J|A|C|S
COMMUNICATIONS

Published on Web 07/29/2006

Localized Protein Interaction Surfaces on the EntB Carrier Protein Revealed by Combinatorial Mutagenesis and Selection

Jonathan R. Lai,[†] Michael A. Fischbach,[†,‡] David R. Liu,[‡] and Christopher T. Walsh*,[†]

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received May 9, 2006; E-mail: christopher_walsh@hms.harvard.edu

Carrier protein domains (80 to 100 residues in length) are central to the polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) enzymatic assembly lines that produce many metabolites of medicinal interest.[1,2] Because substrates for biosynthetic operations are presented on carrier proteins as covalently attached thioesters (through a 4′-phosphopantetheine cofactor), a detailed understanding of protein−protein interactions between carrier proteins and other domains is required for reprogramming of NRPS/PKS machinery.[3,4] However, little is known about how much of a particular carrier protein's surface participates in interdomain interactions and whether different domains recognize separate regions of a carrier protein. Here we report the identification of a protein interaction surface on the EntB aryl carrier protein (EntB−ArCP) for phosphopantetheinyl transferases (PPTases), such as EntD and Sfp, by combinatorial mutagenesis and selection. This protein interaction surface is highly localized, consisting of just two surface residues, and is distinct from the previously identified interface for the downstream elongation module, EntF.

Enterobactin (**1**) is an iron-chelating siderophore produced by *Escherichia coli* upon iron starvation.[5] The enterobactin synthetase consists of four protein components, EntBDEF, that use three molecules each of 2,3-dihydroxybenzoate (DHB) and serine to produce **1** via NRPS logic (Figure 1a).[5a,b] The ArCP domain of EntB (EntB−ArCP) must participate in three well-timed protein−protein interactions during the biosynthetic reaction cascade (Figure 1b): (i) with EntD (or other PPTases) during phosphopantetheinylation; (ii) with EntE during activation of DHB and thiolation onto the phosphopantetheine arm of holo-EntB−ArCP; and (iii) with EntF during condensation of DHB (presented on the EntB pantetheine) with serine. We recently reported an in vivo selection for EntB function by plating *E. coli* onto iron-deficient media.[6] This strategy allowed us to rapidly process large ($>10^6$) EntB mutant libraries for their ability to support production of **1** in vivo. We used this selection together with combinatorial mutagenesis of C-terminal regions of EntB to map an interaction interface on EntB−ArCP for EntF.

Using the EntB crystal structure as our guide (Figure 2a),[7] we designed and prepared three libraries of mutants that collectively span the N-terminal portions of EntB−ArCP: helix 1 (library H1) and the long loop between helix 1 and helix 2 (libraries L1A and L1B). In library H1, noncore residues in helix 1 were allowed to vary between WT and Ala by partial codon variation (due to the degeneracy of the genetic code, a 3rd and 4th residue was permitted at some positions).[8,9] For libraries L1A and L1B, residues in regions 225−235 and 236−244, respectively, were subjected to a similar randomization scheme. Selection for clones that produce **1** was then

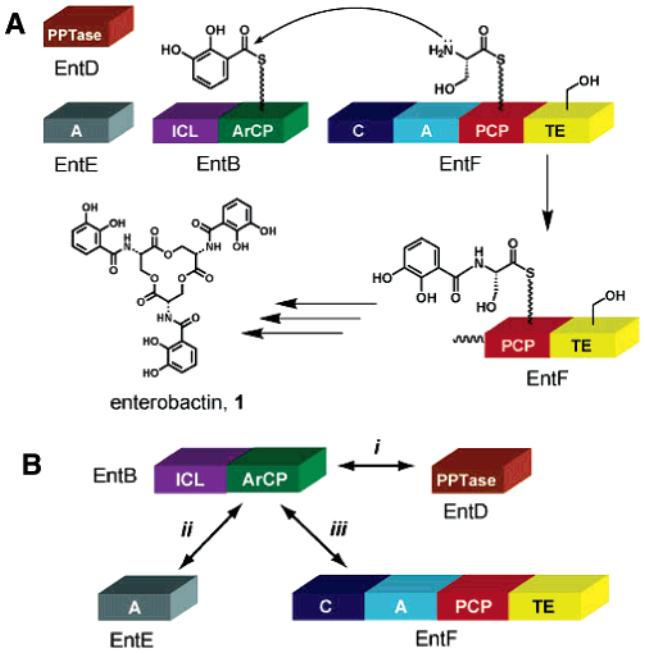

Figure 1. (A) Schematic of the enterobactin synthetase, consisting of four proteins: EntBDEF. Abbreviations for domain functions not specified in the text: A, adenylation; ICL, isochorismate lyase; C, condensation; PCP, peptidyl carrier protein; TE, thioesterase. (B) Protein−protein interactions required for enterobactin production. EntB−ArCP must contact (i) EntD (or other PPTases), (ii) EntE, and (iii) EntF at various points during the biosynthetic cycle.

achieved by plating the libraries onto minimal media made iron-deficient by the addition of the metal chelator 2,2′-dipyridyl.[6]

Over 65 nonredundant surviving clones from each library were isolated and sequenced. From these data, WT/Ala ratios for each position, defined as the number of times WT was observed to the number of times Ala was observed,[8] were determined. The degree of conservation for each residue was classified as high (WT/Ala ≥ 20), intermediate (6 < WT/Ala < 20), or low (WT/Ala ≤ 6). Only five residues fell into the intermediate or high conservation categories (Table 1). Figure 2b shows the surface of EntB−ArCP color coded according to these classifications, including data compiled from our previous report.[6]

The sequencing results revealed that the residues G242 and D244 form a conserved, surface-exposed patch that immediately precedes the phosphopantetheinylated S245 (Figure 2b). This cluster corresponds to the interaction surface on EntB−ArCP for PPTases, such as EntD.[5c] We found that EntB−ArCP G242A or D244R mutants are poor substrates for EntD.[9] The ArCP mutant D244A is still efficiently phosphopantetheinylated by EntD in vitro but cannot be recognized by the broad substrate PPTase Sfp from *B. subtilis*.[9]

[†] Department of Biological Chemistry and Molecular Pharmacology.
[‡] Howard Hughes Medical Institute and Department of Chemistry and Chemical Biology.

10.1021/ja063238h CCC: $33.50

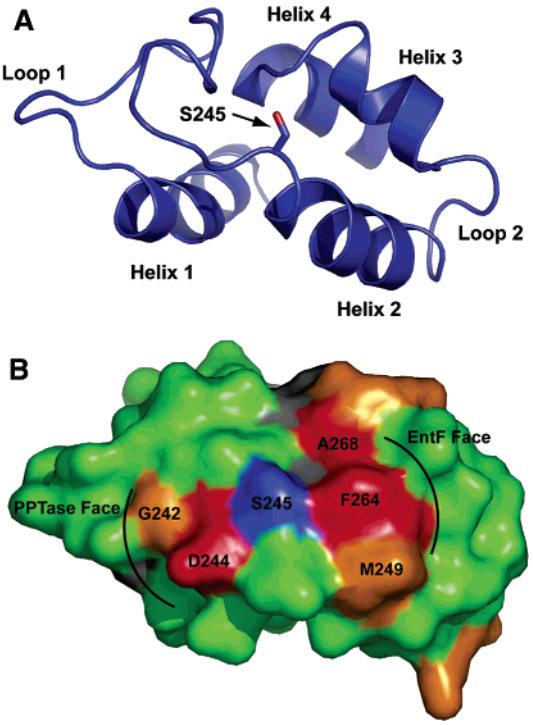

Figure 2. (A) Ribbon representation of the EntB−ArCP structure. (B) Surface of EntB−ArCP color coded for degree of conservation, where red is high, orange is intermediate, and green is low. Serine 245 is shown in blue. The residues that comprise the differential PPTase and EntF interaction faces are indicated.

Table 1. WT/Ala Ratios for Selected Residues on EntB−ArCP

residue	WT/Ala	residue	WT/Ala
D234	13.7	G242	17.8
L238	>64.0	L243	46.0
		D244	24.0

Mutation of this conserved Asp, which immediately precedes the phosphopantetheinylated serine, has been reported to disrupt PPTase recognition in EntB and other systems.[2b,7] The interaction surface on EntB−ArCP for PPTase recognition is distinct from that of EntF, which we have previously characterized.[6] Each interaction surface is located on a separate side of S245, and each one is comprised of residues from different structural elements. These observations suggest that PPTases and EntF recognize distinct and highly localized interaction faces on EntB−ArCP.[10] Therefore, it should be possible to alter the recognition properties of EntB−ArCP for one of these synthetase components while leaving interactions with the other unaffected.

Three other residues displayed intermediate or high conservation: L238, L243, and D234. The residues L238 and L243, located on the loop, point toward the carrier protein core.[9] The high WT/Ala ratios at these positions are likely due to the role of the Leu side chain in maintaining the stability of the EntB−ArCP fold. Aspartate at position 234 was preferred about 14-fold over Ala, presumably because D234 participates in charge−charge interactions with K215 and R219 of helix 1.[9]

Collectively, we now have scanned ~80% of the EntB−ArCP surface using a combinatorial mutagenesis and selection scheme. Overall, the majority of EntB−ArCP surface residues were highly tolerant to mutation.[9] Thirty-six of 44 total surface residues that were examined here and in our earlier report showed low conservation.[6,9] This result implies that the majority of EntB−ArCP surface residues are not involved in interactions with other synthetase components.

We have not identified an interaction surface on EntB−ArCP for the upstream adenylation domain, EntE. The format of the genetic selection may be such that the activation of DHB and subsequent thiolation onto holo-EntB is not rate-limiting for in vivo production of **1**. Nonetheless, we and others have found that aryl carrier proteins from EntBDEF and related synthetases are surprisingly impervious to mutation while maintaining their ability to be recognized by free-standing adenylation domains in vitro.[3c,7] Thus, the interface for EntE may be malleable for presentation of aminoacyl-*O*-AMP to the pantetheinyl arm of EntB.

The results presented herein suggest that future efforts to reprogram NRPS and PKS assembly lines by engineering selective carrier protein interactions should focus on interaction "hot spots", similar to those on EntB−ArCP for EntD/Sfp and EntF. This process could be facilitated by directed evolution approaches that target these regions. It remains to be determined if all carrier proteins utilize similar interaction faces; studies to test this hypothesis are currently underway.

Acknowledgment. This work was funded by the NIH (AI042738, GM020011, and GM047467 to C.T.W. and GM065400 to D.R.L.) and the Howard Hughes Medical Institute (D.R.L.). J.R.L. was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation. M.A.F. acknowledges a predoctoral fellowship from the Hertz Foundation. We thank Andrew Gulick for helpful discussions.

Supporting Information Available: Full sequence analysis data, biochemical assays, and procedures for protein purification and library production. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Walsh, C. T. *Science* **2004**, *303*, 1805−1810.
(2) (a) Koglin, A.; Mofid, M. R.; Lohr, F.; Schafer, B.; Rogov, V. V.; Blum, M. M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dotsch, V. *Science* **2006**, *312*, 273−276. (b) Worsham, L. M. S.; Earls, L.; Jolly, C.; Langston, K. G.; Trent, M. S.; Ernst-Fonberg, M. L. *Biochemistry* **2003**, *42*, 167−176.
(3) (a) Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Biochemistry* **2001**, *40*, 2326−2331. (b) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723−731. (c) Marshall, C. G.; Burkhart, M. D.; Meray, R. K.; Walsh, C. T. *Biochemistry* **2002**, *41*, 8429−8437.
(4) (a) Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Science* **1999**, *284*, 482−485. (b) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 275−280.
(5) (a) Crosa, J. H.; Walsh, C. T. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 223−249. (b) Gehring, A. M.; Mori, I.; Walsh, C. T. *Biochemistry* **1998**, *37*, 2648−2659. (c) Gehring, A. M.; Bradley, K. A.; Walsh, C. T. *Biochemisty* **1997**, *36*, 8495−8503.
(6) Lai, J. R.; Fischbach, M. A.; Liu, D. R.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5314−5319.
(7) Drake, E. J.; Nicolai, D. A.; Gulick, A. M. *Chem. Biol.* **2006**, *13*, 409−419.
(8) Weiss, G. A.; Watanabe, C. K.; Zhong, A.; Goddard, A.; Sidhu, S. S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8950−8954.
(9) See Supporting Information.
(10) High-resolution structural data on protein−protein complexes involving EntB−ArCP will be required to confirm these interaction interfaces.

JA063238H