

Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics

Atanasio, S., Deganutti, G. & Reynolds, C. A.

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Atanasio, S, Deganutti, G & Reynolds, CA 2020, 'Addressing free fatty acid receptor 1 (FFAR1) activation using supervised molecular dynamics', *Journal of Computer-Aided Molecular Design*, vol. 34, no. 11, pp. 1181–1193.

<https://dx.doi.org/10.1007/s10822-020-00338-6>

DOI 10.1007/s10822-020-00338-6

ISSN 0920-654X

ESSN 1573-4951

Publisher: Springer

The final publication is available at Springer via <http://dx.doi.org/10.1007/s10822-020-00338-6>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Addressing Free Fatty Acid Receptor 1 (FFAR1) Activation using Supervised Molecular Dynamics

Silvia Atanasio¹, Giuseppe Deganutti^{1,2*} and Christopher A. Reynolds^{1,2}

¹School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

²New address: Centre for Sport, Exercise and Life Sciences, Coventry University, Alison Gingell Building, CV1 5FB, Coventry, UK

*Author for correspondence: ad5288@coventry.ac.uk

ORCID Christopher A Reynolds - 0000-0001-9267-5141

ORCID Giuseppe Deganutti - 0000-0001-8780-2986

ABSTRACT

The free fatty acid receptor 1 (FFAR1, formerly GPR40), is a potential G protein-coupled receptor (GPCR) target for the treatment of type 2 diabetes mellitus (T2DM), as it enhances glucose-dependent insulin secretion upon activation by endogenous long-chain free fatty acids. The presence of two allosterically communicating binding sites and the lack of the conserved GPCR structural motifs challenge the general knowledge of its activation mechanism. To date, four X-ray crystal structures are available for computer-aided drug design. In this study, we employed molecular dynamics (MD) and supervised molecular dynamics (SuMD) to deliver insights into the (un)binding mechanism of the agonist MK-8666, and the allosteric communications between the two experimentally determined FFAR1 binding sites. We found that FFAR1 extracellular loop 2 (ECL2) mediates the binding of the partial agonist MK-8666. Moreover, simulations showed that the agonists MK-8666 and AP8 are reciprocally stabilized and that AP8 influences MK-8666 unbinding from FFAR1.

Keywords: G protein-coupled receptors (GPCRs); FFAR1; GPR40; Molecular dynamics; MD; Supervised Molecular dynamics; SuMD.

1. INTRODUCTION

The free fatty acid receptor 1 (FFAR1, formerly GPR40) is a G protein-coupled receptor (GPCR) naturally activated by saturated C₁₂-C₁₆ or unsaturated C₁₈-C₂₀ free fatty acids [1]. These long chain free fatty acids (LCFA) act as full agonists to enhance glucose-stimulated insulin secretion from pancreatic β cells [2], and secretion of the incretins Glucagon-like peptide 1 (GLP-1) and Gastric inhibitory polypeptide (GIP) from intestinal enteroendocrine cells [3]. The development of drugs activating FFAR1 represent a new potential therapeutic approach against type 2 diabetes mellitus (T2DM - counting for ~ 90% of all diabetes cases [4]) as the activation of its alternative signaling [5, 6] may avoid the negative side effects characterizing the currently approved drugs, such as hypoglycemia and weight gain [7].

To date, the structure-based design of new compounds [8, 9] can exploit four crystallographic FFAR1 complexes (Table S1). In all these structures the receptor is captured in the inactive state, probably due to the stabilizing mutations introduced and the insertion of lysozyme T4 into intracellular loop 3 (ICL3) [10]. Moreover, the glycine-rich C-terminal segment of transmembrane helix 7 (TM7) and helix 8 (H8) are not resolved. Even though FFAR1 is organized in the GPCR structural hallmark seven transmembrane domain (TMD), the conserved structural motifs NP^{7.50}xxY (Figure S1), DR^{3.50}Y, and P^{5.50}I^{3.40}F^{6.44} as well as the “toggle switch” W^{6.48} [11, 12], are not present.

Before any FFAR1 structural data were available (Table S1), a pharmacological body of evidence led to the hypothesis of up to three different binding sites [13]. In this scenario, full agonists would bind a different site to the partial agonists, with positive heterotropic cooperativity between different binding sites. Indeed, the partial agonist fasiglifam (TAK-875, Figure 1, discontinued in phase III clinical trials due to liver toxicity) is able to amplify the agonistic activity of endogenous LCFA by binding to a largely internal FFAR1 allosteric site [14]. The first published FFAR1 structure [15] located the bound TAK-875 enclosed between the top halves of TM3 and TM4 (Figure 1), with the carboxylic moiety pointing into a hydrophilic cavity partially buried from the extracellular bulk by extracellular loop 2 (ECL2). TAK-875 participates in an extended charge network involving R183^{5.39}, R258^{7.35}, E172^{ECL2}, Y91^{3.37}, S187^{5.43}, N241^{6.52}, and N244^{6.55}. The disruption of the E145^{4.64} - R183^{5.39} and E172^{ECL2} - R258^{7.35} salt bridges upon TAK-875 binding have been proposed as contributing to receptor activation [13, 16], and corroborated by mutagenesis data on Y91^{3.37} and N244^{6.55} [14]. In 2017 Lu J. *et al.* [17] resolved FFAR1 in a ternary complex with the partial agonist MK-8666 and the agonist AP8 (Figure 1), revealing a second external binding site at the protein/lipid interface, which is responsible for the recognition of full agonists. While MK-8666 inserts in the same crevice and forms the same interaction pattern as TAK-875 (site A1 in Figure 1), AP8 accommodates in a cleft shaped by TM3, TM4, and TM5 (site A2 in Figure 1), where I130^{4.49}, L133^{4.52}, V134^{4.53}, and L190^{5.46} form a hydrophobic pocket for the terminal trifluoromethoxyphenyl moiety of the ligand and Y44^{2.42}, Y114^{ICL2}, and S123^{4.42} anchor the carboxylate through polar interactions (Figure 1). A direct comparison between the binary FFAR1:TAK-875 and the ternary FFAR1:MK-8666:AP8 complexes shows TM4 and TM5 sliding approximately one-half of a helical turn in opposite directions to create a deeper cleft for AP8, which in turn stabilizes ICL2 in a helix conformation. The subsequent crystal structure of the binary complex between FFAR1 and the full agonist Compound 1 (Figure 1) bound to site A2 sheds further light into the cross-talk between the different binding sites [10]; it shows a rearrangement of site A1 in which D175^{5.31} inserts between R183^{5.39} and R258^{7.35} and moves ECL2 towards TM6 and TM7. This conformational rearrangement in the absence of partial agonists occupying site A2 prevents the formation of an additional proposed solvent exposed binding site between TM1 and TM7 (site A3 in Figure 1b) [15, 18].

In this complex scenario, endogenous LCFA are thought to bind to the site A2. An orthosteric binding site exposed to the membrane (site A2), responsible for full activation of the receptor, could have evolved along with a different structural mechanism of transduction, finely tuned by an allosteric communication network as intriguingly suggested by the absence of conserved class A structural motifs. Indeed, a chemical signal (the binding of LCFA) originating from the membrane, rather than from the extracellular environment (the case for almost all the other class A GPCRs) could underlie the FFAR1 divergence from the consensus structure, and the possibility of a 2:1 stoichiometry between LCFA and FFAR1 (with sites A1 and A2 both occupied simultaneously).

Here we address the FFAR1 activation first describing a putative binding mechanism for the formation of the binary FFAR1:MK-8666 complex using supervised molecular dynamics (SuMD) simulations. To date, few ligands accessing a

GPCRs orthosteric site from the membrane have been addressed with atomistic unbiased simulations [19, 20]. Then, insights on possible interactions contributing to the allosteric communication between site A1 and A2 from equilibrium MD simulations performed on different FFAR1 complexes (**Table S2**) are presented. Finally, we applied a modified version of SuMD to simulate the MK-8666 unbiased unbinding from the binary FFAR1:MK-8666 and ternary FFAR1:MK-8666:AP8 complexes. Overall, this study produces mechanistic working hypothesis on the allosteric regulation and activation of FFAR1.

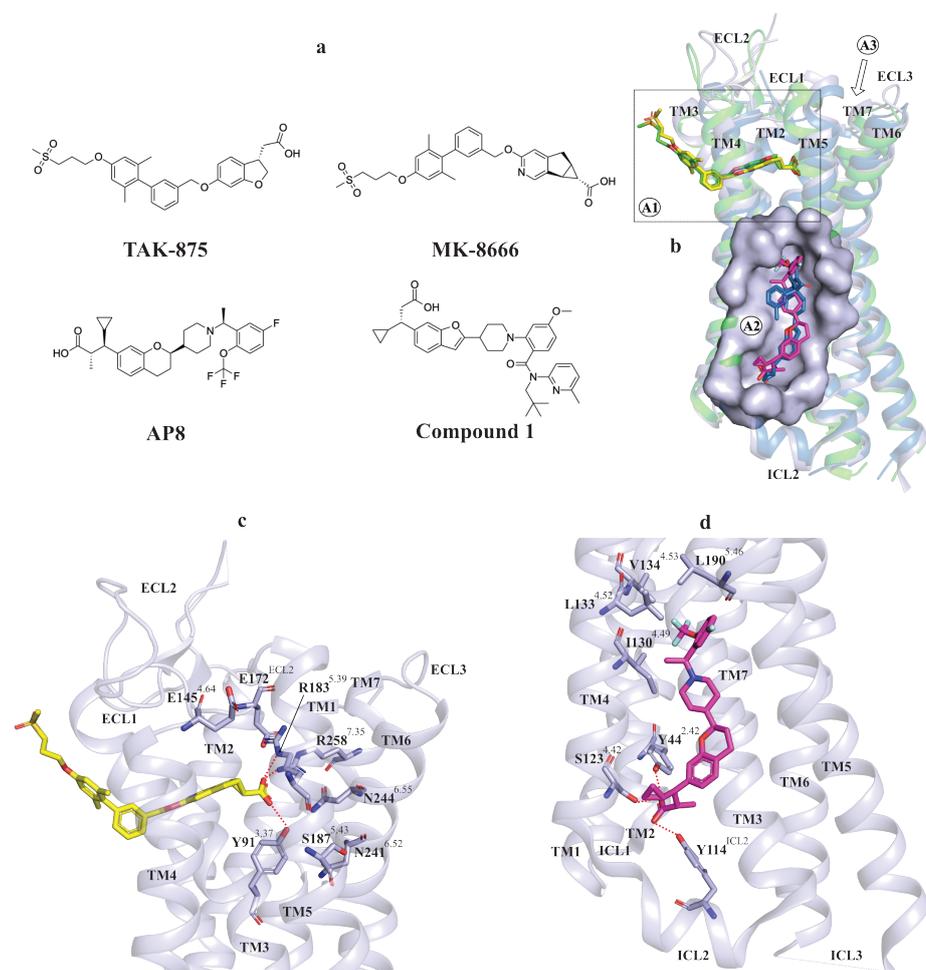
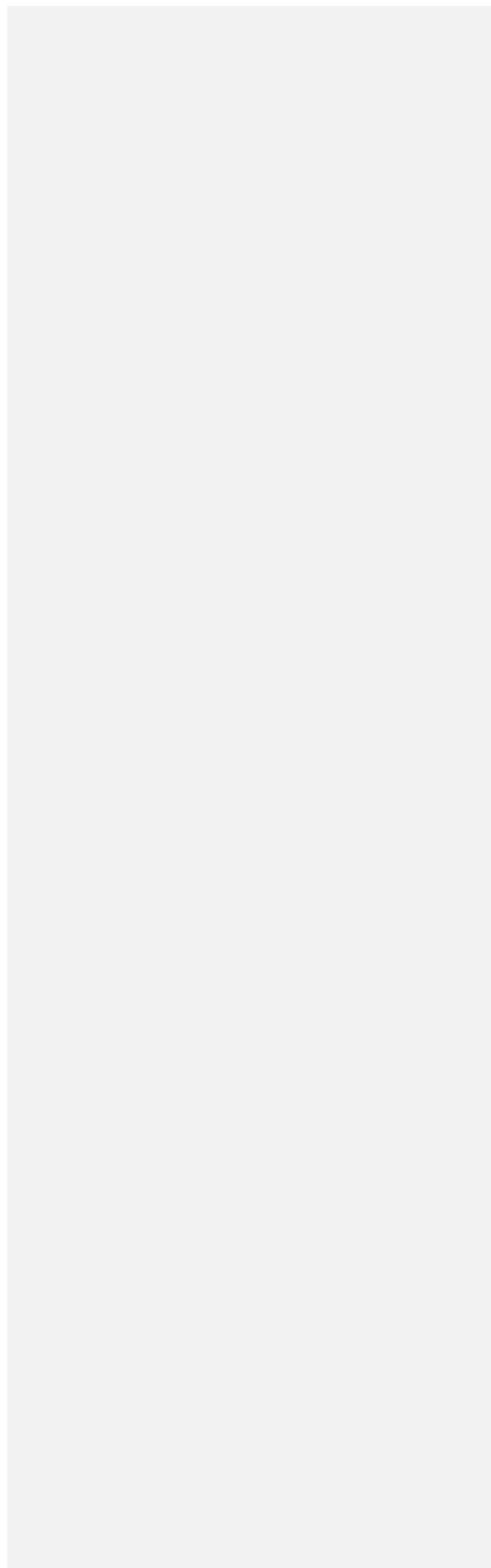


Figure 1. Allosteric binding sites A1 and A2 in FFAR1. (a) Chemical structure of the ligands co-crystallized with FFAR1; (b) binding sites of the different FFAR1 ligands: TAK-875 (green) and MK-8666 (yellow) bind to site A1, while the AP8 (magenta) and Compound 1 (blue) bind to site A2 (surface representation). The receptor is shown in cartoon representation (PDB ID: 4PHU in green, 5TZY in light-blue, and 5KW2 in sky-blue). Binding site details of (c) site A1 with

MK-8666 (yellow) and (d) site A2 with AP8 (magenta). The receptor (PDB ID: 5TZY) is shown in cartoon representation, while important residues are shown in stick representation.



2. MATERIALS AND METHODS

2.1. System Preparation for the MD

Since FFAR1 is in the inactive state in all the available crystallographic structures (possibly due to the stabilizing mutations and the insertion of T4 lysozyme into ICL3 [10]), the intermediate active structure was retrieved from the GPCRdb website (<http://www.gpcrdb.org>) [21]. All the systems reported in **Table S2** were prepared using a combination of High-Throughput Molecular Dynamics (HTMD) [22] and Tool Command Language (TCL) scripts. The disulfide bond between residues C79^{3.25} and C170^{6CL2} was included. The receptor orientation was determined from the Orientations of Proteins in Membranes (OPM) database entry 5TZY (<http://opm.phar.umich.edu/>) [23]. The hydrogen atoms were added using PDB2PQR [24] and PROPKA [25] software (considering a simulated pH of 7.0) coupled with visual inspection. The receptor was embedded in a 85 Å x 85 Å 1-palmitoyl-2-oleyl-sn-glycerol-3-phospho-choline (POPC) bilayer using the Visual Molecular Dynamics (VMD) Membrane Builder plugin 1.1 (<http://www.ks.uiuc.edu/Research/vmd/plugins/membrane/>) through an insertion method [26] with overlapping lipids removed (protein-lipid distance cut-off of 0.6 Å). The addition of the TIP3P water molecules to the simulation box (85 Å x 85 Å x 100 Å) was carried out using the VMD Solvate plugin v.1.5 (<http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/>). Finally, sodium and chloride ions were added to neutralize the system and mimic an ionic strength of 0.150 M using the VMD Autoionize plugin 1.3 (<http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/>).

2.2. System Equilibration and MD Settings

The equilibration and productive simulations were performed using ACEMD [27], employing the CHARMM36 force field [28, 29]. The ligands AP8 and MK-8666 (**Figure 1**) were parameterized in the CGenFF force field [30], with the exception of the AP8 rotatable C-O bond linking the trifluoromethoxy group to the fluorobenzene ring, and the MK-8666 rotatable bond involving the carboxylic acid and the cyclopropyl ring. These dihedral terms were respectively optimized using HTMD parametrize [22] and the VMD plugin fftk [31], after fragmentation of the molecule.

The Langevin thermostat [32] (target temperature: 300 K; low damping: 1 ps⁻¹; positional restraints on protein atoms: 1 kcal mol⁻¹ Å⁻²) and the Berendsen barostat [33] (target pressure: 1 atm) were used for the equilibration in isothermal-isobaric conditions (NPT) through a *four-stage* procedure employing an integration time step of 2 fs. The following *four-stage* procedure was used. *First stage*: clashes between protein and lipid atoms were reduced through 2000 conjugate-gradient minimization steps; *Second stage*: a 2 ns long MD simulation was run with positional restraints of 1 kcal mol⁻¹ Å⁻² on protein and lipid phosphorus atoms; *Third stage*: a 20 ns of MD simulation was performed restraining only the protein atoms; *Fourth stage*: positional restraints were applied only to the protein backbone alpha carbons, for a further 30 ns.

Productive trajectories were computed with an integration time step of 4 fs in the canonical ensemble (NVT) at 300 K using a thermostat damping of 0.1 ps⁻¹. The M-SHAKE algorithm [34] was used to constrain the covalent bonds involving hydrogen atoms. A 9 Å cut-off distance was set for electrostatic interactions, with a switching function applied beyond 7.5 Å. All simulations performed are summarized in **Table S2**. The particle mesh Ewald summation method (PME) [35] with a mesh spacing of 1.0 Å was employed to handle long-range Coulombic interactions.

2.3. SuMD - binding

The supervised MD (SuMD) [36, 37] is an adaptive sampling method that uses a tabu-like algorithm to monitor the distance between the centers of masses (or the geometrical centers) of the ligand and the predicted binding site, during short classical MD simulations. After each simulation, the distances (collected at regular time intervals) are fitted to a linear function and if the resulting slope is negative then the next simulation step starts from the last set of coordinates and velocities produced, otherwise, the simulation is restarted by randomly assigning the atomic velocities. This approach allows simulating the formation of intermolecular complexes in timescales that are one or two orders of magnitudes faster than the correspondent classical (unsupervised) MD simulations. Importantly, the sampling is gained without the introduction of any energetic bias.

MK-8666 was placed about 30 Å from its binding site, at the extracellular water/membrane interface, and the resulting system was prepared as reported in section 2.1. The distance between the MK-8666 (tetrahydrocyclopropa[4,5]cyclopenta[1,2-c]pyridine-6-carboxylic) moiety and the FFAR1 centroid computed on residues F87^{3.33} and F142^{4.61} (centroids distance) was supervised during successive 300 ns-long MD time windows. A single replica was run until the ligand reached the receptor, then, two different replicas were seeded (Replica 1 and Replica 2 in **Video S1**).

AP8 was placed about 30 Å from its binding site A2, in the intracellular POPC layer, and the resulting system was prepared as reported in section 2.1. The distance between AP8 and FFAR1 (centroid computed on residues A98^{3.44} and A99^{3.45}) was supervised during successive 300 ns-long MD time windows. No productive SuMD trajectories were sampled due to the presence of stable POPC residues at the FFAR1 site A2 (**Figure S2**).

2.4. SuMD - unbinding

Recently, the SuMD approach has been successfully applied to the unbinding of GPCR ligands [38]. In this study, giving the extended electrostatic network characterizing MK-8666 in the bound state, we have further modified and tested the protocol by supervising, besides the centroids distance, the number of water oxygen atoms that are within 4 Å from protein atoms involved in hydrogen bonds with the ligand (**Figure S3**). At the end of each MD time window, the simulation was considered productive (and the MD was restarted from it) if the slopes of both the two plotted linear functions were positive, or the last centroids distance was longer than the one from the previous productive time window. Otherwise, the coordinates from the last productive MD time window were used and the velocities were reassigned. The general rationale for this water supervision is that the displacement of ligands strongly stabilized by ionic interactions should be facilitated if the hydration of protein hydrophilic spots interacting with the ligand is enhanced.

Differently from the original SuMD binding algorithm, the length (Δt) of the short simulations performed increased along the unbinding pathway, according to the formula:

$$\Delta t = \Delta t_0 N t_{i2} \quad (1)$$

Δt_0 is the duration of the very first MD time window and Nt_i represents a factor that is picked from three user-defined values (Nt_1 , Nt_2 , and Nt_3), according to the last ligand-protein distance detected. Three distance threshold values (D_1 , D_2 , and D_3) were set and the ligand-protein distance (r_{L2}) at the end of each MD run was compared to these threshold values, allowing a decision on the value of the Nt_i factor according to the following conditions:

$$r_{L2} \leq D_{12} \rightarrow Nt_i = 1 \quad (2)$$

$$D_{12} < r_{L2} \leq D \rightarrow Nt_{i2} = Nt_{12} \quad (3)$$

$$D < r_{L2} \leq D_3 \rightarrow Nt_i = Nt \quad (4)$$

$$D_{32} < r_{L2} \rightarrow Nt_{i2} = Nt_{32} \quad (5)$$

For the FFAR1:MK-8666 unbinding, the initial time window length was 100 ps, with Nt_1 , Nt_2 , and Nt_3 set to 2, 4, and 8. Values of 10 Å, 14 Å, and 18 Å were used as D_1 , D_2 , and D_3 distances. The goal of increasing the simulation time window (Δt in **Equation 1**) along the unbinding pathway is to facilitate the sampling of metastable states, which could otherwise be poorly visited.

The unbinding was iterated until no ligand-protein van der Waals contact was detected by means of the GetContacts scripts tools (<https://getcontacts.github.io>). The ligand-protein distance and the number of water oxygen atoms within 4 Å of protein donor/acceptor atoms were computed employing PLUMED 2 [39]. After each productive MD time window, the GetContacts script (<https://getcontacts.github.io/index.html>) was employed to detect and update the protein atoms involved in hydrogen bonds with the ligand, considering a distance of 3.5 Å and an angle value of 120° as geometrical cut-offs. Notably, if no hydrogen bond between the ligand and the protein was present at the end of a productive MD time window, then protein atoms involved in water-mediated or van der Waals interactions were considered.

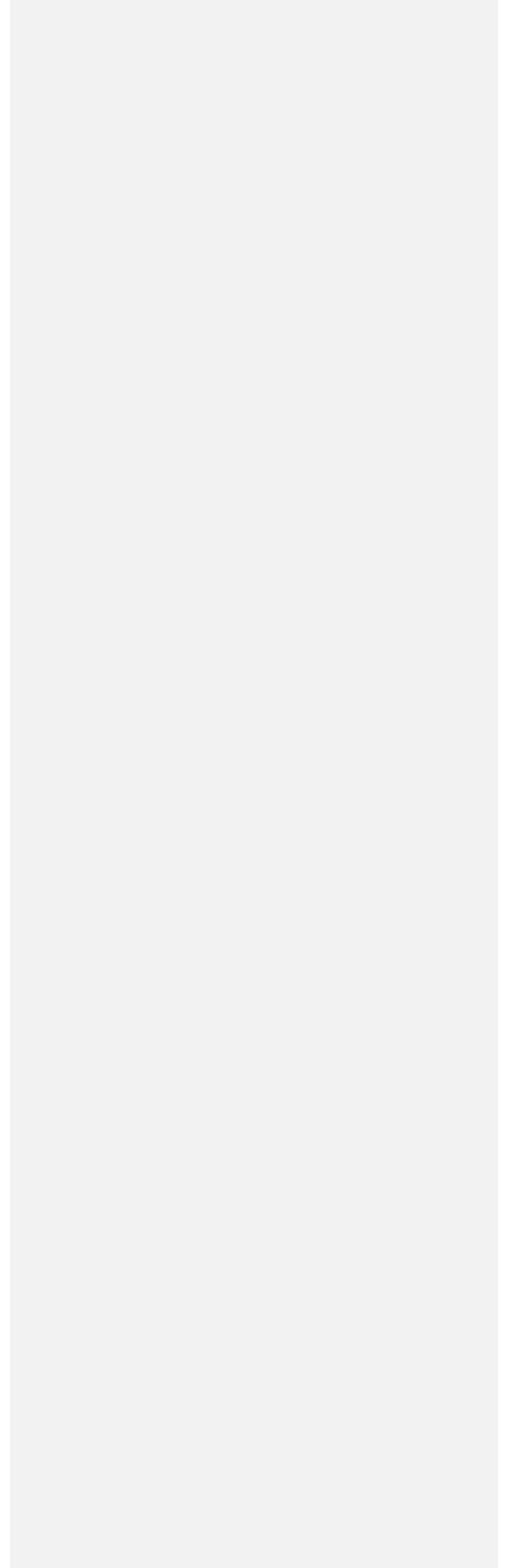
The FFAR1:MK-8666 and FFAR1:MK-8666:AP8 complexes were prepared for simulations as described in paragraph 2.1.

2.5. Analysis of the MD trajectories

Ligand-protein hydrogen bonds and van der Waals contacts were detected using the GetContacts scripts (<https://getcontacts.github.io/index.html>), setting a donor-acceptor distance of 3.3 Å and an angle value of 150° as geometrical cut-offs. Root mean square deviations (RMSD), root mean square fluctuations (RMSF), and dihedral angles were computed using VMD [40]. Distances between atoms were computed using PLUMED 2 [39] or VMD. The MMPBSA.py [41] script, from the AmberTools17 suite (The Amber Molecular Dynamics Package, at <http://ambermd.org/>), was used to compute molecular mechanics energies combined with the generalized Born and surface area continuum solvation (MM/GBSA) method [42], after transforming the CHARMM psf topology files to an Amber prmtop format using ParmEd (ParmEd documentation at <http://parmed.github.io/ParmEd/html/index.html>). The Poisson-Boltzmann surface area (PBSA) was not employed to speed up the computation, *i.e.* the membrane was not implicitly modelled.

2.6. General

Throughout the text, the Ballesteros-Weinstein GPCR universal numbering [43] is used alongside the normal residue numbers.



3. RESULTS AND DISCUSSION

3.1. Intermediate interactions between MK-8666 and ECL2 drive the FFAR1:MK-8666 complex formation

The structural data indicates a probable MK-8666 binding route to site A1 between TM3 and TM4 (**Figure 1**), no further experiments have yet confirmed this. Indeed, an alternative path between TM4 and TM5 was not ruled out [15]. Overall, the SuMD simulations corroborate the first hypothesis, as a binding path between TM3 and TM4 (**Figure 2**) was sampled without any important steric barrier (differently from AP8 binding to site A2, **Figure S2**). From the unbound state, the ligand approached the FFAR1 extracellular vestibule from the bulk solvent, before moving to the membrane/bulk interface and establishing more extensive contacts with the receptor (**Video S1**). Metastable states (M1 in the energy landscape reported in **Figure 2**) were characterized by MK-8666 carboxylate hydrogen bonds with FFAR1 residue S157^{ECL2} and the backbone, while hydrophobic contacts were formed with L158^{ECL2}, L171^{ECL2}, W174^{ECL2}, P80^{3,26}, and F142^{4,61} (**Figure 2c,d**). A conformational rearrangement of ECL2, with W174^{ECL2} acting as a sort of gatekeeper, allowed MK-8666 deeper into the receptor through the crevice between TM3 and TM4 to reach site A1, where it formed electrostatic interactions with R183^{5,39}, R258^{7,35}, and Y240^{6,51}. FFAR1 ECL2, which is rich of hydrophilic spots, could drive also the binding of LCFA from the membrane, stabilizing and channeling the polar head groups toward the final bound state. Interestingly, SuMD binding trajectories showed the disruption of an ionic lock between R183^{5,39} and E145^{4,64} upon binding of the partial agonist (**Video S1**); this ionic lock is one of the proposed switches for FFAR1 partial activation [13, 15–17]. Mutagenesis studies pointed out the E145^{4,64} - R183^{5,39} and E172^{ECL2} - R258^{7,35} salt bridges as possible alternative molecular switches involved in partial agonists binding to site A1 and the inactive/active transition of FFAR1. Our results corroborate this hypothesis, as the simulated association mechanism of the partial agonists MK-8666 showed the rupture of the E145^{4,64} - R183^{5,39} interaction, and the partial break of the E172^{ECL2} - R258^{7,35} ionic interaction, upon MK-8666 binding (**Video S1**).

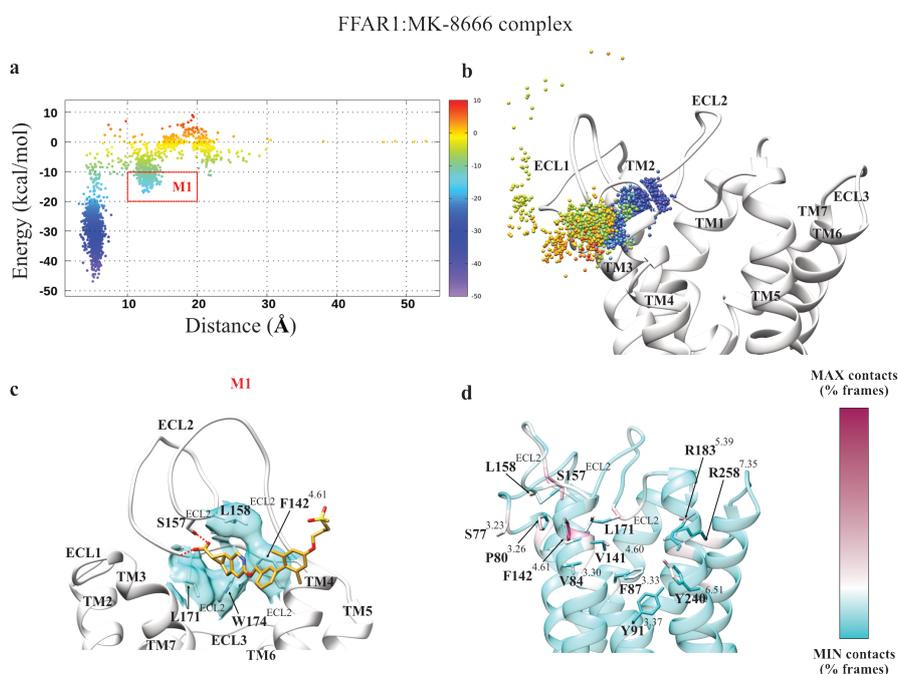


Figure 2. FFAR1:MK-8666 SuMD binding simulations. (a) Energy landscape showing the FFAR1:MK-8666 interaction energy during the recognition; the x-axis denotes the distance between the ligand and the receptor centroids; (b) MK-8666 centroid positions colored according to the interaction energy with FFAR1; (c) a representation of the FFAR1:MK-8666 interactions in the metastable states M1; (d) FFAR1 residues engaged during the simulated binding.

3.2. MK-8666 and AP8 are reciprocally stabilized during MD simulations

Simulations performed on the FFAR1 binary and ternary complexes (receptor bound to MK-8666 and AP8, **Table S1**, **Table S2**) highlighted several differences in the systems' dynamics. Interestingly, MK-8666 bound to site A1 gained stabilization in presence of AP8 (site A2), as shown by the shift of interaction energy towards lower values and the decrease of RMSF (**Figure 3**). This stabilization could be ascribed to improved contacts with FFAR1 residues S77^{3,23}, P80^{3,26}, V84^{3,30}, located at the top of TM3, and L140^{4,59} (**Figure 4**).

AP8, on the other hand, even though less influenced, was characterized by a RMSF decrease of the trifluoromethyl group (**Figure 3b**). This part of the molecule is in contact with the backbone of Y91^{3,37}, which possibly stabilizes the ligand in the ternary complex. A further comparison of the Root Mean Square Fluctuation (RMSF) between the different systems (**Figure S4**) indicates diminished flexibility of ECL2 and the top of TM4 upon binding of MK-8666, possibly influencing the site A2 and therefore AP8 stability. The AP8 interaction pattern in the ternary complex, compared to the FFAR1:AP8 binary one, was characterized by more contacts with L189^{5,45}, L190^{5,46}, Y144^{ICL2}, Y44^{2,42}, P40^{2,38}, and fewer

interactions with A92^{3,38} and Y91^{3,37}. Interestingly, in the ternary complex MK-8666 formed transient hydrogen bonds with the latter residue, losing interaction with Y240^{6,51} (Table S3, Figure 4).

In our model, the energy stabilization experienced by MK-8666 (Figure 3) in the ternary complex is putatively due to increased contacts with the top of TM3 and, to less extent, with TM4 (Figure 4). AP8, instead, formed better interactions with residues located at TM2, along with a general reorganization of the contacts with TM3, TM4, and TM5 (Figure 4).

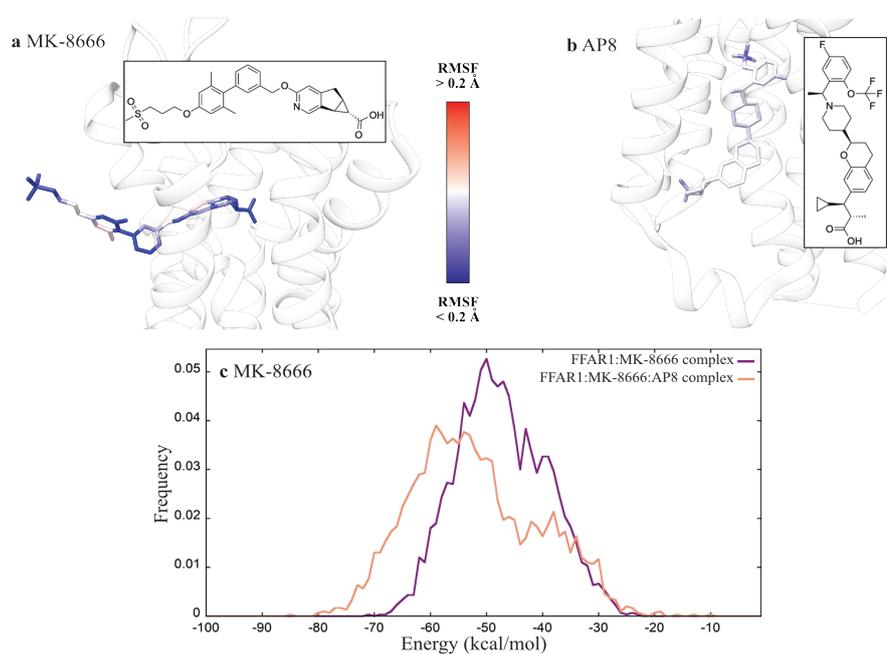


Figure 3. Ligand RMSF (a,b) and GBSA energy (c) distributions. (a) MK-8666 RMSF comparison between FFAR1:MK-8666 and FFAR1:MK-8666:AP8 complexes; (b) AP8 RMSF comparison between FFAR1:AP8 and FFAR1:MK-8666:AP8 complexes; the RMSF values (a,b) were computed after superposing the MD trajectories on the initial ligand coordinates (blue atoms had lower mobility in the ternary complex, while red atoms had higher mobility). (c) MK-8666 GBSA energy comparison between FFAR1:MK-8666 (purple) and FFAR1:MK-8666:AP8 complexes (pink). The shift toward lower energy values (of about 10 kcal/mol) of the FFAR1:MK-8666:AP8 complex in the GBSA distribution plot (c) indicates the increase in stability of MK-8666 in the presence of AP8.

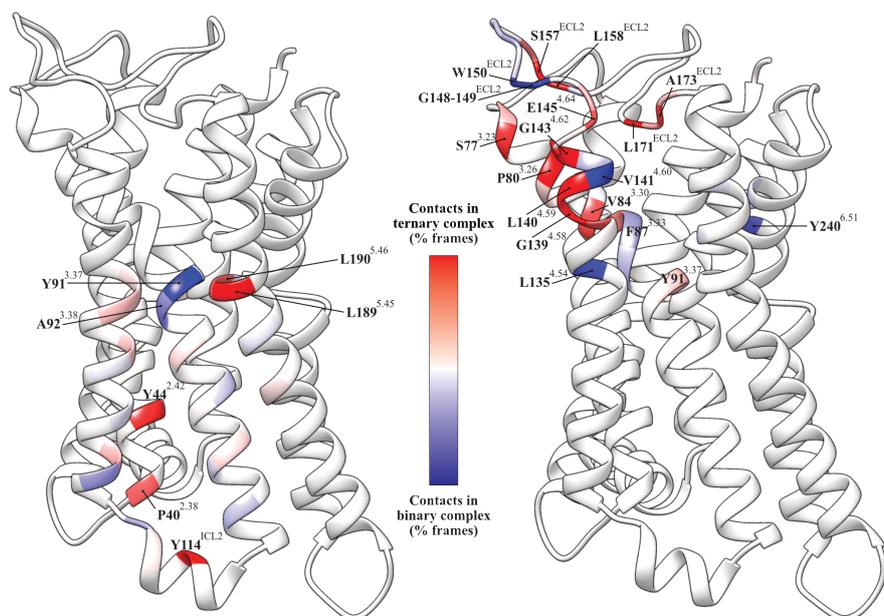


Figure 4. Protein-ligands contact differences between the FFAR1:MK-8666:AP8 ternary complex and the FFAR1:AP8 (left) and FFAR1:MK-8666 (right) binary complexes. In this figure, the receptor is shown in cartoon representation. The colour indicates whether a given residue makes more contacts in the binary complex (blue) or more contacts in the ternary complex (red). The contact residues are P40^{2.38}, Y44^{2.42}, Y91^{3.37}, A92^{3.38}, Y114^{CL2}, L189^{5.45}, and L190^{5.46} on the left, and S77^{3.23}, P80^{3.26}, V84^{3.30}, F87^{3.33}, Y91^{3.37}, L135^{4.54}, G139^{4.58}, L140^{4.59}, V141^{4.60}, G143^{4.62}, E145^{4.64}, G148^{ECL2}, G149^{ECL2}, W150^{ECL2}, S157^{ECL2}, L158^{ECL2}, L171^{ECL2}, A173^{ECL2}, and Y240^{6.51} on the right.

3.3. MK-8666 modifies the TM6 dynamic

The main hallmark of the class A GPCR activation is the outward movement of TM6, which, in the resting position, is stabilized by the ionic lock between R^{3.50} and E^{6.30} that is common to many GPCRs [11]. In FFAR1 the basic residue K219^{6.30} (part of RRR motif) cannot take part in this inter-helix salt bridge with R104^{3.50}. Moreover, TM6 in FFAR1 lacks the “toggle switch” residue W^{6.48}, which strongly influences GPCR activation [44]; position 6.48 is 81% tryptophan across all class A GPCRs. Position 6.48 in FFAR1 is occupied by V237^{6.48}, which lies just above the A2 binding site (Figure S5). The comparison of the V237^{6.48} dihedral angle distribution shows that the presence of AP8 bound to site A1 (binary complex FFAR1:AP8) changed the rotameric state of this residue (Figure S6) blocking a specific configuration, while the presence of MK-8666 favored a bimodal distribution. However, this conformational state of V237^{6.48} did not drive the overall flexibility of TM6. Indeed, as a general view, the presence of MK-8666 increased the FFAR1 TM6 flexibility (Figure 5). While the apo receptor and the binary FFAR1:AP8 complex have a similar dynamic in this region, the partial agonists bound to site A1 led to a wider TM6 tilt. It follows that the apoFFAR1 and the FFAR1:AP8 complex appeared constrained

compared to the FFAR1:MK-8666 complex. In the presence of both ligands (FFAR1:MK-8666:AP8 complex) TM6 appeared more prone to outward movements, while in the FFAR1:MK-8666 complex TM6 experienced numerous closed conformations (Figure 5). This possibly facilitates the recruitment of intracellular effectors ($G_{i/11}$, G_s proteins, and β -arrestins 1 and 2) as the TM6 outward movement is the key feature shaping the binding site of the cognate proteins.

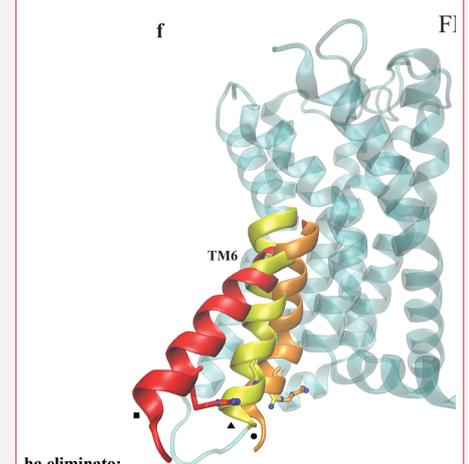
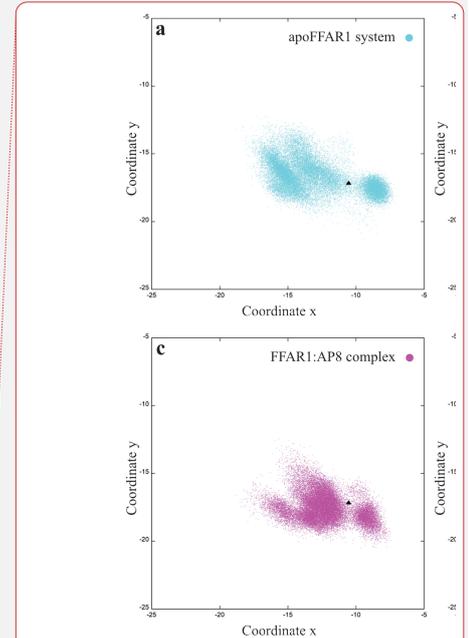
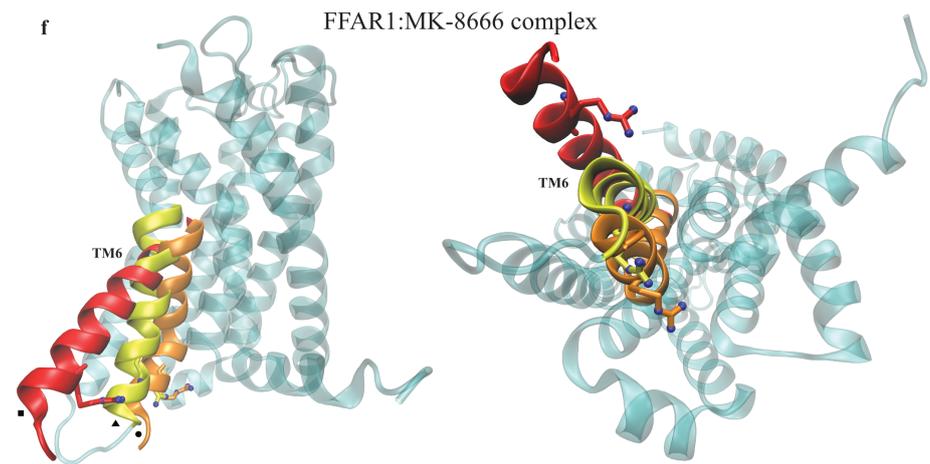
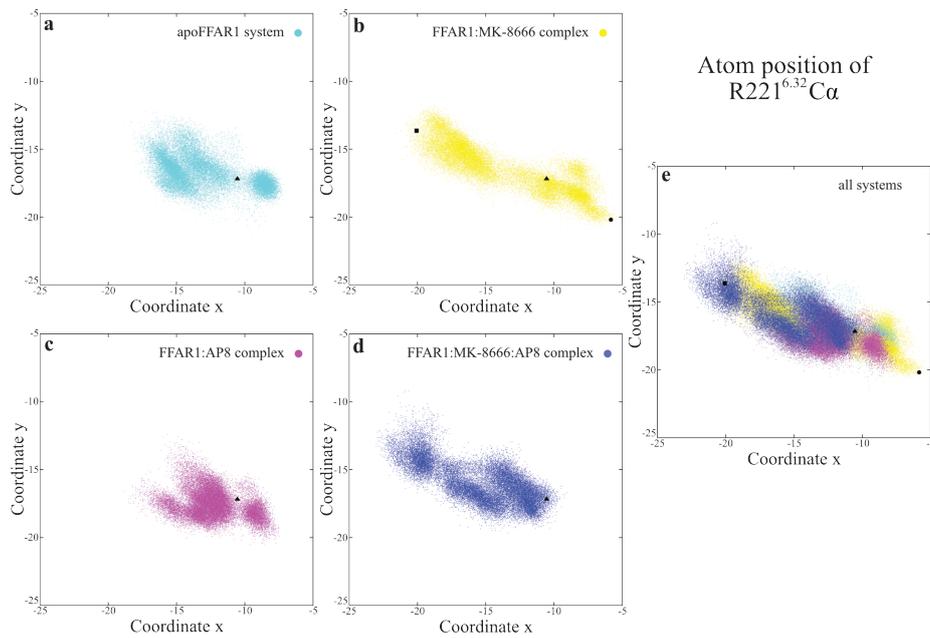


Figure 5. Atom positions of the C α atom of R221^{6.32}. The position of R221^{6.32} (shown in sticks representation) was taken to represent the movement of TM6, which is implicated in activation. Plots (a) to (e) present the x and y coordinates of the C α atom of R221^{6.32} respectively from (a) the apoFFAR1 (cyan), (b) the FFAR1:MK-8666 complex (yellow), (c) the FFAR1:AP8 complex (magenta), and (d) the FFAR1:MK-8666:AP8 ternary complex (blue). On the right, all systems are reported in the same plot (e). The triangle indicates the original PDB position of R221^{6.32} in each system, while the circle and the square illustrate the position of R221^{6.32} respectively in the closest and most open position of TM6 in the FFAR1:MK-8666 complex (b,f). TM6 is shown as a ribbon and colored according to the three positions previously reported (the triangle indicates the original PDB position of R221^{6.32}, the circle and the square show the position of R221^{6.32} respectively in the closest and most open position of TM6 in the FFAR1:MK-8666 complex).

3.4. AP8 influences MK-8666 unbinding from FFAR1

The putative FFAR1:MK-8666 dissociation mechanism (**Video S2**), simulated using a modified version of SuMD, was sampled both in the presence and absence of AP8 bound to site A2. In both of these systems, the ligand left receptor site A1 following an unbinding pathway between TM3 and TM4 that resembles the binding pathway (**Figure 6a**). During the unbinding from the binary complex, MK-8666 made more contacts with residues F87^{3.33}, V84^{3.30}, L138^{4.57}, G139^{4.58}, V141^{4.60}, F142^{4.61}, and R258^{7.35}, while in the ternary complex the partial agonist (MK-8666) engaged more residues located at the top of TM3 (A83^{3.29}, V81^{3.27}, P80^{3.26}, and S77^{3.23}) and on ECL2 (L144^{4.63} and L158^{ECL2}), as shown in **Figure 6c** and **Figure S7**. This shift in the barycenter of the interactions favored the retaining of either one of the two electrostatic interactions with R183^{5.39} or R258^{7.35} during the early step of unbinding. In the absence of AP8, indeed, MK-8666 moved from the bound state straight to the metastable configuration M1 (**Figure 6a,b**), where it engaged R258^{7.35}. In the ternary complex, instead, the ligand experienced several metastable states along the path characterized by the ionic interaction with R183^{5.39} (metastable states M2 in **Figure 6a,b**). Numerous stable states scattered between the bound state and the final unbound state of the FFAR1:MK-8666:AP8 complex suggest an influence of the full agonist on the dissociation path from site A1.

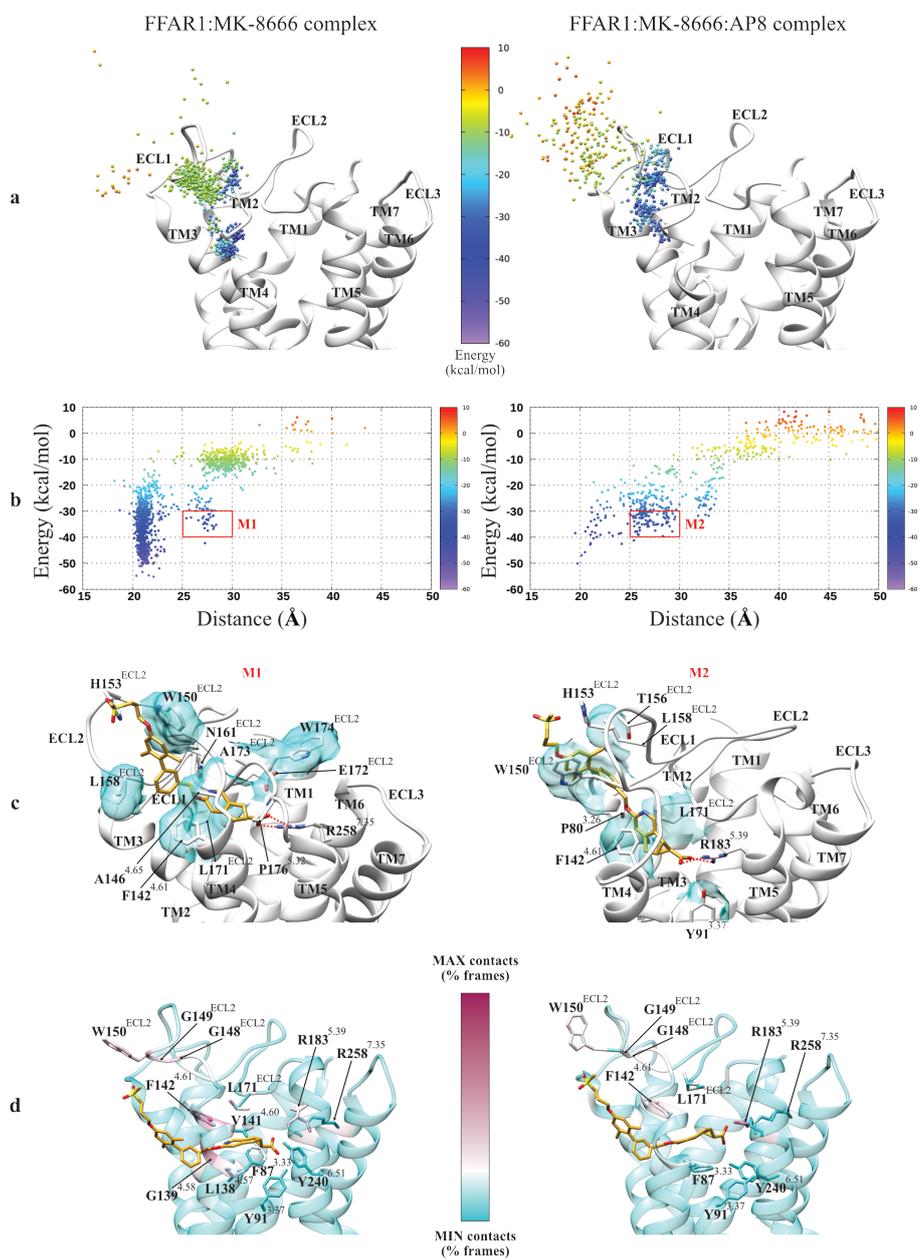
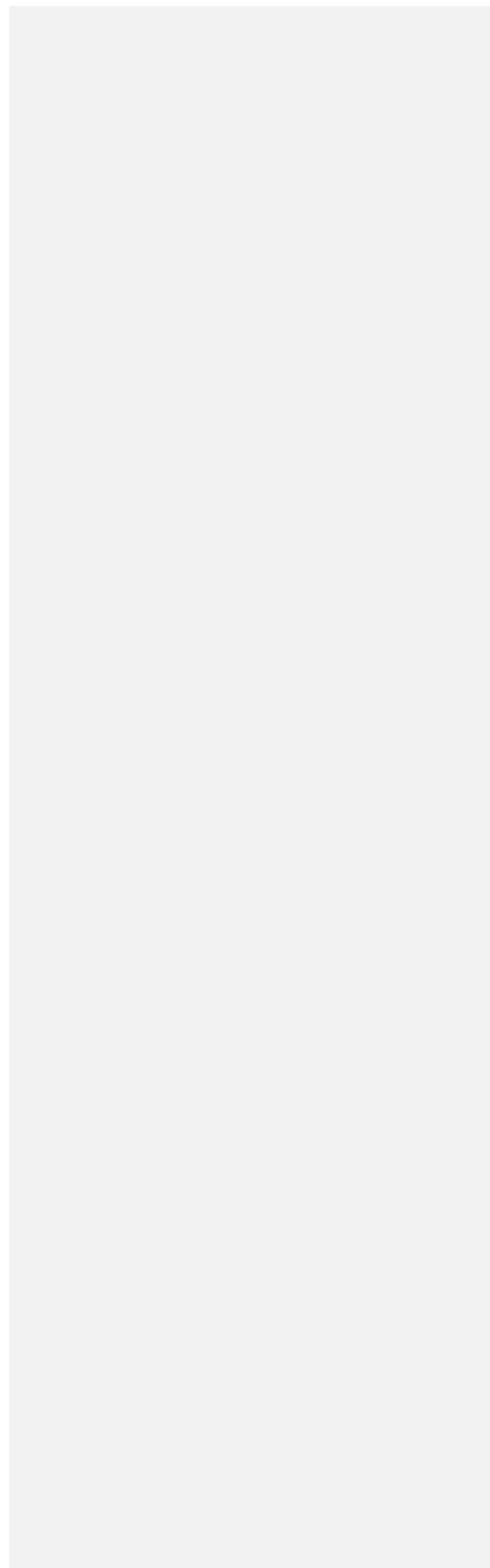


Figure 6. FFAR1:MK-8666 (site A1) SuMD unbinding simulations in the absence (left) and presence (right) of AP8 bound to site A2. (a) MK-8666 centroid positions colored according to the interaction energy with FFAR1 in the absence

(left) and presence (right) of AP8; (b) Energy landscape showing the FFAR1:MK-8666 interaction energy during the dissociation in the absence (left) and presence (right) of AP8. The distance is between the ligand and the receptor centroids; the receptor centroid lies about 20 Å below binding site A1. (c) FFAR1:MK-8666 interactions in the metastable states M1 and M2; (d) FFAR1 residues engaged during the simulated binding in the absence (left) and presence (right) of AP8.



4. CONCLUSION

The present computational work addresses the activation of FFAR1 by allosteric ligands. FFAR1, which is involved in glycemic control, diverges from the other class A GPCRs as it presents several distinctive structural features. X-ray structures of FFAR1 unveiled the presence of two distinct binding sites, namely A1 and A2, responsible respectively for the binding of partial and full agonists, and in reciprocal allosteric communication, where site A2 is a novel site on the external lipid facing surface of the TM bundle. FFAR1 does not bear the conserved structural motifs NP^{7.50}xxY, DR^{3.50}Y, and P^{5.50}l^{3.40}F^{6.44}, the highly conserved “toggle switch” W^{6.48} or the ionic lock between R^{3.50} and E^{6.30} that is common to many class A GPCRs [11, 12]. These characteristics, and the absence of a crystallized fully active reference state, ensure that the study of the FFAR1 activation mechanism is challenging. The pursuit of novel drugs able to tackle T2DM will increasingly consider FFAR1, as the structural knowledge of its activation and the allosteric mechanism is being unveiled by means of a plethora of different approaches. The X-ray crystal structures of FFAR1 paved the way for computational studies and rational structure-based drug design. This will likely lead to the development of new chemotypes capable to overcome the hepatotoxicity that prevented the approval of FFAR1 as therapeutics so far.

Our findings about the binding mechanism of TAK-875 to site A1 from the lipid/water interface highlighted the role of the extracellular vestibule, and ECL2 in particular, is in line with observations for the other GPCRs, where ECLs play a fundamental role in ligands binding (and functionality), providing the first interactions and favorably orienting the incoming molecule prior to the orthosteric site [45, 46].

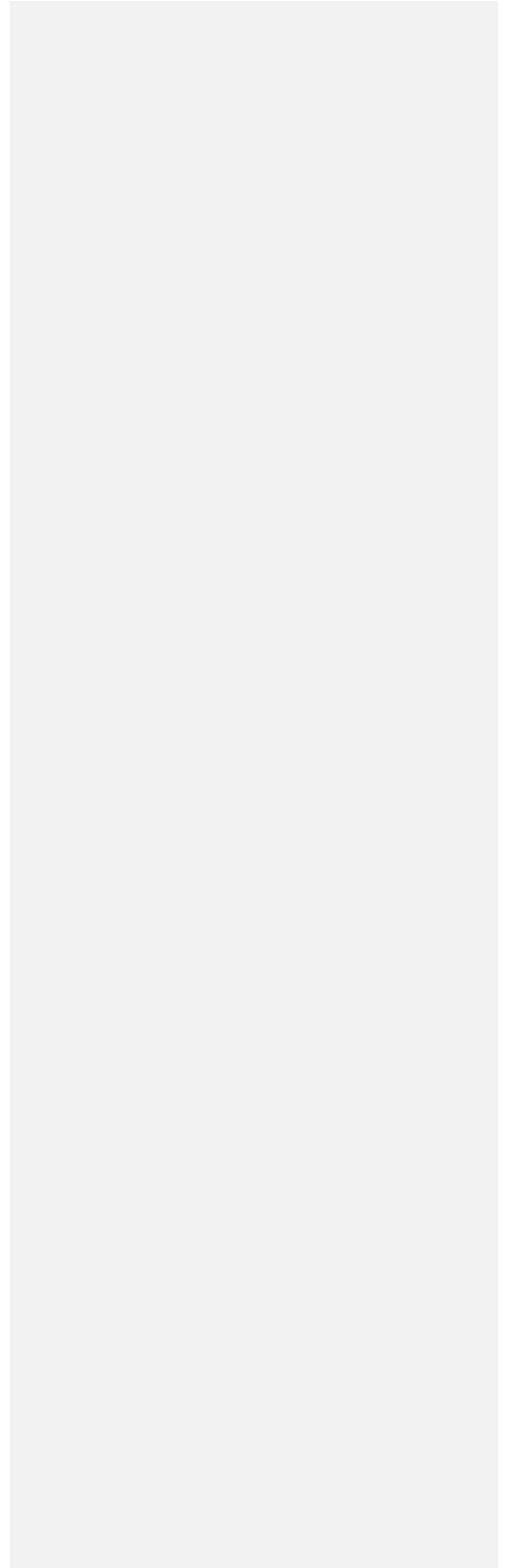
The allosteric cross-talking between FFAR1 sites A1 and A2 has created difficulties in classifying one of them as the orthosteric one. On one hand, site A1 is located within the TMD, consistent with the orthosteric site of other GPCRs, but trigger a partial activation, site A2, instead, is linked to a full activation but is located outside the TMD at the membrane interface as other GPCR allosteric sites characterized so far. Plausibly, in light of its exposure to the membrane, site A2 could be responsible for LCFA recognition under physiological conditions that are close to homeostasis, while the binding to the less (kinetically) accessible site A2 could take place when the local LCFA concentration increases. Intriguingly, the partial agonist TAK-875 is more effective in recruiting β -arrestins 1 and 2 than the endogenous ligands palmitate or oleate, which instead trigger coupling to the G_{q/11} path [47]. Such deference in signaling profiles may underlie distinct metabolic regulation exerted by site A1 and site A2.

According to MD simulations of the binary FFAR1:AP8 and FFAR1:MK-8666 complexes and the ternary FFAR1:MK-8666:AP8 complex, TM3 and TM4 could be fundamental for the allosteric communication between binding sites. The simulations pointed out a possible role for Y91^{3.37} at the interface between the sites A1 and A2 [48], as the presence of AP8 slightly favored the hydrogen bonding with the MK-8666 carboxylate, thereby stabilizing the latter.

Focusing on other structural elements of FFAR1, TM6 showed variable degree of flexibility in the different complexes. As a general view, the presence of MK-8666 bound to site A1 increased TM6 mobility, with the greater flexibility occurring in the ternary complex FFAR1:MK-8666:AP8. In the FFAR1 crystal structures, H8 was not resolved. The lack of structure is likely due to the numerous glycine residues in the primary sequence of the receptor. Simulations confirmed this instability, as the modelled H8 was unstable and unstructured in the many MD replicas. However, a slight tendency to keep helicity was detected when AP8 was bound to A2, probably due to a general stabilization of the intracellular side of the receptor.

Finally, SuMD unbinding of MK-8666 highlighted different mechanisms in the presence or absence of AP8. More precisely, the MK-8666 contacts with the top of TM3 were improved in the presence of AP8, and this may stabilize the ligand and hinder its dissociation, consistent with the slower dissociation rates measured for partial agonists in the ternary FFAR1 complexes [13]. For the first time, the SuMD protocol was extended to a second metric of the system, beside the distance between the ligand-protein centroids. This implementation could facilitate the unbiased simulation of the unbinding of ligands that, like MK-8666, are particularly stabilized by buried (or numerous) hydrogen bonds.

In conclusion, this work delivered computational insights on some important aspects of the poorly known FFAR1 activation and allostery, providing a plethora of working hypotheses that we hope will be experimentally explored in future efforts.



REFERENCES

1. & Kotarsky K, Nilsson NE, Flodgren E, et al (2003) A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem Biophys Res Commun* 301:406–10.
2. & Itoh Y, Kawamata Y, Harada M, et al (2003) Free fatty acids regulate insulin secretion from pancreatic beta cells through GPR40. *Nature* 422:173–6.
3. & Edfalk S, Steneberg P, Edlund H (2008) Gpr40 is expressed in enteroendocrine cells and mediates free fatty acid stimulation of incretin secretion. *Diabetes* 57:2280–7.
4. & World Health Organization (2016) Diabetes mellitus fact sheet. <https://www.who.int/news-room/fact-sheets/detail/diabetes>
5. & Watterson KR, Hudson BD, Ulven T, Milligan G (2014) Treatment of Type 2 Diabetes by Free Fatty Acid Receptor Agonists. *Front Endocrinol (Lausanne)* 5:137.
6. & Li Z, Xu X, Huang W, Qian H (2018) Free Fatty Acid Receptor 1 (FFAR1) as an Emerging Therapeutic Target for Type 2 Diabetes Mellitus: Recent Progress and Prevailing Challenges. *Med Res Rev* 38:381–425.
7. & Burant CF, Viswanathan P, Marcinak J, et al (2012) TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet (London, England)* 379:1403–11.
8. & Tikhonova IG (2017) Application of GPCR Structures for Modelling of Free Fatty Acid Receptors. *Handb Exp Pharmacol* 236:57–77.
9. & Tikhonova IG, Poerio E (2015) Free fatty acid receptors: structural models and elucidation of ligand binding interactions. *BMC Struct Biol* 15:16.
10. & Ho JD, Chau B, Rodgers L, et al (2018) Structural basis for GPR40 allosteric agonism and incretin stimulation. *Nat Commun* 9:1645.
11. & Latorraca NR, Venkatakrishnan AJ, Dror RO (2017) GPCR Dynamics: Structures in Motion. *Chem Rev* 117:139–155.
12. & Weis WI, Kobilka BK (2018) The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu Rev Biochem* 87:897–919.
13. & Lin DCH, Guo Q, Luo J, et al (2012) Identification and pharmacological characterization of multiple allosteric binding sites on the free fatty acid 1 receptor. *Mol Pharmacol* 82:843–59.
14. & Yabuki C, Komatsu H, Tsujihata Y, et al (2013) A novel antidiabetic drug, faglifam/TAK-875, acts as an ago-allosteric modulator of FFAR1. *PLoS One* 8:e76280.
15. & Srivastava A, Yano J, Hirozane Y, et al (2014) High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* 513:124–7.
16. & Sum CS, Tikhonova IG, Costanzi S, Gershengorn MC (2009) Two arginine–glutamate ionic locks near the extracellular surface of FFAR1 gate receptor activation. *J Biol Chem* 284:3529–36.
17. & Lu J, Byrne N, Wang J, et al (2017) Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat Struct Mol Biol* 24:570–577.
18. & Lückmann M, Trauelsen M, Bentsen MA, et al (2019) Molecular dynamics-guided discovery of an ago-allosteric modulator for GPR40/FFAR1. *Proc Natl Acad Sci U S A* 116:7123–7128.
19. & Stanley N, Pardo L, Fabritiis G De (2016) The pathway of ligand entry from the membrane bilayer to a lipid G

protein-coupled receptor. *Sci Rep* 6:22639.

20. & Jakowiecki J, Filipek S (2016) Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *J Chem Inf Model* 56:2457–2466.
21. & Isberg V, Mordalski S, Munk C, et al (2016) GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 44:D356–D364.
22. & Doerr S, Harvey MJ, Noé F, De Fabritiis G (2016) HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* 12:1845–52.
23. & Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22:623–5.
24. & Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665-7.
25. & Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* 7:525–37.
26. & Sommer B (2013) Membrane Packing Problems: A short Review on computational Membrane Modeling Methods and Tools. *Comput Struct Biotechnol J* 5:e201302014.
27. & Harvey MJ, Giupponi G, Fabritiis G De (2009) ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* 5:1632–9.
28. & Klauda JB, Venable RM, Freites JA, et al (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* 114:7830–43.
29. & Huang J, MacKerell AD (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34:2135–45.
30. & Vanommeslaeghe K, Hatcher E, Acharya C, et al (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31:671–90.
31. & Mayne CG, Saam J, Schulten K, et al (2013) Rapid parameterization of small molecules using the Force Field Toolkit. *J Comput Chem* 34:2757–70.
32. & Loncharich RJ, Brooks BR, Pastor RW (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* 32:523–35.
33. & Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690.
34. & Kräutler V, Van Gunsteren WF, Hünenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* 22:501–508.
35. & Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
36. & Sabbadin D, Moro S (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *J Chem Inf Model* 54:372–6.
37. & Cuzzolin A, Sturlese M, Deganutti G, et al (2016) Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J Chem Inf Model* 56:687–705.
38. & Deganutti G, Moro S, Reynolds CA (2020) A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. *J Chem Inf Model* 60:1804–1817.

39. & Tribello GA, Bonomi M, Branduardi D, et al (2014) PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 185:604–613.
40. & Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–8, 27–8.
41. & Miller BR, McGee TD, Swails JM, et al (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* 8:3314–21.
42. & Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–61.
43. & Isberg V, de Graaf C, Bortolato A, et al (2015) Generic GPCR residue numbers – aligning topology maps while minding the gaps. *Trends Pharmacol Sci* 36:22–31.
44. & Katritch V, Fenalti G, Abola EE, et al (2014) Allosteric sodium in class A GPCR signaling. *Trends Biochem Sci* 39:233–44.
45. & Wheatley M, Wootten D, Conner MT, et al (2012) Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* 165:1688–1703.
46. & Peeters MC, van Westen GJP, Li Q, IJzerman AP (2011) Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol Sci* 32:35–42.
47. & Mancini AD, Bertrand G, Vivot K, et al (2015) β -Arrestin Recruitment and Biased Agonism at Free Fatty Acid Receptor 1. *J Biol Chem* 290:21131–40.
48. & Teng D, Chen J, Li D, et al (2020) Computational Insights into Molecular Activation and Positive Cooperative Mechanisms of FFAR1 Modulators. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.0c00030>

LIST OF ABBREVIATIONS

FFAR1 = Free fatty acid receptor 1 &

GPCR = G protein-coupled receptor &

T2DM = Type 2 diabetes mellitus &

LCFA = Long-chain free fatty acids &

MD = Molecular dynamics &

cMD = Classic molecular dynamics &

SuMD = Supervised molecular dynamics &

GLP-1 = Glucagon-like peptide 1 &

GIP = Gastric inhibitory polypeptide &

ICL1-3 = Intracellular loop 1-3 &

TM1-7 = Transmembrane 1-7 &

H8 = Helix 8 &

TMD = Transmembrane domain &

ECL1-3 = Extracellular loop 1-3 &

POPC = 1-palmitoyl-2-oleyl-sn-glycerol-3-phospho-choline &

RMSD = Root mean square deviations &

RMSF = Root mean square fluctuation &