Distinct binding hotspots for natural and synthetic agonists of FFA4 from in silico approaches.

Guillaume Patient, Corentin Bedart, N. A. Khan, Nicolas Renault, Amaury Farce

RESEARCH ARTICLE

molecular
informatics

Distinct binding hotspots for natural and synthetic agonists of FFA4 from in silico approaches

Guillaume Patient[1] 🆔 | **Corentin Bedart[1]** | **Naim. A Khan[2]** 🆔 | **Nicolas Renault[1]** | **Amaury Farce[1]** 🆔

[1]University of Lille, Inserm, CHU Lille, U1286 – INFINITE–Institute for Translational Research in Inflammation, Lille, France

[2]U1231 Inserm, Equipe NuTox, AgroSup, Université de Bourgogne, Dijon, France

Correspondence
Amaury Farce, University of Lille, Inserm, CHU Lille, U1286 – INFINITE–Institute for Translational Research in Inflammation, F-59000 Lille, France.
Email: amaury.farce@univ-lille.fr

Abstract

FFA4 has gained interest in recent years since its deorphanization in 2005 and the characterization of the Free Fatty Acids receptors family for their therapeutic potential in metabolic disorders. The expression of FFA4 (also known as GPR120) in numerous organs throughout the human body makes this receptor a highly potent target, particularly in fat sensing and diet preference. This offers an attractive approach to tackle obesity and related metabolic diseases. Recent cryo-EM structures of the receptor have provided valuable information for a potential active state although the previous studies of FFA4 presented diverging information. We performed molecular docking and molecular dynamics simulations of four agonist ligands, TUG-891, Linoleic acid, α-Linolenic acid, and Oleic acid, based on a homology model. Our simulations, which accumulated a total of 2 μs of simulation, highlighted two binding hotspots at Arg99$^{2.64}$ and Lys293 (ECL3). The results indicate that the residues are located in separate areas of the binding pocket and interact with various types of ligands, implying different potential active states of FFA4 and a highly adaptable binding intra-receptor pocket. This article proposes additional structural characteristics and mechanisms for agonist binding that complement the experimental structures.

KEYWORDS
fatty acid sensing, free fatty acid receptor, homology modeling, molecular dynamic, TUG-891

1 | INTRODUCTION

G-protein coupled receptors (GPCR) are a diverse class of membrane proteins that are involved in various cellular processes; mainly signal transduction via secondary messengers, and sensing of extracellular stimuli. GPCRs have a core transmembrane domain (TM domain) composed of 7 helices that often harbors a ligand binding pocket. Upon activation, the helices undergo conformational changes inducing signaling pathways downstream to the coupled intracellular heterotrimeric G-protein which dissociates into its subunit α and βγ complex [1,2]. GPCRs have various ligand types and are classified into subfamilies sharing common structural and physiological features.

Deorphanization has enabled the identification of four endogenous free fatty acids receptors (FFAs) (Figure 1). They are classified in the rhodopsin-like
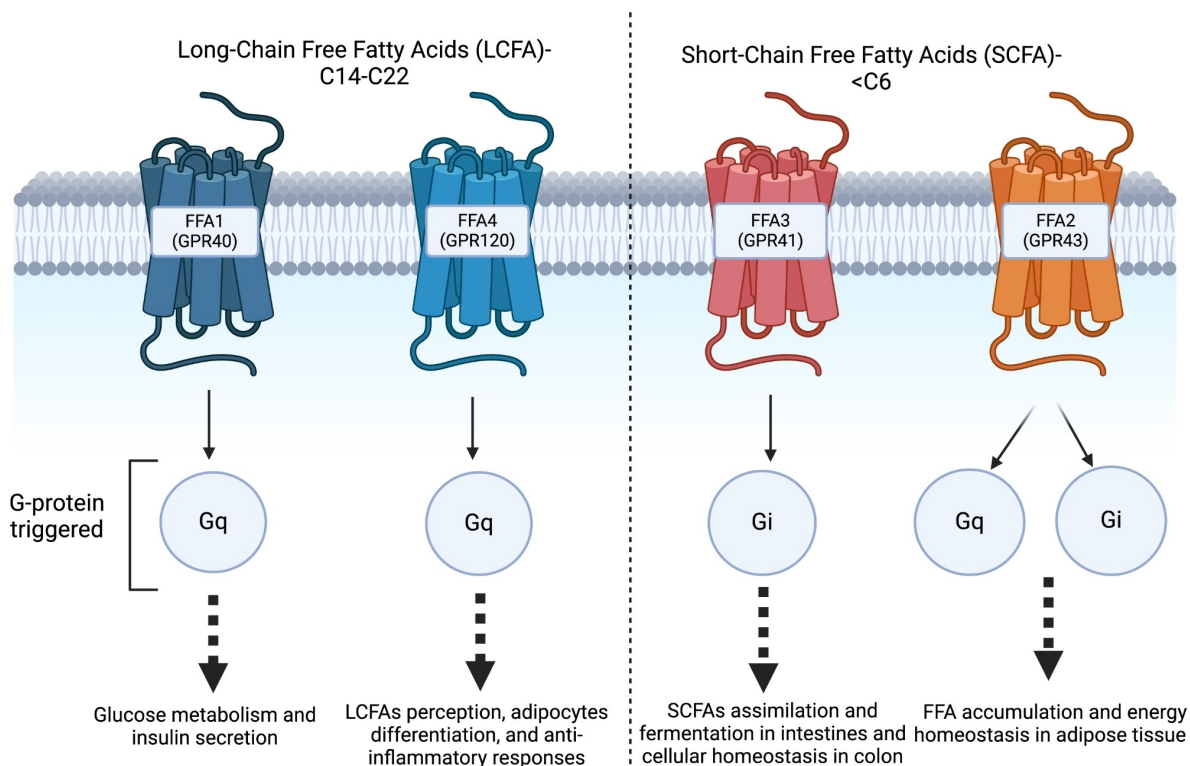
FIGURE 1 Signaling pathways of FFAs: ligand affinity and specific G-protein activation of each FFA (created with BioRender.com).

family (class A) and are involved in lipid metabolism and energy homeostasis. FFA2, previously known as GPR43, and FFA3, previously GPR41, are specific for short chain fatty acids (SCFAs) resulting from gut microbiota fermentation [3,4]. On the other hand, FFA1 (GPR40) and FFA4 (GPR120) specifically bind long chain fatty acids (LCFAs) [5,6]. FFA1-FFA3 share a relatively high sequence identity of 30–40% whereas, although overlapping ligand affinity with FFA1, FFA4 shows low similarity with the FFA family. It also presents two isoforms differing by 16 residues in the intracellular loop 3 (ICL3).

The deorphanization of FFA4 in 2005 [7] has uncovered its metabolic implications. In lipid metabolism such as adipogenesis and lipid storage, it regulates glucose uptake and glucagon-like peptide 1 (GLP-1) secretion critical in diabetes mellitus [8] as well as anti-inflammatory effects induced through inhibiting the production of pro-inflammatory cytokines and chemokines in immune cells [9,10]. Although labeled as a lipid receptor, FFA4 also revolves around fat sensing and taste perception through its expression in human taste bud cells (TBC) [11]. FFA4 is involved in oro-sensory detection of fatty acids in the diet. Its activation by the synthetic agonist ligand TUG-891 modulates dietary fat

preference in mice and curtails weight intake through the gut-tongue-brain axis activation that triggers satiety [12]. The ability to modulate diet preference by activating FFA4 constitutes a novel approach to tackle obesity and related metabolic disorders [13].

Understanding the structure and function of FFA4 is of great importance to unravel its molecular mechanisms and to explore its therapeutic potential. The lack of depth in experimental determination of three-dimensional structures has limited FFA4 pharmacological research and restrained potent agonist design up to very recently [14]. This limitation has led to the use of computational methods, mainly homology modeling, to generate models of FFA4 based on the available experimental structures of related class A GPCRs [15–17]. Here, we present a computational approach using comparative modeling and molecular dynamics to gain insights into the structure and dynamics of FFA4. To investigate the dynamic behavior and ligand interactions of FFA4, we performed MD simulations with FFA4-ligand complexes for a total simulation time of 2.5 μs. Structural insights from our computational study described in this paper support complementary information to the experimental data that were made available through published cryo-EM structures [18,19]

and furthermore contribute to the understanding of the ligand binding mechanisms. We concentrate here on agonists of synthetic (TUG-891) or natural origin (linoleic acid, α-linolenic acid, and oleic acid).

2 | MATERIAL AND METHODS

2.1 | Homology modeling

2.1.1 | Sequence retrieval and template selection

The 361-amino acid sequence of human FFA4 (Mass = 40 494 Da) was retrieved from UniProt [20] under reference Q5NUL3 for the canonical (short) isoform. A BLAST search was performed on the PDB to select templates identified on sequence similarity and structural conservation (class A GPCR). The choice of templates was guided by factors such as overall sequence identity, coverage, alignment quality and structural quality of template (resolution, co-crystallized agents/ligands, and active or inactive state of template).

2.1.2 | Alignment and homology modeling

The target sequence (FFA4) and selected template structures (10 X-ray crystal structures and 2 cryo-EM templates) were aligned using the multiple sequence alignment tool Clustal W [21,22]. A manual adjustment has been made to improve the alignment quality, particularly in regions with low sequence conservation such as the region of TM5 to TM6. Homology modeling was performed using MODELLER [23] with the aligned target-template sequence serving for generating a three-dimensional model of FFA4. Two models of FFA4 were generated to assess the impact of its activation states on its structure (inactive state and active state) [24–26]. The modeling process involved the construction of a preliminary model based on the templates followed by energy minimization and refinement including loops and addition of N and C termini. Evaluation of the models was performed depending on stereochemical quality and validated using PROCHECK and Ramachandran diagrams [27] (Figure S2). Preliminary MD studies were performed on the models to refine the unfolded N and C termini. Depending on the activation of the model, a $G_{q/11}$ protein was added to stabilize the activated helix bundle following its positioning from an active state template. Generic Ballesteros and Weinstein numbers were assigned to residues based on the most conserved residues in GPCR families for each transmembrane helix [28].

2.2 | Docking and molecular dynamics

2.2.1 | Docking

Endogenous and synthetic ligands of FFA4 were obtained through the IUPHAR/BPS database [29]. Ligands TUG-891, α-Linolenic acid (LNA), Linoleic acid (LA) and Oleic acid (OA) (Table 1) were docked using Auto-Dock Vina [30] in the binding pocket described previously [16,31] with $Arg99^{2.64}$ as the main ionic interaction and aromatic cage further down the pocket formed by $Trp277^{6.48}$, $Phe115^{3.29}$, Trp104 and $Phe304^{7.36}$. To facilitate the docking process, a grid box was generated around $Arg99^{2.64}$ and centered around the whole 7TMH domain. The dimensions of the grid box were set to encompass the putative ligand-binding pocket based on the available experimental data or predictions from structural analysis. 10 conformations of each ligand were generated and their fit with the binding mode hypothesis was assessed visually. The most plausible binding poses based on a combination of highest docking scores and best orientation of ligand-protein interactions observed for each ligand was selected for MD studies.
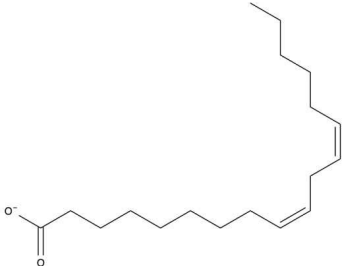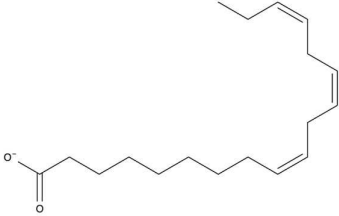
2.2.2 | Molecular dynamics

MD studies were performed using YASARA [36] on FFA4/$G_{q/11}$ in complex with the docked ligands for a total of 1 µs each in bilipid layer and a water box, using AMBER 14 IPQ forcefield [37] and a time step of 2×2.5 fs using the standard protocols from YASARA. Snapshots of the system were recorded at regular intervals of 250 ps, the system was analyzed using YASARA and ligand-protein interactions were interpreted using RStudio and SINAPs [38]. RMSD of the backbone-aligned or ligand-aligned protein and RMSF of the protein were obtained to visualize molecular shifting and protein/binding pocket conformations clusters. The analysis of the trajectories and the clustering of samples (ε = 1.5 Å) were done with CPPTRAJ from AmberTools [39].

3 | RESULTS AND DISCUSSION

Homology modeling of FFA4 based on templates of GPCR from the class A family allowed us to create 2 models based on its supposed activation state (active and inactive state). Based on a blast search on the full

T A B L E 1   Agonist ligands docked in homology generated model of FFA4 [32–35].

TUG-891
Agonist / pEC$_{50}$ = 7
pka = 4.6

Oleic Acid (OA)
Agonist / pEC$_{50}$ = 4.7
pka = 4.8

Linoleic Acid (LA)
Agonist / pEC$_{50}$ = 5.9
pka = 4.8

α-Linolenic Acid (LNA)
Agonist / pEC$_{50}$ = 5.5
pka = 4.8

sequence of isoform 2 of hFFA4, 10 x-ray crystallographic structures were pre-selected as templates (Figure S1). The alignment of FFA4 sequence with the templates gave a maximum identity sequence of 24.6 % (6TOD template) [24] leading to homology modeling with an antagonist bound Orexin 1 receptor template (inactive state). The low identity of sequences in the TM5 and TM6 resulted in poorly defined helices. TM5 and TM6 have long but heterogenous lengths throughout the GPCR rhodopsin-like family and TM6 is considered a "macroswitch" for activation of GPCR [40] and the key intracellular loop 3 (ICL3) between them contributes directly to the screening and signal transduction of the coupled G-protein [41]. A manual adjustment of TM5 and TM6 alignment has been carried out from a multiple sequence alignment within class A family (Figure S1) and MD studies were performed to refine the helices. Conservation of generic residues in all the TM helices enabled us to manually adjust the modeling of TM.

The homology modeling of the active model was performed using Cholecystokinin Receptor (CCKB; PDB identifier 7F8 W). CCKB cryo-EM structure is in active state co-crystallized with an agonist and protein G$_{q/11}$ corresponding to the desired activation state of FFA4 for our study. The identity sequence is 30.1 % and the template has a resolution of 3.10 Å. Alignment of the 7TM domains was performed and the G-protein was grafted (Figure 2a). MD simulations were performed with and

without G-protein to assess the implication of G protein during ligand binding and downward signaling pathway through micro-switches and molecular interactions. Manual adjustment of ICL3 by modifying torsions was performed to allow the graft of the G-protein. MD of the active model enabled 1.5 supplementary helical turns to form at the intracellular end of TM6 (Figure 2b).

3.1 | Key features of inactive and active state models

General comparison of both models does not highlight disruption in geometry or transmembrane helices orientation or length. The inactive model exhibits longer TM with the highest Δ in TM1 of the inactive model being 11 amino acids longer resulting in a shorter N-terminus. The folding of the N-terminus in both models obtained from MD study shows a more flexible terminus for the active model due to a longer unfolded chain. TM2 is presented as a key region for the ligand-induced activation of FFA4 based on studies of ligand binding to Arg99$^{2.64}$. This residue is located at the extracellular end of TM2 and oriented inwards the TM region for both models. No notable differences have been highlighted. TM3 and TM4 are of the same length for both models and a disulfide bond between Cys111$^{3.25}$ on TM3 and Cys194 (ECL2) between TM4 and TM5 stabilizes the beta hairpin structure of the ECL2 which forms a lid on

(a)

(b)

FIGURE 2 (a) Sequence alignment and secondary structure of active state model of FFA4 based on cholecystokinin receptor B (CCKB) and cryo-EM structure PDB: 8IYS of FFA4 (highlighted regions on the sequences indicate TM helix; conserved AA generic residues of Ballesteros and Weinstein in bold) [28]; (b) Structure of CCKB template (left); MD-refined homology model of FFA4 in its active form (middle) and cryo-EM 8IYS (right) illustrated with UCSF Chimera [42].

the extracellular end of the 7TM domain. Studies of the ECL2 highlighted the stabilizing effect of the lid on the pre-bonded ligand in some GPCRs.

ICL3 between TM5 and TM6 is shorter by 2 residues in the inactive state model compared to the active state but presents a higher flexibility in MD studies. The steric

hindrance and ionic interaction between Arg253 (ICL3) and Asp180 (Gα subunit) contribute to immobilizing ICL3 in the active state model. It is described as the main interacting and signal transduction pathway for activation of secondary messengers such as G protein. The relevance of the loop is shown by the isoform 1 (FFA4 L) selective secondary messenger pathway activation in concordance with its 16 supplementary residues in ICL3 (FFA4S co activates G protein and ß-arrestin2 whereas FFA4 L recruits only ß-arrestin 2) [34]. Direct contact between intracellular loops have also been described as determinant for G protein activation for some GPCRs, however the orientation of TM3/TM4 and TM5/TM6 and the depth of the helices in the intracellular region for both the inactive and active state of FFA4 seem to make this hypothesis less plausible.

TM5 and TM6 constitute the longest helices and are often heterogeneous throughout the GPCR class A family. The length of these helices are often crucial for GPCR activation as it implies micro-switches across those transmembrane domains as well as inward and outward movement of TM6 which is a common feature of conformational change during agonist binding [43]. TM5 presents no structural discrepancy to the contrary of TM6 displaying a shorter domain by 3 residues mainly at the extracellular region. The outcome of this structural variation is a longer and more flexible ECL3

between TM6 and TM7. The relevance of ECL3 has been highlighted throughout the studies carried out on the ligand binding and MD of FFA4 presented in this paper by an ionic interaction with Lys293 present on ECL3 (Figure 3a). The degree of flexibility of ECL3 is essential for the orientation of Lys293 toward the TM domain as observed for the active state model, in contrast with its extracellular facing in the inactive model due to geometrical constraints induced by the collapsing of residues Asn291 and Phe292 in the helical domain of TM6 (Figure 3b).

A conformational shift is also observed for the E/DRY motif present in the GPCR class 2 family that indicates the active or inactive state of the receptor. An intra helical salt bridge between Arg135$^{3.50}$ and Glu134$^{3.49}$ characterizes the inactive conformation of the receptor whereas in the active state, Arg135$^{3.50}$ favors an ionic interaction with Tyr227$^{5.58}$ [44]. The absence of Tyr residues on the TM6 unhinges the E/DRY mechanism involving interaction between TM3 and TM6, resulting in the inward/backward movement of TM6 upon activation. This can explain the limited movements between both models for the intracellular tip of TM6 and the pivot around Trp277$^{6.48}$ resulting in a lateral movement of the extracellular region of TM6 for the activated model. The involvement of the NPxxY motif on the TM7 for the active state model can be assessed by the shift of
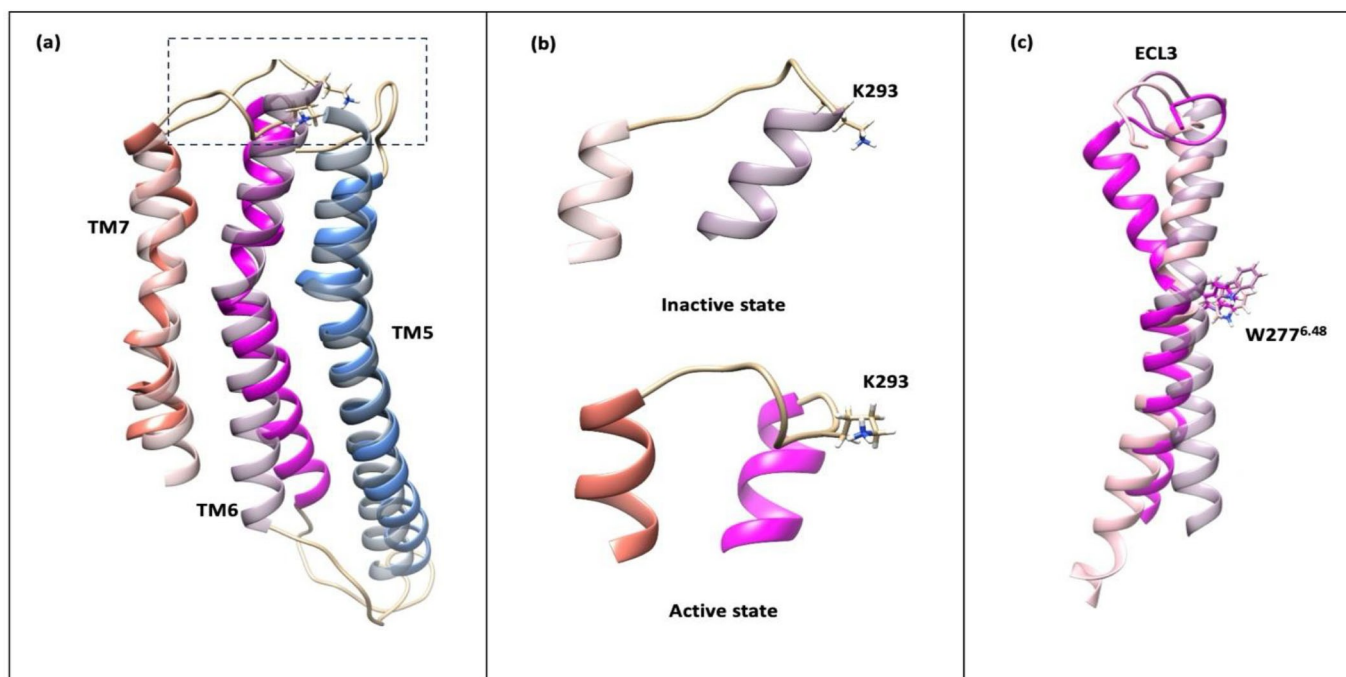
FIGURE 3 (a) Superimposition of TM hotspots domains (TM5, TM6, TM7) for active (dodger blue, magenta, and salmon) and inactive HM (grey blue, grey pink and grey salmon); (b) dissimilar flexibility of ELC3 ensues different orientation of Lys293; (c) Kink motion of TM6 centered onto Trp277$^{6.48}$ and induced by ligand binding with ECL3 facing towards the binding pocket. (active HM = magenta, inactive HM = pink, cryo-EM 8IYS = grey pink).

Tyr321[7.53] towards Arg135[3.50] therefore tightening the distance between TM3 and TM7 (Figure 4a). The affinity of Arg135[3.50] to either Tyr227[5.58] or Tyr321[7.53] may correspond to an intermediate state of activation in the switch from TM3-TM5 to TM3-TM7 with the anchoring of the G-protein to TM6 resulting in a backward shift of the TM5-TM6 as well as of ICL3 to accommodate the Gα subunit.
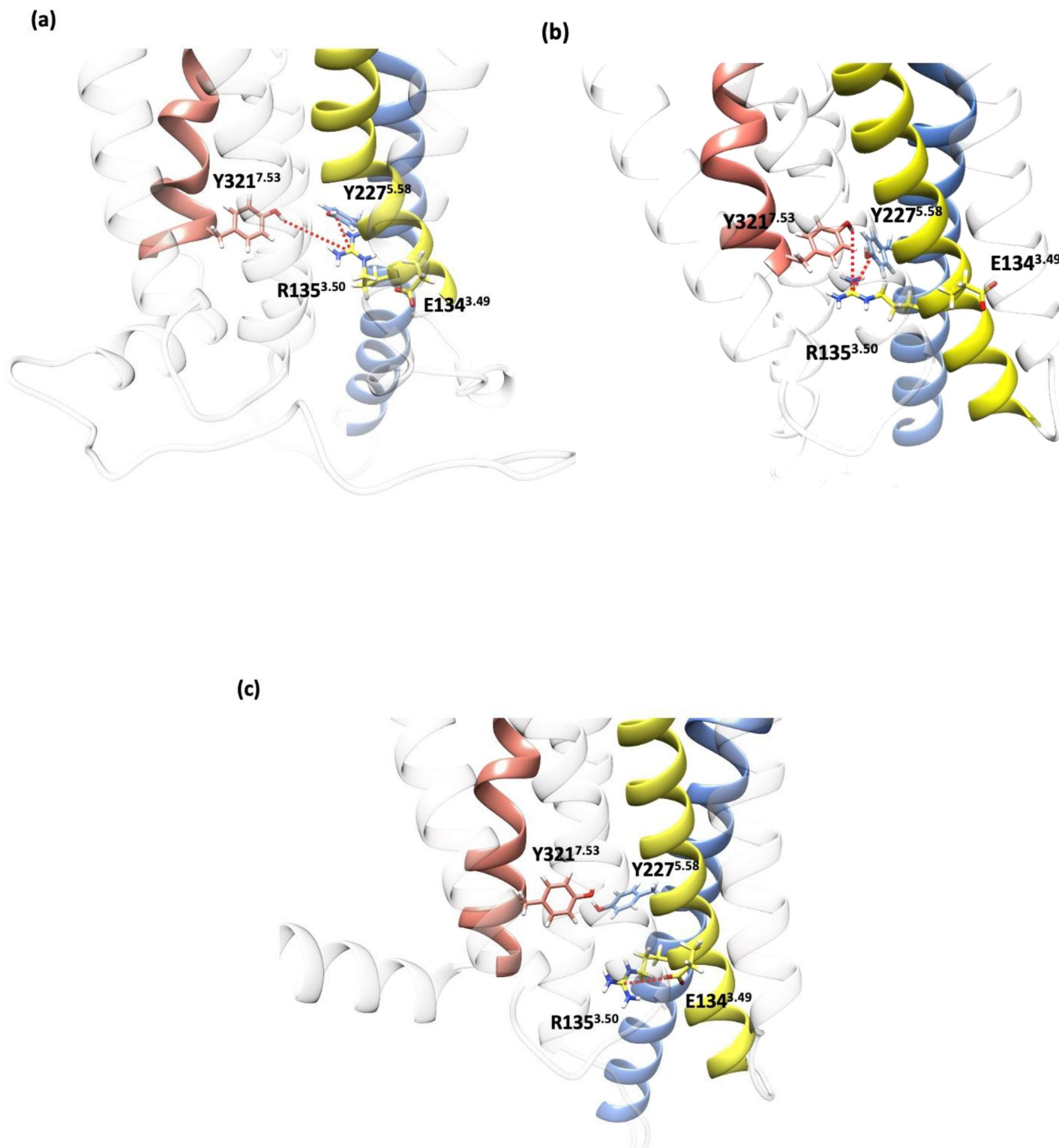
FIGURE 4 (a) Active state FFA4 homology model E/DRY and NPxxY interactions Arg135[3.50]-Tyr227[5.58]-Tyr321[7.53]; (b) Cryo-EM structure of FFA4-TUG891 showcasing active conformation TM3-TM5-TM7 interaction; (c) Inactive FFA4 homology model exhibits the ionic lock Arg135[3.50]-Glu134[3.49].

3.2 | Comparison with cryo-EM

The backbone mean RMSD compared to all cryo-EM structures [18,19] of the active and inactive model is 1.038 Å and 1.088 Å respectively, which is lower than the resolutions of the templates used for the homology modeling of both models (6TOD = 2.11 Å and 7F8 W = 3.10 Å). The low RMSD indicates that our model in both states is structurally close to the experimental data. However, RMSD comparison with our inactive model can be biased due to different activation states with the cryo-EM structures. The ligands tested in the experimental study revolved around long chain fatty acids recognition except for TUG-891. The comparison of the experimental data is mostly relevant for our active state model. The analysis of the protein backbone of the cryo-EM structures of FFA4 cryo-EM confirms the initial findings made by comparative modeling concerning the length of the transmembrane helices and the manual adjustments of TM5 and TM6 fits with the experimental data apart from the extracellular tip of TM6 (Table 2). The outward movement of TM6 in the case of the active state is confirmed with the intracellular tip being similar for the cryo-EM in contrast with the inactive homology model being inwards towards the TM (Figure 3c).

The lack of experimental data for the N and C-termini as well as the intra/extracellular loops does not enable validation of the following hypothesis derived from our models:

· The N-terminus of FFA4 folds back on the extracellular TM domain and forms an ionic lock parallel to the ECL2 hairpin stabilized by TM3.
· ICL2 is not implied in the interaction with and activation of G protein.
· ECL2 shares similar structure in both active/inactive states.
· ICL3 is the main signaling pathway for G-protein anchoring and activation. Outward movement of ICL3

from inactive to active state is nonetheless well described in experimental data.
· ECL3 is more flexible in the active state with an inward orientation of the loop to offer an anchoring point in the ligand binding site.
· Helix-8 structured from residues Cys326 to Cys335.

All cryo-EM structures are assumed to be in active state with G-protein grafts associated with the receptor backbone. Comparison of the E/DRY and NPxxY motif between the experimental data (Figure 4b) and our homology models conclude to a TM3/TM5 to TM3/TM7 shift although direct interactions cannot be assumed between Arg135$^{3.50}$ and Tyr321$^{7.53}$. The inactive state homology model exhibits the ionic lock Arg135$^{3.50}$-Glu134$^{3.49}$ in contrast with all the experimental cryo-EM structures, thus validating the conformational difference (Figure 4c).

3.3 | Molecular dynamics and induced-fit docking

The preliminary docking of TUG-891 in the binding pocket was carried out according to the protocol described in the literature [16], taking care of keeping the main ionic interaction between Arg99$^{2.64}$ and the carboxylic acid group of the ligand. Visual verification of the binding site assured the orientation of Arg99$^{2.64}$ toward the binding pocket. It gave the desired ionic interaction (Figure 5), but the stability of this binding mode was tested through MD studies. The results pinpointed the steric hindrance around Arg99$^{2.64}$ as the expected docking pose with the carboxylic acid function facing the arginine residue and an 'L' conformation of TUG-891 further down the TM domain could not be achieved. The inward movement of TM3 and the ECL2 hairpin structure stabilized through a disulfide bond on the extracellular tip of TM3 creates a clutter limiting the flexibility of TM2 for ligand binding. Docking of LNA and

TABLE 2 Comparison of transmembrane helices length for cryo-EM and homology model.

|	Generic Color (Figure 2b)	Generic residue	Homology Model-7F8 W			Cryo-EM-8IYS			Difference
			Start	End	Length	Start	End	Length	
TM1	Cyan	N58$^{1.50a}$	E43$^{1.35}$	V65$^{1.57}$	22	R36$^{1.28}$	R67$^{1.59}$	31	9
TM2	Red	D85$^{2.50a}$	A73$^{2.38}$	R99$^{2.64}$	26	G72$^{2.37}$	E102$^{2.67}$	30	4
TM3	Yellow	R136$^{3.50a}$	P108$^{3.22}$	R145$^{3.59}$	37	G107$^{3.21}$	G146$^{3.60}$	39	2
TM4	Green	W163$^{4.50a}$	R152$^{4.39}$	F176$^{4.63}$	24	G151$^{4.38}$	F176$^{4.63}$	25	1
TM5	Cornflower Blue	P219$^{5.50a}$	W207$^{5.38}$	S244$^{5.75}$	37	T200$^{5.31}$	V243$^{5.74}$	43	6
TM6	Purple	P279$^{6.50a}$	Q258$^{6.29}$	Q290$^{6.61}$	32	H251$^{6.23}$	N291$^{6.62}$	40	8
TM7	Salmon	P318$^{7.50a}$	P300$^{7.32}$	M323$^{7.55}$	23	W299$^{7.31}$	N322$^{7.54}$	23	0
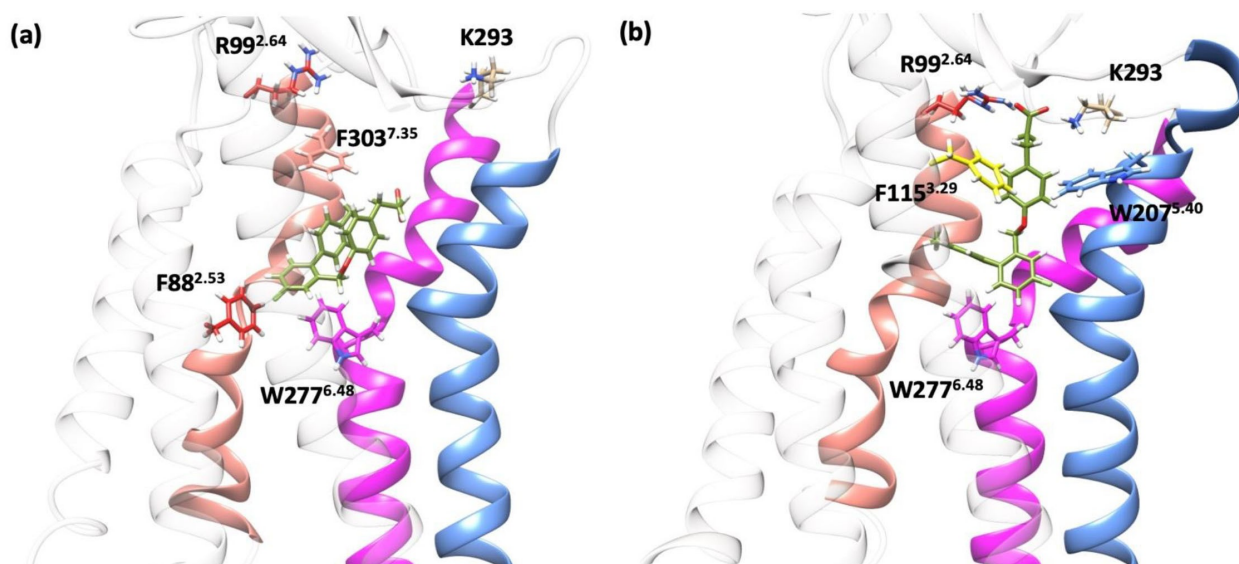
FIGURE 5 (a) Preliminary docking of TUG-891(sea green) centered in the aromatic cage; (b) TUG-891 conformation obtained and confirmed after 3 replicates of 100 ns MD simulation. TM5 = Cornflower blue, TM6 = Magenta and TM7 = Salmon.

OA in the binding pocket gave a much more plausible ligand pose with the carboxylic acid facing toward Arg99[2.64] and the flexible hydrophobic tail of both fatty acids deeper down the TM domain. The tail of both endogenous ligands gave significant insights with possible hydrophobic interacting residues mostly located on TM5, TM6 and TM7. The surface analysis of the binding pocket indicates a deep binding site with a trigger switch residue W277[6.48] at the bottom. Entries have been identified between TM4-TM5 through the bilipid membrane or on top of the extracellular region around ECL3 but hindered due to the hairpin loop ECL2. Complementary to the main ionic interaction Arg99[2.64]-Carboxylic acid of TUG-891, residues in a 5 Å radius around ligands were highlighted for possible interactions. Main residues highlighted consisting mainly of Phe and Ile residues resulted in π-stacking or hydrophobic interactions with the aromatic fragments of TUG-891. Docking of endogenous ligands of FFA4 with α-Linolenic Acid (LNA) and Oleic Acid (OA) also resulted in the carboxylic acid facing towards Arg99[2.64] and the flexible hydrophobic tail down the TM domain.

The MD studies of TUG-891, LNA and OA have mitigated results with the rapid loss of the main ionic interaction between the carboxylic acid function and Arg99[2.64] except with LNA. The acidic moiety is unable to achieve a stable interaction with the arginine and the ligand remains in the pocket solely due to hydrophobic interactions. All replicas resulted in the loss of the Acid-Arg99[2.64] interaction which led us to a different approach. The RMSD analysis of the ligands highlighted the low movement of TUG891 and OA (Figure S3). We

screened the binding pocket for negatively charged amino acids that would result in a more stable interaction, but no suitable site was identified. Hence, TUG-891 was placed deeper inside the binding pocket with no ionic interactions. The V shape conformation of TUG-891 still had the carboxylic acid function facing upward the binding pocket (Figure 5a). We identified a new binding hotspot localized throughout the TM5-TM6-TM7 region with a stable salt bridge with Lys293 located on ECL3 (Figure 5b). The 'L' shape conformation described in the literature [16] and confirmed by experimental data (cryo-EM structures 8G59 and 8ID8) concords with the MD simulations data of TUG-891. Multiple conformations of TUG-891 have been tested to ascertain the binding site and comparative studies between Arg99[2.64] and Lys293 have shown a higher stability with the latter with a lower RMSD of the ligand. Ligand-protein contact analysis gave us insights about possible binding sites of FFA4 and the interaction stability. LNA exhibited a more stable interaction with Arg99[2.64] than TUG-891 or even other endogenous ligands OA and LA. However, docking conformation 2 of LNA oriented toward Lys293 presented a higher stability, shown by a lower ligand RMSD. The cryo-EM structures validate the binding site around Lys293 although a direct interaction including the latter cannot be visualized due to the lack of this residue in the PDB files [18,19]. The residues highlighted by the cryo-EM data exhibits primarily hydrophobic properties for the recognition of unsaturated fatty acids with a surprising unfavorable carboxylic acid-Asp/Glu side chain proximity. This interaction is suggested to occur through the implication of a water molecule bridge and

Asn291 residue acting as the positive end of this quadrapoles interaction. The B-factor for the experimental data indicates that the ECL2 and ECL3 are highly flexible and may be important for the receptor's function and ligand binding (Figure S6) [40]. In line with ECL3, the extracellular end of TM6 and TM7 could allow conformational change in response to ligand binding. The high B-factor of the extracellular region of FFA4 suggests a poorly structured region and this supports the hypothesis of the entrance of the ligand taking place by the top of the receptor as opposed to the membrane-diffusion as observed in FFA1 (PDB: 5TZR) [41]. The cryo-EM structures and homology models overall support the fact that FFA4 has a deep but narrow binding pocket that is accessible mainly by the top (Figure S7). Interestingly, the extracellular side of FFA4 has a large number of arginines and lysines, in keeping with the acidic nature of the natural ligands, which may be a further argument in favor of a top opening of the binding site.

The lack of accessibility of Arg99$^{2.64}$ for ligand interaction in both the experimental and *in silico* data makes the comprehension of the binding and activation mechanism of FFA4 challenging. Published mutagenesis studies formerly highlighted Arg99$^{2.64}$ as critical for FFA4 activity [9]. Based on our finding of Lys293 as a potential binding site, we performed thorough MD simulations on a set of 10 reference ligands (IUPHAR/BPS database) and 3 additional endogenous ligands described by the experimental data [42,43] (Figure S4). Comprehension of a binding site based on ligand properties has emerged from the molecular dynamic studies with this selection of natural or endogenous and saturated or unsaturated backbone ligands.

The endogenous ligands such as oleic acid (Figure 6) interact with Arg99$^{2.64}$ apart from OA presenting dual binding to Arg99$^{2.64}$ and Lys293. The residue interaction maps of the natural ligands show specific residues of TM3: Met118$^{3.32}$ and Thr119$^{3.33}$ exhibits π-stacking with unsaturated or halogen moiety of the ligands whereas the Phe211$^{5.42}$ on TM5 interacts with the CH3-termini of the carbon chain. With a chain length above C18, uncorrelated to unsaturation, ECL2 residues Thr195 and Leu196 promotes ligand orientation in the binding pocket through hydrogen bonding with the carboxylic acid head or in some cases prevents the escape of the ligand for shorter but more flexible fatty acids (grifolic acid or myristic acid) (Table S1). The implication of ECL2 is not compelling although confirmed in our homology model for the binding of natural endogenous ligands with longer carbon chains (Figure S5). The high flexibility of this loop could imply its conformational shift because of the ligand binding down the pocket. Furthermore,
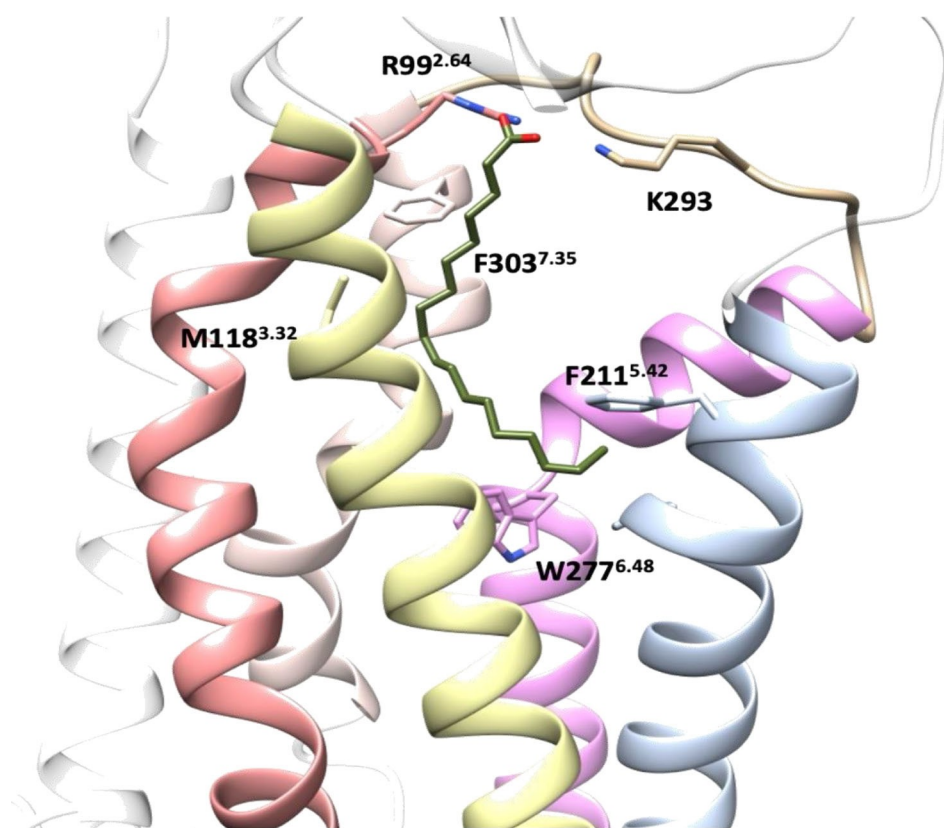

FIGURE 6 Conformation of OA (olive) after 100 ns MD simulation interacting by a SB with Arg99$^{2.64}$.

TABLE 3 Comparative study of TUG-891/FFA4 interaction between experimental data (cryo-EM and mutagenesis data) and homology study (HB=Hydrogen Bond; SB=Salt Bridge; VdW=Van der Waals'; π-π=Hydrophobic π-stacking; Halogen=Halogen bond; repulsive contact=Negative-negative interaction; Critical=Critical to receptor's activity).

AA side chains		8IYS	8G59	Homology Model	Mutagenesis study [31]
Arg	$99^{2.64}$	–	–	–	Critical
Trp	104 (ECL1)	–	–	–	Critical
Phe	$115^{3.29}$	–	–	–	Critical
Met	$118^{3.32}$	VdW	VdW	Halogen	–
Ile	$126^{3.40}$	VdW	VdW	–	–
Leu	$173^{4.60}$	VdW	–	VdW	–
Trp	198 (ECL2)	HB	HB	–	–
Glu	$204^{5.35}$	repulsive contact	repulsive contact	–	–
Trp	$207^{5.38}$	VdW	VdW	π-π	Critical
Asp	$208^{5.39}$	repulsive contact	repulsive contact	–	–
Phe	$211^{5.42}$	π-π	π-π	π-π	Critical
Val	$212^{5.43}$	–	–	VdW	–
Trp	$277^{6.48}$	–	–	π-π	Critical
Pro	$279^{6.50}$	–	–	VdW	–
Ile	$280^{6.51}$	VdW	VdW	VdW	–
Ile	$284^{6.55}$	VdW	VdW	–	–
Ile	$287^{6.58}$	–	VdW	–	–
Asn	$291^{6.62}$	–	HB	–	–
Lys	293 (ECL3)	–	–	SB	–
Phe	$303^{7.35}$	VdW	VdW	π-π	–
Phe	$304^{7.36}$	–	–	–	Critical
Val	$306^{7.38}$	–	–	VdW	–
Val	$307^{7.39}$	VdW	VdW	VdW	–

information derived from the mutagenesis study of FFA4 activity [30] has not discovered residues on ECL2 as critical for FFA4 activity. Our model suggests an intermediate state of FFA4 resulting in the ligand-entry from the extracellular media down the binding pocket. Lys293 serves as an anchor to stabilize the ligand in the hydrophobic transmembrane domain. The inward movement of Lys293 and by extension ECL3 causes an outward movement of the upper end of TM6 widening the binding pocket volume (Figure S7). This results in a greater proximity of the ligand to residues on TM6 and TM7 discarding the unfavorable interaction with Glu204$^{5.35}$ and Asp208$^{5.39}$. Complementary to Lys293, Trp207$^{5.38}$ could stabilize the ligand conformation by hydrophobic interactions and is confirmed to play a pivotal role too in FFA4 activity in biological assays.

Our main interest was for the synthetic TUG-891. The cryo-EM data pinpoint an interacting pole at the top of TM5 between the carboxylic acid moiety and residues Glu204$^{5.35}$/Asp208$^{5.39}$ resulting in an unfavorable energetic interaction counterbalanced with hydrogen

bonding with Trp198 (ECL2). The analysis of TUG-891 interactions during the MD runs highlights the flexible nature of the binding pocket of FFA4 (Table 3). Based on the clustering of TUG-891 on the pooled, backbone-aligned conformations of the complex from the 6 MD runs, we identified two major conformations accounting for 89% of the conformations. The largest represents nearly three fourths of the complexes alone (Table 4).

The analysis of the interaction distances with the main residues was realized to compare the two most populated clusters (Figure 7). The two main binding hotspots for the acid of the ligands, Arg99$^{2.64}$ and Lys293, are consistently closer to the acid of TUG-891 in our MD simulations than in either experimental structure. Interestingly, the latter is in a range of distance consistent with a salt bridge, indicating a stable binding preference for this residue over the arginine (Figure 8). This residue is absent from the experimental data or very incomplete and pointing toward the outside, much as in our preliminary model. The differences in the orientation of ECL2 and the kink of TM6 compared to the experimental

TABLE 4 Clustering of TUG-891 conformations obtained from MD simulations.

Cluster	Frames	%	RMSD	Standard deviation
0	1777	73.9	1.13	0.4
1	368	15.3	1.1	0.36
2	95	3.9	0.9	0.3
3	49	2	0.9	0.28
4	40	1.7	1.15	0.34
5	38	1.6	0.8	0.19
6	26	1.1	1	0.26
7	6	0.2	0	0
8	6	0.2	1.15	0.31
9	1	0.01	0	0

structures point to a reorganization of the extracellular part of FFA4 to offer a more stable interaction to the acidic group of the ligand upon binding. The experimental data also diverge from our findings for the unfavorable Acid-Acid interaction with Glu204[5.34] and Asp208[5.39]. Proximity of the ligand in the TM5 region is confirmed through the results of our MD simulations but no direct interaction is exhibited contrary to the cryo-EM with Asp208[5.39]. Glu204[5.34] found on the extracellular end of TM5 does not interact with the ligand in our MDs and comforts even more the flexibility of the extracellular end of the binding pocket by being consistently much further from TUG-891. The differences between our homology model and the cryo-EM structures may be the result of a thermodynamic Vs kinetic stability of the complex. The aromatic residues Trp198, Trp207[5.38] and Trp277[6.48] displays median distances to the ligand in accordance with their initial position, with Trp198 and Trp207[5.38] at a greater distance than Trp277[6.48]. Due to the different position of TUG-891 in the binding site, the interaction with Trp198 is not seen during the MD runs. However, Trp207[5.38] keeps a similar role with a consistent π-π interaction. Trp277[6.48] is also able to form a π-π interaction, although the distance between the centers of the aromatics is not fully fixed. This residue is able to move away from the ligand rather freely, with a marked preference for an interacting position. It is at the base of the kink in TM6 and in close proximity with Phe216[5.47] to stabilize it further.

4 | CONCLUSION

Our *in-silico* study of FFA4 demonstrates a new possibility of interaction in the binding site. Lys293 preferentially interacts with synthetic acids and Arg99[2.64] with endogenous ligands. This preference may be rooted in the distinct physicochemical properties of these ligands (saturation, flexibility, and backbone chain length) with synthetic acids that are more lipophilic with aromatic rings and a shorter backbone, able to form stronger interactions with Lys293. The involvement of ECL2 and ECL3 in agonist binding could contribute to further elucidate the mechanisms underlying FFA4 ligand specificity and signaling. The extracellular loop 2 appears to play a role in orienting longer chain ligands within the binding pocket while restraining shorter ligands from escaping, and the extracellular loop 3 contributing to the capture of the ligands and the stabilization of the receptor-ligand complex. The homology model showcases a different active state of FFA4 from the experimental structures that offer complementary information on agonist binding. It could provide access to different structural properties for the selective preferences of ligands in the binding pocket of FFA4 and thus improve ligand design. The identification of Lys293/Trp207[5.38] and Arg99 as binding points and the implication of ECL2 and ECL3 in agonist binding provide valuable insights into the structure-function relationships of FFA4. These findings have significant implications for the design of FFA4-targeted drugs and resulted in the construction of a pharmacophore model for Virtual High Throughput Screening to design potent novel FFA4 agonists.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Guillaume Patient http://orcid.org/0000-0002-3940-9116
Naim. A Khan http://orcid.org/0000-0002-8930-9332
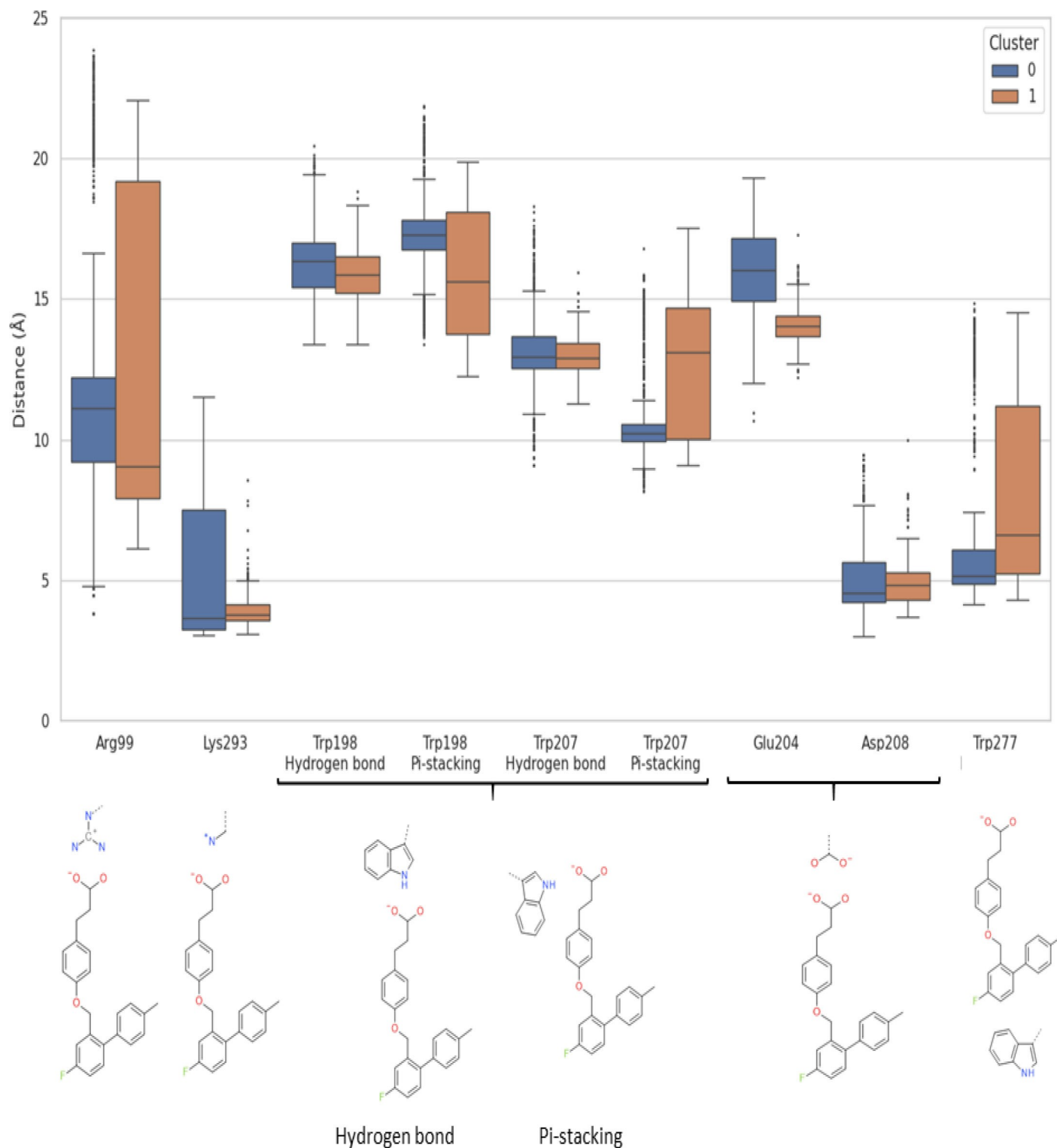*Amaury Farce* http://orcid.org/0000-0003-3992-9629

FIGURE 7 Distance between TUG-891 and the main residues of the binding site throughout the pooled MDs for cluster 0 and 1. For a full analysis of all MD generated conformations, refer to Figure S8.

REFERENCES
1. D. M. Rosenbaum, S. G. F. Rasmussen, B. K. Kobilka, *Nature* **2009**, *459*, 356–363.
2. B. K. Kobilka, *Biochim. Biophys. Acta (BBA) - Biomembr.* **2007**, *1768*, 794–807.
3. G. Milligan, L. A. Stoddart, N. J. Smith, *Br. J. Pharmacol.* **2009**, *158*, 146–153.
4. C. Tang, S. Offermanns, in *Free Fatty Acid Receptors* (Eds: G. Milligan, I. Kimura), Springer International Publishing, Cham **2016**, pp. 205–220.
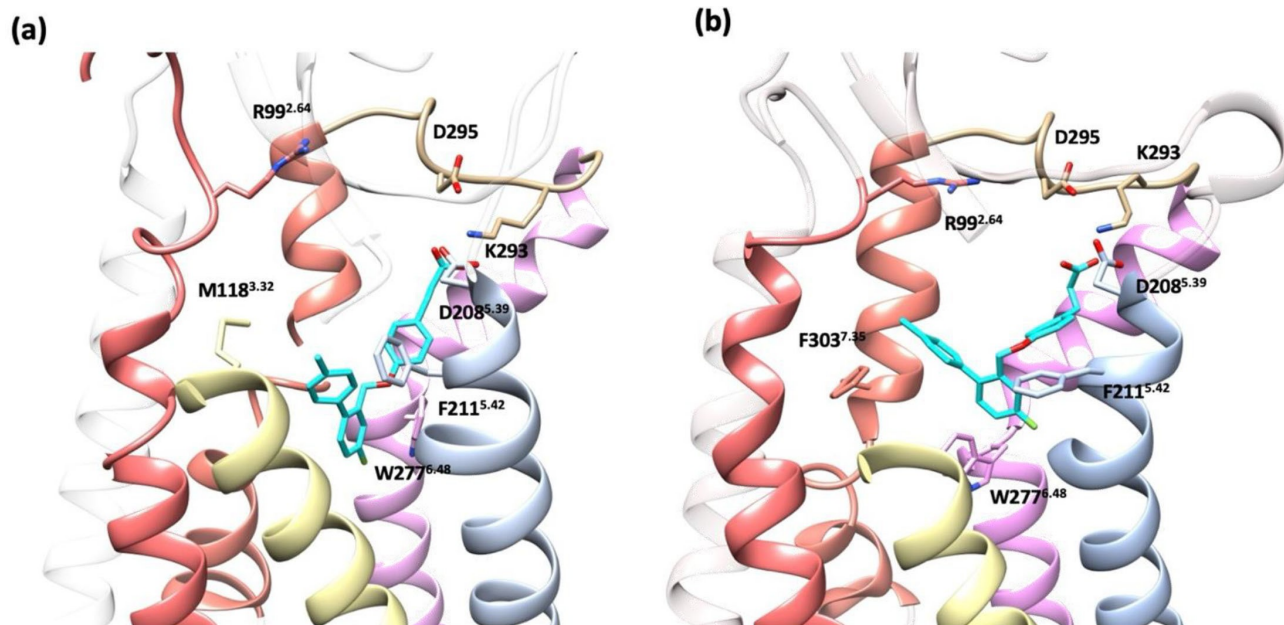
FIGURE 8 TUG-891-FFA4 conformations representative of (a) cluster 0; (b) cluster 1.

5. T. Hara, I. Kimura, D. Inoue, A. Ichimura, A. Hirasawa, in *Reviews of Physiology, Biochemistry and Pharmacology, Vol. 164* (Eds: B. Nilius, S. G. Amara, R. Lill, S. Offermanns, T. Gudermann, O. H. Petersen, R. Jahn), Springer International Publishing, Cham **2013**, pp. 77–116.

6. C. P. Briscoe, M. Tadayyon, J. L. Andrews, W. G. Benson, J. K. Chambers, M. M. Eilert, C. Ellis, N. A. Elshourbagy, A. S. Goetz, D. T. Minnick, P. R. Murdock, H. R. Sauls, U. Shabon, L. D. Spinage, J. C. Strum, P. G. Szekeres, K. B. Tan, J. M. Way, D. M. Ignar, S. Wilson, A. I. Muir, *J. Biol. Chem.* **2003**, *278*, 11303–11311.

7. A. Hirasawa, K. Tsumaya, T. Awaji, S. Katsuma, T. Adachi, M. Yamada, Y. Sugimoto, S. Miyazaki, G. Tsujimoto, *Nat. Med.* **2005**, *11*, 90–94.

8. G. Carullo, S. Mazzotta, M. Vega-Holm, F. Iglesias-Guerra, J. M. Vega-Pérez, F. Aiello, A. Brizzi, *J. Med. Chem.* **2021**, *64*, 4312–4332.

9. D. Y. Oh, E. Walenta, T. E. Akiyama, W. S. Lagakos, D. Lackey, A. R. Pessentheiner, R. Sasik, N. Hah, T. J. Chi, J. M. Cox, M. A. Powels, J. Di Salvo, C. Sinz, S. M. Watkins, A. M. Armando, H. Chung, R. M. Evans, O. Quehenberger, J. McNelis, J. G. Bogner-Strauss, J. M. Olefsky, *Nat. Med.* **2014**, *20*, 942–947.

10. D. Y. Oh, S. Talukdar, E. J. Bae, T. Imamura, H. Morinaga, W. Fan, P. Li, W. J. Lu, S. M. Watkins, J. M. Olefsky, *Cell* **2010**, *142*, 687–698.

11. C. Martin, P. Passilly-Degrace, D. Gaillard, J.-F. Merlin, M. Chevrot, P. Besnard, *PLoS ONE* **2011**, *6*, e24014.

12. B. Murtaza, A. Hichami, A. S. Khan, B. Shimpukade, T. Ulven, M. H. Ozdener, N. A. Khan, *J. Lipid Res.* **2020**, *61*, 133–142.

13. A. S. Khan, A. Hichami, B. Murtaza, M.-L. Louillat-Habermeyer, C. Ramseyer, M. Azadi, S. Yesylevskyy, F. Mangin, F. Lirussi, J. Leemput, J.-F. Merlin, A. Schmitt, M. Suliman, J. Bayardon, S. Semnanian, S. Jugé, N. A. Khan, *Cell. Mol. Gastroenterol. Hepatol.* **2023**, *15*, 633–663.

14. G. Milligan, B. Shimpukade, T. Ulven, B. D. Hudson, *Chem. Rev.* **2017**, *117*, 67–110.

15. S. R. Villa, R. K. Mishra, J. L. Zapater, M. Priyadarshini, A. Gilchrist, H. Mancebo, G. E. Schiltz, B. T. Layden, *J. Invest. Med.* **2017**, *65*, 1116–1124.

16. S. Pant, V. Ravichandiran, *Molecular Understanding of GPR120 Agonist Binding Using Homology Modeling and Molecular Dynamics*, Chemistry, **2021**.

17. I. G. Tikhonova, in *Free Fatty Acid Receptors* (Eds: G. Milligan, I. Kimura), Springer International Publishing, Cham **2016**, pp. 57–77.

18. C. Mao, P. Xiao, X.-N. Tao, J. Qin, Q.-T. He, C. Zhang, S.-C. Guo, Y.-Q. Du, L.-N. Chen, D.-D. Shen, Z.-S. Yang, H.-Q. Zhang, S.-M. Huang, Y.-H. He, J. Cheng, Y.-N. Zhong, P. Shang, J. Chen, D.-L. Zhang, Q.-L. Wang, M.-X. Liu, G.-Y. Li, Y. Guo, H. E. Xu, C. Wang, C. Zhang, S. Feng, X. Yu, Y. Zhang, J.-P. Sun, *Science* **2023**, *380*, eadd6220.

19. S. Zhang, B. L. Roth, *Cell Res.* **2023**, *33*, 657–658.

20. The UniProt Consortium: A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. Ebenezer, J. Fan, P. Garmiri, L. J. Da Costa Gonzales, A. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimo, G. Argoud-Puy, A. H. Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M.-C. Blatter, J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. De Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-

Nouspikel, F. Jungo, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, J. Zhang, *Nucl. Acids Res.* **2023**, *51*, D523–D531.

21. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.

22. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucl. Acids Res.* **1994**, *22*, 4673–4680.

23. B. Webb, A. Sali, *CP Bioinform.* **2016**, *54*

24. M. Rappas, A. A. E. Ali, K. A. Bennett, J. D. Brown, S. J. Bucknell, M. Congreve, R. M. Cooke, G. Cseke, C. de Graaf, A. S. Doré, J. C. Errey, A. Jazayeri, F. H. Marshall, J. S. Mason, R. Mould, J. C. Patel, B. G. Tehan, M. Weir, J. A. Christopher, *J. Med. Chem.* **2020**, *63*, 1528–1543.

25. F. Madeira, M. Pearce, A. R. N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, R. Lopez, *Nucl. Acids Res.* **2022**, *50*, W276–W279.

26. M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez, *Nucl. Acids Res.* **2010**, *38*, W695–W699.

27. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

28. J.A. Ballesteros, H. Weinstein, 1995, Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors, 366, in Methods in Neurosciences, S.C. Sealfon, Elsevier.

29. S. D. Harding, J. F. Armstrong, E. Faccenda, C. Southan, S. P. H. Alexander, A. P. Davenport, M. Spedding, J. A. Davies, *Nucl. Acids Res.* **2024**, *52*, D1438–D1449.

30. O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455–461.

31. B. D. Hudson, B. Shimpukade, G. Milligan, T. Ulven, *J. Biol. Chem.* **2014**, *289*, 20345–20358.

32. B. Shimpukade, B. D. Hudson, C. K. Hovgaard, G. Milligan, T. Ulven, *J. Med. Chem.* **2012**, *55*, 4511–4515.

33. C. P. Briscoe, A. J. Peat, S. C. McKeown, D. F. Corbett, A. S. Goetz, T. R. Littleton, D. C. McCoy, T. P. Kenakin, J. L. Andrews, C. Ammala, J. A. Fornwald, D. M. Ignar, S. Jenkinson, *Br. J. Pharmacol.* **2006**, *148*, 619–628.

34. S.-J. Watson, A. J. H. Brown, N. D. Holliday, *Mol. Pharmacol.* **2012**, *81*, 631–642.

35. X. Pan, H. Wang, C. Li, J. Z. H. Zhang, C. Ji, *J. Chem. Inf. Model.* **2021**, *61*, 3159–3165.

36. E. Krieger, G. Vriend, *J. Comput. Chem.* **2015**, *36*, 996–1007.

37. D. S. Cerutti, W. C. Swope, J. E. Rice, D. A. Case, *J. Chem. Theory Comput.* **2014**, *10*, 4515–4534.

38. C. Bedart, N. Renault, P. Chavatte, A. Porcherie, A. Lachgar, M. Capron, A. Farce, *J. Chem. Inf. Model.* **2022**, *62*, 1425–1436.

39. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan, K. M. Merz, *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191.

40. A. S. Hauser, A. J. Kooistra, C. Munk, F. M. Heydenreich, D. B. Veprintsev, M. Bouvier, M. M. Babu, D. E. Gloriam, *Nat. Struct. Mol. Biol.* **2021**, *28*, 879–888.

41. F. Sadler, N. Ma, M. Ritt, Y. Sharma, N. Vaidehi, S. Sivaramakrishnan, *Nature* **2023**, *615*, 734–741.

42. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

43. Q. Zhou, D. Yang, M. Wu, Y. Guo, W. Guo, L. Zhong, X. Cai, A. Dai, W. Jang, E. I. Shakhnovich, Z.-J. Liu, R. C. Stevens, N. A. Lambert, M. M. Babu, M.-W. Wang, S. Zhao, *eLife* **2019**, *8*, e50279.

44. G. E. Rovati, V. Capra, R. R. Neubig, *Mol. Pharmacol.* **2007**, *71*, 959–964.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only.
It will not be published as part of main.

Homology
Model of FFA4

Distinct binding
hotspots of FFA4