

## **Structural basis for the ligand recognition and signaling of free fatty acid receptors**

Xuan Zhang<sup>1</sup>, Abdul-Akim Guseinov<sup>2+</sup>, Laura Jenkins<sup>3+</sup>, Kunpeng Li<sup>4</sup>, Irina G. Tikhonova<sup>2\*</sup>, Graeme Milligan<sup>3\*</sup>, Cheng Zhang<sup>1\*</sup>

<sup>1</sup>Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA15261, USA

<sup>2</sup>School of Pharmacy, Medical Biology Centre, Queen's University Belfast, Belfast BT9 7BL, Northern Ireland, United Kingdom

<sup>3</sup>Centre for Translational Pharmacology, School of Molecular Biosciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, United Kingdom

<sup>4</sup>Cryo-EM core facility, Case Western Reserve University, OH44106, USA.

<sup>+</sup>Contributed equally.

\*Corresponding authors: CZ ([chengzh@pitt.edu](mailto:chengzh@pitt.edu)), IGT ([i.tikhonova@qub.ac.uk](mailto:i.tikhonova@qub.ac.uk)) and GM ([graeme.milligan@glasgow.ac.uk](mailto:graeme.milligan@glasgow.ac.uk))

## Abstract

Free fatty acid receptors 1-4 (FFA1-4) are class A G protein-coupled receptors (GPCRs). FFA1-3 share substantial sequence similarity whereas FFA4 is unrelated. Despite this FFA1 and FFA4 are activated by the same range of long chain fatty acids (LCFAs) whilst FFA2 and FFA3 are instead activated by short chain fatty acids (SCFAs) generated by the intestinal microbiota. Each of FFA1, 2 and 4 are promising targets for novel drug development in metabolic and inflammatory conditions. To gain insights into the basis of ligand interactions with, and molecular mechanisms underlying activation of, FFAs by LCFAs and SCFAs, we determined the active structures of FFA1 and FFA4 bound to the polyunsaturated LCFA docosahexaenoic acid (DHA), FFA4 bound to the synthetic agonist TUG-891, as well as SCFA butyrate-bound FFA2, each complexed with an engineered heterotrimeric G<sub>q</sub> protein (miniG<sub>q</sub>), by cryo-electron microscopy. Together with computational simulations and mutagenesis studies, we elucidated the similarities and differences in the binding modes of fatty acid ligands with varying chain lengths to their respective GPCRs. Our findings unveil distinct mechanisms of receptor activation and G protein coupling. We anticipate that these outcomes will facilitate structure-based drug development and underpin future research to understand allosteric modulation and biased signaling of this group of GPCRs.

**(205 words)**



## Introduction

Free fatty acids are bioactive lipids comprising a carboxylic acid head group and an aliphatic hydrocarbon chain with various lengths. In humans, and many other species, they can activate a group of G protein-coupled receptors (GPCRs) including free fatty acid receptors 1-4 (FFA1-4 receptors) and GPR84 to regulate metabolic homeostasis and immunity<sup>1-3</sup>. Among them, FFA1 (GPR40) and FFA4 (GPR120) mainly sense long-chain fatty acids (LCFAs) with more than 12 carbons while FFA2 (GPR43) and FFA3 (GPR41) primarily sense short-chain fatty acids (SCFAs) with less than 6 carbons<sup>3,4</sup>. Representative LCFA ligands of FFA1 and FFA4 include  $\omega$ -3 and  $\omega$ -6 polyunsaturated fatty acids (PUFAs)<sup>5,6</sup>. Meanwhile, SCFA ligands of FFA2 and FFA3 are mainly produced in the gut as products of microbiota-mediated fermentation and include acetate, propionate, and butyrate<sup>7-12</sup>.

FFAs play critical roles in both immunity and metabolism<sup>3,4,10</sup>. FFA1 signaling induced by LCFAs in pancreatic  $\beta$  cells can facilitate insulin secretion<sup>13</sup>, making it a promising drug target for type 2 diabetes mellitus (T2D)<sup>14,15</sup>. Although the FFA1-selective agonist TAK-875, also named fasiglifam, exhibited promising antidiabetic effects in clinical studies, it failed in phase III trials due to liver toxicity<sup>16</sup>. However, other FFA1 activators are still being pursued for the treatment of T2D<sup>14</sup>. FFA4, which has been described as the  $\omega$ -3 PUFA receptor, is highly expressed in adipose tissue and macrophages. It mediates anti-inflammatory and other beneficial effects of  $\omega$ -3 PUFAs such as docosahexaenoic acid (DHA) in those tissues and cells<sup>3-5,17-21</sup>. FFA4 is also considered as a new drug target for diabetes<sup>21-23</sup>. FFA4 selective or FFA1/FFA4 dual agonists<sup>24-26</sup> may hold the promise of becoming a new class of antidiabetic drugs with additional anti-inflammatory benefits<sup>22,26</sup>. Interestingly, in addition to their functions in metabolism and immunity, both FFA1 and FFA4, particularly FFA4, have been suggested to function as lipid taste receptors<sup>27-29</sup>. On the other hand, FFA2 and FFA3 are expressed in adipocytes and a range of immune cells. Their unique ligand preference of SCFAs produced by the fermentation of dietary fiber in the lower gut has led to intensive research on their roles at the interface of host and gut microbiota<sup>30,31</sup>. Previous studies suggested that many beneficial effects of gut microbiota on the host, including the resolution of inflammation<sup>32</sup>, suppression of fat accumulation<sup>33</sup>, and protection from viral and bacterial pathogens<sup>34,35</sup>, are mainly mediated by the SCFA-FFA2 signaling axis. Therefore, FFA2 and FFA3, especially FFA2, are considered as new promising therapeutic targets for metabolic disorders, including obesity and diabetes, and inflammatory diseases<sup>3,10,36-38</sup>. However, compared to FFA1 and FFA4, fewer synthetic ligands have been reported for FFA2 and FFA3, which may suggest certain obstacles in developing small molecule ligands for these two SCFA receptors.

Phylogenetic analysis suggests that FFA4 diverged early from other FFAs (**Fig. S1**). As a result, while FFA1-3 are structurally related with high sequence similarity, FFA4 shares very little sequence similarity with FFA1-3<sup>39</sup>. This implies distinct ligand recognition and signaling mechanisms for FFA4 and other FFAs. Regarding G protein coupling, FFA1 is a highly promiscuous GPCR that is capable of coupling to all four G protein families:  $G_s$ ,  $G_{i/o}$ ,  $G_{q/11}$ , and  $G_{12/13}$ <sup>39,40</sup>. FFA2 and FFA4 both can signal through  $G_{i/o}$  and  $G_{q/11}$ <sup>10,39-41</sup>. For FFA4, a human splice variant (FFA4<sup>Long</sup>) has been identified with an additional 16-amino acid segment in the intracellular loop 3 (ICL3). This isoform is unable to induce  $G_{q/11}$  signaling but is capable of coupling to  $\beta$ -arrestins<sup>42,43</sup>. Crystal structures of highly engineered forms of FFA1 bound to synthetic agonists including TAK-875 and positive allosteric modulators (PAMs) have been

reported<sup>44-46</sup>, and the receptor in those structures stayed in the inactive state. To understand the molecular mechanisms by which LCFAs and SCFAs act on and activate their FFAs, we determined active structures of DHA-bound FFA1, butyrate-bound FFA2, and DHA-bound FFA4 in complex with an engineered heterotrimeric G<sub>q</sub> protein (miniG<sub>q</sub>)<sup>47</sup> by cryo-electron microscopy (cryo-EM). To further examine how FFA4 recognizes different ligands, we also solved a cryo-EM structure of the FFA4-miniG<sub>q</sub> complex with the most widely employed synthetic FFA4 agonist, TUG-891<sup>48</sup>. These structures revealed diverse modes of ligand recognition by FFAs. Together with computational simulations and mutagenesis studies, these studies highlight similarities and differences in modes of binding of the fatty acid ligands of varying chain length to their corresponding GPCRs.

## Results

### Overall structures of FFA signaling complexes

We used the wild type human FFA1, FFA2, and FFA4 in our structural studies. For FFA4, there are two isoforms and we chose the canonical short form since FFA4<sup>Long</sup> doesn't couple to G<sub>q/11</sub><sup>43</sup>. The miniG<sub>q</sub> protein contains an engineered miniG<sub>aq</sub> subunit<sup>49</sup> with the N-terminal 35 amino acids replaced by their corresponding residues in G<sub>ai</sub>. The same miniG<sub>q</sub> protein has been successfully used to obtain cryo-EM structures of several other G<sub>q</sub>-coupled GPCRs<sup>47,50-53</sup>. To further stabilize the FFA-miniG<sub>q</sub> complexes, we assembled the complexes using the NanoBit tethering strategy in insect Sf9 cells<sup>54</sup> together with an antibody fragment, scFv16, which has been developed previously to stabilize the G<sub>i</sub> heterotrimer<sup>55</sup>. The structures of DHA-bound FFA1 and FFA4, TUG-891-bound FFA4, and butyrate-bound FFA2 with miniG<sub>q</sub> were determined to overall resolutions of 3.4 Å, 3.2 Å, 3.1 Å, and 3.1 Å, respectively (**Fig. 1a**, **Fig. S2-5**, **Table S1-2**).

The majority of residues in the three FFAs, miniG<sub>q</sub>, and scFv16 were modeled based on the robust cryo-EM density maps. We also modeled several cholesterol and lipid molecules to fit strong density maps surrounding the transmembrane domains (TMDs) of FFA2 and FFA4 (**Fig. S6**). In the two structures of FFA4, the density of the intracellular region of transmembrane helix 4 (TM4) and intracellular loop 2 (ICL2) is relatively weak, indicating a high degree of flexibility (**Fig. 1b**). In contrast, ICL2 forms a helical structure in both FFA1 and FFA2 (**Fig. 1b**). A large part of the extracellular loop 2 (ECL2) of FFA2 is also not modeled due to weak density. Noticeably, in the structures of all three receptors, no helix 8 after TM7 was modeled due to very weak density. This suggests that the C-terminal region after TM7 in all three receptors is highly mobile when coupled with G proteins.

As for the ligands, the density maps for TUG-891 and DHA in FFA4 were sufficiently clear to enable modeling of the entire ligands (**Fig. S2-3**). Additionally, the density of butyrate in FFA2 was also strong (**Fig. S5**). However, due to the small size of butyrate and the limited resolution of the cryo-EM map, functional data was necessary to complement cryo-EM map information for accurate ligand modeling. In the case of FFA1, we modeled DHA in the orthosteric site based on a partial density map (**Fig. S4**). Further discussion on this topic will be provided in the subsequent content.

### Binding of DHA and the synthetic agonist TUG-891 to FFA4

Both DHA and TUG-891 bind to a pocket formed among the extracellular regions of TM3-7 of FFA4 (**Fig. 2a**). The amino-terminal (N-terminal) region of FFA4 preceding TM1 folds inside the TMD and

directly interacts with the ligands (**Fig. 2a**), resulting in almost complete shielding of the ligand-binding pocket from the extracellular milieu (**Fig. 2b**). This is similar to the N-terminal region of DP2, a GPCR that binds the fatty acid ligand prostaglandin D<sub>2</sub> (PGD<sub>2</sub>), and that forms a well-folded structure that participates in ligand interactions <sup>56,57</sup>.

The cryo-EM density is strong for the  $\omega$ -3 unsaturated chain of DHA, while the density for the carboxylate group is comparatively weaker. Nevertheless, we were able to model the entire DHA molecule that well fits the density. Our structure revealed that the  $\omega$ -3 unsaturated chain of DHA adopts a 'L' shape binding pose, which enables it to penetrate deeply into a binding pocket that is rich in aromatic residues (**Fig. 2a**). The carboxylate head group of DHA, on the other hand, extends outward towards the extracellular milieu (**Fig. 2a**). The six carbon-carbon double bonds in DHA are surrounded by, and potentially form, extensive  $\pi$ - $\pi$  interactions with aromatic residues F27<sup>N</sup>, F28<sup>N</sup>, F88<sup>2.53</sup>, F115<sup>3.29</sup>, W198<sup>ECL2</sup>, W207<sup>5.38</sup>, F211<sup>5.42</sup>, W277<sup>6.48</sup>, and F303<sup>7.35</sup> (superscripts represent Ballesteros-Weinstein numbering <sup>58</sup>) (**Fig. 2c**). Hydrophobic residues I280<sup>6.51</sup>, I284<sup>6.55</sup>, I287<sup>6.58</sup>, and L288<sup>6.59</sup> lining one side of TM6, together with M118<sup>3.32</sup>, L196<sup>ECL2</sup>, and I287<sup>6.58</sup> form additional hydrophobic interactions with DHA to further stabilize lipid binding (**Fig. 2c**).

TUG-891 (3-(4-((4-fluoro-4-methyl-[1,1-biphenyl]-2-yl)methoxy)phenyl)-propanoic acid) binds to FFA4 in a similar 'L' shape binding pose as DHA, overlapping extensively with the  $\omega$ -3 chain of DHA and with the ortho-biphenyl moiety of TUG-891 defining the bottom of the binding pocket (**Fig. 2a**). This leads to the observation of similar sets of hydrophobic and  $\pi$ - $\pi$  interactions between the three benzene rings of TUG-891 and FFA4 (**Fig. 2d**). Mutagenesis studies we performed previously are in accord with these observations. Alteration of F88<sup>2.53</sup>, F115<sup>3.29</sup>, W207<sup>5.38</sup>, F211<sup>5.42</sup>, W277<sup>6.48</sup> (each to A) and F303<sup>7.35</sup> (to H) (and also here to A, **Fig. 2e**) resulted in either a complete lack of response to TUG-891 or a greater than 100-fold reduction in potency <sup>59</sup>. The mutation W198A also lacked response to both TUG-891 and DHA (**Fig. 2e, Fig. S7**). However, although mutations to Ala of F27<sup>N</sup> and F28<sup>N</sup> each lacked response to TUG-891 (**Fig. 2e, Fig. S7**), these results could not be interpreted because, although well expressed following transient transfection in HEK293 cells (**Fig. S7c**), each of these mutants failed to reach the cell surface (**Fig. S7d**). Mutation to Ala of either I280<sup>6.51</sup> or I284<sup>6.55</sup> also produced a greater than 100-fold (I280A) <sup>59</sup> or a more modest but still significant (I284A) reduction in potency for TUG-891 (**Fig. 2e, Fig. S7a**). We previously observed a similar pattern of effects of these mutations for the  $\omega$ -3 PUFA  $\alpha$ -linolenic acid (*all-cis*-9,12,15-octadecatrienoic acid) <sup>59</sup>. Although not altering the potency of DHA (**Fig. 2e, Fig. S7a**), a notable feature of the I284A mutant was that it reduced the efficacy of DHA, such that it acted as a partial agonist compared to TUG-891 at this mutant (**Fig. S7b**), a feature that was not observed for the other mutants studied.

An additional difference in the binding mode of TUG-891 and DHA is that the linking ether oxygen of TUG-891 forms a hydrogen bond with T119<sup>3.33</sup>, which is absent in the DHA-bound FFA4 (**Fig. 2c and d**). Notably, mutation of T119A significantly impaired the ability of TUG-891 to activate FFA4, in both  $\beta$ -arrestin interaction (**Fig. S7a**) and, particularly, G<sub>q</sub>-mediated Ca<sup>2+</sup> elevation assays <sup>59</sup>, indicating an important role of this hydrogen bond in TUG-891 binding and function. No such effects of this mutation on the function of either DHA (**Fig. 2e, Fig. S7a**) or  $\alpha$ -linolenic acid <sup>59</sup> was observed. A distinct group of sulphonamide-based FFA4 agonists have been reported <sup>39,60</sup>. Among them, TUG-1197 (2-(3-(pyridin-

2-yloxy)phenyl)-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide)) showed a large, greater than 10-fold, reduction in potency at the T119A mutant and greatly reduced efficacy in comparison to TUG-891 (**Fig. 2e, Fig. S7a-b**). Docking of this ligand to the obtained structures of FFA4 suggests a similar binding pose as TUG-891 and a clear interaction of the sulphonamide, which overlaps location with the ether oxygen of TUG-891, with T119 (**Fig. 2d**).

Intriguingly, in the structure, the carboxylate group of the phenyl-propanoic acid of TUG-891 is positioned in proximity to E204<sup>5.35</sup> and W198<sup>ECL2</sup> (**Fig. 2d**). Remarkably, the cryo-EM density between E204<sup>5.35</sup> and W198<sup>ECL2</sup> appears to be continuous (**Fig. 2f**), which raises the possibility that a water molecule may be present between these two residues to facilitate extensive polar interactions between the carboxylate group of TUG-891 and E204<sup>5.35</sup> and W198<sup>ECL2</sup> of FFA4. The mutant E204A modestly but significantly reduced potency of TUG-891 but not of DHA (**Fig. S7a**) whereas a W198A mutant was not activated by either TUG-891 or DHA (**Fig. 2e, Fig. S7a**).

The carboxylate group of DHA, which is associated with weak cryo-EM density, is modeled close to polar residues R22<sup>N</sup> and R24<sup>N</sup> from the N-terminal region and N291<sup>ECL3</sup> from ECL3 (**Fig. 2c**). However, single mutations R22A and R24A or the double mutation R22A/R24A did not reduce the potency of DHA or either of the synthetic agonists TUG-891 and TUG-1197 (**Fig. 2e, Fig. S7a**). Therefore, the binding of DHA to FFA4 is mainly driven by hydrophobic and  $\pi$ - $\pi$  interactions. Nevertheless, it is to be noted that the overall binding pocket of DHA exhibits a negatively charged potential, which may help to position the carboxylate group of DHA at the extracellular surface (**Fig. 2g**). A similar charge interaction-facilitated lipid recognition mechanism has also been suggested for other lipid GPCRs including prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) receptors and lysophospholipid GPCRs<sup>56,57</sup>. However, although DHA adopts a similar binding orientation as PGE<sub>2</sub> and lysophospholipids, their binding sites are located differently (**Fig. S8a**). The pockets of PGE<sub>2</sub> and lysophospholipids form among TM1-TM2-TM3-TM7 or TM2-TM3-TM5-TM6-TM7, while the pocket of DHA in FFA4 forms among TM3-TM4-TM5-TM6-TM7 (**Fig. S8a**). To the best of our knowledge, no other lipid GPCRs have been shown to have lipid binding pockets at similar locations as that of DHA in FFA4.

During the preparation of our manuscript, other research groups published structures of FFA4 bound to several LCFAs, including eicosapentaenoic acid (EPA), an  $\omega$ -3 PUFA, and TUG-891<sup>61</sup>. The structure of EPA bound to FFA4 showed a highly similar binding pose to that of DHA observed in our structure, especially with regard to their  $\omega$ -3 chains (**Fig. 2h**). However, DHA is two carbons longer than EPA and contains an additional double bond. As a consequence, the carboxylate group of DHA extends further towards the extracellular surface above E204<sup>5.35</sup>, while the carboxylate group of EPA is located below E204<sup>5.35</sup> (**Fig. 2h**). Furthermore, we observed a slightly different binding mode of TUG-891 in our cryo-EM structure, which is strongly supported by clear cryo-EM density, compared to the published structure (**Fig. S8b**). The overall position of TUG-891 in the published structure is closer to the extracellular surface compared to that in our structure (**Fig. S8b**). As a result, the carboxylate group of TUG-891 in our structure is too distant from N291<sup>ECL3</sup> to form a hydrogen bond. Consistent with this, we did not observe an effect of the N291A mutant on the potency of TUG-891 (**Fig. 2e, Fig. S7a**) and such a mutant was not reported in the published study<sup>61</sup>. In addition, the hydrogen bond between the ether oxygen of TUG-891 and T119<sup>3.33</sup> in our structure is absent in the published structure (**Fig. S8b**). The discrepancies



in TUG-891 binding in the two structures may indicate a high degree of mobility of TUG-891 in FFA4. Nevertheless, despite these discrepancies, our mutagenesis studies demonstrated the important role of T119 in the action of TUG-891 (**Figs. S7a, S8b**) and the positioning of the ortho-biphenyl, which is key feature of TUG-891 and related synthetic FFA4 agonists, is entirely in accord with our earlier mutagenesis studies <sup>59</sup>.

### **Distinct mechanisms of DHA recognition by FFA1 and FFA4**

Despite their similar ligand recognition profiles as LCFA receptors, FFA1 and FFA4 exhibit little sequence similarity and a distant phylogenetic relationship (**Fig. S1**). To investigate whether FFA1 utilizes a distinct mechanism to recognize DHA, we sought to obtain a cryo-EM structure of miniGq-coupled FFA1 bound to DHA. However, while the overall resolution of the structure reached 3.4 Å (**Fig. S4**), modeling DHA proved to be challenging.

Previous structural studies on FFA1 with synthetic agonists and allosteric modulators <sup>44-46</sup> have identified two binding sites: 'Site 1' is located in the extracellular region within the 7TM as the putative orthosteric site for synthetic agonists TAK-875 <sup>46</sup> and MK-8666 <sup>44</sup>, while 'Site 2' is located on the surface of the 7TM above ICL2 as the allosteric site for the synthetic ago-positive allosteric modulators (ago-PAMs) AP8 <sup>44</sup> and Compound 1 (3-benzyl-4-(cyclopropyl-(4-(2,5-dichlorophenyl)thiazol-2-yl)amino)-4-oxobutanoic acid) <sup>45</sup>. Similar allosteric sites have also been identified for C5a receptor <sup>62</sup> and β2-adrenergic receptor <sup>63</sup>. Evidence from a previous study showed that TAK-875 exhibited positive cooperativity with the LCFA ligand of FFA1 γ-linolenic acid (γ-LA), suggesting that 'Site 1' is not the primary site for γ-LA <sup>64</sup>.

In our structure, we observed weak density in 'Site 1' for DHA (**Fig. 3a**). We further performed local refinement focusing on the receptor to improve this density, which allowed us to model the part of DHA from C1 to C8 together with the carboxylate group in 'Site 1' (**Fig. 3a**). The region of DHA from C9 to C22 was assigned with zero occupancy in the structure since it couldn't be well modeled. In our structural model, the carboxylate group of DHA forms salt bridges with R183<sup>5.39</sup> and R258<sup>7.35</sup>, while the carbon chain from C1 to C8 forms hydrophobic and π-π interactions with surrounding residues F87<sup>3.33</sup>, L138<sup>4.57</sup>, F142<sup>4.61</sup>, W174<sup>ECL2</sup>, and L186<sup>5.42</sup> of FFA1 (**Fig. 3a**). A very recent study also reported a similar partial cryo-EM density for DHA in 'Site 1' of FFA1 <sup>65</sup>. It is to be noted that the density of most residues in 'Site 1' after local refinement was sufficiently clear to allow unambiguous modeling, suggesting that the relatively weaker density of DHA was likely due to the high flexibility of DHA in this site.

Interestingly, strong density in 'Site 2' indicated the presence of a lipid molecule (**Fig. 3b**). It is possible that 'Site 2' is another binding site for DHA in FFA1. However, this site can also accommodate other lipid molecules, making it challenging to confirm it as the specific binding site for DHA. Nonetheless, since the cryo-EM sample contained a high concentration of DHA, it was the most prevalent lipid present. Thus, we fitted the density observed with a DHA molecule (**Fig. 3b**), and the binding pose of DHA in this site highly resembles that of the ago-PAM AP8 <sup>44,45</sup>. The carboxylate group of DHA forms hydrogen bonds with two tyrosine residues Y44<sup>2.42</sup> and Y114<sup>ICL2</sup>, while the carbon chain forms hydrophobic interactions with hydrophobic residues from TM3-5 of FFA1 (**Fig. 3b**). Interestingly, we also observed

a hydrogen bond between DHA and Y243 from the  $\alpha 5$  of mini-G<sub>aq</sub>, which is the major receptor interaction site in mini-G<sub>aq</sub> (**Fig. 3b**).

In the bile acid receptor GPBAR, the endogenous lipid ligand bile acid binds to a site formed between TM3 and TM4 above ICL2, similar to 'Site 2' in FFA1<sup>66</sup>. However, bile acid can also bind to a more conventional orthosteric site located in the extracellular region of GPBAR<sup>66</sup>. It is possible that DHA binds to FFA1 similarly to bile acid in GPBAR. 'Site 1' in FFA1 serves as the primary site for DHA, where DHA exhibits a high flexibility, while 'Site 2' in FFA1 serves as the secondary site for DHA.

Neither 'Site 1' nor 'Site 2' in FFA1 is conserved in FFA4, providing further evidence of the distant phylogenetic relationship between FFA1 and FFA4, despite their similar ligand preferences. In FFA4, the carboxylate group of DHA is positioned near the ligand entrance at the extracellular surface (**Fig. 2c**). In our previous studies on DP2, we observed that the prostaglandin PGD<sub>2</sub> adopts a 'polar-group-in' binding pose in DP2 with its carboxylate group buried deep within the binding pocket, while another prostaglandin, PGE<sub>2</sub>, adopts a 'polar-group-out' binding pose in PGE<sub>2</sub> receptors (EPs) with its carboxylate group positioned near the extracellular surface, and these two different binding poses of prostaglandins are facilitated by the distinct charge potentials of the binding pockets in DP2 and EPs<sup>56,57</sup>. Similarly, DHA in FFA4 adopts a 'polar-group-out' binding pose in a negatively charged environment (**Fig. 2g**), although the role of the charge potential in DHA binding is not clear.

## Recognition of SCFAs by FFA2

As anticipated from the relatedness of FFA2 to FFA1<sup>67</sup>, the overall structures of these two receptors are similar. Also, butyrate and the carboxylate head group of DHA in 'Site 1' of FFA1 are very close if the structures of FFA1 and FFA2 are aligned (**Fig. 4a**). In the structure of FFA2-butyrate, the carboxylate group of butyrate is coordinated by a pair of adjacent arginine residues, R180<sup>5,39</sup> and R255<sup>7,35</sup> (**Fig. 4b**). These two arginine residues are highly conserved among FFA1-3<sup>67,68</sup>. Indeed, the same pair of residues in FFA1 interact with the carboxylate group of TAK-875<sup>46</sup>. In FFA2, mutating R180<sup>5,39</sup> to other amino acids including Ala and Lys, as well as mutating R255<sup>7,35</sup> to Ala, eliminates the response to SCFAs<sup>68</sup>. Furthermore, the mutation to Ala of H242<sup>6,55</sup>, which interacts with R255<sup>7,35</sup> to organize the binding pocket for the carboxylate of SCFAs (**Fig. 4b**), also abolishes SCFA function<sup>68</sup>. These mutations eliminate the binding of SCFAs rather than simply affecting ligand function, as evidenced by the fact that SCFAs are unable to compete for binding with an FFA2 orthosteric antagonist, the affinity of which is only slightly reduced compared to the wild type receptor at each of the R180A, R255A, and H242A mutants of FFA2<sup>69</sup>. Interestingly, while R180<sup>5,39</sup> and R255<sup>7,35</sup> are conserved in FFA1, H242<sup>6,55</sup> in FFA2 is replaced by N244<sup>6,55</sup> in FFA1, which does not interact with R<sup>5,39</sup> and R<sup>7,35</sup>.

In addition to R180<sup>5,39</sup> and R255<sup>7,35</sup>, two other polar residues from ECL2, Q148<sup>ECL2</sup> and Y165<sup>ECL2</sup>, are also close to butyrate (**Fig. 4b**). Y165<sup>ECL2</sup> forms a hydrogen bond with the carboxylate of butyrate, while Q148<sup>ECL2</sup> interacts with and potentially stabilizes R180<sup>5,39</sup> (**Fig. 4b**). We have previously demonstrated that substituting Q148<sup>ECL2</sup> to a glutamate residue results in a reduction of potency for SCFAs<sup>70</sup>. It has an even more dramatic effect on larger synthetic FFA2 agonists including Compound 1 (3-benzyl-4-(cyclopropyl-(4-(2,5-dichlorophenyl)thiazol-2-yl)amino)-4-oxobutanoic acid), where the agonist function is all but ablated<sup>70</sup>. This may be due to the repulsion of negatively charged carboxylate groups

in the mutated E148 residue and agonists. Also, the mutation of Y165A resulted in a nearly 15-fold reduction of the potency of SCFAs<sup>70</sup>, indicating the important role of this residue in ligand binding as well.

The opening between TM3 and TM4 of FFA1 that allows for binding of TAK-875 is closed in FFA2, resulting in a small pocket in FFA2 with limited space to accommodate the carbon chains of fatty acids (**Fig. 4c**). This may explain the selectivity of FFA2 for SCFAs over LCFAs. When comparing the structures of FFA1 and FFA2 coupled with miniGq, it becomes apparent that while TM3 aligns well between the two, TM4 in FFA2 shifts towards TM3 in comparison to FFA1 (**Fig. 4d**). H140<sup>4,56</sup> in TM4 of FFA2 appears to play an important role in determining the selectivity of chain length of fatty acids. Whilst alteration of this residue to Ala reduces the potency of SCFAs, it enables binding and function of the C6 fatty acid caproate and, to a more modest degree, C8 caprylate<sup>68</sup>. H140<sup>4,56</sup> forms a hydrogen bond with E182<sup>5,41</sup> in FFA2 (**Fig. 4d**), which is replaced by S185<sup>5,41</sup> in FFA1. The longer side chain of E182<sup>5,41</sup> may force H140<sup>4,56</sup> together with TM4 to be positioned towards TM3. The H140<sup>4,56</sup>-E182<sup>5,41</sup> pair also defines the size of the hydrophobic pocket involving C141<sup>4,57</sup>, V144<sup>4,60</sup>, V179<sup>5,38</sup>, and L183<sup>5,42</sup> in FFA2 that accommodates the hydrophobic tail of butyrate (**Fig. 4d**). Notably, the mutations V179A and L183A do not change the potency of SCFAs but cause ~10-fold reduction of the potency for larger synthetic FFA2 agonists Compound 1 and Compound 2 ((*R*)-3-(cyclopentylmethyl)-4-(cyclopropyl-(4-(2,6-dichlorophenyl)thiazol-2-yl)amino)-4-oxobutanoic acid)<sup>70</sup>. C141<sup>4,57</sup> is at the bottom of the SCFA binding pocket and forms van der Waals interactions with the last carbon of butyrate. This residue in bovine FFA2 is replaced by Gly. The C141G mutation in human FFA2 leads to an altered ligand preference for longer saturated and unsaturated C5 and C6 fatty acids, similar to that of bovine FFA2<sup>71</sup>. In previous studies, two tyrosine residues in FFA2, Y90<sup>3,29</sup> and Y238<sup>6,51</sup>, were suggested to participate in the binding of SCFAs<sup>69,70</sup>. Mutating each residue to alanine significantly reduced the potency of SCFAs<sup>69,70</sup>. In the structure of FFA2-butyrate, Y90<sup>3,29</sup> forms direct hydrophobic interactions with the short chain of butyrate, while Y238<sup>6,51</sup> forms hydrogen-bonding and cation- $\pi$  interactions with the critical arginine residue R255<sup>7,35</sup> (**Fig. 4b**), potentially stabilizing it in the appropriate position to interact with butyrate. Additionally, Y90<sup>3,29</sup> is stabilized by a hydrogen bond with K65<sup>2,60</sup> (**Fig. 4b**), and mutations of K65<sup>2,60</sup> to either alanine or glutamate substantially reduce the potency of SCFAs<sup>72</sup>. This indicates the important role of the K65<sup>2,60</sup>-Y90<sup>3,29</sup> pair in SCFA binding. Interestingly, K65<sup>2,60</sup> of human FFA2 is replaced by R65<sup>2,60</sup> in mouse FFA2. It is notable that SCFAs including butyrate display lower potency at mouse FFA2 compared to human FFA2<sup>73</sup>. Consistent with this, a K65R mutation in human FFA2 resulted in a small but still significant reduction of potency of SCFAs<sup>72</sup>.

### **Molecular determinants of ligand recognition by FFAs revealed by MD simulations**

To evaluate the stability and dynamics of the receptor-ligand interactions in FFA2 and FFA4, we performed molecular dynamics (MD) simulations of the two receptors in agonist-bound and unbound forms in a water-lipid bilayer. Guided by the protonation state prediction, D208<sup>5,39</sup> was protonated in the simulations of the FFA4 complex with TUG-891, which reduced the repulsion of the carboxyl groups of nearby E204<sup>5,35</sup>, D208<sup>5,39</sup> and TUG-891. Whilst mutation of E204 to Ala reduced the potency of TUG-891 this alteration did not alter the potency of DHA. By contrast, whilst mutation of D208 to Ala was without effect on potency of TUG-891 this alteration significantly increased the potency of DHA (**Fig.**

**S7a**), consistent with differences in the detailed location of the carboxyl of the synthetic and fatty acid ligands (**Fig. 5**). The receptor and agonists had small fluctuations, with DHA having a higher mobility, in all the three 1  $\mu$  MD simulations (**Table S3**). From the average ligand-residue interaction energy, DHA and TUG-891 showed electrostatic attraction to nearby FFA4 residues R24<sup>N</sup> and R22<sup>N</sup> and repulsion to E204<sup>5.35</sup> (**Fig. 5a and b**). However, mutation of either or both R24<sup>N</sup> and R22<sup>N</sup> residues to alanine did not alter ligand potency (**Fig. 2e, Fig. S7a**). In the simulations, in the absence of the agonists, R22<sup>N</sup> and R24<sup>N</sup> interact with E204<sup>5.35</sup> and D208<sup>5.39</sup>, thus, occluding ligand access to the interhelical hydrophobic cavity. TUG-891, with its carboxyl group located deeper in the binding cavity, is further stabilized by electrostatic interactions with W198<sup>ECL2</sup>. The MD simulations allowed us to observe water clusters at the extracellular vestibule of the orthosteric binding site of FFA4 (**Fig. 5a and b**) which mediate polar interactions between the agonists, E204<sup>5.35</sup>, D208<sup>5.39</sup>, W198<sup>ECL2</sup>, R24<sup>N</sup> and R22<sup>N</sup>. This supports the presence of water as suggested by the cryo-EM density (**Fig. 2f**).

According to ligand fragment interaction energy calculations, the carboxyl group of the FFA4 agonists is further stabilized by electrostatic attraction to electron-deficient aromatic hydrogens of F25<sup>N</sup> and F27<sup>N</sup> and van der Waals interactions with their aromatic rings (**Fig. 5a, Table S4**). The first two double bonds of DHA demonstrate high mobility and do not form persistent interactions with surrounding residues. Deeper inside the pocket, double bonds 3-5 form stable van der Waals interactions with F115<sup>3.29</sup>, M118<sup>3.32</sup>, and T119<sup>3.33</sup> (**Fig. 5a**). The last double bond of DHA is engaged in van der Waals interactions with F211<sup>5.42</sup>, I280<sup>6.51</sup>, and the ‘toggle switch’, W277<sup>6.48</sup> (**Fig. 5a**). A similar picture is observed for TUG-891, with its last two aromatic rings, having interaction energy with these residues, along with I287<sup>6.58</sup> and I284<sup>6.55</sup> (**Fig. 5b**), mutation of which to Ala, as noted earlier, reduces potency of TUG-891 (**Fig. S7a**).

In FFA2, butyrate is stabilized by strong electrostatic interactions with R180<sup>5.39</sup> and R255<sup>7.35</sup>, together with Y238<sup>6.51</sup>, Y90<sup>3.33</sup> and Q148<sup>ECL2</sup> (**Fig. 5c**), as suggested by the cryo-EM structure. In the simulations, we also observed electrostatic stabilization of the ligand by H242<sup>6.55</sup> (**Fig. 5c**). The hydrogen bond between H140<sup>4.56</sup> and E182<sup>5.41</sup> suggested by cryo-EM was found to be persistent throughout the MD simulations.

No large movements of the helices in FFA2 and FFA4 occurred upon the removal of the agonist and miniG<sub>q</sub> protein during 1  $\mu$ s simulations. However, we did observe the start of deactivation processes of the receptors. We saw an increase in mobility of the aromatic residues at positions 5.47, 6.44 and 6.48 (**Fig. S9**) associated with GPCR activation<sup>74,75</sup>. In addition, the formation of the ‘ionic lock’ involving E<sup>3.49</sup> and R<sup>3.50</sup> of the ERY motif, and the conformational change of the microswitch residue at position 7.53 of the NPxxY motif, both leading to an inactive state of GPCRs, were observed in the receptors (**Fig. S10**). These changes were observed in all simulations lacking the miniG<sub>q</sub> protein, usually to a greater extent in the systems without the agonist.

In summary, the MD simulations support the importance of hydrophobic and aromatic contacts deep inside the interhelical cavity in FFA4 as opposed to the polar contacts at the extracellular cavity of FFA2.

## Activation mechanisms of FFAs



As inactive structures of FFA4 have not been experimentally solved, we took an inactive structure model (FFA4-AF) obtained from the GPCRdb database <sup>76</sup>, which was generated using an AlphaFold-based multi-state prediction protocol <sup>77</sup>, in our structural comparison analysis. Our analysis showed significant conformational changes at the cytoplasmic region, including the outward and inward displacements of TM6 and TM7, respectively, as observed in the activation of other class A GPCRs <sup>75,78,79</sup>, when comparing the active DHA-bound structure and the inactive structure of FFA4 (**Fig. 6a**). The extracellular region exhibits rather modest conformational differences except for the N-terminal region. In the FFA4-AF, the N-terminal region sticks out and does not interact with the rest of the receptor (**Fig. 6a**). This contrasts the 'U' shape N-terminal segment that is buried inside the 7-TM bundle in the active FFA4, which is likely stabilized by the interaction between the N-terminal residue F27 and DHA (**Fig. 1c and 6a**).

At the bottom of the ligand-binding pocket of inactive FFA4, a triad of aromatic residues F216<sup>5,47</sup>, F274<sup>6,44</sup>, and W277<sup>6,48</sup> form an aromatic cluster with extensive  $\pi$ - $\pi$  interactions (**Fig. 6b**). In the active structure of FFA4-DHA, the long chain of DHA reaches this motif and causes rearrangements of the three residues (**Fig. 6b**). In many Class A GPCRs, W<sup>6,48</sup> and F<sup>6,44</sup> constitute a conserved 'activation switch' microdomain, and conformational rearrangement of this microdomain serves as a crucial step in the activation mechanism. <sup>78,80-82</sup>. Indeed, the movement of F274<sup>6,44</sup> and W277<sup>6,48</sup> breaks the continuous helical structure of TM6 of FFA4 (**Fig. 6b**), resulting in an outward displacement of the cytoplasmic segment of TM6, a hallmark of GPCR activation.

When we tried to compare the AlphaFold-predicted inactive structure of FFA2 obtained from the GPCRdb database (FFA2-AF) with our active structure of FFA2-butyrate, surprisingly, we found that FFA2-AF closely resembles the active conformation of FFA2 with only subtle differences at the cytoplasmic region, indicative of an active conformation of FFA2-AF (**Fig. 6c**). This finding complicates the examination of conformational changes during receptor activation. Nonetheless, we observed inward movement of TM5 and TM7 at the extracellular region in the FFA2-butyrate structure compared to the ligand-free structure of FFA2-AF, which is likely due to interactions between butyrate and the two arginine residues R180<sup>5,39</sup> and R255<sup>7,35</sup> (**Fig. 6c**). Similar inward movement of the extracellular segment of TM5 was also observed in the  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR) during receptor activation due to hydrogen bonds between agonists and two serine residues of TM5, leading to the rearrangement of the P<sup>5.50</sup>/I<sup>3.40</sup>/F<sup>6.44</sup> motif at the core region of the 7-TM bundle and outward movement of TM6 <sup>83,84</sup>. The PIF motif functions as a molecular microswitch in the activation of some class A GPCRs <sup>79,85</sup>. We hypothesize that FFA2 adopts a similar activation mechanism, where the agonist-induced inward movement of TM5 leads to the rearrangement of the core triad motif P191<sup>5.50</sup>/T97<sup>3.40</sup>/F231<sup>6.44</sup> and the outward movement of TM6.

For FFA1, the AlphaFold-predicted inactive structure (FFA1-AF) also displayed subtle conformational differences compared to the active miniG<sub>q</sub>-coupled FFA1 in the intracellular region, suggesting that FFA1-AF may adopt an active-like conformation (**Fig. 6d**). In addition, the DHA binding site in FFA1 cannot be definitively resolved at this time, making it difficult to speculate on the mechanism of DHA-mediated activation of the receptor. If DHA binds to 'Site 2', it is plausible that it stabilizes the helical structure of ICL2, similar to FFA1 ago-PAMs, to position it for interaction with G protein. Furthermore,

DHA interacts directly with mini-G<sub>αq</sub> at 'Site 2' (**Fig. 3b**), indicating that it may also function to directly stabilize the interactions between FFA1 and G<sub>q</sub>.

### Insights into the coupling of G<sub>q</sub> to FFAs

In our structural studies, we used the miniG<sub>q</sub> variant of the heterotrimeric G<sub>q</sub> due to its greater propensity to form stable complexes with FFAs in our experiments. Consistent with other GPCR-G protein complexes, the C-terminal α5 helix<sup>86</sup> of miniG<sub>αq</sub> served as the primary interaction site for the receptors (**Fig. 7a**). Remarkably, the orientations of α5 of miniG<sub>αq</sub> with respect to the three receptors were highly similar (**Fig. 7a**), with each receptor forming similar sets of interactions with the wavy hook region in α5 of mini-G<sub>αq</sub> (**Fig. S11**). The C-terminal α5 helix of mini-G<sub>αq</sub> is the main interaction site for the receptors. As mentioned previously, we did not observe strong cryo-EM density for a putative helix 8 (H8) in any of the four structures. This suggests a disordered C-terminal region following TM7 in the three active receptors. However, we did observe direct interactions between the intracellular end of TM7 of each receptor and mini-G<sub>αq</sub> (**Fig. S11**), underscoring an important role of TM7 in G<sub>q</sub>-coupling to each of the three receptors.

Although the interactions at the α5 helix of miniG<sub>αq</sub> are similar across all three receptors, they engage in distinct interactions with other regions of miniG<sub>q</sub>. Specifically, for FFA1 and FFA2, the intracellular loop 2 (ICL2) forms a helical structure to directly interact with miniG<sub>q</sub>. In both receptors, a conserved tyrosine residue, Y114<sup>ICL2</sup> in FFA1 and Y117<sup>ICL2</sup> in FFA2, sticks toward the 7-TM core and forms a hydrogen bond with Y243 of mini-G<sub>αq</sub> (**Fig. 7b**). Additionally, on the opposite side of ICL2, V115<sup>ICL2</sup> and L119<sup>ICL2</sup> in FFA2, or L112<sup>ICL2</sup> in FFA1, form hydrophobic interactions with mini-G<sub>αq</sub> residues L34, V79, F228, and I235 (**Fig. 7b**). A similar set of hydrophobic interactions are also observed in the structures of the muscarinic acetylcholine receptor 1 (M1R)-G<sub>11</sub><sup>87</sup> and the serotonin receptor 5-HT<sub>2A</sub>-miniG<sub>q</sub> complexes<sup>47</sup>. However, in FFA4, a large part of ICL2 is disordered, and no similar hydrophobic interactions are observed (**Fig. 1b**). Nevertheless, R145<sup>ICL2</sup> of FFA4 forms hydrogen bonds with the side chain of N239 and the main chain carbonyl of I235 in the α5 helix of mini-G<sub>αq</sub>.

ICL3 is another region in the three receptors that interacts differently with miniG<sub>q</sub>. In FFA1 and FFA2, ICL3 is positioned close to the wavy hook region of mini-G<sub>αq</sub> (**Fig. 7c**). S212<sup>ICL3</sup> in FFA1 forms a hydrogen bond with D233 of miniG<sub>αq</sub>, while R217<sup>ICL3</sup> in FFA2 forms hydrogen bonds with D233 and Q237 of miniG<sub>αq</sub> (**Fig. 7c**). However, in FFA4, TM5 extends by two helical turns compared to that of FFA1 and FFA2, resulting in ICL3 of FFA4 being positioned near the α4 helix and the α4-β6 loop of mini-G<sub>αq</sub> (**Fig. 7c**). In this position, a segment of ICL3 in FFA4 adopts a helical structure. In the structure of FFA4-miniG<sub>q</sub> with DHA, residues Y247<sup>ICL3</sup> and R254<sup>ICL3</sup> of FFA4 form hydrogen bonds with the carbonyl groups of the mini-G<sub>αq</sub> residues C211 and T202 backbone, respectively, while H251<sup>ICL3</sup> of FFA4 forms a hydrogen bond with D199 of mini-G<sub>αq</sub> (**Fig. 7c**). Additionally, V255<sup>ICL3</sup> and L241<sup>5.72</sup> of FFA4 pack against I210 of mini-G<sub>αq</sub> to form hydrophobic interactions (**Fig. 7c**).

The conformation of TM5 and ICL3 may account for the inability of FFA4<sup>Long</sup>, the long form of FFA4, to induce G<sub>q/11</sub> signaling<sup>43,61</sup>. FFA4<sup>Long</sup> contains an insertion of 16 additional amino acids after Q232<sup>5.63</sup>, which would further extend TM5 and potentially cause a severe steric clash with G<sub>αq</sub>. This clash is clearly

visible if we superimpose the AlphaFold-predicted structure of FFA4<sup>Long</sup> onto FFA4 in our structure (Fig. 7d).

## Discussion

### Lipid recognition by FFAs

The different binding pockets in FFA2 and FFA4 for their endogenous SCFA and LCFA ligands clearly explain their preferences of free fatty acids<sup>39,40</sup>. For FFA2, the small size of the binding pocket can only accommodate LCFAs with very short carbon chains. For FFA4, the binding pocket is much larger to accommodate LCFAs. It seems that the binding of LCFAs to FFA4 is largely driven by hydrophobic and  $\pi$ - $\pi$  interactions since the mutations of polar residues near the carboxylate group of DHA did not significantly reduce the potency. SCFAs and medium-chain fatty acids (MCFAs) with shorter carbon chains would result in less contacts with FFA4 and thus weaker potency for this receptor. On the other hand, despite the small observed binding pocket of FFA2, a number of larger synthetic FFA2 selective and orthosteric agonists have been identified and studied<sup>70,88</sup>. It will be interesting in time to explore how FFA2 recognize those large ligands.

The molecular basis for the lipid recognition of FFA1 is not readily clear based on our structural data. 'Site 1' and 'Site 2' are both potential binding sites for DHA. 'Site 1' but not 'Site 2' is rich in aromatic residues, which would favor the binding of PUFAs to FFA1. In addition, the two arginine residues R158<sup>5,39</sup> and R258<sup>7,35</sup> that are highly conserved in FFA1-3 form salt bridges with the carboxylate group of DHA (Fig. 3a). Mutations of these two residues significantly reduced the potency of DHA and another LCFA,  $\gamma$ -linoleic acid ( $\gamma$ -LA)<sup>65</sup>, further suggesting that 'Site 1' is the primary binding site for DHA. However, the strong cryo-EM density in 'Site 2' in our structure implies potential binding of DHA at this site as well, which may serve as the secondary binding site for DHA. Nevertheless, neither 'Site 1' nor 'Site 2' is large enough to accommodate the entire LCFAs with 20 carbons or more. The mechanism for the selectivity of FFA1 for LCFAs over SCFAs still needs further investigation.

### Insight into drug development on FFAs

Over the past decade, a growing body of research has established the critical roles of FFAs in regulating metabolism and immunity. Studies have also provided evidence suggesting that agonists of FFA1, FFA2, and FFA4 have the potential to treat metabolic and inflammatory diseases. However, only the FFA1 agonist TAK-875 has completed all three phases of clinical trials. While TAK-875 showed promising results in improving glycaemic control with a low risk of hypoglycemia in a phase III trial for diabetics, its further development was halted due to liver toxicity concerns<sup>89</sup>. The underlying mechanism behind TAK-875's liver toxicity is still not entirely clear, with studies suggesting both direct hepatotoxicity in an FFA1-dependent manner and metabolite-induced inhibition of hepatic transporters and mitochondrial respiration<sup>90,91</sup>. To advance future drug development, it is crucial to determine whether TAK-875's liver toxicity is related to the activation of FFA1. The two well-defined ligand binding sites in FFA1 offer opportunities for designing chemically diverse FFA1 ligands. If TAK-875's liver toxicity is caused by its metabolites, developing new FFA1 agonists with novel chemical scaffolds may provide a solution to this issue.

Compared to FFA1, the development of synthetic agonists for FFA2 and FFA4 toward the clinic has been limited, with only a small number of FFA2 ligands having been reported<sup>39,70,88,92</sup>. By contrast, although the chemical diversity of synthetic FFA4 agonists has been rather limited<sup>22</sup>, many analogs of the TUG-891 phenyl-propionic scaffold have been generated to improve the drug-like characteristics of this ligand for further assessment of their effects on the regulation of glucose homeostasis and other disease indications. A surprising feature of the observed binding of TUG-891 and DHA within FFA4 was the absence of interaction between the carboxylate of agonists and R99<sup>2,64</sup>. This had been widely anticipated based on earlier modelling and mutagenesis studies. Initial modelling studies linked to the development of TUG-891 predicted an ionic interaction with this residue<sup>48,93</sup> and subsequent mutagenesis to R99Q confirmed the importance of this residue as the agonists were unable to activate this mutant. However, R99<sup>2,64</sup> does not interact directly with agonists but rather forms polar interactions with D30<sup>N</sup> and E43<sup>1,35</sup> and a cation- $\pi$  interaction with F304<sup>7,36</sup> (**Fig. S12**). By doing so, R99<sup>2,64</sup> likely stabilizes the aromatic network of F25<sup>N</sup>, F27<sup>N</sup>, F28<sup>N</sup>, F115<sup>3,29</sup> and F303<sup>7,35</sup> that forms a lid to the hydrophobic pocket. Mutation of R99<sup>2,64</sup>, therefore, could break this aromatic network and lead to the exposure of the hydrophobic pocket to water destabilizing agonist binding.

Our work and another recent structural study<sup>61</sup> show that the carboxyl group of ligands does not form specific interactions with FFA4; rather binding and ligand location is driven through hydrophobic interactions. The observed orthosteric binding pocket of FFA4, as revealed by our structures, offers valuable templates for designing new agonists for this receptor using computer-aided and AI-driven drug design (CADD and AIDD) approaches. Another group of FFA4 agonists indeed don't contain the carboxyl group but have a sulphonamide or amide moiety as the central core linking the aromatic rings to form a 'L' shape<sup>22,39</sup>. Interestingly, when docking one such ligand, TUG-1197, into our FFA4 structure, the sulphonamide formed an H-bond with T119<sup>3,33</sup> and its position overlapped with the O-linker of TUG-891 (**Fig. 2d**). It is hence noteworthy that TUG-1197 displayed markedly reduced potency at a T119A mutant of FFA4, and although more modest in extent the potency of TUG-891 was also reduced at T119A.

In contrast to FFA4, key residues of the orthosteric binding pocket of FFA2 were highly aligned with those predicted by previous mutagenesis studies<sup>68</sup>. Given the challenges of developing potent and selective orthosteric FFA2 activators, there has been interest in the availability and design of selective allosteric agonists of FFA2<sup>22,39</sup>. Although nothing is currently known about their mode of binding, structural studies akin to those reported herein, and the ability to use computational tools to predict allosteric binding sites<sup>94</sup>, it is likely that rapid progress could be made. This will also allow the development of more 'drug-like' allosteric regulators of FFA2 and potentially also of the other SCFA receptor FFA3 where no high potency synthetic ligands are currently available<sup>39</sup>, even as tool compounds, to better explore the biology and potential patho-physiological functions of this receptor. In both FFA2 and FFA4, we observed strong cryo-EM density in sites similar to the 'Site 2' observed in FFA1, where we modeled PI4P (phosphatidylinositol-4-phosphate) and palmitic acid to fit the density (**Fig. S6**). These observations suggest the possibility of developing allosteric modulators for FFA2 and FFA4 targeting this site. In the case of FFA2, we also observed a large cavity at the extracellular region right above the butyrate binding pocket. This cavity represents another potential allosteric site, which

we refer to as 'Site 3'. Interestingly, if the structure of FFA2 is aligned with the structure of the muscarinic receptor M2R bound to a positive allosteric modulator (PAM) named LY2119620, the allosteric site of LY2119620 highly overlaps with the putative 'Site 3' in FFA2. As noted earlier there are a number of FFA2 PAMs reported<sup>40</sup>. They may target the two potential allosteric sites, 'Site 2' and 'Site 3', revealed by our structures. Further structural studies are needed to fully understand the molecular mechanisms of allosteric modulation in FFAs.

## Acknowledgement

We thank the cryo-EM facility at the University of Pittsburgh partly supported by the grants S10 OD025009 (Krios) and S10 OD019995 (Falcon 2/3 camera) from the National Institutes of Health (NIH) in the USA. We thank Dr. Sudha Chakrapani for oversight of the cryo-EM core facility at Case Western Reserve University. This work was supported by the NIH grant R35GM128641 to C.Z., the Medical Research Council (U.K.) grant MR/X010198/1 to G.M., the Biotechnology and Biological Sciences Research Council (U.K.) grants BB/R001480/1 and BB/S000453/1 to G.M. and BB/R007101/1 to I.G.T..

This project made use of computational time on Kelvin-2 supported by Engineering and Physical Sciences Research Council (EPSRC) (grant no. [EP/T022175/1](#) and [EP/W03204X/1](#)) and ARCHER2 granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (<https://www.hecbiosim.ac.uk>), supported by EPSRC (grant no. [EP/R029407/1](#) and [EP/W03204X/1](#)).

## Contributions

C.Z., G.M., and I.G.T. conceived the project and designed the research with X.Z. X.Z. performed protein expression and purification studies, screened cryo-EM grids, collected cryo-EM data, and processed the data under the supervision of C.Z. L.J. performed pharmacological and mutagenesis studies, A-A.G. performed computational analysis. G.M. and I.G.T. analyzed data. K.L. assisted in the cryo-EM data collection. C.Z. wrote the manuscript together with G.M. and I.G.T. with the help from X.Z.

## Competing Interests

GM is co-founder and a director of both Caldan Therapeutics (<https://www.caldantherapeutics.com/>) and KelticPharmaTherapeutics (<https://keltic-pharma.com/>) which both have interests in the development of FFA4 activators. The other authors declare no competing financial interests.

## Data availability

The 3D cryo-EM density maps of FFA signaling complexes have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-41013 and EMD-41014 for the DHA-FFA1-miniGq complex, EMD-41010 for the butyrate-FFA2-miniGq complex, and EMD-41007 and EMD-41008 for the DHA- and TUG-891-FFA4-miniGq complexes, respectively. Atomic coordinates for the atomic models have been deposited in the Protein Data Bank (PDB) under the accession numbers 8T3V, 8T3S, 8T3Q, and 8T3Q for the DHA-FFA1-miniGq complex, the butyrate-FFA2-miniGq complex, and the DHA- and TUG-891-FFA4-miniGq complexes, respectively.



## Methods

### Protein expression and purification

Human FFA1, FFA2, and FFA4 were cloned into the pFastBac vector (Thermo Fisher) with the LargeBit protein<sup>54</sup> fused to the C-terminus of each receptor. The miniGaq subunit<sup>49</sup> was cloned into the pFastBac vector. Human Gβ<sub>1</sub> was fused with an N-terminal His<sub>6</sub>-tag and a C-terminal HiBiT subunit connected with a 15-amino acid linker, which was cloned into pFastBac dual vector (Thermo Fisher) together with human Gγ<sub>2</sub>.

The scFv16 was expressed in High Five cells using Bac-to-Bac expression system. To purify the protein, the cell supernatant was collected and loaded onto Ni-NTA resins. Following nickel affinity chromatography, the protein was further purified by size exclusion chromatography using a Superdex 200 Increase 100/300 GL column (GE Healthcare). The purified scFv16 were pooled, concentrated and stored at -80°C until use.

Free fatty acid receptors, miniGaq and Gβ<sub>1</sub>γ<sub>2</sub> were co-expressed in Sf9 insect cells using Bac-to-Bac method. Sf9 cells were infected with three types of viruses at the ratio of 1:1:1 for 48 h at 27 °C. After being cultured for 48 hours, the cells were harvested and frozen at -80 °C for further protein purification. Cell pellets were thawed in lysis buffer containing 20 mM HEPES, pH7.5, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, 2.5 μg/ml leupeptin, 300 μg/ml benzamidine. The complexes of DHA-bound or TUG-891-bound FFA4 with miniG<sub>q</sub> were assembled on cell membranes by the addition of 10 μM DHA or 10 μM TUG-891. For the FFA1-miniG<sub>q</sub> and FFA2-miniG<sub>q</sub> complexes, 10 μM DHA and 1 mM butyrate were added to stimulate the formation of signaling complexes. To facilitate complex formation, 25 mU/ml Apyrase (NEB), and 100 μM TCEP was added and incubated at room temperature for 2 h. The cell membranes were isolated by centrifugation at 25,000 g for 40 min and then resuspended in solubilization buffer containing 20 mM HEPES, pH7.5, 100 mM NaCl, 0.5% (w/v) lauryl maltose neopentylglycol (LMNG, Anatrace), 0.1% (w/v) cholesteryl hemisuccinate (CHS, Anatrace), 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, 12.5 mU/ml Apyrase, 10 μM or 1 mM ligands, 2.5 μg/ml leupeptin, 300 μg/ml benzamidine, 100 μM TECP at 4 °C for 2 h. The supernatant was collected by centrifugation at 25,000 g for 1 h and incubated with nickel Sepharose resin (GE Healthcare) at 4 °C overnight. The resin was washed with a buffer A containing 20 mM HEPES, pH 7.5, 100 mM NaCl, 0.05% (w/v) LMNG, 0.01% (w/v) CHS, 20 mM imidazole, and 10 μM or 1 mM ligands, 2.5 μg/ml leupeptin, 300 μg/ml benzamidine, 100 μM TECP. The complex was eluted with buffer A containing 400 mM imidazole. The eluate was supplemented with 2mM CaCl<sub>2</sub> and loaded onto anti-Flag M1 antibody resin. After wash, the complex was eluted in buffer A containing 5 mM EDTA and 200 μg/ml FLAG peptide and concentrated using an Amicon Ultra Centrifugal Filter (MWCO, 100 kDa). Finally, a 1.3 molar excess of scFv16 was added to the elution. The sample was then loaded onto a Superdex 200 Increase 10/300 column (GE Healthcare) pre-equilibrated with buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 0.00075% (w/v) LMNG, 0.00025% (w/v) GDN, 0.00015% (w/v) CHS, 10 μM or 1 mM ligands and 100 μM TECP. Peak fractions of the complex were collected and concentrated to 5-10 mg/ml for cryo-EM experiments.

### Cryo-EM sample preparation and data acquisition

For cryo-EM grids preparation of DHA-FFA1-miniGq complex, butyrate-FFA2-miniGq complex and TUG-891-FFA4-miniGq complex, 3  $\mu$ l of the protein complex was applied onto 300 mesh R1.2/1.3 UltrAuFoil Holey gold support films (Quantifoil). For cryo-EM grids preparation of DHA-FFA4-miniGq complex, 3  $\mu$ l of the purified complex was applied to glow-discharged holey carbon grids (Quantifoil, Au300 R1.2/1.3). Grids were plunge-frozen in liquid ethane using Vitrobot Mark IV (Thermo Fischer Scientific).

All cryo-EM data were collected using Titan Krios transmission electron microscope, equipped with a Gatan K3 Summit direct electron detector and an energy filter. For TUG-891-FFA4-miniGq complex and DHA-FFA4-miniGq complex, cryo-EM movie stacks were recorded with a nominal defocus setting in the range of -1.0 to -1.8  $\mu$ m using SerialEM<sup>95</sup> with beam-tilt image-shift data collection strategy with a  $3 \times 3$  pattern and 3 shot per hole. A total of 4,968 movies for the dataset of TUG-891-FFA4-miniGq complex and 10,040 movies for three datasets of DHA-FFA4-miniGq complex were collected in the correlated double sampling (CDS) super-resolution mode of the K3 camera at a nominal magnification of 105,000 $\times$  yielding a physical pixel size of 0.828 Å. Each stack was dose-fractionated to 52 frames with a total dose of 55 e<sup>-</sup>/Å<sup>2</sup>. For DHA-FFA1-miniGq complex and butyrate-FFA2-miniGq complex, 12,349 movies and 15,371 movies were collected with a nominal magnification of 105,000 $\times$  using the SerialEM software running a  $3 \times 3$  image shift pattern and 3 shot per hole. Micrographs were recorded using a super-resolution mode at a calibrated pixel size of 0.826 Å and a defocus range of -0.8 to -2.5  $\mu$ m. Each stack was dose-fractionated to 50 frames with a total dose of 61.6 e<sup>-</sup>/Å<sup>2</sup>.

### **Data processing, 3D reconstruction and modeling building**

Image stacks were subjected to patch motion correction using cryoSPARC<sup>96</sup>. The contrast transfer function (CTF) parameters were calculated using the patch CTF estimation tool in cryoSPARC.

For the TUG-891-FFA4-miniGq complex, a total of 4,928,436 particles were auto-picked and then subjected to 2D classification to discard poorly defined particles. After ab initio reconstruction and heterogeneous refinement, 391,203 particles were subjected to non-uniform refinement and local refinement, which generated a map with an indicated global resolution of 3.06 Å at a Fourier shell correlation (FSC) of 0.143. For the DHA-FFA4-miniGq complex, a threshold of CTF fit resolution of more than 4 Å was used to exclude low-quality micrographs. Each dataset was processed separately with autopicking and 2D classification. After ab initio reconstruction and heterogeneous refinement, 380,284 particles were subjected to non-uniform refinement and local refinement, which generated a map with an indicated global resolution of 3.14 Å at a Fourier shell correlation (FSC) of 0.143. For the DHA-FFA1-miniGq complex and the butyrate-FFA2-miniGq complex, a threshold of CTF fit resolution of more than 4 Å was used to exclude low-quality micrographs, respectively. A total of 7,942,319 particles and 12,559,721 particles were auto-picked and then subjected to 2D classification to discard bad particles. After ab initio reconstruction and heterogeneous refinement, 305,318 particles and 393,952 particles were subjected to non-uniform refinement and local refinement, which generated a map with an indicated global resolution of 3.39 Å and 3.07 Å at a Fourier shell correlation (FSC) of 0.143. Local resolutions of density maps were estimated in cryoSPARC.



The Alphafold-predicted structures of FFA1, FFA2, and FFA4 served as initial models for model rebuilding and refinement against the electron microscopy map. The model was initially docked into the electron microscopy density map using Chimera <sup>97</sup>. This step was followed by iterative manual adjustment and rebuilding in COOT <sup>98</sup>. Real space refinement and Rosetta refinement were then carried out using Phenix programs <sup>99</sup>. To validate the model, MolProbity was employed for assessing its structural statistics <sup>100</sup>.

For the preparation of structural figures, Chimera and PyMOL (<https://pymol.org/2/>) were utilized. The final refinement statistics can be found in Supplementary Table 1. To evaluate the degree of overfitting during the refinement process, the final model was refined against one of the half-maps. The resulting map versus model FSC curves were compared with both half-maps and the full model. Surface coloring of the density map was achieved using UCSF Chimera <sup>97</sup>.

### **Mutagenesis and bioluminescence resonance energy transfer (BRET)-based arrestin-3 recruitment assays**

All cell culture reagents and TMB substrate solution were from Thermo Fisher Scientific (Loughborough, UK). Polyethylenimine (PEI) [linearpoly(vinyl alcohol) (MW-25000)] was from Polysciences (Warrington, PA). Molecular biology enzymes and reagents were from Promega (Southampton, UK). TUG-891 and docosahexaenoic acid (DHA) were from Tocris Biosciences (Bristol, UK). TUG-1197 was a gift from Trond Ulven (University of Copenhagen).

A plasmid encoding the short isoform of human FFA4 fused at its C terminus to enhanced yellow fluorescent protein (eYFP) and containing an N-terminal FLAG epitope tag was generated as described previously <sup>101</sup>. Mutations were introduced into the FFA4 sequence using the QuikChange method (Stratagene), and in all cases the presence of the mutation was verified through sequencing. The NanoLuc luciferase coding sequence was subcloned after PCR amplification (using primers designed to add XbaI and NotI sites) into an arrestin-3-pcDNA3 plasmid (arrestin-3-NLuc).

All FFA4 constructs were expressed in HEK293T cells, which were maintained in Dulbecco's modified Eagle's medium supplemented with 0.292 g/l L-glutamine, 1% penicillin/streptomycin mixture, and 10% heat-inactivated fetal bovine serum at 37 °C in a 5% CO<sub>2</sub> humidified atmosphere. To express receptors, cells were transfected using PEI. The day before transfection  $2 \times 10^6$  cells were plated into 10 cm dishes. Plasmid DNA was then combined with PEI (in 1:6 ratio) in 500 µl of 150 mM NaCl, thoroughly mixed then incubated for 10 min at room temperature. Cell medium was changed and the DNA-PEI mixture was added to the medium in a dropwise manner.

For BRET assays, HEK293T cells were seeded in 10 cm<sup>2</sup> dishes and transiently co-transfected with wild type or each of the indicated FFA4 mutants, each with a FLAG epitope tag engineered into the N-terminal domain and eYFP fused at its C terminus, and arrestin-3-NLuc at a 50:1 ratio respectively using PEI. Control cells were transfected with arrestin-3-NLuc only. After 24 h cells were detached by incubating with trypsin-EDTA and seeded at  $5 \times 10^4$  cells/well in poly-D-lysine coated white 96-well plates, then incubated overnight at 37°C. Cells were washed once with pre-warmed (37°C) HBSS and incubated in HBSS for 30–60 min at 37°C. The luciferase substrate coelenterazine-h (Prolume) was added to a final concentration of 5 µM and the plate incubated for 10 min at 37°C protected from light.

Agonists were added at the relevant concentrations in triplicate and the plate incubated for a further 5 min at 37°C, then the emissions at 475 nm and 535 nm were read on a PHERAstar FS. The net BRET ratio (mBRET) was calculated as follows: [(signal 535 nm/signal 475 nm) - (signal nanoluc luciferase only 535 nm/signal nanoluc luciferase only 475 nm)] \*1000. Direct measures of e-YFP fluorescence determined total expression levels of FFA4 receptor variants.

### **Cell surface enzyme-linked immunosorbent assays**

Cell surface expression of receptors was quantified by live-cell ELISA. The same co-transfected cells used for BRET studies were seeded at  $5 \times 10^4$  cells/well in poly-D-lysine-coated clear 96-well plates and incubated overnight at 37°C. Cells were incubated with primary antibody (rabbit polyclonal anti-FLAG 1:1000) in culture medium for 1 h at 37°C, then washed three times with DMEM-HEPES and incubated with secondary antibody (Horseradish peroxidase (HRP)-sheep anti-rabbit IgG 1:10,000) for 1 h at 37°C protected from light. Cells were then washed three times with warmed (37°C) PBS. Finally, PBS was removed and 100 µL/well room temperature TMB substrate was added. The plate was incubated for 15 min at room temperature protected from light, then the absorbance at 620 nm was read on a POLARStar® Omega.

### **Phylogenetic analysis**

Phylogenetic analysis of full-length sequences of human class A GPCRs (312 in total) was done using the sequences and tools provided by GPCRdb web server<sup>102</sup>. Unrooted phylogenetic tree with increasing node order was built using FigTree v1.4.4<sup>103</sup>. All GPCRs were colored according to their GPCRdb ligand type. Nodes with descendants sharing the same GPCRdb ligand type were colored accordingly, with the remaining nodes colored grey.

### **Molecular Docking**

The protein structures were prepared with the protein preparation module, and the structure of TUG-1197 was assessed with the ligand preparation module of Schrodinger software. TUG-891 from the FFA4-TUG-891 complex were selected to center the docking box. Receptor docking grids with the receptor van der Waals radius scaling of 1.0 were generated. Docking poses were obtained and evaluated with the Glide program<sup>104-107</sup>. The OPLS\_2005 force field was used in all calculations.

### **MD simulations**

The cryo-EM structures of FFA4 bound to TUG-891 or DHA, and FFA2 bound to butyrate were prepared using Schrodinger Maestro 2021-3<sup>108</sup>. All molecules except the GPCR and, if necessary, the ligand and G-protein, were removed. Missing loops and sidechain atoms of the proteins were filled using knowledge-based homology modelling of the Prime module<sup>109-111</sup> with amino acid sequences taken from the UniProt database<sup>112</sup>. The N- and C- termini of the receptors were extended by up to 3 residues as per the UniProt sequence, and the added residues were minimized using the 3D builder of Maestro. The N- and C-termini of the receptors and the G-protein were capped with acetyl and N-methyl groups, respectively. The obtained structures were analyzed using the Maestro Protein Reports tool and strong steric clashes were removed by local geometry minimization.

The protonation states of amino acids at pH 7.4 were predicted by PROPKA 3 as a separate application<sup>113,114</sup> and as implemented in the Maestro Protein Preparation workflow. Thus, D208<sup>5,39</sup> in the FFA4-TUG-891 complex and H140<sup>4,56</sup> in the FFA2-butyrate complex were kept protonated in the simulations. All non-protonated histidine residues were taken as a  $\delta$ -tautomer except H242<sup>6,55</sup> in the FFA2-butyrate complex, which was taken as an  $\epsilon$ -tautomer as it reduced the root mean square deviation of the ligand atoms.

The input for membrane simulations was prepared using the CHARMM-GUI server<sup>115-123</sup>. The receptor was oriented in membrane using the PPM server within CHARMM-GUI<sup>124</sup>. The ligand parameter files were created by the Antechamber utility of the CHARMM-GUI server with the AM1BCC charge scheme and GAFF2 atom types. The receptor was placed in the 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer membrane sized 100 Å x 100 Å and 22.5 Å solvent layer on each side of the membrane. For receptor-miniGq systems, 120 Å x 120 Å membrane bilayer was used. The solvent contained 150 mM NaCl, and the total number of atoms was between 100,000 and 130,000 atoms for receptor-only systems and between 240,000 and 260,000 atoms for the receptor-miniGq systems.

The minimization, equilibration and production were done using the PMEMD program from the Amber20 package [AMBER 20] using the ff19SB<sup>125</sup>, LIPID21<sup>126</sup>, GAFF2<sup>127</sup> force fields and the TIP3P model for the protein, membrane lipids, ligands and water, respectively. The nonbonded interaction cut-off was set as 9 Å.

As per the recommended CHARMM-GUI protocol, the initial energy minimization included 5000 steepest descent steps followed by 5000 more steps using the conjugated gradient method. The equilibration steps followed the same pattern of gradually decreasing the force constants of positional restraints but were conducted with longer simulation length than recommended by the CHARMM-GUI protocol. Heating to 310 K was done in the NVT ensemble with 1 fs timestep in two consecutive equilibration steps, each 2.5 ns long. This was followed by a 2.5 ns NPT equilibration with Berendsen barostat and a 1 fs timestep<sup>128</sup>. The next two steps of NPT equilibration used 2 fs timestep and lasted for 10 ns each, and the final step in which only the protein backbone was restrained lasted 20 ns. The production was done with Langevin thermostat<sup>129</sup> with a friction coefficient of 1.0 ps<sup>-1</sup> and Berendsen barostat with timestep of 2 fs. The production was done in 3 replicas of 1  $\mu$ s for the empty and ligand-bound receptor systems and 1 replica of 1  $\mu$ s for the ligand bound receptor-miniGq systems. The snapshots of the simulations were saved every 50'000 steps, or 0.1 ns, in simulations.

The results of simulations were analysed using CPPTRAJ from the Amber20 package and MDAnalysis<sup>130,131</sup>. The residue–ligand interaction energy was calculated using the “namdenergy.tcl” script v1.6 for VMD<sup>132</sup> and the NAMD2 program with cut-off and switch parameters of 9 Å and 7.5 Å, respectively<sup>133,134</sup>. Forcefield parameters were taken from the AMBER parameter file used for simulations. Energy calculation was done for every 10<sup>th</sup> snapshot, or every 1.0 ns of simulation.

## References

- 1 Ichimura, A., Hirasawa, A., Hara, T. & Tsujimoto, G. Free fatty acid receptors act as nutrient sensors to regulate energy homeostasis. *Prostaglandins Other Lipid Mediat* **89**, 82-88, doi:10.1016/j.prostaglandins.2009.05.003 (2009).
- 2 Hara, T. *et al.* Role of free fatty acid receptors in the regulation of energy metabolism. *Biochim Biophys Acta* **1841**, 1292-1300, doi:10.1016/j.bbalip.2014.06.002 (2014).
- 3 Alvarez-Curto, E. & Milligan, G. Metabolism meets immunity: The role of free fatty acid receptors in the immune system. *Biochem Pharmacol* **114**, 3-13, doi:10.1016/j.bcp.2016.03.017 (2016).
- 4 Kimura, I., Ichimura, A., Ohue-Kitano, R. & Igarashi, M. Free Fatty Acid Receptors in Health and Disease. *Physiol Rev* **100**, 171-210, doi:10.1152/physrev.00041.2018 (2020).
- 5 Oh, D. Y. *et al.* GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* **142**, 687-698, doi:10.1016/j.cell.2010.07.041 (2010).
- 6 Briscoe, C. P. *et al.* The orphan G protein-coupled receptor GPR40 is activated by medium and long chain fatty acids. *J Biol Chem* **278**, 11303-11311, doi:10.1074/jbc.M211495200 (2003).
- 7 Lee, W. J. & Hase, K. Gut microbiota-generated metabolites in animal health and disease. *Nat Chem Biol* **10**, 416-424, doi:10.1038/nchembio.1535 (2014).
- 8 Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* **12**, 661-672, doi:10.1038/nrmicro3344 (2014).
- 9 Rooks, M. G. & Garrett, W. S. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol* **16**, 341-352, doi:10.1038/nri.2016.42 (2016).
- 10 McCarville, J. L., Chen, G. Y., Cuevas, V. D., Troha, K. & Ayres, J. S. Microbiota Metabolites in Health and Disease. *Annu Rev Immunol* **38**, 147-170, doi:10.1146/annurev-immunol-071219-125715 (2020).
- 11 Brown, A. J. *et al.* The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem* **278**, 11312-11319, doi:10.1074/jbc.M211609200 (2003).
- 12 Le Poul, E. *et al.* Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *J Biol Chem* **278**, 25481-25489, doi:10.1074/jbc.M301403200 (2003).
- 13 Itoh, Y. *et al.* Free fatty acids regulate insulin secretion from pancreatic beta cells through GPR40. *Nature* **422**, 173-176, doi:10.1038/nature01478 (2003).
- 14 Li, Z., Zhou, Z. & Zhang, L. Current status of GPR40/FFAR1 modulators in medicinal chemistry (2016-2019): a patent review. *Expert Opin Ther Pat* **30**, 27-38, doi:10.1080/13543776.2020.1698546 (2020).
- 15 Mancini, A. D. & Poitout, V. The fatty acid receptor FFA1/GPR40 a decade later: how much do we know? *Trends Endocrinol Metab* **24**, 398-407, doi:10.1016/j.tem.2013.03.003 (2013).
- 16 Mancini, A. D. & Poitout, V. GPR40 agonists for the treatment of type 2 diabetes: life after 'TAKing' a hit. *Diabetes Obes Metab* **17**, 622-629, doi:10.1111/dom.12442 (2015).
- 17 Nakamoto, K. *et al.* DHA supplementation prevent the progression of NASH via GPR120 signaling. *Eur J Pharmacol* **820**, 31-38, doi:10.1016/j.ejphar.2017.11.046 (2018).
- 18 Zhao, C. *et al.* DHA Sensor GPR120 in Host Defense Exhibits the Dual Characteristics of Regulating Dendritic Cell Function and Skewing the Balance of Th17/Tregs. *Int J Biol Sci* **16**, 374-387, doi:10.7150/ijbs.39551 (2020).
- 19 Arantes, E. L. *et al.* Topical Docosahexaenoic Acid (DHA) Accelerates Skin Wound Healing in Rats and Activates GPR120. *Biol Res Nurs* **18**, 411-419, doi:10.1177/1099800415621617 (2016).

- 20 Chen, J., Wang, D., Zong, Y. & Yang, X. DHA Protects Hepatocytes from Oxidative Injury through GPR120/ERK-Mediated Mitophagy. *Int J Mol Sci* **22**, doi:10.3390/ijms22115675 (2021).
- 21 Milligan, G., Alvarez-Curto, E., Hudson, B. D., Prihandoko, R. & Tobin, A. B. FFA4/GPR120: Pharmacology and Therapeutic Opportunities. *Trends Pharmacol Sci* **38**, 809-821, doi:10.1016/j.tips.2017.06.006 (2017).
- 22 Carullo, G. *et al.* GPR120/FFAR4 Pharmacology: Focus on Agonists in Type 2 Diabetes Mellitus Drug Discovery. *J Med Chem* **64**, 4312-4332, doi:10.1021/acs.jmedchem.0c01002 (2021).
- 23 Wang, X. *et al.* Discovery of Novel and Selective G-Protein Coupled Receptor 120 (GPR120) Agonists for the Treatment of Type 2 Diabetes Mellitus. *Molecules* **27**, doi:10.3390/molecules27249018 (2022).
- 24 Halder, S., Kumar, S. & Sharma, R. The therapeutic potential of GPR120: a patent review. *Expert Opin Ther Pat* **23**, 1581-1590, doi:10.1517/13543776.2013.842977 (2013).
- 25 Bianchini, G. *et al.* A new synthetic dual agonist of GPR120/GPR40 induces GLP-1 secretion and improves glucose homeostasis in mice. *Biomed Pharmacother* **139**, 111613, doi:10.1016/j.biopha.2021.111613 (2021).
- 26 Zhang, X. & Macielag, M. J. GPR120 agonists for the treatment of diabetes: a patent review (2014 present). *Expert Opin Ther Pat* **30**, 729-742, doi:10.1080/13543776.2020.1811852 (2020).
- 27 Galindo, M. M. *et al.* G protein-coupled receptors in human fat taste perception. *Chem Senses* **37**, 123-139, doi:10.1093/chemse/bjr069 (2012).
- 28 Iwata, S., Yoshida, R. & Ninomiya, Y. Taste transductions in taste receptor cells: basic tastes and moreover. *Curr Pharm Des* **20**, 2684-2692, doi:10.2174/13816128113199990575 (2014).
- 29 Cartoni, C. *et al.* Taste preference for fatty acids is mediated by GPR40 and GPR120. *J Neurosci* **30**, 8376-8382, doi:10.1523/JNEUROSCI.0496-10.2010 (2010).
- 30 Kim, M. H., Kang, S. G., Park, J. H., Yanagisawa, M. & Kim, C. H. Short-chain fatty acids activate GPR41 and GPR43 on intestinal epithelial cells to promote inflammatory responses in mice. *Gastroenterology* **145**, 396-406 e391-310, doi:10.1053/j.gastro.2013.04.056 (2013).
- 31 Cani, P. D., Everard, A. & Duparc, T. Gut microbiota, enteroendocrine functions and metabolism. *Curr Opin Pharmacol* **13**, 935-940, doi:10.1016/j.coph.2013.09.008 (2013).
- 32 Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282-1286, doi:10.1038/nature08530 (2009).
- 33 Kimura, I. *et al.* The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nat Commun* **4**, 1829, doi:10.1038/ncomms2852 (2013).
- 34 Zhao, Y. *et al.* GPR43 mediates microbiota metabolite SCFA regulation of antimicrobial peptide expression in intestinal epithelial cells via activation of mTOR and STAT3. *Mucosal Immunol* **11**, 752-762, doi:10.1038/mi.2017.118 (2018).
- 35 Antunes, K. H. *et al.* Microbiota-derived acetate protects against respiratory syncytial virus infection through a GPR43-type 1 interferon response. *Nat Commun* **10**, 3273, doi:10.1038/s41467-019-11152-6 (2019).
- 36 Bindels, L. B., Dewulf, E. M. & Delzenne, N. M. GPR43/FFA2: physiopathological relevance and therapeutic prospects. *Trends Pharmacol Sci* **34**, 226-232, doi:10.1016/j.tips.2013.02.002 (2013).
- 37 Kim, S., Kim, Y. M. & Kwak, Y. S. A novel therapeutic target, GPR43; where it stands in drug discovery. *Arch Pharm Res* **35**, 1505-1509, doi:10.1007/s12272-012-0900-6 (2012).
- 38 Kimura, I., Inoue, D., Hirano, K. & Tsujimoto, G. The SCFA Receptor GPR43 and Energy Metabolism. *Front Endocrinol (Lausanne)* **5**, 85, doi:10.3389/fendo.2014.00085 (2014).



- 39 Milligan, G., Shimpukade, B., Ulven, T. & Hudson, B. D. Complex Pharmacology of Free Fatty Acid Receptors. *Chem Rev* **117**, 67-110, doi:10.1021/acs.chemrev.6b00056 (2017).
- 40 Grundmann, M., Bender, E., Schamberger, J. & Eitner, F. Pharmacology of Free Fatty Acid Receptors and Their Allosteric Modulators. *Int J Mol Sci* **22**, doi:10.3390/ijms22041763 (2021).
- 41 Husted, A. S. *et al.* Autocrine negative feedback regulation of lipolysis through sensing of NEFAs by FFAR4/GPR120 in WAT. *Mol Metab* **42**, 101103, doi:10.1016/j.molmet.2020.101103 (2020).
- 42 Moore, K., Zhang, Q., Murgolo, N., Hosted, T. & Duffy, R. Cloning, expression, and pharmacological characterization of the GPR120 free fatty acid receptor from cynomolgus monkey: comparison with human GPR120 splice variants. *Comp Biochem Physiol B Biochem Mol Biol* **154**, 419-426, doi:10.1016/j.cbpb.2009.08.005 (2009).
- 43 Watson, S. J., Brown, A. J. & Holliday, N. D. Differential signaling by splice variants of the human free fatty acid receptor GPR120. *Mol Pharmacol* **81**, 631-642, doi:10.1124/mol.111.077388 (2012).
- 44 Lu, J. *et al.* Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat Struct Mol Biol* **24**, 570-577, doi:10.1038/nsmb.3417 (2017).
- 45 Ho, J. D. *et al.* Structural basis for GPR40 allosteric agonism and incretin stimulation. *Nat Commun* **9**, 1645, doi:10.1038/s41467-017-01240-w (2018).
- 46 Srivastava, A. *et al.* High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **513**, 124-127, doi:10.1038/nature13494 (2014).
- 47 Kim, K. *et al.* Structure of a Hallucinogen-Activated Gq-Coupled 5-HT(2A) Serotonin Receptor. *Cell* **182**, 1574-1588 e1519, doi:10.1016/j.cell.2020.08.024 (2020).
- 48 Shimpukade, B., Hudson, B. D., Hovgaard, C. K., Milligan, G. & Ulven, T. Discovery of a potent and selective GPR120 agonist. *J Med Chem* **55**, 4511-4515, doi:10.1021/jm300215x (2012).
- 49 Wan, Q. *et al.* Mini G protein probes for active G protein-coupled receptors (GPCRs) in live cells. *J Biol Chem* **293**, 7466-7473, doi:10.1074/jbc.RA118.001975 (2018).
- 50 Jiang, W. & Zheng, S. Structural insights into galanin receptor signaling. *Proc Natl Acad Sci U S A* **119**, e2121465119, doi:10.1073/pnas.2121465119 (2022).
- 51 Xia, R. *et al.* Cryo-EM structure of the human histamine H(1) receptor/G(q) complex. *Nat Commun* **12**, 2086, doi:10.1038/s41467-021-22427-2 (2021).
- 52 Meyerowitz, J. G. *et al.* The oxytocin signaling complex reveals a molecular switch for cation dependence. *Nat Struct Mol Biol* **29**, 274-281, doi:10.1038/s41594-022-00728-4 (2022).
- 53 Zhang, S. *et al.* Molecular basis for selective activation of DREADD-based chemogenetics. *Nature* **612**, 354-362, doi:10.1038/s41586-022-05489-0 (2022).
- 54 Duan, J. *et al.* Cryo-EM structure of an activated VIP1 receptor-G protein complex revealed by a NanoBiT tethering strategy. *Nat Commun* **11**, 4121, doi:10.1038/s41467-020-17933-8 (2020).
- 55 Koehl, A. *et al.* Structure of the micro-opioid receptor-Gi protein complex. *Nature* **558**, 547-552, doi:10.1038/s41586-018-0219-7 (2018).
- 56 Wang, L. *et al.* Structures of the Human PGD(2) Receptor CRTH2 Reveal Novel Mechanisms for Ligand Recognition. *Mol Cell* **72**, 48-59 e44, doi:10.1016/j.molcel.2018.08.009 (2018).
- 57 Liu, H. *et al.* Molecular basis for lipid recognition by the prostaglandin D(2) receptor CRTH2. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2102813118 (2021).
- 58 Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods in Neurosciences* **25**, 366-428, doi:[https://doi.org/10.1016/S1043-9471\(05\)80049-7](https://doi.org/10.1016/S1043-9471(05)80049-7) (1995).

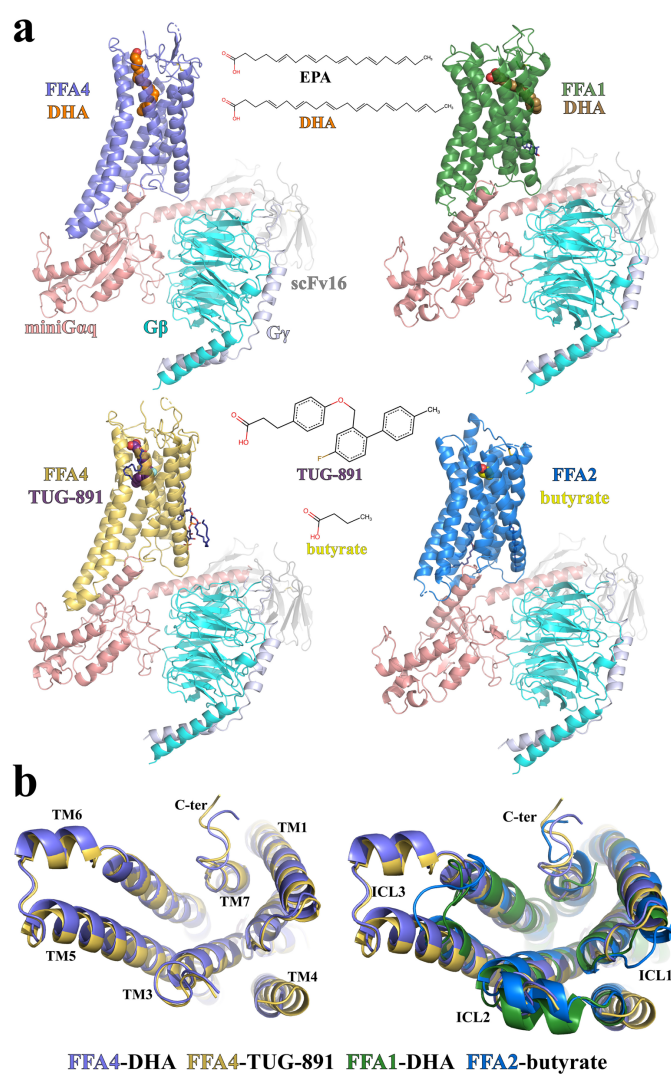
- 59 Hudson, B. D., Shimpukade, B., Milligan, G. & Ulven, T. The molecular basis of ligand interaction at free fatty acid receptor 4 (FFA4/GPR120). *J Biol Chem* **289**, 20345-20358, doi:10.1074/jbc.M114.561449 (2014).
- 60 Azevedo, C. M. *et al.* Non-Acidic Free Fatty Acid Receptor 4 Agonists with Antidiabetic Activity. *J Med Chem* **59**, 8868-8878, doi:10.1021/acs.jmedchem.6b00685 (2016).
- 61 Mao, C. *et al.* Unsaturated bond recognition leads to biased signal in a fatty acid receptor. *Science*, eadd6220, doi:10.1126/science.add6220 (2023).
- 62 Liu, H. *et al.* Orthosteric and allosteric action of the C5a receptor antagonists. *Nat Struct Mol Biol* **25**, 472-481, doi:10.1038/s41594-018-0067-z (2018).
- 63 Liu, X. *et al.* An allosteric modulator binds to a conformational hub in the beta(2) adrenergic receptor. *Nat Chem Biol* **16**, 749-755, doi:10.1038/s41589-020-0549-2 (2020).
- 64 Yabuki, C. *et al.* A novel antidiabetic drug, fasiglifam/TAK-875, acts as an ago-allosteric modulator of FFAR1. *PLoS One* **8**, e76280, doi:10.1371/journal.pone.0076280 (2013).
- 65 Kumari, P., Inoue, A., Chapman, K., Lian, P. & Rosenbaum, D. M. Molecular mechanism of fatty acid activation of FFAR1. *Proc Natl Acad Sci U S A* **120**, e2219569120, doi:10.1073/pnas.2219569120 (2023).
- 66 Yang, F. *et al.* Structural basis of GPCR activation and bile acid recognition. *Nature* **587**, 499-504, doi:10.1038/s41586-020-2569-1 (2020).
- 67 Stoddart, L. A., Smith, N. J. & Milligan, G. International Union of Pharmacology. LXXI. Free fatty acid receptors FFA1, -2, and -3: pharmacology and pathophysiological functions. *Pharmacol Rev* **60**, 405-417, doi:10.1124/pr.108.00802 (2008).
- 68 Stoddart, L. A., Smith, N. J., Jenkins, L., Brown, A. J. & Milligan, G. Conserved polar residues in transmembrane domains V, VI, and VII of free fatty acid receptor 2 and free fatty acid receptor 3 are required for the binding and function of short chain fatty acids. *J Biol Chem* **283**, 32913-32924, doi:10.1074/jbc.M805601200 (2008).
- 69 Sergeev, E. *et al.* Non-equivalence of Key Positively Charged Residues of the Free Fatty Acid 2 Receptor in the Recognition and Function of Agonist Versus Antagonist Ligands. *J Biol Chem* **291**, 303-317, doi:10.1074/jbc.M115.687939 (2016).
- 70 Hudson, B. D. *et al.* Defining the molecular basis for the first potent and selective orthosteric agonists of the FFA2 free fatty acid receptor. *J Biol Chem* **288**, 17296-17312, doi:10.1074/jbc.M113.455337 (2013).
- 71 Hudson, B. D. *et al.* Chemically engineering ligand selectivity at the free fatty acid receptor 2 based on pharmacological variation between species orthologs. *FASEB J* **26**, 4951-4965, doi:10.1096/fj.12-213314 (2012).
- 72 Sergeev, E. *et al.* A single extracellular amino acid in Free Fatty Acid Receptor 2 defines antagonist species selectivity and G protein selection bias. *Sci Rep* **7**, 13741, doi:10.1038/s41598-017-14096-3 (2017).
- 73 Hudson, B. D., Tikhonova, I. G., Pandey, S. K., Ulven, T. & Milligan, G. Extracellular ionic locks determine variation in constitutive activity and ligand potency between species orthologs of the free fatty acid receptors FFA2 and FFA3. *J Biol Chem* **287**, 41195-41209, doi:10.1074/jbc.M112.396259 (2012).
- 74 Filipek, S. Molecular switches in GPCRs. *Curr Opin Struct Biol* **55**, 114-120, doi:10.1016/j.sbi.2019.03.017 (2019).
- 75 Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *Elife* **8**, doi:10.7554/eLife.50279 (2019).
- 76 Pandey-Szekeres, G. *et al.* GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Res* **51**, D395-D402, doi:10.1093/nar/gkac1013 (2023).

- 77 Heo, L. & Feig, M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins* **90**, 1873-1885, doi:10.1002/prot.26382 (2022).
- 78 Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **87**, 897-919, doi:10.1146/annurev-biochem-060614-033910 (2018).
- 79 Hauser, A. S. *et al.* GPCR activation mechanisms across classes and macro/microscales. *Nat Struct Mol Biol* **28**, 879-888, doi:10.1038/s41594-021-00674-7 (2021).
- 80 Manglik, A. & Kruse, A. C. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry* **56**, 5628-5634, doi:10.1021/acs.biochem.7b00747 (2017).
- 81 Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* **21**, 541-551, doi:10.1016/j.sbi.2011.06.002 (2011).
- 82 Trzaskowski, B. *et al.* Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr Med Chem* **19**, 1090-1109, doi:10.2174/092986712799320556 (2012).
- 83 Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **469**, 175-180, doi:10.1038/nature09648 (2011).
- 84 Ring, A. M. *et al.* Adrenaline-activated structure of beta2-adrenoceptor stabilized by an engineered nanobody. *Nature* **502**, 575-579, doi:10.1038/nature12572 (2013).
- 85 White, K. L. *et al.* Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure* **26**, 259-269 e255, doi:10.1016/j.str.2017.12.013 (2018).
- 86 Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* **9**, 60-71, doi:10.1038/nrm2299 (2008).
- 87 Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G. & Kobilka, B. K. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* **364**, 552-557, doi:10.1126/science.aaw5188 (2019).
- 88 Hansen, A. H. *et al.* Discovery of a Potent Thiazolidine Free Fatty Acid Receptor 2 Agonist with Favorable Pharmacokinetic Properties. *J Med Chem* **61**, 9534-9550, doi:10.1021/acs.jmedchem.8b00855 (2018).
- 89 Kaku, K., Enya, K., Nakaya, R., Ohira, T. & Matsuno, R. Efficacy and safety of fasiglifam (TAK-875), a G protein-coupled receptor 40 agonist, in Japanese patients with type 2 diabetes inadequately controlled by diet and exercise: a randomized, double-blind, placebo-controlled, phase III trial. *Diabetes Obes Metab* **17**, 675-681, doi:10.1111/dom.12467 (2015).
- 90 Kim, M. *et al.* Fasiglifam (TAK-875), a G Protein-Coupled Receptor 40 (GPR40) Agonist, May Induce Hepatotoxicity through Reactive Oxygen Species Generation in a GPR40-Dependent Manner. *Biomol Ther (Seoul)* **26**, 599-607, doi:10.4062/biomolther.2017.225 (2018).
- 91 Otieno, M. A. *et al.* Fasiglifam (TAK-875): Mechanistic Investigation and Retrospective Identification of Hazards for Drug Induced Liver Injury. *Toxicol Sci* **163**, 374-384, doi:10.1093/toxsci/kfx040 (2018).
- 92 Forbes, S. *et al.* Selective FFA2 Agonism Appears to Act via Intestinal PYY to Reduce Transit and Food Intake but Does Not Improve Glucose Tolerance in Mouse Models. *Diabetes* **64**, 3763-3771, doi:10.2337/db15-0481 (2015).
- 93 Tikhonova, I. G. Application of GPCR Structures for Modelling of Free Fatty Acid Receptors. *Handb Exp Pharmacol* **236**, 57-77, doi:10.1007/164\_2016\_52 (2017).
- 94 Ciancetta, A. *et al.* Probe Confined Dynamic Mapping for G Protein-Coupled Receptor Allosteric Site Prediction. *ACS Cent Sci* **7**, 1847-1862, doi:10.1021/acscentsci.1c00802 (2021).
- 95 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).

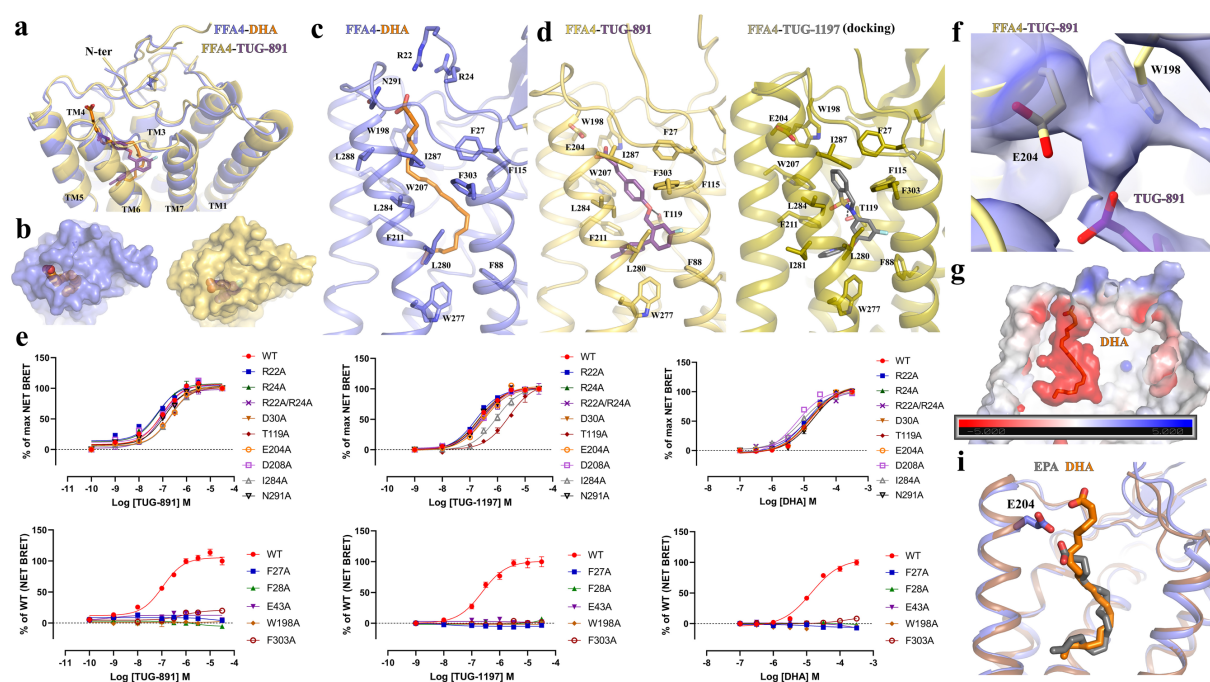


- 96 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
- 97 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 98 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
- 99 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 100 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
- 101 Hudson, B. D. *et al.* The pharmacology of TUG-891, a potent and selective agonist of the free fatty acid receptor 4 (FFA4/GPR120), demonstrates both potential opportunity and possible challenges to therapeutic agonism. *Mol Pharmacol* **84**, 710-725, doi:10.1124/mol.113.087783 (2013).
- 102 Pandey-Szekeres, G. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* **46**, D440-D446, doi:10.1093/nar/gkx1109 (2018).
- 103 FigTree v. 1.4.4 (Institute of Evolutionary Biology, University of Edinburgh, 2018).
- 104 Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739-1749, doi:10.1021/jm0306430 (2004).
- 105 Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**, 1750-1759, doi:10.1021/jm030644s (2004).
- 106 Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **49**, 6177-6196, doi:10.1021/jm051256o (2006).
- 107 Schrödinger Release 2021-3: Glide (2021).
- 108 Schrödinger Release 2021-3: Maestro (2021).
- 109 Schrödinger Release 2021-3: Prime (2021).
- 110 Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351-367, doi:10.1002/prot.10613 (2004).
- 111 He, X., Man, V. H., Ji, B., Xie, X. Q. & Wang, J. Calculate protein-ligand binding affinities with the extended linear interaction energy method: application on the Cathepsin S set in the D3R Grand Challenge 3. *J Comput Aided Mol Des* **33**, 105-117, doi:10.1007/s10822-018-0162-6 (2019).
- 112 UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531, doi:10.1093/nar/gkac1052 (2023).
- 113 Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **7**, 525-537, doi:10.1021/ct100578z (2011).
- 114 Sondergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput* **7**, 2284-2295, doi:10.1021/ct200133y (2011).
- 115 Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2**, e880, doi:10.1371/journal.pone.0000880 (2007).

- 116 Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* **97**, 50-58, doi:10.1016/j.bpj.2009.04.013 (2009).
- 117 Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004, doi:10.1002/jcc.23702 (2014).
- 118 Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**, 405-413, doi:10.1021/acs.jctc.5b00935 (2016).
- 119 Kim, S. *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem* **38**, 1879-1886, doi:10.1002/jcc.24829 (2017).
- 120 Lee, J. *et al.* CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J Chem Theory Comput* **15**, 775-786, doi:10.1021/acs.jctc.8b01066 (2019).
- 121 Lee, J. *et al.* CHARMM-GUI supports the Amber force fields. *J Chem Phys* **153**, 035103, doi:10.1063/5.0012280 (2020).
- 122 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865, doi:10.1002/jcc.20945 (2008).
- 123 Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614, doi:10.1002/jcc.21287 (2009).
- 124 Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* **40**, D370-376, doi:10.1093/nar/gkr703 (2012).
- 125 Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **16**, 528-552, doi:10.1021/acs.jctc.9b00591 (2020).
- 126 Dickson, C. J., Walker, R. C. & Gould, I. R. Lipid21: Complex Lipid Membrane Simulations with AMBER. *J Chem Theory Comput* **18**, 1726-1736, doi:10.1021/acs.jctc.1c01217 (2022).
- 127 He, X., Man, V. H., Yang, W., Lee, T. S. & Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J Chem Phys* **153**, 114502, doi:10.1063/5.0019056 (2020).
- 128 Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **81**, 3684-3690, doi:10.1063/1.448118 (1984).
- 129 Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics* **65**, 1409-1419, doi:10.1080/00268978800101881 (2006).
- 130 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319-2327, doi:10.1002/jcc.21787 (2011).
- 131 Gowers, R. *et al.* in *Proceedings of the 15th Python in Science Conference*. 98-105.
- 132 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38, doi:10.1016/0263-7855(96)00018-5 (1996).
- 133 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).
- 134 Phillips, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **153**, 044130, doi:10.1063/5.0014475 (2020).

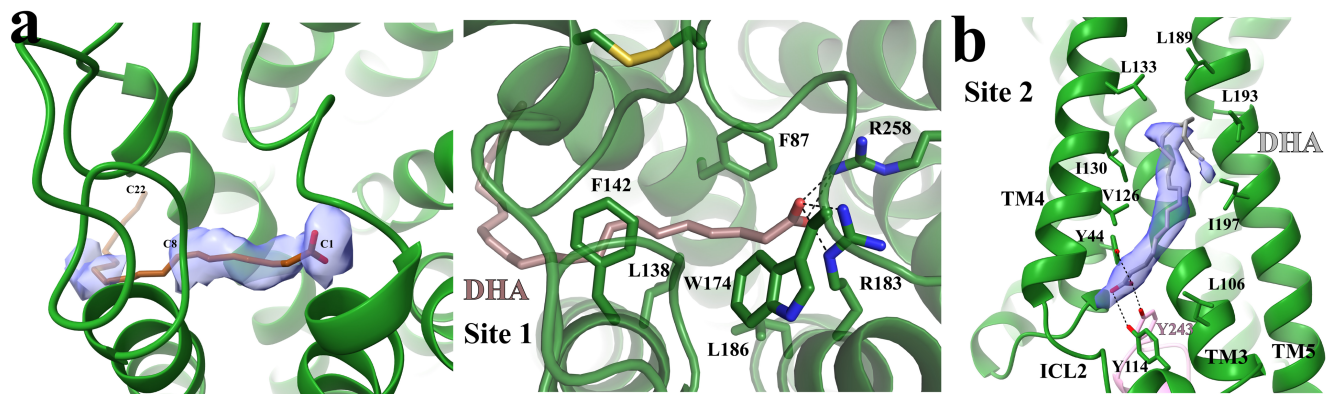


**Figure 1. Overall structures of FFA1, FFA2, and FFA4 signaling complexes. (a)** Overall structures of DHA-FFA4 (slate), DHA-FFA1 (green), TUG-891-FFA4 (dark yellow) and butyrate-FFA2 (blue), each in complex with miniG<sub>q</sub>, are shown, as are the chemical structures of the bound ligands. miniG<sub>αq</sub>, G<sub>β</sub> and G<sub>γ</sub> subunits are colored in salmon, cyan and light blue, respectively. ScFv16 is colored grey. The LCFA eicosapentaenoic acid (EPA) is also shown for comparison to DHA (see main text for discussion). **(b)** Comparison of the above structures as seen from the intracellular face.

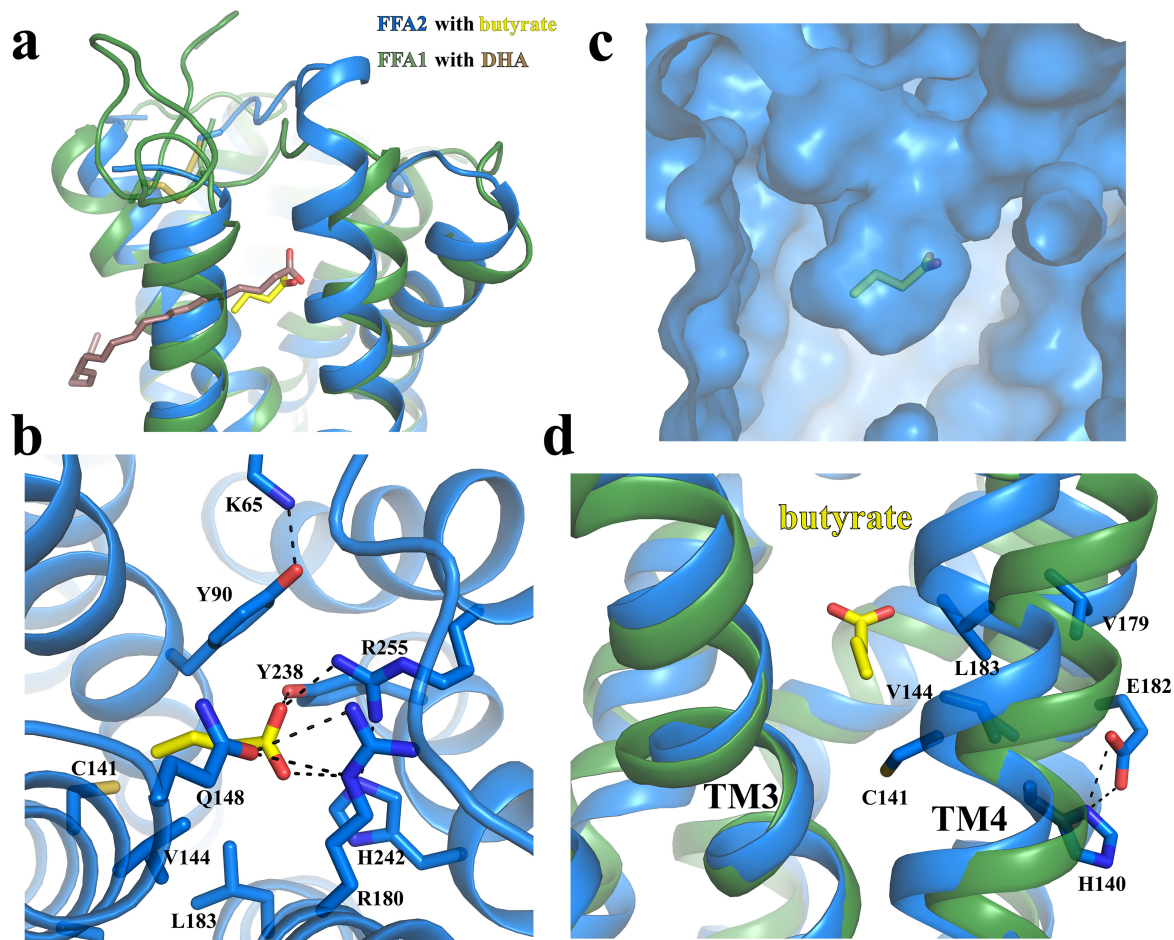


**Figure 2. Ligand binding in FFA4.** (a-d) Details of the interactions of DHA (orange) and TUG-891 (purple) with FFA4. Panel **a** illustrates the general positions of the two ligands whilst **b** highlights the closed nature of the occupied ligand binding pockets. Details of key residues of the binding pockets are highlighted for DHA (**c**) and TUG-891 (**d, left**). TUG-1197 docked into the FFA4 structure (**d, right**) highlights the important role of T119 and the similarity of the binding mode of TUG-1197 and TUG-891 at the bottom of the pocket. (**e**) Various point mutants of FFA4 generated and assessed for the ability of each of TUG-891, TUG-1197 and DHA to promote interactions with arrestin-3. See **Fig. S7** for quantitation. In concert with the large-scale mutagenesis studies reported previously<sup>59</sup>, this provides a comprehensive analysis of the orthosteric binding pocket of FFA4. (**f**) Continuous electron density observed between W198 and E204. The cryo-EM map is contoured at the level of 0.13. (**g**) Negative charge potential of the FFA4 binding pocket with DHA. (**h**) Additional length of DHA compared to EPA and the position of the DHA carboxylate above and beyond E204.

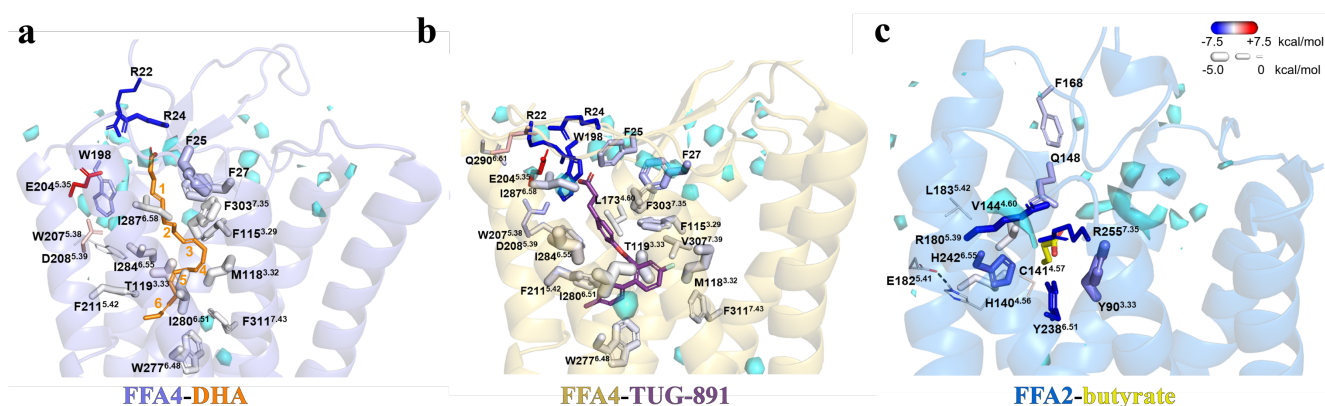




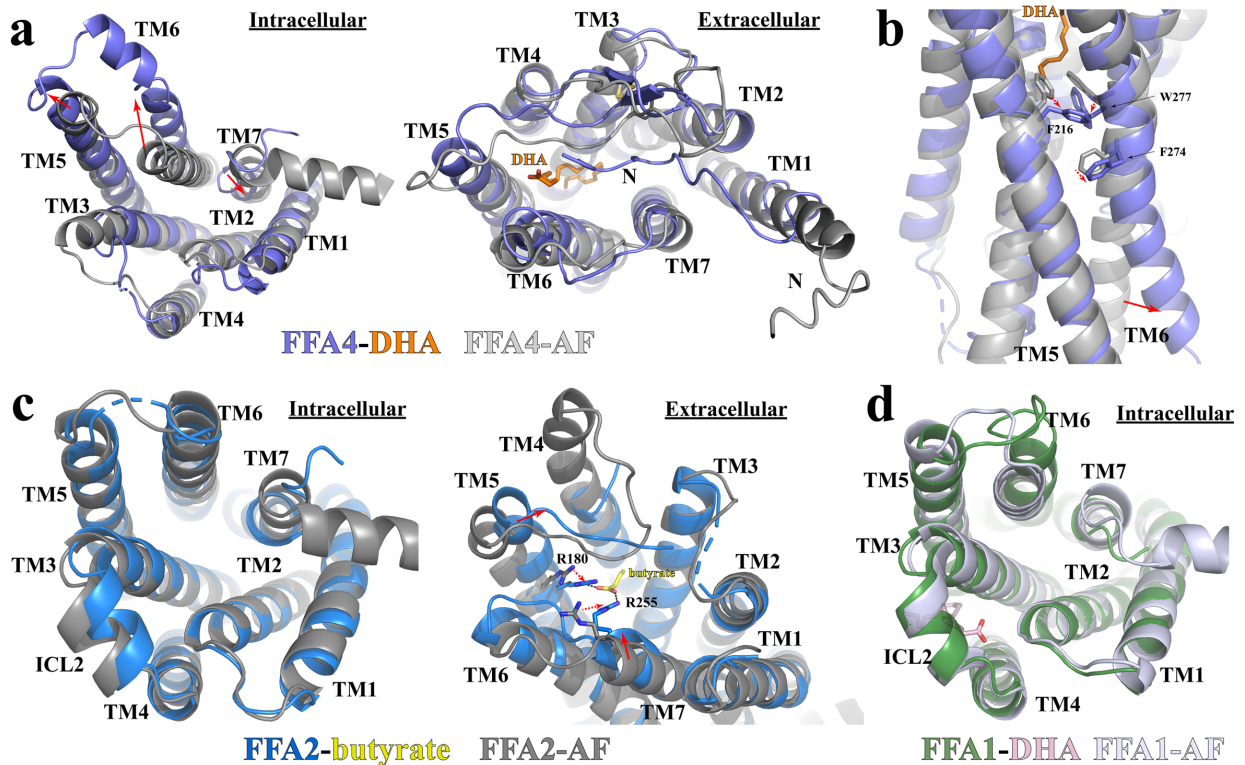
**Figure 3. Potential DHA binding sites in FFA1. (a)** DHA binding in 'Site 1'. The partial cryo-EM density map of DHA colored in light blue in the left panel is contoured at the level of 0.07. C1, C8 and C22 atoms of DHA are labeled. The occupancy of DHA C9-C22 was assigned as zero due to a lack of density. The details of interactions between DHA and FFA1 in 'Site 1' are shown in the right panel. DHA is colored brown. **(b)** Putative DHA binding in 'Site 2'. The strong cryo-EM density map in this site is contoured at the level of 0.12. The modeled DHA molecule is colored grey. Polar interactions are shown as black dashed lines. FFA1 is colored green.



**Figure 4. Recognition of butyrate by FFA2.** (a) Structural alignment of FFA2-butyrates and FFA1-DHA. The carboxylate of butyrate occupies the equivalent position to the carboxylate of DHA. (b) Details of interactions between butyrate and FFA2. (c) Overall shape of the butyrate binding pocket. (d) Differences in location of TM3 and TM4 in FFA1 and FFA2. In all panels, FFA1 and FFA2 are colored in green and blue, respectively, whilst DHA and butyrate are colored in brown and yellow, respectively. Polar interactions are shown as black dashed lines.

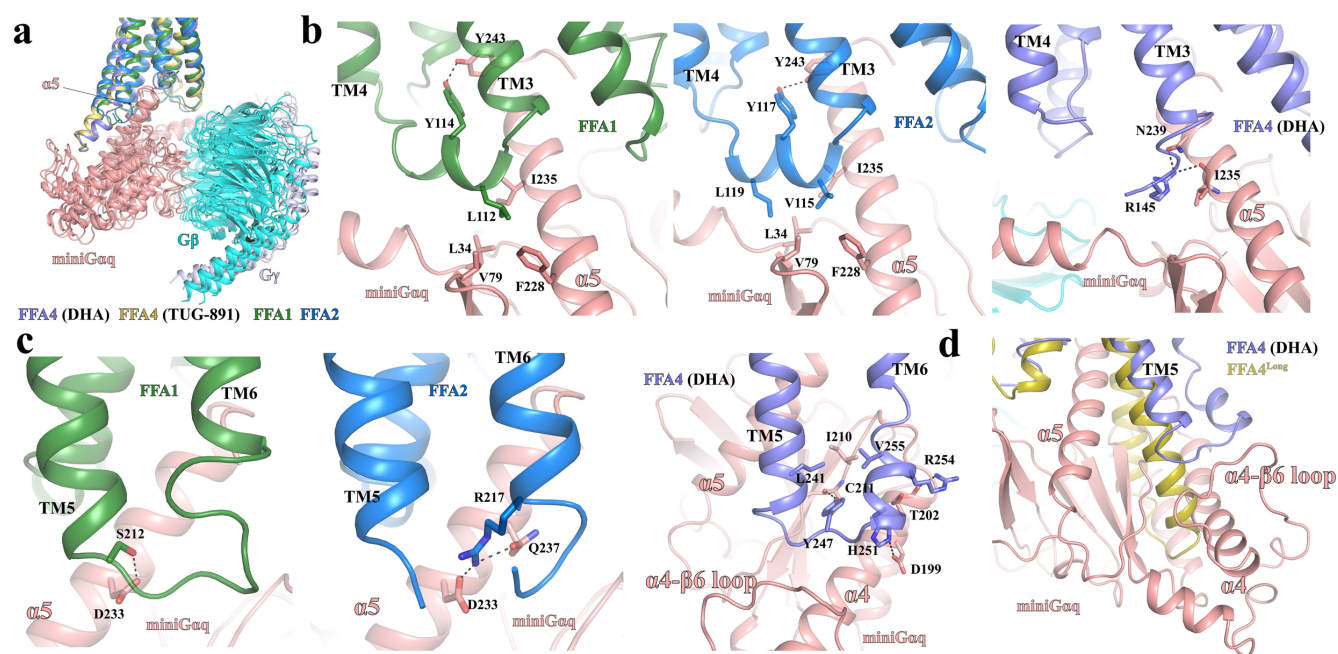


**Figure 5. Agonist recognition of FFA4 and FFA2 probed by MD simulations.** (a) FFA4-DHA, (b) FFA4-TUG-891, and (c) FFA2-butyrate complexes. A representative frame is shown with key residues forming contacts with DHA (orange) with DHA carbon-carbon double bonds numbered 1-6, TUG-891 (purple) and butyrate (yellow) in stick representation. The size and color of the residues correspond to the average strength of van der Waals and electrostatic interactions with the agonist, respectively. Water clusters observed in the MD simulations are shown in the cyan surface-like representation. The superscripts in the amino acid labels denote the Ballesteros–Weinstein generic GPCR residue numbering.



**Figure 6. Activation of FFAs.** (a) Superimposition of the active DHA-bound FFA4 structure (slate) to the AlphaFold predicted inactive FFA4 structure FFA4-AF (light grey) viewed from the intracellular (left) and the extracellular (right) sides. (b) Residues involved in the receptor activation at the core region of FFA4. (c) Superimposition of the active butyrate-bound FFA2 structure (blue) to the AlphaFold predicted FFA2 structure FFA2-AF (dark grey) viewed from the intracellular (left) and the extracellular (right) sides. (d) Superimposition of the active DHA-bound FFA1 structure (green) to the AlphaFold predicted FFA1 structure FFA1-AF (light blue) viewed from the intracellular side. Red solid and dash arrows represent conformational changes of TMs and individual residues, respectively, from the AlphaFold predicted structures to the active agonist-bound structures of FFA1, FFA2, and FFA4.



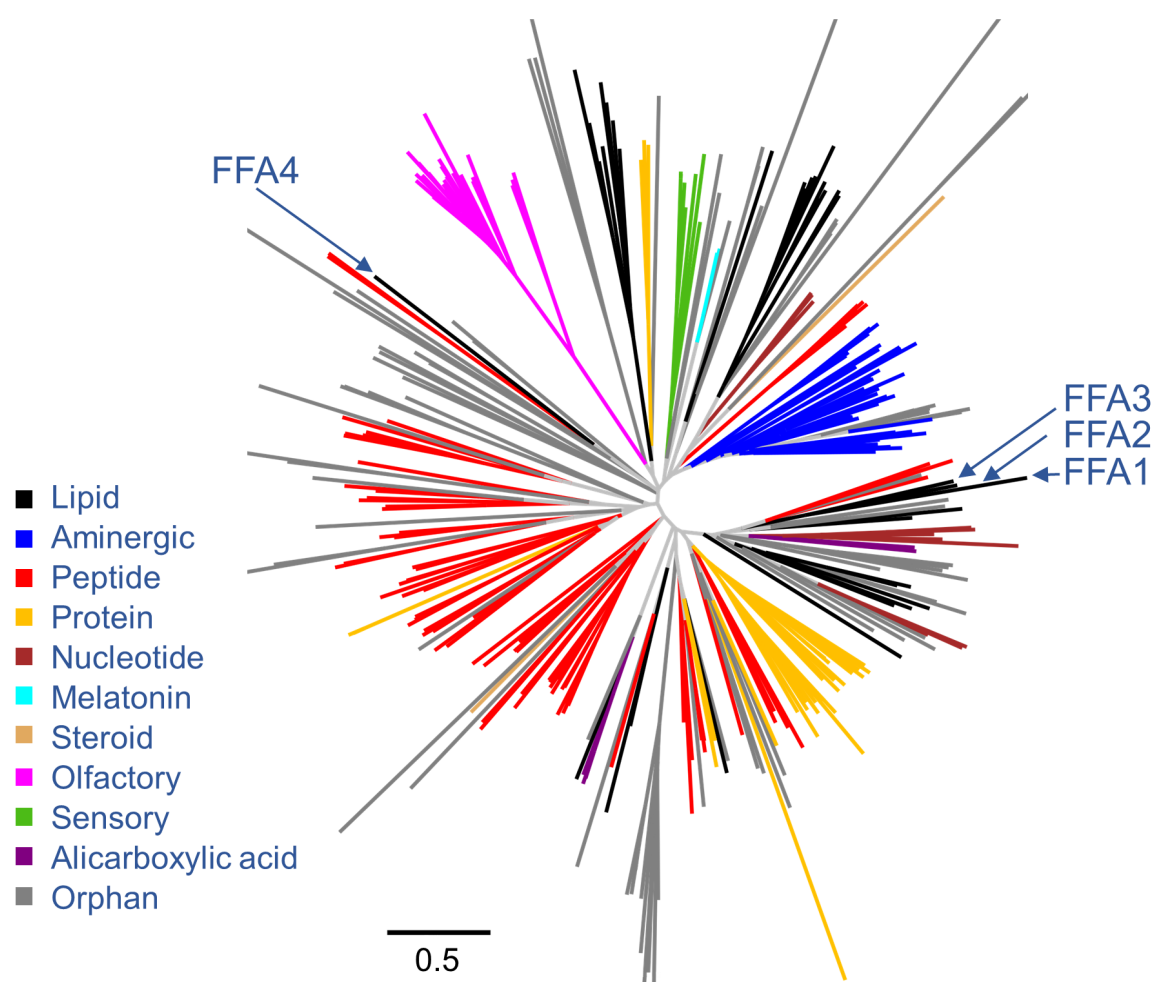


**Figure 7. Differences in the coupling of miniG<sub>q</sub> to FFAs.** (a) Alignment of the structures of FFA1, FFA2, and FFA4 coupled with miniG<sub>q</sub> based on the receptors. (b) Differences in the interactions between miniG<sub>αq</sub> and ICL2 of FFA1, FFA2, and FFA4. (c) Differences in the interactions between miniG<sub>αq</sub> and ICL3 of FFA1, FFA2, and FFA4. (d) Superimposition of the AlphaFold predicted structure of FFA4<sup>Long</sup> to the structure of DHA-bound FFA4 coupled with miniG<sub>q</sub>. MiniG<sub>αq</sub>, G<sub>β</sub> and G<sub>γ</sub> subunits are colored in salmon, cyan and light blue, respectively. The colors of receptors and ligands are indicated in each panel.

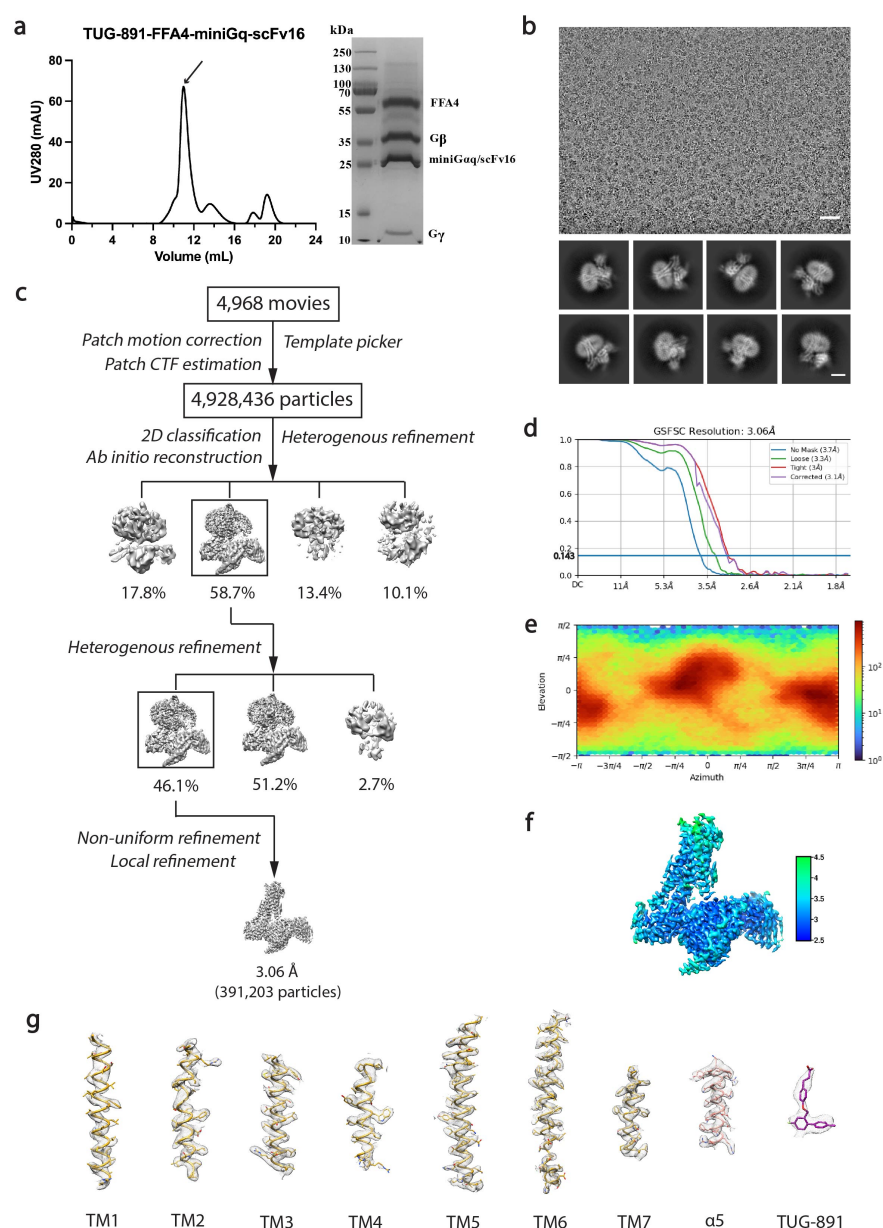
## **Supplemental figures**

### **Structural basis for the ligand recognition and signaling of free fatty acid receptors**

Xuan Zhang, Abdul-Akim Guseinov, Laura Jenkins, Kunpeng Li, Irina G. Tikhonova, Graeme Milligan, Cheng Zhang

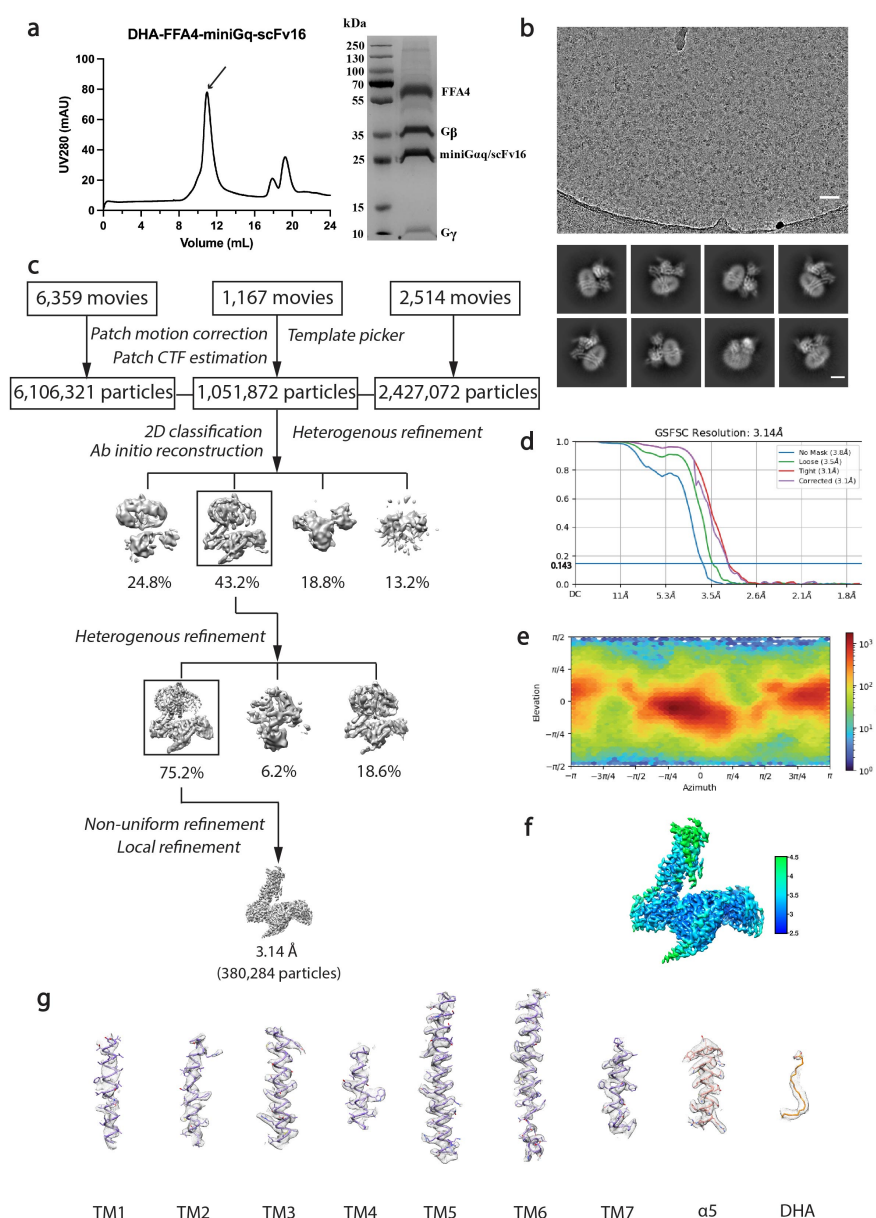


**Figure S1. Phylogenetic analysis of the FFA receptor family.** Clustering of full-length sequences of human class A GPCRs (312 in total) was done using the sequences and tools provided by the GPCRdb web server (<https://gpcrdb.org/>). An unrooted phylogenetic tree with an increasing node order was built using FigTree v1.4.4(<http://tree.bio.ed.ac.uk/>). All GPCRs were colored according to their GPCRdb ligand type. Nodes with descendants sharing the same GPCRdb ligand type were colored accordingly, with the remaining nodes colored grey.



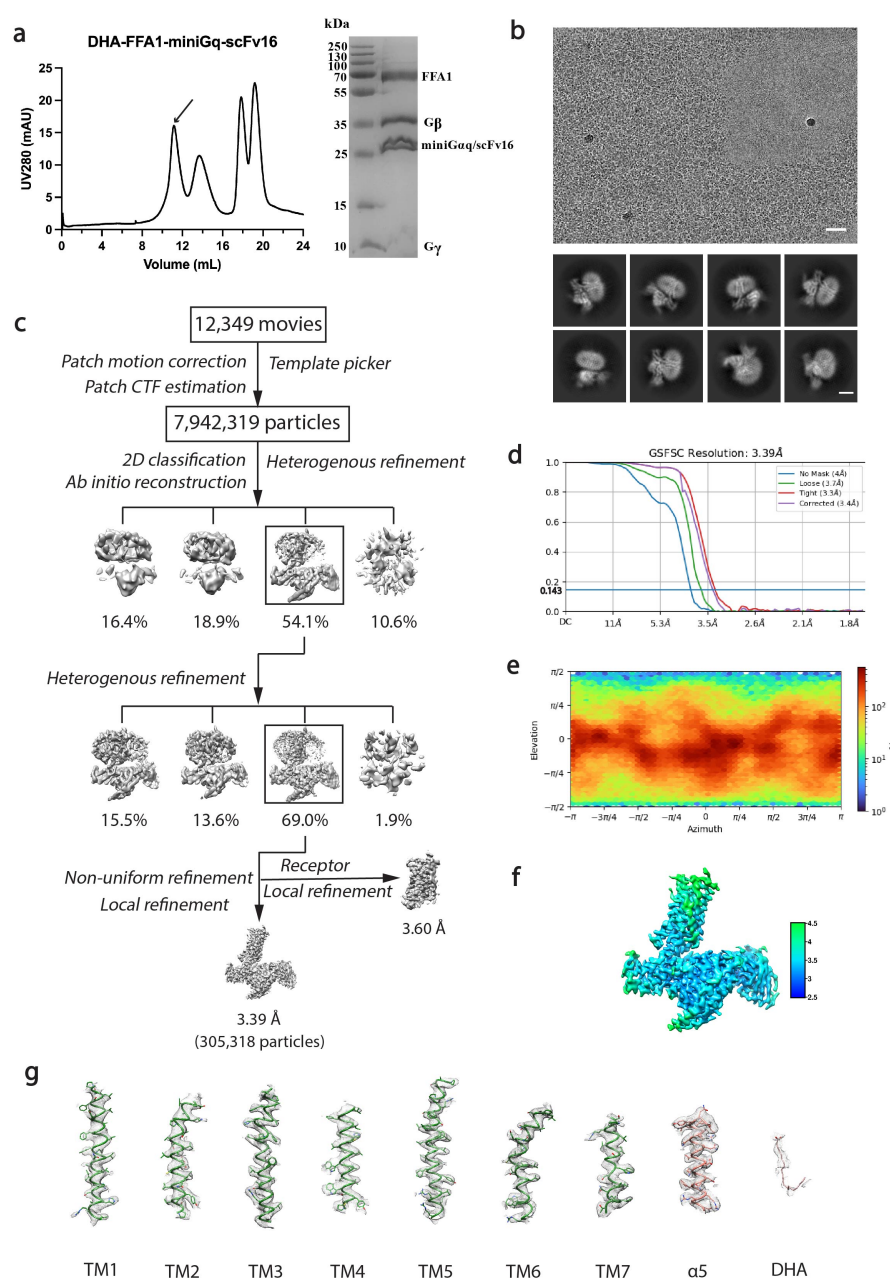
**Figure S2. Purification of the FFA4-miniGq complex with TUG-891 and cryo-EM data processing.**

(a) Size-exclusion chromatography profile and SDS-PAGE analysis of the purified TUG-891-FFA4-miniGq complex. (b) Representative cryo-EM micrograph (scale bar: 50 nm) and 2D class averages (scale bar: 5 nm). (c) Cryo-EM image processing workflow for the TUG-891-FFA4-miniGq complex. (d) Gold-standard Fourier shell correlation (FSC) curve showing an overall resolution is 3.06 Å at FSC=0.143. (e) Angular distribution of the particles used in the final reconstruction. (f) Density map according to local resolution estimation. (g) Cryo-EM density maps and models of the seven transmembrane helices (TM1-7), α5 helix of miniGq and the ligand of TUG-891 bound FFA4-miniGq complex are shown. The EM density is shown at 0.147 threshold.

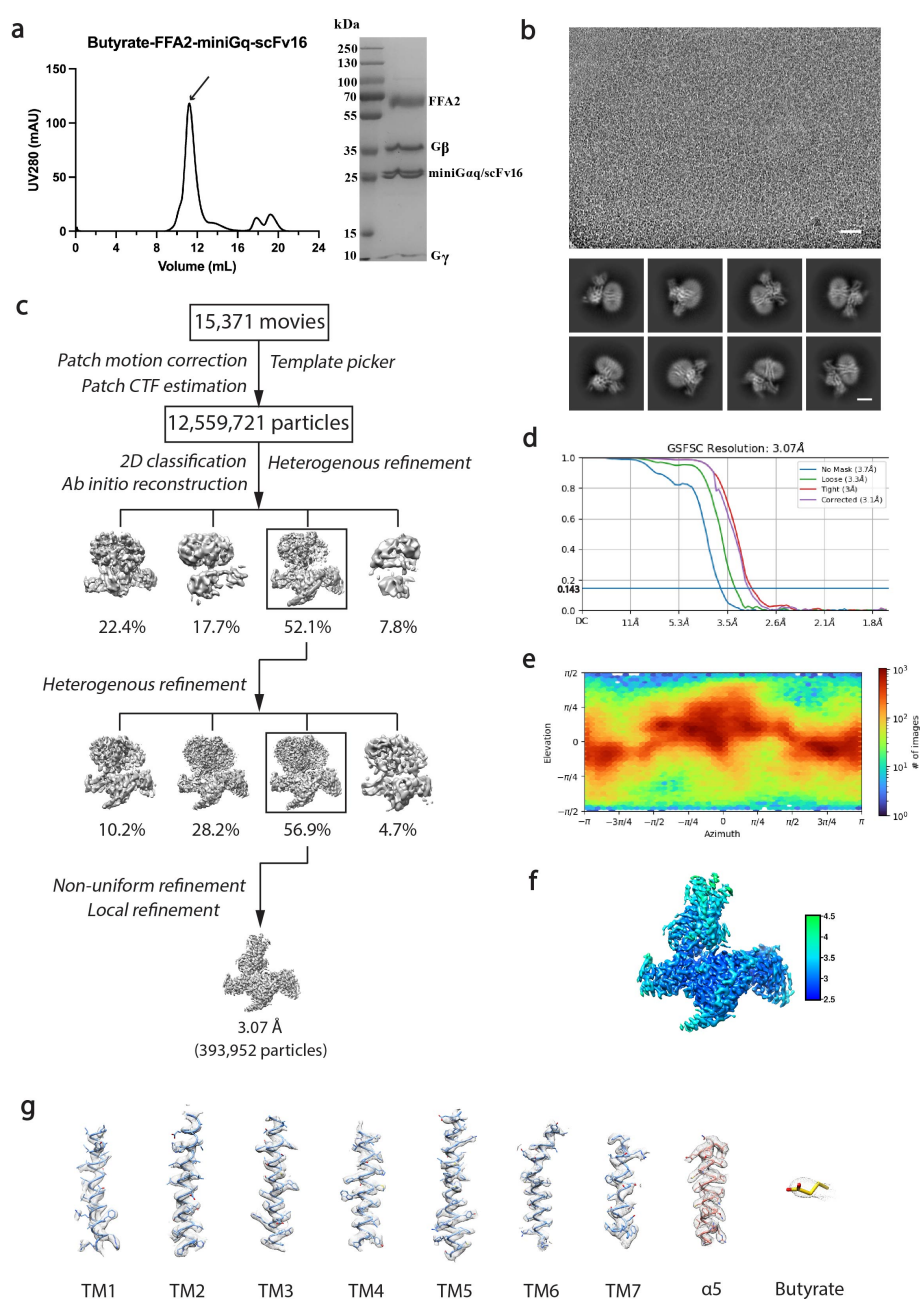


**Figure S3. Purification of the FFA4-miniGq complex with DHA and cryo-EM data processing.** (a) Size-exclusion chromatography profile and SDS-PAGE analysis of the purified DHA-FFA4-miniGq complex. (b) Representative cryo-EM micrograph (scale bar: 50 nm) and 2D class averages (scale bar: 5 nm). (c) Cryo-EM image processing workflow for the DHA-FFA4-miniGq complex. (d) Gold-standard Fourier shell correlation (FSC) curve showing an overall resolution is 3.14 Å at FSC=0.143. (e) Angular distribution of the particles used in the final reconstruction. (f) Density map according to local resolution estimation. (g) Cryo-EM density maps and models of the seven transmembrane helices (TM1-7),  $\alpha 5$  helix of miniGq and the ligand of DHA bound FFA4-miniGq complex are shown. The EM density is shown at 0.132 threshold.

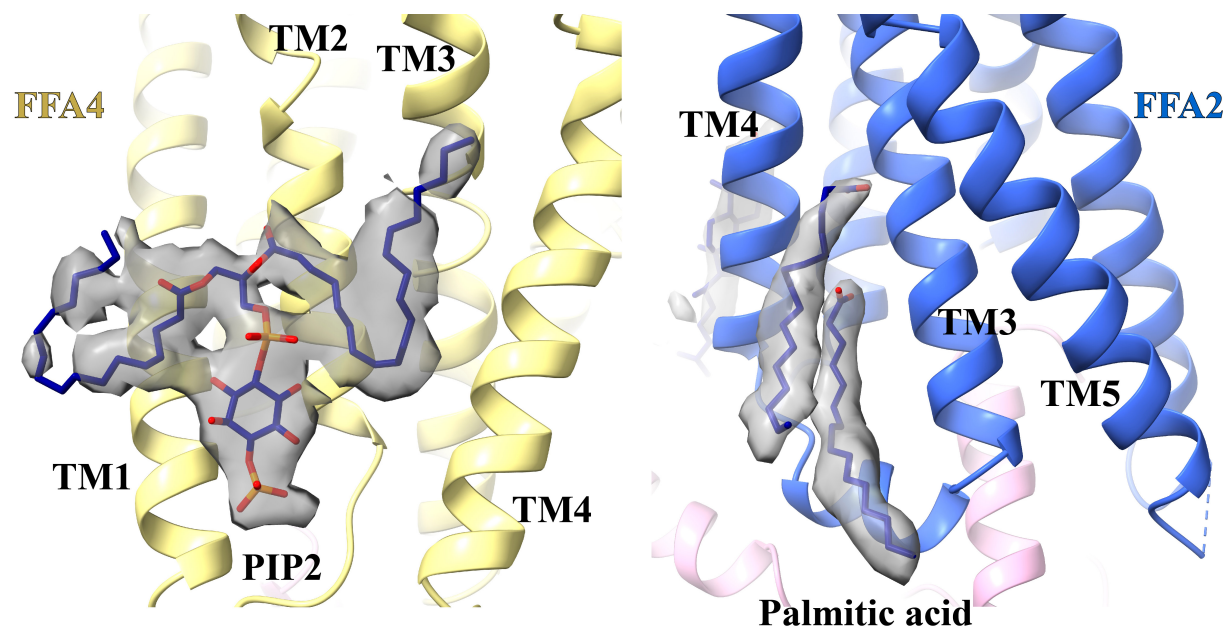




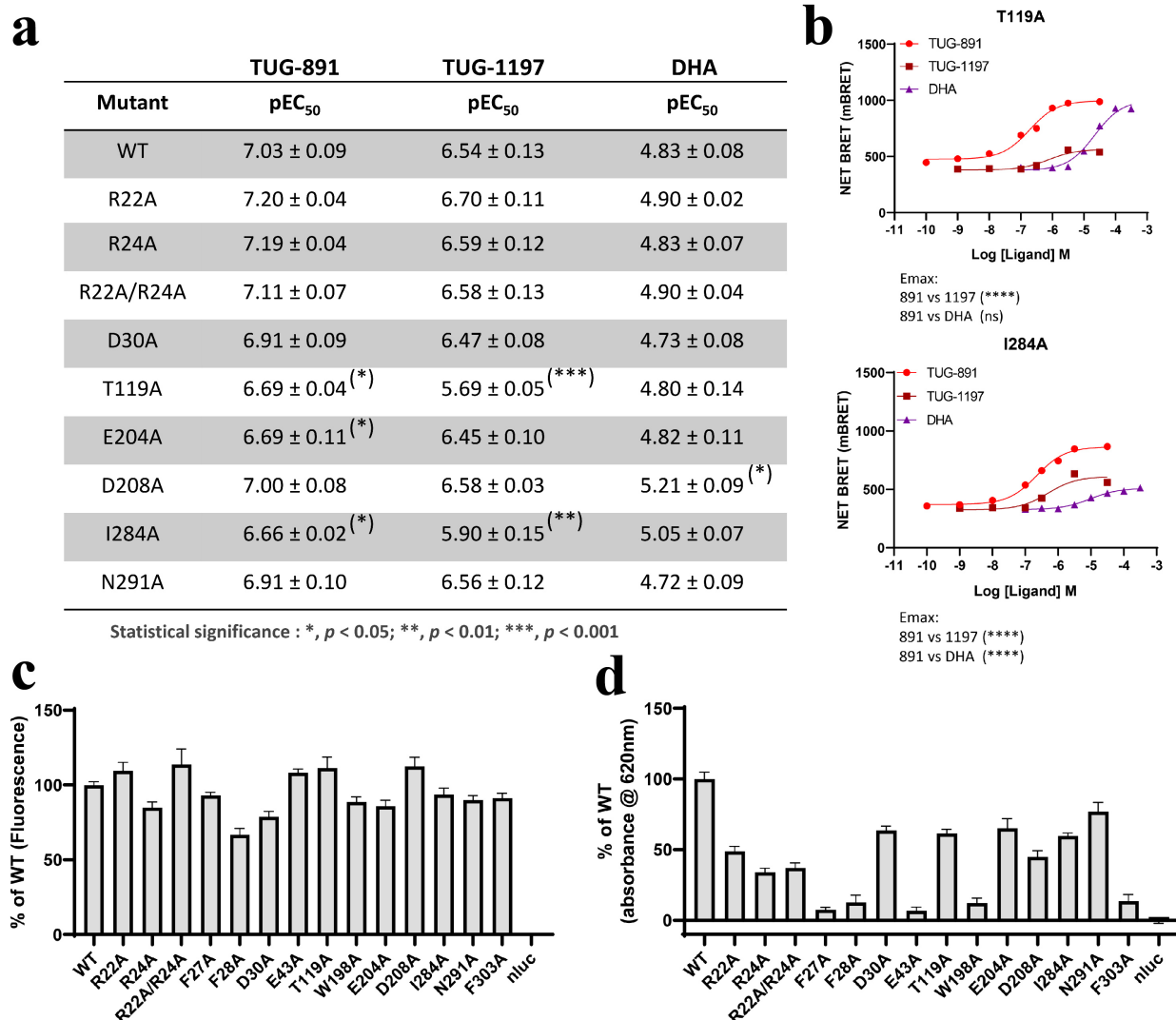
**Figure S4. Purification of the FFA1-miniGq complex with DHA and cryo-EM data processing.** (a) Size-exclusion chromatography profile and SDS-PAGE analysis of the purified DHA-FFA1-miniGq complex. (b) Representative cryo-EM micrograph (scale bar: 50 nm) and 2D class averages (scale bar: 5 nm). (c) Cryo-EM image processing workflow for the DHA-FFA1-miniGq complex. (d) Gold-standard Fourier shell correlation (FSC) curve showing an overall resolution is 3.39 Å at FSC=0.143. (e) Angular distribution of the particles used in the final reconstruction. (f) Density map according to local resolution estimation. (g) Cryo-EM density maps and models of the seven transmembrane helices (TM1-7), α5 helix of miniGq and the ligand of DHA bound FFA1-miniGq complex are shown. The EM density is shown at 0.128 threshold.



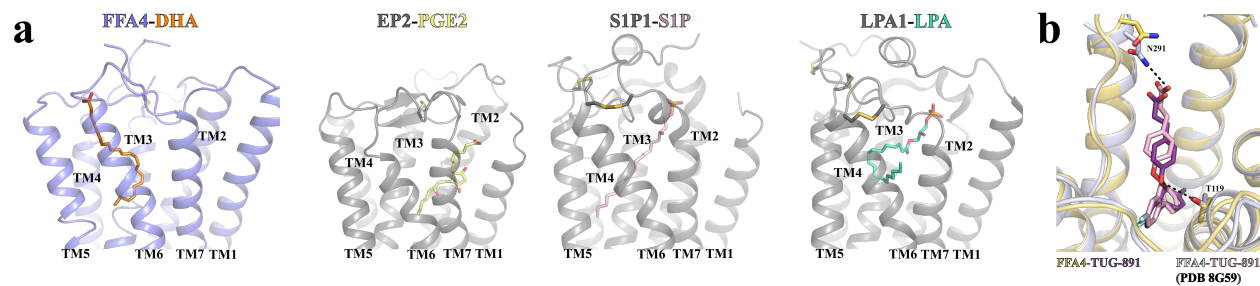
**Figure S5. Purification of the FFA2-miniGq complex with butyrate and cryo-EM data processing.** (a) Size-exclusion chromatography profile and SDS-PAGE analysis of the purified Butyrate-FFA2-miniGq complex. (b) Representative cryo-EM micrograph (scale bar: 50 nm) and 2D class averages (scale bar: 5 nm). (c) Cryo-EM image processing workflow for the Butyrate-FFA2-miniGq complex. (d) Gold-standard Fourier shell correlation (FSC) curve showing an overall resolution is 3.07 Å at FSC=0.143. (e) Angular distribution of the particles used in the final reconstruction. (f) Density map according to local resolution estimation. (g) Cryo-EM density maps and models of the seven transmembrane helices (TM1-7),  $\alpha$ 5 helix of miniGq and the ligand of Butyrate bound FFA2-miniGq complex are shown. The EM density is shown at 0.152 threshold.



**Figure S6. Lipid molecules surrounding 7TMs of FFA2 (blue) and FFA4 (yellow).** PIP2 and palmitic acid molecules shown as dark blue sticks are modeled to fit the cryo-EM density (dark grey). The cryo-EM density of PIP2 is contoured at level 0.12. The cryo-EM density of palmitic acid is contoured at level 0.15.

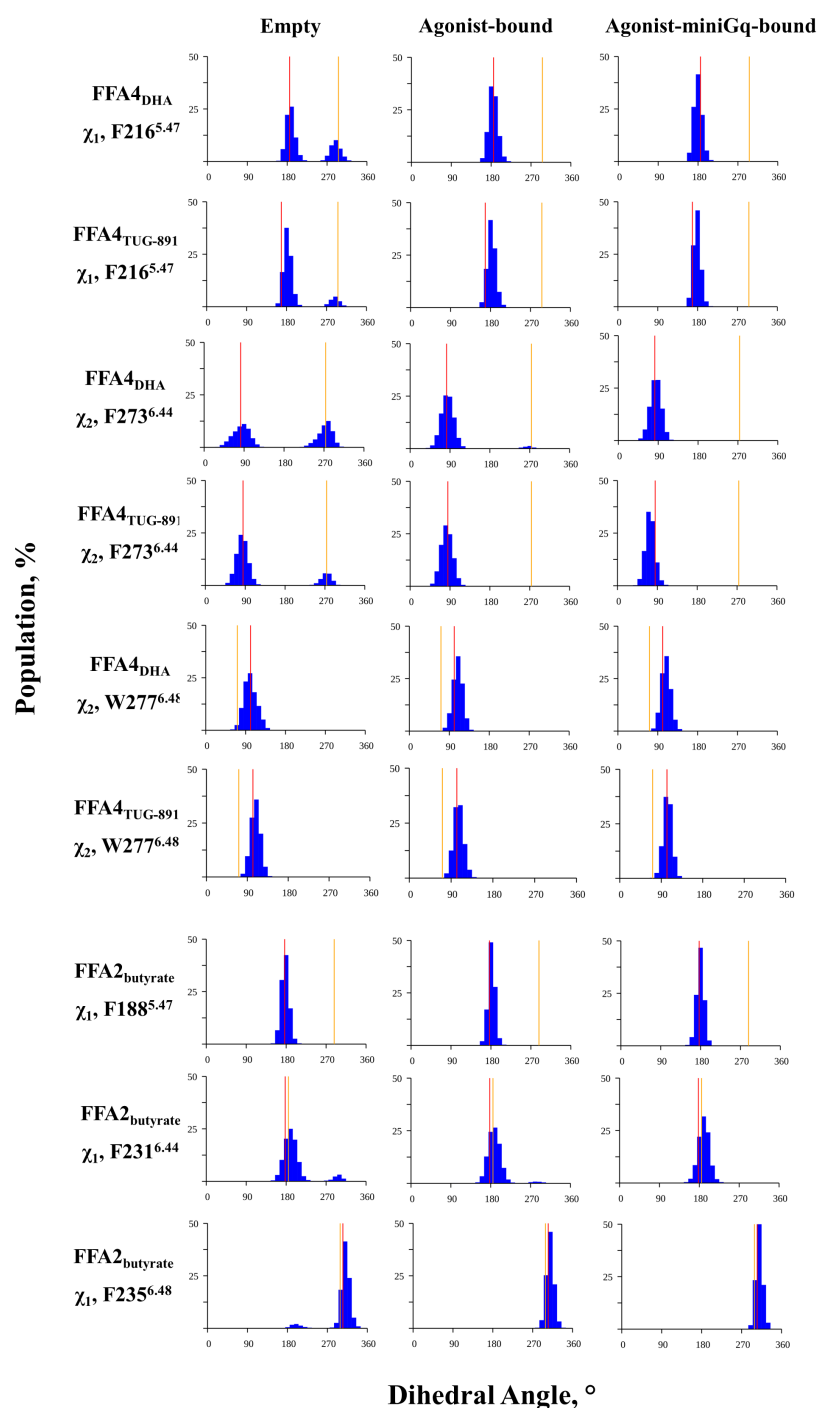


**Figure S7. Mutagenesis studies on FFA4.** (a) Point mutants of FFA4 were compared to wild type (WT) in arrestin 3 interaction studies to explore potential alterations in potency for TUG-891, TUG-1197 and DHA. Data are means  $\pm$  S.E.M. for  $n = 3$  or more. Significantly different from wild type at  $p < 0.05$  \*,  $< 0.01$  \*\* and  $0.001$  \*\*\*. (b) Concentration-response curves for TUG-891, TUG-1197 and DHA at T119A (upper panel) and I284A (lower panel) highlights the reduced efficacy of DHA and TUG1197 at I284A but only for TUG-1197 at T119A. (c) whilst all FFA4-eYFP mutants were expressed effectively as measured by eYFP fluorescence, (d) only a subset were effectively delivered to the cell surface as quantified by ELISA measurements. We were thus unable to define the effect of mutation to Ala of each of F27, F28 and E43 (see text for details).

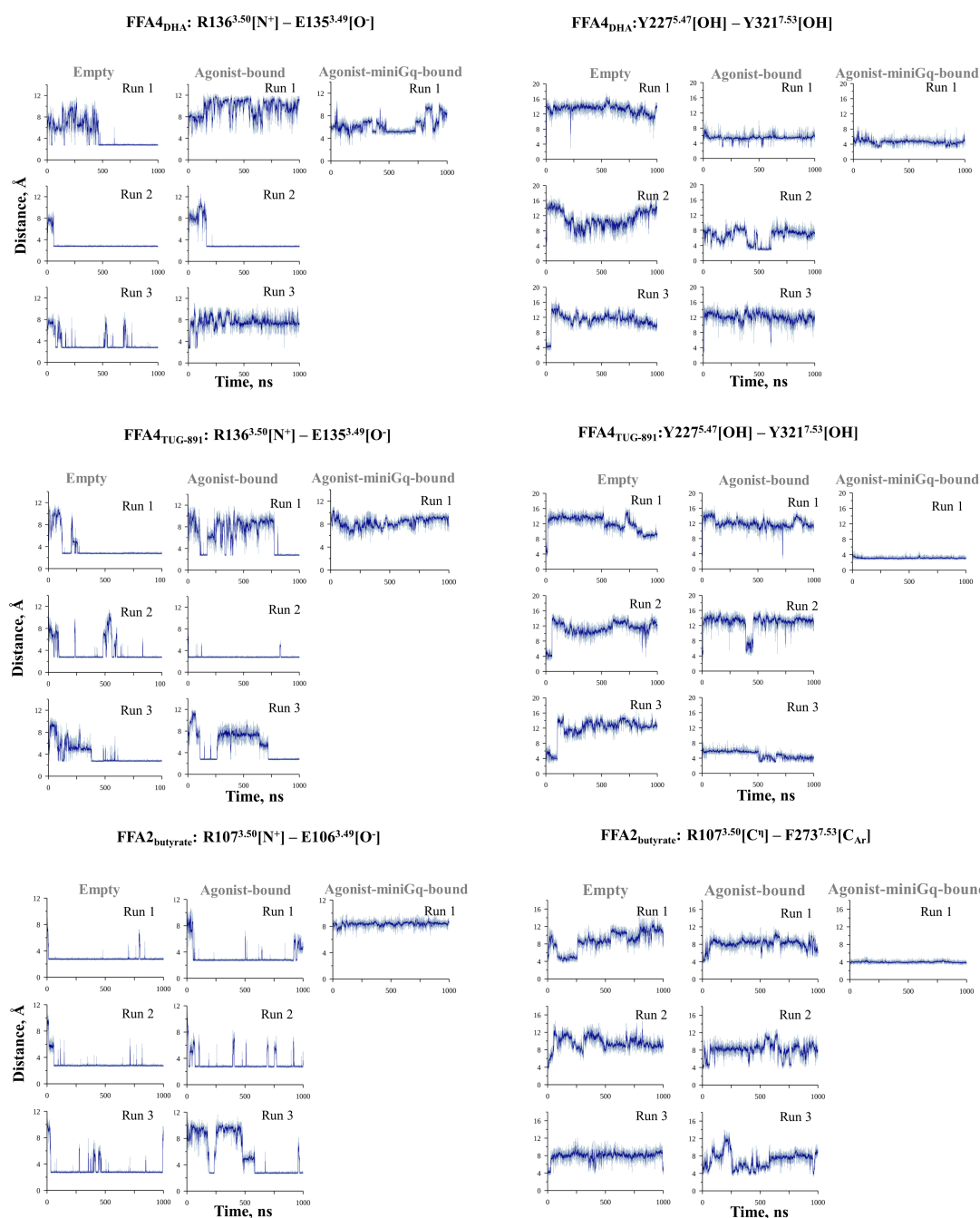


**Figure S8. Structural comparison of FFA4 with (a) other lipid GPCRs and (b) published structure of FFA4 with TUG-891.** EP2 is the receptor for the prostaglandin E2 (PGE2). S1P1 and LPA1 are receptors for the lysophospholipids S1P and LPA, respectively. The PDB IDs of the structures of EP2-PGE2, S1P1-S1P, and LPA1-LPA are 7XC2, 7TD3, and 7TD0, respectively. TUG-891 in our structure and in the published structure (PDB ID 8G59) is colored purple and pink, respectively.

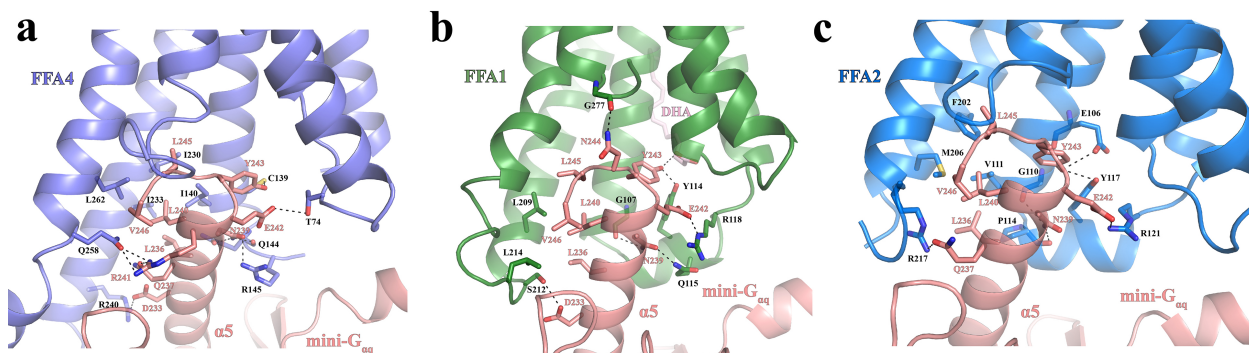




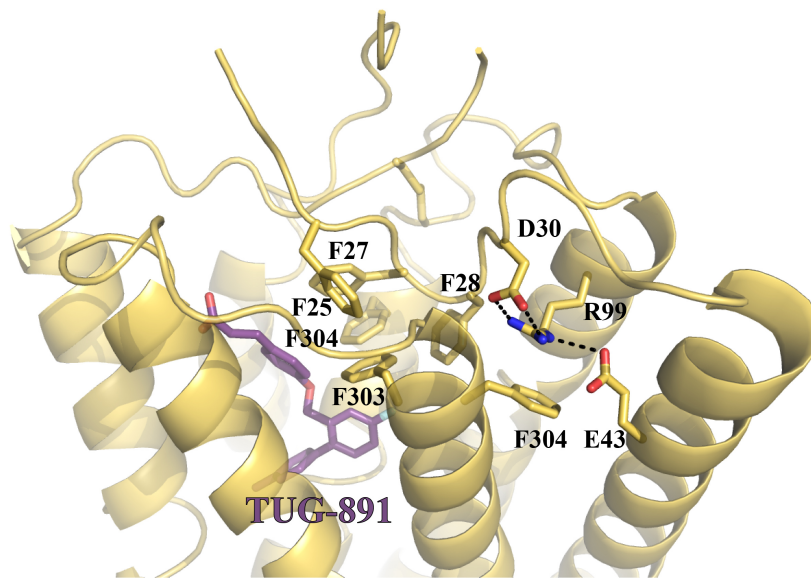
**Figure S9. Dihedral angle population in MD simulations.** The dihedral  $\chi_1$  or  $\chi_2$  angles of the aromatic residues in positions 5.47, 6.44, 6.45 and 6.48 are shown from the three 1  $\mu$ s MD simulations of FFA4 and FFA2 in the empty, agonist- and agonist-miniG<sub>q</sub>-bound forms. The populations are calculated with bin width of 10°. The red and orange lines show a  $\chi_1$  /  $\chi_2$  value observed in the cryoEM active and AlphaFold inactive structures, respectively.



**Figure S10. Evolution of atom distances in MD simulations.** The distance between atoms of conserved amino acids known to be important for GPCR activation are shown from the three 1  $\mu$ s MD simulations of FFA4 and FFA2 in the empty, agonist- and agonist-miniG<sub>q</sub>-bound forms. The sets of atoms used to measure the minimum inter-atomic distance in each frame are given in brackets. The thin line represents the raw values while bold line represents the shifting average of 10 consecutive frames.



**Figure S11. Interactions between FFAs and the  $\alpha 5$  helix of mini-G $\alpha_q$ .** (a) Interactions between FFA4 and the  $\alpha 5$  helix of mini-G $\alpha_q$ . Specifically, hydrophobic interactions form among the mini-G $\alpha_q$  residues L236, L240, L245, and V246 and the FFA4 residues I140<sup>3,54</sup>, I230<sup>5,61</sup>, I233<sup>5,64</sup>, and L262<sup>6,33</sup>. In addition, Y243, E242, and D233 of mini-G $\alpha_q$  form polar interactions with the side chains of FFA4 residues C139<sup>3,53</sup>, T74<sup>2,39</sup>, and R240<sup>5,71</sup>, respectively. Furthermore, N239 of mini-G $\alpha_q$  forms polar interactions with FFA4 residues R145<sup>ICL2</sup> and Q144<sup>ICL2</sup>, whereas R241 and Q237 of mini-G $\alpha_q$  form a polar interaction network with FFA4 residue Q258<sup>6,29</sup>. (b) Interactions between FFA1 and the  $\alpha 5$  helix of mini-G $\alpha_q$ . Specifically, hydrophobic interactions form among the mini-G $\alpha_q$  residues L236, L240, L245, and V246 and the FFA1 residues L209<sup>5,65</sup> and L214<sup>ICL3</sup>. Y243, E242, N239, and D233 of mini-G $\alpha_q$  form polar interactions with the side chains of FFA1 residues Y114<sup>ICL2</sup>, R118<sup>ICL2</sup>, Q115<sup>ICL2</sup>, and S212<sup>ICL3</sup>, respectively. Furthermore, N244 and N239 of mini-G $\alpha_q$  forms polar interactions with the carbonyl groups of FFA1 residues G277<sup>7,54</sup> and G107<sup>3,53</sup>, respectively. Y243 of mini-G $\alpha_q$  also forms a hydrogen bond with DHA that may be bound at 'Site 2'. (c) Interactions between FFA2 and the  $\alpha 5$  helix of mini-G $\alpha_q$ . Specifically, hydrophobic interactions form among the mini-G $\alpha_q$  residues L236, L240, L245, and V246 and the FFA2 residues M206<sup>5,65</sup> and F202<sup>5,61</sup>. In addition, E242 and Q237 of mini-G $\alpha_q$  form polar interactions with the side chains of FFA2 residues R121<sup>ICL2</sup> and R217<sup>6,30</sup>, respectively. Furthermore, Y243 of mini-G $\alpha_q$  forms polar interactions with FFA2 residues E106<sup>3,49</sup> and Y117<sup>ICL2</sup>, whereas 239 of mini-G $\alpha_q$  form polar interactions with the mainchain carbonyl groups of FFA4 residue G110<sup>3,53</sup> and P114<sup>ICL2</sup>.



### Figure S12. Interactions mediated by R99 of FFA4 and nearby aromatic network

In contrast to initial predictions based on homology modeling and mutagenesis R99<sup>2,64</sup> does not interact directly with the carboxylate of TUG-891. Rather it acts to shape the binding pocket through interactions with D30<sup>N</sup> and E43<sup>1,35</sup> and a cation- $\pi$  interaction with F304<sup>7,36</sup>.

**Table S1. Cryo-EM data collection and refinement statistics 1**

	(EMD-41007, PDB 8T3O)	(EMD-41008, PDB 8T3Q)
Data collection and processing	TUG-891-FFA4-miniGq	DHA-FFA4-miniGq
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	55	55
Defocus range (μm)	-1.0 to -1.8	-1.0 to -1.8
Pixel size (Å)	0.828	0.828
Symmetry imposed	C1	C1
Initial particle images (no.)	4,928,436	9,585,265
Final particle images (no.)	391,203	380,284
Map resolution (Å)	3.06	3.14
FSC threshold	0.143	0.143
Map resolution range (Å)	2.5-4.5	2.5-4.5
Refinement		
Model resolution (Å)	3.0	3.1
FSC threshold	0.143	0.143
Model composition		
Non-hydrogen atoms	9112	9032
Protein residues	1158	1158
Ligand	1	1
R.m.s. deviations		
Bond lengths (Å)	0.005	0.008
Bond angles (°)	0.802	1.160
Validation		
MolProbity score	1.65	2.00
Clashscore	8.65	11.84
Rotamer outliers (%)	0.21	1.44
Ramachandran plot		
Favored (%)	96.93	95.79
Allowed (%)	3.07	4.21
Disallowed (%)	0	0

**Table S2. Cryo-EM data collection and refinement statistics 2**



	(EMD-41013, PDB 8T3V)/(EMD-41014)	(EMD-41010, PDB 8T3S)
Data collection and processing	DHA-FFA1-miniGq/Local	butyrate-FFA2-miniGq
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	61.6	61.6
Defocus range (μm)	-0.8 to -2.5	-0.8 to -2.5
Pixel size (Å)	0.826	0.826
Symmetry imposed	C1	C1
Initial particle images (no.)	7,942,319	12,559,721
Final particle images (no.)	305,318	393,952
Map resolution (Å)	3.39/3.6	3.07
FSC threshold	0.143	0.143
Map resolution range (Å)	2.5-4.5	2.5-4.5
Refinement		
Model resolution (Å)	3.4	3.0
FSC threshold	0.143	0.143
Model composition		
Non-hydrogen atoms	8745	8772
Protein residues	1136	1119
Ligand	1	1
R.m.s. deviations		
Bond lengths (Å)	0.007	0.007
Bond angles (°)	0.873	0.797
Validation		
MolProbity score	1.69	1.60
Clashscore	11.48	9.17
Rotamer outliers (%)	0.44	0.53
Ramachandran plot		
Favored (%)	97.41	97.45
Allowed (%)	2.59	2.55
Disallowed (%)	0	0

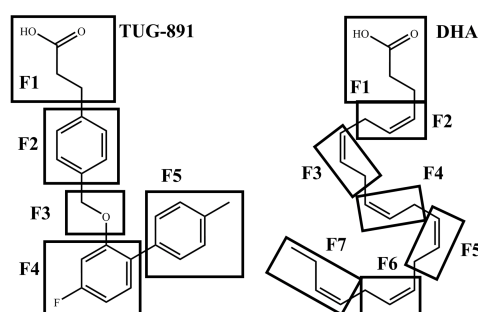
**Table S3. The average root mean square deviation and fluctuation (RMSD and RMSF) of the receptor systems.** RMSD and RMSF values were calculated for the receptor Cα atoms, the 7-transmembrane bundle (7TMB) Cα atoms and the ligand non-hydrogen atoms (Ligand) from the three 1μ MD simulations for receptor-only systems and one 1 μs MD simulation for receptor-miniGq systems.

RMSD values were averaged within each replica and combined between different replicas to yield the mean  $\pm$  standard deviation values given below. RMSF values are also given as mean  $\pm$  standard deviation between all replicas.

Receptor Systems	RMSD, Å			RMSF, Å		
	C $\alpha$	7TMB-C $\alpha$	Ligand	C $\alpha$	7TMB-C $\alpha$	Ligand
FFA4 <sub>DHA</sub>	3.6 $\pm$ 0.2	1.3 $\pm$ 0.0	3.9 $\pm$ 1.1	1.4 $\pm$ 0.1	0.7 $\pm$ 0.0	1.9 $\pm$ 0.2
FFA4 <sub>empty</sub>	3.7 $\pm$ 0.1	1.5 $\pm$ 0.0		1.6 $\pm$ 0.2	0.8 $\pm$ 0.1	
FFA4 <sub>DHA_Gq</sub>	3.4	1.2	4.0	1.4	0.7	2.2
FFA4 <sub>TUG-891</sub>	3.7 $\pm$ 0.3	1.2 $\pm$ 0.0	1.9 $\pm$ 0.2	1.3 $\pm$ 0.1	0.7 $\pm$ 0.0	1.0 $\pm$ 0.1
FFA4 <sub>empty_TUG-891</sub>	3.2 $\pm$ 0.4	1.3 $\pm$ 0.1		1.5 $\pm$ 0.2	0.8 $\pm$ 0.1	
FFA4 <sub>TUG-891_Gq</sub>	2.8 $\pm$	1.2	1.9	1.1	0.6	1.0
FFA2 <sub>butyrate</sub>	2.6 $\pm$ 0.2	1.1 $\pm$ 0.0	2.2 $\pm$ 0.4	1.2 $\pm$ 0.0	0.7 $\pm$ 0.0	1.6 $\pm$ 0.3
FFA2 <sub>empty</sub>	2.8 $\pm$ 0.1	1.3 $\pm$ 0.2		1.2 $\pm$ 0.1	0.8 $\pm$ 0.1	
FFA2 <sub>butyrate_Gq</sub>	2.2	1.2	1.9	0.9	0.6	1.5

**Table S4. The average ligand fragment-residue interaction energy.** The interaction energy involving electrostatic (ele) and van der Waals (vdW) components in kcal/mol was calculated from three 1 $\mu$  MD simulations of the FFA4-agonist complexes. The fragmentation of the agonists is shown below. The residues having noticeable interactions with a given ligand fragment, i.e. having total energy below -1 kcal/mol or above 1 kcal/mol are shown in bold. Among them, residues having repulsive interactions, i.e. positive total energy, are colored red.

The values were averaged within each replica and combined between different replicas to yield the mean  $\pm$  standard deviation values given below.



TUG-891, F1				TUG-891, F2			
	Ele	vdW	total		ele	vdW	total
R22	-45 $\pm$ 21	1.0 $\pm$ 1.0	-44 $\pm$ 20	I284	-0.3 $\pm$ 0.0	-2.0 $\pm$ 0.0	-2.3 $\pm$ 0.0
R24	-33 $\pm$ 5	-0.4 $\pm$ 0.2	-33 $\pm$ 5	F27	-0.2 $\pm$ 0.0	-0.8 $\pm$ 0.3	-1.1 $\pm$ 0.3
W198	-9.1 $\pm$ 2.8	-0.3 $\pm$ 0.1	-9.5 $\pm$ 2.9	W207	-0.3 $\pm$ 0.0	-0.6 $\pm$ 0.1	-0.9 $\pm$ 0.2
D208	-2.3 $\pm$ 0.2	-0.3 $\pm$ 0.0	-2.6 $\pm$ 0.2	F303	-0.1 $\pm$ 0.0	-0.6 $\pm$ 0.1	-0.7 $\pm$ 0.1
F27	-1.7 $\pm$ 0.0	-0.7 $\pm$ 0.1	-2.4 $\pm$ 0.1	F115	-0.2 $\pm$ 0.1	-0.5 $\pm$ 0.3	-0.7 $\pm$ 0.4
I287	-0.8 $\pm$ 0.6	-1.4 $\pm$ 0.1	-2.2 $\pm$ 0.7	L196	-0.1 $\pm$ 0.0	-0.4 $\pm$ 0.1	-0.5 $\pm$ 0.2
N291	-1.9 $\pm$ 0.5	-0.1 $\pm$ 0.0	-1.9 $\pm$ 0.5	I287	0.1 $\pm$ 0.0	-0.6 $\pm$ 0.1	-0.5 $\pm$ 0.1
F25	-1.1 $\pm$ 2.1	-0.7 $\pm$ 0.1	-1.8 $\pm$ 2.1	L173	0.2 $\pm$ 0.1	-0.7 $\pm$ 0.1	-0.5 $\pm$ 0.0
W207	-0.6 $\pm$ 0.1	-0.2 $\pm$ 0.0	-0.8 $\pm$ 0.1	F211	-0.2 $\pm$ 0.0	-0.3 $\pm$ 0.1	-0.5 $\pm$ 0.1
L173	-0.2 $\pm$ 0.1	-0.2 $\pm$ 0.1	-0.4 $\pm$ 0.1	F25	-0.2 $\pm$ 0.0	-0.2 $\pm$ 0.1	-0.4 $\pm$ 0.1
F115	-0.1 $\pm$ 0.2	-0.1 $\pm$ 0.1	-0.2 $\pm$ 0.2	I280	0.2 $\pm$ 0.0	-0.5 $\pm$ 0.3	-0.4 $\pm$ 0.2
F303	-0.0 $\pm$ 0.0	-0.1 $\pm$ 0.1	-0.1 $\pm$ 0.1	T283	-0.0 $\pm$ 0.0	-0.3 $\pm$ 0.1	-0.3 $\pm$ 0.2
L288	0.3 $\pm$ 0.1	-0.4 $\pm$ 0.5	-0.1 $\pm$ 0.6	T119	-0.0 $\pm$ 0.1	-0.3 $\pm$ 0.1	-0.3 $\pm$ 0.0
I280	-0.1 $\pm$ 0.0	-0.0 $\pm$ 0.0	-0.1 $\pm$ 0.0	D208	-0.0 $\pm$ 0.0	-0.1 $\pm$ 0.0	-0.2 $\pm$ 0.0
<b>L196</b>	<b>2.6 <math>\pm</math> 0.6</b>	<b>-1.4 <math>\pm</math> 0.2</b>	<b>1.2 <math>\pm</math> 0.5</b>	E204	-0.1 $\pm$ 0.0	-0.0 $\pm$ 0.0	-0.2 $\pm$ 0.0
<b>Q290</b>	<b>1.7 <math>\pm</math> 0.9</b>	<b>-0.0 <math>\pm</math> 0.0</b>	<b>1.7 <math>\pm</math> 0.9</b>	I281	-0.1 $\pm$ 0.0	-0.0 $\pm$ 0.0	-0.1 $\pm$ 0.0
<b>E204</b>	<b>28 <math>\pm</math> 1</b>	<b>-0.4 <math>\pm</math> 0.0</b>	<b>27.2 <math>\pm</math> 1.4</b>	L288	-0.0 $\pm$ 0.0	-0.1 $\pm$ 0.1	-0.1 $\pm$ 0.1
TUG-891, F3				TUG-891, F4			
	Ele	vdW	total		ele	vdW	total
F115	-0.9 $\pm$ 0.3	-0.4 $\pm$ 0.2	-1.2 $\pm$ 0.4	M118	-0.7 $\pm$ 0.3	-2.2 $\pm$ 0.2	-2.9 $\pm$ 0.1
T119	-0.8 $\pm$ 0.2	-0.4 $\pm$ 0.1	-1.2 $\pm$ 0.3	I280	-0.2 $\pm$ 0.1	-1.5 $\pm$ 0.2	-1.6 $\pm$ 0.2
I284	-0.4 $\pm$ 0.0	-0.4 $\pm$ 0.1	-0.7 $\pm$ 0.1	V307	-0.4 $\pm$ 0.2	-1.0 $\pm$ 0.2	-1.4 $\pm$ 0.5
T283	-0.7 $\pm$ 0.2	-0.0 $\pm$ 0.0	-0.7 $\pm$ 0.2	F88	-0.3 $\pm$ 0.0	-0.7 $\pm$ 0.1	-1.1 $\pm$ 0.1
F27	-0.3 $\pm$ 0.1	-0.1 $\pm$ 0.1	-0.4 $\pm$ 0.1	T119	0.4 $\pm$ 0.0	-1.3 $\pm$ 0.7	-0.9 $\pm$ 0.8
F303	-0.2 $\pm$ 0.0	-0.1 $\pm$ 0.0	-0.4 $\pm$ 0.1	W277	-0.3 $\pm$ 0.1	-0.4 $\pm$ 0.1	-0.7 $\pm$ 0.1
F211	-0.1 $\pm$ 0.0	-0.2 $\pm$ 0.0	-0.3 $\pm$ 0.1	F311	-0.1 $\pm$ 0.1	-0.4 $\pm$ 0.1	-0.6 $\pm$ 0.2
L120	-0.2 $\pm$ 0.1	-0.0 $\pm$ 0.0	-0.2 $\pm$ 0.1	F303	-0.0 $\pm$ 0.1	-0.5 $\pm$ 0.2	-0.6 $\pm$ 0.2

F25	-0.2 ± 0.0	-0.0 ± 0.0	-0.2 ± 0.0	T310	0.1 ± 0.3	-0.6 ± 0.2	-0.5 ± 0.5
W207	-0.1 ± 0.0	-0.1 ± 0.0	-0.1 ± 0.0	F115	0.3 ± 0.1	-0.7 ± 0.1	-0.5 ± 0.1
TUG-891, F5				DHA, F1			
	Ele	vdW	total		ele	vdW	total
<b>F211</b>	<b>-0.7 ± 0.1</b>	<b>-1.9 ± 0.5</b>	<b>-2.6 ± 0.5</b>	<b>R24</b>	<b>-55 ± 29</b>	<b>-0.2 ± 0.7</b>	<b>-55 ± 28</b>
<b>T119</b>	<b>-0.2 ± 0.1</b>	<b>-1.7 ± 0.2</b>	<b>-1.9 ± 0.2</b>	<b>R22</b>	<b>-33 ± 23</b>	<b>0.3 ± 0.3</b>	<b>-32 ± 23</b>
<b>W277</b>	<b>-0.2 ± 0.2</b>	<b>-1.4 ± 0.2</b>	<b>-1.6 ± 0.4</b>	<b>F25</b>	<b>-2.1 ± 3.7</b>	<b>-1.2 ± 0.2</b>	<b>-3.4 ± 3.8</b>
<b>I280</b>	<b>-0.3 ± 0.2</b>	<b>-1.3 ± 0.1</b>	<b>-1.6 ± 0.2</b>	<b>W198</b>	<b>-2.3 ± 1.1</b>	<b>-0.1 ± 0.1</b>	<b>-2.4 ± 1.1</b>
<b>M118</b>	<b>-0.5 ± 0.2</b>	<b>-0.5 ± 0.2</b>	<b>-1.0 ± 0.4</b>	<b>F27</b>	<b>-0.8 ± 0.7</b>	<b>-0.6 ± 0.2</b>	<b>-1.5 ± 0.9</b>
I281	-0.0 ± 0.1	-0.9 ± 0.1	-0.9 ± 0.2	I287	0.1 ± 0.1	-0.6 ± 0.4	-0.5 ± 0.3
S123	0.1 ± 0.1	-1.1 ± 0.1	-0.9 ± 0.1	F115	-0.0 ± 0.0	-0.1 ± 0.1	-0.1 ± 0.1
N215	0.2 ± 0.1	-0.9 ± 0.2	-0.7 ± 0.1	F303	-0.0 ± 0.0	-0.0 ± 0.0	-0.1 ± 0.0
G122	0.3 ± 0.1	-0.9 ± 0.1	-0.6 ± 0.2	W207	-0.0 ± 0.0	-0.0 ± 0.0	-0.0 ± 0.1
F216	-0.0 ± 0.1	-0.5 ± 0.2	-0.6 ± 0.2	<b>Q290</b>	<b>0.9 ± 0.7</b>	<b>-0.0 ± 0.0</b>	<b>0.9 ± 0.7</b>
V212	-0.1 ± 0.0	-0.3 ± 0.3	-0.4 ± 0.3	<b>D208</b>	<b>1.8 ± 1.6</b>	<b>-0.0 ± 0.0</b>	<b>1.7 ± 1.6</b>
I126	0.3 ± 0.0	-0.8 ± 0.1	-0.4 ± 0.1	<b>L196</b>	<b>4.4 ± 3.0</b>	<b>-1.5 ± 0.1</b>	<b>2.9 ± 2.9</b>
F88	-0.1 ± 0.0	-0.2 ± 0.1	-0.4 ± 0.1	<b>E204</b>	<b>11 ± 7</b>	<b>-0.2 ± 0.1</b>	<b>11 ± 7</b>
DHA, F2				DHA, F3			
	Ele	vdW	total		ele	vdW	total
<b>R24</b>	<b>-0.8 ± 0.1</b>	<b>-0.2 ± 0.3</b>	<b>-1.1 ± 0.4</b>	F27	-0.1 ± 0.0	-0.6 ± 0.0	-0.7 ± 0.0
R22	-0.8 ± 0.6	-0.1 ± 0.1	-0.9 ± 0.6	F115	-0.0 ± 0.1	-0.6 ± 0.3	-0.6 ± 0.2
F27	-0.1 ± 0.1	-0.7 ± 0.0	-0.8 ± 0.1	F25	0.0 ± 0.1	-0.5 ± 0.5	-0.4 ± 0.4
I287	-0.1 ± 0.1	-0.6 ± 0.4	-0.8 ± 0.5	I284	0.0 ± 0.0	-0.4 ± 0.2	-0.4 ± 0.2
F25	0.2 ± 0.3	-0.9 ± 0.1	-0.7 ± 0.2	L196	-0.0 ± 0.0	-0.3 ± 0.1	-0.4 ± 0.1
L196	-0.0 ± 0.0	-0.4 ± 0.1	-0.5 ± 0.1	L173	0.0 ± 0.0	-0.4 ± 0.1	-0.3 ± 0.0
F115	-0.0 ± 0.0	-0.4 ± 0.5	-0.4 ± 0.5	F303	-0.0 ± 0.0	-0.3 ± 0.2	-0.3 ± 0.2
F303	-0.0 ± 0.0	-0.4 ± 0.3	-0.4 ± 0.3	W207	-0.1 ± 0.0	-0.1 ± 0.1	-0.2 ± 0.1
DHA, F4				DHA, F5			
	Ele	vdW	total		ele	vdW	total
<b>F115</b>	<b>-0.1 ± 0.0</b>	<b>-1.0 ± 0.1</b>	<b>-1.1 ± 0.1</b>	<b>M118</b>	<b>-0.2 ± 0.0</b>	<b>-1.2 ± 0.2</b>	<b>-1.4 ± 0.2</b>
M118	0.1 ± 0.0	-0.6 ± 0.0	-0.5 ± 0.0	T119	0.1 ± 0.0	-0.8 ± 0.1	-0.8 ± 0.1
T119	0.1 ± 0.0	-0.5 ± 0.0	-0.4 ± 0.1	F115	-0.1 ± 0.0	-0.5 ± 0.1	-0.7 ± 0.2
F303	-0.1 ± 0.0	-0.3 ± 0.2	-0.4 ± 0.2	I280	0.1 ± 0.0	-0.6 ± 0.1	-0.5 ± 0.0
F27	-0.0 ± 0.0	-0.4 ± 0.0	-0.4 ± 0.0	F303	-0.2 ± 0.1	-0.2 ± 0.1	-0.4 ± 0.2
F25	0.0 ± 0.1	-0.3 ± 0.4	-0.3 ± 0.4	V307	-0.0 ± 0.0	-0.3 ± 0.0	-0.3 ± 0.0
DHA, F6				DHA, F7			
	Ele	vdW	total		ele	vdW	total
<b>M118</b>	<b>-0.4 ± 0.0</b>	<b>-0.8 ± 0.0</b>	<b>-1.1 ± 0.0</b>	<b>G122</b>	<b>-0.7 ± 0.1</b>	<b>-0.8 ± 0.1</b>	<b>-1.5 ± 0.2</b>
<b>T119</b>	<b>-0.1 ± 0.0</b>	<b>-1.1 ± 0.1</b>	<b>-1.1 ± 0.1</b>	<b>W277</b>	<b>-0.1 ± 0.1</b>	<b>-1.2 ± 0.2</b>	<b>-1.3 ± 0.1</b>
I280	-0.0 ± 0.0	-0.8 ± 0.1	-0.8 ± 0.1	<b>F211</b>	<b>-0.2 ± 0.0</b>	<b>-1.0 ± 0.1</b>	<b>-1.2 ± 0.1</b>
W277	-0.2 ± 0.0	-0.4 ± 0.1	-0.6 ± 0.1	<b>I280</b>	<b>-0.2 ± 0.0</b>	<b>-0.9 ± 0.1</b>	<b>-1.0 ± 0.0</b>
F88	-0.1 ± 0.0	-0.4 ± 0.0	-0.5 ± 0.0	S123	-0.1 ± 0.0	-0.7 ± 0.1	-0.8 ± 0.1
F211	-0.0 ± 0.0	-0.4 ± 0.1	-0.4 ± 0.1	N215	-0.0 ± 0.0	-0.7 ± 0.1	-0.7 ± 0.1
G122	0.1 ± 0.0	-0.3 ± 0.0	-0.2 ± 0.1	I281	0.0 ± 0.0	-0.6 ± 0.1	-0.6 ± 0.1
V307	-0.1 ± 0.0	-0.1 ± 0.0	-0.2 ± 0.0	I126	0.1 ± 0.0	-0.6 ± 0.1	-0.5 ± 0.1

