Chemical biology
Edited by Helge B. Bode
A detailed view on 1,8-cineol biosynthesis by Streptomyces clavuligerus

Jan Rinkel, Patrick Rabe, Laura zur Horst and Jeroen S. Dickschat*

Abstract

The stereochemical course of the cyclisation reaction catalysed by the bacterial 1,8-cineol synthase from Streptomyces clavuligerus was investigated using stereospecifically deuterated substrates. In contrast to the well investigated plant enzyme from Salvia officinalis, the reaction proceeds via (S)-linalyl diphosphate and the (S)-terpinyl cation, while the final cyclisation reaction is in both cases a syn addition, as could be shown by incubation of (2-13C)geranyl diphosphate in deuterium oxide.

Introduction

Among all classes of natural products the climax of structural diversity and complexity is reached within the largest, the terpenoids. An estimated number of 75,000 different compounds are known from all kinds of organisms including plants [1], bacteria [2-5], fungi [6] and, as recently shown, even social amoebae [7]. These molecules are all made from only a handful of linear and achiral precursors such as geranyl diphosphate (GPP, monoterpenes), farnesyl diphosphate (FPP, sesquiterpenes) and geranylgeranyl diphosphate (GGPP, diterpenes). Terpene cyclases (type I) contain a trinuclear (Mg^2+)3 cluster in their active site that is stabilised by binding to several highly conserved motifs including the aspartate-rich motif (DDXXD) and the NSE triad (ND(L,V)XXXXXE, modified in plants to a DTE triad: DD(L,V)XXXXXE) [8]. Their substrates bind with the diphosphate portion to the (Mg^2+)3 cluster and via hydrogen bridges to a highly conserved arginine (diphosphate sensor) and a RY dimer [9]. The substrate is ionised by diphosphate abstraction and the resulting allyl cation undergoes a domino reaction via a series of cationic intermediates and a final deprotonation or attack of water to yield a terpene hydrocarbon or alcohol. This reaction cascade proceeds in a hydrophobic cavity from which water is excluded to enable carbocation chemistry in an aqueous environment. Furthermore, the hydrophobic cavity provides a template that arranges the substrate in a certain conformation to determine the formation of a specific product. Single residues such as phenylalanines are involved in the stabilisation of cationic intermediates, e.g., by cation–π interactions [8-10]. The overall process usually generates an en-
antiomerically pure (poly)cyclic terpene with several stereo-
genic centres. A large variety of carbon skeletons is accessible, e.g., more than 120 skeletons each representing various stereoisomers and constitutional isomers with different positioning of olefinic double bonds or alcohol functions are known just for sesquiterpenes [11]. The structural diversity of terpenoids can be further increased by the action of tailoring enzymes such as cytochrome P450 monoxygenases and acyl transferases [12,13]. Very few cases are known in which terpene cyclases generate an achiral product as exemplified by the monoterpen e 1,8-cineol (eucalyptol, 1) and the sesquiterpenes germacrene B (2) and α-humulene (3) (Figure 1).

![Figure 1: Selection of achiral terpenes.](image)

A direct 1,6-cyclisation of the monoterpen precursor GPP to 1 is prevented by the topological constraints associated with the (2E) geometry which necessitates the isomerisation of GPP (4) to linalyl diphosphate (LPP, 5) followed by an anti,endo-SN′-cyclisation [14], but the stereochemical course of this reaction is not readily clear and may proceed via either enantiomer of the α-terpinyl cation (6, Scheme 1). Isotopic labelling experiments currently experience a revival [15] and are a very powerful method to follow the enzyme mechanisms of terpene cyclases [16-24] including the stereochemical courses of the cyclisation reactions (reviewed in [25]). While the stereochemical course of the GPP cyclisation to 1 has been investigated for the 1,8-cineol synthase from Salvia officinalis [26,27], it is unknown for the bacterial enzyme that was recently reported from Streptomyces clavuligerus [28]. Here we describe isotopic labelling experiments that gave insights into the cyclisation mechanism of the bacterial 1,8-cineol synthase.

**Results**

**The absolute configuration of the intermediate terpinyl cation**

While the two possible cyclisation pathways via (R)- and (S)-6 to 1 cannot be distinguished with unlabelled GPP, its two enantiotopic protons at C-1 (indicated by H\textsubscript{R} for the pro-\textsubscript{R} hydrogen and H\textsubscript{S} for the pro-\textsubscript{S} hydrogen) end up in diastereotopic positions of 1. Thus, a labelling experiment using the deuterated substrates (R)-(1-\textsuperscript{2}H)GPP (H\textsubscript{R}=\textsuperscript{2}H, H\textsubscript{S}=H) and (S)-(1-\textsuperscript{2}H)GPP (H\textsubscript{R}=H, H\textsubscript{S}=\textsuperscript{2}H) can give insights whether the cyclisation proceeds via (R)- or (S)-6, by determination in which of the distinguishable diastereotopic positions the label ends up. The synthesis of the two enantiomers of (1-\textsuperscript{2}H)GPP (Scheme S1, Supporting Information File 1) was performed by

![Scheme 1: Cyclisation of GPP to 1 via the (R)-terpinyl cation ((R)-6, left) or the (S)-terpinyl cation ((S)-6, right).](image)
Alpine borane reduction [29] (both enantiomers of this reagent are commercially available) of (1-^2^H)geranial to (R)- and (S)-(1-^2^H)geraniol that were obtained with high enantiomeric excess (>95% ee) as determined by Mosher ester analysis (Figure S1, Supporting Information File 1). The alcohols were subsequently converted into the corresponding diphosphates using triethylammonium phosphate in trichloroacetoneitrile [30,31]. The gene encoding the 1,8-cineol synthase [28] was cloned into the yeast-to-*Escherichia coli* shuttle vector pYE-Express by homologous recombination in yeast [32], followed by expression in *E. coli* BL21. The protein was purified by Ni^{2+}-NTA affinity chromatography and used to convert both (R)- and (S)-(1-^2^H)GPP into (^2^H)-1 (in agreement with the findings described in reference [28], 1 is the only product from unlabelled GPP as was shown by GC–MS, Figure S2, Supporting Information File 1). The obtained products were analysed by HSQC spectroscopy (Figure 2). While for unlabelled 1 a 2:2 signal intensity is observed for the crosspeaks representing the

![Figure 2](image.png)

**Figure 2:** Partial HSQC spectra showing the region of crosspeaks for H_A and H_B connected to C-3 and C-5 of A) unlabelled 1, B) (^2^H)-1 obtained by enzymatic conversion of (R)-(1-^2^H)GPP, and C) (^2^H)-1 obtained by enzymatic conversion of (S)-(1-^2^H)GPP. The indicated chemical shift data are for unlabelled 1 (for full data cf. Table S1, Supporting Information File 1).
two pairs of enantiotopic hydrogens $H_A$ and $H_B$ connected to carbons C-3 and C-5, the sample obtained from (R)-(1-$^{13}$C)GPP gave a 1:2 ratio of signal intensities (i.e., $H_A = ^2$H), while the sample from (S)-(1-$^2$H)GPP resulted in ratio of 2:1 by peak integration (i.e., $H_B = ^2$H), indicating the cyclisation via (S)-LPP ((S)-5) and the (S)-terpinyl cation ((S)-6) (Scheme 1, right).

**Syn versus anti addition in the final ring closure**

The final cyclisation step from (S)-7 via 8 to 1 can in principle proceed either via a syn or an anti addition to the olefinic double bond, requiring a protonation of the original C-2 of GPP. To distinguish between these alternatives (2-$^{13}$C)GPP was synthesised from sulcatone (Scheme S2, Supporting Information File 1) and converted by the 1,8-cineol synthase in deuterium oxide. The obtained product was analysed by HSQC spectroscopy (Figure 3A), showing that deuterium is taken up into the exo position at the $^{13}$C-labelled C-2 of 1 (indicated by $H'$), while the endo position ($H''$) is occupied by the proton from the substrate, resulting in a strong crosspeak [22]. Furthermore, deuterium incorporation at C-2 was indicated by a strongly enhanced triplet in the $^{13}$C NMR spectrum due to $^{13}$C-$^2$H-spin coupling (Figure S4, Supporting Information File 1) [11,20,23]. This finding is in agreement with a syn addition to the olefinic double bond of (S)-7 in the final cyclisation step. It is possible that the proton is directly transferred from the protonated hydroxy function in (S)-7 to C-2 (Figure 3B). Alternatively, a deprotonation of (S)-7 to the hypothetical neutral intermediate α-terpineol followed by reprotonation at C-2 from the Si face can be assumed, but these two alternatives cannot be distinguished based on the labelling experiments described here.

**Discussion**

Plant and bacterial terpene cyclases show important structural differences [33]. While plant monoterpane synthases
composed of α and β domains and exhibit a quaternary α2β2 structure [34,35], bacterial mono- and sesquiterpene cyclases are monodomain enzymes (α) [9,10,36]. Accordingly, also the 1,8-cineol synthases from Salvia officinalis and from Streptomyces clavuligerus are not related and have evolved independently. While the plant enzyme was shown to convert GPP via \((R)-LPP ((R)-5)\) and the \((R)\)-terpinyl cation ((R)-6) into 1 [27], the experiments described here revealed a different course for the bacterial enzyme via (S)-6. This finding is particularly interesting, because it reflects the frequent observation that the (chiral) products of bacterial terpene cyclases represent the opposite enantiomers as usually generated by plant enzymes [24,37,38]. Both enzymes from Salvia officinalis and from Streptomyces clavuligerus share the syn addition in the final cyclisation step which can be rationalised by a direct intramolecular proton transfer, circumventing the need of a low-energy neutral intermediate such as \(\alpha\)-terpineol. However, in case of the sesquiterpene ethers corvol ethers A (19) and B (18) a reprotonation step was shown to proceed from the opposite face than the preceding attack of water, thus excluding a direct proton transfer from oxygen to the neighbouring carbon (reactions from 12 to 14 in Scheme 2). Conclusively, reprotonation of the neutral intermediate 13 is possible in this case [21].

**Experimental**

**Cloning and homologous recombination**

Cells of Streptomyces clavuligerus ATCC 27064 were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany). Cultivation was done in liquid 65 Gym medium (4 g yeast extract, 4 g glucose, 10 g malt extract, 1 L water, pH 7.2) and isolation of genomic DNA was performed using a standard protocol. The gene WP_003952918 encoding the 1,8-cineol synthase was amplified using forward primer (ATGCCCGCCGGCCACGAAGA) and reversed primer (TCACCAAGGGGTGGTGCC). The isolated PCR product was elongated for homologous recombination with pYE-Express in yeast with a second set of primers (GGCAGCCATATGGCTAGCATGACTGGTGGAATGCCCGCCGGCCACGAAGA and TCTCAGTGGTGTTGGTGCTCGAGTTCACCAAGGGGTGGTGCC). This elongated product was transformed in Saccha-

---

**Scheme 2:** Mechanism for the cyclisation of FPP to corvol ethers A (19) and B (18). WMR: Wagner-Meerwein rearrangement.
romyces cerevisiae FY34 together with linearised vector pYE-Express [32] (EcoRI and HindIII digestion) using the LiOAc/SS carrier DNA protocol [39]. Transformed cells were plated on SM-URA medium (20 g glucose, 1.7 g yeast nitrogen base, 5 g ammonium sulphate, 0.77 g nutritional supplement minus uracil, 24 g agar, 1 L water) and grown for 3 days at 28 °C. Plasmids were isolated using the kit Zymoprep Yeast Plasmid Miniprep II (Zymo Research, Irvine, USA), shuttled in *E. coli* BL21 by electroporation and confirmed by sequencing.

**Incubation experiments with (1R)- and (1S)-(1-2H)GPP**

A preculture of *E. coli* BL21 cells carrying the plasmid pYE_WP003952918 in 2YT medium (16 g trypton, 10 g yeast extract, 5 g NaCl, 1 L water, pH 7.2) was grown overnight to inoculate a 2YT main culture (2 L). The cultures were shaken at 160 rpm, 37 °C until OD<sub>600</sub> = 0.4 was reached. Prior to induction with IPTG (0.4 mM), the cultures were cooled down to 18 °C. After incubation overnight (160 rpm, 18 °C), cells were harvested by centrifugation (5000 g, 4 °C, 45 min) and resuspended in binding buffer (20 mL; 20 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 mM NaCl, 20 mM imidazole, 1 mM MgCl<sub>2</sub>, pH 7.0). The cells were crushed by ultra-sonification (6 × 1 min, 4 °C) and the cell debris pellet was separated by centrifugation (10 min, 15500 g, 4 °C). The soluble fraction was loaded onto a Ni<sup>2+</sup>-NTA affinity column (Novagen) and treated with binding buffer (2 × 10 mL). The target protein was then eluted with elution buffer (2 × 10 mL; 20 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 0.5 M imidazole, 1 mM MgCl<sub>2</sub>, pH 7.0) and used directly for incubations with 5 mg of (1R)- and (1S)-(1-2H)GPP, solved in incubation buffer (50 mM Tris/HCl, 10 mM MgCl<sub>2</sub>, 20% v/v glycerol, pH 8.2) to reach a final substrate concentration of 0.2 mg/mL. The enzyme reaction was incubated for 2 h at 28 °C, overlaid with 400 μL (2H<sub>6</sub>)benzene and further incubated overnight. The organic phase was separated, dried over MgSO<sub>4</sub> and the extract was directly analysed by GC–MS and NMR.

**Incubation experiment with (2-13C)GPP**

Enzyme purification starting from an *E. coli* expression culture (0.5 L) was performed as described above. The last washing fraction was substituted with <sup>2</sup>H<sub>2</sub>O-based binding buffer and elution was done with <sup>2</sup>H<sub>2</sub>O-based elution buffer. The first elution fraction was incubated with (2-13C)GPP (0.8 mg) for 16 h at 28 °C. The enzyme reaction was extracted with (2H<sub>6</sub>)benzene (0.6 mL), dried with MgSO<sub>4</sub> and the extract was analysed directly by GC–MS and NMR.

**NMR spectroscopy**

To record NMR spectra, instruments AV Avance DME-500 (500 MHz), DPX-400 (400 MHz) and AV III HD Cryo (700 MHz) from Bruker were used. Solvent signals were used to reference the spectra (1H NMR, residual proton signals: (2H<sub>6</sub>)benzene δ = 7.16; 13C NMR: (2H<sub>6</sub>)benzene δ = 128.06 [40].

**GC–MS analysis**

An Agilent 7890B gas chromatograph equipped with a HP5-MS silica column (30 m, 0.25 mm inner diameter, 0.50 μm film) connected to an Agilent 5977A inert mass selective detector was used to acquire GC–MS data. Instrumental settings were: (1) inlet pressure: 77.1 kPa, He: 23.3 mL/min, (2) transfer line: 250 °C, (3) electron energy: 70 eV. The GC was set to 50 °C starting temperature for 5 min, then increasing with 5 °C per minute to 320 °C and holding this temperature for another 5 min. The injection volume was 2 μL and the inlet was operating in split mode (10:1, 60 s valve time). Helium was used as the carrier gas at 1 mL/min. Retention indices were determined against a homologous series of n-alkanes (C<sub>8</sub>−C<sub>10</sub>).

**Synthesis of (2-13C)geranyl diphosphate**

(2-13C)Geraniol was synthesized as reported previously [41]. The synthetic (2-13C)geraniol (16 mg, 0.072 mmol, 1.0 equiv) was dissolved in dry THF (0.3 mL) and PbB<sub>3</sub> (8.1 mg, 0.029 mmol, 0.4 equiv) was added at 0 °C. The solution was stirred for 45 min at room temperature. The reaction mixture was hydrolyzed by addition of ice cold water and extracted three times with pentane. The combined organic layers were dried over MgSO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was used for phosphorylation.

In a second flask, to a solution of (n-Bu<sub>4</sub>)<sub>2</sub>PO<sub>4</sub> (97 mg, 0.11 mmol, 1.5 equiv) in dry CH<sub>2</sub>Cl<sub>2</sub> (1.0 mL) the crude product of the allyl bromide (1.0 equiv) was added and the reaction mixture was stirred for 2 h at room temperature and then concentrated under reduced pressure. The colorless oil was loaded onto an ion exchange column (DOWEX 50W-X8, NH<sub>4</sub><sup>+</sup> form). Elution of the product was performed by addition of two column volumes of ion exchange buffer (0.03 M NH<sub>4</sub>HCO<sub>3</sub> in 2% iPrOH/H<sub>2</sub>O). Freeze drying yielded the product as a white solid (14.1 mg, 0.04 mmol, 55%).

1H NMR (500 MHz, H<sub>2</sub>O) δ 5.37 (dt, 1<sup>1</sup>J(1H,CH) = 156.7 Hz, 3<sup>1</sup>J(1H,H) = 6.9 Hz, 1H, CH<sub>3</sub>), 5.16–5.11 (m, 1H, 1 × CH<sub>3</sub>), 4.44–4.38 (m, 2H, 1 × CH<sub>2</sub>), 2.12–2.06 (m, 2H, 1CH<sub>2</sub>), 2.06–2.00 (m, 2H, 1CH<sub>2</sub>), 1.65 (d, 3<sup>2</sup>J(CH,H) = 5.2 Hz, 3H, 1 × CH<sub>3</sub>), 1.62 (s, 3H, 1 × CH<sub>3</sub>), 1.56 (s, 3H, 1 × CH<sub>3</sub>) ppm; 13C NMR (125 MHz, H<sub>2</sub>O) δ 141.9 (d, 1<sup>1</sup>J(C,C) = 72.7 Hz, 1 × C<sub>q</sub>), 133.7 (1 × C<sub>p</sub>), 124.1 (1 × CH<sub>2</sub>), 119.6 (d, 3<sup>2</sup>J(C,H) = 8.2 Hz, 1 × CH<sub>2</sub>), 38.8 (d, 2<sup>2</sup>J(C,C) = 2.7 Hz, 1 × CH<sub>2</sub>), 25.6 (d, 2<sup>2</sup>J(C,C) = 2.9 Hz, 1 × CH<sub>2</sub>), 24.8 (1 × CH<sub>2</sub>), 16.9 (1 × CH<sub>3</sub>), 15.6 (d, 2<sup>2</sup>J(C,C) = 1.2 Hz, 1 × CH<sub>3</sub>) ppm; 31P NMR (202 MHz, H<sub>2</sub>O) δ −10.0 (1 × P), −10.6 (1 × P) ppm.


License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at:
doi:10.3762/bjoc.12.225
Computational methods in drug discovery

Sumudu P. Leelananda and Steffen Lindert*

Abstract
The process for drug discovery and development is challenging, time consuming and expensive. Computer-aided drug discovery (CADD) tools can act as a virtual shortcut, assisting in the expedition of this long process and potentially reducing the cost of research and development. Today CADD has become an effective and indispensable tool in therapeutic development. The human genome project has made available a substantial amount of sequence data that can be used in various drug discovery projects. Additionally, increasing knowledge of biological structures, as well as increasing computer power have made it possible to use computational methods effectively in various phases of the drug discovery and development pipeline. The importance of in silico tools is greater than ever before and has advanced pharmaceutical research. Here we present an overview of computational methods used in different facets of drug discovery and highlight some of the recent successes. In this review, both structure-based and ligand-based drug discovery methods are discussed. Advances in virtual high-throughput screening, protein structure prediction methods, protein–ligand docking, pharmacophore modeling and QSAR techniques are reviewed.

Introduction
Bringing a pharmaceutical drug to the market is a long term process that costs billions of dollars. In 2014, the Tufts Center for the Study of Drug Development estimated that the cost associated with developing and bringing a drug to the market has increased nearly 150% in the last decade. The cost is now estimated to be a staggering $2.6 billion dollars. The probability of a failure in the drug discovery and development pipeline is high and 90% of the drugs entering clinical trials fail to get FDA approval and reach the consumer market. Approximately 75% of the cost is due to failures that happen along the drug discovery and design pipeline [1]. Nowadays with faster high-throughput screening (HTS) experiments, which can assay thousands of molecules with robotic automation, human labor associated with screening of compounds is no longer necessary. However, HTS is still expensive and requires a lot of resources of targets and ligands. These resources are frequently not available in academic settings. Additionally, many pharmaceutical companies are now looking for ways that can avoid screening of
Figure 1: Schematic representation of a computer-aided drug discovery (CADD) pipeline. CADD methods are broadly classified into structure-based and ligand-based methods. Structure-based methods require the 3D information of the target to be known. Ligand-based methods are used when the 3D structure of the target is not known. They use information about the molecules that bind to the target of interest. Hits are identified, filtered and optimized to obtain potential drug candidates that will be experimentally tested in vitro.
developed frequently. It has immense potential and promise in the drug discovery workflow. In this review we give an overview of structure-based and ligand-based methods used in CADD, focusing on recent successes of CADD in the pharmaceutical industry. We outline structure prediction tools that are routinely used in structure-based drug discovery, widely used docking algorithms, scoring functions, virtual high-throughput screening, lead optimization and methods of assessment of ADME properties of drugs.

**Review**

**Structure-based drug discovery (SBDD)**

If the three-dimensional structure of a disease-related drug target is known, the most commonly used CADD techniques are structure-based. In SBDD the therapeutics are designed based on the knowledge of the target structure. Two commonly used methods in SBDD are molecular docking approaches and de novo ligand (antagonists, agonists, inhibitors, etc. of a target) design. Molecular dynamics (MD) simulations are frequently used in SBDD to give insights into not only how ligands bind with target proteins but also the pathways of interaction and to account for target flexibility. This is especially important when drug targets are membrane proteins where membrane permeability is considered to be important for drugs to be useful [10,11].

Successes have been reported for SBDD and it has contributed to many compounds reaching clinical trials and get FDA approvals to go into the market [12]. HIV-1 (Human Immunodeficiency Virus 1) protease is a prime drug target for anti-AIDS therapeutics. In the early 1990s many approved HIV protease inhibitors were developed to target HIV infections using structure-based molecular docking. It was a ground breaking success at that time and made it possible for HIV infected individuals to live longer than they could have without the treatment [13,14].

Saquinavir is one of the first HIV-1 protease targeted drugs to reach the market (Figure 2a and 2c) [15]. Amprenavir is another drug that was developed to target HIV-1 protease that was also developed influenced by SBDD (Figure 2b and 2d) [16]. In another study, structure-based computational methods have developed to target HIV-1 protease that was also developed influenced by SBDD (Figure 2b and 2d) [16].

---

*Figure 2: FDA approved drugs Saquinavir and Amprenavir for the treatment of HIV infections. (a) The structure of Saquinavir in complex with HIV-1 protease (3OXC) (b) the structure of Amprenavir in complex with HIV-1 protease (3NU3) (c) the molecular structure of Saquinavir and (d) the molecular structure of Amprenavir. Amprenavir and Saquinavir target HIV-1 protease and, in part, have been discovered through structure-based computer aided drug discovery methods.*
been used to predict binding sites, which are important for inhibitor binding, in AmpC beta lactamase which have been experimentally verified [17]. FDA approved Dorzolamide is a carbonic anhydrase II inhibitor which is used in the treatment of glaucoma and was developed using structure-based tools (Figure 3) [7,8].

Protein structure determination
All structure-based methods rely on the three-dimensional target structure. The most common way to determine a protein structure is by X-ray crystallography and NMR spectroscopy. Recently, cryo-electron microscopy (cryoEM) has experienced a ‘resolution revolution’, leading to an increasing number of near-atomic resolution structures [18]. Experimental methods such as X-ray crystallography and NMR spectroscopy are associated with cost and time constraints, and are also limited by experimental challenges. X-ray crystallography is only possible if the target protein can be crystallized. Some proteins, for example membrane proteins which account for about 60% of the approved drug targets today [19], are usually difficult to crystallize, thus experimental methods are not always successful in determining their structures [20]. One of the disadvantages of NMR is that it generally is limited to smaller proteins. Attempts are continuously being made to overcome these challenges and limitations of experimental methods [21].

SBDD methods rely on the protein structure and in the cases where the target structure is not possible to be determined by experimental methods, computational methods become useful. Determining structures from sequences using computational methods is a powerful tool that can bridge the sequence–structure gap. Importance of protein structure prediction methods and their role in drug discovery pipeline are well reviewed in literature [22-25]. Several methods have been used for protein structure prediction including homology modeling [26,27], threading approaches [28], and ab initio folding [29,30]. Several computational protein structure prediction tools that are commonly used are listed in Table 1. Large-scale genomic protein structure modeling has also been accomplished [31,32].

Homology (comparative) Modeling
Homology modeling is a popular computational structure prediction method for obtaining the 3D coordinates of structures. It is well known that the protein structure remains more conserved than the sequence during evolution [33,34]. The basis for homology modeling is the fact that evolutionary-related proteins often share similar structures. Knowing structures that have amino acid sequences similar to the target sequence of interest, can assist in predicting the target structure, function and even possible binding and functional sites of the structure.

In homology modeling, the first task is to find a homologous sequence to the sequence of interest. To do that, the sequence is compared against a database of protein sequences where the three-dimensional structures are known [35]. NCBI Basic Local Alignment Search Tool (BLAST) is one of the most popular bioinformatics sequence alignment tools used with sequence similarity searches [36]. Once a homologous protein structure for the sequence has been identified, building the models for the target structure is done using comparative modeling algorithms [37]. The models built are evaluated and refined. Assessment of the general stereochemistry of a protein structure, such as satisfaction of bond lengths and angle restraints of generated models, is done in the model evaluation stage. Once the models are verified to be acceptable in terms of their stereochemistry, they are then evaluated using 3D profiles or scoring functions that were not used in their generation. It is generally possible to
use homology modeling to predict the structure of a protein sequence that has over 40% identity to a protein of a known three-dimensional structure. When the sequence similarity drops below 30%, homology modeling is not reliable enough for structure prediction [26]. Success of homology modeling is well documented and has been continuously shown in CASP (Critical Assessment of protein Structure Prediction) which is a biannual competition aimed at determining protein structure using computational methods [38,39].

Homology modeling is commonly applied in structure-based drug discovery to predict target structures that are important in diseases [40,41]. Homology modeling of HIV protease from a distantly-related structure has been used in the design of inhibitors for this structure [42]. Similarly, M antigen structure prediction by homology modeling has given insights into function by revealing that the structures and domains are similar to fungal catalases [43]. One of the pioneering comparative modeling servers developed in the early 1990s, which is still popular today, is the SWISS-MODEL server [44,45]. MODELLER is also a popular comparative modeling program that is available as a server and also as a standalone program [46]. HHpred which is available as a server uses hidden Markov model (HMM) profiles for the detection of homology and structure prediction [47].

Fold recognition (threading)
Fold recognition or threading methods are used to identify proteins that do not have any sequence similarities but still have similar folds [35,48]. Fold recognition is based on the fact that over billions of years of protein structure evolution, considerable sequence divergence is observed but only small overall structural changes have occurred in protein folds [49]. Here the sequence of a known protein structure is replaced by the query sequence of the target of interest for which the structure is not known. The new “threaded” structure is then evaluated using various scoring methods [50,51]. This process is repeated for all experimentally determined 3D structures in a database and the best fit structure for the query sequence is obtained [35]. This process of identifying the best structure corresponding to the target sequence is known as fold recognition and has been used in structure-based drug discovery studies [48]. GenTHREADER is a popular fold recognition program that uses neural networks for the evaluation of the alignments [52]. MUSTER is a freely available webserver that generates sequence-template alignments for a query sequence and identifies best structure matches from the PDB [53]. In addition to sequence profile alignments, it also uses multiple structure information as well. DescFold is another webserver which employs SVM-based machine learning algorithms in protein fold recognition [54].

Ab initio (de novo) modeling
Ab initio or de novo modeling is employed when there is no sufficiently homologous structure to use comparative modeling. De novo protein modeling does not rely on a template structure. It models the target structure solely based on the sequence. Ab initio structure prediction implemented in Rosetta is a popular

| Table 1: Some of the popular structure prediction tools, methods of prediction and their availability. |
|---|---|---|---|
| Tool | Method | Availability | Citation |
| Homology | 3D-JIGSAW | Fragment-based assembly | server | [58] |
| MODELLER | Satisfaction of spatial restraints | server/download | [46] |
| HHpred | Pairwise comparison of profile HMMs | server/download | [47] |
| RaptorX | Single/multi-template threading, alignment quality prediction | server | [59,60] |
| Swiss model | Fragment-based assembly and local similarity | server | [44] |
| Phyre2 | Advanced remote homology detection, effect of amino acid variants | server | [61,62] |
| Fold recognition | MUSTER | Profile-profile alignment with multiple structural information | server | [53] |
| GenTHREADER | Sequence alignment, threading evaluation by neural networks | server/download | [52] |
| I-TASSER | Iterative template fragment assembly | server/download | [63] |
| Ab initio | QUARK | Replica-exchange MC and optimized knowledge-based force field | server | [57] |
| Rosetta/Robetta | Fragment assembly, simulated annealing | server/download | [55,64,65] |
| I-TASSER | Fragment assembly | server/download | [63] |
| CABS-FOLD | User provided distance restraints from sparse experimental data | server | [66] |
| EVfold | Calculate evolutionary variation by co-evolved residue pairs | server | [67] |
De novo structure prediction technique [55]. Here a knowledge-based scoring function is used to guide a fragment-based Monte Carlo search in conformation space. This method will generate a protein-like structure having centroid atoms to represent the side chains. Another step follows to refine this centroid-based structure using an all-atom refinement function in order to relax the structure. Rosetta protein structure prediction methods have shown successes in CASP experiments [56]. Ab initio structure prediction server QUARK, developed by the Zhang group has also shown great success in recent CASP experiments [57]. QUARK uses atomic knowledge-based potential functions and models are built from small residue fragments by replica exchange Monte Carlo simulations. In both CASP9 and CASP10, QUARK was the number one ranked server in the template free modeling category outperforming the Rosetta server though Rosetta remains to be one of the most popular methods of ab initio structure prediction. Many other ab initio structure prediction software packages have been developed in the last three decades and some of the popular ones are listed in Table 1.

De novo modeling with sparse experimental restraints
Ab initio prediction of protein structures starting from the sequence is challenging and success is often limited to only small proteins [65]. However, ab initio structure prediction can be guided by the use of sparse experimental data [68]. NMR information has been used in many studies to intelligently guide protein structure prediction [69-71]. NMR Nuclear Overhauser Effect (NOE) data and chemical shifts have been used in combination with Rosetta ab initio structure prediction to obtain better protein structure predictions [72]. Freely available CABS-FOLD uses a reduced representation approach and lets the user provide experimental distant restraints in ab initio structure prediction [66]. This method was successful when tested in CASP6 for targets for which the necessary NMR data already existed [69]. NMR data is not the only form of experimental data that can be used in ab initio structure prediction. With the EM-Fold method it is possible to obtain atomic level protein structures using only the protein sequence information and medium-resolution electron density maps [73,74]. Sparse electron paramagnetic resonance (EPR) spectroscopy data has also been used in high-resolution de novo structure prediction [75-77].

Protein and small molecule databases
Information about drug molecules and target structures is critical in using SBDD tools and many repositories collect and store such information about small molecules and target proteins. PubChem, a small molecule repository is available through NIH which contains millions of biologically relevant small molecules [78]. ZINC is a virtual high-throughput screening compound library which is a free public resource [79,80]. This database contains over 35 million molecules that are purchasable and are available in 3D formats. These molecules have all been pre-processed and are ready for docking. DrugBank has about 5000 small molecules and more than 3000 of these are experimental drugs [81]. There are over 800 compounds in DrugBank that are FDA approved.

The Protein Databank (PDB), which was first introduced in 1970s, is a global resource that contains a wealth of 3D information about experimentally determined biological macromolecules [82,83]. The structures in the PDB are individual macromolecules, protein–DNA/RNA or protein–ligand complexes. Experimental methods used in structure determination are mostly X-ray crystallography and NMR spectroscopy. As of 2016, the PDB databank contains around 120,000 biological macromolecular structures that have been deposited. It has structural information on over 20,000 bound ligand molecules as well. Swiss-Prot is a database which has non-redundant protein sequences which are manually annotated to contain descriptions such as functional information of protein sequences and post-translational modifications [84]. PDB and Swiss-Prot are both general purpose biological databases.

There are other databases that contain specific biological information as well. The BIND database contains protein complex information and biomolecular interactions [85]. BindingDB contains measured binding affinity information of proteins that are considered to be targets for drugs [86]. This database contains over one million binding data points.

Binding pocket identification and volume calculation
Once a protein’s three-dimensional structure is known, finding binding pockets on that protein is an important next step in structure-based drug discovery. It can give indications of where small molecules can bind to target structures, which are associated with diseases, contributing to increase or decrease of target activity. Binding sites in target proteins can be experimentally determined; for example using site-directed mutagenesis or X-ray crystallography. There are also a variety of computational binding pocket identifying algorithms available for the drug discovery scientific community [87].

Binding pocket predicting algorithms can be grouped into two broad categories; geometry-based and energy-based methods. In many cases the binding pocket is considerably larger than all the other pockets in a target and it has been found that in 83% of enzymes that are single chain, the ligands bind to the largest pocket in the enzyme [88]. According to this finding, the binding pockets of a target could be predicted by the geometry of the target. Therefore the size of the pocket is important for
function as well. One of the geometry-based binding site identification servers is 3V [89]. Even though for some cases the largest pocket or cleft of a protein is its binding pocket, it is not necessarily true for all target proteins. Energy-based methods have been developed to address this issue and have shown more success than geometry-based methods [90]. In Q-SiteFinder a van der Waals probe is used and the interaction energy between the probe and the protein is found in order to identify binding sites of the protein [91]. The SiteHound program is another energy-based method that uses two kinds of probes; a carbon probe and a phosphate probe which are used to identify the binding sites for drug-like molecules and phosphorylated ligands (such as ATP) respectively [92]. The best ligand binding site identified in HIV-1 protease by SiteHound is shown in Figure 4. This ligand binding site is the known inhibitor binding site in HIV-1 protease. Another energy-based method, FTMAP, uses 16 probes in identifying hot spots in structures and was more recently extended to include any user provided small molecule as an additional probe [93,94]. Many other binding pocket finding programs exist. PEP-SiteFinder [95], SiteMap available through Schrodinger [96] and MolSite [97] are a few of these programs.

When the binding pocket of a target is known one significant characteristic to be calculated is its binding pocket volume. With this information elimination of ligands that are too bulky to fit in the pocket can be done during the lead identification process. One algorithm that calculates the volume of a binding pocket is POVME (POcket Volume MEasurer) [98]. McVol is another standalone program that can identify and calculate the volume binding cavities in protein structures by using a Monte Carlo algorithm [99].

Scoring functions used in docking
In molecular docking, how well a drug binds to its target is determined by the binding affinity prediction of the pose. This is done by scoring. Scoring is used to evaluate and rank the target–ligand complexes predicted by docking algorithms. Scoring functions are used in SBDD for scoring and evaluating protein–ligand interactions [100,101]. The scoring function used by docking algorithms is a crucial part of the algorithm. It is used in the exploration of the binding space of the ligand and also in the evaluation of target–ligand complexes in molecular docking. The scoring functions can be categorized into knowledge-based [102,103], force-field based [104,105], empirical [106,107] and consensus [108,109]. These will be discussed below. The accuracy of different scoring functions has been evaluated in the literature [110-113]. These comparative studies that evaluate docking method scoring functions use evaluation criteria such as binding pose, binding affinity and ranking of true binders [114]. Wang et al. evaluated the performance of fourteen different scoring functions using 800 protein–ligand complexes in the PDBbind database [113]. The performance was evaluated by the predicted binding affinities of protein–ligand complexes by different scoring functions. According to this study X-Score, DrugScore and ChemScore were among the best performing scoring functions. Ferrara et al. used nine scoring functions and assessed the performance of these functions using 189 protein–ligand complexes [110]. They found that ChemScore shows the best correlation for experimental binding energies and predicted binding scores. In another study done by Marsden et al., calculated binding free energies with knowledge-based potential function Bleep agreed best with experimental binding constants [115].

Knowledge-based scoring functions
Knowledge-based scoring functions are statistical potentials and are derived from experimentally determined protein–ligand information. The frequency of occurrence of interactions of a large number of target–ligand complexes are used to generate these potentials. The basis of these potentials is the Boltzmann distribution. The frequency of occurrence of atom pairs is converted into a potential using Boltzmann’s distribution of states. Since these potentials use target–ligand complex data already available, they are highly dependent on the dataset used to create them. DrugScore uses knowledge-based potential functions to predict binding affinity [116]. Other knowledge-based scoring functions include the PMF (Potential of Mean Force) [117] and Bleep [118].

Force-field-based scoring functions
Force-field based energy functions are developed using classical molecular mechanics. Electrostatic (coulombic) interactions and van der Waals interactions (Lennard-Jones potential)
contribute to the interaction energy between a target–ligand complex. Two of the most widely used molecular mechanical force-fields are CHARMM [119] (Chemistry at HArvard Macromolecular Mechanics) and AMBER [120] (Assisted Model Building and Energy Refinement) which have been built mainly for molecular dynamics simulations. The molecular docking program DOCK [121] uses force-field based scoring functions derived from molecular dynamics force-field AMBER.

Empirical scoring functions

Empirical scoring functions are obtained by using data from experimentally determined structures and fitting this information to parameters. The idea here is that the binding free energy is calculated as the weighted sum of terms that are uncorrelated. These terms can be the number of hydrogen bonds, hydrophobic effect, and different types of contacts and their types etc. Regression analysis is usually done to obtain weights of the terms using experimental target–ligand complexes with known binding free energy data [122]. Unlike knowledge-based scoring functions, which are obtained by directly converting frequency of occurrence of different interactions into potentials using Boltzmann principle, these functions take into account multiple terms or contributions and find the best weights for each term using regression analyses. The HYDE scoring function is an empirical energy function which is a part of BioSolveIT tools [123]. Here the binding energy of a target–ligand complex is solely estimated by a hydrogen bond term and a dehydration energy term. ChemScore [122] and SCORE [124] are two other scoring functions that are also empirical.

Consensus-based scoring functions

Current scoring functions are not perfect and no one scoring function can do well in every docking complex studied. Consensus scoring was introduced to combine different scoring functions in the hope that it can balance out errors and improve accuracy [125]. Consensus scoring function X-CSCORE [126] was developed by combining three different empirical scoring functions, namely Bohm’s scoring function [127], SCORE and ChemScore. Another example of consensus scoring is MultiScore [128]. This score function is a combination of eight different scoring functions and have shown improved protein–ligand binding affinities.

Protein–ligand docking algorithms

In docking, predictions are made on how intermolecular complexes are formed between a target and a ligand. These algorithms search for the best target–ligand poses with the right conformational state and relative orientation. The algorithms also crudely estimate the binding affinities of the target–ligand complexes in terms of scoring. The docking algorithms therefore comprise a search algorithm that searches the conformational space to find docking poses and a scoring function to predict the affinity of the ligand in that pose. Computationally docking a target structure to a molecule is a challenging process. Even when target flexibility is ignored there are still a huge number of ways a molecule can be docked. The total number of possible modes increases exponentially as the size of the two docked molecules increases. Therefore efficient search methods that are fast and effective, and reliable scoring functions are critical components of docking algorithms.

Once a target protein structure is known and a potential drug binding site has been identified, small molecules that bind to this site need to be determined. In drug discovery, docking algorithms are used to find the best fit between a target and a small molecule drug. Docking algorithms require a target protein structure and a library of small molecules. The target protein structure is usually determined using experimental methods such as X-ray crystallography and NMR, or else it is computationally modeled. Molecular docking aims to predict the binding mode and binding affinity of a protein–ligand complex. A library of small molecules is virtually placed (docked) into the desired protein–target binding site and thousands of possible poses of binding are obtained and evaluated. The pose which is scored with the lowest energy is predicted to be the best possible binding mode. The models are evaluated using a scoring function and the poses are ranked and a group of high ranking compounds are chosen for the next step of experimental verification.

One of the very first studies that developed algorithms to evaluate docking poses by looking at steric overlaps was published in the 1980s [129]. Ever since then many docking algorithms have been developed [130-135]. Popular molecular docking programs include Glide [131], Fred [136], AutoDock3 [137], AutoDock Vina [134], GOLD [138] and FlexX [139]. AutoDock3 uses an empirical scoring function that has five terms. These terms are weighted with experimental target–ligand data. It can model side chain flexibility of the target molecule. AutoDock Vina is the new generation of AutoDock. The scoring function used in AutoDock Vina is a hybrid scoring function that combines knowledge-based and empirical scoring functions [134]. GOLD uses a force-field based scoring function and allows the ligand to be fully flexible. It allows target side chain flexibility to be taken into account. FlexX is an incremental fragment-based docking algorithm where the conformational space sampling is done using a tree-search method. It is an ensemble method that can incorporate target structure flexibility. The scoring function is a modified version of empirical Bohm’s scoring. Glide is a highly popular
The best docking algorithm should be the one with the best scoring function and the best searching algorithm. The performance of various docking algorithms has been evaluated and they are able to generate docked ligand conformations that are similar to experimental complexes [146]. Compared to co-crystallized X-ray structures of target–ligand complexes docking results can sometimes even predict poses with RMSDs of less than 1 Å [147]. Measuring RMSD (root mean square deviation) is the most common way to compare the structural similarity between two superimposed structures. RMSD is given by:

$$RMSD = \sqrt{\frac{1}{n} \sum_{x=1}^{n} d_x^2}$$

where $n$ is the number of atom pairs and $d_x$ is the distance between the two atoms in the $x^{th}$ atom pair.

However, it is important to note that no single docking method performs well for all targets and the quality of docking results is highly dependent on the ligand and the binding site of interest [148-150]. The best four binding poses predicted for the known inhibitor Dorzolamide binding to carbonic anhydrase II obtained by AutoDock Vina are shown in Figure 5.

Preprocessing of target and ligands

Target and ligand preparation steps are crucial and are often done before docking is performed to ensure good screening results [151]. In experimental methods such as X-ray crystallography the hydrogen atoms of structures are not generally present. However, the presence of these atoms and the locations of these bonds are important for molecule docking algorithms. Additionally, the target protein structures, if used without preprocessing, can give rise to potential issues due to missing residues, atom clashes, crystallographic waters and alternate locations. In target preprocessing, missing atoms such as hydrogen are added and atomic clashes are removed. The same is true for the ligands that are used. During ligand preprocessing, ligand three-dimensional geometries are predicted. All possible ionization, stereoisomeric and tautomer states are assigned [152]. The protonation states of structures are also important in prediction of docking poses because protonation states affect how ligands bind to the binding site [100]. Optimizing protonation states of binding pockets and also positions of polar hydrogens can lead to identifying the most native-like docking poses.

SPORES is one program that is used for the preprocessing of proteins for protein–ligand docking. It can generate different protonated states, tautomeric states and stereoisomers for protein structures [152]. LigPrep from the Schrodinger Suite [153] allows to obtain all-atom 3D structures of ligands. It is available through the maestro interface or by command line. A web-based ligand topology generating server, PRODRG, can generate 3D coordinates for ligands that are of equal or better quality than other methods [154].

De novo ligand design

By using fragment-based de novo ligand design it is possible to assemble molecules that are drug-like with much less search space having to be explored. In some cases, de novo drug design is less successful in generating drug candidates compared to other methods such as high-throughput virtual screening methods for large databases. One limitation of this approach can be attributed to its high complexity. When a high-resolution target structure is available, ligand growing programs
such as biochemical and organic model builder (BOMB) can be used to design ligands that bind to the target without using ligand databases [155,156]. Using BOMB it is possible to grow molecules by adding substituents into a core structure. It has been possible to design inhibitors for Escherichia coli RNS polymerase using the de novo drug design program SPROUTS [157]. In another study that used the SPROUT program, novel inhibitors were developed for Enterococcus faecium ligase VanA using hydroxyethylamine as the base template structure [158]. It is generally necessary to synthesize molecules that are obtained by de novo drug design. Whereas when using virtual screening methods, since the screening is usually done with databases of commercially available molecules, it is possible to purchase these molecules without the need to synthesize them. LigMerge is another novel algorithm that can generate novel ligands for drug targets [159]. It uses known ligands of a target and generates models with similar chemical features by finding the maximum common substructure of known ligands. Chemical groups of superimposed ligands are attached to the common substructure. This produces different molecules that have features of the known ligands. The algorithm is able to identify novel ligands for several known drug targets that have predicted affinities higher than their known binders. The AutoGrow software is a drug molecule optimizing program. It can be used to optimize ligands according to various properties and binding affinities and is available to download [160]. If two fragments bind to two non-overlapping nearby sites on a target protein, these fragments can be joined to obtain a possible new drug molecule. In the SILCS (site identification by ligand competitive saturation) method, molecular dynamics simulations are used to identify fragments that bind to a target [161,162]. SILCS uses explicit molecular dynamics simulations where the target molecule is simulated in an aqueous solution that contains different fragments. Using multiple simulations SILCS determines high probability binding areas of the target for the different fragments which can be used in fragment-based drug design. Another de novo ligand design program is LigBuilder which is available for download [163,164]. Here the ligands are either grown or linked by user’s choice, and an empirical scoring function is used to estimate binding affinities.

**Structure-based virtual high-throughput screening (VHTS)**

Structure-based virtual high-throughput screening (VHTS) is large scale in silico screening of drug molecules in databases of small molecule compounds for a target of interest. Here a target is “screened” against a library of drug-like molecules and binding affinities of the ligands to the target are estimated using the scoring functions described previously. In addition to finding the best docking pose VHTS also ranks docking results according to their predicted affinities for the entire database. Since large databases are screened, it is important that the target–ligand docking algorithms used in VHTS are both fast and sufficiently accurate, to be able to identify a subset of possible drug compounds. Small molecules that are predicted to bind to the target with high affinity can be identified. These “hits” are then generally further optimized and subsequently tested experimentally. With improving computational resources and parallel processing cluster availability, it is now possible to screen millions of compounds within a matter of hours or days. Because of VHTS it is possible to experimentally test a rather small number of molecules. Testing thousands of available compounds in databases experimentally may no longer be necessary.

Structure-based high-throughput screening has been used to identify inhibitors of protein kinase CK2 targeting its ATP binding site [165]. CK2 is an important target in developing antitumor drugs. About 400,000 compounds have been screened, from which 12 hits were selected for evaluations using in vitro assays. Out of these hits a novel drug was identified which was able to inhibit CK2 enzymatic activity with an IC_{50} of 80 nM. At the time it was discovered, this drug was considered one of the most potent drugs for a protein kinase. In another study, VHTS was used to show proliferator-activated target agonistic behavior with Sulfonylureas and Glinides binding [166]. This finding has implications in the treatment of type-2 diabetes and it was also confirmed by experimental assays. Recently, virtual screening has been used to find antiviral inhibitors that target the Ebola virus. This study found a lead candidate from the TCM database that shows a decrease in activity for the protein encoded by the virus [3]. This promising candidate also shows good pharmacokinetic properties.

VHTS was also used to find a novel quinolinol that binds to MDM2 in a fashion mimicking the binding of p53 and inhibit the MDM2-p53 interaction [167]. This inhibition can activate p53 in cancer cells. There are many more examples where VHTS has been used in drug discovery studies [168-172].

**Target flexibility in molecular docking**

Experimental methods such as X-ray crystallography represent proteins as static structures. However proteins are dynamic systems and show internal motions. The functionality of proteins is governed by their internal dynamics. In order to explain their function it is important to understand their dynamic characteristics [173]. In the first molecular docking attempts in the 1980s, the rigidity of the target protein was always assumed [129]. This has not changed significantly until recently; some newer docking algorithms can account for target flexibility. It is important to account for target protein flexibility because protein structures are dynamic in nature and their structures change...
upon binding of drug molecules. These changes may involve overall backbone structure rearrangements or they can be subtle where only the side chains near the ligand binding site change to accommodate the bound ligand. However, this dynamic nature of targets is frequently ignored and protein flexibility is underrepresented in CADD. In conventional docking algorithms the target is held rigid while the ligand molecule is generally assumed to be flexible. This rigid body docking of ligands to the target is not realistic and can give misleading results because targets are actually able to freely undergo side chain and backbone movements as a result of ligand binding by an induced fit mechanism [174]. Two approaches that can be taken to account target flexibility are induced fit docking methods and ensemble-based screening methods.

In induced fit docking the target protein structures are modeled as flexible, not rigid. They are able to accommodate induced fit that is caused by the ligand molecule binding to it. Schrödinger has introduced induced fit docking protocols through Glide [174]. RosettaLigand also accounts for target flexibility and shows success in predicting target–ligand poses. All residue side chain flexibility of the ligand binding pocket and target backbone flexibility are taken into account by RosettaLigand. This method uses a Monte-Carlo-based minimization algorithm and the Rosetta full-atom energy function [175].

Ensemble-based docking is an alternative method to induced fit docking. With ensemble-based screening methods there is no need to choose flexible residues of interest to binding [176-180]. The relaxed complex scheme (RCS) method uses structure dynamics and docking algorithms in combination to account for target flexibility [181,182]. One successful application of RCS was reported for HIV integrase. MD simulations that were performed with the holo-structure of HIV integrase bound to a known ligand showed signs of a novel binding pocket opening in close proximity to its active site [183]. RCS ligand docking showed that this binding site is a possible binding pocket for drug molecules. This finding paved the way for the development of raltegravir to treat HIV infection which was later approved by the FDA (Figure 6) [184,185]. These ensemble-based screens use an experimentally determined starting target structure and use various methods to generate target structure ensembles. These methods include molecular dynamics simulations [186], Monte Carlo simulations [187], enhanced sampling [177] or just simply experimental ensembles from NMR or multiple crystal structures, to generate an ensemble of conformations based on the starting target structure. By doing so, conformations of the target structure that are relevant to the biological function which are not accessible by experimental structure determination, can be obtained. The trajectories can be clustered to obtain a set of representative conformations. The difference here is that instead of using one structure, an ensemble of structures is used in docking. Due to the range of conformations generated by these methods a more representative set of small molecules can now bind to the ensemble [6,188,189].

**Figure 6:** The molecular structure of Raltegravir. Raltegravir is an FDA approved drug used in the treatment of HIV infection.

**Enhanced sampling methods**

Ensemble-based methods which are typically employed to account for target flexibility use enhanced sampling methods. One of the tools that is extensively and routinely used to understand protein motions and conformational space that is accessible for protein structures is molecular dynamics (MD) simulations [190]. The most widely used molecular dynamics software packages are NAMD [191], GROMACS [192] and AMBER [120]. The typical time-scale of a molecular dynamics simulation is in the order of nanoseconds to microseconds. However to capture biologically important conformational transitions, frequently it is important to probe dynamics in the order of milliseconds. Simulating on that timescale makes it possible to overcome high-energy barriers in some important biological transitions. With conventional all-atom MD simulations and typically available computational resources this can be a very time consuming process. Millisecond-scale MD simulations are possible with high speed supercomputers, although most computational scientists do not have access to such powerful machines [193]. This is considered a major limitation of MD. It may take months to complete a one microsecond molecular dynamics run on a system having around 25000 atoms on 24 processors [194]. Enhanced sampling methods have been introduced to address this issue [176,195]. With enhanced sampling methods it is possible to find conformational states that are relevant to the function of proteins that are not explored in conventional MD. Enhanced sampling methods introduce a bias on the system being simulated. Several methods of enhanced sampling are introduced in literature, including: accelerated molecular dynamics [196,197], metadynamics [198,199], umbrella sampling [200] and temperature-accelerated molecular dynamics [201,202].
Accelerated molecular dynamics (aMD) simulations reduce the energy barrier of wells or in other words raise the energies of the wells that are below a certain threshold energy [203]. This leaves the high-energy states above the cutoff unaffected. When the original energy of the system is below the calculated energy, an additional potential term is added (a boost potential), thereby allowing energy barriers to be smaller. This makes it possible for the system to access conformations which are not accessible without the energy barrier reduction [203-205].

In metadynamics a history-dependent bias potential (which is a function of a set of collective variables) or a force is added to the Hamiltonian of the system to accelerate the system in consideration by pushing it from the local energy minimum [199]. It is important that the collective variables used can describe the initial, final and intermediates states. Commonly used collective variables are interatomic angles, dihedrals and distances. By doing so, it is possible to sample rare events that are otherwise not sampled by conventional MD. Finding the set of collective variables however is challenging especially when the simulated biological system is more complex. Recently, induced fit docking has been coupled with metadynamics to predict protein–ligand complexes in a reliable way. By incorporating metadynamics with induced fit methods, the predictive power of these methods can be enhanced without requiring too much computational resources [4].

The umbrella sampling technique is used to calculate free energy differences in systems [200]. An additional energy term or a bias potential is introduced to the system along a reaction coordinate. This bias potential can then drive the system from the reactant state to the product state. Each of the intermediate states is simulated by MD. Most of the time, for reasons of simplicity, bias potentials are applied as harmonic potentials [200,206].

In temperature-accelerated MD, the system simulation is done at a high enough temperature which makes it possible to accelerate the sampling. Temperature accelerated MD has been used in the study of ligand dissociation from the inducer binding pocket in the Lac repressor protein [202]. By using this method, it was possible to sample the dissociation trajectories in a relatively short period of time to capture the ligand dissociation. The replica exchange method runs a number of independent replicas in different ensembles of the systems at different temperatures and allows exchange of replica coordinates to take place between these ensembles [207]. This method can also enhance sampling in cases where the energy landscape of a system has many minima and where it is not possible to cross the barriers between them during standard simulation times.

Rigorous binding free energy calculations
Rigorous binding free energy calculations can be used to more precisely estimate the binding affinity of target–ligand complexes and these affinities can be used to rank the fit of drug molecules for a particular target. Binding affinities can be used to infer how drug binding will be affected by target mutations [208]. The potency of a drug is assumed to be directly related to the target–drug molecule binding affinity. Therefore it is important to be able to accurately predict the target–ligand binding affinity [209]. Currently the most accurate approaches to calculate binding free energies are rigorous approaches [210]. Monte Carlo algorithms and molecular dynamics simulations are used for generating ensemble averages to model complexes in the presence of explicit water molecules using classical force-fields. Two rigorous binding free energy approaches are the free energy perturbation (FEP) methods and thermodynamic integration (TI) methods [211-215]. These methods are much more accurate than virtual screening. Both of these methods are rigorous alchemical (non-physical) transformations, where the transformation happens via an alchemical pathway of states in a thermodynamic cycle (Figure 7). By using these intermediate states, the starting state of a biological system can be transformed into another state. Turning off atom charges is one example of an intermediate pathway of an alchemical thermodynamic cycle. The binding free energy is computed as a sum of all the steps in the cycle from unbound to bound.

Free energy perturbation is one of the most popular molecular simulation-based free energy calculation methods [213]. It was first introduced by Zwanzig in the 1950s [216]. This method uses statistical mechanics as well as molecular dynamics and Monte Carlo simulations. It requires that there are a sufficient number of alchemical intermediate states especially if the end state perturbation is large. An alternative to FEP is thermodynamic integration. In thermodynamic integration a coupling parameter is introduced to define a series of non-physical intermediate states. The free energy change of two states is then calculated by integrating the derivative of potential energy over all coupling parameters [215]. The energy calculation methods employed in docking algorithms are fast and therefore useful in screening large databases of molecules. Rigorous free energy based methods are not suggested for screening large databases since they are much more time consuming. Even though energy calculations used in virtual high-throughput screening experiments can lead to the identification of hits, they are not reliable in predicting accurate binding affinities. Therefore they cannot be reliably used in lead optimization [112,217]. Recently, free energy calculation guided (FEP-guided) lead optimization has started to evolve [156]. The novel method BEDAM (binding energy distribution analysis method), based on statistical
Figure 7: An example alchemical thermodynamic cycle for a protein–ligand binding free energy calculation. The protein is shown in blue spheres. The ligand, depicted in solid black, indicates there are no coulombic or van der Waals (VDW) interactions with the environment. The ligand, depicted in solid orange, indicates there are coulombic and VDW interactions with its environment. The systems that are subjected to simulations in each cycle are highlighted in blue boxes. All simulations are run in a water environment. The first step is to add restraints between ligand and the protein in order to keep the ligand confined to the binding pocket and to avoid the ligand leaving the pocket when its interactions are removed. The systems with restraints turned on are indicated by red hexagons. In the next step the coulombic and VDW interactions of the ligand are removed. This step is followed by the removal of the restraints applied to the ligand. Next the coulombic and VDW interactions of the ligand are turned on such that the ligand is in contact with solvent. Summing up the free energy changes along the thermodynamic cycle would give the protein–ligand binding free energy.

mechanics, is used to calculate binding free energies of target-ligand complexes [218]. BEDAM is an implicit solvent method that is implemented using Hamiltonian replica exchange molecular dynamics. Recently BEDAM showed success in the SAMPL4 (statistical assessment of the modeling of proteins and ligands) challenge in predicting free energies of binding for a set of octa-acid host–guest complexes [219]. VM2 is another method used in target–ligand binding energy calculations which falls between rigorous free energy calculation methods and approximate docking and scoring algorithms in its complexity [220]. It is an implicit solvent method and uses empirical forcefields. Its implementation is based on mining minima end point method (M2). In this method the binding site is taken to be fully flexible and the other parts of the target are kept fixed. Due to the flexibility of the binding site, it can adapt according to different bound ligands. The free energy is estimated to be the sum of all local energy minima.

Lead optimization and assessment of ADME and drug safety
When hits are obtained for a target structure by screening small molecule databases, the next step usually is lead optimization. During lead optimization, the effectiveness of promising hits obtained is generally enhanced while at the same time obtaining the desired pharmacological profiles to reach the required affinity, pharmacokinetic properties, drug safety, and ADME (absorption, distribution, metabolism, and excretion/elimination) properties. By increasing the affinity of a drug to the target its potency (efficacy) can be increased. The free energy of binding of a drug is a measurement of the potency of a drug to the target of interest. This could be done by doing alchemical free energy calculations in complex with running molecular dynamics simulations. One simulation starts with the target–ligand bound complex and slowly removes the ligand, and the other slowly removes the ligand from the solution. It is possible to
find chemical changes of a possible drug candidate that can improve its potency using alchemical free energy calculations. This is done by gradually converting one atom of the ligand to another and calculating the binding affinity. These affinity changes with atom modifications can be used as guides for improving potency of drug candidates [194].

The permeability of a drug through the intestines and solubility are both important factors that affect drug absorption [221]. Therefore, in silico prediction of solubility and membrane permeability of drugs is an important part of lead optimization [222]. If an orally administrated drug has poor solubility or a high dissolution rate, the drug tends to be excreted by the body without entering the blood stream. This causes the drug to be inefficient and can even cause other biological side effects. To experimentally measure the solubility, the synthesis of the drug is needed which is a time consuming process. However, predicting solubility using computational methods is fast. It is possible to perform solubility calculations on large molecule libraries without needing a lot of computational resources. The solubility data can assist medicinal chemists to evaluate the drug candidates without having to synthesize molecules at all. This greatly reduces the costs of molecule synthesis and time for experimental solubility measurements. Huynh et al. used an in silico method for the prediction of solubility of docetaxel (DTX), an anti-cancer molecule used to treat various types of cancer [223]. In this study solubility parameters for DTX were obtained using MD simulations. This in silico model was in agreement with the experimental solubility of DTX. Simulation-based approaches are frequently used in computational permeability prediction [224,225]. In one study, trajectories obtained by molecular dynamic simulations have been used to obtain diffusion coefficients of permeation of drug-like molecules through the blood-brain barrier [225]. In silico approaches to predict drug solubility in both aqueous media and DMSO are discussed in a review [226].

Human intestinal absorption of a candidate drug is of high importance because it can affect the bioavailability of a drug. According to the Lipinski’s ‘Rule of 5’, poor absorption or permeation is more likely when: there are more than 10 H-bond acceptors, more than 5 H-bond donors, Log P is over 5, and the molecular weight is over 500 [227]. There are extensions of the Rule of 5 in predicting drug-likeness as well [228]. One such extension later proposed is the ‘Rule of 3’ which was used in the construction of fragment libraries for lead generation [229]. These rules are generalized rules for evaluating the drug-likeness and bioavailability of compounds. Various statistical and mathematical models have been based on these rules and their extensions. Machine learning algorithms such as neural networks have been used in the prediction of drug-likeness and bioavailability [230,231].

QikProp is an ADME program offered by Schrodinger that predicts pharmacologically relevant and physically significant descriptors for small drug-like molecules [232]. The VolSurf package can be used to calculate ADME properties and generate ADME models [233]. These ADME models can then be used to predict the behavior of novel molecules. It can also be used to find molecules with similar ADME properties as active ligands of interest. FAF-Drugs2 is an ADME and toxicity filtering tool that can calculate physicochemical properties, toxic and unstable groups, and key functional components [234]. Even though many possible drug molecules go to experimental verification stage or even animal models, they do not reach clinical trials. This is mostly due to the fact the drugs have poor pharmacokinetic properties and toxicity [235]. Thus filters for ADME properties are important for drug screening [236]. Computational ADME methods have advanced greatly in the last few decades and pharmaceutical companies are showing great interest in this area [237].

**Ligand-based drug design (LBDD)**

The main alternative to SBDD is LBDD. In the case where the potential drug target structure is unknown and predicting this structure using methods such as homology modeling or ab initio structure prediction is challenging or undesirable, the alternative protocol to use is Ligand-based drug design [238,239]. Importantly, however, this method relies on the knowledge of small molecules that bind to the target of interest. Pharmacophore modeling, molecular similarity approaches and QSAR (quantitative structure–activity relationship) modeling are some popular LBDD approaches [240]. In molecular similarity methods, the molecular fingerprint of known ligands that bind to a target is used to find molecules with similar fingerprints through screening molecular libraries [241]. In ligand-based pharmacophore modeling, common structural features of ligands that bind to a target are used to do the screening [242]. QSAR is a computational method that models the relationship between structural features of ligands that bind to a target and the corresponding biological activity effect [243].

**Similarity searches**

The main idea of similarity-based or fingerprint-based approaches is to select novel compounds based on chemical and physical similarity to known drugs for the target. Ligand similarity search methods are simple but effective approaches based on the theory that structurally similar molecules tend to have similar binding properties [244]. These similarity measures do not take into account information about activities of known binders of the target. G-protein-coupled target GPR30 specific
agonist that activates GPR30 was developed using similarity searches. The final similarity score that was used comprised a 2D score and a 3D structure similarity component [245-247].

Pharmacophore modeling
A pharmacophore is a molecular framework that defines the essential features responsible for the biological activity of a compound. When structural information about the drug target is limited or not known, pharmacophore models may be built using the structural characteristics of active ligands that bind to the target [248]. When 3D information of the target structure is known this binding site information can also be used in generating pharmacophore models [242]. Pharmacophore models that use chemical features such as acidic/basic residues and hydrogen bond acceptors and donors are found to be the most effective models [248]. Pharmacophore modeling has also been used in virtual screening of drugs in large databases [249]. There are programs developed to identify and generate pharmacophore models such as DISCO, GASP and Catalyst. It has been reported that GASP and Catalyst perform better than DISCO in reproducing the pharmacophore models [250]. One naturally occurring anti-cancer molecule identified using QSAR is I3C (indole-3-carbinol). However, this molecule has never gone past clinical trials due to its low potency. This active compound was optimized using ligand-based pharmacophore modeling to develop highly potent analog SR13668 which is a novel drug that shows to be highly potent against several cancer types [5]. Pharmacophore model construction steps can be summarized as follows:

1. The active compounds known to be binding to the desired target, that are also known to have the same interaction mechanism, are identified either by a literature search or a database search.
2. (a) For a 2D pharmacophore model essential atom types and their connectivity are defined (b) For a 3D pharmacophore model the conformations are defined using IUPAC nomenclature.
3. Ligand alignment or superimposition is used to find common features required in binders.
4. Pharmacophore model building.
5. Ranking of the pharmacophore models and selecting the best models.
6. Validation of pharmacophore models.

QSAR (quantitative structure–activity relationships)
QSAR methods are based on statistics that correlate activities of target drug interactions with various molecular descriptors. The basis of the QSAR method is the fact that structurally similar molecules tend to show similar biological activity [251]. These models describe mathematically how the activity response of a target, that binds a ligand, varies with the structural features of the ligand. QSAR is obtained by calculating the correlation between experimentally determined biological activity and various properties of small ligand binders [243]. QSAR relationships can be used to predict the activity of new drug molecule analogs.

In order to quantify the activity of a drug molecule, several values can be used. Half maximal inhibitory concentration \( (IC_{50}) \) and inhibition constant \( (K_i) \) are the most commonly used measures. QSAR models, unlike the pharmacophore models, can be used to find the positive or negative effect of a particular feature of a drug molecule to its activity. QSAR methods have been used successfully on various drug targets such as carbonic anhydrase [252,253], thrombin [254,255] and renin [256]. Different machine learning techniques have also been used in constructing QSAR models [257-259]. In classical or 2D QSAR methods, the biological activity is correlated to physical and chemical properties such as electronic hydrophobic and steric features of compounds [260]. In more advanced 3D QSAR methods, in addition to physical and geometric features of active drug molecules, quantum chemical features are also used. Recently QSAR models have also been developed for membrane systems [261].

The basic steps (Figure 8) of the QSAR method can be summarized as follows:

1. The active molecules that bind to the desired drug target and their activities are identified through a database search, a literature search, or HTS experiments.
2. Identification of structural or physicochemical molecular features (fingerprint) affecting biological activity (e.g. bond, atom, functional group counts, surface area etc.).
3. Building of a QSAR between the biological activity and the identified features of the drug molecules.
4. Validation of the QSAR biological activity predictive power.
5. Use of the QSAR model to optimize the known active compounds to maximize the biological activity.
6. The new optimized drug molecule activities are tested experimentally.

Success of a QSAR depends on the molecular descriptors selected and the ability of these models to predict biological activity. If there is not enough activity data to extract patterns, QSARs cannot perform well. Therefore, this method requires a certain minimum amount of training data in order to build a good predictive model and it is often linked to high-throughput screening. Statistical methods have been used in linear QSAR to pick molecular descriptors that are important in predicting the
Some of the drugs that are on the market with the help of ligand-based drug discovery are Zolmitriptan, Norfloxacin and Losartan [8]. Norfloxacin is a drug that is used in urinary tract infections and was developed using a QSAR model and approved by the FDA in 1986 [269]. Losartan [270] is used to treat hypertension and Zolmitriptan [271] is used as a treatment to migraine (Figure 9).

One difference between pharmacophore models and QSAR is that the pharmacophore model is constructed based on the necessary or essential features of an active ligand, whereas QSAR takes into account not only the essential features but also the features that affect the activity. One important structural feature used in both the pharmacophore model and in QSAR is the volume of the binding site. It is well established that the binding pocket volume has a big influence on the biological activity. In the cases where the binding pocket volume is known, elimination of molecules that are too large to fit in the binding pocket can be done in early stages of drug discovery process (see section “Binding pocket identification and volume calculation”).

Role of machine learning in LBDD

Machine learning algorithms can be trained to identify patterns in data and used to do predictions on test data sets. These algorithms are extensively applied in the field of biology and drug discovery [272-275]. Machine learning is used in many stages in the drug discovery pipeline including in the QSAR analysis stage [276]. Support vector machine (SVM) based algorithms are commonly used and have been shown to have high predictive power. SVM are often used for classification of sets of biological data. For example, they can be used to distinguish be-

Figure 9: A few drugs discovered with the help of ligand-based drug discovery tools. (a) Zolmitriptan: used as a treatment to migraine (b) Norfloxacin: used in urinary tract infections and (c) Losartan: used to treat hypertension.
between molecules that have high affinity for a target and those that have no affinity. Machine learning based scoring functions can also be used in structure-based drug discovery to predict target–ligand interactions and binding affinities [277]. Compared to conventional scoring functions, machine learning based scoring functions have often shown comparable or even improved performance. Moreover these algorithms can be trained to distinguish active drugs from decoys that do not have known drug activity [278]. Artificial neural networks (ANNs) have been used in drug discovery as a powerful predictive tool for non-linear systems [279]. For example, ANNs were used to construct the QSAR of a set of known aldose reductase inhibitors and biological activities of new molecules were predicted based on the QSAR [280]. Docking algorithms were then used to find novel inhibitors that bind to aldose reductase. ANN-based prediction models are also used in predicting biotoxicity of molecules as well [281].

**Conclusion**

In the past 10 years the identification rate of disease-associated targets has been higher than the therapeutics identification rate. With considerable rise in the number of drug targets, computational methods such as protein structure prediction methods, virtual high-throughput screening and docking methods have been used to accelerate the drug discovery process, and are routinely used in academia and in the pharmaceutical industry. These methods are well established and are now a valuable integral part of the drug discovery pipeline and have shown great promise and success. It is cheaper and faster to computationally predict and filter large molecular databases and to select the most promising molecules to be optimized. Only the molecules predicted to have the desired biological activity will be screened in vitro. This saves money and time because the risk of committing resources on possibly unsuccessful compounds that would otherwise be tested in vitro is reduced.

Structure-based and ligand-based virtual screening methods are popular with most of the applications being directed towards enzyme targets [282]. Even though structure-based methods are more frequently used, ligand-based methods have led to the discovery of an impressive number of potent drugs. In SBDD knowing the three-dimensional structure of the target of interest is required. However, in some cases it is not possible to determine structures of targets using conventional experimental methods due to experimental challenges. In the cases where experimental methods fail, computational methods become useful and potentially necessary for SBDD [23]. In the absence of an experimentally determined structure or a computationally generated model for a target of interest LBDD tools can be used. These tools require the knowledge of active drugs that bind to the target. LBDD tools such as 2D and 3D similarity searches, QSAR and pharmacophore modeling have proven successful in lead discovery.

Experimental methods usually represent proteins as static structures. However proteins are highly dynamic in character and protein dynamics play an important role in their functions. Computational modeling of the flexible nature of proteins is of great interest and various ensemble-based methods in structure-based drug discovery have emerged [178]. Molecular dynamics simulations are widely used in generating target ensembles that can be subsequently used in molecular docking [178]. Docking tools have been developed with different scoring functions and search algorithms. Comparative studies have been performed to evaluate these scoring functions and docking algorithms in docking pose selection and virtual screening [112,144,283]. There is no one superior tool that works for all target–ligand systems. The quality of docking results is highly dependent on the ligand and the binding site of interest [148-150].

VHTS methods are useful to screen large small molecule repositories fast and pick a smaller number of possible drug-like molecules for testing. By reducing the number of possible molecules that need to be tested experimentally, these methods can help to greatly cut the cost associated with drug discovery process. Studies have shown that with VHTS it is possible to identify molecules that are not observed with conventional high-throughput screening (HTS) experiments [284]. Thus VHTS methods are frequently used to complement HTS methods. The molecules selected by both of these methods are more likely to be possible drug candidates and should be considered when selecting hits.

Absorption, distribution, metabolism and elimination/excretion properties, commonly abbreviated ADME, as well as toxicity are important for the ultimate success or failure of a possible drug candidate. Adverse effects in animal models or even clinical trials can be reduced by filtering drug candidates by their ADME properties in early stages. Another important fact to consider in drug safety studies is how one drug can affect the metabolic stability of another drug [285]. Some drug–drug interactions (DDI) could lead to serious health effects; therefore, predicting these effects is important but challenging. The prescription antihistamine terfenadine and antifungal drug ketoconazole are two examples of drugs that should not be co-administered [286]. Terfenadine–ketoconazole drug–drug interactions results in cardiotoxicity. Computational methods such as pharmacokinetic modeling and predicting drug–drug interactions using large DDI interaction databases are successful and are both cost and time saving as well [287,288].
Currently hybrid structure-based and ligand-based methods are also gaining popularity. These combined (ligand-based and structure-based) drug discovery methods are of interest because they can amplify the advantages of both methods and improve the protocols [289-291]. One example is the hybrid docking protocol HybridDock, which incorporates both structure-based and ligand-based methods [291]. This hybrid method shows significantly improved performance in both binding pose and binding affinity prediction.

CADD has had a significant impact on the discovery of various therapeutics that are currently helping treat patients. Despite the successes, CADD also faces challenges such as accurate identification and prediction of ligand binding modes and affinities [8]. One of the challenging areas in drug discovery is the phenomenon of drug polymorphism [292]. Drug polymorphism occurs when a drug has different forms which are identical chemically but differ structurally. This can have a great impact on the success of a drug. Different polymorphic forms of a drug, which have different solid-state structures, can differ in solubility, stability and dissolution rates. Drug polymorphism can affect the bioavailability, efficacy and toxicity of a drug. One polymorphic form that is responsible for a particular drug effect may differ if a different polymorphic form of the same drug is administered. Using techniques such as spectroscopy it is possible to characterize drugs having different polymorphic forms.

Protein–protein interactions (PPIs) pose another challenge in drug discovery. PPIs are involved in many cellular processes and biological functions that are linked to diseases. Therefore, small molecule drugs that aim at PPIs are important in drug discovery [293]. It is of interest to develop therapeutics that can either disrupt or stabilize these interactions. However, it is challenging to design inhibitors that can directly interrupt PPIs. Common drug design usually targets a specific binding site on a protein of interest. However, protein–protein interacting surfaces have larger interfaces and are more exposed. Therefore their binding sites are often not well defined. Finding the sites that can be aimed at in PPI inhibition is therefore challenging and of great importance.

Through collaboration with the Drug Design Data Resource (D3R), which is a project funded by the National Institutes for Health, pharmaceutical companies are able to release their previously unreleased drug discovery data to the scientific community. This project allows the scientists all over the world to use new high-quality data in improving computer-aided drug discovery and design, and also to speed up the progress. The field of CADD is continuously evolving with improvements being made in each and every area. Some of the focus areas are scoring functions, search algorithms for molecular docking and virtual screening, optimization of hits, and assessment of ADME properties of possible drug candidates. With the current successes there is a promising future for computational methods to aid in the discovery of many more therapeutics in the future.

References

1. Tollman, P. A Revolution in R&D. How genomics and genetics are transforming the biopharmaceutical industry; 2001.


Beilstein J. Org. Chem. 2016, 12, 2694–2718. 2717
License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.12.267
A non-canonical peptide synthetase adenylates 3-methyl-2-oxovaleric acid for auriculamide biosynthesis

Daniel Braga¹,², Dirk Hoffmeister¹ and Markus Nett*³

Abstract
Auriculamide is the first natural product known from the predatory bacterium Herpetosiphon aurantiacus. It is composed of three unusual building blocks, including the non-proteinogenic amino acid 3-chloro-L-tyrosine, the α-hydroxy acid L-isoleucic acid, and a methylmalonyl-CoA-derived ethane unit. A candidate genetic locus for auriculamide biosynthesis was identified and encodes four enzymes. Among them, the non-canonical 199 kDa four-domain nonribosomal peptide synthetase, AulA, is extraordinary in that it features two consecutive adenylation domains. Here, we describe the functional characterization of the recombinantly produced AulA. The observed activation of 3-methyl-2-oxovaleric acid by the enzyme supports the hypothesis that it participates in the biosynthesis of auriculamide. An artificially truncated version of AulA that lacks the first adenylation domain activated this substrate like the full-length enzyme which shows that the first adenylation domain is dispensable. Additionally, we provide evidence that the enzyme tolerates structural variation of the substrate. α-Carbon substituents significantly affected the substrate turnover. While all tested aliphatic α-keto acids were accepted by the enzyme and minor differences in chain size and branches did not interfere with the enzymatic activity, molecules with methylene α-carbons led to low turnover. Such enzymatic plasticity is an important attribute to help in the perpetual search for novel molecules and to access a greater structural diversity by mutasynthesis.

Findings
Herpetosiphon aurantiacus is a filamentous, Gram-negative bacterium with a facultative saprophytic predatory behaviour [1,2]. For a more profound insight into the predation strategies among bacteria, along with the underlying chemistry, the complete genome of H. aurantiacus 114-95T (ATCC 23779, DSM 785) was sequenced and analysed [3]. Present as one circular
chromosome and two circular plasmids, the 6.8 Mb genome of *H. aurantiacus* encodes as many as 14 biosynthesis gene clusters corresponding to 6.6% (0.45 Mb) of the genome. This capacity highlights this microorganism as a promising source of natural products. Genes for nonribosomal peptide synthetases (NRPSs) were found to be preponderant, either solely or organised in combination with polyketide synthase (PKS) genes, representing four and five clusters, respectively. Two PKS and three putative bacteriocin gene clusters complete the total set involved in the biosynthesis of natural products. Contrasting the high number of biosyntheses deduced from genomic data, knowledge on the actual natural products is limited. Recently, the dipeptide auriculamide (1, Figure 1), and the diterpene *O*-methylkolavelool were observed in cultures of *H. aurantiacus* 114-95T, providing initial evidence for the assumed secondary metabolome of this species [4-6]. Within the entire genus, 1 is only the second PKS/NRPS-derived molecule to be described, following the report on siphonazole (2, Figure 1) [7].

Retrobiosynthetic analysis allowed the identification of a 14,130 bp-gene cluster, now referred to as *aul*-cluster (Figure 2), which putatively encodes two NRPSs (AulA and AulB) and one PKS (AulC) possessing domains that collectively allow and plausibly explains the assembly of 1. A gene for a type-II thioesterase is also found at the 3' portion of the *aul*-cluster that may help unload misacylated carrier protein domains [8,9].

Contrasting the standard layout of NRPSs, the amino acid sequence of one of the deduced NRPSs, termed AulA (1818 aa, 199 kDa), reveals the peculiar chimeric A₁-A₂-KR-PCP architecture (Figure 2) [10]. Of particular interest, the occurrence of two sequential adenylation (A) domains is a very rare feature and only preceded by PyrG from *Streptomyces pyridomyceticus* [11].

Since the lack of a genetic system for *H. aurantiacus* makes the use of reverse genetics prohibitive, we sought to provide biochemical evidence for the participation of this unusual NRPS in the biosynthesis of 1. AulA is suggested to incorporate L-isoleucic acid (= 2-hydroxy-3-methylvaleric acid). The domain architecture indicates the substrate undergoes no other chemical modification besides a reductive step after being tethered to the PCP domain by the PKS-type ketoreductase domain (KR), as reported for other natural products, such as pyrindomycin [11], cereulide, valinomycin [12], and bacillaene [13]. Hence, the molecule to be recognized and activated by AulA would be 3-methyl-2-oxovaleric acid (3).

Seminal work with gramicidin synthetase from *Bacillus brevis* led to the identification of ten positions within an A domain (PheA), collectively referred to as nonribosomal code [14], that control substrate selectivity. Further research started to establish a relationship between this code and structural require-
ments of the monomers to be recognised and incorporated to form the product [15,16]. In silico tools to identify the nonribosomal code, namely PKS/NRPS Analysis [17] and NRPSpredictor2 [18], are often accurate for the analysis of bacterial NRPSs. Yet, in our case, none retrieved any result after the analysis of AulA-A1. A subsequent manual inspection further revealed that the acyl-activating consensus motif is hardly conserved in AulA-A1. Moreover, the strictly invariant residue Asp413, which is essential for adenylate binding [14] was replaced by a tyrosine residue in AulA-A1. We hence concluded that this domain cannot function as an adenylating enzyme and is likely skipped during the biosynthetic assembly. Inspection of AulA-A2 with PKS/NRPS Analysis [17] and NRPSpredictor2 [18] yielded the nonribosomal code G-I-F-W-L-G-A-S-G-- (Table 1). Although the last position was not detected, evidences support its occupancy by a remarkably conserved lysine residue (K517 in PheA) [19], whose side chain counters the negative charge of the substrate’s carboxylic group [14,20]. Also, the relationship between the expected substrate and the nonribosomal code of AulA-A2 posed itself as a conundrum. The first position (D235 in PheA) is normally indicative of the substrate class to be used by the NRPS. Curiously, in AulA-A2 this corresponds to a glycine residue, associated with the activation of anthranilic acid [21,22] and diverts from what is often observed for the activation of aliphatic or aromatic α-keto acids, where the nonribosomal code starts with a valine residue [23]. In the face of this preliminary analysis it remained elusive if and how AulA-A1 would contribute to the biosynthetic auriculamide, e.g., through structural support for the catalytic role of AulA-A2, as noticed with a fungal A domain [24]. In order to evaluate the individual biosynthetic contribution of each A domain, we assembled two constructs to express aulA both as full-length gene and as an artificial open reading frame solely encoding the AulA-A2 domain, the α-keto reductase domain, and the terminal carrier protein. Independently, E. coli KRX was transformed with both constructs for heterologous expression. Our biochemical in vitro results highlight AulA as apt to take part on a NRPS/PKS complex for the biosynthesis of auriculamide, given the activation of 3-methyl-2-oxovaleric acid by its second adenylation domain. Our results also contribute to hone algorithms used to predict substrates from nonribosomal codes. Moreover, we describe how this enzyme is pliant to minor structural variations of that molecule, enabling future attempts to generate auriculamide analogues as potential new drug candidates.

To probe their enzymatic activity, the two purified AulA fusion proteins were subjected to the ATP-[32P]pyrophosphate exchange assay. In this assay, the protein is incubated with a potential substrate, ATP and radioactive pyrophosphate. The reversible back exchange of [32P]pyrophosphate into ATP is quantified by scintillation counting after solid phase capture of ATP on activated charcoal [25]. Both recombinant AulA variants tested against the assumed substrate 3 led to similar turnover (Figure 3a) which demonstrates that the A1 domain is not essential for adenylation of 3 and PCP loading. Further functional characterization was carried out using the native four-domain enzyme A11-A22-KR-PCP. Maximum turnover of 3 was observed at pH 7.0 and 30 °C. For more insight into the structural requirements of substrates, we assayed AulA against different molecules similar to 3, varying the functional group at the α-carbon, position and number of methyl substituents, and chain length (Figure 3b, compounds 4-10). As anticipated, the presence of an α-carbonyl notably influenced a successful adenylation. In the case of the tested α-keto acids, the differences in the chain size or position of the methyl group did not seem to play a role, as demonstrated by the equal enzymatic preference for 3 (261,000 cpm), 4 (249,000 cpm) and 5 (275,000 cpm). Conversely, the activation of α-hydroxy acids was not uniform. 2-Hydroxy-4-methylvaleric acid (6) could also be recognized by the NRPS, albeit a slightly lower radiolabel exchange (210,000 cpm) followed its incubation with the enzyme. Interestingly, the assay of 2-hydroxy-3-methylbutyric acid (7) resulted in a major decrease in the radiolabel exchange (60,000 cpm) when compared to its α-keto acid analogue. Molecules possessing a methylene α-carbon were not suitable substrates for AulA. 4-Methylpent-2-enoic acid (8) was only modestly activated (89,000 cpm), while reactions with 9 (24,000 cpm) and 10 (28,000 cpm) resulted in negligible substrate turnover.

Our biochemical in vitro results highlight AulA as apt to take part on a NRPS/PKS complex for the biosynthesis of auriculamide, given the activation of 3-methyl-2-oxovaleric acid by its second adenylation domain. Our results also contribute to hone algorithms used to predict substrates from nonribosomal codes. Moreover, we describe how this enzyme is pliant to minor structural variations of that molecule, enabling future attempts to generate auriculamide analogues as potential new drug candidates.

Table 1: Deduced nonribosomal code for H. aurantiacus AulA-A2 and the comparison with other α-keto acid activating NRPSs.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>GenBank Accession #</th>
<th>Nonribosomal code</th>
<th>Substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PksJ</td>
<td>P40806</td>
<td>V G M W N G A S V K</td>
<td>4-methyl-2-oxovaleric acid</td>
</tr>
<tr>
<td>CesA-A1</td>
<td>ABD14711</td>
<td>V G M W N G T S I K</td>
<td>4-methyl-2-oxovaleric acid</td>
</tr>
<tr>
<td>CesB-A1</td>
<td>ABD14712</td>
<td>V G M W N G V S V K</td>
<td>2-oxovaleric acid</td>
</tr>
<tr>
<td>PyrG-A2</td>
<td>AEFS3080</td>
<td>V G M T I G A S G K</td>
<td>3-methyl-2-oxovaleric acid</td>
</tr>
<tr>
<td>AulA-A2</td>
<td>ABX05055</td>
<td>G I F W L G A S G K</td>
<td>3-methyl-2-oxovaleric acid</td>
</tr>
</tbody>
</table>

Figure 3: Testing of AulA (A₁₋₂-KR-PCP) and a truncated variant (A₂-KR-PCP) in the ATP-[³²P]pyrophosphate exchange assay using 3-methyl-2-oxovaleric acid as a substrate. Relative activities are referenced to the A₂-KR-PCP enzyme (a). Substrate specificity of recombinant AulA in the ATP-[³²P]pyrophosphate exchange assay (b). All bar diagrams depict the substrate-dependent exchange based on the arithmetic mean of triplicate reactions. Error bars represent the standard deviations. Substrates with chiral centres were tested as racemic mixtures. PIPES buffer was used as negative control.
Supporting Information

Supporting Information File 1
Complete experimental details.
[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-12-274-S1.pdf]

Acknowledgements
D.B. gratefully acknowledges a doctoral fellowship from the International Leibniz Research School (ILRS Mibintact). We thank Wiebke Hanke and Dr. Hirokazu Kage for a preliminary analysis of AulA.

References

License and Terms
This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.12.274
Chemical probes for competitive profiling of the quorum sensing signal synthase PqsD of *Pseudomonas aeruginosa*

Michaela Prothiwa, Dávid Szamosvári, Sandra Glasmacher and Thomas Böttcher*

Full Research Paper

Address:
Department of Chemistry, Konstanz Research School Chemical Biology, University of Konstanz, 78457 Konstanz, Germany

Email:
Thomas Böttcher - thomas.boettcher@uni-konstanz.de

* Corresponding author

Keywords:
activity-based probes; PqsD; protein labelling; *Pseudomonas aeruginosa*; quinolones

Abstract

The human pathogen *Pseudomonas aeruginosa* uses the *pqs* quorum sensing system to coordinate the production of its broad spectrum of virulence factors to facilitate colonization and infection of its host. Hereby, the enzyme PqsD is a virulence related quorum sensing signal synthase that catalyzes the central step in the biosynthesis of the *Pseudomonas* quinolone signals HHQ and PQS. We developed a library of cysteine reactive chemical probes with an alkyne handle for fluorescence tagging and report the selective and highly sensitive in vitro labelling of the active site cysteine of this important enzyme. Interestingly, only one type of probe, with a reactive α-chloroacetamide was capable of covalently reacting with the active site. We demonstrated the potential of our probes in a competitive labelling platform where we screened a library of synthetic HHQ and PQS analogues with heteroatom replacements and found several inhibitors of probe binding that may represent promising scaffolds for the development of customized PqsD inhibitors as well as a chemical toolbox to investigate the activity and active site specificity of the enzyme.

Introduction

The emergence of multi-drug resistant bacterial strains urges the rapid discovery of new antibiotics and the development of novel antiinfective strategies [1]. One of the leading causes for nosocomial infections is the opportunistic human pathogen *Pseudomonas aeruginosa*, which, by chronic infections, also poses a major threat for cystic fibrosis patients [2,3]. *P. aeruginosa* deploys numerous virulence factors such as toxins, extracellular enzymes, and small molecule factors that are responsible for the bacterium’s ability to invade the host and cause a broad spectrum of different diseases [4,5]. The production of these virulence factors is coordinated on population level by several layers of hierarchically interconnected quorum sensing systems [6]. Quorum sensing signals are released from the cells and accumulate in a growing bacterial population to a certain threshold by which they start inducing the production of virulence factors. This simple signaling strategy thus regulates bacterial behaviour in dependence of population density. One of these quorum sensing systems, the *pqs* system, uses 2-alkyl-4-quinolones (AQs) as signals of which the *Pseudomonas* quinolone signal (PQS) and its biosynthetic precursor 2-heptyl-
4-quinolone (HHQ) are the two best studied AQs (Figure 1A) [7]. A variety of virulence factors are under control of the pqs quorum sensing system, including the production of elastase, pyocyanin, PA-IL lectin, and rhamnolipids, as well as populations dynamic behaviours such as biofilm formation. However, the exact roles of the different AQs are still not completely understood [6,8].

Besides HHQ and PQS, in total more than 50 structurally related AQs have been detected in P. aeruginosa [9]. Key to this large diversity of natural AQs are their common biosynthesis steps by enzymes encoded in the pqsABCDE operon [10]. The biosynthesis of AQs has been matter of a long-standing debate that could only recently be resolved. Although HHQ could be produced in vitro by a PqsD catalyzed “head-to-head” decarboxylative Claisen condensation of activated anthranilic acid with \( \beta \)-keto fatty acid derivatives [10,11], isotope labelled feeding experiments indicated an entirely different mechanism for its biosynthesis [12]. This mechanism has been elucidated step by step in recent efforts by the work of various research groups. Hereby, PqsA activates anthranilic acid to anthraniloyl-CoA which is transferred to PqsD which catalyzes the condensation with malonyl-CoA to form 2-aminobenzoylacetate (2-ABA) [13]. The PqsBC complex finally generates HHQ or other AQs in a decarboxylative condensation reaction of 2-ABA with fatty acids loaded on PqsC (Figure 1B) [14].

For the condensation step of an anthraniloyl residue with malonyl-CoA by PqsD, a cysteine residue (Cys112) is involved in the formation of a covalent thioester intermediate. We were speculating that activity-based electrophilic probes may be applicable to target this enzyme in vitro which could allow to study its active site reactivity in greater detail and apply a competitive labelling platform to discover potential PqsD inhibitors.

Results and Discussion
Electrophilic activity-based probes

The primary structure of PqsD comprises in total six cysteines. However, only one of them, Cys112, is engaging in the catalytic process forming a covalent reaction intermediate. We thus aimed at exploring the possibility to selectively label the active site cysteine residue using chemical probes.

Activity-based protein profiling (ABPP) has become a powerful tool to study protein function and elucidate targets of pro-
tein-reactive natural products in complex proteomes [15-18]. Various types of probes with an electrophilic core have been applied as tools for in vitro and in situ experiments of activity-based protein profiling [19-21]. ABPP uses probes with a reactive chemical group selectively targeting the active site of an enzyme and a reporter group that allows in-gel imaging and/or affinity enrichment of target enzymes [22].

We thus synthesized a small library of chemical probes with electrophilic $\alpha$-chloroacetamide, $\alpha,\beta$-unsaturated amide, and $\alpha,\beta$-unsaturated ketone moieties as protein reactive groups, which have been reported to exhibit selectivity for active site cysteines [19] (Supporting Information File 1, Figure S1). Each probe was equipped with a terminal alkyne handle for in-gel analysis by fluorescence tagging via click chemistry with a corresponding rhodamine azide. Variations of linker length and side group decorations between the reactive group and the alkyne handle were introduced to investigate potential differences in selectivity. Different alkyne amines were used to generate $\alpha,\beta$-unsaturated amide probes $\text{UA1–3}$ by reaction with acrylic acid chloride (Figure 2A) and $\alpha$-chloroacetamide probes $\text{CA1–3}$ by reaction with chloroacetyl chloride (Figure 2B). The $\alpha,\beta$-unsaturated ketone probe $\text{UK1}$ was synthesized via the Weinreb–Nahm amide in a Grignard reaction with vinylmagnesium bromide (Figure 2C). An overview of the small ABPP probe library is given in Figure 2D.

**Active site specific labelling of PqsD**

Next, we were interested to investigate if any of the ABPP probes was capable of labelling PqsD. Therefore, we cloned the $\text{pqsD}$ gene of *Pseudomonas aeruginosa* PAO1 into an expression vector encoding an N-terminal strep-tag. The protein was heterologously expressed in *Escherichia coli* BL21 followed by affinity purification by an ÄKTA chromatography system.

---

**Figure 2:** Synthesis of electrophilic ABPP probes. A) Synthesis of $\alpha,\beta$-unsaturated amide probes $\text{UA1–3}$. B) Synthesis of $\alpha$-chloroacetamide probes $\text{CA1–3}$. C) Synthesis of $\alpha,\beta$-unsaturated ketone $\text{UK1}$. D) Structures of the ABPP probe library.
equipped with a StrepTrap HP column. The individual probes were incubated with purified PqsD for 30 min and a rhodamine fluorescent reporter tag was appended by click chemistry. The remaining non-covalently bound probe and excess reporter tag were removed by SDS-polyacrylamide gel electrophoreses (SDS-PAGE), and labelling of the protein was visualized by fluorescence imaging. Consistency of protein levels was checked by coomassie staining (Supporting Information File 1, Figures S2 and S3). While the α,β-unsaturated amide probes UA1–3 and the α,β-unsaturated ketone UK1 only resulted in very weak or no labelling, all three α-chloroacetamide probes CA1–3 gave a strong fluorescent signal in the gel (Figure 3A).

In order to investigate the selectivity of the probes, we constructed a PqsD C112A mutant, where the active site cysteine was replaced by alanine. The purified mutant PqsD C112A exhibited only low background labeling for some probes but not comparable to labelling of the wild type protein by CA1–3, indicating that the probes were selectively targeting the active site (Figure 3A). Concentration series with a dose-down of the three CA probes showed that labelling of PqsD was concentration dependent and the two most potent probes CA1 and CA2 resulted in significant labelling at concentrations as low as 200 nM. Mass spectrometric analysis of a tryptic digest of CA2 labelled wild type PqsD resulted in an additional mass corresponding to a probe modified cysteine residue Cys112 confirming that the CA probes indeed covalently labelled the active site cysteine. Only one peptide was detected with another cysteine residue (Cys138) modified by the probe compared to 64 detected peptides for CA2 labelled Cys112 underlining the selectivity of our probes (Figure 3C). These results indicate that probes CA1–3 are specific and covalently bind to Cys112 of PqsD and are thus, to the best of our knowledge, the first account of activity-based probes targeting and selectively labelling the active site of PqsD.

Interestingly, variations in the probe structure had little impact on labeling intensity and specificity. Although each α-chloroacetamide probe had one closely related α,β-unsaturated acetamide counterpart, only the reactive group but not the structure of the probe or its side groups determined active-site labeling. These findings are surprising, as all three reactive groups are known to bind to cysteines which indicate a fine-tuned nucleophilic reactivity of the active site cysteine Cys112. The fine-tuned nucleophilicity towards our probes is supported by calculations of a mechanistic model where Cys112 is activated by deprotonation by His257 [23]. Our results may also partially explain the potent inhibition of a PqsD inhibitor described in the literature which was discovered in silico and had been equipped with an α-chloroacetyl group [24].

Inhibition of PqsD has been proposed as promising antivirulence strategy leading to disruption of AQ signaling and thereby to global down-regulation of virulence factor production [11]. Consequently, PqsD has become a highly attractive target and a great amount of work pioneered by the Hartmann group has resulted in inhibitor discovery using a combination of in vitro assay, in silico modelling and chemical lead optimization.
Examples of successful inhibitors are represented by the scaffolds of various 2-benzamidobenzoic acids [11,25,26], 2-nitrophenyl derivatives [27-29], ureidothiophene-2-carboxylic acids [24,30], and catechol-based compounds [31].

Many promising in vitro inhibitors based on these leads have been described and importantly, some of them also displayed in situ activity by reducing signal production and biofilm formation in live cultures of *P. aeruginosa* [28,31]. Recently, a synergistic dual PqsD and PqsR inhibitor was developed which also led to a marked decrease in the production of the virulence factors pyocyanine and pyoverdine [32].

So far only laborious enzyme-based assays, docking studies or modelling resulted in new scaffolds. We were thus interested, if our probes could be applied as a simple tool to discover novel scaffolds or chemical PqsD-binding motifs.

**Competitive screening approach**

In order to test if our probes could be used as a competitive labelling platform, we used the well-known unspecific cysteine reactive agent N-ethylmaleimide (NEM) and dosed NEM in increasing concentrations to wild-type PqsD before applying the probe CA2 followed by click chemistry with the fluorophore. A decreasing labelling intensity at increasing NEM concentrations indicates that NEM also blocks the active site cysteine Cys112 at concentrations above 5 µM and thereby prevents covalent attachment of the probe to PqsD (Figure 3D).

We were thus interested to investigate if we could demonstrate the value of our activity-based probes in a competitive screening approach by identifying potentially new scaffolds for PqsD inhibitors. We have recently reported the discovery of inhibitors of the virulence factor elastase of *P. aeruginosa* by a library of synthetic HHQ and PQS derivatives with systematic heteroatom replacements [33]. Because the interactions of PqsD with AQtS is still not entirely understood, we reasoned that HHQ or PQS analogues may be promising scaffolds for inhibitor development and we thus aimed to screen this library competitively against the active site specific probe CA2. Therefore, we further refined the library by the synthesis of two additional HHQ analogues and implemented improved synthetic strategies.

In detail, we synthesized 2-heptylquinolin-4(1H)-one (HHQ, 1) and 2-heptyl-3-hydroxyquinolin-4-one (PQS, 10) according to the described procedures by McGlacken et al. [34] and Hradil et al. [35], respectively. We previously described the synthesis of HHQ and PQS derivatives with nitrogen in position 1 exchanged by oxygen and sulfur [33]. In our efforts to optimize the synthesis of these heteroatom derivatives we used a one-pot reaction for the synthesis of 2-heptyl-chromen-4-one (1-O-HHQ, 4) which includes esterification, Baker–Venkataraman rearrangement and subsequent acid-catalyzed ring closure to afford the 1-O-HHQ in 60% yield [36] (Scheme 1).

The 2-heptyl-3-hydroxychromen-4-one (1-O-PQS, 13) was previously synthesized from the chroman-4-one 8 which was produced by base-catalyzed Knoevenagel reaction from 2-hydroxyacetophenone with octanal. Although the starting material was readily available, the reaction gave the product only in low yield (20–30%) and separation of the starting material from the product could be difficult especially for multigram scale approaches. Since 1-O-HHQ (4) was now easily available, we used 4 as starting point for the 1-O-PQS (13) synthesis by two different approaches. First, we tried to synthesize chroman-4-one 8 by hydrogenation of 4. We found that ammonium formate (NH₄HCO₂) as mild hydrogen source with Pd/C gave chroman-4-one 8 in a clean reaction with a good yield of 76% whereas the direct use of H₂ with Pd/C gave mainly the fully reduced 2-heptylchromane. Compound 8 can be applied for the synthesis of 1-O-PQS (13) as described in [33]. A new, direct way in which 1-O-HHQ (4) can be used for the synthesis of 1-O-PQS (13) was explored by epoxidation with subsequent ring opening in 41%. Thus, 1-O-PQS (13) could be produced in just two steps with an overall yield of 25% (Scheme 1). The synthesis of 1-S-PQS (15) was previously accomplished by a two-step synthesis of thiocroman-4-one 9 and following oxidation to give 1-S-PQS (15) in 12% overall yield [33]. In our attempts to synthesize 9 more efficiently, we used a method described by Olah et al. [37] starting from commercially available (E)-deca-2-enolic acid and thiophenol which gave 9 in 53% yield without the use of microwave assistance (Scheme 1). Thionation of the 4-position of 1, 4, 6, 10, 13 and 15 using P₄S₁₀ in pyridine under reflux conditions gave the 4-thiones 2, 5, 7, 11, 14 and 16 in yields between 60–80%, respectively (Scheme 1). The HHQ-oxide (3) was synthesized from HHQ (1) by conversion in the benzyl-protected chinolinol form (3a) and oximation with hydroxylamine hydrochloride similar to the described method used for the synthesis of the PQS-oxide 12 [33]. The entire compound library of HHQ and PQS analogues is presented in Figure 4. Further details on the syntheses are given in the Supporting Information File 1.

All compounds were screened at an initial concentration of 240 µM in a competitive experiment against probe CA2. With the compounds added as DMSO stocks, solubility of the compounds was not an issue at these concentrations. PqsD was hereby pre-incubated with the compounds for 30 min, followed by the addition of the probe. A compound interacting tightly with the active site would hinder the probe from binding to the active site. Thus, a reduced labelling intensity in a competitive screening approach by identifying potentially new scaffolds for PqsD inhibitors was developed which also led to a marked decrease in the production of the virulence factors pyocyanine and pyoverdine [32].

So far only laborious enzyme-based assays, docking studies or modelling resulted in new scaffolds. We were thus interested, if our probes could be applied as a simple tool to discover novel scaffolds or chemical PqsD-binding motifs.

**Competitive screening approach**

In order to test if our probes could be used as a competitive labelling platform, we used the well-known unspecific cysteine reactive agent N-ethylmaleimide (NEM) and dosed NEM in increasing concentrations to wild-type PqsD before applying the probe CA2 followed by click chemistry with the fluorophore. A decreasing labelling intensity at increasing NEM concentrations indicates that NEM also blocks the active site cysteine Cys112 at concentrations above 5 µM and thereby prevents covalent attachment of the probe to PqsD (Figure 3D).

We were thus interested to investigate if we could demonstrate the value of our activity-based probes in a competitive screening approach by identifying potentially new scaffolds for PqsD inhibitors. We have recently reported the discovery of inhibitors of the virulence factor elastase of *P. aeruginosa* by a library of synthetic HHQ and PQS derivatives with systematic heteroatom replacements [33]. Because the interactions of PqsD with AQtS is still not entirely understood, we reasoned that HHQ or PQS analogues may be promising scaffolds for inhibitor development and we thus aimed to screen this library competitively against the active site specific probe CA2. Therefore, we further refined the library by the synthesis of two additional HHQ analogues and implemented improved synthetic strategies.

In detail, we synthesized 2-heptylquinolin-4(1H)-one (HHQ, 1) and 2-heptyl-3-hydroxyquinolin-4-one (PQS, 10) according to the described procedures by McGlacken et al. [34] and Hradil et al. [35], respectively. We previously described the synthesis of HHQ and PQS derivatives with nitrogen in position 1 exchanged by oxygen and sulfur [33]. In our efforts to optimize the synthesis of these heteroatom derivatives we used a one-pot reaction for the synthesis of 2-heptyl-chromen-4-one (1-O-HHQ, 4) which includes esterification, Baker–Venkataraman rearrangement and subsequent acid-catalyzed ring closure to afford the 1-O-HHQ in 60% yield [36] (Scheme 1).
screening experiment indicates a potential hit compound (Figure 5A).

Strikingly, half of the derivatives were significantly active in the competitive screening and abolished probe labelling at 240 µM, while the other compounds had no such effect at this initial concentration (Figure 5B). Active compounds were found in pairs of HHQ and their corresponding PQS derivatives and either comprised a 4-thionated HHQ (2, 5, 7) or PQS (11, 14, 16) scaffold or an oxime group in position 4 (3 and 12).

The eight active compounds were tested in a concentration-dependent experiment in order to assess their potency in inhibiting probe binding. Interestingly, the HHQ derivative 3 with an oxime group was significantly more active than its PQS counterpart 12, with the lowest activity. In contrast, the 4-thionated PQS derivatives 11, 14 and 16 were always more active than their corresponding HHQ analogues (2, 5, and 7), and efficiently blocked probe labelling already around 24 µM. These results indicate that the 3-OH group was important for the activity of the 4-thionated compounds. In order to exclude any adverse effects of the compounds on the click chemistry, we performed control experiments where the most active derivatives were added directly before the last step of the click protocol. Intense labelling in this control group indicated that click chemistry was not affected by the compounds (Figure 5D).

To assess the stability of the probe, we incubated CA2 with two of the most active compounds, 11 and 14. NMR and MS data indicate that the probe was not chemically modified even after 18 and 24 h so that the potential inactivation of the probe by the compound scaffold during protein labelling could be ruled out (Supporting Information File 1, Figures S4 and S5).

Our 4S-PQS analogues thus represent a promising novel scaffold that inhibits the labelling of PqsD by an active-site-directed probe in the lower micromolar range. We have previously described these compounds (11, 14, and 16) as potent inhibitors of the virulence factor elastase (LasB) of Pseudomonas aeruginosa [33]. However, the mechanism of inhibition was by direct binding to the active site of elastase. While elastolytic activity was completely inhibited with 11 even in situ we could not confirm any significantly large inhibition of rhamnolipid or pyocyanin production. Nevertheless, our new compounds may be useful scaffolds for the future development of a novel generation of PqsD inhibitors.

**Conclusion**

Electrophilic probes represent powerful tools for investigating protein reactivity and discovering customized enzyme inhibitors. We discovered α-chloroacetamide probes selectively labelling the active site cysteine residue of the Pseudomonas aeruginosa quorum sensing signal synthase PqsD. While these
findings may guide the future development of covalent PqsD inhibitors, we could also demonstrate the value of the probes as tools for investigating the reactivity of PqsD and apply them in a competitive screening approach. These led to the novel class of 4S-PQS analogues as potent in vitro inhibitors of the active-site labelling of PqsD. In combination, our probes and their inhibitors represent a valuable toolkit for investigating this important virulence-related enzyme.
Figure 5: Competitive profiling platform. A) Schematic representation of the competitive labelling strategy with an alkyne probe (green) and potential inhibitors (blue). Rh = rhodamine. B) Initial screening of our small library of HHQ (1–9) and PQS (10–16) analogues against the active site specific chemical probe CA2. C) Concentration dependent competition experiment where the probe concentration is held constant and PqsD is pre-treated with varying inhibitor concentrations. D) Click chemistry control where the compounds were added shortly before CuII-salt addition in the click protocol. Cont.: DMSO control.

Supporting Information
Supporting Information File 1
Syntheses, and full compound characterization, experimental methods, and probe labelling. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-12-277-S1.pdf]

Acknowledgements
We thank Prof. Andreas Marx and his group for their generous support. We gratefully acknowledge funding by the Emmy Noether program of the Deutsche Forschungsgemeinschaft (DFG), EU FP7 Marie Curie Zukunftskolleg Incoming Fellowship Program – University of Konstanz grant no. 291784, Fonds der Chemischen Industrie (FCI), Konstanz Research School of Chemical Biology (KoRS-CB), and CRC969 (DFG). MP was supported by a Carl Zeiss Ph.D. fellowship and DS received a Ph.D. fellowship from KoRS-CB. We thank Jing Qiao for helping with protein purification and gel-based labelling experiments and Martin Mex and Daniel Hammler for help with mass spectrometry. We also acknowledge the help of the Proteomics-facility and NMR core facility of the University of Konstanz.

References
2794–2801. doi:10.1021/cb400530d


Negri, M.; Plaza, A.; Steinbach, A.; Müller, R.; Hartmann, R. W.

Böttcher, T.; Pitscheider, M.; Sieber, S. A. doi:10.1038/nchembio707G


2012, J. Heterocycl. Chem.
141–144. doi:10.1002/jhet.5570360121


Weerapana, E.; Simon, G. M.; Cravatt, B. F. Nat. Methods 2009, 6, 405–407. doi:10.1038/nmeth.91

Dubern, J.-F.; Diggle, S. P. doi:10.1073/pnas.96.20.11229


2012, MedChemComm


Hartmann, R. W.; Empting, M.; Steinbach, A.; Hartmann, R. W. doi:10.1016/j.ejmech.2014.02.014

Hartmann, R. W. doi:10.1016/j.ejmech.2014.11.055


2012, J. Heterocycl. Chem.


2013, Protein Cell
Biochemical and structural characterisation of the second oxidative crosslinking step during the biosynthesis of the glycopeptide antibiotic A47934

Veronika Ulrich¹, Clara Brieke¹ and Max J. Cryle*¹,²,³

Full Research Paper

Address:
¹Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany,
²EMBL Australia, Monash University, Clayton, Victoria 3800, Australia
³The Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology and ARC Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia

Email:
Max J. Cryle* - max.cryle@monash.edu

* Corresponding author

Keywords:
crystal structure; cytochrome P450; glycopeptide antibiotic; peptide; phenolic coupling

Introduction

The glycopeptide antibiotics (GPAs) are a series of highly modified heptapeptide natural products and are highly effective antibiotics against Gram-positive bacteria, where they affect their function by preventing the correct crosslinking of the peptidoglycan cell wall [1]. Produced by bacteria, these compounds derive their efficacy from their unique three-dimen-
sional structure, which in turn enables them to bind to the dipeptide terminus of the peptidoglycan precursor lipid II [1,2]. This three-dimensional structure is generated by the high degree of crosslinking exhibited by the glycopeptide antibiotics: in the case of the two most widely known natural examples (vancomycin and teicoplanin) this includes three and four crosslinks, respectively, which occur between the side chains of aromatic residues [3] within the parent heptapeptide (Figure 1b) [4]. This degree of crosslinking in turn renders the total synthesis of GPAs as unfeasible for production and hence both first and second generation GPAs in clinical use are all entirely derived from in vivo biosynthesis [1,2].

The biosynthesis of GPAs is based around the initial synthesis of the linear heptapeptide by a type-I non-ribosomal peptide synthetase (NRPS) [5,6] and its subsequent modification by cytochrome P450 monoxygenases [7-9], which install the crosslinks that provide the unique structure and hence activity of the GPAs (Figure 1a) [4]. Later diversification of the completely crosslinked peptide aglycones is the major source of diversity in natural GPAs, and occurs against the completed peptide aglycones [1,10,11]. The installation of the crosslinks has received significant attention both in vitro [12-26] and in vivo [27-32] techniques, largely due to the synthetic challenge that these modifications represent. In vivo studies initially confirmed that the cytochrome P450s, known as the Oxy enzymes, are each responsible for the installation of a single ring in the GPA aglycones and that there is a conserved order of activity in both type-I and type-IV GPAs. In type-I GPA biosynthesis OxyB acts first to install the C-O-D ring (between residues 4/6), followed by D-O-E ring installation (between residues 2/4) catalysed by OxyA and finally formation of the AB ring (between residues 5/7), catalysed by OxyC [28,30-32]. In type-IV systems, where there is an extra ring present between residues 1 and 3 (the F-O-G ring), this is installed by OxyE, which acts between OxyB and OxyA in the cyclisation cascade [27]. In vivo experiments also hinted towards the activity of the Oxy enzymes against the substrate peptides whilst

![Figure 1](image-url)

**Figure 1:** (a) Schematic representation of the biosynthesis of A47934 by the heterotrimeric non-ribosomal peptide synthetase; the 7 modules of the A47934 NRPS machinery are distributed over 4 proteins (StaA to StaD) and exhibit the typical NRPS domain architecture with adenylation (A; purple), peptidyl carrier protein with phosphopantetheine linker (PCP; green), condensation (C; red), epimerisation (E; dark blue), P450-recruitment (X; blue) and thioesterase (TE; light grey) domain; the peptide is shown at its distinct stages of biosynthesis; the amino acid cyclisation steps are depicted with arrows and the corresponding Oxy enzyme; (b) the structures of the glycopeptide antibiotics A47934, teicoplanin and vancomycin, with cross-links highlighted (blue) and standard ring nomenclature shown for A47934 (magenta).
they remain bound to the NRPS [29], and in vitro experiments performed with OxyB from the vancomycin biosynthesis pathway confirmed that the Oxy enzymes do indeed act against peptides when these are bound to peptidyl carrier protein (PCP) domains [26]. More recently, it has been shown that the activity of the Oxy enzymes is actually reliant upon an additional conserved domain present within the final module of GPA NRPS machineries, known as the X-domain [16]. Characterisation of this domain has shown that it is a modified, catalytically inactive condensation-type domain and that this domain is capable of forming 1:1 complexes with the Oxy enzymes from GPA biosynthesis [16]. More importantly, with the exception of OxyBvan, the activity of Oxy enzymes in vitro has also been shown to be highly dependent on the presence of the X-domain fused to the peptidyl carrier protein domain [16]. This in turn has, for the first time, allowed the characterisation of the second cyclisation step, catalysed by OxyA, from the teicoplanin system [13,16,17]. These results showed that OxyA, in contrast to OxyB, is highly selective for the correct stereochemistry of the peptide C-terminal residue and generally displays a higher selectivity for the structure of the substrate peptides [13,17].

Recent in vitro studies performed with the teicoplanin-related A47934 (sta) GPA biosynthetic machinery from *Streptomyces toyocaensis* [33] (Figure 1a) have revealed that the X-domain is in fact far from an innocent bystander during peptide oxidation and that switching this domain to other homologues can affect the selectivity of the Oxy enzymes for their peptide substrates [12]. Combined with the fact that only a single OxyA enzyme has been successfully characterised to date [13,14,16,17], we resolved to make a detailed structural and biochemical analysis of the OxyA homologue from the A47934 system, named StaF, to investigate not only some of the mechanistic features of the OxyA reaction but also the role of the X-domain on the activity of this enzyme and to determine whether the recruitment domain can also affect peptide selectivity for later Oxy enzymes in the GPA cyclisation cascade.

**Results and Discussion**

**Spectral analysis of StaF**

Spectral analysis of P450s allows determination of their potential catalytic competence. The UV–visible spectrum of StaF exhibited a Soret maximum at $\lambda = 421$ nm and $\beta/\alpha$ bands at $\lambda = 539$ and 566 nm, respectively (Figure 2a). This corresponds to the absorption spectra characteristic for P450s in the low-spin state, indicating the heme moiety of StaF to be present in its water-bound ferric form. Equivalent spectra were observed for related P450s such as StaH, OxyA*tei*, OxyB*tei* and OxyE*tei*, which are also involved in GPA cyclisation reactions [12,14,19,22]. Reduction of StaF by addition of sodium dithionite led to conversion of ferric to ferrous heme, which was accompanied by shift of the Soret maximum to $\lambda = 422$ nm and of the $\beta/\alpha$ bands to $\lambda = 532$ and 559 nm in the UV–visible spectrum. Upon saturation of ferrous StaF with carbon monoxide, two major peaks appeared at $\lambda = 420$ and 450 nm, respectively, as well as a broad minor peak at $\lambda = 548$ nm (Figure 2a). The peaks at $\lambda = 420$ and 450 nm are caused by different protonation states of the thiol side chain of the proximal heme ligand cysteine: P450 enzymes displaying a protonated thiol ligand (as indicated by a peak at $\lambda = 420$ nm) are catalytically inactive, whilst a catalytically competent P450 enzyme with a thiolate-ligated heme exhibits the signature $\lambda = 450$ nm absorption peak [34]. The fact that peaks at both $\lambda = 420$ and 450 nm appear in the spectrum of StaF indicates that this P450 is present in both incompetent as well as competent states. It has been shown that...
the inactive form can convert into an active species upon substrate binding [35], however the true catalytic competence of StaF was subsequently determined by substrate turnover assays.

**Interaction analysis of StaF with the A47934 X-domain**

StaF, as a member of the group of P450s involved in GPA cyclisation, is anticipated to be recruited to the NRPS machinery through interaction with the X-domain being present in the final NRPS module as has been demonstrated for other Oxy homologues [12,13,16]. In order to determine if StaF is also recruited by the A47934 X-domain (X$_{sta}$) we analysed their interaction by analytical SEC. This method is suitable for interaction analysis with P450s as interaction partner, as not only the typical protein absorption at $\lambda = 280$ nm, but also the heme-specific absorption at $\lambda = 415$ nm can be monitored. The X$_{sta}$ construct has previously been shown to form a tight interaction with StaH, the P450 responsible for the first (C-O-D) phenolic coupling reaction in A47934 biosynthesis [12].

Prior to analysis by SEC, a mixture of StaF and a 3-fold excess of X$_{sta}$ as well as each individual protein was incubated in appropriate buffer (50 mM Tris pH 7.4 and 150 mM NaCl) at room temperature for 30 min to allow complex formation to occur between StaF and X$_{sta}$. Analysis of StaF alone (MW of 47.3 kDa) resulted in overlapping peaks with absorption at $\lambda = 280$ and 415 nm at an elution volume of 13.1 mL, whereas when X$_{sta}$ alone (MW of 53.2 kDa) was analysed a $\lambda = 280$ nm peak at an elution volume of 12.6 mL was observed. The mixture of StaF with X$_{sta}$ led to the appearance of overlapping peaks with absorption at both $\lambda = 280$ and 415 nm at an earlier elution volume of 12.1 mL (Figure 2b), which indicates that the heme-specific $\lambda = 415$ nm absorption peak has shifted to an earlier elution volume and that can be explained through formation of a complex between StaF and X$_{sta}$. Thus, we conclude that StaF is, in addition to StaH [12], also recruited to the NRPS machinery through interaction with X$_{sta}$ [16]. The fact that the StaF and X$_{sta}$ mixture shows a single peak upon gel filtration analysis argues for the StaF and X$_{sta}$ molecules being in constant exchange. This is significantly different to the interaction behaviour of StaH and X$_{sta}$, where the interaction of StaH to X$_{sta}$ was strong enough to result in two X$_{sta}$ populations, one bound to StaH and the other free in solution [12].

Studies on the biosynthesis of teicoplanin and the vancomycin-type chloroeremomycin GPA showed decreasing affinity of the P450s to the X-domain with later positions in the GPA cyclisation cascade [13,16], and our results from the A47934 system would appear to follow these trends.

**Reconstitution of in vitro StaF activity**

On the basis of StaF being a catalytic competent P450 and its interaction with X$_{sta}$, we attempted to reconstitute the activity of this enzyme (Figure 3). In the activity assay we initially employed a teicoplanin-like heptapeptide exhibiting L-Hpg (hydroxyphenylglycine) instead of L-Dpg (3,5-dihydroxyphenylglycine) at position 3 and 7 (abbreviated as Tei7-L-Hpg; Figure 4), which served as suitable substrate as A47934 and teicoplanin exhibit the same amino acid composition of their parent peptide [13,15-18,36]. The linear Tei7-L-Hpg peptide as well as the mono- and bicyclic products based on P450-catalysed turnover have been analysed in earlier studies [13,16,17]. Prior to the activity assay the substrate was loaded onto the A47934 PCP-X di-domain construct exhibiting maltose binding protein as N-terminal fusion partner (MBP-PCP-X$_{sta}$) using the R4-4 mutant of the promiscuous phosphopantetheinyl transferase Sfp [37]. Subsequently, triplicate turnover assays of StaF both including and excluding StaH were performed using the redox system composed of palustrisredoxin B A105V/palustrisredoxin reductase/NADH to ensure electron supply to the P450s (Figure 3) [38]. NADH was additionally regenerated throughout the assay via a glucose/glucose oxidase couple. The assay was stopped by cleaving the peptide from the PCP-X constructs using excess of methyamine and the peptide was then purified by solid phase extraction before being subjected to HPLC–MS analysis [15,17].

The StaF activity was first investigated using a linear Tei7-L-Hpg peptide loaded onto the A47934 PCP-X di-domain construct both in the absence and presence of StaH. Only linear peptide was detected in the samples lacking StaH, which is in line with previous in vitro and in vivo experiments that indicate that the presence of the C-O-D ring is a prerequisite for the activity of subsequent P450 enzymes, such as StaF (Figure 4, Table 1 entry 1) [13,16,22,27-29,32]. Both mono- and bicyclic peptide products could be detected in samples with StaH included in the turnover assay: given that we have demonstrated that StaH is capable of producing a C-O-D ring containing peptide from a linear precursor [12] and the lack of StaF activity against linear peptide substrates, we conclude that the formation of the bicyclic peptide is due to the activity of both StaH and StaF (first by StaH installing the C-O-D ring and then subsequent formation of the D-O-E ring by StaF, Figure 4). The level of activity observed for StaF was lower than for OxyA$_{tei}$ which is most likely explained by the fact that a significant proportion of StaF was isolated with a protonated – and hence catalytically inactive – heme thiolate ligand. Furthermore, it is also possible that product inhibition could also playing a role in reducing substrate turnover in this system.

**Characterising the substrate specificity of StaF**

After showing that StaF installs the D-O-E crosslink between amino acids D-Tyr$_2$ and D-Hpg$_4$ on the Tei7-L-Hpg$_7$ peptide, we were interested in probing the substrate specificity of StaF
Figure 3: Complete workflow for the Cytochrome P450 activity assay used in this study. 1) Loading of the substrate (Tei7-L/D-Hpg\(_7\) is depicted) onto a conserved serine residue of the PCP-domain using the R4-4 mutant of the promiscuous phosphopantetheinyl transferase Sfp. The substrate peptide is attached to the PCP-domain via a coenzyme A-derived phosphopantetheine moiety. Excess of substrate is removed via centrifugation using centrifugal filter units with an appropriate MWCO. 2) Subsequently, the activity assay is performed using StaH and StaF together with the redox system composed of palustrisredoxin reductase (PuR), palustrisredoxin B A105V (PuxB) and NADH, in which StaH catalyses C-O-D ring formation between D-Hpg\(_4\) and L-Tyr\(_6\) and StaF catalyses ring D-O-E ring formation between D-Tyr\(_2\) and D-Hpg\(_4\). 3) The reaction is quenched by the addition of methylamine, which cleaves off the peptide from the phosphopantetheine linker thus liberating the peptide methylamide. 4) The peptide is purified by solid phase extraction (SPE) and 5) analysed by LC–MS in single ion monitoring (SIM) mode.

Figure 4: (a) StaF activity against different peptide substrates and using NRPS constructs; the activity of StaF with Tei7-L-Hpg\(_7\) (magenta), Tei7-D-Hpg\(_7\) (green), Pek7-rac-Hpg\(_7\) (orange) and Act7-rac-Hpg\(_7\) (blue) were determined; all peptides were bound to wildtype and hybrid PCP-X constructs derived from the A47934 (magenta) and teicoplanin NRPS (blue); yield was calculated based on the integrated peak area of bicyclic peptide divided by the sum of the integrated peak areas of monocyclic and bicyclic peptide observed by HPLC–MS (SIM) and is depicted in %; the calculation is based on turnover assay triplicates and the standard deviation is shown. (b) Structures of the peptides used as substrates for StaF, being Tei7-L/D-Hpg\(_7\) (1), Pek7-rac-Hpg\(_7\) (2) and Act7-rac-Hpg\(_7\) (3); R = CoA or methylamine.
Table 1: StaF turnover activity.

<table>
<thead>
<tr>
<th>Entry</th>
<th>fus-PCP-X&lt;sup&gt;a&lt;/sup&gt;</th>
<th>peptide</th>
<th>StaF activity&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MBP sta sta</td>
<td>Tei7-L-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>23.8% ± 0.3%</td>
</tr>
<tr>
<td>2</td>
<td>MBP sta sta</td>
<td>Tei7-D-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2.3% ± 0.5%</td>
</tr>
<tr>
<td>3</td>
<td>MBP sta sta</td>
<td>Pek7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2.3% ± 0.2%</td>
</tr>
<tr>
<td>4</td>
<td>MBP sta sta</td>
<td>Act7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>25.5% ± 2.2%</td>
</tr>
<tr>
<td>5</td>
<td>MBP tei tei</td>
<td>Tei7-L-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>5.7% ± 0.4%</td>
</tr>
<tr>
<td>6</td>
<td>MBP tei tei</td>
<td>Tei7-D-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2.3% ± 0.6%</td>
</tr>
<tr>
<td>7</td>
<td>MBP tei tei</td>
<td>Pek7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>1.2% ± 0.2%</td>
</tr>
<tr>
<td>8</td>
<td>MBP tei tei</td>
<td>Act7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>24.9% ± 1.5%</td>
</tr>
<tr>
<td>9</td>
<td>MBP sta tei</td>
<td>Tei7-L-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2.9% ± 0.5%</td>
</tr>
<tr>
<td>10</td>
<td>MBP sta tei</td>
<td>Tei7-D-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>1.2% ± 0.5%</td>
</tr>
<tr>
<td>11</td>
<td>MBP sta tei</td>
<td>Pek7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>0.3% ± 0.1%</td>
</tr>
<tr>
<td>12</td>
<td>MBP sta tei</td>
<td>Act7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>7.8% ± 0.9%</td>
</tr>
<tr>
<td>13</td>
<td>MBP sta tei</td>
<td>Tei7-L-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>21.1% ± 0.6%</td>
</tr>
<tr>
<td>14</td>
<td>MBP sta tei</td>
<td>Tei7-D-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>5.4% ± 0.3%</td>
</tr>
<tr>
<td>15</td>
<td>MBP tei sta</td>
<td>Pek7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>2.6% ± 0.5%</td>
</tr>
<tr>
<td>16</td>
<td>MBP tei sta</td>
<td>Act7-rac-Hpg&lt;sub&gt;7&lt;/sub&gt;</td>
<td>31.5% ± 2.5%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Spheres correspond to the N-terminal fusion partner (abbreviated as fus; MBP; shown in grey). PCP- (middle sphere) and X-domain (C-terminal sphere). PCP-/X-domains from A47934 NRPS are shown in red, PCP-/X-domains from teicoplanin NRPS are shown in blue. <sup>b</sup>Effective StaF activity = integrated peak areas of bicyclic product/sum of integrated peak areas of mono- and bicyclic product observed by HPLC–MS (single ion monitoring). Mean activity and standard deviation were calculated based on turnover assay triplicates (shown in %).

and hence we analysed StaF activity on different substrates bound to MBP-PCP-X<sub>sta</sub> (Figure 4, Table 1, entries 2–4). We found that StaF activity was dramatically reduced with a teicoplanin-like heptapeptide exhibiting the 7<sup>th</sup> amino acid in the unnatural D-configuration (Tei7-D-Hpg<sub>7</sub>, Figure 4). This indicates that the incorrect stereochemistry of the C-terminal amino acid residue hinders cyclisation of amino acids 2 and 4 in spite of these being localised towards the N-terminus of the
peptide, and mimics the behaviour observed for the only other OxyA homologue characterised to date, OxyA\textsubscript{tei} [13]. This behaviour is in contrast to that of StaH and other OxyB homologues, which exhibit similar activity on both Tei7-L-Hpg\textsubscript{7} and Tei7-D-Hpg\textsubscript{7} peptides [12,13,16,17]; StaH even shows a preference for the incorrect peptide diastereomer under specific conditions [12]. These results provide hints of more stringent substrate specificity at later stages of the GPA cyclisation cascade and we hence investigated StaF activity against altered peptide substrates, including pekiskomycin- (Pek) and actinomycin-like (Act) heptapeptides. These peptides differ to A47934 and teicoplanin in the amino acid residues present in positions 1 and 3 of the peptide (Pek: 1, 3; Act: 3; Figure 4) [1,17,39]. Both Pek- and Act-heptapeptides exhibit a Hpg residue at position 7 instead of L-Dpg, but with a racemic mixture of L- and D-Hpg\textsubscript{7} (Pek7-rac-Hpg\textsubscript{7}, Act7-rac-Hpg\textsubscript{7}) due to an inability to resolve the diastereomers via HPLC. We observed very little activity of StaF against Pek7-rac-Hpg\textsubscript{7} (Figure 4, Table 1, entry 3), which is similar to the results observed for OxyA\textsubscript{tei}. This could be explained by the significant differences in the structures of the amino acids at positions 1 and 3 of the peptide, given that these are in the direct locale of the residues involved in the D-O-E ring [17]. In contrast, StaF (following StaH-catalysed monomycyclisation of the linear peptide) showed similar activity against Act7-rac-Hpg\textsubscript{7} (Figure 4, Table 1, entry 4) compared to Tei7-L-Hpg\textsubscript{7}, which indicates a preference for hydrophobic amino acids with bulky side chains. This is clearly different to the broad substrate specificity shown by the preceding enzyme StaH, which accepts peptides including Tei7-L/D-Hpg\textsubscript{7}, Pek7-rac-Hpg\textsubscript{7} as well as Act7-rac-Hpg\textsubscript{7} as shown here [12]. Similar results were obtained for OxyB and OxyA from the teicoplanin system and it appears that the substrate specificity of the P450s is decreased when acting on later steps of the GPA cyclisation cascade [17]. This also makes the identification of an active OxyA homologue from a type-I GPA producer (vancomycin/pekiskomycin type) [1] of great importance to test the substrate selectivity of these homologues against altered peptide substrates.

Impact of the A47934 X-domain on StaF activity

Previously, it has been shown that StaF exhibits high activity against peptide substrates presented by the PCP-X-di-domain from teicoplanin biosynthesis, whilst low activity was achieved on PCP-X constructs from the A47934 biosynthetic machinery. Through domain exchange of PCP-X constructs from the A47934 and teicoplanin NRPS system, it was discovered that the A47934 X-domain was responsible for the low levels of StaH activity [12]. In order to analyse if this effect is maintained over the subsequent amino acid cyclisation reactions in A47934 biosynthesis, we tested StaF activity using the same constructs all exhibiting MBP as N-terminal fusion partner: a PCP-X construct from A47934 biosynthesis (MBP-PCP-X\textsubscript{sta}, Table 1, entries 1–4), a PCP-X construct from teicoplanin biosynthesis (MBP-PCP-X\textsubscript{tei}, Table 1, entries 5–8) and hybrid PCP-X constructs from A47934 and teicoplanin biosynthesis (MBP-PCP\textsubscript{sta}/X\textsubscript{tei}, Table 1, entries 9–12; MBP-PCP\textsubscript{tei}/X\textsubscript{sta}, Table 1, entries 13–16) [12]. The influence of each individual PCP-X construct was tested with Tei7-L/D-Hpg\textsubscript{7}, Pek7-rac-Hpg\textsubscript{7}; Act7-rac-Hpg\textsubscript{7} and Act7-rac-Hpg\textsubscript{7} peptides. The presentation of Tei7-D-Hpg\textsubscript{7} and Pek7-rac-Hpg\textsubscript{7} by MBP-PCP-X\textsubscript{tei}, MBP-PCP\textsubscript{sta}/X\textsubscript{tei} and MBP-PCP\textsubscript{tei}/X\textsubscript{sta} did not lead to a change in their acceptance by StaF, with both peptides not accepted as substrates (Figure 4, Table 1, entries 6, 7, 10, 11, 14, and 15). In case of Tei7-L-Hpg\textsubscript{7}, StaF was active when the substrate was bound to MBP-PCP-X\textsubscript{sta} as described above, and we also observed a similar activity level when the peptide was presented by the MBP-PCP\textsubscript{tei}/X\textsubscript{sta} construct (Figure 1, entry 13). However, against our expectations, StaF activity did not increase on Tei7-L-Hpg\textsubscript{7} when bound to PCP-X constructs exhibiting the teicoplanin X-domain (MBP-PCP-X\textsubscript{tei}/PCP\textsubscript{sta}-X\textsubscript{tei}) but rather showed a significant decrease (Figure 4, Table 1, entries 5 and 9). StaF activity is clearly diminished when using the non-matched X-domain. In order to determine if this effect was maintained with different peptide substrates, we analysed StaF activity using the Act7-rac-Hpg\textsubscript{7} peptide: StaF activity was highest with MBP-PCP\textsubscript{tei}/X\textsubscript{sta}, and was considerably decreased with MBP-PCP\textsubscript{sta}/X\textsubscript{tei} (Figure 4, Table 1, entries 16, 4, 8 and 12, respectively). Thus, the negative effect of the teicoplanin X-domain on StaF activity with Act7-rac-Hpg\textsubscript{7} is not as clear as for Tei7-L-Hpg\textsubscript{7}. An explanation for this could lie in the observation that Act7-rac-Hpg\textsubscript{7} seems to be a very good substrate for StaF (as it is for OxyA\textsubscript{tei}), possibly due to the increased conformational flexibility of the phenylalanine residue at position 3 of the peptide when compared to the Hpg residue present in the teicoplanin-like peptide (Figure 4). In spite of this, Tei7-L-Hpg\textsubscript{7} is the peptide with the highest structural similarity to the natural substrate and the fact that StaF activity on Tei7-L-Hpg\textsubscript{7} was obtained only with PCP-X constructs exhibiting the A47934 X-domain indicates that StaF is dependent on the corresponding X-domain from its own NRPS when using teicoplanin-like peptides. Comparison of StaH and StaF activity reveals that while StaH exhibits only low to moderate activity on substrates bound to PCP-X constructs with the A47934 X-domain, presence of the A47934 X-domain in PCP-X constructs appears to be essential for StaF activity. In case of StaH, the high affinity of the A47934 X-domain likely hinders reorganisation of the P450/NRPS complex and hence can be trapped in states that display sub-optimal substrate orientation in the P450 active site [12]. In the case of StaF, it now seems clear that the natural X-domain is in fact the best system.
for peptide cyclisation, although different combinations of X-domain and peptide substrate can be identified that afford atypical levels of in vitro activity (StaF: Act7-rac-Hpg7 peptide and A47934 X-domain; StaH: Tei7-D-Hpg7 peptide and A47934 X-domain). Thus, our findings highlight the importance of the X-domain in GPA cyclisation reactions and provide further indication that its role appears to be more than just recruitment of the P450 to the substrate, but also ensuring proper substrate orientation via the PCP-domain in the P450 active site.

Structure and active site architecture of StaF

In order to gain insight into the structure-function relationship of StaF, we attempted to structurally characterise the protein. We were able to determine the crystal structure of StaF to a resolution of 2.1 Å and 2.2 Å using different cryo-protectant solutions (ethylene glycol and glycerol) and by solving the phase problem through molecular replacement using OxyEtei (PDB ID: 3O1A) as search model (Table 2) [22]. Both structures exhibit a core RMSD of only 0.2 Å, indicating that the structures are practically identical. Manual comparison also did

Table 2: Crystallographic data for StaF.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>StaF Native ethylene glycol</th>
<th>StaF Native glycerol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P3_2_1 (152)</td>
<td>P3_2_1</td>
</tr>
<tr>
<td>Cell dimensions a, b, c (Å)</td>
<td>110.1, 110.1, 93.7</td>
<td>109.7, 109.7, 93.9</td>
</tr>
<tr>
<td>Molecules/asymmetric unit</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>X-ray source</td>
<td>SLS X10SA</td>
<td>SLS X10SA</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9792</td>
<td>0.9792</td>
</tr>
<tr>
<td>Resolution (Å)^a</td>
<td>50.0–2.1</td>
<td>50.0–2.2</td>
</tr>
<tr>
<td>Rmerge^a</td>
<td>0.07 (0.31)</td>
<td>0.10 (0.44)</td>
</tr>
<tr>
<td>a因子</td>
<td>19.8 (4.2)</td>
<td>16.8 (5.5)</td>
</tr>
<tr>
<td>Completeness (%)^a</td>
<td>95.1 (91.6)</td>
<td>98.4 (95.5)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.2</td>
<td>9.6</td>
</tr>
<tr>
<td>Wilson B-factor (Å^2)</td>
<td>27.6</td>
<td>35.3</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>36206</td>
<td>31857</td>
</tr>
<tr>
<td>Resolution in refinement</td>
<td>50.0–2.1</td>
<td>50.0–2.2</td>
</tr>
<tr>
<td>Rwork/Rfree (%)</td>
<td>19.4 / 22.9</td>
<td>19.1 / 22.0</td>
</tr>
<tr>
<td>No. of atoms</td>
<td>3690</td>
<td>3576</td>
</tr>
<tr>
<td>Protein</td>
<td>3315</td>
<td>3269</td>
</tr>
<tr>
<td>Heme</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>64</td>
<td>–</td>
</tr>
<tr>
<td>Glycerol</td>
<td>–</td>
<td>48</td>
</tr>
<tr>
<td>Water</td>
<td>268</td>
<td>216</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>35.1</td>
<td>35.5</td>
</tr>
<tr>
<td>Heme</td>
<td>21.5</td>
<td>22.0</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>53.0</td>
<td>–</td>
</tr>
<tr>
<td>Glycerol</td>
<td>–</td>
<td>59.8</td>
</tr>
<tr>
<td>Water</td>
<td>43.8</td>
<td>41.7</td>
</tr>
<tr>
<td>RMSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.193</td>
<td>1.107</td>
</tr>
<tr>
<td>Ramachandran statistics§</td>
<td>97.3/2.2/0.5</td>
<td>97.8/1.7/0.5</td>
</tr>
<tr>
<td>Ramachandran statistics¶</td>
<td>97.1/2/3</td>
<td>97.6/2/6</td>
</tr>
<tr>
<td>PDB Code</td>
<td>5EX8</td>
<td>5EX9</td>
</tr>
</tbody>
</table>

^aNumbers in parentheses correspond to the highest resolution shell (2.2–2.1 Å; 2.3–2.2 Å). ^bRwork = Σ |Fo|–|Fc|/Σ |Fo|, calculated from the working reflection set; Rfree calculated in the same manner using the 5% test set reflections. ^cCalculated by PROCHECK; percentage of the protein residues in favored/ allowed/ disallowed regions. ^dCalculated by MOLPROBITY; percentage of the protein residues in most favored regions; disallowed residues and percentage of bad rotamers. ^eResidues in disallowed region: E331 (disordered loop region), F382 (active site residue, clearly defined density). ^fResidues in disallowed region: A329 (disordered loop region), F382 (active site residue, clearly defined density).
Figure 5: Structural analysis of StaF: (A) overall structure of StaF, with the heme moiety depicted using sticks and specific helices coloured and labelled; (B) view of the active site of StaF, with residues close to the heme moiety shown as sticks and labelled, with the colour scheme and labelling retained from panel (A); (C) an overlay of the StaF structure (orange) on the Xtei-OxyBtei complex (pale cyan/blue; PDB ID: 4TX3); (D) an overlay of StaF (orange) on the structure of OxyAtei (yellow) showing the location of the N-terminal regions of other molecules within the crystal lattice for both StaF (purple) and OxyAtei (green).

not reveal any important differences between them and thus we used the highest resolution structure for analysis (PDB ID: 5EX8; ethylene glycol cryo-protectant solution).

The StaF structure is well resolved and adopts the typical structure of a cytochrome P450 [34], which consists predominantly of α-helices (12 in total: labelled A to L, including two additional helices labelled A’ and J’, Figure 5A). The core of the P450, the four-helix bundle, is present in StaF and comprises helices D, E, I and L. Two β-sheet regions are observed on the side of the protein opposite to the core 4-helix bundle of the P450, with β-1 exhibiting 4 strands and β-2 exhibiting 2 strands. The most interesting structural feature of StaF is the long A’ helix at the N-terminus, which forms the ceiling of the active site. This helix seems to be specific for D-O-E ring catalysing P450s as it was only observed once before in OxyAtei, the D-O-E ring forming P450 from teicoplanin biosynthesis [14]. The centre of the active site is occupied by a heme moiety, which is sandwiched between helix I and L, the loops connecting helices B and C, K and L as well as the loop connecting the last β-strand of β-1 and the J’ helix. The thiolate side chain of Cys342 serves as proximal ligand for the heme.
and is found in the conserved P450 heme-binding sequence (FGHGxHxCLG) in the K–L loop. The heme propionate moieties also interact with the protein through ionic interactions: His93 (2.7 Å), Arg97 (2.8 Å), His283 (2.7 Å), Arg285 (2.7 and 2.9 Å) and His340 (2.8 Å).

The architecture of the active site involves the I helix, the B–C loop and the loop connecting the J’ helix and the last strand of β-1. Its ceiling is formed by the A’, F and G helices and the C-terminal loop of the protein (Figure 5B). Phe382, present in the long C-terminal loop that impinges on the active site, adopts an unusual Ramachandran conformation. As this conformation is also found for OxyA_{tel} (PDB ID: 5HH3) and forms a portion of the active site, this is likely to be of importance for the activity of these enzymes. The other region where Ramachandran outlines are present in the structure of StaF (329-331) is in the region prior to the crucial heme-coordinating cysteine residue Cys342, which is a region of poorly defined electron density. The I helix contains the conserved residues responsible for controlling protonation during oxygen activation of the P450 catalytic cycle (Asp235 and Gln236) [34]. Residues projecting into the active site are Thr86 in the β–C loop, Gly231 in the I helix and Asp279 and Thr282 in the loop connecting the J’ helix and β-1. These residues make the active site more polar than those of OxyB/OxyC homologues, whilst aromatic amino acids are concentrated at the B–C-loop side of the active site, with Trp81 and Phe84 in the B–C loop and Phe228 in the I helix. This distribution of polar and hydrophobic residues in the active site is clearly different from the arrangement in related P450s such as StaH (PDB ID: 5EX6), OxyB_{tel} (PDB ID: 4TXF) and OxyB_{van} (PDB ID: 1LG9) [12,19,40,41], where hydrophobic residues were concentrated in the middle of the active site around the heme, and was only previously observed in OxyA_{tel} (PDB ID: 5HH3) [14].

**Structural comparison to other P450s**

The presence of the additional A’ helix and the distinct distribution of polar and aromatic amino acid residues in the active site sets StaF and OxyA_{tel} apart from other structurally characterised examples of P450s involved in GPa cyclisation reactions. Comparison of StaF and OxyA_{tel} (PDB ID: 5HH3) reveals very similar structures with a core rmsd of 1.2 Å. Major differences include the length of the N-terminus, which is shorter for StaF, the conformation of the B–C loop, which exhibits a helical part in OxyA_{tel} in contrast to StaF, and the position of the F and G helices, which are drawn down towards the centre of the protein in StaF closing the active site to a greater extent than observed for OxyA_{tel}. In both the StaF structures, the N-terminal (tag) region of a symmetry-related molecule forms a loop above the heme, which likely leads to the open conformation of the B–C loop region (Figure 5D). One of the protein chains in the asymmetric unit of the OxyA_{tel} structure also displays an interaction with the N-terminus of another protein chain, although in this case there is direct coordination between the N-terminal amine nitrogen and the heme iron. This different binding mode leads to minor changes in the orientation of various amino acid side chains within the active site of OxyA_{tel} when compared to StaF as well as the opening of the F-G helices and alterations to the I-helix packing (Figure 5D). An attempt to reengineer the protein construct to shorten the N-terminal protein tag and to redesign the sequence to resemble that of a PCP domain both lead to proteins that failed to crystallise either under the original conditions or in broad screens. Thus, it would appear as though OxyA homologues require active site interactions in order to stabilise their structures sufficiently to enable crystallisation, which is in contrast with other Oxy homologues. The importance of active site interactions may also provide an indication why OxyA enzymes appear to have higher degrees of substrate specificity than OxyB homologues.

The StaF structure is similar to the structures of other Oxy homologues that have been solved [4], including OxyE_{tel} (PDB ID: 3OIA/3O03) [21,22] and OxyB_{tel} in complex with the X-domain (PDB ID: 4TX3) [16] with an rmsd of under 2.0 Å (Table 3). Other P450 enzymes with high structural similarity to StaF are those from secondary metabolism and involve oxidative functionalisation of large substrates, such as pravastatin (CYP105AS, PDB ID: 4OQS) [42], oleandomycin (OleP, PDB ID: 4XE3) [43], mycinamicin (MycG, PDB ID: 2YCA) [44] and filipin (CYP105P1, PDB ID: 3E5L) [45] (Table 3). StaF also shows moderate levels of structural similarity to other P450s that oxidise carrier protein-bound substrates, including the fattyacyl-ACP oxidase P450\_BiO (PDB ID: 3EDJ) [46-48] and the aminocacyl-PCP hydroxylases OxD (PDB ID: 3MGX) and P450\_axy (PDB ID: 4PXH) [49,50] (Table 3). Central to the Oxy/X-domain interaction is the PRDD-region, which is found at the beginning of the F-helix in the Oxy enzymes [16]. This motif is conserved in the Oxy enzymes, and the two Asp residues located in this region form numerous contacts to the X-domain [16]. Overlaying the structure of StaF onto the OxyB_{tel}/X-domain complex structure shows that the interface expected between StaF and the teicoplanin X-domain would appear to be a favourable one, although this is clearly not the case based on the data from in vitro activity assays (Figure 5C). Sequence-based comparisons of OxyA_{tel} and StaF (Figure 6) as well as the A47934 and teicoplanin X-domains (Figure 7) also do not provide a clear indication of the grounds of the selectivity of StaF for the A47934 X-domain over that from the teicoplanin system. However, the discovery that peptides can be accepted by StaF when presented by the teicoplanin X-domain if the correct peptide sequence is selected (specifically the
Table 3: Top ranking structures homologous to StaF as identified by a Dali search.

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Chain</th>
<th>RMSD Ca [Å]</th>
<th>Z-score</th>
<th>% Identity</th>
<th>Description (donor organism)</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HH3</td>
<td>A</td>
<td>1.2</td>
<td>55.9</td>
<td>79</td>
<td>OxyA&lt;sub&gt;tei&lt;/sub&gt; (Actinoplanes teichomyceticus)</td>
<td>[14]</td>
</tr>
<tr>
<td>5HH3</td>
<td>C</td>
<td>1.6</td>
<td>55.7</td>
<td>80</td>
<td>OxyA&lt;sub&gt;tei&lt;/sub&gt; (Actinoplanes teichomyceticus)</td>
<td>[14]</td>
</tr>
<tr>
<td>3OO3</td>
<td>A</td>
<td>1.8</td>
<td>46.3</td>
<td>48</td>
<td>CYP165D3 (Actinoplanes teichomyceticus)</td>
<td>[21]</td>
</tr>
<tr>
<td>3OO3</td>
<td>A</td>
<td>2.0</td>
<td>46.3</td>
<td>48</td>
<td>CYP165D3 (Actinoplanes teichomyceticus)</td>
<td>[22]</td>
</tr>
<tr>
<td>4TX3</td>
<td>A</td>
<td>1.9</td>
<td>44.0</td>
<td>41</td>
<td>OxyB&lt;sub&gt;tei&lt;/sub&gt; in complex with the X-domain (Actinoplanes teichomyceticus)</td>
<td>[16]</td>
</tr>
<tr>
<td>1LG9</td>
<td>A</td>
<td>2.6</td>
<td>43.6</td>
<td>38</td>
<td>CYP165B3 (Nocardia orientalis)</td>
<td>[41]</td>
</tr>
<tr>
<td>5EX6</td>
<td>A</td>
<td>2.2</td>
<td>43.4</td>
<td>41</td>
<td>StaH (Streptomyces toyoaensis NRRL15009)</td>
<td>[12]</td>
</tr>
<tr>
<td>1UED</td>
<td>A</td>
<td>2.1</td>
<td>43.2</td>
<td>33</td>
<td>CYP165C3 (Nocardia orientalis)</td>
<td>[40]</td>
</tr>
<tr>
<td>4OQS</td>
<td>A</td>
<td>2.2</td>
<td>42.1</td>
<td>34</td>
<td>CYP105AS1 (Amycolatopsis orientalis)</td>
<td>[42]</td>
</tr>
<tr>
<td>4TVF</td>
<td>A</td>
<td>2.1</td>
<td>41.2</td>
<td>42</td>
<td>OxyB&lt;sub&gt;tei&lt;/sub&gt; (Actinoplanes teichomyceticus)</td>
<td>[19]</td>
</tr>
<tr>
<td>4XE3</td>
<td>B</td>
<td>2.4</td>
<td>41.1</td>
<td>30</td>
<td>OleP (Streptomyces antibioticus)</td>
<td>[43]</td>
</tr>
<tr>
<td>2YCA</td>
<td>A</td>
<td>2.3</td>
<td>40.8</td>
<td>28</td>
<td>MysG (Micromonospora griseorubida)</td>
<td>[44]</td>
</tr>
<tr>
<td>3ESL</td>
<td>A</td>
<td>2.5</td>
<td>40.7</td>
<td>34</td>
<td>CYP105P1 (Streptomyces avermitilis)</td>
<td>[45]</td>
</tr>
<tr>
<td>3EDa</td>
<td>B</td>
<td>2.4</td>
<td>38.5</td>
<td>25</td>
<td>P450&lt;sub&gt;bio&lt;/sub&gt; (CYP107H1, Bacillus subtilis)</td>
<td>[47]</td>
</tr>
<tr>
<td>3MGX</td>
<td>B</td>
<td>2.9</td>
<td>35.6</td>
<td>22</td>
<td>OxyD (CYP146, Amycolatopsis orientalis)</td>
<td>[51]</td>
</tr>
<tr>
<td>4PXH</td>
<td>E</td>
<td>3.0</td>
<td>35.5</td>
<td>19</td>
<td>PCP&lt;sub&gt;7&lt;/sub&gt;-P450&lt;sub&gt;bio&lt;/sub&gt; complex (CYP163B3, Streptomyces sp. ACTA 2897)</td>
<td>[49]</td>
</tr>
</tbody>
</table>

*<sup>a</sup>Included for purposes of comparison.

Figure 6: Sequence alignment of StaF and OxyA<sub>tei</sub>. Protein secondary structure was derived from the StaF crystal structure (PDB ID: 5EX8) and is shown above the alignment (α-helices = blue, β-sheets = magenta). The PRDD-region, which has been shown to be crucial for interaction with the X-domain is highlighted in an orange box.
Figure 7: Sequence alignment of the A47934 (sta) and teicoplanin (tei) X-domain; secondary structure was derived from the Xtei-OxyBtei-complex (PDB ID: 4TX3) and is shown above the alignment (α-helices = blue, β-sheets = magenta); the residues crucial for interaction with cytochrome P450s are shown in orange and both the crossover I region (purple) and the crossover II region (green) are highlighted.

Act7-rac-Hpg7 peptide) shows that the peptide plays a significant role in the formation of a catalytically competent state of StaF. This cannot be explained by the current structures that we have access to from GPA biosynthesis. This also clearly indicates the importance of characterising substrate-bound Oxy structures in future, although this remains a challenging task.

Conclusion
In this study we characterised the activity and structure of StaF, the D-O-E ring forming Oxy enzyme from A47934 biosynthesis. This is only the second characterised example of these types of P450s, after the teicoplanin homologue OxyAtei. StaF adopts the canonical P450 fold and strongly resembles the structure of OxyAtei, with both exhibiting the long additional A' helix at the protein’s N-terminus. Spectral analysis of StaF showed that it exhibits the typical P450 absorption spectra, but with only half of the StaF species being in the catalytically competent state upon reduction and CO-complexation. Despite this, we successfully reconstituted the StaF activity in vitro and could show that the substrate specificity of StaF is not as broad as for OxyS catalysing the C-O-D ring formation, in agreement with the results from OxyAtei. Additionally, we could show that StaF interacts with the A47934 X-domain, indicating that StaF is, along with other related Oxy enzymes, recruited by the X-domain to the A47934 NRPS machinery. The interaction of StaF to Xsta appears to be weaker than the interaction of StaH to Xsta. We have previously shown that the strong StaH/Xsta interaction is the cause for poor substrate turnover of StaH of substrates bound to PCP-X constructs exhibiting the A47934 X-domain. In contrast, the weaker interaction of StaF to Xsta helps to explain why StaF exhibits higher levels of activity against substrates bound to PCP-X construct exhibiting the A47934 X-domain. Taking into account the weaker binding of StaF to Xsta, we postulate that the weaker interaction of this
complex allows substrate reorganisation after initial complex formation, which ensures proper substrate orientation in the active site. These results highlight the importance of testing different peptide/protein carrier constructs for in vitro GPA cyclisation assays and show that different Oxy homologues, such as StaH and StaF, can display significantly different reactivity and specificity despite their similar sequences, structures and substrates. Such insights will be crucial in future identification of an optimal system for the in vitro generation of GPAs.

**Experimental Cloning**

The gene encoding StaF was obtained from genomic DNA [33] and was amplified by PCR using specific primers (fwd: 5'-CACCATGTTCGAGGAGATCAACGTCGTC-3', rev: 5'-CTACCAAGTGGACGACGGCTTC-3') for cloning into pET151d (Life Technologies) using TOPO-cloning. The plasmid was sequenced using T7 promoter and terminator primers. StaF was expressed with an N-terminal hexahistidine-tag and under the control of the T7 promoter. The StaH construct (pET28a StaH) as well as all NRPS constructs (pET MBP-PCP1c,Xsta1c, pET MBP-PCP,PCP1c,Xtei1c, pET MBP-PCP1c,Xsta1c, pET MBP-PCP1c,Xtei1c, pET NCL-4 MBP,Xsta1c) were employed from a previous study – Ulrich et al. (2016) [12].

**Expression and purification**

StaF. For the expression of StaF, a starter culture of *E. coli* KRX cells (Promega), which had been transformed with pET151d StaF, was grown at 37 °C overnight. This was used for the inoculation of 6 × 2 L TB medium plus 100 mg/L ampicillin with 1% (v/v) of starter culture. This expression culture was incubated at 37 °C and 90 rpm until an OD600 of 0.4 was reached. At this point, 25 mg/L δ-aminolevulinic acid was added and the temperature was decreased to 18 °C. The culture was further grown until an OD600 = 0.6–0.8, at which the expression was induced with 0.1% rhamnose and 0.1 mM IPTG. After overnight expression, cells were harvested by centrifugation at 1000 × g for 1 min. Subsequently, the Ni-NTA resin was incubated with the cleared lysate for 1 h and rotation. The supernatant was then removed by centrifugation as described above, before the Ni-NTA resin was washed with 10 × CV of Ni-NTA wash buffer for 5 min with rotation. Prior to transfer of the Ni-NTA resin into column format, the Ni-NTA wash buffer was removed by centrifugation as described above and resuspended in 2 × CV Ni-NTA wash buffer. StaF was finally eluted using 3 × CV of Ni-NTA elution buffer (50 mM Tris pH 8, 300 mM NaCl, 300 mM imidazole).

For anion exchange chromatography (AEC), the Ni-NTA elution buffer was exchanged with AEC buffer A (see below) using illustra NAP-25 columns (GE Healthcare, Chalfont St Giles, UK) and concentrated using vivaspin® centrifugal concentrators with a 30 kDa MWCO (Sartorius, Göttingen, Germany). AEC was then performed using a Resource™ Q (6 mL) column (GE Healthcare, Chalfont St Giles, UK) connected to an Äkta pure 25 system with 50 mM Tris pH 7.4, 20 mM NaCl as AEC buffer A and 50 mM Tris pH 7.4, 1 M NaCl as AEC buffer B at rt. The column was equilibrated with AEC buffer A, before the protein solution was applied onto the column. The column was then washed with 5 × CV of AEC buffer A, before StaF was eluted using a gradient of 20 × CV of 0 to 100% AEC buffer B. Appropriate elution fractions were pooled and concentrated as described above, after analysis by SDS-PAGE.

Additionally, StaF was further purified by size-exclusion chromatography (SEC) using a Superose 12 (300 mL) column connected to an Äkta pure 12 system at rt. The column was equilibrated with SEC buffer (50 mM Tris pH 7.4, 150 mM NaCl), before the protein solution was applied onto the column. StaF was then eluted using SEC buffer. The elution fractions were again analysed by SDS-PAGE, appropriate fractions were pooled and concentrated as described above. Determination of the protein concentration was performed spectrophotometrically using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and the calculated extinction coefficient of the protein at λ = 280 nm. Furthermore, the protein identity was confirmed by MALDI–TOF MS peptide fingerprinting of a tryptic digest of excised protein bands from SDS-PAGE analysis. StaF was finally stored in SEC buffer in aliquots, which were first flash frozen in liquid nitrogen before being stored at −80 °C. The yield of purified StaF was 28 nanomoles (1.3 mg) per L of expression culture.

Ni-NTA Agarose resin (Macherey-Nagel, Düren, Germany) was equilibrated twice with the 10-fold column bed volume (CV) of Ni-NTA wash buffer (50 mM Tris pH 8, 300 mM NaCl, 10 mM imidazole). This was achieved through resuspension of the resin in the Ni-NTA wash buffer and subsequent removal of the Ni-NTA wash buffer after centrifugation at 1000g for 1 min. Subsequently, the Ni-NTA resin was incubated with the cleared lysate for 1 h and rotation. The supernatant was then removed by centrifugation as described above, before the Ni-NTA resin was washed with 10 × CV of Ni-NTA wash buffer for 5 min with rotation. Prior to transfer of the Ni-NTA resin into column format, the Ni-NTA wash buffer was removed by centrifugation as described above and resuspended in 2 × CV Ni-NTA wash buffer. StaF was finally eluted using 3 × CV of Ni-NTA elution buffer (50 mM Tris pH 8, 300 mM NaCl, 300 mM imidazole).

For anion exchange chromatography (AEC), the Ni-NTA elution buffer was exchanged with AEC buffer A (see below) using illustra NAP-25 columns (GE Healthcare, Chalfont St Giles, UK) and concentrated using vivaspin® centrifugal concentrators with a 30 kDa MWCO (Sartorius, Göttingen, Germany). AEC was then performed using a Resource™ Q (6 mL) column (GE Healthcare, Chalfont St Giles, UK) connected to an Äkta pure 25 system with 50 mM Tris pH 7.4, 20 mM NaCl as AEC buffer A and 50 mM Tris pH 7.4, 1 M NaCl as AEC buffer B at rt. The column was equilibrated with AEC buffer A, before the protein solution was applied onto the column. The column was then washed with 5 × CV of AEC buffer A, before StaF was eluted using a gradient of 20 × CV of 0 to 100% AEC buffer B. Appropriate elution fractions were pooled and concentrated as described above, after analysis by SDS-PAGE.

Additionally, StaF was further purified by size-exclusion chromatography (SEC) using a Superose 12 (300 mL) column connected to an Äkta pure 12 system at rt. The column was equilibrated with SEC buffer (50 mM Tris pH 7.4, 150 mM NaCl), before the protein solution was applied onto the column. StaF was then eluted using SEC buffer. The elution fractions were again analysed by SDS-PAGE, appropriate fractions were pooled and concentrated as described above. Determination of the protein concentration was performed spectrophotometrically using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and the calculated extinction coefficient of the protein at λ = 280 nm. Furthermore, the protein identity was confirmed by MALDI–TOF MS peptide fingerprinting of a tryptic digest of excised protein bands from SDS-PAGE analysis. StaF was finally stored in SEC buffer in aliquots, which were first flash frozen in liquid nitrogen before being stored at −80 °C. The yield of purified StaF was 28 nanomoles (1.3 mg) per L of expression culture.
StaH, MBP-PCP<sub>sta</sub>-X<sub>sta</sub>, MBP-PCP<sub>tei</sub>-X<sub>tei</sub>, MBP-PCP<sub>tei</sub>-X<sub>sta</sub>, MBP-PCP<sub>tei</sub>-X<sub>tei</sub>-X<sub>sta</sub>, X<sub>sta</sub>. Purification of before mentioned proteins was performed as described by Ulrich et al. (2016) [12].

Spectral analysis of StaF

StaF was analysed spectroscopically in a concentration of 2.5 μM in 50 mM Tris pH 8.0 at 30 °C using a Jasco V-650 spectrophotometer and the SpectraManager software in order to determine the potentially catalytic active species. Spectral analysis was performed from 390 to 600 nm with 0.2 nm increments from the ferric protein (as purified), the ferrous protein, which had been reduced through the addition of 10 μL of a saturated Na<sub>2</sub>SO<sub>4</sub> solution, and of the ferrous P450, which had been saturated with CO through bubbling of 60 mL CO gas using a syringe through the cuvette filled with protein solution.

Protein interaction studies

The interaction analysis of StaF with the A47934 X-domain (X<sub>sta</sub>) was done by analytical size-exclusion chromatography (SEC) using a Superose 12 10/300 GL column connected to an Äkta pure 25 system and the unicorn 6.4 software. The Superose 12 column had been calibrated using Gel Filtration Standard from Bio-Rad (Catalogue number 151-1901) resulting in following elution volumes: 670,000 Da at 8.23 mL, 158,000 Da at 11.27 mL, 44,000 Da at 13.01 mL, 17,000 Da at 14.62 mL for and 1,350 Da at 19.23 mL. 50 mM Tris pH 7.4 and 150 mM NaCl was used as SEC buffer. This method was appropriate for the analysis of the P450 – X-domain interaction as both the protein specific absorption at λ = 280 nm as well as the heme absorption at approximately λ = 415 nm could be monitored. Interaction of StaF and X<sub>sta</sub> was detected through a significant shift of the heme peak at λ = 415 nm to earlier elution volume when StaF and X<sub>sta</sub> were analysed together compared to individual analysis of StaF. Prior to analysis, 33.3 μM StaF and 100 μM X<sub>sta</sub> were incubated at RT for 30 min in SEC buffer in a reaction volume of 100 μL. The reaction was then analysed with the flow rate set to 1 mL/min and detection of the absorption at λ = 280 and 415 nm. Individual analysis of StaF and X<sub>sta</sub> served as controls.

P450 assay activity

An in vitro phenolic coupling assay was performed in order to determine the StaF activity. As substrates the teicoplanin-like NH<sub>2</sub>-D-Hpg-D-Tyr-L-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Tei7-L-Hpg<sub>3</sub>, D/L-Hpg<sub>2</sub>)), the pekiskomycin-like NH<sub>2</sub>-D-Ala-D-Tyr-L-Glu-D-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Pek7(D/L-Hpg<sub>2</sub>)), and the actinoidin-like NH<sub>2</sub>-D-Hpg-D-Tyr-L-Phe-D-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Ac77(D/L-Hpg<sub>2</sub>)) heptapeptide were used, which were synthesised according to Brieke et al. [18,36]. It has to be noted that the Hpg residue at position 7 of all heptapeptides is highly racemisation prone. In case of the Tei7 peptide effective separation by preparative HPLC was possible, so that pure L-Hpg and D-Hpg<sub>7</sub> peptide could be used [13]. The diastereomers of Pek7 and Ac77 were not separated by preparative HPLC, so that both peptides were used with a racemic mixture of D/L-Hpg[17]. The activity assay was performed as described in Brieke and Peschke et al. [17] with the first step being the loading of the substrate peptide onto the PCP-X construct (MBP-PCP<sub>sta</sub>-X<sub>sta</sub>, MBP-PCP<sub>tei</sub>-X<sub>tei</sub>, MBP-PCP<sub>tei</sub>-X<sub>sta</sub>, MBP-PCP<sub>tei</sub>-X<sub>tei</sub>-X<sub>sta</sub>) using the R4-4 mutant of the promiscuous phosphonatecinyl transferase Sfp, subsequently, the actual activity assay was performed using palustrisredoxin B (A105V), palustrisredoxin reductase and NADH as P450 electron source [38], and finally the peptides were purified by solid-phase extraction and analysed HPLC–MS [17]. The StaF activity assays with Tei7-L-Hpg as substrate were performed both with and without StaH. All other StaF activity assays were always performed together with StaH.

StaF protein crystallisation

Crystals were grown using hanging drop vapour diffusion at 4 °C. The StaF protein (140 μM) was mixed (1:1) with the reservoir solution (0.1 M phosphate/citrate buffer (pH 4.2), 1.2 M Na<sub>2</sub>PO<sub>4</sub>, 0.3 M K<sub>2</sub>HPO<sub>4</sub>; final pH 5.2) and equilibrated against the reservoir solution. After 10 days red diamonds (<150 μm length) had formed. The crystals were passed through a cocrystallization solution (0.1 M phosphate/citrate buffer (pH 4.2), 1.2 M Na<sub>2</sub>PO<sub>4</sub>, 0.3 M K<sub>2</sub>PO<sub>4</sub> and either 25% (v/v) glycerol, or 25% (v/v) ethylene glycol) and then flash cooled in liquid nitrogen for data collection. Two native data sets using different cryoprotectant solutions were collected at the X10SA beamline at the Swiss Light Source at the Paul Scherrer Institute (Villigen, Switzerland, λ = 0.9792 Å) with the crystals kept at 100 K during data collection. The data was processed using the XDS program suite [52]. The space group of the crystals was P<sub>2</sub>1<sub>1</sub>(1) with a single P450 molecule per asymmetric unit. The StaF structure was solved using molecular replacement with the program Phaser [53] and a search model consisting of OxyEt<sub>163</sub> (Protein Data Bank code 3O1A, Chain A) [22], residues 2-384 and heme. Iterative manual model building and refinement were performed using the programs COOT [54] and REFMAC [55] with TLS refinement [56] following a simulated annealing protocol in CNS [57,58]. During several rounds of refinement with REFMAC and manual rebuilding, ethylene glycol or glycerol and solvent molecules were included in the models where appropriate. TLS input files were generated using the TLS-Motion Determination Server [59,60]. Structure validation was performed using MolProbity [61] and PROCHECK [62]. Structure-based sequence alignments were carried out with SSM [63] as implemented in COOT and comparisons to known structures performed with DaliLite [64].
All structural figures were prepared using PyMol [65]. Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank (PDB) under accession codes 5EX8 (ethylene glycol cryoprotectant solution) and 5EX9 (glycerol cryoprotectant solution).

Supporting Information

Supporting Information File 1

HPLC–MS analysis of StaF turnover activity of Tei7-L-Hpg7 (a) and Act7-rac-Hpg7 (b) bound to MBP-PCP-Xsta.

[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-12-284-S1.pdf]

Acknowledgements

The authors are thankful to Madeleine Peschke for preparing MBP-PCP-Xsta to Alexa Koch for assistance in cloning and protein preparation; to Melanie Müller for mass spectral analysis; and to Christopher Roome for IT support. Diffraction data were collected at the Swiss Light Source, beamline X10SA, Paul Scherrer Institute, Villigen, Switzerland. We thank the Heidelberg team for data collection and the PXII staff for their support in setting up the beamline. M.J.C. is grateful for the Heidelberg team for data collection and the PXII staff for their support in setting up the beamline.

References

A postsynthetically 2’-“clickable” uridine with arabino configuration and its application for fluorescent labeling and imaging of DNA

Heidi-Kristin Walter1, Bettina Olshausen2, Ute Schepers2 and Hans-Achim Wagenknecht*1

Full Research Paper

Address:
1 Institute of Organic Chemistry, Karlsruhe Institute of Technology (KIT), Fritz-Haber-Weg 6, 76131 Karlsruhe, Germany and
2 Institute of Toxicology and Genetics, Karlsruhe Institute of Technology (KIT), H.-v.-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

Email:
Hans-Achim Wagenknecht* - Wagenknecht@kit.edu

* Corresponding author

Keywords:
dyes; fluorescence; nucleic acid; oligonucleotide

doi:10.3762/bjoc.13.16
Received: 17 October 2016
Accepted: 03 January 2017
Published: 20 January 2017

This article is part of the Thematic Series "Chemical biology".

Received: 17 October 2016
Accepted: 03 January 2017
Published: 20 January 2017

© 2017 Walter et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract
The arabino-configured analog of uridine with a propargyl group at the 2’-position was synthesized and incorporated into DNA by solid-phase chemistry. The fluorescence quantum yields of DNA strands that were postsynthetically modified by blue and green emitting cyanine-styryl dyes were improved due to the arabino-configured anchor. These oligonucleotides were used as energy transfer donors in hybrids with oligonucleotides modified with acceptor dyes that emit in the yellow-red range. These combinations give energy transfer pairs with blue–yellow, blue–red and green–red emission color changes. All combinations of arabino- and ribo-configured donor strands with arabino- and ribo-configured acceptor strands were evaluated. This array of doubly modified hybrids was screened by their emission color contrast and fluorescence quantum yield. Especially mixed combinations, that means donor dyes with arabino-configured anchor with acceptor dyes with ribo-configured anchor, and vice versa, showed significantly improved fluorescence properties. Those were successfully applied for fluorescent imaging of DNA after transport into living cells.

Introduction
The “click”-type reactions [1], in particular the 1,3-dipolar cycloaddition between alkynes and azides (CuAAC) is a broadly applied strategy for postsynthetic oligonucleotide modification since both reactive groups are not present in nucleic acids [2-5]. Although Huisgen described the uncatalyzed reaction yielding 1,2,3-triazoles already in the 1960s [6], the bioorthogonality with respect to proteins and nucleic acids emerged after Sharpless [7] and Meldal [8] had reported that catalysis by Cu(I) enhances not only reaction rates but improves also regioselectivity. The formation of oligonucleotide oxidation side products by Cu(I) is avoided by the use of chelating Cu(I) ligands, in particular tris(1-benzyl-1H-1,2,3-triazol-4-
configured uridine analog 3. The 3′- and 5′-hydroxy functions of nucleoside 3 were selectively protected by the Markiewicz silyl ether [22]. The central step of the whole synthetic procedure was the alkylation of the 2′-OH function of nucleoside 4 by propargylic bromide which worked in 65% yield in the presence of NaH as base. After removal of the silyl protecting group from nucleoside 5, the 5′-position of nucleoside 2 was again protected by 4,4′-dimethoxytrityl chloride (DMTr-Cl) and, finally, the 3′-position of nucleoside 6 was phosphitylated. Remarkably, the overall yield of phosphoramidite 7 with the optimized conditions over the described five steps is 54%. Automated DNA synthesis with 7 as building block required a slightly extended coupling time of 10 min. The phosphoramidite for the “clickable” nucleoside 1 is commercially available. After preparation, the detritylated oligonucleotides DNA1a (“a” = arabino) and DNA1r (“r” = ribo) were cleaved from the resin and deprotected with conc. NH4OH at 45 °C for 16 h. The lyophilized oligonucleotides were reacted with the azide-modified dyes D1–D4 in the presence of Cu(I) and TBTA, as mentioned above. The reaction was performed in H2O/DMSO/i-BuOH 3:3:1 and was completed after 1.5 h at 60 °C. The modified oligonucleotides were purified by ethanol precipitation in the presence of EDTA to remove copper ions and subsequently by semi-preparative HPLC. Finally, the modified oligonucleotides were identified by MALDI–TOF mass spectrometry (see Supporting Information File 1) and annealed with the corresponding unmodified counterstrand.

The four fluorophores D1 [23], a blue emitter excitable at 389 nm, D2 [24], D3 [19], and D4 [24], all green emitters excitable at 450–460 nm, that were “clicked” to the oligonucleotides DNA1a and DNA1r belong to our recently established class of cyanine-styryl dyes that show a unique combination of optical properties [25], including suitable brightness and fluorescence quantum yields, large Stokes’ shifts compared to conventionally applied Cy3 and Cy5, and most importantly, excellent photostabilities. D1–D4 were representatively chosen since they will serve as energy donors in the energy transfer-based DNA systems (vide infra). The corresponding dye azides were synthesized as previously described [19,23,24]. The modified double strands (ds) DNA2aD1 to DNA2aD4 were compared with their structural counterpart among the duplexes DNA2rD1 to DNA2rD4 with respect to their optical properties (UV–vis absorption and fluorescence, see Supporting Information File 1), fluorescence quantum yields ΦF and melting temperatures Tm (Table 1). The reference duplexes of DNA1a and DNA1r annealed with the unmodified complementary strand showed Tm values of 61.0 °C and 62.0 °C, respectively. This small difference tracks well with the general observation that arabino-configured nucleic acids in general show lower stabilities than the ribo-configured ones. With the attached dyes, the
Scheme 2: Synthesis of phosphoramidite 7 and modified DNA. a) TIPDSiCl₂, pyridine, 2 h at 0 °C, 16 h at rt, 89%; b) 1. NaH, THF, 0 °C, 15 min, 2. propargyl bromide, rt, 18 h, 65%; c) TBAF, THF, rt, 5 min, 99%; d) DMTr-Cl, pyridine, rt, 5 h, 99%; e) 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite, (iPr)₂NEt, CH₂Cl₂, rt, 3 h, 95%; f) automated DNA synthesis; g) D₁–D₄, sodium ascorbate, TBTA, (CH₃CN)₄CuPF₆, H₂O/DMSO/t-BuOH 3:3:1, 1.5 h, 60 °C; annealing with counterstrand for 10 min at 90 °C and slow cooling to rt. For structures of D₁–D₄ see Scheme 3.

Table 1: Melting temperatures (T_m) and fluorescence quantum yields (Φ_F) of singly modified DNA2aD₁–DNA2rD₄.

<table>
<thead>
<tr>
<th>dye</th>
<th>λ_exc (nm)</th>
<th>DNA2a... T_m [°C]</th>
<th>Φ_F</th>
<th>DNA2r... T_m [°C]</th>
<th>Φ_F</th>
</tr>
</thead>
<tbody>
<tr>
<td>...D₁</td>
<td>389</td>
<td>61.9</td>
<td>0.096a</td>
<td>65.7</td>
<td>0.052a</td>
</tr>
<tr>
<td>...D₂</td>
<td>462</td>
<td>61.7</td>
<td>0.452b</td>
<td>64.0</td>
<td>0.266b</td>
</tr>
<tr>
<td>...D₃</td>
<td>450</td>
<td>64.2</td>
<td>0.136c</td>
<td>65.1</td>
<td>0.122c</td>
</tr>
<tr>
<td>...D₄</td>
<td>462</td>
<td>64.2</td>
<td>0.156d</td>
<td>65.2</td>
<td>0.087d</td>
</tr>
</tbody>
</table>

arabino-modified duplexes show a smaller stabilization effect by the dyes than the corresponding ribo-modified duplexes. The stabilization of dsDNA2a ranges only from 0.7 °C for D₁ to 3.2 °C for D₃ and D₄, whereas the stabilizing effects for dsDNA2r are more diverse, ranging from 2.0 °C for D₁ to 3.7 °C for D₃. Obviously, the dye interactions with double-stranded DNA do slightly depend on the type of dye. In D₁ and D₂, the pyridinium part is connected to the rest of the dye by its 4-position, in D₃ and D₄ via its 2-position. The latter connectivity has a larger stabilizing influence on the DNA2a double strands. The fluorescence quantum yields of dsDNA2a are all

129
higher than the corresponding ones of dsDNA2r. Especially in case of D2 ΦF could be significantly improved from 27% to 45%, and in case of D4 from 9% to 16%. This is remarkable and clearly shows that the arabino-configured nucleoside 2 provides the structurally optimized anchor for fluorescent dye interactions with the DNA. Obviously, placing the dyes into the major groove led them find a better orientation than in the minor groove, with respect to the DNA helix with enhanced fluorescence intensities.

The dyes D1–D4 as energy donors were combined with dyes D5–D9 as energy acceptors (Scheme 3). This approach follows our concept of “DNA/RNA traffic lights” [17,19,25] that are energy transfer-based nucleic acid probes that can be used in molecular beacons [26], especially for vesicular microRNA imaging in living cancer cells [27], and for siRNA transport imaging [28]. Donor and acceptor dyes are combined in an interstrand and diagonal orientation to promote best possible energy transfer. In particular, we combined each of the eight

\[ \text{DNA2a/DNA2r} \]

**Scheme 3:** Structures of donor dyes D1–D4 as modifications of DNA2a and DNA2r and structures of acceptor dyes D5–D9 as modifications of DNA3a and DNA3r yielding energy transfer-based nucleic acid probes.
oligonucleotides DNA2a and DNA2r modified with D1–D4 with each of the ten oligonucleotides DNA3a and DNA3r modified with D5 [29], D6 [25], D7 [30], D8 [19] and D9 [29]. In detail, the blue emitting dye D1 combines to a blue–yellow fluorophore pair (D1/D5) and to blue–red emitting pairs (D1/D6–D1/D9). The green emitting dyes D2–D4 result all in green–red emitting pairs (D2/D7–D4/D9). The combination of dyes D2–D4 with D5 and D6 is not meaningful for this concept since the fluorescence of the donors and absorption of the acceptors show broad spectral overlays and therefore selective excitation is not possible.

For each of the previously described dye combinations, we probed all four combinations of arabino- and ribo-configured donor strands (DNA2a and DNA2r) with acceptor strands (DNA3a and DNA3r). This array of doubly modified DNA duplexes was screened by their emission color contrast C = IAc/IDo (fluorescence intensity near maximum of acceptor divided by fluorescence intensity near maximum of donor) and for the fluorescence quantum yield ΦF in the range of the acceptor emission (Table 2). The comparison with our previously applied approach to link both donor and acceptor dyes at ribo-configured nucleoside 1 (in DNA2r and DNA3r) revealed that the emission color contrasts are not improved when they are both anchored at the arabino-configured nucleoside 2 (in DNA2a and DNA3a). There are only very few exceptions; especially the combination DNA2aD4–DNA3aD8 yields a red-to-green contrast of 83 compared to 45 in case of the ribo-configured DNA2rD4–DNA3rD8. Additionally, the fluorescence quantum yield was improved from 39% to 53%. The dye combination D2/D8 nicely demonstrates the effect of the arabino-configured attachment because the corresponding duplexes show all enhanced quantum yields whereas the pure ribo configuration in DNA2rD2–DNA3rD8 quenches its fluorescence significantly. The latter example (DNA2rD2–DNA3rD8) shows an altered absorbance of the two dyes which gives an important photophysical insight. An efficient energy transfer between two dyes requires the selective excitation of an uncoupled donor in the proximity to an uncoupled acceptor. Hyper-/hypochromicity and/or shifted absorbance of the dyes indicate excitonic (ground state) interactions between the dyes which interfere with the energy transfer between them [31].

Among the tested combinations, there are some remarkable examples in this array in which mixed energy transfer duplexes, meaning the combination of donor dyes linked to arabino-configured nucleosides (DNA2a) with acceptor dyes attached to

<table>
<thead>
<tr>
<th>Do → Ac</th>
<th>DNA3a and DNA3r</th>
<th>DNA2a and DNA2r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA2a</td>
<td>DNA2r</td>
</tr>
<tr>
<td></td>
<td>C   ΦF</td>
<td>C   ΦF</td>
</tr>
<tr>
<td>D5</td>
<td>a–a  35</td>
<td>0.146a</td>
</tr>
<tr>
<td></td>
<td>a–r  198</td>
<td>0.606</td>
</tr>
<tr>
<td></td>
<td>r–a  129</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>r–r  70</td>
<td>0.217</td>
</tr>
<tr>
<td>D6</td>
<td>a–a  15</td>
<td>0.146b</td>
</tr>
<tr>
<td></td>
<td>a–r  48</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>r–a  11</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>r–r  40</td>
<td>0.206</td>
</tr>
<tr>
<td>D7</td>
<td>a–a  44</td>
<td>0.273c</td>
</tr>
<tr>
<td></td>
<td>a–r  85</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>r–a  46</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>r–r  93</td>
<td>0.340</td>
</tr>
<tr>
<td>D8</td>
<td>a–a  109</td>
<td>0.606e</td>
</tr>
<tr>
<td></td>
<td>a–r  80</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>r–a  215</td>
<td>0.719</td>
</tr>
<tr>
<td></td>
<td>r–r  87</td>
<td>0.545</td>
</tr>
<tr>
<td>D9</td>
<td>a–a  60</td>
<td>0.307g</td>
</tr>
<tr>
<td></td>
<td>a–r  58</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>r–a  215</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>r–r  69</td>
<td>0.245</td>
</tr>
</tbody>
</table>

aλexc = 389 nm, λem = 515–800 nm; bλexc = 389 nm, λem = 525–800 nm; cλexc = 389 nm, λem = 550–800 nm; dλexc = 435 nm, λem = 550–800 nm; eλexc = 430 nm, λem = 560–800 nm; fλexc = 430 nm, λem = 540–800 nm; gλexc = 389 nm, λem = 530–800 nm; hλexc = 423 nm, λem = 550–800 nm.
ribo-configured nucleosides (DNA3r) and vice versa (DNA2r with DNA3a) yield significantly enhanced emission color contrasts. As a representative example, the fluorescence color readout for the combinations of D1 with D5 (Figure 1) ranges from green (DNA2aD1–DNA3aD5) to orange/red (DNA2rD1–DNA3aD5). Especially, the combination DNA2aD1–DNA3rD5 revealed a yellow-to-blue contrast of 198 and a quantum yield of 61%. For the blue–red emitting dye combinations the highest red-to-blue contrast of 215 and the highest quantum yield of 71% is achieved in DNA2rD1–DNA3aD8. Finally, among the broadest array of green–red fluorophore pairs there are a few remarkable duplexes with superior energy transfer parameters. Representatively, it is noteworthy that the combination DNA2aD3–DNA3rD7 gives a red-to-green contrast of 177 (and a quantum yield of 32%), and the combination DNA2rD4–DNA2aD8 shows a quantum yield of 53% (and a red-to-green contrast of 86).

In order to test the functionality of the respective dyes as FRET pairs in DNA duplexes for imaging in cells, four representative duplexes, DNA2aD1–DNA3rD5, DNA2rD1–DNA3aD8, DNA2aD2–DNA3aD8 and DNA2rD4–DNA3aD8, were tested in HeLa cells. 5 × 10^4 HeLa cells were transiently transfected with 15 pmol of the above mentioned DNA duplexes and Screenfect®, for 24 hours at a concentration, which was not toxic for the cells (see cytotoxicity test in Supporting Information File 1), and imaged by confocal fluorescent microscopy using the excitation wavelength of the energy donor (D1, λ_{exc} = 405 nm, D2, λ_{exc} = 488 nm, D4, λ_{exc} = 488 nm). To analyze the energy transfer to the energy acceptor the fluorescence of the energy donor (D1, λ_{em} = 435–470 nm (blue), D2, λ_{em} = 490–550 nm (green), Figure 2, left column) and the respective energy acceptor dye (D5, λ_{em} = 575–750 nm (yellow), D8, λ_{em} = 575–750 nm (red), Figure 2, middle column) was detected. In comparison to non-transfected control cells specific fluorescent staining could be observed in the perinuclear region, indicating that all dyes tested were endocytosed by the cells. The DNA duplexes preferentially accumulated in endosomal/lysosomal vesicles. The fluorescence of the energy donors, D1, D2 and D4 (Figure 2, left column), as well as the fluorescence of the energy acceptors, D5 and D8 (Figure 2, middle column), could be detected showing that fluorescence energy was transferred from the donor to the acceptor in the respective FRET pairs in the endosomal vesicles. This suggested that the DNA duplexes were still intact after transfection into cells.

**Conclusion**

The phosphoramidite 7 bearing the arabino-configured analog of uridine 2 that is additionally propargylated at the 2’-position was easily synthesized from commercially available nucleoside precursor 3 in 54% yield over five steps. The fluorescence quantum yields of oligonucleotides that were postsynthetically modified by the blue emitting dye D1 and the green-emitting dyes D2–D4 were improved due to the arabino-configured anchor 2 in comparison to the conventional ribo-configured uridine 1. This rather small structural difference allows the attached fluorophores to point into the major groove. Thereby optimized dye–DNA orientations result in higher fluorescence quantum yields of these single dye modifications. The modified oligonucleotides with dyes D1–D4 were applied as energy donors together with the correspondingly modified oligonucleotides bearing the acceptor dyes D5–D9. All dyes belong to our

---

**Figure 1:** Representative demonstration of the fluorescence readout differences between the four arabino/ribo combinations of D1 (donor) and D5 (acceptor). Left: Fluorescence of DNA2aD1–DNA3aD5; 2.5 µM DNA in 50 mM Na-Pi buffer, 250 mM NaCl, pH 7, λ_{exc} = 391 nm. Right: Corresponding image of cuvettes excited by a handheld UV lamp.
Figure 2: Confocal microscopy of HeLa cells after transfection with DNA2aD1–DNA3rD5 (row 1), DNA2rD1–DNA3aD8 (row 2), DNA2aD2–DNA3aD8 (row 3) and DNA2rD4–DNA3aD8 (row 4). The visualization was performed using a Leica TCS-SPE (DMi8) inverted microscope with an APO 63×/1.30 oil objective. For DNA2aD1–DNA3rD5 $\lambda_{\text{exc}} = 405$ nm (UV laser), $\lambda_{\text{em}} = 435–470$ nm (blue) and 575–750 nm (yellow), for DNA2rD1–DNA3aD8 $\lambda_{\text{exc}} = 405$ nm (UV laser), $\lambda_{\text{em}} = 415–550$ nm (blue) and 575–750 nm (red), for DNA2aD2–DNA3aD8 $\lambda_{\text{exc}} = 488$ nm (argon ion laser), $\lambda_{\text{em}} = 490–550$ nm (green) and 550–675 nm (red), for DNA2rD4–DNA3aD8 $\lambda_{\text{exc}} = 488$ nm (argon ion laser), $\lambda_{\text{em}} = 490–550$ nm (green) and 675–800 nm (red), scale bar = 20 µm.
DNA2aD1 to DNA2aD4, DNA2rD1 to DNA2rD4, DNA3aD5 to DNA3aD9 and DNA3rD5 to DNA3rD9 were purified using reversed-phase Supelcosil™ LC-C18 column (250 × 10 mm, 5 µm) on a Shimadzu HPLC system (autosampler, SIL-10AD, pump LC-10AT, controller SCL-10A, diode array detector SPD-M10A). Purification was confirmed by MS (MALDI–TOF) on a Biiflex-IV spectrometer from Bruker Daltonics in the linear negative mode (matrix: 1:9 mixture of diaminonium hydrogencitrate (100 g/L) and a saturated 3-hydroxypicolinic acid solution (10 g/L in 50% acetonitrile in water)). DNA concentrations were measured by their absorbance in water at 260 nm on a ND-1000 spectrometer from NanoDrop in the nucleic acid mode.

**Synthesis of 4.** 1-Deoxy-1-(uracil-1-yl)-β-D-arabinofuranose (3, 1.00 g; 4.10 mmol) was dried under reduced pressure for 1 h and was then dissolved in dry pyridine (5 mL). The reaction mixture was cooled to 0 °C and TIPDSiCl₂ (1.44 mL, 4.51 mmol) was slowly added. After 2 h, the reaction mixture was warmed to room temperature and stirred overnight. The solvent was removed under reduced pressure and the remaining solid was purified by flash chromatography (SiO₂, 0 → 50% EtOAc in CH₂Cl₂). 1.78 g (3.66 mmol, 89%) of 4 as a colorless solid were obtained. Spectral data were in accordance with the literature [32].

**Synthesis of 5.** Under argon atmosphere 4 (1.02 g, 2.10 mmol) was dissolved in dry THF (20 mL) and cooled to 0 °C with an ice bath. Then NaH (0.168 g, 4.20 mmol of 60% dispersion in mineral oil) was added and the reaction mixture was stirred for 15 min at 0 °C. The reaction mixture was warmed to room temperature and propargyl bromide (0.94 mL, 1.25 g, 8.40 mmol) was added slowly within 30 minutes. The reaction was stirred for 18 h at room temperature and quenched by adding distilled water (10 mL). The mixture was extracted with ethyl acetate (two times 100 mL). The combined organic layers were washed with saturated NaHCO₃ solution and then dried over Na₂SO₄. The solvent was removed under reduced pressure and the residue was purified by column chromatography (SiO₂, 0–40% EtOAc in hexane) to obtain 5 (0.716 g, 1.37 mmol, 65%) as a colorless foam. Rₚ 0.40 (hexane/EtOAc 1:1); ¹H NMR (400 MHz, CDCl₃) δ 8.48 (s, 1H, NH), 7.62 (d, J = 8.1 Hz, 1H, H-6), 6.22 (d, J = 6.0 Hz, 1H, H-1'), 5.85 (m, 1H, OH-3'), 5.69 (d, J = 8.2 Hz, 1H, H-5), 4.37 (dd, J = 7.7 Hz, 6.1 Hz, 1H, H-1'), 4.28–4.19 (m, 3H, OCH₃), 4.07 (dd, J = 13.2 Hz, 2.4 Hz, 1H, H-5'), 3.73 (dt, J = 8.6 Hz, 2.6 Hz, 1H, H-4'), 2.42 (t, J = 2.4 Hz, 1H, CH), 1.11–0.95 (m, 28H, & 4× CH₂) ppm; ¹³C NMR (75 MHz, CDCl₃) δ 160.1 (C-4), 150.4 (C-2), 140.9 (C-6), 101.9 (C-5), 82.2 (C-1'), 82.2 (C-4'), 80.4 (C-2'), 78.9 (C-CH₃), 75.5 (CH), 72.4 (C-3'), 60.5 (C-5'), 59.5 (OCH₃), 17.6–17.0 (8 CH₃), 13.6–12.5 (4 CH₂) ppm; FAB–MS m/z (%): 525.2 (65) [M + H]⁺; FAB–HRMS
FAB m/z: [M + H]^+ calcd for C_{24}H_{41}N_{2}O_{5}Si^2^+; 525.2447; found, 525.2447.

Synthesis of 2. Under an Ar atmosphere 5 (0.366 g, 1.30 mmol) was dissolved in dry THF (17 mL) 1 M tetrabutylammonium fluoride in THF (3.28 mL, 3.28 mmol) was added. The reaction was stirred for 5 min at room temperature. The reaction solution was directly poured onto a short silica plug and eluted with CH_{2}Cl_{2}/MeOH 5:1. The solvent was removed under reduced pressure and the crude product was purified by column chromatography (SiO_{2}; CH_{2}Cl_{2}/MeOH 10:1) to afford 2 (0.166 g, 89% as a colorless foam. R_{f} 0.24 (CH_{2}Cl_{2}/MeOH 9:1); ^{1}H NMR (300 MHz, DMSO-d_{6}) δ 11.33 (s, 1H, NH), 7.65 (d, J = 8.1 Hz, 1H, H-6), 6.13 (m, 1H, H-1'), 5.63 (m, 1H, OH-3') 5.60 (d, J = 8.1 Hz, 1H, H-5), 5.02 (m, 1H, OH-5'), 4.16 (d, J = 2.4 Hz, 2H, OCH), 4.13–4.02 (m, 2H, H-2', H-3'), 3.72–3.49 (m, 3H, H-4', 2 H-5'), 3.44 (t, J = 7.1 Hz, 2H, OCH), 3.89 (dt, J = 7.0 Hz, 5.7 Hz, 1H, H-2'), 4.28–4.10 (m, 3H, OCH), 3.93 (dd, J = 7.1 Hz, 3.6 Hz, 1H, H-3'), 3.52 (dd, J = 10.8 Hz, 3.6 Hz, 1H H-5'), 3.46 (dd, J = 10.8 Hz, 3.6 Hz, 1H H-5'), 2.49 (t, J = 2.4 Hz, 1H, CH) ppm; ^{13}C NMR (126 MHz, DMSO-d_{6}) δ 163.10 (C-4), 150.39 (C-6), 141.90 (C-2), 141.90 (C-6), 135.50 (C-1'), 135.50 (C-1'), 128.29 (C-4'), 128.29 (C-4'), 113.41 (C-7), 107.80 (C-2), 83.6 (C-5), 83.3, 81.0, 79.2, 77.4, 75.8, 74.0, 61.6, 59.0, 55.4 ppm; FAB−MS m/z (%): 242.3 (100) [M]^+ − CH_{2}CCH (propargyl).

Synthesis of 6. Under an inert gas atmosphere 2 (0.543 g, 1.26 mmol) was dissolved in dry pyridine (14 mL), 4,4'-dimethoxytrityl chloride (0.510 g, 1.51 mmol) was added in one portion and the reaction mixture was then stirred for 5 h at room temperature. The reaction was quenched by adding MeOH (5 mL) and the solvents were removed under reduced pressure. The residue was dissolved in EtOAc (20 mL). The organic layer was washed with 1 M aqueous NaHCO_{3} solution (3 times 20 mL), dried over Na_{2}SO_{4}, and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (SiO_{2}; CH_{2}Cl_{2}/MeOH 99:1 + 0.1% NEt_{3}) to afford 6 (0.729 g, 1.25 mmol, 99%) as a colorless foam. R_{f} 0.24 (CH_{2}Cl_{2}/MeOH 9:1); ^{1}H NMR (400 MHz, CDCl_{3}) δ 7.75 (d, J = 8.1 Hz, 1H, H-6), 7.44–7.22 (m, 9H, DMTr), 6.88–6.81 (m, 4H, DMTr), 6.28 (d, J = 5.8 Hz, 1H, H-1'), 5.43 (d, J = 8.2 Hz, 1H, H-5), 4.39 (dd, J = 7.0 Hz, 5.7 Hz, 1H, H-2'), 4.28–4.10 (m, 3H, OCH), 3.89 (dt, J = 7.1 Hz, 3.6 Hz, 1H, H-3'), 3.52 (dd, J = 10.8 Hz, 3.6 Hz, 1H H-5'), 3.46 (dd, J = 10.8 Hz, 3.6 Hz, 1H, H-5'), 2.49 (t, J = 2.4 Hz, 1H, CH) ppm; ^{13}C NMR (101 MHz, DMSO-d_{6}) δ 163.09, 158.81, 150.39, 144.52, 141.7, 135.5, 135.5, 130.3, 130.2, 128.3, 128.2, 127.3, 113.4, 101.7, 87.0, 83.6, 83.3, 81.0, 79.2, 77.4, 75.8, 74.0, 61.6, 59.0, 55.4 ppm; FAB−MS m/z (%): 855.1 (68) [M + H]^+; FAB−HRMS m/z: [M + H]^+ calcd for C_{33}H_{33}N_{2}O_{5}S^2^+; 585.2231; found, 585.2231.

Synthesis of 7. In a round bottom flask 6 (0.196 g, 0.34 mmol) was dried overnight under vacuum and then dissolved in dry CH_{2}Cl_{2} (5 mL) under an Ar atmosphere. N,N-Diisopropylethylamine (175 µL, 1.01 mmol) and 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (119 µL, 0.50 mmol) were added. The reaction mixture was stirred for 3 h at room temperature and then directly purified by column chromatography (SiO_{2}; CH_{2}Cl_{2}/acetone 5:1 + 0.1% NEt_{3}). 6 (0.253 g, 0.32 mmol, 95%) was obtained as a colorless foam. R_{f} 0.56 (CH_{2}Cl_{2}/acetone 5:1); APPI−MS m/z (%): 785.6 (70) [M + H]^+.

Preparation, purification and characterization of DNA. All oligonucleotides were synthesized on an Expedite 8909 Synthesizer from Applied Biosystems (ABI) using standard phosphoramidite chemistry. Reagents and CPG (1 µmol) were purchased from Proligo. The commercially available ribo-configured 2'-O-propargyluridine was purchased from ChemGenes. For the arabino-configured building block a slightly extended coupling time of 10 minutes was used. After preparation, the trityl-off oligonucleotides were cleaved from the resin and deprotected with conc. NH_{4}OH at 45 °C for 16 h.

Click reaction with modified oligonucleotides. To the lyophilized alkyl-modified DNA sample were added water (100 µL), sodium ascorbate (25 µL of 0.4 M in water), tris[(1-benzy1-1H-1,2,3-triazol-4-yl)methyl]amine (34 µL of 0.1 M in DMSO/t-BuOH 3:1), dye azide (114 µL of 0.01 M in DMSO/t-BuOH 3:1) and tetrakis(acenonitrile) copper(I) hexafluorophosphate (17 µL of 0.1 M in DMSO/t-BuOH 3:1). The reaction mixture was kept at 60 °C for 1.5 h. After cooling to room temperature, the DNA was precipitated by adding Na_{2}EDTA (150 µL of 0.05 M in water), sodium acetate (450 µL of 0.3 M in water) and ethanol (10 mL, 100%) and stored at −32 °C for 16 h. After centrifugation, the supernatant was removed and the residue washed two times with cold ethanol (2 mL, 80%). The dried DNA pellet was then further purified via HPLC as further described in Supporting Information File 1.

Cell experiments and confocal fluorescence microscopy. Human cervix carcinoma cells (HeLa cells) were cultured in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% fetal calf serum and 1% penicillin/streptomycin at 37 °C in a 5% CO_{2} atmosphere. 24 h before transfection 5 × 10^6 HeLa cells per well were seeded in an 8-well chamber slide (µ Slide 8 well ibiTreat, IBIDI, Martinsried, Germany) in 200 µL of media. For the transfection 15 pmol of the respective DNA duplexes were diluted in ScreenFect®A dilution buffer (Incella, Egggenstein-Leopoldshafen, Germany) to a final volume of 9 µL. 12 µL of a 1:10 dilution of ScreenFect®A in dilution buffer were added to the diluted DNA and rapidly mixed. A subsequent incubation time of 20 min at room temperature allowed the formation of lipoplexes (liposome–DNA complexes). The transfection mixture was then added to the cells.
The cells were incubated for 24 h with the respective transfection mixture at 37 °C in a 5% CO₂ atmosphere. The visualization of the DNA duplexes was performed by confocal laser scanning microscopy using a Leica TCS SPE (DMi8) inverted microscope with an ACS APO 63×/1.30 oil objective. Fluorophores were excited using an UV laser (405 nm) for duplexes DNA2aD1–DNA3rD5 and DNA2rD1–DNA3aD8 and an argon ion laser (488 nm) for duplexes DNA2aD2–DNA3aD8 and DNA2rD4–DNA2aD8. The emission detection bandwidths were at 435–470 nm (blue) and 575–750 nm (yellow) for DNA2aD1–DNA3rD5, 415–550 nm (blue) and 575–750 nm (red) for DNA2rD1–DNA3aD8, 490–550 nm (green) and 550–675 nm (red) for DNA2aD2–DNA3aD8, 490–550 nm (green) and 675–800 nm (red) for DNA2rD4–DNA2aD8. Using the acquisition software Leica Application Suite (LAS) X 2.0.1.14392, the picture ratio was adjusted to 1024 × 1024 pixels 8 bit depth.

Supporting Information
Supporting File 1
Additional data and spectra. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-16-S1.pdf]

Acknowledgements
Financial support by the Deutsche Forschungsgemeinschaft (Graduiertenkolleg 2039-1) and grant Wa 1386/17-1) and by Helmholtz program BIF-TM is gratefully acknowledged. B.O. was supported by a fellowship of the Landesgraduiertenförderung.

References

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.13.16
Versatile synthesis of the signaling peptide glorin

Robert Barnett¹, Daniel Raszkowski², Thomas Winckler² and Pierre Stallforth*¹

Abstract

We present a versatile synthesis of the eukaryotic signaling peptide glorin as well as glorinamide, a synthetic analog. The ability of these compounds to activate glorin-induced genes in the social amoeba Polysphondylium pallidum was evaluated by quantitative reverse transcription PCR, whereby both compounds showed bioactivity comparable to a glorin standard. This synthetic route will be useful in conducting detailed structure–activity relationship studies as well as in the design of chemical probes to dissect glorin-mediated signaling pathways.

Introduction

The emergence of multicellularity from unicellular ancestors is considered a major evolutionary transition [1]. This transition has occurred not only once, in fact more than 25 independent instances of this event are known. The resulting increase in biological complexity requires fine-tuned differentiation and cell–cell communication mechanisms. The social amoebae are exquisite organisms to study the emergence of multicellularity since they can exist in both a unicellular and a multicellular stage with a well-orchestrated developmental cycle linking the two [2]. The unicellular amoebae feed on bacteria and divide by binary fission. Upon depletion of their food source, they aggregate to form a multicellular organism. Eventually, they culminate in fruiting bodies to spread some of the population as dormant spores into the environment. Secondary metabolites often constitute the key signaling molecules in these developmental processes [3,4]. For instance, aggregation of the amoebae is initiated by pulses of chemoattractive, low-molecular weight signaling molecules – so-called acrasins [5]. Additionally, it has been shown that natural products are also involved in interspecies interactions of social amoebae and bacteria [6-10]. A detailed investigation of both inter- and intraspecies interactions will give insight into the fundamentals
of cell signaling and access to a rich source of novel natural products.

We describe a practical synthesis of the modified dipeptide glorin (1, Figure 1), the assumed acrasin for many of the early-diverged species of social amoebae. While numerous species of social amoebae such as Polysphondylium pallidum, Dictyostelium fasciculatum [11], and D. caveatum [12], amongst others respond chemotactically to glorin (1), the acrasin has only been isolated from P. violaceum [13]. Despite its crucial role in the initiation of multicellularity, little is known about glorin’s biosynthesis, signaling pathways, or degradation. To facilitate further studies, our chemical route allows for a facile synthesis of glorin derivatives and glorin-based chemical probes.

Here, we report the synthesis of glorin (1), as well as the novel synthetic analog glorinamide 2: a compound with comparable bioactivity, that is hydrolytically – and thus metabolically – more stable than glorin (1) [14]. The molecules were shown to be bioactive, effectively mediating the induction of gene expression during early development, as determined by quantitative reverse-transcription PCR (RT-qPCR).

Results and Discussion
Synthesis of glorin and glorinamide
Two glorin syntheses have been published [15,16], one of which lacked sufficient data to be reproducible and the other one displayed limited versatility. Therefore, we focused on designing a robust synthesis that would allow for facile access to glorin derivatives required for structure–activity relationship studies. Eventually, these studies can lead to the construction of various chemical probes to identify the unknown glorin receptor.

Syntheses of glorin and glorinamide (Scheme 1) started from commercially available L-ornithine (3) and benzoxycarbonyl-

![Figure 1: Glorin (1) and glorinamide 2.](image)

![Scheme 1: Synthesis of glorin (1) and glorinamide 2.](image)

Scheme 1: Synthesis of glorin (1) and glorinamide 2. Reagents and conditions: a) TMSCl, MeOH, rt, 12 h; b) NaOEt, EtOH, 0 °C to rt, 30 min, 97% over two steps; c) HCl, p-TsOH, toluene, Dean–Stark conditions, 3 h, 76%; d) for 7a: NaOEt, EtOH, 0 °C to rt, 30 min, 76%, for 7b: H2NET, THF, rt, 16 h, 75%; e) for 8a: isobutyl chloroformate, NMM, 4, DMF, −15 °C to rt, 2 h, 69%, for 8b: HBTU, Et3N, 4, DMSO, rt, 3 h, 69%; f) Pd/C, H2, MeOH, rt, 1 h, 74% 9a, 97% 9b; g) iPr2EtN, DMAP, propionic anhydride, DCM, rt, 2 h, 92% 1, 97% 2.
protected L-glutamic acid 5. In an improvement on previous two-step syntheses, L-ornithine δ-lactam 4 was synthesized from L-ornithine in a one-pot procedure, whereby the latter was converted into the corresponding methyl ester using trimethylsilyl chloride in methanol [17] and cyclization was achieved under basic conditions using sodium ethoxide; the lactam was prone to racemization under strongly basic conditions, this was avoided by short reaction times with sodium ethoxide, and by avoiding strongly basic reaction conditions in subsequent steps. A key challenge in the syntheses of glorin and analogs is the differentiation between the α- and the γ-carboxylic acid groups of L-glutamic acid for selective esterification or amidation. α-Selective functionalization was achieved via synthesis of oxazolidinone 6 from Cbz-L-glutamic acid (5) with paraformaldehyde and p-toluenesulfonic acid under dehydrating conditions [18]. Opening of the oxazolidinone with sodium ethoxide as nucleophile thus yielded ester 7a, while addition of ethylamine yielded amide 7b. Subsequent amide bond formation with lactam 4 using isobutyl chloroformate or HBTU as coupling reagents furnished the protected dipeptides 8a and 8b, respectively. Hydrogenolysis with hydrogen gas and palladium on charcoal gave free amines 9a and 9b. Glorin (1) and glorinamide 2 were then obtained by treating amines 9a and 9b, respectively, with propionic anhydride. The main advantage of our synthesis over previous syntheses is that ours allows for late-stage functionalization of the α-amino group of the glutamic acid moiety. This is particularly useful for the rapid generation of different α-amide analogues.

Biological activity of glorin and glorinamide

The bioactivities of synthetic glorin (1) and glorinamide 2 were assayed by their ability to elevate the expression of the glorin-induced gene PPL_09347 in the social amoeba P. pallidum, as previously determined [11]. We chose the gene PPL_09347, which encodes the Dictyostelium discoideum ortholog of profilin I, an actin binding protein required for the actin cytoskeleton organization, as a model gene. Upon glorin (1) exposure, this gene was found to be up-regulated by about 50-fold [11]. To this end, the compounds were dissolved in DMSO/water and added in two 1 μM portions (30 min apart) to a suspension of starving P. pallidum. After 1 h, the cells were harvested and total RNA was extracted. The differential regulation of PPL_09347 was determined by RT-qPCR. Synthetic glorin (1) and glorinamide 2 led to similar expression of PPL_09347, comparable to commercial glorin as positive control (Figure 2). While a small baseline induction of PPL_09347 without test compounds is always observed (Figure 2, top), induction by glorin and glorinamide led to a significant increase in expression level above baseline (Figure 2, bottom).

Conclusion

In summary, we have devised a versatile and robust synthetic route to glorin that allows for a wide range of derivatizations. Glorin, as well as the hydrolytically more stable derivative glorinamide, were shown to display comparable glorin-induced gene expression in Polysphondylium pallidum. In future this synthesis will facilitate the construction of a library of glorin derivatives for a detailed structure–activity relationship study. Ultimately, we wish to synthesize chemical probes of glorin for the identification of the unknown glorin receptor.
Experimental

Quantitative reverse transcription PCR: P. pallidum PN500 cells were cultured in association with Escherichia coli K12 cells. Cells were harvested before first signs of aggregation became visible. Cells were washed three times in 17 mM phosphate buffer (pH 6.2) to remove bacteria and suspended at 2 × 10^7 cells/mL in 17 mM phosphate buffer (pH 6.2) with shaking at 150 rpm. After 1 h of starvation, 1 µM glorin (Phoenix Pharmaceuticals, Burlingame CA, USA), synthetic glorin (I), or glorinamide 2 (100 µM stock solutions with 3% DMSO) or water were added to the cells every 30 min for 1 h. Cells were centrifuged for 30 min after the last addition and stored in pellets of 2 × 10^7 cells at −80 °C.

Total RNA was prepared using the QIAGEN RNeasy kit and cDNA was synthesized using the QIAGEN Omniscript kit and an oligo(dT) primer. Expression of the glorin-induced model gene PPL_09347 was determined by RT-qPCR as described [11,19]. Expression of PPL_09347 in different cDNA samples was standardized to the reference gene glyceraldehyde-3-phosphate dehydrogenase (gpdA, SACGB accession number PPL_12017; http://sacgb.fli-leibniz.de/cgi/index.pl).

Supporting Information

Supporting Information File 1
Detailed experimental procedures, compound characterization data, and copies of NMR spectra. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-27-S1.pdf]

Acknowledgements

We thank A. Perner and H. Heinecke (Hans Knöll Institute, Jena) for MS and NMR measurements. We are grateful for financial support from the Leibniz Association, and the Fonds der Chemischen Industrie (VCI) (P.S.). This work was also supported by the DFG-funded graduate school of excellence Jena School for Microbial Communication (R.B. and D.R.).

References

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.13.27

Revaluation of biomass-derived furfuryl alcohol derivatives for the synthesis of carbocyclic nucleoside phosphonate analogues

Bemba Sidi Mohamed, Christian Périgaud and Christophe Mathé*

Abstract

The racemic synthesis of new carbocyclic nucleoside methylphosphonate analogues bearing purine bases (adenine and guanine) was accomplished using bio-sourced furfuryl alcohol derivatives. All compounds were prepared using a Mitsunobu coupling between the heterocyclic base and an appropriate carbocyclic precursor. After deprotection, the compounds were evaluated for their activity against a large number of viruses. However, none of them showed significant antiviral activity or cytotoxicity.

Introduction

Biomass is a valuable resource in search of renewable organic carbon sources for the future and can be used to produce a range of chemical building blocks. The latter products can further be transformed to value-added compounds that are suitable for supplement or replacement to oil-derived chemicals. One such building block is furfuryl alcohol which is obtained through the catalytic hydrogenation of furfural; the latter is obtained from the dehydration of xylose, a 5-carbon sugar derived from vegetal biomass. Furfuryl alcohol finds widespread application in the chemical industries and for example is employed for the production of synthetic fibers, fine chemicals, etc. In fine organic chemicals synthesis, furfuryl alcohol is a raw material for the production of tetrahydrofurfuryl alcohol, which is an intermediate for the synthesis of 1,2- and 2,5-pentanediols and their derivatives and an agent for the manufacture of fragrance, vitamin C and lysine [1,2]. Furfuryl alcohol is also the source of 4-hydroxy-2-cyclopentenone, which, in enantiopure form, has been used as an intermediate for the synthesis of natural products and pharmaceutical drugs [3]. Recently, racemic (+/−)-4-hydroxy-2-cyclopentenone has found application in the synthesis of nucleoside analogues [4,5] and some of the products have shown interesting antiviral activities. As part of our studies on carbocyclic nucleoside phosphonates [6] as potential anti-HIV agents [7,8], we envisioned to use bio-sourced racemic (+/−)-4-O-protected 2-cyclopentenone for the synthesis of hitherto unknown carbocyclic nucleoside methylphosphonates (Figure 1) bearing purine bases (adenine and guanine) in order to evaluate their antiviral properties.
Results and Discussion
Synthesis of precursors (+/-)-5 and (+/-)-7
The synthesis began with the preparation of racemic 4-O-TBDMS-2-cyclopentenone (1) which was obtained in two steps from commercially available furfuryl alcohol following a reported procedure (Scheme 1) [9]. Addition of the carbanion, generated in situ from dimethyl methylphosphonate and n-butyllithium in dry THF at -78 °C, to compound (+/-)-1 gave cyclopentenyl alcohol (+/-)-2 stereoselectively through a 1,2-addition mechanism. The stereochemistry of (+/-)-2 may be ascribed to a nucleophilic attack of the incoming nucleophile from the less-hindered face due to the presence of the silyl protective group. 1H and 13C NMR spectra were in accordance with the presence of a sole diastereoisomer.

The reduction of compound (+/-)-2 readily provided the saturated derivative (+/-)-3 and the free tertiary hydroxy groups in (+/-)-2 and (+/-)-3 were then protected using acetic anhydride. Finally, removal of the TBDMS groups afforded the appropriate carbocyclic precursors, namely, (+/-)-1-((dimethoxyphosphoryl)methyl)-3-hydroxycyclopentyl acetate (5) and (+/-)-1-((dimethoxyphosphoryl)methyl)-4-hydroxycyclopent-2-en-1-yl acetate (7). The protection of the tertiary hydroxy groups was necessary in order to avoid competing side reactions during the coupling reaction under Mitsunobu conditions [10]. Compounds (+/-)-5 and (+/-)-7 were used as suitable precursors for the synthesis of the target carbocyclic methylphosphonates.

Synthesis of cyclopentyl carbocyclic methylphosphonates (+/-)-12 and (+/-)-13
The synthesis of the target compounds was accomplished using a Mitsunobu reaction (Scheme 2) [11]. The coupling of (+/-)-5 with bis-Boc-adenine or 2-amino-6-chloropurine in the presence of disopropyl azodicarboxylate (DIAD) and PPh3, provided the N9 carbocyclic nucleosides (+/-)-8 and (+/-)-9 as racemic mixtures. No concomitant formation of the N7 regiosomer was observed.
The removal of Boc and acetyl groups from (+/-)-8 afforded the carbocyclic phosphonate (+/-)-10. Subsequently, the phosphonoester protecting groups were cleaved in the presence of TMSCl and NaI in a CH3CN/DMF mixture to give (+/-)-12 in 61% yield. The treatment of (+/-)-9 with K2CO3 in methanol at room temperature gave compound (+/-)-11 which upon reaction with TMSBr in DMF led to a deprotection of the diester groups as well as a concomitant hydrolysis of the methoxy group. Both compounds, (+/-)-12 and (+/-)-13, were obtained as their sodium salts, after reversed phase column chromatography and ion exchange chromatography.

Synthesis of cyclopentenyl carbocyclic methylphosphonates

Compounds (+/-)-16 and (+/-)-21

The synthesis of the cyclopentenyl carbocyclic derivatives was envisioned from the precursor (+/-)-7 (Scheme 3) using Mitsunobu conditions. A coupling reaction of (+/-)-7 with bis-Boc-adenine [12] afforded the desired adduct (+/-)-14 with 56% yield. The removal of the Boc groups was achieved following a similar protocol as developed for compound 8.

Surprisingly, the treatment of compound (+/-)-14 under acidic conditions lead to the removal of the Boc group accompanied with an unexpected transposition of the allyl moiety and a concomitant loss of AcOH leading to compound (+/-)-15 with 85% yield. The 1,3-allylic transposition of the hindered tertiary alcohol group under acidic conditions has not been reported yet for such compounds. It may conceivably that such a transposition occurs through the formation of an allylic carbocation which upon reaction with water as an incoming nucleophile, afforded compound (+/-)-15 with a higher substituted double bond. After deprotection of the phosphonate diester groups, structural assignments of (+/-)-16 were based upon 1H and 13C NMR spectra and correlation experiments, which showed some characteristic features compared to the parent derivative (+/-)-14. In particular, the chemical shifts of C2’, C3’ and C1’ carbon atoms showed differences (Table 1). In case of com-
Scheme 3: Reagents and conditions: (a) N6-bis-Boc-adenine, PPh3, DIAD, THF, rt, 56%; (b) TFA, Cl(CH2)2Cl, rt, 85%; (c) TMSCl, NaI, CH3CN, 40 °C to rt, 60%.

Table 1: Selected chemical shifts in 13C NMR.\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>(+/-)-14</th>
<th>(+/-)-16</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3'</td>
<td>134.5</td>
<td>81.5</td>
</tr>
<tr>
<td>C2'</td>
<td>138.2</td>
<td>125.6</td>
</tr>
<tr>
<td>C1'</td>
<td>88.7</td>
<td>143.0</td>
</tr>
</tbody>
</table>

\(^a\)\(\delta\) in ppm of C2', C3' and C1'.

Additionally, in order to confirm the stereochemistry of the allylic alcohol in position 3', a NOESY correlation experiment was accomplished with compound (+/-)-16 (Figure 3). We have observed a correlation between proton H8 and H3' confirming the orientation of the 3'-OH.

Figure 3: Selected NOESY correlations for compound (+/-)-16.

From the results obtained with adenine, we envisioned to synthesize the parent nucleoside of (+/-)-16 bearing guanine as the base. We chose as a precursor of the heterocyclic base the commercially available 2-amino-6-methoxypurine which upon treatment with Boc2O afforded the suitable heterocyclic precursor 17 (Scheme 4). The coupling reaction of (+/-)-7 and 17 using the Mitsunobu reaction gave a separable mixture of N9/N7 regioisomers (+/-)-18 and (+/-)-19 with 55% and 5% yield, respectively. After purification, compound (+/-)-18 was treated under acidic conditions to remove the Boc group as well as to induce a transposition of the allyl moiety in a similar manner to the one previously observed with compound (+/-)-14. Finally, treatment of (+/-)-20 with TMSBr in DMF led to the cleavage of the phosphonoester groups and concomitant hydrolysis of the methoxy group. Compound (+/-)-21 was obtained as sodium salt after reversed phase column chromatography and ion exchange chromatography and its structural assignment was based upon \(^1\)H and \(^13\)C NMR spectra and correlation experiments. It is noteworthy that the carbocyclic methylphosphonates (+/-)-16 and (+/-)-21 have structural similarity with carbonucleosides belonging to the neplanocin family, in particular, with neplanocin F [13].

Compounds (+/-)-26 and (+/-)-27

Based on the observation that carbocyclic methylphosphonates derived from the Mitsunobu coupling reaction are sensitive to treatment in acidic medium, the synthesis of the target compounds (+/-)-26 and (+/-)-27 was envisaged through the use of precursors without acid-labile protecting groups. Thus, reaction of (+/-)-7 with \(\text{N}^6\)-Bz-adenine [14] or 2-amino-6-chloropurine in the presence of PPh3 and DIAD in THF provided the Mitsunobu adducts (+/-)-22 and (+/-)-23 with 17% and 53% yield, respectively (Scheme 5). A lower yield was observed for the coupling reaction of \(\text{N}^6\)-Bz-adenine with (+/-)-7.
compared the reaction with of N^6-bis-Boc-adenine (17% versus 56%, Scheme 3). In both cases, no formation of the N7 alkylation product was observed.

**Scheme 4**: Reagents and conditions: (a) i) Boc_2O, DMAP, THF, rt; ii) K_2CO_3, MeOH, 75%; (b) PPh_3, DIAD, THF, rt, 55% for 18; 5% for 19; (c) TFA, Cl(CH_2)_2Cl, rt, 51%; (d) TMSBr, DMF, 0 °C to rt, 59%.

The treatment of (+/-)-22 and (+/-)-23 with K_2CO_3 in methanol at room temperature afforded compounds (+/-)-24 and (+/-)-25. Cleavage of the ester groups and in case of (+/-)-25, hydrolysis of the methoxy group, was achieved by reaction with TMSBr in DMF. Compounds (+/-)-26 and (+/-)-27 were obtained as their sodium salts after reversed phase column chromatography and ion exchange chromatography.

**Conclusion**

In summary, we have developed a methodology for the synthesis of carbocyclic nucleoside phosphonate analogues through the use of bio-sourced furfuryl alcohol derivatives. The methodology involved the preparation of the proper carbocyclic phosphonate precursors which upon Mitsunobu reaction with the appropriate heterocyclic bases afforded the protected target intermediates. Some unsaturated derivatives have shown instability in acidic medium and underwent an unexpected 1,3-allylic transposition giving rise to carbocyclic nucleoside phosphonates having structural similarity with carbonucleosides belonging to the neplanocin family. All the newly synthesized compounds were evaluated for their antiviral properties against HIV-1, Zika virus, Dengue-2 virus, HSV-1, HSV-2 and Chikungunya virus. However, none of them showed significant antiviral or cytotoxic activities. The absence of biological activity may be attributed to various factors, such as inability to enter
cells or to behave as substrates for intracellular enzymes catalyzing phosphorylation, as well as a lack of inhibition of viral polymerases by their diphospho–phosphonate forms.

Supporting Information
Supporting Information File 1
Synthetic details and characterization data of new compounds. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-28-S1.pdf]

Supporting Information File 2
Copies of NMR spectra for the synthesized compounds. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-28-S2.pdf]

Acknowledgements
B.S.M. is particularly grateful to the Fondation de Coopération Scientifique Infectiopole Sud for doctoral fellowship.

References
Polyketide stereocontrol: a study in chemical biology

Kira J. Weissman

Abstract

The biosynthesis of reduced polyketides in bacteria by modular polyketide synthases (PKSs) proceeds with exquisite stereocontrol. As the stereochemistry is intimately linked to the strong bioactivity of these molecules, the origins of stereochemical control are of significant interest in attempts to create derivatives of these compounds by genetic engineering. In this review, we discuss the current state of knowledge regarding this key aspect of the biosynthetic pathways. Given that much of this information has been obtained using chemical biology tools, work in this area serves as a showcase for the power of this approach to provide answers to fundamental biological questions.

Introduction

Reduced polyketides and their derivatives form the basis for a number of medicines in current clinical usage, notably antiinfectives [1] (e.g., erythromycin A (1) and its semi-synthetic derivatives azithromycin (2), clarithromycin (3), telithromycin (4) ([2] and others) and anticancer compounds (e.g., ixabepilone (5) [3]), a semi-synthetic derivative of the natural product epothilone B (6)) (Figure 1). Given the medical and economic importance of these compounds, there is significant interest in trying to generate new versions of polyketides for evaluation as drug leads. The significant bioactivity of these compounds derives from their complex structures (particularly when compared to the typical products of chemical synthesis [4]), which incorporate both high functional group density and rich stereochemistry. These features, coupled with the fact that the majority of reduced polyketides are macrocyclic, result in significant in-built conformational constraints. As a consequence, these molecules present their diverse functionality in a defined way in three dimensions, allowing them to bind their biological targets with useful affinity (10⁻⁷ to 10⁻⁹ M [4]).
Erythromycin A (1, Figure 1) is the prototypical polyketide, as its biosynthesis has been studied most heavily to date. The structure incorporates 10 stereocenters, and so in principle, 1024 (2^{10}) different stereoisomers are possible. Yet, nature reliably assembles only one stereoisomer (at least at detectable levels), at once revealing the strict stereocontrol underpinning the pathway and the importance of synthesizing this particular version. Indeed, the crystal structure of erythromycin A (1) bound to the 50S ribosomal subunit of the eubacterium *Deinococcus radiodurans* [5], shows a suite of interactions between the ribosomal bases and multiple chiral functional groups of the polyketide macrolactone, including the hydroxy groups at C-6, C-11 and C-12, and the desosamine appended to the hydroxy group at C-5 (whose positioning in 3D depends on the hydroxy stereochemistry) (Figure 2).

This intimate link between polyketide stereochemistry and biological activity makes the control of stereochemistry an attractive research area for attempts to generate new polyketide structures by synthetic biology [6]. The aim of this review is to trace how our understanding of these feature of the biosynthesis has developed, and more specifically, the critical role that an array of chemical biology approaches [7] has played in furnishing the underlying data. These include, but are not limited to, the synthesis of isotopically-labeled precursors and the analysis of the resulting labeling patterns, characterization by assays in vitro of wild type and mutant recombinant enzymes in the presence of synthetic substrates, and genetic engineering of model systems coupled with analysis of product structures by gas-chromatography/mass spectrometry (GC–MS) and liquid chromatography (LC)–MS.

**Review**

**Biosynthesis of complex polyketides by modular PKSs and stereochemical considerations**

The reduced or complex class of polyketides is assembled in bacteria by gigantic multienzymes called polyketide synthases (PKSs), in a process resembling fatty acid biosynthesis by the mammalian fatty acid synthase (FAS) [8] from which the PKSs likely evolved [9]. In both cases, simple acyl-CoA building blocks are concatenated head-to-tail to construct linear chains. Several features distinguish these two pathways, however: PKSs use a wider range of both initial building blocks (referred to as ‘starter units’) and chain extension units than their FAS counterparts which notably results in branching of the chains, the degree of reduction of the initially formed C3-keto intermediates is variable (whereas in FA biosynthesis, full reduction to the fatty acyl group occurs systematically), and polyketides are most typically released in cyclic form, whereas fatty acids are liberated as carboxylic acids. The much more complicated
biosynthetic control in PKSs is achieved by successive action of multiple FAS-like modules (hence the name ‘modular PKS’ for this type of system), each of which carries out a single round of chain extension and chemical tailoring of the resulting intermediate.

Each PKS module incorporates three functional domains necessary for chain growth (Figure 3): an acyl transferase (AT) which selects the appropriate precursor from the cellular pool, a ketosynthase (KS) which extends the chain via a Claisen-like decarboxylative condensation, and a non-catalytic acyl carrier protein (ACP) to which the intermediates are covalently tethered through a phosphopantetheine prosthetic group. The modules can also incorporate a variable complement of the processing activities which act in each cycle of FA biosynthesis, including ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains; these activities lead successively to hydroxy groups, olefinic moieties or saturated methylene groups at specific positions in the polyketide chains. Building of the polyketide core is typically terminated by a thioesterase (TE) domain situated at the end of the final PKS multienzyme, which releases the product by hydrolysis or more usually macrolactonization, using an internal hydroxy nucleophile. This PKS-free intermediate (6-deoxyerythronolide B in the case of erythromycin biosynthesis, Figure 3) is then frequently modified by a series of so-called ‘post-PKS enzymes’ (e.g., methyl transferases, hydroxylases, and glycosyl transferases), to achieve its final bioactive form [10].

Nature has, in fact, evolved two distinct types of modular PKSs, referred to as cis-AT (including the erythromycin PKS (Figure 3)) and trans-AT (Figure 4). The principle distinguishing feature for trans-AT systems is the absence of an AT domain integrated into the subunits, as the activity is instead present as a discrete protein which acts iteratively to furnish extender unit to the modules [11]. Other characteristic features include unusual domain orderings, duplicated and inactive domains, atypical enzymatic functions, and modules distributed between two subunits (so-called ‘split modules’). This architectural divergence in all likelihood reflects independent evolutionary paths of the two types of systems [12], although more recent evidence indicates that some trans-AT PKSs may have evolved from a cis-AT parent [13].

In terms of stereochemical considerations, however, they are largely the same for the two systems, as stereochemistry can be introduced at several points within the pathways. For example, although fatty acids are constructed primarily from malonyl-CoA units, the AT domains of cis-AT PKSs exhibit specificity towards a number of branched extender units (including methylmalonyl-CoA, ethylmalonyl-CoA, hydroxymalonyl-ACP, methoxymalonyl-ACP, etc. [15,16]), thus incorporating pendant functionality into the polyketide skeleton (C-2-methyl, C-2-ethyl, C-2-hydroxy and C-2-methoxy groups, respectively). In the case of erythromycin A (1) (Figure 2 and Figure 3), for example, the C-2-methyl groups resulting from use of methylmalonyl-CoA exhibit both possible stereochemistries. In
Figure 3: Schematic of the biosynthetic pathway leading to erythromycin A (1) in the bacterium *Saccharopolyspora erythraea*. The first stage of biosynthesis occurs on a modular polyketide synthase (PKS) incorporating three gigantic multienzyme polypeptides, DEBS 1, 2 and 3. Each of these subunits contains two chain extension modules, while DEBS 1 additionally incorporates a loading module to initiate the biosynthesis, and DEBS 3, a termination module, consisting of a thioesterase (TE) domain. Each of the chain extension modules includes three essential domains (ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP)), and a variable complement of processing activities (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)). The number and character of the variable domains correlates precisely with the structure of the resulting intermediates (the building block added at each stage is color-coded to match the module responsible). Following release of the chain from the PKS, the first free intermediate, 6-deoxyerythronolide B, is further modified by a series of post-PKS enzymes to yield the final, bioactive metabolite.
methyl configurations in the final polyketide products could arise by judicious choice by the PKS AT domains of one or the other enantiomer. The first information on extender unit selection in polyketide biosynthesis was provided in the mid-1980s via feeding of isotopically-labeled precursors to whole cells of the erythromycin producer *Saccharopolyspora erythraea*, leading to the generation of isotopically-labeled (2S)- and (2R)-methylmalonyl-CoA in situ [23]. When a precursor of (2S)-methylmalonyl-CoA, [2-2H2,2-13C]propionate, was used, analysis of the products by difference 13C{1H, 2H} NMR provided evidence for isotopic labeling at C-2, C-4, and C-10 of the macrolide ring. This result was consistent with incorporation of (2S)-methylmalonyl-CoA during the second, fifth, and sixth chain extension cycles, with inversion of configuration at the C-2 center as found for fatty acid biosynthesis (vide infra) [24]. However, attempts to illuminate the origin of the remaining centers by feeding of ethyl [2-2H2,2-13C]-succinate to produce labeled (2R)-methylmalonyl-CoA in situ, were inconclusive.

Access to the erythromycin PKS (DEBS) multienzymes as pure proteins [25] allowed extender unit preference in *cis*-AT PKSs to be investigated under more controlled conditions in vitro. Critically, the researchers were able to generate exclusively 14C-(2S)- or (2R)-methylmalonyl-CoA (7 and 8, respectively) by enzymatically removing the enantiomeric substrate under conditions designed to minimize spontaneous epimerization. Using the resulting enantiomeric materials, it was then shown by autoradiography that acylation of all six DEBS proteins is highly specific for the (2S)-isomer (7) [26], implying that the six AT domains present in the multienzymes select exclusively this stereoisomer (Figure 5).

Subsequent studies in vitro with a model recombinant protein, DEBS 1-TE (Figure 6), confirmed that this preference is also exercised during chain extension [27]. The DEBS 1-TE protein was created by joining the terminal TE domain to the end of the bimodular first subunit, DEBS 1 [28], to cause release of the
Figure 5: Determination of the stereochemistry of extender unit selection by the AT domains of modular PKS. a) Enantiomerically pure $^{14}$C-labeled (2$S$)-methylmalonyl-CoA (7) and (2$R$)-methylmalonyl-CoA (8) were generated enzymatically from a racemic mixture by consumption of the opposite enantiomer under conditions designed to minimize spontaneous epimerization. b) Results of labeling of the DEBS proteins with pure $^{14}$C-(2$R$)- or (2$S$)-methylmalonyl-CoA. Only incubation with (2$S$)-methylmalonyl-CoA (7) produced radioactive labeling of the three DEBS proteins (the signal obtained for DEBS 1 with (2$R$)-methylmalonyl-CoA (8) is due to the presence of a small amount of propionyl-CoA contaminant, which labeled the loading module. Image adapted from [26].

Figure 6: Creation by genetic engineering of the DEBS 1-TE model system. The region of the eryAIII gene encoding the thioesterase (pink) was relocated to the end of gene eryAI. The resulting protein, DEBS 1-TE, produces a small triketide lactone 9 instead of the heptaketide 6-deoxyerythronolide B. The two methyl centers in lactone 9 are of opposite stereochemical configuration, and thus DEBS 1-TE represented an ideal system for elucidating the origin of the two configurations. In the presence of a suitable starter unit such as propionyl-CoA, (2$R$)-methylmalonyl-CoA as extender unit and NADPH (the cofactor for the KR domains), no product was observed. However, when (2$S$)-methylmalonyl-CoA was provided instead, the product was obtained at a satisfactory rate, showing that both modules
select this isomer. Thus, the idea that one of the methyl configurations arises from use of (2R)-methylmalonyl-CoA by the analyzed modules and the second from use of the (2S)-isomer, was now firmly excluded. Given the high level of homology among many AT domains from cis-PKSs [29], it is likely that all such acyl transferases exhibit the same stereospecificity.

The AT domains operate by a ping-pong bi-bi mechanism [30], in which the initially formed acyl-O-AT intermediate is subject to nucleophilic attack by the terminal phosphopantetheine thiol of the ACP domain. Recent steady-state kinetic analysis of an AT domain sourced from DEBS module 3 (Figure 3) has provided evidence that the specificity for (2S)-methylmalonyl-CoA is expressed during the first half reaction of the ping-pong mechanism (i.e., formation of the methylmalonyl-O-AT intermediate) [30]. Substrate preference can be rationalized, at least in part, by bioinformatics which has revealed several sequence motifs correlating with building block choice (whether for starter or extender units, malonyl or branching extender units) [31-37], in combination with structure elucidation at high resolution of AT5 from the DEBS PKS, which was solved in the presence of acetate (Figure 7) [38].

For example, extender unit-specific ATs contain positively charged residues in the active site (R667 and H745, DEBS AT5 numbering) capable of forming salt bridges with the carboxyl group of the building block, while these are non-polar amino acids in starter-unit specific ATs. The choice of methylmalonyl-CoA over malonyl-CoA is correlated with a YASH motif some 100 residues downstream of the active site serine, whereas malonyl-CoA specific ATs exhibit an alternative HAFH sequence (Figure 7a) [38]. In the AT5 crystal structure, the Tyr, Ser and His all lie within the active site (the His is the second member of the catalytic dyad). This leads to a model in which the C-2-methyl of methylmalonyl-CoA forms favorable hydrophobic interactions with the Tyr while being sterically accommodated by the relatively small Ser (Figure 7b). Finally, stereospecificity for the (2S)-isomer appears to lie in steric clashes that would occur between a (2R)-methyl group and both the Ser and His of the YASH motif. Nonetheless, efforts in vivo to convert methylmalonyl-CoA-specific ATs into malonyl-CoA-specific ATs by exchange of these key sequence motifs resulted only in promiscuous ATs capable of recognizing both extender units [32,34,36], revealing that further elements of the AT active site contribute to specificity.

![Figure 7: Model for substrate selection by AT domains.](image)
In terms of the stereochemistry of the less common extender units, labeling studies indicate that the (2S) isomer of ethylmalonyl-CoA is also used [39], which correlates with it originating predominantly from the reductive carboxylation of crotonyl-CoA [40]. Several extender units including aminomalonyl-ACP [41] and hydroxy-/methoxymalonyl-ACP [42] are generated via multi-step pathways from a primary metabolite, with the intermediates tethered to a discrete ACP domain. The building blocks are then transferred onto the AT domains of the PKS, and from there to the downstream integral ACP to participate in chain extension. Based on the presumed biosynthetic origin of these extender units (from L-serine and from a glycolytic intermediate (in all likelihood 1,3-biphospho-D-glycerate), respectively), it was initially proposed that the (2S)-isomer of aminomalonyl-ACP and the (2R)-isomers of hydroxyl-/methoxymalonyl-ACP [39] are employed. However, more recent crystallographic work on the zwittermicin pathway [43] in which hydroxymalonyl-ACP is used as extender unit, has raised some uncertainty over the hydroxymalonyl stereochemistry, as the (2S)-isomer would appear to fit better within the investigated AT structure. Indeed, selection of the (2S)-isomer, and correspondingly, the (2R) isomer of aminomalonyl-ACP, would simplify the biosynthetic mechanism in a number of polyketide pathways, as subsequent epimerization of the resulting pendant centers (vide infra) would not be required.

Ketosynthases

The next step in the biosynthetic cycle is KS-catalyzed chain extension. This reaction occurs by nucleophilic attack of an enolate generated by decarboxylation of an ACP-bound extender unit onto the starter unit or chain extension intermediate attached to the active site cysteine of the KS domain. The face of the enolate which is used for the attack determines whether the reaction occurs with retention or inversion of configuration at the C-2 center relative to the starting material (Figure 8).

In the related FAS enzymes, this reaction has been shown to proceed with inversion of stereochemistry at the extender unit C-2 [24]. Circumstantial evidence for this same condensation stereochemistry in cis-AT PKSs was obtained for at least a subset of modules in the DEBS PKS by the feeding studies in Sac. erythraea cited previously, but direct proof that inversion occurs was provided by experiments in vitro with DEBS 1-TE [46]. In this study (Figure 9), (2RS)-[2-2H]methylmalonyl-CoA (10) was prepared and provided to DEBS 1-TE (along with starter unit butyryl-CoA (11) and NADPH (12)), knowing that solely the (2S)-isomer would be utilized. Analysis by mass spectrometry and NMR of the triketide lactone product 13 showed that only a single deuterium label was retained at the C-2 position bearing the D-configured methyl group (generated by module 2), while no labeling was observed at C-4 bearing the L-configured methyl group (generated by module 1). The opposite labeling pattern was obtained when biosynthesis was carried out with unlabeled (2RS)-methylmalonyl-CoA in D2O.

These labeling patterns are consistent with inversion of stereochemistry occurring in both modules 1 and 2 as in fatty acid biosynthesis without cleavage of the C-2–H bond (giving directly the D-configuration at C-2 observed in the final product), but show that an additional epimerization step must occur in module 1 to yield the L-methyl stereochemistry present at C-4 (thus explaining the loss of deuterium from the 2-position when (2RS)-[2-2H]methylmalonyl-CoA was used, and its incorporation from solvent in the presence of unlabeled extender) (mechanism III, Figure 9).

Although these experiments established the stereochemistry of condensation, showing it to furnish directly the D-methyl groups of polyketides, the origin of the epimerization activity in module 1 remained obscure. Shortly thereafter, the results of genetic engineering experiments carried out on DEBS KS1 implicated this domain as the seat of this activity, with a downstream KR then choosing between the two methyl configurations presented to it by the KS. More specifically, when KS1 was paired with the remaining domains of DEBS module 2 (AT2, KR2 and ACP2), and the hybrid module sandwiched between the DEBS loading module and the TE, the resulting construct produced a diketide 14 with opposite stereochemistry to that normally generated by module 1 (Figure 10a) [47]. This result was taken to show that KS1 can produce both methyl stereochemistries, but that in the hybrid 1/2 diketide synthase, the selectivity of KR2 for the unpimerized methyl configuration masks the KS1 epimerase activity. Subsequent work seemed to strengthen the idea that KS1 acts as an epimerase [48]. In this case (Figure 10b), the loading module-KS1 portion of DEBS 1 was grafted onto DEBS 3 (whose two modules 5 and 6 generate the unpimerized methyl configuration) to generate a hybrid PKS called TKS-AR1, and the stereochemistry of the resulting triketide lactones 15 and 16 established by NMR. This analysis showed that the methyl group arising from the hybrid 1/5 module was epimerized in 50% of the product 16 and that this change in stereochemistry was propagated to the ketoreduction in module 6, despite the lack of methyl group epimerization in this module. Thus, it appeared from these experiments that introduction of KS1 into a normally non-epimerizing context was sufficient to alter the methyl configuration, consistent with its role as an epimerase. However, as it has now been clearly established that it is instead the KR domains that possess this activity, it must be assumed that the engineered synthase suffered a significant change in architecture which allowed the epimerization to happen spontaneously, perhaps by
Figure 8: Proposed mechanism for KS-catalyzed chain extension, based on extrapolation from studies on homologous enzymes from animal FAS [44]. The reaction encompasses two stages overall: a) acyl transfer, and b) the Claisen-like condensation. From the stereochemical perspective, the important aspect of the mechanism is that the C-2-methyl stereochemistry is set by the direction of attack of the enolate nucleophile on the acyl enzyme carbonyl (reaction bii). (Although several elements of this mechanism differ from that proposed more recently in [45], including the roles of the His residues in the acyl transfer reaction, and whether decarboxylation proceeds with initial attack by a water molecule, these do not have stereochemical consequences).

providing water with increased access to the chain extension intermediates. How the KR domains were shown to participate in epimerization will be detailed below.

Ketoreductases
KRs catalyze the stereospecific reduction of the C-3-ketone groups arising from the chain extension reaction, to give both possible stereoisomers of the resulting hydroxy groups. The direction of reduction is intrinsic to the KR domains, as the majority of KR domains transplanted by genetic engineering into alternative contexts have maintained their native stereospecificity [49-51]. Incubation of enzymatically-generated, chirally deuterated NADPH (both (4R)- and (4S)-[4-2H]NADPH) with modules 1, 2, 5 and 6 from the DEBS PKS and analysis of the resulting products by GC-MS, showed that all of the KR domains are specific for the 4′-pro-S hydride of the nicotinamide cofactor [52,53], as found for fatty acid biosynthesis [54,55]. Given the high sequence similarity among KR domains from modular PKS
systems, it is likely that this hydride specificity is common to all of them. Indeed, the 8 KR structures solved to date (7 from *cis*-AT PKSs [56-62] and 1 from a *trans*-AT PKS [63]) show the domains to adopt the same overall fold and share a conserved active site architecture. These analyses have revealed the KRs to be monomeric proteins containing a catalytic subdomain and a catalytically-inactive structural subdomain, both of which exhibit a Rossmann fold. Within the catalytic subdomain, all reductase active KRs possess the active site tetrad of Tyr, Ser, Lys and Asn [64] characteristic of the short-chain dehydrogenase/reductase (SDR) superfamily [65,66], and bind the NADPH cofactor in the same orientation so that it presents its 4′-pro-S hydride to the active site. Therefore, the alternative directions of keto-reduction (referred to as A- and B-type [67] to avoid ambiguity, as the *R*/S designations can vary depending on the relative priority of the functional groups) are thought to arise from opposite modes of binding into the common active center (i.e., the binding modes are related by a 180° rotation around the axis of the target carbonyl, with either the re or si face of the C-3-keto group presented to NADPH) (Figure 11).

Achieving these alternative modes of binding necessitates that the substrate enter from one or the other side of the KR, as appropriate. Several sequence motifs (referred to here as the ‘Caffrey motifs’) correlating with the direction of reduction and therefore presumably guiding substrate entry, were initially identified by comparative sequence analysis [64,67], and shown subsequently by structural analysis to occupy positions proximal to the active site [56-62]. The strongest indicator for a B-type KR domain is an LDD motif in the region between amino acids 88 and 103 (numbered as in [67]) which is absent from *A*-type KR domains (B-type KR domains in *trans*-AT PKSs appear only to conserve the second D [63]). These residues lie on a flexible loop (the ‘lid loop’) adjacent to the

---

**Figure 9:** Experiment in vitro to determine the stereochemistry of condensation in modular PKS [46]. Use of specifically C-2-deuterium labeled extender unit **10** during biosynthesis with DEBS 1-TE (alongside starter unit butyryl-CoA **11** and NADPH **12**), resulted in a labeling pattern in the triketide lactone product **13**, which allowed discrimination between the four possible mechanisms for condensation in modules 1 and 2 of the PKS (the C-2 methyl center of the product is established by module 2 and the C-4 center by module 1). The obtained pattern (exclusive deuterium labeling at the C-2 position) was consistent with mechanism III (boxed) – inversion of stereochemistry in both modules as found for fatty acid synthase, with an additional epimerization occurring in module 1 to give the observed final configuration.
active site. Additional amino acids in the 134–149 region, specifically P144 and N148, correlate with B-type KR's, while W141, which is located on the opposite side of the substrate-binding groove to the LDD motif, is most strongly indicative of an A-type KR. Nonetheless, despite the availability of multiple ketoreductase structures, the role of these residues in shepherding the substrates into their correct orientations remains unclear, possibly because none of the KR's was co-crystallized as a ternary complex with both native polyketide intermediate and cofactor.
To date, two alternative mechanisms have been proposed to account for substrate positioning. In the first [57], ‘southeast’ entry (A-type reduction) is the default, and from this direction the phosphopantetheine arm of the ACP can contact the conserved W. In B-type KRs, on the other hand, the southeast side of the active site is blocked by an interaction between the LDD and the ‘lid helix’ (a mobile \(\alpha\)-helix adjacent to the NADPH cofactor), which prevents the phosphopantetheine arm from slipping between them. The intermediate therefore enters the active site from the ‘northwest’ side, where the phosphopantetheine can make favorable interactions with the conserved Leu. In the alternative proposal [62], the direction of reduction is controlled by a divergent degree of ordering within the active sites of A- and B-type domains. In A-type KRs, cofactor binding generates a well-organized and catalysis-ready active site, in which a key residue (Met in the solved structure upon which the mechanism was based [62]) blocks entry from the northwest, allowing the substrate to penetrate the active site groove only from the southeast. The characteristic W of this type of KR points into the southeast entry channel, where it may help orient the phosphopantetheine cofactor by hydrogen bonding. In contrast, in B-type KRs, cofactor binding is loose, allowing in principle the polyketide to enter from both sides of the channel. However, only binding of substrate from the northwest side results in a catalysis-competent conformation of the active site. In this model, the LDD motif does not interact directly with substrate, but may contribute to substrate-assisted assembly of the active site. (For more recent ideas on substrate guiding, see [69]).

Although KRs catalyze reduction from one or the other direction in their native contexts, for many KRs, this strict control is at least partially lost in vitro. Assays of KR activity have been carried out with model synthetic substrates in the context of native and engineered modules [48,70,71] and with KRs obtained as isolated domains [58,61-63,68,72-76]. In the majority of cases, the substrate used was the synthetically accessible \((2\text{RS})-2\text{-methyl-3-oxopentanoic acid} \text{N-acetylcysteamine} (\text{NAC})\) thioester (’\(\beta\)-keto diketide’) 17 – a racemic analogue of the diketide generated by condensation of a propionyl starter unit and a \((2\text{S})\)-methylmalonyl extender unit. NAC was chosen as the activating group because it mimics the terminal portion of
the phosphopantetheine cofactor to which the chain extension intermediates are normally tethered. The stereochemistry of the reduction products was typically established by GC–MS and comparison to authentic synthetic standards, or alternatively by LC–MS. Analysis of results obtained with KR domains from the DEBS [68], tylosin (Tyl) [68,72] and amphotericin [58,61] PKSs (Figure 12), showed that when the KR domains selected the correct stereoisomer at the C-2 methyl position, reduction occurred almost exclusively in the native direction; the same result was obtained for certain of these KR domains with diketide and triketide intermediates generated enzymatically in situ on ACP domains [77], a process leading only to the correct C-2 methyl isomer (vide infra). However, when the incorrect methyl isomer was chosen and reduced (which in some cases was the kinetically favored outcome [68]), reduction occurred in both the native and reverse directions (Figure 12). Thus, in these instances, a change in methyl stereochemistry was sufficient to flip the substrate in the active site, suggesting that the energetic differences between the two binding modes are minor. (The caveat with these results is that reduction might still have followed the natural course even in the presence of the ‘wrong’ methyl stereochemistry if the substrates more closely resembled the native ones and/or the substrates were attached to an ACP domain (the result, for example, of tethering (2R)-2-methyl-3-oxopentanoate to an ACP, has not been tested)). In any case, these data encouraged the view that mutation of a few key residues in the KR active sites might be used to alter reduction stereochemistry.

Site-directed mutagenesis can indeed modify the stereochemical outcome of ketoreduction, at least in vitro, showing that the altered residues do play some role in stereocontrol. This is notably the case for changes introduced into the two Caffrey motifs. For example, swapping the B-type motifs of DEBS KR1

---

Figure 12: Assays in vitro to evaluate the stereospecificity of recombinant KR domains. A series of KR domains of all types (A1, A2, B1 and B2, as indicated) were investigated using a common substrate, racemic diketide derivatized as its N-acetylcysteamine (SNAC) thioester 17. When the KR domains selected the appropriate substrate (C-2 non-epimerized (NE) for the type 1 domains and C-2 epimerized (E) for the type 2 domains), reduction occurred almost exclusively in the correct direction. However, when substrate bearing the non-native C-2 stereochemistry was chosen, the opposite direction of reduction was also observed. The percentages of each product obtained are shown.
for characteristic A-type residues, yielded a KR which catalyzed exclusively A-type reduction of β-keto diketide \(17\) [73]. Unexpectedly, however, carrying out the reverse changes with the A-type DEBS KR2 produced a mutant KR with increased wild type behavior towards the model substrate, and thus the targeted residues cannot be the sole determinants of the direction of ketoreduction. Similarly, high-throughput mutagenesis of the Caffrey motifs [74] and other residues identified as potentially participating in stereocontrol by sequence and/or structural analysis [58,61], failed to produce consistent results, with certain mutations leading to the predicted shift in stereochemistry, others again strengthening wild type behavior, and still others having no effect on the stereochemical outcome. Most importantly, introducing the same Caffrey motifs mutations into DEBS KRs 1 and 2 housed within DEBS 1-TE produced no discernable stereochemical switch in vivo [78]. These results clearly show that within the context of an intact PKS multienzyme, other factors override the effects of these mutations; possibilities include the specificity of one or more downstream domains acting against stereochemically altered intermediates, increased hydrolytic removal of stalled chains via the proof-reading activity of a cluster-associated TEII domain [79-81], or constraints imposed on substrate orientation due to the fact that the intermediates are tethered to ACP domains which are themselves covalently linked to the KRs; which if any of these mechanisms predominates remains to be determined. In the meantime, it has proven more successful to swap entire KR domains both within and between PKS multienzymes, as a means to achieve rationale alteration of C-3-hydroxy stereochemistry [49-51].

As noted previously, following chain extension in certain modules, the initial D-2-methyl group undergoes an epimerization reaction to yield the L-methyl. In mechanistic terms, epimerization involves removal of the C-2 proton and delivery of proton to C-2 from the opposite face of the resulting, planar enol/enolate intermediate. Monitoring by NMR of the rate of epimerization of a model C-3-ketoacyl ester (ethyl 2-methylacetoacetate) showed this reaction to be rapid at room temperature \(\left(t_{1/2} = 4.7\ \text{min}\right)\) [82]. Thus, during polyketide biosynthesis, there must be some mechanism to protect the intermediate from spontaneous epimerization following chain extension, both as it is passed between the KS and KR active sites and within the KR prior to ketoreduction (the alternative possibility that epimerization occurs in all modules but that the KRs select the correct isomer is excluded by the in vitro studies with DEBS 1-TE (Figure 9) [46], as no deuterium from the deuterated extender unit would have been retained in the triketide lactone product). Sequestering of intermediates by the ACP domains has been proposed as a source of configurational stability at C-2 [82]. However, as all direct study of this question to date by NMR has failed to reveal any direct contact between modular PKS ACP domains and their attached substrates [83-85], the origin of this stabilization remains unknown.

KRAs were first suggested to act as epimerases – despite the fact that no other SDR enzyme exhibits this activity – based on structural analysis [57]. This proposal led to the classification of PKS KRs working on C-2 methylated substrates into six distinct categories – KRs catalyzing A- and B-type reduction in the absence of epimerization (A1 and B1, respectively), KRs catalyzing both epimerization and the two senses of ketoreduction (A2 and B2), reductive- and epimerization-inactive KRs (C1), and KRs catalyzing epimerization in the absence of reduction (C2). (KRAs operating on substrates lacking C-2 methyl groups are referred to as A0 and B0). The first direct proof for this activity was provided by studies on reconstituted modules (combinations of individually purified KS-AT didomains, ACPS and KRs) [77]. In brief, the notable finding of this work was that the stereochemical outcome both at C-2 and C-3 of the products correlated not with the modular origin of the KS-AT and ACP domains, but with that of the KR. For example, combining KS-AT and ACP from DEBS module 6 (which produces a non-epimerized methyl and an A-type C-3-hydroxy) with DEBS KR1 in the presence of starter unit (synthetic propionyl-SNAC), NADPH (12) and (2RS)-methylmalonyl-CoA, resulted in diketide with the stereochemistry at both the C-2 methyl and ACP domains working on C-2 methylated substrates into six distinct categories – KRs catalyzing both epimerization and the two senses of ketoreduction (A2 and B2), reductive- and epimerization-inactive KRs (C1), and KRs catalyzing epimerization in the absence of reduction (C2). (KRAs operating on substrates lacking C-2 methyl groups are referred to as A0 and B0). The first direct proof for this activity was provided by studies on reconstituted modules (combinations of individually purified KS-AT didomains, ACPS and KRs) [77]. In brief, the notable finding of this work was that the stereochemical outcome both at C-2 and C-3 of the products correlated not with the modular origin of the KS-AT and ACP domains, but with that of the KR. For example, combining KS-AT and ACP from DEBS module 6 (which produces a non-epimerized methyl and an A-type C-3-hydroxy) with DEBS module 6 KR, yielded diketide incorporating the stereochemistry of module 6 (unepimerized C-2 methyl and A-type reduction at C-3) (Figure 13). Thus, these experiments provided the first conclusive evidence that certain KR domains can control the stereochemistry at both C-2 and C-3 of the chain extension intermediates.

The most convincing evidence has emerged from a series of so-called ‘equilibrium isotope exchange (EIX)’ experiments [86-88] – another clear illustration of the power of chemistry to elucidate key aspects of stereocontrol. In these assays (Figure 14a), the epimerization activity of select KRs (DEBS KR1, nystatin (Nys) KR1 and rifamycin (Rif) KR; was demonstrated directly by incubating them with the stereochemically characteristic of module 6 (unepimerized C-2 methyl and A-type reduction at C-3) (Figure 13). Thus, these experiments provided the first conclusive evidence that certain KR domains can control the stereochemistry at both C-2 and C-3 of the chain extension intermediates.
Figure 13: Assays in vitro which provided the first direct evidence that KR domains act as epimerases [77]. Biosynthesis in these experiments was carried out by reconstituted modules comprising KS-AT didomains and isolated KR and ACP domains. The KS was charged with the propionate starter unit from propionyl-SNAC, and the ACP with extender unit by the AT domain using (2\text{S})-methylmalonyl-CoA. The product diketides were hydrolyzed from the ACP domains and their stereochemistries were determined by chiral GC–MS. Reconstituted DEBS modules 1 and 6 (red and yellow, respectively) gave predominantly the expected products (epimerized C-2 methyl and B-type alcohol stereochemistry for module 1; non-epimerized C-2 methyl and A-type alcohol stereochemistry for module 6 (indicated in bold)), while exchanging only the KR domain caused the reconstituted modules to produce the product characteristic of the introduced KR (for example (line 3), KR\text{1} in place of KR\text{6} resulted in the native product of module 1). Thus, the KRs were shown to control the stereochemistry at both the C-2 and C-3 positions of the chain extension intermediates.

Reduced forms. Under these conditions, time-dependent washout of deuterium from the C-2 position (above background) occurred for epimerizing KRs as they are capable of racemizing this position once the C-3-keto is present, while the label remained intact for two model, non-epimerizing KRs (DEBS KR\text{6} and Tyl KR\text{1}), as confirmed by LC–MS analysis [89] of the reduced products (while chiral GC–MS was used to confirm that no change in configuration of the reduced product occurred).

This assay was subsequently extended to demonstrate the intrinsic epimerase activity of specific non-reducing KRs [87]. In this ‘tandem EIX’ format (Figure 14b), the ketoacyl substrate for the KR to be assayed is generated transiently from the appropriate reduced product by a second, validated non-epimerizing KR, at which point, the intrinsic epimerase activity of the target KR is again evidenced by time-dependent washout of the C-2 deuterium label. Using this coupled assay, epimerase activity was established for two natively non-reducing (C2-type) KRs (DEBS and pikromycin (PIKS) KR\text{3}), as well as redox-defective mutants of DEBS KR\text{1} obtained by site-directed inactivation of the NADPH-binding site.

The tandem assay strategy was also used to try to identify residues potentially participating in the epimerization reaction [88]. This is an intriguing question, as comparative sequence analysis [57,90] fails to reveal any residues which are differentially and strictly conserved in epimerizing KRs relative to non-
epimerizing KRs, which could serve as catalytic general acids and bases. In these in vitro experiments, it was possible to decouple the role of the active site Tyr and Ser residues in the reduction and epimerization reactions by assaying two natively redox-inactive but epimerizing KRs (DEBS and PIKS KRs 3; C2-type), along with a validated redox active but epimerization inactive KR domain (DEBS KR6). The fact that the mutant KRs lost a substantial percentage of their epimerase activity implicates both of these residues in the epimerization reaction. On the other hand, it is not clear how the same residues can function in both capacities – what, for example, inhibits the reduction occurring prior to epimerization if identical amino acids are involved? One possibility, which has not previously been discussed in the literature, is that in fact, epimerizing KRs bind their substrates in two distinct modes. In the first, which is only available to the substrate bearing the non-epimerized methyl center, the thioester and C-3-keto groups are aligned so that the $pK_a$ of the C-2 proton is suitably depressed, allowing facile catalysis by KR residues or alternatively abstraction by an available water molecule. The resulting enol/enolate could tautomerize spontaneously back to the original substrate or its epimer, with only the epimerized substrate possessing the required C-2 methyl stereochemistry for subsequent reduction. This epimer would then bind in a second mode common to all KRs of the same type (i.e., either A- or B-), in which it is positioned properly relative to the reductive catalytic apparatus. In this way, the KRs could effectively discriminate between substrates bearing the two methyl stereochemistries. This mechanism might be borne out by the first crystal structures of epimerizing KRs in the presence of native substrate, and such data are eagerly anticipated. In the meantime, in the absence of a clear mechanistic basis for epimerization, it has been shown...
possible to rationally alter the methyl stereochemistry (both introducing and removing C-2 epimerization) by the whole-sale exchange of KR domains in the context of model PKS systems [51,61,91], although the efficiency of such experiments remains generally low.

**Dehydratases**

PKS DHs are members of the double hot dog (DHD) family of enzymes, in which the active site in one of the two fused single-hot dog subdomains is inactive [92-95]. The DHs catalyze the elimination of water from the polyketide intermediates to form double bonds which are typically trans (E) in configuration, although cis (Z) alkenes are also present in a significant fraction of structures [94]. Studies on the evolutionary related DHs from animal FAS which produce exclusively trans double bonds [96] have demonstrated that this reaction proceeds with overall syn elimination of the pro-(2S) hydrogen and the (3R)-hydroxy group [97,98], while biochemical and stereochemical experiments on this class of enzymes suggest a catalytic mechanism in which a single histidine plays the role of both general acid and base [97-99]. Extending this proposal to PKS DHs which operate on C-2 methylated intermediates, implies that only D-methylated ((2R), unepimerized) compounds will be substrates for the DHs, as then the C-2 proton is of the correct stereochemistry. In this model (Figure 15), whether trans or cis double bonds are obtained directly by syn elimination depends on the hydroxy configuration, with (3R)-hydroxy groups (B-type ketoreduction) leading directly to trans double bonds and (3S)-hydroxy groups (A-type) giving cis double bonds [95]. These alternative reaction courses could be achieved by a common mode of binding into the DH active site, where all that differs is the direction in which the remainder of the chain (R in Figure 15) points [95]. To summarize: According to this proposal, the KR domains ultimately determine whether or not dehydration can occur (it should not occur for intermediates in which the C-2 methyl is epimerized, because the C-2 proton is inaccessible to the DH catalytic apparatus) and the stereochemistry of the resulting double bonds (via A- or B-type reduction on unepimerized chains; or in other words, (2R,3R) intermediates yield trans double bond sterechemistry, and (2R,3S), cis double bond sterechemistry).

A number of experiments reported to date support the origin of trans double bonds from a B-type hydroxy precursor. Specifically, studies in vitro with recombinant DH domains from the DEBS [100] and nanchangmycin [101] PKSs on ACP-tethered substrates generated in situ from reconstituted modules acting on synthetically-prepared diketide SNACs, showed that both domains generated trans double bonds from the corresponding (2R,3R)-2-methyl-3-hydroxyacyl chains. In the case of the DEBS DH, no reaction was observed for the three diastereomeric substrates (e.g., (2S,3R), (2R,3S) and (2S,3S)). A model DH from the Tyl PKS was also assayed with a panel of substrates, and found to recognize B-type and not A-type alcohols, generating exclusively trans double bonds [102].

On the other hand, the origin of cis double bonds is much less obvious. The only clear evidence to date for cis bond formation by a specific module comes from studies of the phoslactomycin PKS; here, only diketide incorporating a cis double bond was shown to productively prime the second chain-extension module, implying that it must be the product of the first module (which incorporates an A-type KR) [103]. Cis double bond formation was also explored by studies in vitro with a DH from module 10 of the Rif PKS [94]. In this case, the natural DH substrate had previously been shown to contain (2S,3S)-2-methyl-3-hydroxyacyl functionality [104], which is not in accord with the (2R,3S) stereochemistry postulated for cis-double bond pre-

![Figure 15: Model for DH-catalyzed generation of trans and cis double bonds by syn elimination from substrates bearing C-2-methyl groups (the side chains shown are those found in DEBS DHs [92]). Both substrates incorporate a (2R)- (non-epimerized, NE) methyl group, but differ in the relative orientation of the C-3 H and R groups within the active site (boxed). In both cases, however, the hydroxy group is held in place by interactions with Tyr and Asp residues. Syn dehydration of the product derived from B-type reduction (indicated) directly yields a trans double bond, while syn dehydration of an A-type hydroxy substrate directly gives a cis double bond. Adapted from [95].](image-url)
cursors. Indeed, dehydration by recombinant Rif DH of this substrate (18) tethered to the native ACP domain resulted in a trans double bond, while none of the diastereomeric substrates were active (Figure 16a). Intriguingly, the diastereospecificity of the dehydration was completely reversed when the polyketide chain was attached to a non-cognate ACP from DEBS, with the (2R,3R) isomer (19) now being dehydrated to a trans double bond (Figure 16b). This is apparently the first example of such reversal of diastereospecificity due to the nature of the thioester conjugate. Taken together, these results agree with a common syn dehydration mechanism for PKS DHs, but the requirement that the abstracted proton be positioned equivalently relative to

Figure 16: Stereospecificity of dehydration by Rif DH$_{10}$ [94]. a) The four possible diastereomeric diketide-ACP substrates were prepared by stereospecific C-3 reduction (and C-2-methyl epimerization in the case of the A2 and B2-type KR$_6$s) of (2RS)-2-methyl-3-ketopentanoyl-Rif ACP$_{10}$ 18. Only the substrate bearing (2S,3S)-2-methyl-3-hydroxy stereochemistry was accepted as substrate by Rif DH$_{10}$, yielding the trans double bond isomer. b) The alternative (2R,3R)-2-methyl-3-hydroxy substrate was prepared on DEBS ACP$_6$ via condensation of propionyl and methylmalonyl building blocks by KS$_6$, followed by stereospecific reduction by Tyl KR$_1$. Syn dehydration again yielded the trans double bond.
the conserved His for both the (2S,3S) and (2R,3R) polyketide chains, requires that the substrate enters from the opposite side of the DH active site. Nonetheless, this alternative direction of entry mechanism could explain the surprising observation that some A-type KR domains are found in modules producing trans double bonds [105].

In any case, the obtained data failed to shed light on the origin of this cis double bond in the final rifamycin structure, although clearly it arises from isomerization of an initially formed trans alkene. Indeed, further studies on the Rif PKS and a handful of other systems have revealed that at least a subset of cis double bonds in the products arise from mechanisms other than direct DH-catalyzed dehydration. These include isomerization by either integral enoyl isomerase domains (present in trans-AT PKSs only) [21,22,106] or by post-PKS domains [107-109], and TE-mediated formation from a B-type alcohol precursor [110]. It may thus be the case that all PKS DHs produce trans double bonds. Consistent with this idea, comparative sequence analysis and the resolution of six DH crystal structures to date (1 from DEBS (trans-double bond producing module) [92], 4 from the curacin PKS (1 cis and 3 trans) [93], and 1 from the Rif PKS (cis) [94]) have not revealed any notable differences between DHs giving rise to trans double bonds and those apparently responsible for direct cis double bond formation. In terms of the catalytic mechanism, a two-base mechanism has been proposed based on the crystal structure of the DEBS DH [92], in which the conserved His acts as a general base to deprotonate at C-2, while an Asp residue serves as a general acid to stabilize the C-3 hydroxy leaving group. However, only the His has been shown by site-directed mutagenesis to be essential [111,112], and so definitive proof of whether the classic one-base mechanism mentioned earlier [97-99] or alternative two-base mechanism applies [113] remains to be obtained.

Enoyl reductases

The enoyl reductase domains act on trans double bonds, producing fully-saturated methylene groups. In fatty acid biosynthesis by animal FAS, this reaction proceeds with attack of the 4′-pro-R hydride of NADPH on the 3-re face of the unsaturated thioester intermediate, with stereospecific protonation at the 2-si face, giving an overall syn addition [114]. In the case of polyketide chains, when a C-2 methyl substituent is present, enoyl reduction has stereochemical consequences, producing both the (2R)- and (2S)-configurations depending on which side of the double bond is protonated. As for the KR domains, by comparative sequence analysis of PKS ERs, a correlation was uncovered between the presence of specific residues and the direction of reduction [115] (Figure 17). When the identified position, which lies some 90 residues upstream of the conserved NADPH-binding motif, is occupied by a tyrosine residue, the methyl branch has an S configuration. In domains producing the alternative R configuration, this residue is most often valine, but also alanine or phenyalanine.

The role of these residues in stereocontrol was evaluated in vivo by site-directed mutagenesis of a derivative of DEBS 1-TE in which the KR domain of module 2 was replaced with the ‘reductive loops’ (DH-KR-ER tridomain) sourced from the DEBS and RAPS PKSs (giving TKS-ery4 and TKS-rap13, respectively) [115]. Native TKS-ery4 produces a triketide lactone 20 with a (2S)-methyl, while TKS-rap13 yields the alternative (2R)-methyl. Dramatically, when the conserved Y of the DEBS ER in TKS-ery4 was replaced with V, the resulting lactone 21 incorporated exclusively the (2R)-methyl (Figure 18a). This result showed that this residue is involved in ER stereocontrol. Nevertheless, the equivalent mutation introduced into the ER of TKS-rap13 (V to Y) did not result in the predicted change in stereochemistry at C-2 to S, with only parental product obtained. In subsequent experiments [116], 4 additional residues characteristic of (2S) specific domains were introduced simultaneously into the RAPS ER within the same model system, but this yielded only a small overall shift in stereochemical outcome.

On the other hand, mutagenesis of a putative catalytic residue (a Lys) without changing the Val had a more dramatic effect on stereochemistry [116]. To explain this result, it is proposed that the Lys serves as proton donor at C-2, and in its absence, there is less control of the face to which the proton is added from solvent to the carbonyl intermediate. Based on the high-resolution structure of a representative ER from the spinosyn PKS [60], this mechanism has been extended to account for the role of the conserved Tyr (Figure 18b). In the solved structure, the Lys and the Tyr lie on opposite sides of the active site cleft from one another, in appropriate positions to protonate the C-2 carbon of a bound polyketide substrate. When the Tyr is present, it acts as the proton donor, but in its absence (as in the native PKSs in which V is instead present), the Lys delivers its proton from the opposite side of the polyketide substrate (thus explaining the reversal in stereochemistry observed with the Y to V mutant in TKS-ery4). Clearly, however, simply introducing Tyr at the appropriate position into (2R)-producing ERs (as in the experiments with TKS-rap13) is not sufficient to override proton donation by the Lys, and so rational manipulation of ER stereochemistry by site-directed mutagenesis awaits identification of further stereochemical determinants in ER active sites.

Bioinformatics-guided structure elucidation

The strong correlations between certain sequence motifs present in PKS domains and the stereochemistry of the resulting polyketide chains has been exploited in several cases to predict
and/or corroborate absolute stereochemical assignments made on newly-discovered natural products (for example, elansolid [117], the disciformycins [118], hygroabafilomycin [119], phormidolide [120], and haprolid [121]), either by manual inspection of domain sequences or by more sophisticated methods including hidden Markov model (HMM)-based sequence classification (e.g., ScoreDiff [90]). This is notably the case for the direction of ketoreduction by assignment of KR as either A- or B-type. Such analyses are relatively straightforward when the canonical ‘Caffrey’ motifs are present (for example, the LDD motif indicative of B-type KR, with the second D being most diagnostic, particularly for trans-AT PKSs) and therefore these predictions can be an important complement to full structure elucidation. On the other hand, some KR possess sequence features of both A- and B-type KR, and so confident assignment is not possible [90,120]. In addition, predicting the configuration of the adjacent C-2-methyl groups (i.e., whether epimerization occurs or not) remains unreliable at present [90], but this situation may improve with the incorporation of additional sequences of epimerizing KR domains into sequence classification programs.

In terms of predicting double bond stereochemistry, using the direction of ketoreduction as a guide is not an infallible method (as explained earlier, although A-type reduction is often correlated to cis-double bond formation and B-type reduction, to trans, the exact opposite outcome has been observed in multiple systems). This situation would be improved by elucidation of the complete set of molecular mechanisms underlying cis-double bond formation. For the ER domains, as described previously, the observed methyl configuration correlates quite strongly with the residue at a specific sequence position (Y = S configuration; not Y = R) [115], and this has already proved useful for correctly predicting methyl stereochemistry (see for example, [122,123]).

**Conclusion**

As illustrated in this review, the tools of chemical biology coupled with molecular biological techniques have played a critical part in elucidating fundamental aspects of stereocontrol in modular polyketide biosynthesis. Given that our molecular understanding of these determinants remains incomplete – notably for the processing KR, DH and ER domains – this ap-
Figure 18: a) PKS engineered to test the role of the ER stereospecificity residues [115]. TKS-ERY4 was created by replacing the KR domain of DEBS module 2 (within the context of DEBS 1-TE) with the 'reductive loop' (DH-ER-KR) from DEBS module 4. This PKS gives rise to lactone 20 incorporating a (2S)-methyl group, consistent with the presence of the conserved Y in the active site. Site-directed mutation of this Y to the alternative V resulted in a complete shift in stereochemistry to give lactone 21, with (2R)-methyl stereochemistry. b) Model for reduction by ER domains from the 4′-pro-R hydride of NADPH. When the Tyr is present, it acts as the proton donor following reduction to give the (2S)-methyl. In its absence, the Lys residue on the opposite side of the active site acts as the general acid, yielding the (2R)-methyl.

References
Solid-phase enrichment and analysis of electrophilic natural products

Frank Wesche¹, Yue He¹ and Helge B. Bode*¹,²

Abstract
In search for new natural products, which may lead to the development of new drugs for all kind of applications, novel methods are needed. Here we describe the identification of electrophilic natural products in crude extracts via their reactivity against azide as a nucleophile followed by their subsequent enrichment using a cleavable azide-reactive resin (CARR). Using this approach, natural products carrying epoxides and α,β-unsaturated enones as well as several unknown compounds were identified in crude extracts from entomopathogenic Photorhabdus bacteria.

Introduction
Microorganisms are a major source for novel natural products and the subsequent development of new drugs for all kinds of applications [1,2]. For example, the discovery of the fungal natural product cyclosporine as an immunosuppressant drug facilitated modern organ transplantation [3,4].

The increasing sensitivity of analytical methods, especially in mass spectrometry, enables the detailed analysis of various natural product producing microbes. Much more compounds have been identified than originally thought, but often these are produced only at a very low level. This is also reflected by the genome sequences of bacteria and fungi that often encode numerous biosynthesis gene clusters (BGC) with most of the corresponding natural products unknown for several reasons: these BGCs are silent under standard laboratory conditions [5], the compounds are too labile for isolation or they are produced in amounts still below the detection limit of modern mass spectrometers.

Therefore it is desirable to have multiple and complementary methods available that allow the detection of several different natural product classes. Besides the traditional chemical
screening and bioactivity-guided isolation, the exploitation of the inherent properties of natural products is also feasible. Consequently, simple functional groups of natural products like dehydroalanine [6,7], ketones, aldehydes [8,9], carboxylic acids [8,9], amines [8-10], thiols [8,9], alcohols [11], epoxides [12], terminal alkynes [13,14] and azides [15] can be targeted to introduce a label. Such labels might increase the visibility in UV or MS detection in liquid chromatography coupled to UV or mass spectrometry. Alternatively, natural products can be immobilized on reactive resins by making use of their chemical functionality and can be eluted after washing off all non-desired substances [8,9,11,13,15].

The recently introduced cleavable azide-reactive resin (CARR (2), Scheme 1) is such a resin able to react with a broad range of azides [15]. Thus, the metabolic fate of azide-containing biosynthesis intermediates or building blocks can be studied and natural products containing these azides can be identified.

Herein we describe the application of the CARR enrichment for the detection of electrophilic natural products using azide as the nucleophile. Indeed, we could detect epoxystilbene 1 and three glidobactins from different Photobauds strains and some of these compounds are produced in such low amounts that they are not detectable in standard crude extracts.

Results and Discussion
In order to evaluate whether the CARR approach is suitable for epoxide detection, we azidated commercially available trans-stilbene oxide as representative model compound for 1 affording the vicinal azido alcohol, 2-azido-1,2-diphenylethanol (Supporting Information File 1, Figure S2) [16,17]. The azidation was carried out with sodium azide and ammonium chloride in 80% MeOH under reflux overnight (Supporting Information File 1, Figure S2). The azido alcohol was then incubated with CARR (2) in acetonitrile overnight followed by extensive washing of the resin with methanol and dichloromethane. After disulfide-bond cleavage with a solution of tris(2-carboxyethyl)phosphine (TCEP) in a 1:5:10 mixture of phosphate-buffered saline (PBS)/chloroform/methanol at pH 7, the filtrate was analyzed by HPLC–MS. As expected the corresponding mass m/z 493.2 [M + H]+ of the cleaved cycloaddition product could be directly detected in the base-peak chromatogram (BPC) showing the characteristic fragmentation pattern of the CARR adducts (Supporting Information File 1, Figure S3) [15]. Additionally, the detection limit of the model epoxide was investigated in a complex environment. For this, defined amounts of trans-stilbene oxide were added to liquid Lysogeny Broth (LB) medium. The obtained methanolic Amberlite XAD-16 extracts were azidated, enriched and analyzed. Up to a final concentration of 5 µg/L (≈25 nmol/L) the epoxide could be detected (Supporting Information File 1, Figure S4). Following these results, we tested the method with an XAD extract of P. luminescens TT01 (Figure 1).

The obtained XAD extract was treated the same way as trans-stilbene oxide, and any potentially containing epoxide should be converted into the corresponding vicinal azido alcohol. Afterwards the azidated extract was incubated with 2, the disulfide bond was cleaved and the filtrate subsequently analyzed by HPLC–MS. To our surprise, the BPC showed one distinct peak at 8.1 min with a characteristic MS² fragmentation pattern of a derivatized azide and a mass of m/z 567.2 [M + H]+. This mass corresponds to the calculated mass of 3 (Figure 1) derived from the derivatization of epoxystilbene 1, an oxidized isopropylstilbene derivative from this strain. Since only a single peak could
C-labeled medium prior to the azidation and cultivated in *P. luminescens* TT01 was for further structural confirmation, without formation of the other possible regioisomer. 1 epoxygate addition took place only on the less hindered position in be seen within the chromatogram, we assume that the conjugate addition took place only on the less hindered position in epoxide 1 without formation of the other possible regioisomer. For further structural confirmation, *P. luminescens* TT01 was cultivated in 13C-labeled medium prior to the azidation and enrichment procedure indeed confirming the incorporation of 17 carbon atoms in 3 (Supporting Information File 1, Figure S5). Furthermore, the molecular formula of the cleaved azide–alkyne cycloaddition product 3 was confirmed by HPLC–HRMS (calcd. mass: \(m/z\) 567.2636 [M + H]⁺, found: \(m/z\) 567.2635 [M + H]⁺, ∆ppm = 0.1). Compound 1 could hardly be detected in extracts from standard growth media but was detected from infected insects and media mimicking the insect hemolymph [18-20]. Clearly the biosynthesis of this compound is strictly regulated and moreover it is a highly labile compound that is probably rapidly degraded [21]. Only a very weak signal of \(m/z\) 271.1 [M + H]⁺ could be detected at 8.5 min, which can be associated with 1.

Encouraged by these results and especially the high sensitivity of the method, the azidation was performed with XAD extracts of three additional *Photorhabdus* strains (*Photorhabdus* PB45.5, *Photorhabdus* PB 68.1 and *Photorhabdus temperata* subsp. *thracensis* DSM 15199). Here, an even lower amount of 3 could be detected in *Photorhabdus* PB 68.1 (Supporting Information File 1, Figure S6), whereas in *Photorhabdus* PB45.5 and *Photorhabdus temperata* subsp. *thracensis* DSM 15199 nothing was visible at all, suggesting that the appropriate gene for the biosynthesis of 1 is either missing or silent in these species.

Upon a detailed look at the chromatograms, different masses with the characteristic fragmentation pattern of derivatized azides could be found (Table 1 and Supporting Information File 1, Figures S7–S9). The molecular formula obtained from HRMS data indicates that three glidobactin derivatives namely glidobactin A (4) [22], cepafungin I (5) [23,24] and luminmycin D (6) [25,26] were enriched (Scheme 2). Glidobactins are well-known proteasome inhibitors that react with a conserved threonine residue in the β5 subunit of the proteasome [24]. From *Photorhabdus* strains they have previously

**Table 1:** Additionally found masses in the tested strains, calculated molecular formulas of possible azide–alkyne cycloaddition products, and the molecular formulas of the putative parent compounds derived from subtraction of the azide and CARR-derived moiety (C13H9N4O2S).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Compd.</th>
<th>(t_R) min</th>
<th>[M + H]⁺ found</th>
<th>Calcd. molecular formula</th>
<th>[M + H]⁺ calcd.</th>
<th>∆ppm</th>
<th>Molecular formula natural product</th>
<th>Natural product</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. luminescens</em> TT01</td>
<td>3</td>
<td>8.1</td>
<td>567.2635</td>
<td>C30H39N4O6S</td>
<td>567.2636</td>
<td>0.1</td>
<td>C17H16O3</td>
<td>1</td>
</tr>
<tr>
<td><em>Photorhabdus</em> PB 68.1</td>
<td>–</td>
<td>8.6</td>
<td>491.2103</td>
<td>C27H31N4O2S</td>
<td>491.2071</td>
<td>2.2</td>
<td>C14H10O</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>8.0</td>
<td>493.2256</td>
<td>C27H33N4O2S</td>
<td>493.2268</td>
<td>2.0</td>
<td>C14H12O</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>8.5</td>
<td>517.2483</td>
<td>C28H37N4O2S</td>
<td>517.2479</td>
<td>2.0</td>
<td>C15H16O3</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8.1</td>
<td>567.2635</td>
<td>C30H39N4O6S</td>
<td>567.2636</td>
<td>0.1</td>
<td>C17H16O3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9.0</td>
<td>817.4630</td>
<td>C40H38N4O6S</td>
<td>817.4641</td>
<td>2.0</td>
<td>C27H44N4O20</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>9.3</td>
<td>831.4801</td>
<td>C41H40N4O8S</td>
<td>831.4797</td>
<td>1.1</td>
<td>C26H36N4O20</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 1: Additionally found masses in the tested strains, calculated molecular formulas of possible azide-alkyne cycloaddition products, and the molecular formulas of the putative parent compounds derived from subtraction of the azide and CARR-derived moiety (C<sub>13</sub>H<sub>19</sub>N<sub>4</sub>O<sub>2</sub>S). (continued)

<table>
<thead>
<tr>
<th>Strain</th>
<th>MW</th>
<th>Formula</th>
<th>MW</th>
<th>Formula</th>
<th>Δ</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photorhabdus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PB 45.5</td>
<td>10.7</td>
<td>551.3617</td>
<td>10.8</td>
<td>551.3626</td>
<td>1.5</td>
<td>unknown</td>
</tr>
<tr>
<td>9</td>
<td>10.1</td>
<td>819.5144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Photorhabdus temperata subsp.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>thracensis</strong></td>
<td>8.0</td>
<td>493.2258</td>
<td>9.5</td>
<td>489.2525</td>
<td>2.0</td>
<td>unknown</td>
</tr>
<tr>
<td>DSM 15199</td>
<td>10.7</td>
<td>551.3617</td>
<td>9.8</td>
<td>815.4859</td>
<td>1.5</td>
<td>unknown</td>
</tr>
</tbody>
</table>

**Scheme 2:** Structures of glidobactin derivatives (glidobactin A (4), cepafungin I (5) and luminmycin D (6)) before and after azidation and azide enrichment procedure (7, 8, 9). The MS<sup>2</sup> spectrum indicates that azidation of glidobactins only took place on the reactive site that is also targeted by the proteasome (Supporting Information File 1, Figure S11) [24].

Conclusion

The combination of the reactivity-guided introduction of an azide functionality into electrophilic natural products and the subsequent azide enrichment on a solid phase facilitates the detection of epoxides and α,β-unsaturated enones in XAD extracts of *Photorhabdus*. Epoxystilbene (1) and glidobactins have never been observed before in XAD extracts of *Photorhabdus* grown under standard conditions. Most likely this is due to very low production levels of these compounds thus illustrating the power of this method. We deem that many more electrophilic compounds were just overlooked in the past due to their low concentrations. In combination with labeling...
experiments even the nature of the parent natural product could be revealed. Moreover, a possible scale-up of this procedure should enable the preparative purification of yet unidentified compounds as well as the structural confirmation of the identified structures. This approach can also be applied to extracts of other bacteria, fungi and plants and can give at least hints on new electrophilic natural products, where their reactivity against azide might also reflect their biological activity.

Supporting Information
Supporting Information File 1
Materials, methods and supplementary figures. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-43-S1.pdf]

Acknowledgements
The authors are grateful to Dr. A. J. Perez for providing CARR and helpful discussions, as well as Lei Zhao for providing purified glidobactin A. This work was supported by a European Research Council starting grant under grant agreement number 311477.

References

License and Terms
This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.13.43
Investigation of the action of poly(ADP-ribose)-synthesising enzymes on NAD\(^+\) analogues

Sarah Wallrodt, Edward L. Simpson and Andreas Marx*

**Abstract**

ADP-ribosyl transferases with diphtheria toxin homology (ARTDs) catalyse the covalent addition of ADP-ribose onto different acceptors forming mono- or poly(ADP-ribose)ylated proteins. Out of the 18 members identified, only four are known to synthesise the complex poly(ADP-ribose) biopolymer. The investigation of this posttranslational modification is important due to its involvement in cancer and other diseases. Lately, metabolic labelling approaches comprising different reporter-modified NAD\(^+\) building blocks have stimulated and enriched proteomic studies and imaging applications of ADP-ribosylation processes. Herein, we compare the substrate scope and applicability of different NAD\(^+\) analogues for the investigation of the polymer-synthesising enzymes ARTD1, ARTD2, ARTD5 and ARTD6. By varying the site and size of the NAD\(^+\) modification, suitable probes were identified for each enzyme. This report provides guidelines for choosing analogues for studying poly(ADP-ribose)-synthesising enzymes.

**Introduction**

ADP-ribosyl transferases with diphtheria toxin homology [1] (ARTDs), also termed poly(ADP-ribose) polymerases (PARPs), form an enzyme family of 18 human members [2] that mediate their widespread functions in cellular homeostasis through the catalysis of ADP-ribosylation [3,4]. This posttranslational modification received considerable attention within the last decade [5,6] and has been linked to tumour biology, oxidative stress, inflammatory, and metabolic diseases [7]. Using NAD\(^+\) as a substrate, ARTDs covalently transfer ADP-ribooses onto themselves or different targets forming mono(ADP-ribose)ylated proteins. Some ARTDs are in particular able to elongate these initial units with additional NAD\(^+\) molecules to build a complex, highly charged biopolymer called poly(ADP-ribose) (PAR, Figure 1). These polymers consist of up to 200 units of ADP-ribose and may branch every 20 to 50 monomers [8-10]. To date, only four ARTD members were found to...
accomplish the synthesis of PAR, namely the DNA-dependent ARTD1 and ARTD2 as well as the tankyrases ARTD5 and ARTD6 [2,3].

ARTD1 as the founding member is the best investigated enzyme of ARTDs and is considered the main source of cellular PAR [11]. ARTD1 and its closest relative ARTD2 comprise DNA-binding domains and their activity is stimulated by binding to different types of DNA breaks [12]. They fulfill functions in DNA repair, genome maintenance, transcription, and metabolic regulation [11,13]. The tankyrases ARTD5 and ARTD6 also exhibit a unique domain structure consisting of multiple ankyrin repeats mediating protein–protein interactions [13]. Tankyrases are involved in telomere homeostasis, Wnt/β-catenin signalling, glucose metabolism, and cell cycle progression [14].

Remarkable efforts have been undertaken to develop tools and assays for studying PARylation on a molecular level and to understand the complex processes and interactions of the involved ARTDs. Recently, the employment of NAD⁺ analogues resulted in the development of powerful applications for

Figure 1: NAD⁺ is used as a substrate by ARTDs to form MARylated and PARylated proteins. Depicted are alkyne- and dye-modified NAD⁺ analogues 1–6 that are applied in this study.
Figure 2: Workflow of the ADP-ribosylation assay. The protein of interest (POI) is ADP-ribosylated by the respective ARTD and by NAD⁺, NAD⁺ analogue or a 1:1 mixture. Then, copper(I)-catalysed azide–alkyne click reaction (CuAAC) is performed and mixture is resolved by SDS PAGE.
Table 1: Acceptance of alkyne-modified NAD\(^+\) analogues 1–4 by different ARTDs without or with competition of natural substrate.\(^a\) \(\checkmark\) = analogue is well processed, (\(\checkmark\)) = analogue is processed with lower efficiency, \(\times\) = analogue is not processed.

<table>
<thead>
<tr>
<th>NAD(^+) analogue</th>
<th>Nat. NAD(^+)</th>
<th>ARTD1</th>
<th>ARTD2</th>
<th>ARTD5</th>
<th>ARTD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>1:1</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>2</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>1:1</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>3</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
<tr>
<td>1:1</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
<tr>
<td>4</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
<tr>
<td>1:1</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
<td>(\times)</td>
</tr>
</tbody>
</table>

\(\text{\(a\)All gels are depicted in Supporting Information File 1, Figure S1 and Figure S2.}\)

Figure 3: SDS PAGE analysis of ADP-ribosylation of histone H1.2 with ARTD1, ARTD2, ARTD5 and ARTD6 using NAD\(^+\) analogue 1. Upper panel shows Coomassie Blue staining; lower panel shows TMR fluorescence. Experimental details are provided in Supporting Information File 1. *Unspecific staining of H1.2 in lanes 3 results from non-catalytic bond formation of NAD\(^+\) analogues with the protein.

Similarly in histone ADP-ribosylation (Supporting Information File 1, Figure S1) and in auto(ADP-ribos)ylation (Figure S2). As known from previous work [15,17], ARTD1 was not able to process 7- and 8-modified NADs 3 and 4 and so does ARTD2 (Supporting Information File 1, Figure S1, lanes 9 to 14 and Figure S2, lanes 7 to 10). In both assays, only small amounts of modified PAR was formed with the 6-modified derivative 2 and in the absence of natural NAD\(^+\) (Supporting Information File 1, Figure S1, lane 7 and Figure S2, lane 5), when compared in parallel with 2-modified analogue 1. However, a strong signal is detected in a mixture containing NAD\(^+\) (Figure S1, lane 8 and Figure S2, lane 6). Application of compound 1 results in the strongest signal and is competitive towards natural substrate (Figure S1, lanes 4 to 5 and Figure S2, lanes 3 to 4).

Also in ARTD5- and ARTD6-catalysed ADP-ribosylation (panel c and d), analogues 3 and 4 were not used as substrates (Supporting Information File 1, Figure S1, lanes 9 to 10 and Figure S2, lanes 7 to 10). In contrast, compounds 1 and 2 were both used by both enzymes for PAR formation, even in the absence of natural NAD\(^+\). In case of ARTD5, derivative 1 seems to be slightly better processed than 2 in histone ADP-
ribozymation, whereas in case of ARTD6 both are used as substrates in both assays with similar efficiencies.

Dye-modified NAD\textsuperscript{+} analogues

Because the alkyne-tag induces only small alterations to the NAD\textsuperscript{+} scaffold, we also investigated how these enzymes would act on bulkier substitutions. For this purpose, we selected bulky, dye-modified NAD\textsuperscript{+} analogues 5 and 6, which were previously prepared by our group [21], in order to have a direct, fluorescent read-out. The outcome is summarised in Table 2 and the SDS PAGE gels obtained are depicted in Figure 4 and Supporting Information File 1, Figures S3 and S4.

As shown in Figure 4 and Supporting Information File 1, Figure S4b, ARTD2 processes analogue 5 in a competitive manner and fluorescent and Coomassie Blue stained polymer chains are formed in the absence and the presence of natural substrate (Figure 4, lanes 4 to 5 and Figure S4b, lanes 3 to 4). Unlike ARTD1 (Supporting Information File 1, Figure S3a), little fluorescent signal is obtained with compound 6 in ARTD2 catalysed histone PARylation in the absence of natural NAD\textsuperscript{+} (Figure 4, lane 7) and in ARTD2 automodification (Figure S4b, lane 5).

ARTD5 showed decreased incorporation of the larger substituted analogues 5 and 6. During automodification, both compounds failed to form detectable, fluorescent PAR chains (Figure 4 and Supporting Information File 1, Figure S4c, lanes 3 to 6). In general, it can be concluded that ARTD5 showed less activity in automodification compared to the other ARTDs [26].

Table 2: Acceptance of dye-modified NAD\textsuperscript{+} analogues 5 and 6 by different ARTDs without or with competition of natural substrate.\textsuperscript{a} \textsuperscript{b} \textsuperscript{c} \textsuperscript{=} analogue is well processed, (\textsuperscript{=}) = analogue is processed with lower efficiency, \textsuperscript{x} = analogue is not processed.

<table>
<thead>
<tr>
<th>NAD\textsuperscript{+} analogue</th>
<th>Nat. NAD\textsuperscript{+}</th>
<th>ARTD1</th>
<th>ARTD2</th>
<th>ARTD5</th>
<th>ARTD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>1:1</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>1:1</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

\textsuperscript{a}All gels are depicted in Supporting Information File 1, Figure S3 and Figure S4. \textsuperscript{b}6 is accepted in H1.2 ADP-ribosylation with little efficiency, but not in automodification. \textsuperscript{c}Analogues are not accepted in automodification.

Figure 4: SDS PAGE analysis of ADP-ribosylation of histone H1.2 with ARTD2, ARTD5 and ARTD 6 using NAD\textsuperscript{+} analogues 5 and 6. Upper panel shows Coomassie Blue staining; lower panel shows TMR fluorescence. Experimental details are provided in Supporting Information File 1. *High unspecific staining of H1.2 in lanes 3 and 6 results from non-catalytic bond formation of NAD\textsuperscript{+} analogues with the protein.
Nevertheless, analogue 6 was somewhat processed using the histone-based assay as seen by fluorescent and Coomassie-blue-stained polymers in the absence of natural substrate and increased polymer in the presence of natural NAD\(^+\) (Figure 4, lanes 7 to 8). The fluorescence observed in the presence of 5 is similar to the background signal indicating poor processing of 5 (Figure 4, lanes 4 to 5).

In case of ARTD6, both analogues were used for the ADP-ribose-sylation of histone (Figure 4, lanes 3 to 8) and in automodification (Supporting Information File 1, Figure S4d, lanes 3 to 6) with similar efficiency.

Conclusion

In this paper, we investigated the scope of PAR synthesising enzymes, namely ARTD1, ARTD2, ARTD5 or ARTD6 for using modified NAD\(^+\) analogues. It was found that NAD\(^+\) analogues 1 and 2 modified with alkyne groups in adenine position 2 and 6 are used by all these enzymes to a certain extent, whereas the employed substitutions in adenine at position 7 and 8 completely abrogated the processing towards PAR. The DNA-dependent ARTDs ARTD1 and ARTD2 can process 2-modified analogues best as also sterically demanding compounds such as dye-modified 4-modifed analogues. Because ARTD5 and ARTD6 exhibit different constraints for metabolising bulky 2-modified analogue 5, this behaviour could be used to discriminate their activity in a cellular context. By choosing the best NAD\(^+\) substrate for each enzyme more reliable and valuable insights into PARylation can be achieved and will help to decipher these processes in more detail.

Supporting Information

Supporting Information File 1

Additional figures, synthesis of compounds and biochemical methods.
[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-49-S1.pdf]

Acknowledgements

Financial support by Konstanz Research School Chemical Biology is gratefully acknowledged. S. W. acknowledges the ‘Beilstein-Institut zur Förderung der Chemischen Wissenschaften’ and E. L. S. the RISE programme of the German Academic Exchange Service for stipends.

References

2. Hottiger, M. O. Mol Cell 2015, 58, 1134. doi:10.1016/j.molcel.2015.06.001

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.13.49
Isoxazole derivatives as new nitric oxide elicitors in plants

Anca Oancea¹, Emilian Georgescu², Florentina Georgescu³, Alina Nicolescu⁴,⁵, Elena Iulia Oprita¹, Catalina Tudora¹, Lucian Vladulescu³, Marius-Constantin Vladulescu³, Florin Oancea⁶ and Calin Deleanu*⁴,⁵

Abstract
Several 3,5-disubstituted isoxazoles were obtained in good yields by regiospecific 1,3-dipolar cycloaddition reactions between aromatic nitrile oxides, generated in situ from the corresponding hydroxyimidoyl chlorides, with non-symmetrical activated alkynes in the presence of catalytic amounts of copper(I) iodide. Effects of 3,5-disubstituted isoxazoles on nitric oxide and reactive oxygen species generation in Arabidopsis tissues was studied using specific diaminofluoresceine dyes as fluorescence indicators.

Introduction
Isoxazoles are an interesting class of N-heterocyclic compounds intensely studied mainly due to their wide range of biological activity [1,2]. Isoxazole compounds show antiviral [3,4], antithrombotic [5-9], analgesic [9], COX-2 inhibitory [10,11], anti-inflammatory [9,11], antinociceptive [12] and anticancer [13] activities. Several isoxazole derivatives have GABAₐ antagonist [14] and T-type Ca²⁺ channel blocking activities [15]. Commercial drugs featuring an isoxazole moiety include the COX-2 inhibitor Valdecoxib and the β-lactam antibiotics Cloxacillin and Dicloxacillin. An isoxazole derivative, namely
3,5-difluorophenyl-[3-methyl-4-(methylsulfonyl)isoxazol-5-yl]methanone, was recently reported as an inducer of nitric oxide producing elicitor in plants [16, 17]. Nitric oxide (NO), which has been demonstrated to be a major gasotransmitter in mammals, is also involved in the orchestration of various plant physiological responses, playing an important role in the regulation of interactions between plant and microorganisms and in plant defense mechanisms against stresses [18, 19]. Consequently, there is interest in the biological evaluation of further isoxazole derivatives.

Many synthetic approaches towards the isoxazole core include the reactions of hydroxylamine with aryl-[β]-diketones [20], α,β-unsaturated carbonyl compounds [21], or α,β-unsaturated nitriles [22], and 1,3-dipolar cycloaddition reactions between alkenes or alkynes and nitrile oxides [23-25].

Nitrile oxides are known as reactive 1,3-dipoles involved in 1,3-dipolar cycloaddition reactions with various dipolarophiles generating five-membered heterocyclic compounds, such as isoxazoles, isoxazolines, oxadiazoles, oxadiazolines, dioxazolidines etc. [23-25]. Intermediate nitrile oxides are usually generated in situ by the oxidative dehydrogenation of aldoximes in the presence of various oxidants [26-29], or by the dehydrohalogenation of hydroxymoyl halides promoted by organic or inorganic bases [30-32]. A less used synthetic procedure involves the oxidative dehydration of primary nitro compounds with isocyanates in the presence of tertiary alkylamines [33].

Generally, the cycloaddition reactions of nitrile oxides to alkenes yield isoxazolines or a mixture of isoxalines and isoxazoles. Cycloaddition reactions of nitrile oxides to alkynes yield isoxazoles directly, without a catalyst, but the yields of isoxazole products are quite low because of side reactions and both regioisomers are generally obtained [23-25]. The one-pot 1,3-dipolar cycloaddition reaction of a nitrile oxide, generated in situ from the corresponding hydroxymoyl chloride, with an in situ brominated electron-deficient alkene led to the intermediate bromoisoxazoline from which, by loss of HBr, a 3,5-disubstituted isoxazole derivative is formed as major regioisomer [34]. Based on the copper(I)-catalyzed click reactions of organic azides with terminal acetylenes [35], different copper(I)-catalyzed synthetic procedures towards isoxazole derivatives were developed [36,37].

As part of our continued efforts to develop simple synthetic routes towards bioactive heterocyclic compounds [38-43], we report here the synthesis of several 3,5-isoxazole derivatives, bearing benzo[1,3]dioxole and thiophene scaffolds respectively, as well as their inductor effect on the generation of nitric oxide and reactive oxygen species in plant tissues. The benzo[1,3]dioxole framework is a constituent of some fragrances and flavors [44], and several bioactive compounds with a broad spectrum of applications [45-48]. The thiophene is a core system of a large number of bioactive molecules such as antineoplastic agents [49], non-steroidal anti-inflammatory drugs [50] or compounds with antibacterial activities against several Gram-positive strains [51]. Thus, 3,5-disubstituted isoxazole derivatives were obtained by regioselective 1,3-dipolar cycloaddition reactions of aromatic nitrile oxides to non-symmetrical activated alkyne derivatives in the presence of catalytic amounts of copper(I) iodide.

Results and Discussion

Synthesis of 3,5-disubstituted isoxazole derivatives

The regioselective copper(I)-catalyzed 1,3-dipolar cycloaddition reactions of aromatic nitrile oxides, generated in situ from the corresponding crude imidoyl chlorides, and non-symmetrically activated alkynes led to 3,5-disubstituted isoxazole derivatives. For this, aromatic aldehydes 1 are first converted to the corresponding aldoximes 2, via reactions with hydroxylamine, and the crude reaction products are transformed to the corresponding imidoyl chlorides 3 which are directly used in the next step without purifications. Catalytic amounts of copper(I) iodide and a base, such as KHCO₃, are added to an aqueous solution containing the crude imidoyl chlorides 3 and the non-symmetrically activated alkynes 4. The in situ generated aromatic nitrile oxides 5 undergo an addition to copper(I) acetyldienes formed in situ to give the 3,5-disubstituted isoxazoles 6–11 as single isomers, in moderate to good yields (Scheme 1).

All reactions between crude imidoyl chlorides 3 and the non-symmetrical activated alkynes 4 are carried out in aqueous solutions at room temperature. The final 3,5-disubstituted isoxazoles are easily separated by filtration, washed with water and recrystallised. The synthesized 3,5-disubstituted isoxazoles are presented in Table 1.

3,5-Disubstituted isoxazoles 6, 7, 9 and 10 (Table 1) have been previously prepared by different synthetic procedures, but compounds 6 [52], 7 [49] and 10 [53] have been only partly characterized, while for compound 9 [54] no characterization data have been reported.

3,5-Disubstituted isoxazole structures were unambiguously assigned on the basis of chemical and spectral analysis (IR, ¹H, ¹³C and ¹⁵N NMR spectra). NMR spectra clearly indicated the presence of only one regioisomer for all synthesized 3,5-disubstituted isoxazoles. Signals in ¹H and ¹³C NMR spectra were fully assigned based on H,C-HSQC and H,C-HMBC experiments. The 3,5-disubstitution was also experimentally proven.
Scheme 1: Synthetic route to 3,5-disubstituted isoxazoles.

Table 1: Reported 3,5-disubstituted isoxazoles.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Ar</th>
<th>R</th>
<th>mp (º)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Me</td>
<td>159–160; 141–142</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>OEt</td>
<td>100–101.5</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Ph</td>
<td>135–137</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Me</td>
<td>113–115</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>OEt</td>
<td>78–80</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Ph</td>
<td>90–91</td>
<td>74</td>
<td></td>
</tr>
</tbody>
</table>

by NOE experiments which indicated through space interaction between the H-4 proton and the protons from both 3- and 5-substituents.

**Biological activity**

We investigated the inductor effects of 3,5-disubstituted isoxazoles 6–11 on NO and reactive oxygen species (ROS) production in plant tissues. Usually, NO and ROS, such as O₂⁻, OH⁻ and H₂O₂, together are required to induce the activation of various defense-related enzymes in plants [55]. Plant cells contain oxygen radical detoxifying enzymes and nonenzymatic antioxidants which have an essential role in protection of plant cells from oxidative damages at the sites of ROS and NO generation [56,57]. Measuring the ROS and NO levels in plant tissues is often difficult due to high reactivity and extremely short physiological half-life of these free radicals [58,59].

In this work, generation of both NO and ROS was proven by fluorescence microscopy using specific fluorescence indicators that help to exactly define the sites of generation. We used Arabidopsis thaliana wild type seeds, cultivated for six weeks in laboratory in Arasystem [60]. Arabidopsis thaliana was selected as model organisms for NO inductors because this is the flowering plant with the largest amount of knowledge on cellular and molecular biology and it has a relatively short life cycle. The Arabidopsis leaves were sprayed with fine suspensions of isoxazole inductors 6–11 at two different concentrations (10 μg/mL and 50 μg/mL, respectively) and collected after 24 h. Collected leaves were washed with distilled water and incubated with the specific fluorescence indicator for histochernical analysis of ROS and NO by fluorescence microscopy. Arabidopsis leaves untreated with inductor suspensions have been used as negative controls. As positive control, we used chitosan (β-1,4 linked glucosamine) with average molecular weight, a fungal elicitor with known effect as NO and ROS inductor on A. thaliana [61].

Intracellular ROS was visualized using 2',7'-dichlorodihydrofluorescein diacetate (H₂DCF DA) as fluorescence indicator. The method is based on the oxidation of the non-fluorescent probe of 2',7'-dichlorodihydrofluorescein diacetate to the highly fluorescent 2',7'-dichlorofluorescein diacetate [62,63].

Intracellular NO was visualized using 4-amino-5-methylamino-2',7'-dichlorodihydrofluorescein diacetate (DAF-FM diacetate), a non-fluorescent compound, which reacts with NO to form a
fluorescent benzotriazole and does not react with any ROS [64-67].

Fluorescence microscopy images of all 3,5-disubstituted isoxazoles-treated Arabidopsis leaves showed a pronounced presence of ROS at both concentrations of inductors (10 μg/mL and 50 μg/mL). Strong fluorescence densities were observed especially at higher concentration (50 μg/mL) of 3,5-disubstituted isoxazoles. Intensities of fluorescence revealed differences between tested compounds. These compounds can be listed in the increasing order of ROS generation efficacy as follows: 11 > 10 > 9 > 7 > 8 > 6 (Figure S2 in Supporting Information File 1).

Similarly, images of fluorescence microscopy revealed the presence of NO in all 3,5-disubstituted isoxazoles-treated Arabidopsis leaves, at both concentrations (10 μg/mL and 50 μg/mL), especially at higher concentration of compounds (50 μg/mL). The NO releasing capacity of newly synthesized (50 μg/mL), especially at higher concentration of compounds.

Fluorescence data indicate that the 3,5-disubstituted isoxazoles 6–11, particularly 9–11, are involved in NO and ROS production in Arabidopsis treated leaves.

Conclusion
Several 3,5-disubstituted isoxazoles were obtained by the convenient, regioselective 1,3-dipolar cycloaddition reactions of aromatic nitrile oxides, generated in situ from the crude imidoyl chlorides, with non-symmetrical activated alkynes in the presence of catalytic amounts of copper(I) iodide. The effect of 3,5-disubstituted isoxazoles in generation of ROS and NO in plant tissues was investigated by fluorescence microscopy. The obtained data indicate that some of these compounds are chemical elicitors that induce NO and ROS generation in plant tissues and could activate various defense mechanisms in plants. Further research is in progress to assess the in planta mechanism of NO generation by these compounds.

Acknowledgments
This work was supported by the Ministry of National Education – Research Activity, UEFISCUD, Project PN-II-PT-PCCA-2013-4-0267 – SAFE-SEL, contract 186/2014. Access to research infrastructure developed in the “Petru Poni” Institute of Macromolecular Chemistry through the European Social Fund for Regional Development, Competitiveness Operational Programme Axis 1, Project InoMatPol (ID P.36_570, Contract 142/10.10.2016, cod MySMIS: 107464) is gratefully acknowledged.

References

Supporting Information
Supporting Information File 1
Experimental procedures, characterization data, IR, 1H, 13C and 15N NMR data for all new compounds and fluorescence microscopy images of NO and ROS generation for all 3,5-disubstituted isoxazoles-treated Arabidopsis leaves.
[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-65-S1.pdf]


63. Chen, Y.; Mo, H.-Z.; Hu, L.-B.; Li, Y.-Q.; Chen, J.; Yang, L.-F. 

  doi:10.1021/ac9801723

  373–375. doi:10.1248/cpb.46.373

  doi:10.1002/(SICI)1521-3773(19991102)38:21<3209::AID-ANIE3209>3.0.CO;2-6


**License and Terms**

This is an Open Access article under the terms of the Creative Commons Attribution License 
(http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: 
(http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: 
doi:10.3762/bjoc.13.65
An eco-compatible strategy for the diversity-oriented synthesis of macrocycles exploiting carbohydrate-derived building blocks

Sushil K. Maurya*1,2,§ and Rohit Rana1,2

Abstract

An efficient, eco-compatible diversity-oriented synthesis (DOS) approach for the generation of library of sugar embedded macrocyclic compounds with various ring size containing 1,2,3-triazole has been developed. This concise strategy involves the iterative use of readily available sugar-derived alkyne/azide–alkene building blocks coupled through copper catalyzed azide–alkyne cycladdition (CuAAC) reaction followed by pairing of the linear cyclo-adduct using greener reaction conditions. The eco-compatibility, mild reaction conditions, greener solvents, easy purification and avoidance of hazards and toxic solvents are advantages of this protocol to access this important structural class. The diversity of the macrocycles synthesized (in total we have synthesized 13 macrocycles) using a set of standard reaction protocols demonstrate the potential of the new eco-compatible approach for the macrocyclic library generation.

Introduction

Macrocycles offer very complex molecular architectures with a diverse range of ring sizes decorated with many functional groups found application in pharmaceuticals, agrochemicals, cosmetics and materials science [1-4]. Carbohydrate-embedded macrocycles represent an important class of macrocyclic compounds in which at least two bonds from a monosaccharide residue form a part of the macrocyclic rings and have shown important biological properties [5-12]. For example, macrocyclic aminoglycoside analogues have shown binding with the trans-activating region (TAR) RNA of the human immunodefi-
access to multiple building blocks by incorporating diversity-tional groups in the carbohydrate precursor allows for easy tionalities which enable them to establish catalytic sites as well as poly-functionalized derivatives via CuAAC reaction. The approach enables the rapid synthesis of carbohydrate conjugates in which the heterocyclic triazolyl ring serves as a shackle for joining the carbohydrate building blocks. Further, these carbohydrate conjugates decorated with appropriate coupling partner can be paired through ring closing metathesis (RCM) reaction. Carrying out the metathesis processes in green solvents is a major challenge. Unfortunately, halogenated solvents such as dichloromethane (DCM), 1,2-dichloroethane (DCE) or aromatics such as benzene and toluene are the most frequently used solvents for metathesis reactions whereas these solvents possess serious health and environmental hazards.

Here we report a novel application of the popular build-couple-pair (B/C/P) strategy [4,18,54,55,59,60] for the synthesis of sugar embedded macrocycles by iterative use of carbohydrate derived building blocks armed with azide/alkyne–alkyne functionalities. The building blocks were coupled via 1,3-dipolar cycloaddition (click reaction) iteratively through the development of a greener base-free Cu(I)-catalyzed azide–alkyne cycloaddition reaction. The cycloadducts were then converted to macrocycles by Ru-catalyzed cyclization reaction using greener and non-hazards reaction conditions.

### Results and Discussion

There are several DOS strategies to generate a collection of diverse molecules among them three-phase build-couple-pair (B/C/P) is one of the most frequently used. The B/C/P strategy involves build phase in which different building blocks were synthesized incorporating different diversity elements. These different building blocks were then combined together in the couple phase to give the substrates for the next phase. Finally, in the pair phase various functional-group-compatible reactions were used to generate distinct molecular scaffolds. The build-couple-pair strategy using iterative couple steps (B/C/P or B/C/C/P etc.) to increase the diversity of scaffolds accessed from the sets of building blocks has been exploited in recent times [59-63]. Also, simple and economical polyfunctional substrates available in abundance from the natural resources are ideal starting materials in DOS, which aims at providing quick access to libraries of diverse molecules. To exploit the strategy, it was envisioned that different sugars could serve as precursor for the necessary azide–alkyne and azide–alkene functionalities and could be connected through a sequence of protection-deprotection-functionalization reactions at appropriate position (Figure 1). D-glucose, D-xylose and L-arabinose were used as the key starting materials for the DOS protocol. It was expected
that each given sugar building block (generated in the building phase of the DOS) could be attached through Cu-catalyzed azide–alkyne cycloaddition (CuAAC) reaction (couple phase). Noteworthy, herein we utilized the CuAAC reaction as a medium to couple different building blocks assembled iteratively to generate a 1,2,3-triazole moiety. This 1,2,3-triazole moiety linked as a spacer due to its inherent properties including stability towards acid–base hydrolysis, active participation in H-bonding, dipole–dipole and π-stacking interactions [37,64-66]. The reaction would then afford a range of acyclic precursors, which could then undergo the intramolecular cyclization reaction to furnish the macrocyclic compounds (pair phase). In the pair phase, CuAAC adducts were cyclized using a Ru-catalyzed metathesis reaction utilizing Grubbs second generation catalysts under greener reaction conditions (Figure 1).

Figure 1: Build-couple-pair (B/C/P) strategy for macrocycles.

Build phase: preparation of building blocks

The alkyne-alkene (1a–f) and azide-alkene (2a–d) building blocks were synthesized in multigram scale following known literature procedures (Figure 2). The experimental details of the various building blocks used for DOS can be found in Supporting Information File 1.

Couple phase: Copper-catalyzed azide–alkyne cycloaddition (CuAAC)

After having ready requisite building blocks our next goal was to assemble them iteratively to synthesize macrocyclic library (Scheme 1). All the reactions were monitored after an interval of 2 and 4 hours and if required than after 24 hours for the optimizations; the conversion in the reaction was calculated by comparing the ratio of integration of the terminal alkyne proton

Figure 2: Different building blocks used for DOS.
in the propargyl building block and the characteristic triazole–alkene proton in the cyclo-adducts. The click reaction proceeds under various conditions with a plenty of sources of Cu(I) \[19\]. We have selected copper iodide (CuI) as Cu(I) source for the CuAAC. Initially we tried the reaction using CuI as catalyst and DIPEA as a base for the cycloaddition of the alkyne (1a) azide and (2a) building blocks in acetonitrile at room temperature. Pleasingly, the reaction resulted in excellent conversion (by \(^1\)H NMR) in two hours with 70\% isolated yield whereas addition of triethylamine in acetonitrile resulted in 65\% yield. For developing greener conditions for the cycloaddition reaction, a control experiment with alkyne (1a) and azide (2a) in water at room temperature reacted up to 24 hours but in the absence of copper catalyst and base, only 6\% conversion was observed (measured by \(^1\)H NMR; formation of two products were observed in the ratio of 77:23). Reaction in water at 70 °C under the above conditions gave a 33\% conversion with a 63\% selectivity for the product. Complete disappearance of starting materials after 24 hours with the formation of exclusively one product in 45\% yield was observed when the reaction was performed at room temperature in water using 5 mol \% CuI. Another reaction under similar conditions using CuI and DIPEA resulted in a lower yield of 35\% after 24 hours. The next reaction was performed in water at 70 °C using 5 mol \% catalysts in absence of a base. Interestingly, we observed complete disappearance of starting substrate in two hours with an excellent isolated yield of 95\% for the exclusive product whereas addition of DIPEA under similar conditions resulted in the low yield of 48\%. It is worth mentioning that the formation of the other regiosomer was not observed when the reactions were performed at 90 °C and 110 °C in water using 5 mol \% CuI as catalyst. These results confirm the essential role of copper required for the high conversion and selectivity of the products (Table 1).

After screening various reaction conditions we found a “greener” protocol for the CuAAC reaction in water under mild heating and the use of base was eliminated. We have utilized this methodology for the synthesis of a range of cycloadducts (3a–m, Table 2) via iterative coupling of carbohydrate derived azide and propargyl building blocks to be used as metathesis substrates for the synthesis of novel sugar embedded macrocyclic molecules. Cycloaddition of xylose derived azide building blocks containing a primary azido group (2a) produced similar yields (i.e., 3a, 3b and 3d) with xylose and arabinose derived building blocks containing a propargyl ether group on the primary OH group (1a and 1c) or xylose derived building block containing a propargyl group on the secondary OH group (1b). Further, a comparatively lower yield for the cycloaddition reaction (3e) was obtained when both building blocks used contain a secondary azide group (2b) and a propargyl ether on the secondary OH group (1c). We have observed relatively low yields (3e and 3f) when we used a combination of glucose (1e and 2c) and xylose (1b and 2a) derived building block whereas

---

**Scheme 1:** Cycloaddition reaction of alkyne-azide building block.

---

**Table 1:** Optimization of the reaction conditions for the cycloaddition.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Solvent</th>
<th>Base</th>
<th>Catalyst (CuI, mol %)</th>
<th>Temperature</th>
<th>Time (hours)</th>
<th>Yield %a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACN</td>
<td>TEA</td>
<td>5</td>
<td>ambient</td>
<td>2</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>ACN</td>
<td>DIPEA</td>
<td>5</td>
<td>ambient</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>H(_2)O</td>
<td>–</td>
<td>–</td>
<td>ambient</td>
<td>24</td>
<td>6(^b) (77% selectivity for 3a)(^c)</td>
</tr>
<tr>
<td>4</td>
<td>H(_2)O</td>
<td>–</td>
<td>–</td>
<td>70 °C</td>
<td>24</td>
<td>33(^b) (63% selectivity for 3a)(^c)</td>
</tr>
<tr>
<td>5</td>
<td>H(_2)O</td>
<td>–</td>
<td>5</td>
<td>ambient</td>
<td>24</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>H(_2)O</td>
<td>DIPEA</td>
<td>5</td>
<td>ambient</td>
<td>24</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>H(_2)O</td>
<td>–</td>
<td>5</td>
<td>70 °C</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>H(_2)O</td>
<td>DIPEA</td>
<td>5</td>
<td>70 °C</td>
<td>2</td>
<td>48</td>
</tr>
</tbody>
</table>

\(^a\)Isolated yield after column chromatography; \(^b\)conversion and \(^c\)product selectivity was measured by \(^1\)H NMR.
<table>
<thead>
<tr>
<th>Alkyne</th>
<th>Azide</th>
<th>Cycloadduct</th>
<th>Yield %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>2a</td>
<td><img src="3a.png" alt="Image" /></td>
<td>95</td>
</tr>
<tr>
<td>1b</td>
<td>2a</td>
<td><img src="3b.png" alt="Image" /></td>
<td>94</td>
</tr>
<tr>
<td>1b</td>
<td>2b</td>
<td><img src="3c.png" alt="Image" /></td>
<td>75</td>
</tr>
<tr>
<td>1c</td>
<td>2a</td>
<td><img src="3d.png" alt="Image" /></td>
<td>90</td>
</tr>
<tr>
<td>1e</td>
<td>2a</td>
<td><img src="3e.png" alt="Image" /></td>
<td>76</td>
</tr>
</tbody>
</table>
Table 2: Copper catalyzed azide-alkyne cycloaddition. (continued)

<table>
<thead>
<tr>
<th>1b</th>
<th>2c</th>
<th>3f</th>
<th>78</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e</td>
<td>2c</td>
<td>3g</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1d</td>
<td>2c</td>
<td>3h</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>2d</td>
<td>3i</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>2d</td>
<td>3j</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Copper catalyzed azide-alkyne cycloaddition. (continued)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Structure" /></td>
<td>95%</td>
</tr>
<tr>
<td><img src="image" alt="Structure" /></td>
<td>77%</td>
</tr>
<tr>
<td><img src="image" alt="Structure" /></td>
<td>67%</td>
</tr>
</tbody>
</table>

aMethod: CuI (5 mol %), water, 70 °C, 2 h.

an excellent yield was obtained (3g and 3h) when both coupling partners were derived from glucose (1e, 1d and 2c) irrespective of the position of the propargyl group on the primary OH (1e) or secondary OH group (1d). Next we thought of exploring the effect of protecting groups on the feasibility of the reaction and the yields and various building blocks with free OH groups were selected. It is worth mentioning that we did not observe any significant change in the reaction rate. Yields were relatively high (3i, 3j, 3k and 3l) when we used combination of azide and propargyl building blocks containing at least one free OH group and generally yields were not influenced by the position of the azide or propargyl group onto building blocks. However, when both building blocks used for the cycloaddition containing a free OH group (1f and 2d), the yield for the product was significantly low (3m). In conclusion, the CuAAC reaction of xylose derived building blocks gave relatively higher yields (3a and 3b), except when both building blocks contain a secondary azide and a secondary propargyl ether group (3c). Cycloaddition of building blocks derived from glucose and xylose worked better with non-protected OH groups (3i, 3j and 3k) than with protected (3e and 3f). Whereas glucose–glucose did work better with protected OH groups (3g and 3h) than with non-protected (3l and 3m).

Pair phase: macrocyclization via Ru-catalyzed ring closing metathesis (RCM)

In the pair phase the range of linear substrates derived by CuAAC were cyclized via Ru-catalyzed ring closing metathesis reaction (Table 3). In general, RCM conditions used in this study proved to be very robust and delivered the macrocyclic product in moderate yields. To begin our RCM endeavor, we performed the macrocyclization reaction on cycloadduct 3a in dichloromethane (10 mM) heating at 50 °C with 2 mol % second generation Grubbs catalyst. The reaction was incomplete after two hours and required an additional catalyst loading of 2 mol % and 1 mol % after every two hours. However, when we performed RCM reaction with 5 mol % catalyst under similar conditions, the reaction was completed in two hours with
Table 3: Application of ring-closing metathesis reactions in the synthesis of macrocycles.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Method(^a) (mol %; time; yield)</th>
<th>RCM product</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>A (5; 2 h; 63%) B (5; 2 h; 84%)</td>
<td><img src="image" alt="4a" /></td>
</tr>
<tr>
<td>3b</td>
<td>A (5; 2 h; 85%) B (5; 2 h; 94%)</td>
<td><img src="image" alt="4b" /></td>
</tr>
<tr>
<td>3c</td>
<td>A (5; 2 h; 88%) B (5; 3 h; 94%)</td>
<td><img src="image" alt="4c" /></td>
</tr>
<tr>
<td>3d</td>
<td>A (5; 2 h; 70%) B (5; 2 h; 90%)</td>
<td><img src="image" alt="4d" /></td>
</tr>
<tr>
<td>3e</td>
<td>A (5+5; 3 h; 88%) B (5+3; 3 h; 40%)</td>
<td><img src="image" alt="4e" /></td>
</tr>
<tr>
<td>3f</td>
<td>A (5+3; 3 h; 83%) B (5; 2 h; 39%)</td>
<td><img src="image" alt="4f" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3g</td>
<td>A (5; 2 h; 77%)</td>
<td>B (5+3; 3 h; 92%)</td>
</tr>
<tr>
<td>3h</td>
<td>A (5+3; 3 h; 82%)</td>
<td>B (5; 2 h; 92%)</td>
</tr>
<tr>
<td>3i</td>
<td>A (5; 2 h; 95%)</td>
<td>B (5; 2 h; 19%)</td>
</tr>
<tr>
<td>3j</td>
<td>A (5; 2 h; 84%)</td>
<td>B (5; 2 h; 56%)</td>
</tr>
<tr>
<td>3k</td>
<td>A (5; 2 h; 81%)</td>
<td>B (5; 2 h; 96%)</td>
</tr>
<tr>
<td>3l</td>
<td>A (5; 2 h; 53%)</td>
<td>B (5; 2 h; 61%)</td>
</tr>
</tbody>
</table>
61% isolated yield. The reaction was performed on the same substrate (i.e., 3a) under high dilution (1 mM) with 5 mol % catalyst at 50 °C in dichloromethane and pleasingly we observed completion of the reaction in two hours with 63% isolated yield. Halogenated solvents are not preferred because of associated health and safety hazards and “greener” solvents for RCM reactions are always required. Ethyl acetate can be chosen as a “green, inexpensive and easily available reaction medium” for metathesis to synthesize this important yet synthetically challenging class of molecules [20]. Therefore, our next attention turns towards using ethyl acetate as “greener” solvent for the macrocyclization reaction. A reaction under high dilution (1 mM) with 5 mol % catalysts at 75 °C in ethyl acetate resulted in macrocycle 4a in 84% yield (Table 3). The structure of 4a was confirmed by 1H NMR based on the disappearance of the signal corresponding to the allyl group (from the starting material) and appearance of multiplet near δ 5.65 ppm for the alkenyl protons. Moreover, the complete structural assignment was done with the help of 2D NMR. It is worth mentioning here that the reaction proceeded with excellent selectivity for the trans product (confirmed by 2D NMR).

Next we performed the macrocyclization reaction with a range of metathesis precursors (3b–m) using dichloromethane and ethyl acetate solvents. Many of the RCM reactions were clean, however, to few the catalyst was added portion-wise until completion of the reaction judged by TLC analysis. The results are summarized in Table 3. Relatively better yields were observed in ethyl acetate compared to dichloromethane when the metathesis precursor consists of pentose (xylose and/or arabinose) building blocks irrespective of the position of the allyl group on the primary or secondary OH group (4a–d). However, yields were significantly lower in ethyl acetate when metathesis precursors were consisting of glucose with protected OH groups and xylose building blocks (4e, 4f). Interestingly, metathesis substrate with both building blocks made-up of glucose with protected OH groups gave significantly better yield in ethyl acetate (4g, 4h). Considerably low yields were observed in ethyl acetate when the metathesis substrate contains a free secondary OH group (4i, 4j). Whereas yield was quite high in ethyl acetate when metathesis substrate contains a free primary OH group (4k). Metathesis yields were relatively higher in ethyl acetate when both glucose derived building blocks were used containing either one free OH group (4l) or two free OH groups (4m). Most notably RCM reactions in ethyl acetate produce almost the same or even better yields than in dichloromethane in most cases (apart from 4e, 4f, 4i, 4j) which confirms ethyl acetate as a viable, greener, inexpensive and easily available alternative to the highly hazardous chlorinated solvent which is a traditionally and most frequently used solvent for RCM reactions.

To check the effect of purity of the cycloadduct on the rate and feasibility of subsequent RCM reactions and on isolated yield obtained in the individual steps, we explored the feasibility of the RCM reaction without isolating the product at the couple phase. Compounds 4e–h were synthesized without purifying the respective cycloaddition products. The second generation Grubbs catalyst catalyzed RCM reaction was performed using the crude substrate in ethyl acetate at 75 °C (Table 4). Interestingly, isolated yields for 4e, 4f and 4h were comparable to the yields obtained when they were synthesized in two separate steps. However, the yield obtained in case of 4g was significantly lower in case of the direct reaction compared to when the compound was synthesized via the two-step process (Table 4).

Lastly, the macrocycle 4m was acetylated in pyridine using acetyl chloride and a catalytic amount of DMAP to furnish diacetate 5. The 1H NMR analysis of 5 clearly showed presence of two singlets at δ 2.08 and 2.06 ppm integrating for three protons each corresponding to acetate methyl groups. Acetate groups were further confirmed by 13C NMR wherein signals corresponding to two carbonyl groups apparent at δ 170.9, 170.5 ppm and two methyl groups at δ 21.0 and 20.9 ppm. The product was further confirmed by mass spectrometry (Scheme 2).
Table 4: Feasibility studies of cycloaddition and RCM reaction in single and two-step protocol.

<table>
<thead>
<tr>
<th>RCM product</th>
<th>CuAAC Yield (%)</th>
<th>Grubbs cat. (mol %)</th>
<th>RCM yield (%)</th>
<th>Combined yield (%)</th>
<th>Grubbs cat. (mol %)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4e</td>
<td>76</td>
<td>(5+3)</td>
<td>40</td>
<td>31</td>
<td>(5+5)</td>
<td>32</td>
</tr>
<tr>
<td>4f</td>
<td>78</td>
<td>5</td>
<td>39</td>
<td>30</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>4g</td>
<td>92</td>
<td>(5+3)</td>
<td>92</td>
<td>85</td>
<td>5</td>
<td>49</td>
</tr>
<tr>
<td>4h</td>
<td>91</td>
<td>5</td>
<td>92</td>
<td>84</td>
<td>5</td>
<td>80</td>
</tr>
</tbody>
</table>

*Isolated yield after column chromatography.*

Scheme 2: Acetylation of macrocycle 4m.

Conclusion

In conclusion we report a novel and green route to synthesize sugar embedded macrocycles (in total we have synthesized 13 macrocycles with 17 to 19-membered rings) which involves CuAAC reaction and Ru-catalyzed RCM reaction. The CuAAC reaction were performed in water and produce moderate yields. Thus, we have successfully demonstrated novel application of build-couple-pair (B/C/P) strategy in DOS and synthesized 13 new macrocycles (4a–m). This synthetic method represents a significant advantage over current routes for sugar embedded macrocycles where reactions are rapid, eco-friendly without compromise in yield and selectivity.

Supporting Information

Supporting Information File 1

Experimental details and analytical data.
[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-110-S1.pdf]

Acknowledgments

S.K.M. thanks Director, CSIR-IHBT for providing the necessary facilities and for financial support (MLP0066) and Ujjwal J. Goswami for his assistance. R.R. acknowledges the junior research fellowship from CSIR-New Delhi. CSIR-IHBT communication no. for this publication is 4064.

References


License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at: doi:10.3762/bjoc.13.110
2-Methyl-2,4-pentanediol (MPD) boosts as detergent-substitute the performance of β-barrel hybrid catalyst for phenylacetylene polymerization

Julia Kinzel¹,§, Daniel F. Sauer²,§, Marco Bocola¹, Marcus Arlt¹, Tayebeh Mirzaei Garakani¹,³, Andreas Thiel², Klaus Beckerle², Tino Polen⁴, Jun Okuda² and Ulrich Schwaneberg*¹,³

Abstract

Covering hydrophobic regions with stabilization agents to solubilize purified transmembrane proteins is crucial for their application in aqueous media. The small molecule 2-methyl-2,4-pentanediol (MPD) was used to stabilize the transmembrane protein Ferric hydroxamate uptake protein component A (FhuA) utilized as host for the construction of a rhodium-based biohybrid catalyst. Unlike commonly used detergents such as sodium dodecyl sulfate or polyethylene polyethylene glycol, MPD does not form micelles in solution. Molecular dynamics simulations revealed the effect and position of stabilizing MPD molecules. The advantage of the amphiphilic MPD over micelle-forming detergents is demonstrated in the polymerization of phenylacetylene, showing a ten-fold increase in yield and increased molecular weights.

Introduction

The combination of a transition metal catalyst and a protein by either dative, supramolecular or covalent means leads to so-called artificial metalloenzymes or biohybrid catalysts [1,2]. Using a non-natural catalyst, the scope of natural enzymes can be expanded or the activity improved. Recent examples are the construction of metatheases [3,4], asymmetric transfer...
hydrogenases [5,6], Diels-Alderases [7-10], an enzyme for carbon–silicon bond formation [11], a phenylacetylene polymerase [12,13] and others [14-17].

A challenge to overcome are unintended substrate–protein interactions, e.g., repulsion of polar substrates with polar amino acid residues [18]. Furthermore, nonpolar substrates are poorly soluble in water and often build a second phase or require a cosolvent. For proteins, these conditions are challenging. The interaction of solvents with the protein can destroy the three dimensional structure and cause protein precipitation [19-21]. To avoid precipitation when using nonpolar substrates, the protein concentration usually is decreased leading to a loss in activity. As an example, the polymerization of phenylacetylene was achieved in water by using the robust β-barrel protein nitrobindin. The selectivity in the polymerization of phenylacetylene was influenced with the protein as second ligand sphere [12,13]. The catalyst achieved a cis/trans ratio of 91:9 in the organic solvent tetrahydrofuran (THF) or being bound on a protein surface without a defined protein environment [12]. By mutations within the cavity of the protein, the ratio was almost inverted to cis/trans 18:82 [13]. Nevertheless, the productivity remained low due to the decreased protein concentration.

A strategy to increase the stability of proteins is the use of whole-cell catalysts. Cells usually show increased stability towards cosolvents, pH and elevated temperatures [22,23]. A recent example in the field of artificial metalloenzymes was shown by Ward and co-workers, who used an artificial metathesis in an in vivo approach. These first attempts are promising to generate artificial whole-cell catalysts. Nevertheless, the productivity with a turnover number of 6 (with respect to the metal content) is yet low [4].

Here, we present a new strategy based on the robust β-barrel protein Ferric hydroxamate uptake protein component A (FhuA, $T_m$ 60–65 °C, refolding after heating up to 85 °C, THF up to 40 vol % tolerated) [19,24-27]. FhuA is one of the largest known outer membrane proteins consisting of 22 antiparallel β-sheets, which are connected through long extracellular loops and short periplasmic turns. After removal of the barrel-plugging “cork” domain ($Δ1$-160), the formed pore (2.5–3.0 nm) is sufficiently large to harbor sterically demanding catalysts and substrates [28,29]. As a transmembrane protein, FhuA needs stabilization of its hydrophobic transmembrane region in an aqueous environment, which is naturally covered by phospholipids in the outer membrane of Escherichia coli (E. coli) [30]. For extraction of membrane proteins, commonly micelle-forming detergents such as sodium dodecyl sulfate (SDS), polyethylene–polyoxethyleneglycol (PE–PEG), sugar glycosides or polyoxethylenes are applied [24,25,28,31,32]. SDS is an efficient detergent for membrane protein solubilization, but is leading to protein unfolding as a drawback. Disadvantageous of detergents is the tremendous reduction of selectivity due to denaturing the protein or the reduction of productivity by detergent micelles since hydrophobic compounds are most likely located inside the hydrophobic micelle core. Recently, the small amphiphilic alcohol 2-methyl-2,4-pentanediol (MPD) was shown to successfully stabilize membrane proteins and enable characterization of protein modifications [33,34]. Polymerization of phenylacetylene in the presence of MPD molecules as refolding agent was carried out, reaching higher molecular weights and yields compared to catalysis with the micelle-forming refolding reagent PE–PEG. Minimum of MPD molecules was analyzed by molecular dynamics studies to enable refolding of SDS-denatured transmembrane protein FhuA $ΔCVF^{tev}$ [29].

This report aims to demonstrate the importance of the right choice of the membrane protein stabilizer for biohybrid catalysis.

Results and Discussion

For solubilizing the transmembrane protein FhuA $ΔCVF^{tev}$ PE–PEG and MPD were applied as stabilizing agent and phenylacetylene polymerization was performed as model reaction (Figure 1).

Molecular dynamics (MD) simulations reveal an optimal minimum number of ≈200 MPD molecules for shielding the hydrophobic transmembrane region of FhuA $ΔCVF^{tev}$

MD simulations of FhuA $ΔCVF^{tev}$ were performed in a box with varying numbers of MPD molecules from 126 MPD, 189 MPD, 252 MPD to 378 MPD molecules as stabilizing cosolvent to investigate the molecular dynamics of protein structure stabilization, how a small amphiphilic molecule could stabilize a transmembrane protein such as FhuA $ΔCVF^{tev}$. All simulations started with a random distribution of MPD, but after a few nanoseconds, the MPD molecules start to cluster around the hydrophobic transmembrane region. Membrane proteins are normally stabilized by incorporation in a protecting membrane layer formed by ionic detergent molecules such as lipids, SDS or nonionic glycolipids. In contrast, in MD simulations with the two highest concentrations MPD forms a small layer of around 200 MPD molecules. The layer is completely covering the transmembrane region and forms a soluble complex, as can be seen in Figure 2A and B). Using less MPD molecules leads to an insufficient coverage (Figure 2C and D) and thus less stabilization of the membrane protein FhuA $ΔCVF^{tev}$. The theoretical calculations are in line with the experimental findings, that
Figure 1: Phenylacetylene polymerization of FhuA ΔCVF[29] refolded in a polymer or small amphiphilic molecule. Refolding agents are essential to solubilize transmembrane proteins and keep membrane proteins refolded by shielding hydrophobic residues in aqueous environments. Refolding of the open channel protein FhuA ΔCVF[29] was on one hand achieved with polyethylene–polyethylene glycol (PE–PEG), which is a micelle-forming detergent. In contrast, refolding of FhuA ΔCVF[29] with 2-methyl-2,4-pentanediol (MPD) prevents micelle formation and leads to increased yield and molecular weight of the corresponding polymer.

Figure 2: Hydrophobic transmembrane region of FhuA ΔCVF[29] stabilized by ≈200 MPD molecules. MPD is illustrated as mainly cyan molecules, water molecules are mainly red. A) A belt of 209 MPD molecules is located close to the transmembrane area. FhuA ΔCVF[29] with 22,374 water and 378 MPD molecules was used as starting condition, in which most MPD molecules diffused away. B) Water molecules in the first solvation sphere (<5 Å) of FhuA ΔCVF[29] are shown to visualize that the transmembrane area of FhuA ΔCVF[29] is completely water free in MD simulations using 378 MPD molecules. C) MD simulations of FhuA ΔCVF[29] with 12,208 water and 126 MPD molecules show that a saturation of the transmembrane region could not be achieved, leading to an incomplete coverage of the hydrophobic belt. D) Water molecules in the first solvation sphere (<5 Å) are partly covering the hydrophobic belt of FhuA ΔCVF[29] using 126 MPD molecules. MPD, 2-methyl-2,4-pentanediol.
FhuA ΔCVF<sup>tev</sup> is properly folded using refolding buffer with 50 mM MPD, which was confirmed by CD spectroscopy (Figure S1, Supporting Information File 1).

**2-Methyl-2,4-pentanediol stabilizes FhuA ΔCVF<sup>tev</sup> up to eight weeks**

Keeping membrane proteins properly folded outside of a biological membrane is a challenging task. Detergents are needed to refold the applied membrane proteins after their extraction from the natural bilayer environment [35-40]. In case of FhuA, so far, refolding has been reported by protecting its hydrophobic transmembrane region in the presence of a detergent such as octylpolyoxyethylene (oPOE) or block copolymer such as PE–PEG [19,28,31,41-44]. Although PE–PEG improves protein solubility, polymerization reactions utilizing FhuA ΔCVF<sup>tev</sup> as protein host in the presence of this copolymer go along with losses in yield due to its micelle-forming property, leading to the need for other types of detergents. Therefore, using a small amphiphilic molecule as an alternative to polymeric detergents is desirable in order to overcome this limitation (Table 1).

In this study, we used the water-miscible amphipathic alcohol MPD (118.18 g/mol) as stabilizing agent in addition to the commonly used PE–PEG [39,45]. The method, originally developed by Michaux and colleagues, consists of using amphipathic cosolvents to refold SDS-denatured proteins and enable them to regain their 3D structure [33,46]. Using MPD is not only beneficial for the polymerization process, but also enables the use of characterization techniques such as transmission electron microscopy and atomic force microscopy. Polymeric detergents are effective protein-stabilizing agents mainly at high concentrations. In contrast, the polymerization using FhuA ΔCVF<sup>tev</sup> could be achieved at lower millimolar concentrations of MPD, which binds tightly to the channel protein. A buffer containing 50 mM MPD was used in the experiments, which contains more than 3 times of the minimum required value for FhuA ΔCVF<sup>tev</sup> (see MD simulation results, Figure 2C and D), ensuring the long-term stability of the protein (Figure S1, Supporting Information File 1). The aforementioned features are consistent with results from circular dichroism (CD) spectroscopy (Figure S1, Supporting Information File 1), showing that FhuA ΔCVF<sup>tev</sup> is correctly folded even up to eight weeks.

**Coupling efficiency of the rhodium catalyst to FhuA ΔCVF<sup>tev</sup> is more than 90%**

The rhodium catalyst 1 bearing a maleimide group was attached to FhuA ΔCVF<sup>tev</sup> for the generation of the biohybrid catalyst [Rh]-FhuA ΔCVF<sup>tev</sup> as previously reported for the Grubbs–Hoveyda type [29,47] or copper complexes [10] (Scheme 1).

FhuA ΔCVF<sup>tev</sup> was dissolved in a solution containing 1.25% SDS. The state of FhuA ΔCVF<sup>tev</sup> is partially unfolded. The catalyst 1 easily accesses the thiol group (Cys545, numbering based on FhuA WT with PDB ID 1BY3 [24]) intro-

---

**Table 1:** Comparison of common solubilizing agents for membrane proteins.<sup>a</sup>

<table>
<thead>
<tr>
<th>Refolding agent</th>
<th>Activity</th>
<th>Selectivity</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS [33]</td>
<td>++</td>
<td>–</td>
<td>unfolding property</td>
</tr>
<tr>
<td>oPOE [19,29]</td>
<td>+</td>
<td>+</td>
<td>costly, micelle formation</td>
</tr>
<tr>
<td>PE–PEG [29]</td>
<td>+</td>
<td>++</td>
<td>bulky, micelle formation</td>
</tr>
<tr>
<td>MPD [33,34]</td>
<td>++</td>
<td>++</td>
<td>small, amphiphatic alcohol, water-miscible</td>
</tr>
</tbody>
</table>

<sup>a</sup>SDS, sodium dodecyl sulfate; oPOE, octylpolyoxyethylene; PE–PEG, polyethylene–polyethylene glycol; MPD, 2-methyl-2,4-pentanediol. ++, very good; +, beneficial; –, non-beneficial.

---

**Scheme 1:** Coupling of [Rh]-1 to the open channel protein FhuA ΔCVF<sup>tev</sup>. SDS, sodium dodecyl sulfate; THF, tetrahydrofuran.
duced for maleimide thiol coupling and a high coupling efficiency is achieved (Figure S2, Supporting Information File 1). After coupling, the excess catalyst is removed by washing the protein residue with THF. The dried biohybrid conjugate 2 is dissolved in water and refolded. As refolding reagents, the block copolymer PE–PEG and amphiphilic MPD are used, respectively. Refolding is achieved by dialysis of the protein in a solution containing the particular refolding agents. The structural integrity of FhuA ΔCVFtev was confirmed with CD spectroscopy (Figure 3).

When either PE–PEG or MPD is applied, the CD spectra for the biohybrid conjugate 2 show typical features of a β-barrel structure (maximum around 195 nm, minimum around 215 nm) [48], indicating a successful refolding of the transmembrane protein FhuA ΔCVFtev with both reagents.

The coupling efficiency was determined by fluorescence titration of the cysteine function of 2 (Cys545) using the fluorescence dye ThioGlo® 1 (fluorescent thiol reagent, Figure S2, Supporting Information File 1). More than 90% of the cysteines are occupied, showing a very high coupling efficiency of the rhodium catalyst. Further, the biohybrid conjugate was analyzed by MALDI–TOF mass spectrometry prior to digestion of 2 with the protease of the Tobacco Etch Virus (TEV) [29,47,49]. Even though the calculated mass of 6,301 Da for the FhuA ΔCVFtev fragment containing Cys545 and the metal catalyst (=6 kDa) could not be observed, the MALDI–TOF mass spectra indicate the successful conjugation of the catalyst by an increase of the molecular weight of 116 Da corresponding to the maleimide group (Figure 4). In studies with other catalysts attached to FhuA ΔCVFtev, the addition of water to the maleimide ring was observed [10,47]. During digestion or ionization, also cleavage

Figure 4: MALDI–TOF mass spectra of apo FhuA ΔCVFtev (red; calculated m/z = 5902.6; found: m/z = 5911.7) and 2 (black; m/z = 6028.5 is assigned to the FhuA fragment containing the maleimide function after water addition). Possible fragmentation of the [Rh] catalyst is indicated. FhuA ΔCVFtev was analyzed after digestion by protease from Tobacco Etch Virus (TEV).
Polymerization of phenylacetylene
The synthesized and characterized biohybrid conjugate based on FhuA ΔCVF<sup>Rev</sup> was used to polymerize phenylacetylene (3, Table 2).

Polymerization of phenylacetylene in THF at 25 °C yields in 65% polymer with M<sub>n</sub> = 5,300 and a high cis-content of 90% (Table 2, entry 1). If the rhodium catalyst is not present, FhuA ΔCVF<sup>Rev</sup> itself is not able to convert the substrate, as expected (Table 2, entry 2). The polymerization reaction of 3 with the biohybrid catalyst 2 is strongly dependent on the choice of the stabilization reagent. In case of PE–PEG, FhuA ΔCVF<sup>Rev</sup> precipitation is observed. Filtering of the solution shows similar results as the reaction with the precipitate present, indicating a deactivation of the catalyst or restricted access of the substrate to the active site. The isolated polymer yield is approximately 5% (Table 2, entry 3). Polymer analysis with gel permeation chromatography (GPC) shows only an oligomeric fraction (M<sub>n</sub> up to 800 g/mol). Applying the refolding reagent MPD, the solution stays clear and turns turbid over time. The yellow to orange color indicates successful polymer formation. The isolated polymer was analyzed by GPC, showing a nearly seven-fold increased molecular weight (M<sub>n</sub> = 5,500) compared to the polymerization reaction with PE–PEG (Table 2, entry 4). Further, the isolation is easier due to facilitated removal of the MPD compared to the polymeric refolding reagent PE–PEG. The isolated yield increased from 5% to 52%. This is related to the increased FhuA ΔCVF<sup>Rev</sup> stability in the presence of hydrophobic substrates. The hydrophobic phenylacetylene interacts with the micelles formed by the PE–PEG refolding reagent, causing the protein precipitation. Experiments utilizing dynamic light scattering (DLS) revealed an interaction of the PE–PEG micelles with the phenylacetylene, showing a decrease in the size distribution of micelles (Figure S3, Supporting Information File 1). Increasing the MPD concentration (up to c(MPD) = 200 mM) did not lead to a significant increased polymer yield.

The selectivity of the polymerization was affected by the FhuA scaffold. Due to the fact that a catalyst is covalently attached inside of a protein scaffold and surrounded by amino acid residues. The protein free catalyst 1 shows a high cis-selectivity (90%). The biohybrid conjugate almost inverts the selectivity, showing 70% trans-selectivity independent from the choice of detergent (Table 2, entry 3 and entry 4). Based on the results of cis/trans ratios not detergents, but the FhuA scaffold leads to changes in selectivity and emphasizes the position of the catalyst inside the barrel. Similar findings were made by Hayashi and co-workers, utilizing the soluble protein nitrobindin as protein scaffold. Upon anchoring of the catalyst to the nitrobindin mutant, the selectivity drastically changed. Further, the group gradually influenced the selectivity by changing the direct environment of the catalyst by introducing sterically demanding amino acids in the protein cavity [13]. Additionally, FhuA ΔCVF<sup>Rev</sup> is stable over the time. As reported by Hayashi and co-workers, the polymerase based on nitrobindin loses structural integrity after 12 hours, resulting in a loss of cis/trans selectivity [13]. The membrane protein FhuA

---

### Table 2: Results of phenylacetylene (3) polymerization catalyzed by biohybrid conjugate 2.<sup>a</sup>

<table>
<thead>
<tr>
<th>Entry&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Catalyst</th>
<th>Stabilization agent&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Isolated yield (%)</th>
<th>M&lt;sub&gt;n&lt;/sub&gt;&lt;sup&gt;d&lt;/sup&gt; (g/mol)</th>
<th>PDI&lt;sup&gt;d&lt;/sup&gt;</th>
<th>trans/cis&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1</td>
<td></td>
<td>13 mg (65)</td>
<td>5,300</td>
<td>4.6</td>
<td>10:90</td>
</tr>
<tr>
<td>2&lt;sup&gt;g&lt;/sup&gt;</td>
<td>FhuA ΔCVF&lt;sup&gt;Rev&lt;/sup&gt;</td>
<td>PE–PEG or MPD</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3&lt;sup&gt;g&lt;/sup&gt;</td>
<td>2</td>
<td>PE–PEG</td>
<td>&lt;1 mg (5)</td>
<td>800</td>
<td>6.0</td>
<td>70:30</td>
</tr>
<tr>
<td>4&lt;sup&gt;η&lt;/sup&gt;</td>
<td>2</td>
<td>MPD</td>
<td>10 mg (52)</td>
<td>5,500</td>
<td>2.9</td>
<td>75:25</td>
</tr>
</tbody>
</table>

<sup>a</sup>THF, tetrahydrofuran; PE–PEG, polyethylene–polyethylene glycol; MPD, 2-methyl-2,4-pentanediol. <sup>b</sup>Buffer: Water containing NaPi (pH 8, 10 mM) and EDTA (1 mM). c(3) = 0.1 M; V<sub>total</sub> = 2 mL. <sup>c</sup>c(c(PE–PEG)) = 0.125 mM; c(MPD) = 50 mM. <sup>d</sup>Determined by GPC. <sup>e</sup>Determined with <sup>1</sup>H NMR spectroscopy. <sup>f</sup>Reaction in THF. <sup>g</sup>Reaction in buffer, containing 10% (v/v) THF. <sup>η</sup>1H NMR spectroscopy.
ACVF$^{\text{tev}}$ in MPD shows stability for more than three days under the reaction conditions and therefore is leading to significantly increased yields.

Conclusion
In conclusion, we successfully demonstrated the use of MPD as a small-molecule stabilizer for utilization of the biohybrid catalyst 2 in phenylacetylene polymerization. The small detergent MPD stabilizes the transmembrane protein FhuA ACVF$^{\text{tev}}$ in aqueous solution without forming micelles. The structural integrity was proven by CD spectroscopy. Applying MPD as a stabilizing agent, an approximately ten-fold increase in yield of poly(phenylacetylene) was obtained compared to reactions in PE–PEG containing solutions. MD simulations revealed the refolding-supporting behavior of the MPD molecules shielding the hydrophobic transmembrane regions of FhuA ACVF$^{\text{tev}}$.

This finding makes the use of membrane proteins more attractive. When using other stabilizing agents, micelle formation decreases the activity by building up an additional diffusion barrier. Furthermore, the formed micelles are influenced by the substrate leading to protein precipitation. The usage of the amphiphilic stabilizer MPD avoided protein precipitation leading to increased yields.

The membrane protein FhuA is robust towards external influences such as increased temperatures and pH values. The catalyst and substrate scope in biohybrid catalysis can be fine-tuned choosing a suitable stabilizing agent as shown in this report. These results may inspire the tailoring of membrane proteins as catalysts in the field of biohybrid catalysis.

Experimental comments
General comments
All used chemicals were of analytical grade or higher quality, purchased from Applichem (Darmstadt, Germany) or Sigma-Aldrich Chemie (Taufkirchen, Germany).

All operations were performed under an inert atmosphere of argon or nitrogen using standard Schlenk or glove box techniques if not mentioned otherwise. Water and other solvents were degassed by using the “freeze-pump-thaw” technique. THF was obtained dry and degassed from a SPS 800 from MBraun (Garching, Germany). Chloroform-$d_1$ was dried over calcium hydride, distilled, degassed and stored in a glove box.

NMR spectra were recorded on a Bruker DRX 400 spectrometer ($^1$H, 400.1 MHz). Chemical shifts were referenced internally by using the residual solvent resonances [50]. MALDI–TOF MS spectra were recorded on an Ultraflex III TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA, US). GPC was measured on an Agilent Series 1100 (Midland, ON, Canada), equipped with two SDV linear N columns of 8 × 300 mm and 8 × 600 mm measures and 5 µm pore size, in THF at 30 °C against a poly(styrene) standard. Dynamic light scattering was performed with a Zetasizer Nano Line (Malvern Instruments, Worcestershire, UK). The rhodium catalyst 2 was synthesized according to literature procedures [12]. Phenylacetylene (3) is commercially available and was used as received. All other chemicals were used as received if not mentioned otherwise.

Expression and extraction of FhuA ACVF$^{\text{tev}}$ with TEV protease
Expression of FhuA ACVF$^{\text{tev}}$ from T7 expression vector pR-IβA1 was performed using the E. coli B$^4$ BL21 (DE3) omp8 strain as expression host according to previous descriptions [29,34,51]. FhuA ACVF$^{\text{tev}}$ was extracted from E. coli with SDS as solubilizing agent as described previously [19,29,34]. Refolding of FhuA ACVF$^{\text{tev}}$ in 1.25% SDS was performed by dialysis against 0.125 mM PE–PEG or 50 mM MPD, respectively [29,34]. Protein concentration was determined by bicinchoninic acid reaction (PierceTM BCA Protein Assay Kit, Thermo Fisher Scientific, Darmstadt, Germany). Refolding buffers are defined as 10.0 mM sodium phosphate buffer pH 8.0 and 1.0 mM EDTA with the addition of 0.125 mM PE–PEG (PE–PEG buffer) or 50.0 mM MPD (MPD buffer) for the purpose of this article.

Cleavage of FhuA ACVF$^{\text{tev}}$ with TEV protease
For analysis of the modification of Cys545 of FhuA ACVF$^{\text{tev}}$ with MALDI–TOF mass spectrometry, two cleavage sites of the TEV protease (ENLYFQ|G) were introduced in the extracellular loop regions 7 and 8 [29]. Protease cleavage was performed as described previously [12,29].

MD simulations
Simulations were based on the X-ray crystal structure of the β-barrel membrane channel protein FhuA WT co-crystallized with the detergent n-octyl-2-hydroxyethyl sulfoxide [24]. The N-terminal cork domain (residue 1–160) blocking the channel was removed. The amino acid exchanges of the hybrid catalyst model FhuA ACVF$^{\text{tev}}$, namely cysteine at position 545, valine at position 548, phenylalanine at position 501 and two flanking TEV-protease recognition sequences in loop 7 and loop 8 [29]. Protease cleavage was performed as described previously [12,29].
water molecules and 126, 189, 252 or 378 randomly distributed MPD molecules as cosolvent [52-56]. The MD calculations (75 ns each) were performed using the AMBER99 force field for the protein and GAFF for MPD cosolvent. The electrostatic interactions were calculated using a 8 Å cut-off and Particle Mesh Ewald [57] for long range electrostatics at pH 7.4 and a density of 0.997 g/mL. The hydrophobic membrane area was covered by an average of 200 MPD molecules in the last 10 ns of the MD simulations, avoiding direct water contact.

Coupling and purification

To a degassed solution of FhuA ΔCVF (5–6 mg/mL) in aqueous SDS solution (1% (w/w) SDS, pH 8 adjusted with NaHCO₃), rhodium catalyst 2 (10 equiv) in THF (10% (v/v)) was added dropwise. The mixture was stirred for 16 h at room temperature. The solvent was removed in vacuo and the residue washed with THF (3 × 15 mL). The residue was dried in vacuo and dissolved in water. Refolding of the biohybrid conjugate was achieved as described above for the apo protein.

Polymerization of phenylacetylene

To an aqueous solution of refolded 2 (2 mL, 10 μM, refolded with either PE–PEG or MPD) in air atmosphere, phenylacetylene (3) in THF (10% (v/v) THF, final concentration of 3 = 0.1 M) was added. The mixtures were stirred at room temperature. After the appropriate reaction time, the polymer was extracted with chloroform, dried in vacuo, washed with water and analyzed by ¹H NMR and GPC as reported previously [12,13].

Supporting Information

Supporting Information File 1
CD spectra of unmodified FhuA ΔCVF directly after refolding and after eight weeks, Thioglo® 1 titration and DLS results. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-148-S1.pdf]

Acknowledgements

The German Federal Ministry of Education and Research (BMBF) is kindly acknowledged for financial support in the framework of the BMBF-Forschertandem “Chiral Membranes”. T.M.G. thanks the Alexander von Humboldt-Stiftung for financial support. We gratefully acknowledge the financial support by the Biology Science Center (BioSC) and the Deutsche Forschungsgemeinschaft (DFG) through the International Research Training Group “Selectivity in Chemo- and Biocatalysis” (SeleCa), and Umicore, Frankfurt (Dr. A. Doppiu), for a generous gift of rhodium precursor.

References

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the Beilstein Journal of Organic Chemistry terms and conditions: (http://www.beilstein-journals.org/bjoc).

The definitive version of this article is the electronic one which can be found at: doi:10.3767/bjoc.13.148
Remarkable functions of \( sn \)-3 hydroxy and phosphocholine groups in 1,2-diacyl-\( sn \)-glycerolipids to induce clockwise (+)-helicity around the 1,2-diacyl moiety: Evidence from conformation analysis by \(^1\)H NMR spectroscopy

Yoshihiro Nishida\(^1\), Mengfei Yuan\(^1\), Kazuo Fukuda\(^1\), Kaito Fujisawa\(^1\), Hirofumi Dohi\(^1\) and Hirotaka Uzawa\(^2\)

Abstract

Cell-membrane glycerolipids exhibit a common structural backbone of asymmetric 1,2-diacyl-\( sn \)-glycerol bearing polar head groups in the \( sn \)-3 position. In this study, the possible effects of \( sn \)-3 head groups on the helical conformational property around the 1,2-diacyl moiety in the solution state were examined. \(^1\)H NMR Karplus relation studies were carried out using a series of 1,2-dipalmitoyl-\( sn \)-glycerols bearing different \( sn \)-3 substituents (namely palmitoyl, benzyl, hydrogen, and phosphates). The \(^1\)H NMR analysis indicated that the helical property around the 1,2-diacyl moiety is considerably affected by these \( sn \)-3 substituents. The \( sn \)-3 hydroxy group induced a unique helical property, which was considerably dependent on the solvents used. In CDCl\(_3\) solution, three staggered conformers, namely gt(\( + \)), gg(\( - \)) and tg, were randomized, while in more polar solvents, the gt(\( + \)) conformer with (+)-helicity was amplified at the expense of gg(\( - \)) and tg conformers. The \( sn \)-3 phosphocholine in phosphatidylcholine exhibited a greater effect on the gt(\( + \)) conformer, which was independent of the solvents used. From the \(^1\)H NMR analysis, the helical conformational properties around the 1,2-diacyl moiety conformed to a simple empirical rule, which permitted the proposal of a conformational diagram for 1,2-dipalmitoyl-\( sn \)-glycerols in the solution states.

Introduction

Glycerophospholipids, constituting the basic elements of cytoplasm bilayer membranes, are responsible for several cell functions [1-3]. These chiral biomolecules have an asymmetric \( sn \)-glycerol backbone. Although \( sn \)-glycerol is symmetric, an \( sn \)-3 phosphate group makes it chiral with an (\( R \))-configuration at the \( sn \)-2 position [4]. Such molecular chirality is crucial to
not only their biological activities but also for their metaphysical properties, as glycerophospholipids comprise elements of fluid membrane [5] and nanoscale vesicles called liposomes [6].

In addition, the chiral \( sn \)-glycerol backbone is composed of acyclic polyols that produce several conformers through the free rotation about each of the C–C single bonds. For example, the free rotation about the \( sn \)-1,2 and \( sn \)-2,3 C–C bonds furnishes nine conformers by the combination of three staggered rotamers, namely \( gt \) (gauche–trans), \( gg \) (gauche–gauche) and \( tg \) (trans–gauche, Figure 1). Conformational flexibility often leads to the ambiguous characterization of acyclic molecules, thereby making it difficult to precisely examine their biological activities. This observation is applicable for cell-membrane glycerophospholipids that have been targets in numerous conformational studies [7-15].

Cell-membrane glycerophospholipids are known to adopt the \( gt(+) \) and \( gg(-) \) conformations around the 1,2-diacyl moiety (Figure 1). From X-ray crystallography data, a common structure in which the 1,2-diacyl chains are aligned in parallel is observed, which adopts either the \( gt(+) \) or \( gg(-) \) conformer [7,10,12]. An analogous conformation has been reportedly observed among \( \alpha \)-glycosyl 1,2-diacyl-\( sn \)-glycerols in the solution state [16]. Probably, the two gauche conformers, namely \( gt(+) \) and \( gg(-) \), are stabilized in a manner so as to permit stacking interactions between the 1,2-diacyl chains.

In our previously reported circular dichroism (CD) studies [17,18], helical conformational properties of a series of 1,2-dibenzoyl-\( sn \)-glycerols bearing different \( sn \)-3 substituting groups were examined. As shown in Figure 1, \( gt(+) \) is one of the gauche conformers with a right-handed (\( + \))-helicity around 1,2-diol, while \( gg(-) \) is another gauche conformer with an antipodal left-handed (\( - \))-helicity. Harada and Nakanishi [19] reported the dibenzoate chirality CD methodology, which helps in the analysis of the chirality originating from the disparity between these two helical conformers. We have found thereby that the 1,2-dibenzoyl moiety favors the right-hand screwed \( gt(+) \) conformer over the left-handed one [17]. The \( gt(+) \)-preference was kept irrespective of the \( sn \)-3 substituting groups and the solvents used. Moreover, a relation in the order as \( gt(+) > gg(-) > tg \) was maintained. On the other hand, the intensity of exciton couplet CD bands changed remarkably among the 1,2-dibenzoyl-\( sn \)-glycerols [18], indicating that the disparity between \( gt(+) \) and \( gg(-) \) conformers varies widely by influences from \( sn \)-3 groups.

Helical properties constitute one of the major factors in determining the molecular chirality [20] of not only proteins and nucleic acids but also simpler biomolecules [17-19] such as acyclic \( sn \)-glycerols and glycerophospholipids [8,21]. In this study, the helical properties of four 1,2-dipalmitoyl-\( sn \)-glycerols 1–4 (Scheme 1) are examined; these 1,2-dipalmitoyl-\( sn \)-glycerols are composed of different substituents (X) at the \( sn \)-3 position, and each of them serves as a representative model for the 1,2-diacyl-\( sn \)-glycerols, as categorized in Scheme 1. Although the exciton chirality CD methodology is not applicable for these 1,2-diacyl-\( sn \)-glycerolipids without an appropriate UV/CD chromophore, \( ^1 \)H NMR spectroscopy will permit the precise determination of their helical conformational properties.

\[ \text{Figure 1: (a) Structures of cell-membrane glycerophospholipids with a common asymmetric 1,2-diacyl-} \text{sn-glycerol-3-phosphate structure and (b) the conformational equilibrium among three staggered conformers, namely} \text{gt(+), gg(-) and tg around the 1,2-diacyl moiety.} \]
Results and Discussion

1. Helical conformational properties of tripalmitin 1 and 3-O-benzyl 1,2-dipalmitoyl-
   sn-glycerol (2) in CDCl₃ solutions

First, the helical property of tripalmitin 1 (entry 1, Table 1) is examined according to a previously reported method [18]. Briefly, fractional populations (%) of the three staggered

\begin{align}
  J_{H1proR,H2}^{gt}(\text{Hz}) &= 3.1gt(+) + 2.8gg(+) + 10.7tg \\
  J_{H1proR,H2}^{gg}(\text{Hz}) &= 10.7gt(+) + 0.9gg(+) + 5.0tg \\
  J_{H1proR,H2}^{tg}(\text{Hz}) &= 2.5gt(+) + 2.3gg(+) + 10.6tg \\
  J_{H1proR,H2}^{gt}(\text{Hz}) &= 10.2gt(+) + 1.3gg(+) + 5.8tg
\end{align}

(1) (2)

From the conformer populations (%), the “helicity index” is determined according to the method previously reported by our group [18].

The result in entry 1 (Table 1) indicates that tripalmitin 1 favors gt(+) with right-handed (+)-helicity compared to gg(−) with left-handed helicity (helical disparity = +6%−7%). According to our previously reported study [18], the disparity, as estimated from Equation 2, is linear with respect to the magnitude and intensity of exciton coupling CD bands, indicating that the 1,2-diacyl moiety in 1 exhibits (+)-chirality corresponding to the equilibrium imbalance between gt(+) and gg(−) conformers as indicated by the helicity index (entry 1 in Table 1). The helical volume of 1 (76% by Equation 2 and 81% by Equation 1) indicates that this glycerolipid favors the two helical conformers in addition to the antiperiplanar tg conformer (ca. 25% by Equation 2) at equilibrium.

Next, the helical property of chiral 3-O-benzyl derivative 2 is examined. In our previously reported CD study [17], the intensity of the exciton couplet CD bands for 3-O-benzyl-1,2-dibenzoyl-sn-glycerol is greater than those of 3-palmitoyl-1,2-dibenzoyl-sn-glycerol. From the preceding result, the replacement of the sn-3 palmitoyl group in 1 with a benzyl ether is expected to enhance the helical property. As can be seen from the result of 2...

---

**Table 1:** ¹H NMR data and helical conformational properties of tripalmitin 1 and 3-O-benzyl derivative 2 in the solution state.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Compound (head X = )</th>
<th>Solvent²</th>
<th>¹H NMR data</th>
<th>Populations (%) of staggered conformers in sn-1,2 position</th>
<th>Helicity index in sn-1,2 position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>δ (ppm)</td>
<td>H1proS H1proR gt(+) gg(−) tg gt(+) gg(−) tg</td>
<td>Equation 1 Equation 2 Equation 2 (Equation 1)</td>
</tr>
<tr>
<td>1</td>
<td>1³ (palmitoyl)</td>
<td>CDCl₃</td>
<td>4.15 6.0 4.29 4.4</td>
<td>44 37 19 41 35 24 + 6 (7) 76 (81)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (-CH₂Ph)</td>
<td>CDCl₃</td>
<td>4.19 6.4 4.34 3.8</td>
<td>52 37 11 49 34 17 + 15 (15) 83 (89)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C/M (10:1)</td>
<td></td>
<td>4.19 6.5 4.34 3.8</td>
<td>53 36 11 50 33 17 + 17 (17) 83 (89)</td>
<td></td>
</tr>
</tbody>
</table>

²C/M (v/v) represents the ratios of the mixed solvents CDCl₃ (C) and methanol-d₄ (M). °Discrimination between H₁proR and H₁proS as well as the acquisition of their ¹H NMR data are carried out according to our previously reported studies [23,24] and in the Materials and methods section of this paper.
(Table 1, entries 2 and 3), the helical disparity (+15%, Equation 1 and Equation 2) increases with the introduction of a benzyl group. This result is in good agreement with our expectation. In addition, the helical volume (%) was increased by 7–8% as compared with that of 1. The 3-O-benzyl group apparently enhances the (+)-chirality around the 1,2-diacyl moiety.

To examine the possible effects of solvents, the helical property of 2 is also examined in a mixed solvent containing ca. 10% methanol-$d_4$ in CDCl$_3$ (C/M 10:1, v/v). The result in entry 3 (Table 1) indicates that the helical property of 2 is marginally affected by protic solvents.

2. Helical conformational property of chiral 1,2-dipalmitoyl-$sn$-glycerol (3) using different solvents

Next, the helical property of 1,2-dipalmitin 3 with a hydroxy (OH) group in the $sn$-3 position is examined. This compound is selected as a representative model of 1,2-diacyl-$sn$-glycerols, which play essential roles in the metabolism and anabolism of glycerolipids [25-28]. Compound 3 is prepared by the catalytic hydrogenolysis of benzyl ether (for the synthetic details, see Supporting Information File 1).

The $^1$H NMR spectrum of 3 in a CDCl$_3$ solution (Figure 2a) shows a pair of double doublet signals of H1$proR$ ($\delta$ 4.32 ppm) and H1$proS$ ($\delta$ 4.23 ppm), which exhibit a spectral feature similar to that of 1 [23]. On the other hand, the signals of H3$proR$ and H3$proS$ in 3 collapse in a narrow region around $\delta$ 3.73 ppm. These observations are in good agreement with the $^1$H NMR data of 3 reported by Vilceze and Bittman [29].

From the analysis of the $^1$H NMR data using Equations 1 and 2, 1,2-dipalmitin 3 in CDCl$_3$ exhibits a very unique helical conformational property. That is, the populations of the gt($+$) and gg($-$) conformers are almost equal to give a helical disparity of around 0% (Table 2, entries 1 and 2). A helical volume of around 75% (Equation 2) is analogous to that observed in 1. In contrast to the $^1$H NMR data of 2, those of 3 showed remark-
able changes in the “mixed solvents” containing methanol-d$_4$ in CDCl$_3$. With the addition of methanol-d$_4$, the H$_1$proR and H$_1$proS signals shift to high and low fields, respectively (Figure 2b). Simultaneously, the H$_3$ signals shift upfield by 0.04 ppm. The shift of these H$_1$ signals increases with an increase in the content of methanol-d$_4$ in the mixed solvents, while the H$_3$ signals are marginally changed; thereafter, their positions are maintained at $\delta$ 3.69 ppm (Figure 2b). As shown in Table 2, entries 1–6, the change in the chemical shifts is related to that in the vicinal coupling constants, indicative of a change in the dynamic conformations occurring around the 1,2-diacyl moiety in 3.

From the analysis of the $^1$H NMR data using the Karplus equations (Equation 1 and 2), an equilibrium shift mainly occurs between the gt(+) and tg conformers. In the mixed solvents with high methanol-d$_4$ contents, the population of the gt(+) conformer seemingly increases at the expense of the tg conformer. The population of the gg(−) conformer decreases by several percent after the addition of ca. 10% of methanol-d$_4$ (Table 2, entry 3). Thereafter, the gg(−) population remains constant at around 35% irrespective of the solvents.

Because of the shift in the equilibrium from tg to gt(+) in the mixed solvents with high methanol contents, the helical disparity (%) and helical volume (%) increase. With an increase in the methanol-d$_4$ content to 17% (C/M 5:1), the helical property of 3 becomes similar to that of 2 (Figure 3). Although this change seems to be saturated in the mixed solvent containing 33% methanol-d$_4$ (C/M 2:1, v/v), the addition of one aliquot of D$_2$O to this solution further changes the gt(+) and tg populations by a few percent (Table 2, entry 6 and Figure 3). Moreover, the H$_2$ signal of 3 shifts downfield by 0.03 ppm in the presence of D$_2$O, although this signal marginally changes in the mixed solvents without D$_2$O.

From the $^1$H NMR spectra in Figure 2, a part of 3 is isomerized to 1,3-isomer during storage in solutions. To examine the possible effects from this isomer, the isomerization is promoted up to 50%, and the $^1$H NMR spectrum of the isomeric mixture is analyzed. This experiment indicates that the presence of the 1,3-isomer marginally affects the $^1$H NMR signals of 3.

As shown in Table 1, entries 2 and 3, the solvents marginally affect the $^1$H NMR signals of 2. Clearly, sn-3 OH plays an essential role in the conformational dynamics, as shown above. The dynamic change is probably caused by solvation by methanol-d$_4$ and/or D$_2$O around the 3-OH group as well as the increasing polarity of the mixed solvent. As judged from the chemical shift change in the H$_3$ signals, the solvation is possibly saturated in the mixed solvent with 10% methanol-d$_4$ (C/M = 10:1). In the solvent containing more than 33% methanol-d$_4$ (C/M = 2:1), the solvation by methanol-d$_4$ might be partly replaced with D$_2$O.

Hamilton et al. [30] employed $^{13}$C NMR spectroscopy to examine the dynamic molecular behavior of 1,2-dilauroyl-sn-glycerol located in liposomes mixed with glycerophospholipids. Their

![Figure 3](https://example.com/figure3.png)
$^{13}$C NMR analysis revealed that the hydration occurring around the carbonyl groups in the 1,2-diacyl moiety triggers the dynamics of the molecular alignments in liposomes. Probably, an analogous phenomenon related to the solvation around sn-3 OH was observed. Thus, solvation is thought to play a key role in the dynamic conformation change around the 1,2-diacyl moiety.

3. Helical conformational properties of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (4, DPPC) and other glycerophospholipids in the solution state

The current $^1$H NMR analysis is extended to four 1,2-dipalmitoyl-sn-glycero phospholipids (Scheme 2) bearing different terminal groups (Y). Large portions of their $^1$H NMR data were collated by Hauser et al. [10]. In our experiment, the $^1$H NMR data of phosphatidylcholine 4 are obtained using the mixed solvent C/M = 10:1.

As shown in Figure 4, the $^1$H NMR spectrum of 4 shows a pair of well-separated double doublet signals of H1proR (δ 4.14 ppm) and H1proS (δ 4.40 ppm). Compared to the other 1,2-diacyl-sn-glycerols 1–3, this phospholipid exhibits a higher vicinal coupling constant to H1proR ($^3$JH1R,H2 = 7.2 Hz) and a lower one to H1proS ($^3$JH1S,H2 = 3.5 Hz). In addition, the difference in the chemical shift ($\Delta\delta$ = 0.26 ppm) between the H1proR and H1proS signals increases in 4. These observations predict that the 1,2-diacyl moiety in 4 exhibits an extremely unique conformational property.

Figure 4: Partial $^1$H NMR spectrum of 4 in a mixture of CDCl$_3$ and methanol-d$_4$ (C/M = 10:1, v/v).

In fact, the $^1$H NMR Karplus analysis indicates that the helical disparity of 4 increases above 30% (Table 3, entries 1 and 2); the disparity is greater than that observed thus far in previously reported studies [16-18]. When previously reported $^1$H NMR data for 4 are examined [8,10,31], the strong (+)-chirality is independent of the solvents used (Table 3, entries 1–4). Moreover, the data in entries 5–7 (Table 3) indicate that this property is commonly observed in the glycerophospholipids listed in Scheme 2, indicating that an sn-3 phosphate group plays a key role. From Table 3, the sn-3 phosphate group can also simultaneously increase the helical volume (%). The helical volumes (%) of 4 using Equation 1 nearly reach the theoretical limit (100%). This result is in good agreement with the conformational properties of cell-membrane glycerophospholipids reported previously [10-15]. On the other hand, in our calculations using Equation 2 as the advanced Karplus equation [18], the helical volumes of these glycerophospholipids are around 90%, which permits the presence of the tg conformer by ca. 10%. Note, that the tg conformer is crucial [32,33] because the antiperiplanar relation is thought to deform lamellar phases and trigger membrane fusion.

With respect to the antiperiplanar tg conformer, Hauser et al. [10] examined the effect of self-assembly using 1,2-dihexanoyl (C6) homologs of glycerophospholipids. They added these acyl homologs into D$_2$O at concentrations less than or greater than the critical micellar concentration. In their $^1$H NMR spectroscopy analysis, the tg conformer is almost absent under the self-assembled conditions [10]. In addition, in our calculation by Equation 2, the helical volume (%) reaches the theoretical limit (100%), and the helical disparity (%) is greater 40% [18]. Probably, cell-membrane glycerophospholipid 4 can adopt the unusual rotational mode, where the 1,2-diacyl chains swing between gt(+) and gg(−) conformers. However, such extraordinary rotation would be possible only when molecules are located under self-assembled conditions.
Table 3: $^1$H NMR data of 1,2-dipalmitoyl-sn-glycero-3-phospholipids and their helical conformational properties in solution states.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Compound</th>
<th>Solventa</th>
<th>$^1$H NMR δ (ppm)</th>
<th>$^3$J (Hz)</th>
<th>Populations (%) of staggered conformers around sn-1,2</th>
<th>Helicity index in sn-1,2 position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>H1proR</td>
<td>H1proS</td>
<td></td>
<td>Equation 1</td>
</tr>
<tr>
<td>1</td>
<td>4 (DPPC)</td>
<td>CDCl$_3$</td>
<td>4.13$^b$</td>
<td>4.40$^b$</td>
<td>66</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>C/M (10:1)</td>
<td>4.14</td>
<td>4.40</td>
<td>62</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>C/M (2:1)</td>
<td>4.16$^c$</td>
<td>4.42$^c$</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>CD$_2$OD</td>
<td>4.18$^d$</td>
<td>4.42$^d$</td>
<td>61</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>DPPE$^c$</td>
<td>C/M (2:1)</td>
<td>4.18$^e$</td>
<td>4.40</td>
<td>59</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>DPPS$^c$</td>
<td>C/M (4:3)</td>
<td>4.19$^e$</td>
<td>4.43</td>
<td>64</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>DPPA$^c$</td>
<td>C/M (2:1)</td>
<td>4.21$^e$</td>
<td>4.40</td>
<td>61</td>
<td>33</td>
</tr>
</tbody>
</table>

$^a$C/M (v/v) represents the ratios of the mixed solvents CDCl$_3$ (C) and methanol-d$_4$ (M). $^b$H NMR data obtained from a database of Spectral Database for Organic Chemistry (SDBS). No. 16108HSP-45-792 in http://sdbs.db.aist.go.jp/sdbs/vgi-bin/direct_frame_top.cgi [31]. $^c$H NMR data from a paper of Hauser et al. [10]. $^d$H NMR data from a paper of Bruzik et al. [8].

4. General trend in the helical conformational properties of 1,2-dipalmitoyl-sn-glycerols 1–4 in the solution state

By plotting the helical disparity (%) obtained by Equation 2 against the population (%) of the gt(+) conformers for glycerolipids 1–4 examined herein, a linear relation (\(y = 1.34x - 50.8, R^2 = 0.976\)) is obtained (Figure 5).

From the linearity, we obtain Equation 3 and Equation 4:

\[
\text{Helical disparity} \% = \frac{[\text{gt}(+) - \text{gg}(-)]}{1.34}\% - 37.9
\]

\[
\text{Population} \% \text{ of gt}(+) = 2.94\times[50.8 - \text{gg}(-)]\%
\]

Equation 3 indicates that the helical disparity (%) increases as a function of gt(+) population (%). Equation 4 indicates that the population (%) of the gt(+) conformer increases at the expense of the gg(−) conformer. When the rule of 100 > gt(+) > 0 (%) is applied to Equation 4, the gg(−) population can assume values in a narrow range between 25% and 51%. At a gg(−) population of 25%, the gt(+) population and helical volume (%) reach their theoretical limits (75% and 100%, respectively). At a gg(−) population of 51%, the gt(+) population reaches 0% (tg = 49%).

When the gt(+) population is arbitrarily changed between 30% (B1 section) and 75% (C2 section) in these empirical formulae, a diagram shown in Figure 6 is obtained. The derived diagram is apparently useful for summarizing the overall helical conformational properties of the four 1,2-dipalmitoyl-sn-glycerols 1–4.

In this diagram, an intersection, denoted by B2, is observed, indicating that the helical disparity becomes 0% when both gt(+) and gg(−) populations are 38%. At this point, the helical volume is 76%, and the tg population is 24%. 1,2-Dipalmitin 3 exhibits a similar behavior when dissolved in CDCl$_3$ (Table 2, entry 2). When methanol-d$_4$ is added to the CDCl$_3$ solution of 3, the gt(+) population increases from 37% up to 50% at the expense of the gg(−) and tg conformers. The observed change is well reproduced in this diagram. Glycerocephospholipid 4 shows the largest gt(+) population (64%) in the CDCl$_3$ solution (Table 3, entry 1). A similar situation is denoted by a section C1, where the populations of gt(+), gg(−) and tg are 64%, 29% and 7%, respectively. These values are in good agreement with the experimental results (Table 3, entry 1).

In Table 4, the applicability of Equation 3 and Equation 4 is evaluated using α-D- and α-L-glucopyranosyl 1,2-dipalmitoyl-sn-glycerols (Table 4, entries 1–4). The helical conformational
properties of these α-glycolipids are determined by Equation 2 applying the $^1$H NMR data reported in a preceding paper [16]. The results of the $^1$H NMR analyses are compared with those calculated by Equation 4. Entries 1–4 (Table 4) indicate that Equation 4 can reproduce also the helical conformational properties of these α-glycolipids.

Conclusion

In this study, a $^1$H NMR spectroscopy analysis of 1,2-dipalmitoyl-sn-glycerols 1–4 in the solution state was carried out to elucidate their helical conformational properties around the 1,2-diacyl moiety. In addition, the possible effects from the substituents at the sn-3 position were evaluated. In the current analysis, the chiral $^2$H-labeled triacylglycerols [23,24] provided a key basis to discriminate between the H1proR and H1proS signals (Materials and methods). Throughout this study, each of the 1,2-dipalmitoyl-sn-glycerols 1–4 exhibited a unique helical property, indicating that not only sn-configurations but also sn-3 substituents govern the helical conformational property around the 1,2-diacyl moiety. The biological systems in nature effectively utilize the sn-3 substituents. For example, the sn-3 OH group in 1,2-diacyl-sn-glycerols is essential for the dynamic conformational behavior, which possibly plays major roles in their biological functions as transmembrane second messengers [25-30,34]. The sn-3 phosphocholine in phosphatidylcholine induced strong (+)-chirality regardless of the solvents used, which should considerably contribute to their functions as activators of membrane-bound glycoproteins [35-37]. The helical conformational properties observed in the four 1,2-dipalmitoyl-sn-glycerols (Scheme 1) conformed to an empirical rule, as shown in Equation 3 and in the diagram shown in Figure 6. This rule revealed that the helical disparity (%)...
linely changes by the function of gt(+) populations, albeit in an allowed range. Probably, the range between B2 and C1 sections in the diagram covers the conformational properties of most 1,2-diacyl-sn-glycero-3-phosphocholines in the solution state. The conformational properties in this region can be characterized by the relation of gt(+) ≥ gg(−) > tg (%), which has been commonly observed here in our preceding studies [16–18].

The 1H NMR spectroscopy analysis was carried out in organic solvents. It is possible that the conclusions obtained herein deviate from those examined under physiological conditions. For example, glycerophospholipids are located in self-assembled lamellar structures that show liquid crystalline properties. Plasma membranes comprise glycerophospholipids which interact with other membrane components such as glycoproteins and sterols [38,39]. Moreover, natural glycerolipids are composed of heterogeneous acyl chains with different alkyl lengths and alkenyl –C=C– bonds. Thus, it will be of high significance in extensional studies to evaluate the helical conformational properties of 1,2-diacyl-sn-glycero-3-phosphocholines in the solvent mixture of CDCl3 and methanol-d4 (C/M = 10:1).

### Materials and Methods

#### Model compounds

Tripalmitin 1 was prepared together with chirally deuterated sn-glycero-3-phosphocholine (4 DPPC) was purchased from Tokyo Kasei Co. Ltd. and used without purification. All the compounds studied here have chemical purities over 95% (1H NMR) except for 3 which isomerizes into the 1,3-diacyl isomer during storage in CDCl3 solution.

<table>
<thead>
<tr>
<th>Table 4: Helical conformational properties of α-D- and α-L-glucopyranosyl 1,2-dipalmitoyl-sn-glycerols in the solvent mixture of CDCl3 and methanol-d4 (C/M = 10:1).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

*aAbbreviations: α-D- or α-L-Glc = α-D- or α-L-glucopyranoside, 1H NMR data in our preceding study [16] are analyzed with Equation 2; calculated values (%) from Equation 4 by adapting the gt population (%) in the 1H NMR spectroscopy analysis.

**Acquisition of the 1H NMR spectral data of H1proR and H1proS signals**

Each of the four glycerolipids 1–4 is dissolved in either CDCl3 or the mixed solvents containing methanol-d4 in CDCl3 (deuterium content > 99.5%) at ca. 10 mM concentrations. 1H NMR spectroscopy is measured on a JEOL 400 MHz or 500 MHz instruments at temperatures between 22–25 °C. Chemical shifts (δ, ppm) and coupling constants (J Hz) of H1proR and H1proS signals are obtained manually with 1H NMR spectra expanded in the region between δ 4.0 ppm and δ 4.5 ppm. The manual process is of high significance for the current 1H NMR analysis since a peak top by computer system does not always point at a weighted center correctly.

The discrimination between H1proR and H1proS signals is another crucial process. In our former studies [22,23], chiral 2H-labelled triacylglycerols were prepared (Scheme 3) and applied for the assignment of these diastereomeric protons, namely H1proR and H1proS. The results have shown an empirical relation between the two H1 signals; the H1proS signals appear downfield from the H1proR signals (δ H1proS > δ H1proR ppm) and have lower smaller coupling constants (J H1proR,H2 > J H1proS,H2 Hz). This rule is maintained among 1,2-diacetyl-, 1,2-dipalmitoyl-, and 1,2-dibenzoyl-sn-glycerols and substituents at the sn-3 position. The validity of this rule is confirmed in a comparative analysis using circular dichroism (CD) spectroscopy [17,18]. The current study applies these relations established in our preceding 1H NMR and CD studies.

**Calculation of fractional populations (%) of three staggered conformers around the 1,2-diacyl group with a Karplus relation**

A general Karplus equation of Haasnoot et al. [40] is extended into the simultaneous linear equations Equation 1 [22] and Equation 2 [18].
Scheme 3: Chirally $^2$H-labelled tripalmitins (1S)- and (1R)-$^1$-$^2$H-1

From the vicinal coupling constants ($^3J$ Hz) of H1proR and H1proS signals, the fractional populations (%) of the three staggered conformers are calculated. Equation 1 is a standard equation, in which the three staggered conformers have the dihedral angles of ± 60° or 180° around 1,2-diols.

Equation 2 is an advanced equation [18], which is optimized for the analysis of 1,2-diacyl-sn-glycerols in the solution state. The results by Equation 1 and Equation 2 produce some deviations each other. In general, Equation 1 tends to overestimate the population (%) of gt(+) and gg(−) conformers by 3–5% compared to those by Equation 2. The current study applies both Equation 1 and Equation 2 in parallel while the main discussion utilizes the results by Equation 2 as the advanced equation.

Definition of ‘helicity index’, ‘helical disparity (%)’ and ‘helical volume (%)’

The ‘helicity index’ [18] comprises three items, namely ‘(+)-’ or ‘(−)-’ sign’, ‘helical disparity (%)’ and ‘helical volume (%)’. The helical disparity (%) is the difference in populations (%) between gt(+) and gg(−) conformers. The disparity has either a ‘(+)-’ or ‘(−)-’ sign’, which corresponds to the sign of exciton couplet CD bands. When the gt(+) conformer is preferred over the gg(−) conformer, the sign is positive. The absolute value in the helical disparity (%) corresponds to the magnitude of the exciton couplet CD bands.

The helical volume (%) is the summation of gt(+) and gg(−) conformers. The volume expresses to what extent a given glycerolipid can adopt the two helical conformers around the 1,2-diacyl moiety. The helical volume (%) may reach the theoretical limit (100%) under self-assembled conditions [18].

Supporting Information

Supporting Information File 1
Experimental and copies of spectra. [http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-13-196-S1.pdf]

Acknowledgements

This work was supported by Grant in Aid from the Japan Society of the Promotion of Science (KAKENHI 25450146, 16K07711). We thank all staffs at Center for Analytical Instrumentation of Chiba University for their technical supports for NMR and other spectroscopic measurements. The authors would like to thank Enago (http://www.enago.jp) for the English language review.

References


