Structural and biochemical characterization of 20β-hydroxysteroid dehydrogenase from *Bifidobacterium adolescentis* strain L2-32

Heidi L. Doden^{1,2}, Rebecca M. Pollet³, Sean M. Mythen^{1,2}, Zdzislaw Wawrzak⁴, Saravanan Devendran^{1,2}, Isaac Cann^{1,2,5}, Nicole M. Koropatkin³, and Jason M. Ridlon^{1,2,6,7}*

From the ¹Microbiome Metabolic Engineering Theme, Carl R. Woese Institute for Genomic Biology, Urbana, IL 61801; ²Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ³Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109; ⁴Northwestern Synchrotron Research Center–LS-CAT, Northwestern University, Argonne, IL 60439; ⁵Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ⁶Cancer Center of Illinois, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ⁷Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801;

Figure S1. Active site mutant thin layer chromatography (TLC) isothermal titration calorimetry (ITC) and circular dichroism (CD).

Figure S2. Structural analysis of 20β-HSDH.

Figure S3. Sequence alignment of 20β -HSDH found in *B. adolescentis* and other structurally similar SDR members.

Figure S4. SDR family subset with extended N-terminus multiple sequence alignment.

Figure S5. Gel filtration chromatography of 20β-HSDH truncation mutants.

Figure S6. Model of 20β-HSDH N-terminal truncation effects.



Figure S1. Active site mutant thin layer chromatography (TLC) isothermal titration calorimetry (ITC) and circular dichroism (CD). (A) Wild-type and active site mutant 20 β -HSDH overnight reaction products separated by TLC. (1) Cortisol standard, (2) 20 β -dihydrocortisol standard, (3) 11 β -hydroxyandrostenedione standard, (4) WT + NADH + cortisol, (5) WT + NADH + 11 β -OHAD, (6) S181A + NADH + cortisol, (7) Y200A + NADH + cortisol. (B) Ligand binding order of S181A and Y200A studied

by ITC. Left panel is 2 mM NADH binding to S181A, middle panel is 1 mM cortisol binding to S181A with 2 mM NADH, right panel is 2 mM NADH binding to Y200A. (C) CD spectra of purified recombinant 20β -HSDH and its active site mutants.



Figure S2. Structural analysis of 20β-HSDH. (A) Fo-Fc omit map of NADH from Chain A and E shown at 2 sigma. (B) 2Fo-Fc map of NADH and surrounding backbone chains from Chain A and E shown at 2 sigma. (C) Alignment of 5 of the monomers from the NADH-bound structure showing the flexibility of loop 235-245 near where cortisol is predicted to bind. The 3 monomers not shown look similar to those shown. (D) Comparison of 20β-HSDH (green) to 3α , 20β-HSDH from *Streptomyces hydrogenans (pink)* The Gly-rich region is highlighted in brighter colors near the adenine of NADH and the side chains of the residues making up the catalytic tetrad are shown. Ser181 of the catalytic tetrad has been mutated to an alanine in 20β-HSDH. (E) Superposition of apo (yellow) and S181A mutant holo structure with NADH bound (green).

	1 10	20	η1 222 Τ 3	η2 <u>τ</u> <u>εεεεε</u>	10 - β1	50	x1 . <u>0000000</u> 60	β2 70 70 η3 οοοο
1 adol 2 desm 3 3GDF 4 3CTM 5 4J2H 6 400L 7 5HA5 8 3ENN 9 1HXH 10 2HSD	MAVESSQIPEKTVEQI MG.EIESYCNP GSHMSSE	EFDERYPLD EFYAVYPKD M.PG KELGPL.PT SDGPF.PK	KWKDS AWKDP QQATK KAPTL ATPQL	NYSILDKF SYRLMDKF HESLLDQL SKNVLDLF SMMTSPIFI PNSVFDMF MEI GPGSMFI NI	MRGRKGFVTG LKGKKGFVTG LKGKVVVVTG LKGKVASVTG LSGRALVTG LSGRALVTG LSGRALVTG LTGRKALVTG LGRKALVTG LGRKALVTG LGGKTVIITG	AAGGIGRN. GGGGIGRN ASGGRCMG SSGGIGWA ASRGIGQS GGGGIGFA AGRGICAA ATGGICAA GASGVCLE GARGICAE	AAALAQAG TAAAWAEAG EAARGCAEMG VAEAYAQAG IAVALAEAG IASSGAAEQG IARALHAQG VVKLLIGEG AARQAVAAG	ADVALVDLPSQE ADVALVDIPASK AAVAITYA.SRA ADVAIWYN.SH. AVAVTAR.TVE GDVALLYR.SAP ARVILVDIDGTA AIVGLHGTREEK AKVAFSDINEAA ARVVLADVLDEE
	α2 22222222222222222	β3	عع	α3 000000000		β4	٥٥	000000000000000000000000000000000000000
1 adol 2 desm 3 3GDF 4 3CTM 5 4J2H 6 400L 7 5HA5 8 3ENN 9 1HXH 10 2HSD	80 90 DKLTELAKDMSERFGT DRLEPLAKEMSERFGT QGAEENVKELEKTYGT .PADEKAEHLQKTYG .GLAET.RALIEKTGF .NMEERSAELAKRFGT AKAQADALTAKG GQQLAAELGF GQATARELGT	I O O NVI AL TCD K V VPL YCC K A KAYKCQ VH S KAYKCQ VH S KAYKCQ VK V K S Y Q C E VAEGHALD R I F VF PAN R SMF V R HD AA R Y Q H LD	VTD TV VSD EE VDS YE ISD PK VRD VE VTD RD VTD RD LSD RE VSS EA VTI EE	110 QVAELKTQ QVAELKTQ QVAALKDV SCEKLVKD SVEETISQ ACASVTRA SVKQAIEA AVAALADD AVAALADD AVAALADD DWTLVMAA DWQRVVAY	120 VADFCQIDAF VADFCQIDAF EKDFGTIDVF AEGLGGLDV LSRFGGLDVI LSRFGGLDVI LSRFGGVDIL QRRLGTLNVL REEFGSVDGL	L 3 0 F L NAG VC LMI H I NAG VC LMI I ANAG VC LMI VANAG VT WT VNNAG FENV' VNNAG VAGR VNNAG VAGR VNNAG ILLP VNNAG IST G	140 GDDVDVY SG.ILDGSV QGPEIDVDNY RP.SFDVDE S.SINPDYPL AA.FDQPEAV GL.FVRMSD GD.METGRL MF.LETESV	150 EVWITRTINININ EAWNHVVVQVDIDIN DSWNKIISVDIN ALWDTIVSTNIK EAWHKTQSVNIH EOWDRVIGVNLE EDWDRVIGVNLT EDFSRLLKINTE ERFRKVVEINT
	α4 <u>000000000000000000000000000000000000</u>	β.	80	TT 190	2000	α5 200000000 210	η ⁴ 2202222000. 220	β6
1 adol 2 desm 3 3GDF 4 3CTM 5 4J2H 6 400L 7 5HA5 8 3ENN 9 1HXH 10 2HSD	α4 000000000000000000000000000000000000	β IGHG GSLIF IGHG GSLIF IGHG GSLVI IG.K GSSI IG.K GSLVI IG.K GSSI GSS	80 TASLS TSSLS TSSLS TSSLS TSSLS TSSLS TSSLS TSSLS TSSLS LCSVA ITSIV MASVA ISSAA	TT 190 GHNANYMMG GYNANFIG GKIVNI. SYVGI. SYVGI. GFVSGG. GFVSGG. GVTGNP. SWLPIE. GLMGLA.	2000 200 GSPTPVNAMGA GGPSPVCAMGT .PQLQAPMNT .PT.AVPMGA STAGMVS STAGMVS GQANMCA QYAGMSA LISSMGA	US 210 210 TKAAIMEHS AKAGISOLA AKAGCIHMA SKSGLLGVT SKGALGVT SKGALSIT SKAAVVHES SKGAIRSIT SKAAVSALT SKAAVSALT SKAAVSALT	n4 220 220 RYLAAALAK. RYMGAGLAP. RSLANEWRD. KSLAIEWAA. RSLARDWRN. QVMARDLAP. KSLAQEIAS. RAAALSCRKQ KSLAQEIAS.	β6 230 230 230 230 230 230 230 230
1 adol 2 desm 3 3GDF 4 3CTM 5 4J2H 6 400L 7 5HA5 8 3ENN 9 1HXH 10 2HSD	α4 000000000000000000000000000000000000	β IGHGGSLIF IGHGGSLIF IGHGGSLII IG.KGSLII IG.KGSLII IG.KGSLII IG.KGSLII IG.KGSLII IG.KGSLII IG.KGSLII IG.GGSIVN IG.GGSIN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GGNN IG.GNN IG	80 TASLS TSSLS TSSLS TSSLS TSSLS TSSLS TTSIV MASVS TTSIV MASVS TT	TT 190 GHNANYMM GYNANFIG GKIVNI. SYVGI. SYVGI. GFVSGG. GFVSGG. GVTGNP. SWLPIE. GLMGLA. 2000 260	2000 200 SSPTPVNA¥GA SGPSPVQYGT .PQLQAP¥NT .PT.AVP¥GA STAG¥VV GQAN¥CA QYAG¥SA LTSS¥GA LTSS¥GA LTSS¥GA LTSS¥GA	US 210 210 TKAAIMEHS AKAGISOLA AKAGCIHMA AKAACIHLA SKGALLGVI SKGALLGVI SKGALRSLI SKGALGS KAAVSALI SKGALGS KAAVSALI X 1000 280	n4 220 220 RYLAAALAK. RYMGAGLAP. RSLANEWRD. KSLAIEWAA. RSLANEWRN. OVMARDLAP. KSLAQEIAS. RAAALSCRKQC KSLAQEIAS. RAAALSCRKQC XIAAVELGT.	β6 230 DGIRSNTISPG YGIRVNTISPG F ARVNTISPG F ARVNTISPG HNIRVNAIAPG F ARVNTISPG HGIRVNAVAPG RNVTVNCIAPG GYAIRVNSIHPD DRIRVNSVHPG

Figure S3. Sequence alignment of 20 β -HSDH found in *B. adolescentis* and other structurally similar SDR members. Red highlights indicate identical residues, green highlights and stars indicate active site residues, blue boxes indicate conserved residues. Secondary structural elements of apo-20 β -HSDH are displayed on the top of the alignment. The abbreviations of protein names are as follows: adol, 20 β -HSDH from *Bifidobacterium adolescentis* strain L2-32; desm, 20 β -HSDH from *Butyricicoccus desmolans* ATCC 43058; remaining labels correspond to PDB IDs. The sequences were aligned with Clustal Omega (1.2.4) and secondary structure rendered by ESPript 3.0 web server.

Adol WP_003810233.1 AVE---SSQI--PEKTVEQI-FDERYPLDKWKDSNYSILDKFSMRGR-KGFVTGAAGGLG 54 A0A1A9H573 HELPX N-----WREKESOK-VAVITGASSGI U4WVT8 HELPX -----MGVGEKEEKKESQK-VAVI<mark>TG</mark>ASSGI<mark>G</mark> 26 -----VGEK---EEKK<mark>E</mark>SQK-VAII<mark>TG</mark>ASSGIC A0A083YF40 HELPX 26 -----VGEK---EEKK<mark>E</mark>SQK-VAVI<mark>TC</mark>ASSGI T2SNM1 HELPX 26 -----VGEK---EEKKESQK-VAII<mark>TG</mark>ASSGI A0A0B2EPQ3 HELPX 26 A0A0U1A5J2 9MYCO GHWLPFSHPOVLAAATTELIDAVSGNOPGRGLRRAEMGKSRRPFEDO-LVVI<mark>TG</mark>GGSGI<mark>G</mark> 235 -----GRAR-SVVI<mark>TG</mark>ASRGL<mark>G</mark>34 A0A0N8HB56 9ACTN -----MAEPRSGDERSCGKR-TVVI<mark>TG</mark>ASRGL<mark>G</mark> 27 A0A101BC04 9MYCO A0A2D6MTV7_9DELT SGM-----TRRSRSP-----ARISEEDKMKGALGYEGK-TVVI<mark>TG</mark>AASGMC 51 A0A2E5YH49 9DELT PGM-----KRKRRSS----TRISEGYKMKNALGYEGK-TVVLTGAASGM 51 A0A1X1T1G8_9MYCO -----MTGIDGLWRHLGYHCR-RVVV<mark>TG</mark>CASGI<mark>G</mark> 28 -----MRI-LGHGYPGIDLKGA-RVLI<mark>TG</mark>AGRGI<mark>G</mark>28 A0A379BZG6 9NOCA -----MNL-FSSRDHLARLDGA-LVVV<mark>TG</mark>GARGI<mark>G</mark> 28 A0A3P8L1B1 TSUPA -----M-ADSTTIGVRVRDK-VIVI<mark>TG</mark>GARGI A0A2Z5YFI0 MYCMR 26 A-R---GWTV--P----PK-----QSTTFNSGDLEIKMMDIKGK-TVVI<mark>TG</mark>ASRGI A4ETZ5 9RHOB 55 -----MNFEGMKMSNMQGK-VVFI<mark>TG</mark>ASRGI A3JLM7 9RHOB 26 -----P----P-----RAGLIDDEGNYTMDMTGK-TVMITGASRGIG 40 A3STW2 SULSN -----MTRNAHHLRQAMTR-TILI<mark>TG</mark>TSSGF<mark>G</mark>26 A0A0A3XQW2 BRAJP A0A103KGN7 9BURK -----MRLRLSNLACWKTLDMSK-TILI<mark>TG</mark>ASSGF<mark>G</mark> 30 A0A174GG33 9BACT GHRER-GDKA--VTAPCEPAEVQGGTGE---RHGRRLERGAVAPGSA-WALV<mark>TG</mark>AGSGI 113 U4E9H0 9VIBR -----MKTSTDKTEVNIMK-TAFI<mark>TG</mark>ATSGI 26 U4KB55_9VIBR -----MKTSTDKTEVNIMK-TAFI<mark>TG</mark>ATSGI<mark>G</mark> 26 -----MQYVYCDNMVKLANK-VVLI<mark>TG</mark>ASSGL<mark>G</mark> 27 A0A1E3LA45 9BACL A0A3S4VN67 MYCAU ----MTKWTAADVPDOSGR-VAIVTGANTGIC 27 -----MRWTAADLPS<mark>FAGR-TVVV<mark>TC</mark>ANSGL<mark>G</mark>26</mark> A0A2X1S640 MYCXE -----MAWKPSEIPDQSGR-TVVI<mark>TG</mark>ANGGI A0A378YLY9_9NOCA 26 A0A174QRR8 BACVU GIISKIKSKL--SYKEVTPL-YMDDLL---RDAYQTTSVVGGS<mark>LKGR-IALV<mark>TG</mark>ATSGI</mark> 57 D6D469 9BACE SIKKYLKRAF--V----F-LLHGIPERHVIANITKLAPNEMLKGR-TALI<mark>TG</mark>ATSGI<mark>G</mark>51 053547 MYCTU -----TINTTDLSGK-VAVVTGAAAGLC 33 ----TTNTTDLSGK-VAVV<mark>TG</mark>AAAGL<mark>G</mark> 33 A0A1R3Y4F5 MYCBO Q0S7K5 RHOJR -----MNAV-ADRDVNVDGK-VAVV<mark>TG</mark>AGSGL<mark>G</mark>26 -----MFMNAV-ADRDVNVDGK-VAVV<mark>TG</mark>AGSGL J1RDG0_9NOCA 28 -----RRTAPSHAGR-CVLV<mark>TG</mark>GASGL<mark>G</mark> A0A376F856 9MICO 26 A0A0M3C339 9SPHI -----IIDYNRTGMDISKLFDLSGK-TAII<mark>TG</mark>GAAGI<mark>G</mark>36 3BHDP RUMGV -----MNFGGFIMGR<mark>F</mark>DEK-IMLV<mark>TG</mark>ATSGI<mark>G</mark>26 -----MHVNGTHDPAQLPLAGR-TALV<mark>TG</mark>GGRGL<mark>G</mark> 29 A2WJD3 9BURK -----FGGSRRTRYA-DAVV<mark>TG</mark>AGSGI A0A378YJ86 9NOCA 27 A0A3B8M323 9ACTN -----GDMSNRLNGL-TAIISGGARGO

Extended N-terminus

UniProt ID

Figure S4. SDR family subset with extended N-terminus multiple sequence alignment. A Clustal Omega (1.2.4) alignment was performed with 555 sequences labeled as SDR family, hydroxysteroid dehydrogenase, and of bacterial origin from the UniProt database. This excerpt from the total sequence alignment includes 39 sequences with a ≥ 10 amino acid residue extended N-terminus. Conserved residues are highlighted, bolded residues indicate the *B. adolescentis* 20 β -HSDH extended N-terminus, and the line indicates the beginning of the *B. adolescentis* 38-residue extended N-terminus.



Figure S5. Gel filtration chromatography of 20β-HSDH truncation mutants. Native molecular weight estimates were based on elution time of 10 mg/ml WT and truncated 20β-HSDH from a Superose 6 10/300 GL analytical column. Fractions were collected and analyzed on SDS-PAGE for visualization. (A) WT, (B) 17-truncated, (C) 21-truncated, (D) 38-truncated. Numbers indicate elution time of standard proteins (1) thyroglobulin, (2) γ -globulin, (3) ovalbumin, (4) myoglobin, (5) vitamin B₁₂.



Figure S6. Model of 20β-HSDH N-terminal truncation effects. Oligomerization is shown by the number of blue circles, or subunits, and enzymatic activity is depicted by the color of the explosion shape. WT 20β-HSDH has 100% relative activity and is likely tetrameric. When the first 17 residues are truncated, the protein exhibits 62% activity, but remains a tetramer. When 21 residues are truncated, the protein loses quaternary structure and has only 2% activity. The full 38-residue truncation results in no activity and a mixed dimeric and monomeric form.