

1 **An enhanced chemo-enzymatic method for loading substrates onto carrier**
2 **protein domains**

3

4 Tiia Kittilä¹ and Max J. Cryle^{1,2,3*}

5 *1. Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse*
6 *29, 69120 Heidelberg, Germany.*

7 *2. The Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology,*
8 *Monash University, Clayton, Victoria 3800, Australia*

9 *3. EMBL Australia, Monash University, Clayton, Victoria 3800, Australia.*

10 * Address correspondence to: A/Prof Dr. Max Cryle (max.cryle@monash.edu)

11 **Abstract**

12 Non-ribosomal peptide synthetase (NRPS) machineries produce many medically relevant peptides that
13 cannot be easily accessed by chemical synthesis. Thus, understanding NRPS mechanism is of crucial
14 importance to allow efficient redesign of these machineries in order to produce new compounds.
15 During NRPS-mediated synthesis, substrates are covalently attached to PCPs, and studies of NRPSs
16 are impeded by difficulties in producing PCPs loaded with substrates. Different approaches to load
17 substrates on to PCP domains have been described, but all suffer from difficulties in either the
18 complexity of chemical synthesis or low enzymatic efficiency. Here, we describe an enhanced chemo-
19 enzymatic loading method that combines two approaches into a single, highly efficient one-pot
20 loading reaction. First, D-pantetheine and ATP are converted into dephospho-coenzyme A via the
21 actions of two enzymes from coenzyme A (CoA) biosynthesis. Next, phosphoadenylates are
22 dephosphorylated using alkaline phosphatase to allow linker attachment to PCP domain by Sfp mutant
23 R4-4, which is inhibited by phosphoadenylates. This route does not depend on activity of the
24 commonly problematic dephospho-CoA kinase, and therefore offers an improved method for substrate
25 loading onto PCP domains.

26 **Keywords:** carrier protein; post-translational modification; non-ribosomal peptide synthetase;
27 phosphopantetheinyl transferase; coenzyme A.

28

29 **Introduction**

30 Non-ribosomal peptide synthetases (NRPSs) produce a wide variety of bioactive molecules, often with
31 complex chemical structures. (Al Toma et al. 2015; Süssmuth and Mainz 2017) Due to this
32 complexity, many such compounds of medicinal interest are still produced through fermentation,
33 which makes understanding NRPS function highly important for future efforts in compound
34 reengineering. NRPS machineries typically utilize a modular architecture of repeating domains to
35 synthesise their peptide products, with the core domains for peptide synthesis being adenylation (A)
36 domains, condensation (C) domains and peptidyl carrier protein (PCP) domains (Figure 1) (Süssmuth
37 and Mainz 2017). A domains select and activate amino acids in an ATP-dependant process and load
38 them onto the phosphopantetheine (PPE) arm of a neighbouring PCP domain, where the amino acid
39 remains tethered as a thioester (Kittilä et al. 2016a). This PCP-bound aminoacyl thioester is then
40 accepted by a C domain, where an upstream PCP-bound amino acid or peptide is condensed to form a
41 new peptide bond with concomitant transfer of the upstream molecule onto the downstream
42 aminoacyl-PCP. By reproducing these catalytic steps, and the introduction of additional tailoring
43 domains, multiple modules can in turn generate peptides whose length is governed by the number of
44 modules within the NRPS (Hur et al. 2012; Payne et al. 2017). Once the peptide has been fully
45 synthesised, a thioesterase (TE)-domain typically cleaves the peptide from the NRPS machinery
46 (Horsman et al. 2016).

47 Given the significance of the compounds produced by NRPS systems, these assembly lines are
48 important targets for structural and biochemical characterisation (Süssmuth and Mainz 2017).
49 Studying NRPS systems is challenging however due to the complexity of NRPS machineries and the
50 requirement for intermediates to be covalently loaded onto PCP domains before they are recognized as
51 substrates by any catalytic domains (Kittilä et al. 2016a). A domains can be used to load substrates
52 onto PCP domains, but unfortunately A domains can be very selective for their substrates and
53 therefore it is not possible to load a wide range of compounds *via* this route (Ehmann et al. 2000a;
54 Henderson et al. 2014; Mitchell et al. 2012; Villiers and Hollfelder 2009). In some cases, substrates
55 bound to a small molecule mimic of the PPE-linker (N-acetylcystamine thioesters, SNACs) are

56 accepted (Ehmann et al. 2000b; Luo et al. 2002; Roche and Walsh 2003), but this approach lacks
57 general utility as it removes crucial protein-protein interactions and not all NRPS systems accept
58 SNAC-bound substrates (Figure 2). Thus, one of the biggest breakthroughs in NRPS characterization
59 has been the identification of a promiscuous phosphopantetheinyl transferase (PPTase) that can load
60 PCP domains with substrates that are coupled to coenzyme A (CoA): the enzyme Sfp from *Bacillus*
61 *subtilis* (Beld et al. 2014; Belshaw et al. 1999; Lambalot et al. 1996; Quadri et al. 1998). This enzyme
62 can load unnatural amino acids and peptides onto PCP domains *in vitro*, which has enabled detailed
63 characterization of the enzymatic steps during NRPS catalysis. As the Sfp enzyme is promiscuous
64 towards the compounds coupled to CoA, this route has also been used for protein crosslinking, which
65 has been valuable in crystallographic studies (Haslinger et al. 2014; Liu et al. 2011) as well as for
66 protein labelling (La Clair et al. 2004; Shute et al. 2005; Yin et al. 2006).

67 The major limitation of this PCP loading strategy is the fact that the essential handle for Sfp function –
68 CoA – has limited solubility in solvents appropriate for chemical synthesis, which in turn makes the
69 synthesis of CoA linked substrates challenging – especially when the substrates are hydrophobic in
70 nature. In addition, CoA coupled substrates are not suitable for *in vivo* studies due to poor membrane
71 permeability. Thus, modified approaches to overcome these limitations have been developed. A less
72 commonly used approach is to use a mutant of the Sfp enzyme to allow substrate loading beyond CoA
73 coupled compounds. The R4-4 Sfp mutant developed by the Yin group was an important development
74 in the field, as it allows any substrate-linker combinations that are coupled to adenosine 5'-
75 diphosphate (ADP) to be loaded onto PCP domains (Figure 2) (Zou and Yin 2009). The advantage of
76 this approach is that ADP coupling removes any restrictions on the linker region between the substrate
77 and PCP domain. However, a more versatile route for loading carrier proteins is a chemo-enzymatic
78 route where substrates/probes are synthesized as D-pantetheine derivatives (Figure 2). These can be
79 converted into CoA derivatives by the use of three promiscuous enzymes from the CoA biosynthesis
80 pathway: pantothenate kinase (PanK or CoaA), phosphopantetheine adenylyltransferase (PPAT or
81 CoaD) and dephospho-CoA kinase (DPCK or CoaE) (Nazi et al. 2004). The derivatives thus formed
82 can then be loaded to PCP domains using Sfp. This route allows a wide range of compounds to be

83 loaded on to carrier proteins as demonstrated both *in vivo* (Clarke et al. 2005; Meier et al. 2006) and *in*
84 *vitro* (Hur et al. 2009; Meier et al. 2006; Reimer et al. 2016; van Wyk and Strauss 2007).

85 Although this enzymatic route to CoA-bound substrates is highly functional, several groups have
86 reported problems with the DPCCK enzyme, which may contribute to low functionality of the loading
87 cascade (Francois et al. 2006; Rootman et al. 2010; Strauss et al. 2010). To eliminate this problem, we
88 sought to combine the chemo-enzymatic route to CoA analogues together with the use of the R4-4
89 mutant of Sfp, as this would eliminate the need for DPCCK by virtue of the ability of the Sfp R4-4
90 mutant to load PCP domains with dephospho-CoA coupled substrates (Figure 2). We have explored
91 this strategy and in doing so have been able to establish a simplified loading route for PCP domains
92 that is fast, simple and removes the problematic DPCCK enzyme from the PCP loading process.

93 **Materials and methods**

94 **Cloning:** The gene encoding *ppat* from *E.coli* was obtained as a synthetic gene (ThermoFisher,
95 Genart[®]) and excised from the pEX-K plasmid using NdeI and HindIII restriction enzymes (New
96 England Biolabs). The gene encoding *panK* was amplified from *E. coli* strain DH10 β using whole cell
97 PCR including primers GGGAAATTCCATATGACCGCCAGAAACATGCTTATGAG (forward, NdeI
98 site underlined) and GCGGTAGAAGAGGTCAGACTACGCAAATAAAAGCTTGGAT (reverse,
99 HindIII site underlined) and GoTaq[®] Green Premix (Promega). Both genes were then subcloned into
100 pET-28(a) vector (Novagen), with inserts and vector cut using NdeI and HindIII-HF restriction
101 enzymes, ligated with T4 ligase. A synthetic gene codon optimized for expression in *Escherichia coli*
102 encoding *PCP6_{com}* from *Streptomyces lavendulae* (Chiu et al. 2001) was obtained from Eurofins
103 genomics. The *PCP6_{com}* gene was amplified by PCR using Phusion high fidelity polymerase (NEB)
104 and primers TATTACCCATGGCAGGCGGTCTGTGATCC (forward, NcoI site underlined) and
105 AATAACTCGAGACCAGTCTCGGGCAGGCTTGCTGCTTCTTCGGC (reverse, XhoI site
106 underlined). *PCP6_{com}* was subcloned into a modified pET vector (Bogomolovas et al. 2009) containing
107 a thioredoxin (Trx) solubility tag followed by a TEV cleavage site and a C-terminal His6-tag under the
108 control of a T7 promoter. All plasmids were used to transform BL21.DE3 Gold cells for protein

109 expression. Cloning of Trx-PCP_{2_{tei}} was performed as described previously (Kittila et al. 2017). The
110 plasmid encoding the Sfp mutant R4-4 was provided by the Yin group (Sunbul et al. 2009).

111 **Protein expression and purification:** The Sfp mutant R4-4 was prepared following a protocol
112 previously published by the Walsh laboratory (Yin et al. 2006); Trx-PCP_{2_{tei}} was also prepared as
113 described previously (Kittila et al. 2017). A single colony was used to inoculate 20-100 mL Luria-
114 Bertani (LB) medium supplemented with kanamycin (50 mg/L) and grown over night at 37°C. 4 L of
115 LB-medium was supplemented with kanamycin (50 mg/L) and inoculated with the overnight culture
116 (1 %). After the OD₆₀₀ reached 0.5-0.8 (37°C, 80 rpm), protein expression was induced by the addition
117 of 0.1 mM isopropyl β-D-1-thiogalactopyranoside. Cultures were then allowed to grow overnight
118 (18°C, 80 rpm), after which the cells were harvested by centrifugation (5500 g, 10 min, 4°C) and the
119 cell pellets resuspended in lysis buffer (50 mM Tris·HCl pH 7.4, 50 mM NaCl, 10 mM imidazole).
120 Resuspended cells were flash frozen and stored at -80°C prior to purification.

121 Cells and proteins were kept on ice or in a cold room throughout the purification; purification steps
122 performed using an Äkta system were performed at RT. Prior to cell lysis, protein inhibitor cocktail
123 tablets (Sigmafast™, EDTA-free, Sigma-Aldrich) were added to the thawed cells. Cells were lysed by
124 four passes through a microfluidizer (Microfluidics) and the lysate cleared by centrifugation (38800 g,
125 1 h, 4 °C). The cleared lysate was incubated with Ni-NTA beads (4 mL/construct, 1 h, 4°C, Macherey-
126 Nagel) with gentle shaking. Beads were washed twice with wash buffer (50 mM Tris·HCl, pH 7.4, 300
127 mM NaCl, 10 mM imidazole) and the bound protein eluted with elution buffer (50 mM Tris·HCl, pH
128 7.4, 300 mM NaCl, 300 mM imidazole); the elution fraction for PPAT was dialyzed against dialysis
129 buffer (50 mM Tris·HCl, pH 8.0, 150 mM NaCl, overnight). PanK and PPAT were further purified by
130 size exclusion chromatography using a Superose 12 column connected to an Äkta PURE system (GE
131 Healthcare Life Sciences, 50 mM Tris·HCl, pH 7.4, 150 mM NaCl, 2 mM MgCl₂). Protein purity was
132 assessed using 15 % SDS-PAGE, with selected fractions combined and concentrated before being
133 aliquoted, flash frozen and stored at -80°C. Protein concentration was determined by A₂₈₀ absorption
134 (Nanodrop 2 000c, ε_(PanK) = 45380 M⁻¹ cm⁻¹, ε_(PPAT) = 8480 M⁻¹ cm⁻¹, ε_(Trx-PCP_{6com}) = 15470 M⁻¹ cm⁻¹). The
135 identity of the purified proteins was confirmed by MALDI-TOF MS peptide mass fingerprinting.

136 **Activity assays:** Activity of PanK was determined using an NADH coupled colorimetric ADP assay
137 (see (Beinker et al. 2005)): 1 μ M PanK was mixed in assay buffer (50 mM Tris, pH 7.4, 20 mM KCl,
138 5 mM MgCl₂, 1 mM EDTA, 2 mM TCEP, 0.1 g/L BSA and 0.2 mM NADH, V_{tot} = 500 μ L) with 4 U
139 pyruvate kinase/ 6 U lactic dehydrogenase (Pyruvate kinase/ Lactic dehydrogenase from rabbit muscle
140 (P0294), Sigma-Aldrich) and 0.8 mM phosphoenolpyruvate (Sigma-Aldrich). The reaction was
141 allowed to equilibrate after which 0.5 mM D-pantetheine was added. The reaction was initiated by
142 addition of 1 mM ATP, with the decrease in NADH absorption at 340 nm monitored over time using a
143 Jasco V-650 spectrophotometer (Jasco).

144 The activity of PPAT was determined using an NADH coupled PP_i assay (Kittilä et al. 2016b): 2 μ M
145 PanK, 0.5 mM D-pantetheine, 2 mM ATP and the components for PP_i detection were mixed in 100
146 mM Tris-HCl, pH 7.4, 1 mM MgCl₂, 0.1 mM EDTA and incubated for 10 min at 25°C to allow the
147 formation of phosphopantetheine. 2 μ M PPAT was subsequently added and the decrease in NADH
148 absorption was observed spectroscopically.

149 Product formation was confirmed by LCMS measurements: briefly, 2.5 μ M PanK, 0.5 mM reduced D-
150 pantetheine and 2.5 mM ATP were mixed in assay buffer (50 mM Tris-HCl, pH 7.4, 20 mM KCl, 10
151 mM MgCl₂) and incubated for 30 min (37°C). A sample was taken and 25 μ M of PPAT was added
152 with 1 mM ATP to the remaining sample. The reaction was incubated for 30 min (37°C) and a further
153 sample was taken. From both samples the proteins were denatured at 95°C and the resultant
154 precipitate removed by centrifugation. The clarified supernatant was then analyzed by LCMS on a
155 Shimadzu 2020 system (column: XBridge BEH300 Prep C18 column, 10 μ m, 4.6 x 250 mm, Waters;
156 gradient: 2 % AcN for 4 min, 2-10 % AcN in 11 min, 10-40 % AcN in 10 min; flow rate: 1 mL/min)
157 using positive and negative ionization modes. Product formation was then determined using extracted
158 ion chromatograms.

159 **PCP loading (conventional method):** 120 μ M Trx-PCP_{6,com}, 240 μ M CoA or dephospho-CoA and 50
160 nM Sfp mutant R4-4 were mixed in loading buffer (50 mM Tris, pH 7.4, 15 mM MgCl₂, 2 mM DTE);
161 the reaction was incubated for 2 h at 30°C and quenched by the addition of 20 mM EDTA. PCP
162 loading efficiency was determined using MALDI-TOF-MS (intact mass determination, Axima

163 Confidence or Performance, Shimadzu). For examining loading inhibition, 5 mM ATP/ADP/AMP, 4
164 μM PanK or 5 μM PPAT were added to the loading reaction.

165 **PCP loading (one-pot loading reaction):** 120 μM Trx-PCP_{6_{com}} or Trx-PCP_{2_{tei}}, 2.5 μM PanK and 240
166 μM reduced D-pantetheine, panthenol or pantothenic acid were mixed in assay buffer (25 mM
167 Tris-HCl, pH 7.4, 5 mM MgCl₂, 0.5 mM DTE). The reaction was initiated by the addition of 0.5 mM
168 ATP. After 5-15 min incubation (RT), 20 μM PPAT was added and reaction incubated for a further
169 10-30 min (RT). Next, 0.4 U/ μL of alkaline phosphatase (from calf intestine, grade I (Ref
170 10108146001), Roche) was added and reaction incubated for an additional 10 min (RT). The linker
171 was then enzymatically loaded onto the relevant PCP domain by the addition of 0.5 μM Sfp R4-4
172 mutant (2 h, RT). The reaction was quenched by addition of 20 mM EDTA and the loading efficiency
173 analyzed by MALDI-TOF-MS (intact mass determination).

174 **Substrate loading onto PCP domain:** 100 μM Trx-PCP_{2_{tei}} was loaded with 300 μM (D)-tyrosine-4-
175 hydroxyphenylglycine-pantetheine (for details of the synthesis see Kittilä et al. 2017) as described
176 above, but with the incubation times reduced to protect the thioester bond from hydrolysis (PanK 5
177 min, PPAT 10 min, CIP 10 min, Sfp R4-4 15 min). Loaded Trx-PCP_{2_{tei}} was washed four times with
178 50 mM Tris-HCl (pH 7.4) (Amicon® ultra ultracentrifugal filters, MWCO 10 000, 0.5 mL, Merck
179 Millipore) to remove excess peptide. Dipeptide was cleaved from Trx-PCP_{2_{tei}} with excess of
180 methylamine and sample neutralized with formic acid. Precipitate was removed via centrifugation and
181 supernatant analyzed with UPLC-MS (LCMS-8050, Shimadzu) with single ion monitoring (Column:
182 Acquity UPLC Peptide BEH C18, 1.7 μM , 2.1 x 100 mm, Waters; Gradient: 5 % acetonitrile for 0.5
183 min, 5-20 % acetonitrile in 20 min; flow rate 0.5 mL/min). A reaction without Trx-PCP_{2_{tei}} (negative
184 control) was handled with the same protocol.

185 **Results and discussion**

186 In our initial experiments, we tested the activity of the PanK and PPAT enzymes prior to optimizing
187 the PCP loading assay. Both enzymes were expressed in *Escherichia coli* and purified in a simple two-
188 step procedure to generate high yields of both purified enzymes (PanK = 0.7 mmol/L, PPAT = 3.8

189 mmol/L). The enzymatic activities of PanK and PPAT were then assessed using spectroscopic NADH
190 coupled assays (Beinker et al. 2005; Kittilä et al. 2016b). Both enzymes were shown to be active under
191 the assay conditions and product formation was confirmed using LCMS (Figure 3), with the
192 phosphorylated pantetheine needed as a substrate for PPAT produced enzymatically from D-
193 pantetheine by PanK.

194 After ensuring the production of dephospho-CoA from D-pantetheine, the one-pot PCP-loading
195 reaction was explored. As an acceptor protein, a PCP domain of the NRPS machinery from
196 complestatin biosynthesis (Trx-PCP_{6com} module 6, expressed as a thioredoxin fusion protein to
197 improve yield) was utilized (Chiu et al. 2001). The functionality of the loading reaction was
198 determined by measuring intact mass using MALDI-TOF ($\Delta_{m/z} = 358$ for phosphopantetheine loaded
199 versus the unloaded PCP domain). Surprisingly no loading of the PCP domain was detected in the
200 one-pot loading reaction, although it has been reported that R4-4 mutant of Sfp is able to load the
201 PPE-linker to PCP domains using dephospho-CoA as a substrate (Sunbul et al. 2009). After reviewing
202 the substrate binding pocket of Sfp (Mofid et al. 2004), it appeared plausible that the mutation to allow
203 the wider substrate selectivity of the R4-4 mutant enzyme abolished its ability to distinguish between
204 CoA and phosphoadenylates. If this were the case, then ATP, ADP and AMP could potentially bind to
205 Sfp, preventing phosphopantetheine loading on the PCP domain. To test potential Sfp inhibition, we
206 assessed how the different components of the one-pot loading reaction (ATP, ADP, AMP PanK,
207 PPAT) affect the conventional loading reaction, where the linker is derived from CoA. The
208 components were added individually to the loading reaction although they are not needed for linker
209 loading from CoA. No PPE-loading was detected in the presence of phosphoadenylates, whilst the
210 addition of PanK or PPAT did not influence the loading reaction (Figure 4). Therefore, the hypothesis
211 of phosphoadenylates inhibiting Sfp mutant R4-4 seemed to be the most likely explanation for the lack
212 of loading from our initial one-pot reaction.

213 To overcome this unwanted inhibition, we introduced alkaline phosphatase from calf-intestinal (CIP)
214 into the loading reaction. CIP is commonly used to dephosphorylate DNA as it cleaves phosphates
215 from the 5'end of DNA. Addition of CIP to the PCP-loading reaction was therefore expected to

216 dephosphorylate the problematic phosphoadenylylates but not the desired dephospho-CoA product. This
217 strategy proved to be successful and loading of PCP domains could now be achieved (Figure 5). As all
218 of the loading enzymes (PanK, PPAT, Sfp) have been reported to be promiscuous, a wide variety of
219 pantetheine analogs should be able to be loaded onto PCP domains. To test this, the one-pot loading
220 reaction was also tested using D-pantothenic acid and D-panthenol to provide the linker regions for
221 PCP loading. Both compounds were converted to dephospho-CoA analogs and loaded successfully on
222 the PCP domain using our established one-pot loading method (Figure 5).

223 To demonstrate the functionality of this loading strategy towards different PCP domains and that the
224 loading strategy can also be used for substrate loading, loading of a PCP domain derived from the
225 teicoplanin producing NRPS machinery (Trx-PCP_{2_{tei}}, module 2 (Kittila et al. 2017)) was studied. Trx-
226 PCP_{2_{tei}} was first successfully loaded using the one-pot loading strategy with the linker derived from
227 pantetheine (Figure 6A). Next, we successfully loaded linked substrates – pantetheine dipeptide
228 thioesters – onto the PCP domain using our established method (Figure 6B). Furthermore, we have
229 been able to employ these loaded substrates to study catalytic steps in synthesis of glycopeptide
230 antibiotics (Kittila et al. 2017).

231 In conclusion, the R4-4 mutant of Sfp enables loading of PCP domains in a one-pot loading reaction
232 after dephosphorylation of phosphoadenylylates. The main advantage of this new PCP loading method is
233 that the phosphorylation of dephospho-CoA catalyzed by DPCK is not needed to achieve PCP loading.
234 The reaction catalyzed by DPCK is reversible (Strauss et al. 2010) and removal of the pyrophosphate
235 formed in the step catalyzed by PPAT could well be required for efficient product formation when
236 DPCK is present. In addition, the activity of DPCK has been reported to reduce over time (Rootman et
237 al. 2010) and some groups have reported problems in achieving high activity through protein
238 purification (Francois et al. 2006). Therefore, removing the need for this step is likely to prove
239 advantageous in loading substrates onto PCP domains for future studies on NRPS systems, with the
240 protocol described herein proving to be a simple and robust method to achieve this essential
241 modification of PCP domains.

242 **Acknowledgements**

243 The authors are grateful to G. Stier (BZH-Heidelberg) for the thioredoxin fusion protein vector; to J.
244 Yin (Georgia State University) for the R4-4 Sfp expression plasmid; to Melanie Müller and Marion
245 Gradl (MPIMF-Heidelberg) for mass spectral analysis; to Melanie Schoppet (MPIMF-Heidelberg/
246 Monash University) for dipeptide thioester substrate synthesis; and to Alexa Weinmann and Veronika
247 Ulrich (MPIMF-Heidelberg) for assistance with construct cloning and protein preparation. T.K. is
248 grateful for the support of the Deutsche Akademischer Austausch Dienst (DAAD Graduate School
249 Scholarship Program); M.J.C. is grateful for the support of the Deutsche Forschungsgemeinschaft
250 (Emmy-Noether Program, CR 392/1-1), Monash University and the EMBL Australia program.

251

252 **References**

- 253 Al Toma, R.S., Brieke, C., Cryle, M.J., and Süßmuth, R.D. 2015. Structural aspects of
254 phenylglycines, their biosynthesis and occurrence in peptide natural products. *Nat Prod Rep* 32, 1207-
255 1235.
- 256 Beinker, P., Schlee, S., Auvula, R., and Reinstein, J. 2005. Biochemical Coupling of the Two
257 Nucleotide Binding Domains of ClpB: covalent linkage is not a prerequisite for chaperon activity
258 *Journal of Biological Chemistry* 280, 37965-37973.
- 259 Beld, J., Sonnenschein, E.C., Vickery, C.R., Noel, J.P., and Burkart, M.D. 2014. The
260 phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Natural*
261 *Product Reports* 31, 61-108.
- 262 Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. 1999. Aminoacyl-CoAs as probes of condensation
263 domain selectivity in nonribosomal peptide synthesis. *Science* 284, 486-489.
- 264 Bogomolovas, J., Simon, B., Sattler, M., and Stier, G. 2009. Screening of fusion partners for high
265 yield expression and purification of bioactive viscotoxins. *Protein Expression and Purification* 64, 16-
266 23.
- 267 Chiu, H.-T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T., and Khosla, C. 2001.
268 Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proceedings of*
269 *the National Academy of Sciences* 98, 8548-8553.
- 270 Clarke, K.M., Mercer, A.C., La Clair, J.J., and Burkart, M.D. 2005. In vivo reporter labeling of
271 proteins via metabolic delivery of coenzyme A analogues. *Journal of the American Chemical Society*
272 127, 11234-11235.
- 273 Ehmann, D.E., Shaw-Reid, C.A., Losey, H.C., and Walsh, C.T. 2000a. The EntF and EntE adenylation
274 domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP
275 transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences* 97,
276 2509-2514.
- 277 Ehmann, D.E., Trauger, J.W., Stachelhaus, T., and Walsh, C.T. 2000b. Aminoacyl-SNACs as small-
278 molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol* 7,
279 765.

280 Francois, J.A., Starks, C.M., Sivanuntakorn, S., Jiang, H., Ransome, A.E., Nam, J.W., Constantine,
281 C.Z., and Kappock, T.J. 2006. Structure of a NADH-insensitive hexameric citrate synthase that resists
282 acid inactivation. *Biochemistry* 45, 13487-13499.

283 Haslinger, K., Brieke, C., Uhlmann, S., Sieverling, L., Süßmuth, R.D., and Cryle, M.J. 2014. The
284 Structure of a Transient Complex of a Nonribosomal Peptide Synthetase and a Cytochrome P450
285 Monooxygenase. *Angew Chem, Int Ed* 53, 8518-8522.

286 Henderson, J.C., Fage, C.D., Cannon, J.R., Brodbelt, J.S., Keatinge-Clay, A.T., and Trent, M.S. 2014.
287 Antimicrobial Peptide Resistance of *Vibrio cholerae* Results from an LPS Modification Pathway
288 Related to Nonribosomal Peptide Synthetases. *ACS Chemical Biology* 9, 2382-2392.

289 Horsman, M.E., Hari, T.P.A., and Boddy, C.N. 2016. Polyketide synthase and non-ribosomal peptide
290 synthetase thioesterase selectivity: logic gate or a victim of fate? *Natural Product Reports* 33, 183-202.

291 Hur, G.H., Meier, J.L., Baskin, J., Codelli, J.A., Bertozzi, C.R., Marahiel, M.A., and Burkart, M.D.
292 2009. Crosslinking Studies of Protein-Protein Interactions in Nonribosomal Peptide Biosynthesis.
293 *Chemistry & Biology* 16, 372-381.

294 Hur, G.H., Vickery, C.R., and Burkart, M.D. 2012. Explorations of catalytic domains in non-
295 ribosomal peptide synthetase enzymology. *Natural Product Reports* 29, 1074-1098.

296 Kittila, T., Kittel, C., Tailhades, J., Butz, D., Schoppet, M., Buttner, A., Goode, R.J.A., Schittenhelm,
297 R.B., van Pee, K.-H., Süßmuth, R.D., *et al.* 2017. Halogenation of glycopeptide antibiotics occurs at
298 the amino acid level during non-ribosomal peptide synthesis. *Chemical Science* 8, 5992-6004.

299 Kittilä, T., Mollo, A., Charkoudian, L.K., and Cryle, M.J. 2016a. New Structural Data Reveal the
300 Motion of Carrier Proteins in Nonribosomal Peptide Synthesis. *Angewandte Chemie International*
301 *Edition* 55, 9834-9840.

302 Kittilä, T., Schoppet, M., and Cryle, M.J. 2016b. Online Pyrophosphate Assay for Analyzing
303 Adenylation Domains of Nonribosomal Peptide Synthetases. *ChemBioChem* 17, 576-584.

304 La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. 2004. Manipulation of
305 carrier proteins in antibiotic biosynthesis. *Chemistry & Biology* 11, 195-201.

306 Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R.,
307 Khosla, C., and Walsh, C.T. 1996. A new enzyme superfamily - the phosphopantetheinyl transferases.
308 *Chem Biol* 3, 923-936.

309 Liu, Y., Zheng, T., and Bruner, Steven D. 2011. Structural basis for phosphopantetheinyl carrier
310 domain interactions in the terminal module of nonribosomal peptide synthetases. *Chemistry &*
311 *Biology* 18, 1482-1488.

312 Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A., and Walsh, C.T. 2002. Timing of
313 epimerization and condensation reactions in nonribosomal peptide assembly lines: Kinetic analysis of
314 phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* 41, 9184-9196.

315 Meier, J.L., Mercer, A.C., Rivera, H., and Burkart, M.D. 2006. Synthesis and Evaluation of
316 Bioorthogonal Pantetheine Analogues for in Vivo Protein Modification. *Journal of the American*
317 *Chemical Society* 128, 12174-12184.

318 Mitchell, C.A., Shi, C., Aldrich, C.C., and Gulick, A.M. 2012. Structure of PA1221, a Nonribosomal
319 Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* 51,
320 3252-3263.

321 Mofid, M.R., Finking, R., Essen, L.O., and Marahiel, M.A. 2004. Structure-Based Mutational
322 Analysis of the 4'-Phosphopantetheinyl Transferases Sfp from *Bacillus subtilis*: Carrier Protein
323 Recognition and Reaction Mechanism^{†,‡}. *Biochemistry* 43, 4128-4136.

324 Nazi, I., Koteva, K.P., and Wright, G.D. 2004. One-pot chemoenzymatic preparation of coenzyme A
325 analogues. *Anal Biochem* 324, 100-105.

326 Payne, J.A.E., Schoppet, M., Hansen, M.H., and Cryle, M.J. 2017. Diversity of nature's assembly lines
327 - recent discoveries in non-ribosomal peptide synthesis. *Molecular BioSystems* 13, 9-22.

328 Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. 1998.
329 Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein
330 domains in peptide synthetases. *Biochemistry* 37, 1585-1595.

331 Reimer, J.M., Aloise, M.N., Harrison, P.M., and Martin Schmeing, T. 2016. Synthetic cycle of the
332 initiation module of a formylating nonribosomal peptide synthetase. *Nature* 529, 239-242.

333 Roche, E.D., and Walsh, C.T. 2003. Dissection of the EntF condensation domain boundary and active
334 site residues in nonribosomal peptide synthesis. *Biochemistry* *42*, 1334-1344.

335 Rootman, I., de Villiers, M., Brand, L.A., and Strauss, E. 2010. Creating Cellulose-Binding Domain
336 Fusions of the Coenzyme A Biosynthetic Enzymes to Enable Reactor-Based Biotransformations.
337 *ChemCatChem* *2*, 1239-1251.

338 Shute, T.S., Matsushita, M., Dickerson, T.J., La Clair, J.J., Janda, K.D., and Burkart, M.D. 2005. A
339 site-specific bifunctional protein labeling system for affinity and fluorescent analysis. *Bioconjug*
340 *Chem* *16*, 1352-1355.

341 Strauss, E., de Villiers, M., and Rootman, I. 2010. Biocatalytic Production of Coenzyme A Analogues.
342 *ChemCatChem* *2*, 929-937.

343 Sunbul, M., Marshall, N.J., Zou, Y., Zhang, K., and Yin, J. 2009. Catalytic Turnover-Based Phage
344 Selection for Engineering the Substrate Specificity of Sfp Phosphopantetheinyl Transferase. *Journal of*
345 *Molecular Biology* *387*, 883-898.

346 Süssmuth, R.D., and Mainz, A. 2017. Nonribosomal Peptide Synthesis—Principles and Prospects.
347 *Angewandte Chemie International Edition* *56*, 3770-3821.

348 van Wyk, M., and Strauss, E. 2007. One-pot preparation of coenzyme A analogues via an improved
349 chemo-enzymatic synthesis of pre-CoA thioester synthons. *Chemical Communications*, 398-400.

350 Villiers, B.R.M., and Hollfelder, F. 2009. Mapping the Limits of Substrate Specificity of the
351 Adenylation Domain of TycA. *ChemBioChem* *10*, 671-682.

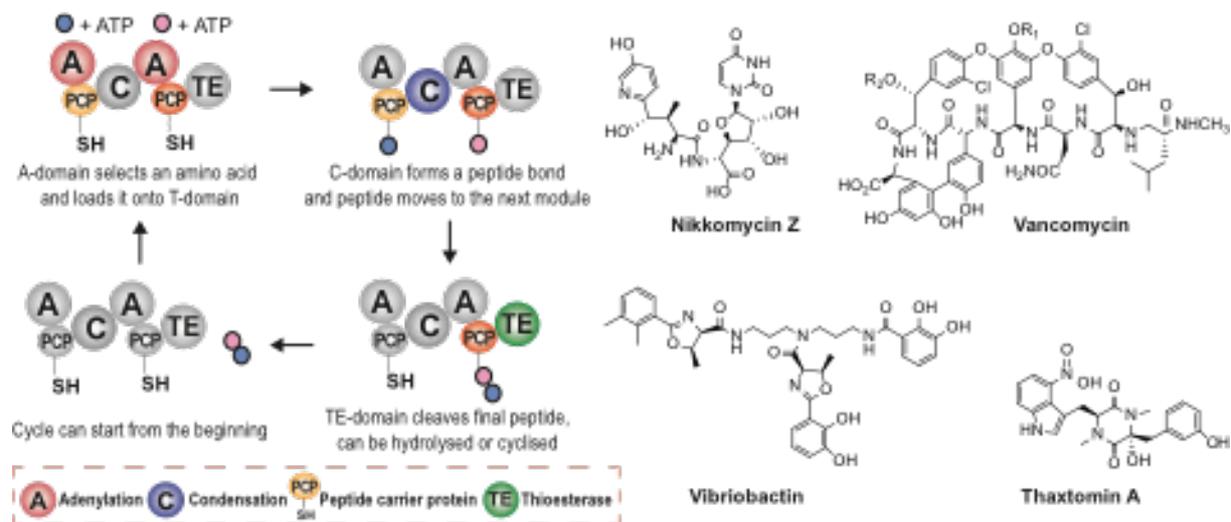
352 Yin, J., Lin, A.J., Golan, D.E., and Walsh, C.T. 2006. Site-specific protein labeling by Sfp
353 phosphopantetheinyl transferase. *Nat Protocols* *1*, 280-285.

354 Zou, Y., and Yin, J. 2009. Phosphopantetheinyl transferase catalyzed site-specific protein labeling
355 with ADP conjugated chemical probes. *Journal of the American Chemical Society* *131*, 7548-7549.

356

357

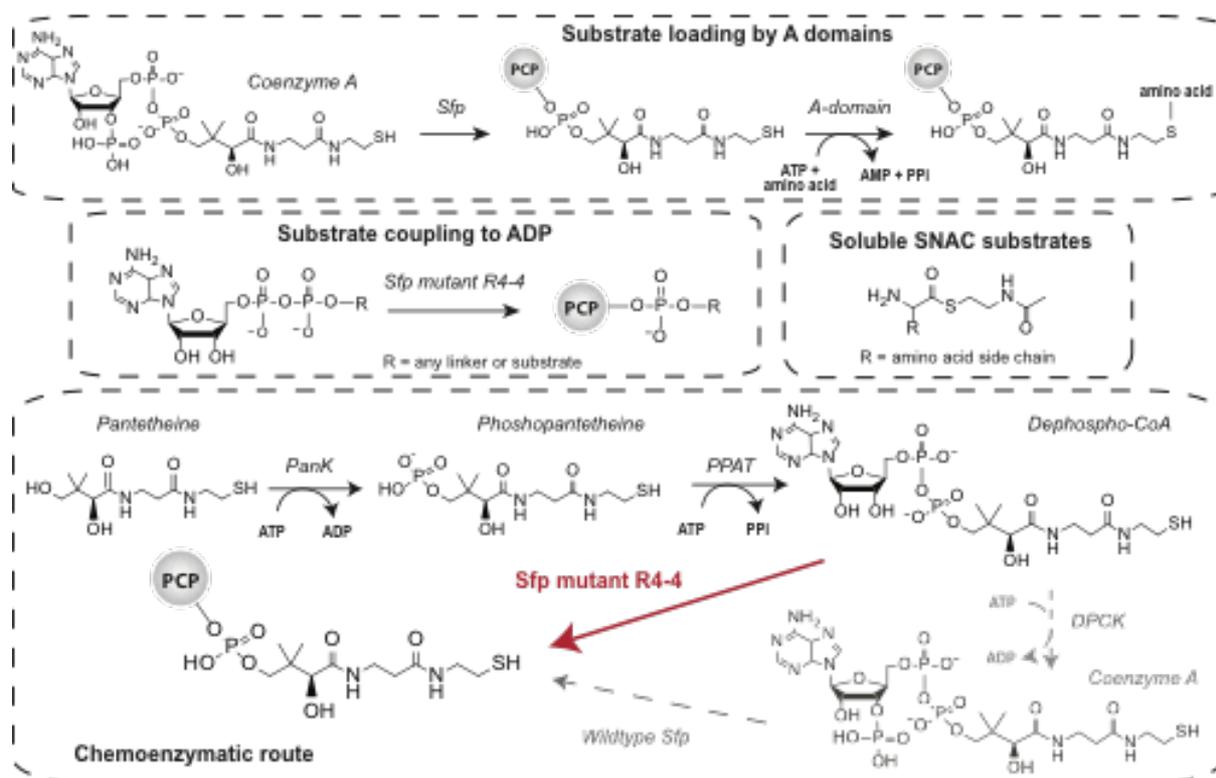
358 **Figures**



359

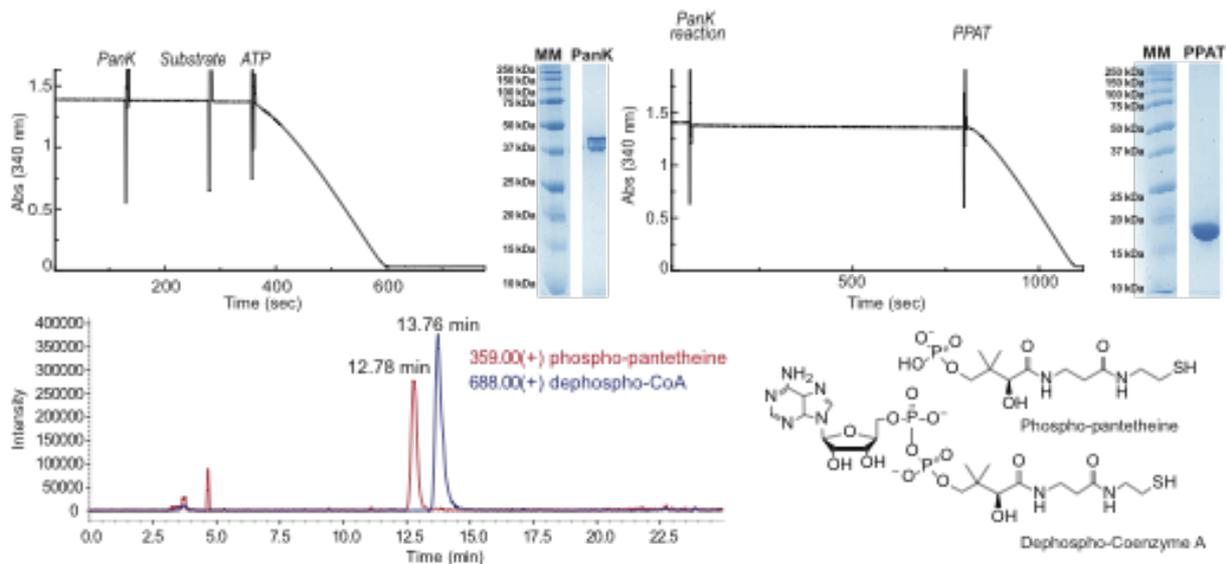
360 **Figure 1.** Non-ribosomal peptide synthesis and examples of important peptide natural products
 361 produced by NRPS machineries. NRPS synthesis utilizes a modular architecture that is based upon
 362 repeating catalytic domains. Initial amino acid selection and activation is an ATP-dependent process
 363 that is performed by adenylation (A) domains. The activated amino acid is then transferred onto a
 364 peptidyl carrier protein (PCP) domain, which serve as attachment points for all intermediates during
 365 NRPS-mediated biosynthesis. Peptide bond formation is catalyzed by condensation (C) domains,
 366 which bind two PCP domains and transfer the upstream substrate onto the amino acid bound on the
 367 downstream PCP-domain with concomitant formation of a new peptide bond.

368



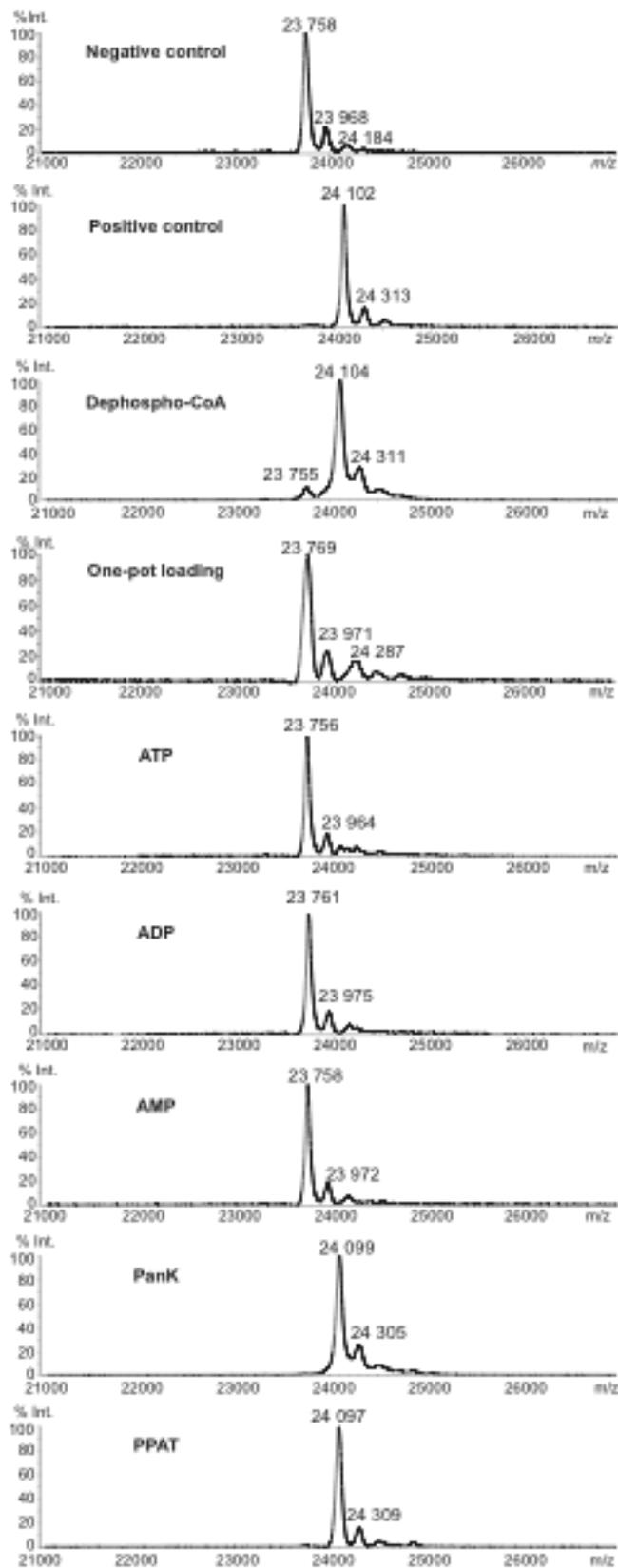
369

370 **Figure 2.** Loading substrates onto PCP domains. Loading substrates onto PCP domains is one of the
 371 biggest challenges in NRPS research and therefore several approaches to achieve loaded PCP domains
 372 have been established. A domains can be used to load substrates to PCP domains *in vitro*. First, the
 373 linker region needs to be attached in a reaction catalyzed by the phosphopantetheine transferase
 374 enzyme Sfp before an A domain can attach the substrate to the free thiol group of the linker.
 375 Substrates can also be coupled to coenzyme A chemically and Sfp can be used to load them to PCP
 376 domains. A mutant Sfp enzyme (R4-4) loads substrates coupled to ADP, which allows loading of
 377 varied linker regions in addition to different substrates. The majority of NRPS domains only accept
 378 PCP-bound substrates, but in some cases soluble SNAC-substrates can also be accepted. However,
 379 acceptance of such derivatives must be tested for each system separately and their use removes
 380 important protein-protein interactions between NRPS domains. Substrates can also be coupled to
 381 pantetheine, which can in turn be converted to coenzyme A using three enzymatic steps (chemo-
 382 enzymatic route). In this work, we explored an improved loading route in which a chemo-enzymatic
 383 route is combined with the use of Sfp mutant R4-4. This route is shown with a red arrow, with the
 384 steps being bypassed shown in grey. PanK = pantothenate kinase, PPAT = phosphopantetheine
 385 adenylyltransferase, DPKC = dephospho-CoA kinase.



386

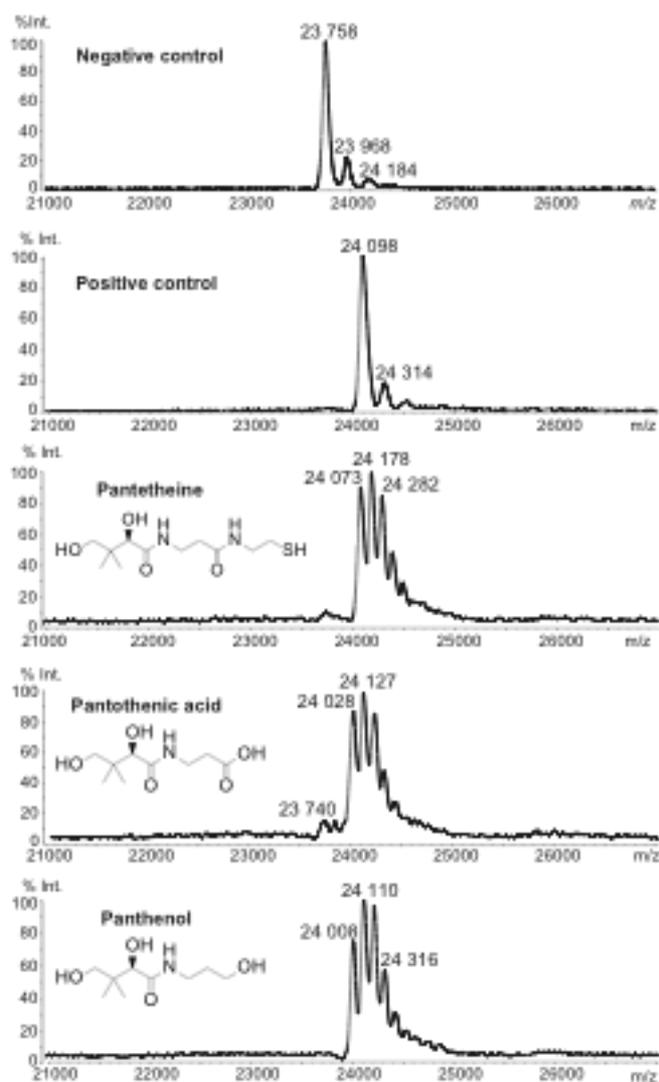
387 **Figure 3.** PanK and PPAT activity. Both enzymes were initially purified to homogeneity (PanK = 39.3
 388 kDa, PPAT = 20.3 kDa), with enzymatic activities analyzed using NADH coupled spectroscopic
 389 assays. The PanK reaction was initiated by addition of ATP and subsequent production of ADP was
 390 detected. PPAT adenylates phosphopantetheine that is produced by PanK: thus, a PanK reaction was
 391 incubated in assay buffer with components to detect PP_i formation. As PanK does not produce PP_i , a
 392 steady baseline was detected prior to PPAT addition. Product formation was confirmed by LCMS
 393 measurement, with traces showing a combined reaction with PanK and PPAT. No remaining starting
 394 material (D-pantetheine) was detected.



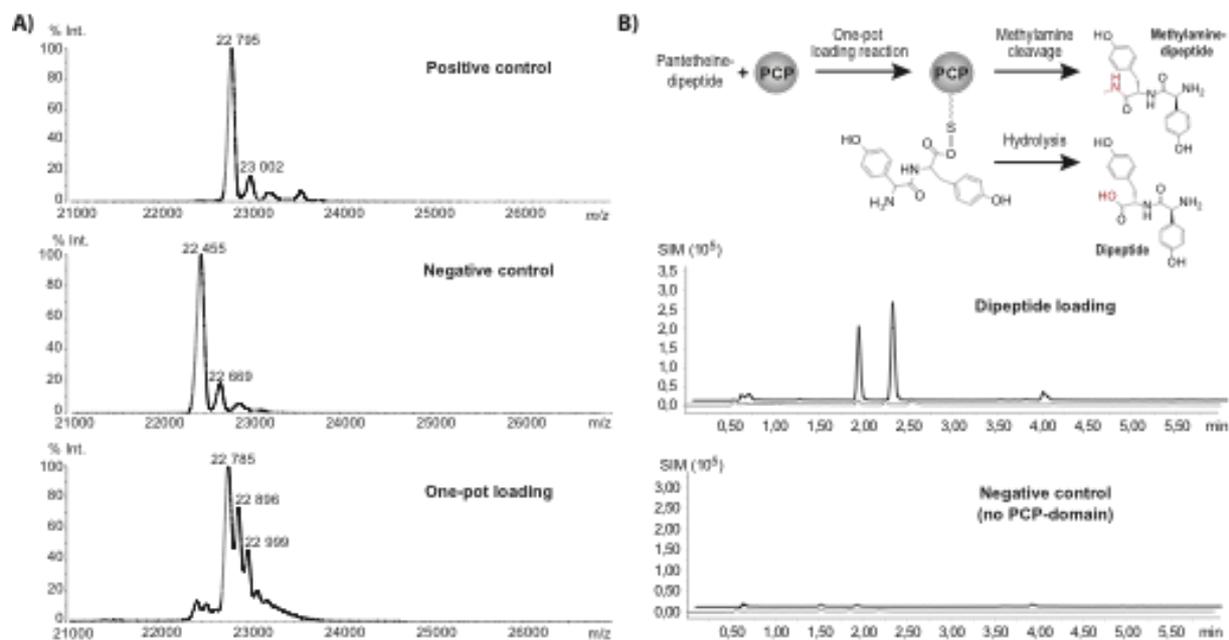
395

396 **Figure 4.** Inhibition of the one-pot PCP loading reaction. The Sfp mutant R4-4 can load Trx-PCP_{6com}
 397 domain using CoA (positive control) or dephospho-CoA – however, the initial one-pot loading
 398 reaction tested did not afford successful PCP loading despite the demonstrated functionality of all

399 enzymes. Inhibition of Sfp R4-4 was studied by addition of phosphoadenylates or other enzymes from
400 the loading cascade into a loading reaction utilizing CoA. No loading was observed in the presence of
401 phosphoadenylates, whilst the additional enzymes did not affect loading. Trx-PCP_{6com} = 23750 Da (-
402 Met); PPE-loaded Trx-PCP_{6com} = 24108 Da (-Met).



403
404 **Figure 5.** One-pot loading reaction including calf intestine alkaline phosphatase (CIP). Addition of
405 CIP to the one-pot loading reaction allowed different linkers to be successfully loaded onto Trx-
406 PCP_{6com}. CIP was stored in 3.2 M ammonium sulfate leading to formation of sulfate adducts (sulfate =
407 96 g/mol); linker loading from CoA by Sfp R4-4 was used as a positive control. Trx-PCP_{6com} = 23750
408 Da (-Met), loading with CoA/pantetheine = 24108 Da (-Met), loading with pantothenic acid = 24049
409 Da (-Met), loading with panthenol = 24035 Da (-Met).



410

411 **Figure 6.** Loading of Trx-PCP_{2_{tei}}. A) A one-pot loading reaction was used to load Trx-PCP_{2_{tei}} with
 412 linker derived from pantetheine to ensure the functionality of the loading cascade towards different
 413 PCP domains. Trx-PCP_{2_{tei}} = 22475 Da (-Met), loaded Trx-PCP_{2_{tei}} = 22833 Da (-Met). B) Trx-PCP_{2_{tei}}
 414 was loaded with dipeptide coupled to a pantetheine linker using the established one-pot loading
 415 cascade. After the loading reaction, the substrate was cleaved from the linker via addition of
 416 methylamine and analyzed via UPLC-MS. Two peaks were observed because of dipeptide
 417 racemization during synthesis. Hydrolyzed dipeptide ((D)-tyrosine-4-hydroxyphenylglycine, 331
 418 g/mol) trace shown in grey, methylamine cleaved dipeptide (344 g/mol) shown in black.

419