**Title**

Characterization and Engineering of an Atypical Polyketide Synthase for Chemical Production



**Author**

Hagen, Andrew Ronald

**Publication Date**

2014

Peer reviewed|Thesis/dissertation

# Characterization and Engineering of an Atypical Polyketide Synthase for Chemical Production

by

Andrew Ronald Hagen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair
Professor Michelle Chang
Professor Cheryl Kerfeld

Fall 2014

Characterization and Engineering of an Atypical Polyketide Synthase for Chemical Production



by Andrew Ronald Hagen

# Abstract

Characterization and Engineering of an Atypical Polyketide Synthase for Chemical Production

by

Andrew Ronald Hagen

Doctor of Philosophy in Microbiology
University of California, Berkeley

Professor Jay D. Keasling, Chair

Polyketide synthases (PKSs) are remarkable enzymes. The diverse products from naturally ocurring PKSs have saved lives due to their useful medicinal qualities and inspired chemists to push the envelope on size and complexity in their synthetic pursuits. More recently, these enzymes have been modified to produce "unnatural" natural products—including close analogs of naturally occurring products as well as molecules wholly unknown to nature. These engineering pursuits have been enabled by an increasingly sophisticated understanding of PKS mechanics as well as technologies of the genomics era like inexpensive DNA sequencing and synthesis. In the doctoral work presented in this thesis, I will describe studies that have expanded our understanding of PKS biochemistry and demonstrated the production of an extremely economic and environmentally important small molecule from an engineered polyketide synthase.

Chapter 1 will begin with a discussion of the current state of the art and challenges associated with engineering these enzymes as well as offering opinions about routes forward to narrow the considerable gap between the promise and reality of engineered PKSs. This will be followed by a brief review of the PKS system responsible for borrelidin biosynthesis with a focus on features that make it unusual and attractive for engineering purposes.

In Chapter 2, I will present a study describing the first *in vitro* characterization of the unusual carboxyacyl-processing borrelidin polyketide synthase. It was found that the loading module has a specific requirement for carboxylated substrates of a specific stereochemistry, whereas the first extension module is comparatively promiscuous in its ability to extend carboxy- and descarboxy substrates with no apparent stereoselectivity.

The following chapter describes engineering efforts which exploit borrelidin PKS' diacid processing abilities towards the production of the commodity chemical, adipic acid. Extensive intermediate analysis using mass spectrometry revealed an unexpected difficulty in dehydrating the 3-hydroxyadipic-ACP intermediate, which was overcome in part by judicious chimeric junction selection and utilization of a dehydratase domain that

may better tolerate carboxylated substrates. Provision of a thioesterase led to the first demonstration of adipic acid production from an engineered polyketide synthase.

In the final data chapter, I describe a preliminary characterization of a module in the borrelidin cluster which performs three iterative condensations rather than the single extension performed in canonical type I modular PKS systems. It was proven *in vitro* for the first time that the isolated module is necessary and sufficient for iteration to occur and that the identity of the starter substrate has a profound effect on the distribution of extension products. Based on observations from time course experiments and co-incubation with the downstream extension module, a model for chain-length control in this unusual module is proposed.

The last chapter provides a brief summary of the principal findings described in this work and suggests experiments that will form the foundation of future efforts to improve our understanding of and ability to engineer these fascinating molecular assembly lines.

To the scientists whose shoulders I have stood upon and the friends and family whose shoulders I have leaned on

# Table of Contents

# **1.** Chapter 1 – Introduction

*Including material from published work:* Poust, S., **Hagen, A.**, Katz, L., & Keasling, J. D. (2014). Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Current opinion in biotechnology*, 30, 32-39.

Engineered modular polyketide synthases (PKSs) have the potential to be an extraordinarily effective retrosynthesis platform. Native PKSs assemble and tailor simple, readily available cellular acyl-CoAs into large, complex, chiral molecules (**Figure 1.1**). By successfully rearranging existing polyketide modules and domains, one could exquisitely control chemical structure from DNA sequence alone. As an example of the diverse biosynthetic potential of PKSs, we have concluded that approximately 20 of the roughly 150 commodity chemicals tracked by the petrochemical market information provider ICIS could be produced by mixing and matching naturally occurring PKS domains (ISIS Chemicals commodity and product finder; URL: http://www.icis.com/chemicals/channel-info-finder/). To form these chemicals, engineered PKSs would load acyl-CoAs accessible in *Escherichia coli* and other industrially relevant hosts, perform a programmed number of extension and then release products using previously published mechanisms. However, this potential has only just begun to be realized as the compounds that have been made using engineered PKSs (**Figure 1.2**) represent a small fraction of the potentially accessible chemical space. We envision a future in which a single design algorithm, using a molecule of interest as input, successfully combines natural PKS sequences to produce the desired molecule.

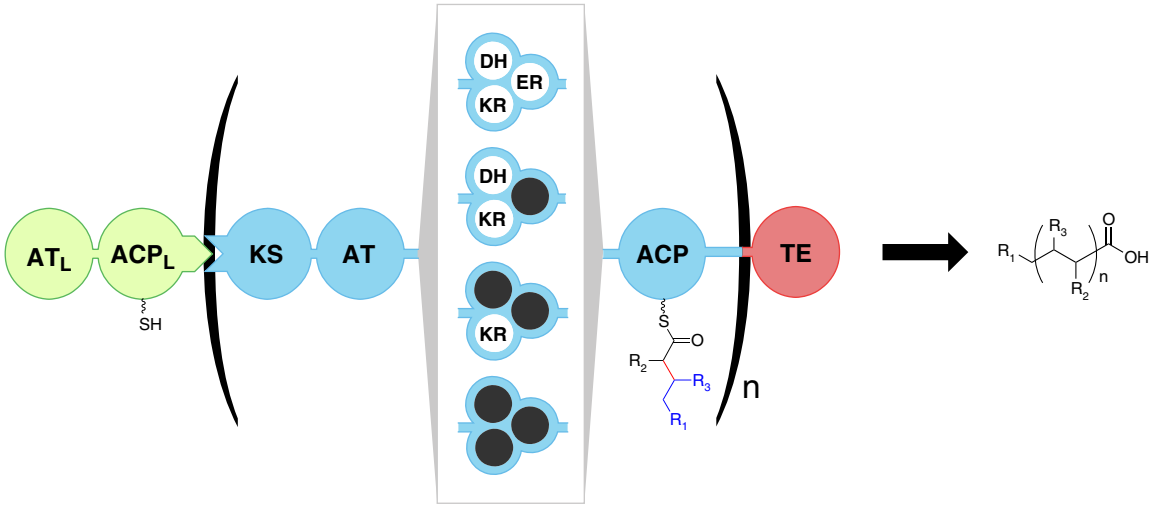



**Figure 1.1**. Symbolic representation of a general type I modular polyketide synthase and a general polyketide product. Biosynthesis begins with the selection of an acyl-CoA by the loading acyl transferase ($AT_L$) and subsequent transfer to the phosphopantetheine arm of the loading acyl carrier protein ($ACP_L$). A variable number of extension modules (represented as 'n') perform successive elongations of the enzyme-bound intermediates with downstream malonyl or methylmalonyl-ACPs (loaded by their cognate AT domains) via Claisen condensations catalyzed by ketosynthase (KS) domains. Accessory domains often present within a module such as ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains determine the final oxidation state and stereochemistry of the β-carbonyl attached to the ACP. The bond formed by the KS is represented on the ACP in red; the portion of the extended polyketide chain formed from the malonyl moiety is shown in black; and the portion from the upstream module(s) is shown in blue. In both the general polyketide product and in the intermediate attached to the ACP, $R_1$ represents the acyl chain from upstream module(s) or from the loading module and $R_2$ represents the side-chain on the extender unit, $R_3$ may be the oxygen of a ketone, an (*R*) or (*S*) hydroxyl or hydrogen, depending on the number of reducing domains in the particular module employed. Linker-mediated interactions promote chain transfer when modules are contained within different polypeptides. At the end of the synthesis, the final product is released as a free acid or as a lactone by a thioesterase. The bond entirely within the parentheses of the general product may be a double bond if a KR–DH pair is present within a module.

## 1.1. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform

### 1.1.1. The state of the art

Many compounds with no natural analogs have been produced using engineered PKSs. Examples include triketide lactones (McDaniel, 1997; Menzella et al., 2005; Pieper et al., 1995), pyrones (Hughes and Keatinge-Clay, 2011), linear branched carboxylic acids (Guo et al., 2010; Yuzawa et al., 2013), and advanced intermediates in the synthesis of epothilone (Menzella et al., 2010) (**Figure 1.2**). However, production titers from engineered PKSs are usually low in heterologous hosts such as *Streptomyces coelicolor* or *E. coli*, likely because the activities of the engineered enzymes are greatly reduced relative to their native counterparts, or the engineered enzymes are not solubly expressed in the host. For example, in the most extensive attempt to produce a wide variety of triketide lactones, Menzella and co-workers rearranged PKS domains from different clusters in a total of 154 bimodular combinations (Menzella et al., 2005). Remarkably, half produced detectable amounts of product, demonstrating the potential of PKSs as a platform to produce many different molecules. However, the great majority of functional constructs produced very little product and only a few produced compounds at titers above 10 mg/L. As a comparison, a 'wild type' strain of *Saccharopolyspora erythraea* can produce erythromycin at 660 mg/L (Hamedi et al., 2004). Menzella and coworkers subsequently used the knowledge of functional proteins gained from their earlier work to engineer trimodular PKS combinations (Menzella et al., 2007), where they were able to achieve a much higher proportion of functional proteins but, once again, the same problem remained with most combinations producing only a small amount of product.

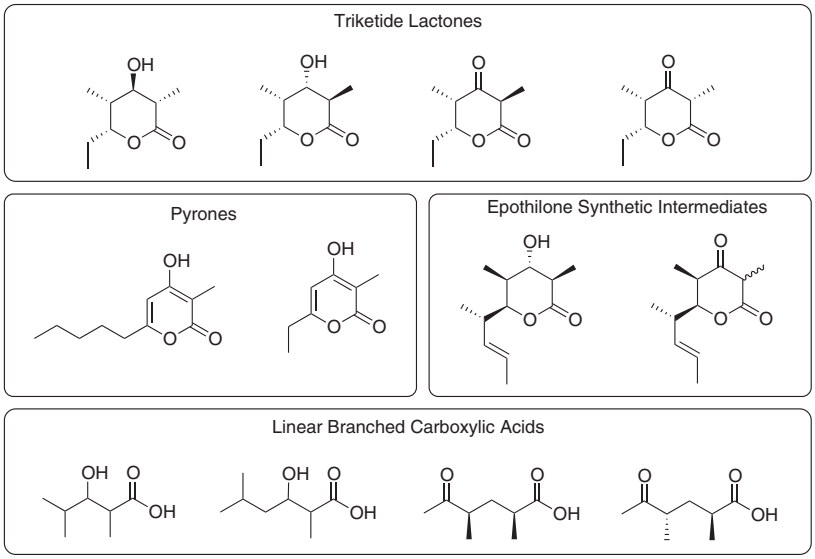



**Figure 1.2**. Example compounds with no natural analogs produced by engineered polyketide synthases. These include triketide lactones, pyrones, linear branched carboxylic acids, and advanced intermediates in the synthesis of epothilone.

While this work demonstrates modularity and the chemistry available for engineered PKSs, the knowledge of how to universally combine modules without compromising titers or kinetics remains to be elucidated. Expanding upon existing PKS engineering efforts to produce analogs of natural compounds (recently reviewed (Winter and Tang, 2012)), we believe that designing PKSs to make compounds of interest from the bottom up, quickly, and at titers of at least 10 mg/L will drive the field forward. This titer is high enough to allow for NMR-based structural characterization and for activity assays. Little work in the area of reverse engineering of PKSs to produce compounds of interest with no natural analogs has been published, but the tools to produce such molecules are clearly beginning to become available.

### 1.1.2. How to develop polyketide synthase based retrosynthesis?

We propose two main thrusts toward improving PKS-based retrosynthesis. First is the development of a detailed scientific understanding of PKS domain specificity, reaction mechanisms, structures, and domain–domain protein interactions. Knowledge in this area has advanced considerably in the last 20 years. For more information on this area, we direct the reader to recent reviews (Keatinge-Clay, 2012; Khosla, 2009). As part of these efforts, engineering specificity of individual domains, especially acyl transferases, is an ongoing pursuit, recently reviewed by Dunn and Khosla (Dunn and Khosla, 2013).

A second thrust toward PKS-based retrosynthesis is improvement of the design–build–test–learn cycle for PKSs (**Figure 1.3**), which we believe has great, as yet under-developed potential. Here we address areas within each of the parts of this cycle:

i. **Design**: reliable methods for engineering chimeric PKS proteins.
ii. **Build**: fast DNA assembly and soluble protein production in PKS expression hosts.
iii. **Test**: analytical methods for product detection, finding kinetic bottlenecks and high throughput screening.
iv. **Learn**: incorporating lessons learned though iterations of this cycle toward development of a single retro-synthesis algorithm.

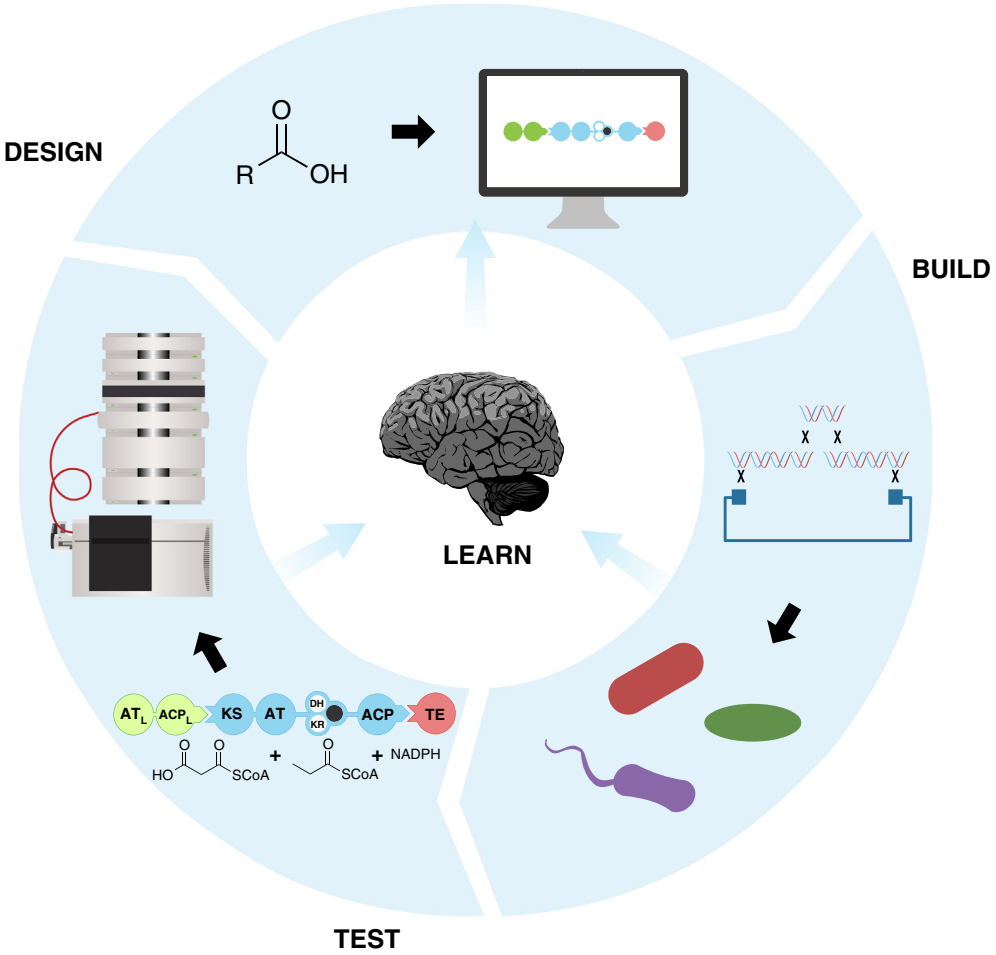



**Figure 1.**3. The iterative design-build-test cycle for engineering polyketide synthases. We define each portion of the cycle in the following ways. Design: sequence level design of chimeric PKS proteins to produce a desired molecule of interest. Build: DNA assembly of desired chimeric PKS sequences and soluble protein production in PKS expression hosts. Test: product detection via analytical methods and high throughput screening. Learn: incorporating lessons learned though iterations of this cycle toward development of a retrosynthesis algorithm.

### 1.1.3. Design

Developing PKS-based retrosynthesis will invariably involve the construction of chimeric proteins, and many different strategies have been successfully employed to produce them. Initial efforts relied on sequence conservation to determine the boundaries of domains, which were used to covalently fuse several different modules, as well as to swap KR domains, ACP domains and DH-KR didomains (McDaniel, 1997). A conserved YRVXW sequence between KS and AT domains was identified using limited proteolysis and has been used to construct KS–AT chimeric proteins (Chen et al., 2006; Kim et al., 2004b). Structure-based methods have also been used to determine functional domain boundaries. For example, an EEAPERE motif between the KS and AT domain was used by Chen and coworkers to dissociate these domains into functional stand-alone proteins (Chen et al., 2007).

Another strategy to facilitate nascent polyketide chain transfer between modules *in trans* is the incorporation of cognate interpolypeptide docking domains, often from heterologous clusters. Successful applications of this strategy include using class 1 docking domains from the erythromycin cluster to facilitate acyl-chain transfer between the ACP of module 1 and the KS domain of module 3 of the erythromycin cluster (Weissman, 2006), and incorporating linker domains from the phoslactomycin cluster between the ACP and KS of PikAI, the first polypeptide in the pikromycin cluster, creating two peptides that would be brought together post-translationally by the linkers to facilitate transfer between the dissected PikAI protein pairs (Yan et al., 2009). Most recently, cyanobacterial docking domains, which bind each other in a distinct manner and have a different structure from class 1 linker domains have also been used to facilitate transfer between the last two modules of the pikromycin PKS cluster *in trans* (Whicher et al., 2013). These constructions have facilitated transfer *in trans* with modules that either normally interact *in cis* (Weissman, 2006; Yan et al., 2009) or have their own native linker domains (Whicher et al., 2013); in the more general case of using docking domains to facilitate transfer between modules from different clusters, achieving titers above 10 mg/L remains difficult, as discussed above for the work of Menzella and coworkers (Menzella et al., 2005).

The techniques above employ coarse-grained strategies: they rely on rationally combining a few DNA sequences and assaying for any resulting activity, which may be low (additional examples of PKS engineering have been reviewed by Chen and Khosla (Chen and Khosla, 2008)). These existing strategies also have no clear path toward improving an inactive or marginally active construct. A more refined strategy that could improve the design of chimeric PKSs is profiling the active and inactive members of large libraries of chimeric PKSs, recombined at different domain junctions (KS–AT, ACP–KS, etc.) using different library construction techniques (as described in the build section). Sequence and activity information derived from library screening experiments could be then incorporated into a retrosynthesis algorithm, as detailed in the learn section. Natural selection currently produces new PKSs by screening recombination libraries: computational modeling suggests that recombination is of central importance in generating new polyketide compounds (Callahan et al., 2009). This is borne out in sequence analysis of PKS gene clusters, which suggests that recombination both within clusters and between both inactive and active clusters is a source of new metabolite diversity (Jenke-Kodama, 2005; Jenke-Kodama et al., 2006). By learning from selection experiments in the laboratory, one could develop more refined PKS construction techniques.

These kinds of selection experiments are beginning to be performed in a laboratory setting. Kim and coworkers utilized DNA shuffling and screening using a bioassay to substitute portions of the DEBS loading domain in the pikromycin cluster (Kim et al., 2004a). However, despite successes with mixing and matching strategies for iterative type II PKSs to make new metabolites (McDaniel et al., 1995), these experiments have yet to become mainstream in modular type I PKS engineering. Nonetheless, very recently Sugimoto and coworkers have utilized recombination within the type I aureothin cluster to produce the related compound luteoreticulin, for which there is no known naturally

producing strain (Sugimoto et al., 2014). Random mutagenesis can also be used to improve chimeric proteins after they are artificially recombined in the laboratory. Fischbach and coworkers used this technique to quickly improve non-ribosomal peptide synthetase (NRPS) chimeric proteins by screening with a bioassay (Fischbach et al., 2007). NRPSs are thiotemplated megasynthases similar to PKSs, but use amino acids instead of acyl-CoAs as extender units. It remains to be seen whether similar results would be obtained using directed evolution with PKS domain chimeras. More PKS laboratory evolution studies could create the knowledge necessary to inform design of a retrosynthesis algorithm.

### 1.1.4. Build

The cost of gene synthesis has recently fallen at an exponential rate and is now low enough to allow for codon optimization of genes the size of PKSs (Cost Per Base of DNA Sequencing and Synthesis; URL: http://www.synthesis.cc/library/carlson_cost%20per_ba- se_oct_2012.png).

However, synthesis can still be expensive for construction of large combinatorial libraries of PKSs. Developments in DNA software like j5 (Hillson et al., 2012), which designs scarless combinatorial constructs that can be assembled using techniques such as Gibson (Gibson et al., 2009), or Golden Gate Assembly (Engler et al., 2009) will be helpful in construction of combinatorial libraries of chimeric PKS proteins. Technologies like MAGE (Wang and Church, 2011), could also be used to generate insertions or deletions up to 30 base pairs or more in a high-throughput fashion. Cross platform robotic automation technologies such as PR–PR (Linshiz et al., 2013, 2014), which can take the output from programs like j5 and then set up the required PCRs using many robotics platforms may facilitate large-scale projects and new approaches to PKS engineering.

We have found that initial proof-of-principle studies with engineered PKSs are best done *in vitro*, either in cell lysates containing overexpressed protein(s) or preferably with purified protein(s). *In vitro* systems allow all required substrates (acyl-CoAs, reducing power, etc.) to be supplied exogenously, obviating the need for initial host engineering to supply these precursors *in vivo*. Additionally, *in vitro* systems allow for higher concentrations of protein than *in vivo* systems and reduces background signal, which may be necessary if the initial engineered protein constructs are sub optimally active. Engineering PKSs for production of a wide array of molecules will ultimately necessitate the use of domains and modules from a variety of source organisms. Such constructed chimeric PKSs will be inherently ‘non-monophyletic,’ complicating the choice of a suitable host for initial prototyping. Some examples of the use of chaperones to improve PKS solubility (Mutka et al., 2006) or achieve modest improvements in specific activity (Betancor et al., 2008) for *E. coli* expressed proteins have been reported. However, we feel that as the many factors that contribute to successful heterologous protein expression are still poorly understood, it is unlikely that a ‘super host’ (such as a highly engineered *E. coli* strain) capable of predictable success in soluble expression of all engineered PKSs will emerge in the near term; more probable is the need to prepare a ‘suite of production hosts’ (Baltz, 2010). As there is currently no *a priori* way to know which host would have the highest likelihood of soluble, functional expression, we focus on recent tools

that allow for rapid sampling of optimized, genetically tractable hosts to find ones amenable to pilot studies.

Natural product discovery through the cloning and expression of metagenomic DNA libraries has spurred the development of several vectors with broad host ranges for *E. coli*, *Pseudomonas* and *Bacillus*, among others (Aakvik et al., 2009; Kakirde et al., 2011; Troeschel et al., 2012). Commensurate with the creation of such vectors have been developments of high-throughput transformation and conjugation techniques (Martinez et al., 2004), which would be crucial for adequately sampling large combinatorial libraries of enzyme variants. The recent discovery of a plasmid and the development of inducible promoter systems in *Myxococcus* (Iniesta et al., 2012; Zhao et al., 2008) may lead to increased utility of this genus, as perhaps will the development of a thermophilic, fast-growing myxobacterial strain that successfully produced the NRPS/PKS-based myxochromide at levels substantially higher than the native host (Perlova et al., 2009). Recent examples of host engineering in *Streptomyces* have focused on measures designed to eliminate background production of native secondary metabolites and otherwise reduce the genome in order to increase production of heterologous products and simplify their detection (Gomez-Escribano and Bibb, 2011; Komatsu et al., 2010). Codon-optimization of synthetic DNA to allow for sampling of many different hosts may be challenging and require production of additional tRNAs which could be chromosomally integrated under constitutive promoters. Additional host features that may promote solubility would be expression of native and heterologous chaperones which have been shown in some cases to increase production of non-native products (Betancor et al., 2008; Mutka et al., 2006). Upon demonstration of *in vitro* production of a target compound from soluble protein, *in vivo* production would be attempted in the appropriate host; additional pathway engineering would be needed to supply required acyl-CoAs and a suitable phosphopantetheinyl transferase, if necessary. Efforts in engineering *E. coli* for type I polyketide production were recently reviewed by Yuzawa and coworkers (Yuzawa et al., 2012).

### 1.1.5. Test

In our experience, many engineered PKSs are insoluble when expressed in heterologous hosts like *E. coli*. Fusing fragments of green fluorescent protein to libraries of candidate chimeric proteins and screening for solubility using a split GFP system (Cabantous and Waldo, 2006) or by utilizing a colony filtration blot (Cornvik et al., 2005) could be used to prescreen libraries to remove insoluble members, further library screening could then be performed with lower-throughput analytical approaches such as GC/MS and LC/MS. This two-step approach would allow for larger libraries to be effectively screened. Once a soluble, active chimeric protein has been found, analytical techniques have been developed to find enzymatic bottlenecks through detection of covalent intermediates attached to acyl carrier protein peptides (Bumpus and Kelleher, 2010; Dorrestein et al., 2006). These techniques are accessible via a triple quadrupole mass spectrometer; many institutions have core proteomic facilities capable of performing these techniques, yet they appear underutilized in the PKS literature.

High-throughput screens for activities of individual domains have been exploited to engineer domains within modular synthases (recently exemplified in the work of Zhang and coworkers (Zhang et al., 2013)). However, engineering entire PKSs requires detection of the final product, and we believe mass spectrometry is the best technology for this application. Several new technologies have been developed that will aid in greatly improving mass spectrometry throughput. New platforms such as Agilent's RapidFire High-throughput MS Systems use solid phase extraction to achieve ten times the throughput of conventional mass spectrometry screening methods, with sample processing times between 6 and 10 s. Many solid phase chemistries are available and a variety of extraction/separation methods are in development. Nanostructure-Initiator Mass Spectrometry (NIMS) coupled with acoustic deposition has been used to screen enzymes acting on carbohydrates in a high-throughput format (Greving et al., 2012); building on NIMS demonstrations utilizing non-carbohydrate substrates (de Rond et al., 2013), this technology has potential application to screening PKSs libraries. Other developments in mass spectrometry such as nano-desorption electrospray ionization, which has been used to obtain mass spectra *in situ* from single bacterial colonies, also have potential to dramatically reduce the laborious sample preparation of traditional approaches (Watrous et al., 2012). The additional throughput of the next generation of mass spectrometry technologies has great potential to enable screening libraries of chimeric PKSs—a new approach to the PKS design–build– test–learn cycle.

### 1.1.6. Learn

At present, functional chimeric proteins are usually reported, but we suspect that nonfunctional or insoluble chimeric proteins are generally not published. This likely publication bias does not allow the issues of insolubility or lack of function to be addressed systematically, as failure can be as important for generating knowledge as success. We propose the creation of an online database for reporting of unpublished non-functional proteins, perhaps through research gate (URL: http://www.researchgate.net/), DOE's KnowledgeBase (URL: http://kbase.science.energy.gov/) or a similar site. This database would allow for failure analysis by bioinformatics methods, and should bring a new perspective to PKS engineering. Bioinformatic efforts so far (reviewed recently (Boddy, 2014)), focus on identifying or classifying PKS modules and domains, not necessarily yielding information that is directly applicable to engineering PKSs or in failure analysis. Expanding this database to include documentation of failures and successes at every stage of the design–build–test–learn cycle could serve as the foundation for a retrosynthesis algorithm. This kind of algorithm could allow non-experts to use PKSs for production of compounds of interest, expanding the role of engineered PKSs from niche science to a widespread chemical tool. As an example of the potential of developing algorithms based on an iterative design–build–test–learn PKS engineering approach, Chandran and coworkers developed an algorithm to activate several nonfunctional modules from their earlier Menzella work (Menzella et al., 2005) by swapping the nonfunctional KSs with other KSs that accepted the substrate of interest, observing 160-fold to 1300-fold increases in titers (Chandran et al., 2006).

### 1.1.7. Conclusions

The gap between the promise of type I PKSs as a retro-synthesis platform and the current state of the art is large. The promise continues to grow as new PKS modules and

chemistries are discovered, expanding the potential range of accessible chemicals. Techniques and technologies have been developed that make every stage in the retro-synthesis design–build–test–learn cycle more tractable. Many strategies for the design of chimeric proteins as well as methods for library design offer several options when constructing novel PKSs. DNA synthesis, the increasing power of DNA design software and laboratory automation are enabling scientists to make cloning much less of an obstacle. Many genetically tractable hosts are available to facilitate soluble protein expression. Mass spectrometers are getting better every year, and new techniques are being developed which allow higher throughput and less sample preparation, allowing for library-based approaches in the design–build–test–learn cycle. By both building on the existing polyketide science and harnessing technologies developing in other areas, we hope to see engineered PKSs with commercial applications in medicine and industrial products within 10 years.

## 1.2. Borrelidin: An atypical polyketide synthase

Borrelidin was originally isolated and characterized as an antibiotic effective against the genus *Borrelia* of the spirochete phylum in 1949 (Berger et al., 1949). Since then, several additional biological activities have been described for borrelidin including inhibition of angiogenesis, and as an anti-malarial and anti-viral agent (Dickinson et al., 1965; Ishiyama et al., 2011; Wakabayashi et al., 1997). Borrelidin is an 18-membered macrolide which was structurally characterized in 1989, fifty years after its discovery (Anderson et al., 1989) and found to contain a nitrile group as well as an unusual carboxycyclopentane moiety (**Figure 1.4**). In 2004, Olano and co-workers elucidated borrelidin's biosynthesis from a modular type I PKS gene cluster and made the striking observation that only six extension modules are somehow responsible for the formation of the nonaketide product—indicating the repeated use of one or more modules (Olano et al., 2004a).

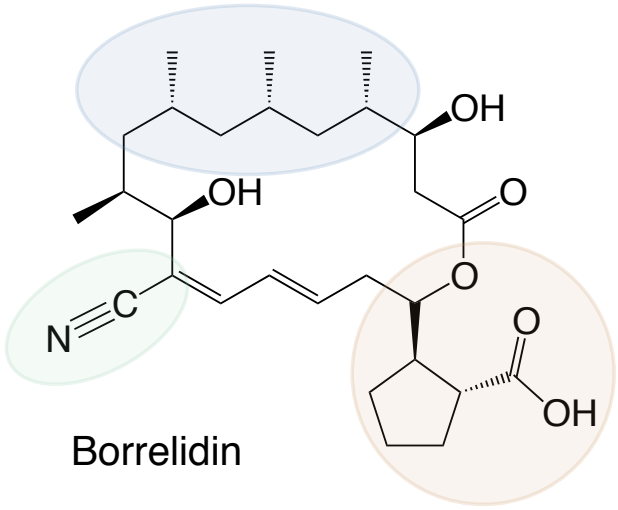



**Figure 1.4**. Structure of borrelidin with some unusual features highlighted. The *t*1,2-cyclopentanedicarboxylic acid starter is in brown, the nitrile moiety is in green and the trimethyl group installed by the iterative module is in blue.

The nitrile group is derived from a methyl branch at the C-12 position in a pathway involving the action of a cytochrome P450 monooxygenase (BorI), an aminotransferase (BorJ) and a dehydrogenase (BorK) (Olano et al., 2004b). Outside of the cyanogenic glycosides used in plant defense, nitrile-containing compounds are rare in biology (Fleming and Fleming, 1999) and it was recently shown that the nitrile moiety is essential for borrelidin's activity as a tRNA synthetase inhibitor (Schulze et al., 2014). Post-PKS modifications such as oxidations, glycosylations, halogenations etc. are, however, quite commonplace (reviewed by (Olano et al., 2010)).

In contrast, borrelidin is the only PKS known to begin with the *t*1,2-cyclopentanedicarboxylic acid starter, and furthermore only one other PKS from a *trans*-AT myxobacterial cluster (etnangien) is known that loads and extends a carboxyacyl starter (Menche et al., 2008). Diacid PKS products that do not result from a post-synthetic oxidation mechanism (e.g. in nystatin (Volokhan et al., 2006)), therefore are

exceedingly rare. In Chapter 2, I propose an evolutionary origin of borrelidin's ability to load such a 4-carboxy starter unit, however it remains unclear if specific adaptations were required to allow efficient extension and processing of this unusual intermediate.

Finally, the iterative use of a module that incorporates and fully reduces three methylmalonyl extender units challenges the notion of clear distinctions between modular and iterative type I PKS systems. Though a growing number of examples of such iterative modules in otherwise modular systems have been described in recent years (see Chapter 4 for examples), the overwhelming majority of type I modular PKS clusters described to date obey the rules of colinearity.

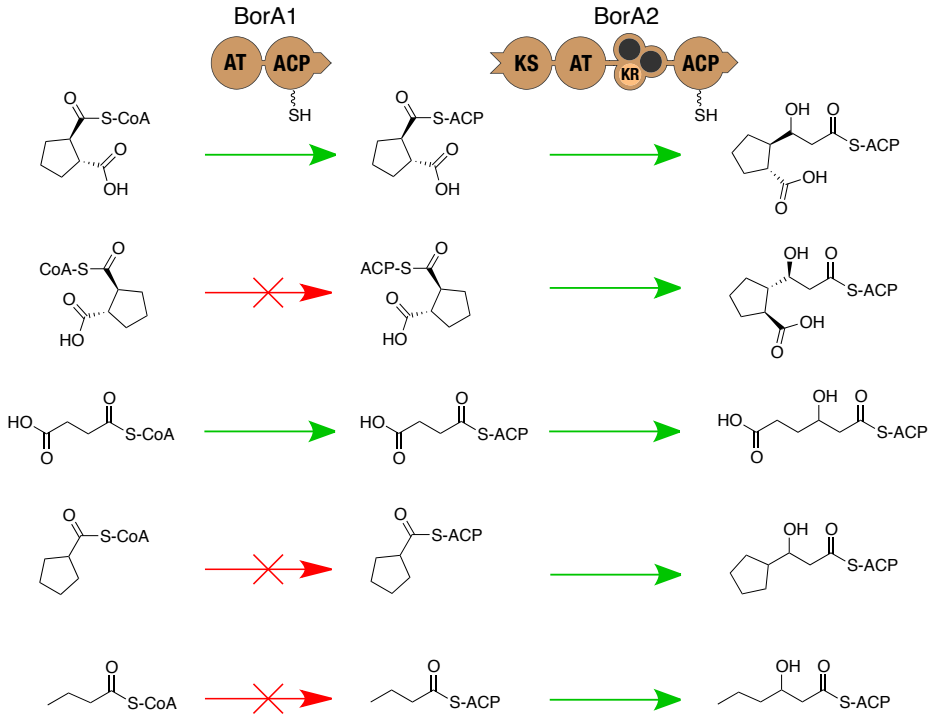
Despite the wealth of molecules that can be produced by typical PKS systems, understanding and exploiting the idiosyncrasies of the borrelidin synthase should open the door to incorporating expanded chemical functionalities (diacids) and provide a more facile means of production for long-chain PKS products.

# **2.** Chapter 2 – Characterization of carboxyacyl substrate tolerance in the loading and first extension modules of Borrelidin PKS

*Based on published work:* **Hagen, A.**, Poust, S., de Rond, T., Yuzawa, S., Katz, L., Adams, P. D., ... & Keasling, J. D. (2014). In Vitro Analysis of Carboxyacyl Substrate Tolerance in the Loading and First Extension Modules of Borrelidin Polyketide Synthase. *Biochemistry*, *53*(38), 5975-5977.

## 2.1. Abstract

The borrelidin polyketide synthase (PKS) begins with a carboxylated substrate and, unlike typical decarboxylative loading PKSs, retains the carboxy group in the final product. The specificity and tolerance of carboxyacyl substrate incorporation in type I PKSs have not been explored. Here, we show that the first extension module is promiscuous in its ability to extend both carboxy- and non-carboxy- acyl substrates. However, the loading module has a requirement for substrates containing a carboxy moiety. Unlike typical 3-carboxy loading AT domains (e.g. that load malonyl-CoA and methylmalonyl-CoA), the carboxy group does not undergo decarboxylation *in situ*. Thus, the loading module is the basis for the observed specific incorporation of carboxylated starter units by the borrelidin PKS.



## 2.2. Introduction

Type I polyketide synthase (PKS) enzymes are modular "assembly lines" which give rise to large and complex molecules through successive condensations of 3-carboxyacyl-CoA substrates (recently reviewed (Khosla et al., 2014)). The "textbook" understanding of polyketide biosynthesis initiation in type I PKS systems proceeds through two main mechanisms (reviewed (Hertweck, 2009)). In the first, as is present in the well-studied erythromycin PKS (Khosla et al., 2007), an AT-ACP didomain loading module selects an acyl starter unit which is then transferred to and extended by the downstream module. In the second mechanism, the AT loads a 3-carboxyacyl substrate (e.g., malonyl- or methylmalonyl-CoA), which is subsequently decarboxylated by an accompanying KSQ domain prior to chain translocation. Besides the presence of an upstream KSQ domain, AT domains in such a tripartite system are identified by the presence of a conserved arginine residue thought to interact with the carboxylate moiety of the substrate (Long et al., 2002). Some loading modules (e.g. in lipomycin PKS) have been shown to promiscuously load a variety of acyl-CoA substrates (Yuzawa et al., 2013), however there have been few such *in vitro* studies examining substrate tolerance in early PKS modules.

The structure of borrelidin suggests it uses a different mechanism than the two described above for polyketide biosynthesis initiation (**Figure 2.1**) (Lumb et al., 1965). Borrelidin is to our knowledge the only *cis*-AT PKS proven to begin with a loaded carboxyacyl substrate (*trans*1,2-cyclopentanedicarboxylic acid, *t*1,2-CPDA) that is extended without *in situ* decarboxylation by a KSQ domain. This was demonstrated by the production of borrelidin after feeding *t*1,2-CPDA racemate to strains deficient in starter unit biosynthesis (Olano et al., 2004b). Subsequent work demonstrated that the borrelidin PKS can process a variety of non-natural starter substrates by the production of several borrelidin analogs in a starter unit deficient strain fed with a panel of over 40 acyl- and carboxyacyl acids (Moss et al., 2006). Only diacids with a particular geometric arrangement of carboxylates were found to serve as starter units, which the authors suggest is a result of selectivity imposed by the PKS loading module (BorA1). However, in the *in vivo* system, it was not possible to determine whether this selectivity resulted from failure of the substrate to enter the cell; inability of the endogenous CoA ligase to recognize the substrate; diversion of the substrate or its CoA thioester into other metabolic pathways; nor could the authors determine at what stage in the polyketide biosynthesis this selectivity is imposed.

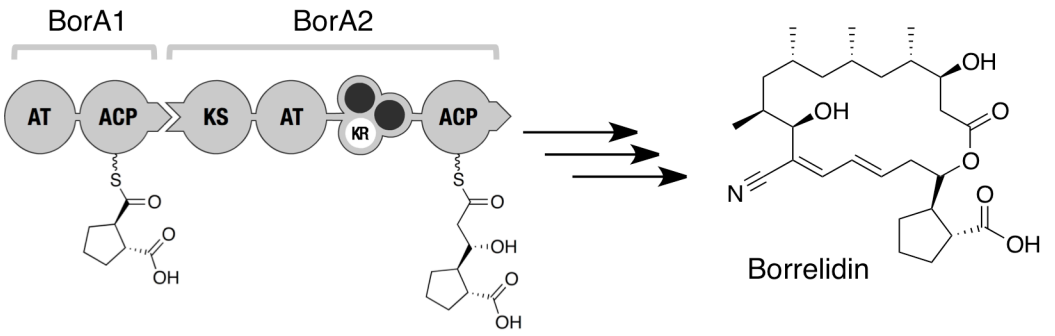


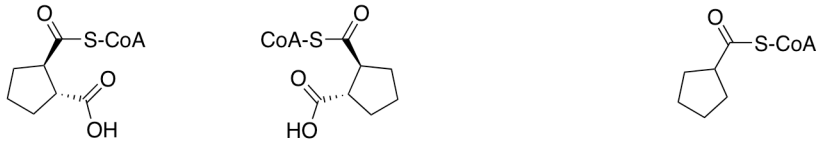

**Figure 2.1.** Partial schematic of borrelidin biosynthesis highlighting the loading and first extension modules

To better understand substrate selectivity towards acyl and carboxyacyl substrates in the borrelidin PKS, we heterologously expressed and purified the loading and first extension modules and tested their activity *in vitro* with a panel of CoA substrates. Because pendant polyketide products are covalently attached to the phosphopantheteine arm of the acyl carrier proteins, tandem mass spectrometry methods such as the "PPant ejection assay" (Meluzzi et al., 2008) provide a convenient and sensitive way to qualitatively monitor enzyme activity. Briefly, the tryptic peptide containing the acylated-PPant prosthetic group is selected by a quadrupole mass filter and subjected to collision induced dissociation (CID) to eject the (acyl)PPant arm which is detected as a daughter ion via a quadrupole or time of flight mass analyzer.

## 2.3. Results

### 2.3.1. Loading and extension assays reveal BorA1's strict requirement for carboxyacyl substrates

We performed loading/extension assays in which the loading and extension modules (BorA1 and BorA2) were co-incubated with different starter CoA substrates along with malonyl-CoA and NADPH (the extender unit and reducing substrate for BorA2). We tested *t*1,2-CPDA CoA (CPDA-CoA), borrelidin's natural starter along with an aliphatic analog, cyclopentanemonocarboxylic acid CoA (CPMA-CoA) (**Figure 2.2**).




(*1R*, *2R*) and (*1S*, *2S*)-cylcopentanedicarboxylic acid CoA ("CPDA-CoA")

cyclopentanemonocarboxylic acid CoA ("CPMA-CoA")

**Figure 2.2** Synthetic substrates used in this study

As expected, when provided the CPDA-CoA substrate, its extended/reduced product, the 3-(2-carboxycyclopentyl)-3-hydroxypropionyl moiety (CPDA-ext) accumulated on the BorA2 ACP (**Figure 2.3B**) in a time-dependent manner. In contrast, the CPMA-CoA substrate was not accepted by the loading AT as neither CPMA-ACP nor its extension product (CPMA-ext) were detected (**Figure 2.3 A,B**). Performing the loading/extension assay with two additional substrates—succinyl-CoA and another aliphatic analog, butyryl-CoA surprisingly revealed that while succinyl-ACP was not detected on the loading ACP (which we attribute to instability, see **Figure 2.4**), we infer its loading and extension by the detection of its extension product, 3-hydroxyhexanedioic-ACP (Succ-ext). As with the CPMA-CoA substrate, when butyryl-CoA was tested, neither butyryl-ACP nor its extension product (Buty-ext) were detected.

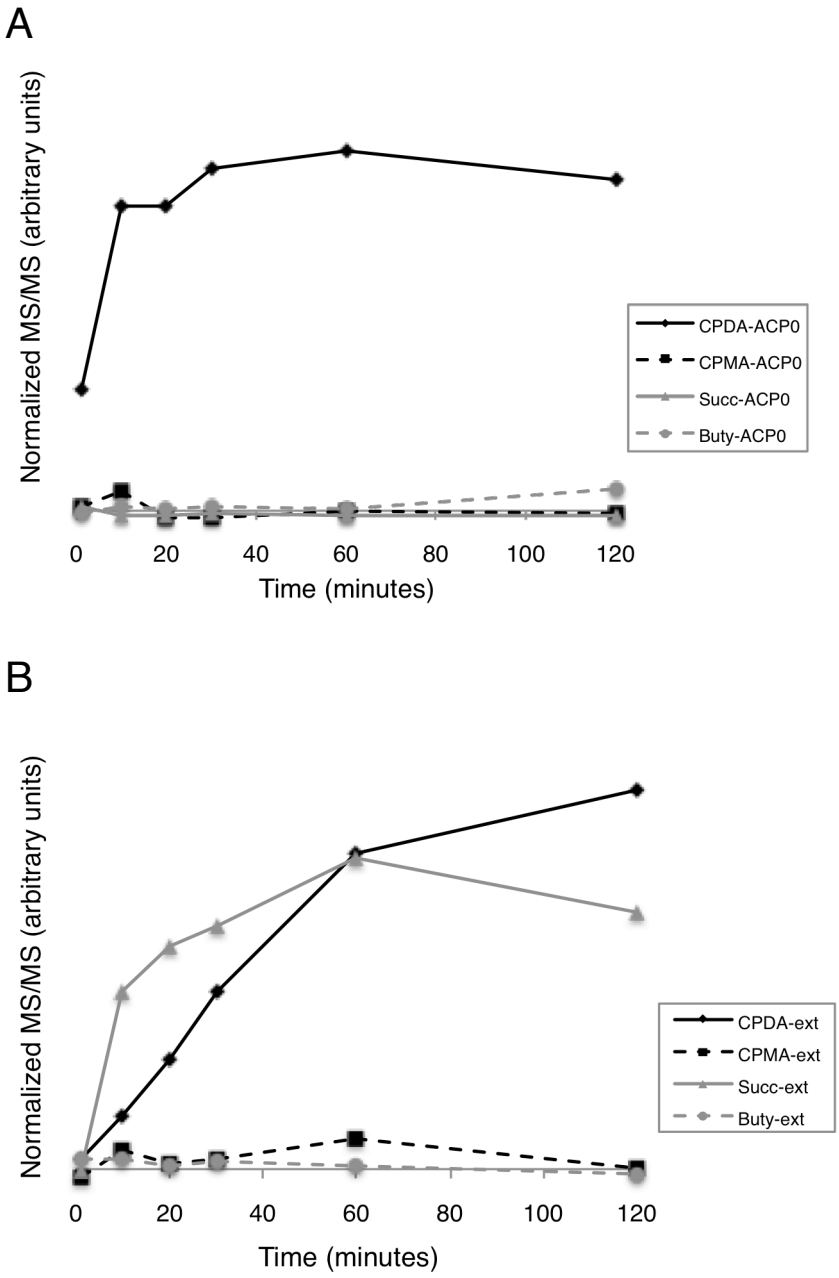



**Figure 2.3.** Loading/extension assay with BorA1 and BorA2 showing normalized MS/MS transitions of (A) ACP0 and (B) ACP1

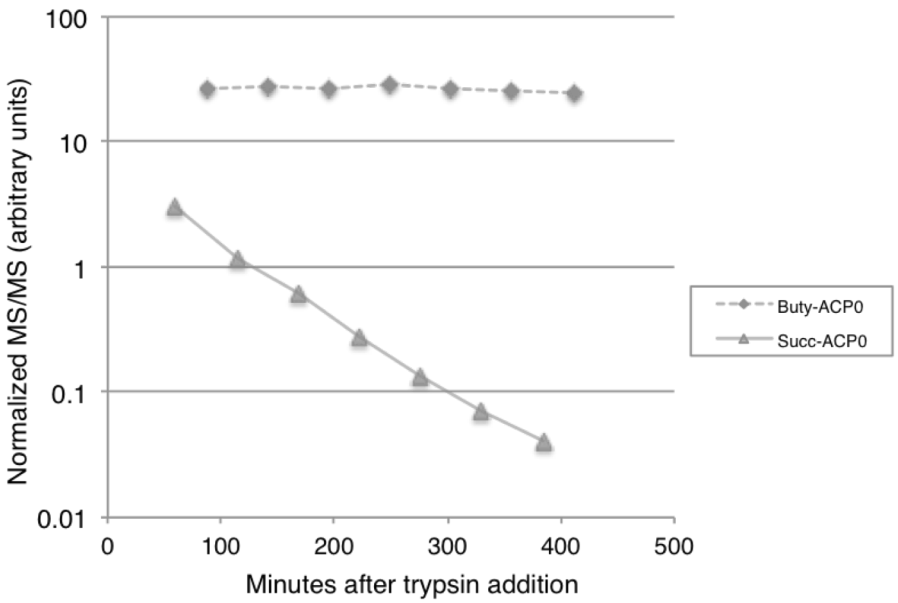



**Figure 2.4.** Comparison of butyryl-ACP0 and succinyl-ACP0 stability. Shown are normalized respective MS/MS intensities for acyl-ACP0 transitions. Note log scale on y-axis

### 2.3.2. Chain extension assays reveal broad substrate tolerance of BorA2

We next focused on the substrate specificity of BorA2 by performing chain extension assays using acylated loading ACP (ACP0) as a starting substrate. In these experiments, the loading ACP was expressed in its apo form as a monodomain and then phosphopantetheinylated *in vitro* with the same panel of four CoA substrates via Sfp (Quadri et al., 1998). These (carboxy)acyl-ACP substrates were incubated in the presence of (holo) BorA2 and product formation was monitored using the PPant ejection assay. Because both CPDA-CoA enantiomers generated during the synthesis were presumably loaded onto the apo ACP by Sfp, we investigated if BorA2 preferentially extends one enantiomer over the other by refining the liquid chromatography method with a shallower gradient in an attempt to resolve the different possible diastereomer products.

Interestingly, examination of CPDA-ACP0 extension revealed two clear peaks repeatedly observed in an approximately 2:3 ratio (**Figure 2.5A**, top panel) and analysis with high-resolution time of flight mass spectrometry confirmed the two peaks have the same *m/z* value and isotopic distribution (expected *m/z* 967.1270, z=+3; **Figure 2.6A,B**). In contrast, when examining the CPDA samples from the loading/extension experiment using this method, only one peak was detected (**Figure 2.5A**, lower panel), suggesting only one enantiomer (presumably *1R, 2R*) was loaded by the AT. Surprisingly, extension products for the CPMA- and butyryl-ACP substrates were also detected at levels comparable to those of the native CPDA-ACP extension product (**Figure 2.5B**). Taken together, these data show that while the loading AT domain of BorA1 only accepts

dicarboxylates with a stereospecific arrangement as proposed by Moss et. al (Moss et al., 2006), BorA2 is comparatively permissive in its ability to extend substrates with and without carboxy groups or ring structures and with little apparent stereoselectivity. No β-keto-ACP intermediates were detected for any substrates, indicating the ketoreductase activity is rapid and substrate agnostic (data not shown).

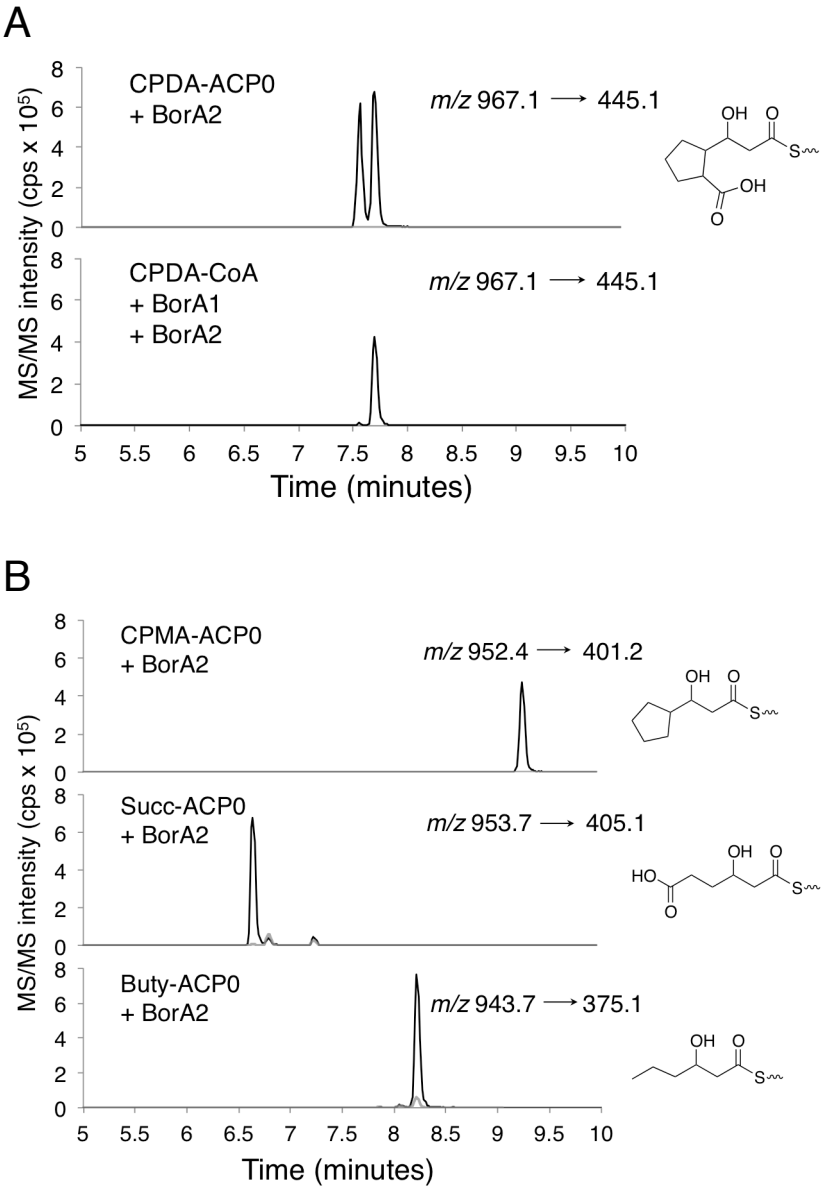
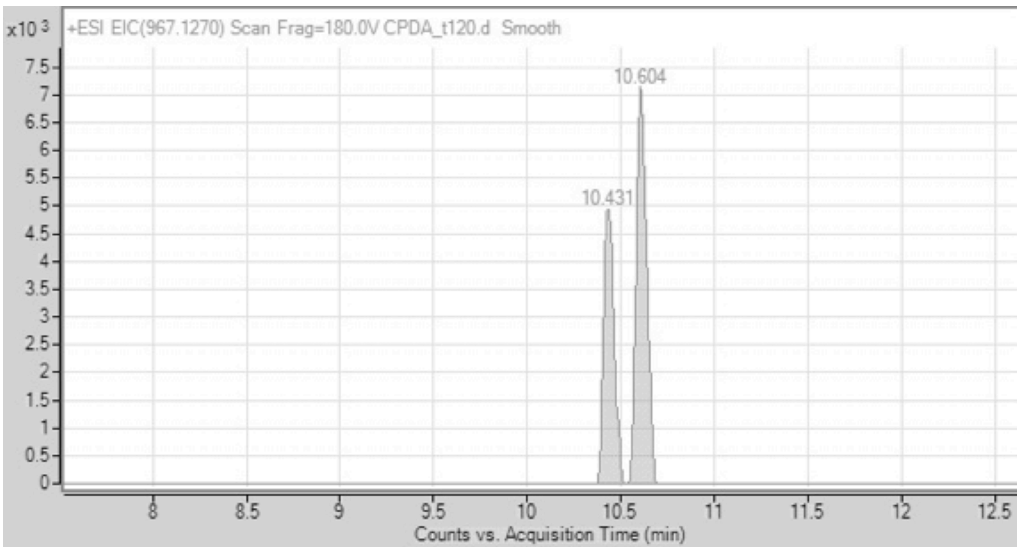



**Figure 2.5.** MS/MS chromatograms and structures of extension products (stereochemistry not shown for simplicity). (A) Comparison of CPDA-ext products when generated by extension of CPDA-ACP0 synthetically loaded by Sfp (top panel) or from AT domain-mediated loading (bottom panel). (B) Chromatograms of respective products for different acyl-ACP0 extension reactions. Negative controls (where appropriate) are shown in grey lines

A




B

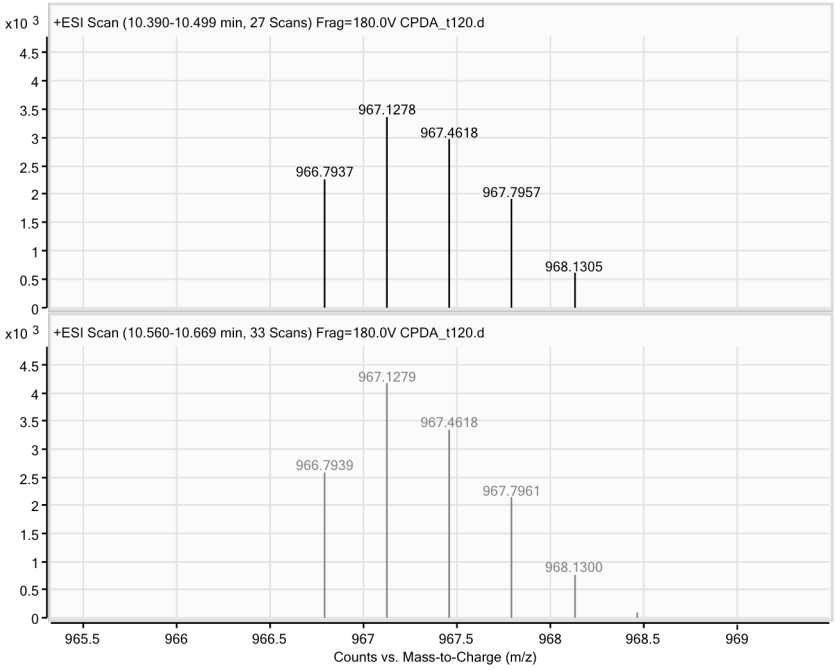



**Figure 2.6.** (A) Extracted ion chromatogram (967.127 -/+ 0.02) of translocation/extension reaction using CPDA-ACP0 as substrate. (B) Mass spectra of above peaks (top panel: peak 1, RT 10.4; bottom panel: peak 2, RT 10.6)

### 2.3.3. The N-terminus of BorA1 is a degenerate ketosynthase domain

We observed that some PKS domain analysis programs (e.g. MAPSI (Tae et al., 2009), SBSPKS (Anand et al., 2010)) but not all (e.g. DoBiscuit (Ichikawa et al., 2013))

annotate the first approximately 300 residues of BorA1 as a KS domain. The presence of such a domain would indicate that BorA1's domain structure is that of a typical KSQ-harboring loading module. A multiple sequence alignment was created using the putative KS domain from BorA1 and biochemically verified KSQ domains (**Figure 2.7**). This alignment indicates BorA1's N-terminus shares homology with authentic KSQ domains, however the lack of a recognizable active site indicates that it is not catalytically competent to perform decarboxylation.

```
BorA1_1-310         ----------MTGSAVSAPFLQ-----PPEPVSGHSERKS-----------------D
MonAI_17-430        DPIAVVGMACRLPGAPDPDAFWRLLSEGRSAVSTAPPERRRADSGLHGPGGYLDRIDGFD
NidA1_23-430        DAIAVIGMSCRFPGAPGTAEFWQLLSSGADAVVTAADGRRR---------GTIDAPADFD
TylORF1_43-451      GEIAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWR---------AALEDHAGFD
                              :.*:.    *                .                  *

BorA1_1-310         PVLLVGAGRRARMADAVR--------AAGAQAGIDPAVLRRTRAT------------LIT
MonAI_17-430        ADFFHISPREAVAMDPQQRLLLELSWEALEDAGIRPPTLARSRTGVFVGAFWDDYTDVLN
NidA1_23-430        AAFFGMSPREAAATDPQQRLVLELGWEALEDAGIVPESLRGEAASVFVGAMNDDYA-TLL
TylORF1_43-451      AGFFGMNARQAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYA-VLT
                    . ::    * *   *. .          *  :*** *  *    :              :

BorA1_1-310         AGSAGAAG--------------RLAAALRLTGATISLDTRETPTLLALHLAAQALRAGDT
MonAI_17-430        LRAPGAVTRHTMTGVHRSILANRISYAYHLAGPSLTVDTAQSSSLVAVHLACESIRSGDS
NidA1_23-430        HRAGAPTDTYTATGLQHSMIANRLSYFLGLRGPSLVVDTQQSSSLVAVALAVESLRGGTS
TylORF1_43-451      RRSAVSAGGYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGET
                      :  ..               *::    * *.:: :*: ::.:*:*: ** :::* * :

BorA1_1-310         SYAVVGAE---LPDGNCALILARQSAATAEG------AVPQAIVRTTTADRTTTADHAPA
MonAI_17-430        DIAFAGGVNLICSPRTTELAAARFGGLSAAGRCHTFDARADGFVRGEGGGLVVLKPLAAA
NidA1_23-430        GIALAGGVNLVLAEEGSA-AMERVGALSPDGRCHTFDARANGYVRGEGGAIVVLKPLADA
TylORF1_43-451      SLAVAGGVNLILTEESTT-VMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAA
                    . *..*.     .         * .. :. *      * .:. **   .  ..      *

BorA1_1-310         PDDHG---SPAREAPHATRTLSPGITQAPAEGFPGLLATLHDDTPLRPTAVTE---HGS-
MonAI_17-430        RRDGDTVYCVIRGSAVNSDGTTDGITLPSGQAQQDVVRLACRRARITPDQVQYVELHGTG
NidA1_23-430        LADGDRVYCVVRGVATGNDGGGPGLTVPDRAGQEAVLRAACDQAGVRPADVRFVELHGTG
TylORF1_43-451      LADGDRVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTG
                      * .   .  .  .  .     .:* .   .   ::      : : .  *     **:

BorA1_1-310         -------DATTVLVLLDQPQDAAPAAPLPWVVSAPHTRALRATA-------ATLAVHLD-
MonAI_17-430        TPVGDPIEAAALGAALG--QDAARAVPLAVGSAKTNVGHLEAAAGIVGLLKTALSIHHRR
NidA1_23-430        TPAGDPVEAEALGAVYGTGRPA--NEPLLVGSVKTNIGHLEGAAGIAGFVKAALCLHERA
TylORF1_43-451      TRAGDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGE
                           :* :: .  .    :    **      .:   * .:*       ::*.:.

BorA1_1-310         -------TTPAAPADVAHTLLTARPDRHRAAVVGADRATLTDGLRALATGGDAPHL----
MonAI_17-430        LAPSLNFTTPNPAIPLADLGLTVQQDLADWP--RPEQPLIA-GVSSFGMGGTNGHVVVA-
NidA1_23-430        LPASLNFETPNPAIPLERLRLKVQTAHAALQPGTGGGPLLA-GVSAFGMGGTNCHVVLEE
TylORF1_43-451      LVPSLNFSTPNPDIPLDDLRLRVQTERQEWNE-EDDRPRVA-GVSSFGMGGTNVHLVIAE
                           ** .   :    * ..             . :: *: ::. **   *:

BorA1_1-310         --
MonAI_17-430        --
NidA1_23-430        TP
TylORF1_43-451      AP
```

**Figure 2.7.** Multiple sequence alignment comparing the putative KS domain from BorA1 to authentic KSQ domains (monensin, niddamycin and tylactone). Position of active site glutamine residue is boxed

## 2.4. Discussion

In the study described in this chapter, we have provided the first biochemical characterization of a type I PKS loading and extension module system that processes a carboxyacyl substrate. We showed that BorA1 has a requirement for carboxyacyl-CoA substrates of a particular stereospecificity whereas BorA2 is comparatively promiscuous in its ability to extend both acyl- and carboxyacyl substrates with no apparent stereoselectivity. It remains unknown whether BorA2 is unique in its ability to extend carboxyacyl substrates and would be an interesting line of investigation to pursue using methods described in this study.

Demonstration of succinyl-CoA loading and extension in these in vitro studies is seemingly at odds with the precursor-directed biosynthesis findings of Moss and co-workers wherein the succinate analog of borrelidin was never detected. One possibility is that borrelidin PKS expression only takes place when succinyl-CoA levels are low enough to not compete with CPDA-CoA. Alternatively, it is possible that succinyl-CoA is occasionally loaded but is subjected to editing, perhaps by the type II thioesterase in the borrelidin cluster whose function is as yet, unassigned.

As the majority of PKS starter units are achiral, stereospecificity in loading ATs has not been extensively investigated, though it is well-documented in extension ATs (Marsden et al., 1994). The biosynthetic route to *trans*-1,2-CPDA-CoA remains hypothetical, so it is unclear if the loading AT in borrelidin PKS evolved to discriminate against one of two stereoisomers that may be present, or if only one species is biosynthesized and the stereoselectivity is incidental. Nevertheless, this *in vitro* study has demonstrated that a loading AT domain can discriminate between different stereoisomers and therefore has the capacity to impose another layer of specificity in starter unit selection.

The observation that the N-terminus of BorA1 is a degenerated KSQ domain and the presence of the aforementioned conserved arginine in the AT domain is highly suggestive of an evolutionary origin of this atypical module. BorA1 may have originally functioned as a typical 3-carboxyacyl loading/decarboxylating module whose substrate specificity shifted to a 4-carboxy substrate such as CPDA-CoA. Because canonical KSQ-mediated decarboxylation is not possible on such substrates, selective pressure on the BorA1's putative KSQ domain may have have been relaxed, leading to its degeneration.

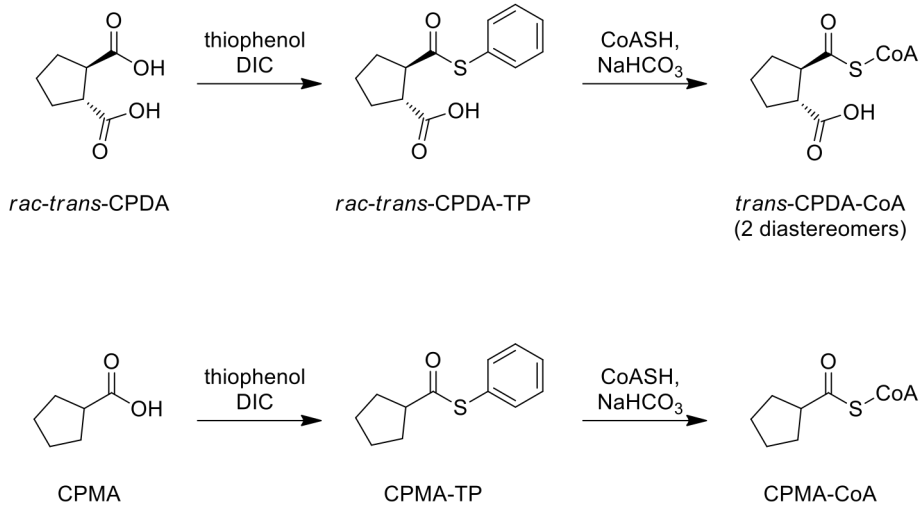
Finally, we have shown that BorA1 and BorA2 can produce a range of partially reduced linear dicarboxylic acid intermediates. Of particular interest is the 3-hydroxyadipic-ACP intermediate, which is three biochemical transformations from the commodity chemical, adipic acid. Having established borrelidin PKS as a viable starting point, the next chapter will be dedicated to engineering this system via reductive loop engineering and addition of a thioesterase domain towards the production of this highly relevant molecule.

## 2.5. Materials and Methods

### 2.5.1. Reagents and Chemicals

HisPur cobalt resin was purchased from Thermo Scientific, Bradford reagent was from Bio-Rad and SDS-PAGE gels from Life Technologies. Solvents (hexanes, ethyl acetate, dichloromethane and methanol) were purchased from EDH; all other reagents were purchased from Sigma-Aldrich.

### 2.5.2. Chemical synthesis and NMR data




**Supplmenetary Figure 2.1**. Synthetic approach

### General synthetic procedures

$^{1}H$ NMR and $^{13}C$ NMR were obtained on a Bruker AVQ 400MHz spectrometer at the UC Berkeley College of Chemistry NMR facility, funded in part by NSF grant CHE-0130862. Chemical shifts are reported in ppm relative to residual solvent signal ($\delta^{1}H$=7.26 and $\delta^{13}C$=77.16 for $CDCl_3$, $\delta^{1}H$=4.79 $D_2O$).

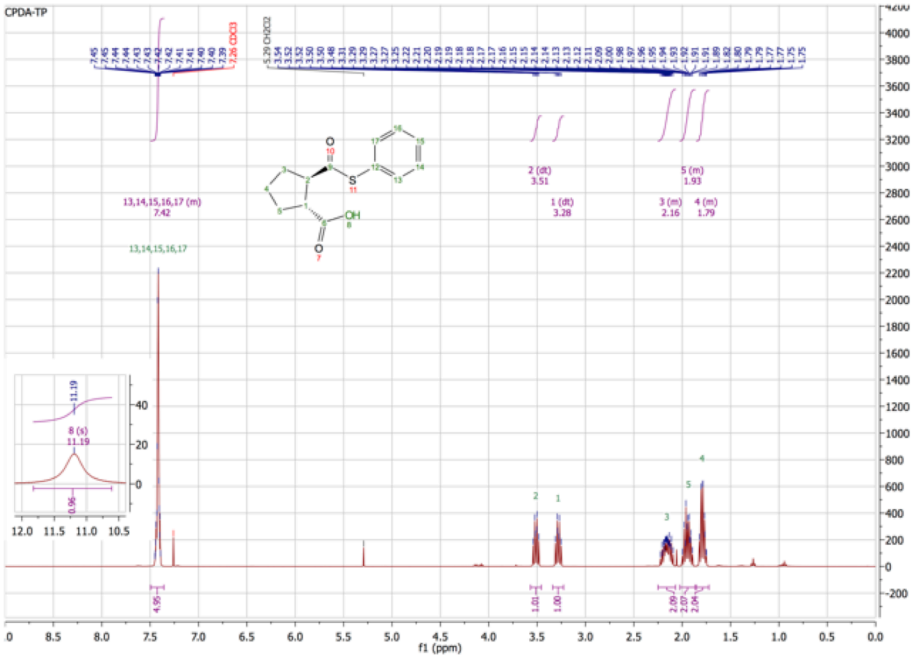
Column chromatography was performed on a Teledyne Isco Combiflash Rf, with RediSep Rf Gold 24g normal phase silica columns.
Preparative HPLC was performed on an Agilent 1260 infinity series instrument with a ZORBAX SB-C18 semi-prep column (50mm × 9.4 mm internal diameter, 5 μM particle size).

Separation conditions: Solvent A = 50 mM NaH2PO4 (pH = 4.2) in water, Solvent B = methanol. The following gradient was used: 0 min, 2.5% B; 2.5 min, 2.5% B; 11 min

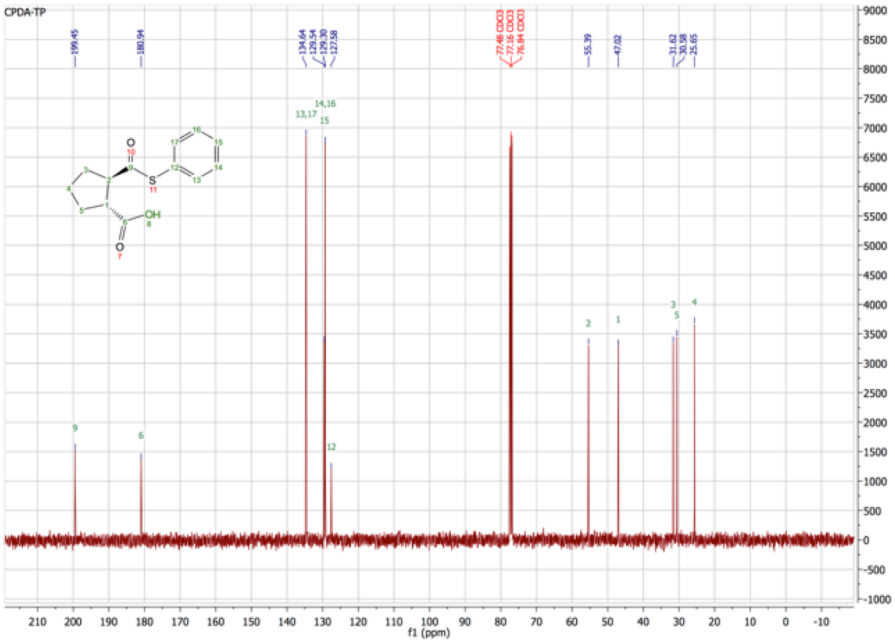
11% B; 16min, 30% B; 17min, 95% B; 19min, 95% B; 20min, 2.5% B; 22min, 2.5% B. Flow rate was 10mL/min. Fraction collection was triggered by Abs(260).
Desalting conditions: Solvent A = water, Solvent B = methanol. The following gradient was used: 0 min, 2.5% B; 2.5 min, 2.5% B; 5 min, 15% B; 6 min, 95% B; 8 min, 95% B; 8.50 min 2.5% B; 10 min, 2.5% B. Flow rate was 10mL/min. Fraction collection was triggered by Abs(260nm).
Acyl-CoA working stocks were made up at 1.4mM in water**,** using absorbance at 260 nm for quantification ($\varepsilon$ = 16,400 $mM^{-1}$ $cm^{-1}$).

**CPDA-TP**

A 100mL flame-dried round-bottom flask with 20 mL acetone was cooled to 0 °C on ice, and charged with 2 mmol *trans*-cyclopentanedicarboxylic acid (316 mg) and 1.5 eq. thiophenol (308µL). 1.125 eq. diisopropylcarbodiimide (DIC) (312 µL) was diluted into 4.5 mL acetone and added to the stirring solution dropwise over the course of 2 h, during which the solution turned from colorless to cloudy white. The solution was stirred on ice for another 2 h, and then quenched with 100mL 5% aqueous ammonium chloride. The mixture was extracted with 2 × 100mL diethyl ether, concentrated *in vacuo*, and purified by flash chromatography (20:80 EtOAc:Hexanes) to afford the title compound as a viscous colorless oil (179.3 mg, 0.71 mmol, 36% yield).
$^{1}$H NMR (400 MHz, $CDCl_3$) δ 11.19 (s, 1H), 7.49 – 7.36 (m, 5H), 3.51 (dt, *J* = 8.8, 7.4 Hz, 1H), 3.28 (dt, *J* = 9.0, 7.4 Hz, 1H), 2.25 – 2.07 (m, 2H), 2.02 – 1.87 (m, 2H), 1.85 – 1.72 (m, 2H). $^{13}$C NMR (101 MHz, $CDCl_3$) δ 199.45, 180.94, 134.64, 129.54, 129.30, 127.58, 55.39, 47.02, 31.62, 30.58, 25.65.
The bis-thiophenolate was also formed in this reaction. It eluted from the flash column at 40:60 EtOAc:Hexanes to afford a colorless oil. 188.7 mg (0.55 mmol, 28% yield)




$^{1}$H NMR




$^{13}$C NMR

**Supplementary Figure 2.2.** NMR spectra of CPDA-TP

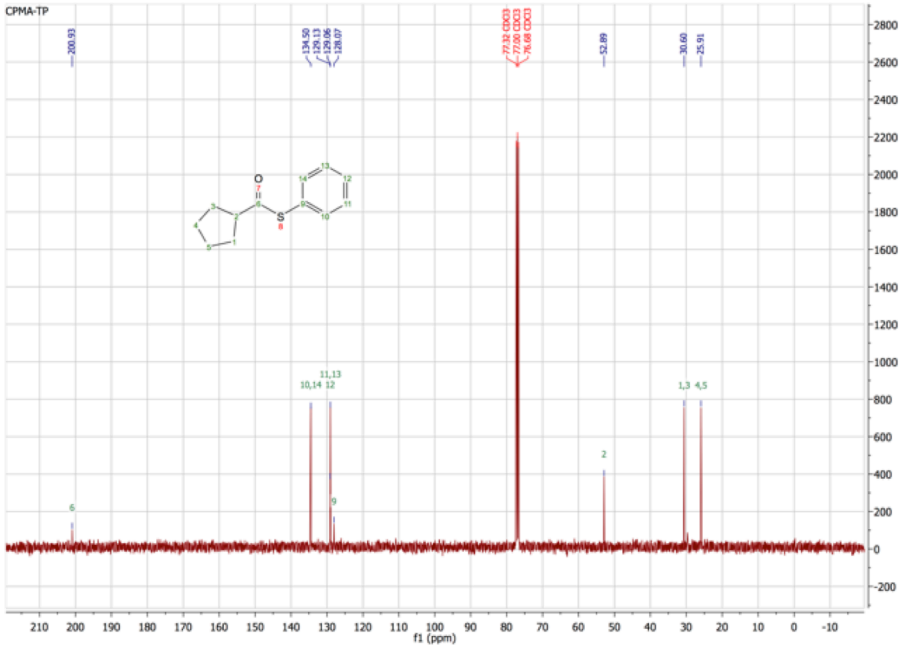
**CPMA-TP**

A 250mL flame-dried round-bottom flask with 50 mL acetone was cooled to 0 °C on ice, and charged with 5 mmol cyclopentanecarboxylic acid (542 µL) and 1.5 eq. thiophenol (770 µL). 1.5 eq neat DIC was added to the stirring solution dropwise over the course of 15 min. The solution was stirred on ice for another 2 h, and then quenched with 100mL 5% aqueous ammonium chloride. The mixture was extracted with 2 × 100mL diethyl ether, concentrated *in vacuo*, and purified by flash chromatography (10:90 EtOAc:Hexanes) to afford the title compound as a colorless oil (157.7 mg, 0.76 mmol, 15% yield).

$^{1}$H NMR (400 MHz, $CDCl_3$) δ 7.50 – 7.32 (m, 5H), 3.10 (p, $J$ = 7.9 Hz, 1H), 2.04 – 1.83 (m, 4H), 1.81 – 1.66 (m, 2H), 1.69 – 1.54 (m, 2H). $^{13}$C NMR (101 MHz, $CDCl_3$) δ 200.93, 134.50, 129.13, 129.06, 128.07, 52.89, 30.60, 25

The major product of the reaction appears to be an adduct between DIC and thiophenol that could possibly be avoided by letting the acid react with DIC before adding thiophenol.


$^{1}$H NMR

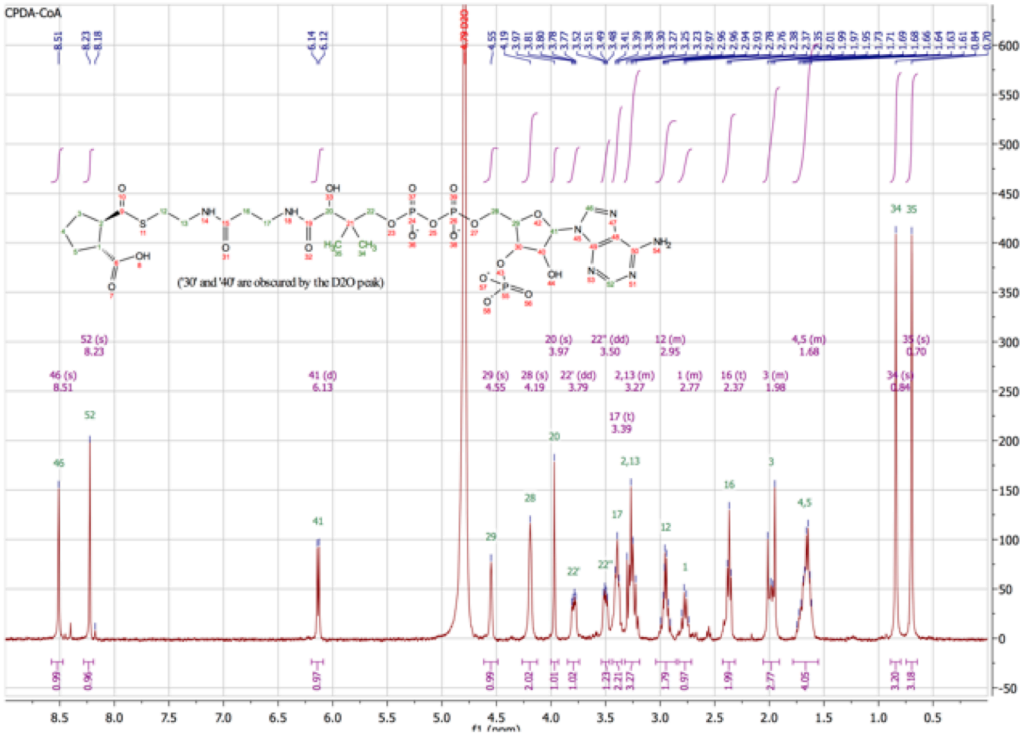



$^{13}$C NMR

**Supplementary Figure 2.3.** NMR spectra of CPMA-TP

**CPDA-CoA**

500 µL 0.5M aqueous sodium bicarbonate was cooled to 0 °C on ice in a reaction tube, and to it was added 25 µmol coenzyme A (sodium salt, hydrate, ~20 mg) and 5 eq *trans*-CPDA-TP (125 µmol, 31.3 mg). The mixture was stirred on ice for 1 h and allowed to warm up to room temperature for 1 h. 250 µL 1M HCl was added, and the mixture extracted with 3 × 1mL ethyl acetate. The aqueous layer was diluted to a final volume of 2 with water and purified by prep-HPLC as described above. The product fraction was lyophilized and desalted by prep-HPLC as described above, to afford a fluffy white powder. $^{1}$H NMR (400 MHz, $D_2O$) δ 8.51 (s, 1H), 8.23 (s, 1H), 6.13 (d, $J$ = 6.6 Hz, 1H), 4.55 (s, 1H), 4.19 (s, 2H), 3.97 (s, 1H), 3.79 (dd, $J$ = 10.0($^{3}J_{C\text{-}P}$), 4.6 Hz, 1H), 3.50 (dd, $J$ = 9.9($^{3}J_{C\text{-}P}$), 4.7 Hz, 1H), 3.39 (t, $J$ = 6.5 Hz, 2H), 3.33 – 3.19 (m, 3H), 3.04 – 2.85 (m, 2H), 2.83 – 2.71 (m, 1H), 2.37 (t, $J$ = 6.5 Hz, 2H), 2.06 – 1.91 (m, 2H), 1.78 – 1.55 (m, 4H), 0.84 (s, 3H), 0.70 (s, 3H).

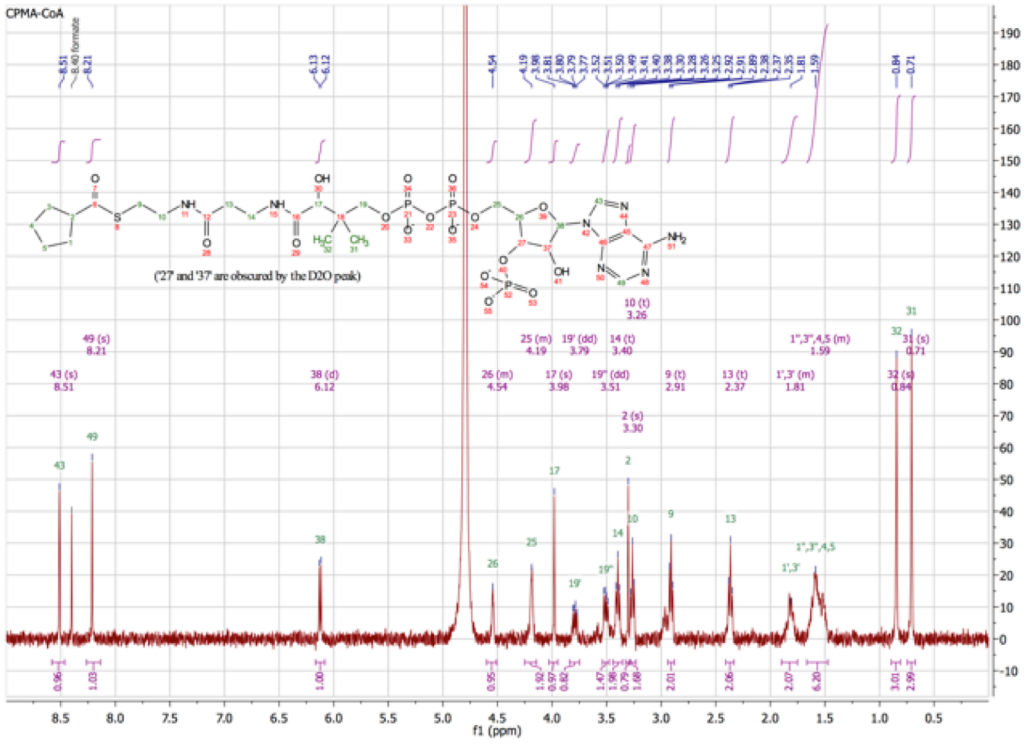



$^{1}$H NMR

**Supplementary Figure 2.4.** NMR spectrum of CPDA-CoA

**CPMA-CoA**

500 µL 0.5M aqueous sodium bicarbonate was cooled to 0 °C on ice in a reaction tube, and to it was added 25 µmol coenzyme A (sodium salt, hydrate, ~20 mg) and 5 eq *trans*-CPDA-TP (125 µmol, 31.3 mg). The mixture was stirred on ice for 1 h and allowed to warm up to room temperature for 1 h. 250 µL 1M HCl was added, and the mixture extracted with 3 × 1mL ethyl acetate. The aqueous layer was diluted to a final volume of 2 mL with water and purified by prep-HPLC as described above. The product fraction was lyophilized and desalted by prep-HPLC as described above, to afford a fluffy white powder. $^{1}$H NMR (400 MHz, $D_2O$) δ 8.51 (s, 1H), 8.21 (s, 1H), 6.12 (d, $J$ = 6.8 Hz, 1H), 4.60 – 4.51 (m, 1H), 4.25 – 4.14 (m, 2H), 3.98 (s, 1H), 3.79 (dd, $J$ = 9.6($^{3}J_{C\text{-}P}$), 5.1 Hz, 1H), 3.51 (dd, $J$ = 9.6($^{3}J_{C\text{-}P}$), 5.4 Hz, 1H), 3.40 (t, $J$ = 6.6 Hz, 2H), 3.30 (s, 1H), 3.26 (t, $J$ = 6.2 Hz, 2H), 2.91 (t, $J$ = 6.3 Hz, 2H), 2.37 (t, $J$ = 6.6 Hz, 2H), 1.90 – 1.75 (m, 2H), 1.67 – 1.47 (m, 6H), 0.84 (s, 3H), 0.71 (s, 3H)




$^{1}$H NMR

**Supplementary Figure 2.5.** NMR spectrum of CPDA-CoA

### 2.5.3. Plasmid list and construction

**pARH090**: MBP sequence from a pMal vector (New England Biolabs) was amplified via PCR such that an N-terminal 6xHis tag and C-terminal TEV protease cleavage site were introduced with flanking EcoRI/NdeI sites. PCR product was digested with EcoRI/NdeI and ligated into a similarly digested modified version of pBbS2k-RFP (Lee et al., 2011) in which the NdeI site normally present in the TetR gene was abolished

**pARH098**: The BorA1 ORF was codon-optimized for *E. coli* and synthesized by DNA 2.0 with flanking NdeI/BamHI sites. The NdeI/BamHI fragment was digested out of the parent vector and ligated into similarly digested pARH090

**pARH100**: The ketosynthase domain (boundaries as determined by MAPSIDB (Tae et al., 2009)) and remainder of BorA2's ORF were codon-optimized for *E. coli* and independently synthesized by DNA 2.0. These two fragments were PCR amplified from their parent vectors and assembled via Gibson into pARH090 digested with NdeI/BamHI

**pARH110**: The region of pARH098 which includes the ACP and C-terminal linker (residues 682-876 of the native BorA1 ORF) was PCR amplified to include flanking NdeI/XhoI sites and ligated into similarly digested pET28a vector (Novagen)

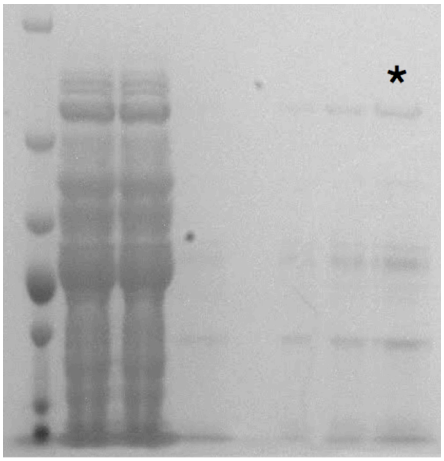
### 2.5.4. Protein expression and Purification

**MBP-BorA1** and **MBP-BorA2** were expressed by introducing plasmids pARH098 and pARH100 respectively into strain K207-3 (Menzella et al., 2007). Cells (2L) were grown at 37C in Terrific Broth to an O.D. of approximately 0.6 and then 50 ng/ml anhydrotetracycline and 250 uM isopropyl-β-D- galactopyranoside (IPTG) were added to induce expression of PKS proteins and Sfp, respectively. Cultures continued incubation at 18C for 20 hours and then were harvested by centrifugation (10000g, 10 min) and resuspended in lysis/wash buffer (300 mM NaCl, 50 mM phosphate, pH 7.4, 10 mM imidazole) supplemented with 0.1 mg/ml lysozyme. Cells were lysed by passing several times through an Emulsiflex C3 homogenizer (Avestin) and cellular debris was removed by centrifugation (15000g, 30 minutes). Cobalt resin (4 ml) was added to the supernatant and mixed at 4C for one hour before being applied to a fritted column. Resin was washed until flow-through resulted in no color change when mixed with Bradford reagent. Proteins were eluted with several resin volumes of elution buffer (300 mM NaCl, 50 mM phosphate, pH 7.4, 300 mM imidazole) and concentrated via spin filtration (Amicon, 100 kDa MWCO). MBP-BorA1 was exchanged into storage buffer (50 mM phosphate, pH 7.4, 8% glycerol) using a PD-10 column (GE Lifesciences), flash frozen in liquid nitrogen and stored at -80C. To remove the MBP tag on MBP-BorA2, this protein was first exchanged into cleavage buffer (75 mM NaCl ,12.5 mM phosphate, pH 7.4, 2.5 mM imidazole) using a PD-10 column and then incubated 20 hours at 4C with 1:100 w/w His-tagged TEV protease (purified using a simplified protocol from Tropea et al (Tropea et al., 2009)). After cleavage, 6 ml cobalt resin was added to the mixture and incubation continued with gentle mixing for one hour. Cleaved BorA2 protein was collected in flow-through after application to a fritted column and then concentrated, buffer-exchanged and flash frozen as described above for MBP-BorA1. We encountered tenacious contaminants in the MBP-BorA1 protein prep which we attempted to remove with additional purification steps including anion-exchange and size-exclusion

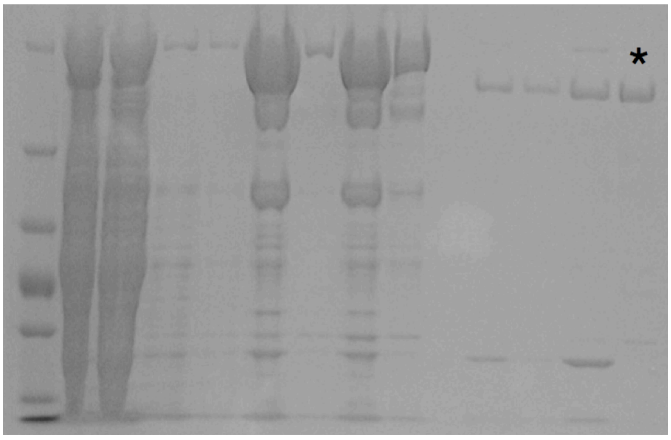
chromatography. Such polishing steps were mostly ineffective in the enrichment of target protein and were accompanied with significant loss of activity and therefore the prep as described above was used.

**apo-ACP0 monodomain** was expressed by introducing plasmid pARH110 into strain BL21(DE3). Cells (1L) were grown, induced and harvested as above except for the omission of anhydrotetracycline and supplementation of additional IPTG (500 uM final concentration). Cells were resuspended in lysis/wash buffer (50 mM Tris-HCl, pH 7.4, 10 mM imidazole, 1 mM DTT), lysed and fractionated as above. Protein supernatant was incubated with 2.5 ml cobalt resin, washed and eluted as above except for differing in the elution buffer composition (50 mM Tris-HCl, pH 7.4, 300 mM imidazole, 1 mM DTT). Eluate was concentrated and exchanged into buffer containing 50 mM Tris-HCl, pH 7.4 and 1 mM DTT via spin filtration (Amicon, 10 kDa MWCO), exchanged into buffer A (see below) with a PD-10 column and then applied to a HiTrap-Q anion exchange column. Chromatography buffers were composed of 50 mM phosphate, pH 7.4, 8% glycerol and 0 or 500 mM NaCl, respectively for buffers A and B; protein eluted at approximately 200 mM NaCl. Fractions containing protein were pooled, concentrated with a new spin filter, flash frozen in liquid nitrogen and stored at -80C.

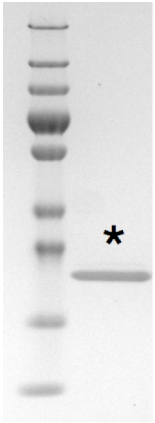
All protein concentrations were measured via Bradford assay using BSA to generate a standard curve and using SDS-PAGE/densitometry analysis to gauge purity and final concentrations of protein of interest (percent purity multiplied by total protein concentration). BorA2 and apo-ACP0 proteins were estimated at >95% purity while MBP-BorA1 was approximately 20% pure.



MBP-BorA1



MBP-BorA2



apo-ACP0

**Supplementary Figure 2.6.** SDS-PAGE gel of protein purifications. Asterisks appear above band corresponding to target protein and indicate fraction used in study

### 2.5.5. Experimental setup and sample processing

**Loading/extension assay:** A master mix (final concentrations: 1 uM each MBP-BorA1 and BorA2, 0.2 mM malonyl-CoA, 1 mM NADPH, 1 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to four tubes, to which 0.2 mM final concentration loading-CoA substrate was added and incubated at room temperature. At time points corresponding to 1, 10, 20, 30, 60, 120 minutes, aliquots were removed from each tube and snap frozen in a dry ice/ethanol bath to quench the reactions. Samples were thawed in the presence of 1:20 w/w trypsin and digested at 37C for six hours prior to injection.

**Translocation/extension assay:** An "ACP master mix" (final concentrations: 5 uM apoACP0, 1 uM sfp, 10 mM MgCl2, 1 mM NADPH, 1 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to four tubes, to which 0.1 mM final concentration loading-CoA was added and incubated at 30C for 30 minutes. Separately, an "extension master mix" was created (5 uM BorA2, 0.2 mM malonyl-CoA, 1mM NADPH in 100 mM phosphate buffer pH 6.8). Upon completion of the ACP0 acylation reaction, extension master mix was aliquoted into tubes containing acylated ACP0. After 120 minutes, tubes were frozen as above. A control extension reaction was included for each acyl-ACP0 substrate in parallel wherein an extension master mix created as above was subjected to 20 minutes of trypsin digestion (see below) prior to addition to acyl-ACP0 substrate. Samples were processed as described above.

**Acyl-ACP0 stability assay:** apoACP0 was acylated using sfp as described above with succinyl-CoA and butyryl-CoA used as substrates. Samples were incubated in the presence of 1:10 w/w trypsin, digested for one hour at 37C and then maintained at 37C and re-injected at indicated time points in order to detect the transient succinyl-ACP0 intermediate.

### 2.5.6. Analytical methods

For MRM data, all samples were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 1-2 µg of total peptide was injected onto a Sigma (St. Louis, MI) Ascentis Peptide Express C-18 column (2.1 mm x 50 mm) via an autosampler. For the loading/extension assay, a 15.5-minute method was used and consisted of a 400 µL/minute flow rate, starting with 95 % Buffer A (2% Acetonitrile, 0.1% formic acid) and 5% Buffer B (98 % Acetonitrile, 0.1% formic acid) for 1.2 minutes, followed by a rapid increase to 30% Buffer B in 1 minute and then a gradual increase to 55% B in 4.8 minutes. This was followed by a sharp increase to 90% B, where it was held for 2 minutes, followed by a quick ramp back down to 5% B, where it was subsequently held for 4 minutes to allow for column equilibration for the next run. For the translocation/extension assay, a 20.5 minute method was used with the same flow-rate and buffer compositions as above beginning with 5% buffer B for 1.2 minutes followed by a rapid rise to 25% over 1 minute and then a very slow rise to 36% over 10 minutes. After the slow gradient step, buffer B was rapidly increased to 90%, held, and dropped back down to re-equilibrate the column as above. The peptides eluting from the column were ionized by using an Turbo V Ion source (curtain gas flow: 20 l/min, temperature: 400 C, ion spray voltage: 4,800 V, ion source gas flow: 50 l/min, entrance potential: 10

V) operating in positive-ion mode. Methods were designed and data collected in Analyst 1.5.1 and data was quantified in MultiQuant 2.1 (AB Sciex).

| ID | Q1 | Q3 | Declustering potential | Collision energy |
|---|---|---|---|---|
| ACP0_ctrl | 1151.08 | 1431.71 (y15) | 115 | 61.3 |
| Holo-ACP0 | 730.70 | 261.12 | 50 | 40 |
| CPDA-ACP0 | 800.74 | 401.18 | 50 | 40 |
| CPMA-ACP0 | 778.77 | 357.25 | 50 | 40 |
| Succ-ACP0 | 780.71 | 361.12 | 50 | 40 |
| Butyryl-ACP0 | 765.73 | 331.16 | 50 | 40 |
| ACP1_ctrl | 680.38 | 846.48 (y8) | 125 | 40 |
| CPDA-ext-ACP1 | 967.12 | 445.19 | 50 | 44 |
| CPMA-ext-ACP1 | 952.46 | 401.20 | 50 | 44 |
| Buty-ext-ACP1 | 943.79 | 375.19 | 50 | 44 |
| Holo-ACP1 | 905.76 | 261.12 | 50 | 44 |
| Succ-ext-ACP1 | 953.77 | 405.16 | 50 | 44 |

**Supplementary Table 1.** Mass spectrum parameters

All transitions were monitored using a collision cell exit potential of 10 V.

ACP0_ctrl peptide: TVAAVLEADPAGTADAVAPDTAFK
ACP1_ctrl peptide: VVESVAFGVPSLR

ACP0 apo peptide: EMGLGSLSAVR
ACP1 apo peptide: AAIGPDSSFHAIGFDSLTAVELR
(site of phosphopantetheinylation underlined)

Time of Flight (TOF) MS data was collected on an Agilent 1290 UHPLC system coupled to an Agilent 6550 Q-TOF mass spectrometer operating in MS1 mode. 0.5 microgram (µg) of protein was injected and separated on a Sigma Ascentis Peptide Express C-18 column (2.1 mm × 100 mm, 2.7 um particle size) at a flow rate of 400 µl/min with gradient conditions identical to the 20.5 minute method described above except for the initial 5% buffer B time was 5 minutes. Prior to trypsin digestion, reactions were desalted via drop dialysis using VSWP02500 membranes (Millipore) according to manufacturer's recommendations. The peptides eluting from the column were ionized by using an Agilent Jet Stream source (sheath gas flow: 11 l/min, sheath gas temperature: 250 C, nozzle voltage: 1,000 V, nebulizing pressure: 35 psi, chamber voltage: 5000 V) operating in positive-ion mode. The data were acquired with MassHunter B.05.00 operating in MS1 mode within 300 m/z to 1400 m/z mass range.

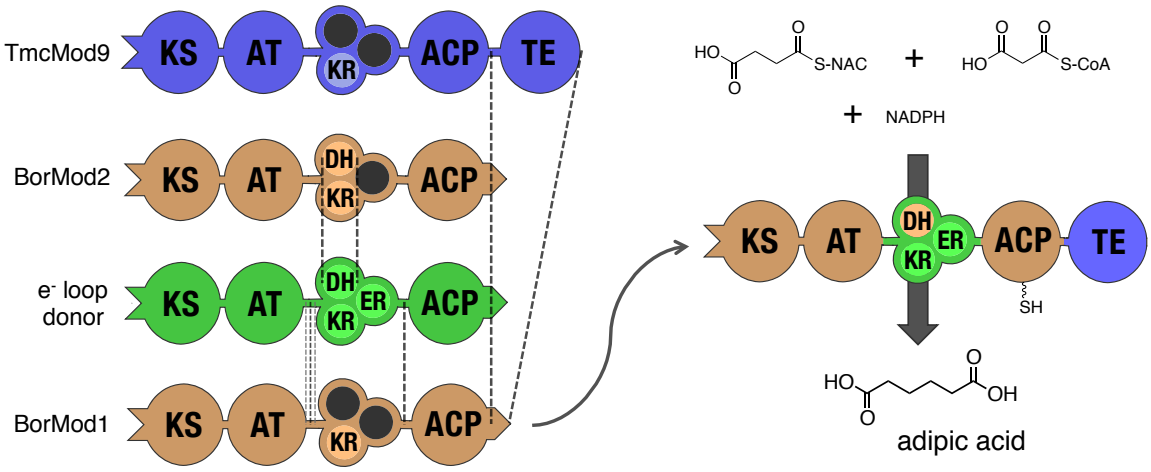
**Data analysis**

For loading/extension assay, raw data for each transition was normalized by dividing a transition's peak area by that of a control peptide present in BorA1, but which does not participate in catalysis and should therefore be invariant across samples ("ACP0_ctrl"). Background subtraction was performed for each transition by subtracting that transition's averaged values from negative control samples.

# **3.** Chapter 3 – Engineering the extension module BorMod1 for production of adipic acid

## 3.1. Abstract

Modular type I polyketide synthases (PKSs) are an attractive platform for production of a wide variety of chemicals. Here, we identify the important commodity chemical adipic acid as a target molecule for production via an engineered PKS extension module. Analysis of the intermediates during covalent catalysis unexpectedly revealed the activity of the dehydratase domain was rate-limiting and significant engineering effort was directed towards overcoming this kinetic bottleneck. By iterating through the design-build-test cycle, we demonstrate production of adipic acid from a succinate analog starter unit by a highly engineered PKS extension module.



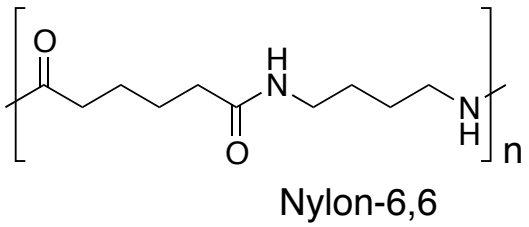
***Mr. McGuire**:* *I want to say one word to you. Just one word.*
***Benjamin**:* *Yes, sir.*
***Mr. McGuire**:* *Are you listening?*
***Benjamin**:* *Yes, I am.*
***Mr. McGuire**:* ***Plastics.***

The Graduate (1967)

## 3.2. Introduction

An unambiguous sedimentary hallmark of the modern age to be unearthed by future geologists will no doubt be plastic. Beyond the physical presence of this recalcitrant polymeric material, contributions to global warming from its production will also have played a role in ushering in the so-called Anthropocene epoch of Earth's history. The ubiquity and associated externalities of this material class are well-exemplified in particular by nylon.

Nylon is an aliphatic polyamide first introduced by DuPont in 1938 for use in tooth brush filaments (Kohan, 1995). Desirable properties such as good chemical resistance and high stiffness, tensile strength and melting temperature lend themselves to nylon's use in many industrial and commercial applications (automotive, sport, home interior etc.) (Palmer, 2001). Nylons are polymers of diacids and diamines or monomers with mixed acid and amine functionalities; the resulting nylon is named based on the number of carbons between the two acid and two amine groups. Nylon-6,6 is formed from adipic acid (hexane-1,6-dioic acid) and hexane-1,6-diamine while nylon-6 is formed from ring-opening polymerization of caprolactam and these two nylons together account for 90% of nylon use (E. I. du Pont de Nemours & Co., 1998; Honeywell, 1998). All three of these monomers are produced primarily from cyclohexanone and therefore are ultimately derived from petroleum feedstocks (benzene) (Alessi et al., 1997; Niu et al., 2002). This presents a problem from a sustainability standpoint as well as making the price of these commodity chemicals linked to the inherent volatility of the petroleum market.




Nylon-6,6

**Figure 3.1** Molecular structure of Nylon-6,6

An unfortunate consequence of adipic acid production (and hexane-1,6-diamine which is itself derived from adipic acid) is the stoichiometric production of $NO_x$ gases during the oxidation of cyclohexanone using nitric acid. Produced at rates of 2.3-3.3 million tons per year worldwide, it has been estimated that 10% of anthropogenic emissions of the potent greenhouse gas, $N_2O$, result from this chemical process alone (Alini et al., 2007).

"Green chemistry" methods using more benign oxidizing agents like hydrogen peroxide have been developed (Sato, 1998), however they have not been commercialized nor do they obviate the need for petroleum feedstocks. The advantage of a route to adipic acid without the associated greenhouse gas emissions and/or based on renewable feedstocks has motivated the exploration of several biological-based strategies for adipic acid production.

*cis, cis*-muconic acid ((2*Z*,4*Z*)-Hexa-2,4-dienedioic acid) is an intermediate in the catabolic pathways for certain aromatic compounds (Harwood and Parales, 1996) and can be catalytically converted to adipic acid via hydrogenation. A *Pseudomonas putida* strain was recently reported to accumulate *cis, cis*-muconic acid to 130 mM in fed-batch culture at an impressive 96% molar product yield (van Duuren et al., 2012). Encouragingly, life-cycle analysis (LCA) suggest an environmental benefit from this route compared to traditional adipic acid production, though the majority of this benefit can only be realized if renewable phenolics (e.g. from lignin) are readily obtainable as it would otherwise still depend on petroleum feedstocks (van Duuren et al., 2011). Frost and co-workers circumvented aromatic feedstocks entirely by diverting intermediates in the shikimate pathway ultimately into the production of *cis, cis*-muconic acid and demonstrated the first production of this adipic acid intermediate from glucose feedstocks (Niu et al., 2002).

To avoid the costly hydrogenation of adipic acid intermediates like *cis, cis*-muconic acid, direct production of adipic acid is preferred. Several companies have proposed various anabolic means of creating adipic acid generally from central metabolites (Burgard et al., 2012; Picataggio and Beardslee, 2012). These rely on condensations of small acyl-CoAs to arrive at either a six-carbon aliphatic monoacid, which must undergo omega-oxidation, or creation of a β-ketoadipate intermediate (e.g. from lysine biosynthesis), which must undergo full reduction at the beta-position. While seemingly viable, these pathways remain hypothetical as several required enzyme activities like specific monooxygenases and reductases have yet to be found. In contrast, Verdezyne has recently demonstrated bio-based production of adipic acid at pilot scale using long-chain fatty acids as a feedstock (URL: http://verdezyne.com/products/adipic-acid/). In this catabolic route, fatty acids undergo omega-oxidation leading to the generation of long chain carboxyacyl-CoAs which are then subjected to β-oxidation until arriving at adipic-CoA which is hydrolyzed to form the acid. An advantage of this platform is the possibility of creating longer chain, higher value diacids such as sebacic (8C) and dodecanedioic (12C) if beta-oxidation is terminated earlier.

The biosynthetic logic of using a polyketide synthase for production of adipic acid, in a sense is the opposite of Verdezyne's approach. Rather than start with a long chain carboxyacyl substrate and reduce the size, the objective is to start with a short-chain diacid and increase the size. Because the linear backbones of polyketide intermediates are extended two carbons at a time, biosynthesis must proceed from either the two-carbon oxalic acid or four-carbon succinic acid. This loaded intermediate would undergo two or one rounds of extension with full reduction using the unbranched malonyl-CoA extender to arrive at the six-carbon adipic-ACP intermediate which is then released by the action of a hydrolytic thioesterase. Due to its important role in the TCA cycle,

succinate/succinyl-CoA is readily available in organisms capable of aerobic respiration (e.g. common production hosts like *E. coli*, *Saccharomyces cerevisiae* and Actinobacteria) as is the malonyl-CoA extension unit. Therefore production of adipic acid using a PKS and succinyl-CoA starter is relatively host and feedstock agnostic, as minimal metabolic engineering is necessary to ensure adequate precursor supply and most feedstocks would presumably flow through central metabolism to provide the requisite precursors. Another advantage of using a PKS system, is the extensibility inherent in its modular nature. For example longer diacids could be accessed by usage of additional (or iterative) modules (see Chapter 4) or even totally novel adipic acid analogs could be created with alpha-substitutions (e.g. methyl- fluoro- or allyl groups) that may endow their polymers with unique attributes such as cross-linkability (Epstein, 1979; Walker et al., 2013).

As described in Chapter 2, loading and extension of carboxylated substrates in PKS systems is exceedingly rare. Significant effort was invested in the initial years of this study to develop a loading module that could process a carboxyacyl starter unit. The first strategy involved sequencing and bioprospecting from the spirofungin cluster in *Streptomyces violaceusniger* as the structure of its product indicated it may start with a diacid. This approach was abandoned in light of elucidation of the biosynthesis of the highly similar reveromycin in *Streptomyces* sp. SN-593 which demonstrated the diacid is generated by a post-synthetic oxidation mechanism and would therefore not be useful (Takahashi et al., 2011). Another strategy was based on engineering succinate specificity into an adenylation domain using the enzyme design algorithm, Rosetta (Richter et al., 2011). These domains are typically found in another class of megasynthases—the non-ribosomal peptide synthetase—and activate amino acids via adenylation prior to thioesterification and entry into thiotemplated biosynthesis (Stachelhaus and Marahiel, 1995). Technical hurdles from lack of solubility, inefficient mutagenesis throughput and difficulties in analytical methods precluded success with this strategy. Yet another strategy involving an elaborate, multi-stage selection was designed for evolving a succinate adenylating enzyme, however the concurrent and serendipitous discovery that the loading module of borrelidin PKS could accept succinyl-CoA and be extended/reduced to the 3-hydroxyadipic-ACP intermediate by the extension module, led to the wholesale adoption of the borrelidin PKS system (see Chapter 2 for detail).

In order to proceed from the 3-hydroxyadipic-ACP intermediate to adipic acid, additional beta-processing and hydrolytic chain release is required. We therefore generated chimeric versions of borrelidin's first extension module, BorA2 (hereafter referred to as BorMod1), which contain the full complement of reducing domains required to fully reduce at the beta position, and then append a thioesterase domain capable of releasing the linear product. Reductive loop swaps were among the earliest and most successful demonstrations of modularity in type I PKS systems (Donadio et al., 1993; Gaisser et al., 2003; McDaniel et al., 1999; Yoon et al., 2002). This may be in part due to the belief that reductive loops function as integral units as indicated by limited proteolysis experiments and recent structural studies (Aparicio et al., 1994; Dutta et al., 2014; Hong et al., 2005). Despite these examples, no prescriptive rules have been developed to guide successful reductive loop swaps and the most extensive, combinatorial study of reductive loop

swaps to date ultimately concluded, “no single donor [module] and no single pair of splice sites were found to be reliably optimal to effect a given alteration” (Kellenberger et al., 2008). We therefore developed a small library of donor reductive loops and candidate splice sites and created chimeric versions of BorMod1 in an attempt to create adipic-ACP. In absence of a thioesterase, intermediates covalently attached to the PKS could be monitored and rate-limiting steps identified and upon satisfactory production of adipic-ACP, a thioesterase was introduced to produce adipic acid.

## 3.3. Results

### 3.3.1. Initial selection of reductive loops and chimeric junctions

Five reductive loops were selected based on three criteria. First, the loop must contain the full complement of reducing domains (ketoreductase, dehydratase and enoyl reductase), which are necessary to fully reduce the β-ketone to the fully saturated methylene.

The second criterion was that the reducing loop is from a so-called "standalone" module in which the open reading frame or "subunit" contains just a single module (as opposed to bimodular structures such as those found in erythromycin PKS (Donadio et al., 1991)). Anecdotal evidence indicates that heterologous expression of standalone modules tends to produce more soluble protein, likely due to the fact that they are expressed in their native context and not as truncations from larger, multimodular subunits (S. Yuzawa, R. Phelan personal correspondence). It is, however, unknown what impact the context of the module has on likelihood of soluble expression when reductive loops are dissected out and used in chimeric constructs.

The final criterion was that the module harboring the reductive loop naturally incorporates a malonate extender unit as previous work has suggested a reduction in catalytic efficiency and relaxed stereospecificity result when KR domains are presented with an alpha carbon differentially substituted than the KR's normal substrate (e.g. a desmethyl substrate when the KR normally acts on an alpha-substituted intermediate) (McDaniel et al., 1999; Zheng et al., 2013).

Selected reductive loops and relevant information is presented in **Table 3.1**

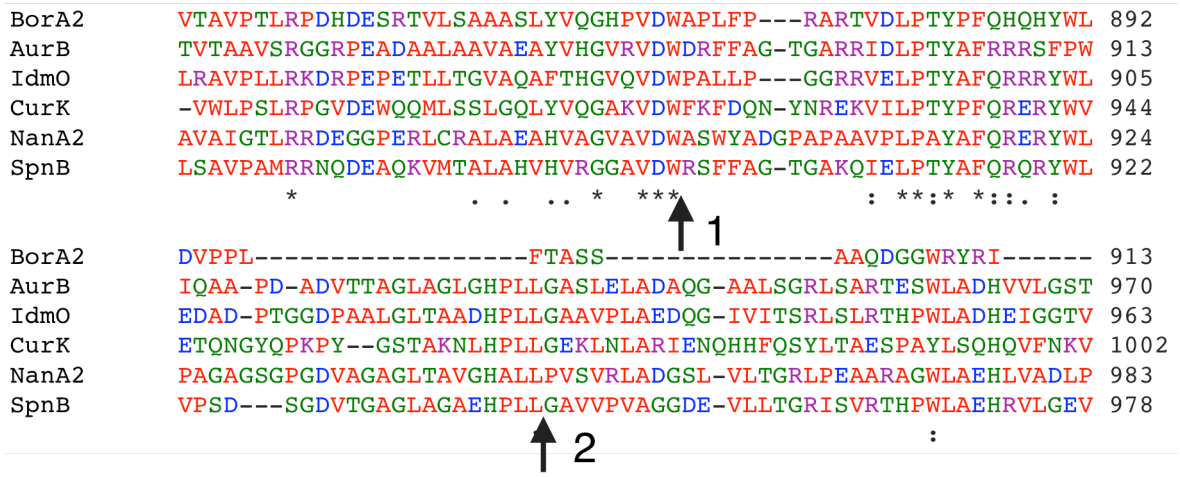
| Gene | PKS cluster | Host organism |
|---|---|---|
| AurB | Aureothin | *Streptomyces thioluteus* |
| CurK | Curacin | *Lyngbya majuscula* |
| IdmO | Indanomycin | *Streptomyces antibioticus* |
| NanA2 | Nanchangmycin | *Streptomyces nanchangensis* |
| SpnB | Spinosad | *Saccharopolyspora spinosa* |

**Table 3.1** Gene name, PKS cluster and host organism of reductive loops used

The most comprehensive study on reductive loop swaps (Kellenberger et al., 2008) concluded that "no single donor and no single pair of splice sites were found to be reliably optimal to effect a given alteration [in oxidation state]." There is, however, evidence that recombination drives the evolution of PKS diversity (Jenke-Kodama et al., 2006) and therefore two N-terminal junctions and a single C-terminal junction were initially selected at locations of high conservation as chimeric boundaries for the construction of the reductive loops (see **Figure 3.1**). The ten-member library (five

reductive loop variants and two junctions) was assembled with scarless cloning and heterologously expressed in *E. coli.*

N-terminal junction:




C-terminal junction:


**Figure 3.2**. Excerpts of a multiple sequence alignment of full modules used. Junction locations are indicated by arrows

### 3.3.2. Soluble expression and initial activity test of loop swap library

The ten library members were heterologously expressed in *E. coli* and solubility was assessed via SDS-PAGE analysis of whole cell lysates (see Materials and Methods for detail).

Surprisingly, the ratio of soluble to insoluble protein expression remains relatively high (as compared to that of wild type BorMod1) for the majority of the chimeric PKSs (**Figure 3.2**). While there are subtle differences between junction 1 and junction 2 constructs, the source of the reductive loop itself seems to be the major determinant of soluble protein expression with some (e.g. AurB) performing well and others (CurK, NanA2) producing very little soluble protein.

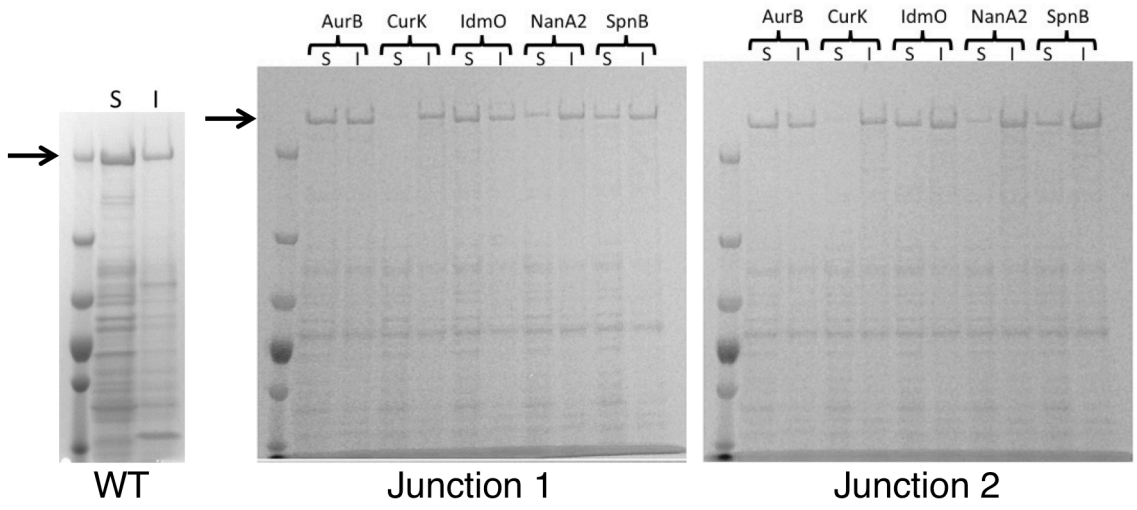



**Figure 3.3.** SDS-PAGE analysis (whole cell lysate) of reductive loop swap library protein solubility Shown is soluble (S) and insoluble (I) protein fractions corresponding to wildtype BorA2 protein and different reductive loop swaps. Black arrow indicates protein of interest

The low solubility of CurK constructs (notably, from the only non-Actinobacteria host organism) precluded their purification and therefore this reductive loop was abandoned. The remaining eight constructs were purified with affinity chromatography and their extension/reduction activity was assayed via the phosphopantetheine ejection assay (see section 2.2).

Prior, unpublished work demonstrated that a succinyl-SNAC analog was readily accepted by wild type BorA2 and extended to the 3-hydroxyadipic-ACP intermediate and was used as the starter substrate along with malonyl-CoA and NADPH supplied as the extender unit and reducing power, respectively. The results of this preliminary experiment are plotted in **Figure 3.3**.

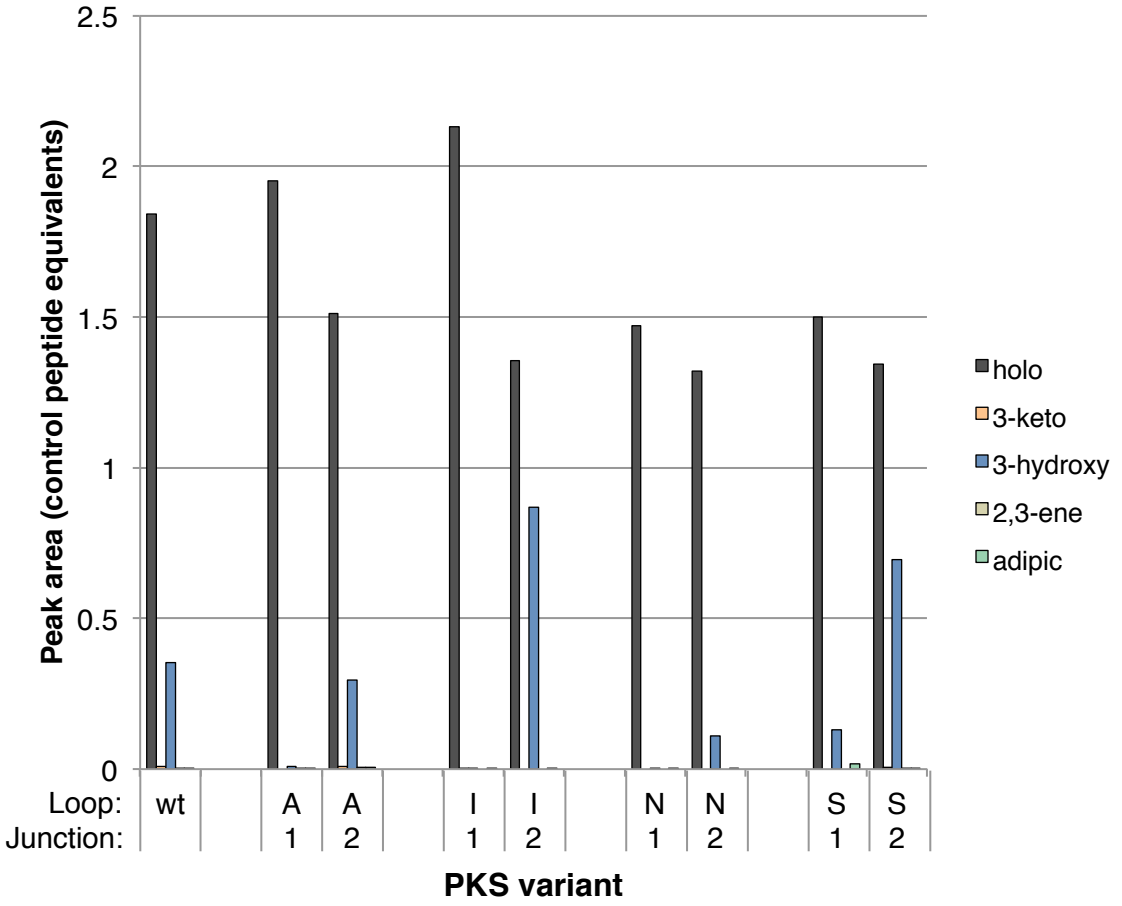



**Figure 3.4**. Extension intermediate analysis of initial PKS variants. The identity of each variant is indicated by the combination of loop source (**A**urB, **I**dmO, **N**anA2, **S**pnB) and junction (**1** or **2**)

Encouragingly, six out of eight constructs remain catalytically competent for condensation as indicated by the presence of extended intermediates. For these six constructs, however, the majority of the extended intermediates are the 3-hydroxyadipic-ACP and little to no adipic-ACP is observed. These data indicate that either for unanticipated reasons, these intermediates are not readily detectable, or that the reductive processing is stalled at the dehydratase step.

### 3.3.3. BorDH2 *in trans* is competent for dehydration and increases the proportion of adipic-ACP

Chapter 2 describes a study wherein it was found that the presence or absence of a terminal carboxy group can have profound influence on the activity of PKS domains. We wondered if the dehydratase domains from the reductive loop variants were not competent to dehydrate the carboxy-bearing 3-hydroxyadipic-ACP and therefore sought to test the ability of a different dehydratase domain which does normally process such a substrate.

Because of its proximity to a terminal carboxylic acid group (see **Figure 3.4**), the first DH domain in the borrelidin cluster, BorDH2, was chosen.

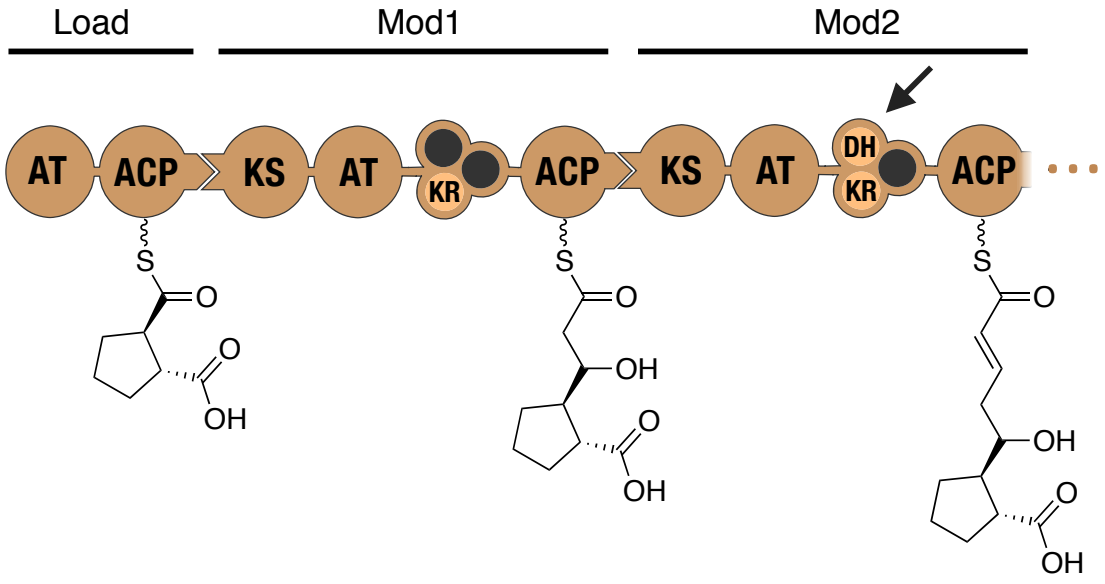



**Figure 3.5**. Excerpt of the borrelidin PKS cluster's domain structure and intermediates (adapted from (Olano et al., 2004a). Arrow indicates BorDH2 domain.

Prior work demonstrated that BorDH2 can be heterologously expressed as a monodomain and is competent to dehydrate surrogate substrates in the form of *N*-acetylcysteamine thioesters, though substantial amounts of enzyme were required and significant conversion was only observed after 16 hours of incubation (Vergnolle et al., 2011). Additionally, previous work by Guo and coworkers has shown that DH domains can be active *in trans* in *in vitro* reactions, supporting this strategy (Guo et al., 2010). We generated a similar construct and purified BorDH2 and provided it *in trans* to the wild type BorMod1 (along with requisite succinyl-SNAC starter, extender and reducing power) in an attempt to observe the 2,3-ene-adipic-ACP intermediate. Indeed, as shown in **Figure 3.5**, provision of BorDH2 *in trans* leads to the detection of the transition corresponding to 2,3-ene-adipic-ACP in a dose-dependent manner.

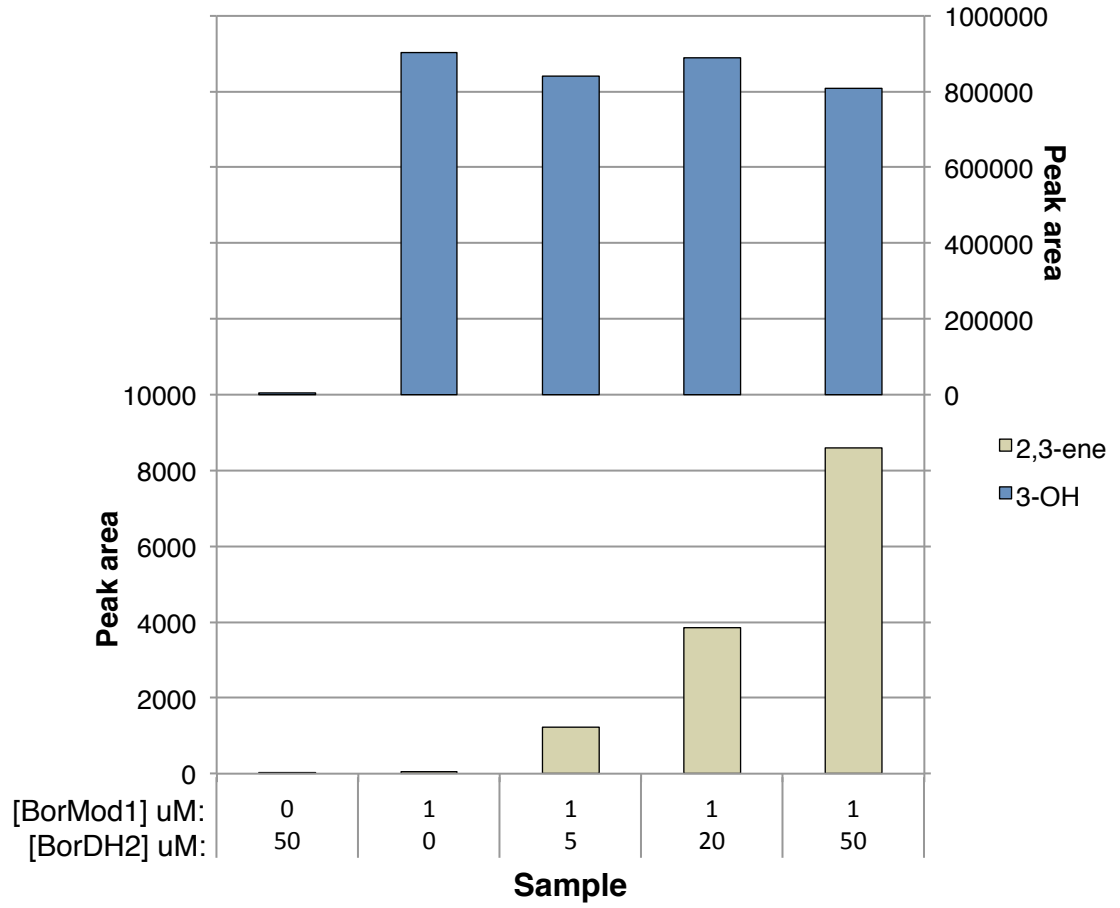



**Figure 3.6**. Peak areas of the 2,3-ene-adipic-ACP intermediate under different combinations of BorMod1 and BorDH2 proteins. Note different y-axes.

This peak is not observed in absence of the BorMod1 nor when BorDH2 is omitted from the reaction and was verified by LC-TOF analysis to have the correct exact mass and isotopic distribution (data not shown).

Despite this proof of principle that BorDH2 can dehydrate the 3-hydroxyadipic-ACP intermediate, conversion is apparently quite low (note different y-axis scales in above figure). Even in the unlikely scenario that the 2,3-ene intermediate is very poorly detected relative to the 3-OH intermediate, there is no decreasing trend in the 3-OH intermediate's peak area when BorDH2 concentration is increased as would be expected if significant conversion was taking place.

We next sought to test provision of BorDH2 *in trans* to the reductive loop swap library. Unlike wild type BorMod1, these modules contain an enoyl reductase and therefore production of the adipic-ACP intermediate will be observed if and only if both the enoyl reductase activity occurs as well as the dehydratase activity from BorDH2 *in trans*

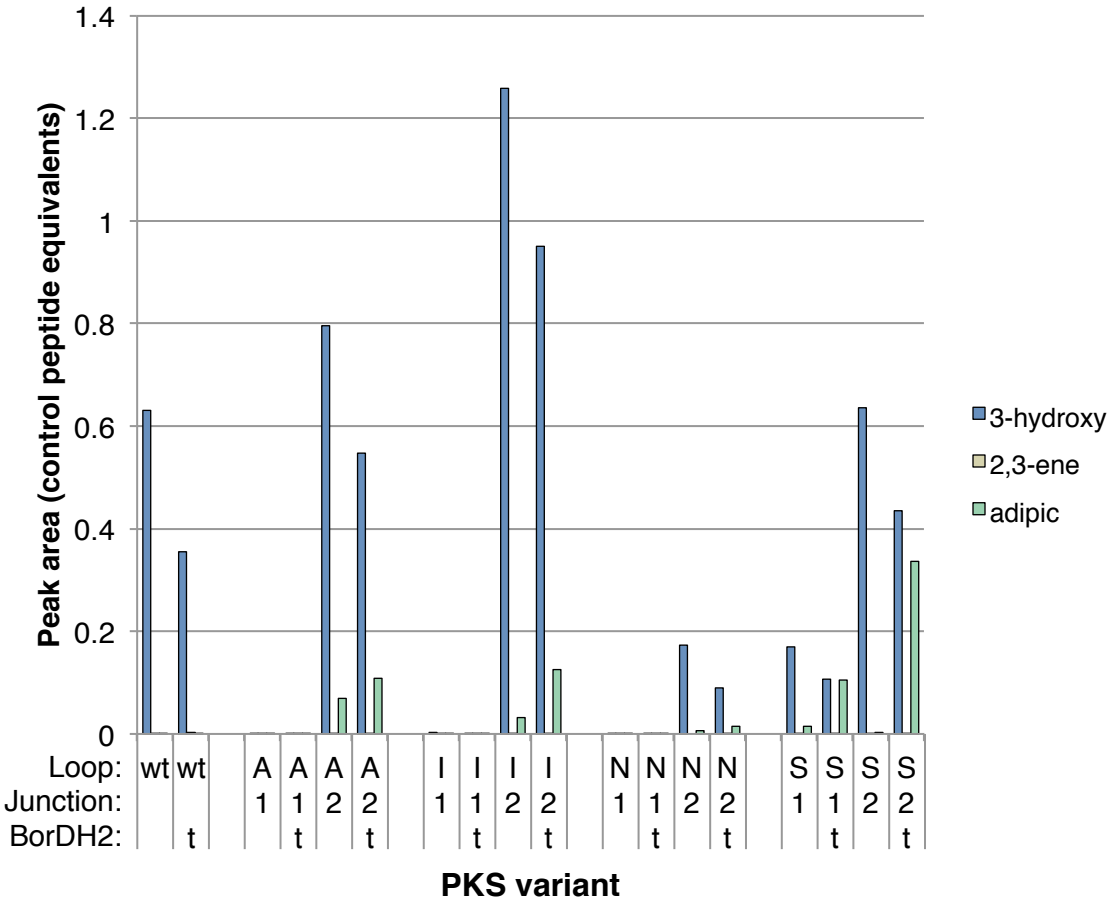



**Figure 3.7**. Extension intermediate analysis of initial PKS variants when BorDH2 is provided *in trans* (50 uM, indicated by the presence of a "t"). Because the 3-keto intermediate was not previously nor presently detected, it is omitted for clarity.

Indeed, when BorDH2 is provided *in trans* at super-stoichiometric levels, adipic-ACP intermediate is detected at higher levels compared to those without BorDH2. A particularly interesting case is the comparison between S2 and S2t, where provision of the dehydratase increases adipic-ACP production from nearly undetectable levels to the highest level amongst all variants.

Interestingly, adipic-ACP is now detected in samples where it was not previously observed (compare variant A2 between **Figures 3.3** and **3.6**). This is likely explained by drastically better sensitivity after switching mass spectrometers (from AB SCIEX 2000 QTRAP to 4000 QTRAP) and refining various mass spec parameters (see 3.5.5 for details). Generally speaking such inter-instrument/inter-run variability is controlled for to an extent by normalizing intermediate signals to the control peptide, however at low signal intensities such as those in the preliminary experiment described in section 3.3.2, noise from normal instrument operation has a higher likelihood of obscuring authentic signals. As a comparison, the mean control peptide peak area in the experiment described above is nearly 35-fold higher than that of the preliminary experiment.

No significant accumulation of the 2,3-ene-ACP intermediate is observed when BorDH2 is provided (except in the case of wild type BorMod1). This, along with the observed production of adipic-ACP in all loop variants, suggests that the 2,3-ene intermediate is readily processed by the enoyl reductase domains present *in cis*.

### 3.3.4. BorDH2 *in cis* further increases proportion of adipic-ACP

Having demonstrated that BorDH2 provided *in trans* is capable of promoting adipic-ACP formation, we wondered if this was a property unique to this particular dehydratase domain or if it the results are simply because it is provided in stoichiometric excess. To test this, BorDH2 was swapped into a subset of the reductive loop library in order to replace the native DH domain. The A2, I2 and S2 variants were selected because they showed the best overall production of adipic-ACP when BorDH2 was provided *in trans* and we were particularly interested in the S2 variant since in the absence of BorDH2, adipic-ACP production is undetectable.

In order to define chimeric boundaries, a multiple sequence alignment was constructed using DH domains from AurB, IdmO, SpnB, BorMod2 and EryMod4 as this DH has been crystallographically characterized (PDB: 3el6 (Keatinge-Clay, 2008)). Junction selection was again guided by regions of homology and was further informed by examination of the crystal structure, which confirmed the core structured regions of the domain were flanked by the junctions (data not shown).


**Figure 3.8**. Multiple sequence alignments of relevant dehydratase domains. Junction locations are indicated by arrows.

SDS-PAGE analysis indicated that soluble expression of the protein was not significantly impacted by the DH swap (data not shown). After purification, these DH swapped variants were compared to previous constructs as before via intermediate analysis after extension of succinyl-SNAC. Results are shown in **Figure 3.8**.

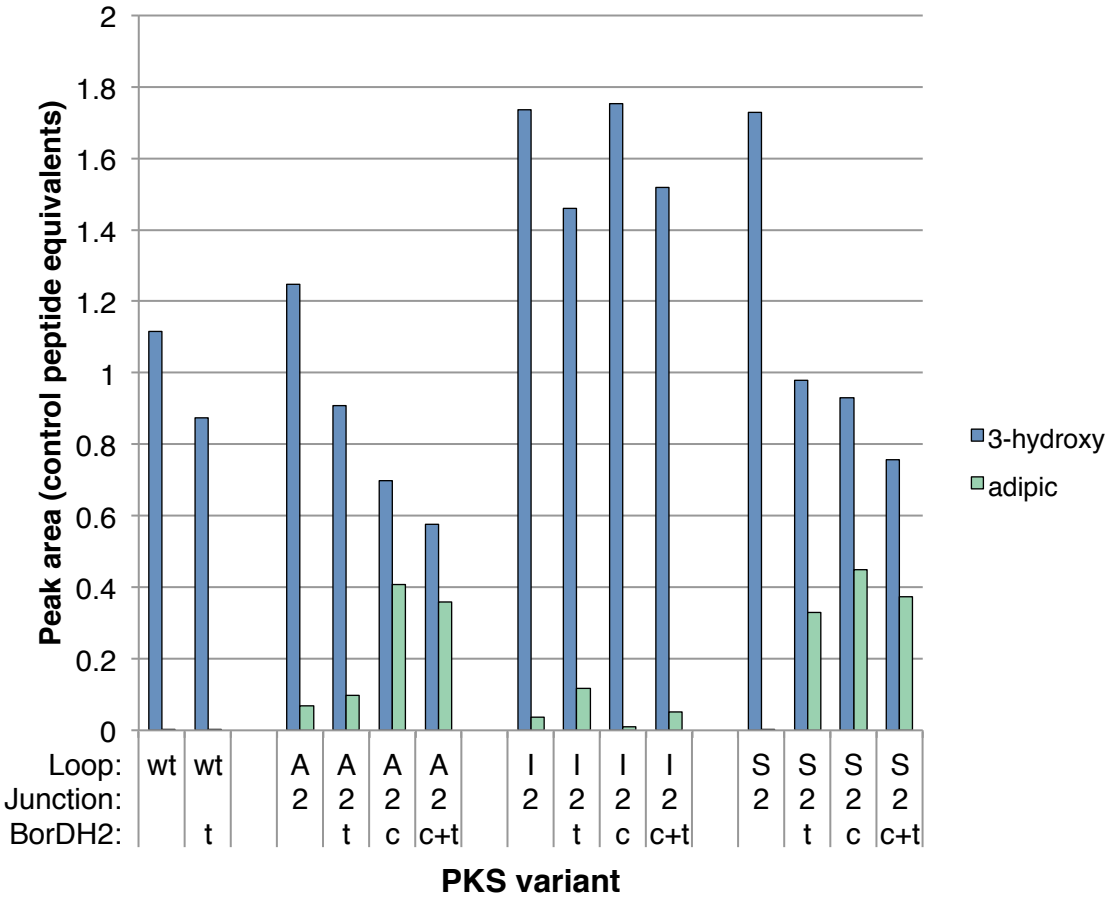



**Figure 3.9**. Extension intermediate analysis with BorDH2 *in cis* (indicated by a "c"), *in trans* (50 uM, indicated by a "t") or *in cis* plus additional added *in trans* ("c+t"). As before, keto and 2,3-ene intermediates were not readily detected and are omitted for clarity.

DH swapped variants clearly promote the formation of adipic-ACP compared to non-swapped variants (e.g. compare A2 to A2c) and are superior to conditions were the DH is provided *in trans* (e.g. compare A2t to A2c). Provision of DH *in trans* to DH swapped constructs does not further increase the amount of adipic-ACP intermediate, and may even have a slight inhibitory effect as indicated by a decrease in the total amount of extension intermediate detected when included in all reactions.

### 3.3.5. *In vitro* production of adipic acid using thioesterase-appended PKS variants

Having demonstrated that the highly engineered PKS variants A2c and S2c can accumulate appreciable amounts of adipic-ACP intermediate, we designed constructs to include a thioesterase (TE) domain in order to liberate acyl-ACP intermediates and produce free adipic acid.

#### 3.3.5.1. Selection of a stereospecific thioesterase

Previous experiments showed that the TE from the well-studied erythromycin PKS (eryTE) readily hydrolyzed the 3-hydroxyadipic-ACP intermediate when appended to the wild type BorMod1 protein (data not shown). Because the majority of the ACP-bound intermediates for even the A2c and S2c constructs are not fully reduced to adipic-ACP (keeping in mind the usual caveats about comparing MRM transitions), and given that the kinetics of adipic-ACP formation were likely slow, premature hydrolysis of incompletely reduced intermediates was a significant concern. We therefore sought a thioesterase that would be less likely to process the 3-hydroxyadipic-ACP intermediate in an attempt to bias the product distribution of free acids towards the fully reduced adipic acid.

The TE from the tautomycetin cluster (ttmTE) is a hydrolase (as opposed to a cyclase) and substrate promiscuous, though interestingly it exhibits high stereoselectivity with a several hundred-fold preference for *(R)*- β-hydroxy acyl substrates compared to the corresponding *(S)* stereoisomer (Scaglione et al., 2010). We reasoned that if the 3-hydroxyadipic-ACP intermediate generated by the various ketoreductase domains in our PKS variants was in the *(S)* configuration, they would be recalcitrant to hydrolysis and therefore allow full reductive processing prior to release. The mass spectrometry-based methods used in the current study are not well-suited for delineating stereochemistry, however previous structural studies have identified "fingerprint" regions that allow one to reliably determine the stereochemical outcome of β-ketone reduction (Keatinge-Clay, 2007). A partial alignment of the AurB and SpnB ketoreductase domains is shown in **Figure 3.9**.

```
Ttm-KR9      LETAGAR-VTVAACDVTDRDALAQVVSEIPEEWPLRTVVHAAGVMGDPTLLTALTPRLLR
Tyl-KR1      LRGHGCE-VVHAACDVAERDALAALVTA----YPPNAVFHTAGILDDAV-IDTLSPESFE
AurB-KR2     LRESGAHSVRAVACDCVDRTAVADLLASIPDEHPLTAVVHTVGVVDDGV-LETMTPERID
SpnB-KR2     LTAYGAE-VSLQACDVADRETLAKVLASIPDEHPLTAVVHAAGVLDDGV-SESLTVERLD
```

**Figure 3.10**. Excerpt of a multiple sequence alignment of relevant ketoreductase domains. The (L/V)DD motif indicative of "B type" KR domains is boxed. An "A type" KR from tautomycetin (Ttm-KR9) is included for comparison.

The presence of the signature (L/V)DD motif (shared by the structurally and biochemically characterized tylosin KR1) indicates that AurB and SpnB possess "B type" ketoreductases and are therefore expected to generate the desired *(S)*-3-OH intermediate. Identification of these KRs as B type is expected as it has been previously observed that

the majority of KR domains found in modules harboring DH and ER domains are B type (Caffrey, 2003).

### 3.3.5.2. Preliminary production and detection of adipic acid

The ttmTE domain including upstream linker regions was introduced into the constructs immediately following the terminus of the acyl-carrier protein as this junction was shown to lead to successful hydrolysis of the 3-hydroxyadipic-ACP intermediates using eryTE, as mentioned above. A2c-TE and S2c-TE constructs were purified and production of adipic acid was tested using an experimental setup similar to that used in previous sections (succinyl-SNAC, malonyl-CoA and NADPH as starter, extender and reducing power, respectively) and samples were analyzed using LC/MS/MS (see materials and methods for details).

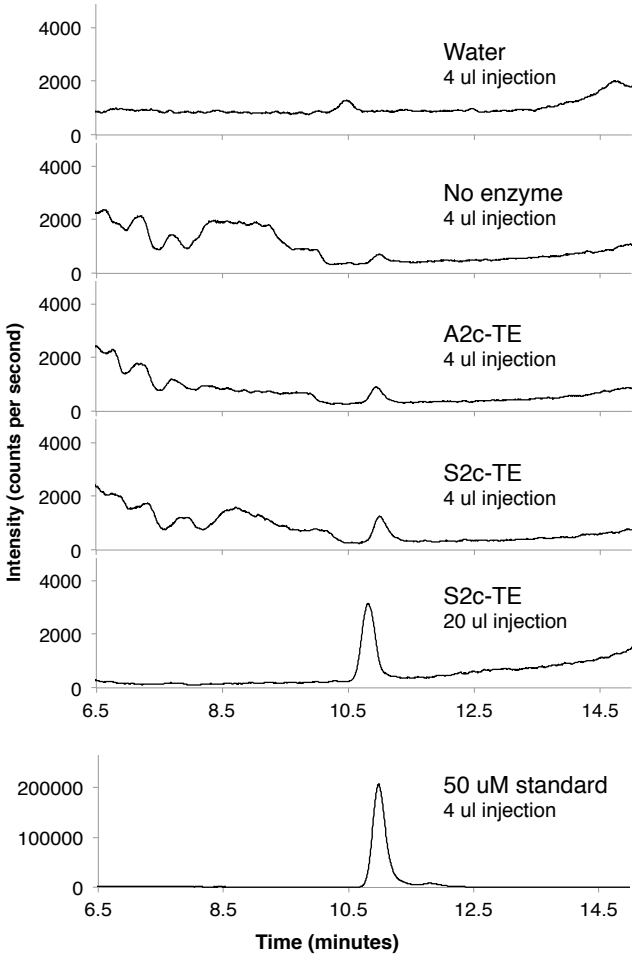



**Figure 3.11**. LC/MS/MS chromatogram (*m/z* 145 / 83 transition) of *in vitro* adipic acid production from TE-appended constructs. Note different y-axis scale for 50 uM standard.

The data in this preliminary experiment suggest that adipic acid is produced by the engineered PKS variants that are appended with the tautomycetin thioesterase domain, however the amount is barely detectable above noise and comparison to the 50 uM

standard indicates that production is very low. Additionally, the presence of a peak in the control reaction lacking enzyme (corresponding volume substituted with enzyme storage buffer) though small, precludes high confidence in the production data.

To verify the results obtained in the preliminary experiment, we repeated the experiment after refining our reaction set up, sample processing and analytical method. In this second experiment the S2c-TE construct alone was tested as it performed better in the preliminary experiment and its concentration in the reaction mixture was doubled from 5 uM to 10 uM. We suspected the filtration columns used in sample processing were the source of the putative contaminating adipic acid as they were previously shown to introduce contamination of other analytes (J. Zhang, personal correspondence; the water in the preliminary experiment was not filtered). Potential contamination from the spin filters was mitigated by rinsing with LC-MS grade water prior to use. Finally, the liquid chromatography method was refined to increase sensitivity, additional controls were included and a semi-pure synthetic standard of the 3-hydoxy-adipic acid was used to generate MS/MS transitions which were also monitored (see materials and methods for details). Results of this experiment are shown in **Figure 3.11**.

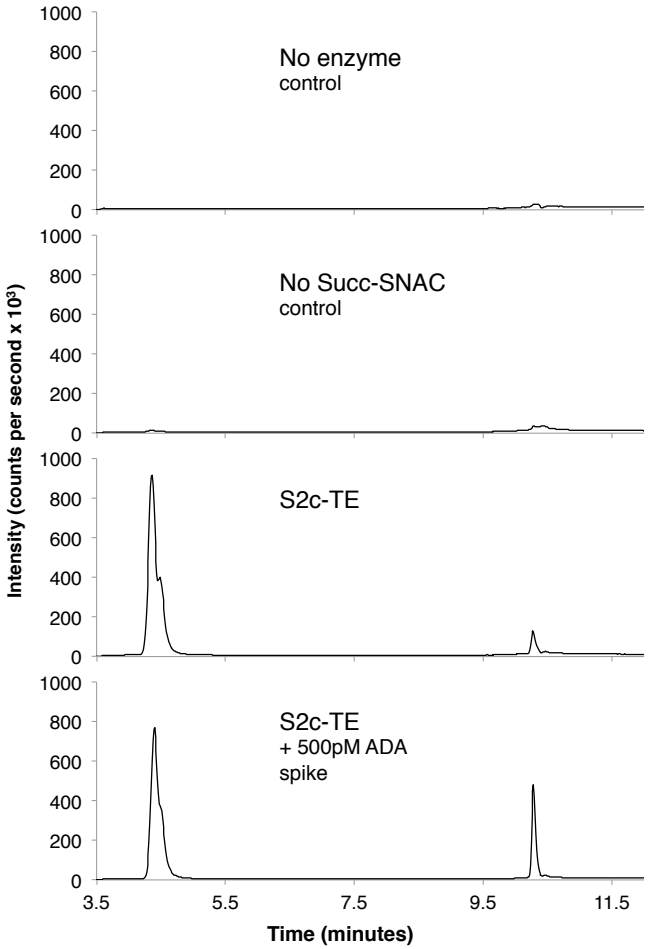



**Figure 3.12**. LC/MS/MS chromatogram (total ions from all MS/MS transitions of 3-hydroxyadipic acid and adipic acid) for different reactions. Adipic acid abbreviated as "ADA".

The results from the above experiment unambiguously confirm that the engineered S2c-TE variant is capable of adipic acid production. Not only is the signal to noise ratio for the adipic acid peak now significantly greater, spiking in a small amount of the authentic adipic acid standard and then re-running the sample increases the peak size (bottom panel). Despite the use of a supposedly stereoselective thioesterase, there is a large peak corresponding to 3-hydroxyadipic acid (retention time approx. 4.5) indicating that a large proportion of this partially reduced intermediate is hydrolyzed by the TE.

### 3.3.6. Refined N-terminal junction and final construct

Despite junction 2 PKS variants generally working the best (especially when BorDH2 swapped), closer sequence and structural analysis indicated that junction 2 constructs are truncated by approximately 15 residues (depending on how domain boundaries are annotated) at the N-terminus of the dehydratase domain (see **Figure 3.12**). These residues are distal to the active site and protein interface where the ACP protein presumably docks and are clearly not essential, however their influence on the overall tertiary structure and kinetics of PKS enzymes is unclear.

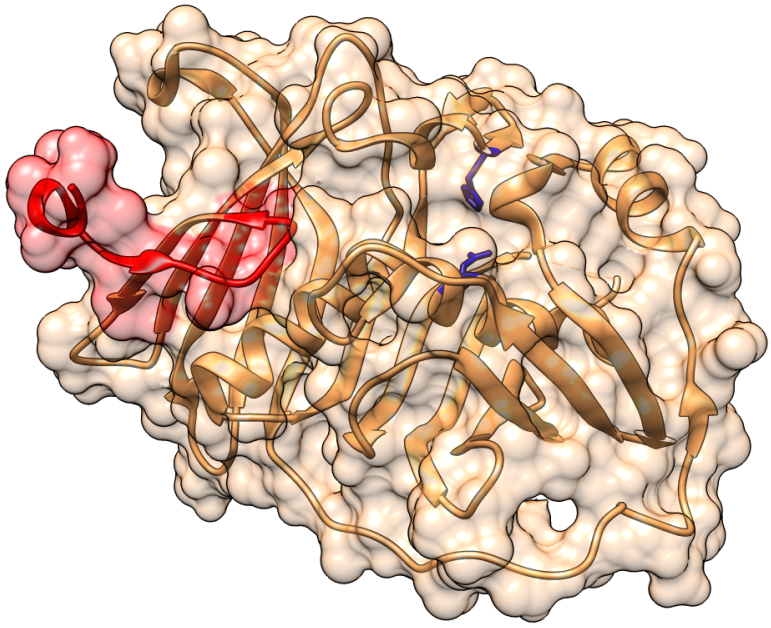


**Figure 3.13**. Structure of Ery4DH (PDB: 3el6). Homologous residues (as determined by sequence alignment) to those deleted in junction 2 constructs are colored in red. Active site His/Asp are colored blue.

Therefore, a new N-terminal junction was selected intermediary to junctions 1 and 2 ("junction 1.5"). This location immediately follows the post-AT linker region which is believed to be important for proper KS-AT domain orientation (Tang et al., 2006) and also restores the DH domain N-terminal truncations as shown in **Figure 3.13**.

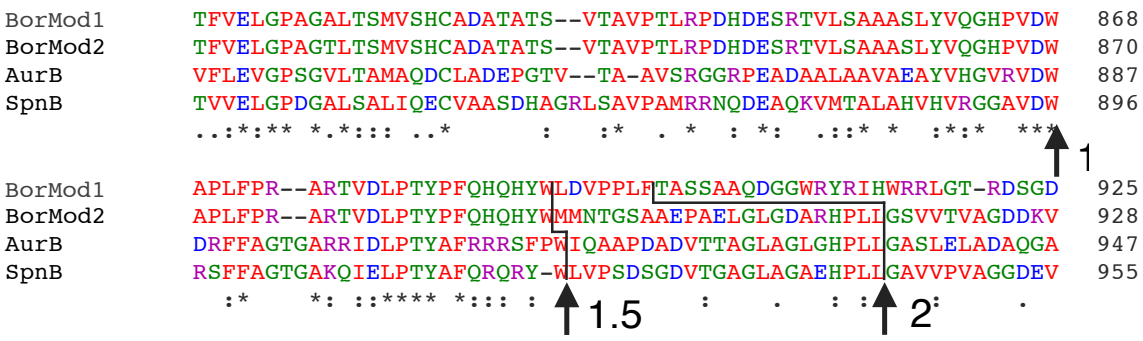



**Figure 3.14**. Excerpt of a multiple sequence alignment of relevant reductive loops (including BorMod2 which harbors BorDH2). Junction locations are indicated by arrows and lines.

Junction 1.5 variants were created for a subset of the reductive loop library, which included the best performing AurB and SpnB loop sources. We were curious to see the effect of the new junction in isolation as well as in combination with the superior BorDH2 swaps and therefore all four combinations (A/S and null/+ BorDH2) were constructed and tested with the intermediate analysis as before. To minimize inter-injection variability and provide a more robust comparison between PKS variants, this experiment was performed with technical triplicates. Results are shown in **Figure 3.14**.

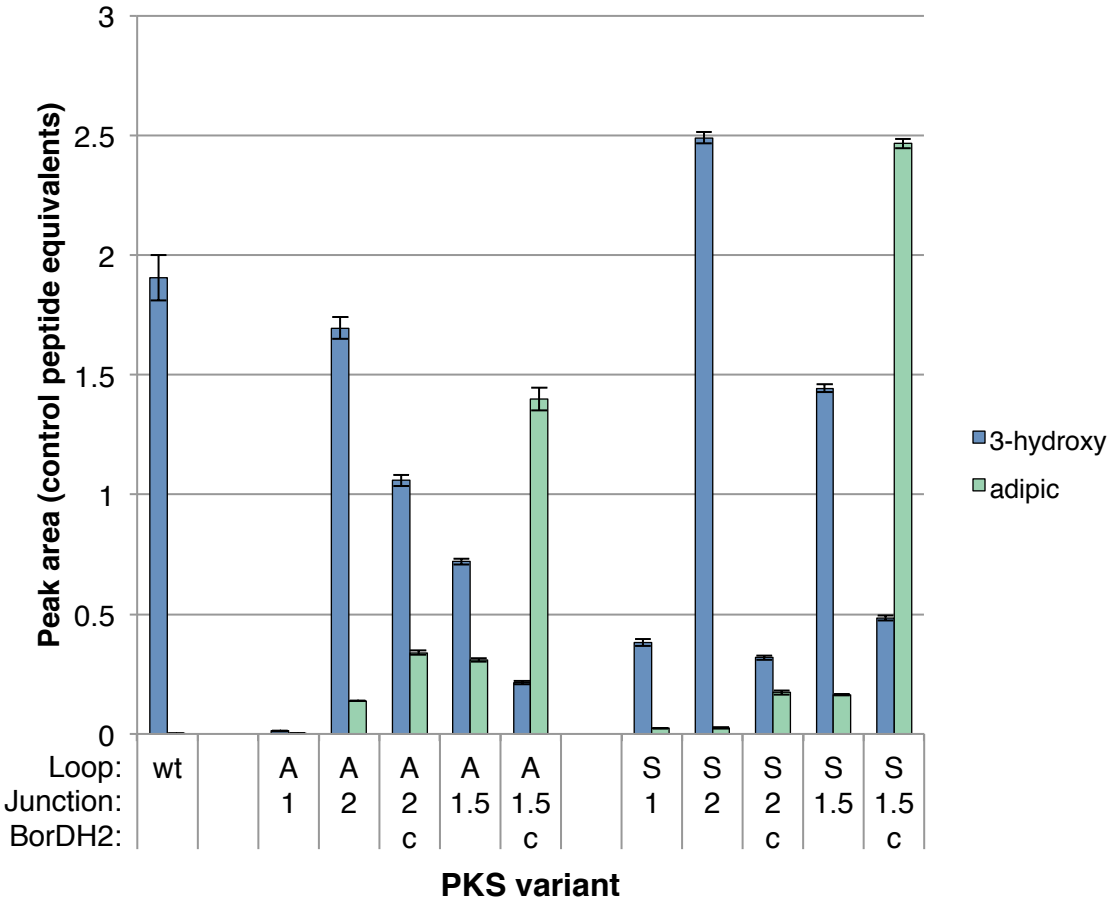



**Figure 3.15**. Extension intermediate analysis including junction 1.5 variants. As before, minimally or undetectable intermediates not shown for clarity. Peak area reported as mean for technical replicates with error bars representing standard error.

The new junction is superior than junction 1 or junction 2 as gauged by total production of adipic-ACP intermediate and strikingly, the combination of the new junction with the BorDH2 swap has a synergistic effect as evidenced by nearly complete intermediate conversion to adipic-ACP in the case of A1.5c and S1.5c constructs. The activity of these variants in conjunction with the tautomycetin thioesterase remain to be tested.

## 3.4. Discussion

In this chapter, using advanced analytical methods and iterations through the design-build-test cycle, I have demonstrated the first production of adipic acid by a highly engineered polyketide synthase. This was facilitated by the identification of an extension module, BorMod1, which naturally extends with malonyl-CoA, is capable of processing carboxyacyl substrates (see Chapter 2) and that is also amenable to engineering efforts as evidenced by robust solubility during heterologous protein expression of chimeric constructs. In section 3.3.1 and 3.3.2 I described the strategy for selection of reductive loops and performed initial solubility and activity tests for the ten-member library. The solubility test showed that in this system, the source of the reductive loop is the major determinant of solubility while the location of the splice junction is secondary. Because these protein sequences are heterologously expressed, it is unclear whether this is simply an artifact from codon-optimization or if there is perhaps a phylogenetic or other explanation for this observation.

Initial activity tests indicated that replacing the reductive loop from BorMod1 with a library of reductive loops from fully reducing modules does not significantly compromise the catalytic competence of the module for the extension reaction. This lends further support to the idea that the reductive loop functions as an "integral unit" apart from the core catalytic activity of the acyltransferase and ketosynthase domains in the module (Aparicio et al., 1994; Dutta et al., 2014; Hong et al., 2005), and that the chimeric junctions used in this study did not significantly perturb the module's tertiary structure such that condensation is precluded.

Unfortunately, data showed that the activity of the dehydratase domain is rate-limiting as indicated by the accumulation of the 3-hydroxyadipic-ACP intermediate. We hypothesized this is the result of the DH domain's inability to process a carboxyacyl intermediate as such substrates are exceedingly rare in PKS systems (see Chapter 2 for more detail). Alternate hypotheses are that the acyl carrier protein native to BorMod1 was unable to productively interact with the DH domains from the donor reductive loops either because of incompatibilities in the docking interface or perturbation of the module's tertiary structure that would disallow access of the highly mobile ACP domain to the DH domains.

The BorDH2 swapped constructs significantly improve dehydratase activity for the A2 and S2 variants, though its effect on the I2 construct is comparatively negligible for unknown reasons. Because the DH swap was performed at junctions flanking the structurally conserved catalytic core of the domain, it is unlikely that the overall tertiary structures of the modules were altered in these highly engineered variants compared to their parent constructs—thus arguing against the tertiary structure perturbation hypothesis. While BorDH2 processes a carboxyacyl substrate which may make it better suited to dehydrate the 3-hydroxyadipic-ACP intermediate, its natural substrate is an 8-, not a 6-carboxy intermediate (see **Figure 3.4**) and therefore substrate specificity arguments may not necessarily apply. With regard to the ACP-DH interaction hypothesis, previous work has shown that modular PKS systems likely evolve from gene

duplication/speciation events (Jenke-Kodama and Dittmann, 2009) and the borrelidin cluster in particular has several stretches of repeated sequences shared between modules (Olano et al., 2004b). It would therefore seem likely that the non-cognate ACP-DH pairs resulting from the reductive loop swaps would be more likely to be productive when the ACP and DH pair are from the same PKS cluster, as is the case with the DH-swapped variants. Ultimately, these hypotheses are not mutually exclusive and it is difficult to form strong conclusions about their individual or collective contribution to inefficiencies in the system absent more rigorous biochemical or modeling studies.

Addition of a thioesterase to the PKS variants converted them into de facto assembly lines capable of multiple turnover production of free adipic acid. Intermediate analysis revealed the dehydratase activity was a significant bottleneck in reductive processing and we identified a TE that should be less active against the putative (*S*)-3-hydroxyadipic-ACP substrate compared to its native (*R*)-hydroxy intermediate (Scaglione et al., 2010). While adipic acid was detected in these *in vitro* production experiments, the majority of the product profile was the incompletely reduced 3-hydroxyadipic acid. It is possible that this indicates the fingerprint analysis which suggested the ketoreductase domains in these PKS variants generate an (*S*)-intermediate was incorrect. The more likelier explanation, however, is that the small amount of activity the TE normally demonstrates towards the (*S*)-stereoisomer is sufficient to hydrolyze the 3-hydroxyadipic-ACP intermediate in the timespan of the experiment. In the original study by Scaglione et al, the stereoselectivity of this domain was tested on diffusible SNAC substrate analogs and it is possible that the stereoselectivity is altered when the reactions are intramolecular as when the intermediates are ACP-bound.

Encouragingly, in section 3.3.6, I described how refined N-terminal junction selection along with the BorDH2 swap led to near complete conversion of intermediates to adipic-ACP. Limited throughput for cloning and protein expression of these engineered PKSs precluded close investigation of C-terminal junctions, but it is possible that catalysis could be further improved by optimizing this chimeric boundary as well. Additionally, thioesterase engineering may further bias product distribution towards adipic acid and such efforts would be aided by the availability of crystal structures for multiple TEs ((Kudo et al., 2006; Tsai et al., 2001, 2002)), as well as high-throughput screening methods developed in-house (S. Poust, manuscript in preparation). Near-term future experiments will attempt to produce adipic acid using the engineered extension module alongside the borrelidin loading module which is capable of initiating assembly line biosynthesis with succinyl-CoA (as described in Chapter 2). Optimized constructs will be put into a heterologous host such as *E. coli* with the goal of producing adipic acid from renewable sugar feedstocks.

## 3.5. Materials and methods

### 3.5.1. Reagents and Chemicals

HisPur cobalt resin was purchased from Thermo Scientific, Bradford reagent was from Bio-Rad and SDS-PAGE gels from Life Technologies.

### 3.5.2. Chemical synthesis and NMR data

Solvents (hexanes, ethyl acetate, dichloromethane and methanol) were purchased from EDH; all other reagents were purchased from Sigma-Aldrich or as indicated.

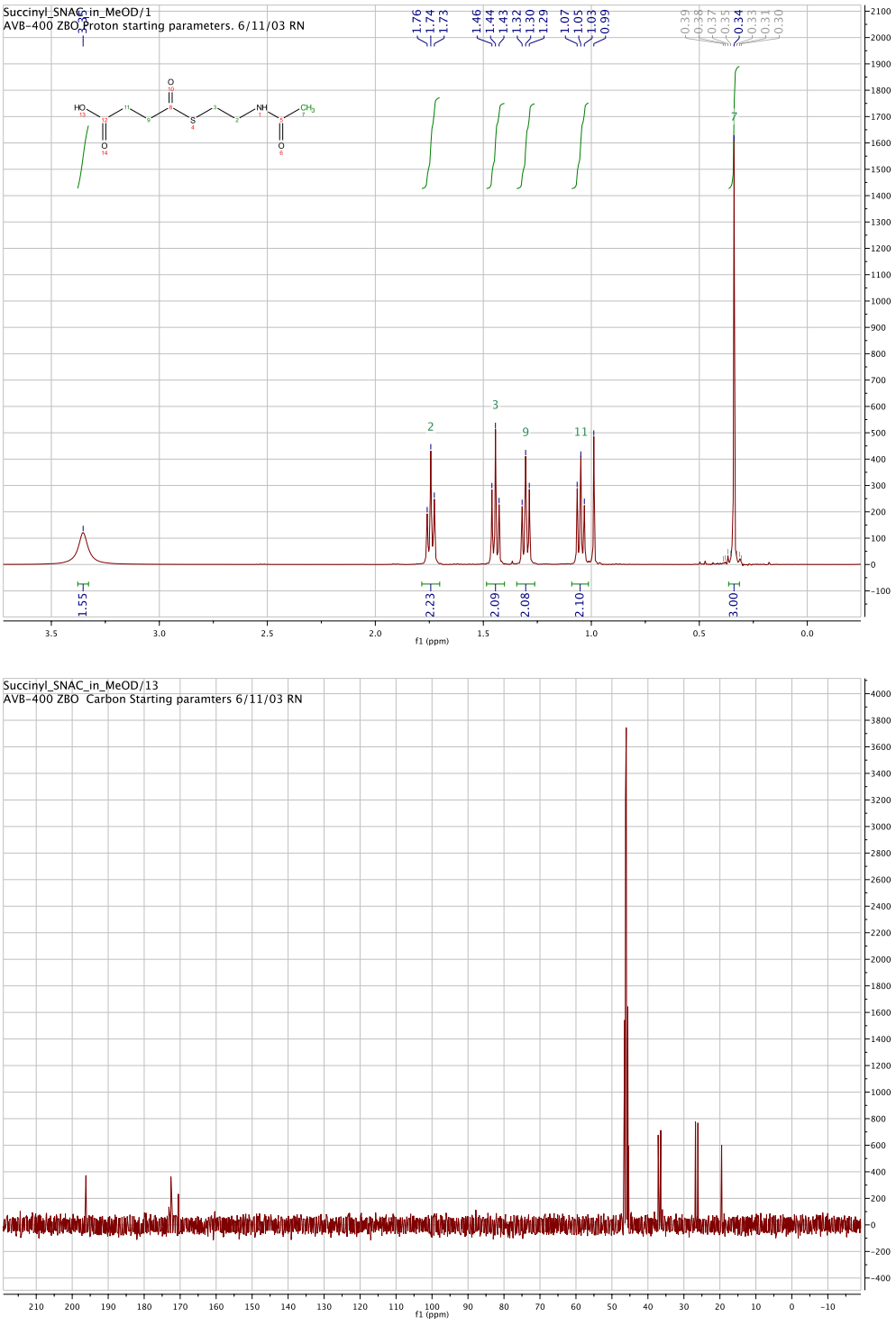
Column chromatography was performed on a Teledyne Isco Combiflash Rf, with RediSep Rf Gold normal phase silica columns.

Gas chromatography – electron impact mass spectrometry (GC-EIMS) was performed on a Agilent5973 - HP6890 GC-MS using a 30 meter db5-ms column

$^{1}$H NMR and $^{13}$C NMR were obtained on a Bruker AVB 400 MHz spectrometer and a Bruker AV 500 MHz spectrometer at the UC Berkeley College of Chemistry NMR facility, funded in part by NSF grant CHE-0130862. Chemical shifts are reported in ppm relative to residual solvent signal ($\delta^{1}$H=3.31 and $\delta^{13}$C=49.0 for Methanol-$d_4$, $\delta^{1}$H=2.05 and $\delta^{13}$C=29.84 for Acetone-$d_6$).

**Succinyl-SNAC**

A 100 ml round-bottom flask was charged with 1 g of succinic anhydride and dissolved in a minimal volume of dichloromethane (DCM). 1 eq. *N*-acetylcysteamine (1.07 ml) was added dropwise to the stirring solution. After overnight incubation at ambient temperature with stirring, the mixture was extracted several times with saturated aqueous sodium bicarbonate solution. The pH of this solution was lowered to approximately 6 with dropwise addition of 1 molar hydrochloric acid in order to protonate unreacted *N*-acetylcysteamine. The mixture was extracted several times with DCM to remove *N*-acetylcysteamine and then the pH of the aqueous solution was lowered to approximately 1.5, again with dropwise addition of 1M HCl. To protonate the title compound. This was extracted several times with ethyl acetate (EtOAc), dried with the addition of sodium sulfate and filtered into a round-bottom flask. The solution was concentrated *in vacuo* to afford a fluffy white powder (0.663 g, 3.02 mmol, 30.4% yield).




**Supplementary Figure 3.1. $^{1}$H (top) and $^{13}$C (bottom) NMR spectra of succinyl-*N*-acetylcysteamine (succinyl-SNAC) in MeOD**

**3-OH-adipic acid standard**

*trans-β*-hydromuconic acid was purchased from Alfa Aesar.

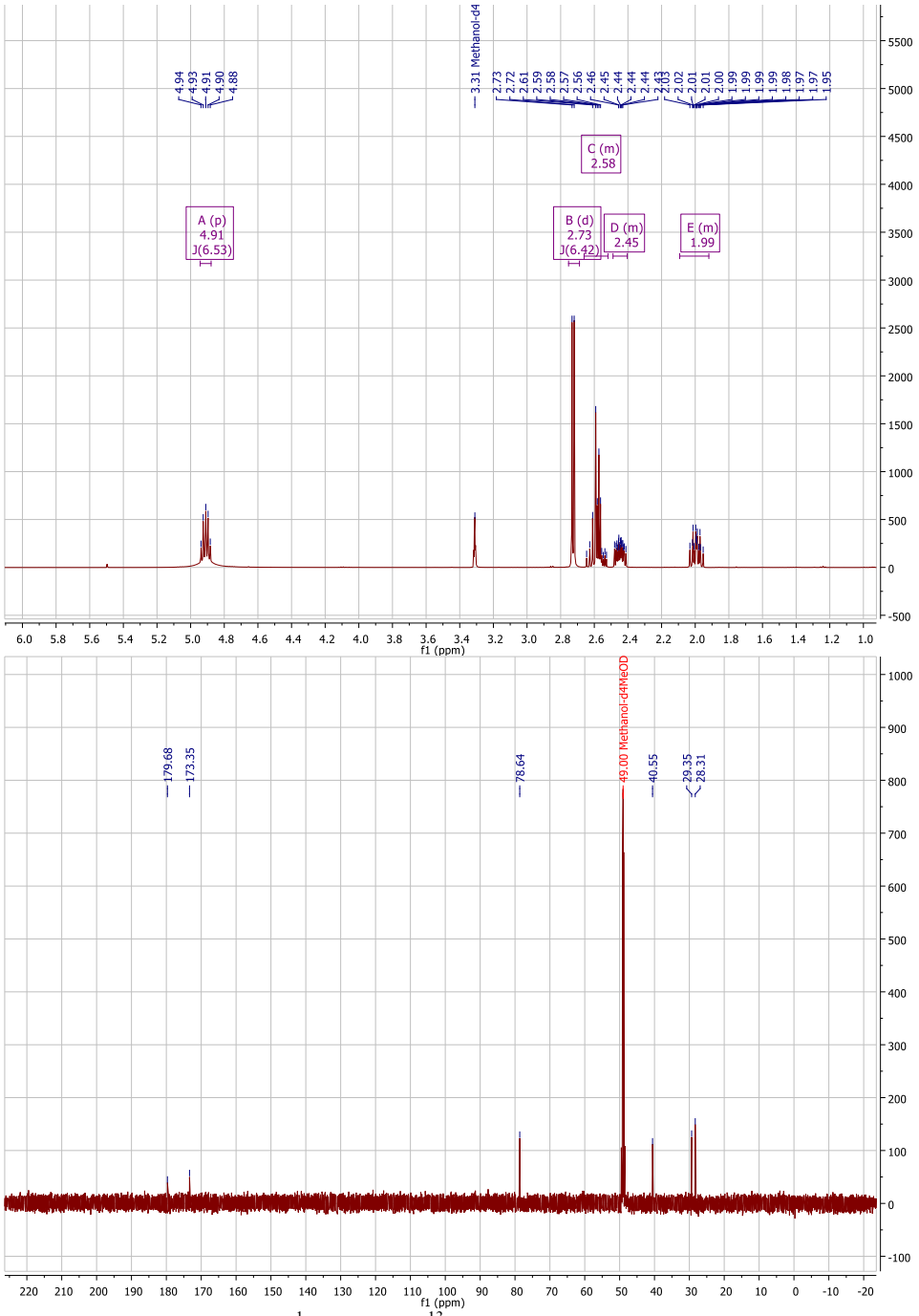
**Solvent-free synthesis of 2-(γ-butyrolactone)acetic acid**

2-(γ-butyrolactone)acetic acid (systematic name: 2-(5-oxotetrahydrofuran-2-yl)acetic acid), InChI=1S/C6H8O4/c7-5(8)3-4-1-2-6(9)10-4/h4H,1-3H2,(H,7,8)

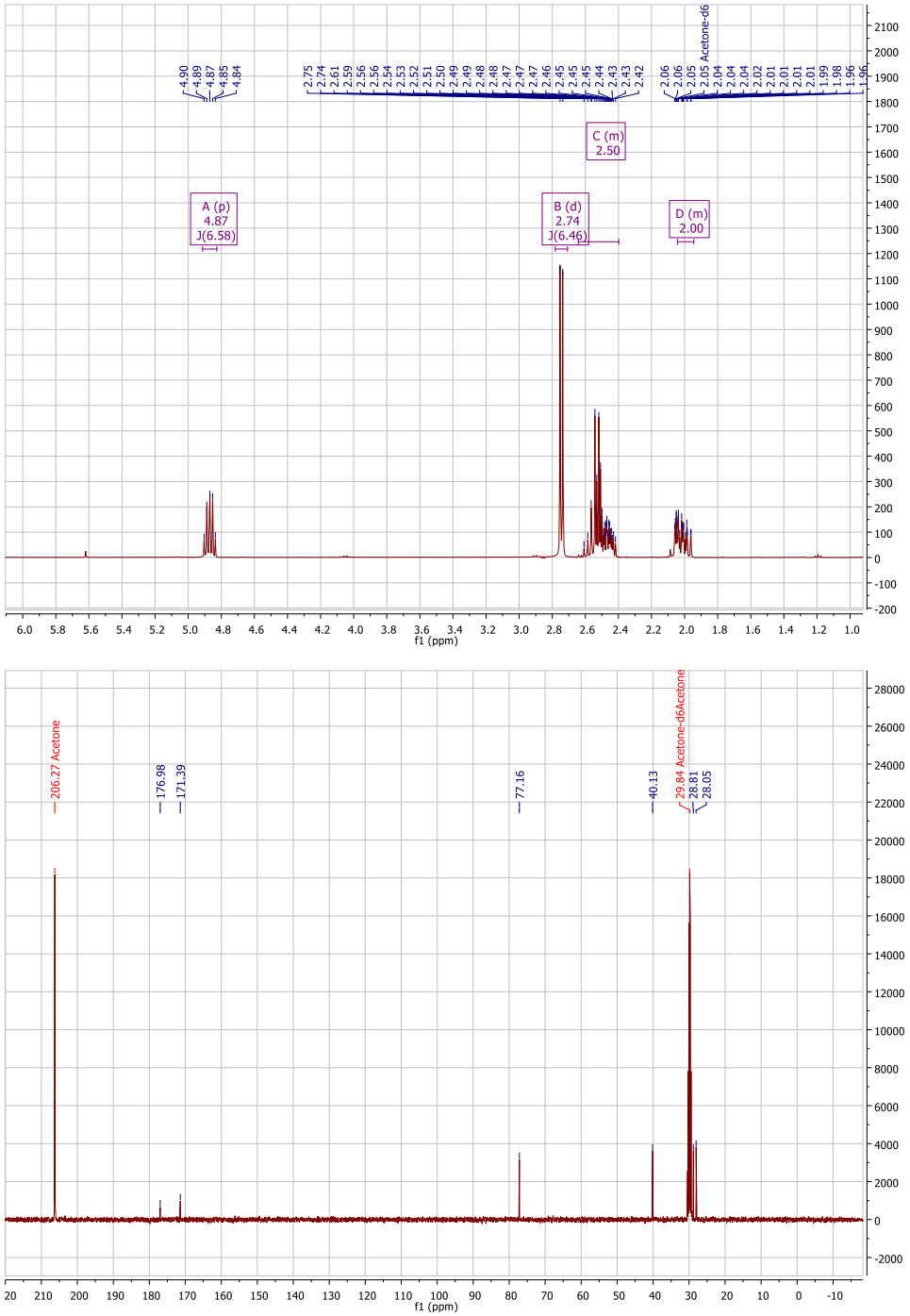
2g of *trans-β*-hydromuconic acid (13.88mmol) and 4g of silica gel (60 Å – 200 mesh) were mixed in a 50mL round-bottom flask with stir bar. The free-flowing mixture was heated to 200°C in a sand bath while gently stirring. The reaction was monitored by pipetting a few milligrams of the hot mixture into 1mL of dichloromethane (DCM), of which 40 µL was treated with 10 µL of *N*,*O*-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) and analyzed by gas chromatography-mass spectrometry (GC-MS). 3 hours into the reaction the mixture starts to turn yellow. After 24 hours, all starting material had been consumed. The mixture was cooled to room temperature and extracted with 50 mL DCM and filtered. The filtrant was extracted with another 50 mL of DCM. The light yellow filtrate was evaporated under reduced pressure and purified by flash chromatography (70:30 Ethyl acetate:Hexane) to afford the title compound as a viscous slightly yellow liquid that solidified upon standing (474mg, 3.29 mmol, 24% yield)
$^{1}$H NMR (500 MHz, MeOD) δ 4.91 (p, $J$ = 6.5 Hz, 1H), 2.73 (d, $J$ = 6.4 Hz, 2H), 2.66 – 2.52 (m, 2H), 2.49 – 2.40 (m, 1H), 2.09 – 1.92 (m, 1H). $^{13}$C NMR (126 MHz, MeOD) δ 179.68, 173.35, 78.64, 49.00, 40.55, 29.35, 28.31.
EIMS (TMS derivative): 201 (7%, $(M\text{-}Me)^{+}$), 159 (27%), 157 (54%), 117 (11%), 101 (8%), 85 (17%), 76 (7%), 75 (100%), 73 (53%), 59 (9%)

**Hydrolysis of 2-(γ-butyrolactone)acetic acid to yield 3-hydroxyadipic acid**

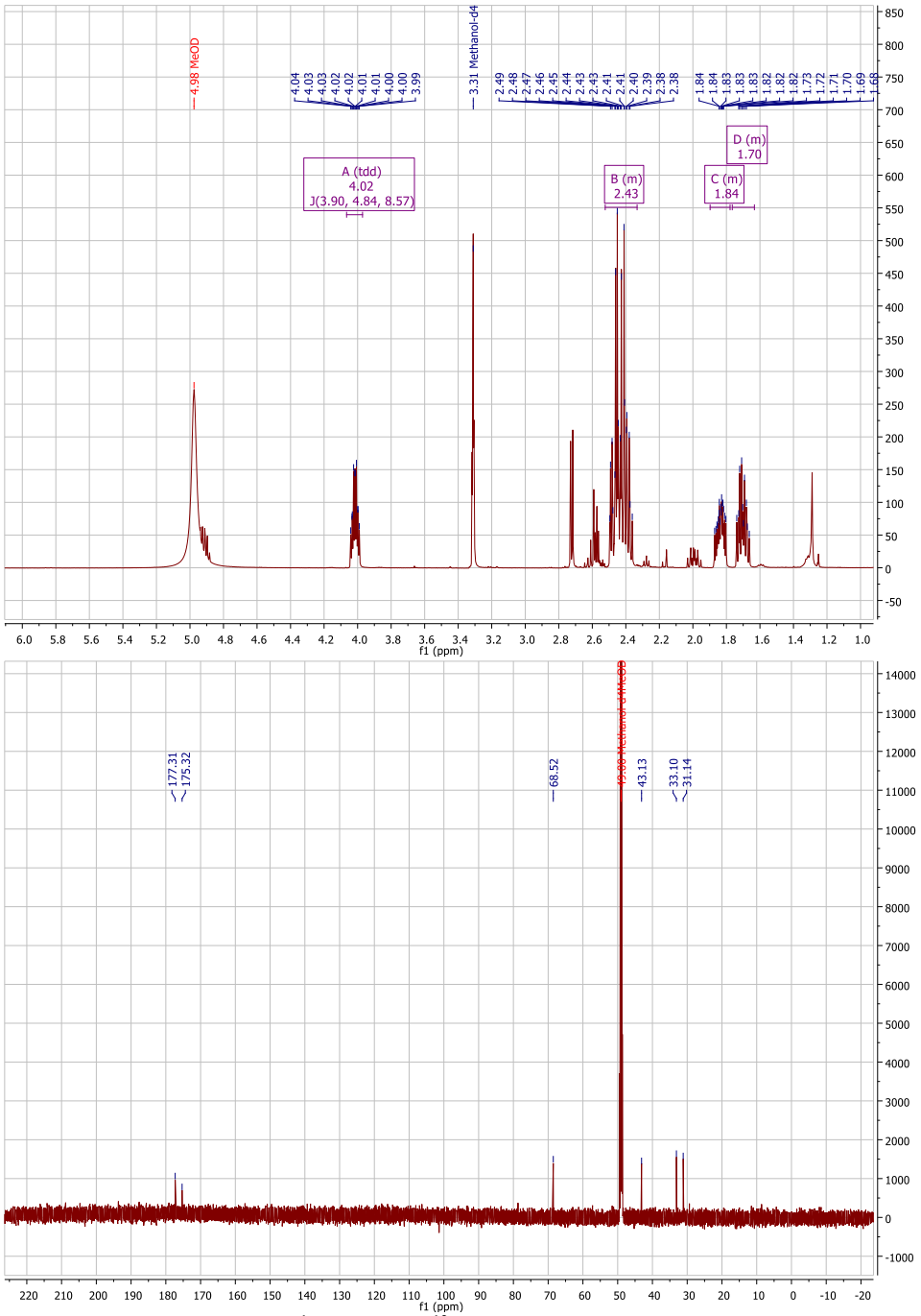
3-hydroxyadipic acid (systematic name: 3-hydroxyhexanediocid acid)
InChI=1S/C6H10O5/c7-4(3-6(10)11)1-2-5(8)9/h4,7H,1-3H2,(H,8,9)(H,10,11)
30mg of 2-(γ-butyrolactone)acetic (208 µmol) was dissolved in 10.4 mL 0.1 M aqueous potassium hydroxide (5 eq.), distributed among the wells of a 96-well PCR plate and heated to 99°C for 3 hours in an Applied Biosciences Venti thermocycler with heated lid (110°C). The solution was consolidated and acidified to pH 3 using 6 M hydrochloric acid. The solution was flash frozen in liquid nitrogen and lyophilized to dryness (~24 h). The remaining powder was extracted with 2×2mL acetone and filtered through a pipette filter (KCl has negligible solubility in acetone). At this point 1 µL of the solution was diluted down to a final volume of 40 µL and derivatized with 10 µL BSTFA to yield the GC-MS chromatogram below. The remaining solution was evaporated under reduced pressure at room temperature to yield the title compound as a white powder (26.7mg, 165 µmol, 79% yield). Re-subjecting the product to GC-MS shows increasing amounts of 2-(γ-butyrolactone) acetic acid over time, suggesting that 3-hydroxyadipic acid spontaneously re-lactonizes at room temperature. Hence, the NMR spectrum below shows 2-(γ-butyrolactone)acetic acid as an impurity.
$^{1}$H NMR (500 MHz, MeOD) δ 4.02 (tdd, $J$ = 8.6, 4.8, 3.9 Hz, 1H), 2.52 – 2.33 (m, 4H), 1.90 – 1.78 (m, 1H), 1.77 – 1.63 (m, 1H). $^{13}$C NMR (126 MHz, MeOD) δ 179.68, 173.35, 78.64, 49.00, 40.55, 29.35, 28.31.
EIMS (tris(TMS) derivative): 363 (32% $(M\text{-}Me)^{+}$), 247 (26%), 233 (11%), 203 (12%), 149 (14%), 147 (55%), 133 (10%), 129 (24%), 75 (27%), 73 (100%)



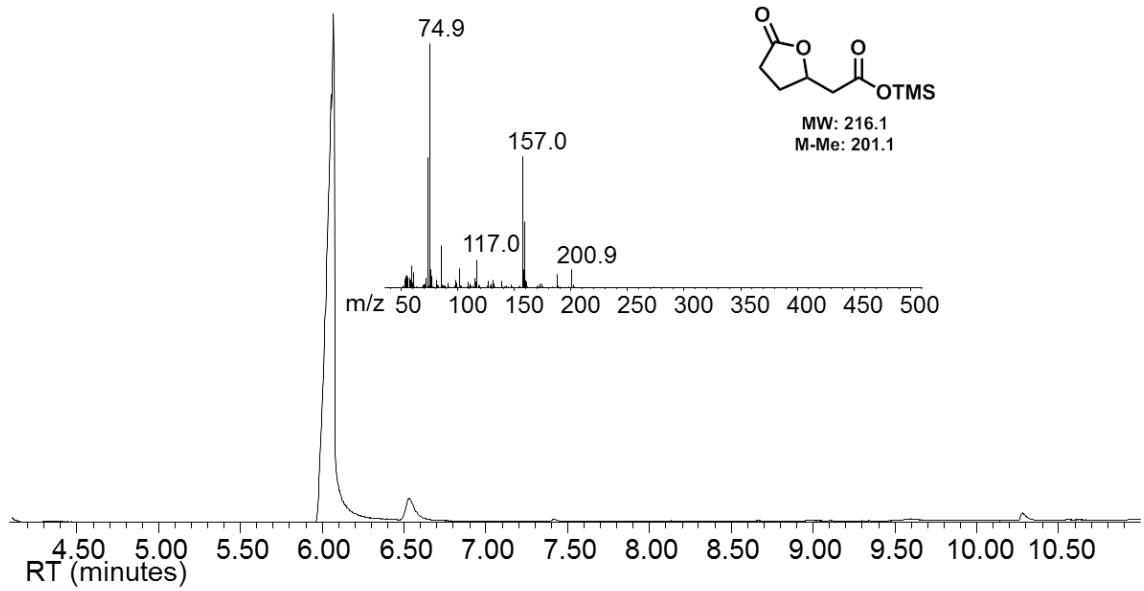

**Supplementary Figure 3.2**. $^{1}H$ (top) and $^{13}C$ (bottom) NMR spectra of 2-(γ-butyrolactone)acetic acid in MeOD



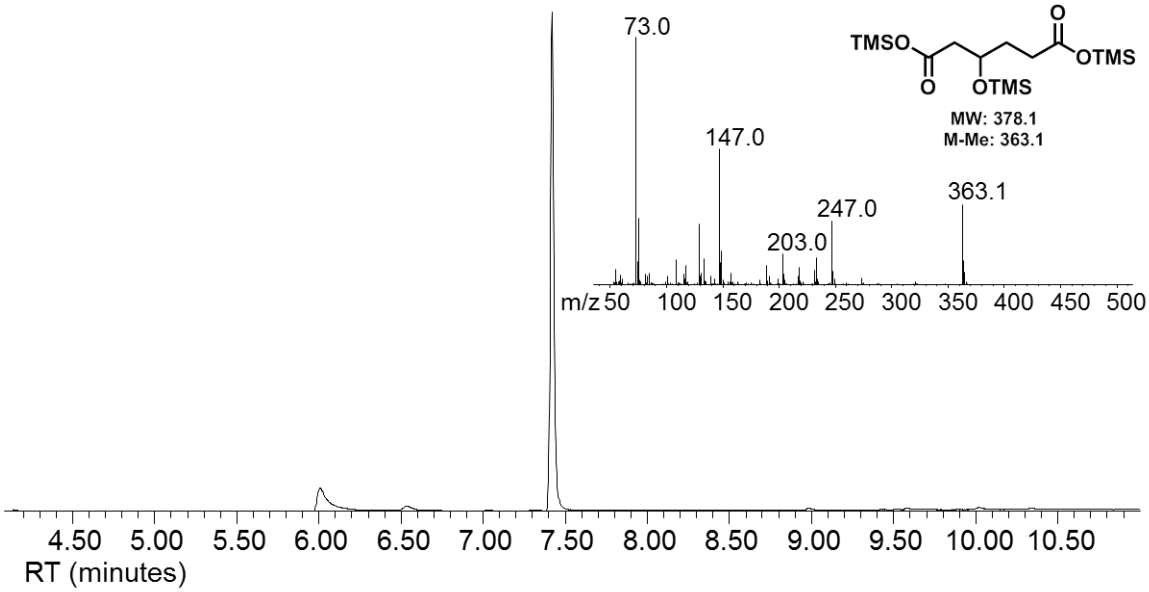

**Supplementary Figure 3.3**. $^{1}H$ (top) and $^{13}C$ (bottom) NMR spectra of 2-(γ-butyrolactone)acetic acid in acetone-d6. Because the pentuplet at δ=4.9 ppm overlaps with a methanol solvent peak, spectra for the same sample in acetone were also acquired.




**Supplementary Figure 3.4**. $^{1}H$ (top) and $^{13}C$ (bottom) NMR spectra of 3-hydroxyadipic acid in MeOD.




**Supplementary Figure 3.5**. GC-MS chromatograms and EI-MS spectra of 2-(γ-butyrolactone)acetic acid (TMS derivative).




**Supplementary Figure 3.6**. GC-MS chromatograms and EI-MS spectra of 3-hydroxyadipic acid (TMS derivative).

### 3.5.3. Plasmid list and construction

#### 3.5.3.1. Reductive loop library

**AurB**, **IdmO**, **CurK** and **SpnB** DNA was generously provided by Ryan Phelan. Codon-optimization and synthesis of AurB was performed by Genscript; codon-optimization and synthesis of IdmO, CurK and SpnB was performed by the Joint Genome Institute (JGI). **NanA2** DNA was generously provided by Satoshi Yuzawa; codon-optimization and synthesis was performed by DNA2.0.

**General design and construction strategy**

Amino acid sequences for various modules were aligned using the MUSCLE or Clustal Omega algorithms (Edgar, 2004; Sievers et al., 2014). The j5 algorithm and Device Editor graphical user interface were used to design DNA assemblies (Hillson et al., 2012). All DNA pieces were amplified via PCR with either Q5 or Phusion polymerases (New England BioLabs) according to manufacturer's recommendations. Gel-extracted DNA was assembled via Gibson cloning using Gibson Assembly® master mix (New England BioLabs). In the case of construct A1.5 (pARH159), Gibson assembly failed and sequence was introduced by oligonucleotides using "round the horn" PCR with pARH137 (A2) as a template. A similar strategy was used to create (A,S)1.5c constructs (pARH163, 164 respectively) starting from (A,S)2c constructs (pARH147, 149 respectively). A complete list of plasmids and open reading frames appears in the appendix.

#### 3.5.3.2. pARH150 - BorDH2 monodomain

BorDH2 domain boundaries were selected after (Vergnolle et al., 2011) and DNA was codon-optimized and synthesized as a gBlock (Integrated DNA Technologies) and ligated into the pET28a vector (Novagen) to yield an N-terminal hexahistidine tagged construct.

### 3.5.4. Protein expression tests and purification

#### 3.5.4.1. SDS-PAGE analysis

Plasmids were introduced into *E. coli* strain BAP1 (Pfeifer, 2001) and cells (5 ml) were grown at 37C in lysogeny broth (LB) to an O.D. of approximately 0.6 and then 60 ng/ml anhydrotetracycline and 200 uM isopropyl-β-D- galactopyranoside (IPTG) were added to induce expression of PKS proteins and Sfp, respectively. Cultures continued incubation at 18C for 24 hours and then 0.5 ml was harvested by centrifugation (10000g, 3 min) and resuspended in lysis buffer (300 mM NaCl, 50 mM phosphate, pH 6.8, 10 mM imidazole) supplemented with 0.1 mg/ml lysozyme. Suspensions were lysed by sonication and insoluble debris was pelleted (20000g, 20 min, 4C). The supernatant was then transferred to a fresh tube and the pellet comprising insoluble protein and cellular debris was resuspended in 500 ul lysis buffer. 30 ul of soluble and insoluble fractions was combined with 10 ul 4x LDS loading buffer (Life Technologies) and 2 ul 1M dithiothreitol. Protein was denatured at 95C for 10 minutes and then loaded on to a 3-8% tris-acetate gel and run per manufacturer's recommendations (Life Technologies).

#### 3.5.4.2. Purification of PKS constructs:

Cultures were grown and induced as above except for the following differences: The culture volumes were increased to 500 ml, terrific broth (TB) was used and cells were induced at O.D. of approximately 1.0. Cells were lysed by passing several times through an Emulsiflex C3 homogenizer (Avestin) and cellular debris was removed by centrifugation (15000g, 30 minutes). Cobalt resin (2-3 ml) was added to the supernatant and mixed at 4C for one hour before being applied to a fritted column. Resin was washed until flow-through resulted in no color change when mixed with Bradford reagent. Proteins were eluted with several resin volumes of elution buffer (300 mM NaCl, 50 mM phosphate, pH 6.8, 200 mM imidazole) and concentrated via spin filtration (Amicon, 100 kDa MWCO). Concentrated eluate was exchanged into storage buffer (50 mM phosphate, pH 6.8, 10% glycerol) using a PD-10 column (GE Life Sciences), and then further concentrated prior to being flash frozen in liquid nitrogen and stored at -80C.

#### 3.5.4.3. Purification of BorDH2:

BorDH2 monodomain was purified as above with the following exceptions: 1000 ml TB and 3 ml of cobalt resin were used. Protein was concentrated with a 10 kDa MWCO filter and stored as a 50% glycerol solution at -20C after buffer exchange.

### 3.5.5. Experimental setup and sample processing

#### 3.5.5.1. Intermediate analysis of reductive loop library

A master mix (final concentrations: 0.5 mM succinyl-SNAC, 0.5 mM malonyl-CoA, 1 mM NADPH, 2.5 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to separate tubes, to which 1uM final concentration of each respective PKS variant was added and incubated at room temperature overnight (~16 hr).

#### 3.5.5.2. Intermediate analysis of BorMod1 with BorDH2 *in trans*

A master mix (final concentrations: 1 mM succinyl-SNAC, 0.2 mM malonyl-CoA, 1 mM NADPH, 2.5 mM TCEP, 5 uM BorMod1 in 100 mM phosphate buffer pH 6.8) was aliquoted to separate tubes. Equal volumes of variously concentrated BorDH2 protein was added such that final concentrations of BorDH2 were 0, 5, 20 and 50 uM. An otherwise identical reaction wherein BorA2 protein was omitted was set up as a negative control. Reactions were incubated at room temperature for two hours.

#### 3.5.5.3. Intermediate analysis of reductive loop library with BorDH2 *in trans*

A master mix (final concentrations: 1 mM succinyl-SNAC, 0.2 mM malonyl-CoA, 1 mM NADPH, 2.5 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to separate tubes, to which 5 uM final concentration of each respective PKS variant was added. One half the volume of these respective mixes was aliquoted to a fresh set of tubes to which 50 uM final concentration of BorDH2 was added. Reactions were incubated at room temperature overnight (~16 hr).

#### 3.5.5.4. Intermediate analysis of library with BorDH2 *in cis* and *in trans*

Experiment was set up as above (BorDH2 *in trans*…) except final concentrations of PKS variants were 1 uM for unintended reasons.

### 3.5.5.5. Adipic acid production with ttmTE variants

100 microliter reactions containing final concentrations: 1 mM succinyl-SNAC, 0.5 mM malonyl-CoA, 1 mM NADPH, 1 mM TCEP and 10 uM enzyme in 100 mM phosphate buffer pH 6.8 were set up and incubated at room temperature overnight (~16hr).

### 3.5.5.6. Intermediate analysis of junction 1.5 variants -/+ BorDH2 *in cis*

A master mix (final concentrations: 1 mM succinyl-SNAC, 0.5 mM malonyl-CoA, 1 mM NADPH, 2.5 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to separate tubes, to which 5uM final concentration of each respective PKS variant was added and incubated at room temperature overnight (~16 hr).

For intermediate analysis, samples were digested with 1:20 w/w porcine trypsin (Sigma-Aldrich) for 4-6 hours at 37C. For adipic acid production, samples were filtered through 3K molecular weight cut off spin filters (Amicon) which were washed prior to use in later experiments by filtration of 500 ul of LC-MS.

### 3.5.6. Intermediate analysis analytical methods

### 3.5.6.1. Initial activity screen:

Samples (section 3.3.2) were analyzed on an AB Sciex (Foster City, CA) 2000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 0.5 µg of total peptide was injected onto a Sigma (St. Louis, MI) Ascentis Peptide Express C-18 column (0.5 mm x 150 mm) via an autosampler.

A 33-minute method was used and consisted of a 10 µL/minute flow rate, starting with 95 % Buffer A (2% Acetonitrile, 0.1% formic acid) and 5% Buffer B (98 % Acetonitrile, 0.1% formic acid) for 1 minute, followed by an increase to 80% Buffer B in 15 minutes where it was held for 2 minutes, followed by a quick ramp back down to 5% B, where it was subsequently held for 14 minutes to allow for column equilibration for the next run. The peptides eluting from the column were ionized by an AB Sciex nanospray source (curtain gas flow: 20 l/min, temperature: ambient, ion spray voltage: 4,800 V, ion source gas flow: 20 l/min, entrance potential: 10 V) operating in positive-ion mode.

| ID | Q1 | Q3 | Declustering potential | Collision energy |
|---|---|---|---|---|
| ACP1_ctrl | 680.38 | 846.48 (y8) | 100 | 19.1 |
| Holo-ACP1 | 905.76 | 261.12 | 100 | 41.8 |
| Keto-ADA-ACP1 | 953.11 | 403.15 | 100 | 41.8 |
| hydroxy-ADA-ACP1 | 953.77 | 405.16 | 100 | 41.8 |
| 2,3-ene-ADA-ACP1 | 947.78 | 387.16 | 100 | 41.8 |
| ADA-ACP1 | 948.45 | 389.17 | 100 | 41.8 |

**Supplementary Table 3.1**. Mass spectrum parameters for initial intermediate analysis experiments

**3.5.6.2. Subsequent experiments:**
Samples (sections 3.3.3 through 3.3.6) were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 1-2 µg of total peptide was injected onto a Sigma (St. Louis, MI) Ascentis Peptide Express C-18 column (2.1 mm x 50 mm) via an autosampler. A 20.5-minute method was used with the same buffer compositions as above but with a flow-rate of 400 ul/min. The method begins with 5% buffer B for 1.2 minutes followed by a rapid rise to 25% over 1 minute and then a very slow rise to 36% over 10 minutes. After the slow gradient step, buffer B was rapidly increased to 90%, held, and dropped back down to re-equilibrate the column as above. The peptides eluting from the column were ionized by a Turbo V Ion source (curtain gas flow: 20 l/min, temperature: 400 C, ion spray voltage: 4,800 V, ion source gas flow: 50 l/min, entrance potential: 10 V) operating in positive-ion mode.

| ID | Q1 | Q3 | Declustering potential | Collision energy |
|---|---|---|---|---|
| ACP1_ctrl | 680.38 | 846.48 (y8) | 125 | 40 |
| Holo-ACP1 | 905.76 | 261.12 | 50 | 44 |
| Keto-ADA-ACP1 | 953.11 | 403.15 | 50 | 44 |
| hydroxy-ADA-ACP1 | 953.77 | 405.16 | 50 | 44 |
| 2,3-ene-ADA-ACP1 | 947.78 | 387.16 | 50 | 44 |
| ADA-ACP1 | 948.45 | 389.17 | 50 | 44 |

**Supplementary Table 3.2**. Mass spectrum parameters for later intermediate analysis experiments

For all experiments, transitions were monitored using a collision cell exit potential of 10 V.

ACP1_ctrl peptide: VVESVAFGVPSLR

ACP1 apo peptide: AAIGPDSSFHAIGFDSLTAVELR
(site of phosphopantetheinylation underlined)

Methods were designed and data collected in Analyst 1.5.1 and data was quantified in MultiQuant 2.1 (AB Sciex)

**Data analysis**
Raw data for each transition was normalized by dividing a transition's peak area by that of a control peptide present in BorMod1, but which does not participate in catalysis and should therefore be invariant across samples ("ACP1_ctrl") to generate values in "control peptide equivalents."

**3.5.7. Adipic acid analytical methods**

Adipic acid (commercially available) and 3-hydroxyadipic acid (crude synthesis) were directly infused into the mass spectrometer operating in negative mode and a scan was conducted to identify product ions during adjustment of relevant acquisition parameters

### 3.5.7.1. Preliminary adipic acid production experiment

Samples (sections 3.3.5) were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. Indicated volumes of each reaction were injected onto a Phenomenex (Torrance, CA) Kinetex XB C-18 column (3 mm x 100 mm, 1.7u) via an autosampler. A 35 minute method was used with a flow-rate of 140 ul/min and started with 90 % Buffer A (water, 0.1% formic acid) and 10% Buffer B (methanol, 0.1% formic acid) for 1.2 minutes followed by a rise to 100% buffer B over 13.8 minutes where it was held for 2 minutes and then a return to 10% for 18 minutes to re-equilibrate the column. Analytes eluted from column were ionized using a Turbo V Ion source (curtain gas flow: 20 l/min, temperature: 400 C, ion spray voltage: -4,500 V, ion source gas flow: 60 l/min, entrance potential: -10 V) operating in negative-ion mode.

| ID | Q1 | Q3 | Declustering potential | Collision energy |
|---|---|---|---|---|
| ADA127 | 145 | 127 | -45 | -18 |
| ADA101 | 145 | 101 | -45 | -18 |
| ADA83 | 145 | 83 | -45 | -18 |
| ADA81 | 145 | 81 | -45 | -18 |
| ADA145 | 145 | 145 | -45 | -18 |

**Supplementary Table 3.3**. Mass spectrum parameters for initial adipic acid detection

### 3.5.7.2. Subsequent adipic acid production experiment

Samples were analyzed as above with the following differences: 10 ul of reactions was injected onto a Phenomenex Kinetex XB C-18 column (2.1 mm X 50 mm, 2.6u). A 16-minute method was used with a flow-rate of 200 ul/min and the same buffers as above. It started with 2.5% buffer B for 3 minutes followed by a rise to 90% B over 3 minutes where it was held for 1 minutes before returning to 2.5% B for 8 minutes to re-equilibrate the column.

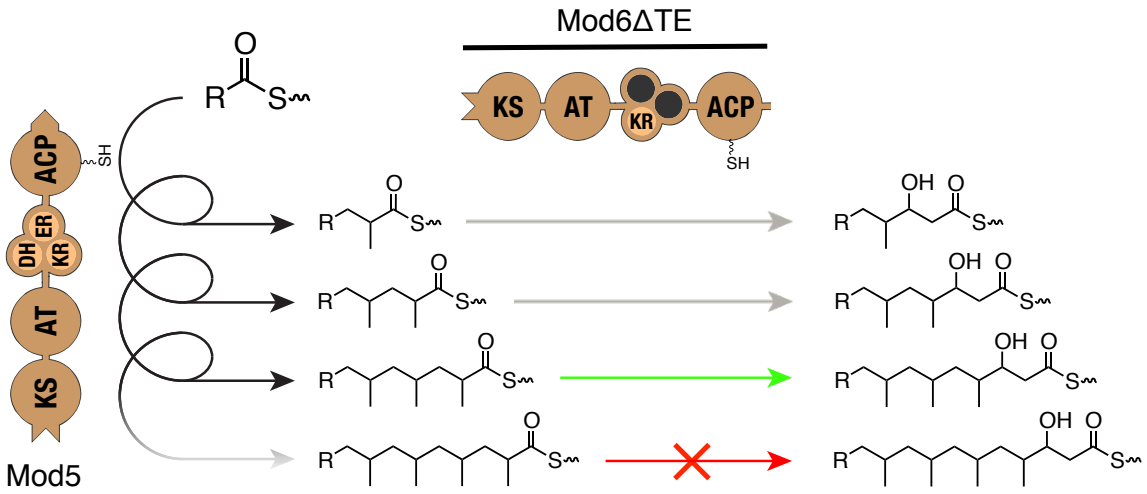
| ID | Q1 | Q3 | Declustering potential | Collision energy |
|---|---|---|---|---|
| ADA127 | 145 | 127 | -45 | -18 |
| ADA101 | 145 | 101 | -45 | -18 |
| ADA83 | 145 | 83 | -45 | -18 |
| ADA145 | 145 | 145 | -45 | -18 |
| 3OH143 | 161 | 143 | -45 | -18 |
| 3OH101 | 161 | 101 | -45 | -18 |
| 3OH99 | 161 | 99 | -45 | -18 |

**Supplementary Table 3.4**. Mass spectrum parameters for adipic acid and 3-OH adipic acid detection

# **4.** Chapter 4 – Characterization of an iterative type I PKS module

### 4.1. Abstract

The organization of modules in type I modular polyketide synthase systems is typically reflected in the structure of the product it creates, termed "colinearity." In recent years, a growing list of clusters have been identified in which this defining biosynthetic logic does not apply and one or more modules in an otherwise modular system can act iteratively to catalyze more than a single round of extension. The molecular determinants that allow such programmed iteration to proceed and with high fidelity are currently unknown. Here we heterologously express and purify an iterative module from the borrelidin system and perform *in vitro* characterization of its substrate range and product profile. We demonstrate that the isolated module is necessary and sufficient for iteration to take place, the module accepts a variety of substrate analogs and that the identity of the starter substrate has a profound impact on the number of iterations that are catalyzed by the module. Furthermore, we show that iteration is not dependent on the acyltransferase domain nor the identity of the extender unit and propose a model for chain length regulation based on kinetic data and interaction with the downstream acceptor module.



## 4.2. Introduction

One of the most striking features of type I modular PKS systems is the parallel between a product's structure and the organization of modules in the PKS proteins responsible for the product's synthesis (Donadio et al., 1991). This so-called "colinearity" can allow for the deduction of a product's structure from sequence information alone and conversely, to predict the modular organization of the biosynthetic cluster responsible for the synthesis of a structurally characterized PKS product. Because of the tight one module, one extension/reduction correlation, PKSs were celebrated as a platform well-suited for engineering. Indeed, many examples of "re-programming" such systems via replacement of loading modules, whole module insertion/deletion, domain swaps and mutagenesis led to production of numerous novel analogs of naturally occurring metabolites (McDaniel, 1997; Menzella et al., 2005; Zhang et al., 2006). Such exuberance has been tempered by the widely recognized phenomenon that such re-programming efforts are usually associated with significant penalties in terms of enzyme kinetics and overall product titers and therefore the extent of modularity in these systems has been revisited (Khosla et al., 2009).

In recent times, as more PKS clusters are sequenced and compared to their expected product, the colinearity paradigm of "modular" type I PKSs has also been challenged. As far back as 1990, scientists at Glaxo Group Research reported on alterations in product profiles from a mutagenized *Streptomyces* isolate wherein the intermediate had apparently bypassed the penultimate module of the cluster but was nonetheless extended and cyclized by the final module (International symposium on genetics of industrial microorganisms et al., 1990). Later, in an unmutagenized strain of the Myxobacterium *Sorangium cellulosum* a variety of analogs of the usual 16-membered macrolactone, epothilone, were isolated including 14-membered (epothilone K) and 18-membered (epothilone I) variants (Hardt et al., 2001). Comparison to the epothilone gene cluster (Julien et al., 2000; Molnar et al., 2000) indicated that the pendant epothilone K intermediate had bypassed module 7 while the epothilone I intermediate had undergone an additional round of extension/reduction catalyzed by module 5. Such events have been called "skipping" and "stuttering" respectively, and have been observed in a variety of other PKSs, including the well-characterized erythromycin cluster (Wilkinson et al., 2000). It should be emphasized that these analogs are always present as a minor component of the total PKS product pool and therefore must be regarded as aberrant, though it has been proposed that such analogs can drive evolution of PKS diversity if they undergo positive selection (Meiser et al., 2008; Moss et al., 2004).

In contrast to the above examples of aberrant loss of colinearity, a growing list of examples of *programmed* violations of colinearity are known. The biosynthetic gene cluster for stigmatellin, produced by the Myxobacterium *Stigmatella aurantiaca*, was deduced by Gaitatzis and co-workers and was found to have one less module than expected for the undecaketide product (Gaitatzis et al., 2002). The following year, analysis of the aureothin gene cluster by He et al revealed that the loaded *para*-aminobenzoic acid undergoes two rounds of extension and reduction before transfer to the next module (He and Hertweck, 2003). Finally in 2004, Olano et al described the biosynthesis of borrelidin (also described in Chapter 2), whose intermediate undergoes a

remarkable three extensions with full reduction by the penultimate module in order to form the nonaketide product from six extension modules (see **Figure 4.1**) (Olano et al., 2004a). No stigmatellin or aureothin analogs have been reported and "exhaustive HPLC and LCMS analysis" was unable to find borrelidin-related hepta-, octa- or decaketide analogs (Olano et al., 2003) and therefore these repeated extensions must be considered programmed events quite apart from module stuttering. Subsequent experiments using gene fusions and heterologous expression provided the first functional confirmation that these modules act iteratively and do not act modularly in a head-to-tail multimeric state or require additional host factors such as an intermediary acyl carrier protein (He and Hertweck, 2005). Other examples of clusters with modules that perform programmed iteration include neoaureothin, lankacidin and DKxanthene (Meiser et al., 2008; Tatsuno et al., 2007; Traitcheva et al., 2007).

Iteration in type I PKS modules can be considered a reversion to a more fatty-acid synthase (FAS)-like state. Indeed, extensive phylogenetic analysis indicates that the ubiquitous FAS was the ancient progenitor of PKSs (Jenke-Kodama, 2005) and therefore the evolution of modularity can be viewed as a gain-of-function event. In an iterative system like FAS, after the first round of extension intermediates on the ACP are transacylated to the sole ketosynthase, which allows the ACP to be re-loaded with another extender unit and the next round of extension/reduction occur. In contrast, in canonical type I modular PKS systems, each ACP must interact with the KS in its own module ($KS_n$-$ACP_n$) in order for condensation to occur, and then subsequently interact with the downstream module's KS ($ACP_n$-$KS_{n+1}$) in order to translocate the growing acyl-chain. How a multimodular PKS system maintains unidirectional biosynthesis, or more specifically, how back transfer to the $KS_n$ domain is prevented, has been a long standing question in the field. One proposal based on the observed high-occupancy of active sites in PKS domains is termed "congestion control" (Hong et al., 2009). In this model, the chain extension and translocation steps in polyketide biosynthesis are highly coordinated and back transfer is precluded by a highly occupied ketosynthase domain. In contrast, the Khosla group has proposed a "ratchet system" to prevent back transfer (Kapur et al., 2012) which does not invoke KS domain occupancy. They found the acyl carrier proteins in erythromycin PKS have distinct epitopes for chain-extension and chain-translocation interactions with their intra- ($KS_n$-$ACP_n$) and intermodular ($ACP_n$-$KS_{n+1}$) KSs, respectively. Back transfer is therefore precluded by a mismatch in the chain-translocation epitopes of $ACP_n$ for $KS_n$. A simple fragment exchange in the Helix I region between ACP proteins allowed a second round of extension to proceed, supporting their model, though it is unclear why iteration stopped at the second cycle. The most recent, and perhaps definitive model is based on the examination of different conformational states of an isolated, intact PKS module using electron cryo-microscopy (cryoEM) (Dutta et al., 2014; Whicher et al., 2014). The structure of PikAIII from the pikromycin cluster reveals a central "reaction chamber" occupied by the ACP during the extension/reduction reactions. Remarkably, upon completion of intermediate processing, large structural rearrangements "flip-out" the ACP into a position in which chain translocation to the next module can occur. In this way, back transfer is precluded based on geometric constraints alone, and intriguingly the primary determinant of the ACP's

position was not dictated by protein-protein interactions, but rather the identity of the tethered substrate.

The specific molecular determinants that dictate the relative location of an ACP remain to be elucidated for iterative as well as canonical PKS modules. Furthermore, no *in vitro* studies have been performed on an iterative module from an otherwise modular type I PKS system. By removing the module from its normal cellular milieu in such studies, precise control over provision and identity of starter and extender units, reducing power and reaction time can be exerted. We therefore purified the iterative module (BorMod5) from the borrelidin cluster with the aim of performing the first *in vitro* characterization of an iterative module in the hopes of shedding some light on these unanswered questions.

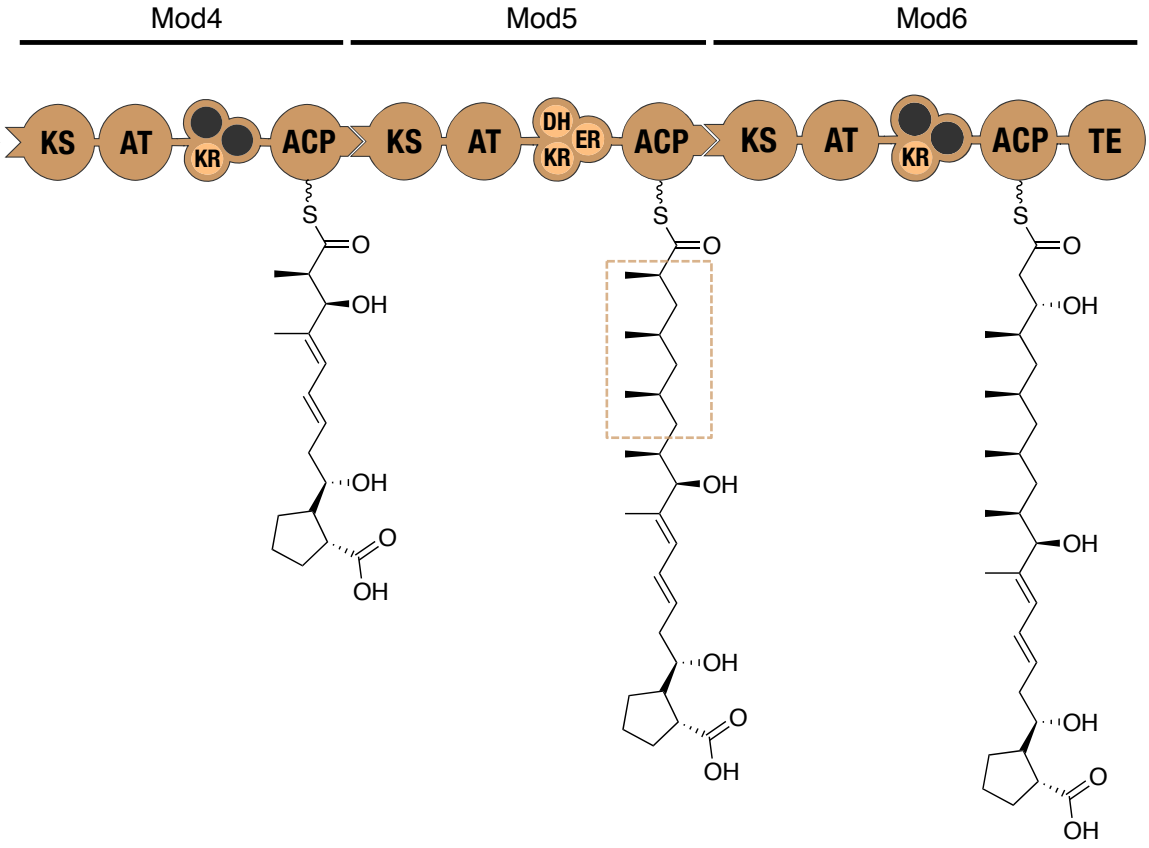



**Figure 4.1**: Excerpt of the borrelidin PKS cluster's domain structure and intermediates including the iterative BorMod5 (adapted from (Olano et al., 2004a). Chain extensions/reductions catalyzed by BorMod5 are boxed.

## 4.3. Results

### 4.3.1. Soluble expression and purification of BorMod5 as an MBP fusion

BorMod5 was *E. coli* codon-optimized and cloned into two vector backbones to test expression. The first, pET28a, is driven by a strong T7 promoter while the second, pARH090 (see section 2.5.3), uses the pTet promoter and expresses the protein with an N-terminal maltose binding protein (MBP) domain which anecdotally promotes solubility of PKS proteins in *E. coli* (data not shown). Solubility was assessed via SDS-PAGE analysis of whole cell lysates and are shown in **Figure 4.2**.

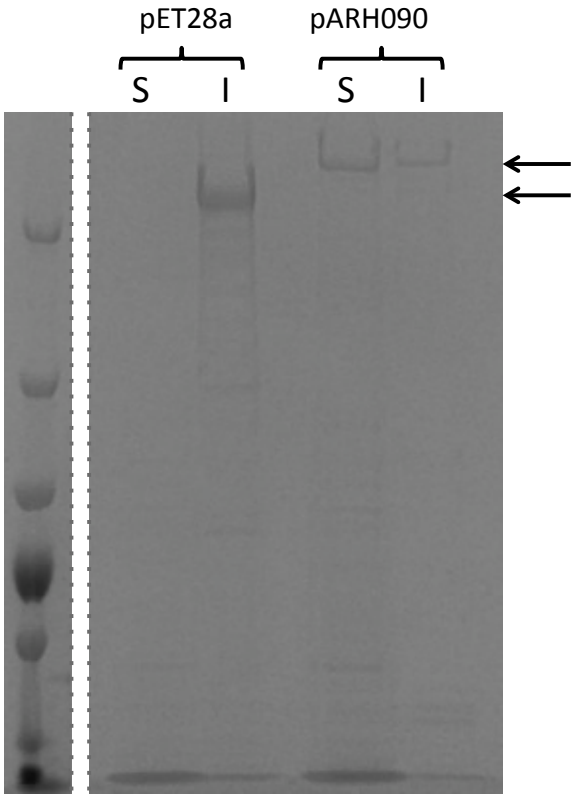



**Figure 4.2**. SDS-PAGE analysis (whole cell lysate) of proteins expressed from different vectors. Shown is soluble (S) and insoluble (I) protein fractions. Black arrow indicates protein of interest (note shift in migration from MBP fusion)

The SDS-PAGE analysis reveals the protein is expressed in a partially soluble form only from pARH090 whereas expression from pET28a is completely insoluble. Protein was expressed and purified as the intact MBP fusion and used for subsequent experiments (see Materials and Methods for details)

### 4.3.2. BorMod5 is sufficient to demonstrate iteration *in vitro*

In order to test if (MBP-)BorMod5 alone is necessary and sufficient to perform iterative extension, we incubated purified BorMod5 protein with a substrate analog, isobutyryl-*N*-acetylcysteamine as a starter unit along with methylmalonyl-CoA and NADPH as extender unit and reducing power, respectively. Intermediate analysis was performed

using liquid chromatography time-of-flight mass spectrometry (LC-TOF) operating in MS1 mode to allow direct comparison between peak intensities.

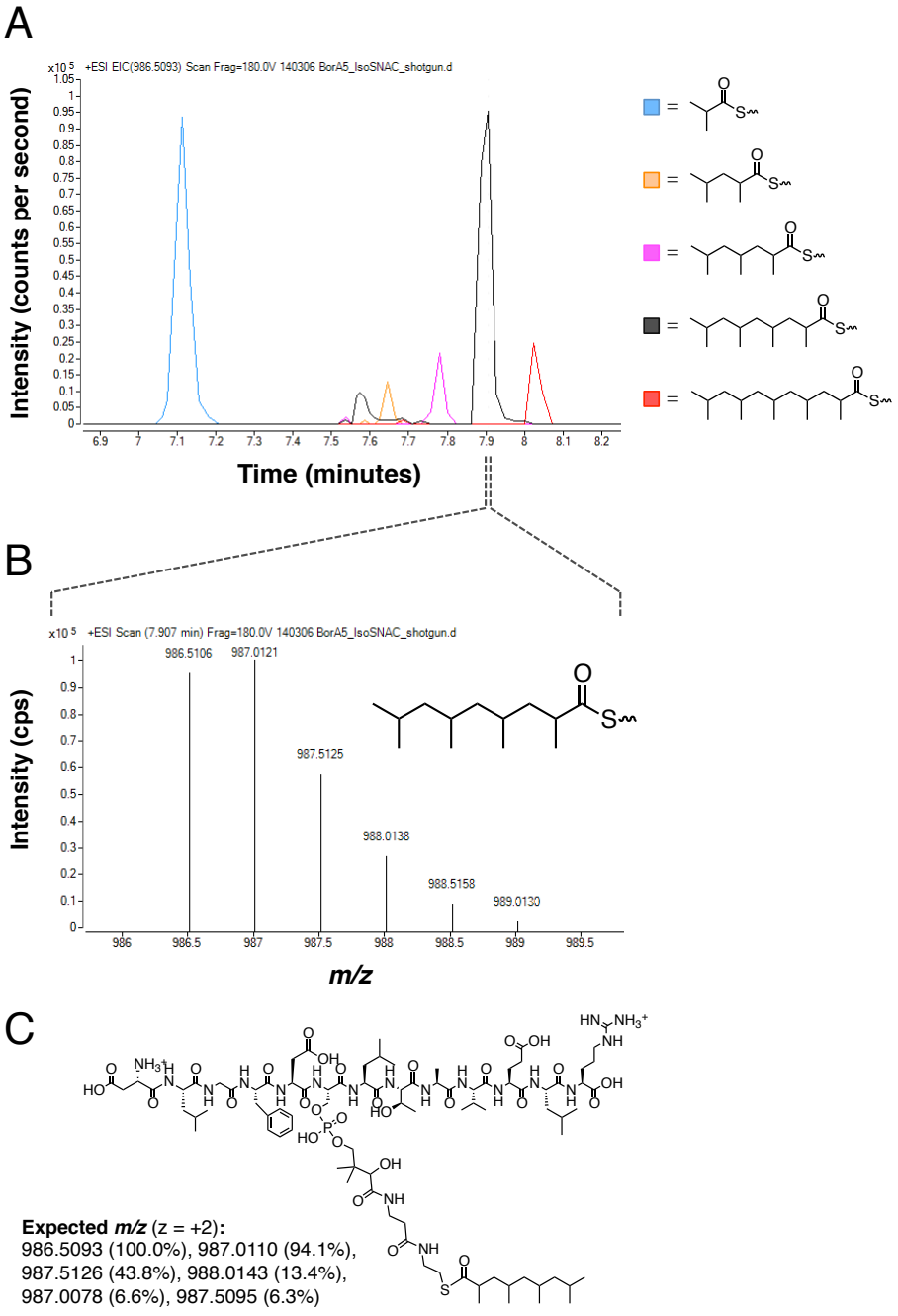



**Figure 4.3**. LC-TOF analysis of extension intermediates. (A) Combined extracted ion chromatograms of indicated ions. (B) Mass spectrum of indicated peak region. (C) Molecular structure and isotopic distribution of analyte shown in B.

The results in **Figure 4.3** clearly demonstrate BorMod5 is capable of iteratively extending the isobutyryl-SNAC substrate and a distribution of extension products (as well as unextended isobutyryl-ACP) is observed. Interestingly, the largest peak corresponds to the product of three extension cycles, which is the reported natural activity of BorMod5, however a minor peak belonging to product of four extension cycles is also observed demonstrating that BorMod5 can "overshoot" the number of iterations it naturally performs. Masses corresponding to partially reduced intermediates were never unambiguously detected, suggesting that beta-processing occurs rapidly relative to the putative ACP to KS "back-transfer" step presumably required to initiate the next round of condensation (Kapur et al., 2012).

### 4.3.3. BorMod5 accepts a variety of SNAC substrates and the extension distribution profile is dependent on the identity of the SNAC starter

Isobutyryl-SNAC, like BorMod5's normal upstream intermediate, has an α-methyl group and can therefore be neatly superimposed onto the upstream intermediate. Having demonstrated that BorMod5 is capable of accepting and iteratively extending such a "truncated" natural substrate, we were curious to see if other substrates of varying lengths and substituents (e.g. a hydroxyl) could also be accepted and what, if any, is the impact on the profile of extension products generated. Three additional SNAC substrates, acetyl-, propionyl- and a partially reduced diketide-SNAC were tested in an extension assay performed as above and the extension product profile is shown in **Figure 4.4**.

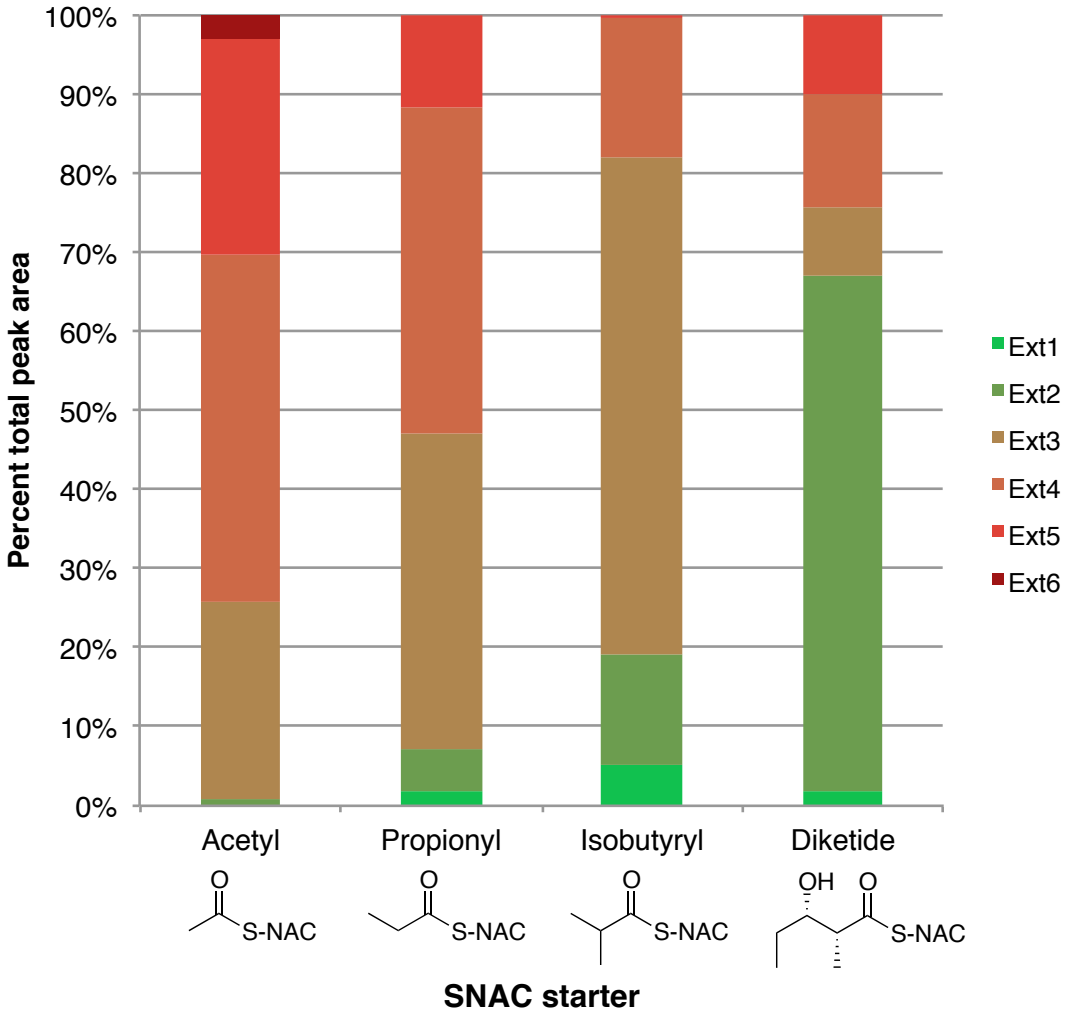



**Figure 4.4**. Extension intermediate distribution plotted as percent total peak area for a variety of SNAC starters. "Ext(n)" denotes product of n extensions.

The data show that the identity of the SNAC starter substrate has a profound effect on the distribution of extension products. As before, isobutyryl-SNAC's most abundant intermediate is that of the three extensions, however propionyl-SNAC which differs only in the absence of an alpha-methyl group, has an approximately equal amount of three and four extension products (40% and 41%, respectively) as well as a significant amount of the five extension product. Furthermore, comparison of the linear two-carbon acetyl-SNAC to linear three-carbon propionyl-SNAC shows a bias towards products of higher numbers of extensions with the major species being that of four extensions and an appreciable amount of five and even six extensions. Taken together, the product distribution of these three SNACs would suggest the general trends that length of carbon backbone and number of methyl branches inversely correlate with number of extension cycles, however the diketide-SNAC behaves differently. This SNAC's major product results from only two extensions, yet despite having a long, α-methyl substituted

backbone, it has an appreciable amount of four and five extension intermediates as well. This distribution is especially curious given the similarity of the diketide-SNAC to the natural upstream intermediate in terms of identity and stereochemistry of the substituents at the alpha and beta positions (compare with intermediate on ACP4 in **Figure 4.1**).

### 4.3.4. Time course of extension product formation reveals intermediate length-kinetics dependency

The experiments described in previous sections, demonstrate the interesting and profound effects the identity of the SNAC starter substrate can have on the distribution of extension products. We were curious to see how the profile of products changes with respect to time, and therefore performed a time course experiment wherein a master reaction mixture was set up using isobutyryl-SNAC as a starter and aliquots were removed and quenched at regular intervals prior to analysis. Results of this experiment are shown in **Figure 4.5**.

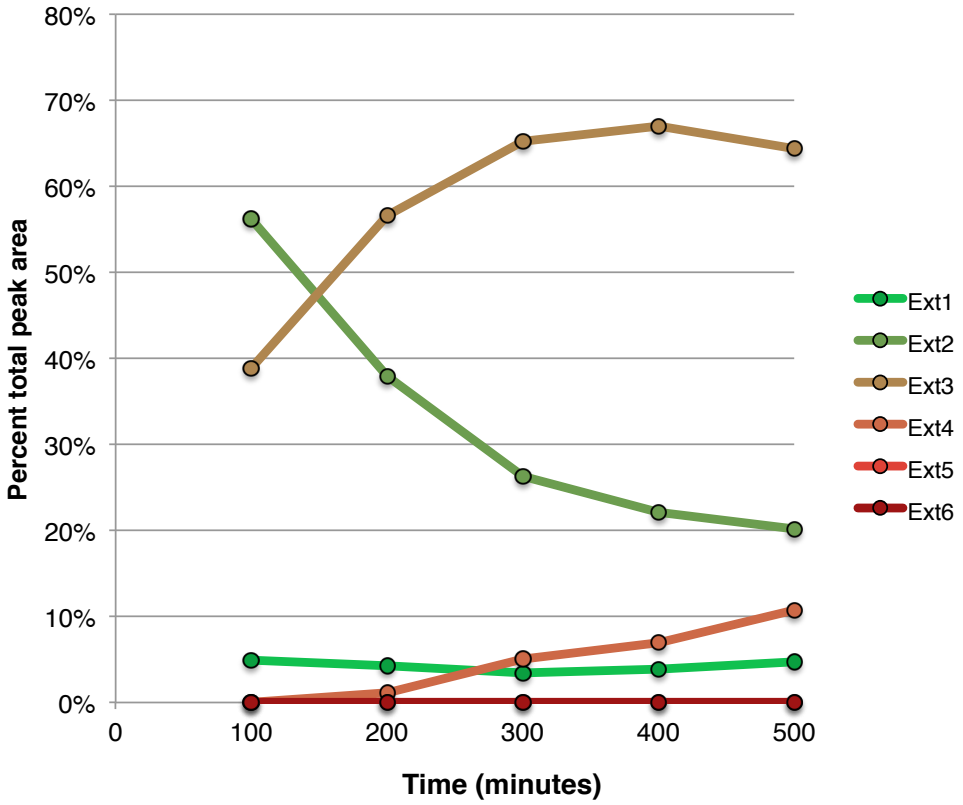


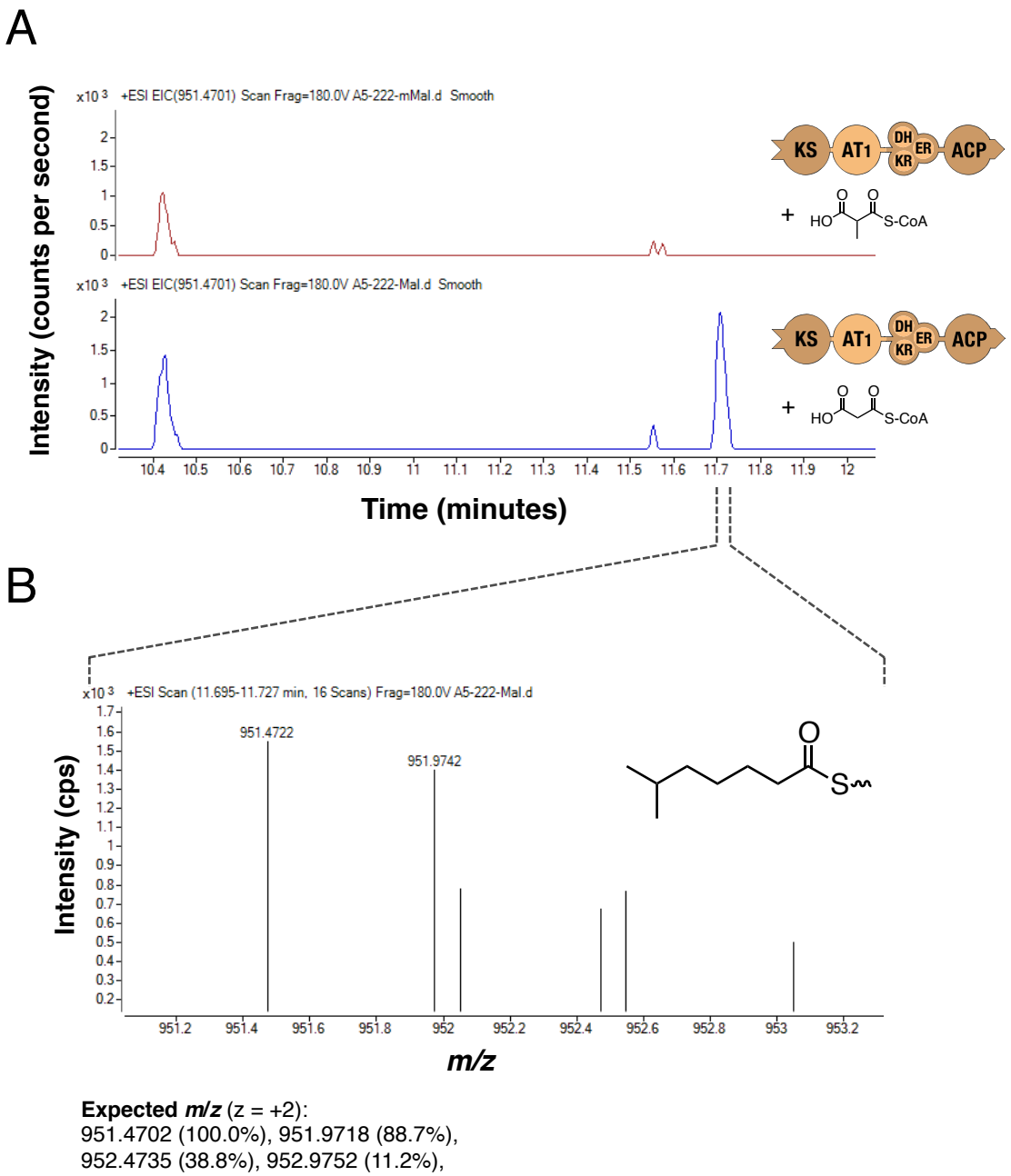

**Figure 4.5**. Time course analysis of extension intermediate distribution using the isobutyryl-SNAC starter plotted as percent total peak area. "Ext(n)" denotes product of n extensions.

As expected, the data reveal that the distribution of extension products is dynamic with respect to time. At the earliest measured time point of 100 minutes, Ext2 predominates but by 200 minutes, enough of this intermediate has been extended that now Ext3 dominates. The Ext1 intermediate stays low and relatively constant as a percentage of the total intermediate profile, which may indicate that it is a kinetically preferred substrate that is rapidly extended. This is in contrast to Ext3 whose accumulation begins

leveling off between 300 and 400 minutes and begins to decline only at later time points as the Ext4 intermediate slowly increases. Taken together, these data indicate BorMod5 more rapidly extends smaller substrates and provide some initial insight into a possible mechanism by which this module regulates the number of iterations it performs.

### 4.3.5. A BorMod5 acyltransferase swap iteratively extends with malonate and reveals AT substrate promiscuity

The observations of a seeming kinetic preference for less extended substrates could indicate a length or number of methyl-branches dependencies on the iterative activity of BorMod5. In order to gauge this module's amenability for engineering as well as gain insight in the role of methyl branches in intermediate processing, we performed a domain swap wherein the wild type methylmalonate-incorporating acyl transferase ("BorAT5") domain was replaced by the malonate-incorporating AT domain from borrelidin's first extension module ("BorAT1"). Chimeric junctions were selected using a recently developed method which has been shown to have minimal impact on catalytic efficiency (S. Yuzawa, manuscript in preparation). To test activity of this new chimeric BorMod5 construct, "BorMod5-AT1," isobutyryl-SNAC was used as a starter and in separate reactions, either methylmalonyl-CoA or malonyl-CoA were provided as extender units. As shown in **Figure 4.6**, the molecular ion corresponding to two iterations with malonate extensions is observed (expected m/z 951.4702, observed m/z 951.4722 for most abundant isotope), proving that the AT-swapped mutant maintains the ability to iterate and is capable of extension using this desmethyl extender unit. This ion was not observed in the reaction in which methylmalonyl-CoA was provided as an extender and the retention time is in the expected range, ruling out the possibility that this is a background peptide.




**Expected *m/z*** (z = +2):
951.4702 (100.0%), 951.9718 (88.7%),
952.4735 (38.8%), 952.9752 (11.2%),

**Figure 4.6**. LC-TOF analysis comparing products of methylmalonyl-CoA vs. malonyl-CoA extender reactions using AT-swapped BorMod5-AT1 mutant. (A) Extracted ion chromatogram of twice extended (malonate) product. (B) Mass spectrum of indicated peak region and expected isotopic distribution.

Unexpectedly, the AT-swapped mutant is still capable of iterative extension using methylmalonyl-CoA as shown in **Figure 4.7**.

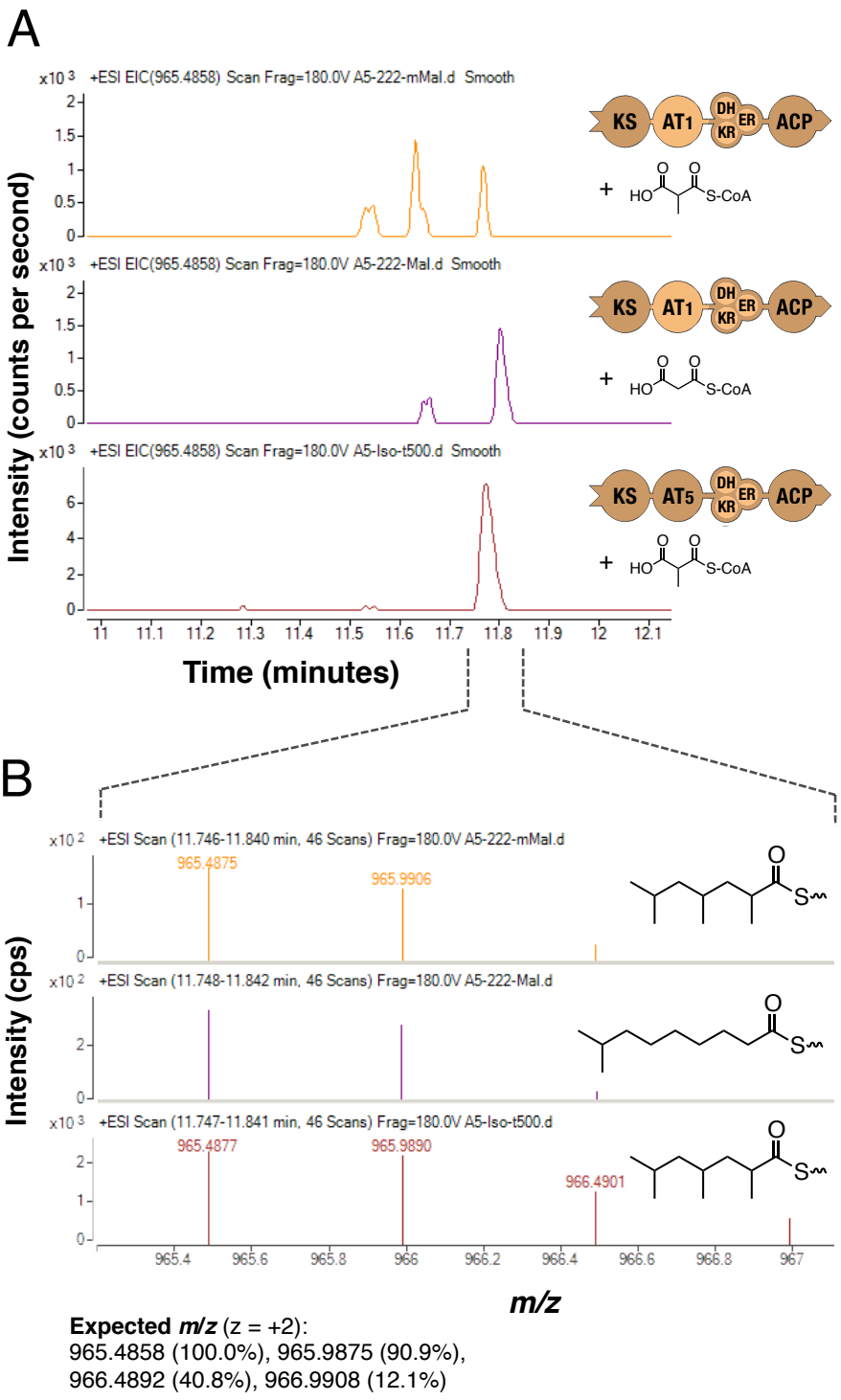



**Figure 4.7**. LC-TOF analysis comparing products of methylmalonyl-CoA vs. malonyl-CoA extender reactions using AT-swapped BorMod5-AT1 mutant and including wt BorMod5 (protein variants and extenders indicated to right of upper panels). (A) Extracted ion chromatogram of indicated ion. (B) Mass spectrum of indicated peak region and expected isotopic distribution for the different products.

**Figure 4.7** shows that the isomeric Ext2 (methylmalonate) and Ext3 (malonate) are both observed in their respective reactions. The peak for this molecular ion (*m*/*z* 965.4858), found at the same retention time as in the methylmalonate extensions with wt BorMod5 (lower panel), confirms its identity and it is interesting to observe a slight retention time shift in the Ext3 (malonate) isomer. This result suggests that AT1 from BorMod1 is promiscuous and can load both extender units at least at the concentrations used in this *in vitro* study (500 uM). Such promiscuity in selection of extender units has been documented before as in AT3 from the epothilone PKS (Gerth et al., 2000) but has not previously been reported for any ATs from the borrelidin cluster.

To test whether AT1 from BorMod1 is inherently promiscuous or whether the observed promiscuity is a result of changing the AT domain's modular context, we tested wild type BorMod1's ability to extend with methylmalonate in an extension assay using succinyl-SNAC as a starter substrate (see Chapter 3 for details). Results are shown in **Figure 4.8**.

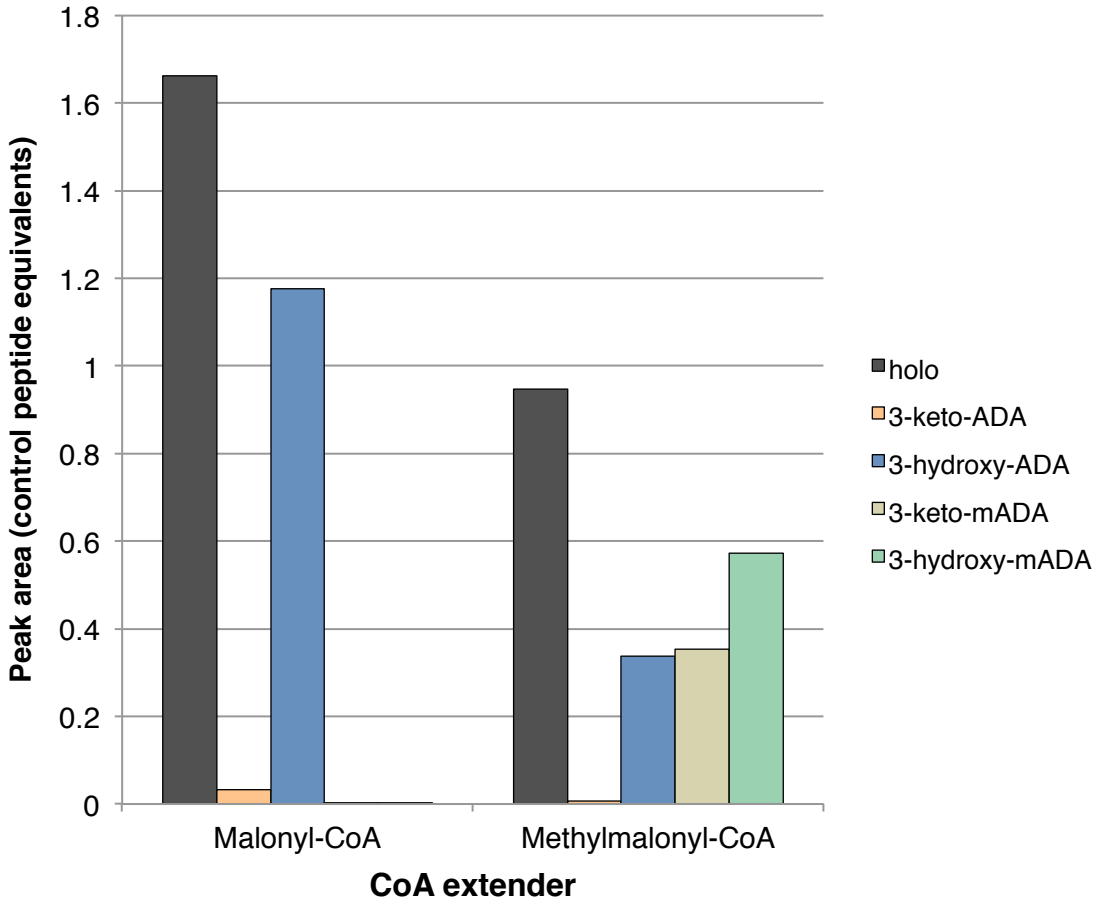



**Figure 4.8**. Extension intermediate analysis using BorMod1 and different CoA extender units. "ADA" and "mADA" refer to "adipic acid" and "2-methyladipic acid," respectively

Analysis of extension intermediates on BorMod1 shows the formation of 3-keto-2-methyl adipic acid and the reduced 3-hydroxy-2-methyl adipic acid ("3-keto-mADA" and "3-hydroxy-mADA," respectively) indicating that the acyltransferase does indeed load methylmalonyl-CoA in its native context and condensation can proceed with this intermediate. The substantial amount of the 3-keto-mADA intermediate remaining indicates the ketoreductase does not efficiently process this intermediate—likely due to the presence of the α-methyl group. Interestingly, an appreciable amount of the desmethyl 3-hydroxy-ADA intermediate was also detected, possibly indicating that the methylmalonyl-CoA provided is contaminated with malonyl-CoA or perhaps that the protein was purified in a pre-loaded malonyl-ACP state as has been previously observed (S. Poust, personal correspondence; (Belecki and Townsend, 2013)).

### 4.3.6. Co-incubation with BorMod6ΔTE indicates a possible gatekeeping mechanism for ensuring product fidelity

The observations in sections 4.3.2 and 4.3.3 that BorMod5 produces a distribution of extension products and can "overshoot" its usual three iterations (at least for the substrates tested) suggest that the downstream acceptor, BorMod6, likely plays a role in ensuring product fidelity. Previous studies have shown that modules acting late in PKS biosynthesis tend to be more tolerant of atypical upstream intermediates (Wu et al., 2000), however we hypothesized that BorMod6 may be acting as a "gatekeeper" and only accepts the pendant borrelidin intermediate after BorMod5 has performed three iterations. In order to test this idea, BorMod6 was heterologously expressed and purified from *E. coli* as a C-terminal truncation lacking the thioesterase domain ("BorMod6ΔTE") so that acyl-ACP intermediate analysis could be used to see what intermediates are found on this module in absence of the normal cyclase/hydrolase activity of the TE domain. SDS-PAGE analysis revealed the protein expressed abundantly and solubly as an N-terminal MBP fusion and was purified to near homogeneity after removal of MBP (see materials and methods for details). An experiment was performed wherein BorMod5 was pre-incubated with isobutyryl-SNAC, methylmalonyl-CoA extender and NADPH for variable amounts of time and then added to a mixture containing BorMod6ΔTE, malonyl-CoA (its natural extender) and additional NADPH (see materials and methods for protein expression/purification and details of experimental setup). The objective was to accumulate different distributions of extension products on BorMod5 prior to exposure and further incubation with BorMod6ΔTE and then ascertain whether some of these intermediates are preferentially accepted and extended by BorMod6ΔTE —thereby providing some indication if the hypothesized gatekeeping is taking place. Results are shown in **Figure 4.9**.

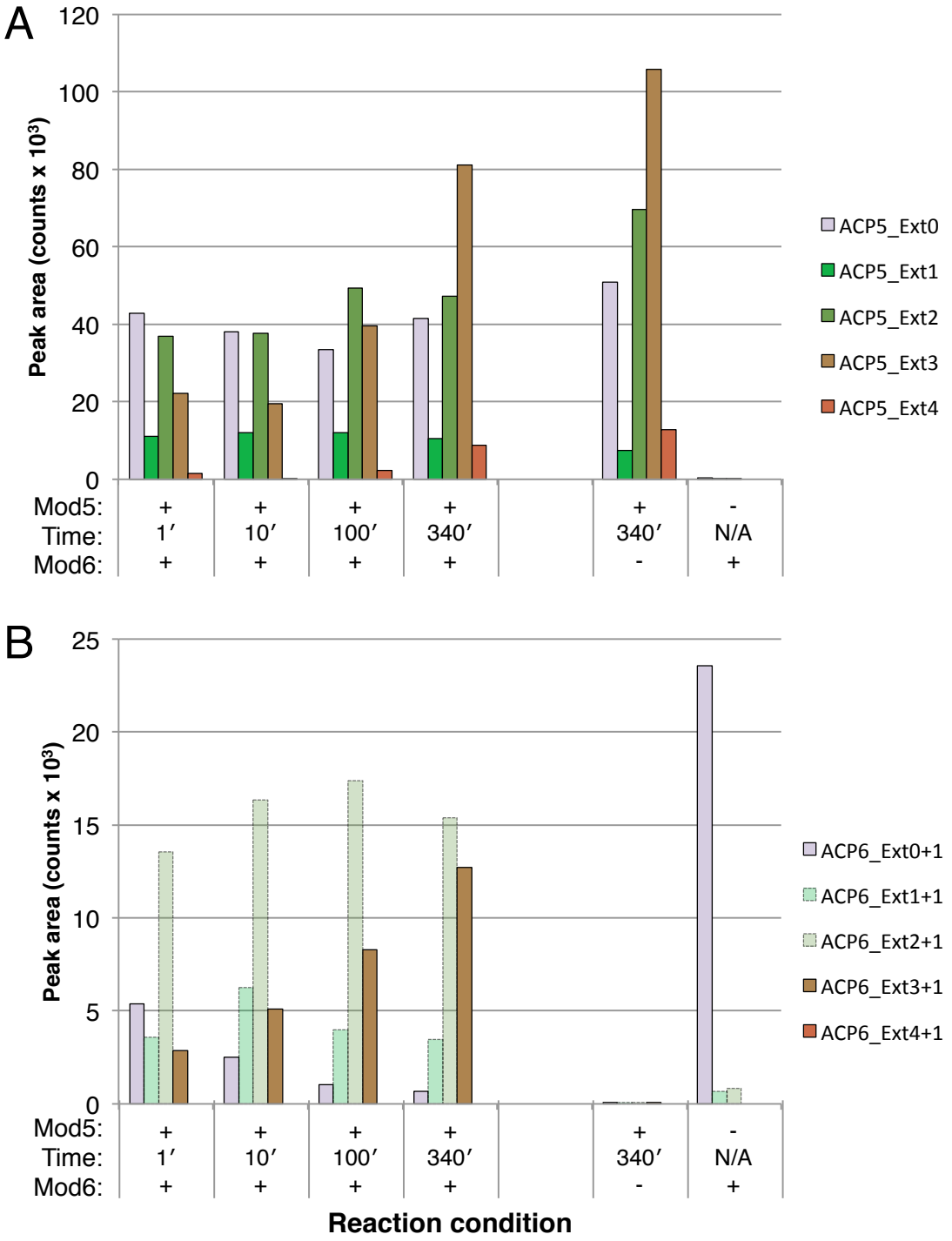



**Figure 4.9**. Extension intermediate analysis of (A) BorMod5 and (B) BorMod6ΔTE intermediates. Reaction conditions are indicated by presence or absence of respective modules and BorMod5 "pre-incubation" time (minutes). Identity of intermediates in (B) follows a convention where "Ext(n)" indicates number of extensions catalyzed by BorMod5 and "+1" indicates the product of the extension of that upstream intermediate by BorMod6ΔTE (see materials and methods for details). Translucent bars with dashed outline indicate uncertainty in data due to contaminating peaks.

As shown in (A) above, the accumulation and distribution of extension intermediates does not appear to be significantly affected by co-incubation with BorMod6ΔTE, at least at the 340 minute time point (compare "+,340,+" to "+, 340,-" and corresponding time points from previous experiments). This observation would seem to argue against BorMod6ΔTE being only able to accept/extend the Ext3 product of BorMod5 which would otherwise be depleted in the "+,340,+" reaction. However in (B), a time-dependent formation of the ACP6_Ext3+1 intermediate is observed and no accumulation of the ACP6_Ext4+1 intermediate is seen despite detectable formation of ACP5_Ext4. Unfortunately, it was not possible to accurately quantify the accumulation of putative ACP6_Ext1+1 and ACP6_Ext2+1 intermediates due to the presence of abundant contaminating peaks (indicated with translucent bars that are outlined in dashed lines in (B)), though careful analyses of the mass spectra do indicate they are present at some level (data not shown). The presence of the ACP6_Ext0+1 intermediate in the reaction "-,N/A,+" wherein BorMod6ΔTE is incubated with the SNAC starter, CoAs and NADPH but without BorMod5 indicates that isobutyryl-SNAC can be loaded onto BorMod6ΔTE and subsequently extended. Furthermore, the ubiquitous and abundant presence of ACP5_Ext0 and ACP6_Ext0+0 (not shown) indicate that isobutyryl-SNAC can non-enzymatically transacylate onto the 4'Ppant prosthetic group of the ACPs, further complicating analysis by attenuating enzyme activity.

Despite the technical limitations in this experimental setup, these data suggest that BorMod6 readily accepts the Ext3 intermediate and not the "overshot" Ext4 intermediate, though it remains inconclusive whether BorMod6 gatekeeps against less extended intermediates. Detection of ACP6_Ext0+1 at the very least indicates that once loaded onto the BorMod6 ketosynthase, a short intermediate is capable of being extended, even if it remains unclear whether such a short intermediate is ever normally transferred to this module. As will be subsequently discussed, method refinement and a more sophisticated experimental setup should be able to unambiguously resolve these uncertainties.

## 4.4. Discussion

The results presented in this chapter represent a preliminary, but to our knowledge, first *in vitro* characterization of an iterative module from an otherwise modular type I PKS system. Initial experiments examining expression of the module revealed the profound impact the expression system has on solubility of the protein. When expressed from the T7 promoter in the medium-high copy pET28a plasmid, the protein is highly expressed but completely insoluble. In contrast, when expressed as an N-terminal maltose-binding protein fusion from a low-copy plasmid using the pTet promoter (and native *E. coli* RNA polymerase rather than T7 polymerase), solubility is promoted to above 50%. It is impossible to determine what the individual or collective contribution to solubility these differences make by comparing just two plasmids expressing a single construct, but nonetheless one wonders if previous *in vitro* investigations were obstructed by lack of protein solubility if modules were overexpressed from the ubiquitous pET system (Menzella et al., 2006; Murli et al., 2003; Watanabe and Oikawa, 2007).

In section 4.3.2 we described how the simple provision of a starter SNAC substrate, the natural extender unit (methylmalonyl-CoA) and reducing power in the form of NADPH is sufficient for iterative extension to occur. These results demonstrate that BorMod5 is in and of itself necessary and sufficient for iteration to occur and does not depend on other factors (e.g. an intermediary acyl carrier protein) to mediate the repeated extension steps in full support of previous *in vivo* studies. It is as yet unclear the exact mechanism by which an extended intermediate on the acyl carrier protein is transferred to the ketosynthase in order to allow acyltransferase-mediated loading of the next extender unit and subsequent extension, nor is it clear whether the transfer occurs to the KS *in cis* or the *trans*-KS belonging to the presumed homo-dimeric partner module though evidence from previous studies is highly suggestive of *trans*-KS ("interchenar") transfer, again, in the aureothin PKS system (Busch et al., 2012). The possibility that BorMod5 forms higher order structures (i.e. head-to-tail associations of multiple homodimers) and intermediate extension proceeds vectorially would seem unlikely in light of previous *in vivo* studies involving protein fusions (Olano et al., 2003), however analytical gel filtration could easily be used to determine the oligomeric state of the enzyme complex. The observation of an intermediate corresponding to four extensions shows that this module can also "overshoot" the normal three iterations that occur during borrelidin biosynthesis and therefore does not *sensu stricto* count the number of iterations it performs. Similar *in vivo* observations were made when the iterative module AurA from aureothin PKS was expressed in isolation (Busch et al., 2013). More likely is a dependency on some feature of the intermediate itself, such as chain-length found in iterative type II PKSs (mediated by a dedicated chain-length factor domain (Tang et al., 2003)) or possibly some other structural feature like number of methyl-branches.

In the subsequent experiment (section 4.3.3) wherein different acyl-SNAC starter units were provided to the isolated module, it was shown that BorMod5 can accept and iteratively extend a variety of SNACs and interestingly, the distribution of extension products is affected by the identity of the starter. The general trend is that shorter starters are extended more times than longer and branched starters, and it is interesting to note

that the predominant extension intermediate observed for each SNAC starter tends to have 9-10 carbons in the linear backbone and 3-4 methyl branches. The time course experiment described in section 4.3.4 provides additional support to the chain-length or number of methyl branches mechanism for regulation of iteration. It is shown that the smaller or less extended the intermediate is, the more rapidly it is consumed and converted into longer intermediates with the rate of extension apparently slowing dramatically upon reaching three extensions. While this seems like a plausible explanation for chain-length regulation, it is difficult to reconcile with the fact that the normal upstream intermediate presented to Mod5 by ACP4 is a pentaketide (see **Figure 4.1**) and therefore already has a longer chain length than any observed SNAC extension intermediate. Nevertheless, given that the pentaketide intermediate possesses no methyl-branches and contains a double bond at the 2,3 position which may serve as a structural "handle," it could be possible that "counting" by chain-length or number of methyl branches in either case begins at zero for the natural substrate.

Swapping the methylmalonyl-specific acyltransferase domain from BorMod5 with malonyl-specific AT1 from borrelidin PKS' first extension module still permits iterative extension. This proves that the iteration demonstrated by BorMod5 is in no way dependent on the identity of the AT domain nor which extender unit is used. Only the twice and thrice extended intermediates are observed (at an approximately 3:2 ratio), however given the overall reduced activity of the chimeric enzyme, it is difficult to draw any conclusions about iteration regulation with this system. Interestingly, the reductive loop seems to efficiently reduce the intermediates despite the lack of methyl branches at the alpha position as no partially reduced intermediates were observed. This could indicate that back transfer is slow relative to the rate of β-processing even at the presumed reduced rates for the desmethyl-extended unnatural substrates. Additionally, there may be a mechanism by which back transfer is precluded until complete reduction to the β-methylene has occurred as has been previously suggested (Hong et al., 2009; Whicher et al., 2014). Experiments in which NADPH is omitted or present at reduced concentrations, or a more rapid quench method would likely shed light on these unknowns. An unexpected observation was that the AT-swapped mutant is still able to load and extend with methylmalonyl-CoA—indicating AT promiscuity, which was subsequently confirmed when tested in its native modular context. This is curious given AT1 has the canonical motif indicating malonyl-CoA specificity as opposed to a hybrid malonyl/methylmalonyl motif observed in known promiscuous AT domains such as epothilone's AT3 (Gerth et al., 2000). Whether this is relevant in the context of *in vivo* borrelidin biosynthesis is unclear given the extended incubation times and high CoA concentrations used in these *in vitro* experiments. Using CoAs at concentrations in a physiological range and competition experiments as well as more rigorously controlled biochemical studies would be more definitive.

In the final experiment presented in this chapter the downstream module, BorMod6, is used to gain insight into chain-length regulation in the borrelidin PKS system by a hypothetical gatekeeping mechanism. Despite some shortcomings in the experimental setup and analytical method, some tentative conclusions can be made. The data suggests that BorMod6 may, to some degree accept BorMod5's Ext1 and Ext2 products, however

it would seem to prefer the Ext3 product. This is supported by the time-dependent increase in the total amount of the Ext3+1 intermediate found on BorMod6ΔTE which also increases as a percent of the total intermediates found on BorMod6ΔTE. If it were the case that the lesser-extended intermediates were transferred to BorMod6ΔTE at the same rate as Ext3 intermediates, the expectation would be that the Ext1+1 and Ext2+1 intermediates would always make up a higher proportion of total intermediates since Ext1 and Ext2 necessarily form prior to Ext3. Furthermore, the observation that no Ext4+1 intermediates are found would suggest that BorMod6ΔTE actively gatekeeps against this intermediate. While this would ensure no decaketide borrelidin analog is ever formed, it would effectively prevent further synthesis of borrelidin as the assembly line would be stalled with a dead-end intermediate at BorMod5. It is difficult to rationalize the adaptive benefit of such strict gatekeeping absent a mechanism by which improperly extended intermediates are released from the PKS enzyme. This may be accomplished by the type II thioesterase present in the borrelidin cluster whose function is as yet undetermined (Olano et al., 2004a).

Taken together, the data obtained in these studies suggest a model for chain-length fidelity in borrelidin PKS consistent with the recent cryoEM studies of Whicher and colleagues. During the normal course of borrelidin biosynthesis, BorMod5's ACP can have a variety of extended intermediates attached, dependent on the number of iterations that have taken place. The specific identity of such intermediates biases ACP5 either towards remaining in the module's "reaction chamber," thereby promoting back transfer to KS5, or towards the "flipped-out" conformation that would allow interaction with KS6 and permit chain translocation. This may explain why longer intermediates are apparently extended at a reduced rate relative to short intermediates—the ACP spends more time outside the reaction chamber primed for forward transfer, making it less likely to back transfer from within the chamber. An extra layer of regulation could be achieved by BorMod6, which may be less likely to accept a short intermediate presented to it by an ACP that has prematurely exited the reaction chamber, than an intermediate that has been extended the appropriate number of times. The specific epitopes that dictate the ACP localization remain unclear, though chain-length and/or number of methyl groups appear likely candidates.

Future characterization studies using more refined experimental setups and analytical methods could address many of these remaining questions. Non-enzymatic transacylation of SNAC substrates onto ACP5 and ACP6 could be circumvented by initiating biosynthesis using acyl-ACP4—readily created by purifying the monodomain in its apo form and using Sfp and acyl-CoAs to load it (see Chapter 2 for examples). Similarly, BorMod5 could be purified in its apo form and acylated with substrates of varying lengths and methyl branches to more precisely see which get transferred to BorMod6ΔTE. Using a null ketosynthase mutant (e.g. lacking the acyl-accepting cysteine residue) or omission of extender unit in this system would preclude back transfer and extension of loaded substrates and therefore allow even tighter control over which acyl-ACP5 substrates are presented to BorMod6. Finally, the full-length version of BorMod6 harboring its natural TE would allow the system to be multiple-turnover and therefore may provide a more realistic picture of enzyme activity and specificity

compared to the single-turnover conditions required for intermediate analysis using the described proteomics methods. The presence of contaminating ions complicates analysis of certain intermediates and therefore refined liquid chromatography methods with longer, slower gradients would help deconvolute these peaks. Injection of higher amounts of more homogenously purified proteins would likely raise the signal to noise ratio for analytes and therefore also reduce ambiguity in the data.

It is interesting to note that every iterative module identified to date is a standalone module. The implications from a mechanistic or evolutionary perspective are unclear, however we suggest this makes such modules excellent candidates for cryoEM studies such as the ones described, and would doubtless reveal a more complete picture of the mechanism and capabilities of type I "modular" systems.

## 4.5. Materials and Methods

### 4.5.1. Reagents and Chemicals

HisPur cobalt resin was purchased from Thermo Scientific, Bradford reagent was from Bio-Rad and SDS-PAGE gels from Life Technologies. Solvents (hexanes, ethyl acetate, dichloromethane and methanol) were purchased from EDH; all other reagents were purchased from Sigma-Aldrich.

### 4.5.2. Chemical synthesis and NMR data

SNAC substrates were generously provided by S. Yuzawa and were synthesized after a similar procedure described in (Zhou et al., 2008). Briefly, carboxylic acids were incubated with N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC-HCl), 4-dimethylaminopyridine (DMAP) and *N*-acetylcysteamine.

### 4.5.3. Plasmid list and construction

**pARH130**: The protein sequence encoding BorA5 was codon-optimized for *E. coli* and synthesized by Genewiz into a pUC57 holding vector. Flanking NdeI/XhoI restriction sites were used to liberate the open reading frame from the plasmid and was subcloned into a similarly digested pARH090 (see Chapter 2).

**pARH130b**: The insert used in the creation of pARH130 (above) was ligated into a similarly digested pET28a vector (Novagen).

**pARH135:** The pUC57::BorA6 plasmid from Genewiz was digested with NdeI and BamHI (silently encoded into the "GS" residues at amino acid positions 1450-1451, BorA6 numbering) to liberate a C-terminal truncation insert lacking the thioesterase domain. This insert was ligated into a similarly digested pARH090 (see Chapter 2).

**pARH170:** Nucleotides 3467-4768 of pARH130 were replaced by nucleotides 3395-4681 of pARH100 (see 2.5.3). The j5 algorithm and Device Editor graphical user interface were used to design DNA assemblies (Hillson et al., 2012). All DNA pieces were amplified via PCR with either Q5 or Phusion polymerases (New England BioLabs)

according to manufacturer's recommendations. Gel-extracted DNA was assembled via Gibson cloning using Gibson Assembly® master mix (New England BioLabs).

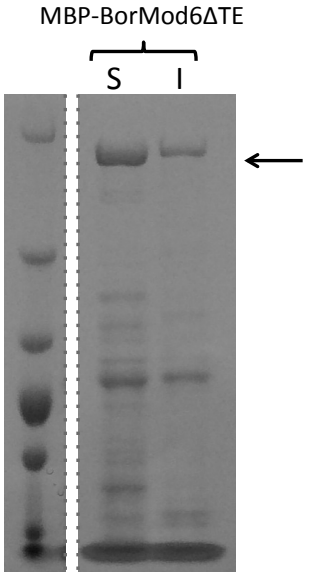
### 4.5.4. Protein expression tests and purification

#### 4.5.4.1. SDS-PAGE analysis

Solubility analysis via SDS-PAGE was performed as described in section 3.5.4.1 using the respective plasmids pARH130 and pARH130b

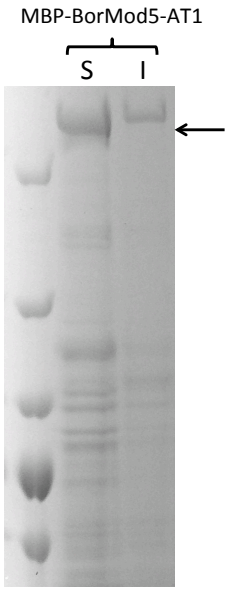
#### 4.5.4.2. Purification of PKS constructs

**MBP-BorMod5** was expressed and cell pellets processed in a manner similar to that described for MBP-BorA2 (see 2.5.4) using plasmid pARH130. Clarified lysate was applied to a 5 ml HisTrap FF column (GE lifesciences) and eluted with a gradient of up to 500 mM imidazole. Eluate was concentrated via spin filtration (Amicon, 100 kDa MWCO) and exchanged into Buffer A (50 mM phosphate, 0 mM NaCl, 8% glycerol) using a PD-10 column (GE lifesciences). This was then applied to a 1 ml HiTrap Q HP anion exchange column and eluted with a gradual gradient of increasing NaCl (Buffer B = 50 mM phosphate, 500 mM NaCl, 8% glycerol). Eluate was concentrated again via spin filtration and snap frozen. **MBP-BorMod6ΔTE** was expressed, purified and the MBP domain subsequently removed as described for MBP-BorA2 (see 2.5.4) using plasmid pARH135.




**Supplementary Figure 4.1**. SDS-PAGE analysis of Mod6ΔTE solubility. Shown is soluble (S) and insoluble (I) protein fractions. Black arrow indicates protein of interest

**MBP-BorMod5-AT1** was purified in a manner as described in 3.5.4.2 from a 200 ml culture using plasmid pARH170.




**Supplementary Figure 4.2**. SDS-PAGE analysis of MBP-BorMod5-AT1 solubility. Shown is soluble (S) and insoluble (I) protein fractions. Black arrow indicates protein of interest

### 4.5.5. Experimental setup and sample processing

#### 4.5.5.1. Initial isobutyryl-SNAC extension with MBP-BorMod5

A 30 ul reaction mixture (final concentrations: 5 mM isobutyryl-SNAC, 0.5 mM methylmalonyl-CoA, 1 mM NADPH, 2.5 mM TCEP, 1 uM MBP-BorMod5 in 100 mM phosphate buffer pH 7.4) was incubated at room temperature overnight (~16 hr).

#### 4.5.5.2. Extension of other acyl-SNAC substrates with MBP-BorMod5

40 ul reactions were set up with the same final concentrations used in 4.5.5.1 but with different SNAC substrates and was incubated at room temperature overnight (~16 hr).

#### 4.5.5.3. Time course of isobutyryl-SNAC extension

A 250 ul reaction mixture was created with the same final concentrations used in 4.5.5.1 and incubated at room temperature. At indicated time points, 40 ul aliquots were removed from the master mix, transferred to thin-walled PCR tubes and snap-frozen in a dry ice/ethanol bath.

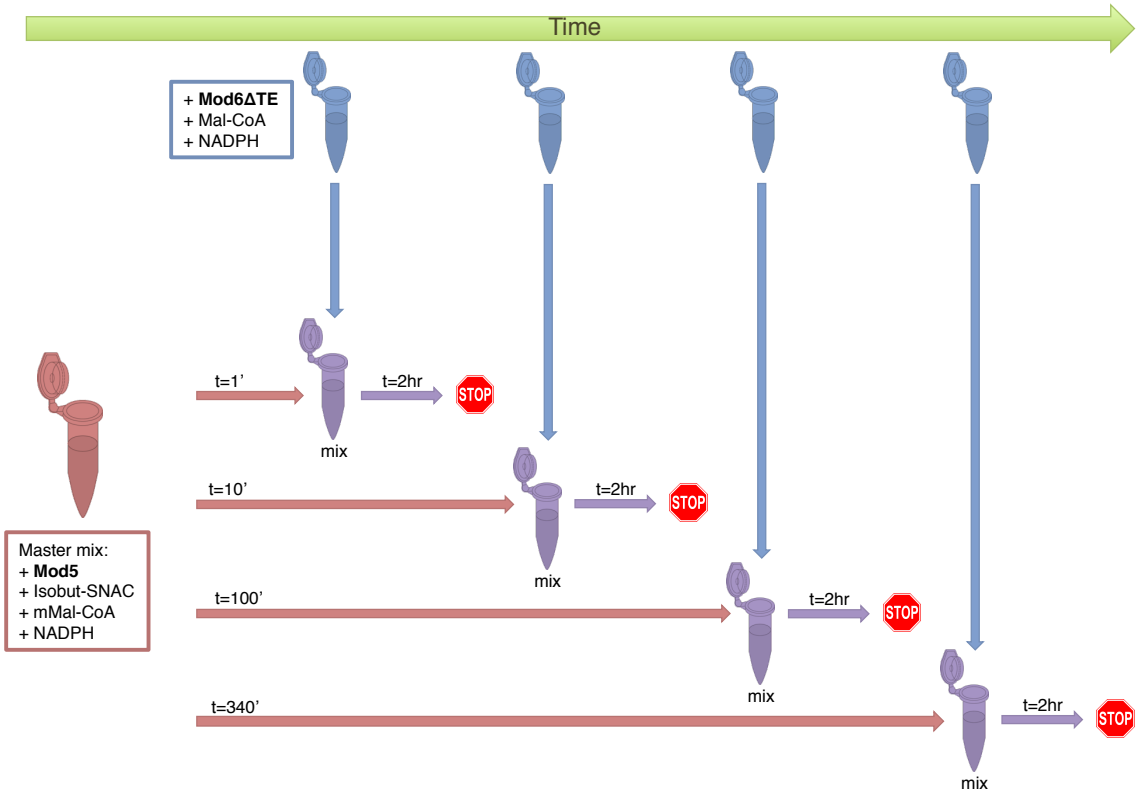
#### 4.5.5.4. Extension with AT-swapped MBP-BorMod5-AT1

50 ul reactions were created with the same final concentrations used in 4.5.5.1 and either malonyl-CoA or methylmalonyl-CoA was provided as an extender unit. Reactions were incubated at room temperature overnight (~16 hr).

#### 4.5.5.5. Co-incubation with BorMod6ΔTE

A 100 ul master mix created with the same final concentrations used in 4.5.5.1 was prepared to create the "M5" master mix. At indicated time points, 20 ul of this mixture was combined with 20 ul of freshly prepared "M6" which contained the same final concentrations of all components but the BorMod6ΔTE protein and malonyl-CoA as PKS

protein and extension unit, respectively. After a further 2 hr incubation, samples were snap-frozen in a dry ice ethanol bath. A schematic of this experimental setup follows:




**Supplementary Figure 4.3**. Schematic of reaction set up described in section 4.3.6.

### 4.5.6. Sample processing and analytical methods

Samples were desalted via drop dialysis using VSWP02500 membranes (Millipore) according to manufacturer's recommendations against a 50 mM ammonium bicarbonate solution (pH ~8.0) and subsequently trypsinized with 1:20 w/w porcine trypsin (Sigma-Aldrich) for 4-6 hours at 37C. For the time course experiment described in section 4.5.5.3, trypsin was added prior to thawing the samples and subsequent dialysis in order to mitigate resumption of enzyme activity during desalting.

Time of Flight (TOF) MS data was collected on an Agilent 1290 UHPLC system coupled to an Agilent 6550 Q-TOF mass spectrometer operating in MS1 mode. Approximately one microgram (µg) of protein was injected and separated on a Sigma Ascentis Peptide Express C-18 column (2.1 mm × 100 mm, 2.7 um particle size) at a flow rate of 400 µl/min. The run began with 95% Buffer A (2% acetonitrile in water + 0.1% formic acid) and 5% Buffer B (98% acetonitrile in water + 0.1% formic acid) for 5 minutes. Buffer B was increased to 35% over the course of 5.5 minutes and then rapidly increased to 80% over the course of one minute where it was held for 3.5 minutes and the flow rate was increased to 600 µl/min. Buffer B was then rapidly decreased to 5% where it was held

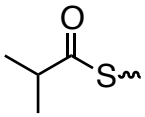
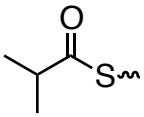
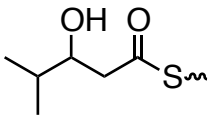
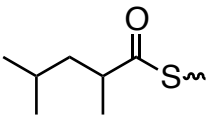
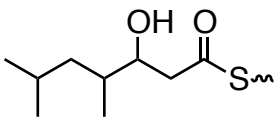
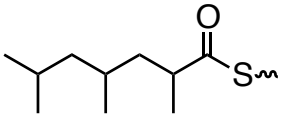
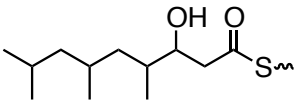
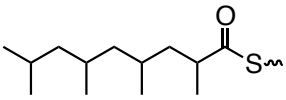
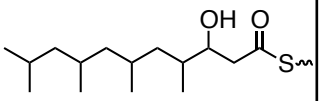
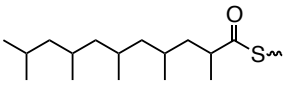
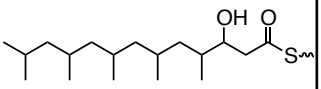
for two minutes and the flow rate was returned to 400 µl/min to re-equilibrate the column. The peptides eluting from the column were ionized by using an Agilent Jet Stream source (sheath gas flow: 11 l/min, sheath gas temperature: 250 C, nozzle voltage: 1,000 V, nebulizing pressure: 35 psi, chamber voltage: 5000 V) operating in positive-ion mode. The data were acquired with MassHunter B.05.00 operating in MS1 mode within 300 m/z to 1700 m/z mass range. *m/z* values and predicted isotopic distributions for molecular ions were calculated using ChemDraw. Peak areas for extracted ion chromatograms of most abundant isotopes were quantified using TOF Quantitative Analysis B.05.00.

ACP5 apo peptide: DLGFDSLTAVELR
Holo *m/z* 888.4179 (z=+2)

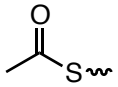
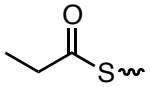
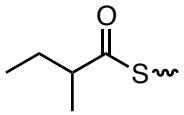
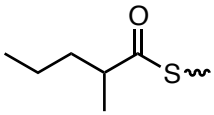
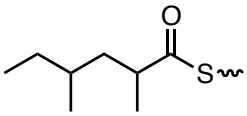
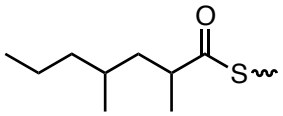
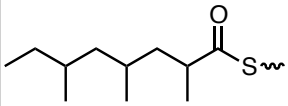
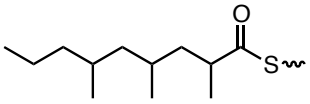
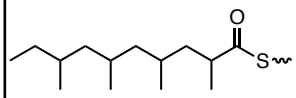
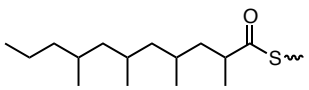
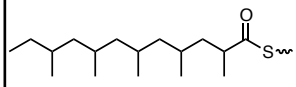
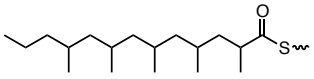
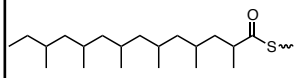
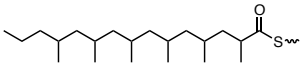
ACP6 apo peptide: IGADHEFLALGFDSLTSIELR
Holo *m/z* 882.4303 (z=+3)
(site of phosphopantetheinylation underlined)

Tables of molecular ions and corresponding m/z values follow:

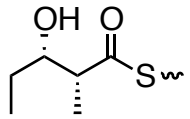
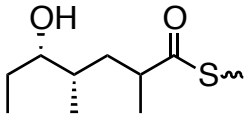
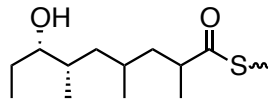
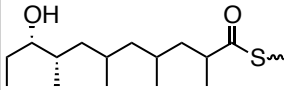
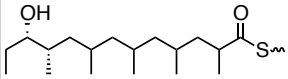
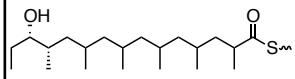
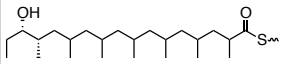
Isobutyryl-SNAC / mMal-CoA extension products Mod5, Mod6

| ID | Structure (ACP5) | m/z z=+2 | ID | Structure (ACP6) | m/z z=+3 |
|---|---|---|---|---|---|
| Ext0 |  | 923.4389 | Ext0+0 |  | 905.7776 |
| | | | Ext0+1 |  | 920.4530 |
| Ext1 |  | 944.4623 | Ext1+1 |  | 934.4686 |
| Ext2 |  | 965.4858 | Ext2+1 |  | 948.4843 |
| Ext3 |  | 986.5093 | Ext3+1 |  | 962.4999 |
| Ext4 |  | 1007.5328 | Ext4+1 |  | 976.5156 |
| | | | | | |

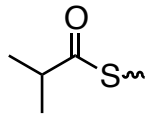
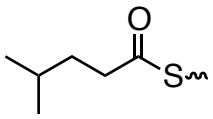
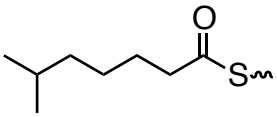
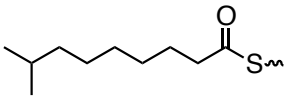
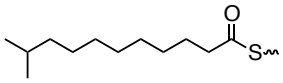
Various-SNACs / mMal-CoA extension products Mod5

| ID | Acetyl-SNAC Structure (ACP5) | m/z z=+2 | ID | Propionyl-SNAC Structure (ACP5) | m/z z=+2 |
|---|---|---|---|---|---|
| Ext0 |  | 909.4232 | Ext0 |  | 916.4310 |
| Ext1 |  | 930.4467 | Ext1 |  | 937.4545 |
| Ext2 |  | 951.4702 | Ext2 |  | 958.4780 |
| Ext3 |  | 972.4936 | Ext3 |  | 979.5015 |
| Ext4 |  | 993.5171 | Ext4 |  | 1000.5249 |
| Ext5 |  | 1014.5406 | Ext5 |  | 1021.5484 |
| Ext6 |  | 1036.0657 | Ext6 |  | 1043.0736 |

Various-SNACs cont. / mMal-CoA extension products Mod5

| ID | Diketide-SNAC Structure (ACP5) | m/z z=+2 | | | |
|---|---|---|---|---|---|
| Ext0 |  | 945.4520 | | | |
| Ext1 |  | 966.4754 | | | |
| Ext2 |  | 987.4989 | | | |
| Ext3 |  | 1008.5224 | | | |
| Ext4 |  | 1029.5459 | | | |
| Ext5 |  | 1051.0710 | | | |
| Ext6 |  | 1072.0945 | | | |

Isobutyryl-SNAC / Mal-CoA extension products Mod5

| ID | Structure (ACP5) | m/z z=+2 | | | |
|---|---|---|---|---|---|
| Ext0 |  | 923.4389 | | | |
| Ext1 |  | 937.4545 | | | |
| Ext2 |  | 951.4702 | | | |
| Ext3 |  | 965.4858 | | | |
| Ext4 |  | 979.5015 | | | |
| | | | | | |
| | | | | | |

# **5.** Chapter 5 – Summary, conclusions and future directions

In the doctoral work presented in this thesis, I have described studies that have expanded our understanding of PKS biochemistry and demonstrated the production of an economically and environmentally important small molecule, adipic acid.

The first data chapter describes how expression and purification of the first two modules of borrelidin PKS allowed for the first *in vitro* studies of a carboxyacyl processing PKS system. Using a variety of carboxy- and descarboxyacyl substrates, the range of accepted substrates for these modules was interrogated. It was discovered that the loading module was specific for carboxylated substrates of a specific geometric configuration and would not accept descarboxy analogs, however it is otherwise promiscuous in its ability to accept simpler, non-cyclic substrates like succinyl-CoA in addition to its natural *trans*-1,2-cyclopentanedicarboxylic acid starter unit. In contrast, the extension module, when provided synthetically loaded ACP substrates was comparatively promiscuous in its ability to extend carboxy and descarboxy substrates with no apparent stereoselectivity. It remains to be determined if such tolerance to carboxy groups in PKS intermediates required specific adaptations in the evolution of the borrelidin cluster or whether other modules could also process such substrates. This could be easily tested by performing similar synthetic loading/extension assays using modules from clusters that naturally process simpler substrates such as erythromycin PKS. An interesting observation based on sequence analysis suggests that carboxyacyl processing PKS systems may have a recognizable signature in the form of a degenerated KSQ domain. It would be interesting to mine the sequence databases for such a signature in an attempt to identify additional carboxyacyl processing PKS systems.

Building on the results described in Chapter 2, the first extension module of the borrelidin cluster was engineered to contain a full, catalytically competent reductive loop. The purpose of creating such gain-of-function mutants was to produce the important commodity chemical adipic acid from readily available succinyl- and malonyl-CoA precursors. In contrast to previous PKS engineering efforts where success is determined by the presence or absence of final product, intermediate analysis was extensively used in order to "debug" the sub-optimally active PKS. Such analysis was critical in identifying an unexpected failure of the dehydratase domain to process the intermediate and by iterating through a design-build-test cycle, the behavior of the enzyme was improved to the point where the highly engineered module, assembled with parts from four disparate clusters, was able to produce adipic acid. Despite this proof of principle, the kinetics of the system have yet to be studied, nor has production of adipic acid using the loading module and succinyl-CoA to initiate biosynthesis rather than a SNAC substrate analog been attempted. Further improvements could no doubt be made to the enzyme by finer sampling of chimeric boundaries and with a thioesterase better tuned to the kinetics of adipic-ACP formation to prevent premature hydrolysis of the partially reduced intermediates. *In vivo* production of adipic acid has been a long-standing milestone of this research and each iterative improvement in enzyme activity is a step towards this goal.

In the final data chapter, I have performed the first *in vitro* characterization of an iterative module from an otherwise type I modular PKS cluster. Such experiments allowed for the provision of different acyl-SNAC substrates and sensitive analysis of product formation at different time points using time-of-flight mass spectrometry. The data indicates that the specific identity of the intermediate present on the acyl-carrier protein has a profound impact on the rate of chain-elongation as well as the distribution of extension products that accumulate. Additional experiments with the downstream extension module, though preliminary, suggest two layers of regulation work in concert to ensure product fidelity in this iterative system, consistent with previously proposed models. The specific molecular determinants of iteration remain unknown, however the solubility and *in vitro* activity of BorMod5 suggest it is a tractable system for subsequent biochemical and structural studies to further this line of inquiry.

# **6.** References


Aakvik, T., Degnes, K.F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Völker, U., Ellingsen, T.E., and Valla, S. (2009). A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiol. Lett. *296*, 149–158.

Alessi, V., Penzo, R., Slater, M.J., and Tessari, R. (1997). Caprolactam production: A comparison of different layouts of the liquid-liquid extraction section. Chem. Eng. Technol. *20*, 445–454.

Alini, S., Basile, F., Blasioli, S., Rinaldi, C., and Vaccari, A. (2007). Development of new catalysts for N2O-decomposition from adipic acid plant. Appl. Catal. B Environ. *70*, 323–329.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., and Mohanty, D. (2010). SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res. *38*, W487–W496.

Anderson, B.F., Herlt, A.J., Rickards, R.W., and Robertson, G.B. (1989). Crystal and molecular structures of two isomorphous solvates of the macrolide antibiotic borrelidin: absolute configuration determination by incorporation of a chiral solvent in the crystal lattice. Aust. J. Chem. *42*, 717–730.

Aparicio, J.F., Caffrey, P., Marsden, A.F., Staunton, J., and Leadlay, P.F. (1994). Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. J. Biol. Chem. *269*, 8524–8528.

Baltz, R.H. (2010). *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J. Ind. Microbiol. Biotechnol. *37*, 759–772.

Belecki, K., and Townsend, C.A. (2013). Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Enediyne Polyketide Synthase CalE8. J. Am. Chem. Soc. *135*, 14339–14348.

Berger, J., Jampolsky, L.M., and Goldberg, M.W. (1949). Borrelidin, a new antibiotic with antiborrelia activity and penicillin enhancement properties. Arch. Biochem. *22*, 476–478.

Betancor, L., Fernández, M.-J., Weissman, K.J., and Leadlay, P.F. (2008). Improved Catalytic Activity of a Purified Multienzyme from a Modular Polyketide Synthase after Coexpression with *Streptomyces* Chaperonins in *Escherichia coli* . ChemBioChem *9*, 2962–2966.

Boddy, C.N. (2014). Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. J. Ind. Microbiol. Biotechnol. *41*, 443–450.

Bumpus, S.B., and Kelleher, N.L. (10). Accessing natural product biosynthetic processes by mass spectrometry. Curr Opin Chem Biol *12*, 475–482.

Burgard, A.P., Pharkya, P., and Osterhout, R.E. (2012). Microorganisms for the production of adipic acid and other compounds (Patent US8088607).

Busch, B., Ueberschaar, N., Sugimoto, Y., and Hertweck, C. (2012). Interchenar Retrotransfer of Aureothin Intermediates in an Iterative Polyketide Synthase Module. J. Am. Chem. Soc. *134*, 12382–12385.

Busch, B., Ueberschaar, N., Behnken, S., Sugimoto, Y., Werneburg, M., Traitcheva, N., He, J., and Hertweck, C. (2013). Multifactorial Control of Iteration Events in a Modular Polyketide Assembly Line. Angew. Chem. *125*, 5393–5397.

Cabantous, S., and Waldo, G.S. (2006). In vivo and in vitro protein solubility assays using split GFP. Nat Meth *3*, 845–854.

Caffrey, P. (2003). Conserved Amino Acid Residues Correlating With Ketoreductase Stereospecificity in Modular Polyketide Synthases. ChemBioChem *4*, 654–657.

Callahan, B., Thattai, M., and Shraiman, B.I. (2009). Emergent gene order in a model of modular polyketide synthases. Proc Natl Acad Sci U A *106*, 19410–19415.

Chandran, S.S., Menzella, H.G., Carney, J.R., and Santi, D.V. (2006). Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. Chem Biol *13*, 469–474.

Chen, A.Y., and Khosla, C. (2008). Protein Engineering of Modular Polyketide Synthases. In Protein Engineering Handbook, (Wiley-VCH Verlag GmbH & Co. KGaA), pp. 797–827.

Chen, A.Y., Schnarr, N.A., Kim, C.-Y., Cane, D.E., and Khosla, C. (2006). Extender Unit and Acyl Carrier Protein Specificity of Ketosynthase Domains of the 6-Deoxyerythronolide B Synthase. J. Am. Chem. Soc. *128*, 3067–3074.

Chen, A.Y., Cane, D.E., and Khosla, C. (2007). Structure-Based Dissociation of a Type I Polyketide Synthase Module. Chem. Biol. *14*, 784–792.

Cornvik, T., Dahlroth, S.-L., Magnusdottir, A., Herman, M.D., Knaust, R., Ekberg, M., and Nordlund, P. (2005). Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. Nat Meth *2*, 507–509.

Dickinson, L., Griffiths, A.J., Mason, C.G., and Mills, R.F.N. (1965). Anti-Viral Activity of Two Antibiotics Isolated from a Species of Streptomyces. Nature *206*, 265–268.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675–679.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M., and Katz, L. (1993). An erythromycin analog produced by reprogramming of polyketide synthesis. Proc. Natl. Acad. Sci. *90*, 7119–7123.

Dorrestein, P.C., Bumpus, S.B., Calderone, C.T., Garneau-Tsodikova, S., Aron, Z.D., Straight, P.D., Kolter, R., Walsh, C.T., and Kelleher, N.L. (2006). Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry †. Biochemistry (Mosc.) *45*, 12756–12766.

Dunn, B.J., and Khosla, C. (2013). Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. J. R. Soc. Interface *10*, 20130297–20130297.

Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structure of a modular polyketide synthase. Nature *510*, 512–517.

Van Duuren, J.B.J.H., Brehmer, B., Mars, A.E., Eggink, G., dos Santos, V.A.P.M., and Sanders, J.P.M. (2011). A limited LCA of bio-adipic acid: Manufacturing the nylon-6,6 precursor adipic acid using the benzoic acid degradation pathway from different feedstocks. Biotechnol. Bioeng. *108*, 1298–1306.

Van Duuren, J.B.J.H., Wijte, D., Karge, B., Martins dos Santos, V.A.P., Yang, Y., Mars, A.E., and Eggink, G. (2012). pH-stat fed-batch process to enhance the production of cis, cis-muconate from benzoate by *Pseudomonas putida* KT2440-JD1. Biotechnol. Prog. *28*, 85–92.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

E. I. du Pont de Nemours & Co., I., Wilmingotn, Del., (1998). Zytel Nylon Resins, Product Guide and Properties.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS ONE *4*, e5553.

Epstein, B.N. (1979). Tough thermoplastic nylon compositions (Patent US4174358).

Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T., and Liu, D.R. (2007). Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proc Natl Acad Sci U A *104*, 11951–11956.

Fleming, F.F., and Fleming, F.F. (1999). Nitrile-containing natural products. Nat. Prod. Rep. *16*, 597–606.

Gaisser, S., Kellenberger, L., Kaja, A.L., Weston, A.J., Lill, R.E., Wirtz, G., Kendrew, S.G., Low, L., Sheridan, R.M., Wilkinson, B., et al. (2003). Direct production of

ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of *Streptomyces avermitilis* ATCC31272. Org. Biomol. Chem. *1*, 2840.

Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J. Biol. Chem. *277*, 13082–13090.

Gerth, K., Steinmetz, H., HöFle, G., and Reichenbach, H. (2000). Studies on the Biosynthesis of Epothilones: The Biosynthetic Origin of the Carbon Skeleton. J. Antibiot. (Tokyo) *53*, 1373–1377.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gomez-Escribano, J.P., and Bibb, M.J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters: Streptomyces host for heterologous expression of gene clusters. Microb. Biotechnol. *4*, 207–215.

Greving, M., Cheng, X., Reindl, W., Bowen, B., Deng, K., Louie, K., Nyman, M., Cohen, J., Singh, A., Simmons, B., et al. (2012). Acoustic deposition with NIMS as a high-throughput enzyme activity assay. Anal. Bioanal. Chem. *403*, 707–711.

Guo, X., Liu, T., Valenzano, C.R., Deng, Z., and Cane, D.E. (2010). Mechanism and Stereospecificity of a Fully Saturating Polyketide Synthase Module: Nanchangmycin Synthase Module 2 and Its Dehydratase Domain. J. Am. Chem. Soc. *132*, 14694–14696.

Hamedi, J., Malekzadeh, F., and Saghafi-nia, A.E. (2004). Enhancing of erythromycin production by *Saccharopolyspora erythraea* with common and uncommon oils. J. Ind. Microbiol. Biotechnol. *31*, 447–456.

Hardt, I.H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H., and Höfle, G. (2001). New Natural Epothilones from *Sorangium cellulosum* , Strains So ce90/B2 and So ce90/D13: Isolation, Structure Elucidation, and SAR Studies. J. Nat. Prod. *64*, 847–856.

Harwood, C.S., and Parales, R.E. (1996). The β-ketoadipate pathway and the biology of self-identity. Annu. Rev. Microbiol. *50*, 553–590.

He, J., and Hertweck, C. (2003). Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. Chem. Biol. *10*, 1225–1232.

He, J., and Hertweck, C. (2005). Functional Analysis of the Aureothin Iterative Type I Polyketide Synthase. ChemBioChem *6*, 908–912.

Hertweck, C. (2009). The Biosynthetic Logic of Polyketide Diversity. Angew. Chem. Int. Ed. *48*, 4688–4716.

Hillson, N.J., Rosengarten, R.D., and Keasling, J.D. (2012). j5 DNA Assembly Design Automation Software. ACS Synth. Biol. *1*, 14–21.

Honeywell, M., N.J. (1998). Capron Nylon Homopolymers for Molding and Extrusion.

Hong, H., Appleyard, A.N., Siskos, A.P., Garcia-Bernardo, J., Staunton, J., and Leadlay, P.F. (2005). Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry: Limited proteolysis and MS of modular PKSs. FEBS J. *272*, 2373–2387.

Hong, H., Leadlay, P.F., and Staunton, J. (2009). The changing patterns of covalent active site occupancy during catalysis on a modular polyketide synthase multienzyme revealed by ion-trap mass spectrometry: Changing patterns of covalent active-site occupancy. FEBS J. *276*, 7057–7069.

Hughes, A.J., and Keatinge-Clay, A. (2011). Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Func-tional Showcasing of Streptomyces coelicolor MatB. Chem. Biol. *18*, 165–176.

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S., and Fujita, N. (2013). DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. *41*, D408–D414.

Iniesta, A.A., Garcia-Heras, F., Abellon-Ruiz, J., Gallego-Garcia, A., and Elias-Arnanz, M. (2012). Two Systems for Conditional Gene Expression in *Myxococcus xanthus* Inducible by Isopropyl-*D*-Thiogalactopyranoside or Vanillate. J. Bacteriol. *194*, 5875–5885.

International symposium on genetics of industrial microorganisms, Société française de microbiologie., and Section de microbiologie industrielle et de biotechnologie. (1990). Abstract book (Paris (28 rue du Docteur Roux, 75724): Société Française de microbiologie).

Ishiyama, A., Iwatsuki, M., Namatame, M., Nishihara-Tsukashima, A., Sunazuka, T., Takahashi, Y., Ōmura, S., and Otoguro, K. (2011). Borrelidin, a potent antimalarial: stage-specific inhibition profile of synchronized cultures of Plasmodium falciparum. J. Antibiot. (Tokyo) *64*, 381–384.

Jenke-Kodama, H. (2005). Evolutionary Implications of Bacterial Polyketide Synthases. Mol. Biol. Evol. *22*, 2027–2039.

Jenke-Kodama, H., and Dittmann, E. (2009). Evolution of metabolic diversity: Insights from microbial polyketide synthases. Phytochemistry *70*, 1858–1866.

Jenke-Kodama, H., Börner, T., and Dittmann, E. (2006). Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium *Streptomyces avermitilis*. PLoS Comput. Biol. *2*, e132.

Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000). Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. Gene *249*, 153–160.

Kakirde, K.S., Wild, J., Godiska, R., Mead, D.A., Wiggins, A.G., Goodman, R.M., Szybalski, W., and Liles, M.R. (2011). Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene *475*, 57–62.

Kapur, S., Lowry, B., Yuzawa, S., Kenthirapalan, S., Chen, A.Y., Cane, D.E., and Khosla, C. (2012). Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. Proc. Natl. Acad. Sci. *109*, 4110–4115.

Keatinge-Clay, A. (2008). Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. J. Mol. Biol. *384*, 941–953.

Keatinge-Clay, A.T. (2007). A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. Chem. Biol. *14*, 898–908.

Keatinge-Clay, A.T. (2012). The structures of type I polyketide synthases. Nat. Prod. Rep. *29*, 1050.

Kellenberger, L., Galloway, I.S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J., and Leadlay, P.F. (2008). A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases. ChemBioChem *9*, 2740–2749.

Khosla, C. (2009). Structures and Mechanisms of Polyketide Synthases. J Org Chem *74*, 6416–6420.

Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A., and Cane, D.E. (2007). Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. Annu. Rev. Biochem. *76*, 195–221.

Khosla, C., Kapur, S., and Cane, D.E. (2009). Revisiting the modularity of modular polyketide synthases. Curr. Opin. Chem. Biol. *13*, 135–143.

Khosla, C., Herschlag, D., Cane, D.E., and Walsh, C.T. (2014). Assembly line polyketide synthases: mechanistic insights and unsolved problems. Biochemistry (Mosc.) *53*, 2875–2883.

Kim, B.S., Sherman, D.H., and Reynolds, K.A. (2004a). An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. Protein Eng Sel *17*, 277–284.

Kim, C.-Y., Alekseyev, V.Y., Chen, A.Y., Tang, Y., Cane, D.E., and Khosla, C. (2004b). Reconstituting Modular Activity from Separated Domains of 6-Deoxyerythronolide B Synthase. Biochemistry (Mosc.) *43*, 13892–13898.

Kohan, M.I. (1995). Nylon plastics handbook (Munich ; New York : Cincinnati: Hanser Publishers ; Distributed in the USA and in Canada by Hanser/Gardner Publications).

Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E., and Ikeda, H. (2010). Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc. Natl. Acad. Sci. *107*, 2646–2651.

Kudo, F., Kitayama, T., Kakinuma, K., and Eguchi, T. (2006). Macrolactam formation catalyzed by the thioesterase domain of vicenistatin polyketide synthase. Tetrahedron Lett. *47*, 1529–1532.

Lee, T.S., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., and Keasling, J.D. (2011). BglBrick vectors and datasheets: a synthetic biology platform for gene expression. J. Biol. Eng. *5*, 1–14.

Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J.D., and Hillson, N.J. (2013). PaR-PaR Laboratory Automation Platform. ACS Synth. Biol. *2*, 216–222.

Linshiz, G., Stawski, N., Goyal, G., Bi, C., Poust, S., Sharma, M., Mutalik, V., Keasling, J.D., and Hillson, N.J. (2014). PR-PR: Cross-Platform Laboratory Automation System. ACS Synth. Biol. *3*, 515–524.

Long, P.F., Wilkinson, C.J., Bisang, C.P., Cortés, J., Dunster, N., Oliynyk, M., McCormick, E., McArthur, H., Mendez, C., Salas, J.A., et al. (2002). Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. Mol. Microbiol. *43*, 1215–1225.

Lumb, M., Macey, P.E., Spyvee, J., Whitmarsh, J.M., and Wright, R.D. (1965). Isolation of vivomycin and borrelidin, two antibiotics with anti-viral activity, from a species of *Streptomyces* (C2989). Nature *206*, 263–265.

Marsden, A.F., Caffrey, P., Aparicio, J.F., Loughran, M.S., Staunton, J., and Leadlay, P.F. (1994). Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. Science *263*, 378–380.

Martinez, A., Kolvek, S.J., Yip, C.L.T., Hopke, J., Brown, K.A., MacNeil, I.A., and Osburne, M.S. (2004). Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. Appl. Environ. Microbiol. *70*, 2452–2463.

McDaniel, R. (1997). Engineered intermodular and intramodular polyketide synthase fusions. Chem. Biol. *4*, 667–674.

McDaniel, R., Ebert-Khosla, S., Hopwood, D.A., and Khosla, C. (1995). Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. Nature *375*, 549–554.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M., and Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide

synthase to produce a library of novel “unnatural” natural products. Proc. Natl. Acad. Sci. *96*, 1846–1851.

Meiser, P., Weissman, K.J., Bode, H.B., Krug, D., Dickschat, J.S., Sandmann, A., and Müller, R. (2008). DKxanthene Biosynthesis—Understanding the Basis for Diversity-Oriented Synthesis in Myxobacterial Secondary Metabolism. Chem. Biol. *15*, 771–781.

Meluzzi, D., Zheng, W.H., Hensler, M., Nizet, V., and Dorrestein, P.C. (2008). Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. Bioorg. Med. Chem. Lett. *18*, 3107–3111.

Menche, D., Arikan, F., Perlova, O., Horstmann, N., Ahlbrecht, W., Wenzel, S.C., Jansen, R., Irschik, H., and Müller, R. (2008). Stereochemical Determination and Complex Biosynthetic Assembly of Etnangien, a Highly Potent RNA Polymerase Inhibitor from the Myxobacterium *Sorangium cellulosum*. J. Am. Chem. Soc. *130*, 14234–14243.

Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., and Santi, D.V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. *23*, 1171–1176.

Menzella, H.G., Reisinger, S.J., Welch, M., Kealey, J.T., Kennedy, J., Reid, R., Tran, C.Q., and Santi, D.V. (2006). Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. J. Ind. Microbiol. Biotechnol. *33*, 22–28.

Menzella, H.G., Carney, J.R., and Santi, D.V. (2007). Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. Chem. Biol. *14*, 143–151.

Menzella, H.G., Carney, J.R., Li, Y., and Santi, D.V. (2010). Using Chemobiosynthesis and Synthetic Mini-Polyketide Synthases To Produce Pharmaceutical Intermediates in Escherichia coli. Appl. Environ. Microbiol. *76*, 5221–5227.

Molnar, I., Schupp, T., Ono, M., Zirkle, R.E., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., et al. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem. Biol. *7*, 97–109.

Moss, S.J., Martin, C.J., and Wilkinson, B. (2004). Loss of colinearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. Nat. Prod. Rep. *21*, 575.

Moss, S.J., Carletti, I., Olano, C., Sheridan, R.M., Ward, M., Math, V., Nur-E-Alam, M., Braña, A.F., Zhang, M.Q., Leadlay, P.F., et al. (2006). Biosynthesis of the angiogenesis inhibitor borrelidin: directed biosynthesis of novel analogs. Chem. Commun. 2341.

Murli, S., Kennedy, J., Dayem, L.C., Carney, J.R., and Kealey, J.T. (2003). Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. J. Ind. Microbiol. Biotechnol. *30*, 500–509.

Mutka, S.C., Carney, J.R., Liu, Y., and Kennedy, J. (2006). Heterologous Production of Epothilone C and D in *Escherichia coli* . Biochemistry (Mosc.) *45*, 1321–1330.

Niu, W., Draths, K.M., and Frost, J.W. (2002). Benzene-Free Synthesis of Adipic Acid. Biotechnol. Prog. *18*, 201–211.

Olano, C., Wilkinson, B., Moss, S.J., Braña, A.F., Méndez, C., Leadlay, P.F., and Salas, J.A. (2003). Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase. Chem. Commun. 2780.

Olano, C., Wilkinson, B., Sánchez, C., Moss, S.J., Sheridan, R., Math, V., Weston, A.J., Braña, A.F., Martin, C.J., Oliynyk, M., et al. (2004a). Biosynthesis of the Angiogenesis Inhibitor Borrelidin by *Streptomyces parvulus* Tü4055: Cluster Analysis and Assignment of Functions. Chem. Biol. *11*, 87–97.

Olano, C., Moss, S.J., Braña, A.F., Sheridan, R.M., Math, V., Weston, A.J., Méndez, C., Leadlay, P.F., Wilkinson, B., and Salas, J.A. (2004b). Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: insights into nitrile formation: Nitrile formation during borrelidin biosynthesis. Mol. Microbiol. *52*, 1745–1756.

Olano, C., Méndez, C., and Salas, J.A. (2010). Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. Nat. Prod. Rep. *27*, 571.

Palmer, R.J. (2001). Polyamides, Plastics. In Encyclopedia of Polymer Science and Technology, John Wiley & Sons, Inc., ed. (Hoboken, NJ, USA: John Wiley & Sons, Inc.),.

Perlova, O., Gerth, K., Kuhlmann, S., Zhang, Y., and Müller, R. (2009). Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate *Corallococcus macrosporus* GT-2. Microb. Cell Factories *8*, 1.

Pfeifer, B.A. (2001). Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. Science *291*, 1790–1792.

Picataggio, S., and Beardslee, T. (2012). Biological methods for preparing adipic acid (Patent US20120021474).

Pieper, R., Luo, G., Cane, D.E., and Khosla, C. (1995). Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. Nature *378*, 263–266.

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

Richter, F., Leaver-Fay, A., Khare, S.D., Bjelic, S., and Baker, D. (2011). De novo enzyme design using Rosetta3. PLoS One *6*, e19230.

De Rond, T., Peralta-Yahya, P., Cheng, X., Northen, T.R., and Keasling, J.D. (2013). Versatile synthesis of probes for high-throughput enzyme activity screening. Anal. Bioanal. Chem. *405*, 4969–4973.

Sato, K. (1998). A "Green" Route to Adipic Acid: Direct Oxidation of Cyclohexenes with 30 Percent Hydrogen Peroxide. Science *281*, 1646–1647.

Scaglione, J.B., Akey, D.L., Sullivan, R., Kittendorf, J.D., Rath, C.M., Kim, E.-S., Smith, J.L., and Sherman, D.H. (2010). Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. Angew. Chem. Int. Ed. *49*, 5726–5730.

Schulze, C.J., Bray, W.M., Loganzo, F., Lam, M.-H., Szal, T., Villalobos, A., Koehn, F.E., and Linington, R.G. (2014). Borrelidin B: Isolation, Biological Activity, and Implications for Nitrile Biosynthesis. J. Nat. Prod. 141117141413001.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Stachelhaus, T., and Marahiel, M.A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. FEMS Microbiol. Lett. *125*, 3–14.

Sugimoto, Y., Ding, L., Ishida, K., and Hertweck, C. (2014). Rational Design of Modular Polyketide Synthases: Morphing the Aureothin Pathway into a Luteoreticulin Assembly Line. Angew. Chem. Int. Ed. *53*, 1560–1564.

Tae, H., Sohng, J.K., and Park, K. (2009). Development of an analysis program of type I polyketide synthase gene clusters using homology search and profile hidden Markov model. J. Microbiol. Biotechnol. *19*, 140–146.

Takahashi, S., Toyoda, A., Sekiyama, Y., Takagi, H., Nogawa, T., Uramoto, M., Suzuki, R., Koshino, H., Kumano, T., Panthee, S., et al. (2011). Reveromycin A biosynthesis uses RevG and RevJ for stereospecific spiroacetal formation. Nat. Chem. Biol. *7*, 461–468.

Tang, Y., Tsai, S.-C., and Khosla, C. (2003). Polyketide chain length control by chain length factor. J. Am. Chem. Soc. *125*, 12708–12709.

Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E., and Khosla, C. (2006). The 2.7-A crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. *103*, 11124–11129.

Tatsuno, S., Arakawa, K., and Kinashi, H. (2007). Analysis of Modular-iterative Mixed Biosynthesis of Lankacidin by Heterologous Expression and Gene Fusion. J. Antibiot. (Tokyo) *60*, 700–708.

Traitcheva, N., Jenke-Kodama, H., He, J., Dittmann, E., and Hertweck, C. (2007). Non-Colinear Polyketide Biosynthesis in the Aureothin and Neoaureothin Pathways: An Evolutionary Perspective. ChemBioChem *8*, 1841–1849.

Troeschel, S.C., Thies, S., Link, O., Real, C.I., Knops, K., Wilhelm, S., Rosenau, F., and Jaeger, K.-E. (2012). Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. J. Biotechnol. *161*, 71–79.

Tropea, J.E., Cherry, S., and Waugh, D.S. (2009). Expression and Purification of Soluble His6-Tagged TEV Protease. In High Throughput Protein Expression and Purification, S.A. Doyle, ed. (Humana Press), pp. 297–307.

Tsai, S.-C., Miercke, L.J., Krucinski, J., Gokhale, R., Chen, J.C.-H., Foster, P.G., Cane, D.E., Khosla, C., and Stroud, R.M. (2001). Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. Proc. Natl. Acad. Sci. *98*, 14808–14813.

Tsai, S.-C., Lu, H., Cane, D.E., Khosla, C., and Stroud, R.M. (2002). Insights into Channel Architecture and Substrate Specificity from Crystal Structures of Two Macrocycle-Forming Thioesterases of Modular Polyketide Synthases. Biochemistry (Mosc.) *41*, 12598–12606.

Vergnolle, O., Hahn, F., Baerga-Ortiz, A., Leadlay, P.F., and Andexer, J.N. (2011). Stereoselectivity of Isolated Dehydratase Domains of the Borrelidin Polyketide Synthase: Implications for *cis* Double Bond Formation. ChemBioChem *12*, 1011–1014.

Volokhan, O., Sletta, H., Ellingsen, T.E., and Zotchev, S.B. (2006). Characterization of the P450 Monooxygenase NysL, Responsible for C-10 Hydroxylation during Biosynthesis of the Polyene Macrolide Antibiotic Nystatin in *Streptomyces noursei*. Appl. Environ. Microbiol. *72*, 2514–2519.

Wakabayashi, T., Kageyama, R., Naruse, N., Tsukahara, N., Funahashi, Y., Kitoh, K., and Watanabe, Y. (1997). Borrelidin is an angiogenesis inhibitor; disruption of angiogenic capillary vessels in a rat aorta matrix culture model. J. Antibiot. (Tokyo) *50*, 671–676.

Walker, M.C., Thuronyi, B.W., Charkoudian, L.K., Lowry, B., Khosla, C., and Chang, M.C.Y. (2013). Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. Science *341*, 1089–1094.

Wang, H.H., and Church, G.M. (2011). Multiplexed Genome Engineering and Genotyping Methods. In Methods in Enzymology, (Elsevier), pp. 409–426.

Watanabe, K., and Oikawa, H. (2007). Robust platform for de novo production of heterologous polyketides and nonribosomal peptides in *Escherichia coli*. Org. Biomol. Chem. *5*, 593.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M., et al. (2012). Mass spectral molecular networking of living microbial colonies. Proc. Natl. Acad. Sci. *109*, E1743–E1752.

Weissman, K.J. (2006). The Structural Basis for Docking in Modular Polyketide Biosynthesis. ChemBioChem *7*, 485–494.

Whicher, J.R., Smaga, S.S., Hansen, D.A., Brown, W.C., Gerwick, W.H., Sherman, D.H., and Smith, J.L. (2013). Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis. Chem. Biol. *20*, 1340–1351.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature *510*, 560–564.

Wilkinson, B., Foster, G., Rudd, B.A., Taylor, N.L., Blackaby, A.P., Sidebottom, P.J., Cooper, D.J., Dawson, M.J., Buss, A.D., Gaisser, S., et al. (2000). Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. Chem. Biol. *7*, 111–117.

Winter, J.M., and Tang, Y. (2012). Synthetic biological approaches to natural product biosynthesis. Curr. Opin. Biotechnol. *23*, 736–743.

Wu, N., Kudo, F., Cane, D.E., and Khosla, C. (2000). Analysis of the Molecular Recognition Features of Individual Modules Derived from the Erythromycin Polyketide Synthase. J. Am. Chem. Soc. *122*, 4847–4852.

Yan, J., Gupta, S., Sherman, D.H., and Reynolds, K.A. (2009). Functional Dissection of a Multimodular Polypeptide of the Pikromycin Polyketide Synthase into Monomodules by Using a Matched Pair of Heterologous Docking Domains. ChemBioChem *10*, 1537–1543.

Yoon, Y.J., Beck, B.J., Kim, B.S., Kang, H.-Y., Reynolds, K.A., and Sherman, D.H. (2002). Generation of Multiple Bioactive Macrolides by Hybrid Modular Polyketide Synthases in Streptomyces venezuelae. Chem. Biol. *9*, 203–214.

Yuzawa, S., Kim, W., Katz, L., and Keasling, J.D. (2012). Heterologous production of polyketides by modular type I polyketide synthases in Escherichia coli. Curr. Opin. Biotechnol. *23*, 727–735.

Yuzawa, S., Eng, C.H., Katz, L., and Keasling, J.D. (2013). Broad Substrate Specificity of the Loading Didomain of the Lipomycin Polyketide Synthase. Biochemistry (Mosc.) *52*, 3791–3793.

Zhang, K., Nelson, K.M., Bhuripanyo, K., Grimes, K.D., Zhao, B., Aldrich, C.C., and Yin, J. (2013). Engineering the Substrate Specificity of the DhbE Adenylation Domain by Yeast Cell Surface Display. Chem. Biol. *20*, 92–101.

Zhang, X., Chen, Z., Li, M., Wen, Y., Song, Y., and Li, J. (2006). Construction of ivermectin producer by domain swaps of avermectin polyketide synthase in *Streptomyces avermitilis*. Appl. Microbiol. Biotechnol. *72*, 986–994.

Zhao, J., Zhong, L., Shen, M., Xia, Z., Cheng, Q., Sun, X., Zhao, G., Li, Y., and Qin, Z. (2008). Discovery of the Autonomously Replicating Plasmid pMF1 from *Myxococcus fulvus* and Development of a Gene Cloning System in *Myxococcus xanthus*. Appl. Environ. Microbiol. *74*, 1980–1987.

Zheng, J., Piasecki, S.K., and Keatinge-Clay, A.T. (2013). Structural Studies of an A2-Type Modular Polyketide Synthase Ketoreductase Reveal Features Controlling α-Substituent Stereochemistry. ACS Chem. Biol. *8*, 1964–1971.

Zhou, Y., Meng, Q., You, D., Li, J., Chen, S., Ding, D., Zhou, X., Zhou, H., Bai, L., and Deng, Z. (2008). Selective Removal of Aberrant Extender Units by a Type II Thioesterase for Efficient FR-008/Candicidin Biosynthesis in *Streptomyces* sp. Strain FR-008. Appl. Environ. Microbiol. *74*, 7235–7242.

# 7. Appendices

## 7.1. Plasmids used and relevant nucleotide sequences

| Name | Alias | Product | Chapter described |
|---|---|---|---|
| pARH090 | pBbS2k::6xHisMBP-RFP | N/A | 2, 4 |
| pARH098 | pBbS2k::6xHisMBP-BorA1_v2 | MBP-BorA1 | 2 |
| pARH100 | pBbS2k::6xHisMBP-BorA2 | MBP-BorA2 | 2 |
| pARH110 | pET28a::BorA1_ACPdock | ACP0 monodomain | 2 |
| pARH130 | pBbS2k::6xHisMBP-BorA5 | MBP-BorMod5 | 4 |
| pARH130b | pET28a::BorA5 | BorMod5 | 4 |
| pARH135 | pBbS2k::6xHisMBP-BorA6ΔTE | MBP-BorA6ΔTE | 4 |
| pARH136 | pBbS2k::6xHisMBP-BorA2eloopswap-A1 | MBP-A1 | 3 |
| pARH137 | pBbS2k::6xHisMBP-BorA2eloopswap-A2 | MBP-A2 | 3 |
| pARH138 | pBbS2k::6xHisMBP-BorA2eloopswap-I1 | MBP-I1 | 3 |
| pARH139 | pBbS2k::6xHisMBP-BorA2eloopswap-I2 | MBP-I2 | 3 |
| pARH140 | pBbS2k::6xHisMBP-BorA2eloopswap-N1 | MBP-N1 | 3 |
| pARH141 | pBbS2k::6xHisMBP-BorA2eloopswap-N2 | MBP-N2 | 3 |
| pARH142 | pBbS2k::6xHisMBP-BorA2eloopswap-S1 | MBP-S1 | 3 |
| pARH143 | pBbS2k::6xHisMBP-BorA2eloopswap-S2 | MBP-S2 | 3 |
| pARH147 | pBbS2k::6xHisMBP-BorA2-A2-BorDH2 | MBP-A2c | 3 |
| pARH148 | pBbS2k::6xHisMBP-BorA2-I2-BorDH2 | MBP-I2c | 3 |
| pARH149 | pBbS2k::6xHisMBP-BorA2-S2-BorDH2 | MBP-S2c | 3 |
| pARH150 | pET28a::BorDH2 | BorDH2 monodomain | 3 |
| pARH154 | pBbS2k::6xHisMBP-BorA2-A2-BorDH2-ttmTE | MBP-A2c-TE | 3 |
| pARH156 | pBbS2k::6xHisMBP-BorA2-S2-BorDH2-ttmTE | MBP-S2c-TE | 3 |
| pARH159 | pBbS2k::6xHisMBP-BorA2-A1.5 | MBP-A1.5 | 3 |
| pARH162 | pBbS2k::6xHisMBP-BorA2-S1.5 | MBP-S1.5 | 3 |
| pARH163 | pBbS2k::6xHisMBP-BorA2-A1.5-BorDH2 | MBP-A1.5c | 3 |
| pARH164 | pBbS2k::6xHisMBP-BorA2-S1.5-BorDH2 | MBP-S1.5c | 3 |
| pARH170 | pBbS2k::6xHisMBP-BorA5_Mut222 | MBP-BorMod5-AT1 | 3 |

### 7.1.1. pARH090

```
LOCUS       pBbS2k_6xHisMBP_RFP       6025 bp ds-DNA   circular     27-JAN-2012
DEFINITION  tetR and tet promoter amplified from Sung's pZB vector.
SOURCE
  ORGANISM
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     gene            896..1993
                     /label=MBP
                     /ApEinfo_fwdcolor=#e3b760
                     /ApEinfo_revcolor=#e3b760
     CDS             complement(join(143..187,189..769))
                     /vntifkey="4"
                     /label=tetR
                     /ApEinfo_fwdcolor=pink
                     /ApEinfo_revcolor=pink
     misc_feature    2741..2869
                     /vntifkey="21"
                     /label=dbl\term
                     /ApEinfo_fwdcolor=#d4c7ff
                     /ApEinfo_revcolor=#d4c7ff
     CDS             2042..2716
                     /vntifkey="4"
                     /label=RFP
                     /ApEinfo_fwdcolor=pink
                     /ApEinfo_revcolor=pink
     promoter        770..842
                     /vntifkey="30"
                     /label=tet\promoter\region
                     /ApEinfo_fwdcolor=#cea656
                     /ApEinfo_revcolor=#cea656
     protein_bind    791..809
                     /vntifkey="31"
                     /label=Q1
                     /ApEinfo_fwdcolor=#3edbff
                     /ApEinfo_revcolor=#3edbff
     protein_bind    821..837
                     /vntifkey="31"
                     /label=Q2
                     /ApEinfo_fwdcolor=#3edbff
                     /ApEinfo_revcolor=#3edbff
     misc_marker     complement(5231..6025)
                     /vntifkey="22"
                     /note="encodes nptII (aka AphA, neoR), gives kan and neo
                     resistance, with native promoter in on AatII and SpeI from
                     Lutz vector pZE2PLtetO-1 MCS2"
                     /label=kan/neoR
                     /ApEinfo_fwdcolor=pink
                     /ApEinfo_revcolor=pink
     terminator      5100..5205
                     /vntifkey="43"
                     /label=T0
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
     rep_origin      complement(join(2876..3293,3296..5099))
                     /vntifkey="33"
                     /note="pSC101* aka pMPP6, gives plasmid number 3-4 copies
                     per cell, BglII site in pSC101* ori has been deleted by
                     quick change agatcT changed to agatcA giving pSC101**"
                     /label=pSC101**\ori
                     /ApEinfo_fwdcolor=pink
                     /ApEinfo_revcolor=pink
     modified_base   188..188
                     /label=I>I
                     /ApEinfo_fwdcolor=#386ae9
                     /ApEinfo_revcolor=#386ae9
     modified_base   3294..3295
                     /label=pSC101***
                     /ApEinfo_fwdcolor=#386ae9
                     /ApEinfo_revcolor=#386ae9
     protein_bind    1994..2035
```

```
                     /label=TEV cleavage site
                     /ApEinfo_fwdcolor=#3edbff
                     /ApEinfo_revcolor=#3edbff
ORIGIN
        1 gcgaaacgat cctcatcctg tctcttgatc agatcatgat cccctgcgcc atcagatcct
       61 tggcggcaag aaagccatcc agtttacttt gcagggcttc ccaaccttac cagagggcgc
      121 cccagctggc aattccgacg tcttaagacc cactttcaca tttaagttgt ttttctaatc
      181 cgcatatAat caattcaagg ccgaataaga aggctggctc tgcaccttgg tgatcaaata
      241 attcgatagc ttgtcgtaat aatggcggca tactatcagt agtaggtgtt tccctttctt
      301 ctttagcgac ttgatgctct tgatcttcca atacgcaacc taaagtaaaa tgccccacag
      361 cgctgagtgc atataatgca ttctctagtg aaaaaccttg ttggcataaa aaggctaatt
      421 gattttcgag agtttcatac tgtttttctg taggccgtgt acctaaatgt acttttgctc
      481 catcgcgatg acttagtaaa gcacatctaa aacttttagc gttattacgt aaaaaatctt
      541 gccagctttc cccttctaaa gggcaaaagt gagtatggtg cctatctaac atctcaatgg
      601 ctaaggcgtc gagcaaagcc cgcttatttt ttacatgcca atacaatgta ggctgctcta
      661 cacctagctt ctgggcgagt ttacgggttg ttaaaccttc gattccgacc tcattaagca
      721 gctctaatgc gctgttaatc actttacttt tatctaatct agacatcatt aattcctaat
      781 ttttgttgac actctatcgt tgatagagtt attttaccac tccctatcag tgatagagaa
      841 aagaattctt taagaaggag atataATGCA TCATCATCAT CATCACagca gcggcAAAAT
      901 CGAAGAAGGT AAACTGGTAA TCTGGATTAA CGGCGATAAA GGCTATAACG GTCTCGCTGA
      961 AGTCGGTAAG AAATTCGAGA AAGATACCGG AATTAAAGTC ACCGTTGAGC ATCCGGATAA
     1021 ACTGGAAGAG AAATTCCCAC AGGTTGCGGC AACTGGCGAT GGCCCTGACA TTATCTTCTG
     1081 GGCACACGAC CGCTTTGGTG GCTACGCTCA ATCTGGCCTG TTGGCTGAAA TCACCCCGGA
     1141 CAAAGCGTTC CAGGACAAGC TGTATCCGTT TACCTGGGAT GCCGTACGTT ACAACGGCAA
     1201 GCTGATTGCT TACCCGATCG CTGTTGAAGC GTTATCGCTG ATTTATAACA AAGATCTGCT
     1261 GCCGAACCCG CCAAAAACCT GGGAAGAGAT CCCGGCGCTG GATAAAGAAC TGAAAGCGAA
     1321 AGGTAAGAGC GCGCTGATGT TCAACCTGCA AGAACCGTAC TTCACCTGGC CGCTGATTGC
     1381 TGCTGACGGG GGTTATGCGT TCAAGTATGA AAACGGCAAG TACGACATTA AAGACGTGGG
     1441 CGTGGATAAC GCTGGCGCGA AAGCGGGTCT GACCTTCCTG GTTGACCTGA TTAAAAACAA
     1501 ACACATGAAT GCAGACACCG ATTACTCCAT CGCAGAAGCT GCCTTTAATA AAGGCGAAAC
     1561 AGCGATGACC ATCAACGGCC CGTGGGCATG GTCCAACATC GACACCAGCA AAGTGAATTA
     1621 TGGTGTAACG GTACTGCCGA CCTTCAAGGG TCAACCATCC AAACCGTTCG TTGGCGTGCT
     1681 GAGCGCAGGT ATTAACGCCG CCAGTCCGAA CAAAGAGCTG GCAAAAGAGT TCCTCGAAAA
     1741 CTATCTGCTG ACTGATGAAG GTCTGGAAGC GGTTAATAAA GACAAACCGC TGGGTGCCGT
     1801 AGCGCTGAAG TCTTACGAGG AAGAGTTGGC GAAAGATCCA CGTATTGCCG CCACTATGGA
     1861 AAACGCCCAG AAAGGTGAAA TCATGCCGAA CATCCCGCAG ATGTCCGCTT TCTGGTATGC
     1921 CGTGCGTACT GCGGTGATCA ACGCCGCCAG CGGTCGTCAG ACTGTCGATG AAGCCCTGAA
     1981 AGACGCGCAG ACTgattacg atatcccaac gaccgaaaac ctttacttcc agggcCATAT
     2041 Ggcgagtagc gaagacgtta tcaaagagtt catgcgtttc aaagttcgta tggaaggttc
     2101 cgttaacggt cacgagttcg aaatcgaagg tgaaggtgaa ggtcgtccgt acgaaggtac
     2161 ccagaccgct aaactgaaag ttaccaaagg tggtccgctg ccgttcgctt gggacatcct
     2221 gtccccgcag ttccagtacg gttccaaagc ttacgttaaa cacccggctg acatcccgga
     2281 ctacctgaaa ctgtccttcc cggaaggttt caaatgggaa cgtgttatga acttcgaaga
     2341 cggtggtgtt gttaccgtta cccaggactc ctccctgcaa gacggtgagt tcatctacaa
     2401 agttaaactg cgtggtacca acttcccgtc cgacggtccg gttatgcaga aaaaaaccat
     2461 gggttgggaa gcttccaccg aacgtatgta cccggaagac ggtgctctga aaggtgaaat
     2521 caaaatgcgt ctgaaactga aagacggtgg tcactacgac gctgaagtta aaaccaccta
     2581 catggctaaa aaaccggttc agctgccggg tgcttacaaa accgacatca aactggacat
     2641 cacctcccac aacgaagact acaccatcgt tgaacagtac gaacgtgctg aaggtcgtca
     2701 ctccaccggt gcttaaggat ccaaactcga gtaaggatct ccaggcatca aataaaacga
     2761 aaggctcagt cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt gaacgctctc
     2821 tactagagtc acactggctc accttcgggt gggcctttct gcgtttatac ctagggtacg
     2881 ggttttgctg cccgcaaacg ggctgttctg gtgttgctag tttgttatca gaatcgcaga
     2941 tccggcttca gccggtttgc cggctgaaag cgctatttct tccagaattg ccatgatttt
     3001 ttccccacgg gaggcgtcac tggctcccgt gttgtcggca gctttgattc gataagcagc
     3061 atcgcctgtt tcaggctgtc tatgtgtgac tgttgagctg taacaagttg tctcaggtgt
     3121 tcaatttcat gttctagttg ctttgtttta ctggtttcac ctgttctatt aggtgttaca
     3181 tgctgttcat ctgttacatt gtcgatctgt tcatggtgaa cagctttgaa tgcaccaaaa
     3241 actcgtaaaa gctctgatgt atctatcttt tttacaccgt tttcatctgt gcaATtggac
     3301 agttttccct ttgatatgta acggtgaaca gttgttctac ttttgtttgt tagtcttgat
     3361 gcttcactga tagatacaag agccataaga acctcagatc cttccgtatt tagccagtat
     3421 gttctctagt gtggttcgtt gtttttgcgt gagccatgag aacgaaccat tgagatcata
     3481 cttactttgc atgtcactca aaaattttgc ctcaaaactg gtgagctgaa tttttgcagt
     3541 taaagcatcg tgtagtgttt ttcttagtcc gttatgtagg taggaatctg atgtaatggt
     3601 tgttggtatt ttgtcaccat tcatttttat ctggttgttc tcaagttcgg ttacgagatc
     3661 catttgtcta tctagttcaa cttggaaaat caacgtatca gtcgggcggc ctcgcttatc
     3721 aaccaccaat ttcatattgc tgtaagtgtt taaatcttta cttattggtt tcaaaaccca
     3781 ttggttaagc cttttaaact catggtagtt attttcaagc attaacatga acttaaattc
     3841 atcaaggcta atctctatat ttgccttgtg agttttcttt tgtgttagtt cttttaataa
     3901 ccactcataa atcctcatag agtatttgtt ttcaaaagac ttaacatgtt ccagattata
     3961 ttttatgaat ttttttaact ggaaaagata aggcaatatc tcttcactaa aaactaattc
     4021 taatttttcg cttgagaact tggcatagtt tgtccactgg aaaatctcaa agcctttaac
     4081 caaaggattc ctgatttcca cagttctcgt catcagctct ctggttgctt tagctaatac
```

```
     4141 accataagca ttttccctac tgatgttcat catctgagcg tattggttat aagtgaacga
     4201 taccgtccgt tctttccttg tagggttttc aatcgtgggg ttgagtagtg ccacacagca
     4261 taaaattagc ttggtttcat gctccgttaa gtcatagcga ctaatcgcta gttcatttgc
     4321 tttgaaaaca actaattcag acatacatct caattggtct aggtgatttt aatcactata
     4381 ccaattgaga tgggctagtc aatgataatt actagtcctt ttcccgggtg atctgggtat
     4441 ctgtaaattc tgctagacct ttgctggaaa acttgtaaat tctgctagac cctctgtaaa
     4501 ttccgctaga cctttgtgtg tttttttgt ttatattcaa gtggttataa tttatagaat
     4561 aaagaaagaa taaaaaaaga taaaaagaat agatcccagc cctgtgtata actcactact
     4621 ttagtcagtt ccgcagtatt acaaaaggat gtcgcaaacg ctgtttgctc ctctacaaaa
     4681 cagaccttaa aacctaaag gcttaagtag caccctcgca agctcgggca aatcgctgaa
     4741 tattcctttt gtctccgacc atcaggcacc tgagtcgctg tcttttcgt gacattcagt
     4801 tcgctgcgct cacggctctg gcagtgaatg ggggtaaatg gcactacagg cgccttttat
     4861 ggattcatgc aaggaaacta cccataatac aagaaaagcc cgtcacgggc ttctcagggc
     4921 gttttatggc gggtctgcta tgtggtgcta tctgactttt tgctgttcag cagttcctgc
     4981 cctctgattt tccagtctga ccacttcgga ttatcccgtg acaggtcatt cagactggct
     5041 aatgcaccca gtaaggcagc ggtatcatca acaggcttac ccgtcttact gtccctagtg
     5101 cttggattct caccaataaa aaacgcccgg cggcaaccga gcgttctgaa caaatccaga
     5161 tggagttctg aggtcattac tggatctatc aacaggagtc caagcgagct ctcgaacccc
     5221 agagtcccgc tcagaagaac tcgtcaagaa ggcgatagaa ggcgatgcgc tgcgaatcgg
     5281 gagcggcgat accgtaaagc acgaggaagc ggtcagccca ttcgccgcca agctcttcag
     5341 caatatcacg ggtagccaac gctatgtcct gatagcggtc cgccacaccc agccggccac
     5401 agtcgatgaa tccagaaaag cggccatttt ccaccatgat attcggcaag caggcatcgc
     5461 catgggtcac gacgagatcc tcgccgtcgg gcatgcgcgc cttgagcctg gcgaacagtt
     5521 cggctggcgc gagcccctga tgctcttcgt ccagatcatc ctgatcgaca agaccggctt
     5581 ccatccgagt acgtgctcgc tcgatgcgat gtttcgcttg gtggtcgaat gggcaggtag
     5641 ccggatcaag cgtatgcagc cgccgcattg catcagccat gatggatact ttctcggcag
     5701 gagcaaggtg agatgacagg agatcctgcc ccggcacttc gcccaatagc agccagtccc
     5761 ttcccgcttc agtgacaacg tcgagcacag ctgcgcaagg aacgcccgtc gtggccagcc
     5821 acgatagccg cgctgcctcg tcctgcagtt cattcagggc accggacagg tcggtcttga
     5881 caaaaagaac cgggcgcccc tgcgctgaca gccggaacac ggcggcatca gagcagccga
     5941 ttgtctgttg tgcccagtca tagccgaata gcctctccac ccaagcggcc ggagaacctg
     6001 cgtgcaatcc atcttgttca atcat
//
```

### 7.1.2. MBP-BorA1

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGACCGGCAGCGCAGTTTCCGCCCCATTCCTGC
AACCACCAGAACCAGTCAGCGGCCACAGCGAGCGCAAATCCGATCCGGTCCTGCTGGTGGGTGCGGGTCGT
CGTGCACGCATGGCGGATGCAGTCCGTGCCGCTGGTGCTCAAGCTGGTATCGATCCGGCCGTTCTGCGTCG
TACCCGTGCAACCCTGATTACCGCGGGTAGCGCGGGTGCAGCAGGCCGTTTGGCAGCGGCACTGCGTCTGA
CCGGCGCAACCATCTCGCTGGATACGCGCGAAACCCCGACGTTGCTGGCCTTGCACCTGGCAGCGCAAGCA
CTGCGTGCGGGCGACACCTCGTATGCGGTTGTTGGTGCAGAATTGCCGGACGGTAACTGTGCGTTGATCCT
GGCACGTCAGTCGGCAGCGACCGCAGAGGGCGCTGTCCCGCAGGCAATCGTGCGTACGACTACGGCGGATC
GTACTACCACTGCTGATCACGCGCCAGCGCCTGATGATCACGGTAGCCCGGCACGTGAAGCCCCGCATGCG
ACCCGTACGCTGAGCCCGGGTATTACCCAGGCTCCGGCCGAGGGTTTCCCGGGTCTGCTGGCGACCCTGCA
CGATGATACCCCGCTGCGTCCGACCGCCGTGACGGAGCATGGCTCTGACGCTACCACGGTCCTGGTTCTGC
TGGACCAGCCGCAAGACGCGGCTCCGGCTGCCCCTCTGCCGTGGGTTGTGAGCGCACCGCACACCCGTGCT
CTGCGTGCAACTGCCGCGACCCTGGCAGTTCATTTGGATACGACGCCGGCAGCTCCAGCGGACGTCGCGCA
CACGCTGCTGACCGCTCGCCCGGACCGTCACCGTGCCGCGGTTGTGGGCGCAGACCGCGCGACGCTGACCG
ACGGCCTGCGTGCGCTGGCGACTGGTGGTGATGCACCGCACCTGGTGCATGGCACCGCGACCGGTAGCCCA
CGCCCTGTATTCGTGTTTCCTGGCCAAGGCTCCCAGTGGCCTGGTATGGCCGCAGAACTGCTGGAAACCTC
TGAGCCTTTTCATGACAGCGTTCACGCTTGCGCGGACGCCCTGGCAGAGTTCGTCGATTGGAGCGTTCTGG
ACGTTCTGCGCCAAGCTCCGGACGCACCGCCGTTGCGTCGCGTCGACGTTTTGCAGCCGACTTTGTGGGCG
ACGATGGTCAGCCTGGCGGAAGTTTGGCGCAGCTATGGTGTAGAGCCGGCAGCGGTGGTCGGTCATTGTTG
CGGTGAGATCGCGGCTGCACAAGTGGCGGGTGCGCTGGACATGCGTGATGCGGCTCGCCTGCTGGCGCACC
GTAGCCGTGCATGGTTGCGTTTGGTGGGTAAGGGCACGGTTATTTCCGTGGCGACCAGCGGTCAGGACATC
ACCCGTCGTATGGCTGCGTGGCCGGACAGCGTGGAGCTGGCGGCTCTGAACGGCCCACGTAGCGTTGCTCT
GGCGGGTCCGCCGGACGTGCTGGATGGTATCGTCAACGACCTGACGGATCAGGGTATCCACGCGAAGCGTA
TTCCGGGCGTCGATACGGTTGGCCACTGTAGCCAGGTGGAAGTACTGCGTGATCACCTGTTGGATGTCCTG
CGTCCGGTGAGCCCGCGTCCTGCCGCAGTCCCGTTTTACAGCACGGTTGATGGTACGGAACGTGACACCAC
GACCCTGGACACCGACTACTGGTACTTGAACACGCGCAGCCAAGTCCGCTTTCACCAAGCAGTTCGCAATC
TGCTGGCAGCGGGTCATCGTAGCTTTGTGGAAGTCAGCCCTCACCCGCTGCTGGGTGCCAGCATTGAGGAT
ACGGCAGCGGAGTTTGGCCTGGACGATGTCGCAGCCGTTGGTACGCTGCGTCGCGGTCAGGGTGGCACCCG
CCGTGTGCTGACCAGCGTTGCTGAGGCCTATGTGCACGGCATTGATATTGATTTCACGCCGGCATTCACCG
GCACCACCCCGAATCGTATTGATCTGCCGACGGTCGAGGATCATGGCATTGAGGGCCACGGCGATGATGGC
GGTGAAACCTGGACCGACCGTGTTCGTACCCTGCCGGATGAGCAGCGCGAAGAGGCCCTGCTGGACTTGGT
GTGCCGTACTGTTGCAGCGGTGCTGGAAGCGGACCCAGCCGGTACCGCAGACGCTGTTGCCCCGGACACCG
CGTTCAAAGAAATGGGCTTGGGTTCTCTGTCCGCGGTGCGCCTGCGCAATGGTCTGCGCGAGGCGACGGGT
GCCCATCTGCCGGCAACCATCGCTTACGACCATCCGACCCCGGCAGCCCTGGCGCGTCACCTGGCAATGAC
GCTGTTCGATGCGACGGGCGCAGCGCCAGCGGTTCCGGCACCGAGCCGCGACGACGAGCCGATTGACGCAG
AAACCGCAGTTCTGACTGCCCTGGAACGCGCTGACGAAGCTCTGGAACGTCTGCGTGCTCCGCACGCACGT
ACTCCACGTCAAGAAACCGGCCGTCGTATCGATGAGTTGCTGCGTTCCCTGACGGACAAGGCTCGCCGCAT
GCGCCAGGCGGACGCCGTGGATGACGTGGATGATCCAGCGACCGATCGTTTCGCTGCGGCAACCGACGACG
AAATGTTCGAGCTGCTGGAGAAACGTTTTGGCATTAGCTAA

### 7.1.3. MBP-BorA2

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTACCGCGAGCAGCGCAGCGCAGGATGGCGGTTGGCGTTA
TCGTATTCACTGGCGTCGCCTGGGTACCCGCGATAGCGGCGACCGCCTGAGCGGTCGTTGGTTGTTGTTGG
TCCCGGAAAGCGACGGCACCGAGCCGTGGGTAGAAGGCGCAGAGAAGATGTTGGCCGAGCGTGGTTGCGAG

GTGGTTCACGTTCCGATCGCGGCCACGGCGGACCGCGACGCAATGGTTGGTGCCGTGCGCGAGTCCGTTGA
AGATGGCCGTGTCGACGGTGTTCTGAGCCTGCTGGCCCTGGATGGTCGCCCGCACCCGGATGCTGCGGCAG
TGCCGACCGGTCTGGTTGCGACCGCCCAGGTTGTGCAAGTCAGCGACGAACTGGGTATTGGTCCGTTGTGG
GTGGCGACCCGTCAGGCAGTTAGCGTTGATGGCGCGGATGAGGCGGACGGTGCTGGCCGCACCCGTAAAGC
GGATGATCCGGCGGATGTCGCCCAGGCAGCGGTGTGGGGTTTGGGTCGCGTCGCGGCTCTGGAGAAGCCGC
GTCTGTGGGGTGGCCTGGTGGATCTGCCTGCGCGTGCGGATGAGCGTATGCGTGACCTGGTCGCGCAAGCG
CTGACTGCGCCGGATGCAGAGGACCAACTGGCGGTCCGTGCGGACGGTATTGCCGTGCGTCGCTTGGTCCG
CTCCGCGGCAAGCGCACCGGCCGATGACTGGCAACCGAGCGGCACCGTTCTGGTGACGGGTGGCACGGGCG
GTGTTGGTGCGAATGTCGCGCGTTGGCTGGTCACGCAGGATATTCAGCATCTGCTGTTGGTTAGCCGCCGT
GGTCCGGACGCACCGGGTGCTGCGGAGCTGCTGGCGGAACTGTCGGCGTCGGGCACGAGCGTCACGATCGA
GCCGTGCGACGTGACCGACGCGGATGCGGTCCGTCGTCTGATCGGTGCGGTCCCAGCGGAACGTCCGCTGA
GCACGGTTGTGCACGCAGCCGGTGTGCTGGACGACTGTCTGATCGATGCCCTGACCCCGCAACGCCTGGCT
GCGGCGCTGGAAGTTAAAGCAAAGGGTGCCCTGAATCTGCATGAGGCGGCTGGTGAGGCCCACTTGGTCCT
GTTTAGCAGCCTGGCTGGTACGACCGGTACGAAAGGTCAAGGCAACTACGCAGCCGCAAACGCGTACCTGG
ATGCACTGGCTGAACGCCGTCGTGCAGACGGCCTGCCGGCCACCAGCGTCGCATGGGGCGCATGGCAAGGT
GCCGGTATGGTGGCCGACGCAGCGGTTGCACACCGTACCCGTCGTTACGGTCTGCCACTGATGAGCCCGGA
TCGTGCTGTCGCGACTCTGCGCCAAGTGATGGCAGAGCCGGTGGCTACCCAGGTCGTTGCAGATGTCGACT
GGCAGCGTTTCGTCGCAGATTTTACCGCAGTGCGTCCGAGCCGTCTGCTGGCCGATCTGCCGGAAGTGCGC
TCTCTGGGCGAGCAACGTAAAGACGGTCCTGGCGGCCAAGGTGAAGAGGACGGTCTGGCGAGCAAACTGGC
TGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGG
GCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCGCA
GTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCACCC
GACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACACGG
CTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCG
CGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGA
CTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.4. ACP0 monodomain

atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatgCGTACCCT
GCCGGATGAGCAGCGCGAAGAGGCCCTGCTGGACTTGGTGTGCCGTACTGTTGCAGCGGTGCTGGAAGCGG
ACCCAGCCGGTACCGCAGACGCTGTTGCCCCGGACACCGCGTTCAAAGAAATGGGCTTGGGTTCTCTGTCC
GCGGTGCGCCTGCGCAATGGTCTGCGCGAGGCGACGGGTGCCCATCTGCCGGCAACCATCGCTTACGACCA
TCCGACCCCGGCAGCCCTGGCGCGTCACCTGGCAATGACGCTGTTCGATGCGACGGGCGCAGCGCCAGCGG
TTCCGGCACCGAGCCGCGACGACGAGCCGATTGACGCAGAAACCGCAGTTCTGACTGCCCTGGAACGCGCT
GACGAAGCTCTGGAACGTCTGCGTGCTCCGCACGCACGTACTCCACGTCAAGAAACCGGCCGTCGTATCGA
TGAGTTGCTGCGTTCCCTGACGGACAAGGCTCGCCGCATGCGCCAGGCGGACGCCGTGGATGACGTGGATG
ATCCAGCGACCGATCGTTTCGCTGCGGCAACCGACGACGAAATGTTCGAGCTGCTGGAGAAACGTTTTGGC
ATTAGCtaa

### 7.1.5. MBP-BorMod5

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGACCAGCGTGGGTCGCCGCCAACTGCATGTTC
ACCCGGAAGACCACGATCACCATGCCAAACGCACAGGTCTGAGTATGGGTGACGAACAGAAACTGCGCACC
TACCTGCGTCGTGTGACAGCCGATTTAGCCGACGTTACCGAACGCCTGCAGCGTGCCGAAGACAAAAATGC
AGAGCCGATTGCCATTGTGGGCATGGGTTGTCGCTATCCTGGCGGTGTTCGTAGCCCGGAGGAGTTTTGGA
ATCTGTTAGATGAGGGCGTGGACGCAGTTGCCGGCTTTCCTGAGGATCGTGGCTGGGACCTGGAGAATCTG
TACGATCCGGACCCGGACGAACCGGGTAAATGCTATGCACGCGAGGGCGGTTTTCTGTATGACGCCGGTGA
GTTCGATGCCGCCTTCTTTGGTATCAGCCCTCGCGAAGCACTGAGCATGGACCCTCAACAGCGCCTGCTGC
TGGAGTGCAGTTGGAGTGCACTGGAGCGTGCAGGTATTGACCCGGGCAGCTTACGCGGCAAGGACGTTGGC
GTTTATGTGGGTGCATGGAACAGCAACTATGGTCGCGGTGGCGGCGCCGAAAGCAGTGAGGGCCACCTGCT
GACAGGCAACGCCAGCAGTGTTGTTAGCGGTCGCGTTGCCTACGTTCTGGGTCTGGAAGGCCCGGCAGTTA
CCATTGACACCGCCTGTAGTAGTAGCCTGGTTGGCCTGCACCTGGCCGCACAAGCACTGCGTAGCGGTGAA
TGTGGCCTGGCATTAGCAGGTGGTGTGACAGTTATGAGTACCCCGCTGAGTCTGGTTAGCTTTAGCCGCCA
ACGTGGTCTGGCCCAAGACGGCCGTAGTAAGGCCTTTAGTGCAGATGCCGACGGTATGGGCATGGCAGAAG
GCGTGGGTGTGCTGGTGCTGGAGCGTCTGAGTGAAGCACGTCGCAACGGCCACGAAGTGCTGGCAGTGCTG
CGTAGTAGCGCCGTTAATCAGGATGGTGCCAGCAACGGTCTGAGCGCCCCTAATGGTCCTGCACAGCAGCG
TGTTATCCAGAGTGCCCTGACAGTGGGTCGCCTGGCACCTAGTGATATTGACGTGGTTGAGGCACACGGTA
CCGGTACCGCACTGGGCGACCCGATCGAGGCACAGGCCCTGTTAGCCACCTATGGCCGTGGTCGCGACGCA
GACCGTCCTTTATGGCTGGGTAGCGTGAAGAGCAATATCGGTCACACCCAAGCAGCAGCCGGTGTGGCAGG
CGTTATTAAAATGGTGCTGGCCCTGCGTAAAGGTGTTTTACCGCGCACCCTGCATGTGGATGAGCCGACAG
GTGAAGTTGATTGGGACAGCGGCGCAGTTCGTCTGTTAACCGAAGCACGCGAATGGCCTAGCGGCGAAGGT
CGTGTGCGTCGTGCCGGTGTGAGTAGTTTTGGCATTAGCGGTACCAATGCACACGTTATCGTTGAAGAGGC
ACCGGAAGAAGAGCCTCGTCCGGAGGCACCGAGCGTTGATGTGGTGCCTTGGGTGTTAAGCGCCCGTAGTG
CAGAGGCATTACGCGAACAAGCCGCACGCCTGGCCAGTGTGGCCGGCGGTCTGAATGTGGTGGATGTTGGC
TGGAGTCTGGCAAGCACCCGTGCCGCCTTTGAACACCGTGCCGTGGTGGTGGGTCGTGATAGCGAGGAATT
AGTGAGCGGTCTGAGCAGCGTTAGTGGCGTTGAAGTTGGCGTGGGCGTGGGTGCAGGTGGCGGCGTGGTTT
TAGTTTTCGCCGGTCAGGGCTGCCAGTGGGTTGGCATGGGTCGTGAGCTGCTGGGTAGCAGTTTAGTGTTC
GCCGAGAGTATGCGCGAGTGCGCAGCCGCATTAAGTCCTTTCGTGGACTTCAGTGTTGTGGACGTTTTAGG
CAGCGCAGGTGAACTGGGCCGTGTGGAAGTGGTGCAGCCGGCACTGTGGGCCGTTATGGTGAGCCTGGCCC
GTGTTTGGCGTAGCTGGGGTGTTCCGGTTGCCGCCGTTGTGGGTCATAGCCAGGGCGAAATCGCCGCAGCA
ACCGTTGCAGGCGCATTAAGTGTTGGTGACGCCGCCCGCGTTGTTGCACTGCGTAGCCGCCTGATTGCCGA
ACGCCTGAGCGGTTTAGGCGGCATGGTTAGTGTTGCCCTGAGTCGCGAGCGTGTGGTGAGTCTGATTGCAG
GTGTGCCGGGTGTGAGCGTGGCAGCAGTGAACGGTAGCAGTAGCACAGTTGTGAGTGGCGAGGCAGCAGGT
CTGGAACGCGTTTTAGCAGCCTGTGTTAGTAGCGGCGTTCGTGCACGCCGCATCGATGTGGATTACGCCAG
CCATAGCGTTCAGGTTGAGCTGATTCGTGAGGAACTGCTGGGTGTGTTAGACGGCATCGTGCCGCGCAGCG
GTGAGATTCCGTTTGTGAGCACCGTGACCGGTGAGCGTATTGACACAGTGGAGCTGGGTGCAGAGTACTGG
TACCGTAACCTGCGCCAGACCGTTGAGTTCCAGGCCAGCGTTCAGACACTGCTGGCACAGGGCCATCAGGT
TTTCCTGGAAAGTAGCCCGCACCCGGTGTTAACCGTGGGCATCGAAGAAACCGTTCATGAGAGTGCCGCAC
AGGCAGTGGTTCTGGGCAGCCTGCGTCGTGATGAAGGCGCCTTAACACGTCTGGTGACCAGTGCCGGCGAA
GCCTGGGCACGTGGCGTTCCTGTGGATTGGGCCGGCATGTTAGCAGGTGGCCGCCGTGTGGAACTGCCTAC
CTACCCGTTTTTACGCGAACGTCTGTGGTTAGAACCGAGTCGCAGCCGTACAGGCAATTTAAACATGGCAG
GCCTGGTGGAAGCCGGCCATGAAATTTTACCGGCAGCCGTTGAGCTGCCGGGCGAACAATGGGTGTGGACC

GGCGAATTAAGTCTGAGCGCATACCCGTGGCTGGCAGACCATCAAGTGCTGGGTCAAACCCTGGTGCCTGG
CGTTGCATGGGTGGAGCTGGCATTACACGCAGGCCATCAGCTGGGCTTCGGTAGCGTGGAAGAACTGACAC
TGCAGGCACCTCTGGTGCTGGGTGAGAGCGACGCAGTGCAAGTTCGTGTGGTTGTGAGTGACCTGGGCGAA
AGCGATCGTCGTGCCGTGAGCGTTCACAGTCGCGGCGACGATCAGACCTGGGTGACACACGCCGAAGGCTT
TCTGACCGCAAAAGGCGCCCAGCCGGAAACAATGGCCGTGTGGCCGCCTAGTGGCGCAGAGCCTGTGGAAG
CCGATGGCTTTTATGAACGTCTGGCCGATGCCGGTTACCATTACGGTCCGGTTTTCCAGGGTGTGAGCAAG
GTGTGGCGTGCCGGTGAAGAAATATACGCCGAGGTTGGTCTGCTGGACGACGCAGATGTTGACGGTTTCGG
CATCCATCCGGCCTTACTGGATGCCGCATTACAAACCGCGTACGTGGCCCAGCGTGGCCCTGCAGAGACCA
AATTACCGTTCGCCTTTGGCGATGTGCAACTGTTCGCCACCGGCGCCCGTAGTTTACGCGTTCGCGTTAGC
CCTGCAGCCCAGCAAGGCATGGCATGGGAAGCCTGGGACCCGACCGGTCTGCCTGTGTTTAGCCTGGGTTA
CCTGGCAACACGTCCGGTTGATCGTGGTCAGCTGACCGTGAAACGCCCGGAGAGTTTATTTAAAGTGGCAT
GGGACGAGACAGTTCCGGTTGTTGGTAATGCCACCGCCGCACATGGTGTTGTGCTGGGTGATGATCCTTTT
GCCCTGGGCGCAGCATTACGCGCCGCAGGTTGGGAAGTGGGTGCAGCCCCGGAGCCTGCAAGCGCAGATAC
CGCCGCCGAAGTTTTATTACTGCCGTGTACCGCACCGGGCGAGCCTGACGCGGACCTGCCTACAGCAGTGC
GCGCAGTTACCGCCCGTGTGCTGGGTGTGCTGCAAGAATGGCTGGCAGATGAGCGTCTGGCCGGTACCCGT
CTGGCCGTTGTGACACGTAATGCCCTGCCGGGTGACTTACTGCATAGCCCTGTTTGGGGCCTGGTGCGTAG
CGCACAGACCGAAAATCCGGGCCGCATCACATTAGTTGATTTAGATGATCATCCTGATAGCGCAGCCGTTT
TAGCAGAAGCAGTTCAAAGTGATGAGCCGCGTATTATGGTGCGCGAAGGCCGTCCTACAGCCGCCCGTCTG
GTGCGTGCCACCGCACCTGAATTAGTGCCGCCGGCAGGTGCCGATGCATGGCGCCTGGAGATTACAGAACC
GGGCACCTTTGACAACCTGACCCTGGGCGTGTACCCGCATGCCGAGAAGACCCTGGCCGATAATGAGGTGC
GCGTGGCAGTGCACGCAGGCGGTCTGAACTTCCACGATGTGGTTGCCGCACTGGGCATGGTTGAAGATGAC
CTGACCTTAGGTCGCGAGGCAGCCGGCGTTGTTGTTGAAGTGGGTGACGCAGTGCCTGACCTGACCCCGGG
TGATCACGTGATGGGTATTCTGAGCAGCGGTTTCGGCCCTCTGGCAGTTACCGACCACCGCTATCTGGCAC
GTATGCCGGAAGGTTGGACATTCGCACAGGCCGCAAGCGTTCCGGCCGCCTTCTTAACAGCCTATTATGGT
CTGTGCGACCTGGGTGGCATTCGCGCAGGTGACCGCGTGTTAATTCACGCAGCCGCCGGTGGCGTTGGCAT
GGCAGCAGTGCAGATCGCCCGCCACCTGGGTGCAGAAGTTTTTGGTACCGCCAGCCCGCGTAAGTGGGGTG
CATTACGTGCACTGGGCCTGGACGACGCCCATCTGAGCAGTAGCCGTACCTTAGACTTCGAGCAGGAGTTT
CTGGATGCAACCGATGGTCGCGGCGTTGATCTGGTGCTGAATAGCCTGGCCCGTGAGTTCGTTGACGCAAG
TCTGCGCCTGATGCCGGGCGGTGGTCGCTTTGTGGATATGGGCAAGACAGACATCCGTCGTCCGGAACAGG
TTGCAGAAGATCATGGTGGTGTGGCCTACCAGGCATTTGATCTGGTTGAAGCAGGCCCTCAGCGCACAGGT
GAGATGCTGGCAGAGATTGTTCGCCTGTTCCAGGCAGGCGCATTTCGCCCGTTACCGATTACCCAGTGGGA
TGTGCGCCGCGCACCTGAGGCATTTCGTCACATCAGTCAGGCCAAACACATTGGCAAGATCGTGTTAACAG
TTCCGCGCCCGATCGACACCGACGGTACCGTGATGGTTACCGGCGCCACAGGCACACTGGGTGGTTTCGTT
GCCCGTCATCTGGTGACACACCACGGTATTCGCCGTTTACTGTTAGTGAGCCGCAGTGCAGAGCGTACCGA
CCTGGTGCGCGAACTGACAGAATTAGGCGCCGATGTTACATGGGCAAGCTGTGATCTGGCAGACGCCACAG
CCGTTGAAGAAACAGTGCGTAGCGTGGATGAGCGCCACCCGTTAGTTGCCGTGGTGCATAGCGCCGGTGTG
CTGGATGACGGTGTGATTGACAAGCAGAGTCCGGAACGTCTGGACACCGTGATGCGCCCTAAGGTGGATGC
AGCATGGAACTTACACCGTCTGCTGGATAACGCACCGTTAGCCGACTTCGTTCTGTTTAGCAGCGCCAGCG
GCGTGCTGGGCGGCGCCGGCCAAAGCAATTACGCAGCCGCCAACGCCTTTCTGGACGCACTGGCCGAGCAC
CGTCGCGCACAAGGCCTGGCAGGTCAAGCACTGGCATGGGGCTTATGGAGCGATCGTAGCACCATGACAGG
CCAGCTGGGTAGTACCGAACTGGCCCGTATCGCACGTAACGGTGTTGCAGAAATGAGTGAAACCGAAGGTC
TGGCACTGTTTGACGCAGCCCGTGATACAGCAGAGGCCGTTCTGCTGCCGATGCACCTGGATGTTGCCCGC
CTGCGTAGCCGTAATGGTGAGGTGCCTGCCGTGTTTCGTCGTCTGATTCATGCCACCGCACGTCGCACCGC
CAGTACCGCCGTTCGCAGTGCAGGTTTAGAACAGCAGTTAGCCAGTCTGAGCGGCCCTGAACGTACCGAAC
TGTTACTGGGCCTGGTGCGTGATCACGCCGCAGCAGTGTTAGGCCACGGTACCAGTGACGCAGTGAGTCCG
GACCGCCCGTTTCGTGATCTGGGTTTCGACAGTCTGACCGCAGTGGAGCTGCGTAATCGTTTTGCCGCCCT
GACCGGCCTGCGTTTACCGGCCACACTGGTTTTCGACCACCCTAGTCCGACAGCACTGGCAGGTCACCTGG
CCGGCTTATTAGGCGCAGCCACACCTAGCGCCGCCGAACCTGTGCTGGCAGCAGTTGGTCGTCTGCGTGCC
GATTTACGCAGTCTGACACCTGATGCAGAAGGTGCAGAGGATGTGACCATTCAGCTGGAAGCCCTGCTGGC
CGAGTGGCGTGAAGCCGCAGAGAAACGTGCACCGGAAGCCGTGGGCGATGAGGACCTGAGCACAGCCACAG
ACGATGAAATATTCGCCCTGGTGGATAGTGAATTAGGTGAAGCATGA

### 7.1.6. MBP-BorA6ΔTE

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGACCGCAGAAGCAAGCCAAGACAAGCTGCGTG
ATTACCTGCGCAAAACCCTGGCCGATCTGCGCACCACAAAACAGCGCCTGCGCGACACAGAGCGTCGTGCC
ACAGAACCGGTTGCAATCGTTGGTATGAGCTGTCGCTTACCGGGCGATGTGCGCACCCCTGAGCGTTTTTG
GGAACTGCTGGATACAGGCACCGACGCACTGACCCCTTTACCTACCGACCGCGGCTGGAACCTGGATACCG
CCTTCGATGACGAGCGTCCGTATCGCCGCGAGGGTGGCTTCTTATATGATGCCGGTCGCTTTGACGCCGAA
TTTTTCGGCATCAGCCCTCGTGAAGCCCTGGCAATGGACCCGCAGCAGCGCTTATTACTGGAAAGCAGTTG
GGAGGCAATTGAACATGCCCGTATCGATCCTCGCAGTCTGCACGGTAGCCGTACCGGCGTGTGGTTTGGTA
CCATTGGCCAGGATTACTTTAGCCTGTTTGCCGCCAGTGGTGGCGAACATGCCAACTACCTGGCAACCGCC
TGCAGTGCAAGCGTGATGAGTGGCCGTGTTAGCTATGTGCTGGGCCTGGAAGGTCCTGCCGTGACAGTGGA
TACAGCCTGCAGTAGTAGTCTGGTTGCACTGCATAGTGCCGTGCAAGCACTGCGTAGCGGTGAATGTGAAT
TAGCCCTGGCAGGTGGTGCCACCGTGATGGCAACCCCGACAGTGTTCACCGCATTTAGTCACCAACGTGGT
TTAGCAGGCGATGGCCGCTGCAAGGCATTTGCCGCAGGCGCCGATGGCGCCGGTTTTGCCGAAGGTGTTGG
TGTGCTGGTGTTAGAACGCTTAAGCGTGGCCCGTCGTAATGGCCACCGCGTTCTGGCAGTTGTTCGTGGCA
GTGCCGTGAACCAGGATGGTGCAAGCAATGGCCTGACCGCCCCGAATGGTCCTAGTCAGCAGCGTGTGATC
CGTGCAGCATTAGCAAATGCACGCCTGGCACCGGAGGACGTGGACGCAGTTGAGGGCCACGGTACCGGTAC
CAGCCTGGGCGATCCGATCGAGGCCCAGGCACTGCTGGCAACCTACGGTCGCGGTCGTGACGCCGAACGTC
CGCTGTGGTTAGGCAGCGTGAAAAGCAACATCGGTCATGCACAGGCAGCCGCCGGTGTGGCCGGCGTTATC
AAAATGGTGTTAGCCATGGAAAAAGGCCGTTTACCTCGTACCCTGCATGTTGACGAACCGAGCGGTGAGGT
GGATTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGCGATTGGCCTAGCGGTGAAGGTCGTGTGC
GTCGCGCAGGTGTTAGCAGCTTCGGCATCAGTGGCACCAATGCCCACGTGATCATTGAGGAACCTCAAGAA
GAAGAGGCGGCACCTGACAGCAGCGCAAGTGGCGCAGTTCCGTGGGTGCTGAGCGCACGCAGTGCCGAAGC
ACTGCAAGCATTAGCCAGCCAGCTGGCCGACCACAGCGCAAAAAGCAGTCCTGTGGATGTTGGCTGGAGTC
TGGTTAGCACCCGCGCCGCATTTGAACATCGTGCCGTGGTTGTTGGTCGCGGTCGTGATGAATTAGTTCGC
GGTCTGAGCGAAGTTGCCCAGGGCCGCGGTGTTCGCGGCGTGGCAAGTAGTGCAAGTGGCGGCTTAGCATT
TGTGTTTGCAGGTCAGGGCAGTCAGCGTCTGGGCATGGGTCGTGGCCTGTATGAGCGTTTTCCGGTTTTCG
CCGAGGCATTTGATGAAGTGTGTGGCCGCGTTGGTCCTGGCGTGCGTGAAGTTGTGTTTGGCAGTGATGCC
GGCGAACTGGATCGCACCGTTTGGGCACAGGCAGGTCTGTTTGCCCTGGAAGTTGCACTGTTCCGCTTATT
AGAGAGCTGGGGTGTGCGTCCTGGCTGCTTAATCGGCCATAGCGTTGGTGAGTTAAGCGCAGCATGTGTTG
CCGGCCTGTGGAGTCTGGAGGACGCATGCCGCGTGGTGGCAGCCCGCGCCCGCTTAATGCAGGCCTTACCG
GCCGGCGGTGTTATGGTTGCAGTTCGCGCAGAAGCAGGCGAACTGGCAGGTTTCCTGGGCGAGGACGTGGT
GATTGCCAGCGTGAATGCCCCTGGCCAGGTTGTTATCGCAGGTCCTGAGGGCGGTGTTGAACGTGTGGTTG
CCGCCTGTGGTGCCCGTAGTCGTCGTCTGGCCGTTAGTCATGCCTTTCACAGTCCGCTGGTTGAACCGATG
TTAGGTGAGTTTCGCCGCGTTGTGGAAAGCGTGGCATTTGGTGTGCCTAGCTTACGCGTTGTTAGCAACGT
GACCGGTGCATGGGTAGACCCGGAAGAATGGGGCACACCTGAGTATTGGGTTCGCCAGGTTCGTGAACCGG
TTCGCTTCGCAGATGGTGTGGCAACCCTGCTGGATGCCGGCGTTCGTACATTCGTTGAACTGGGCCCGGCC
GGTACATTAACCAGCATGGTGAGCCACTGTGCAGATGCCACAGCCACAAGTGTTACCGCAGTGCCGACATT
ACGTCCGGACCATGATGAAAGCCGCACAGTGCTGAGTGCCGCCGCAAGCTTATACGTGCAAGGTCACCCGG
TGGATTGGGCACCGCTGTTCCCGCGTGCACGTACCGTGGATCTGCCTACCTACCCGTTCCAACACCAACAC
TACTGGATGGAGAGCGCAGCACGCCCGACAGTTGAAGATACCCCTCGTGAGCCGCTGGATGGCTGGACCCA
CCGTATCGATTGGGTGCCTTTAGTGGATGAGGAGCCTGCACCTGTTCTGGCAGGCACCTGGCTGCTGGTTC
GCCCTGAAGAAGGCCCGCGTCCTTTAGCAGATGCCGTGGCAGACGCCCTGACCCGTCATGGCGCAAGCGTG

GTTGAAGCCGCACGCGTGCCTCATCAGAGTGACACAGAGCTGACAGGCGTTGTTAGTCTGCTGGGCCCGGG
TGCCGATGGTGATGGTGGTCTGGATGCAACCCTGCGTCTGGTTCAAGATTTAGCCACCGCAGGCAGCACCG
CCCCGTTATGGATTGTGACCAGCGGCGCAGTTGCCGTTGGCACCAGTGATACAGTGCCGAATCCGGAGCAG
GCAACCTTATGGGGTCTGGCCCGTGCCGCCGCAACCGAGTGGCCGGGTTTAGGTGCCGCACGCATTGATTT
ACCGGCCGACTTAACAGAGCAGGTGGGCCGCCGCCTGTGTGCACGTCTGCTGGATCGTAGTGAACAAGAAA
CAGCCGTGCGTCAGGCCGGTGTGTTCGCACGTCGCCTGGTTCGTGCCCGCACAAGCGATGGTCGCTGGACA
CCTCGTGGTACCGTTCTGGTTACCGGTGGTACCGGCGCACTGGCAGGTCACGTGGCACGTTGGCTGGCCGA
AGAGGGGGCAGAACATATCGTTCTGGCAGGCCGCCGCGGTCCGGATGGTCAAGGCGCAGAAGCATTACGCG
CAGACCTGGTGGCAGCAGGCGTGAAAGCCACAATTGTGCGTTGCGATGTGGCAGACCGCGATGCCGTTCGC
CTGCTGTTAGACGCCCATCGCCCTAGCGCCATCGTTCACACAGCAGGCGTGGTTGATGATGGCCTGTTAAC
CAGCCTGACCCCTGCACAGGTTGAGCGTGTGCTGCGCCCGAAGTTACTGGGCGCCCGCAACCTGCACGAAT
TAACCCGTGACCGCGAACTGGACGCATTCGTGCTGTTTAGTAGCCTGGCAGGCGTTCTGGGTGGTGCCGGC
CAGGCCAATTATGCAGCCGCCAATGCATATCTGGACGCATTAGCCGCACACCGCACAGCCCATGGTCTGCC
TGCCGCAAGTTTAGCATGGGGTCCTTGGGAGGGTGACGGCATGGCCGCAGCCCAAGAAGCAGCAGACCGCC
TGCGCCGTAGTGGTCTGACACCGCTGCCTCCGGAACAAGCCGTGCGTGCCTTAGGCCGTGGCCATGGTCCT
CTGGTGGTTGCCGATGCAGATTGGGCACGTTTAGCAGCCGGCAGTACCCAGCGCCTGCTGGATGAGCTGCC
TGAGGTTCGTGCAGTGCGTCCGGCCGAGCCTGCAGTTGGTCAGCGCCCGGACTTACCGGCACGTTTAGCAG
GTCGTCCGGCCGAAGAACAAAGCGCCGTGTTACTGGAAGCCGTGCGCGAAGAAATCGCCGCAGTTCTGCGC
TATGCAGATCCGGCACGCATTGGCGCCGATCACGAGTTCCTGGCCCTGGGCTTCGATAGCCTGACCAGCAT
CGAATTACGTAACCGCCTGGCCACCCGTATTGGTTTAACACTGCCGGCAACACTGACACTGGAACAGCGCA
CACCTGCCGGCCTGGCAGCCCATTTACGCGAGCGTATTGCCGATCGCCCGGTTggatcctaa

### 7.1.7. MBP-A1

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTCGATTGGGACCGCTTTTTCGCAGGCACCGGTGCTCGTCGCATTGATCTGCCGACGTACGCATTT
CGTCGCCGTTCGTTCCCGTGGATCCAGGCAGCACCGGATGCAGACGTGACCACGGCAGGTCTGGCTGGTCT
GGGTCATCCGCTGCTGGGTGCAAGCCTGGAACTGGCAGATGCCCAAGGTGCAGCTCTGTCGGGCCGTCTGA
GCGCACGTACCGAATCCTGGCTGGCAGATCACGTGGTTCTGGGTTCAACCCTGGTTCCGGGTACGGCAGTC

GTGGAAATGGCAGTCCGTGCAGGTGCTGAAACCGGTTGCGGTCGCCTGGCAGAACTGACGCAGGAAGCGCC
GCTGGCAGTTCCGGAACGTGGTGCAGTCCATCTGCAGGTTCGTGTTGGTCCGGCAGGTGAACAAGGTCACC
GTCCGGTCGGTGTGTATTCCCGTCCGGAAGACGCCGAACCGGATGAACCGTGGGCATGTCATGCTCGTGGC
GTGCTGGCACCGGAAGCAGCACCGGTTCCGGCAGGTACCGGCGGTGCTTGGCCGCCGTCAGGTGCAGAACC
GGTGCCGCTGGATGGCTTTTATGAACGCCTGGCAGCTGAAGGTTTTGCGTACGGCCCGGCATTCCAGGGTC
TGACCCGCGCCTGGCGTCTGGGCGATGAAGTTCTGGCAGAAATTACGCTGCCGGAAGGTGCATGCAGCGGT
GCTGACCGTTACGGTGTGCATCCGGCACTGCTGGATGCAGCACTGCACACCGCACTGCTGAAAGAAGAAGC
TTCAGATACGTCGCAGGTTCGCATCCCGTTTGCCTGGCATGAAGTCTCGTTCCACGGCGGTAGCGCACCGG
TGCTGCGTGCACGTCTGACCCCGTCTGGTACCGACACCGTTAGTCTGGCCCTGTGGGATGAACATGGTACG
CCGGTGGCAAGCGTTGGCTCTCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTGCGTGCTACCCGTACGCA
CGACACCCTGTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGCACGTTGCGCAGTCCTGG
GTGATGACGAACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCGCACTGGAAAGTGCCGAT
CCGGTGCCGGAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGACCGTGCAGACGCTGCACG
TAGCCTGACCGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACGCTGGGCAACCACGCGTC
TGGCTCTGGTGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGGATCTGCCGGCAGCAGCT
GTCTGGGGTCTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTGGCTGATCTGGACGGTGA
TACCGCATCGGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACCGCAGCTGGCACTGCGTG
AAGGTGCAGTTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGGTGCCGCCGCCGGGTACG
CGCGATTGGCACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTGACCCCGTTTCCGGAAGC
AGCAGCTCCGCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCTGAACTTCCGTGATGTTG
TCATGGCACTGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAGGTATCGTGACCGAAGCA
GGTCCGGGTGTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGATGGTTGCATCGGTCCGGT
TGCCGTGGTTGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCCGCAGGCAGCTTCAGTGC
CGGTTACCTTCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTCCGGGTGATCGTGTCCTG
GTGCATGCAGCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCACCTGGGTGCTGAAGTTTT
TGCAACCGCTGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGACGATCATCTGGCGAGCT
CTCGCACCGATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTATCGATGTCGTGCTGAAT
TCTCTGGCCGGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGCCGTTTTATTGAAATGGG
CAAAACCGATATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGTCTACCGTGCCTTCGACC
TGATGGATGGTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTCTGTTCGAAGGTGGCAAA
ATTCAGCTGGTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTTCGCTATTTCGCACAAGC
TCGTCATGTGGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGGTACCGTGCTGGTTACGG
GTGCATCCGGTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATGATGTGCGTCACCTGCTG
CTGGCATCACGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAACTGCGTGAATCTGGTGC
ACATAGTGTGCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGACCTGCTGGCGTCTATTC
CGGATGAACATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATGGCGTGCTGGAAACCATG
ACGCCGGAACGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCATCTGCACGAACTGACCCG
CGACCGTGATCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGGTAGTGCAGCTCAGGCAA
ACTACGCGGCCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATCACGGTCTGCCGGCAACC
TCTCTGGCTTGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGTCGTGCAGACCTGGATCG
TATGCAACGCTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTTTGACGCTGCACTGGCAG
CAGGTCGTCCGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCGCTGCGGGTGGCGGTTCT
CTGCCGGCACCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCGCTGCCGGCTGCGGATGC
GCTGCGCGGTCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAAC
TGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGT
TTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGAC
TCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTA
CCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTC
GGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCG
CTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGG
GCATCAGCCGCTAA

### 7.1.8. MBP-A2

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTGCAAGCCTGGAACTGGCAGATGCCCAAGGTGCAGC
TCTGTCGGGCCGTCTGAGCGCACGTACCGAATCCTGGCTGGCAGATCACGTGGTTCTGGGTTCAACCCTGG
TTCCGGGTACGGCAGTCGTGGAAATGGCAGTCCGTGCAGGTGCTGAAACCGGTTGCGGTCGCCTGGCAGAA

CTGACGCAGGAAGCGCCGCTGGCAGTTCCGGAACGTGGTGCAGTCCATCTGCAGGTTCGTGTTGGTCCGGC
AGGTGAACAAGGTCACCGTCCGGTCGGTGTGTATTCCCGTCCGGAAGACGCCGAACCGGATGAACCGTGGG
CATGTCATGCTCGTGGCGTGCTGGCACCGGAAGCAGCACCGGTTCCGGCAGGTACCGGCGGTGCTTGGCCG
CCGTCAGGTGCAGAACCGGTGCCGCTGGATGGCTTTTATGAACGCCTGGCAGCTGAAGGTTTTGCGTACGG
CCCGGCATTCCAGGGTCTGACCCGCGCCTGGCGTCTGGGCGATGAAGTTCTGGCAGAAATTACGCTGCCGG
AAGGTGCATGCAGCGGTGCTGACCGTTACGGTGTGCATCCGGCACTGCTGGATGCAGCACTGCACACCGCA
CTGCTGAAAGAAGAAGCTTCAGATACGTCGCAGGTTCGCATCCCGTTTGCCTGGCATGAAGTCTCGTTCCA
CGGCGGTAGCGCACCGGTGCTGCGTGCACGTCTGACCCCGTCTGGTACCGACACCGTTAGTCTGGCCCTGT
GGGATGAACATGGTACGCCGGTGGCAAGCGTTGGCTCTCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTG
CGTGCTACCCGTACGCACGACACCCTGTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGC
ACGTTGCGCAGTCCTGGGTGATGACGAACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCG
CACTGGAAAGTGCCGATCCGGTGCCGGAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGAC
CGTGCAGACGCTGCACGTAGCCTGACCGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACG
CTGGGCAACCACGCGTCTGGCTCTGGTGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGG
ATCTGCCGGCAGCAGCTGTCTGGGGTCTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTG
GCTGATCTGGACGGTGATACCGCATCGGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACC
GCAGCTGGCACTGCGTGAAGGTGCAGTTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGG
TGCCGCCGCCGGGTACGCGCGATTGGCACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTG
ACCCCGTTTCCGGAAGCAGCAGCTCCGCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCT
GAACTTCCGTGATGTTGTCATGGCACTGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAG
GTATCGTGACCGAAGCAGGTCCGGGTGTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGAT
GGTTGCATCGGTCCGGTTGCCGTGGTTGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCC
GCAGGCAGCTTCAGTGCCGGTTACCTTCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTC
CGGGTGATCGTGTCCTGGTGCATGCAGCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCAC
CTGGGTGCTGAAGTTTTTGCAACCGCTGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGA
CGATCATCTGGCGAGCTCTCGCACCGATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTA
TCGATGTCGTGCTGAATTCTCTGGCCGGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGC
CGTTTTATTGAAATGGGCAAAACCGATATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGT
CTACCGTGCCTTCGACCTGATGGATGGTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTC
TGTTCGAAGGTGGCAAAATTCAGCTGGTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTT
CGCTATTTCGCACAAGCTCGTCATGTGGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGG
TACCGTGCTGGTTACGGGTGCATCCGGTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATG
ATGTGCGTCACCTGCTGCTGGCATCACGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAA
CTGCGTGAATCTGGTGCACATAGTGTGCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGA
CCTGCTGGCGTCTATTCCGGATGAACATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATG
GCGTGCTGGAAACCATGACGCCGGAACGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCAT
CTGCACGAACTGACCCGCGACCGTGATCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGG
TAGTGCAGCTCAGGCAAACTACGCGGCCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATC
ACGGTCTGCCGGCAACCTCTCTGGCTTGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGT
CGTGCAGACCTGGATCGTATGCAACGCTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTT
TGACGCTGCACTGGCAGCAGGTCGTCCGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCG
CTGCGGGTGGCGGTTCTCTGCCGGCACCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCG
CTGCCGGCTGCGGATGCGCTGCGCGGTCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTT
GGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCA
GCTTCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGT
ATGAAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACA
ACTGGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAG
CCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAG
ACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTT
CCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.9. MBP-I1

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTGGATTGGCCAGCCCTGCTGCCGGGTGGCCGTCGCGTGGAGCTGCCAACGTACGCCTTTCAACGT
CGCCGTTACTGGCTGGAAGACGCCGATCCAACCGGTGGTGACCCAGCGGCGCTGGGCCTGACGGCGGCGGA
CCACCCGCTGCTGGGTGCGGCGGTTCCACTGGCGGAAGATCAGGGCATTGTGATTACCAGTCGCCTGAGTC
TGCGTACCCATCCGTGGCTGGCGGATCACGAAATTGGTGGTACGGTTCTGCTGCCAGGTGCCGGTCTGGTT

GAAATTGCGCTGCGTGCGGGCGATGAAGTGGGTTGCGGCCGTGTTGAAGAACTGACCCTGGAAATTCCGCT
GGTTGTTCCACAGGAAGGTGGTGTTACCGTGCAAATTCGTGTTGGTGCCCCAGATGAAAGTGGTTGGCGTC
CAATGACGGTGCATAGTCGTACCGACCCAGAAGAAGAATGGACCCGCCATGTTAGCGGCGTGCTGAGTCCG
GATGTGCCAACCGAACGTTACGATCTGGGTGCCTGGCCACCAGCCGGTGCGACCCCGGTGGAACTGGATGG
TTTTTACGAAGCCTATGCGCGTCTGGGTTACGCCTACGGTCCAAGTTTTCAAGGCCTGCGTGCCGCGTGGC
GTCGTGGTGATGAAGTTTTCGCCGAAGTGAGTCTGCCAGTGGAAGAACAGGAAACCGCCGGTCGCTTTACG
CTGCATCCAGCCCTGCTGGACGCGGCGCTGCAAAGCGCCGGTGCGGGCGCCTTTTTTGATAGTGGTGGTAG
TATGCGCCTGCCGTTTGCCTGGAGTGGTGTGAGTGTGTTTGCGGCGGGTGCGAGTACCGTTCGTGTGCGTC
TGAGTCCAGCGGGTCCAGATGCCGTTACCGTTGCCCTGGCGGACCCAACCGGTGCCCCGGTTGCGCTGGTT
GAGCGTCTGCTGATTCCAGAAATGAGCCCGGAACAACTGGAACGTGTTCGCGGTGAGGAAAAAGAAGCCCC
ATACGTGCTGGATTGGGTTCCGGTGGAAGTGCCAGCCGATGACCTGGTTCGCCCAGAGCGTTGGACGCTGC
TGGGCGGTGCCGACGCGGGTGTGGGCCTGGATGTTGCGGGTGCCTTCGCCAGTCTGGAACCAAGTGATGGT
GCGCCGGAATTTGTTGTTCTGCCGTGTGTGCCGCCAACGAGTCCAACCCGTGCCGCGGATGTTCGTCAAAG
TACCCTGCAAGCCCTGACCGTTCTGCAAAATTGGGTGACCGATGAGCGCCATGCCGATAGTCGTCTGGTTC
TGGTGACCCGTCGTGCCGTTGGTGTTGGTGCCCATGATGATGTGCCGGATCTGACGCATGCCGCCCTGTGG
GGTCTGGTGCGTAGTGCCCAAACCGAAAATCCAGGTCGTTTTCTGCTGGTTGATCTGGATGAAGGCGCGGA
GCTGGCCGAGGTGCTGCCAGGCGCGCTGGGTAGCGGTGAAAGTCAAGTGGCCGTGCGTGCGGGTCGCGTTC
TGGCGGCCCGTCTGGCCCGTAGCGGCAGCGGTGGTGCGGAACTGGTGCCACCAGCCGGCGCCCCATGGCGT
CTGGATACGACCAGTCCAGGTACGCTGGAAAACCTGGCCCTGGTGCCAAGTGCCGAAGAACCACTGGGTCC
ACTGGATGTGCGTGTTAGTGTTCGTGCGGCGGGTCTGAATTTTCGCGATGTTCTGATTGCGCTGGGCATGT
ATCCGGGCGATGCCCGTATGGGTGGTGAAGGTGCGGGTGTTGTGACCGATGTTGGCAGTGAAGTGACGACC
CTGGCCCCAGGTGACCGTGTGATGGGTATGCTGAGTAGTGCCTTTGGTCCAACCGCCGTGAGTGATCATCG
CGCGCTGGTTCGCGTTCCGGACGATTGGAGTTTCGAACAAGCCGCGAGTGTGCCAACCGTTTTCGCCACCG
CCTACTATGGCCTGGTGGATCTGGCCGAACTGCGTGCCGGTCAAAGTGTGCTGGTGCATGCCGCCGCGGGT
GGCGTTGGTATGGCCGCGGTGCAACTGGCGCGTCATCTGGGTGCGGAAGTGTTTGGTACGGCCAGTACGGG
TAAATGGGATAGCCTGCGCGCCGGTGGTCTGGACGCCGAACATATTGCCAGTAGTCGTACGGTTGAATTTG
AAGAAACCTTTCTGGCGGCGACGGCCGGTCGTGGTGTTGATGTGGTGCTGGACAGCCTGGCCGGTGAGTTC
GTTGATGCCAGTCTGCGCCTGCTGCCGCGTGGTGGTCGCTTTGTGGAAATGGGTAAAGCCGATATCCGTGA
TGCCGAACGTGTGGCGGCCGATCATCCAGGTGTGACCTACCGTAGTTTTGATCTGCTGGAAGCCGGCCTGG
ATCGTTTTCAAGAAATTCTGACCGAGGTGGTGCGTCTGTTTGAACGTGGCGTGCTGCGCCATCTGCCAGTG
ACCGCCTGGGATGTTCGCCGTGCCGCGGAGGCCTTTCGTTTTGTGAGTCAAGCCCGTCATGTGGGCAAAAA
CGTTCTGGTTATGCCACGTGTGTGGGACCGTGACGGTACGGTGCTGATTACCGGTGGCACGGGTGCCCTGG
GCGCGCTGGTGGCCCGTCATCTGGTGGCCGAACATGGTATGCGTAATGTTCTGCTGGCCGGTCGCCGTGGC
GTTGATGCCCCAGGTGCCCGTGAACTGCTGGCCGAGCTGGAAACGGCGGGTGCCCAAGTGAGCGTTGTTGC
GTGTGATGTTGCGGACCGTGATGCCGTTGCCGAACTGATTGCCAAGGTGCCAGTGGAACATCCACTGACGG
CGGTTGTTCACACCGCGGGTGTTGTGGCGGATGCCACCCTGACCGCGCTGGACGCCGAACGTGTGGATACG
GTTCTGCGCGCGAAAGTTGACGCGGTGCTGCATCTGCATGAGGCCACCCGTGGCCTGGACCTGGCCGGCTT
TGTGCTGTTTAGTAGTGCGAGTGGTATTTTTGGTAGTCCAGGTCAAGGTAATTATGCGGCGGCGAATAGTT
TTATCGACGCCTTTGCCCATCATCGTCGCGCCCAAGGCCTGCCAGCCCTGAGTCTGGCCTGGGGTCTGTGG
GCCCGTACCAGTGGCATGGCGGGTCAGCTGGGTCACGACGACGTGGCCCGCATTAGTCGCACGGGCCTGGC
CCCAATTACCGATGATCAGGGTATGGCCCTGCTGGATGCGGCCCTGGGTGCGGGTCGTCCGCTGCTGGTTC
CGGTTCGTCTGGATCGTGCCGCCCTGCGTAGCCAAGCGACGGCCGGTACGCTGCCACCAATCCTGCGTGGT
CTGGTTCGCGCCACCGTGCGTCGTGCGGCGAGTACGGCCGCGGCGCAGGGTCCAAGCCTGGCGGAGCGTCT
GGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGC
TGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACC
GCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCA
CCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACA
CGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAA
GCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCG
CGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.10. MBP-I2

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTGCGGCGGTTCCACTGGCGGAAGATCAGGGCATTGT
GATTACCAGTCGCCTGAGTCTGCGTACCCATCCGTGGCTGGCGGATCACGAAATTGGTGGTACGGTTCTGC
TGCCAGGTGCCGGTCTGGTTGAAATTGCGCTGCGTGCGGGCGATGAAGTGGGTTGCGGCCGTGTTGAAGAA

CTGACCCTGGAAATTCCGCTGGTTGTTCCACAGGAAGGTGGTGTTACCGTGCAAATTCGTGTTGGTGCCCC
AGATGAAAGTGGTTGGCGTCCAATGACGGTGCATAGTCGTACCGACCCAGAAGAAGAATGGACCCGCCATG
TTAGCGGCGTGCTGAGTCCGGATGTGCCAACCGAACGTTACGATCTGGGTGCCTGGCCACCAGCCGGTGCG
ACCCCGGTGGAACTGGATGGTTTTTACGAAGCCTATGCGCGTCTGGGTTACGCCTACGGTCCAAGTTTTCA
AGGCCTGCGTGCCGCGTGGCGTCGTGGTGATGAAGTTTTCGCCGAAGTGAGTCTGCCAGTGGAAGAACAGG
AAACCGCCGGTCGCTTTACGCTGCATCCAGCCCTGCTGGACGCGGCGCTGCAAAGCGCCGGTGCGGGCGCC
TTTTTTGATAGTGGTGGTAGTATGCGCCTGCCGTTTGCCTGGAGTGGTGTGAGTGTGTTTGCGGCGGGTGC
GAGTACCGTTCGTGTGCGTCTGAGTCCAGCGGGTCCAGATGCCGTTACCGTTGCCCTGGCGGACCCAACCG
GTGCCCCGGTTGCGCTGGTTGAGCGTCTGCTGATTCCAGAAATGAGCCCGGAACAACTGGAACGTGTTCGC
GGTGAGGAAAAAGAAGCCCCATACGTGCTGGATTGGGTTCCGGTGGAAGTGCCAGCCGATGACCTGGTTCG
CCCAGAGCGTTGGACGCTGCTGGGCGGTGCCGACGCGGGTGTGGGCCTGGATGTTGCGGGTGCCTTCGCCA
GTCTGGAACCAAGTGATGGTGCGCCGGAATTTGTTGTTCTGCCGTGTGTGCCGCCAACGAGTCCAACCCGT
GCCGCGGATGTTCGTCAAAGTACCCTGCAAGCCCTGACCGTTCTGCAAAATTGGGTGACCGATGAGCGCCA
TGCCGATAGTCGTCTGGTTCTGGTGACCCGTCGTGCCGTTGGTGTTGGTGCCCATGATGATGTGCCGGATC
TGACGCATGCCGCCCTGTGGGGTCTGGTGCGTAGTGCCCAAACCGAAAATCCAGGTCGTTTTCTGCTGGTT
GATCTGGATGAAGGCGCGGAGCTGGCCGAGGTGCTGCCAGGCGCGCTGGGTAGCGGTGAAAGTCAAGTGGC
CGTGCGTGCGGGTCGCGTTCTGGCGGCCCGTCTGGCCCGTAGCGGCAGCGGTGGTGCGGAACTGGTGCCAC
CAGCCGGCGCCCCATGGCGTCTGGATACGACCAGTCCAGGTACGCTGGAAAACCTGGCCCTGGTGCCAAGT
GCCGAAGAACCACTGGGTCCACTGGATGTGCGTGTTAGTGTTCGTGCGGCGGGTCTGAATTTTCGCGATGT
TCTGATTGCGCTGGGCATGTATCCGGGCGATGCCCGTATGGGTGGTGAAGGTGCGGGTGTTGTGACCGATG
TTGGCAGTGAAGTGACGACCCTGGCCCCAGGTGACCGTGTGATGGGTATGCTGAGTAGTGCCTTTGGTCCA
ACCGCCGTGAGTGATCATCGCGCGCTGGTTCGCGTTCCGGACGATTGGAGTTTCGAACAAGCCGCGAGTGT
GCCAACCGTTTTCGCCACCGCCTACTATGGCCTGGTGGATCTGGCCGAACTGCGTGCCGGTCAAAGTGTGC
TGGTGCATGCCGCCGCGGGTGGCGTTGGTATGGCCGCGGTGCAACTGGCGCGTCATCTGGGTGCGGAAGTG
TTTGGTACGGCCAGTACGGGTAAATGGGATAGCCTGCGCGCCGGTGGTCTGGACGCCGAACATATTGCCAG
TAGTCGTACGGTTGAATTTGAAGAAACCTTTCTGGCGGCGACGGCCGGTCGTGGTGTTGATGTGGTGCTGG
ACAGCCTGGCCGGTGAGTTCGTTGATGCCAGTCTGCGCCTGCTGCCGCGTGGTGGTCGCTTTGTGGAAATG
GGTAAAGCCGATATCCGTGATGCCGAACGTGTGGCGGCCGATCATCCAGGTGTGACCTACCGTAGTTTTGA
TCTGCTGGAAGCCGGCCTGGATCGTTTTCAAGAAATTCTGACCGAGGTGGTGCGTCTGTTTGAACGTGGCG
TGCTGCGCCATCTGCCAGTGACCGCCTGGGATGTTCGCCGTGCCGCGGAGGCCTTTCGTTTTGTGAGTCAA
GCCCGTCATGTGGGCAAAAACGTTCTGGTTATGCCACGTGTGTGGGACCGTGACGGTACGGTGCTGATTAC
CGGTGGCACGGGTGCCCTGGGCGCGCTGGTGGCCCGTCATCTGGTGGCCGAACATGGTATGCGTAATGTTC
TGCTGGCCGGTCGCCGTGGCGTTGATGCCCCAGGTGCCCGTGAACTGCTGGCCGAGCTGGAAACGGCGGGT
GCCCAAGTGAGCGTTGTTGCGTGTGATGTTGCGGACCGTGATGCCGTTGCCGAACTGATTGCCAAGGTGCC
AGTGGAACATCCACTGACGGCGGTTGTTCACACCGCGGGTGTTGTGGCGGATGCCACCCTGACCGCGCTGG
ACGCCGAACGTGTGGATACGGTTCTGCGCGCGAAAGTTGACGCGGTGCTGCATCTGCATGAGGCCACCCGT
GGCCTGGACCTGGCCGGCTTTGTGCTGTTTAGTAGTGCGAGTGGTATTTTTGGTAGTCCAGGTCAAGGTAA
TTATGCGGCGGCGAATAGTTTTATCGACGCCTTTGCCCATCATCGTCGCGCCCAAGGCCTGCCAGCCCTGA
GTCTGGCCTGGGGTCTGTGGGCCCGTACCAGTGGCATGGCGGGTCAGCTGGGTCACGACGACGTGGCCCGC
ATTAGTCGCACGGGCCTGGCCCCAATTACCGATGATCAGGGTATGGCCCTGCTGGATGCGGCCCTGGGTGC
GGGTCGTCCGCTGCTGGTTCCGGTTCGTCTGGATCGTGCCGCCCTGCGTAGCCAAGCGACGGCCGGTACGC
TGCCACCAATCCTGCGTGGTCTGGTTCGCGCCACCGTGCGTCGTGCGGCGAGTACGGCCGCGGCGCAGGGT
CCAAGCCTGGCGGAGCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGA
AGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGA
TTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCG
GCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGA
TGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAG
CGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAA
CCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGA
GCTGGGCATCAGCCGCTAA

### 7.1.11. MBP-N1

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaaccttt acttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTCGACTGGGCGTCGTGGTATGCGGATGGTCCGGCGCCGGCGGCAGTTCCACTGCCGGCATACGCG
TTCCAGCGTGAACGCTATTGGTTGCCGGCAGGTGCGGGTAGCGGTCCGGGTGATGTTGCGGGTGCGGGTTT
GACGGCGGTGGGTCATGCGCTGCTGCCGGTGAGCGTTCGTCTGGCCGACGGTAGCCTGGTTCTGACCGGTC
GCCTGCCGGAAGCTGCACGTGCGGGCTGGCTGGCGGAACACCTGGTGGCGGACCTGCCGTTGTTGCCGGGC

ACGGTGTTGGTGGAGTGGGTCCTGCGTGCAGCGGACGAGGCCGGCTGTGGCGGTGTCGAAGAACTGGCGCT
GCAGGTTCCGGTCGCCCTGCCGGTTTCTGGCGGTCTGGTTATCCAGGTCGTGGTCGACGCTGCGGAAGGTG
ACGGTCGTCGCCCGGTGCGTGTTCACAGCCGTCCGGAGGAAGATAGCGGTGCGCCGGATGCGTGGGTTTGC
CACGTGAGCGGCACGCTGCTGCCAGGTGTTGCCGGTCCGGTCCCGCCGAGCGGTCCGGGCGGTGCGTGGCC
ACCGCCAGGTGCGCGTCCTGCAGCTATTGATGGCTTTTACGAACGTGCTGAAGCAGCAGGTTACGGTTATG
GTGCCTTTTTCCGTGGTCTGACCAATGTGTGGCACGACGGCGAGGATACCCTGGCAGAGGTGGTGCTGCCG
AAAGAAGCGGCAGAACAAGCGGGTGGCTTCGGTATTCACCCGGCACTGCTGGACGCGGCCATGCAACCGGT
GCTGCTGGCTGGCCAGCTGCGTCAGTGCGCAGCGGCTGCTGGTGCGGACACTGCGAGCGGCACCGTTCTGC
TGCCGTTTACCTGGTCGGGCGTGCGTCTGTGGGCGGGTGGTGCAACTCGTTTGCGTGTGCGTCTGAGCCCT
CGTCCGGAGGGCCTGCGCGTGCTGCTGGCGGATGCAACCGGTGCTCCGGTTCTGACGGCGGACGCTGTCGC
TCTGCGTGAAACGGGTGTTCAGCAGTTGCGTGCGAGCAGCCGTGTCCGTGGTAGCCACGGTCTGTTCGCCG
TTGAATGGGTGCCGCCTTTGAGCGCGACCGCAGGTGGTACGGCACCGGCAACCCTGGCCGTGCTGGGTGAT
GATGCGCCGGATCTGGCTGACGCCGACCGCTACCCGGATCTGGATGCTCTGTTCCGCGCAGTCGCGGACGG
TGCGCCTGCACCGGATGTCGTTATCGCAAGCGTCCGCACGGGCAATGACCCTGCGGGTAGCGACACCGGTC
TGGCCACGGCGCGTCGCACCCTGACCCTGGCGCAGGAGTGGCTGGCGGGTTCTGGTGCGGATGGTGCGCGC
CTGGCAGTGGTTACCCGTAGCGCAATTCGCACGGGCGACGACGGTCAAGAGCGCGTGGTTCCGAGCGCCGC
TGCGGTCTGGGGTCTGATGCGTAGCGCACAGACTGAACACCCGGGTCGCTTTGTGCTGATCGATGAGGATA
CGGATAGCACCGAGAACATTCTGGAAGCAGTCCGTACGGACGAACCTCAATTAGCCCTGCGTGGCGGTCGT
GCGCTGGTTCCGCGCATGGCGCGTGTGGATGCGGAGCCGGAGCTGACCGCACCAAGCGGTGAGCGCGCGTG
GCACGTTGCTGCGGGCAAAACCGGTCCGGATGATCTGACGGCCGTGCCGAGCCCGCGTGCGTCTGCTCCGC
TGGCGCCAGGTCAAGTCCGTATCGCGGTTCGTGCGGCAGGTCTGAATTTCCGTGATGCCCTGATTGCTCTG
GATATGTACCCGGACGCGAGCGCGAGCATTGGTAGCGAGGGTGCCGGTGTCGTGCTGGAAGTGTCCGAGGG
TGTTGCCGGCGTCGCGGTGGGCGATCGTGTTATGGGTCTGTTTAATGACGCATTTGGTCCGGTGGCTGTTG
CGGATGCGCGCATGGTTGCACCGGTCCCAGATGGTTGGTCCTTTCGTGAAGCTGCGGCTGCGCCGGTTGCG
TTCCTGACCGCCTGGTACGGTCTGGTTGACCTGGGTGGCCTGAGCTCCGGTGAAACCGTGGTCATCCACGG
CGCAGCCGGTGGTGTTGGCATGGCTGCCGTGCAGGTAGCACGTCACCTGGGTGCCGAGGTGTTCGCGACCG
CGAGCCCGGCTAAACACCCGGTGTTGGAGGGTATGGGTGTTGATGCGGCACACCGTGCAAGCTCTCGTGAC
CTGGGTTTCGAAGCAGCGTTTTCCTCCGCGACCGGTGGTCGCGGTGTCGACGTTGTCCTGAATAGCCTGGC
GGGTGAGTTCACCGACGCGAGCCTGCGCTTGCTGGCACCGGGTGGTCGCCTGATCGAAATGGGTAAGACCG
ATGTCCGTGACCCGGACCAAGTGGCTCGTGAGCACAGCGTTGCGTATCGTGCATTTGACTTGATCGCGGAC
GCCGGTCCGGAGCGCATCGGTCAGCTGCTGGCGGCTCTGGGTGAACGCTTTGCGGACGGTGCGTTTACCCC
GTTGCCGGTTACCGGTTGGCGTCTGGGTCAGGCCCGTCAAGCGCTGCGTCAACTGAGCCAAGCGCGTCACA
CCGGCAAGCTGGTGCTGGACGTTGATCCGGCACCTGATCCAGACGGCACGGTCCTGATCACGGGTGGCACG
GGCACCTTGGGTGGTCTGATTGCGGAGCACCTGGTTCGCTCTCGCGGCGTCCGCCACTTGCTGTTGTTGTC
TCGCCGTGGTCCGGATGCGCCTGGTGCGGAGGAGCTGACTGCCCGTCTGACCGAGTTGGGTGCGCGCGTTC
GTGTGGCAGCGGTTGACGTCGGTGACGCCACCGCGCTGGGCGAAGCCGTGGCGGGTGTGGACCCGGCACAC
CCGTTGACCGGCGTTGTTCACGCTGCCGGCGTGGTCGCAGATGCAATGCTGCCGAGCCAGGATGACGAGCG
TCTGGTGGCGGCATGGTCCGCGAAAGCGGCTGCCGCAGCTCGCCTGCACGACGCCACGGCGGGTCTGCCGC
TGGGTATGTTTGTTCTGTTTTCCAGCTTTGCCAGCACCCTGGGCACCGCGGGTCAGGCGAACTATGCCGCT
GCGAATGCGTACTGCGATGCGCTGGTTGAACGTCGCCATGCTGAAGGTCTGCCGGGTGTTTCCGTTAGCTG
GGGCCTGTGGAGCGCCGCAAGCGGTCTGACCGGTGGCTTGACCGAGGCAGACGTGGCTCGCATTGCACGTC
AGGGTATTGTCCCGAACAGCACGGAGCAGGGCTATGACCTGTTTGATGCCGCGCTGGGCCACGGTCGCCCA
GCCCTGTTGGCCCTGAACCTGGATACGCGTGCGCTGGCCGCACAGCCGGTTGCGGCGCTGCCGGCACCGCT
GCGTGCGCTGGCCGCTGACGCCCAAGCTGCCGGTGCTCGCAGCGGCGGTGCTGCGGCGCGTCCTACGGCCG
CAGCCGCGGAGGAGCCGGCAGATTGGGCAGCGCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCG
GTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGA
TAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTC
TGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCAC
GAACAACTGGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTT
GGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATA
TGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTG
GACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.12. MBP-N2

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTCCGGTGAGCGTTCGTCTGGCCGACGGTAGCCTGGTTCT
GACCGGTCGCCTGCCGGAAGCTGCACGTGCGGGCTGGCTGGCGGAACACCTGGTGGCGGACCTGCCGTTGT
TGCCGGGCACGGTGTTGGTGGAGTGGGTCCTGCGTGCAGCGGACGAGGCCGGCTGTGGCGGTGTCGAAGAA

CTGGCGCTGCAGGTTCCGGTCGCCCTGCCGGTTTCTGGCGGTCTGGTTATCCAGGTCGTGGTCGACGCTGC
GGAAGGTGACGGTCGTCGCCCGGTGCGTGTTCACAGCCGTCCGGAGGAAGATAGCGGTGCGCCGGATGCGT
GGGTTTGCCACGTGAGCGGCACGCTGCTGCCAGGTGTTGCCGGTCCGGTCCCGCCGAGCGGTCCGGGCGGT
GCGTGGCCACCGCCAGGTGCGCGTCCTGCAGCTATTGATGGCTTTTACGAACGTGCTGAAGCAGCAGGTTA
CGGTTATGGTGCCTTTTTCCGTGGTCTGACCAATGTGTGGCACGACGGCGAGGATACCCTGGCAGAGGTGG
TGCTGCCGAAAGAAGCGGCAGAACAAGCGGGTGGCTTCGGTATTCACCCGGCACTGCTGGACGCGGCCATG
CAACCGGTGCTGCTGGCTGGCCAGCTGCGTCAGTGCGCAGCGGCTGCTGGTGCGGACACTGCGAGCGGCAC
CGTTCTGCTGCCGTTTACCTGGTCGGGCGTGCGTCTGTGGGCGGGTGGTGCAACTCGTTTGCGTGTGCGTC
TGAGCCCTCGTCCGGAGGGCCTGCGCGTGCTGCTGGCGGATGCAACCGGTGCTCCGGTTCTGACGGCGGAC
GCTGTCGCTCTGCGTGAAACGGGTGTTCAGCAGTTGCGTGCGAGCAGCCGTGTCCGTGGTAGCCACGGTCT
GTTCGCCGTTGAATGGGTGCCGCCTTTGAGCGCGACCGCAGGTGGTACGGCACCGGCAACCCTGGCCGTGC
TGGGTGATGATGCGCCGGATCTGGCTGACGCCGACCGCTACCCGGATCTGGATGCTCTGTTCCGCGCAGTC
GCGGACGGTGCGCCTGCACCGGATGTCGTTATCGCAAGCGTCCGCACGGGCAATGACCCTGCGGGTAGCGA
CACCGGTCTGGCCACGGCGCGTCGCACCCTGACCCTGGCGCAGGAGTGGCTGGCGGGTTCTGGTGCGGATG
GTGCGCGCCTGGCAGTGGTTACCCGTAGCGCAATTCGCACGGGCGACGACGGTCAAGAGCGCGTGGTTCCG
AGCGCCGCTGCGGTCTGGGGTCTGATGCGTAGCGCACAGACTGAACACCCGGGTCGCTTTGTGCTGATCGA
TGAGGATACGGATAGCACCGAGAACATTCTGGAAGCAGTCCGTACGGACGAACCTCAATTAGCCCTGCGTG
GCGGTCGTGCGCTGGTTCCGCGCATGGCGCGTGTGGATGCGGAGCCGGAGCTGACCGCACCAAGCGGTGAG
CGCGCGTGGCACGTTGCTGCGGGCAAAACCGGTCCGGATGATCTGACGGCCGTGCCGAGCCCGCGTGCGTC
TGCTCCGCTGGCGCCAGGTCAAGTCCGTATCGCGGTTCGTGCGGCAGGTCTGAATTTCCGTGATGCCCTGA
TTGCTCTGGATATGTACCCGGACGCGAGCGCGAGCATTGGTAGCGAGGGTGCCGGTGTCGTGCTGGAAGTG
TCCGAGGGTGTTGCCGGCGTCGCGGTGGGCGATCGTGTTATGGGTCTGTTTAATGACGCATTTGGTCCGGT
GGCTGTTGCGGATGCGCGCATGGTTGCACCGGTCCCAGATGGTTGGTCCTTTCGTGAAGCTGCGGCTGCGC
CGGTTGCGTTCCTGACCGCCTGGTACGGTCTGGTTGACCTGGGTGGCCTGAGCTCCGGTGAAACCGTGGTC
ATCCACGGCGCAGCCGGTGGTGTTGGCATGGCTGCCGTGCAGGTAGCACGTCACCTGGGTGCCGAGGTGTT
CGCGACCGCGAGCCCGGCTAAACACCCGGTGTTGGAGGGTATGGGTGTTGATGCGGCACACCGTGCAAGCT
CTCGTGACCTGGGTTTCGAAGCAGCGTTTTCCTCCGCGACCGGTGGTCGCGGTGTCGACGTTGTCCTGAAT
AGCCTGGCGGGTGAGTTCACCGACGCGAGCCTGCGCTTGCTGGCACCGGGTGGTCGCCTGATCGAAATGGG
TAAGACCGATGTCCGTGACCCGGACCAAGTGGCTCGTGAGCACAGCGTTGCGTATCGTGCATTTGACTTGA
TCGCGGACGCCGGTCCGGAGCGCATCGGTCAGCTGCTGGCGGCTCTGGGTGAACGCTTTGCGGACGGTGCG
TTTACCCCGTTGCCGGTTACCGGTTGGCGTCTGGGTCAGGCCCGTCAAGCGCTGCGTCAACTGAGCCAAGC
GCGTCACACCGGCAAGCTGGTGCTGGACGTTGATCCGGCACCTGATCCAGACGGCACGGTCCTGATCACGG
GTGGCACGGGCACCTTGGGTGGTCTGATTGCGGAGCACCTGGTTCGCTCTCGCGGCGTCCGCCACTTGCTG
TTGTTGTCTCGCCGTGGTCCGGATGCGCCTGGTGCGGAGGAGCTGACTGCCCGTCTGACCGAGTTGGGTGC
GCGCGTTCGTGTGGCAGCGGTTGACGTCGGTGACGCCACCGCGCTGGGCGAAGCCGTGGCGGGTGTGGACC
CGGCACACCCGTTGACCGGCGTTGTTCACGCTGCCGGCGTGGTCGCAGATGCAATGCTGCCGAGCCAGGAT
GACGAGCGTCTGGTGGCGGCATGGTCCGCGAAAGCGGCTGCCGCAGCTCGCCTGCACGACGCCACGGCGGG
TCTGCCGCTGGGTATGTTTGTTCTGTTTTCCAGCTTTGCCAGCACCCTGGGCACCGCGGGTCAGGCGAACT
ATGCCGCTGCGAATGCGTACTGCGATGCGCTGGTTGAACGTCGCCATGCTGAAGGTCTGCCGGGTGTTTCC
GTTAGCTGGGGCCTGTGGAGCGCCGCAAGCGGTCTGACCGGTGGCTTGACCGAGGCAGACGTGGCTCGCAT
TGCACGTCAGGGTATTGTCCCGAACAGCACGGAGCAGGGCTATGACCTGTTTGATGCCGCGCTGGGCCACG
GTCGCCCAGCCCTGTTGGCCCTGAACCTGGATACGCGTGCGCTGGCCGCACAGCCGGTTGCGGCGCTGCCG
GCACCGCTGCGTGCGCTGGCCGCTGACGCCCAAGCTGCCGGTGCTCGCAGCGGCGGTGCTGCGGCGCGTCC
TACGGCCGCAGCCGCGGAGGAGCCGGCAGATTGGGCAGCGCGTCTGGCTGCGTTGCCGGAAGCAGACCGTC
GCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATC
GGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGAC
CGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACC
ACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTG
GATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCT
GAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCG
AAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.13. MBP-S1

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTGGACTGGCGTAGTTTCTTCGCGGGTACCGGTGCGAAACAAATTGAGCTGCCAACCTACGCCTTT
CAACGTCAACGTTACTGGCTGGTGCCAAGCGACAGTGGTGATGTGACGGGTGCCGGTCTGGCCGGTGCCGA
ACACCCACTGCTGGGTGCGGTTGTTCCGGTTGCGGGTGGTGACGAAGTGCTGCTGACCGGCCGTATTAGTG
TGCGTACCCATCCATGGCTGGCGGAGCACCGTGTTCTGGGCGAGGTTATTGTGGCCGGCACGGCCCTGCTG

GAAATTGCCCTGCATGCCGGCGAACGTCTGGGCTGTGAGCGTGTGGAGGAACTGACCCTGGAAGCCCCGCT
GGTTCTGCCGGAACGTGGTGCGATTCAAGTGCAACTGCGCGTGGGTGCCCCAGAAAATAGTGGTCGCCGTC
CGATGGCCCTGTATAGTCGTCCGGAAGGTGCCGCGGAACATGATTGGACCCGCCATGCCACGGGTCGCCTG
GCCCCAGGCCGTGGCGAAGCCGCCGGTGATCTGGCGGATTGGCCAGCCCCAGGTGCGCTGCCAGTGGATCT
GGATGAATTTTACCGTGACCTGGCCGAGCTGGGTCTGGAATACGGTCCAATTTTTCAAGGTCTGAAGGCGG
CGTGGCGTCAAGGTGACGAAGTGTACGCCGAAGCCGCCCTGCCGGGTACCGAAGATAGTGGTTTTGGTGTT
CATCCAGCCCTGCTGGACGCGGCGCTGCACGCGACGGCGGTTCGTGATATGGATGATGCCCGCCTGCCATT
TCAATGGGAAGGTGTGAGTCTGCATGCGAAAGCGGCCCCAGCGCTGCGCGTGCGTGTTGTGCCAGCGGGCG
ATGACGCGAAAAGTCTGCTGGTTTGTGACGGCACGGGTCGTCCAGTGATTAGTGTGGATCGTCTGGTTCTG
CGCAGTGCCGCCGCGCGTCGTACGGGTGCCCGTCGCCAGGCCCACCAAGCGCGCCTGTACCGTCTGAGTTG
GCCGACCGTTCAGCTGCCAACCAGTGCCCAGCCGCCGAGTTGTGTTCTGCTGGGTACCAGTGAAGTGAGTG
CCGATATTCAAGTGTATCCAGACCTGCGTAGCCTGACCGCGGCCCTGGATGCCGGTGCCGAGCCACCGGGT
GTGGTTATCGCCCCGACCCCACCAGGCGGTGGTCGCACGGCGGATGTGCGTGAAACCACCCGTCATGCGCT
GGACCTGGTGCAAGGTTGGCTGAGTGATCAACGTCTGAATGAAAGTCGTCTGCTGCTGGTGACCCAAGGCG
CCGTGGCCGTGGAACCAGGTGAGCCGGTTACGGATCTGGCGCAAGCCGCGCTGTGGGGCCTGCTGCGTAGT
ACCCAAACCGAACACCCAGATCGTTTTGTTCTGGTTGATGTTCCAGAACCGGCGCAACTGCTGCCGGCGCT
GCCAGGCGTTCTGGCCTGTGGTGAACCACAACTGGCCCTGCGTCGCGGTGGCGCCCATGCCCCACGTCTGG
CGGGTCTGGGCAGTGATGACGTTCTGCCAGTTCCGGATGGTACCGGTTGGCGTCTGGAAGCCACCCGCCCA
GGTAGCCTGGACGGTCTGGCGCTGGTGGATGAACCGACCGCGACGGCCCCACTGGGCGACGGTGAAGTGCG
TATTGCCATGCGTGCCGCGGGTGTTAACTTCCGTGATGCCCTGATTGCGCTGGGCATGTATCCGGGCGTGG
CCAGTCTGGGTAGTGAAGGTGCCGGCGTGGTTGTTGAAACGGGCCCAGGTGTTACGGGCCTGGCCCCAGGC
GACCGTGTGATGGGTATGATTCCAAAAGCGTTTGGTCCGCTGGCGGTGGCGGATCATCGTATGGTGACCCG
TATTCCAGCCGGTTGGAGTTTTGCCCGTGCGGCCAGTGTTCCAATTGTGTTTCTGACCGCCTACTACGCCC
TGGTTGACCTGGCGGGTCTGCGTCCGGGTGAAAGTCTGCTGGTTCACAGTGCGGCGGGTGGCGTTGGCATG
GCCGCGATCCAACTGGCGCGTCATCTGGGTGCGGAAGTGTATGCCACGGCGAGTGAGGATAAATGGCAAGC
CGTGGAACTGAGTCGTGAACATCTGGCGAGTAGTCGTACCTGTGATTTTGAACAACAATTTCTGGGCGCCA
CGGGCGGTCGCGGCGTTGATGTGGTGCTGAATAGTCTGGCCGGTGAGTTCGCCGACGCCAGCCTGCGTATG
CTGCCGCGTGGTGGCCGTTTTCTGGAGCTGGGTAAAACCGATGTGCGTGACCCAGTTGAAGTTGCCGATGC
CCATCCAGGCGTTAGTTACCAAGCCTTTGATACGGTGGAAGCCGGTCCACAACGCATTGGTGAGATGCTGC
ATGAGCTGGTTGAGCTGTTTGAAGGTCGTGTGCTGGAACCACTGCCAGTGACGGCCTGGGACGTTCGTCAA
GCCCCAGAAGCCCTGCGTCACCTGAGTCAAGCGCGCCATGTTGGTAAACTGGTTCTGACGATGCCACCAGT
GTGGGATGCGGCCGGTACGGTTCTGGTTACCGGCGGTACGGGTGCCCTGGGTGCCGAAGTGGCCCGTCATC
TGGTTATTGAGCGTGGTGTTCGTAATCTGGTGCTGGTGAGTCGTCGTGGCCCAGCCGCGAGCGGTGCCGCG
GAGCTGGTTGCCCAACTGACCGCCTACGGTGCGGAAGTTAGTCTGCAAGCCTGTGATGTTGCCGATCGTGA
AACCCTGGCCAAAGTGCTGGCGAGTATTCCGGATGAACATCCGCTGACGGCGGTTGTGCATGCGGCGGGTG
TTCTGGACGATGGTGTTAGTGAGAGTCTGACCGTTGAACGTCTGGACCAAGTTCTGCGTCCAAAAGTGGAT
GGTGCCCGCAATCTGCTGGAACTGATTGATCCAGATGTGGCGCTGGTTCTGTTTAGTAGTGTGAGTGGCGT
GCTGGGTAGTGGTGGTCAAGGTAACTATGCCGCCGCGAATAGCTTTCTGGATGCGCTGGCCCAACAACGTC
AAAGTCGCGGTCTGCCAACGCGTAGCCTGGCCTGGGGTCCGTGGGCCGAGCATGGTATGGCCAGTACCCTG
CGCGAGGCCGAACAAGATCGTCTGGCCCGCAGTGGCCTGCTGCCAATTAGTACCGAAGAGGGTCTGAGTCA
ATTCGATGCCGCGTGCGGTGGTGCGCATACCGTGGTTGCGCCAGTGCGTTTTAGTCGCCTGAGTGATGGTA
ATGCCATTAAGTTCAGTGTGCTGCAAGGTCTGGTGGGTCCACATCGTGTGAATAAAGCGGCCACCGCGGAT
GATGCCGAAAGTCTGCGTAAACGCCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCT
GGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCC
ACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAG
CTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGT
TATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCC
TGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCG
TGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGAC
CAATGAGCTGGGCATCAGCCGCTAA

### 7.1.14. MBP-S2

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaaccttt acttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTGCGGTTGTTCCGGTTGCGGGTGGTGACGAAGTGCT
GCTGACCGGCCGTATTAGTGTGCGTACCCATCCATGGCTGGCGGAGCACCGTGTTCTGGGCGAGGTTATTG
TGGCCGGCACGGCCCTGCTGGAAATTGCCCTGCATGCCGGCGAACGTCTGGGCTGTGAGCGTGTGGAGGAA

CTGACCCTGGAAGCCCCGCTGGTTCTGCCGGAACGTGGTGCGATTCAAGTGCAACTGCGCGTGGGTGCCCC
AGAAAATAGTGGTCGCCGTCCGATGGCCCTGTATAGTCGTCCGGAAGGTGCCGCGGAACATGATTGGACCC
GCCATGCCACGGGTCGCCTGGCCCCAGGCCGTGGCGAAGCCGCCGGTGATCTGGCGGATTGGCCAGCCCCA
GGTGCGCTGCCAGTGGATCTGGATGAATTTTACCGTGACCTGGCCGAGCTGGGTCTGGAATACGGTCCAAT
TTTTCAAGGTCTGAAGGCGGCGTGGCGTCAAGGTGACGAAGTGTACGCCGAAGCCGCCCTGCCGGGTACCG
AAGATAGTGGTTTTGGTGTTCATCCAGCCCTGCTGGACGCGGCGCTGCACGCGACGGCGGTTCGTGATATG
GATGATGCCCGCCTGCCATTTCAATGGGAAGGTGTGAGTCTGCATGCGAAAGCGGCCCCAGCGCTGCGCGT
GCGTGTTGTGCCAGCGGGCGATGACGCGAAAAGTCTGCTGGTTTGTGACGGCACGGGTCGTCCAGTGATTA
GTGTGGATCGTCTGGTTCTGCGCAGTGCCGCCGCGCGTCGTACGGGTGCCCGTCGCCAGGCCCACCAAGCG
CGCCTGTACCGTCTGAGTTGGCCGACCGTTCAGCTGCCAACCAGTGCCCAGCCGCCGAGTTGTGTTCTGCT
GGGTACCAGTGAAGTGAGTGCCGATATTCAAGTGTATCCAGACCTGCGTAGCCTGACCGCGGCCCTGGATG
CCGGTGCCGAGCCACCGGGTGTGGTTATCGCCCCGACCCCACCAGGCGGTGGTCGCACGGCGGATGTGCGT
GAAACCACCCGTCATGCGCTGGACCTGGTGCAAGGTTGGCTGAGTGATCAACGTCTGAATGAAAGTCGTCT
GCTGCTGGTGACCCAAGGCGCCGTGGCCGTGGAACCAGGTGAGCCGGTTACGGATCTGGCGCAAGCCGCGC
TGTGGGGCCTGCTGCGTAGTACCCAAACCGAACACCCAGATCGTTTTGTTCTGGTTGATGTTCCAGAACCG
GCGCAACTGCTGCCGGCGCTGCCAGGCGTTCTGGCCTGTGGTGAACCACAACTGGCCCTGCGTCGCGGTGG
CGCCCATGCCCCACGTCTGGCGGGTCTGGGCAGTGATGACGTTCTGCCAGTTCCGGATGGTACCGGTTGGC
GTCTGGAAGCCACCCGCCCAGGTAGCCTGGACGGTCTGGCGCTGGTGGATGAACCGACCGCGACGGCCCCA
CTGGGCGACGGTGAAGTGCGTATTGCCATGCGTGCCGCGGGTGTTAACTTCCGTGATGCCCTGATTGCGCT
GGGCATGTATCCGGGCGTGGCCAGTCTGGGTAGTGAAGGTGCCGGCGTGGTTGTTGAAACGGGCCCAGGTG
TTACGGGCCTGGCCCCAGGCGACCGTGTGATGGGTATGATTCCAAAAGCGTTTGGTCCGCTGGCGGTGGCG
GATCATCGTATGGTGACCCGTATTCCAGCCGGTTGGAGTTTTGCCCGTGCGGCCAGTGTTCCAATTGTGTT
TCTGACCGCCTACTACGCCCTGGTTGACCTGGCGGGTCTGCGTCCGGGTGAAAGTCTGCTGGTTCACAGTG
CGGCGGGTGGCGTTGGCATGGCCGCGATCCAACTGGCGCGTCATCTGGGTGCGGAAGTGTATGCCACGGCG
AGTGAGGATAAATGGCAAGCCGTGGAACTGAGTCGTGAACATCTGGCGAGTAGTCGTACCTGTGATTTTGA
ACAACAATTTCTGGGCGCCACGGGCGGTCGCGGCGTTGATGTGGTGCTGAATAGTCTGGCCGGTGAGTTCG
CCGACGCCAGCCTGCGTATGCTGCCGCGTGGTGGCCGTTTTCTGGAGCTGGGTAAAACCGATGTGCGTGAC
CCAGTTGAAGTTGCCGATGCCCATCCAGGCGTTAGTTACCAAGCCTTTGATACGGTGGAAGCCGGTCCACA
ACGCATTGGTGAGATGCTGCATGAGCTGGTTGAGCTGTTTGAAGGTCGTGTGCTGGAACCACTGCCAGTGA
CGGCCTGGGACGTTCGTCAAGCCCCAGAAGCCCTGCGTCACCTGAGTCAAGCGCGCCATGTTGGTAAACTG
GTTCTGACGATGCCACCAGTGTGGGATGCGGCCGGTACGGTTCTGGTTACCGGCGGTACGGGTGCCCTGGG
TGCCGAAGTGGCCCGTCATCTGGTTATTGAGCGTGGTGTTCGTAATCTGGTGCTGGTGAGTCGTCGTGGCC
CAGCCGCGAGCGGTGCCGCGGAGCTGGTTGCCCAACTGACCGCCTACGGTGCGGAAGTTAGTCTGCAAGCC
TGTGATGTTGCCGATCGTGAAACCCTGGCCAAAGTGCTGGCGAGTATTCCGGATGAACATCCGCTGACGGC
GGTTGTGCATGCGGCGGGTGTTCTGGACGATGGTGTTAGTGAGAGTCTGACCGTTGAACGTCTGGACCAAG
TTCTGCGTCCAAAAGTGGATGGTGCCCGCAATCTGCTGGAACTGATTGATCCAGATGTGGCGCTGGTTCTG
TTTAGTAGTGTGAGTGGCGTGCTGGGTAGTGGTGGTCAAGGTAACTATGCCGCCGCGAATAGCTTTCTGGA
TGCGCTGGCCCAACAACGTCAAAGTCGCGGTCTGCCAACGCGTAGCCTGGCCTGGGGTCCGTGGGCCGAGC
ATGGTATGGCCAGTACCCTGCGCGAGGCCGAACAAGATCGTCTGGCCCGCAGTGGCCTGCTGCCAATTAGT
ACCGAAGAGGGTCTGAGTCAATTCGATGCCGCGTGCGGTGGTGCGCATACCGTGGTTGCGCCAGTGCGTTT
TAGTCGCCTGAGTGATGGTAATGCCATTAAGTTCAGTGTGCTGCAAGGTCTGGTGGGTCCACATCGTGTGA
ATAAAGCGGCCACCGCGGATGATGCCGAAAGTCTGCGTAAACGCCTGGCTGCGTTGCCGGAAGCAGACCGT
CGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGAT
CGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGA
CCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGAC
CACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATT
GGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTC
TGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACC
GAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.15. MBP-A2c

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGT
GTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGC
TGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGAC

CTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACC
GGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCC
TGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCG
CCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCC
TCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAG
CAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGT
CTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTAC
GCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCG
ACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTGCGT
GCTACCCGTACGCACGACACCCTGTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGCACG
TTGCGCAGTCCTGGGTGATGACGAACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCGCAC
TGGAAAGTGCCGATCCGGTGCCGGAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGACCGT
GCAGACGCTGCACGTAGCCTGACCGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACGCTG
GGCAACCACGCGTCTGGCTCTGGTGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGGATC
TGCCGGCAGCAGCTGTCTGGGGTCTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTGGCT
GATCTGGACGGTGATACCGCATCGGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACCGCA
GCTGGCACTGCGTGAAGGTGCAGTTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGGTGC
CGCCGCCGGGTACGCGCGATTGGCACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTGACC
CCGTTTCCGGAAGCAGCAGCTCCGCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCTGAA
CTTCCGTGATGTTGTCATGGCACTGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAGGTA
TCGTGACCGAAGCAGGTCCGGGTGTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGATGGT
TGCATCGGTCCGGTTGCCGTGGTTGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCCGCA
GGCAGCTTCAGTGCCGGTTACCTTCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTCCGG
GTGATCGTGTCCTGGTGCATGCAGCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCACCTG
GGTGCTGAAGTTTTTGCAACCGCTGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGACGA
TCATCTGGCGAGCTCTCGCACCGATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTATCG
ATGTCGTGCTGAATTCTCTGGCCGGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGCCGT
TTTATTGAAATGGGCAAAACCGATATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGTCTA
CCGTGCCTTCGACCTGATGGATGGTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTCTGT
TCGAAGGTGGCAAAATTCAGCTGGTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTTCGC
TATTTCGCACAAGCTCGTCATGTGGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGGTAC
CGTGCTGGTTACGGGTGCATCCGGTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATGATG
TGCGTCACCTGCTGCTGGCATCACGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAACTG
CGTGAATCTGGTGCACATAGTGTGCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGACCT
GCTGGCGTCTATTCCGGATGAACATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATGGCG
TGCTGGAAACCATGACGCCGGAACGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCATCTG
CACGAACTGACCCGCGACCGTGATCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGGTAG
TGCAGCTCAGGCAAACTACGCGGCCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATCACG
GTCTGCCGGCAACCTCTCTGGCTTGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGTCGT
GCAGACCTGGATCGTATGCAACGCTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTTTGA
CGCTGCACTGGCAGCAGGTCGTCCGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCGCTG
CGGGTGGCGGTTCTCTGCCGGCACCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCGCTG
CCGGCTGCGGATGCGCTGCGCGGTCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGA
CCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCT
TCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATG
AAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACT
GGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCC
GCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACC
GCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCT
GACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.16. MBP-I2c

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaaccttt acttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGT
GTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGC
TGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGAC

CTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACC
GGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCC
TGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCG
CCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCC
TCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAG
CAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGT
CTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTAC
GCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCG
ACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGCTGATTCCAGAAATGAGCCCGGAACAACTGGAA
CGTGTTCGCGGTGAGGAAAAAGAAGCCCCATACGTGCTGGATTGGGTTCCGGTGGAAGTGCCAGCCGATGA
CCTGGTTCGCCCAGAGCGTTGGACGCTGCTGGGCGGTGCCGACGCGGGTGTGGGCCTGGATGTTGCGGGTG
CCTTCGCCAGTCTGGAACCAAGTGATGGTGCGCCGGAATTTGTTGTTCTGCCGTGTGTGCCGCCAACGAGT
CCAACCCGTGCCGCGGATGTTCGTCAAAGTACCCTGCAAGCCCTGACCGTTCTGCAAAATTGGGTGACCGA
TGAGCGCCATGCCGATAGTCGTCTGGTTCTGGTGACCCGTCGTGCCGTTGGTGTTGGTGCCCATGATGATG
TGCCGGATCTGACGCATGCCGCCCTGTGGGGTCTGGTGCGTAGTGCCCAAACCGAAAATCCAGGTCGTTTT
CTGCTGGTTGATCTGGATGAAGGCGCGGAGCTGGCCGAGGTGCTGCCAGGCGCGCTGGGTAGCGGTGAAAG
TCAAGTGGCCGTGCGTGCGGGTCGCGTTCTGGCGGCCCGTCTGGCCCGTAGCGGCAGCGGTGGTGCGGAAC
TGGTGCCACCAGCCGGCGCCCCATGGCGTCTGGATACGACCAGTCCAGGTACGCTGGAAAACCTGGCCCTG
GTGCCAAGTGCCGAAGAACCACTGGGTCCACTGGATGTGCGTGTTAGTGTTCGTGCGGCGGGTCTGAATTT
TCGCGATGTTCTGATTGCGCTGGGCATGTATCCGGGCGATGCCCGTATGGGTGGTGAAGGTGCGGGTGTTG
TGACCGATGTTGGCAGTGAAGTGACGACCCTGGCCCCAGGTGACCGTGTGATGGGTATGCTGAGTAGTGCC
TTTGGTCCAACCGCCGTGAGTGATCATCGCGCGCTGGTTCGCGTTCCGGACGATTGGAGTTTCGAACAAGC
CGCGAGTGTGCCAACCGTTTTCGCCACCGCCTACTATGGCCTGGTGGATCTGGCCGAACTGCGTGCCGGTC
AAAGTGTGCTGGTGCATGCCGCCGCGGGTGGCGTTGGTATGGCCGCGGTGCAACTGGCGCGTCATCTGGGT
GCGGAAGTGTTTGGTACGGCCAGTACGGGTAAATGGGATAGCCTGCGCGCCGGTGGTCTGGACGCCGAACA
TATTGCCAGTAGTCGTACGGTTGAATTTGAAGAAACCTTTCTGGCGGCGACGGCCGGTCGTGGTGTTGATG
TGGTGCTGGACAGCCTGGCCGGTGAGTTCGTTGATGCCAGTCTGCGCCTGCTGCCGCGTGGTGGTCGCTTT
GTGGAAATGGGTAAAGCCGATATCCGTGATGCCGAACGTGTGGCGGCCGATCATCCAGGTGTGACCTACCG
TAGTTTTGATCTGCTGGAAGCCGGCCTGGATCGTTTTCAAGAAATTCTGACCGAGGTGGTGCGTCTGTTTG
AACGTGGCGTGCTGCGCCATCTGCCAGTGACCGCCTGGGATGTTCGCCGTGCCGCGGAGGCCTTTCGTTTT
GTGAGTCAAGCCCGTCATGTGGGCAAAAACGTTCTGGTTATGCCACGTGTGTGGGACCGTGACGGTACGGT
GCTGATTACCGGTGGCACGGGTGCCCTGGGCGCGCTGGTGGCCCGTCATCTGGTGGCCGAACATGGTATGC
GTAATGTTCTGCTGGCCGGTCGCCGTGGCGTTGATGCCCCAGGTGCCCGTGAACTGCTGGCCGAGCTGGAA
ACGGCGGGTGCCCAAGTGAGCGTTGTTGCGTGTGATGTTGCGGACCGTGATGCCGTTGCCGAACTGATTGC
CAAGGTGCCAGTGGAACATCCACTGACGGCGGTTGTTCACACCGCGGGTGTTGTGGCGGATGCCACCCTGA
CCGCGCTGGACGCCGAACGTGTGGATACGGTTCTGCGCGCGAAAGTTGACGCGGTGCTGCATCTGCATGAG
GCCACCCGTGGCCTGGACCTGGCCGGCTTTGTGCTGTTTAGTAGTGCGAGTGGTATTTTTGGTAGTCCAGG
TCAAGGTAATTATGCGGCGGCGAATAGTTTTATCGACGCCTTTGCCCATCATCGTCGCGCCCAAGGCCTGC
CAGCCCTGAGTCTGGCCTGGGGTCTGTGGGCCCGTACCAGTGGCATGGCGGGTCAGCTGGGTCACGACGAC
GTGGCCCGCATTAGTCGCACGGGCCTGGCCCCAATTACCGATGATCAGGGTATGGCCCTGCTGGATGCGGC
CCTGGGTGCGGGTCGTCCGCTGCTGGTTCCGGTTCGTCTGGATCGTGCCGCCCTGCGTAGCCAAGCGACGG
CCGGTACGCTGCCACCAATCCTGCGTGGTCTGGTTCGCGCCACCGTGCGTCGTGCGGCGAGTACGGCCGCG
GCGCAGGGTCCAAGCCTGGCGGAGCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGA
CCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCT
TCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATG
AAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACT
GGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCC
GCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACC
GCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCT
GACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.17. MBP-S2c

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGT
GTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGC
TGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGAC

CTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACC
GGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCC
TGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCG
CCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCC
TCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAG
CAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGT
CTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTAC
GCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCG
ACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGGTTCTGCGCAGTGCCGCCGCGCGTCGTACGGGT
GCCCGTCGCCAGGCCCACCAAGCGCGCCTGTACCGTCTGAGTTGGCCGACCGTTCAGCTGCCAACCAGTGC
CCAGCCGCCGAGTTGTGTTCTGCTGGGTACCAGTGAAGTGAGTGCCGATATTCAAGTGTATCCAGACCTGC
GTAGCCTGACCGCGGCCCTGGATGCCGGTGCCGAGCCACCGGGTGTGGTTATCGCCCCGACCCCACCAGGC
GGTGGTCGCACGGCGGATGTGCGTGAAACCACCCGTCATGCGCTGGACCTGGTGCAAGGTTGGCTGAGTGA
TCAACGTCTGAATGAAAGTCGTCTGCTGCTGGTGACCCAAGGCGCCGTGGCCGTGGAACCAGGTGAGCCGG
TTACGGATCTGGCGCAAGCCGCGCTGTGGGGCCTGCTGCGTAGTACCCAAACCGAACACCCAGATCGTTTT
GTTCTGGTTGATGTTCCAGAACCGGCGCAACTGCTGCCGGCGCTGCCAGGCGTTCTGGCCTGTGGTGAACC
ACAACTGGCCCTGCGTCGCGGTGGCGCCCATGCCCCACGTCTGGCGGGTCTGGGCAGTGATGACGTTCTGC
CAGTTCCGGATGGTACCGGTTGGCGTCTGGAAGCCACCCGCCCAGGTAGCCTGGACGGTCTGGCGCTGGTG
GATGAACCGACCGCGACGGCCCCACTGGGCGACGGTGAAGTGCGTATTGCCATGCGTGCCGCGGGTGTTAA
CTTCCGTGATGCCCTGATTGCGCTGGGCATGTATCCGGGCGTGGCCAGTCTGGGTAGTGAAGGTGCCGGCG
TGGTTGTTGAAACGGGCCCAGGTGTTACGGGCCTGGCCCCAGGCGACCGTGTGATGGGTATGATTCCAAAA
GCGTTTGGTCCGCTGGCGGTGGCGGATCATCGTATGGTGACCCGTATTCCAGCCGGTTGGAGTTTTGCCCG
TGCGGCCAGTGTTCCAATTGTGTTTCTGACCGCCTACTACGCCCTGGTTGACCTGGCGGGTCTGCGTCCGG
GTGAAAGTCTGCTGGTTCACAGTGCGGCGGGTGGCGTTGGCATGGCCGCGATCCAACTGGCGCGTCATCTG
GGTGCGGAAGTGTATGCCACGGCGAGTGAGGATAAATGGCAAGCCGTGGAACTGAGTCGTGAACATCTGGC
GAGTAGTCGTACCTGTGATTTTGAACAACAATTTCTGGGCGCCACGGGCGGTCGCGGCGTTGATGTGGTGC
TGAATAGTCTGGCCGGTGAGTTCGCCGACGCCAGCCTGCGTATGCTGCCGCGTGGTGGCCGTTTTCTGGAG
CTGGGTAAAACCGATGTGCGTGACCCAGTTGAAGTTGCCGATGCCCATCCAGGCGTTAGTTACCAAGCCTT
TGATACGGTGGAAGCCGGTCCACAACGCATTGGTGAGATGCTGCATGAGCTGGTTGAGCTGTTTGAAGGTC
GTGTGCTGGAACCACTGCCAGTGACGGCCTGGGACGTTCGTCAAGCCCCAGAAGCCCTGCGTCACCTGAGT
CAAGCGCGCCATGTTGGTAAACTGGTTCTGACGATGCCACCAGTGTGGGATGCGGCCGGTACGGTTCTGGT
TACCGGCGGTACGGGTGCCCTGGGTGCCGAAGTGGCCCGTCATCTGGTTATTGAGCGTGGTGTTCGTAATC
TGGTGCTGGTGAGTCGTCGTGGCCCAGCCGCGAGCGGTGCCGCGGAGCTGGTTGCCCAACTGACCGCCTAC
GGTGCGGAAGTTAGTCTGCAAGCCTGTGATGTTGCCGATCGTGAAACCCTGGCCAAAGTGCTGGCGAGTAT
TCCGGATGAACATCCGCTGACGGCGGTTGTGCATGCGGCGGGTGTTCTGGACGATGGTGTTAGTGAGAGTC
TGACCGTTGAACGTCTGGACCAAGTTCTGCGTCCAAAAGTGGATGGTGCCCGCAATCTGCTGGAACTGATT
GATCCAGATGTGGCGCTGGTTCTGTTTAGTAGTGTGAGTGGCGTGCTGGGTAGTGGTGGTCAAGGTAACTA
TGCCGCCGCGAATAGCTTTCTGGATGCGCTGGCCCAACAACGTCAAAGTCGCGGTCTGCCAACGCGTAGCC
TGGCCTGGGGTCCGTGGGCCGAGCATGGTATGGCCAGTACCCTGCGCGAGGCCGAACAAGATCGTCTGGCC
CGCAGTGGCCTGCTGCCAATTAGTACCGAAGAGGGTCTGAGTCAATTCGATGCCGCGTGCGGTGGTGCGCA
TACCGTGGTTGCGCCAGTGCGTTTTAGTCGCCTGAGTGATGGTAATGCCATTAAGTTCAGTGTGCTGCAAG
GTCTGGTGGGTCCACATCGTGTGAATAAAGCGGCCACCGCGGATGATGCCGAAAGTCTGCGTAAACGCCTG
GCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCT
GGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCG
CAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCAC
CCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACAC
GGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAG
CGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGC
GACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.18. BorDH2 monodomain

atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagcCATATGGGTTCTGC
GGCGGAACCGGCTGAACTGGGTCTGGGTGACGCTCGCCACCCGCTGCTGGGTAGCGTTGTTACCGTGGCTG
GCGATGACAAAGTTGTGTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTT
CTGGACGCCGTTCTGCTGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTG
CCCGGTTGTTCACGACCTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAA
TGGCAGTAGGTGCACCGGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCA
GAACACGAATGGACCCTGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGC
CGGTGCCTGGCCGCCGCCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACG
GTTACCACTACGGCCCTCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAA
ATTGTGCTGCCGGAAGCAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGT
GCTGCACGCAGCTCGTCTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGG
GCGTGACCCTGCGTACGCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATC
CGTCTGGACGTGACCGACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGACCCTGCGTCCGGTCTC
CGCGGGTCAGCTGATGGCTGTGCCGCGTGACTAA

### 7.1.19. MBP-A2c-TE

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaaccttt acttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGT
GTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGC
TGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGAC

CTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACC
GGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCC
TGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCG
CCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCC
TCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAG
CAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGT
CTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTAC
GCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCG
ACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTGCGT
GCTACCCGTACGCACGACACCCTGTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGCACG
TTGCGCAGTCCTGGGTGATGACGAACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCGCAC
TGGAAAGTGCCGATCCGGTGCCGGAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGACCGT
GCAGACGCTGCACGTAGCCTGACCGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACGCTG
GGCAACCACGCGTCTGGCTCTGGTGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGGATC
TGCCGGCAGCAGCTGTCTGGGGTCTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTGGCT
GATCTGGACGGTGATACCGCATCGGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACCGCA
GCTGGCACTGCGTGAAGGTGCAGTTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGGTGC
CGCCGCCGGGTACGCGCGATTGGCACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTGACC
CCGTTTCCGGAAGCAGCAGCTCCGCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCTGAA
CTTCCGTGATGTTGTCATGGCACTGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAGGTA
TCGTGACCGAAGCAGGTCCGGGTGTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGATGGT
TGCATCGGTCCGGTTGCCGTGGTTGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCCGCA
GGCAGCTTCAGTGCCGGTTACCTTCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTCCGG
GTGATCGTGTCCTGGTGCATGCAGCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCACCTG
GGTGCTGAAGTTTTTGCAACCGCTGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGACGA
TCATCTGGCGAGCTCTCGCACCGATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTATCG
ATGTCGTGCTGAATTCTCTGGCCGGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGCCGT
TTTATTGAAATGGGCAAAACCGATATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGTCTA
CCGTGCCTTCGACCTGATGGATGGTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTCTGT
TCGAAGGTGGCAAAATTCAGCTGGTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTTCGC
TATTTCGCACAAGCTCGTCATGTGGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGGTAC
CGTGCTGGTTACGGGTGCATCCGGTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATGATG
TGCGTCACCTGCTGCTGGCATCACGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAACTG
CGTGAATCTGGTGCACATAGTGTGCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGACCT
GCTGGCGTCTATTCCGGATGAACATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATGGCG
TGCTGGAAACCATGACGCCGGAACGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCATCTG
CACGAACTGACCCGCGACCGTGATCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGGTAG
TGCAGCTCAGGCAAACTACGCGGCCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATCACG
GTCTGCCGGCAACCTCTCTGGCTTGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGTCGT
GCAGACCTGGATCGTATGCAACGCTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTTTGA
CGCTGCACTGGCAGCAGGTCGTCCGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCGCTG
CGGGTGGCGGTTCTCTGCCGGCACCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCGCTG
CCGGCTGCGGATGCGCTGCGCGGTCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGA
CCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCT
TCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATG
AAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAAct
cgagCACGCTGAACCACAAGAAGCTCAGAGCGATTATTTTGGTGAACTTTTCCTGCAAGCAATGCGCACAG
GTGAACTGGCGCAGGCCCAGCAGCTTATGGCCGGTGCCGCGCAATTGCGCTTGAAATATGGTGATCCGGCT
GGTCCGGAGGCTGTGCCGGAGATTGTGCGTCTGGGCCGTGGGCAGCTGGGCCCGCAACTGATTCTGGTCTG
CCCGACCGTGATGACCACGGGCCCTCAGGTATATAGCCGCCTTGCAGAAGAGCTGGATGCGGGCCGCCGTG
TCAGCGCGTTGGTGCCACCGGGGTTTCACGGTGGTCAGGCGTTGCCGGCCACGCTGACAGTGCTGGTGCGC
TCTCTGGCAGACGTCGTCCAGGCGGAAGTAGCTGACGGCGAGTTCGCTCTGGCGGGGCACAGTTCGGGGGG
TGTAGTGGCGTACGAGGTGGCCCGTGAACTGGAGGCTcgtgggctggcgccccgcggaGTTGTCCTGATCG
ACAGCTACAGTTTTGACGGCGATGGTGGCCGCCCAGAAGAACTGTTCCGCTCGGCGCTGAATGAACGTTTT
GTGGAATATTTGCGCCTGACCGGCGGCGGGAACCTGAGCCAGCGCATTACGGCTCAGGTTTGGTGCTTGGA
ACTGCTGCGTGGCTGGCGTCCGGAAGGATTGACGGCGCCAACACTGTATGTTCGTCCGGCGCAACCTCTGG
TGGAGCAGGAAAAACCTGAGTGGCGCGGTGACGTGCTGGCCGCAATGGGGCAGGTCGTGGAAGCGCCGGGA

GATCATTTCACGATTATTGAGGGCGAACACGTCGCATCAACGGCCCACATCGTCGGTGACTGGCTGCGCGA
AGCCCACGCACATTATTCCACCGAAGGCTGGGGCGGTGGCCTCCGTACCGAAGAATAA

### 7.1.20. MBP-S2c-TE

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGATGTGCCGCCACTGTTTGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGT
GTTCGCTGGCCGCCTGGCGCTGCGTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGC
TGCCGGCGACCGCCTTCCTGGAACTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGAC

CTGACCCTGCACCGCCCGCTGGTTGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACC
GGAGGCGGACGGTCGTCGCGAGGTACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCC
TGCACGCCGCGGGTCTGCTGGCTTCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCG
CCGGAAGCTCAGGCCGTGGACCTGGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCC
TCTGTTTCAGGGTGTTCGTGCTGCCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAG
CAGCAGGCGCCGACGCAGCACGTTACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGT
CTGGGCGCCTTCCGTGAACGCTCTGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTAC
GCGTGGTGCAACCGCGGTGCGCGCGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCG
ACACCGCCGACCGTCCGGTGCTGACTGCCGAGAGCCTGGTTCTGCGCAGTGCCGCCGCGCGTCGTACGGGT
GCCCGTCGCCAGGCCCACCAAGCGCGCCTGTACCGTCTGAGTTGGCCGACCGTTCAGCTGCCAACCAGTGC
CCAGCCGCCGAGTTGTGTTCTGCTGGGTACCAGTGAAGTGAGTGCCGATATTCAAGTGTATCCAGACCTGC
GTAGCCTGACCGCGGCCCTGGATGCCGGTGCCGAGCCACCGGGTGTGGTTATCGCCCCGACCCCACCAGGC
GGTGGTCGCACGGCGGATGTGCGTGAAACCACCCGTCATGCGCTGGACCTGGTGCAAGGTTGGCTGAGTGA
TCAACGTCTGAATGAAAGTCGTCTGCTGCTGGTGACCCAAGGCGCCGTGGCCGTGGAACCAGGTGAGCCGG
TTACGGATCTGGCGCAAGCCGCGCTGTGGGGCCTGCTGCGTAGTACCCAAACCGAACACCCAGATCGTTTT
GTTCTGGTTGATGTTCCAGAACCGGCGCAACTGCTGCCGGCGCTGCCAGGCGTTCTGGCCTGTGGTGAACC
ACAACTGGCCCTGCGTCGCGGTGGCGCCCATGCCCCACGTCTGGCGGGTCTGGGCAGTGATGACGTTCTGC
CAGTTCCGGATGGTACCGGTTGGCGTCTGGAAGCCACCCGCCCAGGTAGCCTGGACGGTCTGGCGCTGGTG
GATGAACCGACCGCGACGGCCCCACTGGGCGACGGTGAAGTGCGTATTGCCATGCGTGCCGCGGGTGTTAA
CTTCCGTGATGCCCTGATTGCGCTGGGCATGTATCCGGGCGTGGCCAGTCTGGGTAGTGAAGGTGCCGGCG
TGGTTGTTGAAACGGGCCCAGGTGTTACGGGCCTGGCCCCAGGCGACCGTGTGATGGGTATGATTCCAAAA
GCGTTTGGTCCGCTGGCGGTGGCGGATCATCGTATGGTGACCCGTATTCCAGCCGGTTGGAGTTTTGCCCG
TGCGGCCAGTGTTCCAATTGTGTTTCTGACCGCCTACTACGCCCTGGTTGACCTGGCGGGTCTGCGTCCGG
GTGAAAGTCTGCTGGTTCACAGTGCGGCGGGTGGCGTTGGCATGGCCGCGATCCAACTGGCGCGTCATCTG
GGTGCGGAAGTGTATGCCACGGCGAGTGAGGATAAATGGCAAGCCGTGGAACTGAGTCGTGAACATCTGGC
GAGTAGTCGTACCTGTGATTTTGAACAACAATTTCTGGGCGCCACGGGCGGTCGCGGCGTTGATGTGGTGC
TGAATAGTCTGGCCGGTGAGTTCGCCGACGCCAGCCTGCGTATGCTGCCGCGTGGTGGCCGTTTTCTGGAG
CTGGGTAAAACCGATGTGCGTGACCCAGTTGAAGTTGCCGATGCCCATCCAGGCGTTAGTTACCAAGCCTT
TGATACGGTGGAAGCCGGTCCACAACGCATTGGTGAGATGCTGCATGAGCTGGTTGAGCTGTTTGAAGGTC
GTGTGCTGGAACCACTGCCAGTGACGGCCTGGGACGTTCGTCAAGCCCCAGAAGCCCTGCGTCACCTGAGT
CAAGCGCGCCATGTTGGTAAACTGGTTCTGACGATGCCACCAGTGTGGGATGCGGCCGGTACGGTTCTGGT
TACCGGCGGTACGGGTGCCCTGGGTGCCGAAGTGGCCCGTCATCTGGTTATTGAGCGTGGTGTTCGTAATC
TGGTGCTGGTGAGTCGTCGTGGCCCAGCCGCGAGCGGTGCCGCGGAGCTGGTTGCCCAACTGACCGCCTAC
GGTGCGGAAGTTAGTCTGCAAGCCTGTGATGTTGCCGATCGTGAAACCCTGGCCAAAGTGCTGGCGAGTAT
TCCGGATGAACATCCGCTGACGGCGGTTGTGCATGCGGCGGGTGTTCTGGACGATGGTGTTAGTGAGAGTC
TGACCGTTGAACGTCTGGACCAAGTTCTGCGTCCAAAAGTGGATGGTGCCCGCAATCTGCTGGAACTGATT
GATCCAGATGTGGCGCTGGTTCTGTTTAGTAGTGTGAGTGGCGTGCTGGGTAGTGGTGGTCAAGGTAACTA
TGCCGCCGCGAATAGCTTTCTGGATGCGCTGGCCCAACAACGTCAAAGTCGCGGTCTGCCAACGCGTAGCC
TGGCCTGGGGTCCGTGGGCCGAGCATGGTATGGCCAGTACCCTGCGCGAGGCCGAACAAGATCGTCTGGCC
CGCAGTGGCCTGCTGCCAATTAGTACCGAAGAGGGTCTGAGTCAATTCGATGCCGCGTGCGGTGGTGCGCA
TACCGTGGTTGCGCCAGTGCGTTTTAGTCGCCTGAGTGATGGTAATGCCATTAAGTTCAGTGTGCTGCAAG
GTCTGGTGGGTCCACATCGTGTGAATAAAGCGGCCACCGCGGATGATGCCGAAAGTCTGCGTAAACGCCTG
GCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCT
GGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCG
CAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCAC
CCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAActcgagCACGCTGAACCACAAGAAGCTCAGAG
CGATTATTTTGGTGAACTTTTCCTGCAAGCAATGCGCACAGGTGAACTGGCGCAGGCCCAGCAGCTTATGG
CCGGTGCCGCGCAATTGCGCTTGAAATATGGTGATCCGGCTGGTCCGGAGGCTGTGCCGGAGATTGTGCGT
CTGGGCCGTGGGCAGCTGGGCCCGCAACTGATTCTGGTCTGCCCGACCGTGATGACCACGGGCCCTCAGGT
ATATAGCCGCCTTGCAGAAGAGCTGGATGCGGGCCGCCGTGTCAGCGCGTTGGTGCCACCGGGGTTTCACG
GTGGTCAGGCGTTGCCGGCCACGCTGACAGTGCTGGTGCGCTCTCTGGCAGACGTCGTCCAGGCGGAAGTA
GCTGACGGCGAGTTCGCTCTGGCGGGGCACAGTTCGGGGGGTGTAGTGGCGTACGAGGTGGCCCGTGAACT
GGAGGCTcgtgggctggcgccccgcggaGTTGTCCTGATCGACAGCTACAGTTTTGACGGCGATGGTGGCC
GCCCAGAAGAACTGTTCCGCTCGGCGCTGAATGAACGTTTTGTGGAATATTTGCGCCTGACCGGCGGCGGG
AACCTGAGCCAGCGCATTACGGCTCAGGTTTGGTGCTTGGAACTGCTGCGTGGCTGGCGTCCGGAAGGATT
GACGGCGCCAACACTGTATGTTCGTCCGGCGCAACCTCTGGTGGAGCAGGAAAAACCTGAGTGGCGCGGTG
ACGTGCTGGCCGCAATGGGGCAGGTCGTGGAAGCGCCGGGAGATCATTTCACGATTATTGAGGGCGAACAC

GTCGCATCAACGGCCCACATCGTCGGTGACTGGCTGCGCGAAGCCCACGCACATTATTCCACCGAAGGCTG
GGGCGGTGGCCTCCGTACCGAAGAATAA

### 7.1.21. MBP-A1.5

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaaccttt acttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGATTCAAGCGGCACCGGATGCCGATGTGACGACGGCTGGTCTTGCTGGTCTTGGTCATCC
TCTCCTTGGTGCAAGCCTGGAACTGGCAGATGCCCAAGGTGCAGCTCTGTCGGGCCGTCTGAGCGCACGTA
CCGAATCCTGGCTGGCAGATCACGTGGTTCTGGGTTCAACCCTGGTTCCGGGTACGGCAGTCGTGGAAATG

GCAGTCCGTGCAGGTGCTGAAACCGGTTGCGGTCGCCTGGCAGAACTGACGCAGGAAGCGCCGCTGGCAGT
TCCGGAACGTGGTGCAGTCCATCTGCAGGTTCGTGTTGGTCCGGCAGGTGAACAAGGTCACCGTCCGGTCG
GTGTGTATTCCCGTCCGGAAGACGCCGAACCGGATGAACCGTGGGCATGTCATGCTCGTGGCGTGCTGGCA
CCGGAAGCAGCACCGGTTCCGGCAGGTACCGGCGGTGCTTGGCCGCCGTCAGGTGCAGAACCGGTGCCGCT
GGATGGCTTTTATGAACGCCTGGCAGCTGAAGGTTTTGCGTACGGCCCGGCATTCCAGGGTCTGACCCGCG
CCTGGCGTCTGGGCGATGAAGTTCTGGCAGAAATTACGCTGCCGGAAGGTGCATGCAGCGGTGCTGACCGT
TACGGTGTGCATCCGGCACTGCTGGATGCAGCACTGCACACCGCACTGCTGAAAGAAGAAGCTTCAGATAC
GTCGCAGGTTCGCATCCCGTTTGCCTGGCATGAAGTCTCGTTCCACGGCGGTAGCGCACCGGTGCTGCGTG
CACGTCTGACCCCGTCTGGTACCGACACCGTTAGTCTGGCCCTGTGGGATGAACATGGTACGCCGGTGGCA
AGCGTTGGCTCTCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTGCGTGCTACCCGTACGCACGACACCCT
GTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGCACGTTGCGCAGTCCTGGGTGATGACG
AACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCGCACTGGAAAGTGCCGATCCGGTGCCG
GAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGACCGTGCAGACGCTGCACGTAGCCTGAC
CGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACGCTGGGCAACCACGCGTCTGGCTCTGG
TGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGGATCTGCCGGCAGCAGCTGTCTGGGGT
CTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTGGCTGATCTGGACGGTGATACCGCATC
GGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACCGCAGCTGGCACTGCGTGAAGGTGCAG
TTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGGTGCCGCCGCCGGGTACGCGCGATTGG
CACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTGACCCCGTTTCCGGAAGCAGCAGCTCC
GCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCTGAACTTCCGTGATGTTGTCATGGCAC
TGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAGGTATCGTGACCGAAGCAGGTCCGGGT
GTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGATGGTTGCATCGGTCCGGTTGCCGTGGT
TGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCCGCAGGCAGCTTCAGTGCCGGTTACCT
TCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTCCGGGTGATCGTGTCCTGGTGCATGCA
GCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCACCTGGGTGCTGAAGTTTTTGCAACCGC
TGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGACGATCATCTGGCGAGCTCTCGCACCG
ATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTATCGATGTCGTGCTGAATTCTCTGGCC
GGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGCCGTTTTATTGAAATGGGCAAAACCGA
TATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGTCTACCGTGCCTTCGACCTGATGGATG
GTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTCTGTTCGAAGGTGGCAAAATTCAGCTG
GTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTTCGCTATTTCGCACAAGCTCGTCATGT
GGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGGTACCGTGCTGGTTACGGGTGCATCCG
GTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATGATGTGCGTCACCTGCTGCTGGCATCA
CGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAACTGCGTGAATCTGGTGCACATAGTGT
GCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGACCTGCTGGCGTCTATTCCGGATGAAC
ATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATGGCGTGCTGGAAACCATGACGCCGGAA
CGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCATCTGCACGAACTGACCCGCGACCGTGA
TCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGGTAGTGCAGCTCAGGCAAACTACGCGG
CCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATCACGGTCTGCCGGCAACCTCTCTGGCT
TGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGTCGTGCAGACCTGGATCGTATGCAACG
CTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTTTGACGCTGCACTGGCAGCAGGTCGTC
CGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCGCTGCGGGTGGCGGTTCTCTGCCGGCA
CCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCGCTGCCGGCTGCGGATGCGCTGCGCGG
TCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGG
GCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCT
CTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTA
CGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGA
CGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAG
CCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAG
CAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCC
GCTAA

### 7.1.22. MBP-S1.5

ATGCATCATCATCATCATCACAGCAGCGGCAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTGATTACGA
TATCCCAACGACCGAAAACCTTTACTTCCAGGGCCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGCTGGTGCCAAGCGACAGTGGTGATGTGACGGGTGCCGGTCTGGCCGGTGCCGAACACCC
ACTGCTGGGTGCGGTTGTTCCGGTTGCGGGTGGTGACGAAGTGCTGCTGACCGGCCGTATTAGTGTGCGTA
CCCATCCATGGCTGGCGGAGCACCGTGTTCTGGGCGAGGTTATTGTGGCCGGCACGGCCCTGCTGGAAATT

GCCCTGCATGCCGGCGAACGTCTGGGCTGTGAGCGTGTGGAGGAACTGACCCTGGAAGCCCCGCTGGTTCT
GCCGGAACGTGGTGCGATTCAAGTGCAACTGCGCGTGGGTGCCCCAGAAAATAGTGGTCGCCGTCCGATGG
CCCTGTATAGTCGTCCGGAAGGTGCCGCGGAACATGATTGGACCCGCCATGCCACGGGTCGCCTGGCCCCA
GGCCGTGGCGAAGCCGCCGGTGATCTGGCGGATTGGCCAGCCCCAGGTGCGCTGCCAGTGGATCTGGATGA
ATTTTACCGTGACCTGGCCGAGCTGGGTCTGGAATACGGTCCAATTTTTCAAGGTCTGAAGGCGGCGTGGC
GTCAAGGTGACGAAGTGTACGCCGAAGCCGCCCTGCCGGGTACCGAAGATAGTGGTTTTGGTGTTCATCCA
GCCCTGCTGGACGCGGCGCTGCACGCGACGGCGGTTCGTGATATGGATGATGCCCGCCTGCCATTTCAATG
GGAAGGTGTGAGTCTGCATGCGAAAGCGGCCCCAGCGCTGCGCGTGCGTGTTGTGCCAGCGGGCGATGACG
CGAAAAGTCTGCTGGTTTGTGACGGCACGGGTCGTCCAGTGATTAGTGTGGATCGTCTGGTTCTGCGCAGT
GCCGCCGCGCGTCGTACGGGTGCCCGTCGCCAGGCCCACCAAGCGCGCCTGTACCGTCTGAGTTGGCCGAC
CGTTCAGCTGCCAACCAGTGCCCAGCCGCCGAGTTGTGTTCTGCTGGGTACCAGTGAAGTGAGTGCCGATA
TTCAAGTGTATCCAGACCTGCGTAGCCTGACCGCGGCCCTGGATGCCGGTGCCGAGCCACCGGGTGTGGTT
ATCGCCCCGACCCCACCAGGCGGTGGTCGCACGGCGGATGTGCGTGAAACCACCCGTCATGCGCTGGACCT
GGTGCAAGGTTGGCTGAGTGATCAACGTCTGAATGAAAGTCGTCTGCTGCTGGTGACCCAAGGCGCCGTGG
CCGTGGAACCAGGTGAGCCGGTTACGGATCTGGCGCAAGCCGCGCTGTGGGGCCTGCTGCGTAGTACCCAA
ACCGAACACCCAGATCGTTTTGTTCTGGTTGATGTTCCAGAACCGGCGCAACTGCTGCCGGCGCTGCCAGG
CGTTCTGGCCTGTGGTGAACCACAACTGGCCCTGCGTCGCGGTGGCGCCCATGCCCCACGTCTGGCGGGTC
TGGGCAGTGATGACGTTCTGCCAGTTCCGGATGGTACCGGTTGGCGTCTGGAAGCCACCCGCCCAGGTAGC
CTGGACGGTCTGGCGCTGGTGGATGAACCGACCGCGACGGCCCCACTGGGCGACGGTGAAGTGCGTATTGC
CATGCGTGCCGCGGGTGTTAACTTCCGTGATGCCCTGATTGCGCTGGGCATGTATCCGGGCGTGGCCAGTC
TGGGTAGTGAAGGTGCCGGCGTGGTTGTTGAAACGGGCCCAGGTGTTACGGGCCTGGCCCCAGGCGACCGT
GTGATGGGTATGATTCCAAAAGCGTTTGGTCCGCTGGCGGTGGCGGATCATCGTATGGTGACCCGTATTCC
AGCCGGTTGGAGTTTTGCCCGTGCGGCCAGTGTTCCAATTGTGTTTCTGACCGCCTACTACGCCCTGGTTG
ACCTGGCGGGTCTGCGTCCGGGTGAAAGTCTGCTGGTTCACAGTGCGGCGGGTGGCGTTGGCATGGCCGCG
ATCCAACTGGCGCGTCATCTGGGTGCGGAAGTGTATGCCACGGCGAGTGAGGATAAATGGCAAGCCGTGGA
ACTGAGTCGTGAACATCTGGCGAGTAGTCGTACCTGTGATTTTGAACAACAATTTCTGGGCGCCACGGGCG
GTCGCGGCGTTGATGTGGTGCTGAATAGTCTGGCCGGTGAGTTCGCCGACGCCAGCCTGCGTATGCTGCCG
CGTGGTGGCCGTTTTCTGGAGCTGGGTAAAACCGATGTGCGTGACCCAGTTGAAGTTGCCGATGCCCATCC
AGGCGTTAGTTACCAAGCCTTTGATACGGTGGAAGCCGGTCCACAACGCATTGGTGAGATGCTGCATGAGC
TGGTTGAGCTGTTTGAAGGTCGTGTGCTGGAACCACTGCCAGTGACGGCCTGGGACGTTCGTCAAGCCCCA
GAAGCCCTGCGTCACCTGAGTCAAGCGCGCCATGTTGGTAAACTGGTTCTGACGATGCCACCAGTGTGGGA
TGCGGCCGGTACGGTTCTGGTTACCGGCGGTACGGGTGCCCTGGGTGCCGAAGTGGCCCGTCATCTGGTTA
TTGAGCGTGGTGTTCGTAATCTGGTGCTGGTGAGTCGTCGTGGCCCAGCCGCGAGCGGTGCCGCGGAGCTG
GTTGCCCAACTGACCGCCTACGGTGCGGAAGTTAGTCTGCAAGCCTGTGATGTTGCCGATCGTGAAACCCT
GGCCAAAGTGCTGGCGAGTATTCCGGATGAACATCCGCTGACGGCGGTTGTGCATGCGGCGGGTGTTCTGG
ACGATGGTGTTAGTGAGAGTCTGACCGTTGAACGTCTGGACCAAGTTCTGCGTCCAAAAGTGGATGGTGCC
CGCAATCTGCTGGAACTGATTGATCCAGATGTGGCGCTGGTTCTGTTTAGTAGTGTGAGTGGCGTGCTGGG
TAGTGGTGGTCAAGGTAACTATGCCGCCGCGAATAGCTTTCTGGATGCGCTGGCCCAACAACGTCAAAGTC
GCGGTCTGCCAACGCGTAGCCTGGCCTGGGGTCCGTGGGCCGAGCATGGTATGGCCAGTACCCTGCGCGAG
GCCGAACAAGATCGTCTGGCCCGCAGTGGCCTGCTGCCAATTAGTACCGAAGAGGGTCTGAGTCAATTCGA
TGCCGCGTGCGGTGGTGCGCATACCGTGGTTGCGCCAGTGCGTTTTAGTCGCCTGAGTGATGGTAATGCCA
TTAAGTTCAGTGTGCTGCAAGGTCTGGTGGGTCCACATCGTGTGAATAAAGCGGCCACCGCGGATGATGCC
GAAAGTCTGCGTAAACGCCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGA
AGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGA
TTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCG
GCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGA
TGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAG
CGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAA
CCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGA
GCTGGGCATCAGCCGCTAA

### 7.1.23. MBP-A1.5c

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGATGATGAATACTGGCTCGGCCGCCGAACCGGCCGAATTAGGTCTCGGCGACGCCCGTCA
TCCTCTCCTGGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGTGTTCGCTGGCCGCCTGGCGCTGC
GTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGCTGCCGGCGACCGCCTTCCTGGAA

CTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGACCTGACCCTGCACCGCCCGCTGGT
TGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACCGGAGGCGGACGGTCGTCGCGAGG
TACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCCTGCACGCCGCGGGTCTGCTGGCT
TCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCGCCGGAAGCTCAGGCCGTGGACCT
GGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCCTCTGTTTCAGGGTGTTCGTGCTG
CCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAGCAGCAGGCGCCGACGCAGCACGT
TACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGTCTGGGCGCCTTCCGTGAACGCTC
TGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTACGCGTGGTGCAACCGCGGTGCGCG
CGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCGACACCGCCGACCGTCCGGTGCTG
ACTGCCGAGAGCCTGGTTAGTCGTCCGGTCTCCGCACGTCAGCTGCGTGCTACCCGTACGCACGACACCCT
GTTTCGTCTGGATTGGGTGGAAACCACGATTACGCCGGCAGCTGCACGTTGCGCAGTCCTGGGTGATGACG
AACTGGCAGGTGCACTGTCTGTGCCGGCATTCGCTGACCTGGCCGCACTGGAAAGTGCCGATCCGGTGCCG
GAACTGGTTCTGTATCCGTGTCTGGGCGATGACGCAGAAGATGACCGTGCAGACGCTGCACGTAGCCTGAC
CGCGCGTGTCCTGGGTGTGCTGCAGGCGTGGGTTGCCGATGAACGCTGGGCAACCACGCGTCTGGCTCTGG
TGACCCGTGGCGCAATGTCGGTGACCGACCGTGAACAGGTTACGGATCTGCCGGCAGCAGCTGTCTGGGGT
CTGGTGCGTAGCGCGCAAGCCGAACATCCGGGTCGCTTTGTGCTGGCTGATCTGGACGGTGATACCGCATC
GGCAGCAGCACTGCCGGGTATTCTGGCTGCAAGCGGTGATGAACCGCAGCTGGCACTGCGTGAAGGTGCAG
TTCTGGTCCCGCGTCTGGCGCGCGGTGTCCCGAGTGGTACCCTGGTGCCGCCGCCGGGTACGCGCGATTGG
CACCTGGAACTGACCGGCGGTGGCACGGTTGATGACCTGGCACTGACCCCGTTTCCGGAAGCAGCAGCTCC
GCTGGCACCGGGTCAAGTGCGTGTTGCAGTCCGTGCAGCAGGTCTGAACTTCCGTGATGTTGTCATGGCAC
TGGGTATGGTGGATGACCGCCGTGCACTGGGTGGCGAAATTGCAGGTATCGTGACCGAAGCAGGTCCGGGT
GTTACGGGTTTTGCACCGGGTGACCGTGTGTTCGGTCTGGCAGATGGTTGCATCGGTCCGGTTGCCGTGGT
TGATCATCGTCTGATTGCACGCATCCCGGAAGGCTGGTCCTTTCCGCAGGCAGCTTCAGTGCCGGTTACCT
TCCTGACGGCGTATTACGGTCTGGTTGACCTGGCCGGTGTCCGTCCGGGTGATCGTGTCCTGGTGCATGCA
GCAGCAGGTGGCGTGGGTATGGCTGCAGTTCAACTGGCACGTCACCTGGGTGCTGAAGTTTTTGCAACCGC
TGGTCCGGCAAAATGGGACACGGTGCGTGCACTGGGTATTGATGACGATCATCTGGCGAGCTCTCGCACCG
ATGAATTTGAAACGCGTTTCGCCGCAGAAGACGGTGGCCGTGGTATCGATGTCGTGCTGAATTCTCTGGCC
GGCGAAATGGCGGATGCAAGTCTGCGTCTGGTTCGTCCGGGTGGCCGTTTTATTGAAATGGGCAAAACCGA
TATCCGCGACGCGGATGAAGTGGCTGCGGCCTATGAAGGTGTTGTCTACCGTGCCTTCGACCTGATGGATG
GTGGCGCGGAATGTATTGCCCGCATCTTTGCAGAACTGCTGGCTCTGTTCGAAGGTGGCAAAATTCAGCTG
GTTCCGGTCACCACGTGGGATGTTCGCCAGGCACCGGAAGCTTTTCGCTATTTCGCACAAGCTCGTCATGT
GGGCAAAATCGTTCTGACCGTCCCGCCGGCATGGGATCCGGAAGGTACCGTGCTGGTTACGGGTGCATCCG
GTGGCGTTGCAGCTCATCTGGTTCGTCACCTGGTCCGCACCCATGATGTGCGTCACCTGCTGCTGGCATCA
CGCCGTGGTCCGGACGCAGAGGGTATGGATGAACTGATTGCCGAACTGCGTGAATCTGGTGCACATAGTGT
GCGCGCGGTTGCCTGCGACTGTGTCGATCGCACCGCAGTGGCTGACCTGCTGGCGTCTATTCCGGATGAAC
ATCCGCTGACCGCCGTGGTTCACACGGTCGGTGTCGTGGACGATGGCGTGCTGGAAACCATGACGCCGGAA
CGCATCGACGCGGTGTTTCGTCCGAAAGCGGATGGCGCCTGGCATCTGCACGAACTGACCCGCGACCGTGA
TCTGGCAGCATTCGCTGTGTGTAGTTCCGTTGCCGGTACGCTGGGTAGTGCAGCTCAGGCAAACTACGCGG
CCGCAAATGCGTTTCTGGACGCACTGGCTGCACATCGCCGTGATCACGGTCTGCCGGCAACCTCTCTGGCT
TGGGGCATGTGGGCCGGTACGGGTGGCATGGCAGCAAATCTGAGTCGTGCAGACCTGGATCGTATGCAACG
CTCCGGTATCTCAGGCCTGTCGACCGAAGAAGGCCTGGCACTGTTTGACGCTGCACTGGCAGCAGGTCGTC
CGGTGTGGCTGCCGGCACGTCTGGATGCAAAAGCACTGCGTACCGCTGCGGGTGGCGGTTCTCTGCCGGCA
CCGCTGCGCGGTCTGGTCCATGTGCCGGCAGCAGACGCAGGTCCGCTGCCGGCTGCGGATGCGCTGCGCGG
TCGTCTGGCTGCGTTGCCGGAAGCAGACCGTCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGG
GCGTGCTGGGCCATGAAACGCGTGCGGCGATCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCT
CTGACCGCAGTTGAACTGCGCAATCTGCTGACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTA
CGACCACCCGACCCTGAGCTCCCTGGCGGACCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGA
CGGACACGGCTGCAGACCTGTTGGCGGAATTGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAG
CCGGAAGCGCGTGCACGTATCGGTCGCCGTCTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAG
CAGCCGCGACTTGAAATCTGCCAGCCGTACCGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCC
GCTAA

### 7.1.24. MBP-S1.5c

ATGCATCATCATCATCATCACagcagcggcAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGGCCCACGAGGACAAACTCAGACATCTGCTCA
AACGCGTAAGCGCGGAGTTGGACGACACCCAGCGTCGTGTTCGTGAAATGGAGGAGAGCGAGCGTGAACCG
ATCGCGATTGTTGGCATGTCGTGCCGTCTGCCGGGTGGTGTGAATAGCCCGGGTGAATTTTGGAGCTTGCT
GGAAGCTGGTACGGATGCAGTTTCTGAGTTCCCGCGTGACCGTGGCTGGGATGTCGAAAATCTGTATGATC
CGGACCCGGATGCACCTGGTCGCTCCTACGTCCGTGAAGGCGGTTTCCTGGACGGTGCCGGTCAATTTGAT
GCCGCCTTCTTCGGTATTAGCCCTCGTGAGGCACTGGCCATGGACCCGCAACAGCGCCTGCTGCTGGAGTG
CTCCTGGGAAGCAATTGAGCGTAGCCGCATCGATCCGAAAACTCTGCATGGTAGCCGTACGGGTGTTTTCG
CGGGTAGCAACTGGCAAGACTACAATACGCTGCTGCTGAATGCAGAGGAACGTAGCCAGTCGTACCTGGCG
ACCGGTGCGTCGGGCAGCGTCCTGTCTGGCCGTGTCAGCTATACCCTGGGTATGGAAGGTCCGGCGATTAC
GGTGAACACCGCTTGTTCCTCCAGCCTGGTGGCAGTGCACTTGGCAGCGCGTAGCCTGCGTGCGGGCGAGT
GTGACCTGGCGTTGGCTGGCGCAGTCACCGTTATGAGCACCCCGCAGCTGCCGGTTGCGTTTTCTCGTCAG
CGTGGTCTGGCTCCGGACGGTCGTAGCAAGGCGTTCGCGGTTAGCGCGGACGGTATGGGTTTTGGTGAAGG
TGTGGGCGTTCTGGTCCTGGAGCGTCTGAGCGTCGCACGCCGCAATGGTCATCGCGTCCTGGCAGTTGTGC
GTGGTTCCGCGGTGAACCAAGATGGTGCGTCCAACGGCTTGACGGCACCGAATGGCCCGAGCCAACAACGT
GTCATTCGTGCGGCGCTGGCAAGCGCCGGTCTGGGTCCGGCTGACGTGGATGTTGTCGAAGCGCACGGCAC
TGGCACCCGCCTGGGCGACCCGATCGAGGCGCAGGCACTGTTGGCCACCTATGGCCGTGGCCGTGATGCGG
AGCGTCCGCTGTGGTTGGGCAGCGTGAAGAGCAACATCGGTCACGCACAGGCTGCAGCGGGCGTTGCGGGT
GTGATCAAAATGGTCCTGGCGATGGAGAAGGGTCGCCTGCCACGTACGCTGCATGTGGATGAGCCAAGCGG
TGAAGTGGACTGGGACAGCGGTGCCGTTCGCCTGCTGACCGAAGCCCGTGATTGGCCGAGCGAAGAAGGTC
GTTTGCGTCGCGCTGGCGTGTCCAGCTTCGGCATCTCTGGCACCAACGCGCACGTGATCATTGAAGAGGCG
CCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGGA
AGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGTT
GGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCTG
GTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTCT
GGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCGG
TGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAGC
GACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTCG
TCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCAT
GCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGCC
CTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGGA
TGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGTG
TGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGAG
CCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCAG
CAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGCG
AACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGGC
CCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCCC
GACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGCC
ATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCGACCTACCCGTTCCAACAT
CAACACTATTGGATGATGAATACTGGCTCGGCCGCCGAACCGGCCGAATTAGGTCTCGGCGACGCCCGTCA
TCCTCTCCTGGGTAGCGTTGTTACCGTGGCTGGCGATGACAAAGTTGTGTTCGCTGGCCGCCTGGCGCTGC
GTACCCATCCATGGCTGGCGGATCACACGGTTCTGGACGCCGTTCTGCTGCCGGCGACCGCCTTCCTGGAA

CTGGCGGTTCGTGCGGGTGAGGAGGTTAGCTGCCCGGTTGTTCACGACCTGACCCTGCACCGCCCGCTGGT
TGTGCCGGAACGCGGTGCGGTTCAGGTACAAATGGCAGTAGGTGCACCGGAGGCGGACGGTCGTCGCGAGG
TACGTGTTTACTCTCGCCCGGACGACGACGCAGAACACGAATGGACCCTGCACGCCGCGGGTCTGCTGGCT
TCTGCGGCTACCGCAGAACCGGCTGTGGCGGCCGGTGCCTGGCCGCCGCCGGAAGCTCAGGCCGTGGACCT
GGATGGTTTCTACGCTGGTCTGGCAGAACACGGTTACCACTACGGCCCTCTGTTTCAGGGTGTTCGTGCTG
CCTGGCGTCTGGGTGACGACGTTCTGGCGGAAATTGTGCTGCCGGAAGCAGCAGGCGCCGACGCAGCACGT
TACGGTATGCACCCAGCACTGCTGGACGCGGTGCTGCACGCAGCTCGTCTGGGCGCCTTCCGTGAACGCTC
TGAAGAAAAATATCTGCCGTTCGCGTGGGAGGGCGTGACCCTGCGTACGCGTGGTGCAACCGCGGTGCGCG
CGCGTATTAGCCGCGCCGGTACCGATGCAATCCGTCTGGACGTGACCGACACCGCCGACCGTCCGGTGCTG
ACTGCCGAGAGCCTGGTTCTGCGCAGTGCCGCCGCGCGTCGTACGGGTGCCCGTCGCCAGGCCCACCAAGC
GCGCCTGTACCGTCTGAGTTGGCCGACCGTTCAGCTGCCAACCAGTGCCCAGCCGCCGAGTTGTGTTCTGC
TGGGTACCAGTGAAGTGAGTGCCGATATTCAAGTGTATCCAGACCTGCGTAGCCTGACCGCGGCCCTGGAT
GCCGGTGCCGAGCCACCGGGTGTGGTTATCGCCCCGACCCCACCAGGCGGTGGTCGCACGGCGGATGTGCG
TGAAACCACCCGTCATGCGCTGGACCTGGTGCAAGGTTGGCTGAGTGATCAACGTCTGAATGAAAGTCGTC
TGCTGCTGGTGACCCAAGGCGCCGTGGCCGTGGAACCAGGTGAGCCGGTTACGGATCTGGCGCAAGCCGCG
CTGTGGGGCCTGCTGCGTAGTACCCAAACCGAACACCCAGATCGTTTTGTTCTGGTTGATGTTCCAGAACC
GGCGCAACTGCTGCCGGCGCTGCCAGGCGTTCTGGCCTGTGGTGAACCACAACTGGCCCTGCGTCGCGGTG
GCGCCCATGCCCCACGTCTGGCGGGTCTGGGCAGTGATGACGTTCTGCCAGTTCCGGATGGTACCGGTTGG
CGTCTGGAAGCCACCCGCCCAGGTAGCCTGGACGGTCTGGCGCTGGTGGATGAACCGACCGCGACGGCCCC
ACTGGGCGACGGTGAAGTGCGTATTGCCATGCGTGCCGCGGGTGTTAACTTCCGTGATGCCCTGATTGCGC
TGGGCATGTATCCGGGCGTGGCCAGTCTGGGTAGTGAAGGTGCCGGCGTGGTTGTTGAAACGGGCCCAGGT
GTTACGGGCCTGGCCCCAGGCGACCGTGTGATGGGTATGATTCCAAAAGCGTTTGGTCCGCTGGCGGTGGC
GGATCATCGTATGGTGACCCGTATTCCAGCCGGTTGGAGTTTTGCCCGTGCGGCCAGTGTTCCAATTGTGT
TTCTGACCGCCTACTACGCCCTGGTTGACCTGGCGGGTCTGCGTCCGGGTGAAAGTCTGCTGGTTCACAGT
GCGGCGGGTGGCGTTGGCATGGCCGCGATCCAACTGGCGCGTCATCTGGGTGCGGAAGTGTATGCCACGGC
GAGTGAGGATAAATGGCAAGCCGTGGAACTGAGTCGTGAACATCTGGCGAGTAGTCGTACCTGTGATTTTG
AACAACAATTTCTGGGCGCCACGGGCGGTCGCGGCGTTGATGTGGTGCTGAATAGTCTGGCCGGTGAGTTC
GCCGACGCCAGCCTGCGTATGCTGCCGCGTGGTGGCCGTTTTCTGGAGCTGGGTAAAACCGATGTGCGTGA
CCCAGTTGAAGTTGCCGATGCCCATCCAGGCGTTAGTTACCAAGCCTTTGATACGGTGGAAGCCGGTCCAC
AACGCATTGGTGAGATGCTGCATGAGCTGGTTGAGCTGTTTGAAGGTCGTGTGCTGGAACCACTGCCAGTG
ACGGCCTGGGACGTTCGTCAAGCCCCAGAAGCCCTGCGTCACCTGAGTCAAGCGCGCCATGTTGGTAAACT
GGTTCTGACGATGCCACCAGTGTGGGATGCGGCCGGTACGGTTCTGGTTACCGGCGGTACGGGTGCCCTGG
GTGCCGAAGTGGCCCGTCATCTGGTTATTGAGCGTGGTGTTCGTAATCTGGTGCTGGTGAGTCGTCGTGGC
CCAGCCGCGAGCGGTGCCGCGGAGCTGGTTGCCCAACTGACCGCCTACGGTGCGGAAGTTAGTCTGCAAGC
CTGTGATGTTGCCGATCGTGAAACCCTGGCCAAAGTGCTGGCGAGTATTCCGGATGAACATCCGCTGACGG
CGGTTGTGCATGCGGCGGGTGTTCTGGACGATGGTGTTAGTGAGAGTCTGACCGTTGAACGTCTGGACCAA
GTTCTGCGTCCAAAAGTGGATGGTGCCCGCAATCTGCTGGAACTGATTGATCCAGATGTGGCGCTGGTTCT
GTTTAGTAGTGTGAGTGGCGTGCTGGGTAGTGGTGGTCAAGGTAACTATGCCGCCGCGAATAGCTTTCTGG
ATGCGCTGGCCCAACAACGTCAAAGTCGCGGTCTGCCAACGCGTAGCCTGGCCTGGGGTCCGTGGGCCGAG
CATGGTATGGCCAGTACCCTGCGCGAGGCCGAACAAGATCGTCTGGCCCGCAGTGGCCTGCTGCCAATTAG
TACCGAAGAGGGTCTGAGTCAATTCGATGCCGCGTGCGGTGGTGCGCATACCGTGGTTGCGCCAGTGCGTT
TTAGTCGCCTGAGTGATGGTAATGCCATTAAGTTCAGTGTGCTGCAAGGTCTGGTGGGTCCACATCGTGTG
AATAAAGCGGCCACCGCGGATGATGCCGAAAGTCTGCGTAAACGCCTGGCTGCGTTGCCGGAAGCAGACCG
TCGCCGTGCGGTCTTGGACCTGGTTGAAGAACTGGTGCTGGGCGTGCTGGGCCATGAAACGCGTGCGGCGA
TCGGTCCGGATAGCAGCTTCCACGCGATTGGTTTCGACTCTCTGACCGCAGTTGAACTGCGCAATCTGCTG
ACCGTTCGTCTGGGTATGAAGCTGCCGGCGACTCTGGTGTACGACCACCCGACCCTGAGCTCCCTGGCGGA
CCACTTGCACGAACAACTGGTTATTGATGGTACCCCGATGACGGACACGGCTGCAGACCTGTTGGCGGAAT
TGGATGCGTTGGCAGCCCGCCTGGCAGCGGTCGGTCTGGAGCCGGAAGCGCGTGCACGTATCGGTCGCCGT
CTGAAGGATATGCAGACCGCGTGTGAACCGCGCTCCGAAAGCAGCCGCGACTTGAAATCTGCCAGCCGTAC
CGAAGTGCTGGACTTCCTGACCAATGAGCTGGGCATCAGCCGCTAA

### 7.1.25. MBP-BorMod5-AT1

ATGCATCATCATCATCATCACAGCAGCGGCAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGA
TAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTG
AGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTT
CCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCG
CTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATC
CCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAG
ACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACAC
ATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAA
CGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACC
GCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGG
AAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACT
GCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTgattacga
tatcccaacgaccgaaaacctttacttccagggcCATATGACCAGCGTGGGTCGCCGCCAACTGCATGTTC
ACCCGGAAGACCACGATCACCATGCCAAACGCACAGGTCTGAGTATGGGTGACGAACAGAAACTGCGCACC
TACCTGCGTCGTGTGACAGCCGATTTAGCCGACGTTACCGAACGCCTGCAGCGTGCCGAAGACAAAAATGC
AGAGCCGATTGCCATTGTGGGCATGGGTTGTCGCTATCCTGGCGGTGTTCGTAGCCCGGAGGAGTTTTGGA
ATCTGTTAGATGAGGGCGTGGACGCAGTTGCCGGCTTTCCTGAGGATCGTGGCTGGGACCTGGAGAATCTG
TACGATCCGGACCCGGACGAACCGGGTAAATGCTATGCACGCGAGGGCGGTTTTCTGTATGACGCCGGTGA
GTTCGATGCCGCCTTCTTTGGTATCAGCCCTCGCGAAGCACTGAGCATGGACCCTCAACAGCGCCTGCTGC
TGGAGTGCAGTTGGAGTGCACTGGAGCGTGCAGGTATTGACCCGGGCAGCTTACGCGGCAAGGACGTTGGC
GTTTATGTGGGTGCATGGAACAGCAACTATGGTCGCGGTGGCGGCGCCGAAAGCAGTGAGGGCCACCTGCT
GACAGGCAACGCCAGCAGTGTTGTTAGCGGTCGCGTTGCCTACGTTCTGGGTCTGGAAGGCCCGGCAGTTA
CCATTGACACCGCCTGTAGTAGTAGCCTGGTTGGCCTGCACCTGGCCGCACAAGCACTGCGTAGCGGTGAA
TGTGGCCTGGCATTAGCAGGTGGTGTGACAGTTATGAGTACCCCGCTGAGTCTGGTTAGCTTTAGCCGCCA
ACGTGGTCTGGCCCAAGACGGCCGTAGTAAGGCCTTTAGTGCAGATGCCGACGGTATGGGCATGGCAGAAG
GCGTGGGTGTGCTGGTGCTGGAGCGTCTGAGTGAAGCACGTCGCAACGGCCACGAAGTGCTGGCAGTGCTG
CGTAGTAGCGCCGTTAATCAGGATGGTGCCAGCAACGGTCTGAGCGCCCCTAATGGTCCTGCACAGCAGCG
TGTTATCCAGAGTGCCCTGACAGTGGGTCGCCTGGCACCTAGTGATATTGACGTGGTTGAGGCACACGGTA
CCGGTACCGCACTGGGCGACCCGATCGAGGCACAGGCCCTGTTAGCCACCTATGGCCGTGGTCGCGACGCA
GACCGTCCTTTATGGCTGGGTAGCGTGAAGAGCAATATCGGTCACACCCAAGCAGCAGCCGGTGTGGCAGG
CGTTATTAAAATGGTGCTGGCCCTGCGTAAAGGTGTTTTACCGCGCACCCTGCATGTGGATGAGCCGACAG
GTGAAGTTGATTGGGACAGCGGCGCAGTTCGTCTGTTAACCGAAGCACGCGAATGGCCTAGCGGCGAAGGT
CGTGTGCGTCGTGCCGGTGTGAGTAGTTTTGGCATTAGCGGCACCAACGCGCACGTGATCATTGAAGAGGC
GCCTGAAGAGGGCGAAGAACCGGAGTCTGACGCCGGTGGTGTTGTGCCGTGGGTTTTGTCTGCTCGTACGG
AAGGCGCGCTGCAGGCCCAGGCTGTTCAGTTGAGCGAGTTCGTTGGTGAGAGCTCGCCGGTCGACGTGGGT
TGGTCCCTGGTTTCTACTCGTGCGGCCTTCGAGCATCGTGCGGTGGTGGTTGGCCGTGGTCGCGACGAGCT
GGTACGTGGTCTGAGCGAGGTCGCGCAGGGTCGTGGTGTTCGTGGCGTGGCAAGCAGCGCGTCCGGTGGTC
TGGCTTTCGTTTTCGCAGGCCAGGGTAGCCAGCGCCTGGGCATGGGTCGCGGCTTGTATGAGCGTTTCCCG
GTGTTTGCAGAAGCGTTTGATGAGGTGTGTGGTCGCGTTGGTCCGGGCGTACGCGAGGTTGTTTTTGGTAG
CGACGCGGGTGAGCTGGACCGTACCGTTTGGGCACAAGCCGGCCTGTTCGCCCTGGAAGTCGCGCTGTTTC
GTCTGTTGGAGAGCTGGGGCGTTCGTCCGGGTTGCCTGATTGGCCACTCCGTCGGCGAGCTGTCCGCGGCA
TGCGTGGCGGGTCTGTGGAGCCTGGAGGACGCGTGCCGCGTTGTGGCCGCTCGTGCACGTCTGATGCAGGC
CCTGCCGGCTGGCGGCGTCATGGTTGCTGTTCGCGCGGAAGCTGGCGAGCTGGCCGGTTTCCTGGGTGAGG
ATGTGGTGATTGCAAGCGTGAACGCGCCAGGTCAAGTCGTGATCGCGGGTCCGGAAGGCGGCGTCGAGCGT
GTGGTTGCTGCATGCGGTGCCCGTTCGCGCCGTCTGGCGGTGAGCCATGCGTTCCATTCCCCGCTGGTGGA
GCCGATGCTGGGTGAGTTTCGCCGTGTCGTTGAGAGCGTGGCATTTGGCGTCCCGAGCCTGCGTGTCGTCA
GCAACGTGACCGGTGCGTGGGTTGATCCGGAAGAGTGGGGCACCCCGGAGTACTGGGTGCGTCAGGTGCGC
GAACCGGTCCGTTTTGCGGATGGTGTCGCGACGTTGCTGGATGCCGGCGTCCGTACCTTTGTTGAGTTGGG
CCCAGCGGGTGCGCTGACCAGCATGGTTAGCCACTGTGCCGACGCGACTGCGACGTCTGTGACCGCCGTCC
CGACGCTGCGTCCTGACCACGACGAAAGCCGTACGGTTCTGAGCGCTGCAGCCAGCCTGTATGTGCAGGGC
CATCCGGTTGATTGGGCGCCTCTGTTCCCACGCGCACGCACCGTTGATCTGCCTACCTACCCGTTTTTACG
CGAACGTCTGTGGTTAGAACCGAGTCGCAGCCGTACAGGCAATTTAAACATGGCAGGCCTGGTGGAAGCCG
GCCATGAAATTTTACCGGCAGCCGTTGAGCTGCCGGGCGAACAATGGGTGTGGACCGGCGAATTAAGTCTG

AGCGCATACCCGTGGCTGGCAGACCATCAAGTGCTGGGTCAAACCCTGGTGCCTGGCGTTGCATGGGTGGA
GCTGGCATTACACGCAGGCCATCAGCTGGGCTTCGGTAGCGTGGAAGAACTGACACTGCAGGCACCTCTGG
TGCTGGGTGAGAGCGACGCAGTGCAAGTTCGTGTGGTTGTGAGTGACCTGGGCGAAAGCGATCGTCGTGCC
GTGAGCGTTCACAGTCGCGGCGACGATCAGACCTGGGTGACACACGCCGAAGGCTTTCTGACCGCAAAAGG
CGCCCAGCCGGAAACAATGGCCGTGTGGCCGCCTAGTGGCGCAGAGCCTGTGGAAGCCGATGGCTTTTATG
AACGTCTGGCCGATGCCGGTTACCATTACGGTCCGGTTTTCCAGGGTGTGAGCAAGGTGTGGCGTGCCGGT
GAAGAAATATACGCCGAGGTTGGTCTGCTGGACGACGCAGATGTTGACGGTTTCGGCATCCATCCGGCCTT
ACTGGATGCCGCATTACAAACCGCGTACGTGGCCCAGCGTGGCCCTGCAGAGACCAAATTACCGTTCGCCT
TTGGCGATGTGCAACTGTTCGCCACCGGCGCCCGTAGTTTACGCGTTCGCGTTAGCCCTGCAGCCCAGCAA
GGCATGGCATGGGAAGCCTGGGACCCGACCGGTCTGCCTGTGTTTAGCCTGGGTTACCTGGCAACACGTCC
GGTTGATCGTGGTCAGCTGACCGTGAAACGCCCGGAGAGTTTATTTAAAGTGGCATGGGACGAGACAGTTC
CGGTTGTTGGTAATGCCACCGCCGCACATGGTGTTGTGCTGGGTGATGATCCTTTTGCCCTGGGCGCAGCA
TTACGCGCCGCAGGTTGGGAAGTGGGTGCAGCCCCGGAGCCTGCAAGCGCAGATACCGCCGCCGAAGTTTT
ATTACTGCCGTGTACCGCACCGGGCGAGCCTGACGCGGACCTGCCTACAGCAGTGCGCGCAGTTACCGCCC
GTGTGCTGGGTGTGCTGCAAGAATGGCTGGCAGATGAGCGTCTGGCCGGTACCCGTCTGGCCGTTGTGACA
CGTAATGCCCTGCCGGGTGACTTACTGCATAGCCCTGTTTGGGGCCTGGTGCGTAGCGCACAGACCGAAAA
TCCGGGCCGCATCACATTAGTTGATTTAGATGATCATCCTGATAGCGCAGCCGTTTTAGCAGAAGCAGTTC
AAAGTGATGAGCCGCGTATTATGGTGCGCGAAGGCCGTCCTACAGCCGCCCGTCTGGTGCGTGCCACCGCA
CCTGAATTAGTGCCGCCGGCAGGTGCCGATGCATGGCGCCTGGAGATTACAGAACCGGGCACCTTTGACAA
CCTGACCCTGGGCGTGTACCCGCATGCCGAGAAGACCCTGGCCGATAATGAGGTGCGCGTGGCAGTGCACG
CAGGCGGTCTGAACTTCCACGATGTGGTTGCCGCACTGGGCATGGTTGAAGATGACCTGACCTTAGGTCGC
GAGGCAGCCGGCGTTGTTGTTGAAGTGGGTGACGCAGTGCCTGACCTGACCCCGGGTGATCACGTGATGGG
TATTCTGAGCAGCGGTTTCGGCCCTCTGGCAGTTACCGACCACCGCTATCTGGCACGTATGCCGGAAGGTT
GGACATTCGCACAGGCCGCAAGCGTTCCGGCCGCCTTCTTAACAGCCTATTATGGTCTGTGCGACCTGGGT
GGCATTCGCGCAGGTGACCGCGTGTTAATTCACGCAGCCGCCGGTGGCGTTGGCATGGCAGCAGTGCAGAT
CGCCCGCCACCTGGGTGCAGAAGTTTTTGGTACCGCCAGCCCGCGTAAGTGGGGTGCATTACGTGCACTGG
GCCTGGACGACGCCCATCTGAGCAGTAGCCGTACCTTAGACTTCGAGCAGGAGTTTCTGGATGCAACCGAT
GGTCGCGGCGTTGATCTGGTGCTGAATAGCCTGGCCCGTGAGTTCGTTGACGCAAGTCTGCGCCTGATGCC
GGGCGGTGGTCGCTTTGTGGATATGGGCAAGACAGACATCCGTCGTCCGGAACAGGTTGCAGAAGATCATG
GTGGTGTGGCCTACCAGGCATTTGATCTGGTTGAAGCAGGCCCTCAGCGCACAGGTGAGATGCTGGCAGAG
ATTGTTCGCCTGTTCCAGGCAGGCGCATTTCGCCCGTTACCGATTACCCAGTGGGATGTGCGCCGCGCACC
TGAGGCATTTCGTCACATCAGTCAGGCCAAACACATTGGCAAGATCGTGTTAACAGTTCCGCGCCCGATCG
ACACCGACGGTACCGTGATGGTTACCGGCGCCACAGGCACACTGGGTGGTTTCGTTGCCCGTCATCTGGTG
ACACACCACGGTATTCGCCGTTTACTGTTAGTGAGCCGCAGTGCAGAGCGTACCGACCTGGTGCGCGAACT
GACAGAATTAGGCGCCGATGTTACATGGGCAAGCTGTGATCTGGCAGACGCCACAGCCGTTGAAGAAACAG
TGCGTAGCGTGGATGAGCGCCACCCGTTAGTTGCCGTGGTGCATAGCGCCGGTGTGCTGGATGACGGTGTG
ATTGACAAGCAGAGTCCGGAACGTCTGGACACCGTGATGCGCCCTAAGGTGGATGCAGCATGGAACTTACA
CCGTCTGCTGGATAACGCACCGTTAGCCGACTTCGTTCTGTTTAGCAGCGCCAGCGGCGTGCTGGGCGGCG
CCGGCCAAAGCAATTACGCAGCCGCCAACGCCTTTCTGGACGCACTGGCCGAGCACCGTCGCGCACAAGGC
CTGGCAGGTCAAGCACTGGCATGGGGCTTATGGAGCGATCGTAGCACCATGACAGGCCAGCTGGGTAGTAC
CGAACTGGCCCGTATCGCACGTAACGGTGTTGCAGAAATGAGTGAAACCGAAGGTCTGGCACTGTTTGACG
CAGCCCGTGATACAGCAGAGGCCGTTCTGCTGCCGATGCACCTGGATGTTGCCCGCCTGCGTAGCCGTAAT
GGTGAGGTGCCTGCCGTGTTTCGTCGTCTGATTCATGCCACCGCACGTCGCACCGCCAGTACCGCCGTTCG
CAGTGCAGGTTTAGAACAGCAGTTAGCCAGTCTGAGCGGCCCTGAACGTACCGAACTGTTACTGGGCCTGG
TGCGTGATCACGCCGCAGCAGTGTTAGGCCACGGTACCAGTGACGCAGTGAGTCCGGACCGCCCGTTTCGT
GATCTGGGTTTCGACAGTCTGACCGCAGTGGAGCTGCGTAATCGTTTTGCCGCCCTGACCGGCCTGCGTTT
ACCGGCCACACTGGTTTTCGACCACCCTAGTCCGACAGCACTGGCAGGTCACCTGGCCGGCTTATTAGGCG
CAGCCACACCTAGCGCCGCCGAACCTGTGCTGGCAGCAGTTGGTCGTCTGCGTGCCGATTTACGCAGTCTG
ACACCTGATGCAGAAGGTGCAGAGGATGTGACCATTCAGCTGGAAGCCCTGCTGGCCGAGTGGCGTGAAGC
CGCAGAGAAACGTGCACCGGAAGCCGTGGGCGATGAGGACCTGAGCACAGCCACAGACGATGAAATATTCG
CCCTGGTGGATAGTGAATTAGGTGAAGCATGA