Title

Essential Role of Loop Dynamics in Type II NRPS Biomolecular Recognition



Journal

ACS Chemical Biology, 17(10)

ISSN

1554-8929

Authors

Corpuz, Joshua C

Patel, Ashay

Davis, Tony D

et al.

Publication Date

2022-10-21

DOI

10.1021/acschembio.2c00523

Peer reviewed

HHS Public Access

Author manuscript

*ACS Chem Biol*. Author manuscript; available in PMC 2023 October 21.

Published in final edited form as:

*ACS Chem Biol*. 2022 October 21; 17(10): 2890–2898. doi:10.1021/acschembio.2c00523.

# The essential role of loop dynamics in type II NRPS biomolecular recognition

**Joshua C. Corpuz**[a], **Ashay Patel**[a], **Tony D. Davis**[a], **Larissa M. Podust**[b], **J. Andrew McCammon**[a,c], **Michael D. Burkart**[a,*]

[a]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0332, USA

[b]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0332, USA

[c]Department of Pharmacology, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093, USA

## Abstract

Non-ribosomal peptides play a critical role in the clinic as therapeutic agents. To access greater chemical diversity of therapeutics, the non-ribosomal peptide synthetase (NRPS) has been targeted for engineering through combinatorial biosynthesis; however, this has been met with limited success in part due to the lack of proper protein-protein interactions between non-cognate proteins that enable non-ribosomal peptide biosynthesis. Herein, we report our use of chemical biology to enable X-ray crystallography, molecular dynamics (MD) simulations, and biochemical studies in the elucidation of binding specificities between peptidyl carrier proteins (PCPs) and adenylation (A) domains. Specifically, we determined X-ray crystal structures of a type II PCP crosslinked to its cognate A domain, PigG and PigI, and of PigG crosslinked to a non-cognate PigI homolog, PltF. The crosslinked PCP-A domain structures possess large protein-protein interfaces that predominantly feature hydrophobic interactions, with specific electrostatic interactions that orient the substrate for active site delivery. MD simulations of the PCP-A domain complexes and unbound PCP structures provide a dynamical evaluation of the transient interactions formed at PCP-A domain interfaces, which confirms the previously hypothesized role of a PCP loop as a crucial recognition element. Finally, we demonstrate that the interfacial interactions at the PCP loop 1 region can be modified to control PCP binding specificity through gain-of-function mutations. This work suggests that conformational preferences and loop dynamism account for improved shape complementary in the PCP-A domain interaction. Ultimately, these studies show how crystallographic, biochemical, and computational methods can be used to rationally re-engineer NRPSs for non-cognate interactions.

*Corresponding author: Michael D. Burkart, mburkart@ucsd.edu.

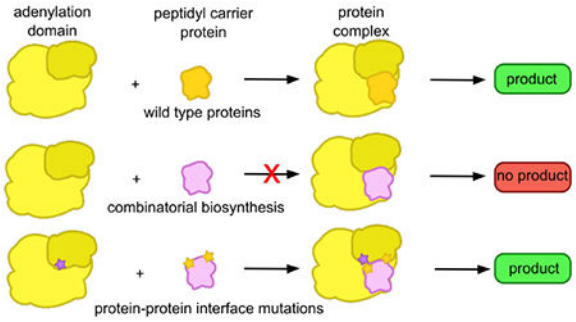
Supporting information

The following materials and methods are available free of charge via the internet at https://pubs.acs.org/: protein expression and purification; chemical probe synthesis; crystallization and structure determination of PCP-A domain complexes; mutagenesis and assays; MD simulations and analysis; and protein-protein interface design.

Competing interests

The authors have declared no conflicts of interest.

## Graphical Abstract




## Introduction

To date, nearly half of approved drugs are directly derived from or based upon natural products.[1] Natural product biosynthesis yields a vast library of structurally complex, highly functionalized, stereodense compounds. Among these compounds, non-ribosomal peptides (NRPs) have established a privileged status as drug-like molecules.[2,3] Consequently, considerable effort has been expended in the studies of engineering non-ribosomal peptide synthetases (NRPSs), with the expectation for novel therapeutics and the associated biosynthetic routes to produce them.[4] One such promising example includes identification of recombination boundaries in megasynthase (type I) NRPSs for the combinatorial biosynthesis of NRPSs.[27,28] Though an exciting conceptual opportunity, directed biosynthesis in these pathways has proven elusive for the rapid and scalable production of non-native natural products. These difficulties can largely be attributed to our limited understanding of the molecular basis of natural product biosynthesis; namely the sensitive and precise interplay of the key elements and proteins, that govern natural product biosynthesis.[5] The work reported herein represents one promising approach toward rational engineering of NRPS pathways.

How standalone (type II) NRPS enzymes associate with each other has been a central theme in our biosynthetic studies of proline-derived natural products. While naturally, these metabolites have been found to possess cellular roles of signaling molecules, pigments, and toxins, these NRPs also possess antitumor, antibacterial, and immunosuppressant therapeutic properties.[6] Clinically relevant examples of proline-derived natural products include prodigiosin, pyoluteorin, and chlorozidine (Fig 1A).[6] The proline moieties present in these metabolites are incorporated into these natural products through the coordinated effort of multiple enzymes. Specifically, these pathways utilize type II NRPS enzymatic domains to prepare the pyrrolidine ring originating from proline. Due to their ability to biosynthesize complex molecules with potentially medically useful bioactivities, these pathways have represented a promising target for engineering to create new natural products.

Each of these proline-incorporating pathways employ similar type II NRPS proteins with an adenylation (A) domain acting as a gatekeeper that activates L-proline in an ATP-dependent manner to transfer proline to the phosphopantetheine (PPant) arm of the peptidyl carrier protein (PCP).[2,7] The PCP is a small ~10 kDa protein that is post-

translationally phosphopantetheinylated to *holo*-PCP that, through the thiol terminus of its 4'-phosphopantetheine prosthetic group, can tether peptide substrates to the carrier protein via a thioester linkage.[7] Once loaded with proline, the PCP shuttles the prolyl scaffold to a dehydrogenation (DH) domain, which catalyzes the FAD-dependent oxidation of the proline moiety in the biosyntheses of anatoxin, prodigiosin, pyoluteorin, and chlorizidine. The resulting pyrrole species can undergo further chemical modification before transfer and elaboration by a downstream enzyme (Fig 1A).[6,8] In the type II NRPS system responsible for prodigiosin biosynthesis, the PCP, PigG, is loaded with a prolyl moiety by a cognate A domain, PigI (Fig 1B). Similarly, the NRPS that produces pyoluteorin includes a PCP, PltL, that is converted to prolyl-PltL by the A domain, PltF (Fig 1C). Prior studies of these systems have revealed the central role of proper protein-protein interactions in the substrate-charging of the PCP.[9] Further experimental work showed that both cognate (PigI) and noncognate (PltF) A domains can aminoacylate PigG, while PltL reacts exclusively with PltF, despite high sequence homology and conserved PCP structures. Understanding the precise mechanism of specificity between these highly conserved PCP and A domain interactions has been seen by our team as a key to understanding the rules by which these enzymes function.

Development of the mechanism-based chemical trapping of PCP-A domain complex laid foundation for elucidation of the co-crystal structure for cognate PltL-PltF.[9,10] Here we report the cognate PigG-PigI and non-cognate PigG-PltF complexes and the analysis of protein-protein interactions in each PCP-A domain complex in context of cross-species activity data obtained previously by the NMR titration studies and alanine scanning mutagenesis in the PCP loop 1 region.[9] We also applied molecular dynamics (MD) simulations to the PltL-PltF, PigG-PigI, PigG-PltF, holo-PigG, and holo-PltL structures to compare dynamism in the PCP loop 1 region. Finally, sequence alignment and structural superimposition-guided mutagenesis in the non-cognate PltL-PigI interface resulted in 20-fold enhancement of the PltL-PigI activity.

## Results

### Crystal structures of the PigG-PigI and PigG-PltF complexes.

In order to elucidate the molecular basis of PigG's promiscuity towards homologous A domains, we solved the X-ray crystal structures of the cognate PigG-PigI and the noncognate PigG-PltF complexes. A mechanism-based crosslinker[10] was used to trap *holo*-PigG in association with either PigI or PltF, which stabilized these complexes and facilitated their crystallization. The PigG-PigI crystal diffracted to 1.61 Å resolution, while the PigG-PltF crystal diffracted to 2.46 Å resolution. Initial phases for both data sets were obtained using the adenylation domain, PltF, from a previously reported PCP-A domain complex crystal structure (PDB ID: 6O6E)[10],which was used as a search model for molecular replacement. The NMR solution structure of PigG (PDB ID: 5JDX)[9] was then used to fit PigG into the remaining electron density. PigG and PigI crystallized as a single complex in the asymmetric unit, whereas two PigG-PltF complexes constituted a single asymmetric unit.

The A domains, PigI and PltF, comprise an N-terminal subdomain ($A_{core}$) and a C-terminal subdomain ($A_{sub}$). In the crosslinked complexes, both A domains have been resolved in the thiolation conformation (Fig 2A), in which both subdomains are positioned to form a protein interface capable of binding a partner PCP.[5] The $A_{core}$ of PigI and PltF are responsible for binding substrates ATP and L-proline, whereas the $A_{sub}$ contains residues that are important for catalysis of both the adenylation and thiolation half-reactions. In both crosslinked structures, PigG maintains the conserved 4 α-helix bundle with the PPant attached to Ser36 of PigG and extended into the active site of the A domain, where it is covalently linked to the prolyl-adenosine vinylsulfonamide (Pro-AVSN) chemical probe (Fig 1D, Fig S1).

The active site of the PigG-PigI and PigG-PltF structures resembles the active site of the previously reported crosslinked PltL-PltF complex (PDB ID: 6O6E).[10] Similar to what is observed in the PltL-PltF structure, the $A_{sub}$ domain catalytic lysine responsible for adenylation in PltF and PigI, Lys486 and Lys477, respectively, are 24 Å and 25 Å away from the active site, in accordance with the proposed domain reorganization process that occurs between the catalysis of the adenylation and thiolation half-reactions.[11,12]

### The protein-protein PCP-A domain interface.

The protein-protein interface between PigG and both A domains is formed by PigG loop 1, which is a region of 20 residues connecting helices 1 and 2 (Fig 2), that mainly contacts the $A_{sub}$ domain. The protein-protein interfaces of both PigG-PigI and PigG-PltF feature specific hydrogen-bonding and hydrophobic interactions. Loop 1 of PigG interacts with PigI utilizing a network of hydrogen bonds that include interactions between the sidechain carboxamide of PigG's Asn32 and the backbone carbonyl of PigI's Ile462, the sidechain amine of PigI's Lys450 and the backbone carbonyls of PigG's Leu31 and Asn32, the amide of PigI's Asn464 and the backbone carbonyls of PigG's Phe14 and Ile33, and PigG's Asp35 sidechain carboxylate with the sidechain amide of PigI's Asn396 (Fig 2B). The protein-protein interface of PigG-PltF features a similar hydrogen-bonding network that includes hydrogen bonding of the sidechain amide of PigG's Asn32 with the sidechain carboxyl of PltF's Asp471 and the backbone carbonyls of PltF's Ile469 and Ile470, the amine of PltF's Lys472 and the backbone carbonyl of PigG's Gln13 and Phe14. Arg404 of PltF forms a salt-bridge interaction with Asp35 of PigG (Fig 2E). Both the PigG-PigI and PigG-PltF interfaces share a common hydrophobic interaction, where an aliphatic residue, either Leu447 of PigI or Ile454 of PltF, sits inside a hydrophobic pocket formed by Leu15, Thr26, Ile31, Ile33, and Gly21 of PigG's loop 1 (Fig 2C). The PigG-PltF crystal structure shows an additional hydrophobic interaction involving the aromatic sidechain of Phe231 of PltF, which sits in between helix 1 and 3 of PigG (Fig 2F).

Alanine scanning mutagenesis of PigI and PltF residues that directly participate in molecular recognition demonstrates the essential role of specific interactions in mediating PCP-A domain complexation (Fig S2). Disruption of specific hydrogen-bonding and hydrophobic interactions by mutagenesis results in stunted activity with PigG. Interestingly, the F231A mutant of PltF exhibited an increased rate of aminoacylation of PigG. Previous NMR titration experiments suggested that PigG binds more tightly to the noncognate PltF than

its cognate partner PigI,[9] which may be attributed to hydrophobic interactions involving F231A. Furthermore, the same NMR titration studies reveal significant chemical shift perturbations (CSPs) in loop 1 residues of PigG upon binding to PigI and PltF, which include PigG Leu31, Asn32 and Asp35 upon PigI binding and PigG Leu31 upon PltF binding.[9] These perturbed residues are also observed at the interfaces of the crosslinked PigG-PigI and PigG-PltF structures, supporting that these residues are involved in forming a productive interaction (Fig S3). Lastly, these findings are consistent with the previously reported PltL-PltF structure, which was crosslinked in a similar manner using an identical proline AVSN probe.[10] The PCP, PltL, interacts through its loop 1 to PltF's $A_{sub}$ domain, forming similar hydrogen-bonding and hydrophobic interactions found in the PigG-PigI and PigG-PltF interfaces (Fig S3). Interestingly, unlike PigG, PltL is specific towards its cognate A domain, PltF, and is not prolylated efficiently by PigI despite the formation of similar PCP-A domain interfaces.[9]

## Contribution of the PCP loop 1 in protein partner specificity.

In order to tease out the differences in the molecular basis of partner protein specificity of PigG and PltL, the solution NMR structures of standalone (unbound) *holo*-PigG (PDB 5JDX) and *holo*-PltL (PDB 2N5H) were compared to the structures of bound PCPs observed in the X-ray crystal structures of the crosslinked PCP-A domain complexes. Comparison of the structures revealed dependence of the PCP-A domain complex formation on the PCPs' loop 1 dynamics (Fig 3). Closer inspection of the PltL-PltF structure reveals a change in the orientation of the N-terminal region of PltL's loop 1 that distinguishes the conformations of loop 1 in the bound and unbound states. This conformational change creates a hydrophobic pocket that is filled with PltF's Ile454 (Fig 3L). In the bound structure, the N-terminal portion of loop 1 of PltL was disordered.[10] When compared to standalone PigG, PigG in the PigG-PigI and the PigG-PltF structures show a similar, but more modest change in the conformation of the PigG's loop 1 (residues 17-25) (Fig 3B, Fig 3G). The beginning of loop 1 slightly shifts to form a hydrophobic pocket that accommodates PigI's Leu447 or PltF's Ile454, in the PigG-PigI and PigG-PltF complexes (Fig 3B, Fig 3G).

Furthermore, the PltL loop 1 (22 residues long) is two residues longer than PigG's 20-residue loop 1. PltL has an additional Pro21 and Ser22 (Fig S4, Fig S5A); this proline residue is not found at the corresponding position in loop 1 of PigG. A comparison of the solution NMR structures of PigG and PltL demonstrates that this proline kinks PltL's loop 1 in a manner that distinguishes it from PigG's loop 1 and may play a role in PCP-A domain specificity (Fig 3C, Fig 3H, Fig 3M, Fig S5A). Collectively, we speculate that PCP-A domain specificity may be defined by the differences in loop length and amino acid composition that affect conformational dynamics of the N-terminal part of loop 1.

## MD simulations of the PCP-A domain interface.

To elucidate the PCP loop 1 dynamics, MD simulations were performed on *holo*-PigG, *holo*-PltL, prolyl-PigG, and prolyl-PltL as well as the PigG-PigI, PigG-PltF, and PltL-PltF complexes (Fig 3). Initial root mean squared fluctuation (RMSF) analysis revealed the N-terminal portion of the PCP loop 1 of both *holo*-PltL and *holo*-PigI experienced larger backbone and side chain fluctuations compared to the rest of loop 1 (Fig 3D, Fig 3I, Fig

3N). MD simulations of each PCP-A domain complex revealed lower backbone fluctuations in all loop 1 regions relative to the unbound PCP (Fig 3E, Fig 3J, Fig 3O). Bound PltL in the PltL-PltF complex, however, has higher fluctuations in the beginning of loop 1 compared to bound PigG in the PigG-PigI and PigG-PltF complex (Fig 3E, Fig 3J, Fig 3O). These increased fluctuations suggest a lack of a single low energy conformation of the beginning of the PltL loop 1. This is also supported by the PltL-PltF crystal structure, where the N-terminal portion of the PltL loop 1 lacks electron density.[10]

Analysis of PCP-A domain interactions sampled over the course of MD simulations finds that the interface is largely stabilized by hydrophobic interactions between PCP and A domain residues and additional contacts absent in the X-ray structures of PigG-PigI and PltF-PigG are observed computationally (Fig 4). The most frequently sampled example of such a contact is the interaction of PltL's Leu40 and PltF's Val403 of the PltL-PltF complex (Fig 4B). PltL's Leu40 is also frequently in hydrophobic contact with PltF's Phe303 and Leu401 and engages in backbone hydrogen bonding with Glu380 and Glu405. In addition, PltL's Tyr51 forms a sidechain hydrogen bond with PltF's Gln260. Lastly, Lys24 of PltL is found in proximity to Lys457. Unlike PltL-PltF, the most frequently sampled contact in simulation of PigG-PigI and PigG-PltF is a hydrogen bonding interaction (Fig 4B). In the case of PigG-PigI simulations, this interaction involves the carboxylate of PigG's Asp35 and the backbone of PigI's Gly373 (Fig 4A), while according to simulations of the noncognate PigG-PltF complex, a similar interaction involves Asp38 of PigG and Val403 of PltF (Fig 4A). Computer simulations of PigG-PigI indicate that the PigG-PigI interface features a greater number of polar residues than the interfaces of either PigG-PltF or PltL-PltF, including the following PCP-A domain contacts: Asp35 of PigG with Gly373 of PigI, Leu15 of PigG with Asn464 of PigI, and Asp35 of PigG with Asp 186 of PigI.

### Rational design of enhanced protein-protein interface.

Based on the sequence alignments, structural superimpositions, and MD simulations, we identified residues that could be targeted to introduce and enhance noncognate activity between PltL and PigI (Fig S5B, Fig S5C). Site directed mutagenesis was performed to swap residues between PigG and PltL to give PltL mutants (mPltL) W37L, G38N, and ΔP21 ΔS22 as well as between PltF and PigI to give PigI mutants (mPigI) K212F, N396R, L447I. The PltL ΔP21 ΔS22 and PltL W37L mutants did not significantly affect activity with PigI, whereas the G38N mutation increased prolyl-PltL formation to 32% from the initial 15% between wild-type PigI and PltL (Fig 5). Combination of single mutations to form the PltL double mutants generally did not affect activity, whereas surprisingly the PltL ΔP21 ΔS22 G38N mutation increased conversion to 59%. Remarkably, combination of all three PltL mutations increased conversion to 72%, where PltF had a calculated turnover rate of 4.54 mPltL/hr (Table S2), where the turnover rate of PigI with PltL was calculated to be 0.92 PltL/hr. Of the individual PigI mutations, the single N396R mutation significantly improved activity with WT PltL to 74% conversion, with a turnover rate of 6.17 PltL/hr (Fig 5, Table S2). The PigI double and triple mutations did not provide any more prolyl-PltL formation compared to the single N396R mutation. Subsequent incubation of the best mutants PigI N396R with PltL ΔP21 ΔS22 W37L G38N provided an increased product formation at 93%

completion with a calculated turnover rate of 18.87 mPltL/hr, which is a 20-fold increase compared to the wild-type proteins.

To better understand how loop dynamics of PltL are altered by mutagenesis, we performed simulations of the holo and prolyl forms of PltL ΔP21 ΔS22 and PltL ΔP21 ΔS22 W37L G38N mutants. To generate coordinates for the deletion mutants we used RoseTTAFold, an AI-driven structure prediction tool to predict the conformation of the altered loops,[20] as we anticipated that residue deletions would necessarily influence the backbone conformational preferences of the loop. Interestingly and perhaps unsurprisingly, a template-based comparative modeling approach using RosettaCM predicted approximately identical structures of the PltL mutant (Fig S6). Analysis of the resulting RoseTTAFold-determined structures demonstrated that the deletion of PltL's Ser21 and Pro22 residues have a dramatic effect on the geometry of the loop 1 (Fig 6). Comparison to MD simulations of the PltL ΔP21 ΔS22 W37L G38N mutant also decreases the dynamism observed in the loop 1 region. Ultimately, these modeled structures feature a loop orientation that more closely resembles that of PigG's loop 1 arrangement in both the unbound NMR structure as well as the crosslinked PigG-PigI and PigG-PltF structures.

## Discussion

X-ray crystal structures, mutagenesis studies, and computer simulations demonstrate that the PCP-A domain interfaces are largely stabilized by hydrophobic interactions. The interfaces of PltL-PltF, PigG-PigI, and PigG-PltF measure in terms of buried surface area, 847, 769, and 955 $Å^2$, respectively, and are larger than the interfaces observed between type II acyl carrier proteins and type II fatty acid synthases, which are dominated by electrostatic and polar interactions (Fig S7).[14–17] Taken together these findings suggest that type II NRPS proteins may have evolved to form relatively hydrophobic interactions to maintain orthogonality between carrier protein-mediated secondary metabolic pathways and type II fatty acid primary metabolism.[18]

Although mainly a hydrophobic interface, a key electrostatic interaction between the PCPs and A domains appear to help form productive complexes. This phenomenon is evidenced by the effect of the introduction of a positively charged residue, N396R (Fig S5C), at the entrance of PigI's substrate binding pocket, which yielded a mutant PigI that had ~6-fold higher turnover rate when treated with wild type PltL; this mutation introduces a long-range electrostatic interaction involving the arginine's guanidinium group of N396R and the phosphate of the PPant cofactor.[19] Sequence alignments of 75 A domains reveal that positively charged residues, like Lys212 of PigI or Arg404 of PltF, at the entrance of the PPant binding tunnel are conserved amongst type I and type II NRPS A domains (Fig S8, Fig S9). Structural analysis of previously solved PCP-A domain crystal structures also reveal a positively charged amino acid interacting with the phosphate of the PPant, thus supporting the sequence alignments and the importance of PPant recognition in PCP-A domain binding (Fig S10).[10,21–25]

Conversely, a single G38N mutation of PltL increases its activity toward PigI. Surprisingly, subsequent mutation additions of ΔP21 ΔS22 and W37L to the G38N mutation revealed a

synergistic 4-fold increase in activity with PigI. The G38N mutation seems to add additional hydrogen bonding interactions with the A domain (Fig S5B), whereas the ΔP21 ΔS22 mutation changes the loop 1 region 1 structure to adopt a conformation more accessible to the A domain hydrophobic residue. The PCP loop 1 length is generally between 17-21 residues from 84 aligned PCP sequences, however, PltL was the only PCP with a loop 1 length of 22 residues (Fig S11). Lastly, decreasing the size and thus steric clash of the hydrophobic residue (PltL W37L) involved in forming the PCP loop 1 hydrophobic pocket may also enhance PCP and A domain recognition.

The stepwise increase in activity observed in additional PltL mutations may provide hints towards the mechanism of PCP-A domain binding. First, initial PCP attraction to the A domain can be influenced by the long-range electrostatic attraction between the PPant phosphate and positively charged residues at the entrance to the PPant tunnel (PltF Arg404, PigI Lys212, PigI N396R). Subsequently, adjacent hydrogen bonds may begin to form between the PCP loop 1 region (PigG Asn32, PltL Gly38, and PltL G38N) with the $A_{sub}$ domain (PigI Ile462 Asn464, PltF Ile469 Ile470 Asp471). This intermediate binding conformation may prime the PCP loop 1 hydrophobic pocket (PltL ΔP21 ΔS22) to more readily accommodate the $A_{sub}$ domain hydrophobic residue (PigI Leu447, PltF Ile454). Lastly, the PCP loop 1 hydrophobic pocket is then formed (PigG Leu31, PltL Trp37, PltL W37L) and accessed by the A domain hydrophobic residue. The order of PCP-A domain binding can be hypothesized as follows, where 1) long-range electrostatics attract the two proteins, then 2) adjacent specific hydrogen bonding interactions form, which allows for 3) neighboring hydrophobic interactions to develop through loop 1 movements, thus creating an optimized protein-protein interface that can enables PCP and A domain binding and for thiolation to occur.

The addition of positive charge at the PPant tunnel entrance has proven to be significant and powerful in enhancing a non-cognate PCP-A domain interaction, which may be utilized as an initial step in not only designing a new PCP-A domain interface, but also other carrier protein dependent pathways such as fatty acid synthases and polyketide synthases, which generally rely on electrostatic interactions to form a protein-protein interface.[18] Although the A domain prolylation activity of our designed PCP-A domain interface is still two orders of magnitude slower than the wild-type complex,[26] further interface design can be performed to reach wild-type levels of prolylation.

While this work improves upon the design of an enhanced PCP-A domain interaction, this method of structure-based rational design may be incorporated to other PCP and partner protein interactions. Notably, the condensation (C) domain is commonly the next partner protein the PCP interacts with, which condenses the upstream and downstream peptidyl moieties attached to the PCP; thus, a noncognate PCP-C domain complex may require interface enhancements for binding during combinatorial biosynthesis. Recent structural analysis of the PCP-C domain complex has begun revealing the interface dependence of hydrophobic interactions, which can be used as a starting point towards structure-based rational design of noncognate PCP and C domain complexes. We envision that the protein-protein interface design of PCP-A domain interfaces and other PCP-partner protein

interfaces will enhance future success in the combinatorial biosynthesis of carrier protein dependent pathways.

## Conclusion

The X-ray crystal structures, mutagenesis studies, and MD simulations in conjunction with prior NMR experiments and X-ray crystallography demonstrate the integral role of carrier protein loop 1 dynamics in mediating molecular recognition and catalysis in type II NRP biosynthesis. This work provides insights into the forces that govern the noncovalent interactions that promote PCP-A domain complexation. Perhaps most importantly, these studies provide evidence that rational interface design guided by structural, chemical, and computational biology can be used to manipulate or alter the binding specificities of proteins responsible for the biosynthesis of complex natural products.

## Methods

### Molecular dynamics simulation structure preparation.

The following carrier proteins were modeled in their *holo* and prolyl-forms of standalone carrier proteins, PigG, PltL and two mutant PltL variants, ΔP21, ΔS22-PltL and ΔP21, ΔS22, W37L, G38N-PltL. In addition, the following PCP-A domain complexes were simulated, PltL-PltF, PigG-PigI, and PigI-PltF. For simulation work, the coordinates for the wildtype PCPs, PigG and PltL were generated using previously reported solution-phase NMR structures of holo-PltL (PDB ID: 2N5H)[7], prolyl-PltL (PDB ID: 2N5I)[7], and holo-PigG (PDB ID: 5JDX)[6]. The crosslinked PCP-A didomain complexes reported herein and previously, PltL-PltF (PDB ID: 6O6E)[1], PigG-PigI, and PigG-PltF were used to generate initial coordinates for simulations of the three PCP-A complexes. The prolyl-PCPs and PCP-A complexes possess a substrate or intermediate species that could exist in either a neutral or protonated form; in all case both forms were considered. In total, 20 systems were subjected to MD simulation.

### Machine learning structure prediction.

The variants of PltL, ΔP21, ΔS22-PltL and ΔP21, ΔS22, W37L, G38N-PltL, were prepared using RoseTTAFold using Robetta (https://robetta.bakerlab.org)[8], a continually evaluated (through CAMEO) protein structure prediction service. Schrodinger's Protein Preparation Wizard (https://www.schrodinger.com/protein-preparation-wizard and https://www.schrodinger.com/prime) were used to add missing C-, N-terminal residues and missing side chains not resolvable from the experimental density, and hydrogen atoms were added all heavy atoms to cap all open valences, to predict the protonation states of the titratable residues in each structure assuming a pH of 7.4, and to optimize their orientation, all waters resolved crystallographically were preserved for computer simulation. Histidine protonation states were inspected by hand.

### Interface design of PltL and PigI

Interface design between PltL and PigI was prepared by introducing single mutations as described previously in alanine scanning of PigI and PltF. Expression and purification of

each PigI mutant was performed similarly to as described previously for PigI and PltF alanine scanning, whereas each PltL mutant was expressed and purified similarly as *holo*-PigG and *holo*-PltL.

The designed PigI with the designed PltL initial assays contained 50 mM Tris pH 7.5, 12.5 mM $MgCl_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.05 mM *holo*-PltL, and 0.0238 mM mPigI. The reaction was incubated at 25 °C, 300 rpm, for 1 hour and quenched with 1.3% formic acid. The reaction was prepared, passed, and analyzed through the same HPLC protocol as outlined in Alanine scanning of PigI and PltF.

The time course assays of the designed PigI with the designed PltL contained 50 mM Tris pH 7.5, 12.5 mM $MgCl_2$, 2.0 mM TCEP, 5 mM ATP, 5 mM L-Pro, 0.079 mM *holo*-PltL, and 0.0238 mM mPigI. Each time course utilized between 5-7 time points and was repeated in at least duplicate. Each time point was quenched with 1.67% formic acid and subjected to HPLC analysis identically to Alanine scanning and initial Interface design assays.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Y. Su for MS analyses, A. Mrse and X. Huang for NMR assistance, and G. Meigs, J. Holton, and J. Tanamachi for X-ray data collection assistance. This work was supported by NIH R01 GM095970 and R01 GM031749. J.C.C. was supported by F31 GM13761601 and T32 GM008326. We thank the San Diego Supercomputing Center for access to the Triton Shared Computing Cluster for resources used to perform computer simulations. J.C.C. and A.P. contributed equally to the experimental, analytical, and manuscript writing of this project. J.C.C., A.P., T.D.D., L.M.P., J.A.M., and M.D.B. contributed to writing and editing of the manuscript.

## References

(1). Newman DJ; Cragg GM Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. J. Nat. Prod 2020, 83 (3), 770–803. [PubMed: 32162523]

(2). Hur GH; Vickery CR; Burkart MD Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. Nat. Prod. Rep 2012, 29 (10), 1074. [PubMed: 22802156]

(3). Finking R; Marahiel MA Biosynthesis of Nonribosomal Peptides. Annual Review of Microbiology 2004, 58 (1), 453–488.

(4). Brown AS; Calcott MJ; Owen JG; Ackerley DF Structural, Functional and Evolutionary Perspectives on Effective Re-Engineering of Non-Ribosomal Peptide Synthetase Assembly Lines. Nat. Prod. Rep 2018, 35 (11), 1210–1228. [PubMed: 30069573]

(5). Corpuz JC; Sanlley JO; Burkart MD Protein-Protein Interface Analysis of the Non-Ribosomal Peptide Synthetase Peptidyl Carrier Protein and Enzymatic Domains. Synthetic and Systems Biotechnology 2022, 7 (2), 677–688. [PubMed: 35224236]

(6). Jaremko MJ; Davis TD; Corpuz JC; Burkart MD Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. Nat. Prod. Rep 2020, 37 (3), 355–379. [PubMed: 31593192]

(7). Beld J; Sonnenschein EC; Vickery CR; Noel JP; Burkart MD The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. Nat. Prod. Rep 2014, 31 (1), 61–108. [PubMed: 24292120]

(8). Acharya A; Yi D; Pavlova A; Agarwal V; Gumbart JC Resolving the Hydride Transfer Pathway in Oxidative Conversion of Proline to Pyrrole. Biochemistry 2022, 61 (3), 206–215. [PubMed: 35072459]

(9). Jaremko MJ; Lee DJ; Patel A; Winslow V; Opella SJ; McCammon JA; Burkart MD Manipulating Protein–Protein Interactions in Nonribosomal Peptide Synthetase Type II Peptidyl Carrier Proteins. Biochemistry 2017, 56 (40), 5269–5273. [PubMed: 28920687]

(10). Corpuz JC; Podust LM; Davis TD; Jaremko MJ; Burkart MD Dynamic Visualization of Type II Peptidyl Carrier Protein Recognition in Pyoluteorin Biosynthesis. RSC Chem. Biol 2020, 1 (1), 8–12. [PubMed: 33305272]

(11). Gulick AM Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. ACS Chem. Biol 2009, 4 (10), 811–827. [PubMed: 19610673]

(12). Mayerthaler F; Feldberg A-L; Alfermann J; Sun X; Steinchen W; Yang H; Mootz HD Intermediary Conformations Linked to the Directionality of the Aminoacylation Pathway of Nonribosomal Peptide Synthetases. RSC Chem. Biol 2021, 2 (3), 843–854. [PubMed: 34458813]

(13). Jaremko MJ; Lee DJ; Opella SJ; Burkart MD Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL. Journal of the American Chemical Society 2015, 137 (36), 11546–11549. [PubMed: 26340431]

(14). Chen A; Re RN; Burkart MD Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. Nat. Prod. Rep 2018, 35 (10), 1029–1045. [PubMed: 30046786]

(15). Misson LE; Mindrebo JT; Davis TD; Patel A; McCammon JA; Noel JP; Burkart MD Interfacial Plasticity Facilitates High Reaction Rate of E. Coli FAS Malonyl-CoA:ACP Transacylase, FabD. Proc Natl Acad Sci USA 2020, 117 (39), 24224–24233. 10.1073/pnas.2009805117. [PubMed: 32929027]

(16). Mindrebo JT; Misson LE; Johnson C; Noel JP; Burkart MD Activity Mapping the Acyl Carrier Protein: Elongating Ketosynthase Interaction in Fatty Acid Biosynthesis. Biochemistry 2020, 59 (38), 3626–3638. [PubMed: 32857494]

(17). Mindrebo JT; Patel A; Kim WE; Davis TD; Chen A; Bartholow TG; La Clair JJ; McCammon JA; Noel JP; Burkart MD Gating Mechanism of Elongating β-Ketoacyl-ACP Synthases. Nat Commun 2020, 11 (1), 1727. 10.1038/s41467-020-15455-x. [PubMed: 32265440]

(18). Chen A; Re RN; Burkart MD Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. Nat. Prod. Rep 2018, 35 (10), 1029–1045. [PubMed: 30046786]

(19). Zhou H-X; Pang X Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. Chem. Rev 2018, 118 (4), 1691–1741. [PubMed: 29319301]

(20). Baek M; DiMaio F; Anishchenko I; Dauparas J; Ovchinnikov S; Lee GR; Cong Q; Kinch LN; Schaeffer RD; Millan C; Park H; Adams C; Glassman CR; DeGiovavnni A; Pereira JH; Rodrigues AV; van Dijk AA; Ebrecht AC; Opperman DJ; Sagmeister T; Buhlheller C; Pavkov-Keller T; Rathinaswamy MK; Dalwadi U; Yip CK; Burke JE; Garcia CK; Grishin NV; Adams PD; Read RJ; Baker D Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021, 373 (6557), 871–876. [PubMed: 34282049]

(21). Mitchell CA; Shi C; Aldrich CC; Gulick AM Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. Biochemistry 2012, 51 (15), 3252–3263. [PubMed: 22452656]

(22). Sundlov JA; Shi C; Wilson DJ; Aldrich CC; Gulick AM Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. Chemistry & Biology 2012, 19 (2), 188–198. [PubMed: 22365602]

(23). Drake EJ; Miller BR; Shi C; Tarrasch JT; Sundlov JA; Leigh Allen C; Skiniotis G; Aldrich CC; Gulick AM Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. Nature 2016, 529 (7585), 235–238. [PubMed: 26762461]

(24). Reimer JM; Aloise MN; Harrison PM; Martin Schmeing T Synthetic Cycle of the Initiation Module of a Formylating Nonribosomal Peptide Synthetase. Nature 2016, 529 (7585), 239–242. [PubMed: 26762462]

(25). Miyanaga A; Kurihara S; Chisuga T; Kudo F; Eguchi T Structural Characterization of Complex of Adenylation Domain and Carrier Protein by Using Pantetheine Cross-Linking Probe. ACS Chem. Biol 2020, 15 (7), 1808–1812. [PubMed: 32608966]

(26). Thomas MG; Burkart MD; Walsh CT Conversion of L-Proline to Pyrrolyl-2-Carboxyl-S-PCP during Undecylprodigiosin and Pyoluteorin Biosynthesis. Chem Biol. 2002, 9 (2), 171–184. [PubMed: 11880032]

(27). Bozhüyük KAJ; Fleischhacker F; Linck A; Wesche F; Tietze A; Niesert C-P; Bode HB De Novo Design and Engineering of Non-Ribosomal Peptide Synthetases. Nature Chem 2018, 10 (3), 275–281. [PubMed: 29461518]

(28). Bozhüyük KAJ; Linck A; Tietze A; Kranz J; Wesche F; Nowak S; Fleischhacker F; Shi Y-N; Grün P; Bode HB Modification and de Novo Design of Non-Ribosomal Peptide Synthetases Using Specific Assembly Points within Condensation Domains. Nat. Chem 2019, 11 (7), 653–661. [PubMed: 31182822]

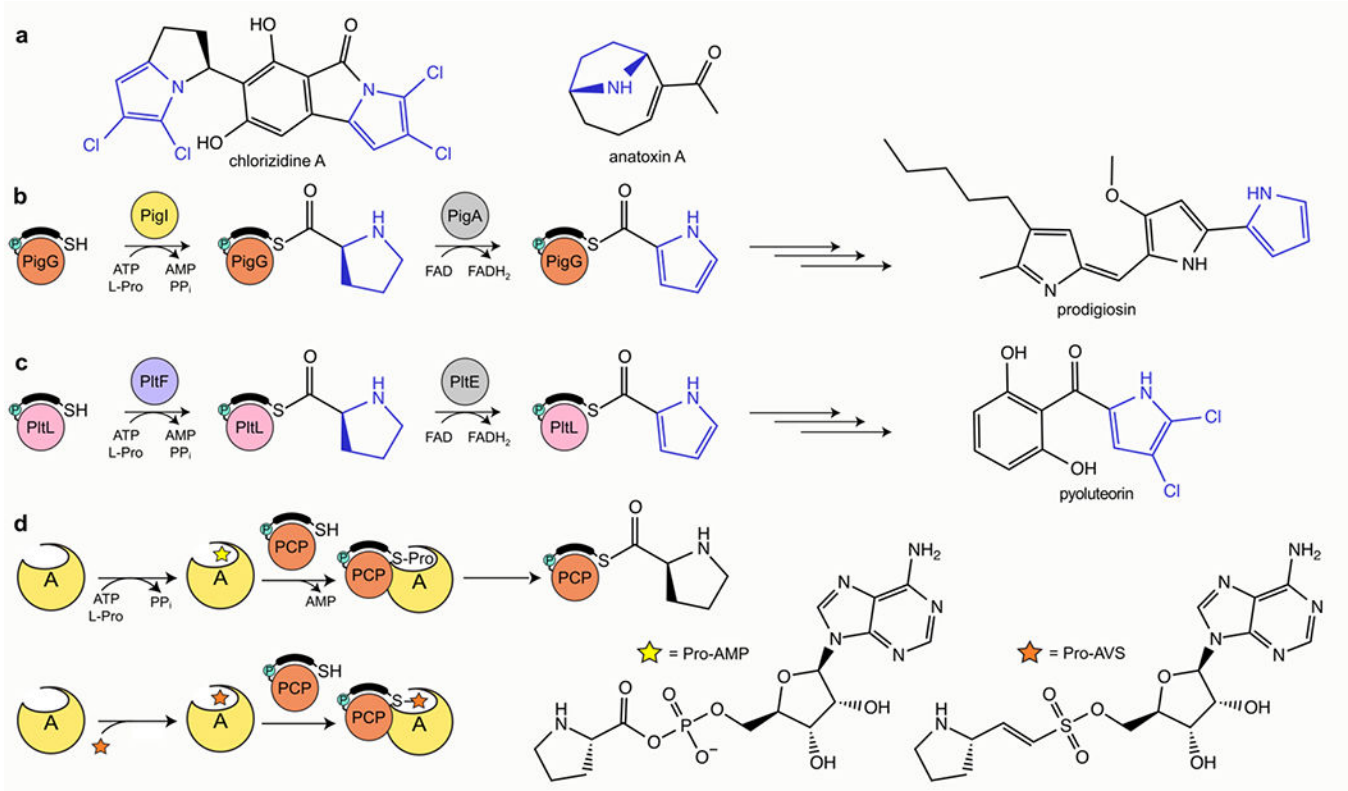



**Figure 1.**

Prevalence and preparation of proline in NRP biosynthesis. (a) Examples of NRPs that apply a functionalized proline ring. Proline is activated and dehydrogenated in (b) prodigiosin and (c) pyoluteorin biosynthesis. (d) Adenylation of L-proline to form a proline-adenosine monophosphate intermediate (yellow star) and thiolation of the holo-PCP to form prolyl-PCP. Below, the PCP-A domain complex can also be trapped with a mechanism-based inhibitor, proline adenosine vinylsulfonamide (orange star).

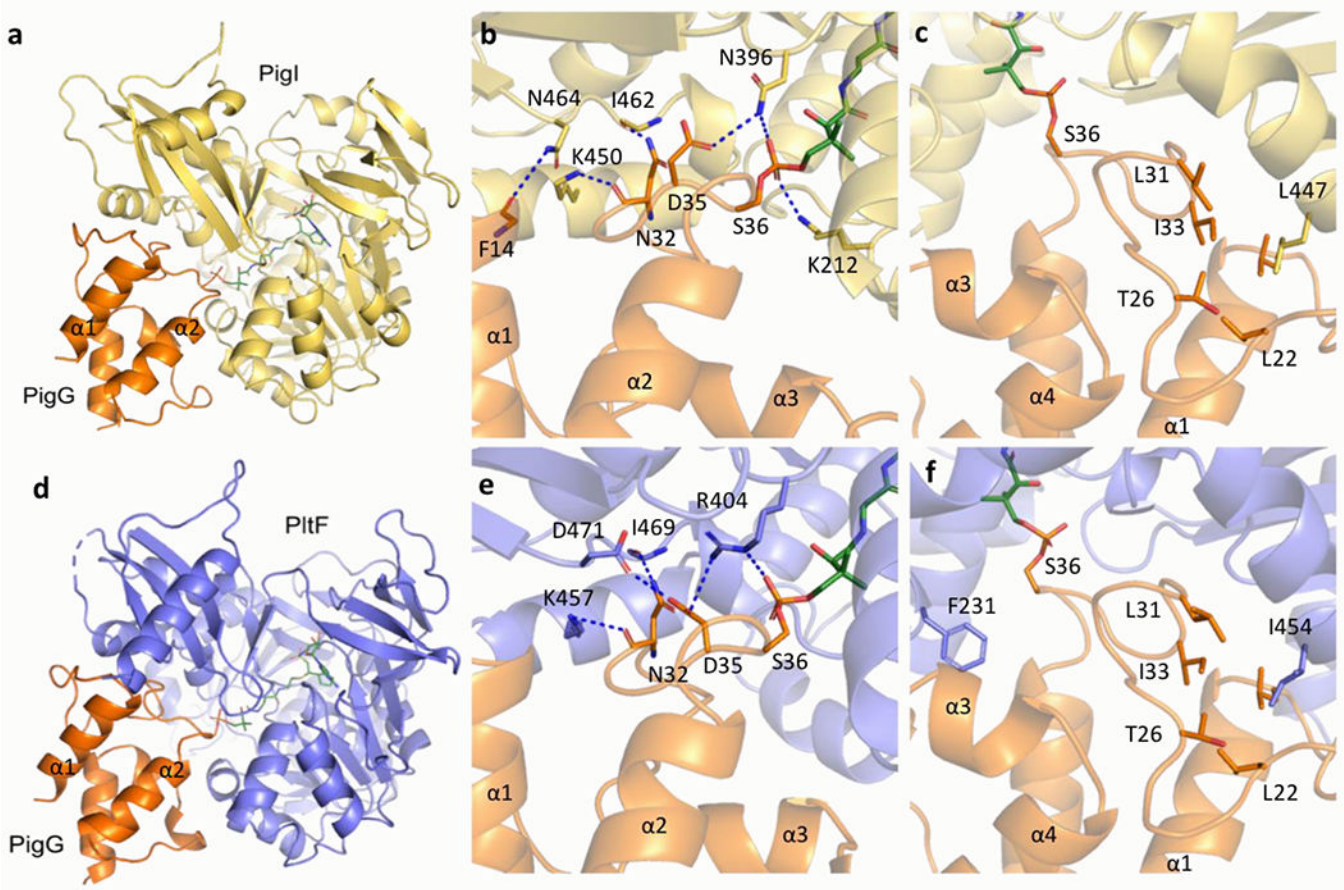



**Figure 2.**

PCP-A domain X-ray crystal structure, interface analysis, and validation. Overall X-ray crystal structure of the PigG-PigI (a) and PigG-PltF (d) complex trapped with the Pro-AVS probe. Zoom in of the PCP-A domain protein-protein interface of (b,e) hydrogen bonding and electrostatic interactions and (c,f) hydrophobic interactions. Panels b,e and c,f are rotated ~180° from each other.

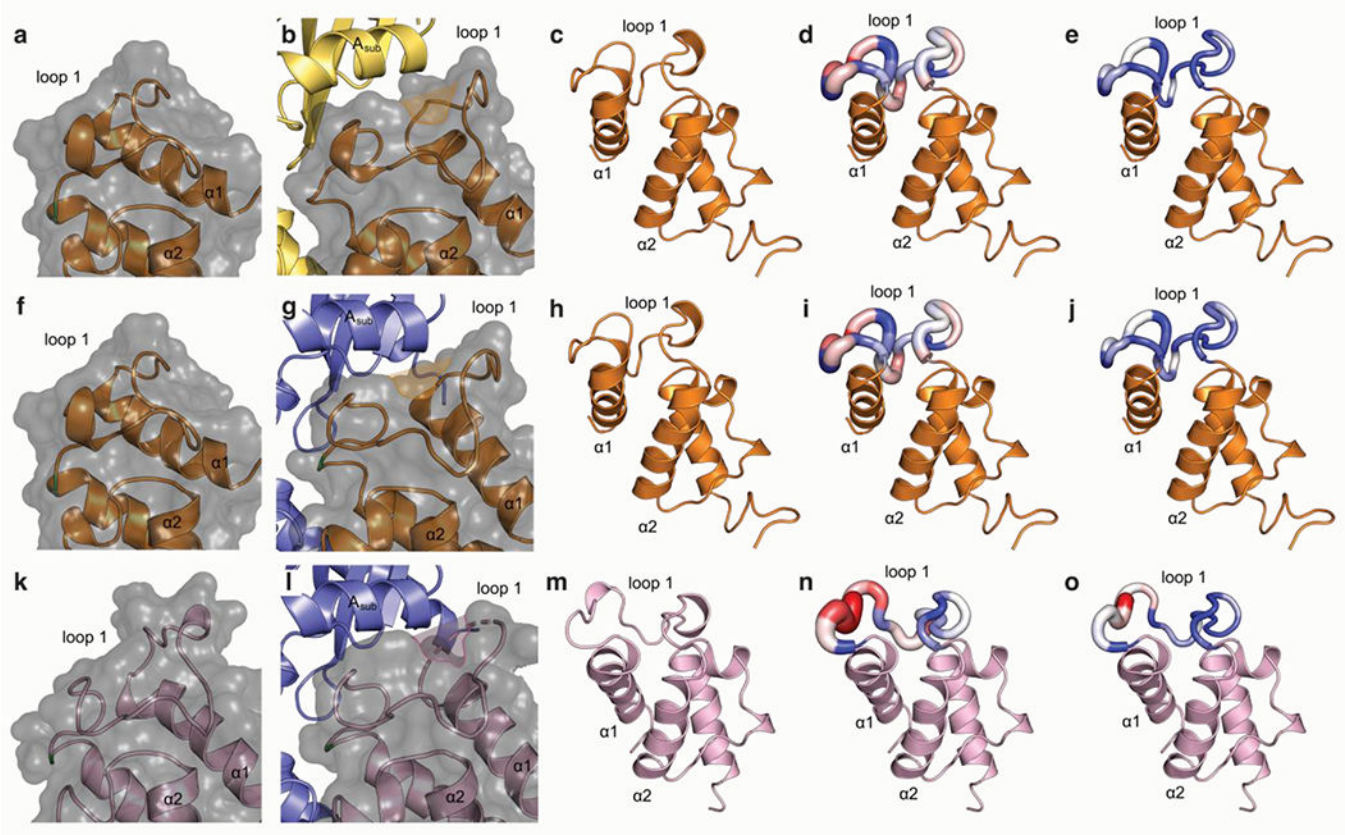



**Figure 3. Analysis of loop dynamics.**

(a, f, k) Illustrations of the loop orientation and molecular surface of the PigG (a, f) and PltL (k) in their unbound states. Images generated using NMR structures of PigG (PDB: 5JDX) and PltL (2N5H). (b, g, i). Illustrations of the loop orientation and molecular surface of the PigG (a, f) and PltL (k) in their PigI, PltF, and PltF-bound states. Loops have undergone structural reorganization upon PCP-A domain complex formation as highlighted in orange or pink to show the sideview of the hydrophobic pocket. Images generated using the X-ray structures of PigG-PigI and PigG-PltF (reported herein) as well as PltL-PltF (PDB: 606E). (c, h, m) NMR solution structures of PigG (c, h) and PltL (m) oriented to clearly present the PCPs' key loop. (d, i, n) Root mean square fluctuations (RMSFs) of the standalone PigG (d & h) and PltL (k) sampled over the course of MD simulations mapped onto the loops of the PCPs. (e, j, k) RMSFs of the PigI and PltF-bound PigG (e, h, respectively) and PltF-bound PltL (k) sampled over the course of MD simulations mapped onto the loops of the PCPs. Color spectrum from blue to white to red shows increasing backbone RMSFs calculated on a per-residue basis, whereas loop thickness from thin to thick indicates increasing sidechain RMSFs calculated on a per-residue basis. Note that PigG, PltL, PigI, and PltF are colored orange, pink, yellow, and purple, respectively.

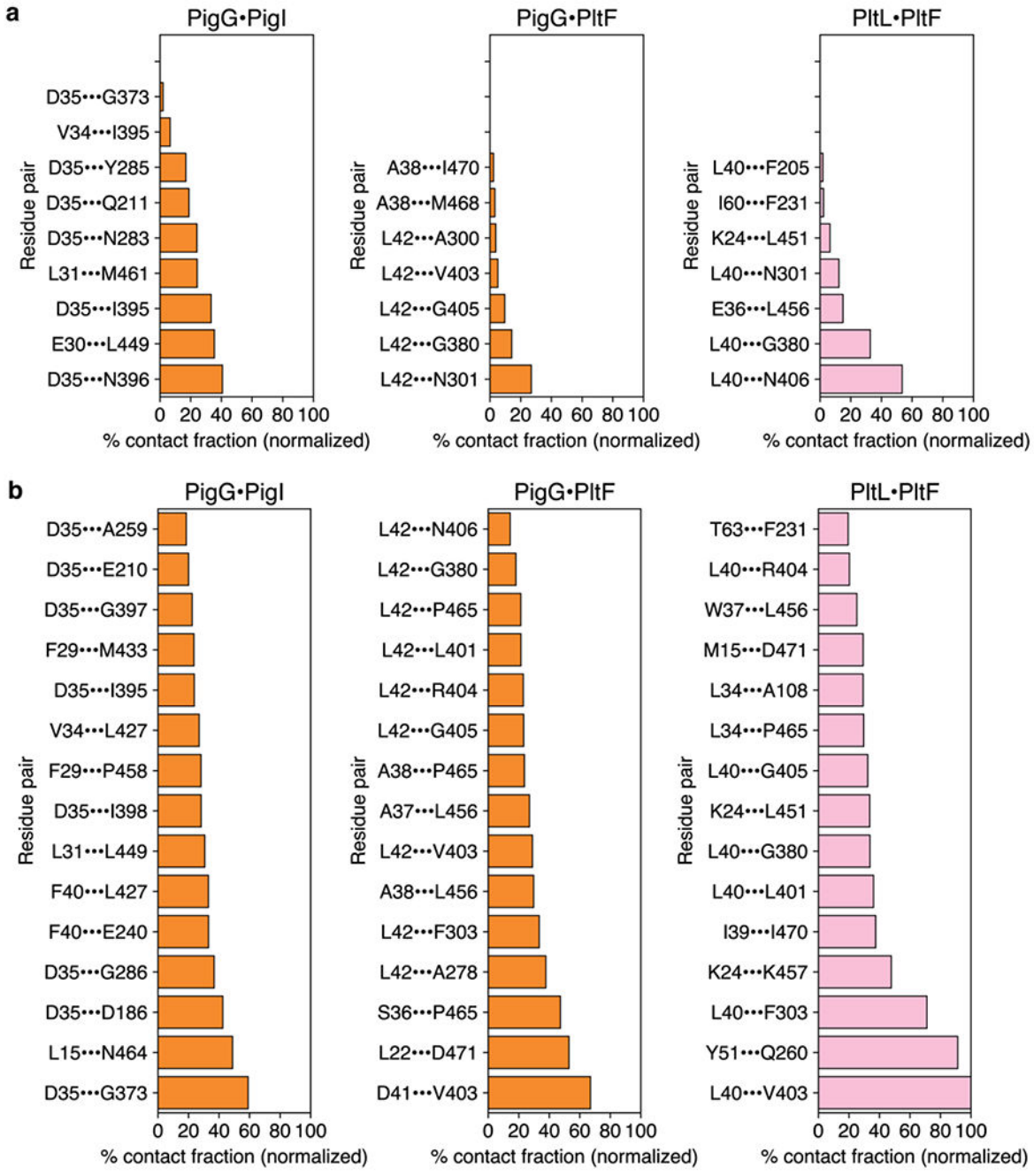



**Figure 4. Time-resolved analysis of the contacts formed between PCP and A domains.**

Contacts from the initial conditions and contacts formed during the course of the simulations are shown in panels (a) and (b), respectively. The PCP-A domain complex is labeled above each bar graph. Residues listed on the left side of the interaction pair correspond to PCP and those on the right side of the pair correspond to the A domain. The bar graphs indicate the contact fraction normalized relative to the most frequently sampled contact. The contact fraction is defined as the total fraction of simulation data in which a residue pair is engaged in an intermolecular contact. A distance criterion of 3.0 Å or less between a pair of

heavy atoms defines such a contact. Only pairwise contacts with contact fractions (prior to normalization) greater than or equal to 0.10 are included in the plots above. For brevity, no more than 15 contacts are shown in each plot. See SI for complete data.

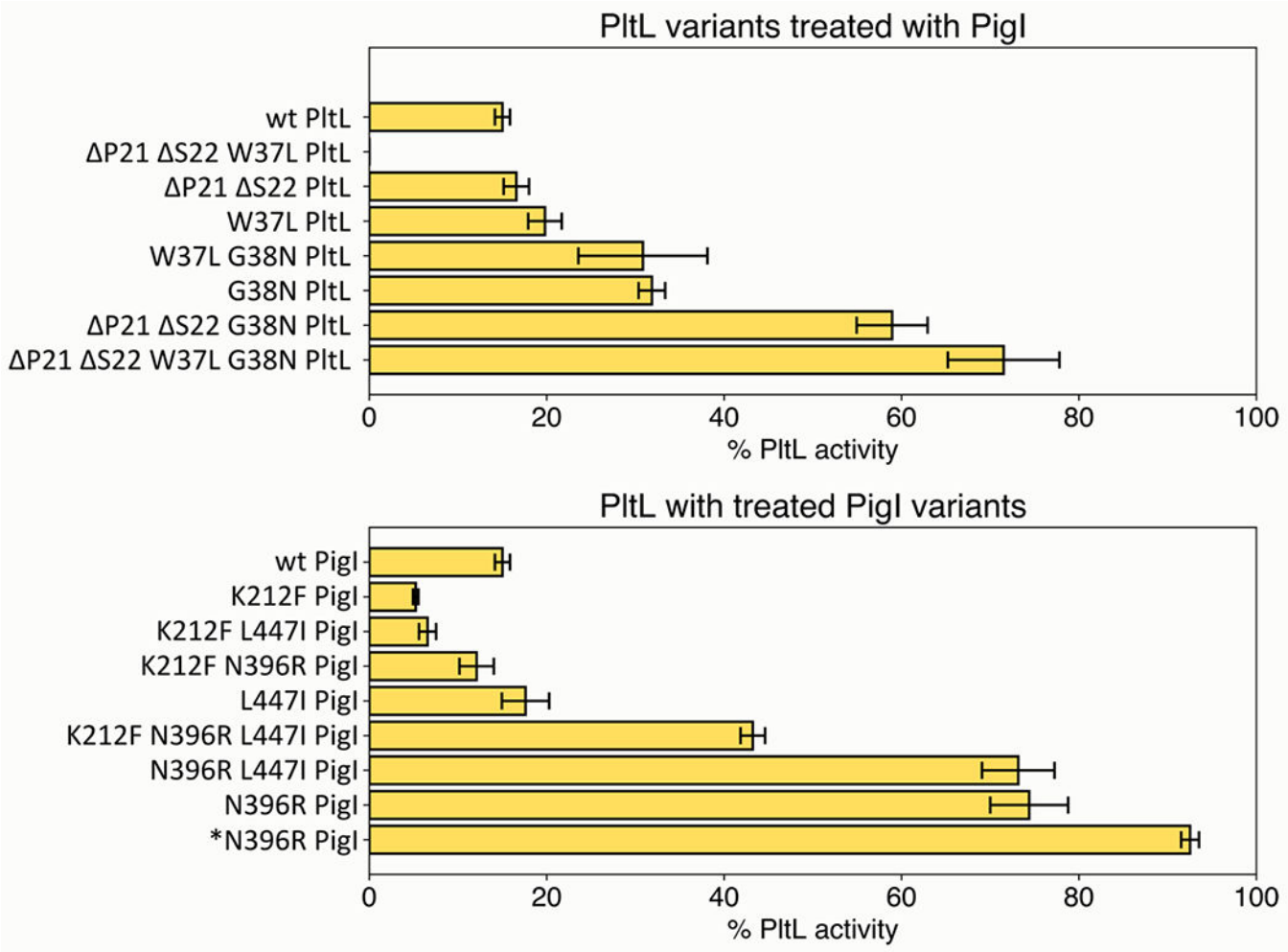



**Figure 5.**

The initial activity of rationally designed PigI or PltL mutants. Mutant PltL was incubated with WT PigI (top) and WT PltL was also incubated with mutant PigI (bottom). Activities were monitored using HPLC, and % prolyl-PltL was calculated as described previously.[10] The *N396R-PigI was incubated with the PltL ΔP21 ΔS22 W37L G38N mutant.

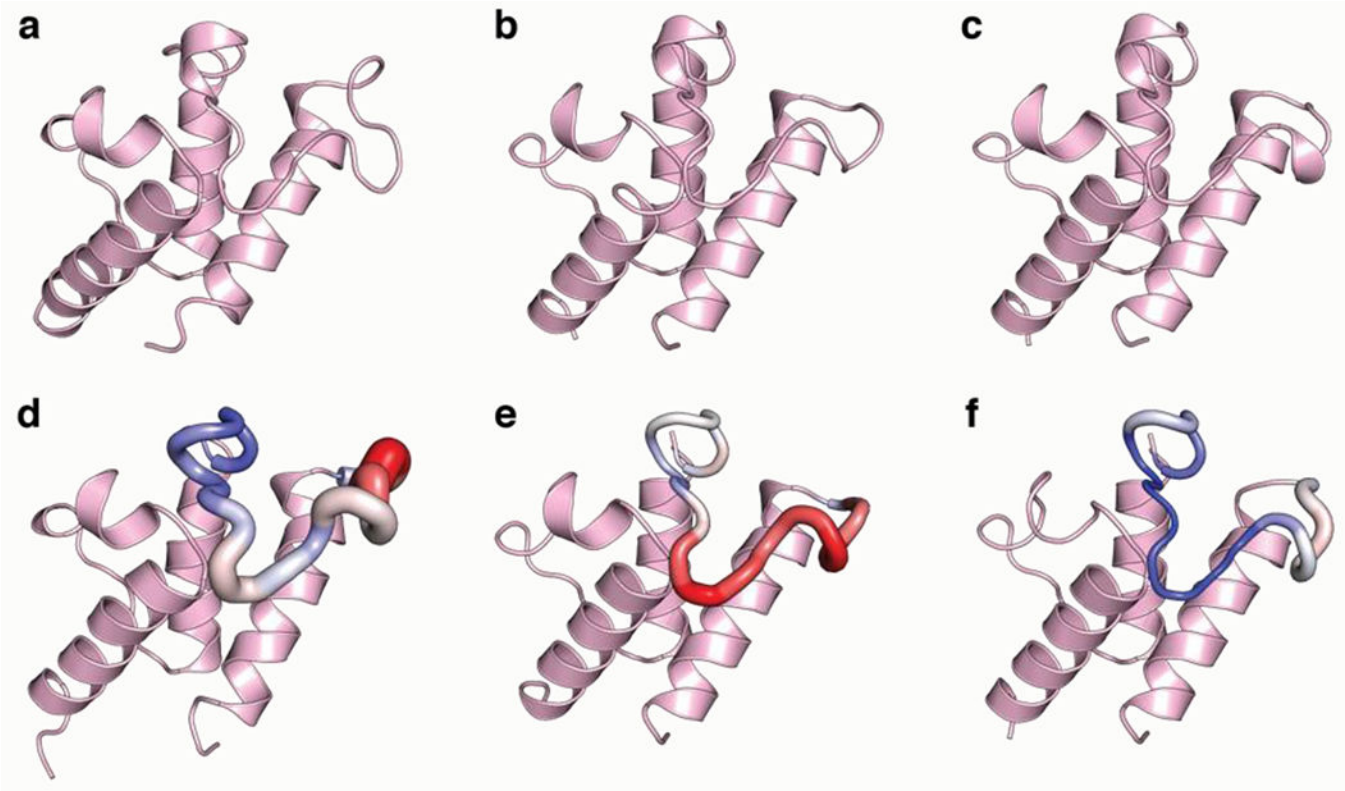



**Figure 6. Analysis of the loop conformation and dynamics of mutant PltL.**

NMR structure of wt prolyl-PltL (a), and computational models of ΔP21 ΔS22-PltL (b), and ΔP21 ΔS22 W37L G38N PltL (c). Root mean square fluctuations (RMSFs) of wt prolyl-PltL (d), and computational models of ΔP21 ΔS22-PltL (e), and ΔP21 ΔS22 W37L G38N PltL (f) sampled over the course of MD simulations mapped onto the loops of the PCPs. Color spectrum from blue to white to red shows increasing backbone RMSFs calculated on a per-residue basis, whereas loop thickness from thin to thick indicates increasing sidechain RMSFs calculated on a per-residue basis.