Widening the bottleneck: Heterologous expression, purification, and characterization of the *Ktedonobacter racemifer* minimal type II polyketide synthase in *Escherichia coli*

Joshua G. Klein, Yang Wu, Bashkim Kokona,* Louise K. Charkoudian*

*to whom all correspondences should be addressed: bkokona@haverford.edu, lcharkou@haverford.edu

Abstract

Enzyme assemblies such as type II polyketide synthases (PKSs) produce a wide array of bioactive secondary metabolites. While the molecules produced by type II PKSs have found remarkable clinical success, the biosynthetic prowess of these enzymes has been stymied by 1) the inability to reconstitute the bioactivity of the minimal PKS enzymes *in vitro* and 2) limited exploration of type II PKSs from diverse phyla. To begin filling this unmet need, we expressed, purified, and characterized the ketosynthase chain length factor (KS-CLF) and acyl carrier protein (ACP) from *Ktedonobacter racemifer* (*Kr*). Using *E. coli* as a heterologous host, we obtained soluble proteins in titers signifying improvements over previous KS-CLF heterologous expression efforts. Characterization of these enzymes reveals that *Kr*ACP has self-malonylating activity. Sedimentation velocity analytical ultracentrifugation (SV-AUC) analysis of *holo*-*Kr*ACP and *Kr*KS-CLF indicates that these enzymes do not interact *in vitro*, suggesting that the acylated state of these proteins might play an important role in facilitating biosynthetically relevant interactions. These results lay important groundwork for optimizing the interaction between *Kr*KS-CLF and *Kr*ACP and exploring the biosynthetic potential of other non-actinomycete type II PKSs.

Introduction

Microorganisms produce secondary metabolites, which are small organic molecules that are not essential to the organism's growth and reproduction but rather confer an evolutionary advantage. These molecules, often referred to as "natural products," have been repurposed as pharmaceutical agents and biological probes. Over 60% of all drugs approved between 1981 and 2014 originated from natural products and their synthetic derivatives, with notable examples including erythromycin (antibiotic), lovastatin (cholesterol medication), rapamycin (immunosuppressant), doxorubicin (anticancer agent) and vancomycin (antibiotic).[1–3] Polyketides represent a remarkably successful class of natural products. Molecules in this class are manufactured by multi-enzyme pathways called polyketide synthases (PKSs), which are encoded by biosynthetic gene clusters (BGCs). The chemical diversity programmed by polyketide BGCs has been expanded through engineering efforts, as shown by the alteration of type I PKSs via precursor-directed biosynthesis, substrate specificity engineering, and combinatorial biosynthesis.[4,5] By comparison, progress towards diversifying type II PKSs, which are comprised of discrete proteins that act iteratively to manufacture polyaromatic polyketides,

has been limited. This is in part due to a scarcity of *in vitro* studies on these systems due to difficulties in obtaining soluble biosynthetic proteins for structural and functional characterization. Nonetheless, type II PKSs remain a particularly attractive source of new chemical diversity because of their outstanding track record in producing pharmaceutically-relevant molecules.[1,6,7]

The core proteins involved in type II polyketide biosynthesis include the heterodimeric ketosynthase chain length factor (KS-CLF) and the acyl carrier protein (ACP) (**Fig 1**). These enzymes act collaboratively and iteratively to produce a nascent polyketide chain. The reactive β-keto chain is converted into a structurally complex molecule through the action of tailoring enzymes, including cyclases and ketoreductases, giving rise to the final branching, oxidation state, and cyclization pattern of the polyaromatic product. The remarkable chemical diversity observed in this class of molecules does not originate from monomer diversity; rather, it results from variations in chain length and tailoring reactions. In principle, molecules of novel structure and bioactivity can be created by altering the identity of the KS-CLF (chain length) and tailoring enzymes (oxidation state, cyclization pattern), or by varying the KSIII (starter units). However, fundamental hurdles remain that limit our ability to engineer type II PKSs to make molecules of novel structure and function. For example, we have a limited understanding of how type II PKSs efficiently channel polyketide intermediates to the appropriate enzymes.[8] In addition, unnatural protein-protein and protein-substrate interactions are often insufficient to generate catalytically efficient synthases.[9,10] Engineering efforts are further hampered by the large and growing deficit between genomic data and functionally annotated enzymes.[5]

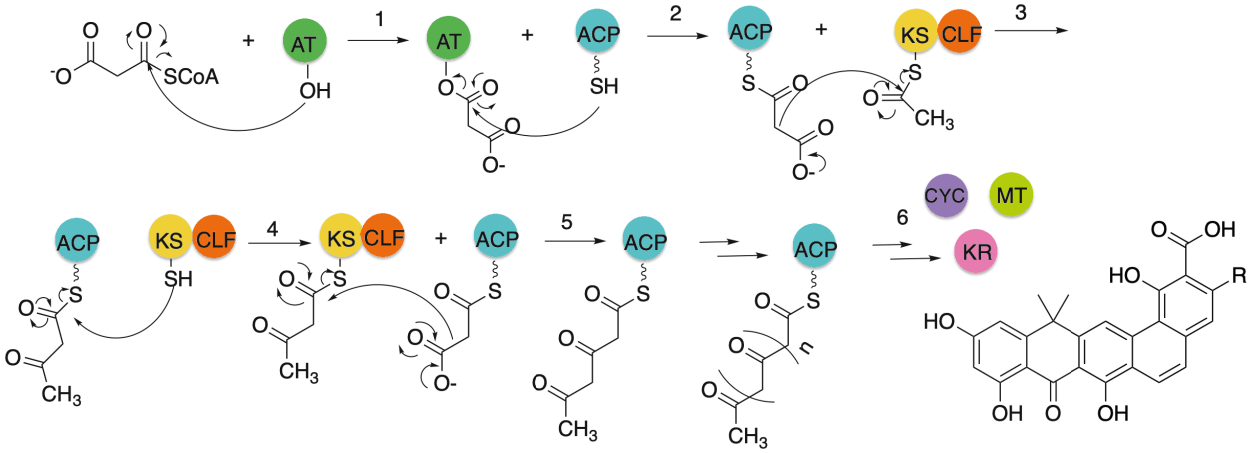

Fig 1. Overview of the role of the KS-CLF and ACP in type II polyketide biosynthesis. 1. If necessary, an acyl transferase (AT) reacts with malonyl-coenzyme A (malonyl-CoA) via a nucleophilic acyl substitution reaction to malonylate the AT. 2. The terminal thiol group on the ACP phosphopantetheine arm (represented as a squiggly line) reacts via a nucleophilic acyl substitution reaction to malonylate the ACP. If the ACP is "self-malonylating", the terminal thiol can directly load the malonyl-CoA. 3. Malonylated-ACP reacts with the acylated active thiol on the KS-CLF via a decarboxylative Claisen-like condensation reaction to initiate the formation of the polyketide chain on the ACP. 4. The polyketide chain is transferred to the KS-CLF via an acyl substitution reaction. 5. The β-keto-bound KS-CLF reacts with malonylated ACP via a decarboxylative Claisen-like condensation to further elongate the polyketide chain; this process repeats until the polyketide chain is elongated to the programmed chain length, which in part is directed by the steric cavity of the KS-CLF.[11] 6. Tailoring enzymes such as cyclases (CYC), methyltransferases (MT), and ketoreductases (KR) act on the elongated polyketide chain to form the final natural product.

In an effort to study the evolution of BGCs on a cluster-wide scale, we previously created a catalog of genes that encode putative type II PKS enzymes.[11] This work, led by Hillenmeyer in

collaboration with our research team, unveiled that the enormous structural and biological diversity of type II polyketide antibiotics originated from primary metabolic fatty acid synthases (FASs).[11] In our study, we noted that all characterized polyketide BGCs (sequenced gene clusters whose secondary metabolite products had been structurally characterized as of 2015) were from actinobacteria. We also revealed that the divergence of type II PKSs from FASs predated major speciation events, suggesting that type II PKSs diverged from FASs well before the actinobacterial phylum had formed. This led us to recognize the clade of uncharacterized, anciently diverged type II PKS BGCs from non-actinobacterial species as unexplored space opportune for the mining of novel molecules and enzymes.

We were particularly excited about the possibility of characterizing KS-CLFs from non-actinobacteria. This heterodimer plays an essential role in the manufacturing of type II PKSs yet has long evaded robust expression and purification methods. The challenges posed by the lack of a robust heterologous expression platform for the type II PKSs has hampered progress towards understanding and engineering these powerful biosynthetic systems.[12] We hypothesized that the ancient type II PKS BGCs harbored in non-actinobacterial species might be amenable to expression in *E. coli* because they share a closer evolutionarily history to *E. coli* FAS BGCs. Recent groundbreaking work by Takano and co-workers confirmed that these non-actinobacterial type II PKS gene clusters are indeed ripe for bioprospecting.[13] The researchers showed that type II PKS enzymes from non-actinomycetes can be expressed and effectively manufacture polyketides in *E. coli*. Moreover, the authors showed that various tailoring enzymes can be substituted into the reconstituted type II PKS assembly to alter the final polyketide. These findings provide an exciting platform for future synthetic biology efforts and inspire us to pivot our research focus towards gaining a molecular understanding of the structure and function of core type II PKS enzymes.

Herein, we present initial results on the characterization of type II PKS biosynthetic enzymes from the non-actinobacterium *Ktedonobacter racemifer* DSM 44963. Isolated from Italian forest soil, *K. racemifer* is a gram-positive bacterium belonging to the Chloroflexi phylum[14] that harbors a putative type II PKS BGC belonging to the ancient resistomycin clade.[11] We viewed its evolutionary proximity to the *E. coli* FAS as a harbinger that the core enzymes might be amenable to expression in *E. coli*. We report optimized expression conditions for *Kr*ACP and *Kr*KS-CLF, the *K. racemifer* minimal PKS components, in *E. coli*. Successful expression of *Kr*ACP and *Kr*KS-CLF enabled their biophysical characterization and set the stage for future investigations into their biocatalytic potential.

Results

Expression and purification of *K. racemifer* minimal PKS enzymes

The *K. racemifer* minimal PKS enzymes, *Kr*ACP and *Kr*KS-CLF, were successfully expressed and purified from the *E. coli* BAP1[15] and BL21 competent cell lines, respectively (**Fig 2**). Titers of 16 mg/L for *Kr*ACP and 3 mg/L for *Kr*KS-CLF were obtained from purification. Expression of *Kr*ACP in the BAP1 cell line, which harbors the broad substrate phosphopantetheinyl transferase from *B. subtilis* Sfp, afforded ~50 % of the ACP with the phosphopantetheine (Ppant) arm installed ("*holo-Kr*ACP") and ~50 % *apo-Kr*ACP. (**Fig S1, S2**). The mixture of *holo-Kr*ACP and *apo-Kr*ACP was confirmed by high-performance liquid chromatography (HPLC) analysis and is supported by literature characterization of other ACPs (**Fig S3**).[16,17] Presence of *holo-Kr*ACP was also confirmed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis, as the ~27 kDa band corresponding to a disulfide bond-linked *holo-Kr*ACP dimer is observed under non-reducing conditions and not in the presence of 5% v/v β-mercaptoethanol (BME; **Fig 2A**).

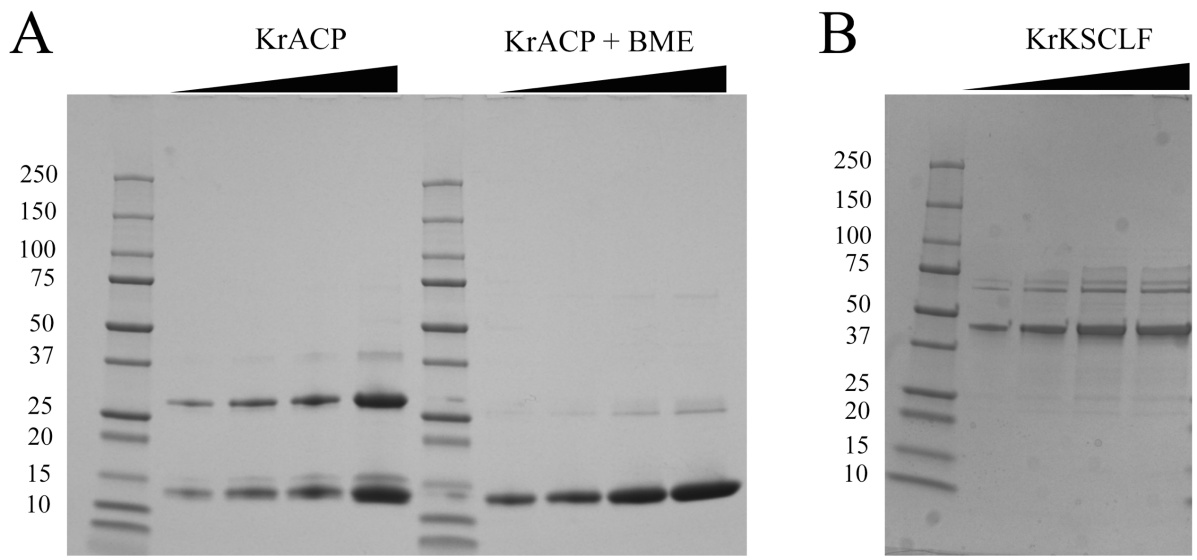

Fig 2. Characterization of *K. racemifer* minimal PKS enzymes by SDS-PAGE. A. *Kr*ACP concentration gradients run both with and without the presence of 5% v/v BME. Lanes: 1. Protein standards ladder; 2. *Kr*ACP (31 μM, 5 μL) run under non-reducing conditions; 3. *Kr*ACP (31 μM, 10 μL) run under non-reducing conditions; 4. *Kr*ACP (31 μM, 15 μL) run under non-reducing conditions; 5. *Kr*ACP (31 μM, 20 μL) run under non-reducing conditions; 6. Protein standards ladder; 7. *Kr*ACP (31 μM, 5 μL) run under reducing conditions; 8. *Kr*ACP (31 μM, 10 μL) run under reducing conditions; 9. *Kr*ACP (31 μM, 15 μL) run under reducing conditions; 10. *Kr*ACP (31 μM, 20 μL) run under reducing conditions. The disappearance of the ~27 kDa band upon the addition of reducing agent indicates that this band represents the *holo-Kr*ACP disulfide dimer. B. *Kr*KS-CLF purified under reducing conditions. 1. Protein standards ladder; 2. *Kr*KS-CLF (5 μL, 2.6 μM); 3. *Kr*KS-CLF (10 μL, 2.6 μM); 4. *Kr*KS-CLF (15 μL, 2.6 μM); 5. *Kr*KS-CLF (20 μL, 2.6 μM). The ~47 kDa band was excised and confirmed to contain *Kr*KS and *Kr*CLF monomers via tandem proteolysis mass spectrometry (Figures S4). Purification under reducing conditions was necessary to avoid aggregation and purification of the nickel column elute via anion exchange was required to remove ~25 kDa *E. coli* protein contaminants (see Results and Supporting Information for details).

Far UV-Circular Dichroism spectroscopy (CD) was used to determine the secondary structure of *Kr*ACP (**Fig 3A**). Similar to other ACPs studied in our laboratory, solvated *Kr*ACP is mostly helical with two negative peaks at 222 nm and 208 nm and a positive peak at 192 nm (**Fig 3A**).[18] Pre- and post-thermal denaturation spectra showed that unfolding is reversible with greater than 80% recovery in signal following denaturation (**Fig 3A**). Thermal unfolding appeared to be a cooperative two-state process without observable intermediates with a midpoint $T_m = 66.7 \pm 1.8$ °C, enthalpy of unfolding $\Delta H_m = 118.7 \pm 13.4$ kJ mol^{-1}, and change in entropy $\Delta S_m = 1.8 \pm 0.2$ kJ (K^{-1} mol^{-1}) (**Fig 3A inset**). Sedimentation velocity experiments with the analytical ultracentrifuge (SV-AUC) indicates that *Kr*ACP sediments mostly as a monomer with a molecular weight of ~12 kDa (theoretical molecular weight of *holo Kr*ACP is 12.2 kDa; **Fig 3B**). The sedimentation of *Kr*ACP in primarily monomeric form is consistent with what has been observed for other type II PKS ACPs.[18]

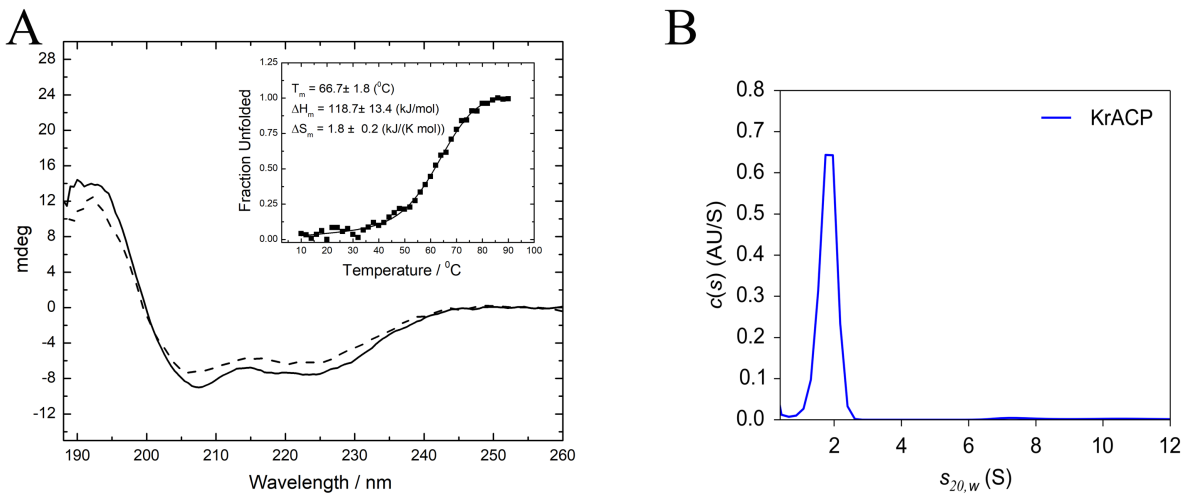

Fig 3. Characterization of *Kr*ACP. A. Circular dichroism (CD) spectra and melting temperature experiments of *Kr*ACP. The CD spectrum of purified *Kr*ACP is given by the solid line, while the CD spectrum of *Kr*ACP after melting temperature experiments is given by the dashed line. Melting temperature curve measured at 222 nm and subsequent extracted thermodynamic parameters are given. The similarity in pre- and post-heating CD spectra indicate that *Kr*ACP folds to regain nearly all of its secondary structure after heating to 90 °C. B. SV-AUC profile of *Kr*ACP under reducing conditions. *Kr*ACP shows as a single species in SV-AUC with a molecular weight of ~12 kDa, indicating that the protein is a monomer in solution and that any disulfide bonds between *holo-Kr*ACP monomers are reduced through the presence of dithiothreitol (1 mM DTT) in the buffer (50 mM sodium phosphate, pH 7.6, 150 mM NaCl).

Initial *Kr*KS-CLF purification attempts under non-reducing conditions led to protein aggregation. The addition of BME to the samples reduced aggregation, suggesting that aggregate formation was at least partially caused by nonspecific cysteine interactions between *Kr*KS-CLF monomers. As such, the expression protocol was optimized to include the addition of reducing agent in all purification and storage buffers. SDS-PAGE gel electrophoresis of the reducing buffer-purified *Kr*KS-CLF revealed three distinct bands at ~ 85 kDa, 47 kDa, and 25 kDa (**Fig S4).** The highest molecular weight (MW) band, corresponding to a mass of approximately 85 kDa, was attributed to the *Kr*KS-CLF heterodimer. The ~47 kDa band and ~25 kDa bands were excised for analysis by tandem-proteolysis mass spectrometry. The ~47 kDa band was determined to be composed of both the *Kr*KS and *Kr*CLF monomers (**Fig S5**). The ~25 kDa band was partially composed of the *Kr*KS and *Kr*CLF monomers, but primarily composed of

three proteins native to *E. coli*: FKBP-type peptidyl prolytl cis-trans isomerase SlyD (NCBI accession number P0A9K9), hydroxyethylthiazole kinase ThiM (NCBI accession number P76423), and triosephosphate isomerase TpiA (NCBI accession number P0A858) (**Fig S6**). This ~25 kDa species was also observed in the SV-AUC profile of *Kr*KS-CLF sample prior to further purification of the sample under reducing conditions by size exclusion chromatography (SEC; **Figs S4 and S7)**. Concentration dependence SV-AUC revealed that the ratio of *Kr*KS-CLF to 25 kDa proteins remained unchanged regardless of the concentration of sample loaded. These findings indicate that the identified *E. coli* proteins co-purified with *Kr*KS-CLF and co-precipitated with the *Kr*KS-CLF heterodimer to the extent that we were not able to separate via SEC (in absence of salt and reducing agent). Inclusion of 150 mM NaCl and 1 mM DTT in the SEC buffer made it possible to purify *Kr*KS-CLF away from the *E. coli* contaminants. Subsequent analysis of the purified *Kr*KS-CLF heterodimer by CD indicates that the secondary structure is comprised of both β-sheets and α-helices (**Fig 4A**). Whereas *Kr*KS and *Kr*CLF run as a monomer-dimer equilibrium via SDS PAGE (**Fig 2**), the proteins sediment as the *Kr*KS-CLF heterodimer in solution (**Fig 4B**).

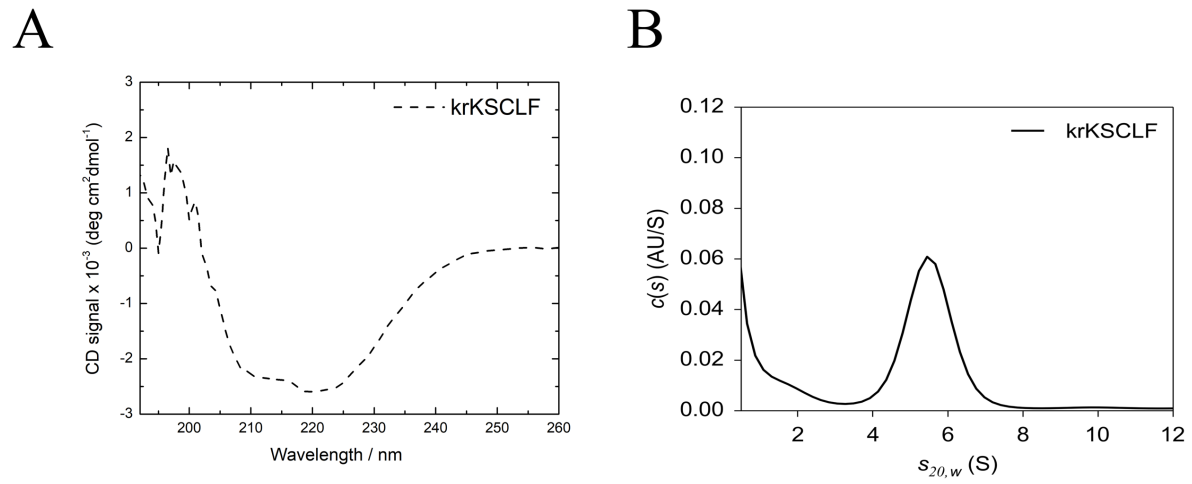

Fig 4. Characterization of *Kr*KS-CLF. A. The CD spectrum of purified *Kr*KS-CLF indicates that the heterodimer is comprised of both β-sheets and α-helices. B. SV-AUC analysis of purified *Kr*KS-CLF. The SV-AUC profile of purified *Kr*KS-CLF reveals the presence of a major species of ~98.7 kDa, which is roughly consistent with the molecular weight of the *Kr*KS-CLF heterodimer (theoretical molecular weight of 92.6 kDa)

Malonylation of *Kr*ACP

To initiate biosynthesis, the Ppant arm of *holo*-ACP must be primed with an initial malonyl extender unit (**Fig 1**). For type II PKSs studied *in vitro* or via heterologous expression in *E. coli*, the malonylation of *holo*-ACP is often facilitated by a dedicated AT within the BGC or by the malonyl transferase endogenous to the *E. coli* host system, FabD.[19–21] However, some ACPs have previously been reported to malonylate without the presence of a malonyl transferase.[22] The ability of *Kr*ACP to undergo this so-called self-malonylation was assessed by characterizing the products of *Kr*ACP and malonyl-CoA in the presence and absence of FabD. The reaction solutions were purified by SEC, and the fractions corresponding to *Kr*ACP were collected for analysis by liquid chromatography-mass spectrometry (LCMS) (**Fig S8**). LCMS

samples were analyzed using a modified Ppant ejection assay.[23] Both reaction solutions analyzed by this Ppant ejection assay revealed species with *m/z* values of 347.1 Da, corresponding to the cyclized malonyl-Ppant arm (**Fig 5**). The presence of malonylated *Kr*ACP in both samples was additionally confirmed by the deconvolution of higher *m/z* species corresponding to the molecular weight of malonylated *Kr*ACP (**Fig S9**). Taken together, these data suggest that *Kr*ACP can load malonyl-CoA without the presence of a transferase partner. The physiological relevance of this self-malonylation mechanism has yet to be determined.

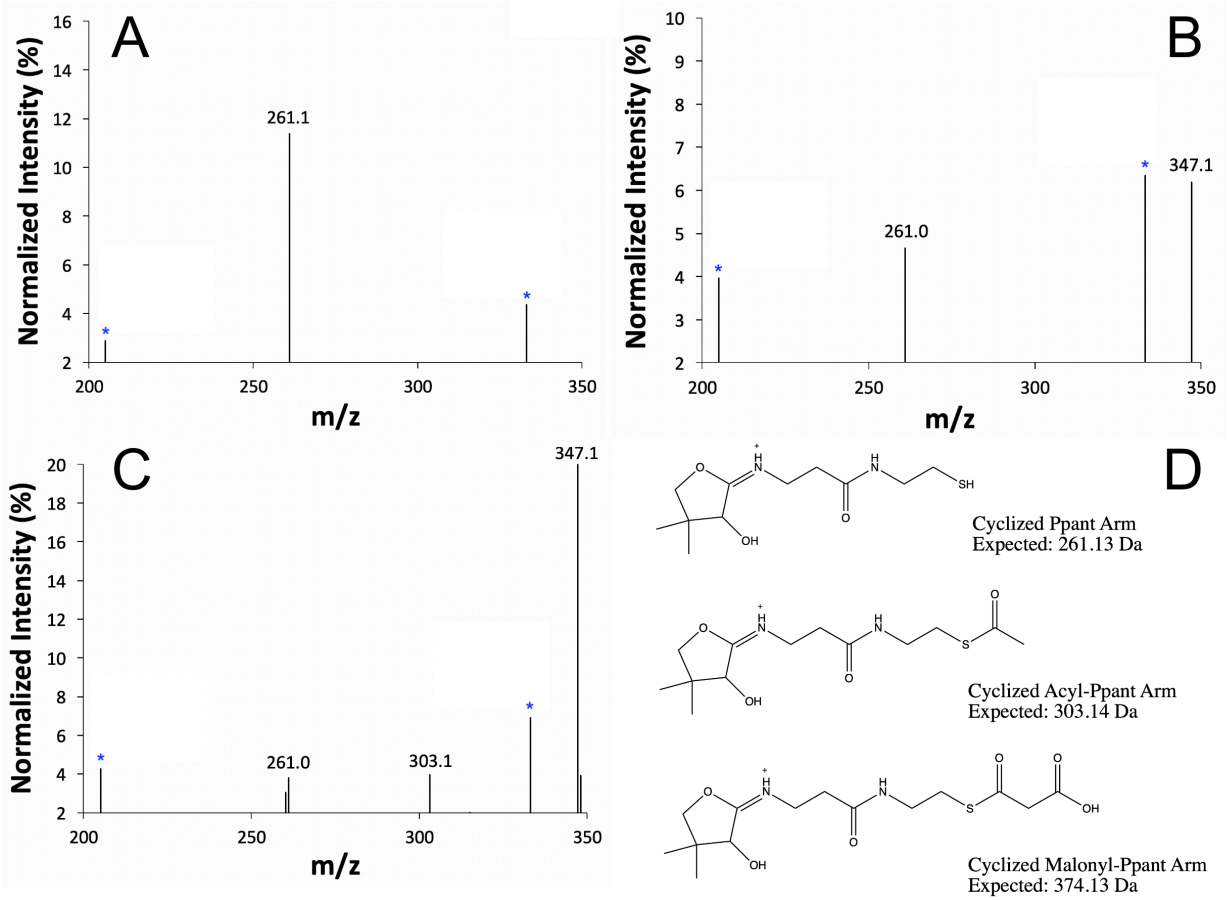

Fig 5. *Kr*ACP can load malonyl-CoA in absence of an acyl transferase. A. Ppant ejection of *Kr*ACP shows the presence of *holo-Kr*ACP, represented by the 261.1 Da species corresponding to the cyclized Ppant arm. B. Ppant ejection of *Kr*ACP incubated with malonyl-CoA shows the appearance of a 347.1 Da species corresponding to the cyclized, malonylated Ppant arm, indicating that *holo-Kr*ACP can load malonyl-CoA in the absence of a transferase. C. Ppant ejection of *Kr*ACP with malonyl-CoA and FabD shows the presence of both 261.1 Da and 347.1 Da peaks, with the addition of a 303.1 Da peak corresponding to a cyclized acyl-Ppant arm, which forms upon the decarboxylation of malonyl- *Kr*ACP. D. Expected species in mass spectra resulting from cyclization of *Kr*ACP Ppant arm. Peaks marked by a blue asterisk were present in all spectra, including buffer, and likely represent background noise apparent at high electron multiplier voltage.

Lack of observed *holo-Kr*ACP/*Kr*KS-CLF interaction *in vitro*

To investigate the interactions between *holo-Kr*ACP and *Kr*KS-CLF *in vitro*, both enzymes were studied by SV-AUC under reducing conditions. As noted above, *Kr*ACP

sediments as a single species corresponding to the *Kr*ACP monomer (**Fig 3B**) and the *Kr*KS-CLF as a heterodimer (**Fig 4B**). A combined sample consisting of both *Kr*ACP (100 µM) and *Kr*KS-CLF (4 µM) showed no significant shift in the higher MW peak corresponding to the *Kr*KS-CLF heterodimer (**Fig 6**), suggesting a lack of interaction between *Kr*KS-CLF and *holo-Kr*ACP under the conditions studied.

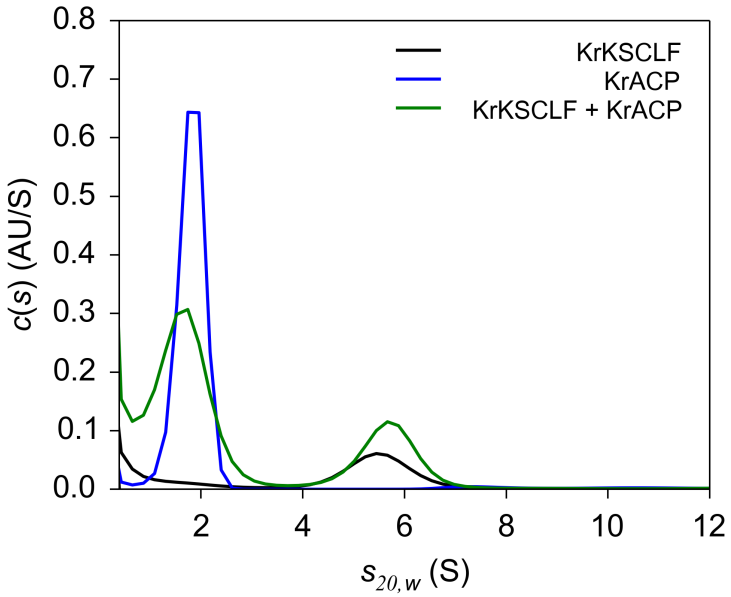

Fig 6. SV-AUC analysis of *K. racemifer* minimal PKS enzymes suggests a lack of interaction between *Kr*KS-CLF and *Kr*ACP *in vitro*. SV-AUC analysis of minimal PKS enzymes showed no significant shift in the ~98.7 kDa peak corresponding to *Kr*KS-CLF upon the addition of *Kr*ACP, indicating a lack of interaction between the enzymes under the investigated conditions. Conditions: 2.9 µM *Kr*KS-CLF, 140 µM *Kr*ACP, and 4 µM *Kr*KS-CLF combined with 110 µM *Kr*ACP. Samples were in 50 mM sodium phosphate, pH 7.2, 150 mM NaCl, and 1 mM DTT.

It is possible that the apparent lack of interplay between *holo-Kr*ACP and *Kr*KS-CLF is due to the inability of SV-AUC to capture the transient and weak interactions under the conditions studied. A small binding interface, held together by a few polar residues, was recently reported between the *Photorhabdus luminescens* anthraquinone KS-CLF and ACP,[21] and could be at play in the *K. racemifer* system as well. It is also possible that structural features of *Kr*KS-CLF limit its interactions with *Kr*ACP. To further explore *Kr*KS-CLF via protein modeling, the *Kr*KS-CLF amino acid sequence was threaded onto the solved actKS-CLF from *S. coelicolor* (**Fig 7**).[24] The 17 Å amphipathic polyketide tunnel between Phe116 and Cys169 in actKS-CLF has been shown to play a vital role in the elongation of the polyketide chain, with widening of the tunnel by site-directed mutagenesis corresponding to an increase in polyketide chain length.[24] The threaded structure of *Kr*KS-CLF shows the presence of a similar polyketide tunnel between residues Leu125 and Cys172; however, the width of this modelled *Kr*KS-CLF amphipathic polyketide tunnel measures 22.5 Å. It is currently unclear if this predicted difference in tunnel shape affects how *Kr*KS-CLF binds to *Kr*ACP. A recent structure of the highly reducing type II PKS KS-CLF Iga11-Iga12 enables us to draw some additional comparisons (**Fig S10**). First, both the polyketide tunnel size and charge of residues lining the cavity of Iga11-Iga12 were implicated in directing the length and structure of the manufactured product.[25] In the Iga system, β-ketoacylation-ACP elongation is inhibited by Asp113, which does not appear to be conserved in *Kr*KS-CLF nor in other aromatic polyketide-synthesizing type II PKSs.[24,25] However, the

hydrogen bonding amino acids lining the IgaPKS KS-CLF tunnel that stabilize the ACP Ppant arm (Thr303 and Thr305) appear to be conserved in *Kr*KS-CLF. Second, the catalytic triad (Cys170-His301-His336) of Iga11 shifted upon acylation and ACP binding.[25] These comparisons support the possibility that *Kr*KS-CLF can accommodate the biosynthesis of a polyaromatic polyketide, but that the acylation state of *Kr*KS-CLF and/or *Kr*ACP could play a pivotal role in ACP-KS-CLF binding. This result is consistent with the *P. luminescens* anthraquinone KS-CLF, which was reported to require significant rearrangements to accommodate an octaketide[21] and highlights that the acylation state of the heterodimer can influence its structure. The importance of enzyme acylation state in carrier protein interactions has also been observed in other biosynthetic systems, including type I PKSs[26] and non-ribosomal peptide synthetases (NRPSs).[27]

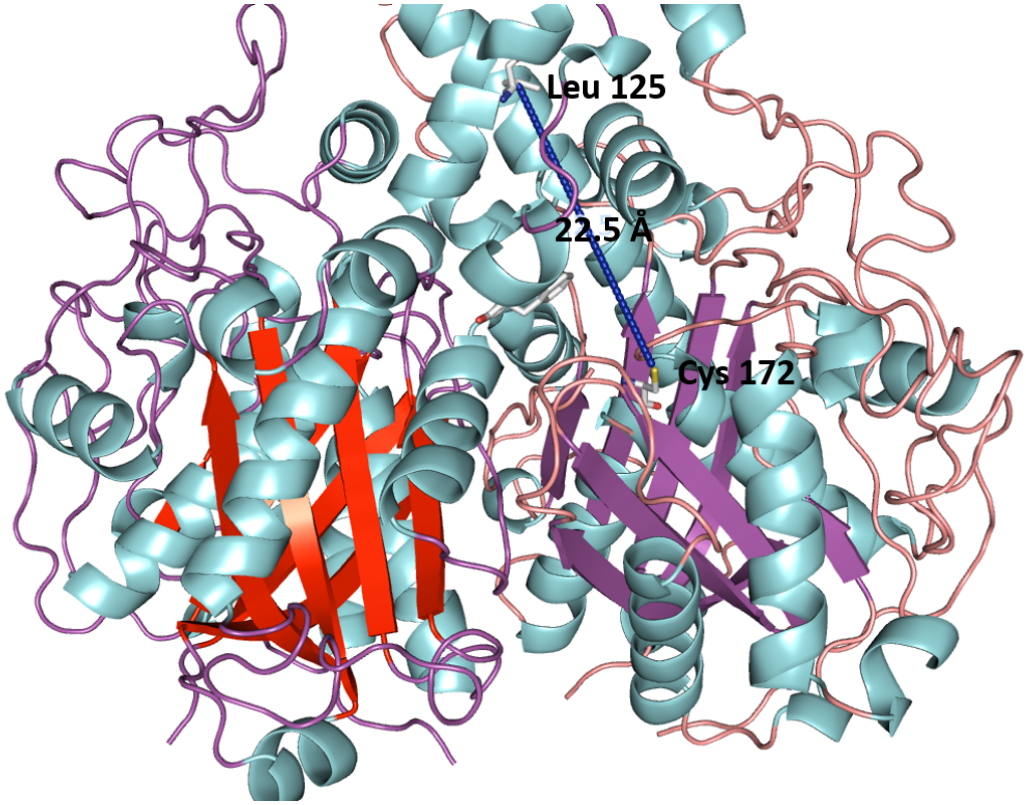

Fig 7. The *Kr*KS-CLF sequence threaded onto the solved actKS-CLF crystal structure (PDB 1TQY).[24] The threaded *Kr*KS-CLF structure reveals the heterodimer might harbor a similar polyketide tunnel to that of the previously studied in actKS-CLF. However, the length of the proposed *Kr*KS-CLF polyketide tunnel is 22.5 Å in comparison to the 17 Å polyketide tunnel of actKS-CLF.[24] It is possible that the length of the *Kr*KS-CLF polyketide tunnel and/or the acylation state of *Kr*KS-CLF affects its ability to interact with *holo-Kr*ACP.

Discussion

The inability to heterologously express KS-CLF enzymes in *E. coli* has been a longstanding bottleneck in the chemical engineering of type II PKSs. Recent developments have shown the ability of yeast and *E. coli* to serve as heterologous hosts for total type II PKS gene cluster reconstitution; however, the biochemical properties of the heterologously expressed

minimal PKS enzymes from these systems have not been investigated.[13,28,29] The heterologous expression of the *K. racemifer* minimal PKS enzymes *Kr*ACP and *Kr*KS-CLF in *E. coli* is a viable route to the molecular-level study of minimal PKS enzymes from type II PKSs. Titers of 3 mg/L obtained from heterologous expression of *Kr*KS-CLF in the *E. coli* BL21 cell line are a marked improvement over the heterologous expression of other KS-CLFs in *S. coelicolor* CH999, which often affords less than 1 mg/L of purified protein from a much more resource-intensive process.[24,30]

A workflow to gain access to high quantities of purified type II PKS enzymes opens the door to important molecular-level biochemical studies that can complement ongoing *in vivo* characterization of non-actinomycete type II PKS pathways. For example, recent pioneering work by the Takano Lab revealed that ancient orphaned type II PKS gene clusters from non-actinomycetes can be harnessed to synthesize type II polyketide upon reconstitution in *E. coli*.[13] The *in vitro* biochemical studies of various ancient orphaned ACPs and KS-CLFs, including *Kr*ACP and *Kr*KS-CLF, complement this work by enabling access to the structural causes behind ACP/KS-CLF compatibility. Information such as protein sequence, secondary structure, oligomeric state, and acylated state can be pooled to create a catalog of ACPs and KS-CLFs that in turn could be used to predict compatibility and design hybrid synthases capable of making novel chemical diversity.

While we were able to heterologously express and purify both components of the *K. racemifer* minimal PKS in *E. coli*, we did not observe any interaction between *Kr*ACP and *Kr*KS-CLF *in vitro*. The recent crystal structure of the *Photorhabdus luminescens* anthraquinone ACP-KS-CLF highlights the small binding interface between ACP and KS-CLF, facilitated by only a few polar contacts.[21] It is likely that *Kr*KS-CLF and *Kr*ACP also interact in a weak and transient manner and that these interactions could not be detected at the low concentrations of proteins used in our SV-AUC studies. It is also possible that the size and charge of the amphipathic polyketide tunnel of *Kr*KS-CLF might play a role. The ability to obtain soluble *Kr*KS-CLF enables further investigation of this possibility through site-directed mutagenesis experiments coupled with biophysical studies. Moreover, the role of acylation state in facilitating KS-CLF – ACP interactions can be studied through pairing the chemoenzymatic acylation of the KS-CLF and/or ACP with biophysical studies using previously established methods.[26,27]

Notwithstanding the lack of observed *Kr*ACP/*Kr*KS-CLF interactions, the results presented herein lay the foundation for further investigation of *Kr*ACP and *Kr*KS-CLF as potential components of a hybrid PKS. Additionally, the methods described herein can be applied to explore other non-actinomycete type II PKS enzymes. The ability of *Kr*ACP to self-malonylate may represent an advantage of calling upon this ACP in chemical engineering efforts as it could participate in biosynthesis in the absence of a native malonyl transferase. Future ACPs heterologously expressed from non-actinomycete strains can be similarly tested for self-malonylation activity; subsequent sequence alignments between ACPs with and without the ability to self-malonylate could elucidate the structural motifs necessary for self-malonylation. It will be important to conduct kinetic analyses on enzyme-catalyzed versus self-malonylated PKSs because the rate of polyketide production in reconstituted PKSs has been shown to effectively mirror the rate of ACP malonylation.[31,32]

Taken together, our results support that non-actinomycete type II PKS enzymes can be robustly expressed and purified in *E. coli*. Access to these purified enzymes can be leveraged to increase knowledge of the structural determinants behind ACP/KS-CLF compatibility. A large catalog of such information can enable more accurate prediction of compatibility between both characterized and uncharacterized type II PKS ACPs and KS-CLFs that have been identified through sequence analyses.

Materials and Methods

Construction of expression plasmids

The expression plasmid for the *E. coli* FAS AT FabD was provided as a gift from the Campopiano Research Group at University of Edinburgh. *Ktedonobacter racemifer* genomic DNA was purchased from the DSMZ-German Collection of Microorganisms and Cell Cultures (DSM 44963). Plasmids pYW2 and p*Kr*ACP (encoding the expression of *Ktedonobacter racemifer* KS-CLF and ACP, respectively) were constructed by inserting amplified genes encoding the relevant enzymes into the NdeI/EcoRI restriction cut sites of pET28a vector via Gibson Assembly for pYW2 and restriction digestion / ligation for p*Kr*ACP. See Supporting Information for the protein sequences of FabD, *Kr*ACP and *Kr*KS-CLF.

The gene sequence encoding the expression of *Kr*KS-CLF was amplified from the gDNA using the following primers (NdeI and EcoRI endonuclease cut sites are underlined): 5' CCTGGTGCCGCGCGGCAGC<u>CATATG</u>CGCCGTGTCGTTATCTCTG 3' and 5' AGCTTGTCGACGGAGCTC<u>GAATTC</u>CTAGGCCCACGTGCGCAGT 3'. Thermal cycling conditions were as follows (40 ng DNA, Phusion polymerase): 1. 98 °C for 1 min, 2. 98 °C for 10 sec, 3. 72 °C for 2 min, 4. Repeat steps 2-3 a total of 34 times, 5. 72 °C for 10 min, 6. 4 °C forever. Amplified DNA was purified using the Zymoclean Recovery Kit (Zymo Research) and cloned into the NdeI and EcoRI endonuclease cut sites of linearized pET28a vector DNA via Gibson Assembly as follows. The insert and vector DNA were combined in a 3:1 insert:vector molar ratio in a volume of 5 µL and added 15 µL of Gibson Assembly Master Mix (New England BioLabs), consisting of T5 Exonuclease, Phusion Polymerase, Taq Ligase, dNTPs and $MgCl_2$ in Tris-HCl buffer. The resulting mixture was incubated for 1 hr at 50 °C and transformed into DH5alpha competent cells and plated on LB agar supplemented with kanamycin (50 µg/mL). Plasmid DNA was isolated from colonies using a Qiagen QIAprep Spin Miniprep Kit and sequenced using standard T7 primers (Eurofins Genomics).

The gene sequence encoding the expression of *kr*ACP was amplified from the gDNA using the following primers (NdeI and EcoRI endonuclease cut sites are underlined): 5' AAAAAA<u>CATATG</u>GCTAAAGATTCGGG 3' and 5' AAAAAA<u>GAATTC</u>TCATACGCTCTGGAC 3'.
Thermal cycling conditions were as follows (40 ng DNA, Phusion polymerase): 1. 98 °C for 1 min, 2. 98 °C for 10 sec, 3. 65 °C for 30 sec, 4. 72 °C for 30 sec, 4. Repeat steps 2-4 for a total of 34 times, 5. 72 °C 10 for minutes, 6. 4 °C forever. Amplified DNA was purified using the Zymoclean Recovery Kit (Zymo Research). Vector DNA was obtained by digesting pET28a vector with NdeI and EcoRI endonuclease enzymes, treating linearized vector with calf intestinal phosphatase (New England Biolabs) at 16 °C for 30 minutes, and subsequently purified with the

Zymo Resaerch DNA Clean and Concentrator Kit (Zymo Research). Vector and insert were mixed in a 1:1 molar ratio with T4 ligase (0.25 µL; Thermo Fisher), 10X T4 ligase buffer (2 µL) and ddH_2O to a total volume of 20 µL. After 3 hrs of incubating at 16 °C, the ligase was heat deactivated for 10 minutes at 65 °C, transformed into DH5alpha chemically competent cells, and plated on LB agar supplemented with kanamycin (50 µg/mL). Plasmid DNA was isolated from colonies using a Qiagen QIAprep Spin Miniprep Kit and sequenced using standard T7 primers (Eurofins Genomics).

Expression and purification of krACP, krKS-CLF, and FabD

Expression plasmids were transformed into chemically competent BAP1 cells[19] (p*Kr*ACP) or BL21 cells (pYW2; *Kr*KS-CLF), pFabD (*E. coli* FAS AT) for expression. Seed cultures (10 mL LB, 50 µg kanamycin/mL) were grown at 37 °C with shaking and then added to production cultures (1 L LB, 50 µg kanamycin/mL). Production cultures were grown at 37 °C with shaking until OD600 was between 0.4 – 0.6. Each 1 L culture was induced with 250 µL of 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 18 °C for 18 – 21 h with shaking. Cells were harvested by centrifugation (5000 RPM, 4 °C, 20 min), resuspended in lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.6) and sonicated on ice using an XL-200 Microson sonicator (40 % A, 30 sec cycles, 10 min). Supernatant was clarified by centrifugation (13000 RPM, 1 hr) and incubated overnight on a nutating mixer at 4 °C with 2 mL nickel-NTA agarose slurry equilibrated in lysis buffer. After collecting the flow through, resin was washed with 75 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.6), 100 mL of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 30 mM imidazole, pH 7.6), and protein was eluted from the resin with 10 mL of elution buffer (50 mM sodium phosphate, 100 mM NaCl, 300 mM imidazole, pH 7.6). For *Kr*KS-CLF purification, 1 mM DTT and 10% glycerol were added to lysis, wash, and elution buffers. Proteins were concentrated using 3 kDa (*Kr*ACP, FabD) and 30 kDa (*Kr*KS-CLF) Amicon Ultra (Millipore, 15 mL capacity) molecular weight cutoff (MWCO) centricons (4000 RPM, 4 °C) to A_{280} readings of 0.8-1.2. Prior to flash freezing, *Kr*ACP and FabD were dialyzed into 50 mM sodium phosphate, pH 7.6, using Thermo Fisher Slide-A-Lyzer dialysis cassettes. For SV-AUC and SEC studies, *Kr*KS-CLF sample was dialyzed into reducing buffer (50 mM sodium phosphate, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, pH 7.2). Glycerol (10 % v/v) was added to *Kr*KS-CLF samples. All protein aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

*Kr*KS-CLF purification by size exclusion chromatography

A GE Healthcare Superdex 75 increase 10/300 GL size exclusion column was equilibrated with 50 mM sodium phosphate, pH 7.6, 150 mM NaCl and 2 mM DTT. Next, *Kr*KS-CLF (600 µL, 49 µM) was injected using an ÄKTA Pure 25 L FPLC system. Flow rate was held at 0.8 mL/min and absorbance measured at 280 nm. Fractions (1 mL) were collected and those with a strong absorbance at 280 nm (fractions 4-6) were combined and concentrated by centrifugation for SV-AUC analysis.

SDS-PAGE gel electrophoresis

Bio-Rad™ Mini-PROTEAN® TGX™ 10-well 30 µL 4-20 % precast polyacrylamide gels were used for SDS-PAGE gel electrophoresis. *Kr*KS-CLF was dialyzed and diluted to 2x running concentration in reducing buffer (50 mM sodium phosphate, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, pH 7.2). *Kr*ACP was dialyzed and diluted to 2x running concentration in 50 mM sodium phosphate, pH 7.6. Protein samples were further diluted 1:1 v/v to running concentration with 2x SDS-PAGE loading dye (100mM Tris-HCl, 4% w/v sodium dodecyl sulfate, 0.2% w/v bromophenol blue, 20% v/v glycerol, pH 6.8) to reach a final concentration of 1x SDS-PAGE loading dye. For *Kr*ACP samples run under reducing conditions, 5% v/v BME was added to the protein sample. Gels were run in tris-glycine running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS, pH 8.3) at 120 V for 90 min. Gels were washed with ddH_2O and stained for 15 minutes in Thermo Scientific™ GelCode Blue Safe Protein Stain (Coomassie G-250). After staining, gels were de-stained in ddH_2O overnight or until sufficient band contrast had been reached prior to imaging.

Tandem proteolysis mass spectrometry

Liquid chromatography – tandem mass spectrometry (LC-MS/MS) was performed by the Proteomics and Metabolomics Facility at the Wistar Institute (Philadelphia, PA). A Q Exactive Plus mass spectrometer (ThermoFisher Scientific) coupled with a Nano-ACQUITY UPLC system (Waters) were used for the analyses. Samples were digested in gel with trypsin and injected onto a UPLC Symmetry trap column (180 µm i.d. x 2 cm packed with 5 µm C18 resin; Waters). Tryptic peptides were separated by reversed phase HPLC on a BEH C18 nanocapillary analytical column (75 µm i.d. x 25 cm, 1.7 µm particle size; Waters) using a 95 min gradient formed by solvent A (0.1 % formic water in water) and solvent B (0.1 % formic in acetonitrile). A 30-min blank gradient was run between sample injections to minimize carryover. Eluted peptides were analyzed by the mass spectrometer set to repetitively scan *m/z* from 400 to 2000 in positive ion mode. The full MS scan was collected at 70,000 resolution followed by data-dependent MS/MS scans at 175,000 resolution on the 20 most abundant ions exceeding a minimum threshold 20,000. Peptide match was set as “preferred, exclude isotopes” option and charge-state screening were enabled to reject singly and unassigned charged ions. Peptide sequences were identified using the MaxQuant 1.6.1.0 program. MS/MS spectra were searched against a custom *E. coli* UniProt protein database using full tryptic specificity with up to two missing cleavages, static carboxamidomethylation (57.02146) of Cys and variable oxidation (15.99491) of Met. Consensus identification lists were generated with false discovery rates of 1% at protein, peptide and site levels. Data were plotted by the Wistar Institute.

High-performance liquid chromatography (HPLC)

HPLC data were collected using a Varian ProStar High-Performance Liquid Chromatography instrument equipped with a Macherey-Nagel NUCLEOSIL® C4 column. The instrument was equipped with two solvents: Solvent A (H_2O + 0.1% trifluoroacetic acid) and Solvent B (acetonitrile + 0.1% trifluoroacetic acid). The column was equilibrated for 10 minutes

at a 1 mL/min flow rate with 90% Solvent A and 10% Solvent B, after which 25 µL of 50 µM *Kr*ACP was injected into the column. The column was then run for 30 minutes, during which the solvent gradient linearly increased from 10% Solvent B to 100% Solvent B (3% min^{-1}). The eluent was measured by absorbance at 235 nm. Data were plotted in Microsoft Excel (v.15.28).

Liquid chromatography-mass spectrometry and phosphopantetheine ejection assay

LCMS data were collected in the positive mode on an Agilent Technologies InfinityLab G6125B LC/MS coupled with Agilent 1260 Infinity II LC system with a Waters XBridge Protein BEH C4 reverse phase column. The following solvent gradient was used (solvent A = water + 0.1% formic acid; solvent B = acetonitrile + 0.1% formic acid): 0–1 min 95% A, 3.1 min 5% A, 4.52 min 5% A, 4.92–9 min 95% A. Samples were run using a capillary voltage of 3000 V and a fragmentation voltage of 75V. For the Ppant ejection assay, the fragmentation voltage was adjusted to 250V. Resulting spectra were deconvoluted using ESIProt.[33] Data were plotted in Microsoft Excel (v.15.28).

*Kr*ACP malonylation experiments

To a microcentrifuge tube, *Kr*ACP (2.6 mM, 26.9 µL) was diluted to a volume of 500 µL in 50 mM sodium phosphate, pH 7.6, to reach a target concentration of 140 µM *Kr*ACP. To a second tube, *Kr*ACP (2.6 mM, 26.9 µL) and malonyl-CoA (50 mM, 10 µL) were added and diluted to a target volume of 500 µL in sodium phosphate, pH 7.6, to reach target concentrations of 140 µM *Kr*ACP and 1 mM malonyl-CoA. To a third tube, *Kr*ACP (2.6 mM, 26.9 µL), malonyl-CoA (50 mM, 10 µL), and FabD (56.1 µM, 13.4 µL) were mixed in 50 mM sodium phosphate, pH 7.6 to reach target concentrations of 140 µM *Kr*ACP, 1 mM malonyl-CoA, and 1.5 µM FabD in a total volume of 500 µL. All solutions were incubated at 23 ºC for 30 min.

The 500 µL reaction mixtures were loaded onto a GE Healthcare Superdex 75 increase 10/300 GL size exclusion column attached to a ÄKTA Pure 25 L FPLC system. Samples were run at a flow rate of 0.8 mL/min and absorbance was measured at 280 nm. Fractions were collected at 1 mL, and Fraction 8 (corresponding to 12.7-13.7 mL volume eluted) was collected for the Ppant ejection assay analysis. 25µL samples of the collected fractions were run on LCMS using a capillary voltage of 3000 V and a fragmentation voltage of 250 V. Data were plotted in Microsoft Excel (v.15.28).

Sedimentation velocity analytical ultracentrifugation (SV-AUC)

Sedimentation velocity experiments were performed using a Beckman Optima XL-A ultracentrifuge equipped with a 4-hole An-60 Ti rotor. Temperature-corrected partial specific volumes, densities, and viscosities were calculated using Sednterp (beta version).[34] Data were analyzed using the dc/dt method $DCDT^{+}$ (v.2.4.0), and a model-independent continuous c(s) distribution from Sedfit (v. 15.3).[35] Confidence intervals for temperature-corrected sedimentation coefficient, s(*20,w*), diffusion coefficient, D(*20,w*), and molecular mass (kDa) were computed using bootstrap method, CI 90% (± 1.65 sigma) implemented in $DCDT^{+}$ (v.2.4.0).

SV-AUC experiments used *Kr*KS-CLF purified by size exclusion chromatography in elution buffer (50 mM sodium phosphate, pH 7.6) with the presence (**Fig 6**) or absence of salt (150 mM NaCl) and reducing agent (2 mM DTT). An SV-AUC concentration gradient profile of *Kr*KS-CLF prior to gel filtration chromatography purification was also taken (**Fig S4B**). For *Kr*KS-CLF purified in the presence of salt and reducing agent, *Kr*KS-CLF concentration was 2.9 μM and *Kr*ACP concentration was 140 μM. A mixed sample containing both *Kr*KS-CLF (4 μM) and *Kr*ACP (110 μM) was also analyzed. All samples were diluted to the required starting absorbance between 0.3 and 1.0 AU and had a final concentration of 150 mM NaCl and 1 mM DTT. SV-AUC runs used two-channel Epon charcoal-filled centerpieces containing 450 μL protein samples and 450 μL buffer as reference. Sedimentation velocity boundaries were measured at a speed of 42,000 rpm at 20 ºC using a step size of 0.003 cm, a delay time of 0 sec, and a total of 150 scans. Samples were monitored at 280 nm.

Sample heterogeneity was determined using model-independent continuous c(s) distribution where regularization of the distribution by the maximum entropy with the parameter constrained value at 0.95. The s and D parameters required to solve the Lamm equation are calculated with the approximation that all species in a solution for a given sample have similar signal-average frictional coefficient f/f_0. The distributions were plotted using the Gussi interface (v. 1.0.3) implemented in Sedfit (v. 15.3).

Circular dichroism spectroscopy (CD)

CD spectra were collected using an Aviv Model 410A circular dichroism spectropolarimeter. Proteins were dialyzed in absence of DTT to prevent any light scattering from DTT. *Kr*KS-CLF was dialyzed overnight into 50 mM sodium phosphate, pH 7.6, and *Kr*ACP was dialyzed overnight into 10 mM Tris-HCl, pH 7.6. Protein samples were diluted to 200 μL to final concentrations of 13 μM in 50 mM sodium phosphate, pH 7.6 for *Kr*KS-CLF and 50 μM into 10 mM Tris-HCl, pH 7.6 for *Kr*ACP. Prior to injection, samples were filtered using a 0.2 μM low protein-binding filter with a HT Tuffryn Membrane (Pall Corporation). Samples were injected into a High Precision Quartz SUPRSIL cuvette with 0.1 cm pathlength (Hellma Analytics). The spectropolarimeter was purged with nitrogen for one hour, and CD spectra were collected at 25 ºC with a range of 260-180 nm using a bandwidth of 1 nm, a 0.5 nm step size, and an averaging time of 3 sec. Changes in signal at 222 nm were followed as function of temperature using the following parameters; 10-90 ºC, 2-min equilibration, heating rate 2 ºC min^{-1}, 30 sec signal averaging time, and 1 nm bandwidth. Pre- and post-Tmelt spectra were smoothed using a smoothing function implemented in the Aviv software, applying a window width of 11 data points, degree 2. Thermal stability data were analyzed using CDpal (v. 2.18) using a two-state unfolding model N↔D with the standard assumption that $\Delta C_P = 0$. Estimated errors for each fitted parameter were calculated using a Jackknifing method implemented in CDpal (v.2.18). Data were plotted in Origin (v.8.6.0). The resulting spectra were converted to units of mean residue ellipticity (MRE) using the amino acid sequence of the proteins, smoothed, and normalized to an MRE of 0 at 250 nm.

Acknowledgements

We are grateful for generous support from the National Science Foundation CAREER Award (CHE-1652424 to L.K.C.) and the Henry Dreyfus Teacher-Scholar Award (TH-10-2020 to L.K.C.). We would like to thank Grayson S. Hamrick and Sharon C. Nwankpa for their technical support with protein expression and purification, Professor Robert Fairman (Haverford Biology) for generously offering us the use of the analytical ultracentrifuge for this work, Professor Dominic Campopiano (University of Edinburgh) for providing the FabD expression plasmid, and Dr. Maureen Hillenmeyer (Hexagon Bio) for helpful discussions).

References

(1) Hertweck, C. *Angew. Chemie Int. Ed.* **2009**, *48* (26), 4688.
(2) Zhang, Z.; Pan, H.-X.; Tang, G.-L. *F1000Research* **2017**, *6*, 172.
(3) Lancini, G.; Demain, A. L. In *The Prokaryotes*; Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 257–280.
(4) Williams, G. J. *Curr. Opin. Struct. Biol.* **2013**, *23* (4), 603.
(5) Wong, F. T.; Khosla, C. *Curr. Opin. Chem. Biol.* **2012**, *16* (1–2), 117.
(6) Lowry, B.; Walsh, C.; Khosla, C. *Synlett* **2015**, *26* (08), 1008.
(7) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. *Nat. Prod. Rep.* **2007**, *24* (1), 162.
(8) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. *Biochemistry* **2014**, *53* (18), 2875.
(9) Gulick, A. M.; Aldrich, C. C. *Nat. Prod. Rep.* **2018**, *35*, 1156.
(10) Schaub, A. J.; Moreno, G. O.; Zhao, S.; Truong, H. V.; Luo, R.; Tsai, S. C. *Methods Enzymol.* **2019**, *622*, 375.
(11) Hillenmeyer, M. E.; Vandova, G. A.; Berlew, E. E.; Charkoudian, L. K. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (45), 13952.
(12) Fujii, I. *Nat. Prod. Rep.* **2009**, *26* (2), 155.
(13) Cummings, M.; Peters, A. D.; Whitehead, G. F. S.; Menon, B. R. K.; Micklefield, J.; Webb, S. J.; Takano, E.; Id, G. F. S. W.; Binuraj, R.; Menon, K.; Micklefield, J.; Webb, S. J.; Takano, E. *PLOS Biol.* **2019**, *17* (7), e3000347.
(14) Chang, Y.; Land, M.; Hauser, L.; Chertkov, O.; Del Rio, T. G.; Nolan, M.; Copeland, A.; Tice, H.; Cheng, J.-F.; Lucas, S.; Han, C.; Goodwin, L.; Pitluck, S.; Ivanova, N.; Ovchinikova, G.; Pati, A.; Chen, A.; Palaniappan, K.; Mavromatis, K.; Liolios, K.; Brettin, T.; Fiebig, A.; Rohde, M.; Abt, B.; Göker, M.; Detter, J. C.; Woyke, T.; Bristow, J.; Eisen, J. A.; Markowitz, V.; Hugenholtz, P.; Kyrpides, N. C.; Klenk, H.-P.; Lapidus, A. *Stand. Genomic Sci.* **2011**, *5* (1), 97.
(15) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science.* **2001**, *291* (5509), 1790.
(16) Hill, R. B.; Mackenzie, K. R.; Flanagan, J. M.; Cronan, J. E.; Prestegard, J. H. *Protein Expr. Purif.* **1995**, *6* (4), 394.
(17) Chan, Y. A.; Thomas, M. G. *Biochemistry* **2010**, *49* (17), 3667.
(18) Rivas, M. A.; Courouble, V. C. V. C.; Baker, M. C.; Cookmeyer, D. K. L.; Fiore, K. E.; Frost, A. J.; Godbe, K. N.; Jordan, M. R.; Krasnow, E. N.; Mollo, A.; others; Nawal, S.; Ridings, S. T.; Sawada, K.; Shroff, K. D.; Studnitzer, B.; Thiele, G. A. R.; Sisto, A. C.; Huff, A. R.; Fairman, R.; Johnson, K.; Beld, J.; Kokona, B.; Charkoudian, L. K. *AIChE J.* **2018**, *64* (12), 4308.
(19) Pfeifer, B. A.; Khosla, C. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 106.
(20) Das, A.; Szu, P.-H.; Fitzgerald, J. T.; Khosla, C. *J. Am. Chem. Soc.* **2010**, *132* (26), 8831.
(21) Bräuer, A.; Zhou, Q.; Grammbitter, G. L. C.; Schmalhofer, M.; Rühl, M.; Kaila, V. R. I.; Bode, H. B.; Groll, M. *Nat. Chem.* **2020**, [Epub ahead of print]
(22) Arthur, C. J.; Szafranska, A.; Evans, S. E.; Findlow, S. C.; Burston, S. G.; Owen, P.; Clark-Lewis, I.; Simpson, T. J.; Crosby, J.; Crump, M. P. *Biochemistry* **2005**, *44* (46), 15414.
(23) Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. *Biochemistry* **2006**, *45* (42), 12756.
(24) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. *Nat Struct Mol Biol* **2004**, *11* (9), 888.
(25) Du, D.; Katsuyama, Y.; Horiuchi, M.; Fushinobu, S.; Chen, A.; Davis, T. D.; Burkart, M. D.; Ohnishi, Y.

Nat. Chem. Biol. **2020**, [Epub ahead of print].
(26) Charkoudian, L. K.; Liu, C. W.; Capone, S.; Kapur, S.; Cane, D. E.; Togni, A.; Seebach, D.; Khosla, C. *Protein Sci.* **2011**, *20* (7), 1244.
(27) Kokona, B.; Winesett, E. S.; von Krusenstiern, A. N.; Cryle, M. J.; Fairman, R.; Charkoudian, L. K. *Anal Biochem* **2016**, *495* (3), 42.
(28) Jakočiūnas, T.; Klitgaard, A. K.; Romero-Suarez, D.; Petzold, C. J.; Gin, J. W.; Tong, Y.; Frandsen, R. J. N.; Weber, T.; Lee, S. Y.; Jensen, M. K.; Keasling, J. D. *bioRxiv*, **2019**, *5* (1), 11.
(29) Liu, X.; Hua, K.; Liu, D.; Wu, Z. L.; Wang, Y.; Zhang, H.; Deng, Z.; Pfeifer, B. A.; Jiang, M. *ACS Chem. Biol.* **2019**, [Epub ahead of print].
(30) Szu, P. H.; Govindarajan, S.; Meehan, M. J.; Das, A.; Nguyen, D. D.; Dorrestein, P. C.; Minshull, J.; Khosla, C. *Chem. Biol.* **2011**, *18* (8), 1021.
(31) Arthur, C. J.; Szafranska, A.; Evans, S. E.; Findlow, S. C.; Burston, S. G.; Owen, P.; Clark-Lewis, I.; Simpson, T. J.; Crosby, J.; Crump, M. P. *Biochemistry* **2005**, *44* (46), 15414.
(32) Matharu, A. L.; Cox, R. J.; Crosby, J.; Byrom, K. J.; Simpson, T. J. *Chem. Biol.* **1998**, *5*, (12), 669.
(33) Winkler, R. *Rapid Commun. Mass Spectrom.* **2010**, *24* (3), 285.
(34) Schuck, P.; Zhao, H. *Methods* **2011**, *54* (1), 1.
(35) Schuck, P. *Analyst* **2016**, *141* (14), 4400.

Supporting Information for

Opening the bottleneck: Heterologous expression, purification, and characterization of the *Ktedonobacter racemifer* minimal type II polyketide synthase in *Escherichia coli*

Joshua G. Klein, Yang Wu, Bashkim Kokona, Louise K. Charkoudian

Index of Supplementary Information and Figures

Supplementary Information

Supplementary Figures

Protein sequences of expressed proteins

*Kr*ACP

MGSSHHHHHHSSGLVPRGSHMAKDSGQATVSSRVYEILRPYAEGIALTPETHLSDDLNIDSIELVEVGVALE
KEFSNKRFTITGLKSCPTVQDLVKLVEQTVAAEQVQSV*

*Kr*KS-CLF

MGSSHHHHHHSSGLVPRGSHMRRVVISGLGVVSPAGIGKEAFWQNLLDGVSGAVALDRVTCSPLFGHHEF
GAQAVCEVTGFDPVTHHVPTAYHSADRFIQFAFAAVHQAFLDAHLDQDTWDVSRVGVTLATAICGTQTL

DLEFTKSTNGGKEAFHSEGISPFLYSAAMGNSAALAVGSRYGLQGECATLSTGCIAGLDAISYAYESIAYGD
HDVMIAGASEAPITPITIAAFDIINCLSHHSDPTTASRPFSVDRDGFVLSEGCGIVVLEELEHALARNAPIYAEI
LGCDVTEHAVHMTDMSPEGRDLARAITGALGKANVEPEAIDFVNAHGTSTPQNDFFESLALKTSLGQQAL
TIPVNSTKSMVGHALAAASAVEVVACAMSMQTQRIHPTINLVNPDPRCDLDYVPNHARSHDVQRMLTTAS
GFSGLHAAMVLESYSTREEMMARIVITGMGVISPYGIGSGILWEKLLAGENGLKPLTTFETSHIQCRVGGQL
LDFRPEAYLSPRLIRKIDRFSTFGLISAYLALQDAGLLSDGKKPVWTQQEQHSHRVGITVGNNLGGWEFAER
ELRHLWALGPRDVSPHMATAWFPAAVQGNMSIYFGIKGIGRTFLSDRASGALAIMHAADCLQRGRADIML
AGGTEAPFSPYAALCYETSGLMSKKAVTGSPETYRPFDEAHDGLVAGEGAAFFILERAEDAEKRGATILAE
VAGWASTNDGYHPVQPAPDGERYAAAMTRAMQRADASADEVDCLFAAGSAVPDEDVSETRAIHLALRE
AVRRVPVATPKSAFGNLFGAAFPVDMAIALLAMQHRVLPATLHLDQAAPGCDLDYVPQTPRSVDHLDRCL
INARGIGGANASVLLRTWA

E. coli FabD

MGSSHHHHHHSSGLVPRGSHMTQFAFVFPGQGSQTVGMLADMAASYPIVEETFAEASAALGYDLWALTQ
QGPAEELNKTWQTQPALLTASVALYRVWQQQGGKAPAMMAGHSLGEYSALVCAGVIDFADAVRLVEMR
GKFMQEAVPEGTGAMAAIIGLDDASIAKACEEAAEGQVVSPVNFNSPGQVVIAGHKEAVERAGAACKAA
GAKRALPLPVSVPSHCALMKPAADKLAVELAKITFNAPTVPVVNNVDVKCETNGDAIRDALVRQLYNPVQ
WTKSVEYMAAQGVEHLYEVGPGKVLTGLTKRIVDTLTASALNEPSAMAAALEL

Protein threading and modeling

Plasmid nucleotide sequences were translated using the web program ExPasy. This nucleotide sequence was threaded by the web program Phyre 2.0. Modeling and analysis of the threaded sequence was performed using PyMol.

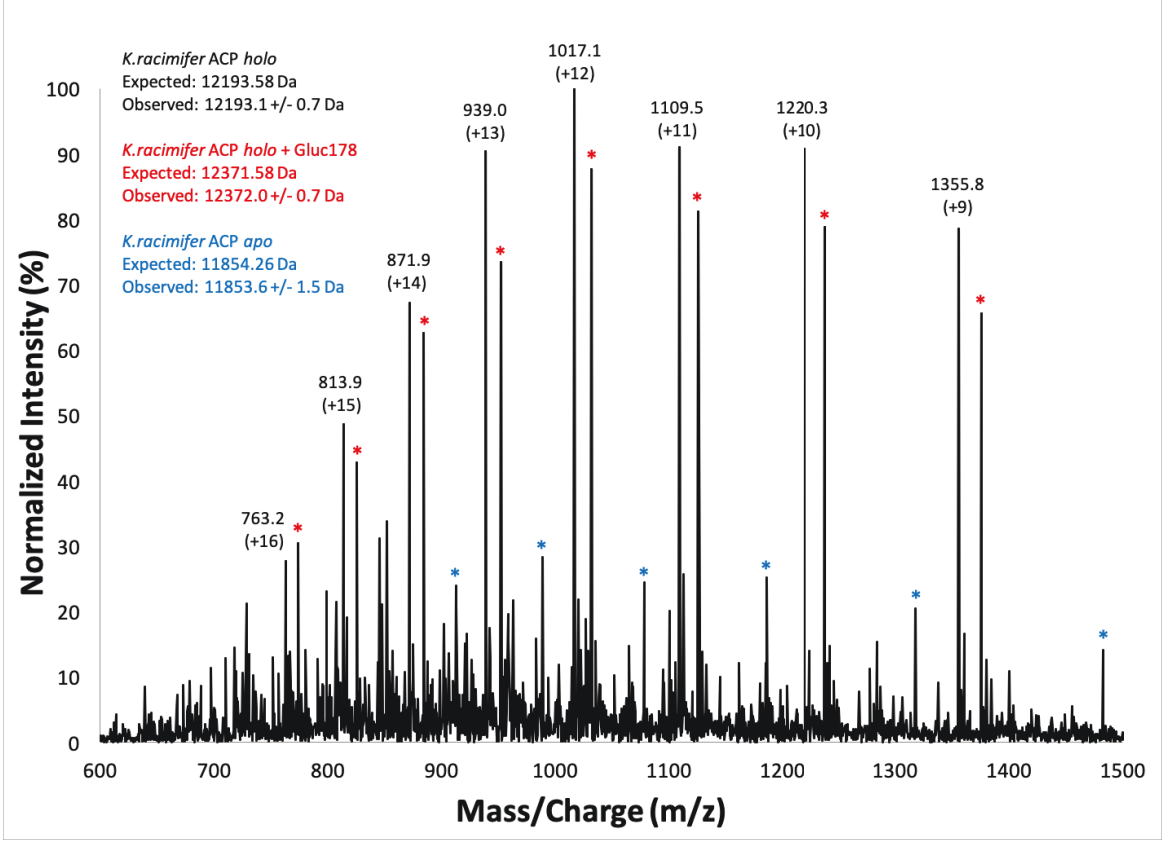

Fig S1. LC-MS spectrum of purified krACP. The *Kr*ACP LS-MS spectrum shows both *holo* and *apo-Kr*ACP post-purification in the *E. coli* BAP1 competent cell line. All theoretical molecular weights are calculated with the removal of a single methionine residue.

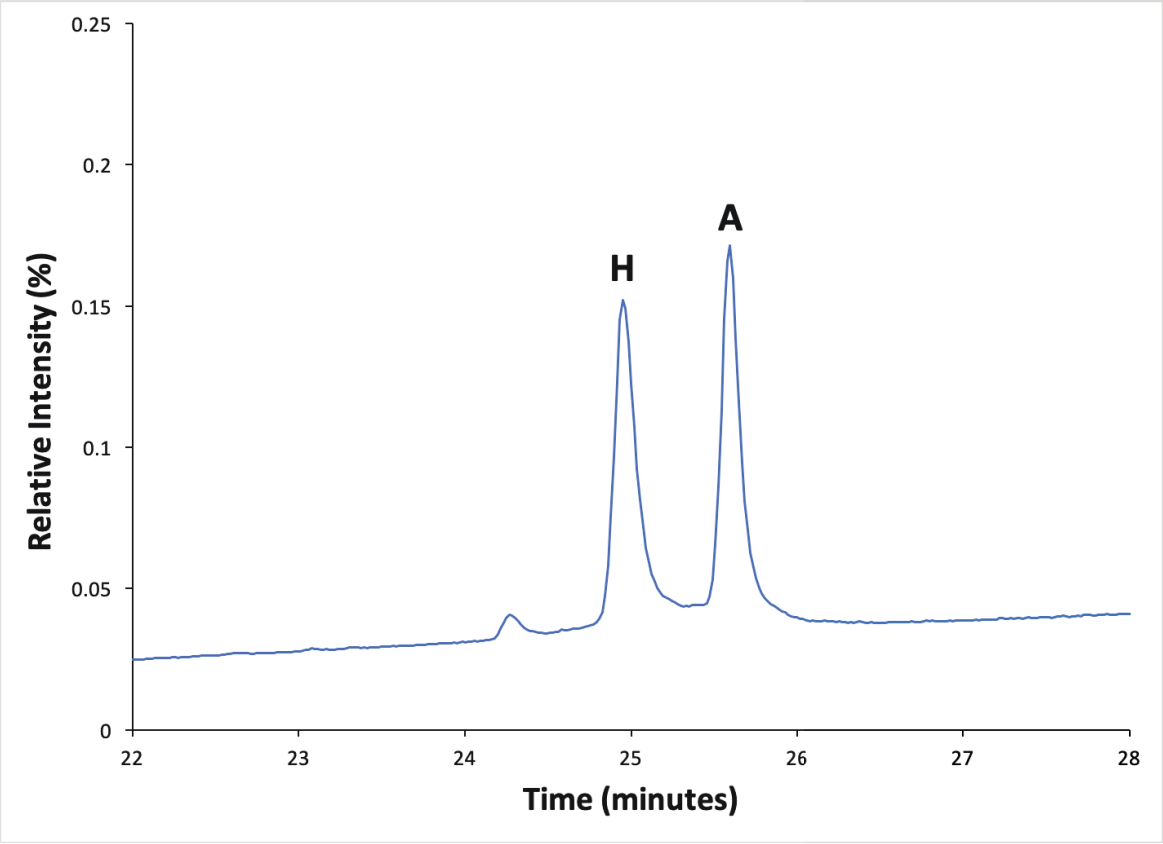

Fig S2. High-pressure liquid chromatography profile of *Kr*ACP. H: Peak corresponding to *holo-Kr*ACP. A: Peak corresponding to *apo-Kr*ACP. The HPLC profile of purified *Kr*ACP confirms the presence of both *holo* and *apo-Kr*ACP in an approximate 1:1 ratio.

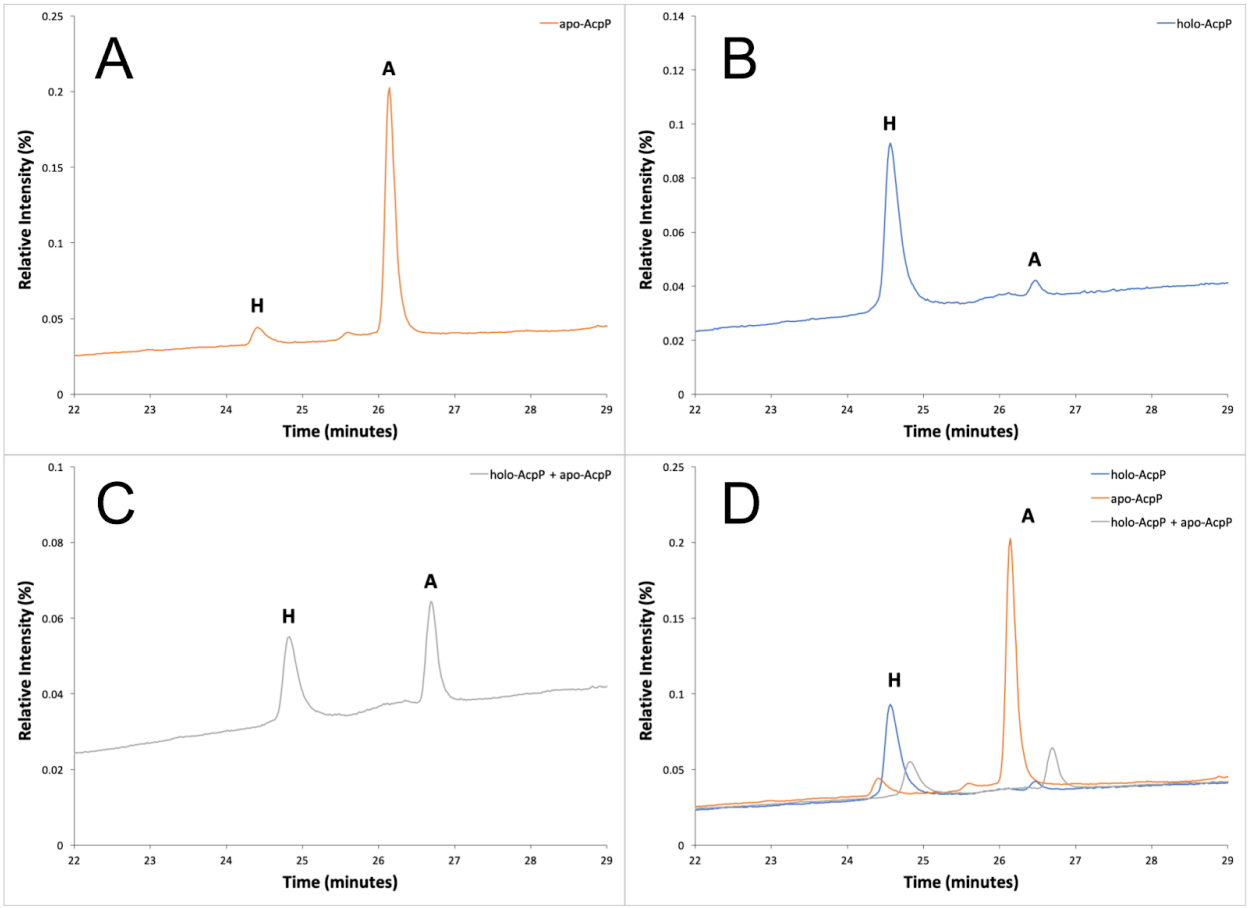

Fig S3. High-pressure liquid chromatography spiking experiments with *E. coli* AcpP. H corresponds to *holo*-AcpP peak and A corresponds to *apo*-AcpP peak. A. HPLC profile of a primarily *apo*-AcpP sample (70 µM) shows a minor peak at 24.41 minutes corresponding to *holo*-AcpP and a major peak at 26.14 minutes corresponding to *apo*-AcpP; B. HPLC profile of a primarily *holo*-AcpP sample (70 µM) shows a major peak at 24.59 minutes corresponding to *holo*-AcpP and a minor peak at 26.47 minutes corresponding to *apo*-AcpP; C. HPLC profile of a mixed sample containing (52.5 µM) *holo*-AcpP and (17.5 µM) *apo*-AcpP shows major peaks at 24.80 minutes corresponding to *holo*-AcpP and at 26.62 minutes corresponding to *apo*-AcpP; D. Overlaid HPLC profiles of subfigures S3 A-C.

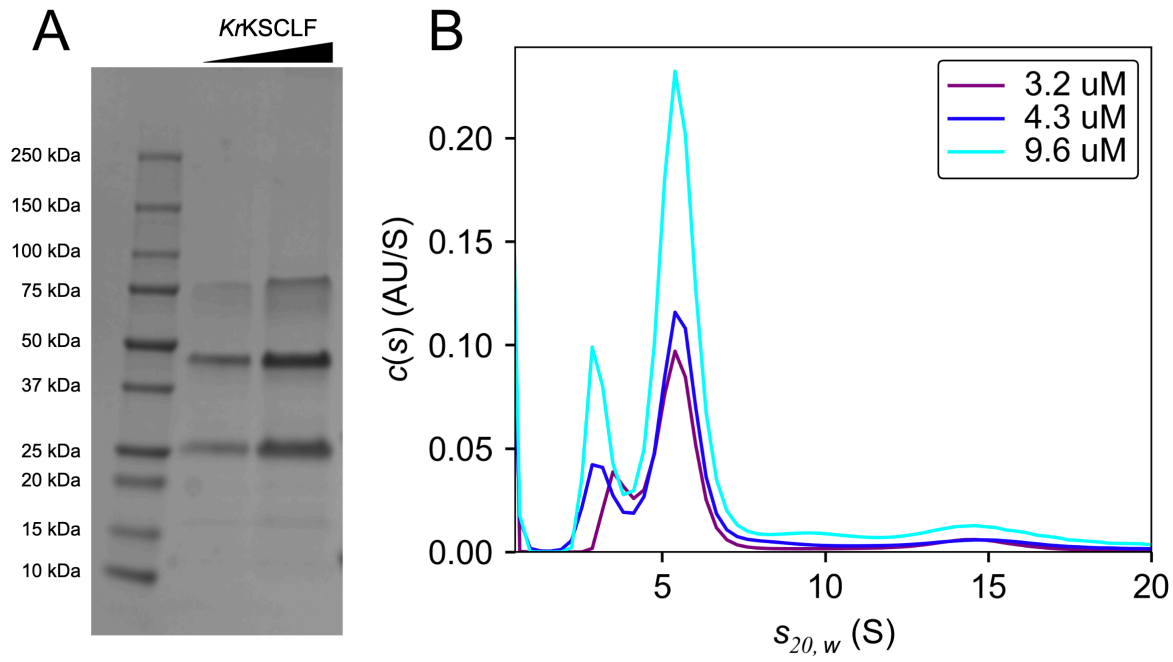

Fig S4. Pre-SEC SDS-PAGE and SV-AUC analysis of *Kr*KS-CLF. A. SDS-PAGE of *Kr*KS-CLF pre-SEC purification. Lanes: 1. Protein standards ladder; 2. *Kr*KS-CLF (10 µL, 3.7 µM); 3. *Kr*KS-CLF (20 µL, 3.7 µM). B. SV-AUC profile of *Kr*KS-CLF pre-SEC purification at increasing concentrations. The distribution between the 25.0 kDa peak and the 98.7 kDa peak of SEC-purified *Kr*KS-CLF does not change with increasing *Kr*KS-CLF concentration. This indicates that the protein(s) represented by the 25.0 kDa peak is/are not in equilibrium with the protein(s) presented by the 98.7 kDa peak, and that the two peaks do not represent different oligomeric states of the same protein. This 25.0 kDa peak persists after SEC purification in 50 mM sodium phosphate, pH 7.6, but can be separated upon addition of 150 mM NaCl and 2 mM DTT to the equilibration buffer.

Fig S5. Tandem proteolysis mass spectrometry profile of ~47 kDa SDS-PAGE Band. The ~47 kDa SDS-PAGE band was digested with proteolytic enzymes and the resulting mass spectra fragments was compared to the mass of expected proteolysis cut sites. Comparing the tandem proteolysis mass spectrometry profile of the ~47 kDa band to expected proteolysis cut sites in the full *Kr*KS-CLF sequence, all expected proteolysis fragments from both *Kr*KS and *Kr*CLF sequences are present in the band. Blank fragments were either too small or too large to detect via this method.

Fig S6. Tandem proteolysis mass spectrometry profile of ~25 kDa SDS-PAGE band. The ~25 kDa SDS-PAGE band was digested with proteolytic enzymes and the resulting mass spectra fragments was compared to the mass of expected proteolysis cut sites. Comparing tandem proteolysis mass spectrometry profile of the ~25 kDa band to expected proteolysis cut sites in the full *Kr*KS-CLF sequence, most expected proteolysis fragments from both *Kr*KS and *Kr*CLF are present in the band. However, the band was mostly comprised of the *E. coli* proteins SlyD, ThiM, and TpiA.

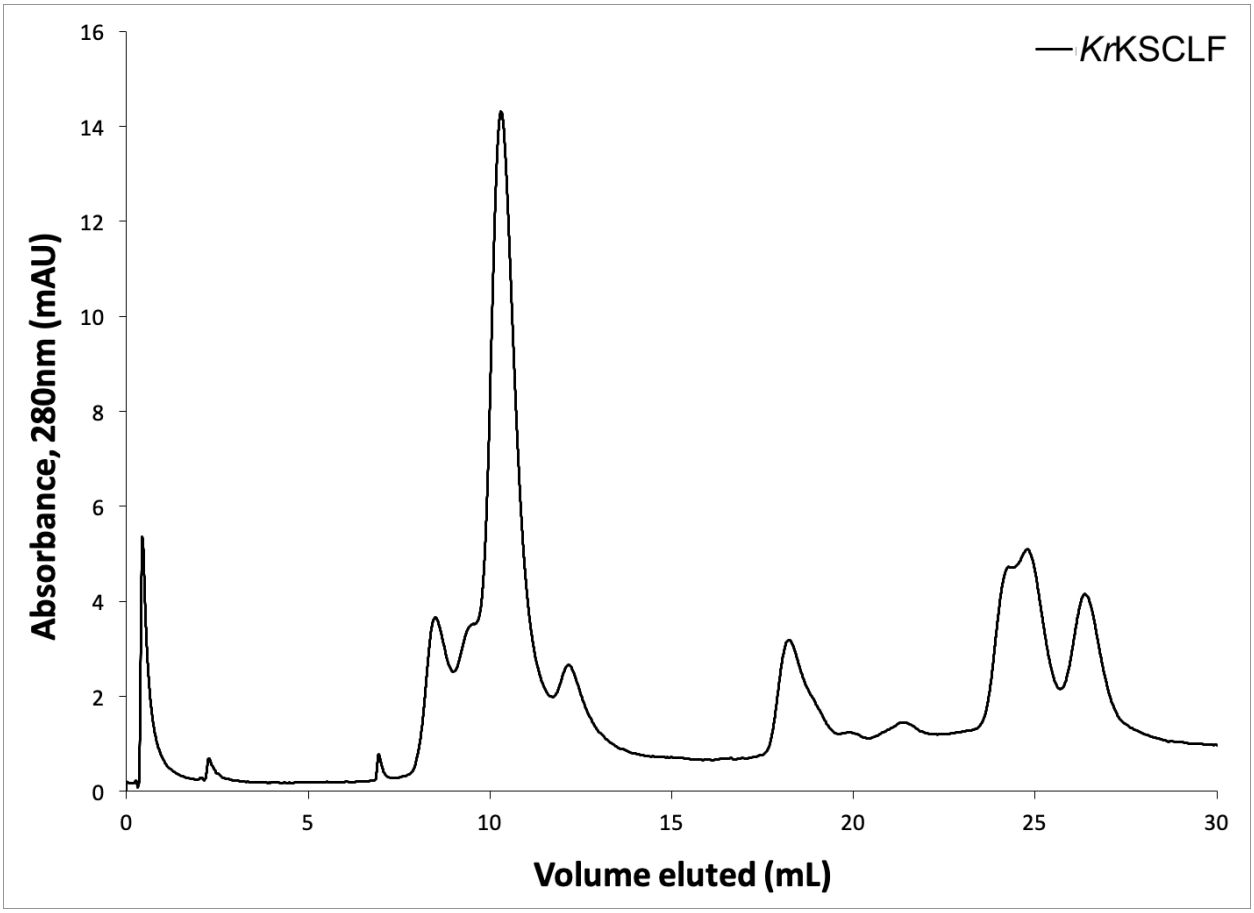

Fig S7. SEC purification of *Kr*KS-CLF. The SEC profile of *Kr*KS-CLF shows a major species at a high molecular weight, attributed to the *Kr*KS-CLF heterodimer. This peak (Fractions 4+5, corresponding to 8.7-10.7 mL eluted) was collected and further analyzed by SDS-PAGE and SV-AUC.

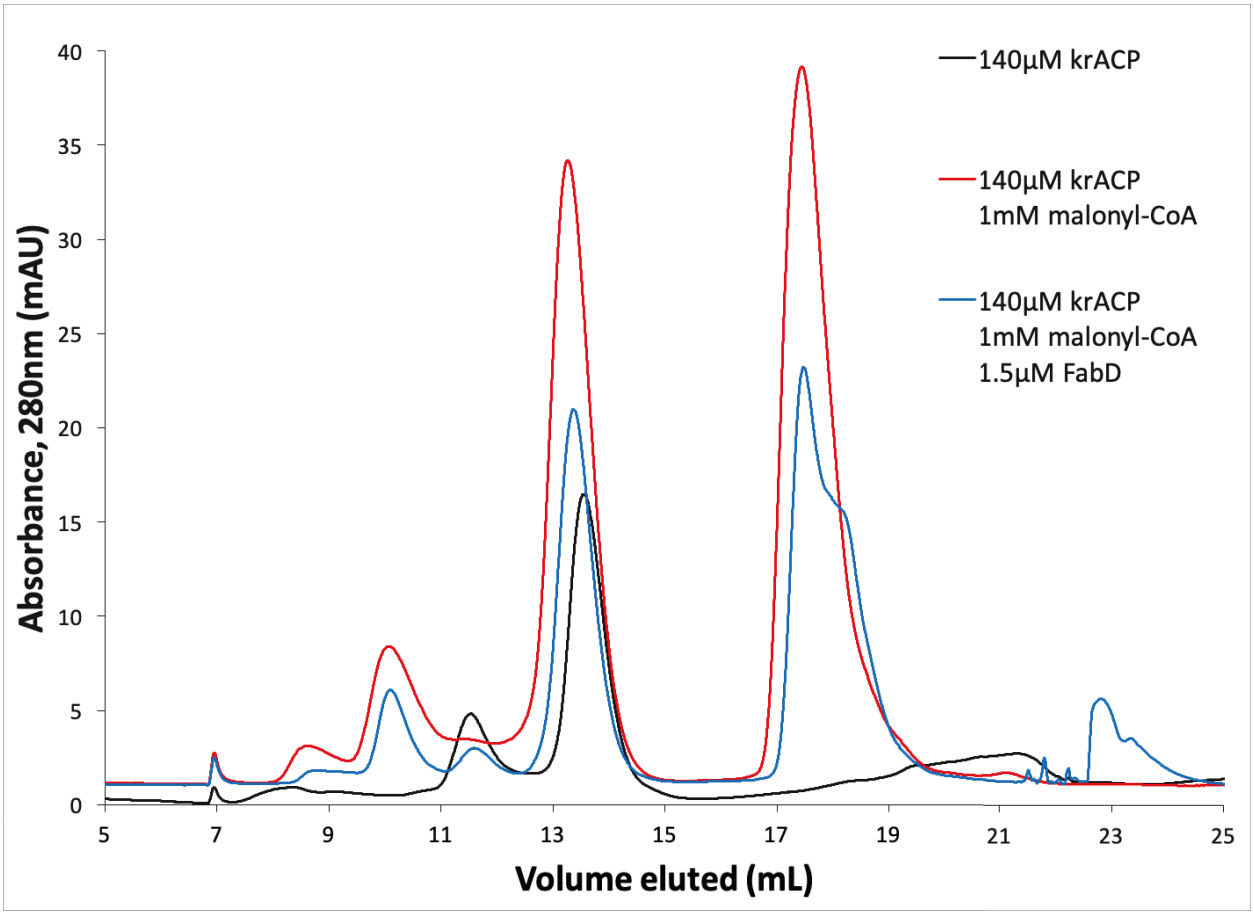

Fig S8. SEC Purification of *Kr*ACP malonylation solutions. The fraction corresponding to 12.7-13.7 mL volume eluted was collected and analyzed by LC-MS.

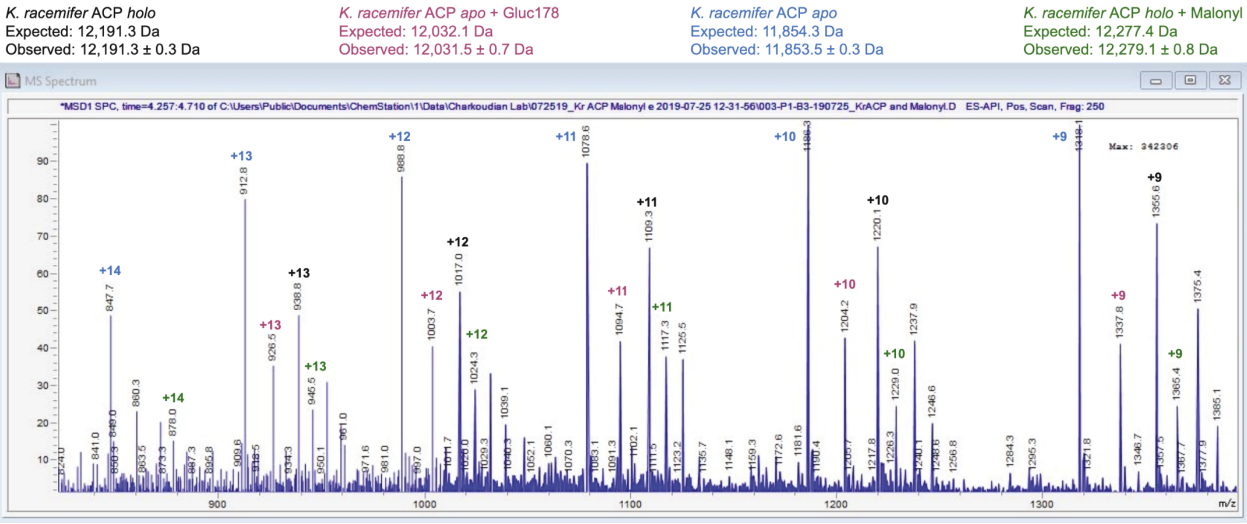

Fig S9. LC-MS spectrum of *Kr*ACP after incubation with malonyl-CoA. The *Kr*ACP LC-MS spectrum shows the presence of *apo*- (red, blue), *holo*- (black), and *malonyl*- (green) *Kr*ACP after incubation with malonyl-CoA. All theoretical molecular weights are calculated with the removal of a single methionine residue.

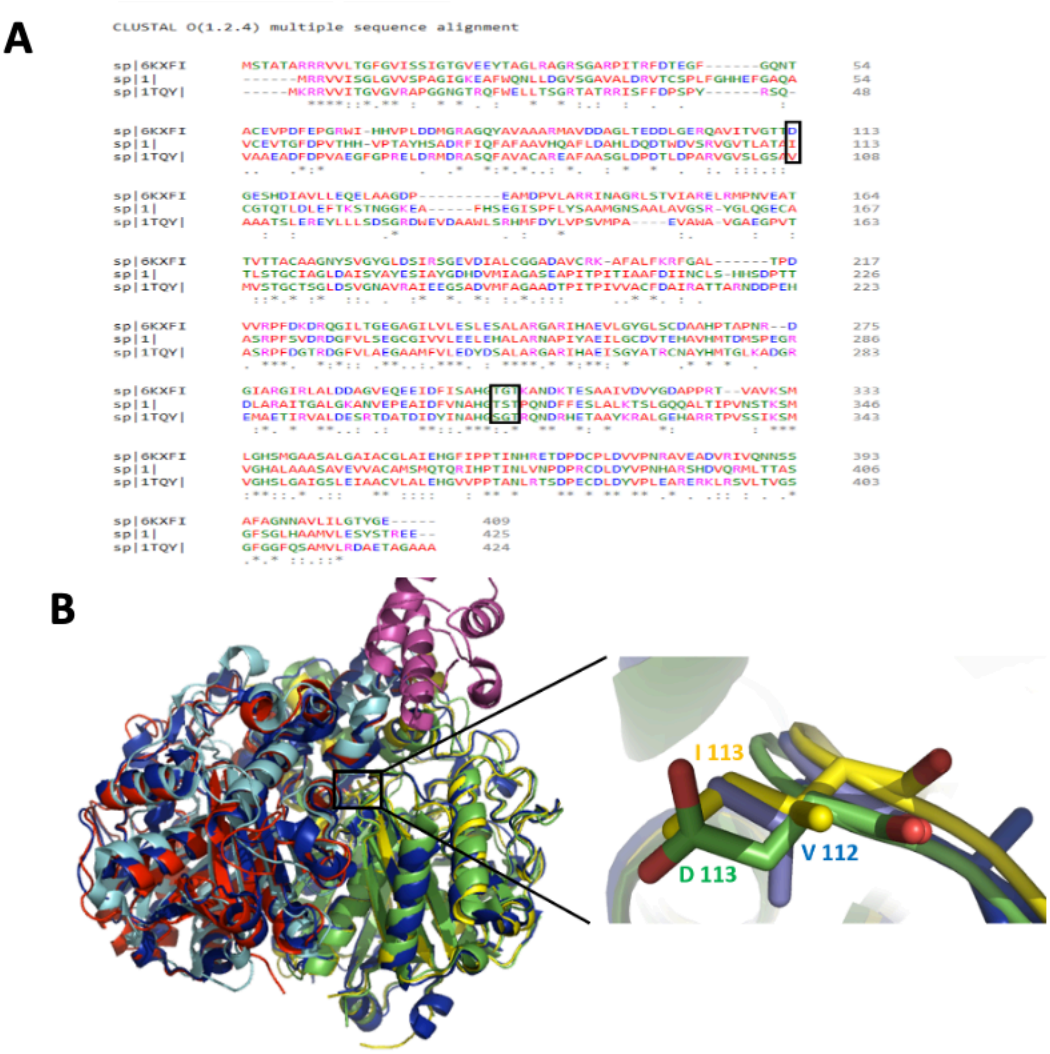

Figure S10. Sequence (A) and structure (B) alignments of *Kr*KS, actKS (1TQY) and Ig11 (6KXFI). Green (Iga11), yellow (KrKS), blue (actKS).

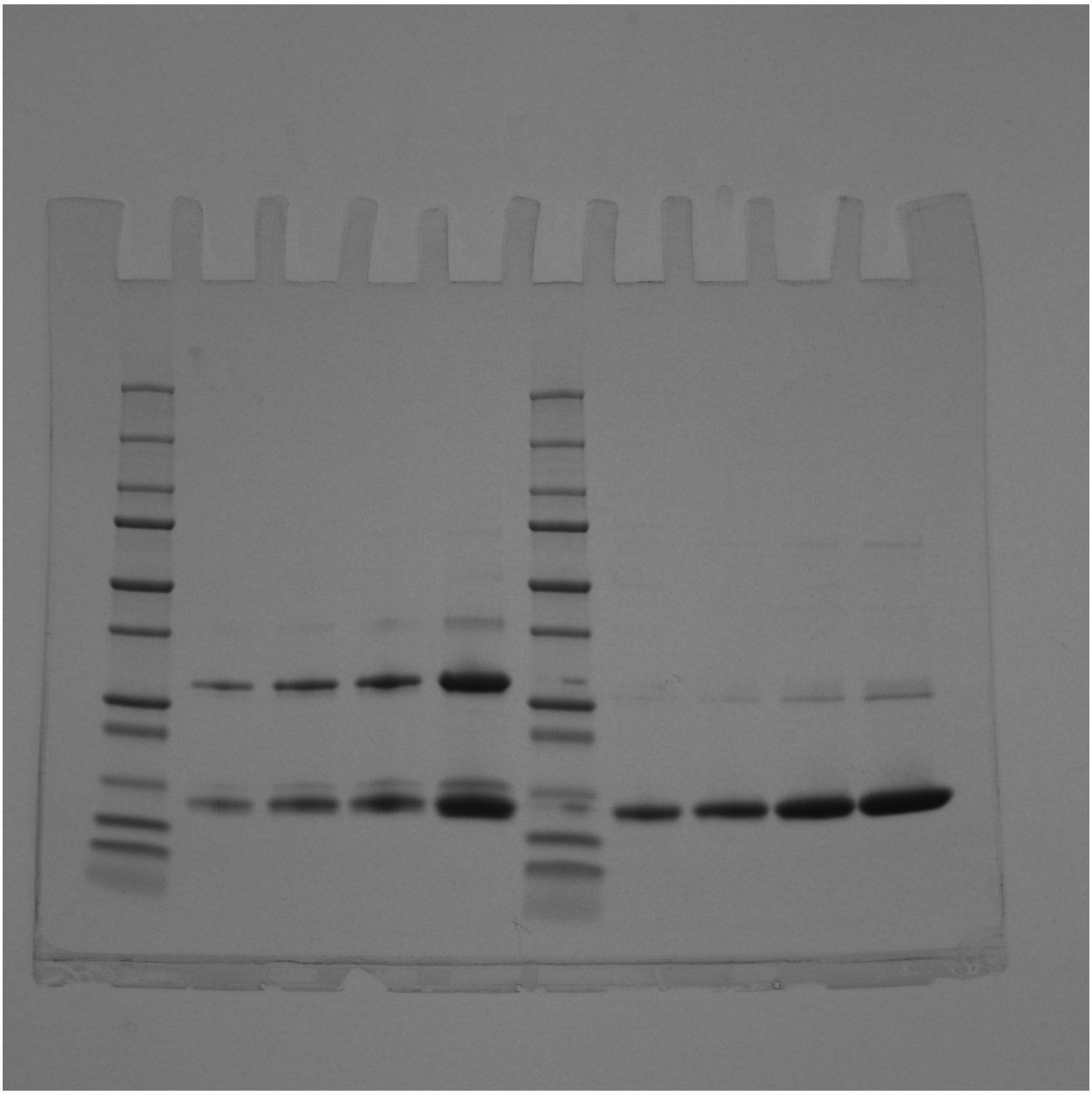

Fig S11. Uncropped Gel of Fig 2A. *Kr*ACP concentration gradients run both without and with the presence of 5% v/v BME. Lanes: 1. Protein standards ladder; 2. *Kr*ACP (31.4 μM, 5 μL) run under non-reducing conditions; 3. *Kr*ACP (31.4 μM, 10 μL) run under non-reducing conditions; 4. *Kr*ACP (31.4 μM, 15 μL) run under non-reducing conditions; 5. *Kr*ACP (31.4 μM, 20 μL) run under non-reducing conditions; 6. Protein standards ladder; 7. *Kr*ACP (31.4 μM, 5 μL) run under reducing conditions; 8. *Kr*ACP (31.4 μM, 10 μL) run under reducing conditions; 9. *Kr*ACP (31.4 μM, 15 μL) run under reducing conditions; 10. *Kr*ACP (31.4 μM, 20 μL) run under reducing conditions.

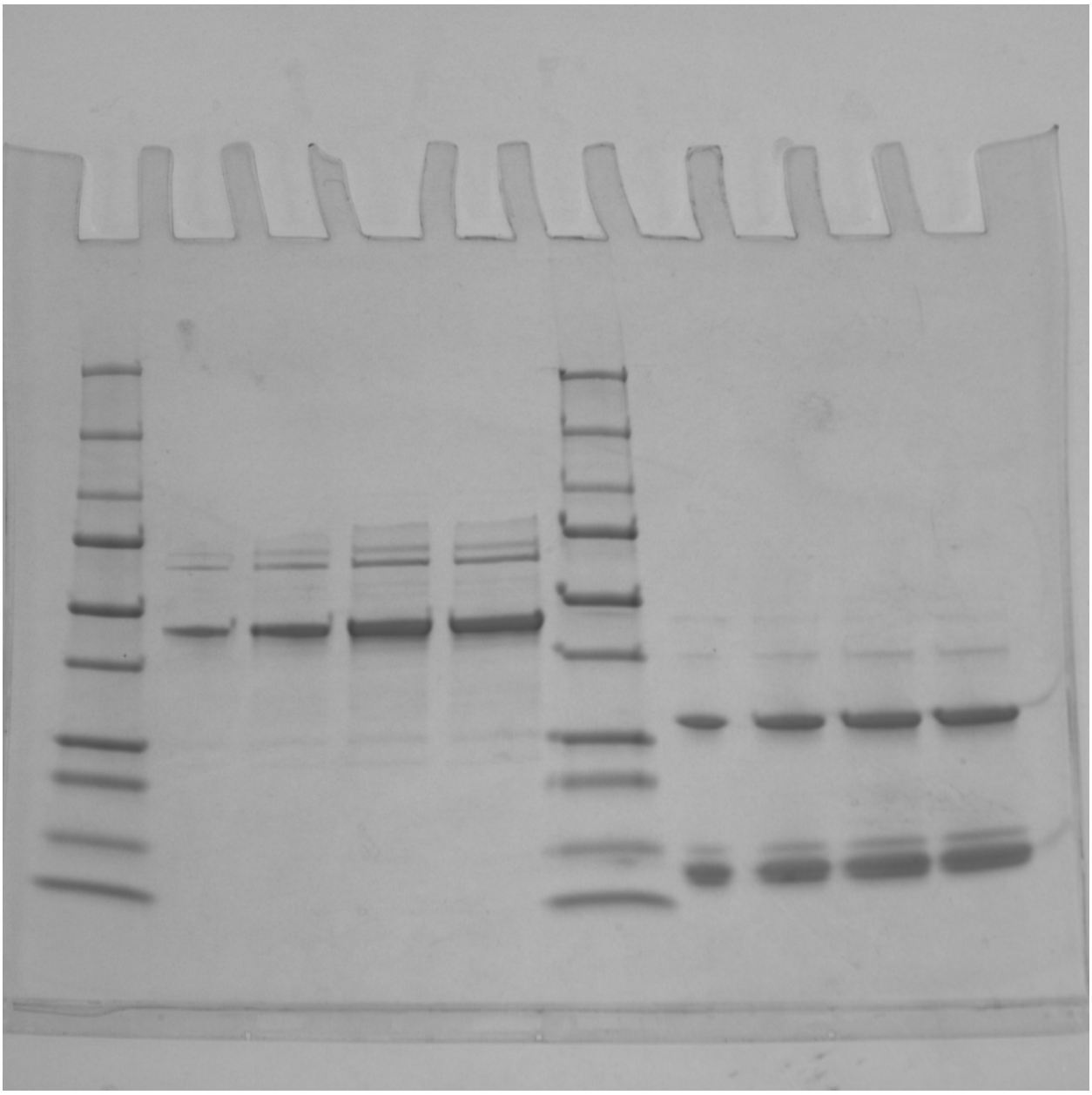

Fig S12. Uncropped Gel of Fig 2B. *Kr*KS-CLF purified under reducing conditions. 1. Protein standards ladder; 2. *Kr*KS-CLF (5 µL, 2.6 µM); 3. *Kr*KS-CLF (10 µL, 2.6 µM); 4. *Kr*KS-CLF (15 µL, 2.6 µM); 5. *Kr*KS-CLF (20 µL, 2.6 µM); 6. Protein standards ladder; 7. *Kr*ACP (5 µL, 25 µM); 8. *Kr*ACP (10 µL, 25 µM); 9. *Kr*ACP (15 µL, 25 µM); 10. *Kr*ACP (20 µL, 25 µM).

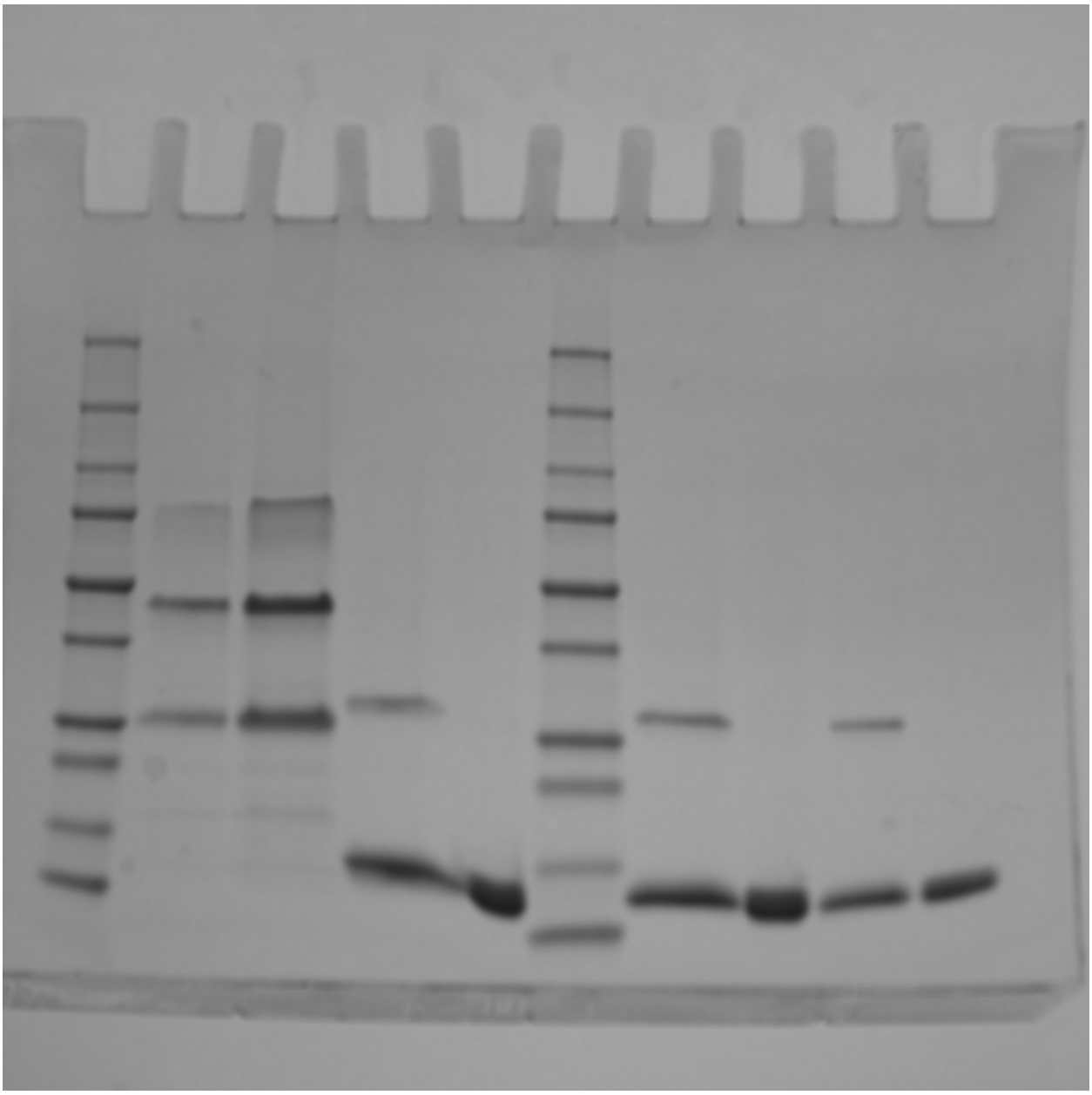

Fig S13. Uncropped Gel of S4A Fig. *Kr*KS-CLF purified under reducing conditions pre-SEC purification. 1. Protein standards ladder; 2. *Kr*KS-CLF (10 µL, 3.7 µM); 3. *Kr*KS-CLF (20 µL, 3.7 µM); 4. *Kr*ACP (15 µL, 25 µM); 5. *Kr*ACP + 5% v/v BME (15 µL, 25 µM); 6. Protein standards ladder; 7. *Kr*ACP (10 µL, 25 µM); 8. *Kr*ACP + 5% v/v BME (10 µL, 25 µM); 9. *Kr*ACP (5 µL, 25 µM); 10. *Kr*ACP + 5% v/v BME (5 µL, 25 µM).