

ANALYZING THE ROLE OF COVALENT MODIFICATIONS TO AN ARYL
CARRIER PROTEIN IN DIRECTING SEQUENTIAL INTERACTIONS IN
YERSINIABACTIN SYNTHETASE

by
Andrew Goodrich

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
June, 2016

Abstract

Nonribosomal peptide synthetases (NRPSs) are enzymatic systems found in bacteria and fungi responsible for the production of complex secondary metabolites. NRPSs are able to generate a remarkably diverse array of natural products from simple starting materials. This is due to their modular nature and assembly-line organization in which the growing product is passed from module to module with an individual monomer incorporated at each step.

The modular nature of NRPSs makes them an attractive tool for synthetic biologists to generate novel natural products with desirable pharmaceutical or industrial properties. In principle, existing modules could be rearranged in a combinatorial fashion to produce an enormous number of new products. However, in practice this has not been successful, likely due to a lack of understanding of the mechanisms underlying NRPS synthesis.

Each module is made up of at least three domains whose combined function leads to the selection, activation, and incorporation of a small molecule into the growing product.

Domains called carrier proteins are first modified from an inactive apo form to an active holo form in which a 4'-phosphopantetheine (PP) moiety is attached to a conserved serine. Adenylation domains then select and activate a small molecule using ATP via formation of an acyl-adenylate and then load the small molecule onto a holo-carrier protein via formation of a thioester with the terminal thiol of the PP arm. Condensation domains then catalyze amide bond formation between substrates loaded on adjacent carrier proteins. During synthesis, individual domains must move relative to one another. Domains are also subject to both small and large-scale conformation changes. NRPS synthesis is therefore a complex process involving the interplay of catalysis, covalent

modification, and conformational rearrangements. Understanding how these processes are orchestrated to achieve efficient synthesis will be necessary for rationally redesigning these systems.

Here, I present my work aimed at dissecting the role of covalent modifications to carrier proteins in altering their structure and how this, in turn, modulates interactions with adenylation domains. In order to study the structure of a carrier protein in all of its forms, we first developed a novel method to characterize the loaded form by nuclear magnetic resonance (NMR) spectroscopy. We then solved the solution structures of a CP in all of its forms and characterized the NMR dynamics of the holo and substrate-loaded forms. Finally, we characterized the interaction between a carrier protein and an adenylation domain by fluorescence anisotropy, isothermal titration calorimetry, and NMR titration. Our results show that covalent modifications alter the structure and dynamics of the carrier protein and prosthetic moieties in a way that provides directionality to the interaction with the adenylation domain that parallels the chemical steps of elongation and thus promotes efficient synthesis.

Advisor: Dominique Frueh

Reader: James Stivers

Acknowledgements

I would first like to thank my advisor, Dr. Dominique Frueh, for his support and encouragement during my time in his lab and especially for his guidance in mastering and utilizing NMR. His willingness to share his expertise and guide my development as a spectroscopist were instrumental to my success as a graduate student and for that I am extremely grateful.

I would also like to thank the members of my thesis committee, Dr. Craig Townsend, Dr. James Stivers, and Dr. Joel Toman, for their input and assistance throughout my graduate career.

I would like to thank Dr. Mario Amzel for fruitful and enlightening discussions about thermodynamics and binding.

Finally, I would like to thank my parents, Mike and Karen, for their unending love and support, not only during graduate school but throughout my entire life. Without their unwavering support, none of this would have been possible.

Table of Contents

Abstract	ii
Acknowledgements	iv
List of Tables	vii
List of Figures	viii
Chapter 1	Introduction to nonribosomal peptide synthetases1
	Figures.....8
Chapter 2	A nuclear magnetic resonance method for probing molecular influences of substrate loading in nonribosomal peptide synthetase carrier proteins..... 12
	Methods..... 18
	Figures..... 29
Chapter 3	Solution structure of a nonribosomal peptide synthetase carrier protein loaded with its substrate reveals transient, well-defined contacts38
	Introduction.....39
	Results and Discussion42
	Conclusions.....55
	Methods.....58
	Figures.....67
Chapter 4	The solution structure of apo-ArCP reveals pre-formed binding sites for the phosphopantetheine arm and tethered substrate.....85
	Introduction.....86
	Results and Discussion87
	Conclusion95
	Methods.....96
	Figures.....97
Chapter 5	Covalent modifications to carrier proteins modulate interactions with catalytic domains110
	Introduction.....111
	Results.....113
	Discussion117
	Methods.....126

Figures.....	134
Bibliography	146
<i>Curriculum Vitae</i>	153

List of Tables

Table 3-1	NMR Structure Statistics for Holo-ArCP	69
Table 3-2	NMR Structure Statistics for Loaded-ArCP	70
Table 4-1	NMR Structure Statistics for Apo-ArCP	109
Table 5-1	Thermodynamic Parameters for Interactions with YbtE	145

List of Figures

Figure 1-1	Modular organization of nonribosomal peptide synthetases	8
Figure 1-2	Domain organization of nonribosomal peptide synthetases and schematic of synthesis.....	9
Figure 1-3	Central role of carrier proteins in NRPS synthesis	10
Figure 1-4	Domain alternation in adenylation domains	11
Figure 2-1	Spectroscopic signature of substrate loading and hydrolysis	29
Figure 2-2	Overlay of HN-HSQC of holo- and loaded-ArCP	30
Figure 2-3	HN-HSQC shows no detectable interaction between holo-ArCP, salicylate, and ATP	31
Figure 2-4	Time course of substrate-loading monitored by NMR	32
Figure 2-5	SrfTEII catalyzed substrate hydrolysis monitored by NMR.....	33
Figure 2-6	Uncatalyzed hydrolysis monitored by NMR	34
Figure 2-7	SrfTEII catalyzed hydrolysis regenerates holo-ArCP.....	35
Figure 2-8	Loaded-ArCP is stable for the time-course of a 3D-NOESY	36
Figure 2-9	Chemical shift perturbation profile of salicylate loading	37
Figure 3-1	Role of ArCP in yersiniabactin synthesis	67
Figure 3-2	Solution structures of holo- and loaded-ArCP	68
Figure 3-3	NMR bundles for holo- and loaded-ArCP	71
Figure 3-4	NOESY strips involving the phosphopantetheine arm in holo-ArCP	72
Figure 3-5	NOESY strips of residues in holo-ArCP displaying NOESY cross-peaks with the phosphopantetheine arm	73
Figure 3-6	NOESY strips involving the phosphopantetheine arm in loaded-ArCP ...	74
Figure 3-7	NOESY strips involving salicylate in loaded-ArCP	75

Figure 3-8	NOESY strips of residues in loaded-ArCP displaying NOESY cross-peaks with the phosphopantetheine arm	76
Figure 3-9	NOESY strips of residues in loaded-ArCP displaying NOESY cross-peaks with salicylate	77
Figure 3-10	Visualization of distance constraints involving prosthetic groups in holo- and loaded-ArCP	78
Figure 3-11	Change in conformation of prosthetic arm upon salicylate loading leads to different electrostatic surfaces in holo- and loaded-ArCP	79
Figure 3-12	Comparison of holo- and loaded-ArCP in complex with an adenylation domain	80
Figure 3-13	NMR dynamics of holo- and loaded-ArCP	81
Figure 3-14	Dynamics visualization of holo- and loaded-	82
Figure 3-15	ArCP Model-free analysis of holo- and loaded-ArCP	83
Figure 3-16	Chemical shift perturbations between major and minor forms in holo- and loaded-ArCP	84
Figure 4-1	Cartoon representation of the structure of apo-ArCP	97
Figure 4-2	Relative orientation of the four major helices in apo-ArCP	98
Figure 4-3	Comparison of apo-ArCP with holo- and loaded-ArCP	99
Figure 4-4	Repositioning of helix 3 upon phosphopantetheinylation and substrate loading	100
Figure 4-5	Apo-ArCP has a tightly-packed hydrophobic core and mostly hydrophilic surface	101
Figure 4-6	Solvent-exposed hydrophobic residues are present near the phosphopantetheine attachment site	102
Figure 4-7	Solvent-exposed residues form the binding sites for the phosphopantetheine arm and tethered substrate	103
Figure 4-8	Apo-EntB shows a similar pattern of solvent-exposed hydrophobic residues to apo-ArCP	104

Figure 4-9	Apo-TycC3 shows two potential binding surfaces for the phosphopantetheine arm105
Figure 4-10	Solvent-exposed hydrophobic residues on the structure of PCP7 _{Teic}106
Figure 4-11	Solvent-exposed hydrophobic residues on the structure of PCP1107
Figure 4-12	Loop 1 conformation in apo-, holo-, and loaded-ArCP108
Figure 5-1	Affinity of YbtE for apo-, holo-, and SalNH-ArCP as determined by fluorescence anisotropy132
Figure 5-2	Thermodynamic parameters of the interaction between apo-ArCP and YbtE as determined by ITC133
Figure 5-3	Thermodynamic parameters of the interaction between holo-ArCP and YbtE as determined by ITC134
Figure 5-4	Thermodynamic parameters of the interaction between SalNH-ArCP and YbtE as determined by ITC135
Figure 5-5	Spectroscopic signature of YbtE interacting with apo-ArCP136
Figure 5-6	Spectroscopic signature of YbtE interacting with holo-ArCP137
Figure 5-7	Residues from apo-ArCP affected by titration with YbtE138
Figure 5-8	Residues from holo-ArCP affected by titration with YbtE.....139
Figure 5-9	ITC fails to show an interaction between ArCP and Cy1141
Figure 5-10	Comparison of the thermodynamics of the interactions between YbtE and apo-, holo-, and SalNH-ArCP142

Chapter 1-Introduction to Nonribosomal Peptide Synthetases

Nonribosomal peptide synthetases (NRPSs) are enzymatic systems found in bacteria and fungi that are responsible for the production of diverse secondary metabolites. These complex natural products serve a wide variety of functions, including scavenging trace metals from the environment (including host organisms) and killing other bacteria or fungi. These natural functions have made NRPs valuable pharmaceutical agents and many antibiotics, anti-cancer agents, and immunosuppressants are NRPs or derivatives thereof¹. The products made by NRPSs cover an enormous amount of chemical space. They vary greatly in both size and composition and include many chemical moieties not found in ribosomally produced peptides, including heterocycles, D-amino acids and aryl acids. Despite their dissimilar structures, NRPs are all produced by a conserved mechanism in which the final products are built in an assembly-line fashion out of simple starting materials. Central to this synthetic strategy is the modular nature of NRPSs (Figure 1-1A). Each module, composed of multiple domains, is responsible for selecting, activating, and incorporating a single small molecule into the growing peptide before passing it to the downstream module¹. This modular nature has made NRPS an attractive target for synthetic biologists, as swapping modules with differing substrate specificities could, in principle, generate new peptides with novel pharmaceutical or industrial properties (Figure 1-1B). However, this has not yet emerged as an efficient means of generating novel natural products, most likely due to the complex interplay between catalysis, covalent modification, and protein-protein interactions that underlies NRPS synthesis and is still poorly understood.

A canonical module is composed of three different domains: an adenylation (A) domain, a carrier protein (CP), and a condensation (C) domain (Figure 1-2A). The CP first has to be converted from an inactive apo form to an active holo form via covalent attachment of a 4'-phosphopantetheine (PP) arm to a conserved serine by a phosphopantetheinyl transferase (PPTase) (Figure 1-2B). A domains then select a small molecule to be incorporated and, using ATP, activate the substrate by forming a high-energy acyl-adenylate. The activated substrate is then loaded onto the holo-CP via formation of a thioester with the PP arm (Figure 1-2C). The CP delivers the tethered substrate to an upstream C domain, which catalyzes peptide bond formation between the tethered monomer and an intermediate product loaded on an upstream CP. The CP then delivers the elongated product to a downstream C domain, where it is unloaded onto a downstream CP. This process continues until the final product is formed (Figure 1-2D) and a thioesterase (TE) domain catalyzes hydrolysis to form a linear product or macrocyclization¹.

NRPS synthesis is a highly complex process that involves CPs cycling between two or more biochemical states, catalytic domains carrying out multiple chemical reactions and undergoing large-scale conformational changes, and a set of many transient protein-protein interactions (Figure 1-3). All of these processes take place while a labile intermediate is covalently attached to the synthetase via the PP of a carrier protein. Understanding how the interplay between covalent modification, catalysis, and protein-protein interactions is orchestrated to optimize the transfer of labile peptide intermediates between active sites is key to delineating the mechanism of NRPS synthesis and providing a rationale for reengineering these systems to produce novel products.

Carrier proteins play a central role in NRPS synthesis, as intermediate products are covalently tethered to CPs via the PP arm when they are shuttled between multiple active sites. As such, CPs exist in many different biochemical forms (apo, holo, monomer-loaded, intermediate-loaded) and must interact with many different catalytic domains (PPTase, A domain, upstream and downstream C domains). Discerning the mechanism of NRPS synthesis thus requires knowledge of the structure of CPs in all the different forms and an understanding of the energetics governing the entire set of protein-protein interactions in which they participate.

The first structure of an apo-CP from a NRPS system² (tyrocidine synthetase) showed that NRPS CPs fold into the four-helix bundle that typifies the structure of CPs from NRPS, polyketide synthase (PKS), and fatty acid synthase (FAS) systems³. The four-helix bundle is comprised of 3 longer helices ($\alpha 1$, $\alpha 2$, and $\alpha 4$) that run parallel/antiparallel to one another in an up, down, down arrangement and a shorter helix ($\alpha 3$) that lies almost perpendicular to the other three. The loops connecting the helices are of variable length, with loop 1 being very long and dynamic, loop 2 being shorter but still flexible, and loop 3 being short and relatively rigid. Single-turn helices are also frequently found within loop 1. The conserved serine onto which the PP arm is covalently attached is found at the junction between loop 1 and helix 2.

Attachment of the PP arm to generate a holo-CP has been reported to have a wide range of effects on the structure of CPs. Some studies have reported that phosphopantetheinylation has no effect on the structure and found no evidence of an interaction between the CP and PP arm^{2,4} while others have reported that holo-CPs exist in a conformational equilibrium between two distinct structures, both of which show a strong interaction with the PP arm

but differ in the interaction surface⁵. We found⁶ (Chapter 3) that the PP arm has a well-defined interaction with a CP but is still highly dynamic and likely interconverts between a rigid, bound state and a flexible, unbound state. A recent solution structure of a holo-CP from a different NRPS system also found a well-defined interaction between the CP and PP arm⁷. In Chapter 4, I attempt to use the structures of apo-CPs to reconcile the disparate reports and suggest that the model we proposed, in which the PP arm rapidly interconverts between an unbound and one (or more) unbound forms, is common to all NRPS CPs.

Structures of a loaded-CP from a NRPS system were only solved recently^{6,7}. The labile nature of the thioester bond between the loaded substrate and the PP arm limits the lifetime of a loaded-CP to the point that it can not be characterized by traditional NMR methods. To overcome this limitation, two complementary strategies were developed and utilized. Our lab developed a method to generate the loaded form directly *in situ* in the NMR tube⁸ (Chapter 2) while the Burkart lab utilized a one-pot enzymatic reaction to activate and attach a nonhydrolyzable analog of substrate-loaded pantetheine to a CP⁷. Both studies found that NRPS CPs interact directly with the tethered substrate, wherein the substrate lies along a shallow groove formed by residues from loop 1, helix 2, and helix 3. We also found that substrate loading rigidifies the PP arm, although it still maintains some flexibility and, like the unloaded PP arm of a holo-CP, interconverts between a bound and unbound form. Finally, we found that the different placement of the PP arm in the holo and loaded structures obscures or exposes different binding surfaces and may, in this way, modulate interactions with A domains.

Structural and dynamic studies have now provided a detailed view of NRPS CPs in many of their biochemical states. We now understand how these covalent modifications alter the

structure of the CP itself and how they interact with the core of the CP. However, these structural and dynamic changes only gain significance when we consider how they influence protein-protein interactions and must therefore consider how these modifications modulate binding between CPs and catalytic domains.

Adenylation domains are large (~55-60 kDa) catalytic domains responsible for first selecting and activating small molecule substrates and then loading them onto holo-CPs in two distinct reactions⁹. A domains are comprised of two subdomains, a large N-terminal domain and a smaller C-terminal domain, and the active site for the chemical reactions lies at the interface of the two subdomains (Figure 1-4)⁹. Notably, the C-terminal domain contributes a different set of residues to the active site for each of the two reactions⁹. Crystallographic studies have shown that A domains adopt the so-called adenylation conformation for the first reaction, activation of the substrate with ATP to form an acyl-adenylate⁹. The C-terminal domain then rotates almost 140 degrees to adopt the thiolation conformation, in which it binds to a holo-CP and catalyzes substrate loading⁹. A third arrangement, the open conformation, has also been observed and is proposed to be present before substrates (amino acid and ATP) are bound⁹. A domains can thus exist in at least three conformations that differ in the relative orientation of the N- and C-terminal domains as they progress through their catalytic cycle (Figure 1-4).

Interaction between A domains and CPs has thus far only been observed with an A domain in the thiolation conformation and a holo-CP. The interaction between these two domains is transient and does not crystallize on its own. Multiple groups have used suicide inhibitors to trap holo-CPs in complex with an A domain and have solved crystal structures of a complex between an isolated CP and isolated A domain¹⁰, an A domain:CP fusion protein¹¹,

and the interaction within the context of full modules^{12,13}. These structures have consistently identified interactions between the N-terminal domain of A domains and helices 2 and 3 of a CP. The first two structures also claimed to identify an interaction between the C-terminal domain and loop 1 of the CP, but this has not been described in all reports of A domain:CP complexes^{12,13}. The nature of the interaction between an A domain and an apo- or loaded-CP is not yet known.

Years of work studying NRPSs has yielded a thorough biochemical description of nonribosomal peptide synthesis and begun to formulate a structural basis for understanding the mechanism of NRP synthesis. Structures have provided snapshots of individual domains, di-domains, and entire modules in various states^{2,6,7,10,12-14}. However, what is necessary for a complete mechanistic understanding of NRP synthesis is a measurement of the forces driving transitions *between* these various states. Only a limited number of studies have addressed how substrate binding or competition amongst binding partners modulates the structure and interactions of NRPS domains^{14,15}.

Here, I present my work investigating how covalent modifications to a an aryl carrier protein (ArCP) from the yersiniabactin synthetase system influence the structure and dynamics of the CP itself and how this, in turn, modulates interactions with partner catalytic domains. In order to study all forms of ArCP, I first developed a novel method for studying loaded forms of NRPS CPs by NMR⁸ (Chapter 2). I then solved the solution structure of apo-ArCP (Chapter 4) and solution structure and dynamics of holo-, and loaded-ArCP⁶ (Chapter 3), and, finally, characterized the interactions between the three forms of ArCP and the A domain YbtE by fluorescence anisotropy, isothermal titration calorimetry, and NMR titrations (Chapter 5). What emerges is a picture in which covalent

modifications to a CP alter the structure and dynamics of both the CP and the covalent modifications themselves. This, in turn, modulates the interactions with YbtE in a manner that provides a directionality to the set of protein-protein interactions that parallels the chemical steps of peptide elongation. The work presented here begins to address the mechanisms underlying protein-protein interactions in NRPS synthesis and will provide a foundation for studies involving additional catalytic domains.

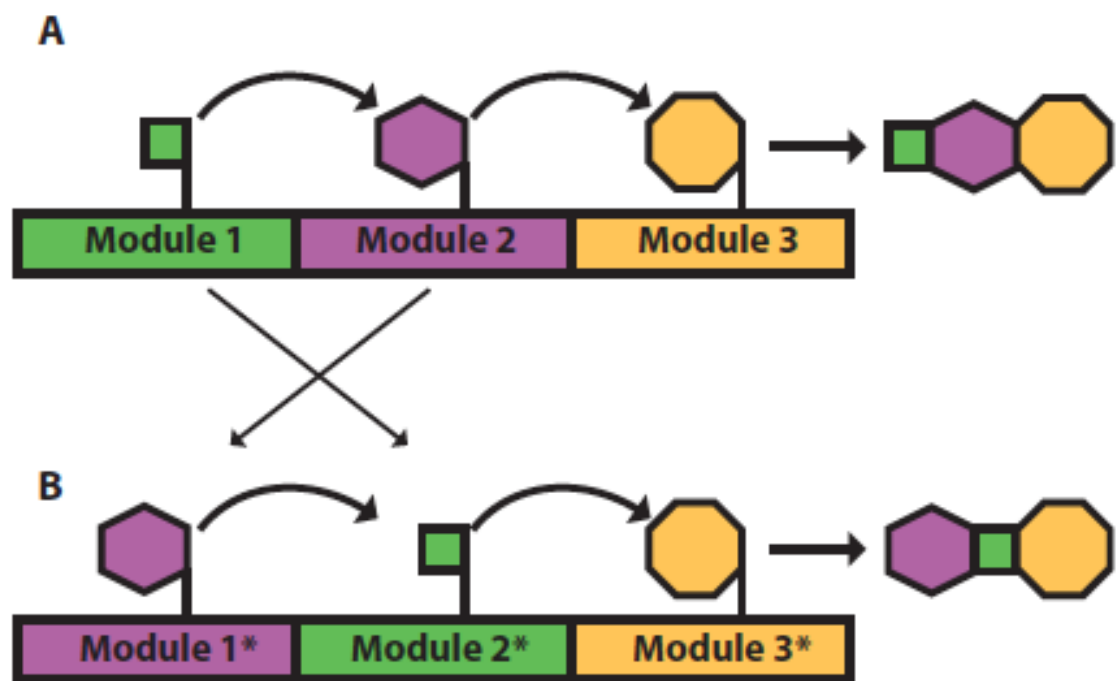


Figure 1-1. A) NRPSs are organized into modules. Each module is specific for a single small molecule, shown as a green square, purple hexagon, or orange octagon. The growing product is passed from module to module and monomers are incorporated in the order in which the modules are arranged. B) Since the order of the modules dictates the order in which small molecules are incorporated, swapping the order of modules could create novel products with desirable pharmaceutical or industrial properties.

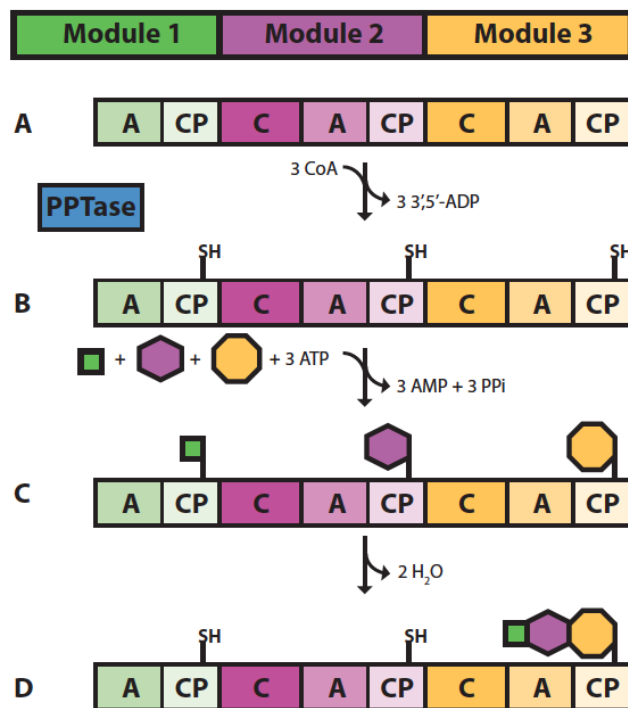


Figure 1-2. A) Each module is made up of a number of domains, including adenylation (A), carrier protein (CP), and condensation (C) domains. The green module is an initiation module and therefore has no C domain. B) CPs must be activated from an apo to a holo form by a phosphopantetheinyl transferase (PPTase). These enzymes catalyze the attachment of 4'-phosphopantetheine (PP) to a conserved serine found on all CPs. C) Adenylation domains select a small molecule substrate, activate it with ATP by forming an acyl-adenylate, and load it onto the PP arm via formation of a thioester. D) Condensation domains catalyze formation of amide bonds between products loaded on adjacent carrier proteins until the final product is attached to the terminal CP. This regenerates the holo form of the CPs so another round of synthesis can occur. The final product will be hydrolyzed as a linear product or released as a macrocycle by a thioesterase domain (not shown).

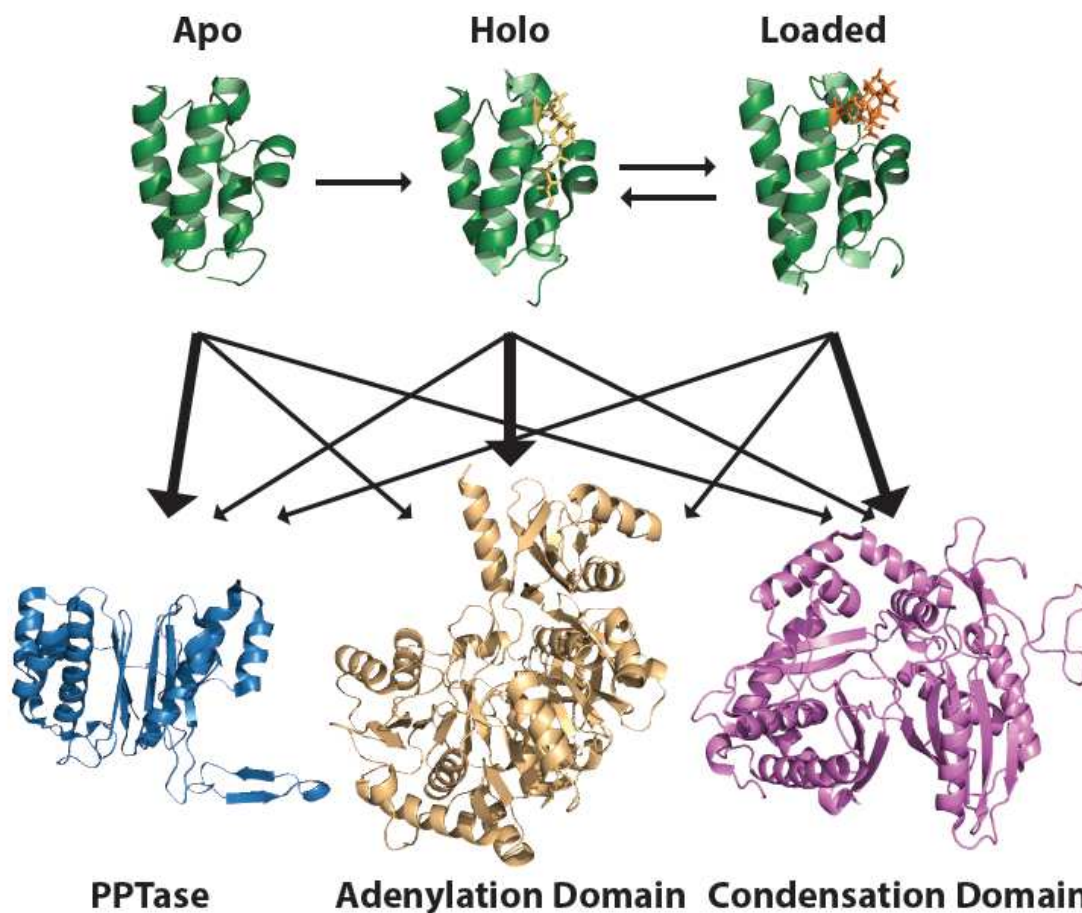


Figure 1-3. Central role of carrier proteins in NRPS synthesis. Carrier proteins are involved in every step of NRPS synthesis, from priming to elongation. They are first covalently activated from an apo to a holo form. During the course of elongation, they become loaded with a substrate and then donate the substrate to a downstream carrier protein, regenerating the holo form. These steps require interactions with a PPTase, an adenylation domain, and a condensation domain, at a minimum. Each form of a carrier protein is a substrate for a different enzyme, indicated with the bold arrow. However, every form may interact with every enzyme and it is unknown how, or if, enzymes discriminate between the various forms.

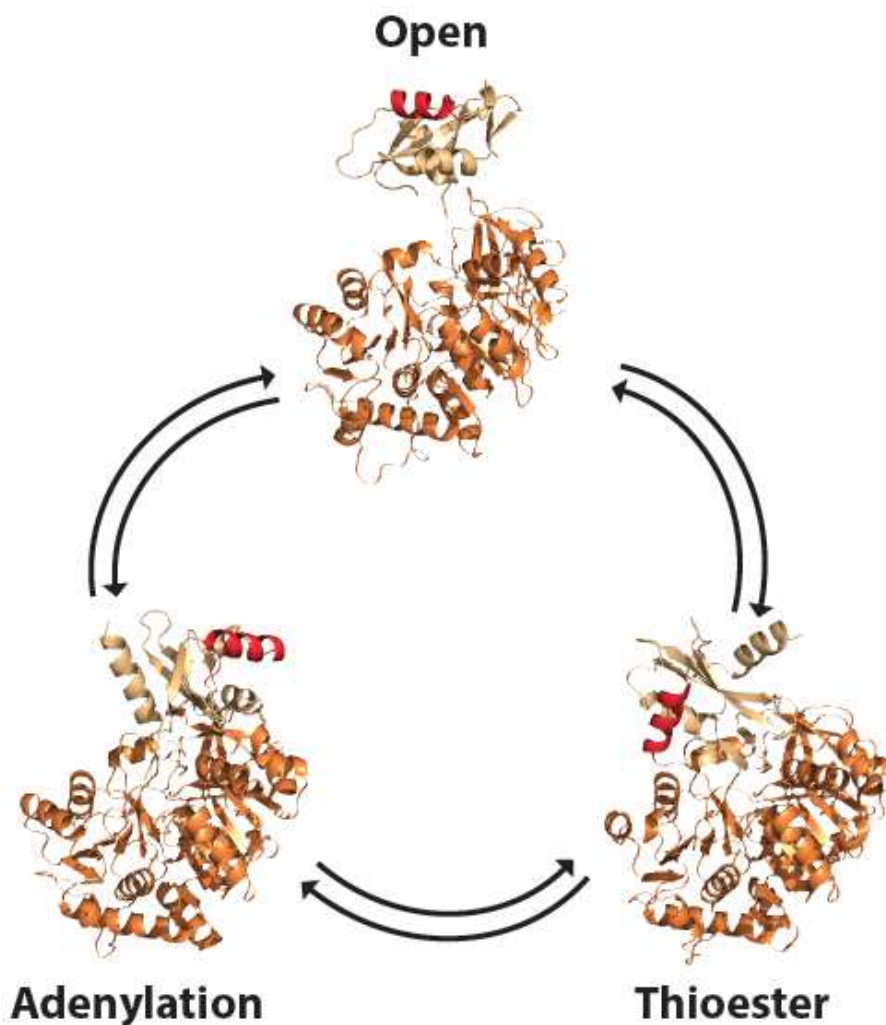


Figure 1-4. Domain alternation in adenylation domains. Adenylation (A) domains have been found to exist in at least three different conformations, each of which is proposed to play a different role in catalysis. The open conformation (luciferase, PDB 1LCI) is expected to be the major conformation present before substrates have been bound. The adenylation conformation (DhbE, PDB 1MDF) catalyzes formation of the acyl-adenylate and the thioester conformation (EntE, PDB 3RG2) binds to a holo-carrier protein and catalyzes substrate loading. Structures were aligned based on their N-terminal domains and a homologous helix is highlighted in red in each structure to show the changing position of the C-terminal domains.

Chapter 2-A nuclear magnetic resonance method for probing molecular influences of substrate loading in nonribosomal peptide synthetase carrier proteins

This chapter was published with minor modifications in *Biochemistry*, January 2015, by Andrew Goodrich and Dominique Frueh.

ABSTRACT: Carrier proteins (CPs) play a central role in nonribosomal peptide synthetases (NRPSs) as they shuttle covalently attached substrates between active sites. Understanding how the covalent attachment of a substrate (loading) influences the molecular properties of CPs is key to determining the mechanism of NRPS synthesis. However, structural studies have been impaired by substrate hydrolysis. Here, we used nuclear magnetic resonance spectroscopy to monitor substrate loading of a CP and to overcome hydrolysis. Our results reveal the spectroscopic signature of substrate loading and provide evidence of molecular communication between an NRPS carrier protein and its covalently attached substrate.

Nonribosomal peptide synthetases (NRPSs) are bacterial and fungal enzymatic systems that produce complex secondary metabolites from simple starting materials such as amino or aryl acids, many of which have found use as antibiotics and anti-cancer agents¹⁶. NRPSs possess a remarkable assembly line architecture, in which substrates are covalently attached to contiguous modules and condensed to form the final product. Each module is comprised of a core set of conserved domains and a long standing goal of the field is swapping domains or modules with differing substrate specificities so as to generate novel pharmaceuticals¹⁷. Unfortunately the molecular mechanisms of NRPS synthesis, and particularly domain communication, remain largely unknown, impeding progress in reprogramming NRPS assembly lines. Amongst NRPS domains, carrier proteins (CPs) play a central role as they tether the substrates to the assembly line and, hence, they visit many catalytic domains during NRPS synthesis. CPs are first converted from an inactive apo to an active holo form via covalent attachment of a 4'-phosphopantetheine arm (PP) onto a conserved serine. Next, adenylation (A) domains catalyze both substrate adenylation and thioester bond formation between the activated substrate and the PP of holo carrier proteins to generate a substrate loaded CP. Finally, condensation domains catalyze the peptide bond formation between two substrates loaded on neighboring CPs to extend the peptide. NMR and crystallographic studies indicate that NRPS modules are not rigid, but their domains are subject to inter- and intra-domain dynamics^{5,9,14,18,19}. Moreover, attachment of the PP altered the structure and dynamics of an isolated CP⁵. Studies of structurally related fatty acid synthases (FAS) and polyketide synthases (PKS) have implicated substrate loading in influencing large scale domain rearrangements^{20,21}. However, the lability of NRPS substrate thioester bonds has precluded similar studies of

loaded NRPS carrier proteins. Understanding how CPs efficiently orchestrate sequential, transient interactions with partner domains and elucidating the role of tethered substrates in modulating these interactions is of vital importance to understanding NRPS assembly line synthesis and, ultimately, rationally redesigning these systems. Here, we exploited the non-invasive nature of nuclear magnetic resonance (NMR) to overcome hydrolysis and study a loaded aryl carrier protein (ArCP) from yersiniabactin synthetase. Our results reveal that NRPS ArCPs interact either directly or indirectly with the substrates attached at the end of the 20 Å long PP.

In the yersiniabactin synthetase system, the free-standing A domain YbtE initiates synthesis by loading salicylate (Sal) onto the holo aryl carrier protein of the multidomain protein HMWP2²². The excised ArCP was overexpressed in *E. coli* BL21 (DE3) ΔEntD cells (a gift of Drs. Chalut and Guilhot) and purified to yield pure, homogeneous apo ArCP. Apo ArCP was phosphopantetheinylated *in vitro* and purified to obtain holo ArCP. To study the loaded form of ArCP, two major obstacles had to be addressed: hydrolysis and transthioation from thiol containing reducing agents²³, which were necessary to prevent disulfide bond formation in holo ArCP. Transthioation was avoided by using tris(2-carboxyethyl)phosphine as a reducing agent. Although hydrolysis of thioesters is slower than transthioation, it was rapid enough to preclude quantitative analysis of loaded ArCP. Indeed, when ArCP was loaded with Sal (confirmed by MALDI mass spectrometry (MALDI MS)) and freshly purified, NMR spectra featured signals of both holo ArCP and a previously unobserved form. Unfortunately, the new signals decreased over time, raising the possibility that they were an artifact of sample preparation and not reporters of substrate loading. Purified, loaded ArCP was therefore unsuitable for NMR studies.

To allow for prolonged detection of loaded ArCP, we exploited the isotope editing ability of NMR, and we generated the loaded form *in situ* to bypass the need for purifying loaded ArCP. Sal and ATP (2 mM each) were added in large excess to ^{15}N -labeled holo ArCP (300 μM). No interaction between holo ArCP and the reagents was detected in HN-HSQC spectra (Figure 2-3). We then added catalytic quantities of YbtE (100 nM) and collected a series of HN-HSQC. At such concentrations, the binding of YbtE to ArCP does not induce shifts in NMR signals. During the reaction, signals of loaded ArCP increase while signals of holo ArCP decay (as shown for the phosphopantetheinylation site, Ser52, in Figure 2-1 A1, B1-4). MALDI MS confirmed the conversion into ArCP loaded with Sal. To demonstrate that all signals observed by NMR report on substrate loading and not some undesired side effects, we induced substrate hydrolysis by enzymatic catalysis. We added the promiscuous thioesterase SrfTEII²⁴ to purified loaded ArCP, and the NMR signals of loaded ArCP disappeared as those of holo ArCP reappeared (Figure 2-1 A2, B5-7). The HN-HSQC collected following the completion of SrfTEII-catalyzed hydrolysis (confirmed by MALDI MS) overlays perfectly with that of holo ArCP, demonstrating that the new peaks observed upon incubation of holo ArCP with ATP, Sal, and YbtE are indeed the spectroscopic signature of substrate loading (Figure 2-7). Comparison between the kinetics of uncatalyzed hydrolysis (Figure 2-6) and substrate loading suggested that hydrolysis may be compensated by continuous reloading of Sal on ArCP. Indeed, spectra of loaded ArCP recorded after 5 days displayed no significant regeneration of holo ArCP in presence of YbtE, Sal, and ATP, whereas *in the absence* of YbtE, Sal, and ATP, ~57% of purified loaded ArCP hydrolyses over that time (Figures 2-8 and 2-6). Thus, adding catalytic amounts of the cognate A domain overcomes hydrolysis of loaded monomers for a duration

amenable for NMR studies. This method complements other strategies used to circumvent hydrolysis in loaded carrier proteins (SI, S6).

The changes we observe in NMR spectra of loaded ArCP may reflect many molecular events. NMR signals in HN-HSQC report on the local chemical environment of the amide group of each residue. Differences between spectra of holo and loaded ArCP, so-called chemical shift perturbations (CSPs), can indicate a variety of events, such as direct contact with the salicylate moiety, repositioning of the PP, structural alterations, and modulation of dynamics. Figure 2-2A shows an overlay of holo and loaded ArCP spectra. Simple inspection reveals that no massive structural rearrangement occurs upon loading, since most signals are the same in both spectra. However, a significant number of peaks shift markedly, indicating that loading Sal impacts the related amide groups. To gain insights into the origins of the CSPs of holo/loaded ArCP, we assigned the ^1H and ^{15}N resonances of apo ^{15}N - ^{13}C -ArCP using conventional experiments^{25,26} and transposed the assignment of apo ArCP to holo ^{15}N -ArCP and to loaded ^{15}N -ArCP with NOESY-HN-HSQC spectra. Figure 2-2B, C, and D highlight four peaks that show varying degrees of CSP. Mapping the largest CSPs (Figure 2-9) onto the secondary structure of ArCP as determined with chemical shift indexing^{27,28}, reveals distinct clusters of residues affected by loading (Figure 2E). CSPs of residues around the phosphopantetheinylation site (Ser52, N-terminus of helix 2) likely reflect a change in conformation of the PP. CSPs of residues in helix 3 together with residues in the middle of helix 2 may indicate a direct substrate interaction reminiscent of that observed in PKS and FAS acyl carrier proteins^{29,30}. In both FAS and PKS, changing tethered substrates repositions helices 2 and 3^{31,32}, and it was suggested that these structural variations may modulate binding events²¹. Various NMR experiments can

test if salicylate alters the structure and/or dynamics of ArCP and modulates protein-protein interactions. Our newly designed conditions will permit such measurements while ArCP remains loaded.

In summary, we used NMR to directly monitor NRPS substrate loading, thereby providing the first atomic level description of this process. We found that NRPS substrates directly or indirectly interact with their cognate carrier proteins. Whether substrates bind to CPs or induce conformational fluctuations, substrate loading is expected to modulate the binding affinity of CPs toward partner catalytic domains. Decades of biochemical studies have demonstrated interplay between carrier proteins and their substrates during catalytic steps involving various domains^{33,34} and our method provides a framework for investigating the molecular determinants of this interplay.

Methods

Cloning of ArCP 14-93

The *Y. pestis* irp2 gene (Accession Number AAM85957) fragment coding for residues 14-93 of the protein HMWP2 (courtest Dr. Christopher T. Walsh, Harvard Medical School) was PCR amplified using the primers ArCPStartD14KpnI (5'-CTCAGGATTCGCTGGGTACCGACAACCGCCACGCGGC-3') to introduce a KpnI cut site and ArCPEndE93 (5'-TATCGAGTCATTCTCGAGCCGCCTACTCAGGCGACCGGC-3') to introduce an XhoI cut site and stop codon. The PCR product and target vector pETRP1B-GB1-TEV were digested with KpnI and XhoI, gel purified, and extracted. The fragment coding for residues 14-93 was ligated into pETRP1B-GB1-TEV (courtesy Dr. Wolfgang Peti, Brown University) to yield pETRP1B-GB1-TEV-ArCP14-93. The resulting plasmid directs production of residues 14-93 of HMWP2 with an N-terminal GB1 tag followed by a hexahistidine tag and a TEV cleavage site. The DNA sequence of pETRP1B-GB1-TEV-ArCP14-93 was confirmed by sequencing. Following TEV cleavage, a GT sequence remains at the N-terminus.

Cloning of YbtE

The gene coding for the full length YbtE (Accession Number CAA21394) was cloned in two steps. The YbtE gene was PCR amplified from pPROEX-1 (courtesy Dr. C.T. Walsh) using the primers YbtE-XhoI-Reverse (5'-TGTCTCGAGTCGGTTTGCGCTTATTGGGCAG-3') to introduce a XhoI cut site and YbtE-NdeI-Forward (5'-CTGTATTTTGAGGGCGCCCATATGAATTCTTCC-3') which overlapped with an existing NdeI cut site. The PCR product and target vector

pETRP1B were digested with NdeI and XhoI and gel purified. The insert was ligated into the vector to give the plasmid pETRP1B-YbtE coding for full length YbtE with an N-terminal hexahistidine tag followed by a TEV cleavage site. The YbtE gene was PCT amplified from pETRP1B-YbtE using YbtE-XhoI-Reverse and YbtE_KpnI_Forward (5'TTCATCATCATCATGGTACCGAAATCTTTATTTT-3'). The PCR product and target vector pETRP1B-GB1-TEV were digested with XhoI and KpnI and gel purified. The extracted products were ligated to yield pETRP1B-GB1-TEV-YbtE. This plasmid directs production of full length YbtE with an N-terminal GB1 tag followed by a hexahistidine sequence and a TEV cleavage site. The sequence of pETRP1B-GB1-TEV-YbtE was confirmed by sequencing. Following cleavage by TEV, a GH cloning artifact remains at the N-terminus.

Protein Expression and Purification

Unless otherwise noted, the pH listed for each buffer is the pH at 4 °C.

pETRP1B-GB1-TEV-ArCP14-93 was transformed into competent *E. coli* BL21 (DE3) ΔEntD cells (courtesy Dr.s Chault and Guilhot, CNRS, Toulouse, France). A 2 ml culture of Luria broth (LB) with 50 µg/ml kanamycin was inoculated with a single transformed colony and grown at 37 °C and 225 rpm for 5-6 hours. 500 µl of this culture was added to 200 ml M9 minimal media with 1 g/L ¹⁵NH₄Cl as the sole nitrogen source for ¹⁵N-labeled samples and grown at 37 °C overnight. The 200 ml culture was added to 800 ml media and growth continued at 37 °C. At an optical density of 0.8 (600 nm), isopropyl β-D-1 thiogalactopyranoside (IPTG) was added to 0.5 mM to induce protein expression. Cells were harvested 3-3.5 hours after induction at an optical density of 1.2-1.3. The cell pellet was resuspended in lysis buffer (50 mM Tris, pH 7.4, 325 mM MgCl₂, 20 mM imidazole,

2% CHAPS (w/v), 2 mM phenylmethylsulfonyl fluoride (PMSF), 1 mg/ml lysozyme) and lysed using a French pressure cell at 18,000 PSI. DNase I was added to the lysate to 10 μ g/ml and the lysate incubated on ice for 15 minutes. The lysate was clarified by centrifugation at 26,900 xg for 30 minutes. The supernatant was diluted 5-fold in 50 mM Tris, pH 7.4, 20 mM imidazole to reduce the concentration of CHAPS to 0.4% (w/v) and the concentration of $MgCl_2$ to 65 mM and loaded onto a 5 ml HisTrap HP column (GE Healthcare). The column was washed with 30 ml His. Buffer A (50 mM Tris, pH 7.4, 1 M NaCl, 20 mM imidazole) and eluted with a linear gradient reaching 100% His. Buffer B (50 mM Tris, pH 7.4, 1 M NaCl, 500 mM imidazole) over 20 column volumes (CV) at a flow rate of 2 ml/min using an Akta purifier (GE Healthcare). Fractions containing GB1-TEV-ArCP14-93 were identified by SDS-PAGE, pooled, and dialyzed against 2 L Dialysis Buffer A (50 mM K_2HPO_4 , pH 7.0, 2 mM EDTA) at 4 °C overnight. The dialyzed sample was loaded onto a 5 ml HiTrap Q HP column (GE Healthcare) using a peristaltic pump and washed with 20 ml Dialysis Buffer A. The flow-through containing GB1-ArCP-14-93 was collected and 1 OD_{280} TEV protease added per 20 OD_{280} sample to remove the GB1 and hexahistidine tags, and the sample dialyzed against 2 L Dialysis Buffer B (50 mM Tris, pH 8.0, 0.3 M NaCl, 20 mM imidazole, 2 mM EDTA) at 4 °C overnight. Complete digestion of the sample to produce Gb1-His6 and ArCP was verified by SDS-PAGE. The dialyzed sample was loaded onto a 5 ml HisTrap HP column and washed with 20 ml Dialysis Buffer B. The flow-through containing ArCP was collected, concentrated using a 3K MWCP centrifugal filter (Millipore) and loaded onto a Superdex 75 16/60 pg column (GE Healthcare) that had been pre-equilibrated with 1.2 column volumes of NMR buffer (50 mM ACES, pH 6.80 at 22 °C, 150 mM NaCl, 500 μ M TCEP, 1 mM $MgCl_2$). Peak fractions

were identified by SDS-PAGE and stored dilute at 4 °C. NMR samples of apo ArCP were concentrated to 0.3 mM. Concentration was determined by measuring absorbance at 280 nm and using an extinction coefficient of 20970/M*cm.

SrgTEII was grown at 37 °C to an optical density of 0.6 and then chilled for 1 hour in an ice bath. IPTG was then added to 1 mM and growth continued at 18 °C. Cells were harvested after 16 hours and the cell pellet resuspended in His. Buffer C (50 mM Tris, pH 8, 0.5 M NaCl, 30 mM imidazole, 130 µg/ml lysozyme). Cells were lysed by sonication (15 cycles of 30 seconds on, 60 seconds off) and DNase I added to 3 µg/ml. The lysate was clarified by centrifugation at 26,900 xg for 30 minutes and the supernatant loaded onto a 5 ml HisTrap HP column. The column was washed with 15 CV His. Buffer C and eluted with a linear gradient of His. Buffer D (50 mM Tris, pH 8.0, 0.5 M NaCl, 0.5 M imidazole) to 100% over 27 CV at a flow rate of 3 ml/min. Peak fractions were identified by SDS-PAGE and dialyzed against SEC buffer (50 mM HEPES, pH 7.0, 100 mM NaCl, 0.1 mM TCEP). The dialyzed sample was run on a Superdex 75 16/60 pg column that had been equilibrated with 1.5 CV SEC buffer and a single peak was observed. Peak fractions were pooled, flash frozen in liquid N₂, and stored at -80 °C. Concentration was determined by measuring absorbance at 280 nm and using an extinction coefficient of 16,960/M*cm.

pETRP1B-GB1-TEV-YbtE was transformed into *E.coli* BL21 (DE3) cells and grown in LB with 50 µg/ml kanamycin at 37 °C to an OD of 0.6. IPTG was added to 0.5 mM to induce protein expression and growth continued at 18 °C for 16 hours before harvesting. Cell pellets were resuspended in YbtE lysis buffer (50 mM Tris, pH 8.0, 0.5 M NaCl, 20

mM imidazole, 2 mM PMSF, 1 mg/ml lysozyme) and lysed using a French pressure cell at 18,000 psi. DNase I was then added to the lysate to 20 µg/ml and incubated on ice for 15 minutes. The lysate was clarified by centrifugation at 26,900xg for 30 minutes. The supernatant was loaded onto a 5 ml HisTrap HP column using a peristaltic pump, washed with 30 ml His. Buffer A, and eluted with a linear gradient to 100% His. Buffer B over 20 CV at a flow rate of 2 ml/min. Peak fractions were identified by SDS-PAGE, pooled, concentrated using a 30K MWCO centrifugal filter (Millipore), and loaded onto a Superdex 75 16/60 pg column that had been equilibrated with 1.2 CV NMR buffer. Peak fractions were identified by SDS-PAGE and pooled. TEV protease was added to cleave the GB1 tag and the pool dialyzed against 50 mM Tris, pH 8.0, 0.1 M NaCl, 2 mM EDTA at 4 °C overnight. The dialyzed sample was run on a Superdex 75 16/60 pg column that had been equilibrated with 1.2 CV NMR buffer. Fractions containing YbtE were identified by SDS-PAGE, pooled, and stored dilute at 4 °C. Concentration was determined by measuring absorbance at 280 nm and using an extinction coefficient of 52,370/M*cm.

His₆-tagged Sfp was overexpressed in *E. coli* BL21 (DE3) cells and grown in LB with 50 µg/ml kanamycin at 37 °C to an optical density of 0.6. IPTG was added to a final concentration of 0.5 mM to induce protein expression and growth continued for an additional 4 hours. Cells were harvested and the cell pellet was resuspended in Sfp lysis buffer (50 mM Tris, pH 7.4, 0.5 M NaCl, 20 mM imidazole, 0.4 mg/ml lysozyme, 2 mM PMSF) and lysed using a French pressure cell at 18,000 psi. DNaseI was added to the lysate to 20 µg/ml and incubated on ice for 15 minutes. The lysate was clarified by centrifugation at 26,900xg for 30 minutes. The supernatant was loaded onto a 5 ml HisTrap HP column

using a peristaltic pump, washed with 30 ml His. Buffer A, and eluted with a linear gradient to 100% His. Buffer B over 20 CV at a flow rate of 2 ml/min. Peak fractions were identified by SDS-PAGE and pooled. The pooled fractions were run on a Superdex 75 16/60 pg equilibrated with 1.2 CV 50 mM Tris, pH 8, 100 mM NaCl. Fractions containing Sfp were identified by SDS-PAGE, concentrated to 160 μ M, and separated into 100 μ l aliquots. Aliquots were flash frozen in liquid N₂ before storing at -80 °C. Concentration was determined by measuring absorbance at 280 nm and using an extinction coefficient of 28,880/M*cm.

Phosphopantetheinylation of apo ArCP

Apo ArCP was phosphopantetheinylated *in vitro* in multiple 0.5 ml reactions containing 50 μ M apo ArCP, 75 mM Tris, pH 7.5 at 22 °C, 10 mM MgCl₂, 2 mM DTT, 250 μ M Coenzyme A trilithium salt (Sigma), and 500 nM of the PPTase Sfp. Reactions were incubated at 30 °C for 4 hours and then combined and concentrated. Holo ArCP was then purified by size exclusion chromatography on a Superdex 75 16/60 pg that had been pre-equilibrated with 1.2 column volumes NMR buffer. Fractions containing holo ArCP were identified by SDS-PAGE, pooled, and stored dilute at 4 °C.

MALDI-TOF Mass Spectrometry Measurements

All MALDI-TOF experiments were performed using a Voyager DE-STR (Applied Biosystems) instrument in linear mode. Sinapic acid (10 mg/ml in 50% acetonitrile/0.1% trifluoroacetic acid (TFA)) was used as a matrix. Samples were prepared by 50-fold

dilution into 0.1% TFA. Masses were calibrated using the Protein Calibration Standard 1 (Bruker) with standards ranging from 5734.51-16952.30 Da.

NMR Experiments

All samples were prepared in NMR buffer with 90% H₂O/10% D₂O at a final concentration of 0.3-0.4 mM with DSS for internal referencing.

To determine the optimal conditions for ArCP long term solubility, we performed a buffer screen with an incomplete factorial design approach³⁵. This approach allowed us to effectively sample a large number of solution conditions in a small number of experiments (48). The long-term solubility of ArCP was tested at 8 pHs (5.0, 5.5 (sodium acetate), 6.0, 6.5, 7.0 (sodium phosphate or HEPES when containing magnesium), 7.5, 8, and 8.5 (TrisHCl), all buffer at 100 mM), 5 salt types (NaCl, KCl, MgCl₂, MgSO₄, and Na₂SO₄), 5 salt concentrations (10 mM, 50 mM, 100 mM, 200 mM, and 300 mM), and 4 additives (BSA 20 M, Arg/Glu at 25 or 50 mM, and no additive). To test the solubility, 900 nl hanging drops (600 nl 150 μ M ArCP in 20 mM HEPES, pH 6.72, 150 mM NaCl, 1 mM DTT and 300 nl test solution) were set up in a 96 well plate with 100 μ l of the corresponding test solution in the well. Drops were imaged after 19 days using a Rock Imager (Formulatrix) and scored based on visible precipitation. The scores were analyzed using the procedure of Ducat and collaborators³⁵. The screen indicated that ArCP was most soluble in buffers with high pH (7.5 or 8) and poorly soluble in buffers with a pH below 7. Initial NMR experiments were thus performed in a buffer with a pH of 7.5. However, the pH of the buffer was lowered to 6.8 to allow for detection of two resonances that are

otherwise invisible due to solvent exchange line-broadening. At lower pH, the solubility of ArCP is unsuitable for NMR studies. Thus, the pH of 6.8 is a compromise between protein solubility and spectroscopic properties.

All spectra were collected at 25 °C on a 600 MHz Bruker Avance III spectrometer equipped with a QCI cryoprobe. Unless otherwise noted, 2D-HN-HSQC spectra were recorded with spectral widths of 16.019 ppm for proton, centered at 4.696 ppm, and 26 ppm for nitrogen, centered at 117 ppm, with a data matrix of 1024×128 complex points. A recycling delay of 1 s was used and 16 scans were accumulated, amounting to a measurement time of 1 hour 20 minutes. Each spectrum was linear predicted and zero-filled to a final size of $2048 (^1\text{H}) \times 256 (^{15}\text{N})$ points.

Assignment of apo ArCP was performed using a 0.3 mM ^{15}N , ^{13}C -labeled sample and a standard suite of triple resonance experiments^{25,26}, namely HNCA (8 scans, $2048 (^1\text{H}, 16.019 \text{ ppm at } 4.705 \text{ ppm}) \times 28 (^{15}\text{N}, 35 \text{ ppm at } 118 \text{ ppm}) \times 64 (^{13}\text{C}, 35 \text{ ppm at } 53 \text{ ppm})$ complex points, 19.5 hrs), HNC(8 scans, $2048 (^1\text{H}, 16.019 \text{ ppm at } 4.705 \text{ ppm}) \times 20 (^{15}\text{N}, 118 \text{ ppm at } 4.705 \text{ ppm}) \times 50 (^{13}\text{C}, 173 \text{ ppm at } 4.705 \text{ ppm})$ complex points, 11 hrs), HNCACO (8 scans, $2048 (^1\text{H}, 16.109 \text{ ppm at } 4.705 \text{ ppm}) \times 23 (^{15}\text{N}, 35 \text{ ppm at } 118 \text{ ppm}) \times 50 (^{13}\text{C}, 15 \text{ ppm at } 173 \text{ ppm})$ complex points, 2 days 2.5 hrs), HNCOCA (32 scans, $2048 (^1\text{H}, 16.019 \text{ ppm at } 4.705 \text{ ppm}) \times 29 (^{15}\text{N}, 35 \text{ ppm at } 118 \text{ ppm}) \times 75 (^{13}\text{C}, 35 \text{ ppm at } 53 \text{ ppm})$ complex points, 1 day 23.5 hrs), HNCACB (40 scans, $2048 (^1\text{H}, 16.019 \text{ ppm at } 4.705 \text{ ppm}) \times 20 (^{15}\text{N}, 35 \text{ ppm at } 118 \text{ ppm}) \times 75 (^{13}\text{C}, 70 \text{ ppm at } 38 \text{ ppm})$ complex points, 3 days 11 hrs). All 3D spectra were linear predicted once and zero-filled to the nearest power of two.

3D-NOESY-HN-HSQC (16 scans, 2048 (^1H , 16.019 ppm at 4.696 ppm) \times 40 (^{15}N , 26 ppm at 117 ppm) \times 120 (^1H , 11 ppm at 4.696 ppm) complex points, mixing time of 90 ms, 4 days 12 hrs) were recorded for ^{15}N samples of apo, holo, and loaded ArCP. The assignment of apo ArCP was transposed to the signals of holo and loaded ArCP when the nOe cross-peaks enabled unambiguous identification. All 3D spectra were linear predicted once and zero-filled to the nearest power of two.

All NMR spectra were processed using NMRPipe and analyzed using CARA. Intensity measurements were made using the batch integration mode in CARA and plotted and analyzed using MATLAB.

Monitoring ArCP Loaded with Salicylate by NMR

Observation of loading was performed by first adding salicylic acid (100 mM stock in NMR buffer, pH 6.80) to 2 mM to holo ArCP and recording an HN-HSQC to verify lack of interaction (Figure 2-3). ATP (100 mM stock in NMR buffer, pH 6.80) was then added to 2 mM and an HN-HSQC recorded to verify lack of interaction (Figure 2-3). YbtE (in NMR buffer) was then added to 100 nM and a series of HN-HSQC (8 scans, 1024(^1H , 16.019 ppm at 4.696 ppm) \times 128(^{15}N , 26 ppm at 117 ppm) complex points, 40 minutes) recorded. Time points reported for the time course of loading were chosen as the midpoint of each spectrum (Figure 2-4).

Monitoring Hydrolysis of Loaded ArCP by NMR

Uncatalyzed and SrfTEII-catalyzed hydrolysis were monitored using loaded ArCP freshly purified by size exclusion chromatography (using a Superdex 75 16/60 pg column equilibrated with 1.2 CV of NMR buffer). The uncatalyzed hydrolysis of purified loaded ArCP was monitored by collecting HN-HSQC spectra over the course of 15 days. Time points in Figure S4 correspond to the starting time of each spectrum. For SrfTEII catalyzed hydrolysis, the buffer of SrfTEII was first exchanged from HEPES to ACES by repeated dilutions and concentrations in NMR buffer using an Ultracel-3K (Millipore) centrifugal filter until a 1600-fold buffer exchange had been achieved. SrfTEII was added to a final concentration of 500 nM, and a series of HN-HSQC spectra was collected and analyzed in the same manner as described for the loading reaction (Figure 2-5). SrfTEII is a promiscuous type II thioesterase, routinely used to hydrolyze NRPS thioester substrates.²⁴

Regeneration of Loaded Carrier Proteins Versus Amide and Ester Mimics of Loaded Carrier Proteins

Our method complements existing protocols for studying loaded CPs. In order to avoid hydrolysis, it is possible to prepare carrier proteins loaded with their substrates but with ester and amide bonds instead of labile thioester bonds. To do so, the (protected) substrate is first coupled to ethylenediamine (to make amide mimics) or ethanolamine (to make esters). Second, chemical synthesis is used to produce derivatives of pantetheine.³⁶ Next, the pantetheine derivatives are incubated with three enzymes, Pank, Ppat, and DpcK, to produce CoA derivatives^{36,37}. Finally, incubation with a phosphopantetheine transferase (e.g. Sfp) leads to a carrier protein loaded with its substrate via an amide (or ester) bond in place of a thioester bond³⁸⁻⁴⁰. The enzymatic reactions can be combined in a one-pot

reaction³⁶. Clearly, for studies of substrate monomers, our method using *in situ* regeneration of loaded CPs is both easier and cheaper. Carrier proteins are simply incubated with substrates and catalytic amounts of adenylation domains. Further, carrier proteins are loaded with their substrates via native thioester bonds and the influences of amide or ester bonds need not be considered. However, our method may be ill-suited for studies of CPs loaded with complex *intermediates* (e.g. polypeptides instead of amino acid monomers), as generating the loaded intermediate *in situ* would require adding other catalytic domains in addition to adenylation domains. Thus, the combined catalytic efficiency of multiple enzymes would have to overcome hydrolysis in order to regenerate a CP loaded with the appropriate intermediate. Moreover, this procedure could produce a mixture of monomer and intermediate loaded carrier proteins, hindering analysis and interpretation. Thus, for studies of CP loaded with intermediates, non-hydrolysable analogs are most likely best suited, whereas studies of CPs loaded with monomeric substrates may be facilitated when using our method or adaptations thereof.

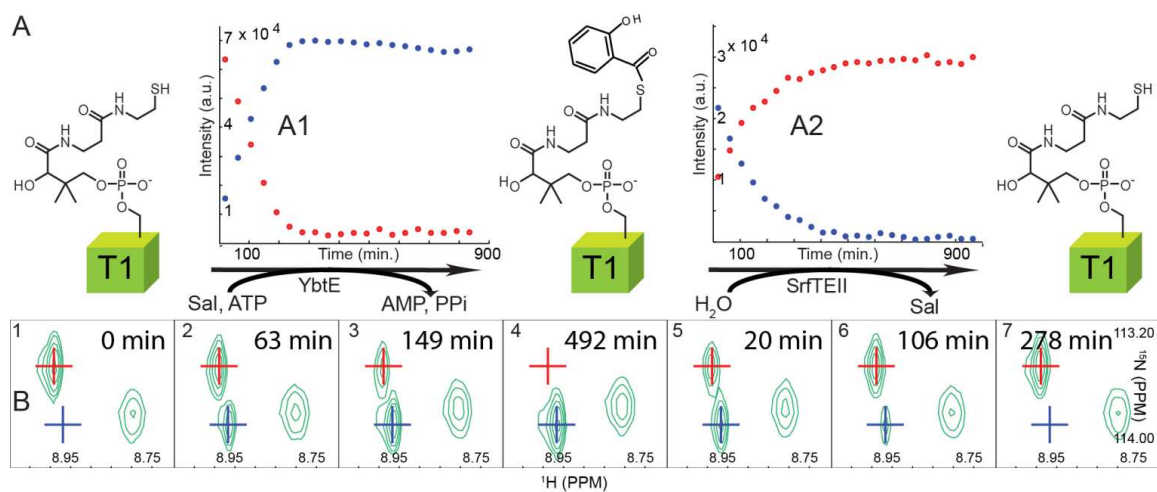


Figure 2-1. (A) Reactions catalyzed by YbtE and SrfTEII and their time-course (A1, A2, respectively), here monitored by the signal of Ser52, the phosphopantetheinylation site. (B) Signal of Ser52 during the loading reaction (B1-B4) and following the addition of SrfTEII after purification to remove YbtE and substrates (B5-B7). Blue: loaded ArCP; red: holo ArCP.

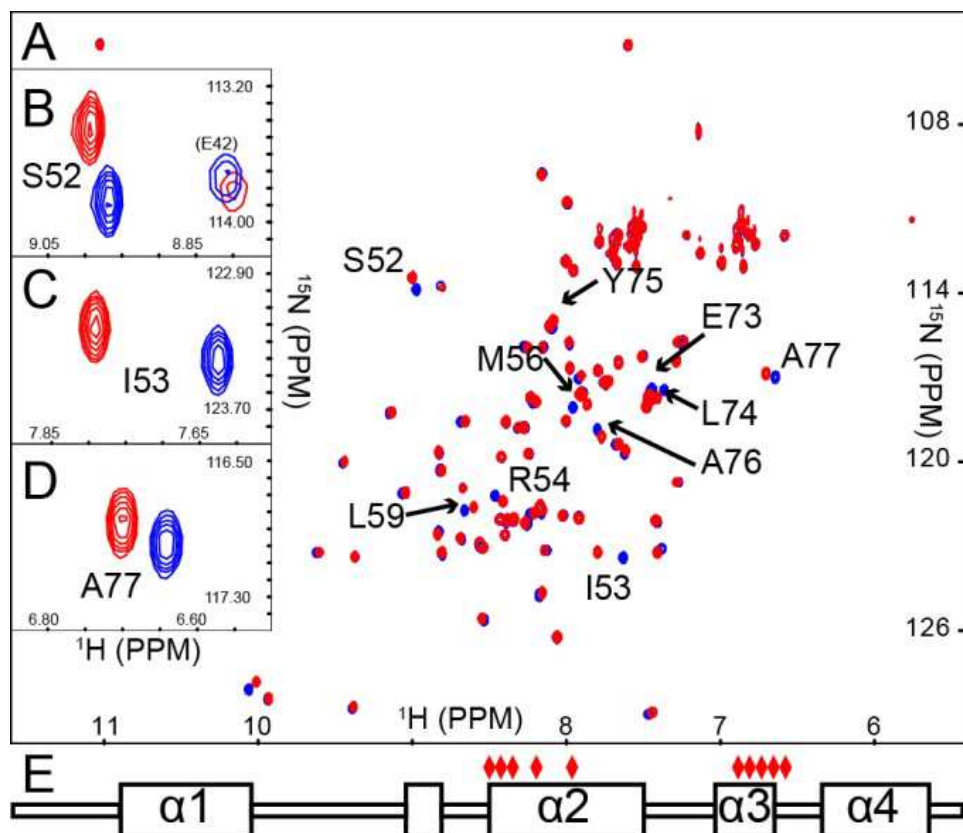


Figure 2-2. (A) Overlay of holo (red) and salicylate loaded (blue) ArCP HN-HSQC. Signals with CSP greater than 1 standard deviation above the median are indicated. Insets show zooms on the signals of (B) Ser52, (C) Ile53, and (D) Ala77. (E) Mapping the residues with significant CSPs on the secondary structure of ArCP shows two distinct clusters of residues spanning both helices 2 and 3.

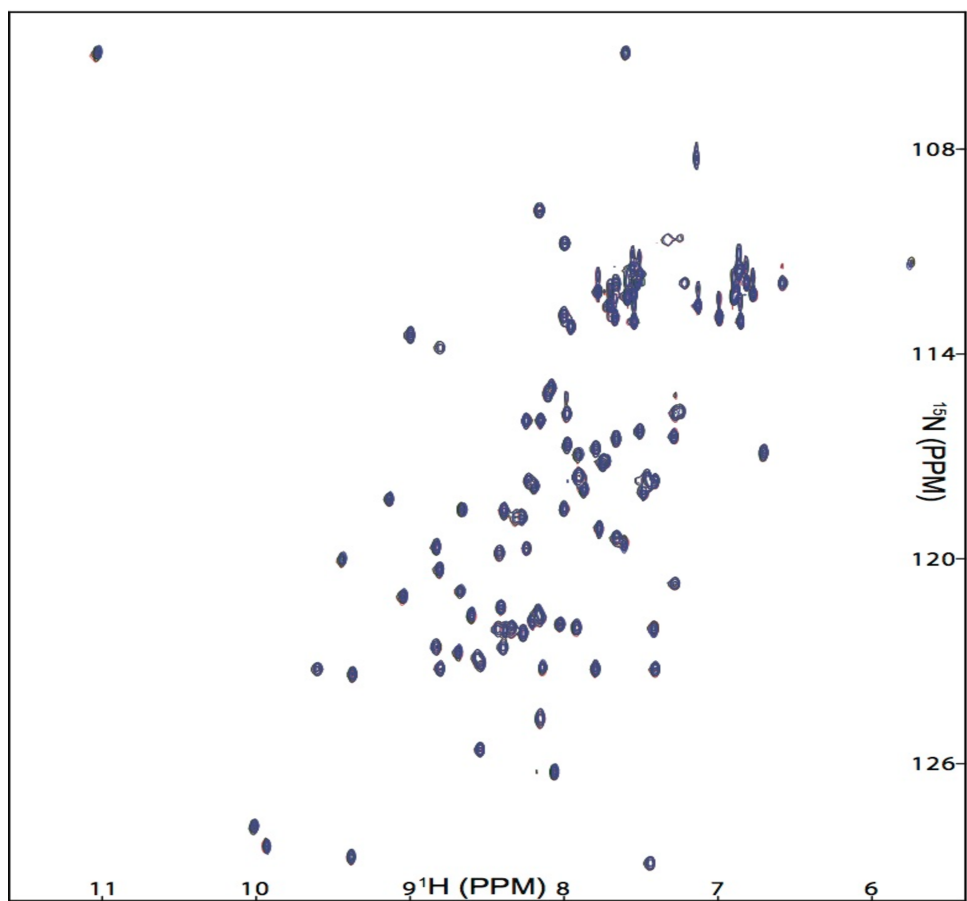


Figure 2-3. Overlay of HN-HSQC of holo ArCP (Red) and holo ArCP following the addition of salicylate (Green) and ATP (Purple) to 2 mM each. No CSPs could be detected.

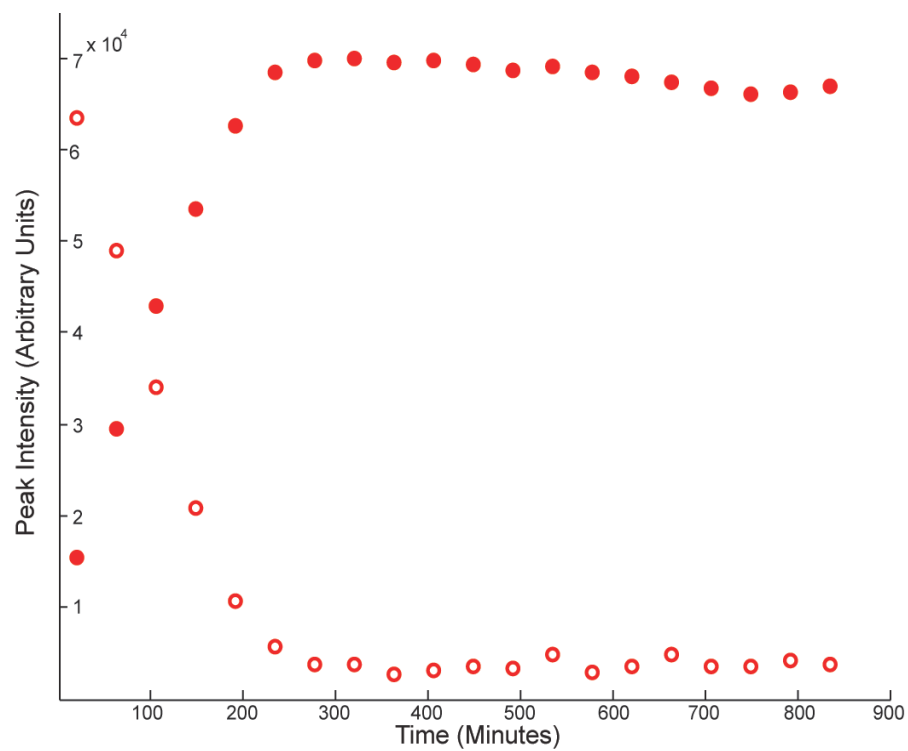


Figure 2-4. Time course of the loading reaction monitored with the signals of S52. The signal of holo ArCP (open circles) decreases over time as the signal of loaded ArCP (filled circles) increases.

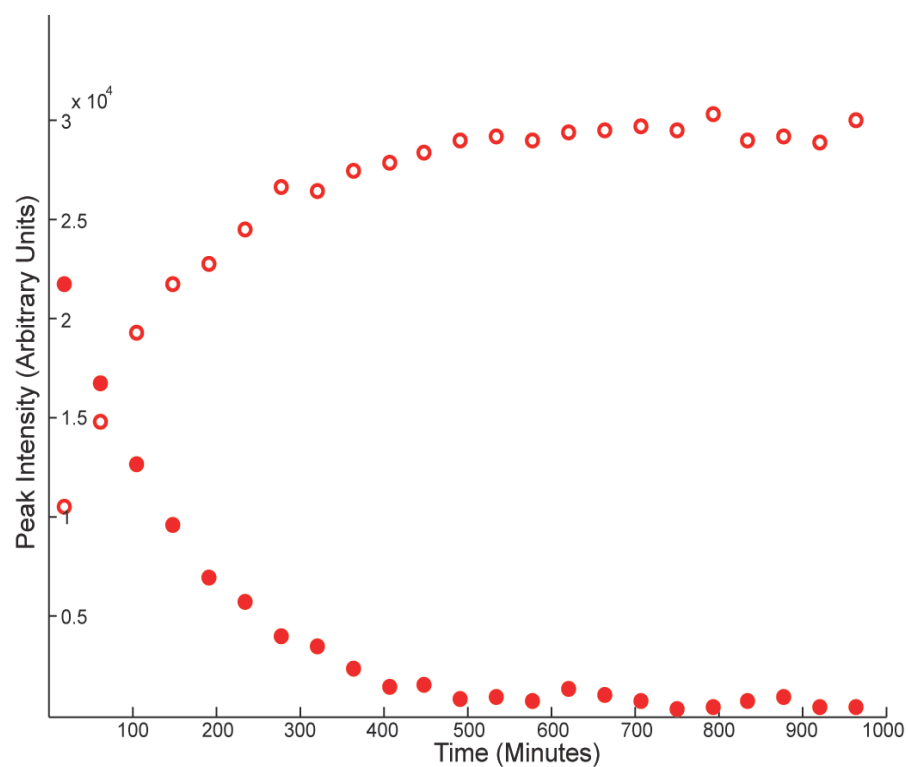


Figure 2-5. Time course of substrate hydrolysis catalyzed by SrfTEII showing the rapid decay of the signal of S52 in loaded ArCP (filled circles) with a corresponding increase in the signal of holo ArCP (open circles).

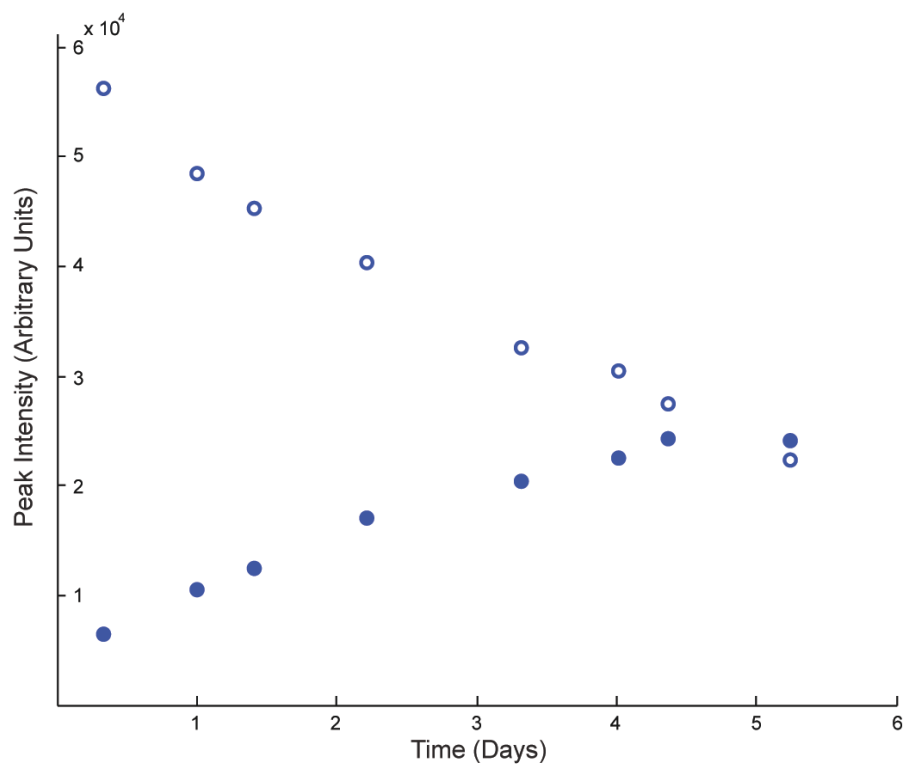


Figure 2-6. Representative time course of uncatalyzed hydrolysis of loaded ArCP, showing the decay of the signal of loaded ArCP for residue S52 (open circles) and buildup of the corresponding holo signal (filled circles).

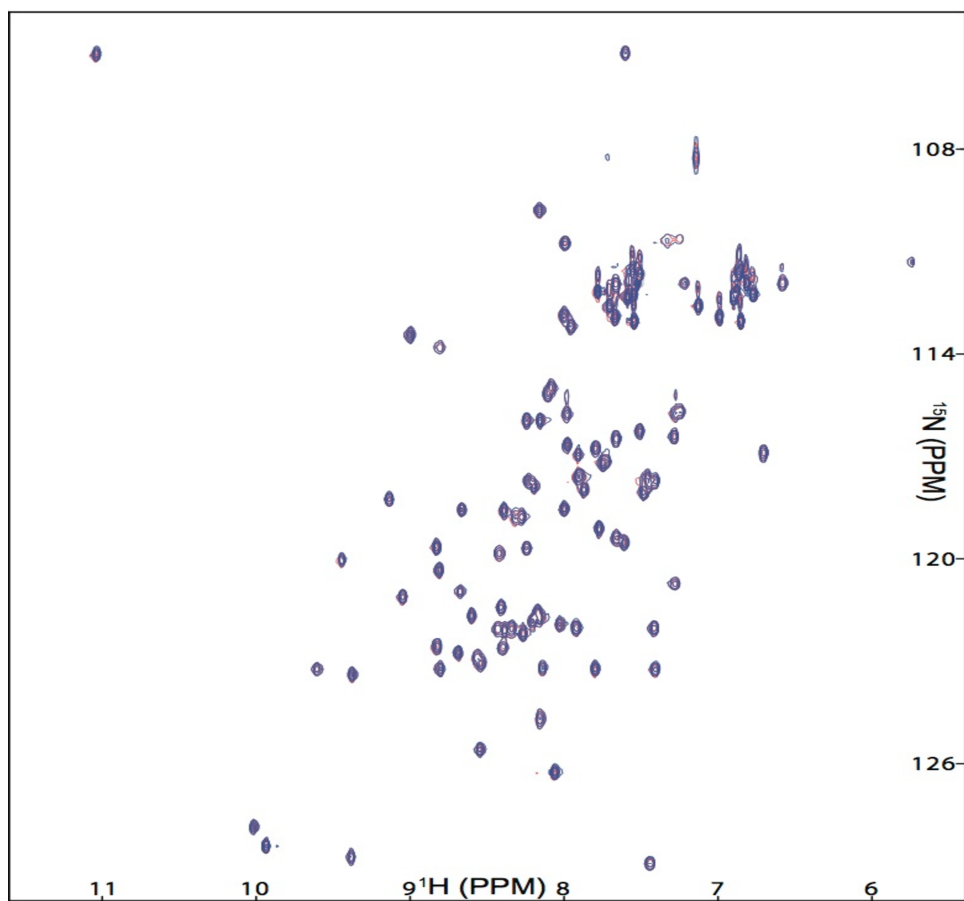


Figure 2-7. Overlay of HN-HSQC of pure holo ArCP (Red) and holo ArCP regenerated by the action of the thioesterase SrfTEII (Purple). The spectra are identical within experimental error, demonstrating that the signals observed in loaded ArCP indeed denote the effect of substrate loading.

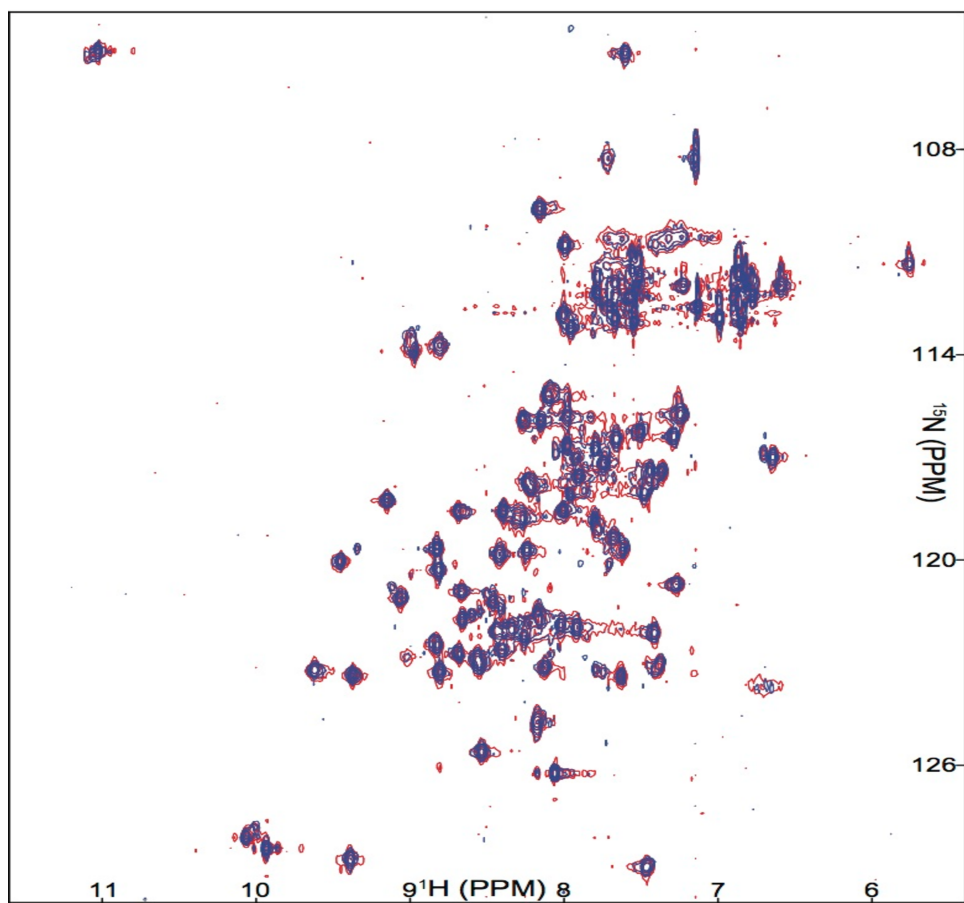


Figure 2-8. Overlay of HN-HSQC of salicylate loaded ArCP immediately after loading reached completion (Red) and after 5 days (Purple).

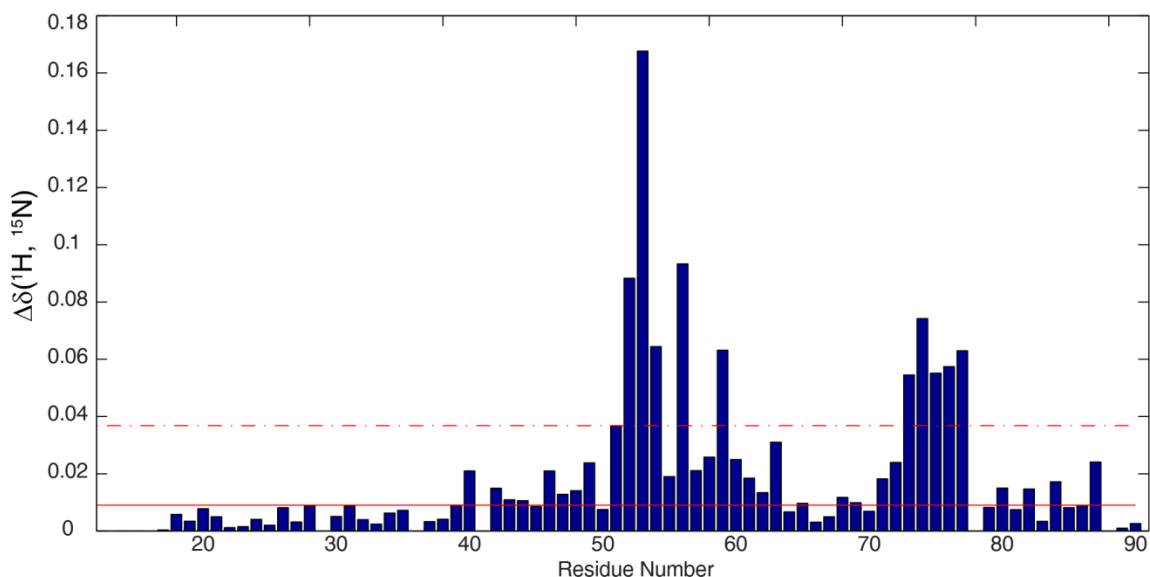


Figure 2-9. Chemical Shift Perturbation comparing chemical shifts of holo-ArCP and ArCP loaded with Sal. The reported chemical shift differences are calculated with $\Delta\delta(^1\text{H}, ^{15}\text{N}) = ((\Delta\delta_{\text{H}})^2 + (1/5(\Delta\delta_{\text{N}})^2))^{1/2}$ where $\Delta\delta_i$ is the chemical shift difference between the two species for nucleus i. The horizontal solid line indicates the median (0.009 ppm) and the dashed line indicates one standard deviation above the median (0.036 ppm). Residues reported in Figure 1 of the main text are those above the dashed line.

Chapter 3-Solution structure of a nonribosomal peptide synthetase carrier protein loaded with its substrate reveals transient, well-defined contacts

This chapter was published with minor modification in *The Journal of the American Chemical Society*, September 2015, by Andrew Goodrich, Bradley Harden, and Dominique Frueh

Abstract

Nonribosomal peptide synthetases (NRPSs) are microbial enzymes that produce a wealth of important natural products by condensing substrates in an assembly line manner. The proper sequence of substrates is obtained by tethering them to phosphopantetheinyl arms of holo carrier proteins (CPs) via a thioester bond. CPs in holo and substrate-loaded forms visit NRPS catalytic domains in a series of transient interactions. A lack of structural information on substrate-loaded CPs has hindered our understanding of NRPS synthesis. Here, we present the first structure of an NRPS aryl carrier protein loaded with its substrate via a native thioester bond, together with the structure of its holo form. We also present the first quantification of NRPS CP backbone dynamics. Our results indicate that prosthetic moieties in both holo and loaded forms are in contact with the protein core, but they also sample states in which they are disordered and extend in solution. We observe that substrate loading induces a large conformational change in the phosphopantetheinyl arm, thereby modulating surfaces accessible for binding to other domains. Our results are discussed in the context of NRPS domain interactions.

Introduction

Nonribosomal peptide synthetases (NRPSs) are enzymatic systems found in bacteria and fungi responsible for the production of a myriad of secondary metabolites. These systems are capable of generating exceptionally complex and diverse natural products from simple starting materials such as amino and aryl acids by utilizing a modular architecture. Multiple modules are arranged in an assembly-line fashion to comprise the full synthetase, an organization also encountered in related fatty acid synthetases (FASs) and modular polyketide synthetases (PKSs)^{20,41–48}. Each module within an NRPS is comprised of at least three core domains whose combined action leads to the selection, activation, and incorporation of a single small molecule into the growing peptide^{1,49,50}. NRPS modules select starting materials from a pool of hundreds of small molecules including the 20 standard L-amino acids, small aryl acids, and D-amino acids. Assembling these small molecules in a combinatorial fashion creates the potential to generate enormous chemical and functional diversity. A central aspect of this successful strategy is the covalent tethering of chemical substrates to the assembly line, which occurs through so-called carrier proteins.

Each module within a NRPS is typically composed of at least an adenylation (A) domain, a condensation (C) domain, and a thiolation domain, also called a carrier protein (CP). CPs must first be activated via attached of a 4'-phosphopantetheine (PP) moiety to a conserved serine by a phosphopantetheinyl transferase⁵¹. The PP provides a thiol by which activated monomers and intermediate products are covalently tethered to the synthetase. A domains load the substrates onto CPs by catalyzing two distinct reactions. First, the select the amino or aryl acid to be incorporated and activate it using ATP via formation of a high energy

acyl-adenylate. Second, they loaded the activated monomer onto CPs via formation of a thioester with the PP. C domains then catalyze amide bond formation between products loaded on adjacent CPs, passing intermediates from an upstream donor CP to a downstream acceptor CP and extending the peptide by a single monomer. After the final monomer is incorporated, a thioesterase domain found in the final module releases the peptide via hydrolysis or macrocyclization. During synthesis, NRPS carrier proteins must interact with at least three different catalytic domains: a PPTase, an A domain, and one or more C domains. Crystallographic and nuclear magnetic resonance (NMR) studies have indicated that these interactions do not occur within the confines of a rigid assembly of NRPS domains but through a succession of transient interactions involving a dynamic quaternary structure^{9,10,14,18,52,53}. Understanding the role the chemical modifications of CPs play in orchestrating this series of transient protein-protein interactions is key to elucidating the molecular mechanism of NRPS synthesis.

Progress in understanding the molecular influences of PP and substrates on the function of carrier proteins has been impeded by the lack of structural information on substrate-loaded NRPS carrier proteins. There are several NMR and crystal structures of apo- and holo-CPs either isolated or together with other NRPS domains^{2,4,5,11,14,18,50,54–58}. Of these, the only two solution structures of isolated holo carrier proteins have produced contradictory results with respect to the influence of the PP moiety. In the first system, PP was found not only to bind to its CP but also to dramatically influence conformational fluctuations⁵. In the second study, the protein core of CP was thought to interact only weakly, if at all, with the PP, leading to the conclusion that phosphopantetheinylation does not affect the carrier protein in a relevant manner⁴. Perhaps more importantly, there are currently no structures

available of substrate-loaded CPs from an NRPS and it remains unclear whether the substrate is simply tethered on an unstructured PP arm or if it interacts with its carrier protein. Lack of success in studying structures of loaded carrier proteins results in part from rapid hydrolysis of the thioester bond, so we recently proposed a means to bypass this limitation for monomeric substrates⁸. Determining if the loaded substrate directly interacts with the protein core of a CP or remains unbound will shape our understanding of the role it plays in directing protein-protein interactions.

In order to study the influence of substrate loading on the structure of a CP, we determined the solution structures of the aryl carrier protein (ArCP) from yersiniabactin synthetase in its holo and substrate-loaded forms. Yersiniabactin (Figure 3-1E) is an iron chelator and virulence factor for *Yersinia pestis*, the causative agent of the bubonic plague, and its biosynthesis has been extensively characterized^{22,59-64}. The yersiniabactin synthetase system is comprised of the stand-alone A domain YbtE, the multidomain proteins HMWP2 (Figure 3-1A) and HMWP1, and the reductase YbtU. ArCP is the first carrier protein involved in yersiniabactin synthesis. It composes the N-terminal 100 residues of HMWP2, which also contains two cyclization domains (Cy1 and Cy2), an adenylation domain (A), two peptidyl carrier proteins (PCP1 and PCP2), and an epimerization domain (E) (Figure 3-1A). Cyclization domains are related to condensation domains yet catalyze a cyclodehydration in addition to condensation. As a starter carrier protein, ArCP has a relatively simple lifecycle. Following activation of carrier proteins to holo forms, YbtE loads ArCP with salicylic acid while the A domain of HMWP2 loads PCP1 with cysteine (Figure 3-1B). Cy1 then catalyzes peptide bond formation and cyclization between the substrates loaded on ArCP and PCP1, regenerating holo-ArCP and producing PCP1 now loaded with a 2-

hydroxyphenylthiazoline moiety (Figure 3-1C). Two forms of ArCP communicate with catalytic domains during synthesis: holo-ArCP is a substrate for YbtE and a product for Cy1, where ArCP loaded with salicylate (hereafter referred to as ‘loaded-ArCP’) is a product of YbtE and a substrate for Cy1.

We have previously established a method for analysis of ArCP loaded with salicylate in its native thioester form by nuclear magnetic resonance spectroscopy⁸. The method exploits the non-invasive nature of NMR and isotope editing to allow for sustained measurements on loaded-ArCP. Loaded-ArCP is generated in situ and the appropriate concentration of substrates and adenylation domain, Sal and YbtE, ensure that the regeneration of loaded-ArCP outperforms hydrolysis without influencing the chemical shift perturbations (CSPs) observed in the substrate-loaded form. We found that these CSPs occur along multiple distinct structural elements of the core of the protein. However, CSPs cannot differentiate among a direct interaction with the substrate, structural changes in the ArCP, a modulation of protein dynamics, or some combination of these effects. Here, we present the solution structures of ArCP in the holo and salicylate-loaded forms together with a characterization of the dynamics of holo and loaded-ArCP, including the PP. We show that holo-ArCP has a transient yet well-defined interaction with the PP and that the protein core of ArCP has a direct interaction with loaded salicylate, and we discuss the role these observations may play in influencing protein-protein interactions in NRPS systems.

Results and Discussion

Acyl, aryl, and peptidyl carrier proteins play a central role in PKSs, FASs, and NRPSs as they shuttle substrates between various catalytic sites during synthesis. In the past few

years, various structural and functional studies have suggested that NRPSs are not rigid assemblies but are subject to inter-domain dynamics. Notably, rather than simply swinging the phosphopantetheinyl arm between active sites, the entire carrier protein is expected to visit partner domains in a series of transient domain interactions. Understanding the biosynthesis of all associated products thus requires an understanding of the molecular parameters that modulate these successive transient domain interactions^{9,14,18,52,53}. Here, we will describe how both the phosphopantetheinyl arm and its tethered substrate interact with the core of an NRPS aryl carrier protein in its holo and substrate-loaded form. We found that although these interactions are well-defined they are nevertheless transient as a substantial population of both holo- and loaded-ArCP possesses prosthetic groups subject to large amplitude motions. Our results provide novel insights into interactions between carrier proteins and PP and tethered substrates, and our observations will be discussed within the framework of domain interactions.

Structures of holo- and loaded-ArCP

Holo-ArCP and loaded-ArCP both display the right-handed helical bundle fold typical of carrier proteins^{2,65,66} (Figure 3-2A-D). This fold consists of three long helices with an up, down, down arrangement (α 1, A19-E31, α 2, S52-K64, and α 4, L80-M87) and an additional shorter helix (α 3, L71-A76). Here, secondary structure boundaries are defined as observed in the mean structure of holo-ArCP. The post-translational modification site, S52, is located at the N-terminal end of α 2. In ArCP, the four helices are packed predominantly by a hydrophobic core substantiated with aromatic interactions (F62, W83, Y67, W58, and W61 partially). In addition, in holo-ArCP, a salt bridge can form between E31 and R54 (α 1/ α 2) although R54 may also interact with D51 with different rotamers.

$\alpha 1$ and $\alpha 2$ are linked by loop1, which is well-defined (Figure 3-3) and features a single-turn helix, αI (L45-A48). Loop1 is held in place by a number of hydrophobic interactions with $\alpha 1$, $\alpha 3$, and $\alpha 4$ and likely with an ionic interaction between E41 and R25 in the holo form. These interactions allow for a large number of distance constraints to be detected, which anchors loop1 to the helical bundle core of ArCP in the NMR ensemble. $\alpha 3$ lies in between loop 2 (G66-T70) and the very short loop 3 (A77-T79). Overall, ArCP in both the holo and loaded forms adopts a well-defined and compact protein fold.

A large number of NOESY cross-peaks revealed direct contacts between PP and ArCP in holo-ArCP and between PP, Sal, and ArCP in loaded-ArCP. In holo-ArCP, a total of 50 cross-peaks involving PP could be assigned unambiguously (Figures 3-4, 3-5, and 3-10). They include 14 nOe's within the PP arm. These nOe's have the same sign as the source peaks (so-called diagonal signals in NOESY experiments) which demonstrates that the PP arm is subject to the same molecular tumbling as the carrier protein. In addition, 36 nOe's between the protein core and PP could be identified. These nOe's denote a contact between PP and ArCP. Together, these cross-peaks were used as constraints to define the conformation of PP and to position it on the surface of ArCP. Similarly, NOESY cross-peaks permitted the determination of the structure of PP tethered to salicylate in loaded-ArCP (Figures 3-6, 3-7, 3-8, 3-9, and 3-10). Here, 12 strong and unambiguous nOe's were observed between loaded-ArCP and Sal and 8 with PP. In addition, 11 constraints could be determined within the Sal-PP moiety. As for holo-ArCP, the NOESY cross-peaks are indicative of a stable, well-defined interaction between loaded-ArCP and its prosthetic group.

Substrate loading alters the conformation of the PP arm in a dramatic manner and modifies the surface of the protein sustained by $\alpha 2$, $\alpha 3$, and nearby regions. In holo-ArCP, the phosphopantetheinyl arm is extended and docks on ArCP along $\alpha 2$ on one side and $\alpha 3$ on the other side (Figure 3-2A,B). Such a conformation masks a large area of the solvent-exposed surface of $\alpha 2$ and $\alpha 3$. This surface involves residues S52, I53, M56, and L59 on $\alpha 2$, and L71, R72, and Y75 on loop 2 and $\alpha 3$. Upon substrate loading, the phosphopantetheinyl arm adopts a curled conformation that accommodates salicylate binding (Figure 3-2, C,D, and F). Sal binds at the surface of ArCP in a region defined by the N-terminal end of $\alpha 2$, the C-terminal end of loop 1, and the C-terminal end of $\alpha 3$. Together, Sal and PP cover a surface involving residues L50-I53, M56, R72, Y75, A76, and I46. Hence, the change of conformation induced by tethering of salicylate exposes a region previously covered by PP in holo-ArCP and masks a new region of ArCP. Both regions are involved in domain communication and the significance of this alteration in surface access will be discussed within the framework of domain interactions below.

Comparison of the protein cores of holo- and loaded-ArCP reveals that subtle changes occur upon substrate loading. As underlined, the overall core adopts the same fold in holo- and loaded forms. Three helices, $\alpha 1$, $\alpha 2$, and $\alpha 4$, are slightly shorter after substrate loading. Accordingly, there is a slight variation in the relative distances between helices 1, 2, and 3. The N-terminal end of $\alpha 2$ moves towards $\alpha 1$ and away from $\alpha 3$, in a manner that accommodates substrate docking in loaded-ArCP (Figure 3-2B and D). While doing so, $\alpha 2$ rotates slightly and the phosphate group on S52 is moved towards the outside of the protein (perhaps best seen in Figures 3-10 and 3-11), and the helices in the bundle become more parallel to one another. Of note: in loaded-ArCP, the angles of residues in αI are interpreted

as helical by some software packages (molmol, see e.g. Figure 3-12C) but not by others (PyMOL, see Figure 3-2C and D). The spatial arrangement of the residues involved is however rather well-maintained in holo- and loaded-ArCP (Figure 3-2E). A major difference between the NMR ensembles of holo-ArCP and loaded-ArCP lies in a change of conformation of loop 1, between L41 and N46 (Figure 3-2E). This region packs against the helical bundle with hydrophobic interactions towards the core, but also with interactions involving side-chains of $\alpha 1$ and $\alpha 4$ that are more peripheral. In holo-ArCP, the latter provide more NMR constraints while in the loaded form constraints with the core dominate. This change in conformation shifts the mean positions of the amide proton of E41 by 4 angstroms. As a consequence, the ionic interaction previously mentioned between E41 and R25 can only occur in the holo form. Loop 1 has been shown to be involved in a number of interactions with NRPS catalytic domains^{5,10,11,54,67-69} and its change in conformation is discussed further below.

Dynamics in the ArCP core

The protein cores of holo- and loaded-ArCP are mainly rigid on fast time-scales but undergo local conformation fluctuations at slower time-scales. To assess the dynamics of holo- and loaded-ArCP, we measured the nitrogen longitudinal (Figure 3-13A) and transverse (Figure 3-13B) relaxation rates (R_1 and R_2 , respectively) as well as the heteronuclear NOE between amide protons and nitrogens (HN-NOE, Figure 3-13C). R_2 is sensitive to both motions in picosecond and nanosecond time scale and in microsecond or slower time scales. R_1 is sensitive exclusively to picosecond to nanosecond motions. HN-NOE is a direct reporter of picosecond to nanosecond fluctuations in bond orientations. We applied the Lipari-Szabo formalism⁷⁰ as developed by Palmer and co-workers⁷¹ and as

implemented by Fushman and co-workers^{72,73} to obtain order parameters, S^2 , that provide a measure of the amplitude of ps-ns motions (a value of 1 indicates rigidity and 0 denote complete disorder). Figures 3-13D and 3-14 reveal that all helices and a large part of the connecting loops are relatively rigid for both forms and only a few selected residues display increased flexibility. They are L34 and T35 in the beginning of loop 1, D51, the residue preceding the phosphopantetheinylation site, and R68 and L69, both in loop 2. No fewer than nine studies reported flexibility in the N-terminal half of loop 1 in FAS and PKS acyl carrier proteins (ACPs)⁷⁴⁻⁸². We speculate that such flexibility may be required to achieve different interactions with different NRPS partner domains. The position of L34 (i-18 with respect to the conserved serine) was found to be at the interface between an adenylation domain and a related aryl carrier protein (EntB)¹⁰.

D51 stands out as the single residue with marked fast dynamics at the end of loop 1 in both holo and loaded forms (Figure 3-13D). This is a rather critical observation as D51 is the predecessor of the PP site, a position that has been shown to be actively involved in many domain interactions^{10,11,14,18,56,57,67,68} and even hypothesized to be participating in enzymatic activity⁶⁷, so this position must have access to many conformations to satisfy its role in these protein interactions. Upon re-inspecting previously reported relaxation reports and order parameters, we found that this position is flexible in five ACPs⁷⁶⁻⁸². This conserved flexibility likely reflects the versatile role that the position preceding the PP site plays during synthesis.

Fast internal motions have also been detected in the region encompassing R68 and L69 (loop 2) in FAS and PKS ACPs. Flexibility has been consistently probed in loop 2^{74-76,78,79,82}, sometimes extending to the adjacent $\alpha 3$ ^{80,81}. $\alpha 3$ has often been found to be subject to

conformational exchange and its conformation as well as its relative orientation are modulated upon interactions with substrates and partner domains^{5,74,75,79}. The conserved flexibility of loop 2 likely permits modulation of the relative orientation between helices $\alpha 2$ and $\alpha 3$.

Overall, substrate loading does not affect the flexibility of ArCP in a dramatic manner. Nevertheless, comparison of order parameters indicates a trend for a rigidification at ps-ns time-scales for a few residues. The small amplitude of these effects, however, prevents us from commenting on their significance.

Slower conformation fluctuations (μ s) increase the magnitude of R_2 for two groups of residues (Figure 3-13B, color coded in Figure 3-14A and B). A18, A19, and D20 are located at the N-terminal end of $\alpha 1$ and Q37, H40, E42, S43, and L45 are all in the center of loop 1, up to the single turn helix, αI . All residues display R_{ex} in the Model Free analysis (Figure 3-15). These two groups are remote in the structures of ArCP and these fluctuations likely reflect separate events.

The conformational fluctuations observed in loop 1 are likely relevant to ArCP's activity. Conformational exchange has been detected in corresponding regions of FAS and PKS ACPs^{76,82} and other regions of loop 1^{80,81,83}. Most strikingly, the two residues that display the largest R_2 rates are H40 and E42 that flank E41 (Figure 3-14A and B). E41 could not be detected, presumably because its NMR resonances have been broadened by exchange beyond detection. We mentioned previously that E41 can form a salt bridge in holo-ArCP but not in loaded-ArCP because the surrounding region of loop 1 changes conformation upon substrate loading. We observe an overall reduction of R_{ex} upon substrate loading, indicating that the substrate affects related conformational fluctuations (Figures 3-13B and

3-15). Indeed, salicylate docks in the vicinity of α I, which signals the end of the malleable region and the docking of Sal may modulate the dynamics of loop 1. In summary, the change in conformation that we observed when comparing holo- and loaded-ArCP structures results from a change in conformational equilibria and is not a static effect. In this context, the flexibility of T35 and L34 discussed above may reflect a hinge in this region, used to provide malleability to the region encompassing Q37-L45. Various residues in loop 1 have been shown to interact with NRPS catalytic domains^{5,10,11,54,67-69}

Transient interactions between ArCP and its prosthetic groups

Our data reveal that, although they are well-defined, the interactions we see between ArCP and its prosthetic groups are transient in nature. Our relaxation data indicate that the NMR signals of PP reflect an extreme amount of motional averaging, characteristic of disorder (Figure 3-13, A, B, and C). However, NOE's between ArCP and its prosthetic moieties as well as NOE's within PP with the same sign as those of the protein indicate that PP is bound to ArCP. Together, these observations indicate an average of NMR parameters⁸⁴ due to an equilibrium between a bound form of holo-ArCP, b-holo-ArCP, and a form in which PP does not bind and is disordered, u-holo-ArCP (unbound). Likewise, loaded-ArCP exists in both bound and unbound forms.

The detection of positive NOE's with signals that seemingly display high motional averaging is reminiscent of the transferred NOE's that occur for small molecules binding to a large protein⁸⁵⁻⁹⁰. The detection of NOE's between PP and ArCP through signals of the unstructured form of PP is a special case of population averaged NOE's in which the NOE of u-holo-ArCP is zero. Similar effects have been discussed for NOE's in protein cores and even for *E. coli* ACP^{77,91}. Our results indicate that holo-ArCP and loaded-ArCP are subject

to equilibria between bound and unbound forms (Figure 3-14A and C). Further studies are necessary to characterize the populations and time-scales of these equilibria.

In addition to our investigation of ArCP's flexibility and conformational fluctuations, we have observed signals indicative of a minor conformer of ArCP in both holo and loaded forms. Similar observations have been made for a PCP⁵ and many PKS and FAS ACPs^{74,75,78-81,83,91,92}. In one study, the second conformer was shown to involve a second binding site for the PP arm and was accompanied by alterations in $\alpha 3$ ⁷⁵. We observe a second set of signals for PP in holo-ArCP and loaded-ArCP and a similar scenario cannot be excluded for ArCP. A limited chemical shift perturbation (Figure 3-16) indicates that these conformers are likely subject to subtle changes in conformations and do not belong to an unfolded or unfolded-like state. Further studies are needed to characterize these minor conformers.

Implications for domain communication

All our observations can be revisited within the context of domain interactions and catalytic steps that occur during biosynthesis. Following post-translational modification, ArCP interacts with the adenylation domain YbtE to harvest salicylate as well as with the cyclization domain Cy1 to catalyze the condensation of salicylate with cysteine^{22,23}. In this chain of events, holo-ArCP is the substrate of YbtE and loaded-ArCP is its product. Conversely, loaded-ArCP is the substrate for Cy1 and holo-ArCP is a product of Cy1.

As noted above, interactions between the core of the protein and the tethered salicylate alter the conformation of the PP between holo- and loaded-ArCP. This change in PP conformation in turn alters the electrostatic surface presented by loop 1 and helices $\alpha 2$ and $\alpha 3$. Loop 1, $\alpha 2$, and $\alpha 3$ are all involved in binding with other domains and altering the

solvent exposed surfaces as described necessarily modulates domain affinities. Indeed, studies of related PKS ACPs have shown that changes in surface potentials could explain the success or failure in ACP domain swaps in 6-deoxyerythronolide B synthase⁹³. Inspection of the surface potential of ArCP reveals a positively charged region defined by the second half of $\alpha 2$, loop 2, and R72 in $\alpha 3$, as well as a negatively charged region delimited by the C-terminal end of loop 1, the beginning of $\alpha 2$, and part of $\alpha 3$ and loop 2 (Figure 3-11). In holo-ArCP, much of the positively charged area is covered by the PP arm, while the negatively charged region is accessible. Upon substrate loading, part of the positively charged area is covered by the PP arm, while the negatively charged surface is obfuscated by Sal and the end of the PP arm. We hypothesize that such a dramatic modification of the surface potential in a region consistently involved in domain binding^{5,10,11,68,69} likely participates in modulating the binding affinity of ArCP towards its partner domains.

The equilibrium between unbound and bound forms of holo-ArCP may be modulated by various mechanisms during interactions with adenylation domains. Like all adenylation domains, YbtE contains a large N-terminal sub-domain, A(N), and a smaller C-terminal sub-domain A(C). YbtE catalyzes two distinct steps. First, salicylate is selected and adenylated to produce an activated mixed anhydride Sal-AMP. Second, the thiol group of holo-ArCP reacts with the activated carbonyl to form a thioester bond and, hence, to load ArCP with salicylate. Crystallographic and biochemical studies have shown that the relative orientations of A(N) and A(C) changes by about 140° during the two-step reaction, with an adenylation conformation A_A active for adenylation and a thioester conformer A_T responsible for thioesterification^{9,52}. The proposed mechanism invokes a transition from A_A

to A_T upon completion of the adenylation reaction and prior to binding holo-ArCP. Thioesterification requires PP to be extended towards the adenylate, e-holo-ArCP, and thus the relevant complex consists of e-holo-ArCP bound to A_T (see figure 3-12B). Our observations of equilibria between bound and unbound forms of holo-ArCP raise questions as to whether the bound form is used during A/CP recognition and which mechanisms are compatible with such an equilibrium. An active participation of b-holo-ArCP would mean that the complex competent for thioesterification, e-holo-ArCP/ A_T , is obtained following formation of an encounter complex involving b-holo-ArCP, whereas a passive participation would require A_T to select e-holo-ArCP through u-holo-ArCP.

The active participation of b-holo-ArCP would relate to a mechanism that has recently been proposed for type II FAS ACPs upon crystallization of a type II ACP trapped in complex with a partner protein, FabA²¹. Substrates in type II ACPs are buried in the cleft between $\alpha 2$ and $\alpha 3$. Burkart and coworkers propose that FabA recognizes ACP with a surface that includes PP. Following formation of this encounter complex, the substrate is excised from ACP and PP extends towards the active site. In our system, this would mean that the adenylation domain first recognizes the surface provided by $\alpha 2$, $\alpha 3$, and PP before rearrangements occur to yield the final e-holo-ArCP/ A_T complex.

Combined insight from existing structures and the structures described here suggest how the positioning of PP may influence interactions with A domains and, in turn, how this affects each scenario described above. Complexes of e-holo-ArCP bound to A_T have been crystallized, notably with an aryl carrier protein closely related to ArCP, EntB-ArCP¹⁰. Loop 1 in EntB-ArCP interacts with A(C) at positions corresponding to residues D51, Q47, E42, and L34 in ArCP (Figure 3-12B), whereas $\alpha 2$ interacts both with A(C) at positions

R54 and I53 and with A(N) at positions M56, H60, and K64. A simple structure alignment of b-holo-ArCP onto EntB-ArCP reveals that b-holo-ArCP could maintain interactions with an A_T domain through A(C) but not A(N) (Figure 3-12A). Thus, an encounter complex between A_T and b-holo-ArCP would necessarily occur with a different domain organization than that seen in Figure 3-12B. Given the existence of dramatic reorientations between A(N) and A(C) when going from A_A to A_T, it is very likely that A_T maintains some malleability, and a complex between b-holo-ArCP and A_T may be accommodated. Clearly, a more direct conformation selection mechanism, in which A_T selects e-holo-ArCP from the u-holo-ArCP ensemble, would not be subject to any such constraints.

In either unbound or bound form, ArCP can interact with A(C). This is an important observation given that the domain reorientation in A domains was proposed to provide a means of shuttling carrier proteins between binding sites^{10,18}. If A(C) is used to shuttle holo-ArCP towards A(N), an interaction between A(C) and ArCP such as that shown in Figure 3-12 would likely shift the equilibrium towards b-holo-ArCP, as many conformations of u-holo-ArCP would be incompatible with binding. Upon reaching a conformation A_T, the adenylation domain must then select for e-holo-ArCP. This model would correspond to a sequence of conformational selection events and the equilibrium between b-holo-ArCP and u-holo-ArCP may then be a means to ensure simultaneously an interaction between A(C) and ArCP (with b-holo-ArCP paying an entropic cost and opening an interaction surface) while maintaining access to an extended form that must be selected through the unbound form for catalysis. Many mechanisms are compatible with our observations and further investigations will be necessary to decipher the mechanism of NRPS synthesis.

The loaded form of ArCP in bound form prevents interaction between ArCP and A(N) but allows for a stable interaction with A(C). Loaded-ArCP is the product of the thiolation reaction, and hence loaded-ArCP must dissociate from the A domain. Inspection of Figure 3-12C demonstrates that the curled conformation of PP in the bound form would prevent any functional interaction between b-loaded-ArCP and A_T. Thus, b-loaded-ArCP likely helps ensure that NRPS synthesis moves productively to the next step. Like b-holo-ArCP, b-loaded-ArCP can interact with A(C). Again, this observation is compatible with A(C) shuttling loaded-ArCP towards the next catalytic partner, here Cyl.

Inspection of the conformation of loop1 within the context of A/CP complex suggests that substrate loading may help ArCP interact with the C-terminal subdomain of YbtE. Figure 3-12D emphasizes the changes in conformation that occur in loop 1 for holo-ArCP, loaded-ArCP, and holo-EntB-ArCP. It is immediately apparent that loop 1 moves towards A(C) when comparing holo-ArCP, loaded-ArCP, and EntB-ArCP. Loop 1 simultaneously adopts a conformation that is increasingly open with respect to the ArCP core protein. With EntB-ArCP representing an optimal interaction, this observation suggests that loaded-ArCP adopts a conformation more suitable for an interaction with A(C). Thus, the bound form of loaded-ArCP may affect the affinity towards A domains in two manners, first by disrupting interactions with A(N) and adopting a conformation incompatible with a competent complex, and second by promoting an interaction between ArCP and A(C) by stabilizing loop 1 in an open form. Substrate loading would then actively contribute to release from A(N) and shuttling of ArCP by A(C). Further studies are needed to probe this hypothesis. The bound form of loaded-ArCP likely modulates interactions with partner cyclization or condensation domains, while access to the extended form needed for catalysis is provided

by the unbound form. ArCP is the first carrier protein of the yersiniabactin synthetase and, as such, Sal-loaded-ArCP is exclusively a substrate-donor carrier protein. There are no structures of condensation domains in complex with substrate-donor carrier proteins and we cannot make detailed mechanistic predictions for loaded-ArCP as those we discussed for holo-ArCP. Fortunately, biochemical studies were conducted specifically with ArCP or EntB-ArCP, and they identified residues in $\alpha 1$, loop 1, and $\alpha 3$ that are necessary for communication with condensation domains^{69,94}. These residues were identified through combinatorial mutagenesis and selection and correspond to M56, L71, Y75, as well as I53, A76, and N44. They span the solvent accessible surface of N-terminal $\alpha 2$ (I53 and M56), C-terminal $\alpha 3$ (Y75 and A76), and one residues in loop 1 (N44). Strikingly, all these residues interact with PP or Sal in b-loaded-ArCP. It may well be that selection was achieved by optimizing interactions with the tethered substrate. If so, this suggests that bound loaded-ArCP is used as a substrate for an encounter complex. Access to e-loaded-ArCP through its unbound form would then be used to reach the Cy (or C) domain active site, which is located far from the surface of these domains. Clearly, our observations must be revisited once the structure of a donor ArCP in complex with a cyclization or condensation domain is available.

In summary, the structures of b-holo-ArCP and b-loaded-ArCP that we have determined, together with our observations of equilibria between bound and free forms, set a framework to elucidate mechanisms for domain recognition and domain rearrangements during NRPS synthesis.

Conclusions

We have presented the first solution structure of a nonribosomal peptide synthetase carrier protein loaded with its substrate, the backbone dynamics analysis of an NRPS carrier protein, and the first solution structure of an aryl carrier protein, both in holo and loaded forms.

The solution structures of ArCP indicate that the phosphopantetheinyl arm interacts with the protein core of ArCP both in holo and in loaded forms. Substrate loading induces a large conformational change in the PP arm that alters the nature of the protein surface in the surrounding regions, which are involved in domain recognition. In holo-ArCP, PP lies in an extended conformation between helices $\alpha 2$ and $\alpha 3$, whereas in loaded-ArCP PP curls back to allow for substrate binding in a region near the phosphopantetheinylation site, defined by loop 1, $\alpha 2$, and $\alpha 3$. The repositioning of the PP arm modulates access to regions of ArCP with distinct electrostatic potentials, providing a rationale for altering the binding affinity of ArCP to its partner domains.

Joint analysis of NOESY spectra and NMR spin relaxation indicate that ArCP interacts with its prosthetic groups in a transient yet well-defined manner. Our findings indicate that both holo- and loaded-ArCP undergo conformational equilibria between unbound and bound forms. Interactions with PP have been occasionally observed in NMR studies of NRPS, PKS, and FAS carrier proteins and the (well-established) motional averaging of NMR parameters we discussed can be applied to revisit observations made for these systems.

Many molecular properties have been discussed within the context of NRPS, PKS, and FAS domain communication and our results suggest that the dynamics of the prosthetic group must be considered as well. That is, binding to carrier proteins must occur through a

conformational selection of unbound or bound forms, potentially with subsequent induced fits or conformation selection events. In addition, we have observed dynamics in the protein cores of holo- and loaded-ArCP that cover all time-scales, in agreement with multiple reports. We discussed the relevance of flexible and malleable regions in ArCP within the context of interactions with partner domains, with emphasis on adenylation domains. Notably we have found that substrate loading affects conformational fluctuations occurring in loop 1, a region involved in domain communication. Together, our results substantiate a mechanism for NRPS synthesis that relies on the interplay between chemical modification and modulation of molecular properties, such as dynamics and surface potentials, to direct domain communication.

Methods

Cloning of ArCP 14-93 and Sfp into pET-Duet-1

[illegible]

sequencing. This plasmid directs production of Sfp containing a C-terminal hexahistidine tag and residues 14-93 of HMWP2 with an N-terminal GB1 tag followed by a hexahistidine tag and a TEV cleavage site. Following TEV cleavage of GB1-TEV-ArCP14-93, a GT sequence remains at the N-terminus of ArCP14-93.

Cloning of YbtE, ArCP 14-93 (without Sfp), Sfp, their expression and purification were performed as previously described⁸.

Purification of holo-ArCP Made from Coexpression with Sfp

The purification of holo-ArCP resulting from coexpression with Sfp is identical to that reported for apo ArCP⁸ except for the following modifications. pET-DUET-SfpHis6-GB1-TEV-ArCP14-93 was transformed into Δ EntD cells (courtesy Drs. Chalut and Guilhot, CNRS, Toulouse, France). Amipicillin was used instead of kanamycin. Following overnight cell growth at 37°C, the temperature was lowered to 18 °C when reaching an optical density of 0.4. At an optical density of 0.6, IPTG was added to 0.5 mM and growth continued at 18 °C. Cells were harvested 4-4.5 hours after induction at an optical density of 1.1-1.2. Following digestion by TEV protease and HisTrap purification, dithiothreitol (DTT) was added to a final concentration of 10 mM.

In vitro phosphopantetheinylation of apo ArCP was performed as described previously. Completion of the phosphopantetheinylation reaction was confirmed by HN-HSQC.

NMR Data for Assignment and Structure Determination

All spectra were collected at 25 °C on a 600 MHz Bruker Avance III spectrometer equipped with a QCI cryoprobe. All NMR spectra were processed using NMRPipe⁹⁵ and analyzed using CARA⁹⁶. Details of NMR acquisitions are described below

To ensure long-term stability of NMR samples, samples were buffer exchanged (>125-fold) into freshly prepared NMR buffer containing 0.05% (w/v) sodium azide by repeated concentration and dilution immediately prior to use. D₂O was added to all samples to a final concentration of 10% and DSS was used for internal referencing. For samples in D₂O, NMR buffer was prepared as described above in 99.8% D₂O (Aldrich Chemistry) and the pH adjusted using sodium deuteroxide (40% in D₂O, 99.5% D, Cambridge Isotope Laboratories) to a pH of 6.40 (pD=6.80).

¹⁵N-Holo-ArCP used for the holo and loaded samples was generated in vitro using apo ArCP, purified Sfp, and unlabeled coenzyme A. The NOESY-HN-HSQC of holo-ArCP was run on a 520 μM sample in the standard NMR buffer containing 1 mM DTT instead of 500 μM TCEP. To prepare the loaded sample, salicylic acid and ATP were added to 2 mM and YbtE added to 100 nM to 400 μM ArCP and the loading reaction monitored by HN-HSQC. The NOESY-HN-HSQC was begun immediately upon completion of loading. 3D-NOESY-HN-HSQC (16 scans, 20148 (¹H, 16.109 ppm at 4.696 ppm) × 40 (¹⁵N, 26 ppm at 117 ppm) × 120 (¹H, 11 ppm at 4.696 ppm) complex points, mixing time of 90 ms, 4 days 12 hrs) were recorded for sample of apo-, holo-, and loaded-ArCP. All spectra were linear-predicted once and zero-filled to the nearest power of two.

^{13}C , ^{15}N apo ArCP has previously been assigned⁸ and its assignment was used as a starting point for assigning holo- and loaded-ArCP resonances.

^{13}C , ^{15}N -Holo-ArCP was generated by coexpression with Sfp such that the phosphopantetheine cofactor is also labeled. A first sample was prepared in buffered H_2O to complete resonance assignments and collect dihedral angle constraints. The following experiments were run on a 320 μM samples in 90% H_2O /10% D_2O : HNCO (16 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 75 (^{13}C , 11 ppm at 174 ppm) complex points, 10 hrs 20 mins), HNCA (16 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 100 (^{13}C , 30 ppm at 52 ppm) complex points, 13 hrs 44 mins), HNCACB (32 scans, 2048 (^1H , 16.109 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 150 (^{13}C , 60 ppm at 42 ppm) complex points, 1 day 18 hrs 14 mins), HNCACO (32 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 75 (^{13}C , 11 ppm at 75 ppm) complex points, 20 hrs 55 mins), and HcccoNH (16 scans, 2048 (^1H , 16.109 ppm at 4.698 ppm) \times 28 (^{15}N , 26 ppm at 117 ppm) \times 50 (^1H , 7 ppm at 4.698 ppm) complex points, 1 day 7 hrs 25 mins).

A 390 μM sample was prepared in D_2O by repeated concentration and dilution in NMR buffer prepared in D_2O until a 900-fold dilution had been achieved. This sample was used for aromatic side-chain resonance assignment and to collect distance constraints involving aliphatic and aromatic protons. The following experiments were run in D_2O : 2D-hbCBcgcdHD (192 scans, 2048 (^1H , 16.1092 ppm at 4.696 ppm) \times 52 (^{13}C , 22.002 ppm at 30 ppm) complex points, 3 hrs 18 mins) and 3D-HC-HSQC-NOESY (16 scans, 2048 (^1H ,

16.0192 ppm at 4.696 ppm) \times 100 (^1H , 13 ppm at 4.696 ppm) \times 50 (^{13}C , 50.0003 at 69.5 ppm) complex points, mixing time of 90 ms, 4 days 15 hrs 29 mins).

^{13}C , ^{15}N -Loaded-ArCP was prepared by adding salicylic acid and ATP to 2 mM (final concentration) and YbtE to 100 nM (final concentration) to 413 μM holo-ArCP and monitoring the loading reaction by HN-HSQC. After completion of the loading reaction, the sample was diluted 15-fold in identical buffer to lower the concentration of adenosine monophosphate (AMP) and pyrophosphate (PPi) produced by the loading reaction and additional YbtE added to increase the final concentration to 250 nM. This was concentrated to a final ArCP concentration of 360 μM . This sample was used to complete resonance assignments and collect dihedral angle constraints. The following experiments were performed on a sample in 90% H_2O /10% D_2O : HNCQ (16 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 75 (^{13}C , 11 ppm at 174 ppm) complex points, 10 hrs 20 mins), HNCA (16 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 100 (^{13}C , 30 ppm at 52 ppm) complex points, 13 hrs 44 mins), HNCACB (32 scans, 2048 (^1H , 16.109 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 150 (^{13}C , 60 ppm at 42 ppm) complex points, 1 day 18 hrs 14 mins), HNCACO (32 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 64 (^{15}N , 26 ppm at 117 ppm) \times 75 (^{13}C , 11 ppm at 75 ppm) complex points, 20 hrs 55 mins). Following the HNCACO, the sample was buffer exchanged 150-fold into identical buffer as before to remove AMP and PPi and concentrated to 400 μM and the following experiment performed: HcccNH (16 scans, 2048 (^1H , 16.019 ppm at 4.698 ppm) \times 28 (^{15}N , 26 ppm at 117 ppm) \times 60 (^1H , 7 ppm at 4.698 ppm) complex points, 1 day 14 hrs 7 mins).

A sample in D₂O was used for aromatic side-chain resonance assignment and to collect distance constraints involving aliphatic and aromatic protons. Loading was performed with 50 μ M holo-ArCP, 50 nM YbtE, 2 mM ATP, and 500 μ M ¹³C-salicylate in a total volume of 3.95 ml in 90% H₂O/10% D₂O and loading monitored by HN-HSQC. After loading was complete, the sample was buffer exchanged 560-fold into NMR buffer prepared in D₂O containing 2 mM ATP and 500 μ M ¹³C-salicylate and concentrated to a final loaded-ArCP concentration of 360 μ M and the following experiments run: 2D-hbCBcgcdHD (800 scans, 2048 (¹H, 16.1092 ppm at 4.696 ppm) \times 52 (¹³C, 22.0002 ppm at 30 ppm) complex points, 13 hrs 39 mins) and 3D-NOESY-HC-HSQC (16 scans, 2048 (¹H, 16.1092 ppm at 4.696 ppm) \times 100 (¹H, 13 ppm at 4.696 ppm) \times 50 (¹³C, 50.0003 at 69.5 ppm) complex points, mixing time of 90 ms, 4 days 15 hrs 54 mins).

Relaxation Experiments

¹⁵N-Holo-ArCP was generated by coexpression with Sfp such that the phosphopantetheine cofactor is also labeled. T1 and {HN}-heteronuclear-NOESY experiments were run on 300 μ M (T1 and het-NOE) and 380 μ M (T2) holo samples with the following parameters: T1: 24 scans, 2048 (¹H, 16.1092 ppm at 4.700 ppm) \times 128 (¹⁵N, 26 ppm at 117 ppm) complex points, 3s recycling delay, and relaxation delays of 0, 0.510, 1.02, and 1.53 seconds, collected in that order; T2: 16 scans, 2048 (¹H 16.0192 ppm at 4.696 ppm) \times 128 (¹⁵N, 26 ppm at 117 ppm) complex points, 4s recycling delay, and relaxation delays of 0.0, 0.140, 0.350, 0.210, 0.070, and 0.280 seconds collected in that order; {HN}-heteronuclear-NOESY: 100 scans, 2048 (¹H, 16.0192 ppm at 4.700 ppm) \times 78 (¹⁵N, 26 ppm at 117 ppm)

complex points, 5s recycling delay for reference experiment and 2s recycling delay followed by 3s saturation by 120° ^1H pulses every 5 ms for saturation experiment.

Salicylate-loaded ArCP was generated by incubating 50 μM ^{15}N -holo-ArCP with 2 mM unlabeled salicylic acid, 2 mM ATP, and either 100 nM (R1, {HN}-Heteronuclear-NOESY) YbtE in 3.5 ml total volume or 20 nM (R2 measurement) YbtE in 3.75 ml total volume and monitoring loading by HN-HSQC. Upon completion of loading, the sample was concentrated to 1 ml, diluted 15-fold in NMR buffer with 2 mM unlabeled salicylic acid and 2 mM ATP, and concentrated to a final protein concentration of 300 μM (R1, {HN}-Heteronuclear NOESY) or 390 μM (R2). T1, T2, and {HN}-heteronuclear-NOESY experiments were run with the following parameters: T1: 24 scans, 2048 (^1H , 16.1092 ppm at 4.697 ppm) \times 128 (^{15}N , 26 ppm at 117 ppm) complex points, 3s recycling delay, and relaxation delays of 0, 1.53, 0.510, and 1.02 seconds, collected in that order; T2: 16 scans, 2048 (^1H 16.0192 ppm at 4.696 ppm) \times 128 (^{15}N , 26 ppm at 117 ppm) complex points, 4s recycling delay, and relaxation delays of 0.140, 0.350, 0.210, 0.070, 0.280 and 0.0 seconds collected in that order; {HN}-heteronuclear-NOESY: 100 scans, 2048 (^1H , 16.0192 ppm at 4.696 ppm) \times 110 (^{15}N , 26 ppm at 117 ppm) complex points, 5s recycling delay for reference experiment and 2s recycling delay followed by 3s saturation by 120° ^1H pulses every 5 ms for saturation experiment with saturation and reference experiments collected in an interleaved manner.

Relaxation parameters were fit using the program nlinLS, part of the NMRPipe software package⁹⁵. In all experiments, line shapes in the ^1H dimension were fit using a Gaussian function and line shapes in the ^{15}N dimension were fit using a Fourier-transformed,

apodized, exponentially decaying sinusoid. Residues 20, 29, 34, 40, 88 and 91 were excluded from fitting in holo-ArCP due to severe overlap. In loaded-ArCP, residues 20, 29, 32, 34, 75, 88, and 91 were excluded for the same reason.

Lipari-Szabo model-free analysis was performed with the program ROTDIF^{72,73}. ROTDIF fits both the overall rotational diffusion tensor as well as the model-free parameters at each residue, including R_{ex} . Here, a first pass was performed to identify residues with order parameters of 0.75 or below. These residues were then excluded from analysis when fitting the global rotational diffusion tensor but were included when fitting the model-free parameters. In holo-ArCP the excluded residues were 14, 15, 51, 69, 90, and 93. In loaded-ArCP residues 14, 15, 51, 69 and 93 were excluded.

The “sausage” representations of Figure 13 were created using the PyMol⁹⁷ “putty” feature after replacing each residue’s b-factor with its corresponding value of $1-S^2$.

Structure Calculation

Assignment of NOESY cross-peaks was performed manually using CARA⁹⁶. 1636 unambiguous restraints were assigned for holo-ArCP and 1314 for loaded-ArCP. In addition, 131 and 142 angle constraints were obtained with TALOS-N⁹⁸. Structure calculations were performed using CYANA version 2.1⁹⁹. For the final structure calculation, 100 structures were calculated using 50,000 steps. The final CYANA target functions were 2.65 (for holo) and 2.99 (for loaded). There were no distance violations bigger than 0.5 Å and no angle violations larger than 3.5° in either NMR ensemble. The

average rmsd to mean for these conformers were 0.37 Å (backbone) and 0.88 Å (heavy) for holo and 0.36 Å (backbone) and 0.85 Å (heavy) for loaded. Other rmsd's are described in the reminder of the text. The 20 structures with the lowest target function were chosen for water refinement in explicit solvent using CNS. Modified parameter and topology files were generated using the ACPYPE web application and refinement run using modified RECOORD scripts. The NMR ensembles were analyzed with the protein structure validation suite, PSVS, that includes PROCHECK_NMR¹⁰⁰ and MolProbity^{101,102}. Ramachandran statistics (PROCHECK) are: 93.6% in most favored region, 6.4% in additionally allowed, 0.0% in generously allowed, and 0.01% (G49 in loop1) in disallowed region for holo-ArCP and 92.4%, 7.6%, 0.0% and 0.0%, respectively, for loaded-ArCP. See also Tables 3-1 and 3-2.

Surface potentials were generated using the APBS Tools2.1¹⁰³ plugin for PyMOL using the default parameters. Input files were generated from pdb files of holo- and loaded-ArCP in which the phosphopantetheinylated serine was replaced with a standard serine using the pdb2pqr^{104,105} online server.

Structures were analyzed with PyMOL⁹⁷ and MOLMOL¹⁰⁶.

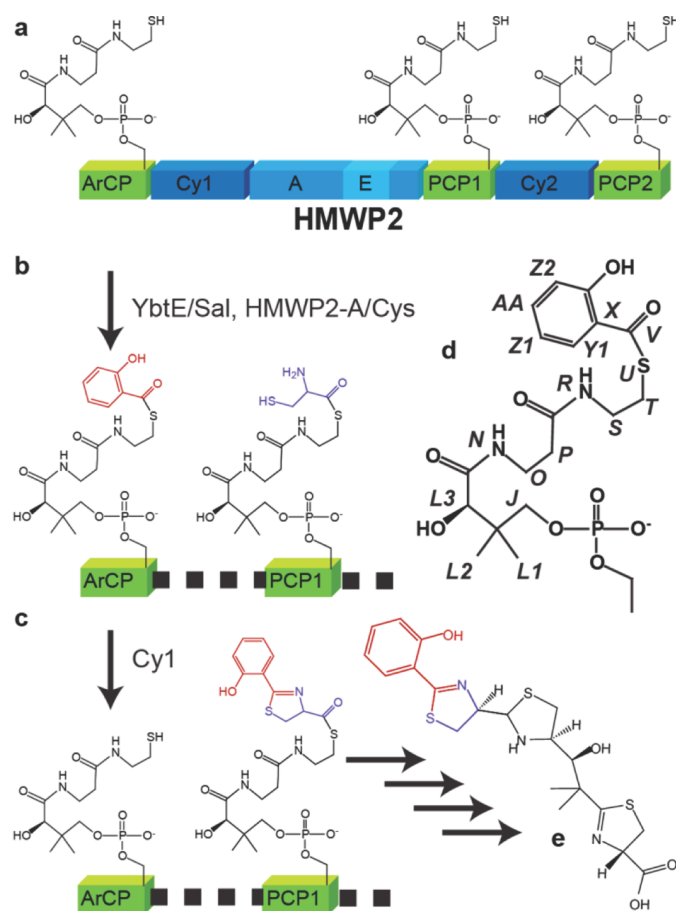


Figure 3-1. Role of ArCP in yersiniabactin synthesis. (A) ArCP is the first carrier protein domain HMWP2, which also comprises two cyclization domains (Cy1 and Cy2), two peptidyl carrier proteins (PCP1 and PCP2), an adenylation domain for cysteine (A), and an epimerization domain embedded in the A domain (E). (B) The A domain loads cysteine onto PCP1 and PCP2 and the stand alone A domain YbtE loads salicylate onto ArCP. (C) Cy1 catalyzes the condensation and cyclodehydration of Sal and Cys, forming 2-hydroxyphenylthiazoline on PCP1 and returning ArCP to its holo form. Not shown: Cy2, HMWP1 (a mixed PKS/NRPS protein), and YbtU complete the synthesis of yersiniabactin (E). (D) Nomenclature used to assigning phosphopantetheine and salicylate is shown in italics.

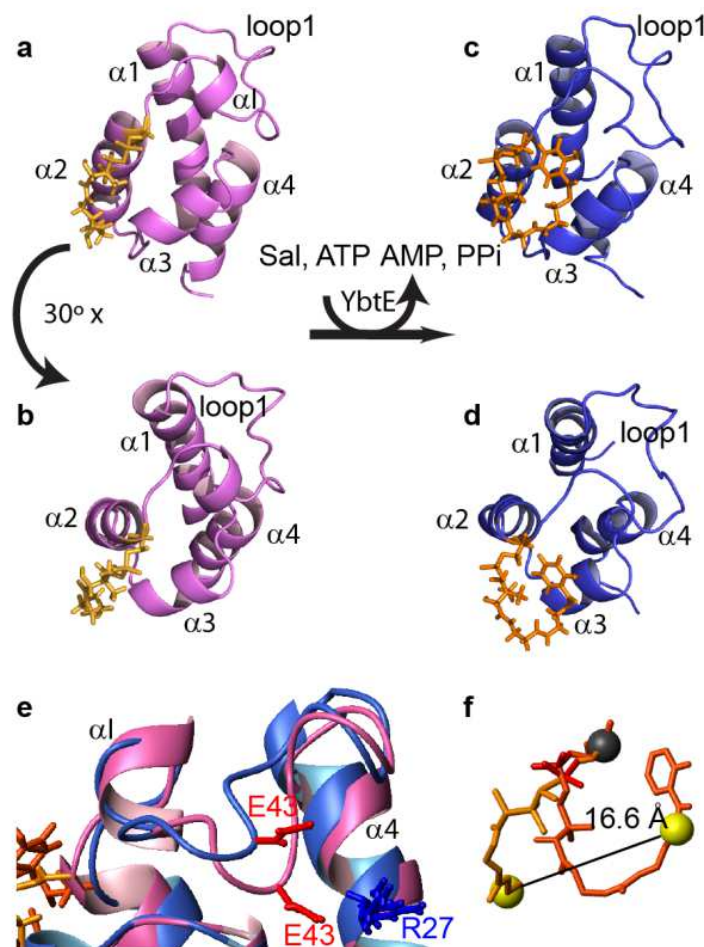


Figure 3-2. Solution structures of holo (a,b) and loaded (c,d) ArCP. The lowest energy conformer of the NMR ensemble is shown for each form of the protein under two different views. e) detail of loop1 shown for holo (pink) and loaded (blue) ArCP. Structures were aligned with each other using helices $\alpha 1$ through $\alpha 4$. f) mean structures of Ser-PP from holo-ArCP (orange) and Ser-PP-Sal from loaded-ArCP (orange-red). The moieties were translated to overlay Ser 52 C γ of each form.

NMR structure statistics for holo-ArCP	
Violations (mean and s.d.) ^a	
Distance constraints (Å)	0.32 +/- 0.07
Dihedral angle constraints (°)	1.88 +/- 0.41
Max. dihedral angle violation (°)	3.12
Max. distance constraint violation (Å)	0.43
R. m. s. deviations geometry ^c	
Bond lengths (Å)	0.017
Bond angles (°)	1.4
Average pairwise r.m.s.d. [residues 19–88] (Å) ^a	
Heavy	0.88 +/- 0.07
Backbone	0.37 +/- 0.06
Ramachandran Statistics ^b	
Most favoured	92.5
Additionally allowed	7.5
Generously allowed	0.0
Disallowed	0.0

Table 3-1. NMR structure statistics from holo-ArCP. a, from CYANA 2.1. b, from PROCHECK_NMR residues 17-90. c, From PSVS.

NMR structure statistics for loaded-ArCP	
Violations (mean and s.d.) ^a	
Distance constraints (Å)	0.30 +/- 0.07
Dihedral angle constraints (°)	2.81 +/- 0.27
Max. dihedral angle violation (°)	3.38
Max. distance constraint violation (Å)	0.47
R. m. s. deviations geometry ^c	
Bond lengths (Å)	0.017
Bond angles (°)	1.6
Average pairwise r.m.s.d. [residues 19–88] (Å) ^a	
Heavy	0.85 +/- 0.09
Backbone	0.36 +/- 0.08
Ramachandran Statistics ^b	
Most favoured	89.4
Additionally allowed	10.6
Generously allowed	0.0
Disallowed	0.0

Table 3-2. NMR structure statistics for loaded-ArCP. a, from CYANA 2.1. b, from PROCHECK_NMR residues 16-89. c, From PSVS.

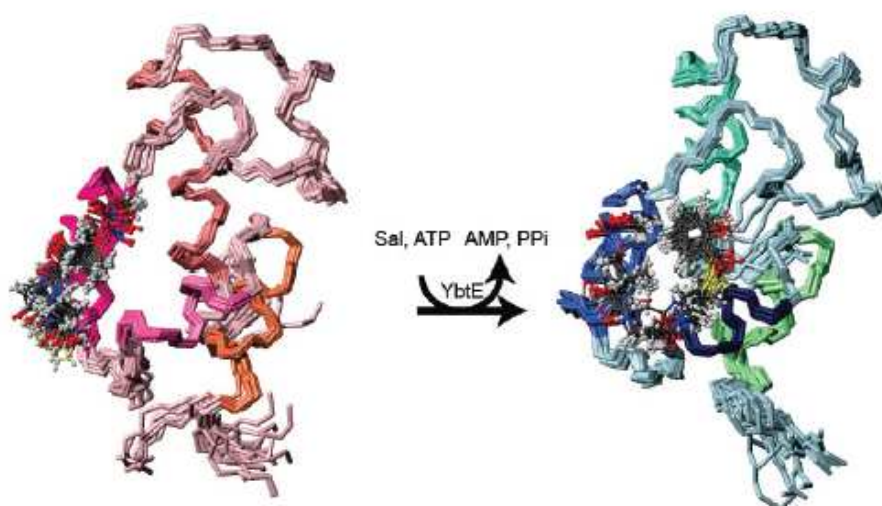


Figure 3-3. NMR bundles for holo-ArCP (pink) and loaded-ArCP (blue). In holo-ArCP, alpha helices are colored as follows: $\alpha 1$: brown, $\alpha 2$: magenta, $\alpha 3$: pink, $\alpha 4$: beige. The corresponding colors for loaded-ArCP are $\alpha 1$: cyan, $\alpha 2$: sky blue, $\alpha 3$: royal blue, $\alpha 4$: green. Alignments made with C $^{\alpha}$ of residues in helices only. The NMR structure of the holo-ArCP core is defined by 1636 distance constraints and the protein core of loaded-ArCP is defined by 1314 distance constraints. The residues masked by PP in holo-ArCP are S52, I53, M56, and L59 on $\alpha 2$ and L71, R72, and Y75 on loop 2 and $\alpha 3$. In loaded-ArCP, Sal and PP cover a surface involving residues L50-I53, M56, R72, Y75, A76, and I46.

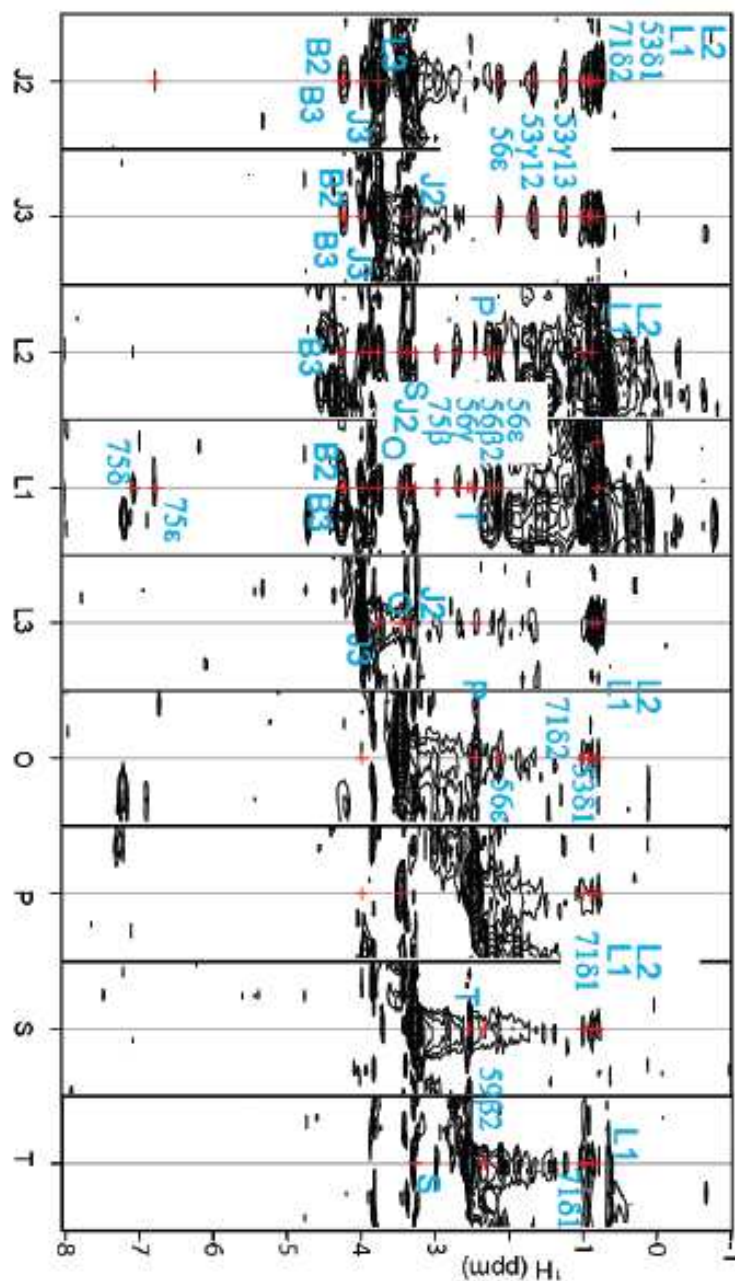


Figure 3-4. NOESY strips involving PP in holo-ArCP. The labels for PP moieties (Roman alphabet) are defined in Figure 3-1. Labls are positioned between two strips when the corresponding signals are observed in both. For corwded regions, the order of the labels from top to bottom refers to the nearest series of signals, again from top to bottom.

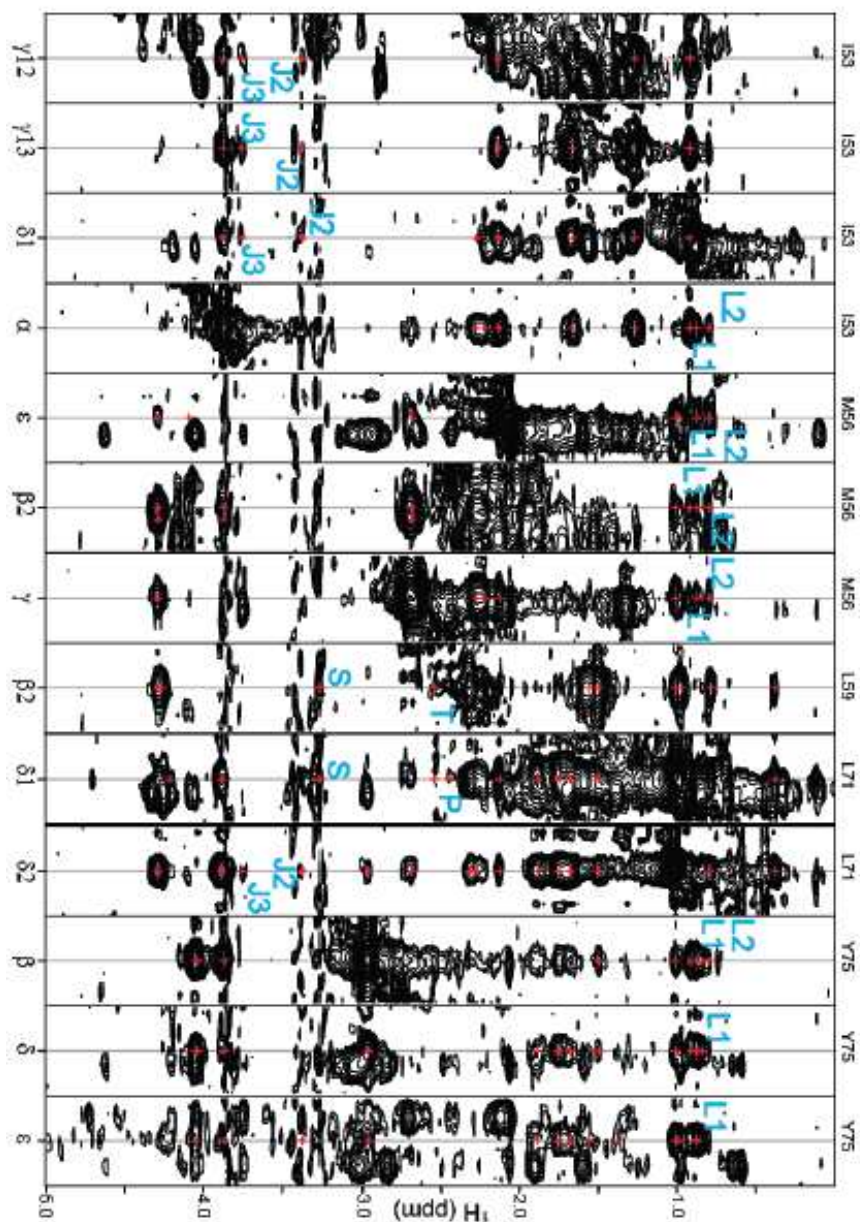


Figure 3-5. NOESY strips of residues in holo-ArCP displaying NOESY cross-peaks with PP. Unlabeled cross-peaks denote cross-peaks within the protein core. A total of 50 cross-peaks involving PP could be assigned unambiguously from peaks seen in Figures 3-7 and 3-8. These include 14 nOe's within the PP arm and 36 nOe's between the protein core and PP.

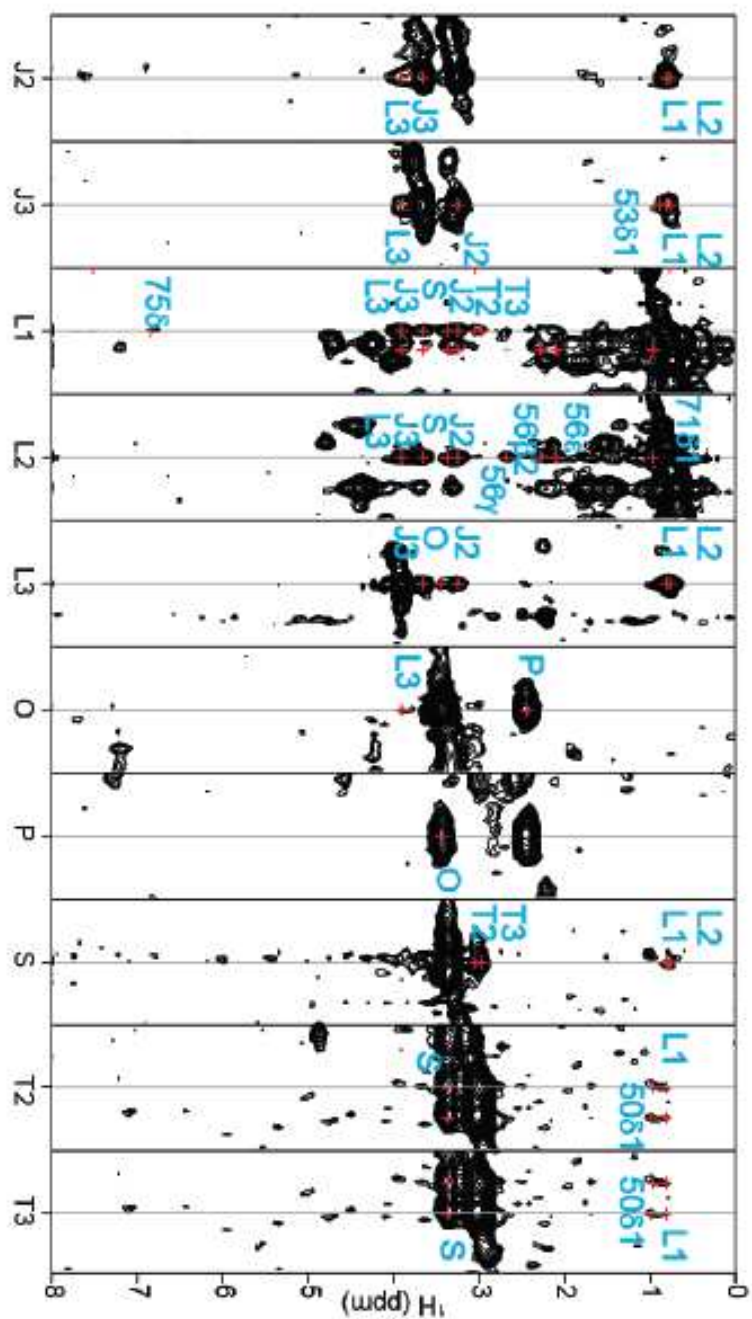


Figure 3-6. NOESY strips involving PP in loaded-ArCP. The labels for PP moieties are defined in figure 3-1. For crowded regions, the order of labels from top to bottom refers to the nearest series of signals, again from top to bottom.

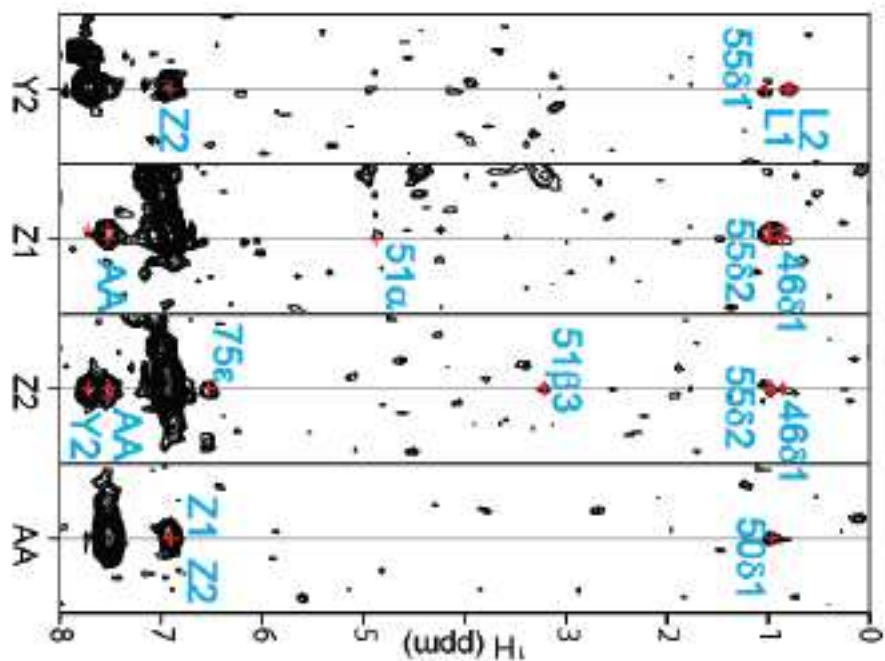


Figure 3-7. NOESY strips involving salicylate in loaded-ArCP. The labels for salicylate moieties are defined in figure 3-1. For crowded regions, the order of labels from top to bottom refers to the nearest series of signals, again from top to bottom.

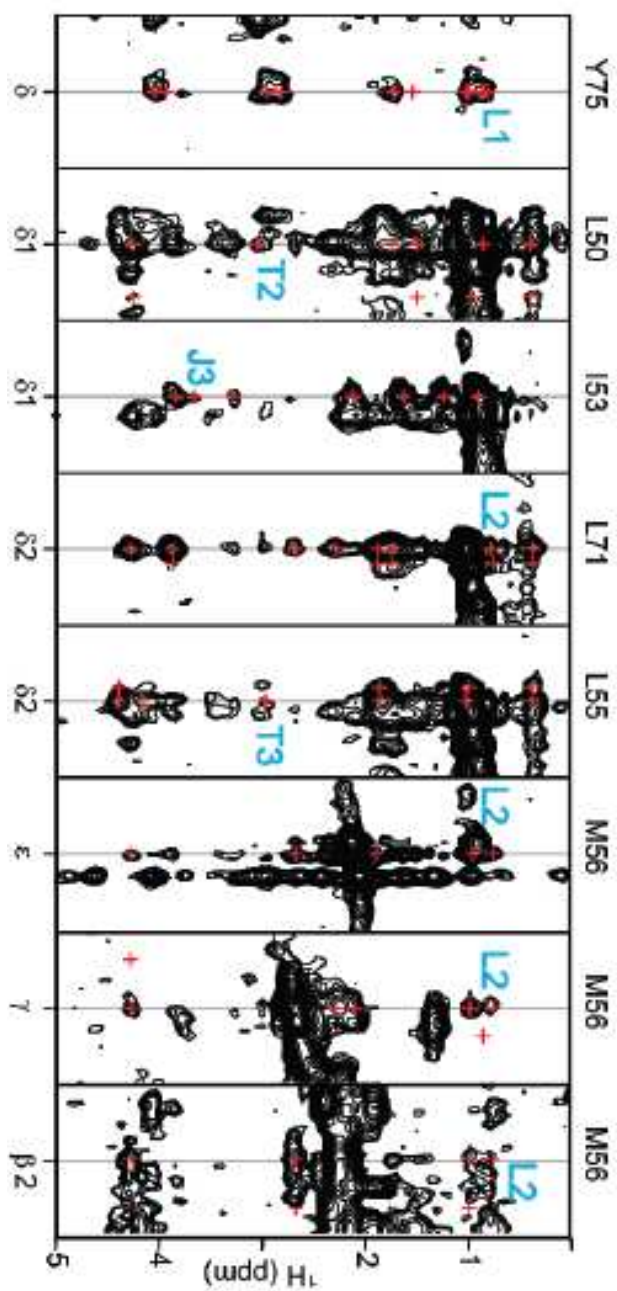


Figure 3-8. NOESY strips of residues in loaded-ArCP displaying NOESY cross-peaks with PP. Unlabeled cross-peaks denote cross-peaks within the protein core. 8 nOe's were observed between loaded-ArCP and PP.

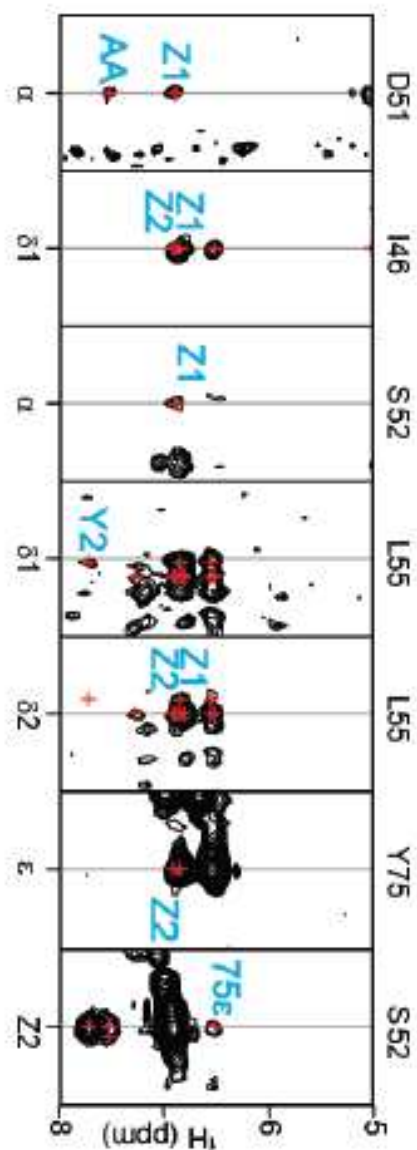


Figure 3-9. NOESY strips of residues in loaded-ArCP displaying NOESY cross-peaks with salicylate. Unlabeled cross-peaks denote cross-peaks within the protein core. 12 nOe's were observed between loaded-ArCP and salicylate. The strip of Z2 (in salicylate) is shown to confirm the assignment of a NOESY cross-peak from Y75 ϵ that overlaps with another cross-peak. 9 constraints could also be determined within the Sal-PP moiety.

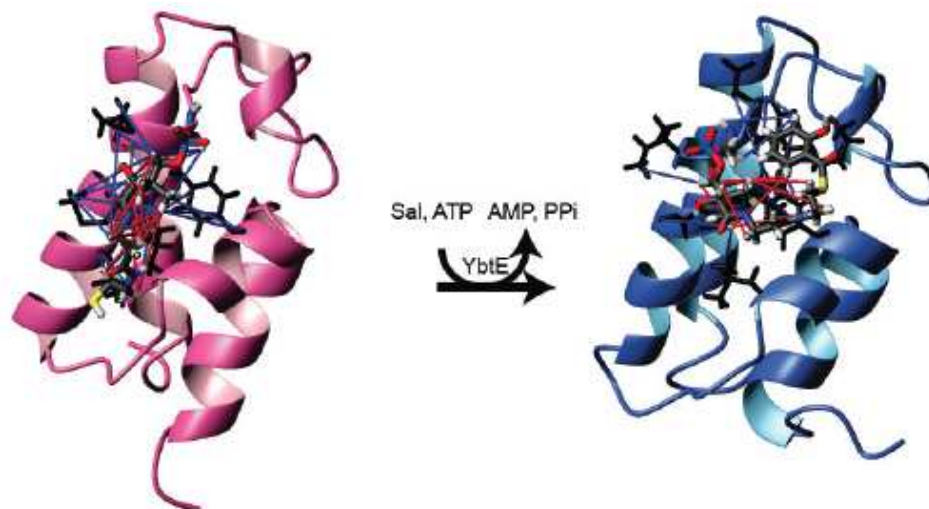


Figure 3-10. Visualization of distance constraints involving prosthetic groups in holo-ArCP (left) and loaded-ArCP (right). Lines in blue denote distance between the core protein and the prosthetic group, while lines in red denote distances within the prosthetic groups. Side chains of residues involved are show in black. Distance constraints are displayed on the lowest energy conformers.

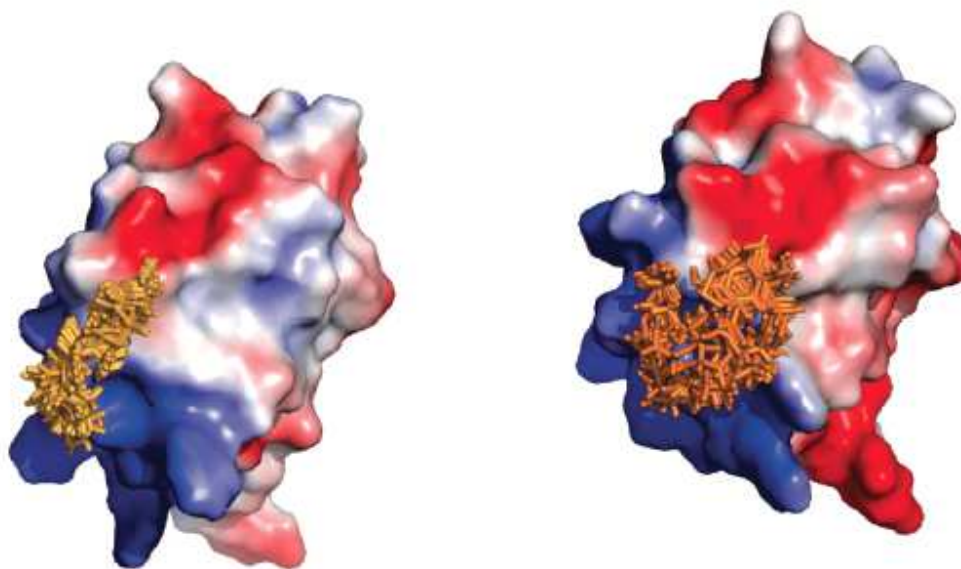


Figure 3-11. Change in conformation of prosthetic arm upon salicylate loading leads to different electrostatic surfaces in holo (left) and loaded (right) ArCP. Blue, positive charges; red, negative charges. Upon loaded, the change of the conformation in PP modifies access to both the positively charged and negatively charged surfaces. Surface potentials were generated using the APBS Tools2.1 plugin for PyMOL.

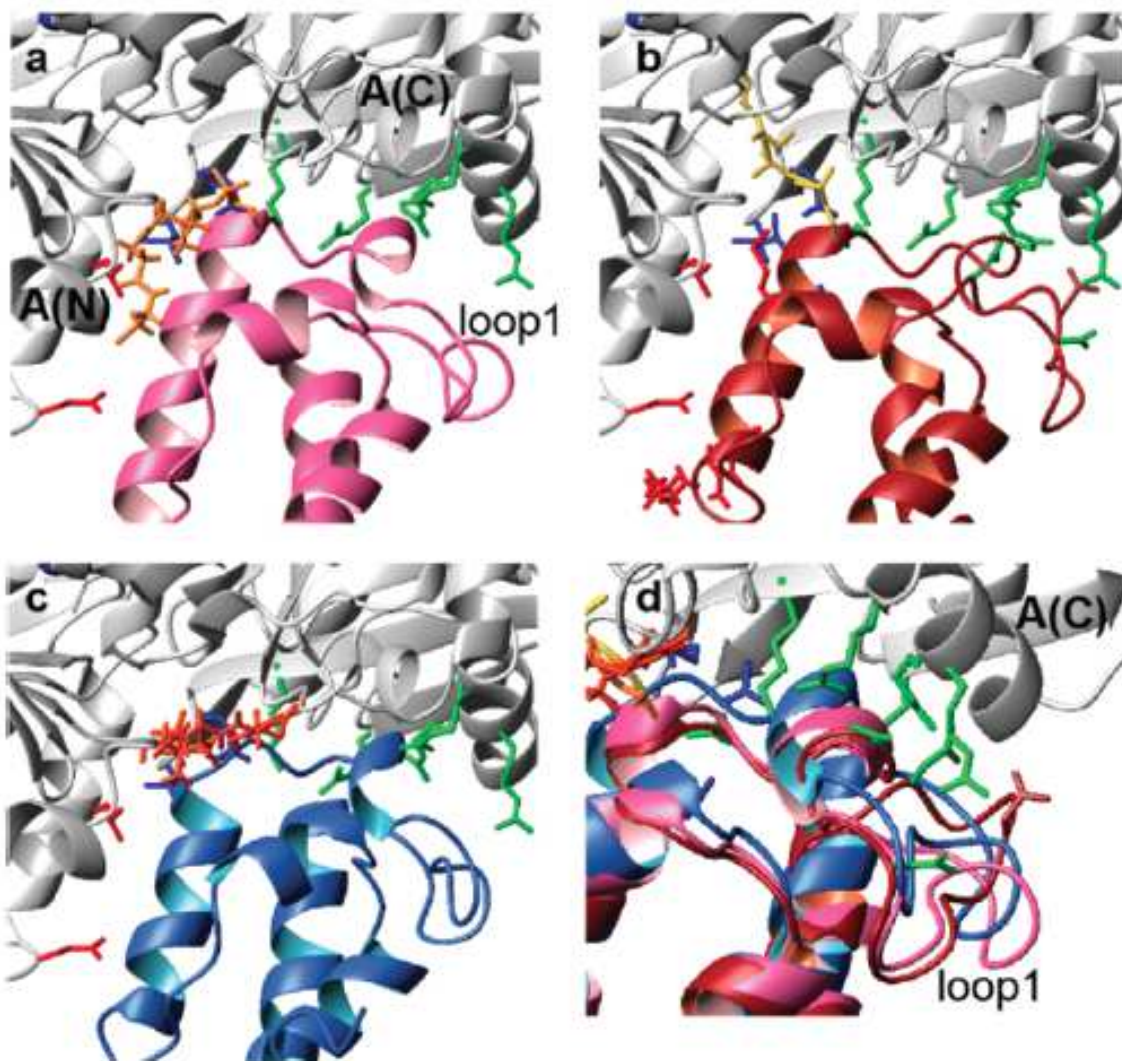


Figure 3-12. Comparison of holo- (a, in pink) and loaded-ArCP (c, in blue) in complex with an adenylation domain in A_T conformation (in white). The original structure of EntB-ArCP (brown) in complex with EntE is shown in b) (2ROG). In b), EntB-ArCP sidechains 2, those in blue show interactions between A(C) and ArCP, and those in green denote interactions between A(C) and loop1. The same color scheme was used for side-chains of EntE that are displayed in a-c). d) Detail showcasing changes in the conformation of loop1. The orientation in d is obtained by rotation of 30° around the Y vertical axis.

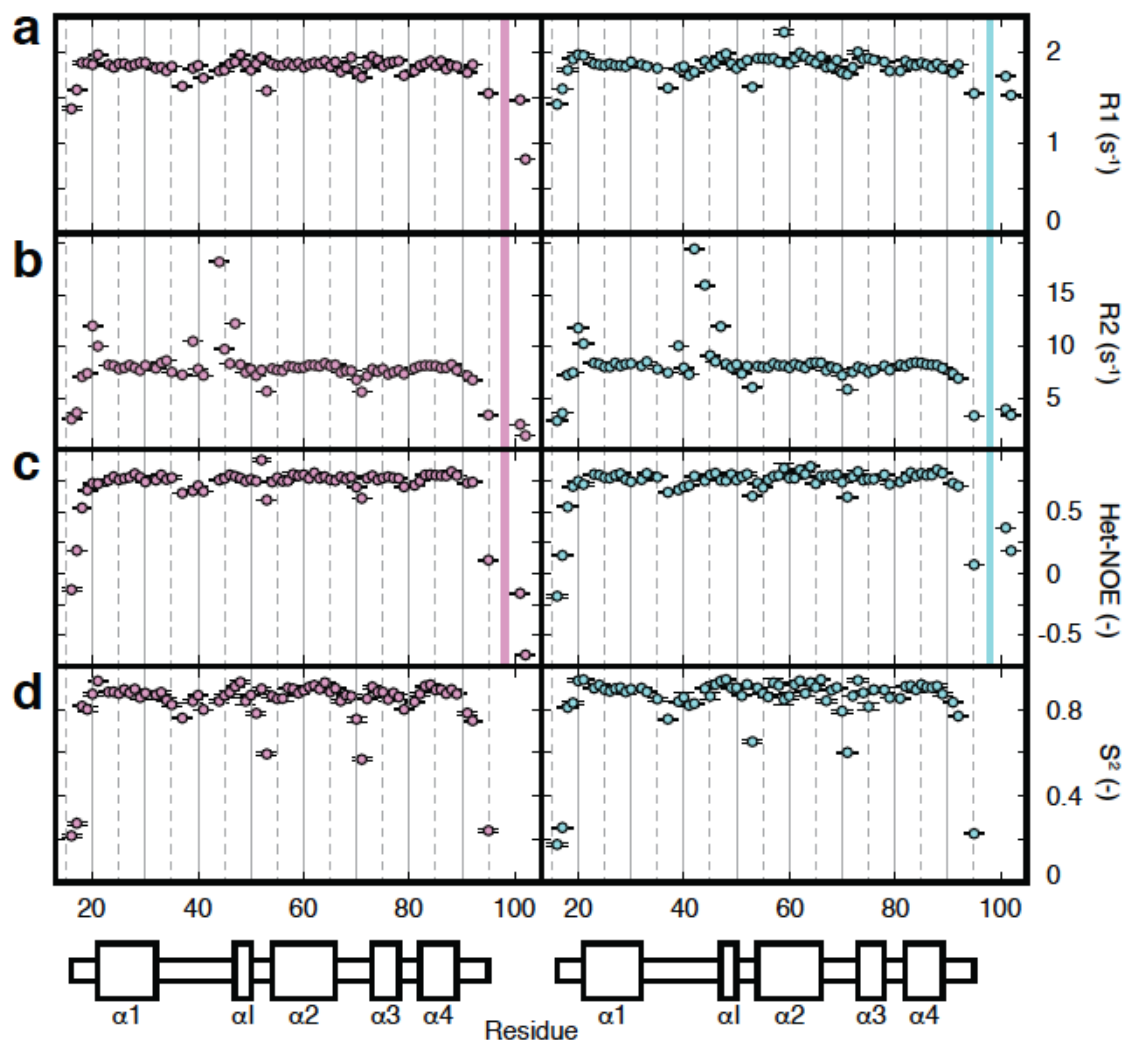


Figure 3-13. NMR dynamics of holo and loaded-ArCP. (a-d) Residue-specific NMR relaxation parameters for holo (left, magenta) and loaded (right, cyan). The secondary structure of ArCP is illustrated below the plots (a) R1 relaxation rates (b) R2 relaxation rates (c) Heteronuclear NOE parameterized by $I_{\text{sat}}/I_{\text{ref}}$ where I_{sat} and I_{ref} are the amplitudes of signals in the proton-saturated and reference experiment respectively. (d) Order parameter S^2 .

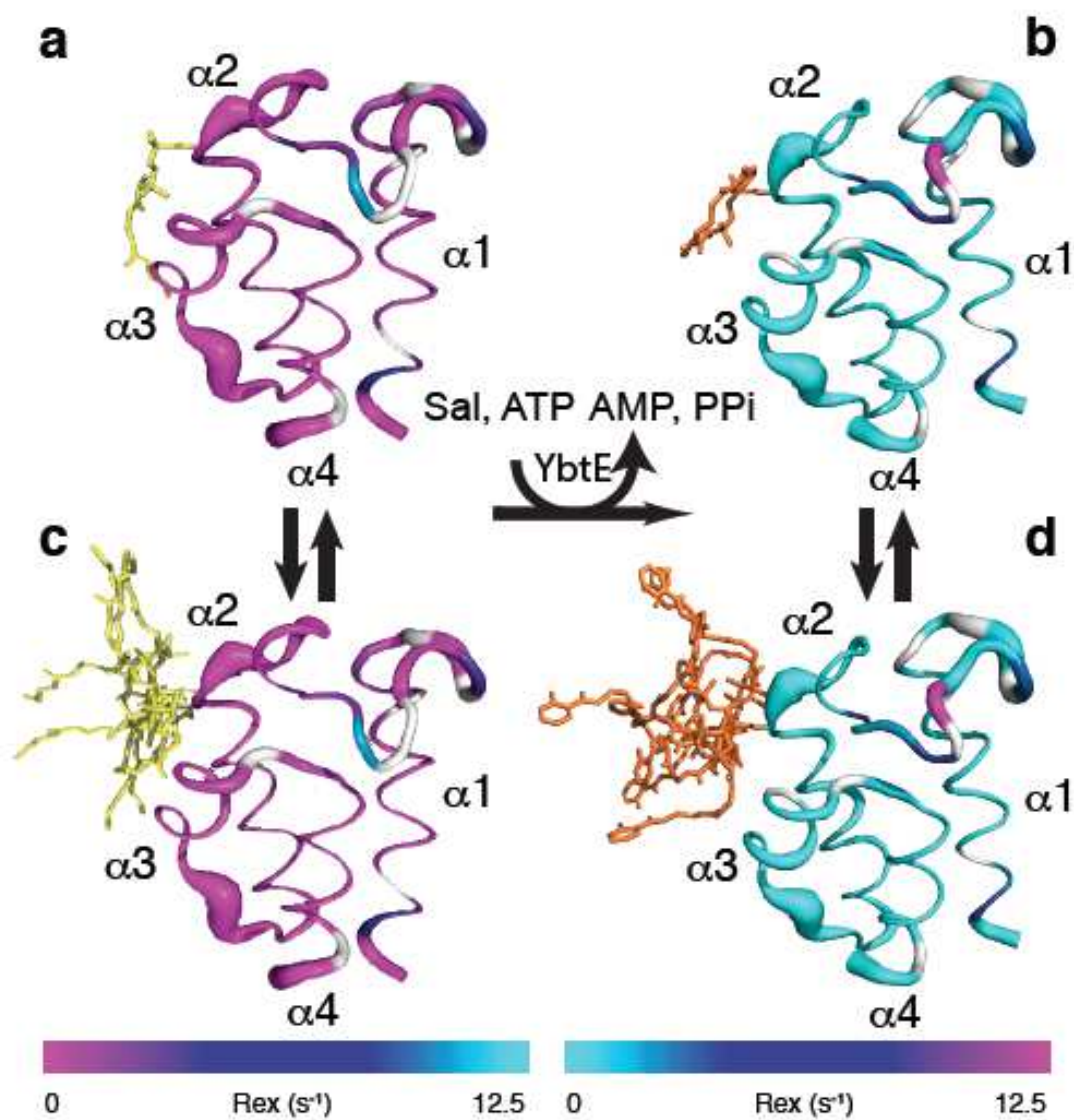


Figure 3-14. Dynamics visualization of holo (magenta) and loaded (cyan) ArCP, residues 18-93. A thicker ribbon corresponds to a reduced order parameter and increased flexibility. Colors represent the R_{ex} parameters fit during Lipari-Szabo analysis. Data are not available for residues in white due to overlap (e) holo-ArCP with the PP arm in its bound state (f) loaded-ArCP with the PP arm in its bound state (g) holo-ArCP with the PP arm in its unbound state (h) loaded-ArCP with the PP arm in its unbound state.

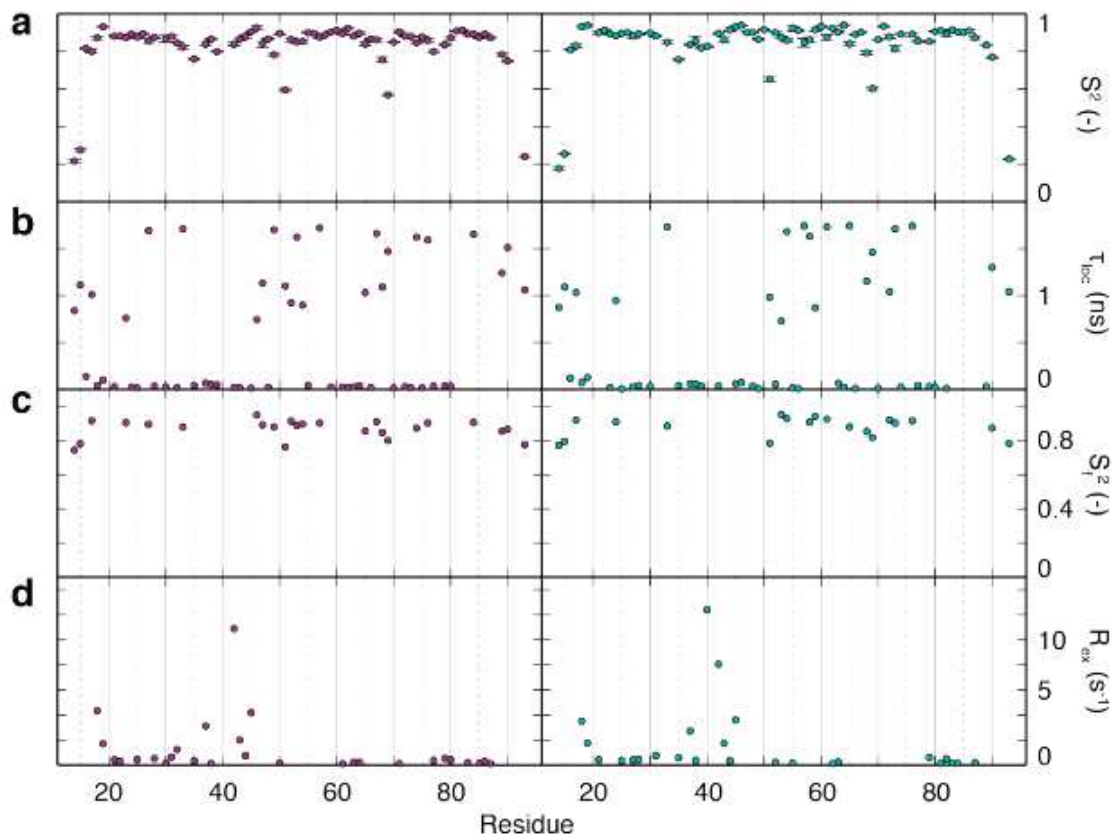


Figure 3-15. Model-free analysis of holo (left) and loaded (right) ArCP. The analysis was performed using the software ROTDIF. ROTDIF selects which parameters to use when fitting based on the Akaike Information Criterion. For a given residue, parameters not selected have not been plotted. (A) Overall (S^2) or slow time-scale (S_s^2 , when S_f^2 is present) order parameters (B) Effective correlation time (C) Fast time scale order parameters (S_f^2) (D) Exchange contribution to transverse relaxation (R_{ex}).

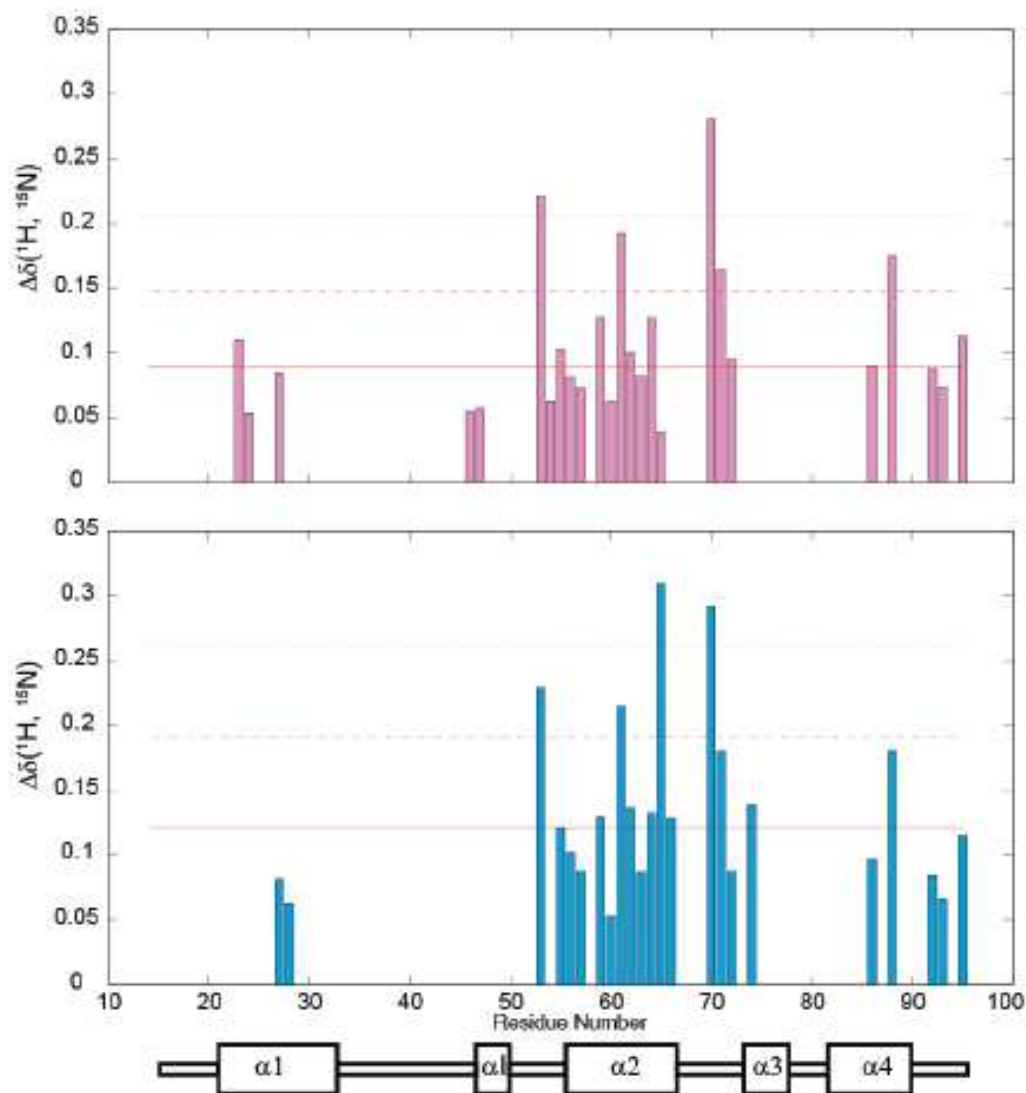


Figure 3-16. Chemical shift perturbations between major and minor forms in holo (top) and loaded (bottom) ArCP. The solid horizontal lines indicate the median (0.09 for holo, 0.12 for loaded) and the median plus one standard deviation (std = 0.06 for holo, std = 0.07 for loaded), respectively. The reported chemical shift differences are calculated with $\Delta\delta(^1\text{H}, ^{15}\text{N}) = ((\Delta\delta\text{H})^2 + (1/5)(\Delta\delta\text{N})^2)^{1/2}$, where $\Delta\delta\text{I}$ is the chemical shift difference between the two species for nucleus i.

Chapter 4-The solution structure of apo-ArCP reveals pre-formed binding sites for the phosphopantetheine arm and tethered substrate

ABSTRACT: Carrier proteins (CPs) play a central role in nonribosomal peptide (NRP) synthesis. CPs are first modified from an inactive (apo) to active (holo) state via attachment of a phosphopantetheine (PP) arm. Substrate monomers are then directly attached to the PP arm via formation of a thioester to generate the loaded form. CPs interact with many different catalytic domains during synthesis and understanding how each of these modifications influences the structure of a CP will provide insight into the mechanism of NRP synthesis. In Chapter 3, I described the structures of holo- and substrate-loaded ArCP. Here, I describe the structure of apo-ArCP and thereby provide a structural view of ArCP as it exists in all of its possible forms. Comparison with the holo structure shows how attachment of the PP arm alters the conformation of helix 3 and examination of solvent-exposed hydrophobic and aromatic residues shows that binding sites for both the PP arm and loaded salicylate are pre-formed in the apo structure.

In Chapter 2, I described the development of a novel method that enabled our lab to study the salicylate-loaded form of ArCP using NMR and in Chapter 3 I described the solution structures and dynamics of holo- and loaded-ArCP. The combined results from those studies offered us novel insight into how substrate loading influences the structure and dynamics of the CP to which it is tethered and allowed us to propose a means by which rearrangements of the PP moiety influence interactions with an adenylation domain. However, for a complete description of the influence of post-translational modifications on the structure of ArCP and their potential influence on modulating interactions with catalytic domains, the solution structure of ArCP in its unmodified (apo) form is needed. Comparison of the apo and holo forms will allow us to dissect the role of the PP arm in influencing the structure of ArCP and, in turn, how this influences interactions with catalytic domains.

In this chapter, I describe the solution structure of ArCP in its apo form. Comparing the structure of apo-ArCP with holo-ArCP shows how hydrophobic interactions between the geminal methyl groups of the PP arm and aliphatic residues on ArCP lead to structural rearrangements that reposition regions of ArCP known to interact with catalytic domains. An examination of solvent-exposed hydrophobic residues shows that the binding sites for both the PP arm and tethered salicylate are pre-formed in the apo form and that this seems to be a general feature in carrier proteins. Finally, I compare the position of loop 1 between the apo, holo, and loaded forms and discuss the implications for modulating interactions with YbtE and PPTases.

Results and Discussion

The solution structure for apo-ArCP was solved using CYANA 2.1⁹⁹ with 1128 distance constraints (14.1 per residue) obtained from a HN-NOESY-HSQC in H₂O and a HC-HSQC-NOESY in D₂O along with 141 dihedral angle constraints from TALOS+²⁷ based on backbone chemical shifts. The 20 structures with the lowest target function were taken out of 100 calculated structures. This bundle had an average target function of 1.98 ± 0.13 Å with no distance constraint violations greater than 0.5 Å and no dihedral angle constraint violations greater than 5°. The structural bundle has a backbone RMSD of 0.38 ± 0.10 Å and a heavy atom RMSD of 0.94 ± 0.10 Å, demonstrating that the structures calculated are both consistent with the constraints used to generate them and well-defined.

The structures of apo-ArCP show the well-known four-helix bundle typical of carrier proteins¹⁰⁷ (Figure 4-1). The bundle shows three longer helices, helix 1 (A18-L32), helix 2 (S52-L65), and helix 4 (L80-L88) that are mostly parallel to one another and run in an up-down-down fashion with a fourth, shorter helix, helix 3 (L71-L76), lying at a large angle to the others (Figure 4-2). S52, at the N-terminus of helix 2, is the phosphopantetheinylation site. Helices 1 and 2 are connected by a long loop (loop 1) while the other helices are connected by shorter loops (2 and 3). Finally, within loop 1 lies an additional single-turn helix, α I, which is also found in the holo and salicylate-loaded structures⁶.

Comparison of the apo structure with that of holo- and loaded-ArCP shows that the overall structure is conserved between all three forms (Figure 4-3) and that the post-translational modifications only change the conformation of loop 1 and cause subtle rearrangement of the helices with respect to one another. The most notable of these is the repositioning of helix 3 with respect to helix 2 in the holo form. Phosphopantetheinylation of apo-ArCP

causes helix 3 to move slightly towards the core of ArCP, as can be seen in Figure 4-4A. This is most likely driven by hydrophobic interactions between the geminal methyl groups at the base of the PP arm and aliphatic or aromatic residues found on helix 3, specifically L71 and Y75. Indeed, the methyl groups of the PP arm show nOe's to the methyl groups of L71 and the α , β , and ring protons of Y75. Upon loading, salicylate docks on the surface of ArCP in between helices 2 and 3, increasing the space between the helices and returning helix 3 close to its position in the apo form (Figure 4-4B). In the loaded form the methyl groups of L71 seem poised to interact with the methyls of M56 and L59, but not with the geminal methyls of the PP arm. In both the holo and loaded forms of ArCP, the loop connecting helices 2 and 3 was found to be flexible. This flexibility allows the repositioning of helices 2 and 3 relative to each other as ArCP is converted among its various forms. This subtle repositioning likely contributes to the modulation of interactions with the A domain YbtE as helices 2 and 3 likely form the interaction surface for YbtE (Chapter 5).

Another notable difference between the structure of apo-ArCP and those of holo- and loaded-ArCP is the length of helix 1 (Figure 4-3). In apo-ArCP, helix 1 begins at residue A18 (as recognized by PyMOL), while in the other forms this helix begins at R16. The results of the dynamics measurements shown in Chapter 3 indicate that residues near the end of helix 1 (A18-D20) are dynamic on a timescale slower than molecular tumbling (μ s-ms) and we proposed that this may be due to transient unfolding at the N-terminus of this helix⁶. In apo-ArCP, fewer interresidue nOe's were observed for residues H17 and A18 than in holo- and loaded-ArCP, consistent with this region being dynamic and largely disordered.

As in holo- and loaded-ArCP, the four-helix bundle is held together by hydrophobic interactions contributed by the large aromatic (F, W, and Y) and aliphatic (I, L, M, and V) residues of the four amphipathic helices (Figure 4-5A). Further, hydrophobic interactions between L34, L39, L45, and L50 on loop 1 and residues within the core of the protein anchor loop 1 to the surface and define its position relative to the four helices. Figure 4-5B shows that apo-ArCP has a predominantly hydrophilic surface surrounding a tightly packed, hydrophobic core.

Closer inspection of the apo structure, however, shows that not all aliphatic or aromatic residues in apo ArCP are buried within the core. Figure 4-6 shows that residues I46 from loop 1 (on α I), I53, M56, and L59 from helix 2, and L71 and Y75 from helix 3 are all surface exposed in the calculated structures. Indeed, while the methyl or aromatic moieties of these residues all show strong nOe's to other residues within this group, they show few or no nOe's to other aliphatic residues, even immediate neighbors. For example, the methyl groups of I46 show strong nOe's to Y75 but only one very weak nOe to the adjacent and solvent protected L45. These seven residues form an isolated and self-consistent hydrophobic patch on the surface of ArCP.

Examination of the literature and comparison of the apo structure with that of holo- and loaded-ArCP suggest a role for each of the residues in this cluster. I53 (i+1 with respect to the PP site, S52) and M56 (i+4) are known to be important for interaction with phosphopantetheinyl transferases (PPTase), based on a co-crystal structure of a carrier protein and Sfp⁶⁸. M56, L59 (i+7), and L71 (i+19) are all found to be covered by the PP arm in the structure of holo-ArCP (Figure 4-7A)⁶. The position at i+4 is thus important for mediating interactions with the PPTase and the PP arm. Finally, I46 (i-6) and Y75 (i+23)

both have a direct interaction with the tethered salicylate in the structure of loaded-ArCP (Figure 4-7B)⁶. This demonstrates that the binding sites for the PPTase, PP arm itself, and tethered substrate are all pre-formed in the apo structure. This is in contrast to earlier reports that a binding site for the tethered substrate was not apparent in the first apo peptidyl carrier protein structure solved²; however, the authors of that study did not have a structure of a loaded NRPS CP at their disposal and may have been looking for a deep binding pocket instead of a shallow surface.

The observation that solvent-exposed hydrophobic residues contribute to ArCP's function suggests that they may play a key role in the function of all CPs. Indeed, residues at these positions are found to interact with the PP arm or tethered substrate in other holo- and loaded-CP structures¹⁰⁸. Examining the structure of other apo NRPS CPs (those of EntB from enterobactin synthetase⁵⁴, TycC3 from tyrocidine synthetase^{2,5}, PCP7 from teicoplanin synthetase⁴, and PCP1 from yersiniabactin synthetase (Bradley Harden, personal communication)) shows that the position of these solvent-exposed residues corresponds to similar positions in ArCP or rationalizes alternative binding sites for the PP arm. This conserved feature suggests that the interaction between the PP arm and CP is driven at least in part by hydrophobic interactions between the methylene moieties of the PP arm and these hydrophobic and aromatic residues.

In Figures 4-8 through 4-11 used in the discussion that follows, all residues not colored in cyan are solvent-exposed hydrophobic or aromatic residues.

In all structures, the residue at $i+1$ is hydrophobic and is solvent-exposed. This is to be expected as this residue mediates the interaction with the PPTase⁶⁸ and apo-CPs are inactive until they are phosphopantetheinylated.

EntB, which is also an aryl carrier protein, shows a similar pattern of solvent-exposed residues to ArCP⁵⁴. The residues at positions i-6 (Ile), i+4 (Met), and i+19 (Phe) are all hydrophobic and solvent-exposed (Figure 4-8). Residues i+7 and i+8 are both alanines and may substitute together for the leucine found in ArCP at i+7 (Figure 4-8). Position i+23 is also occupied by an alanine (Figure 4-8). The substitution of the smaller alanine for a tyrosine may compensate for the extra hydroxyl group found in EntB's tethered substrate, 2,3-dihydroxybenzoate (DHB), compared with ArCP's substrate, salicylic acid. Thus, binding sites for both the PP arm and DHB seem to be present in EntB as well.

The apo and holo structures of TycC3 were originally solved in 2000 and at that time were proposed to be identical, based on very similar patterns of nOe's found for the two forms². However, it was also noted that small peaks were present in HN-HSQC's of both apo- and holo-TycC3, suggesting that both forms interconvert between a major and minor species. In 2006, solution structures of the major and minor states were solved⁵. This work suggested that the apo and holo forms share a common major state (A/H) but the apo form interconverts between the A/H state and an additional state (A) that is distinct from that of the holo (H). Notably, the position of the PP arm was proposed to change between the A/H and H state of holo-TycC3. In the A/H state, the PP arm interacted with residues near the beginning of loop 1, while in the H state it was found it interact with residues along helices 2 and 3. In the original apo structure solved in 2000, aliphatic residues at i+4 (Met) and i+20 (Leu) are solvent-exposed, as in ArCP (Figure 4-9A). As in EntB, residues i+7 and i+8 are both alanines (Figure 4-9A). This shows that a pre-formed binding site for the PP arm is present in apo-TycC3, consistent with that found in ArCP. The residues at i-6 (Phe) and i+24 (also Phe) are also solvent-exposed aromatics, suggesting that a binding site for

the tethered substrate (tyrosine) is also present (Figure 4-9A). Further, the position at i-19, the residue proposed to interact with the PP arm in the A/H state, is a solvent-exposed valine (Figure 4-9B). This suggests that this valine may be part of an additional binding site. This additional hydrophobic patch is also seen in other CPs (discussed below) and suggests that multiple PP binding sites can be present on a single CP.

The apo structure of PCP7_{Teic} from the teicoplanin synthetase seems to show the same two PP binding sites as those proposed for TycC3^{4,5}. Similar to ArCP, i+4 (Met), i+7 (Ile), and i+20 (Ile) form a contiguous hydrophobic patch along helices 2 and 3 (Figure 4-10A). PCP7_{Teic} also shows a hydrophobic patch spanning the C-terminus of helix 1 and the N-terminus of helix 2 and comprised of residues at positions i+2 (Leu), i-18 (Leu), and i-19 (Ile) (Figure 4-10B). The location of these residues is consistent with the second PP binding site found in TycC3. The authors of this study also solved the structure of holo-PCP7_{Teic}, but were unable to identify any nOe's between the core of the protein and the PP arm. Additionally, experiments with a paramagnetically-labeled PP arm were also inconclusive. The authors claimed that relaxation increased solely as a function of distance away from the PP attachment site, as would be expected for a completely unconstrained PP arm. However, their data appears to show enhanced relaxation at the N-terminus of helix 3, near residue i+20, suggesting that the PP arm does bind in a manner close to that seen in ArCP. Further, it is possible that the lack of nOe's and the seemingly inconclusive paramagnetic relaxation enhancement data could be explained by rapid interconversion between an unbound state and *two* bound states, in which the PP arm interacts specifically yet transiently with the two sites described above.

Reinspection of apo-ArCP shows that the residue at i-20 is also a solvent-exposed leucine, showing that this alternative binding site also seems to be present in ArCP. While we did not find any evidence of an interaction between this region of holo-ArCP and the PP arm in our NOESY data, some signals from the PP arm show peak doubling, indicating the PP arm can adopt multiple different conformations. It is possible that the PP arm is binding in this region, but the population is too low for the interaction to be observed.

The position of solvent-exposed aliphatic and aromatic residues in apo-ArCP, TycC3, and PCP7_{Teic} could be used to rationalize varying observations about the placement (or lack thereof) of the PP arm in the various existing structures. It may therefore be possible to predict binding sites for the PP arm based on these solvent-exposed hydrophobic patches. In the structure of apo-PCP1, for which a holo structure is not yet available, residues at i-7 (Phe) and i+23 (Phe) are both solvent-exposed and may form a binding site for loaded cysteine (Figure 4-11A). As shown in the structure of loaded-ArCP (Chapter 3), the curled conformation that the PP arm adopts in the loaded form positions the tethered substrate near the PP phosphate and the conserved aspartic acid at i-1. These negative charges may interact with the positively charged amine of the loaded amino acid and this charge-charge interaction may be a conserved feature of CPs loaded with amino acids. The residue at position i+19 (Leu) seems poised to form a binding site for the PP arm; however, the residues at positions i+4 and i+7 are both threonine, and the PP binding site identified in ArCP is decidedly more hydrophilic in PCP1 due to these substitutions (Figure 4-11A). Instead, the residues at i+2 (Leu), i-18 (Leu), and i-19 (Leu) seem to form an alternative binding site for the PP arm (Figure 4-11B), one that is consistent with that found for the A/H state from TycC3 and the second potential site from PCP7_{Teic}. The holo structure of

PCP1 has not been solved yet and it will be interesting to see how the actual structure compares with the prediction made here.

In Chapter 3, when comparing the structures of holo- and loaded-ArCP, it was shown that the position of a region of loop 1, namely that containing residues H40-S43 and centered on E41, was both subject to structural fluctuations on a μ s time-scale and changed position upon substrate loading. We attributed that change in position to a change in conformational equilibrium caused by substrate loading, wherein direct interactions between I46 of loop 1 and the tethered salicylate stabilized one conformation of loop 1 and therefore altered the observed position of E41 relative to that in the holo form⁶. Figure 4-12 shows an overlay of the position of E41 in apo-, holo-, and loaded-ArCP and shows that loop 1 in holo-ArCP adopts a conformation in between that seen in apo- and loaded-ArCP. One means of rationalizing this observation is that the position of loop 1 in the three forms lies along a trajectory between two states, one of which (state A) most resembles the apo form and the other of which (state B) most resembles the loaded form. The observed position of E41 in a given structure then reflects the relative populations in a conformational equilibrium. Comparison of the structures indeed shows a shift from state A towards state B as ArCP is modified from the apo form to the holo form and finally to the loaded form.

Conclusion

In this chapter, I presented the solution structures of apo-ArCP. Comparisons of apo-ArCP with holo- and loaded-ArCP show that covalent modifications cause only modest changes to the overall structure. These comparisons show that the binding sites for the PP arm in holo-ArCP and tethered salicylate in loaded-ArCP are already present in the apo structure. Comparison with structures of apo-CPs from other NRPS systems seems to show that pre-

formed PP and tethered substrate binding sites seem to be a general feature of apo-CPs from NRPSs. In the following chapter, I will discuss how the structures of apo-, holo, and loaded-ArCP, together with binding studies, inform our understanding of protein-protein interactions in the yersiniabactin synthetase system.

Methods

Details about HNCA, HNCOC, HNCACO, HNCOC, and NOESY-HN-HSQC experiments can be found in the Methods section of Chapter 2. For assignment of aromatic residues, hbCBcgcdHD (288 scans, 2048 (^1H , 16.019 ppm at 4.696 ppm) \times 26 (^{13}C , 22 ppm at 30 ppm) complex points, 4 hrs 55 mins), hbCBcgcdceHE (672 scans, 2048 (^1H , 16.019 ppm at 4.696 ppm) \times 26 (^{13}C , 22 ppm at 30 ppm) complex points,), and HC-HSQC-NOESY (16 scans, 2048 (^1H , 16.019 ppm at 4.696 ppm) \times 100 (^1H , 13 ppm at 4.696 ppm) \times 50 (^{13}C , 50.0003 ppm at 69.5 ppm) complex points, mixing time of 90 ms, 4 days 15 hrs 54 mins) were run on a sample in 100% D_2O .

Structure Calculation

Assignment of NOESY cross-peaks was performed manually using CARA. 1128 unambiguous restraints were assigned for apo-ArCP. In addition, 141 angle constraints were obtained with TALOS-+. Structure calculations were performed using CYANA version 2.1. For the final structure calculation, 100 structures were calculated using 50,000 steps. The final CYANA target functions was 1.98. There were no distance violations bigger than 0.5 Å and no angle violations larger than 5° in either NMR ensemble. The average rmsd to mean for these conformers were 0.38 (backbone) and 0.98 Å (heavy). The 20 structures with the lowest target function were chosen for water refinement in explicit solvent using CNS and refinement run using RECOORD scripts.

Structures were analyzed with PYMOL.

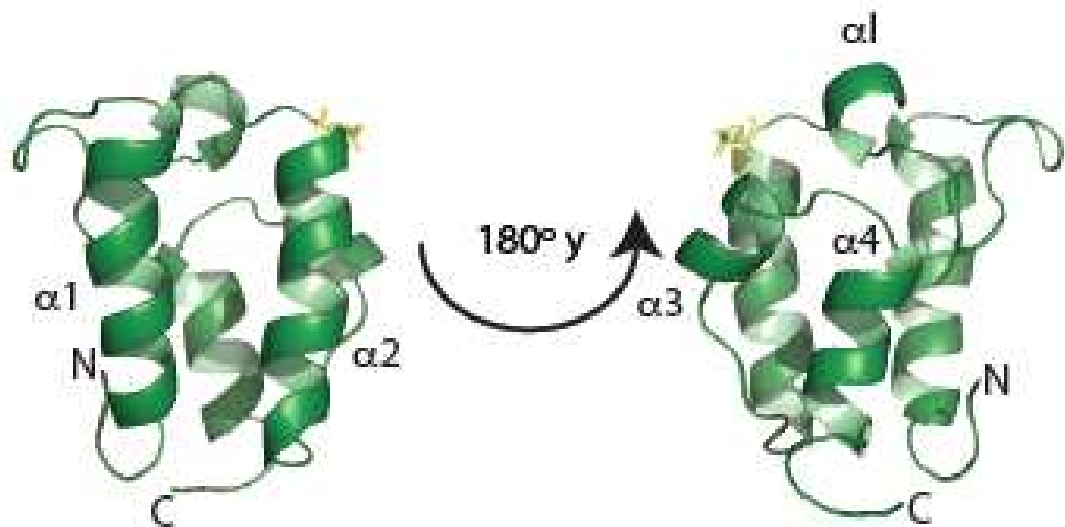


Figure 4-1. Cartoon representation of the structure of apo-ArCP. This structure shows the canonical four-helix bundle typical of carrier proteins, with three longer helices ($\alpha1$, $\alpha2$, and $\alpha4$) running parallel to one another in an up, down, down configuration and a fourth, shorter helix ($\alpha3$) lying at a large angle to the other three. A single-turn helix (αI) is found within loop 1 connecting helices 2 and 3. The conserved serine is shown in yellow sticks at the N-terminus of $\alpha2$.

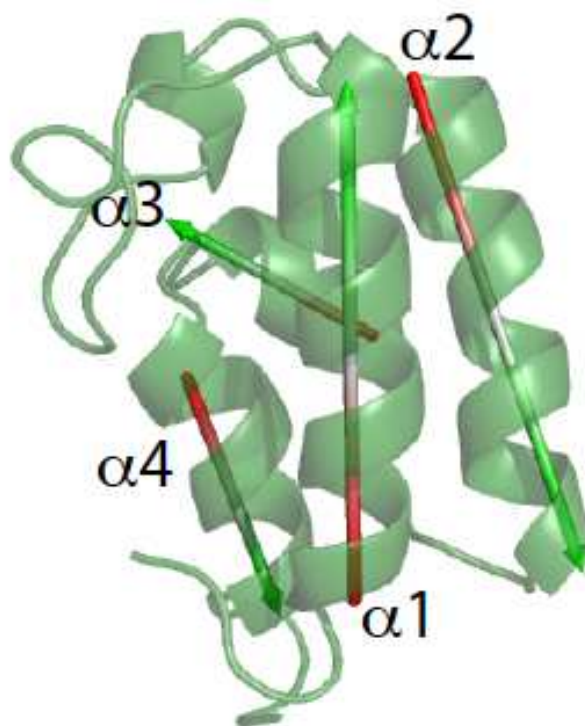


Figure 4-2. Cartoon representation highlighting the topology and relative orientation of the four major helices in apo-ArCP. $\alpha1$, $\alpha2$, and $\alpha4$ are nearly parallel to each other while $\alpha3$ lies nearly perpendicular to the rest.

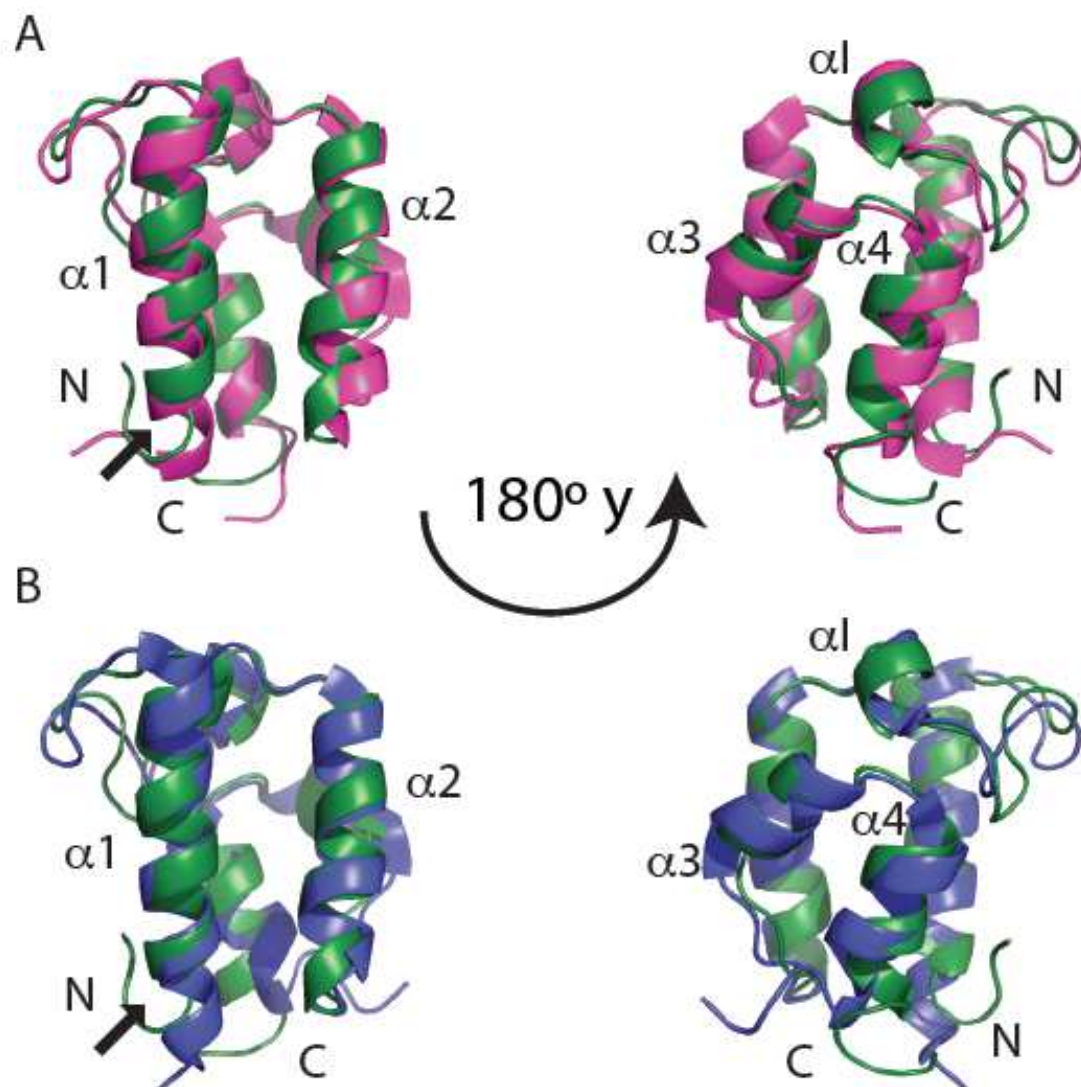


Figure 4-3. Comparison of the structures of apo-ArCP (green) with A) holo-ArCP (magenta, 2N6Y) and B) loaded-ArCP (blue, 2N6Z). Overall, there are no major structural differences between apo-ArCP and either holo- or loaded-ArCP and most differences result in minor alterations in the relative positioning of the helices. The black arrow at the N-terminus indicates where helix 1 is shortened by a half-turn in apo-ArCP.

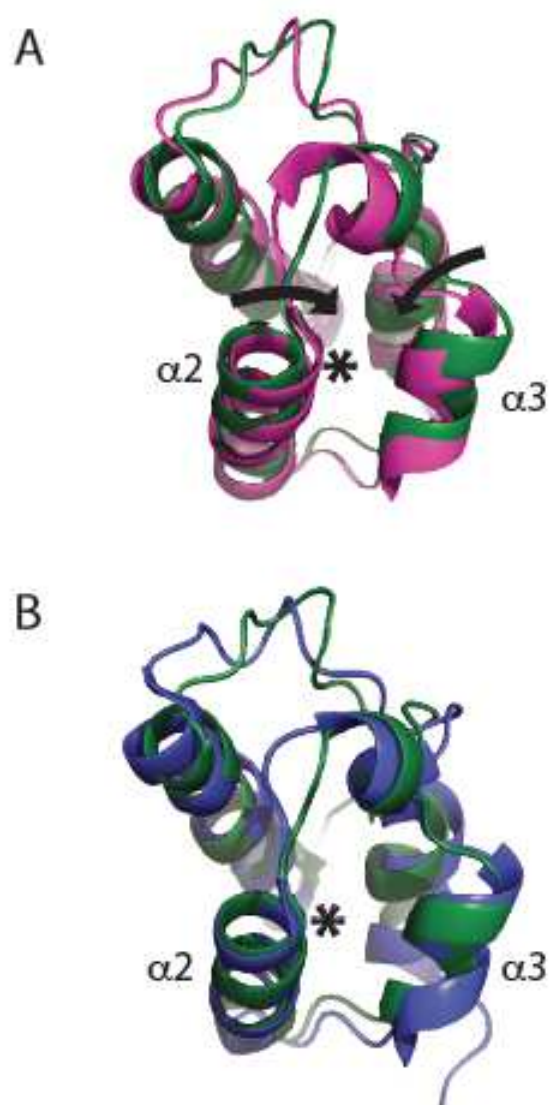


Figure 4-4. Zoom in on the relative position of helix 3 when comparing apo-ArCP with A) holo-ArCP and B) loaded-ArCP. The asterisk denotes the position of the conserved serine at the N-terminus of helix 2. In holo-ArCP, phosphopantetheinylation causes helices 2 and 3 to move towards each other relative to their positions in apo-ArCP. Salicylate loading causes the helices to return back to the position found in apo-ArCP in order to accommodate docking of the tethered salicylate.

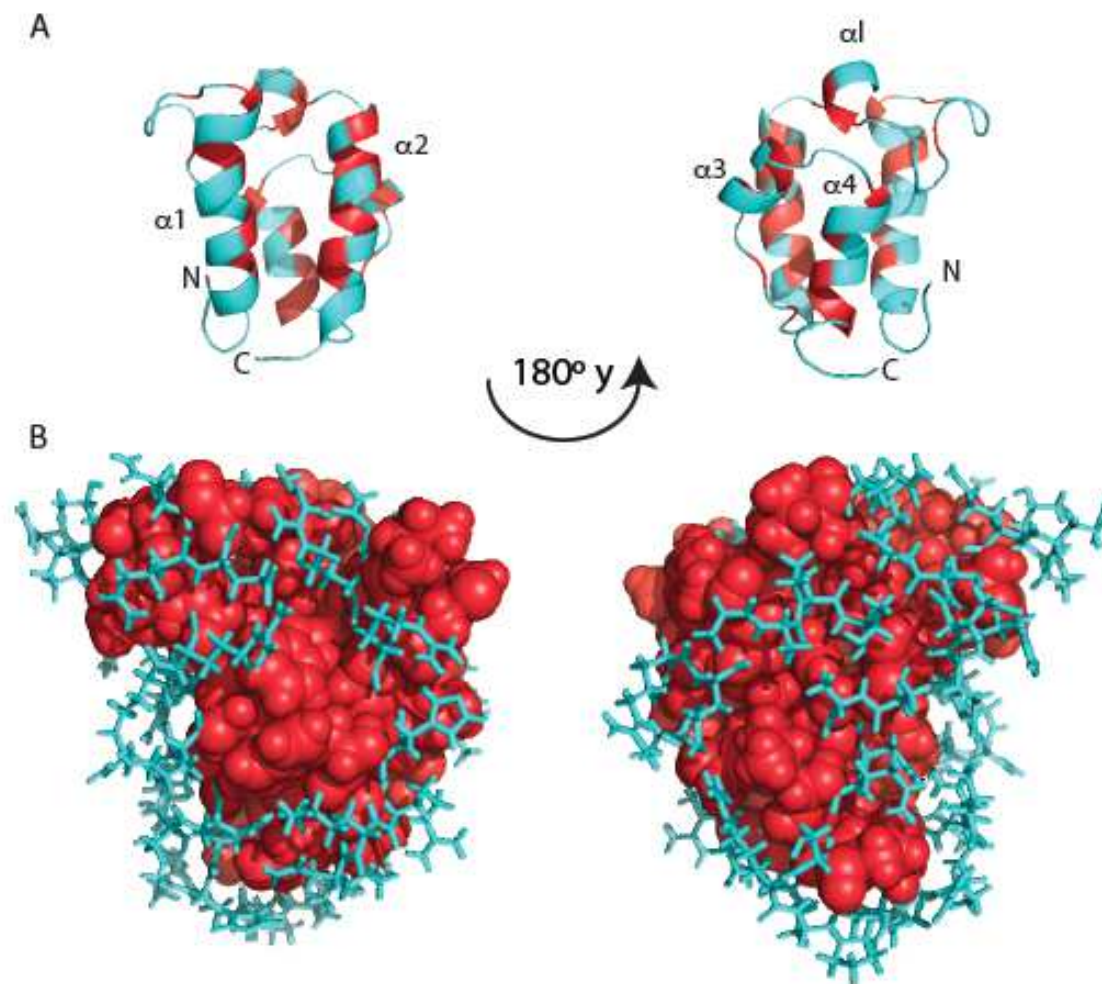


Figure 4-5. A) Cartoon representations of apo-ArCP highlighting the amphipathic nature of the four major helices and loop 1. The large aromatic (F, W and Y) and large hydrophobic (I, L, M, V) residues are colored in red while all others are colored in cyan. B) Sphere and stick representation of apo-ArCP visualizing the hydrophobic core. The orientations are the same as those shown in A). The aromatic and hydrophobic residues form a tightly-packed core and the surface is decorated by hydrophilic amino acids, although surface-exposed hydrophobic patches are evident.

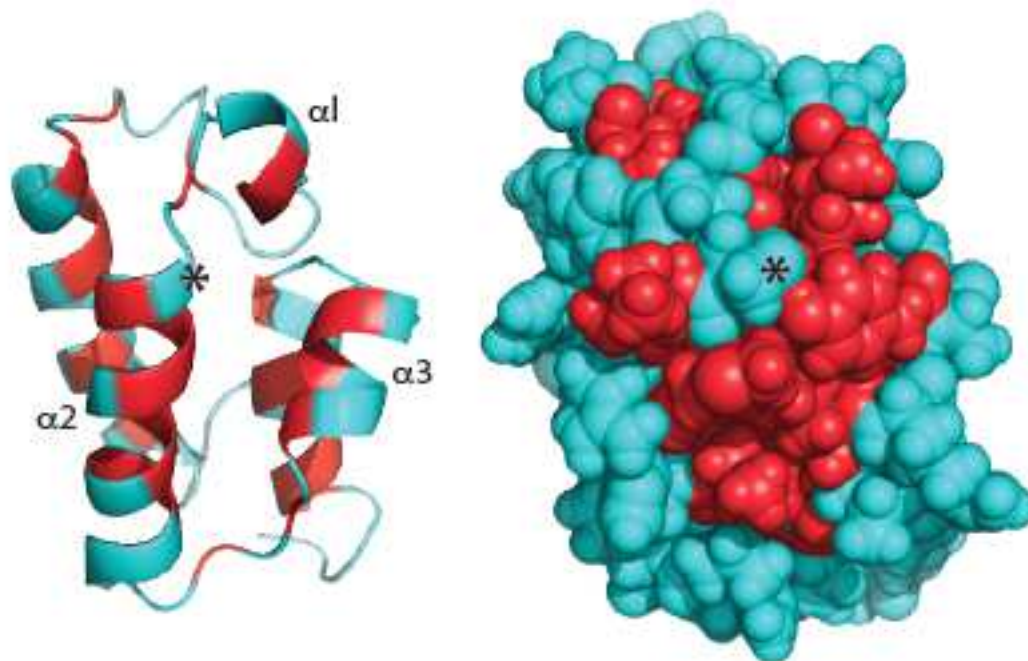


Figure 4-6. Cartoon (left) and sphere (right) representations of apo-ArCP highlighting the solvent-exposed hydrophobic residues that surround the conserved serine (denoted by an asterisk). Both views are shown in the same orientation. Coloring is the same as in Figure 4-5. Hydrophobic residues along $\alpha 2$, $\alpha 3$, and $\alpha 1$ in the vicinity of the PP attachment site are present on the surface of apo-ArCP.

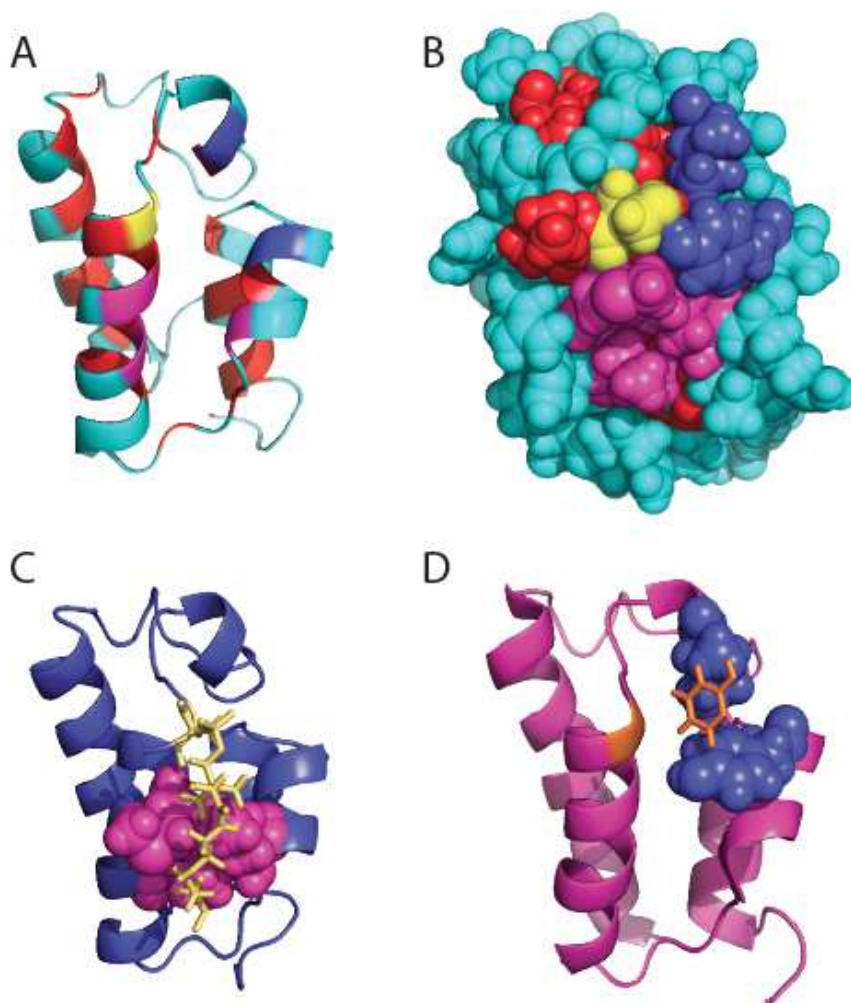


Figure 4-7. A) and B) show the same view as Figure 4-6 with key residues highlighted by color. The PP attachment site is in yellow, solvent-exposed hydrophobic residues that interact with the PP arm are shown in magenta, and residues that interact with salicylate are shown in blue. The residue colored in red immediately adjacent to the conserved serine is residue $i+1$, which mediates the interaction with Sfp. C) Magenta residues from B) shown on the structure of holo-ArCP with the PP arm shown in yellow. These residues form the binding surface for the PP arm. D) Blue residues from B) shown on the structure of loaded-ArCP with salicylate (PP arm not shown for clarity) in orange. The interaction surfaces for the PP arm and salicylate are clearly pre-formed in apo-ArCP.

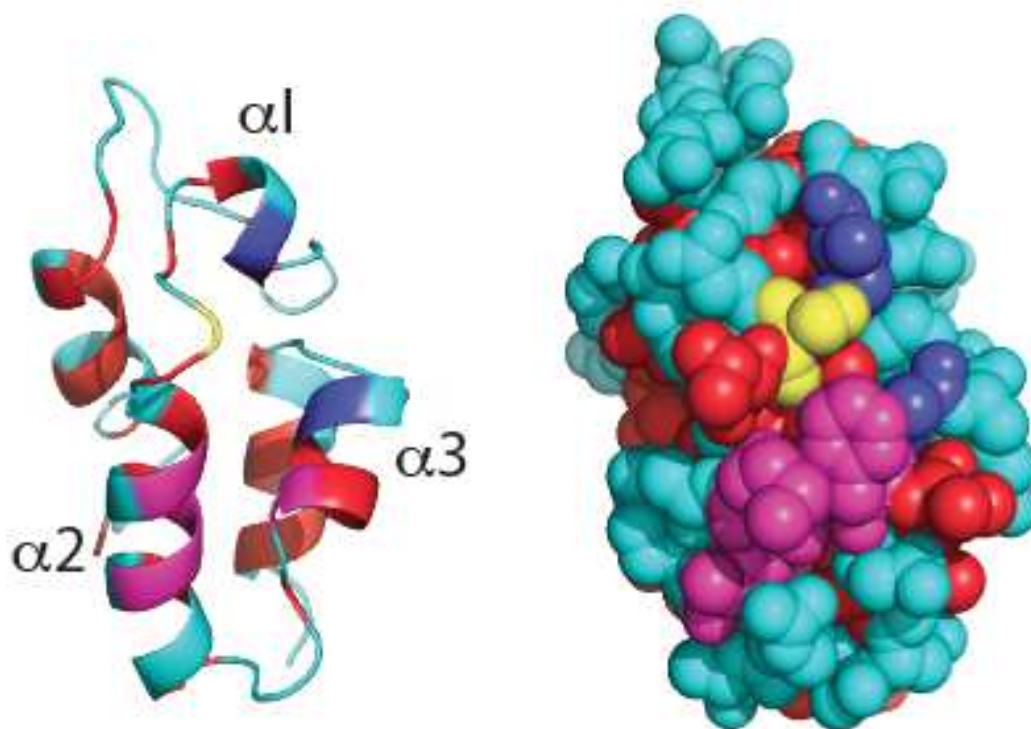


Figure 4-8. Structure of apo-EntB-ArCP shown in the same orientation as apo-ArCP in Figure 4-7. Hydrophilic residues are colored in cyan and all other colors except yellow denote hydrophobic residues. The conserved serine is colored in yellow, putative PP binding surface (i+4, i+7, i+8, and i+19) in magenta, and putative DHB binding surface (i-6 and i+23) colored in blue. EntB-ArCP appears to show similar pre-formed PP and DHB interaction surfaces as apo-ArCP.

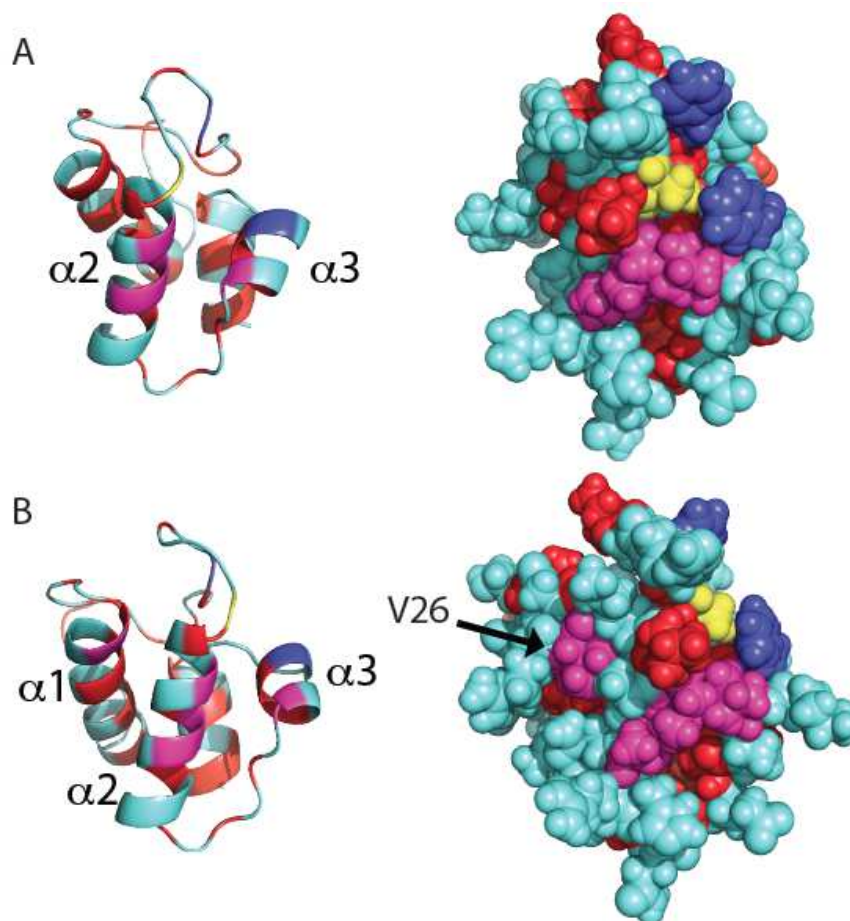


Figure 4-9. Cartoon and sphere representations of apo-TycC3 in A) the same orientation as apo-ArCP in Figure 4-7 and B) an alternative view center on $\alpha 2$ and focusing on the C-terminus of $\alpha 1$. Coloring is as in Figure 4-8. A) Residues i+4, i+7, i+8, and i+20, shown in magenta, form the PP binding surface proposed for the H state of holo-TycC3 and consistent with that found for holo-ArCP. Residues i-6 and i+24, shown in blue, are solvent-exposed phenylalanines and form a potential interaction surface for a tethered substrate. B) A surface-exposed valine at the C-terminus of helix 1 is proposed to form an alternative PP binding site in the A/H form of holo-TycC3.

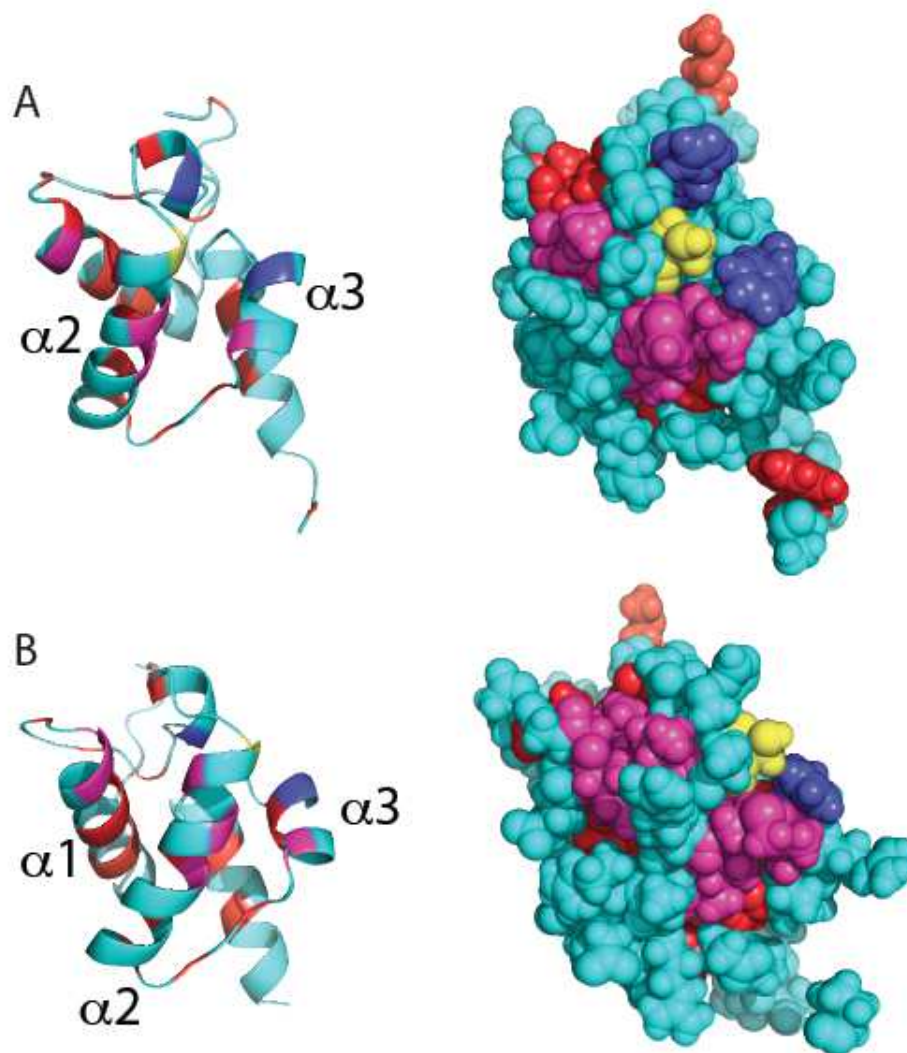


Figure 4-10. Cartoon and sphere representations of PCP7_{Teic} shown with the same color scheme and in the same orientations as in Figure 4-9. A) View highlighting residues (i+4, i+7, and i+20, magenta) that form a putative PP binding surface similar to that found in apo-ArCP. i-6 (Phe) and i+24 (Phe) shown in blue are solvent-exposed aromatic residues that form a potential binding surface for the tethered substrate. B) Alternative view of PCP7_{Teic} shows an additional solvent-exposed hydrophobic surface comprised of residues i+2, i-18, and i-19 that form a putative PP interaction surface consistent with that proposed for the A/H state of holo-TycC3.

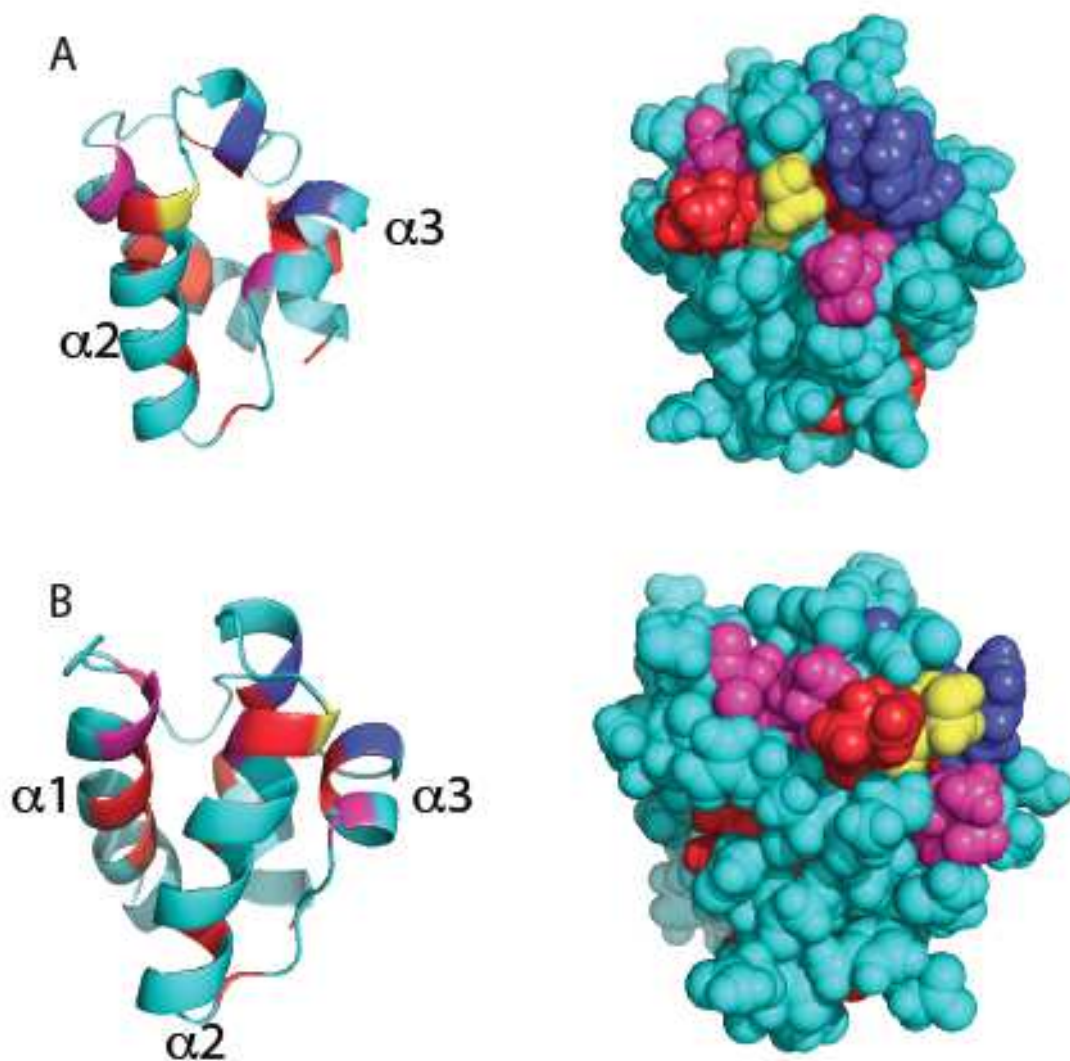


Figure 4-11. Cartoon and sphere representations of PCP1 shown with the same color scheme and in the same orientations as in Figure 4-9. A) apo-PCP1 seems to lack the same hydrophobic PP binding site as apo-ArCP, as residues 1+4 and i+7 are both threonines. Residues i-7 (Phe) and i+23 (Phe) form a potential binding site for tethered cysteine. B) Residues i+2, i-18, and i-19 (shown in magenta) form an alternative PP binding surface consistent with that proposed for the A/H state of holo-TycC3.

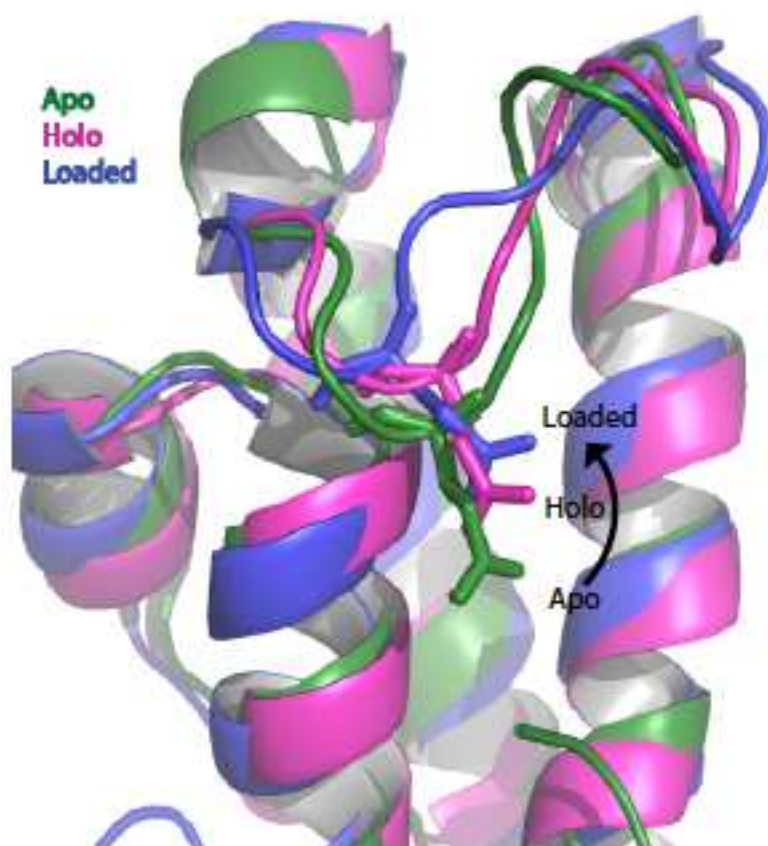


Figure 4-12. Zoom in on loop 1 of apo-, holo- and loaded-ArCP. Structures were aligned based on the four major helices. E41 is shown in sticks. The position of loop 1 in holo-ArCP lies in between the positions observed for apo- and loaded-ArCP.

NMR structure statistics for apo-ArCP	
Violations (mean and s.d.) ^a	
Distance constraints (Å)	0.38 +/- 0.03
Dihedral angle constraints (°)	2.52 +/- 0.56
Max. dihedral angle violation (°)	3.86
Max. distance constraint violation (Å)	0.46
R. m. s. deviations geometry ^b	
Bond lengths (Å)	0.016
Bond angles (°)	1.5
Average pairwise r.m.s.d. [residues 21–90] (Å) ^a	
Heavy	0.94 +/- 0.10
Backbone	0.38 +/- 0.10
Ramachandran Statistics ^c	
Most favoured	93.0
Additionally allowed	7.0
Generously allowed	0.0
Disallowed	0.0

Table 4-1: NMR structure statistics for apo-ArCP. a from CYANA 2.1⁹⁹, b From PSVS¹⁰¹, c from ProCheck¹⁰².

Chapter 5-Structural rearrangements induced by covalent modification to carrier proteins modulate interactions with catalytic domains

ABSTRACT: Nonribosomal peptide (NRP) synthesis is a complex process involving the interplay between covalent protein modifications, conformational changes, catalysis, and protein-protein interactions (PPIs). Delineating the mechanisms involved in orchestrating these various processes will deepen our understanding of domain-domain communication in NRPSs and lay the groundwork for the rational reengineering of NRPSs by swapping domains handling different substrates to generate novel natural products. While many structural and biochemical studies of NRPSs have been performed, few have focused on the thermodynamic parameters governing interactions in these systems. Here, I present my work aimed at understanding how covalent modifications to ArCP influences its interactions with YbtE. Using fluorescence anisotropy, isothermal titration calorimetry, and NMR titrations, I show that covalent modifications modulate the strength and nature of these PPIs in a manner that provides a logical directionality that permits efficient NRP synthesis. This is the first study analyzing binding between a carrier protein in all of its forms and a catalytic domain. The results will provide a basis for guiding the reengineering and optimization of artificial NRPS systems.

Introduction

The modular nature of nonribosomal peptide synthetases (NRPSs) makes them an attractive target for bioengineering efforts, as in principle modules could be rearranged to produce novel products. Done in a combinatorial manner, this strategy could provide a limitless number of new natural products that could be tested for use as antibiotics or anti-cancer agents. However, efforts to utilize this strategy have been limited by poor yields and it has not yet emerged as a viable means of producing novel peptides.

Rational redesign of these systems necessitates an understanding of the molecular mechanisms governing NRPS synthesis. A full description of NRPS synthesis will incorporate structural studies, focusing on proteins of a size ranging from individual domains up to multi-module proteins, with biochemical and biophysical studies describing the thermodynamics of these systems. The work described in the previous two chapters has outlined our contributions to elucidating the structural and dynamic changes to carrier proteins upon various post-translational modifications. This builds upon previous efforts investigating the structure and dynamics of individual adenylation domains⁹, condensation domains^{19,109}, and CPs^{2,4-6,54,108}, a CP/TE didomain¹⁴, A/CP complexes^{10,11}, and a full module with an apo CP¹⁸. Recent efforts have further expanded our structural understanding of NRPS synthesis by providing additional snapshots of full modules as they progress through NRP synthesis^{12,13}.

While these efforts have contributed significantly towards developing a structural understanding of NRPS synthesis and provided a number of snapshots of possible states, few have focused on the energetics governing transitions between these states. The study of the apo-CP/TE didomain showed that a conformational exchange between bound (CP

interacting with TE) and unbound forms was modulated by an interaction with a phosphopantetheinyl transferase (PPTase)¹⁴. It was also found that the terminal thiol of the PP arm of a holo-CP/A di-domain was protected when the A domain was bound to an adenylate mimic, suggesting that substrate binding by A domains promotes binding of the holo-CP and burial of the PP arm in the active site¹¹⁰. In order to monitor how molecular events modulate domain affinities, and how exogenous domains may still communicate with non-cognate partners in artificial NRPSs, it is necessary to first characterize the interaction of each state of a CP (apo, holo, and substrate-loaded) with each of its potential native interaction partners.

Here, I used a combination of fluorescence anisotropy (FA), isothermal titration calorimetry (ITC), and NMR titrations to characterize the interactions between the adenylation domain, YbtE, and the apo, holo, and salicylate-loaded forms of ArCP. I also investigated the interaction between holo and salicylate-loaded ArCP and Cy1 using ITC. At the onset of these studies, we hypothesized that the catalytic domains would interact most strongly with the form of ArCP that serves as a substrate for its chemical reaction (holo-ArCP for YbtE, loaded-ArCP for Cy1) and weakly for the form that is a product of the reaction (loaded-ArCP for YbtE, holo-ArCP for Cy1) or inactive (apo). Fluorescence anisotropy titrations, ITC experiments, and NMR titrations all confirmed our expectation that YbtE would preferentially bind to holo-ArCP over apo-ArCP. However, NMR titrations revealed that apo- and holo-ArCP interact with YbtE with the same set of residues and thus raises questions about how YbtE discriminates between the two forms. Further, FA and ITC experiments showed that YbtE has a similar affinity for both holo- and loaded-ArCP, a result seemingly inconsistent with our initial hypothesis. An examination of the

thermodynamics provided by the ITC experiments along with the structural and dynamics studies presented in the previous chapters leads me to propose a model whereby subtle structural rearrangements of the ArCP core and modulation of binding interfaces by the PP arm leads to differential binding modes that provide a logical directionality to this set of PPIs and permit efficient synthesis.

Results

In order to characterize the relative affinities of each of the forms of ArCP for the A domain YbtE, we first used fluorescence anisotropy (FA) due to the capacity to perform multiple titrations in parallel in a 384-well plate. Further, wild type ArCP does not feature any cysteine, which facilitates their engineering for controlled incorporation of fluorescent probes. Using site-directed mutagenesis, a single cysteine was engineered at position 16 of ArCP, at the very N-terminal end of helix 1. This site was chosen because helix 1 has never been found to be involved in interactions with catalytic domains and this residue is relatively rigid on a ps-ns timescale based on its high order parameter (see Chapter 3). Thus, a fluorescent probe at this position will not interfere with binding events and will be a good reporter of molecular tumbling. Apo-ArCP_R16C was labeled with fluorescein maleimide and divided into three aliquots, one of which remained apo, one of which was converted to the holo form, and the last of which was modified with a nonhydrolyzable amide analog of salicylate-loaded phosphopantetheine (referred to as SalNH-ArCP). The latter allowed for monitoring binding of YbtE with its product, and thus for comparing the affinity between YbtE and its substrate and product.

The results of the fluorescence anisotropy titrations are shown in Figure 5-1. As expected, YbtE shows a relatively weak interaction with apo-ArCP compared with holo-ArCP, with

apparent K_D 's of $68.6 \pm 9.62 \mu\text{M}$ and $7.22 \pm 1.34 \mu\text{M}$, respectively, based on fitting to a one-site specific binding model. This was expected because the apo form is inactive and therefore not a substrate for YbtE. However, we also found that YbtE binds to holo-ArCP and SalNH-ArCP with very similar affinities (K_D 's of 7.22 ± 1.34 vs. $4.70 \pm 1.10 \mu\text{M}$). We had anticipated the interaction with SalNH-ArCP, a *product* of the chemical reaction performed by YbtE, to be comparatively weak; however, our FA titrations show this not to be the case.

In order to both verify the results of the FA titrations and gain insight into the thermodynamics of each interaction, we performed ITC experiments with each form. All thermodynamic parameters are listed in Table 5-1. In these experiments, either apo-, holo-, or SalNH-ArCP was in the cell and YbtE in the syringe. The baseline-corrected data for each titration is shown in the top panel of Figures 5-2, 3 and 4 and the integrated signals fit to a 1:1 model are shown in the lower panels of the same figures. Prior to fitting, the integrated heat from the final four points of each titration was averaged and subtracted from all points to account for the heat of dilution.

The results of the ITC experiments recapitulate the trends observed in the FA experiments. That is, YbtE binds relatively weakly to apo-ArCP ($11.6 \pm 1.89 \mu\text{M}$) and interacts with holo- and SalNH-ArCP with similar affinities ($2.70 \pm 0.37 \mu\text{M}$ and $1.86 \pm 0.18 \mu\text{M}$, respectively). For holo- and SalNH-ArCP, the K_D 's determined from the ITC experiments are also in good agreement with those found by FA. For apo-ArCP, neither FA nor ITC results fit well to a 1:1 binding model and the results will only be used qualitatively.

The entropic and enthalpic contributions to each of these interactions are starkly different.

The interaction between apo-ArCP and YbtE is endothermic, with a measured ΔH of

+0.610 \pm 0.035 kcal/mol. This indicates that fewer favorable interactions are present in the complex than in the isolated proteins and their association is driven purely by a positive change in entropy, calculated to be 7.38 \pm 1.10 kcal/mol. An interpretation of the thermodynamics of these interactions will be given in the discussion section below. In contrast to the interaction between apo-ArCP and YbtE, binding of YbtE to holo-ArCP is highly exothermic. A positive change in entropy also contributes to this interaction, although to a lesser extent than with apo-ArCP. Finally, a negative change in enthalpy and positive change in entropy both contribute to the interaction between SalNH-ArCP and YbtE. In this case, the contribution of enthalpy to the free energy of binding is lower than that seen for holo-ArCP (1.421 \pm 0.021 kcal/mol vs. 3.501 \pm 0.084 kcal/mol) and the interaction with SalNH-ArCP has a more favorable change in entropy. This suggests that even though holo-ArCP and SalNH-ArCP have similar affinities for YbtE, they interact with YbtE via rather different mechanisms.

Finally, to identify the residues on ArCP that mediate the interaction with YbtE, we performed NMR titrations with the apo and holo forms. In these experiments, ^{15}N -labeled ArCP was held at constant concentration and unlabeled YbtE was titrated. An HN-HSQC was collected at each concentration of YbtE. SalNH-ArCP was not available when the titrations were performed.

Figures 5-5 and 5-6 show the spectroscopic signature of complex formation for apo- and holo-ArCP, respectively. First, it can be seen that all signals are not affected in a uniform manner, demonstrating that binding is specific to a subset of residues. Additionally, comparison of signal line-shapes and positions during apo- and holo-ArCP titrations confirms that YbtE preferentially binds to holo-ArCP over apo-ArCP. The signals in the

apo-ArCP titration show large changes in chemical shift but small changes in intensity upon interaction, indicative of fast exchange on the NMR time-scale. In contrast, at equivalent concentrations of YbtE, signals of holo-ArCP broaden and disappear into the noise without a significant change in position at equivalent concentrations of YbtE, characteristic of intermediate exchange and demonstrating a tighter interaction than that with apo-ArCP.

In Figures 5-7 and 5-8, residues affected by the interaction are plotted onto the structures of apo- and holo-ArCP. For apo-ArCP, the signals are colored according to the concentration of YbtE at which they passed a threshold chemical shift perturbation while for holo-ArCP they are colored based on the concentration at which the signal became undetectable. Surprisingly, the patterns shown for apo- and holo-ArCP are remarkably similar. For both forms, all residues affected cluster along the solvent-exposed faces of helices 2 and 3. Figures 5-7 and 5-8 show that these residues form a continuous patch along the surface of ArCP that is nearly identical between the two forms. This demonstrates that YbtE interacts with apo- and holo-ArCP at the same surface despite having different affinities for the two forms and significantly different thermodynamics for the two interactions.

ITC experiments were also performed to attempt to characterize the interaction between Cy1 and holo- or SalNH-ArCP and the results are shown in Figure 5-9. Even with 100 μ M holo- or SalNH-ArCP in the cell, no interaction could be detected. One possible explanation for this result is that Cy1 needs to first bind to the downstream CP, PCP1, before it can bind ArCP. It is also possible that the Cy1 construct we are using is simply inactive and new constructs need to be tested.

Discussion

Prior to our investigations, little was known about whether or how NRPS catalytic domains discriminate between the various forms in which CPs exist. Before structures of multi-domains became available, there were two competing models of NRPS synthesis. In the first model, NRPS were expected to display a rigid organization with the long, flexible PP arm swinging between catalytic sites. In the second model, NRPSs acted as bead on a string and domains would randomly interact with each other during synthesis. Crystallography, NMR, and CryoEM have since shown that the reality lies somewhere in between both models: NRPSs seem to adopt a series of transient, well-defined quaternary conformations during synthesis. However, it is still unclear how NRPSs stabilize the relevant conformation in each catalytic step. One possibility is that catalytic domains do not discriminate between the carrier protein forms, but instead the CPs randomly visit the various active sites available and chemistry occurs when the relevant pair of CP and catalytic domain is formed. Alternatively, chemical modifications of CPs may modulate interactions between CPs and catalytic domains. Here, I show how covalent modifications modulate both the affinity and the nature of domain-domain interactions.

We originally hypothesized that the strength of the interactions between ArCP and the catalytic domains would reflect ArCP's role as either a substrate or product of the various chemical reactions. That is, holo-ArCP should interact strongly with YbtE but weakly with Cy1, SalNH-ArCP should interact strongly with Cy1 but weakly with YbtE, and apo-ArCP should interact weakly with both since it is a substrate for the phosphopantetheinyl transferase YbtD but not YbtE or Cy1. While we found that YbtE preferentially interacts with the holo form over the apo form, we also found that it binds to holo- and SalNH-ArCP

with similar affinity. Further, despite our FA, ITC, and NMR titration results indicating a preference for holo-ArCP over the apo form, the NMR titrations showed that YbtE interacts with the two forms at very similar interfaces. Comparison of the measured dissociation constants and interaction surfaces alone does not provide a rationale for understanding how YbtE discriminates between these two forms, nor do the binding affinities rationalize the tight association between YbtE and SalNH-ArCP. Instead, a consideration of the entropic and enthalpic contributions to each interaction along with the structures presented in the previous two chapters is necessary to understand how the biochemical state of ArCP modulates the affinity for YbtE in a manner that reflects its role as a substrate or product of the loading reaction.

As shown in Figure 5-10 and listed in Table 5-1, although the apparent K_D 's for the three forms are all within an order of magnitude and are effectively identical for the holo and loaded forms, the relative contribution of enthalpy and entropy to each interaction varies widely. For apo-ArCP, the interaction is enthalpically unfavorable and is therefore driven exclusively by a favorable change in entropy. For holo-ArCP, changes in entropy and enthalpy contribute almost equally to the interaction, while for SalNH-ArCP, the interaction is driven predominantly by a favorable change in entropy but is still exothermic. In order to interpret these thermodynamic parameters mechanistically, it is necessary to consider the individual factors that contribute to the overall changes of enthalpy and entropy upon interaction. Changes in enthalpy arise from making or breaking chemical interactions, such as van der Waal's contacts, hydrogen bonds, and charge-charge interactions¹¹¹. When more bonds are formed than broken, the change in enthalpy will be

negative and drive an interaction (as for holo- and SalNH-ArCP), while the opposite is true if more bonds are broken than formed (apo-ArCP).

Changes in entropy come from a greater number and wider variety of sources and can be attributed to changes involving the solvent (ΔS_{solv}) or the interacting molecules themselves (ΔS_{mole})¹¹¹. ΔS_{mole} can be broken down into changes in conformational entropy (ΔS_{conf}), translational and rotational entropy, and residual entropy¹¹¹. When considering complex formation, ΔS_{conf} refers to the number of conformations (tertiary structures) each binding partner can adopt. The total change in entropy upon association is the sum of ΔS_{solv} and ΔS_{mole} and will incorporate changes in entropy from both ArCP and YbtE. The following discussion will focus on ΔS_{solv} , which can be calculated^{112,113}, and ΔS_{conf} , which will be discussed qualitatively.

A positive ΔS_{solv} is generally considered to come from a release of ordered water molecules away from the surfaces of two interaction proteins and into bulk solution¹¹¹. This reflects the burial of nonpolar groups at the protein-protein interface and can be calculated as a function of temperature based on the changes in solvent-accessible surface area (ΔASA) of polar and nonpolar groups as $\Delta S_{\text{solv}}(T) = 0.45 * \Delta \text{ASA}_{\text{nonpolar}} * \ln(T/385) - 0.26 * \Delta \text{ASA}_{\text{polar}} * \ln(T/335)$ in units of $\text{cal} * \text{mol}^{-1} * \text{K}^{-1}$.^{112,114} While there is no structure available of an ArCP:YbtE complex, there is a co-crystal structure of the homologs EntE and EntB from the enterobactin system with EntE in the thioester conformation¹⁰. This arrangement is expected to maximize the interaction surface between A domain and CP. Calculating ΔS_{solv} based on this crystal structure and $T=300\text{K}$ (temperature at which ITC experiments were performed) gives a ΔS_{solv} of 17.1 kcal/mol. This value likely only holds

for apo- and holo-ArCP, which were shown by NMR titration to interact with YbtE at the same surface as that found in the EntE:EntB co-crystal structure and may be lower for SalNH-ArCP, which we propose has a smaller interaction surface.

ΔS_{conf} reflects the total number of conformations available to a protein in different states. When a discrete number of conformations can be identified, this can be calculated as $\Delta S = k_B \ln(n_2/n_1)$, where n_i represents the number of conformations available in state i ¹¹. As can be seen, if fewer conformations are available in state 2 than state 1 (for example, bound vs. unbound), the contribution to the overall entropy will be negative and will disfavor binding. A domains are comprised of two sub-domains, a large N-terminal domain and a smaller C-terminal domain, and adopt a number of different conformations that differ in the relative orientation of the sub-domains⁹. The so-called adenylation conformation is proposed to catalyze formation of the acyl-adenylate. The C-terminal domain is then proposed to rotate 140° to form the thioester conformation, which interacts with holo-CPs to catalyze substrate loading. The number of states available to YbtE must therefore be considered. Additionally, the number of conformations available to ArCP varies depending on the form. Apo- and holo-CPs have been reported to exist in two states⁵, named A and A/H for the apo form and A/H and H for the holo form(see chapter 4), although the significance of this observation has recently been debated⁶⁸. In addition to having multiple conformers of the core protein, holo-ArCP harbors a highly flexible PP arm that interconverts between a bound and unbound form. Further, in the unbound state the PP arm itself exists in a disordered state sampling multiple conformations as evidenced by NMR relaxation parameters (Chapter 3). Apo-ArCP has no modifications and likely has fewer possible conformations than holo-ArCP. Finally, the salicylate-loaded PP arm is also

dynamic and also likely adopts multiple conformations, although it is less flexible than the PP arm of holo-ArCP based on the NMR dynamics measurements presented in Chapter 3. All of these dynamic events must be taken into account when discussing binding.

The interaction between apo-ArCP and YbtE is endothermic, indicating that fewer favorable interactions are present in the complex than in the free proteins, and complex formation is therefore driven entirely by favorable changes in entropy. As noted above, ΔS_{solv} contributes 17.1 kcal/mol, significantly higher than the 7.38 ± 1.10 kcal/mol calculated from the ITC data. I propose that selection of a single conformation of YbtE upon binding accounts for a portion of the difference between ΔS_{solv} and the measured entropy via a negative change in conformational entropy. This pre-existing conformational exchange by YbtE may also rationalize the endothermic nature of the interaction. As described in Chapter 1, A domains have crystallized in at least three different conformations and may access many more in solution. However, only one, the thioester conformation, is expected to interact with holo-ArCP and, due to the similarity in binding interfaces identified in the NMR titrations, probably the apo form as well. Comparison of the solvent-accessible surface area of the C-terminal domain in the adenylation (calculated for DhbE¹¹⁵, another homolog of YbtE and EntE) and thioester conformations (EntE¹⁰) shows that an additional 206.7 square angstroms is buried in the adenylation conformation, which must therefore feature more intradomain interactions. Hence, the transition from the adenylation to thiolation conformation may therefore require an input of energy, making it an endothermic process. The amount of energy necessary to make this transition is apparently greater than that provided by the interaction with apo-ArCP, making the overall process of the conformational change plus the interaction with apo-ArCP endothermic.

Additionally, selecting for the thioester conformation out of multiple possible conformations would also negatively contribute to the overall change in entropy, as $n_2 < n_1$, making $\ln(n_2/n_1)$ negative. Based on these considerations, release of water into bulk solution is likely the main driving force for the interaction between apo-ArCP and YbtE. In contrast to apo-ArCP, the interaction between holo-ArCP and YbtE is driven almost equally by entropy and enthalpy. The contribution of entropy to this interaction is less than for the interaction with apo-ArCP. As we showed in Chapter 3, the PP arm on holo-ArCP is highly dynamic and likely interconverts between a bound form and many unbound forms, only one of which is present upon an interaction with an A domain⁶ (referred to as e-holo-ArCP in Chapter 3). Therefore, the association of YbtE and holo-ArCP likely selects for a single conformation of YbtE *and* a single conformation (out of many) of holo-ArCP, leading to a further reduction in ΔS . Notably, despite using the same interaction as apo-ArCP as indicated by NMR titrations, this interaction becomes strongly exothermic, not endothermic. The PP arm itself is only expected to interact weakly with the A domain, based on the crystal structure of the EntE:EntB complex¹⁰ and biochemical data showing that free phosphopantetheine is a poor substrate for A domains¹¹⁶, so this is unlikely to be the source of the additional favorable interactions. In Chapter 4, I described how phosphopantetheinylation causes a minor rearrangement of helices 2 and 3 due to a hydrophobic interaction between residues at the N-terminus of helix 3 and the methyl groups on the PP arm. According to our NMR titrations, helices 2 and 3 comprise the interaction surface for YbtE. Phosphopantetheinylation thus appears to properly position the residues in this region for an optimal interaction with an A domain, leading to stronger

interactions and a negative enthalpy that more than compensates for the entropy lost due to freezing out the motion of the PP arm.

Finally, while loaded-ArCP binds to YbtE with similar affinity as holo-ArCP, the thermodynamics of the interaction are quite different and provide insight into the nature of the interaction (Figure 5.10). This interaction is exothermic, although less so than the interaction between holo-ArCP and YbtE by ~ 2.1 kcal/mol. To achieve a similar overall binding affinity, this interaction is more entropically favorable than that of holo-ArCP. Again, the structures and dynamics described in Chapter 3 allow us to rationalize this observation. In the loaded form, ArCP interacts directly with the tethered substrate. This interaction positions the PP arm such that the PP arm and substrate obscure helices 2 and 3, but not loop 1. In the co-crystal structure of EntE and EntB, loop 1 of EntB appears to interact with the C-terminal domain of EntE¹⁰. We proposed that the interaction with the substrate breaks the interaction with the N-terminal domain while maintaining an interaction with the C-terminal domain. We further proposed that this would free the C-terminal domain to rotate relative to the N-terminal domain in order to deliver the CP to the condensation domain⁶, an idea supported by recent crystal structures^{12,13}. Based on this model, the loaded-ArCP:YbtE interaction would have a smaller interface than that with holo-ArCP, forming fewer interactions and therefore having a smaller change in enthalpy, while also freeing the C-terminal domain to sample multiple conformations, increasing the number of states possible and therefore increasing the entropy of the bound form relative to the complex formed between YbtE and holo-ArCP.

The sum of these results allows us to build a model of how covalent modifications provide directionality to the set of protein-protein interactions necessary for substrate loading. First,

phosphopantetheinylation of apo-ArCP repositions residues at a pre-formed binding site, optimizing the interactions and providing the additional energy that results in preferential binding of YbtE to holo-ArCP over apo-ArCP. Substrate loading then obscures part of this binding surface, partially disrupting the protein-protein interaction but freeing the C-terminal domain to rotate. In the context of a full module, this would allow the A domain to deliver the loaded-CP to the next catalytic domain. The combination of NMR structural and dynamic studies with FA, ITC, and NMR titrations presented throughout this work thus allows us to propose a model in which covalent modifications to ArCP impart a directionality to the set of interactions with YbtE in a manner that parallels the chemical steps of elongation.

The work described here serves as a starting point for analyzing the influence of additional substrates on the interaction between catalytic domains and carrier proteins. All of our titrations were done with free YbtE; however, for loading to occur, YbtE would contain an activated salicyl-adenylate in its active site. Our studies with free YbtE will serve as a point of reference for titrations done in the presence of salicylate, ATP, or the nonhydrolyzable salicyl-adenylate mimic Sal-AMS.

Finally, the binding studies done here between a wild-type A domain and its cognate CP will also guide efforts to reengineer NRPS systems. Interactions between A domains and non-cognate CPs or mutant A domains with altered substrate specificity and their cognate CPs can be optimized to match the binding affinities and thermodynamics measured for ArCP and YbtE. The mechanistic insights into NRPS synthesis provided by our work combining NMR structural and dynamic data with FA, ITC, and NMR titrations thus lay

the foundation for engineering artificial NRPS systems and the production of novel secondary metabolites.

Methods

Expression and Purification of PanK, Ppat, and DPCK

Unless otherwise noted, pH listed for each buffer is the pH at 4 °C.

Expression and purification of PanK, Ppat, and DPCK is identical except where noted.

Plasmids encoding for PanK, Ppat, and DPCK (courtesy of Dr. Craig Townsend) were transformed into *E. coli* BL21 (DE3) cells and plated on luria broth (LB) agar with kanamycin. A single colony was selected and added to 15 ml LB with kanamycin and grown at 37 °C with shaking at 250 rpm overnight. The following day, 10 ml of the overnight culture was added to 1 L LB with kanamycin and growth continued at 37 °C with shaking. At OD₆₀₀=0.6, cultures were placed in an ice bath and allowed to cool to 15 °C. IPTG was then added to a final concentration of 1 mM and growth continued at 16 °C with shaking for 16 hours. Cells were harvested by centrifugation and pellets flash frozen in liquid nitrogen. Pellets were stored at -80 °C until needed.

To begin purification, cell pellets were resuspended in 50 ml lysis buffer (50 mM Tris, pH 8.0, 0.5 M NaCl, 30 mM imidazole 200 µg/ml lysozyme, 2 µg/ml DNase I) and lysed using either a microfluidizer or French pressure cell. The lysate was clarified by centrifugation at 27,000 xg for 30 minutes at 4 °C. The supernatant was filtered and loaded onto a 5 ml HisTrap HP column. The column was washed with 16 column volumes (CV) His Buffer A (50 mM Tris, pH 8.0, 0.5 M NaCl, 30 mM imidazole) on an Akta purifier while collecting fractions. The column was eluted with a linear gradient from 0-100% His Buffer B (50 mM Tris, pH 8.0, 0.5 M NaCl, 0.5 M imidazole) over 25 CV while collecting fractions.

Fractions were analyzed by SDS-PAGE and fractions containing desired protein were pooled and dialyzed against 2 L 20 mM Tris, pH 7.5, 0.1 M NaCl at 4 °C overnight.

The following day, the samples were concentrated to 2 ml or less and loaded onto a Superdex 200 16/60 pg size exclusion column that had been equilibrated with 20 mM Tris, pH 7.5, 0.1 M NaCl and run at max flow-rate (1.2 ml/min) while collecting fractions. Fractions were analyzed by SDS-PAGE and fractions containing pure target protein pooled.

Removal of Contaminating Coenzyme A from Ppat and PanK

Ppat and PanK co-purify with coenzyme A, which will result in contamination of samples with holo-CP. ATP is a competitive inhibitor of Coenzyme A and was used to remove bound Coenzyme A from Ppat and PanK. To remove coenzyme A, Ppat and PanK were buffer exchanged into 20 mM Tris, pH 7.5, 0.1 M NaCl, 20 mM ATP by repeated concentration and dilution in this buffer until a 1000-fold exchange had been achieved. Samples were then exchanged into 20 mM Tris, pH 7.5, 0.1 M NaCl, 10% glycerol (v/v) by repeated concentration and dilution until a 1,000-fold exchange had been achieved.

DPCK was also exchanged into 20 mM Tris, pH 7.5, 0.1 M NaCl, 10% glycerol (v/v) in the same manner. Samples were then concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C until needed. Concentrations were determined by UV-vis absorbance at 280 nm. Extinction coefficients used were 45380/M*cm (PanK), 8480/M*cm (Ppat), and 16960/M*cm (DPCK).

Modification of Apo-ArCP with Nonhydrolyzable Amide Mimic for ITC

A 10 ml reaction in 100 mM Tris, pH 7.5 at 22 °C, 10 mM MgCl₂, and 100 mM NaCl with 50 µM apo-ArCP, 150 µM nonhydrolyzable amide mimic, 5 mM ATP, 500 nM Sfp, 500 nM PanK, 500 nM Ppat, and 500 nM DPCCK was prepared and incubated at room temperature for 4 hours. The sample was then concentrated and run on a Superdex 75 16/60 pg (GE Healthcare) size exclusion column that had been equilibrated with ITC buffer (50 mM ACES, pH 6.80 at 22 °C, 150 mM NaCl, 1 mM MgCl₂, 2 mM TCEP) and peak fractions collected. Fractions were analyzed by SDS-PAGE and completion of reaction confirmed using MALDI.

ITC Experiments

For all ITC experiments, ITC buffer (50 mM ACES, pH 6.80 at 22 °C, 150 mM NaCl, 1 mM MgCl₂, 2 mM TCEP) was used.

ITC experiments were performed using a VP-ITC Microcalorimeter (MicroCal).

YbtE was purified as described in Chapter 2.

To prepare YbtE or Cy1 for titrations, samples were exchanged >100-fold into freshly prepared ITC buffer by repeated concentration and dilution in a 30K MWCO centrifugal filter (Millipore). Samples were then filtered through a 0.45 µm filter to remove any precipitate that had formed during concentration. A 5 µl sample was taken and diluted 5-fold into ITC buffer. 5 µl of this was then diluted 20-fold into 6.3 M guanidinium-HCl. 5 µl of ITC buffer was also diluted 20-fold into 6.3 M guanidinium-HCl to use as a blank. A₂₈₀ was measured and the concentration calculated using extinction coefficients of 52370/M*cm (YbtE) and 80000/M*cm (Cy1). Samples were then diluted to the final working concentration and stored on ice until needed.

Samples of ArCP were exchanged >100-fold into freshly prepared ITC buffer by repeated concentration and dilution in a 3K MWCO centrifugal filter. Samples were filtered through a 0.45 μm filter to remove any precipitate that had formed during concentration. A 5 μl sample was taken and diluted 20-fold into 6.3 M guanidinium-HCl. A blank was prepared as described above. A_{280} was measured and the concentration calculated using an extinction coefficient of 20970/M*cm. Samples were diluted to the final working concentration and stored on ice until needed.

All samples were degassed prior to use.

For experiments with YbtE, ArCP (40 μM) was in the cell and YbtE (720 μM) in the syringe and the following settings used: Cell temperature: 27 $^{\circ}\text{C}$; Reference power: 20 $\mu\text{cal/second}$; Initial delay: 600 seconds; Stirring speed: 300 rpm; Feedback mode: High; Equilibration options: Fast equil./Auto. The injection schedule was as follows: one 2 μl injection with 150 second delay followed by 24 12 μl injections with 300 second delays. Data was analyzed using Origin.

For experiments with Cy1, ArCP (100 μM) was in the cell and Cy1 (1.2 mM) in the syringe and the following settings used: Cell temperature: 27 $^{\circ}\text{C}$; Reference power: 4 $\mu\text{cal/second}$; Initial delay: 600 seconds; Stirring speed: 300 rpm; Feedback mode: High; Equilibration options: Fast equil./Auto. The injection schedule was as follows: one 2 μl injection with 150 second delay followed by 20 15 μl injections with 300 second delays. For titration of Cy1 into buffer, the cell contained only ITC buffer but all settings were identical. For titrations of buffer into ArCP, the syringe contained only ITC buffer but all settings were identical. Raw data was integrated in Origin and plotted using Prism 5.

Cloning of ArCP_R16C

A cysteine was introduced into GB1-ArCP-14-93 using primers ArCP_14-93_R16C_For (5'-CACGAAAATCTTTATTTTCAAGGTACCGACAACCTGCCACGCGGCTG-3') and ArCP_14-93_R16C_Rev (5'-CAGCCGCGTGGCAGTTGTCTGGTACCTTGAAAATAAAGATTTTCGTG-3') using the Phusion High-Fidelity Master Mix with High Fidelity Buffer according to the protocol recommended by the manufacturer (New England Biolabs). The reaction contained 5% DMSO and used the following PCR protocol: 1 round of 30 seconds at 98 °C; 30 rounds of 10 seconds melting at 98 °C, 30 seconds annealing at 72 °C, 5 minutes extension at 72 °C; 1 round of 10 minutes extension at 72 °C. Colonies containing the proper mutation were identified by sequencing.

Expression and purification of ArCP_R16C are the same as for apo-ArCP 14-93 (Chapter 2) except dialysis buffers contain 2 mM DTT and the final size exclusion buffer was FA Labeling Buffer (10 mM HEPES, pH 7.5, 150 mM NaCl, 500 μ M TCEP).

Fluorescent Labeling of apo-ArCP_R16C

Purified apo-ArCP_R16C (5 mg) in FA Labeling Buffer was concentrated to a volume of ~3 ml and Triton X-100 added to a final concentration of 1% (w/v). 1 mg of fluorescein-5-maleimide (Vector Laboratories) dissolved in 30 μ l DMSO was then added to apo-ArCP_R16C. Tubes were wrapped in parafilm and aluminum foil and rocked at 4 °C overnight.

The following day, β -mercaptoethanol was added to a final concentration of 10 mM and rocking continued at 4 °C for one hour. Excess dye was removed by repeated concentration

and dilution in FA Labeling Buffer and the sample concentrated to 2 ml. The sample was then run on a Superdex 75 16/60 pg that had been equilibrated with ITC buffer. Peak fractions were analyzed by SDS-PAGE and fractions containing ArCP were pooled. This pool was divided into three aliquots: one remained apo, one was phosphopantetheinylated as described in Chapter 2, and the other was modified with a nonhydrolyzable amide mimic as described above.

Protein concentration and degree of labeling were calculated according to the manufacturer's protocols. Degree of labeling was found to be 0.282 (apo-ArCP), 0.352 (holo-ArCP), and 0.344 (SalNH-ArCP). Glycerol was added to 10% (w/v) and samples flash frozen in liquid nitrogen and stored at -80 °C until needed. After thawing and before use in titrations, samples were dialyzed against freshly prepared ITC buffer such that a >10,000-fold dilution of glycerol was achieved.

To prepare samples for fluorescence anisotropy titrations, YbtE was buffer exchanged >100-fold into freshly prepared ITC buffer by repeated concentration and dilution. The sample was filtered and concentration determined by diluting 5 μ l 20-fold in 6.3 M guanidinium-HCl and measuring the A_{280} . A series of 2X YbtE samples was then created from this stock. A 2X stock of fluorescently labeled ArCP was made by diluting concentrated samples to 80 nM in ITC buffer. Final samples were made by mixing 30 μ l 2X YbtE with 30 μ l 2X ArCP in 1.5 ml microcentrifuge tubes. The final concentration of ArCP was 40 nM. The final concentrations of YbtE were 0, 1, 2, 5, 10, 15, 20, 30, 40, 60, 80, 100, 150, 200, and 300 μ M for titrations with apo-ArCP and 0, 0.25, 0.5, 0.75, 1, 2, 4, 6, 8, 10, 15, 20, 30, 40, 60, 80, and 100 μ M for titrations with holo- and SalNH-ArCP. Samples were briefly spun in a table top centrifuge and incubated in the dark for 20

minutes. Samples were then further mixed by pipetting, 40 μ l added to a 384-well plate (Corning #3575), and incubated in the dark for an additional 20 minutes before reading.

Measurements were performed on a Tecan M1000 Infinite PRO plate reader. The G-factor was determined using 1 nM fluorescein in 0.01 M NaOH and the anisotropy readings averaged across five wells. The excitation wavelength was 470 nm with 5 nm bandwidth. The emission wavelength was 518 nm with 8 nm bandwidth. 200 flashes were used for each point. Background fluorescence was averaged across 5 wells containing only ITC buffer. Data was analyzed using Prism.

NMR Titrations of Apo- and Holo-ArCP with YbtE

^{15}N -labeled apo-ArCP was purified as previously described (Chapter 2). ^{15}N -labeled holo-ArCP was generated as previously described (Chapter 2). YbtE was purified as previously described (Chapter 2).

To prepare for titrations, ArCP and YbtE were both buffer exchanged >200-fold into NMR buffer + 0.05% NaN_3 . Samples were then concentrated and filtered. Concentration of YbtE and ArCP was determined as described for ITC experiments. For all experiments, ArCP was held at fixed concentration of 100 μM and the concentration of YbtE was varied. All samples contained 10% D_2O and DSS for referencing.

For holo-ArCP, 8 HN-HSQC were collected in the following order described by the ratio of ArCP:YbtE: 1:0, 1:4, 1:2, 1:1, 1:0.5, 1:0.25, 1:0.375, 1:0.125. Each HN-HSQC was collected with 32 scans and $2048 (^1\text{H}, 16.1095 \text{ ppm at } 4.698 \text{ ppm}) \times 128 (^{15}\text{N}, 30 \text{ ppm at } 117 \text{ ppm})$ complex points for a total acquisition time of 2 hours, 40 minutes. All spectra

were apodized and zero-filled to the nearest power of 2. Spectra were analyzed using CARA.

For apo-ArCP, 6 HN-HSQC spectra were collected in the same manner as for holo-ArCP and in the following order: 1:0, 1:1, 1:0.5, 1:0.25, 1:0.375, 1:0.125. Spectra were analyzed using CARA.

Calculating Solvent Accessible Surface Area

Solvent accessible surface (SAS) area was calculated using the AreaMol program in CCP4 with a probe solvent molecule radius of 1.4 angstroms. For the EntE:EntB complex, SAS was calculated for the full complex (PDB 3RG2) and PDBs generated from 3RG2 in which one of the binding partners was removed. For the PDB generated for EntB from 3RG2, the phosphopantetheine arm was removed from the PDB file.

For analysis of DhbE, SAS was calculated for the full DhbE molecule (PDB 1MDF) and the isolated C-terminal domain, beginning at the conserved lysine in the linker between the sub-domains.

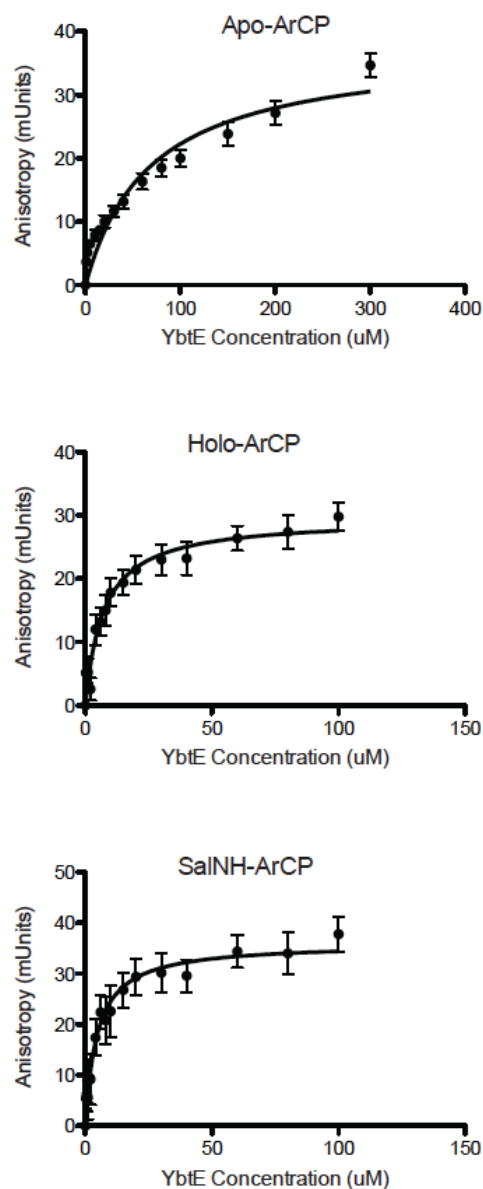


Figure 5.1. Results of fluorescence anisotropy experiments in which the various forms of ArCP were titrated with YbtE. Note the different scale of the x-axis for the titration with apo-ArCP. Fits to a 1:1 specific binding model in Prism 5 gave K_D 's of $68.6 \pm 9.62 \mu\text{M}$ (apo-ArCP), $7.22 \pm 1.34 \mu\text{M}$ (holo-ArCP), and $4.70 \pm 1.10 \mu\text{M}$ (SalNH-ArCP). Each point shows the average \pm standard error for six (apo-ArCP) or five (holo- and SalNH-ArCP) titrations.

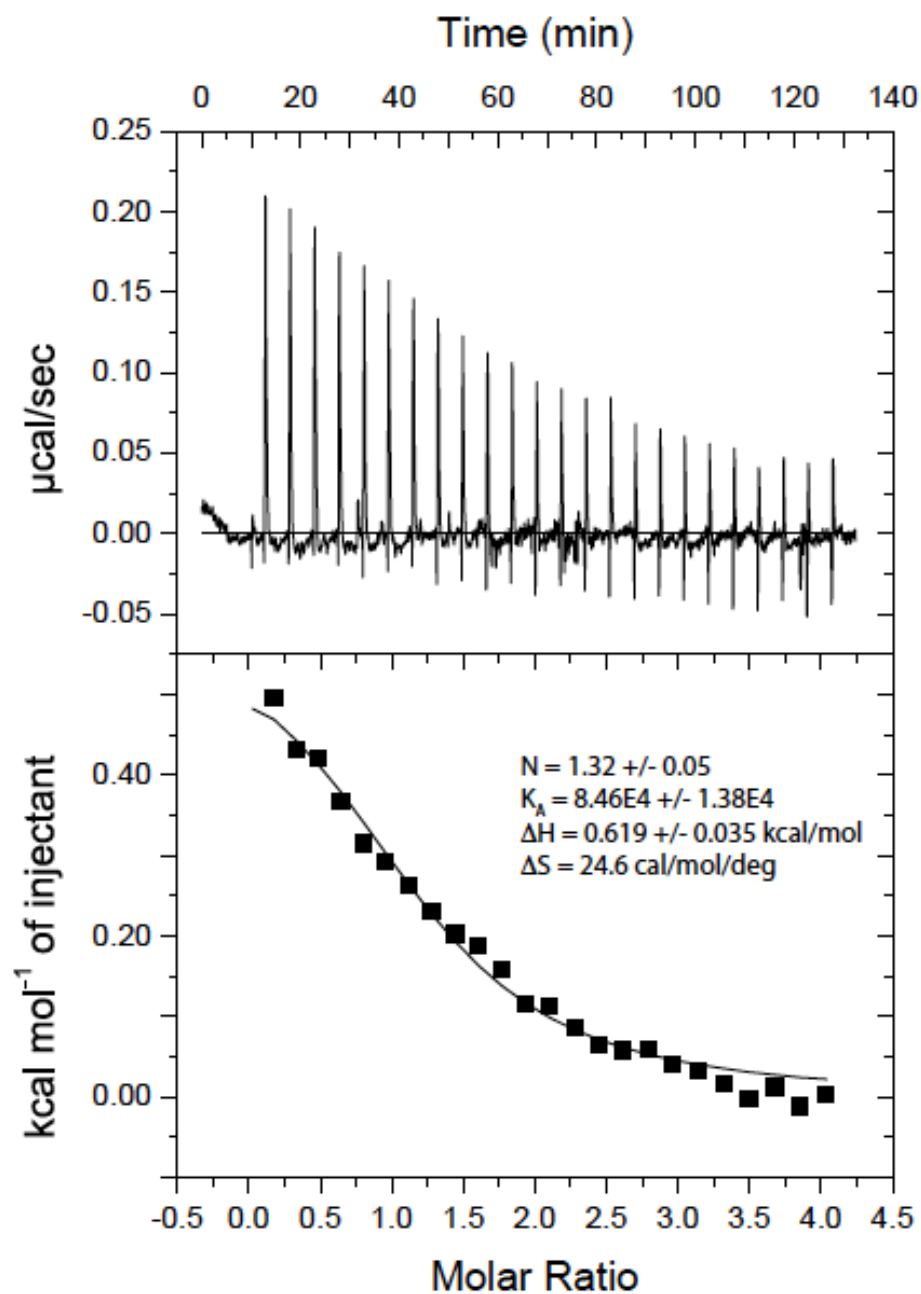


Figure 5-2. Results of the ITC experiment in which YbtE was titrated into apo-ArCP. The top panel shows the baseline corrected raw data. The bottom panel shows the integrated heats fit to a 1:1 model with the stoichiometry, association constant, enthalpy, and entropy of the interaction. The integrated heats of the last four injections were averaged and subtracted from all points prior to fitting in Origin.

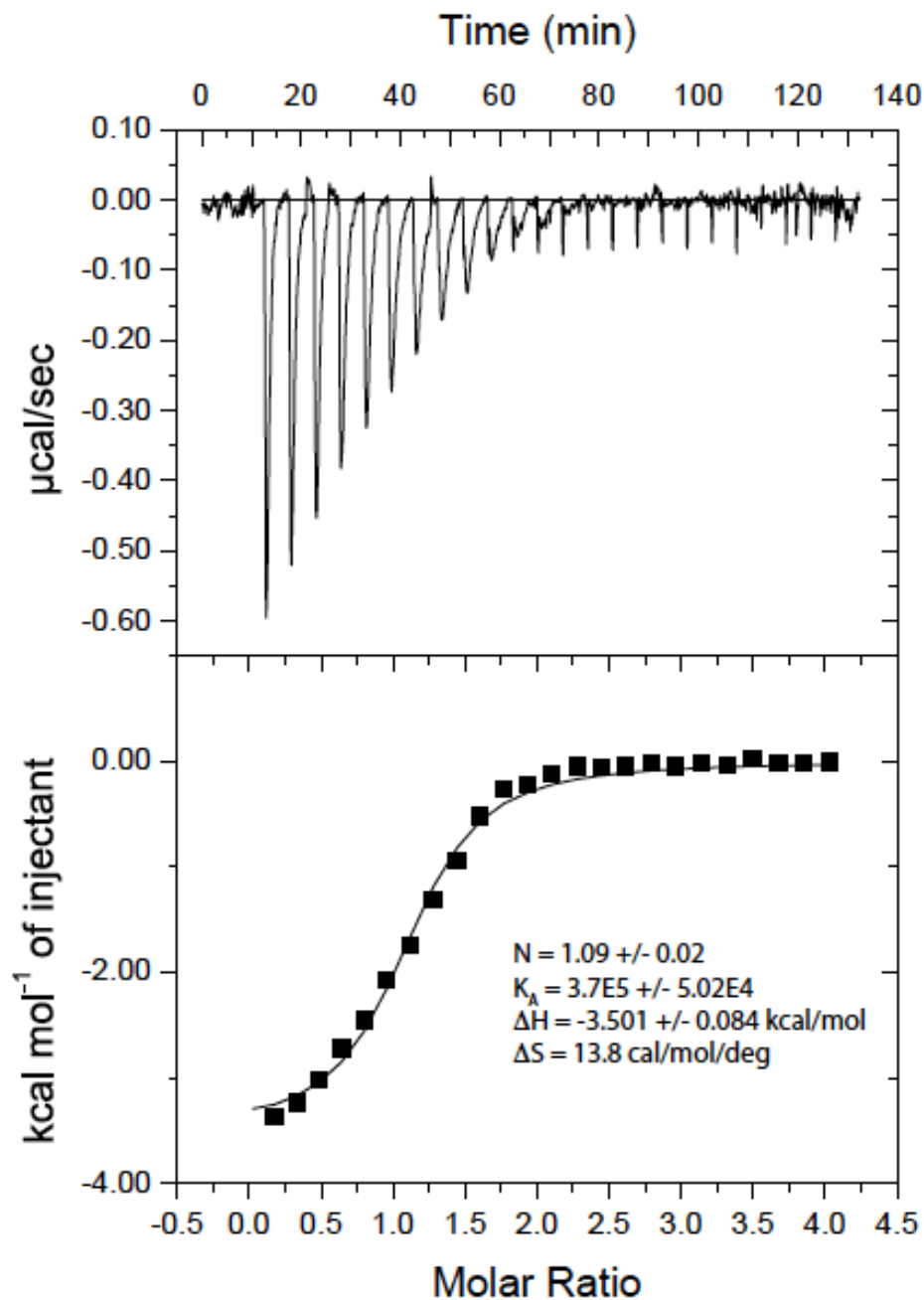


Figure 5-3. Results of the ITC experiment in which YbtE was titrated into holo-ArCP. The top panel shows the baseline corrected raw data. The bottom panel shows the integrated heats fit to a 1:1 model with the stoichiometry, association constant, enthalpy, and entropy of the interaction. The integrated heats of the last four injections were averaged and subtracted from all points prior to fitting in Origin.

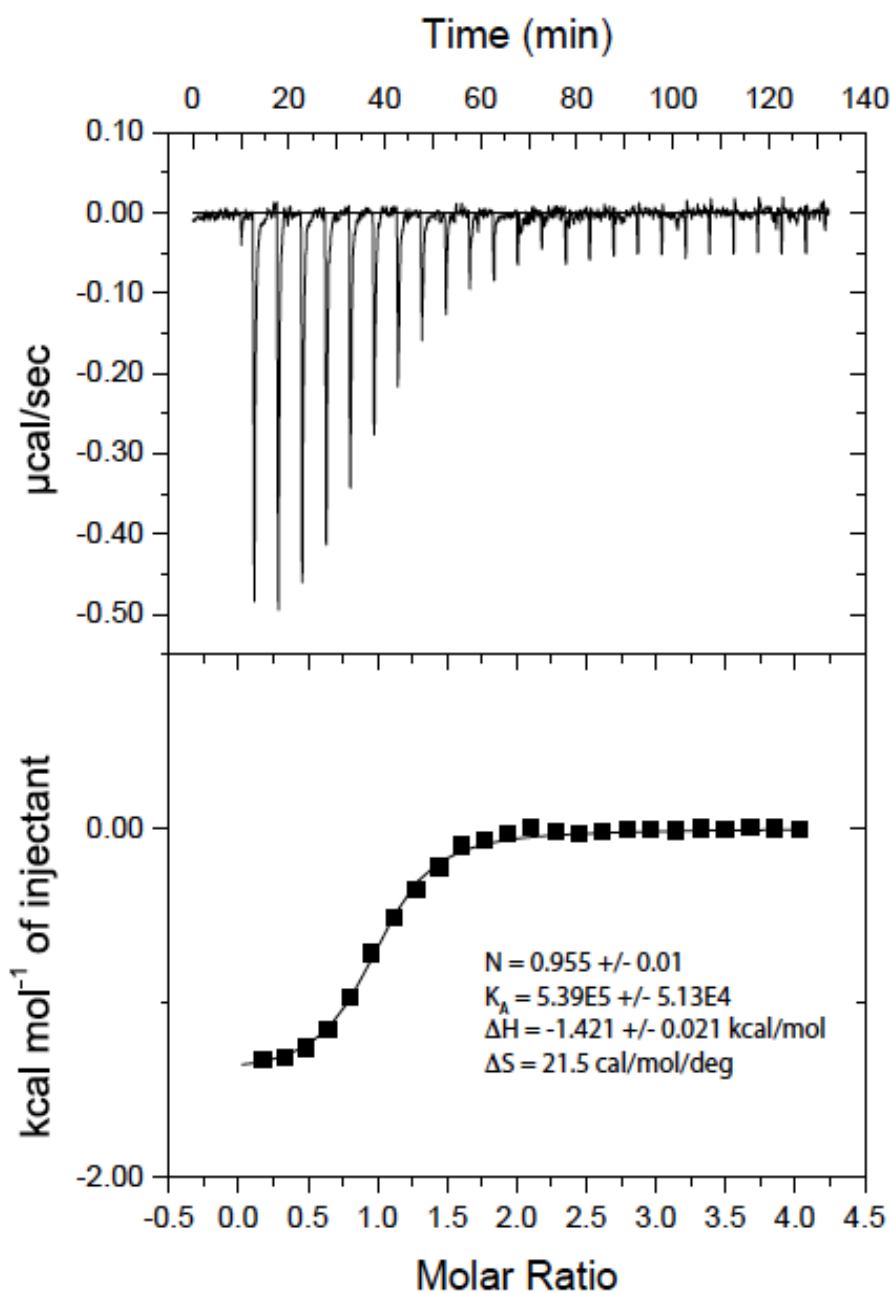


Figure 5-4. Results of the ITC experiment in which YbtE was titrated into SalNH-ArCP. The top panel shows the baseline corrected raw data. The bottom panel shows the integrated heats fit to a 1:1 model with the stoichiometry, association constant, enthalpy, and entropy of the interaction. The integrated heats of the last four injections were averaged and subtracted from all points prior to fitting in Origin.

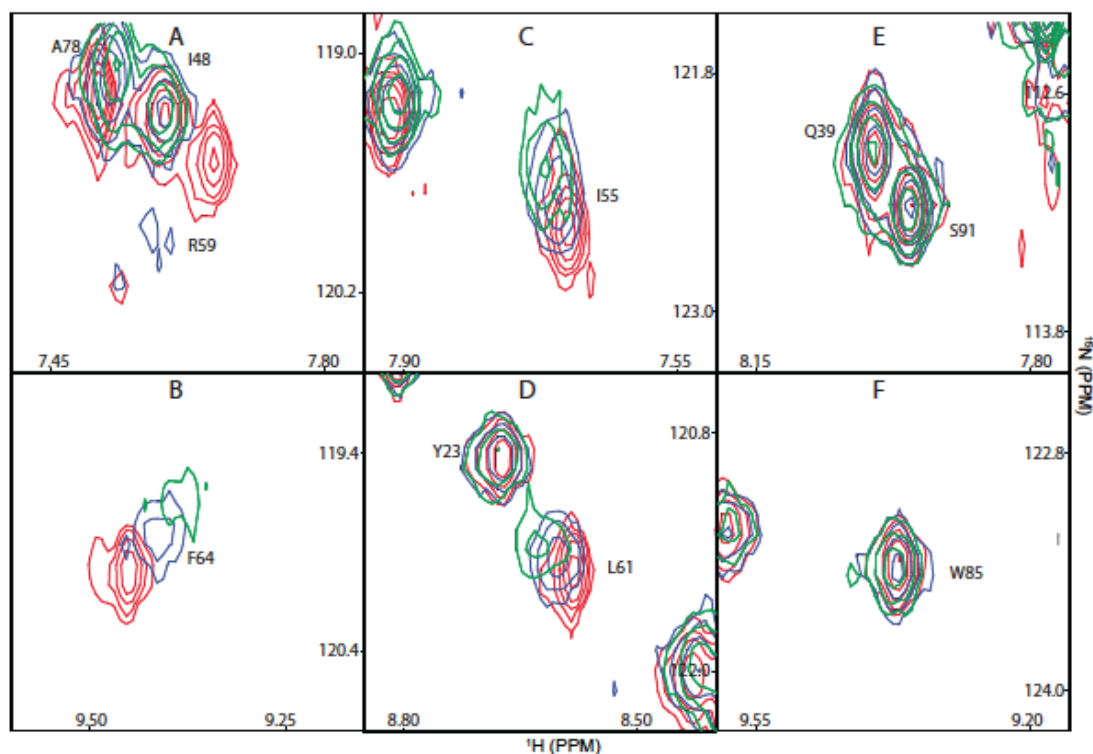


Figure 5-5. Signals of apo-ArCP from HN-HSQC showing a variety of responses to increasing concentrations of YbtE. **Red:** Free apo-ArCP. **Blue:** 1:0.25 apo-ArCP:YbtE. **Green:** 1:0.5 apo-ArCP:YbtE. Panels A and B show two signals that shift dramatically and show a marked decrease in intensity as the concentration of YbtE increases. C and D show two signals whose position changes modestly. Panels E and F show signals for three residues whose appearance is wholly unaffected by the addition of YbtE. The differential responses of these residues indicates a specific binding interaction between apo-ArCP and YbtE and allows us to identify a potential interaction surface.

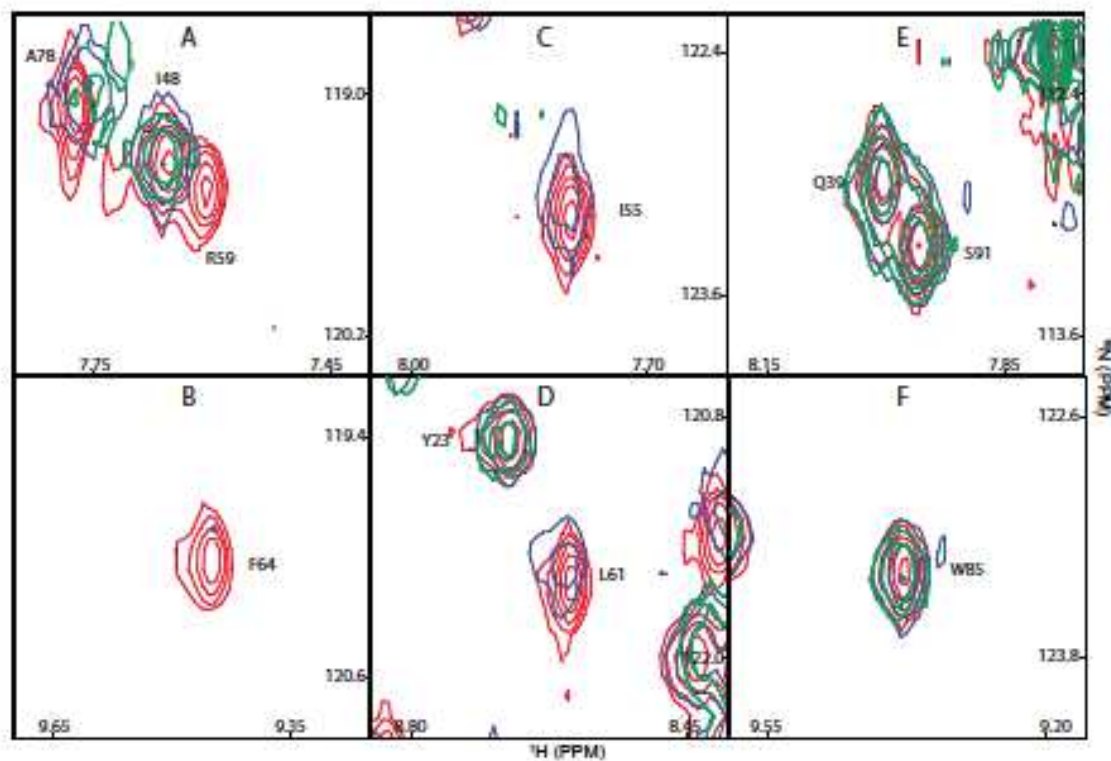


Figure 5-6. Signals of holo-ArCP from HN-HSQC showing a variety of responses to increasing concentrations of YbtE. **Red:** Free holo-ArCP. **Blue:** 1:0.25 holo-ArCP:YbtE. **Green:** 1:0.5 holo-ArCP:YbtE. All panels show the same residues as in Figure 5-5. Panels A and B show two signals that disappear from the spectrum at low concentrations of YbtE. C and D show two signals whose position changes modestly before disappearing from the spectrum. Panels E and F show signals for three residues whose appearance is wholly unaffected by the addition of YbtE. The differential responses of these residues indicates a specific binding interaction between holo-ArCP and YbtE and allows us to identify a potential interaction surface. The response of the signals shown here is indicative of intermediate exchange and indicates a tighter interaction between holo-ArCP and YbtE than apo-ArCP and YbtE.

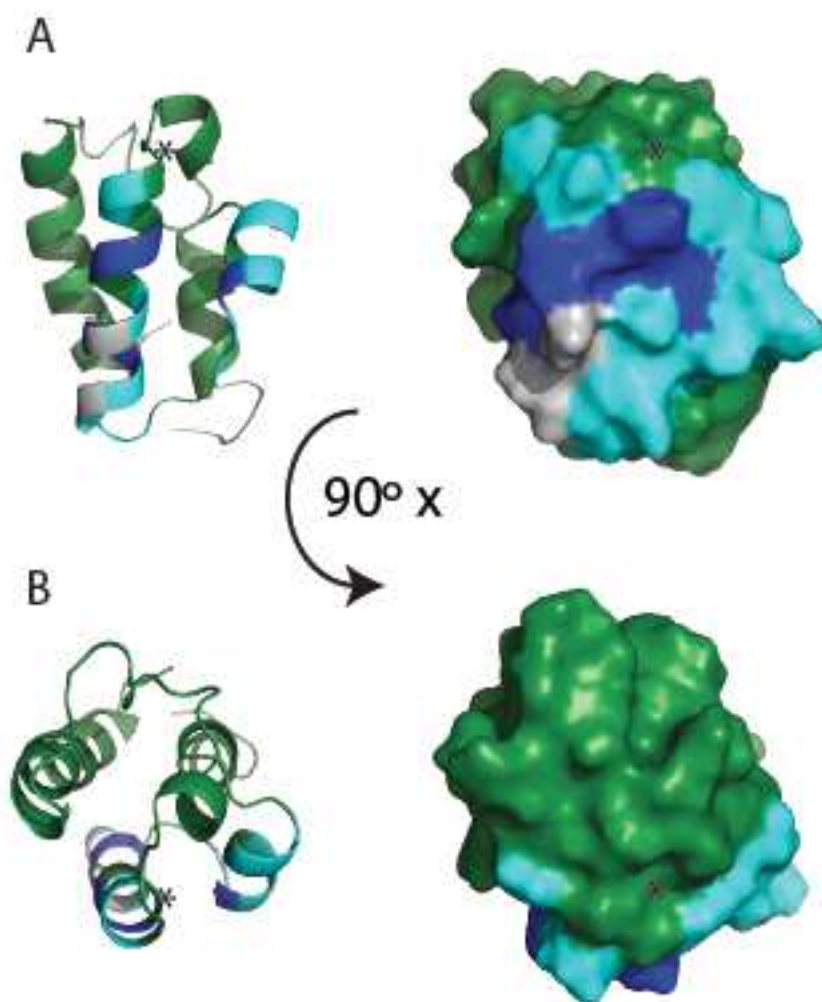
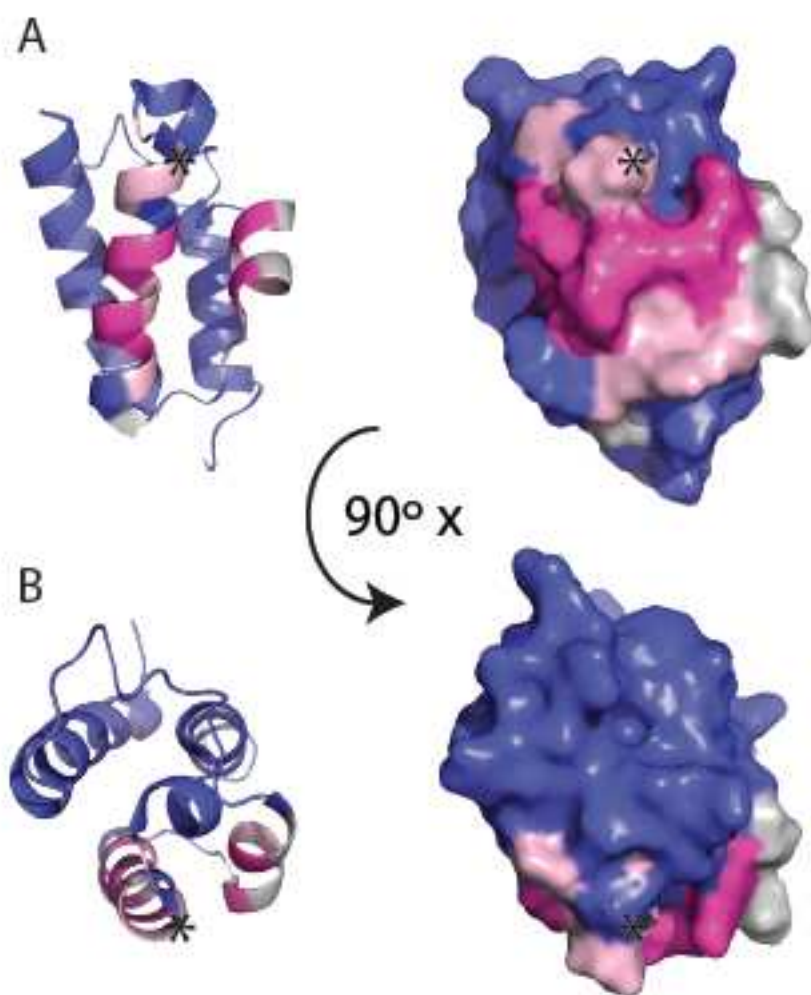


Figure 5.7. Residues from apo-ArCP affected by titration with YbtE. Residues whose signals showed little or no response to the presence of YbtE are colored in green. Residues whose signals passed a threshold chemical shift perturbation are colored by the apo-ArCP:YbtE ratio at which that threshold was passed: Blue: 1:0.125; Cyan: 1:0.25; Grey: 1:0.375. A) Residues likely involved in the interaction cluster along helices 2 and 3, near the PP attachment site (highlighted with an asterisk) and form a contiguous patch along the surface of apo-ArCP. B) No residues on loop 1 were identified as being involved in mediating the interaction between apo-ArCP and YbtE.

Figure 5.8. Residues from holo-ArCP affected by titration with YbtE. Residues whose signals showed little or no response to the presence of YbtE are colored in TV blue. Residues whose signals dropped below a threshold intensity are colored by the holo-ArCP:YbtE ratio at which that threshold was passed: Red: 1:0.25; Pink: 1:0.375; Grey: 1:0.5. A) Residues likely involved in the interaction cluster along helices 2 and 3, near the PP attachment site (highlighted with an asterisk) and form a contiguous patch along the surface of holo-ArCP. B) No residues on loop 1 were identified as being involved in mediating the interaction between holo-ArCP and YbtE. These titrations were done in the absence of substrates for the adenylation reaction (salicylate and ATP) or an adenylation mimic, which may further alter the nature of the interaction. Interaction between loop 1 and YbtE may also be strengthened by substrate loading, which repositions loop 1 via interaction with the tethered substrate. Figure on page 140.



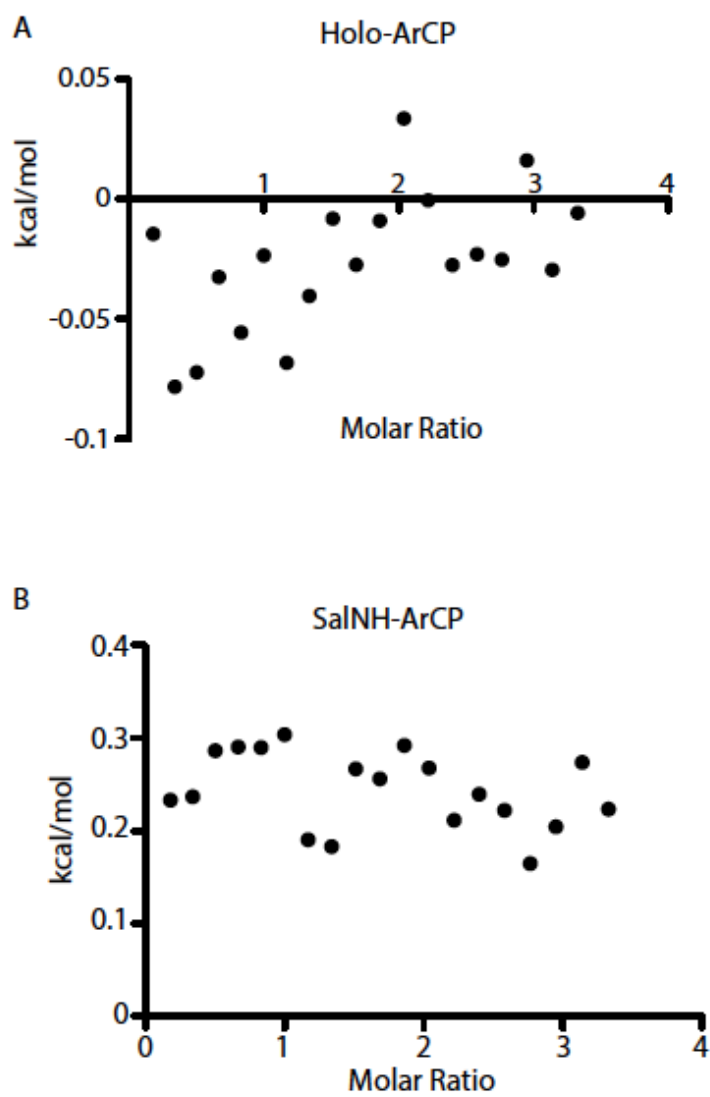


Figure 5-9. Results of ITC experiments with Cy1 and A) holo-ArCP or B) SalNH-ArCP. Integrated heats from blank titrations of Cy1 into buffer and buffer into holo- or SalNH-ArCP were subtracted from the integrated heats from the actual titrations and the resulting heats plotted in Prism 5. Neither experiment shows a pattern consistent with a specific interaction between Cy1 and ArCP under these conditions and with these constructs.

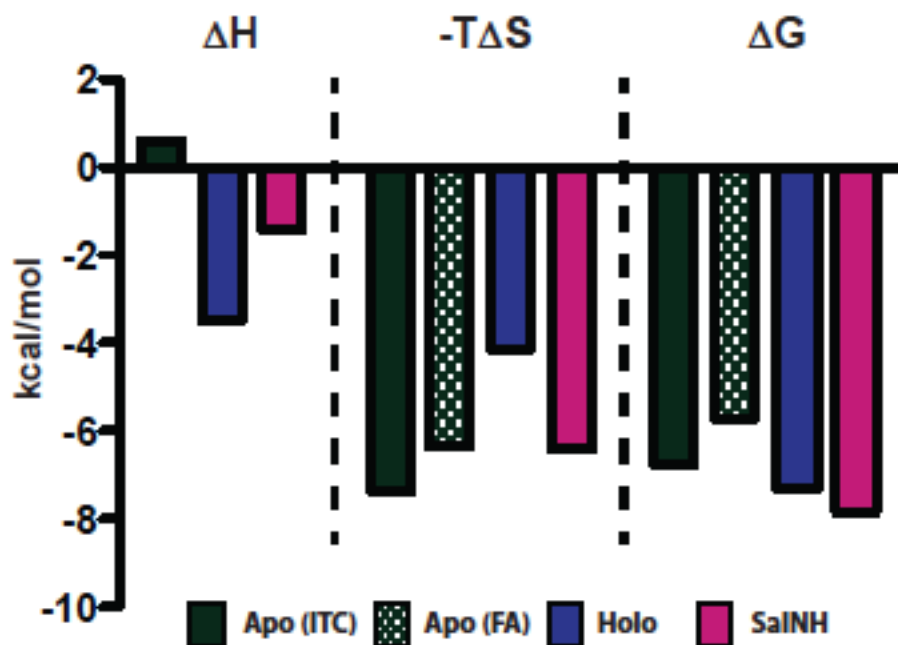


Figure 5-10. Comparison of the thermodynamic parameters for the interaction between YbtE and apo- (Green), holo- (blue), and SalNH-ArCP (pink). The solid green bars show $-T\Delta S$ and ΔG for apo-ArCP based on the ITC results. As the fluorescence anisotropy experiments gave a rather different K_D (68.6 μM (FA) vs. 11.6 μM) and therefore ΔG , this will affect the calculated change in entropy. The checkered green bars show the ΔG as calculated using a K_D of 68.6 μM and $-T\Delta S$ based on this ΔG and the ΔH obtained from the ITC experiment. The difference between these two calculations does not substantially change the entropy:enthalpy balance and therefore does not alter the interpretation. This figure clearly shows that covalent modifications to ArCP alter the enthalpy/entropy balance for the interaction with YbtE.

	Apo (FA)	Apo (ITC)	Holo (FA)	Holo (ITC)	SalNH (FA)	SalNH (ITC)
K_D (μ M)	68.6 ± 9.62	11.8 ± 1.92	7.22 ± 1.34	2.70 ± 0.37	4.70 ± 1.10	1.86 ± 0.18
ΔG (kcal/mol)	-5.71 ± 0.80	-6.76 ± 1.10	-7.06 ± 1.31	-7.64 ± 1.05	-7.31 ± 1.71	-7.87 ± 0.762
ΔH (kcal/mol)	0.619 ± 0.035	0.619 ± 0.035	-3.501 ± 0.084	-3.501 ± 0.084	-1.421 ± 0.021	-1.421 ± 0.021
$-T\Delta S$ (kcal/mol)	-6.33 ± 0.80	-7.38 ± 1.10	-3.56 ± 1.31	-4.14 ± 1.05	-5.89 ± 1.71	-6.45 ± 0.76

Table 5-1. List of binding constants and thermodynamic parameters determined by fluorescence anisotropy (FA) and isothermal titration calorimetry (ITC).

Bibliography

1. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBiochem* **3**, 490–504 (2002).
2. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A. & Holak, T. A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407–18 (2000).
3. Lai, J. R., Koglin, A. & Walsh, C. T. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869–79 (2006).
4. Haslinger, K., Redfield, C. & Cryle, M. J. Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins* **83**, 711–21 (2015).
5. Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–6 (2006).
6. Goodrich, A. C., Harden, B. J. & Frueh, D. P. Solution Structure of a Nonribosomal Peptide Synthetase Carrier Protein Loaded with Its Substrate Reveals Transient, Well-Defined Contacts. *J. Am. Chem. Soc.* **137**, 12100–9 (2015).
7. Jaremko, M. J., Lee, D. J., Opella, S. J. & Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *J. Am. Chem. Soc.* **137**, 11546–9 (2015).
8. Goodrich, A. C. & Frueh, D. P. A nuclear magnetic resonance method for probing molecular influences of substrate loading in nonribosomal peptide synthetase carrier proteins. *Biochemistry* **54**, 1154–6 (2015).
9. Gulick, A. M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–27 (2009).
10. Sundlov, J. a, Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188–98 (2012).
11. Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. (2012).
12. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
13. Reimer, J. M., Aloise, M. N., Harrison, P. M. & Martin Schmeing, T. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242 (2016).
14. Frueh, D. P. *et al.* Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903–906 (2008).
15. Koglin, A. *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* **454**, 907–11 (2008).
16. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–10 (2004).
17. Stachelhaus, T., Schneider, A. & Marahiel, M. A. Rational design of peptide

- antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69–72 (1995).
18. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–63 (2008).
 19. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **425**, 3137–50 (2013).
 20. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–4 (2014).
 21. Nguyen, C. *et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427–31 (2014).
 22. Gehring, A. M., Mori, I., Perry, R. D. & Walsh, C. T. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637–50 (1998).
 23. Suo, Z., Walsh, C. T. & Miller, D. A. Tandem Heterocyclization Activity of the Multidomain 230 kDa HMWP2 Subunit of *Yersinia pestis* Yersiniabactin Synthetase: Interaction of the 1–1382 and 1383–2035 Fragments †. *Biochemistry* **38**, 14023–14035 (1999).
 24. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14083–8 (2002).
 25. Sattler, M. & Schleucher, J. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
 26. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65 (2000).
 27. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
 28. Wishart, D. & Sykes, B. The ¹³C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J. Biomol. NMR* **4**, (1994).
 29. Mayo, K. H. & Prestegard, J. H. Acyl carrier protein from *E. coli*. Structural characterization of short-chain acylated acyl carrier proteins by NMR. *Biochemistry* **24**, 7834–7838 (1985).
 30. Roujeinikova, A. *et al.* X-Ray Crystallographic Studies on Butyryl-ACP Reveal Flexibility of the Structure around a Putative Acyl Chain Binding Site. *Structure* **10**, 825–835 (2002).
 31. Płoskoń, E. *et al.* Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis. *Chem. Biol.* **17**, 776–85 (2010).
 32. Evans, S. E. *et al.* Probing the Interactions of early polyketide intermediates with the Actinorhodin ACP from *S. coelicolor* A3(2). *J. Mol. Biol.* **389**, 511–28 (2009).
 33. Miller, D. A. & Walsh, C. T. Yersiniabactin Synthetase: Probing the Recognition

- of Carrier Protein Domains by the Catalytic Heterocyclization Domains, Cy1 and Cy2, in the Chain-Initiating HMWP2 Subunit †. *Biochemistry* **40**, 5313–5321 (2001).
34. Marshall, C. G., Burkart, M. D., Keating, T. A. & Walsh, C. T. Heterocycle Formation in Vibriobactin Biosynthesis: Alternative Substrate Utilization and Identification of a Condensed Intermediate †. *Biochemistry* **40**, 10655–10663 (2001).
 35. Ducat, T., Declerck, N., Gostan, T., Kochoyan, M. & Déméné, H. Rapid determination of protein solubility and stability conditions for NMR studies using incomplete factorial design. *J. Biomol. NMR* **34**, 137–51 (2006).
 36. Worthington, A. S. & Burkart, M. D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **4**, 44–6 (2006).
 37. Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal. Biochem.* **324**, 100–5 (2004).
 38. Sieber, S. A., Walsh, C. T. & Marahiel, M. A. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J. Am. Chem. Soc.* **125**, 10862–6 (2003).
 39. Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. & Flugel, R. S. Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* **1**, 309–15 (1997).
 40. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–9 (1999).
 41. Khosla, C. Harnessing the Biosynthetic Potential of Modular Polyketide Synthases. *Chem. Rev.* **97**, 2577–2590 (1997).
 42. Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science* **311**, 1263–7 (2006).
 43. Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254–61 (2007).
 44. Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288–90 (2007).
 45. Khosla, C. Structures and mechanisms of polyketide synthases. *J. Org. Chem.* **74**, 6416–20 (2009).
 46. Khosla, C., Kapur, S. & Cane, D. E. Revisiting the modularity of modular polyketide synthases. *Curr. Opin. Chem. Biol.* **13**, 135–43 (2009).
 47. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–7 (2014).
 48. Smith, J. L., Skiniotis, G. & Sherman, D. H. Architecture of the polyketide synthase module: surprises from electron cryo-microscopy. *Curr. Opin. Struct. Biol.* **31**, 9–19 (2015).
 49. Finking, R. & Marahiel, M. A. Biosynthesis of Nonribosomal Peptides 1. *Annu. Rev. Microbiol.* **58**, 453–488 (2004).
 50. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **15**, 781–792 (2007).

51. Lambalot, R. H. *et al.* A new enzyme superfamily — the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
52. Yonus, H. *et al.* Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J. Biol. Chem.* **283**, 32484–91 (2008).
53. Strieker, M., Tanović, A. & Marahiel, M. a. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–40 (2010).
54. Drake, E. J., Nicolai, D. A. & Gulick, A. M. Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem. Biol.* **13**, 409–19 (2006).
55. Allen, C. L. & Gulick, A. M. Structural and bioinformatic characterization of an *Acinetobacter baumannii* type II carrier protein. *Acta Crystallogr. D. Biol. Crystallogr.* **70**, 1718–25 (2014).
56. Haslinger, K. *et al.* The structure of a transient complex of a nonribosomal peptide synthetase and a cytochrome P450 monooxygenase. *Angew. Chem. Int. Ed. Engl.* **53**, 8518–22 (2014).
57. Liu, Y., Zheng, T. & Bruner, S. D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–8 (2011).
58. Lohman, J. R. *et al.* The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. *Proteins* **82**, 1210–8 (2014).
59. Gehring, A. M. *et al.* Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573–586 (1998).
60. Pelludat, C., Rakin, A., Jacobi, C. A., Schubert, S. & Heesemann, J. The Yersiniabactin Biosynthetic Gene Cluster of *Yersinia enterocolitica*: Organization and Siderophore-Dependent Regulation. *J. Bacteriol.* **180**, 538–546 (1998).
61. Keating, T. A., Miller, D. A. & Walsh, C. T. Expression, Purification, and Characterization of HMWP2, a 229 kDa, Six Domain Protein Subunit of Yersiniabactin Synthetase †. *Biochemistry* **39**, 4729–4739 (2000).
62. Keating, T. A., Suo, Z., Ehmann, D. E. & Walsh, C. T. Selectivity of the Yersiniabactin Synthetase Adenylation Domain in the Two-Step Process of Amino Acid Activation and Transfer to a Holo-Carrier Protein Domain †. *Biochemistry* **39**, 2297–2306 (2000).
63. Suo, Z., Tseng, C. C. & Walsh, C. T. Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. *Proc. Natl. Acad. Sci.* **98**, 99–104 (2001).
64. Miller, D. A., Luo, L., Hillson, N., Keating, T. A. & Walsh, C. T. Yersiniabactin Synthetase : A Four-Protein Assembly Line Producing the Nonribosomal Peptide / Polyketide Hybrid Siderophore of *Yersinia pestis*. **9**, 333–344 (2002).
65. Holak, T. A., Kearsley, S. K., Kim, Y. & Prestegard, J. H. Three-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations. *Biochemistry* **27**, 6135–6142 (1988).
66. Crump, M. P. *et al.* Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **36**, 6000–8 (1997).

67. Parris, K. D. *et al.* Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883–895 (2000).
68. Tufar, P. *et al.* Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chem. Biol.* **21**, 552–62 (2014).
69. Zhou, Z., Lai, J. R. & Walsh, C. T. Directed evolution of aryl carrier proteins in the enterobactin synthetase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11621–6 (2007).
70. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
71. Mandel, A. M., Akke, M. & Palmer, III, A. G. Backbone Dynamics of *Escherichia coli* Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **246**, 144–163 (1995).
72. Berlin, K., Longhini, A., Dayie, T. K. & Fushman, D. Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J. Biomol. NMR* **57**, 333–52 (2013).
73. Walker, O., Varadan, R. & Fushman, D. Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from (15)N relaxation data using computer program ROTDIF. *J. Magn. Reson.* **168**, 336–45 (2004).
74. Li, Q., Khosla, C., Puglisi, J. D. & Liu, C. W. Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. *Biochemistry* **42**, 4648–57 (2003).
75. Sharma, A. K., Sharma, S. K., Surolia, A., Surolia, N. & Sarma, S. P. Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from *Plasmodium falciparum* provides insight into the mechanism of activation of ACPs. *Biochemistry* **45**, 6904–16 (2006).
76. Kim, Y., Kovrigin, E. L. & Eletr, Z. NMR studies of *Escherichia coli* acyl carrier protein: dynamic and structural differences of the apo- and holo-forms. *Biochem. Biophys. Res. Commun.* **341**, 776–83 (2006).
77. Kim, Y. & Prestegard, J. . Measurement of vicinal couplings from cross peaks in COSY spectra. *J. Magn. Reson.* **84**, 9–13 (1989).
78. Chan, D. I. *et al.* NMR solution structure and biophysical characterization of *Vibrio harveyi* acyl carrier protein A75H: effects of divalent metal ions. *J. Biol. Chem.* **285**, 30558–66 (2010).
79. Zornetzer, G. A., Tanem, J., Fox, B. G. & Markley, J. L. The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. *Biochemistry* **49**, 470–7 (2010).
80. Lim, J. *et al.* Solution structures of the acyl carrier protein domain from the highly reducing type I iterative polyketide synthase CalE8. *PLoS One* **6**, e20549 (2011).
81. Lim, J., Xiao, T., Fan, J. & Yang, D. An Off-Pathway Folding Intermediate of an Acyl Carrier Protein Domain Coexists with the Folded and Unfolded States under Native Conditions. *Angew. Chemie* **126**, 2390–2393 (2014).
82. Findlow, S. C., Winsor, C., Simpson, T. J., Crosby, J. & Crump, M. P. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **42**, 8423–33 (2003).
83. Lim, J. *et al.* Rigidifying acyl carrier protein domain in iterative type I PKS CalE8

- does not affect its function. *Biophys. J.* **103**, 1037–44 (2012).
84. Wennerström, H. Nuclear magnetic relaxation induced by chemical exchange. *Mol. Phys.* **24**, 69–80 (1972).
 85. Balaram, P., Bothner-By, A. A. & Breslow, E. Localization of tyrosine at the binding site of neurophysin II by negative nuclear Overhauser effects. *J. Am. Chem. Soc.* **94**, 4017–4018 (1972).
 86. Balaram, P., Bothner-By, A. A. & Dadok, J. Negative nuclear Overhauser effects as probes of macromolecular structure. *J. Am. Chem. Soc.* **94**, 4015–4017 (1972).
 87. Clore, G. . & Gronenborn, A. . Theory and applications of the transferred nuclear overhauser effect to the study of the conformations of small ligands bound to proteins. *J. Magn. Reson.* **48**, 402–417 (1982).
 88. Lippens, R. M., Cerf, C. & Hallenga, K. Theory and experimental results of transfer-NOE experiments. 1. The influence of the off rate versus cross-relaxation rates. *J. Magn. Reson.* **99**, 268–281 (1992).
 89. Nageswara Rao, B. D. *Nuclear Magnetic Resonance Part A: Spectral Techniques and Dynamics. Methods Enzymol.* **176**, (Elsevier, 1989).
 90. Nirmala, N. ., Lippens, G. . & Hallenga, K. Theory and experimental results of transfer NOE experiments. II. The influence of residual mobility and relaxation centers inside the protein on the size of transfer NOEs. *J. Magn. Reson.* **100**, 25–42 (1992).
 91. Yangmee, K., Ohlrogge, J. B. & Prestegard, J. H. Motional effects on NMR structural data. *Biochem. Pharmacol.* **40**, 7–13 (1990).
 92. Oswood, M. C., Kim, Y., Ohlrogge, J. B. & Prestegard, J. H. Structural homology of spinach acyl carrier protein and Escherichia coli acyl carrier protein based on NMR data. *Proteins* **27**, 131–43 (1997).
 93. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16**, 2093–107 (2007).
 94. Lai, J. R., Fischbach, M. A., Liu, D. R. & Walsh, C. T. A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5314–9 (2006).
 95. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, (1995).
 96. Keller, R. *The computer aided resonance assignment tutorial. Goldau, Switz. Cantina Verlag* (2004).
 97. DeLano, W. L. The PyMOL Molecular Graphics System. *Schrödinger LLC www.pymol.org Version 1.*, <http://www.pymol.org> (2002).
 98. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–41 (2013).
 99. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–78 (2004).
 100. Nederveen, A. J. *et al.* RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662–72 (2005).

101. Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778–95 (2007).
102. Laskowski, R., Rullmann, J. A., MacArthur, M., Kaptein, R. & Thornton, J. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, (1996).
103. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–41 (2001).
104. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–5 (2007).
105. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–7 (2004).
106. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
107. Koglin, A. & Walsh, C. T. Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep.* **26**, 987–1000 (2009).
108. Jaremko, M. J., Lee, D. J., Opella, S. J. & Burkart, M. D. Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. *J. Am. Chem. Soc.* **137**, 11546–9 (2015).
109. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–6 (2002).
110. Zettler, J. & Mootz, H. D. Biochemical evidence for conformational changes in the cross-talk between adenylation and peptidyl-carrier protein domains of nonribosomal peptide synthetases. *FEBS J.* **277**, 1159–71 (2010).
111. Brady, G. P. & Sharp, K. A. Entropy in protein folding and in protein-protein interactions. *Curr. Opin. Struct. Biol.* **7**, 215–21 (1997).
112. Hilser, V. J., García-Moreno E, B., Oas, T. G., Kapp, G. & Whitten, S. T. A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* **106**, 1545–58 (2006).
113. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G. & Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* **6**, 352–8 (2010).
114. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G. & Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* **6**, 352–8 (2010).
115. May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12120–5 (2002).
116. Marshall, C. G., Burkart, M. D., Meray, R. K. & Walsh, C. T. Carrier protein recognition in siderophore-producing nonribosomal peptide synthetases. *Biochemistry* **41**, 8429–37 (2002).

PLACE OF BIRTH

Columbus, OH

EDUCATION

Johns Hopkins University School of Medicine, Baltimore, MD

Ph. D. Focus: Biochemistry, Cellular and Molecular Biology

The Ohio State University, Columbus, OH

June, 2010

B.S. Focus: Biochemistry, With Honors and Research Distinction in Biochemistry, Cum Laude

RESEARCH EXPERIENCE

Johns Hopkins University School of Medicine, Baltimore, MD

August 2010-Present

Graduate Research

The Ohio State University, Columbus, OH

March 2008-August 2010

Undergraduate Research

Working in the lab of Professor Karin Musier-Forsyth to complete my Honors Thesis, I used fluorescence anisotropy, FRET, and electrophoretic mobility shift assays to investigate the role of the zinc fingers of the HIV-1 Gag protein in its nucleic acid chaperone activity and how this activity regulates the viral life cycle.

TEACHING EXPERIENCE

Teaching Assistant, Molecular Biology and Genomics, Fall 2012, 2013, and 2014: Responsible for developing homework assignments and exams as well as grading assignments. Held office hours for students needing additional help and a review session focusing on common techniques used in molecular biology.

Teaching Assistant, Macromolecular Structure and Analysis, Fall 2013: Headed weekly sessions with first year graduate students to review topics such as DNA, RNA, and protein structure and methods for their analysis, such as x-ray crystallography, NMR, and mass spectrometry.

Teaching Assistant, Biochemistry 521, Fall 2009, Winter 2010: Supervised undergraduate biochemistry students in lab while performing experiments to learn the basics of pipetting, spectrophotometry, chromatography, protein purification, and cloning.

Peer Mentor, Chemistry 123 Research Experience Enhances Learning Program, Spring 2009: Responsible for guiding two groups of four first year undergraduates through the process of designing, performing, analyzing, and presenting a research project meant to further their understanding of the material presented in class and expose them to research.

AWARDS AND HONORS

American Heart Association Mid-Atlantic Affiliate Predoctoral Fellowship, 07/2014-Present

Arts and Sciences Certificate for Excellence in Scholarship, The Ohio State University, May 2010

First Place-Best Poster, Biological Sciences Division, Denman Undergraduate Research Forum, The Ohio State University, May 2010

President's Salute to Undergraduate Achievement, The Ohio State University, April 2010

Arts and Sciences Undergraduate Research Scholarship, The Ohio State University, 09/2009-06/2010

College of Biological Sciences Mayer Award, The Ohio State University, 2009- Not accepted

Department of Biochemistry Summer Undergraduate Research Program Award, The Ohio State University, 2009- Not accepted

National Science Foundation Research Experience for Undergraduates Fellowship, 06/2009-08/2009

Arts and Sciences Undergraduate Research Scholarship, The Ohio State University, 01/2009-06/2009

University Scholarship, The Ohio State University, 2006-2010

POSTER PRESENTATIONS

NMR Studies of the Structural Influence of Phosphopantetheinylation in Nonribosomal Peptide Synthetase Carrier Proteins and Impact on Binding Affinities

Andrew C. Goodrich, Dominique P. Frueh

Protein Society Symposium, Barcelona, Spain, July 2015

Biochemical State of the Aryl Carrier Protein Directs Sequential Protein-Protein Interactions in the Yersiniabactin Synthetase System

Andrew C. Goodrich, Dominique P. Frueh

Biophysical Society Annual Meeting, Baltimore, MD, February 2015

Molecular Interactions During Substrate Loading in the Aryl Carrier Protein from Yersinibactin Synthetase

Andrew C. Goodrich, Dominique P. Frueh

ICMRBS, Dallas, TX, August 2014

Role of Zinc Fingers of HIV-1 Gag in Nucleic Acid Binding and Chaperone Function.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Denman Undergraduate Research Forum, Columbus, OH, May 2010.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Meek Undergraduate Poster Session, Columbus, OH, May 2010.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

RNA Rustbelt Meeting, Mt. Sterling, OH, October 2009.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Ohio Science and Engineering Alliance Student Research Forum, Columbus, OH, August 2009.

A Short-Term Investigation of Combined Sewer Overflows (CSOs) and their Impact on Water Quality in the Olentangy River.

Andrew C. Goodrich, Daniel Hannah, Ted M. Clark.

Central Regional Meeting of the American Chemical Society, Cleveland, OH, May 2009.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Central Regional Meeting of the American Chemical Society, Cleveland, OH, May 2009.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Denman Undergraduate Research Forum, Columbus, OH, May 2009.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Meek Undergraduate Poster Session, Columbus, OH, May 2009.

Probing the Role of HIV-1 Gag in Regulation of Reverse Transcription.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Biological, Mathematical, and Physical Sciences Undergraduate Research Forum, Columbus, OH, April 2009.

Regulation of Reverse Transcription by HIV-1 Gag.

Andrew C. Goodrich, Christopher P. Jones, Karin Musier-Forsyth.

Rustbelt RNA Meeting, Mt. Sterling, OH, October 2008.

PUBLICATIONS

Goodrich, A.C., Harden, B.J., and Frueh, D.P. (2015) Solution Structure of a Nonribosomal Peptide Synthetase Carrier Protein Loaded with Its Substrate Reveals Transient, Well-Defined Contacts. *Journal of the American Chemical Society*, 137, 12100-12109. doi: 10.1021/jacs.5b07772

Goodrich, A.C. & Frueh, D.P. (2015) A Nuclear Magnetic Resonance Method for Probing Molecular Influences of Substrate Loading in Nonribosomal Peptide Synthetase Carrier Proteins. *Biochemistry*, 54, 1154-6. doi: 10.1021/bi501443r.

Goodrich, A.C., Nichols, S.R., & Frueh, D.P. (2014) Resonance Assignments and Structure Determination of Large and Challenging Proteins. *Encyclopedia of Magnetic Resonance*, 3. doi: 10.1002/9780470034590.emrstm1359.

Frueh, D. P., Goodrich, A. C., Mishra, S. H., & Nichols, S. R. (2013) NMR methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology*, 23, 734-9. doi:10.1016/j.sbi.2013.06.016.