Supporting Information.

Probing the Binding Site of Bile Acids in TGR5

Authors: Antonio Macchiarulo, Antimo Gioiello, Charles Thomas, Thijs W.H. Pols, Roberto Nuti, Cristina Ferrari, Nicola Giacchè, Francesca De Franco, Mark Pruzanski, Johan Auwerx, Kristina Schoonjans, Roberto Pellicciari.

Table of Contents:

Figure S3. Immunofluorescence staining experiments in CHO cells confirming the presence of the wild-type and mutant TGR5 protein (red staining) at the cell membrane. Nuclei were counterstained with DAPI (Blue).

Figure S4. Binding poses of *S*-EMCA (INT-777, **4**) as resulting from the second round of docking studies. a) Binding mode 3 compliant to mutagenesis data; b) binding mode 4 not compliant to mutagenesis data (presence of hydrogen bonding with Asn76; c) binding modes 5-7 not compliant to mutagenesis data (lack of hydrogen bonding with Asn93).

Figure S5. a) ^{1}H-NMR and b) ^{13}C-NMR spectra of **6**.

Figure S6. Receptor atom shells of energy minimization: no constrain force constant applied to first shell (red atoms); 200 kcal/mol Å^2 force constant applied to second shell (green atoms); 400 kcal/mol Å^2 force constant applied to outer shell (gray atoms).

Experimental Parts.

Homology modelling. The crystal structure of bovine rhodopsin (PDB code: 1L9H) was used as template to build a homology model of human TGR5. The primary sequences of the bovine rhodopsin GPCR and human TGR5 share a 15% of identities (percentage of residues that are identical between the sequences) and a 26% of similarities (percentage of residues according to the BLOSUM 62 similarity matrix). The sequence alignment was guided by the highly conserved amino acid residues including the D(E)RY motif (Asp3.49, Arg3.50, and Tyr3.51) and three proline residues (Pro4.60, Pro6.50, and Pro7.50) in the TM segments of the GPCRs, using Prime software of Schrödinger. Starting from the alignment of these residues, the alpha helixes were constrained to avoid presence of insertions/deletions in them. Since cysteine residues in TM3 and EL2 are conserved between rhodopsin receptor and TGR5, a disulfide bridge linking such residues was constructed to constrain the geometry around the EL2.

The homology model of TGR5 was submitted to minimization protocols using OPLS-2005 all-atom force field. The Surface Generalized Born (SGB) continuum solvation model was used to mimic the solvation effect. The energy minimization was performed using a cycle of 500 steps of Steepest Descendent algorithm followed by several cycles of truncated-Newton algorithm until reaching a gradient of 0.01 Kcal/mol. A geometric validation of the 3D model was carried out using Procheck and Verify3D servers.[1] If present, bad geometries were manually corrected and the structure minimized again with the above protocol.

Docking calculations. All ligands were prepared with the ligand preparation tool implemented in the Schrödinger Suite. In particular, we considered the ionization states at the physiological pH of 7.0 for each compound. Docking calculations into the binding site of TGR5, as defined by Asn93 and Glu169 residues, were carried out using Glide 5.5 and the extra precision (XP) protocol. During these experiments, the induced fit protocol of Glide was employed to take into account ligand

induced conformational rearrangements of binding site residues. In particular, the van der Waals radii of receptor atoms endowed with a partial charge less than 0.25 were scaled by a factor of 0.70 to reduce possible steric clashes with ligand atoms. Two solutions for each ligand, namely the top scored one and its most diverse solution among the best ranked solutions, were stored for the analysis of the binding modes (Tables S1 and S2). In particular, the calculation of the root mean square deviation (RMSD) of heavy atoms from the top scored solution was used to identify the second solution of docking experiments.

Starting from the binding mode 2 (*head-to-tail*) of *S*-EMCA (**4**), second round of docking studies was preceded by energy minimization of binding site residues, after removing the ligand from the binding cleft of TGR5. In particular, receptor atoms lying within a first shell of 8Å from the ligand were submitted to energy minimization without restrains; receptor atoms lying between 8Å and 14Å from the ligand defined a second shell and were constrained on their atomic coordinates, using a force constant of 200 kcal/mol $Å^2$; all receptor atoms external to the former two shells were tethered on their atomic coordinates using a force constant of 400 kcal/mol $Å^2$ (Figure S6). The energy minimization cycle was carried out using Macromodel with the OPLS-2005 force field and the SGB continuum solvation model. The truncated-Newton algorithm was used with 1000 maximum number of iterations and a threshold of 0.01 kcal/mol as energy convergence criteria. Induced fit docking of **4** was then carried out using a looser steric factor, with the reduction of van der Waals radii of receptor atoms with a partial charge less than 0.25 by a factor of 0.50 (Table S3). Overall, this strategy was motivated by the fact that BAs have a large and rigid steroid scaffold, making difficult docking studies into the narrow binding pocket of TGR5 as resulting from the homology modelling.

Mutagenesis and Immunofluorescence. Mutants of TGR5 were generated by site-directed mutagenesis using specific primers and standard cloning techniques. For immunofluorescence, cells

were grown on Labtek II chamber slides (Nunc) and fixed with Shandon Formal Fixx (Thermo Scientific). Cells were stained using our home-made TGR5 antibody,[2] followed by Cy3-labeled anti-rabbit antibodies. Nuclei were subsequently stained with 4′,6-diamidino-2-phenylindole (DAPI), and cells were embedded with 1,4-diazabicyclo(2.2.2)octane (DABCO). Images (Figure S3) were acquired with a LSM700 confocal microscope (Zeiss).

Biological assay. The CREB-luciferase reporter construct (Stratagene) contained four copies of the CRE enhancer sequence and was transfected in CHO cells with or without (mutant) TGR5 and CMV-β-gal expression vector to correct for transfection efficiency (Clontech) using JetPEI (Polyplus Transfection). Luciferase activity was measured with the luciferase assay system (Promega) in the Victor X4 (PerkinElmer).

Synthetic chemistry. The synthesis of *S*-EMCA (**4**) and 6-ECDCA (**5**) have been performed as previously reported.[3,4] 3α,7α-dihydroxy-6α-ethyl-24-nor-5β-cholanylamine (**6**) was prepared from 6-ECDCA (**5**) according to our previously reported procedure. [5] Thus, the 3,7-diacetyl-protected 6-ECDCA was converted into the corresponding acyl azides *via* the acyl chloride intermediate (thionyl chloride) followed by treatment with aqueous sodium azide. The crude acyl azide mixture was then refluxed overnight in EtOH to give the corresponding *N*-carbethoxyamino derivative, which upon treatment with methanolic sodium hydroxide at reflux gave the desired product **6**, in 53% yield.

References

1. (a) Laskowski, R. A.; Moss, D. S.; Thornton, J. M., Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **1993,** *231* (4), 1049-67; (b) Bowie, J. U.; Luthy, R.; Eisenberg, D., A method to identify protein sequences

that fold into a known three-dimensional structure. *Science* **1991,** *253* (5016), 164-70; (c) Luthy, R.; Bowie, J. U.; Eisenberg, D., Assessment of protein models with three-dimensional profiles. *Nature* **1992,** *356* (6364), 83-5.

2. Pols, T. W.; Nomura, M.; Harach, T.; Lo Sasso, G.; Oosterveer, M. H.; Thomas, C.; Rizzo, G.; Gioiello, A.; Adorini, L.; Pellicciari, R.; Auwerx, J.; Schoonjans, K., TGR5 activation inhibits atherosclerosis by reducing macrophage inflammation and lipid loading. *Cell Metab* **2011,** *14* (6), 747-57.

3. Pellicciari, R.; Gioiello, A.; Macchiarulo, A.; Thomas, C.; Rosatelli, E.; Natalini, B.; Sardella, R.; Pruzanski, M.; Roda, A.; Pastorini, E.; Schoonjans, K.; Auwerx, J., Discovery of 6alpha-ethyl-23(S)-methylcholic acid (S-EMCA, INT-777) as a potent and selective agonist for the TGR5 receptor, a novel target for diabesity. *J Med Chem* **2009,** *52* (24), 7958-61

4. Pellicciari, R.; Fiorucci, S.; Camaioni, E.; Clerici, C.; Costantino, G.; Maloney, P. R.; Morelli, A.; Parks, D. J.; Willson, T. M., 6alpha-ethyl-chenodeoxycholic acid (6-ECDCA), a potent and selective FXR agonist endowed with anticholestatic activity. *J Med Chem* **2002,** *45* (17), 3569-72.

5. (a) Pellicciari, R.; Costantino, G.; Camaioni, E.; Sadeghpour, B. M.; Entrena, A.; Willson, T. M.; Fiorucci, S.; Clerici, C.; Gioiello, A., Bile acid derivatives as ligands of the farnesoid X receptor. Synthesis, evaluation, and structure-activity relationship of a series of body and side chain modified analogues of chenodeoxycholic acid. *J Med Chem* **2004,** *47* (18), 4559-69.

Table S1: Binding poses of S-EMCA (INT-777, **4**) as resulting from docking studies.

Pose	RMSD (Å)	IFD Score (kcal/mol)[a]	Binding Mode
#1	**0.00**	**-622.3**	**1** (*tail-to-head pose)*[b]
#2	0.50	-622.3	1
#3	0.47	-622.1	1
#4	0.99	-621.5	1
#5	4.36	-618.6	1
#6	8.81	-618.0	2
#7	8.71	-617.7	2
#8	**8.84**	**-617.7**	**2** (*head-to-tail pose)* [b]
#9	8.66	-617.5	2

[a] IFD Score = 1.0*(GlideScore) + 0.05*(Prime_Energy); [b] Selected binding poses are typed in bold.

Table S2: Binding poses of 6-ECDCA (Obeticholic acid, OCA, INT-747, **5**) as resulting from docking studies.

Pose	RMSD (Å)	IFD Score (kcal/mol)[a]	Binding Mode
#1	**0.00**	**-621.8**	**1** (*tail-to-head pose)*[b]
#2	0.57	-621.4	1
#3	0.36	-621.3	1
#4	0.72	-621.3	1
#5	8.27	-617.7	2
#6	**8.35**	**-617.7**	**2** (*head-to-tail pose)*[b]
#7	8.13	-617.6	2
#8	1.80	-617.6	1
#9	8.27	-617.5	2

[a] IFD Score = 1.0*(GlideScore) + 0.05*(Prime_Energy); [b]Selected binding poses are typed in bold.

Table S3: Binding poses of S-EMCA (INT-777, **4**) as resulting from the second round of docking studies.

Pose	RMSD (Å)	IFD Score (kcal/mol)[a]	Binding Mode (N93A- N76A Compliance)
#1	**0.00**	**-634.1**	**3 (Compliance)** [b]
#2	1.29	-633.6	4 (No-Compliance)
#3	0.53	-633.6	3 (Compliance)
#4	1.22	-633.5	4 (No-Compliance)
#5	0.68	-633.5	3(Compliance)
#6	4.02	-633.5	5 (No-Compliance)
#7	1.22	-632.9	4 (No-Compliance)
#8	5.39	-631.9	6 (No-Compliance)
#9	2.99	-631.8	7 (No-Compliance)

[a] IFD Score = 1.0*(GlideScore) + 0.05*(Prime_Energy); [b]Selected binding pose is typed in bold.

Figure S1

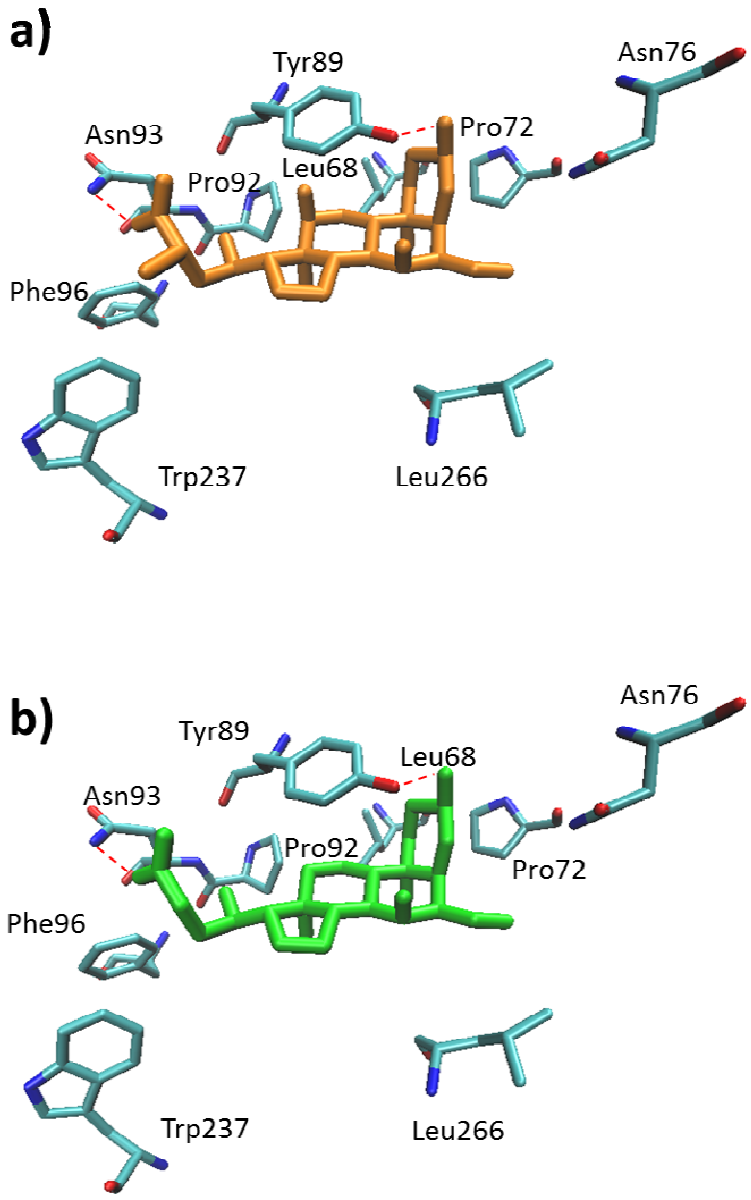

a)
Tyr89
Asn76
Asn93
Pro72
Pro92
Leu68
Phe96
Trp237
Leu266
b)
Asn76
Tyr89
Leu68
Asn93
Pro92
Pro72
Phe96
Trp237
Leu266

Figure S2

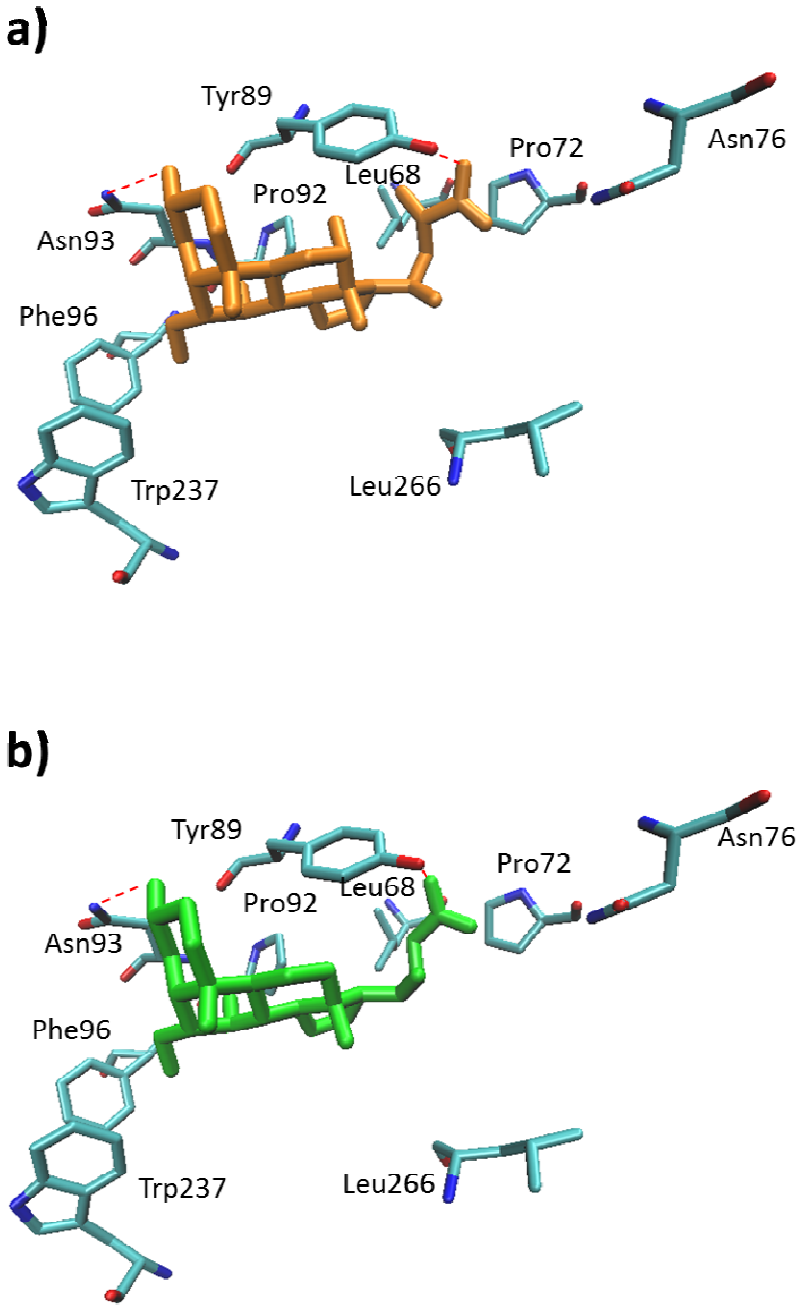
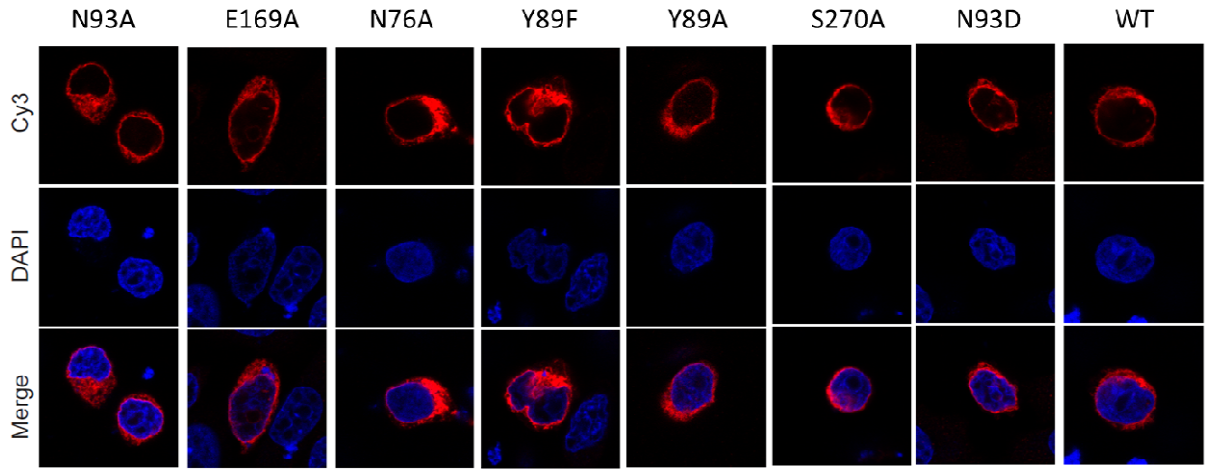

a)
Tyr89
Pro72
Asn76
Leu68
Pro92
Asn93
Phe96
Trp237
Leu266
b)
Tyr89
Asn76
Pro72
Leu68
Pro92
Asn93
Phe96
Trp237
Leu266

Figure S3



Immunofluorescence staining experiments in CHO cells confirming the presence of the wild-type and mutant TGR5 protein (red staining) at the cell membrane.. Nuclei were counterstained with DAPI (Blue).

Figure S4

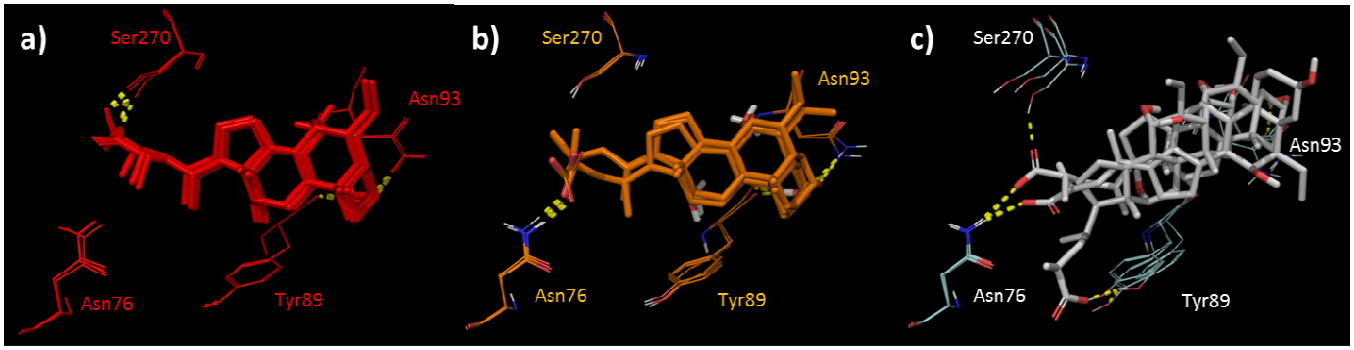

a)
Ser270
Asn93
Asn76
Tyr89
b)
Ser270
Asn93
Asn76
Tyr89
c)
Ser270
Asn93
Asn76
Tyr89

Figure S5a

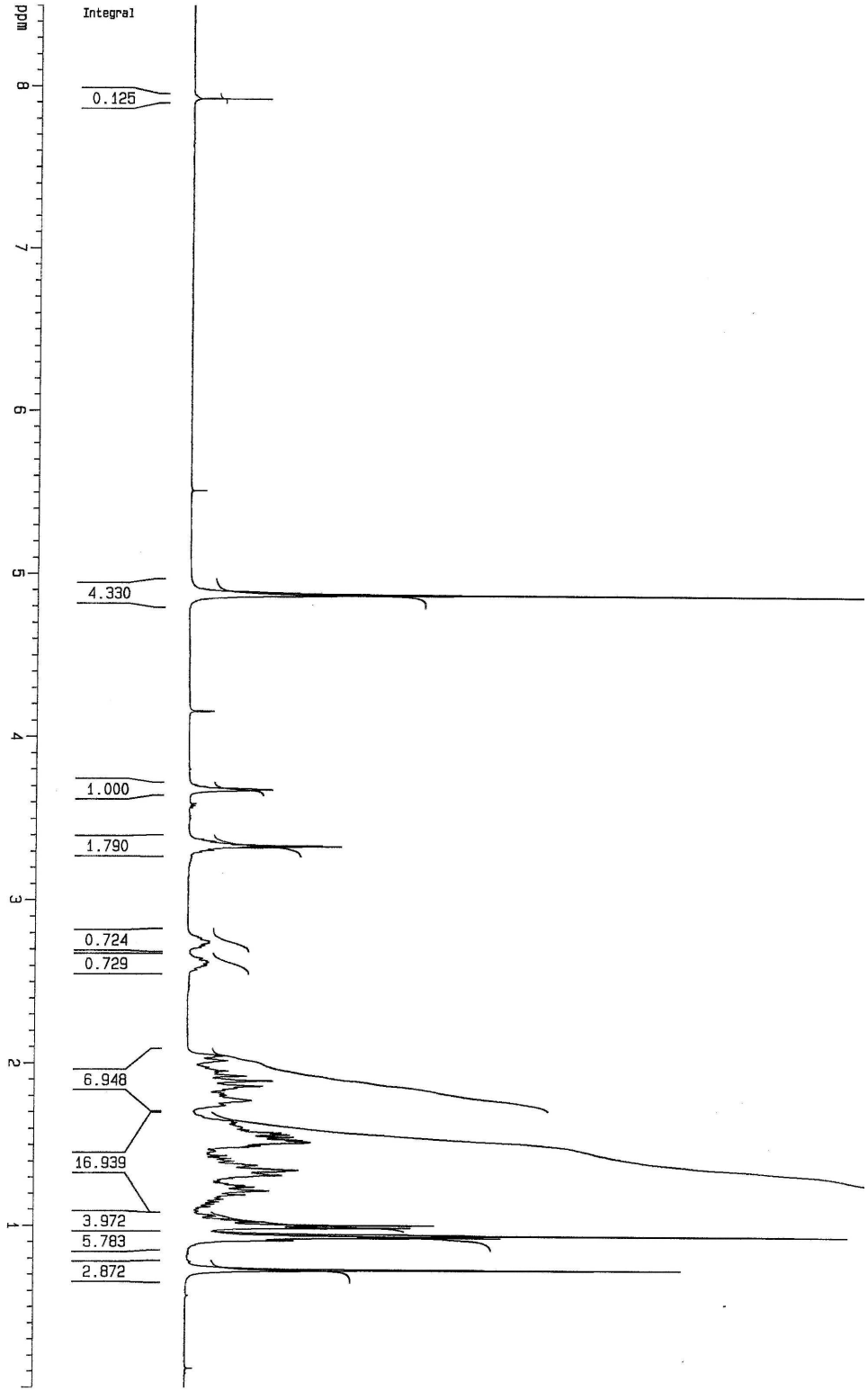

ppm
Integral
0.125
4.330
1.000
1.790
0.724
0.729
6.948
16.939
3.972
5.783
2.872

Figure S5b

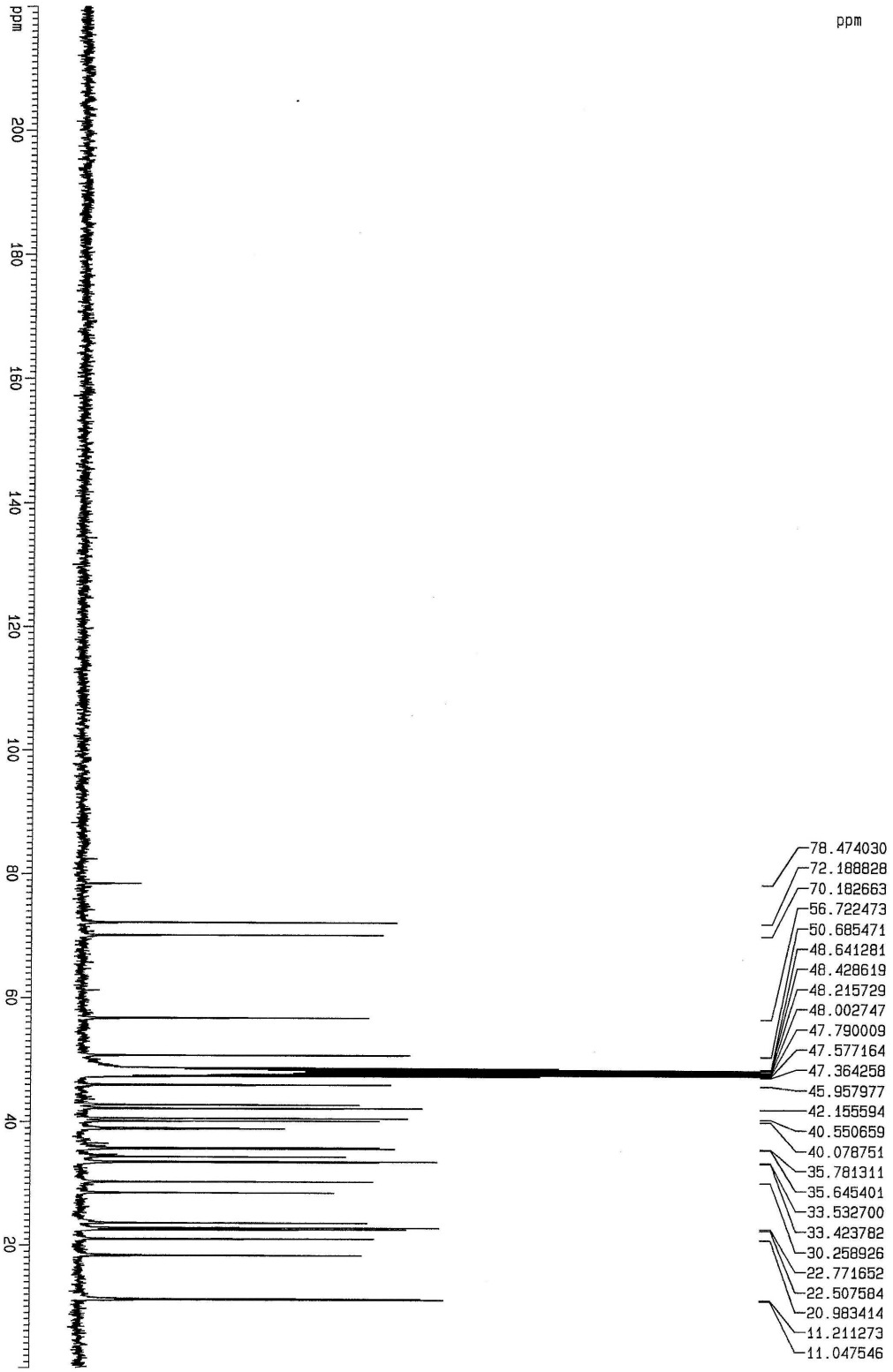

ppm
ppm
200
180
160
140
120
100
80
60
40
20
78.474030
72.188828
70.182663
56.722473
50.685471
48.641281
48.428619
48.215729
48.002747
47.790009
47.577164
47.364258
45.957977
42.155594
40.550659
40.078751
35.781311
35.645401
33.532700
33.423782
30.258926
22.771652
22.507584
20.983414
11.211273
11.047546

Figure S6