Reports

BioTechniques

Efficient and robust preparation of tyrosine phosphorylated intrinsically disordered proteins

Pavel Brázda[1,2], Ondrej Šedo[1], Karel Kubíček[1] & Richard Štefl[*,1,2]

ABSTRACT

Intrinsically disordered proteins (IDPs) are subject to post-translational modifications. This allows the same polypeptide to undertake different interaction networks with different consequences, ranging from regulatory signalling networks to formation of membraneless organelles. We report a robust method for co-expression of modification enzyme and SUMO-tagged IDP with subsequent purification procedure allowing production of modified IDP. The robustness of our protocol is demonstrated on a challenging system, RNA polymerase II C-terminal domain (CTD), that is a low-complexity repetitive region with multiple phosphorylation sites. *In vitro* phosphorylation approaches fail to yield multiple-site phosphorylated CTD, whereas our *in vivo* protocol allows to rapidly produce near homogeneous phosphorylated CTD at a low cost. These samples can be used in functional and structural studies.

METHOD SUMMARY

Production of phosphorylated intrinsically disordered proteins (IDPs) remains a challenge. We report a robust method for co-expression of kinase and SUMO-tagged IDP with subsequent purification procedure allowing production of modified IDP. Our *in vivo* protocol allows to rapidly produce highly phosphorylated IDPs at a low cost which can be used in functional and structural studies.

KEYWORDS

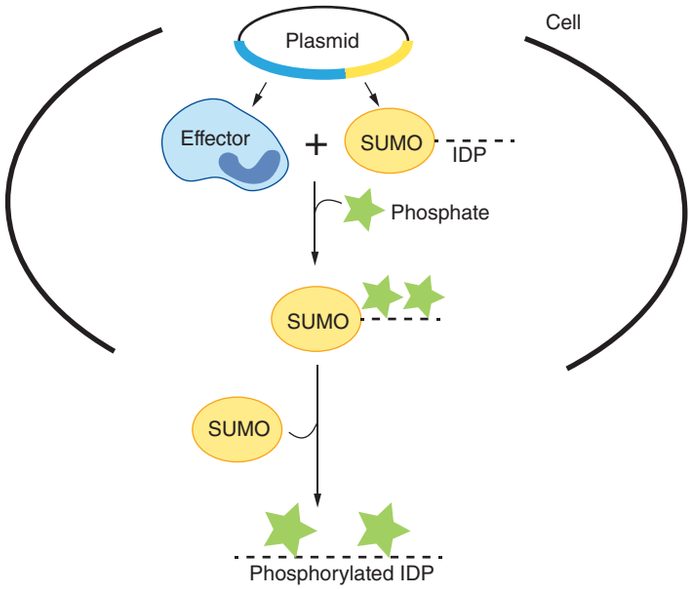
C-terminal domain • co-expression • CTD • IDP • intrinsically disordered proteins • phosphorylation • purification • RNA polymerase II

*[1]CEITEC–Central European Institute of Technology, Masaryk University, CZ-62500 Brno, Czech Republic; [2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, CZ-62500 Brno, Czech Republic; *Author for correspondence: richard.stefl@ceitec.muni.cz*

BioTechniques 67: 16-22 (July 2019) 10.2144/btn-2019-0033

GRAPHICAL ABSTRACT

Both effector kinase and SUMO-fused IDP target are recombinantly co-expressed. The phosphorylated peptide is subsequently purified using the affinity and ion-exchange chromatography that yields peptides with homogeneous phosphorylations.



Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are in the center of attention, as they not only occupy a significant portion of proteomes [1], but also violate 'one protein-one structure-one function' paradigm [2]. Thus, they do not adopt a well ordered structure [2], but rather an ensemble of dynamic structures interconverting on a number of timescales and some may undergo disorder to order transitions in the presence of natural ligands. IDPs and IDRs are subject to post-translational modifications (PTM), which allows the same polypeptide to undertake different interaction networks with different consequences, ranging from regulatory signalling networks to formation of membrane-less organelles through phase separation (e.g., liquid droplets). IDPs are involved in transcription, translation, splicing, cellular signalling and cell cycle processes [3–6]. Beside that IDPs are associated with numerous human diseases including cancer, cardiovascular diseases, amyloidoses, neurodegenerative diseases, and diabetes [7]. Therefore, unravelling structural and functional information about these proteins is of biomedical importance as they represent a novel class of drug targets that aims to modulate protein–protein interactions.

A number of strategies have been developed to aid the production of IDPs [8], whereas the production of modified (e.g., phosphorylated IDPs) still remains a challenge. Current protocols involve several ways how to produce post-translationally modified protein or peptide. The first option is to subject a pre-purified protein to enzymatic assay, which will introduce the desired modification. For example, Portz and colleagues [9] phosphorylated the purified CTD by the P-TEFb. The final product must be re-purified after *in vitro* phosphorylation and the reactions require a great deal of condition optimization and even then ▶

the enzymatic reaction is rarely driven to completion [10,11]. Alternatively, a solid-phase synthesized fragment with PTM can be fused to a purified protein. For example, a lipidated solid-phase synthesized peptide was conjugated via chemically selective ligation to Ras protein expressed in standard *E. coli* host [12]. This method is limited to conjugation with terminal residues [13–15].

Serine phosphorylation can be studied via engineering a residue mimicking the phosphomark (e.g. glutamate). The phosphoserine–glutamate substitution is imperfect due to a different atomic size (P is bigger than C), number of oxygen atoms, and overall charge at neutral pH. Nevertheless, several studies show a good approximation between modified and phosphorylated wild-type protein behaviour in biochemical assays [16–18]. The site-specific phosphorylation can also be introduced using nonsense codon suppression technique. The termination signal of amber codon is suppressed during translation and site-specific unnatural amino acid is placed in the growing polypeptide in 'suppression-friendly' *Escherichia coli* strains [19]. Production of modified protein both *in vivo* and *in vitro* suffer from a low yield and efficiency of suppression [20–22].

Last option is to co-express two distinct proteins in the same cell. This strategy relies on two independent plasmids using distinct antibiotic resistances and origins of replication. If the same antibiotic resistance is used, only one of the two constructs is maintained [23]. The propagation of both plasmids is ensured by using compatible replicons, where there is no competition between them [24]. Several studies employed this method to produce heterodimer complex [25], chaperon and its target [26] and kinase with its ligand [27]. The downside of this co-expression approach is requirement for optimization of vectors and their design, expression and purification for a particular set of proteins [28]. Additionally, maintenance of multiple plasmids each with its own resistance and protein puts *E. coli* cells under huge metabolic stress [29–31].

Here, we describe an alternative approach in which the modification enzyme and target are expressed via a single plasmid with the use of solubility enhancement tag. We utilised pETDuet-1 vector to co-express phosphotyrosine kinase c-Abl [32,33] with its target fused to SUMO and FLAG tags. The use of tags is important as they prevent proteolysis and aggregation of IDPs [34]. SUMO tag enhances the overall target solubility and expression and can be cleaved to remove the SUMO moiety using SUMO-specific proteases such as Ulp1 [35–39]. FLAG tag prevents proteolysis and offers additional detection and purification tool [34]. Thanks to these tags, we are able to produce short as well as long peptides that are fully phosphorylated. As a target for c-Abl, we chose very challenging model of IDP, the C-terminal domain (CTD) of the largest RNA polymerase II subunit, Rpb1. The CTD protrudes from RNA polymerase II core and plays an essential role in transcription initiation, elongation, termination, mRNA co-transcriptional processing, chromatin modification and DNA repair [40–42]. The CTD domain is highly repetitive and features characteristics of IDRs with a consensus repetitive motif of YSPTSPS [43–45]. Dynamic post-translational modifications of each non-proline residue enables CTD to act as a landing pad for recruitment, exchange or displacement of various transcription and processing factors [46–49].

MATERIAL & METHODS

Preparation of the constructs

The synthetic Abl kinase and target CTD genes were produced by GeneArt (Thermo Fisher Scientific). The $(CTD)_2$ and $(CTD)_{13}$ genes were N-terminally fused with 6xHis tag followed by SUMO solubility tag. The construct with 13 repetitions of CTD was also C-terminally FLAG tagged to promote its stability [18]. The kinase and target genes were inserted in pETDuet-1 vector using NcoI and XhoI. Therefore, the kinase is placed in the MCS1 (multiple cloning site), whereas the target gene in the MCS2. The $(CTD)_{13}$ construct was cloned using BamHI and XhoI restriction sites. All plasmids used along with gene sequences can be found in Table S1.

Protein expression

All constructs were transformed into BL21-CodonPlus (DE3)-RIL bacterial strain. The transformants were grown in LB or minimal M9 media ([U-^{13}C] D-glucose and $^{15}NH_4Cl$ for isotopically labelled samples) for $(CTD)_2$ and $(CTD)_{13}$, respectively and induced with 0.1 mM IPTG. Protein expression was carried out at 16°C. Cells were harvested, resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 20 mM BME, pH 7.3) and sonicated. We note that the expression temperature and time may initially be optimized to yield the largest amount of phosphorylated peptide. The biomass was sonicated on ice with 1s pulse, 4 s off with 40% amplitude for total sonication time 6.5 min using Q700 sonicator (Qsonica, LLC). The cell lysate was cleared by centrifugation and loaded on 5 ml HisTrap FF Crude column (GE Healthcare) connected to AKTA purifier 10 with UV-900 flow cell. The column was washed with 4 CV (column volumes) of lysis buffer and then buffer was exchanged to the cleavage buffer (100 mM NH_4HCO_3, pH 8). An on-column cleavage using Ulp1 protease (construct donated by Christopher D Lima) was performed [35,36]. The cleaved peptides were eluted with the cleavage buffer in 5 CV. Fractions containing the mix of peptides were pooled and 10x diluted with low-salt buffer (10 mM NH_4HCO_3, pH 8) and loaded onto 5 ml HiTrap Q HP (GE Healthcare) column. The column was washed with 4 CV of low-salt buffer. The gradient elution with high-salt buffer (1 M NH_4HCO_3, pH 8) was performed for 10 and 20 CV, at 2 ml/min with gradient of 12.5 and 100 min for the $(CTD)_2$ or $(CTD)_{13}$ constructs, respectively. The collected fractions were analysed by mass spectrometry. The peptides were then thoroughly lyophilized. The protocol yielded phosphorylated peptides with a 99% purity and the final yield of ~1 mg.l^{-1} of media.

Nuclear magnetic resonance experiments

The ~1.5 mM uniformly ^{15}N, ^{13}C-labelled $(pY1\text{-}CTD)_2$ peptide was measured in 20 mM Na_2HPO_4, pH 6 (90% H_2O/10% D_2O) at a temperature of 10°C. Phosphorus-detected experiments were performed on a 600 MHz Bruker AVANCE III HD spectrometer equipped with a QCI Cryoprobe. The 1D ^{31}P spectra were recorded with a standard Bruker pulse program, using a 30° excitation pulse, a 1.5 s recycle delay and a WALTZ-16 power-gated composite-pulse proton decoupling. Set of standard 2D experiments (^{1}H-^{15}N HSQC, aliphatic and aromatic ^{1}H-^{13}C HSQC) were recorded on a Bruker AVANCE 700 MHz spectrometer equipped with 5 mm TXI probe. The spectra were processed with Topspin 3.2 (Bruker BioSpin).

Mass spectrometric peptide characterization

The purified peptides were subjected to MALDI-TOF mass spectrometric analysis using an Ultraflextreme instrument (Bruker Daltonics, Bremen, Germany) equipped with a stainless-steel sample target. The CTD peptides were analysed in the reflectron negative ion detection mode using alpha-cyano-4-hydroxycinnamic acid (Bruker) as a MALDI matrix, while the $(pY1\text{-}CTD)_{13}$ was analysed in linear positive ion detection mode with ferulic acid (Sigma Aldrich) as a MALDI matrix. Calf Intestinal Alkaline Phosphatase kit (Invitrogen) was used for $(pY1\text{-}CTD)_{13}$ dephosphorylation according to guidelines of the manufacturer. The localization of phosphorylation sites and relative quantification of the peptide forms was done by LC-MS/MS using RSLCnano system (SRD-3400, NCS-3500RS CAP, WPS-3000 TPL RS) connected to Orbitrap Elite hybrid spectrometer (Thermo Fisher Scientific). Prior to LC separation, peptides were online concentrated and desalted using trapping column (100 μm × 30 mm) filled with 3.5-μm X-Bridge BEH 130 C18 sorbent (Waters). After washing of trapping column with 0.1% fluoroacetic acid, the peptides were eluted (flow 300 nl/min) from the trapping column onto an analytical column (Acclaim Pepmap100 C18, 3 μm particles, 75 μm × 500 mm; Thermo Fisher Scientific) by 50 min nonlinear gradient program (1–56% of mobile phase B; mobile phase A: 0.1% fluoroacetic acid in water; mobile phase B: 0.1% fluoroacetic acid in 80% acetonitrile). Equilibration of the trapping column and the column was done prior to sample injection to sample loop. The analytical column outlet was directly connected to the Digital PicoView 550 (New Objective) ion source with sheath gas option and SilicaTip emitter (New Objective; FS360-20-15-N-20-C12) utilization. ABIRD (Active Background Ion Reduction Device, ESI Source Solutions) was installed. MS data were acquired in targeted mode with 10 scan events measured in Orbitrap analyzer in total, covering survey scan and three MS/MS scans at different relative fragmentation energies (10, 25 and 35 %) for 3 peptide forms (0x phosphorylated – *m/z* 940.9968, 1x phosphorylated – *m/z* 980.4800, 2x phosphorylated – *m/z* 1020.9631; nine MS/MS scans in total). The scan range was 350–2000 or 100–2000 *m/z*, resolution 60000 or 15000 (at 400 *m/z*), target value of 1 × 106 or 5 × 104, and maximum injection time of 200 or 500 ms for the survey or MS/MS scan, respectively. The isolation window for MS/MS fragmentation was set to 2 *m/z*.

RESULTS & DISCUSSION

Divide-and-conquer strategy utilizing short truncated phospho-peptides prepared by solid-phase synthesis comprise a standard strategy to study interactions of modified IDP with their receptor proteins. In the case of the CTD, a number of studies used short truncated peptides (cca 14–16 residues in length) derived from the CTD repetitive heptad elements (YSPTSPS) [50–60]. These peptides were phosphorylated in different registers and scrutinized for binding affinity and mechanisms of binding towards their effector molecules. The results were then extrapolated for the entire CTD lengths. In some cases, the CTD fragments were truncated to suboptimal sizes and did not fully occupy the binding site of the effector proteins [52,61,62]. Often, the CTD binding proteins are large entities or even multimerize and form together with the CTD large molecular complexes, called CTDsomes [18]. Therefore, the CTD interactome studies require the use of the full-length or large fragment of the modified CTDs. Unfortunately, these long CTD peptides are beyond the limits of current state-of-the-art solid-phase synthetic approaches. We set out to employ *in vivo* system to overcome this issue, and it offers production of modified peptides with native N- and ▶

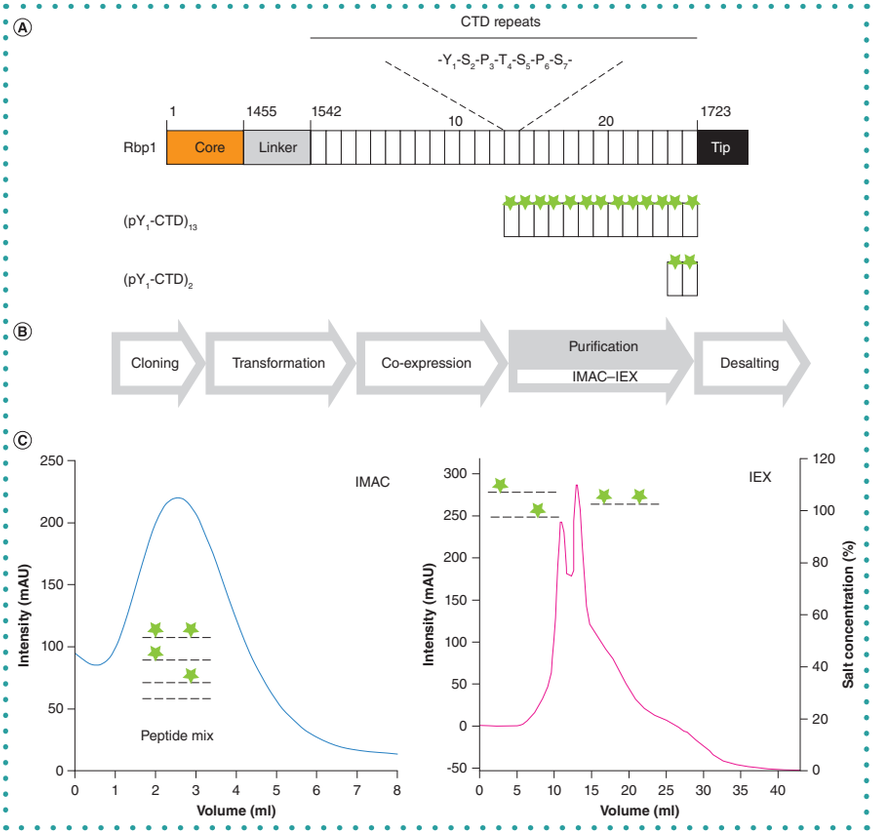


Figure 1. Peptide production work-flow. (A) Overview of Rpb1 and the used $(CTD)_2$ and $(CTD)_{13}$ constructs. Produced peptides herein referred as $(pY1\text{-}CTD)_2$ and $(pY1\text{-}CTD)_{13}$. (B) Scheme of the phosphorylated peptide production. Briefly, both effector and target substrate are cloned into pETDuet-1 vector, followed by plasmid transformation. After expression, the substrate is purified and the product is desalted upon lyophilization. (C) The purification procedure is simple, yet robust and involves the following steps: lysed cells are loaded on immobilized metal affinity chromatography (IMAC) column, where a mixture of phosphorylated and non-phosphorylated peptides are eluted after SUMO cleavage in a single peak. This mixture is loaded on IEX, where the peptide isoforms are separated. The first peak represents the a single-phosphorylated isoform, while the downstream peak represents the fully phosphorylated form $(pY1\text{-}CTD)_2$. Peptides are shown as dashed lines, the phosphate marks are indicated as green stars.

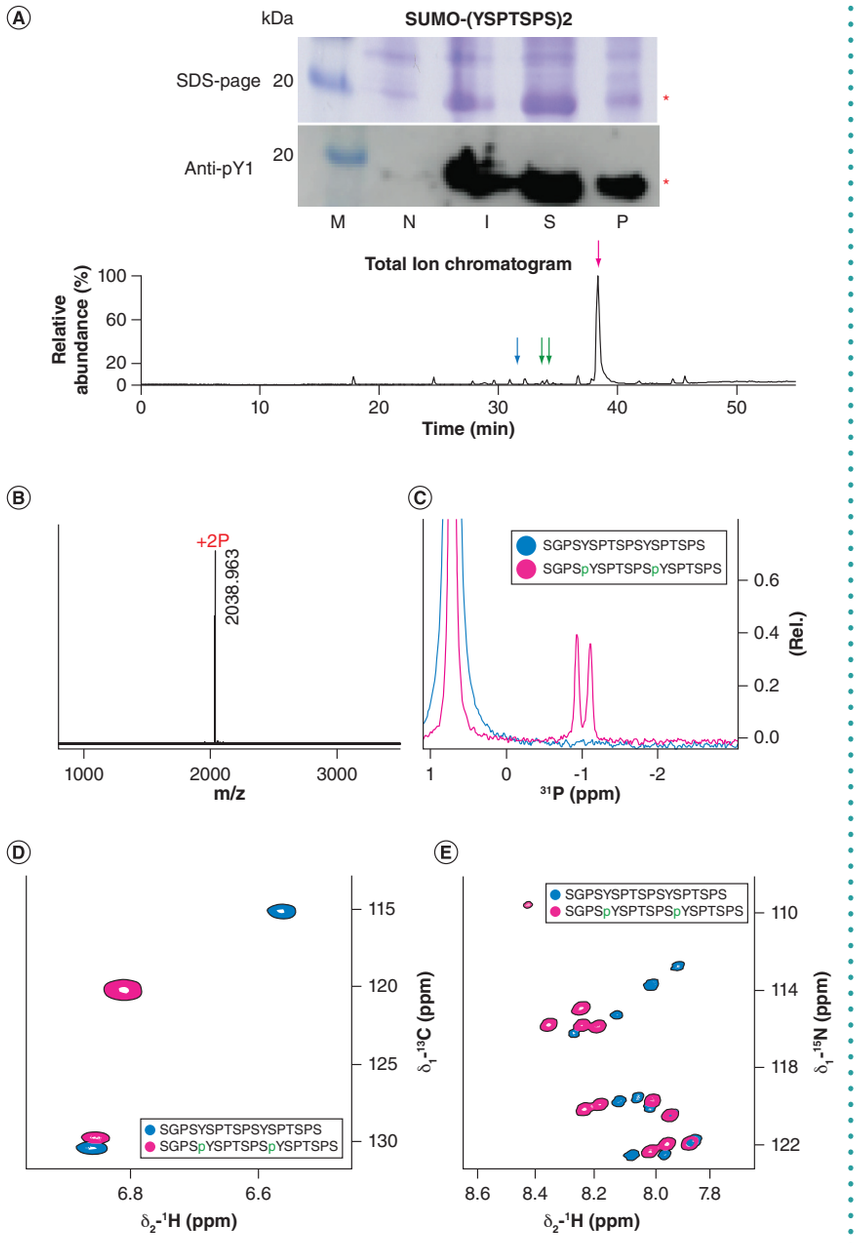


Figure 2. Production of $(pY1\text{-}CTD)_2$. (A) Upper Expression and solubility tests of the SUMO-$(CTD)_2$ construct. Expression and solubility tests analysed by SDS page and western blot with antibody against CTD phosphotyrosine (clone 3D12, MABE350 (Merck)). SUMO-$(CTD)_2$ is indicated by red asterisk. Lower, Total ion chromatogram from LC-MS analysis showing relative abundance of $(pY1\text{-}CTD)_2$ in the purified sample. Peaks corresponding to unphosphorylated $(CTD)_2$, singly phosphorylated and doubly phosphorylated variant $(pY1\text{-}CTD)_2$ are highlighted by blue, green, and red arrows, respectively. (B) MALDI-TOF mass spectra indicating presence of doubly phosphorylated peptide ($[M\text{-}H]^-$ theoretical = 2038,892). (C) Overlay of 1D ^{31}P spectra of phosphorylated (red) and non-phosphorylated (blue) peptide variant. Phosphorylations are depicted in green. The large peak around 0.8 ppm corresponds to residual phosphate buffer. (D) Overlay of 2D aromatic $^{1}H\text{-}^{13}C$ HSQC spectra of phosphorylated (red) and non-phosphorylated (blue) peptide variant. The e- and δ-carbons of tyrosine side-chains are perturbed due to the proximity of the phosphate group. Phosphorylations are depicted in green. (E) Overlay of the 2D $^{1}H\text{-}^{15}N$ HSQC spectra of phosphorylated (red) and non-phosphorylated (blue) peptide variants. Phosphorylations are depicted in green. I: Induced cells; M: marker; N: Non-induced cells; P: Pellet; S: Supernatant.

▶ C-termini with possibility of isotopic enrichment and residue-specific modification (depending on the specificity of modification enzyme) at a low cost. Additionally, this system can be applied to any peptide sequence including repetitive ones and any size of the peptide. The purified peptides are free of salts and organic solvents, in contrast to the solid-phase synthetized ones.

Our experiments with co-expression systems showed the best results for the system with two separated monocistronic transcripts each under its own promoter, as implemented in pETDuet-1. Having two separate promoters increases the expression compared to a bi-cistronic operon with single T7 promoter [63]. We also found that placing the c-Abl kinase in MCS1 resulted in a higher overall yield compared to situation where it was cloned in MCS2. This is likely the result of non-equal expression of both genes. The expression of the first one is slightly impaired as the MCS1 region lacking the T7 terminator is susceptible to the transcriptional read-through [64]. Thus, by cloning the c-Abl kinase in MCS1 we lowered its expression level in favor of its CTD substrate. Production of both the kinase and the CTD substrate *in vivo* eliminates cumbersome screening of reaction conditions. The CTD is difficult to produce as IDPs are prone to proteolysis and aggregation [34,65–67]. Several studies showed the improvement of solubility and stability by fusing the IDPs to a fusion partner e.g. SUMO [68,69], GST [70] or MBP [9]. We used SUMO tag, which facilitates expression, solubility and subsequent purification with highly active and site-specific Ulp1 protease [37–39].

We probed the ability of this expression approach by preparation of two constructs, $(CTD)_2$ and $(CTD)_{13}$. Two repeats of the CTD consensus heptad, $(pY1\text{-}CTD)_2$, was widely used in previous studies as it represents a minimal binding motif in some cases [50–60]. The large fragment of 13 CTD heptads, $(pY1\text{-}CTD)_{13}$, corresponds to the half of length of yeast's CTD, and it was used in structural studies only recently [18] (Figure 1A). Viability tests in yeast showed strong phenotypes for the CTD that was truncated to less than 13 repeats [71]. In both cases, we followed the scheme outlined in Figure 1B which takes about one week of

experimental work. The procedure starts with cloning of the kinase and substrate into the pETDuet-1 vector using MCS1 (NcoI/HindIII) and MCS2 (NdeI/XhoI) sites, respectively. The resulting plasmid was verified by sequencing. Expression and solubility for both substrates were assessed from the SDS-PAGE gels (Figure 2A & 3A). We note that expression conditions might be varied for a different kinase-substrate pair, e.g., time, temperature or inductor concentration. The biomass is then purified on affinity and ion-exchange (IEX) column using standard FPLC machine (Figure 1C). The elution from the affinity column contains a pool of peptide isoforms, which are subsequently separated on IEX. Our protocol enables to separate a mixture of phospho-peptides with a single phosphate resolution for shorter fragment and low-intermediate-high phosphorylated species for the $(CTD)_{13}$ peptides (Figure 3B). Peaks representing fully phosphorylated forms of $(pY1\text{-}CTD)_2$ and $(pY1\text{-}CTD)_{13}$ were analysed for purity on LC-MS. The purity of (pY1-CTD)2 has been 99% (Figure 2A), which is comparable to purities offered by peptide synthesizing companies.

Subsequently, eluted fractions were assayed for phosphorylation by western blot with antibody against the phosphorylated CTD on Y1 (anti-phospho-RNAPII [Tyr1] Ab, clone 3D12), where both constructs $(CTD)_2$ and $(CTD)_{13}$ were positive for phosphotyrosine. Next, we used 1D ^{31}P NMR phosphorus spectra to independently assess the presence of the phosphorylation. The spectra clearly show the frequency of two phosphorylations in respect to the free phosphate from the buffer, whereas no peaks are observed for the non-phosphorylated form (Figure 2C). These observations are further supported by mass spectrometry data for $(pY1\text{-}CTD)_2$ (Figure 2B) and $(pY1\text{-}CTD)_{13}$ (Figure 3B), and treatment of the $(pY1\text{-}CTD)_{13}$ with alkaline phosphatase (Figure 3B). The $(pY1\text{-}CTD)_{13}$ product is near homogenous in terms of number of phosphorylations. Further improvement may be achieved by optimizing the expression temperature and time, and using an ultra-high resolution IEX column. Next, we tested the identity of phosphorylated residues by additional NMR experiments. The aromatic 2D ^{1}H-^{13}C-HSQC experiment identified the tyrosine as the residue that possesses the phosphorylation mark. The phosphate group

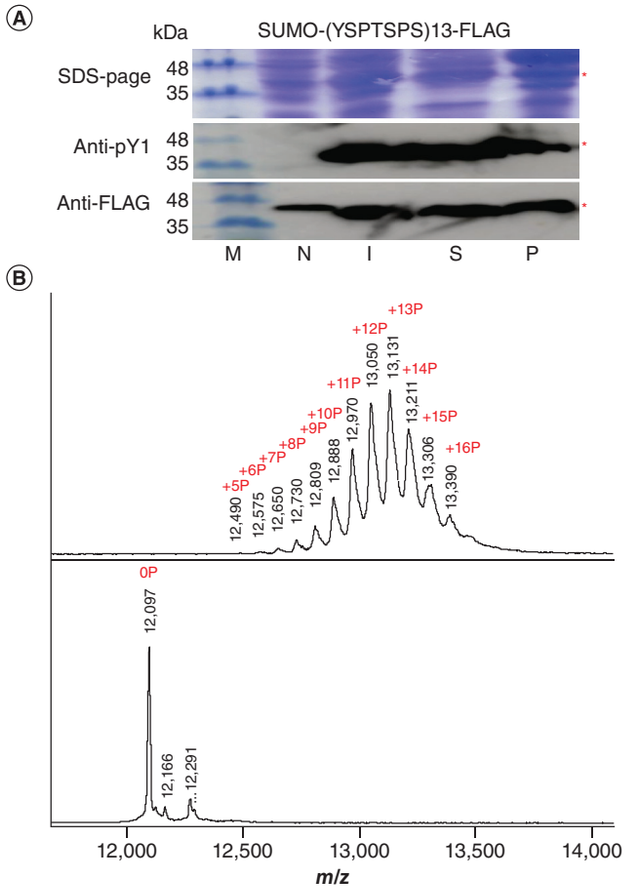


Figure 3. Production of $(pY1\text{-}CTD)_{13}$. (A) Expression and solubility tests of the SUMO-$(CTD)_{13}$-FLAG construct. Presence of the FLAG tag and phosphorylated tyrosine was verified by western blot analysis with antibody against FLAG epitope (Monoclonal ANTI-FLAG®M2, clone M2, Merck) and phosphotyrosine (Anti-phospho-RNAPII (Tyr1) Antibody, clone 3D12, Merck), respectively. SUMO-(YSPTSPS)13 is indicated by red asterisk.(B) Top, heterogeneity of the purified phosphorylated $(pY1\text{-}CTD)_{13}$ peptide analyzed by MALDI-TOF MS. The 13x phosphorylated peptide is the most abundant. Additional phosphorylations (>13+) correspond to tyrosines of the flanking sequences of the peptide construct. (B) Bottom, analysis of $(pY1\text{-}CTD)_{13}$ after treatment of phosphatase assays. Dominant peak corresponds to the non-phosphorylated form, $(Y1\text{-}CTD)_{13}$.
I: Induced cells; M: marker; N: Non-induced cells; P: Pellet; S: Supernatant.

alters the chemical environment of the neighbouring atoms in the tyrosine side-chain. As a result, the chemical shifts of aromatic Cδ and Ce atoms in the tyrosine side-chain display significant perturbations (Figure 2D). MS/MS fragmentation independently and unambiguously pinpointed the tyrosine phosphorylations (Supplementary Figure 1). This finding is supported by the specificity of c-Abl kinase in literature [32,33,72].

The CTD as an IDP might be subjected to proteolytic degradation [65,67]. Indeed, our initial trial with peptides lacking N- and C-terminal tags suffered from significant proteolysis. The level of proteolysis is significantly greater for longer than for the short peptides. To prevent this, we fused SUMO tag to the N-terminus for both constructs and added an additional FLAG tag signal peptide to the C-terminus in case of $(pY1\text{-}CTD)_{13}$. To monitor the peptide stability, we measured NMR spectra for the $(pY1\text{-}CTD)_2$ (Figure 2E). The spectrum shows that the number of residues corresponds to the theoretical sequence. The $(pY1\text{-}CTD)_{13}$ has a C-terminal FLAG tag, which was detected by western blot using the anti-flag monoclonal antibody (clone M2, Merck) (Figure 3B).

IDPs and IDRs that are subjected to posttranslational modifications are associated with human diseases, including cancer, cardiovascular diseases, amyloi- ▶

doses, neurodegenerative diseases, and diabetes [7]. Therefore, unravelling structural and functional information about these proteins is of biomedical importance as they represent a novel class of drug targets that aims to modulate protein-protein interactions. Here, we combined co-expression of tyrosine kinase with its substrate, the CTD of RNAPII, and demonstrated that the modified substrate can be expressed and purified to homogeneity via a simple and robust protocol. The pair of modification enzyme and target can be specifically tailored depending on the requested modifications of IDP. To phosphorylate serines of the CTD, BRD4 can be used as the kinase module in the expression system [73], whereas the use of cyclin-dependent kinases (Cdks), e.g., CDK7 and CDK9 [74–76], is challenging and would require additional modification of the *E.coli* expression system, as their activity is mainly regulated by association with regulatory subunits and their PTMs. We believe that future biochemical, biophysical and structural studies of modified IDPs will benefit from our robust protocol for *in vivo* preparation and purification of modified peptides.

FUTURE PERSPECTIVE

We believe that future biochemical, biophysical and structural studies of modified IDPs will benefit from our robust protocol for *in vivo* preparation and purification of modified peptides. Our method paves the way for further development to obtain high yields of selectively phosphorylated tyrosine, serine and threonine residues contained in the IDPs and IDRs. The future development lies in advancement of eukaryotic expression systems that will facilitate production of more complex IDPs with multiple modifications.

AUTHOR CONTRIBUTIONS

PB carried out sample preparation, performed and analyzed NMR experiments, and wrote the manuscript. OŠ performed and interpreted the MS analyses. RS and KK conceived and designed the project, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank D Potesil for LC-MS/MS analyses and data processing, and J Novacek for technical support of NMR measurements. We thank CD Lima for gift of pSMT3-pET28b and pULP-pET28b plasmids.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 649030). This publication reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains. The results of this research have been acquired within CEITEC 2020 (LQ1601) project with financial contribution made by the Ministry of Education, Youths and Sports of the Czech Republic within special support paid from the National Programme for Sustainability II funds. This work was also supported by the Czech Science Foundation (15-17670S and 18-11397S to RS). We acknowledge the CIISB research infrastructure project LM2015043 funded by MEYS CR for the financial support of the measurements at the Proteomics Core Facility and at the Josef Dadok National NMR Centre, and for their support with obtaining scientific data presented in this paper. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

OPEN ACCESS

SUPPLEMENTARY DATA

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2019-0033

REFERENCES

1. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337(3), 635–645 (2004).
2. Uversky VN. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta BBA – Proteins Proteomics* 1834(5), 932–951 (2013).
3. Dunker AK, Lawson JD, Brown CJ *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* 19(1), 26–59 (2001).
4. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6(3), 197–208 (2005).
5. Yoon M-K, Mitrea DM, Ou L, Kriwacki RW. Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochem. Soc. Trans.* 40(5), 981–988 (2012).
6. Harlen KM, Churchman LS. Subgenic Pol II interactomes identify region-specific transcription elongation regulators. *Mol. Syst. Biol.* 13(1), 900 (2017).
7. Uversky VN. Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. *Front. Mol. Biosci.* 1, 6 (2014).
8. Lebendiker M, Danieli T. Production of prone-to-aggregate proteins. *FEBS Lett.* 588(2), 236–246 (2014).
9. Portz B, Lu F, Gibbs EB *et al.* Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* 8, 15231 (2017).
10. London JW, Shaw LM, Fetterolf D, Garfinkel D. A systematic approach to enzyme assay optimization, illustrated by aminotransferase assays. *Clin. Chem.* 21(13), 1939–1952 (1975).
11. Peck SC. Analysis of protein phosphorylation: methods and strategies for studying kinases and substrates. *Plant J.* 45(4), 512–522 (2006).
12. Mejuch T, Waldmann H. Synthesis of lipidated proteins. *Bioconjug. Chem.* 27(8), 1771–1783 (2016).
13. Hejjaoui M, Butterfield S, Fauvet B *et al.* Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: α-synuclein phosphorylation at tyrosine 125. *J. Am. Chem. Soc.* 134(11), 5196–5210 (2012).
14. Ottesen JJ, Huse M, Sekedat MD, Muir TW. Semisynthesis of phosphovariants of Smad2 reveals a substrate preference of the activated TβRI kinase †. *Biochemistry.* 43(19), 5698–5706 (2004).
15. Ludwig C, Pfeiff M, Linne U, Mootz HD. Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew. Chem. Int. Ed Engl.* 45(31), 5218–5221 (2006).
16. Tarrant MK, Cole PA. The chemical biology of protein phosphorylation. *Annu. Rev. Biochem.* 78(1), 797–825 (2009).
17. Chacko BM, Qin B, Correia JJ, Lam SS, de Caestecker MP, Lin K. The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. *Nat. Struct. Biol.* 8(3), 6 (2001).
18. Jasnovidova O, Klumpler T, Kubíček K, Kalynych S, Plevka P, Stefl R. Structure and dynamics of the RNAPII CTDsome with Rtt103. *Proc. Natl. Acad. Sci.* 114(42), 11133–11138 (2017).
19. Pirman NL, Barber KW, Aerni HR *et al.* A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nat. Commun.* 6(1), (2015).
20. Kato Y. Tunable translational control using site-specific unnatural amino acid incorporation in ***Escherichia coli***. *PeerJ.* 3 (2015).
21. Chen S, Maini R, Bai X, Nangreave RC, Dedkova LM, Hecht SM. Incorporation of phosphorylated tyrosine into proteins: ***in vitro*** translation and study of phosphorylated IκB-α and its interaction with NF-κB. *J. Am. Chem. Soc.* 139(40), 14098–14108 (2017).
22. Katoh T, Passioura T, Suga H. Advances in ***in vitro*** genetic code reprogramming in 2014–2017. *Synth. Biol.* 3(1), (2018).
23. Schmidt CM, Shis DL, Nguyen-Huu TD, Bennett MR. Stable maintenance of multiple plasmids in ***E. coli*** using a single selective marker. *ACS Synth. Biol.* 1(10), 445–450 (2012).
24. Tolia NH, Joshua-Tor L. Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods.* 3(1), 55–64 (2006).
25. Johnston K, Clements A, Venkataramani RN, Trievel RC, Marmorstein R. Coexpression of proteins in bacteria using t7-based expression plasmids: expression of heteromeric cell-cycle and transcriptional regulatory complexes. *Protein Expr. Purif.* 20(3), 435–443 (2000).
26. Bross P, Andersen BA, Winter V *et al.* Co-overexpression of bacterial GroESL chaperonins partly overcomes non-productive folding and tetramer assembly of E. coli-expressed human medium-chain acyl-CoA dehydrogenase (MCAD) carrying the prevalent disease-causing K304E mutation. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* 1182(3), 264–274 (1993).
27. Khokhlatchev A, Xu S, English J, Wu P, Schaefer E, Cobb MH. Reconstitution of mitogen-activated protein kinase phosphorylation cascades in bacteria efficient synthesis of active protein kinases. *J. Biol. Chem.* 272(17), 11057–11062 (1997).

28. Busso D, Peleg Y, Heidebrecht T *et al.* Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: a benchmarking study. *J. Struct. Biol.* 175(2), 159–170 (2011).
29. Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS. Plasmid-encoded protein: The principal factor in the 'metabolic burden' associated with recombinant bacteria. *Biotechnol. Bioeng.* 35(7), 668–681 (1990).
30. Birnbaum S, Bailey JE. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli. Biotechnol. Bioeng.* 37(8), 736–745 (1991).
31. Corchero JL, Villaverde A. Plasmid maintenance in ***Escherichia coli*** recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng.* 58(6), 625–632 (1998).
32. Baskaran R, Dahmus ME, Wang JY. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc. Natl. Acad. Sci.* 90(23), 11167–11171 (1993).
33. Baskaran R, Chiang GG, Wang JYJ. Identification of a binding site in c-Abl tyrosine kinase for the C-terminal repeated domain of RNA polymerase II. *Mol. Cell Biol.* 16, 9 (1996).
34. Minde DP, Halff EF, Tans S. Designing disorder: tales of the unexpected tails. *Intrinsically Disord. Proteins* 1(1), e26790 (2013).
35. Mossessova E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell.* 5(5), 865–876 (2000).
36. Lima CD, Mossessova E. Rapidly cleavable sumo fusion protein expression system for difficult to express proteins [Internet]. (2011). www.freepatentsonline.com/7910364.html.
37. Prakash A, Parsons SJ, Kyle S, McPherson MJ. Recombinant production of self-assembling β-structured peptides using SUMO as a fusion partner. *Microb. Cell Factories* 11(1), 92 (2012).
38. Sadr V, Saffar B, Emamzadeh R. Functional expression and purification of recombinant Hepcidin25 production in *Escherichia coli* using SUMO fusion technology. *Gene* 610, 112–117 (2017).
39. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci. Publ. Protein Soc.* 15(1), 182–189 (2006).
40. Corden JL. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chem. Rev.* 113(11), 8423–8455 (2013).
41. Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26(19), 2119–2137 (2012).
42. Jeronimo C, Watanabe S, Kaplan CD, Peterson CL, Robert F. The histone chaperones FACT and Spt6 restrict H2A.Z from intragenic locations. *Mol. Cell.* 58(6), 1113–1123 (2015).
43. Corden JL, Cadena DL, Ahearn JM, Dahmus ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc. Natl Acad. Sci. USA* 82(23), 7934–7938 (1985).
44. Allison LA, Wong JK, Fitzpatrick VD, Moyle M, Ingles CJ. The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae, Drosophila melanogaster,* and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.* 8(1), 321–329 (1988).
45. Meinhart A. A structural perspective of CTD function. *Genes Dev.* 19(12), 1401–1415 (2005).
46. Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* 23(9), 771–777 (2016).
47. Suh H, Ficarro SB, Kang U-B, Chun Y, Marto JA, Buratowski S. Direct ANALYSIS OF PHOSPHORYLATION SITES on the Rpb1 C-terminal domain of RNA polymerase II. *Mol. Cell.* 61(2), 297–304 (2016).
48. Schüller R, Forné I, Straub T *et al.* Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol. Cell.* 61(2), 305–314 (2016).
49. Jasnovidova O, Stefl R. The CTD code of RNA polymerase II: a structural view: The CTD code of RNA polymerase II. *Wiley Interdiscip. Rev. RNA* 4(1), 1–16 (2013).
50. Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc. Natl. Acad. Sci. USA* 102(49), 17636–17641 (2005).
51. Zhang Y, Kim Y, Genoud N *et al.* Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Mol. Cell.* 24(5), 759–770 (2006).
52. Lunde BM, Reichow SL, Kim M *et al.* Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* 17(10), 1195–1201 (2010).
53. Sun M, Larivière L, Dengl S, Mayer A, Cramer P. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J. Biol. Chem.* 285(53), 41597–41603 (2010).
54. Liu J, Zhang J, Gong Q *et al.* Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. *J. Biol. Chem.* 286(33), 29218–29226 (2011).
55. Mayer A, Heidemann M, Lidschreiber M *et al.* CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science* 336(6089), 1723–1725 (2012).
56. Kubíček K, Cerna H, Holub P *et al.* Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* 26(17), 1891–1896 (2012).
57. Tudek A, Porrua O, Kabzinski T *et al.* Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Mol. Cell.* 55(3), 467–481 (2014).
58. Ni Z, Xu C, Guo X *et al.* RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat. Struct. Mol. Biol.* 21(8), 686–695 (2014).
59. Jasnovidova O, Krejcikova M, Kubíček K, Stefl R. Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep.* 18(6), 906–913 (2017).
60. Nemec CM, Yang F, Gilmore JM *et al.* Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. *Proc. Natl. Acad. Sci. USA* 114(20), E3944–E3953 (2017).
61. Verdecia MA, Bowman ME, Ping K, Hunter T, Noel JP. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* 7(8), 6 (2000).
62. Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell.* 11(6), 1549–1561 (2003).
63. Kim K-J, Kim H-E, Lee K-H *et al.* Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Sci. Publ. Protein Soc.* 13(6), 1698–1703 (2004).
64. Mädje K, Schmölzer K, Nidetzky B, Kratzer R. Host cell and expression engineering for development of an E. coli ketoreductase catalyst: enhancement of formate dehydrogenase activity for regeneration of NADH. *Microb. Cell Factories* 11(1), 7 (2012).
65. Hwang PM, Pan JS, Sykes BD. A PagP fusion protein system for the expression of intrinsically disordered proteins in *Escherichia coli*. *Protein Expr. Purif.* 85(1), 148–151 (2012).
66. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van Y-YJ, Dunker AK. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* 51(37), 7250–7262 (2012).
67. Goda N, Matsuo N, Tenno T *et al.* An optimized N^{pro}-based method for the expression and purification of intrinsically disordered proteins for an NMR study. *Intrinsically Disord. Proteins* 3(1), e1011004 (2015).
68. Mo Q, Fu A, Lin Z, Wang W, Gong L, Li W. Expression and purification of antimicrobial peptide AP2 using SUMO fusion partner technology in ***Escherichia coli***. *Lett. Appl. Microbiol.* 67(6), 606–613 (2018).
69. Satakarni M, Curtis R. Production of recombinant peptides as fusions with SUMO. *Protein Expr. Purif.* 78(2), 113–119 (2011).
70. Czudnochowski N, Bösken CA, Geyer M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* 3(1), (2012).
71. West ML, Corden JL. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140(4), 1223–1233 (1995).
72. Baskaran R, Escobar SR, Wang JYJ. Nuclear c-Abl Is a COOH-terminal repeated domain (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation. *Cell Growth Differ.* 10, 387–396 (1999).
73. Devaiah BN, Lewis BA, Cherman N *et al.* BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA polymerase II carboxy-terminal domain. *Proc. Natl. Acad. Sci. USA* 109(18), 6927–6932 (2012).
74. Boehning M, Dugast-Darzacq C, Rankovic M *et al.* RNA polymerase II clustering through CTD phase separation. *Nat. Struct. Mol. Biol.* 25(9), 833–840 (2018).
75. Price DH. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* 20(8), 2629–2634 (2000).
76. Lu F, Portz B, Gilmour DS. The C-terminal domain of RNA polymerase II is a multivalent targeting sequence that supports drosophila development with only consensus heptads. *Mol. Cell.* 73(6), 1232–1242.e4 (2019).