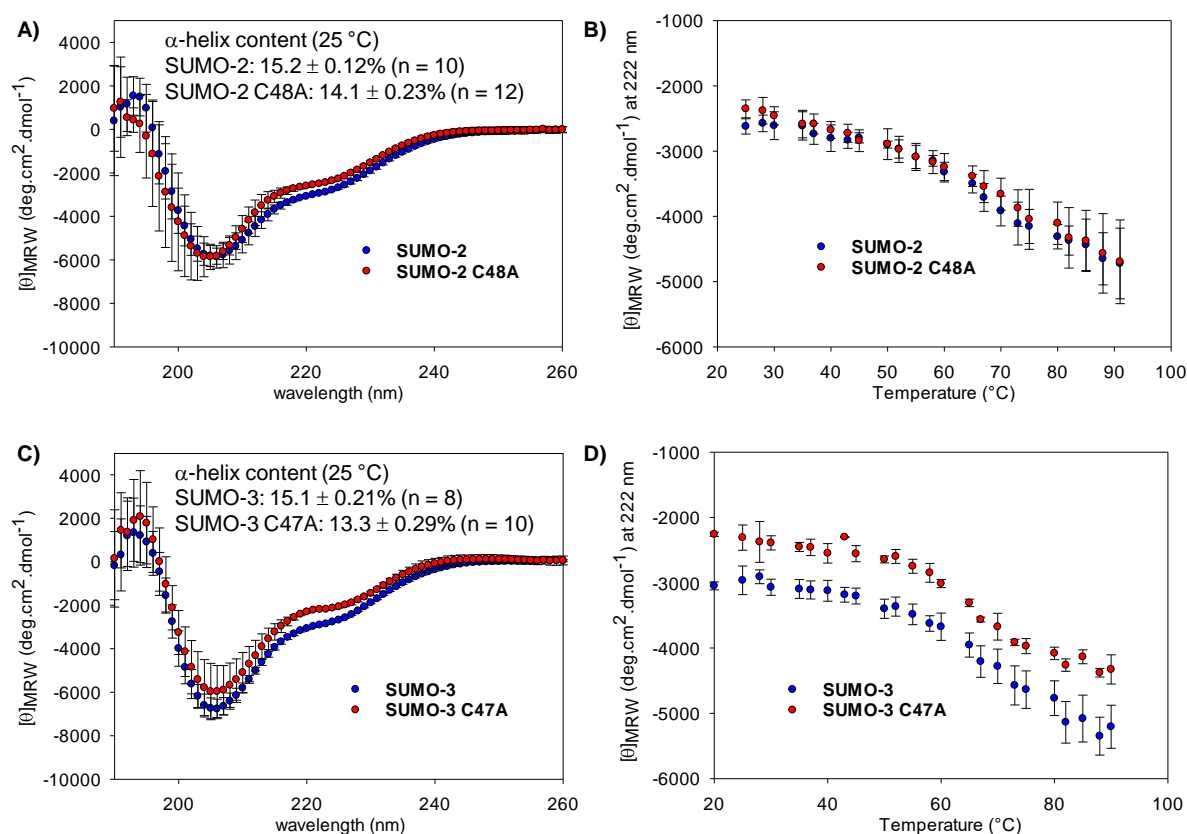


Figure 4. CD analysis of synthetic SUMO-2, SUMO-2 C48A, SUMO-3 and SUMO-3 C47A proteins. A) Analysis of SUMO-2 (n = 10) and SUMO-2 C48A (n = 12) at 25 °C. B) Thermal denaturation of SUMO-2 (n = 5) and SUMO-2 C48A (n = 7) ($[\theta]_{\text{MRW}}$ at 222 nm). C) Analysis of SUMO-3 (n = 8) and SUMO-3 C47A (n = 10) at 25 °C. D) Thermal denaturation of SUMO-3 (n = 9) and SUMO-2 C47A (n = 5) ($[\theta]_{\text{MRW}}$ at 222 nm). Conditions: 10 mM phosphate buffer pH 7.2.

The samples were carefully degassed during 30 min with argon and immediately transferred under argon into the CD cuvette. Protein concentration for thermal denaturations : SUMO-2 12.9-28.6 μM , SUMO-2 C48A 12.0-32.5 μM , SUMO-3 12.5-40.4 μM , SUMO-3 C47A 13.5-34.7 μM . Protein concentration for CD analysis at 25 °C: SUMO-2 13.8-25.3 μM , SUMO-2 C48A 9.1-31.6 μM , SUMO-3 14.4-24.0 μM , SUMO-3 C47A 25.1-34.7 μM . The data correspond to the mean and standard deviation.

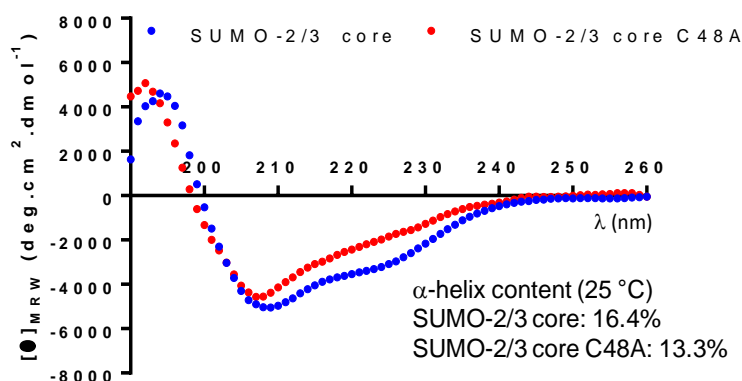


Figure 5. CD analysis of synthetic SUMO-2/3 core and SUMO-2/3 core C48A proteins. Conditions: 10 mM phosphate buffer pH 7.2. The samples were carefully degassed during 30 min with argon and immediately transferred under argon into the CD cuvette. Protein concentration: SUMO-2/3 core 21.2 μM , SUMO-2/3 core C48A 22.4 μM .

Effect of the Cys→Ala modification on SUMO-2/3 conjugation assays

Having characterized the effect of the Cys→Ala modification on the secondary structure and thermal stability of SUMO-2 and SUMO-3 proteins, we next performed a classical SUMO biochemical conjugation assay to see if the modification influences the enzyme-catalyzed formation of the isopeptidic bond using RanGAP as the target protein (see Supporting Information, Figure S30). We used for this purpose reagents from a commercially available conjugation kit. The different SUMO analogues (i.e. the recombinant proteins from the kit r_2 -SUMO-2/3, synthetic wild-type SUMO-2/3 and synthetic desulfurized SUMO-2(3) C48A (C47A) were efficiently transferred with no apparent difference in the amount of RanGAP-SUMO conjugate formed whether SUMO-2 or SUMO-3 were desulfurized or not. Unexpectedly, isolated desulfurized proteins SUMO-2 C48A and SUMO-3 C47A were poorly recognized by the polyclonal anti-SUMO-2/3 antibody provided with the kit.

Since the enzymatic conjugation assay described above could be biased by the differential recognition of the SUMO analogues by the anti-SUMO-2/3 antibody, we performed the FRET-based quantitative bioconjugation assay described in Fig. 6.^{68,69} In this assay, the cyan fluorescent protein (CFP)-labelled RanGAPtail target is conjugated enzymatically to the yellow

fluorescent protein (YFP)-labelled SUMO-3 protein. The CFP protein absorbs light at 430 nm and emits a fluorescence signal at 485 nm. When CFP and YFP moieties are in close proximity (e.g. when RanGAPtail is SUMOylated), the energy transfer from CFP to YFP results in an emission signal at 527 nm. Therefore, the intensity ratio of the emission at 527 nm and the emission at 485 nm is representative of the amount of RanGAPtail SUMOylated. The assay enables the monitoring of the conjugation reaction in real-time and in parallel. With this assay, we analyzed a commercially available and recombinant SUMO-3 protein and its chemically desulfurized analogue (Fig. 7A) as well as synthetic SUMO-3 and SUMO-3 C47A proteins (Fig. 7B). The different analogues were used as competitors of the YFP-labelled SUMO-3 substrate. The capacity of the synthetic and recombinant SUMO-3 proteins to compete for the YFP-SUMO-3 substrate is the same. The data also show that the Cys→Ala modification has no significant effect on the capacity of SUMO-3 to be conjugated to RanGAPtail, either with the recombinant protein or the synthetic one at different competitor/ YFP-labelled SUMO-3 ratios.

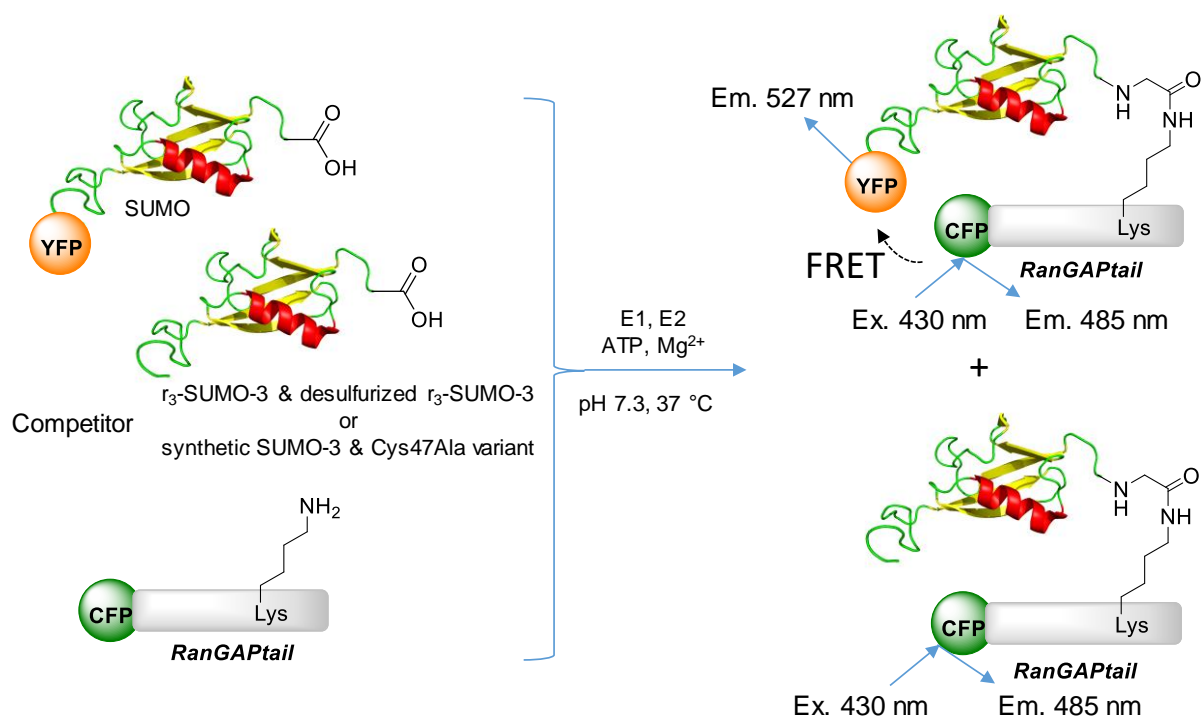


Figure 6. Principle of the FRET-based enzymatic conjugation assay using RanGAPtail as target.

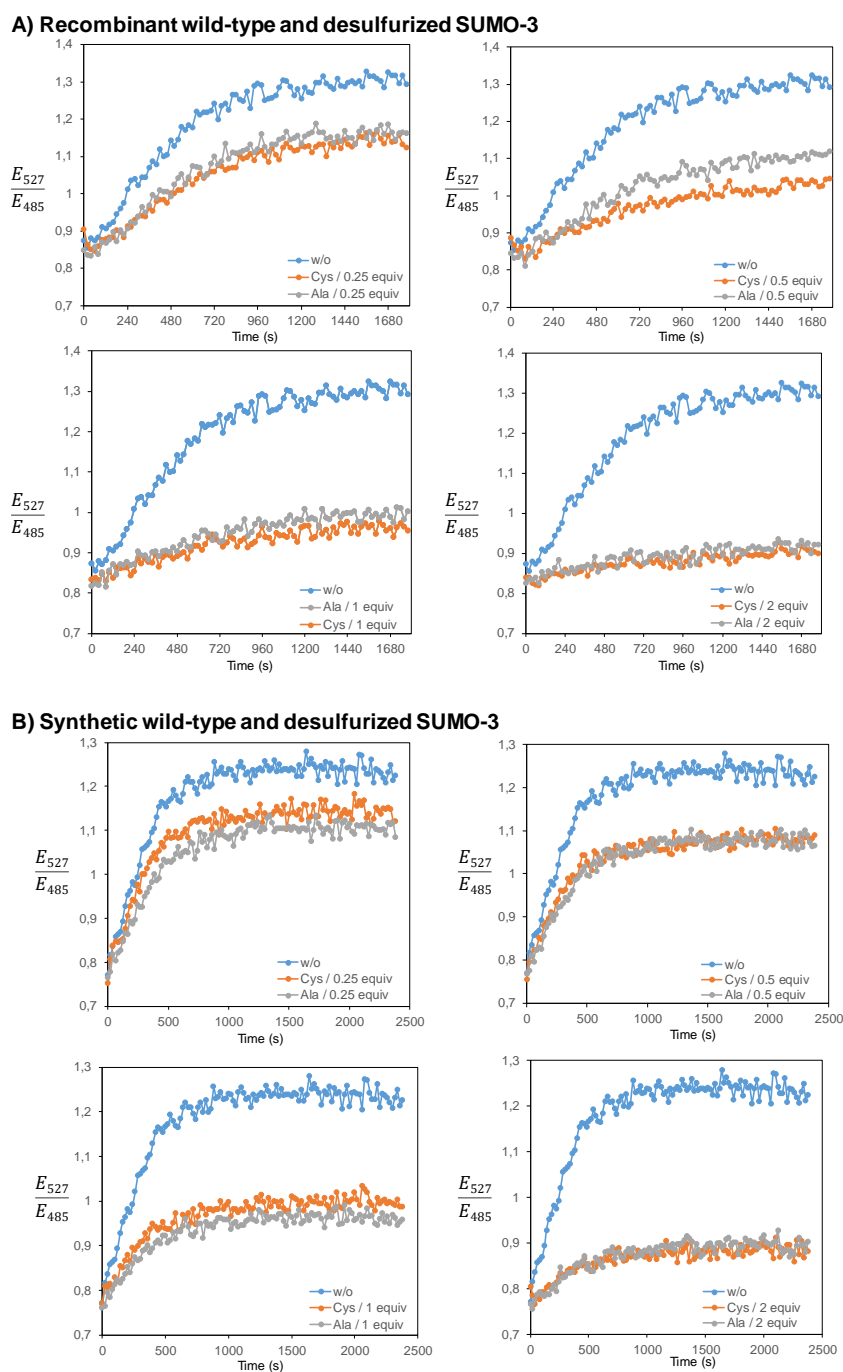
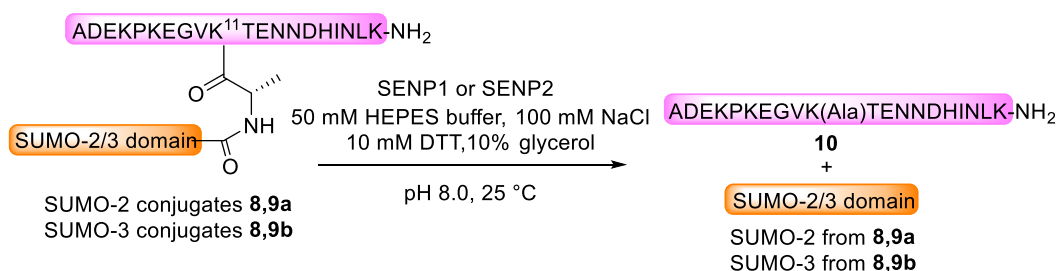


Figure 7. FRET-based enzymatic conjugation assay using RanGAPtail as target.^{68,69} A) Competition with a commercially available and recombinant SUMO-3 protein and its chemically desulfurized analogue. B) Competition with synthetic SUMO-3 and SUMO-3 C47A proteins. Conditions: 15 nM E1, 20 nM E2, 300 nM YFP-SUMO-3, 300 nM CFP-RanGAPtail and competitors (75 nM-600 nM) were incubated at 37°C for 10 min in conjugation medium (110 mM KOAc, 20 mM HEPES pH 7.3, 2 mM Mg(OAc)₂, 1 mM EGTA, supplemented with 1 μg/mL of each leupeptin, pepstatin and aprotinin, 1 mM DTT, 0.05% Tween®-20 and 0.2 mg/mL ovalbumin). In vitro SUMOylation reaction was started upon addition of 1 mM ATP. Every 20 s, the ratio $\frac{E_{527}}{E_{485}}$ was determined upon excitation at 430 nm.

Effect of the Cys→Ala modification on the SUMO-2 and 3 deconjugation reaction

We next examined the effect of Cys→Ala modification on the cleavage of synthetic SUMO-2 and 3 conjugates by SENP1 and SENP2 enzymes (Scheme 6). The enzymatic cleavage yields peptide **10** and the SUMO-2 or SUMO-3 domains. The rate of the deconjugation reaction could be easily monitored by LC-MS (Fig. 8).

Scheme 6. SENP1 or SENP2-catalyzed deconjugation of SUMO-2 conjugates **8,9a** and SUMO-3 conjugates **8,9b**



The data presented in Fig. 8 show that SUMO-2 and SUMO-2 C48A conjugates **8,9a** are cleaved at the same rate, whatever SENP enzyme is used. Therefore, SUMO-2 domain is recognized by the SENP1 and SENP2 enzymes whenever it features an internal Cys or not. In contrast, SUMO-3 conjugate **8b** is cleaved by both enzymes significantly faster than the C47A analog **9b**. In this case, the Cys→Ala modification alters the functionality of SUMO-3 domain. These results are in full accordance with the CD analyses, which showed that of the two proteins SUMO-2 and SUMO-3, SUMO-3 is the more sensitive to the Cys→Ala modification.

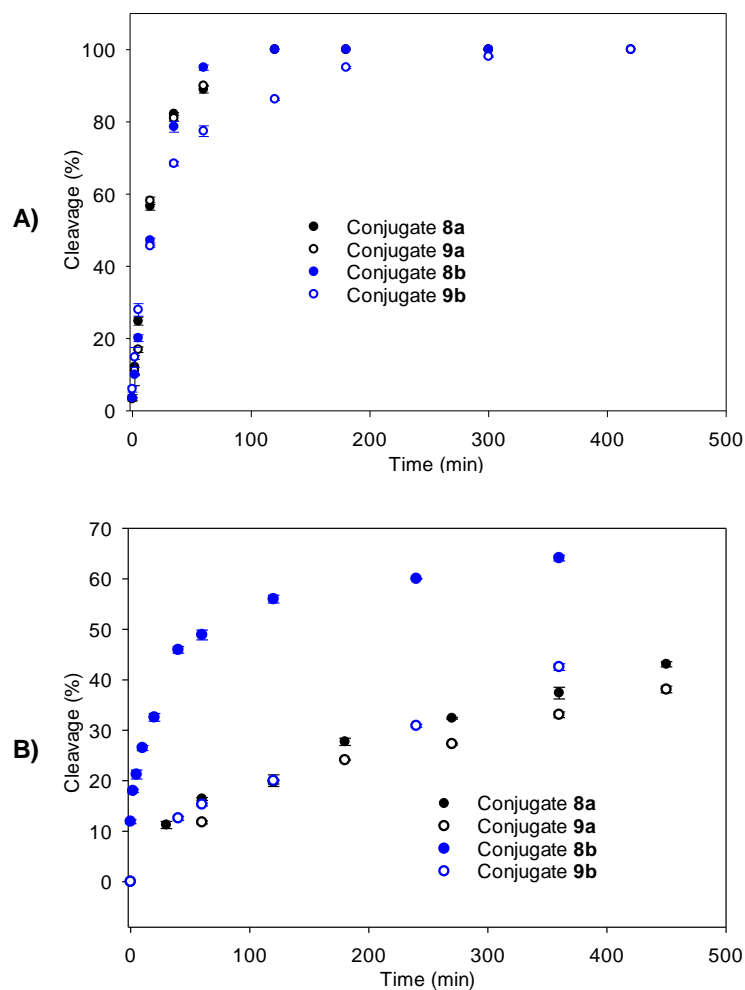


Figure 8. Deconjugation of SUMO conjugates **8,9a** and **8,9b** by SENP1 and SENP2 enzymes. A) SENP1-catalyzed deconjugation. B) SENP2-catalyzed deconjugation. Conditions: 50 mM HEPES buffer pH 8.0, 100 mM NaCl, 10 mM DTT, 10% glycerol, 25 °C, conjugate 0.5 mg/mL, final volume 100 μ L, SENP1 5.0 ng per cleavage reaction, SENP2 0.4 units per cleavage reaction. The deconjugation reaction was performed under nitrogen atmosphere, quenched with 10% aqueous acetic acid and analyzed by UPLC-MS. The experiments were performed in triplicate. The data correspond to the mean and standard deviation.

CONCLUSIONS

In conclusion, we have successfully prepared a series of SUMO-2 and SUMO-3 proteins that enabled investigating the impact of the SUMO-2/3 domain desulfurization on the secondary structure and thermal stability of SUMO-2 and 3 proteins, the rate of SUMO-2 and 3 conjugation to RanGAP substrates and on the deconjugation of SUMO-2 and 3 conjugates by SENP1 and SENP2 enzymes. Contrary to SUMO-2, the fold of SUMO-3 is significantly affected by the Cys \rightarrow Ala modification. Likewise, the rate of cleavage of SUMO-3 conjugates by SENP1 and SENP2 is reduced when the internal Cys residue is desulfurized into Ala. The CD analysis

of the wild-type and desulfurized SUMO-2, SUMO-3 and SUMO-2/3 core proteins show the existence of a stabilizing interaction between the tail and the core domain in SUMO-2 which is less apparent for SUMO-3. Taken together, our results show that mutating the internal Cys residue in SUMO-3 must be considered with caution. They also reveal subtle differences between SUMO-2 and 3 that cannot be considered as being the same although they sequences are highly similar. Therefore, we show here that SUMO-2 and SUMO-3 are distinct proteins and should not be considered as redundant.

NOTES

- i) The composition in amino acids of some commercially available recombinant SUMO-2/3 proteins was found to differ significantly from the canonical sequences of SUMO-2/3 shown in Fig. 1A. Therefore, each purchased recombinant product was trypsinized and analyzed by MALDI-TOF mass spectrometry to determine its sequence independently of the information provided by the manufacturer (see Supporting Information).
- ii) The recombinant SUMO-2 and SUMO-3 products used for SDS-PAGE and CD analyses, i.e. r₁-SUMO-2 and r₁-SUMO-3, correspond to the canonical sequences for SUMO-2/3 shown in Fig. 1A, see Supplementary Information.
- iii) For SUMO-2/3 core protein, we use the SUMO-2 numbering of amino acid residues. The Cys residue is therefore at position 48.

ACKNOWLEDGEMENTS

We thank the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation for financial support (PhD fellowship to Jennifer Bouchenna). We thank Jean-Jacques Lacapère and Florine Cavelier for useful discussions. We thank Frauke Melchior for giving us access to the FRET SUMO conjugation assay.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at DOI: #####.

Experimental procedures and characterization data for all peptides and proteins (LC-MS, MALDI-TOF, CD and proteomic analyses) (PDF).

References

- (1) *Topics in Current Chemistry. Protein Ligation and Total Synthesis*; Liu, L., Ed.; Springer International Publishing: New York, 2015; Vol. I & II.
- (2) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoet, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods. Mechanisms, Catalysis, Scope and Limitations. *Chem. Rev.* **2019**, *12*, 7328-7443.

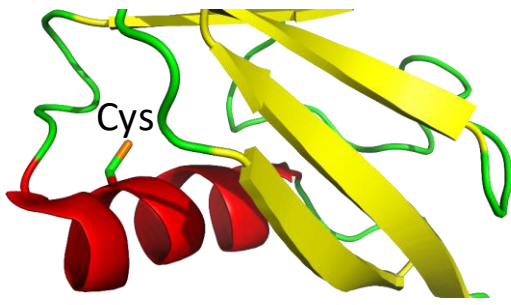
- (3) Agouridas, V.; El Mahdi, O.; Cargoët, M.; Melnyk, O. A Statistical View of Protein Chemical Synthesis Using NCL and Extended Methodologies. *Bioorg. Med. Chem.* **2017**, *25*, 4938-4945.
- (4) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L. One-Pot Native Chemical Ligation of Peptide Hydrazides Enables Total Synthesis of Modified Histones. *Org. Biomol. Chem.* **2014**, *12*, 5435-5441.
- (5) Chiang, K. P.; Jensen, M. S.; McGinty, R. K.; Muir, T. W. A Semisynthetic Strategy to Generate Phosphorylated and Acetylated Histone H2B. *ChemBioChem* **2009**, *10*, 2182-2187.
- (6) Yu, R. R.; Mahto, S. K.; Justus, K.; Alexander, M. M.; Howard, C. J.; Ottesen, J. J. Hybrid Phase Ligation for Efficient Synthesis of Histone Proteins. *Org. Biomol. Chem.* **2016**, *14*, 2603-2607.
- (7) Jbara, M.; Maity, S. K.; Morgan, M.; Wolberger, C.; Brik, A. Chemical Synthesis of Phosphorylated Histone H2A at Tyr57 Reveals Insight into the Inhibition Mode of the SAGA Deubiquitinating Module. *Angew. Chem. Int. Ed.* **2016**, *55*, 4972-4976.
- (8) Thompson, R. E.; Liu, X.; Ripoll-Rozada, J.; Alonso-Garcia, N.; Parker, B. L.; Pereira, P. J. B.; Payne, R. J. Tyrosine Sulfation Modulates Activity of Tick-Derived Thrombin Inhibitors. *Nat. Chem.* **2017**, *9*, 909-917.
- (9) Watson, E. E.; Liu, X.; Thompson, R. E.; Ripoll-Rozada, J.; Wu, M.; Alwis, I.; Gori, A.; Loh, C. T.; Parker, B. L.; Otting, G. et al. Mosquito-Derived Anophelin Sulfoproteins Are Potent Antithrombotics. *ACS Cent. Sci.* **2018**, *4*, 468-476.
- (10) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically Ubiquitylated Histone H2B Stimulates hDot1L-Mediated Intranucleosomal Methylation. *Nature* **2008**, *453*, 812-816.
- (11) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* **2009**, *48*, 8090-8094.
- (12) El Oualid, F.; Merckx, R.; Ekkebus, R.; Hameed, D. S.; Smit, J. J.; de Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. Chemical Synthesis of Ubiquitin, Ubiquitin-Based Probes, and Diubiquitin. *Angew. Chem. Int. Ed.* **2010**, *49*, 10149-10153.
- (13) Kumar, K. S.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. Total Chemical Synthesis of Di-Ubiquitin Chains. *Angew. Chem. Int. Ed.* **2011**, *49*, 9126-9131.
- (14) Kumar, K. S.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. Total Chemical Synthesis of a 304 Amino Acid K48-Linked Tetraubiquitin Protein. *Angew. Chem. Int. Ed.* **2011**, *50*, 6137-6141.
- (15) Sakamoto, I.; Tezuka, K.; Fukae, K.; Ishii, K.; Taduru, K.; Maeda, M.; Ouchi, M.; Yoshida, K.; Nambu, Y.; Igarashi, J. et al. Chemical Synthesis of Homogeneous Human Glycosyl-Interferon-Beta That Exhibits Potent Antitumor Activity in Vivo. *J. Am. Chem. Soc.* **2012**, *134*, 5428-5431.
- (16) Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y. Chemical Synthesis of an Erythropoietin Glycoform Containing a Complex-Type Disialyloligosaccharide. *Angew. Chem. Int. Ed.* **2012**, *51*, 3567-3572.
- (17) Unverzagt, C.; Kajihara, Y. Chemical Assembly of N-Glycoproteins: A Refined Toolbox to Address a Ubiquitous Posttranslational Modification. *Chem. Soc. Rev.* **2013**, *42*, 4408-4420.
- (18) Reif, A.; Siebenhaar, S.; Troster, A.; Schmalzlein, M.; Lechner, C.; Velisetty, P.; Gottwald, K.; Pohner, C.; Boos, I.; Schubert, V. et al. Semisynthesis of Biologically Active Glycoforms of the Human Cytokine Interleukin 6. *Angew. Chem. Int. Ed.* **2014**, *53*, 12125-12131.

- (19) Boll, E.; Drobecq, H.; Ollivier, N.; Raibaut, L.; Desmet, R.; Vicogne, J.; Melnyk, O. A Novel PEG-Based Solid Support Enables the Synthesis of >50 Amino-Acid Peptide Thioesters and the Total Synthesis of a Functional SUMO-1 Peptide Conjugate. *Chem. Sci.* **2014**, *5*, 2017-2022.
- (20) Boll, E.; Drobecq, H.; Ollivier, N.; Blanpain, A.; Raibaut, L.; Desmet, R.; Vicogne, J.; Melnyk, O. One-Pot Chemical Synthesis of Small Ubiquitin-Like Modifier (SUMO) Protein-Peptide Conjugates Using *Bis*(2-Sulfanylethyl)Amido Peptide Latent Thioester Surrogates *Nat. Protoc.* **2015**, *10*, 269-292.
- (21) Seenayah, M.; Jbara, M.; Mali, S. M.; Brik, A. Convergent Versus Sequential Protein Synthesis: The Case of Ubiquitinated and Glycosylated H2B. *Angew. Chem. Int. Ed.* **2015**, *54*, 12374-12378.
- (22) Pan, M.; Gao, S.; Zheng, Y.; Tan, X.; Lan, H.; Tan, X.; Sun, D.; Lu, L.; Wang, T.; Zheng, Q. et al. Quasi-Racemic X-Ray Structures of K27-Linked Ubiquitin Chains Prepared by Total Chemical Synthesis. *J. Am. Chem. Soc.* **2016**, *138*, 7429-7435.
- (23) Drobecq, H.; Boll, E.; Senechal, M.; Desmet, R.; Saliou, J. M.; Lacapere, J. J.; Mougel, A.; Vicogne, J.; Melnyk, O. A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide-Protein Conjugates. *Bioconjugate Chem.* **2016**, *27*, 1540-1546.
- (24) Tang, S.; Liang, L.-J.; Si, Y.-Y.; Gao, S.; Wang, J.-X.; Liang, J.; Mei, Z.; Zheng, J.-S.; Liu, L. Practical Chemical Synthesis of Atypical Ubiquitin Chains by Using an Isopeptide-Linked Ub Isomer. *Angew. Chem. Int. Ed.* **2017**, *56*, 13333-13337.
- (25) Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. Synthetic, Structural and Biological Studies of the Ubiquitin System: The Total Chemical Synthesis of Ubiquitin. *Biochem. J.* **1994**, *299* (Pt 1), 151-158.
- (26) Ogunjobi, O.; Ramage, R. Ubiquitin: Preparative Chemical Synthesis, Purification and Characterization. *Biochem. Soc. Trans.* **1990**, *18*, 1322-1323.
- (27) Huang, Y.-C.; Chen, C.-C.; Gao, S.; Wang, Y.-H.; Xiao, H.; Wang, F.; Tian, C.-L.; Li, Y.-M. Synthesis of L- and D-Ubiquitin by One-Pot Ligation and Metal-Free Desulfurization. *Chem. Eur. J.* **2016**, *22*, 7623-7628.
- (28) Qu, Q.; Pan, M.; Gao, S.; Zheng, Q.-Y.; Yu, Y.-Y.; Su, J.-C.; Li, X.; Hu, H.-G. A Highly Efficient Synthesis of Polyubiquitin Chains. *Adv. Sci.* **2018**, *5*, 1800234.
- (29) Bondalapati, S.; Eid, E.; Mali, S. M.; Wolberger, C.; Brik, A. Total Chemical Synthesis of SUMO-2-Lys63-Linked Diubiquitin Hybrid Chains Assisted by Removable Solubilizing Tags. *Chem. Sci.* **2017**, *8*, 4027-4034.
- (30) Pan, M.; Zheng, Q.; Ding, S.; Zhang, L.; Qu, Q.; Wang, T.; Hong, D.; Ren, Y.; Liang, L.; Chen, C. et al. Chemical Protein Synthesis Enabled Mechanistic Studies on the Molecular Recognition of K27-Linked Ubiquitin Chains. *Angew. Chem. Int. Ed.* **2019**, *58*, 2627-2631.
- (31) Wucherpfennig, T. G.; Pattabiraman, V. R.; Limberg, F. R. P.; Ruiz-Rodríguez, J.; Bode, J. W. Traceless Preparation of C-Terminal α -ketoacids for Chemical Protein Synthesis by α -ketoacid-hydroxylamine Ligation: Synthesis of SUMO2/3. *Angew. Chem. Int. Ed.* **2014**, *53*, 12248-12252.
- (32) Chatterjee, C.; McGinty, R. K.; Pellois, J. P.; Muir, T. W. Auxiliary-Mediated Site-Specific Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* **2007**, *46*, 2814-2818.
- (33) Geurink, P. P.; El Oualid, F.; Jonker, A.; Hameed, D. S.; Ovaa, H. A General Chemical Ligation Approach Towards Isopeptide-Linked Ubiquitin and Ubiquitin-Like Assay Reagents. *ChemBioChem* **2012**, *13*, 293-297.
- (34) Weller, C. E.; Dhall, A.; Ding, F.; Linares, E.; Whedon, S. D.; Senger, N. A.; Tyson, E. L.; Bagert, J. D.; Li, X.; Augusto, O. et al. Aromatic Thiol-Mediated Cleavage of N-O Bonds Enables Chemical Ubiquitylation of Folded Proteins. *Nat. Commun.* **2016**, *7*, No.

- (35) Weller, C. E.; Huang, W.; Chatterjee, C. Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *Chembiochem* **2014**, *15*, 1263-1267.
- (36) Spasser, L.; Brik, A. Chemistry and Biology of the Ubiquitin Signal. *Angew. Chem. Int. Ed.* **2012**, *51*, 6840-6862.
- (37) van Tilburg, G. B.; Elhebieshy, A. F.; Ovaa, H. Synthetic and Semi-Synthetic Strategies to Study Ubiquitin Signaling. *Curr. Opin. Struct. Biol.* **2016**, *38*, 92-101.
- (38) Matunis, M. J.; Coutavas, E.; Blobel, G. A Novel Ubiquitin-Like Modification Modulates the Partitioning of the Ran-GTPase-Activating Protein Rangap1 between the Cytosol and the Nuclear Pore Complex. *J. Cell. Biol.* **1996**, *135*, 1457-1470.
- (39) Lapenta, V.; Chiurazzi, P.; van der Spek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C. SMT3A, a Human Homologue of the *S. Cerevisiae* Smt3 Gene, Maps to Chromosome 21qter and Defines a Novel Gene Family. *Genomics* **1997**, *40*, 362-366.
- (40) Praefcke, G. J.; Hofmann, K.; Dohmen, R. J. SUMO Playing Tag with Ubiquitin. *Trends Biochem. Sci.* **2012**, *37*, 23-31.
- (41) McManus, F. P.; Lamoliatte, F.; Thibault, P. Identification of Cross Talk between Sumoylation and Ubiquitylation Using a Sequential Peptide Immunopurification Approach. *Nat. Protoc.* **2017**, *12*, 2342.
- (42) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive Compilation of SUMO Proteomics. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 581.
- (43) Hickey, C. M.; Wilson, N. R.; Hochstrasser, M. Function and Regulation of SUMO Proteases. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 755.
- (44) Flotho, A.; Melchior, F. Sumoylation: A Regulatory Protein Modification in Health and Disease. *Annu. Rev. Biochem.* **2013**, *82*, 357-385.
- (45) Jia, Y.; Claessens, L. A.; Vertegaal, A. C. O.; Ovaa, H. Chemical Tools and Biochemical Assays for SUMO Specific Proteases (Senps). *ACS Chem. Biol.* **2019**, 10.1021/acscchembio.1029b00402.
- (46) Kost, L. J.; Mootz, H. D. A FRET Sensor to Monitor Bivalent SUMO-SUMO Interactions in SUMO Chain Binding. *Chembiochem* **2018**, *19*, 177-184.
- (47) Xu, Y.; Plechanovova, A.; Simpson, P.; Marchant, J.; Leidecker, O.; Kraatz, S.; Hay, R. T.; Matthews, S. J. Structural Insight into SUMO Chain Recognition and Manipulation by the Ubiquitin Ligase Rnf4. *Nat. Commun.* **2014**, *5*, 4217.
- (48) Mulder, M. P. C.; Merckx, R.; Witting, K. F.; Hameed, D. S.; El Atmioui, D.; Lelieveld, L.; Liebelt, F.; Neefjes, J.; Berlin, I.; Vertegaal, A. C. O. et al. Total Chemical Synthesis of SUMO and SUMO-Based Probes for Profiling the Activity of SUMO-Specific Proteases. *Angew. Chem. Int. Ed.* **2018**, *57*, 8958-8962.
- (49) Melnyk, O.; Vicogne, J. Total Chemical Synthesis of SUMO Proteins. *Tetrahedron Lett.* **2016**, *57*, 4319-4324.
- (50) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. Dual Native Chemical Ligation at Lysine. *J. Am. Chem. Soc.* **2009**, *131*, 13592-13593.
- (51) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Synthesis of K48-Linked Diubiquitin Using Dual Native Chemical Ligation at Lysine. *Chem. Commun.* **2010**, *46*, 7199-7201.
- (52) Siman, P.; Karthikeyan, S. V.; Nikolov, M.; Fischle, W.; Brik, A. Convergent Chemical Synthesis of Histone H2B Protein for the Site-Specific Ubiquitination at Lys34. *Angew. Chem. Int. Ed.* **2013**, *52*, 8059-8063.
- (53) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776-779.
- (54) Haj-Yahya, N.; Haj-Yahya, M.; Castañeda, C. A.; Spasser, L.; Hemantha, H. P.; Jbara, M.; Penner, M.; Ciechanover, A.; Fushman, D.; Brik, A. Modifying the Vicinity of the Isopeptide Bond to Reveal Differential Behavior of Ubiquitin Chains with Interacting

- Proteins: Organic Chemistry Applied to Synthetic Proteins. *Angew. Chem. Int. Ed.* **2013**, *52*, 11149-11153.
- (55) McGinty, R. K.; Köhn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. Structure-Activity Analysis of Semisynthetic Nucleosomes: Mechanistic Insights into the Stimulation of Dot1l by Ubiquitylated Histone H2B. *ACS Chem. Biol.* **2009**, *4*, 958-968.
- (56) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123*, 526-533.
- (57) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed.* **2007**, *46*, 9248-9252.
- (58) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenicke, R.; Becker, J. Structure Determination of the Small Ubiquitin-Related Modifier SUMO-1. *J. Mol. Biol.* **1998**, *280*, 275-286.
- (59) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Bis*(2-Sulfanylethyl)Amino Native Peptide Ligation. *Org. Lett.* **2010**, *12*, 5238-5241.
- (60) Gong, L.; Millas, S.; Maul, G. G.; Yeh, E. T. Differential Regulation of Sentrinized Proteins by a Novel Sentrin-Specific Protease. *J. Biol. Chem.* **2000**, *275*, 3355-3359.
- (61) Yeh, E. T. H.; Gong, L.; Kamitani, T. Ubiquitin-Like Proteins: New Wines in New Bottles. *Gene* **2000**, *248*, 1-14.
- (62) Nishida, T.; Kaneko, F.; Kitagawa, M.; Yasuda, H. Characterization of a Novel Mammalian SUMO-1/Smt3-Specific Isopeptidase, a Homologue of Rat Axam, Which Is an Axin-Binding Protein Promoting B-Catenin Degradation. *J. Biol. Chem.* **2001**, *276*, 39060-39066.
- (63) Sommer, S.; Weikart, N. D.; Linne, U.; Mootz, H. D. Covalent Inhibition of SUMO and Ubiquitin-Specific Cysteine Proteases by an in Situ Thiol-Alkyne Addition. *Bioorg. Med. Chem.* **2013**, *21*, 2511-2517.
- (64) Dheur, J.; Ollivier, N.; Vallin, A.; Melnyk, O. Synthesis of Peptide Alkylthioesters Using the Intramolecular *N,S*-Acyl Shift Properties of *Bis*(2-Sulfanylethyl)Amido Peptides. *J. Org. Chem.* **2011**, *76*, 3194-3202.
- (65) Ollivier, N.; Vicogne, J.; Vallin, A.; Drobecq, H.; Desmet, R.; El-Mahdi, O.; Leclercq, B.; Goormachtigh, G.; Fafeur, V.; Melnyk, O. A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.* **2012**, *51*, 209-213.
- (66) Bhattacharya, S.; Ainaravaru, S. R. K. Mechanical Softening of a Small Ubiquitin-Related Modifier Protein Due to Temperature Induced Flexibility at the Core. *J. Phys. Chem. B* **2018**, *122*, 9128-9136.
- (67) Greenfield, N.; Fasman, G. D. Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry* **1969**, *8*, 4108-4116.
- (68) Bossis, G.; Chmielarska, K.; Gartner, U.; Pichler, A.; Stieger, E.; Melchior, F. A Fluorescence Resonance Energy Transfer-Based Assay to Study SUMO Modification in Solution. *Methods Enzymol.* **2005**, *398*, 20-32.
- (69) Stankovic-Valentin, N.; Kozaczkiwicz, L.; Curth, K.; Melchior, F. An in Vitro FRET-Based Assay for the Analysis of SUMO Conjugation and Isopeptidase Cleavage. *Methods Mol. Biol.* **2009**, *497*, 241-251.

TOC



Total Chemical Synthesis of All SUMO- 2/3 Dimer Combinations

Jennifer Bouchenna, Magalie Sénéchal, Hervé Drobecq,
Jérôme Vicogne,* Oleg Melnyk*

^a University of Lille, CNRS, Institut Pasteur de Lille, INSERM U1019, UMR CNRS 8204, Centre d'Immunité et d'Infection de Lille, F-59000 Lille, France.

email of corresponding author:

oleg.melnyk@ibl.cnrs.fr

Jerome.vicogne@ibl.cnrs.fr

ABSTRACT

One hallmark of protein chemical synthesis is its capacity to access proteins that living systems can hardly produce. This is typically the case for proteins harboring post-translational modifications such as ubiquitin or ubiquitin-like modifiers. Various methods have been developed for accessing polyubiquitin conjugates by semi or total synthesis. Comparatively, the preparation of small-ubiquitin like modifier (SUMO) conjugates and more particularly of polySUMO scaffolds is much less developed. We describe hereinafter a synthetic strategy for accessing all SUMO-2/3 dimer combinations.

KEY-WORDS

Protein chemical synthesis, SUMO dimers, SUMO-2, SUMO-3, selective desulfurization.

Introduction

Protein chemical synthesis is a powerful mean for accessing proteins harboring a large variety of post-translational modifications.¹⁻²⁶ Among the recent achievements in the field, the production of polyubiquitin chains^{7-11, 26-28} best illustrates the capacity of modern chemical synthesis to address the production of proteins of exceptional size and branching complexity with an atom-by-atom control of their structure.²⁹ Synthetic polyubiquitins open the possibility to investigate precisely how these modifications modulate protein structure and functions.³⁰⁻³² The chemical synthesis of small ubiquitin-like modifiers (SUMO)¹²⁻¹⁴ and SUMO conjugates^{15, 16, 21-25, 33} has been described too but the works in this area are few and far between the reports describing the synthesis of polyubiquitin scaffolds.³⁴ For example, the chemical synthesis of di-ubiquitin or tetra-ubiquitin chains has been reported by Brik and coworkers in 2011,^{8,9} while the chemical synthesis of homogeneous and branched polySUMO chains remains to be addressed.

Like ubiquitin (Ub), SUMO modifiers are attached by their C-terminal glycine residue to the side-chain of exposed lysine residues (Lys), usually located in SUMOylation consensus motifs within the target proteins. Five SUMO isoforms have been identified so far, among which SUMO-1-3 are constitutively expressed in all eukaryotic cells.^{35, 36} SUMO modification modulates the localization and the activity of the target proteins by changing protein-protein interactions and/or by competing with other lysine modifications such as ubiquitination or acetylation.^{37, 38} SUMO-2/3 isoforms differ significantly from SUMO-1 by their sequence and by containing an internal SUMOylation site enabling the formation of polySUMO-2/3 chains *in vivo* (Fig. 1A). Naturally, deciphering the role of polySUMO-2/3 modification stimulates huge efforts worldwide but progress in this area is complicated by the fact that SUMO conjugates and, in particular polySUMO chains, are branched proteins that cannot be precisely

programmed using classical recombinant techniques. Therefore, the chemical synthesis of well-defined polySUMO-2/3 scaffolds is a significant goal to pursue.

SUMO-2 and SUMO-3 feature a high sequence similarity and differ by only a few amino acids located in their N-terminal tail (Fig. 1A). However, we recently showed that SUMO-2 and SUMO-3 have different conformational and biochemical properties.³³ Contrary to SUMO-3, the structure of SUMO-2 is stabilized by a weak interaction occurring between the N-terminal tail and the SUMO core domain (Fig. 1B). Since SUMO-2 and SUMO-3 are not equivalent, polySUMO-2/3 chains might have their structure and function modulated by the proportion of each SUMO isoform within the chain. Another point to consider is the presence into the tail of SUMO-2 of the SUMOylation site, Lys¹¹. Although the exact nature of the SUMO-2 tail-core interaction remains to be established, its existence raise the question of the impact of Lys¹¹ modification on the structure and function of polySUMO-2/3 chains. Indeed, the properties of polySUMO-2/3 conjugates might vary not only with the type of SUMO isoforms that they are made of, but also with the order by which they are concatenated. For example, the properties of SUMO-2-SUMO-3 and SUMO-3-SUMO-2 dimers might not be same because the N-terminal tail of SUMO-2 is unmodified in the former case, while it hosts a SUMO modifier in the latter case (Fig. 1C).

An important step for investigating how the composition of polySUMO-2/3 chains impacts their properties is to access these molecules in homogeneous form and substantial amount, i.e., at mg scale. We describe hereinafter a rapid and robust access to all combinations of SUMO-2/3 dimers. The chemical synthesis of these proteins is challenging due to their size (~21 kDa). The described strategy gives also access to C-terminal SUMO-2/3 dimer hydrazides. By being useful protein thioester surrogates,³⁹⁻⁴² SUMO-2/3 dimer hydrazides allow to consider in the future the chemical synthesis of polySUMO scaffolds of even higher complexity.

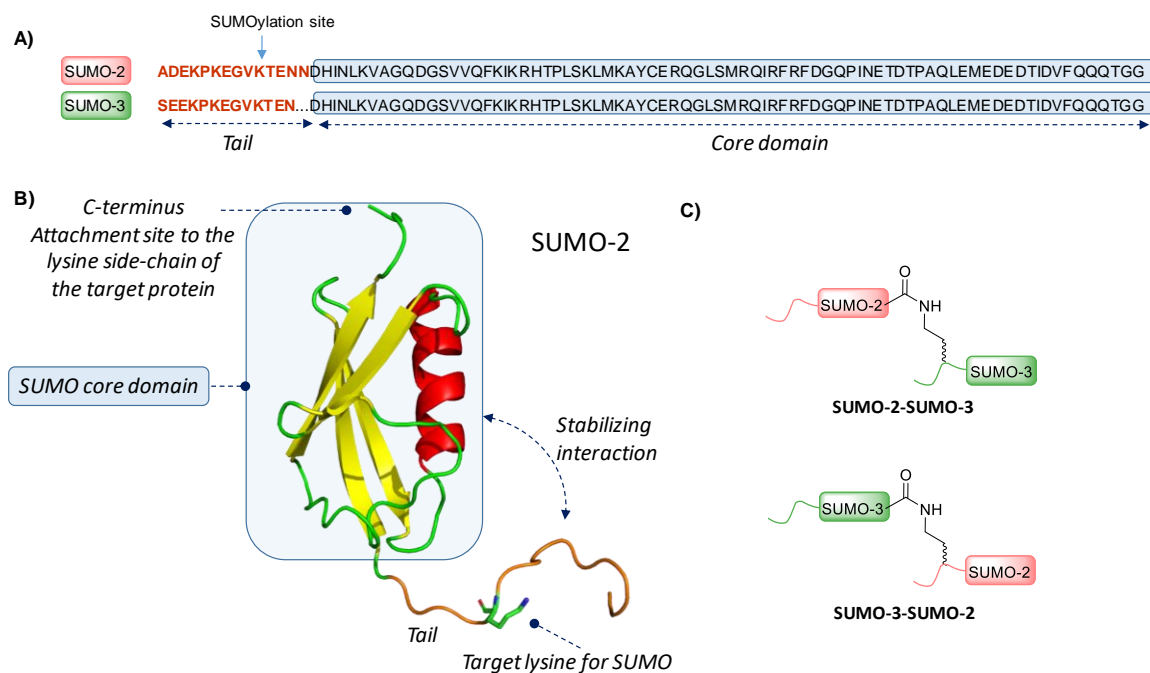


Figure 1. A) Primary structure of SUMO-2 and SUMO-3 proteins showing the tail and core domain. B) Tertiary structure of SUMO-2 (NMR structure, pdb entry 2N1W). C) Schematic representation of SUMO-2-SUMO-3 and SUMO-3-SUMO-2 dimers.

Results and discussion

Synthetic strategy

The strategy shown in Fig. 2 was adopted for accessing all four SUMO-2/3 dimers (illustrated with the synthesis of SUMO-2-SUMO-3). According to this strategy, the SUMO-2/3 dimers are assembled in one-pot by ligating three peptide segments. Two of these are produced by solid phase peptide synthesis (SPPS) owing to their length (~45AA) and are used to assemble the left SUMO domain highlighted in pink. The third segment in green used in the one-pot process corresponds to the right SUMO domain. It is produced in one-pot too and features an internal Lys(Cys) residue at position 11 to attach the SUMO domain coming from the first ligation step. Therefore, SUMO-2/3 dimers are accessed through two successive one-pot processes. This strategy enables limiting the number of resolutive HPLC purification steps that are known to induce important mass losses.^{43, 44}

The SUMO-2/3 dimers produced by this way need to be desulfurized subsequently to remove the thiol function in-between the two SUMO domains.⁴⁵ In the case of SUMO-2/3 dimers, a global desulfurization is not recommended because we recently showed that the desulfurization of the central Cys residue and its conversion into Ala results in the destabilization of the SUMO-2/3 domains, SUMO-3 being especially sensitive to this modification.³³ Therefore, the synthesis of SUMO-2/3 conjugates requires selectively removing the thiol from the thiol amino acid used to form the isopeptidic bond by NCL. The performance

of a selective desulfurization of SUMO conjugates featuring one SUMO domain has been described in a few studies.^{22, 24, 33} These reactions are performed in native conditions. The selectivity relies on a difference in accessibility between the thiol to be removed, which is solvent exposed, and the Cys thiol internal to the SUMO domain, which is buried in the hydrophobic core of the folded SUMO domain. However, extending the principle of a selective desulfurization to SUMO dimers for which the isopeptidic bond is located in between two folded domains had to be established.⁴⁶ Note that the burying of isopeptidic bonds in some polyubiquitin conjugates has been observed by Liu and coworkers.¹⁰

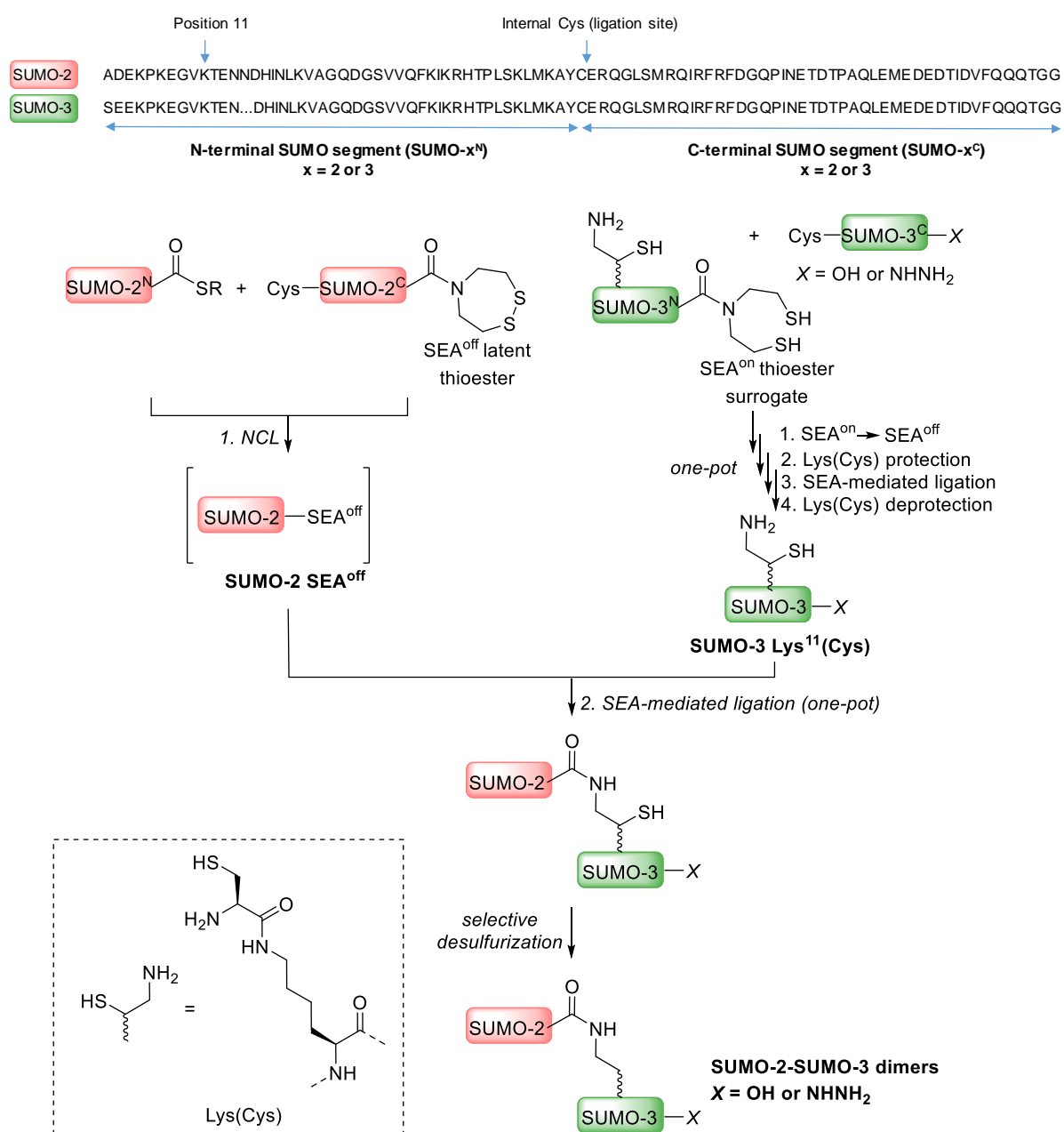


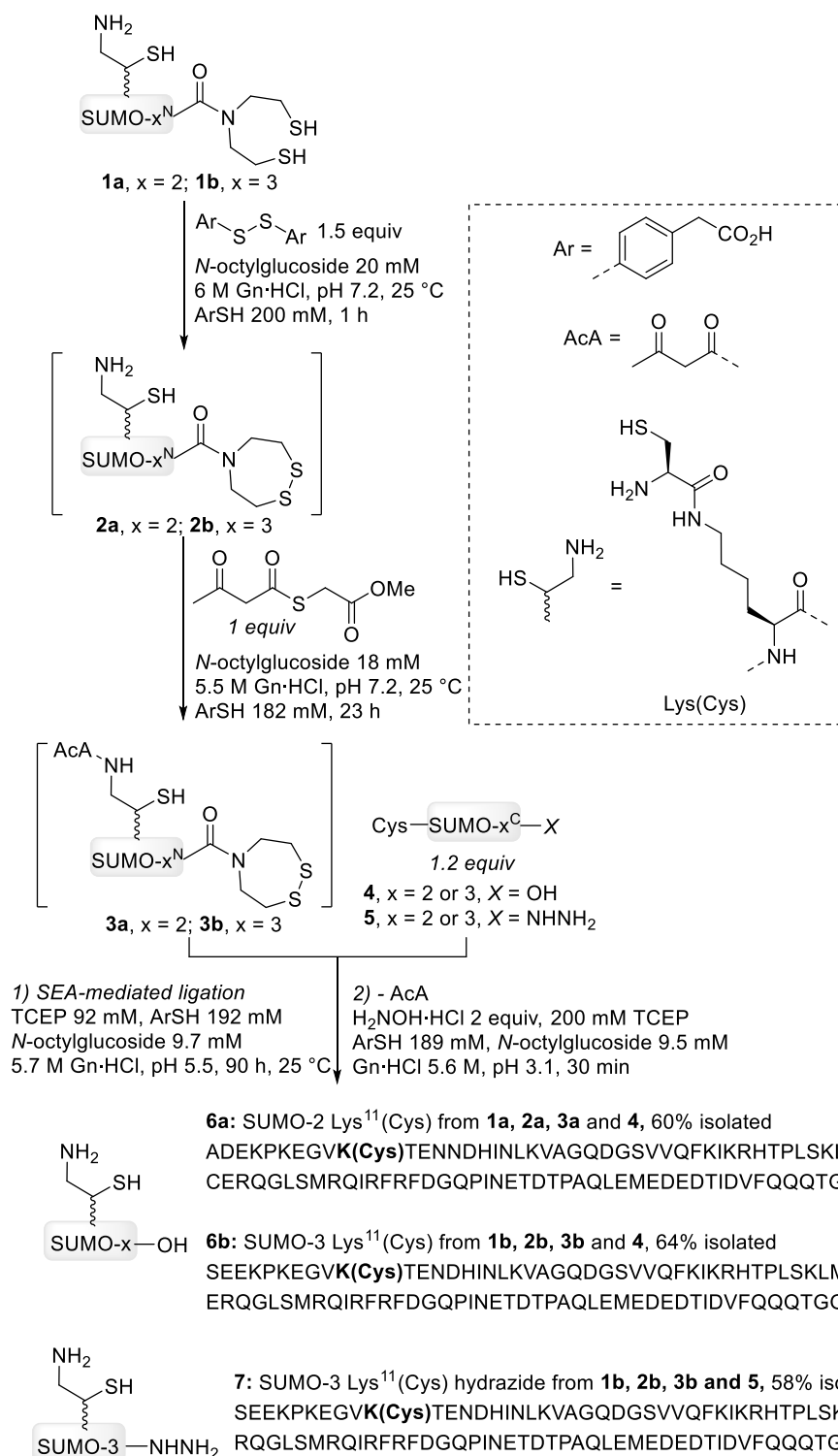
Figure 2. General synthetic strategy toward SUMO dimers taking the synthesis of SUMO-2-SUMO-3 as an example.

Synthesis of the starting peptide segments

According to the one-pot strategy depicted in Fig. 2, the first ligation step enables the assembly of SUMO-2 or SUMO-3 domains equipped with a C-terminal SEA functionality. The synthesis of the peptide segments used during this step, i.e. SUMO- x^N and SUMO- x^C ($x = 2$ or 3), were produced by conventional 9-fluorenylmethoxycarbonyl (Fmoc) SPPS using bis(2-sulfanylethyl)amino (SEA) ChemMatrix® resin as already described elsewhere.^{12, 14}

The one-pot synthesis of the SUMO-2 and 3 domains equipped with a Lys(Cys) residue at position 11, i.e. SUMO-2 Lys¹¹(Cys) and SUMO-3 Lys¹¹(Cys) (the peptide segment in green in Fig. 2), required novel developments that are described in Scheme 1. The starting SEA^{on} peptide segment **1a,b** was produced from bis(2-sulfanylethyl)amino (SEA) ChemMatrix® resin using Fmoc-Lys[Boc-Cys(Trt)]-OH for position 11. Such bifunctional peptide cannot be used as such for the assembly of the SUMO-2/3 Lys¹¹(Cys) domains because it can cyclize or oligomerize during the ligation with segments **4** or **5**. Therefore, the SEA^{on} group was first inactivated *in situ* by adding 4-mercaptophenylacetic acid (MPAA) disulfide to the reaction vessel. MPAA disulfide oxidizes the SEA^{on} group into the corresponding SEA^{off} cyclic disulfide by a thiol-disulfide exchange mechanism. The MPAA arylthiol produced during this step contributes to the catalysis of subsequent ligation reactions.⁴⁷ Then, a thioester derivative of acetoacetic acid (AcA) was added to protect temporarily the Lys¹¹(Cys) residue through an NCL reaction.⁴⁸ A tight control of the stoichiometry is required during this step as an excess of AcA thioester would inevitably cap the Cys segment introduced in the subsequent step. The AcA-protected segment **3a,b** was not isolated and was ligated directly with segment **4** or **5** by adding an excess of MPAA and tris(2-carboxyethyl)phosphine (TCEP). Finally, the AcA group was removed in one-pot by adding two equivalents of hydroxylamine hydrochloride to the ligation mixture. The presence of non-ionic detergent *N*-octylglucoside in the reaction mixture was found to be mandatory to avoid the partial precipitation of the peptides. Peptide segments SUMO-2/3 Lys¹¹(Cys) **6a,b** were isolated in good overall yields (58-60% after HPLC purification) and purity (see Supporting Information). The strategy was also applied successfully to the synthesis of a SUMO-3 Lys¹¹(Cys) derivative featuring a C-terminal hydrazide group, i.e., SUMO-3 Lys¹¹(Cys) **7**.

Scheme 1. One-pot synthesis of SUMO-2 or 3 peptide segments equipped with a Lys(Cys) residue at position 11

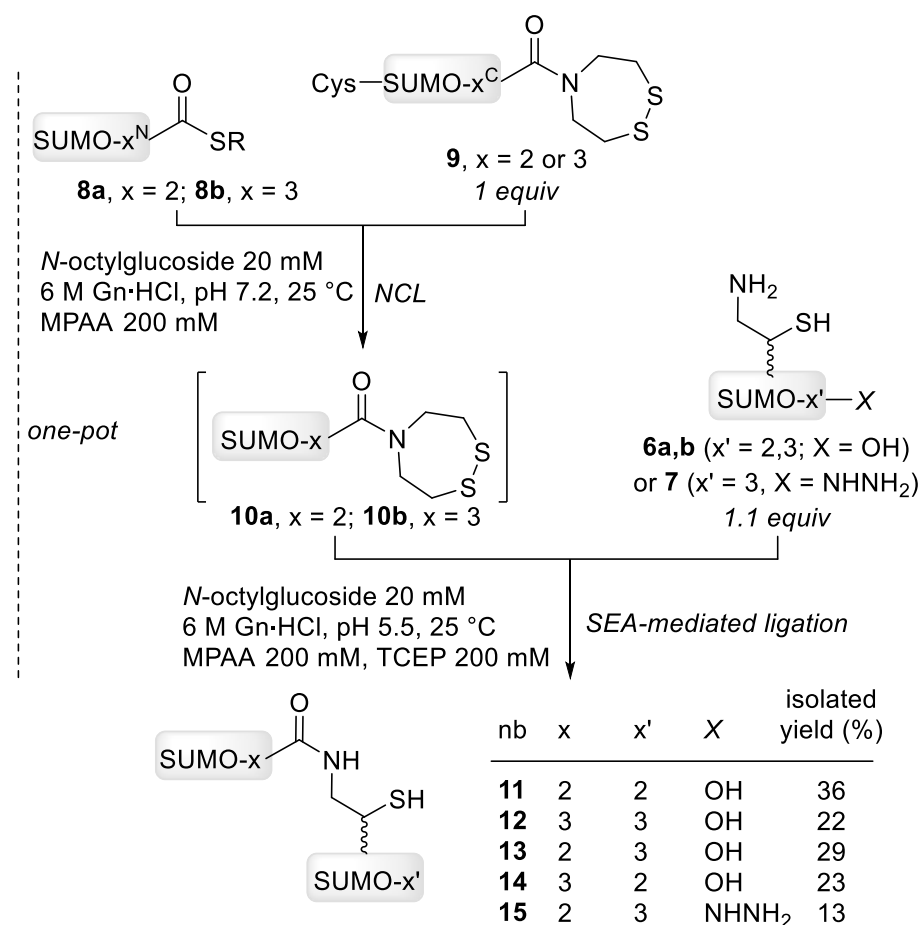


Assembly of SUMO-2/3 dimers

The successful synthesis of peptide segments **6a,b** and **7** set the stage for the assembly of the SUMO-2/3 dimers according to Scheme 2, which is an application of the general strategy

depicted in Fig. 2.⁴⁹ As for the synthesis of SUMO-2/3 Lys¹¹(Cys) domains, the use of *N*-octylglucoside and 6 M Gn·HCl as additives ensured the solubilization of starting materials and products throughout the process. The different 21 kDa SUMO-2/3 dimers **11-15** were produced successfully and isolated by HPLC in excellent purity (Fig. 3). They migrate to an apparent molecular weight of 30 kDa by SDS-PAGE (Fig. 4). These molecules that are composed of 183-185 amino acid residues are a novel illustration of the powerfulness of *N,S*-acyl shift thioester surrogates for accessing challenging proteins.^{50, 51}

Scheme 2. Chemical synthesis of SUMO-2/3 dimers



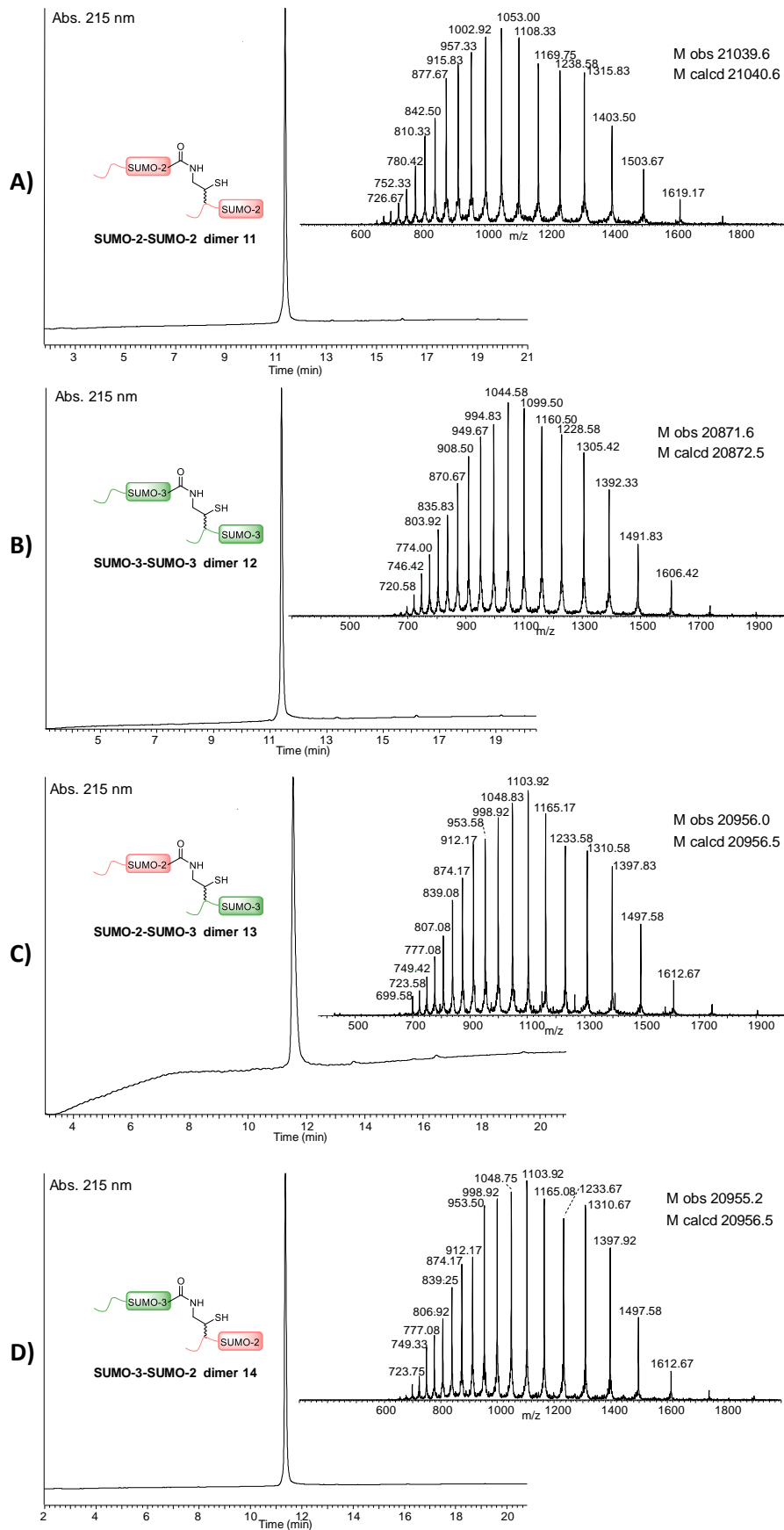


Figure 3. LC-MS analysis of isolated SUMO-2/3 dimers 11-14.

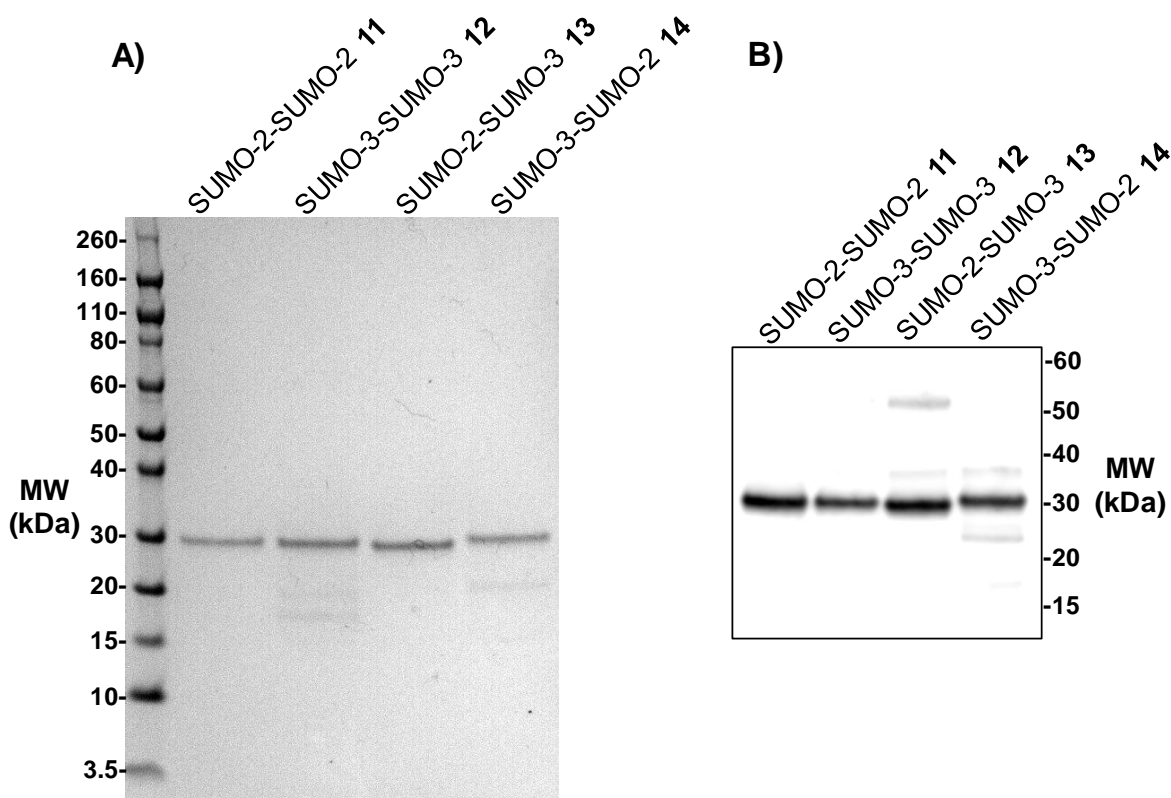
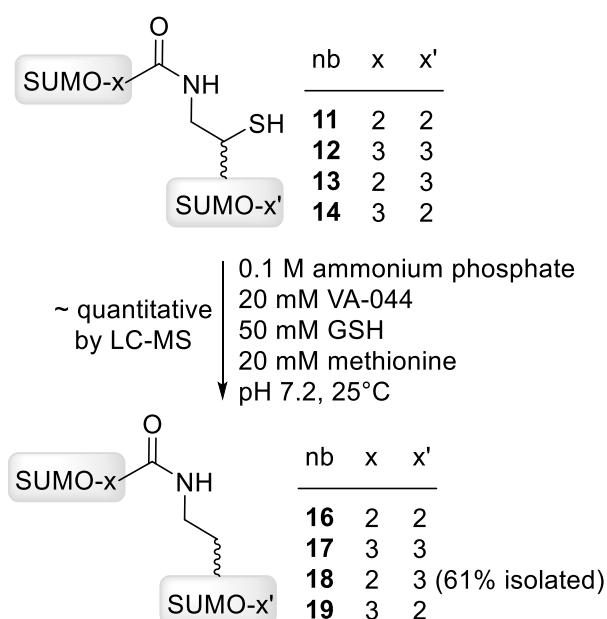


Figure 4. SDS-PAGE analysis of SUMO-2/3 dimers **11-14**. A) Coomassie staining, 1 μ g of each protein. B) Western-blot analysis using anti-SUMO-2/3 polyclonal antibody, 100 ng of each protein. MW: apparent molecular weight markers in kDa.

Selective desulfurization of SUMO-2/3 dimers 11-14

The successful synthesis of the different SUMO-2/3 dimers set the stage for the desulfurization experiments (Scheme 3). The TCEP-induced desulfurization of all four dimers in native conditions proceeded with a total selectivity for the Cys thiol in between the two SUMO domains, as demonstrated by extensive proteomic analysis of the desulfurized products (see Supporting Information). In a control experiment, SUMO 2/3 dimer **13** was desulfurized in 6 M Gn-HCl. In this case, all Cys residues were converted to Ala. Thus under native conditions, the Cys thiol used for linking the SUMO domains remains highly accessible to the solvent whenever the composition of the dimer. The efficient protection of internal Cys residues within SUMO domains under native conditions shows that both SUMO domains are properly folded. SUMO-2-SUMO-3 dimer **13** was desulfurized on a preparative scale to produce the desulfurized analog **18** in 61% isolated yield.

Scheme 3. Selective desulfurization of SUMO-2/3 dimers 11-14



Analysis of SUMO-2-SUMO-3 dimers **13** and **18**

SUMO-2-SUMO-3 dimers **13** and **18** were analysed by circular dichroism (CD) as shown in Fig. 5. The CD analysis indicates that the secondary structure of the dimers and in particular their α -helical content (18%) are highly similar. The α -helical content found for the dimers is consistent with the CD α -helical content of individual SUMO-2 and SUMO-3 domains which is 15%.³³ Moreover, the CD spectra of SUMO dimers **13** and **18** are similar to the reconstituted CD spectrum for SUMO-2-SUMO-3 dimer calculated using the experimental CD spectra obtained for individual SUMO-2 and SUMO-3 domains.³³

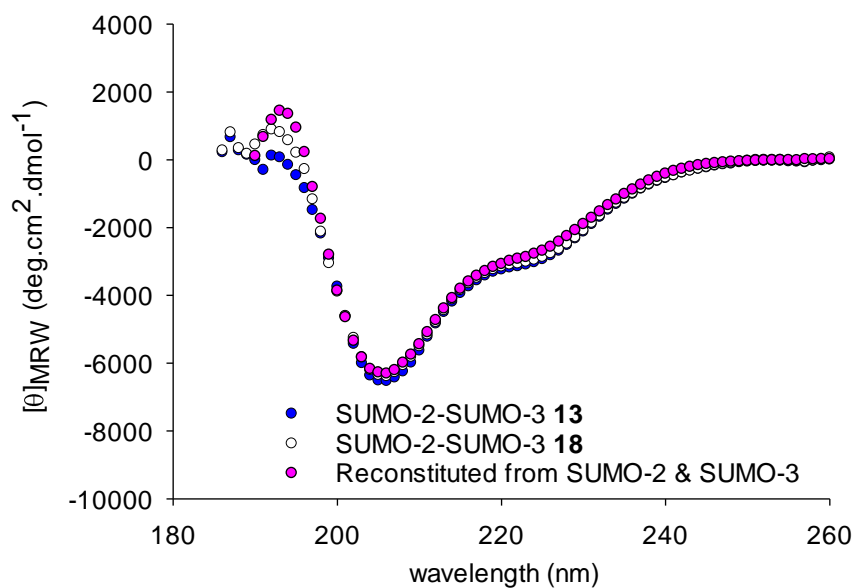


Figure 5. CD analysis of SUMO-2/3 dimers **13** and **18**. Comparison with a reconstituted spectrum calculated using the CD spectra of individual SUMO-2 and SUMO-3 proteins (data taken from ref³³).

We further examined the cleavage of SUMO-2-SUMO-3 dimer **18** by Sentrin-specific protease 1 (SEN1⁵²) and 2 (SEN2^{53, 54}), which are human SUMO-1, 2 and 3 specific deconjugating enzymes (Fig. 6).^{55, 56} SUMO-2-SUMO-3 dimer **18** was cleaved by both SENP enzymes to produce SUMO-2 and SUMO-3 Lys¹¹(Ala) proteins that were separated SDS-PAGE and identified by both Coomassie staining and Western blot analysis at an apparent molecular weight of ~15 kDa. The breakdown of SUMO-2-SUMO-3 dimer **18** into individual SUMO-2/3 proteins was confirmed by LC-MS analysis of the cleavage mixture. Taken together, the CD signature of SUMO-2-SUMO-3 dimer **18** and its cleavage by SENP1 and SENP2 enzymes show that the dimer is properly folded and functional.

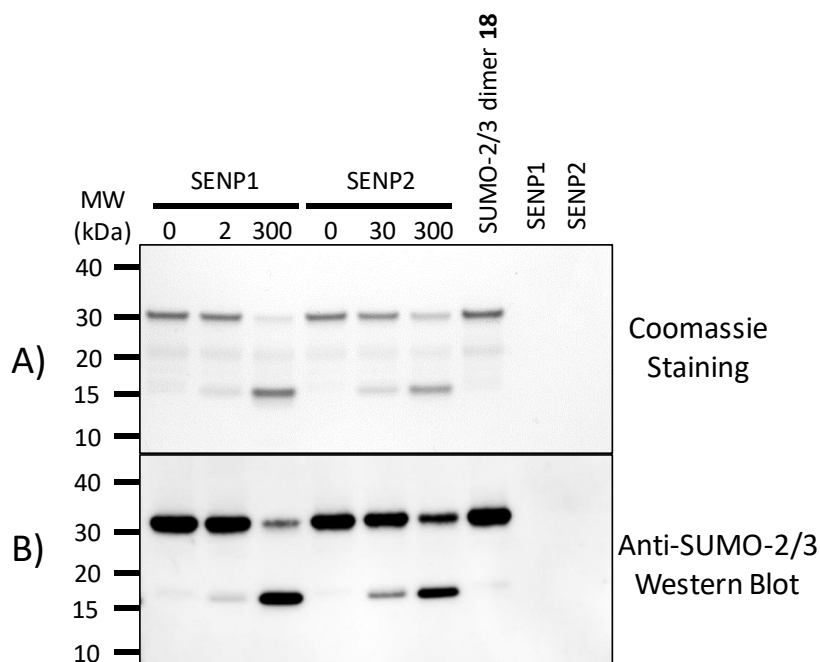


Figure 6. Cleavage of SUMO-2-SUMO-3 dimer **18** by SENP1 and SENP2 deconjugating enzymes (time in minutes). A) Coomassie staining, 2 µg of SUMO-2-SUMO-3 dimer **18**. B) Western-blot analysis using anti-SUMO-2/3 polyclonal antibody, 100 ng of SUMO-2-SUMO-3 dimer **18**. MW: apparent molecular weight markers in kDa.

CONCLUSIONS

We describe in this report a simple and efficient access to functional SUMO-2/3 dimers that is compatible with the production of SUMO dimers featuring a C-terminal hydrazide functionality. The preparation of SUMO dimer hydrazides opens the possibility to access more elaborated polySUMO conjugates by exploiting the thioester surrogate properties of the hydrazide group. The desulfurization of the thiol amino acid used to produce the isopeptidic bond has been performed with a high degree of selectivity, regardless the SUMO-2/3 dimer combination examined. The production of SUMO-2/3 dimers at the mg scale is an important step toward further investigations aiming at determining the effect of SUMO dimer composition and order of concatenation on their biophysical and biochemical properties.

ACKNOWLEDGEMENTS

We thank the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation for financial support (PhD fellowship to Jennifer Bouchenna). We thank Jean-Jacques Lacapère and Florine Cavelier for useful discussions.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at DOI: #####.

Experimental procedures and characterization data for all peptides and proteins (LC-MS, MALDI-TOF, and proteomic analyses) (PDF).

References

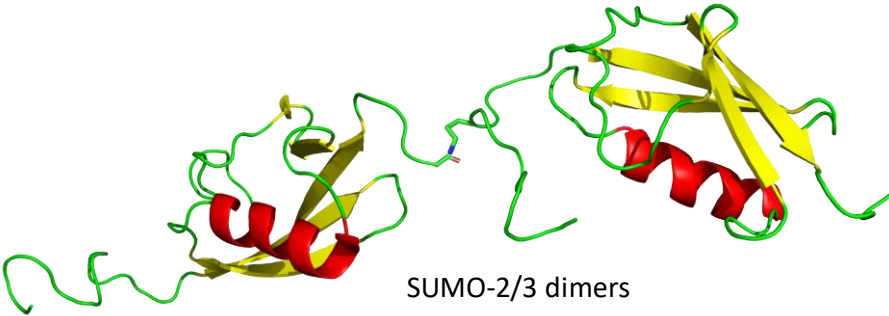
- (1) Li, J., Li, Y., He, Q., Li, Y., Li, H., and Liu, L. (2014) One-Pot Native Chemical Ligation of Peptide Hydrazides Enables Total Synthesis of Modified Histones. *Org. Biomol. Chem.* *12*, 5435-5441.
- (2) Chiang, K. P., Jensen, M. S., McGinty, R. K., and Muir, T. W. (2009) A Semisynthetic Strategy to Generate Phosphorylated and Acetylated Histone H2B. *ChemBioChem* *10*, 2182-2187.
- (3) Yu, R. R., Mahto, S. K., Justus, K., Alexander, M. M., Howard, C. J., and Ottesen, J. J. (2016) Hybrid Phase Ligation for Efficient Synthesis of Histone Proteins. *Org. Biomol. Chem.* *14*, 2603-2607.
- (4) Jbara, M., Maity, S. K., Morgan, M., Wolberger, C., and Brik, A. (2016) Chemical Synthesis of Phosphorylated Histone H2A at Tyr57 Reveals Insight into the Inhibition Mode of the SAGA Deubiquitinating Module. *Angew. Chem. Int. Ed.* *55*, 4972-4976.
- (5) Thompson, R. E., Liu, X., Ripoll-Rozada, J., Alonso-Garcia, N., Parker, B. L., Pereira, P. J. B., and Payne, R. J. (2017) Tyrosine Sulfation Modulates Activity of Tick-Derived Thrombin Inhibitors. *Nat. Chem.* *9*, 909-917.
- (6) Watson, E. E., Liu, X., Thompson, R. E., Ripoll-Rozada, J., Wu, M., Alwis, I., Gori, A., Loh, C. T., Parker, B. L., Otting, G., et al. (2018) Mosquito-Derived Anophelin Sulfoproteins Are Potent Antithrombotics. *ACS Cent. Sci.* *4*, 468-476.
- (7) El Oualid, F., Merckx, R., Ekkebus, R., Hameed, D. S., Smit, J. J., de Jong, A., Hilkmann, H., Sixma, T. K., and Ovaa, H. (2010) Chemical Synthesis of Ubiquitin, Ubiquitin-Based Probes, and Diubiquitin. *Angew. Chem. Int. Ed.* *49*, 10149-10153.
- (8) Kumar, K. S., Spasser, L., Erlich, L. A., Bavikar, S. N., and Brik, A. (2011) Total Chemical Synthesis of Di-Ubiquitin Chains. *Angew. Chem. Int. Ed.* *49*, 9126-9131.
- (9) Kumar, K. S., Bavikar, S. N., Spasser, L., Moyal, T., Ohayon, S., and Brik, A. (2011) Total Chemical Synthesis of a 304 Amino Acid K48-Linked Tetraubiquitin Protein. *Angew. Chem. Int. Ed.* *50*, 6137-6141.
- (10) Pan, M., Gao, S., Zheng, Y., Tan, X., Lan, H., Tan, X., Sun, D., Lu, L., Wang, T., Zheng, Q., et al. (2016) Quasi-Racemic X-Ray Structures of K27-Linked Ubiquitin Chains Prepared by Total Chemical Synthesis. *J. Am. Chem. Soc.* *138*, 7429-7435.
- (11) Tang, S., Liang, L.-J., Si, Y.-Y., Gao, S., Wang, J.-X., Liang, J., Mei, Z., Zheng, J.-S., and Liu, L. (2017) Practical Chemical Synthesis of Atypical Ubiquitin Chains by Using an Isopeptide-Linked Ub Isomer. *Angew. Chem. Int. Ed.* *56*, 13333-13337.
- (12) Boll, E., Drobecq, H., Ollivier, N., Raibaut, L., Desmet, R., Vicogne, J., and Melnyk, O. (2014) A Novel PEG-Based Solid Support Enables the Synthesis of >50 Amino-Acid Peptide Thioesters and the Total Synthesis of a Functional SUMO-1 Peptide Conjugate. *Chem. Sci.* *5*, 2017-2022.
- (13) Wucherpennig, T. G., Pattabiraman, V. R., Limberg, F. R. P., Ruiz-Rodríguez, J., and Bode, J. W. (2014) Traceless Preparation of C-Terminal α -ketoacids for Chemical Protein Synthesis by α -Ketoacid-Hydroxylamine Ligation: Synthesis of SUMO2/3. *Angew. Chem. Int. Ed.* *53*, 12248-12252.

- (14) Boll, E., Drobecq, H., Ollivier, N., Blanpain, A., Raibaut, L., Desmet, R., Vicogne, J., and Melnyk, O. (2015) One-Pot Chemical Synthesis of Small Ubiquitin-Like Modifier (SUMO) Protein-Peptide Conjugates Using *Bis*(2-Sulfanylethyl)Amido Peptide Latent Thioester Surrogates *Nat. Protoc.* *10*, 269-292.
- (15) McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G., and Muir, T. W. (2008) Chemically Ubiquitylated Histone H2B Stimulates hDot1L-Mediated Intranucleosomal Methylation. *Nature* *453*, 812-816.
- (16) Ajish Kumar, K. S., Haj-Yahya, M., Olschewski, D., Lashuel, H. A., and Brik, A. (2009) Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* *48*, 8090-8094.
- (17) Sakamoto, I., Tezuka, K., Fukae, K., Ishii, K., Taduru, K., Maeda, M., Ouchi, M., Yoshida, K., Nambu, Y., Igarashi, J., et al. (2012) Chemical Synthesis of Homogeneous Human Glycosyl-Interferon-Beta That Exhibits Potent Antitumor Activity in Vivo. *J. Am. Chem. Soc.* *134*, 5428-5431.
- (18) Murakami, M., Okamoto, R., Izumi, M., and Kajihara, Y. (2012) Chemical Synthesis of an Erythropoietin Glycoform Containing a Complex-Type Disialyloligosaccharide. *Angew. Chem. Int. Ed.* *51*, 3567-3572.
- (19) Unverzagt, C., and Kajihara, Y. (2013) Chemical Assembly of N-Glycoproteins: A Refined Toolbox to Address a Ubiquitous Posttranslational Modification. *Chem. Soc. Rev.* *42*, 4408-4420.
- (20) Reif, A., Siebenhaar, S., Troster, A., Schmalzlein, M., Lechner, C., Velisetty, P., Gottwald, K., Pohner, C., Boos, I., Schubert, V., et al. (2014) Semisynthesis of Biologically Active Glycoforms of the Human Cytokine Interleukin 6. *Angew. Chem. Int. Ed.* *53*, 12125-12131.
- (21) Seenaiyah, M., Jbara, M., Mali, S. M., and Brik, A. (2015) Convergent Versus Sequential Protein Synthesis: The Case of Ubiquitinated and Glycosylated H2B. *Angew. Chem. Int. Ed.* *54*, 12374-12378.
- (22) Drobecq, H., Boll, E., Senechal, M., Desmet, R., Saliou, J. M., Lacapere, J. J., Mougel, A., Vicogne, J., and Melnyk, O. (2016) A Central Cysteine Residue Is Essential for the Thermal Stability and Function of SUMO-1 Protein and SUMO-1 Peptide-Protein Conjugates. *Bioconjugate Chem.* *27*, 1540-1546.
- (23) Chatterjee, C., McGinty, R. K., Pellois, J. P., and Muir, T. W. (2007) Auxiliary-Mediated Site-Specific Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* *46*, 2814-8.
- (24) Geurink, P. P., El Oualid, F., Jonker, A., Hameed, D. S., and Ovaas, H. (2012) A General Chemical Ligation Approach Towards Isopeptide-Linked Ubiquitin and Ubiquitin-Like Assay Reagents. *Chembiochem* *13*, 293-297.
- (25) Weller, C. E., Dhall, A., Ding, F., Linares, E., Whedon, S. D., Senger, N. A., Tyson, E. L., Bagert, J. D., Li, X., Augusto, O., et al. (2016) *Nat. Commun.* DOI: 10.1038/ncomms12979.
- (26) Pan, M., Zheng, Q., Ding, S., Zhang, L., Qu, Q., Wang, T., Hong, D., Ren, Y., Liang, L., Chen, C., et al. (2019) Chemical Protein Synthesis Enabled Mechanistic Studies on the Molecular Recognition of K27-Linked Ubiquitin Chains. *Angew. Chem. Int. Ed.* *58*, 2627-2631.
- (27) Qu, Q., Pan, M., Gao, S., Zheng, Q.-Y., Yu, Y.-Y., Su, J.-C., Li, X., and Hu, H.-G. (2018) A Highly Efficient Synthesis of Polyubiquitin Chains. *Adv. Sci.* *5*, DOI: 10.1002/advs.201800234.
- (28) Xu, L., Huang, J.-F., Chen, C.-C., Qu, Q., Shi, J., Pan, M., and Li, Y.-M. (2018) Chemical Synthesis of Natural Polyubiquitin Chains through Auxiliary-Mediated Ligation of an Expressed Ubiquitin Isomer. *Org. Lett.* *20*, 329-332.

- (29) Bondalapati, S., Jbara, M., and Brik, A. (2016) Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nat. Chem.* 8, 407-418.
- (30) McGinty, R. K., Chatterjee, C., and Muir, T. W. (2009) Semisynthesis of Ubiquitylated Proteins. *Methods Enzymol.* 462, 225-243.
- (31) Chatterjee, C., and Muir, T. W. (2010) Chemical Approaches for Studying Histone Modifications. *J. Biol. Chem.* 285, 11045-11050.
- (32) Spasser, L., and Brik, A. (2012) Chemistry and Biology of the Ubiquitin Signal. *Angew. Chem. Int. Ed.* 51, 6840-6862.
- (33) Bouchenna, J., Sénéchal, M., Drobecq, D., Stankovic-Valentin, N., Vicogne, J., and Melnyk, O. (2019) The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis. *Bioconjugate Chem.* 30, 2684-2696.
- (34) Melnyk, O., and Vicogne, J. (2016) Total Chemical Synthesis of SUMO Proteins. *Tetrahedron Lett.* 57, 4319-4324.
- (35) Matunis, M. J., 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. *J. Cell. Biol.* 135, 1457-1470.
- (36) Lapenta, V., Chiurazzi, P., van der Spek, P., Pizzuti, A., Hanaoka, F., and Brahe, C. (1997) SMT3A, a Human Homologue of the *S. Cerevisiae* Smt3 Gene, Maps to Chromosome 21qter and Defines a Novel Gene Family. *Genomics* 40, 362-366.
- (37) Cappadocia, L., and Lima, C. D. (2018) Ubiquitin-Like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* 118, 889-918.
- (38) Vertegaal, A. C. O. (2011) Uncovering Ubiquitin and Ubiquitin-Like Signaling Networks. *Chem. Rev.* 111, 7923-7940.
- (39) Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P., and Liu, L. (2013) Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protocols* 8, 2483-2495.
- (40) Fang, G.-M., Wang, J.-X., and Liu, L. (2012) Convergent Chemical Synthesis of Proteins by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* 51, 10347-10350.
- (41) Fang, G.-M., Li, Y.-M., Shen, F., Huang, Y.-C., Li, J.-B., Lin, Y., Cui, H.-K., and Liu, L. (2011) Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* 50, 7645-7649.
- (42) Flood, D. T., Hintzen, J. C. J., Bird, M. J., Cistrone, P. A., Chen, J. S., and Dawson, P. E. (2018) Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.* 57, 11634-11639.
- (43) Loibl, S. F., Harpaz, Z., Zitterbart, R., and Seitz, O. (2016) Total Chemical Synthesis of Proteins without HPLC Purification. *Chem. Sci.* 7, 6753-6759.
- (44) Raibaut, L., El Mahdi, O., and Melnyk, O. (2015) Solid Phase Protein Chemical Synthesis. *Top. Curr. Chem.* 363, 103-154.
- (45) Wan, Q., and Danishefsky, S. J. (2007) Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed.* 46, 9248-9252.
- (46) Keusekotten, K., Bade, V. N., Meyer-Teschendorf, K., Sriramachandran, A. M., Fischer-Schrader, K., Krause, A., Horst, C., Schwarz, G., Hofmann, K., Dohmen, R. J., et al. (2014) Multivalent Interactions of the SUMO-Interaction Motifs in Ring Finger Protein 4 Determine the Specificity for Chains of the SUMO. *Biochem. J.* 457, 207-214.
- (47) Johnson, E. C., and Kent, S. B. H. (2006) Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* 128, 6640-6646.

- (48) Boll, E., Ebran, J. P., Drobecq, H., El-Mahdi, O., Raibaut, L., Ollivier, N., and Melnyk, O. (2015) Access to Large Cyclic Peptides by a One-Pot Two-Peptide Segment Ligation/Cyclization Process. *Org. Lett.* 17, 130-133.
- (49) Ollivier, N., Vicogne, J., Vallin, A., Drobecq, H., Desmet, R., El-Mahdi, O., Leclercq, B., Goormachtigh, G., Fafeur, V., and Melnyk, O. (2012) A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.* 51, 209-213.
- (50) Agouridas, V., El Mahdi, O., Cargoët, M., and Melnyk, O. (2017) A Statistical View of Protein Chemical Synthesis Using NCL and Extended Methodologies. *Bioorg. Med. Chem.* 25, 4938-4945.
- (51) Sun, H., and Brik, A. (2019) The Journey for the Total Chemical Synthesis of a 53 kDa Protein. *Acc. Chem. Res.*, 10.1021/acs.accounts.9b00372.
- (52) Gong, L., Millas, S., Maul, G. G., and Yeh, E. T. (2000) Differential Regulation of Sentrinized Proteins by a Novel Sentrin-Specific Protease. *J. Biol. Chem.* 275, 3355-3359.
- (53) Yeh, E. T. H., Gong, L., and Kamitani, T. (2000) Ubiquitin-Like Proteins: New Wines in New Bottles. *Gene* 248, 1-14.
- (54) Nishida, T., Kaneko, F., Kitagawa, M., and Yasuda, H. (2001) Characterization of a Novel Mammalian SUMO-1/Smt3-Specific Isopeptidase, a Homologue of Rat Axam, Which Is an Axin-Binding Protein Promoting B-Catenin Degradation. *J. Biol. Chem.* 276, 39060-39066.
- (55) Hickey, C. M., Wilson, N. R., and Hochstrasser, M. (2012) Function and Regulation of SUMO Proteases. *Nat. Rev. Mol. Cell Biol.* 13, 755-766.
- (56) Sommer, S., Weikart, N. D., Linne, U., and Mootz, H. D. (2013) Covalent Inhibition of SUMO and Ubiquitin-Specific Cysteine Proteases by an in Situ Thiol-Alkyne Addition. *Bioorg. Med. Chem.* 21, 2511-2517.

TOC



METHODS IN MOLECULAR BIOLOGY

Article 1: The problem of aspartimide formation during protein chemical synthesis using SEA-mediated ligation

Publication Date: December 12, 2019
DOI: 0000-0002-3863-5613

The problem of aspartimide formation during protein chemical synthesis using SEA-mediated ligation

Jennifer Bouchenna, Magalie Sénéchal, Hervé Drobecq, Jérôme Vicogne & Oleg Melnyk

CNRS, Pasteur Institute of Lille, University of Lille, UMR CNRS 8204, Lille, France.
Correspondence should be addressed to O.M. (oleg.melnyk@ibl.cnrs.fr).

Summary

Aspartimide formation often complicates the solid phase synthesis of peptides. Much less discussed is the potential occurrence of this side-reaction during the coupling of peptide segments using chemoselective peptide bond forming reactions such as the native chemical ligation and extended methods. Here we describe how to manage this problem using *bis*(2-sulfenylethyl)amido (SEA)-mediated ligation.

Key words: Aspartimide, *bis*(2-sulfenylethyl)amido (SEA)-mediated ligation, SPPS

1. Introduction

Aspartimide formation is a well-known side-reaction, which often complicates the solid phase peptide synthesis (SPPS⁽¹⁾). This side-reaction is due to the carboxylic acid group of the aspartic acid side-chain, which can react with the α -nitrogen of the next residue to form a succinimide moiety called aspartimide. The aspartimide can eventually be opened by nucleophiles, typically water or hydroxide ion, to produce a peptide having a natural backbone or an isopeptidic structure depending on which aspartimide carbonyl is attacked by the nucleophile. Note that the opening of an aspartimide yields usually the isopeptide are the major product. Note also that aspartimides are prone to epimerization. Therefore, aspartimide formation can ultimately result in the formation of several byproducts. Several tools are now available for minimizing this side reaction during SPPS such as optimized Fmoc deprotection cocktails, side-chain protecting groups for aspartic acid or backbone protecting groups for the residue following Asp.⁽²⁾

Aspartimide byproduct formation can also spontaneously occur in aqueous solution at neutral pH and as such is one of the mechanisms that lead to the degradation of proteins in vivo with aging. (3) The occurrence of this side-reaction in proteins, (3) peptide therapeutics (4) and model peptides (5-7) featuring Asn or Asp residues has been the subject of numerous studies. The occurrence of Asp/Asn-Gly dipeptides in proteins, which are particularly prone to aspartimide formation, is on average 2.7 units per protein (1.2 for Asp-Gly, 1.5 for Asn-Gly, data extracted from UniRef50 (8)). Therefore, such aspartimide-prone sites are frequent and can considerably complicate the synthesis of most proteins (see Note 1), regardless of the type of ligation used. (9)

The chemical synthesis of SUMO-2 and SUMO-3 proteins has been described in a few works using either ketoacid-hydroxylamine ligation (10) or optimized SPPS protocols. (11) In each case, the SPPS of the peptide segments was complicated by aspartimide byproduct formation due to the presence of two Asp-Gly units in SUMO-2 and 3 domains (Figure 1). Aspartimide byproduct formation during SPPS could be suppressed by using Fmoc-Asp(OtBu)-(Dmb)Gly-OH dipeptide unit for introducing the Asp and Gly residues.

During our investigations on the chemical synthesis of SUMO-2 and 3 proteins by SEA-mediated ligation, (12) we also found that the use of Fmoc-Asp(OtBu)-(Dmb)Gly-OH dipeptide unit during the SPPS of the peptide segments was mandatory for suppressing aspartimide byproduct formation. Performing the ligation of the peptide segments under classical experimental conditions (pH 5.5, 4-mercaptophenylacetic acid (MPAA), (13) 37 °C) resulted in the concomitant formation of the target SUMO protein and of a -18 uma contaminant (~ 20% by MALDI-TOF), which could not be separated by HPLC. Interestingly, lowering the temperature of the ligation mixture to 25 °C enabled greatly minimizing the side-reaction and the isolation of the target proteins in good yield and purity. (14,15)

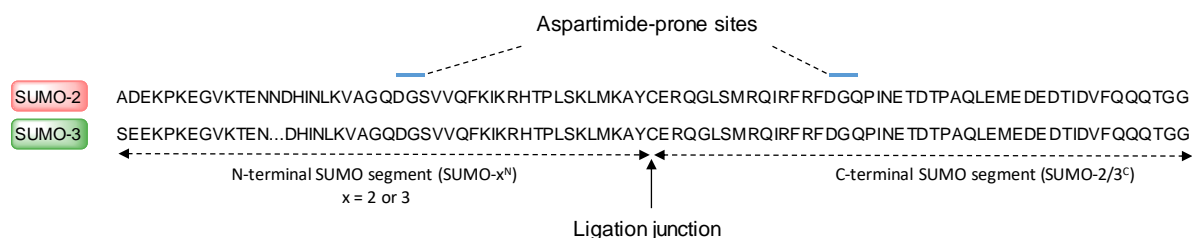


Figure 1. Sequences of SUMO-2 and 3 proteins showing the presence of an Asp-Gly dipeptide unit on both sides of the ligation site.

This protocol details the chemical synthesis of SUMO-2 and 3 proteins by SEA-mediated ligation at 25 °C in the presence of MPAA. The presence in SUMO proteins of a Cys residue

in central position enabled their assembly by ligating two peptide segments of about 45 amino acids (Figure 1). The N-terminal peptide segments, i.e., SUMO- x^N ($x = 2$ or 3) were produced by 9-fluorenylmethoxycarbonyl (Fmoc) SPPS using bis(2-sulfanylethyl)amino (SEA) ChemMatrix® resin (Fig. 2A).^(16,17) The C-terminal peptide segment SUMO-2/3^C is the same for both proteins. It was produced by Fmoc SPPS starting from 4-hydroxymethylphenoxyacetyl (HMPA) ChemMatrix® resin (Fig. 2B).

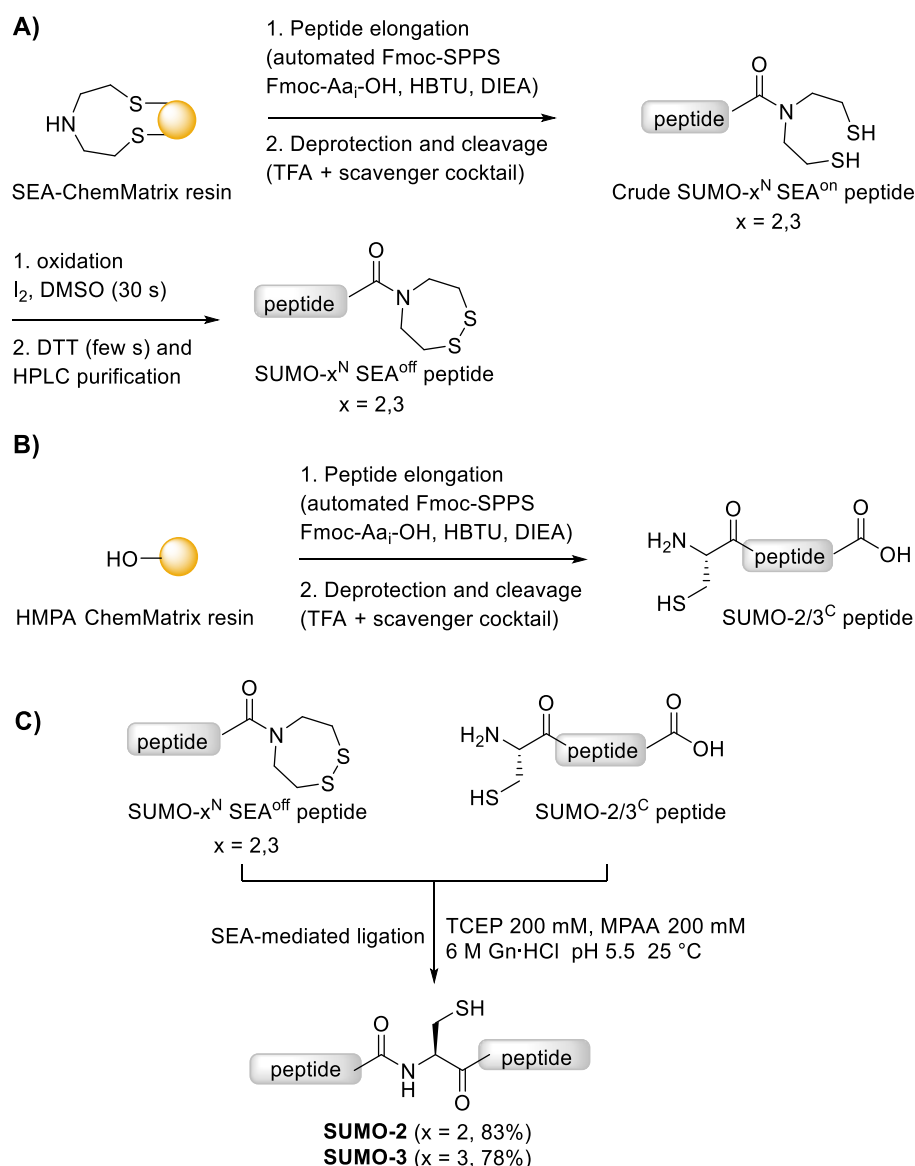


Figure 2. Synthesis of SUMO-2 and 3 proteins by SEA-mediated ligation. A) SPPS of the N-terminal segments using SEA-ChemMatrix® resin. B) SPPS of the C-terminal segment using 4-hydroxymethylphenoxyacetyl (HMPA) ChemMatrix® resin. C) Assembly of SUMO-2 and 3 proteins by SEA-mediated ligation at 25 °C.

2. Materials & Instruments

All organic solvents and chemicals used in this protocol should be handled inside a chemical fume hood with appropriate personal protective equipment (lab coat, gloves and protective glasses). TFA is strongly corrosive and toxic.

2.1 General

1. Synthesis of *bis*(2-sulfanylethyl)aminotriyl ChemMatrix (SEA ChemMatrix) resin was carried out as described elsewhere. **(16,17)**
2. HMPA ChemMatrix resin (IRIS Biotech, Lot 08K11-20-03-043).
3. Dichloromethane RPE ACS stabilised with amylene (DCM; Carlo Erba, Cat. No. 463314).
4. *N,N*-Dimethylformamide for peptide synthesis (DMF; Carlo Erba, Cat. No. P0343521).
5. 1-Methyl-2-pyrrolidinone (NMP; Carlo Erba, Cat. No. P0873521)
6. Resins were conditioned in DCM (3 x 2 min, 3 mL) and then in DMF (3 x 2 min, 3 mL) in a manual SPPS glass reactor prior to their use.
7. *N* α -Fmoc protected amino acids were obtained from Iris Biotech GmbH: Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(O*t*Bu)-OH, Fmoc-Asp(O*t*Bu)[Dmb-Gly]-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Val-OH, Fmoc-Cys(*S**t*Bu)-OH or Fmoc-Cys(Trt)-OH.
8. *N*-[(dimethylamino)-1*H*-1,2,3-triazolo-[4,5-*b*]pyridin-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide (HATU; Novabiochem, Cat. No. 01-62-0041).
9. *N*-[(1*H*-benzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide (HBTU; IRIS Biotech, Cat. No. RL-1030).
10. *N,N*-Diisopropylethylamine (DIEA; SDS, Cat. No. 0403516).
11. Acetaldehyde (Aldrich, Cat. No. 110078), 2 vol-% in DMF.
12. *p*-Chloranil (Acros, Cat. No. 213561000), 2 wt-% in DMF.
13. 2,4,6-Trinitrobenzenesulfonic acid (TNBS, TIC Europe N.V, Cat. No. 2508-19-2)
14. *N,N*-diisopropylcarbodiimide (DIC, Acros, Cat. No. 446181000)
15. 4-(dimethylamino)pyridine (DMAP, Aldrich chemie, Cat. No. 10 770-0)
16. Acetic anhydride (Ac₂O) was purchased from Biosolve and Fisher-Chemical.
17. UV spectrophotometer
18. Precision quartz cell (length 10 mm, quartz suprasil, Hellma, Cat. No. 117104-05)
19. Diethyl ether RPE stabilised with 2,6-*bis*(1,1-dimethylethyl)-4-methylphenol (Carlo Erba, Cat. No. 447523)

20. Piperidine for peptide synthesis (Biosolv, Cat. No. 16183302).
21. Thioanisole (Fluka, Cat. No. 88470).
22. Thiophenol (Alfa Aesar, Cat. No.10201873)
23. Trifluoroacetic acid (TFA; Biosolv, Cat. No. 20233301).
24. Triisopropylsilane (TIS; Aldrich, Cat. No. 23378-1).
25. 1,2-Ethanedithiol, 99% (EDT, Aldrich, Cat. No. E360-0).
26. *n*-Heptane (Biosolv, Cat. No. 0805202).
27. Water was purified with a Milli-Q Ultra Pure Water Purification System.
28. Dithiothreitol (DTT; Acros Organics, Cat. No. 16568-0050).
29. Iodine (Acros Organics, Cat. No. 38705100).
30. 0.1% TFA in water, Optima LC/MS (Fisher Chemical, Cat. No. LS119-212).
31. 0.1% TFA in acetonitrile, Optima LC/MS (Fisher Chemical, Cat. No. LS119-212).
32. Sodium dihydrogen phosphate dihydrate ($\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, Prolabo, Cat. No. 28015294).
33. Disodium hydrogen phosphate dihydrate ($\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, Prolabo, Cat. No. 28015294).
34. Guanidine hydrochloride ($\text{Gn} \cdot \text{HCl}$; Aldrich, Cat. No. G4505).
35. Glacial acetic acid (AcOH) (Carbo Erba, Cat. No. 401422).
36. Nitrogen gas ($\text{O}_2 < 3 \text{ ppm}$, Air Liquide).
37. Argon (Alphagaz Airliquide).
38. Hydrochloric acid (HCl ; Sigma Aldrich, Cat. No. 32-0331-500 mL).
39. Sodium hydroxide (NaOH ; VWR, Cat. No. 28240292).
40. *Tris*(2-carboxyethyl)phosphine hydrochloride ($\text{TCEP} \cdot \text{HCl}$; Aldrich, Cat. No. C4706-10G).
41. 4-mercaptophenylacetic acid (MPAA; Alfa Aesar, Cat. No. H27658).
42. Conical tubes (15 mL, 50 mL).
43. Thermomixer comfort (Eppendorf) equipped with a block heater (25 °C)
44. A set of adjustable pipettes (0.5–10, 2–20, 10–50, 10–100, 100–1,000 μL , Eppendorf)
45. Pipette tips, 0.5–10, 2–20, 20–200, 100–1,000 μL (VWR).
46. pH meter
47. Vortex shaker.
48. Microfuge tubes (1.5 mL safe-lock tubes; Eppendorf).

2.2 HPLC analysis & purification

1. TFA 10% (vol/vol) in water. Add carefully 10 mL of TFA to 90 mL of deionized water.
2. Eluent A for semi-preparative HPLC. TFA 0.1% (vol/vol) in deionized water. Add carefully 10 mL of 10% aqueous TFA (vol/vol) to 990 mL of deionized water.

3. Eluent B for semi-preparative HPLC. HPLC 80% (vol/vol) containing 0.1% (vol/vol) TFA. Add 10 mL of 10% aqueous TFA (vol/vol) to 990 mL of ACN 80% (vol/vol) in deionized water.
4. Analytical UPLC-MS system: Dionex UltiMate 3000/ LCQ Fleet Ion Trap, Dionex DA detector (215-280 nm), Corona Veo charged aerosol detector, ES+, m/z range 300–2000, capillary voltage 2.8 kV, cone voltage 10 V, tube lens 75 V, capillary voltage temperature 350 °C.
5. Analytical UPLC-MS column: Acquity UPLC peptide BEH300 C18 reverse-phase column (2.1 × 100 mm; pore size 300 Å; particle size: 1.7 µm).
6. Semi-preparative HPLC system (Waters 600 controller, UV 2487 Detector, 215 nm, TL 105 HPLC column heater).
7. Semi-preparative HPLC column: Waters XBridge BEH300 C18 reverse-phase column (10 × 250 mm; pore size 300 Å; particle size: 5 µm).

2.3 Lyophilization

1. Lyophilizer (Christ Gamma 2-20) equipped with a manifold.
2. Lyophilizer flasks (VWR, 300 mL, Cat. No. 88516; 600 mL, Cat. No. 88517).
3. Liquid nitrogen
4. Dewar

2.4 MALDI-TOF analysis

1. MALDI-TOF mass spectrometer (Autoflex Speed, Bruker).
2. 2,5-dihydrobenzoic acid (DHB, Aldrich, Cat. No. 490799).
3. A solution of DHB 20 mg/mL in 50% aqueous acetonitrile containing 0.1% TFA.

3. Solid phase peptide synthesis of the peptide segments

1. Column peptide synthesizer (MultiPep Intavis AG Bioanalytical instrument equipped with a 6-column module)
2. Synthesis columns (Intavis, Cat. No. 35760)
3. Reaction columns (Intavis, Cat. No. 34274)
4. Manual peptide synthesis vessels
5. Fmoc amino acid solutions, Fmoc-AA-OH 0.54 M in DMF.
6. DIEA solution, DIEA 35% in NMP (vol/vol).
7. HBTU solution: HBTU 0.54 M in DMF.
8. Piperidine solution 20% in DMF (vol/vol), see **Note 2**.
9. Capping solution, acetic anhydride 10%/DIEA 5%/DMF 85% (vol/vol/vol), see **Note 3**.

3.1 Coupling of first amino acid to the SEA-Trt ChemMatrix® resin

1. Weigh the amino acid Fmoc-Tyr(OtBu)-OH (459 mg, 1.00 mmol, 10 equiv) in a plastic tube.
2. Weigh 377 mg (0.95 mmol, 9.5 equiv) of HATU in a plastic tube, see **Note 4**.
3. Add 5 mL of DMF to the tubes and agitate with a vortex shaker to dissolve the reagents.
4. Transfer the HATU/DMF solution to the tube containing the amino acid solution with an adjustable pipette.
5. Add 348 μ L (2.00 mmol, 20 equiv) of DIEA to the tube with an adjustable pipette and agitate with a vortex shaker for 1 min. The solution becomes yellow upon addition of the base.
6. Transfer the activated amino acid/DMF solution to the reaction column.
7. Agitate with a laboratory flask shaker for 1 h 30 min and drain the column.
8. Aspirate 5-10 mL of DMF, agitate for 2 min and then drain the column.
9. Repeat Step 8 twice.
10. Take a sample of the resin and perform the chloranil test which must be negative (colorless resin beads).
11. Acetylate remaining amino groups eventually present on the Fmoc-Tyr(OtBu)-SEA-Trt ChemMatrix resin by adding 5 mL of acetic anhydride/DIEA/DMF capping solution to the resin. Shake for 10 min and drain, see **Note 5**.
12. Repeat Step 11 twice.
13. Wash the resin with DMF (3 x 2 min, 5 mL).
14. Wash the resin with DCM (3 x 2 min, 5 mL).
15. Aspirate 5 mL of diethyl ether, shake for 2 min and then drain the column.
16. Repeat Step 15 twice.
17. Determine the loading of the Fmoc-Tyr(OtBu)-SEA Trt ChemMatrix resin by treating the resin with the 20% piperidine solution in DMF and UV quantification of the formed dibenzofulvene-piperidine adduct at 290 nm as described elsewhere. **(16)** Typical loading for Fmoc-Tyr(OtBu)-SEA-Trt ChemMatrix resin: 0.22 mmol/g.

3.2 Coupling of first amino acid to the HMPA ChemMatrix® resin

1. Weigh HMPA ChemMatrix® resin (500 mg).
2. Weigh the amino acid Fmoc-Gly-OH (297 mg, 1.00 mmol) in a plastic tube.
3. Weigh 12.2 mg (0.100 mmol) of DMAP.
4. Add 5 mL of DMF to the tube containing the amino acid to dissolve it.
5. Add 78.3 μ L (0.500 mmol) of DIC to the tube containing the amino acid solution with an adjustable pipette.
6. Transfer the activated amino acid/DIC solution to the reaction column.
7. Add DMAP in one portion to the reaction column.

8. Agitate with a laboratory flask shaker for 2 h at room temperature.
9. Aspirate 5-10 mL of DMF, agitate for 2 min and drain the column.
10. Repeat Step 9 twice.
11. Wash the resin with DMF (3 x 2 min, 5 mL).
12. Wash the resin with DCM (3 x 2 min, 5 mL).
13. Add 5 mL of diethyl ether, shake for 2 min and drain the column.
14. Repeat Step 13 twice
15. Determine the loading of the Fmoc-Gly-HMPA ChemMatrix® resin after treating an aliquot with 20% piperidine solution in DMF and reading the absorbance of the supernatant at 290 nm. Typical loading for Fmoc-Gly-HMPA ChemMatrix® resin: 0.58 mmol/g.

3.3 First elongation using the automated peptide synthesizer

Elongate Fmoc-Tyr(OtBu)-SEA-Trt ChemMatrix resin and Fmoc-Gly-HMPA ChemMatrix® resin using the automated peptide synthesizer, see **Note 6**.

Elongated sequence using Fmoc-Tyr(OtBu)-SEA-Trt ChemMatrix resin (SUMO-2^N and SUMO-3^N): SVVQFKIKRHTPLSKLMKA

Elongated sequence using Fmoc-Gly-HMPA ChemMatrix® resin (SUMO-2/3^C): QPINETDTPAQLEMEDEDTIDVFQQQTG

3.4 Manual coupling of Fmoc-Asp(OtBu)[Dmb]Gly-OH

1. Weigh Fmoc-Asp(OtBu)[Dmb]Gly-OH (186 mg, 0.300 mmol, 3 equiv) in a plastic tube.
2. Weigh 108 mg of HATU (108 mg, 0.284 mmol, 2.85 equiv) in the plastic tube.
3. Add 3 mL of DMF to the tube containing HATU and agitate with a vortex shaker to dissolve the reagent.
4. Transfer the HATU/DMF solution to the tube containing the amino acid solution with an adjustable pipette.
5. Add 104 µL of DIEA (0.600 mmol, 6 equiv) to the tube with an adjustable pipette and agitate with a vortex shaker for 1 min. The solution becomes yellow upon addition of the base.
6. Transfer the activated amino acid/DMF solution to the reaction column.
7. Agitate with a laboratory flask shaker for 2 h and then drain the column.
8. Aspirate 5-10 mL of DMF, agitate for 2 min and drain the column.
9. Repeat Step 8 twice
10. Take a sample of the resin and perform a TNBS test that must be negative (colorless resin beads), see **Note 7**.
11. Optional: analysis of the elongated peptide by MALDI-TOF after cleavage and deprotection, see Steps 12-14.

12. Treat an aliquot of the peptidyl resin with 200 μ L of 20% piperidine to remove the N-terminal Fmoc group.
13. Remove the piperidine solution with 1 mL of diethyl ether using a Pasteur pipette and dry the resin under a gentle flow of argon.
14. Perform the TFA cleavage for 1 h by adding 70 μ L of the cleavage mixture to the resin aliquot: TFA/triisopropylsilane (TIS)/thioanisole/H₂O/thiophenol: 87.5/5/2.5/2.5/2.5 by vol (1 mL) for SEA ChemMatrix[®] resin, or TFA/TIS/H₂O/ethanedithiol (EDT): 90/5/2.5/2.5 by vol (1 mL) for HMPA-ChemMatrix[®] resin.
Take 1 μ L of the cleavage mixture, dilute with water (50 μ L) and analyze by MALDI-TOF to confirm the successful elongation of the peptide chain including the coupling of the dipeptide unit. Elongated sequence using Fmoc-Tyr(OtBu)-SEA-Trt ChemMatrix resin (SUMO-2^N and SUMO-3^N): DGSVVQFKIKRHTPLSKLMKAY [M+H]⁺ calc. 2663.45, found 2664.86.
Elongated sequence using Fmoc-Gly-HMPA ChemMatrix[®] resin (SUMO-2/3^C): DGQPINETDTPAQLEMEDEDIDVFQQQTGG [M+H]⁺ calc. 3421.48, found 3422.10.
See **Note 8**.
15. Continue the synthesis by acetylating unreacted amino groups eventually present on the HMPA or SEA ChemMatrix[®] peptidyl resins by adding 5 mL of acetic anhydride/DIEA/DMF capping solution to the solid support. Shake for 10 min and drain
16. Repeat Step 15 twice
17. Add 5 mL of DMF to the column shake for 2 min and drain the column.
18. Repeat Step 17 twice

3.5 Second elongation using automated peptide synthesizer

Elongate the peptidyl resins using the automated peptide synthesizer.

Elongated sequence for SEA-Trt ChemMatrix resin: ADEKPKEGVKTENNDHINLKVAGQ for SUMO-2^N, SEEKPKEGVKTENDHINLKVAGQ for SUMO-3^N.

Elongated sequence for HMPA ChemMatrix[®] resin: CERQGLSMRQIRFRF for SUMO-2/3^C.

3.6 Cleavage step

1. In a graduated cylinder add 8.75 mL of TFA, 0.25 mL of deionized water, , 0.25 mL of thiophenol, 0.25 mL of thioanisole and 0.50 mL of TIS for peptides SUMO-x^N (x = 2, 3).
2. In a graduated cylinder add 9 mL of TFA, 0.25 mL of deionized water, 0.25 mL of EDT and 0.50 mL of TIS for peptide SUMO-2/3^C.
3. Transfer each peptidyl resin in a different manual peptide synthesis vessel.

4. Add the TFA solution to the manual peptide synthesis vessel and agitate at room temperature for 2 h
5. In the meantime, prepare three 300 mL Erlenmeyer flasks, each containing a mixture of 150 mL of diethyl ether and 150 mL of *n*-heptane. Add magnetic bars and place the Erlenmeyer flasks in an ice bath.
6. Precipitate the peptides by adding the peptide solutions dropwise to the ice-cold diethyl ether/*n*-heptane mixture with magnetic stirring.
7. Add 5 mL of TFA to each peptide reactor, shake the bead suspensions for 2 min and then add this solution dropwise to the Erlenmeyer flask.
8. Repeat Step 7, see **Note 9**.
9. Transfer the suspensions to centrifuge tubes and centrifuge at 1,900 × *g* for 10 min, then pour carefully the supernatant into an Erlenmeyer flask.
10. Add 40 mL of cold diethyl ether 50%/n-heptane 50% (vol/vol) to the peptide, triturate the peptide pellet and centrifuge again at 1,900 × *g* for 10 min, then pour carefully the supernatant into the Erlenmeyer flask.
11. Carefully flush the tubes with nitrogen to evaporate the residual solvent.
12. Dissolve the crude peptides in TFA 0.1% (vol/vol) in deionized water and place the peptide solutions immediately at -20°C overnight, see **Note 10**.
13. Lyophilize for 2 days.
14. Analyze the crude peptide by UPLC-MS and MALDI-TOF mass spectrometry

3.7 Oxidation SEA^{on} peptides. Purification of SUMO-x^N SEA^{off} peptides (x = 2,3)

1. Prepare the semi-preparative HPLC system, which should be ready before starting the oxidation of the peptides.
2. Rapidly weigh 100-150 mg of iodine in a 5 mL volumetric flask. Immediately add the DMSO up to 5 mL. Calculate the concentration of iodine, see **Note 11**.
3. Prepare a solution of DTT 10 mg/mL in deionised water.
4. Dissolve the crude peptide in phosphate buffer/Gn-HCl 6 M at room temperature (peptide concentration ~ 0.25 mM), see **Note 12**.
5. Add 2 equivalents of iodine in DMSO to the peptide solution. The solution becomes yellow indicating that an excess of iodine has been added to the peptide solution.
6. Vortex the solution for 30 s, see **Note 13**.
7. Then immediately add a few drops of the DTT solution with a Pasteur pipette until the yellow color disappears, indicating that the excess of iodine has been decomposed.

3.8 HPLC purification of SUMO-2/3^C and SUMO-2/3^N peptides.

1. Purify immediately by semi preparative HPLC.

2. Gradient used for the HPLC purification of peptide **SUMO-2^N**: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.
3. Yield for peptide **SUMO-2^N**: 34.4 mg of crude product furnished (11.2 mg) of **SUMO-2^N**, Calculated for M (average mass) 5281.17, observed 5281.13 by LC-MS after deconvolution.
4. Gradient used for the HPLC purification of peptide **SUMO-3^N**: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-25% B in 5 min, then 25-35% B in 40 min, flow rate 6 mL/min, 50 °C, UV detection at 215 nm.
5. Yield for peptide **SUMO-3^N**: 30.8 mg of crude product furnished (11.2 mg) of **SUMO-3^N**, Calculated for M (average mass) 5197.09, observed 5197.17 by LC-MS after deconvolution.
6. Gradient used for the HPLC purification of peptide **SUMO-2/3^C**: eluent A water containing 0.1% of TFA, eluent B CH₃CN/water : 4/1 by vol containing 0.1% of TFA, gradient: 0-20% B in 5 min, then 20-40% B in 40 min, flow rate 6 mL/min, 50°C, UV detection at 215 nm
7. Yield for peptide **SUMO-2/3^C**: 31.3 mg of crude product furnished (10.6 mg) of **SUMO-2/3^C**, Calculated for M (average mass) 5331.82, observed 5331.95 by LC-MS after deconvolution.
8. Analyze the collected fractions by analytical UPLC-MS and pool the pure fractions in a 50 mL plastic tube.
9. Place the tubes in liquid nitrogen to freeze the peptide solutions and lyophilize for two or three days to obtain the purified **SUMO-2^N**, **SUMO-3^N**, **SUMO-2/3^C** peptides segments, see **Note 10**.

4. Synthesis of SUMO-2 and 3

1. Weigh 5.38 mg (1 equiv, 0.825 μmol, 4 mM) of SUMO-x^N SEA^{off} peptide (x = 2 or 3) in a 1.5-mL microfuge tube.
2. Weigh 4.94 mg (1 equiv, 0.825 μmol, 4 mM) of cysteinyl peptide SUMO-2/3^C in a 1.5-mL microfuge tube.
3. Weigh 13.42 mg (79.77 μmol, 200 mM final concentration) of MPAA in a 1.5-mL microfuge tube.
4. Weigh 31.6 mg (110 μmol, 200 mM final concentration) of TCEP in a 1.5-mL microfuge tube.
5. Prepare 2 mL of 0.1 M sodium phosphate buffer (pH 7.2) / 6 M Gn-HCl.

6. Transfer 1 mL of AcOH glacial, 1 mL of HCl 1 N and 1 mL of NaOH 6 N in three different 1.5-mL microfuge tubes.
7. Cap all the tubes and transfer them to the glove box under nitrogen atmosphere.
8. Add 0.551 mL of the sodium phosphate buffer from Step 5 to the microfuge tube containing TCEP (see Step 4). Vortex until dissolution of the TCEP and transfer 0.399 mL of this solution to the microfuge tube containing MPAA (see Step 3). Add NaOH 6 N (30 μ L) with the micropipette and vortex until dissolution of the MPAA. The final reagent concentration is 181 mM for MPAA and 181 mM for TCEP.
9. Measure the pH of the solution prepared in Step 8 and adjust to pH 5.5 by adding NaOH 6 N or HCl 1 N with the micropipette (2-20 μ L).
10. Add 0.205 mL of the solution prepared in Step 8 to the microfuge tube containing the SUMO-x^N SEA^{off} peptide (see Step 1). Vortex to dissolve the peptide and then add this solution to the microfuge tube containing the cysteinyl peptide SUMO-2/3^C (see Step 2). Vortex to dissolve the peptide.
11. Place the reaction tube from Step 10 in a thermomixer kept at 25 °C with agitation.
12. To monitor the ligation reaction (just after preparing the reaction mixture at Step 10 and every 6-7 hour until ~33 h of reaction, transfer 1.5 μ L of the reaction mixture to a microfuge tube and quench the reaction by acidifying the sample with 100 μ L of 10% aqueous AcOH (vol/vol). Vortex and remove the sample from the glove box.
13. Remove MPAA present in the reaction mixture by extracting the aqueous phase with 1 mL of diethyl ether. Remove the diethyl ether phase with a Pasteur pipette.
14. Repeat Step 13 twice.
15. Inject 10 μ L of the sample isolated in Step 14 into the analytical UPLC-MS system to confirm that the starting peptide segments have been consumed to produce the ligation product SUMO-2 or 3.
16. After 33 h, remove the reaction tube from the glove box.
17. Transfer the contents of the reaction tube into a 5-mL microfuge tube.
18. Add 0.500 mL of phosphate buffer / 6 M Gn·HCl to the reaction mixture and 70.5 μ L of glacial acetic acid to a final concentration of 5-10 % (vol/vol) with a micropipette to acidify the solution down to a pH of 2-3 (pH paper).
19. Extract the MPAA from the aqueous phase with 2 mL of diethyl ether as done in Step 13. Remove the diethyl ether phase with a Pasteur pipette.
20. Repeat Step 19 twice.
21. Remove the dissolved diethyl ether by flushing the tube with nitrogen (~2 min)
22. Purify the target SUMO-2 or 3 proteins using a Waters XBridge BEH300 C18 reverse-phase column (10 x 250 mm; pore size 300 Å; particle size : 5 μ m) and a semi-preparative HPLC system; Gradient : 0 % eluent B in eluent A to 20 % (vol/vol) eluent

- B in eluent A over 5 min, and then 20 % (vol/vol) eluent B in eluent A to 40 % (vol/vol) eluent B in eluent A over 60 min (UV detection at 215 nm). Analyze the fractions using the MALDI-TOF spectrometer. Mix 1.2 μL of the DHB solution with 1 μL of the sample on the MALDI plate and let dry at room temperature in air.
23. Pool the pure fractions in a 50 mL plastic tube, freeze the solution with liquid nitrogen and lyophilize it for two or three days. The reaction yields 4.20 mg (78% yield) of SUMO-2 and 4.46 mg of SUMO-3 (83% yield).
 24. Analyze the purified products by UPLC-MS. ACQUITY UPLC peptide BEH C18 300 \AA 1.7 μm 2.1 mm \times 150 mm, 50 $^{\circ}\text{C}$. Flow 0.400 mL/min, eluent A 0.1% trifluoroacetic acid in water, eluent B 0.1% trifluoroacetic acid in 100% acetonitrile. Gradient from 0% eluent B to 70% eluent B in 20 min.
 25. SUMO-2: Calculated for M (average mass) 10477.24, observed 10476.71 after deconvolution
 26. SUMO-3: Calculated for M (average mass) 10393.66, observed 10393.84 after deconvolution.

Notes

Note 1. Aspartimide byproduct formation can occur during NCL.**(18)** For example, the assembly of Ras protein (166 amino acid residues) by NCL yielded a product having 18 ± 2 uma less than expected.**(19)** The mass loss occurred during the last ligation and could be due to a loss of water or ammonia. The site of dehydration or deamination was not identified. For another example, see ref **(20)**.

Note 2. The solution can be stored or used for up to 2 weeks at room temperature.

Note 3. This solution can be used up to 24 h after preparation. For the automated peptide elongation step, this solution is renewed every day.

Note 4. HATU is an uronium-based coupling reagent with reacts efficiently with amines and thus can cause undesired peptide termination if present in excess. To avoid the presence of unreacted HATU during the coupling step, the amino acid is used in slight molar excess and the amino acid is activated prior to the coupling step.

Note 5. The capping step is mandatory even if the chloranil test is negative, because the sensitivity of the test for the secondary amino group of the SEA resin is lower than for α -amino groups.

Note 6. The synthesis of the ~ 50 amino acids SEA^{off} peptide segments **SUMO-x^N** is performed using the SEA-Trt-ChemMatrix[®] resin. The synthesis program must consider the large swelling volume of the starting SEA-Trt resin and the significant mass increase of the peptidyl resin that occurs during the peptide elongation step.

Note 7. We recommend to confirm the coupling of Fmoc-Asp(OtBu)[Dmb]Gly-OH by cleaving an aliquot in TFA and analyzing the cleaved peptide by MALDI-TOF.

Note 8. MALDI-TOF analysis of the cleaved peptides might show the expected peaks and also products of higher mass (+ 150 uma) due to the incomplete removal of the Dmb group.

Note 9. We recommend that the volume of the diethyl ether/*n*-heptane solution be at least 20 times that of the peptide solution in TFA.

Note 10. The peptide solution can be frozen quickly using liquid nitrogen. The tube must be checked before starting the lyophilization because it can crack during this step.

Note 11. Iodine is available as granules. Therefore, weighting a given amount of iodine precisely is difficult to achieve.

Note 12. SUMO-x^N SEA^{off} peptides (x = 2,3) have a limited solubility and the use of Gn·HCl, which is a classical additive for improving the solubility of the peptide segments, was mandatory in this case. Iodine oxidation can be performed in the absence of Gn·HCl and at mM concentrations when solubility is not an issue.

Note 13. The duration of the iodine treatment should not exceed 30 s, otherwise iodine can react with some amino acid residues such as tyrosines. For the same reason, the amount of iodine which is added to the peptide solution must not exceed two equivalents. This oxidation procedure is not compatible with the presence of free cysteine thiols on the peptide chain.

Acknowledgements

This study was supported by the National Center for Scientific Research (CNRS), the University of Lille and the Pasteur Institute of Lille.

References

- (1) Merrifield RB Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* 1963;85:2149-54.
- (2) Behrendt R, White P, Offer J Advances in Fmoc solid-phase peptide synthesis. *J Pept Sci* 2016;22:4-27.
- (3) Voorter CE, de Haard-Hoekman WA, van den Oetelaar PJ, Bloemendal H, de Jong WW Spontaneous peptide bond cleavage in aging alpha-crystallin through a succinimide intermediate. *J Biol Chem* 1988;263:19020-3.
- (4) Joshi AB, Sawai M, Kearney WR, Kirsch LE Studies on the mechanism of aspartic acid cleavage and glutamine deamidation in the acidic degradation of glucagon. *J Pharm Sci* 2005;94:1912-27.
- (5) Geiger T, Clarke S Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J Biol Chem* 1987;262:785-94.
- (6) Stephenson RC, Clarke S Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J Biol Chem* 1989;264:6164-70.

- (7) Tyler-Cross R, Schirch V Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. *J Biol Chem* 1991;266:22549-56.
- (8) Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 2007;23:1282-8.
- (9) Agouridas V, El Mahdi O, Diemer V, Cargoet M, Monbaliu J-CM, Melnyk O Native chemical ligation and extended methods. Mechanisms, catalysis, scope and limitations. *Chem Rev* 2019;12:7328-443.
- (10) Wucherpennig TG, Pattabiraman VR, Limberg FRP, Ruiz-Rodríguez J, Bode JW Traceless preparation of C-terminal α -ketoacids for chemical protein synthesis by α -ketoacid-hydroxylamine ligation: Synthesis of SUMO2/3. *Angew Chem Int Ed* 2014;53:12248-52.
- (11) Mulder MPC, Merckx R, Witting KF, Hameed DS, El Atmioui D, Lelieveld L, Liebelt F, Neefjes J, Berlin I, Vertegaal ACO et al. Total chemical synthesis of SUMO and SUMO-based probes for profiling the activity of SUMO-specific proteases. *Angew Chem Int Ed* 2018;57:8958-62.
- (12) Ollivier N, Dheur J, Mhidia R, Blanpain A, Melnyk O *Bis*(2-sulfanylethyl)amino native peptide ligation. *Org Lett* 2010;12:5238-41.
- (13) Johnson EC, Kent SBH Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 2006;128:6640-6.
- (14) Bouchenna J, Sénéchal M, Drobecq H, Vicogne J, Melnyk O Total Chemical Synthesis of All SUMO-2/3 Dimer Combinations. *Bioconjugate Chem* 2019;30:2967-73.
- (15) Bouchenna J, Sénéchal M, Drobecq D, Stankovic-Valentin N, Vicogne J, Melnyk O The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis. *Bioconjugate Chem* 2019;30:2684-96.
- (16) Boll E, Drobecq H, Ollivier N, Blanpain A, Raibaut L, Desmet R, Vicogne J, Melnyk O One-pot chemical synthesis of small ubiquitin-like modifier (SUMO) protein-peptide conjugates using *bis*(2-sulfanylethyl)amido peptide latent thioester surrogates *Nat Protoc* 2015;10:269-92.
- (17) Boll E, Drobecq H, Ollivier N, Raibaut L, Desmet R, Vicogne J, Melnyk O A novel PEG-based solid support enables the synthesis of >50 amino-acid peptide thioesters and the total synthesis of a functional SUMO-1 peptide conjugate. *Chem Sci* 2014;5:2017-22.
- (18) Dawson PE, Muir TW, Clark-Lewis I, Kent SBH Synthesis of proteins by native chemical ligation. *Science* 1994;266:776-79.
- (19) Becker CF, Hunter CL, Seidel R, Kent SBH, Goody RS, Engelhard M Total chemical synthesis of a functional interacting protein pair: the protooncogene H-Ras and the Ras-binding domain of its effector c-Raf1. *Proc Natl Acad Sci U S A* 2003;100:5075-80.
- (20) Cowper B, Shariff L, Chen W, Gibson SM, Di W-L, Macmillan D Expanding the scope of N \rightarrow S acyl transfer in native peptide sequences. *Org Biomol Chem* 2015;13:7469-76.