Bacteria on steroids: the enzymatic mechanism of an NADH-dependent dehydrogenase that regulates the conversion of cortisol to androgen in the gut microbiome

Rafael C. Bernardi^{1,‡}, Heidi L. Doden^{2,3,‡}, Marcelo C.R. Melo^{1,4}, Saravanan Devendran^{2,3}, Rebecca M. Pollet⁵, Sean M. Mythen^{2,3}, Shiva Bhowmik⁶, Scott A. Lesley⁶, Isaac Cann^{2,3,7,8}, Zaida Luthey-Schulten^{1,9}, Nicole M. Koropatkin^{3,5}, Jason M. Ridlon^{2,3,7,8,10,*}

1. Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL;

2. Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL;

3. Microbial Metabolic Engineering Theme, Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL;

4. Present Address: Departments of Psychiatry and Microbiology, Perelman School of Medicine; Department of Bioengineering, University of Pennsylvania, Philadelphia, PA;

5. Department of Microbiology & Immunology, University of Michigan at Ann Arbor, Ann Arbor, MI;

6. Department of Integrative and Computational Biology, Scripps Research Institute, San Diego, CA;

7. Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL;

8. Department of Microbiology, University of Illinois at Urbana-Champaign, IL, USA; Cancer Center of Illinois, Urbana IL;

9. Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL;

10. Department of Microbiology & Immunology, Virginia Commonwealth University School of Medicine, Richmond, VA;

KEYWORDS: gut microbiome, steroid metabolism, dehydrogenase, X-ray crystallography, QM/MM, molecular dynamics, free energy, string method

ABSTRACT:

The ability to metabolize both endogenous and exogenous compounds to a variety of metabolic products is not exclusive to our human cells. In fact, the bacterial communities that inhabit our digestive system are responsible for a network of steroid transformations that can produce hormones in the gut, which are then absorbed to act in the host. These communities have been shown to impact our health in numerous ways, affecting disease predisposition, pathogenesis, physical fitness, and dietary responsiveness. Steroid biotransformations by gut bacteria are predicted to impact the host endocrine system. A particular set of transformations facilitated by microbial enzymes has been shown to result in the formation of 11-oxy-androgens from host-derived cortisol. Since androgens have been implicated in disease and immune modulations, understanding the structure and catalytic mechanism of enzymes involved in cortisol metabolism is a key step to hasten the development of strategies that reduce the formation of disease-promoting bioactive steroids in certain individuals. Here, we combine experimental and computational techniques to describe DesC, an enzyme capable of creating 20 α -dihydrocortisol and siphoning cortisol away from pathways that produce androgens. DesC diverges significantly from previously described bacterial and eukaryotic counterparts, catalyzing an NADH-dependent 20 α -hydroxysteroid dehydrogenase reaction but presenting little sequence and structure similarity to them. The structural information obtained by X-ray crystallography and hybrid QM/MM simulations, validated through mutagenesis studies, show the reaction occurs through a multi-step proton relay mechanism. Free energy calculations were then used to describe the kinetics of the reaction mechanism. The mechanistic information presented here can be employed in the development of therapeutics to divert microbial pathways away from disease-promoting steroids.

1. INTRODUCTION:

Bacterial communities have a wide influence on human health and play important roles in disease predisposition, pathogenesis, immune regulation, and dietary responsiveness(1–5). Gut microbes are estimated to be numerically equal to mammalian host cells, although their gene content is estimated to dwarf our own with 99% of functional genes in the human body being microbial. It is no surprise then that the gut microbial consortium has evolved biochemical pathways to biotransform, synthesize, or regulate the production of host endogenous signaling molecules including catecholamines(6), γ -aminobutyric acid (GABA)(7), serotonin(8), and bile acids(9, 10). The contribution of the vertebrate microbiome to behavior(11), blood pressure(12), immune system(13), appetite(14), and even mating behavior are only starting to be addressed(15). In addition, the contribution of host-associated bacterial metabolism of steroid hormones is coming into focus. Steroid hormones are essential for regulation of various physiological functions, including cellular communication, metabolism, inflammatory response, and stress response(16). One class in particular, the endogenous glucocorticoids, is synthesized from cholesterol in the adrenal cortex(17) and released into the blood to influence cells throughout the body. Upon binding to host nuclear receptors they affect cellular processes through genomic and non-genomic mechanisms(16). Thus, investigating the chemistry of steroid biotransformation in the gut microbiome is crucial for a better understanding of the symbiotic relationship between us and our microbiota(18).

Cortisol, the major glucocorticoid in humans, is converted to various derivatives by enzymes throughout our body. The local concentrations of cortisol are regulated by enzymes that reversibly modify functional groups on the steroid nucleus and side chain, saturate the steroid rings, and conjugate the steroid with sulfate or glucuronide(19–21). In this way, ratios of active

vs. inactive glucocorticoids can be locally and rapidly altered in peripheral tissues. Intriguingly, gut microbes possess multiple cortisol-altering enzymes that have co-evolved separately from eukaryotic counterparts(21–26). Microbes are therefore expected to play an important role in the complex steroid metabolome of the host. Indeed, accumulating evidence demonstrates human gut microbes are an important component of the host endocrine system(20). Of particular relevance is the observation that peripheral side-chain cleavage of cortisol to 11 β -hydroxyandrostenedione (11 β -OHAD) occurs by some combination of host and microbial input (27–29). The androgens produced from cortisol side-chain cleavage function similarly to testosterone by binding and activating androgen receptor (30). Recently, these novel androgens have been shown to be relevant to human physiology and pathology, including castration-resistant prostate cancer (29, 31), polycystic ovary syndrome (30, 32), and potentially hypertension (22, 33–36).

The human gut bacterium *Clostridium scindens* ATCC 35704 was previously shown to side-chain cleave cortisol(37). However, the lack of knowledge about the metabolic pathway responsible for androgen production severely limits our ability to mechanistically study androgen formation by the human microbiome *in vivo*. We previously discovered a cortisol-inducible gene cluster (*desABCD*) in *C. scindens* ATCC 35704, encoding NADH-dependent 20 α -hydroxysteroid dehydrogenase (20 α -HSDH, DesC)(38) and heterotetrameric steroid-17,20-desmolase (DesAB)(39) which produce 20 α -dihydrocortisol and 11 β -OHAD from cortisol, respectively. Furthermore, the 20 α -dihydrocortisol DesC product is not a substrate for DesAB(39), suggesting DesC acts as a metabolic “switch” regulating side-chain cleavage of cortisol. Importantly, pharmaceutical glucocorticoids such as dexamethasone and prednisone are substrates for side-chain cleavage, generating metabolites that drive prostate cancer cell proliferation(40, 41). Therefore, understanding the structure and catalytic mechanism of enzymes involved in cortisol

metabolism is key to hasten the development of strategies to reduce drug metabolism and alter androgen formation by the human microbiome.

Figure 1. Overview of the metabolism of cortisol by the gut bacterium *Clostridium scindens*. The obligate anaerobic bacterium, *C. scindens* ATCC 35704, resides in the large intestine. This bacterium expresses a gene cluster (*desABCD*) encoding DesAB (no structure available) which removes the side-chain of cortisol, forming 11β-hydroxyandrostenedione. DesC (new structure presented in this manuscript) is an NAD(H)-dependent 20α-hydroxysteroid dehydrogenase that catalyzes the reversible oxidoreduction of the C20 ketone. The androgenic end product of DesAB is implicated in prostate cancer, and may affect innate and adaptive immune function because many immune cells express androgen receptor. By inhibiting the host enzyme 11β-hydroxysteroid dehydrogenase 1 & 2 isoforms, these metabolites may also affect blood pressure and inflammation in the GI tract. The formation of 20α-dihydrocortisol in the host is associated with disorders such as Cushing's syndrome; however, little is known about the physiological or pathophysiological role of 20-reduced forms of cortisol. Moreover, DesC appears to regulate substrate availability to DesAB.

Here, we combine multiple experimental and computational approaches to investigate the enzymatic mechanism of *C. scindens* ATCC 35704 DesC, a potential regulator of the pro-androgenic DesAB pathway (**Fig. 1**). Our findings reveal the structure of this enzyme with

atomistic detail as well as its interaction with both NADH and cortisol substrates. Classical and hybrid QM/MM molecular dynamics simulations elucidate the intricate reaction mechanism including a hydride transfer from NADH to cortisol and a proton relay that finally leads to a 20 α -dihydrocortisol enzymatic product. Molecular dynamics simulations are a powerful tool to investigate the nanoscopic world, particularly when in combination with experimental methods. Simulation can be used to investigate structural properties from folding(42) to chemical reactions(43).

2. RESULTS AND DISCUSSION

2.1 X-ray crystallography. To understand the structure and mechanism of DesC, the apo-form of the recombinant enzyme (rDesC) was crystallized via vapor diffusion sitting drop at pH 6.75 in a buffer containing 0.2M MgCl₂, 24% PEG 400, 0.1M HEPES, resulting in a resolution of 2.0Å. X-ray diffraction data can be found in **Table S1**. DesC is a member of the zinc-dependent medium chain dehydrogenase family that catalyzes zinc-dependent oxidation of primary and secondary alcohols using NAD(P)⁺ as the proton and hydride acceptor(44). DesC displays the typical bilobal topology comprised of a nucleotide-binding Rossmann fold domain and a catalytic domain(45, 46) (**Fig. 2A**). Across the alcohol dehydrogenase superfamily, the size and architecture of the catalytic domain varies greatly, with the medium chain dehydrogenases having a larger and more complex catalytic domain(44). While this family of enzymes varies in sequence conservation, the Rossmann fold is the most conserved feature among short-, medium-, and long-chain dehydrogenases(44). In the two-domain structure of DesC, the Rossmann fold is located proximal to the C-terminus, and comprised of β -strands 11-16 and flanked by α -helices 6-11 (**Fig. 2B**). The predicted catalytic Zn²⁺

ion (**Fig. 2C**) is octahedrally coordinated by two water oxygens and residues of the catalytic domain including the side chains of His73, Glu74, Glu156, and a cysteinesulfonic acid at Cys45.

Like many prokaryotic alcohol dehydrogenases, DesC is a tetramer (**Fig. 2A**). The enzyme crystallized with one monomer in the asymmetric unit, and the tetrameric assembly is formed by crystallographic symmetry; therefore no differences are observed among the monomers that comprise the biological assembly of DesC in this model.

The tetramer displays the catalytic domains at the periphery of the complex. Two Rossmann folds align along the first two-fold axis to create a continuous β -sheet comprised of 6 β -strands from each monomer across the dimeric interface. The second two-fold axis situates two adjacent catalytic domains such that a helical loop created by $\alpha 4$ and a 3_{10} helix and stabilized by a second Zn^{2+} (Zn finger) extends towards the adjacent monomer (**Fig. 2D**). Referred to as the lobe loop, this structural feature is conserved among many alcohol dehydrogenase including those from mammals(47–49). The Zn^{2+} at this position is located $\sim 21\text{\AA}$ from the catalytic cleft and is tetrahedrally coordinated by 4 Cys (105/108/111/119). Single alanine substitutions of Cys105, Cys108, Cys111, Cys115 residues composing the zinc-finger were sufficient to result in loss of quaternary structure (**Table S2**), confirming the structural role of the zinc-finger in dimer-dimer interaction.

The medium-chain dehydrogenases/reductases (MDRs) that are close structural and sequence homologs of DesC include the threonine dehydrogenases (TDH) from *Pyrococcus horikoshii* (PhTDH; PDB 2DFV, $2.0\text{\AA}$ RMSD for 333 Ca, $Z=40.5$) and *Thermococcus kodakaraensis* (TkTDH; PDB 3GFB, $2.3\text{\AA}$ RMSD for 334 Ca, $Z=39.4$), see **Fig. S1**. The mechanism for the TDHs and many MDRs is believed to proceed via an ordered bi-bi reaction whereby NAD^+ binds first and is the last to leave(47, 48, 50). Binding of NAD^+ likely initiates the conformational change

that in part closes the active site to engage the catalytic machinery. The crystal structures of PhTDH and TkTDH were determined with NAD⁺(47, 48). An overlay of DesC and TDH structures (**Fig. S1**) demonstrates conservation of the structural and catalytic Zn²⁺-binding residues (**Fig. S2**) as well as the NADH binding pocket. Not shared between these two is a hydrophobic patch of four tryptophan residues that is positioned near the lobe loop and conserved among the thermophilic threonine dehydrogenases(47, 48).

Figure 2. Depiction of DesC highlighting functional features. **A.** DesC structure with amino acid chains shown in cartoon representation and colored differently to show the bilobal topology. Structural waters are shown in red spheres, and two catalytic Zn²⁺ ions are shown as gray spheres. **B.** Same view of the complex colored by secondary structures: α -helices in purple and β -strands in green. **C.** Detailed view of the catalytic Zn²⁺ ion in octahedral coordination by water molecules and catalytic site residues. Amino acid residues are shown in licorice representation where carbon atoms are colored in gray, oxygen in red, nitrogen in blue and phosphorous in yellow. **D.** Detailed view of the helical loop that extends towards the adjacent monomer.

Conserved features among the NADH binding pocket include Asp204 in DesC which is Glu199 in PhTDH and hydrogens bonds with the O2 and O3 of the adenosine ribose. This residue is shared among bacterial alcohol dehydrogenases and imparts specificity for the co-factor(47, 51).

Also of importance is Arg204 of PhTDH, which extends towards the pyrophosphate and is equivalent with K209 in DesC, though in our model this residue is positioned somewhat further from the predicted location of the pyrophosphates. Mutagenesis of Glu199 and Arg204 in PhTDH demonstrates these residues are significant determinants in NAD(H) binding. The nucleotide-binding GxGxxG motif is also conserved among these enzymes although Pro178 of PhTDH is replaced by Gly183 in DesC. Finally, in PhTDH, Thr44 and His50 are believed to be part of the proton relay that connects the active site with bulk solvent, and these are conserved in DesC as Ser47 and His50. However, in the current crystal structure of DesC Ser47 and His50 are not aligned to interact. While the structures of DesC, PhTDH and TkTDH share a very similar fold and conservation of important catalytic residues, our overlay reveals a longer extension of α -helix (residues 58-66) and the loop connecting β 1/2 (residues 9-19) in DesC. This region is near the entrance to the active site cleft, but $\sim 20\text{\AA}$ from the predicted site of hydride transfer to the nicotinamide ring. It is possible that this loop plays some role in selecting or restricting substrate. The structure of DesC was deposited to the protein data bank(52) (PDB) with accession code 4OH1.

2.2 Docking. Since structural data was obtained only for the apo-DesC, we used homologous structures available within the PDB in order to fit an NADH molecule to the binding pocket of a monomer of DesC, employing a previously established protocol(53). With BLAST(54), we obtained homologous structures (PDB IDs: 4ILK, 4EJ6, 4A2C, 3QE3, 3GFB, 2DQ4, 2DFV, 2D8A, 1PL7, 1E3J) from the PDB. The alignment (**Fig. 3A,B**) and placing of the NADH molecule on DesC binding site was performed using VMD(55). Leveraging advanced run options of QwikMD(56), the structure of the ligand was minimized in the pocket together with atoms of nearby DesC residues, while maintaining the structure of most of DesC static.

Figure 3. Characterization of the cortisol binding pocket of DesC. **A.** Homologous proteins from the PDB were superimposed to highlight structure similarity and binding pose of multiple ligands. All proteins are shown in cartoon representation while ligands are shown in orange licorice representation. **B.** Same as previous, with alternate view of the binding site. **C.** Final binding poses of NADH and cortisol to DesC. The protein is shown in cartoon representation. **D.** Space-filling representation of DesC shows the alignment of cortisol deep in a cleft. **E.** Detailed view of cortisol and NADH binding near the Zn^{2+} in the active site. Substrates are shown in ball-and-stick representations. Carbon atoms in gray, oxygen in red, nitrogen in blue and hydrogen in white. **F.** Detailed view of tyrosines 101 and 124 stabilizing cortisol's binding pose. **G.** Comparison of circular dichroism spectra (190-260 nm) between wild type (WT) recombinant DesC, Y101R mutant, and Y124R mutant. **H.** Isothermal titration calorimetry of Y101R mutant in the presence of 1.5 mM NADH (upper left panel), and 1.5 mM NADH + 0.5 mM cortisol (upper right panel) and Y124R mutant in the presence of 1.5 mM NADH (lower left panel), and 1.5 mM NADH + 0.5 mM cortisol (lower right panel). **I.** Effect of binding pocket mutations on K_m for cortisol in the presence of saturating NADH. **J.** Relative activity (%) of WT vs. binding pocket mutants when cortisol and NADH are saturating.

The structure of cortisol was also fitted to the most probable binding site using VMD (Fig. 3C).

The docking was performed by a combined manual and computational approach(57), where VMD

was used to position some of the reference atoms of the cortisol, while NAMD(58), through its QwikMD interface(56), was used to minimize the structure of the complex. As reference atoms we chose the cortisol atoms that were known to participate in the enzymatic reaction, positioning them near the Zn^{2+} and the NADH hydride donor (**Fig. 3E**). Also, cortisol was aligned in a cleft (**Fig. 3D**), which was shown to contain a more hydrophobic character, typical of ligand binding sites. The complex was then solvated, and the net charge of the system was neutralized in a 0.15 mol/l sodium chloride solution, using QwikMD to follow predetermined protocols employed in previous enzymatic studies(53). The solvated system was subjected to 100 ns of equilibrium MD simulation, using the CHARMM36 force field(59), along with the TIP3 water model(60). The simulations were performed with periodic boundary conditions, in the NpT ensemble, with temperature maintained at 300 K and pressure at 1 bar. Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method(61). All MD simulations were executed with the GPU-accelerated NAMD molecular dynamics package(58).

An MD simulation protocol can be used to investigate the stability of a molecular system, as well as predict the behavior and function of proteins or protein domains(62). Here, after 100 ns of MD simulation, we observed that cortisol was stable in the predicted catalytic cleft. After analyzing the MD trajectory, we observed that the cortisol molecule was stabilized by interactions with two tyrosine residues, namely Tyr101, and Tyr124 (**Fig. 3F**). Simulations revealed that the mostly hydrophobic cleft was fundamental in maintaining the complex structure stability, with cortisol fluctuating very little in the cleft. Physical chemical properties of the interactions were calculated using VMD. Particularly, in a previously described protocol(63–65), we use radial distribution function to observe the how stable was the molecule in the binding pocket. To validate the importance of Tyr101 and Tyr124 in substrate-binding, we performed site-directed

mutagenesis, observing loss of enzyme activity at $96.01 \pm 0.67\%$ and $81.82 \pm 0.31\%$ activity, respectively, without altering secondary structural features as determined by circular dichroism (CD)spectroscopy (**Fig. 3G**). Isothermal titration calorimetry (ITC) experiments showed that there was no effect on NADH binding with rDesCY101R ($K_d = 35.08 \mu\text{M}$) and rDesCY124R ($K_d = 39.06 \mu\text{M}$). However, binding affinity of cortisol to rDesCY101R and rDesCY124R was ablated, based on ITC data (**Fig. 3**). Additionally, mutation of tyrosine to arginine in the putative binding pocket altered activity and kinetic parameters (**Table S3**). The experimental data support the binding of cortisol to the hydrophobic pocket predicted by the computational study. Particularly, the experimental data shows that by mutating the aforementioned residues, the structure did not change, but interactions with cortisol did as evidenced by altered ITC and kinetic constants which were not observed with NADH.

2.3 Reaction Mechanism. After examining the enzyme active site, we sought to determine the catalytic mechanism involved in 20α -dehydrogenation. For that, we employed a combination of hybrid quantum mechanics (QM) and classical molecular mechanics (MM) simulations. We also integrated a state-of-the-art enhanced sampling method, which was recently implemented in NAMD(66). Hybrid QM/MM has been the main tool used to study processes that cannot be explained by classical MM methods alone(43, 67, 68). In biological contexts, QM/MM calculations are often used in enzymology(69, 70) as well as to investigate polarizable molecules in different environments(71, 72). In the QM/MM molecular dynamics (MD) approach, the system of interest is typically investigated in a water bath (**Fig. 4A**), as in a regular classical MD simulation, with the exception that part of the system is treated at the quantum-mechanical level(66). This approach offers an augmented MD simulation, where a selection of atoms is investigated at sub-atomic resolution in order to account for electronic effects. Specifically, we

employed an electrostatic embedded QM/MM approach, where the classical part of the system was treated with the CHARMM36 force field(59), while the QM part of the system was treated at semiempirical PM7 level(73). The QM region was chosen to be relatively large, with 573 atoms, because of the elongated shape of the catalytic pocket, and large presence of aromatic groups that could affect the electronic structure near the reaction site. To ensure the presence of a shell of QM atoms around the reaction site, a total of 36 QM-MM bonds were created using a link-atom approach(74). Our computational studies were then supported by biochemical assays.

To investigate possible reaction mechanisms, we ran a 20 ps long QM/MM MD equilibration to prepare the system, and used a QM/MM steered MD (SMD) approach(66, 75–78) to induce the reaction to occur according to the chosen mechanisms we describe below. These selected reaction mechanisms were designed to determine multiple aspects of this complex reaction, one being the participation of a water molecule to mediate the interaction between the Zn^{2+} ion and substrates, as previously suggested for similar enzymes(79). Our QM/MM SMD only forced the initial steps of the reactions, allowing the remaining steps to occur freely in case they were energetically favorable. The simulations revealed that the most probable scenario would be the reaction mechanism without the participation of a water molecule, having a concerted proton transfer from S47 and a hydride from NADH to the C20 ketone oxygen of cortisol. The QM/MM SMD simulations revealed a proton relay that was responsible for reestablishing the amino acid protonation states in the catalytic site (**Fig. 4B**). The proton relay was composed of the replenishing of S47 by a proton transfer from the ribosyl of NAD^+ , followed by the protonation of the ribosyl group by His50 (which was previously in a protonated state). His50 can once again become protonated when NAD^+ leaves the catalytic pocket for a new enzymatic cycle, a situation in which His50 becomes largely exposed to the bulk solvent.

Figure 4. Description of QM/MM calculation of the reaction catalyzed by DesC. **A.** Depiction of the entire simulated system, showing DesC in a simplified water box. **B.** Detailed view of the active site showing all proton transfers that take place during the reaction. DesC is shown in cartoon representation, while atoms in the QM region are shown in ball-and-stick representations. Carbon atoms in gray, oxygen in red, nitrogen in blue and hydrogen in white. **C.** Plot showing the progression of all collective variables tracked during the string method optimization of the reaction coordinate. Collective variable pairs labeled with shades of the same color describe the distance between a hydrogen atom and its donor and acceptor carbon or oxygen atoms. The numbers correspond to the labelled reactions in section B of the figure. The interatomic distances show the proton relay progression during the catalytic process. **D.** Matching plot showing the same reaction coordinate and the results of the eABF free energy calculation. The first energy barrier corresponds to the proton exchange from NADH to cortisol. **E.** Multiple detailed views of the active site show the progression of the overall reaction. Blue dotted rectangles and arrows indicate which stage of the reaction is depicted.

To test this mechanism, we generated S47A and S47T mutations that resulted in loss of catalytic function (100% vs. 98.52 ± 0.08 % loss of activity, respectively) (**Fig 5A,E**). While S47T mutant bound NADH ($K_d = 66.67 \mu\text{M}$), cortisol lost binding affinity (**Fig 5D**). By contrast, S47A showed high affinity for cortisol based on ITC measurements ($K_d = 11.56 \mu\text{M}$) in the presence of NADH ($K_d = 97.0 \text{ nM}$) (**Fig 5C**). Hydrogen-bonding between the C21-hydroxy group of cortisol and the NADH is predicted to be responsible for the orders of magnitude difference in binding affinity between S47A, which binds cortisol, and S47T, which does not. There are no statistically significant differences in CD spectra between DesCWT and S47A and S47T (**Fig 5B**).

Figure 5. Role of Serine 47 in catalysis. **A.** Depiction of active site with bound reactants. The two residues highlighted were mutated to confirm their importance to binding and enzymatic activity. **B.** Comparison of circular dichroism spectra (190-260 nm) between wild type (WT) recombinant DesC, S47A mutant, S47T mutant, and H50A mutant. **C.** Isothermal titration calorimetry of S47A mutant in the presence of 1.5 mM NADH (left panel), and 1.5 mM NADH + 0.5 mM cortisol (right panel). **D.** Effect of active-site mutations on K_m for cortisol in the presence of saturating NADH. **E.** Relative activity (%) of WT vs. active-site mutants when cortisol and NADH are saturating.

With the experimental validation of the possible mechanism, we moved to optimize the reaction pathway and understand the atomistic details of the enzymatic mechanism. The string method(80) was chosen for the task. This method is an optimization algorithm that relies on iterations of biased QM/MM MD simulations to find the chemical reaction path with the smallest energetic barrier. It does so by tracking “collective variables”, a set of measurements made on the system to describe the transformation being examined. In our case, the collective variables were chosen to be distances between relevant atoms in the QM region of the simulation.

The string method was initialized with the QM/MM SMD simulation as a first guess of the reaction path. A set of representative snapshots of the chemical reaction was selected to describe the entire process, from initial to final state, creating a set of “images” that represent a copy of the entire simulated system at a different stage of the reaction. At each iteration, multiple independent MD simulations are initiated from each image, allowing the atomic systems to explore the energy surface and drift to local minima. Next, average values for the collective variables are determined. Here, these values are cartesian coordinates, that are then used to constrain the systems to keep consecutive images approximately equidistant in the collective variables space, smoothing the reaction coordinate. Convergence is achieved when the iterations do not produce significant changes in the mean values of the collective variables. Taking advantage of NCSA’s Blue Waters supercomputer, we ran 750 replicas for our string optimization, revealing the pathway of the reaction (**Fig 4C**). The string optimization shows that the initial reaction that triggers the enzymatic process is the hydride donation to the cortisol side-chain C20, followed by the S47 proton donation to cortisol. The formation of 20 α -dihydrocortisol triggers the proton relay we described above.

2.4 Free Energy calculation. To investigate the free energy barrier that limits the start of the reaction, namely the activation free energy (ΔG_{act}), we perform a parallel extended-Adaptive Biasing Force (eABF) calculation(81), following established protocols(66, 82). Using the final images of the optimized string we can define a continuous path that represents the entire enzymatic reaction. The eABF method can then be used to calculate the free energy changes along this path by constraining the system to path collective variables S and Z, where S indicates progression along the reaction path, and Z indicates a perpendicular distance to the path. Using NCSA's Blue Waters supercomputer, we used a parallel strategy that initiated 600 walkers from the different images, and were thus able to conduct extensive sampling over the defined path. The multiple-walker method is so-called due to the fact that each replica executes a random walk along the path(83). Here, 10,000 MD steps were performed by each walker, accumulating 6,000,000 QM/MM ABF steps, or 3 ns of QM/MM MD simulation. The simulations revealed that only the initial step of the reaction, namely the hydride transfer from NADH to cortisol C20, imposes an energetic barrier in the order of 6 kcal/mol (**Fig 4D**). The 3 proton transfers that follow the hydride transfer were found to be energetically favorable. The entire reaction process was found to be exergonic, with a free energy difference between reactants and products (ΔG) in the order of -25 kcal/mol. It is worth mentioning that this free energy difference only accounts for the reaction itself, and not for binding and rearrangements of the structure that occur prior and after the catalytic reaction. The ITC data however has also shown that the binding is exothermic, making the initial binding event, as expected, part of the overall favorable energy of the reaction.

The enhanced sampling protocol adopted here greatly accelerated the dynamics of the system, allowing us to balance the gain in precision and the increased cost of QM/MM simulations. The simulations revealed that all reaction steps can occur in quick succession due to a stabilization of

the reaction site, which is partially imposed by the presence of a Zn^{2+} ion. After the initial reaction step, all reactions in the proton relay are highly favorable because all protons are already aligned in hydrogen bonds. This stability is noticeable in the snapshots shown in **Fig 4E**, which reveal very little difference in the catalytic region during the entire reaction process.

3. CONCLUSIONS

Here we reported the crystal structure and proposed a catalytic mechanism for DesC, a corticosteroid NADH-dependent 20α -HSDH expressed by *Clostridium scindens* ATCC 35704. Bacterial metabolism of host and synthetic corticosteroids may have wide ranging effects on the immune system(84–87), the cardiovascular system(20), the structure of the gut microbiome(88, 89), and the urogenital tract(41, 90). Their importance for human health brings enzymes catalyzing 20-hydroxysteroids significant pharmaceutical value.

The expression of DesC by *C. scindens* ATCC 35704 highlights its importance as a member of the human ‘sterolbiome’(39, 91), the collective gene pathways responsible for steroid biotransformations by host-associated bacteria. In addition, DesC is part of a gene cluster encoding steroid-17,20-desmolase encoded by *desA* and *desB* genes(38). Steroid-17,20-desmolase catalyzes side-chain 17,20-lyase activity, which is the first step to converting cortisone to 11-keto-androstenedione, an androgen-receptor ligand(41). Thus, *C. scindens* ATCC 35704 contributes to the sterolbiome through the formation of 11-oxy-androgens and 20-oxy-steroids. Importantly, *C. scindens* is one of a few species of clostridia capable of converting host primary bile acids to disease-causing secondary bile acids such as deoxycholic acid and lithocholic acid. *C. scindens* sterolbiome enzymes such as the rate-limiting bile acid 7α -dehydratase (BaiE)(92) steroid-17,20-desmolase (DesAB)(39), and 20α -HSDH (DesC) emerge as potential therapeutic targets.

In nature, reversible conversion of the C20 oxygen of C21 corticosteroids is catalyzed by diverse pyridine nucleotide-dependent 20 α -hydroxysteroid dehydrogenases. DesC, the NAD(H)-dependent 20 α -HSDH described in the present study, diverges significantly from previously described bacterial counterparts. Strains of *Escherichia coli* (DH5 α , K12, E132) encode a homotetrameric (28 kDa monomer) complex with NADH-dependent 20 α -HSDH activity belonging to the classical short-chain dehydrogenase/reductase (SDR) family, which does not require metal ions(93). The enzyme was previously reported as a 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (KduD), and has broad substrate specificity for both sugars and C₂₁-corticosteroids. Interestingly, KduD and DesC share only 16% amino acid identity. A second example is the dimeric (34 kDa monomer) NADPH-dependent 20 α -HSDH purified from the protozoan *Tetrahymena pyriformis*, which was structurally and functionally distinct from bacterial as well as mammalian 20 α -HSDHs(94). The *T. pyriformis* 20 α -HSDH was highly specific for 17 α -progesterone, which inhibits the growth of this protozoan. Mammalian enzymes reported to have 20 α -HSDH activity are monomeric with *M_r* values between 33 kDa to 55 kDa, except for estradiol 17 α -dehydrogenase (dual 20 α -HSDH activity), which is a dimer. Many mammalian enzymes are polyfunctional, catalyzing several regio- and stereo-specific reactions at more than one steroid keto group (e.g. 3 α , 20 α -HSDH from human liver dihydrodiol dehydrogenase). Thus, the DesC represents a unique and now well-characterized bacterial 20 α -HSDH.

In summary, we have employed a combination of state-of-the-art computational tools(95) and experiments to reveal the structure and proposed catalytic mechanism of DesC, a key steroid-metabolizing enzyme found in the gut microbiome (**Fig. 1**). X-ray crystallography was utilized to obtain the apo-DesC structure (**Fig. 2**), which was then used as the starting point of a computational study that involved the identification of the DesC catalytic pocket, as well as the docking of both

NADH and cortisol to this pocket (**Fig. 3**). Biochemical assays were then employed to confirm the putative pocket. As it was hypothesized for the TDHs and many MDRs, our ITC data shows that NADH is the first substrate to bind the enzyme, given that cortisol does not bind to DesC unless NADH is present at the catalytic pocket. With the structure of DesC and its ligands now available, a hybrid QM/MM MD approach(66) was applied to investigate the reaction mechanism of DesC. The approach, which combined steered molecular dynamics, string method optimization, and extended adaptive biasing force, with hybrid QM/MM calculations, then revealed that NADH first donates a hydride to cortisol, which triggers a proton relay that provides a proton to cortisol, forming the 20 α -dihydrocortisol enzymatic product. The proton relay then stabilizes the reaction pocket by a series of proton transfers that starts in a protonated histidine (**Fig. 4**). The mechanism was confirmed by biochemical assays, where amino acids that were revealed critical in the simulations were mutated (**Fig. 5**). The enzymatic process was found to be exergonic, although the free energy difference calculated between reactants and products only accounts for the reaction itself and not for other steps of the DesC mechanism. Our findings can be applied in the development of approaches to modulate DesC activity, while our protocols can be employed to investigate other enzymes that are biomedically and biotechnologically relevant. The successful combination of experimental and computational approaches presented here reinforces the power and the need of combining these techniques.

4. MATERIALS AND METHODS

4.1 Bacterial Strains and Materials. *Clostridium scindens* ATCC 35704 has been maintained as -80°C glycerol stocks in our laboratory. *Escherichia coli* DH5 α (turbo) competent cells were from New England Biolab (Ipswich, MA), *E. coli* BL21 CodonPlus (DE3)-RIPL was purchased from Stratagene (La Jolla, CA). The pET-51b vector was obtained from Novagen (San Diego, CA). Restriction enzymes were purchased from New England Biolab (Ipswich, MA), QIAprep Spin Miniprep kit was obtained from Qiagen (Valencia, CA). Isopropyl β -D-1-thiogalactopyranoside (IPTG) was purchased from Gold Biotechnology (St. Louis, MO). Strep-Tactin® resin was purchased from IBA GmbH, (Goettingen, Germany). Cortisol and 20 α -dihydrocortisol were purchased from Steraloids (Newport, RI). Amicon Ultra centrifugal filter units with 10-kDa MWCO were obtained from Millipore (Billerica, MA). All other reagents were of the highest possible purity and were purchased from Fisher Scientific (Pittsburgh, PA).

4.2 Gene cloning, expression and protein purification. The open reading frame (EDS07887) encoding 20 α -hydroxysteroid dehydrogenase (20 α -HSDH) amplified from *C. scindens* genomic DNA with primers reported in **Table S4**, using high fidelity Phusion polymerase and cloned into pET51b as previously described. The sequence correctness of DNA insert in the recombinant plasmid was confirmed by DNA sequencing at W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign.

For protein expression, the correct recombinant plasmid extracted from the *E. coli* DH5 α cells was transformed into *E. coli* BL21 CodonPlus (DE3)-RIPL chemically competent cells by the heat shock method and grown overnight at 37 °C on LB agar plates supplemented with ampicillin (100 μ g/mL) and chloramphenicol (50 μ g/mL). After 16 hours, five isolated colonies were used to

inoculate 10 mL of fresh LB medium supplemented with antibiotics and grown at 37 °C for 6 hours with vigorous aeration. The pre-cultures were then added to fresh LB medium (1 L), supplemented with the same antibiotics at the same concentrations, and grown with vigorous aeration at 37 °C. At OD600 of 0.3, isopropyl β -D-thio-galactopyranoside (IPTG) was added to each culture at a final concentration of 0.1 mM and the temperature was decreased to 16 °C. Following 16 hours of culturing, cells were pelleted by centrifugation (4,000 x g, 30 min, 4 °C) and re-suspended in 30 mL of binding buffer (20 mM Tris-HCl, 150 mM NaCl, 20% glycerol, 10mM 2-Mercaptoethanol pH 7.9). The cell suspension was subjected to four passages through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada), and the cell lysate was clarified by centrifugation at 20,000 x g for 30 min at 4°C.

The recombinant 20 α -HSDH was then purified using Strep-Tactin® resin (IBA GmbH, Goettingen, Germany) as per manufacturer's protocol. The recombinant protein was eluted using an elution buffer composed of 20 mM Tris-HCl, 150 mM NaCl, 20% glycerol, 10mM 2-mercaptoethanol pH 7.9 and 2.5 mM d-desthiobiotin. The protein purity was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and protein bands were visualized by staining with Coomassie brilliant blue G-250. The protein concentrations were calculated based on the molecular mass and computed extinction coefficient.

4.3 Protein Crystallization. Selenomethionine derivatized DesC protein from *C. scindens* ATCC 35704 yielded crystals from JCSG Core Suite (Qiagen Sciences, MD) using a drop volume of 0.2 mL and a 1:1 (vol/vol) ratio of reservoir to protein in 20 mM HEPES pH 8.0 solution equilibrated against 100 μ L of reservoir solution. The apo-form of the recombinant DesC crystallized via sitting drop vapor diffusion at 20°C against a crystallization solution containing

0.2M MgCl₂, 24% PEG 400, 0.1M HEPES at pH 6.75. The crystals were cryoprotected with addition of 10% 1,2 ethanediol before harvesting and flash- freezing with liquid nitrogen. X-ray data were collected at BL14-1 at the Stanford Synchrotron Radiation Lightsource. The x-ray data were indexed and integrated with XDS(96) and scaled with XSCALE(97). Multiwavelength anomalous dispersion (MAD) from the selenium signal used to determine the structure using the programs with SHELX(98, 99). Model refinement was performed with Refmac(100, 101).

4.4 Mutational analysis of DesC to identify key residues for its activity and binding to substrate. In order to identify the key amino acid residues that may be involved in both activity and binding to cortisol, an amino acid sequence alignment was carried out with known zinc dependent medium chain dehydrogenase sequences obtained from NCBI. Mutations were made in the pET-51b containing the wild type DesC using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent, Santa Clara) according to the manufacture's protocol. The primers used in the site-directed mutagenesis study are summarized in **Table S4**. Mutated plasmids were transformed into *E. coli* XL10 competent cells by heat-shock and plated onto lysogeny broth (LB) solidified with Bacto-agar (Difco) containing 100 µg/mL ampicillin sodium salt, and the plates were incubated at 37 °C overnight. Individual colonies were cultivated in 5ml of LB medium supplemented with ampicillin (100 µg/mL) overnight at 37 °C, and plasmids were extracted using a QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA). Plasmids containing the expected mutations were confirmed by DNA sequencing (W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign). The mutated proteins were expressed and purified using the same procedure described above for wild type. The proteins were quantified based on their molecular mass and theoretical extinction coefficient.

4.5 Determination of structural integrity of the DesC wild type and mutants by circular dichroism (CD) spectra. Determination of circular dichroism (CD) spectra for both DesC wild type and its site-directed mutant proteins was carried out using a J-815 circular dichroism spectropolarimeter (Jasco, Tokyo, Japan). Protein samples were prepared at a concentration of 0.2 mg/ml in 10 mM KH₂PO₄ buffer (pH 7.5). For the measurements, a quartz cell with a path length of 0.1 cm was used. CD scans were carried out at 25°C from 190 nm to 260 nm at a speed of 50 nm/min with a 0.1-nm wavelength pitch, with five accumulations. Data files were analyzed on the DICHROWEB online server (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) using the CDSSTR algorithm with reference set 4, which is optimized for analysis of data recorded in the range of 190 nm to 240 nm.

4.6 Isothermal titration calorimetric analysis of the DesC. Isothermal titration calorimetric (ITC) analysis was performed using a VP-ITC microcalorimeter with a 1.4 mL cell volume from MicroCal, Inc. The proteins were dialyzed with phosphate buffer (50 mM sodium phosphate, 150 mM NaCl, 10% glycerol, pH 7.5) and the pyridine nucleotide (NADH) and cortisol were dissolved in the same buffer. The proteins (50 µM) with and without pyridine nucleotide were then injected with 28 successive 10 µL aliquots of cortisol (0.5 mM) at 300-s intervals. The data were fitted to a nonlinear regression model using a single binding site (MicroCal Origin software). The thermodynamic parameters were calculated using the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$), and the relationship $\Delta G = -RT\ln(K_a)$.

4.7 DesC enzyme assay. DesC enzyme activity for both wild type and mutants was determined as described previously. Linearity of enzyme activity with respect to time and enzyme concentration was determined aerobically by monitoring the oxidation of NADH at 340 nM ($\epsilon=6,220 \text{ M}^{-1}\text{cm}^{-1}$) in the presence of cortisol. The standard reaction mixture contained 50 mM phosphate buffer, pH 7.5, 50 μM of cortisol, 150 μM NADH, 0.05 μM enzyme, and buffer to a final volume of 0.5 ml. The reaction was started by the addition of the enzyme. The initial velocities of enzyme were plotted against the substrate concentrations, and the kinetic parameters were estimated by fitting the data to the Michaelis-Menten equation by non-linear regression method using the enzyme kinetics module in GraphPad Prism (GraphPad Software, La Jolla, CA).

4.8 NADH and cortisol docking to DesC. Using BLAST(54), we obtained homologous structures (PDB IDs 4ILK, 4EJ6, 4A2C, 3QE3, 3GFB, 2DQ4, 2DFV, 2D8A, 1PL7, 1E3J) within the PDB. The alignment and placing of both NADH and cortisol molecules on their binding sites was performed using VMD(55). Employing advanced run options of QwikMD(56), the structure of the ligands was minimized in the pocket together with the nearby DesC residues while maintaining the structure of most of DesC static. PyContact(102) was then used to analyze the contact interface.

4.9 Classical molecular dynamics simulations. MD simulations were performed employing the GPU-accelerated NAMD molecular dynamics package(58). The simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for temperature and pressure coupling, the latter kept at 1 bar. A distance cut-off of 12.0 Å was applied to short-range non-bonded interactions, whereas long-range

electrostatic interactions were treated using the particle-mesh Ewald (PME) method(61). The equations of motion were integrated using the r-RESPA multiple time step scheme(103) to update the van der Waals interactions every step and electrostatic interactions every two steps. The time step of integration was chosen to be 2 fs, and before MD simulations, the system was submitted to an energy minimization protocol for 5,000 steps. An MD simulation with position restraints in the protein backbone atoms and ligands non-hydrogen atoms was performed for 10 ns. To allow for a total relaxation of the system and to make sure ligands were stable in the DesC pocket, a 100 ns simulation in equilibrium, where no external forces were applied, was performed. The MD protocol served to pre-equilibrate the system before the hybrid QM/MM simulations.

4.10 Hybrid QM/MM calculations. Quantum mechanics (QM) is crucial to investigate subatomic mechanisms that occur in biology. However, studying entire biomolecular systems quantum mechanically is computationally prohibitive. Traditionally, NAMD(58) employs classical molecular mechanics (MM) to determine the movement of a molecule, solving Newton's equations of motion and treating atoms as spheres and bonds as springs. Combining both approaches, hybrid QM/MM methods employ the quantum mechanics formalism to key regions of the biological system, while using molecular mechanics approach to include the effects of the surrounding area. NAMD's QM/MM interface(66) can be combined with many molecular dynamics protocols, such as enhanced sampling and free energy calculations(82). This combination was crucial in solving the mechanism of DesC. Employing advanced run options of VMD's(55) QwikMD(56) plugin, our in silico approach followed established protocols that were previously employed to investigate the mechanism of setting of the genetic code(66).

In summary, we first perform a short 20 ps long hybrid QM/MM MD simulation using NAMD(66) and MOPAC(104, 105), with a 2.0 fs integration time step. This equilibration simulation was followed by a hybrid QM/MM SMD(106, 107) simulation, with 0.5 fs integration time step, where possible reaction mechanisms were probed. The most favorable reaction mechanism was then tested experimentally and also using more advanced enhanced sampling techniques, namely string simulations. All simulations were carried out using the charge-shift method to treat link atoms, and a “shift” function was applied to surrounding classical partial charges in the electrostatic embedding scheme(66).

4.11 Hybrid QM/MM string simulations. To study a transformation that occurs in a biomolecular system, such as a chemical reaction(66) or a conformational change(108, 109), one defines collective variables such as distances between atoms or between centers of mass of groups of atoms, or even angles between subdomains of a molecular structure(110, 111). The collective variables are used to track the changes in the system as it undergoes the transformation being studied, and can be used to define a reaction coordinate.

The reaction coordinate defined in this study describes the chemical reaction carried out by DesC, and it was optimized by the string method, an iterative method that adapts the reaction coordinate as to fit it to the path of least resistance. Collective variables were defined as distances between atoms relevant for the reaction. From the initial state, before the chemical reaction has occurred, to the final state, after all transformations have taken place, multiple “images” are selected as representative stages of the overall transformation. Over multiple iterations, the string method initiates sets of short, unbiased, and independent QM/MM MD simulations from each image, creating full copies of the system that explore conformations the system can explore along

the reaction coordinate. Then, each set of QM/MM MD simulations is used to re-define the image, leading to a smooth path from initial to final state. Iterations are performed until average consecutive changes converge to a stable value.

Here we have launched 750 multiple independent MD simulations per image, for a total of 25 images. The total simulation time was 4.5 ns, which was enough to converge the simulations in the collective variables space. Collective variables were analyzed using in-house python scripts implemented in Jupyter notebooks. Following previously established protocols(*112, 113*), we analyzed the simulations in direct comparison it with experimental data.

4.11 Hybrid QM/MM free energy calculations. After optimizing the reaction coordinate, we applied the extended adaptive biasing force (eABF) method to estimate the free energy change along the reaction. In eABF, the QM/MM MD simulation is constrained to occur along the reaction coordinate using two new collective variables called “path collective variables” S and Z. The former describes the transformation of the system from start to end, and latter describes transformations “perpendicular” to the reaction coordinate, allowing minor adjustments and optimizations not obtained during the string optimization. We leveraged NAMD’s scalability to initiate multiple QM/MM MD simulations along the reaction coordinate, leading to a multiple-walker eABF calculation that produced extensive sampling over the chemical reaction path.

This approach allowed us to perform 3 ns of hybrid QM/MM simulation, which would not be feasible without the parallel distribution of eABF sampling. Specifically, 375 NCSA Blue Waters XE nodes were used to simulate 750 replicas (walkers) of the system in order to streamline the free energy calculation process.

ASSOCIATED CONTENT

Supporting Information Available

AUTHOR INFORMATION

Corresponding Author

*jmridlon@illinois.edu

Author Contributions

‡These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We would like to thank Prof. Klaus Schulten (in memoriam), whose contribution and support to this work, particularly at its early stages, is of inestimable value. R.C.B. and M.C.R.M. are supported by the National Institutes of Health (NIH) grant P41-GM104601. R.C.B. is also supported by the National Science Foundation (NSF) grant MCB-1616590. We gratefully acknowledge the financial support provided to J.M.R. for new faculty startup through the Department of Animal Sciences at the University of Illinois at Urbana-Champaign (Hatch ILLU-538-916), and the support from the National Institutes of Health grant R01GM134423. Molecular dynamics simulations made use of NCSA Blue Waters supercomputer as part of the Illinois allocation grant “ILL_baxs”. The state of Illinois and the National Science Foundation (awards OCI-0725070 and ACI-1238993) support Blue Waters sustained-petascale computing project. We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. The structural determination work (PDB ID: 4OH1) was supported by the NIH, National Institute of General Medical Sciences (NIGMS), Protein Structure Initiative [U54 GM094586]. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

REFERENCES

1. I. Cann, R. C. Bernardi, R. I. Mackie, Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. *Environ. Microbiol.* **18**, 307–310 (2016).
2. L. A. David *et al.*, Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* **505**, 559 (2014).
3. P. J. Turnbaugh, J. I. Gordon, The core gut microbiome, energy balance and obesity. *J. Physiol.* **587**, 4153–4158 (2009).
4. L. F. Milles, K. Schulten, H. E. Gaub, R. C. Bernardi, Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science (80-.).* **359**, 1527–1533 (2018).
5. I. Cann *et al.*, Thermophilic degradation of hemicellulose, a critical feedstock in the production of bioenergy and other value-added products. *Appl. Environ. Microbiol.* **86** (2020).
6. M. Lyte, Microbial endocrinology. *Gut Microbes.* **5**, 381–389 (2014).
7. P. Strandwitz *et al.*, GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol.* **4**, 396–403 (2019).
8. J. M. Yano *et al.*, Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell.* **161**, 264–276 (2015).
9. J. M. Ridlon, S. C. Harris, S. Bhowmik, D. J. Kang, P. B. Hylemon, Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes.* **7** (2016), pp. 22–39.
10. P. Lefebvre, B. Cariou, F. Lien, F. Kuipers, B. Staels, Role of bile acids and bile acid receptors in metabolic regulation. *Physiol. Rev.* **89** (2009), pp. 147–191.
11. K. G. Jameson, C. A. Olson, S. A. Kazmi, E. Y. Hsiao, Toward Understanding Microbiome-

- Neuronal Signaling. *Mol. Cell* (2020), , doi:10.1016/j.molcel.2020.03.006.
12. D. J. Morris, A. S. Brem, Role of gut metabolism of adrenal corticosteroids and hypertension: clues gut-cleansing antibiotics give us. *Physiol. Genomics*. **51**, 83–89 (2019).
  13. L. V. Hooper, D. R. Littman, A. J. Macpherson, Interactions between the microbiota and the immune system. *Science* (80-.). **336** (2012), pp. 1268–1273.
  14. M. van de Wouw, H. Schellekens, T. G. Dinan, J. F. Cryan, Microbiota-Gut-Brain Axis: Modulator of Host Metabolism and Appetite. *J. Nutr.* **147**, 727–745 (2017).
  15. G. Sharon *et al.*, Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20051–20056 (2010).
  16. O. Bereshchenko, S. Bruscoli, C. Riccardi, Glucocorticoids, Sex Hormones, and Immunity. *Front. Immunol.* **9**, 1332 (2018).
  17. D. W. Cain, J. A. Cidlowski, Immune regulation by glucocorticoids. *Nat. Rev. Immunol.* **17**, 233–247 (2017).
  18. J. M. Ridlon, Conceptualizing the vertebrate sterolbiome. *Appl. Environ. Microbiol.* (2020).
  19. C. Woods, J. W. Tomlinson, in *Advances in Experimental Medicine and Biology* (Springer New York LLC, 2015; http://link.springer.com/10.1007/978-1-4939-2895-8_16), vol. 872, pp. 353–380.
  20. D. J. Morris, J. M. Ridlon, Glucocorticoids and gut bacteria: “The GALF Hypothesis” in the metagenomic era. *Steroids*. **125** (2017), pp. 1–13.
  21. I. A. Macdonald, V. D. Bokkenheuser, J. Winter, A. M. McLernon, E. H. Mosbach, Degradation of steroids in the human gut. *J. Lipid Res.* **24**, 675–700 (1983).
  22. J. Benach *et al.*, Structure of bacterial 3 β /17 β -hydroxysteroid dehydrogenase at 1.2 Å resolution: a model for multiple steroid recognition. *Biochemistry*. **41**, 14659–68 (2002).

23. D. Ghosh, Z. Wawrzak, C. M. Weeks, W. L. Duax, M. Erman, The refined three-dimensional structure of 3 alpha,20 beta-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases. *Structure*. **2**, 629–40 (1994).
24. S. Devendran, C. Méndez-García, J. M. Ridlon, Identification and characterization of a 20 β -HSDH from the anaerobic gut bacterium *Butyricicoccus desmolans* ATCC 43058. *J. Lipid Res.* **58**, 916–925 (2017).
25. H. L. Doden *et al.*, Structural and biochemical characterization of 20 β -hydroxysteroid dehydrogenase from *Bifidobacterium adolescentis* strain L2-32. *J. Biol. Chem.* **294**, 12040–12053 (2019).
26. N. A. Stokes, P. B. Hylemon, Characterization of delta 4-3-ketosteroid-5 beta-reductase and 3 beta-hydroxysteroid dehydrogenase in cell extracts of *Clostridium innocuum*. *Biochim. Biophys. Acta.* **836**, 255–61 (1985).
27. A. C. Swart, K. H. Storbeck, 11 β -hydroxyandrostenedione: Downstream metabolism by 11 β HSD, 17 β HSD and SRD5A produces novel substrates in familiar pathways. *Mol. Cell. Endocrinol.* **408** (2015), pp. 114–123.
28. C. H. L. Shackleton, M. S. Neres, B. A. Hughes, P. M. Stewart, C. E. Kater, 17-Hydroxylase/C17,20-lyase (CYP17) is not the enzyme responsible for side-chain cleavage of cortisol and its metabolites. *Steroids*. **73**, 652–656 (2008).
29. K.-H. Storbeck *et al.*, 11 β -Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: a putative role in castration resistant prostate cancer? *Mol. Cell. Endocrinol.* **377**, 135–46 (2013).
30. E. Pretorius, W. Arlt, K.-H. Storbeck, A new dawn for androgens: Novel lessons from 11-oxygenated C19 steroids. *Mol. Cell. Endocrinol.* **441**, 76–85 (2017).

31. E. Pretorius *et al.*, 11-Ketotestosterone and 11-Ketodihydrotestosterone in Castration Resistant Prostate Cancer: Potent Androgens Which Can No Longer Be Ignored. *PLoS One*. **11**, e0159867 (2016).
32. A. F. Turcu, R. J. Auchus, Clinical significance of 11-oxygenated androgens. *Curr. Opin. Endocrinol. Diabetes. Obes.* **24**, 252–259 (2017).
33. D. J. Morris, S. A. Latif, M. P. Hardy, A. S. Brem, Endogenous inhibitors (GALFs) of 11beta-hydroxysteroid dehydrogenase isoforms 1 and 2: derivatives of adrenally produced corticosterone and cortisol. *J. Steroid Biochem. Mol. Biol.* **104**, 161–8 (2007).
34. A. P. Wade, J. D. H. Slater, A. E. Kellie, M. E. Holliday, Urinary Excretion of 17-Ketosteroids Following Rectal Infusion of Cortisol. *J. Clin. Endocrinol. Metab.* **19**, 444–453 (1959).
35. S. C. Harris *et al.*, Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243T. *Gut Microbes*. **9**, 523–539 (2018).
36. P. de Prada, K. D. Setchell, P. B. Hylemon, Purification and characterization of a novel 17 alpha-hydroxysteroid dehydrogenase from an intestinal Eubacterium sp. VPI 12708. *J. Lipid Res.* **35**, 922–9 (1994).
37. V. D. Bokkenheuser, G. N. Morris, A. E. Ritchie, L. V. Holdeman, J. Winter, Biosynthesis of Androgen from Cortisol by a Species of *Clostridium* Recovered From Human Fecal Flora. *J. Infect. Dis.* **149**, 489–494 (1984).
38. J. M. Ridlon *et al.*, *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. *J. Lipid Res.* **54**, 2437–49 (2013).
39. S. Devendran, S. M. Mythen, J. M. Ridlon, The *desA* and *desB* genes from *Clostridium scindens* ATCC 35704 encode steroid-17,20-desmolase. *J. Lipid Res.* **59**, 1005–1014

- (2018).
40. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*. **570**, 462–467 (2019).
  41. L. K. Ly *et al.*, Bacterial steroid-17,20-desmolase is a taxonomically rare enzymatic pathway that converts prednisone to 1,4-androstenedione-3,11,17-trione, a metabolite that causes proliferation of prostate cancer cells. *J. Steroid Biochem. Mol. Biol.* **199**, 105567 (2020).
  42. M. C. R. Melo, R. C. Bernardi, T. V. a Fernandes, P. G. Pascutti, GSAFold: a new application of GSA to protein structure prediction. *Proteins*. **80**, 2305–10 (2012).
  43. H. M. Senn, W. Thiel, QM/MM methods for biological systems. *At. Approaches Mod. Biol. from Quantum Chem. to Mol. Simulations*. **268**, 173–290 (2007).
  44. K. L. Kavanagh, H. Jörnval, B. Persson, U. Oppermann, Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).
  45. P. J. Baker *et al.*, Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. *Proc. Natl. Acad. Sci.* **106**, 779–784 (2009).
  46. L. Esposito *et al.*, Crystal Structure of a Ternary Complex of the Alcohol Dehydrogenase from *Sulfolobus solfataricus* † , ‡. *Biochemistry*. **42**, 14397–14407 (2003).
  47. K. Ishikawa, N. Higashi, T. Nakamura, T. Matsuura, A. Nakagawa, The First Crystal Structure of l-Threonine Dehydrogenase. *J. Mol. Biol.* **366**, 857–867 (2007).
  48. A. Bowyer *et al.*, Structure and function of the l-threonine dehydrogenase (TkTDH) from the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Struct. Biol.* **168**, 294–304

- (2009).
49. H. Eklund *et al.*, Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**, 27–59 (1976).
  50. T. Tressel, R. Thompson, L. R. Zieske, M. I. Menendez, L. Davis, Interaction between L-threonine dehydrogenase and aminoacetone synthetase and mechanism of aminoacetone production. *J. Biol. Chem.* **261**, 16428–37 (1986).
  51. M. G. Rossmann, D. Moras, K. W. Olsen, Chemical and biological evolution of a nucleotide-binding protein. *Nature*. **250**, 194–199 (1974).
  52. H. M. Berman, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
  53. R. C. Bernardi, I. Cann, K. Schulten, Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels*. **7**, 1–8 (2014).
  54. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
  55. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
  56. J. V Ribeiro *et al.*, QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci. Rep.* **6**, 26536 (2016).
  57. B. C. Goh *et al.*, Computational Methodologies for Real-Space Structural Refinement of Large Macromolecular Complexes. *Annu. Rev. Biophys.* **45** (2016), doi:10.1146/annurev-biophys-062215-011113.
  58. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–802 (2005).
  59. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field

- Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
60. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
  61. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
  62. R. C. Bernardi *et al.*, Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy. *J. Am. Chem. Soc.* **141**, 14752–14763 (2019).
  63. R. C. Bernardi, D. E. B. Gomes, P. G. Pascutti, A. S. Ito, A. T. Ota, Theoretical studies on water-tetracaine interaction. *Int. J. Quantum Chem.* **106**, 1277–1282 (2006).
  64. R. C. Bernardi *et al.*, Water solvent and local anesthetics: A computational study. *Int. J. Quantum Chem.* **107**, 1642–1649 (2007).
  65. R. C. Bernardi *et al.*, Density functional and molecular dynamics simulations of local anesthetics in 0.9% NaCl solution. *Mol. Simul.* **33**, 1135–1141 (2007).
  66. M. C. R. Melo *et al.*, NAMD goes quantum: An integrative suite for hybrid simulations. *Nat. Methods.* **15**, 351–354 (2018).
  67. M. W. Van Der Kamp, A. J. Mulholland, Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry.* **52**, 2708–2728 (2013).
  68. E. Rosta, M. Klähn, A. Warshel, Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *J. Phys. Chem. B.* **110**, 2934–2941 (2006).
  69. V. R. I. Kaila, M. Wikström, G. Hummer, Electrostatics, hydration, and proton transfer

- dynamics in the membrane domain of respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6988–93 (2014).
70. H. M. Senn, W. Thiel, QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **11**, 182–187 (2007).
  71. R. C. Bernardi, P. G. Pascutti, Hybrid QM/MM Molecular Dynamics Study of Benzocaine in a Membrane Environment: How Does a Quantum Mechanical Treatment of Both Anesthetic and Lipids Affect Their Interaction. *J. Chem. Theory Comput.* **8**, 2197–2203 (2012).
  72. Y. S. Mendes *et al.*, The structural dynamics of the flavivirus fusion Peptide-membrane interaction. *PLoS One*. **7**, e47596 (2012).
  73. J. J. P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **19**, 1–32 (2013).
  74. U. C. Singh, P. A. Kollman, A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH₃Cl + Cl⁻ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **7**, 718–730 (1986).
  75. C. Schoeler *et al.*, Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **5**, 5635 (2014).
  76. T. Verdorfer *et al.*, Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J. Am. Chem. Soc.* **139** (2017), doi:10.1021/jacs.7b07574.
  77. S. M. Sedlak, L. C. Schendel, H. E. Gaub, R. C. Bernardi, Streptavidin/biotin: tethering

- geometry defines unbinding mechanics. *Sci. Adv.* **submitted** (2019).
78. S. M. Sedlak *et al.*, Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. *Nano Lett.* **19**, 3415–3421 (2019).
  79. G. V. Dhoke, M. D. Davari, U. Schwaneberg, M. Bocla, QM/MM Calculations Revealing the Resting and Catalytic States in Zinc-Dependent Medium-Chain Dehydrogenases/Reductases. *ACS Catal.* **5**, 3207–3215 (2015).
  80. A. C. Pan, D. Sezer, B. B. Roux, Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B.* **112**, 3432–3440 (2008).
  81. H. Fu, X. Shao, C. Chipot, W. Cai, Extended Adaptive Biasing Force algorithm. An on-the-fly implementation for accurate free-energy calculations. *J. Chem. Theory Comput.* **12**, 3506–3513 (2016).
  82. R. C. Bernardi, M. C. R. Melo, K. Schulten, Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta.* **1850**, 872--877 (2015).
  83. J. Comer, J. C. Phillips, K. Schulten, C. Chipot, Multiple-replica strategies for free-energy calculations in NAMD: Multiple-walker adaptive biasing force and walker selection rules. *J. Chem. Theory Comput.* **10**, 5276–5285 (2014).
  84. V. G. Thackray, Sex, Microbes, and Polycystic Ovary Syndrome. *Trends Endocrinol. Metab.* **30** (2019), pp. 54–65.
  85. L. Yurkovetskiy *et al.*, Gender bias in autoimmunity is influenced by microbiota. *Immunity.* **39**, 400–412 (2013).
  86. J. G. M. Markle *et al.*, Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity. *Science (80-.).* **339**, 1084–1088 (2013).

87. E. García-Gómez, B. González-Pedrajo, I. Camacho-Arroyo, Role of sex steroid hormones in bacterial-host interactions. *Biomed Res. Int.* **2013** (2013), doi:10.1155/2013/928290.
88. R. Menon *et al.*, Diet complexity and estrogen receptor β status affect the composition of the murine intestinal microbiota. *Appl. Environ. Microbiol.* **79**, 5763–5773 (2013).
89. L. Xiao *et al.*, A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1**, 1–6 (2016).
90. A. F. Turcu, J. Rege, R. J. Auchus, W. E. Rainey, 11-Oxygenated androgens in health and disease. *Nat. Rev. Endocrinol.* **16** (2020), pp. 284–296.
91. J. M. Ridlon, J. S. Bajaj, The human gut sterolbiome: bile acid-microbiome endocrine aspects and therapeutics. *Acta Pharm. Sin. B.* **5**, 99–105 (2015).
92. S. Bhowmik *et al.*, Structure and functional characterization of a bile acid 7 α dehydratase BaiE in secondary bile acid synthesis. *Proteins Struct. Funct. Bioinforma.* **84**, 316–331 (2016).
93. A. Tubeleviciute, M. G. Teese, J. Jose, Escherichia coli kduD encodes an oxidoreductase that converts both sugar and steroid substrates. *Appl. Microbiol. Biotechnol.* **98**, 5471–85 (2014).
94. A. Inazu *et al.*, Purification and characterization of a novel dimeric 20 α -hydroxysteroid dehydrogenase from Tetrahymena pyriformis. *Biochem. J.* **297** (Pt 1, 195–200 (1994).
95. J. R. Perilla *et al.*, Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* **31**, 64–74 (2015).
96. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 133–144 (2010).
97. K. Diederichs, Some aspects of quantitative analysis and correction of radiation damage.

- Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 96–101 (2006).
98. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.* **64** (2008), pp. 112–122.
  99. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485 (2010).
  100. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).
  101. M. D. Winn, G. N. Murshudov, M. Z. Papiz, Macromolecular TLS Refinement in REFMAC at Moderate Resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
  102. M. Scheurer *et al.*, PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. *Biophys. J.* **114**, 577–583 (2018).
  103. R. Bernardi *et al.*, NAMD User’s Guide. *Urbana, Illinois, USA Theor. Comput. Biophys. Group, Beckman Institute, Univ. Illinois* (2018).
  104. J. J. P. Stewart, MOPAC: A semiempirical molecular orbital program. *J. Comput. Aided. Mol. Des.* **4**, 1–103 (1990).
  105. J. D. C. Maia *et al.*, GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *J. Chem. Theory Comput.* **8**, 3072–3081 (2012).
  106. C. Schoeler *et al.*, Mapping Mechanical Force Propagation Through Biomolecular Complexes. *Nano Lett.* **15**, 7370–7376 (2015).
  107. J. Seppälä *et al.*, Skeletal Dysplasia Mutations Effect on Human Filamins’ Structure and Mechanosensing. *Sci. Rep.* **7**, 4218 (2017).

108. L. V. B. Hoelz *et al.*, Dynamical behaviour of the human β 1 -adrenoceptor under agonist binding. *Mol. Simul.* **37**, 907–913 (2011).
109. L. V. B. Hoelz *et al.*, The role of helices 5 and 6 on the human β 1-adrenoceptor activation mechanism. *Mol. Simul.* **38**, 236–240 (2012).
110. G. Fiorin, M. L. Klein, J. Hénin, Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).
111. J. Hénin, G. Fiorin, C. Chipot, M. L. Klein, Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **6**, 35–47 (2010).
112. R. C. Bernardi *et al.*, Molecular dynamics study of biomembrane/local anesthetics interactions. *Mol. Phys.* **107**, 1437–1443 (2009).
113. T. J. K. Haataja *et al.*, Non-syndromic Mitral Valve Dysplasia Mutation Changes the Force Resilience and Interaction of Human Filamin A. *Structure*. **27**, 102-112.e4 (2019).