Electronic Supplementary Material (ESI) for Green Chemistry.

Expanding the reaction space of aldolases using hydroxypyruvate as nucleophile substrate

Supporting Information

Véronique de Berardinis,[a] Christine Guérard-Hélaine,[b] Ekaterina Darii,[a] Karine Bastard,[a] Virgil Hélaine,[b] Aline Mariage,[a] Jean-Louis Petit,[a] Nicolas Poupard,[b] Israel Sanchez-Moreno,[b] Mark Stam,[a] Thierry Gefflaut,[b] Marcel Salanoubat,[a] and Marielle Lemaire[b]

[a] Commissariat à l'Energie Atomique et aux énergies alternatives, Direction de la Recherche Fondamentale, Institut de Génomique, Genoscope, UMR 8030, 91057 Evry, France ; Université d'Evry Val d'Essonne, UMR 8030, 91057 Evry, France ; Centre National de la Recherche Scientifique, UMR8030, 91057 Evry, France

[b] Clermont Université, Université Blaise Pascal, ICCF, BP 10448, F-63000 Clermont-Ferrand, France ; CNRS, UMR 6296, BP 80026, F-63177 Aubière, France.

1. Materials

Pyruvate (PA), D- and D,L-glyceraldehyde (GA), and 2-ketogluconate (**1**) were purchased from Sigma-Aldrich. D- and L-lactaldehyde[1] and hydroxypyruvate (HPA)[2] were obtained as already published. Isopropyl-β-D-thiogalactoside (IPTG) was purchased from Alexis Biochemicals. Lactate dehydrogenase (LDH) from rabbit muscle , Glygly, HEPES, triethanolamine, sodium phosphate monobasic dihydrate and Trizma hydrochloride were purchased from Sigma-Aldrich. Oligonucleotides were from Sigma-Genosys. *E. coli* BL21Star (DE3) plysE strains for protein expression were from Invitrogen.

Nuclear magnetic resonance (NMR) spectra were measured in deuterated water (D_2O) on a Bruker AC-400 spectrometer, operating at 400 MHz for ^1H and 100 MHz for ^{13}C. Chemical shifts (δ) are reported in ppm relative to TMS signal. Coupling constant values (*J*) are given in Hertz. TLC analyses were performed on silica gel 60 F_{254} plates from Merck using a mixture of vanillin (0.3 M) and H_2SO_4 (0.5 M) in EtOH as developer agent.

Electrospray ionization mass spectra (ESI-MS) were recorded on a micro q-tof Micromass (3000 V) and high resolution mass spectra (HR-MS) were recorded on the same instrument with an internal lock mass (H_3PO_4) and an external lock mass (Leu-enkephalin).

2. Methods

2.1. Method S1. Set up of the candidate collection to be screened.

The collection of aldolases candidate was developed in a three-step process.

(1), a group of proteins was created by selecting in the EC number list (Table S1) known enzymes of interest. The chosen aldolases were those using dihydroxyacetone phosphate (DHAP), dihydroxyacetone (DHA), phosphoenol pyruvate, or PA as donor substrates on the basis of substrate structure similarity. Deoxyribose-5-phosphate aldolase (DERA) described to be flexible towards donors, like DHA-dependent aldolase (FSA) was also included. Then a reference set of experimentally characterized enzymes from this EC number list was selected from published data. This set was compared to UniprotKB or from metagenomics bacterial encoding genes using low stringency parameters (> 30 % identity with the reference set enzymes). This resulted in the selection of 11,717 candidate enzymes. (2), a clustering based on 80 % identity was applied to organize them into putative iso-functional groups. On this basis a set of 1,148 candidate proteins from 313 different prokaryote species was selected. (3) Among the candidate proteins, 731 genes (64 %) from ~230 species were successfully cloned in an expression vector using automated platform facilities (Method S2). Out of these, 571 proteins were significantly overexpressed (79 %) as observed on SDS gel (Method S2). Despite the reduction of the number of enzymes, this protein set still covered most of the selected family's diversity (Figure S1).

The choice of the primers was based on the Primer3 program.[3]. Specific extensions are added to the primers for cloning into our plasmid pET22b(+) (Novagen) modified for ligation independent cloning.[4]

The second small collection of putative HPA aldolases was built as already described above but HPA aldolases from the first screening (Table S2) were used as reference set.

2.2. Method S2. Cloning, expression and cell free extract preparation.

All steps, from primers purchase to cell lysate preparation, were performed in 96 micro-wells plates. The cloning was done into plasmid pET22b(+), the protein production in *E. coli BL21*Star (DE3) strains and cell lysate preparations were carried out as previously described.[5] After centrifugation to clarify the cell extract, the supernatants were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to check for recombinant protein production. Negative control lysate for enzymatic assays was done by transforming *E. coli BL21*Star (DE3) strains by plasmid pET22b(+). Protein concentrations were determined by the Bradford method with bovine serum albumin as the standard (Bio-Rad). The samples were stored at -80°C. Primers used for His-tagged cloned and the 60 aldolases from other Pfam families are described in Table S9. Final hits from the screening were individually sequenced and new genes have been deposited to the European Nucleotide Archive (ENA): aldolase from *Pseudomonas aeruginosa* DSM 50071 was identified as Accession#LN811419 (UniprotKB id attributed: A0A081HJP9), aldolase from *Sinorhizobium meliloti* DSM 30135 was identified as Accession#LN811421 (UniprotKB id attributed: A0A0F4Y2G3), aldolase from *Pseudomonas putida* DSM 291 was identified as Accession#LN811420 (UniprotKB id attributed: A0A0E3VZZ4).

2.3. Method S3. Screening assay and LC/MS analysis.

Biochemical assays were performed in 96-well microplates. Enzyme assays were performed in a final volume of 100 µL containing 1 mM dithiothreitol, 3 µL cell lysate (0.05 to 0.1 mg mL^{-1} of total proteins), 10 mM of each substrate, 50 mM GlyGly buffer, pH 7.5). Assays were performed overnight at room temperature and then reactions were stopped by adding 1 % TFA. After centrifugation, a 1/20 dilution in water was done before LC/MS injection by transferring 10 µL of each well of acidified reaction media in a 96-well daughter microplate. The standards were prepared as described above replacing enzyme cell lysate by *E. coli* BL-21 blank cell lysate.

2.4. Method S4. LC-MS analysis of 2-ketogluconate.

LC/MS/MS analyses were performed on a Hybrid triple quadrupole-linear ion trap mass spectrometer (QTRAP 5500 from ABSciex, Toronto, Canada) equipped with an ESI source and coupled to a Dionex UltiMate 3000 RS LC system (Thermo Scientific Dionex Corporation, Sunnyvale, CA, USA).

The QTRAP mass spectrometer was operated in the ESI negative ion mode with the following parameters: ion source (IS) 4500 V, curtain gas (CUR) 20 a.u., temperature (TEM) 500°C, gas 1 (GS1) 45 a.u., gas 2 (GS2) 60 a.u., CID – medium.

MS/MS experiments were carried out using MRM scan type. The ESI-MS/MS methods were built using two specific transitions for 2-ketogluconate.

The reversed-phase method was used on an ACQUITY UPLC BEN C18 column (1.7 µm, 2.1 x 150 mm) from Waters (Milford, USA). The flow rate was 300 µL/min, the sample injection volume was 10 µL. The column was thermostated at 50°C. The mobile phase A was 10 mM ammonium carbonate in water, and the mobile phase B was acetonitrile. Linear gradient from 0 to 100% of B was applied over a period of 7 min.

The data processing was performed using Analyst software (ABSciex). On each microtiter plate, *E. coli* cell lysate was used as negative control. Enzymes were considered as positive when the signal was at least 3 fold higher than that of the negative control.

MRM conditions for the detection of 2-ketogluconate

	Q1 m/z	Q3 m/z	DP(Volts)	CE (Volts)	CXP (Volts)
2-ketogluconate	192.9	103	-80	-16	-11
	192.9	59	-80	-28	-9

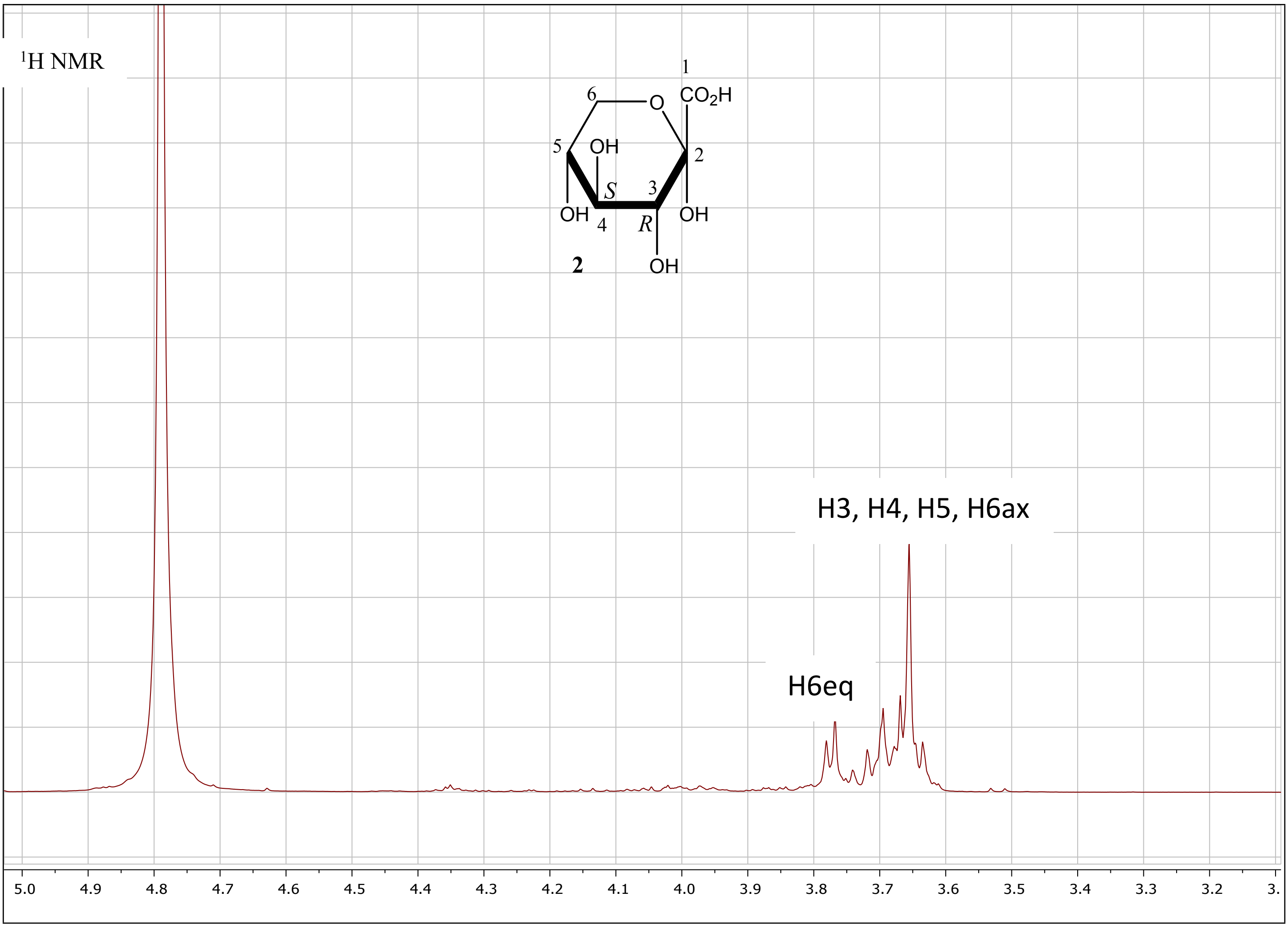
2.5. Method S5. Enzyme purification.

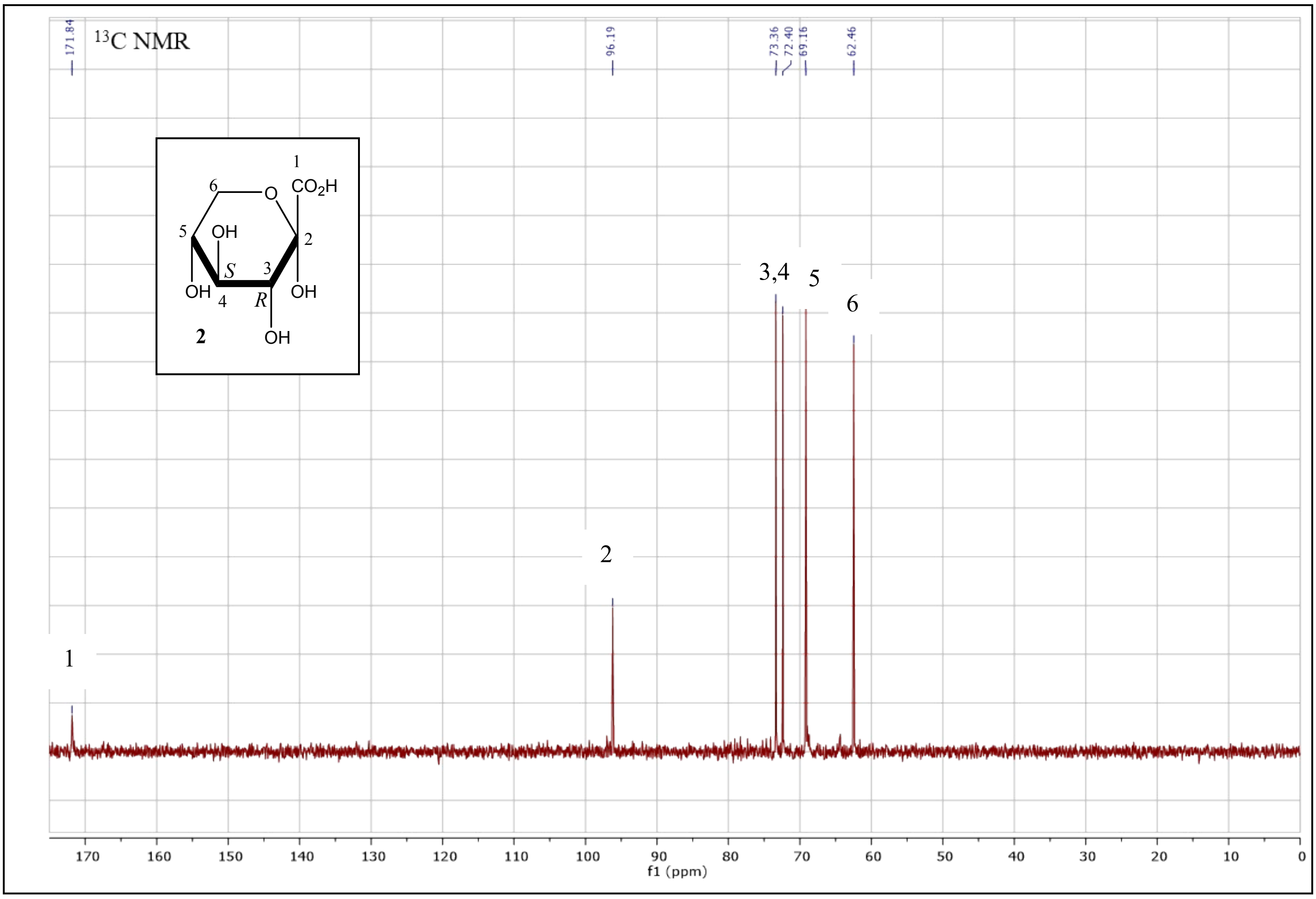
Selected His-tagged enzymes were purified by loading the crude extract onto a Ni-NTA column (QIAGEN), according to the manufacturer’s instructions. The enzyme purity was analysed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS PAGE). Protein concentrations were determined by the Bradford method with bovine serum albumin as the standard (Bio-Rad). Purified enzymes were stored at -80°C.

2.6. Method S6. NMR spectra

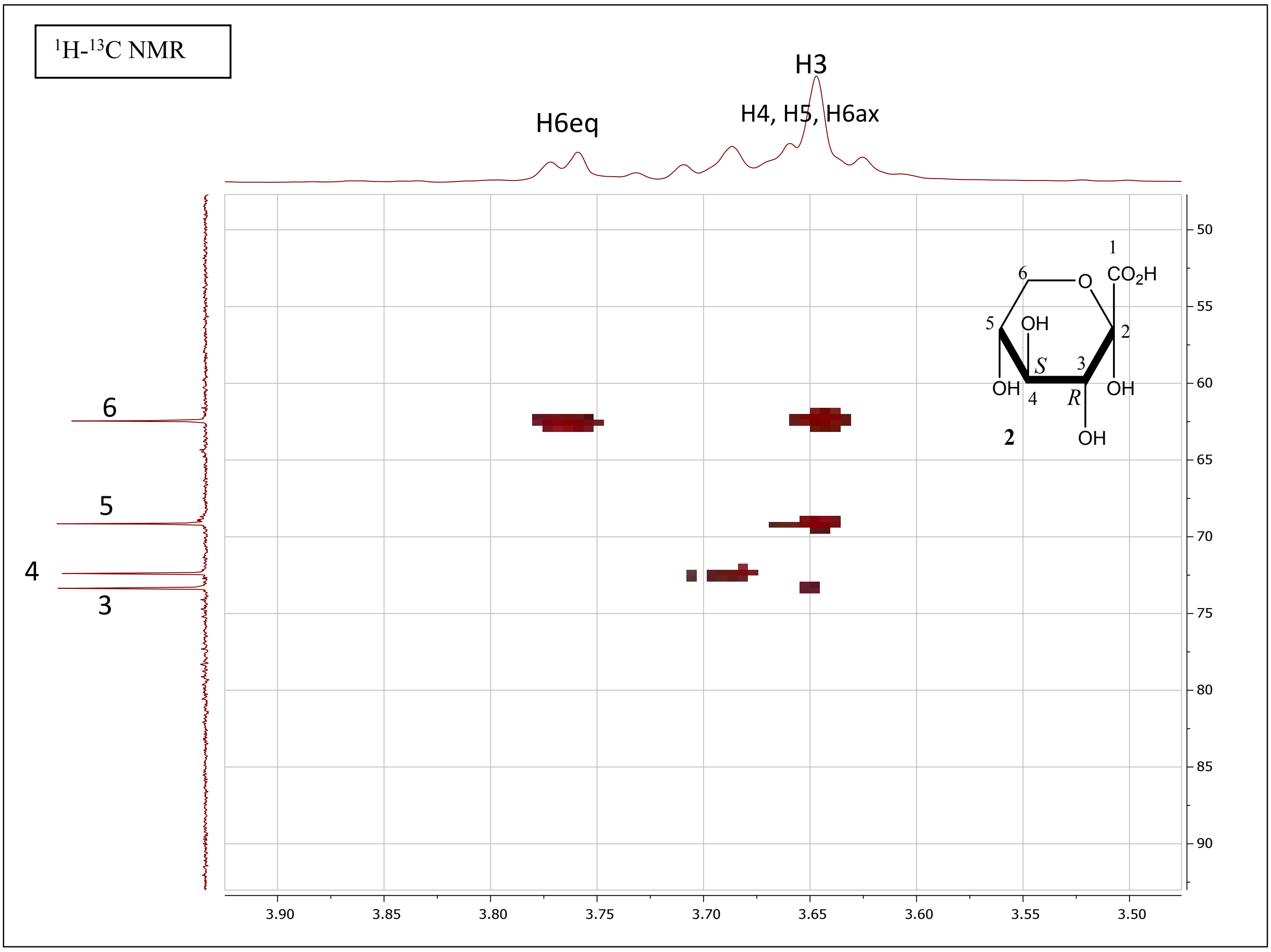
2.5.1 Reactions involving the 22 selected HPA aldolases with HPA or PA as nucleophile substrate and D-GA as electrophile.

- Compound **2**: D-xylo-2-hexulosonic acid (2-ketogalactonic acid).

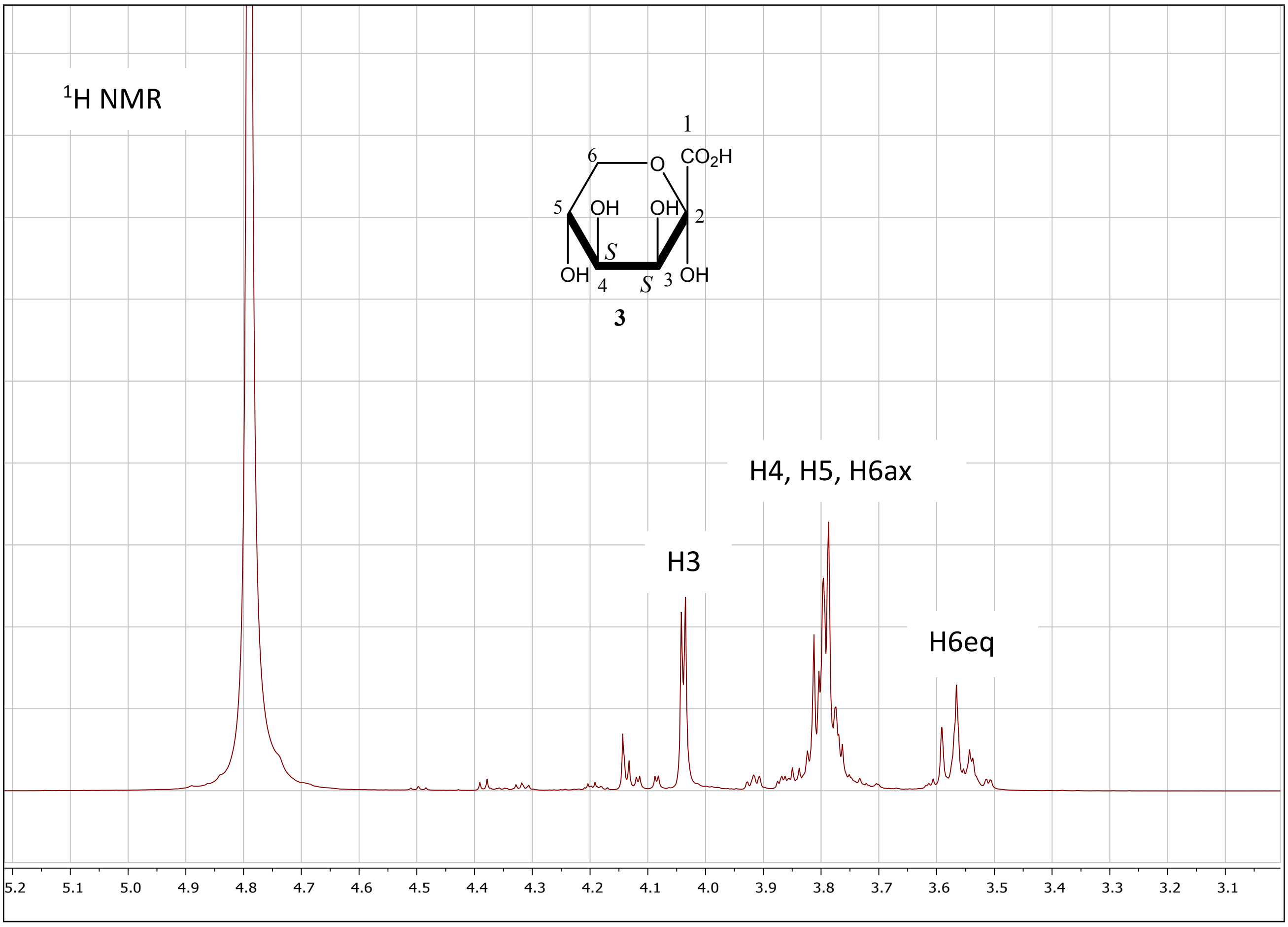


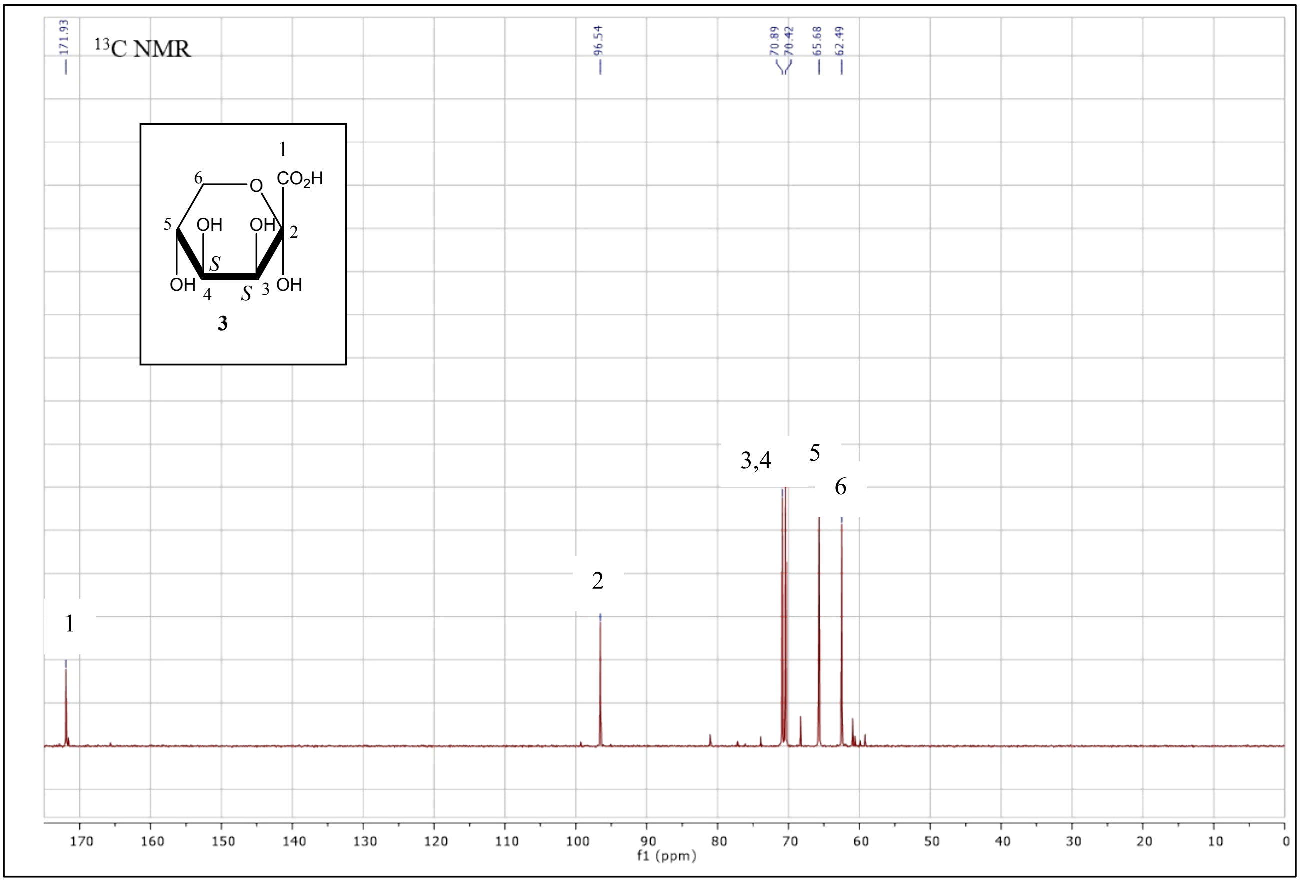


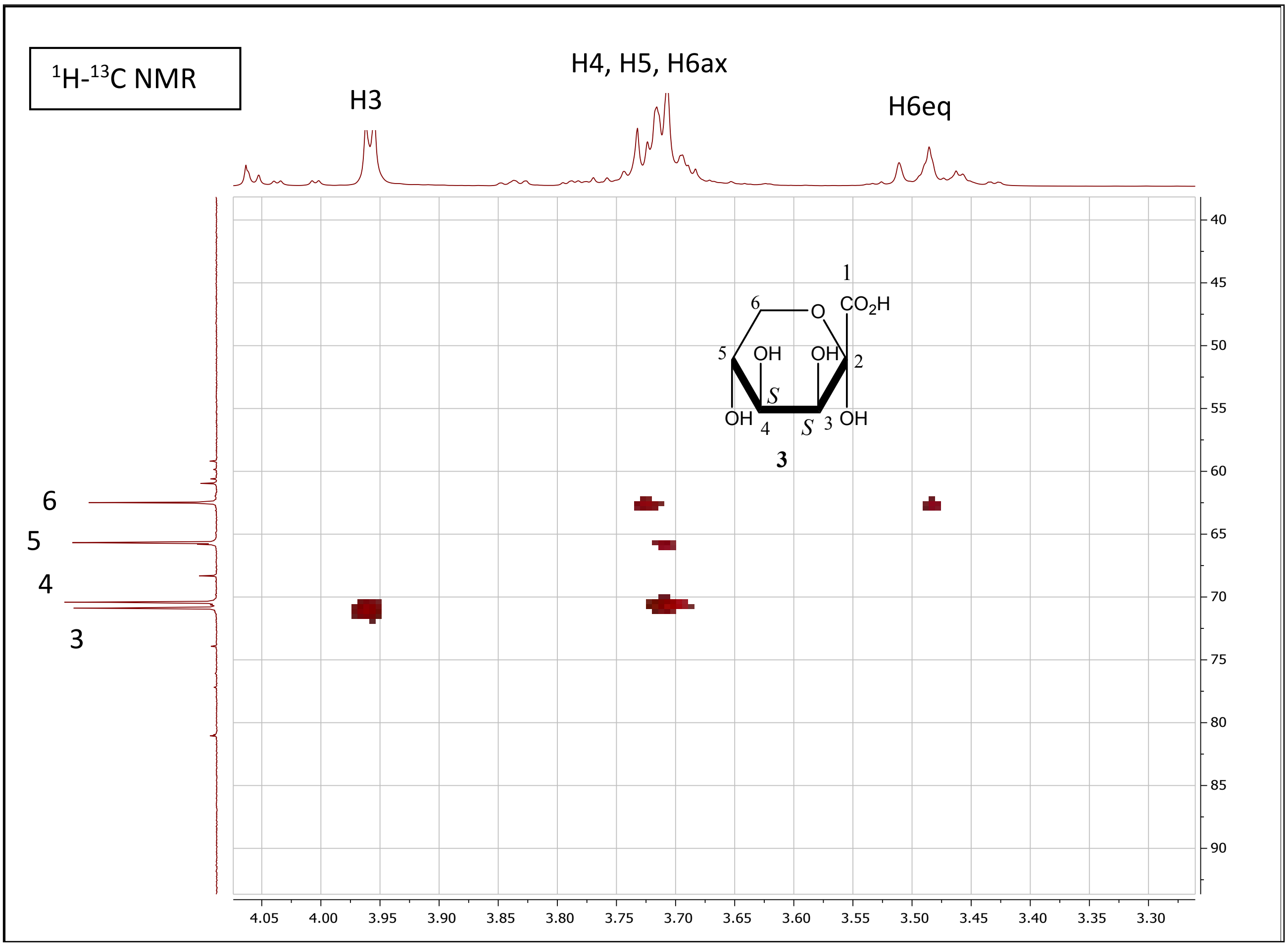
¹³C NMR
171.84
96.19
73.36
72.40
69.16
62.46
1
CO2H
6
O
5
OH
2
S
3
OH
4
R
OH
OH
2
1
2
3,4
5
6
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)



- Compound **3**: D-lyxo-2-hexulosonic acid (2-ketogulonic acid).


^{1}H NMR
1
6
O
CO$_2$H
5
OH
OH
2
S
OH
4
S
3
OH
3
H4, H5, H6ax
H3
H6eq
5.2
5.1
5.0
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1


171.93
96.54
70.89
70.42
65.68
62.49
^{13}C NMR
1
6
O
CO$_2$H
5
OH
OH
2
S
OH
4
S
3
OH
3
3,4
5
6
2
1
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

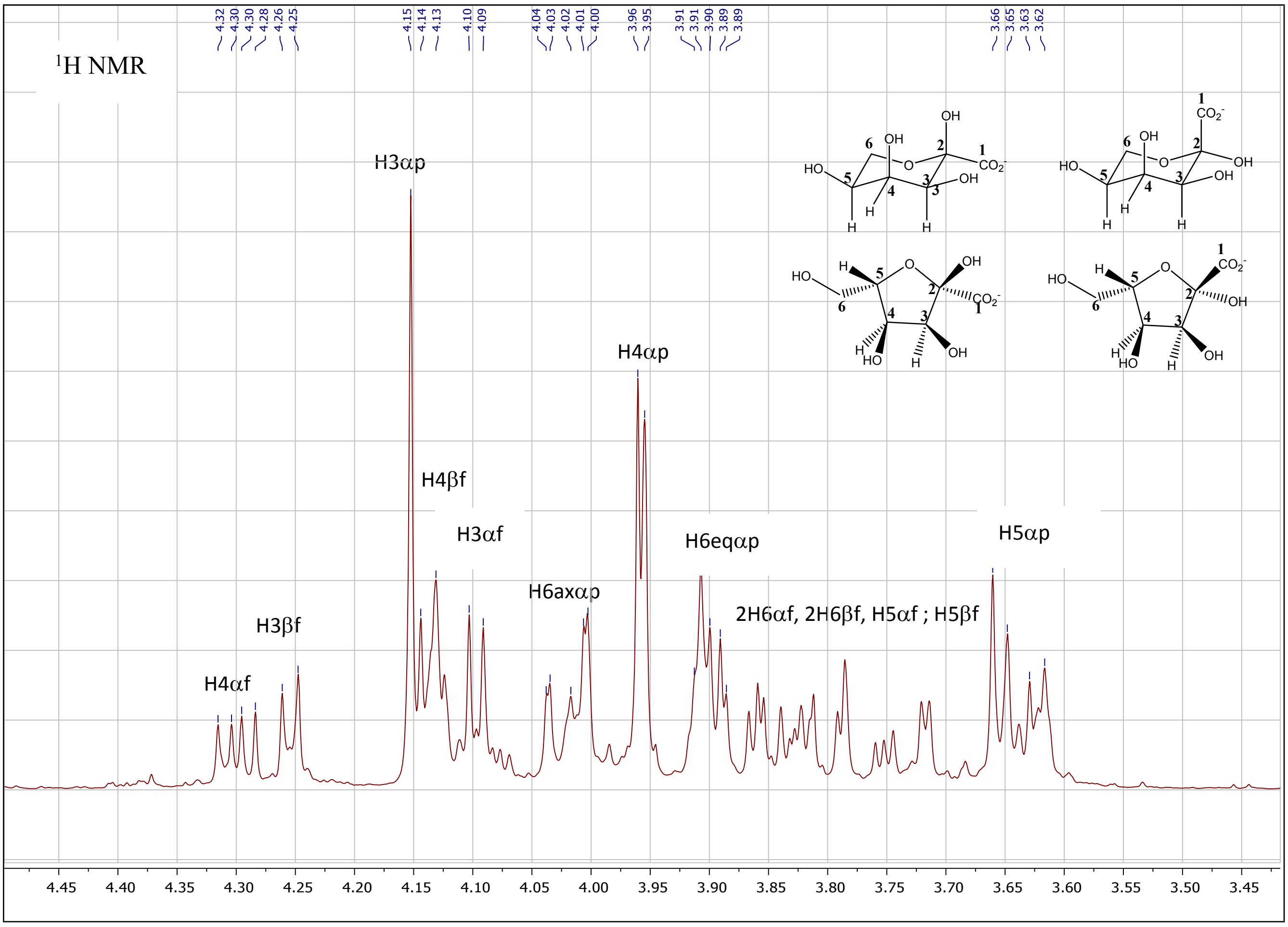


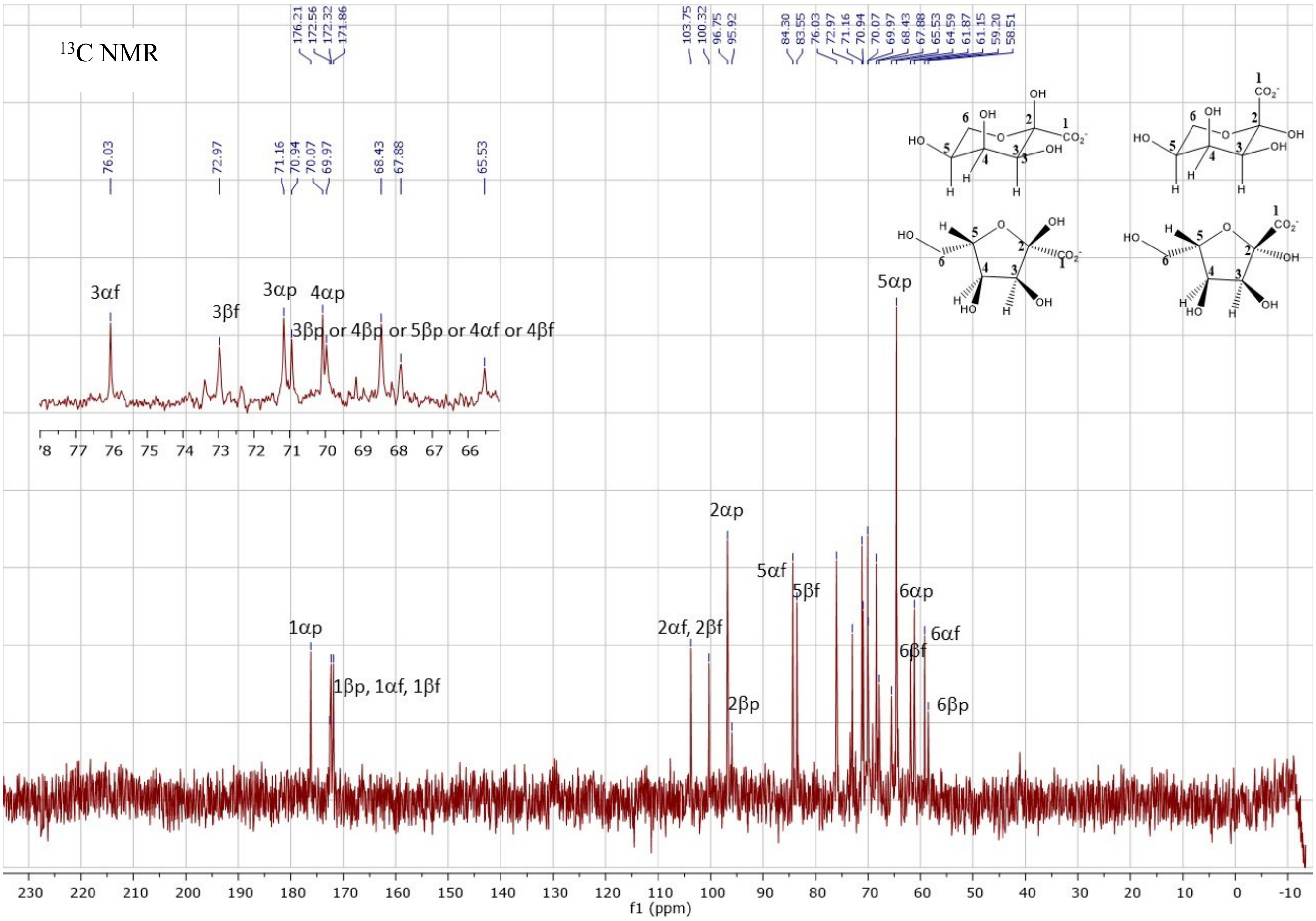
2-keto-3-deoxy-D-gluconic acid **4**: ^{13}C NMR (100 MHz, D$_2$O): δ 172.78 (C1), 95.13 (C2), 68.16 (C4), 64.35 (C5), 64.49 (C6); 33.10 (C3).
2-keto-3-deoxy-D-galactonic acid **5**: ^{13}C NMR (100 MHz, D$_2$O): δ 173.23 (C1), 95.64 (C2); 70.54 (C4), 68.59 (C5); 63.30 (C6); 38.67 (C3).

13C NMR data given for decarboxylated compounds **1** to **5**:
^{13}C NMR (100 MHz, D$_2$O) δ: D-arabinonate 181.05 (C1), 73.75 (C3), 73.40 (C2), 72.72 (C4), 64.60 (C5); D-xylonate 180.00 (C1), 74.24 (C2), 74.22 (C3 or C4), 73.89 (C4 or C3), 63.90 (C5); D-lyxonate: 181.09 (C1), 75.08 (C2), 73.23 (C3 or C4), 72.65 (C4 or C3), 64.30 (C5). 2-deoxy-D-ribonate: 180.24 (C1), 74.38 (C3), 69.71 (C4), 62.42 (C5), 40.57 (C2); 2-deoxy-D-lyxonate: 179.96 (C1), 73.91 (C3), 69.17 (C4), 62.77 (C5), 41.13 (C2).

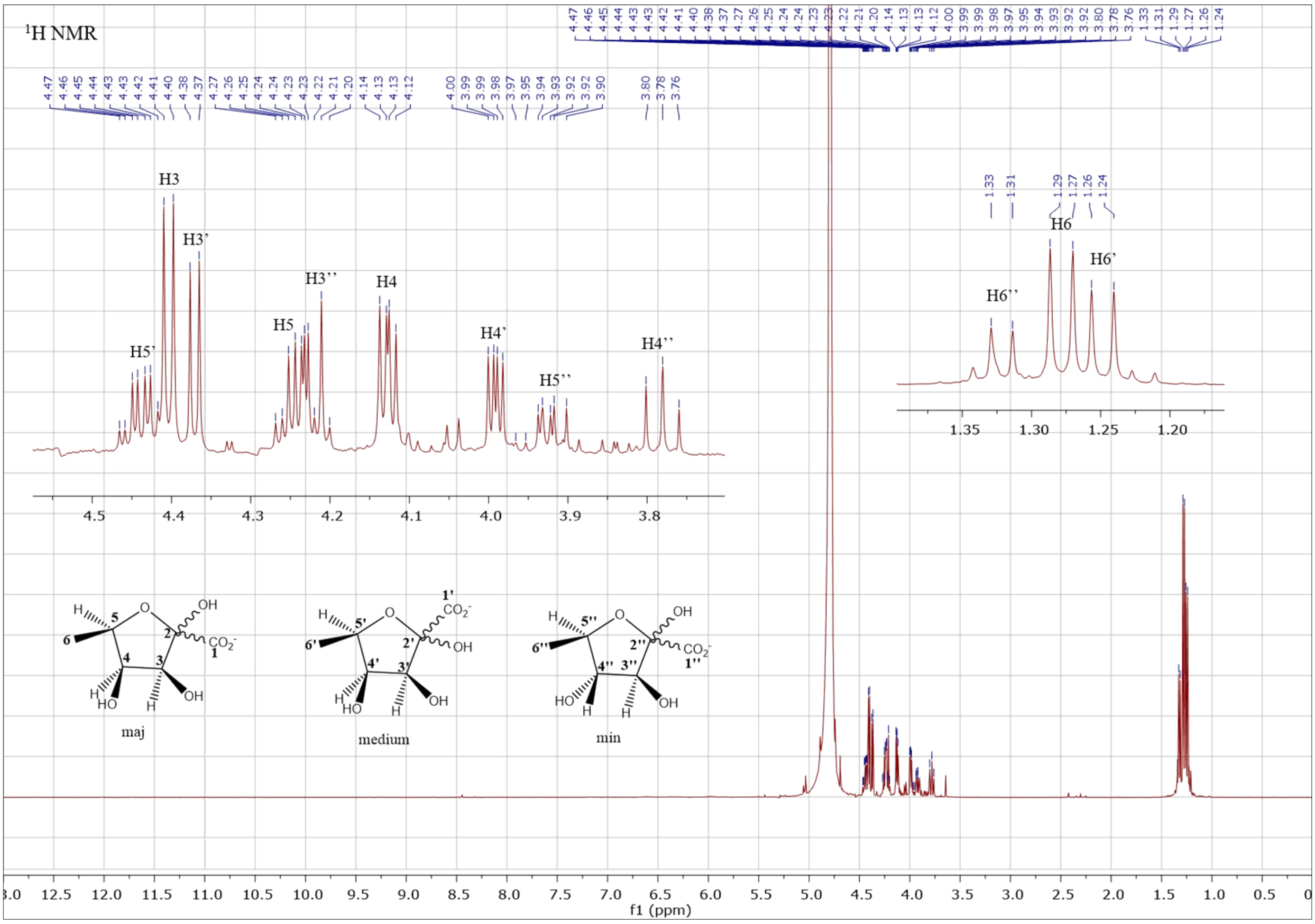
2.5.2 *Reactions involving aldolase A5VH82 with HPA as nucleophile substrate and L-GA, L- or D-lactaldehyde or D-erythrose as electrophile.*

- Compound **6**: L-ribo-2-hexulosonic acid.

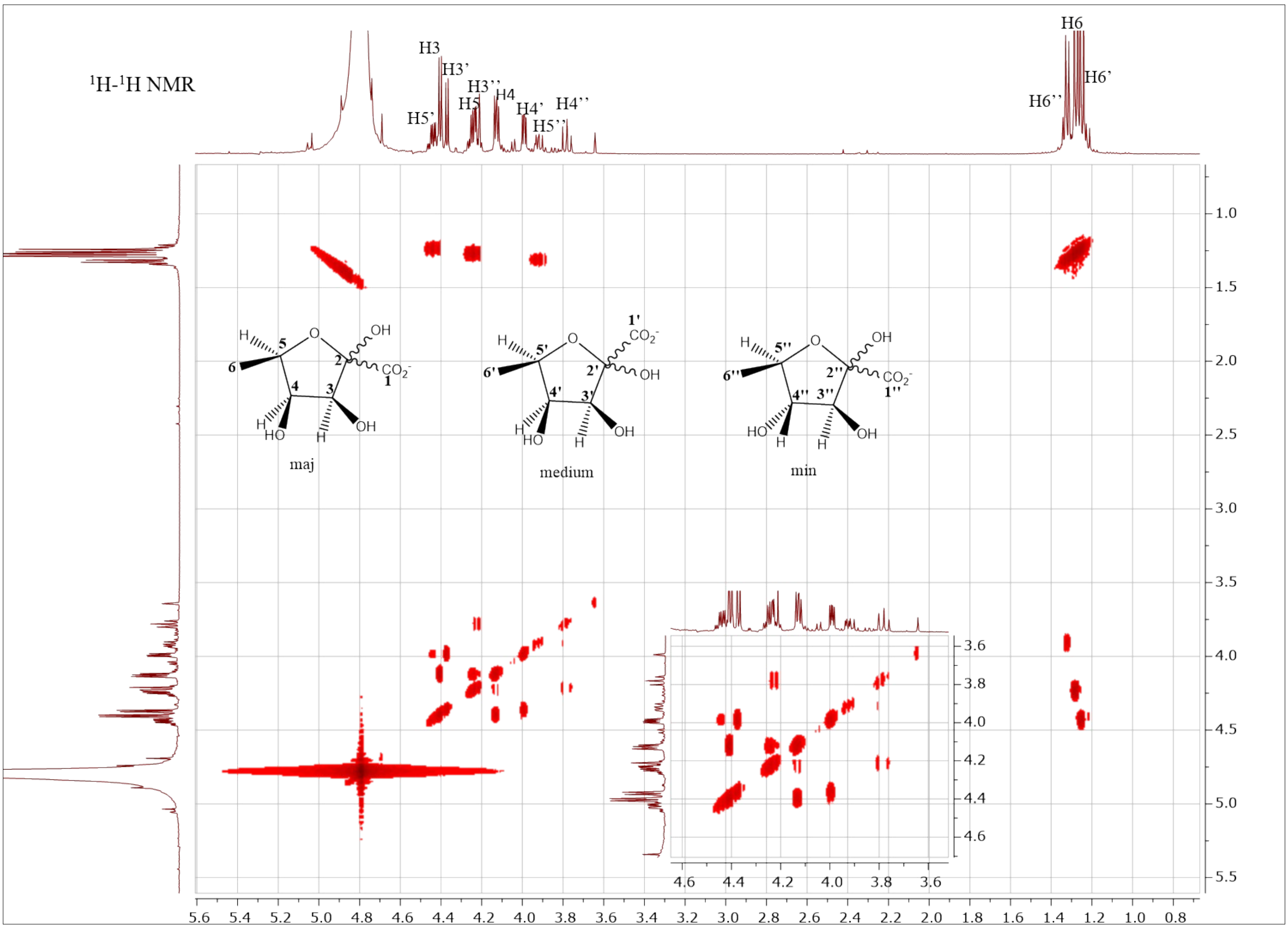




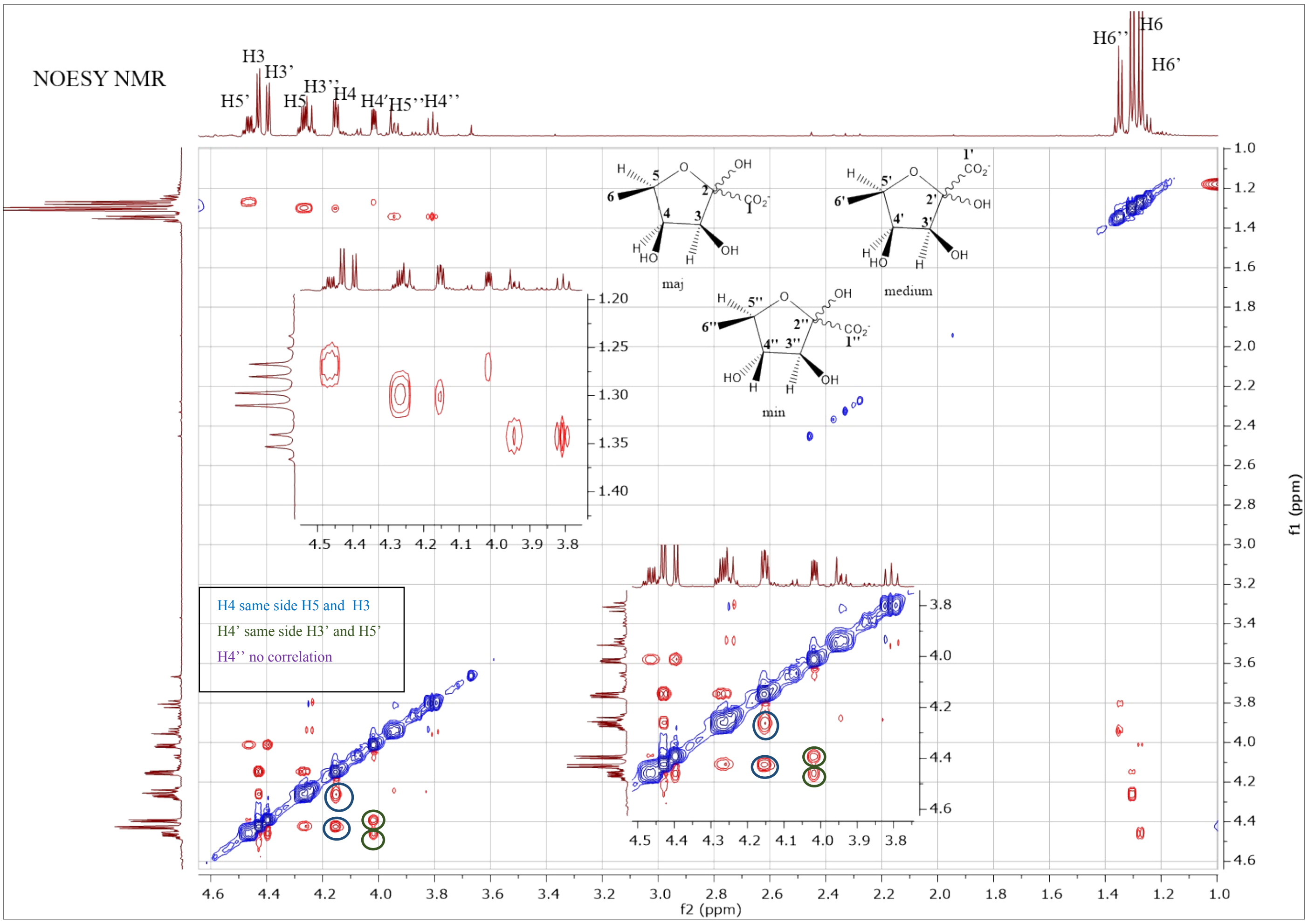
- Compounds **7a** and **7b**: 5-deoxy-D-lyxo-2-hexulosonic acids.

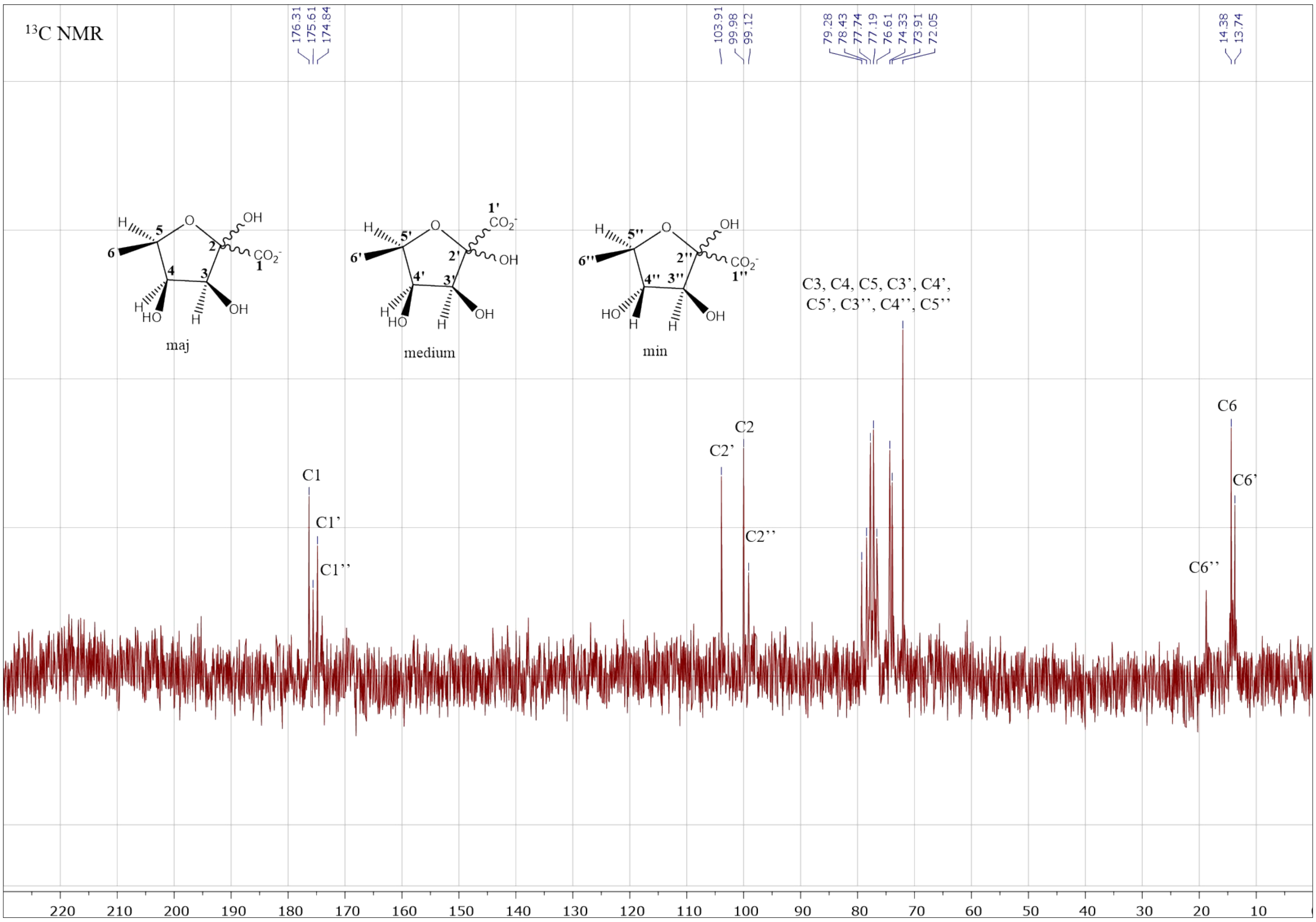


¹H NMR
4.47 4.46 4.45 4.44 4.43 4.43 4.42 4.41 4.40 4.38 4.37 4.27 4.26 4.25 4.24 4.24 4.23 4.23 4.22 4.21 4.20 4.14 4.13 4.13 4.12 4.00 3.99 3.99 3.98 3.97 3.95 3.94 3.93 3.92 3.92 3.90 3.80 3.78 3.76 1.33 1.31 1.29 1.27 1.26 1.24
H3
H3'
H3''
H4
H5
H5'
H4'
H5''
H4''
H6
H6'
H6''
4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8
1.35 1.30 1.25 1.20
maj
medium
min
f1 (ppm)

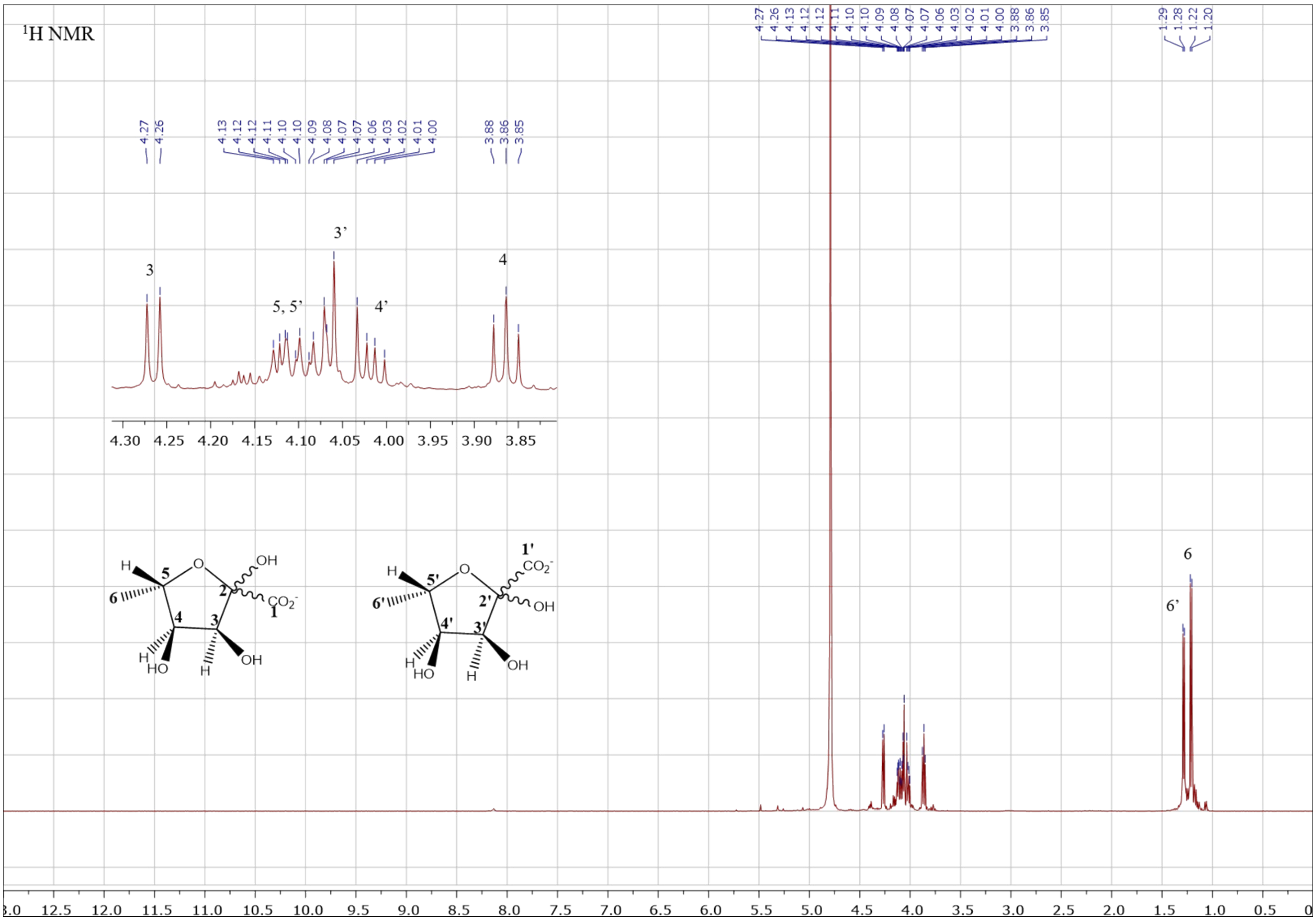


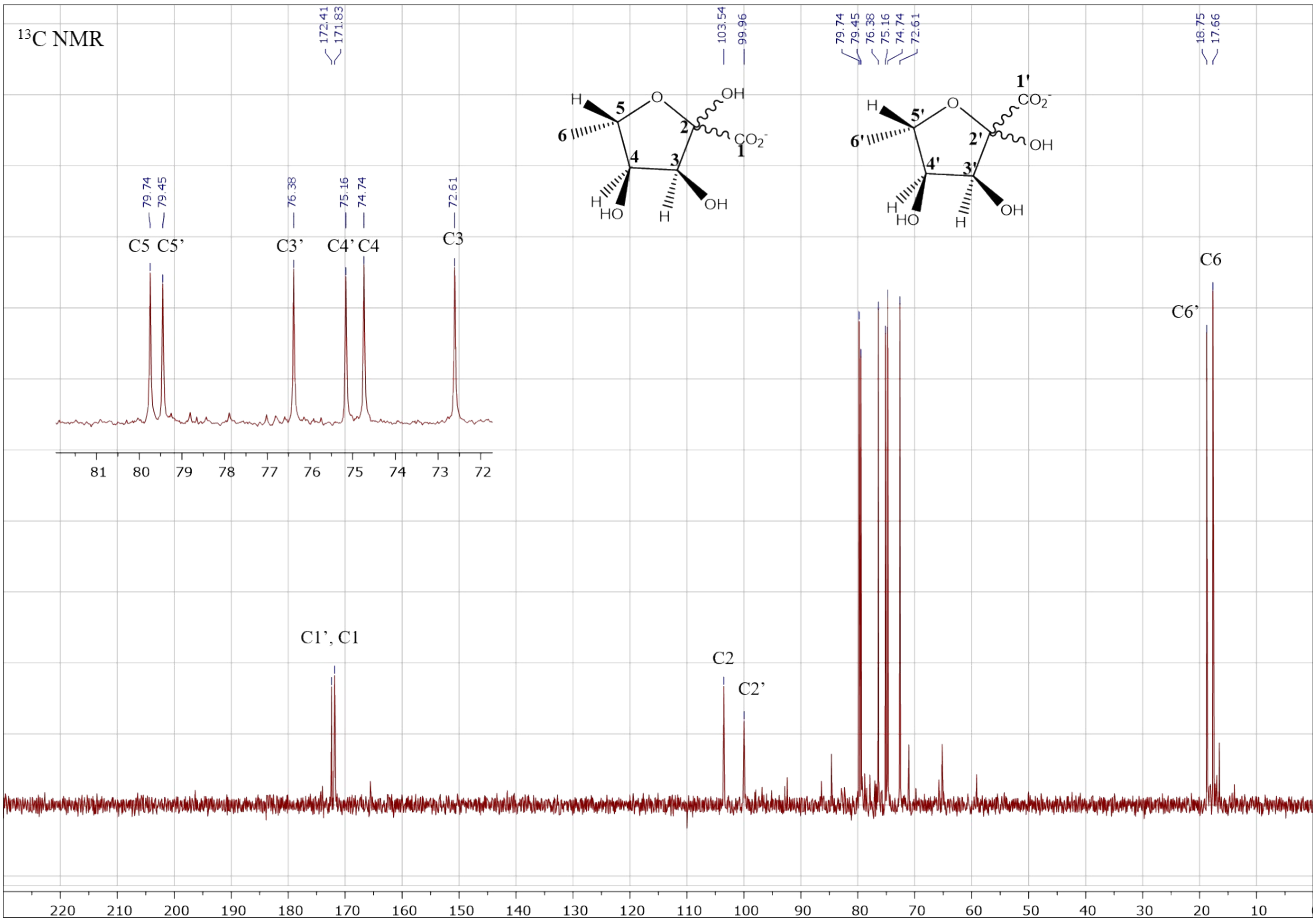
¹H-¹H NMR
H3
H3'
H3''
H5
H4
H5'
H4'
H5''
H4''
H6
H6'
H6''
maj
medium
min


NOESY NMR
H3
H3'
H5'
H5
H3''
H4
H4'
H5''
H4''
H6''
H6
H6'
maj
medium
min
H4 same side H5 and H3
H4' same side H3' and H5'
H4'' no correlation
f1 (ppm)
f2 (ppm)

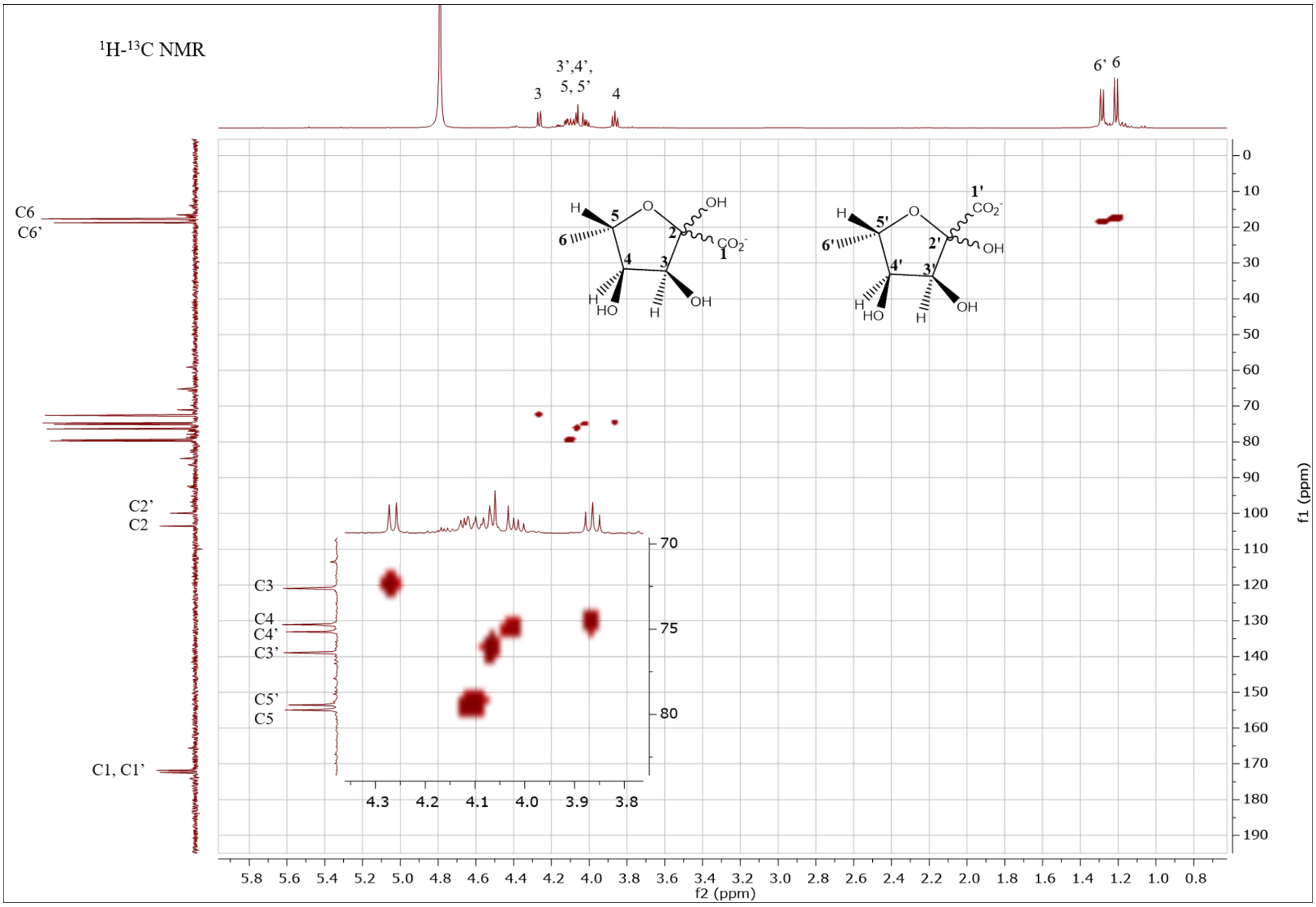

13C NMR
176.31
175.61
174.84
103.91
99.98
99.12
79.28
78.43
77.74
77.19
76.61
74.33
73.91
72.05
14.38
13.74
maj
medium
min
C3, C4, C5, C3', C4', C5', C3'', C4'', C5''
C1
C1'
C1''
C2
C2'
C2''
C6
C6'
C6''

- Compound **8** : 5-deoxy-L-ribo-2-hexulosonic acid.



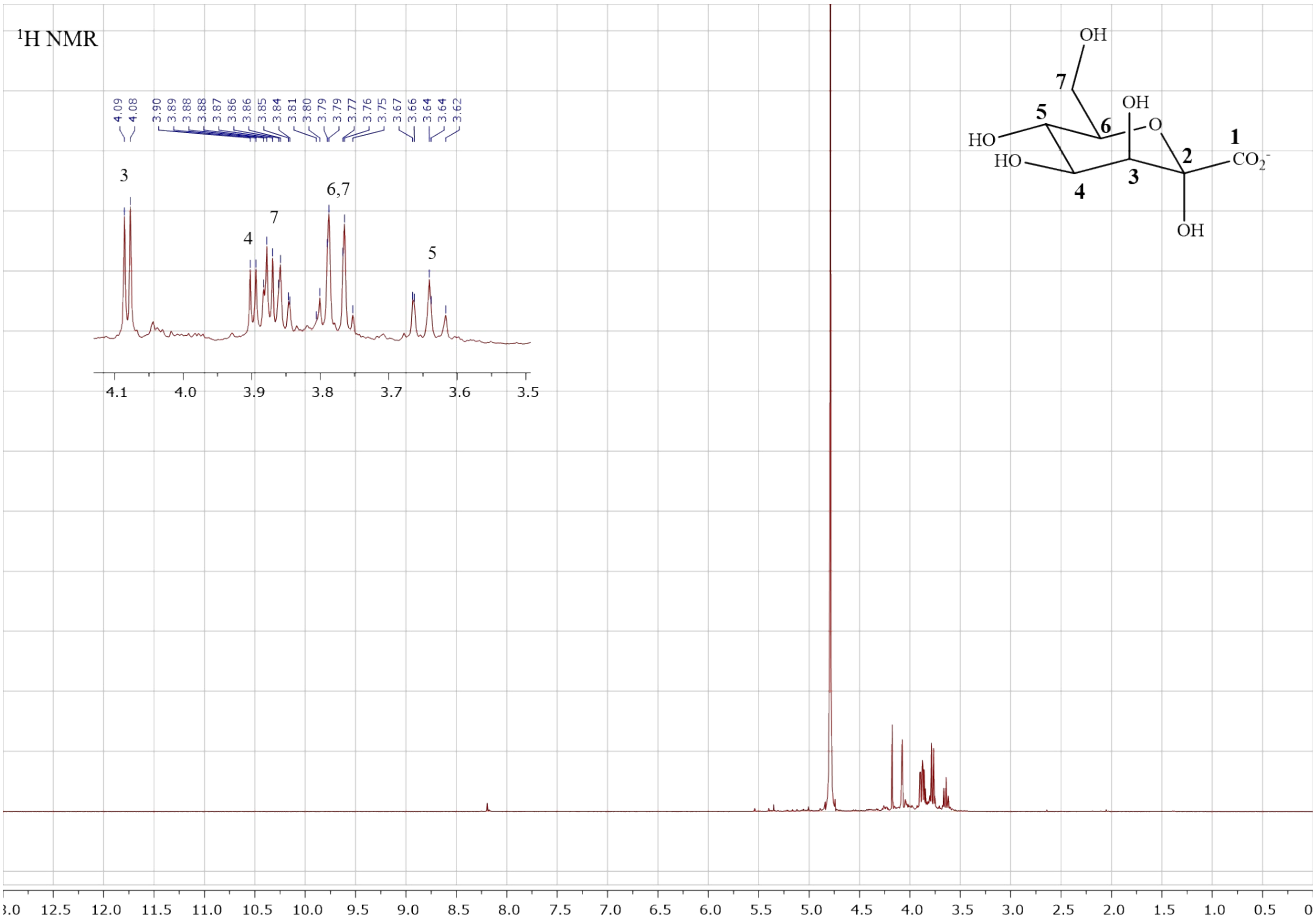


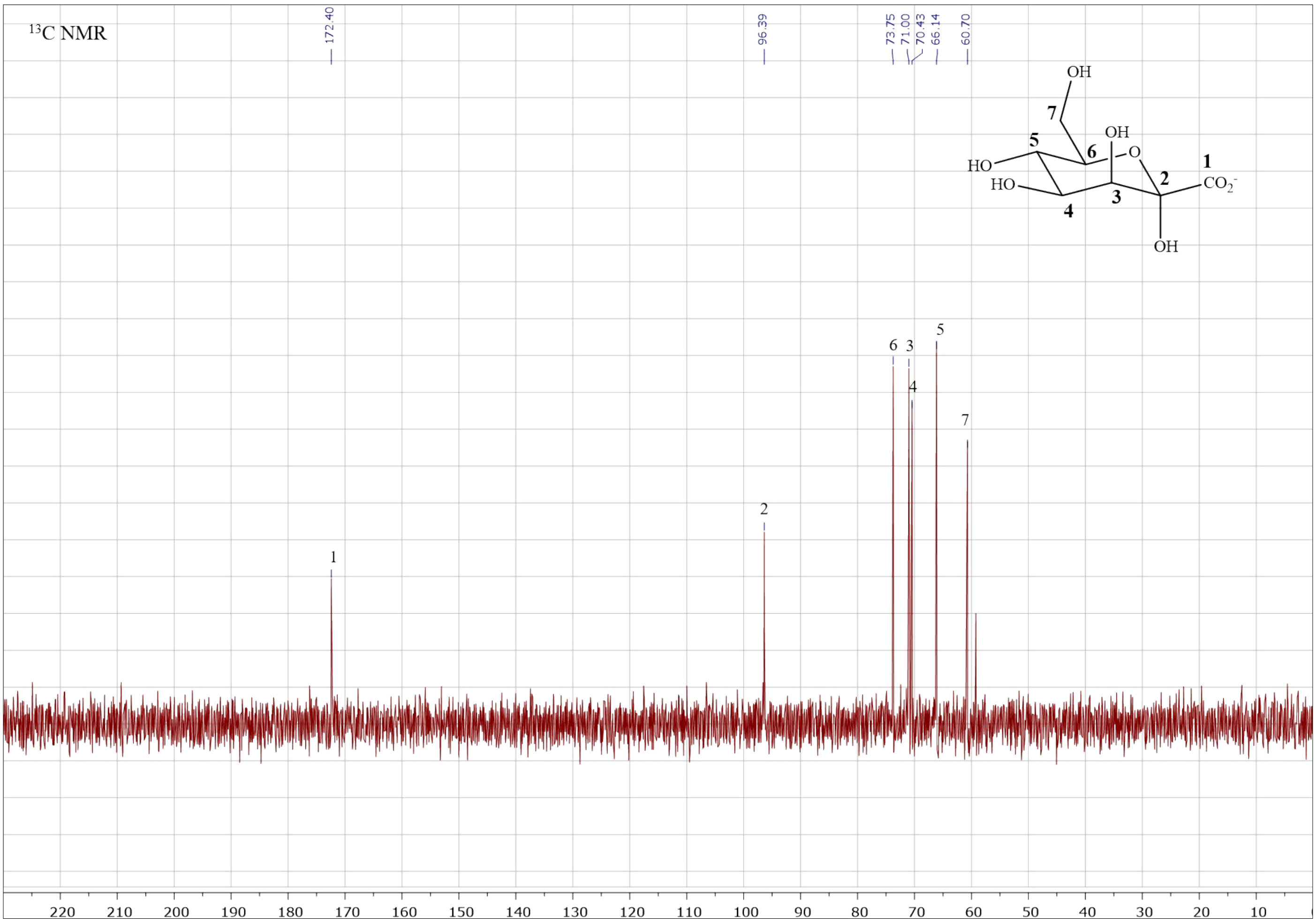
^{13}C NMR
172.41
171.83
103.54
99.96
79.74
79.45
76.38
75.16
74.74
72.61
18.75
17.66
C5 C5'
C3'
C4' C4
C3
C1', C1
C2
C2'
C6
C6'
1'
2'
3'
4'
5'
6'
1
2
3
4
5
6
OH
CO2-
HO
H
O
81 80 79 78 77 76 75 74 73 72
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10

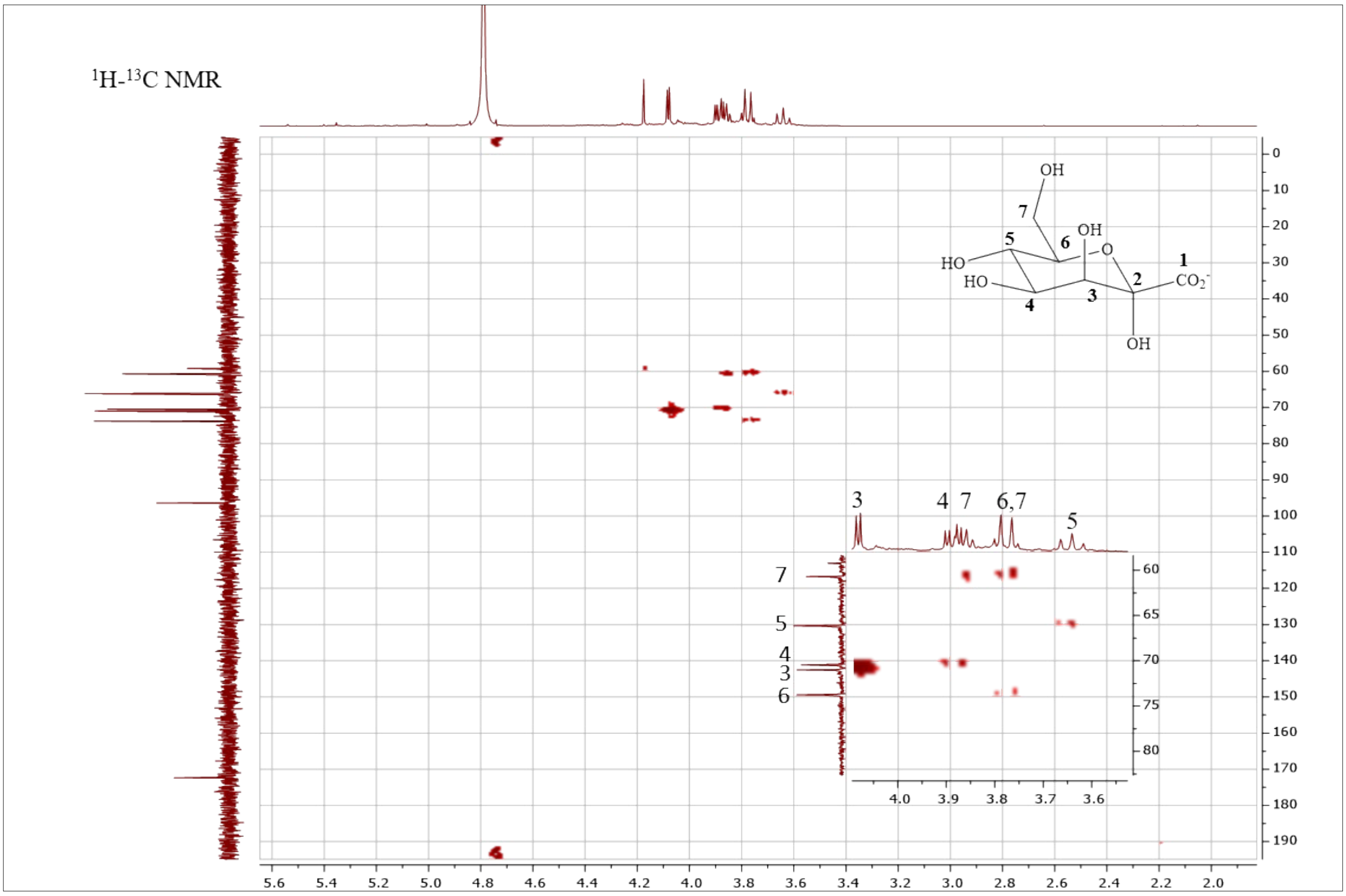


^{1}H-^{13}C NMR
3
3',4',
5, 5'
4
6' 6
C6
C6'
C2'
C2
C1, C1'
C3
C4
C4'
C3'
C5'
C5
1'
CO2-
OH
HO
H
O
70
75
80
4.3 4.2 4.1 4.0 3.9 3.8
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190
f1 (ppm)
5.8 5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8 3.6 3.4 3.2 3.0 2.8 2.6 2.4 2.2 2.0 1.8 1.6 1.4 1.2 1.0 0.8
f2 (ppm)

- Compound **9**: D-manno-2-heptulosonic acid.



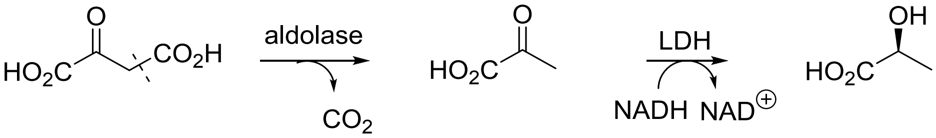

^{13}C NMR
172.40
96.39
73.75
71.00
70.43
66.14
60.70
1
2
6 3
4
5
7
OH
7
5
OH
6
O
1
HO
2
CO$_2^-$
HO
3
4
OH
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10


^{1}H-^{13}C NMR
OH
7
5
OH
6
O
1
HO
2
CO$_2^-$
HO
3
4
OH
3
4 7
6,7
5
7
5
4
3
6
4.0 3.9 3.8 3.7 3.6
60 65 70 75 80
5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8 3.6 3.4 3.2 3.0 2.8 2.6 2.4 2.2 2.0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190

Determination of configurations of **6-9**
Compounds **6** and **9** were decarboxylated following the protocol already described for **1-5** to give known L-ribonate, and D-mannonate respectively.
^{13}C NMR (100 MHz, D_2O) δ: L-ribonate 179.55 (C1), 74.94 (C2), 74.71 (C3), 72.98 (C4), 64.29 (C5); D-mannonate 178.76 (C1), 73.54 (C2), 70.77 (C4), 70.46 (C5), 70.37 (C3), 62.95 (C6).
Compound **8** was also decarboxylated and submitted to a subsequent lactonisation step to give the known 5-deoxy-L-ribono-1,4-lactone. ^{13}C NMR (100 MHz, D_2O): δ 178.78 (C1), 83.73 (C3), 72.83 (C4), 69.01 (C2), 17.40 (C5).

2.7. Method S7. Aldolase activity assay and purification table of the three selected aldolases.

Aldolase activity was spectrophotometrically measured by enzymatic quantification of PA formed during the decarboxylation of oxalacetic acid (see scheme below).[6] This assay was run at room temperature following NADH absorbance at 340 nm for 10 min in the presence of various amounts of assayed aldolase, 0.5 mM oxalacetate, 1 mM Mn^{2+}, 0.5 mM NADH, LDH (50 U) in 50 mM GlyGly buffer, pH 8. Absorbance decrease at 340 nm was proportional to the amount of PA produced in the reaction (ε_{NADH} = 6220 cm^{-1} M^{-1}). One unit of aldolase activity was defined as the amount catalysing the decarboxylation of 1 µmol of oxalacetate per min in these conditions.



	Purification	Total activity (U)	Total protein quantity (mg)	Specific activity (U mg^{-1})
B2T1L6	IMAC	345	30	12
A0A081HJP9	IMAC	1515	15	101
A5VH82	IMAC	150	7.5	20

Purification of recombinant aldolases (from 200 mL of cell culture, retroaldol activity corresponding to the decarboxlation of oxalacetate)

2.8. Method S8. Donor substrates conversion levels with glycolaldehyde as the acceptor.

0.5 mg of aldolase (solution after IMAC purification followed by a desalting column) was added to a solution (1 mL, final volume) containing 1 mM Mn^{2+}, 50 mM HPA or PA and 50 mM glycolaldehyde in 50 mM GlyGly buffer, pH 8. The resulting solution was shaked at room temperature for 1 h. The relative rates of conversion were determined by monitoring the concentration of HPA or PA after 1 h by using lactate dehydrogenase / NADH.

2.9. Method S9. Molecular modeling.

3D models of the 19 candidates and of all the members of the PF03328 family were built by homology modeling using as template the crystal structures of: (i) *E. coli* 2-dehydro-3-deoxy-galactarate aldolase in complex with pyruvate (PDB code: 1DXG), (ii) the citrate lyase beta subunit complex with oxaloacetate (1Z6K), (iii) a putative *E. coli* HpcH aldolase (2VWT) and (iv) *E. coli* HpaI aldolase in complex with pyruvate and succinic semialdehyde (4B5U). Above the 4772 sequences of the family, we have chosen to model only sequence with % of sequence identity with one of the template structure superior to 29 %. We obtained 1481 models. All 3D-models were aligned and the conservation of the active site residues was calculated.
Docking simulations were performed using HPA in *E. coli* HpcH/HpaI active site in presence of Mg^{2+} ion and of succinic semialdehyde using Autodock Vina software. The system was then submitted to Steepest Descent energy minimization using MMTK software and AMBER parameters for residues.

3. Tables

Table S1. EC numbers of aldolases used as Reference Set

EC number	**Gene name / Protein name**	**Annotation**	**Type**
4.1.2.	*fsa*/Fsa	fructose-6-phosphate aldolase (FSA)	DHA aldolases
4.1.2.4	*deoC*/DERA	Deoxyribose-phosphate aldolase (DERA)	others
4.1.2.13	*fba*A/FBPA	Fructose-*bis*phosphate aldolase	DHAP aldolases
4.1.2.14	*eda*/KHG/KDPG, KDGA	2-dehydro-3-deoxy-phosphogluconate aldolase	Pyruvate aldolase or analogs
4.1.2.17	*fuc*A/FucA	L-fuculose-phosphate aldolase	DHAP aldolases
4.1.2.19	*rha*D/RhaD	Rhamnulose-1-phosphate aldolase	DHAP aldolases
4.1.2.20	*gar*L/GarL	2-dehydro-3-deoxyglucarate aldolase	Pyruvate aldolase or analogs
4.1.2.21	*dgo*A/KDPG	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	Pyruvate aldolase or analogs
4.1.2.29	*iol*J/DKGP	5-dehydro-2-deoxyphosphogluconate aldolase	DHAP aldolases
4.1.2.40	*gat*Y,*kba*Y/TBPA	Tagatose-bisphosphate aldolase	DHAP aldolases
4.1.3.3	*nan*A/NanA	N-acetylneuraminate aldolase	Pyruvate aldolases or analogs
4.1.3.16	*eda*/KHG, KDPG, KDGA	4-hydroxy-2-oxoglutarate aldolase	Pyruvate aldolases or analogs
4.1.3.39	*bph*I/HOA, MhpE	4-hydroxy-2-oxovalerate aldolase	Pyruvate aldolase or analogs
2.5.1.54	*aro*G/DAHP	3-deoxy-7-phosphoheptulonate synthase	PEP aldolases or analogs
2.5.1.55	*kds*A/KDO	3-deoxy-8-phosphooctulonate synthase	PEP aldolases or analogs

Table S2. Relative rates for comparison of donor substrates

UniprotKB id	HPA	Pyruvate
B2T1L6	40	100
A0A081HJP9	30	100
A5VH82	100	90

Relative rate of HPA or pyruvate conversion towards glycolaldehyde as acceptor substrate. Calculations were based on enzymatic titration of HPA or PA after 1h of reaction (50 mM HPA or PA, 50 mM glycolaldehyde, 1 mM Mn^{2+}, 0.5 mg protein, 50 mM GlyGly buffer pH 8).

Table S3. Protein pairwise sequence comparisons between hydroxypyruvate aldolases to evaluate their sequence diversity

UniprotKB id	A5VAX1	B4XH86	A5VH82	B2TGJ0	A9CJL2	Q9RYM1	A9E0U5	A7HU24	I7DKY0	Q0K203	Q2K7V6	A3K1H1	A3SLS0	A0A081HJP9	B2T1L6	Q1QUJ4	A9BQY3	B7NJZ1	Q16DB5
A5VAX1		36	32	58	52	64	44	31	46	59	55	47	50	61	44	50	50	49	46
B4XH86			29	39	39	41	35	25	39	37	38	36	36	38	36	39	36	36	35
A5VH82				31	36	36	33	36	32	33	35	33	34	33	32	34	31	32	33
B2TGJ0					50	66	44	32	48	58	52	48	54	59	43	57	53	48	50
A9CJL2						57	39	31	58	58	47	45	57	38	41	45	44	46	54
Q9RYM1							45	32	54	67	59	51	57	72	42	58	57	50	53
A9E0U5								28	41	44	50	44	41	42	36	40	39	44	37
A7HU24									32	27	32	33	30	30	33	33	30	31	33
I7DKY0										50	45	59	59	54	42	49	49	46	58
Q0K203											54	48	52	62	40	52	50	46	49
Q2K7V6												56	44	59	38	49	50	45	45
A3K1H1													45	50	40	47	47	41	43
A3SLS0														54	44	50	53	49	55
A0A081HJP9															40	56	58	47	54
B2T1L6																42	41	41	43
Q1QUJ4																	69	50	44
A9BQY3																		45	46
B7NJZ1																			43
Q16DB5																			

Table S4. Gene cluster organizations of HPA/PA aldolases compare to the reference *E. coli* enzymes HpcH, GarL or RhmA gene neighborhoods.

UniprotKB id	Annotation	% of identity with			Genomic context
		HpcH (B1IS70)	GarL (P23522)	RhmA (P76469)	
B2TGJ0	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	60	32	55	HpcH genomic context
A0A081HJP9	putative uncharacterized protein	65	47	58	HpcH genomic context
A5VAX1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	61	49	50	no known genomic context
B4XH86	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	44	36	39	no known genomic context
A5VH82	hpch/hpai aldolase family protein	36	32	34	no known genomic context
A9CJL2	putative aldolase	58	46	51	no known genomic context
Q9RYM1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	67	47	59	no known genomic context
A9E0U5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	48	44	41	no known genomic context
A7HU24	hpch/hpai aldolase family protein	31	31	33	no known genomic context
I7DKY0	hpch/hpai aldolase family protein	52	46	51	no known genomic context
Q0K203	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	60	46	59	no known genomic context
Q2K7V6	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	61	45	50	no known genomic context
A3K1H1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	51	41	48	no known genomic context
A3SLS0	hpch/hpai aldolase family protein	52	49	49	no known genomic context
B7NJZ1	2-dehydro-3-deoxyglucarate aldolase	48	99	44	GarL genomic context
B2T1L6	2-dehydro-3-deoxyglucarate aldolase	42	41	39	no known genomic context
Q1QUJ4	2-dehydro-3-deoxyglucarate aldolase	52	49	52	no known genomic context
A9BQY3	2-dehydro-3-deoxyglucarate aldolase	54	45	50	RhmA genomic context

Q16DB5	4-hydroxy-2-oxovalerate aldolase	48	43	51	no known genomic context

Table S5. Second set of HPA aldolases from PF03328 assayed with HPA as donor and GA as acceptor by LC/MS analysis. The ratio between the product signal and the negative control (*E. coli* lysate) is reported in the last column

UniprotKB id	**Protein annotation**	**Organism**	**HPA + GA**
B6XAH9	putative uncharacterized protein	*Providencia alcalifaciens*	8
Q1R1H4	HpcH/HpaI aldolase	*Chromohalobacter salexigens*	3
E0E947	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Actinobacillus pleuropneumoniae*	3
F4KUN3	4-hydroxy-2-oxovalerate aldolase	*Haliscomenobacter hydrossis*	4
A9IPR8	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Bordetella petrii*	4
B1G064	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Burkholderia graminis*	6
B1G3M5	2-dehydro-3-deoxyglucarate aldolase	*Burkholderia graminis*	3
B1K8T9	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Burkholderia cenocepacia*	4
B7WZH5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Comamonas testosteroni*	5
B9JRF3	aldolase	*Agrobacterium vitis*)	6
B9Z3Z6	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Pseudogulbenkiania ferrooxidans*	11
C6BIY7	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Ralstonia pickettii*	3
D4DXK2	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Serratia odorifera*	3
A0A0E3VZZ4	hpai	*Pseudomonas putida*	7
A0A0F4Y2G3	4-hydroxy-2-oxovalerate aldolase	*Sinorhizobium meliloti*	3
A5VED8	2-dehydro-3-deoxyglucarate aldolase	*Sphingomonas wittichii*	4
A9BTL5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Delftia acidovorans*	3
Q9HWQ3	putative uncharacterized protein	*Pseudomonas aeruginosa*	4

(HPA): hydroxypyruvate; (GA): glyceraldehyde

Table S6. Set of aldolases representatives of other pyruvate aldolase Pfam families (PF01081, PF00682 and PF00701) assayed with HPA and PA as donor and glyceraldehyde as acceptor. The ratio between the product signal and the negative control (*E. coli* lysate) are reported in PA and HPA columns. *E.coli* enzymes from PF03328 were used as positive controls

Uniprot ID	**Annotation**	**Organism**	**PA**	**HPA**	**PFAM family**
P76469	**2-keto-3-deoxy-l-rhamnonate aldolase (RhmA)**	**Escherichia *coli K12***	**72**	**36**	**PF03325**
P23522	**5-keto-4-deoxy-d-glucarate aldolase (GarL)**	***Escherichia coli K12***	**75**	**26**	**PF03328**
B1IS70	**4-hydroxy-2-oxo-heptane-1,7-dioate aldolase (HpcH)**	***Escherichia coli C***	**126**	**25**	**PF03328**
B8G187	4-hydroxy-2-oxovalerate aldolase	*Desulfitobacterium hafniense*	1	1	PF00682
A0A0S3K3Y1	4-hydroxy-2-oxovalerate aldolase	*Vibrio alginolyticus*	18	1	PF00682
A1HNW3	4-hydroxy-2-oxovalerate aldolase	*Thermosinus carboxydivorans*	1	1	PF00682
B1KJK9	4-hydroxy-2-oxovalerate aldolase	*Shewanella woodyi*	1	1	PF00682
D5DK17	4-hydroxy-2-oxovalerate aldolase	*Bacillus megaterium*	1	1	PF00682
D9S0Z5	4-hydroxy-2-oxovalerate aldolase	*Thermosediminibacter oceani*	1	1	PF00682
E1R9T2	4-hydroxy-2-oxovalerate aldolase	*Spirochaeta smaragdinae*	1	1	PF00682
F6CPD7	4-hydroxy-2-oxovalerate aldolase	*Desulfotomaculum kuznetsovii*	11	1	PF00682
F7NHX5	4-hydroxy-2-oxovalerate aldolase	*Acetonema longum*	1	1	PF00682
A0A0E2GPU2	4-hydroxy-2-oxovalerate aldolase	*Acinetobacter radioresistens*	1	1	PF00682
C8WX09	4-hydroxy-2-oxovalerate aldolase	*Alicyclobacillus acidocaldarius subsp. acidocaldarius*	1	1	PF00682
D5WW76	4-hydroxy-2-oxovalerate aldolase	*Kyrpidia tusciae*	1	1	PF00682
Q1LNS4	4-hydroxy-2-oxovalerate aldolase 1	*Ralstonia metallidurans*	1	1	PF00682
D3FA61	4-hydroxy-2-oxovalerate aldolase	*Conexibacter woesei*	1	1	PF00682
P0A6L4	**n-acetylneuraminate lyase (NanA)**	***Escherichia coli K12***	**2**	**1**	**PF00701**
Q3ID27	n-acetylneuraminate lyase (aldolase)	*Pseudoalteromonas haloplanktis*	**23**	**1**	**PF00701**
A0A0H2UR28	n-acetylneuraminate lyase	*Streptococcus pneumoniae*	2	1	PF00701
E7S791	n-acetylneuraminate lyase, putative	*Streptococcus agalactiae*	1	1	PF00701
Q5LEN8	putative n-acetylneuraminate lyase	*Bacteroides fragilis*	2	1	PF00701
Q64P99	putative n-acetylneuraminate lyase	*Bacteroides fragilis*	10	1	PF00701
Q6F6U2	dihydrodipicolinate synthase	*Acinetobacter sp*	1	1	PF00701
A0A0K6LIU7	dihydrodipicolinate synthase	*Bacillus cereus*	1	1	PF00701
Q64TM6	dihydrodipicolinate synthase	*Bacteroides fragilis*	1	1	PF00701
A5EKM8	dihydrodipicolinate synthase	*Bradyrhizobium sp*	1	1	PF00701
A4XJU0	dihydrodipicolinate synthase	*Caldicellulosiruptor saccharolyticus*	0	1	PF00701
Q74GT6	dihydrodipicolinate synthase	*Geobacter sulfurreducens*	1	1	PF00701
A5ULF8	dihydrodipicolinate synthase	*Methanobrevibacter smithii*	1	1	PF00701
A0A0R0LVT3	dihydrodipicolinate synthase	*Neisseria gonorrhoeae*	2	1	PF00701
Q3SRA2	dihydrodipicolinate synthase	*Nitrobacter winogradskyi*	1	1	PF00701
A9DWX7	dihydrodipicolinate synthase	*Oceanibulbus indolifex*	2	1	PF00701

B2U6L4	dihydrodipicolinate synthase	*Ralstonia pickettii*	11	1	PF00701
A4WNS1	dihydrodipicolinate synthase	*Rhodobacter sphaeroides*	1	1	PF00701
Q2RP09	dihydrodipicolinate synthase	*Rhodospirillum rubrum*	1	1	PF00701
Q04796	dihydrodipicolinate synthase	*Bacillus subtilis*	2	1	PF00701
Q7USB5	dihydrodipicolinate synthase	*Rhodopirellula baltica*	11	1	PF00701
P0A955	**khg/kdpg aldolase**	***Escherichia coli (strain K12)***	**43**	**1**	**PF01081**
A0KF01	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Aeromonas hydrophila*	61	1	PF01081
A0QYU3	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Mycobacterium smegmatis*	5	1	PF01081
A1HMS5	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Thermosinus carboxydivorans*	14	1	PF01081
A0A098Z728	keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	*Streptococcus pneumoniae*	49	1	PF01081
A6B1X9	khg/kdpg aldolase	*Vibrio parahaemolyticus*	56	1	PF01081
I7EZU8	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Phaeobacter inhibens*	46	1	PF01081
B2UCA9	kdpg and khg aldolase	*Ralstonia pickettii*	23	1	PF01081
B2UI88	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Ralstonia pickettii*	19	1	PF01081
B7LPJ4	multifunctional 2-keto-3-deoxygluconate 6-phosphate aldolase and 2-keto-4-hydroxyglutarate aldolase and oxaloacetate decarboxylase	*Escherichia fergusonii*	55	1	PF01081
A0A0C5CKJ5	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Bacillus coagulans*	30	1	PF01081
C2JUN9	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Lactobacillus rhamnosus*	7	1	PF01081
P0A955	**khg/kdpg aldolase**	***Escherichia coli (strain K12)***	**43**	**1**	**PF01081**
Q1J2P1	2-keto-3-deoxy-6-phosphogluconate aldolase	*Deinococcus geothermalis*	8	1	PF01081
Q2G569	2-keto-3-deoxy-phosphogluconate aldolase	*Novosphingobium aromaticivorans*	25	1	PF01081
A0A0H1EM44	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Streptococcus agalactiae*	33	1	PF01081
Q3IH59	putative 4-hydroxy-2-oxoglutarate aldolase / 2-dehydro-3-deoxyphosphogluconate aldolase	*Pseudoalteromonas haloplanktis*	6	1	PF01081
Q3IL59	multifunctional: 2-keto-3-deoxygluconate 6-phosphate aldolase; 2-keto-4-hydroxyglutarate aldolase; oxaloacetate decarboxylase	*Pseudoalteromonas haloplanktis*	54	1	PF01081
Q4ZXE6	kdpg and khg aldolase	*Pseudomonas syringae*	16	1	PF01081
Q5WD90	2-dehydro-3-deoxyphosphogluconate aldolase	*Bacillus clausii*	64	1	PF01081
Q5XD30	4-hydroxy-2-oxoglutarate aldolase	*Streptococcus pyogenes*	27	1	PF01081
Q7N9X3	khg/kdpg aldolase	*Photorhabdus luminescens*	17	1	PF01081
Q88P29	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Pseudomonas putida*	47	1	PF01081
Q8EMY7	2-keto-3-deoxy-6-phospho-gluconate aldolase	*Oceanobacillus iheyensis*	56	1	PF01081

(HPA): hydroxypyruvate; (PA): pyruvate; (GA): glyceraldehyde; known enzymes from *E. coli* of each Pfam family are in bold and used as reference

Table S7. Primers used for aldolase gene cloning

Selected Uniprot KB Id	**Annotation**	**Organism**	**Tag status**	**Genome from Genoscope strain collection used for PCR**	**Primer 5'**	**Primer 3'**
P76469	2-keto-3-deoxy-L-rhamnonate aldolase	*Escherichia coli*K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAACGCATTATTAAGCAATCCC	GTGTAATGGATAGTGATCTTAATAACTACCTTTTATGCGTGGC
P23522	5-keto-4-deoxy-D-glucarate aldolase	*Escherichia coli* K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAATAACGATGTTTTCCCG	GTGTAATGGATAGTGATCTTATTTTTTAAAGGTATCAGC
B1IS70	4-hydroxy-2-oxo-heptane-1,7-dioate aldolase	*Escherichia coli* C	Histag	DSM 1576	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGAAAACAGTTTTAAAGCGG	GTGTAATGGATAGTGATCTTAATACACGCCGGGCTTC
P0A6L2	4-hydroxy-tetrahydrodipicolinate synthase	*Escherichia coli* K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATTTCACGGGAAGTATTGTC	GTGTAATGGATAGTGATCTTACAGCAAACCGGCATG
P0A6L4	N-acetylneuraminate lyase	*Escherichia coli* K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGCAACGAATTTACGTG	GTGTAATGGATAGTGATCTTACCCGCGCTCTTGCATC
P0A955	KHG/KDPG aldolase	*Escherichia coli* K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAAAAACTGGAAAACAAGTGCAG	GTGTAATGGATAGTGATCTTACAGCTTAGCGCCTTCTAC
P51020	4-hydroxy-2-oxovalerate aldolase	*Escherichia coli* K12	Histag	*Escherichia coli* K12	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAACGGTAAAAAACTTTA	GTGTAATGGATAGTGATCTTATTTGTTGTTGCGCAGATCC
A0A081HJP9	putative uncharacterized protein	*Pseudomonas aeruginosa* 2192	Histag	DSM 50071	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGACCTGCCCGTCAATC	GTGTAATGGATAGTGATCTTAGCCGTAGACCGAGGAAC
Q0K203	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Ralstonia eutropha* (strain ATCC 17699 / H16 / DSM 428 / Stanier 337)	Histag	DSM 428	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCCCGCAAACAATCCC	GTGTAATGGATAGTGATCTTAGTACGCGGCGGAAGG
B2TGJ0	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Burkholderia phytofirmans* (strain DSM 17436 / PsJN)	Histag	DSM 17436	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATTCCCTGCCGCTTAAC	GTGTAATGGATAGTGATCTTAGTATCCGCCTTTGGG
B4XH86	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Actinobacillus pleuropneumoniae*	Histag	DSM 13472	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGCAATCGTAAAAGATT	GTGTAATGGATAGTGATCTTATTTCTCTAATTCTGCTT

A5VH82	hpch/hpai aldolase family protein	*Sphingomonas wittichii* (strain RW1 / DSM 6014)	Histag	DSM 6014	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAATAAGGTACGAACAT	GTGTAATGGATAGTGATCTTAAAGCGCCCTGGCGTC
B3B2X4 named B7NJZ1 in *Escherichia coli* IAI39	2-dehydro-3-deoxyglucarate aldolase	*Escherichia coli* O157:H7	Histag	*Escherichia coli* IAI39	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAATAACGATGTTTTCCCG	GTGTAATGGATAGTGATCTTATTTTTTAAAGGTATCAGC
A9CJL2	aldolase	*Agrobacterium tumefaciens* (strain C58 / ATCC 33970)	Histag	ATCC 33970	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCCCGCTCCCGAAAAC	GTGTAATGGATAGTGATCTTAATAGCCCGATCTGGA
A5VAX1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Sphingomonas wittichii* (strain RW1 / DSM 6014 / JCM 10273)	Histag	DSM 6014	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAACGACTTCAAGACCG	GTGTAATGGATAGTGATCTTAATAGATGGTTGCTCCGC
A9E0U5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Oceanibulbus indolifex* HEL-45	Histag	DSM 14862	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAAACACCCAGAGAACAAG	GTGTAATGGATAGTGATCTTACGTCTCAGAAATCTTGTCG
A3SLS0	hpch/hpai aldolase family protein	*Roseovarius nubinhibens* ISM	Histag	DSM 15170	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCCCGCGCCCACCAAC	GTGTAATGGATAGTGATCTTAGTCATCGGTGCGCCA
B2T1L6	2-dehydro-3-deoxyglucarate aldolase	*Burkholderia phytofirmans* (strain DSM 17436 / PsJN)	Histag	DSM 17436	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAGCACCTTCACCAATC	GTGTAATGGATAGTGATCTTATGACCTTACCTCCTGCAAC
Q16DB5	4-hydroxy-2-oxovalerate aldolase	*Roseobacter denitrificans* (strain ATCC 33942)	Histag	DSM 7001	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCCCGCGCCACTGAAC	GTGTAATGGATAGTGATCTTAACCCTGTTTCCACTTGGC
Q1QUJ4	2-dehydro-3-deoxyglucarate aldolase	*Chromohalobacter salexigens* (strain DSM 3043 / ATCC BAA-138)	Histag	DSM 3043	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCAACTGCCCCGCAAC	GTGTAATGGATAGTGATCTTAGCCCTTGTAGCGTGA
A7HU24	hpch/hpai aldolase family protein	*Parvibaculum lavamentivorans* (strain DS-1 / DSM 13023)	Histag	DSM 13023	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGCGGAAGCCGTGTCT	GTGTAATGGATAGTGATCTTATCCAGCATTTTTGTTC
A9BQY3	2-dehydro-3-deoxyglucarate aldolase	*Delftia acidovorans* (strain DSM 14801 / SPH-1)	Histag	DSM 14801	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCAACTGCCTGCCAAC	GTGTAATGGATAGTGATCTTAGTAGGTGCGGGATGC
A3K1H1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Sagittula stellata* E-37	Histag	DSM 11524	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGACCTGCCCAAGAAC	GTGTAATGGATAGTGATCTTAGGACCATTCGGCGCG

Q2K7V6	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Rhizobium etli* (strain CFN 42 / ATCC 51251)	Histag	DSM 11541	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATGAACTTCCCGTCAATC	GTGTAATGGATAGTGATCTTAAGTTGCAAATCGGGC
Q9RYM1	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Deinococcus radiodurans* (strain ATCC 13939 / DSM 20539)	Histag	DSM 20539	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATCCGCAACCGATGAAG	GTGTAATGGATAGTGATCTTAATACACGCTGCCGCC
I7DKY0 (former A9EU74)	hpch/hpai aldolase family protein	*Phaeobacter gallaeciensis* (strain 2.10)	Histag	DSM 17395	AAAGAAGGAGATAGGATCATGcccgcacccaaaaat	GTGTAATGGATAGTGATCTTActtcgatgtccagtccg
B6XAH9	putative uncharacterized protein	*Providencia alcalifaciens* DSM 30120	no histag	DSM 30120	AAAGAAGGAGATAGGATCATGTATAAGGAGTTCACCGTGGATCTG	GTGTAATGGATAGTGATCTTAGTACACACTCGAGCCGCTG
Q1R1H4	HpcH/HpaI aldolase	*Chromohalobacter salexigens* (strain DSM 3043 / ATCC BAA-138)	no histag	DSM 3043	AAAGAAGGAGATAGGATCATGAAACGACACAACCCC	GTGTAATGGATAGTGATCTTAGACGGTGTCGAAGTCC
E0E947	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Actinobacillus pleuropneumoniae* serovar 1 str. 4074	no histag	DSM 13472	AAAGAAGGAGATAGGATCATGGCAATCGTAAAAGATTTTT	GTGTAATGGATAGTGATCTTATTTCTCTAATTCTGCTTTAA
F4KUN3	4-hydroxy-2-oxovalerate aldolase	*Haliscomenobacter hydrossis* (strain ATCC 27775 / DSM 1100 / LMG 10767 / O)	no histag	DSM 1100	AAAGAAGGAGATAGGATCATGGAAGTAACAAAAAATAAATTC	GTGTAATGGATAGTGATCTTAATACCTGGTATTTGCACT
A9IPR8	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Bordetella petrii* (strain ATCC BAA-461 / DSM 12804 / CCUG 43448)	no histag	DSM 12804	AAAGAAGGAGATAGGATCATGGATATCCTGACCAATCAATTC	GTGTAATGGATAGTGATCTTAATAGGGCCCGCCAGG
B1G064	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Burkholderia graminis* C4D1M	no histag	DSM 17151	AAAGAAGGAGATAGGATCATGTCCTTGCCGCAAAACACATTCAAGCG	GTGTAATGGATAGTGATCTTAGTACCCGCCTTTCACGGCC
B1G3M5	2-dehydro-3-deoxyglucarate aldolase	*Burkholderia graminis* C4D1M	no histag	DSM 17151	AAAGAAGGAGATAGGATCATGAGCACTTTCACGAATCCCCTCAAG	GTGTAATGGATAGTGATCTTATGCTCTTACCTCCTGCAACGCCTG
B1K8T9	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Burkholderia cenocepacia* (strain MC0-3)	no histag	DSM 16553	AAAGAAGGAGATAGGATCATGCAGATTCCGTCGAATGTCTTCAAGG	GTGTAATGGATAGTGATCTTAATACGTGCCGTCGCCCTGCTTCACG
B7WZH5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Comamonas testosteroni* KF-1	no histag	DSM 14576	AAAGAAGGAGATAGGATCATGCCTGCATACAACCCGTTCAAGACCGC	GTGTAATGGATAGTGATCTTAGTAGGCCGCACCAGGCCGAGCCG

B9JRF3	aldolase	*Agrobacterium vitis* (strain S4 / ATCC BAA-846)	no histag	ATCC BAA-846	AAAGAAGGAGATAGGATCATGCCCGCACCCGTCAACGGTTTCAA	GTGTAATGGATAGTGATCTTAATAACCGCCAGTCGCCTCTGGTCCTGC
B9Z3Z6	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Pseudogulbenkiania ferrooxidans* 2002	no histag	DSM 18807	AAAGAAGGAGATAGGATCATGCACCTCCCCACCAACCCAT	GTGTAATGGATAGTGATCTTAGTACACGCCGGACGAACCG
C6BIY7	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Ralstonia pickettii* (strain 12D)	no histag	DSM 6297	AAAGAAGGAGATAGGATCATGCAACTCCCGGTGAACAACTTCAAGCG	GTGTAATGGATAGTGATCTTAATACACACTGGCGGCGGGCTGGG
D4DXK2	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Serratia odorifera* DSM 4582	no histag	DSM 4582	AAAGAAGGAGATAGGATCATGTTAACCAATGCCTTCAAGCTGGCG	GTGTAATGGATAGTGATCTTAGTAGACCCCGGAAGGTTGCGTCGGC
D7F5M7 named as A0A0E3VZZ4 in *Pseudomonas putida* DSM 291	hpai	*Pseudomonas putida* (strain ATCC 12633/DSM 291)	no histag	DSM 291	AAAGAAGGAGATAGGATCATGGACATGCCCATCAACCACTTC	GTGTAATGGATAGTGATCTTAGTAGGCACCACCCCCGGCAC
H0GA44 named A0A0F4Y2G3 in *Sinorhizobium meliloti* DSM 30135	4-hydroxy-2-oxovalerate aldolase	*Sinorhizobium meliloti* CCNWSX0020	no histag	DSM 30135	AAAGAAGGAGATAGGATCATGCCCGCCCCCAAAAACACCTTCAAAGC	GTGTAATGGATAGTGATCTTAATAGCCGGATTCTATCCGTGTCGGTC
A5VED8	2-dehydro-3-deoxyglucarate aldolase	*Sphingomonas wittichii* (strain RW1 / DSM 6014 / JCM 10273)	no histag	DSM 6014	AAAGAAGGAGATAGGATCATGAACGCGCTCAAGACG	GTGTAATGGATAGTGATCTTAGCCCTTCAGCGCGTC
A9BTL5	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	*Delftia acidovorans* (strain DSM 14801 / SPH-1)	no histag	DSM 14801	AAAGAAGGAGATAGGATCATGCCAGCCCACAACCCCTTCA	GTGTAATGGATAGTGATCTTAATACGCAGCCCCGAGGCGC
Q9HWQ3	putative uncharacterized protein	*Pseudomonas aeruginosa* (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228)	no histag	DSM 22644	AAAGAAGGAGATAGGATCATGGACCTGCCCGTCAATCGC	GTGTAATGGATAGTGATCTTAGCCGTAGACCGAGGAACTGGCGGAC
Q5WD90	2-dehydro-3-deoxyphosphogluconate aldolase	*Bacillus clausii* KSM-K16	no histag	DSM 8716	AAAGAAGGAGATAGGATCATGAGTTTAGACAAGCTTTTGTCTTCTG	GTGTAATGGATAGTGATCTTAGTTGCCCCTTGCCGT

A0KF01	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Aeromonas hydrophila* (strain ATCC 7966 / NCIB 9240)	no histag	DSM 30187	AAAGAAGGAGATAGGATCATGCAGAACTGGAAAGTGAC	GTGTAATGGATAGTGATCTTAGGATGCGACCAGGGC
G2TP95 (former C1P814)	2-dehydro-3-deoxyphosphogluconate /4-hydroxy-2-oxoglutarate aldolase	*Bacillus coagulans* 36D1	no histag	DSM 1	AAAGAAGGAGATAGGATCATGAAAAAAATGCAGACTTT	GTGTAATGGATAGTGATCTTAGGCATATTGACCCCTC
C2JUN9	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Lactobacillus rhamnosus* LMS2-1	no histag	DSM 20021	AAAGAAGGAGATAGGATCATGCAAAAATACACCTTTCTAAGCC	GTGTAATGGATAGTGATCTTATCTCCTAGATTGTTTGATTTG
I7EZU8 (former A9FVJ2)	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Phaeobacter gallaeciensis* (strain ATCC 700781 / DSM 17395 / CIP 105210)	no histag	DSM 17395	AAAGAAGGAGATAGGATCATGGCCATCACCCCACAA	GTGTAATGGATAGTGATCTTAGCGTGCCAGTTGGCT
Q88P29	2-dehydro-3-deoxyphosphogluconate /4-hydroxy-2-oxoglutarate aldolase	*Pseudomonas putida* KT2440	no histag	DSM 6125	AAAGAAGGAGATAGGATCATGCCCATGAGCCAAGGA	GTGTAATGGATAGTGATCTTAGTTGGCGTCCAGCAG
A1HMS5	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Thermosinus carboxydivorans* Nor1	no histag	DSM 14886	AAAGAAGGAGATAGGATCATGACCAAGACCGAAGTTG	GTGTAATGGATAGTGATCTTAGAAGAACCGTCCCTCTGACAA
B2UI88	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	*Ralstonia pickettii strain* 12j	no histag	DSM 6297	AAAGAAGGAGATAGGATCATGGGTATCGATTCCGTC	GTGTAATGGATAGTGATCTTACCGCTTGAGCGCGGC
Q3IH59	putative 4-hydroxy-2-oxoglutarate aldolase / 2-dehydro-3-deoxyphosphogluconate aldolase	*Pseudoalteromonas haloplanktis* strain TAC125	no histag	strain TAC125	AAAGAAGGAGATAGGATCATGACTCGCTTTTCTGAACTTATG	GTGTAATGGATAGTGATCTTACTTTAATGCTTGAATACACGAG
Q3DJ30	2-dehydro-3-deoxyphosphogluconate/4-hydroxy-2-oxoglutarate aldolase	*Streptococcus agalactiae* 515	no histag	strain NEM316	AAAGAAGGAGATAGGATCATGATAAAATCAAAGATTTTG	GTGTAATGGATAGTGATCTTAATTCAGATTTAGCTGTATTTTG
Q1J2P1	2-keto-3-deoxy-6-phosphogluconate aldolase	*Deinococcus geothermalis* (strain DSM11300)	no histag	DSM 11300	AAAGAAGGAGATAGGATCATGGTTGATCTGCTCAGCG	GTGTAATGGATAGTGATCTTATGGATTGGTTGCGCACTG

Q8EMY7	2-keto-3-deoxy-6-phospho-gluconate aldolase	*Oceanobacillus iheyensis* (strain DSM 14371 / JCM 11309 / KCTC 3954)	no histag	DSM 14371	AAAGAAGGAGATAGGATCATGTCAGTAATAGAAGGTATTCAAGAAG	GTGTAATGGATAGTGATCTTAAACTTCTTTTGCGCTTTG
Q2G569	2-keto-3-deoxy-phosphogluconate aldolase	*Novosphingobium aromaticivorans* (strain DSM 12444)	no histag	DSM 12444	AAAGAAGGAGATAGGATCATGAAGCCGATCGAAACTATC	GTGTAATGGATAGTGATCTTACTGCCCCAATGCCGA
Q4ZXE6	kdpg and khg aldolase	*Pseudomonas syringae* (strain B728a)	no histag	DSM 10604	AAAGAAGGAGATAGGATCATGACACAGAACGAAAATAATCAGC	GTGTAATGGATAGTGATCTTAGTCAAACAGCGCCAT
B2UCA9	kdpg and khg aldolase	*Ralstonia pickettii* strain 12J	no histag	DSM 6297	AAAGAAGGAGATAGGATCATGACCTGGACAACCCAC	GTGTAATGGATAGTGATCTTAAGTGCGGTCGGATTG
A5MQQ9	keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	*Streptococcus pneumoniae* SP19-BS75	no histag	DSM 20566	AAAGAAGGAGATAGGATCATGACCAAATCAGATACGATTATTGAAC	GTGTAATGGATAGTGATCTTATCTGAGTGTAATATACTGTTGGGC
Q7N9X3	khg/kdpg aldolase	*Photorhabdus luminescens subsp. laumondii* (strain TT01)	no histag	DSM 15139	AAAGAAGGAGATAGGATCATGATGAAACAGATGATTGAACAGTTACG	GTGTAATGGATAGTGATCTTAAGTATTCACCAAATCCACC
A6B1X9	khg/kdpg aldolase	*Vibrio parahaemolyticus* AQ3810	no histag	DSM 10027	AAAGAAGGAGATAGGATCATGAAAGATCTGAACCAACAAC	GTGTAATGGATAGTGATCTTATTCGATAATTCCCGCAAC
B7LPJ4	multifunctional 2-keto-3-deoxygluconate 6-phosphate aldolase and 2-keto-4-hydroxyglutarate aldolase and oxaloacetate decarboxylase	*Escherichia fergusonii* (strain ATCC 35469 / DSM 13698 / CDC 0568-73)	no histag	ATCC 35469	AAAGAAGGAGATAGGATCATGAAAAACTGGAAAACAAGTGCG	GTGTAATGGATAGTGATCTTACAACTTCGCGCCTTC
Q3IL59	multifunctional: 2-keto-3-deoxygluconate 6-phosphate aldolase; 2-keto-4-hydroxyglutarate aldolase; oxaloacetate decarboxylase	*Pseudoalteromonas haloplanktis* strain TAC125	no histag	strain TAC125	AAAGAAGGAGATAGGATCATGAGTATTGAAAAAATATTAG	GTGTAATGGATAGTGATCTTATTTTTGACTCCTATTTACTTCACAAG
Q5XD30	4-hydroxy-2-oxoglutarate aldolase	*Streptococcus pyogenes serotype M6* (strain ATCC BAA-946 / MGAS10394)	no histag	DSM 20565	AAAGAAGGAGATAGGATCATGTCCAGAGGTGATGATATGG	GTGTAATGGATAGTGATCTTATTGTCGCAAACTTTTGAT

Q16DB5	4-hydroxy-2-oxovalerate aldolase	*Roseobacter denitrificans* (strain ATCC 33942)	no histag	DSM 7001	AAAGAAGGAGATAGGATCATGCCCGCGCCACTGAAC	GTGTAATGGATAGTGATCTTAACCCTGTTTCCACTTGGC
Q1VG27 named as A0A0S3K3Y1 in *Vibrio alginolyticus* DSM 2171	4-hydroxy-2-oxovalerate aldolase	*Vibrio alginolyticus* 12G01	no histag	DSM 2171	AAAGAAGGAGATAGGATCATGAGCAATTTTAAGCAACG	GTGTAATGGATAGTGATCTTAAGAATTCATCACATTTCTCTCTTG
A4NYS5	n-acetylneuraminate lyase	*Haemophilus influenzae* 22.4-21	no histag	DSM 11121	AAAGAAGGAGATAGGATCATGCGTGATTTAAAAGGTATTTTC	GTGTAATGGATAGTGATCTTATGACAAAAATTTCGCTTTC
Q5E738	n-acetylneuraminate lyase	*Vibrio fischeri* (strain ATCC 700601 / ES114)	no histag	DSM 507	AAAGAAGGAGATAGGATCATGAAAAAGTTAACCGGTT	GTGTAATGGATAGTGATCTTATTGTATGAAACCAGCGTTTCTC
Q3ID27	n-acetylneuraminate lyase (aldolase)	*Pseudoalteromonas haloplanktis* strain TAC125	no histag	strain TAC125	AAAGAAGGAGATAGGATCATGAATAAACTTACTGGACTGATCG	GTGTAATGGATAGTGATCTTACGCTGGTTGGTGATCC

4. Figures

Figure S1. Distribution into Pfam families of the proteins of the collection

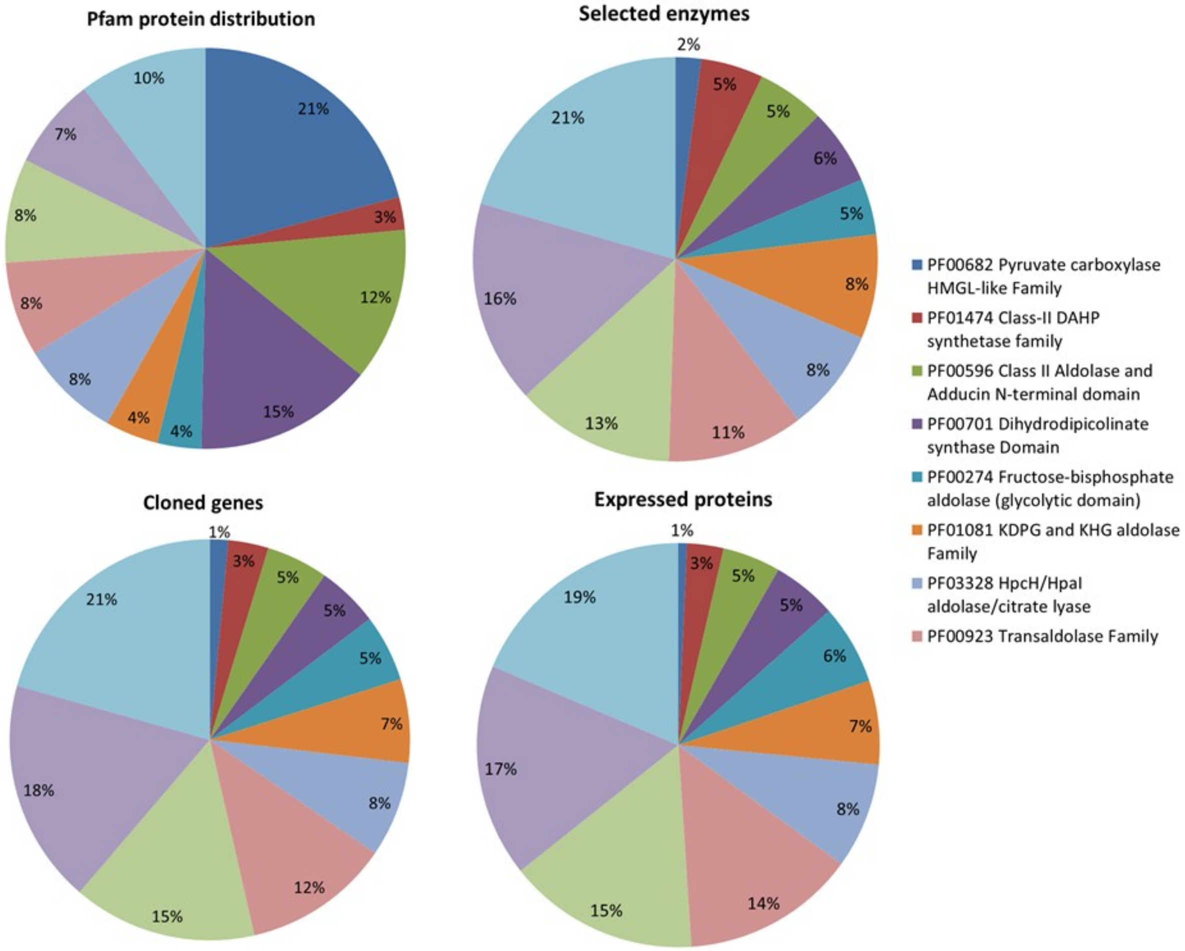
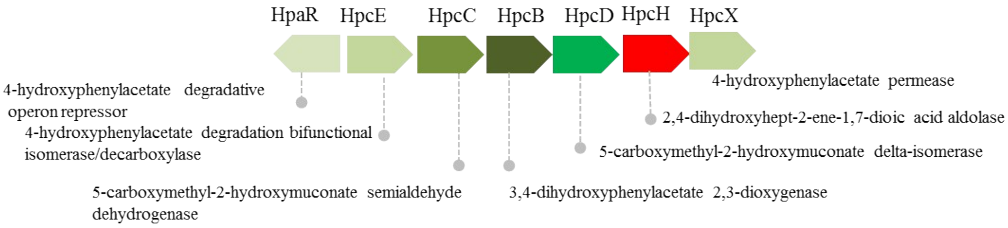
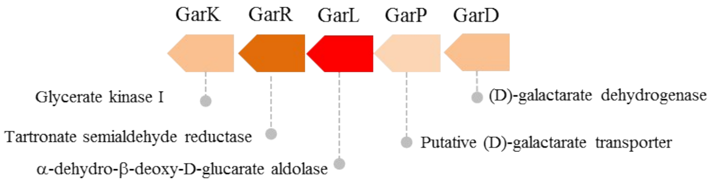


Figure S2. Genomic organisation of the gene clusters of RhmA and GarL from *E coli* K12 and HpcH from *E. coli* C

HpcH from *E. coli* C (B1IS70): 4-hydroxyphenylacetate catabolism



GarL from *E. coli* K12 (P23522) : (D)-galactarate catabolism



RhmA from *E. coli* K12 (P76469): unknown metabolic pathway

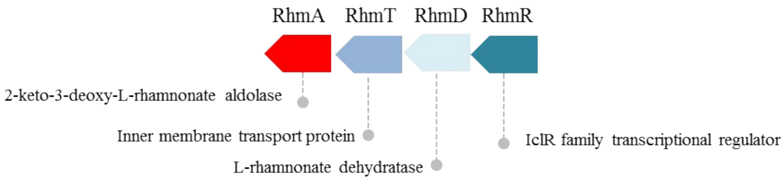
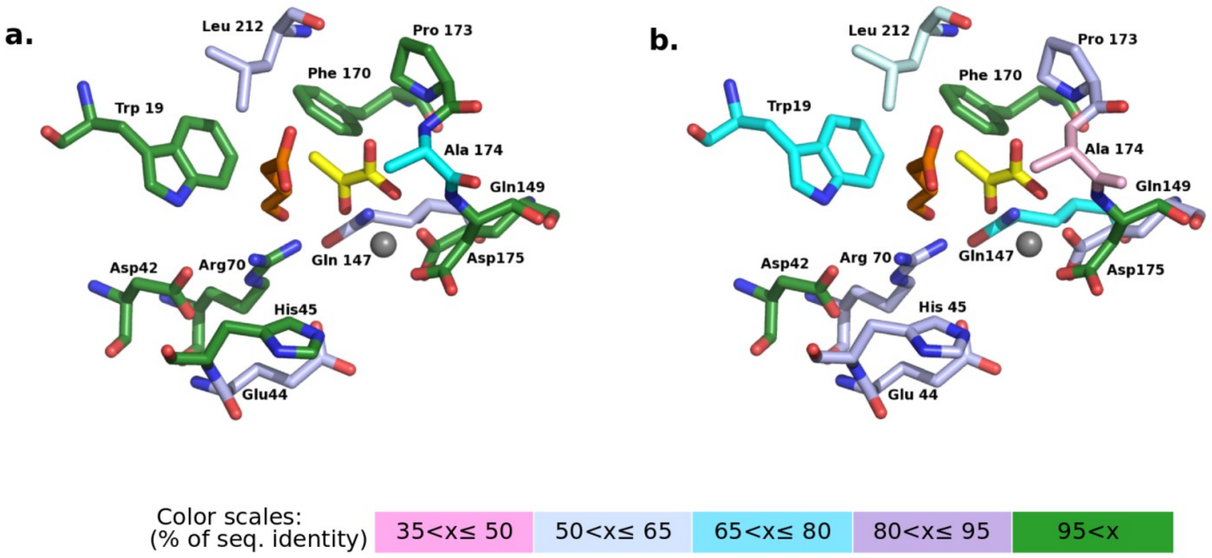


Figure S3. Representation of the active site conservation among the 19 candidates (**a.**) and among a subset of 1481 proteins of the PF03328 Pfam family (**b.**)



The degree of conservation was mapped on the *E. coli* Hpch/HpaI active site[7] with color: $35<x \leq 50\%$ in pink, $50<x \leq 65\%$ in light blue, $65< x \leq 80\%$ in cyan, $80<x \leq 95\%$ in purple and $> 95\%$ in dark green. Mg^{2+} is represented as a gray sphere. The bound pyruvate substrate is represented in yellow, while the succinic semialdehyde is in orange. For more clarity, the water ligands were omitted.

---

[1] B. Zagalak, P.F. Frey, G.L. Karabatsos, R.H. Abeles, *J. Biol. Chem.* **1966**, *241*, 3028-3035.

[2] C. Wikner, L. Meshalkina, U. Nilsson, M. Nikkola, Y. Lindqvist, M. Sundstrom, G. Schneider, *J. Biol. Chem.* **1994**, *269*, 32144-32150.

[3] S. Rozen, H. Skaletsky, *Methods Mol. Biol.* **2000**, *132*, 365-386.

[4] C. Aslanidis, P. J. de Jong, *Nucleic Acids Res.* **1990**, *18*, 6069-6074.

[5] C. Vergne-Vaxelaire, F. Bordier, A. Fossey, M. Besnard-Gonnet, A. Debard, A. Mariage, V. Pellouin, A. Perret, J. Petit, M. Stam, M. Salanoubat, J. Weissenbach, V. de Berardinis, A. Zaparucha, *Adv. Synth. Catal.* **2013**, *355*, 1763-1779.

[6] H. Hara, E. Masai, K. Miyauchi, Y. Katayama, M. Fukuda, *J. Bacteriol.* **2003**, *185*, 41-50.

[7] M. Coincon, W. Wang, J. Sygusch, S. Y. K. Seah, *J. Biol. Chem.* **2012**, *287*, 36208-36221.