

# Acyl carrier protein structural classification and normal mode analysis

David C. Cantu, Michael J. Forrester, Katherine Charov, and Peter J. Reilly\*

Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011

Received 2 December 2011; Accepted 20 February 2012

DOI: 10.1002/pro.2050

Published online 28 February 2012 proteinscience.org

**Abstract:** All acyl carrier protein primary and tertiary structures were gathered into the ThYme database. They are classified into 16 families by amino acid sequence similarity, with members of the different families having sequences with statistically highly significant differences. These classifications are supported by tertiary structure superposition analysis. Tertiary structures from a number of families are very similar, suggesting that these families may come from a single distant ancestor. Normal vibrational mode analysis was conducted on experimentally determined freestanding structures, showing greater fluctuations at chain termini and loops than in most helices. Their modes overlap more so within families than between different families. The tertiary structures of three acyl carrier protein families that lacked any known structures were predicted as well.

**Keywords:** acyl carrier protein; normal mode analysis; primary structure; protein family; tertiary structure; ThYme

## Introduction

Acyl carrier proteins (ACPs) usually have 70 to 100, but occasionally more, amino acid residues, and they are usually linked through an interior serine residue to the terminal phosphate group of a 4'-phosphopan-

tetheine prosthetic group. In turn, the latter binds fatty acids, polyketides, and other moieties by a thioester bond to its terminal thiol group, activating them for reactions that usually produce longer acyl chains, but also many other compounds (Table I).

ACP molecules have many more anionic than cationic residues and rather few hydrophobic residues.<sup>1</sup> Their tertiary structures feature three generally parallel  $\alpha$ -helices (helices I, II, and IV) with a shorter crosswise  $\alpha$ -helix (helix III). Helix II has many conserved anionic residues and plays an important role in ACP-enzyme interactions.<sup>2</sup> Helix III is not present in all ACP structures, and it displays helix-loop equilibrium conformations.<sup>3</sup>

ACPs are either independent, freestanding structures, or they are covalently bound as part of multimodular enzymes such as fatty acid synthases (FASs), polyketide synthases (PKSs), and non-ribosomal peptide synthases (NRPSs). Acyl chains attached to freestanding ACPs are held within the hydrophobic pocket formed by the  $\alpha$ -helices until they are subjected to reaction,<sup>4-7</sup> when they are expelled into the active site of the enzyme catalyzing the reaction.<sup>8</sup> The mechanism of delivering an acyl substrate to an enzyme active site from ACP's cavity, which may be accompanied by flexing of the ACP, is not completely

---

*Abbreviations:* ACP, acyl carrier protein; AMBER, Assisted Model Building with Energy Refinement; ANM, Anisotropic Network Model server; BLAST, Basic Local Alignment Search Tool; CoA, coenzyme A; FAS, fatty acid synthase; I-TASSER, Iterative Threading ASSEMBly Refinement server; NRPS, nonribosomal peptide synthase; *P*, percentage of  $\alpha$ -carbon atoms of the amino acid residues between two structures compared for calculations; PDB, Protein Data Bank; PKS, polyketide synthase; RMSD, root mean square deviation; ThYme, Thioester active enzyme database; UniProt, Universal Protein resource.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: U.S. National Science Foundation; Grant number: EEC-0813570; Grant sponsors: National Science Foundation Engineering Research Center for Biorenewable Chemicals, headquartered at Iowa State University and including Rice University, the University of California, Irvine, the University of New Mexico, the University of Virginia, the University of Wisconsin-Madison.

\*Correspondence to: Peter J. Reilly, Department of Chemical and Biological Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011-2230. E-mail: reilly@iastate.edu

**Table I.** *Acyl Carrier Protein Families*

Family	Approximate number of sequences <sup>a</sup>	Producing domain	Role	Representative end product (nonexhaustive list)	UniProt accession code	Reference
ACP1	2870	<b>B, E, A<sup>b</sup></b>	Freestanding ACPs: in type II fatty acid synthesis in bacterial long-chain fatty acid synthesis in mitochondria	Fatty acids Long-chain fatty acids Fatty acids	P0A6A8 P0A2W5 P52505	19 20 21
ACP2	180	<b>E</b>	ACP domains of fungal FASs	Fatty acids	P19097	22
ACP3	190	<b>E</b>	ACP domains of animal FASs	Fatty acids	P49327	23
ACP4	50	<b>B</b>	Freestanding ACPs in polyketide synthesis	Bacillaene	Q7PC63	24
ACP5	100	<b>B</b>	Freestanding ACPs in polyketide synthesis	Actinorhodin Frenolicin	Q02054 Q54996	25 26
ACP6	100	<b>B</b>	ACP domains in PKSs	Oxytetracycline Mycoserolic acid	P43677 Q02251	27 28
ACP7	30	<b>E</b>	ACP domains in PKSs	Phthioceranic acids	O07798	29
ACP8	240	<b>B, E</b>	ACP domains in PKSs Conidial yellow pigment	6-Methylsalicylic acid Aflatoxin	P22367 Q12053	30 31
ACP9	80	<b>E</b>	ACP domain in PKSs/chalcone synthases	Acylpyrones	Q03149	32
ACP10	7000	<b>B, E, A</b>	ACP/PCP domain in PKSs/NRPSs and ACP domain in ferrichrome synthases	Surfactin Erythronolide Gramicidin Phthiocerol Tyrocidine	P27206 Q03131 P0C061 Q7TXL8 O30409	34 35 36 37 38
ACP11	120	<b>E</b>	ACP domains in PKSs	Plipastatin Enniatin	O31827 Q00869	39 40
ACP12	280	<b>B</b>	ACP domains in enterobactin synthases	Ferrichrome Bacillaene Lovastatin	Q9P7T1 Q05470 Q9Y8A5	41 42 43
ACP13	220	<b>B, E</b>	ACP domains in isochorismatases	Enterobactin	P0ADI4	44
ACP15	200	<b>B</b>	ACP domains in enterobactin synthases Freestanding ACPs active with malonate decarboxylases	Mycobactins Enterobactin Acetate	P71717 P11454 O06925	45 46 47
ACP16	300	<b>B</b>	Freestanding ACPs active with citrate lyases	Acetyl-CoA	P02903	48
ACP17	250	<b>B</b>	Freestanding D-alanyl carrier proteins	D-Alanyl lipoteichoic acid	P55153	49

<sup>a</sup> Present in the ThYme database at the time of writing this article.

<sup>b</sup> A, archaea; B, bacteria; E, eukaryota. Most prevalent producers bolded.

```

ACP16 -----MEMKIDALAGTLESSDVMVRIGPAAQFGIQLEI--DSIVKQFGAAIEQVVRETLAQLGVKQANVVDDKGALECVLRARVQAAALRAAQQTQLQWSQL
ACP15 MEQITLSFPASRALSGRALAGVVGSGDMEV--LYTAAQSATLNQVITTSVDNSQARWQALFDRNLNLINGLPAGQLIHDGATPGVARIRIEQVFEEAHA-----
ACP1  -----MDNDEIFSKVRSIIESEQLDKKEDE---ITDSRFVEDLNADSLDIYELLYLLEAFDDKIPEN-----EANEFETVGDVVMFVKRRKG-----
ACP4  -----MDKQRIFEVLITNICEVLPEDGHR--FEPEDQLVELGA--DSVDRAEITMVLDELKIPRI-----ELSGVKNIGELAEVLYDKVQSA-----
ACP8  -----RALNILASEVGLSESD---MSDDLVPADYGV--DSLLSLVTGKYREELNLDMDSS-----VPIEHPTVGDVFKRFVTQL-----
ACP2  -----KALEVHLTLVAQKLLKSIEE---VSPQKSIKDLVGGKSTLQNEIIGDLQKEFG--ATPEK-----
ACP12 -----AALREVILPILLESDEP---FD--DDNLIDYGL--DSVRRMALAARWRKVHG--DIDFV-----MLAKNPTIDAW-----
ACP3  -----RDLVEAVAHILGVRDVS--LNAESLADLGL--DSLGMVEVQTLERDYDVMTR-----EIRLLTINKLRELS-----
ACP5  -----MTLLTSLDLLTLRECAEESIDLGGD--VE--DVAFDALGY--DSLALLNTVGRERDYGVLGGD-----AVEKATPRALIEMTNASLTGASPSAGGAARDK-----
ACP13 -----IIAAAFSSLLGCDV---QDADADFPALGG--HSLAMKLAQLSRQVARQVTPG-----QVMVASTVAKLAT-----
ACP9  -----LQDKITSKVSDDLSPISK---INFDPHLKHYGL--DSLLTVQFKSWIDKEFEKPLFTH-----
ACP17 -----MADEAIKNGVLDILADLTGSDDVK---KNLDLNLFTGLLDSMGTVQLLELQSQFQVDPAPVSEFDRKEWDTPNKI IAKVEQAQ-----
ACP11 -----QVTLQIPDGES--VHPTIPLIDQGV--DSLGAVTGTFWFSKQLYLDLPLL-----KVLGGASITDLANE-----
ACP7  -----LDEKIRGCVAKVLQMTAED---VDSKAALADLGV--DSVMTVTLRRQLQTLKIAVPT-----LWHSHTVSHLAVWFA-----
ACP10 -----LELVANAVAEVLGHESAAE--INVRRAFSELGL--DSLNAMALRRLSASTGLRLPAS-----LVFDHPTVTLAQHLRLARL-----
ACP6  -----PARLRQLVAEQVSLILRRT--VDPDRPLPEYGL--DSLGALELRTRITETGIRLAPK-----NVSATVRLGLADHLYEQLPADPAPAAALSSQ-----

```

**Figure 1.** Multiple sequence alignment of single members of each of the 16 ACP families, roughly arranged in order of similarity to each other. ACP1: *Borrelia burgdorferi*; ACP2: *Schizosaccharomyces pombe*; ACP3: *Gallus gallus*; ACP4: *Bacillus subtilis*; ACP5: *Streptomyces rimosus*; ACP6: *Mycobacterium tuberculosis*; ACP7: *Penicillium patulum*; ACP8: *Aspergillus nidulans*; ACP9: *Dictyostelium discoideum*; ACP10: *Saccharopolyspora erythraea*; ACP11: *Aspergillus terreus*; ACP12: *Escherichia coli*; ACP13: *E. coli*; ACP15: *Klebsiella pneumoniae*; ACP16: *K. pneumoniae*; ACP17: *Lactobacillus rhamnosus*.

understood. Those ACPs that are part of enzymes may not hold acyl chains within their hydrophobic pockets,<sup>9</sup> as these chains are less exposed to solvent and to cell membranes.<sup>10</sup>

Freestanding ACPs, at least, appear to be quite flexible, being found in multiple conformers<sup>3,5,11</sup> and having very flexible loops and  $\alpha$ -helices.<sup>12–15</sup>

We have gathered all available ACP primary and tertiary structures into the ThYme database.<sup>16</sup> There we have classified ACPs into families, following the same techniques that we have used earlier with thioesterases<sup>17</sup> and ketoacyl synthases,<sup>18</sup> which are described in Computational Methods. In general, members of a protein family have strong sequence similarity. They should also have tertiary structures that can be superimposed with small root mean square deviations (RMSDs) between corresponding amino acid residues. These similarities may imply that members are descended from a common protein ancestor. Members of different families have primary structures with statistically highly significant differences. However, slight similarities in amino acid residue alignments between ACP families may suggest that those without known tertiary structures are related to those with known structures. We describe this work with ACPs for the first time here.

This article is an account also of two further efforts: (1) a normal mode analysis of experimentally determined tertiary structures of freestanding apo-ACPs to describe their dynamic structures; and (2) the attempted computational prediction of tertiary structures of freestanding ACPs in three families with no known experimentally determined structures.

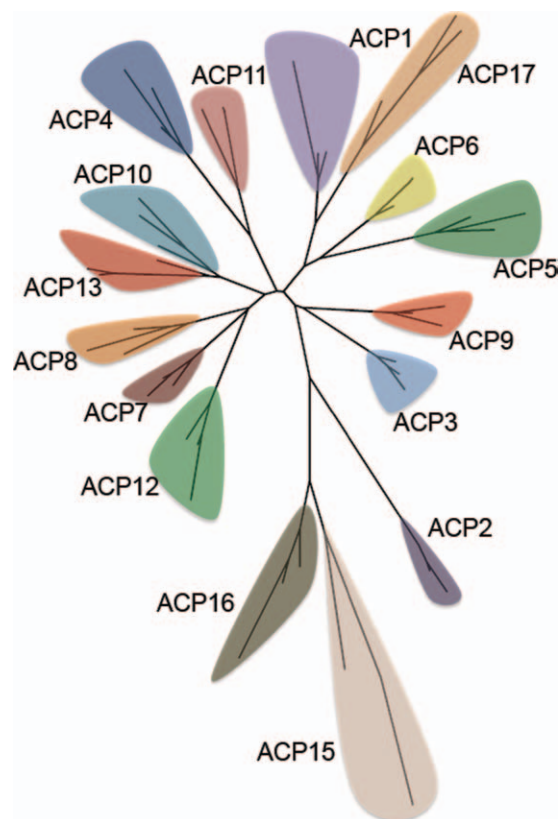
## Results and Discussion

### ACP families

Following the protocol described in the Computational Methods section led to 16 ACP families being defined. A multiple sequence alignment of representative members of these families shows very limited sequence similarity (Fig. 1). Only the serine residue

at position 39 (ACP1 numbering), to which the prosthetic group is attached, is almost completely conserved (except for ACP15). In addition, there is substantial conservation at positions 35 (glycine), 38 (aspartate), and 67 (threonine). A number of positions have almost exclusively hydrophobic residues. In ACP15, the only residue that is completely conserved near the position otherwise occupied by a serine residue is Thr40 (ACP1 numbering).

The same representative sequences of ACP families were subjected to pairwise sequence alignments to identify common residues between two sequences in different families. Even though there is low sequence similarity over the 16 families, there is



**Figure 2.** Phylogenetic tree of ACP families.

substantial similarity from one family to the next (Supporting Information Table S1 and Fig. S1). The relationship of the families to each other is shown by two versions of a phylogenetic tree (Figs. 2 and Supporting Information Fig. S2). They show ACP15 and ACP16 peripheral to the other ACP families, as they are in Figure 1.

Table I summarizes the 16 ACP families, showing (1) approximate numbers of sequences in each family; (2) domains of life that produce each family; (3) whether ACP families are composed of freestanding proteins or are parts of multidomain enzymes; (4) the end products of enzymes with which ACP families interact; (5) representative UniProt accession codes of the ACPs; and (6) related literature.

Families ACP1 through ACP3 are involved with fatty acid synthesis. ACP1 members are freestanding ACPs, present in type II fatty acid synthesis, bacterial long-chain fatty acid synthesis, and mitochondrial fatty acid synthesis. ACP2 and ACP3 proteins are parts of multidomain FASs, involved with fungal and animal type I fatty acid synthesis, respectively. A substantial amount of research has been conducted on these ACPs; much of it is covered in the exhaustive review of Chan and Vogel.<sup>10</sup>

Proteins in ACP4 and ACP5 are freestanding ACPs involved with polyketide synthesis. In ACP4 is AcpK, part of the *pksX* pathway of *Bacillus subtilis* in making bacillaene.<sup>24</sup> ACP5 includes ACPs involved with the synthesis of the antibiotics actinorhodin,<sup>25</sup> frenolicin,<sup>26</sup> and oxytetracycline<sup>27</sup> in various *Streptomyces* species.

ACP6 through ACP9 and also ACP11 include the ACP domains of large multidomain PKSs. Among ACP6 sequences are the ACP domains of larger enzymes involved in making complex lipids found in *Mycobacterium* cell walls, including mycocerosic acid,<sup>28</sup> and other sulfolipids.<sup>29</sup> ACP7 includes the ACP domain of fungal PKS 6-methylsalicylic acid synthase.<sup>30</sup> ACP8 includes mainly fungal sequences; among them are the ACP domains of PKSs involved with aflatoxin production,<sup>31</sup> and the ACP domains of naphthopyrone PKSs that make the yellow pigment in conidia.<sup>32</sup> ACP9 includes the ACP domain of PKS or chalcone synthase *stlA*, which produces acylpyrones.<sup>33</sup> ACP11 includes the ACP domains of lovastatin synthases.<sup>43</sup>

ACP10, the family with the most members, includes the ACP domains of many PKSs, the peptide carrier protein domain of NRPSs, hybrid PKS/NRPS enzymes, and ferrichrome synthases.<sup>34–42</sup> The enzymes in this family make a variety of natural products from secondary metabolism. A representative sample of them is shown in Table I.

Families ACP12 and ACP13 include the ACP domains of enterobactin synthases,<sup>44,46</sup> isochorismatases,<sup>45</sup> and mycobactin synthases.<sup>45</sup> ACP14 has been merged into ACP10 and no longer exists.

The prosthetic group in ACP15 and ACP16 is 2'-(5''-phosphoribosyl)-3'-dephospho-CoA, instead of 4'-phosphopantetheine, linked to an interior serine residue of apo-ACP through its 5''-phospho group and to the acyl molecule with a thioester bond through its terminal thiol group. ACP15 proteins include the ACPs active with malonate decarboxylases in bacteria that convert malonate to acetate and CO<sub>2</sub> as an energy source.<sup>47</sup> ACP16 enzymes include the ACPs active with citrate lyases that convert citrate to oxaloacetate and acetate in bacteria.<sup>48</sup>

Members of ACP17 do not carry acyl groups, but instead they are D-alanyl carrier proteins, as the moiety bound by 4'-phosphopantetheine is D-alanine, which is ligated using adenosine triphosphate to poly(ribitol phosphate). These enzymes are involved with the production of D-alanyl lipoteichoic acid.<sup>49</sup>

The members of all but four families, ACP1, ACP8, ACP10, and ACP13, are produced exclusively by either bacteria or eukaryota. More specifically, virtually all members of ACP2 and ACP8 and all members of ACP7 and ACP11 are produced by fungi (the latter three families all by ascomycota), while all ACP3 members come from animals. ACP9 members are from slime molds. Members of different bacterial phyla produce different members of ACP4, ACP12, ACP15, ACP16, and ACP17. ACP5 and ACP6 members are all from actinobacteria, with the latter being only from *Mycobacterium* species.

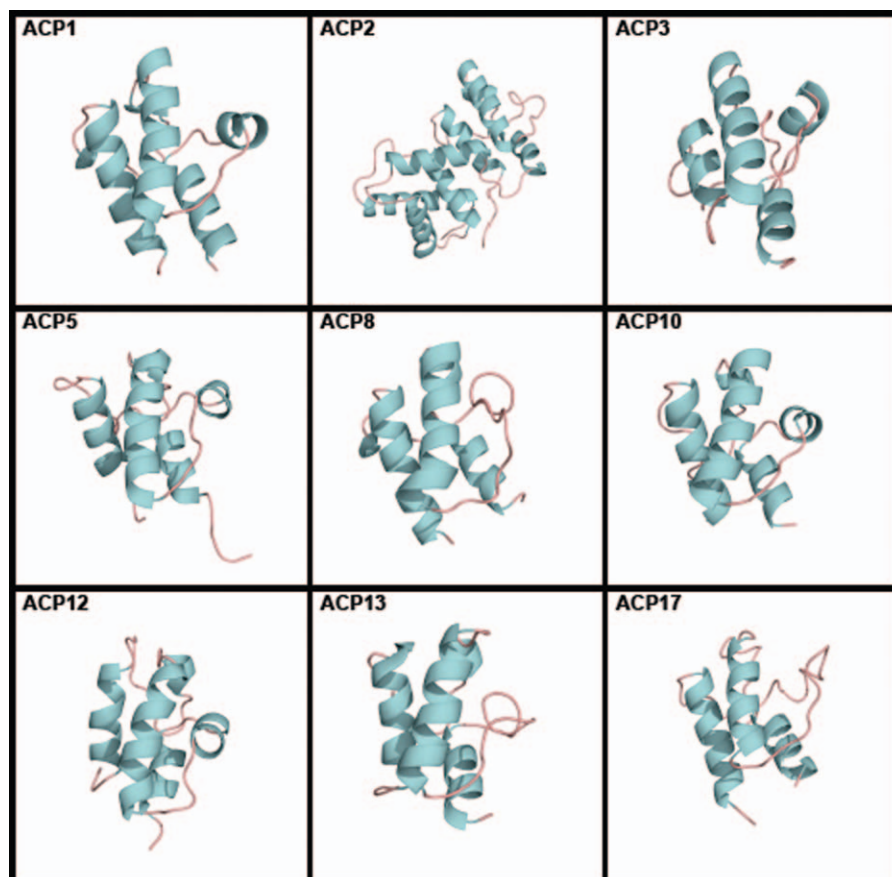
In summary, ACPs have diverged into different families based on primary structures that have statistically highly significant differences. They are either freestanding or are covalently bound to enzymes, they are specific to different substrates, they are produced by different classes of organisms, and they have sharply defined roles.

### Existing ACP tertiary structures

Families ACP1, ACP2, ACP3, ACP5, ACP8, ACP10, ACP12, ACP13, and ACP17 contain members with known tertiary structures (Fig. 3). All but ACP8, ACP12, and ACP13 have more than one known structure. Most known structures and their properties were reviewed by Chan and Vogel.<sup>10</sup> All tertiary structures are tabulated, with links to the Protein Data Bank (PDB), in the ACP section of the ThYme database.

We superimposed one tertiary structure per species within each putative ACP family and calculated their RMSDs using the protocol in the Computational Methods section. ACP2 and ACP17 have multiple known structures of freestanding ACPs or of ACP domains, but they all come from the same species. Therefore, their RMSDs were not calculated, as this would represent structure conservation among the same sequence and not sequences in a family. ACP2 domains in yeast FAS structures contain four extra helices in the C-terminal region,<sup>22</sup> not seen in other ACPs, making them about twice as long as the





**Figure 3.** Tertiary structures of single members of ACP families. ACP1: *Thermus thermophilus* (PDB accession code 1X3O); ACP2: *Saccharomyces cerevisiae* (2UV8); ACP3: *Homo sapiens* (2CG5); ACP5: *Streptomyces roseofulvus* (1OR5); ACP8: *Aspergillus parasiticus* (2KR5); ACP10: *Brevibacillus parabrevis* (2JGP); ACP12: *E. coli* (2FQ1); ACP13: *E. coli* (2ROQ); and ACP17: *Lactobacillus casei* (1HQB).

others (Fig. 2). We found ACP3 and ACP10 domains in enzymes containing them by superimposing free-standing ACPs and extracting the former for RMSD calculations. In ACP10, we chose to superimpose the structures in the A/H conformer, very similar to other ACPs, one of three conformers in which ACP10 structures have been found.<sup>50</sup> Structures within putative families should have low RMSD<sub>ave</sub>

values and high  $P_{ave}$  values (average percentage of  $\alpha$ -carbon atoms of the amino acid residues between two structures compared for calculations). Shown in Table II are the structures used, RMSD<sub>ave</sub> values (ranging from 1.74 Å to 2.02 Å), and  $P_{ave}$  values (ranging from 84.8 to 94.4%). These findings further indicate that the structures represent members in the same families.

**Table II.** Acyl Carrier Protein Tertiary Structures

Family	RMSD <sub>ave</sub> (Å)	$P_{ave}$ (%)	Tertiary structures <sup>a</sup>	Total number of PDB accessions in family
ACP1	2.01	86.4	1HY8, 1KLP, 1LOH, 1X3O, 2DNW, 2EHS, 2FQ0, 2FVE 2KOO, 2KW2, 2KWL, 2L0Q, 2L3V, 2L4B, 2QNW	45
ACP2	—	—	2PFF	6 <sup>b</sup>
ACP3	2.02	94.4	2CG5 <sup>c</sup> , 2PNG	4
ACP5	1.75	84.8	1NQ4, 1OR5, 2AF8	14
ACP8	—	—	2KR5	1
ACP10	1.74	85.9	1DNY, 2GDW, 2JGP <sup>c</sup> , 2JU2, 2VSQ <sup>c</sup>	9
ACP12	—	—	2FQ1	1
ACP13	—	—	2ROQ	1
ACP17	—	—	1DV5	2 <sup>b</sup>

<sup>a</sup> For ACP1, ACP3, ACP5, and ACP10 all structures listed were superimposed to yield reported RMSD<sub>ave</sub> and  $P_{ave}$  values; for others a representative sequence is shown. Only one structure per species within a family was superimposed.

<sup>b</sup> RMSD not calculated, as all resolved ACP domains in existing tertiary structures come from the same species.

<sup>c</sup> Only ACP domain of a larger structure

**Table III.** Normal Mode Analysis Comparisons

Structural comparison	PDB designations	Highest overlap between two modes	RMSD <sub>f-ave</sub> (Å)	P <sub>f-ave</sub> (%)
Same structure, two conformers	2EHS, 2EHT	0.47	0.90	99.1
Same family, two bacterial structures	1T8K, 1X3O	0.42	1.51	80.9
Same family, bacterial and eukaryotic	1T8K, 2QNW	0.47	1.05	95.7
ACP1 to ACP5	1T8K, 2AF8	0.65	2.70	61.2
ACP1 to ACP17	1T8K, 1HQB	0.58	2.08	85.9
ACP5 to ACP17	2AF8, 1HQB	0.64	2.67	71.1

If different families have members with tertiary structures that can be closely aligned, even though their primary structures are different, then they may have a more distant common protein ancestor and may be gathered into the same clan. To determine this, one tertiary structure each from ACP1 (PDB accession code 1X3O), ACP2 (2UV8), ACP3 (2CG5), ACP5 (1OR5), ACP8 (2KR5), ACP10 (2JGP), ACP12 (2FQ1), ACP13 (2ROQ), and ACP17 (1HQB) (Fig. 3) were superimposed. Of these, ACP2, ACP3, ACP8, ACP10, and ACP12 were domains in larger enzymes. The first four helices of the ACP2 structure were well superimposed on the structures of the other families. The superposition of the structures of these nine families resulted in an RMSD<sub>ave</sub> value of 2.34 Å and a P<sub>ave</sub> value of 76.2%, indicating that at least those families may descend from a distant common ancestor and may be gathered into the same clan. The moderate percentages of common residues between different pairs of families (Supporting Information Table S1 and Fig. S1) suggest that other ACP families may also have the same common ancestor. However, this conclusion must await determination of tertiary structures of members of these families.

### Normal mode analysis of freestanding ACP structures

We subjected seven tertiary structures of freestanding apo-ACPs to normal mode analysis, using the Anisotropic Network Model (ANM) web server.<sup>51</sup> We did not attempt to subject ACPs that were part of larger proteins to normal mode analysis, since their structures and fluctuations are most likely affected by their proximity of other parts of the protein. ACP1 members were PDB accession codes 2EHS and 2EHT (two conformers of the same protein) from *Aquiflex aeolicus*, 1T8K from *Escherichia coli*, 1X3O from *Thermus thermophilus*, and 2QNW from *Toxoplasma gondii*. The one ACP5 member was 2AF8 from *Streptomyces coelicolor*, while 1HQB in ACP17 was from *Lactobacillus rhamnosus*. All are produced by bacteria but one; *T. gondii* is an apicomplexan protozoan.

The residue fluctuations of the five slowest modes of each structure are plotted in Supporting Information Figure S3, and ANM mode visualizations of the slowest mode of each structure are shown in Supporting Information Figure S4.

Chain ends, especially the N-terminal one, and loops were predicted to fluctuate more than the three long and parallel  $\alpha$ -helices (helices I, II, and IV), which move very little. The region between helices II and IV, which in ACP1 structures comprises the loop between helices II and III and the crosswise helix III, strongly fluctuated in many structures. In the ACP5 and ACP17 structures, helix III is replaced by a loop. This region was proposed to act as a gatekeeper in the acyl chain delivery process,<sup>52</sup> as the acyl chain is exposed through a fissure following a conformational rearrangement. The acyl chain delivery process is unknown; sword-unsheathed or switchblade-like mechanisms have been proposed. Lower fluctuations are seen in the loop between helices I and II. The N-terminal region of helix II, where the 4'-phosphopantetheine prosthetic group is attached, shows relatively higher fluctuations than the rest of helix II. In general, the vibrations of the first five slowest modes in any structure are moderately to strongly correlated in location, although much less so in magnitude.

We compared the normal modes of two conformers (2EHS and 2EHT) of the same structure in ACP1, of two bacterial structures (1T8K and 1X3O) in ACP1, of bacterial (1T8K) and eukaryotic (2QNW) structures in ACP1, of ACP1 (1T8K) and ACP5 (2AF8) structures, of ACP1 (1T8K) and ACP17 (1HQB) structures, and of ACP5 (2AF8) and ACP17 (1HQB) structures. They were analyzed as explained in the Computational Methods section. Results are summarized in Table III, overlap charts are shown in Supporting Information Figure S5, and residue fluctuation comparisons of the three most overlapped modes for each case are shown in Supporting Information Figure S6.

The two conformers of the same structure showed a very low frame-averaged RMSD (RMSD<sub>f-ave</sub>, as defined in the Computational Methods section), showing that their structural similarity was conserved throughout the motion of the slowest normal mode. Several of their slowest normal modes showed some overlap (Supporting Information Fig. S5). Residue fluctuations (Supporting Information Fig. S6) show that 2EHT has larger amplitudes than 2EHS in the N-terminal region, while 2EHS has larger amplitudes in the central and C-terminal regions. Their fluctuations appear in very similar residue locations.

**Table IV.** Tertiary Structure Prediction Validation

Target structure	Template structure	Amino acid identity (%) <sup>a</sup>	RMSD <sub>p-c</sub> (homology) (Å)	RMSD <sub>p-c</sub> (MD) (Å)
ACP1	ACP5	23.4	1.17	1.42
ACP1	ACP17	20.8	1.44	1.48
ACP5	ACP1	21.7	1.97	1.87
ACP5	ACP17	20.5	2.10	1.96
ACP17	ACP1	20.0	1.73	1.93
ACP17	ACP5	21.3	1.78	1.88

<sup>a</sup> Percentage of same residues in same position between template and target structures divided by the number of residues in the template.

Two different structures within ACP1 display lower overlaps than two conformers of the same structure (Supporting Information Fig. S5). Values of RMSD<sub>f-ave</sub> are higher than between two conformers of the same structure, but still are <2 Å, as would be expected of two structures within a family. In both cases, fluctuations appear conserved in residue location but not so in amplitude (Supporting Information Fig. S6), although they appear more correlated between the bacterial and eukaryotic structures.

Comparing the normal modes between structures in different families showed that the slowest mode of each structure was the most highly overlapped (Supporting Information Fig S5), especially between ACP1 and ACP5 and between ACP5 and ACP17. The slowest modes in each structure display an overall twisting rather than stretching of discrete loops, and overlap highly due to their low flexibility. Their RMSD<sub>f-ave</sub> values are higher than in previous cases, showing structural differences between two families. However, within the most overlapped normal mode, structural differences are not great. Residue fluctuation for modes other than the most overlapped (Supporting Information Fig. S6) are poorly correlated in both amplitude and location, especially between ACP1 and ACP5 and between ACP5 and ACP17, but less so between ACP1 and ACP17.

Even though overlap charts (Supporting Information Fig. S5) do not differ much for comparisons within and between families, residue fluctuations (Supporting Information Fig. S6) are more conserved between two structures within a family than between two structures in different families. However, fluctuation amplitudes can be quite different in different structures produced by organisms in the same genus and, especially noteworthy, in different conformers of the same structure. Differences in fluctuation amplitudes may be due to slight crystal packing effects that could be present in loops.

#### Determination of tertiary structures of freestanding ACPs by computation

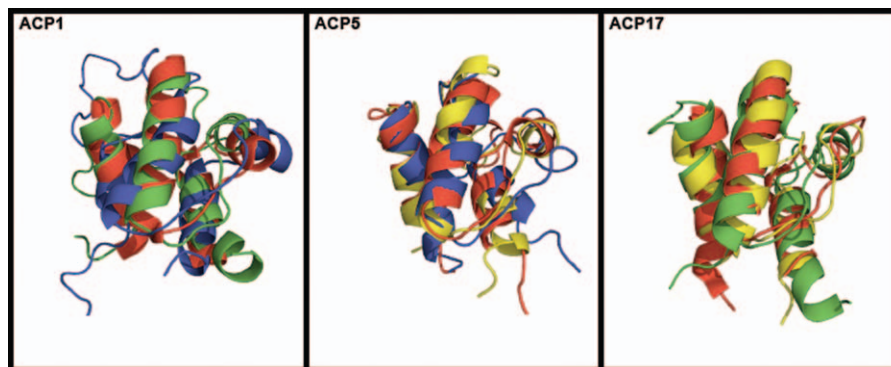
To determine whether the homology modeling and molecular dynamics (MD) protocol described in the Computational Methods section is a trustworthy method of determining tertiary structures of freestanding ACPs whose structures were previously

unknown, we used experimentally determined freestanding structures of two families from ACP1 (2EHS), ACP5 (1OR5), and ACP17 (1HQB) as templates to predict the structure of the third family. We did not use ACP domains as templates or predict unknown ACP domains of large multidomain enzymes, as their folded states may be influenced by nearby sections of the larger protein.

Results are shown in Table IV. The amino acid identities, based on pairwise alignments between individual family members, between template and target structures vary between 20.0 and 23.4%. Two RMSD values between predicted and crystal tertiary structures (RMSD<sub>p-c</sub>) are reported: first the value of the predicted structure by homology modeling to its crystal structure, and second the predicted structure after MD refinement to its crystal structure. RMSD<sub>p-c</sub> values of homology modeling predictions vary between 1.17 Å and 2.10 Å, and RMSD<sub>p-c</sub> values after MD refinement vary between 1.42 Å and 1.96 Å. The predicted structures superimposed to their crystal structures can be seen in Figure 4. As indicated by low RMSD<sub>p-c</sub> values (<2 Å) and by visual inspection, the protocol described can predict freestanding ACP structures when the template used for homology modeling is at least 20% identical in sequence to the target.

We then attempted to predict tertiary structures for sequences in ACP4, ACP15, and ACP16, which contain freestanding members with no known structures. The structure of *Bacillus subtilis* AcpK (UniProt ID Q7PC63) in ACP4 was predicted using as a template 2EHS from ACP1, with which it has an amino acid identity of 25.6%. The predicted structure of ACP4 (Fig. 5) is very similar to structures in other ACP families, and it clearly shows a cavity and three long helices (helices I, II, and IV). When the predicted ACP4 structure was superimposed to structures in ACP1 (2EHS), ACP5 (1OR5), and ACP17 (1HQB), the resulting RMSDs were 1.92 Å, 2.12 Å, and 2.21 Å, respectively.

Subjecting *Malonomonas rubra* malonate decarboxylase (UniProt ID: O06925) from ACP15 and *E. coli* citrate lyase (UniProt ID: P69330) from ACP16 to the homology modeling and MD protocol resulted in unfolded structures that lacked the standard three parallel helices and cavity seen in



**Figure 4.** Predicted tertiary structure of an ACP1 member using ACP5 (green) and ACP17 (blue) tertiary structures as templates, compared with a known ACP1 (red) tertiary structure; of an ACP5 member using ACP1 (yellow) and ACP17 (blue) tertiary structures as templates, compared to a known ACP5 (red) tertiary structure; of an ACP17 member using ACP1 (yellow) and ACP5 (green) tertiary structures as templates, compared to a known ACP17 (red) tertiary structure. Known structure templates: ACP1: 2EHS; ACP5: 1OR5; ACP17: 1HQB.

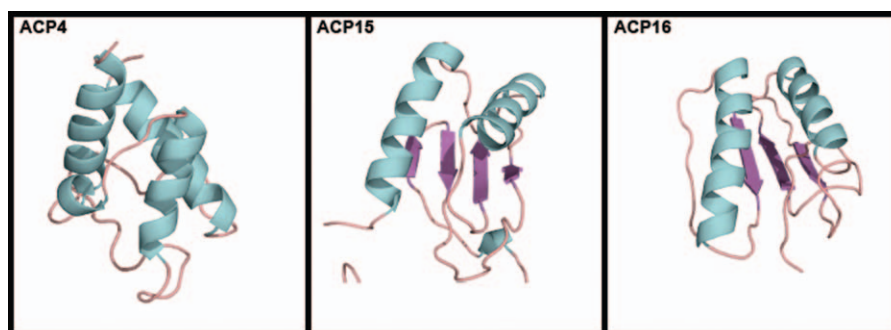
other ACPs. However, they are most likely not intrinsically unstructured. Ten models were then predicted for each sequence using threading programs. Four of the 10 ACP15 models (1, 5, 6, and 8) displayed two main  $\alpha$ -helices and four  $\beta$ -strands (Supporting Information Fig. S7), while seven of the 10 ACP16 models (1, 2, 4–8) showed two main  $\alpha$ -helices with three  $\beta$ -strands. Representative models of the four and seven similar structures of ACP15 and ACP16, respectively, are shown in Fig. 5. When superimposed, the four similar models in ACP15 have  $\text{RMSD}_{\text{ave}}$  and  $P_{\text{ave}}$  values of 0.73 Å and 88.7%, and the seven similar models of ACP16 show values of 0.99 Å and 91.3%.

ACP15 and ACP16 carrier proteins are used in reactions that break substrates into smaller units, unlike other ACPs, which are involved in building large molecules from small units. Also, ACP15 and ACP16 do not use the same 4'-phosphopantetheine prosthetic group present in other ACPs. These facts further suggest that ACP15 and ACP16 may be structurally different from other ACP families, and if this is correct, ACP15 and ACP16 members may descend from different ancestors than most or all of

the other ACPs. The multiple sequence alignment (Fig. 1), phylogenetic tree (Fig. 2, Supporting Information Fig. S2), and threading predictions (Fig. 6, Supporting Information Fig. S7) support this. However, only experimentally determined structures will show if ACP15 and ACP16 are indeed different from other ACP families.

#### Concluding Remarks

Considering that ACPs, except those of ACP2, are all of roughly the same length and in nearly all cases they bind and activate acyl chains prior to their being subjected to enzymatic catalysis, it is noteworthy that they can be separated into 16 families by their highly significant differences in primary structure. It is perhaps equally noteworthy that many of these families have very similar tertiary structures, signifying that they may be descended from a common distant ancestor. We predicted the normal vibrational modes of freestanding ACPs, finding them more conserved within families. Also, we extended knowledge of three families possessing freestanding ACPs, one family by predicting its characteristic tertiary structure to be like those of



**Figure 5.** Tertiary structure of ACP4 (*Bacillus subtilis* AcpK) predicted by homology modeling and MD. Representative models of tertiary structures of ACP15 (*Malonomonas rubra* malonate decarboxylase ACP) and ACP16 (*E. coli* citrate lyase ACP) predicted by threading.



many other ACP families, and two families by apparently showing that they do not have similar structures.

## Computational Methods

### Family identification and phylogeny

Families were identified following the procedures outlined earlier.<sup>16–18</sup> In brief, families were based on sequences with evidence at protein level from the UniProt database<sup>53</sup> or the literature, and populated by use of BLAST<sup>54</sup> (downloadable version 2.2.19) to query the nonredundant (nr) protein sequence database.<sup>55</sup> A cutoff *E*-value of 0.001 (likelihood that similarity between query and compared sequence is due to chance) was used as the exclusion criterion. Families were confirmed with multiple sequence alignments of the retrieved sequences, using MUSCLE 3.6,<sup>56</sup> and with tertiary structure superimpositions.

Phylogenetic trees were built using Molecular Evolutionary Genetics Analysis 5 (MEGA5),<sup>57</sup> based on a multiple sequence alignment containing three sequences per family made with MUSCLE 3.6. The minimum evolution algorithm<sup>58</sup> was used, gaps were treated with pairwise deletion, and an amino acid Jones-Taylor-Thornton matrix<sup>59</sup> was chosen as the model. A bootstrap test with 1000 replicates was performed to further verify the results.

### Tertiary structure superposition and RMSD calculations

In this study, all tertiary structures were superimposed with MultiProt.<sup>60</sup> All RMSD values were calculated between  $\alpha$ -carbon atoms using MATLAB, to consider the most possible  $\alpha$ -carbon atoms in the calculation; MultiProt reports the RMSD for only aligned residues. Values of RMSD (between two structures),  $\text{RMSD}_{\text{ave}}$  (between three or more structures), and  $P_{\text{ave}}$  (the average percentage of  $\alpha$ -carbon atoms of the amino acid residues used to calculate the RMSD between two compared structures) were calculated, as explained in detail in the Supporting Information in Cantu *et al.*<sup>17</sup> When a NMR-resolved structure was superimposed, only the lowest-energy frame was used, chosen following single-point energy calculations with the ff99SB AMBER force field.<sup>61</sup>

### Normal mode analysis

Normal modes of individual ACPs were calculated using the ANM web server,<sup>51</sup> by assuming that a molecule's natural vibrations can be predicted by attaching a spring of uniform force constant to each  $\alpha$ -carbon atom of the molecule and allowing the system to oscillate. The cutoff distance and distance weights were changed until the best correlation between theoretical and experimental B-factors was

achieved. Structures were submitted individually to the ANM server, which computed the vibrational normal modes of each molecule and returned an animated PDB file for the 20 slowest vibrational modes, their associated eigenvectors, and residue fluctuations. No significant crystal packing effects were found when viewing the symmetric molecules of the PDBs submitted.

To compare the normal modes between two structures, overlap charts showing the 20 slowest vibrational modes of two structures were made by taking the dot product of the eigenvectors from both structures, yielding an  $20 \times 20$  overlap chart where each cell ranges in value from 0 (no overlap) to 1 (complete overlap). From this, the highest overlapped mode was chosen, and the animated structure PDB files of such mode were separated into individual PDB files for each frame. Then the frame from one structure was superimposed with the corresponding frame in the other, and RMSD and *P* values was calculated. This was done for each frame, and the averages among all frames were taken, resulting in values of  $\text{RMSD}_{\text{f-ave}}$  and  $P_{\text{f-ave}}$ . This notation is used here to differentiate the RMSD between two structures averaged over different frames in a normal mode (Table III), from  $\text{RMSD}_{\text{ave}}$ , which refers to the average RMSD of superimposing three or more structures (Table II).

### Tertiary structure determination

A homology modeling and MD protocol was used for structure predictions. Homology modeling was done with the I-TASSER server,<sup>62,63</sup> using other ACP structures as templates without alignment or restraints.

The resulting predicted structures were loaded in AmberTools1.4 using the ff99SB AMBER force field. The structures were placed in a 12-Å TIP3P water box,<sup>64</sup> and  $\text{Na}^+$  ions were added to neutralize system charge.

The system was then simulated with the sander module of AMBER.<sup>65</sup> Solvent and ions were minimized first with 1000 steps while restraining the protein with 500 kcal/Å-mol weights. An unrestrained 10,000-step minimization followed. The system was then heated from 0 K to 300 K at constant volume for 25 to 100 ps while weakly restraining the protein with 10 kcal/Å-mol weights. Finally, the system was equilibrated at constant pressure and temperature and run unrestrained for 1.5 to 2 ns.

Sequences were predicted to not be intrinsically unfolded using the IUPred<sup>66</sup> server that bases its predictions by calculating inter-residue energy interactions.

Threading predictions were done using the LOMETS<sup>67</sup> server, which generates tertiary structures from sequences using eight different known threading programs. Unlike homology modeling, no

templates are specified. The 10 best threading models from the output were taken.

RMSDs between predicted structures to experimental crystal structures in the validation runs (Table IV) are labeled as RMSD<sub>p-c</sub> to differentiate it from the other RMSDs used here.

### Acknowledgments

K.C. from Johns Hopkins University was part of a research experiences for undergraduates program associated with the Engineering Research Center. The authors thank Michael T. Zimmermann (Iowa State University) for helpful discussions about normal mode analysis and Christopher D. Warner (The Scripps Research Institute, Florida) for discussions about crystal packing effects and structure.

### References

1. Rock CO, Cronan JE Jr. (1979). Re-evaluation of the solution structure of acyl carrier protein. *J Biol Chem* 254:9778–9785.
2. Zhang YM, Marrakchi H, White SW, Rock CO (2003) The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J Lipid Res* 44:1–10.
3. Sharma AK, Sharma SK, Surolia A, Surolia N, Sarma SP (2006) Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from *Plasmodium falciparum* provides insight into the mechanism of activation of ACPs. *Biochemistry* 45: 6904–6916.
4. Roujeinikova A, Baldock C, Simon WJ, Gilroy J, Baker PJ, Stuitje AR, Rice DW, Slabas AR, Rafferty JB (2002) X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* 10:825–835.
5. Zornetzer GA, Fox BG, Markley JL (2006) Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* 45:5217–5227.
6. Roujeinikova A, Simon WJ, Gilroy J, Rice DW, Rafferty JB, Slabas AR (2007) Structural studies of fatty-(acyl carrier protein) thioesterases reveal a hydrophobic binding cavity that can expand to fit longer substrates. *J Mol Biol* 365:135–145.
7. Chan DI, Stockner T, Tieleman DP, Vogel HJ (2008) Molecular dynamics simulations of the apo-, holo-, and acyl-forms of *Escherichia coli* acyl carrier protein. *J Biol Chem* 283:33620–33629.
8. Rafi S, Novichenok P, Kolappan S, Zhang X, Stratton CF, Rawat R, Kisker C, Simmerling C, Tonge PJ (2006) Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from *Escherichia coli*. *J Biol Chem* 281:39285–39293.
9. Ploskon E, Arthur CJ, Evans SE, Williams C, Crosby J, Simpson TJ, Crump MP (2008) A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains. *J Biol Chem* 283: 518–528.
10. Chan DI, Vogel HJ (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem J* 430:1–19.
11. Kim Y, Prestegard JH (1989) A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* 28:8792–8797.
12. Andrec M, Hill RB, Prestegard JH (1995) Amide exchange rates in *Escherichia coli* acyl carrier protein: correlation with protein structure and dynamics. *Protein Sci* 4:983–993.
13. Park SJ, Kim J-S, Son W-S, Lee BJ (2004) pH-Induced conformational transition of *H. pylori* acyl carrier protein: insight into the unfolding of local structure. *J Biochem (Tokyo)* 135:337–346.
14. Kim Y, Kovrigin EL, Eletr Z (2006) NMR studies of *Escherichia coli* acyl carrier protein: dynamic and structural differences of the apo- and holo-forms. *Biochem Biophys Res Commun* 341:776–783.
15. Zornetzer GA, Tanem J, Fox BG, Markley JL (2010) The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. *Biochemistry* 49:470–477.
16. Cantu DC, Chen Y, Lemons ML, Reilly PJ (2011) ThYme: A database for thioester-active enzymes. *Nucleic Acids Res* 39:D342–D346. Available at: <http://enzyme.cbirc.iastate.edu>.
17. Cantu DC, Chen Y, Reilly PJ (2010) Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci* 19:1281–1295.
18. Chen Y, Kelly EE, Masluk RP, Nelson CL, Cantu DC, Reilly PJ (2011) Structural classification and properties of ketoacyl synthases. *Protein Sci* 20:1659–1667.
19. Vanaman TC, Wakil SJ, Hill RL (1968) The complete amino acid sequence of the acyl carrier protein of *Escherichia coli*. *J Biol Chem* 243:6420–6431.
20. Brozek KA, Carlson RW, Raetz CRH (1996) A special acyl carrier protein for transferring long hydroxylated fatty acids to lipid A in *Rhizobium*. *J Biol Chem* 271: 32126–32136.
21. Cronan JE, Fearnley IM, Walker JE (2005) Mammalian mitochondria contain a soluble acyl carrier protein. *FEBS Lett* 579:4892–4896.
22. Leibundgut M, Jenni S, Frick C, Ban N (2007) Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* 316:288–290.
23. Jayakumar A, Tai MH, Huang WY, Al-Feel W, Hsu M, Abu-Elheiga L, Chirala SS, Wakil SJ (1995) Human fatty acid synthase: properties and molecular cloning. *Proc Natl Acad Sci USA* 92:8695–8699.
24. Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC (2006) Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the *pksX* pathway of *Bacillus subtilis*. *Proc Natl Acad Sci USA* 103:8977–8982.
25. Fernandez-Moreno MA, Martinez E, Boto L, Hopwood DA, Malpartida F (1992) Nucleotide sequences and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J Biol Chem* 267:19278–19290.
26. Bibb MJ, Sherman DJ, Omura S, Hopwood DA (1994) Cloning, sequencing and deduced functions of a cluster of *Streptomyces* genes probably encoding biosynthesis of the polyketide antibiotic frenolicin. *Gene* 142:31–39.
27. Kim E-S, Bibb MJ, Butler MJ, Hopwood DA, Sherman DH (1994) Sequences of the tetracycline polyketide synthase-encoding *otc* genes from *Streptomyces rimosus*. *Gene* 141:141–142.
28. Mathur M, Kolattukudy PE (1992) Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. *bovis* *Bacillus Calmette-Guerin*. *J Biol Chem* 267: 19388–19395.

29. Sirakova TD, Thirumala AK, Dubey VS, Sprecher H, Kolattukugy PE (2001) The *Mycobacterium tuberculosis* *pks2* gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J Biol Chem* 276:16833–16839.
30. Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. *Eur J Biochem* 192:487–498.
31. Feng GH, Leonard TJ (1995) Characterization of the polyketide synthase gene (*pksL1*) required for aflatoxin biosynthesis in *Aspergillus parasiticus*. *J Bacteriol* 177:6246–6254.
32. Watanabe A, Fujii I, Sankawa U, Mayorga ME, Timberlake WE, Ebizuka Y (1999) Re-identification of *Aspergillus nidulans* *wA* gene to code for a polyketide synthase of naphthopyrone. *Tetrahedron Lett* 40:91–94.
33. Austin MB, Saito T, Bowman ME, Haydock S, Kato A, Moore BS, Kay RR, Noel JP (2006) Biosynthesis of *Dicystostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem Biol* 2:494–502.
34. Nakano MM, Marahiel MA, Zuber P (1988) Identification of a genetic locus required for the biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *J Bacteriol* 170:5662–5668.
35. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* 252:675–679.
36. Stein T, Kluge B, Vater J, Franke P, Otto A, Wittmann-Liebold B (1995) Gramicidin S synthase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 5'-phosphopantetheine. *Biochemistry* 34:4633–4642.
37. Azad AK, Sirakova TD, Fernandes ND, Kolattukudy PE (1997) Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem* 272:16741–16745.
38. Mootz HD, Marahiel MA (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* 179:6843–6850.
39. Tsuge K, Ano T, Hirai M, Nakamura Y, Shoda M (1999) The genes *degQ*, *pps*, and *lpa-8* (*sfp*) are responsible for the conversion of *Bacillus subtilis*. *Antimicrob Agents Chemother* 43:2183–2192.
40. Hacker C, Glinski M, Hornbogen T, Doller A, Zocher R (2000) Mutational analysis of the N-methyltransferase domain of the multifunctional enzyme enniatin synthase. *J Biol Chem* 275:30826–30832.
41. Schrettl M, Winkelmann G, Hass H (2004) Ferrichrome in *Schizosaccharomyces pombe*—an iron transport and iron storage compound. *BioMetals* 17:647–654.
42. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 104:305–310.
43. Hendrickson L, Davis CR, Roach C, Nguyen DK, Aldrich T, McAda PC, Reeves CD (1999) Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem Biol* 6:429–439.
44. Gehring AM, Bradley KA, Walsh CT (1997) Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* 36:8495–8503.
45. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for the assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* 5:631–645.
46. Rusnak F, Sakaitani M, Drueckhammer D, Reichert J, Walsh CT (1991) Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the entF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. *Biochemistry* 30:2916–2927.
47. Berg M, Hilibi H, Dimroth P (1996) The acyl carrier protein of malonate decarboxylase of *Malonomas rubra* contains 2'-(5''-phosphoribosyl)-3'-dephosphocoenzyme A as a prosthetic group. *Biochemistry* 35:4689–4696.
48. Bott M, Dimroth P (1994) *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: localization, sequencing, and expression. *Mol Microbiol* 14:347–356.
49. Debabov DV, Heaton MP, Zhang Q, Stewart K, Lamballot RH, Neuhaus FC (1996) The D-alanyl carrier protein in *Lactobacillus casei*: cloning, sequencing, and expression of *dltC*. *J Bacteriol* 178:3869–3876.
50. Koglin A, Mofid MR, Lohr F, Schafer B, Rogov VV, Blum M-M, Mittag T, Marahiel MA, Bernhard F, Dotsch V (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 312:273.
51. Eyal E, Yang L-W, Bahar I (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 22:2619–2627. Available at: <http://ignmtest.cccb.pitt.edu/cgi-bin/annm/annm1.cgi>.
52. Colizzi F, Recanatini M, Cavalli A (2008) Mechanical features of *Plasmodium falciparum* acyl carrier protein in the delivery of substrates. *J Chem Inf Model* 48:2289–2293.
53. UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res* 36:D190–D195. Available at: <http://www.uniprot.org/>.
54. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman, DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. Available at: <http://blast.ncbi.nlm.nih.gov/Blast.cgi/>.
55. National Center for Biotechnology Information (NCBI). The nr Database. Available at <ftp.ncbi.nih.gov>. Last accessed: October, 2011.
56. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. Available at: <http://www.drive5.com/muscle/index.html>.
57. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739. Available at: <http://www.mega-software.net>.
58. Desper R, Gascuel O (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comp Biol* 9:687–705.
59. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282.
60. Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. *Proteins* 56:143–156. Available at: <http://bioinfo3d.cs.tau.ac.il/MultiProt/>.
61. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber

- force fields and development of improved protein backbone parameters. *Proteins* 65:712–725.
62. Zhang Y (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* 69:108–117. Available at: <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>.
  63. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5:725–738. Available at: <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>.
  64. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
  65. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolosvary I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, YE X, Wang J, Hseih M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalchenko A, Kollman PA (2010) AMBER11. University of California, San Francisco. Available at: <http://ambermd.org/>.
  66. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21:3433–3434. Available at: <http://iupred.enzim.hu/>.
  67. Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure predictions. *Nucl Acids Res* 35:3375–3382. Available at: <http://zhanglab.ccmb.med.umich.edu/LOMETS/>.