Host immunomodulatory lipids created by symbionts from dietary amino acids

Sungwhan F. Oh[1,2,8,*], T. Praveena[3,8], Hee Bum Song[4], Ji-Sun Yoo[2], Da-Jung Jung[2], Deniz Erturk-Hasdemir[1], Yoon Soo Hwang[4], ChangWon C. Lee[1], Jérôme Le Nours[3], Hyunsoo Kim[4], Jesang Lee[4], Richard S. Blumberg[5], Jamie Rossjohn[3,6,7*], Seung Bum Park[4,*], and Dennis L. Kasper[1,*]

[1] Department of Immunology, Blavatnik Institute of Harvard Medical School, Boston, USA

[2] Center for Experimental Therapeutics and Reperfusion Injury, Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, USA

[3] Infection and Immunity Program & Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

[4] CRI Center for Chemical Proteomics, Department of Chemistry, Seoul National University, Seoul, Republic of Korea

[5] Division of Gastroenterology, Hepatology and Endoscopy, Department of Medicine, Brigham and Women's Hospital, Boston, USA

[6] Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria, Australia

[7] Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff, UK

[8] Co-first authors

* dennis_kasper@hms.harvard.edu (D.L.K.), sbpark@snu.ac.kr (S.B.P.), jamie.rossjohn@monash.edu (J.R.), sungwhan_oh@hms.harvard.edu (S.F.O.)

ABSTRACT

Symbiotic microbiota–derived small molecules are recognized to critically contribute to intestinal immune maturation and regulation [1]. However, little has been done to define the molecular mechanisms controlling immune development in the host-microbiota-environment. Using a targeted lipidomic analysis and synthetic approach, we carried out a multifaceted investigation of immunomodulatory alpha-galactosylceramides from the human symbiont *Bacteroides fragilis* (BfaGCs). Characteristic terminal branching of BfaGCs is the result of incorporation of branched-chain amino acids (BCAAs) taken up in the host gut by *B. fragilis*. A *B. fragilis* knockout strain that cannot metabolize BCAAs showed reduced branching in BfaGCs, and mice monocolonized with this mutant strain had impaired colonic natural killer T (NKT) cell regulation, implying structure-specific immunomodulatory activity. The sphinganine chain branching of BfaGC is a critical determinant of NKT cell activation, which induces unique immunomodulatory gene expression signatures and effector functions. Co-crystal structure and affinity analyses of CD1d–BfaGC–NKT cell receptor complexes confirmed the unique interaction of BfaGCs as CD1d-restricted ligands. We present a structural- and molecular-level paradigm of immunomodulatory control by endobiotic (symbiont-originated) metabolites through dietary/microbial/immune system interdependence.

MAIN

The commensal microbiota has co-evolved with its mammalian host, playing a critical role in the host's immune development and health [2,3]. Among the microbial metabolites with vast chemical diversity, hosts have developed sophisticated machinery to distinguish molecules from pathogens and symbionts[4,5]. Nonetheless, structural variation of microbial metabolites (either intrinsic [6] or resulting from adaptation to the environment [7]) exists, even in a single species. Delineating the biological actions of chemical homologues based their structure is always challenging.

Among hundreds of microbial species residing in the human gut, *Bacteroides fragilis* has frequently been investigated for molecular mechanisms of symbiont-derived bioactive mediators [3,8]. Unique alpha-galactosylceramides (BfaGCs) were identified from this ubiquitous species [9,10] as a prime example of an endobiotic, immunomodulatory molecule. Monocolonization with *B. fragilis* or oral administration of purified microbially-derived BfaGCs at birth can regulate colonic NKT cell number during development, whereas the NKT cell phenotype of mice monocolonized with a BfaGC-deficient mutant resembles that of GF mice. These observations, along with the report that colonization of gastrointestinal tract with CD1d ligand-producing *Sphingomonas* can impact NKT cell functions[11], led us to further delineate the microbial species-specific molecular structures and functions. To elucidate the structural determinants directing the host's immunomodulatory responses, we have used chemically synthesized molecules to how such structural alterations modulate host NKT cell immune responses. To illuminate the structural basis of this immunomodulation of NKT cells, we studied BfaGC presentation by CD1d to the NKT cell receptor and developed a mechanistic understanding of immunomodulatory control by these symbiont-derived metabolites.

BfaGC synthesis & structural assignment

In contrast to previously characterized aGCs of microbial origin which are immunostimulatory CD1d ligands [12], BfaGCs can functionally antagonize NKT cell function and suppress colonic NKT cell proliferation *in vivo* [10]. The distinct structural differences between BfaGCs and classical aGC agonists (e.g., KRN7000 and OCH, Extended Figure 1), including sphingolipid chain lengths, branching and functional groups, implies a structural basis exists for the immunomodulatory impact of symbiont-derived lipid mediators.

Profiling of *B. fragilis* sphingolipids by liquid chromatography–tandem mass spectrometry (LC-MS/MS) determined that BfaGCs are a mixture of homologous acyl chains (combined chain length of C_{32}–C_{36} in sphinganine and fatty acyl groups, Extended Figure 2A), among which C_{34} BfaGC is the most abundant species (Extended Figure 2B). MS/MS fingerprint assignment three C_{34} isobars confirmed the structures as C_{17} sphinganine/C_{17} fatty acyl (C_{17}/C_{17}) and C_{18} sphinganine/C_{16} fatty acyl (C_{18}/C_{16}) BfaGCs (Extended Figures 3A-D). MS/MS spectra of the three C_{34} isobars confirmed that C_{17}/C_{17} and C_{18}/C_{16} BfaGCs co-elute in the second peak, implying additional structural variation in isobars. Isobaric/isomeric mixtures were also identified in other BfaGCs, with total ceramide backbone carbon numbers of C_{32} to C_{36}, consisting sphinganine and fatty acids with C_{15} to C_{19} carbon chain lengths (Extended Figures 3E-H).

Considering the structural diversity of gut commensal sphingolipids, we first designed and undertook a total organic synthesis of BfaGC structural variants (Figure 1A). We systematically constructed 16 BfaGC analogues (SB2201–SB2216), using a matrix-based approach with acyl and sphingoid building blocks containing isomethyl (ω-2) and anteisomethyl (ω-3) branches. To cover structural variants of prominent BfaGCs, we further synthesized 7 isomers (SB2217–2223) with different branching and functional groups. In total, we prepared 23 BfaGC isomers with different chain lengths (C_{15}–C_{19}) and terminal branching (Extended Figure 4 and Supplementary Table 1).

MS/MS fingerprint showed virtually identical fragmentation patterns for synthetic and biogenic BfaGC (Figure 1B). Co-injection analysis of C_{17}/C_{17} BfaGC isomers with different numbers of isomethyl terminal branches (SB2211, SB2217, and SB2219; Figure 1C) revealed that retention time of BfaGCs is negatively correlated with the number of branches in acyl chains: dibranched BfaGCs (i.e., those branched in both sphinganine and acyl chains) elute first, with subsequent elution of monobranched and unbranched isomers. Dibranched isomers are the major BfaGC species—a finding consistent with previous reports on the acyl chain composition of *B. fragilis* fatty acids [13] and ceramides [14], which have shown that ω-2 and ω-3 branched fatty acids are prominent.

Host dietary BCAA direct BfaGC branching

Branched-chain fatty acids, which are structural components of branched-chain BfaGCs, are relatively rare in eukaryotes but are more widely found in prokaryotes [15]. In some bacterial species, biosynthesis of these fatty acids involves the incorporation of deaminated BCAAs such as valine, leucine, and isoleucine [16,17]. To assess the impact of exogenous BCAAs on BfaGC biosynthesis, we first investigated branched BfaGC production by limiting or supplementing exogenous BCAAs *in vitro*. *B. fragilis* grown in rich medium (Extended Figure 5A) has a BfaGC profile dominated by dibranched-chain isomers; in contrast, *B. fragilis* grown in defined minimal medium without amino acids has a significantly different BfaGC profile that tends more toward mono- and unbranched lipids (Extended Figure 5B). This can be reversed by supplementing the medium with individual BCAAs (Extended Figures 5C–E). Addition of valine, which is converted to C_4 (isobutyl)-CoA *in vivo* [18], produces C_{18}/C_{16} BfaGCs as the major species (Extended Figure 5F), whereas supplementation with leucine, which is converted to C_5 (3-methyl-butyl)-CoA, induces C_{17}/C_{17} BfaGCs (Extended Figure 5G). Isoleucine, which is converted to C_5 (2-methyl-butyl)-CoA, also increases the production of C_{17}/C_{17} BfaGCs (Extended Figure 5H) with slightly shorter retention time—a chromatographic characteristic of di-anteisomethyl BfaGCs such as SB2216

(Supplementary Table 1). These results imply that BCAAs dictate BfaGC structure by being directly incorporated into lipids.

To confirm BCAA conversion to branched BfaGCs by the microbes residing in the host gut lumen, we performed *in vivo* stable isotope tracking. When *B. fragilis*–monocolonized mice are given deuterium-labeled leucine in drinking water, branched-chain BfaGCs with one or two molecules of leucine incorporated are identified in the gut luminal contents by MS/MS fingerprinting of deuterium-incorporated acyl chain(s) (Figure 2A and Extended Figure 5I). The incorporation of heavy leucine generates a MS/MS pattern, which is clearly distinguishable from naturally occurring isotopolog (Extended Figures 5J-K), confirming *in vivo* incorporation of host dietary factors to gut symbiont-derived sphingolipids. Furthermore, BCAA supplementation or removal from the host's diet directly affects the ratio of monobranched to dibranched sphingolipids (Figure 2B), a change reflecting a shift to straight-chain sphingolipids in BCAA-deprived diet *in vivo*. These results provide molecular-level evidence that host dietary components can dictate the structure of commensal molecules biosynthesized in the host's gut lumen.

Branched BfaGCs regulate gut NKT

To investigate the immunomodulatory effects of branched-chain BfaGCs on colonic NKT cell development, the host's diet must be manipulated neonatally in *B. fragilis*–monocolonized animals. In order to block bacterial incorporation of BCAA to BfaGC but not to affect host metabolism of these essential amino acids, we genetically targeted the *B. fragilis* BCAA metabolic pathway. BCAA aminotransferase (*BCAT*) deaminates BCAAs to alpha-keto-carboxylic acids, which are further converted to coenzyme A–conjugated, branched-chain fatty acids [18] and downstream lipids, such as BfaGCs. A pBLAST search identified a *B. fragilis* gene (*BF9343-3671*) as having a high degree of homology with mammalian orthologues. A target gene knockout strain (BF9343-Δ3671) was generated (Extended figure

6A), and *in vitro* and *in vivo* phenotypes were assessed. BF9343-Δ3671 shows a growth pattern comparable to that of the isogenic wild-type strain (Extended Figure 6B), with loss of branched C_{17}/C_{17} BfaGCs *in vitro* that can be rescued by complementation of the deleted gene (Extended Figure 6C).

To assess the impact of the loss of branched-chain BfaGCs on host colonic NKT cell development, the offspring of wild-type and BF9343-Δ3671 monocolonized animals were fed normal (BCAA-sufficient) diet, showing similar level of colonization density (Extended Figure 6D). When monoassociated, wild-type *B.fragilis* produces dibranched BfaGCs as the major component in the host gut. In contrast, BF9343-Δ3671 produces significantly higher portions of monobranched or unbranched forms. (Figure 2C). Furthermore, this knockout *B. fragilis* monocolonized mice had significantly higher levels of colonic NKT cells in adulthood (Figure 2D), a result confirming that branched-chain BfaGC is a necessary factor for colonic NKT cell regulation early in life.

BfaGCs are distinct from typical CD1d ligands

To investigate the relevance of structural variation to host immunomodulatory function, we carried out an *in vitro* panel screening assay of our 21 synthetic BfaGCs in a coculture system with bone marrow–derived dendritic cells (BMDCs) and NKT cells (Figure 3A). Levels of interleukin 2 (IL-2) production in the presence of individual BfaGCs clearly distinguish two groups with different potency in stimulating IL-2 production by NKT cells: strong and weak stimulators. Strikingly, all 10 strong stimulators have branched (either ω-2 or ω-3) sphinganine chains, whereas none of the 11 weak stimulators has such branching. To further assess structure-activity relationship of BfaGC-specific functional groups (sphinganine branching and 3′-hydroxyl group), representative structures of branched sphinganine (SB2217) and unbranched sphinganine (SB2219), which are prominently synthesized by *B. fragilis*, were directly compared with two synthetic isomers with 3′-deoxy (SB2222) or n16-sphinganine (SB2223) structures. A dose-response study confirmed that the branching, not the specific length of the

sphinganine chain is critical, however the 3′-OH group of the acyl chain is dispensable for IL-2 induction (Extended Figure 7A).

SB2217 and SB2219 were chosen for further cellular and *in vivo* assays and potential functional differences between BfaGCs and canonical Th1/Th2-type agonists were investigated *in vivo*. When administered intraperitoneally, KRN7000 (at 16 hours after injection; Extended Figure 7B) induces a high level of serum interferon γ (IFN-γ) and OCH (at 2 hours after injection; Extended Figure 7C) did IL-4. In contrast, SB2217 only weakly induces IFN- γ and very low in IL-4. In addition, SB2217 only weakly induces co-stimulatory molecule expression of splenic dendritic cells, whereas SB2219 did not (Extended Figures 7D–F). Collectively, these results collectively suggest that SB2217 has unique immunomodulatory functions via the CD1d–NKT cell receptor axis. The potential immunomodulatory functions were confirmed in NKT cell–mediated oxazolone colitis: prophylactic administration of SB2217 can protect mice from inflammation, reducing weight loss and tissue damage (Figure 3B, C).

SB2217 induces unique NKT cell signaling

To further elucidate NKT cell responses to SB2217 *in vivo*, we analyzed transcriptomic profiles of splenic NKT cells in mice after intraperitoneal injection of agonists. The SB2217-treated group showed profiles at the entire transcriptome level that were clearly distinct from the profiles of the KRN7000-treated group (Extended Figure 8A) and the OCH or SB2219-treated group (Extended Figure 8B). Differentially expressed genes with SB2217 treatment include those encoding several canonical NKT cytokines such as IL-2, IL-4, and IL-13 as well as the multiple immunoregulatory genes ERG1-3, LRRC32 (the transforming growth factor ß activator), and SOCS1/CISH (Figure 3D). These genes are significantly upregulated in SB2217-treated animals when compared to the OCH- or SB2219-treated groups (Figure 3E). Hallmark pathway enrichment analysis [19] showed enrichment of several immune signaling pathways—e.g., the TNF-α/NF-κB, IL-2/STAT5, and IL-6/JAK/STAT3 pathways—in the SB2217-treated

group as opposed to the vehicle-treated and SB2219-treated groups (Extended Figure 8C). These results confirm that SB2217 is an active functional ligand of CD1d and an activator of NKT cells, inducing distinctive immunomodulatory pathways.

Presentation and recognition of BfaGCs

Next, we established the molecular basis for the presentation and recognition of BfaGCs by determining the crystal structures of the NKT cell receptor in complex with mouse CD1d presenting SB2217 and SB2219 (Supplementary Table 3 and Figure 4A). The clarity of the electron density for the two BfaGCs and the mCD1d–TCR interface enabled an examination of detailed molecular interactions at the interface (Extended Figures 9A-B). Here, and common to the two determined crystal structures, the NKT cell receptor adopted the typical parallel docking topology atop the F'-pocket of the CD1d antigen binding cleft (Figure 4A) [20–23]. In both ternary strucutres, the F'-pocket of CD1d adopted a closed conformation, as previously observed [24]. Within these ternary complexes, the galactose headgroup of the branched and unbranched BfaGCs adopted a highly conserved position similar to that of KRN7000 [23]. Specifically, the galactose headgroup protruded out of the CD1d binding cleft for TCR recognition, whereas the acyl and sphinganine chains were buried deep within the A'- and F'-pockets, respectively (Extended Figure 9C) [20,23]. The TCRα chain dominated the contacts with the bound BfaGCs whereby Asn30α from the complementarity-determining region (CDR) 1α hydrogen bonded the 3''-OH of the galactose moiety, whereas Gly96α and Arg95α from CDR3α formed a hydrogen bond and van der Waals contact with the 2''-OH and the sphinganine chain 3-OH, respectively (Extended Figure 9D and Supplementary Tables 4, 5). The interactions of mCD1d with the two BfaGCs were also largely conserved (Figure 4B); that is, Asp153, Thr156, and Asp80 hydrogen bonded to the 2''-OH, 3''-OH, glycosidic oxygen, and sphinganine 3-OH of the BfaGCs. Thr159 also formed a hydrogen bond with 3'-OH in the acyl chain that is not present in KRN7000. Thus, this interaction compensated for loss of the

197 hydrogen bond between sphinganine 4-OH and Asp80 observed in KRN7000 [20,23]. Further, relative to
198 unbranched SB2219, the branching of the sphinganine chain in SB2217 did not affect the overall position
199 of the sphinganine chain within the F'- pocket of CD1d, and the positions of the CD1d residues
200 contacting the sphinganine chain were also completely conserved (Figure 4C). In contrast, there was a
201 difference in the contacts mediated by the BfaGCs within the F'-pocket of CD1d; compared to what was
202 seen with CD1d–BfaGC, the longer sphinganine chain of KRN7000 protruded further into the F'-pocket
203 of CD1d—resulting in more interactions with CD1d itself and causing conformational changes in the
204 surrounding CD1d residues (Met88 and Trp142, Figure 4D). Accordingly, differences between the
205 interactions of BfaGCs and KRN7000 with the F'-pocket of CD1d likely correlate with differing abilities
206 to be presented by CD1d and therefore with the agonistic properties of these CD1d-restricted ligands.
207 Collectively, these molecular-level insights unambiguously confirm BfaGCs as ligands presented by
208 CD1d and recognized in a conserved manner by the NKT cell receptor.

209 We also measured the binding affinity of the soluble NKT cell receptor for mCD1d *in vitro*,
210 loaded with SB2217 and SB2219, using surface plasmon resonance (SPR). Here, the NKT cell receptor
211 exhibited affinity to mCD1d presenting SB2217 (K_D~ 127 nM) similar to that for KRN7000 (K_D~ 102 nM,
212 Extended Figures 9E-F). In contrast, SB2219-loaded mCD1d showed moderately lower affinity for the
213 NKT cell receptor (K_D~ 246 nM) that was attributable to an observably slower on-rate , thereby
214 suggesting non-stimulatory ligands may possess lower affinities or half lives for the NKT TCR. Indeed,
215 consistent with previous observations[24,25], this result suggests remodeling of the CD1d-SB2219 complex
216 upon NKT cell receptor engagement as well as a functional difference between SB2217 and SB2219.

217 Since the discovery of agelasphin from the marine sponge [26,27] and the development of KRN7000[24]
218 as a prototypic CD1d ligand, the canonical functions of NKT cells have been mostly appreciated in the
219 context of host defense and immunity, and aGCs have been recognized as potent agonists of NKT cell–

mediated immune stimulation. This perspective led to the investigation of various synthetic aGCs as antitumor therapeutic agents [28] and vaccine adjuvants [29]. Recent results have documented NKT cell–mediated immunomodulatory actions [30]. For example, a subset of NKT cells has been functionally characterized as anti-inflammatory [31], and NKT cells can promote IL-10 production by intestinal epithelial cells in a CD1d-dependent manner [32]. Multiple studies have proposed that such functional diversity of NKT cells depends on the structure of ligands as well as on the nature of CD1d-expressing antigen-presenting cells (APCs) and NKT cell subtypes [32–34]. Our study has elucidated unique host immune responses induced by symbiont-derived NKT cell activators, as well as the distinctive interaction of these CD1d ligands with the invariant NKT cell receptor.

Harboring by far the largest number and greatest density of resident bacteria, the host gut lumen absorbs diverse nutrients and their microbial metabolites (secondary metabolites), as well as small molecules primarily synthesized by symbionts (endobiotic metabolites). In this study synthetically prepared chemical homologues enabled us to characterize these endobiotic glycosphingolipids at the molecular level on the basis of their structures. Furthermore, *in vitro* and *in vivo* biological assessment with synthetic ligands revealed that BfaGCs are a distinct type of CD1d ligand and NKT agonist, providing an explanation for their previously reported stimulatory [9] and regulatory [10] natures.

Of considerable interest, we discovered that a change in one component of the host's nutrition (BCAAs) can directly influence the structure of bacterial metabolites and that these structural modifications have an impact on host immunoregulation. We have found that the host, the microbiota, and the environment all make key contributions to the production and function of BfaGCs. Delineation of the molecular-level interdependence of these three contributors is challenging but feasible in the mouse system where variables can be controlled. Using an experimental model of human gut microbiota, we have investigated BfaGC profiles in human microbiota-associated gnotobiotic mice [35].

BfaGCs are produced in the complex microbiota containing *B. fragilis* and the quantity is positively correlated with *B. fragilis* abundance in the mouse gut lumen (Extended figure 10A). BfaGCs are identified in the colon of neonatal HMB mice (Extended figure 10B), supporting the contention that they may also exist in human gut in early life since *B. fragilis* is a ubiquitous human gut symbiont and has a unique colonization pattern in early life [36]. These results collectively imply that BfaGC may be an important contributor to gut immunity. Future studies of essential dietary factors and immunomodulatory mediators, and their synergistic contribution in immune development are warranted.

REFERENCES

1. Surana, N. K. & Kasper, D. L. Deciphering the tête-à-tête between the microbiota and the immune system. *J. Clin. Invest.* **124**, 4197–203 (2014).

2. Skelly, A. N., Sato, Y., Kearney, S. & Honda, K. Mining the microbiota for microbial and metabolite-based immunotherapies. *Nature Reviews Immunology* **19**, 305–323 (2019).

3. Surana, N. K. & Kasper, D. L. The yin yang of bacterial polysaccharides: lessons learned from B. fragilis PSA. *Immunol. Rev.* **245**, 13–26 (2012).

4. Erturk-Hasdemir, D. *et al.* Symbionts exploit complex signaling to educate the immune system. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 26157–26166 (2019).

5. Vatanen, T. *et al.* Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* **165**, 842–853 (2016).

6. d'Hennezel, E., Abubucker, S., Murphy, L. O. & Cullen, T. W. Total Lipopolysaccharide from the Human Gut Microbiome Silences Toll-Like Receptor Signaling. *mSystems* **2**, (2017).

7. Kawahara, K., Tsukano, H., Watanabe, H., Lindner, B. & Matsuura, M. Modification of the structure and activity of lipid A in Yersinia pestis lipopolysaccharide by growth temperature. *Infect. Immun.* **70**, 4092–4098 (2002).

8. Erturk-Hasdemir, D. & Kasper, D. L. Finding a needle in a haystack: Bacteroides fragilis polysaccharide a as the archetypical symbiosis factor. *Annals of the New York Academy of Sciences* **1417**, 116–129 (2018).

9. Wieland Brown, L. C. *et al.* Production of α-Galactosylceramide by a Prominent Member of the Human Gut Microbiota. *PLoS Biol.* **11**, e1001610 (2013).

10. An, D. *et al.* Sphingolipids from a Symbiotic Microbe Regulate Homeostasis of Host Intestinal Natural Killer T Cells. *Cell* **156**, 123–133 (2014).

11. Wingender, G. *et al.* Intestinal microbes affect phenotypes and functions of invariant natural killer T cells in mice. *Gastroenterology* **143**, 418–28 (2012).

12. Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520–525 (2005).

13. Brondz, I. & Olsen, I. Multivariate analyses of cellular fatty acids in Bacteroides, Prevotella, Porphyromonas, Wolinella, and Campylobacter spp. *J. Clin. Microbiol.* **29**, 183–9 (1991).

14. Miyagawa, E., Azuma, R., Suto, T. & Yano, I. Occurrence of free ceramides in Bacteroides fragilis NCTC 9343. *J. Biochem.* **86**, 311–20 (1979).

15. Leo, R. F. & Parker, P. L. Branched-Chain Fatty Acids in Sediments. *Science (80-.).* **152**, 649–650 (1966).

16. Naik, D. N. & Kaneda, T. Biosynthesis of branched long-chain fatty acids by species of Bacillus: relative activity of three alpha-keto acid substrates and factors affecting chain length. *Can. J. Microbiol.* **20**, 1701–8 (1974).

17. Beck, H. C. Branched-chain fatty acid biosynthesis in a branched-chain amino acid aminotransferase mutant of *Staphylococcus carnosus*. *FEMS Microbiol. Lett.* **243**, 37–44 (2005).

18. Kaneda, T. Iso-and Anteiso-Fatty Acids in Bacteria: Biosynthesis, Function, and Taxonomic Significancet. *Microbiol. Rev.* **55**, 288–302 (1991).

19. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).

20. Pellicci, D. G. *et al.* Differential Recognition of CD1d-α-Galactosyl Ceramide by the Vβ8.2 and Vβ7 Semi-invariant NKT T Cell Receptors. *Immunity* **31**, 47–59 (2009).

21. Girardi, E. & Zajonc, D. M. Molecular basis of lipid antigen presentation by CD1d and recognition by natural killer T cells. *Immunol. Rev.* **250**, 167–179 (2012).

22. Rossjohn, J., Pellicci, D. G., Patel, O., Gapin, L. & Godfrey, D. I. Recognition of CD1d-restricted antigens by natural killer T cells. *Nature Reviews Immunology* **12**, 845–857 (2012).

23. Chennamadhavuni, D. *et al.* Dual Modifications of α-Galactosylceramide Synergize to Promote Activation of Human Invariant Natural Killer T Cells and Stimulate Anti-tumor Immunity. *Cell Chem. Biol.* **25**, 571-584.e8 (2018).

24. Li, Y. *et al.* The Vα14 invariant natural killer T cell TCR forces microbial glycolipids and CD1d into a conserved binding mode. *J. Exp. Med.* **207**, 2383–2393 (2010).

25. Wun, K. S. *et al.* A molecular basis for the exquisite CD1d-restricted antigen specificity and functional responses of natural killer T cells. *Immunity* **34**, 327–39 (2011).

26. Natori, T., Koezuka, Y. & Higa, T. Agelasphins, novel α-galactosylceramides from the marine sponge Agelas mauritianus. *Tetrahedron Lett.* **34**, 5591–5592 (1993).

27. Kobayashi, E. *et al.* Enhancing effects of agelasphin-11 on natural killer cell activities of normal and tumor-bearing mice. *Biol. Pharm. Bull.* **19**, 350–3 (1996).

28. Kobayashi, E., Motoki, K., Uchida, T., Fukushima, H. & Koezuka, Y. KRN7000, a novel immunomodulator, and its antitumor activities. *Oncol. Res.* **7**, 529–34 (1995).

29. Li, X. *et al.* Design of a potent CD1d-binding NKT cell ligand as a vaccine adjuvant. *Proc. Natl. Acad. Sci.* **107**, 13010–13015 (2010).

30. Laurent, X. *et al.* Switching Invariant Natural Killer T (iNKT) Cell Response from Anticancerous to Anti-Inflammatory Effect: Molecular Bases. *J. Med. Chem.* **57**, 5489–5508 (2014).

31. Sag, D., Krause, P., Hedrick, C. C., Kronenberg, M. & Wingender, G. IL-10–producing NKT10 cells are a distinct regulatory invariant NKT cell subset. *J. Clin. Invest.* **124**, 3725–3740 (2014).

32. Olszak, T. *et al.* Protective mucosal immunity mediated by epithelial CD1d and IL-10. *Nature* **509**, 497–502 (2014).

33. Brutkiewicz, R. R. CD1d Ligands: The Good, the Bad, and the Ugly. *J. Immunol.* **177**, 769–775 (2006).

34. Joyce, S., Girardi, E. & Zajonc, D. M. NKT cell ligand recognition logic: molecular basis for a synaptic duet and transmission of inflammatory effectors. *J. Immunol.* **187**, 1081–9 (2011).

35. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578–93 (2012).

36. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).

FIGURE LEGENDS

Figure 1. Chemical structure assignment of chain-length and branching variation in BfaGCs. (A) Retrosynthetic analysis of BfaGC structural homologues. BfaGCs with all possible chain lengths and terminal branches in both fatty acyl chains (blue block with R group) and sphingoid chains (red block with R' group) were designed and prepared. (B) An MS/MS spectrum mirror plot of synthetic and biogenic BfaGCs shows essentially identical spectra. (C) Co-injection analyses of *B. fragilis* lipid extract with synthetic BfaGCs assigned the structures of major bacterial glycosphingolipids (C34 BfaGCs). Chromatograms and spectra represent three samples.

Figure 2. BfaGC branching is dictated by host dietary BCAAs, and loss of BCAA utilization in B. fragilis impairs its ability to modulate host colonic NKT cells. (A) Host dietary BCAA is directly incorporated into BfaGCs produced by B. fragilis in the large intestine. The MS/MS fingerprint (m/z = 493) of d6-C34 BfaGC confirms d3-leucine incorporation into both sphinganine and fatty acyl chains. The spectrum represent four samples. (B) The gut luminal BfaGC branching of B. fragilis–monocolonized mice is directly associated with the BCAA availability in the diet (N=11 for baseline and N=10 for d7/d14 groups, $p<0.0001$ for both comparisons). (C) BF9343-Δ3671 monocolonization exhibits significantly less branched C34 BfaGCs (N=4 for each group, $p=0.0079$ for dibranched and $p=0.00043$ for monobranched). (D) BF9343-Δ3671 monocolonization at birth show dysregulated colonic NKT cell levels ($p=0.0050$).

Figure 3. BfaGCs have distinct immunomodulatory signaling and actions. (A) APC-NKT cell coculture with 21 synthetic BfaGCs showed a clear difference in IL-2 inducing efficacy, dependent on their sphinganine branching. Representative structures of BfaGCs with branched-chain sphinganine (SB2217) and with straight-chain sphinganine (SB2219) are shown. Results represents three independent experiments with similar trend. (B-C) When given prophylactically (N=5 for vehicle-treated and N=6 for SB2217-treated group), SB2217 can protect the host from NKT-mediated colitis, shown as less loss of weight (B, p=0.038) and less severe tissue histopathology score (C, p=0.015). Results represent three independent sets with similar trend. (D-E) Transcriptomic profile of splenic NKT cells from SB2217-treated group (mean of N=3) showed a distinct gene expression pattern. A volcano plot (D) and a heat map analysis (E) of differentially expressed, potential immunomodulatory genes with SB2217 treatment.

Figure 4. Crystal structure of 2C12 NKT cell receptor–mCD1d–BfaGC ternary complexes showed conserved and distinct molecular interactions of BfaGCs with mCD1d and the 2C12 TCR. (A) Cartoon representation of 2C12 TCR–mCD1d–BfaGC ternary complex crystal structures. The mCD1d and β2m are colored wheat and brown, respectively, while the 2C12 TCR α- and β-chains are colored purple and violet, respectively. CDR loops are colored as follows: CDR1α, blue; CDR2α, lime green; CDR3α, cyan; CDR1β, yellow; CDR2β, orange; CDR3β, pink. The lipid antigens are represented as sticks: SB2217, blue; SB2219, green. (B) Molecular interactions of SB2217 with mCD1d. Hydrogen bonds are shown as red dashed lines. (C) Superimposition of the sphinganine chains of branched-chain SB2217 and unbranched SB2219. (D) Comparison of interactions with mCD1d, with superimposition of the α1 and α2 domains of the crystal structures of 2C12 TCR–mCD1d–SB2217 (wheat) and 2C12 TCR–mCD1d–KRN7000 (gray) (PDB code: 6BNK). KRN7000 and SB2217 are shown as black and blue sticks, respectively.

METHODS

Mice

All animal procedures were supervised by the Harvard Center for Comparative Medicine and maintained by the Institutional Animal Care and Use Facility. All mice were 5–8 weeks old, and the experimental groups were age-matched with one another. All mice are housed under 12-hour light-dark cycle and controlled climate (temperature: 21 ℃, humidity: 50%).

Germ-free (GF) Swiss-Webster mice were bred and maintained in inflatable plastic isolators. Monocolonized mice were prepared by gavage of breeding pairs with a single bacterial strain (*B. fragilis* wild-type strain NCTC9343 or knockout strain BF9343-Δ3671) and were maintained in isolators to obtain offspring (F1 and later-generation) for experiments. Stool samples from GF and monocolonized mice in isolators were regularly streaked onto plates and grown in both aerobic and anaerobic conditions in order to confirm sterility and freedom from contamination.

Specific pathogen–free (SPF) Swiss-Webster mice were purchased from Taconic. For *ex vivo* coculture assays, *in vivo* cytokine profiling and transcriptomic analysis, conventional C57BL/6 mice were obtained from Taconic.

Human Microbiota associated (HMB) Swiss-Webster mice [35] were bred and maintained in inflatable plastic isolators.

Bacterial culture

Individual bacterial cultures, maintained as frozen stock [37], were first streaked onto plates; a single colony was picked up and inoculated into ~1 mL of deoxygenated rich medium (2% proteose peptone, 0.5% yeast extract, 0.5% NaCl) supplemented with 0.5% D-glucose, 0.5% K_2HPO_4, 0.05% L-cysteine, 5

392 mg/L hemin and 2.5 mg/L vitamin K1) in an anaerobic chamber. Samples were grown overnight,
393 centrifuged, and kept at –80°C until extraction.

394 For *in vitro* BCAA supplementation experiments, a minimal liquid medium[38] (0.1% $(NH_4)_2SO_4$, 0.1%
395 Na_2CO_3, 0.09% KH_2PO_4, 0.09% NaCl, 26.5 mg/L $CaCl_2.2H_2O$, 20 mg/L $MgCl_2.6H_2O$, 10 mg/L $MnCl_2.4H_2O$,
396 1 mg/L $CoCl_2.6H_2O$, 0.5% D-glucose, 0.05% L-cysteine, 5 mg/L hemin, 2.5 mg/L vitamin K_1, 2 mg/L
397 $FeSO_4.7H_2O$ and 5 μg/L vitamin B_{12} was used as basal media, with individual supplementation of 3 mM
398 BCAA (Val, Leu, Ile).

399 For bacterial plating, brain heart Infusion agar plates (3.7% brain heart infusion powder and 1.5% agar)
400 supplemented with 5 mg/L hemin and 2.5 mg/L vitamin K_1 was used. For selection, liquid media and
401 agar plates were supplemented with the following antibiotics: 100 μg/mL ampicillin (*E. coli*), 10 μg/mL
402 erythromycin (*B. fragilis*), 200 μg/mL gentamycin (*B. fragilis*) and 50 ng/mL anhydrotetracycline (*B.*
403 *fragilis*).

404 Lipidomic analyses

405 *Sample preparation.* Samples (about 5mg bacterial cell pellets or 50 mg mouse fecal or colonic contents)
406 were extracted with a methyl tert-butyl ether (MTBE)–methanol–water mixture [39], using perduterated
407 (d35) beta-galactosylceramide (Matreya LLC) as an internal standard.

408 *UHPLC-MS/MS condition.* An UHPLC-MS/MS system (Thermo Scientific Vanquish RP-UPLC connected
409 to a Q Exactive Orbitrap) was used for sphingolipid profiling, updated from previous work [10]. A
410 negative ion mode method was established with parameters of spray voltage, 3kV; sheath gas, 60AU;
411 auxiliary gas, 15AU; capillary temperature, 320°C; probe heater temperature, 400°C; mean collision
412 energy, 22.5AU. Agilent Zorbax C18 column (4.6 mm × 75 mm × 1.8 μm, 600 μL/min) was used for the

gradient LC elution (65% acetonitrile/25% 2-propanol/0.05% formic acid to 80% 2-propanol/10% acetonitrile/0.05% formic acid) over 20 min, at 40 °C.

Targeted lipidomics. A high-resolution (R=70,000 @ m/z 200) MS1 scan (500–1000Da followed by data-dependent acquisition (DDA), or a parallel reaction monitoring (PRM) method was established to acquire MS and MS/MS spectra by Xcalibur 4.0 (Thermo Fisher Scientific). An inclusion list with C_{32}–C_{36} BfaGCs was generated and used for targeted analysis (m/z of individual species for XIC was selected as [M+HCOO-]-). MS/MS spectra of biogenic and synthetic molecules were acquired and directly compared. Relative quantitation of individual BfaGCs was done by quantitation of area under MS1 peaks (relative quantitation and comparison: *in vitro* BfaGC profiling and BCAA manipulation study) or fingerprint MS2 peaks (C_{17}/C_{17} and C_{16}/C_{18} BfaGC identification and quantitation: BfaGC structure assignment and *B. fragilis* WT/KO monocolonization experiment), with normalization by internal standard recovery and sample weight.

Total organic synthesis of BfaGC analogues

Materials. All commercially available reagents were purchased from Sigma-Aldrich, Tokyo Chemical Industry or ThermoFisher Scientific and were used without further purification unless otherwise specified. Solvents were purchased from commercial vendors and used without further purification unless otherwise stated. Dry solvents were prepared with an ultimate solvent purification system CT-SPS-SA (Contour Glass). The progress of reactions was monitored by thin-layer chromatography (TLC) (silica gel 60, F_{254} 0.25 mm). Components on TLC were visualized by treatment of the TLC plates with *p*-anisaldehyde, $KMnO_4$, or phosphomolybdic acid and subsequent heating. The compounds were purified by flash column chromatography on silica gel (230–400 mesh).

Compound characterization. The optical rotations were measured by JP/P-1030 (JASCO) with a sodium lamp (D line, 589 nm). ^{1}H and ^{13}C NMR spectra were determined on a Bruker DRX-300 (Bruker Biospin), Agilent 400-MR DD2 (Agilent Technologies), or Varian Inova-500 (Varian Associates) instrument. Chemical shifts were reported in parts per million (δ, ppm). ^{1}H NMR spectra were calibrated with the residual solvent peak ($CDCl_3$, 7.26 ppm) or tetramethylsilane (TMS, 0.00 ppm) used as the internal standard. ^{13}C NMR spectra were calibrated using the residual solvent peak ($CDCl_3$, 77.23 ppm; CD_3OD, 49.00 ppm). Multiplicity was noted as follows: s (singlet); d (doublet); t (triplet); q (quartet); m (multiplet); dd (doublet of doublet); dt (doublet of triplet); td (triplet of doublet); br s (broad singlet), etc. Coupling constants were reported in Hz. Low-resolution mass spectrometry was assessed by LCMS-2020 (Shimadzu) and LTQ (ThermoFisher Scientific) with electron spray ionization.

Synthesis strategies. To maximize synthetic efficacy and structural diversity, we designed our synthetic route as a combinatorial approach between acyl building blocks and sphingoids with different terminal structures—e.g., ω-2 or ω-3 branches—and normal chains. After the preparation of acyl and sphingoid building blocks, combination of each building block with an amide bond provided all possible BfaGC analogues (SB2201–SB2216) covering full mass ranges (C_{32}–C_{36}) as well as structural diversity at the terminal position of both acyl chain and sphingosine (Extended Figure 3A and Supplementary Table 1). Our newly synthesized 16-member BfaGC analogue library covers a full mass range of major extracted BfaGCs (C_{32}–C_{36}), with excellent structural variation. By combining synthetic building blocks with ω-2 or ω-3 branches or normal chains at the terminal position, we easily generated unbranched, monobranched, and dibranched BfaGC analogues. In this context, we synthesized five additional isobaric C_{34} BfaGC analogues (SB2217–SB2221) to fully cover the structural diversity of C_{17}/C_{17} isomers (Extended Figure 4 and Supplementary Table 1). Also, we synthesized two additional analogues derived from SB2217 to reveal the effect of 3′-OH of acyl chain and branch of sphinganine (SB2222 and SB2223), thus completing

457 the synthesis of a 23-member BfaGC analogue library. The reader is referred to the auxiliary supporting
458 material for details on procedures and proton and ^{13}C NMR spectra.

459 Co-injection (matching) analysis

460 Synthetic BfaGCs (SB2211, SB2217, and SB2219) were individually mixed with total lipid extract of
461 cultured *B. fragilis.*

462 Bacterial mutant generation and complementation

463 A *B. fragilis* NCTC9343 BF9343-3671 mutant (BF9343-Δ3671) was constructed with a pNJR6 suicide vector
464 [40]. The entire BF9343_3671 protein coding region was deleted. Complementation of BF9343_3671 to
465 BF9343-Δ3671 was conducted with a pNBU2_erm-TetR-P1T_DP-GH023 vector [41]. Deletion of targeted
466 locus was confirmed by PCR (Extended Figure 6A).

467 *In vivo* BCAA supplementation and removal

468 *In vivo stable isotope tracking.* Deuterium-labeled leucine (d3-Leu, Millipore Sigma) was dissolved to 1%
469 (w/v) in drinking water and given to *B. fragilis*–monocolonized mice for 7 days, at which point fecal
470 samples were collected. A modified LC-MS/MS method, adding 3 and 6 Dalton larger parent ion in MS1
471 inclusion list, was used to detect d3- and d6-BfaGCs.

472 *Dietary BCAA manipulation. B. fragilis*–monocolonized mice were fed (*ad libitum*) a BCAA-sufficient
473 formulated diet (2.8% total BCAA; Testdiet 5CC7) for 7 days and then switched to a BCAA-deficient
474 formulated diet (Testdiet 58ZX) for the next 7 days. Stool samples were collected at baseline (before diet
475 modification), at day 7, and at day 14, and the ratio between C_{34} dibranched and monobranched BfaGCs
476 was quantified by MS1.

477 Colonic NKT cell analysis

Colonic lamina propria lymphocyte isolation [42]. Conventional (SPF) mice, and mice monocolonized from birth with *B. fragilis* strain BF9343-Δ3671 were euthanized. The large intestines were collected and fat tissue was removed. The intestine was opened longitudinally, and fecal content was removed, cut into 1-inch pieces, and shaken in HBSS containing 2 mM EDTA for 50 min at 37°C. After the removal of epithelial cells, the intestines were washed in HBSS and incubated with RPMI 1640 containing 10% FBS, collagenase type VIII (1 mg/mL), and DNase I (0.1 mg/mL) (Sigma-Aldrich) for 45 min at 37°C under constant shaking. The digested tissues were mixed with FACS buffer (PBS with 2% FBS and 1 mM EDTA), filtered twice through strainers (mesh size, 70 and 40 μm), and used for flow cytometry.

FACS analysis. For staining with the indicated dilution, APC-labeled mouse CD1d tetramer—unloaded or loaded with PBS-57 (1:500; NIH Tetramer Core Facility)—as well as anti-mouse CD3–FITC (1:400), TCRβ–PE (1:400), CD45–PerCP–Cy5.5 (1:200; Biolegend) and cell viability dye (Fixable Viability Dye eFluor™ 780, 1:1000; ThermoFisher) were used. Individual samples were stained for 20 min at 4°C and washed with cold FACS buffer. FACS analysis was performed with a BD FACSCanto system (BD Biosciences), pre-gated with forward and side scatters, a singlet population, and viable cells. The frequencies of CD3+/CD1d tetramer–positive cells from the gated total CD45+ population was enumerated as the target population. Data were analyzed and quantified with FlowJo V10 software (BD Biosciences).

In vitro antigen presentation assay

The IL-2-producing potential of synthetic aGCs KRN7000 (Avanti Polar Lipid), SB2217, and SB2219 was assessed with a slightly modified protocol as described previously [10]. In brief, 5×10^4 BMDCs (bone-marrow monocytes cultured for 7–8 days in the presence of GM-CSF [10 ng/mL]; PeproTech) were pre-incubated with serial dilutions of individual aGCs. After 2 hours, 5×10^4 24.7 NKT hybridoma cells were

added and incubated for 18 hours. The IL-2 levels in culture supernatants were analyzed by ELISA (R&D Systems).

In vivo treatment of aGCs

Serum cytokine ELISA. A 1-μg volume of synthetic NKT cell ligands was injected intraperitoneally into C57BL/6 mice. Serum was collected 2 and 18 hours after injection for measurement of IL-4 and IFN-γ, respectively, by ELISA (Biolegend).

Splenocyte preparation. Eighteen hours after injection of aGCs, spleens of mice were collected and digested in RPMI 1640 containing 10% FBS, 1.5 mM HEPES, and collagenase type IV (1 mg/mL) (Sigma-Aldrich) at 37°C for 1 hour. The digested tissues were mechanically ground and filtered through a strainer (mesh size, 70 μm), yielding a single-cell suspension. After red blood cell lysis, remained splenocytes were used for flow cytometry.

FACS analysis. For splenic dendritic cell analysis, total splenocytes were stained with cell viability dye (eFluor™ 780, 1:1000) and anti-mouse CD16/32 (1:500), MHC class II–PE Cy7 (1:500), CD11c–PerCP–Cy5.5 (1:500), CD40–PB (1:300), CD80–PE (1:500), and CD86–APC (1:500). All antibodies were obtained from Biolegend. FACS analysis was performed by LSR II, and data were processed with FlowJo V10 software.

Oxazolone-mediated colitis

From 1 day before until 2 days after challenge (total, four doses), animals received a daily intraperitoneal injection of vehicle or 1 μg of SB2217 in 0.9% DMSO in PBS solution. Oxazolone (4-ethoxymethylene-2-phenyl-2-oxazolin-5-one; Sigma-Aldrich) was administered intrarectally (through a 3.5F catheter) as a 1% (w/v) solution in 50% ethanol (5 μL/g of body weight).

Animals were monitored daily for body weight and macroscopic health condition. At day 3, mice were euthanized and their colons excised for histopathology scoring by a pathologist blinded to treatment groups. The histologic score represented the combined scores for inflammation and ulceration (0–4, 0 being normal and 4 being most severe).

Transcriptomic analysis

Single-cell suspensions of spleen cells were prepared 2 hours after intraperitoneal injection of aGCs as explained above. Splenic NKT cells were sorted with a BD FACSAria II for RNA preparation. RNAs were prepared with TRIzol (ThermoFisher Scientific) according to the manufacturer's recommendations. RNA libraries were constructed with a SMARTer Stranded Total RNA-Seq Kit v2-Pico Input Mammalian (Takara Bio) and sequenced on an Illumina sequencer. Paired-end reads were trimmed with TrimGalore (v0.4.5). Trimmed reads were aligned to the GENCODE M25 genome with STAR [43] (v2.7.3a). Read counts were obtained with featureCounts [44] (v2.0.0). To identify differentially expressed genes, gene counts were processed with DESeq2 [45] (v1.32.0). Heat maps were generated with the R package pheatmap (v1.0.12) with z-transformation, and other plots were generated with R package ggplot2 [46] (v3.3.5). Gene set enrichment analyses were performed with R packages fgsea [47] (v1.18.0) and msigdbr (v7.4.1).

Structural study

Cloning, expression, and purification of mouse CD1d and the 2C12 TCR. The mCD1d genes including the heavy chain and β2M were cloned into a pFastBac™ vector (Thermo Fisher Scientific) along with BirA and histidine tags. The expression and purification protocols were followed as previously described [48]. In brief, in the baculovirus expression system, viral stocks were prepared using Sf9 cells, and transfection was subsequently carried out in Hi5 cells. The secreted recombinant protein was purified by

543 conventional Ni-NTA affinity chromatography followed by size-exclusion chromatography. The
544 refolding of 2C12 TCR was accomplished by the protocol described previously [23]. In brief, α:β chain
545 inclusion-body proteins (120 mg:144 mg) were added to the refolding buffer containing 8 M urea; 100
546 mM Tris-HCL, pH 8.5; 2 mM EDTA; 0.4 M arginine; 0.5 mM oxidized glutathione; 5 mM reduced
547 glutathione; and 0.2 mM phenylmethylsulfonyl fluoride (PMSF) in the presence of dithiothreitol at 4°C.
548 The refolded 2C12 TCR was purified by a series of steps involving anion-exchange, size-exclusion, and
549 hydrophobic interaction chromatography.
550 *In vitro loading of BfaGCs to mCD1d and 2C12 TCR co-complexation.* Synthetic BfaGCs (SB2217, SB2219)
551 were individually dissolved (0.5 mg/mL in 0.5% tyloxapol). Prior to loading, the solution was vigorously
552 mixed by sonication for 20 min, heated at 60°C for 1 min, and then cooled for 1 min. Individual BfaGCs
553 and mCD1d were mixed at a 3:1 molar ratio and incubated overnight in 10 mM Tris-buffered saline
554 (TBS, pH 8.0) at room temperature. Excess BfaGC was removed by size-exclusion chromatography. The
555 2C12 TCR and mCD1d–BfaGC were mixed in a 1:1 molar ratio and incubated overnight at 4°C in TBS.
556 The formed 2C12 TCR–mCD1d–BfaGC ternary complexes were purified by size-exclusion
557 chromatography on a Superdex S200 10/300 column (GE Health Sciences).
558 *Crystallization and structure determination of ternary complexes.* The generated ternary complexes were
559 concentrated to 5 mg/mL, and crystallization was performed by a hanging drop diffusion method. The
560 crystals grew in 12–16% PEG 3350, 8% Tacsimate (pH 5.0) and were flash-frozen, with 10% (v/v) glycerol
561 as a cryoprotectant. Data were collected on the MX2 beamline at the Australian Synchrotron. The data
562 sets collected were processed with XDS [49] and scaled with SCALA [50]. The structure solution was obtained
563 by molecular replacement with the Phaser-MR program [51] in the Phenix suite, with two separate
564 ensembles: TCR and CD1d without the lipid from the search model of 2C12 TCR–mCD1d–KRN7000
565 complex (PDB code: 6BNK) [23]. The Fo-Fc map was unambiguous, revealing clear electron density for the

566 bound BfaGCs and the 2C12 TCR–CD1d interface. After initial rigid-body refinement, model building
567 was performed with COOT [52]. Subsequent iterative rounds of refinement were performed with BUSTER
568 [53]. The quality of the structures was assessed with the help of the RCSB Protein Data Bank (PDB)
569 validation server. PyMOL molecular visualization software was used for generating all graphic
570 representations (PyMOL Molecular Graphics System, version 2.0, distributed by Schrödinger,;
571 http://www.pymol.org).

572 Surface plasmon resonance analysis

573 All steady-state SPR experiments were performed on a Biacore 3000 instrument (GE Health Sciences) in
574 the presence of a buffer containing 10 mM HEPES-HCl (pH 7.4) and 150 mM NaCl at 25°C. Each
575 mCD1d–BfaGC complex were immobilized on a streptavidin-coated chip to a density of ~2500 response
576 units. The 2C12 TCR was passed over all the flow cells in increasing concentrations ranging from 0
577 through 5 μM at a flow rate of 5 μL/min. The final response was calculated by subtraction of response
578 from mCD1d-endogenous lipid. The equilibrium dissociation constant (K_D) was determined with
579 BIAevaluation software (version 3.1) in the 1:1 Langmuir model. The curve fitting and scientific graphics
580 were drawn with Graphpad prism software.

581 *B. fragilis* and BfaGC profiling of human microbiota-associated (HMB) mice

582 To assess *B. fragilis* change in adult mouse colon, *B. fragilis*-permissive HMB mice (N=5) were
583 preconditioned with antibiotics cocktail (0.5 g/L vancomycin, 0.5 g/L gentamycin, 0.25 g/L metronidazole
584 benzoate and 1 g/L ampicillin in 0.25% DMSO) in drinking water for 24 hours *ad libitum*. After the
585 period, drinking water was exchanged with normal water and animals were kept for 8 hours. 10^9 CFU *B.*
586 *fragilis* in 100 μL culture broth was gavaged individually. Stool samples were collected at 2, 3, 7 days

after gavage. In parallel, HMB mice at neonatal stage (postnatal day 14) were sacrificed and colonic contents were collected.

Bacterial DNA from fecal samples were extracted using DNA QIAamp PowerFecal Pro DNA Kit (Qiagen) according to the manufacture's instructions. qPCR was conducted using CFX-96 Real Time System (Bio-Rad) and SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Abundance of *B. fragilis* was calculated by comparing ratio of amplicons generated from *B. fragilis* specific primers (Leu-3 and Leu-4 [54]) and from pan-bacteria primers (BACT1369F and PROK1492R [55]). *B. fragilis*-specific / pan-bacteria amplicon ratio (ΔCq) of *B. fragilis* monocolonized mice stool samples (average of n=5) was considered as 100% abundance of *B. fragilis*, with which *B. fragilis* abundance of each sample (ΔΔCq) was normalized and converted. Relative abundance of C17/C17 BfaGC was quantified by PRM, normalized with internal standard abundance and sample weight.

Statistics and Reproducibility

All statistical analyses were carried out with Prism software (Graphpad, version 8 or later). Horizontal lines in dot plots represent mean values, except Figure 3C in which horizontal line represent median value, and each dot represents an individual datapoint. For *in vitro* and *ex vivo* experiments, results are shown as mean±SD. *In vivo* experiment results (Figures 2B-2D, 3B-3C and Extended Figures 2B, 8D) are shown as mean±SEM. To determine *P* values for two groups as specified in each figure legend, two-tailed, nonpaired Student's t test was used, except Figure 3C, where two-tailed Mann-Whitney test for non-paremetric values was used. One-way ANOVA was performed for comparisons of more than two groups (Figure 3B), as well as Dunnett's multiple comparison test with adjustment. A simple linear regression was used for Extended Figure 15A. Statistical significance is defined in the figures as follows: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$.

609 All data points presented in the figures are biological replicates in the same set, with exact numbers
610 indicated. Number of replication carried out for the each experiment is described individually. A DNA
611 gel image used in Extended Figure 8A is an uncropped, original result.

612

613 **Reference for Methods**

614 37. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of ROR + regulatory T
615 cells. *Science (80-.).* **349**, 993–997 (2015).

616 38. Varel, V. H. & Bryant, M. P. Nutritional features of Bacteroides fragilis subsp. fragilis. *Appl.*
617 *Microbiol.* **28**, 251–7 (1974).

618 39. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by
619 methyl- *tert* -butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).

620 40. Comstock, L. E. *et al.* Analysis of a capsular polysaccharide biosynthesis locus of Bacteroides
621 fragilis. *Infect. Immun.* **67**, 3525–32 (1999).

622 41. Lim, B., Zimmermann, M., Barry, N. A. & Goodman, A. L. Engineered Regulatory Systems
623 Modulate Gene Expression of Human Commensals in the Gut. *Cell* **169**, 547-558.e15 (2017).

624 42. Olszak, T. *et al.* Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T
625 Cell Function. *Science (80-.).* **336**, 489–493 (2012).

626 43. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

627 44. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning
628 sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

629 45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
630 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

631 46. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2016). Available at:
632 https://ggplot2.tidyverse.org/. (Accessed: 9th March 2021)

47. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2016). doi:10.1101/060012

48. Matsuda, J. L. *et al.* Tracking the response of natural killer T cells to a glycolipid antigen using CD1d tetramers. *J. Exp. Med.* **192**, 741–54 (2000).

49. Kabsch, W. 1 XDS. *Acta Crystallogr. Sect. D* **66**, 125–132 (2010).

50. Evans, P. Scaling and assessment of data quality. in *Acta Crystallographica Section D: Biological Crystallography* **62**, 72–82 (2006).

51. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

53. Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., P. W. & Roversi P, Sharff A., Smart O.S., Vonrhein C., W. T. O. BUSTER. (2017).

54. J, T., C, L., P, S., H, X. & SM, F. Application of quantitative real-time PCR for rapid identification of Bacteroides fragilis group and related organisms in human wound samples. *Anaerobe* **17**, 64–68 (2011).

55. Suzuki, M. T., Taylor, L. T. & DeLong, E. F. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5′-nuclease assays. *Appl. Environ. Microbiol.* **66**, 4605–4614 (2000).

ACKNOWLEDGEMENTS

We thank J. McCoy and E. J. Paik for manuscript preparation, R. T. Bronson for histopathological scoring and S. Iyer, L. Gebremedhin, E. Choi and T. Yanostang for technical assistance. We also thank the staff at the Australian Synchrotron for assistance with data collection. This work was supported by National Institutes of Health (K01-DK102771 and R01-AT010268, S.F.O.; and R01-DK044319, R.S.B.), Department of Defense (W81XWH-19-1-0625, D.L.K.), Brigham and Women's Hospital (Department of Anesthesiology, Perioperative and Pain Medicine Basic Science Grant, S.F.O.), the National Research Foundation of Korea (2014R1A3A2030423 and 2012M3A9C4048780, S.B.P.), and the Australian Research Council (ARC) (CE140100011 and ARC Laureate Fellowship, J.R.; and ARC Future Fellowship, J.L.N.). Graphical images used for the Figure 2 were created with BioRender.com.

AUTHOR CONTRIBUTIONS

S.F.O., D.L.K., and R.S.B. conceived the idea and designed the outline of the research.

S.F.O., H.S., and S.B.P. designed the structures of synthetic BfaGCs; H.S., Y.S.H., H.K., and J. L. synthesized BfaGC molecules.

T.P., J.L.N. and J.R. generated crystals of 2C12 TCR–CD1d–BfaGCs and carried out X-ray crystallography analysis as well as affinity measurements by SPR.

S.F.O., J-S.Y., and C.C.L. designed and carried out all experiments with microbes.

S.F.O., D-J.J., and D.E.H. executed *in vitro*/*in vivo* cytokine assays.

S.F.O. and D-J.J. designed and carried out all animal experiments. J-S. Y. carried out transcriptomic analysis.

S.F.O., S.B.P., J.R. and D.L.K. wrote the manuscript, and all authors contributed to relevant discussion.

DECLARATION OF INTERESTS

S.F.O., R.S.B., and D.L.K. filed a patent on the functions of BfaGCs and related structures (US patent 10,329,315).

S.F.O., S.B.P., and D.L.K. filed a patent on the functions of BfaGCs and related structures (under review).

EXTENDED DATA

Extended figures and figure legends 1-10

CORRESPONDENCE AND REQUEST FOR MATERIALS should be addressed to D.L.K., S.B.P., J.R. or S.F.O.

DATA AVAILABILITY STATEMENT

RNA sequencing results. Raw data for NKT cell transcriptomic analysis was deposited in the NCBI Sequence Read Archive (SRA) with Project number PRJNA750126.

PDB accession codes. The crystal structures of the 2C12 TCR–mCD1d–SB2217/SB2219 ternary complexes were deposited in the PDB with the accession numbers 7M72 and 6XNG, respectively.

Lipidomics data. Lipidomic analysis data containing MS1 scan was deposited to Metabolomics Workbench study number ST001910.

The authors declare that all other data supporting the findings of this study are available within the paper [and its supplementary information files].

690 **Extended Figure Legends**

691 **Extended Figure 1. Molecular structures of prototypic NKT agonist KRN7000, OCH and a**
692 **representative *B. fragilis*-derived aGC (SB2217).**

693 **Extended Figure 2. LC-MS profile of BfaGCs.** (A) Representative extracted ion chromatograms (XICs)
694 of C_{32}–C_{36} BfaGCs. (B) C_{34} BfaGCs are the major component of *B. fragilis* glycosphingolipids (N=5).

695 **Extended Figure 3. LC-MS/MS assignment of C_{34} BfaGC structural variants.** (A) The MS/MS-XIC of
696 total C_{34} BfaGCs (762→698) shows that BfaGCs are isobaric mixtures separated by RP-HPLC. (B, C)
697 MS/MS-XICs of C_{34} BfaGCs reveal co-eluting chemical homologues. Two isobaric species with aliphatic
698 chains of C_{17}/C_{17} (B) and C_{18}/C_{16} (C) were assigned MS/MS fingerprints of 490 and 504, respectively. (D)
699 MS/MS fingerprints of three peaks show a distinct difference in relative intensity between MS/MS
700 fragments of 490 (C_{17}/C_{17}) and 504 (C_{18}/C_{16}), implying that the latter two peaks are a mixture of chain-
701 length homologues. Chromatograms and spectra represent triplicate observations. (E-H) MS/MS spectra
702 of the most abundant peaks of (E) C_{32}, (F) C_{33}, (G) C_{35} and (H) C_{36} BfaGCs. MS/MS fingerprint of 462–518
703 indicates lengths of sphinganine and acyl chains. Spectra are representative of triplicate observation.

704 **Extended Figure 4. Chemical structures of 23 synthetic BfaGCs. (SB2201–SB2223)**

705 **Extended Figure 5. BCAA dictates branching of BfaGCs by direct incorporation *in vivo*.** (A–E) Ratios
706 among differently branched C_{34} BfaGCs (MS1 XIC=762.57, as $[M+HCOO^-]$) are clearly different for *B.*
707 *fragilis* grown in rich medium (A) and *B. fragilis* grown in minimal medium (B). Supplementation with
708 individual BCAAs (C–E) on defined medium increases production of branched-chain (both dibranched
709 and monobranched) BfaGCs. (F–H) MS/MS fingerprints confirm the incorporation of leucine and
710 isoleucine into the C_{17}/C_{17} ceramide backbone (via C5 branched acyl-CoA) and of valine into the C_{18}/C_{16}
711 backbone (via C4 branched acyl-CoA). Chromatograms and spectra are representative of triplicate

observations. (I) An MS/MS-XIC of d_3- and d_6-C_{34} BfaGC shows that deuterium-labeled leucine is actively incorporated into BfaGC. (J-K) MS/MS pattern shows distinctive differences between gut luminal BfaGC (M+3 isotopolog) in (J) presence or (K) absence of d3-leucine, showing MS2 fragments in presence of d3-leucine reflect inclusion of deuterium-labeled leucine in the structure. Chromatograms and spectra are representative results of four mice.

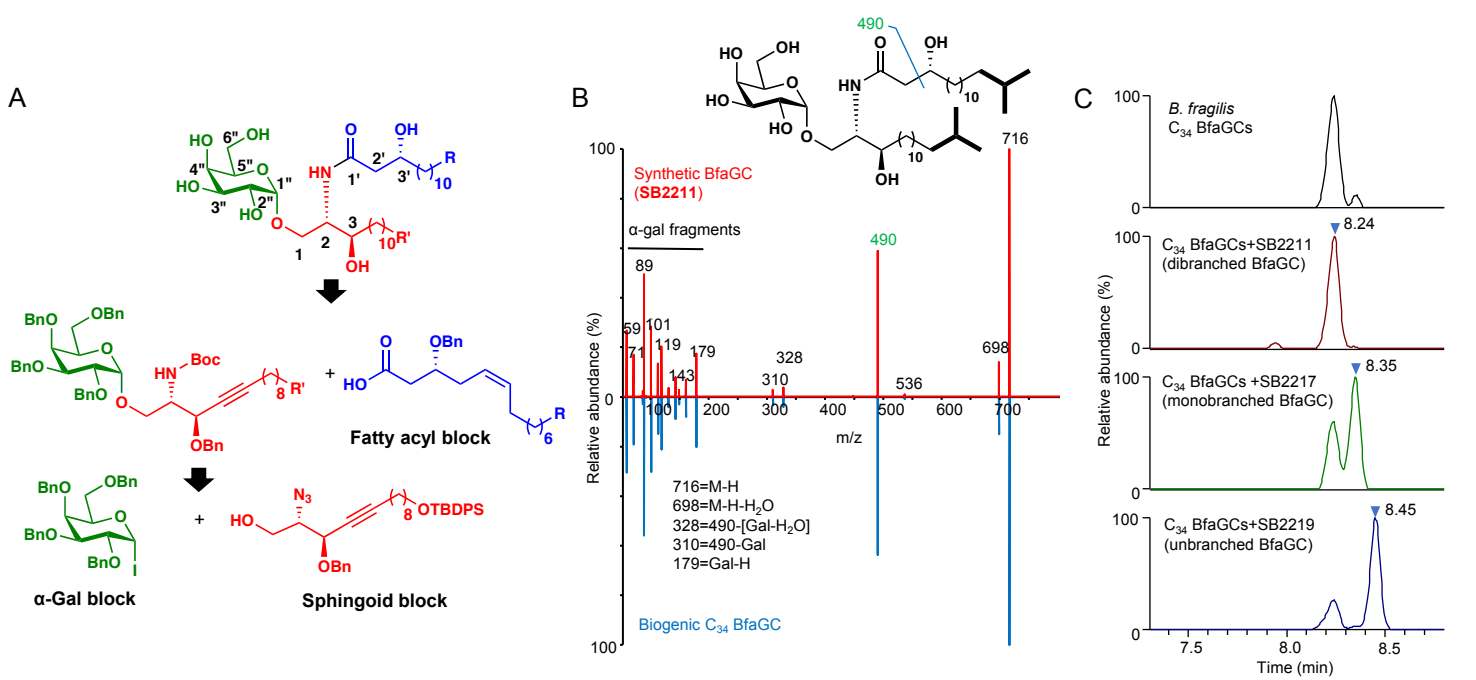
Extended Figure 6. Genetic study of *B. fragilis Bcat* orthologue (BF9343-3671). (A) Confirmation of the target gene deletion by PCR. (B) The knockout strain (BF9343-Δ3671) shows comparable growth pattern to isogenic WT strain (grown in duplicate per group), and a complemented strain of KO strain with empty vector shows same pattern to BF9343-3671 complemented strain. (C) BF9343-Δ3671 complementation can recover the production of di-branched C17/C17 BfaGC production to wild-type level. (D) WT and mutant strain (N=5 for each group) can colonize mouse in comparable density. All results represent of two independent experiments with similar trend. For gel source data, see Supplementary Figure 1.

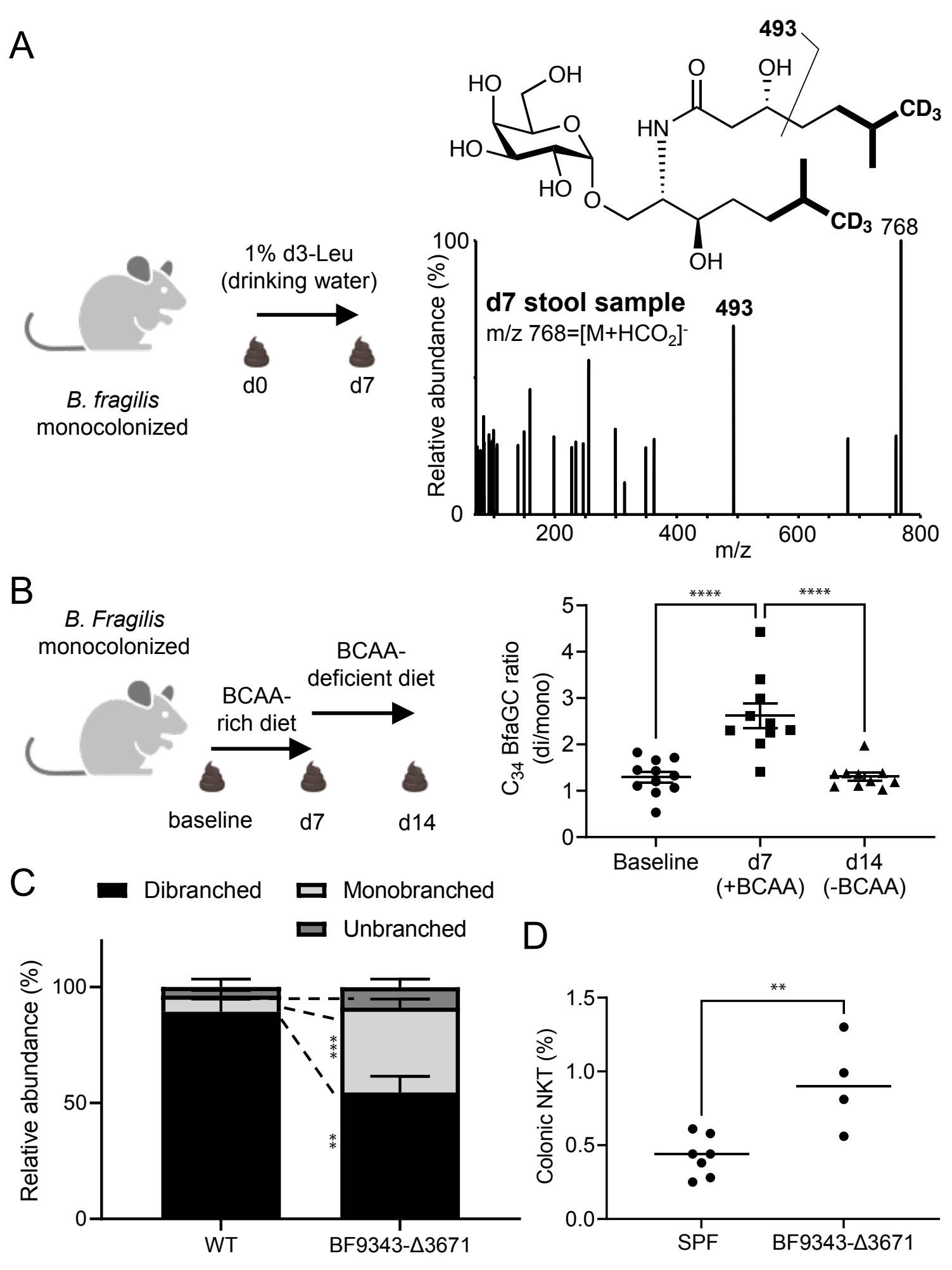
Extended Figure 7. Structure-specific actions of BfaGCs. (A) NKT cell–APC co-culture assays show that branching of sphinganine chain is, but 3'-OH group is not, critical for IL-2 inducing activity. Results are shown in duplicate and represent three independent experiment sets with similar trend ($p=0.017$ for 100nM and $p=0.026$ for 1000nM). (B-C) When injected intraperitoneally (N=5 per group, one sample in OCH group in panel C was lost), unlike Th1- or Th2-skewed prototypic ligands such as KRN7000 or OCH, SB2217 only weakly induce IFN-γ and did not induce IL-4 *in vivo*. (D-F) SB2217 weakly induced expression of co-stimulatory molecules such as CD86, CD40 and CD80 in splenic DCs, where SB2219 did not (N=5 per group).

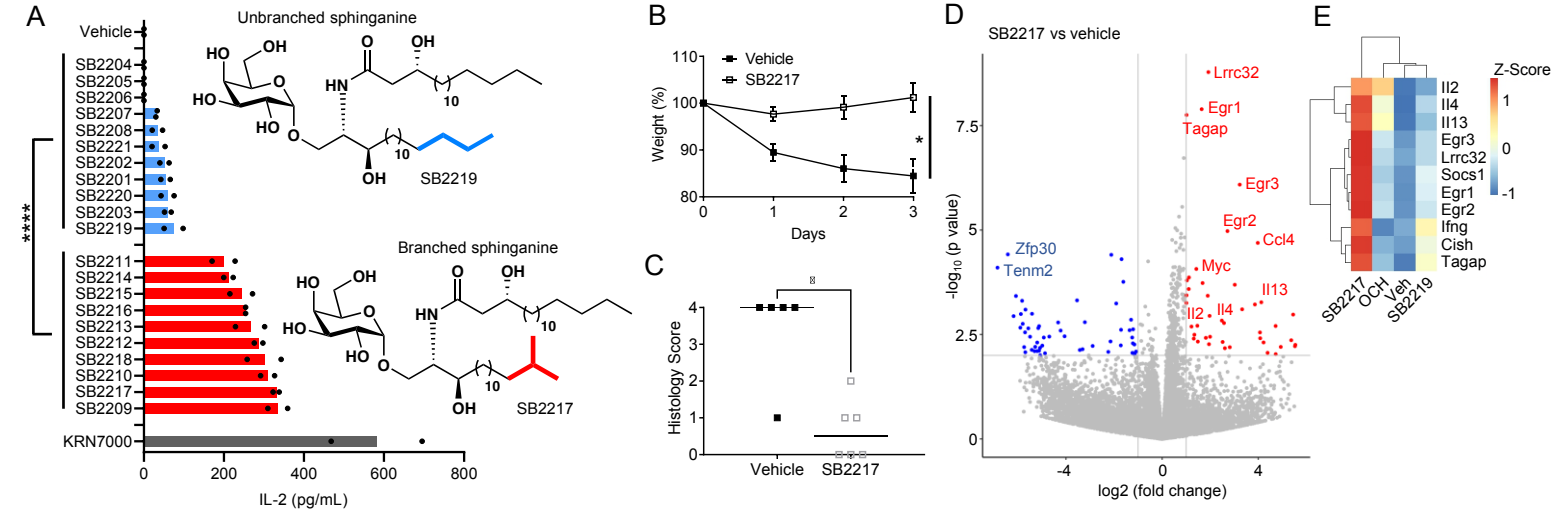
Extended Figure 8. Transcriptomic landscape of splenic NKT cells in responses to agonists. (A) A heatmap shown with the Euclidean distances between different treatment groups. (B) Transcriptomic profile comparison of SB2217, SB2219 and OCH. (C) Pathway enrichment analysis of SB2217 reveals increased expression of immunoregulatory pathways in NKT cells when compared to vehicle or SB2219.

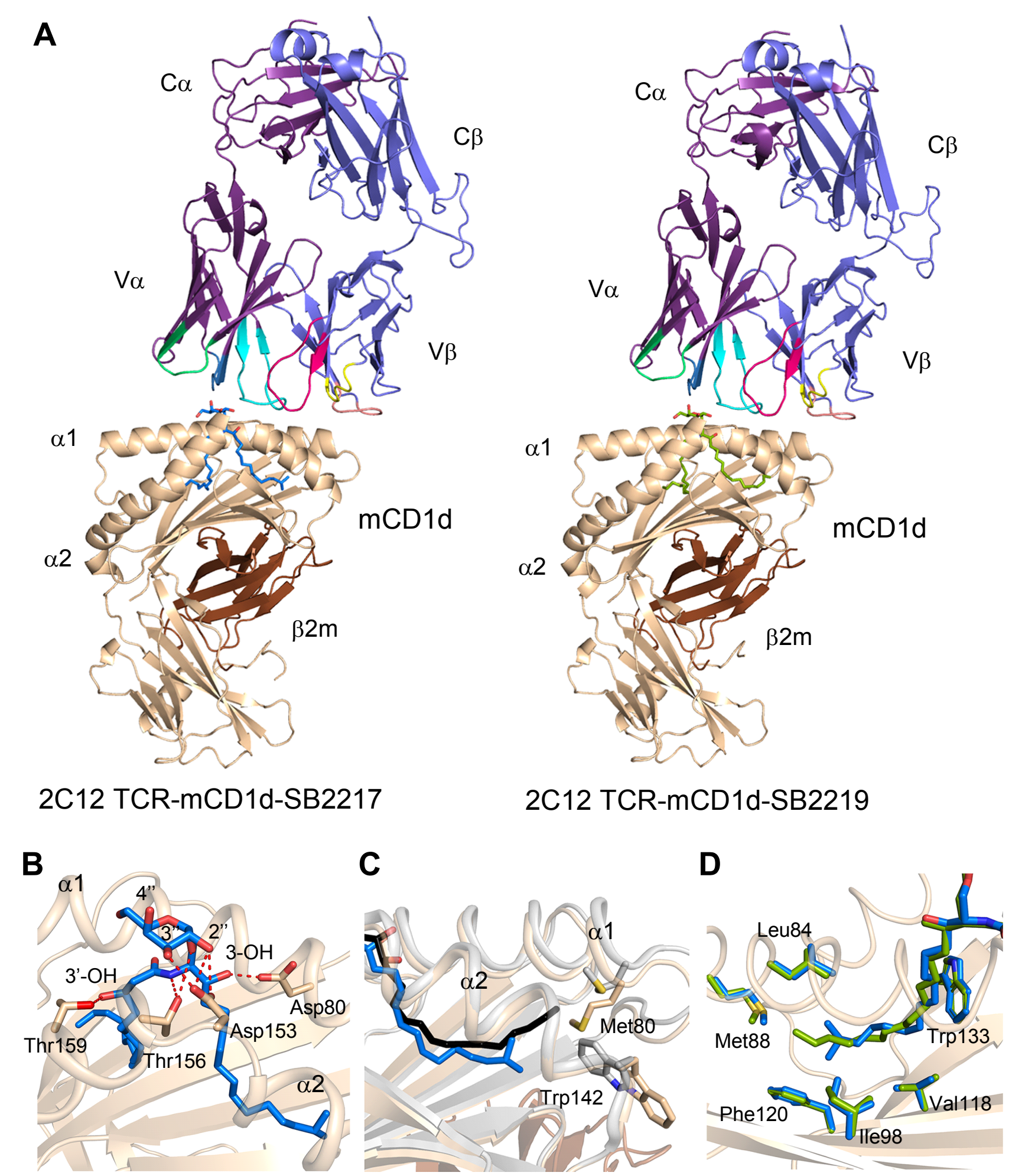
Extended Figure 9. Comparison between SB2217 and SB2219 in mCD1d-BfaGC-2C12 complexes. (A) 2Fo-Fc electron density map contoured at a 0.8σ level of the BfaGCs within each ternary complex. (B) Fo-Fc electron density map (in brown) contoured at a 2.2σ level of the BfaGCs and spacer lipids within each ternary complex. SB2217 is shown as blue and SB2219 is shown as green; Spacer lipids are shown as black sticks. (C) Superimposition of the headgroups of BfaGCs and KRN7000 (PDB code: 6BNK). (D) 2C12 TCR molecular interactions with SB2217 (in blue). mCD1d and CDR loops are colored as in Figure 4A. Hydrogen bonds are shown as red dashed lines. (E-F) The mCD1d–SB2217 complex shows higher affinity to 2C12 TCR than the mCD1d–SB2219 complex. (E) Each SPR datapoint is mean of techincal duplicate and K_D values (mean±SD) were calculated from two independent results, using a single-site binding model with K_D as a shared variable. (F) The sensorgrams are results of single experiment.

Extended Figure 10. BfaGC profile in human microbiota-associated mice. (A) BfaGC and *B. fragilis* abundance shows positive correlation in *B. fragilis*-gavaged HMB mice. Results are from longitudinally collected samples (2, 3 and 7 days after *B. fragilis* oral introduction) from five mice (total N=15). (B) BfaGC (C_{17}/C_{17} dibranched and monobranched) are identified from neonatal (p14) GI contents. Chromatogram and spectrum represent seven samples.


A
B
C
α-Gal block
Sphingoid block
Fatty acyl block
Synthetic BfaGC
(SB2211)
α-gal fragments
716=M-H
698=M-H-H2O
328=490-[Gal-H2O]
310=490-Gal
179=Gal-H
Biogenic C34 BfaGC
Relative abundance (%)
m/z
B. fragilis
C34 BfaGCs
C34 BfaGCs+SB2211
(dibranched BfaGC)
C34 BfaGCs +SB2217
(monobranched BfaGC)
C34 BfaGCs+SB2219
(unbranched BfaGC)
8.24
8.35
8.45
Time (min)


A
493
HO
OH
O
HO
HO
HO
O
HN
O
OH
CD3
CD3
768
OH
1% d3-Leu
(drinking water)
d0
d7
B. fragilis
monocolonized
100
Relative abundance (%)
d7 stool sample
493
m/z 768=[M+HCO2]-
0
200
400
600
800
m/z
B
B. Fragilis
monocolonized
BCAA-
rich diet
BCAA-
deficient diet
baseline
d7
d14
****
****
5
4
3
2
1
0
C34 BfaGC ratio
(di/mono)
Baseline
d7
(+BCAA)
d14
(-BCAA)
C
Dibranched
Monobranched
Unbranched
100
50
0
Relative abundance (%)
***
**
WT
BF9343-Δ3671
D
**
1.5
1.0
0.5
0.0
Colonic NKT (%)
SPF
BF9343-Δ3671


A
Vehicle
SB2204
SB2205
SB2206
SB2207
SB2208
SB2221
SB2202
SB2201
SB2220
SB2203
SB2219
****
SB2211
SB2214
SB2215
SB2216
SB2213
SB2212
SB2218
SB2210
SB2217
SB2209
KRN7000
0
200
400
600
800
IL-2 (pg/mL)
Unbranched sphinganine
SB2219
Branched sphinganine
SB2217
B
Vehicle
SB2217
Weight (%)
110
100
90
80
0
1
2
3
Days
*
C
Histology Score
0
1
2
3
4
Vehicle
SB2217
D
SB2217 vs vehicle
-log10 (p value)
7.5
5
2.5
0
-4
0
4
log2 (fold change)
Lrrc32
Egr1
Tagap
Egr3
Egr2
Ccl4
Myc
Il13
Il4
Il2
Zfp30
Tenm2
E
Z-Score
1
0
-1
Il2
Il4
Il13
Egr3
Lrrc32
Socs1
Egr1
Egr2
Ifng
Cish
Tagap
SB2217
OCH
Veh
SB2219


A
Cα
Cβ
Vα
Vβ
α1
α2
mCD1d
β2m
2C12 TCR-mCD1d-SB2217
2C12 TCR-mCD1d-SB2219
B
α1
4''
3''
2''
3-OH
3'-OH
Asp80
Asp153
Thr159
Thr156
α2
C
α1
α2
Met80
Trp142
D
Leu84
Met88
Trp133
Val118
Phe120
Ile98