Title

Chapter Seventeen Computational structural enzymology methodologies for the study and engineering of fatty acid synthases, polyketide synthases and nonribosomal peptide synthetases

Authors

Schaub, Andrew J
Moreno, Gabriel O
Zhao, Shiji
et al.

Publication Date

2019

DOI

10.1016/bs.mie.2019.03.001

Peer reviewed

HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2020 May 04.

Published in final edited form as:

Methods Enzymol. 2019 ; 622: 375–409. doi:10.1016/bs.mie.2019.03.001.

Computational structural enzymology methodologies for the study and engineering of fatty acid synthases, polyketide synthases and nonribosomal peptide synthetases

Andrew J. Schaub[a], **Gabriel O. Moreno**[b], **Shiji Zhao**[c], **Hau V. Truong**[a], **Ray Luo**[d,*], **Shiou-Chuan Tsai**[e,*]

[a]Department of Chemistry, University of California, Irvine, CA, United States

[b]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, United States

[c]Mathematical, Computational and Systems Biology Program, Center for Complex Biological Systems, University of California, Irvine, CA, United States

[d]Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, CA, United States

[e]Department of Molecular Biology and Biochemistry, Chemistry, Pharmaceutical Sciences, University of California, Irvine, CA, United States

Abstract

Various computational methodologies can be applied to enzymological studies on enzymes in the fatty acid, polyketide, and non-ribosomal peptide biosynthetic pathways. These multi-domain complexes are called fatty acid synthases, polyketide synthases, and non-ribosomal peptide synthetases. These mega-synthases biosynthesize chemically diverse and complex bioactive molecules, with the intermediates being chauffeured between catalytic partners *via* a carrier protein. Recent efforts have been made to engineer these systems to expand their product diversity. A major stumbling block is our poor understanding of the transient protein-protein and protein-substrate interactions between the carrier protein and its many catalytic partner domains and product intermediates. The innate reactivity of pathway intermediates in two major classes of polyketide synthases has frustrated our mechanistic understanding of these interactions during the biosynthesis of these natural products, ultimately impeding the engineering of these systems for the generation of engineered natural products. Computational techniques described in this chapter can aid data interpretation or used to generate testable models of these experimentally intractable transient interactions, thereby providing insight into key interactions that are difficult to capture otherwise, with the potential to expand the diversity in these systems.

*Corresponding authors: ray.luo@uci.edu (R. Luo); sctsai@uci.edu (S-C Tsai).

1. Introduction to computational approaches for studying natural products

1.1 Introduction to natural products

Investigators in the field of natural product chemistry draw from many areas of focus including marine biology, ethnobotany, structural enzymology, genetics, and heterologous expression, to name a few (Dewick, 2009; Kinghorn, 2002). Fatty acids (FAs), polyketides (PKs) and non-ribosomal peptides (NRPs) are medically and industrially useful compounds that are assembled incrementally through the addition of extender units to an initial starter unit by fatty acid synthase (FAS), polyketide synthase (PKS), and nonribosomal peptide synthetase (NRPS) (Chan & Vogel, 2010; Hur, Vickery, & Burkart, 2012; Khosla, Herschlag, Cane, & Walsh, 2014; Staunton & Weissman, 2001a). While there are many high-resolution structures of FAS, PKS and NRPS, our understanding of protein dynamics, conformational changes, protein-protein interactions, and protein-substrate interactions is still limited. The focus of this chapter will be the application and development of computational methodologies for FASs, PKSs and NRPSs, including molecular modeling and molecular dynamic (MD) simulation.

In recent years, the fields of molecular simulation and natural product chemistry have received wide recognition. In 2013, the Nobel Prize in Chemistry was awarded to Drs. Martin Karplus, Michael Levitt, and Arieh Warshel for their contributions in theoretical chemistry that opened up the field for the simulations of macromolecules (Fersht, 2013). In 2015, Drs. Youyou Tu, William Campbell and Satoshi Omura were awarded the Nobel Prize in Physiology and Medicine for their discoveries of two natural products, artemisinin and the polyketide avermectin (Fig. 1) (Van Voorhis, Hooft van Huijsduijnen, & Wells, 2015).

A previous review by Zhang and Rock on the application of computational methods for FASs reviews this subfield up to 2003 (Zhang, Marrakchi, White, & Rock, 2003). Computation works on other classes of natural products include terpenoids, alkaloids, and phenylpropanoids are summarized in other excellent reviews (Ferrer, Austin, Stewart Jr., & Noel, 2008; Gershenzon & Dudareva, 2007; Kochanowska-Karamyan & Hamann, 2010; Matsuda & Abe, 2016; O'Connor & Maresh, 2006). This chapter summarizes some key techniques that have been applied in our group to direct the product outcome of FASs, PKSs and NRPSs. The development and application of these computational methods bridges a major knowledge gap in our understanding of protein dynamics involved in the biosynthesis of these natural products.

1.2 Introduction to enzymatic machinery

FAS, PKS and NRPS are large, multi-domain enzyme complexes (Fig. 2). Their intermediate products, often highly unstable, are shuttled between the catalytic domains *via* acyl carrier proteins (ACPs; in FAS and PKS) or peptidyl carrier proteins (PCPs; in NRPS) in a well-choreographed order that results in the biosynthesis of natural products with high fidelity. ACP and PCP are sequential and structural homologs that share the four-helix bundle fold. The growing intermediate is covalently attached to a conserved serine on the carrier protein (CP). The mature product is ultimately released from the PPant-CP by

cleaving the thioester bond through enzyme-catalyzed hydrolysis or cyclization to generate the final product (Fig. 2) (Du & Lou, 2010).

1.3 Bioinformatics

Traditional computational approaches to studying the structures and functions of FAS, PKS and NRPS are often restricted to bioinformatics, which is not the focus of this review, and is covered in other outstanding reviews (Fischbach & Voigt, 2010; Guider & Moore, 2009; Keller, Turner, & Bennett, 2005; Moore et al., 2002). We will, however, present a brief overview of some tools available for bioinformatics analysis. Traditionally, secondary metabolites were discovered through activity-guided screens. Advances in sequencing and genomics have provided investigators an *in silico* way of identifying secondary metabolites through gene cluster analysis (Helfrich, Reiter, & Piel, 2014). In multi-modular FASs, PKSs, and NRPSs, one extender unit is added for every chain extension, and it is sometimes possible to predict the chemical structure of the final product based on the sequence of genes within the cluster. However, there exist iterative types, which are capable of using the same module more than once for multiple extensions, adding an extra layer of complexity in chemical structure prediction (Gaitatzis et al., 2002). There are several prediction tools available to analyze PKS and NRPS gene clusters, which have been summarized in reviews by Piel and Boddy, and include SMURF, AntiSMASH, NaPDoS, NP.searcher, ClustScan, CLUSEAN, PKMiner, and NRPS-PKS (Adamek, Spohn, Stegmann, & Ziemert, 2017; Blin et al., 2017; Boddy, 2014; Khaldi et al., 2010; Medema et al., 2011; Ziemert et al., 2012). Another excellent resource is Prediction Informatics for Secondary Metabolites (PRISM), a software platform with a web component which can aid in the prediction of final products generated from PKS and NRPS gene clusters (Skinnider et al. , 2015, 2016). These prediction tools provide investigators a method to identify or predict new products.

1.4 Molecular dynamics

Experimental techniques, including X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM), are powerful tools that allow us to visualize biomolecules. These techniques are complementary to each other, with X-ray crystallography providing high-resolution details, NMR providing information on dynamics and cryo-EM providing information on large-scale domain dynamics of large structures, often in the mega-dalton range. Molecular dynamics complements these experimental techniques by offering powerful explanatory and predictive capabilities.

Molecular behavior at the atomic level is described by the time-dependent Schrodinger equation (Eq. 1), for which no analytical solution exists for non-trivial systems, and computationally impracticable even on today's fastest computers (Dror, Dirks, Grossman, Xu, & Shaw, 2012).

$$i\hbar \frac{\partial}{\partial t}\left|\Psi(r,t)\right\rangle = \widehat{H}\left|\Psi(r,t)\right\rangle \quad (1)$$

Using the Born-Oppenheimer approximation, the wave function of a molecule can be described in terms of its nuclear and electronic components (Born & Oppenheimer, 1927).

This allows us to treat atoms as particles, and bonds can be represented as springs, thus providing a framework to represent our target systems classically. In MD simulations, we can obtain particle motion as a function of time, and thus resolve time regimes far beyond traditional structural biology techniques. Because MD simulations can sample different time-scales, we can gain valuable insight into protein-protein and protein-substrate interactions, as well as conformational changes. Karplus et al. classifies simulations into three main types (Karplus & McCammon, 2002). The purpose of the first type is to sample configuration space, and this is used in X-ray crystallography refinement through annealing protocols. The second type is used to describe systems at equilibrium, with the goal of obtaining thermodynamic data. These first two types can also be accomplished through Monte Carlo simulations. The final type is to simulate the actual dynamics over time, for which molecular dynamics (MD) is the only option. Investigators can perform a variety of simulations and analyses using traditional MD programs such as AMBER, CHARMM, GROMACS, and NAMD (Brooks et al., 2009; Case et al., 2018; Kumar et al., 2008; Phillips et al., 2005; Pronk et al., 2013). The frontier area is MD method development, which is still pushing forward based on the new demands of MD users.

Modeling enzymes from FAS, PKS and NRPS presents a challenge, because traditional MD packages are unable to model the long and flexible PPant-bound intermediates observed in these systems. Therefore, a major hurdle that MD faces in these systems is the parameterization of electrostatic and bond properties for the covalently-bound starter units, extender units, and intermediates (Barajas et al., 2015; Jackson et al., 2016; Nguyen et al., 2014).

1.4.1 Development of a PPant force field—Current force fields support the modeling of standard amino acids, nucleic acids, sugars, and lipids (Dickson et al., 2014; Skjevik, Madej, Walker, & Teigen, 2012). A host of non-standard cases can also be investigated, thanks in part to the development of intrinsically disordered protein (IDP), non-canonical amino acid (NCAA), phosphorylated amino acids, and post-translational modified (PTM) force fields (Homeyer, Horn, Lanig, & Sticht, 2006; Khoury et al., 2014; Khoury, Thompson, Smadbeck, Kieslich, & Floudas, 2013; Song, Luo, & Chen, 2017; Song, Wang, et al., 2017). However, at present, no force field exists that is capable of modeling a PPant group bound to a protein. In order to remedy this situation, we recently developed a force field to model PPant-bound intermediates, starter units, and extender units (Schaub et al., 2019, submitted). This represents the first force field specifically designed to probe carrier protein-mediated natural product biosynthesis through *in silico* methods, and provides a framework for an all-inclusive force field capable of handling natural and unnatural building blocks.

Parameters were obtained from the AMBER ff14SB force field when available. Missing parameters were adopted from the General Amber Force Field (GAFF) to serve as a preliminary parameter set to newly defined atom types as was done in the initial LIPID 11 force field (Skjevik et al., 2012). Forcefield_NCAA, another AMBER force field, utilized a similar approach with no additional parameterization, highlighting the promise of such aninitial approach for force field development (Khoury et al., 2014).

Special care was taken during the modeling of the divalent phosphate groups with the employment of Mg^{2}+ and Na + counterions to reproduce experimental values obtained by Schneider, Kabelac, and Hobza (1996). The appropriate basis set to match MP2/aug-cc-pVTZ and experimental values using the B3LYP functional was found to be 6–311 + G(2d,p). Additional diffuse and polarizable basis functions were needed to find good agreement with the O-P bond lengths and angles of the phosphate groups.

Normal mode frequencies were obtained in AMBER using a low mode search as well as from electronic structure calculations (Fig. 3). Calculations were performed at the B3LYP/6–311 + G(2d,p) and MP2/aug-cc-pVDZ levels of theory for all the fragments. Fragments containing parameters that were adopted from GAFF utilized additional electronic structure calculations performed at the MP2/aug-cc-PVTZ level of theory to match the level of theory used in the Forcefield_NCAA, 003, and Amber FB15 force fields (Duan et al., 2003; Khoury et al., 2014; Lee & Duan, 2004; Wang et al., 2017). Projections of normal modes agreed well between Amber and Gaussian. The frequencies observed in the 450–1100 cm^{-1} range include C-O and O-P bond-stretching, O-P-O twisting, O-P-O wagging, and O-P-O scissoring. The GAFF parameters were in good agreement and able to reproduce the normal modes for this range. To perform normal mode analysis on the thioester portion, an *S*-methyl thioacetate fragment was generated. S-C bond stretching was observed at 645 and 749cm^{-1} O-C-S scissoring observed at 439 cm^{-1} and the characteristic intense carbonyl stretch for thioesters at 1720 cm^{-1} at the MP2/aug-cc-pVTZ level of theory.

1.5 Homology modeling

In the absence of an experimentally-determined structure, homology modeling is often utilized to generate a predicted protein structure (Sliwoski, Kothiwale, Meiler, & Lowe, 2014). Critical Assessment of Protein Structure Prediction (CASP) simulations evaluate current homology modeling techniques *via* blind prediction every 2 years. Popular template-based models include Rosetta and I-TASSER (zhanglab.ccmb.med.urnich.edu/I-TASSER/) (Roy, Kucukural, & Zhang, 2010; Song et al., 2013; Yang et al., 2015). In the absence of a suitable template, *ab initio* folding methods are available. DeepMind made national headlines when it outperformed other groups on 43 of the toughest tests with an average margin of 15% accuracy improvement (Evans et al., 2018). QUARK, developed by the Zhang lab, also uses an *ab initio* folding approach and applies a force field-based method (zhanglab.ccmb.med.umich.edu/QUARK/) (Xu & Zhang, 2012).

1.6 Small molecule docking

Small-molecule docking focuses on non-covalent receptor and ligand interactions. However, FAS, PKS and NRPS covalently load their intermediates onto the PPant moiety, which is attached to the CP domain. Therefore, covalent docking approaches (as opposed to non-covalent docking) are needed. A recent review by Scarpino et al. evaluated available covalent docking tools, including AutoDock4, CovDock, FITTED, GOLD, ICM-Pro and MOE (Scarpino, Ferenczy, & Keseru, 2018). The six tools could model covalently bound ligands with 40–70% accuracy, with a lower-success rate for long and flexible ligands that are covalently bound (Abagyan, Totrov, & Kuznetsov, 1994; Corbeil, Williams, & Labute, 2012; Jones, Willett, Glen, Leach, & Taylor, 1997; Morris et al., 2009; Scarpino et al., 2018;

Zhu et al., 2014). Unfortunately, this is the case for the long, flexible PPant moiety; therefore, its covalent docking is a major issue that needs to be addressed. Covalent docking was used by our group (detailed below) to model substrate mimics, and quality was assessed by comparison to known substrate models for which NMR data is present in the protein data bank (PDB) (Shakya et al., 2014).

1.7 Protein-protein docking

The biannual Critical Assessment of Predicted Interactions (CAPRI) evaluates protein-protein docking simulations (Janin, 2002; Janin et al., 2003; Wodak & Janin, 2017). HADDOCK (haddock.science.uu.nl/) and CLUSPRO (cluspro.bu.edu) have consistently performed well in CAPRI competitions (Comeau, Gatchell, Vajda, & Camacho, 2004; Dominguez, Boelens, & Bonvin, 2003; Janin et al., 2003; Kozakov et al., 2013, 2017; Kozakov, Brenke, Comeau, & Vajda, 2006; van Zundert et al., 2016).

1.8 Free energy calculations

The accurate estimation of the free energy of binding for protein-protein and protein-substrate interactions in FAS, PKS and NRPS is crucial for the engineering of these megasynthases. Since there are often many modules, each with several domains, obtaining experimental free energy values can be difficult, if not impossible, using experimental techniques. The Helmholtz free energy, A, can be expressed in terms of a partition function (Eq. 2):

$$\mathrm{A} = -\beta^{-1} \ln Q(N, V, T), \tag{2}$$

where N, V, and T is defined as the canonical ensemble with a fixed number of atoms, volume and temperature. Free energy calculation approaches, including thermodynamic integration (TI) and free energy perturbation (FEP) methods, involve calculating ΔA. TI, FEP and histogram methods can be computationally very expensive as adequate sampling is needed to ensure convergence (Leach, 2001).

Thermodynamic integration (TI) calculates the free energy change by transforming one configuration state of interest into another one, *via* a parameter λ (Michel & Essex, 2010). The formula for TI is (Eq. 3):

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{3}$$

In this equation, $V(\lambda)$ is the λ-coupled Hamiltonian that corresponds to initial state for $\lambda = 0$ and end state for $\lambda = 1$. Due to the difficulty of obtaining an analytical solution, this integral is typically calculated numerically by performing a series of simulations corresponding to different λ points from 0 to 1. Simonson et al. used TI to calculate the proton pK_a shifts of a number of ionizable residues in thioredoxin, where the protonated and deprotonated residues corresponded to initial and end states (Simonson, Carlsson, & Case, 2004).

An alternative approach of calculating free energy difference is Molecular Mechanics Poisson–Boltzmann/Generalized Born Surface Area (MM-PB/GBSA) (Homeyer & Gohlke, 2012). MM-PB/GBSA belongs to thecate-gory of end-state methods, which do not require the simulation of intermediate states as in TI (Miller et al., 2012). This method involves the calculation of (1) the solvation free energy contribution by solving the linearized Poisson Boltzmann or Generalized Born equation, (2) the gas phase energy contribution by performing a molecular mechanics calculation, and (3) the entropy contribution by performing normal mode analysis. In our group, Ellis et al. performed MM-PBSA to calculate the binding free energy between the enzyme DpsC and its native substrate or the recently developed oxetane-based substrate mimics in order to assess the validity of the mimics (Section 2.2.3) (Ellis et al., 2018). MM-PBSA has also been used to assess the contributions of interface residues in FAS (Section 2.1.3).

Although both methods are widely used during free energy calculation, they have certain limitations. TI is theoretically rigorous but computationally expensive. MM-PB/GBSA, on the other hand, has a lower computational cost but limited accuracy, especially for highly charged systems (Wang, Hou, & Xu, 2006). Homeyer et al. developed a user friendly Free Energy Workflow (FEW) allowing users to select different methods depending on the tradeoff between desired accuracy and available computational resources (Homeyer & Gohlke, 2013). There are three methods available in FEW, including MM-PB/GBSA, TI and Linear Interaction Energy (LIE), another end state free energy method (Aqvist, Luzhkov, & Brandsdal, 2002).

2. Applications in the engineering of natural product biosynthesis

2.1 Fatty acid synthase (FAS)

There are two types of FASs. Type I FAS, employed by eukaryotes and some bacteria, links the enzyme domains covalently, forming a “megasynthase” (Maier, Leibundgut, & Ban, 2008; Smith & Tsai, 2007). Type II FAS, found in plants and bacteria, uses stand-alone enzymes. In both types, ACP shuttles the growing intermediates (White, Zheng, & Zhang, 2005). Here we focus on our work of the upstream acyl-CoA carboxylase (ACCase) that provides malonyl-CoA, a building block used by FAS.

2.1.1 Case study 1: Virtual screening to identify AccDS inhibitors—Acyl-CoA carboxylase (ACCase) biosynthesizes malonyl-CoA, a building block of FAS. *Mycobacterium tuberculosis* encodes six ACCase β subunits (accD1–6), and it was not clear why the pathogen needs six ACCases. Previous studies suggested that AccD4, AccD5 and AccD6 are more likely to be involved in providing fatty acids for the cell envelope of *M. tuberculosis* (Takayama, Wang, & Besra, 2005). In order to understand the structures and functions of these ACCases, Lin et al. determined the 2.9 Å *apo* structure of AccD5 (Fig. 4) (Lin et al., 2006). Kinetic studies showed that ccD5 has a high preference for propionyl-CoA, which produces methylmalonyl-CoA, the building block of mycocerosic, phthioceranic, hydroxyphthioceranic, mycosanoic and mycolipenic acids.

AccD5 represents an attractive target for the development of anti-tuberculosis therapeutics. To identify potential inhibitors of AccD5, an *in silico* virtual screening approach was

performed utilizing the National Cancer Institute (NCI) diversity set (1990 compounds) and the ChemDB database from the University of California, Irvine (>5 million compounds) (Chen, Swarnidass, Dou, Bruand, & Baldi, 2005; Lin et al., 2006). Multiple rounds of screening were performed, and the top hits were selected for inhibition assays. A lead compound, NCI-65828, with a K_i of 13.1 μM was identified, which binds AccD5 20 times tighter than the native substrate. This lead compound was used for a second *in silico* round of screening to find similar compounds, which identified that the top ranked compounds all included a naphthalene ring, two phenyl rings, diazo groups and a sulfate which might be important for AccD5 binding (Swarnidass et al., 2009). The above study demonstrates that virtual screening and small molecule docking methods, in combination with experimental approaches, can be applied in structure-based drug/inhibitor design.

2. 1.2 Case study 2: FabA—In the type II FAS from *E. coli,* the ACP is called AcpP, and the dehydratase is called FabA. In 2014, Nguyen et al. reported the crystallographic, NMR and MD studies of AcpP-FabA complex stabilized *via* a mechanism-based cross-linker (PDB ID: 4KEH) (Nguyen et al., 2014). The AcpP-FabA complex exists as a homodimer (Fig. 5). NMR and MD simulations were used to examine the protein dynamics of AcpP when it interacts with FabA. Specifically, Helix III of AcpP-FabA complex undergoes significant conformational change, when compared with the free AcpP structure. We hypothesized that Helix III was important for AcpP-FabA interaction. To test this hypothesis, $^{1}H/^{15}N$ heteronuclear single quantum coherence (HSQC) spectra were obtained of AcpP in the *holo* state or loaded as octanoyl-AcpP. Upon titration of FabA, chemical shift perturbations (CSPs) were observed spanning helices II and III. Using MD simulations coupled with accelerated molecular dynamics (AMD), (M arkwick & M cCammon, 2011) we showed by residual dipolar couplings (RDCs) that the NMR titration data corroborated with the MD simulation results. The AcpP-FabA crystal structure, NMR titrations, and MD simulations shed light on the protein dynamics that govern the sequestration of the growing acyl chain, as well as its subsequent release into the active site of FabA for dehydration.

2.1.3 Case study 3: FabB—Using a strategy similar to the above case that utilized a mechanism-based crosslinking probe to stabilize transient ACP-partner complexes, we successfully stabilized a second type II FAS di-domain complex from *E. coli* between AcpP and the β-ketoacyl-ACP-synthase I (FabB). FabB is responsible for fatty acid chain elongation *via* decarboxylative Claisen condensation. It is the rate-limiting step in fatty acid biosynthesis, and is of interest in biofuel development. Despite of its importance in primary metabolism, the atomic details of its dynamics and interactions with AcpP are lacking, hindering bioengineering efforts involving FabB (Chemler et al., 2015; Gajewski et al., 2017; Zhu & Cronan, 2015).

We solved the 2.4 Å crystal structure of cross-linked AcpP-FabB (Fig. 6) (Milligan et al, 2019, in revision). Similar to the AcpP-FabA structure, the majority of residues on AcpP in contact with its partner protein were primarily located on the conserved Helix II (Fig. 6C), and consisted of salt bridges between acidic residues of AcpP and basic residues of FabB (Byers & Gong, 2007). In contrast to the AcpP-FabA structure, contacts in the AcpP-FabB structure spanned a broader portion of Helix II, thus revealing differences in binding patterns

between AcpP-FabA *versus* AcpP-FabB. The helix III salt bridge interaction was observed in both, supporting the chain flipping mechanism that we previously hypothesized (Nguyen et al., 2014).

Several long MD simulations were performed to identify changes in AcpP protein dynamics upon interacting with FabB. The AcpPs were modeled as *apo* (no PPant), *holo* (empty PPant), C8 acyl-substrate loaded, or C10 acyl-product loaded (Fig. 7). These different trajectories were then compared using principal component analysis (PCA), which revealed a possible mechanism for communication and recognition between AcpP and FabB (Fig. 6D–E). Upon loading of the PPant with a C8 (or C10) FabB substrate, the α6-α7 region of FabB undergoes conformational change, which is in close proximity to interact with another highly flexible helix (α10) of the same FabB monomer. Both motifS also interact with the substrates or products in the active sites; further, α10 interacts with helix III of AcpP. In the *apo* state, AcpP loses its ability to make specific salt bridge contacts observed in the crystal structure (Fig. 7C). However, upon activation to the *holo* state, ACP strengthens its salt bridge contacts with FabB. Therefore, the binding of the substrate to monomer 1 of FabB may trigger the protein conformation/dynamic change of monomer 2, resulting in product release that completes the enzyme turnover of monomer 2.

Free energy calculations were performed to measure the ΔG of AcpP-FabB interaction for wild-type AcpP, D38A AcpP, and R62A and R124A mutants of FabB, using the MM-PBSA method (Fig. 8). A comparison of the ΔΔG values between D38A AcpP, R62A and R124A of FabB, and wt AcpP-FabB shows a decrease in binding affinity for R124A in the *apo* and *holo* states, with the measurements performed in the C8-substrate and C8-product states to be qualitatively decreased or equivalent. The results support that the salt bridge interactions between AcpP and FabB might be important during complex formation, but is less important once the substrate is released *via* the switchblade mechanism into the active site. In sum, the case study provides an example of how MD simulations can be used to model protein dynamics during fatty acid biosynthesis. Finally, we showed that MD-guided mutations change the product lipid profile, which paves a foundation for biofuel development (Milligan et al., 2019, in revision).

2.2 Polyketide synthase

Nature has evolved FAS into PKS to biosynthesize complex natural products, utilizing extra tailoring steps for chemical diversity (Sattely, Fischbach, & Walsh, 2008). PKSs can be categorized into at least three types. Type I PKSs contain “modules,” a collection of enzyme domains that are covalently linked together. Type I PKSs exist as multi-module complexes, or as one single iterative module (Staunton & Weissman, 2001b). Type II PKSs, like their FAS counterparts, are stand-alone enzymes (Case Studies 4–6, Sections 2.2.1–2.2.3). Type III PKSs are also stand-alone enzymes, but one domain performs chain elongation, cyclization and termination (Case Study 7, Section 2.2.4).

2.2.1 Case study 4: Emodin inhibitor study utilizing MD—

Computational techniques have been used extensively in the study of type II PKS biosynthesis in which the proclivity toward spontaneous self-cyclization of the poly-β-ketone intermediate has

hindered mechanistic studies of these systems. In many type II PKSs, the first tailoring step of the elongated linear poly-β-ketone intermediate is the placement and cyclization of the first ring, which is guided by a highly stereo- and regiospecific ketoreduction by a C-9 ketoreductase. In work by Korman et al., computational techniques were used to help explain two puzzling observations about the substrate specificity of the actinorhodin ketoreductase, ActKR (Korman, Tan, Wong, Luo, & Tsai, 2008).

In the search for surrogate substrates suitable for mechanistic and kinetic studies, we found that 2-decalone, which had the same ketone position relative to the first ring of the native substrate, was a much poorer substrate than *trans*-1-decalone, which had a ketone that is offset by one carbon along the first ring. To help shed light on this surprising finding, docking studies were conducted using ICM-Pro (Abagyan, Totrov, & Kuznetsov, 1994b). The docking studies revealed that the stereoisomers of *trans*-1-decalone consistently docked in the same productive orientation, while the 2-decalone stereoisomers could bind in multiple different orientations with significantly weaker binding energies on average when compared with that of *trans*-1-decalone.

After determining the general catalytic mechanism of ActKR using inhibition kinetics with the plant polyketide emodin, Korman et al. co crystallized ActKR with emodin and found that the normally planar emodin was bent ~63° away from planarity within the active site. In order to evaluate the significance of this finding, the geometries of both the bent and linear forms were optimized with Gaussian 03 B3L YP using the 6–311 + G(d,p) basis set, and we found that each conformation persisted through the optimization (Foresman & Frisch, 1996). Next, using Langevin dynamics simulations, we found that the energy difference between the bent and planar forms of emodin was about the same, both when free in solvent and when bound in the ActKR active-site, indicating that the pocket environment selected the bent form without applying steric strain (Lu & Luo, 2003). This study demonstrates the utility of computer simulations in aiding the interpretation of experimental results.

2.2.2 Case study 5: MD-guided homology modeling in a type II PKS—

We solved the crystal structure of anthranilate:CoA ligase (AuaEII), which is responsible for the generation of an uncommon starter unit, anthraniloyl-CoA, for the biosynthesis of aurachins (Fig. 9) (Jackson et al., 2016). AuaE is a homolog of AuaEII and is responsible for transferring anthraniloyl-CoA onto ACP. The basis of regio-specificity between these two highly homologous ligases remained unknown.

We used the crystal structure of AuaEII to build homology models of AuaE in two distinct conformations. Two separate threading templates were used to generate the initial models based on benzoate coenzyme A ligase, which has more than one conformation deposited in the PDB (PDB ID: 4EAT and 2V7B) (Bains & Boulanger, 2007; Thornburg, Wortas-Strom, Nosrati, Geiger, & Walker, 2015). Utilizing a protocol developed by Nurisso et al, short 100 ns simulations were performed to validate the stability of these two distinct conformational poses, and to further refine sidechain geometry (Fig. 9) (Nurisso, Daina, &Walker, 2012). The final 20 ns of each simulation was used to generate average structures using CPPTRAJ, and these were minimized using SANDER (Fig. 9) (Case et al., 2018; Roe & Cheatham, 2013, 2018). The MD-refined models provided a molecular basis for a hinge movement in

AuaE that is absent in AuaEII. This study showed the effectiveness of obtaining additional information on homologous enzymes through MD-guided homology modeling.

2.2.3 Case study 6: The oxetane isosteric probes—Although many *apo* crystal and NMR structures of PKS domains are available, it is difficult to elucidate how PKSs orient and stabilize the growing chain. Specifically, for non-reducing Type I and Type II PKSs, the megasynthases need to stabilize and orient highly reactive poly-P-ketone intermediates while avoiding spontaneous, non-specific cyclization. DpsC from *Streptomyces peucetius* serves both as acyl transferase (AT) and initial ketosynthase (KS) during doxorubicin biosynthesis (Fig. 10A) (Jackson et al., 2018). DpsC uniquely loads propionyl-CoA as the starter unit (as opposed to acetyl-CoA). The molecular basis of starter unit specificity as catalyzed by DpsC was not clear.

We recently designed and synthesized oxetane-polyketide mimics using oxetanes as carbonyl isosteres (Fig. 10B–C) (Ellis et al., 2018; Tsai, 2018). Oxetanes have been shown to mimic both the geometry and dipole properties of carbonyl groups, while sterically preventing intramolecular aldol cyclization at the same time (Bull, Croft, Davis, Doran, & Morgan, 2016; Ellis et al., 2018; Luger & Buschmann, 1984). As a proof of concept, we synthesized an oxetane-based mimic and cocrystallized it with DpsC. The cocrystal structure showed that the oxetane group did indeed orient the mimic in a productive binding mode, posed for the loading of starter unit.

To evaluate the geometric and electronic properties of oxetane mimics, we compared the $\Delta G_{binding}$ of DpsC bound with the oxetane mimic probe *versus* the natural malonate substrate using MD simulations. These simulations were performed in triplicate, and were used for root-mean square fluctuation calculations, free energy calculations, and principal component analysis (PCA). The free energy calculations show that the native malonate-based substrate and the oxetane-based substrate had similar $\Delta G_{binding}$ with no statistically significant differences in their calculated binding affinities. PCA was used to evaluate sampling of the receptor with either substrate present, showing sufficient overlap (Ellis et al., 2018). The above result is promising and the syntheses of higher-order poly-β-ketone mimics (tetra- to dodeca-ketides) are currently underway, which will greatly aid in the visualization of PKSs in action (Tsai, 2018). These oxetane mimic probes can be further loaded to ACP chemoenzymatically, leading to co-crystallization of ACP-probe-partner complexes. These cocrystal structures, coupled with MD simulations, will shed light on the protein-protein interactions and protein dynamics at a molecular level never achieved before.

2.2.4 Case study 7: RpBAS pH MD simulations—The model enzyme for Type III PKSs is chalcone synthase (CHS) from alfalfa *(Medicago sativa)* called MsCHS2, which is responsible for the generation of a tetraketide-derived naringenin chalcone, a precursor involved in the biosynthesis of flavonoid phytoalexins and anthocyanin pigments (Austin & Noel, 2003). Another type III PKS, benzalacetone synthase (RpBAS), biosynthesizes a diketide-derived product, though it shares a 70% sequence similarity with MsCHS2. RpBAS was originally characterized by Abe et al., who found that it produced diketide-derived *p*-hydroxybenzalacetone by a one-step decarboxylative condensation of 4-coumaroyl-CoA with one malonyl-CoA (Abe, Takahashi, Morita, & Noguchi, 2001). Curiously, RpBAS

biosynthesizes a diketide at pH 8, but a triketide at pH 6 (Fig. 11A). It remains a mystery how pH affects the product outcome of RpBAS.

The protonation states of sidechains are influenced primarily by electrostatic contributions of protein-solvent and intra-protein interactions (Stanton & Houk, 2008). Using constant pH MD simulations, we obtained the ratio between protonated and deprotonated states in the constant pH simulations (Fig. 11B).(Schaub et al, 2019, submitted) The MD simulations predicted that the second active site cysteine, Cys197, had its pK_a shifted upfield, which was inconsistent with a previous hypothesis that Cys 197 was catalytic in the deprotonated form (Austin & Noel, 2003). Furthermore, careful attention was given to Cys197, because another plausible model for diketide and triketide selection at different pHs could involve the deprotonation of the phenol group of the bound substrate. The deprotonation may be facilitated by Cys197, although this is doubtful, because our simulations showed that Cys197 remained protonated between pH 6–8. Significantly, the expected pK_a values of two solvent-exposed residues, His257 and His266, were shifted upfield, with pK_as calculated at 6.81, and 7.29, respectively. At pH 6, His257 is 70% protonated and charged, and at pH 9 it is almost 100% in the neutral deprotonated form. Similarly, at pH 6, His266 is 90% protonated and charged, and at pH 9 it is almost 100% in the neutral deprotonated form. Thus, the pH dependence of RpBAS may reflect the protonation states of His257 and His266 (Fig. 11D–F). RMSF calculations were performed for the final 50 ns of each constant pH simulation and the simulations at acidic pHs had greater fluctuations compared to the simulations at basic pHs (Fig. 11C); the fluctuations occur at residues 240–270, the flexible beta-hairpin region that contains both His257 and His266. Based on the MD predictions, we performed alanine and phenylalanine mutations of His257 and His266, and showed that mutations indeed changed the pH dependence of product profile. To the best of our knowledge, this study marked the first MD simulation study of pH dependence of a PKS. The above analysis explains how pH, or protonation states of residues near the active site, may affect the product outcome of type III PKS.

2.3 Nonribosomal peptide synthetase

2.3.1 Case study 8: The termination reductase MxaR—

The termination domains in PKS and NPRS release the covalently attached chain as a linear or macro cycle product. The utilization of these termination domains for the diversification of PKs and NRPs is of great interest to synthetic biology. The myxalamid biosynthesis in *Stigmatella aurantiaca* involves six PKS modules and one terminal NRPS module (Fig. 12). The terminal reductase, MxaA, catalyzes a four-electron, non-processive reduction to biosynthesize myxalamids. Understanding the structure and function of MxaA may enable us to engineer this domain for new biofuel development (Barajas et al., 2015).

We solved the MxaA crystal structure, and performed MD simulations on the *apo* and NADPH bound structures to elucidate MxaA-NADPH interactions (Barajas et al., 2015). Then, a truncated substrate was docked into the active site at the C-terminal subdomain, and a short 100ns simulation was performed (Fig. 13). The substrate-bound simulation resulted in a decrease in RMSD in the active site, supporting a convergence of protein conformation near the active site upon substrate binding. The most noticeable flexible region in the

substrate-bound simulation was in the C-terminal helix-tum-helix motif and we hypothesized that this may be the PCP binding site.

The protein-substrate interactions were analyzed over the trajectories and the contacts measured as time occupancy over the simulation. Key residues were identified for mutagenesis studies. Assays for the mutants utilized decanoyl-CoA, which led to the biosynthesis a medium-chain alcohol, decanal, with the potential for biofuel development. The R1339A mutant biosynthesized decanol four times faster than the wild-type. This study showcased how we m ay utilize MD-directed mutagenesis to develop new biofuels.

3. Summary and future perspectives

3.1 Force field optimization and parameterization

In Section 1.4.1, we described the development of an initial phosphopantetheine force field (Schaub et al, 2019, submitted). It is common in force field development to improve force field parameters after it is utilized in several studies. The initial LIPID11 framework parameter set was primarily based on the General Amber Force Field (GAFF), with additional parameters adopted from the Glycam force field (Skjevik et al., 2012). H ere, we purposely changed the names of atom types so that revisions in the future could be accomplished. Dickson et al. would later optimize the original LIPID11 force field by updating the torsion parameters in the hydrocarbon tails using electronic structure calculated at a higher level with the cc-pVQZ basis set (Dickson et al., 2014). Furthermore, additional fitting to physical observables was performed by fitting the force field to reproduce the heat of vaporization for methyl acetate, which performed poorly with the previous LIPID11 and GAFF parameters. This optimized force field was released as LIPID14.

A major improvement of our PPant force field would be to generate an all-inclusive force field, where extender units, starter units and intermediates could be generated routinely. NORINE, a database of non-ribosomal peptide extender units and non-ribosomal peptides, is available online, which contains 534 monomer units and 1187 peptides (Caboche et al., 2008; Flissi et al., 2016; Pupin et al., 2016). It is impractical to generate force field parameters for every *n*-mer, because this would be computationally infeasible. However, one could generate a library of force fields on demand with minimal input and control by the end-user. This could be accomplished through a pipeline encoded in a software package or a user-friendly web server.

3.2 Domain docking engineering

Our goal of conducting MD simulations for megasynthases is to better understand the protein-substrate and protein-protein interactions so that we can engineer the megasynthases more efficiently for the biosynthesis of engineered natural products. To understand how protein-protein interactions control the product outcome, we need computational tools that can identify and predict binding patterns at the protein-protein interface. A protocol to identify sequence motifs, known as “fingerprints,” was described by Keatinge-Clay and used to define regions of the ketoreductase domain that control the stereochemistry of the reduction product (Keatinge-Clay, 2007). Wood et al. had recently used a similar protocol on

acyl carrier proteins. They identified contact patterns in different families of acyl carrier proteins through sequence and cladogram-based analysis in the *trans*-AT systems (Wood & Keatinge-Clay, 2018). Wood's analysis on the co-evolution between ACP and its downstream partner supports that simply engineering a single domain is not enough. This may also explain why previous engineering efforts that randomly shuffle PKS domains often led to inactive megasynthases.

One possibility to systematically explore the protein-protein interactions in the megasynthases would be to perform isothermal calorimetry (ITC) experiments between protein partners. For example, for modular Type I PKS, between the docking domains of upstream and downstream modules, thus generating a database of binding energies of the docking modules. Then, we can perform free energy calculations *in silica* on the same systems, and link the free energy outcome with the ITC results. Such an approach, if setup correctly with a test set and validation set, could provide a force field capable of predicting binding affinity between the docking domains in modular PKSs.

3.3 Long-range interactions and allostery

In the type II FAS, interdomain communication may play a key role in product outcome. In Section 2.1.3, MD simulations of FabB bound to AcpP showed an increased movement in the α6-α7 region of FabB, which is in close proximity to a highly flexible helix (α10) of the same monomer. The α10 also interacts with AcpP, and supports that protein dynamics of FabB may influence its interaction with AcpP. In order to analyze these dynamic regions in more depth, residue interaction network (RIN) analysis could be performed. Network analysis can also be applied to long-range interactions that allosterically modulate the active sites of domains in megasynthases. Specifically, in megasynthases, many domains contain lid regions that control the product outcome. Performing RIN analysis of the hinge movements of lid regions could potentially lead to a blueprint for protein engineers who wish to modify these lid regions.

3.4 MD simulations on large enzyme complexes over long time scales

While MD simulations have been performed on a homology model of hFAS, none have been performed on larger systems in these pathways. The computational power is available to perform all-atom simulations on system as large as the FAS multi-enzyme complexes found in fungi and bacteria (Ciccarelli et al., 2013). Anselmi et al. recently performed a simulation on the fungal FAS using a coarse-grained model, and showed that ACP shuttling is the result of molecular crowding effects (Anselmi, Grininger, Gipson, & Faraldo-Gomez, 2010). All-atom simulations have the potential to explore long-standing questions in the field. Specifically, are the ACPs coupled and able to interact with each other, or do they function independently? This review shows some promising results that have been achieved using computational biology, but much more can be accomplished. In conclusion, computational biology of megasynthases such as FAS, PKS and NRPS could lead to a greater understanding of the complex interactions in these megasynthases, which in tum can offer a blueprint that guides future efforts in protein engineering toward the development of new therapeutics and biofuels.

Abbreviations

ACCase	acyl-CoA carboxylase
ACP	acyl carrier protein
AcpP	*Escherichia coli* type II fatty acid synthase acyl carrier protein
AT	acyltransferase
AuaEII	anthranilate:CoA ligase
CAPRI	critical Assessment of Predicted Interactions
CASP	critical Assessment of Protein Structure Prediction
Co A	coenzyme A
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
ER	enoylreductase
FA	fatty acid
FAS	fatty acid synthase
KS	ketosynthase
MAT	malonyl-acyl transferase
MD	molecular dynamics
MPT	malonyl/palmitoyl transferase
NCI	National Cancer Institute
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
PCP	peptidyl carrier protein
PK	polyketide
PKS	polyketide synthase
PPant	phosphopantetheine
PPTase	phosphopantetheinyl transferase
TE	thioesterase
***trans*-AT**	*trans*-acyltransferase

References

Abagyan R, Totrov M, & Kuznetsov D (1994). Icm—A new method for protein modeling and design —Applications to docking and structure prediction from the distorted native conformation. Journal if Computational Chemistry, 15, 488–506.

Abe I, T akahashi Y, Morita H, & Noguchi H (2001). Benzalacetone synthase. A novel polyketide synthase that plays a crucial role in the biosynthesis of phenylbutanones in Rheum palmatum. European Journal if Biochemistry, 268, 3354–3359.

Adamek M, Spohn M, Stegmann E, & Ziemert N. (2017). Mining bacterial genomes for secondary metabolite gene clusters. Methods in Molecular Biology, 1520, 23–47. [PubMed: 27873244]

Anselmi C, Grininger M, Gipson P, & Faraldo-Gomez JD. (2010). Mechanism of substrate shuttling by the acyl-carrier protein within the fatty acid mega-synthase. Journal of the American Chemical Society, 132, 12357–12364. [PubMed: 20704262]

Aqvist J, Luzhkov VB, & Brandsdal BO(2002). Ligand binding affinities from MD simulations. Accounts if Chemical Research, 35, 358–365.

Austin MB, & Noel JP. (2003). The chalcone synthase superfamily of type Ill polyketide synthases. Natural Product Reports, 20, 79–110. [PubMed: 12636085]

Bains J, & Boulanger MJ (2007). Biochemical and structural characterization of the paralogous benzoate CoA ligases from Burkholderia xenovorans LB400: Defining the entry point into the novel benzoate oxidation (box) pathway.Journal if Molecular Biology, 373, 965–977.

Barajas JF, Phelan RM, Schaub AJ, Kliewer JT, Kelly PJ, Jackson DR, eta!. (2015). Comprehensive structural and biochemical analysis of the terminal Myxalamid reductase domain for the engineered production of primary alcohols. Chemistry & Biology, 22, 1018–1029. [PubMed: 26235055]

Blin K, Wolf T, Chevrette MG, Lu XW, Schwalen CJ, Kautsar SA, et al. (2017). antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research, 45, W36–W41. [PubMed: 28460038]

Boddy CN (2014). Bioinformatics tools for genome mining of polyketide and nonribosomal peptides. journal if Industrial Microbiology & Biotechnology, 41, 443–450. [PubMed: 24174214]

Born M, & Oppenheimer R (1927). Quantum theory of molecules. Annals if Physics-Berlin, 84, 0457–0484.

Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. (2009). CHARMM: The biomolecular simulation program. Journal if Computational Chemistry, 30, 1545–1614.

Bull j. A., Croft RA, Davis 0A, Doran R, & Morgan KF. (2016). Oxetanes: Recent advances in synthesis, Reactivity, and Medicinal Chemistry. Chemical Reviews, 116, 12150–12233. [PubMed: 27631342]

Byers DM, & Gong H (2007). Acyl carrier protein: Structure-function relationships in a conserved multifunctional protein family. Biochemistry and Cell Biology, 85, 649–662. [PubMed: 18059524]

Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, & Kucherov G (2008). NORINE: A database of nonribosomal peptides. Nucleic Acids Research, 36, D326–D331. [PubMed: 17913739]

Case DA, Ben-Shalom IY, Brozell SR, Derutti DS, Cheatham TE, Cruzeiro VWD, et al. (2018). AMBER 2018. San Francisco: University of California.

Chan DI, & Vogel HJ (2010). Current understanding of fatty acid biosynthesis and the acyl carrier protein. The Biochemical journal, 430, 1–19. [PubMed: 20662770]

Chernler JA, Tripathi A, Hansen DA, O'Neil-Johnson M, Williams RB, Starks C, et al. (2015). Evolution of efficient modular polyketide synthases by homologous recombination. Journal if the American Chemical Society, 137, 10603–10609.

Chen J, Swamidass SJ, Dou Y, Bruand J, & Baldi P (2005). ChemDB: A public database of small molecules and related chemoinformatics resources. Bioinformatics, 21, 4133–4139. [PubMed: 16174682]

Ciccarelli L, Connell SR, Enderle M, Mills DJ, Vonck J, & Grininger M (2013). Structure and conformational variability of the mycobacterium tuberculosis fatty acid synthase multienzyme complex. Structure, 21, 1251–1257. [PubMed: 23746808]

Comeau SR, Gatchell DW, Vajda S, &Camacho CJ (2004). ClusPro: An automated docking and discrimination method for the prediction of protein complexes. Bioinformatics, 20, 45–50. [PubMed: 14693807]

Corbeil CR, Williams CI, & Labute P. (2012). Variability in docking success rates due to dataset preparation. Journal if Computer-Aided Molecular Design, 26, 775–786. [PubMed: 22566074]

Dewick PM (2009). Medicinal natural products: A biosynthetic approach (3rd ed.). Chichester, West Sussex, United Kingdom: Wiley, A John Wiley and Sons, Ltd., Publication.

Dickson CJ, Madej BD, Skjevik AA, Betz RM., Teigen K., Gould IR, et al. (2014). Lipid14: The Amber lipid force field. Journal of Chemical Theory and Computation, 10, 865–879. [PubMed: 24803855]

Dominguez C, Boelens R, & Bonvin AMJJ (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal if the American Chemical Society, 125, 1731–1737.

Dror R0, Dirks RM, Grossman P, Xu HF, & Shaw DE. (2012). Biomolecular simulation: A computational microscope for molecular biology. Annual Review if Biophysics, 41, 429–452.

Du L, & Lou L (2010). PKS and NRPS release mechanisms. Natural Product Reports, 27, 255–278. [PubMed: 20111804]

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. journal if Computational Chemistry, 24, 1999–2012.

Ellis BD, Milligan JC, White AR., Duong V, Altman PX, Mohammed LY, et al. (2018). An Oxetane-based polyketide surrogate to probe substrate binding in a polyketide synthase. journal if the American Chemical Society, 140, 4961–4964.

Evans R, Jumper J, Kirkpatrick J, Sifre L, Green TFG, Qin C, et al. (2018). De novo structure prediction with deep-learning based scoring. In: Thirteenth Critical Assessment if Techniques for Protein Structure Prediction (Abstracts) 1–4.

Ferrer JL, Austin MB, Stewart C Jr., & Noel JP. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiology and Biochemistry, 46, 356–370. [PubMed: 18272377]

Fersht AR (2013). Profile of Martin Karplus, Michael Levitt, and Arieh Warshel, 2013 nobel laureates in chemistry. Proceedings if the National Academy if Sciences if the United States if America, 110, 19656–19657.

Fischbach M, & Voigt CA (2010). Prokaryotic gene clusters: A rich toolbox for synthetic biology. Biotechnology Journal, 5, 1277–1296. [PubMed: 21154668]

Flissi A, Dufresne Y, Michalik j., Tonon L,Janot S, Noe L, et al. (2016). Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. Nucleic Acids Research, 44, D1113–D1118. [PubMed: 26527733]

Foresman JB, & Frisch A (1996). exploring chemistry with electronic structure methods In a guide to using Gaussian. (2nd ed.). Wallingford, CT: Gaussian, Inc.

Gaitatzis N, Silakowski B, Kunze B, Nordsiek G, Blocker H, Hotle G, et al. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. The Journal of Biological Chemistry, 277, 13082–13090. [PubMed: 11809757]

Gajewski J, Buelens F, Serdjukow S, Janssen M, Cortina N, Grubmuller H, et al. (2017). Engineering fatty acid synthases for directed polyketide production. Nature Chemical Biology, 13, 363–365. [PubMed: 28218912]

Gershenzon J, & Dudareva N (2007). The function of terpene natural products in the natural world. Nature Chemical Biology, 3, 408–414. [PubMed: 17576428]

Guider TA, & Moore BS (2009). Chasing the treasures of the sea-Bacterial marine natural products. Current Opinion in Microbiology, 12, 252–260. [PubMed: 19481972]

Helfrich EJN, Reiter S, & Piel J (2014). Recent advances in genome-based polyketide discovery. Current Opinion in Biotechnology, 29, 107. [PubMed: 24762576]

Homeyer N, & Gohlke H. (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Molecular Informatics, 31, 114–122. [PubMed: 27476956]

Homeyer N, & Gohlke. (2013). FEW: A workflow tool for free energy calculations of ligand binding. Journal if Computational Chemistry, 34, 965–973.

Homeyer N, Horn AH, Lanig H, & Sticht H (2006). AMBER force-field parameters for phosphorylated amino acids in different protonation states: Phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. Journal if Molecular Modeling, 12, 281–289.

Hur GH, Vickery CR, & Burkart MD (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Natural Product Reports, 29, 1074–1098. [PubMed: 22802156]

Jackson DR, Shakya G, Patel AB, Mohammed LY, Vasilakis K, Wattana-Arnorn P., et al. (2018). Structural and functional studies of the Daunorubicin priming Ketosynthase DpsC. ACS Chemical Biology, 13, 141–151. [PubMed: 29161022]

Jackson DR, Tu SS, Nguyen M, Barajas JF, Schaub AJ, Krug D, et al. (2016). Structural insights into anthranilate priming during type II polyketide biosynthesis. ACS Chemical Biology, 11, 95–103. [PubMed: 26473393]

Janin J (2002). Welcome to CAPRI: A critical assessment of PRedicted interactions. Proteins: Structure, Function, and Genetics, 47, 257.

Janin J, Henrick K, Moult J, TenEyck L, Sternberg MJE, Vajda S, et al. (2003). CAPRI: A critical assessment of PRedicted interactions. Proteins: Structure, Function, and Bioinformatics, 52, 2–9.

Jones G, Willett P, Glen RC, Leach AR, & Taylor R. (1997). Development and validation of a genetic algorithm for flexible ligand docking. Abstracts of Papers if the American Chemical Society, 214, 154 Comp.

Karplus M, & McCammon JA (2002). Molecular dynamics simulations of biomolecules. Nature Structural Biology, 9, 646–652. [PubMed: 12198485]

Keatinge-Clay AT (2007). A tylosin ketoreductase reveals how chirality is detenmined in polyketides. *Chemistry* & Biology, 14, 898–908.

Keller NP, Turner G, & Bennett JW(2005). Fungal secondary metabolism—From biochemistry to genomics. Nature Reviews. Microbiology, 3, 937–947. [PubMed: 16322742]

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genetics and Biology, 47, 736–741. [PubMed: 20554054]

Khosla C, Herschlag D, Cane DE, & Walsh CT (2014). Assembly line polyketide synthases: Mechanistic insights and unsolved problems. Biochemistry, 53, 2875–2883. [PubMed: 24779441]

Khoury GA, Smadbeck J, Tamamis P, Vandris AC, Kieslich CA, & Floudas CA. (2014). Forcefield_NCAA: Ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. ACS Synthetic Biology, 3, 855–869. [PubMed: 24932669]

Khoury GA, Thompson JP, Smadbeck J, Kieslich CA, & Floudas CA (2013). Forcefield_PTM: Charge and AMBER Forcefield parameters for frequently occurring Post-translational modifications. Journal of Chemical Theory and Computation, 9, 5653–5674. [PubMed: 24489522]

Kinghorn AD (2002). The role of pharmacognosy in modern medicine. Expert Opinion on Pharmacotherapy, 3, 77–79. [PubMed: 11829721]

Kochanowska-Karamyan AJ, & Hamann MT (2010). Marine indole alkaloids: Potential new drug leads for the control of depression and anxiety. Chemical Reviews, 110, 4489–4497. [PubMed: 20380420]

Korman TP, Tan YH, Wong j., Luo R, & Tsai SC. (2008). Inhibition kinetics and emodin cocrystal structure of a type II polyketide ketoreductase. Biochemistry, 47, 1837–1847. [PubMed: 18205400]

Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, et al. (2013). H ow good is automated protein docking? Proteins: Structure, Function, and Bioinformatics, 81, 2159–2166.

Kozakov D, Brenke R, Comeau SR, & Vajda S. (2006). PIPER: An FFT -based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics, 65, 392–406.

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. (2017). The ClusPro web server for protein-protein docking. Nature Protocols, 12, 255–278. [PubMed: 28079879]

Kumar S, Huang C, Zheng G, Bohm E, Bhatele A, Phillips j. C., et al. (2008). Scalable molecular dynamics with NAMD on the IBM blue gene/L system. IBM Journal of Research and Development, 52, 177–188.

Leach AR (2001). Molecular modelling: Principles and applications (2nd ed.). Pearson Education Limited.

Lee MC, & Duan Y (2004). Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins, 55, 620–634. [PubMed: 15103626]

Lin TW, Melgar MM, Kurth D, Swamidass SJ, Purdon j., Tseng T, et al. (2006). Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America, 103, 3072–3077. [PubMed: 16492739]

Lu Q, & Luo R (2003). A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. The Journal of Chemical Physics, 119, 11035–11047.

Luger P, & Buschmann J (1984). Oxetane-The 1st X-ray-analysis of a nonsubstituted 4-membered ring. Journal of the American Chemical Society, 106, 7118–7121.

Maier T, Leibundgut M, & Ban N. (2008). The crystal structure of a mammalian fatty acid synthase. Science, 321, 1315–1322. [PubMed: 18772430]

Markwick PR, & McCammon JA. (2011). Studying functional dynamics in biomolecules using accelerated molecular dynamics. Physical Chemistry Chemical Physics, 13, 20053–20065. [PubMed: 22015376]

Matsuda Y, & Abe I (2016). Biosynthesis of fungal meroterpenoids. Natural Product Reports, 33, 26–53. [PubMed: 26497360]

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. (2011). antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research, 39, W339–W346. [PubMed: 21672958]

Michel J, & Essex JW (2010). Prediction of protein-ligand binding affinity by free energy simulations: Assumptions, pitfalls and expectations. Journal of Computer-Aided Molecular Design, 24, 639–658. [PubMed: 20509041]

Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, & Roitberg AE (2012). MMPBSA.Py: An efficient program for end-state free energy calculations. Journa1 of Chemical Theory and Computation, 8, 3314–3321.

Moore BS, Hertweck C, Hopke JN, Izumikawa M, Kalaitzis JA, Nilsen G, et al. (2002). Plant-like biosynthetic pathways in bacteria: From benzoic acid to chalcone. Journal of Natural Products, 65, 1956–1962. [PubMed: 12502351]

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30, 2785–2791. [PubMed: 19399780]

Nguyen C., Haushalter RW, Lee DJ, Markwick PR, Bruegger j., Caldara-Festin G, et al. (2014). Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. Nature, 505, 427–431. [PubMed: 24362570]

Nurisso A, Daina A, & Walker RC (2012). A practical introduction to molecular dynamics simulations: Applications to homology modeling. Methods in Molecular Biology, 857, 137–173. [PubMed: 22323220]

O'Connor SE, & Maresh JJ (2006). Chemistry and biology of monoterpene indole alkaloid biosynthesis. Natural Product Reports, 23, 532–547. [PubMed: 16874388]

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. (2005). Scalable molecular dynamics with NAMD. Journal of Computational Chemistry, 26, 1781–1802. [PubMed: 16222654]

Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29, 845–854. [PubMed: 23407358]

Pupin M, Esmaeel Q, Flissi A, Dufresne Y, Jacques P, & Leclere V (2016). Norine: A powerful resource for novel nonribosomal peptide discovery. Synthetic and Systems Biology, 1, 89–94.

Roe DR, & Cheatham TE 3rd. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. Journal of Chemical Theory and Computation, 9, 3084–3095. [PubMed: 26583988]

Roe DR, & Cheatham TE 3rd (2018). Parallelization of CPPTRAJ enables large scale analysis of molecular dynamics trajectory data. Journal of Computational Chemistry, 39, 2110–2117. [PubMed: 30368859]

Roy A, Kucukural A, & Zhang Y (2010). I-TASSER: A unified platform for automated protein structure and function prediction. Nature Protocols, 5, 725–738. [PubMed: 20360767]

Sattely ES, Fischbach MA, & Walsh CT (2008). Total biosynthesis: In vitro reconstitution of polyketide and nomibosomal peptide pathways. Natural Product Reports, 25, 757–793. [PubMed: 18663394]

Scarpino A, Ferenczy GG, & Keseru GM (2018). Comparative evaluation of covalent docking *tools*. Journal of Chemical Information and Modeling, 58(7), 1441–1458. 10.1021/acs.jcim.8b00228. [PubMed: 29890081]

Schneider B, Kabelac M, & Hobza P. (1996). Geometry of the phosphate group and its interactions with metal cations in crystals and ab initio calculations. Journal of the American Chemical Society, 118, 12207–12217.

Shakya G, Rivera H Jr., Lee DJ,Jaremko MJ, LaClair JJ, Fox DT, et al. (2014). Modeling linear and cyclic PKS intermediates through atom replacement. Journal of the American Chemical Society, 136, 16792–16799. [PubMed: 25406716]

Simonson T, Carlsson J, & Case DA (2004). Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. Journal of the American Chemical Society, 126, 4167–4180. [PubMed: 15053606]

Skinnider MA, Dejong CA, Rees PN,Johnston CW, Li HX, W ebster ALH, et al. (2015). Genomes to natural products PRediction informatics for secondary metabolomes [PRISM]. Nucleic Acids Research, 43, 9645–9662. [PubMed: 26442528]

Skinnider MA,Johnston CW, Edgar RE, Dejong CA, Merwin NJ, Rees PN., et al. (2016). Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. Proceedings of the National Academy of Sciences of the United States of America, 113, E6343–E6351. [PubMed: 27698135]

Skjevik AA, Madej BD, Walker RC., & Teigen K. (2012). LIPID11: A modular framework for LIPID simulations using amber. The Journal of Physical Chemistry. B, 116, 11124–11136. [PubMed: 22916730]

Sliwoski G, Kothiwale S, Meiler J, & Lowe EW (2014). Computational methods in drug discovery. Pharmacological Reviews, 66, 334–395. [PubMed: 24381236]

Smith S, & Tsai SC (2007). The type I fatty acid and polyketide synthases: A tale of two megasynthases. Natural Product Reports, 24, 1041–1072. [PubMed: 17898897]

Song YF, DiMaio F, Wang RYR, Kim D, Miles C, Brunette TJ., et al. (2013). High-resolution comparative modeling with RosettaCM. Structure, 21, 1735–1742. [PubMed: 24035711]

Song D, Luo R, & Chen HF (2017). The IDP-specific force field ff14IDPSFF improves the conformer sampling of intrinsically disordered proteins. Journa1 of Chemical Information and Modeling, 57, 1166–1178.

Song D, Wang W, Ye W, Ji D, Luo R, & Chen HF. (2017). ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chemical Biology* & Drug Design, 89, 5–15.

Stanton CL, & Houk KN (2008). Benchmarking pK(a) prediction methods for residues in proteins. Journal of Chemical Theory and Computation, 4, 951–966. [PubMed: 26621236]

Staunton J, & Weissman KJ (2001). Polyketide biosynthesis: A millennium review. Natural Product Reports, 18, 380–416. [PubMed: 11548049]

Swamidass SJ, Azencott CA, Lin TW, Gramajo H, Tsai SC, & Baldi P. (2009). Influence relevance voting: An accurate and interpretable virtual high throughput screening method. Journal of Chemical Information and Modeling, 49, 756–766. [PubMed: 19391629]

Takayama K, W ang C, & Besra GS (2005). Pathway to synthesis and processing of mycolic acids in mycobacterium tuberculosis. Clinical Microbiology Reviews, 18, 81–101. [PubMed: 15653820]

Thornburg CK, Wortas-Strom S, Nosrati M, Geiger j. H., & Walker KD (2015). Kinetically and Crystallographically guided mutations of a benzoate CoA ligase (BadA) elucidate mechanism and expand substrate Permissivity. Biochemistry, 54, 6230–6242. [PubMed: 26378464]

Tsai S (2018). The structural enzymology of iterative aromatic polyketide synthases: A critical comparison with fatty acid synthases. Annual Review of Biochemistry, 87, 503–531.

Van Voorhis WC, Hooftvan Huijsduijnen R, &Wells TN. (2015). Profile of William C. Campbell, Satoshi Omura, and Youyou Tu, 2015 Nobel laureates in physiology or medicine. Proceedings of the National Academy of Sciences of the United States of America, 112, 15773–15776. [PubMed: 26699511]

van Zundert GCP, Rodrigues JPGLM., Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. Journal of Molecular Biology, 428, 720–725. [PubMed: 26410586]

Wang JM, Hou TJ, & Xu XJ (2006). Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. Current Computer-Aided Drug Design, 2, 287–306.

Wang LP, McKiernan KA, Gomes J, Beauchamp KA, H ead-Gordon T, Rice JE, et al. (2017). Building a more predictive protein force field: A systematic and reproducible route to AMBER- FB15. The journal of Physical Chemistry. B, 121, 4023–4039. [PubMed: 28306259]

White SW, Zheng J, & Zhang YM (2005). Rock, the structural biology of type II fatty acid biosynthesis. Annual Review of Biochemistry, 74, 791–831.

Wodak SJ, & Janin J (2017). Modeling protein assemblies: Critical assessment of predicted interactions (CAPRI) 15 years hence. 6TH CAPRI evaluation meeting April 17–19 Tel-Aviv, Israel. Proteins: Structure, Function, and Bioinformatics, 85, 357–358.

Wood DAV, & Keatinge-Clay AT (2018). The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. Proteins: Structure, Function, and Bioinformatics, 86, 664–675.

Xu D, & Zhang Y (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Structure, Function, and* Bioinformatics, 80, 1715–1735.

Yang J, Yan R, Roy A, Xu D, Poisson J, & Zhang Y (2015). The I-TASSER suite: Protein structure and function prediction. Nature Methods, 12, 7–8. [PubMed: 25549265]

Zhang YM, Marrakchi H, White SW, & Rock C0 (2003). The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. journal of Lipid Research, 44, 1–10. [PubMed: 12518017]

Zhu K, Borrelli KW, Greenwood JR, Day T, Abel R., Farid RS, et al. (2014). Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. journal of Chemical Information and Modeling, 54, 1932–1940. [PubMed: 24916536]

Zhu L, & Cronan JE (2015). The conserved modular elements of the acyl carrier proteins of lipid synthesis are only partially interchangeable. The Journal of Biological Chemistry, 290, 13791–13799. [PubMed: 25861991]

Ziemert N, Podell S, Penn K, Badger JH, Allen E, &Jensen PR. (2012). The natural product domain seeker NaPDoS: A phylogeny based Bioinforrnatic tool to classify secondary metabolite gene diversity. PLoS One, 7, 1–9.

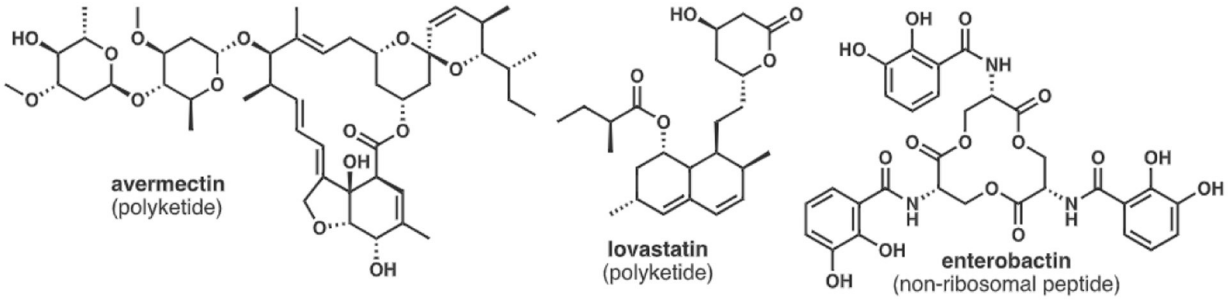

Fig. 1. Examples of natural products biosynthesized by polyketide synthases and nonribosomal peptide synthetases.

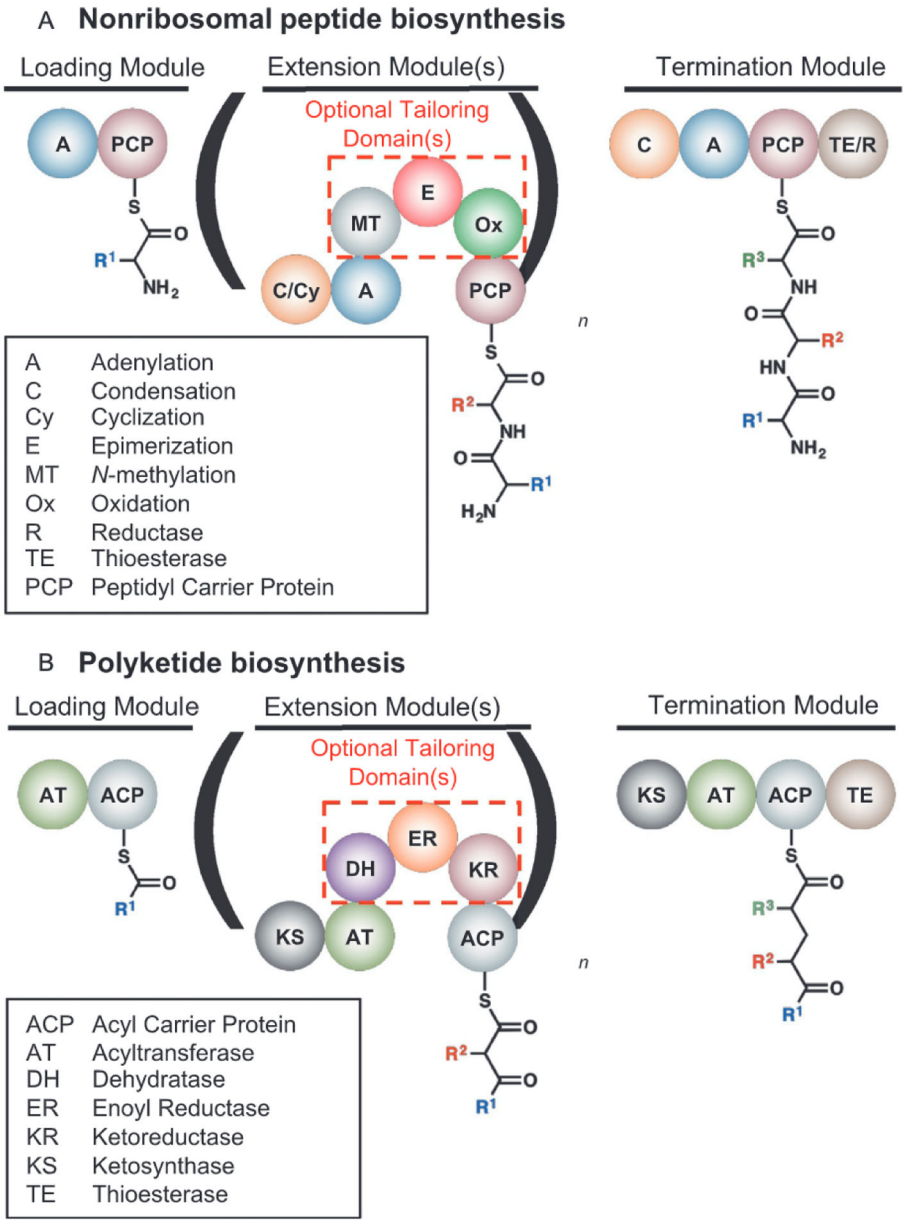

Fig. 2. Examples of assembly line biosynthesis of (A) non-ribosomal peptides in Type A NRPS systems and (B) polyketides in Type I modular PKS systems.

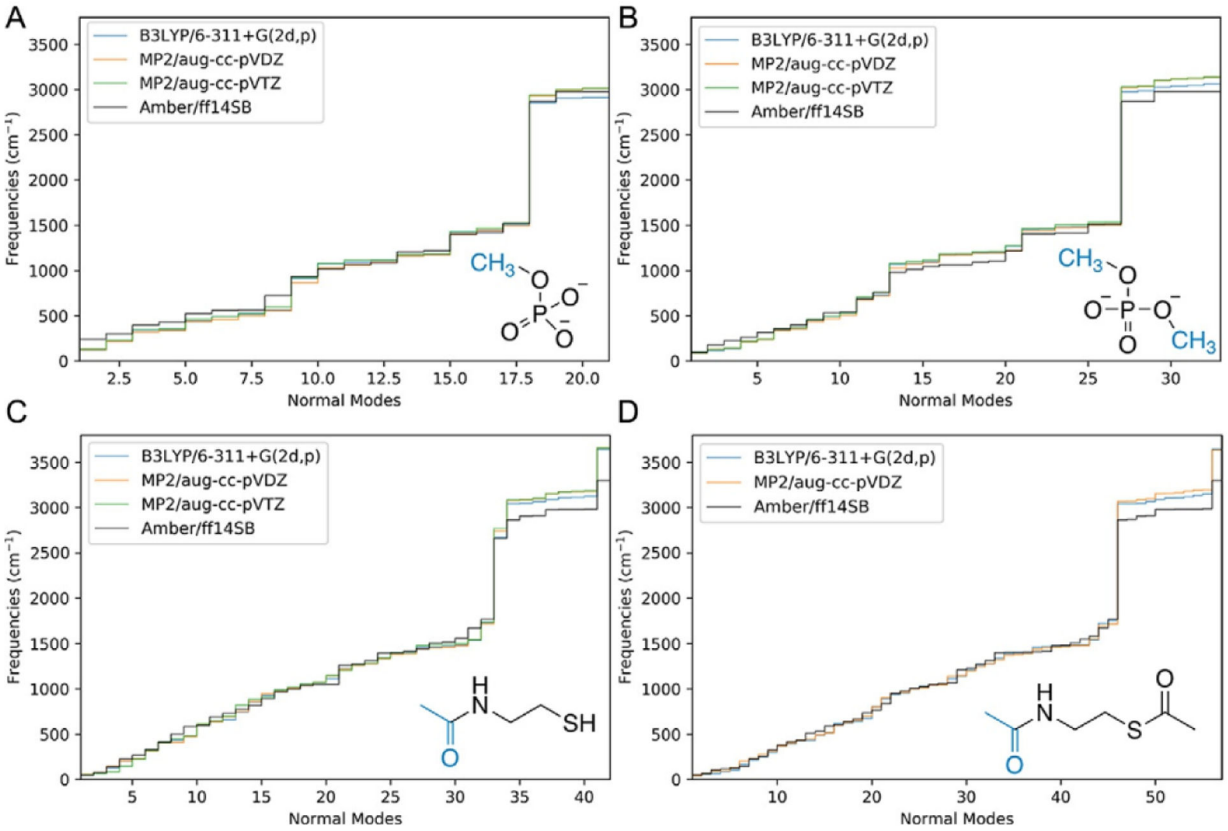

Fig. 3. Normal mode frequency calculations were performed using B3LYP/6–31 1 +G(2d,p), MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels of theory. A low mode search was performed using the Amber ffl 4SB force field with additional phosphopantetheine force field (PFF) parameters.

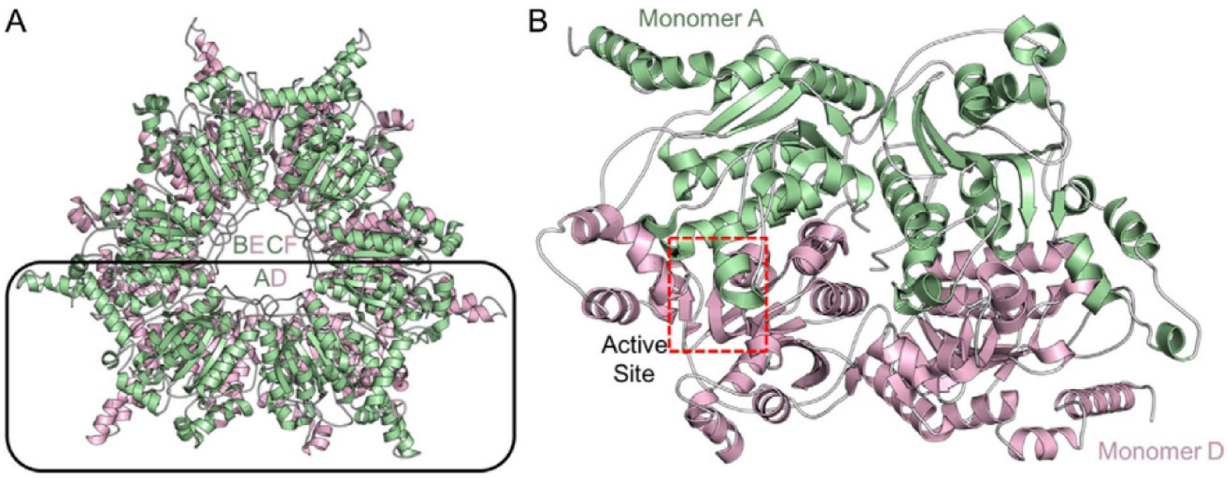

Fig. 4. The ACCDS crystal structure. (A) In ACCDS, six identical monomers (A–F) assemble to form a stack of two trimers (A–C and E–F). (B) The active site is highlighted as a red dashed box and residues at the interface of two monomeric units. Docking simulations were performed in the active site.

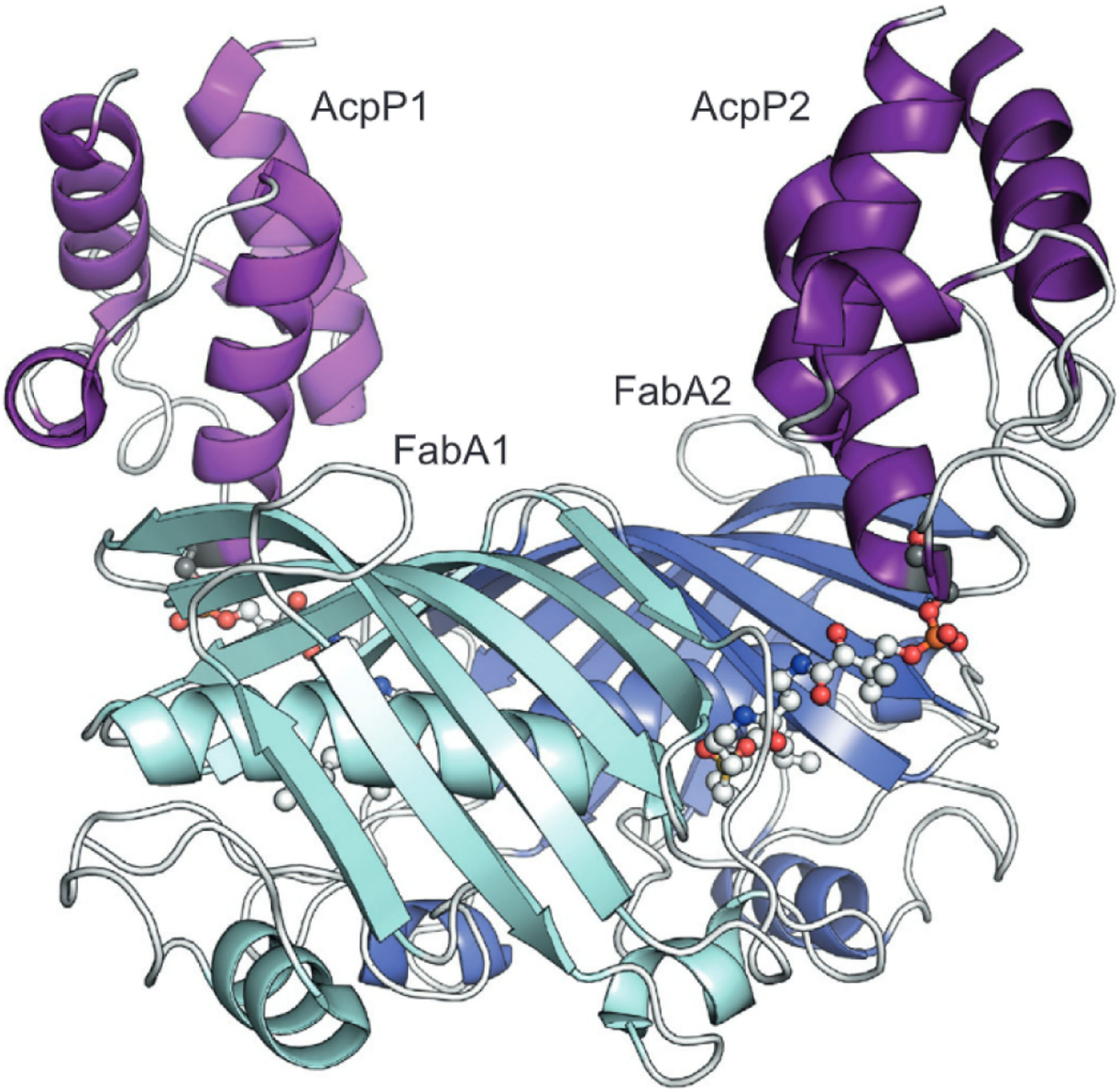

Fig. 5.

The 1.9 A crystal structure of cross-linked AcpP-FabA complex (PDB ID: 4KEH). The contact area between AcpP and FabAS is relatively small, with conserved negatively charged cleft on AcpP interacting with a positive arginine-rich patch on FabA. The cross-linker is shown in ball-and-stick.

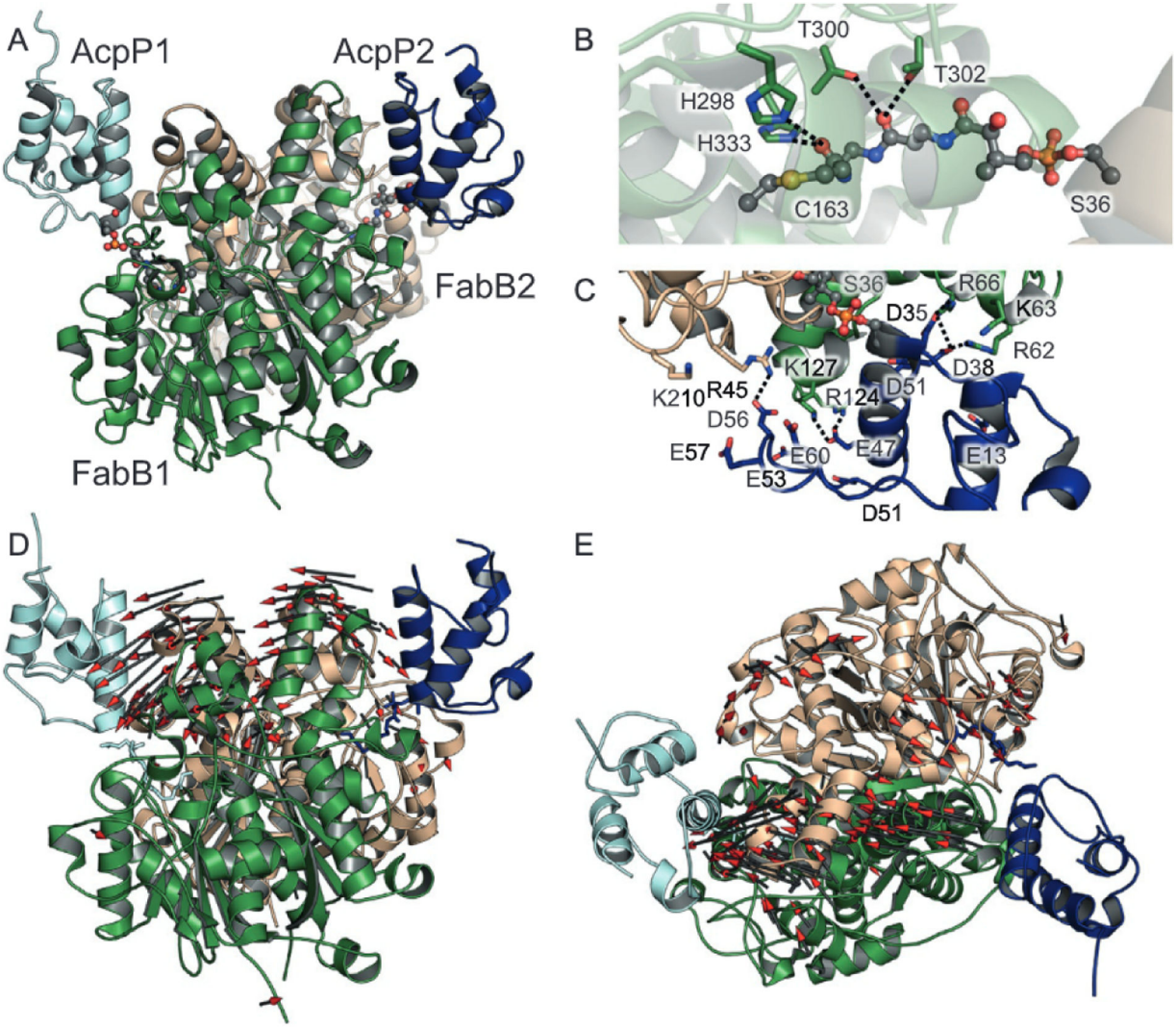

Fig. 6.

(A) The crystal structure of the cross-linked AcpP-FabB structure at 2.4A (PDB ID: 4KOF). (B) FabB active site with the cross-linker. (C) AcpP-FabB interface. (D) Front view of porcupine plot showing PCA mode 1 vectors from starting to final structures superimposed. (E) Side view of the same porcupine plot showing PCA mode 1 vectors.

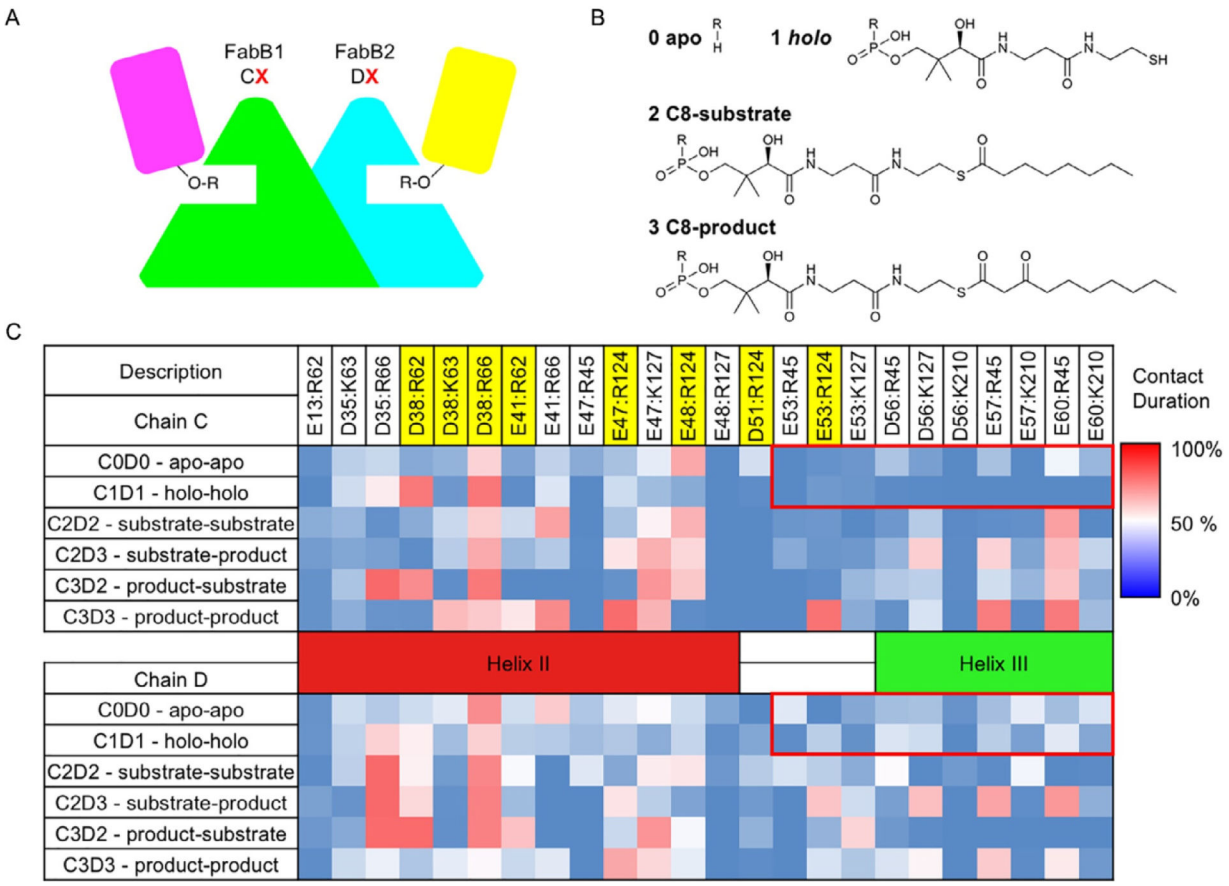

Fig. 7.

(A) A cartoon representation of the FabB-AcpP complex with FabB, (shown in green), $FabB_2$ (shown in cyan), $AcpP_1$ (shown in magenta) and $AcpP_2$ (shown in yellow). (B) The AcpPs were modeled as **1** *apo* (no PPant). **2** *ho/o* (empty PPant), **3** C8 acyl-substrate loaded, or **4** C1 0 acyl-product loaded (C) Protein-protein interactions were analyzed by comparisons between ionic interactions in each trajectory at the protein-protein interface of the FabB-AcpP dimer. Contacts were measured over each 111s trajectory, with a contact being defined as a salt bridge length between two residues < 5 Å. Total contact duration during each simulation is represented in the heat map above with stronger contacts shown in red, and weaker contacts shown in blue. In the apo state, AcpP loses its ability to make specific salt bridge contacts seen in the crystal structure of the FabB-AcpP complex. Interestingly, upon actual loading of PPant with a C8 FabB substrate or product, AcpP helix Ill makes specific strong interactions with FabB. Highlighted in yellow are residues: D38A, R62A, and R124A, which were subjected to further computational inquiry.

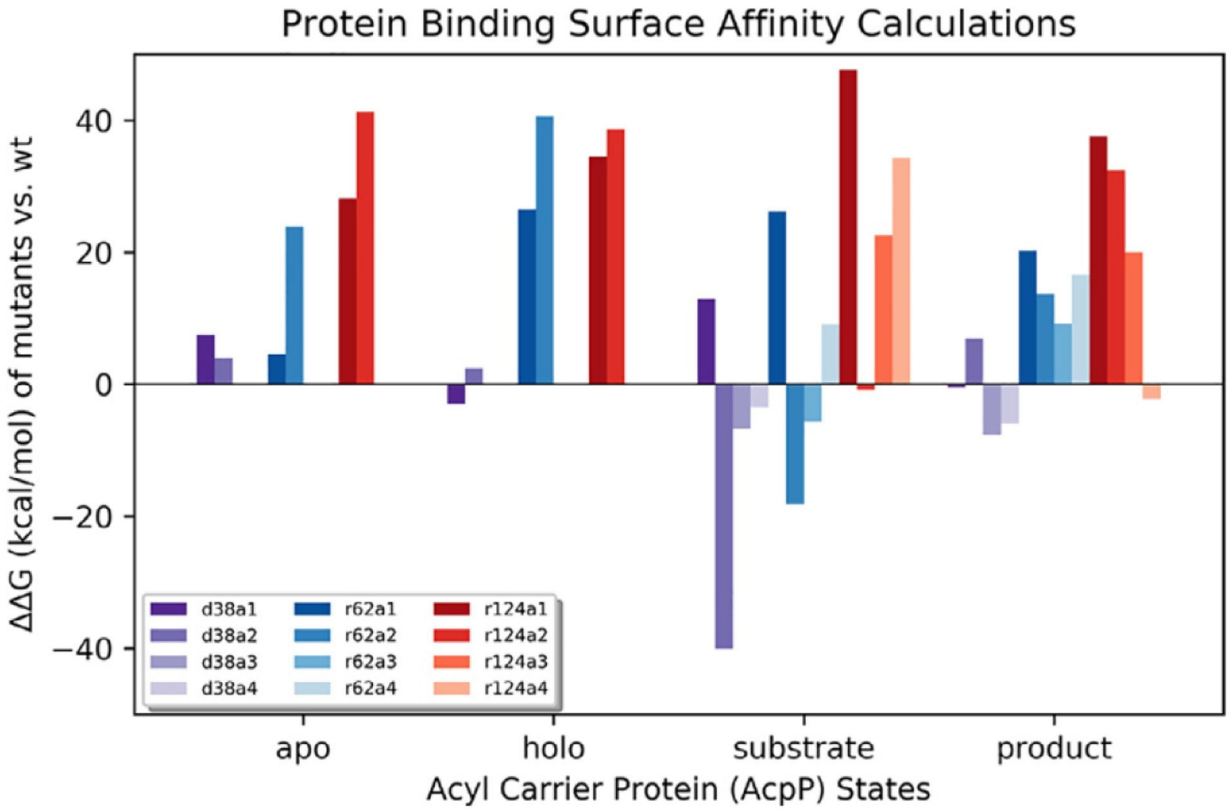

Fig. 8.

Protein binding surface affinity ΔΔG of variants compared to wild-type. Comparisons of ΔG values of the D38A, R62A and R124A variants with wild-type ΔG values are shown above, with the D38A AcpP, R62A FabB and R124A FabB mutants colored in purple, blue and red, respectively. The first and third calculations in each set is measuring the ΔΔG of binding for AcpP1, and the second and fourth calculations in each set are measuring the ΔΔG of binding for AcpP2. For the first and second calculations in each set, both carrier proteins are in identical states.

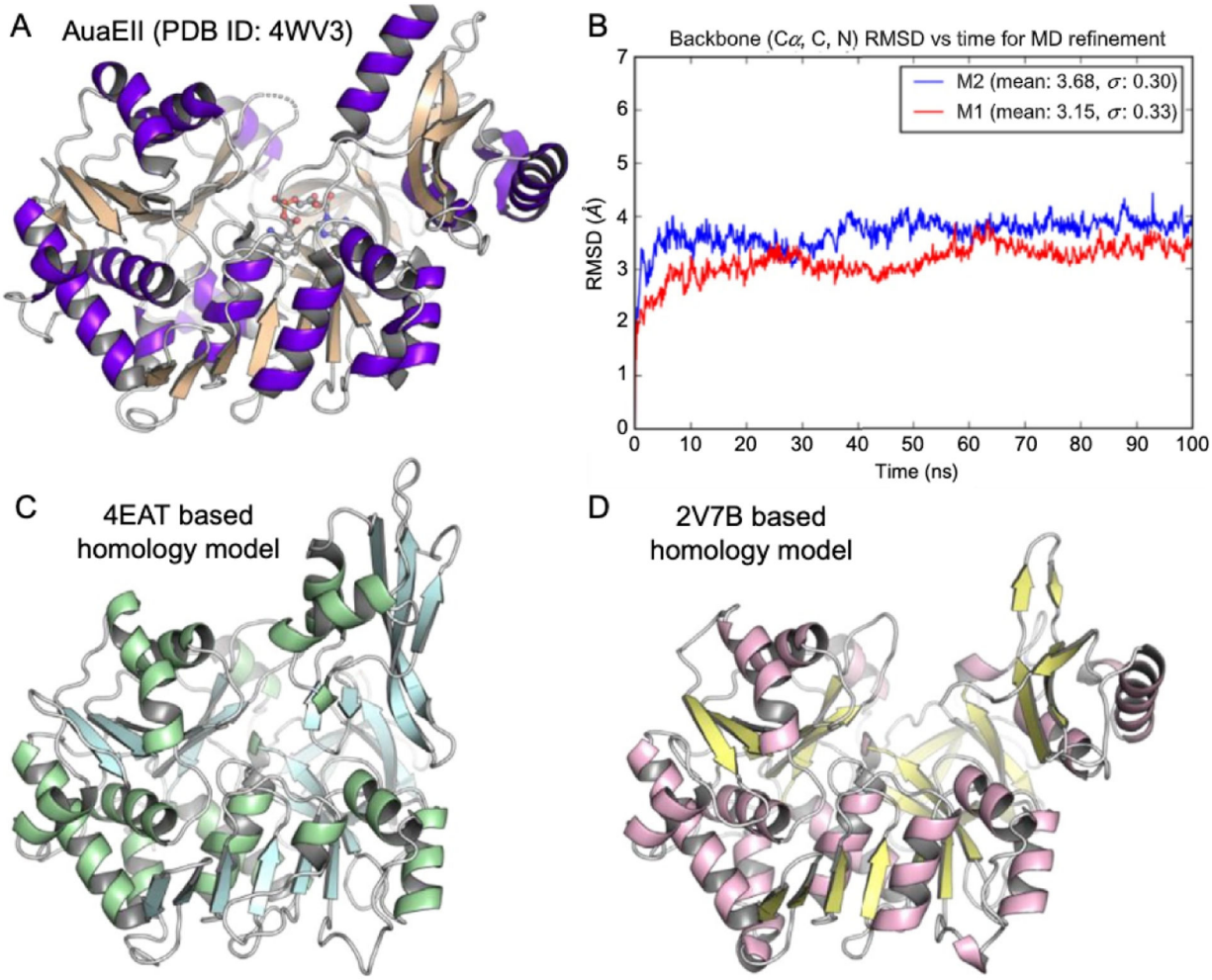

Fig. 9.

(A) The crystal structure of AuaEII bound to anthraniloyi-AMP. (B) 100 ns simulations of AuaE homology models M1 and M2. The final20ns were used for generation of average structures. (C) The M1 structure was generated using PDB ID 4EAT as the threading template. (D) The M2 structure was generated using PDB ID 2V7B as the treading template.

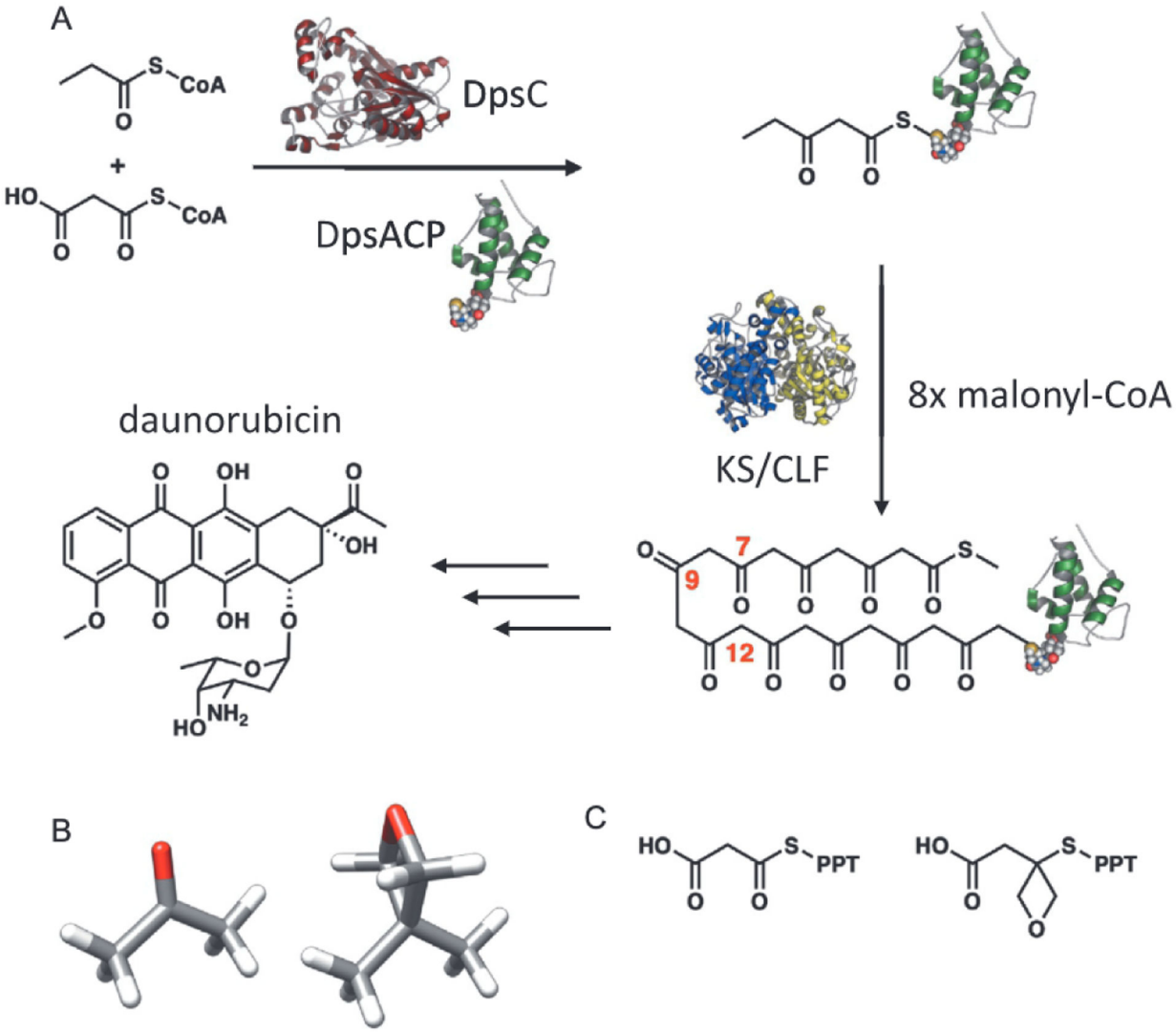

Fig. 10.

(A) DpsC acts as an acyltransferase by transferring to starter unit such as prop-ionyl CoA to the ACP, while also being able to perform the initial chain elongation of the starter unit-loaded ACP with malonyi-CoA. This product is then used as an initial substrate by a ketosynthase chain length factor (KS/CLF) which undergoes eight rounds of elongation. Eventually the biosynthetic pathway ends with the generation of daunorubicin. (B) The carbonyl and oxetane groups have oxygen lone pairs that orient in the same direction, partly due to the C-O-C bond strain. (C) The mimic can be loaded onto phosphopantetheine (PPant), thus mimicking malonyi-PPant.

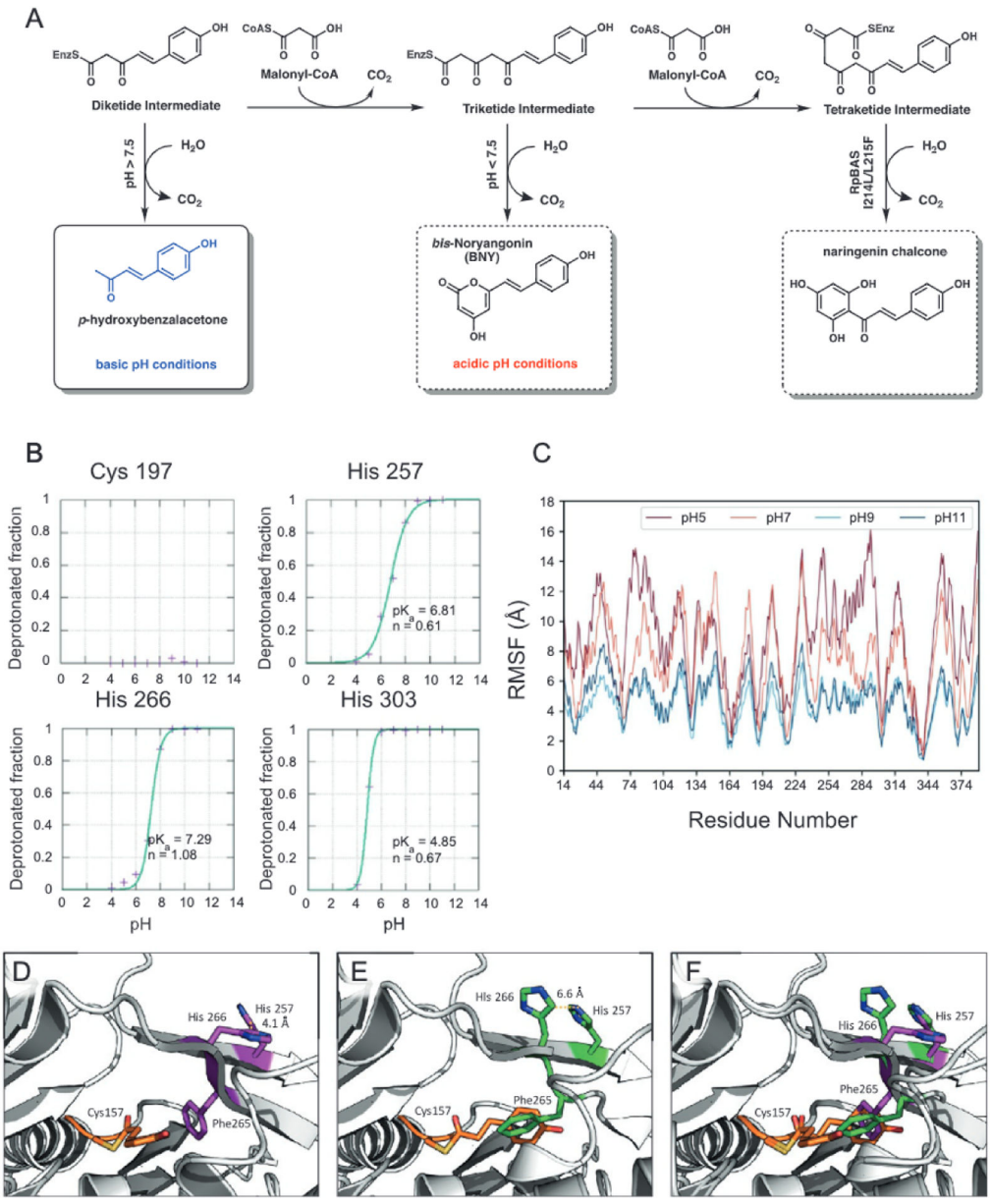

Fig. 11.

(A) RpBAS produces the diketide p-hydroxybenzalacetone under basic pH conditions, and the triketide bis-noryangonin under acidic pH conditions. The double mutant 1214UL21SF functions similarly to MsCHS and produces the tetraketide naringenin chalcone. (B) pKa calculations of titratable sidechains from constant pH MD simulations. (C) RMSF calculations were performed on the final 50 ns of each constant pH simulation. (D) Chain A of the original crystal structure, (E) Chain B of the original crystal structure and (F) Chain A and Chain B of the original crystal structure overlaid to show the different poses of His 257 and His 266.

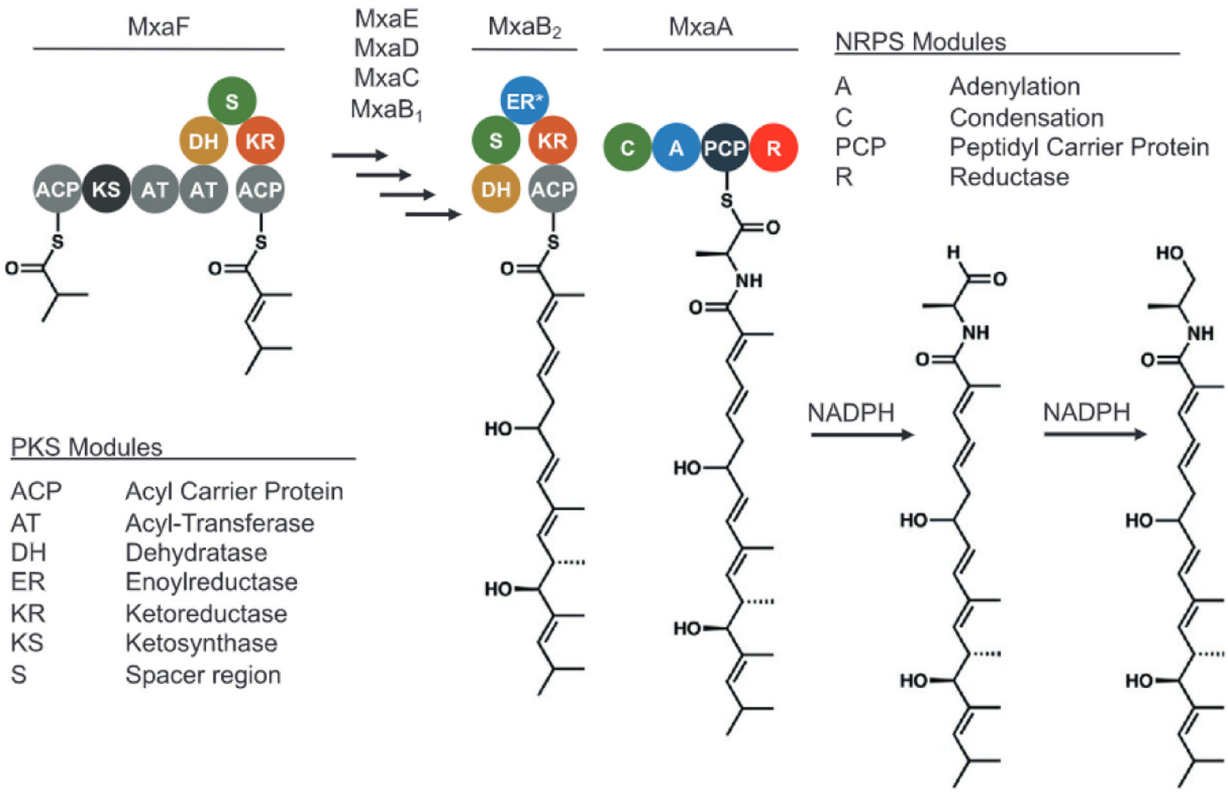

Fig. 12.

Myxalamid biosynthetic pathway: The pathway is composed of six PKS modules (MxaBl-MxaF) and one terminal NRPS module (MxaA) containing a reductase domain (in red).

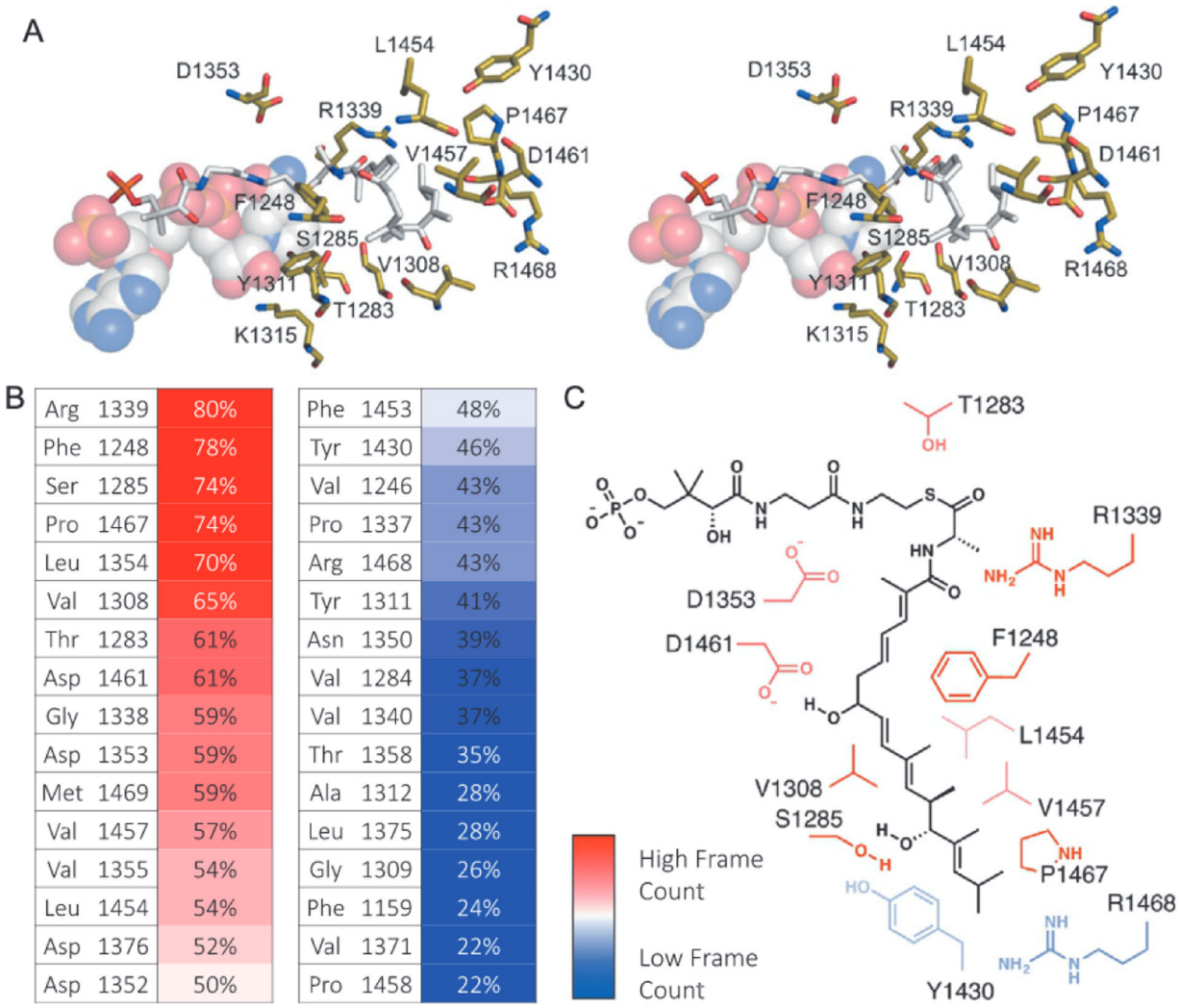

Fig. 13.

(A) Stereo model of the docked truncated substrate. The substrate is shown as white sticks, the cofactor NADPH is shown in spheres, and the active site residues are shown in yellow. (B) Frames from the 100 ns were grouped into 37 clusters, and average representations were assessed for substrate interactions. (C) Residues which had higher occupancy are shown.