Screening a Protein Array with Synthetic Biotinylated Inorganic Polyphosphate To Define the Human PolyP-ome

Cristina Azevedo, Jyoti Singh, Nicole Steck, Alexandre Hofer, Felix A. Ruiz, Tanya Singh, Henning J. Jessen, and Adolfo Saiardi

†Medical Research Council Laboratory for Molecular Cell Biology, University College London, London, WC1E 6BT, United Kingdom
‡Institute of Organic Chemistry, University of Freiburg, Albertstrasse 21, 79104 Freiburg, Germany
§Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Supporting Information

ABSTRACT: Phenotypes are established by tight regulation on protein functions. This regulation can be mediated allosterically, through protein binding, and covalently, through post-translational modification (PTM). The integration of an ever-increasing number of PTMs into regulatory networks enables and defines the proteome complexity. Protein PTMs can occur enzymatically and nonenzymatically. Polyphosphorylation, which is a recently discovered PTM that belongs to the latter category, is the covalent attachment of the linear ortho-phosphate polymer called inorganic polyphosphate (polyP) to lysine residues. PolyP, which is ubiquitously present in nature, is also known to allosterically control protein function. To date, lack of reagents has prevented the systematic analysis of proteins covalently and/or allosterically associated with polyP. Here, we report on the chemical synthesis of biotin-modified monodisperse short-chain polyP (bio-polyP8-bio) and its subsequent use to screen a human proteome array to identify proteins that associate with polyP, thereby starting to define the human polyP-ome.

The human proteome is much more complex than previously recognized, with an ever-increasing number of protein post-translational modifications (PTMs) being identified. Some PTMs are introduced enzymatically, while others are not dependent on enzymes. Polyphosphorylation is a newly discovered nonenzymatic PTM in which the linear polymer inorganic polyphosphate (polyP), which is solely composed of condensed phosphates, is covalently attached to lysine residues through a phosphoramidate P−N bond. This modification results from a nucleophilic attack by a lysine residue of the target protein to an internal phosphate of the polyP chain. The number of orthophosphate monomers in polyP ranges from tens to thousands, making this modification polydisperse or nonuniform. So far, the acid lability of the phosphoramidate functional group and the polymeric nature of polyphosphorylation has prevented its detection by mass spectrometry. Characterization of polyphosphorylation has only been possible in Saccharomyces cerevisiae and through extensive biochemical examinations. The structural determinant for polyphosphorylation resides in a specific cluster designated the PASK domain [Poly-Acidic serine (S) and lysine (K) rich]. However, the required extension of this domain and the importance of each residue for the modification are unknown. Two polyphosphorylation targets were originally identified: Nuclear signal recognition 1 (Nsr1) and its interacting partner Topoisomerase 1 (Top1). Recently, 15 more yeast-validated targets have been identified through bioinformatics analysis.

Inorganic polyphosphate is present in all domains of life: bacteria, archaea, and eukaryotes. In mammals, polyP plays various roles including acting as a hemostasis regulator as a nucleating factor for fibril formation, and is required for bone mineralization. The molecular mechanisms by which polyP exerts these functions are unknown. Two polyphosphorylation targets were originally identified: Nuclear signal recognition 1 (Nsr1) and its interacting partner Topoisomerase 1 (Top1). Recently, 15 more yeast-validated targets have been identified through bioinformatics analysis. Here, we report an approach to systematically identify proteins that associate with polyP, thereby starting to define the human polyP-ome.
length, that is capped at both ends with biotin (bio-polyP8-bio). This chemical tool was used in a human proteome array to identify proteins that associate electrostatically and covalently with polyP, thereby defining the “polyP-ome”.

The synthesis of monodisperse polyP building blocks has been an unsolved problem and only the modification of polydisperse polyP has been reported. Here, we present a bottom-up approach that enables the synthesis of short oligophosphates with a defined composition (Figure 1). This represents the first report on a bidirectional extension of a diphosphate unit with activated trimetaphosphate followed by ring-opening with amine nucleophiles. This strategy enabled the modification of a P8 stretch with P-amidates at the terminal positions introducing an alkyne. The bis-alkyne was used in a 2-fold Huisgen cycloaddition reaction to install biotin labels at both ends. In brief, trimetaphosphate was activated with mesitylenesulfonyl chloride (Ms-Cl) and N-methylimidazole and reacted with substoichiometric amounts of pyrophosphate. The formation of a cyclic ultraphosphate at both termini was analyzed by $^{31}$P NMR [see Figure S1 (spectrum 10) in the Supporting Information] and the expected coupling pattern for the different phosphate units was observed (d, dd, and dt). The intermediate was ring-opened by the addition of propargylamine. All reactions occur in one flask. Next, the alkyne-polyP8-alkyne is click-labeled with biotin, using the Huisgen dipolar cycloaddition.

Overall, the construction of a monodisperse P8 unit with modified termini was achieved in a one-flask reaction with an overall yield of 13%. The alkynylated polyP8 was then subjected to azide–alkyne cycloaddition introducing the biotin labels at both ends. The product precipitated from solution by addition of sodium iodide/acetone and the sodium salt of bio-polyP8-bio was collected in 43% yield.

To identify the human polyP-ome, we screened a commercially available Human Proteome Array slide, containing over 15,000 human proteins expressed and purified from yeast, with bio-polyP8-bio. In 2004, Nomura et al. demonstrated that $^{32}$P-labeled polyP could be specifically transferred to proteins blotted onto membranes. Similarly, the transfer of biotin to recombinant His-Top1(PASK) can be detected upon incubation with bio-polyP8-bio (Figure S2A in the Supporting Information). Moreover, we could detect transfer of polyP to the protein when 1 μM of bio-polyP8-bio was incubated with 0.5 μg His-Top1(PASK) (Figure S2B in the Supporting Information), demonstrating the validity of bio-polyP8-bio as a screening tool to identify polyP clients. Since the exact amount of protein spotted onto the glass chip is unknown and the substrate (glass) is different from nitrocellulose, 20 μM of bio-polyP8-bio were used in the initial screen (Screen A). Under these conditions, 309 proteins were detected with a 1.3 Hit score (HS) cutoff above background, corresponding to ~2% of the library (see Table S1 in the Supporting Information).

We expected to detect targets that associate with polyP through either electrostatic or covalent interactions. To distinguish between these, we relied on a protein mobility shift observed on Bis-Tris Gels/NuPAGE for covalent polyphosphorylation. Conversely, for proteins that interact with polyP through ionic interactions, no mobility shift on Bis-Tris Gels/NuPAGE should be expected. However, a shift
should be observed under native PAGE conditions. In budding yeast, an organism containing high levels of polyP\textsuperscript{24,25} and in which the polyP biosynthetic pathway is characterized, it is possible to study the mobility shift of proteins \textit{in vivo}, comparing between mutants with altered polyP metabolism.\textsuperscript{26} In mammalian cells, this direct assay is not possible because the concentration of polyP is much lower,\textsuperscript{27} and the enzymology of its synthesis is unknown. For this, we relied on an indirect approach: polyP was added exogenously to mammalian protein extracts before assaying the mobility shift. As a positive control, we used the human homologue of the yeast target Nsr1, nucleolin, that possesses a striking PASK domain (see Figure S3A in the Supporting Information). To confirm nucleolin as a target, we expressed it in wild-type yeast. GST-nucleolin has a smeary appearance with a molecular weight that is greater than its predicted size of 77 kDa (see Figure S3B in the Supporting Information). In vtc4Δ, a yeast with impaired polyP synthesis, GST-nucleolin runs as a single band of lower molecular weight. Moreover, exogenously incubating vtc4Δ extracts with polyP results in mobility shift (see Figure S3B). Similarly, in HeLa cell extracts, endogenous nucleolin is shifted upward upon polyP treatment (see Figure S3C in the Supporting Information, as well as Figure 2A). Together, these data indicate that nucleolin is a polyphosphorylation target. The following criteria were used to select the screening hits to authenticate: molecular weight between 40 and 110 kDa, expressed in HeLa, and for which an antibody is available. Furthermore, preference was given to proteins belonging to a family of which more than one member gave a positive hit. To avoid bias, the protein sequence itself was not taken into consideration (i.e., presence of PASK domain). Based on these criteria, we validated 11 hits (see Table 1). We observed that Annexin A6 and A9, FAF1, DPP3, RUVBL1 and RUVBL2, vinculin, and AP1B1 did not show a mobility shift on Bis-Tris Gels/NuPAGE, suggesting that, if there is association with polyP, it is through electrostatic interaction (see Figure 2A and Figure S4A in the Supporting Information). To test this hypothesis, we investigated their mobility under native conditions. FAF1 and DPP3 also failed to show a mobility shift under native conditions, suggesting...
that these proteins do not associate with polyP and are false positives (Figure S4B in the Supporting Information). Annexin A6 showed a mobility shift under native conditions, suggesting that there is an electrostatic association with polyP (see Figure 2B). Annexin A9, vinculin, AP1B1, and RUVBL1 and RUVBL2 antibodies did not recognize the native form of the proteins. To overcome this, we performed gel filtration experiments. Gel filtration analysis of annexin A6 and A9, RUVBL1 and RUVBL2, and AP1B1 in HeLa protein extracts treated with polyP showed that these proteins elute in earlier fractions than untreated proteins, indicating an increase in size as a direct consequence of polyP association or through polyP-dependent polyphosphorylation.

Annexin A6, A9, TTC27, HSP90B1, and eIF2B epsilon showed a clear mobility shift on Bis-Tris Gels/NuPAGE, suggesting that they are covalent polyphosphorylation targets (see Figure 3A).

Our validation revealed that vinculin, FAF1, and DPP3 are false positives. Therefore, we performed a higher stringency screen by decreasing the amount of bio-polyP8-bio 4-fold (Screen B). In doing so, we approximately halved the total number of hits obtained (155 proteins), which corresponds to ~1% of the library (Table S2 in the Supporting Information). Overlap analysis showed that 50% of the proteins (i.e., 75 hits) are common between the screens. Importantly, neither vinculin, FAF1, nor DPP3 were present in the higher stringency screen, suggesting that we succeeded in reducing the number of false positives. Bioinformatics analysis of the combined screens (see the Methods section in the Supporting Information) did not reveal an enrichment of any particular protein family or protein domain. However, the analysis of the gene ontology (GO) biological process revealed an enrichment of proteins involved in metabolic processes and, specifically, in amino acid biosynthesis (see Figure S6 in the Supporting Information).

Until now, we could only identify polyphosphorylation targets that have striking PASK domains such as Nsr1, Top1,5 and Nucleolin, and, to characterize them, we solely relied on an independent repeats produced similar results. Stroke lines are used when the figure was assembled from the same gel/membrane.
the confirmed hits from both screens revealed that, with this new method, we managed to identify proteins, such as TTC27, eIF2B, gelsolin, and EYA1, with very subtle PASK domains and, consequently, less-evident shifts. Indeed, a recently published bioinformatics screen looking for proteins with prominent PASK-domains (one or more 20 amino acid stretches composed of at least 75% D/E/S and at least one K) reported yeast and mammalian targets, with little overlap with our hits. The reason our screen did not identify hits with striking PASK domains resides in the fact that, when expressed and purified from yeast, which is an organism with a very high amount of polyP, these proteins could already be associated with polyP and are therefore less likely to be identified in this screen. An example is nucleolin, identified as a polyphosphorylation target based on its homology with the yeast Nsr1, but was not a hit in our screens. When expressed in WT yeast, nucleolin showed a very high mobility shift (Figure S3B in the Supporting Information) suggesting that, when purified, this protein is already highly polyphosphorylated, preventing it from further modification with bio-polyP8-bio. In contrast, proteins with subtle PASK domains are likely to lose the association with yeast polyp upon purification, because of the known post-extraction instability of polyphosphorylation, and could therefore associate with bio-P8-bio.

Some of the identified polyP targets are members of conserved protein families. To expand the polyP-ome to other family members, we performed sequence analysis on some of the targets. Analysis of HSP90B1 revealed the presence of a possible PASK domain within the “charged linker” region connecting the N-terminus to the middle domain (Figure S7A in the Supporting Information). This region is conserved in HSP90B1’s cytosolic counterparts, HSP90α/β (Figure S7B in the Supporting Information). We tested if HSP90α/β could also be polyphosphorylated. Upon polyP treatment, cytosolic HSP90s (HSP90α/β) show a significant mobility shift on Bis-Tris Gels/NuPAGE gels (Figure S7C in the Supporting Information). Similarly, sequence analysis of AP1B1 revealed a region at the N-terminus, within the head domain, with a few lysine and acidic amino acids (Figures S8A and S8B in the Supporting Information). This sequence is conserved in AP1B1’s closest homologue AP2B1, but not in AP3B1. However, AP3B1, which is a protein previously shown to be polyphosphorylated on serine residues, contains two further regions with strong PASK domain similarity: one within the head and the other within the hinge domains (see Figures S8A and S8B). We tested if AP3B1 could be a target of polyphosphorylation. Myc-AP3B1 showed a mobility shift on NuPAGE upon incubation of transfected HeLa protein extracts with polyP (see Figure S8C). This result suggests that, unlike that observed for AP1B1, the association of AP3B1 with polyP is covalent.

In conclusion, the original use of bio-polyP8-bio has allowed us to start defining the mammalian polyP-ome and, more importantly, will help us to start defining the minimal PASK domain. We identified a few hundred putative polyP-associated targets with subtle PASK domains necessary for polyphosphorylation. This, however, is likely to still represent an underestimation of the total number of targets, since the protein microchip library available lacks 1/4 of the human proteome. Our work confirms that proteins interact with polyP via two distinct modes of action: electrostatic association and covalent modification. Lysine polyphosphorylation as a covalent PTM has only been described very recently, and, thus far, only a handful of proteins have been reported to associate with polyP electrostatically. It will be important to study the precise nature of the electrostatic interactions between polyP and proteins and its regulation.

The bio-polyP8-bio synthesized here (recall Figure 1) consists of only a relatively short polyP chain. However, this new method will enable the bidirectional synthesis of much-longer condensed phosphates, which might help identify additional polyphosphorylation targets, since it is possible that the polyP chain length determines substrate specificity. Moreover, the use of click chemistry will provide useful tools to construct a variety of other compounds, such as fluorescently tagged defined polyP. The defined mass of biopolyP8-bio make it a versatile tool for polyP research that will facilitate the identification of polyphosphorylated lysine residues, since only a restricted range of added mass with different charge states must be analyzed, bypassing one of the main problems that arises from the polymeric nature of this and other PTMs, namely, the unknown size of the modification.
