Published in final edited form as:
Proteins. 2014 July ; 82(7): 1210–1218.

The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins

Jeremy R. Lohman[1], **Ming Ma**[1], **Marianne E. Cuff**[2], **Lance Bigelow**[2], **Jessica Bearden**[2], **Gyorgy Babnigg**[2], **Andrzej Joachimiak**[2], **George N. Phillips Jr.**[3], and **Ben Shen**[1,4,5,*]

[1]Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458

[2]Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439

[3]Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251

[4]Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida 33458

[5]Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, Florida 33458

Abstract

Carrier proteins (CPs) play a critical role in the biosynthesis of various natural products, especially in nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymology, where the CPs are referred to as peptidyl-carrier proteins (PCPs) or acyl-carrier proteins (ACPs), respectively. CPs can either be a domain in large multifunctional polypeptides or standalone proteins, termed Type I and Type II, respectively. There have been many biochemical studies of the Type I PKS and NRPS CPs, and of Type II ACPs. However, recently a number of Type II PCPs have been found and biochemically characterized. In order to understand the possible interaction surfaces for combinatorial biosynthetic efforts we crystallized the first characterized and representative Type II PCP member, BlmI, from the bleomycin biosynthetic pathway from *Streptomyces verticillus* ATCC 15003. The structure is similar to CPs in general but most closely resembles PCPs. Comparisons with previously determined PCP structures in complex with catalytic domains reveals a common interaction surface. This surface is highly variable in charge and shape, which likely confers specificity for interactions. Previous nuclear magnetic resonance (NMR) analysis of a prototypical Type I PCP excised from the multimodular context revealed three conformational states. Comparison of the states with the structure of BlmI and other PCPs reveals that only one of the NMR states is found in other studies, suggesting the other two states may not be relevant. The state represented by the BlmI crystal structure can therefore serve as a model for both Type I and Type II PCPs.

*Correspondence to: Ben Shen, The Scripps Research Institute, 130 Scripps Way, #3A1, Jupiter, FL 33458. shenb@scripps.edu.

Keywords

protein–protein interaction; natural product; biosynthesis; phylogenetics; structural genomics; reductive methylation

INTRODUCTION

Nonribosomal peptides and polyketides are biosynthesized with remarkable fidelity, typically as a single major product after tens to hundreds of enzymatic reactions. The fidelity of nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) comes from what is called the "thiotemplate mechanism" or "assembly-line enzymology."[1] Central to these biosynthetic strategies are carrier proteins (CPs; also called thiolation domains) and more specifically PKS acyl-carrier proteins (ACPs), and NRPS peptidyl-carrier proteins (PCPs), which are ~75 amino acid long domains (predicted bioinformatically) or ~80–95 amino acid long proteins that are posttranslationally modified with a 4′-phosphopantetheinyl (ppant) group from coenzyme A by phosphopantetheinyl transferases (PPTase), also known as holo-ACP or holo-PCP synthases.[2] For both NRPS and PKS, carboxylic acid containing substrates are activated in the form of thioester linkages to the CP ppant, which allows efficient transfer of the substrate from one active site to another. Substrate specificity is conferred on multiple levels, including the enzymatic active sites and specific interactions between the CPs and catalytic domains. Understanding how the protein–protein interactions influence catalysis is an important factor for successful biosynthetic pathway engineering.

PKSs are commonly found as one of two forms, either as standalone enzymes (Type II) or as multiple domains in extended polypeptides (Type I). However, the majority of characterized NRPS are of the Type I architecture, and the existence of Type II NRPS was speculative, until we established the existence of functional Type II NRPS enzymes with the characterization of BlmI, the Type II PCP from the bleomycin biosynthetic pathway in *Streptomyces verticillus* ATCC 15003.[3] We demonstrated BlmI could be phosphopantetheinylated *in vivo* and in vitro by two different PPTases (but not efficiently by *Escherichia coli* PPTases) and loaded with valine by an adenylation domain (A-domain) from an unknown pathway in *S. verticillus*, Val-A. Further establishment of Type II NRPS enzymes comes from our discovery and characterization of a bona fide set of Type II NRPS enzymes, the Type II PCPs SgcC2,[4,5] MdpC2,[6] and KedY2,[7] in the biosynthesis of the enediynes C-1027, maduropeptin and kedarcidin, respectively. We also reported a Type II adenylation enzyme, LnmQ, and Type II PCP, LnmP, initiating the biosynthesis of leinamycin.[8] Together these studies firmly established the existence of functional Type II NRPSs.

The interactions between a PCP and cognate catalytic domains are known to be very important in a few Type I systems.[9,10] In a typical round of peptide bond formation a PCP must interact with an A-domain and the acceptor site of an upstream condensation domain (C-domain) and acceptor donor site of a downstream C-domain. In Type I NRPS these contacts are likely to be very important to ensure that a PCP only interacts with the appropriate catalytic domains. However, in Type II systems the interactions may be

governed by additional features, and the current lack of study of these interactions hampers future efforts for combinatorial biosynthesis.

Structural studies can reveal the features of CPs that compose binding interfaces. While numerous structural studies of Type I and Type II ACPs from PKS and FAS [9 crystal and 21 nuclear magnetic resonance (NMR) reports] have been performed, there are limited studies of NRPS Type I PCPs (4 crystal and 4 NMR) and no reports of Type II PCPs. There is an NMR structure of the D-alanine carrier protein, DltC; however, it does not function as a traditional PCP, as it transfers D-alanine across the membrane for cell wall biosynthesis, and is phosphopantetheinylated by holo-ACP synthase.[11] There are two crystal structures of an aryl-carrier protein (ArCP) from the enterobactin NRPS biosynthetic pathway.[12,13] One as a domain in the full-length EntB, isochorismatase-ArCP, where there are no interacting partners, and may mimic a Type II PCP.[12] The other structure is the EntB ArCP fused to the C-terminal end of the 2,3-dihydroxybenzoic acid adenylation domain, EntE, which functions as a mimic of an NRPS A-PCP didomain.[13] NRPS, crystallization studies revealed the interactions between an A-domain and PCP in a didomain protein named PA1221 in *Pseudomonas aeruginosa*,[14] PCP and donor site of the C-domain for the SrfA-C module from *Bacillus subtilis*,[15] and PCP thioesterase (TE) in a didomain protein from EntF in *E. coli*.[16] In the crystal structure of a PCP-C didomain protein from TycC-5 in *B. subtilis*, the PCP is found in a noncatalytic orientation at the acceptor site that does not reveal insight into the binding interface.[17] Together these structures cover most of the common Type I PCP interactions in NRPS, except for PCP interacting with the C-domain acceptor site in a meaningful way, and for Type II PCP interactions.

NMR structures of Type I PCPs isolated from the NRPS reveal three interchanging conformations, the A state for apo-PCP (without ppant), H state for holo-PCP and a conformation shared by both apo- and holo-PCP, the A/H state.[18,19] These states led the authors to conclude that large conformational changes were necessary for PCP interaction with catalytic partners. NMR analysis of interactions in a PCP-TE didomain protein from EntF[20] and from a protein complex between a Type II TE, SrfTEII, and Type I PCP, TycC-3,[21] also led to the conclusion that the PCP and/or the catalytic domains underwent large conformational changes upon interaction.

Here we present the prototypical crystal structure of a Type II PCP, BlmI. Primary sequence and structural comparisons reveal that ACPs and PCPs are highly divergent in sequence within and between the classes, yet retain almost identical folds and overall structures. This study reveals that BlmI adopts a conformation similar to previously characterized ACPs and PCPs and is in the A/H state in the crystalline form. The observation that all crystallized PCPs to date are in the A/H state, suggests that the heterogeneous structures seen in the NMR experiments result from excising the PCP from the native context. The molecular surface of BlmI is overall neutral with a basic N-terminal tail. Comparison of the PCP-catalytic domain interfaces reveals a common binding surface on one side of the PCP, which will be important for designing future protein–protein interactions in combinatorial biosynthetic approaches.

MATERIALS AND METHODS

Gene cloning and expression and protein purification

Project information and reporting to the PSI Target Track Database used Sesame as the laboratory information system.[22] The full length *blmI* gene from *S. verticillus* ATCC 15003 (NCBI accession, gi: 5326870; locus version: AAD42077.1) was amplified by PCR from genomic DNA with KOD Hot Start DNA polymerase and the following forward and reverse primers: 5′-TTATCCACTTCCAATGT TACCGGTCCCGCTCCCC-3′ and 5′-TACTTCCAATCCA ATGCCATGAGCGCCCCGCGGG-3′, respectively. The amplification buffer was supplemented with betaine to final 2.5 *M* concentration. The PCR product was purified, T4 polymerase treated,[23] cloned into pMCSG57 according to ligation-independent procedures,[24,25] and transformed into *E. coli* BL21(DE3)-Gold strain (Stratagene). For protein production, a bacterial culture was grown at 37°C, at 190 rpm in 1 L of enriched M9 medium[26,27] until it reached optical density$_{600}$ ≈ 1.0. After air-cooling to 4°C over 60 min, methionine biosynthetic inhibitory amino acids (25 mg/L each L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine, L-pheynlalanine) and 90 mg/L of L-selenomethionine (Medicillin, catalog number MD045004D) were added. Gene expression was then induced by addition of isopropyl-β-D-thiogalactoside to 0.5 m*M*. The cells were incubated overnight at 18°C, harvested and resuspended in lysis buffer (500 m*M* NaCl, 5% (v/v) glycerol, 50 m*M* 2-[4–(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 8.0, 20 m*M* imidazole, and 10 m*M* β-mercaptoethanol). Cells were disrupted by lysozyme treatment (1 mg/mL) and sonication, and the insoluble cellular material was removed by centrifugation. The Se-Met protein was purified by Ni-NTA affinity chromatography and the AKTAxpress system (GE Health Systems), with washing with lysis buffer and elution with lysis buffer containing 250 m*M* imidazole. The protein was desalted into lysis buffer without imidazole. This was followed by the cleavage of the His$_6$-tag using recombinant His$_6$-tagged tobacco etch virus protease at 4°C over 48 h, with an additional Ni-NTA purification to remove the protease, uncut protein, and affinity tag. These procedures produced a protein with a Ser-Asn-Ala peptide preceding the N-terminal Met. The pure protein was concentrated to 24.7 mg/ mL using Amicon Ultra-15 concentrators (Millipore, Bedford, MA) in 20 m*M* HEPES pH 8.0 buffer, 250 m*M* NaCl, and 2 m*M* dithiothreitol. Protein concentrations were determined from the absorbance at 280 nm using a molar absorption coefficient (ϵ_{280} = 5500 M^{-1} cm^{-1}) calculated by using the method developed by Gill and von Hippel.[28] A second batch of protein was purified for optimization with reductive alkylation and partial proteolysis.[26] The concentration of protein samples used for crystallization was 14.4 mg/mL for reductive methylated, 14.6 mg/mL for reductive ethylated, and 24.8 mg/mL for partial proteolysis (with chymotrypsin, trypsin, and thermolysin). Individual aliquots of purified protein were stored at −80°C until needed.

Protein crystallization

BlmI was screened against several commercially available crystallization conditions including, MCSG-1–3 (Microlytic, MA) at 24°C and 4°C for the unmodified protein, MCSG-1–4 at 16°C for the reductive alkylated proteins, and PEG/Ion HT (Hampton Research Corp., CA) at 16°C for the partially proteolyzed proteins. Vapor-diffusion sitting drops containing 0.4 μL of protein and 0.4 μL of screening solution and were set up in 96-

well CrystalQuick plates (Greiner Bio-One, Monroe, NC) over wells containing 140 μL screening solution using a Mosquito liquid dispenser (TTP Labtech, Cambridge, MA). Crystals were obtained under several conditions. These crystals were cryoprotected by a brief transfer to the crystallization solution plus 10% ethylene glycol and flash frozen in liquid nitrogen. They were analyzed with synchrotron X-ray radiation near the selenium edge at the Structural Biology Center, sector 19-BM at the Advanced Photon Source.

Data collection, structure determination, and refinement

One crystal was subject to an X-ray fluorescence scan to determine the selenium K absorption edge. Data were collected to 2.1 Å at the peak wavelength (0.97931 Å) for the single best BlmI crystal and processed using the HKL-3000 system.[29] The data-collection strategy was optimized using the HKL-3000 strategy function. One hundred one degree oscillation images were collected at 100 K and a detector (ADSC Q210r) distance of 200 mm. Se-SAD was used to solve the structure as implemented in the HKL-3000 structure solution package. A single Se site was located with SHELXCD, and phases were calculated and improved with iterative rounds of MLPHARE and DM. An initial model was built with ARP/wARP.[30] The model was completed manually with Coot[31] and the structure refined with Refmac5[32] from the CCP4 suite.[33] The structure was analyzed for Ramachandran outliers with MolProbity.[34] The final atomic coordinates and amplitudes have been deposited in the Brookhaven Protein Data Bank, entry code 4NEO.

RESULTS AND DISCUSSION

Bioinformatics analysis of BlmI reveals Type II PCPs are a diverse family of proteins

Typically, the assignment of a CP as either ArCP, PCP, or ACP, comes from the carboxylic acid extending unit that is loaded to it, which is straight forward in Type I systems as the CP is next to the cognate loading domain in primary sequence. However, in the case of Type II CPs, there is no way to infer what enzyme is loading it, making assignment based on sequence homology, or through deductive reasoning in the context of a gene cluster, as for LnmP, SgcC2, MdpB2, and KedY2. In the BLM gene cluster the function of BlmI is not apparent, thus sequence homology was used to deduce the function of BlmI, with subsequent biochemical characterization. 3 Since the report of BlmI as a functional Type II PCP, many others have been discovered, warranting a discussion of the homology between these members.

BlmI shares sequence homology to CPs (Fig. 1). The most closely related proteins are ZbmI (55/68, percent identity/similarity) and TlmI (55/70) from the zorbamycin and tallysomycin biosynthetic pathways, respectively. The conservation of these genes suggests that they have a function in these related biosynthetic pathways, although that function remains to be elucidated. BlmI sequence comparisons fall in the “twilight zone,”[35] where functional assignment is a problem; however, phylogenetic analysis overcomes limitations of pairwise analysis, as shown in Figure 1, where BlmI clades with PCPs. Further analysis reveals that BlmI shares no more homology to CPs in the BLM gene cluster (average percent identity/similarity of 25 ± 3/40 ± 2), than to the other PCPs in Figure 1 (average percent identity of 25 ± 7). BlmI shares lower homology with the ACPs in Figure 1 (average percent identity of

19 ± 3), supporting our assignment of BlmI as a PCP. Interestingly, BlmI is as homologous to the Type II PCPs LnmP and DltC as it is to the ACPs, yet it is as homologous to SgcC2, MdpC2 and KedY2 as it is with the Type I PCPs, which serves to highlight the diversity within the Type II PCP family.

Crystallization, data collection, and refinement

BlmI crystallized in many conditions; however, the best crystals were obtained from reductively methylated BlmI in 0.1 *M* Na-CHES pH 9.5, 30% PEG 400 by vapor diffusion at 16°C, which grew to approximately 200 μm × 180 μm × 100 μm within one week. BlmI crystallized in the trigonal space group $P3_221$ with cell dimensions of $a = b = 73.44$ Å × $c = 43.33$ Å and one molecule per asymmetric unit. The structure was solved by Se-SAD with phasing from a single selenomethionine. Data collection and refinement statistics can be found in Table I. The final refinement statistics are reasonable with an R_{work} of 0.175 and R_{free} of 0.210, and resolution of 2.10 Å. Electron density revealed the presence of residues 3– 85, with residues −2 to 2 and 86 to 90 missing. One molecule of ethylene glycol and one molecule of PEG are visible in the electron density. There is no density for a ppant moiety, which agrees with the observation that *E. coli* PPTases do not efficiently modify BlmI. It is not immediately apparent from the structure why reductive methylation leads to the best crystals, as neither the single lysine nor the N-termini are involved in crystal contacts. Crystallization contacts cover 23.8% of the available surface area as calculated by PISA.[36] The first characterized Type II CPs were ACPs, which resisted crystallization due to their highly negatively charged surfaces.[37] This suggested that other Type II CPs would resist crystallization. However, BlmI is relatively neutral, which may have contributed to the ease in which these crystals were obtained.

The overall structure of BlmI reveals structural similarity to known ACPs and PCPs

BlmI is comprised of a four α-helical bundle similar to other structures of ACPs and PCPs (Figs. 1 and 2). However, the conformation of the loop between α1 and α2 is different between the ACPs and PCPs, in that a short helix immediately follows α1 in ACPs. Helix α1 of BlmI is N-terminally extended by 9 or 10 residues (Fig. 1) and is stabilized by multiple contacts between symmetry mates that may not be structurally relevant in solution. The conserved serine that is phosphopantetheinylated resides at the N-terminal end of α2 (Fig. 1). Dali[38] alignments reveal that BlmI has high structural homology to other crystallized CPs, including the *E. coli* fatty acid ACP [EcACP (PDB 1T8K, root-mean-square deviation (RMSD) 2.0 Å)], *B. subtilis* fatty acid ACP [BsACP (PDB 2X2B, RMSD 1.9 Å)], *E. coli* EntB ArCP (PDB 3RG2, RMSD 2.8 Å; PDB 2FQ1, RMSD 2.8 Å), TycC-5 PCP (PDB 2JGP, RMSD 1.6 Å), SrfA-C PCP (PDB 2VSQ, RMSD 1.8 Å), EntF PCP (PDB 3TEJ, RMSD 2.0 Å), and PA1221 PCP (PDB 4DG9, RMSD 2.6 Å). Structurally, BlmI is most similar to the Type I PCPs, suggesting BlmI can act as a model for the whole PCP family. Overall, this analysis reveals that although ACPs, PCPs, and ArCPs carry different substrates, their overall secondary and tertiary structures are very similar.

Diverse catalytic domains form interactions on the same face of the PCP that is highly variable

As there are crystal structures revealing the interaction of PCPs with the A-domain, donor site of the C-domain and TE, we examined if there was more than one surface of the PCP that interacts. PISA analysis of the interactions reveals, (i) the SrfA-C PCP-C interaction covers 506 $Å^2$ [11% of the PCP accessible surface area (ASA]], (ii) the crystalline EntF PCP-TE interaction covers 781 $Å^2$ (19% of the PCP ASA), (iii) by NMR the EntF PCPTE interaction covers 951 $Å^2$ (18% of the PCP ASA), (iv) the PA1221 PCP-A interaction covers 688 $Å^2$ (16% of the PCP ASA), and (v) the EntB-EntE ArCP-A interaction covers 895 $Å^2$ (15% of the ArCP ASA). The PISA analysis was used to explore the regions of the PCP surface that interacted, by comparison of the ASA of each residue in complex and alone (Fig. 2). This analysis revealed that although the catalytic domains are different, the interactions are roughly on the same surface of the PCP.

This surface is comprised of the loop between α1 and α2, and residues emanating from α2 and α3. For PKS ACPs, α2, which carries the phosphopantetheinylated serine, has been called the “universal recognition helix” and is known to be important for interactions with catalytic domains.[39] Experimental evidence for the importance of the surfaces of α2 and α3 in PCP interactions comes from mutagenic studies of EntB and EntF, where mutagenesis of residues on these helices reduces or abolishes activity.[40–42] Although the overall structure is highly conserved with respect to the positions of the α-helices in CPs, the residues emanating from them is highly variable, creating unique shape and charge distributions (Fig. 3). For example, there is a crevice between α2 and α3 in PA1221, which leads to a very different surface than that of SrfA-C, which is highly protruding and positively charged (Fig. 3). The variable residues making up these surfaces likely serve as a specific recognition features for the catalytic domains. However, these surface features are not likely to be static as the side chains are free to sample other rotameric states, which likely promote specific interactions with other catalytic domains. Future efforts at combinatorial biosynthetic efforts with Type II PCPs will benefit from this analysis.

The A- and H-states observed in NMR are likely artifacts of PCP excision from the Type I modular NRPS context

When compared to the different states found in the NMR structures of TycC-3 we find that BlmI has the highest structural homology with the A/H state (RMSD 2.3 Å, 72 residues; Fig. 2). Furthermore, all of the other PCPs structures to date regardless of whether they carry a phosphopantetheinyl group or are interacting with a partner are found in the A/H state (Fig. 2). The NMR structure of EntF PCP-TE didomain also reveals that the PCP is in the A/H state.[20] The only other structure with the H state is the NMR structure of TycC-3 interacting with the Type II TE SrfTEII, which was used as the model for the state of the PCP.[21] There are no other structures representing the A state of TycC-3. Furthermore, the α3 helix which appears to be important for recognition is not apparent in either of the A or H states. The overwhelming structural evidence suggests that the NMR structures of the A and H states of TycC-3 could be artifacts caused by removing the domain from its native context. Therefore, future experiments aimed at understanding the molecular recognition of PCPs and catalytic domains should be modeled on the state seen in the structure of BlmI.

Acknowledgments

Grant sponsor: National Institute of General Medical Sciences Protein Structure Initiative (to GNP and BS); Grant number: GM094596; Grant sponsor: National Institute of General Medical Sciences Protein Structure Initiative (to MC, LB, JB, GB, and AJ); Grant number: GM094585; Grant sponsor: National Institutes of Health (to BS); Grant number: AI40475; Grant sponsor: U.S. Department of Energy, Office of Biological and Environmental Research (to MC and AJ); Grant number: DE-AC02-06CH11357.

REFERENCES

1. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006; 106:3468–3496. [PubMed: 16895337]
2. Crosby J, Crump MP. The structural role of the carrier protein-active controller or passive carrier. Nat Prod Rep. 2012; 29:1111–1137. [PubMed: 22930263]
3. Du L, Shen B. Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC 15003. Chem Biol. 1999; 6:507–517. [PubMed: 10421758]
4. Van Lanen SG, Dorrestein PC, Christenson SD, Liu W, Ju J, Kelleher NL, Shen B. Biosynthesis of the beta-amino acid moiety of the enediyne antitumor antibiotic C-1027 featuring beta-amino acyl-S-carrier protein intermediates. J Am Chem Soc. 2005; 127:11594–11595. [PubMed: 16104723]
5. Lin S, Van Lanen SG, Shen B. A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. Proc Natl Acad Sci USA. 2009; 106:4183–4188. [PubMed: 19246381]
6. Van Lanen SG, Oh T, Liu W, Wendt-Pienkowski E, Shen B. Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. J Am Chem Soc. 2007; 129:13082–13094. [PubMed: 17918933]
7. Lohman JR, Huang S-X, Horsman GP, Dilfer PE, Huang T, Chen Y, Wendt-Pienkowski E, Shen B. Cloning and sequencing of the kedarcidin biosynthetic gene cluster from Streptoalloteichus sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics. Mol Biosyst. 2013; 9:478–491. [PubMed: 23360970]
8. Tang G-L, Cheng Y-Q, Shen B. Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from *Streptomyces atroolivaceus* S-140. Discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly loadd-alanine. J Biol Chem. 2007; 282:20273–20282. [PubMed: 17502372]
9. Stachelhaus T, Huser A, Marahiel M. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chem Biol. 1996; 3:913–921. [PubMed: 8939706]
10. Weinreb PH, Quadri LE, Walsh CT, Zuber P. Stoichiometry and specificity of in vitro phosphopantetheinylation and aminoacylation of the valine-activating module of surfactin synthetase. Biochemistry. 1998; 37:1575–1584. [PubMed: 9484228]
11. Volkman BF, Zhang Q, Debabov DV, Rivera E, Kresheck GC, Neuhaus FC. Biosynthesis ofd-alanyl-lipoteichoic acid: the tertiary structure of apo-d-alanyl carrier protein. Biochemistry. 2001; 40:7964–7972. [PubMed: 11434765]
12. Drake EJ, Nicolai DA, Gulick AM. Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. Chem Biol. 2006; 13:409–419. [PubMed: 16632253]
13. 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. Chem Biol. 2012; 19:188–198. [PubMed: 22365602]
14. Mitchell CA, Shi C, Aldrich CC, Gulick AM. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry. 2012; 51:3252–3263. [PubMed: 22452656]

15. Tanovic A, Samel SA, Essen L-O, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. Science. 2008; 321:659–663. [PubMed: 18583577]
16. Liu Y, Zheng T, Bruner SD. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. Chem Biol. 2011; 18:1482–1488. [PubMed: 22118682]
17. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen L-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure. 2007; 15:781–792. [PubMed: 17637339]
18. Weber T, Baumgartner R, Renner C, Marahiel MA, Holak TA. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure. 2000; 8:407–418. [PubMed: 10801488]
19. Koglin A, Mofid MR, Löhr F, Schäfer B, Rogov VV, Blum M-M, Mittag T, Marahiel MA, Bernhard F, Dötsch V. Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science. 2006; 312:273–276. [PubMed: 16614225]
20. Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G. Dynamic thiolation-thioesterase structure of a nonribosomal peptide synthetase. Nature. 2008; 454:903–906. [PubMed: 18704088]
21. Koglin A, Löhr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Güntert P, Wagner G, Walsh CT, Marahiel MA, Dötsch V. Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. Nature. 2008; 454:907–911. [PubMed: 18704089]
22. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN, Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: Sesame. J Struct Funct Genomics. 2003; 4:11–23. [PubMed: 12943363]
23. Dieckman L, Gu M, Stols L, Donnelly MI, Collart FR. High throughput methods for gene cloning and expression. Protein Express Purif. 2002; 25:1–7.
24. Eschenfeldt, WH.; Stols, L.; Millard, CS.; Joachimiak, A.; Donnelly, MI. A family of LIC vectors for high-throughput cloning and purification of proteins. In: Doyle, SA., editor. Methods in Molecular biology: high throughput protein expression and purification. Vol. 498. Totowa, NJ: Humana Press; 2009. p. 105-115.
25. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990; 18:6069–6074. [PubMed: 2235490]
26. Kim Y, Babnigg G, Jedrzejczak R, Eschenfeldt WH, Li H, Maltseva N, Hatzos-Skintges C, Gu M, Makowska-Grzyska M, Wu R, An H, Chhor G, Joachimiak A. High-throughput protein purification and quality assessment for crystallization. Methods. 2011; 55:12–28. [PubMed: 21907284]
27. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Express Purif. 2006; 47:446–454.
28. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989; 182:319–326. [PubMed: 2610349]
29. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. Acta Crystallogr D. 2006; 62:859–866. [PubMed: 16855301]
30. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc. 2008; 3:1171–11719. [PubMed: 18600222]
31. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D. 2004; 60:2126–2132. [PubMed: 15572765]
32. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D. 1997; 53:240–255. [PubMed: 15299926]
33. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas J, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. Acta Crystallogr D. 2011; 67:235–242. [PubMed: 21460441]

34. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D. 2010; 66:12–21. [PubMed: 20057044]
35. Rost B. Twilight zone of protein sequence alignments. Protein Eng Des Sel. 1999; 12:85–94.
36. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372:774–797. [PubMed: 17681537]
37. Qiu X, Janson CA. Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions. Acta Crystallogr D. 2004; 60:1545–1554. [PubMed: 15333924]
38. Holm L, Kääriäinen S, Rosenström P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics. 2008; 24:2780–2781. [PubMed: 18818215]
39. Weissman KJ, Hong H, Popovic B, Meersman F. Evidence for a protein–protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase. Chem Biol. 2006; 13:625–636. [PubMed: 16793520]
40. Lai JR, Fischbach MA, Liu DR, Walsh CT. A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. Proc Natl Acad Sci USA. 2006; 103:5314–5319. [PubMed: 16567620]
41. Lai JR, Fischbach MA, Liu DR, Walsh CT. Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. J Am Chem Soc. 2006; 128:11002–11003. [PubMed: 16925399]
42. Zhou Z, Lai JR, Walsh CT. Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. Chem Biol. 2006; 13:869–879. [PubMed: 16931336]
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]
44. Gouet P, Courcelle E, Stuart D, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999; 15:305–308. [PubMed: 10320398]
45. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28:2731–2739. [PubMed: 21546353]

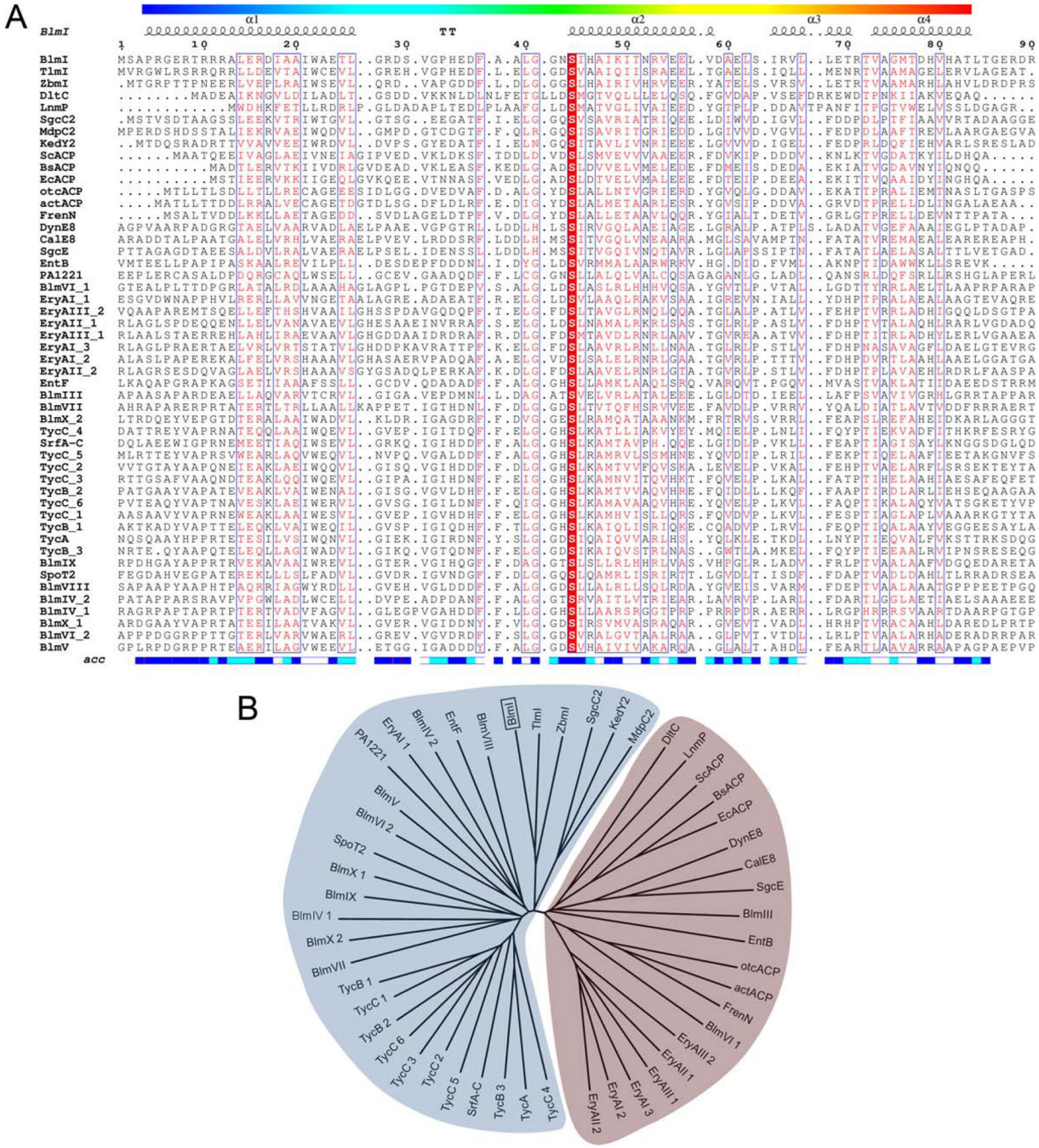

Figure 1.

Sequence alignment and phylogeny of selected ACPs and PCPs. (A) Secondary structure of BlmI is depicted above a sequence alignment. Below is the relative accessible surface area of each residue; white are buried residues, cyan are intermediate, blue are accessible. Completely conserved residues are highlighted with a red background and similar residues are colored red. Alignment created with MUSCLE[43] and rendered with ESPript.[44] (B) Phylogenetic tree revealing BlmI clades with PCPs. The ACPs clade together, and are marked with a bracket. Phylogeny created using the neighbor-joining method and rendered

with MEGA5.[45] PCP or ACP abbreviation (NCBI accession): BlmI (AAD42077), BlmIII (AAG02365), BlmIV (AAG02364), BlmV (AAG02360), BlmVI (AAG02359), BlmVII (AAG02358), BlmVIII (AAG02357), BlmIX (AAG02356), BlmX (AAG02355), ZbmI (ACG60769), ZbmIII (ACG60771), ZbmIV (ACG60772), ZbmV (ACG60775), ZbmVI (ACG60776), ZbmVIIa (ACG60777), ZbmVIIb (ACG60778), ZbmVIII (ACG60781), ZbmIX (ACG60773), ZbmX (ACG60782), TlmI (ABL74952), TlmIII (ABL74946), TlmIV (ABL74945), TlmV (ABL74941), TlmVI (ABL74940), TlmVII (ABL74939), TlmVIII (ABL74938), TlmIX (ABL74937), TlmX (ABL74936), TycA (AAC45928), TycB (AAC45929), TycC (AAC45930), SrfA-C (CAA49818), SgcC2 (AAL06679), KedY2 (AFV52209), MdpC2 (ABY66003), SpoT2 (ABP55169), EntF (AAC73696), EntB (AAC73687), DltC (AAB17659), PA1221 (AAG04610), EryAI (AAV51820), EryAII (AAV51821), EryAIII (AAV51822), FrenN (AAC18109), otcACP(AAA03334), actACP (CAC44202), ScACP (CAA60201), BsACP (CAB13465), EcACP (AAC74178), CalE8 (AAM94794), DynE8 (AAN79725), and SgcE (AAL06699).

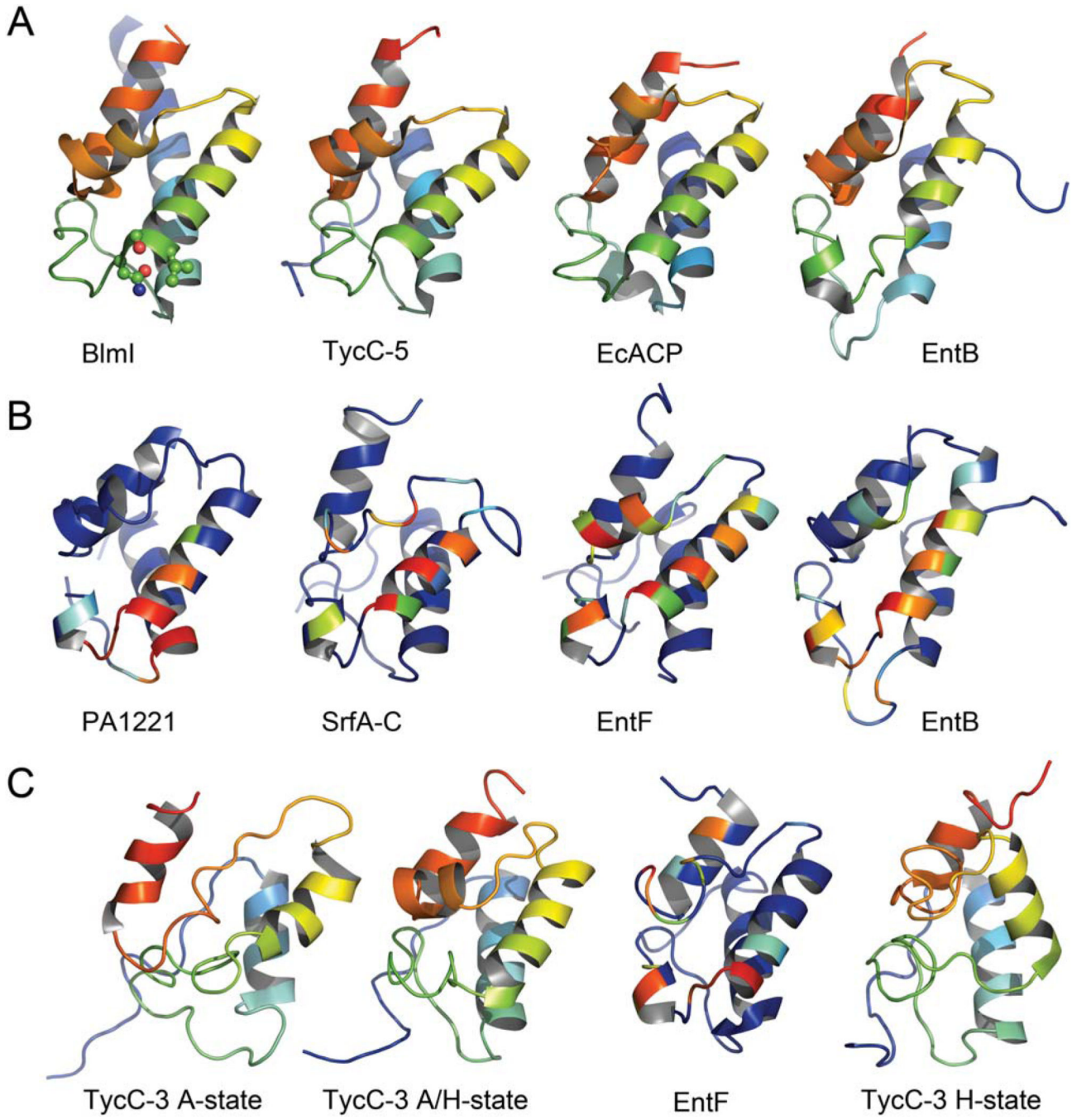

Figure 2.

Structures of the Type II PCP BlmI in comparison to other CPs. All structures are aligned to BlmI using DaliLite,[38] except the states of TycC3, which were aligned in Pymol (Schrödinger, LLC). (A) Crystal structures of apo-CPs with coloring from blue on the N-terminus and red on the Cterminus, as depicted on the secondary structure shown in Figure 1. The BlmI Asn-Ser-Ile ppant site is shown in ball-and-stick at the beginning of α2 for reference. (B) Crystal structures of PCPs interacting with catalytic domains. The PCPs are colored according to the burial of accessible surface area for each residue, with 0% burial in

blue to 100% burial in red. The strictly conserved serine has been artificially set to 100% burial. (C) NMR structures of PCPs. TycC-3 in the A, H, and A/H states are colored as in (A), whereas the EntF PCP is colored as in (B). The NMR structure of the EntF PCP is lacking defined secondary structure for α3, yet the remainder of the structure is almost identical to the crystal structure.

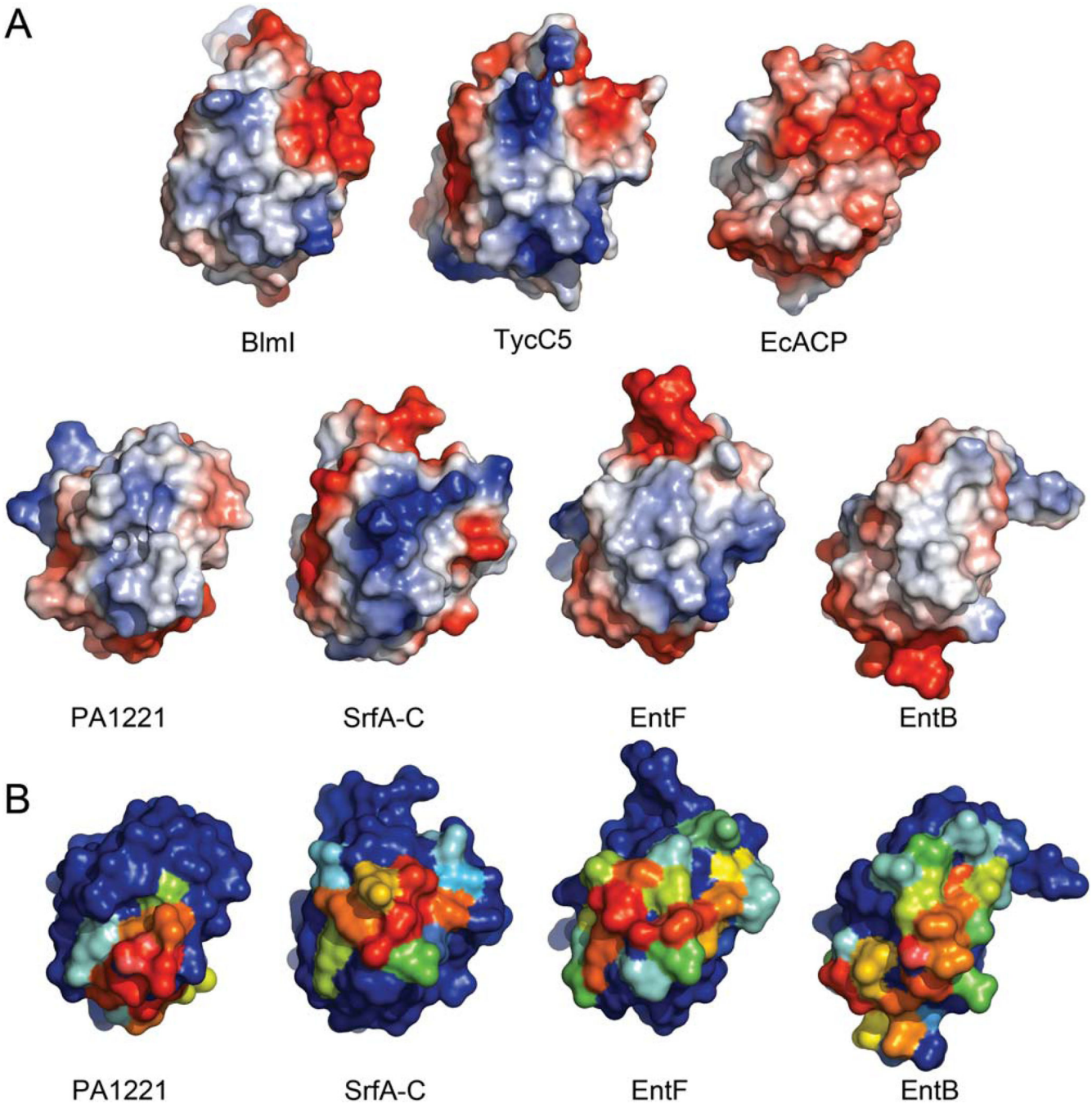

Figure 3.

Surfaces of selected CPs. (A) Electrostatic surfaces with negative charge in red and positive charge in blue and neutral in white. All of the structures are in the same orientation as in Figure 2. (B) The surface of the PCPs with residues buried in interfaces with catalytic domains. The coloring scheme is the same as in (B).

Table I

Data Collection and Refinement Statistics

Space group	$P3_221$
Unit cell (Å)	$a = b = 73.44$, $c = 43.33$
Wavelength (Å)	0.97931
Highest resolution bin (Å)	2.14–2.10
R_{merge} (%)	7.0 (42.9) [5.1 (39.4)]
Completeness (%)	99.7 [99.7]
I/σ_I	12.5 (3.2)
Redundancy	7.1 [3.8]
Wilson B-factor	44.4
Phasing and refinement	
Resolution (Å)	28.0–2.1
Phasing power	1.3
Number of reflections	8041
R_{work}/R_{free} (%)	17.5/21.0 (21.8/22.5)
Number of atoms	
Protein	596
Ligand/ion	11
Water	30
B-factors	
Protein	62.6
Water	68.0
RMSD bond (Å)	0.016
RMSD angle (Å)	1.777
Ramachandran plot (%)	
Most favored	97.1
Allowed	2.9
Disallowed	0
PDB ID	4NEO

$R_{merge} = \Sigma hkl \Sigma i |I_i - \langle I \rangle| / \Sigma hkl \Sigma i |\langle I \rangle|$, where I_i is the intensity for the ith measurement of an equivalent reflection with indices h, k, and l.

$R_{work} = (\Sigma||F_O| - |F_C|)/\Sigma||F_O|$, where F_O and F_C denote observed and calculated structure factors, respectively.

R_{free} was calculated with 5% of the data excluded from refinement.

Numbers in parentheses are for the highest resolution bin.

Numbers in brackets are statistics with the Bijvoet pairs unmerged.