**Clostridium scindens**: a human gut microbe with a high potential to convert glucocorticoids into androgens

Jason M. Ridlon,†§-shigeo Ikegawa,§-joão M. P. Alves,***-biao Zhou,†† akiko Kobayashi,†† takashi Iida,†† kuniko mitamura,§- genzoh tanabe,§ myrna serrano,***-ainee de guzman,§§ patsy cooper,† gregory a. buck,†§ and phillip b. hylemon*†

Department of microbiology and immunology* and center for biological complexity,** virginia commonwealth university, richmond, va 23298; veterans affairs mcguire medical center,† virginia commonwealth university, richmond, va 23249; faculty of pharmaceutical sciences,§ kinki university, kowakae, higashi-osaka 577-8502, japan; department of chemistry,†† college of humanities & sciences, nihon university, sakurajousui, setagaya, tokyo 102-0074, japan; and department of biology,§§ university of virginia, charlottesville, va 22903

Abstract **Clostridium scindens** american type culture collection 35704 is capable of converting primary bile acids to toxic secondary bile acids, as well as converting glucocorticoids to androgens by side-chain cleavage. the molecular structure of the side-chain cleavage product of cortisol produced by *C. scindens* was determined to be 11β-hydroxyandrost-4-ene-3,17-dione (11β-OHA) by high-resolution mass spectrometry, 1h and 13c NMR spectroscopy, and X-ray crystallography. using RNA-seq technology, we identified a cortisol-inducible (19900-fold) operon (desABCd) encoding at least one enzyme involved in anaerobic side-chain cleavage. the desC gene was cloned, overexpressed, purified, and found to encode a 20α-hydroxysteroid dehydrogenase (HSDH). this operon also encodes a putative “transketolase” (desAB) hypothesized to have steroid-17,20-desmolase/oxidase activity, and a possible corticosteroid transporter (desD). RNA-seq data suggests that the two-carbon side chain of glucocorticoids may feed into the pentose-phosphate pathway and are used as a carbon source. the 20α-HSDH is hypothesized to function as a metabolic “rheostat” controlling rates of side-chain cleavage. phylogenetic analysis suggests this operon is rare in nature and the desC gene evolved from a gene encoding threonine dehydrogenase. the physiological effect of 11β-OHAD on the host or other gut microbes is currently unknown.—Ridlon, j. m., s. IKEGAWA, j. M. P. Alves, B. Zhou, A. Kobayashi, T. Iida, k. mitamura, G. tanabe, M. serrano, A. de guzman, P. cooper, G. a. buck, and P. B. hylemon. **Clostridium scindens**: a human gut microbe with a high potential to convert glucocorticoids into androgens. J. lipid Res. 2013. 54: 2437–2449.

Supplementary key words RNA-seq • microbiome • steroid

In the intestinal ecosystem, microbes have evolved enzymes to modify a number of host bile acids and steroids (1, 2). During the enterohepatic circulation (EC), bile salts are synthesized in the liver, concentrated in the gallbladder, and function in the lumen of the small intestine to absorb dietary lipids and limit microbial growth at the site of nutrient uptake (3). Most of these bile salts are actively transported from the lumen of the ileum into the portal circulation and back to the liver; however, roughly 600–800 mg of bile salts escape into the cecum where they are rapidly deconjugated to free bile acids (1, 3). Bile acid 7α/β-dehydroxylating bacteria are organisms capable of converting primary bile acids made by the host to harmful secondary bile acids, deoxycholic acid, and lithocholic acid (1). These bacteria normally comprise a small proportion of the gut microbiota (~10⁹–10¹⁰/g wet weight) and consist of species within the genus *Clostridium* (4).

Host glucocorticoids which enter the large bowel are metabolized by at least one isolate of bacteria with bile acid 7α-dehydroxylation activity (5). The original isolate, *Clostridium strain* “19”, had the capacity to cleave the side chain from 21α,17α-hydroxy glucocorticoids, converting them into C-19 androgens (5, 6). *Clostridium strain* 19 was later named *Clostridium scindens* (type strain American type culture collection (ATCC) 35704) which means “to cut” (6). Previously, our lab purified and characterized an NAD(P)⁺-dependent 20α-hydroxysteroid dehydrogenase (HSDH) from cell extracts of *C. scindens* ATCC 35704 in addition to a partially purified steroid-17,20-desmolase.
organization, though not in function, among a small number of dehydrogenase and the operon appears to be conserved in pears to have evolved from the gene encoding threonine 20 side chain feeding into the pentose-phosphate pathway. The enzyme is synthesized to proceed by transketolation with the two-carbon product that only a fraction of...}

High-resolution mass spectrometry

High-resolution mass spectrometry (HR-MS) using an atmospheric pressure chemical ionization (APCI-MS) was carried out using a JEOLO AccuTOF JMS-T100LC liquid chromatograph-mass spectrometer (JEOLO, Tokyo, Japan) with an APCI source and coupled to an Agilent 1200 series binary pump (Agilent, Santa Clara, CA) in the negative ion mode. HR-APCI-MS of the sample was carried out in the flow injection mode, using methanol as the mobile phase at a flow rate of 1 ml/min. The ionization conditions were as follows: needle voltage, −2 kV; ion guide peak voltage, 2.5 kV; ion source temperature, 80°C; desolvation temperature, 500°C; orifice 1, 2 and 3 voltages, −85, −5, and −15 V respectively; mass range, m/z 50–1,000; nebulizing gas, N2.

NMR spectroscopy

1H and 13C NMR spectra were obtained on a JNM-ECA 800 instrument operated at 800 and 200 MHz, respectively, with CDCl3 containing 0.1% tetramethylsilane (TMS) as the solvent. Chemical shifts were expressed in δ (ppm) relative to TMS, and the following abbreviations were used: s, singlet; d, doublet; br, broad. The 13C distortionless enhancement by polarization transfer (D2EPT, 90°, and 45°) spectrum was measured between CH3, CH2, and CH, and coherence based on their proton environments. In order to further confirm the 1H and 13C signal detected heteronuclear multiple quantum (HMQC coupling) and 1H detected heteronuclear multiple bond correlation (long-range 1H-13C coupling) experiments were also performed.

X-ray crystal structure determination

Colorless crystals of steroid-17,20-desmolase reaction product suitable for X-ray crystal structure determination were grown by recrystallization of the compound from ether–n-hexane gave colorless needles with a melting point of 197–199°C. A crystal having dimensions of 0.25 × 0.10 × 0.05 mm3 was selected. The X-ray intensity measurements were carried out on a Rigaku Micro007HFM-VariMax Saturn 724R CCD system with a confocal X-ray mirror, using graphite-monochromated Mo Kα radiation (λ = 0.7107 Å) at 220 K. The structure was solved using direct methods and refined by a full-matrix least-squares procedure based on F2 using the CrystalStructure crystallographic software package. The crystal structures were refined with anisotropic temperature factors for all nonhydrogen atoms. The positions of hydrogen atoms were generated theoretically, and were refined using the riding model. Crystallographic details are summarized in supplementary Table II.

Bacterial strains, growth conditions, and sterol induction

C. scindens ATCC 35704 was purchased from ATCC. C. scindens ATCC 35704 was stored as a 30% glycerol stock at −80°C prior to this study, and working stocks were cultivated in chopped meat broth. After overnight growth in Brain Heart Infusion broth (BHI), a 5% inoculum was transferred to 100 ml BHI containing 50 µM cortisol. Cells used for RNA isolation were cultivated to O.D. 600nm of 0.4 (mid-log phase), and quenched in a 1:2 v/v solution of RNALater (Ambion, Grand Island, NY) reagent and kept
overnight at 4°C. Cells were pelleted the following day by centrifuga-
tion and stored until processed at −80°C. Induction was verified by HPLC assay (see below). *Escherichia coli* BL21 (DE3) RIL was purchased from Agilent Stratagene (Santa Clara, CA). *E. coli* was cultivated in Luria-Bertani medium containing appropriate antibiotics. Steroids and bile acids were purchased from Sigma-Aldrich (St. Louis, MO) and Steraloids (Newport, RI).

**HPLC assay for steroid-17,20-desmolase activity**

An HPLC assay was used to verify steroid-17,20-desmolase activity in *C. scindens* ATCC 35704 before proceeding with RNA isolation. This assay was also used to determine whether other fecal isolates with bile acid 7α-dehydroxylating activity also possess steroid-17,20-desmolase activity. Strains exhibiting bile acid 7α-dehydroxylating activity were isolated from human volunteer stool as described previously (11). We utilized an HPLC assay to determine whether strains of bile acid 7α-dehydroxylating bacteria have steroid-17,20-desmolase activity. Cortisol and 11β-hydroxyandrostenedione each absorb at 240 nm owing to their 3-oxo-Δ1 structures. Once late log phase was reached, steroids were extracted from the medium with dichloromethane, dried, resuspended in methanol, and separated on an Agilent Eclipse C18 column (5 μm, 4.6 × 250 mm) integrated on an Agi-
lent 1200 series HPLC. The column temperature was maintained at 40°C, the flow rate was 1 ml/min, and the solvent was 50% methanol in HPLC grade water (Sigma-Aldrich). Concentration was determined by peak area compared with a standard curve of peak area versus known concentrations of cortisol and 11β-
hydroxyandrostenedione.

**Purification of mRNA by magnetic bead-capture hybridization**

Cells were quenched with RNALater solution (Ambion) and stored at −80°C until processing. Total RNA was recovered from cells following disruption by bead beating in the presence of acid phenol. RNA was treated several times with DNAse to remove contaminating genomic DNA, and was further purified using the MEGAclear Kit (Ambion). mRNA was enriched by magnetic-bead capture hybridization using custom biotinylated TEG-spaced oligonucleotides designed along the length of the 18s rRNA and 23s rRNA molecules from *C. scindens* ATCC 35704 (supplementary Table 1).

Cells were stored at −80°C in RNALater solution (Ambion). Cells were then collected by centrifugation and washed with lysis buffer (200 mM NaCl, 20 mM EDTA, diethylpyrocarbonate-treated water) and collected again by centrifugation. Cells were resus-
pended in 500 μl lysis buffer and transferred to 2 ml screw-cap bead-beating tubes (Sarstedt, Germany) to which 200 μl zirconium beads, 210 μl 20% SDS solution (Ambion), and 1 ml 5:1 acid pheno-

nol were added. Cells were then disrupted on a Mini-BeadBeater (Biospec Products, Inc., Bartlesville, OK) at maximum speed twice for 1 min, with tubes kept on ice in between treatments. The aqueous and phenol phases were then separated by centrifuga-
tion, and the aqueous phase was washed once with 1 ml 5:1 acid phenol. Nucleic acids in the aqueous phase were then precipi-
tated at −80°C for 20 min by addition of 1/10 vol 5 M ammo-
nium acetate (Ambion), 1 μl Glycoblue (Ambion), and 1 vol ice-cold isopropanol, followed by centrifugation at 13,600 g for 20 min. RNA (>200 hp) was then purified using the MEGAclear kit (Ambion) according to the manufacturer’s instructions. Contaminating genomic DNA was removed by treatment with TURBO DNAse (Ambion) according to the manufacturer’s instructions. MEGAclear and DNAse treatment steps were repeated once. At this point, RNA purity and integrity was checked spectrophotome-
tically by the A260/A280 ratio, the range of pure RNA falling be-
tween 1.7 and 2.1, and integrity checked by separating RNA on a 1.6% denaturing agarose gel. RNA samples with distinct 23S rRNA and 16S rRNA bands were further purified. In addition, residual genomic DNA contamination was determined by RT-
PCR using the RT-for-PCR kit (Clontech) and 200 ng purified RNA.

Dynabeads M-280 Streptavidin (Invitrogen, Grand Island, NY) was made RNase-free according to the manufacturer’s recommen-
dations. Biotinylated oligonucleotides were bound to beads (5 mg) (through streptavidin) by resuspending in 500 μl diethyl
pyrocarbonate-treated 0.5x SSC (0.075 M NaCl; 0.0075 M triso-
dium citrate dehydrate; 10 mM Tris, pH 7.0; 1 mM EDTA) con-
taining 360 pmol each biotinylated oligonucleotide (dena-
tured by heating to 90°C for 5 min and cooling on ice for 3 min). Beads were then captured on a Promega magnetic stand (Pro-
mega, Madison, WI), the supernatant containing unbound probes aspirated, and residual unbound probes washed twice with one volume 0.5x SSC, and twice with 6x SSC (0.9 M NaCl; 0.09 M trisodium citrate dehydrate; 10 mM Tris, pH 7.0; 1 mM EDTA). Efficiency of binding was determined by comparing A260 nm of oligo-mix after bead incubation and washing and comparing to oligo-
mix without bead incubation. Capture-hybridization was per-
formed on a Biorad C1000 thermocycler (Biorad, Hercules, CA). First, 1 μg total RNA suspended in 35 μl 6x SSC in a RNase-
free 0.2 ml PCR tube was heated to 70°C for 5 min to denature RNA, followed by 0°C for 3 min at which point the thermocycler was paused allowing addition of 150 μg oligo-bound beads. The thermocycler was resumed at a temperature of 68°C for 30 min allowing capture hybridization between rRNA molecules and oligo-bound beads. Beads were then captured on a magnetic stand, and the supernatant was combined with two additional (1 vol) washes with 6x SSC, and the enriched mRNA was centrifuged and magnetically captured to remove residual beads before be-
ing precipitated. RNA bound to beads was quantified by A260 nm after removal by three stringent washes with 0.5x SSC containing 10% formamide at 70°C. Enriched mRNA was quantified on a NanoDrop 2000 (Thermo Scientific, Waltham, MA) and stored at −80°C until sequencing.

**Whole transcriptome sequencing**

MRNA-enriched samples were processed using the Roche cDNA synthesis system following the manufacturer’s protocol. Result-
ing cDNA libraries were sequenced in a Roche 454 GS FLX Tita-
nium system (Roche, Indianapolis, IN) using the rapid library protocol according to the manufacturer’s instructions. The se-
quencing yield was of ~300,000 reads for the libraries derived from cortisol-induced cells and from uninduced control cells.

Remaining ribosomal and transfer RNA reads were removed from the sequence dataset by BLAST searches (12) of these reads against a database of known *Clostridium* rRNA and tRNA se-
quences, using a minimum 90% identity cutoff. Non-rRNA reads were then mapped to the *C. scindens* ATCC 35704 genome using Burrows-Wheeler transformation (13), and resulting align-
ments were analyzed by CuffLinks (14) for differential expres-
sion detection.

**Cloning and overexpression of desC gene in E. coli**

Genomic DNA was isolated from *C. scindens* ATCC 35704 by enzymatic treatment, followed by bead beating as described pre-
viously (15). A streptavidin tag was engineered into the reverse primer and the desC gene was PCR amplified, restriction digested, and ligated into the expression vector pSport1.

Primer sequences were as follows: pSportCTSBPEDS07887.1F, 5′-AAGCTTTTGGTGATATCAGAAAGGAGGAAAAAATGAGAC-3′; and pSportCTSBPEDS07887.1R, 5′-GACGTCATTATTGTCACAGTGGGGCTGGGTTGCTCCATTCGT CCATCTTAATTACGATCT TTGC. PCR was performed using the TITANIUM Taq kit (Clontech) on
a Biorad C1000 thermocycler. PCR products and pSport1 were then cut with the appropriate restriction enzymes (New England Biolabs, Ipswich, MA) and ligated with T4 DNA ligase (New England Biolabs). Plasmids were transformed into E. coli DH5α by heat shock at 42°C and cultivated for 1 h in Super Optimal Broth medium (Invitrogen) before plating on LB ampicillin (100 µg/ml). Both strands of the insert were sequenced using T7 promoter and terminator primers. The protein was overexpressed in E. coli BL21(DE3)RIL and purified with Strep-Tactin affinity chromatography. Colonies were grown in LB ampicillin and induced for overexpression by addition of Isopropyl β-D-1-thiogalactopyranoside to 1 µM final. Cultures were screened for expression by Western blot hybridization with the Strep Tag II antibody (IBA, Kansas City, MO). Cell extracts were prepared by treating cells with 5 µg/ml lysozyme in buffer A (20 mM sodium phosphate buffer, pH 7.0; 0.1 M NaCl; 20% glycerol; 10 mM 2-mercaptoethanol) on ice for 1 h followed by two passes through a French pressure cell at 1,500 psi. Cell extract was then centrifuged for 30 min at 16,000 g. The supernatant was then applied to a Strep-Tactin affinity column equilibrated with buffer A. The column was washed with buffer A until protein was not detected in the eluent. Recombinant proteins were eluted in buffer A containing 2.5 mM desthiobiotin.

**20α-HSDH enzyme assay**

20α-HSDH activity was measured by continuous spectrophotometric assay at 340 nm performed in buffer A containing 50 µM cortisol (Steraloids, Inc., Newport, RI) and 150 µM NADH. A molar extinction coefficient of 6,220 M⁻¹ cm⁻¹ was used to determine product concentration. Reaction rates were performed under varying substrate concentrations and above the respective Kₚ values. Initial linear region of the reaction progress curves was utilized for determining the reaction rates. Kₚ and Vₘₐₙ values were determined by Lineweaver-Burk plots of the data.

**Phylogenetic analyses of desC**

Putative orthologs of desC were identified by BLAST (9) search of the C. scindens ATCC 35704 protein sequence against the National Center for Biotechnology Information. Non-redundant protein database, accepting no more than 10,000 matches and employing a maximum E-value cutoff of 1e-20. The resulting sequence set was then filtered to contain at most one strain per species and to remove duplicate sequences and those that were too long or too short, defined respectively as being longer than 1.5 times or shorter than 0.5 times the average sequence length (356 amino acids) of the original sequences. To save computational time, the resulting set of 2,045 protein sequences was submitted to a two-step phylogenetic analysis. For both steps, sequences were aligned using MUSCLE version 3.8.31 (16) and submitted to maximum likelihood analysis using RAxML version 7.2.8 (17); the substitution model employed was WAG (18) using empirical residue frequencies and the γ distribution model of rate heterogeneity (model PROTGAMMAWAGF). The first step consisted of one detailed tree search, in order to get a first overview of phylogeny and identify the region where C. scindens' desC and its nearest neighbors would be located. The 341 sequences from the resulting subtree containing C. scindens' desC and its nearest neighbors plus a few outgroups from another smaller subtree were then submitted to the thorough maximum likelihood tree searches. We performed bootstrap analysis with 100 pseudoreplicates, and bipartition frequencies were then drawn on the best tree out of 20 independent tree searches. Trees were drawn and formatted using TreeGraph2 (19) and Dendroscope (20), and further cosmetic adjustments were done in the Inkscape vector image editor. As a reference species tree of the clostridia, we used the automated phylogenomic analysis from PATRIC, which shows C. scindens as distantly related to most other Clostridium species, being deeply embedded in the Lachnospiraceae family and with several Ruminococcus species in close phylogenetic proximity. The Clostridium genes identified as closely related to C. scindens' desC belong to species only distantly related to this cluster of C. scindens, Ruminococcus, and the Lachnospiraceae. However, no member of the latter two taxa presents a gene closely related to C. scindens' desC. On the other hand, desC grouped with a group that, besides the clostridia (Clostridium leptum DSM 753 ZP_02081557, Clostridium ljungdahlii DSM 13528 YP_003780744 and YP_003780741, Clostridium sp. MSTE9 EJF39065, and Clostridium carboxidivorans P7 ZP_05390048), also contained a sequence from the Spirochaetes Treponema primitia ZAS-1 (ZP_09717055) and the actinobacteria Modestobacter marinus (YP_006366820); sequences in this group were very similar to each other, with BLAST E-values of around 1E-50 or better in comparison to C. scindens' desC. The sister group for this clade is composed of two Thermotogaspecies (Thermotoga maritima MSB8 NP_228110 and Thermotoga petrophila RKU-1 YP_001244210) and a group of Metazoa (15 insects, and one each of Cnidaria, Cephalochordata, and Tunicata); in this case, similarity compared with desC was much lower, with values between 1E-35 and 1E-21. The whole assemblage described above forms a clade with the very high bootstrap support value of 97, and is not closely related to other Firmicutes bacteria present in the tree.

**RESULTS**

**Identification of the side-chain cleavage product of cortisol by C. scindens ATCC 35704**

The steroid product formed by whole cells and cell extracts of C. scindens ATCC 35704 was extracted and purified by a combination of TLC and HPLC. The androgen exhibited an Rₜ of 0.78 on TLC plates using solvent A. The androgen had a retention time of 43.9 min on a semi-preparative C₁₈ reversed-phase HPLC column as compared with cortisol, which eluted at 35.2 min. Reaction product was recrystallized from ether-n-hexane to give colorless needles (mp, 197–199°C; lit*197–199°C) (21). The structural determination was performed by a combination of techniques using HR-MS, ¹H and ¹³C NMR, and X-ray diffraction analysis. HR-MS of the purified compound showed the deprotonated molecule [M-H]⁻ at m/z 301.18038,
which is entirely identical with the calculated mass of 301.18037 (Fig. 1). The molecular structure of the steroid-17,20-desmolase product as 11β-OHA was determined unambiguously by $^1$H and $^{13}$C NMR and X-ray diffraction analysis of crystals of the steroid product (Fig. 2; supplementary Tables II–IV). We did not detect any accumulation of 11β-hydroxytestosterone or 11β-hydroxyepitestosterone from either whole cells or cell extracts of this bacterium. In cell extracts, the reaction proceeded under strictly anaerobic conditions and was not stimulated by NAD$^+/NADP^+$ suggesting a novel steroid-17,20-desmolase/oxidase catalyzing this biotransformation. For more structural details, see supplementary text.

**Identification of cortisol-regulated genes in *C. scindens* ATCC 35704 using RNA-Seq technology**

In previous studies we reported that steroid-17,20-desmolase activity was characteristic of *C. scindens* ATCC 35704, the type strain (6). However, we detected this activity in only three out of 31 strains of 7α-dehydroxylating human gut bacteria, many of which were identified as *C. scindens* by 16s rDNA gene sequencing (supplementary Table V). We chose *C. scindens* ATCC 35704 for identification of cortisol-inducible genes due to the presence of steroid-17,20-desmolase and 20α-HSDH activity and a genome sequence on which to scaffold RNA-Seq data. In order to use RNA-Seq as a method for identifying cortisol-inducible

**Fig. 2.** The molecular structure of the steroid-17,20-desmolase reaction product identified as 11β-hydroxyandrosten-3,17-dione. An Oak Ridge Thermal Ellipsoid Plot (ORTEP) view of the molecule with atomic labeling (thermal ellipsoids are drawn at 50% probability) (See also supplementary Table II for crystallographic details and supplementary Table IV for atomic coordinates, bond angles, and bond lengths) Oxygen atoms are in red, carbon atoms in grey, and hydrogen atoms in white. A: Side view of the molecule. B: Top view of the molecule. C: Standard chemical formula of the molecule (top left), stereochemical formula of the molecule displaying $^{13}$CNMR shifts (ppm) (top right), and stereochemical formula of the molecule displaying $^1$H NMR shifts (ppm) (bottom). See supplementary text and supplementary Table III for further description.
genes in *C. scindens* ATCC 35704, it was necessary to develop a method to enrich the mRNA, as it represents only about 1–5% of the total cellular RNA of bacteria. In this regard, we designed oligonucleotides along the length of the 16s rRNA and 23s rRNA molecules based on genomic sequences from *C. scindens* ATCC 35704 and *C. scindens* VPI 12708 (supplementary Table I) (22). Oligonucleotides were synthesized with biotinylated 3′ TEG spacers allowing for magnetic capture-hybridization of rRNA on Dynabeads linked to streptavidin (23). Before mRNA purification, we measured steroid-17,20-desmolase activity to verify cortisol induction (Fig. 3A). Quantification of the mRNA fraction

Fig. 3. Measurement of cortisol induction, mRNA purification, and effect of cortisol induction on transcriptome of *C. scindens* ATCC 35704. A: Prior to isolation of mRNA for RNA-Seq, whole cells were measured for steroid-17,20-demolase activity by measurement of 11β-OHAD formation (*n* = 3). B: One microgram of total RNA was subjected to bead-capture hybridization (see Materials and Methods for details). We observed ~80% removal of total RNA (Elution) by streptavidin bound magnetic beads coated with biotinylated oligonucleotides designed to bind 16s rRNA and 23s rRNA molecules. C: RNA gel demonstrating reduction in intensity of rRNA bands following bead-capture hybridization. Control (Ctrl) sample (10 μg total RNA) was subjected to bead-capture hybridization with beads lacking biotinylated capture oligos. Enrichment (Enrich) fraction containing enriched mRNA was sequenced. D: Results of RNA-Seq experiments comparing transcriptome data from control cells grown in BHI without cortisol and cortisol-induced cells. Upregulated genes are represented in blue, downregulated in red. Gene annotations from BLAST were used to group genes by general function (see supplementary Dataset I).
(enrichment fraction) as well as the RNA eluted from magnetic beads coupled to our custom biotinylated rRNA capture oligos following capture (elution fraction) showed that our method removed ~80% of the total RNA. Diminution of rRNA band intensity was observed between control total RNA (Ctrl), which used the same procedure as the mRNA enrichment sample (Enrich) except that capture oligos were omitted (Fig. 3B, C). Sequencing of the cDNA libraries resulted in around 300,000 reads using this sequencing technology for each library (cortisol-induced and uninduced control). Filtering of the ribosomal and transfer RNA sequences from these data showed that our rRNA depletion method removed about 90% of the rRNA from the libraries. However, considering that such sequences usually comprise ~95–99% of the total extracted RNA, sequenced libraries still ended up with a 9:1 ratio of rRNA to non-rRNA sequence reads, with a final number of non-rRNA reads of 28,858 for the uninduced library and 27,701 for the cortisol-induced library. A combination of MEGAClear spin column and a pyrosequencing process allowed the removal of small RNAs (<100 bp). Only about 300 reads matched known tRNA sequences.

We have previously reported the characterization of the baiBCDAFGH operon in C. scindens, which encodes genes involved in bile acid-inducible 7α/β dehydroxylation (1). We expected that if our method for isolating mRNA were valid, we should expect to see these genes significantly upregulated in cholic acid-treated cells. Experiments with 50 μM cholic acid induction were performed as described for cortisol (see Materials and Methods). Indeed, the bai operon was upregulated 452-fold as compared with control cells (supplementary Fig. 1).

Identification of a cortisol-inducible operon in C. scindens ATCC 35704

The genome (4.336 kbp) of C. scindens is estimated to encode 3,995 genes, of which a total of 186 genes were upregulated (2- to 1,000-fold) while 97 genes were downregulated (20- to 1,000-fold). Fifteen genes for enzymes in the pentose-phosphate pathway were also upregulated by cortisol induction including: xylulose-5-phosphate isomerase (400-fold), ribulose-5-phosphate isomerase (276-fold), and ribose-5-phosphate isomerase (10-fold) (supplementary Dataset S1). This is expected if the two-carbon side chain of cortisol is feeding into the pentose-phosphate pathway. Fifteen genes for vitamin B12 biosynthesis were upregulated by cortisol (supplementary Dataset S1). The physiological function of B12 in C. scindens cortisol metabolism is currently unknown.

The desC gene encodes an adrenocorticosteroid 20α-HSDH

It has been previously reported that C. scindens ATCC 35704 encodes steroid-inducible 40 kDa 20α-HSDH (7). Indeed, this enzyme has been purified and characterized, though the gene encoding this enzyme has yet to be identified (7, 8). An N-terminal sequence was previously reported for the 20α-HSDH; however, characterization of this and a second gene with matching N-terminal sequences showed that these genes encode enzymes with GADPH activity (supplementary text). Bioinformatic analysis of the desC (EDS07887.1) gene indicated that it encodes an NAD(P) dependent oxidoreductase, M, 38.5 kDa, induced 1,000-fold by cortisol, making this gene product a good candidate for the 20α-HSDH. The desC gene was cloned into pSport1 as an N-terminal streptavidin-tagged (ST) recombinant protein and purified by Strep-Tactin affinity chromatography to apparent electrophoretic homogeneity with a M, of 40 kDa (Fig. 5).

We used several different glucocorticoid substrates to determine the specificity of this enzyme. We observed NADH-dependent reductive activity with cortisol (11β,17α,21-trihydroxy pregn-4-ene-3,20-dione), but not 20α-cortisol (11β,17α,20α,21-tetrahydroxy pregn-4-en3-one), and 20β-cortisol (11β,17α,20β,21-tetrahydroxy pregn-4-en-3-one) was not a substrate for this enzyme. When we tested each of these steroid substrates in the oxidative direction (NAD+) only 20α-cortisol was a substrate. This data suggests that the desC gene encodes an NAD+ dependent 20α-HSDH. The recombinant desC-ST showed an apparent K of 5.35 μM and V, of 126 nmol/min/mg protein for cortisone,
and an apparent $K_m$ of 1.46 μM and $V_{max}$ of 30 nmol/min/mg for cortisol in the presence of NADH. The apparent $K_m$ and $V_{max}$ for 20α-hydroxy cortisol was determined to be 4.54 μM and 24 nmol/min/mg protein, respectively. Substrates lacking a 17α-hydroxy group such as 11β,21-dihydroxyprogren-4-ene-3,20-dione or those lacking a 21-hydroxy group such as 11β,17α-dihydroxyprogren-4-ene-3,20-dione were not substrates. Finally, we did not detect 17α-HSDH activity against the end product of the steroid-17,20-desmolase reaction, 11β-OHA. These data are consistent with characteristics determined for the purified native 20α-HSDH isolated from this bacterium which recognize adrenocorticosteroids with 17α,21-dihydroxy groups (7, 8).

**Phylogenetic analysis of desC**

Annotation of the desC gene product suggests this gene evolved from threonine dehydrogenase (EC 1.1.1.103). We performed an extensive phylogenetic analysis to try to better understand the distribution of this gene in nature, and in particular the gut. Analysis of strains of C. scindens and other species of bile acid 7α/H25β-dehydroxylating bacteria suggest the desC gene is rare even in strains of C. scindens (supplementary Table V). The first step in the phylogenetic analysis of DesC, involving 2,045 sequences, identified a subtree of 341 sequences that contained C. scindens' DesC and its nearest neighbors (Fig. 6A). Thorough analysis of these 341 sequences revealed the phylogenetic placement of DesC among sequences from a few other clostridia, one Actinobacteria (Modestobacter), one Spirochaetes (Treponema), a group of Metazoa (mostly insects), and two Thermotoga species, with the high bootstrap support value of 97 (Fig. 6B). All other Firmicutes sequences present in this subtree are distant from C. scindens' DesC, including another C. scindens sequence (ZP_02432273, also from strain ATCC 35704). This sequence presents a much stronger match (E-value of 0) to the threonine dehydrogenase-like domain cd08234 than DesC (E-value 7e-63), and it is also placed in a group containing many other Firmicutes bacteria. Finally, if we examine genes surrounding desC-like genes in the cluster of microbial genes close to C. scindens' desC and found a high degree of conservation of the transketolase subunits (corresponding to desAB in C. scindens) and the Na+/melibiose transporter (desD), hypothesized to encode a cortisol transport protein (Fig. 6C). Each of the genes in this cluster is associated with organisms that live in anaerobic environments, including the gut of animal species (24–27).

**DISCUSSION**

C. scindens and a small number of species belonging to the genus Clostridium are responsible for significant alterations in the human bile acid pool composition through...
Androgen production by *Clostridium scindens* Q1 2445

Small bowel (37, 38) and inducers of antimicrobial peptides (39). Perturbations in the biliary bile acid pool can be indicative of hepatogastrointestinal diseases such as fat malabsorption (40), gallstones (3), gastrointestinal cancers (41), and possibly type II diabetes (42). Recent studies also illustrate the important role gut microbes play in the levels and profiles of bile acids in various tissues of the body including: liver, kidney, plasma, serum, and heart (43). The physiological significance of these alterations in bile acid profiles in host organs such as heart and kidney is unknown. However, decades of research strongly suggest that secondary bile acids are bile acid 7α/β dehydroxylation (1). Bile acids serve as natural ligands, with varying affinities, for nuclear receptors (farnesoid X receptor, pregnane X receptor, Vitamin D) and the G protein-coupled receptors TGR5 and SIP2 (28–32). Secondary bile acids, products of gut microbial metabolism, are the most potent activators of TGR5 (33, 34). Through activation of these receptors, bile acids regulate their own synthesis, conjugation, transport, and detoxification (reviewed by 30, 31), and are important regulators of lipid, glucose, and energy homeostasis (35, 36). Furthermore, bile acids play an important role in maintaining intestinal barrier function as antimicrobial agents in the small bowel (37, 38) and inducers of antimicrobial peptides (39). Perturbations in the biliary bile acid pool composition can be indicative of hepatogastrointestinal diseases such as fat malabsorption (40), gallstones (3), gastrointestinal cancers (41), and possibly type II diabetes (42). Recent studies also illustrate the important role gut microbes play in the levels and profiles of bile acids in various tissues of the body including: liver, kidney, plasma, serum, and heart (43). The physiological significance of these alterations in bile acid profiles in host organs such as heart and kidney is unknown. However, decades of research strongly suggest that secondary bile acids are
microbiota form multiple metabolites from 11\(\beta\)-OHAD produced by \(C.\ scindens\). The gut microbiota are capable of reducing the 17-oxo group of 11\(\beta\)-OHA to either the \(\alpha\)- or \(\beta\)-configuration generating 11\(\beta\)-hydroxy-epitestosterone or 11\(\beta\)-hydroxy-testosterone, respectively (53, 54). In addition, some members of the gut microbiota can reduce the \(\Delta^1\) bond in ring A to either \(\delta\)- or \(\beta\)-configuration resulting in \(\text{trans}\) or \(\text{cis}\) A/B rings, and can reduce the 3-oxo group to either \(\alpha\)- or \(\beta\)-hydroxy groups (Fig. 7) (2). Thus steroid biotransformation products from one microbe yield further metabolites through the metabolic activity of other members of the human microbiome (Fig. 7). Microbial biotransformation products of host bile acids and steroids may play physiological roles in the host-microbiome axis.

In mammalian tissues, side-chain cleavage of steroids requires molecular oxygen-dependent monooxygenases (9). Side-chain cleavage of cortisol under anaerobic conditions by the gut microbiome has been known since 1959 (48, 55, 56); however, a mechanism has not been suggested until now. We hypothesize that the \(\text{desAB}\) genes encode the steroid-17,20-desmolase/oxidase and that this reaction proceeds by transketolation (Fig. 4 and Fig. 8). Previous work reported that TPP addition stimulates side-chain cleavage in cell extracts of \(C.\ scindens\) ATCC 35704 (7). Indeed, we observed a marked induction by cortisol of a number of genes involved in the biosynthesis of TPP (Fig. 8).

We have identified and characterized a cluster of coexpressed genes induced by cortisol, one of which encodes a 20\(\alpha\)-HSDH. Reduction of the 20-oxo group to either \(\alpha\)- or \(\beta\)-configuration blocks side-chain cleavage of C\(_{21}\)-steroids by \(C.\ scindens\) (23). If the two-carbon fragment of
Androgen production by *Clostridium scindens* Q1 2447

show this gene as a very rare occurrence in clostridia and present only in five *Clostridium* species other than *C. scindens* ATCC 35704 (all of them distantly related to *C. scindens*). This suggests that desC is a novel ortholog derived from duplication (of an ancestral gene encoding a threonine dehydrogenase-like protein) followed by functional divergence and either gene loss in most clostridia or transfer from one species to a few select other species. There is also conservation of genes in the desABCD operon from *C. scindens* ATCC 35704 in organisms whose "desC" genes are most closely related, suggesting this operon may have passed in toto. Alternatively, this gene, and operon, could have been transferred to these organisms from a currently unknown source. Interestingly, multiple strains, which by 16s rDNA gene analysis were identified as *C. scindens*, were found to lack steroid-17,20-desmolase activity. This includes *C. scindens* VPI 12708 (formerly *Eubacterium* sp. strain...
VPI 12708) from which most of our knowledge of the bile acid 7α-dehydroxylation pathway has been elucidated (1). Interestingly, *C. scindens* VPI 12708, but not *C. scindens* ATCC 35704, expresses a 17α-HSDH, suggesting some niche partitioning of host glucocorticoid metabolism within a species of bile acid 7α-dehydroxylating bacteria (34). Reduction of o xo-groups by bacteria may in part serve to regenerate NAD⁺ necessary for fermentation. Host steroids and bile acids can therefore serve as electron sinks.

We now possess tools in the scientific community to uncover the relationship between our resident microbiota and the host. The “omics” revolution allows us to identify and quantify thousands of metabolites (metabolomics), and the genes encoding enzymes (genomics), the level at which genes are transcribed (transcriptomics), or the enzymes themselves (proteomics). However, this data is only as good as our ability to assign function or potential function to the genes. Steroids are important host signaling molecules. It has become clear that microbial metabolism to the genes. Steroids are important host signaling enzymases themselves (proteomics). However, this data is only significant physiological and pathophysiological effects.

Therefore, identification of the genes encoding enzymes responsible for gut microbial biotransformation of host steroids is important for future omics studies of host-gut-microbe interactions in health and disease. The present data suggests that RNA-Seq analysis of inducible genes can aid in the discovery of genes in cultivated organisms known to carry out steroid-biotransforming reactions. It may also be possible to use this approach for gene discovery in mixed populations of microbes by comparing transcriptomes in the presence and absence of a particular gene inducer. Genes of interest can be cloned and characterized even from “unculturables” through a combination of transcriptomics and metagenomics.

The authors would like to thank Dr. Fusae Takamine (University of Ryukyus) for providing bacterial strains.

REFERENCES

1. Ridlon, J. M., D. J. Kang, and P. B. Hylemon. 2006. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**: 241–250.

2. Macdonald, I. A., V. D. Bokkenheuser, J. Winter, A. M. McLernon, and E. H. Mosbach. 1983. Degradation of steroids in the human gut. *J. Lipid Res.* **24**: 675–700.

3. Vlahcevic, Z. R., D. M. Heuman, and P. B. Hylemon. 1996. Physiology and pathophysiology of enterohepatic circulation of bile acids. In Hepatology: A textbook of Liver Disease. 3rd edition. Vol. I. D. Zakim and T. Boyer, editors. Saunders, Philadelphia, PA. 376–417.

4. Berr, F., G. A. Kukal-Ublck, G. Paunggartner, W. Munzig, and P. B. Hylemon. 1996. 7α-dehydroxylating bacteria enhance deoxycholic acid input and cholesterol saturation of bile in patients with gallstones. *Gastroenterology* **111**: 1611–1620.

5. Morris, G. N., J. Winter, E. P. Cato, A. E. Ritchie, and V. D. Bokkenheuser. 1985. *Clostridium scindens* sp. nov., a human intestinal bacterium with desmolytic activity on corticoids. *Int. J. Syst. Ecol. Microbiol.* **35**: 478–481.

6. Winter, J., G. N. Morris, S. O’Rourke-Locascio, V. D. Bokkenheuser, E. H. Mosbach, B. I. Cohen, and P. B. Hylemon. 1984. Mode of action of steroid desmolase and reductases synthesized by *Clostridium scindens* (formerly *Clostridium strain 19*). *J. Lipid Res.* **25**: 1124–1131.

7. Krafft, A. E., J. Winter, V. D. Bokkenheuser, and P. B. Hylemon. 1987. Cofactor requirements of sterol-17,20-desmolase and 20-α hydroxysteroid dehydrogenase activities in cell extracts of *Clostridium scindens*. *J. Steroid Biochem.* **28**: 49–54.

8. Krafft, A. E., and P. B. Hylemon. 1989. Purification and characterization of 20α-hydroxysteroid dehydrogenase from *Clostridium scindens*. *J. Bacteriol.* **171**: 2925–2932.

9. Nelson, D. R. 1999. Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* **369**: 1–10.

10. El-Kabbani, O., U. Dhagat, and A. Har. 2011. Inhibitors of human 20α-hydroxysteroid dehydrogenase (AKR1C1). *J. Steroid Biochem.* **125**: 105–111.

11. Takamine, F., and T. Imamura. 1995. Isolation and characterization of bile-inducible 7α-dehydroxylating bacteria from human feces. *Microbiol. Immunol.* **39**: 11–18.

12. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.

13. Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

14. Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

15. Stöver, B. C., and K. F. Müller. 2010. *TreeGraph 2*: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**: 7.

16. Edgar, R. C. 2004. *MUSCLE*: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

17. Staatakis, A. 2006. *RAXML-VI-HPC*: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2088–2090.

18. Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**: 691–699.

19. Stöver, B. C., and K. F. Müller. 2010. *TreeGraph 2*: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**: 7.

20. Huson, D. H., D. C. Richter, C. Rausch, T. Deculian, F. Markus, and R. Rupp. 2007. *Dendroscope*: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**: 460.

21. Rao, P. 1961. Manganese dioxide oxidation: the optional introduction of Δ⁻ double bond with simultaneous cleavage of dihydroxyacetonate or 17,20-glycolic side chains in Δ⁻ 5-ketoesteroids. *J. Org. Chem.* **26**: 2149–2152.

22. Pang, X., D. Zhou, Y. Song, P. Fei, W. Jiang, Z. Guo, and R. Yang. 2004. Bacterial mRNAs purification by magnetic capture-hybridization method. *Microbiol. Immunol.* **48**: 91–96.

23. Rey, F. E., J. J. Faith, J. Bain, M. J. Muehlbauer, R. D. Stevens, C. B. Newgard, and J. I. Gordon. 2010. Dissecting the in vivo metabolic potential of two human gut acetogen. *J. Biol. Chem.* **285**: 22982–22990.

24. Tanner, R. S., L. M. Miller, and D. Yang. 1993. *Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial RNA homology group I. *Int. J. Syst. Bacteriol.* **43**: 232–236.

25. Liu, J. S., W. L. Borkwill, G. R. Drake, and R. S. Tanner. 2005. *Clostridium carboxidivorans* sp. nov., a solvent-producing *Clostridium* isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium* scatologenes strain SL1 as *Clostridium drakei* sp. nov. *Int. J. Syst. Evol. Microbiol.* **55**: 2085–2091.

26. Gruber, J. R., R. J. Leadbetter, and J. A. Breznak. 2004. Description of *Treponeum azototrophicum* sp. nov. and *Treponeum primitia* sp. nov., the first spirochetes isolated from termites guts. *Appl. Environ. Microbiol.* **70**: 1315–1320.

27. Xiao, J., Y. Luo, J. Xu, S. Xie, and J. Xu. 2011. *Modestobacter marinus* sp. nov., a psychrotolerant actinobacterium from deep-sea sediment, and emended description of the genus *Modestobacter*. *Int. J. Syst. Evol. Microbiol.* **61**: 1710–1714.

28. Machima, M., A. Y. Okamoto, J. J. Repa, H. Tu, R. M. Learned, A. Luk, M. V. Hull, K. D. Lustig, D. J. Mangelsdorf, and B. Shan. 1998. Identification of a nuclear receptor for bile acids. *Science* **284**: 1362–1365.
