BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Chemical biology

Edited by Helge B. Bode

Generated on 06 November 2026, 07:57

Imprint

Beilstein Journal of Organic Chemistry
www.bjoc.org
ISSN 1860-5397
Email: journals-support@beilstein-institut.de

The *Beilstein Journal of Organic Chemistry* is published by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

Beilstein-Institut zur Förderung der Chemischen Wissenschaften
Trakehner Straße 7–9
60487 Frankfurt am Main
Germany
www.beilstein-institut.de

The copyright to this document as a whole, which is published in the *Beilstein Journal of Organic Chemistry*, is held by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften. The copyright to the individual articles in this document is held by the respective authors, subject to a Creative Commons Attribution license.

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

A detailed view on 1,8-cineol biosynthesis by *Streptomyces clavuligerus*

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

Full Research Paper

Open Access

Address:
Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Email:
Jeroen S. Dickschat* - dickschat@uni-bonn.de

* Corresponding author

Keywords:
biosynthesis; enzyme mechanisms; isotopic labelling; stereochemistry; terpenes

Beilstein J. Org. Chem. **2016,** *12,* 2317–2324.
doi:10.3762/bjoc.12.225

Received: 18 August 2016
Accepted: 28 October 2016
Published: 04 November 2016

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

Guest Editor: H. B. Bode

© 2016 Rinkel et al.; licensee Beilstein-Institut.
License and terms: see end of document.

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-^{13}C)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,I,V)XSXXXE, modified in plants to a DTE triad: DD(L,I,V)XTXXXE) [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 stereogenic 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 monooxygenases and acyl transferases [12,13]. Very few cases are known in which terpene cyclases generate an achiral product as exemplified by the monoterpene 1,8-cineol (eucalyptol, **1**) and the sesquiterpenes germacrene B (**2**) and α-humulene (**3**) (Figure 1).

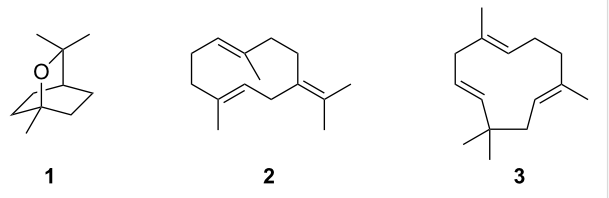


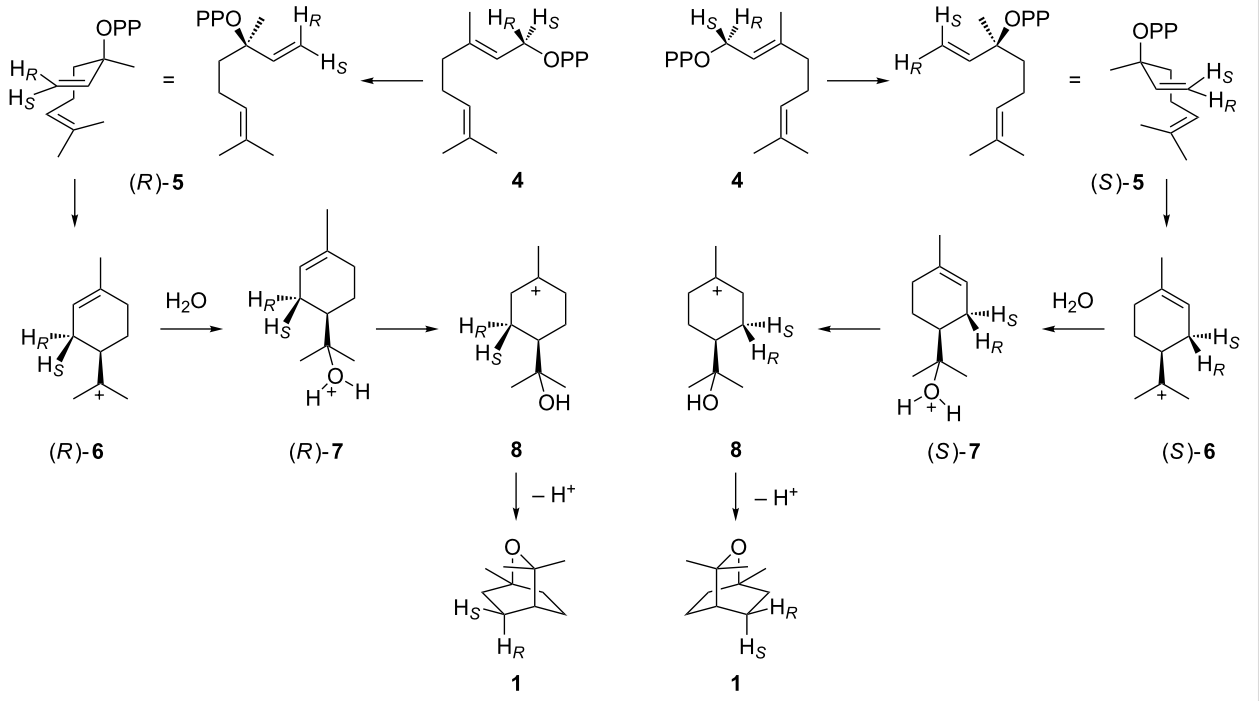
Figure 1: Selection of achiral terpenes.

A direct 1,6-cyclisation of the monoterpene precursor GPP to **1** is prevented by the topological constraints associated with the (2*E*) geometry which necessitates the isomerisation of GPP (**4**) to linalyl diphosphate (LPP, **5**) followed by an *anti*,*endo*-S_N'-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_R for the *pro*-*R* hydrogen and H_S for the *pro*-*S* hydrogen) end up in diastereotopic positions of **1**. Thus, a labelling experiment using the deuterated substrates (*R*)-(1-^{2}H)GPP (H_R = ^{2}H, H_S = H) and (*S*)-(1-^{2}H)GPP (H_R = H, H_S = ^{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-^{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).

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 trichloroacetonitrile [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

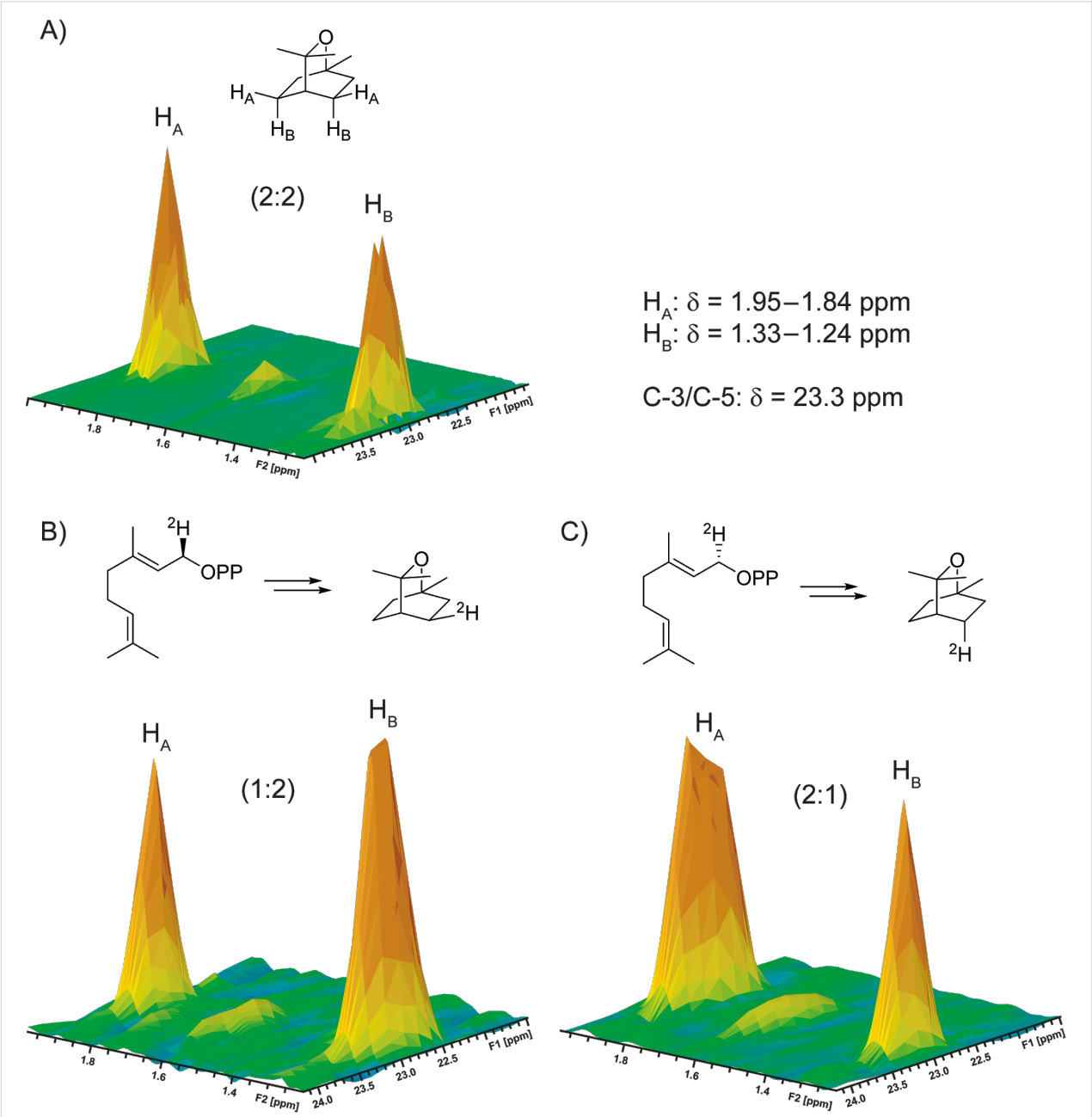


Figure 2: Partial HSQC spectra showing the region of crosspeaks for H_A and H_B connected to C-3 and C-5 of A) unlabelled **1**, B) (^{2}H)-**1** obtained by enzymatic conversion of (*R*)-(1-^{2}H)GPP, and C) (^{2}H)-**1** obtained by enzymatic conversion of (*S*)-(1-^{2}H)GPP. The indicated chemical 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-^{2}H)GPP gave a 1:2 ratio of signal intensities (i.e., $H_A = {}^2H$), while the sample from (*S*)-(1-^{2}H)GPP resulted in ratio of 2:1 by peak integration (i.e., $H_B = {}^2H$), 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 transfered 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 monoterpene synthases are

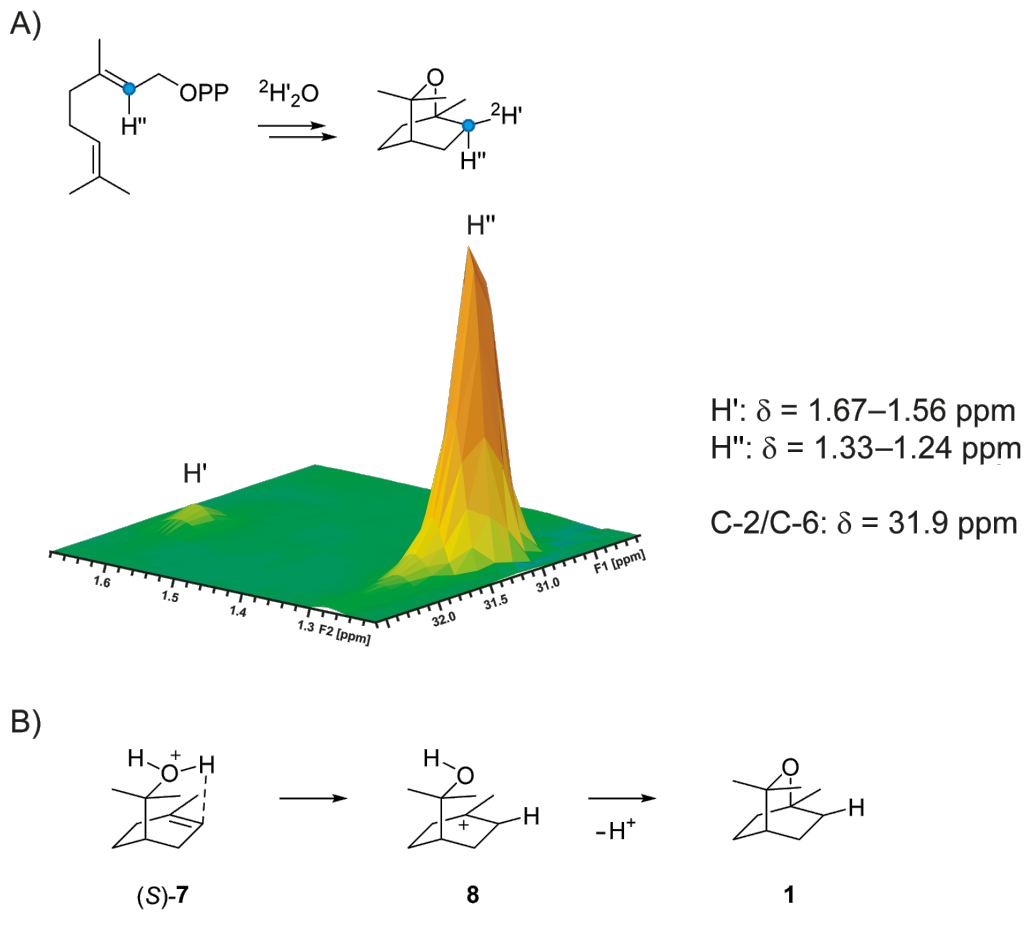


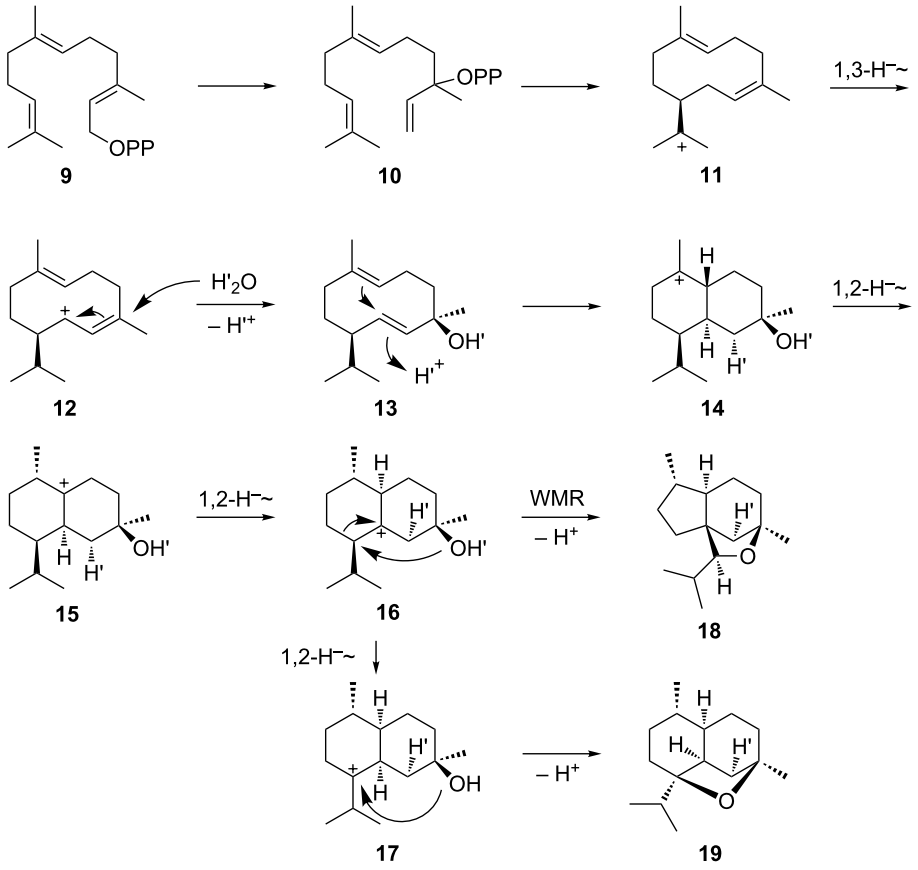
Figure 3: A) Partial HSQC spectrum showing the region of crosspeaks of C-2 with its directly connected hydrogens H' and H'' for compound (2-^{13}C,2-^{2}H)-**1** obtained by enzymatic conversion of (2-^{13}C)GPP in deuterium oxide. Deuterium was specifically incorporated into H' position. The indicated chemical shift data are for unlabelled **1** (for full data cf. Table S1, Supporting Information File 1). Blue circles point to ^{13}C-labelled carbons. B) Intramolecular proton transfer from the protonated hydroxy function in (*S*)-**7** to C-2.

composed of α and β domains and exhibit a quarternary $\alpha_2\beta_2$ 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 α-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 preceeding 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 (TCACCAAGGGGTGGTGGCCC). The isolated PCR product was elongated for homologous recombination with pYE-Express in yeast with a second set of primers (GGCAGCCATATGGCTAGCATGACTGGTGGAATGCCCGCCGGCCACGAAGA and TCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTCACCAAGGGGTGGTGGCCC). 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 FY834 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 (1*R*)- and (1*S*)-(1-^{2}H)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_{600} = 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_2HPO_4, 0.5 mM NaCl, 20 mM imidazole, 1 mM $MgCl_2$, 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^{2+}-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_2HPO_4, 0.5 M NaCl, 0.5 M imidazole, 1 mM $MgCl_2$, pH 7.0) and used directly for incubations with 5 mg of (1*R*)- and (1*S*)-(1-^{2}H)GPP, solved in incubation buffer (50 mM Tris/HCl, 10 mM $MgCl_2$, 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 h, overlaid with 400 µL ($^2H_6$)benzene and further incubated overnight. The organic phase was separated, dried over $MgSO_4$ and directly analysed by GC–MS and NMR.

Incubation experiment with (2-^{13}C)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 ^2H_2O-based binding buffer and elution was done with ^2H_2O-based elution buffer. The first elution fraction was incubated with (2-^{13}C)GPP (0.8 mg) for 16 h at 28 °C. The enzyme reaction was extracted with ($^2H_6$)benzene (0.6 mL), dried with $MgSO_4$ and the extract was analysed directly by GC–MS and NMR.

NMR spectroscopy

To record NMR spectra, instruments AV Avance DMX-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 (^{1}H NMR, residual proton signals: ($^2H_6$)benzene δ = 7.16; ^{13}C NMR: ($^2H_6$)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_8–C_{40}).

Synthesis of (2-^{13}C)geranyl diphosphate

(2-^{13}C)Geraniol was synthesised as reported previously [41]. The synthetic (2-^{13}C)geraniol (16 mg, 0.072 mmol, 1.0 equiv) was dissolved in dry THF (0.3 mL) and PBr_3 (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 $MgSO_4$ 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_4)$_3HP_2O_7$ (97 mg, 0.11 mmol, 1.5 equiv) in dry CH_3CN (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_4^+ form). Elution of the product was performed by addition of two column volumes of ion exchange buffer (0.03 M NH_4HCO_3 in 2% iPrOH/H_2O). Freeze drying yielded the product as a white solid (14.1 mg, 0.04 mmol, 55%).

^{1}H NMR (500 MHz, H_2O) δ 5.37 (dt, $^1J_{(C,H)}$ = 156.7 Hz, $^3J_{(H,H)}$ = 6.9 Hz, 1H, CH), 5.16–5.11 (m, 1H, 1 × CH), 4.44–4.38 (m, 2H, 1 × CH_2), 2.12–2.06 (m, 2H, 1 × CH_2), 2.06–2.00 (m, 2H, 1 × CH_2), 1.65 (d, $^3J_{(C,H)}$ = 5.2 Hz, 3H, 1 × CH_3), 1.62 (s, 3H, 1 × CH_3), 1.56 (s, 3H, 1 × CH_3) ppm; ^{13}C NMR (125 MHz, H_2O) δ 141.9 (d, $^1J_{(C,C)}$ = 72.7 Hz, 1 × C_q), 133.7 (1 × C_q), 124.1 (1 × CH), 119.6 (d, $^3J_{(P,C)}$ = 8.2 Hz, 1 × ^{13}CH), 38.8 (d, $^2J_{(C,C)}$ = 2.7 Hz, 1 × CH_2), 25.6 (d, $^3J_{(C,C)}$ = 2.9 Hz, 1 × CH_2), 24.8 (1 × CH_3), 16.9 (1 × CH_3), 15.6 (d, $^2J_{(C,C)}$ = 1.2 Hz, 1 × CH_3) ppm; ^{31}P NMR (202 MHz, H_2O) δ −10.0 (m, 1 × P), −10.6 (m, 1 × P) ppm.

Supporting Information

Synthesis schemes, Mosher ester analysis of (*R*)- and (*S*)-(1-^{2}H)GPP, gas chromatogram of the enzyme product of 1,8-cineol synthase, ^{13}C NMR of the enzyme product from (2-^{13}C)GPP in deuterium oxide buffer, and full NMR data of **1**.

Supporting Information File 1

Additional material.
[http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-12-225-S1.pdf]

Acknowledgements

This work was funded by the DFG (DI1536/7-1) and by a Ph.D. scholarship of the Fonds der Chemischen Industrie to JR.

References

1. Degenhardt, J.; Köllner, T. G.; Gershenzon, J. *Phytochemistry* **2009,** *70,* 1621–1637. doi:10.1016/j.phytochem.2009.07.030
2. Citron, C. A.; Gleitzmann, J.; Laurenzano, G.; Pukall, R.; Dickschat, J. S. *ChemBioChem* **2012,** *13,* 202–214. doi:10.1002/cbic.201100641
3. Rabe, P.; Citron, C. A.; Dickschat, J. S. *ChemBioChem* **2013,** *14,* 2345–2354. doi:10.1002/cbic.201300329
4. Citron, C. A.; Barra, L.; Wink, J.; Dickschat, J. S. *Org. Biomol. Chem.* **2015,** *13,* 2673–2683. doi:10.1039/C4OB02609H
5. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 857–862. doi:10.1073/pnas.1422108112
6. Quin, M. B.; Flynn, C. M.; Schmidt-Dannert, C. *Nat. Prod. Rep.* **2014,** *31,* 1449–1473. doi:10.1039/C4NP00075G
7. Chen, X.; Köllner, T. G.; Jia, Q.; Norris, A.; Santhanam, B.; Rabe, P.; Dickschat, J. S.; Shaulsky, G.; Gershenzon, J.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 12132–12137. doi:10.1073/pnas.1610379113
8. Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P. *Science* **1997,** *277,* 1815–1820. doi:10.1126/science.277.5333.1815
9. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2014,** *53,* 7652–7656. doi:10.1002/anie.201403648
10. Lesburg, C. A.; Zhai, G.; Cane, D. E.; Christianson, D. W. *Science* **1997,** *277,* 1820–1824. doi:10.1126/science.277.5333.1820
11. Klapschinski, T. A.; Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 10141–10144. doi:10.1002/anie.201605425
12. Bogazkaya, A. M.; von Bühler, C. J.; Kriening, S.; Busch, A.; Seifert, A.; Pleiss, J.; Laschat, S.; Urlacher, V. B. *Beilstein J. Org. Chem.* **2014,** *10,* 1347–1353. doi:10.3762/bjoc.10.137
13. Huber, T.; Weisheit, L.; Magauer, T. *Beilstein J. Org. Chem.* **2015,** *11,* 2521–2539. doi:10.3762/bjoc.11.273
14. Croteau, R.; Alonso, W. R.; Koepp, A. E.; Johnson, M. A. *Arch. Biochem. Biophys.* **1994,** *309,* 184–192. doi:10.1006/abbi.1994.1101
15. Rinkel, J.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2015,** *11,* 2493–2508. doi:10.3762/bjoc.11.271
16. Meguro, A.; Motoyoshi, Y.; Teramoto, K.; Ueda, S.; Totsuka, Y.; Ando, Y.; Tomita, T.; Kim, S.-Y.; Kimura, T.; Igarashi, M.; Sawa, R.; Shinada, T.; Nishiyama, M.; Kuzuyama, T. *Angew. Chem., Int. Ed.* **2015,** *54,* 4353–4356. doi:10.1002/anie.201411923
17. Matsuda, Y.; Mitsuhashi, T.; Lee, S.; Hoshino, N.; Mori, T.; Okada, M.; Zhang, H.; Hayashi, F.; Fujita, M.; Abe, I. *Angew. Chem., Int. Ed.* **2016,** *55,* 5785–5788. doi:10.1002/anie.201601448
18. Okada, M.; Matsuda, Y.; Mitsuhashi, T.; Hoshino, S.; Mori, T.; Nakagawa, K.; Quan, Z.; Qin, B.; Zhang, H.; Hayashi, F.; Kawaide, H.; Abe, I. *J. Am. Chem. Soc.* **2016,** *138,* 10011–10018. doi:10.1021/jacs.6b05799
19. Ye, Y.; Minami, A.; Mandi, A.; Liu, C.; Taniguchi, T.; Kuzuyama, T.; Monde, K.; Gomi, K.; Oikawa, H. *J. Am. Chem. Soc.* **2015,** *137,* 11846–11853. doi:10.1021/jacs.5b08319
20. Burkhardt, I.; Siemon, T.; Henrot, M.; Studt, L.; Rösler, S.; Tudzynski, B.; Christmann, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 8748–8751. doi:10.1002/anie.201603782
21. Rabe, P.; Janusko, A.; Goldfuss, B.; Dickschat, J. S. *ChemBioChem* **2016,** *17,* 146–149. doi:10.1002/cbic.201500543
22. Rabe, P.; Rinkel, J.; Klapschinski, T. A.; Barra, L.; Dickschat, J. S. *Org. Biomol. Chem.* **2016,** *14,* 158–164. doi:10.1039/C5OB01998B
23. Rabe, P.; Pahirulzaman, K. A. K.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2015,** *54,* 6041–6045. doi:10.1002/anie.201501119
24. Rabe, P.; Schmitz, T.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2016,** *12,* 1839–1850. doi:10.3762/bjoc.12.173
25. Dickschat, J. S. *Nat. Prod. Rep.* **2011,** *28,* 1917–1936. doi:10.1039/c1np00063b
26. Wise, M. L.; Savage, T. J.; Katahira, E.; Croteau, R. *J. Biol. Chem.* **1998,** *273,* 14891–14899. doi:10.1074/jbc.273.24.14891
27. Wise, M. L.; Urbansky, M.; Helms, G. L.; Coates, R. M.; Croteau, R. *J. Am. Chem. Soc.* **2002,** *124,* 8546–8547. doi:10.1021/ja0265714
28. Nakano, C.; Kim, H.-K.; Ohnishi, Y. *ChemBioChem* **2011,** *12,* 1988–1991. doi:10.1002/cbic.201100330
29. Edelstein, R. L.; Weller, V. A.; Distefano, M. D.; Tung, J. S. *J. Org. Chem.* **1998,** *63,* 5298–5299. doi:10.1021/jo980304s
30. Thulasiram, H. V.; Phan, R. M.; Rivera, S. B.; Poulter, C. D. *J. Org. Chem.* **2006,** *71,* 1739–1741. doi:10.1021/jo052384n
31. Keller, R. K.; Thompson, R. *J. Chromatogr. A* **1993,** *645,* 161–167. doi:10.1016/0021-9673(93)80630-Q
32. Dickschat, J. S.; Pahirulzaman, K. A. K.; Rabe, P.; Klapschinski, T. A. *ChemBioChem* **2014,** *15,* 810–814. doi:10.1002/cbic.201300763
33. Oldfield, E.; Lin, F.-Y. *Angew. Chem., Int. Ed.* **2012,** *51,* 1124–1137. doi:10.1002/anie.201103110
34. Whittington, D. A.; Wise, M. L.; Urbansky, M.; Coates, R. M.; Croteau, R. B.; Christianson, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 15375–15380. doi:10.1073/pnas.232591099
35. Hyatt, D. C.; Youn, B.; Zhao, Y.; Santhamma, B.; Coates, R. M.; Croteau, R. B.; Kang, C. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 5360–5365. doi:10.1073/pnas.0700915104
36. Baer, P.; Rabe, P.; Citron, C. A.; de Oliveira Mann, C. C.; Kaufmann, N.; Groll, M.; Dickschat, J. S. *ChemBioChem* **2014,** *15,* 213–216. doi:10.1002/cbic.201300708
37. Ding, L.; Goerls, H.; Dornblut, K.; Lin, W.; Maier, A.; Fiebig, H.-H.; Hertweck, C. *J. Nat. Prod.* **2015,** *78,* 2963–2967. doi:10.1021/acs.jnatprod.5b00674
38. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/C5NP00102A
39. Gietz, R. D.; Schiestl, R. H. *Nat. Protoc.* **2007,** *2,* 31–34. doi:10.1038/nprot.2007.13

40. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. *Organometallics* **2010,** *29,* 2176–2179. doi:10.1021/om100106e
41. Rabe, P.; Barra, L.; Rinkel, J.; Riclea, R.; Citron, C. A.; Klapschinski, T. A.; Janusko, A.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2015,** *54,* 13448–13451. doi:10.1002/anie.201507615

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Computational methods in drug discovery

Sumudu P. Leelananda and Steffen Lindert*

Review

Open Access

Address:
Department of Chemistry and Biochemistry, Ohio State University, Columbus, OH 43210, USA

Email:
Steffen Lindert* - lindert.1@osu.edu

* Corresponding author

Keywords:
ADME; computer-aided drug design; docking; free energy; high-throughput screening; LBDD; lead optimization; machine learning; pharmacophore; QSAR; SBDD; scoring; target flexibility

Beilstein J. Org. Chem. **2016,** *12,* 2694–2718.
doi:10.3762/bjoc.12.267

Received: 01 September 2016
Accepted: 22 November 2016
Published: 12 December 2016

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

Guest Editor: H. B. Bode

© 2016 Leelananda and Lindert; licensee Beilstein-Institut.
License and terms: see end of document.

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

ligands that have no possibility of showing success. Therefore, computer-aided drug discovery (CADD) tools are getting a lot of attention in the pharmaceutical industry and academia.

CADD technologies are powerful tools that can reduce the number of ligands that need to be screened in experimental assays. The most popular complementary approach to HTS is the use of virtual (i.e., in silico) HTS. Computer-aided drug discovery and design not only reduces the costs associated with drug discovery by ensuring that best possible lead compound enters animal studies, but it may also reduce the time it takes for a drug to reach the consumer market. It acts as a "virtual shortcut" in the drug discovery pipeline. CADD tools identify lead drug molecules for testing, can predict effectiveness and possible side effects, and assist in improving bioavailability of possible drug molecules. For example, in a recent study of CADD it was found that by introducing a triphenylphosphine group into the base molecule pyridazinone, it is possible to obtain inhibitors for proteasome [2]. Further, analogs have been generated using this starting structure which showed high potency. Many studies show how CADD can influence the development of novel therapeutics [3-6].

CADD methods can be broadly classified into two groups, namely structure-based (SB) and ligand-based (LB) drug discovery (Figure 1). The CADD method used depends on the availability of target structure information. In order to use SBDD tools, information about target structures needs to be known. Target information is usually obtained experimentally by X-ray crystallography or NMR (nuclear magnetic resonance). When neither is available, computational methods such as homology modeling may be used to predict the three-dimensional structures of targets. Knowing the structure makes it possible to use structure-based tools such as virtual high-throughput screening and direct docking methods on targets and possible drug molecules. The affinity of molecules to targets can be evaluated by computing various estimates of binding free energies. Further filtering and optimization of possible drug molecules subsequently follow. The final selected lead molecules are tested in vitro for their activity. When the target structure is not experimentally determined or it is not possible to predict the structure using computational methods, ligand-based approaches are often used as an alternative. These methods, however, rely on the information about known active binders of the target.

CADD has played a significant role in discovering many available pharmaceutical drugs that have obtained FDA approval and reached the consumer market [7-9]. The field of CADD is rapidly improving and new methods and technologies are being

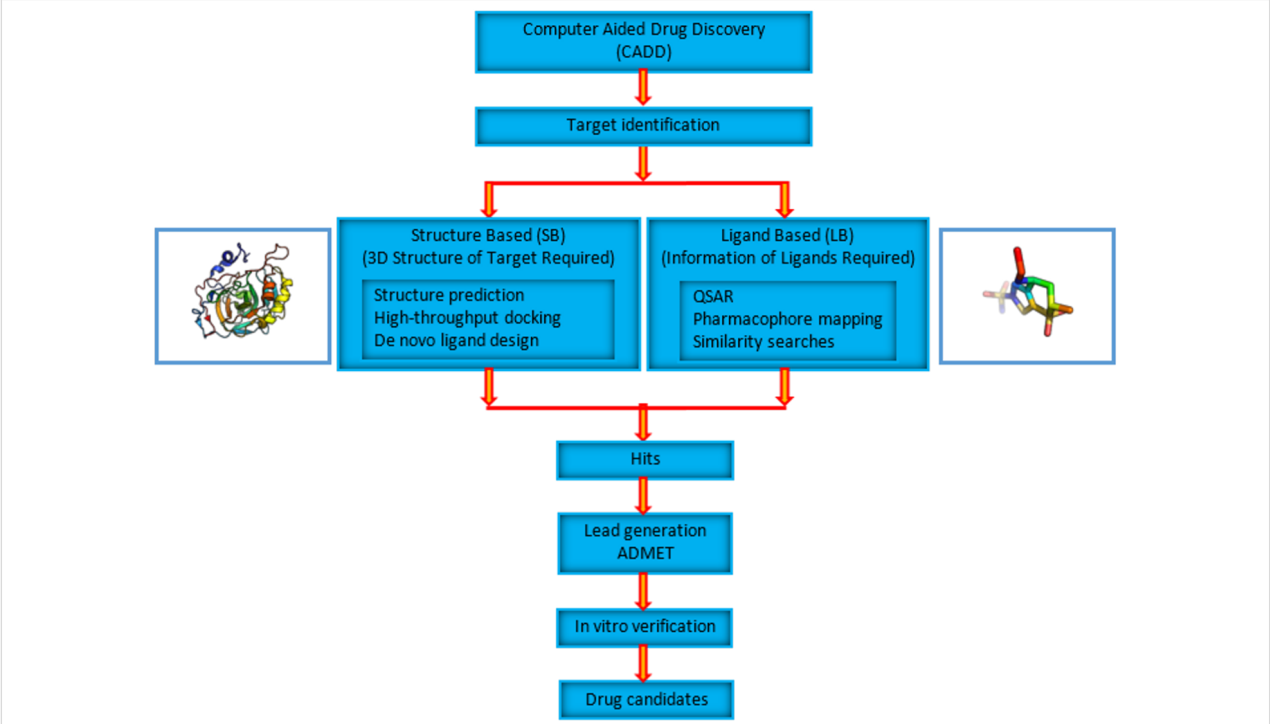


Figure 1: Schematic representation of a computer-aided drug discovery (CADD) pipeline. CADD methods are broadly classified into structure-based and ligand-based methods. Structure-based methods require the 3D information of the target to be known. Ligand-based methods are used when the 3D structure of 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 I) 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

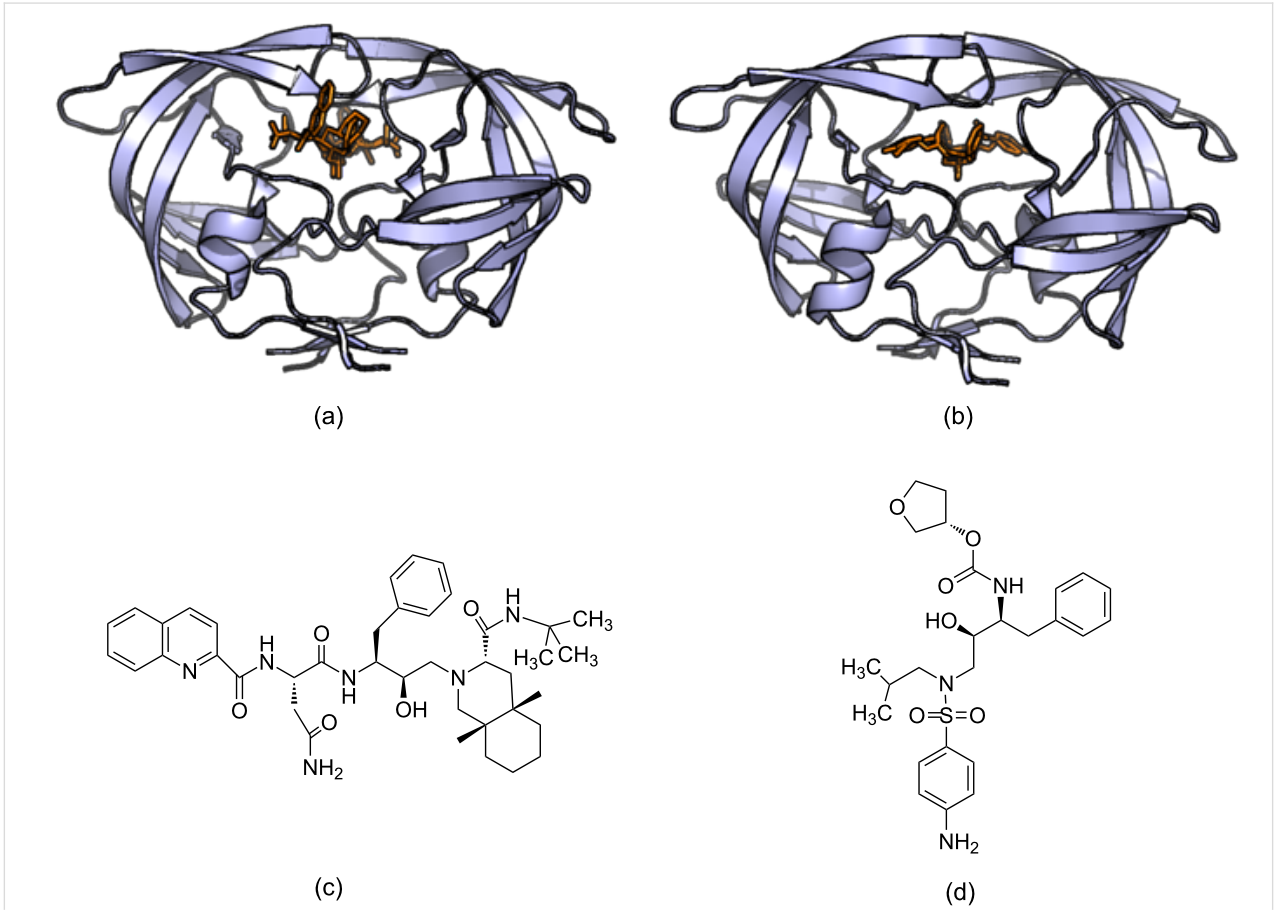


Figure 2: FDA approved drugs Saquinavir and Amprenavir for the treatment of HIV infections. (a) The structure of Saquinavir in complex with HIV-1 protease (3OXC) (b) the structure of Amprenavir in complex with HIV-1 protease (3NU3) (c) the molecular structure of Saquinavir and (d) the molecular structure 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 limitated 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 structure 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

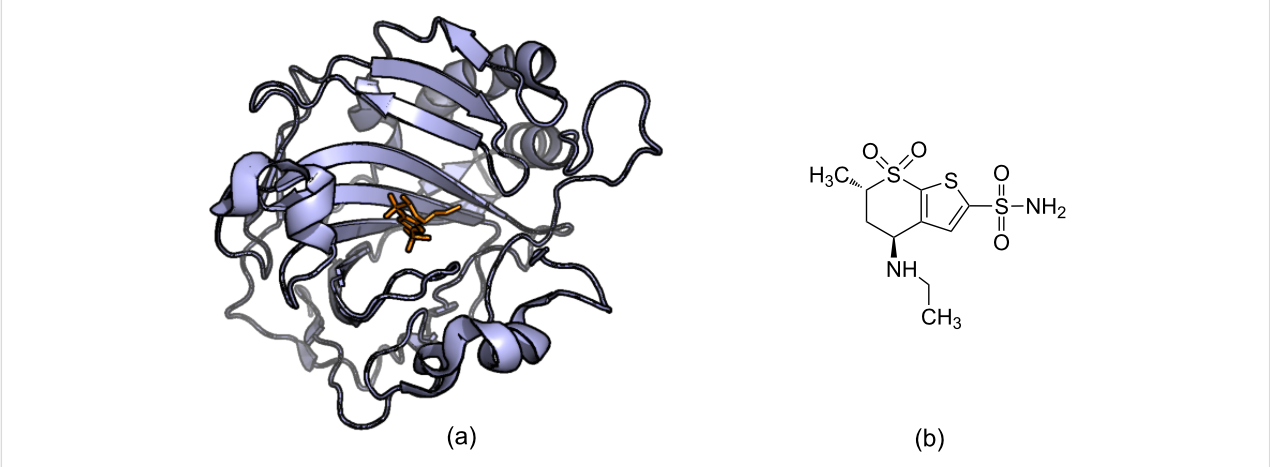


Figure 3: (a) The crystal structure showing the binding of Dorzolamide (orange) to carbonic anhydrase II (purple) (4M2U) (b) the structure of Dorzolamide. Dorzolamide is an FDA approved drug that targets carbonic anhydrase II to treat patience with glaucoma.

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]

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

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.

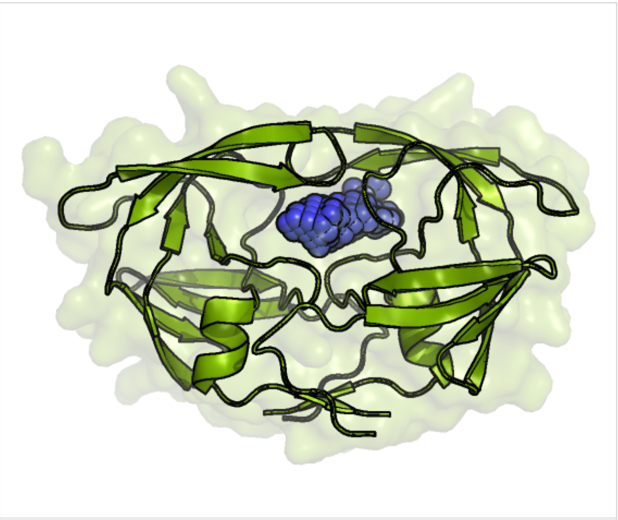


Figure 4: The best ligand binding site identified by SiteHound in HIV-1 protease. The ligand binding pocket is shown in blue spheres and is the known inhibitor binding site of HIV-1 protease.

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

docking algorithm that uses an empirical scoring function [131]. Fred, by OpenEye Scientific, finds protein–ligand docking poses by using a non-stochastic exhaustive method. It uses filters that take shape complementarity into account and the top scoring poses selected are further optimized [136]. Docking algorithms are discussed in detail in reviews [140-142] and comparative assessment of algorithms have been done [112,143-145]. Zhou et al. evaluated the performance of several flexible docking algorithms by calculating enrichment factors for a set of pharmaceutical target–drug complexes and found that Glide XP was superior to other methods tested [144]. The study done by Perola et al. shows that Glide is superior to other methods tested for the prediction of binding poses but virtual screening is mostly target-dependent [112].

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:

$$RMSD = \sqrt{\left(\frac{1}{n}\sum_{x=1}^{n} d_x^2\right)}$$

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.



Figure 5: Binding mode prediction. The known inhibitor Dorzolamide is docked into Carbonic anhydrase II crystal structure (4M2U) (blue) using AutoDock Vina. Four binding poses predicted are shown in green, cyan, red and yellow. The molecular structure of Dorzolamide is shown in Figure 3b.

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 prepossessing 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 anitiviral 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. Schrodinger 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].

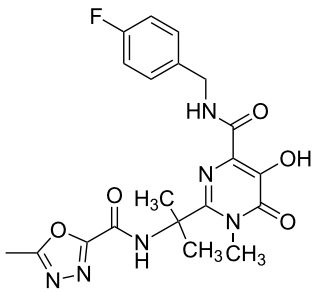


Figure 6: The molecular structure of Raltegravir. Raltegravir is an FDA approved drug used in the treatment of HIV infection.

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

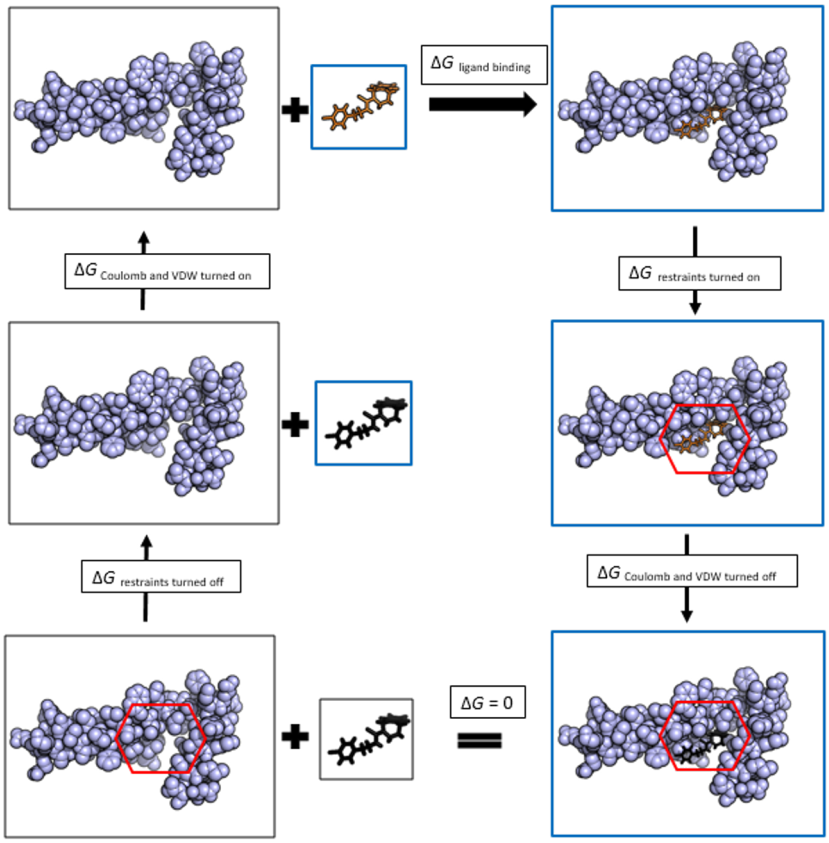


Figure 7: An example alchemical thermodynamic cycle for a protein–ligand binding free energy calculation. The protein is shown in blue spheres. The ligand, depicted in solid black, indicates there are no coulombic or van der Waals (VDW) interactions with the environment. The ligand, depicted in solid 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 force-fields. 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-likeliness 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 pharmaceutically 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

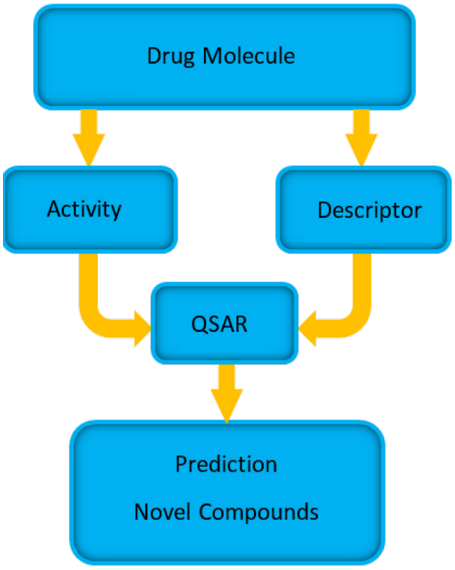


Figure 8: Schematic diagram showing the steps involved in QSAR. Known drug molecule activity and descriptor data is obtained and the mathematical model of QSAR is built such that descriptors can predict the activity of each molecule. The predictive power of models are validated and used in predicting activities of novel compounds.

biological activity. MLR (multivariable linear regression) can be used to find molecular descriptors that have a good correlation with the target–ligand biological activity. It is only possible to use linear regression methods if the activity descriptor relation is linear. However the relationship between biological activity and the molecular descriptors are not always linear [262]. Machine learning approaches such as neural networks and support vector machine methods are used to generate QSAR models to address this issue of non-linear fitting [263-265]. Principal component analysis (PCA) can be used to simplify the complexity by removing the descriptors that are not independent [266]. Once the right set of features is identified and the QSAR is built, these models can be validated using methods such as cross validation [267,268]. QSAR models can be used to predict the biological activity of novel molecules by just using the molecular features. Thus these models can be used to screen a database of molecules to find potential active molecules.

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-

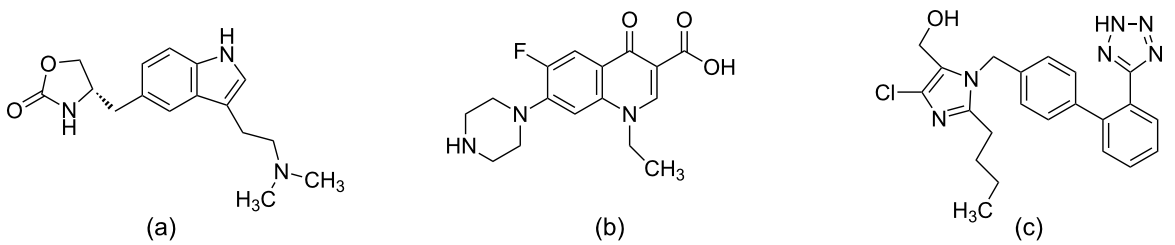


Figure 9: A few drugs discovered with the help of ligand-based drug discovery tools. (a) Zolmitriptan: used as a treatment to migraine (b) Norfloxacin: used in urinary tract infections and (c) Losartan: used to treat hypertension.

tween 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.
2. Yang, L.; Wang, W.; Sun, Q.; Xu, F.; Niu, Y.; Wang, C.; Liang, L.; Xu, P. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 2801–2805. doi:10.1016/j.bmcl.2016.04.067
3. Karthick, V.; Nagasundaram, N.; Doss, C. G. P.; Chakraborty, C.; Siva, R.; Lu, A.; Zhang, G.; Zhu, H. *Infect. Dis. Poverty* **2016,** *5,* No. 12. doi:10.1186/s40249-016-0105-1
4. Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. *J. Chem. Theory Comput.* **2016,** *12,* 2990–2998. doi:10.1021/acs.jctc.6b00201
5. Chao, W. R.; Yean, D.; Amin, K.; Green, C.; Jong, L. *J. Med. Chem.* **2007,** *50,* 3412–3415. doi:10.1021/jm070040e
6. Tran, N.; Van, T.; Nguyen, H.; Le, L. *Int. J. Med. Sci.* **2015,** *12,* 163–176. doi:10.7150/ijms.10826
7. Talele, T. T.; Khedkar, S. A.; Rigby, A. C. *Curr. Top. Med. Chem.* **2010,** *10,* 127–141. doi:10.2174/156802610790232251
8. Clark, D. E. *Expert Opin. Drug Discovery* **2006,** *1,* 103–110. doi:10.1517/17460441.1.2.103
9. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004,** *3,* 935–949. doi:10.1038/nrd1549
10. Wang, Y.; Shaikh, S. A.; Tajkhorshid, E. *Physiology* **2010,** *25,* 142–154. doi:10.1152/physiol.00046.2009
11. Hanson, S. M.; Newstead, S.; Swartz, K. J.; Sansom, M. S. P. *Biophys. J.* **2015,** *108,* 1425–1434. doi:10.1016/j.bpj.2015.02.013
12. Burger, A.; Abraham, D. J. *Burger's Medicinal Chemistry and Drug Discovery, Drug Discovery and Drug Development;* Wiley, 2003.
13. Sham, H. L.; Kempf, D. J.; Molla, A.; Marsh, K. C.; Kumar, G. N.; Chen, C. M.; Kati, W.; Stewart, K.; Lal, R.; Hsu, A.; Betebenner, D.; Korneyeva, M.; Vasavanonda, S.; McDonald, E.; Saldivar, A.; Wideburg, N.; Chen, X.; Niu, P.; Park, C.; Jayanti, V.; Grabowski, B.; Granneman, G. R.; Sun, E.; Japour, A. J.; Leonard, J. M.; Plattner, J. J.; Norbeck, D. W. *Antimicrob. Agents Chemother.* **1998,** *42,* 3218–3224.
14. Jorgensen, W. L. *Science* **2004,** *303,* 1813–1818. doi:10.1126/science.1096361
15. Craig, J. C.; Duncan, I. B.; Hockley, D.; Grief, C.; Roberts, N. A.; Mills, J. S. *Antiviral Res.* **1991,** *16,* 295–305. doi:10.1016/0166-3542(91)90045-S
16. Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *J. Am. Chem. Soc.* **1995,** *117,* 1181–1182. doi:10.1021/ja00108a056
17. Anderson, A. C. *Cell Chem. Biol.* **2003,** *10,* 787–797. doi:10.1016/j.chembiol.2003.09.002
18. Kühlbrandt, W. *Science* **2014,** *343,* 1443–1444. doi:10.1126/science.1251652
19. Yildirim, M. A.; Goh, K.-I.; Cusick, M. E.; Barabasi, A.-L.; Vidal, M. *Nat. Biotechnol.* **2007,** *25,* 1119–1126. doi:10.1038/nbt1338
20. Vyas, V. K.; Ukawala, R. D.; Ghate, M.; Chintha, C. *Indian. J. Pharm. Sci.* **2012,** *74,* 1–17. doi:10.4103/0250-474X.102537

21. Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S. *Curr. Opin. Struct. Biol.* **2008,** *18,* 581–586. doi:10.1016/j.sbi.2008.07.001
22. Grant, M. A. *Comb. Chem. High Throughput Screening* **2009,** *12,* 940–960. doi:10.2174/138620709789824718
23. Lengauer, T.; Zimmer, R. *Briefings Bioinf.* **2000,** *1,* 275–288. doi:10.1093/bib/1.3.275
24. Takeda-Shitaka, M.; Takaya, D.; Chiba, C.; Tanaka, H.; Umeyama, H. *Curr. Med. Chem.* **2004,** *11,* 551–558. doi:10.2174/0929867043455837
25. Whittle, P. J.; Blundell, T. L. *Annu. Rev. Biophys. Biomol. Struct.* **1994,** *23,* 349–375. doi:10.1146/annurev.bb.23.060194.002025
26. Xiang, Z. *Curr. Protein Pept. Sci.* **2006,** *7,* 217–227. doi:10.2174/138920306777452312
27. Krieger, E.; Nabuurs, S. B.; Vriend, G. Homology Modeling. In *Structural Bioinformatics;* Bourne, P. E.; Weissig, H., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA; Vol. 44. doi:10.1002/0471721204.ch25
28. Lemer, C. M.-R.; Rooman, M. J.; Wodak, S. J. *Proteins: Struct., Funct., Bioinf.* **1995,** *23,* 337–355. doi:10.1002/prot.340230308
29. Hardin, C.; Pogorelov, T. V.; Luthey-Schulten, Z. *Curr. Opin. Struct. Biol.* **2002,** *12,* 176–181. doi:10.1016/S0959-440X(02)00306-8
30. Lee, J.; Wu, S.; Zhang, Y. *Ab Initio* Protein Structure Prediction. *From protein structure to function with bioinformatics;* Springer, 2009; pp 3–25. doi:10.1007/978-1-4020-9058-5_1
31. Fischer, D.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94,* 11929–11934. doi:10.1073/pnas.94.22.11929
32. Sánchez, R.; Sali, A. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95,* 13597–13602. doi:10.1073/pnas.95.23.13597
33. Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1980,** *136,* 225–270. doi:10.1016/0022-2836(80)90373-3
34. Illergård, K.; Ardell, D. H.; Elofsson, A. *Proteins: Struct., Funct., Bioinf.* **2009,** *77,* 499–508. doi:10.1002/prot.22458
35. Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* **1991,** *253* (Suppl. 2), 164–170. doi:10.1126/science.1853201
36. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. *Nucleic Acids Res.* **2008,** *36* (Suppl. 2), W5–W9. doi:10.1093/nar/gkn201
37. Liu, T.; Tang, G. W.; Capriotti, E. *Comb. Chem. High Throughput Screening* **2011,** *14,* 532–547. doi:10.2174/138620711795767811
38. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. *Proteins: Struct., Funct., Bioinf.* **2014,** *82,* 1–6. doi:10.1002/prot.24452
39. Kryshtafovych, A.; Fidelis, K.; Moult, J. *Proteins: Struct., Funct., Bioinf.* **2014,** *82,* 164–174. doi:10.1002/prot.24448
40. Cavasotto, C. N.; Phatak, S. S. *Drug Discovery Today* **2009,** *14,* 676–683. doi:10.1016/j.drudis.2009.04.006
41. Chang, C.-e. A.; Ai, R.; Gutierrez, M.; Marsella, M. J. In *Computational Drug Discovery and Design;* Baron, R., Ed.; Springer: New York, NY, 2012; pp 595–613. doi:10.1007/978-1-61779-465-0_35
42. Blundell, T.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. *Eur. J. Biochem.* **1988,** *172,* 513–520. doi:10.1111/j.1432-1033.1988.tb13917.x
43. Guimaraes, A. J.; Hamilton, A. J.; Guedes, H. L. d. M.; Nosanchuk, J. D.; Zancopé-Oliveira, R. M. *PLoS One* **2008,** *3,* e3449. doi:10.1371/journal.pone.0003449
44. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res.* **2003,** *31,* 3381–3385. doi:10.1093/nar/gkg520
45. Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. *Nat. Protoc.* **2008,** *4,* 1–13. doi:10.1038/nprot.2008.197
46. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr Protoc. Bioinformatics;* Unit 5.6; 2006. doi:10.1002/0471250953.bi0506s15
47. Söding, J.; Biegert, A.; Lupas, A. N. *Nucleic Acids Res.* **2005,** *33* (Suppl. 2), W244–W248. doi:10.1093/nar/gki408
48. Mizuguchi, K. *Drug Discovery Today: Targets* **2004,** *3,* 18–23. doi:10.1016/S1741-8372(04)02392-8
49. Ingles-Prieto, A.; Ibarra-Molero, B.; Delgado-Delgado, A.; Perez-Jimenez, R.; Fernandez, J. M.; Gaucher, E. A.; Sanchez-Ruiz, J. M.; Gavira, J. A. *Structure* **2013,** *21,* 1690–1697. doi:10.1016/j.str.2013.06.020
50. McGuffin, L. J. *Comput. Struct. Biol.* **2008,** 37–60. doi:10.1142/9789812778789_0002
51. Jones, D. T.; Taylort, W. R.; Thornton, J. M. *Nature* **1992,** *358,* 86–89. doi:10.1038/358086a0
52. Jones, D. T. *J. Mol. Biol.* **1999,** *287,* 797–815. doi:10.1006/jmbi.1999.2583
53. Wu, S.; Zhang, Y. *Proteins: Struct., Funct., Bioinf.* **2008,** *72,* 547–556. doi:10.1002/prot.21945
54. Yan, R.-X.; Si, J.-N.; Wang, C.; Zhang, Z. *BMC Bioinf.* **2009,** *10,* 416–429. doi:10.1186/1471-2105-10-416
55. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997,** *268,* 209–225. doi:10.1006/jmbi.1997.0959
56. Bradley, P.; Chivian, D.; Meiler, J.; Misura, K. M. S.; Rohl, C. A.; Schief, W. R.; Wedemeyer, W. J.; Schueler-Furman, O.; Murphy, P.; Schonbrun, J.; Strauss, C. E. M.; Baker, D. *Proteins: Struct., Funct., Bioinf.* **2003,** *53,* 457–468. doi:10.1002/prot.10552
57. Xu, D.; Zhang, Y. *Proteins: Struct., Funct., Bioinf.* **2012,** *80,* 1715–1735. doi:10.1002/prot.24065
58. Bates, P. A.; Kelley, L. A.; MacCallum, R. M.; Sternberg, M. J. E. *Proteins: Struct., Funct., Bioinf.* **2001,** *45,* 39–46. doi:10.1002/prot.1168
59. Peng, J.; Xu, J. *Proteins: Struct., Funct., Bioinf.* **2011,** *79,* 161–171. doi:10.1002/prot.23175
60. Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. *Nat. Protoc.* **2012,** *7,* 1511–1522. doi:10.1038/nprot.2012.085
61. Kelley, L. A.; Sternberg, M. J. E. *Nat. Protoc.* **2009,** *4,* 363–371. doi:10.1038/nprot.2009.2
62. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. *Nat. Protoc.* **2015,** *10,* 845–858. doi:10.1038/nprot.2015.053
63. Zhang, Y. *BMC Bioinf.* **2008,** *9,* No. 40. doi:10.1186/1471-2105-9-40
64. Bradley, P.; Misura, K. M. S.; Baker, D. *Science* **2005,** *309,* 1868–1871. doi:10.1126/science.1113801
65. Simons, K. T.; Strauss, C.; Baker, D. *J. Mol. Biol.* **2001,** *306,* 1191–1199. doi:10.1006/jmbi.2000.4459
66. Blaszczyk, M.; Jamroz, M.; Kmiecik, S.; Kolinski, A. *Nucleic Acids Res.* **2013,** *41,* W406–W411. doi:10.1093/nar/gkt462
67. Marks, D. S.; Colwell, L. J.; Sheridan, R.; Hopf, T. A.; Pagnani, A.; Zecchina, R.; Sander, C. *PLoS One* **2011,** *6,* e28766. doi:10.1371/journal.pone.0028766

68. Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Bioinf.* **1998,** *32,* 475–494. doi:10.1002/(SICI)1097-0134(19980901)32:4<475::AID-PROT6>3.0.CO;2-F
69. Latek, D.; Ekonomiuk, D.; Kolinski, A. *J. Comput. Chem.* **2007,** *28,* 1668–1676. doi:10.1002/jcc.20657
70. Thompson, J. M.; Sgourakis, N. G.; Liu, G.; Rossi, P.; Tang, Y.; Mills, J. L.; Szyperski, T.; Montelione, G. T.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 9875–9880. doi:10.1073/pnas.1202485109
71. Li, W.; Zhang, Y.; Kihara, D.; Huang, Y. J.; Zheng, D.; Montelione, G. T.; Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Bioinf.* **2003,** *53,* 290–306. doi:10.1002/prot.10499
72. Bowers, P. M.; Strauss, C. E. M.; Baker, D. *J. Biomol. NMR* **2000,** *18,* 311–318. doi:10.1023/A:1026744431105
73. Lindert, S.; Alexander, N.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. *Structure* **2012,** *20,* 464–478. doi:10.1016/j.str.2012.01.023
74. Lindert, S.; Staritzbichler, R.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. *Structure* **2009,** *17,* 990–1003. doi:10.1016/j.str.2009.06.001
75. Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; Mchaourab, H.; Meiler, J. *Structure* **2008,** *16,* 181–195. doi:10.1016/j.str.2007.11.015
76. Hanson, S. M.; Dawson, E. S.; Francis, D. J.; Van Eps, N.; Klug, C. S.; Hubbell, W. L.; Meiler, J.; Gurevich, V. V. *Structure* **2008,** *16,* 924–934. doi:10.1016/j.str.2008.03.006
77. Hirst, S. J.; Alexander, N.; Mchaourab, H. S.; Meiler, J. *J. Struct. Biol.* **2011,** *173,* 506–514. doi:10.1016/j.jsb.2010.10.013
78. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2015,** *44* (Suppl. D1), D1202–D1213. doi:10.1093/nar/gkv951
79. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J. Chem. Inf. Model.* **2012,** *52,* 1757–1768. doi:10.1021/ci3001277
80. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005,** *45,* 177–182. doi:10.1021/ci049714+
81. Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. *Nucleic Acids Res.* **2006,** *34* (Suppl. 1), D668–D672. doi:10.1093/nar/gkj067
82. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Eur. J. Biochem.* **1977,** *80,* 319–324. doi:10.1111/j.1432-1033.1977.tb11885.x
83. Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L. *Structure* **2012,** *20,* 391–396. doi:10.1016/j.str.2012.01.010
84. Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* **2000,** *28,* 45–48. doi:10.1093/nar/28.1.45
85. Bader, G. D.; Betel, D.; Hogue, C. W. V. *Nucleic Acids Res.* **2003,** *31,* 248–250. doi:10.1093/nar/gkg056
86. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007,** *35* (Suppl. 1), D198–D201. doi:10.1093/nar/gkl999
87. Zheng, X.; Gan, L.; Wang, E.; Wang, J. *AAPS J.* **2013,** *15,* 228–241. doi:10.1208/s12248-012-9426-6
88. Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. *Protein Sci.* **1996,** *5,* 2438–2452.
89. Voss, N. R.; Gerstein, M. *Nucleic Acids Res.* **2010,** *38* (Suppl. 2), W555–W562. doi:10.1093/nar/gkq395
90. Ghersi, D.; Sanchez, R. *J. Struct. Funct. Genomics* **2011,** *12,* 109–117. doi:10.1007/s10969-011-9110-6
91. Laurie, A. T. R.; Jackson, R. M. *Bioinformatics* **2005,** *21,* 1908–1916. doi:10.1093/bioinformatics/bti315
92. Hernandez, M.; Ghersi, D.; Sanchez, R. *Nucleic Acids Res.* **2009,** *37* (Suppl. 2), W413–W416. doi:10.1093/nar/gkp281
93. Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S. *Nucleic Acids Res.* **2012,** *40* (Suppl. Web Server issue), W271–W275. doi:10.1093/nar/gks441
94. Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. *Nat. Protoc.* **2015,** *10,* 733–755. doi:10.1038/nprot.2015.043
95. Saladin, A.; Rey, J.; Thévenet, P.; Zacharias, M.; Moroy, G.; Tufféry, P. *Nucleic Acids Res.* **2014,** *42* (Suppl. Web Server issue), W221–W226. doi:10.1093/nar/gku404
96. Halgren, T. A. *J. Chem. Inf. Model.* **2009,** *49,* 377–389. doi:10.1021/ci800324m
97. Fukunishi, Y.; Nakamura, H. *Protein Sci.* **2011,** *20,* 95–106. doi:10.1002/pro.540
98. Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Mol. Graphics Modell.* **2011,** *29,* 773–776. doi:10.1016/j.jmgm.2010.10.007
99. Till, M. S.; Ullmann, G. M. *J. Mol. Model.* **2010,** *16,* 419–429. doi:10.1007/s00894-009-0541-y
100. Rapp, C. S.; Schonbrun, C.; Jacobson, M. P.; Kalyanaraman, C.; Huang, N. *Proteins: Struct., Funct., Bioinf.* **2009,** *77,* 52–61. doi:10.1002/prot.22415
101. Liu, J.; Wang, R. *J. Chem. Inf. Model.* **2015,** *55,* 475–482. doi:10.1021/ci500731a
102. Huang, S.-Y.; Zou, X. *J. Comput. Chem.* **2006,** *27,* 1866–1875. doi:10.1002/jcc.20504
103. Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000,** *295,* 337–356. doi:10.1006/jmbi.1999.3371
104. Fischer, B.; Fukuzawa, K.; Wenzel, W. *Proteins: Struct., Funct., Bioinf.* **2008,** *70,* 1264–1273. doi:10.1002/prot.21607
105. Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Phys. Chem. Chem. Phys.* **2006,** *8,* 5166–5177. doi:10.1039/B608269F
106. Artemenko, N. *J. Chem. Inf. Model.* **2008,** *48,* 569–574. doi:10.1021/ci700224e
107. Martin, O.; Schomburg, D. *Proteins: Struct., Funct., Bioinf.* **2008,** *70,* 1367–1378. doi:10.1002/prot.21603
108. Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. *J. Chem. Inf. Model.* **2006,** *46,* 380–391. doi:10.1021/ci050283k
109. Teramoto, R.; Fukunishi, H. *J. Chem. Inf. Model.* **2008,** *48,* 288–295. doi:10.1021/ci700239t
110. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. *J. Med. Chem.* **2004,** *47,* 3032–3047. doi:10.1021/jm030489h
111. Huang, S.-Y.; Grinter, S. Z.; Zou, X. *Phys. Chem. Chem. Phys.* **2010,** *12,* 12899–12908. doi:10.1039/C0CP00151A
112. Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins: Struct., Funct., Bioinf.* **2004,** *56,* 235–249. doi:10.1002/prot.20088
113. Wang, R.; Lu, Y.; Fang, X.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2004,** *44,* 2114–2125. doi:10.1021/ci049733j
114. Li, Y.; Han, L.; Liu, Z.; Wang, R. *J. Chem. Inf. Model.* **2014,** *54,* 1717–1736. doi:10.1021/ci500081m
115. Marsden, P. M.; Puvanendrampillai, D.; Mitchell, J. B. O.; Glen, R. C. *Org. Biomol. Chem.* **2004,** *2,* 3267–3273. doi:10.1039/B409570G
116. Velec, H. F. G.; Gohlke, H.; Klebe, G. *J. Med. Chem.* **2005,** *48,* 6296–6303. doi:10.1021/jm050436v

117. Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999,** *42,* 791–804. doi:10.1021/jm980536j
118. Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. *J. Comput. Chem.* **1999,** *20,* 1165–1176. doi:10.1002/(SICI)1096-987X(199908)20:11<1165::AID-JCC7>3.0.CO;2-A
119. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983,** *4,* 187–217. doi:10.1002/jcc.540040211
120. Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981,** *2,* 287–303. doi:10.1002/jcc.540020311
121. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **2001,** *15,* 411–428. doi:10.1023/A:1011115820450
122. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997,** *11,* 425–445. doi:10.1023/A:1007996124545
123. Schneider, N.; Lange, G.; Hindle, S.; Klein, R.; Rarey, M. *J. Comput.-Aided Mol. Des.* **2013,** *27,* 15–29. doi:10.1007/s10822-012-9626-2
124. Wang, R.; Liu, L.; Lai, L.; Tang, Y. *Mol. Model. Ann.* **1998,** *4,* 379–394. doi:10.1007/s008940050096
125. Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999,** *42,* 5100–5109. doi:10.1021/jm990352k
126. Wang, R.; Lai, L.; Wang, S. *J. Comput.-Aided Mol. Des.* **2002,** *16,* 11–26. doi:10.1023/A:1016357811882
127. Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1994,** *8,* 243–256. doi:10.1007/BF00126743
128. Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jørgensen, F. S. *J. Med. Chem.* **2001,** *44,* 2333–2343. doi:10.1021/jm001090l
129. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982,** *161,* 269–288. doi:10.1016/0022-2836(82)90153-X
130. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996,** *261,* 470–489. doi:10.1006/jmbi.1996.0477
131. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004,** *47,* 1739–1749. doi:10.1021/jm0306430
132. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graphics Mod.* **2003,** *21,* 289–307. doi:10.1016/S1093-3263(02)00164-X
133. Goodsell, D. S.; Olson, A. J. *Proteins: Struct., Funct., Bioinf.* **1990,** *8,* 195–202. doi:10.1002/prot.340080302
134. Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010,** *31,* 455–461. doi:10.1002/jcc.21334
135. Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003,** *32,* 335–373. doi:10.1146/annurev.biophys.32.110601.142532
136. McGann, M. *J. Chem. Inf. Model.* **2011,** *51,* 578–596. doi:10.1021/ci100436p
137. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998,** *19,* 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
138. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins: Struct., Funct., Bioinf.* **2003,** *52,* 609–623. doi:10.1002/prot.10465
139. Kramer, B.; Rarey, M.; Lengauer, T. *Proteins: Struct., Funct., Bioinf.* **1999,** *37,* 228–241. doi:10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8
140. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. *Pharmacol. Rev.* **2014,** *66,* 334–395. doi:10.1124/pr.112.007336
141. Irwin, J. J.; Shoichet, B. K. *J. Med. Chem.* **2016,** *59,* 4103–4120. doi:10.1021/acs.jmedchem.5b02008
142. Glaab, E. *Briefings Bioinf.* **2016,** *17,* 352–366. doi:10.1093/bib/bbv037
143. Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. *J. Chem. Inf. Model.* **2006,** *46,* 401–415. doi:10.1021/ci0503255
144. Zhou, Z.; Felts, A. K.; Friesner, R. A.; Levy, R. M. *J. Chem. Inf. Model.* **2007,** *47,* 1599–1608. doi:10.1021/ci7000346
145. Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Proteins: Struct., Funct., Bioinf.* **2004,** *57,* 225–242. doi:10.1002/prot.20149
146. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006,** *49,* 5912–5931. doi:10.1021/jm050362n
147. Elokely, K. M.; Doerksen, R. J. *J. Chem. Inf. Model.* **2013,** *53,* 1934–1945. doi:10.1021/ci400040d
148. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct., Funct., Bioinf.* **2002,** *47,* 409–443. doi:10.1002/prot.10115
149. Schulz-Gasch, T.; Stahl, M. *J. Mol. Model.* **2003,** *9,* 47–57. doi:10.1007/s00894-002-0112-y
150. Stahl, M.; Rarey, M. *J. Med. Chem.* **2001,** *44,* 1035–1042. doi:10.1021/jm0003992
151. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *J. Comput.-Aided Mol. Des.* **2013,** *27,* 221–234. doi:10.1007/s10822-013-9644-8
152. ten Brink, T.; Exner, T. E. *J. Comput.-Aided Mol. Des.* **2010,** *24,* 935–942. doi:10.1007/s10822-010-9385-x
153. *LigPrep,* 2.3; Schrödinger, LLC: New York, NY, 2009.
154. Schüttelkopf, A. W.; Van Aalten, D. M. F. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004,** *60,* 1355–1363. doi:10.1107/S0907444904011679
155. Barreiro, G.; Kim, J. T.; Guimarães, C. R. W.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *J. Med. Chem.* **2007,** *50,* 5324–5329. doi:10.1021/jm070683u
156. Jorgensen, W. L. *Acc. Chem. Res.* **2009,** *42,* 724–733. doi:10.1021/ar800236t
157. Agarwal, A. K.; Johnson, A. P.; Fishwick, C. W. G. *Tetrahedron* **2008,** *64,* 10049–10054. doi:10.1016/j.tet.2008.08.037
158. Sova, M.; Čadež, G.; Turk, S.; Majce, V.; Polanc, S.; Batson, S.; Lloyd, A. J.; Roper, D. I.; Fishwick, C. W. G.; Gobec, S. *Bioorg. Med. Chem. Lett.* **2009,** *19,* 1376–1379. doi:10.1016/j.bmcl.2009.01.034
159. Lindert, S.; Durrant, J. D.; McCammon, J. A. *Chem. Biol. Drug Des.* **2012,** *80,* 358–365. doi:10.1111/j.1747-0285.2012.01414.x
160. Durrant, J. D.; Lindert, S.; McCammon, J. A. *J. Mol. Graphics Modell.* **2013,** *44,* 104–112. doi:10.1016/j.jmgm.2013.05.006
161. Raman, E. P.; Vanommeslaeghe, K.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012,** *8,* 3513–3525. doi:10.1021/ct300088r
162. Faller, C. E.; Raman, E. P.; MacKerell, A. D., Jr.; Guvench, O. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2015,** *1289,* 75–87. doi:10.1007/978-1-4939-2486-8_7
163. Yuan, Y.; Pei, J.; Lai, L. *J. Chem. Inf. Model.* **2011,** *51,* 1083–1091. doi:10.1021/ci100350u
164. Wang, R.; Gao, Y.; Lai, L. *Mol. Model. Ann.* **2000,** *6,* 498–516. doi:10.1007/s0089400060498

165. Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. *J. Med. Chem.* **2003,** *46,* 2656–2662. doi:10.1021/jm030827e
166. Scarsi, M.; Podvinec, M.; Roth, A.; Hug, H.; Kersten, S.; Albrecht, H.; Schwede, T.; Meyer, U. A.; Rücker, C. *Mol. Pharmacol.* **2007,** *71,* 398–406. doi:10.1124/mol.106.024596
167. Lu, Y.; Nikolovska-Coleska, Z.; Fang, X.; Gao, W.; Shangary, S.; Qiu, S.; Qin, D.; Wang, S. *J. Med. Chem.* **2006,** *49,* 3759–3762. doi:10.1021/jm060023+
168. Zhu, J.; Mishra, R. K.; Schiltz, G. E.; Makanji, Y.; Scheidt, K. A.; Mazar, A. P.; Woodruff, T. K. *J. Med. Chem.* **2015,** *58,* 5637–5648. doi:10.1021/acs.jmedchem.5b00753
169. Triballeau, N.; Van Name, E.; Laslier, G.; Cai, D.; Paillard, G.; Sorensen, P. W.; Hoffmann, R.; Bertrand, H.-O.; Ngai, J.; Acher, F. C. *Neuron* **2008,** *60,* 767–774. doi:10.1016/j.neuron.2008.11.014
170. Mueller, R.; Rodriguez, A. L.; Dawson, E. S.; Butkiewicz, M.; Nguyen, T. T.; Oleszkiewicz, S.; Bleckmann, A.; Weaver, C. D.; Lindsley, C. W.; Conn, P. J.; Meiler, J. *ACS Chem. Neurosci.* **2010,** *1,* 288–305. doi:10.1021/cn9000389
171. Lindert, S.; Tallorin, L.; Nguyen, Q. G.; Burkart, M. D.; McCammon, J. A. *J. Comput.-Aided Mol. Des.* **2015,** *29,* 79–87. doi:10.1007/s10822-014-9806-3
172. Liu, Y.-L.; Lindert, S.; Zhu, W.; Wang, K.; McCammon, J. A.; Oldfield, E. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* E2530–E2539. doi:10.1073/pnas.1409061111
173. Henzler-Wildman, K.; Kern, D. *Nature* **2007,** *450,* 964–972. doi:10.1038/nature06522
174. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. *J. Med. Chem.* **2006,** *49,* 534–553. doi:10.1021/jm050540c
175. Meiler, J.; Baker, D. *Proteins: Struct., Funct., Bioinf.* **2006,** *65,* 538–548. doi:10.1002/prot.21086
176. Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A. *Biophys. Chem.* **2014,** *186,* 31–45. doi:10.1016/j.bpc.2013.10.007
177. Sinko, W.; Lindert, S.; McCammon, J. A. *Chem. Biol. Drug Des.* **2013,** *81,* 41–49. doi:10.1111/cbdd.12051
178. Amaro, R. E.; Li, W. W. *Curr. Top. Med. Chem.* **2010,** *10,* 3–13. doi:10.2174/156802610790232279
179. Ellingson, S. R.; Miao, Y.; Baudry, J.; Smith, J. C. *J. Phys. Chem. B* **2015,** *119,* 1026–1034. doi:10.1021/jp506511p
180. Wong, C. F. *Expert Opin. Drug Discovery* **2015,** *10,* 1189–1200. doi:10.1517/17460441.2015.1078308
181. Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput.-Aided Mol. Des.* **2008,** *22,* 693–705. doi:10.1007/s10822-007-9159-2
182. Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *Biopolymers* **2003,** *68,* 47–62. doi:10.1002/bip.10218
183. Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. *J. Med. Chem.* **2004,** *47,* 1879–1881. doi:10.1021/jm0341913
184. Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M. *J. Med. Chem.* **2008,** *51,* 5843–5855. doi:10.1021/jm800245z
185. Hazuda, D. J.; Anthony, N. J.; Gomez, R. P.; Jolly, S. M.; Wai, J. S.; Zhuang, L.; Fisher, T. E.; Embrey, M.; Guare, J. P.; Egbertson, M. S.; Vacca, J. P.; Huff, J. R.; Felock, P. J.; Witmer, M. V.; Stillmock, K. A.; Danovich, R.; Grobler, J.; Miller, M. D.; Espeseth, A. S.; Jin, L.; Chen, I.-W.; Lin, J. H.; Kassahun, K.; Ellis, J. D.; Wong, B. K.; Xu, W.; Pearson, P. G.; Schleif, W. A.; Cortese, R.; Emini, E.; Summa, V.; Holloway, M. K.; Young, S. D. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101,* 11233–11238. doi:10.1073/pnas.0402357101
186. Campbell, A. J.; Lamb, M. L.; Joseph-McCarthy, D. *J. Chem. Inf. Model.* **2014,** *54,* 2127–2138. doi:10.1021/ci400729j
187. Vilar, S.; Costanzi, S. Application of Monte Carlo-Based Receptor Ensemble Docking to Virtual Screening for GPCR Ligands. In *Methods in Enzymology;* Conn, P. M., Ed.; Academic Press, 2013; Vol. 522, pp 263–278. doi:10.1016/B978-0-12-407865-9.00014-5
188. Ivetac, A.; McCammon, J. A. In *Computational Drug Discovery and Design;* Baron, R., Ed.; Springer: New York, NY, 2012; pp 3–12. doi:10.1007/978-1-61779-465-0_1
189. Wells, M. M.; Tillman, T. S.; Mowrey, D. D.; Sun, T.; Xu, Y.; Tang, P. *J. Med. Chem.* **2015,** *58,* 2958–2966. doi:10.1021/jm501873p
190. Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102,* 6679–6685. doi:10.1073/pnas.0408930102
191. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005,** *26,* 1781–1802. doi:10.1002/jcc.20289
192. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999,** *103,* 3596–3607. doi:10.1021/jp984217f
193. Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the 2009 ACM/IEEE Conference on Supercomputing (SC09),* Washington, DC; ACM Press, 2009; pp 1–11.
194. Durrant, J. D.; McCammon, J. A. *BMC Biology* **2011,** *9,* No. 71. doi:10.1186/1741-7007-9-71
195. Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015,** *1850,* 872–877. doi:10.1016/j.bbagen.2014.10.019
196. Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. *Comput. Sci. Discovery* **2011,** *4,* No. 015002. doi:10.1088/1749-4699/4/1/015002
197. Miao, Y.; Feher, V. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2015,** *11,* 3584–3595. doi:10.1021/acs.jctc.5b00436
198. Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008,** *71,* No. 126601. doi:10.1088/0034-4885/71/12/126601
199. Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008,** *100,* 020603. doi:10.1103/PhysRevLett.100.020603
200. Kästner, J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011,** *1,* 932–942. doi:10.1002/wcms.66
201. Selwa, E.; Huynh, T.; Ciccotti, G.; Maragliano, L.; Malliavin, T. E. *Proteins: Struct., Funct., Bioinf.* **2014,** *82,* 2483–2496. doi:10.1002/prot.24612
202. Hu, Y.; Liu, H. *J. Phys. Chem. A* **2014,** *118,* 9272–9279. doi:10.1021/jp503856h
203. Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004,** *120,* 11919. doi:10.1063/1.1755656

204. Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; McCammon, J. A. *J. Chem. Theory Comput.* **2013,** *9,* 4684–4691. doi:10.1021/ct400514p
205. Wereszczynski, J.; McCammon, J. A. In *Computational Drug Discovery and Design;* Baron, R., Ed.; Springer: New York, NY, 2012; pp 515–524. doi:10.1007/978-1-61779-465-0_30
206. Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977,** *23,* 187–199. doi:10.1016/0021-9991(77)90121-8
207. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999,** *314,* 141–151. doi:10.1016/S0009-2614(99)01123-9
208. Chipot, C. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014,** *4,* 71–89. doi:10.1002/wcms.1157
209. Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002,** *41,* 2644–2676. doi:10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O
210. Michel, J.; Essex, J. W. *J. Med. Chem.* **2008,** *51,* 6654–6664. doi:10.1021/jm800524s
211. Michel, J.; Foloppe, N.; Essex, J. W. *Mol. Inf.* **2010,** *29,* 570–578. doi:10.1002/minf.201000051
212. Kollman, P. *Chem. Rev.* **1993,** *93,* 2395–2417. doi:10.1021/cr00023a004
213. Jorgensen, W. L.; Thomas, L. L. *J. Chem. Theory Comput.* **2008,** *4,* 869–876. doi:10.1021/ct800011m
214. Reddy, M. R.; Reddy, C. R.; Rathore, R. S.; Erion, M. D.; Aparoy, P.; Nageswara Reddy, R.; Reddanna, P. *Curr. Pharm. Des.* **2014,** *20,* 3323–3337. doi:10.2174/13816128113199990604
215. Shirts, M. R.; Mobley, D. L.; Brown, S. P. In *Drug Design: Structure- and Ligand-Based Approaches;* Merz, K. M.; Ringe, D.; Reynolds, C. H., Eds.; Cambridge University Press: New York, 2010; pp 61–86.6.
216. Zwanzig, R. W. *J. Chem. Phys.* **1954,** *22,* 1420–1426. doi:10.1063/1.1740409
217. Enyedy, I. J.; Egan, W. J. *J. Comput.-Aided Mol. Des.* **2008,** *22,* 161–168. doi:10.1007/s10822-007-9165-4
218. Gallicchio, E.; Lapelosa, M.; Levy, R. M. *J. Chem. Theory Comput.* **2010,** *6,* 2961–2977. doi:10.1021/ct1002913
219. Gallicchio, E.; Chen, H.; Chen, H.; Fitzgerald, M.; Gao, Y.; He, P.; Kalyanikar, M.; Kao, C.; Lu, B.; Niu, Y.; Pethe, M.; Zhu, J.; Levy, R. M. *J. Comput.-Aided Mol. Des.* **2015,** *29,* 315–325. doi:10.1007/s10822-014-9795-2
220. Chen, W.; Gilson, M. K.; Webb, S. P.; Potter, M. J. *J. Chem. Theory Comput.* **2010,** *6,* 3540–3557. doi:10.1021/ct100245n
221. Bergström, C. A. S. *Basic Clin. Pharmacol. Toxicol.* **2005,** *96,* 156–161. doi:10.1111/j.1742-7843.2005.pto960303.x
222. Fagerberg, J. H.; Karlsson, E.; Ulander, J.; Hanisch, G.; Bergström, C. A. S. *Pharm. Res.* **2015,** *32,* 578–589. doi:10.1007/s11095-014-1487-z
223. Huynh, L.; Grant, J.; Leroux, J.-C.; Delmas, P.; Allen, C. *Pharm. Res.* **2008,** *25,* 147–157. doi:10.1007/s11095-007-9412-3
224. Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; Wang, Y.; Gumbart, J. C. *J. Chem. Inf. Model.* **2016,** *56,* 721–733. doi:10.1021/acs.jcim.6b00022
225. Carpenter, T. S.; Kirshner, D. A.; Lau, E. Y.; Wong, S. E.; Nilmeier, J. P.; Lightstone, F. C. *Biophys. J.* **2014,** *107,* 630–641. doi:10.1016/j.bpj.2014.06.024
226. Balakin, K. V.; Savchuk, N. P.; Tetko, I. V. *Curr. Med. Chem.* **2006,** *13,* 223–241. doi:10.2174/092986706775197917
227. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001,** *46,* 3–26. doi:10.1016/S0169-409X(00)00129-0
228. Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003,** *46,* 1250–1256. doi:10.1021/jm021053p
229. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discovery Today* **2003,** *8,* 876–877. doi:10.1016/S1359-6446(03)02831-9
230. Wang, N.-N.; Dong, J.; Deng, Y.-H.; Zhu, M.-F.; Wen, M.; Yao, Z.-J.; Lu, A.-P.; Wang, J.-B.; Cao, D.-S. *J. Chem. Inf. Model.* **2016,** *56,* 763–773. doi:10.1021/acs.jcim.5b00642
231. Fujiwara, S.-i.; Yamashita, F.; Hashida, M. *Int. J. Pharm.* **2002,** *237,* 95–105. doi:10.1016/S0378-5173(02)00045-5
232. Laoui, A.; Polyakov, V. R. *J. Comput. Chem.* **2011,** *32,* 1944–1951. doi:10.1002/jcc.21778
233. Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000,** *11* (Suppl. 2), S29–S39. doi:10.1016/S0928-0987(00)00162-7
234. Lagorce, D.; Sperandio, O.; Galons, H.; Miteva, M. A.; Villoutreix, B. O. *BMC Bioinf.* **2008,** *9,* No. 396. doi:10.1186/1471-2105-9-396
235. Kennedy, T. *Drug Discovery Today* **1997,** *2,* 436–444. doi:10.1016/S1359-6446(97)01099-4
236. van de Waterbeemd, H.; Gifford, E. *Nat. Rev. Drug Discovery* **2003,** *2,* 192–204. doi:10.1038/nrd1032
237. Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; Wikel, J. H. *J. Pharmacol. Toxicol. Methods* **2000,** *44,* 251–272. doi:10.1016/S1056-8719(00)00109-X
238. Loew, G. H.; Villar, H. O.; Alkorta, I. *Pharm. Res.* **1993,** *10,* 475–486. doi:10.1023/A:1018977414572
239. Mason, J. S.; Good, A. C.; Martin, E. J. *Curr. Pharm. Des.* **2001,** *7,* 567–597. doi:10.2174/1381612013397843
240. Acharya, C.; Coop, A.; Polli, J. E.; MacKerell, A. D. *Curr. Comput.-Aided Drug Des.* **2011,** *7,* 10–22. doi:10.2174/157340911793743547
241. Vogt, M.; Bajorath, J. In *Chemoinformatics and Computational Chemical Biology;* Bajorath, J., Ed.; Humana Press: Totowa, NJ, 2011; pp 159–173.
242. Yang, S.-Y. *Drug Discovery Today* **2010,** *15,* 444–450. doi:10.1016/j.drudis.2010.03.013
243. Verma, J.; Khedkar, V. M.; Coutinho, E. C. *Curr. Top. Med. Chem.* **2010,** *10,* 95–115. doi:10.2174/156802610790232260
244. Klopmand, G. *J. Comput. Chem.* **1992,** *13,* 539–540. doi:10.1002/jcc.540130415
245. Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. *Nat. Chem. Biol.* **2006,** *2,* 207–212. doi:10.1038/nchembio775
246. Lindert, S.; Zhu, W.; Liu, Y.-L.; Pang, R.; Oldfield, E.; McCammon, J. A. *Chem. Biol. Drug Des.* **2013,** *81,* 742–748. doi:10.1111/cbdd.12121
247. Zhu, W.; Zhang, Y.; Sinko, W.; Hensler, M. E.; Olson, J.; Molohon, K. J.; Lindert, S.; Cao, R.; Li, K.; Wang, K.; Wang, Y.; Liu, Y.-L.; Sankovsky, A.; de Oliveira, C. A. F.; Mitchell, D. A.; Nizet, V.; McCammon, J. A.; Oldfield, E. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 123–128. doi:10.1073/pnas.1219899110
248. Lin, S.-K. *Molecules* **2000,** *5,* 987–989. doi:10.3390/50700987
249. Langer, T.; Krovat, E. M. *Curr. Opin. Drug Discovery Dev.* **2003,** *6,* 370–376.

250. Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. *J. Comput.-Aided Mol. Des.* **2002,** *16,* 653–681. doi:10.1023/A:1021954728347
251. Verma, R. P.; Hansch, C. *Chem. Rev.* **2009,** *109,* 213–235. doi:10.1021/cr0780210
252. Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985,** *27,* 493–498.
253. Gupta, S. P.; Kumaran, S. *J. Enzyme Inhib. Med. Chem.* **2005,** *20,* 251–259. doi:10.1080/14756360500067439
254. Kontogiorgis, C. A.; Hadjipavlou-Litina, D. *Curr. Med. Chem.* **2003,** *10,* 525–577. doi:10.2174/0929867033457935
255. Stürzebecher, J.; Prasa, D.; Hauptmann, J.; Vieweg, H.; Wikstrom, P. *J. Med. Chem.* **1997,** *40,* 3091–3099. doi:10.1021/jm960668h
256. Satya, P. G. *Mini-Rev. Med. Chem.* **2003,** *3,* 315–321. doi:10.2174/1389557033488051
257. Mendenhall, J.; Meiler, J. *J. Comput.-Aided Mol. Des.* **2016,** *30,* 177–189. doi:10.1007/s10822-016-9895-2
258. Butkiewicz, M.; Mueller, R.; Selic, D.; Dawson, E.; Meiler, J. *IEEE Trans. Electromagn. Compat.* **2009,** *51,* 255–262.
259. Bleckmann, A.; Meiler, J. *QSAR Comb. Sci.* **2003,** *22,* 722–728.
260. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964,** *86,* 1616–1626. doi:10.1021/ja01062a035
261. Adl, A.; Zein, M.; Hassanien, A. E. *Expert Syst. Appl.* **2016,** *54,* 219–227. doi:10.1016/j.eswa.2016.01.051
262. Karelson, M.; Sild, S.; Maran, U. *Mol. Simul.* **2000,** *24,* 229–242. doi:10.1080/08927020008022373
263. Jain, A. N.; Koile, K.; Chapman, D. *J. Med. Chem.* **1994,** *37,* 2315–2327. doi:10.1021/jm00041a010
264. Darnag, R.; Mostapha Mazouz, E. L.; Schmitzer, A.; Villemin, D.; Jarid, A.; Cherqaoui, D. *Eur. J. Med. Chem.* **2010,** *45,* 1590–1597. doi:10.1021/jm00041a010
265. Mei, H.; Zhou, Y.; Liang, G.; Li, Z. *Chin. Sci. Bull.* **2005,** *50,* 2291–2296. doi:10.1007/BF03183737
266. Ringner, M. *Nat. Biotechnol.* **2008,** *26,* 303–304. doi:10.1038/nbt0308-303
267. Gramatica, P. *QSAR Comb. Sci.* **2007,** *26,* 694–701. doi:10.1002/qsar.200610151
268. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K. *Int. J. Drug Des. Discovery* **2011,** *3,* 511–519.
269. Koga, H.; Itoh, A.; Murayama, S.; Suzue, S.; Irikura, T. *J. Med. Chem.* **1980,** *23,* 1358–1363. doi:10.1021/jm00186a014
270. Duncia, J. V.; Chiu, A. T.; Carini, D. J.; Gregory, G. B.; Johnson, A. L.; Price, W. A.; Wells, G. J.; Wong, P. C.; Calabrese, J. C.; Timmermans, P. B. M. W. M. *J. Med. Chem.* **1990,** *33,* 1312–1329. doi:10.1021/jm00167a007
271. Buckingham, J.; Glen, R. C.; Hill, A. P.; Hyde, R. M.; Martin, G. R.; Robertson, A. D.; Salmon, J. A.; Woollard, P. M. *J. Med. Chem.* **1995,** *38,* 3566–3580. doi:10.1021/jm00018a016
272. Tarca, A. L.; Carey, V. J.; Chen, X.-w.; Romero, R.; Drăghici, S. *PLoS Comput. Biol.* **2007,** *3,* e116. doi:10.1371/journal.pcbi.0030116
273. Sommer, C.; Gerlich, D. W. *J. Cell Sci.* **2013,** *126,* 5529–5539. doi:10.1242/jcs.123604
274. Lima, A. N.; Philot, E. A.; Trossini, G. H. G.; Scott, L. P. B.; Maltarollo, V. G.; Honorio, K. M. *Expert Opin. Drug Discovery* **2016,** *11,* 225–239. doi:10.1517/17460441.2016.1146250
275. Lavecchia, A. *Drug Discovery Today* **2015,** *20,* 318–331. doi:10.1016/j.drudis.2014.10.012
276. Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. *Comput. Chem. (Oxford, U. K.)* **2001,** *26,* 5–14. doi:10.1016/S0097-8485(01)00094-8
277. Ain, Q. U.; Aleksandrova, A.; Roessler, F. D.; Ballester, P. J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015,** *5,* 405–424. doi:10.1002/wcms.1225
278. Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. *J. Chem. Inf. Comput. Sci.* **2003,** *43,* 1882–1889. doi:10.1021/ci0341161
279. Puri, M.; Solanki, A.; Padawer, T.; Tipparaju, S. M.; Moreno, W. A.; Pathak, Y. *Artificial Neural Network for Drug Design, Delivery and Disposition;* Academic Press: Boston, 2016; pp 3–13.
280. Hu, L.; Chen, G.; Chau, R. M.-W. *J. Mol. Graphics Mod.* **2006,** *24,* 244–253. doi:10.1016/j.jmgm.2005.09.002
281. Gao, D.-W.; Wang, P.; Liang, H.; Peng, Y.-Z. *J. Environ. Sci. Health, Part B* **2003,** *38,* 571–579. doi:10.1081/PFC-120023515
282. Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J. *J. Med. Chem.* **2010,** *53,* 8461–8467. doi:10.1021/jm101020z
283. Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. *J. Chem. Inf. Model.* **2009,** *49,* 1455–1474. doi:10.1021/ci900056c
284. Damm-Ganamet, K. L.; Bembenek, S. D.; Venable, J. W.; Castro, G. G.; Mangelschots, L.; Peeters, D. C. G.; McAllister, H. M.; Edwards, J. P.; Disepio, D.; Mirzadegan, T. *J. Med. Chem.* **2016,** *59,* 4302–4313. doi:10.1021/acs.jmedchem.5b01974
285. Vilar, S.; Harpaz, R.; Uriarte, E.; Santana, L.; Rabadan, R.; Friedman, C. *J. Am. Med. Inf. Assoc.* **2012,** *19,* 1066–1074. doi:10.1136/amiajnl-2012-000935
286. Honig, P. K.; Wortham, D. C.; Zamani, K.; Conner, D. P.; Mullin, J. C.; Cantilena, L. R. *J. Am. Med. Inf. Assoc.* **1993,** *269,* 1513–1518.
287. Hudelson, M. G.; Ketkar, N. S.; Holder, L. B.; Carlson, T. J.; Peng, C.-C.; Waldher, B. J.; Jones, J. P. *J. Med. Chem.* **2008,** *51,* 648–654. doi:10.1021/jm701130z
288. Ekins, S.; Wrighton, S. A. *J. Pharmacol. Toxicol. Methods* **2001,** *45,* 65–69. doi:10.1016/S1056-8719(01)00119-8
289. Singh, N.; Chevé, G.; Ferguson, D. M.; McCurdy, C. R. *J. Comput.-Aided Mol. Des.* **2006,** *20,* 471–493. doi:10.1007/s10822-006-9067-x
290. Prathipati, P.; Mizuguchi, K. *J. Chem. Inf. Model.* **2016,** *56,* 974–987. doi:10.1021/acs.jcim.5b00477
291. Huang, S.-Y.; Li, M.; Wang, J.; Pan, Y. *J. Chem. Inf. Model.* **2016,** *56,* 1078–1087. doi:10.1021/acs.jcim.5b00275
292. Bauer, J. F. *J. Validation Technol.* **2008,** *14,* 15.
293. Fry, D. C. *Pept. Sci.* **2006,** *84,* 535–552. doi:10.1002/bip.20608

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

A non-canonical peptide synthetase adenylates 3-methyl-2-oxovaleric acid for auriculamide biosynthesis

Daniel Braga[1,2], Dirk Hoffmeister[1] and Markus Nett[*3]

Letter

Open Access

Address:
[1]Friedrich-Schiller-Universität Jena, Department Pharmaceutical Microbiology at the Hans-Knöll-Institute, Winzerlaer Strasse 2, 07745 Jena, Germany, [2]Friedrich-Schiller-Universität Jena, Junior Research Group Synthetic Microbiology at the Hans-Knöll-Institute, Adolf-Reichwein-Strasse 23, 07745 Jena, Germany and [3]Department of Biochemical and Chemical Engineering, Technical Biology, Technical University Dortmund, Emil-Figge-Strasse 66, 44227 Dortmund, Germany

Email:
Markus Nett[*] - markus.nett@bci.tu-dortmund.de

* Corresponding author

Keywords:
adenylation; auriculamide; biosynthesis; *Herpetosiphon*; nonribosomal peptide synthetase

Beilstein J. Org. Chem. **2016,** *12,* 2766–2770.
doi:10.3762/bjoc.12.274

Received: 02 September 2016
Accepted: 07 December 2016
Published: 16 December 2016

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

Guest Editor: H. B. Bode

© 2016 Braga et al.; licensee Beilstein-Institut.
License and terms: see end of document.

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* 114-95^T (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* 114-95^T, 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 A_1-A_2-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 pyridomycin [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-

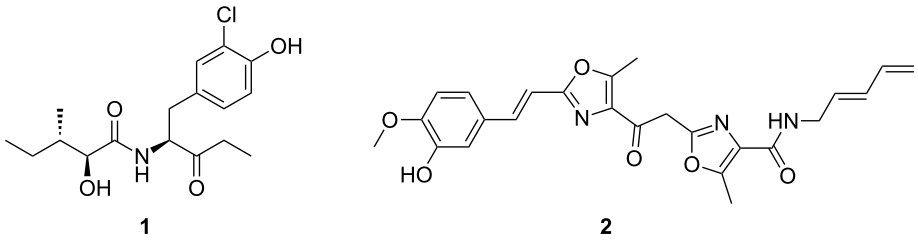


Figure 1: *Herpetosiphon* natural products auriculamide (**1**) and siphonazole (**2**).

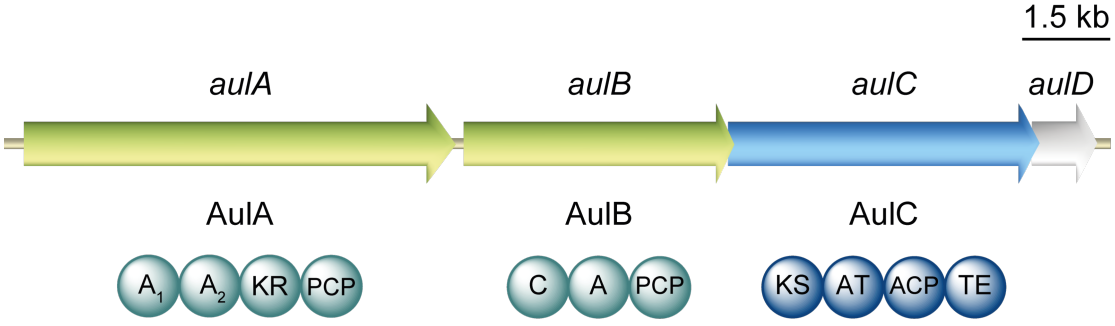


Figure 2: Organisation of the *aul* biosynthetic gene cluster. Circles illustrate the domain architecture of the NRPSs and the PKS present therein. Domains are abbreviated as A, adenylation; ACP, acyl carrier protein; AT, acyl transferase; C, condensation; KR, ketoreductase; KS, ketosynthase; PCP, peptidyl carrier protein; TE, thioesterase. The gene *aulD* encodes a type II thioesterase.

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-A_1. A subsequent manual inspection further revealed that the acyl-activating consensus motif is hardly conserved in AulA-A_1. Moreover, the strictly invariant residue Asp413, which is essential for adenylate binding [14] was replaced by a tyrosine residue in AulA-A_1. We hence concluded that this domain cannot function as an adenylating enzyme and is likely skipped during the biosynthetic assembly. Inspection of AulA-A_2 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 carboxy group [14,20]. Also, the relationship between the expected substrate and the nonribosomal code of AulA-A_2 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-A_2 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-A_1 would contribute to the biosynthesis of auriculamide, e.g., through structural support for the catalytic role of AulA-A_2, 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-A_2 domain, the α-keto reductase domain, and the terminal carrier protein. Independently, *E. coli* KRX was transformed with both constructs for heterologous production of the respective *N*-terminally hexahistidine fusion proteins, which were purified by metal affinity chromatography (Supporting Information File 1, Figure S1).

To probe their enzymatic activity, the two purified AulA fusion proteins were subjected to the ATP-[^{32}P]pyrophosphate exchange assay. In this assay, the protein is incubated with a potential substrate, ATP and radioactive pyrophosphate. The reversible back exchange of [^{32}P]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 A_1 domain is not essential for adenylation of **3** and PCP loading. Further functional characterization was carried out using the native four-domain enzyme A_1-A_2-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-A_2 and the comparison with other α-keto acid activating NRPSs.

Enzyme	GenBank Accession #	Nonribosomal code										Substrate
PksJ	P40806	V	G	M	W	N	G	A	S	V	K	4-methyl-2-oxovaleric acid
CesA-A_1	ABD14711	V	G	M	W	N	G	T	S	I	K	4-methyl-2-oxovaleric acid
CesB-A_1	ABD14712	V	G	M	W	N	G	V	S	V	K	2-oxovaleric acid
PyrG-A_2	AEF33080	V	G	M	T	I	G	A	S	G	K	3-methyl-2-oxovaleric acid
AulA-A_2	ABX05055	G	I	F	W	L	G	A	S	G	K	3-methyl-2-oxovaleric acid

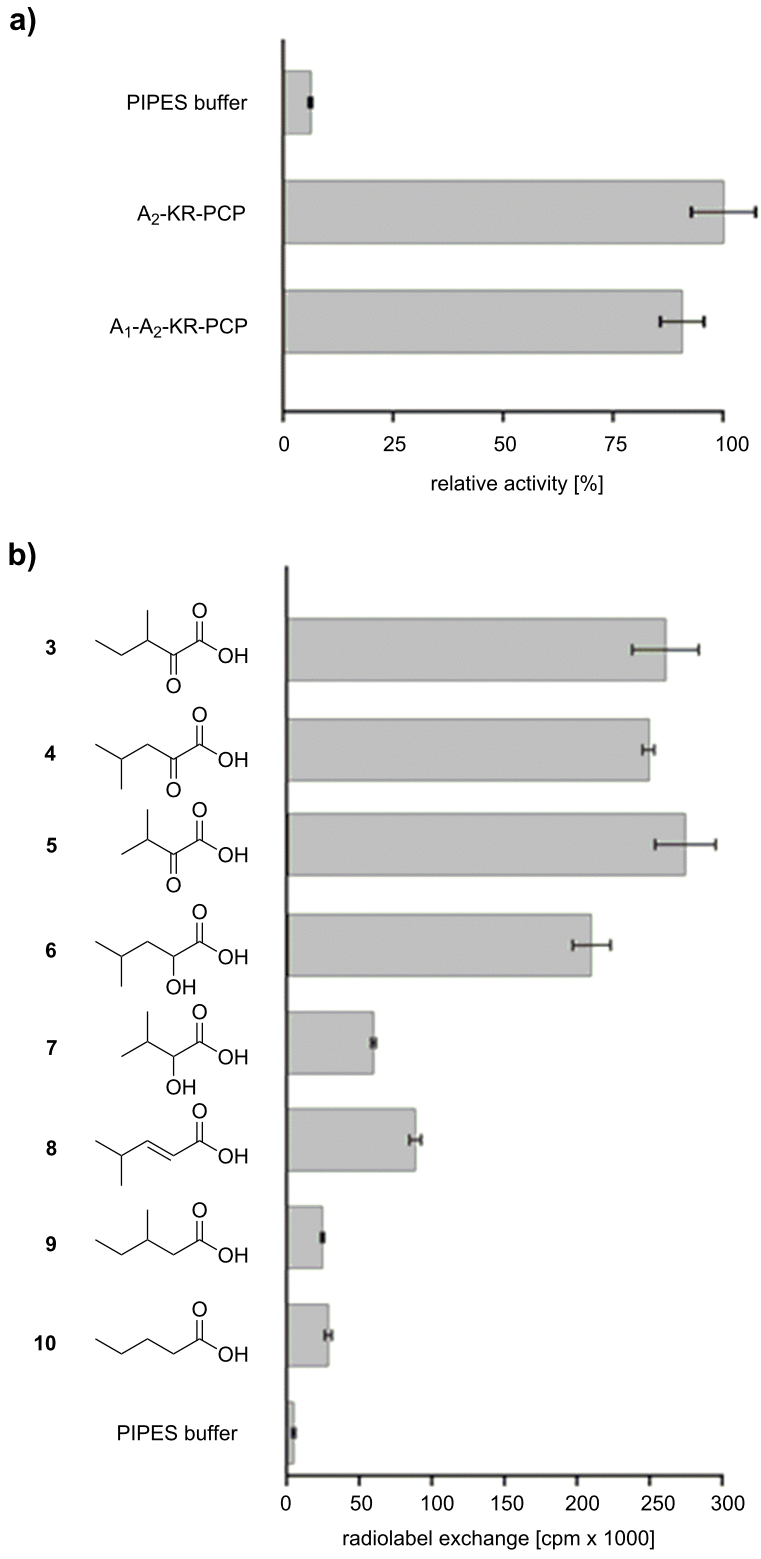


Figure 3: Testing of AulA (A_1-A_2-KR-PCP) and a truncated variant (A_2-KR-PCP) in the ATP-[^{32}P]pyrophosphate exchange assay using 3-methyl-2-oxovaleric acid as a substrate. Relative activities are referenced to the A_2-KR-PCP enzyme (a). Substrate specificity of recombinant AulA in the ATP-[^{32}P]pyrophosphate 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

1. Holt, J.; Lewin, R. *J. Bacteriol.* **1968,** *95,* 2407–2408.
2. Jurkevitch, E. *Microbe* **2007,** *2,* 67–73. doi:10.1128/microbe.2.67.1
3. Kiss, H.; Nett, M.; Domin, N.; Martin, K.; Maresca, J. A.; Copeland, A.; Lapidus, A.; Lucas, S.; Berry, K. W.; Glavina Del Rio, T.; Dalin, E.; Tice, H.; Pitluck, S.; Richardson, P.; Bruce, D.; Goodwin, L.; Han, C.; Detter, J. C.; Schmutz, J.; Brettin, T.; Land, M.; Hauser, L.; Kyrpides, N. C.; Ivanova, N.; Göker, M.; Woyke, T.; Klenk, H. P.; Bryant, D. A. *Stand. Genomic Sci.* **2011,** *5,* 356–370. doi:10.4056/sigs.2194987
4. Schieferdecker, S.; Domin, N.; Hoffmeier, C.; Bryant, D. A.; Roth, M.; Nett, M. *Eur. J. Org. Chem.* **2015,** 3057–3062. doi:10.1002/ejoc.201500181
5. Nakano, C.; Oshima, M.; Kurashima, N.; Hoshino, T. *ChemBioChem* **2015,** *16,* 772–781. doi:10.1002/cbic.201402652
6. Korp, J.; Vela Gurovic, M. S.; Nett, M. *Beilstein J. Org. Chem.* **2016,** *12,* 594–607. doi:10.3762/bjoc.12.58
7. Nett, M.; Erol, Ö.; Kehraus, S.; Köck, M.; Krick, A.; Eguereva, E.; Neu, E.; König, G. M. *Angew. Chem., Int. Ed.* **2006,** *45,* 3863–3867. doi:10.1002/anie.200504525
8. Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 14083–14088. doi:10.1073/pnas.212382199
9. Pfeifer, B.; Hu, Z.; Licari, P.; Khosla, C. *Appl. Environ. Microbiol.* **2002,** *68,* 3287–3292. doi:10.1128/AEM.68.7.3287-3292.2002
10. Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z.; Yamashita, R. A.; Zhang, D.; Zheng, C.; Bryant, S. H. *Nucleic Acids Res.* **2014,** *43* (Suppl. D1), D222–D226. doi:10.1093/nar/gku1221
11. Huang, T.; Li, L.; Brock, N. L.; Deng, Z.; Lin, S. *ChemBioChem* **2016,** *17,* 1421–1425. doi:10.1002/cbic.201600156
12. Magarvey, N. A.; Ehling-Schulz, M.; Walsh, C. T. *J. Am. Chem. Soc.* **2006,** *128,* 10698–10699. doi:10.1021/ja0640187
13. Calderone, C. T.; Bumpus, S. B.; Kelleher, N. L.; Walsh, C. T.; Magarvey, N. A. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 12809–12814. doi:10.1073/pnas.0806305105
14. Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997,** *16,* 4174–4183. doi:10.1093/emboj/16.14.4174
15. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Cell Chem. Biol.* **1999,** *6,* 493–505. doi:10.1016/S1074-5521(99)80082-9
16. Challis, G. L.; Ravel, J.; Townsend, C. A. *Cell Chem. Biol.* **2000,** *7,* 211–224. doi:10.1016/S1074-5521(00)00091-0
17. Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009,** *458,* 181–217. doi:10.1016/S0076-6879(09)04808-3
18. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011,** *39* (Suppl. 2), W362–W367. doi:10.1093/nar/gkr323
19. Kalb, D.; Lackner, G.; Hoffmeister, D. *Fungal Biol. Rev.* **2013,** *27,* 43–50. doi:10.1016/j.fbr.2013.05.002
20. Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003,** *20,* 275–287. doi:10.1039/b111145k
21. Ames, B. D.; Walsh, C. T. *Biochemistry* **2010,** *49,* 3351–3365. doi:10.1021/bi100198y
22. Gao, X.; Chooi, Y.-H.; Ames, B. D.; Wang, P.; Walsh, C. T.; Tang, Y. *J. Am. Chem. Soc.* **2011,** *133,* 2729–2741. doi:10.1021/ja1101085
23. Wackler, B.; Lackner, G.; Chooi, Y. H.; Hoffmeister, D. *ChemBioChem* **2012,** *13,* 1798–1804. doi:10.1002/cbic.201200187
24. Kalb, D.; Lackner, G.; Rappe, M.; Hoffmeister, D. *ChemBioChem* **2015,** *16,* 1426–1430. doi:10.1002/cbic.201500190
25. Linne, U.; Marahiel, M. A. *Methods Enzymol.* **2004,** *388,* 293–315. doi:10.1016/S0076-6879(04)88024-8

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

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

Open Access

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

Beilstein J. Org. Chem. **2016,** *12,* 2784–2792.
doi:10.3762/bjoc.12.277

Received: 16 September 2016
Accepted: 08 December 2016
Published: 20 December 2016

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

Guest Editor: H. B. Bode

© 2016 Prothiwa et al.; licensee Beilstein-Institut.
License and terms: see end of document.

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 β-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-aminobenzoylacetyl-CoA. The thioesterase PqsE hydrolyses the thioester to produce 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-

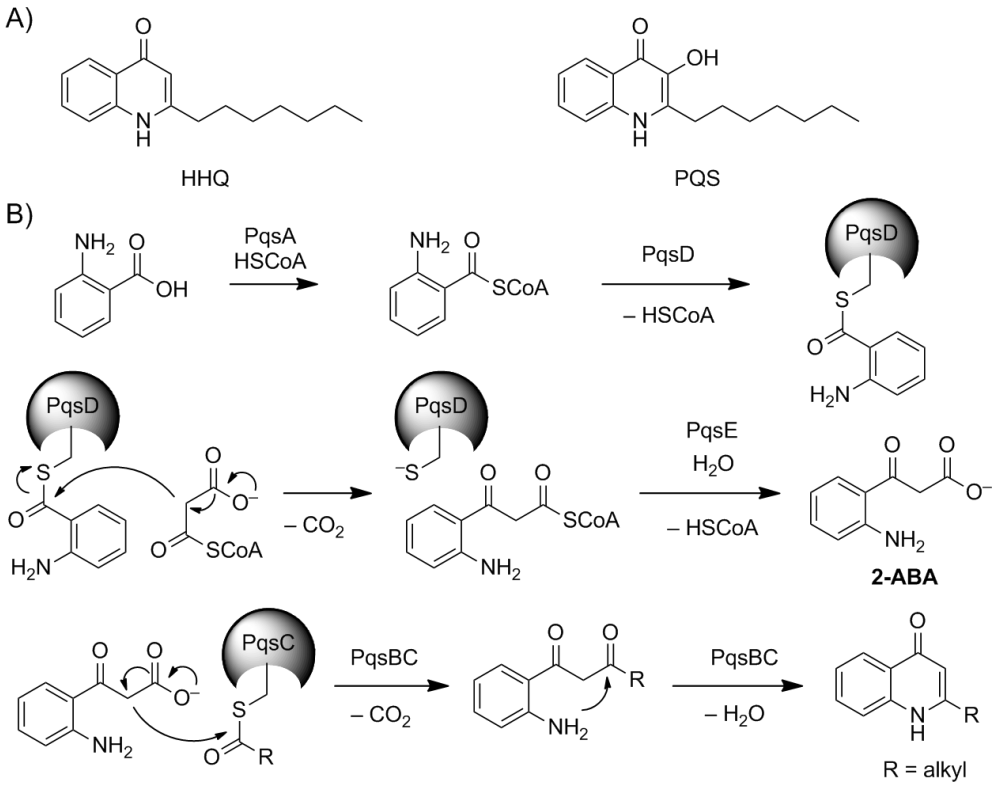


Figure 1: Quinolone signals of *Pseudomonas aeruginosa*. A) Structures of HHQ and PQS. B) Proposed mechanism for the synthesis of 2-alkyl-4-quinolones [12-14].

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 α-chloroacetamide, α,β-unsaturated amide, and α,β-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 α,β-unsaturated amide probes **UA1–3** by reaction with acrylic acid chloride (Figure 2A) and α-chloroacetamide probes **CA1–3** by reaction with chloroacetyl chloride (Figure 2B). The α,β-unsaturated ketone probe **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 *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

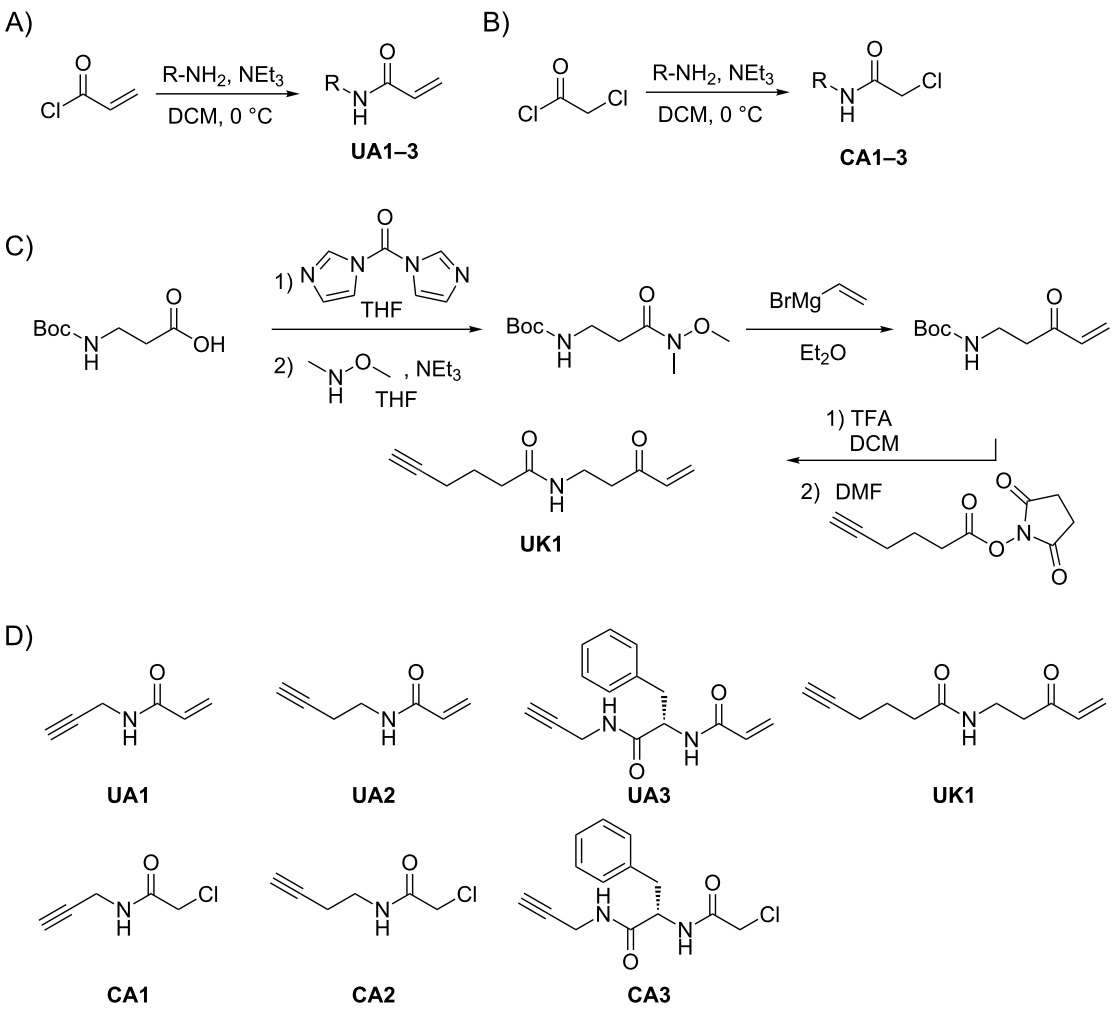


Figure 2: Synthesis of electrophilic ABPP probes. A) Synthesis of α,β-unsaturated amide probes **UA1–3**. B) Synthesis of α-chloroacetamide probes **CA1–3**, and C) synthesis of α,β-unsaturated ketone **UK1**. D) Structures of the ABPP probe library.

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 labeling, 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.

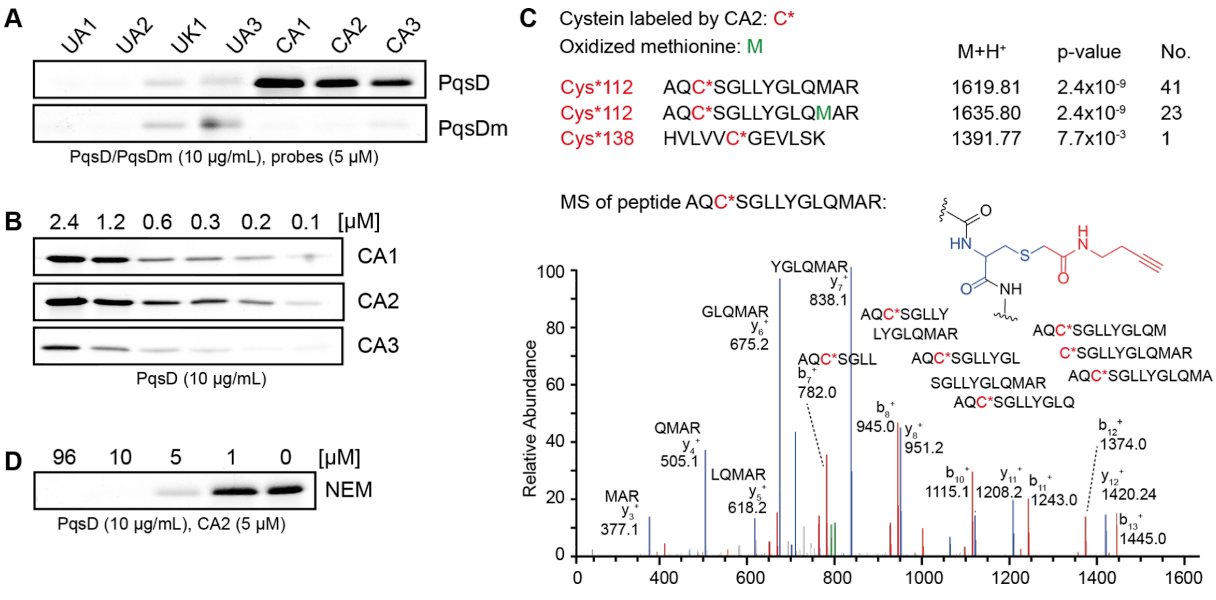


Figure 3: In vitro labeling of PqsD by chemical probes. A) ABPP probe library with wild-type PqsD and PqsD C112A mutant (PqsDm). B) Concentration dependence of labeling by the three active site directed probes. C) Mass spectrometric discovery of tryptic peptide fragments with probe **CA2** attached to the active site Cys112 (No.: number of detected peptides). D) Competitive experiment with *N*-ethylmaleimide and probe **CA2**.

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

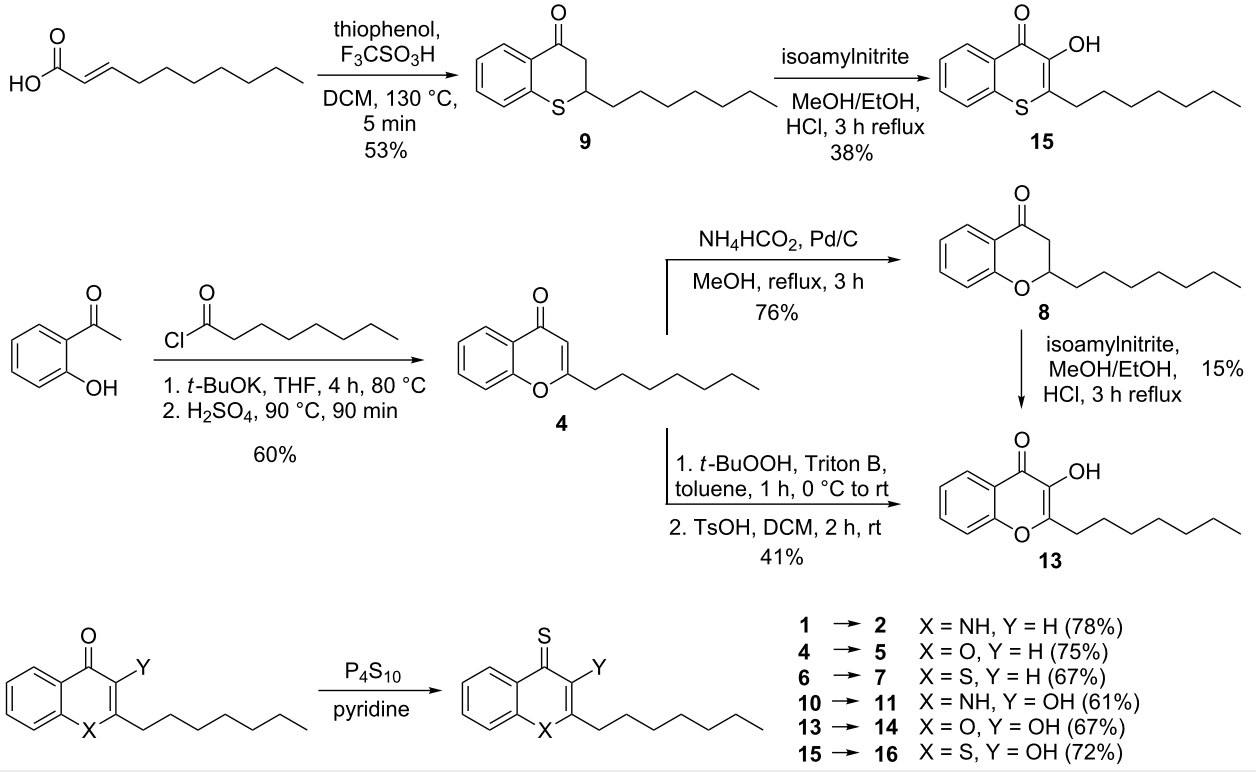
In order to test if our probes could be used as a competitive labelling platform, we used the well-known unspecific cysteine reactive agent *N*-ethylmaleimide (NEM) and dosed NEM in increasing concentrations to wild-type PqsD before applying the probe **CA2** followed by click chemistry with the fluorophore. 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 AQs 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(1*H*)-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 affort 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 (NH_4HCO_2) 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 H_2 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 thiochroman-4-one **9** and following oxidation to give 1-S-PQS (**15**) in 12% overall yield [33]. In our attempt to synthesize **9** more efficiently, we used a method described by Olah et al. [37] starting from commercially available (*E*)-dec-2-enoic 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 P_4S_{10} 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-oxime (**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-oxime **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



Scheme 1: Synthesis of various HHQ and PQS analogues.

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

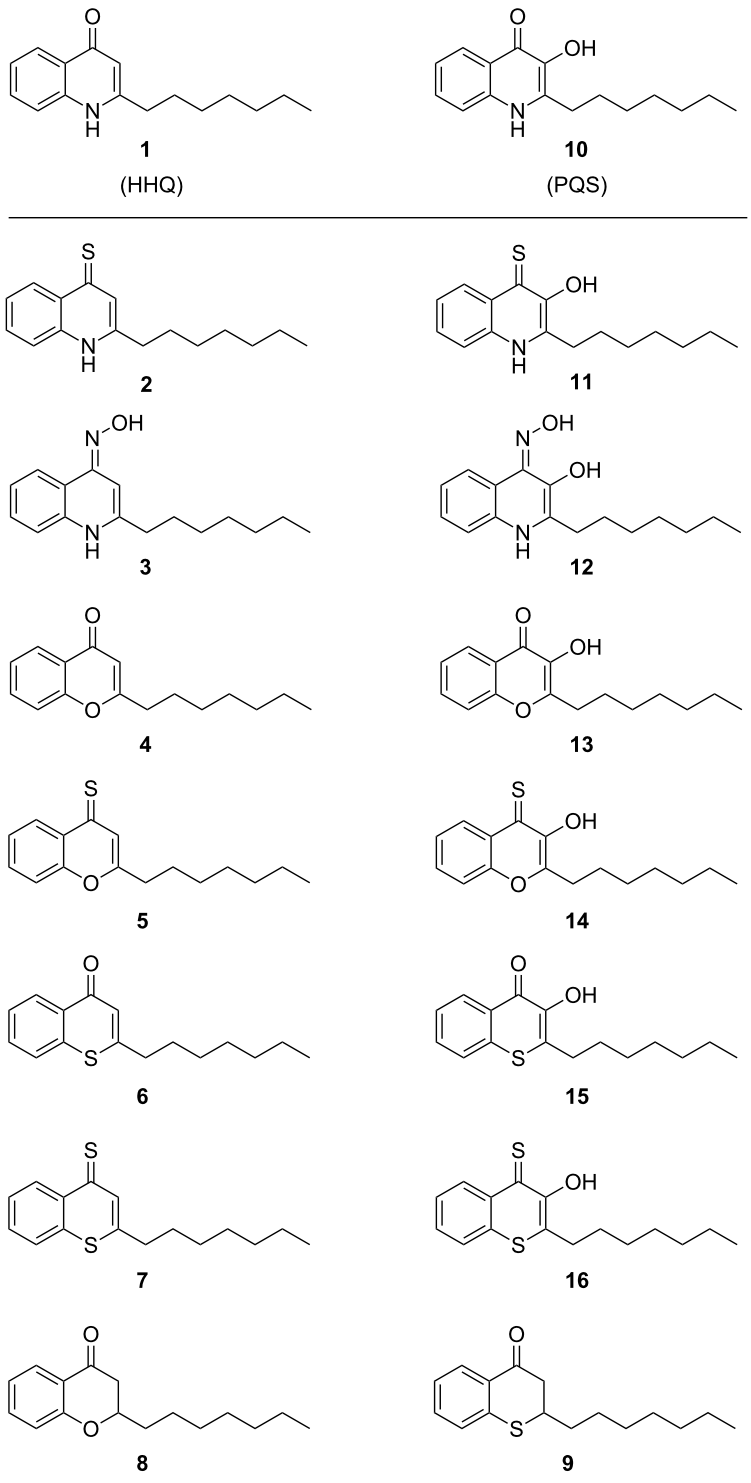


Figure 4: Library of HHQ and PQS analogues.

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.

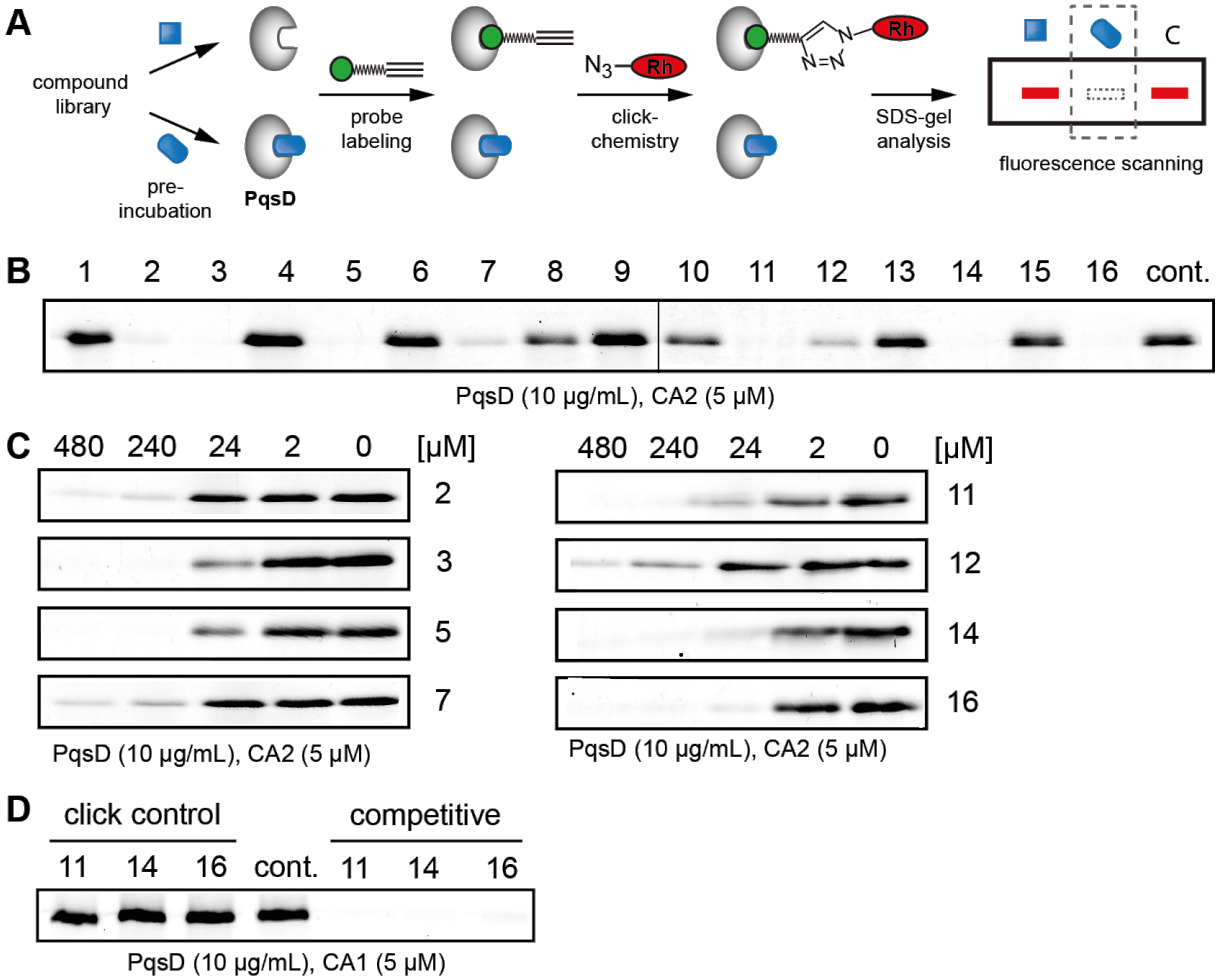


Figure 5: Competitive profiling platform. A) Schematic representation of the competitive labelling strategy with an alkyne probe (green) and potential inhibitors (blue). Rh = rhodamine. B) Initial screening of our small library of HHQ (**1**–**9**) and PQS (**10**–**16**) analogues against the active 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 Cu^{II}-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 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

1. Ventola, C. L. *Pharm. Ther.* **2015,** *40,* 277–283.
2. Davies, J. C. *Paediatr. Respir. Rev.* **2002,** *3,* 128–134. doi:10.1016/S1526-0550(02)00003-3
3. Obritsch, M. D.; Fish, D. N.; MacLaren, R.; Jung, R. *Pharmacotherapy* **2005,** *25,* 1353–1364. doi:10.1592/phco.2005.25.10.1353
4. Strateva, T.; Mitov, I. *Ann. Microbiol. (Heidelberg, Ger.)* **2011,** *61,* 717–732. doi:10.1007/s13213-011-0273-y

5. Bodey, G. P.; Bolivar, R.; Fainstein, V.; Jadeja, L. *Rev. Infect. Dis.* **1983,** *5,* 279–313. doi:10.1093/clinids/5.2.279
6. Lee, J.; Zhang, L. *Protein Cell* **2015,** *6,* 26–41. doi:10.1007/s13238-014-0100-x
7. Pesci, E. C.; Milbank, J. B. J.; Pearson, J. P.; McKnight, S.; Kende, A. S.; Greenberg, E. P.; Iglewski, B. H. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96,* 11229–11234. doi:10.1073/pnas.96.20.11229
8. Dubern, J.-F.; Diggle, S. P. *Mol. BioSyst.* **2008,** *4,* 882–888. doi:10.1039/b803796p
9. Lepine, F.; Milot, S.; Déziel, E.; He, J.; Rahme, L. G. *J. Am. Soc. Mass Spectrom.* **2004,** *15,* 862–869. doi:10.1016/j.jasms.2004.02.012
10. Bredenbruch, F.; Nimtz, M.; Wray, V.; Morr, M.; Müller, R.; Häussler, S. *J. Bacteriol.* **2005,** *187,* 3630–3635. doi:10.1128/JB.187.11.3630-3635.2005
11. Pistorius, D.; Ullrich, A.; Lucas, S.; Hartmann, R. W.; Kazmaier, U.; Müller, R. *ChemBioChem* **2011,** *12,* 850–853. doi:10.1002/cbic.201100014
12. Dulcey, C. E.; Dekimpe, V.; Fauvelle, D.-A.; Milot, S.; Groleau, M.-C.; Doucet, N.; Rahme, L. G.; Lépine, F.; Déziel, E. *Chem. Biol.* **2013,** *20,* 1481–1491. doi:10.1016/j.chembiol.2013.09.021
13. Drees, S. L.; Fetzner, S. *Chem. Biol.* **2015,** *22,* 611–618. doi:10.1016/j.chembiol.2015.04.012
14. Drees, S. L.; Li, C.; Prasetya, F.; Saleem, M.; Dreveny, I.; Williams, P.; Hennecke, U.; Emsley, J.; Fetzner, S. *J. Biol. Chem.* **2016,** *291,* 6610–6624. doi:10.1074/jbc.M115.708453
15. Evans, M. J.; Cravatt, B. F. *Chem. Rev.* **2006,** *106,* 3279–3301. doi:10.1021/cr050288g
16. Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. *Annu. Rev. Biochem.* **2008,** *77,* 383–414. doi:10.1146/annurev.biochem.75.101304.124125
17. Blum, G.; von Degenfeld, G.; Merchant, M. J.; Blau, H. M.; Bogyo, M. *Nat. Chem. Biol.* **2007,** *3,* 668–677. doi:10.1038/nchembio.2007.26
18. Barglow, K. T.; Cravatt, B. F. *Nat. Methods* **2007,** *4,* 822–827. doi:10.1038/nmeth1092
19. Weerapana, E.; Simon, G. M.; Cravatt, B. F. *Nat. Chem. Biol.* **2008,** *4,* 405–407. doi:10.1038/nchembio.91
20. Böttcher, T.; Sieber, S. A. *MedChemComm* **2012,** *3,* 408–417. doi:10.1039/c2md00275b
21. Kato, D.; Boatright, K. M.; Berger, A. B.; Nazif, T.; Blum, G.; Ryan, C.; Chehade, K. A. H.; Salvesen, G. S.; Bogyo, M. *Nat. Chem. Biol.* **2005,** *1,* 33–38. doi:10.1038/nchembio707
22. Böttcher, T.; Pitscheider, M.; Sieber, S. A. *Angew. Chem.* **2010,** *49,* 2680–2698. doi:10.1002/anie.200905352
23. Hutter, M. C.; Brengel, C.; Negri, M.; Henn, C.; Zimmer, C.; Hartmann, R. W.; Empting, M.; Steinbach, A. *J. Mol. Model.* **2014,** *20,* 2255. doi:10.1007/s00894-014-2255-z
24. Sahner, J. H.; Brengel, C.; Storz, M. P.; Groh, M.; Plaza, A.; Müller, R.; Hartmann, R. W. *J. Med. Chem.* **2013,** *56,* 8656–8664. doi:10.1021/jm401102e
25. Weidel, E.; de Jong, J. C.; Brengel, C.; Storz, M. P.; Braunshausen, A.; Negri, M.; Plaza, A.; Steinbach, A.; Müller, R.; Hartmann, R. W. *J. Med. Chem.* **2013,** *56,* 6146–6155. doi:10.1021/jm4006302
26. Hinsberger, S.; de Jong, J. C.; Groh, M.; Haupenthal, J.; Hartmann, R. W. *Eur. J. Med. Chem.* **2014,** *76,* 343–351. doi:10.1016/j.ejmech.2014.02.014
27. Storz, M. P.; Brengel, C.; Weidel, E.; Hoffmann, M.; Hollemeyer, K.; Steinbach, A.; Müller, R.; Empting, M.; Hartmann, R. W. *ACS Chem. Biol.* **2013,** *8,* 2794–2801. doi:10.1021/cb400530d
28. Storz, M. P.; Maurer, C. K.; Zimmer, C.; Wagner, N.; Brengel, C.; de Jong, J. C.; Lucas, S.; Müsken, M.; Häussler, S.; Steinbach, A.; Hartmann, R. W. *J. Am. Chem. Soc.* **2012,** *134,* 16143–16146. doi:10.1021/ja3072397
29. Storz, M. P.; Allegretta, G.; Kirsch, B.; Empting, M.; Hartmann, R. W. *Org. Biomol. Chem.* **2014,** *12,* 6094–6104. doi:10.1039/C4OB00707G
30. Sahner, J. H.; Empting, M.; Kamal, A.; Weidel, E.; Groh, M.; Börger, C.; Hartmann, R. W. *Eur. J. Med. Chem.* **2015,** *96,* 14–21. doi:10.1016/j.ejmech.2015.04.007
31. Allegretta, G.; Weidel, E.; Empting, M.; Hartmann, R. W. *Eur. J. Med. Chem.* **2015,** *90,* 351–359. doi:10.1016/j.ejmech.2014.11.055
32. Thomann, A.; de Mello Martins, A. G. G.; Brengel, C.; Empting, M.; Hartmann, R. W. *ACS Chem. Biol.* **2016,** *11,* 1279–1286. doi:10.1021/acschembio.6b00117
33. Szamosvári, D.; Reichle, V. F.; Jureschi, M.; Böttcher, T. *Chem. Commun.* **2016,** *52,* 13440–13443. doi:10.1039/C6CC06295D
34. Reen, F. J.; Clarke, S. L.; Legendre, C.; McSweeney, C. M.; Eccles, K. S.; Lawrence, S. E.; O'Gara, F.; McGlacken, G. P. *Org. Biomol. Chem.* **2012,** *10,* 8903–8910. doi:10.1039/c2ob26823j
35. Hradil, P.; Hlaváč, J.; Lemr, K. *J. Heterocycl. Chem.* **1999,** *36,* 141–144. doi:10.1002/jhet.5570360121
36. Abdel Ghani, S. B.; Mugisha, P. J.; Wilcox, J. C.; Gado, E. A. M.; Medu, E. O.; Lamb, A. J.; Brown, R. C. D. *Synth. Commun.* **2013,** *43,* 1549–1556. doi:10.1080/00397911.2011.647222
37. Vaghoo, H.; Prakash, G. K. S.; Narayanan, A.; Choudhary, R.; Paknia, F.; Mathew, T.; Olah, G. A. *Org. Lett.* **2015,** *17,* 6170–6173. doi:10.1021/acs.orglett.5b03172

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.12.277

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Biochemical and structural characterisation of the second oxidative crosslinking step during the biosynthesis of the glycopeptide antibiotic A47934

Veronika Ulrich[1], Clara Brieke[1] and Max J. Cryle[*1,2,3]

Full Research Paper

Open Access

Address:
[1]Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany, [2]EMBL Australia, Monash University, Clayton, Victoria 3800, Australia and [3]The 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

Beilstein J. Org. Chem. **2016,** *12,* 2849–2864.
doi:10.3762/bjoc.12.284

Received: 14 September 2016
Accepted: 09 December 2016
Published: 27 December 2016

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

Guest Editor: H. B. Bode

© 2016 Ulrich et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

The chemical complexity and biological activity of the glycopeptide antibiotics (GPAs) stems from their unique crosslinked structure, which is generated by the actions of cytochrome P450 (Oxy) enzymes that affect the crosslinking of aromatic side chains of amino acid residues contained within the GPA heptapeptide precursor. Given the crucial role peptide cyclisation plays in GPA activity, the characterisation of this process is of great importance in understanding the biosynthesis of these important antibiotics. Here, we report the cyclisation activity and crystal structure of StaF, the D-*O*-E ring forming Oxy enzyme from A47934 biosynthesis. Our results show that the specificity of StaF is reduced when compared to Oxy enzymes catalysing C-*O*-D ring formation and that this activity relies on interactions with the non-ribosomal peptide synthetase via the X-domain. Despite the interaction of StaF with the A47934 X-domain being weaker than for the preceding Oxy enzyme StaH, StaF retains higher levels of in vitro activity: we postulate that this is due to the ability of the StaF/X-domain complex to allow substrate reorganisation after initial complex formation has occurred. These results highlight the importance of testing different peptide/protein carrier constructs for in vitro GPA cyclisation assays and show that different Oxy homologues can display significantly different catalytic propensities despite their overall similarities.

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 monooxygenases [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 using 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

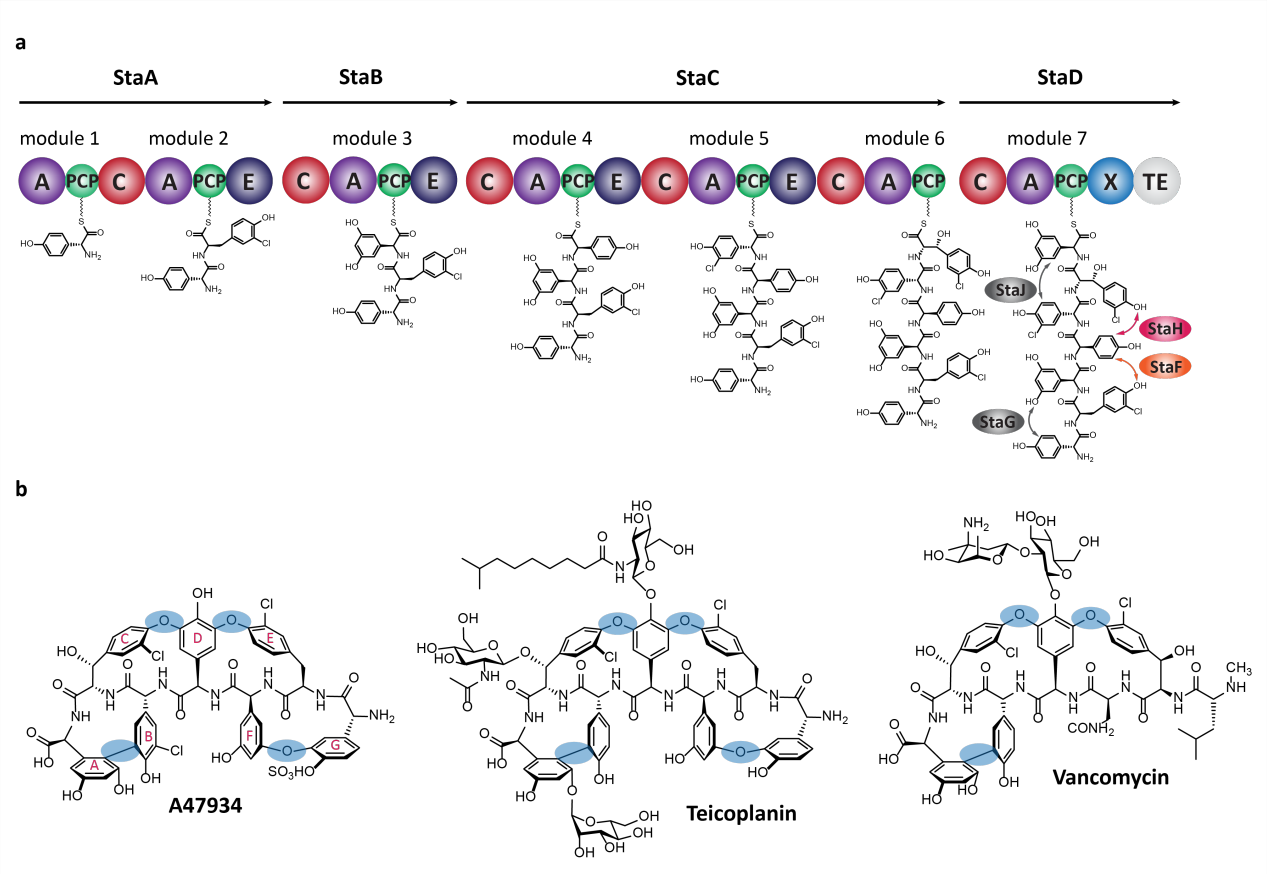


Figure 1: (a) Schematic representation of the biosynthesis of A47934 by the heterotetrameric non-ribosomal peptide synthetase; the 7 modules of the A47934 NRPS machinery are distributed over 4 proteins (StaA to StaD) and exhibit the typical NRPS domain architecture with adenylation (A; purple), peptidyl 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 $OxyB_{van}$, 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.

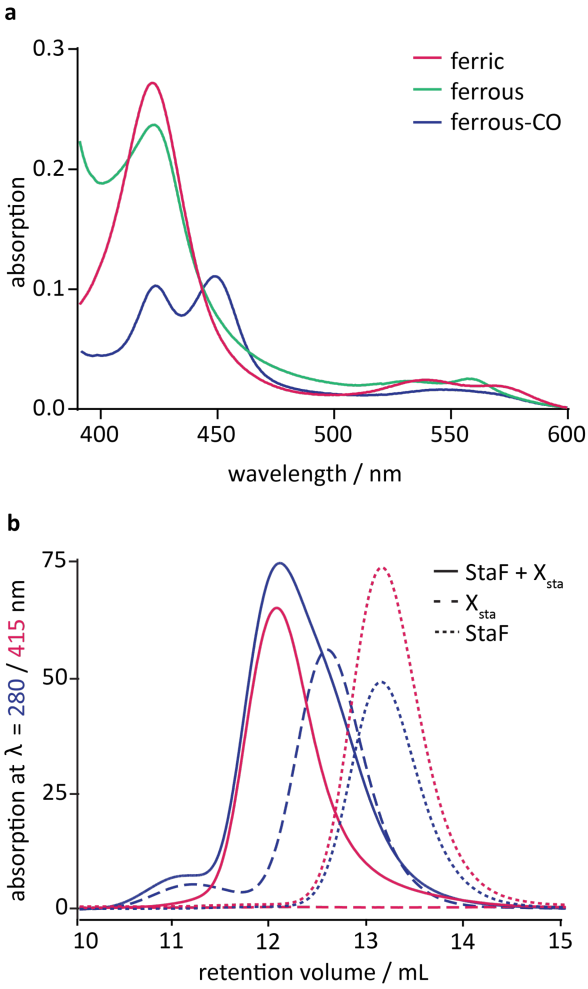


Figure 2: (a) Spectral analysis of StaF, showing the absorption spectra of ferric protein (red), ferrous protein (green) that has been reduced using $Na_2S_2O_4$, and ferrous protein saturated with CO (ferrous-CO; blue) was measured from λ = 390 to 600 nm; (b) interaction analysis of StaF with the A47934 X-domain; analysis of StaF with a 3-fold excess of X_{sta} was investigated by analytical size-exclusion chromatography measuring absorption at λ = 280 nm (blue) and 415 nm (red; heme-specific); analysis of the individual proteins served as control.

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 λ = 421 nm and β/α bands at λ = 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 λ = 422 nm and of the β/α bands to λ = 532 and 559 nm in the UV–visible spectrum. Upon saturation of ferrous StaF with carbon monoxide, two major peaks appeared at λ = 420 and 450 nm, respectively, as well as a broad minor peak at λ = 548 nm (Figure 2a). The peaks at λ = 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 λ = 420 nm) are catalytically inactive, whilst a catalytically competent P450 enzyme with a thiolate-ligated heme exhibits the signature λ = 450 nm absorption peak [34]. The fact that peaks at both λ = 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 λ = 280 nm, but also the heme-specific absorption at λ = 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 λ = 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 λ = 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 λ = 280 and 415 nm at an earlier elution volume of 12.1 mL (Figure 2b), which indicates that the heme-specific λ = 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_7; 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_7 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 methylamine 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_7 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

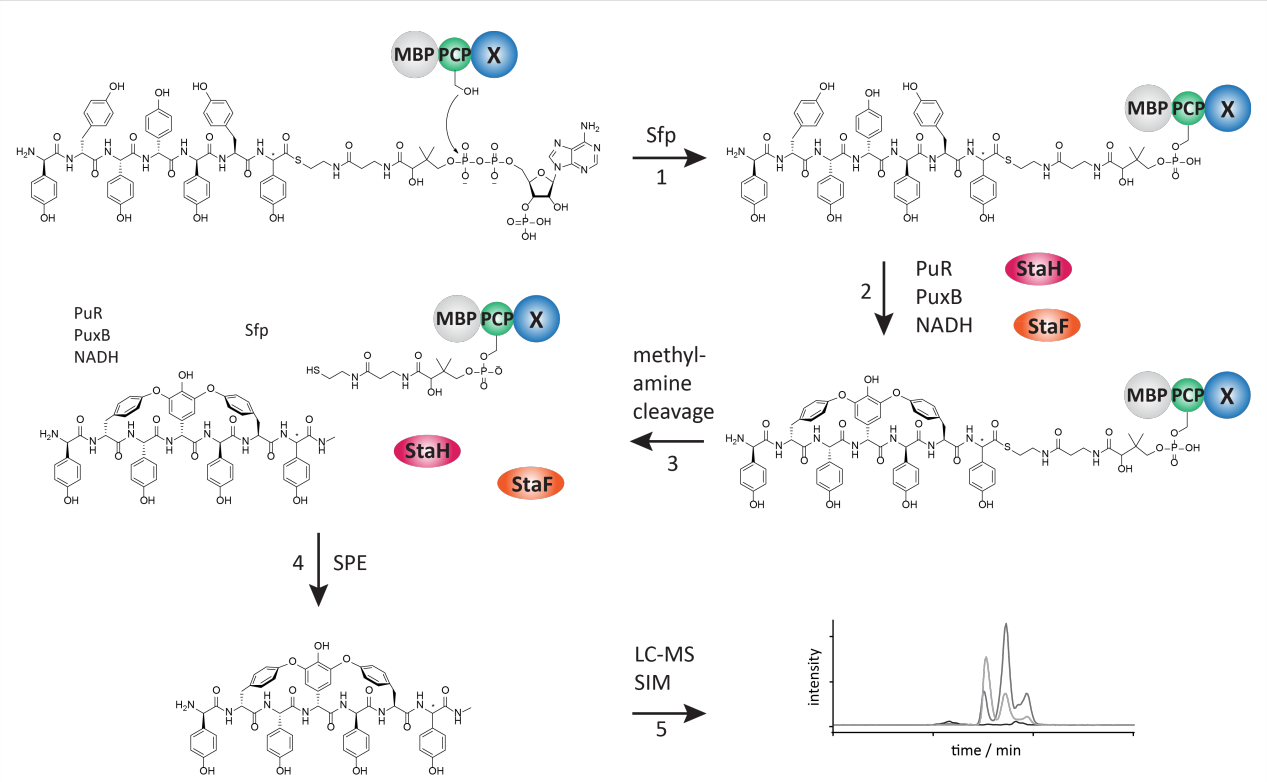


Figure 3: Complete workflow for the Cytochrome P450 activity assay used in this study. 1) Loading of the substrate (Tei7-L/D-Hpg$_7$ is depicted) onto a conserved serine residue of the PCP-domain using the R4-4 mutant of the promiscuous phosphopantetheinyl transferase Sfp. The substrate peptide is attached 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.

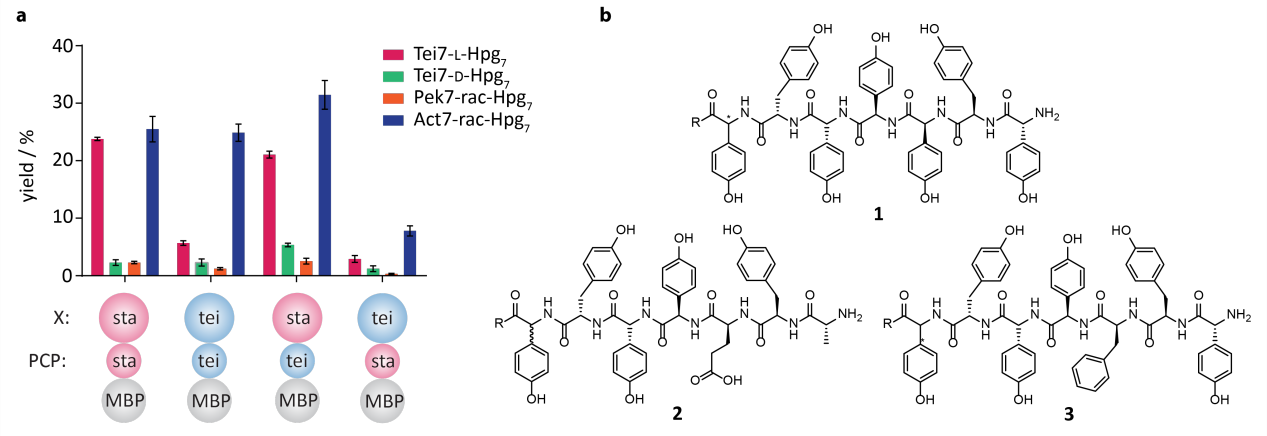


Figure 4: (a) StaF activity against different peptide substrates and using NRPS constructs; the activity of StaF with Tei7-L-Hpg$_7$ (magenta), Tei7-D-Hpg$_7$ (green), Pek7-rac-Hpg$_7$ (orange) and Act7-rac-Hpg$_7$ (blue) were determined; all peptides were bound to wildtype and hybrid PCP-X constructs 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.

Entry	fus-PCP-X[a]	peptide	StaF activity[b]
1	MBP sta sta	Tei7-L-Hpg_7	23.8% ± 0.3%
2	MBP sta sta	Tei7-D-Hpg_7	2.3% ± 0.5%
3	MBP sta sta	Pek7-rac-Hpg_7	2.3% ± 0.2%
4	MBP sta sta	Act7-rac-Hpg_7	25.5% ± 2.2%
5	MBP tei tei	Tei7-L-Hpg_7	5.7% ± 0.4%
6	MBP tei tei	Tei7-D-Hpg_7	2.3% ± 0.6%
7	MBP tei tei	Pek7-rac-Hpg_7	1.2% ± 0.2%
8	MBP tei tei	Act7-rac-Hpg_7	24.9% ± 1.5%
9	MBP sta tei	Tei7-L-Hpg_7	2.9% ± 0.5%
10	MBP sta tei	Tei7-D-Hpg_7	1.2% ± 0.5%
11	MBP sta tei	Pek7-rac-Hpg_7	0.3% ± 0.1%
12	MBP sta tei	Act7-rac-Hpg_7	7.8% ± 0.9%
13	MBP tei sta	Tei7-L-Hpg_7	21.1% ± 0.6%
14	MBP tei sta	Tei7-D-Hpg_7	5.4% ± 0.3%
15	MBP tei sta	Pek7-rac-Hpg_7	2.6% ± 0.5%
16	MBP tei sta	Act7-rac-Hpg_7	31.5% ± 2.5%

[a]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. [b]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_{sta} (Figure 4, Table 1, entries 2–4). We found that StaF activity was dramatically reduced with a teicoplanin-like heptapeptide exhibiting the 7th amino acid in the unnatural D-configuration (Tei7-D-Hpg_7, 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, $OxyA_{tei}$ [13]. This behaviour is in contrast to that of StaH and other OxyB homologues, which exhibit similar activity on both Tei7-L-Hpg_7 and Tei7-D-Hpg_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 actinoidin-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_7 (Pek7-rac-Hpg_7, Act7-rac-Hpg_7) due to an inability to resolve the diastereomers via HPLC. We observed very little activity of StaF against Pek7-rac-Hpg_7 (Figure 4, Table 1, entry 3), which is similar to the results observed for $OxyA_{tei}$. 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 monocyclisation of the linear peptide) showed similar activity against Act7-rac-Hpg_7 (Figure 4, Table 1, entry 4) compared to Tei7-L-Hpg_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_7, Pek7-rac-Hpg_7 as well as Act7-rac-Hpg_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 selectivity of these homologues against altered peptide substrates.

Impact of the A47934 X-domain on StaF activity

Previously, it has been shown that StaH 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_{sta}, Table 1, entries 1–4), a PCP-X construct from teicoplanin biosynthesis (MBP-PCP-X_{tei}, Table 1, entries 5–8) and hybrid PCP-X constructs from A47934 and teicoplanin biosynthesis (MBP-PCP_{sta}-X_{tei}, Table 1, entries 9–12; MBP-PCP_{tei}-X_{sta}, Table 1, entries 13–16;) [12]. The influence of each individual PCP-X construct was tested with Tei7-L/D-Hpg_7, Pek7-rac-Hpg_7 and Act7-rac-Hpg_7 peptides. The presentation of Tei7-D-Hpg_7 and Pek7-rac-Hpg_7 by MBP-PCP-X_{tei}, MBP-PCP_{sta}-X_{tei} and MBP-PCP_{tei}-X_{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_7, StaF was active when the substrate was bound to MBP-PCP-X_{sta}, as described above, and we also observed a similar activity level when the peptide was presented by the MBP-PCP_{tei}-X_{sta} construct (Table 1, entry 13). However, against our expectations, StaF activity did not increase on Tei7-L-Hpg_7 when bound to PCP-X constructs exhibiting the teicoplanin X-domain (MBP-PCP-X_{tei}/PCP_{sta}-X_{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_7 peptide: StaF activity was highest with MBP-PCP_{tei}-X_{sta}, showed a minor decline with MBP-PCP-X_{sta} and MBP-PCP-X_{tei} and was considerably decreased with MBP-PCP_{sta}-X_{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_7 is not as clear as for Tei7-L-Hpg_7. An explanation for this could lie in the observation that Act7-rac-Hpg_7 seems to be a very good substrate for StaF (as it is for $OxyA_{tei}$), 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_7 is the peptide with the highest structural similarity to the natural substrate and the fact that StaF activity on Tei7-L-Hpg_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-Hpg_7 peptide and A47934 X-domain; StaH: Tei7-D-Hpg_7 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 $OxyE_{tei}$ (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.

Data collection	StaF Native ethylene glycol	StaF Native glycerol
Space group	$P3_12_1$ (152)	$P3_12_1$
Cell dimensions *a, b, c* (Å)	110.1, 110.1, 93.7	109.7, 109.7, 93.9
Molecules/asymmetric unit	1	1
X-ray source	SLS X10SA	SLS X10SA
Wavelength (Å)	0.9792	0.9792
Resolution (Å)[a]	50.0–2.1	50.0–2.2
R_{merge}[a]	0.07 (0.31)	0.10 (0.44)
$I/\sigma I$[a]	19.8 (4.2)	16.8 (5.5)
Completeness (%)[a]	95.1 (91.6)	98.4 (95.5)
Redundancy	6.2	9.6
Wilson *B*-factor (Å^2)	27.6	35.3
Refinement		
Unique Reflections	36206	31857
Resolution in refinement	50.0–2.1	50.0–2.2
R_{work}/R_{free}[b](%)	19.4 / 22.9	19.1 / 22.0
TLS-groups	−22–18; 19–75; 79–152; 153–225; 226–326; 327–391	−22–25; 26–75; 79–107; 108–206; 207–320; 321–391
No. of atoms	3690	3576
Protein	3315	3269
Heme	43	43
Ethylene glycol	64	–
Glycerol	–	48
Water	268	216
***B*-factors**		
Protein	35.1	35.5
Heme	21.5	22.0
Ethylene glycol	53.0	–
Glycerol	–	59.8
Water	43.8	41.7
RMSD		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.193	1.107
Ramachandran statistics[c]	97.3/ 2.2/ 0.5[e]	97.8/ 1.7/ 0.5[f]
Ramachandran statistics[d]	97.1/ 2 / 3	97.6/ 2 / 6
PDB Code	5EX8	5EX9

[a]Numbers in parentheses correspond to the highest resolution shell (2.2–2.1 Å; 2.3–2.2 Å). [b]$R_{work} = \sum ||F_o|-|F_c|| / \sum |F_o|$, calculated from the working reflection set; R_{free} calculated in the same manner using the 5% test set reflections. [c]Calculated by PROCHECK; percentage of the protein residues in favored/ allowed/ disallowed regions. [d]Calculated by MOLPROBITY; percentage of the protein residues in most favored regions; disallowed residues and percentage of bad rotamers. [e]Residues in disallowed region: E331 (disordered loop region), F382 (active site residue, clearly defined density). [f]Residues in disallowed region: A329 (disordered loop region), F382 (active site residue, clearly defined density).

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 $OxyA_{tei}$, 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

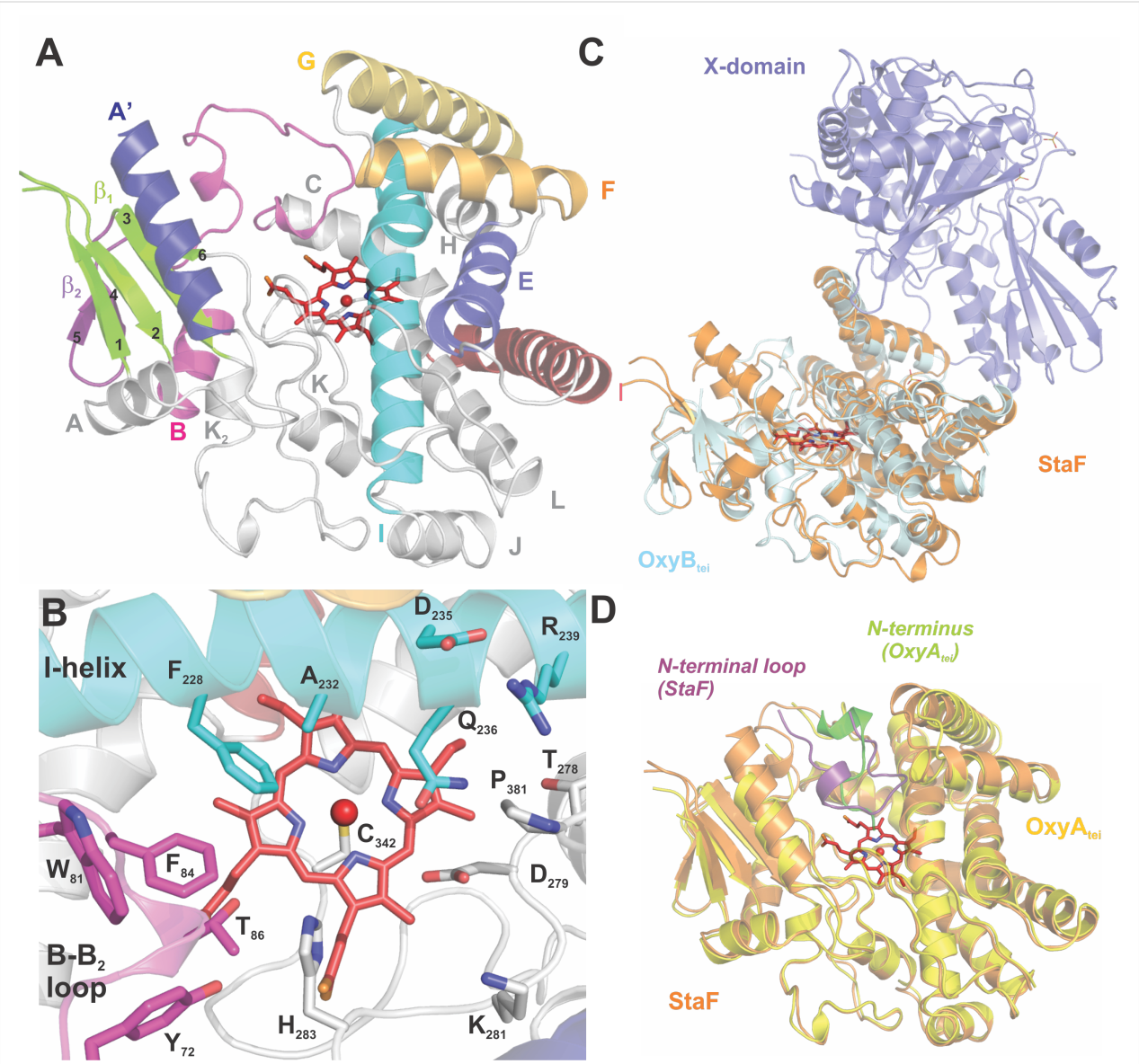


Figure 5: Structural analysis of StaF: (A) overall structure of StaF, with the heme moiety depicted using sticks and specific helices coloured and labelled; (B) view of the active site of StaF, with residues close to the heme moiety shown as sticks and labelled, with the colour scheme and labelling retained from panel (A); (C) an overlay of the StaF structure (orange) on the X_{tei}-$OxyB_{tei}$ complex (pale cyan/ blue; PDB ID: 4TX3); (D) an overlay of StaF (orange) on the structure of $OxyA_{tei}$ (yellow) showing the location of the N-terminal regions of other molecules within the crystal lattice for both StaF (purple) and $OxyA_{tei}$ (green).

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 $OxyA_{tei}$ (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 B–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), $OxyB_{tei}$ (PDB ID: 4TVF) 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 $OxyA_{tei}$ (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 $OxyA_{tei}$ apart from other structurally characterised examples of P450s involved in GPA cyclisation reactions. Comparison of StaF and $OxyA_{tei}$ (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 $OxyA_{tei}$ 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 $OxyA_{tei}$. 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 $OxyA_{tei}$ 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 $OxyA_{tei}$ 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 $OxyE_{tei}$ (PDB ID: 3O1A/3OO3) [21,22] and $OxyB_{tei}$ 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 $P450_{BioI}$ (PDB ID: 3EJD) [46-48] and the aminoacyl-PCP hydroxylases OxyD (PDB ID: 3MGX) and $P450_{sky}$ (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 $OxyB_{tei}$/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 $OxyA_{tei}$ 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.

PDB code	Chain	RMSD Cα [Å]	Z-score	% Identity	Description (donor organism)	Ref
5HH3	A	1.2	55.9	79	$OxyA_{tei}$ (*Actinoplanes teichomyceticus*)	[14]
5HH3	C	1.6	55.7	80	$OxyA_{tei}$ (*Actinoplanes teichomyceticus*)	[14]
3OO3	A	1.8	46.3	48	CYP165D3 (*Actinoplanes teichomyceticus*)	[21]
3O1A	A	2.0	46.3	48	CYP165D3 (*Actinoplanes teichomyceticus*)	[22]
4TX3	A	1.9	44.0	41	$OxyB_{tei}$ in complex with the X-domain (*Actinoplanes teichomyceticus*)	[16]
1LG9	A	2.6	43.6	38	CYP165B3 (*Nocardia orientalis*)	[41]
5EX6	A	2.2	43.4	41	StaH (*Streptomyces toyocaensis* NRRL15009)	[12]
1UED	A	2.1	43.2	33	CYP165C3 (*Nocardia orientalis*)	[40]
4OQS	A	2.2	42.1	34	CYP105AS1 *(Amycolatopsis orientalis)*	[42]
4TVF	A	2.1	41.2	42	$OxyB_{tei}$ (*Actinoplanes teichomyceticus*)	[19]
4XE3	B	2.4	41.1	30	OleP (*Streptomyces antibioticus*)	[43]
2YCA	A	2.3	40.8	28	MycG (*Micromonospora griseorubida*)	[44]
3E5L	A	2.5	40.7	34	CYP105P1 (*Streptomyces avermitilis*)	[45]
3EJD[a]	B	2.4	38.5	25	$P450_{BioI}$ (CYP107H1, *Bacillus subtilis*)	[47]
3MGX[a]	B	2.9	35.6	22	OxyD (CYP146, *Amycolatopsis orientalis*)	[51]
4PXH[a]	E	3.0	35.5	19	PCP_7-$P450_{sky}$ complex (CYP163B3, *Streptomyces sp. ACTA 2897*)	[49]

[a]Included for purposes of comparison.

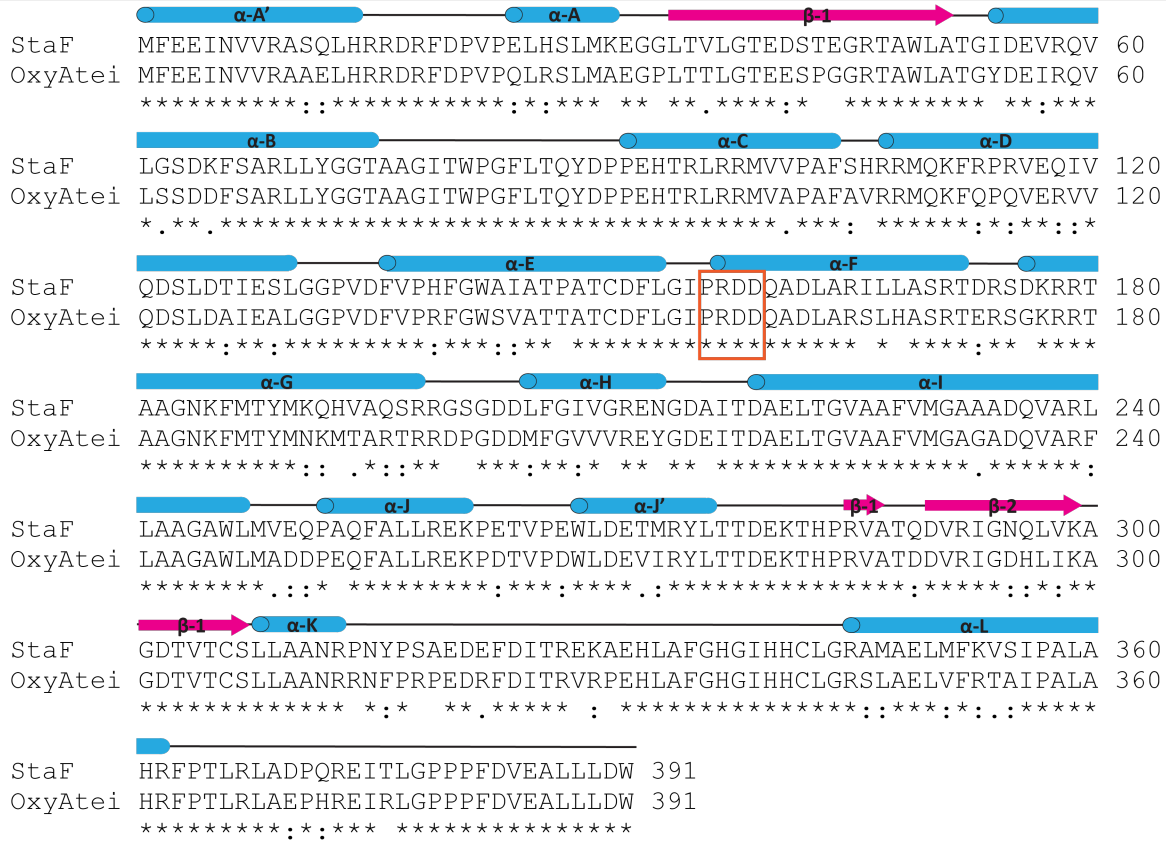


Figure 6: Sequence alignment of StaF and $OxyA_{tei}$. Protein secondary structure was derived from the StaF crystal structure (PDB ID: 5EX8) and is shown above the alignment (α-helices = blue, β-sheets = magenta). The PRDD-region, which has been shown to be crucial for interaction with the X-domain is highlighted in an orange box.

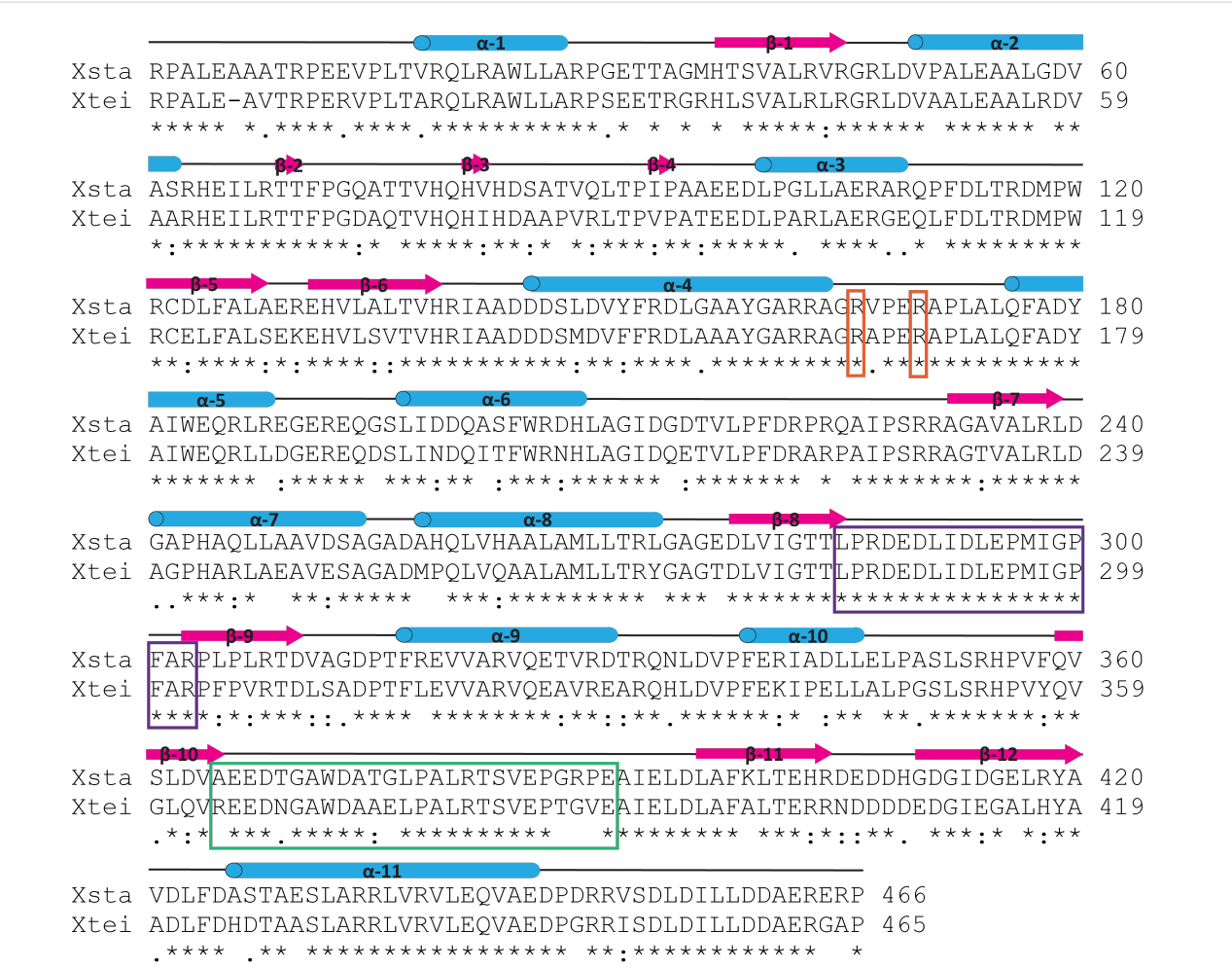


Figure 7: Sequence alignment of the A47934 (sta) and teicoplanin (tei) X-domain; secondary structure was derived from the X_{tei}-$OxyB_{tei}$ complex (PDB ID: 4TX3) and is shown above the alignment (α-helices = blue, β-sheets = magenta); the residues crucial for interaction with cytochrome P450s are shown in orange and both the crossover I region (purple) and the crossover II region (green) are highlighted.

Act7-rac-Hpg_7 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 $OxyA_{tei}$. StaF adopts the canonical P450 fold and strongly resembles the structure of $OxyA_{tei}$, 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 $OxyA_{tei}$. 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 X_{sta} appears to be weaker than the interaction of StaH to X_{sta}. We have previously shown that the strong StaH/X_{sta} 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 X_{sta} 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 X_{sta}, 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'- CTACCAGTCGAGCAGCAGGGCTTC- 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-PCP_{sta}-X_{sta} 1c, pET MBP-PCP_{tei}-X_{tei} 1c, pET MBP-PCP_{tei}-X_{sta} 1c, pET MBP-PCP_{sta}-X_{tei} 1c, pET NCL-4 MBP-X_{sta}) 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 OD_{600} = 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 OD_{600} = 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 at 5000*g* and 4 °C for 10 min and resuspended in lysis buffer (50 mM Tris pH 8, 50 mM NaCl, 10 mM imidazole, 0.5 mM DTE, EDTA-free SIGMAFAST™ Protease Inhibitor Cocktail Tablet).

All purification steps were performed at 4 °C if not stated otherwise. First, the cells were lysed by 3 passes through a microfluidizer (Microfluidics, Westwood, USA), before the lysate was centrifuged at 20,000*g* for 30 min. The cleared lysate was then subjected to Ni-NTA affinity chromatography in batch mode to purify the N-terminally hexahistidine-tagged StaF. Therefore, the Protino® 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 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 was buffer 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 spectroscopically 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 map 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$_{sta}$-X$_{sta}$, MBP-PCP$_{tei}$-X$_{tei}$, MBP-PCP$_{tei}$-X$_{sta}$, MBP-PCP$_{sta}$-X$_{tei}$, X$_{sta}$. 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 $Na_2S_2O_4$ 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$_{sta}$) 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$_{sta}$ was detected through a significant shift of the heme peak at λ = 415 nm to earlier elution volume when StaF and X$_{sta}$ were analysed together compared to individual analysis of StaF. Prior to analysis, 33.3 µM StaF and 100 µM X$_{sta}$ 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$_{sta}$ served as controls.

P450 activity assay

An in vitro phenolic coupling assay was performed in order to determine the StaF activity. As substrates the teicoplanin-like NH_2-D-Hpg-D-Tyr-L-Hpg-D-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Tei7(L-Hpg$_3$, D/L-Hpg$_7$)), the pekiskomycin-like NH_2-D-Ala-D-Tyr-L-Glu-D-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Pek7(D/L-Hpg$_7$)), and the actinoidin-like NH_2-D-Hpg-D-Tyr-L-Phe-D-Hpg-D-Hpg-L-Tyr-D/L-Hpg-C(O)R (Act7(D/L-Hpg$_7$)) 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$_7$ and D-Hpg$_7$ peptide could be used [13]. The diastereomers of Pek7 and Act7 were not separated by preparative HPLC, so that both peptides were used with a racemic mixture of D/L-Hpg$_7$ [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$_{sta}$-X$_{sta}$, MBP-PCP$_{tei}$-X$_{tei}$, MBP-PCP$_{tei}$-X$_{sta}$, MPB-PCP$_{sta}$-X$_{tei}$) using the R4-4 mutant of the promiscuous phosphopantetheinyl 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$_7$ 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_2PO_4, 0.3 M K_2HPO_4; 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 cryoprotectant solution (0.1 M phosphate/citrate buffer (pH 4.2), 1.2 M Na_2PO_4, 0.3 M K_2HPO_4 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*3(1)2(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 OxyE$_{tei}$ (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 performed 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-X_{tei}.
[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_{sta}-X_{sta}; 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 support of the Deutsche Forschungsgemeinschaft (Emmy-Noether Program, CR 392/1-1), Monash University and the EMBL Australia program. This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP170102220).

References

1. Yim, G.; Thaker, M. N.; Koteva, K.; Wright, G. *J. Antibiot.* **2014,** *67,* 31–41. doi:10.1038/ja.2013.117
2. Butler, M. S.; Hansford, K. A.; Blaskovich, M. A. T.; Halai, R.; Cooper, M. A. *J. Antibiot.* **2014,** *67,* 631–644. doi:10.1038/ja.2014.111
3. Al Toma, R. S.; Brieke, C.; Cryle, M. J.; Süssmuth, R. D. *Nat. Prod. Rep.* **2015,** *32,* 1207–1235. doi:10.1039/C5NP00025D
4. Peschke, M.; Gonsior, M.; Süssmuth, R. D.; Cryle, M. J. *Curr. Opin. Struct. Biol.* **2016,** *41,* 46–53. doi:10.1016/j.sbi.2016.05.018
5. Kittilä, T.; Mollo, A.; Charkoudian, L. K.; Cryle, M. J. *Angew. Chem., Int. Ed.* **2016,** *55,* 9834–9840. doi:10.1002/anie.201602614
6. Hur, G. H.; Vickery, C. R.; Burkart, M. D. *Nat. Prod. Rep.* **2012,** *29,* 1074–1098. doi:10.1039/c2np20025b
7. Cryle, M. J. *Metallomics* **2011,** *3,* 323–326. doi:10.1039/c0mt00081g
8. Cryle, M. J.; Stok, J. E.; De Voss, J. J. *Aust. J. Chem.* **2003,** *56,* 749–762. doi:10.1071/CH03040
9. Cryle, M. J.; Brieke, C.; Haslinger, K. *Amino Acids, Pept., Proteins* **2014,** *38,* 1–36. doi:10.1039/9781849737081-00001
10. Yim, G.; Wang, W.; Thaker, M. N.; Tan, S.; Wright, G. D. *ACS Infect. Dis.* **2016,** *2,* 642–650. doi:10.1021/acsinfecdis.6b00105
11. Yushchuk, O.; Ostash, B.; Pham, T. H.; Luzhetskyy, A.; Fedorenko, V.; Truman, A. W.; Horbal, L. *ACS Chem. Biol.* **2016,** *11,* 2254–2264. doi:10.1021/acschembio.6b00018
12. Ulrich, V.; Peschke, M.; Brieke, C.; Cryle, M. J. *Mol. BioSyst.* **2016,** *12,* 2992–3004. doi:10.1039/C6MB00373G
13. Peschke, M.; Haslinger, K.; Brieke, C.; Reinstein, J.; Cryle, M. *J. Am. Chem. Soc.* **2016,** *138,* 6746–6753. doi:10.1021/jacs.6b00307
14. Haslinger, K.; Cryle, M. *FEBS Lett.* **2016,** *590,* 571–581. doi:10.1002/1873-3468.12081
15. Brieke, C.; Kratzig, V.; Peschke, M.; Cryle, M. J. Facile Synthetic Access to Glycopeptide Antibiotic Precursor Peptides for the Investigation of Cytochrome P450 Action in Glycopeptide Antibiotic Biosynthesis. In *Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols;* Evans, S. B., Ed.; Springer New York: New York, NY, 2016; pp 85–102. doi:10.1007/978-1-4939-3375-4_6
16. Haslinger, K.; Peschke, M.; Brieke, C.; Maximowitsch, E.; Cryle, M. J. *Nature* **2015,** *521,* 105–109. doi:10.1038/nature14141
17. Brieke, C.; Peschke, M.; Haslinger, K.; Cryle, M. J. *Angew. Chem., Int. Ed.* **2015,** *54,* 15715–15719. doi:10.1002/anie.201507533
18. Brieke, C.; Kratzig, V.; Haslinger, K.; Winkler, A.; Cryle, M. J. *Org. Biomol. Chem.* **2015,** *13,* 2012–2021. doi:10.1039/C4OB02452D
19. Haslinger, K.; Maximowitsch, E.; Brieke, C.; Koch, A.; Cryle, M. J. *ChemBioChem* **2014,** *15,* 2719–2728. doi:10.1002/cbic.201402441
20. Schmartz, P. C.; Wölfel, K.; Zerbe, K.; Gad, E.; El Tamany, E. S.; Ibrahim, H. K.; Abou-Hadeed, K.; Robinson, J. A. *Angew. Chem., Int. Ed.* **2012,** *51,* 11468–11472. doi:10.1002/anie.201204458
21. Li, Z.; Rupasinghe, S. G.; Schuler, M. A.; Nair, S. K. *Proteins: Struct., Funct., Bioinf.* **2011,** *79,* 1728–1738. doi:10.1002/prot.22996
22. Cryle, M. J.; Staaden, J.; Schlichting, I. *Arch. Biochem. Biophys.* **2011,** *507,* 163–173. doi:10.1016/j.abb.2010.10.017
23. Woithe, K.; Geib, N.; Meyer, O.; Wörtz, T.; Zerbe, K.; Robinson, J. A. *Org. Biomol. Chem.* **2008,** *6,* 2861–2867. doi:10.1039/b805956j
24. Geib, N.; Woithe, K.; Zerbe, K.; Li, D. B.; Robinson, J. A. *Bioorg. Med. Chem. Lett.* **2008,** *18,* 3081–3084. doi:10.1016/j.bmcl.2007.11.093
25. Woithe, K.; Geib, N.; Zerbe, K.; Li, D. B.; Heck, M.; Fournier-Rousset, S.; Meyer, O.; Vitali, F.; Matoba, N.; Abou-Hadeed, K.; Robinson, J. A. *J. Am. Chem. Soc.* **2007,** *129,* 6887–6895. doi:10.1021/ja071038f
26. Zerbe, K.; Woithe, K.; Li, D. B.; Vitali, F.; Bigler, L.; Robinson, J. A. *Angew. Chem., Int. Ed.* **2004,** *43,* 6709–6713. doi:10.1002/anie.200461278
27. Hadatsch, B.; Butz, D.; Schmiederer, T.; Steudle, J.; Wohlleben, W.; Süssmuth, R.; Stegmann, E. *Chem. Biol.* **2007,** *14,* 1078–1089. doi:10.1016/j.chembiol.2007.08.014
28. Stegmann, E.; Pelzer, S.; Bischoff, D.; Puk, O.; Stockert, S.; Butz, D.; Zerbe, K.; Robinson, J.; Süssmuth, R. D.; Wohlleben, W. *J. Biotechnol.* **2006,** *124,* 640–653. doi:10.1016/j.jbiotec.2006.04.009
29. Bischoff, D.; Bister, B.; Bertazzo, M.; Pfeifer, V.; Stegmann, E.; Nicholson, G. J.; Keller, S.; Pelzer, S.; Wohlleben, W.; Süssmuth, R. D. *ChemBioChem* **2005,** *6,* 267–272. doi:10.1002/cbic.200400328
30. Bischoff, D.; Pelzer, S.; Höltzel, A.; Nicholson, G. J.; Stockert, S.; Wohlleben, W.; Jung, G.; Süssmuth, R. D. *Angew. Chem., Int. Ed.* **2001,** *40,* 1693–1696. doi:10.1002/1521-3773(20010504)40:9<1693::AID-ANIE16930>3.0.CO;2-8

31. Bischoff, D.; Pelzer, S.; Bister, B.; Nicholson, G. J.; Stockert, S.; Schirle, M.; Wohlleben, W.; Jung, G.; Süssmuth, R. D. *Angew. Chem., Int. Ed.* **2001,** *40,* 4688–4691. doi:10.1002/1521-3773(20011217)40:24<4688::AID-ANIE4688>3.0.CO;2-M
32. Süssmuth, R. D.; Pelzer, S.; Nicholson, G.; Walk, T.; Wohlleben, W.; Jung, G. *Angew. Chem., Int. Ed.* **1999,** *38,* 1976–1979. doi:10.1002/(SICI)1521-3773(19990712)38:13/14<1976::AID-ANIE1976>3.0.CO;2-3
33. Pootoolal, J.; Thomas, M. G.; Marshall, C. G.; Neu, J. M.; Hubbard, B. K.; Walsh, C. T.; Wright, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 8962–8967. doi:10.1073/pnas.102285099
34. Ortiz de Montellano, P. R. *Cytochrome P450 Structure, Mechanism and Biochemistry;* Springer, 2015; pp 912 ff.
35. Munro, A. W.; Girvan, H. M.; Mason, A. E.; Dunford, A. J.; McLean, K. J. *Trends Biochem. Sci.* **2013,** *38,* 140–150. doi:10.1016/j.tibs.2012.11.006
36. Brieke, C.; Cryle, M. J. *Org. Lett.* **2014,** *16,* 2454–2457. doi:10.1021/ol500840f
37. Sunbul, M.; Marshall, N. J.; Zou, Y.; Zhang, K.; Yin, J. *J. Mol. Biol.* **2009,** *387,* 883–898. doi:10.1016/j.jmb.2009.02.010
38. Bell, S. G.; Xu, F.; Johnson, E. O. D.; Forward, I. M.; Bartlam, M.; Rao, Z.; Wong, L.-L. *J. Biol. Inorg. Chem.* **2010,** *15,* 315–328. doi:10.1007/s00775-009-0604-7
39. Thaker, M. N.; Wang, W.; Spanogiannopoulos, P.; Waglechner, N.; King, A. M.; Medina, R.; Wright, G. D. *Nat. Biotechnol.* **2013,** *31,* 922–927. doi:10.1038/nbt.2685
40. Pylypenko, O.; Vitali, F.; Zerbe, K.; Robinson, J. A.; Schlichting, I. *J. Biol. Chem.* **2003,** *278,* 46727–46733. doi:10.1074/jbc.M306486200
41. Zerbe, K.; Pylypenko, O.; Vitali, F.; Zhang, W.; Rouset, S.; Heck, M.; Vrijbloed, J. W.; Bischoff, D.; Bister, B.; Süssmuth, R. D.; Pelzer, S.; Wohlleben, W.; Robinson, J. A.; Schlichting, I. *J. Biol. Chem.* **2002,** *277,* 47476–47485. doi:10.1074/jbc.M206342200
42. McLean, K. J.; Hans, M.; Meijrink, B.; van Scheppingen, W. B.; Vollebregt, A.; Tee, K. L.; van der Laan, J.-M.; Leys, D.; Munro, A. W.; van den Berg, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 2847–2852. doi:10.1073/pnas.1419028112
43. Montemiglio, L. C.; Parisi, G.; Scaglione, A.; Sciara, G.; Savino, C.; Vallone, B. *Biochim. Biophys. Acta* **2016,** *1860,* 465–475. doi:10.1016/j.bbagen.2015.10.009
44. Li, S.; Tietz, D. R.; Rutaganira, F. U.; Kells, P. M.; Anzai, Y.; Kato, F.; Pochapsky, T. C.; Sherman, D. H.; Podust, L. M. *J. Biol. Chem.* **2012,** *287,* 37880–37890. doi:10.1074/jbc.M112.410340
45. Xu, L.-H.; Fushinobu, S.; Ikeda, H.; Wakagi, T.; Shoun, H. *J. Bacteriol.* **2009,** *191,* 1211–1219. doi:10.1128/JB.01276-08
46. Cryle, M. J. *Biochem. Soc. Trans.* **2010,** *38,* 934–939. doi:10.1042/BST0380934
47. Cryle, M. J.; Schlichting, I. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 15696–15701. doi:10.1073/pnas.0805983105
48. Cryle, M. J.; De Voss, J. J. *Chem. Commun.* **2004,** 86–87. doi:10.1039/B311652B
49. Haslinger, K.; Brieke, C.; Uhlmann, S.; Sieverling, L.; Süssmuth, R. D.; Cryle, M. J. *Angew. Chem., Int. Ed.* **2014,** *53,* 8518–8522. doi:10.1002/anie.201404977
50. Uhlmann, S.; Süssmuth, R. D.; Cryle, M. J. *ACS Chem. Biol.* **2013,** *8,* 2586–2596. doi:10.1021/cb400555e
51. Cryle, M. J.; Meinhart, A.; Schlichting, I. *J. Biol. Chem.* **2010,** *285,* 24562–24574. doi:10.1074/jbc.M110.131904
52. Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66,* 125–132. doi:10.1107/S0907444909047337
53. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007,** *40,* 658–674. doi:10.1107/S0021889807021206
54. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66,* 486–501. doi:10.1107/S0907444910007493
55. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997,** *53,* 240–255. doi:10.1107/S0907444996012255
56. Winn, M. D.; Isupov, M. N.; Murshudov, G. N. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001,** *57,* 122–133. doi:10.1107/S0907444900014736
57. Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998,** *54,* 905–921. doi:10.1107/S0907444998003254
58. Brunger, A. T. *Nat. Protoc.* **2007,** *2,* 2728–2733. doi:10.1038/nprot.2007.406
59. Painter, J.; Merritt, E. A. *J. Appl. Crystallogr.* **2006,** *39,* 109–111. doi:10.1107/S0021889805038987
60. Painter, J.; Merritt, E. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006,** *62,* 439–450. doi:10.1107/S0907444906005270
61. Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66,* 12–21. doi:10.1107/S0907444909042073
62. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993,** *26,* 283–291. doi:10.1107/S0021889892009944
63. Krissinel, E.; Henrick, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004,** *60,* 2256–2268. doi:10.1107/S0907444904026460
64. Holm, L.; Rosenström, P. *Nucleic Acids Res.* **2010,** *38* (Suppl. 2), W545–W549. doi:10.1093/nar/gkq366
65. *The PyMOL Molecular Graphics System,* 1.8; Schrödinger, LLC.

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.12.284

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

A postsynthetically 2'-"clickable" uridine with arabino configuration and its application for fluorescent labeling and imaging of DNA

Heidi-Kristin Walter[1], Bettina Olshausen[2], Ute Schepers[2] and Hans-Achim Wagenknecht[*1]

Full Research Paper

Open Access

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

Beilstein J. Org. Chem. **2017,** *13,* 127–137.
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".

Guest Editor: H. B. Bode

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

Abstract

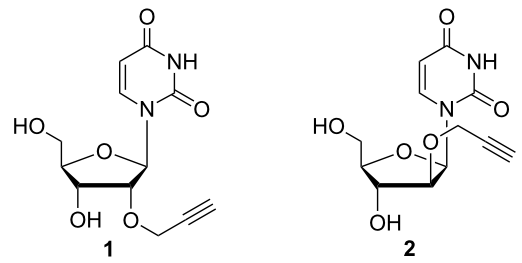
The arabino-configured analog of uridine with a propargyl group at the 2'-position was synthesized and incorporated into DNA by solid-phase chemistry. The fluorescence quantum yields of DNA strands that were postsynthetically modified by blue and green emitting cyanine-styryl dyes were improved due to 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-1*H*-1,2,3-triazol-4-

yl)methyl]amine (TBTA) and better water-soluble derivatives [9,10]. The CuAAC cannot only be applied for conventional postsynthetic oligonucleotide modification in solution but also on solid phase [11] and for the introduction of multiple postsynthetic modifications [12]. The azide groups for CuAAC are typically placed onto the fluorescent dyes since azides are not compatible with phosphoramidite chemistry. The alkyne groups as reactive precursors are attached to the oligonucleotide [13], especially at the 5-position of pyrimidines [13], the 7-position of 7-deazapurines [14], and the 2'-position of ribofuranosides [11,15]. These positions were chosen since they are typically accepted by DNA polymerases in primer extension experiments and PCR [4,16].

To develop fluorescently labelled oligonucleotides that undergo energy transfer reactions [17] we recently applied 2'-propargyl-modified uridine **1** as DNA building block (Scheme 1) [15,18,19]. A simple look on the three-dimensional structure of double-helical DNA elucidates that the positioning of the fluorophores in the major groove may be improved by inversion of the configuration at the 2'-position of the anchor nucleoside sugar. In fact, arabino nucleic acids are an important class of antisense oligonucleotides [20] since their first report [21]. The orientation of the 2'-OH group in the arabino configuration towards the major groove yields hybrids with RNA that show a slightly lower thermal stability compared to DNA/RNA hybrids. In order to evaluate this structural influence for our fluorescently labelled oligonucleotides, we developed and synthesized the 2'-propargyl-modified arabino-configured uridine analog **2**, incorporated it into DNA by automated phosphoramidite chemistry, "clicked" it to a variety of our recently established, photostable cyanine-styryl dyes and probed the fluorescence and energy transfer properties by determination of quantum yields and emission color contrasts.

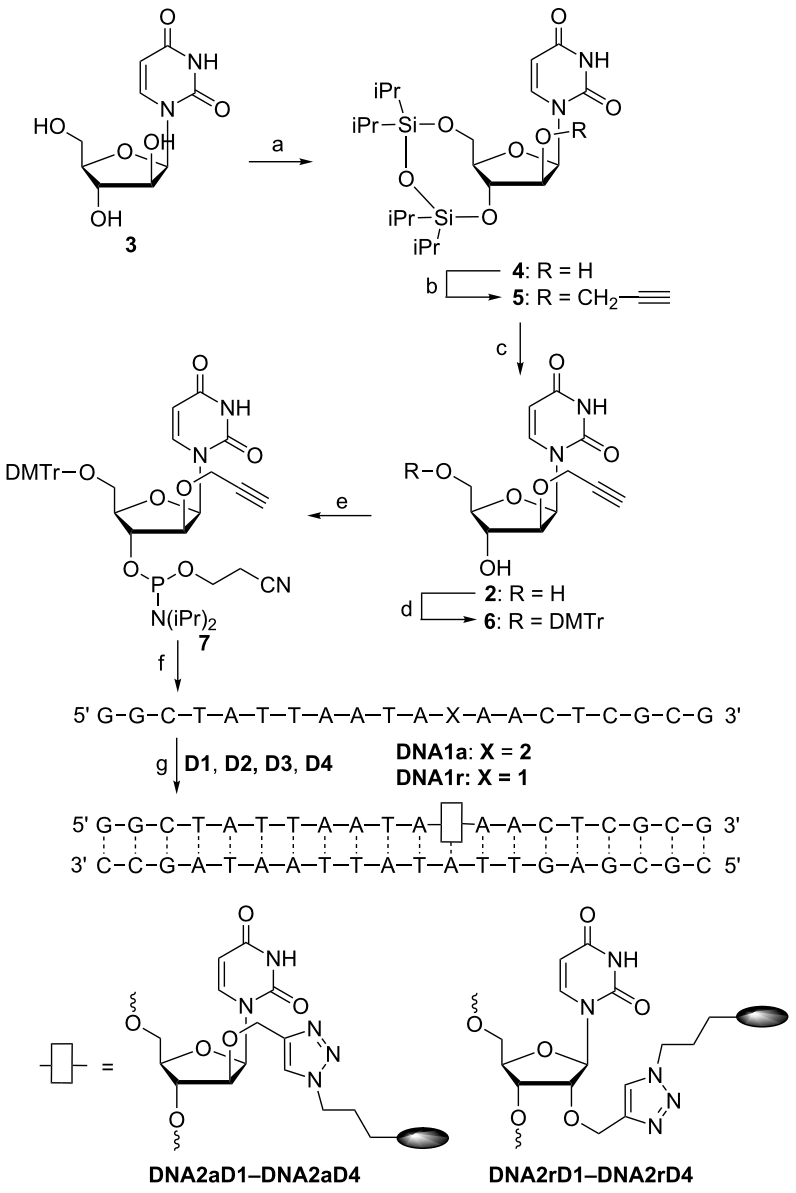


Scheme 1: 2'-Propargylated nucleosides as "clickable" DNA/RNA building blocks with ribo (**1**) and arabino (**2**) configuration.

Results and Discussion

The synthesis of the phosphoramidite **7** (Scheme 2) was straightforward and includes mainly protecting group chemistry since it starts with the commercially available arabino-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. NH_4OH 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 H_2O/DMSO/*t*-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 Φ_F and melting temperatures T_m (Table 1). The reference duplexes of **DNA1a** and **DNA1r** annealed with the unmodified complementary strand showed T_m 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) $TIPDSiCl_2$, 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, $(iPr)_2NEt$, CH_2Cl_2, rt, 3 h, 95%; f) automated DNA synthesis; g) **D1**–**D4**, sodium ascorbate, TBTA, $(CH_3CN)_4CuPF_6$, H_2O/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**.

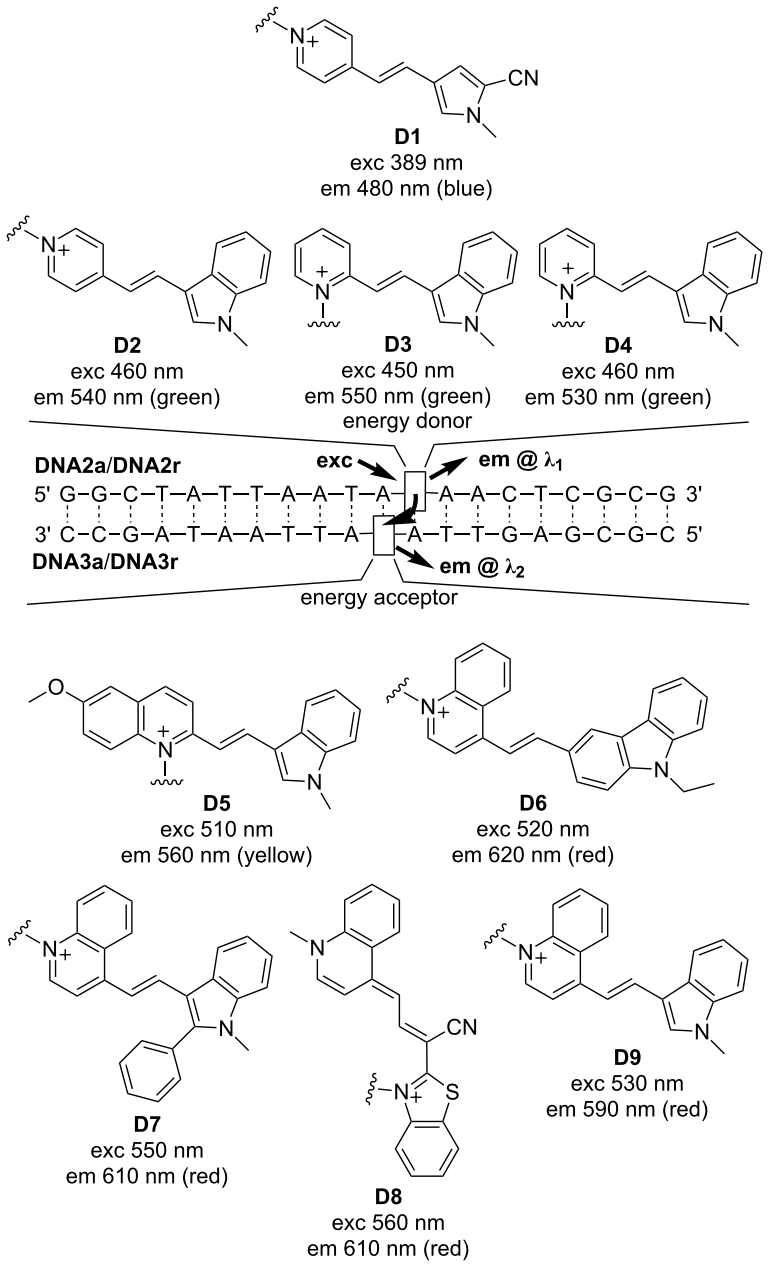
dye	λ_{exc} (nm)	**DNA2a…** T_m [°C]	Φ_F	**DNA2r…** T_m [°C]	Φ_F
…D1	389	61.9	0.096[a]	65.7	0.052[a]
…D2	462	61.7	0.452[b]	64.0	0.266[b]
…D3	450	64.2	0.136[c]	65.1	0.122[c]
…D4	462	64.2	0.156[d]	65.2	0.087[d]

[a]λ_{em} = 404–800 nm; [b]λ_{em} = 477–800 nm; [c]λ_{em} = 480–800 nm; [d]λ_{em} = 473–800 nm.

arabino-modified duplexes show a smaller stabilization effect by the dyes than the corresponding ribo-modified duplexes. The stabilization of ds**DNA2a** ranges only from 0.7 °C for **D1** to 3.2 °C for **D3** and **D4**, whereas the stabilizing effects for ds**DNA2r** 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 ds**DNA2a** are all

higher than the corresponding ones of ds**DNA2r**. Especially in case of **D2** Φ_F 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



Scheme 3: Structures of donor dyes **D1**–**D4** as modifications of **DNA2a** and **DNA2r** and structures of acceptor dyes **D5**–**D8** 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 $C = I_{Ac}/I_{Do}$ (fluorescence intensity near maximum of acceptor divided by fluorescence intensity near maximum of donor) and for the fluorescence quantum yield Φ_F 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 exemptions; 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 2: Fluorescence intensity ratios (color contrast) $C = I_{Ac}/I_{Do}$ and fluorescence quantum yields Φ_F of energy transfer pairs between dyes **D1–D4** in **DNA2a** and **DNA2r** and dyes **D5–D9** in **DNA3a and DNA3r**. The abbreviations a and r are listed in the order according to the duplex formation between **DNA2** (first letter) with **DNA3** (second letter), for instance a–r means **DNA2a**–**DNA3r**.

Do→ Ac↓		**DNA2a** and **DNA2r**							
DNA3a and **DNA3r**		**D1**		**D2**		**D3**		**D4**	
		C	Φ_F	C	Φ_F	C	Φ_F	C	Φ_F
D5	**a–a**	35	0.146[a]	–	–	–	–	–	–
	a–r	198	0.606	–	–	–	–	–	–
	r–a	129	0.224	–	–	–	–	–	–
	r–r	70	0.217	–	–	–	–	–	–
D6	**a–a**	15	0.148[b]	–	–	–	–	–	–
	a–r	48	0.227	–	–	–	–	–	–
	r–a	11	0.127	–	–	–	–	–	–
	r–r	40	0.206	–	–	–	–	––	–
D7	**a–a**	44	0.273[c]	20	0.237[d]	41	0.198[d]	69	0.212[d]
	a–r	85	0.357	36	0.245	177	0.319	39	0.214
	r–a	46	0.213	10	0.210	82	0.218	108	0.268
	r–r	93	0.340	60	0.312	136	0.218	153	0.229
D8	**a–a**	109	0.606[c]	20	0.466[e]	41	0.528[f]	83	0.672[e]
	a–r	80	0.576	12	0.427	43	0.564	48	0.592
	r–a	215	0.719	15	0.378	77	0.549	86	0.534
	r–r	87	0.545	3	0.078	40	0.388	45	0.366
D9	**a–a**	60	0.307[g]	11	0.222[h]	9	0.148[h]	28	0.220[h]
	a–r	58	0.306	23	0.245	62	0.285	27	0.258
	r–a	215	0.240	7	0.132	30	0.237	34	0.184
	r–r	69	0.245	25	0.226	59	0.244	38	0.206

[a]λ_{exc} = 389 nm, λ_{em} = 515–800 nm; [b]λ_{exc} = 389 nm, λ_{em} = 525–800 nm; [c]λ_{exc} = 389 nm, λ_{em} = 550–800 nm; [d]λ_{exc} = 435 nm, λ_{em} = 550–800 nm; [e]λ_{exc} = 430 nm, λ_{em} = 550–800 nm; [f]λ_{exc} = 430 nm, λ_{em} = 540–800 nm; [g]λ_{exc} = 389 nm, λ_{em} = 530–800 nm; [h]λ_{exc} = 423 nm, λ_{em} = 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–DNA3rD5**) 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

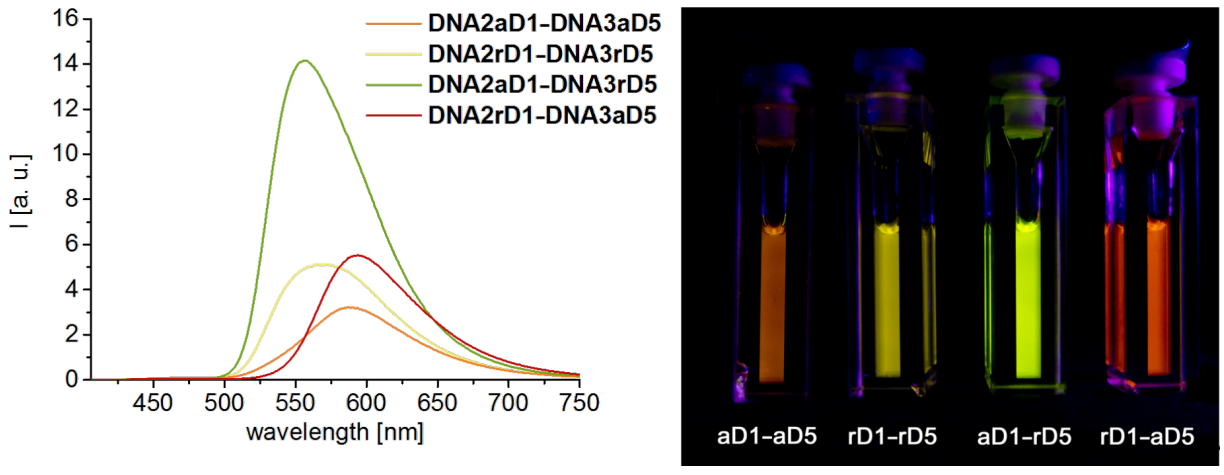


Figure 1: Representative demonstration of the fluorescence readout differences between the four arabino/ribo combinations of **D1** (donor) and **D5** (acceptor). Left: Fluorescence of **DNA2a/rD1–DNA3a/rD5**; 2.5 µM DNA in 50 mM Na-P_i buffer, 250 mM NaCl, pH 7, λ_{exc} = 391 nm. Right: Corresponding image of cuvettes excited by a handheld UV lamp.

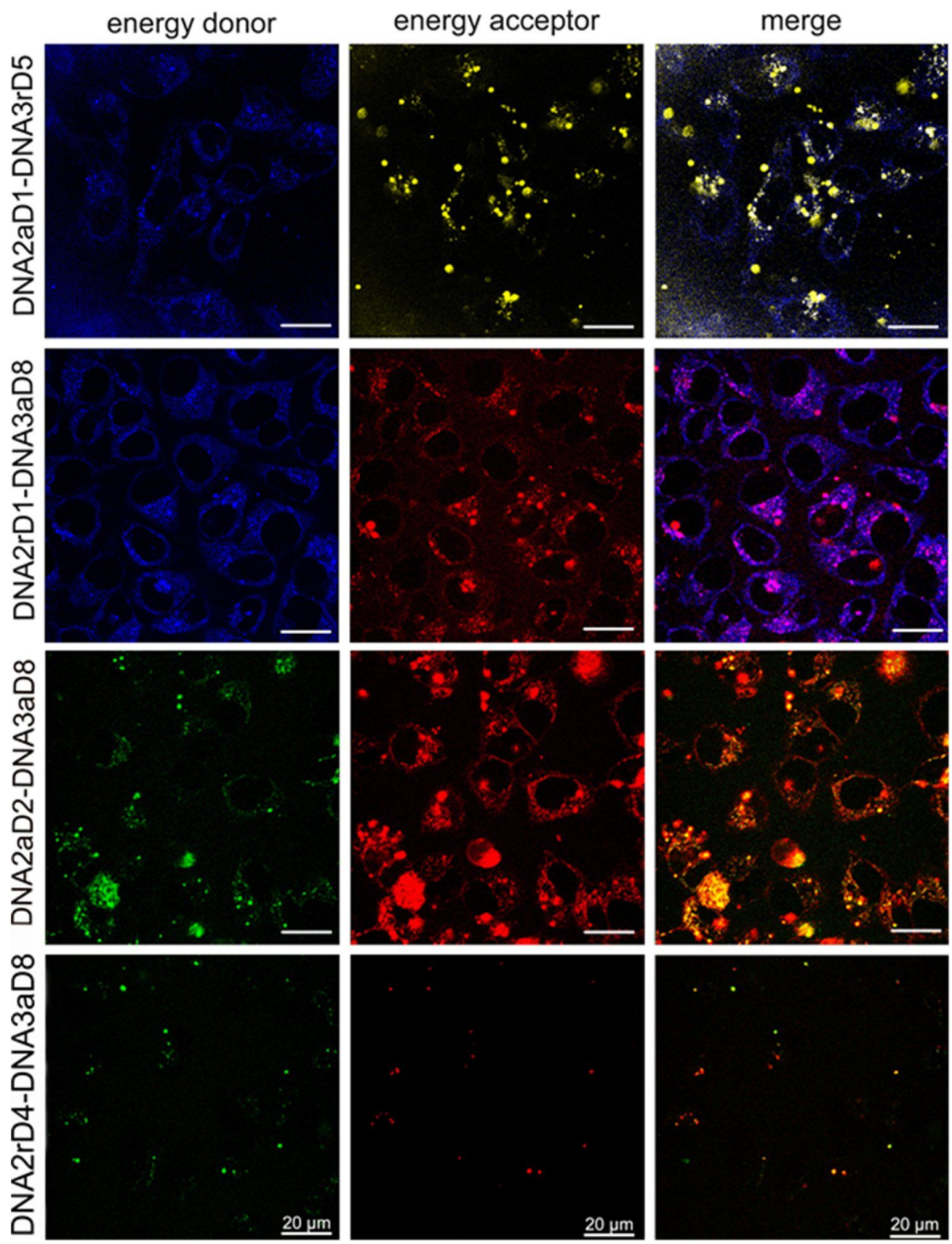


Figure 2: Confocal microscopy of HeLa cells after transfection with **DNA2aD1**–**DNA3rD5** (row 1), **DNA2rD1**–**DNA3aD8** (row 2), **DNA2aD2–DNA3aD8** (row 3) and **DNA2rD4**–**DNA3aD8** (row 4). The visualization was performed using a Leica TCS-SPE (DMi8) inverted microscope with an ACS APO 63×/1.30 oil objective. For **DNA2aD1**–**DNA3rD5** λ_{exc} = 405 nm (UV laser), λ_{em} = 435–470 nm (blue) and 575–750 nm (yellow), for **DNA2rD1**–**DNA3aD8** λ_{exc} = 405 nm (UV laser), λ_{em} = 415–550 nm (blue) and 575–750 nm (red), for **DNA2aD2–DNA3aD8** λ_{exc} = 488 nm (argon ion laser), λ_{em} = 490–550 nm (green) and 550–675 nm (red), for **DNA2rD4**–**DNA3aD8** λ_{exc} = 488 nm (argon ion laser), λ_{em} = 490–550 nm (green) and 675–800 nm (red), scale bar = 20 µm.

recently established class of cyanine-styryl dyes that show excellent photostabilities. The two-by-two combinations of these dyes give energy transfer pairs with blue-to-yellow, blue-to-red and green-to-red emission color changes. For these dye combinations, we probed all four combinations of arabino- and ribo-configured donor strands with arabino- and ribo-configured acceptor strands, and screened this array of doubly modified DNA duplexes by their emission color contrast C and the fluorescence quantum yield Φ_F. This screening revealed that the combination of donor and acceptor dyes does not necessarily yield better optical properties if they are both linked to the arabino-configured nucleoside **2** (compared to the linkage to the ribo-configured nucleoside **1**). However, there are some remarkable examples in this array of duplexes with mixed combinations, that means donor dyes linked to the arabino-configured nucleoside **2** with acceptor dyes linked to the ribo-configured nucleoside **1**, and vice versa, that showed significantly improved emission color contrasts and/or fluorescence quantum yields. Thereby, improved fluorescent nucleic acid probes were elucidated that are suitable not only for nucleic acid imaging of living cells but additionally allow a two-color readout.

Experimental

Materials and methods. Chemicals and dry solvents were purchased from Aldrich, ABCR, and VWR and were used without further purification unless otherwise stated. Unmodified oligonucleotides were purchased from Metabion. TLC was performed on Fluka silica gel 60 F254 coated aluminum foil. FAB mass spectra were measured by the analytical facilitites of the Institute of Organic Chemistry (KIT) using a Finnigan MAT95 in positive ionization mode. NMR spectra were recorded on a Bruker B-ACS-60, Bruker Avance DRX 400 and a Bruker Avance DRX 500 spectrometer in deuterated solvents (^{1}H at 300, 400 or 500 MHz, ^{13}C at 75, 100 or 125 MHz). Chemical shifts are given in ppm relative to TMS. IR spectra were recorded by the analytical facility of the Institute of Organic Chemistry (KIT) on a Bruker IFS88 spectrometer.

Optical-spectroscopic measurements were recorded in NaP_i-buffer solution (10 mM, pH 7) with 250 mM NaCl in quartz glass cuvettes (10 mm). Absorption spectra were recorded with a Varian Cary 100 spectrometer equipped with a 6 × 6 cell changer unit at 20 °C. Fluorescence was measured with a Jobin–Yvon Fluoromax 3 fluorimeter with a step width of 1 nm and an integration time of 0.2 s. All spectra were recorded at 20 °C and are corrected for Raman emission from the buffer solution. Quantum yields were determined with Quantaurus QY C11347 of Hamamatsu.

DNA2aD1 to **DNA2aD4**, **DNA2rD1** to **DNA2rD4**, **DNA3aD5** to **DNA3aD9** and **DNA3rD5** to **DNA3rD9** were purified using a 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 Biflex-IV spectrometer from Bruker Daltonics in the linear negative mode (matrix: 1:9 mixture of diammonium 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 $TIPDSiCl_2$ (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 (SiO_2, 0 → 50% EtOAc in CH_2Cl_2). 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 $NaHCO_3$ solution and then dried over Na_2SO_4. The solvent was removed under reduced pressure and the residue was purified by column chromatography (SiO_2, 0–40% EtOAc in hexane) to obtain **5** (0.716 g, 1.37 mmol, 65%) as a colorless foam. R_f 0.40 (hexane/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) 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'), 5.63 (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, OCH_2, H-3'), 4.07 (dd, J = 13.2 Hz, 2.4 Hz, 1H, H-5$_a$'), 4.00 (dd, J = 13.2 Hz, 2.9 Hz, 1H, H-5$_b$'), 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, 8× CH_3 & 4× CH) ppm; ^{13}C NMR (75 MHz, $CDCl_3$) 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-CH), 75.5 (CH), 72.4 (C-3'), 60.5 (C-5'), 59.5 (OCH_2), 17.6–17.0 (8 CH_3), 13.6–12.5 (4 CH) ppm; FAB–MS m/z (%): 525.2 (65) $[M + H]^+$; FAB–HRMS

FAB *m/z*: $[M + H]^+$calcd for $C_{24}H_{41}N_2O_7Si_2^+$, 525.2447; found, 525.2447.

Synthesis of 2. Under an Ar atmosphere **5** (0.687 g, 1.31 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_2Cl_2/MeOH 5:1. The solvent was removed under reduced pressure and the crude product was purified by column chromatography (SiO_2, CH_2Cl_2/MeOH 10:1) to afford **2** (0.366 g, 1.30 mmol, 99%) as a colorless foam. R_f 0.24 (CH_2Cl_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_2), 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* = 2.3 Hz, 1H, CH) ppm; ^{13}C NMR (126 MHz, DMSO-d_6) δ 163.1 (C-4), 150.4 (C-2), 141.9 (C-6), 100.6 (C-5), 83.2 (C-1'), 82.8 (C-4'), 82.4 (C-2'), 79.5 (C-CH), 77.6 (CH), 72.5 (C-3'), 59.9 (C-5'), 57.5 (OCH_2) ppm; FAB–MS *m/z* (%): 242.3 (100) $[M]^+ - CH_2CCH$ (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_2SO_4, and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (SiO_2, CH_2Cl_2/MeOH 99:1 + 0.1% NEt_3) to afford **6** (0.729 g, 1.25 mmol, 99%) as a colorless foam. R_f 0.13 (CH_2Cl_2/MeOH 50:1); ^{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_2, H-3'), 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$_a$'), 3.46 (dd, *J* = 10.8 Hz, 3.8 Hz, 1H H-5$_b$'), 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* (%): 585.1 (68) $[M + H]^+$; FAB–HRMS *m/z*: $[M + H]^+$calcd for $C_{33}H_{33}N_2O_8^+$, 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_2Cl_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_2Cl_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_2Cl_2/acetone 5:1); APCI–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 **7** 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_4OH at 45 °C for 16 h.

Click reaction with modified oligonucleotides. To the lyophilized alkyne-modified DNA sample were added water (100 µL), sodium ascorbate (25 µL of 0.4 M in water), tris[(1-benzyl-1*H*-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(acetonitrile) 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_2EDTA (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 microsopy. 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 5×10^4 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, Eggenstein-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% CO_2 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 Information 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

1. Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001,** *40,* 2004–2021. doi:10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5
2. Lallana, E.; Riguera, R.; Fernandez-Megia, E. *Angew. Chem., Int. Ed.* **2011,** *50,* 8794–8804. doi:10.1002/anie.201101019
3. Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T. *Angew. Chem., Int. Ed.* **2008,** *47,* 8350–8358. doi:10.1002/anie.200802077
4. El-Sagheer, A. H.; Brown, T. *Chem. Soc. Rev.* **2010,** *39,* 1388–1405. doi:10.1039/b901971p
5. Paredes, E.; Das, S. R. *ChemBioChem* **2011,** *12,* 125–131. doi:10.1002/cbic.201000466
6. Huisgen, R. *Angew. Chem., Int. Ed. Engl.* **1963,** *2,* 565–598. doi:10.1002/anie.196305651
7. Rostovstev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002,** *41,* 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4
8. Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002,** *67,* 3057–3064. doi:10.1021/jo011148j
9. Chan, R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004,** *6,* 2853–2855. doi:10.1021/ol0493094
10. Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; del Amo, D. S.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. *Angew. Chem., Int. Ed.* **2011,** *50,* 8051–8056. doi:10.1002/anie.201101817
11. Astakhova, I. K.; Wengel, J. *Chem. – Eur. J.* **2013,** *19,* 1112–1122. doi:10.1002/chem.201202621
12. Gramlich, P. M. E.; Warncke, S.; Gierlich, J.; Carell, T. *Angew. Chem., Int. Ed.* **2008,** *47,* 3442–3444. doi:10.1002/anie.200705664
13. Gierlich, J.; Burley, G. A.; Gramlich, P. M. E.; Hammond, D. M.; Carell, T. *Org. Lett.* **2006,** *8,* 3639–3642. doi:10.1021/ol0610946
14. Seela, F.; Xiong, H.; Leonard, P.; Budow, S. *Org. Biomol. Chem.* **2009,** *7,* 1374–1387. doi:10.1039/b822041g
15. Berndl, S.; Herzig, N.; Kele, P.; Lachmann, D.; Li, X.; Wolfbeis, O. S.; Wagenknecht, H.-A. *Bioconjugate Chem.* **2009,** *20,* 558–564. doi:10.1021/bc8004864
16. Wenge, U.; Ehrenschwender, T.; Wagenknecht, H.-A. *Bioconjugate Chem.* **2013,** *24,* 301–304. doi:10.1021/bc300624m
17. Holzhauser, C.; Wagenknecht, H.-A. *J. Org. Chem.* **2013,** *78,* 7373–7379. doi:10.1021/jo4010102
18. Schmucker, W.; Wagenknecht, H.-A. *Synlett* **2012,** *23,* 2435–2448. doi:10.1055/s-0032-1317158
19. Bohländer, P. R.; Wagenknecht, H.-A. *Eur. J. Org. Chem.* **2014,** 7547–7551. doi:10.1002/ejoc.201403119
20. Prakash, T. P. *Chem. Biodiversity* **2011,** *8,* 1616–1641. doi:10.1002/cbdv.201100081
21. Giannaris, P. A.; Damha, M. J. *Can. J. Chem.* **1994,** *72,* 909–918. doi:10.1139/v94-118
22. Markiewicz, W. T.; Wiewórowski, M. *Nucleic Acids Res.* **1978,** *1* (Suppl. 1), s186–s190. doi:10.1093/nar/1.suppl_1.s185
23. Bohländer, P. R.; Vilaivan, T.; Wagenknecht, H.-A. *Org. Biomol. Chem.* **2015,** *13,* 9223–9230. doi:10.1039/C5OB01273B
24. Bohländer, P. R.; Wagenknecht, H.-A. *Methods Appl. Fluoresc.* **2015,** *3,* 044003. doi:10.1088/2050-6120/3/4/044003
25. Schwechheimer, C.; Merkel, M.; Bohländer, P. R.; Wagenknecht, H.-A. Synthetic Wavelength-Shifting Fluorescent Probes of Nucleic Acids. In *Modified Nucleic Acids;* Nakatani, K.; Tor, Y., Eds.; Springer International Publishing: Switzerland, 2016; pp 83–100. doi:10.1007/978-3-319-27111-8_4
26. Holzhauser, C.; Wagenknecht, H.-A. *Angew. Chem., Int. Ed.* **2011,** *50,* 7268–7272. doi:10.1002/anie.201101968
27. Bohländer, P. R.; Abba, M. L.; Bestvater, F.; Allgayer, H.; Wagenknecht, H.-A. *Org. Biomol. Chem.* **2016,** *14,* 5001–5006. doi:10.1039/C6OB00691D
28. Holzhauser, C.; Liebl, R.; Göpferich, A.; Wagenknecht, H.-A.; Breunig, M. *ACS Chem. Biol.* **2013,** *8,* 890–894. doi:10.1021/cb3006616
29. Bohländer, P. R.; Wagenknecht, H.-A. *Org. Biomol. Chem.* **2013,** *11,* 7458–7462. doi:10.1039/c3ob41717d
30. Walter, H.-K.; Bohländer, P. R.; Wagenknecht, H.-A. *ChemistryOpen* **2015,** *4,* 92–96. doi:10.1002/open.201402137
31. Barrois, S.; Wörner, S.; Wagenknecht, H.-A. *Photochem. Photobiol. Sci.* **2014,** *13,* 1126–1129. doi:10.1039/C4PP00153B

32. Dioubankova, N. N.; Malakhov, A. D.; Stetsenko, D. A.; Gait, M. J.; Volynsky, P. E.; Efremov, R. G.; Korshun, V. A. *ChemBioChem* **2003,** *4,* 841–847. doi:10.1002/cbic.200300678

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Versatile synthesis of the signaling peptide glorin

Robert Barnett[1], Daniel Raszkowski[2], Thomas Winckler[2] and Pierre Stallforth[*1]

Full Research Paper

Open Access

Address:
[1]Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute – HKI, Junior Research Group Chemistry of Microbial Communication, Beutenbergstr. 11, D-07745 Jena, Germany and [2]Faculty of Biology and Pharmacy, Institute of Pharmacy, Department of Pharmaceutical Biology, University of Jena, Semmelweisstrasse 10, D-07743 Jena, Germany

Email:
Pierre Stallforth* - pierre.stallforth@leibniz-hki.de

* Corresponding author

Keywords:
Dictyostelium; glorin; multicellularity; *Polysphondylium*; signaling molecules; social amoebae

Beilstein J. Org. Chem. **2017,** *13,* 247–250.
doi:10.3762/bjoc.13.27

Received: 29 September 2016
Accepted: 27 January 2017
Published: 08 February 2017

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

Guest Editor: H. B. Bode

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

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 benzyloxycarbonyl-

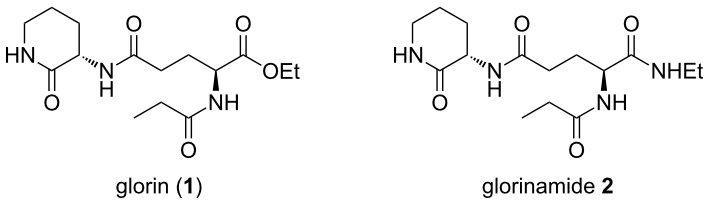
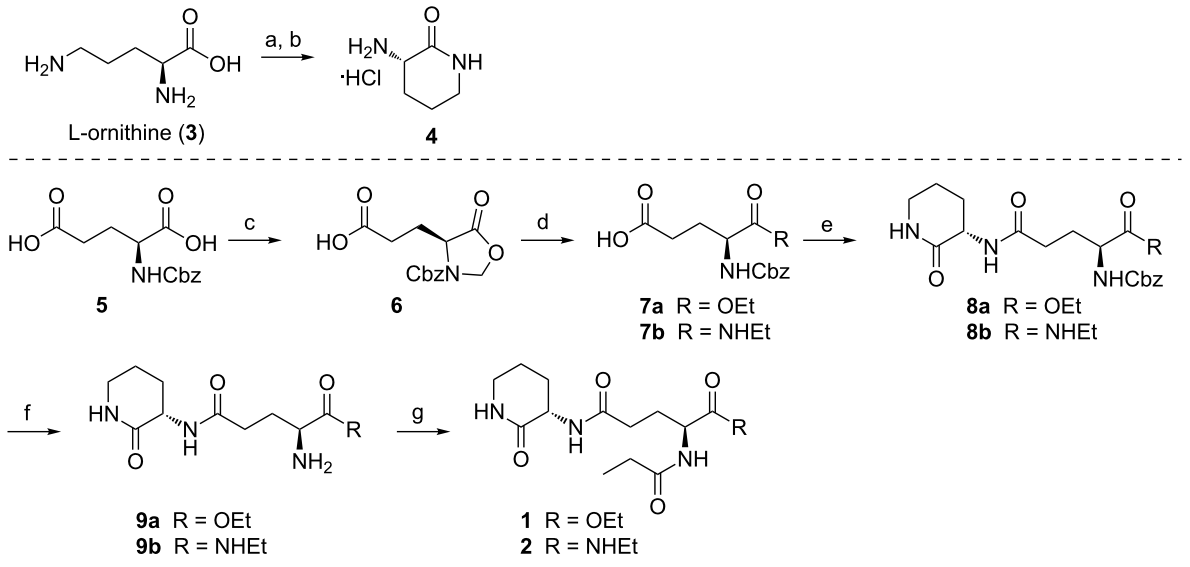


Figure 1: Glorin (**1**) and glorinamide **2**.



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) H_2CO, *p*-TsOH, toluene, Dean–Stark conditions, 3 h, 76%; d) for **7a**: NaOEt, EtOH, 0 °C to rt, 30 min, 76%, for **7b**: H_2NEt, THF, rt, 16 h, 75%; e) for **8a**: isobutyl chloroformate, NMM, **4**, DMF, −15 °C to rt, 2 h, 69%, for **8b**: HBTU, Et_3N, **4**, DMSO, rt, 3 h, 69%; f) Pd/C, H_2, MeOH, rt, 1 h, 74% **9a**, 97% **9b**; g) iPr_2EtN, 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).

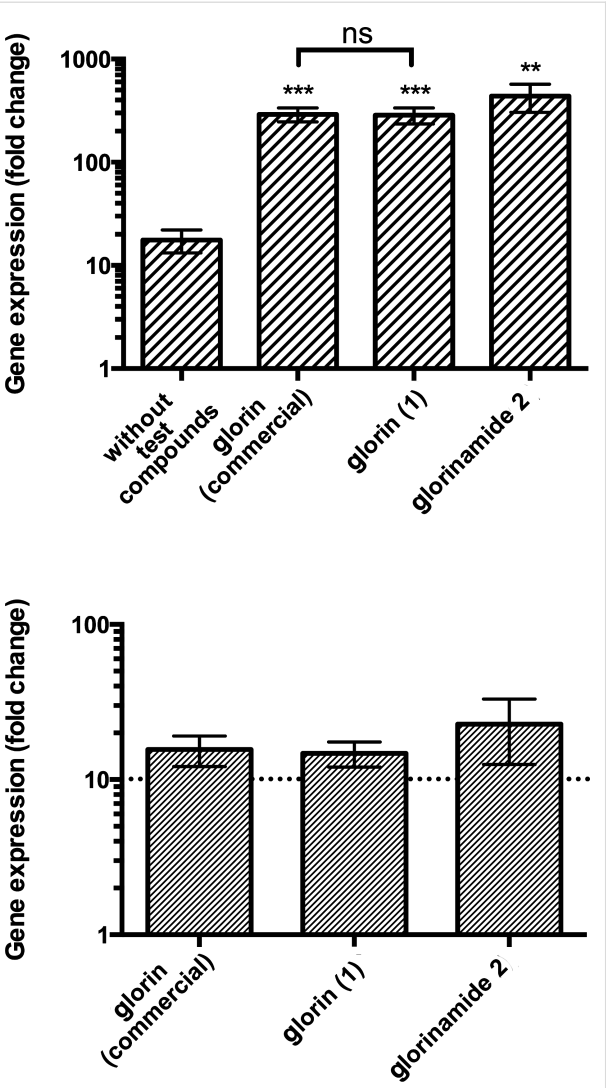


Figure 2: Gene induction in *P. pallidum* (model gene *PPL_09347*) without test compounds (negative control), commercial glorin (positive control), synthetic glorin (**1**) and glorinamide **2**. Top: Gene expression is presented as 'fold change' of gene expression in stimulated vs vegetatively-growing cells. Data are means of three biological replicates ± S.D. **: $p < 0.01$; ***: $p < 0.001$ vs vegetative cells (Student's t-test). Bottom: Gene expression presented as 'fold change' of glorin-stimulated vs unstimulated cells. Data are means of five biological replicates ± S.D. and GraphPad Prism 6 was used for visualization of the results.

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 (**1**), 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

1. Grosberg, R. K.; Strathmann, R. R. *Annu. Rev. Ecol. Evol. Syst.* **2007,** *38,* 621–654. doi:10.1146/annurev.ecolsys.36.102403.114735
2. Chisholm, R. L.; Firtel, R. A. *Nat. Rev. Mol. Cell Biol.* **2004,** *5,* 531–541. doi:10.1038/nrm1427
3. Morris, H. R.; Taylor, G. W.; Masento, M. S.; Jermyn, K. A.; Kay, R. R. *Nature* **1987,** *328,* 811–814. doi:10.1038/328811a0
4. Weijer, C. J. *Curr. Opin. Genet. Dev.* **2004,** *14,* 392–398. doi:10.1016/j.gde.2004.06.006
5. Konijn, T. M.; van de Meene, J. G. C.; Bonner, J. T.; Barkley, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **1967,** *58,* 1152–1154. doi:10.1073/pnas.58.3.1152
6. Klapper, M.; Götze, S.; Barnett, R.; Willing, K.; Stallforth, P. *Angew. Chem., Int. Ed.* **2016,** *55,* 8944–8947. doi:10.1002/anie.201603312
7. Stallforth, P.; Brock, D. A.; Cantley, A. M.; Tian, X.; Queller, D. C.; Strassmann, J. E.; Clardy, J. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 14528–14533. doi:10.1073/pnas.1308199110
8. Jousset, A. *Environ. Microbiol.* **2012,** *14,* 1830–1843. doi:10.1111/j.1462-2920.2011.02627.x
9. Ballestriero, F.; Daim, M.; Penesyan, A.; Nappi, J.; Schleheck, D.; Bazzicalupo, P.; Di Schiavi, E.; Egan, S. *PLoS One* **2014,** *9,* e109201. doi:10.1371/journal.pone.0109201
10. Mazzola, M.; de Bruijn, I.; Cohen, M. F.; Raaijmakers, J. M. *Appl. Environ. Microbiol.* **2009,** *75,* 6804–6811. doi:10.1128/AEM.01272-09
11. Asghar, A.; Groth, M.; Siol, O.; Gaube, F.; Enzensperger, C.; Glöckner, G.; Winckler, T. *Protist* **2012,** *163,* 25–37. doi:10.1016/j.protis.2010.12.002
12. Waddell, D. R. *Nature* **1982,** *298,* 464–466. doi:10.1038/298464a0
13. Shimomura, O.; Suthers, H. L.; Bonner, J. T. *Proc. Natl. Acad. Sci. U. S. A.* **1982,** *79,* 7376–7379. doi:10.1073/pnas.79.23.7376
14. Nogrady, T.; Weaver, D. F. *Medicinal chemistry: a molecular and biochemical approach,* 3rd ed.; Oxford University Press: New York, 2005.
15. Lee, Y. S.; Lee, S. W.; Park, W. J. *Bull. Korean Chem. Soc.* **1987,** *8,* 58–59.
16. Ball, J. B.; Craik, D. J.; Alewood, P. F.; Morrison, S.; Andrews, P. R.; Nicholls, I. A. *Aust. J. Chem.* **1989,** *42,* 2171–2180. doi:10.1071/CH9892171
17. Li, J.; Sha, Y. *Molecules* **2008,** *13,* 1111–1119. doi:10.3390/molecules13051111
18. Silva, L. L.; Joussef, A. C. *J. Nat. Prod.* **2011,** *74,* 1531–1534. doi:10.1021/np200234e
19. Lucas, J.; Bilzer, A.; Moll, L.; Zündorf, I.; Dingermann, T.; Eichinger, L.; Siol, O.; Winckler, T. *PLoS One* **2009,** *4,* e5012. doi:10.1371/journal.pone.0005012

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

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

Bemba Sidi Mohamed, Christian Périgaud and Christophe Mathé*

Full Research Paper

Open Access

Address:
Institut des Biomolécules Max Mousseron (IBMM), UMR 5247, Université de Montpellier, CNRS, ENSCM, cc 1705, Site Triolet, Place Eugène Bataillon, 34095 Montpellier cedex 5, France

Email:
Christophe Mathé* - christophe.mathe@umontpellier.fr

* Corresponding author

Keywords:
analogue; antiviral; carbocyclic; nucleoside; phosphonate

Beilstein J. Org. Chem. **2017,** *13,* 251–256.
doi:10.3762/bjoc.13.28

Received: 25 November 2016
Accepted: 31 January 2017
Published: 09 February 2017

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

Guest Editor: H. B. Bode

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

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.

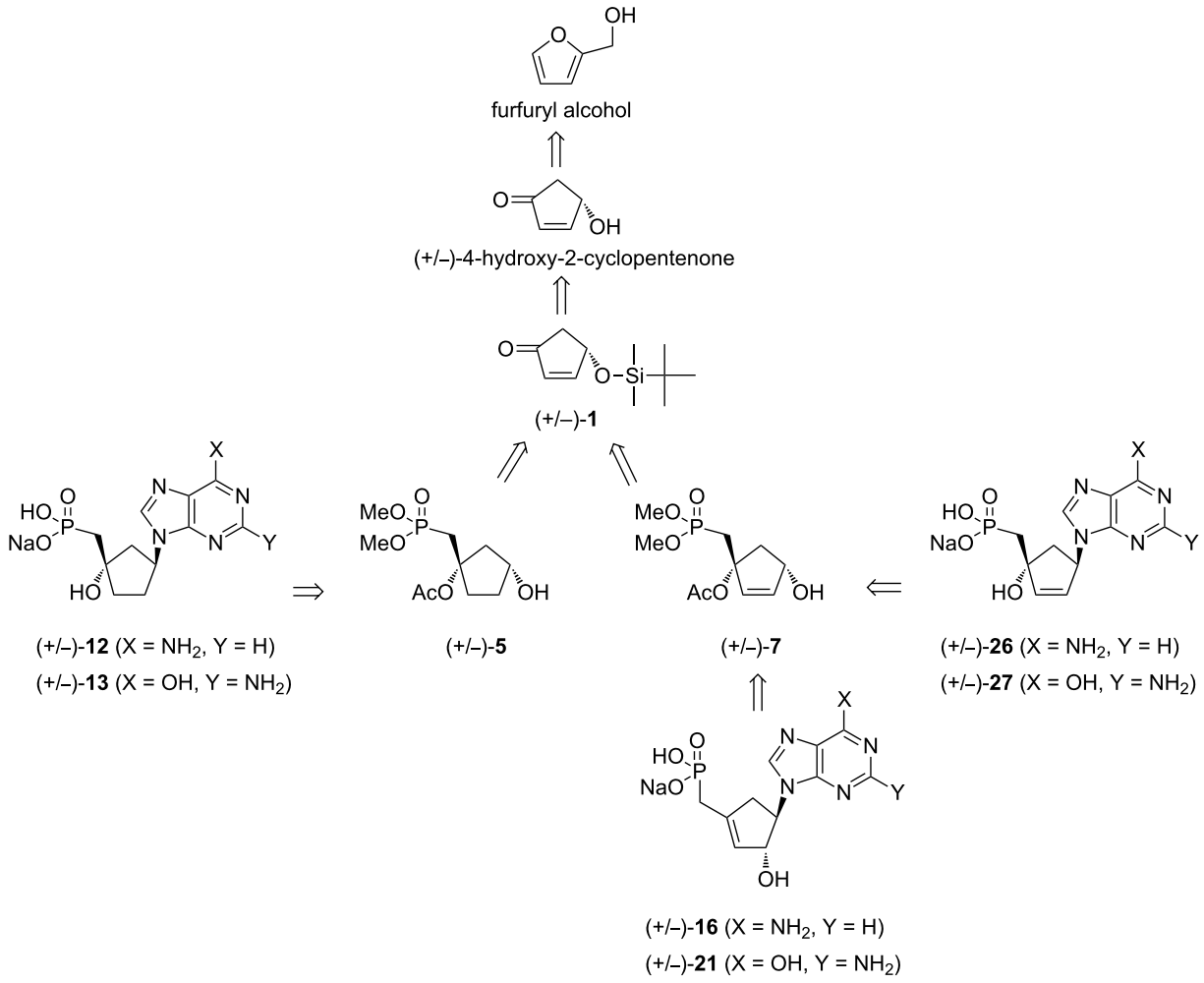


Figure 1: Retrosynthetic pathway for the synthesis of the target carbocyclic nucleoside methylphosphonates.

Results and Discussion

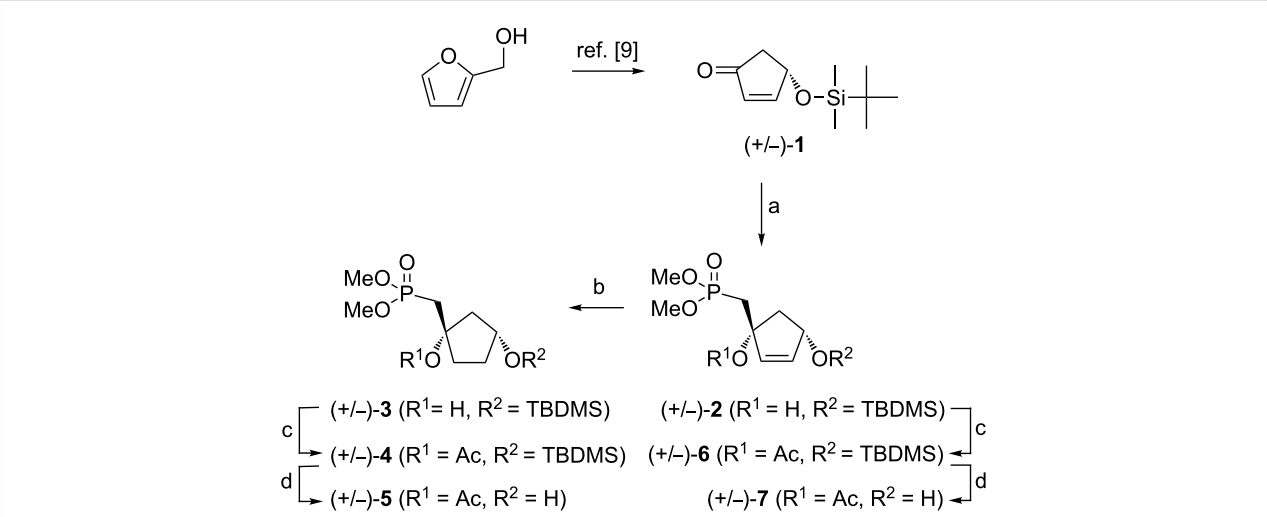
Synthesis of precursors (+/−)-**5** and (+/−)-**7**

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. ^{1}H and ^{13}C NMR spectra were in accordance with the presence of a sole diastereoisomer.

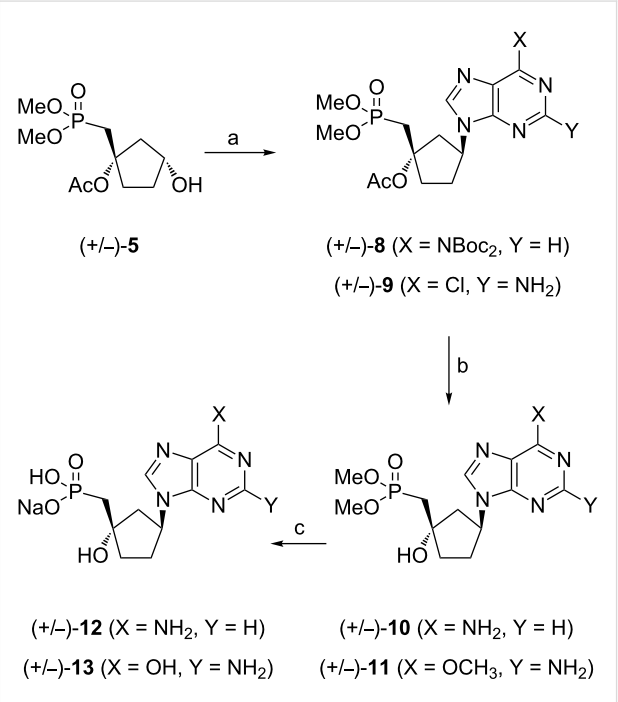
The reduction of compound (+/−)-**2** readily provided the saturated derivative (+/−)-**3** and the free tertiary hydroxy groups in (+/−)-**2** and (+/−)-**3** were then protected using acetic anhydride. Finally, removal of the TBDMS groups afforded the appropriate carbocyclic precursors, namely, (+/−)-1-((dimethoxyphosphoryl)methyl)-3-hydroxycyclopentyl 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 diisopropyl azodicarboxylate (DIAD) and PPh_3, provided the N9 carbocyclic nucleosides (+/−)-**8** and (+/−)-**9** as racemic mixtures. No concomitant formation of the N7 regioisomer was observed.



Scheme 1: Reagents and conditions: (a) $(CH_3O)_2P(O)CH_3$, *n*-BuLi, THF, −78 °C/rt, 2 h, 63%; (b) H_2, Pd/C, MeOH, rt, 96%; (c) Ac_2O, Et_3N, DMAP, Et_2O, rt, 88% (**4**), 88% (**6**); (d) TBAF (1 M), THF, 0 °C, 2 h, 85% (**5**), 76% (**7**).



Scheme 2: Reagents and conditions: (a) N^6-bis-Boc-adenine or 2-amino-6-chloropurine, PPh_3, DIAD, THF, 0 °C to rt, 42% (**8**), 54% (**9**); (b) for **10**: i) **8**, TFA, $Cl(CH_2)_2Cl$, rt; ii) K_2CO_3, MeOH, rt, 60%; for **11**: **9**, K_2CO_3, MeOH, rt, 68%; (c) for **12**: **10**, TMSCl, NaI, CH_3CN, DMF, 40 °C to rt, 61%; for **13**: **11**, TMSBr, DMF, 0 °C to rt, 35%.

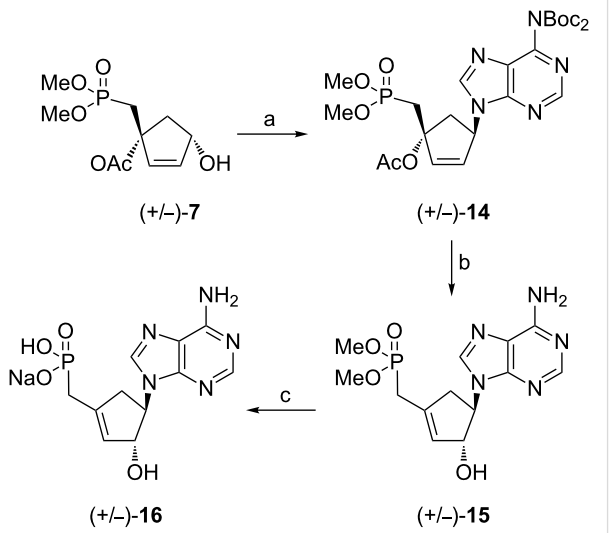
The removal of Boc and acetyl groups from (+/−)-**8** afforded the carbocyclic phosphonate (+/−)-**10**. Subsequently, the phosphonoester protecting groups were cleaved in the presence of TMSCl and NaI in a CH_3CN/DMF mixture to give (+/−)-**12** in 61% yield. The treatment of (+/−)-**9** with K_2CO_3 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 ^{13}C 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) N^6-bis-Boc-adenine, PPh_3, DIAD, THF, rt, 56%; (b) TFA, $Cl(CH_2)_2Cl$, rt, 85%; (c) TMSCl, NaI, CH_3CN, 40 °C to rt, 60%.

pound (+/−)-**14**, the chemical shifts of C2' and C3' are consistent with sp^2-hybridized carbons while for compound (+/−)-**16**, chemical shifts corresponding to sp^2 carbons were detected for C2' and C1' (Figure 2). Furthermore an upfield shift of the signals for C2' was also observed for compounds (+/−)-**14** and (+/−)-**16**, respectively. Inversely, a downfield shift for C1' was observed for compound (+/−)-**14** compared to (+/−)-**16**. These observations are in agreement with a 1,3-allylic transposition under acidic conditions.

Table 1: Selected chemical shifts in ^{13}C NMR.[a]

	(+/−)-**14**	(+/−)-**16**
C3'	134.5	81.5
C2'	138.2	125.6
C1'	88.7	143.0

[a]δ in ppm of C2', C3' and C1'.

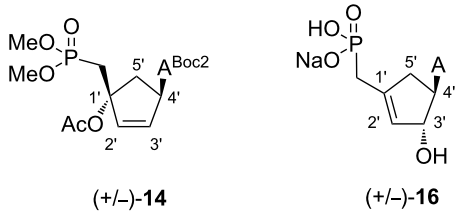


Figure 2: Numbering for **14** and **16**.

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.

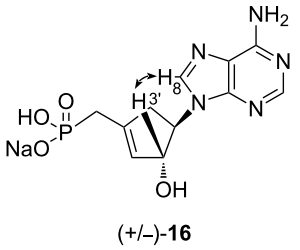
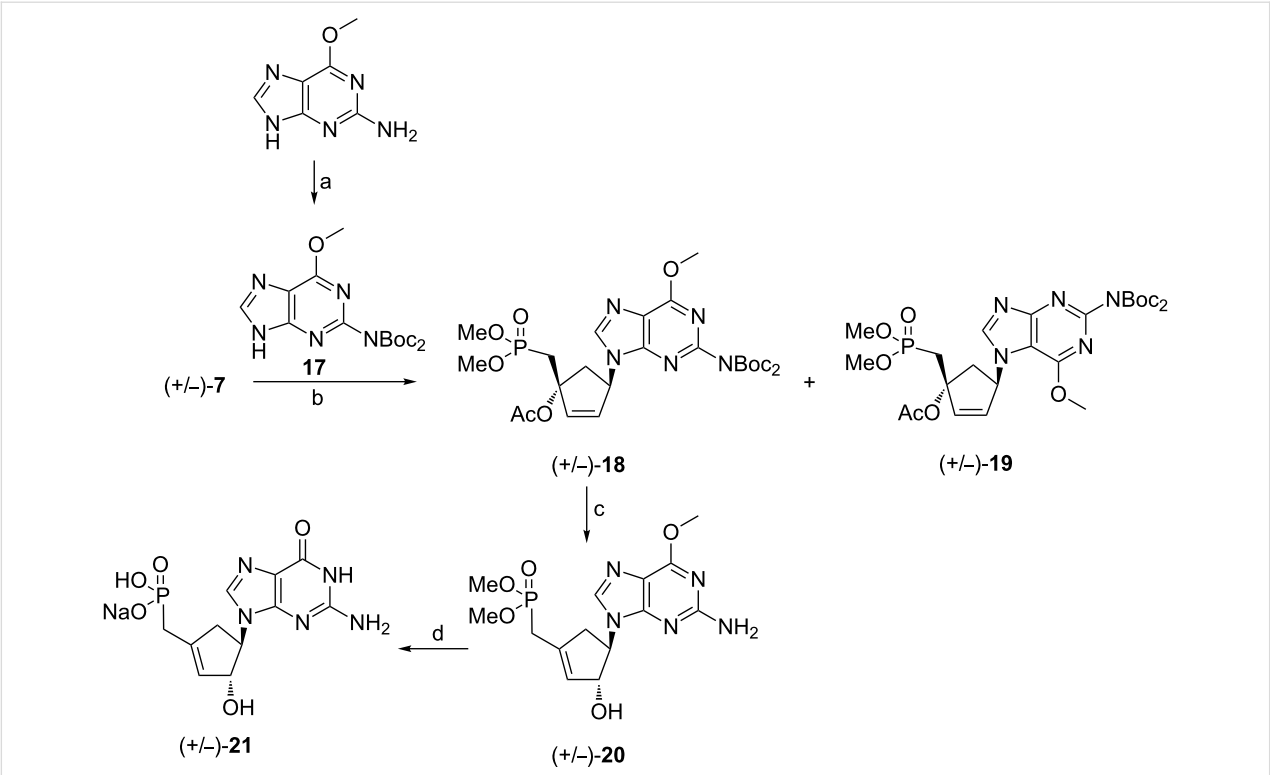


Figure 3: Selected NOESY correlations for compound (+/−)-**16**.

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 Boc_2O 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 assignments were based upon ^1H 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].

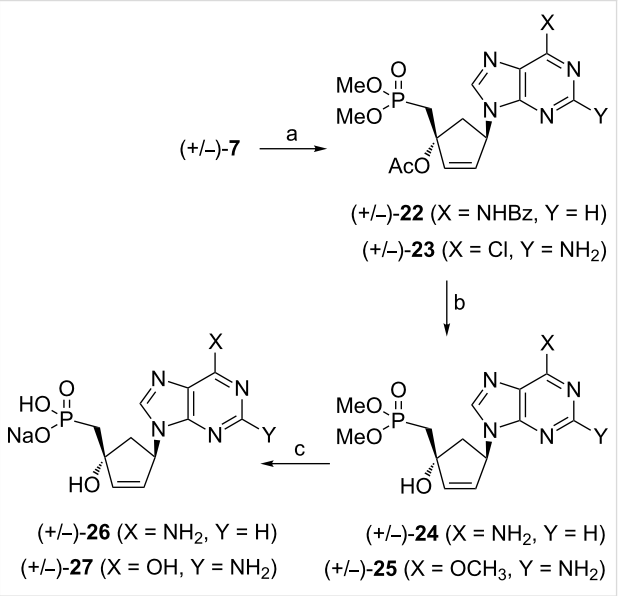
Compounds (+/−)-**26** and (+/−)-**27**

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 N^6-Bz-adenine [14] or 2-amino-6-chloropurine in the presence of PPh_3 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 N^6-Bz-adenine with (+/−)-**7**



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%.

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 5: Reagents and conditions: (a) N^6-Bz-adenine or 2-amino-6-chloropurine, PPh_3, DIAD, THF, 0 °C to rt, 17% for **22**, 53% for **23**; (b) K_2CO_3, MeOH, rt, 34% for **24**, 72% for **25**; (c) TMSBr, DMF, 0 °C to rt, 12% for **26**, 17% for **27**.

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

1. Corma, A.; Iborra, S.; Velty, A. *Chem. Rev.* **2007,** *107,* 2411. doi:10.1021/cr050989d
2. Halilu, A.; Ali, T. H.; Atta, A. Y.; Sudarsanam, P.; Bhargava, S. K.; Hamid, S. B. A. *Energy Fuels* **2016,** *30,* 2216. doi:10.1021/acs.energyfuels.5b02826
3. Roche, S. P.; Aitken, D. J. *Eur. J. Org. Chem.* **2010,** 5339. doi:10.1002/ejoc.201000704
4. Mantione, D.; Aizpuru, O. O.; Memeo, M. G.; Bovio, B.; Quadrelli, P. *Eur. J. Org. Chem.* **2016,** 983. doi:10.1002/ejoc.201501406
5. Ulbrich, K.; Kreitmeier, P.; Vilaivan, T.; Reiser, O. *J. Org. Chem.* **2013,** *78,* 4202. doi:10.1021/jo400409f
6. Uttaro, J.-P.; Broussous, S.; Mathé, C.; Périgaud, C. *Tetrahedron* **2013,** *69,* 2131. doi:10.1016/j.tet.2013.01.011
7. Boojamra, C. G.; Parrish, J. P.; Sperandio, D.; Gao, Y.; Petrakovsky, O. V.; Lee, S. K.; Markevitch, D. Y.; Vela, J. E.; Laflamme, G.; Chen, J. M.; Ray, A. S.; Barron, A. C.; Sparacino, M. L.; Desai, M. C.; Kim, C. U.; Cihlar, T.; Mackman, R. L. *Bioorg. Med. Chem.* **2009,** *17,* 1739. doi:10.1016/j.bmc.2008.12.028
8. Macchi, B.; Romeo, G.; Chiacchio, U.; Frezza, G.; Giofrè, S. V.; Marino-Merlo, F.; Mastino, A. *Top. Med. Chem.* **2015,** *15,* 53. doi:10.1007/7355_2013_28
9. Curran, T. T.; Hay, D. A.; Koegel, C. P.; Evans, J. C. *Tetrahedron* **1997,** *53,* 1983. doi:10.1016/S0040-4020(96)01169-6
10. Mitsunobu, O. *Synthesis* **1981,** 1. doi:10.1055/s-1981-29317
11. Hughes, D. L. *Org. Prep. Proced. Int.* **1996,** *28,* 127. doi:10.1080/00304949609356516
12. Michel, B. Y.; Strazewski, P. *Tetrahedron* **2007,** *63,* 9836. doi:10.1016/j.tet.2007.06.100
13. Rodriguez, S.; Edmont, D.; Mathé, C.; Périgaud, C. *Tetrahedron* **2007,** *63,* 7165. doi:10.1016/j.tet.2007.04.094
14. Milecki, J.; Foldesi, A.; Fischer, A.; Adamiak, W. R.; Chatopadhyaya, J. *J. Labelled Compd. Radiopharm.* **2001,** *44,* 763. doi:10.1002/jlcr.503

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.13.28

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Polyketide stereocontrol: a study in chemical biology

Kira J. Weissman

Review

Open Access

Address:
UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, Avenue de la Forêt de Haye, BP 50184, 54505 Vandœuvre-lès-Nancy Cedex, France

Email:
Kira J. Weissman - kira.weissman@univ-lorraine.fr

Keywords:
chemical biology; polyketide synthases; reduced polyketides; stereocontrol

Beilstein J. Org. Chem. **2017,** *13,* 348–371.
doi:10.3762/bjoc.13.39

Received: 04 November 2016
Accepted: 01 February 2017
Published: 24 February 2017

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

Guest Editor: H. B. Bode

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

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 anti-infectives [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]).

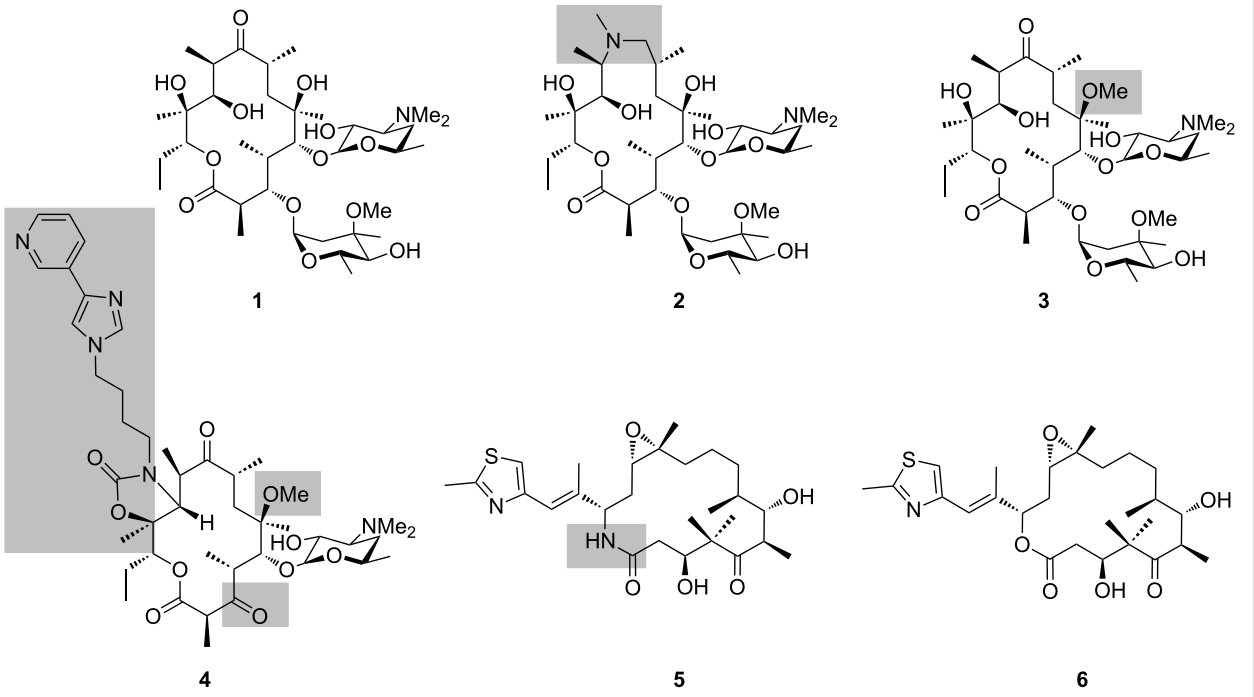


Figure 1: Structures of clinically-relevant polyketides: erythromycin A (**1**), azithromycin (**2**), clarithromycin (**3**), telithromycin (**4**), and ixabepilone (**5**) (a semi-synthetic derivative of epothilone B (**6**)). The structural variations relative to the parent compounds **1** and **6** are indicated by the grey boxes.

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

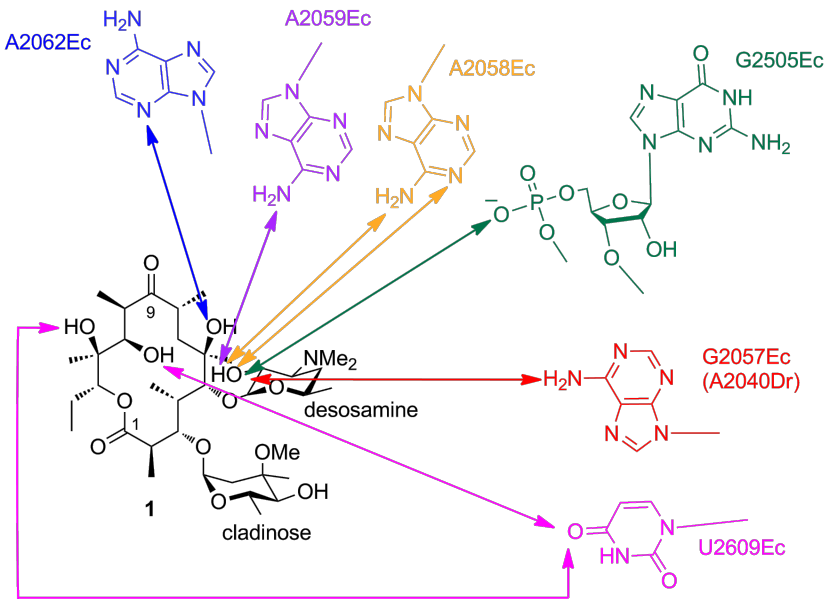


Figure 2: Schematic of erythromycin A (**1**) bound to 23S ribosomal RNA of the 50S subunit of the *Deinococcus radiodurans* (Dr) ribosome. The interactions between the polyketide and the nucleotides (*Escherichia coli* (Ec) numbering) are indicated with colored arrows (reactive groups are less than 4.4 Å apart). Adapted from [5].

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

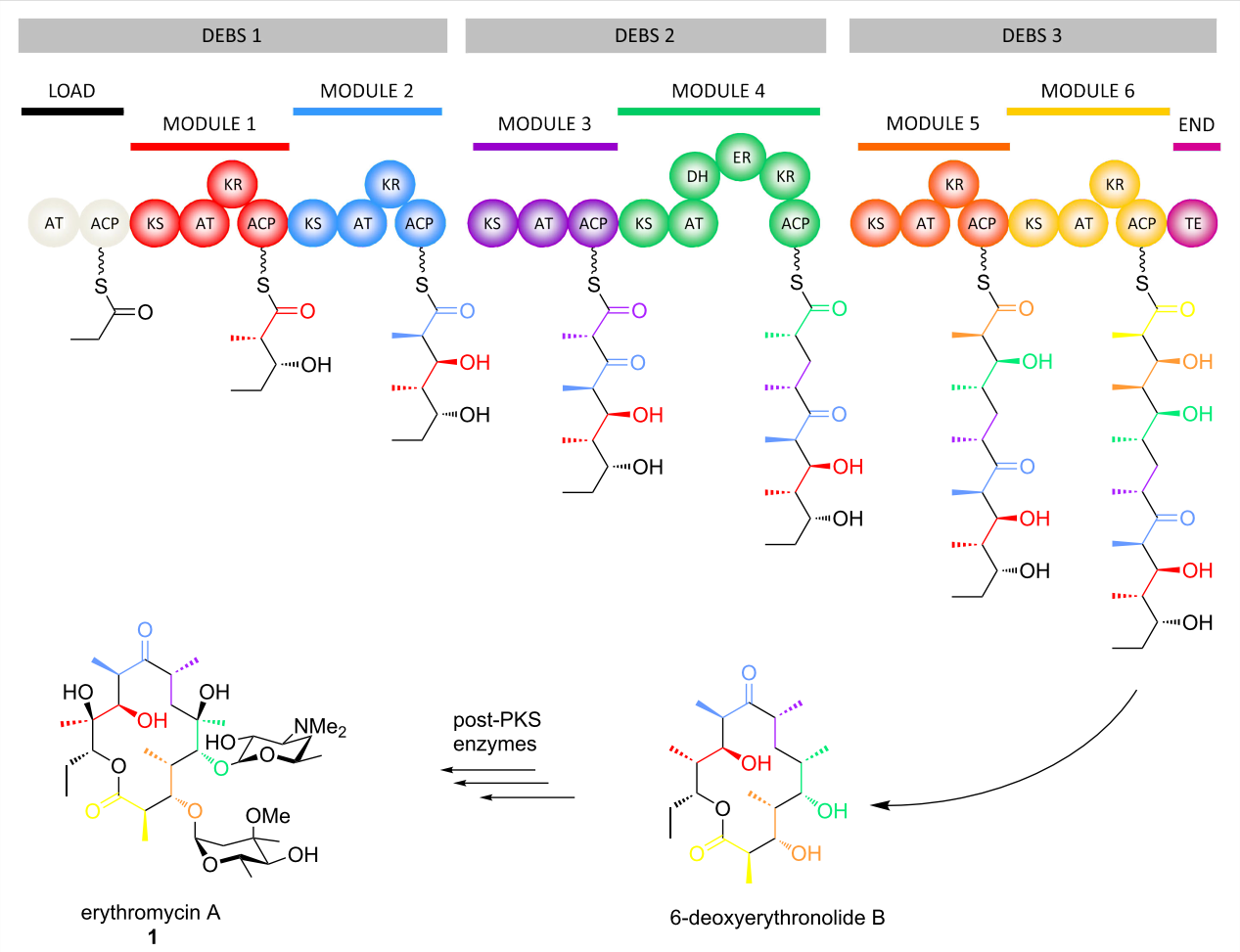


Figure 3: Schematic of the biosynthetic pathway leading to erythromycin A (**1**) in the bacterium *Saccharopolyspora erythraea*. The first stage of biosynthesis occurs on a modular polyketide synthase (PKS) incorporating three gigantic multienzyme polypeptides, DEBS 1, 2 and 3. Each of these subunits 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.

contrast, the majority of *trans*-ATs operating in *trans*-AT PKSs are specific for malonyl-CoA, although exceptions do exist (such as the ethylmalonyl-CoA-specific AT from kirromycin biosynthesis) [17]; C-2-methyl groups in these systems are thus introduced primarily by methyl transferase domains [18], with presumably defined the stereospecificity (the stereochemistry is not always evident, as it can be obscured by subsequent dehydration). The suite of processing reactions also introduces stereochemistry into the molecules: the hydroxy groups resulting from ketoreduction of the initially-formed C-3-ketones exhibit both configurations, dehydration of the hydroxy functionality generates both *cis*- and *trans*-double bonds, and finally, enoyl reduction can produce both configurations at the saturated C2-methyl centers. Other types of processing reactions present in *trans*-AT PKSs and certain *cis*-AT PKSs (for example, pyran synthase domains [19,20], double bond shifting modules [21,22], C-2-hydroxylases [11], etc.) can also have stereochemical consequences, but these will not be treated here as little is known to date about the enzymatic factors controlling the configurational outcomes. Finally, where chirality is introduced, post-PKS processing reactions also proceed with defined stereochemistry, although this aspect will also not be discussed in this article. The following sections will address the role of each of the principal PKS domains in controlling these stereochemical features, highlighting in each case the contribution of chemical biology in illuminating enzymatic function.

Acyl transferases

Pathways to both the (2*R*)- and (2*S*)-isomers of methymalonyl-CoA exist in bacterial cells, and so in principle, the observed

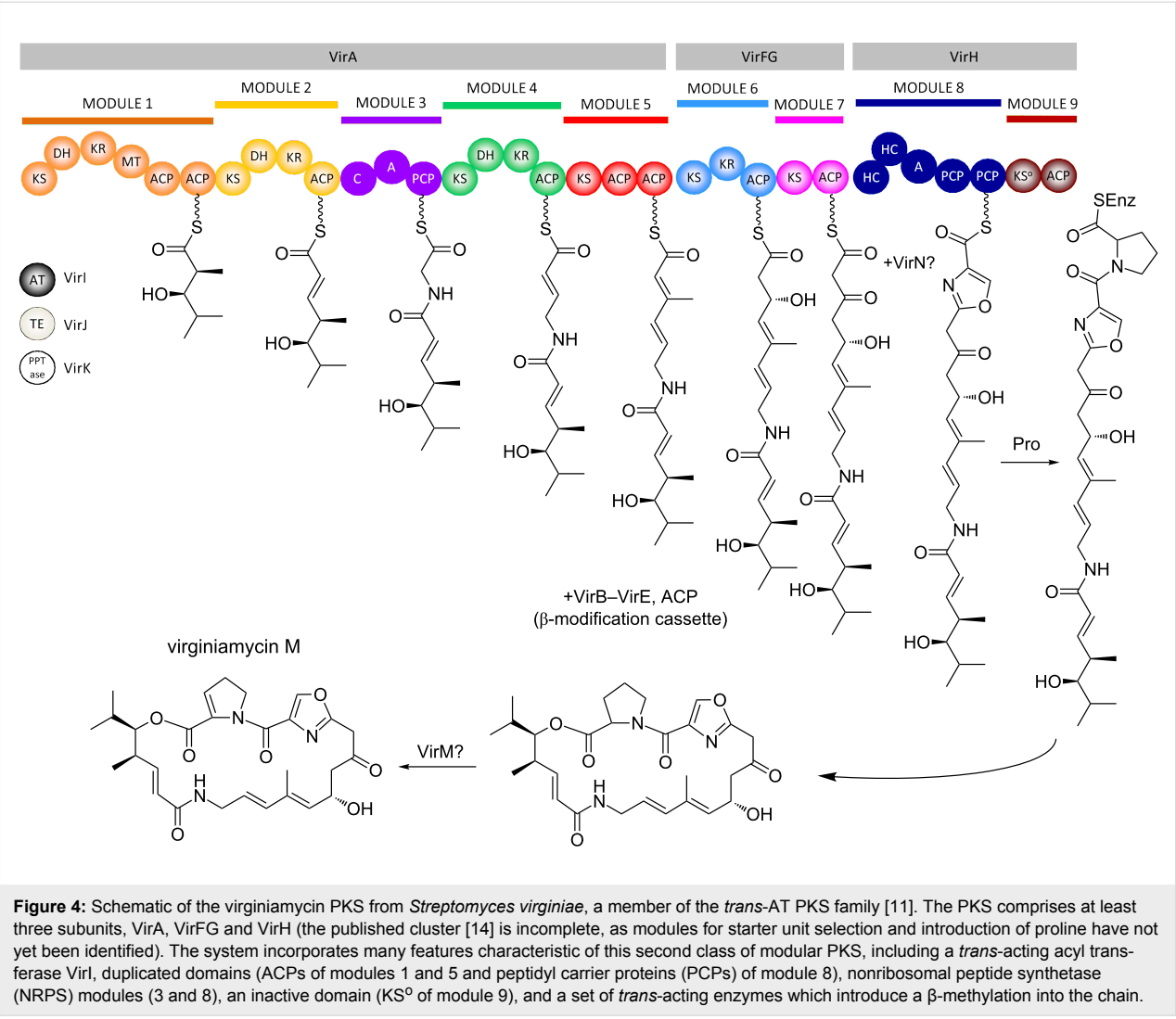


Figure 4: Schematic of the virginiamycin PKS from *Streptomyces virginiae*, a member of the *trans*-AT PKS family [11]. The PKS comprises at least three subunits, VirA, VirFG and VirH (the published cluster [14] is incomplete, as modules for starter unit selection and introduction of proline have not yet been identified). The system incorporates many features characteristic of this second class of modular PKS, including a *trans*-acting acyl transferase VirI, duplicated domains (ACPs of modules 1 and 5 and peptidyl carrier proteins (PCPs) of module 8), nonribosomal peptide synthetase (NRPS) modules (3 and 8), an inactive domain (KS^0 of module 9), and a set of *trans*-acting enzymes which introduce a β-methylation into the chain.

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 (2*R*)- and (2*S*)-methylmalonyl-CoA in situ [23]. When a precursor of (2*S*)-methylmalonyl-CoA, [2-$^2H_2$,2-^{13}C]propionate, was used, analysis of the products by difference $^{13}C\{^1H,^2H\}$ 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 (2*S*)-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-$^2H_2$,2-^{13}C]-succinate to produce labeled (2*R*)-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 ^{14}C-(2*S*)- or (2*R*)-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 (2*S*)-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

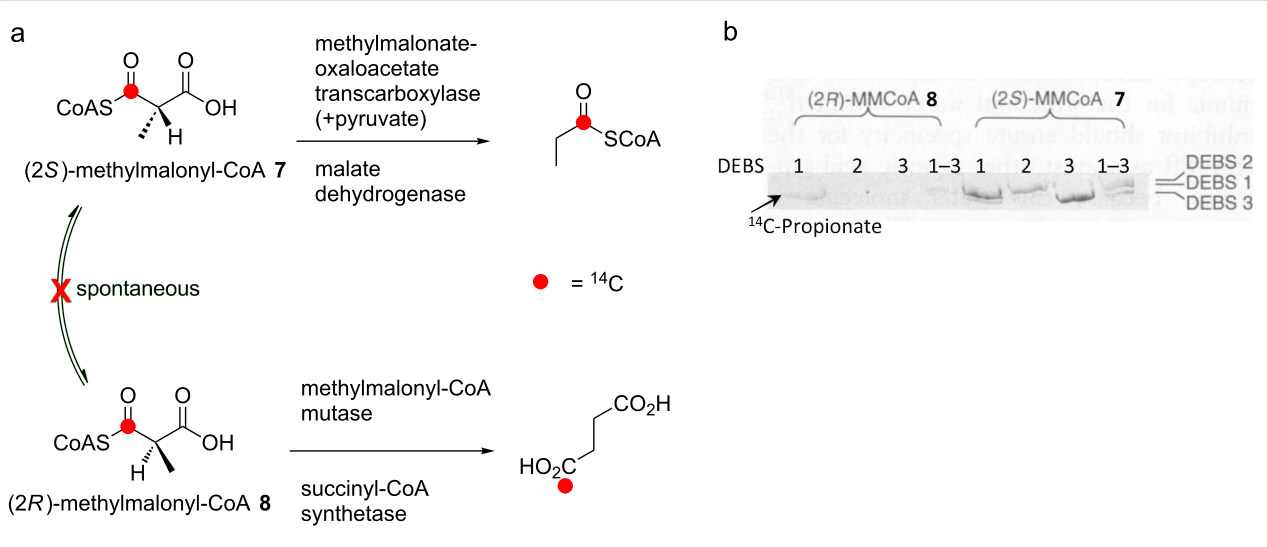


Figure 5: Determination of the stereochemistry of extender unit selection by the AT domains of modular PKS. a) Enantiomerically pure ^{14}C-labeled (2*S*)-methylmalonyl-CoA (**7**) and (2*R*)-methylmalonyl-CoA (**8**) were generated enzymatically from a racemic mixture by consumption of the opposite 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].

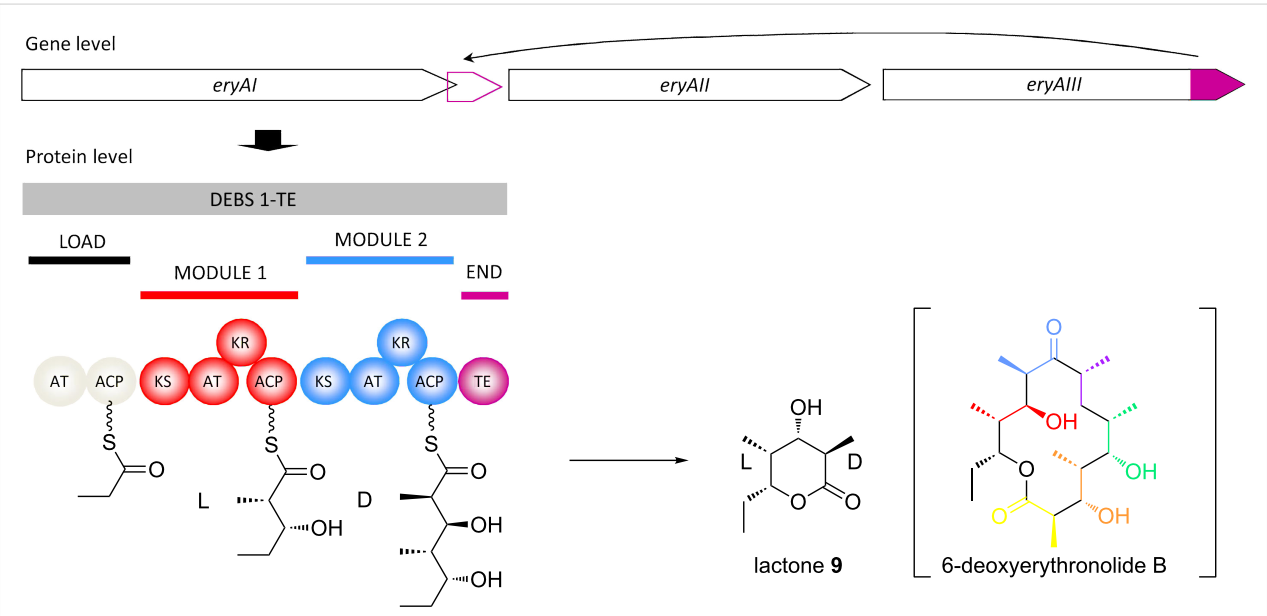


Figure 6: Creation by genetic engineering of the DEBS 1-TE model system. The region of the *eryAIII* gene encoding the thioesterase (pink) was relocated to the end of gene *eryAI*. The resulting protein, DEBS 1-TE, produces a small triketide lactone **9** instead of the heptaketide 6-deoxyerythronolide B. The two methyl centers in lactone **9** are of opposite stereochemical configuration, and thus DEBS 1-TE is an attractive protein for studying the control of stereochemistry.

polyketide at the triketide stage. This modification results in an experimentally tractable δ-lactone **9** instead of the 14-membered macrolide, 6-deoxyerythronolide B. Notably, the two methyl centers in the lactone have opposite configurations (for clarity, that at C-2 will be referred to as D-configured, and that at C-4 as L), 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 (2*R*)-methylmalonyl-CoA by the analyzed modules and the second from use of the (2*S*)-isomer, was now firmly excluded. Given the high level of homology among many AT domains from *cis*-AT 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 (2*S*)-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 AT_5 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 AT_5 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 AT_5 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 (2*S*)-isomer appears to lie in steric clashes that would occur between a (2*R*)-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.

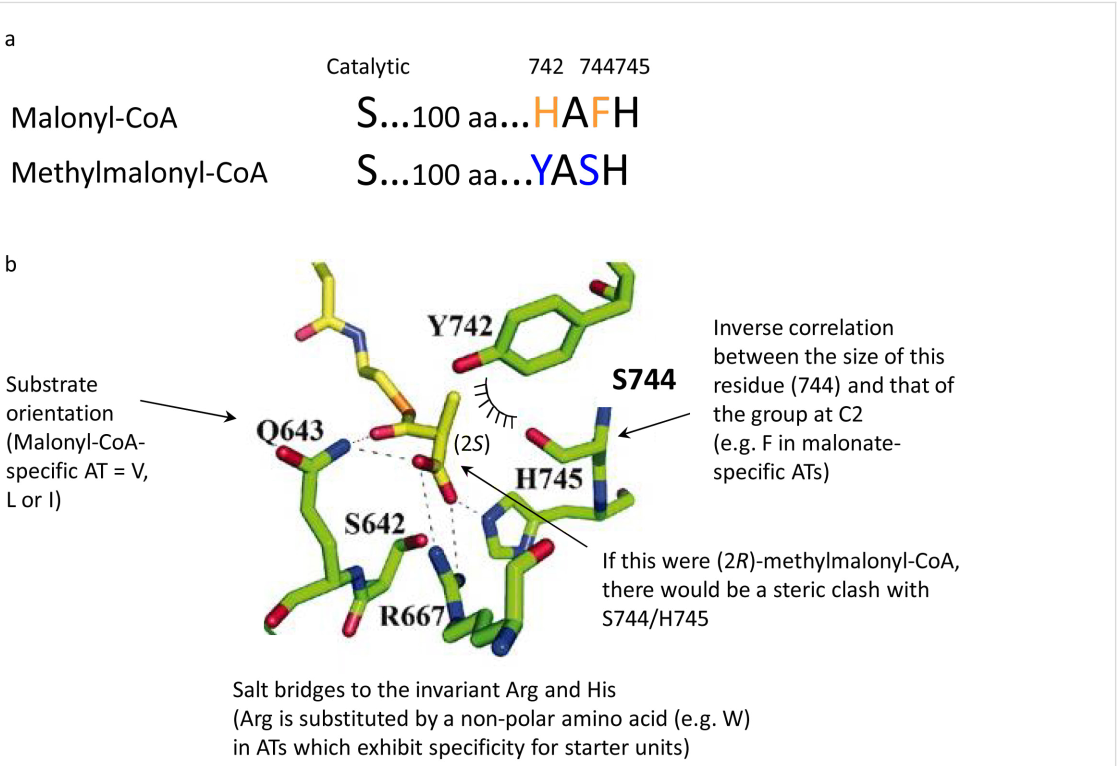


Figure 7: Model for substrate selection by AT domains. a) Sequence motifs in malonyl- and methylmalonyl-CoA-specific ATs which correlate with substrate choice. A HAFH motif is present some 100 amino acids downstream of the active site serine in malonyl-CoA-specific AT domains, while the corresponding sequence is YASH in methylmalonyl-CoA-specific ATs. The numbering is as in DEBS 3 (domain AT_5). b) Model for the molecular basis of specificity for (2*S*)-methylmalonyl-CoA based on the crystal structure of DEBS AT_5 solved in the presence of acetate [38]. This shows notably the proposed role of the Y, S and H residues of the conserved recognition motif. Reprinted with permission from [38]. Copyright 2006, National Academy of Sciences.

In terms of the stereochemistry of the less common extender units, labeling studies indicate that the (2*S*) 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 (2*S*)-isomer of aminomalonyl-ACP and the (2*R*)-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 (2*S*)-isomer would appear to fit better within the investigated AT structure. Indeed, selection of the (2*S*)-isomer, and correspondingly, the (2*R*) 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), (2*RS*)-[2-^{2}H]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 (2*S*)- 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 (2*RS*)-methylmalonyl-CoA in D_2O.

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 (2*RS*)-[2-^{2}H]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 KS_1 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 KS_1 was paired with the remaining domains of DEBS module 2 (AT_2, KR_2 and ACP_2), 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 KS_1 can produce both methyl stereochemistries, but that in the hybrid 1/2 diketide synthase, the selectivity of KR_2 for the *un*epimerized methyl configuration masks the KS_1 epimerase activity. Subsequent work seemed to strengthen the idea that KS_1 acts as an epimerase [48]. In this case (Figure 10b), the loading module-KS_1 portion of DEBS 1 was grafted onto DEBS 3 (whose two modules 5 and 6 generate the unepimerized 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 KS_1 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

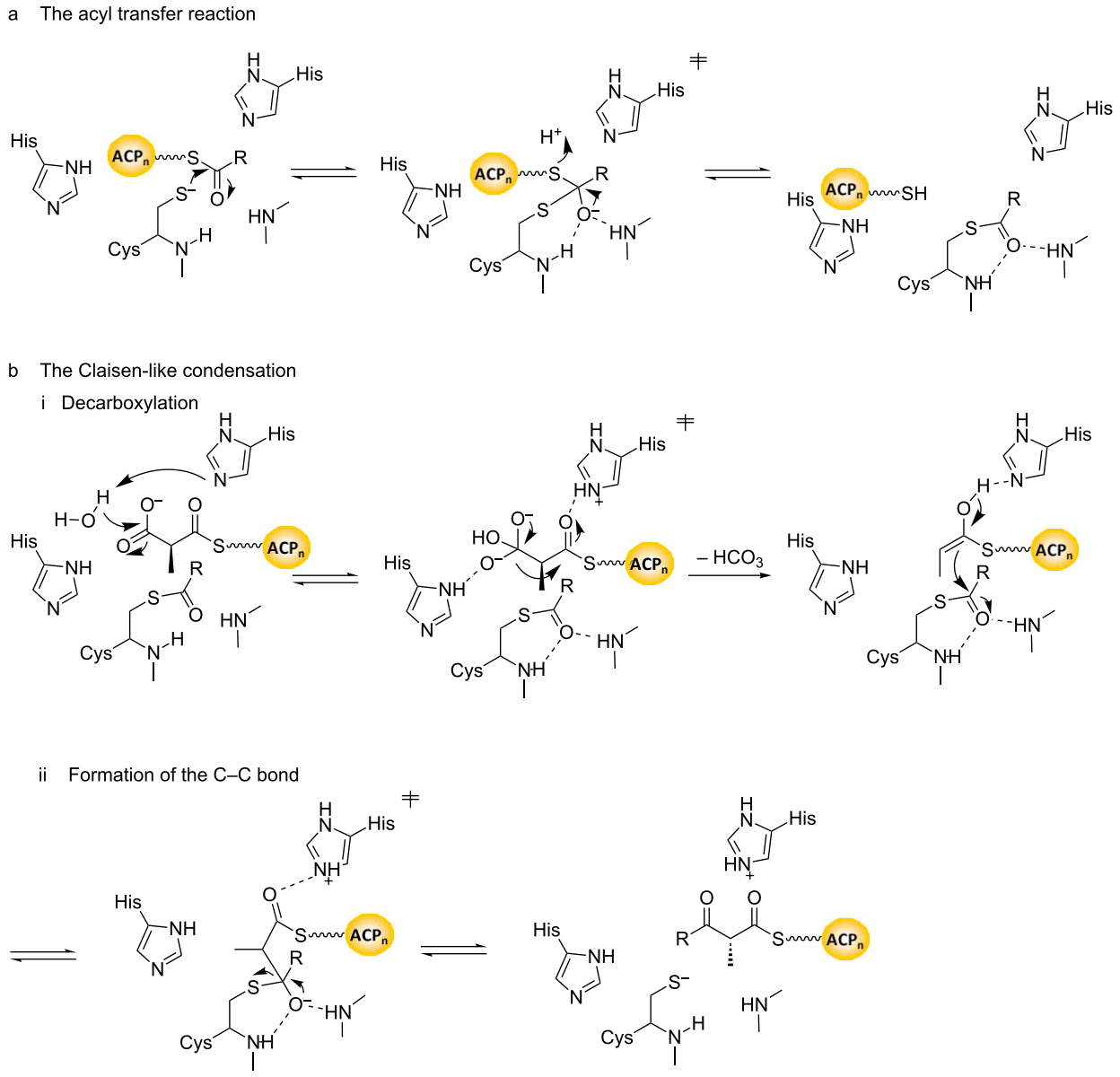
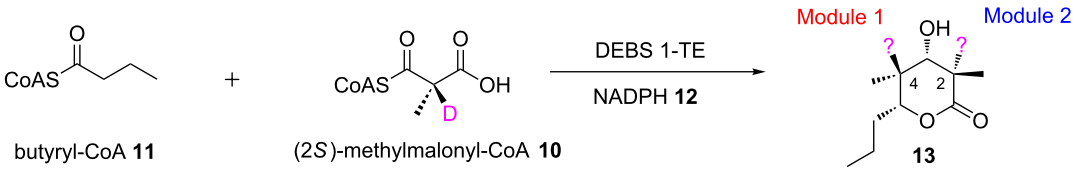


Figure 8: Proposed mechanism for KS-catalyzed chain extension, based on extrapolation from studies on homologous enzymes from animal FAS [44]. The reaction encompasses two stages overall: a) acyl transfer, and b) the Claisen-like condensation. From the stereochemical perspective, the important aspect 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 KRs were shown to participate in epimerization will be detailed below.

Ketoreductases

KR domains 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 KRs transplanted by genetic engineering into alternative contexts have maintained their native stereospecificity [49-51]. Incubation of enzymatically-generated, chirally deuterated NADPH (both (4*R*)- and (4*S*)-[4-^{2}H]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 KRs 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 KRs from modular PKS



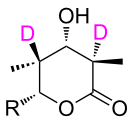
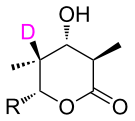
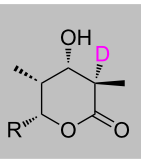
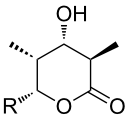
mechanism	condensation with retention	condensation with inversion	epimerization	product
I	Module 1	Module 2	none	
II	Modules 1 + 2		Module 2	
III		Modules 1 + 2	Module 1	
IV	Module 2	Module 1	Modules 1 + 2	

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.

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 ketoreduction (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

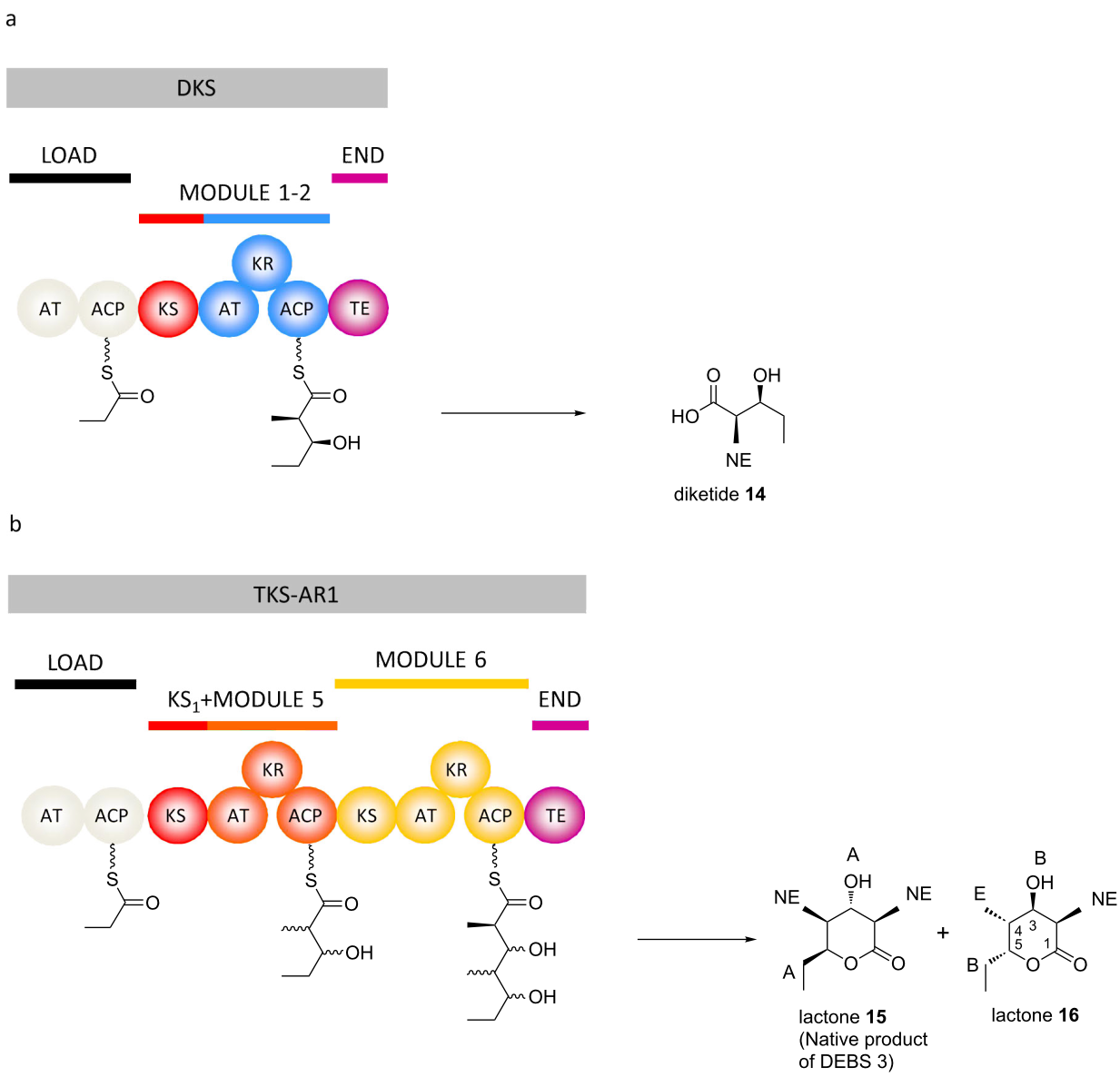


Figure 10: Genetic engineering experiments which suggested a role for the KS domain in epimerization. a) A diketide synthase (DKS) was created by attaching the loading module and KS_1 of DEBS to the remainder of DEBS module 2, which was itself fused to the thioesterase (TE) domain [47]. The resulting construct yielded diketide **14** in which the methyl group at C-2 was not epimerized (NE). As the diketide generated by module 1 normally incorporates an epimerized methyl, this result was taken as evidence that KS_1 can produce both epimerized and unepimerized methyl groups, and that the downstream KR 'chooses' which one is taken on as a substrate for reduction. b) In construct TKS-AR1, the same DEBS loading module-KS_1 region was used to replace the initial KS of DEBS 3 [48]. The resulting protein produced two lactones: lactone **15**, the native product of DEBS 3 in which no methyl epimerization has occurred (NE) and the two hydroxy groups are A-type, and lactone **16**, in which the stereochemistry at the C-4 methyl center generated by module 5 is inverted (E). The presence of this epimerized methyl causes the direction of reduction to reverse (to B-type) in both modules 5 and 6, even though the methyl center produced by module 6 (C-2) is of native, non-epimerized stereochemistry (NE).

active site. Additional amino acids in the 134–149 region, specifically P144 and N148, correlate with B-type KRs, 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 KRs was co-crystallized as a ternary complex with both native polyketide intermediate and cofactor.

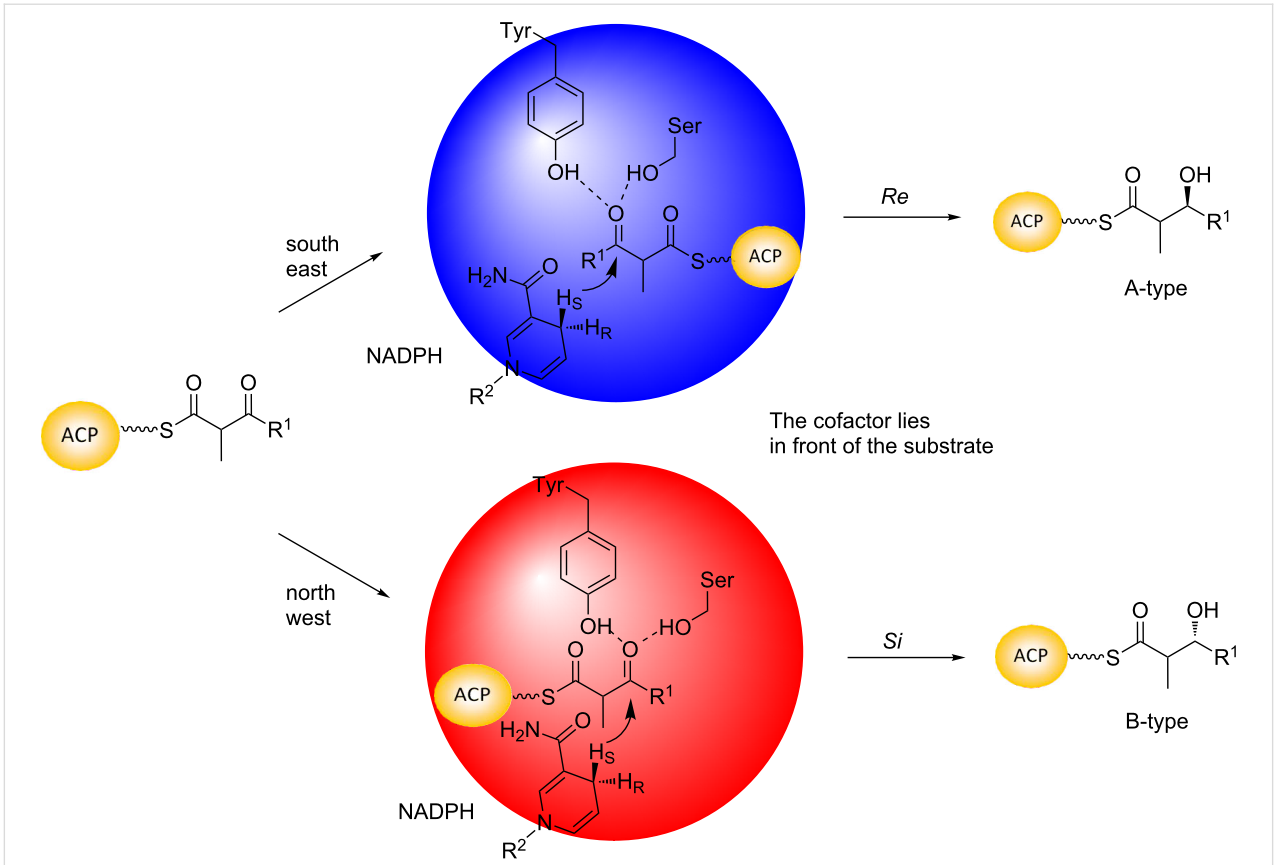


Figure 11: Models for control of the stereochemistry of reduction by KR domains. The two directions of ketoreduction achieved by a conserved catalytic apparatus (i.e. positions of the catalytic tetrad and NADPH cofactor) are obtained by entry of the ACP-bound substrate from one or the other side of the active site [68]. 'South east' entry gives reduction on the *re* face and an 'A-type' product, while the alternative 'north-west' entry yields *si* face reduction, and a 'B-type' hydroxy product.

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 α-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*RS*)-2-methyl-3-oxopentanoic acid *N*-acetylcysteamine (NAC) thioester ('β-keto diketide') **17** – a racemic analogue of the diketide generated by condensation of a propionyl starter unit and a (2*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 KRs from the DEBS [68], tylosin (Tyl) [68,72] and amphotericin [58,61] PKSs (Figure 12), showed that when the KRs 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 KRs 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 (2*RS*)-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 KR_1

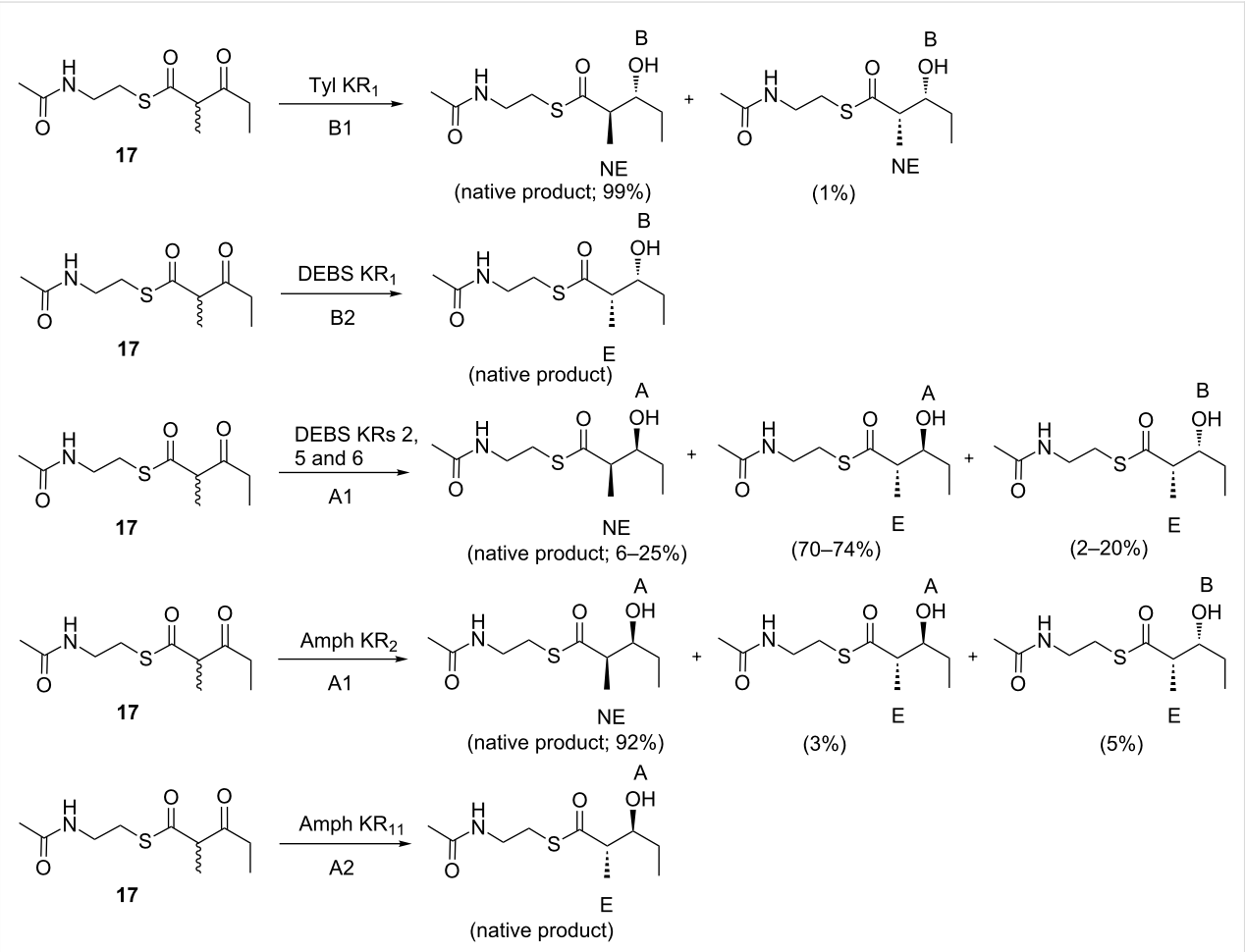


Figure 12: Assays in vitro to evaluate the stereospecificity of recombinant KR domains. A series of KR domains of all types (A1, A2, B1 and B2, as indicated) were investigated using a common substrate, racemic diketide derivatized as its *N*-acetylcysteamine (SNAC) thioester **17**. When the KR domains 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 KR_2 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 ($t_{1/2}$ = 4.7 min) [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.

KRs 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). (KRs 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 KR_1 in the presence of starter unit (synthetic propionyl-SNAC), NADPH (**12**) and (2*RS*)-methylmalonyl-CoA, resulted in diketide with the stereochemistry at both the C-2 methyl group (epimerized) and C-3-hydroxy group (B-type) associated with module 1 (Figure 13). Conversely, mixing the KS-AT and ACP domains from DEBS module 1 (which produces an epimerized methyl and a B-type C-3 hydroxy) with the DEBS module 6 KR, yielded diketide incorporating the stereochemistry 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.

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 KR_1, nystatin (Nys) KR_1 and rifamycin (Rif) KR_7) was demonstrated directly by incubating them with the stereochemically appropriate, configurationally stable reduced product obtained by chemical synthesis, in which C-2 was deuterium labeled (i.e., [2-^{2}H]-2-methyl-3-hydroxypentanoate); the substrate was tethered enzymatically to a model ACP domain sourced from the DEBS PKS. By incubating with $NADP^+$, the redox reaction was run in reverse, establishing an equilibrium between the oxidized (either (2*R*)- or (2*S*)-2-methyl-3-ketoacyl-ACP) and

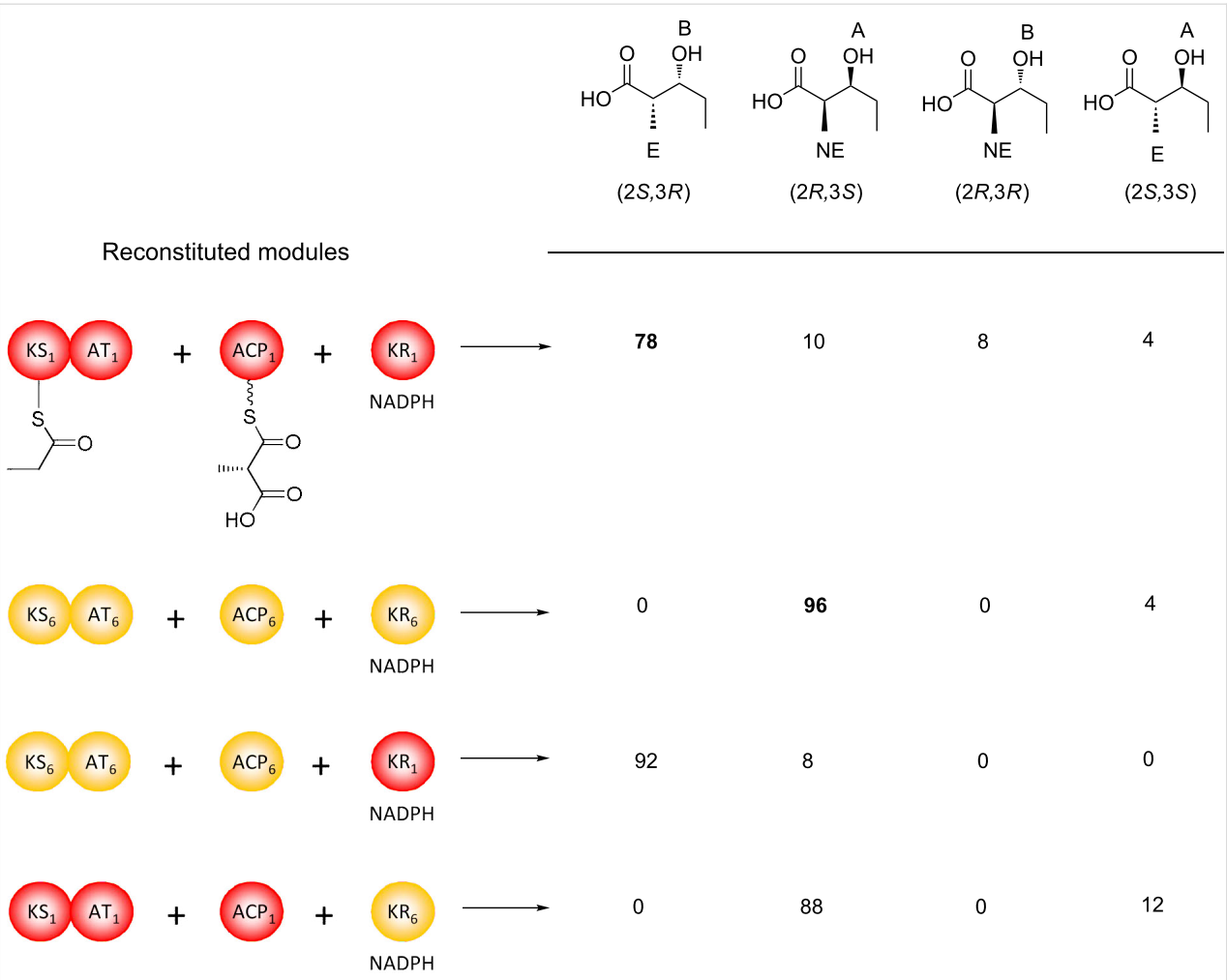


Figure 13: Assays in vitro which provided the first direct evidence that KR domains act as epimerases [77]. Biosynthesis in these experiments was carried out by reconstituted modules comprising KS-AT didomains and isolated KR and ACP domains. The KS was charged with the propionate starter unit from propionyl-SNAC, and the ACP with extender unit by the AT domain using (2*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_1 in place of KR_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_6 and Tyl KR_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) KRs 3), as well as redox-defective mutants of DEBS KR_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-

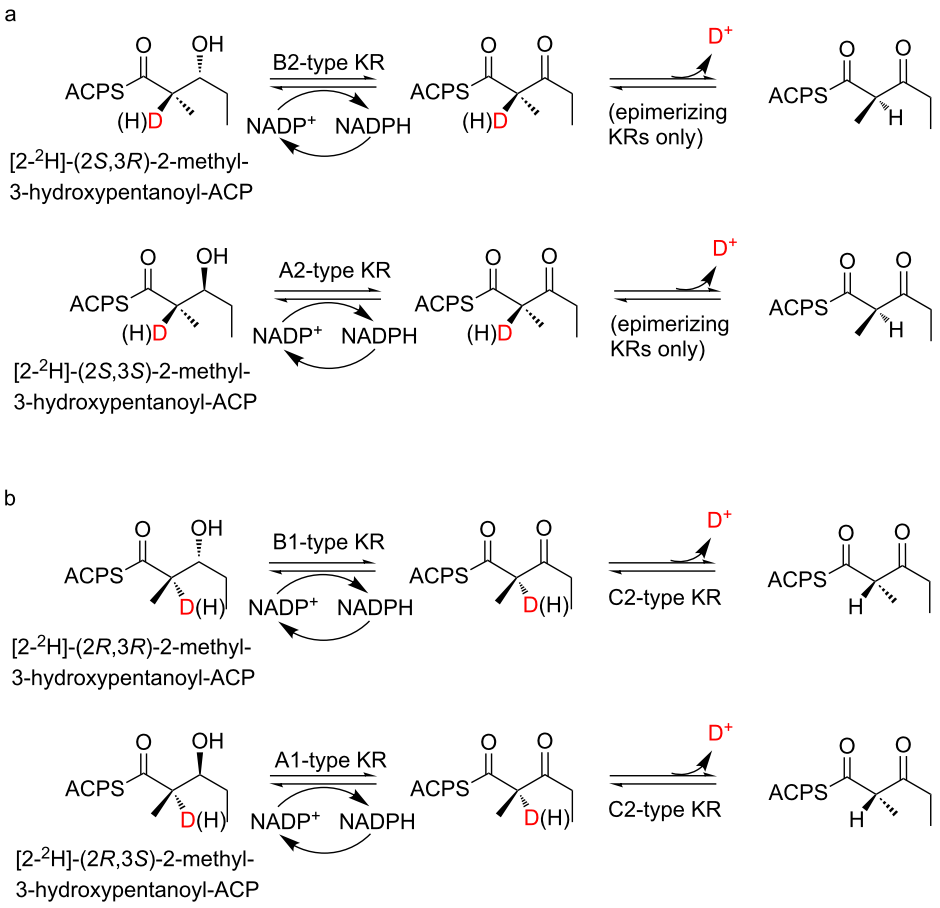


Figure 14: Assays in vitro to demonstrate directly the epimerase activity of PKS KR domains. a) Equilibrium exchange assay [86]. In these assays, an equilibrium is established between stereospecifically deuterated 3-hydroxy diketide-ACP (incorporating either (3*R*)- or (3*S*)-hydroxy stereochemistry as appropriate) and the 3-keto form, which then undergoes KR-catalyzed racemization at the C-2 center. This epimerizing activity is detected by LC–MS via time-dependent washout of deuterium from the reduced product. While A1 or B1-type KRs could catalyze the oxidation of the deuterated compounds, the deuterium would not be lost by subsequent epimerization. b) Tandem equilibrium exchange assay [87]. The aim of this assay is to demonstrate the intrinsic epimerization activity of non-reducing KRs (C2-type). As these are not capable of establishing the initial equilibrium between the C-3 hydroxy and keto forms of the substrate, an additional reducing but non-epimerizing KR (either A1- or B1-type) is added to the assays to carry out these step with (3*R*)- and (3*S*)-hydroxy substrates, respectively. The epimerizing capacity of the KR under study is then detected as for the classical equilibrium exchange assay described above.

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 KR_6). 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*-(2*S*) hydrogen and the (3*R*)-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 ((2*R*), 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 (3*R*)-hydroxy groups (B-type ketoreduction) leading directly to *trans* double bonds and (3*S*)-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 KRs 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, (2*R*,3*R*) intermediates yield *trans* double bond stereochemistry, and (2*R*,3*S*), *cis* double bond stereochemistry).

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 (2*R*,3*R*)-2-methyl-3-hydroxyacyl chains. In the case of the DEBS DH, no reaction was observed for the three diastereomeric substrates (e.g., (2*S*,3*R*), (2*R*,3*S*) and (2*S*,3*S*)). 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 (2*S*,3*S*)-2-methyl-3-hydroxyacyl functionality [104], which is not in accord with the (2*R*,3*S*) stereochemistry postulated for *cis*-double bond pre-

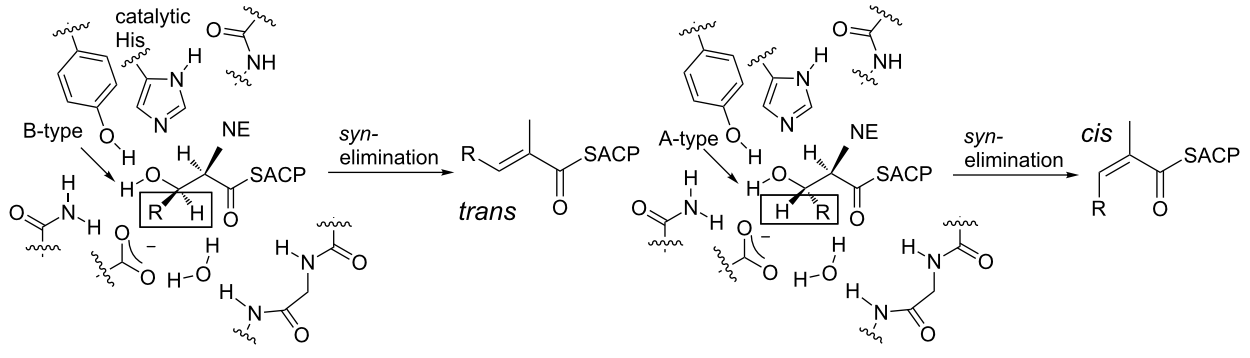


Figure 15: Model for DH-catalyzed generation of *trans* and *cis* double bonds by *syn* elimination from substrates bearing C-2-methyl groups (the side chains shown are those found in DEBS DH_4 [92]). Both substrates incorporate a (2*R*)- (non-epimerized, NE) methyl group, but differ in the relative 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].

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 (2*R*,3*R*) 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

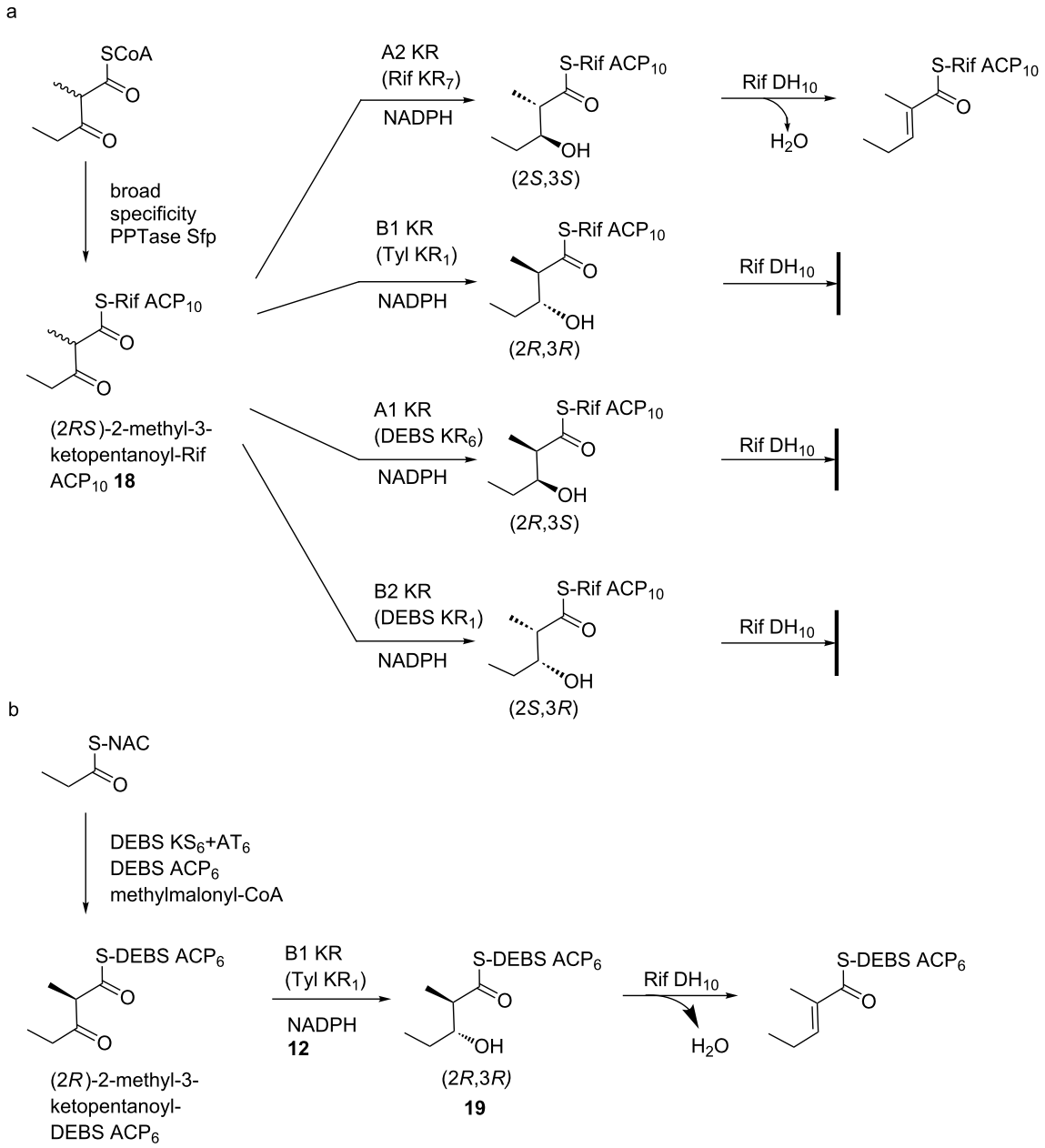


Figure 16: Stereospecificity of dehydration by Rif DH_{10} [94]. a) The four possible diastereomeric diketide-ACP substrates were prepared by stereospecific C-3 reduction (and C-2-methyl epimerization in the case of the A2 and B2-type KRs) of (2*RS*)-2-methyl-3-ketopentanoyl-Rif ACP_{10}. Only the substrate bearing (2*S*,3*S*)-2-methyl-3-hydroxy stereochemistry was accepted as substrate by Rif DH_{10}, yielding the *trans* double bond isomer. b) The alternative (2*R*,3*R*)-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 (2*S*,3*S*) and (2*R*,3*R*) 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 KRs 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 (2*R*)- and (2*S*)-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 tridomains) 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 (2*S*)-methyl, while TKS-rap13 yields the alternative (2*R*)-methyl. Dramatically, when the conserved Y of the DEBS ER in TKS-ery4 was replaced with V, the resulting lactone **21** incorporated exclusively the (2*R*)-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 carbanion 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 (2*R*)-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

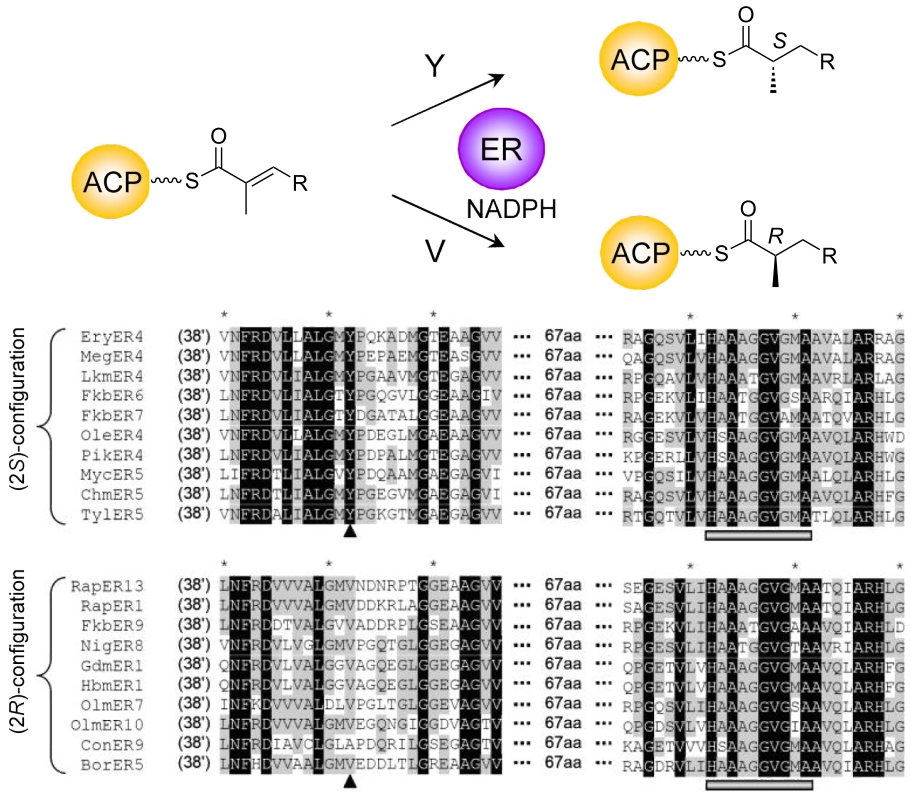


Figure 17: Stereocontrol by PKS ER domains. Sequence motifs correlated with the final stereochemistry of the C-2 methyl group [116]. When a conserved Y is present (indicated with the triangle), a (2*S*)-methyl stereochemistry is observed, while the presence of a conserved V at the same position correlates with (2*R*)-methyl stereochemistry. The grey bar indicates residues involved in binding the NADPH cofactor **12**. Residue numbering is based on that of *E. coli* QOR (PDB ID 1QOR). Reprinted and adapted with permission from [116]. Copyright (2010) American Chemical Society.

and/or corroborate absolute stereochemical assignments made on newly-discovered natural products (for example, elansolid [117], the disciformycins [118], hygrobafilomycin [119], phormidiolide [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 KRs 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 KRs, 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 KRs possess sequence features of both A- and B-type KRs, 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-

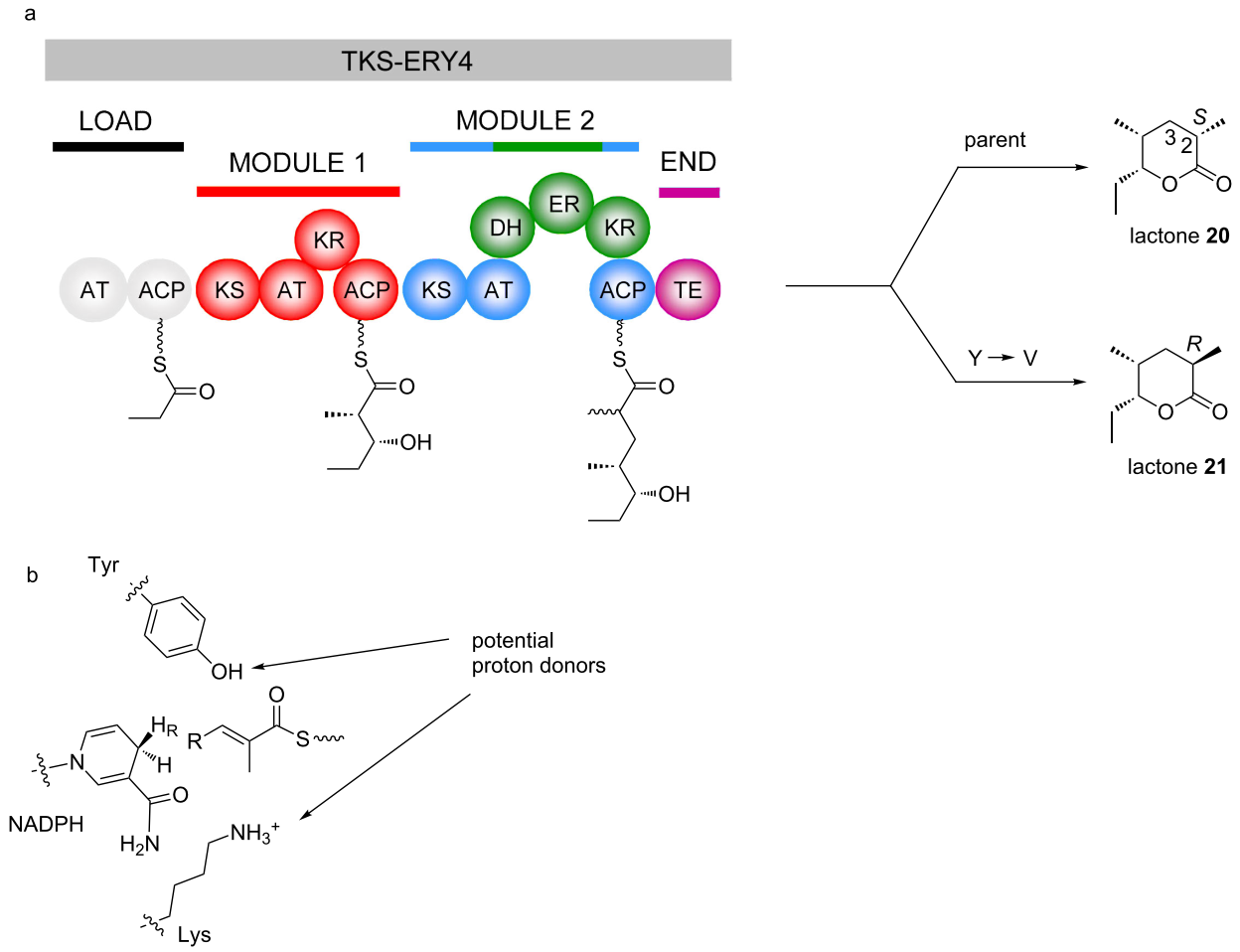


Figure 18: a) PKS engineered to test the role of the ER stereospecificity residues [115]. TKS-ERY4 was created by replacing the KR domain of DEBS module 2 (within the context of DEBS 1-TE) with the 'reductive loop' (DH-ER-KR) from DEBS module 4. This PKS gives rise to lactone **20** incorporating a (2*S*)-methyl 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 (2*R*)-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 (2*S*)-methyl. In its absence, the Lys residue on the opposite side of the active site acts as the general acid, yielding the (2*R*)-methyl.

proach will undoubtedly continue to play an indispensable role in future work in this area.

References

1. Demain, A. L. *J. Ind. Microbiol. Biotechnol.* **2014,** *41,* 185–201. doi:10.1007/s10295-013-1325-z
2. Ma, X.; Ma, S. *Curr. Med. Chem.* **2011,** *18,* 1993–2015. doi:10.2174/092986711795590075
3. Wang, Y.; Cao, S.; Chen, Y. *Anti-Cancer Agents Med. Chem.* **2015,** *15,* 701–720. doi:10.2174/1871520615666150129211901
4. Clardy, J.; Walsh, C. *Nature* **2004,** *432,* 829–837. doi:10.1038/nature03194
5. Schlünzen, F.; Zarivach, R.; Harms, J.; Bashan, A.; Tocili, A.; Albrecht, R.; Yonath, A.; Franceshi, F. *Nature* **2001,** *413,* 814–821. doi:10.1038/35101544
6. Weissman, K. J. *Nat. Prod. Rep.* **2015,** *33,* 203–230. doi:10.1039/C5NP00109A
7. Voices of chemical biology. *Nat. Chem. Biol.* **2015,** *11,* 378–379. doi:10.1038/nchembio.1820
8. Smith, S.; Tsai, S.-C. *Nat. Prod. Rep.* **2007,** *24,* 1041–1072. doi:10.1039/b603600g
9. Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. *Mol. Biol. Evol.* **2005,** *22,* 2027–2039. doi:10.1093/molbev/msi193
10. Olano, C.; Méndez, C.; Salas, J. A. *Nat. Prod. Rep.* **2010,** *27,* 571–616. doi:10.1039/b911956f
11. Helfrich, E. J. N.; Piel, J. *Nat. Prod. Rep.* **2016,** *33,* 231–316. doi:10.1039/C5NP00125K
12. Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008,** *26,* 225–233. doi:10.1038/nbt1379
13. Lohman, J. R.; Ma, M.; Osipiuk, J.; Nocek, B.; Kim, Y.; Chang, C.; Cuff, M.; Mack, J.; Bigelow, L.; Li, H.; Endres, M.; Banigg, G.; Joachimiak, A.; Phillips, G. N., Jr.; Shen, B. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 12693–12698. doi:10.1073/pnas.1515460112

14. Pulsawat, N.; Kitani, S.; Nihira, T. *Gene* **2007,** *393,* 31–42. doi:10.1016/j.gene.2006.12.035
15. Ray, L.; Moore, B. S. *Nat. Prod. Rep.* **2016,** *33,* 150–161. doi:10.1039/C5NP00112A
16. Wilson, M. C.; Moore, B. S. *Nat. Prod. Rep.* **2012,** *29,* 72–86. doi:10.1039/C1NP00082A
17. Musiol, E. M.; Härtner, T.; Kulik, A.; Moldenhauer, J.; Piel, J.; Wohlleben, W.; Weber, T. *Chem. Biol.* **2011,** *18,* 438–444. doi:10.1016/j.chembiol.2011.02.007
18. Stevens, D. C.; Wagner, D. T.; Manion, H. R.; Alexander, B. K.; Keatinge-Clay, A. T. *J. Antibiot.* **2016,** *69,* 567–570. doi:10.1038/ja.2016.66
19. Berkhan, G.; Hahn, F. *Angew. Chem., Int. Ed.* **2014,** *53,* 14240–14244. doi:10.1002/anie.201407979
20. Pöplau, P.; Frank, S.; Morinaka, B. I.; Piel, J. *Angew. Chem., Int. Ed.* **2013,** *52,* 13215–13218. doi:10.1002/anie.201307406
21. Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C. *Angew. Chem., Int. Ed.* **2010,** *49,* 1460–1464. doi:10.1002/anie.200905467
22. Lohr, F.; Jenniches, I.; Frizler, M.; Meehan, M. J.; Sylvester, M.; Schmitz, A.; Gütschow, M.; Dorrestein, P. C.; König, G. M.; Schäberle, T. F. *Chem. Sci.* **2013,** *4,* 4175–4180. doi:10.1039/c3sc51854j
23. Cane, D. E.; Liang, T. C.; Taylor, P. B.; Chang, C.; Yang, C. C. *J. Am. Chem. Soc.* **1986,** *108,* 4957–4964. doi:10.1021/ja00276a042
24. Sedgwick, B.; French, S. J.; Cornforth, J. W.; Gray, R. T.; Kelstrup, E.; Willadsen, P. *FEBS J.* **1977,** *75,* 481–495. doi:10.1111/j.1432-1033.1977.tb11550.x
25. Caffrey, P.; Bevitt, D. J.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1992,** *304,* 225–228. doi:10.1016/0014-5793(92)80624-P
26. Marsden, A. F.; Caffrey, P.; Aparicio, J. F.; Loughran, M. S.; Staunton, J.; Leadlay, P. F. *Science* **1994,** *263,* 378–380. doi:10.1126/science.8278811
27. Wiesmann, K. E. H.; Cortés, J.; Brown, M. J. B.; Cutter, A. L.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1995,** *2,* 583–589. doi:10.1016/1074-5521(95)90122-1
28. Cortes, J.; Wiesmann, K. E.; Roberts, G. A.; Brown, M. J.; Staunton, J.; Leadlay, P. F. *Science* **1995,** *268,* 1487–1489. doi:10.1126/science.7770773
29. Dunn, B. J.; Khosla, C. *J. R. Soc., Interface* **2013,** *10,* 20130297. doi:10.1098/rsif.2013.0297
30. Dunn, B. J.; Cane, D. E.; Khosla, C. *Biochemistry* **2013,** *52,* 1839–1841. doi:10.1021/bi400185v
31. Haydock, S. F.; Aparicio, J. F.; Molnár, I.; Schwecke, T.; Khaw, L. E.; König, A.; Marsden, A. F. A.; Galloway, I. S.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1995,** *374,* 246–248. doi:10.1016/0014-5793(95)01119-Y
32. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2001,** *40,* 15464–15470. doi:10.1021/bi015864r
33. Yadav, G.; Gokhale, R. S.; Mohanty, D. *J. Mol. Biol.* **2003,** *328,* 335–363. doi:10.1016/S0022-2836(03)00232-8
34. Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortés, J.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. *J. Ind. Microbiol. Biotechnol.* **2003,** *30,* 489–494. doi:10.1007/s10295-003-0062-0
35. Minowa, Y.; Araki, M.; Kanehisa, M. *J. Mol. Biol.* **2007,** *368,* 1500–1517. doi:10.1016/j.jmb.2007.02.099
36. Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F. *ACS Chem. Biol.* **2013,** *8,* 443–450. doi:10.1021/cb300505w
37. Bravo-Rodriguez, K.; Klopries, S.; Koopmans, K. R. M.; Sundermann, U.; Yahiaoui, S.; Arens, J.; Kushnir, S.; Schulz, F.; Sanchez-Garcia, E. *Chem. Biol.* **2015,** *22,* 1425–1430. doi:10.1016/j.chembiol.2015.02.008
38. Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103,* 11124–11129. doi:10.1073/pnas.0601924103
39. Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. *Nat. Prod. Rep.* **2009,** *26,* 90–114. doi:10.1039/B801658P
40. Erb, T. J.; Berg, I. A.; Brecht, V.; Müller, M.; Fuchs, G.; Alber, B. E. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 10631–10636. doi:10.1073/pnas.0702791104
41. Emmert, E. A. B.; Klimowicz, A. K.; Thomas, M. G.; Handelsman, J. *Appl. Environ. Microbiol.* **2004,** *70,* 104–113. doi:10.1128/AEM.70.1.104-113.2004
42. Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D. *Gene* **2000,** *251,* 81–90. doi:10.1016/S0378-1119(00)00171-2
43. Park, H.; Kevany, B. M.; Dyer, D. H.; Thomas, M. G.; Forest, K. T. *PLoS One* **2014,** *9,* e110965. doi:10.1371/journal.pone.0110965
44. Witkowski, A.; Joshi, A. K.; Smith, S. *Biochemistry* **2002,** *41,* 10877–10887. doi:10.1021/bi0259047
45. Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. *Biochemistry* **2016,** *55,* 4476–4484. doi:10.1021/acs.biochem.6b00639
46. Weissman, K. J.; Timoney, M.; Bycroft, M.; Grice, P.; Hanefeld, U.; Staunton, J.; Leadlay, P. F. *Biochemistry* **1997,** *36,* 13849–13855. doi:10.1021/bi971566b
47. Böhm, I.; Holzbaur, I. E.; Hanefeld, U.; Cortési, J.; Staunton, J.; Leadlay, P. F. *Cell Chem. Biol.* **1998,** *5,* 407–412. doi:10.1016/S1074-5521(98)90157-0
48. Holzbaur, I. E.; Ranganathan, A.; Thomas, I. P.; Kearney, D. J. A.; Reather, J. A.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. *Cell Chem. Biol.* **2001,** *8,* 329–340. doi:10.1016/S1074-5521(01)00014-X
49. Kao, C. M.; McPherson, M.; McDaniel, R. N.; Fu, H.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **1998,** *120,* 2478–2479. doi:10.1021/ja973913a
50. Kellenberger, L.; Galloway, I. S.; Sauter, G.; Böhm, G.; Hanefeld, U.; Cortés, J.; Staunton, J.; Leadlay, P. F. *ChemBioChem* **2008,** *9,* 2740–2749. doi:10.1002/cbic.200800332
51. Annaval, T.; Paris, C.; Leadlay, P. F.; Jacob, C.; Weissman, K. J. *ChemBioChem* **2015,** *16,* 1357–1364. doi:10.1002/cbic.201500113
52. Yin, Y.; Gokhale, R.; Khosla, C.; Cane, D. E. *Bioorg. Med. Chem. Lett.* **2001,** *11,* 1477–1479. doi:10.1016/S0960-894X(00)00529-1
53. McPherson, M.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **1998,** *120,* 3267–3268. doi:10.1021/ja980028z
54. Almarsson, O.; Bruice, T. C. *J. Am. Chem. Soc.* **1993,** *115,* 2125–2138. doi:10.1021/ja00059a005
55. Glasfeld, A.; Leanz, G. F.; Benner, S. A. *J. Biol. Chem.* **1990,** *265,* 11692–11699.
56. Keatinge-Clay, A. T.; Stroud, R. M. *Structure* **2006,** *14,* 737–748. doi:10.1016/j.str.2006.01.009
57. Keatinge-Clay, A. T. *Chem. Biol.* **2007,** *14,* 898–908. doi:10.1016/j.chembiol.2007.07.009
58. Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T. *Structure* **2010,** *18,* 913–922. doi:10.1016/j.str.2010.04.015
59. Zheng, J.; Keatinge-Clay, A. T. *J. Mol. Biol.* **2011,** *410,* 105–117. doi:10.1016/j.jmb.2011.04.065

60. Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. *Nat. Chem. Biol.* **2012,** *8,* 615–621. doi:10.1038/nchembio.964
61. Zheng, J.; Piasecki, S. K.; Keatinge-Clay, A. T. *ACS Chem. Biol.* **2013,** *8,* 1964–1971. doi:10.1021/cb400161g
62. Bonnett, S. A.; Whicher, J. R.; Papireddy, K.; Florova, G.; Smith, J. L.; Reynolds, K. A. *Chem. Biol.* **2013,** *20,* 772–783. doi:10.1016/j.chembiol.2013.04.014
63. Piasecki, S. K.; Zheng, J.; Axelrod, A. J.; Detelich, M. E.; Keatinge-Clay, A. T. *Proteins: Struct., Funct., Bioinf.* **2014,** *82,* 2067–2077. doi:10.1002/prot.24561
64. Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2003,** *42,* 72–79. doi:10.1021/bi0268706
65. Filling, C.; Berndt, K. D.; Benach, J.; Knapp, S.; Prozorovski, T.; Nordling, E.; Ladenstein, R.; Jörnvall, H.; Oppermann, U. *J. Biol. Chem.* **2002,** *277,* 25677–25684. doi:10.1074/jbc.M202160200
66. Kavanagh, K. L.; Jörnvall, H.; Persson, B.; Oppermann, U. *Cell. Mol. Life Sci.* **2008,** *65,* 3895–3906. doi:10.1007/s00018-008-8588-y
67. Caffrey, P. *ChemBioChem* **2003,** *4,* 654–657. doi:10.1002/cbic.200300581
68. Siskos, A. P.; Baerga-Ortiz, A.; Bali, S.; Stein, V.; Mamdani, H.; Spiteller, D.; Popovic, B.; Spencer, J. B.; Staunton, J.; Weissman, K. J.; Leadlay, P. F. *Chem. Biol.* **2005,** *12,* 1145–1153. doi:10.1016/j.chembiol.2005.08.017
69. Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2016,** *33,* 141–149. doi:10.1039/C5NP00092K
70. Holzbaur, I. E.; Harris, R. C.; Bycroft, M.; Cortés, J.; Bisang, C.; Staunton, J.; Rudd, B. A. M.; Leadlay, P. F. *Chem. Biol.* **1999,** *6,* 189–195. doi:10.1016/S1074-5521(99)80035-0
71. Østergaard, L. H.; Kellenberger, L.; Cortés, J.; Roddis, M. P.; Deacon, M.; Staunton, J.; Leadlay, P. F. *Biochemistry* **2002,** *41,* 2719–2726. doi:10.1021/bi0117605
72. Castonguay, R.; Valenzano, C. R.; Chen, A. Y.; Keatinge-Clay, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 11598–11599. doi:10.1021/ja804453p
73. Baerga-Ortiz, A.; Popovic, B.; Siskos, A. P.; O'Hare, H. M.; Spiteller, D.; Williams, M. G.; Campillo, N.; Spencer, J. B.; Leadlay, P. F. *Chem. Biol.* **2006,** *13,* 277–285. doi:10.1016/j.chembiol.2006.01.004
74. O'Hare, H. M.; Baerga-Ortiz, A.; Popovic, B.; Spencer, J. B.; Leadlay, P. F. *Chem. Biol.* **2006,** *13,* 287–296. doi:10.1016/j.chembiol.2006.01.003
75. Bali, S.; Weissman, K. J. *ChemBioChem* **2006,** *7,* 1935–1942. doi:10.1002/cbic.200600285
76. Häckh, M.; Müller, M.; Lüdeke, S. *Chem. – Eur. J.* **2013,** *19,* 8922–8928. doi:10.1002/chem.201300554
77. Valenzano, C. R.; Lawson, R. J.; Chen, A. Y.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2009,** *131,* 18501–18511. doi:10.1021/ja908296m
78. Kwan, D. H.; Tosin, M.; Schläger, N.; Schulz, F.; Leadlay, P. F. *Org. Biomol. Chem.* **2011,** *9,* 2053–2056. doi:10.1039/c1ob00022e
79. Heathcote, M. L.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2001,** *8,* 207–220. doi:10.1016/S1074-5521(01)00002-3
80. Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A. *J. Biol. Chem.* **2002,** *277,* 48028–48034. doi:10.1074/jbc.M207770200
81. Zhou, Y.; Meng, Q.; You, D.; Li, J.; Chen, S.; Ding, D.; Zhou, X.; Zhou, H.; Bai, L.; Deng, Z. *Appl. Environ. Microbiol.* **2008,** *74,* 7235–7242. doi:10.1128/AEM.01012-08
82. Castonguay, R.; He, W.; Chen, A. Y.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2007,** *129,* 13758–13769. doi:10.1021/ja0753290
83. Charkoudian, L. K.; Liu, C. W.; Capone, S.; Kapur, S.; Cane, D. E.; Togni, A.; Seebach, D.; Khosla, C. *Protein Sci.* **2011,** *20,* 1244–1255. doi:10.1002/pro.652
84. Tran, L.; Broadhurst, R. W.; Tosin, M.; Cavalli, A.; Weissman, K. *J. Chem. Biol.* **2010,** *17,* 705–716. doi:10.1016/j.chembiol.2010.05.017
85. Busche, A.; Gottstein, D.; Hein, C.; Ripin, N.; Pader, I.; Tufar, P.; Eisman, E. B.; Gu, L.; Walsh, C. T.; Sherman, D. H.; Löhr, F.; Güntert, P.; Dötsch, V. *ACS Chem. Biol.* **2012,** *7,* 378–386. doi:10.1021/cb200352q
86. Garg, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2013,** *135,* 16324–16327. doi:10.1021/ja408944s
87. Garg, A.; Xie, X.; Keatinge-Clay, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2014,** *136,* 10190–10193. doi:10.1021/ja5056998
88. Xie, X.; Garg, A.; Keatinge-Clay, A. T.; Khosla, C.; Cane, D. E. *Biochemistry* **2016,** *55,* 1179–1186. doi:10.1021/acs.biochem.6b00024
89. Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. *Biochemistry* **2006,** *45,* 12756–12766. doi:10.1021/bi061169d
90. Kitsche, A.; Kalesse, M. *ChemBioChem* **2013,** *14,* 851–861. doi:10.1002/cbic.201300063
91. Eng, C. H.; Yuzawa, S.; Wang, G.; Baidoo, E. E. K.; Katz, L.; Keasling, J. D. *Biochemistry* **2016,** *55,* 1677–1680. doi:10.1021/acs.biochem.6b00129
92. Keatinge-Clay, A. *J. Mol. Biol.* **2008,** *384,* 941–953. doi:10.1016/j.jmb.2008.09.084
93. Akey, D. L.; Razelun, J. R.; Tehranisa, J.; Sherman, D. H.; Gerwick, W. H.; Smith, J. L. *Structure* **2010,** *18,* 94–105. doi:10.1016/j.str.2009.10.018
94. Gay, D.; You, Y.-O.; Keatinge-Clay, A.; Cane, D. E. *Biochemistry* **2013,** *52,* 8916–8928. doi:10.1021/bi400988t
95. Labonte, J. W.; Townsend, C. A. *Chem. Rev.* **2013,** *113,* 2182–2204. doi:10.1021/cr300169a
96. Steven, A. C.; Baumeister, W.; Johnson, L. N.; Perham, R. N. *Molecular Biology of Assemblies and Machines;* Garland Science: New York, 2016.
97. Sedgwick, B.; Morris, C.; French, S. J. *J. Chem. Soc., Chem. Commun.* **1978,** 193–194. doi:10.1039/c39780000193
98. Schwab, J. M.; Habib, A.; Klassen, J. B. *J. Am. Chem. Soc.* **1986,** *108,* 5304–5308. doi:10.1021/ja00277a040
99. Schwab, J. M.; Klassen, J. B.; Habib, A. *J. Chem. Soc., Chem. Commun.* **1986,** 357–358. doi:10.1039/C39860000357
100. Valenzano, C. R.; You, Y.-O.; Garg, A.; Keatinge-Clay, A.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2010,** *132,* 14697–14699. doi:10.1021/ja107344h
101. Guo, X.; Liu, T.; Valenzano, C. R.; Deng, Z.; Cane, D. E. *J. Am. Chem. Soc.* **2010,** *132,* 14694–14696. doi:10.1021/ja1073432
102. Fiers, W. D.; Dodge, G. J.; Li, Y.; Smith, J. L.; Fecik, R. A.; Aldrich, C. C. *Chem. Sci.* **2015,** *6,* 5027–5033. doi:10.1039/C5SC01505G
103. Alhamadsheh, M. M.; Palaniappan, N.; DasChouduri, S.; Reynolds, K. A. *J. Am. Chem. Soc.* **2007,** *129,* 1910–1911. doi:10.1021/ja068818t
104. You, Y.-O.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2013,** *135,* 7406–7409. doi:10.1021/ja4014776

105. Perlova, O.; Gerth, K.; Kaiser, O.; Hans, A.; Müller, R. *J. Biotechnol.* **2006,** *121,* 174–191. doi:10.1016/j.jbiotec.2005.10.011
106. Gay, D. C.; Spear, P. J.; Keatinge-Clay, A. T. *ACS Chem. Biol.* **2014,** *9,* 2374–2381. doi:10.1021/cb500459b
107. Palaniappan, N.; Alhamadsheh, M. M.; Reynolds, K. A. *J. Am. Chem. Soc.* **2008,** *130,* 12236–12237. doi:10.1021/ja8044162
108. Vergnolle, O.; Hahn, F.; Baerga-Ortiz, A.; Leadlay, P. F.; Andexer, J. N. *ChemBioChem* **2011,** *12,* 1011–1014. doi:10.1002/cbic.201100011
109. Kandziora, N.; Andexer, J. N.; Moss, S. J.; Wilkinson, B.; Leadlay, P. F.; Hahn, F. *Chem. Sci.* **2014,** *5,* 3563–3567. doi:10.1039/C4SC00883A
110. He, H.-Y.; Tang, M.-C.; Zhang, F.; Tang, G.-L. *J. Am. Chem. Soc.* **2014,** *136,* 4488–4491. doi:10.1021/ja500942y
111. Bevitt, D. J.; Staunton, J.; Leadlay, P. F. *Biochem. Soc. Trans.* **1993,** *21,* 30S. doi:10.1042/bst021030s
112. Wu, J.; Zaleski, T. J.; Valenzano, C.; Khosla, C.; Cane, D. E. *J. Am. Chem. Soc.* **2005,** *127,* 17393–17404. doi:10.1021/ja055672+
113. Leesong, M.; Henderson, B. S.; Gillig, J. R.; Schwab, J. M.; Smith, J. L. *Structure* **1996,** *4,* 253–264. doi:10.1016/S0969-2126(96)00030-5
114. Anderson, V. E.; Hammes, G. G. *Biochemistry* **1984,** *23,* 2088–2094. doi:10.1021/bi00304a033
115. Kwan, D. H.; Sun, Y.; Schulz, F.; Hong, H.; Popovic, B.; Sim-Stark, J. C. C.; Haydock, S. F.; Leadlay, P. F. *Chem. Biol.* **2008,** *15,* 1231–1240. doi:10.1016/j.chembiol.2008.09.012
116. Kwan, D. H.; Leadlay, P. F. *ACS Chem. Biol.* **2010,** *5,* 829–838. doi:10.1021/cb100175a
117. Dehn, R.; Katsuyama, Y.; Weber, A.; Gerth, K.; Jansen, R.; Steinmetz, H.; Höfle, G.; Müller, R.; Kirschning, A. *Angew. Chem., Int. Ed.* **2011,** *50,* 3882–3887. doi:10.1002/anie.201006880
118. Surup, F.; Viehrig, K.; Mohr, K. I.; Hermann, J.; Jansen, R.; Müller, R. *Angew. Chem., Int. Ed.* **2014,** *53,* 13588–13591. doi:10.1002/anie.201406973
119. Molloy, E. M.; Tietz, J. I.; Blair, P. M.; Mitchell, D. A. *Bioorg. Med. Chem.* **2016,** *24,* 6276–6290. doi:10.1016/j.bmc.2016.05.021
120. Bertin, M. J.; Vulpanovici, A.; Monroe, E. A.; Korobeynikov, A.; Sherman, D. H.; Gerwick, L.; Gerwick, W. H. *ChemBioChem* **2016,** *17,* 164–173. doi:10.1002/cbic.201500467
121. Steinmetz, H.; Li, J.; Fu, C.; Zaburannyi, N.; Kunze, B.; Harmrolfs, K.; Schmitt, V.; Herrmann, J.; Reichenbach, H.; Höfle, G.; Kalesse, M.; Müller, R. *Angew. Chem., Int. Ed.* **2016,** *55,* 10113–10117. doi:10.1002/anie.201603288
122. Hong, H.; Sambroskyy, M.; Lindner, F.; Leadlay, P. F. *Angew. Chem., Int. Ed.* **2016,** *55,* 1118–1123. doi:10.1002/anie.201509300
123. Zhou, T.; Komaki, H.; Ichikawa, N.; Hosoyama, A.; Sato, S.; Igarashi, Y. *Mar. Drugs* **2015,** *13,* 581–596. doi:10.3390/md13010581

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.13.39

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Solid-phase enrichment and analysis of electrophilic natural products

Frank Wesche[1], Yue He[1] and Helge B. Bode[*1,2]

Full Research Paper

Open Access

Address:
[1]Merck Stiftungsprofessur für Molekulare Biotechnologie, Fachbereich Biowissenschaften, Goethe Universität Frankfurt, Max-von-Laue-Strasse 9, D-60438 Frankfurt am Main, Germany and [2]Buchmann Institute for Molecular Life Sciences (BMLS), Goethe Universität Frankfurt, Max-von-Laue-Strasse 15, D-60438 Frankfurt am Main, Germany

Email:
Helge B. Bode[*] - h.bode@bio.uni-frankfurt.de

* Corresponding author

Keywords:
azides; click chemistry; enrichment; electrophilic natural products; epoxides; glidobactin; *Photorhabdus*; stilbenes

Beilstein J. Org. Chem. **2017,** *13,* 405–409.
doi:10.3762/bjoc.13.43

Received: 04 October 2016
Accepted: 10 February 2017
Published: 02 March 2017

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

Associate Editor: S. Flitsch

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

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

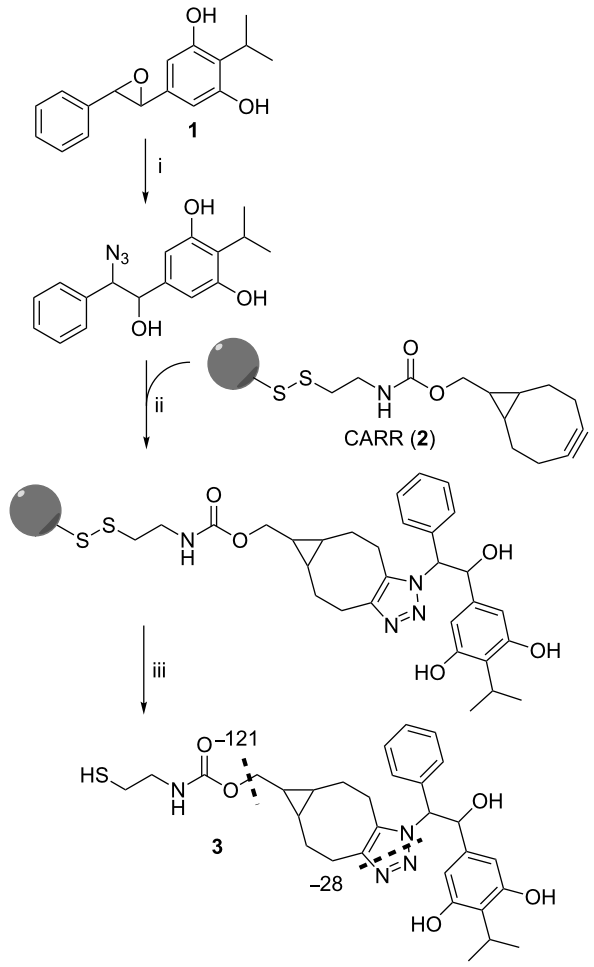
Microorganisms are a major source for novel natural products and the subsequent development of new drugs for all kinds of applications [1,2]. For example, the discovery of the fungal natural product cyclosporine as an immunosuppressant drug facilitated modern organ transplantation [3,4].

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.



Scheme 1: Principle of azidation of XAD extracts from *P. luminescens* TT01 containing **1** and subsequent azide enrichment with CARR (**2**). After the vicinal azido alcohol is covalently bound to the resin through an azide–alkyne cycloaddition, compound **3** is cleaved from the resin and analyzed by HPLC–MS. Reaction conditions: i) NaN_3, NH_4Cl, 80% MeOH in H_2O, reflux overnight; ii) CARR (**2**), ACN, 55 °C, 1 h, then rt, overnight; iii) 5 mM TCEP in PBS/$CHCl_3$/MeOH 1:5:10 (v/v/v), 1 h.

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 *Photorhabdus* 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 MS^2 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

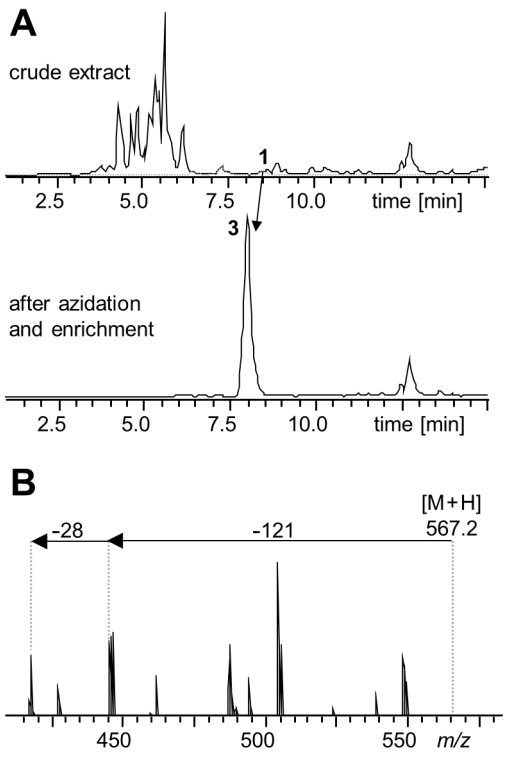


Figure 1: (A) HPLC–MS base peak chromatograms of a crude XAD extract of *P. luminescens* TT01 and after azidation and azide enrichment. The t_R of **1** is indicated in the crude extract. After enrichment, one distinct peak at 8.1 min corresponding to **3** is visible. (B) The characteristic MS^2 fragmentation pattern of the derivatized azide **3** is shown highlighting the neutral loss of carbamate (−121) and dinitrogen (−28) as characteristic fragments for CARR adducts.

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 ^{13}C-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 $[M + H]^+$, Δppm = 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.

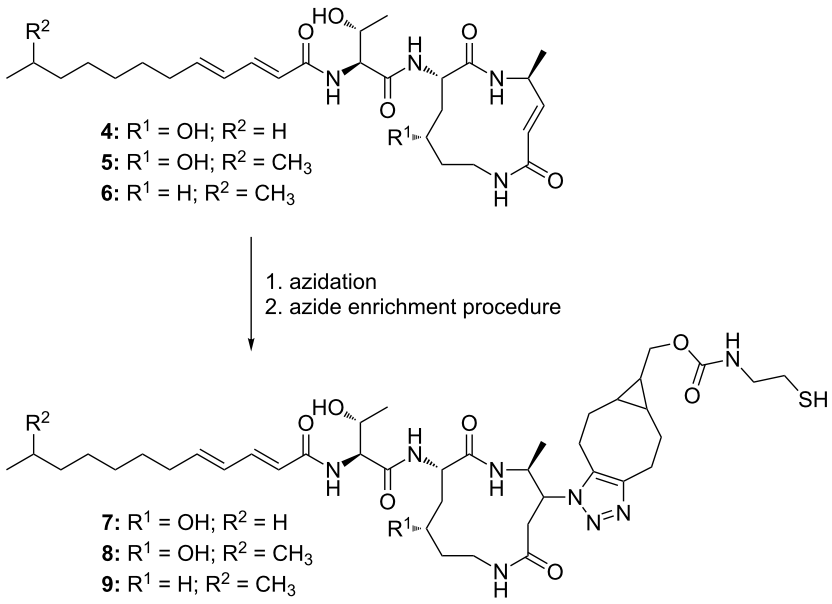
Upon a detailed look at the chromatograms, different masses with the characteristic fragmentation pattern of derivatized azides could be found (Table 1 and Supporting Information File 1, Figures S7–S9). The molecular formula obtained from HRMS data indicates that three glidobactin derivatives namely 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 ($C_{13}H_{19}N_4O_2S$).

Strain	Compd.	t_R min	$[M + H]^+$ found	Calcd. molecular formula	$[M + H]^+$ calcd.	Δppm	Molecular formula natural product	Natural product
P. luminescens TT01	**3**	8.1	567.2635	$C_{30}H_{39}N_4O_5S$	567.2636	0.1	$C_{17}H_{18}O_3$	**1**
Photorhabdus PB 68.1	–	8.6	491.2103	$C_{27}H_{31}N_4O_3S$	491.2071	2.2	$C_{14}H_{10}O$	unknown
	–	8.0	493.2258	$C_{27}H_{33}N_4O_3S$	493.2268	2.0	$C_{14}H_{12}O$	unknown
	–	8.5	517.2483	$C_{26}H_{37}N_4O_5S$	517.2479	2.0	$C_{13}H_{16}O_3$	unknown
	3	8.1	567.2635	$C_{30}H_{39}N_4O_5S$	567.2636	0.1	$C_{17}H_{18}O_3$	**1**
	7	9.0	817.4630	$C_{40}H_{65}N_8O_8S$	817.4641	2.0	$C_{27}H_{44}N_4O_6$	**4**
	8	9.3	831.4801	$C_{41}H_{67}N_8O_8S$	831.4797	1.1	$C_{28}H_{46}N_4O_6$	**5**

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 ($C_{13}H_{19}N_4O_2S$). (continued)

Photorhabdus PB 45.5	–	10.7	551.3617	$C_{29}H_{51}N_4O_4S$	551.3626	1.5	$C_{16}H_{30}O_2$	unknown
	9	9.8	815.4859	$C_{41}H_{67}N_8O_7S$	815.4848	1.3	$C_{28}H_{46}N_4O_5$	**6**
	–	10.1	819.5144	$C_{41}H_{71}N_8O_7S$	819.5161	2.0	$C_{28}H_{50}N_4O_5$	unknown
Photorhabdus temperata subsp. *thracensis* DSM 15199	–	8.0	493.2258	$C_{27}H_{33}N_4O_3S$	493.2268	2.0	$C_{14}H_{14}O$	unknown
	–	9.5	489.2525	$C_{25}H_{37}N_4O_4S$	489.2530	3.9	$C_{12}H_{16}O_2$	unknown
	–	10.7	551.3617	$C_{29}H_{51}N_4O_4S$	551.3626	1.5	$C_{16}H_{30}O_2$	unknown



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^2 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].

only been detected from infected insects [25], low-salt growth media [24] or by heterologous expression of the respective gene cluster in *E. coli* [26]. Their identification in this study was confirmed with pure glidobactin A (**4**) showing the same azide reactivity and retention time compared to **4** from the crude extract (Supporting Information File 1, Figure S10). The comparison of the MS/MS spectra of natural and azidated **4** revealed that the reaction took place at the ring-double bond that is also attacked by the threonine in the proteasome (Supporting Information File 1, Figure S11). To the best of our knowledge, this is the first detection of glidobactins in a supernatant of *Photorhabdus* under standard laboratory conditions, which points out once again the advantage of the enrichment step allowing the detection of otherwise barely detectable components in complex mixtures. A determination of the detection limit for pure glidobactin A (**4**) added into LB medium revealed a detection limit of 10 µg/L (≈20 nmol/L) (Supporting Information File 1, Figure S12) and was comparable to the detection limit of the epoxide (Supporting Information File 1, Figure S4).

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

1. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012,** *75,* 311–335. doi:10.1021/np200906s
2. Cragg, G. M.; Newman, D. J. *Biochim. Biophys. Acta* **2013,** *1830,* 3670–3695. doi:10.1016/j.bbagen.2013.02.008
3. Laupacis, A.; Keown, P. A.; Ulan, R. A.; McKenzie, N.; Stiller, C. R. *Can. Med. Assoc. J.* **1982,** *126,* 1041–1046.
4. Kahan, B. D. *Tex. Heart Inst. J.* **1982,** *9,* 253–266.
5. Rutledge, P. J.; Challis, G. L. *Nat. Rev. Microbiol.* **2015,** *13,* 509–523. doi:10.1038/nrmicro3496
6. Miles, C. O.; Sandvik, M.; Nonga, H. E.; Rundberget, T.; Wilkins, A. L.; Rise, F.; Ballot, A. *Environ. Sci. Technol.* **2012,** *46,* 8937–8944. doi:10.1021/es301808h
7. Cox, C. L.; Tietz, J. I.; Sokolowski, K.; Melby, J. O.; Doroghazi, J. R.; Mitchell, D. A. *ACS Chem. Biol.* **2014,** *9,* 2014–2022. doi:10.1021/cb500324n
8. Carlson, E. E.; Cravatt, B. F. *J. Am. Chem. Soc.* **2007,** *129,* 15780–15782. doi:10.1021/ja0779506
9. Carlson, E. E.; Cravatt, B. F. *Nat. Methods* **2007,** *4,* 429–435. doi:10.1038/nmeth1038
10. Marfey, P. *Carlsberg Res. Commun.* **1984,** *49,* 591–596. doi:10.1007/BF02908688
11. Odendaal, A. Y.; Trader, D. J.; Carlson, E. E. *Chem. Sci.* **2011,** *2,* 760–764. doi:10.1039/C0SC00620C
12. Castro-Falcón, G.; Hahn, D.; Reimer, D.; Hughes, C. C. *ACS Chem. Biol.* **2016,** *11,* 2328–2336. doi:10.1021/acschembio.5b00924
13. Jeon, H.; Lim, C.; Lee, J. M.; Kim, S. *Chem. Sci.* **2015,** *6,* 2806–2811. doi:10.1039/C5SC00360A
14. Ross, C.; Scherlach, K.; Kloss, F.; Hertweck, C. *Angew. Chem., Int. Ed.* **2014,** *53,* 7794–7798. doi:10.1002/anie.201403344
15. Pérez, A. J.; Wesche, F.; Adihou, H.; Bode, H. B. *Chem. – Eur. J.* **2016,** *22,* 639–645. doi:10.1002/chem.201503781
16. Rowland, E. B.; Rowland, G. B.; Rivera-Otero, E.; Antilla, J. C. *J. Am. Chem. Soc.* **2007,** *129,* 12084–12085. doi:10.1021/ja0751779
17. Lupattelli, P.; Bonini, C.; Caruso, L.; Gambacorta, A. *J. Org. Chem.* **2003,** *68,* 3360–3362. doi:10.1021/jo034133p
18. Hu, K.; Li, J.; Li, B.; Webster, J. M.; Chen, G. *Bioorg. Med. Chem.* **2006,** *14,* 4677–4681. doi:10.1016/j.bmc.2006.01.025
19. Kontnik, R.; Crawford, J. M.; Clardy, J. *ACS Chem. Biol.* **2010,** *5,* 659–665. doi:10.1021/cb100117k
20. Crawford, J. M.; Kontnik, R.; Clardy, J. *Curr. Biol.* **2010,** *20,* 69–74. doi:10.1016/j.cub.2009.10.059
21. Eleftherianos, I.; Boundy, S.; Joyce, S. A.; Aslam, S.; Marshall, J. W.; Cox, R. J.; Simpson, T. J.; Clarke, D. J.; ffrench-Constant, R. H.; Reynolds, S. E. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 2419–2424. doi:10.1073/pnas.0610525104
22. Oka, M.; Nishiyama, Y.; Ohta, S.; Kamei, H.; Konishi, M.; Miyaki, T.; Oki, T.; Kawaguchi, H. *J. Antibiot.* **1988,** *41,* 1331–1337. doi:10.7164/antibiotics.41.1331
23. Terui, Y.; Nishikawa, J.; Hinoo, H.; Kato, T.; Shoji, J. *J. Antibiot.* **1990,** *43,* 788–795. doi:10.7164/antibiotics.43.788
24. Stein, M. L.; Beck, P.; Kaiser, M.; Dudler, R.; Becker, C. F. W.; Groll, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 18367–18371. doi:10.1073/pnas.1211423109
25. Theodore, C. M.; King, J. B.; You, J.; Cichewicz, R. H. *J. Nat. Prod.* **2012,** *75,* 2007–2011. doi:10.1021/np300623x
26. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F.; Zhang, Y. *Nat. Biotechnol.* **2012,** *30,* 440–446. doi:10.1038/nbt.2183

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Investigation of the action of poly(ADP-ribose)-synthesising enzymes on NAD$^+$ analogues

Sarah Wallrodt, Edward L. Simpson and Andreas Marx*

Full Research Paper

Open Access

Address:
Department of Chemistry, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany

Email:
Andreas Marx* - andreas.marx@uni-konstanz.de

* Corresponding author

Keywords:
ARTD; click chemistry; NAD$^+$; poly(ADP-ribose); posttranslational modification

Beilstein J. Org. Chem. **2017,** *13,* 495–501.
doi:10.3762/bjoc.13.49

Received: 20 December 2016
Accepted: 23 February 2017
Published: 10 March 2017

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

Guest Editor: H. B. Bode

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

Abstract

ADP-ribosyl transferases with diphtheria toxin homology (ARTDs) catalyse the covalent addition of ADP-ribose onto different acceptors forming mono- or poly(ADP-ribos)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-riboses onto themselves or different targets forming mono(ADP-ribos)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

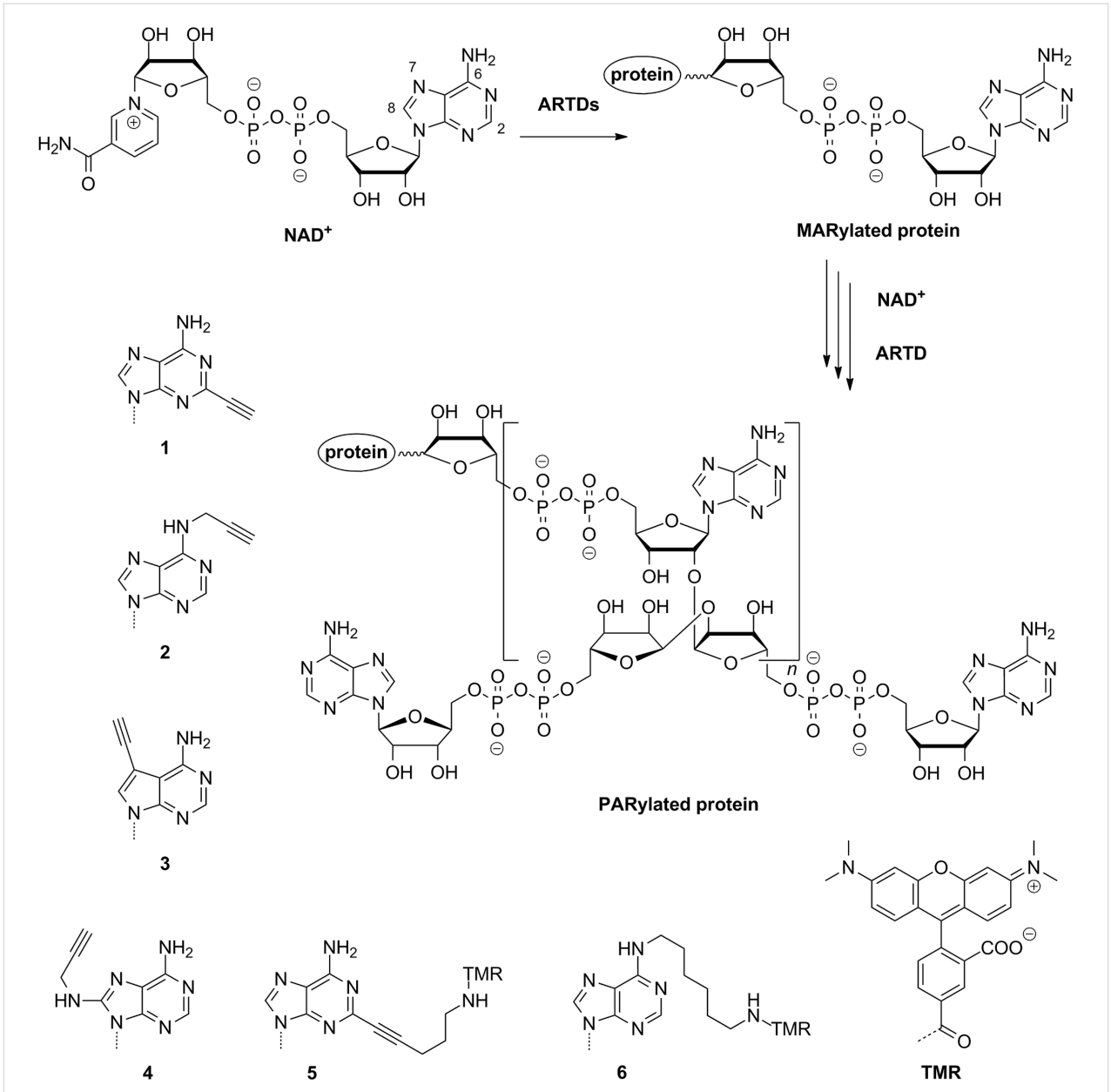


Figure 1: NAD^+ is used as a substrate by ARTDs to form MARylated and PARylated proteins. Depicted are alkyne- and dye-modified NAD^+ analogues **1**–**6** that are applied in this study.

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 fulfil 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

the determination and visualisation of ARTD activity [15-18], the identification of PARylation sites and targets [15,19,20] and the real-time imaging [21] of PARylation processes.

In this report, we systematically compare the substrate scope of the four poly(ADP-ribose)-synthesising enzymes ARTD1, ARTD2, ARTD5 and ARTD6. For this purpose, we tested reporter-modified NAD^+ analogues **1**–**6** (Figure 1) that were previously applied in ARTD1 catalysed ADP-ribosylation [15,17,21]. By investigating them in biochemical assays, we identified sites and sizes of modifications for each enzyme that are well-accepted and competitively used in the presence of natural substrate. In this way, new insights of the enzyme's substrate scope and the applicability of NAD^+ analogues are gained and should thus guide future experiments.

Results and Discussion

Alkyne-modified NAD^+ analogues

First, the position of the reporter group is systematically varied by introducing small, terminal alkyne functionalities at common sites of the adenine base. Upon successful incorporation into PAR, these alkynes serve as handles for copper(I) catalysed azide–alkyne click reaction (CuAAC) [22] with fluorescent dyes. Terminal alkynes are the smallest possible reporter group that allows the selective labelling of poly(ADP-ribose) [17]. As reported, the synthesis of alkyne-modified derivatives **1**–**4** was previously [16,17,23] accomplished by preparing the respective alkyne-modified nucleosides from common precursors and turning them into their corresponding NAD^+ analogues in a two-step procedure (Supporting Information File 1, Scheme S1).

Next, NAD^+ substrate properties were investigated in ADP-ribosylation assays with histone H1.2 as acceptor and in ARTD automodification. For a better comparison, the assay conditions for ARTD2, ARTD5 and ARTD6 were chosen to be similar and were derived from previously established ARTD1 catalysed ADP-ribosylation [21]. Incubation of NAD^+ or NAD^+ analogues with ARTD enzyme in reaction buffer and with or without histone H1.2 as additional acceptor protein were performed at 30 °C to decrease the reported NADase activity of tankyrases [15]. Reaction times were elongated to 1 h, 4 h and 2 h, respectively, to achieve noticeable PAR formation. Moreover, no DNA was added to the tankyrase reactions. Of note, ARTD2 was found to be not activated by short, octameric DNA such as applied in case of ARTD1 and thus activated calf thymus DNA was added to enable ARTD2 catalysed PAR production [24]. After the times indicated, copper-catalysed click conjugations to a fluorophore-containing azide were performed and the reactions were analysed by SDS PAGE. Then, fluorescent signals were detected and compared to the Coomassie Blue stained gels (Figure 2). Each analogue was additionally tested in a 1:1 mixture with natural NAD^+ to explore their competitiveness against natural substrate and all gels contain controls without enzyme. A positive PARylation reaction is indicated by heterogeneous, polymer-modified proteins and/or the reduction of the ARTD band due to automodification. If analogues are successfully incorporated, the polymer chains can additionally be detected in the fluorescence read-out.

For a better comparison, ARTD1-based ADP-ribosylation assays were also performed, because all four analogues have never been tested in parallel before. The outcome of these experiments is summarised in Table 1. For illustration, Figure 3 shows the processing of derivative **1** by all the four ARTDs tested. Of note, it was previously reported [21] that the incubation of proteins with NAD^+ analogues may result in non-enzymatic Schiff base formation of ADP-riboses with lysine residues [25] and can be detected by some minor staining of the involved proteins, which is also visible in some of the investigated reactions.

As expected from the close structural similarity between ARTD1 and ARTD2 (panel a and b), both enzymes behave

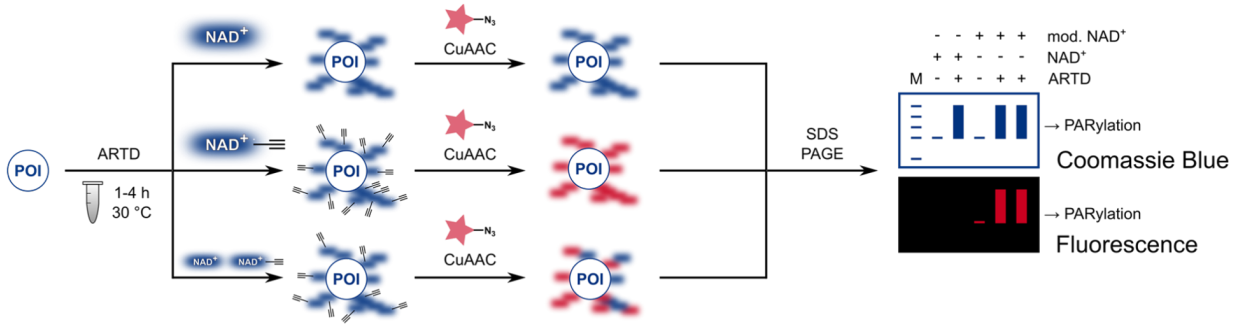


Figure 2: Workflow of the ADP-ribosylation assay. The protein of interest (POI) is ADP-ribosylated by the respective ARTD and by NAD^+, NAD^+ analogue or a 1:1 mixture. Then, copper(I)-catalysed azide–alkyne click reaction (CuAAC) is performed and mixture is resolved by SDS PAGE.

Table 1: Acceptance of alkyne-modified NAD$^+$ analogues **1**–**4** by different ARTDs without or with competition of natural substrate.[a] ✓ = analogue is well processed, (✓) = analogue is processed with lower efficiency, ✗ = analogue is not processed.

NAD$^+$ analogue	Nat. NAD$^+$	ARTD1	ARTD2	ARTD5	ARTD6
1	–	✓	✓	✓	✓
	1:1	✓	✓	✓	✓
2	–	(✓)	(✓)	✓	✓
	1:1	✓	✓	✓	✓
3	–	✗	✗	✗	✗
	1:1	✗	✗	✗	✗
4	–	✗	✗	✗	✗
	1:1	✗	✗	✗	✗

[a]All gels are depicted in Supporting Information File 1, Figure S1 and Figure S2.

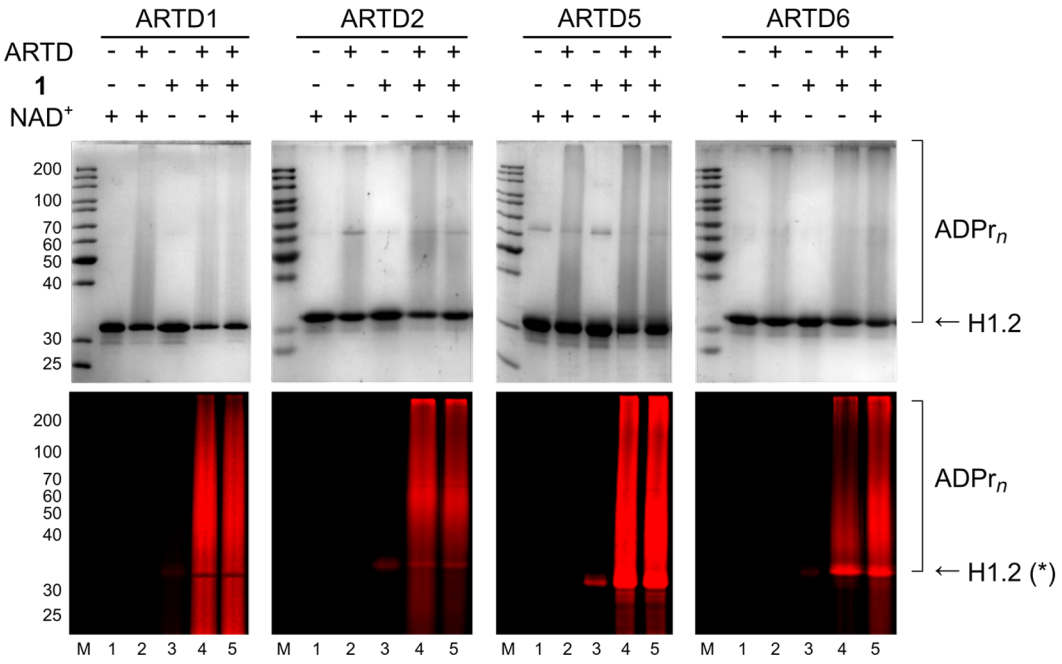


Figure 3: SDS PAGE analysis of ADP-ribosylation of histone H1.2 with ARTD1, ARTD2, ARTD5 and ARTD6 using NAD$^+$ analogue **1**. Upper panel shows Coomassie Blue staining; lower panel shows TMR fluorescence. Experimental details are provided in Supporting Information File 1. *Unspecific staining of H1.2 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-

ribosylation, whereas in case of ARTD6 both are used as substrates in both assays with similar efficiencies.

Dye-modified NAD$^+$ analogues

Because the alkyne-tag induces only small alterations to the NAD$^+$ scaffold, we also investigated how these enzymes would act on bulkier substitutions. For this purpose, we selected bulky, dye-modified NAD$^+$ 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$^+$ (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 automodifcation, 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$^+$ analogues **5** and **6** by different ARTDs without or with competition of natural substrate.[a] ✓ = analogue is well processed, (✓) = analogue is processed with lower efficiency, ✗ = analogue is not processed.

NAD$^+$ analogue	Nat. NAD$^+$	ARTD1	ARTD2	ARTD5	ARTD6
5	–	✓	✓	✗	✓
	1:1	✓	✓	✗	✓
6	–	✗	(✓)[b] ✗	✓[c] ✗	✓
	1:1	✓	✓	✓[c] ✗	✓

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

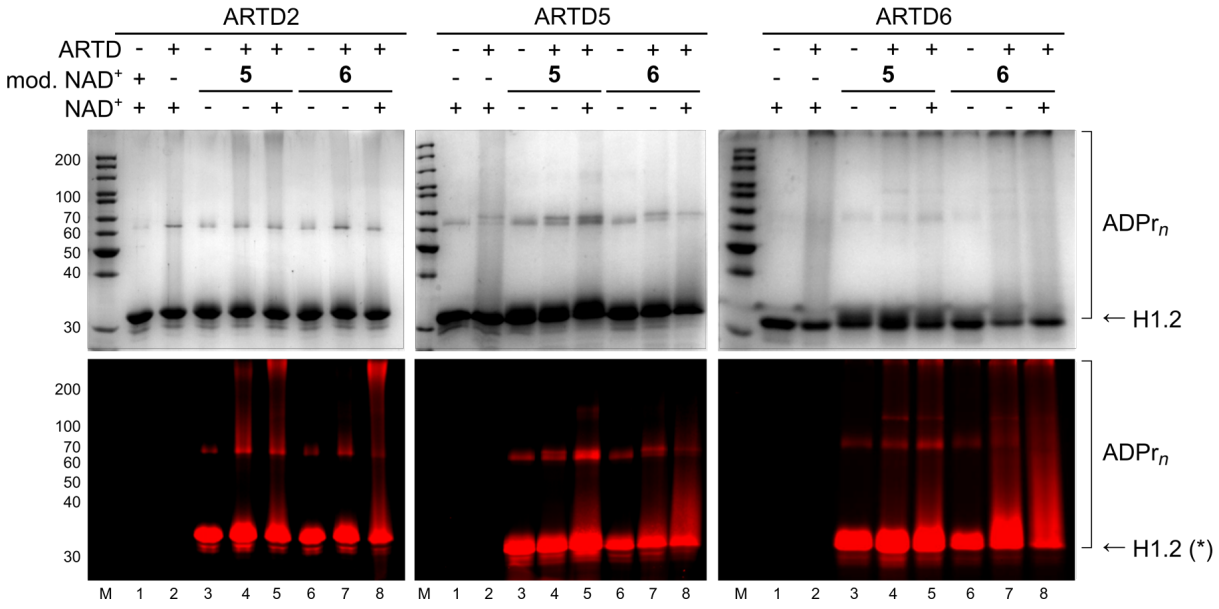


Figure 4: SDS PAGE analysis of ADP-ribosylation of histone H1.2 with ARTD2, ARTD5 and ARTD 6 using NAD$^+$ analogues **5** and **6**. Upper panel shows Coomassie Blue staining; lower panel shows TMR fluorescence. Experimental details are provided in Supporting Information File 1. *High unspecific staining of H1.2 in lanes 3 and 6 results from non-catalytic bond formation of NAD$^+$ 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-ribosylation 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 **5** are processed. Thus, 2-modified analogues are the best choice for the study of these enzymes. On the other hand, 6-modified derivatives should be chosen for the study of the tankyrases ARTD5 and ARTD6. When bulky substitutions are added on the NAD^+ scaffold, tankyrases tolerate better 6-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

1. Hottiger, M. O.; Hassa, P. O.; Lüscher, B.; Schüler, H.; Koch-Nolte, F. *Trends Biochem. Sci.* **2010,** *35,* 208–219. doi:10.1016/j.tibs.2009.12.003
2. Hottiger, M. O. *Mol. Cell* **2015,** *58,* 1134. doi:10.1016/j.molcel.2015.06.001
3. Gibson, B. A.; Kraus, W. L. *Nat. Rev. Mol. Cell Biol.* **2012,** *13,* 411–424. doi:10.1038/nrm3376
4. Ryu, K. W.; Kim, D.-S.; Kraus, W. L. *Chem. Rev.* **2015,** *115,* 2453–2481. doi:10.1021/cr5004248
5. Kraus, W. L. *Mol. Cell* **2015,** *58,* 902–910. doi:10.1016/j.molcel.2015.06.006
6. Virág, L. *Mol. Aspects Med.* **2013,** *34,* 1043–1045. doi:10.1016/j.mam.2013.05.002
7. Bai, P. *Mol. Cell* **2015,** *58,* 947–958. doi:10.1016/j.molcel.2015.01.034
8. Popp, O.; Veith, S.; Fahrer, J.; Bohr, V. A.; Bürkle, A.; Mangerich, A. *ACS Chem. Biol.* **2013,** *8,* 179–188. doi:10.1021/cb300363g
9. Martello, R.; Mangerich, A.; Sass, S.; Dedon, P. C.; Bürkle, A. *ACS Chem. Biol.* **2013,** *8,* 1567–1575. doi:10.1021/cb400170b
10. Mendoza-Alvarez, H.; Chavez-Bueno, S.; Alvarez-Gonzalez, R. *IUBMB Life* **2000,** *50,* 145–149. doi:10.1080/713803695
11. Szántó, M.; Brunyánszki, A.; Kiss, B.; Nagy, L.; Gergely, P.; Virág, L.; Bai, P. *Cell. Mol. Life Sci.* **2012,** *69,* 4079–4092. doi:10.1007/s00018-012-1003-8
12. Simonin, F.; Poch, O.; Delarue, M.; de Murcia, G. *J. Biol. Chem.* **1993,** *268,* 8529–8535.
13. Schreiber, V.; Dantzer, F.; Ame, J.-C.; de Murcia, G. *Nat. Rev. Mol. Cell Biol.* **2006,** *7,* 517–528. doi:10.1038/nrm1963
14. Haikarainen, T.; Krauss, S.; Lehtio, L. *Curr. Pharm. Des.* **2014,** *20,* 6472–6488. doi:10.2174/1381612820666140630101525
15. Jiang, H.; Kim, J. H.; Frizzell, K. M.; Kraus, W. L.; Lin, H. *J. Am. Chem. Soc.* **2010,** *132,* 9363–9372. doi:10.1021/ja101588r
16. Wang, Y.; Rösner, D.; Grzywa, M.; Marx, A. *Angew. Chem., Int. Ed.* **2014,** *53,* 8159–8162. doi:10.1002/anie.201404431
17. Wallrodt, S.; Buntz, A.; Wang, Y.; Zumbusch, A.; Marx, A. *Angew. Chem., Int. Ed.* **2016,** *55,* 7660–7664. doi:10.1002/anie.201600464
18. Bakondi, E.; Bai, P.; Szabó, É.; Hunyadi, J.; Gergely, P.; Szabó, C.; Virág, L. *J. Histochem. Cytochem.* **2002,** *50,* 91–98. doi:10.1177/002215540205000110
19. Carter-O'Connell, I.; Jin, H.; Morgan, R. K.; Zaja, R.; David, L. L.; Ahel, I.; Cohen, M. S. *Cell Rep.* **2016,** *14,* 621–631. doi:10.1016/j.celrep.2015.12.045
20. Gibson, B. A.; Zhang, Y.; Jiang, H.; Hussey, K. M.; Shrimp, J. H.; Lin, H.; Schwede, F.; Yu, Y.; Kraus, W. L. *Science* **2016,** *353,* 45–50. doi:10.1126/science.aaf7865
21. Buntz, A.; Wallrodt, S.; Gwosch, E.; Schmalz, M.; Beneke, S.; Ferrando-May, E.; Marx, A.; Zumbusch, A. *Angew. Chem., Int. Ed.* **2016,** *55,* 11256–11260. doi:10.1002/anie.201605282
22. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002,** *41,* 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4
23. Du, J.; Jiang, H.; Lin, H. *Biochemistry* **2009,** *48,* 2878–2890. doi:10.1021/bi802093g
24. Carter-O'Connell, I.; Jin, H.; Morgan, R. K.; David, L. L.; Cohen, M. S. *J. Am. Chem. Soc.* **2014,** *136,* 5201–5204. doi:10.1021/ja412897a
25. Kun, E.; Chang, A. C.; Sharma, M. L.; Ferro, A. M.; Nitecki, D. *Proc. Natl. Acad. Sci. U. S. A.* **1976,** *73,* 3131–3135. doi:10.1073/pnas.73.9.3131

26. Rippmann, J. F.; Damm, K.; Schnapp, A. *J. Mol. Biol.* **2002,** *323,* 217–224. doi:10.1016/S0022-2836(02)00946-4

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Isoxazole derivatives as new nitric oxide elicitors in plants

Anca Oancea[1], Emilian Georgescu[2], Florentina Georgescu[3], Alina Nicolescu[4,5], Elena Iulia Oprita[1], Catalina Tudora[1], Lucian Vladulescu[3], Marius-Constantin Vladulescu[3], Florin Oancea[6] and Calin Deleanu[*4,5]

Letter

Open Access

Address:
[1]National Institute of Research and Development for Biological Sciences, Spl. Independentei 296, RO-060031 Bucharest, Romania, [2]Research Center Oltchim, Str. Uzinei 1, RO-240050, Ramnicu Valcea, Romania, [3]Research Dept., Teso Spec SRL, Str. Muncii 53, RO-915200 Fundulea, Calarasi, Romania, [4]Petru Poni Institute of Macromolecular Chemistry, Romanian Academy, Aleea Grigore Ghica Voda 41-A, RO-700487 Iasi, Romania, [5]C. D. Nenitescu Centre of Organic Chemistry, Romanian Academy, Spl. Independentei 202-B, RO-060023 Bucharest, Romania and [6]National Research & Development Institute for Chemistry & Petrochemistry – ICECHIM, Spl. Independentei 202, RO-060021 Bucharest, Romania

Email:
Calin Deleanu[*] - calin.deleanu@yahoo.com

* Corresponding author

Keywords:
chemical elicitor; 1,3-dipolar cycloaddition; isoxazole; nitric oxide; nitrile oxide; reactive oxygen species

Beilstein J. Org. Chem. **2017,** *13,* 659–664.
doi:10.3762/bjoc.13.65

Received: 31 December 2016
Accepted: 16 March 2017
Published: 06 April 2017

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

Guest Editor: H. B. Bode

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

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-inflamatory [9,11], antinociceptive [12] and anticancer [13] activities. Several isoxazole derivatives have $GABA_A$ antagonist [14] and T-type Ca^{2+} 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 hydroxyiminoyl 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

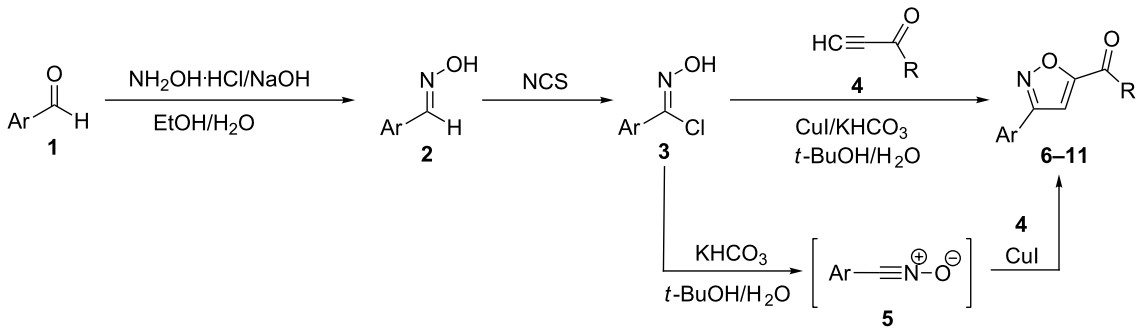
Synthesis of 3,5-disubstituted isoxazole derivatives

The regioselective cooper(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 $KHCO_3$, are added to an aqueous solution containing the crude imidoyl chlorides **3** and the non-symmetrical activated alkynes **4**. The in situ generated aromatic nitrile oxides **5** undergo an addition to copper(I) acetylides 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 recristallysed. The synthesized 3,5-disubstituted isoxazoles are presented in Table 1.

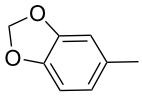
3,5-Disubstituted isoxazoles **6**, **7**, **9** and **10** (Table 1) have been previously prepared by different synthetic procedures, but compounds **6** [52], **7** [49] and **10** [53] have been only partly characterized, while for compound **9** [54] no characterization data have been reported.

3,5-Disubstituted isoxazole structures were unambiguously assigned on the basis of chemical and spectral analysis (IR, ^{1}H, ^{13}C and ^{15}N NMR spectra). NMR spectra clearly indicated the presence of only one regioisomer for all synthesized 3,5-disubstituted isoxazoles. Signals in ^{1}H and ^{13}C 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.

Compound	Ar	R	mp (°)	Yield (%)
6		Me	159–160; 141–142 [51]	81
7		OEt	100–101.5	70
8		Ph	135–137	78
9		Me	113–115	73
10		OEt	78–80	67
11		Ph	90–91	74

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 O_2^-, OH· and H_2O_2, 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 histochemical 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 (H_2DCF 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 3,5-disubstituted isoxazoles followed the series: **9** > **11** > **10** > **8** > **7** = **6** (Figure S3 in Supporting Information File 1).

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, regiospecific 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 fluorescent 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.

Supporting Information

Supporting Information File 1

Experimental procedures, characterization data, IR, ^{1}H, ^{13}C and ^{15}N 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]

Acknowledgments

This work was supported by the Ministry of National Education – Research Activity, UEFISCDI, 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

1. Carlsen, L.; Döpp, D.; Döpp, H.; Duus, F.; Hartmann, H.; Lang-Fugmann, S.; Schulze, B.; Smalley, R. K.; Wakefield, B. J. In *Houben-Weyl Methods in Organic Chemistry;* Schaumann, E., Ed.; Georg Thieme Verlag: Stuttgart, 1992; Vol. E8a, pp 45–204.
2. Banik, U.; Manna, K.; Ghosh, P. S.; Das, M. *Int. J. Institutional Pharm. Life Sci.* **2014,** *4,* 71–78.
3. Lee, Y.-S.; Kim, B. H. *Bioorg. Med. Chem. Lett.* **2002,** *12,* 1395–1397. doi:10.1016/S0960-894X(02)00182-8
4. Srivastava, S.; Bajpai, L. K.; Batra, S.; Bhaduri, A. P.; Maikhuri, J. P.; Gupta, G.; Dhar, J. D. *Bioorg. Med. Chem.* **1999,** *7,* 2607–2613. doi:10.1016/S0968-0896(99)00188-1
5. Pruitt, J. R.; Pinto, D. J.; Estrella, M. J.; Bostrom, L. L.; Knabb, R. M.; Wong, P. C.; Wright, M. R.; Wexler, R. R. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 685–689. doi:10.1016/S0960-894X(00)00097-4
6. Nantermet, P. G.; Barrow, J. C.; Lundell, G. F.; Pellicore, J. M.; Rittle, K. E.; Young, M.; Freidinger, R. M.; Connolly, T. M.; Condra, C.; Karczewski, J.; Bednar, R. A.; Gaul, S. L.; Gould, R. J.; Prendergast, K.; Selnick, H. G. *Bioorg. Med. Chem. Lett.* **2002,** *12,* 319–323. doi:10.1016/S0960-894X(01)00745-4
7. Batra, S.; Srinivasan, T.; Rastogi, S. K.; Kundu, B.; Patra, A.; Bhaduri, A. P.; Dixit, M. *Bioorg. Med. Chem. Lett.* **2002,** *12,* 1905–1908. doi:10.1016/S0960-894X(02)00333-5
8. Batra, S.; Roy, A. K.; Patra, A.; Bhaduri, A. P.; Surin, W. R.; Raghavan, S. A. V.; Sharma, P.; Kapoor, K.; Dikshit, M. *Bioorg. Med. Chem.* **2004,** *12,* 2059–2077. doi:10.1016/j.bmc.2004.02.023
9. Daidone, G.; Raffa, D.; Maggio, B.; Plescia, F.; Cutuli, V. M. C.; Mangano, N. G.; Caruso, A. *Arch. Pharm.* **1999,** *332,* 50–54. doi:10.1002/(SICI)1521-4184(19993)332:2<50::AID-ARDP50>3.0.CO;2-S
10. Talley, J. J. *Prog. Med. Chem.* **1999,** *36,* 201–234. doi:10.1016/S0079-6468(08)70048-1
11. Talley, J. J.; Brown, D. L.; Carter, J. S.; Graneto, M. J.; Koboldt, C. M.; Masferrer, J. L.; Perkins, W. E.; Rogers, R. S.; Shaffer, A. F.; Zhang, Y. Y.; Zweifel, B. S.; Seibert, K. *J. Med. Chem.* **2000,** *43,* 775–777. doi:10.1021/jm990577v
12. Giovannoni, M. P.; Vergelli, C.; Ghelardini, C.; Galeotti, N.; Bartolini, A.; Dal Piaz, V. *J. Med. Chem.* **2003,** *46,* 1055–1059. doi:10.1021/jm021057u
13. Li, W.-T.; Hwang, D.-R.; Chen, C.-P.; Shen, C.-W.; Huang, C.-L.; Chen, T.-W.; Lin, C.-H.; Chang, Y.-L.; Chang, Y.-Y.; Lo, Y.-K.; Tseng, H.-Y.; Lin, C.-C.; Song, J.-S.; Chen, H.-C.; Chen, S.-J.; Wu, S.-H.; Chen, C.-T. *J. Med. Chem.* **2003,** *46,* 1706–1715. doi:10.1021/jm020471r

14. Frølund, B.; Jørgensen, A. T.; Tagmose, L.; Stensbøl, T. B.; Vestergaard, H. T.; Engblom, C.; Kristiansen, U.; Sanchez, C.; Krogsgaard-Larsen, P.; Liljefors, T. *J. Med. Chem.* **2002,** *45,* 2454–2468. doi:10.1021/jm020027o
15. Jung, H. K.; Doddareddy, M. R.; Cha, J. H.; Rhim, H.; Cho, Y. S.; Koh, H. Y.; Jung, B. Y.; Pae, A. N. *Bioorg. Med. Chem.* **2004,** *12,* 3965–3970. doi:10.1016/j.bmc.2004.06.011
16. Monjil, M. S.; Shibata, Y.; Takemoto, D.; Kawakita, K. *Nitric Oxide* **2013,** *29,* 34–45. doi:10.1016/j.niox.2012.12.004
17. Monjil, M. S.; Takemoto, D.; Kawakita, K. *J. Gen. Plant Pathol.* **2014,** *80,* 38–49. doi:10.1007/s10327-013-0493-z
18. Wendehenne, D.; Pugin, A.; Klessig, D. F.; Durner, J. *Trends Plant Sci.* **2001,** *6,* 177–183. doi:10.1016/S1360-1385(01)01893-3
19. Besson-Bard, A.; Pugin, A.; Wendehenne, D. *Annu. Rev. Plant Biol.* **2008,** *59,* 21–39. doi:10.1146/annurev.arplant.59.032607.092830
20. Bandiera, T.; Grünanger, P.; Albini, F. M. *J. Heterocycl. Chem.* **1992,** *29,* 1423–1428. doi:10.1002/jhet.5570290609
21. Cuadrado, P.; González-Nogal, A. M.; Valero, R. *Tetrahedron* **2002,** *58,* 4975–4980. doi:10.1016/S0040-4020(02)00386-1
22. Vicentini, C. B.; Veronese, A. C.; Poli, T.; Guarneri, M.; Giori, P.; Ferretti, V. *J. Heterocycl. Chem.* **1990,** *27,* 1481–1484. doi:10.1002/jhet.5570270555
23. Huisgen, R. In *1,3-Dipolar Cycloaddition Chemistry;* Padwa, A., Ed.; Wiley: New York, 1984; Vol. 1, pp 1–176.
24. Jaeger, V.; Colinas, P. A. Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products. In *Chemistry of Heterocyclic Compounds;* Padwa, A.; Pearson, W. H., Eds.; Wiley: Hoboken, 2002; Vol. 59, pp 363–461.
25. Ajay Kumar, K.; Govindaraju, M.; Jayaroopa, P.; Vasanth Kumar, G. *Int. J. Pharm., Chem. Biol. Sci.* **2012,** *3,* 91–101.
26. Just, G.; Dhal, K. *Tetrahedron* **1968,** *24,* 5251–5269. doi:10.1016/S0040-4020(01)96322-7
27. Grundmann, C.; Dean, J. M. *J. Org. Chem.* **1965,** *30,* 2809–2812. doi:10.1021/jo01019a074
28. Kim, J. N.; Ryu, E. K. *Synth. Commun.* **1990,** *20,* 1373–1377. doi:10.1080/00397919008052851
29. Hassner, A.; Rai, K. M. L. *Synthesis* **1989,** 57–59. doi:10.1055/s-1989-27152
30. Grundmann, C.; Richter, R. *J. Org. Chem.* **1968,** *33,* 476–478. doi:10.1021/jo01265a120
31. Liu, K.-C.; Shelton, B. R.; Howe, R. K. *J. Org. Chem.* **1980,** *45,* 3916–3918. doi:10.1021/jo01307a039
32. Halling, K.; Torssell, K. B. G.; Hazell, R. G. *Acta Chem. Scand.* **1991,** *45,* 736–741. doi:10.3891/acta.chem.scand.45-0736
33. Mukaiyama, T.; Hoshino, T. *J. Am. Chem. Soc.* **1960,** *82,* 5339–5342. doi:10.1021/ja01505a017
34. Xu, J.; Hamme, A. T. *Synlett* **2008,** 919–923. doi:10.1055/s-2008-1042906
35. Tornøe, C. W.; Christiansen, C.; Meldal, M. *J. Org. Chem.* **2002,** *67,* 3057–3064. doi:10.1021/jo011148j
36. Hansen, T. V.; Wu, P.; Fokin, V. V. *J. Org. Chem.* **2005,** *70,* 7761–7764. doi:10.1021/jo050163b
37. Himo, T.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, B. K.; Fokin, V. V. *J. Am. Chem. Soc.* **2005,** *127,* 210–216. doi:10.1021/ja0471525
38. Georgescu, E.; Georgescu, F.; Popa, M. M.; Draghici, C.; Tarko, L.; Dumitrascu, F. *ACS Comb. Sci.* **2012,** *14,* 101–107. doi:10.1021/co2002125
39. Georgescu, E.; Nicolescu, A.; Georgescu, F.; Teodorescu, F.; Marinescu, D.; Macsim, A.-M.; Deleanu, C. *Beilstein J. Org. Chem.* **2014,** *10,* 2377–2387. doi:10.3762/bjoc.10.248
40. Popa, M. M.; Georgescu, E.; Caira, M. R.; Georgescu, F.; Draghici, C.; Stan, R.; Deleanu, C.; Dumitrascu, F. *Beilstein J. Org. Chem.* **2015,** *11,* 1079–1088. doi:10.3762/bjoc.11.121
41. Georgescu, E.; Nicolescu, A.; Georgescu, F.; Shova, S.; Teodorescu, F.; Macsim, A.-M.; Deleanu, C. *Synthesis* **2015,** *47,* 1643–1655. doi:10.1055/s-0034-1380185
42. Georgescu, E.; Nicolescu, A.; Georgescu, F.; Teodorescu, F.; Shova, S.; Marinoiu, A. T.; Dumitrascu, F.; Deleanu, C. *Tetrahedron* **2016,** *72,* 2507–2520. doi:10.1016/j.tet.2016.03.086
43. Paraschivescu, C.C.; Matache, M.; Dobrotă, C.; Nicolescu, A.; Maxim, C.; Deleanu, C.; Fărcăşanu, I. C.; Hădade, N. D. *J. Org. Chem.* **2013,** *78,* 2670–2679. doi:10.1021/jo400023z
44. Fahlbusch, K.-G.; Hammerschmidt, F.-J.; Panten, J.; Pickenhagen, W.; Schatkowski, D.; Bauer, K.; Garbe, D.; Surburg, H. "Flavors and Fragrances". *Ullmann's Encyclopedia of Industrial Chemistry;* Wiley-VCH: Weinheim, 2003.
45. Gupta, O. P.; Nath, A.; Gupta, S. C.; Srivastava, T. N. *Bull. Med. Ethnobot. Res.* **1980,** *1,* 99–106.
46. Rukachaisirikul, T.; Prabpai, S.; Champung, P.; Suksamrarn, A. *Planta Med.* **2002,** *68,* 853–855. doi:10.1055/s-2002-34410
47. Deshpande, S. R.; Nagrale,, S. N.; Patil, M. V.; Chavan, P. S. *Indian J. Pharm. Sci.* **2015,** *77,* 24–33. doi:10.4103/0250-474X.151588
48. Ren, J.; Yang, M.; Liu, H.; Cao, D.; Chen, D.; Li, L.; Tang, L.; He, J.; Chen, Y.-L.; Geng, M.; Xiong, B.; Shen, J. *Org. Biomol. Chem.* **2015,** *13,* 1531–1535. doi:10.1039/C4OB01865F
49. Kamal, A.; Shaik, A. B.; Rao, B. B.; Khan, I.; Kumar, G. B.; Jain, N. *Org. Biomol. Chem.* **2015,** *13,* 10162–10178. doi:10.1039/C5OB01257K
50. Singh, P.; Sharma, P.; Bisetty, K.; Mahajan, M. P. *ARKIVOC* **2011,** *(x),* 55–70. doi:10.3998/ark.5550190.0012.a05
51. Pae, A. N.; Kim, H. Y.; Joo, H. J.; Kim, B. H.; Cho, Y. S.; Choi, K. I.; Choi, J. H.; Koh, H. Y. *Bioorg. Med. Chem. Lett.* **1999,** *9,* 2679–2684. doi:10.1016/S0960-894X(99)00473-4
52. Jiménez, R.; Peréz, L.; Tamariz, J.; Salgado, H. *Heterocycles* **1993,** *35,* 591–598. doi:10.3987/COM-92-S(T)70
53. Bhosale, S.; Kurhade, S.; Prasad, U. V.; Palle, V. P.; Bhuniya, D. *Tetrahedron Lett.* **2009,** *50,* 3948–3951. doi:10.1016/j.tetlet.2009.04.073
54. Inaba, T.; Haas, J.; Shiozaki, M.; Littman, N. M.; Yasue, K.; Andrews, S. W.; Sakai, A.; Fryer, A. M.; Matsuo, T.; Laird, E. R.; Suma, A.; Shinozaki, Y.; Hori, Y.; Imai, H.; Negoro, T. WO 2005058884 A2, June 30, 2005.
55. Delledonne, M.; Xia, Y.; Dixon, R. A.; Lamb, C. *Nature* **1998,** *394,* 585–588. doi:10.1038/29087
56. Kuźniak, E.; Skłodowska, M. *Plant Sci.* **2001,** *160,* 723–731. doi:10.1016/S0168-9452(00)00457-X
57. Pnueli, L.; Liang, H.; Rozenberg, M.; Mittler, R. *Plant J.* **2003,** *34,* 187–203. doi:10.1046/j.1365-313X.2003.01715.x
58. Bryan, N. S.; Grisham, M. B. *Free Radical Biol. Med.* **2007,** *43,* 645–657. doi:10.1016/j.freeradbiomed.2007.04.026
59. Mur, L. A. J.; Mandon, J.; Persijn, S.; Cristescu, S. M.; Moshkov, I. E.; Novikova, G. V.; Hall, M. A.; Harren, F. J. M.; Hebelstrup, K. H.; Gupta, K. J. *AoB Plants* **2013,** *5,* No. pls052. doi:10.1093/aobpla/pls052
60. Weigel, D.; Glazebrook, J. *Arabidopsis: a laboratory manual, Cold Spring Harbor;* Cold Spring Harbor Laboratory Press: New York, 2002; pp 143–170.

61. Srivastava, N.; Gonugunta, V. K.; Puli, M. R.; Raghavendra, A. S. *Planta* **2009,** *229,* 757–765. doi:10.1007/s00425-008-0855-5
62. Chen, X.; Zhong, Z.; Xu, Z.; Chen, L.; Wang, Y. *Free Radical Res.* **2010,** *44,* 587–604. doi:10.3109/10715761003709802
63. Chen, Y.; Mo, H.-Z.; Hu, L.-B.; Li, Y.-Q.; Chen, J.; Yang, L.-F. *PLoS One* **2014,** *9,* e110901. doi:10.1371/journal.pone.0110901
64. Kojima, H.; Nakatsubo, N.; Kikuchi, K.; Kawahara, S.; Kirino, Y.; Nagoshi, H.; Hirata, Y.; Nagano, T. *Anal. Chem.* **1998,** *70,* 2446–2453. doi:10.1021/ac9801723
65. Kojima, H.; Sakurai, K.; Kikuchi, K.; Kawahara, S.; Kirino, Y.; Nagoshi, H.; Hirata, Y.; Nagano, T. *Chem. Pharm. Bull.* **1998,** *46,* 373–375. doi:10.1248/cpb.46.373
66. Kojima, H.; Urano, Y.; Kikuchi, K.; Higuchi, T.; Hirata, Y.; Nagano, T. *Angew. Chem., Int. Ed. Engl.* **1999,** *21,* 3209–3212. doi:10.1002/(SICI)1521-3773(19991102)38:21<3209::AID-ANIE3209>3.0.CO;2-6
67. Kolbert, Z.; Petô, A.; Lehotai, N.; Feigl, G.; Ördög, A.; Erdei, L. *Acta Biol. (Szeged)* **2012,** *56,* 37–41.

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

An eco-compatible strategy for the diversity-oriented synthesis of macrocycles exploiting carbohydrate-derived building blocks

Sushil K. Maurya*[1,2,§] and Rohit Rana[1,2]

Full Research Paper

Open Access

Address:
[1]Natural Product Chemistry and Process Development Division, CSIR- Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, 176 061, India and [2]Academy of Scientific and Innovative Research, CSIR- Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, 176 061, India

Email:
Sushil K. Maurya* - skmaurya@ihbt.res.in

* Corresponding author
§ Phone: +91-1894-230742

Keywords:
carbohydrate; click chemistry; diversity-oriented synthesis; macrocycles; ring-closing metathesis

Beilstein J. Org. Chem. **2017,** *13,* 1106–1118.
doi:10.3762/bjoc.13.110

Received: 27 December 2016
Accepted: 12 May 2017
Published: 09 June 2017

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

Guest Editor: H. B. Bode

© 2017 Maurya and Rana; licensee Beilstein-Institut.
License and terms: see end of document.

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 cycloaddition (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-

ciency virus (HIV); an attractive target for RNA-based drug discovery [13]. Further, macrocyclic glycolipids have shown phosphatase inhibition, cytotoxicity and antiviral activities [12,14]. Generally, the synthesis of these molecules involves a multi-step construction of linear precursors incorporating synthetically compatible functional groups followed by a cyclization in the late stage of the synthesis. The cyclization of the linear precursor is usually achieved by utilizing various ring-closing reactions such as Diels–Alder reactions, [15] aldol reactions, [16] copper-catalyzed azide–alkyne cycloaddition, [17,18] macrolactonization, macrolactamizations, Staudinger ligation or transition-metal-catalyzed coupling reactions [19]. Recently, ring-closing alkyne metathesis (RCAM) [20,21] and ring closing metathesis (RCM) [22-31] have emerged as very powerful tools for macrocyclization including for the preparation of peptidomimetic [17,18,32] glycosides and macrocyclic glycolipids [11]. Similarly, the copper-catalyzed azide–alkyne cycloaddition (CuAAC) reaction has found wide application in medicinal chemistry [33], biology [34,35], polymer chemistry [36], carbohydrates [37-40], peptides [41-44] and in materials science [45-48]. There are several reports wherein different strategies have been developed and used for the synthesis of glycoconjugates [9,49-51], however, the combination of a CuAAC and a RCM reaction has been used very little and rarely combinations of these reactions have been used for the synthesis of sugar-embedded glycoconjugates [52,53]. Further, the linear syntheses of macrocycles based on multistep protocols are not cost-effective and the development of efficient, sustainable, greener and economical methods is highly desired.

Synthetic methods to produce a diverse collection of macrocycles are rare and usually produce only compounds with a similar skeleton [20,33]. However, to achieve a higher hit rate against a broader range of targets libraries of diverse collections of macrocycles are desired [54]. The various diversity elements of a given library should include the molecular size, shape, heteroatoms, functional groups and stereo chemical complexity for selective binding [4]. The diversity-oriented synthesis (DOS), an algorithm in organic chemistry used to generate diverse molecules that include two-directional coupling, ring expansion methods, multidimensional coupling and domain shuffling has been used for the synthesis of small molecules and macrocyclic libraries. Further, several DOS strategies based around build/couple/pair (B/C/P) were developed for the synthesis of compound libraries including macrocycles [18,55]. Carbohydrates as building blocks are inexpensive and easily available commercial products and are well-endowed with functionalities which enable them to establish catalytic sites as well as secondary binding sites [56]. The abundance of various functional groups in the carbohydrate precursor allows for easy access to multiple building blocks by incorporating diversity-oriented synthesis (DOS). These moieties can be easily furnished with alkyne or azide functionality with routine synthetic transformation protocols that allow facile access to mono- 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 [57,58].

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–alkene 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/C/P or B/C/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 alkyne–alkene 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 for coupling 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).

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

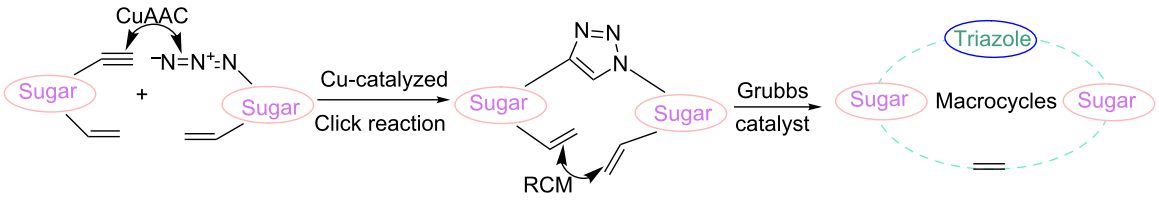


Figure 1: Build-couple-pair (B/C/P) strategy for macrocycles.

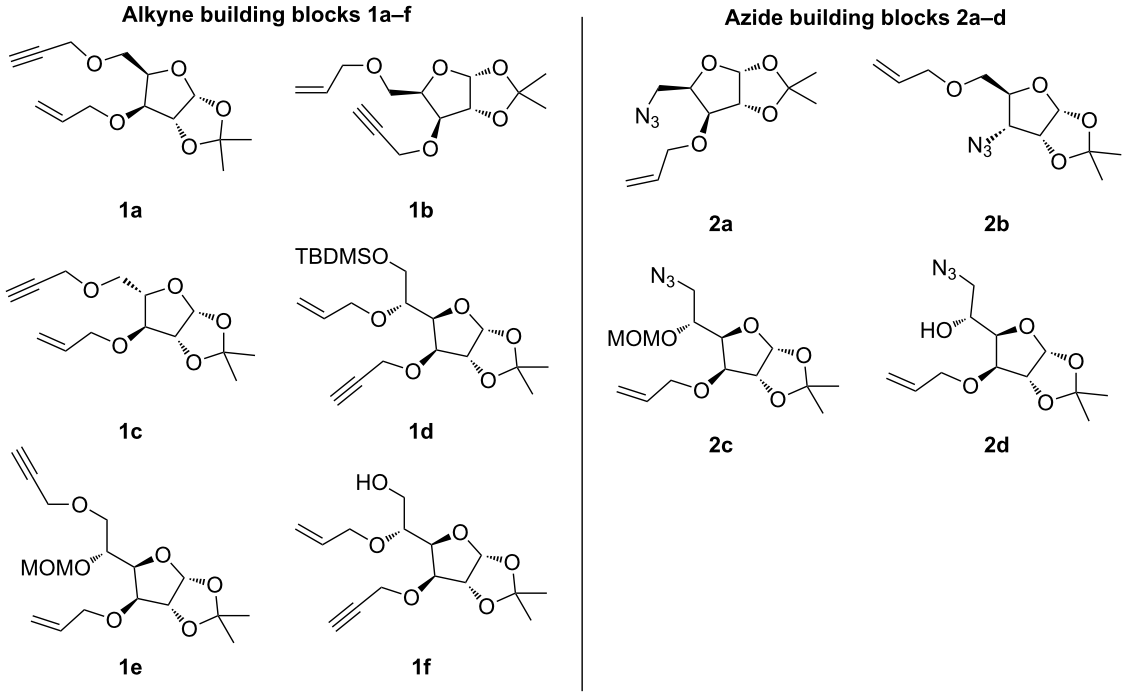
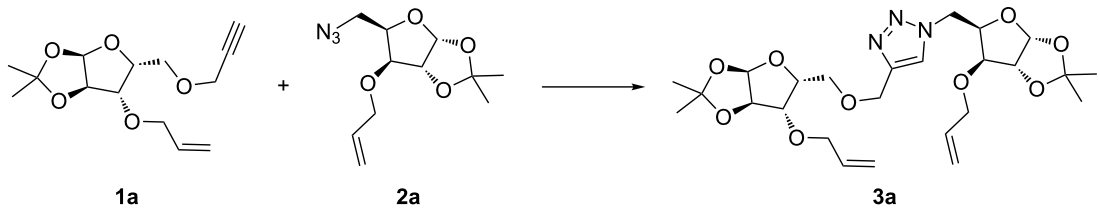


Figure 2: Different building blocks used for DOS.



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

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 regioisomer 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 (**3c**) 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

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

Entry	Solvent	Base	Catalyst (CuI, mol %)	Temperature	Time (hours)	Yield %[a]
1	ACN	TEA	5	ambient	2	65
2	ACN	DIPEA	5	ambient	2	71
3	H_2O	–	–	ambient	24	6[b] (77% selectivity for **3a**)[c]
4	H_2O	–	–	70 °C	24	33[b] (63% selectivity for **3a**)[c]
5	H_2O	–	5	ambient	24	45
6	H_2O	DIPEA	5	ambient	24	35
7	**H_2O**	**–**	**5**	**70 °C**	**2**	**95**
8	H_2O	DIPEA	5	70 °C	2	48

[a]Isolated yield after column chromatography; [b]conversion and [c]product selectivity was measured by ^{1}H NMR.

Table 2: Copper catalyzed azide-alkyne cycloaddition.

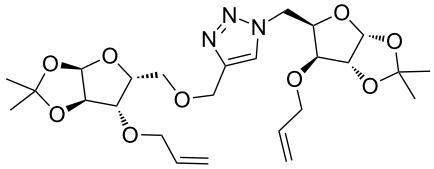
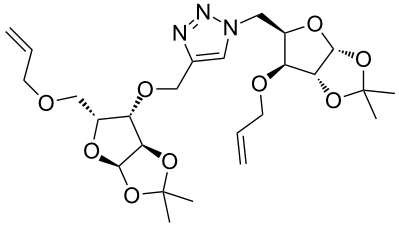
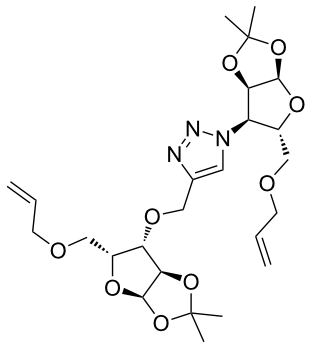
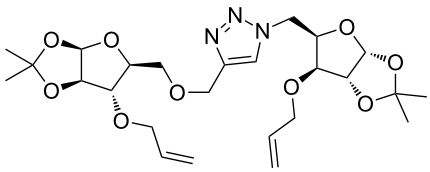
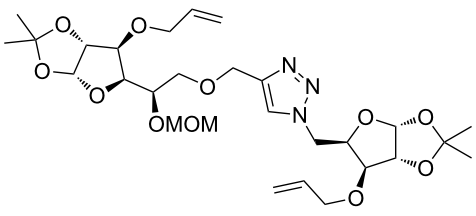
Alkyne	Azide	Cycloadduct[a]	Yield %
1a	**2a**	 **3a**	95
1b	**2a**	 **3b**	94
1b	**2b**	 **3c**	75
1c	**2a**	 **3d**	90
1e	**2a**	 **3e**	76

Table 2: Copper catalyzed azide-alkyne cycloaddition. (continued)

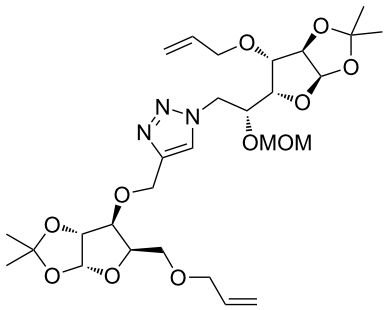
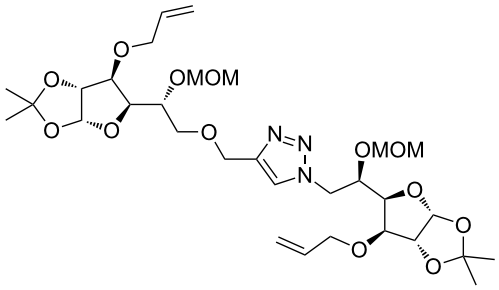
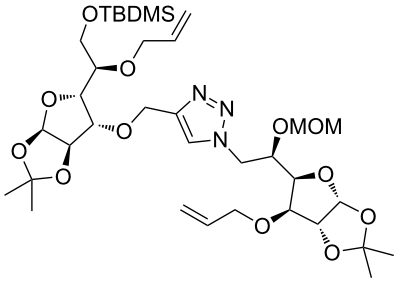
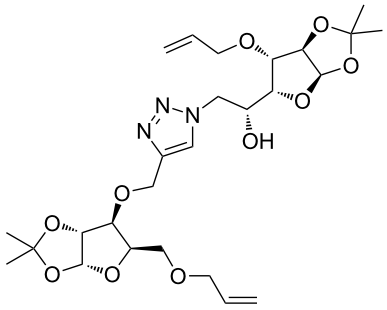
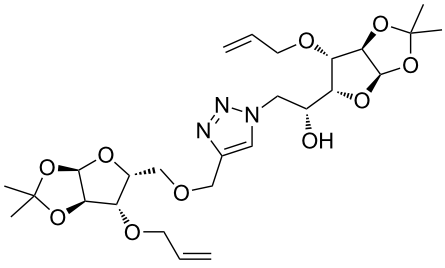
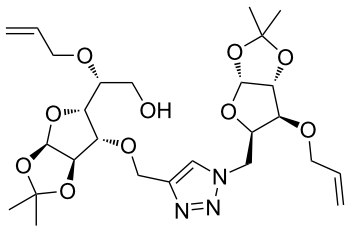
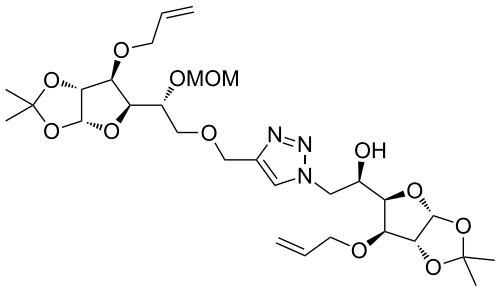
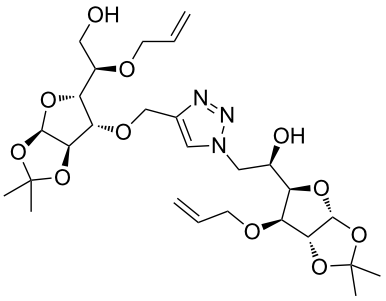
1b	**2c**	 **3f**	78
1e	**2c**	 **3g**	92
1d	**2c**	 **3h**	91
1b	**2d**	 **3i**	85
1a	**2d**	 **3j**	87

Table 2: Copper catalyzed azide-alkyne cycloaddition. (continued)

1f	**2a**	 **3k**	95
1e	**2d**	 **3l**	77
1f	**2d**	 **3m**	67

[a]Method: 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.

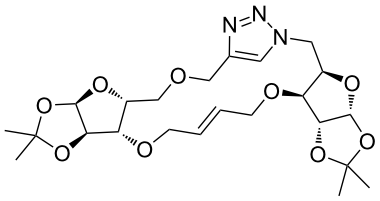
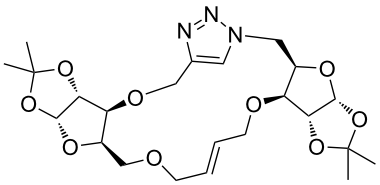
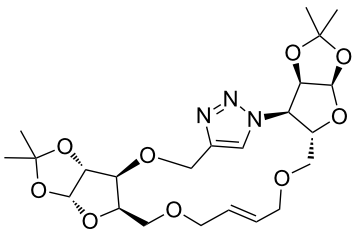
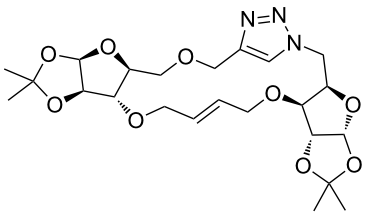
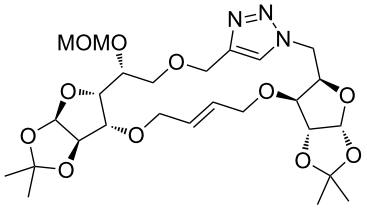
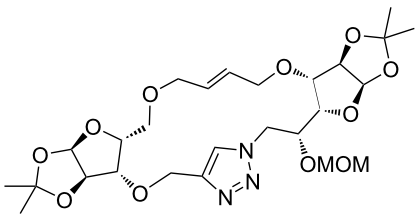
Substrate	Method[a] (mol %; time; yield)	RCM product
3a	A (5; 2 h; 63%) B (5; 2 h; 84%)	 **4a**
3b	A (5; 2 h; 85%) B (5; 2 h; 94%)	 **4b**
3c	A (5; 2 h; 88%) B (5; 3 h; 94%)	 **4c**
3d	A (5; 2 h; 70%) B (5; 2 h; 90%)	 **4d**
3e	A (5+5; 3 h; 88%) B (5+3; 3 h; 40%)	 **4e**
3f	A (5+3; 3 h; 83%) B (5; 2 h; 39%)	 **4f**

Table 3: Application of ring-closing metathesis reactions in the synthesis of macrocycles. (continued)

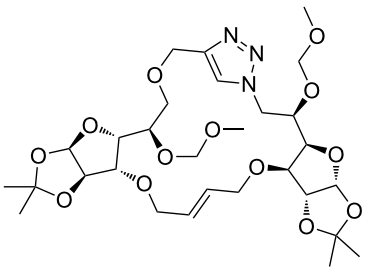
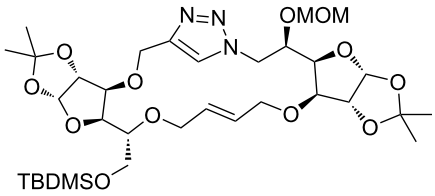
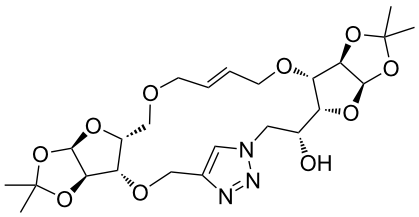
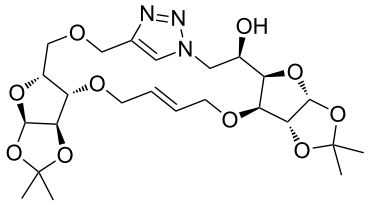
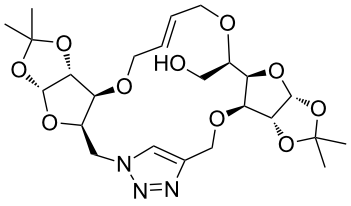
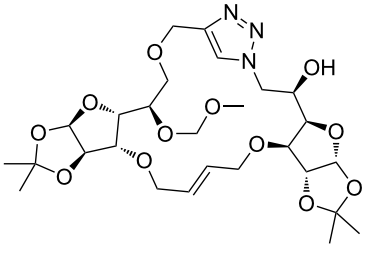
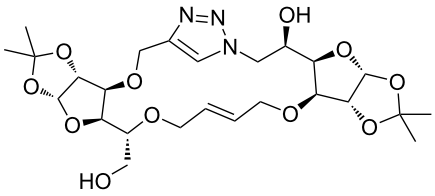
3g	A (5; 2 h; 77%) B (5+3; 3 h; 92%)	 **4g**
3h	A (5+3; 3 h; 82%) B (5; 2 h; 92%)	 **4h**
3i	A (5; 2 h; 95%) B (5; 2 h; 19%)	 **4i**
3j	A (5; 2 h; 84%) B (5; 2 h; 56%)	 **4j**
3k	A (5; 2 h; 81%) B (5; 2 h; 96%)	 **4k**
3l	A (5; 2 h; 53%) B (5; 2 h; 61%)	 **4l**

Table 3: Application of ring-closing metathesis reactions in the synthesis of macrocycles. (continued)

3m	A (5; 2 h; 40%) B (5; 2 h; 55%)	 **4m**

[a]Methods: A: Grubbs second-generation catalyst, CH_2Cl_2, 50 °C; B: Grubbs second-generation catalyst, ethyl acetate, 75 °C.

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 ^{1}H 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.

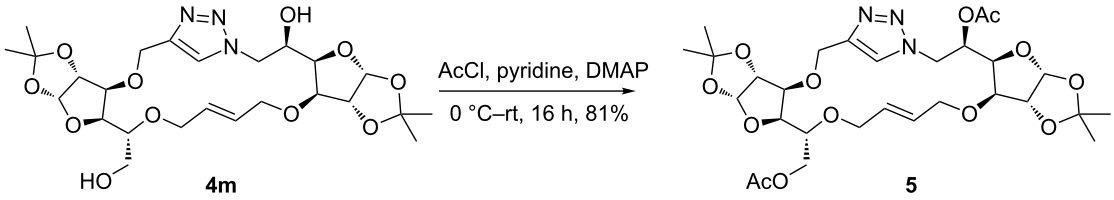
To check the effect of purity of the cycloadduct on the rate and feasibility of subsequent RCM reactions and on isolated yield obtained in the individual steps, we explored the feasibility of the RCM reaction without isolating the product at the couple phase. Compounds **4e**–**h** were synthesized without 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 ^{1}H 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 ^{13}C 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.

RCM product	Two-step protocol[a]				Direct protocol[a]	
	CuAAC Yield (%)	Grubbs cat. (mol %)	RCM yield (%)	Combined yield (%)	Grubbs cat. (mol %)	Yield (%)
4e	76	(5+3)	40	31	(5+5)	32
4f	78	5	39	30	5	29
4g	92	(5+3)	92	85	5	49
4h	91	5	92	84	5	80

[a]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

1. Brandt, W.; Haupt, J. V.; Wessjohann, L. A. *Curr. Top. Med. Chem.* **2010,** *10,* 1361–1379. doi:10.2174/156802610792232060
2. Mallinson, J.; Collins, I. *Future Med. Chem.* **2012,** *4,* 1409–1438. doi:10.4155/fmc.12.93
3. Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug Discovery* **2008,** *7,* 608–624. doi:10.1038/nrd2590
4. Collins, S.; Bartlett, S.; Nie, F.; Sore, H. F.; Spring, D. R. *Synthesis* **2016,** *48,* 1457–1473. doi:10.1055/s-0035-1561414
5. Jarikote, D. V.; Murphy, P. V. *Eur. J. Org. Chem.* **2010,** 4959–4970. doi:10.1002/ejoc.201000491
6. Xie, J.; Bogliotti, N. *Chem. Rev.* **2014,** *114,* 7678–7739. doi:10.1021/cr400035j
7. Madsen, C. M.; Clausen, M. H. *Eur. J. Org. Chem.* **2011,** 3107–3115. doi:10.1002/ejoc.201001715
8. Dörner, S.; Westermann, B. *Chem. Commun.* **2005,** 2852–2854. doi:10.1039/B502682B
9. Adhikary, N. D.; Chattopadhyay, P. *J. Org. Chem.* **2012,** *77,* 5399–5405. doi:10.1021/jo3004327
10. Rao, N. S. J.; Raghunathan, R. *Tetrahedron Lett.* **2015,** *56,* 2669–2673. doi:10.1016/j.tetlet.2015.03.108
11. Fürstner, A. *Eur. J. Org. Chem.* **2004,** 943–958. doi:10.1002/ejoc.200300728
12. Pereda-Miranda, R.; Bah, M. *Curr. Top. Med. Chem.* **2003,** *3,* 111–131. doi:10.2174/1568026033392534
13. Raghunathan, D.; Sánchez-pedregal, V. M.; Junker, J.; Schwiegk, C.; Kalesse, M.; Kirschning, A.; Carlomagno, T. *Nucleic Acids Res.* **2006,** *34,* 3599–3608. doi:10.1093/nar/gkl494
14. Fürstner, A.; Albert, M.; Mlynarski, J.; Matheu, M.; DeClecq, E. *J. Am. Chem. Soc.* **2003,** *125,* 13132–13142. doi:10.1021/ja036521e

15. Nicolaou, K. C.; Snyder, S. A.; Montagnon, T.; Vassilikogiannakis, G. *Angew. Chem., Int. Ed.* **2002,** *41,* 1668–1698. doi:10.1002/1521-3773(20020517)41:10<1668::AID-ANIE1668>3.0.CO;2-Z
16. Casiraghi, G.; Zanardi, F.; Appendino, G.; Rassu, G. *Chem. Rev.* **2000,** *100,* 1929–1972. doi:10.1021/cr990247i
17. Xu, L.; Li, Y.; Li, Y. *Asian J. Org. Chem.* **2014,** *3,* 582–602. doi:10.1002/ajoc.201300245
18. Isidro-Llobet, A.; Georgiou, K. H.; Galloway, W. R. J. D.; Giacomini, E.; Hansen, M. R.; Méndez-Abt, G.; Tan, Y. S.; Carro, L.; Sore, H. F.; Spring, D. R. *Org. Biomol. Chem.* **2015,** *13,* 4570–4580. doi:10.1039/C5OB00371G
19. Wells, J. A.; McClendon, C. L. *Nature* **2007,** *450,* 1001–1009. doi:10.1038/nature06526
20. Beckmann, H. S. G.; Nie, F.; Hagerman, C. E.; Johansson, H.; Tan, Y. S.; Wilcke, D.; Spring, D. R. *Nat. Chem.* **2013,** *5,* 861–867. doi:10.1038/nchem.1729
21. Hansen, E. C.; Lee, D. *J. Am. Chem. Soc.* **2004,** *126,* 15074–15080. doi:10.1021/ja045422d
22. Ghosh, A. K.; Bilcer, G.; Harwood, C.; Kawahama, R.; Shin, D.; Hussain, K. A.; Hong, L.; Loy, J. A.; Nguyen, C.; Koelsch, G.; Ermolieff, J.; Tang, J. *J. Med. Chem.* **2001,** *44,* 2865–2868. doi:10.1021/jm0101803
23. Trnka, T. M.; Grubbs, R. H. *Acc. Chem. Res.* **2001,** *34,* 18–29. doi:10.1021/ar000114f
24. Leeuwenburgh, M. A.; Van der Marel, G. A.; Overkleeft, H. S. *Curr. Opin. Chem. Biol.* **2003,** *7,* 757–765. doi:10.1016/j.cbpa.2003.10.009
25. Schrock, R. R.; Hoveyda, A. H. *Angew. Chem., Int. Ed.* **2003,** *42,* 4592–4633. doi:10.1002/anie.200300576
26. Grubbs, R. H. *Tetrahedron* **2004,** *60,* 7117–7140. doi:10.1016/j.tet.2004.05.124
27. Deiters, A.; Martin, S. F. *Chem. Rev.* **2004,** *104,* 2199–2238. doi:10.1021/cr0200872
28. Hong, S. H.; Sanders, D. P.; Lee, C. W.; Grubbs, R. H. *J. Am. Chem. Soc.* **2005,** *127,* 17160–17161. doi:10.1021/ja052939w
29. Chen, G.-w.; Kirschning, A. *Chem. – Eur. J.* **2002,** *8,* 2717–2729. doi:10.1002/1521-3765(20020617)8:12<2717::AID-CHEM2717>3.0.CO;2-P
30. Monfette, S.; Fogg, D. E. *Chem. Rev.* **2009,** *109,* 3783–3816. doi:10.1021/cr800541y
31. Pérez de Vega, M. J.; García-Aranda, M. I.; Gonáalez-Muñiz, R. *Med. Res. Rev.* **2011,** *31,* 677–715. doi:10.1002/med.20199
32. Isidro-Llobet, A.; Murillo, T.; Bello, P.; Cilibrizzi, A.; Hodgkinson, J. T.; Galloway, W. R. J. D.; Bender, A.; Welch, M.; Spring, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 6793–6798. doi:10.1073/pnas.1015267108
33. Marsault, E.; Peterson, M. L. *J. Med. Chem.* **2011,** *54,* 1961–2004. doi:10.1021/jm1012374
34. Alvarez, R.; Velazquez, S.; San-Felix, A.; Aquaro, S.; De Clercq, E.; Perno, C.-F.; Karlsson, A.; Balzarini, J.; Camarasa, M. J. *J. Med. Chem.* **1994,** *37,* 4185–4194. doi:10.1021/jm00050a015
35. Genin, M. J.; Alwine, D. A.; Anderson, D. J.; Barbachyn, M. R.; Emmert, D. E.; Garmon, S. A.; Graber, D. R.; Grega, K. C.; Hester, J. B.; Hutchinson, D. K.; Morris, J.; Reischer, R. J.; Ford, C. W.; Zurenko, G. E.; Hamel, J. C.; Schaadt, R. D.; Stapert, D.; Yagi, B. H. *J. Med. Chem.* **2000,** *43,* 953–970. doi:10.1021/jm990373e
36. Laurent, B. A.; Grayson, S. M. *J. Am. Chem. Soc.* **2006,** *128,* 4238–4239. doi:10.1021/ja0585836
37. Tiwari, V. K.; Mishra, B. B.; Mishra, K. B.; Mishra, N.; Singh, A. S.; Chen, X. *Chem. Rev.* **2016,** *116,* 3086–3240. doi:10.1021/acs.chemrev.5b00408
38. Chandrasekhar, S.; Rao, C. L.; Nagesh, C.; Reddy, C. R.; Sridhar, B. *Tetrahedron Lett.* **2007,** *48,* 5869–5872. doi:10.1016/j.tetlet.2007.06.062
39. Jogula, S.; Dasari, B.; Khatravath, M.; Chandrasekar, G.; Kitambi, S. S.; Arya, P. *Eur. J. Org. Chem.* **2013,** 5036–5040. doi:10.1002/ejoc.201300548
40. Potopnyk, M. A.; Jarosz, S. "Sweet" Sucrose Macrocycles via a "Click Chemistry" Route. In *Click Chemistry in Glycoscience;* Witczak, Z. J.; Bielski, R., Eds.; Wiley, 2013; p 235. doi:10.1002/9781118526996.ch9
41. Kolb, H. C.; Sharpless, K. B. *Drug Discovery Today* **2003,** *8,* 1128–1137. doi:10.1016/S1359-6446(03)02933-7
42. Whiting, M.; Tripp, J. C.; Lin, Y.-C.; Lindstrom, W.; Olson, A. J.; Elder, J. H.; Sharpless, K. B.; Fokin, V. V. *J. Med. Chem.* **2006,** *49,* 7697–7710. doi:10.1021/jm060754+
43. Wilkinson, B. L.; Bornaghi, L. F.; Houston, T. A.; Poulsen, S.-A. In *Drug Design Research Perspectives;* Kaplan, S. P., Ed.; Nova: Hauppauge: NY, 2007; p 57.
44. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003,** *125,* 3192–3193. doi:10.1021/ja021381e
45. Hawker, C. J.; Fokin, V. V.; Finn, M. G.; Sharpless, K. B. *Aust. J. Chem.* **2007,** *60,* 381–383. doi:10.1071/CH07107
46. Mulla, K.; Shaik, H.; Thompson, D. W.; Zhao, Y. *Org. Lett.* **2013,** *15,* 4532–4535. doi:10.1021/ol402093a
47. Hui, P.; Chandrasekar, R. *Adv. Mater.* **2013,** *25,* 2963–2967. doi:10.1002/adma.201300540
48. Busseron, E.; Coutrot, F. *J. Org. Chem.* **2013,** *78,* 4099–4106. doi:10.1021/jo400414f
49. Campo, V. L.; Ivanova, I. M.; Carvalho, I.; Lopes, C. D.; Carneiro, Z. A.; Saalbach, G.; Schenkman, S.; da Silva, J. S.; Nepogodiev, S. A.; Field, R. A. *Tetrahedron* **2015,** *71,* 7344–7353. doi:10.1016/j.tet.2015.04.085
50. Grimwood, M. E.; Hansen, H. C. *Tetrahedron* **2009,** *65,* 8132–8138. doi:10.1016/j.tet.2009.07.088
51. Ghorai, A.; Padmanaban, E.; Mukhopadhyay, C.; Achari, B.; Chattopadhyay, P. *Chem. Commun.* **2012,** *48,* 11975–11977. doi:10.1039/c2cc36566a
52. Dörner, S.; Westermann, B. *Chem. Commun.* **2005,** 2852–2854. doi:10.1039/b502682b
53. Elchert, B.; Li, J.; Wang, J.; Hui, Y.; Rai, R.; Ptak, R.; Ward, P.; Takemoto, J. Y.; Bensaci, M.; Chang, C.-W. T. *J. Org. Chem.* **2004,** *69,* 1513–1523. doi:10.1021/jo035290r
54. Ciardiello, J. J.; Galloway, W. R. J. D.; O'Connor, C. J.; Sore, H. F.; Stokes, J. E.; Wu, Y.; Spring, D. R. *Tetrahedron* **2016,** *72,* 3567–3578. doi:10.1016/j.tet.2015.10.061
55. Nie, F.; Kunciw, D. L.; Wilcke, D.; Stokes, J. E.; Galloway, W. R. J. D.; Bartlett, S.; Sore, H. F.; Spring, D. R. *Angew. Chem., Int. Ed.* **2016,** *55,* 11139–11143. doi:10.1002/anie.201605460
56. Rapi, Z.; Ozohanics, O.; Tóth, G.; Bakó, P.; Höfler, L.; Nemcsok, T.; Kánya, N.; Keglevich, G. *J. Inclusion Phenom. Macrocyclic Chem.* **2016,** *85,* 19–32. doi:10.1007/s10847-016-0601-8
57. Skowerski, K.; Białecki, J.; Tracz, A.; Olszewski, T. K. *Green Chem.* **2014,** *16,* 1125–1130. doi:10.1039/C3GC41943F and references cited therein.
58. Alder, C. M.; Hayler, J. D.; Henderson, R. K.; Redman, A. M.; Shukla, L.; Shuster, L. E.; Sneddon, H. F. *Green Chem.* **2016,** *18,* 3879–3890. doi:10.1039/C6GC00611F

59. Nielsen, T. E.; Schreiber, S. L. *Angew. Chem., Int. Ed.* **2008,** *47,* 48–56. doi:10.1002/anie.200703073
60. Morton, D.; Leach, S.; Cordier, C.; Warriner, S.; Nelson, A. *Angew. Chem., Int. Ed.* **2009,** *48,* 104–109. doi:10.1002/anie.200804486
61. Marcaurelle, L. A.; Comer, E.; Dandapani, S.; Duvall, J. R.; Gerard, B.; Kesavan, S.; Lee IV, M. D.; Liu, H.; Lowe, J. T.; Marie, J.-C.; Mulrooney, C. A.; Pandya, B. A.; Rowley, A.; Ryba, T. D.; Suh, B.-C.; Wei, J.; Young, D. W.; Akella, L. B.; Ross, N. T.; Zhang, Y.-L.; Fass, D. M.; Reis, S. A.; Zhao, W.-N.; Haggarty, S. J.; Palmer, M.; Foley, M. A. *J. Am. Chem. Soc.* **2010,** *132,* 16962–16976. doi:10.1021/ja105119r
62. Hajduk, P. J.; Galloway, W. R. J. D.; Spring, D. R. *Nature* **2011,** *470,* 42–43. doi:10.1038/470042a
63. Maurya, S. K.; Dow, M.; Warriner, S.; Nelson, A. *Beilstein J. Org. Chem.* **2013,** *9,* 775–785. doi:10.3762/bjoc.9.88
64. Van Maarseveen, J. H.; Horne, W. S.; Ghadiri, M. R. *Org. Lett.* **2005,** *7,* 4503–4506. doi:10.1021/ol0518028
65. Oh, K.; Guan, Z. *Chem. Commun.* **2006,** 3069–3071. doi:10.1039/B606185K
66. Whiting, M.; Muldoon, J.; Lin, Y.-C.; Silverman, S. M.; Lindstrom, W.; Olson, A. J.; Kolb, H. C.; Finn, M. G.; Sharpless, K. B.; Elder, J. H.; Fokin, V. V. *Angew. Chem., Int. Ed.* **2006,** *45,* 1435–1439. doi:10.1002/anie.200502161

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

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

2-Methyl-2,4-pentanediol (MPD) boosts as detergent-substitute the performance of ß-barrel hybrid catalyst for phenylacetylene polymerization

Julia Kinzel[1,§], Daniel F. Sauer[2,§], Marco Bocola[1], Marcus Arlt[1], Tayebeh Mirzaei Garakani[1,3], Andreas Thiel[2], Klaus Beckerle[2], Tino Polen[4], Jun Okuda[2] and Ulrich Schwaneberg[*1,3]

Full Research Paper

Open Access

Address:
[1]Institute of Biotechnology, RWTH Aachen University, Worringer Weg 3, 52074 Aachen, Germany, [2]Institut für Anorganische Chemie, RWTH Aachen University, Landoltweg 1, 52074 Aachen, Germany, [3]DWI - Leibniz Institute for Interactive Materials e.V., Forckenbeckstr. 50, 52056, Aachen, Germany and [4]Institute of Bio- und Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Email:
Ulrich Schwaneberg[*] - u.schwaneberg@biotec.rwth-aachen.de

* Corresponding author
§ Shared first authorship

Keywords:
amphiphilic molecule 2-methyl-2,4-pentanediol; hybrid catalyst; phenylacetylene polymerization; refolding agents; transmembrane protein FhuA

Beilstein J. Org. Chem. **2017,** *13,* 1498–1506.
doi:10.3762/bjoc.13.148

Received: 31 March 2017
Accepted: 13 July 2017
Published: 31 July 2017

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

Guest Editor: H. B. Bode

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

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 polyethyleneglycol, 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 metathease 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–polyethyleneglycol (PE–PEG), sugar glycosides or polyoxyethylenes 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

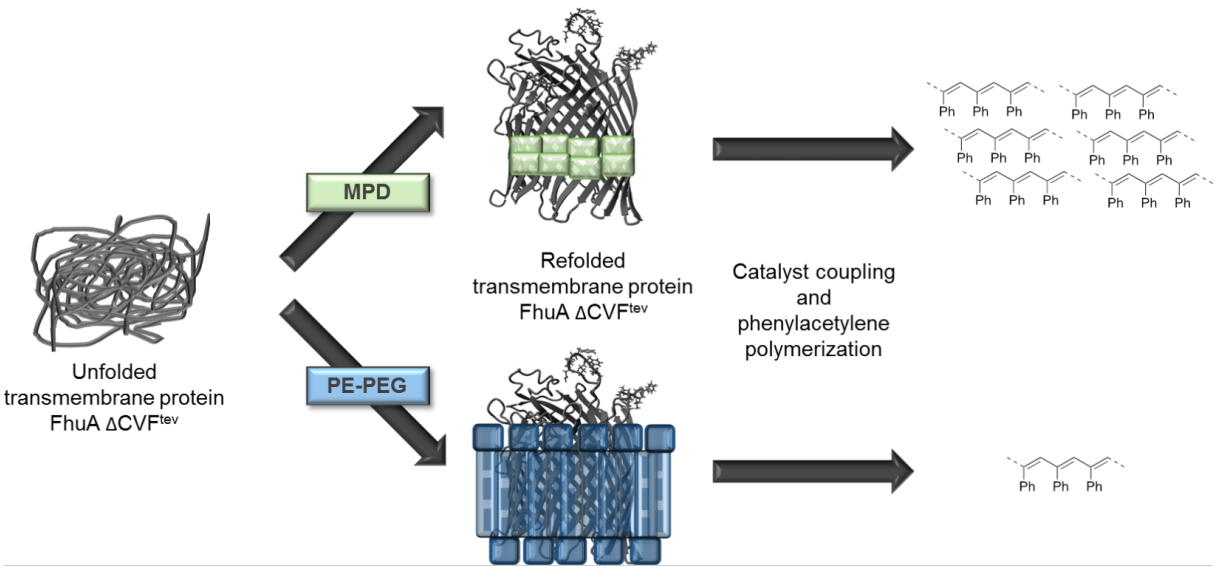


Figure 1: Phenylacetylene polymerization of FhuA ΔCVF^{tev} [29] refolded in a polymer or small amphiphilic molecule. Refolding agents are essential to solubilize transmembrane proteins and keep membrane proteins refolded by shielding hydrophobic residues in aqueous environments. Refolding of the open channel protein FhuA ΔCVF^{tev} was on one hand achieved with polyethylene–polyethyleneglycol (PE–PEG), which is a micelle-forming detergent. In contrast, refolding of FhuA ΔCVF^{tev} with 2-methyl-2,4-pentanediol (MPD) prevents micelle formation and leads to increased yield and molecular weight of the corresponding polymer.

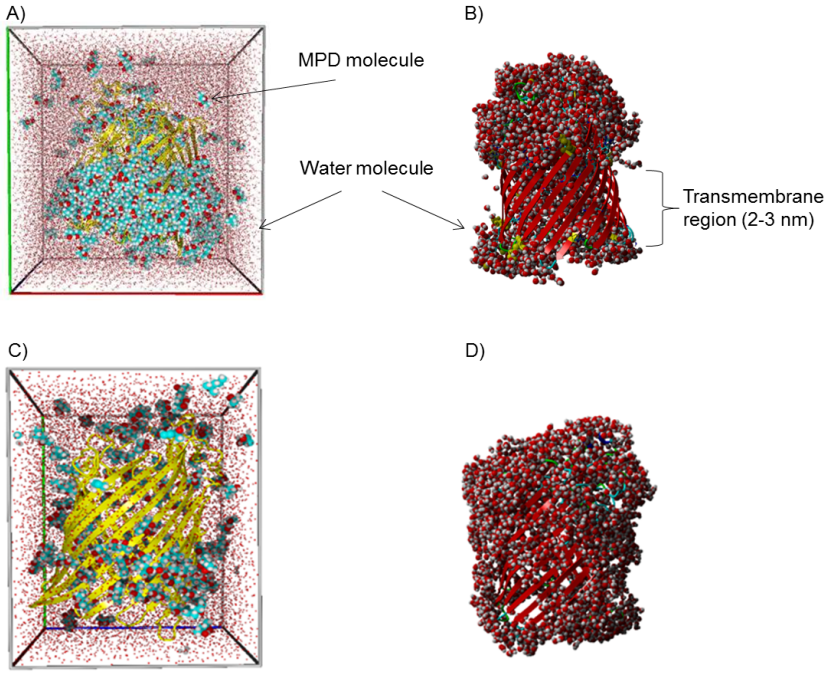


Figure 2: Hydrophobic transmembrane region of FhuA ΔCVF^{tev} [29] stabilized by ≈200 MPD molecules. MPD is illustrated as mainly cyan molecules, water molecules are mainly red. A) A belt of 209 MPD molecules is located close to the transmembrane area. FhuA ΔCVF^{tev} with 22,374 water and 378 MPD molecules 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^{tev} are shown to visualize that the transmembrane area of FhuA ΔCVF^{tev} is completely water free in MD simulations using 378 MPD molecules. C) MD simulations of FhuA ΔCVF^{tev} 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^{tev} using 126 MPD molecules. MPD, 2-methyl-2,4-pentanediol.

FhuA ΔCVF^{tev} 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^{tev} 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^{tev} 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^{tev} 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^{tev} (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^{tev} is correctly folded even up to eight weeks.

Coupling efficiency of the rhodium catalyst to FhuA ΔCVF^{tev} is more than 90%

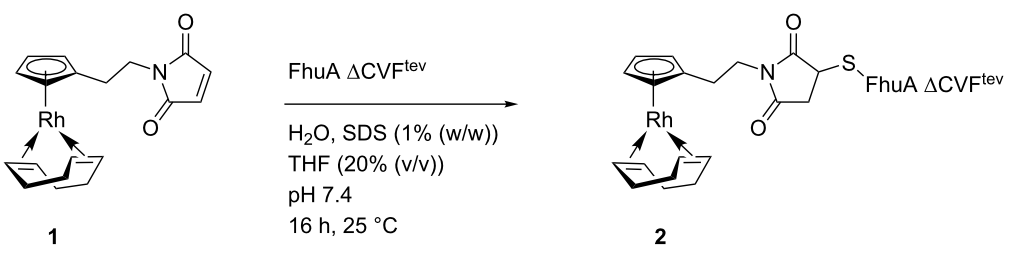
The rhodium catalyst **1** bearing a maleimide group was attached to FhuA ΔCVF^{tev} for the generation of the biohybrid catalyst [Rh]-FhuA ΔCVF^{tev} **2** as previously reported for the Grubbs–Hoveyda type [29,47] or copper complexes [10] (Scheme 1).

FhuA ΔCVF^{tev} was dissolved in a solution containing 1.25% SDS. The state of FhuA ΔCVF^{tev} 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.[a]

Refolding agent	Activity	Selectivity	Comment
SDS [33]	++	−	unfolding property
oPOE [19,29]	+	+	costly, micelle formation
PE–PEG [29]	+	++	bulky, micelle formation
MPD [33,34]	++	++	small, amphipathic alcohol, water-miscible

[a]SDS, sodium dodecyl sulfate; oPOE, octylpolyoxyethylene; PE–PEG, polyethylene–polyethyleneglycol; MPD, 2-methyl-2,4-pentanediol. ++, very good; +, beneficial; −, non-beneficial.



Scheme 1: Coupling of [Rh]-**1** to the open channel protein FhuA ΔCVF^{tev}. 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 ΔCVF^{tev} 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 ΔCVF^{tev} with both reagents.

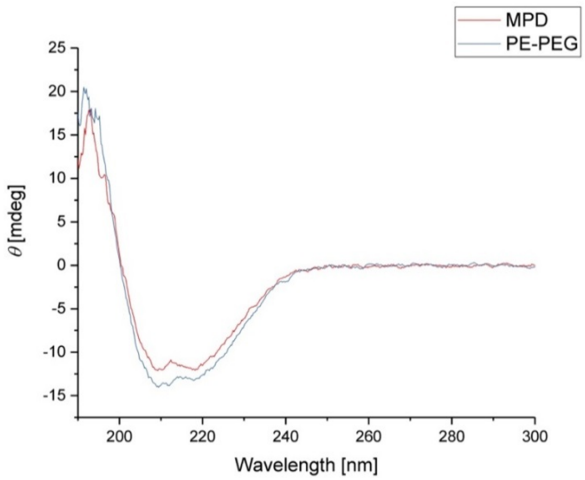


Figure 3: Circular dichroism spectra of **2** refolded in 2-methyl-2,4-pentanediol (MPD, red) and polyethylene–polyethyleneglycol (PE–PEG, blue).

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 ΔCVF^{tev} 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 ΔCVF^{tev}, the addition of water to the maleimide ring was observed [10,47]. During digestion or ionization, also cleavage

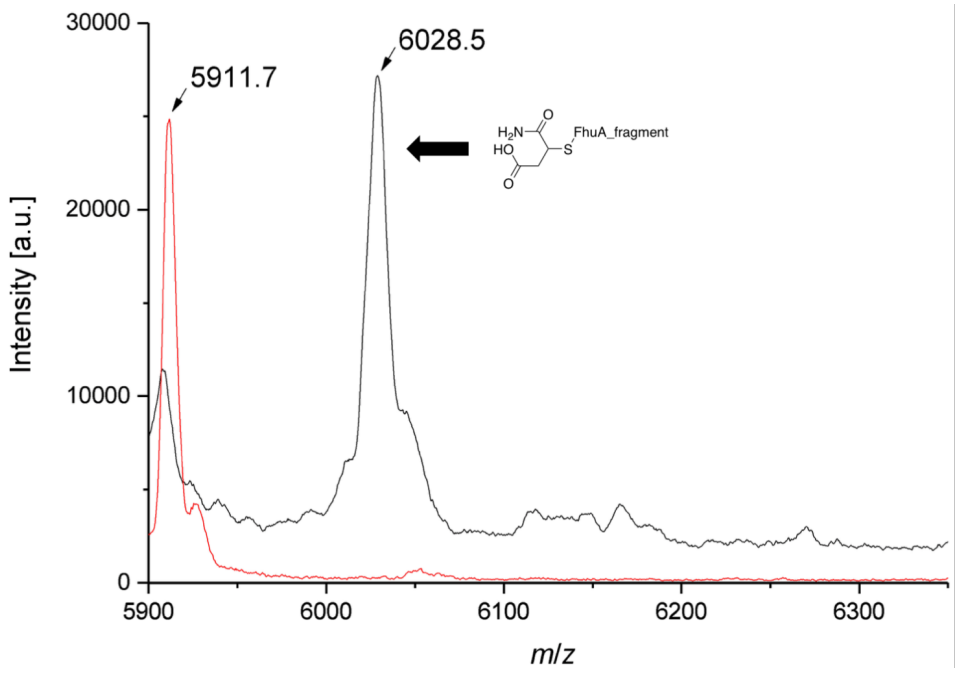


Figure 4: MALDI–TOF mass spectra of apo FhuA ΔCVF^{tev} (red; calculated *m*/*z* = 5902.6; found: *m*/*z* = 5911.7) and **2** (black; *m*/*z* = 6028.5 is assigned to the FhuA fragment containing the maleimide function after water addition). Possible fragmentation of the [Rh] catalyst is indicated. FhuA ΔCVF^{tev} was analyzed after digestion by protease from Tobacco Etch Virus (TEV).

of the amide bond occurs and therefore the metal cannot be observed.

Polymerization of phenylacetylene

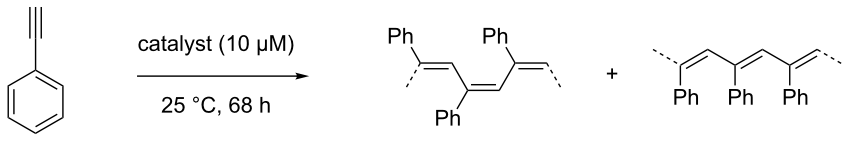
The synthesized and characterized biohybrid conjugate based on FhuA ΔCVF^{tev} was used to polymerize phenylacetylene (**3,** Table 2).

Polymerization of phenylacetylene in THF at 25 °C yields in 65% polymer with M_n = 5,300 and a high *cis*-content of 90% (Table 2, entry 1). If the rhodium catalyst is not present, FhuA ΔCVF^{tev} 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^{tev} 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_n 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_n = 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^{tev} 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^{tev} 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**.[a]



3 *cis*-**4** *trans*-**4**

Entry[b]	Catalyst	Stabilization agent[c]	Isolated yield (%)	M_n[d] (g/mol)	PDI[d]	*trans*/*cis*[e]
1[f]	**1**		13 mg (65)	5,300	4.6	10:90
2[g]	FhuA ΔCVF^{tev}	PE–PEG or MPD	–	–	–	–
3[g]	**2**	PE–PEG	<1 mg (5)	800	6.0	70:30
4[g]	**2**	MPD	10 mg (52)	5,500	2.9	75:25

[a]THF, tetrahydofuran, PE–PEG, polyethylene–polyethylene glycol; MPD, 2-methyl-2,4-pentanediol. [b]Buffer: Water containing NaP_i (pH 8, 10 mM) and EDTA (1 mM). *c*(**3**) = 0.1 M; V_{total} = 2 mL. [c]*c*(PE–PEG) = 0.125 mM; *c*(MPD) = 50 mM. [d]Determined by GPC. [e]Determined with ^{1}H NMR spectroscopy. [f]Reaction in THF. [g]Reaction in buffer, containing 10% (v/v) THF.

ΔCVF^{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 small-molecule stabilizer for utilization of the biohybrid catalyst **2** in phenylacetylene polymerization. The small detergent MPD stabilizes the transmembrane protein FhuA ΔCVF^{tev} in aqueous solution without forming micelles. The structural integrity was proven by CD spectroscopy. Applying MPD as 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 ΔCVF^{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

General comments

All used chemicals used 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 ΔCVF^{tev}

Expression of FhuA ΔCVF^{tev} from T7 expression vector pPR-IBA1 was performed using the *E. coli* B^E BL21 (DE3) omp8 strain as expression host according to previous descriptions [29,34,51]. FhuA ΔCVF^{tev} was extracted from *E. coli* with SDS as solubilizing agent as described previously [19,29,34]. Refolding of FhuA ΔCVF^{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 ΔCVF^{tev} with TEV protease

For analysis of the modification of Cys545 of FhuA ΔCVF^{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 ΔCVF^{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 were introduced using YASARA Structure 13.6.13 as described previously [29] in a detergent-membrane model stabilized by octylpolyoxyethylene (n = 5). To study the interactions of the membrane protein variant FhuA ΔCVF^{tev} with the amphiphilic stabilizing agent MPD, FhuA ΔCVF^{tev} was solvated in a periodic box (size 79.57 × 89.35 × 64.87 or 95.49 × 82.38 × 105.82 with α, β and γ = 90.00°) filled with 12,208 or 22,374 TIP3P

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^{tev} (5–6 mg/mL) in aqueous SDS solution (1% (w/w) SDS, pH 8 adjusted with $NaHCO_3$), 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 ^{1}H NMR and GPC as reported previously [12,13].

Supporting Information

Supporting Information File 1

CD spectra of unmodified FhuA ΔCVF^{tev} 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

1. Steinreiber, J.; Ward, T. R. *Coord. Chem. Rev.* **2008,** *252,* 751–766. doi:10.1016/j.ccr.2007.09.016
2. Lewis, J. C. *ACS Catal.* **2013,** *3,* 2954–2975. doi:10.1021/cs400806a
3. Sauer, D. F.; Gotzen, S.; Okuda, J. *Org. Biomol. Chem.* **2016,** *14,* 9174–9183. doi:10.1039/C6OB01475E
4. Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. *Nature* **2016,** *537,* 661–665. doi:10.1038/nature19114
5. Collot, J.; Gradinaru, J.; Humbert, N.; Skander, M.; Zocchi, A.; Ward, T. R. *J. Am. Chem. Soc.* **2003,** *125,* 9030–9031. doi:10.1021/ja035545i
6. Zimbron, J. M.; Heinisch, T.; Schmid, M.; Hamels, D.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. *J. Am. Chem. Soc.* **2013,** *135,* 5384–5388. doi:10.1021/ja309974s
7. Reetz, M. T.; Jiao, N. *Angew. Chem., Int. Ed.* **2006,** *45,* 2416–2419. doi:10.1002/anie.200504561
8. Podtetenieff, J.; Taglieber, A.; Bill, E.; Reijerse, E. J.; Reetz, M. T. *Angew. Chem., Int. Ed.* **2010,** *49,* 5151–5155. doi:10.1002/anie.201002106
9. Bos, J.; Fusetti, F.; Driessen, A. J. M.; Roelfes, G. *Angew. Chem., Int. Ed.* **2012,** *51,* 7472–7475. doi:10.1002/anie.201202070
10. Osseili, H.; Sauer, D. F.; Beckerle, K.; Arlt, M.; Himiyama, T.; Polen, T.; Onoda, A.; Schwaneberg, U.; Hayashi, T.; Okuda, J. *Beilstein J. Org. Chem.* **2016,** *12,* 1314–1321. doi:10.3762/bjoc.12.124
11. Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. *Science* **2016,** *354,* 1048–1051. doi:10.1126/science.aah6219
12. Onoda, A.; Fukumoto, K.; Arlt, M.; Bocola, M.; Schwaneberg, U.; Hayashi, T. *Chem. Commun.* **2012,** *48,* 9756–9758. doi:10.1039/c2cc35165j
13. Fukumoto, K.; Onoda, A.; Mizohata, E.; Bocola, M.; Inoue, T.; Schwaneberg, U.; Hayashi, T. *ChemCatChem* **2014,** *6,* 1229–1235. doi:10.1002/cctc.201301055
14. Heinisch, T.; Ward, T. R. *Eur. J. Inorg. Chem.* **2015,** 3406–3418. doi:10.1002/ejic.201500408
15. Hoarau, M.; Hureau, C.; Gras, E.; Faller, P. *Coord. Chem. Rev.* **2016,** *308,* 445–459. doi:10.1016/j.ccr.2015.05.011
16. Hayashi, T.; Sano, Y.; Onoda, A. *Isr. J. Chem.* **2015,** *55,* 76–84. doi:10.1002/ijch.201400123
17. Bos, J.; Roelfes, G. *Curr. Opin. Chem. Biol.* **2014,** *19,* 135–143. doi:10.1016/j.cbpa.2014.02.002
18. Matsuo, T.; Imai, C.; Yoshida, T.; Saito, T.; Hayashi, T.; Hirota, S. *Chem. Commun.* **2012,** *48,* 1662–1664. doi:10.1039/c2cc16898g
19. Tenne, S.-J.; Schwaneberg, U. *Int. J. Mol. Sci.* **2012,** *13,* 2459–2471. doi:10.3390/ijms13022459
20. Pace, C. N.; Treviño, S.; Prabhakaran, E.; Scholtz, J. M. *Philos. Trans. R. Soc. London, Ser. B* **2004,** *359,* 1225–1234. doi:10.1098/rstb.2004.1500
21. Timasheff, S. N.; Arakawa, T. *J. Cryst. Growth* **1988,** *90,* 39–46. doi:10.1016/0022-0248(88)90296-5
22. Müller, C. A.; Weingartner, A. M.; Dennig, A.; Ruff, A. J.; Gröger, H.; Schwaneberg, U. *J. Ind. Microbiol. Biotechnol.* **2016,** *43,* 1641–1646. doi:10.1007/s10295-016-1844-5
23. de Carvalho, C. C. C. R. *Biotechnol. Adv.* **2011,** *29,* 75–83. doi:10.1016/j.biotechadv.2010.09.001
24. Locher, K. P.; Rees, B.; Koebnik, R.; Mitschler, A.; Moulinier, L.; Rosenbusch, J. P.; Moras, D. *Cell* **1998,** *95,* 771–778. doi:10.1016/S0092-8674(00)81700-6
25. Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. *Science* **1998,** *282,* 2215–2220. doi:10.1126/science.282.5397.2215

26. Koebnik, R.; Locher, K. P.; Van Gelder, P. *Mol. Microbiol.* **2000,** *37,* 239–253. doi:10.1046/j.1365-2958.2000.01983.x
27. Bonhivers, M.; Desmadril, M.; Moeck, G. S.; Boulanger, P.; Colomer-Pallas, A.; Letellier, L. *Biochemistry* **2001,** *40,* 2606–2613. doi:10.1021/bi001725i
28. Nallani, M.; Onaca, O.; Gera, N.; Hildenbrand, K.; Hoheisel, W.; Schwaneberg, U. *Biotechnol. J.* **2006,** *1,* 828–834. doi:10.1002/biot.200600042
29. Philippart, F.; Arlt, M.; Gotzen, S.; Tenne, S.-J.; Bocola, M.; Chen, H.-H.; Zhu, L.; Schwaneberg, U.; Okuda, J. *Chem. – Eur. J.* **2013,** *19,* 13865–13871. doi:10.1002/chem.201301515
30. Surrey, T.; Jahnig, F. *J. Biol. Chem.* **1995,** *270,* 28199–28203. doi:10.1074/jbc.270.47.28199
31. Güven, A.; Fioroni, M.; Hauer, B.; Schwaneberg, U. *J. Nanobiotechnol.* **2010,** *8,* No. 14. doi:10.1186/1477-3155-8-14
32. Onaca, O.; Sarkar, P.; Roccatano, D.; Friedrich, T.; Hauer, B.; Grzelakowski, M.; Güven, A.; Fioroni, M.; Schwaneberg, U. *Angew. Chem., Int. Ed.* **2008,** *47,* 7029–7031. doi:10.1002/anie.200801076
33. Michaux, C.; Pomroy, N. C.; Prive, G. G. *J. Mol. Biol.* **2008,** *375,* 1477–1488. doi:10.1016/j.jmb.2007.11.026
34. Charan, H.; Kinzel, J.; Glebe, U.; Anand, D.; Garakani, T. M.; Zhu, L.; Bocola, M.; Schwaneberg, U.; Böker, A. *Biomaterials* **2016,** *107,* 115–123. doi:10.1016/j.biomaterials.2016.08.033
35. Frotscher, E.; Danielczak, B.; Vargas, C.; Meister, A.; Durand, G.; Keller, S. *Angew. Chem., Int. Ed.* **2015,** *54,* 5069–5073. doi:10.1002/anie.201412359
36. Pan, Y.; Brown, L.; Konermann, L. *J. Mol. Biol.* **2011,** *410,* 146–158. doi:10.1016/j.jmb.2011.04.074
37. Roman, E. A.; González Flecha, F. L. *Biomolecules* **2014,** *4,* 354–373. doi:10.3390/biom4010354
38. Roussel, G.; Perpète, E. A.; Matagne, A.; Tinti, E.; Michaux, C. *Biotechnol. Bioeng.* **2013,** *110,* 417–423. doi:10.1002/bit.24722
39. Cuesta-Seijo, J. A.; Neale, C.; Khan, M. A.; Moktar, J.; Tran, C. D.; Bishop, R. E.; Pomès, R.; Privé, G. G. *Structure* **2010,** *18,* 1210–1219. doi:10.1016/j.str.2010.06.014
40. Tastan, O.; Dutta, A.; Booth, P.; Klein-Seetharaman, J. *Biochim. Biophys. Acta* **2014,** *1837,* 656–663. doi:10.1016/j.bbabio.2013.11.021
41. Tenne, S.-J.; Kinzel, J.; Arlt, M.; Sibilla, F.; Bocola, M.; Schwaneberg, U. *J. Chromatogr., B* **2013,** *937,* 13–17. doi:10.1016/j.jchromb.2013.07.021
42. Dworeck, T.; Petri, A.-K.; Muhammad, N.; Fioroni, M.; Schwaneberg, U. *Protein Expression Purif.* **2011,** *77,* 75–79. doi:10.1016/j.pep.2010.12.006
43. Mohammad, M. M.; Iyer, R.; Howard, K. R.; McPike, M. P.; Borer, P. N.; Movileanu, L. *J. Am. Chem. Soc.* **2012,** *134,* 9521–9531. doi:10.1021/ja3043646
44. Pawelek, P. D.; Croteau, N.; Ng-Thow-Hing, C.; Khursigara, C. M.; Moiseeva, N.; Allaire, M.; Coulton, J. W. *Science* **2006,** *312,* 1399–1402. doi:10.1126/science.1128057
45. Boisselier, E.; Audet, M.-L.; Cantin, L.; Salesse, C. *BioTechniques* **2011,** *51,* 193–194. doi:10.2144/000113739
46. Michaux, C.; Pouyez, J.; Wouters, J.; Privé, G. G. *BMC Struct. Biol.* **2008,** *8,* No. 29. doi:10.1186/1472-6807-8-29
47. Sauer, D. F.; Bocola, M.; Broglia, C.; Arlt, M.; Zhu, L.-L.; Brocker, M.; Schwaneberg, U.; Okuda, J. *Chem. – Asian J.* **2015,** *10,* 177–182. doi:10.1002/asia.201403005
48. Johnson, W. C., Jr. *Annu. Rev. Biophys. Biophys. Chem.* **1988,** *17,* 145–166. doi:10.1146/annurev.bb.17.060188.001045
49. Carrington, J. C.; Dougherty, W. G. *Proc. Natl. Acad. Sci. U. S. A.* **1988,** *85,* 3391–3395. doi:10.1073/pnas.85.10.3391
50. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. *Organometallics* **2010,** *29,* 2176–2179. doi:10.1021/om100106e
51. Prilipov, A.; Phale, P. S.; Koebnik, R.; Widmer, C.; Rosenbusch, J. P. *J. Bacteriol.* **1998,** *180,* 3388–3392.
52. Krieger, E.; Vriend, G. *Bioinformatics* **2014,** *30,* 2981–2982. doi:10.1093/bioinformatics/btu426
53. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000,** *21,* 1049–1074. doi:10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F
54. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004,** *25,* 1157–1174. doi:10.1002/jcc.20035
55. Jakalien, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002,** *23,* 1623–1641. doi:10.1002/jcc.10128
56. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992,** *13,* 952–962. doi:10.1002/jcc.540130805
57. Krieger, E.; Nielsen, J. E.; Spronk, C. A. E. M.; Vriend, G. *J. Mol. Graphics Modell.* **2006,** *25,* 481–486. doi:10.1016/j.jmgm.2006.02.009

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.13.148

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

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]

Full Research Paper

Open Access

Address:
[1]Nanobiology Course in Graduate School of Advanced Integration Science & Molecular Chirality Research Center, Chiba University, Matsudo 271-8510, Chiba, Japan and [2]Nanomaterials Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba 305-8565, Japan

Email:
Yoshihiro Nishida* - YNishida@faculty.chiba-u.jp

* Corresponding author

Keywords:
cell membrane; chirality; conformation; deuterium labeling; *sn*-glycerol; glycerolipids; glycerophospholipids; helicity; Karplus equation; proton NMR spectroscopy; staggered conformers

Beilstein J. Org. Chem. **2017,** *13,* 1999–2009.
doi:10.3762/bjoc.13.196

Received: 13 February 2017
Accepted: 01 September 2017
Published: 25 September 2017

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

Guest Editor: H. B. Bode

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

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 α-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.

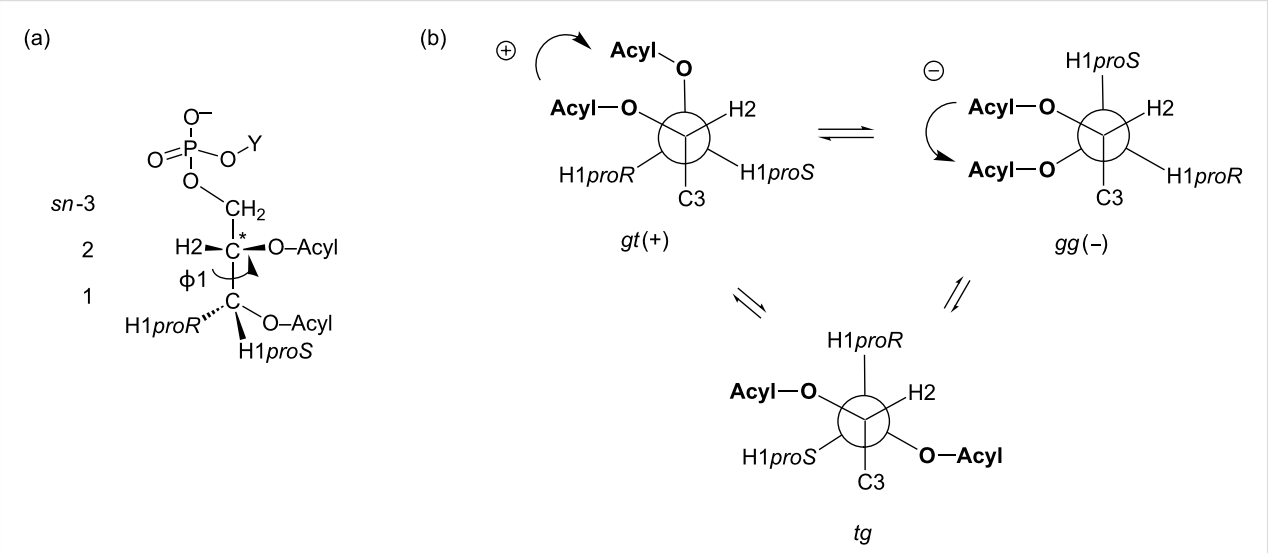
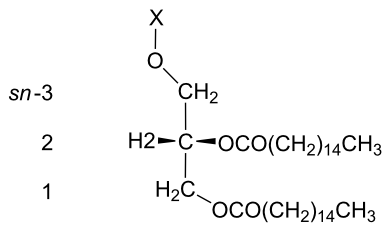


Figure 1: (a) Structures of cell-membrane glycerophospholpids with a common asymmetric 1,2-diacyl-*sn*-glycerol-3-phosphate structure and (b) the conformational equilibrium among three staggered conformers, namely gt(+), gg(−) and tg around the 1,2-diacyl moiety.



X =		types
1:	$-COCH_2(CH_2)_{14}CH_3$	triacylglycerols (TAG)
2:	$-CH_2Ph$	synthetic glycerolipids
3:	-H	1,2-diacyl-*sn*-glycerols (DAG)
4:	$-PO_3^-CH_2CH_2N^+(CH_3)_3$	glycerophospholipids

Scheme 1: The four 1,2-dipalmitoyl-*sn*-glycerols **1–4** examined in this study.

Results and Discussion

1. Helical conformational properties of tripalmitin **1** and 3-*O*-benzyl 1,2-dipalmitoyl-*sn*-glycerol (**2**) in $CDCl_3$ 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 conformers [gt(+), gg(−) and tg] are calculated using two Karplus equations, Equation 1 [22] and Equation 2 [18]. From the conformer populations (%), the "helicity index" is determined according to the method previously reported by our group [18].

$$\begin{aligned} {}^3J_{\mathrm{H1proS,H2}}(\mathrm{Hz}) &= 3.1\mathrm{gt}(+) + 2.8\mathrm{gg}(-) + 10.7\mathrm{tg} \\ {}^3J_{\mathrm{H1proR,H2}}(\mathrm{Hz}) &= 10.7\mathrm{gt}(+) + 0.9\mathrm{gg}(-) + 5.0\mathrm{tg} \\ \mathrm{gt}(+) + \mathrm{gg}(-) + \mathrm{tg} &= 100\ (\%) \end{aligned} \quad (1)$$

$$\begin{aligned} {}^3J_{\mathrm{H1proS,H2}}(\mathrm{Hz}) &= 2.5\mathrm{gt}(+) + 2.3\mathrm{gg}(-) + 10.6\mathrm{tg} \\ {}^3J_{\mathrm{H1proR,H2}}(\mathrm{Hz}) &= 10.2\mathrm{gt}(+) + 1.3\mathrm{gg}(-) + 5.8\mathrm{tg} \\ \mathrm{gt}(+) + \mathrm{gg}(-) + \mathrm{tg} &= 100\ (\%) \end{aligned} \quad (2)$$

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

Entry	Compound (head X =)	Solvent[a]	^{1}H NMR data δ (ppm) ^3J (Hz)		Populations (%) of staggered conformers in *sn*-1,2 position						Helicity index in *sn*-1,2 position		
					Equation 1			Equation 2			Equation 2 (Equation 1)		
			H1*proR*	H1*proS*	gt(+)	gg(−)	tg	gt(+)	gg(−)	tg	Sign (+/−)	Disparity [gt−gg]%	Volume [gt+gg]%
1	**1**[b] (palmitoyl)	$CDCl_3$	4.15 6.0	4.29 4.4	44	37	19	41	35	24	+	6 (7)	76 (81)
2	**2** ($-CH_2Ph$)	$CDCl_3$	4.19 6.4	4.34 3.8	52	37	11	49	34	17	+	15 (15)	83 (89)
3		C/M (10:1)	4.19 6.5	4.34 3.8	53	36	11	50	33	17	+	17(17)	83 (89)

[a]C/M (v/v) represents the ratios of the mixed solvents $CDCl_3$ (C) and methanol-d_4 (M). [b]Discrimination between H_{proR} and H_{proS} as well as the acquisition of their ^{1}H 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 **2** (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*proS* (δ 4.32 ppm) and H1*proR* (δ 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 δ 3.73 ppm. These observations are in good agreement with the ^{1}H NMR data of **3** reported by Vilceze and Bittman [29].

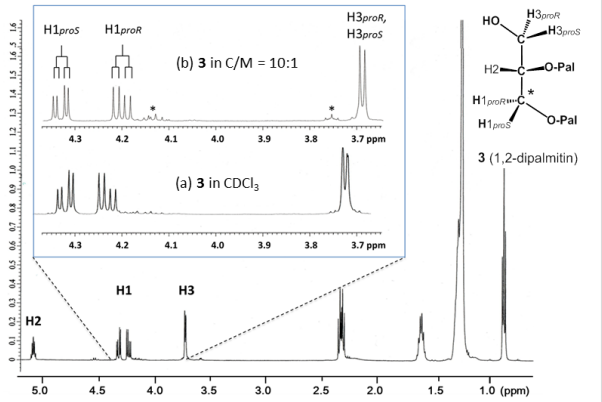


Figure 2: ^{1}H NMR spectra of 1,2-dipalmitin (**3**) in $CDCl_3$ after partial isomerization into the 1,3-isomer. (a) The expanded spectrum of **3** in $CDCl_3$, (b) **3** in a mixed solvent with ca. 10% methanol-d_4 in $CDCl_3$ (C/M ca 10:1, v/v). The signal marked with an asterisk * corressponds to a 1,3-diacyl isomer, which is derived from **3** during storage in a $CDCl_3$ solution.

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-

Table 2: ^{1}H NMR data and helical conformational properties of 1,2-dipalmitin **3** using different solvents.

Entry	Compound (head X =)	Solvent[a]	^{1}H NMR data δ (ppm) ^3J (Hz)		Populations (%) of staggered conformers in *sn*-1,2 position						Helicity index in *sn*-1,2 position		
					Equation 1			Equation 2			Equation 2 (Equation 1)		
			H1*proR*	H1*proS*	gt(+)	gg(−)	tg	gt(+)	gg(−)	tg	Sign (+/−)	Disparity [gt−gg]%	Volume [gt+gg]%
1	**3** (-H)	$CDCl_3$	4.23[b] 5.6	4.33[b] 4.5	40	40	20	35	39	26	–	−4 (0)	74 (80)
2		$CDCl_3$	4.23 5.7	4.32 4.4	41	40	19	37	39	24	–/+	−2 (1)	76 (81)
3		C/M (10:1)	4.20 6.2	4.33 4.0	48	38	13	45	35	20	+	10 (10)	80 (86)
4		C/M (5:1)	4.19 6.4	4.34 3.7	52	38	9	49	35	16	+	14 (14)	84 (90)
5		C/M (2:1)	4.19 6.5	4.37 3.7	53	37	10	50	34	16	+	16 (16)	84 (90)
6		C/M (2:1) + D_2O	4.18 6.6	4.37 3.5	55	38	7	53	34	13	+	19 (17)	87 (93)

[a]C/M (v/v) represents the ratios of the mixed solvents $CDCl_3$ (C) and methanol-d_4 (M). [b]^{1}H NMR data from the study reported by Vilceze and Bittman [29].

able changes in the "mixed solvents" containing methanol-d_4 in $CDCl_3$. With the addition of methanol-d_4, the H1*proR* and H1*proS* signals shift to high and low fields, respectively (Figure 2b). Simultaneously, the H3 signals shift upfield by 0.04 ppm. The shift of these H1 signals increases with an increase in the content of methanol-d_4 in the mixed solvents, while the H3 signals are marginally changed; thereafter, their positions are maintained at δ 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_2O to this solution further changes the gt(+) and tg populations by a few percent (Table 2, entry 6 and Figure 3). Moreover, the H2 signal of **3** shifts downfield by 0.03 ppm in the presence of D_2O, although this signal marginally changes in the mixed solvents without D_2O.

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_2O around the 3-OH group as well as the increasing polarity of the mixed solvent. As judged from the chemical shift change in the H3 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_2O.

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

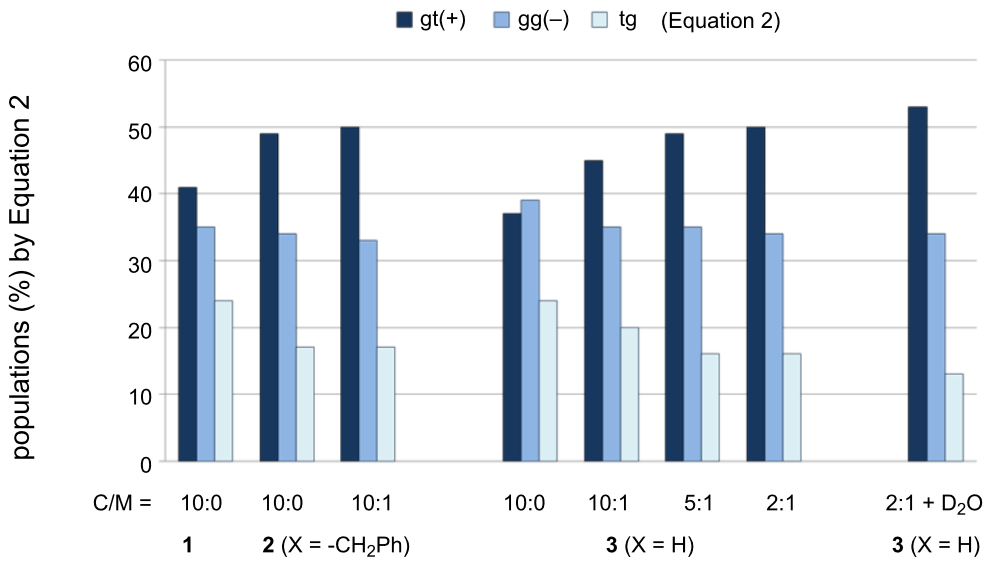
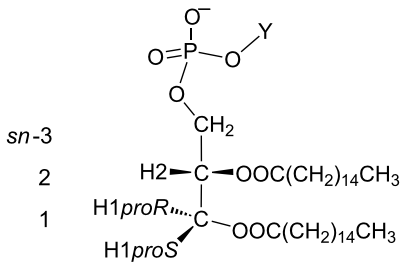


Figure 3: Fractional populations (%) of the three staggered conformers around the *sn*-1,2 C–C single bond in 1,2-dipalmitoyl-*sn*-glycerols **1**–**3** bearing different substituents (X) at the *sn*-3 position. Populations (%) are calculated from Equation 2 and each of the populations possibly includes deviations within ±3% by digital resolution (<0.12 Hz) of ^{1}H NMR spectroscopy (500 MHz).

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



4 (**DPPC**), Y = $-CH_2CH_2N^+(CH_3)_3$

DPPE, = $-CH_2CH_2NH_3^+$

DPPS, = $-CH(NH_3)^+COOH$

DPPA, = -H

Scheme 2: Structures of glycerophospholipids with a common structural skeleton of 1,2-dipalmitoyl-*sn*-glycerol 3-phosphate. Abbreviations: **DPPC** =1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine, **DPPE** = 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine, **DPPS** = 1,2-dipalmitoyl-*sn*-glycero-3-phospho-L-serine, **DPPA** = 1,2-dipalmitoyl-*sn*-glycerol 3-phosphate.

As shown in Figure 4, the ^1H NMR spectrum of **4** shows a pair of well-separated double doublet signals of H1*proR* (δ 4.14 ppm) and H1*proS* (δ 4.40 ppm). Compared to the other 1,2-diacyl-*sn*-glycerols **1**–**3**, this phospholipid exhibits a higher vicinal coupling constant to H1*proR* ($^3J_{H1R,H2}$ = 7.2 Hz) and a lower one to H1*proS* ($^3J_{H1S,H2}$ = 3.5 Hz). In addition, the difference in the chemical shift (Δδ = 0.26 ppm) between the H1*proR* and H1*proS* signals increases in **4**. These observations predict that the 1,2-diacyl moiety in **4** exhibits an extremely unique conformational property.

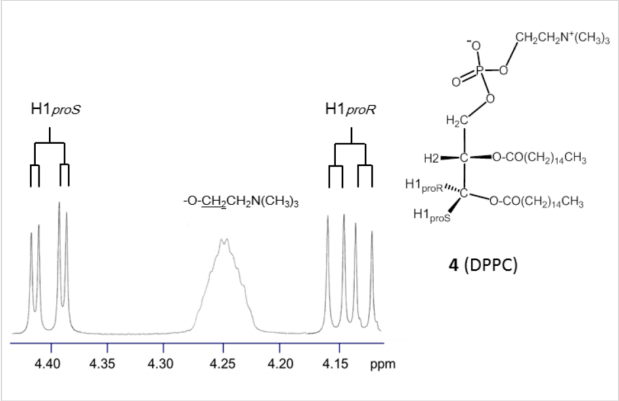


Figure 4: Partial ^1H NMR spectrum of **4** in a mixture of $CDCl_3$ and methanol-d_4 (C/M = 10:1, v/v).

In fact, the ^1H 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 ^1H 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_2O at concentrations less than or greater than the critical micellar concentration. In their ^1H 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.

Entry	Compound	Solvent[a]	^{1}H NMR δ (ppm) ^3J (Hz)		Populations (%) of staggered conformers around *sn*-1,2						Helicity index in *sn*-1,2 position		
					Equation 1			Equtation 2			Equation 2 (Equation 1)		
			H1*proR*	H1*proS*	gt(+)	gg(−)	tg	gt(+)	gg(−)	tg	Sign (+/−)	Disparity [gt−gg]%	Volume [gt+gg]%
1	**4** (**DPPC**)	$CDCl_3$	4.13[b] 7.3	4.40[b] 2.9	66	35	−1	64	30	6	+	34 (31)	94 (101)
2		C/M (10:1)	4.14 7.2	4.40 3.5	62	32	6	59	27	13	+	32 (30)	86 (94)
3		C/M (2:1)	4.16[c] 6.9	4.42[c] 3.1	61	38	1	59	33	8	+	26 (23)	92 (99)
4		CD_3OD	4.18[d] 7.0	4.42[d] 3.2	61	36	3	59	31	10	+	28 (25)	91 (97)
5	**DPPE**[c]	C/M (2:1)	4.18 6.9	4.40 3.4	59	36	5	57	31	12	+	26 (23)	88 (95)
6	**DPPS**[c]	C/M (4:3)	4.19 7.2	4.43 3.0	64	36	0	63	30	7	+	33 (28)	93 (100)
7	**DPPA**[c]	C/M (2:1)	4.21 7.1	4.40 3.5	61	33	6	59	28	13	+	31 (28)	87 (94)

[a]C/M (v/v) represents the ratios of the mixed solvents $CDCl_3$ (C) and methanol-d_4 (M). [b]^{1}H NMR data obtained from a database of Spectral Database for Organic Chemistry (SDBS), No. 16108HSP-45-792 in http://sdbs.db.aist.gojp/sdbs/vgi-bin/direct_frame_top.cgi [31]. [c]^{1}H NMR data from a paper of Hauser et al. [10]. [d]^{1}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 glycero1ipids **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:

$$\text{Helical disparity (\%)} = [\text{gt}(+) - \text{gg}(-)]\% = 1.34[\text{gt}(+)\% - 37.9] \quad (3)$$

$$\text{Population (\%) of gt}(+) = 2.94[50.8 - \text{gg}(-)\%] \quad (4)$$

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. Glycerophospholipid **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

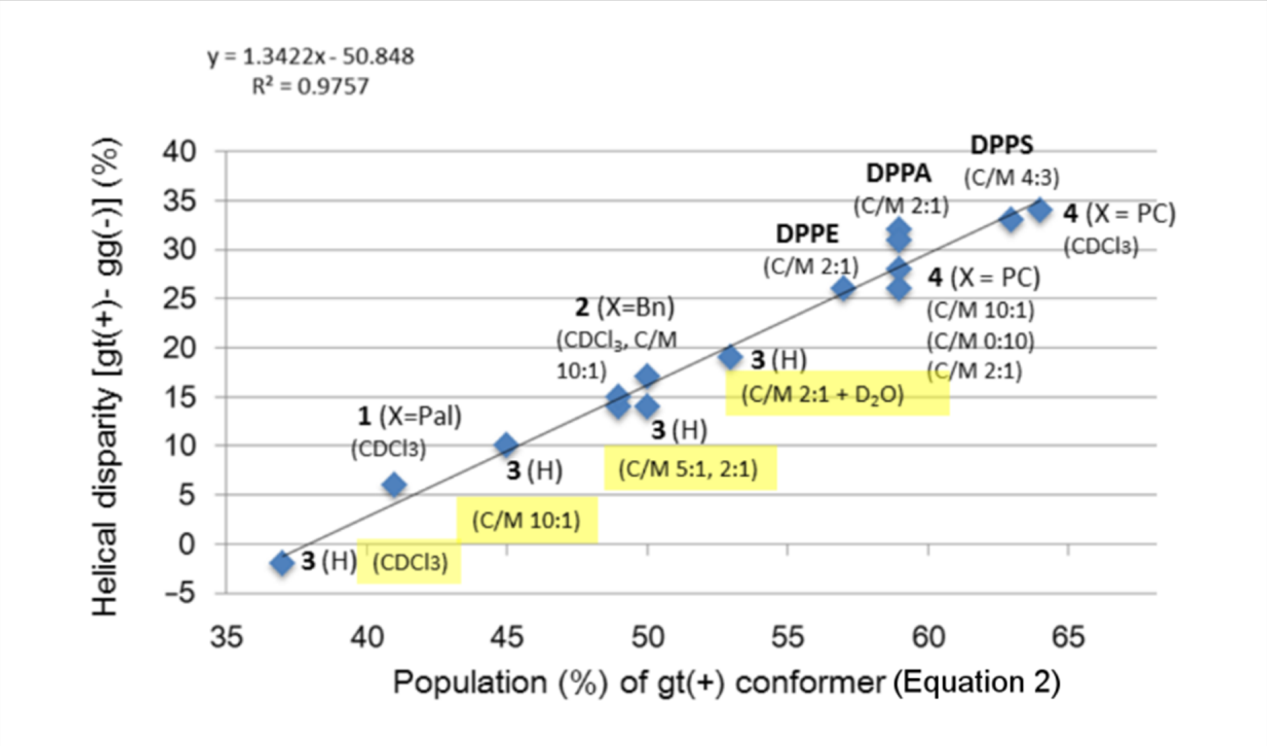


Figure 5: Linear relation between the helical disparity (%) and gt(+) population (%) as observed for the helical conformational properties of 1,2-dipalmitoyl-*sn*-glycerols **1**–**4** in the solution state.

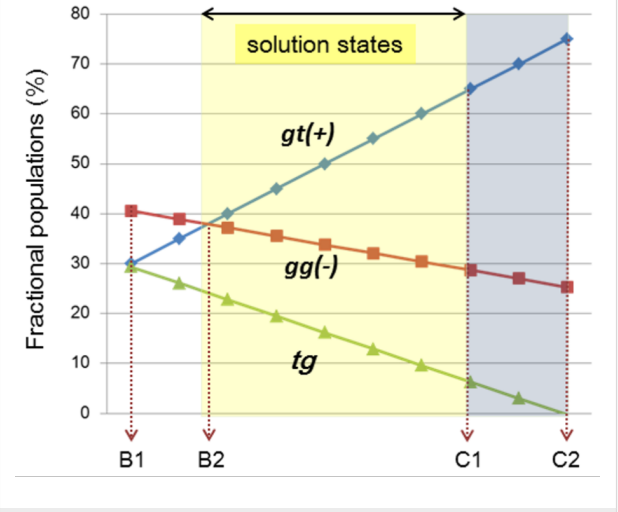


Figure 6: An empirical diagram showing helical conformational properties around 1,2-diacyl moiety in asymmetric 1,2-dipalmitoyl-*sn*-glycerols in solution states.

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 H1*proR* and H1*proS* signals (Materials and methods). Throughout this study, each of the 1,2-diplamitoyl-*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 (%)

Table 4: Helical conformational properties of α-D- and α-L-glucopyranosyl 1,2-dipalmitoyl-*sn*-glycerols in the solvent mixture of $CDCl_3$ and methanol-d_4 (C/M = 10:1).

Entry	Compound[a] (head groups at *sn*-3)	Results[b] (%) from ^{1}H NMR spectroscopic analyses by Equation 2					Calculated values[c] (%) with Equation 4				
		gt	gg	tg	dispariity	volume	gt	gg	tg	disparity	volume
1	α-D-Glc	53	36	11	17	89	53	33	14	20	86
2	6-phosphocholine α-D-Glc	53	36	11	17	89	53	33	14	20	86
3	6-palmitoyl α-D-Glc	49	37	14	12	86	49	34	17	15	83
4	6-phosphocholine α-L-Glc	55	33	12	22	88	55	32	13	23	87

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

linearly 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*-glyceols 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 in our preceding studies [16-18].

The ^{1}H 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*-glycerols assuming these heterogeneous situations which may occur in nature.

Materials and Methods

Model compounds

Tripalmitin **1** was prepared together with chirally deuterated *sn*-glycerols and identified in our former studies [22,23]. 1,2-Dipalmitoyl-*sn*-glycerol (**3**) and its 3-*O*-benzyl derivative **2** were prepared in a reported manner [8,29] (for details, see Supporting Information File 1). 1,2-Dipalmitoyl-*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% (^{1}H NMR) except for **3** which isomerizes into the 1,3-diacyl isomer during storage in $CDCl_3$ solution.

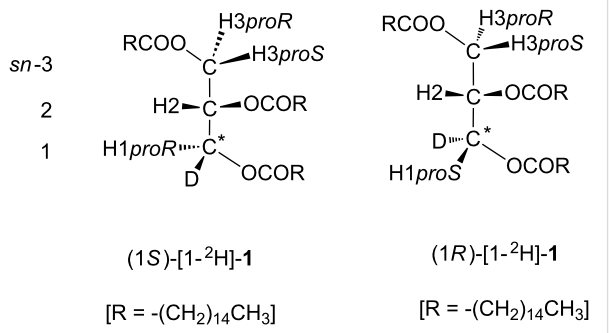
Acquisition of the ^{1}H NMR spectral data of H1*proR* and H1*proS* signals

Each of the four glycerolipids **1**–**4** is dissolved in either $CDCl_3$ or the mixed solvents containing methanol-d_4 in $CDCl_3$ (deuterium content > 99.5%) at ca. 10 mM concentrations. ^{1}H 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 (^3J, Hz) of H1*proR* and H1*proS* signals are obtained manually with ^{1}H 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 ^{1}H NMR analysis since a peak top by computer system does not always point at a weighted center correctly.

The discrimination between H1*proR* and H1*proS* signals is another crucial process. In our former studies [22,23], chiral ^{2}H-labelled triacylglycerols were prepared (Scheme 3) and applied for the assignment of these diastereomeric protons, namely H1*proR* and H1*proS*. The results have shown an empirical relation between the two H1 signals; the H1*proS* signals appear downfield from the H1*proR* signals (δ H1*proS* > δ H1*proR* ppm) and have lower smaller coupling constants ($^3J_{\mathrm{H1}proR,\mathrm{H2}} > {}^3J_{\mathrm{H1}proS,\mathrm{H2}}$ 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 ^{1}H 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 (1*S*)- and (1*R*)-1-[^{2}H]-**1** [23].

From the vicinal coupling constants (^{3}J Hz) of H1*proR* and H1*proS* 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

1. *Cell membrane*. Nature Education; http://www.nature.com/scitable/topicpage/cell-membranes-14052567.
2. "*Glycerophospholipids*". Farooqui, A. A.; http://www.els.net/WileyCDA/ElsArticle/refId-a0000726.html. doi:10.1002/9780470015902.a0000726.pub3
3. van Meer, G.; Voelker, D. R.; Feigenson, G. W. *Nat. Rev. Mol. Cell Biol.* **2008,** *9,* 112–124. doi:10.1038/nrm2330
4. Nomenclature of Lipids. IUPAC-IUB Commission on Biochemical Nomenclature (CBN); http://www.chem.qmul.ac.uk/iupac/lipid.
5. Bangham, A. D.; Horne, R. W. *J. Mol. Biol.* **1964,** *8,* 660–668. doi:10.1016/S0022-2836(64)80115-7
6. Singer, S. J.; Nicolson, G. L. *Science* **1972,** *175,* 720–731. doi:10.1126/science.175.4023.720
7. Hauser, H.; Pascher, I.; Pearson, R. H.; Sundell, S. *Biochim. Biophys. Acta* **1981,** *650,* 21–51. doi:10.1016/0304-4157(81)90007-1
8. Bruzik, K.; Jiang, R. T.; Tsai, M. D. *Biochemistry* **1983,** *22,* 2478–2486. doi:10.1021/bi00279a026
9. Plückthun, A.; DeBony, J.; Fanni, T.; Dennis, E. A. *Biochim. Biophys. Acta* **1986,** *856,* 144–154. doi:10.1016/0005-2736(86)90021-0
10. Hauser, H.; Pascher, I.; Sundell, S. *Biochemistry* **1988,** *27,* 9166–9174. doi:10.1021/bi00426a014
11. Meulendijks, G. H. W. M.; de Haan, J. W.; van Genderen, M. H. P.; Buck, H. M. *Eur. J. Biochem.* **1989,** *182,* 531–538. doi:10.1111/j.1432-1033.1989.tb14860.x
12. Goto, M.; Kodali, D. R.; Small, D. M.; Honda, K.; Kozawa, K.; Uchida, T. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89,* 8083–8086. doi:10.1073/pnas.89.17.8083
13. Hong, M.; Schmidt-Rohr, K.; Zimmermann, H. *Biochemistry* **1996,** *35,* 8335–8341. doi:10.1021/bi953083i
14. Feller, S. E.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2000,** *104,* 7510–7515. doi:10.1021/jp0007843
15. Krishnamurty, S.; Stefanov, M.; Mineva, T.; Begu, S.; Devoissell, J. M.; Goursot, A.; Zhu, R.; Salahub, D. R. *J. Phys. Chem. B* **2008,** *112,* 13433–13442. doi:10.1021/jp804934d

16. Nishida, Y.; Shingu, Y.; Mengfei, Y.; Fukuda, K.; Dohi, H.; Matsuda, S.; Matsuda, K. *Beilstein J. Org. Chem.* **2012,** *8,* 629–639. doi:10.3762/bjoc.8.70
17. Uzawa, H.; Nishida, Y.; Ohrui, H.; Meguro, H. *J. Org. Chem.* **1990,** *55,* 116–122. doi:10.1021/jo00288a024
18. Yuan, M.; Fukuda, K.; Dohi, H.; Uzawa, H.; Nishida, Y. *Tetrahedron: Asymmetry* **2015,** *26,* 1138–1144. doi:10.1016/j.tetasy.2015.08.012
19. Harada, N.; Nakanishi, K. *Circular Dichroic Spectroscopy Exciton Coupling in Organic Stereochemistry;* University Science Books: California, 1983.
20. Carrol, J. D. *Chirality* **2009,** *21,* 354–358. doi:10.1002/chir.20590
21. Mannock, D. A.; Harper, P. E.; Gruner, S. M.; McElhaney, R. N. *Chem. Phys. Lipids* **2001,** *111,* 139–161. doi:10.1016/S0009-3084(01)00153-0
22. Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. *J. Carbohydr. Chem.* **1988,** *7,* 239–250. doi:10.1080/07328308808058917
23. Nishida, Y.; Uzawa, H.; Hanada, S.; Ohrui, H.; Meguro, H. *Agric. Biol. Chem.* **1989,** *53,* 2319–2326. doi:10.1271/bbb1961.53.2319
24. Uzawa, H.; Nishida, Y.; Hanada, S.; Ohrui, H.; Meguro, H. *Chem. Commun.* **1989,** 862–863. doi:10.1039/c39890000862
25. Nishizuka, Y. *Nature (London)* **1984,** *308,* 693–698. doi:10.1038/308693a0
26. Goñi, F. M.; Alonso, A. *Prog. Lipid Res.* **1999,** *38,* 1–48. doi:10.1016/S0163-7827(98)00021-6
27. Coleman, R. A.; Lee, D. P. *Prog. Lipid Res.* **2004,** *43,* 134–176. doi:10.1016/S0163-7827(03)00051-1
28. Carrasco, S.; Mérida, I. *Trends Biochem. Sci.* **2007,** *32,* 27–36. doi:10.1016/j.tibs.2006.11.004
29. Vilcheze, C.; Bittman, R. *J. Lipid Res.* **1994,** *35,* 734–738.
30. Hamilton, J. A.; Bhamidipati, S. P.; Kodali, D. R.; Small, D. M. *J. Biol. Chem.* **1991,** *266,* 1177–1186.
31. Spectral Database for Organic Chemistry (SDBS), Index No. 16108HSP-45-792. http://sdbs.db.aist.gojp/sdbs/vgi-bin/direct_frame_top.cgi.
32. Holopainen, J. H.; Lehtonen, J. Y. A.; Kinnunen, P. K. J. *Biophys. J.* **1999,** *76,* 2111–2120. doi:10.1016/S0006-3495(99)77367-4
33. Samonshina, N. M.; Liu, X.; Brazdova, B.; Franz, A. H.; Samoshin, V. V.; Guo, X. *Pharmaceutics* **2011,** *3,* 379–405. doi:10.3390/pharmaceutics3030379
34. Hishikawa, D.; Hashidate, T.; Shimizu, T.; Shindou, H. *J. Lipid Res.* **2014,** *55,* 799–807. doi:10.1194/jlr.R046094
35. *Phosphatidylcholine*. Christie, W.; http://lipidlibrary.aocs.org/lipids/pc/index.htm. in the AOCS Lipid Library.
36. Ghosh, M. C.; Ray, A. K. *PLoS One* **2013,** *8,* e57919. doi:10.1371/journal.pone.0057919
37. Jang, H.-H.; Kim, D.-H.; Ahn, T.; Yun, C.-H. *Arch. Biochem. Biophys.* **2010,** *493,* 143–150. doi:10.1016/j.abb.2009.10.012
38. Goñi, F. M. *Biochim. Biophys. Acta* **2014,** *1838,* 1467–1476. doi:10.1016/j.bbamem.2014.01.006
39. Lingwood, D.; Simons, K. *Science* **2010,** *327,* 46–50. doi:10.1126/science.1174621
40. Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980,** *36,* 2783–2792. doi:10.1016/0040-4020(80)80155-4

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.13.196