Supporting Information for Publication

Experimental Strategies for Functional Annotation and Metabolism Discovery: Targeted Screening of Solute Binding Proteins and Unbiased Panning of Metabolomes

*Matthew W. Vetting[1,10], Nawar Al-Obaidi[1,10], Suwen Zhao[2,10], Brian San Francisco[3,10], Jungwook Kim[1], Daniel J. Wichelecki [3,4], Jason T. Bouvier[3,4], Jose O. Solbiati[3], Hoan Vu[6], Xinshuai Zhang[3], Dmitry A. Rodionov [7,8], James D. Love[1], Brandan S. Hillerich[1], Ronald D. Seidel[1], *Ronald J. Quinn[6], *Andrei L. Osterman[7], *John E. Cronan[4,9], *Matthew P. Jacobson[2], *John A. Gerlt[3,4,5], *Steven C. Almo[1]

Table S1. Primers used in this study.

Purpose	Primer name	Primer sequence (5′ to 3′)
qRT-PCR	LysR_RTPCR_FOR	CAAGGAAGATGCCGTGGATC
qRT-PCR	LysR_RTPCR_REV	ATAAAGCGCCATATCGACCAGT
qRT-PCR	DctP_RTPCR_FOR	ATCAGCAGTCCGGAGGATTTC
qRT-PCR	DctP_RTPCR_REV	GTCAGATACTTCTGCACCTCG
qRT-PCR	DctQ_RTPCR_FOR	CCGATCTGGAGCGATGAACT
qRT-PCR	DctQ_RTPCR_REV	CGTATTGCCAACTCCACCAG
qRT-PCR	DctM_RTPCR_FOR	CTTTACGTGGCGCTGTGTG
qRT-PCR	DctM_RTPCR_REV	GACCGGTGATGCAATGAAGG
qRT-PCR	VanX_RTPCR_FOR	GAAATGGTGGAAACACGCTATG
qRT-PCR	VanX_RTPCR_REV	GCTCGAAATGCACGGTATGC
Knockout	VanX_KO_H_FOR	CCGGCATCGTGCTGGAATTCGTCACGCCCACC
Knockout	VanX_KO_Hext_REV	CATCCCTCCAGTCCGGCGGTTCGTTTCCAGCAAGC
Knockout	VanX_KO_Bext_FOR	GCTTGCTGGAAACGAACCGCCGGACTGGAGGGATG
Knockout	VanX_KO_B_REV	GTCGCGCCTGCCGGATCCTGCTGCGCC
Knockout	DctP_KO_H_FOR	GGATCTGGCCCTGCTGAATTCGCCTCCCGATGATC
Knockout	DctP_KO_Hext_REV	CCTCAAGGAAGTGAAGGAAGAAAACCGGATGTCTCCTTGTTGTTATG
Knockout	DctP_KO_Bext_FOR	CATAACAACAAGGAGACATCCGGTTTTCTTCCTTCACTTCCTTGAGG
Knockout	DctP_KO_B_REV	CCCCAGCAGGGGATCCACCAGCAAGGCGG
qRT-PCR	Csal0675_left	CACGCGTCGTACCATATTCC
qRT-PCR	Csal0675_right	CGATTGCAGTCGTAGACCAG
qRT-PCR	Csal0676_left	GAGCCGTTCGACTTCCTCTT
qRT-PCR	Csal0676_right	ATCAGCAGCACACCGAGAC
qRT-PCR	Csal0677_left	CATCGAGGGCCTGTTTACC
qRT-PCR	Csal0677_right	TCGTGTAGAGAACGCCATACC
qRT-PCR	Csal0678_left	GGCATCACCGAGGACTACAC
qRT-PCR	Csal0678_right	TCACAGCACACCTCGAACA
qRT-PCR	Csal0679_left	TCCAGACCACCAAGGAAGAG
qRT-PCR	Csal0679_right	GTAATCCGCCGTATCGAGTG
qRT-PCR	Csal0680_left	CGATGAATGACGAACTGGTG
qRT-PCR	Csal0680_right	GCGCTTGATGTAGCGTTTG
qRT-PCR	Csal0681_left	GTGATCTGTGATCCCGAACTG
qRT-PCR	Csal0681_right	GCGTGGCAGATATTCCTTGA
qRT-PCR	Csal16s_left	AAGAAGCACCGGCTAACTCC
qRT-PCR	Csal16s_right	TCCCGATCTCTACGCATTTC
Knockout	Csal0678_EcoRI_Fwd	CTGCTCGAATTCGTACGACAAGATCATGGTC
Knockout	Csal0678_HindIII_Rev	CTGCAAGCTTGACTGGTGGATATGACTGG
Knockout	Csal0679_EcoRI_Fwd	CACAGAATTCAAGGAAATCTGGGAAGAGGAGTAC
Knockout	Csal0679_HindIII_Rev	CAATAAGCTTTCAACAGTACGCTGTTGCCTG
qRT-PCR	LysR_RTPCR_FOR	CAAGGAAGATGCCGTGGATC
qRT-PCR	LysR_RTPCR_REV	ATAAAGCGCCATATCGACCAGT
qRT-PCR	DctP_RTPCR_FOR	ATCAGCAGTCCGGAGGATTTC
qRT-PCR	DctP_RTPCR_REV	GTCAGATACTTCTGCACCTCG
qRT-PCR	DctQ_RTPCR_FOR	CCGATCTGGAGCGATGAACT
qRT-PCR	DctQ_RTPCR_REV	CGTATTGCCAACTCCACCAG
qRT-PCR	DctM_RTPCR_FOR	CTTTACGTGGCGCTGTGTG
qRT-PCR	DctM_RTPCR_REV	GACCGGTGATGCAATGAAGG
qRT-PCR	VanX_RTPCR_FOR	GAAATGGTGGAAACACGCTATG
qRT-PCR	VanX_RTPCR_REV	GCTCGAAATGCACGGTATGC

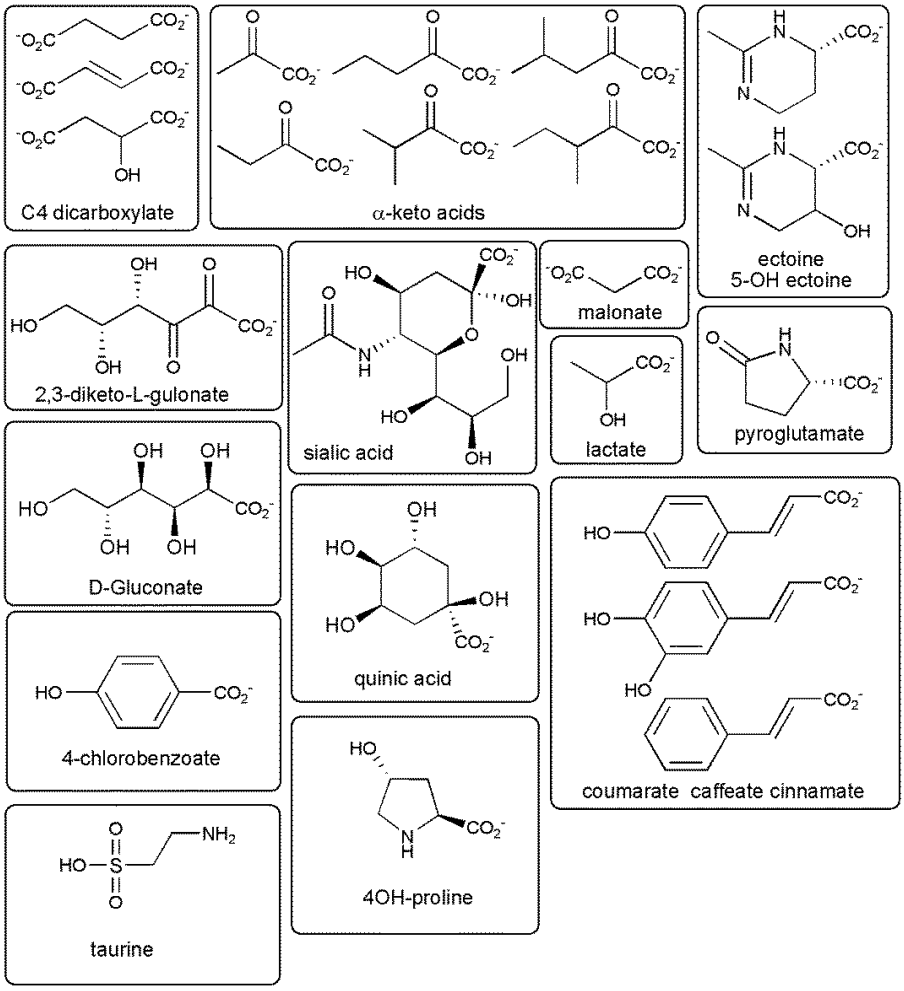


Figure S1

Schematic of the previously determined ligands of TRAP SBPs.

These ligands map to specific clusters within the 10^{-120} TRAP SBP SSN. C4-dicarboxylates (cluster 10, 74), alpha-keto acids (2), ectoine/5-OH ectoine (13), 2,3-diketo-L-gulonate (11), sialic acid (48,64), malonate (17), lactate (220), pyroglutamate (31,166), D-gluconate (43), quinic acid (100), coumarate/caffeate/cinnimate (28), 4-chlorobenzoate (475), taurine (78), 4OH-proline (61).

Table S2 Mapping of ligand, cluster number and their associated colors within the E^{-120} SSN (**Figure 3**)

Node Color	Ligand(s)	Cluster #s in the 10^{-120} SSN
navy	(S)-2-acetolactate	1, 12, 1672
lightskyblue	α-keto acids	2, 191
palegreen	4-carbon diacids	3, 9, 10, 74
sandybrown	D-galacturonate/D-glucuronate	4, 44, 47, 65, 186, 310, 319, 452, 790
deepskyblue	D-glucuronate	5, 30, 161, 585
magenta	2,3-diketo-L-gulonate	11
cyan	ectoine, 5-hydroxyectoine	13, 254, 389, 1038
peru	pyruvate	15
indigo	malonate	17, 35
olive	phenolic acids	21, 28, 139, 180, 279, 429, 1675
darksalmon	5/6-carbon aldonic acids	27, 545
yellow	pyroglutamate	31, 124, 126, 166, 190, 946
pink	ethanolamine	33
green	6-carbon aldonic acids	43, 297
honeydew	-	45
blue	sialic acid	48, 64
olivedrab	L-fuconate/L-galactonate	49, 197, 403
metallic bronze	non-specific sugar acids	50
plum	trans-4-hydroxy-L-proline	61
orange	taurine	78
deep teal	glycerol-3-phosphate	87, 244
deep fir	L-guluronate/L-galacturonate	95, 323
teal	D-mannonate/L-gulonate	97, 223
lightcyan	(R)-pantoate	98
slateblue	D-quinic acid	100
firebrick	D-alanyl-D-alanine	104, 541
black	3-deoxy-D-manno-oct-2-ulosonic acid	107
burlywood	D-erythronate	123, 219
wheat	(putative) lactic acid	147
lightcoral	orotate/dihydrooroate	153
slategray	indole 3-pyruvate/indole 3-acetate	171, 588, 1677
crimson	D-mannuronate	214, 472
fandango	L-erythronate	217
lime	L-lactate	220
silver	D-taluronate/D-mannuronate	267
dove grey	diglycerol phosphate	277
lightseagreen	mandelate	350, 582
beige	(putative) suberate	426
hotpink	4-chlorobenzoate	475
cadetblue	benzoylformate	614
paleturquoise	(putative) D-ribonate	618
eggplant	L-ribonate/D-talonate	1646
royalblue	(putative) D-glycerate	1688
palevioletred	(putative) hydroxylated aromatic carboxylate	1690

Table S3 Number of sequences that map to the 71 clusters for which ligands were determined in this study

Cluster #	Ligand	# of sequences	Cluster #	Ligand	# of sequences
1	(S)-2-acetolactate	**142**	33	ethanolamine	**68**
12	(S)-2-acetolactate	**83**	297	6-carbon aldonic acids	**3**
1672	(S)-2-acetolactate	**1**	49	L-fuconate/L-galactonate	**152**
191	α-keto acids	**12**	197	L-fuconate/L-galactonate	**5**
3	4-carbon diacids	**226**	403	L-fuconate/L-galactonate	**2**
9	4-carbon diacids	**30**	50	non-specific sugar acids	**64**
4	D-galacturonate/D-glucuronate	**70**	87	glycerol-3-phosphate	**32**
30	D-galacturonate/D-glucuronate	**164**	244	glycerol-3-phosphate	**6**
44	D-galacturonate/D-glucuronate	**149**	95	L-guluronate/L-galacturonate	**18**
47	D-galacturonate/D-glucuronate	**67**	323	L-guluronate/L-galacturonate	**3**
65	D-galacturonate/D-glucuronate	**35**	97	D-mannonate/L-gulonate	**9**
186	D-galacturonate/D-glucuronate	**5**	223	D-mannonate/L-gulonate	**11**
310	D-galacturonate/D-glucuronate	**4**	98	(R)-pantoate	**21**
319	D-galacturonate/D-glucuronate	**3**	104	D-alanyl-D-alanine	**29**
452	D-galacturonate/D-glucuronate	**2**	541	D-alanyl-D-alanine	**2**
290	D-galacturonate/D-glucuronate	**3**	107	3-deoxy-D-manno-oct-2-ulosonic acid	**16**
5	D-glucuronate	**74**	123	D-erythronate	**12**
161	D-glucuronate	**8**	219	D-erythronate	**6**
585	D-glucuronate	**2**	147	lactic acid	**8**
254	ectoine, 5-hydroxyectoine	**5**	153	orotate/dihydrooroate	**6**
389	ectoine, 5-hydroxyectoine	**2**	171	indole 3-pyruvate/indole 3-acetate	**4**
1038	ectoine, 5-hydroxyectoine	**1**	588	indole 3-pyruvate/indole 3-acetate	**2**
15	pyruvate	**322**	1677	indole 3-pyruvate/indole 3-acetate	**1**
21	phenolic acids	**35**	214	D-mannuronate	**7**
139	phenolic acids	**8**	472	D-mannuronate	**2**
180	phenolic acids	**5**	217	L-erythronate	**4**
279	phenolic acids	**5**	267	D-taluronate/D-mannuronate	**4**
429	phenolic acids	**2**	277	diglycerol phosphate	**4**
1675	phenolic acids	**1**	350	mandelate	**9**
27	5/6-carbon aldonic acids	**62**	582	mandelate	**2**
545	5/6-carbon aldonic acids	**2**	426	suberate	**2**
124	pyroglutamate	**11**	614	benzoylformate	**2**
126	pyroglutamate	**12**	618	D-ribonate	**2**
190	pyroglutamate	**8**	1646	L-ribonate/D-talonate	**1**
946	pyroglutamate	**2**	1688	D-glycerate	**1**
			1690	phenolic acids	**1**

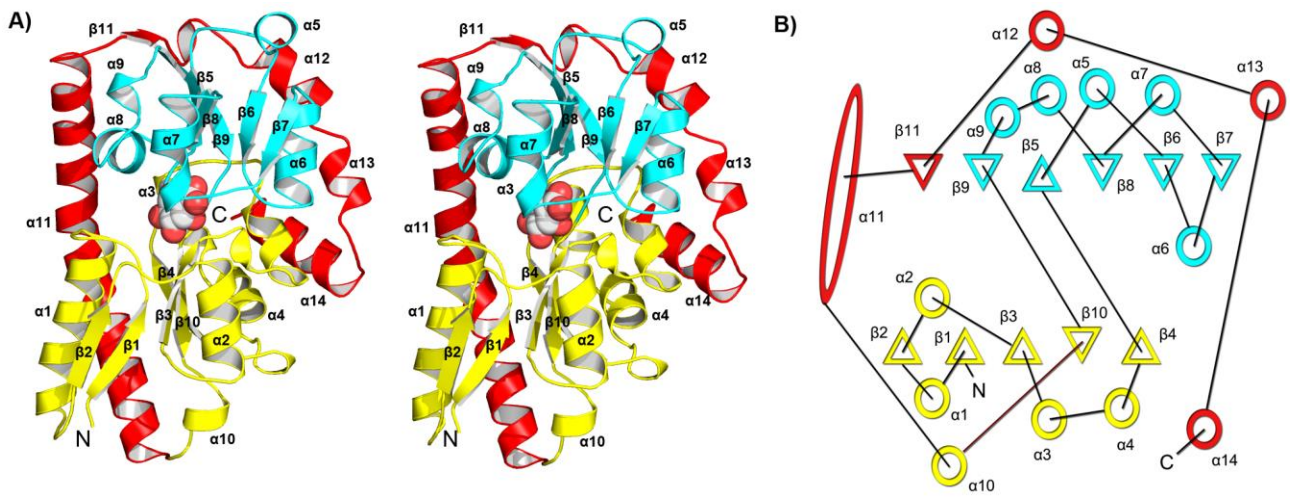


Figure S2

TRAP SBP FOLD.

A) Stereo ribbon diagram of Bamb_6123 from *Burkholderia ambifaria* with bound D-galacturonate (PDBID=4N17). TRAP SBPs are composed of two domains, domain 1 (shown in yellow) and domain 2 (shown in cyan), and a four helix α-helical wrapper (shown in red). **B)** Schematic of the TRAP SBP fold. Strands are shown as circles and α-helicies are show as triangles.

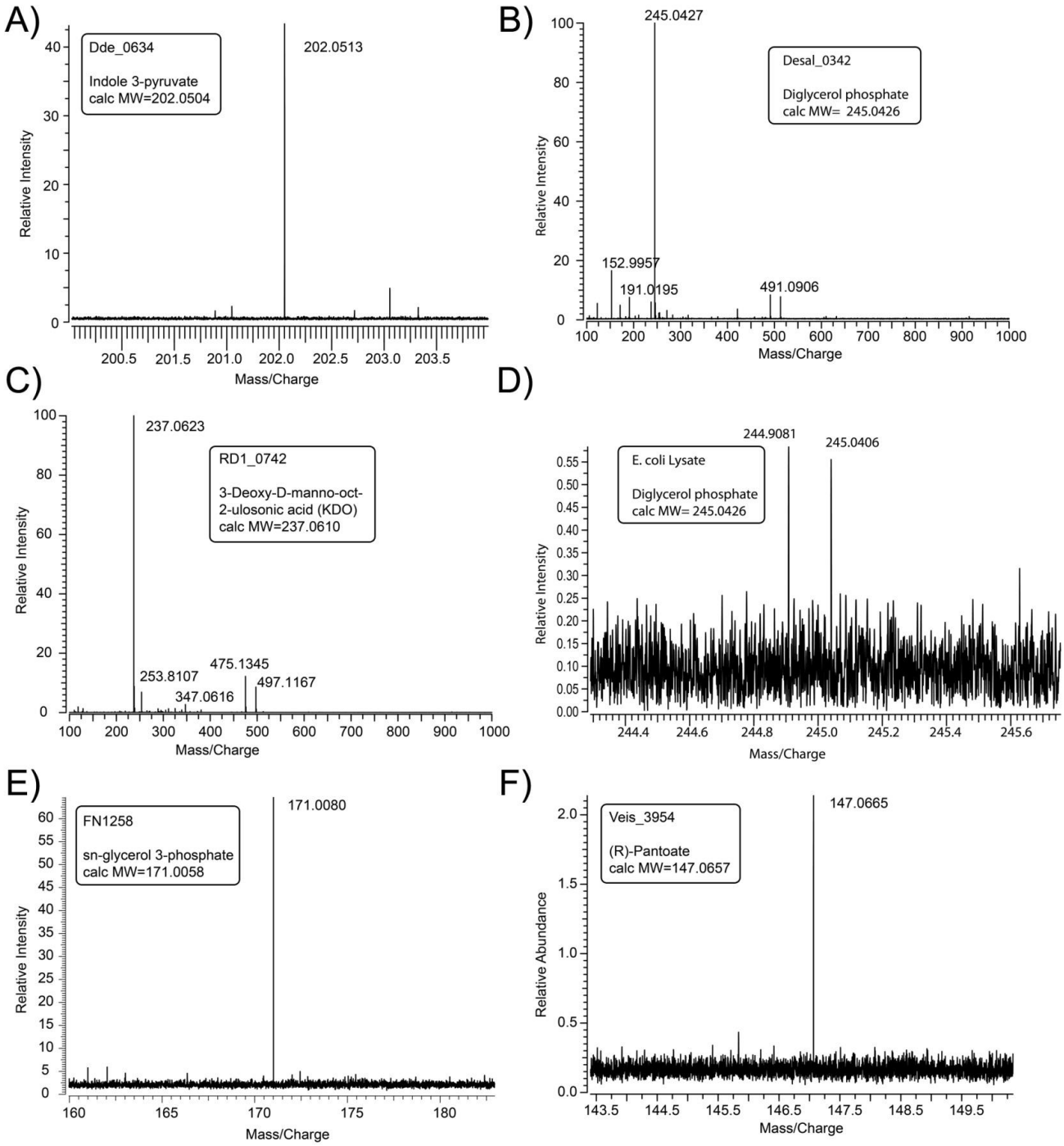

A)
Dde_0634
Indole 3-pyruvate
calc MW=202.0504
202.0513
Relative Intensity
Mass/Charge
B)
245.0427
Desal_0342
Diglycerol phosphate
calc MW= 245.0426
152.9957
191.0195
491.0906
Relative Intensity
Mass/Charge
C)
237.0623
RD1_0742
3-Deoxy-D-manno-oct-
2-ulosonic acid (KDO)
calc MW=237.0610
253.8107
347.0616
475.1345
497.1167
Relative Intensity
Mass/Charge
D)
244.9081
245.0406
E. coli Lysate
Diglycerol phosphate
calc MW= 245.0426
Relative Intensity
Mass/Charge
E)
171.0080
FN1258
sn-glycerol 3-phosphate
calc MW=171.0058
Relative Intensity
Mass/Charge
F)
147.0665
Veis_3954
(R)-Pantoate
calc MW=147.0657
Relative Abundance
Mass/Charge

Figure S3

X-FTMS of TRAP Ligands.

In **panels A-C** and **E-F** free ligands were separated from the TRAP SBP by methanol extraction. All proteins were expressed in *E. coli* and purified by metal affinity and size exclusion chromatography. In **panel D**, cellular extracts of *E. coli* (no expression plasmid) were methanol extracted to obtain a small molecule metabolome sample for X-FTMS analysis. X-FTMS peaks for **A)** indole 3-pyruvate extracted from Dde_0634, **B)** diglycerol phosphate extracted from Desal_0342, **C)** 3-deoxy-D-manno-oct-2-ulosonic acid (KDO) extracted from RD1_0742, **D)** diglycerol phosphate extracted from *E. coli* whole cells, **E)** sn-glycerol 3-phosphate extracted from FN1258 and **F)** (R)-pantoate extracted from Veis_3954.

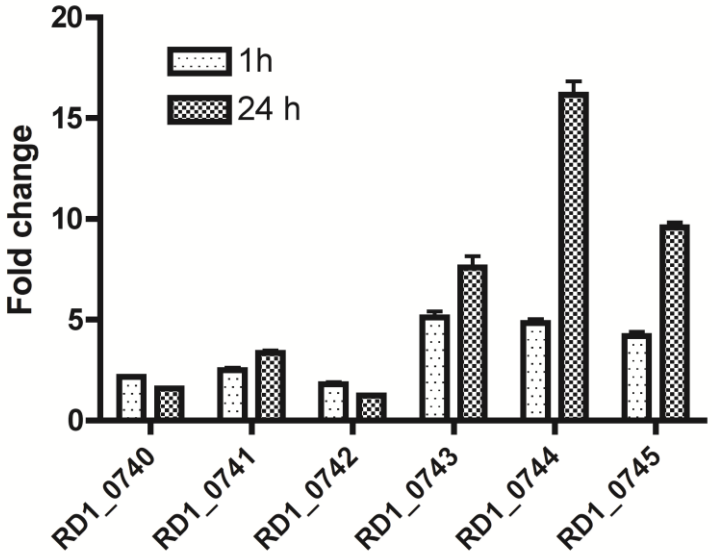
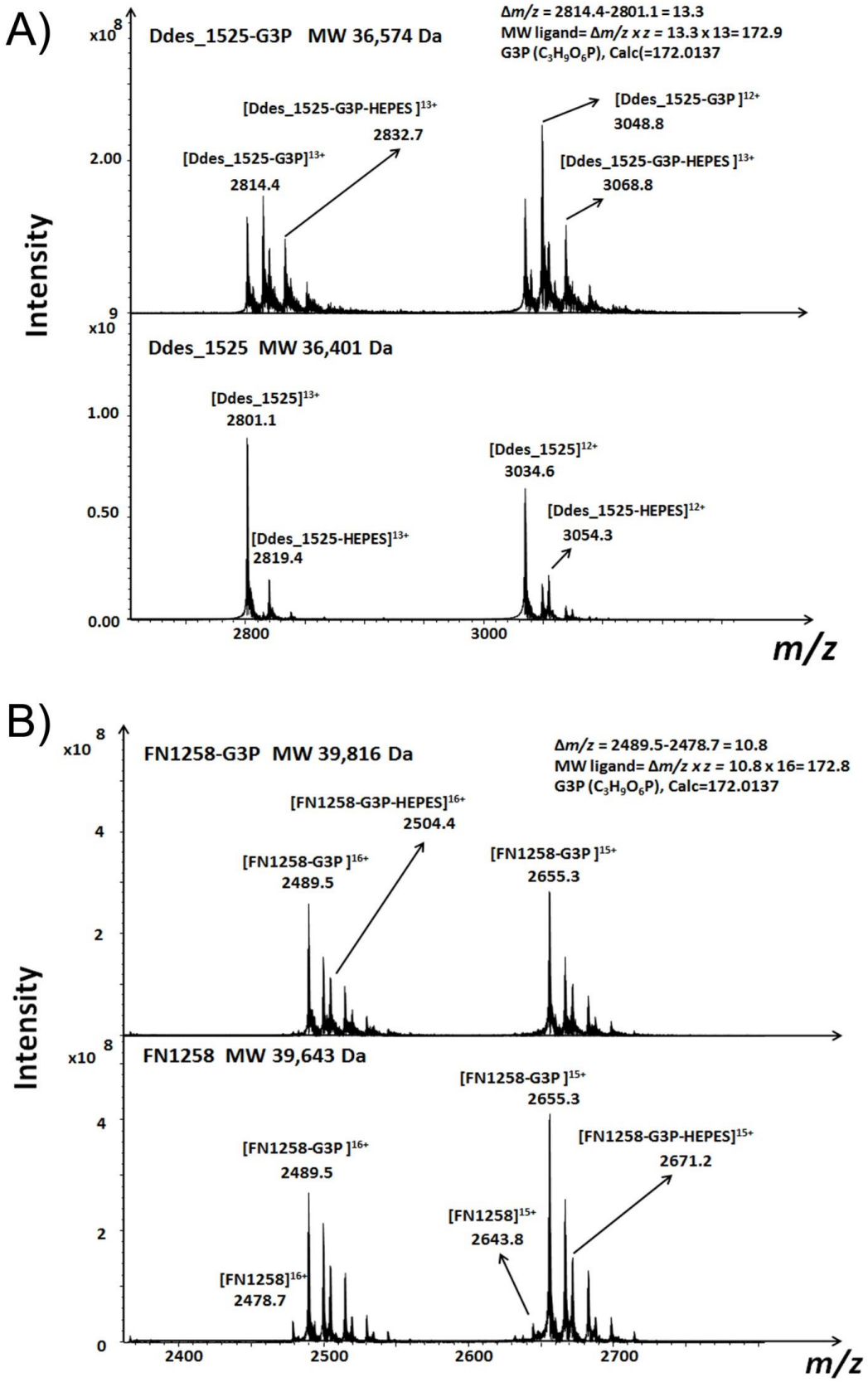
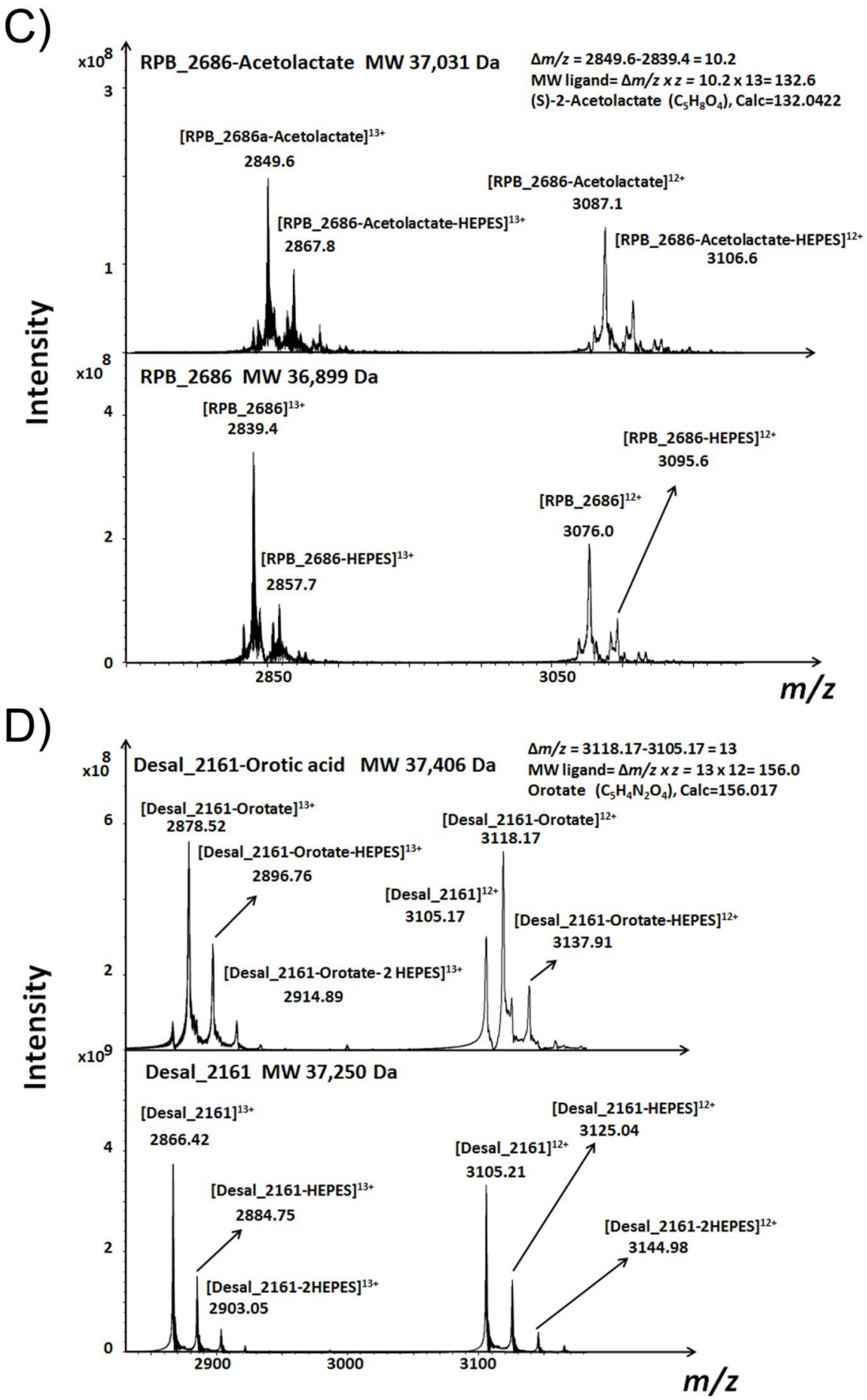


Fig: qRT-PCR for the RD1_0742 neighborhood, 1h and 24 h after inoculation. KDO vs Glucose .

Figure S4

qRT-PCR transcriptional analysis of the up-regulation (fold change in expression) of the KDO TRAP transporter (RD1_0742) and related catabolic genes during incubation of *R. denitrificans* with KDO versus glucose at 1 hr and 24 hr. While these genes are upregulated in the presence of KDO *R. denitrificans* was not able to utilize KDO as a carbon source.


A)
Ddes_1525-G3P MW 36,574 Da
Δm/z = 2814.4-2801.1 = 13.3
MW ligand= Δm/z x z = 13.3 x 13= 172.9
G3P (C3H9O6P), Calc(=172.0137
[Ddes_1525-G3P-HEPES]13+
2832.7
[Ddes_1525-G3P]13+
2814.4
[Ddes_1525-G3P]12+
3048.8
[Ddes_1525-G3P-HEPES]13+
3068.8
Intensity
Ddes_1525 MW 36,401 Da
[Ddes_1525]13+
2801.1
[Ddes_1525]12+
3034.6
[Ddes_1525-HEPES]12+
3054.3
[Ddes_1525-HEPES]13+
2819.4
m/z
B)
FN1258-G3P MW 39,816 Da
Δm/z = 2489.5-2478.7 = 10.8
MW ligand= Δm/z x z = 10.8 x 16= 172.8
G3P (C3H9O6P), Calc=172.0137
[FN1258-G3P-HEPES]16+
2504.4
[FN1258-G3P]16+
2489.5
[FN1258-G3P]15+
2655.3
Intensity
FN1258 MW 39,643 Da
[FN1258-G3P]15+
2655.3
[FN1258-G3P-HEPES]15+
2671.2
[FN1258-G3P]16+
2489.5
[FN1258]15+
2643.8
[FN1258]16+
2478.7
m/z


C)
RPB_2686-Acetolactate MW 37,031 Da
Δm/z = 2849.6-2839.4 = 10.2
MW ligand= Δm/z x z = 10.2 x 13= 132.6
(S)-2-Acetolactate (C5H8O4), Calc=132.0422
[RPB_2686a-Acetolactate]13+
2849.6
[RPB_2686-Acetolactate]12+
3087.1
[RPB_2686-Acetolactate-HEPES]13+
2867.8
[RPB_2686-Acetolactate-HEPES]12+
3106.6
RPB_2686 MW 36,899 Da
[RPB_2686]13+
2839.4
[RPB_2686-HEPES]12+
3095.6
[RPB_2686]12+
3076.0
[RPB_2686-HEPES]13+
2857.7
Intensity
m/z
D)
Desal_2161-Orotic acid MW 37,406 Da
Δm/z = 3118.17-3105.17 = 13
MW ligand= Δm/z x z = 13 x 12= 156.0
Orotate (C5H4N2O4), Calc=156.017
[Desal_2161-Orotate]13+
2878.52
[Desal_2161-Orotate]12+
3118.17
[Desal_2161-Orotate-HEPES]13+
2896.76
[Desal_2161]12+
3105.17
[Desal_2161-Orotate-HEPES]12+
3137.91
[Desal_2161-Orotate-2 HEPES]13+
2914.89
Desal_2161 MW 37,250 Da
[Desal_2161]13+
2866.42
[Desal_2161-HEPES]12+
3125.04
[Desal_2161]12+
3105.21
[Desal_2161-HEPES]13+
2884.75
[Desal_2161-2HEPES]12+
3144.98
[Desal_2161-2HEPES]13+
2903.05
Intensity
m/z

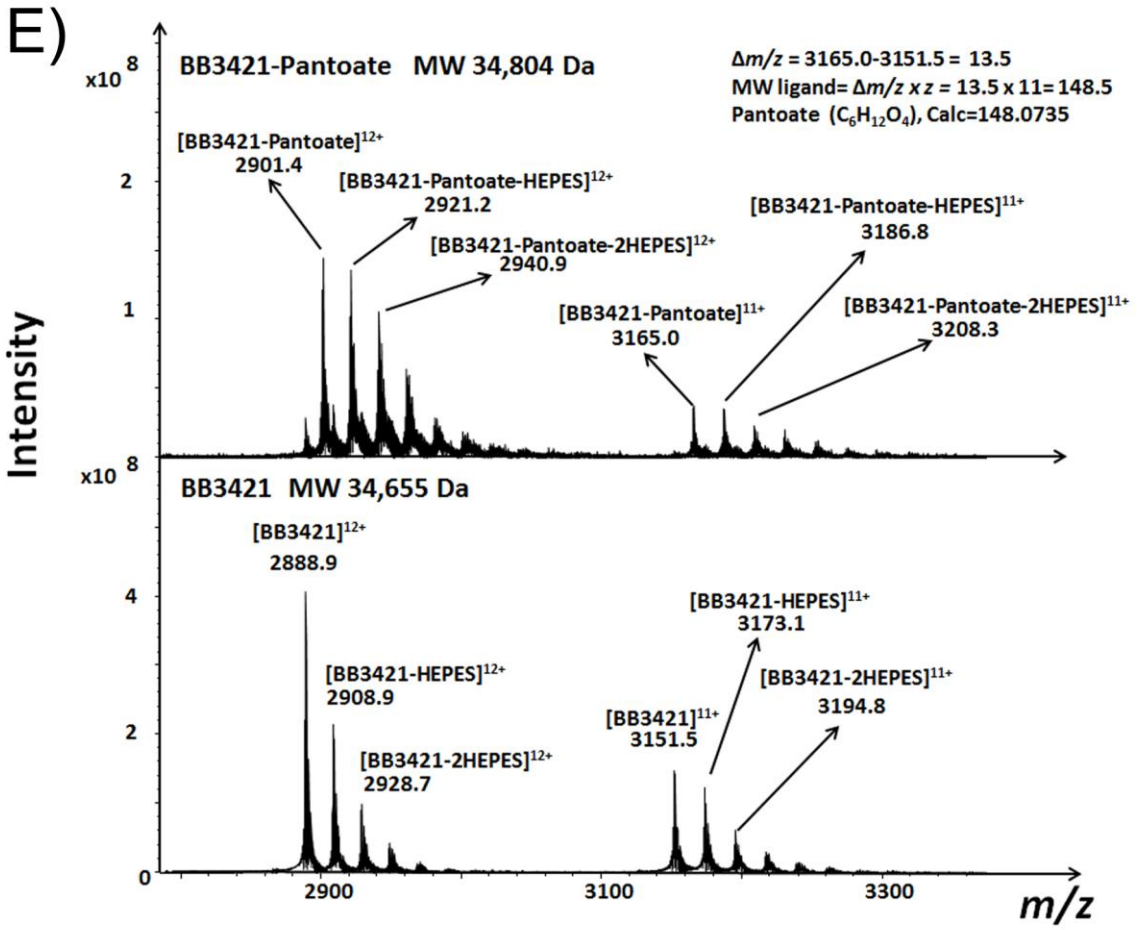


Figure S5

ESI-FTMS of Copurified Ligand TRAP SBP Complexes.

All experiments were performed on a Bruker Solarix 12 Tesla ESI-FTMS. Mass spectra were recorded in the positive ion mode from 50 to 6000 *m/z*. Each spectrum was an average of 32 scans. The Δ*m/z* for the unbound protein and the protein:ligand complex ions was multiplied by the charge state (z) to directly afford the MW of the bound ligand, using the following equation: MW ligand = Δ*m/z* x z. **A)** Ddes_1525 sn-glycerol phosphate complex. **B)** FN1258 sn-glycerol phosphate complex. **C)** RPB_2686 (S)-2-Acetolactate complex. **D)** Desal_2161 orotic acid complex. **E)** BB3421 (R)-pantoate complex.

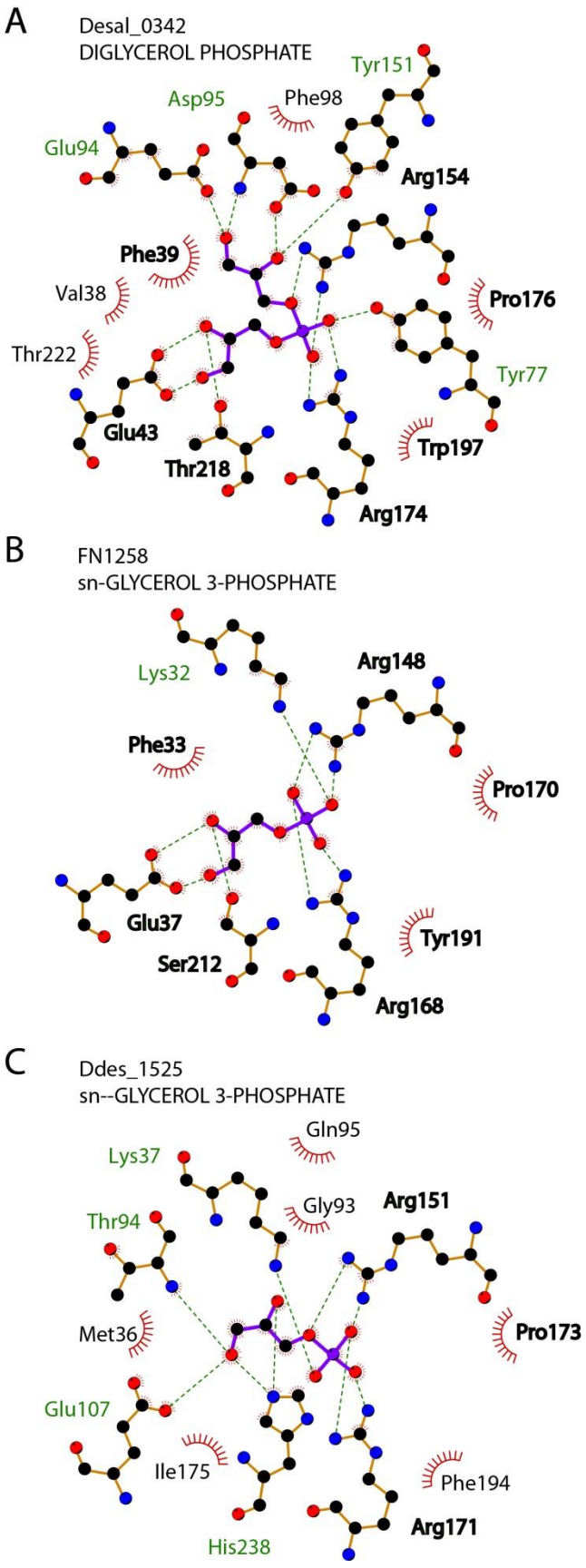


Figure S6

Interactions of TRAP SBPs with glycerol phosphates

A) Desal_0342 diglycerol phosphate complex. **B)** FN1258 sn-glycerol phosphate complex. **C)** Ddes_1525 sn-glycerol 3-phosphate complex. Residues that are similarly positioned between Desal_0342 and FN1258 are shown in BOLD. Residues of Ddes_1525 that are similarly positioned and make similar interactions to FN1258 are shown BOLD. Lys32 of FN1258 and Lys37 of Ddes_1525 do not superimpose.

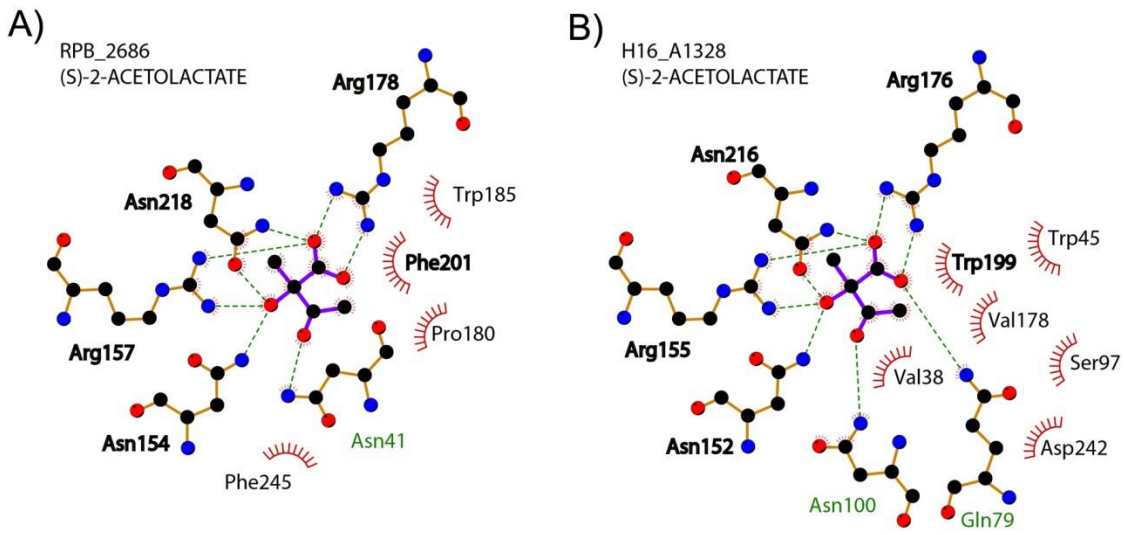


Figure S7

Interactions of TRAP SBPs with (S)-2-Acetolactate

Interactions of **A)** RBP_2686 and **B)** H16_A1328 with (S)-2-Acetolactate. Despite low homology RBP_2686 and H16_A1328 coordinate (S)-2-Acetolatate in similar fashions. Residues that are similarly positioned and make similar interactions between Desal_0342 and FN1258 are shown in BOLD. The structurally conserved interactions are two arginines (Arg178, Arg157, RPB_2686 numbering) and two asparagines (Asn154, Asn218) that coordinate the alpha-hydroxy acid group with six hydrogen bonds, while a non-positionally conserved asparagine (Asn41, RPB_2686 or Asn 100, H16_A1328) coordinates the keto group. H16_A1328 has an additional hydrogen bond between Gln79 (H16_A1328 numbering) and the carboxylate of 2-acetolactate that is not observed in RBP_2686.

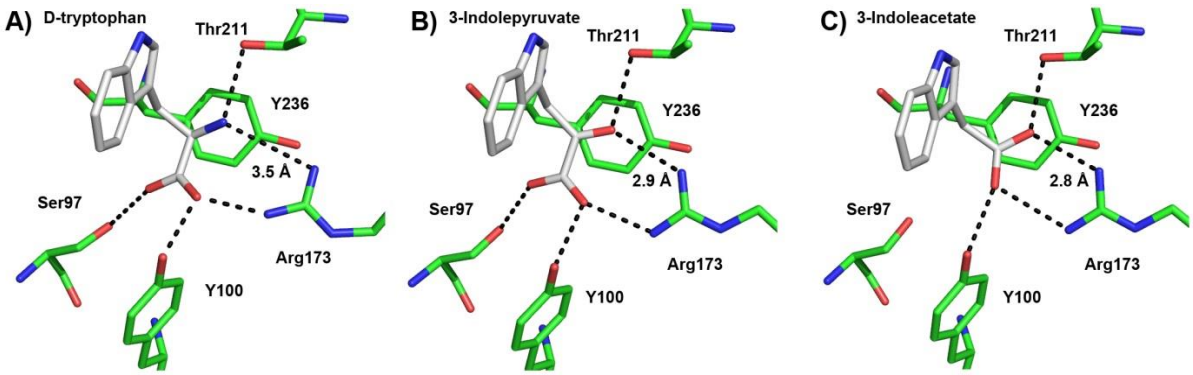


Figure S8.

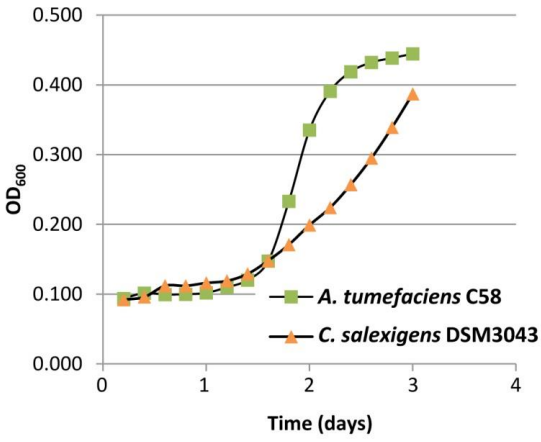
Interactions of the TRAP SBP Dde_0634 with indole acids.

Interactions of **A)** D-tryptophan **B)** 3-Indolepyruvate and **C)** 3-indole acetate with Dde_0634 from *Desulfovibrio alaskensis*. Extensive hydrophobic and polar interactions between the protein and the indole moiety are not shown for clarity. Of note, the D-Trp and 3-IPA complexes are the only TRAP-SBP structures determined in this study where the carboxylate does not participate in two hydrogen bonds with the conserved arginine. Instead the keto acid group/peptide backbone is rotated 90 degrees, with one carboxylate oxygen making a bifurcated hydrogen bond with R173NH2 and Tyr100OH and the other hydrogen bonded to Ser97OG. In contrast, the structure of the IAA complex (2.25 Å) displayed the more classic two hydrogen bonds between the carboxylate and the conserved arginine.

A)

Agrobacterium tumefaciens C58	C-source	N-source
Glucose	+++	
NH_4Cl		+++
Ethanolamine	-	++
Glycine	-	+++
L-glutamate	+++	+++

Chromohalobacter salexigens DSM3043	C-source	N-source
Glucose	+++	
NH_4Cl		+++
Ethanolamine	-	+
Glycine	-	++
L-glutamate	+++	+++



B)

Function/annotation	Locus tag	Fold change
TRAP transporter	Csal_0676 Csal_0677 Csal_0678	80-fold 60-fold 10-fold
Glutamine synthetase	Csal_0679	15-fold
Alcohol dehydrogenase	Csal_0681	50-fold
Aldehyde dehydrogenase	Csal_0680	45-fold
N-formylglutamate amidohydrolase	Csal_0675	125-fold

C)

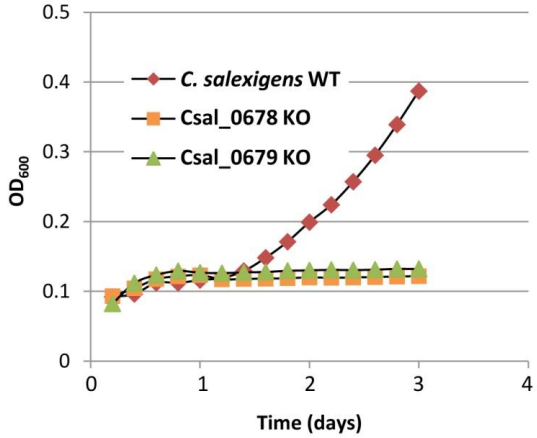


Figure S9.

Analysis of growth of *A. tumefaciens C58* and *C. salexigens* on ethanolamine.

A) Growth characteristics of *A. tumefaciens* and *C. salexigens* with NH_4Cl, ethanolamine, glycine, or L-glutamate (20 mM) as the sole source of nitrogen during aerobic growth. *A. tumefaciens* and *C. salexigens* were unable to utilize ethanolamine or glycine as a carbon source. Growth rate of *A. tumefaciens* (green squares) and *C. salexigens* (orange triangles) with ethanolamine as the sole source of nitrogen. **B)** qRT-PCR transcriptional analysis of the up-regulation (fold change in expression) of the ethanolamine TRAP transporter and related catabolic genes during growth of *C. salexigens* on ethanolamine, relative to growth with NH_4Cl, as the sole source of nitrogen. Fold changes are the averages of at least three biological replicates. **C)** Growth of C. salexigens strains with genetic disruptions in the TRAP SBP (Csal_0678 KO, orange squares) or glutamine synthetase (Csal_0679 KO, green triangles) on ethanolamine as the sole nitrogen source. *C. salexigens* WT is shown for comparison (red diamonds).

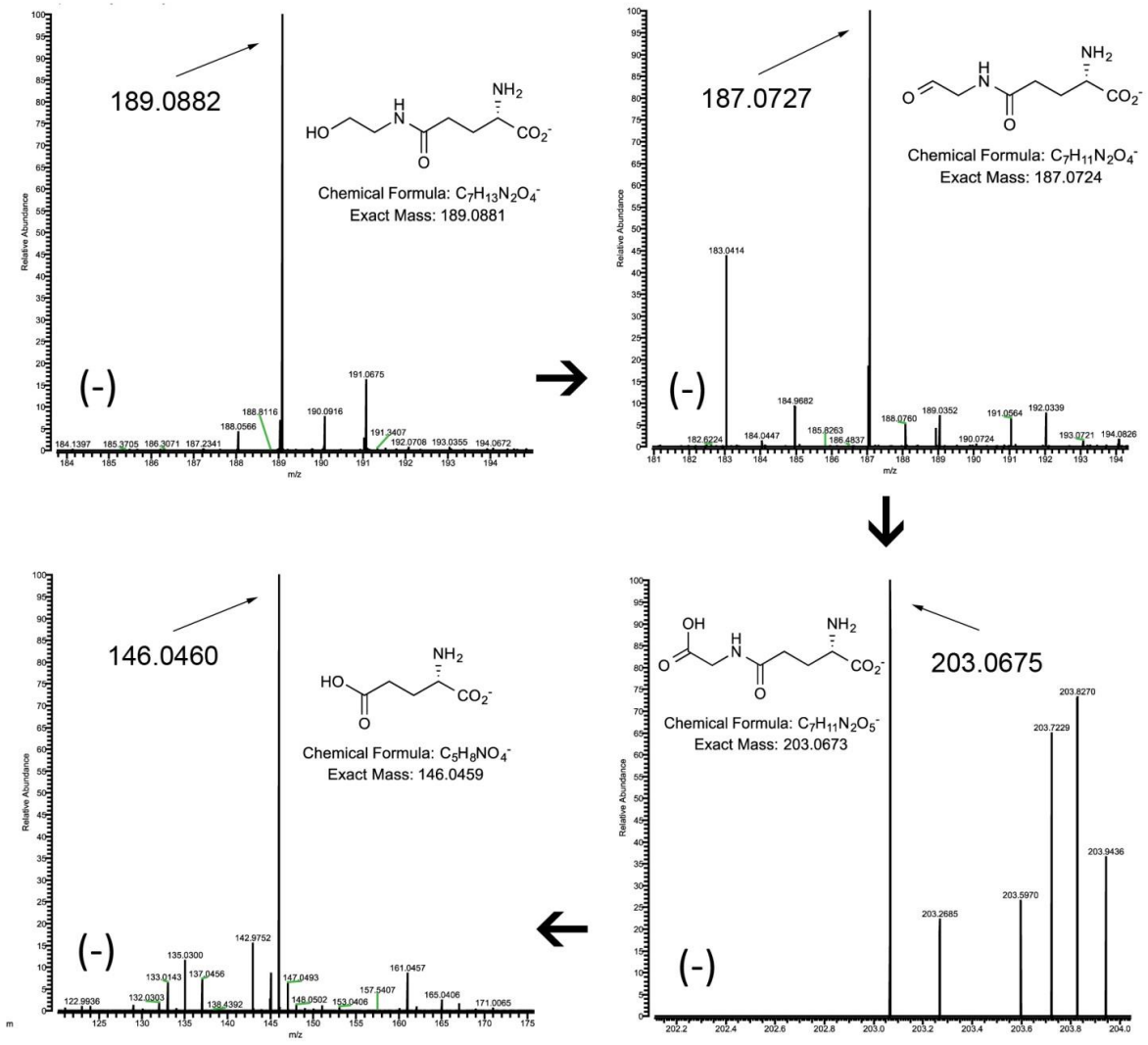


Figure S10.

LC-FTMS *m/z* peaks for ethanolamine utilization pathway intermediates and products.

Intermediates in ethanolamine metabolism detected in cell extracts of *A. tumefaciens* after feeding with 20 mM ethanolamine (shown, t=30 min), analyzed by LC-FTMS in negative (-) ion mode. The mass of each detected metabolite was within ± 0.0003 Da of the exact mass.