BBA - Proteins and Proteomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity☆

Kristjan Bloudoff, T. Martin Schmeing*

Department of Biochemistry, McGill University, Montréal, QC H3G 0B1, Canada

ARTICLE INFO

Keywords:
Condensation domain
Nonribosomal peptide synthetase
Natural products
Megaenzyme
C domain superfamily
Antibiotics

ABSTRACT

Nonribosomal peptide synthetases (NRPSs) are incredible macromolecular machines that produce a wide range of biologically- and therapeutically-relevant molecules. During synthesis, peptide elongation is performed by the condensation (C) domain, as it catalyzes amide bond formation between the nascent peptide and the amino acid it adds to the chain. Since their discovery more than two decades ago, C domains have been subject to extensive biochemical, bioinformatic, mutagenic, and structural analyses. They are composed of two lobes, each with homology to chloramphenicol acetyltransferase, have two binding sites for their two peptidyl carrier protein-bound ligands, and have an active site with conserved motif HHxxxDG located between the two lobes. This review discusses some of the important insights into the structure, catalytic mechanism, specificity, and gatekeeping functions of C domains revealed since their discovery. In addition, C domains are the archetypal members of the C domain superfamily, which includes several other members that also function as NRPS domains. The other family members can replace the C domain in NRP synthesis, can work in concert with a C domain, or can fulfill diverse and novel functions. These domains include the epimerization (E) domain, the heterocyclization (Cy) domain, the ester-bond forming C domain, the fungal NRPS terminal C domain (C_T), the β-lactam ring forming C domain, and the X domain. We also discuss structural and function insight into C, E, Cy, C_T and X domains, to present a holistic overview of historical and current knowledge of the C domain superfamily. This article is part of a Special Issue entitled: Biophysics in Canada, edited by Lewis Kay, John Baenziger, Albert Berghuis and Peter Tieleman.

1. Introduction: nonribosomal peptide synthetases and their domains

1.1. An overview of nonribosomal peptide synthetases

Nonribosomal peptide synthetases (NRPSs) are an interesting family of enzymes which assemble acyl substrates into bioactive secondary metabolites [1–3]. Their nonribosomal peptide products have important and diverse activities. They include antifungals (bacillomycin), antibacterials (daptomycin), antivirals (luzopeptin), antitumors (actinomycin D), siderophores (enterobactin), and immunosuppressants (cyclosporin) [4]. These compounds occupy a huge area of chemical space because NRPSs can use over 500 different acyl monomer substrates (including proteogenic and nonproteogenic amino acids, fatty acids and hydroxy acids), can be co-synthetically or post-synthetically modified, and have linear, cyclic, or branched topologies.

NRPSs themselves are large and elegant macromolecular machines which use a modular, assembly line synthetic logic [5]. NRPSs are organized into modules of ~1100 residues with each module responsible for the addition of one specific amino acid to the peptide product [6]. The number and order of modules typically correspond directly to the number and order of amino acids in the peptide product, though there are many known cases of iterativity, module skipping and intermediate oligomerization. NRPSs can be as small as one module or as large as eighteen modules, with molecular weights > 2 MDa in a single protein chain (Fig. 1). Typically, an elongation module consists of the three domains essential to the elongation of a peptide: the adenylation (A) domain, the peptidyl carrier protein (PCP) domain (alternatively referred to as the thiolation (T) domain) and the condensation (C) domain. Terminal modules also include domains dedicated to releasing the final peptide, such as a thioesterase (Te) or reductase (R) domain. A typical organization of an NRPS is A-PCP-$(C\text{-}A\text{-}PCP)_n$-Te, with additional tailoring domains inserted into some of the modules.

In each round of the peptide synthetic cycle (Fig. 2), the A domain binds specifically to its cognate substrate and first activates it by

☆ This article is part of a Special Issue entitled: Biophysics in Canada, edited by Lewis Kay, John Baenziger, Albert Berghuis and Peter Tieleman.

* Corresponding author.
E-mail address: martin.schmeing@mcgill.ca (T.M. Schmeing).

http://dx.doi.org/10.1016/j.bbapap.2017.05.010
Received 27 March 2017; Received in revised form 5 May 2017; Accepted 12 May 2017

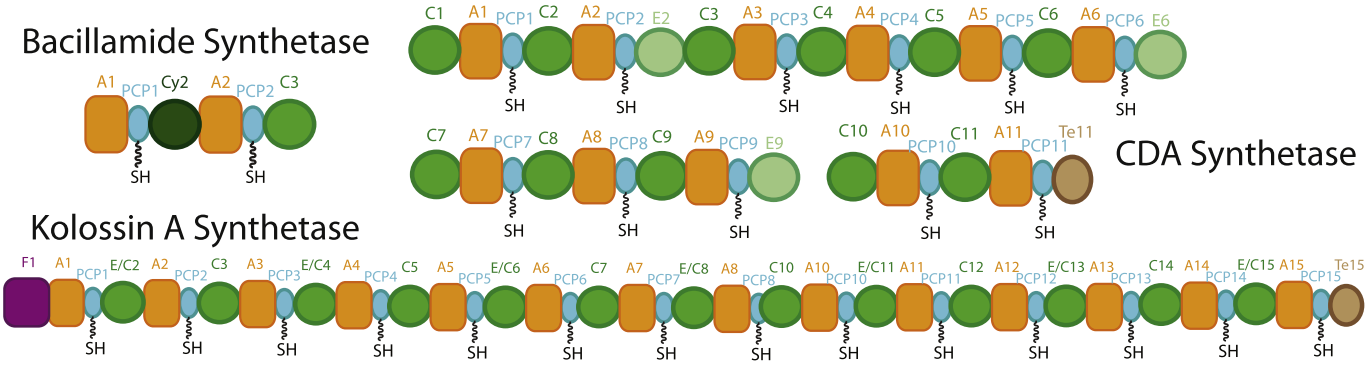

Fig. 1. Schematic representation of some NRPSs. Bacillamide synthetase is a 6-domain, 3-module, single protein NRPS; calcium-dependent antibiotic (CDA) synthetase is a 37-domain, 11-module, three protein NRPS; kolossin A synthetase is a 43-domain, 15-module, single protein NRPS.

adenylation, then transfers it onto the 4′-phosphopantetheinyl (PPE) arm of the PCP domain. The PCP domains of the current and previous modules transport their covalently bound substrates into the active site of the C domain. The C domain catalyzes the formation of a peptide bond between the two substrates, which elongates the growing peptide chain as it transfers to the acceptor PCP domain. After this condensation reaction, the peptide (now attached to the PPE arm of the PCP domain of the current module) is carried to the C domain of the next module and donated in that module's condensation reaction. This both frees the PCP domain to participate in the round of catalysis within the current module and further elongates the peptide chain in the downstream module. The peptide is likewise elongated in each of the subsequent modules, which act like stations in an assembly line, until it reaches the termination module, where after a final elongation, the Te or R domain releases the peptide product.

Significant effort has been put into elucidating how these NRPS domains function individually, as well as how they function within the context of the larger NRPS. Several insightful reviews of the field have recently been published recently, including those by Miller et al. [7], Payne et al. [8] and Weissman [2]. In addition to the desire to fundamentally understand the complicated and elegant way NRPSs function, much interest in NRPSs has been generated because, despite complex molecular mechanisms, the overall synthetic logic is quite simple, which raises the tantalizing possibility of bioengineering: Adding, removing and substituting substrate-binding residues, individual domains and entire modules should all have predictable effects on the peptide products, and allow production of myriads of novel bioactive compounds. There have been successes in bioengineering, including excellent work with the daptomycin pathway [9]. However, bioengineering of NRPSs often fails, and when successful is typically accompanied by a marked decrease in peptide yield [10–12], underlining the fact that our understanding of NRPSs is not yet complete.

1.2. Other common NRPS domains

This review focuses on the condensation domain and related condensation domain superfamily members that have acquired specialized functions. However, as C domains work in concert with the other NRPS domains, a brief discussion of the structure and function of those domains is presented here to provide context:

1.2.1. The adenylation domain

As described above, the A domain is responsible for selecting the cognate substrate and ligating it to the PCP domain to make it into an intermediate competent to participate in reactions catalyzed by other NRPS domains. Thus, the A domain is the main selectivity determinant/

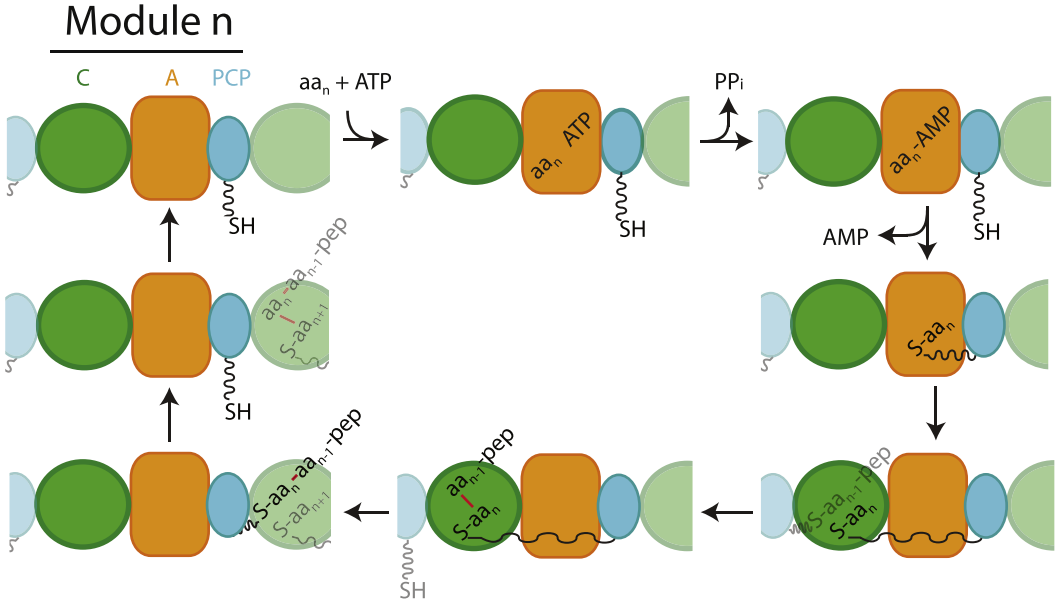

Fig. 2. Schematic representation of the NRPS elongation cycle. The elongation cycle starts with the cognate amino acid and ATP binding to the A domain (orange). The amino acid is then activated by adenylation. The A domain next catalyzes the transfer of the activated amino acid onto the PPE arm covalently linked to the PCP domain (blue). The PCP-linked amino acid of the current module (n) and the PCP-linked amino acid or peptide from the previous module (n-1) are delivered to the current module's C domain (green), where peptide bond formation occurs. The PCP-linked elongated peptide is then delivered to C domain of the next module (n + 1), where it is passed off in that condensation reaction.