

Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases

Niran Roongsawang^a, Siew Ping Lim^a, Kenji Washio^a, Kazufumi Takano^b,
Shigenori Kanaya^b, Masaaki Morikawa^{a,*}

^a Division of Biosphere Science, Graduate School of Environmental Science, Hokkaido University, Kita-ku, Sapporo 060-0810, Japan

^b Department of Material and Life Science, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan

Received 20 June 2005; received in revised form 26 August 2005; accepted 28 August 2005

First published online 9 September 2005

Edited by M.Y. Galperin

Abstract

Condensation (C) domains in the nonribosomal peptide synthetases are capable of catalyzing peptide bond formation between two consecutively bound various amino acids. C-domains coincide in frequency with the number of peptide bonds in the product peptide. In this study, a phylogenetic approach was used to investigate structural diversity of bacterial C-domains. Phylogenetic trees show that the C-domains are clustered into three functional groups according to the types of substrate donor molecules. They are L-peptidyl donors, D-peptidyl donors, and N-acyl donors. The fact that C-domain structure is not subject to optical configuration of amino acid acceptor molecules supports an idea that the conversion from L to D-form of incorporating amino acid acceptor occurs during or after peptide bond formation. L-peptidyl donors and D-peptidyl donors are suggested to separate before separating the lineage of Gram-positive and Gram-negative bacteria in the evolution process.

© 2005 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Nonribosomal peptide synthetases; Nonribosomal peptides; Condensation domain; D-Amino acid, Cyclic lipopeptide

1. Introduction

Nonribosomal peptide synthetases (NRPS) are modular mega-multifunctional enzymes that synthesize an incredibly diverse set of biological active peptides or cyclic lipopeptides [1]. They are antibiotics, biosurfactants, siderophores, and immunosuppressants, as well as anti-tumor and antiviral agents. These valuable biomolecules carry important medical and biotechnological applications. The presence of D-amino acid residues is a hallmark of nonribosomal peptides. This creates structural diversity, constrains their stereochemical conformation, and resists to proteolysis [2]. Biosynthesis of nonriboso-

mal peptides occurs via the action of catalytic units of NRPS, referred to as modules, in the direction from N-terminal to C-terminal ends as well as by ribosome dependent mechanism. Each module is composed of functionally specific domains that are responsible for catalyzing different enzymatic activities [1,3]. They are adenylation (A), thiolation (T), and condensation (C) domains. A-domain is responsible for amino acid recognition and adenylation at an expense of ATP. Then, adenylation amino acid is covalently attached to a phosphopantetheine carrier of the adjacent T-domain. Peptide bond formation of two consecutively bound amino acids to the growing peptide chain is catalyzed by the C-domain. Additionally, a modifying epimerization (E) domain that catalyzes the conversion of L-amino acids to D-isomers is typically associated with

* Corresponding author: Tel./fax: +81 11 706 2253.

E-mail address: morikawa@ees.hokudai.ac.jp (M. Morikawa).

the module incorporating D-amino acid. After the synthesis of linear intermediate peptide, cyclization and release of the product peptide are carried out by a C-terminal thioesterase (Te) domain. This modular organization of NRPS allows artificial alteration of the protein template with an aim to reprogram it for the synthesis of novel peptides with improved properties [4].

Peptide bond formation by C-domains requires two adjacent aminoacyl molecules for initiation or one peptidyl donor and one aminoacyl acceptor molecules during elongation of peptide chain. C-domain is about 450 amino acids in length and is located at the N-terminus of each module. It coincides in frequency with the number of peptide bonds in the product peptide. Although A-domain primarily recognizes and activates specific substrate molecules based on the structure, it is probable that the structure of C-domains is also affected by the structure of substrates donor and acceptor molecules. In fact, an alignment of 74 C-domains indicated a clustering according to the types of reactions catalyzed and chirality of donor and acceptor molecules [5]. They are for L-L condensations (${}^L C_L$; ${}^{\text{donor}} C_{\text{acceptor}}$), D-L condensations (${}^D C_L$), δ -L- α -L condensations, condensations involving N-methylated amino acids, N-acylations, and reactions involving cysteinyl-peptides as donors. Biochemical characterization of the C₅ domain of tyrocidine synthetase has revealed that this domain is indeed a ${}^D C_L$ catalyst [6]. It was suggested that there is no ${}^L C_D$ and ${}^D C_D$ domains. Unfortunately, there were only a limited number of amino acid sequences available for putative ${}^L C_D$ and ${}^D C_D$ in the NRPS to support this idea. So, it remains unclear whether the putative ${}^{L/D} C_D$ constitutes a phylogenetic group different from ${}^{L/D} C_L$. We have recently cloned a gene cluster encoding arthrofactin synthetase (Arf), *arfABC*, from *Pseudomonas* sp. MIS38 [7]. The product lipopeptide, arthrofactin, contains seven D-form amino acid residues and Arf is expected to contain five putative ${}^D C_D$ and one putative ${}^L C_D$. We took advantage of this recent information to re-examine the relationship between the structure of C-domain and donor and acceptor molecules. It was found that ${}^L C_L$ and putative ${}^L C_D$, and ${}^D C_L$ and putative ${}^D C_D$, are not clearly separated in L-peptidyl donors and D-peptidyl donors, respectively. This topology of the tree allowed us to conclude that there is no C-domain, ${}^{L/D} C_D$, which specifically recognizes D-form acceptor molecules.

2. Materials and methods

The amino acid sequences of C-domains in various NRPS were retrieved from publicly accessible databases (<http://www.ncbi.nlm.nih.gov/entrez/>). We adopted the NRPS whose product structure has been determined. They are of anabaenopeptilides (Apd, DDBJ/EMBL/GenBank Accession No. AJ269505), arthrofactin (Arf,

AB107223), bacillomycin D (Bam, AY137375), bacitracin (Bac, AF007865), chloroeremomycin (Cep, T17483), complestatin (Com, AF386507), fengycin (Fen, AF023464), gramicidin (Grs, X61658), iturin A (Itu, AB050629), lichenysin (Lic, U95370), mycosubtilin (Myc, AF184956), nostopeptolide (Nos, AF204805), pyoverdin D (Pvd, U07359), pyoverdin S (Pvs, AE016863), surfactin (Srf, X70356), syringomycin (Syr, AF047828), syringopeptin22 (Syp, AF286216), tyrocidine (Tyc, AF004835), and a condensation enzyme from vibriobactin synthetase (VibH, AAD48879). The sequences of C-domains were aligned by the ClustalW program [8] provided by the DNA Data Bank of Japan, DDBJ.

Phylogenetic trees were constructed using the distance method and the character based method (protein parsimony, PROTPARS program) from the PHYLIP package v3.6 [9]. For the distance method, a distance matrix is calculated by protein distance matrix calculation (PROTDIST program) and the matrix is then transformed into a tree by NEIGHBOR program. In order to verify the accuracy of the tree, multiple data sets were generated using the SEQBOOT program with 1000 bootstrap replicates. A tree was built from each replicate with PROTDIST or PROTPARS program, then bootstrap values were computed with CONSENSE program. The phylogenetic tree was visualized with the Phylodendron program (<http://iubio.bio.indiana.edu/treeapp/treeprint-form.html>). Both methods gave similar tree topology, suggesting that the major clustering did not result from computational artefacts. Then, only the trees constructed by the distance method are presented in this paper.

3. Results and discussion

3.1. C-domains of arthrofactin synthetase

Arf is a NRPS responsible for the biosynthesis of a cyclic lipoundecapeptide, arthrofactin, in *Pseudomonas* sp. MIS38. It consists of three large subunit proteins (ArfA, ArfB, ArfC) that contain totally 11 functional modules. Arf represents a novel type of NRPS architecture that features tandem Te-domains but lacks the internal E-domain [7]. There are 10 peptide forming C-domains (namely from ArfA_C2 to ArfC_C5) in Arf. One additional C-domain (namely ArfA_C1) was identified in the first module of ArfA. This arrangement has also been reported for several NRPS [10–14]. It is suggested that the first amino acid could be initially acylated with a fatty acid in this domain [12]. When D-amino acid residues show up in the peptide products, the available L-isomers will be typically selected by A-domains and then epimerized by downstream E-domain. The peptide intermediate will be then presented to the

next C-domain that is a ${}^{\text{D}}\text{C}_\text{L}$ catalyst. It is unclear if there are any ${}^{\text{L}}\text{C}_\text{D}$ and ${}^{\text{D}}\text{C}_\text{D}$ domains [6]. According to the chirality of amino acid in arthrofactin, D-D-D-D-D-D-L-D-L-L-L, we propose that the second to sixth C-domains in Arf (ArfA_C2/ArfB_C1/ArfB_C2/ArfB_C3/ArfB_C4) are potentially D-specific for both donor and acceptor molecules (namely putative ${}^{\text{D}}\text{C}_\text{D}$). Similarly, ArfC_C1 and ArfC_C3 would be ${}^{\text{D}}\text{C}_\text{L}$, ArfC_C2 would be putative ${}^{\text{L}}\text{C}_\text{D}$, whereas ArfC_C4 and ArfC_C5 would be ${}^{\text{L}}\text{C}_\text{L}$, respectively.

A phylogenetic tree of C-domains in Arf clearly showed two clustered branches according to the chirality of donor molecules rather than acceptor molecules, namely L-peptidyl donors and D-peptidyl donors (Fig. 1). ${}^{\text{L}}\text{C}_\text{L}$ and putative ${}^{\text{L}}\text{C}_\text{D}$, and ${}^{\text{D}}\text{C}_\text{L}$ and putative ${}^{\text{D}}\text{C}_\text{D}$, are not clearly separated in L-peptidyl donors and D-peptidyl donors, respectively. This finding suggests the evolutionary relationship of C-domain for donor molecules rather than acceptors. ArfA_C1 that probably catalyzes N-acylation of the first amino acid is closely related to L-peptidyl donors but forms an independent branch.

Sequence alignments of 10 peptide bond forming C-domains revealed that there are insertion or deletion of amino acid sequences between L-peptidyl donors and D-peptidyl donors (Fig. 2). These differences may reflect on the different substrate specificity for the chirality of donor molecules. Unlike to the NRPS from

Gram-positive bacteria, Arf, Syr, and Syp in *Pseudomonas* do not contain E-domain genes in the synthetase gene clusters [7,11,14]. Then, an external epimerase is supposed to be involved in the formation of holo-NRPS. The differences in the sequence of L- and D-peptidyl donors may attribute to the binding or non-binding with this unknown external epimerase. The amino acid sequence of Arf has updated the previous information about C-domains of Gram-negative bacteria, especially putative ${}^{\text{D}}\text{C}_\text{D}$ domain. We further analyzed the phylogeny of C-domains of various Gram-positive and Gram-negative bacteria.

3.2. C-domains of NRPS from Gram-positive and Gram-negative bacteria

The tree with totally 162 C-domains does indeed show a clustering according to the types of reactions catalyzed and types of bacterial group (Fig. 3). C-domains are apparently grouped into two main functional classes; L-peptidyl donors, ${}^{\text{L}}\text{C}_\text{L}$ and putative ${}^{\text{L}}\text{C}_\text{D}$, and D-peptidyl donors, ${}^{\text{D}}\text{C}_\text{L}$ and putative ${}^{\text{D}}\text{C}_\text{D}$. It was confirmed that L-acceptors (${}^{\text{L}}\text{D}\text{C}_\text{L}$) and D-acceptors (${}^{\text{L}}\text{D}\text{C}_\text{D}$) are not clearly separated. L-peptidyl donors and D-peptidyl donors are subclustered into two groups according to the bacterial group, Gram-positives and Gram-negatives. These results suggest that L-peptidyl donors

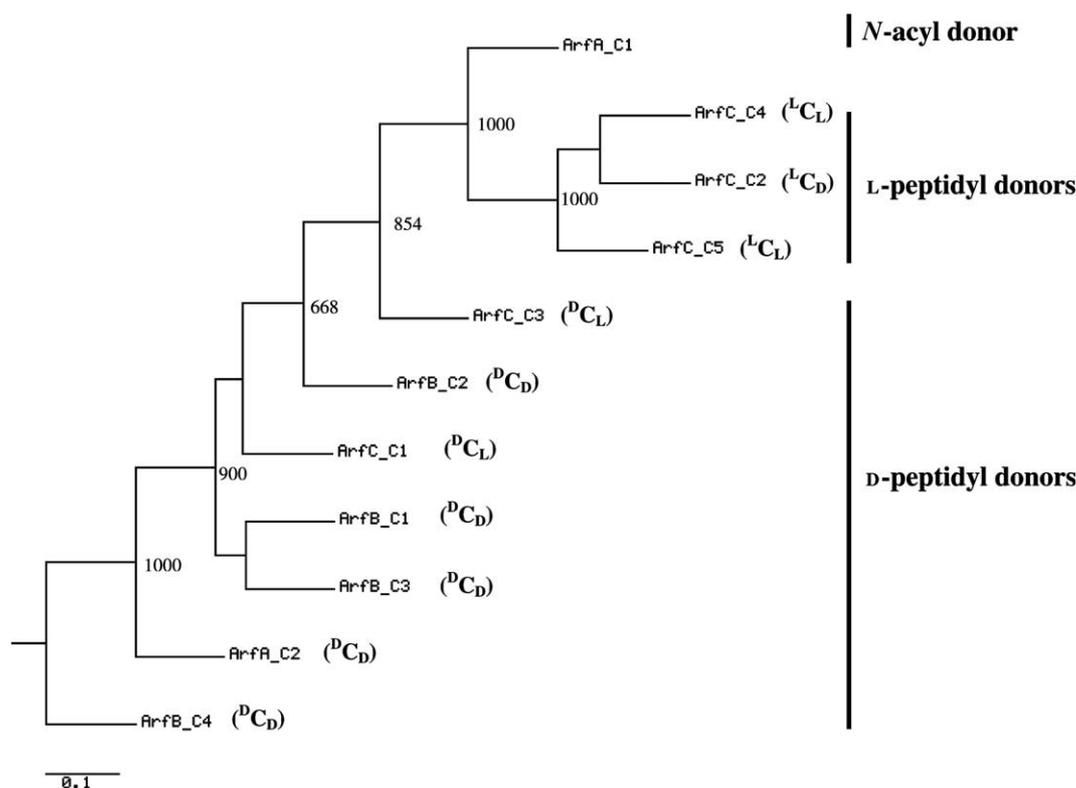


Fig. 1. A phylogenetic tree analysis of C-domains in Arf. The scale bar represents 10 substitutions per 100 amino acids. Bootstrap values higher than 500 are indicated.

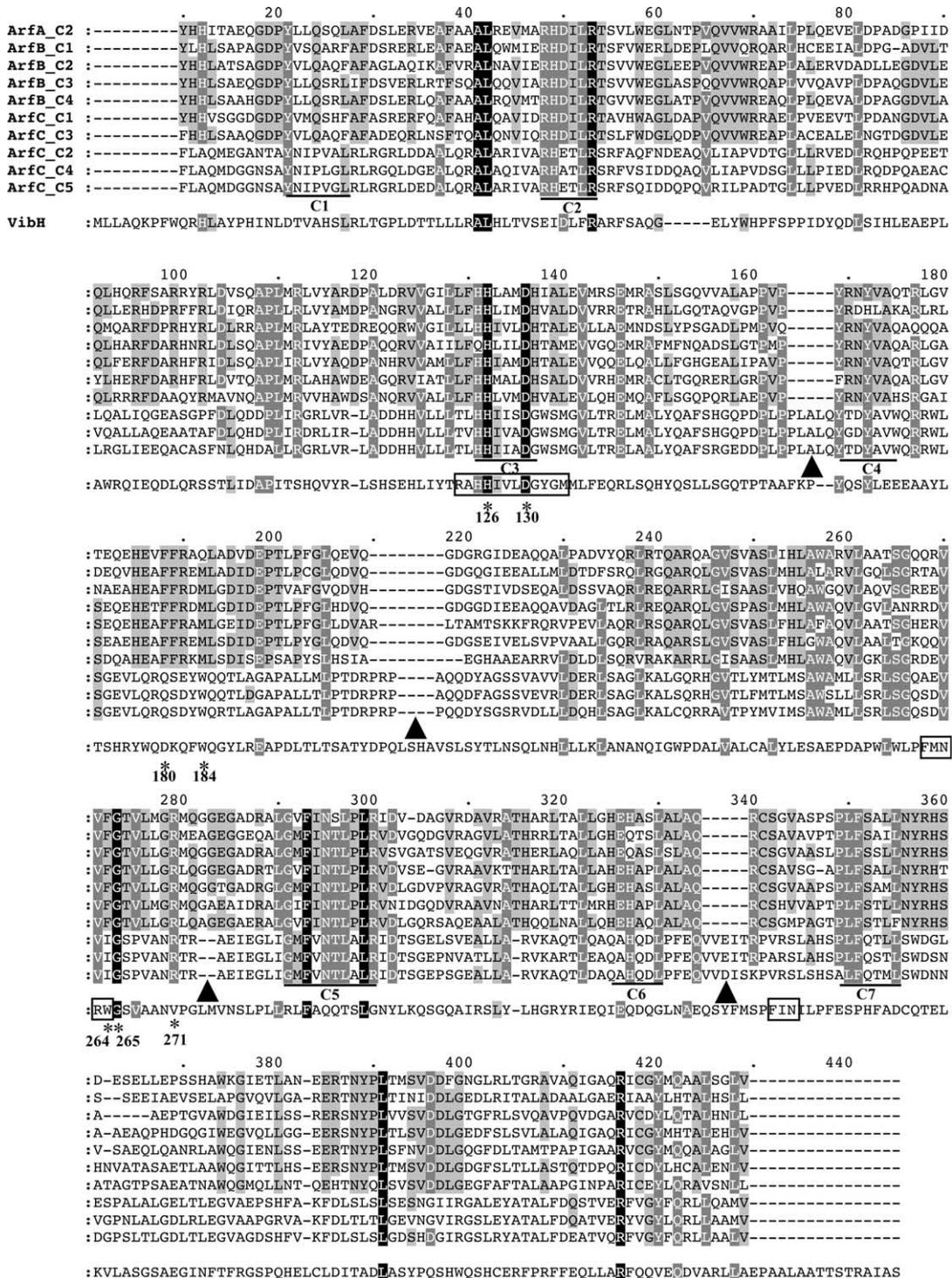


Fig. 2. Amino acid sequence alignment of 10 peptide bond forming C-domains in Arf and VibH. Arfc_C2, Arfc_C4 and Arfc_C5 are L-peptidyl donors whereas other C-domains belong to D-peptidyl donors except for VibH (stand alone C-domain). Conserved core motifs are underlined. Active site residues are boxed in VibH. Type specific peptide insertion and deletion is indicated by arrow heads. The corresponding amino acid residues (His126, Asp130, Asp180, Trp184, Trp264, Gly265, and Val271) for VibH are indicated by asterisks.

and D-peptidyl donors were separated before Gram-positive and Gram-negative bacteria were separated in their evolution. It should be noted that D-peptidyl donors of siderophore synthetase (Pvd, Pvs) in *Pseudomo-*

nas are obviously closer lineage to those of Gram-positive filamentous actinomycetes (Cep, Com) than those of lipopeptides in *Pseudomonas* (Arf, Syr, and Syp). Both Pvd and Pvs contain internal E-domains

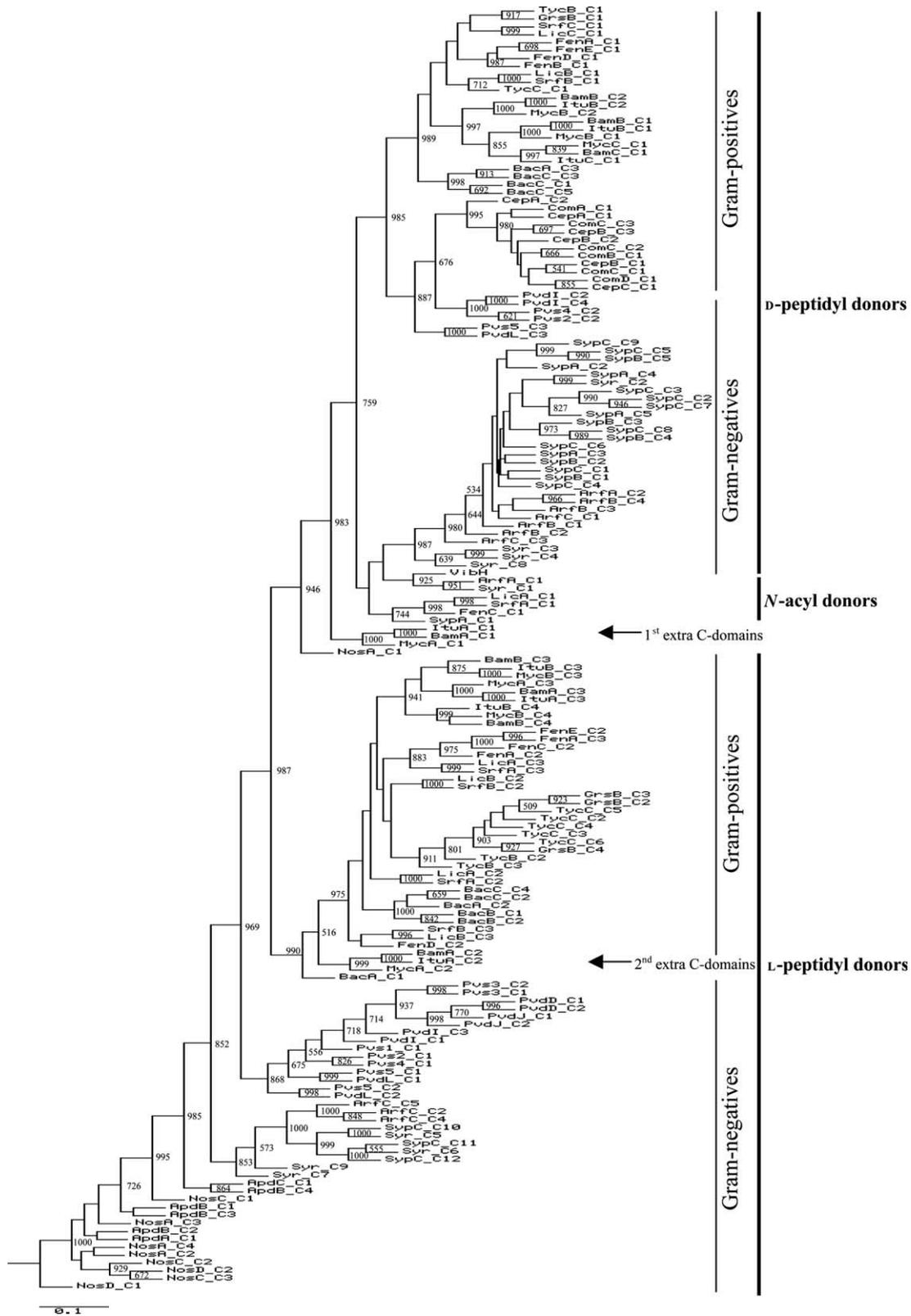


Fig. 3. A phylogenetic tree analysis of 162 C-domains of NRPS from Gram-positive and Gram-negative bacteria. In Gram-positive bacteria, six NRPS for cyclic lipopeptides (Bam, Fen, Itu, Lic, Myc, Srf), and three NRPS for cyclic peptides (Bac, Grs, Tyc) from *Bacillus* strains were analyzed. Two NRPS from filamentous actinomycetes (Cep, Com) were also chosen in this study. In Gram-negative bacteria, a free standing C-domain (VibH) from *Vibrio cholerae*, three NRPS for cyclic lipopeptides (Arf, Syr, Syp) and two NRPS for siderophores (Pvd, Pvs) from *Pseudomonas* spp., and two NRPS for cyclic peptides (Apd, Nos) from cyanobacteria (*Anabaena* sp. and *Nostoc* sp.) were analyzed. The scale bar represents 10 substitutions per 100 amino acids. Bootstrap values higher than 500 are indicated.

similar to Gram-positive NRPS but Arf, Syr, and Syp do not contain E-domains gene in the synthetase gene cluster [7,11,14]. This may be the reason for the fact that Pvd/Pvs and Com/Cep share a similar structure of D-peptidyl donor over the difference in Gram-positive and Gram-negative bacteria.

N-acyl donors use fatty acyl-CoA as their starter unit. The *N*-acyl donor is unexceptionally located at the N-terminal domain of the first subunit NRPS that synthesize cyclic lipopeptides (CYLP) in both Gram-positive and Gram-negative bacteria. It has been suggested that the first amino acid of lipopeptides is initially *N*-acylated with a β -hydroxy fatty acid in this domain [12]. CYLP containing a β -hydroxy fatty acid have various physiological activities, such as enzyme inhibitors (fengycin, plipastatin [13]), biosurfactants (arthrofactin [7], lichenysin [12], surfactin [10]), and phytotoxin (syringomycin [11], syringopeptin22 [14]). Whereas, CYLP with a β -amino fatty acid modification are antifungal substances such as bacillomycin, iturin and mycosubtilin. The first subunit of iturin family synthetase (BamA, ItuA, MycA) exhibits a remarkable complexity with two extra N-terminal C-domains [15–17]. The first extra C-domains (BamA/ItuA/MycA_C1) that have been believed to transfer β -amino fatty acid to a T-domain [15] are close relatives to *N*-acyl donors for β -hydroxy fatty acid (Fig. 3). Meanwhile, the second extra C-domains (BamA/ItuA/MycA_C2) that have been proposed to acylate the activated asparagine with β -amino fatty acid are closer to L-peptidyl donors than *N*-acyl donors for β -hydroxy fatty acid. This finding may suggest the different optical configuration of β -hydroxy fatty acid and β -amino fatty acid. Chirality of asymmetric alpha-carbon in the β -hydroxy fatty acid part in surfactin has been reported to be *R*-form [18]. It is interesting that the SrfA_C1 is closer to D-peptidyl donors than L-peptidyl donors because optical configuration of asymmetric alpha-carbon of D-form amino acid is known to be *R*-form. Then, it is possible that clustering of *N*-acyl donor is also according to the chirality of the molecule. It is of great interest whether β -amino fatty acid in nonribosomal peptides has *R*-form or *S*-form configuration.

In order to analyze the relationship between the structures of C-domains and donor/acceptor molecules in more details, phylogenetic trees were constructed in each Gram-positive and Gram-negative bacterial NRPS. In Gram-positive *Bacillus*, the 1C_L and 2C_L domains are relatively high abundance (Fig. 4(a)). Only 3 putative 2C_D domains were identified in iturin synthetase family (BamB/ItuB/MycB_C1). The 2C_L and putative 2C_D domains were also contained in Cep and Com synthetases of actinomycetes. CepB_C1 (1C_D), CepC_C1 (1C_L) and ComD_C1 (1C_D) were clustered not into L-peptidyl donors but into D-peptidyl donors. This may attribute to the fact that they adopt unusual amino acids as acceptor, HPG and DHPG. Their side chain structures

would affect the structure of C-domains over the enantioselectivity for donor molecules.

Phylogenetic analysis of C-domains from Gram-negative bacteria also showed three clusters according to the types of reactions catalyzed, L-peptidyl donors, D-peptidyl donors, and *N*-acyl donors (Fig. 4(b)). However, C-domains of Syp also showed unusual clustering. SypC_C12 is 2C_L -domain for D-Dab and L-Thr condensation but it belongs to L-peptidyl donors group. In contrast, SypA_C5, and SypC_C3 should be 1C_D but they are members of D-peptidyl donors. In addition, the C-domains for Dhb, which is an achiral molecule, such as SypA_C2 and SypB_C5 are members of D-peptidyl donors, too.

3.3. Evolutionary trace analysis of C-domains

VibH, that is responsible for acylation of amine group in norspermidine with aryl acid dihydroxybenzoate, was classified as a member of *N*-acyl donors (Fig. 3). Since the crystal structure and active site structure of VibH has already been determined [19], evolutionary trace (ET) analysis [20] was applied to extract functionally important residues conserved at different levels of partition among the C-domain groups (figure not shown). The aligned sequences with VibH (1L5A) coordinate were submitted to the University of Cambridge ET server (<http://www-cryst.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html> [21]). ET analysis of 162 C-domains indicated that the major functional clusters (*N*-acyl, L-peptidyl, and D-peptidyl donors) are visible as partition P03. These results showed that the catalytic residue His126 and structural residue Asp130 are conserved in almost all partitions, suggesting the common catalytic mechanism is shared. The Asp130 was replaced with structurally related Glu in PvdL_C2 and Pvs5_C2. Gly265 and Arg271, are also highly conserved. Gly265 lines near the solvent channel of VibH and may be important for catalytic reaction. Substitution of Arg271 to Gln was observed in the first extra C-domains of Gram-positive bacteria. There was a significant difference between L-peptidyl and D-peptidyl donors at position 264. This residue is located in the active site of VibH, just before the conserved Gly265. In the L-peptidyl donors group, the residue 264 was either Val or Ile. The corresponding residue in the D-peptidyl donors group was more bulky hydrophobic Phe in most cases (see also Fig. 2). Tyr264 was also found in GrsB_C1 domain. ET analysis revealed another signature difference at positions 180 (Asp) and 184 (Trp) in alpha-6 of VibH. Trp is highly conserved at 180 and the residue at 184 is Leu with no exception in the D-peptidyl donors. While, the corresponding residues in L-peptidyl donors are in most cases Gln and Trp, respectively. It is interesting that these positions are replaced by Phe or Tyr and Met or Arg, respectively, in the D-peptidyl donors from

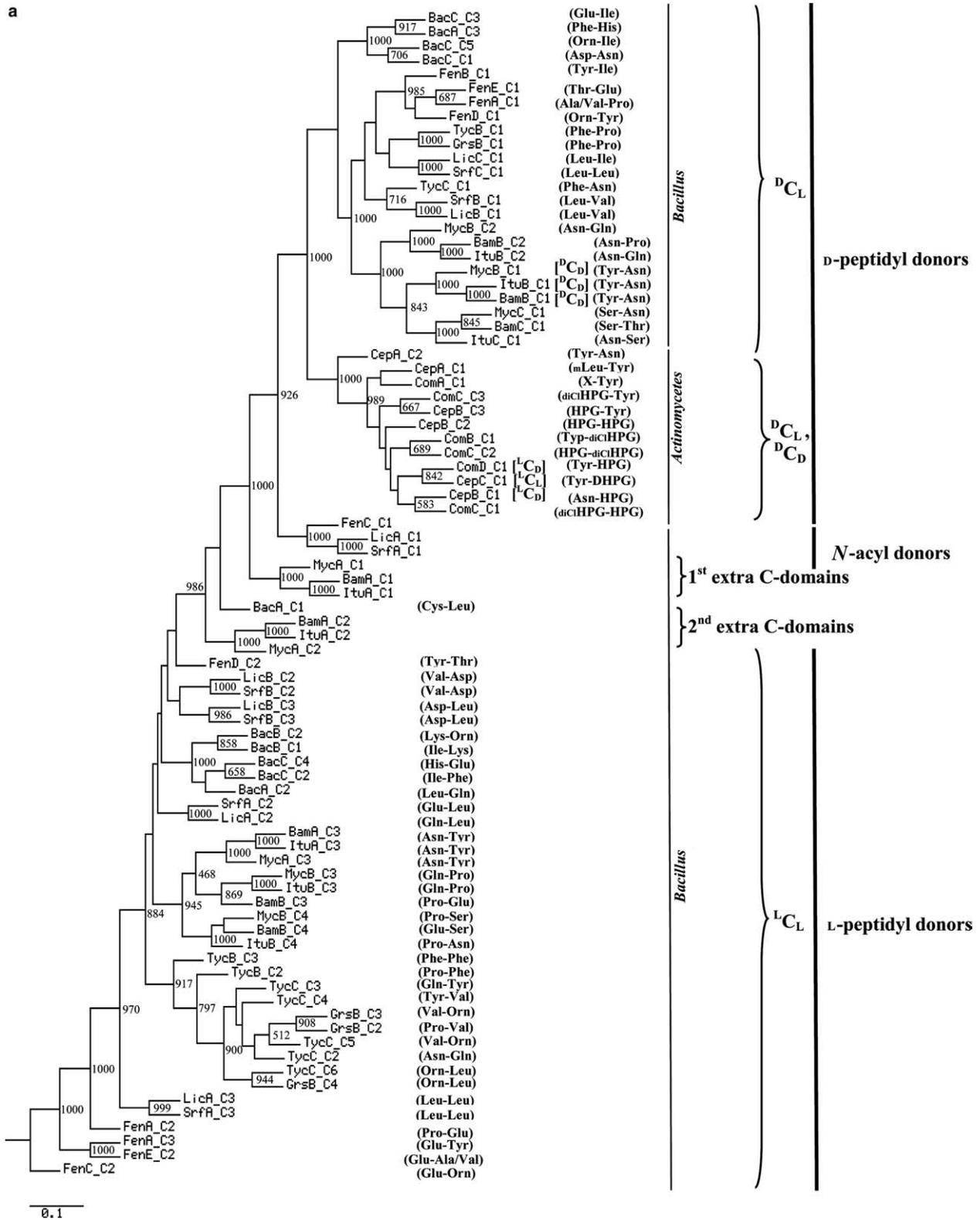


Fig. 4. A phylogenetic tree analysis of C-domains in NRPS from Gram-positive (a) and Gram-negative (b) bacteria. Amino acid donor and acceptor molecules are shown in parentheses. Proteineous amino acids are shown in three letter notation. Unusual amino acids are abbreviated as follows. Ahp, 3-amino-6-hydroxy-2-piperidone; oHAsp, 3-hydroxyaspartic acid; Dab, 2,4-diaminobutyric acid; Dhb, dehydro-2-aminobutyric acid; diClHPG, 3,5, dichloro-hydroxyphenylglycine; HPG, 4-hydroxyphenylglycine; DHPG, 3,5 dihydroxyphenylglycine; ciThr, 4-chloro-threonine; Leu_{Ac}, leucylacetate; mLeu, methylleucine; hfOrn, *N*⁵-formyl-*N*⁵-hydroxyornithine; hTyr, homotyrosine; mePro, 4-methylproline; X, 3,5-dichloro-4-hydroxybenzoylformate. Y indicates the acyl group of pyoverdine which is either a carboxylic acid or an amide. The scale bar represents 10 substitutions per 100 amino acids. Bootstrap values higher than 500 are indicated.

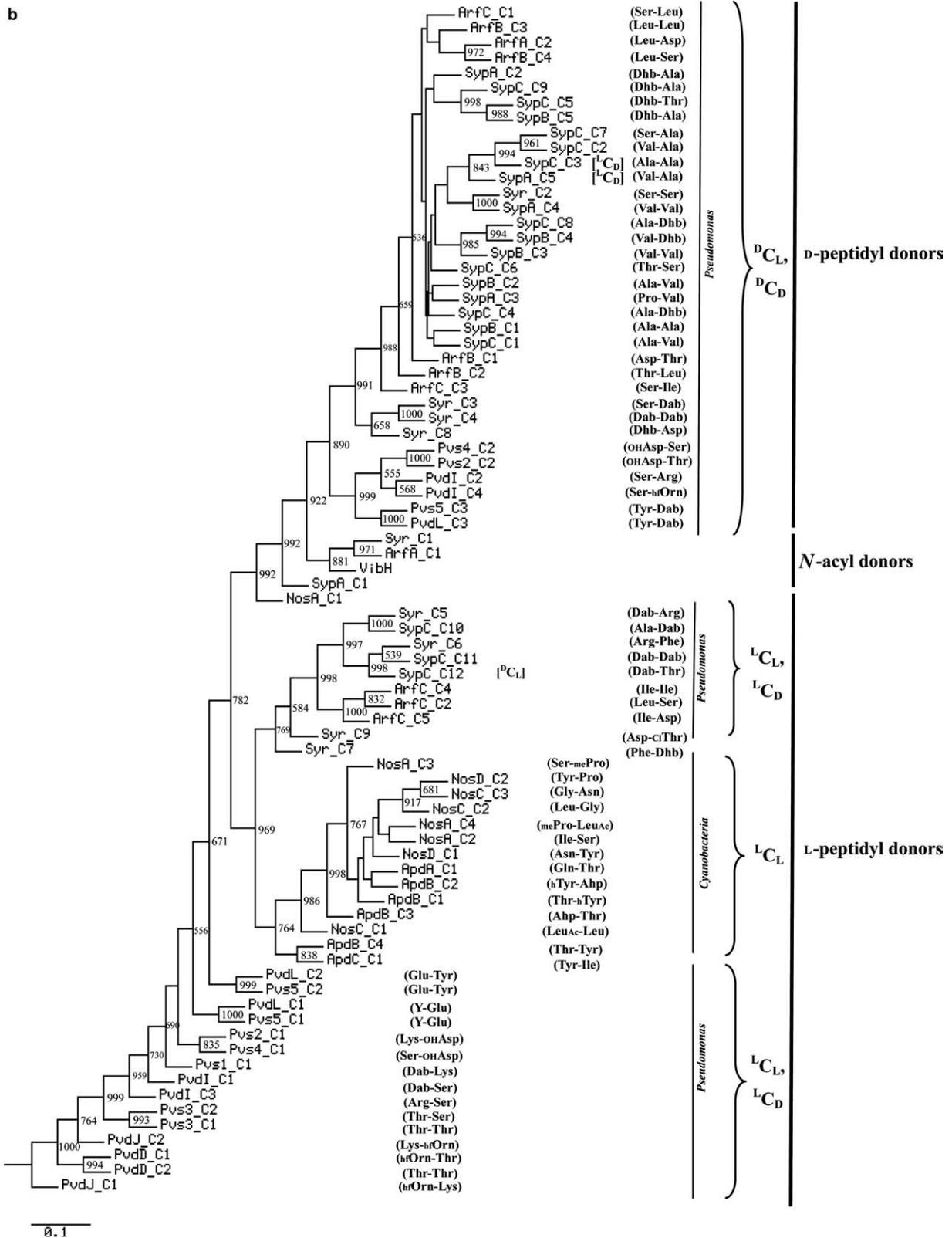


Fig. 4 (continued)

Gram-negative bacteria that do not contain internal E-domain (Arf, Syr, and Syp). Because these corresponding residues in VibH, Asp180 and Trp184, are

lined near the solvent channel, these differences probably relate to the different substrate specificity of C-domains against donor molecules.

4. Conclusion

Phylogenetic analyses of C-domains presented here revealed the structural diversity and relationship with donor and acceptor molecules. This study demonstrates that optical configuration of the donor molecule is critical for the structure of C-domain and also there is no D-acceptor specific C-domain as far as examined. Several type-specific amino acid residues were identified within putative solvent channel of C-domains that may be directly or indirectly responsible for determining substrate specificity. This phylogenetic approach should be useful not only for engineering novel NRPS but also for deducing the chirality of amino acids in product peptides.

Acknowledgements

N.R. acknowledges his post-doctoral fellowship from Japan Society for the Promotion of Science (P04468). This work was supported by the Grants-in-Aid for Scientific Research for Exploratory Research of the MEXT (No. 17510171) and Takeda Science Foundation.

References

- [1] Schwarzer, D., Finking, R. and Marahiel, M.A. (2003) Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* 20, 275–287.
- [2] Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A. and Walsh, C.T. (2002) Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* 41, 9184–9196.
- [3] Marahiel, M.A., Stachelhaus, T. and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97, 2651–2674.
- [4] Stachelhaus, T., Schneider, A. and Marahiel, M.A. (1995) Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* 269, 69–72.
- [5] von Dohren, H., Dieckmann, R. and Pavela-Vrancic, M. (1999) The nonribosomal code. *Chem. Biol.* 6, R273–R279.
- [6] Clugston, S.L., Sieber, S.A., Marahiel, M.A. and Walsh, C.T. (2003) Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C₅ domain of tyrocidine synthetase is a ¹³C_L catalyst. *Biochemistry* 42, 12095–12104.
- [7] Roongsawang, N., Hase, K., Haruki, M., Imanaka, T., Morikawa, M. and Kanaya, S. (2003) Cloning and characterization of the gene cluster encoding arthrofactin synthetase from *Pseudomonas* sp. MIS38. *Chem. Biol.* 10, 869–880.
- [8] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- [9] Felsenstein, J. (1989) PHYLIP: phylogeny inference package. *Cladistics* 5, 164–166.
- [10] Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G. and van Sinderen, D. (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* 8, 821–831.
- [11] Guenzi, E., Galli, G., Grgurina, I., Gross, D.C. and Grandi, G. (1997) Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* 273, 32857–32863.
- [12] Konz, D., Doekel, S. and Marahiel, M.A. (1999) Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J. Bacteriol.* 181, 133–140.
- [13] Lin, T.P., Chen, C.L., Chang, L.K., Tschén, J.S. and Liu, S.T. (1999) Functional and transcriptional analyses of a fengycin synthetase gene, *fenC*, from *Bacillus subtilis*. *J. Bacteriol.* 181, 5060–5067.
- [14] Scholz-Schroeder, B.K., Soule, J.D. and Gross, D.C. (2003) The *sypA*, *sypB*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Mol. Plant Microb. Interact.* 16, 271–280.
- [15] Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F. and Vater, J. (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. USA* 96, 13294–13299.
- [16] Tsuge, K., Akiyama, T. and Shoda, M. (2001) Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* 183, 6265–6273.
- [17] Moyne, A.L., Cleveland, T.E. and Tuzun, S. (2004) Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. *FEMS Microbiol. Lett.* 234, 43–49.
- [18] Nagai, S., Okimura, K., Kaizawa, N., Ohki, K. and Kanatomo, S. (1996) Study on surfactin, a cyclic depsipeptide. II. Synthesis of surfactin B2 produced by *Bacillus natto* KMD 2311. *Chem. Pharm. Bull. (Tokyo)* 44, 5–10.
- [19] Keating, T.A., Marshall, C.G., Walsh, C.T. and Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* 9, 522–526.
- [20] Lichtarge, O., Bourne, H.R. and Cohen, F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257, 342–358.
- [21] Innis, C.A., Shi, J. and Blundell, T.L. (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. *Protein Eng.* 13, 839–847.