On the biosynthetic origin of carminic acid

Rasmussen, Silas A.; Kongstad, Kenneth T; Khorsand-Jamal, Paiman; Kannangara, Rubini Maya; Nafisi, Majse; Van Dam, Alex; Bennedsen, Mads; Madsen, Bjørn; Okkels, Finn; Gottfredsen, Charlotte H.

Published in:
Insect Biochemistry and Molecular Biology

Link to article, DOI:
10.1016/j.ibmb.2018.03.002

Publication date:
2018

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
On the biosynthetic origin of carminic acid

Silas A. Rasmussen, Kenneth T. Kongstad, Paiman Khorsand-Jamal, Rubini Maya Kannangara, Majse Nafisi, Alex Van Dam, Mads Bennedsen, Bjørn Madsen, Finn Okkels, Charlotte H. Gotfredsen, Dan Staerk, Ulf Thrane, Uffe H. Mortensen, Thomas O. Larsen, Rasmus J.N. Frandsen

PII: S0965-1748(18)30111-5
DOI: 10.1016/j.ibmb.2018.03.002
Reference: IB 3039

To appear in: Insect Biochemistry and Molecular Biology

Received Date: 8 July 2017
Revised Date: 9 March 2018
Accepted Date: 13 March 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
1 x acetyl-CoA
7 x malonyl-CoA

Polyketide synthase (PKS)

Flavokermesic acid anthrone

Monooxygenase C-glucosyltransferase

Carminic acid
On the biosynthetic origin of carminic acid

Silas A. Rasmussen a, Kenneth T. Kongstad b, Paiman Khorsand-Jamal b,c, Rubini Maya Kannangara c,d, Majse Nafisi c,d, Alex Van Dam a,e, Mads Bennedsen c, Bjørn Madsen c, Finn Okkels c,f, Charlotte H. Gotfredsen g, Dan Staerk b, Ulf Thrane a, Uffe H. Mortensen a, Thomas O. Larsen a,* and Rasmus J.N. Frandsen a,*

a Department of Bioengineering and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 221 and 223, 2800 Kongens Lyngby, Denmark.
b Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.
c Chr. Hansen A/S, Boege Alle 10-12, 2970 Hoersholm, Denmark.
d Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark.
e Department of Biology, University of Puerto Rico, PO Box 9000, Mayagüez, PR 00681-9000, Puerto Rico.
f Current adresse: Actabio ApS, Kongemarker 11, 4000 Roskilde, Denmark

g Department of Chemistry, Technical University of Denmark, Kemitorvet, Building 207, 2800 Kongens Lyngby, Denmark.

* Corresponding author: Phone: +45 45252632, E-mail: tol@bio.dtu.dk and Phone: +45 45252708, E-mail: rasf@bio.dtu.dk
ABSTRACT

The chemical composition of the scale insect *Dactylopius coccus* was analyzed with the aim to discover new possible intermediates in the biosynthesis of carminic acid. UPLC-DAD/HRMS analyses of fresh and dried insects resulted in the identification of three novel carminic acid analogues and the verification of several previously described intermediates. Structural elucidation revealed that the three novel compounds were desoxyerythrolaccin-0-glucosyl (DE-O-Glc p), 5,6-didehydroxyerythrolaccin 3-O-β-D-glucopyranoside (DDE-3-O-Glp), and flavokermesic acid anthrone (FKA). The finding of FKA in *D. coccus* provides solid evidence of a polyketide, rather than a shikimate, origin of coccid pigments. Based on the newly identified compounds, we present a detailed biosynthetic scheme that accounts for the formation of carminic acid (CA) in *D. coccus* and all described coccid pigments which share a flavokermesic acid (FK) core. Detection of coccid pigment intermediates in members of the *Planococcus* (mealybugs) and *Pseudaulacaspis* genera shows that the ability to form these pigments is taxonomically more widely spread than previously documented. The shared core-FK-biosynthetic pathway and wider taxonomic distribution suggests a common evolutionary origin for the trait in all coccid dye producing insect species.

Keywords: Carminic acid, carmine, anthraquinones, coccid pigment, polyketide, insects, *Dactylopius coccus*, biosynthesis
1. Introduction

Pigments derived from insects and especially coccids (scale insects) have been used by humans since ancient times for dyeing textiles, in cosmetics and in paints, and for coloring foods (Donkin, 1977). The most commonly used coccid dyes include kermesic acid (KA), laccaic acids (LA) and carminic acid (CA), which share a red color hue due to a similar chromophore structure (Łagowska and Golan, 2009). The compounds, or combinations of these, have been reported to be produced by several species of distantly related scale insects (Hemiptera: Coccoidea). Mainly five species, namely Porphyrophora hamelii (Armenian/Ararat cochineal), Kermes vermilio (kermes), Porphyrophora polonica (Polish cochineal), Dactylopius coccus (Mexican cochineal) and Kerria lacca (Indian lac insect) have at various points in history, and at different geographical localities, been utilized by humans for large scale production of coccid dyes (Donkin, 1977). Carminic acid and its aluminum salt carmine (E120) is by many considered as the pinnacle of coccid dyes, based on its hue, light, temperature, and oxidation stability, and the yields by which it can be obtained from natural sources (Dapson, 2007). CA is known to be produced by P. hamelii (Asia Minor), P. polonica (Europe), and D. coccus (Meso and South America), all of which have served as sources for the compound (Wouters and Verhecken, 1989). Present day production is based on D. coccus due to its exceptional high pigment content (16-22% of dry weight), low fat content, and the ease by which the insect can be cultured and harvested from leaves of Opuntia cacti (Donkin, 1977; Downham and Collins, 2000). A thorough introduction to the historical use and geopolitical role of carmine is given by Dapson (Dapson, 2007).

Although insect-derived pigments have been utilized by humans for millennia and remain of significant value within the food colorant market, the underlying biochemistry for their production remains largely unknown. The coccid dyes, such as CA, have by many authors been categorized as polyketides solely based on their structure (Morgan, 2010; Cameron et al., 1978; Pankewitz and Hilker, 2008; Brown, 1975). The biosynthesis mechanisms of formation of polyketides, via the successive condensation of acetyl-CoA and malonyl-CoA units catalyzed by polyketide synthases (PKSs), is well described in bacteria, fungi, and plants (Staunton and Weissman, 2001). However, no animal PKSs have yet been biochemically characterized, even though many insect species are known to contain compounds that potentially may be synthesized via the polyketide pathway. In most cases, the putative polyketides contents have been ascribed to the sequestering of precursors, or the finished compounds, from the insects’ diet (Pankewitz and Hilker, 2008). This situation is seen in Timarcha spp. (leaf beetles) which accumulates anthraquinones from its host plant Galium spp. (Rubiaceae) (Petitpierre, 1981), and in Laetilia coccidivora (pyralid moth) and Hyperaspis trifurcate (coccinellid beetle) larva that accumulate CA by predating on Dactylopius spp. (Eisner et al., 1994). In other cases, the origins of the detected polyketides in insects have been linked to the activity of
endosymbiotic bacteria, such as the production of pederin, a polyketide-peptide hybrid, in *Paederus* spp. (rove beetles), which depends on an endosymbiotic bacterium related to *Pseudomonas aeruginosa* (Piel, 2002; Kellner, 2002). A second example is the facultative endosymbiotic *Rickettsiella* spp. responsible for the production of the polyketide viridaphin A1 in various aphids (*Acyrthosiphon pisum* and *Megoura crassicauda*) (Tsuchida et al., 2010; Horikawa et al., 2011). The biosynthetic origin of coccid pigments in scale insects, however, remains a mystery. The hypothesis that coccid dyes are polyketides has solely been based on their chemical structure, even though the polyketide class is characterized by a shared mode of synthesis rather than shared structural features (Staunton and Weissman, 2001). In fact, for the anthraquinone core of CA, one could envision that this is formed via the shikimate based chorismate/O-succinyl benzoic acid pathway, as described for lucidi, alizarin, and morindone in rubiaceous plants (Leistner, 1973; Han et al., 2001). If CA is formed by this pathway it would not qualify as a polyketide and its formation would not depend on a PKS. However, it is possible to distinguish between the two alternative pathways as the polyketide-based synthesis would include a unique anthrone intermediate, which is not found in the shikimate-based pathway, where the anthraquinone is formed directly. The present study aims at increasing our understanding of the CA origin and its biosynthesis in *D. coccus*. We also aim to elaborate on the previously proposed links to other biosynthetic pathways responsible for the production of coccid dyes within the *Coccoidea* superfamily. In our study, we report the presence of flavokermesic acid anthrone (FKA) in *D. coccus*, which strengthens the hypothesis that coccid pigments are formed via the polyketide pathway, as FKA is the first cyclic intermediate in a polyketide-based pathway.

2. Materials and methods

2.1. Biological material

Adult *D. coccus* specimens were collected from *Opuntia* cacti pads on the Canary Islands, Lanzarote, near the village of Guatiza, June 2012. The insects were transported to Denmark either as live specimens on cacti pads or as dead specimens stored on dry ice. Additional *D. coccus* insects were collected from cacti pads near the city of Arequipa Peru, August 2012, flash frozen in liquid nitrogen, and shipped to Denmark on dry ice. Commercially available dried *D. coccus* insects were supplied by Chr. Hansen A/S. *Coccus hesperidium*, *Pseudococcus longispinus*, *Palmicultor browni*, and *Pseudaulacaspis pentagona* were collected in the greenhouses of the Botanical Garden (Natural History Museum of Denmark, University of Copenhagen) in Copenhagen in June 2014, and identified using the latest available identification keys (Dooley and Dones, 2015; Miller et al., 2014).

2.2. Instrumentation
Chemical analysis of *D. coccus* samples was performed using three different LC-MS setups. UPHLC-DAD-HRMS was performed on a maXis G3 QTOF mass spectrometer (Bruker Daltronics, Bremen, Germany) equipped with an electrospray ionization source coupled to an Ultima 3000 UHPLC-DAD (Dionex).

Separation was performed on a Kinetex C\textsubscript{18} column (150 × 2.1 mm, 2.6 μm, Phenomenex Inc., Torrance, CA, USA) maintained at 40 °C using a linear H\textsubscript{2}O-acetonitrile gradient consisting of A: milliQ H\textsubscript{2}O containing 10 mM formic acid and B: acetonitrile containing 10 mM formic acid from 10 to 100% B in 10 min with a flow rate of 400 μL min\textsuperscript{-1}. The FK anthonre was detected on a HPLC-DAD-HRMS system consisting of an Agilent 1200 chromatograph comprising quaternary pump, degasser, thermostatted column compartment, autosampler, and photodiode array detector (Agilent Technology, Santa Clara, CA, USA) and a Bruker microTOF-Q II mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an electrospray ionization source and operated via a 1:99 flow splitter. Analyses were performed at 40 °C on a Luna C\textsubscript{18}(2) reversed-phase column (150 × 4.6 mm, 3 μm particle size, 100 Å pore size, Phenomenex Inc., Torrance, CA, USA) with a flow rate of 800 μL min\textsuperscript{-1}. HPLC solvent A consisted of H\textsubscript{2}O-acetonitrile 95:5 (v/v) with 0.1% formic acid and solvent B consisted of acetonitrile-H\textsubscript{2}O 95:5 (v/v) with 0.1 % formic acid. Separation was obtained using a linear gradient from 0 to 100% B in 20 minutes. Mass spectra were acquired in negative ionization mode. The search for coccid dye intermediates in the different scale insect species was performed on a 6540 Ultra High Definition UHD Accurate Mass Quadrupole Q-TOF LC/MS system (Agilent Technology, Santa Clara, CA, USA). Separation of the analytes was conducted on a Kinetex XB-C\textsubscript{18} (100 × 4.6 mm i.d. 2.6 μm, Phenomenex Inc., Torrance, CA, USA), column maintained at 35 °C. The analytes were eluted with a flow rate of 400 400 µL min\textsuperscript{-1} using a water-acetonitrile gradient consisting of the following steps: 100% water for 50 s followed by an gradual increase to 18.6% over 60 s, to 37.8% over 60 s, to 52.2% over 120 s, to 54.2% over 70 s, to 90% over 120 s, to 100% over 120 s followed by 60 s at 100% acetonitrile. The column was reconstituted with 100% water for 110 s prior to injection of the subsequent sample.

NMR spectra of 5,6-didehydroxyerythrolaccin 3-O-β-D-glucopyranoside (DDE-3-O-Glc\textsubscript{p}), desoxyerythrolaccin O-glucopyranoside (DE-O-Glcp), and dcll were recorded on a Varian Inova 500 MHz (Varian Inc., Palo Alto, California) using a 5-mm probe. Samples were dissolved in 500 μl DMSO-\textit{d}_6 and referenced to δ\textsubscript{n} at 2.50 ppm and δ\textsubscript{c} at 39.5 ppm. The NMR spectrum of flavokermesic acid anthrone (FKA) was recorded on a Bruker Avance III HD 600 MHz NMR spectrometer (\textsuperscript{1}H operating frequency 600.13 MHz) equipped with a cryogenically cooled 5-mm CPDCH probe-head (Bruker Biospin, Rheinstetten, Germany). The sample was dissolved in acetone-\textit{d}_6 and referenced to δ\textsubscript{n} 2.05 ppm and δ\textsubscript{c} 29.84 ppm. Following structural elucidation of the described compounds, their presence in the original samples was verified using targeted MS analysis.
Chiral GC-MS was performed using 10 μg DDE-3-O-Glc that was hydrolyzed in 10% aqueous HCl for 90 min at 90 °C, dried by a steam of N₂ and dissolved in 40 μL dry pyridine followed by 10 μL N-methyl-bis-trifluoroacetamide (MBTFA) (GC-grade, 99%, Sigma-Aldrich) and heating to 65 °C for 40 min. The sample was cooled to room temperature and subsequently analyzed on a CP-ChiraSil-L-Val GC column (25 m × 0.25 mm × 0.12 μm, Agilent Technology, Santa Clara, CA, USA) programmed to 70−150 °C at 4 °C min⁻¹ using an HP 6890 series GC system and Agilent 5973 mass selective detector. The sample was compared to the standards of D- and L-glucose (Sigma-Aldrich, St. Louis, MO, USA).

2.3. Extraction and isolation
Desoxyerythrolaccin (DE) and DDE-3-O-Glc were purified from 100 g of dried D. coccus. The insects were grinded and extracted with MeOH:H₂O (1:1) followed by liquid-liquid partitioning with EtOAc at pH 3 and concentrated in vacuo. The crude extract was first separated by ion-exchange using an NH₂ flash column. Carboxylic acid containing compounds was retained when washed with 50% aqueous MeCN containing 10 mM ammonium formate. Finally, the acidic compounds (FK, KA, and CA) were eluted with 50% aqueous MeCN adjusted to pH 11 with ammonium hydroxide. DE and DDE-3-O-Glc were purified from the 50% MeCN 10 mM ammonium formate eluate using reversed-phase semi-preparative chromatography on a 250 × 10 mm Luna2 C₁₈ column (Phenomenex, Torrance, CA, USA) using a Gilson HPLC system. Compounds were eluted with a gradient consisting of MilliQ H₂O:MeCN, both containing 50 ppm TFA. FK and dcll were recovered in the alkaline eluent of the NH₂ column. The extract was then further purified on Isolute diol material (Biotage, Uppsala, Sweden) on an Isolera auto flash purification system (Biotage, Uppsala, Sweden) in a step-wise elution from dichloromethane to EtOAc to MeOH to afford dcll and FK.

2.4. Synthesis of flavokermesic acid anthrone
Synthesis of flavokermesic acid anthrone from flavokermesic acid was conducted according to a previously published method by Schätzle, with slight modifications (Schätzle, 2012). In brief, 10 mg flavokermesic acid was dissolved in 1mL glacial acetic acid and 0.2 mL hydriodic acid (57 wt. % in H₂O) in a sealed microwave reactions vial and heated to 50 °C under stirring for 2 h in the dark. This adaption to the method reported by Schätzle 2012 was done to ensure full conversion of the flavokermesic acid to the anthrone with only limited decarboxylation. The hydriodic acid was quenched with 10 mL saturated Na₂S₂O₃ and extracted three times with diethyl ether. The ether phase was dried over MgSO₄ and lyophilized under reduced pressure. The sample was at all times kept in the dark and chemical analyses were performed in amber vials and NMR tubes to minimize the risk of dimerization and oxidation. The formation of the anthrone was confirmed by NMR analysis (Supplementary data Table S1).
2.5. Biosynthetic models

The biosynthetic models for formation of CA and related compounds were formulated using the retrosynthesis approach including commonly accepted enzymatic driven reactions, as described in the BRENDA database (Schomburg et al., 2004), and the available structural data for coccid dyes (Morgan, 2010; Cameron et al., 1978, 1981; Brown, 1975, Peggie et al., 2007; Stathopoulou et al., 2013; Bhide et al., 1969). The models were drawn using ChemDraw 15.9.9.106 (PerkinElmer Informatics, Inc., US).

3. Results

3.1. Detection of compounds in Dactylopius coccus extracts

HPLC-HRMS analysis of the raw extracts from fresh D. coccus showed that the main extractable pigment components were CA, FK, KA and dcll as previously reported (Wouters, J., Verhecken, A., 1989; Peggie et al., 2007; Méndez et al., 2004) (Figure 1A). These compounds were putatively identified based on relative retention time, high-resolution mass, UV/VIS spectra, and MS/MS fragmentation patterns. Spiking with authentic samples of CA and KA confirmed these compounds identity. FK was isolated and subjected to structural elucidation by 2D NMR (Supplementary data Table S1), and the data were in agreement with and confirmed the previously reported structure of FK (Wouters and Verhecken, 1987). CA and dcll were not easily separated using reversed-phase HPLC, and isolation of the two compounds relied on normal-phase flash chromatography on diol substituted silica. The putative dcll was further purified using semi-preparative HPLC and characterized by HR-MS and 2D NMR. Structural elucidation of the compound by 2D NMR experiments (Supplementary data Table S1) showed that dcll was flavokermesic acid 2-C-β-D-glucopyranoside, and the NMR data were in agreement with those reported for dcll (Stathopoulou et al., 2013). The reversed-phase HPLC-based analysis also revealed a previously undescribed major peak, eluting at 13.40 minutes (Figure 1A and 1C). The mass of the corresponding compound equaled the theoretical mass of flavokermesic acid anthrone (FKA), and this identity was confirmed by comparison with a FKA standard, semi-synthesized from authentic FK. In addition, to the metabolites detected using reversed-phase chromatography, strong anion exchange SPE (SAX SPE) were used to identify three non-acidic metabolites (DE, DDE-3-O-Glc and DE-0-Glc), that all displayed UV/VIS spectra with similarities to those reported for FK, KA, dcll and CA (Figure 1B, 1D, 1E and 1F). For unambiguously structure elucidation of DE and DDE-3-O-Glc, the compounds were purified in amounts sufficient for structural elucidation by HRMS and NMR spectroscopy. Purification of the DE-O-Glc compound unfortunately did not yield sufficient quantities for full structural elucidation by NMR.
3.2. Structural elucidation of DE, and the novel compounds DE-O-glucosyl, DDE-O-glucosyl and FKA

DE displayed UV/VIS spectrum (Figure 1E) similar to that of FK, indicating a similar core skeleton. The compound was not retained on a SAX column, suggesting that it lacked the carboxylic acid group found at C-7 in FK. This conclusion was supported by HRMS (m/z 271.0600 [M+H]+, calcd 271.0600, ΔM 0.0 ppm), suggesting a molecular formula of C₁₅H₁₀O₅, i.e., DE lacking CO₂ as compared to FK. Structural elucidation was carried out by ¹H NMR and 2D NMR spectroscopy (Table S1). The ¹H NMR spectrum showed a signal for the OH-group positioned peri to the carbonyl group (δ 13.30, 1-OH), two sets of meta-coupled protons H-5 and H-7 (δ 6.54 and 7.04, respectively, ³J₅₋₇ = 1.9 Hz) and H-2 and H-4 (δ 7.43 and 7.01, respectively, ³J₂₋₄ = 2.5 Hz), and a peri-positioned methyl group (δ 2.81, s, 11-CH₃). The meta-coupling between H-5 and H-7 clearly proves the lack of the carboxylic acid in position 7, and thus HRMS and NMR data supported the compound to be DE, also known as 3-hydroxy-aloesaponarin II (Mehandale et al., 1968), and previously observed in air-dried D. coccus (Sugimoto et al., 1998). Structural analysis of the first novel compound DDE-3-O-Glc revealed that the compound was a hitherto undescribed O-glucoside of 5,6-didehydroxyerythrolaccin (DDE) (Figure 2). The high-resolution mass spectrum of DDE-3-O-Glc suggested a molecular formula of C₂₁H₂₀O₉ ([M+H]+ m/z 417.1180, calcd. C₂₁H₂₁O₉+ 417.1180, ΔM 0.0 ppm; [M-H]− m/z 415.1034, calcd. C₂₁H₁₉O₉− 415.1029, ΔM 1.2 ppm). In addition, the compound exhibited a loss of m/z 162.0528, which is likely due to the loss of a labile O-linked hexose unit. The structural elucidation was carried out based on ¹H NMR and 2D NMR spectroscopy (Supplementary data Table S1). The ¹H spectrum showed resemblance to that of DE, but instead of the two doublets observed for the meta-coupled H-5 and H-7 in DE, signals for H-5 (δ 8.12, dd, 7.5, 1.2 Hz), H-6 (δ 7.78, t, 7.5 Hz), and H-7 (δ 7.75 dd, 7.5, 1.3 Hz) showed the absence of a hydroxyl group at C-6 in DDE-3-O-Glc. Furthermore, a doublet at δ 5.26 (³J₁'-H₂' = 7.6 Hz) for a β-configuration of the anomeric proton H-1' (as well as the ¹³C value of 101.1 ppm for C-1) and the remaining ¹H and ¹³C signals for H-2' to H-6' and C-2' to C-6' (Supplementary data Table S1), are in agreement with a β-D-glucose unit (Bock and Pedersen, 1983). The O-linkage of the sugar was evident from a more deshielded anomeric proton (101.1 ppm) compared to that of C-glucosyl linkages reported for related compounds (Stathopoulou et al., 2013). In addition, a HMBC correlation from H-1' to C-3 (164.6 ppm) further confirmed the O-glucosyl linkage to C-3. Thus, to establish the D- or L-configuration of the glucose moiety, an aliquot was hydrolyzed and analyzed by GC-MS and chiral GC-MS. This confirmed that the hexose moiety was D-glucose (Figure S2 and S3). Thus, the compound was identified as 5,6-didehydroxyerythrolaccin 3-O-β-D-glucopyranoside. Several attempts to purify the second novel compound DE-O-Glc did not yield sufficient quantities to allow acquisition of NMR data. However, the high-resolution mass spectrum of DE-O-Glc suggested a molecular formula of C₂₁H₂₀O₁₀ ([M+H]+ m/z 433.1129, calcd. C₂₁H₂₁O₁₀+ 433.1129, ΔM 0.0 ppm; C₂₁H₂₀O₁₀
(\text{[M+H]}^+ \text{ m/z } 431.0981, \text{ calcd. } \text{C}_{21}\text{H}_{19}\text{O}_{10}^+ \text{ m/z } 431.0984, \Delta M 0.7 \text{ ppm}); and the loss of a hexose moiety (m/z 162.0530) generated a fragment with the same mass as \text{DE}. These results indicate that this compound is an O-glucosylated form of \text{DE} (Figure 2), but the exact position of glucosylation could not be established.

The third novel compound was only detected in extracts from fresh and frozen \textit{D. coccus} and identified as flavokermesic acid anthrone (\text{FKA}) (Figure 2). The high-resolution mass spectrum of \text{FKA} detected in the insect suggested a molecular formula of \text{C}_{16}\text{H}_{12}\text{O}_{6} (\text{[M-H]} \text{ m/z 299.0559, calcd. } \text{C}_{16}\text{H}_{11}\text{O}_{6}^− \text{ m/z 299.0561, } \Delta M 1.0 \text{ ppm}). Positive identification of this compound was achieved by comparison of retention time, high-resolution MS, MS/MS fragmentation pattern, and UV/VIS spectrum for \text{FKA} that had been prepared by \text{chemical semi-synthesis}. See Table S1 for NMR data. Chemical semi-synthesis of \text{FKA} demonstrated that the pure compound is prone to dimerization as well as oxidation to \text{FK} in the presence of oxidants under \textit{in vitro} conditions. The observed spontaneous oxidation to \text{FK} likely explains why only fresh and frozen \textit{D. coccus} was found to contain \text{FKA} while the compound was not detected in dried insects which have been exposed to light, oxygen and other oxidative agents for longer periods of time.

3.3. LC-DAD/MS-based screening of coccid dye production in selected members of the superfamily \textit{Coccoidea}

To analyze the taxonomic distribution of the ability to produce coccid dyes, we collected representatives of four different \textit{Coccoidea} families found in Denmark. The collected species included nymphal states of \textit{Coccus hesperidium}, \textit{Pseudococcus longispinus}, \textit{Palmicultor browni} and \textit{Pseudaulacaspis pentagona}.

Metabolites from approximately 1 g of each of the individual species were extracted and analyzed by LC-MS/DAD. Extracted ion chromatograms (Figure 3) for masses equivalent to the known coccid dye intermediates showed that \textit{C. hesperidium} contained \text{KA}, \text{FK}, \text{dcII} and \text{CA}, while \textit{P. longispinus} and \textit{P. browni} both contained \text{FK} and \text{dcII}. Positive identification was based on accurate mass, retention time, UV/VIS spectra and authentic standards. Analysis of the \textit{P. pentagona} material did not reveal the presence of any coccid dye intermediates in this species.

4. Discussion

Assyrian cuneiform texts document that the utilization of scale insects for the production of pigments and dyes dates back to at least 3000 BC (Dapson 2007). Different civilizations separated by time and geography have relied on a limited number of scale insect species and the pigments they produce (Morgan, 2010; Cameron et al., 1978, 1981; Brown, 1975, Petitpierre, 1981: Peggie et al., 2007; Stathopoulou et al., 2013; Bhide et al., 1969). The long history and traditions associated with coccid dye utilization have resulted in a situation where scientific literature has focused on a very limited number of producing species (\textit{Dactylopius}...
coccus, Kerria lacca, and Kermes vermilio) indicating a discontinuous taxonomic distribution of the trait and unrelated biosynthetic origins for the various pigments. However, the number of shared metabolites and the common FK core structure of the pigments suggest that all coccid dyes are formed via a similar biosynthetic scheme with a common evolutionary origin. If so, then many more scale insect species, i.e. those forming a monophyletic clade with known producers, would also be expected to be able to produce coccid dyes or related metabolites. To this end, we have demonstrated that members of the Planococcus (mealybugs) and Pseudaulacaspis genera also produce coccid dyes. Hence, the ability to form these pigments is indeed more widely taxonomically spread than previously believed. This would support the most parsimonious explanation, being a common evolutionary origin within Coccoidea. Of evolutionary importance here is that Porphyrophora and Dactylopius belong to two vastly different clades of Coccoidea. Porphyrophora belongs to the more primitive Margarodidae and Dactylopius to the Dactylopiidae within the more highly derived Neococcoidea clade. The Margarodidae and Dactylopiidae share a most recent common ancestor 250 million years ago (Vea et al., 2016).

4.1. The biosynthetic origin of coccid dyes
The biosynthetic origin of coccid dyes such as KA, FK, CA, and LA has long been debated (Brown, 1975; Joshi and Lambdin, 1996; Ramirez-Puebla et al., 2010). Several studies have rejected the hypothesis that host plants supply the insects with any of the known coccid dye intermediates. Similarly, we were unable to detect any of these intermediates in Opuntia cacti pads in the case of D. coccus (data not shown). To say nothing about the wide variety of different hosts utilized by many Coccoidea, e.g. over 400 plant species are described as host for Kerria lacca (Sharma et al., 1997). Accordingly, it seems clear that Coccoidea must be able to synthesize the coccid dyes de novo from simple metabolites, e.g. glucose, present in the phloem sap of a wide range of plant species. Several studies have suggested that endosymbiotic bacteria may be responsible for the formation of the coccid dyes, and members of the Wolbachia and Azoarcus bacteria genera have been identified in the scale insects (Brown, 1975; Ramirez-Puebla et al., 2010; Pankewitz et al., 2007). However, none of these studies have proven a direct link between the presence, or activity, of these endosymbionts with the formation of coccid dyes. An alternative hypothesis is that the biosynthetic apparatus is encoded in the insects’ nuclear genome. Several examples exist where complex secondary metabolite’s biosynthetic pathways are encoded by genes in the genome of the producing insect, such as the Drosophila eye pigments drosopterin and ommochromes (Chovnick et al., 1990; Nijhout, 1997). Though no one has yet succeeded in identifying or describing the enzymological- or genetic basis for polyketide biosynthesis in insects, strong evidence does exist in support of nuclear encodement of the enzymes required for producing coccid dyes. However, this has largely been
overlooked in past studies. Such evidence includes the report of a stable yellow color mutant of the normally red *Kerria lacca* (Indian Lac-insect) (Negi, 1954), and a white color mutant incapable of producing LA$s (Chauhan and Teotia, 1973). Dissection of the genetic basis for these two mutations, by classical genetic crossing experiments, showed that the two traits are non-allelic and that they follow simple recessive inheritance (Chaucun, 1977; Chaucun and Mishra, 1977). If endosymbiotic bacteria were involved in catalyzing steps in the formation of the core structure of the coccid dyes, the mutant trait would be expected to be maternally inherited as a result of transfer via eggs (Ferrari and Vavre, 2011). Any conclusions on this matter must await demonstration of the ability of the color mutants to host endosymbiotic bacteria. Independently, support of an insect rather than bacterial origin of CA has been provided by a series of microscopy studies of the hemolymph from various *Dactylopius* spp. These studies described the existence of special granulocytes with a high concentration of secretory (M-) granules containing red pigments and small corpuscles of CA floating freely in the hemolymph (Joshi and Lambdin, 1996; Caselin-Castro et al., 2008, 2010).

4.2. Model for the biosynthesis of carminic acid

Based on its structure, the biosynthesis of CA may be hypothesized to proceed by two different routes (Figure 4). One envisioned biosynthetic scheme involves the polyketide pathway, also known as the acetate/malonate pathway (Figure 4A). A second possible biosynthetic route involves the shikimate based chorismate/O-succinyl benzoic acid pathways (Figure 4B). Both biosynthetic schemes ultimately result in the formation of anthraquinones; however they would be predicted to include different intermediates, e.g. a unique anthrone in the case of a polyketide-based pathway, which can be used to distinguish between the two. Detection of the FK anthrone (FKA) in fresh and frozen *D. coccus* material (Figure 1) provides support for a polyketide rather than shikimate origin of compounds with an FK core. That the anthrone is detected in the current study can likely be attributed to the milder extraction conditions, and the use of fresh material rather than dried insects or dyed textiles that have been used in previous reports on the subject. The anthrone is abundant in the fresh material as evident by the HPLC-HRMS-DAD analysis (Figure 1), but is not detected in dried insects, which is in good agreement with our previous observation that the purified anthrone spontaneous oxidizes. Microbial based reduction of the FK anthraquinone to yield the FKA anthrone is a possible alternative explanation for detection of the anthrone. de Witte and co-workers have previously shown that bacteria isolated from mammalian fecal material are capable of catalyzing the reduction of the anthraquinone rhein to the corresponding anthrone. The currently available data do not allow us to rule out this explanation, and further experiments e.g. feeding experiments are hence required (de Witte et al. 1992).
Based on the detection of FKA and the novel CA related compounds identified in D. coccus, and under the assumption that its formation is not the result of microbial reduction, we here propose a biosynthetic pathway for the formation of CA in D. coccus (Figure 5). This biosynthetic pathway is an elaboration of the models previous proposed by Brown (Brown, 1975) and Morgan (Morgan, 2010), and differs by including additional intermediates and predictions for the required enzymatic activities and co-factors.

The enzymatic machinery, responsible for the formation of polyketides in animals, remains unknown, and several competing hypotheses exists. One possible explanation could be that the involved PKS has been introduced into the genome of scale insects by horizontal gene transfer (HGT) from fungi or bacteria.

Several examples of HGT from fungi to insects have previously been documented such as the carotenoid forming pathway in Acyrthosiphon pisum (pea aphid) (Moran and Jarvik, 2010). Synthesis of the FK core requires the formation of a C7-C12 intermolecular bond in the octaketide backbone (Figure 5). Since fungal type I iterative PKSs have only been described to form either C2-C7 or C6-C11 bonds, fungi are an unlikely donor (Li et al., 2010). Engineered bacterial type II PKS systems have previously been shown to be able to produce FK, known as TMAC in the bacterial literature (Tang et al., 2004). Specifically, TMAC is formed by the combined actions of the minimal actinorhodin PKS (act-KSα, act-KSβ, act-ACP) from Streptomyces coelicolor and the two cyclases (ZhuI and ZhuJ) from Streptomyces sp. No. 1128 (Tang et al., 2004). To settle whether HGT has formed the basis for CA production requires that the responsible genes are identified and analyzed in the context of a high-quality scale insect genome sequence, which is not presently available.

A competing hypothesis for the origin of insect PKSs, and the one we favor, is that the putative PKSs may have evolved from the insect’s endogenous type I fatty acid synthase (FAS). Animal type I FAS and fungal type I iterative PKS are thought to have evolved from a common bacterial type I PKS ancestor (Hopwood and Sherman, 1990; Kroken et al., 2003). The main product of FASs in animals is palmitic acid, a fully reduced C-16 chain. This chain length is equivalent to an octaketide, which is the intermediate required for FKA formation. Converting a FAS to a non-reducing PKS capable of producing a non-reduced linear octaketide would require inactivation of the FAS’s β-ketoreductase (KR) domain combined with a relaxation of the substrate specificity of its β-ketosynthase domain (KS) to allow for non-reduced products to form.

Non-reduced linear polyketides are highly reactive, due to the presence of carbonyl groups on every second carbon atom, and they spontaneously fold into heterocyclic and aromatic structures via the formation of intramolecular C-C bonds. Several studies have shown that non-reduced octaketides spontaneously form the aromatic compounds SEK4 and SEK4b, which contain intermolecular bond configurations that differ significantly from that of FKA (Figure 5) (Fu et al., 1994; Mizuuchi et al., 2009). The fact that we find only FKA, and not SEK4 or SEK4b, in D. coccus extracts (data not shown) suggests that folding of the polyketide chain does not proceed as a spontaneous reaction. Folding control of non-reduced polyketide backbones in
fungal type I iterative non-reducing PKS systems and bacterial type III PKS systems are achieved by a
‘Product Template’ (PT) domain in the PKS (evolved from DH domain) or by trans-acting cyclases and
aromatases, respectively (Shen et al., 1995; Bringmann et al., 2006). In the case of coccid pigments,
controlled folding of the linear octaketide to form FKA could depend on similar mechanisms that would
require additional mutations in the FAS or trans-acting enzymes as presented in Figure 5. It has not escaped
our notice that the mutated FAS hypothesis potentially also can explain the formation of other polyketides
found in insects, such as 5-hydroxy-7-methyl-6-acetylporpurin from Ericoccus spp. (Coccoidea: Ericoccidae)
(Banks and Cameron, 1970), chrysophanol in Galeruca tanaceti (Coleoptera: Chrysomelidae) (leaf beetle)
(Bringmann et al., 2006), and the predicted monomeric precursors of protoaphins in aphids (Brown, 1975).
The listed compounds are all likely also formed from non-reduced octaketide precursors but display
alternative backbone folds and would hence dependent on other cyclases than those involved in coccid
dyes biosynthesis.

In the case of CA formation, the enzymatic steps following formation of FKA are predicted to include two
oxidations and a C-glucosylation. Based on the metabolites detected in D. coccus, monooxygenation of the
central aromatic ring (position C-10), from FKA to FK, likely occur before oxidation of the outer ring
(position C-4) and before C-glucosylation (position C-2), as neither of the detected metabolites contain a C-
4 oxidation without a C-10 oxidation and as all known glucosylated intermediates (e.g. dcII and CA) have
the C-10 oxidation. Several studies of bacterial and fungal systems have shown that efficient in vivo
anthrone oxidation is dependent on specific anthrone oxidases (Chung et al., 2002; Erhlich et al. 2010),
making it likely that FKA to FK conversion is an enzyme dependent reaction in vivo, rather than a
spontaneous reaction. The order of the two subsequent reactions (monooxygenation of C-4 and C-
4-glucosylation at C-2) that ultimately yield CA is unclear as both KA and dcII accumulates. Hence, it is
impossible to determine whether both pathways are active in vivo or whether one represents a shunt.
Monooxygenation of the C-4 position that converts FK to KA is likely catalyzed by either a cytochrome P-
450 or flavin-dependent monooxygenases. The accumulation of several intermediates suggests that the
natural CA biosynthetic pathway is imbalanced; a situation that likely is caused by insufficient flux through
downstream enzymatic steps resulting in the buildup of intermediates.

4.3. Decarboxylation of the FKA core
We also demonstrated the presence of DE, DE-O-Glcp, and DDE-3-O-Glcp in freshly collected insects. DE
has previously been observed in air-dried D. coccus (Sugimoto et al., 1998). Dehydroxy- and O-glucosylated
forms of DE had not previously been reported in D. coccus. The reason why we detect the O-glucosylated
forms may be attributed to the mild extraction conditions and moderate pH compared to previously
reported extraction protocols, which may result in hydrolysis of O-glucosides. Kerria laccas (lac insect) and Austrotachardia acacia (Maskell) are also known to accumulate DEL, erythrolaccin (EL) and iso-erythrolaccin (IEL) in their resin deposits (Chauhan, 1977; Caselin-Castro et al., 2010). The structural similarity and co-occurrence in multiple species suggest a common biosynthetic origin for the FK (C16) and EL (C15) compound families. This is further supported by the observation made by Chauhan and Mishra (Chauhan and Mishra, 1977) who noted that a single mutation in white K. lacca strains affected both the body color (primarily caused by LA) and the resin color (primarily caused by EL) (Bhide et al., 1969). Based on this, we propose that the FK (C16) and EL (C15) compound families are products of the same biosynthetic mechanism in scale insects and that the difference in carbon number is due to decarboxylation of FKA, as presented in Figure 5. A highly similar anthraquinone decarboxylation step has previously been documented in the chrysophanol biosynthetic pathway in G. tanaceti, though the responsible mechanism and timing of the decarboxylation step is unknown (Bringman et al., 2006).

4.4. Evolution of the FK biosynthetic pathways / Variations to the FK biosynthetic pathway
CA has been shown to act as a chemo deterrent that protects the immobile scale insects from predatory ants (Eisner et al., 1980). However, García Gil de Muñoz and co-workers recently extended CA’s biological function by showing that it may contribute to the innate immune system of the scale insect to protect against invading microorganisms (Garcia-Gil De Muñoz et al, 2002, 2005; De La Cruz Hernandez-Hernandez, 2003; García-Gil De Muñoz, 2007). Specifically, this system depends on encapsulation of the invaders by melanization via the rapid formation of eumelanin by polymerization of tyrosine and L-DOPA (Satyavathi et al., 2014; Charles and Killian, 2015). Phenoloxidase (PO) is responsible for catalyzing multiple steps of the melanization cascade: tyrosine to L-DOPA, L-DOPA to dopaquinone, dopamine to dopaminequinone, and dopamine to N-arachidonoyl dopamine. Reactions that generate reactive radicals, including reactive oxygen species (ROS) and DOPA semi-quinones, which in addition to encapsulation may harm intruders (González-Santoyo and Córdoba-Aguilar, 2012). As shown by Garcia et al., PO can also act directly on CA resulting in the formation of insoluble polymers of CA (García-Gil De Muñoz et al, 2005; García-Gil De Muñoz, 2007). This process is attributed to the quinone nature of CA allowing it to participate in redox cycling with ROSs. Relocation of the radical within the conjugated system of CA or FK would allow for activation of multiple positions in the FK core (C4, C8, C10, and C9-OH) as described for other naphthoquinones (Frandsen et al., 2006). This is a very interesting observation as the formation of FK radicals potentially can explain how the tyrosine-derived groups found in LAs are added to the FK core by radical activation of the C10 positions and oxidative coupling with tyrosine or one of its derivates (tyrosol, N-acetyltymamine, tymamine or 2-(4-hydroxyphenyl)ethyl acetate) (Figure 6). Figure 6 summarizes the chemical diversity and the required
enzymatic steps in the form of a meta-biosynthetic pathway accounting for all known coccoid pigments. The model includes five different monoxygenases, acting on C4, C6, C8, C16, two dehydrogenases, a decarboxylase and a C-glucosyltransferase. The extensive chemical diversity and existence of multiple alternative decoration patterns support the hypothesis that the FK forming biosynthetic pathway has a long evolutionary history within the Coccoidea superfamily.

In summary, we propose that formation of CA depends on the activity of a modified fatty acid synthase or polyketide synthase, possibly one or more cyclases/aromatases, one anthrone oxidases, a 'cytochrome P450 monoxygenases'/flavin-dependent monoxygenases', and a C-glucosyl transferase. Validation of the proposed hypothetical biosynthetic schemes and the involved enzyme types naturally depends on future biochemical evidence and mapping of their genetic basis in D. coccus or an endosymbiont organism.

Abbreviations
Flavokermesic acid anthrone (FKA); flavokermesic acid (FK); kermesic acid (KA); carminic acid (CA); C-glucosylated flavokermesic acid (dcII); laccaic acid (LA); desoxyerythrolaccin (DE), 5,6-didehydroxyerythrolaccin (DDE); didehydroxyerythrolaccin 3-O-β-D-glucopyranoside (DDE-3-O-Glc\textsubscript{p}); erythrolaccin (EL); iso-erythrolaccin (IEL)

Acknowledgements
The Danish Innovation Foundation is acknowledged for funding the work via grant no. XXXXXXXX. Center for Synthetic Biology funded by the University of Copenhagen’s Excellence Program for Interdisciplinary Research is acknowledged for technical and financial support. The US NSF for funding AVD’s involvement in the project via grant no. DBI-1306489. We would like to thank Cultivo de la cochinilla en Mala y Guatiza for their assistance with collection the live Dactylopius coccus specimens on Lanzarote. We are also thankful for Jimmy Oluf Olsen, from the Royal Botanical Garden, the Natural History Museum of Denmark, University of Copenhagen, for his assistance with collecting.

Conflicts of interest
Authors SAR, KTK, DS, CHG, UT, UHM, TOL, and RJNF declare no financial nor any competing financial interests. The authors, PKJ, MB and BM are or were employed by the private company Chr. Hansen A/S that produces and sells D. coccus derived carmine as a food pigment in a business-to-business setup.
Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/ToBeInserted

References


De La Cruz Hernandez-Hernandez, F., García-Gil De Muñoz, F., Rojas-Martínez, A., Hernández-Martínez, S.,


Frandsen, R.J.N., Nielsen, N.J., Maolanon, N., Sørensen, J.C., Olsson, S., Nielsen, J., Giese, H., 2006. The biosynthetic pathway for aurofusarin in Fusarium graminearum reveals a close link between the naphthoquinones and naphthopyrones. Mol. Microbiol. 61, 1069–1080


Horikawa, M., Hoshiyama, T. S., Matsuzawa, M., Shugyo, T., Tanaka, M., Suzuki, S., Sato, M., Ito, T.,

Joshi, P.A., Lambdin, P.L., 1996. The ultrastructure of hemocytes in Dactylopius confusus (Cockerell), and the role of granulocytes in the synthesis of cochineal dye. Protoplasma 192, 199–216


Petitpierre, E., 1981. Presence of anthraquiones in the haemolymph of Timarcha. Chrysomela 30, 4


**TABLES AND ARTWORK**

**Figure 1.** HPLC-HRMS analysis of CA related compounds found in *Dactylopius coccus* with annotation of major peaks and UV/VIS spectra for the novel compounds. **A:** Annotated EIC chromatogram for the CA related compounds identified in the raw extract of *D. coccus*, coloring scheme for the EIC is shown below the chromatogram. **B:** Annotated EIC chromatogram for CA related compounds identified in the SAX purified *D. coccus* extract, **C-F:** UV/Vis spectra for DE and the new compounds DDE-3-O-Glcp, DE-O-Glcp, and FKA.
Figure 2. Structures of the new compounds DDE-3-O-Glc, and FKA

5,6-dihydrohydroxyerythrolaccin 3-O-β-D-glucopyranoside
DDE-3-O-Glc

Flavokermesic acid anthrone
FKA
Figure 3. Targeted metabolite analysis of coccid dyes in representatives of the *Coccus*, *Pseudococcus*, *Hemiberlesia*, and *Palmicultor* genera. Extracted ion chromatograms at $m/z$ 313.0351 ± 0.01 equal to the [M-H]" ion of FK (green), at 329.0246 ± 0.01 for the [M-H]" ion of KA (purple), at 475.0875 ± 0.01 for the [M-H]" ion of dcll (red), and at 491.0832 ± 0.01 for the [M-H]" ion of CA (green).
Figure 5. The two theoretical biosynthetic schemes that can lead to the formation of carminic acid with the first stable tricyclic intermediates shown in boxes. Panel A shows the polyketide based pathway while B shows the O-succinyl benzoic bases mechanism.
Figure 5. Model for carminic acid biosynthesis in Dactylopius coccus, including predicted enzyme types, substrates, and co-factors. Compounds in brackets represent hypothetical intermediates that have not been detected. Signatures: ● compound identified based on HRMS data; ●● compound identified by HRMS and NMR data; ●●● Compound identified based on authentic standard and HRMS data. The numbering of carbon atoms refers to their position in the polyketide backbone, counting from the carbon closest to the enzyme upon completion of the catalytic program.

Enzymes: OxidoR = oxidoreductase; MO = monooxygenase (three different a, b and f); UGT = UDP-glucose dependent membrane bound glucosyltransferase; PKS/FAS = polyketide/‘mutated fatty acid’ synthase; Cyclase/aromatase = small-molecule-foldases as found in bacterial type II iterative PKS systems.
Figure 6. A summative meta-model for the biosynthesis of coccid dyes in scale insects, including predicted enzyme types, substrates, and cofactors. Compounds in brackets represent hypothetical intermediates. The meta-model includes one decarboxylation step and six different monooxygenases (MOa-f) reactions, which are likely catalyzed by different enzymes, based on different positions in the FK core that is modified. The oxidative coupling reactions, branching out from the two hypothetical intermediates, are likely unspecific reactions that occur via highly reactive radicals generated by the insect’s innate immune system.
• Two novel coccid pigment intermediates isolated from *Dactylopius coccus*, including flavokermesic acid anthrone.
• Detection of the anthrone for the first time provides solid evidence for biosynthesis via the polyketide biosynthetic pathway.
• Coccid pigments is much wider spread within the Coccoidea superfamily than previously believed, suggesting a common genetic basis and evolutionary origin.