IECM

1st International Electronic Conference on Metabolomics
1-30 November 2016
chaired by Dr. Peter Meikle

sponsored by metabolites

Increasing arginine production in C. glutamicum by rational strain design in combination with metabolomics and proteomics

Frederik Walter[1], Marcus Persicke[1], Aiko Barsch[2]*, Stephanie Kaspar-Schönefeld[2], Heiko Neuweger[2], Nikolas Kessler[2], Jörn Kalinowski[1]

[1] Bielefeld University, Bielefeld, Germany
[2] Bruker Daltonik GmbH, Bremen, Germany

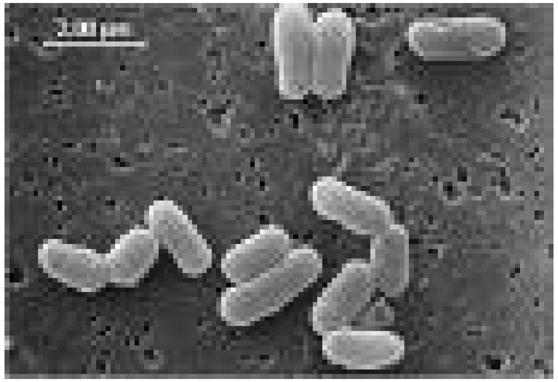
* Corresponding author: aiko.barsch@bruker.com

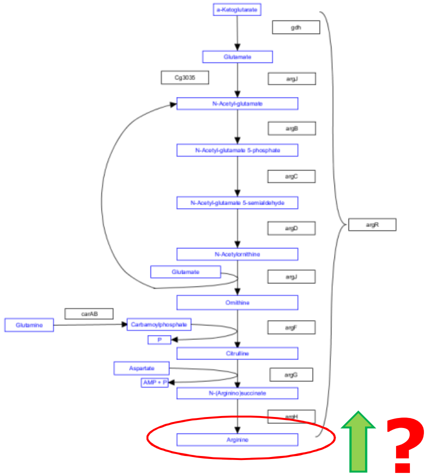
Introduction:

- *Corynebacterium glutamicum* is a biotechnological workhorse for the production of amino acids and other primary metabolites.

- Arginine is a glutamate-derived amino acid that is used in the cosmetic and pharmaceutical industries and as a food additive.

- Aim of this study:
  - To increase arginine production in the “bacterial workhorse” by rational strain design in combination with metabolomics and proteomics





Introduction:

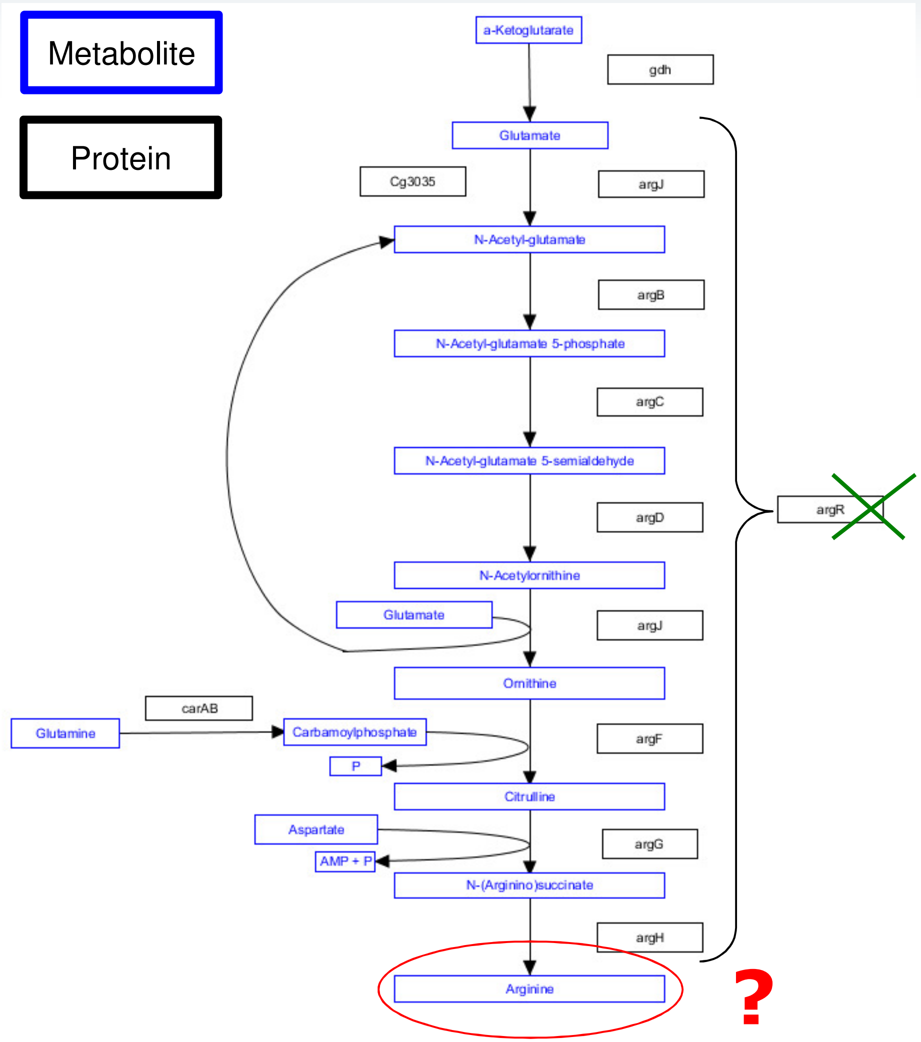
Increasing arginine production in *C. glutamicum* by rational strain design

Are Arginine levels increased if *argR* is deleted?

Wildtype **Repressor deletion**

- Arginine biosynthesis genes are regulated by the repressor *argR*, which is activated by arginine
- Repressor deletion mutant (*argR*): chromosomal deletion of *argR*

Are Arginine levels increased if *argR* is deleted?



Methods:
Label-free Proteomics & non-targeted Metabolomics

Samples

1 Wildtype and 3 mutant strains
5 biological replicates each

Wildtype

Repressor deletion

Tryptic digestion
Injection 500ng on column

Extraction of hydrophilic
metabolites with 80 % MeOH

Data acquisition

Proteomics

nanoLC	90 min - RP
impact II - QTOF	Data dependent AutoMSMS run

Metabolomics

UHPLC	15 min - ZIC HILIC
impact II - QTOF	Data dependent AutoMSMS run

Data processing

Proteomics

MaxQuant	Feature detection and database search
Perseus	Statistical processing

Metabolomics

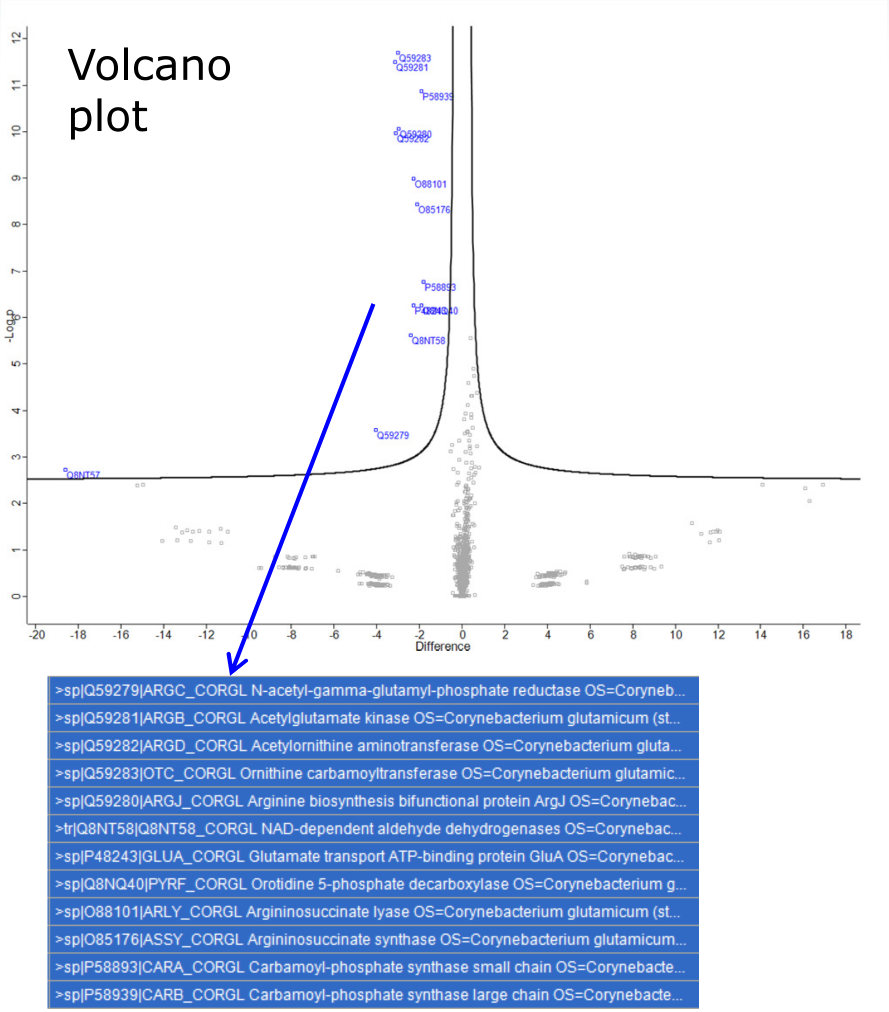
ProfileAnalysis	Feature detection and alignment
MetaboScape	Known and unknown ID, PathwayMapping

Visualization

PathVisio

Results & Discussion:

Label free proteomics: Which proteins are upregulated in the *argR* mutant compared to WT?



>sp|Q59279|ARGC_CORGL N-acetyl-gamma-glutamyl-phosphate reductase OS=Coryneb...
>sp|Q59281|ARGB_CORGL Acetylglutamate kinase OS=Corynebacterium glutamicum (st...
>sp|Q59282|ARGD_CORGL Acetylornithine aminotransferase OS=Corynebacterium gluta...
>sp|Q59283|OTC_CORGL Ornithine carbamoyltransferase OS=Corynebacterium glutamic...
>sp|Q59280|ARGJ_CORGL Arginine biosynthesis bifunctional protein ArgJ OS=Corynebac...
>tr|Q8NT58|Q8NT58_CORGL NAD-dependent aldehyde dehydrogenases OS=Corynebac...
>sp|P48243|GLUA_CORGL Glutamate transport ATP-binding protein GluA OS=Corynebac...
>sp|Q8NQ40|PYRF_CORGL Orotidine 5-phosphate decarboxylase OS=Corynebacterium g...
>sp|O88101|ARLY_CORGL Argininosuccinate lyase OS=Corynebacterium glutamicum (st...
>sp|O85176|ASSY_CORGL Argininosuccinate synthase OS=Corynebacterium glutamicum...
>sp|P58893|CARA_CORGL Carbamoyl-phosphate synthase small chain OS=Corynebacte...
>sp|P58939|CARB_CORGL Carbamoyl-phosphate synthase large chain OS=Corynebacte...

Repressor deletion versus Wildtype

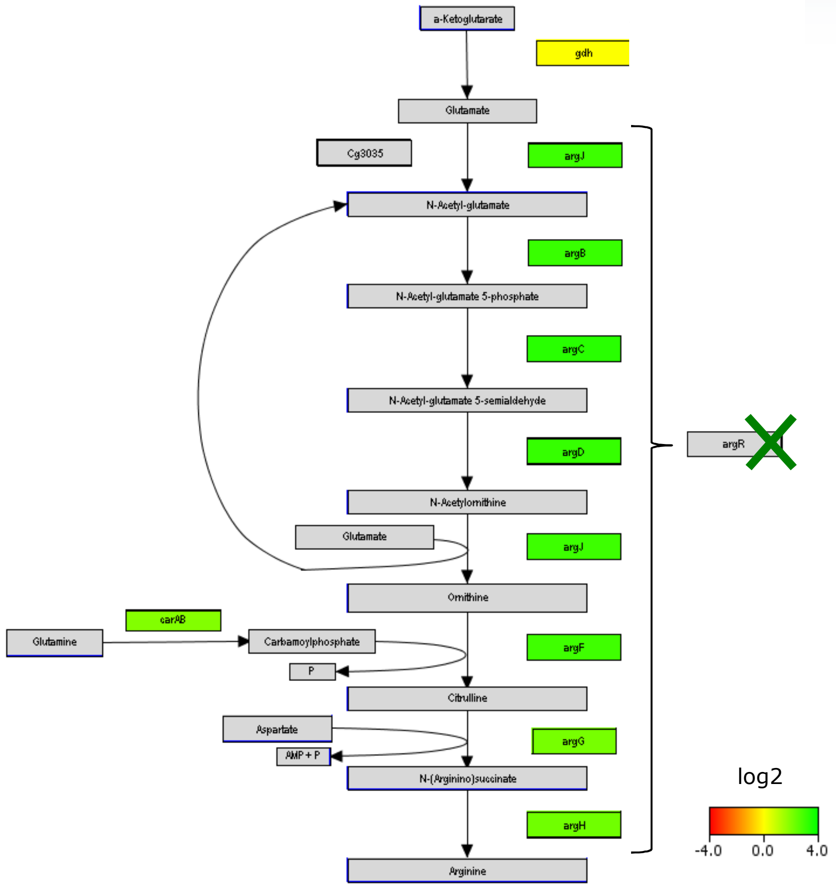
- In total **12 proteins are significantly upregulated** by derepression of arginine biosynthesis compared to wildtype
- **Arginine biosynthesis pathway** is involved in derepression of arginine biosynthesis
- **But how is the complete pathway influenced?**

⇩

Pathway mapping!

Results & Discussion:

Proteomics data clearly shows an **upregulation of the arginine pathway enzymes** in the repressor deletion mutant



Pathway designed with PathVisio

Note: Log2 fold changes have been mapped.
Green -> upregulation WT vs. *argR*
Red -> downregulation WT vs. *argR*

Repressor deletion versus Wildtype

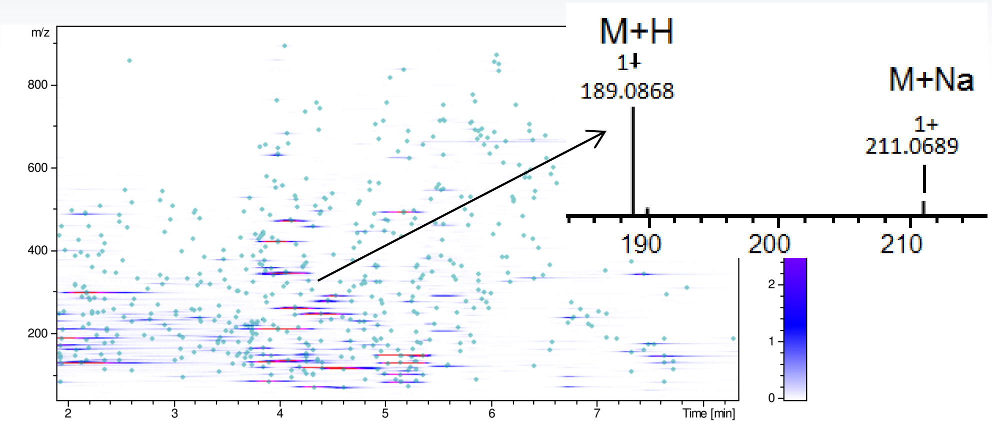
- Derepression of *argR* leads to significant increase in abundance for enzymes involved in arginine pathway (-> “green color coding”)

But do changes on proteome level really result in increased arginine production?

Metabolomics!

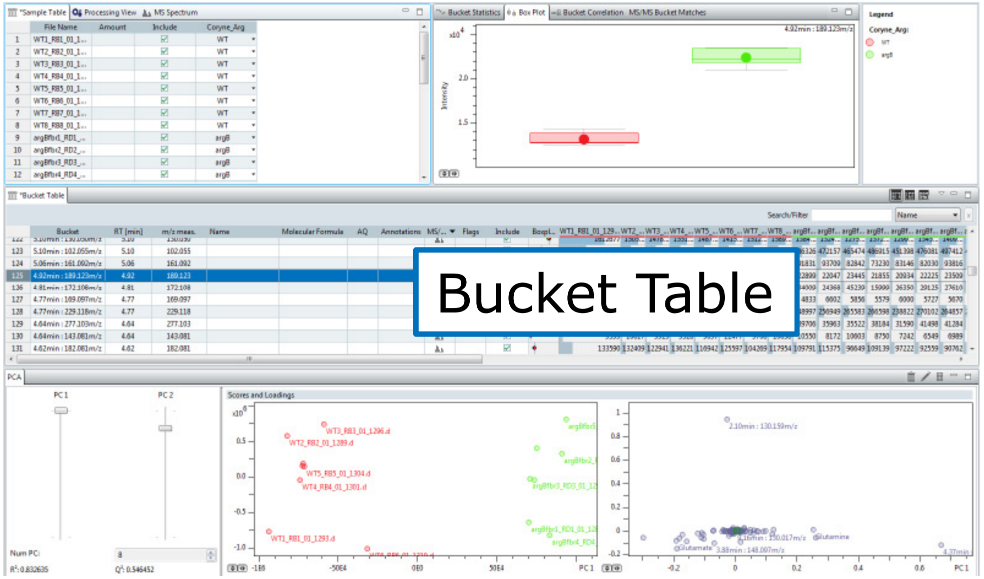
Results & Discussion:

Non-targeted **Metabolic profiling**:
Data pre-processing by novel MetaboScape 2.0



Step 1) Comprehensive feature extraction by "Find Molecular Features" algorithm in ProfileAnalysis

- *Combines adducts, charge states and isotopes* belonging to one compound
- Leads to large data reduction!



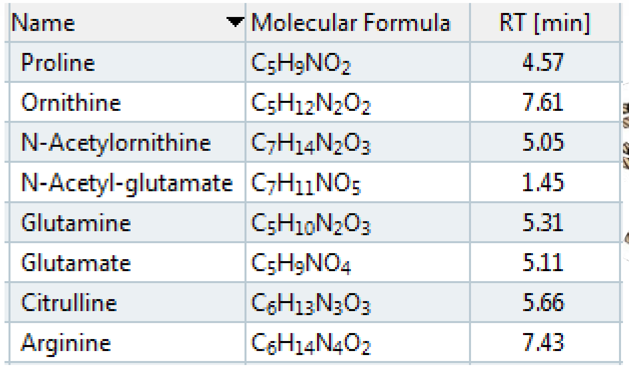
- **Step 2) Aligning** extracted "FMF" features across all samples resulted in corresponding **buckets** for further analysis in novel **MetaboScape** software

Automatic identification of known compound = dereplication

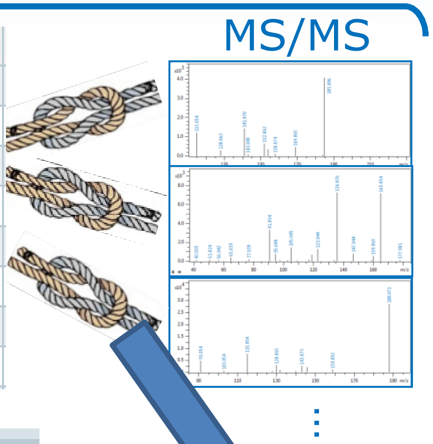
Step 3) Fast and confident dereplication by known Analyte List including MS/MS library spectra

- **Database** of know target compounds consisting of molecular formula, name and retention time (**in simple CSV format**) -> “Analyte List”
- Optionally add **MS/MS** library **spectra**
- Automatically & confidently annotate **buckets** using Analyte List

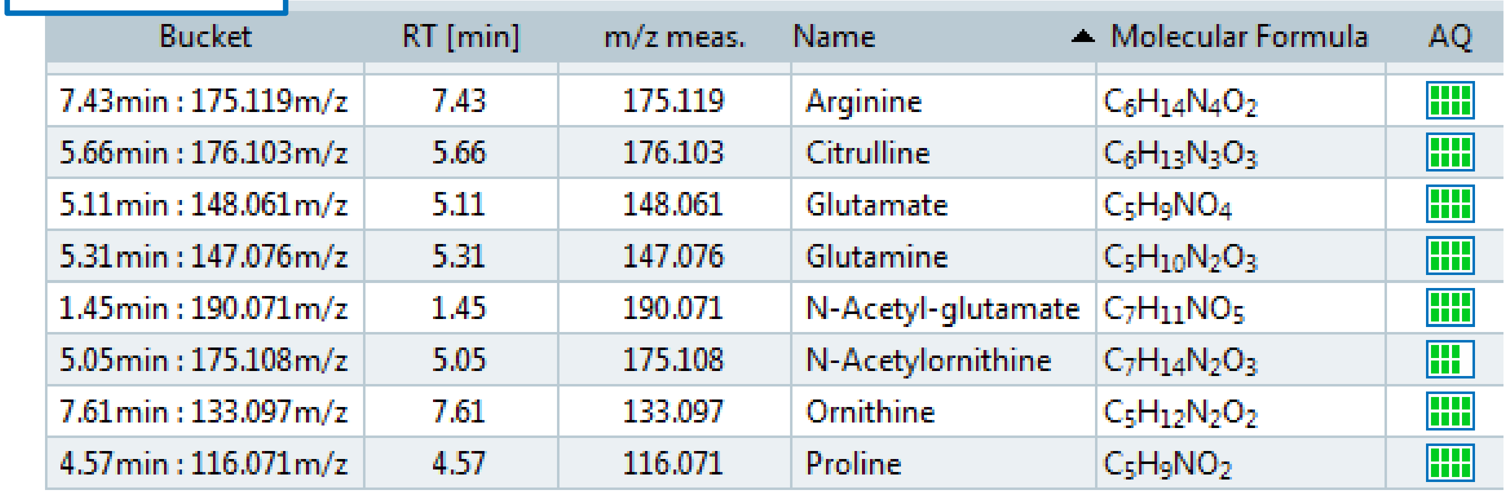
Analyte List



Name	Molecular Formula	RT [min]
Proline	$C_5H_9NO_2$	4.57
Ornithine	$C_5H_{12}N_2O_2$	7.61
N-Acetylornithine	$C_7H_{14}N_2O_3$	5.05
N-Acetyl-glutamate	$C_7H_{11}NO_5$	1.45
Glutamine	$C_5H_{10}N_2O_3$	5.31
Glutamate	$C_5H_9NO_4$	5.11
Citrulline	$C_6H_{13}N_3O_3$	5.66
Arginine	$C_6H_{14}N_4O_2$	7.43



Bucket Table

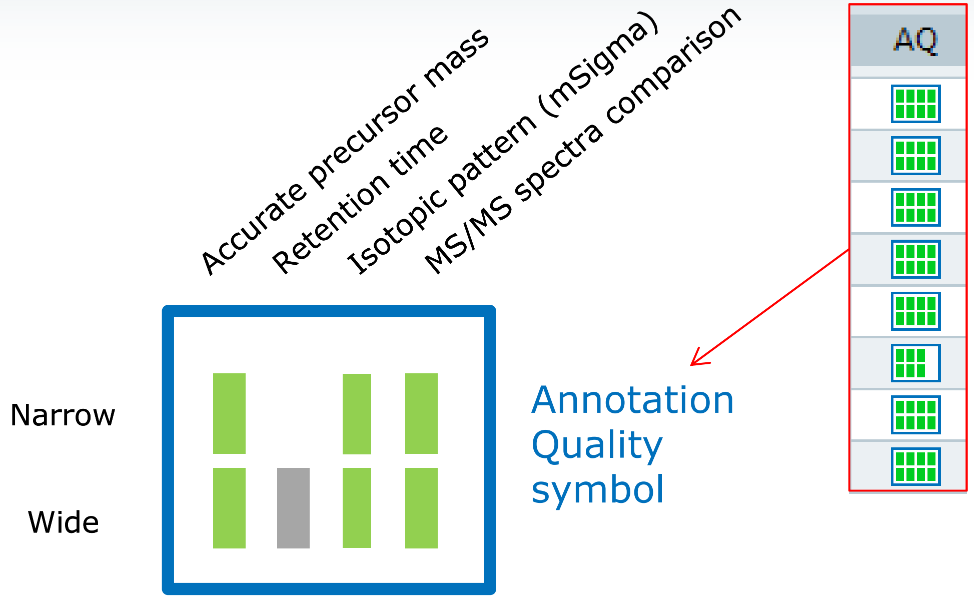


Bucket	RT [min]	m/z meas.	Name	Molecular Formula	AQ
7.43min : 175.119m/z	7.43	175.119	Arginine	$C_6H_{14}N_4O_2$	🟩🟩🟩🟩
5.66min : 176.103m/z	5.66	176.103	Citrulline	$C_6H_{13}N_3O_3$	🟩🟩🟩🟩
5.11min : 148.061m/z	5.11	148.061	Glutamate	$C_5H_9NO_4$	🟩🟩🟩🟩
5.31min : 147.076m/z	5.31	147.076	Glutamine	$C_5H_{10}N_2O_3$	🟩🟩🟩🟩
1.45min : 190.071m/z	1.45	190.071	N-Acetyl-glutamate	$C_7H_{11}NO_5$	🟩🟩🟩🟩
5.05min : 175.108m/z	5.05	175.108	N-Acetylornithine	$C_7H_{14}N_2O_3$	🟩🟩🟩☐
7.61min : 133.097m/z	7.61	133.097	Ornithine	$C_5H_{12}N_2O_2$	🟩🟩🟩🟩
4.57min : 116.071m/z	4.57	116.071	Proline	$C_5H_9NO_2$	🟩🟩🟩🟩

Annotated Bucket Table, including “AQ” **A**nnotation **Q**uality column

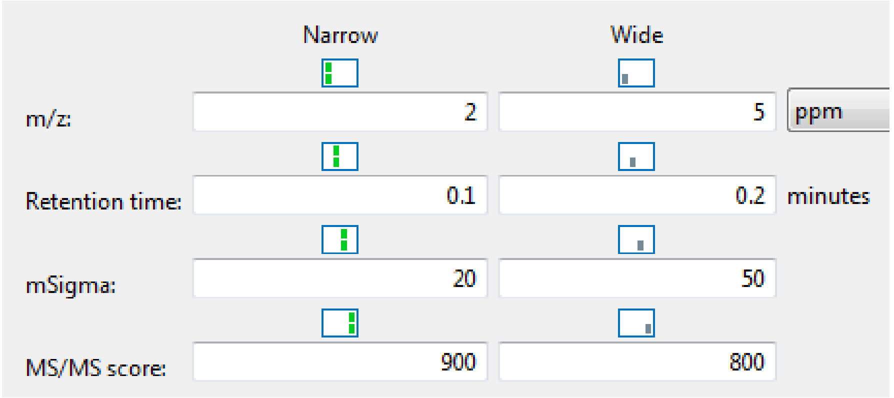
Automatic identification of known compound = dereplication

Confidence provided by “AQ - **A**nnotation **Q**uality” according to user definable criteria



Annotation Quality symbol

MetaboScape streamlines de-replication and annotated data review by highlighting the **Annotation Quality**



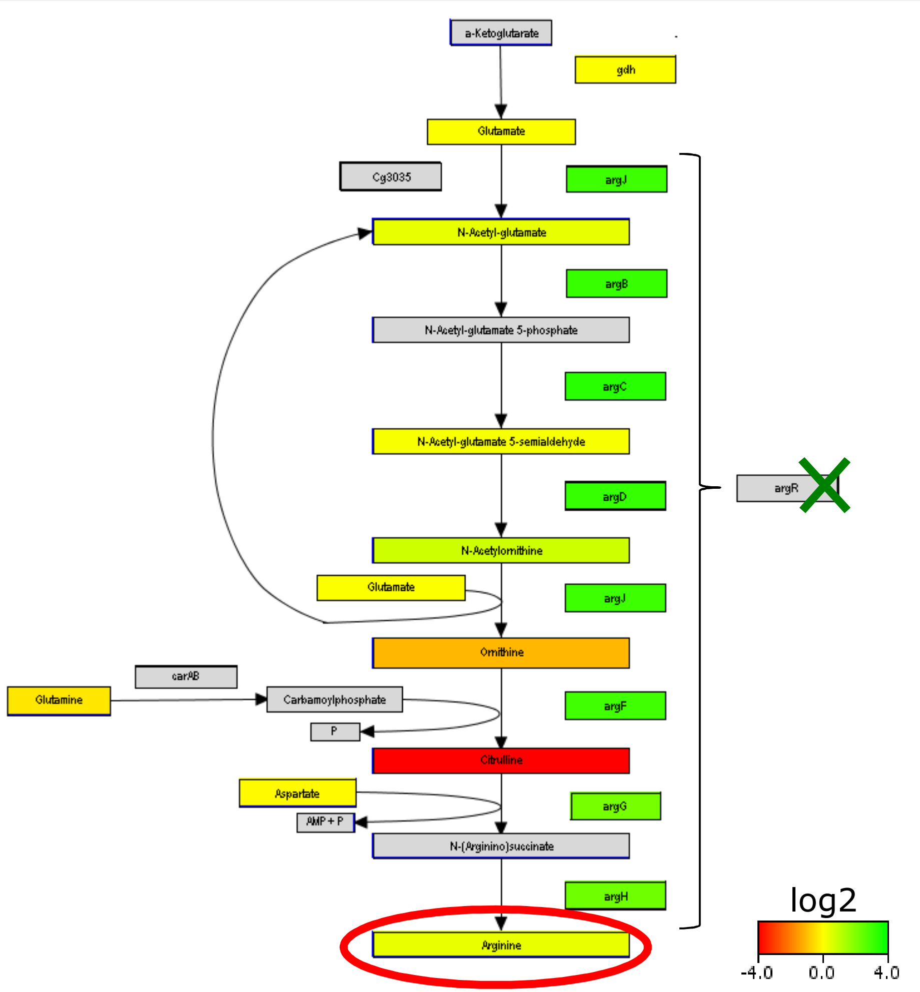
	Narrow	Wide	
m/z:	2	5	ppm
Retention time:	0.1	0.2	minutes
mSigma:	20	50	
MS/MS score:	900	800	

Annotation Quality according to **user definable confidence levels** is based on:

- Accurate precursor mass
- Retention time
- Isotopic pattern (“mSigma” value)
- MS/MS spectra comparison

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results**:** Repressor deletion (*Δ argR*) versus wild type



Pathway designed with PathVisio

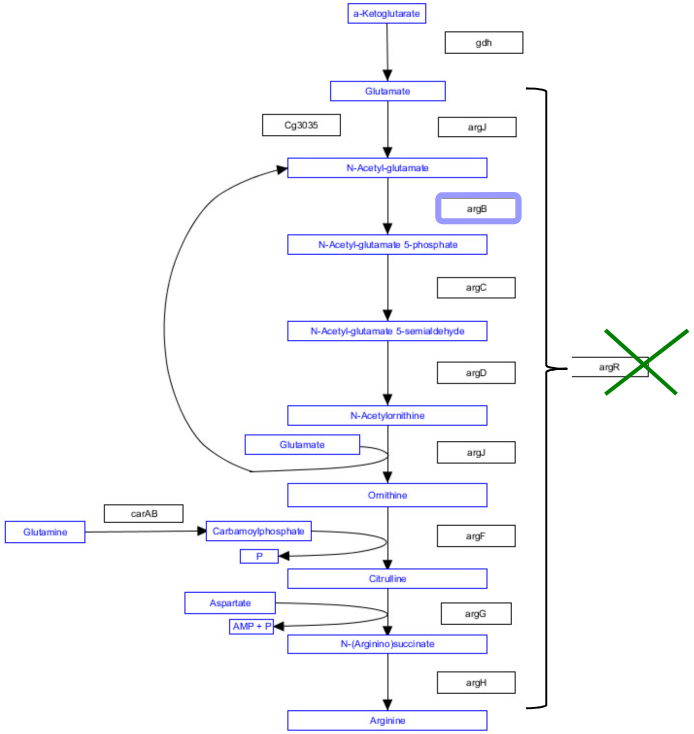
Repressor deletion vs. WT

- Increase in protein abundance for all enzymes involved in pathway
- Derepression of arginine biosynthesis genes alone **does not result in increased arginine production**
- But why?!?
  - It is known that N-acetylglutamate kinase, which is encoded by *argB*, is feedback regulated by arginine.
  - Hypothesis: Arginine production is limited by *argB* feedback-inhibition.

Increasing arginine production in *C. glutamicum* by rational strain design in combination with metabolomics and proteomics

Are Arginine levels increased if *argR* is deleted AND feedback-resistant *argB*^{fbr} allels introduced?

Wildtype **Repressor deletion** **Feedback resistant**



Rational strain design:

- Creation of a Feedback resistant mutant (*argB*): chromosomal deletion of *argR* and introduction of feedback-resistant *argB*^{fbr} allels

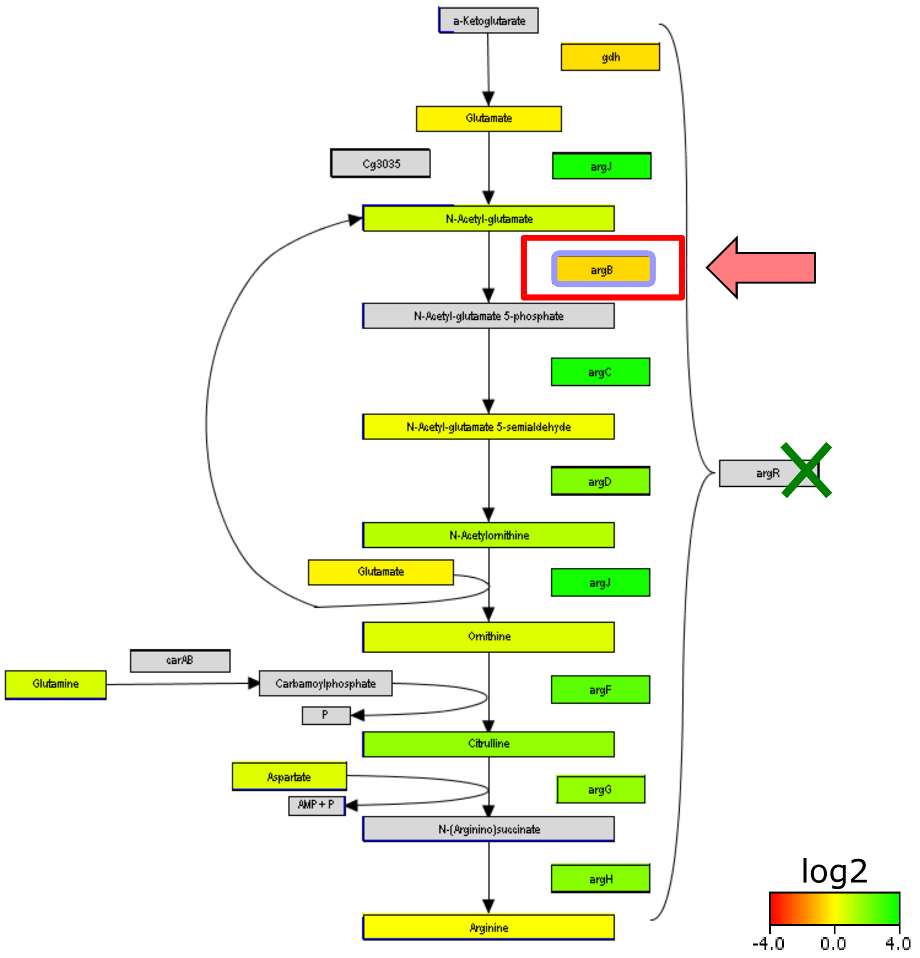
⇩

Are Arginine levels increased if *argR* is deleted AND feedback-resistant *argB*^{fbr} allels introduced?

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results:

Feedback resistant *ΔargR argB*^{fbr} vs. WT



Pathway designed with PathVisio

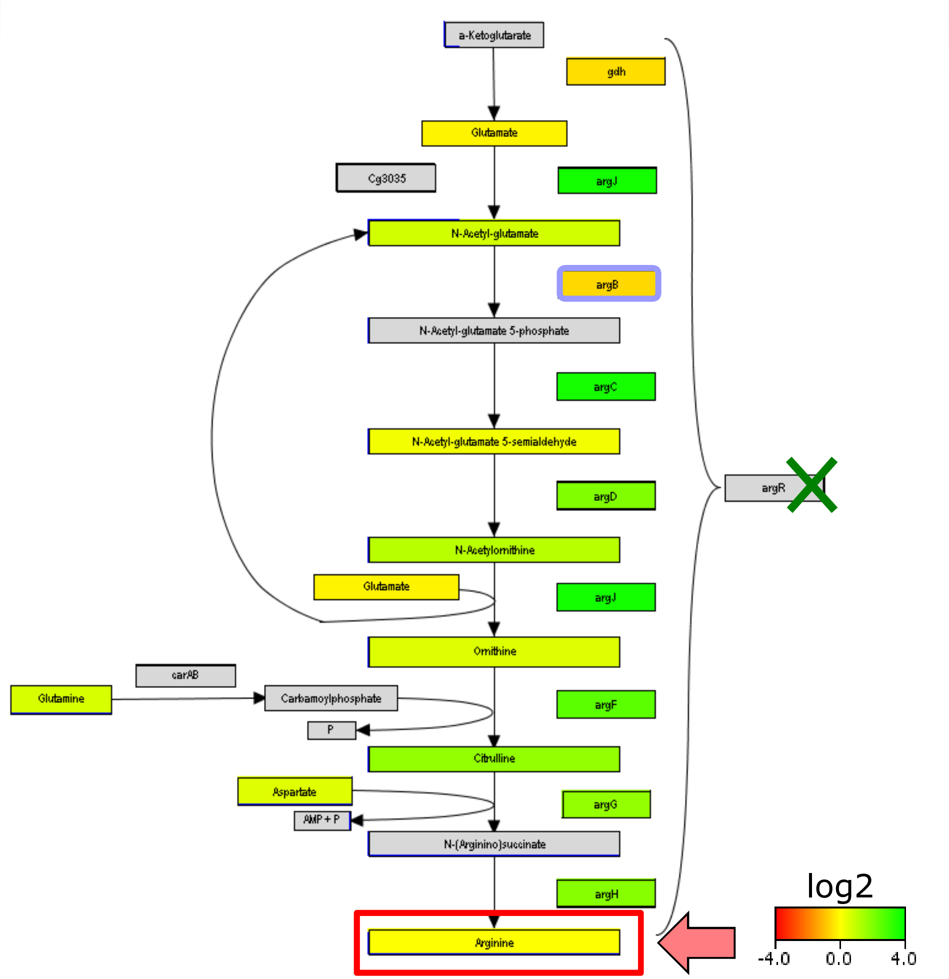
Feedback resistant *argB*-allel *ΔargR argB*^{fbr} vs. WT

- Interestingly, introduction of feedback-resistant *argB*^{fbr} allels results in reduced abundance of *argB* / N-acetylglutamate kinase

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results:

Feedback resistant *ΔargR argB*^{fbr} vs. WT



Pathway designed with PathVisio

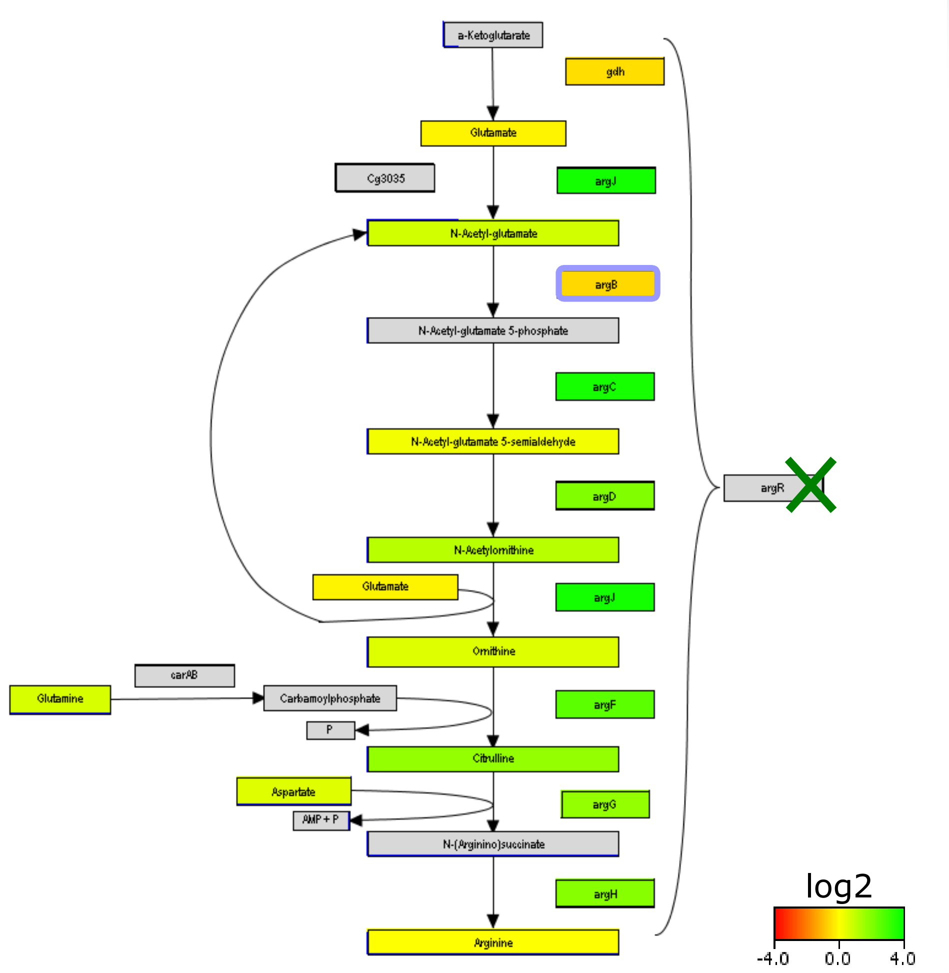
Feedback resistant *argB*-allel *ΔargR argB*^{fbr} vs. WT

- Interestingly, introduction of feedback-resistant *argB*^{fbr} allels results in reduced abundance of *argB* / N-acetylglutamate kinase
- Intracellular levels of arginine are not increased

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results:

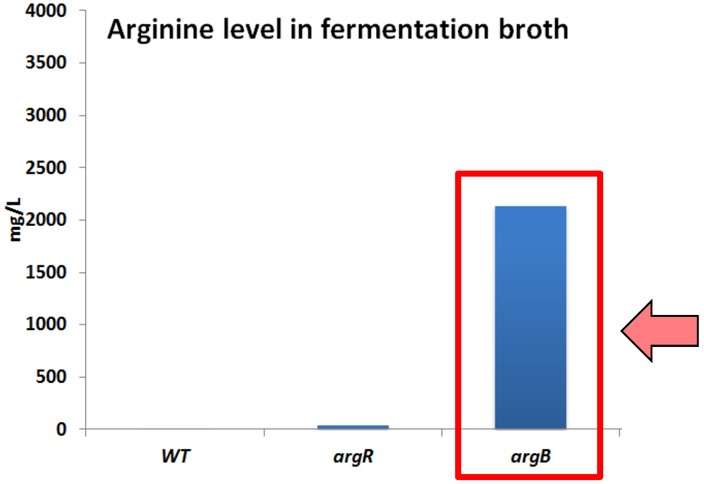
Feedback resistant *ΔargR argB*^{fbr} vs. WT



Pathway designed with PathVisio

Feedback resistant *argB*-allel *ΔargR argB*^{fbr} vs. WT

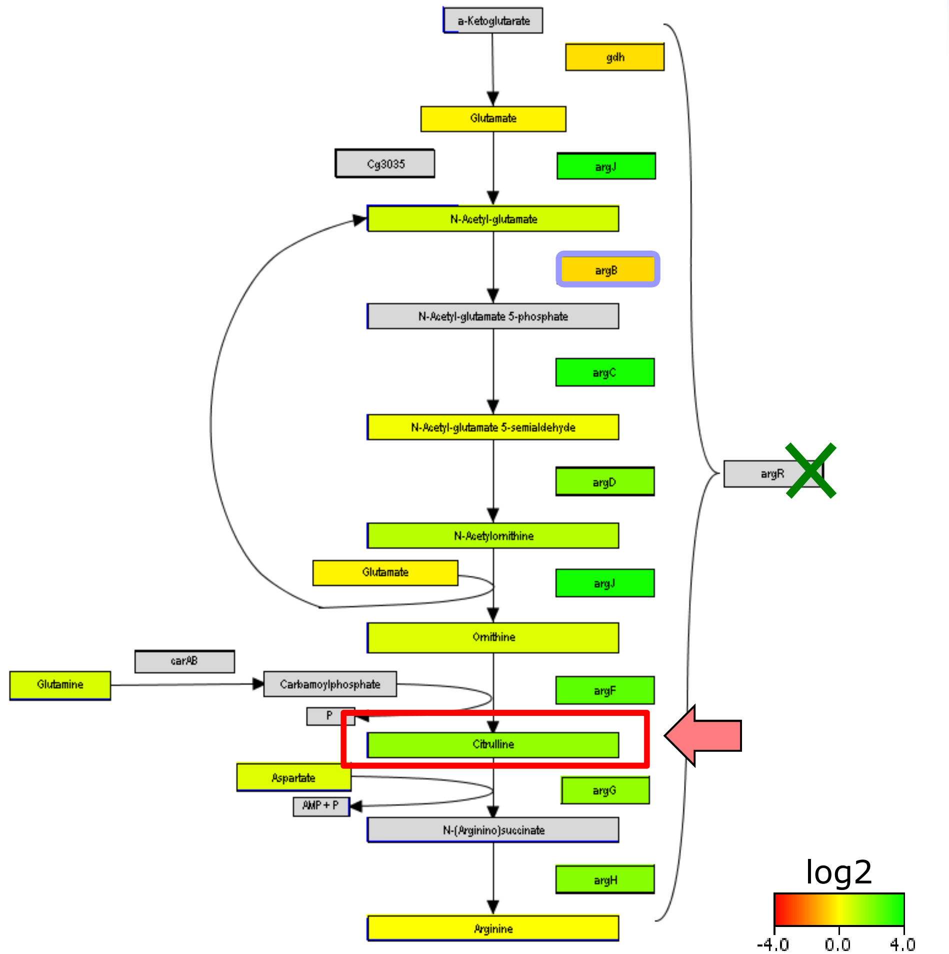
- Interestingly, introduction of feedback-resistant *argB*^{fbr} allels results in reduced abundance of *argB* / N-acetylglutamate kinase
- Intracellular levels of arginine are not increased
- BUT: Arginine is produced and secreted to the fermentation broth:



Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results:

Feedback resistant *ΔargR argB*^{fbr} vs. WT

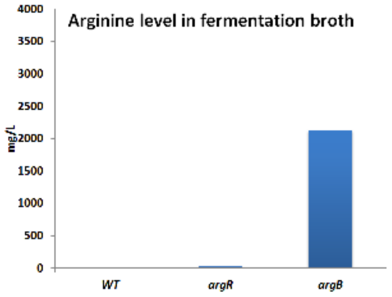


Pathway designed with PathVisio

Feedback resistant *argB*-allel *ΔargR argB*^{fbr} vs. WT

- Interestingly, introduction of feedback-resistant *argB*^{fbr} allels results in reduced abundance of *argB* / N-acetylglutamate kinase
- Intracellular levels of arginine are not increased

- BUT: Arginine is produced and secreted to the fermentation broth:



- Increase of intracellular citrulline level indicates a limitation in the last two steps of the pathway.

Increasing arginine production in *C. glutamicum* by rational strain design using a combination of metabolomics and proteomics

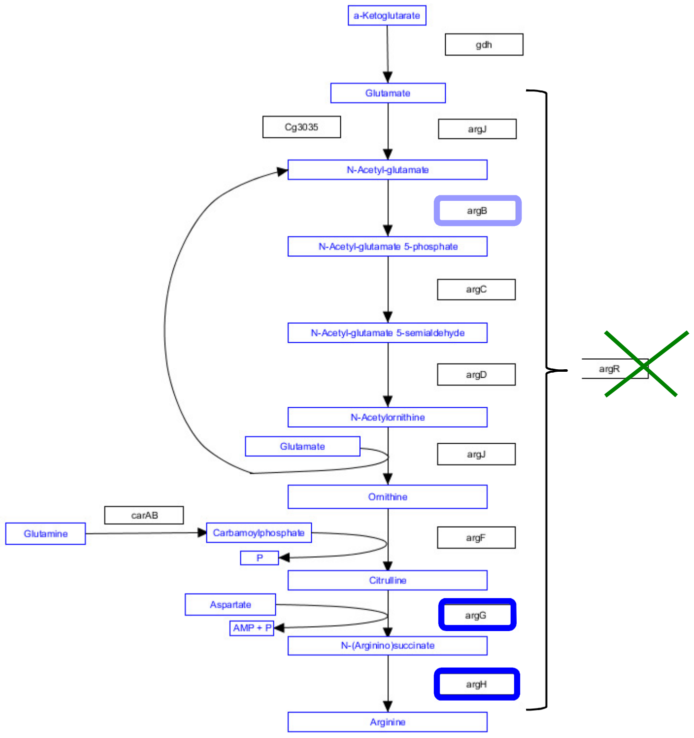
Are Arginine levels increased if *argGH are overexpressed?*

Wildtype

Repressor deletion

Feedback resistant

Debottlenecking by overexpression



Rational strain design:

- Creation of a Debottlenecking by overexpression mutant (*argGH*): chromosomal deletion of *argR* and introduction of feedback-resistant *argB*^{fbr} allels as well as debottlenecking of the last two reactions by overexpression of *argGH*

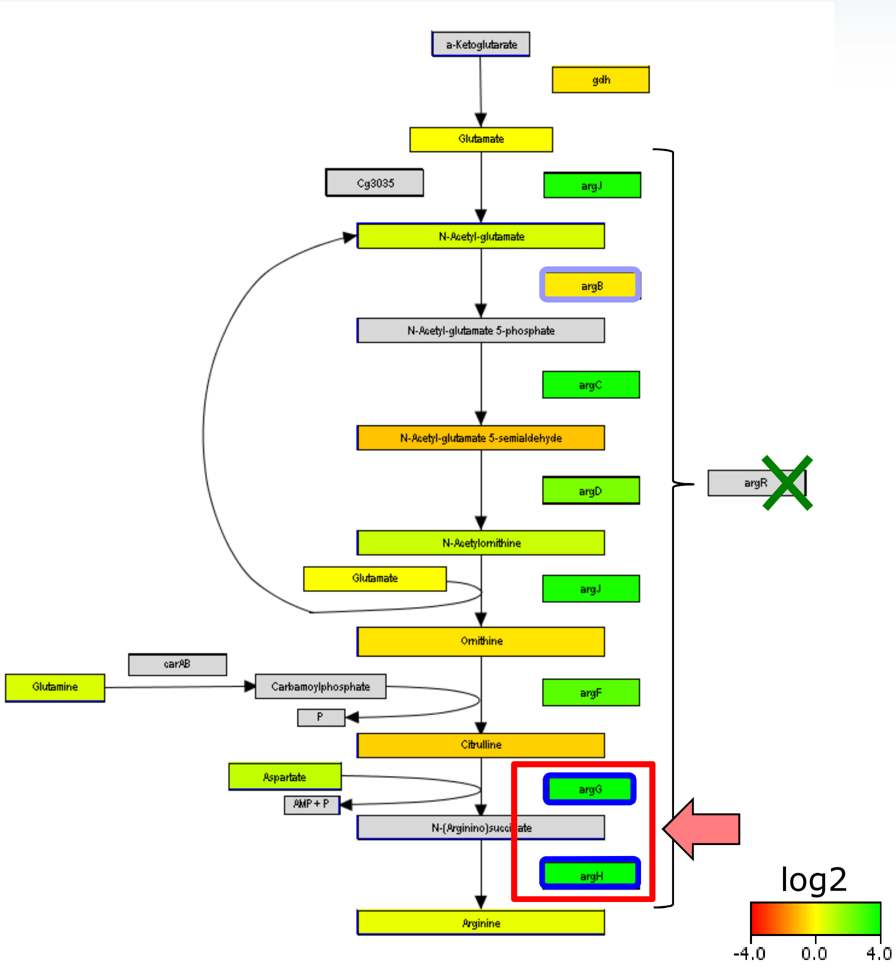
⇩

Are Arginine levels increased in this triple mutant?

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results: Debottlenecking by overexpression

ΔargR argB^{fbr} **pZ8-1::*argGH* vs. WT**



Pathway designed with PathVisio

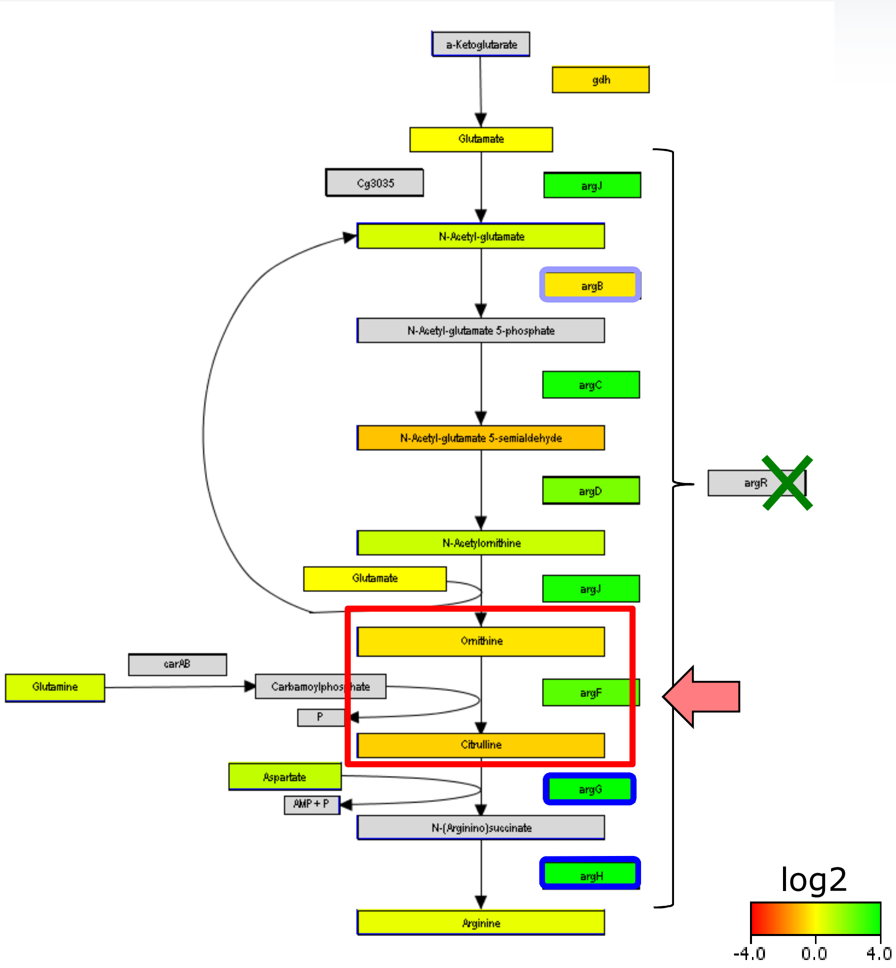
Debottlenecking by overexpression

- Overexpression of *argGH* results in increased abundance of argininosuccinat synthetase and argininosuccinat lyase on proteome level

Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results: Debottlenecking by overexpression

***ΔargR argB^{fbr}* pZ8-1::*argGH* vs. WT**

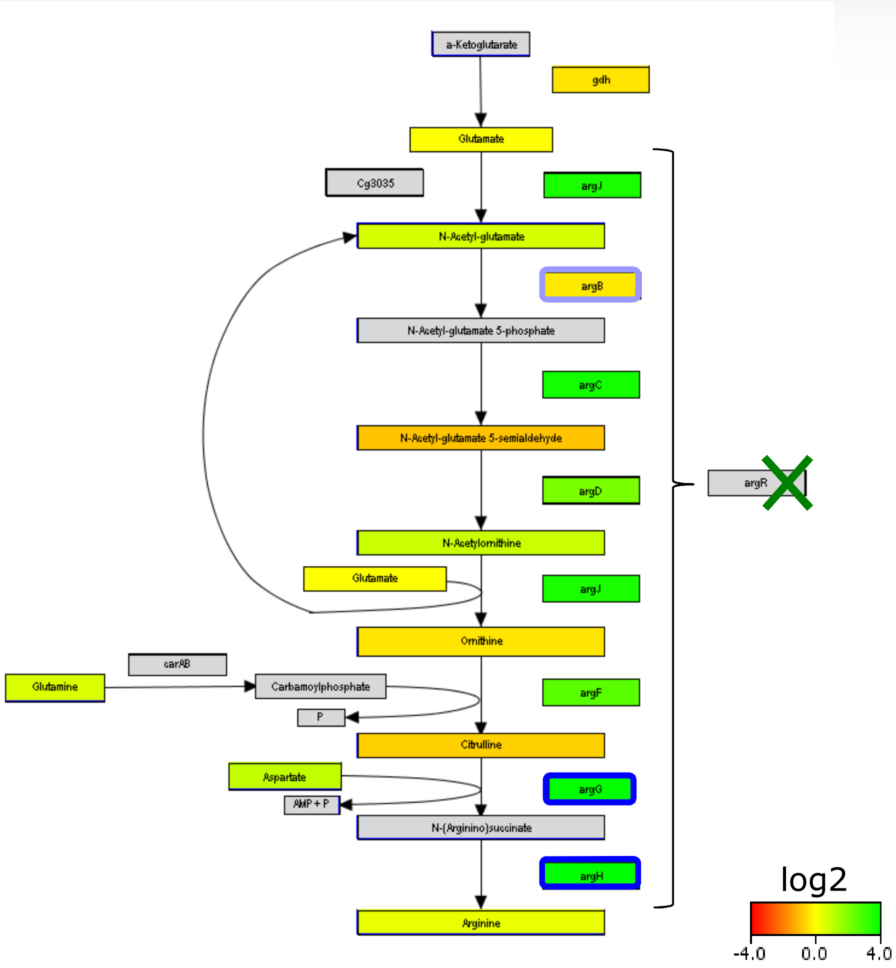


Pathway designed with PathVisio

Debottlenecking by overexpression

- Overexpression of *argGH* results in increased abundance of argininosuccinat synthetase and argininosuccinat lyase on proteome level
- Debottlenecking of last two reactions lowers intracellular concentrations of ornithine and citrulline.

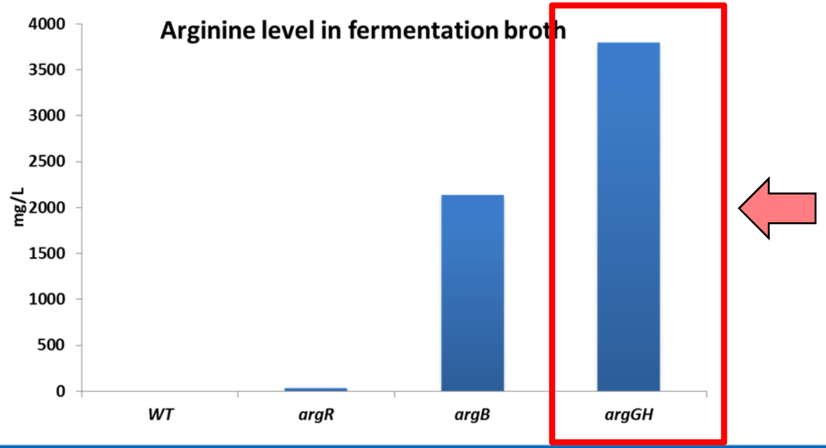
Results & Discussion:
Pathway mapping of **Proteomics and Metabolomics** results: Debottlenecking by overexpression
***ΔargR argB*^{fbr} pZ8-1::*argGH* vs. WT**



Pathway designed with PathVisio

Debottlenecking by overexpression

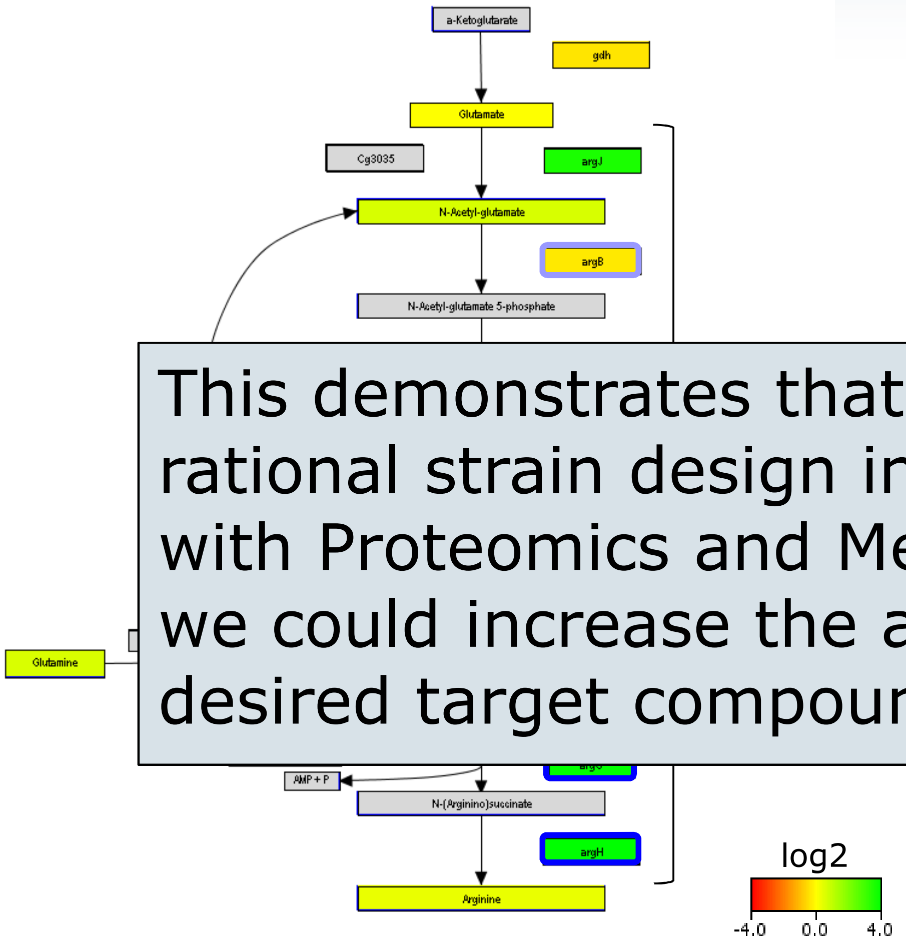
- Overexpression of *argGH* results in increased abundance of argininosuccinat synthetase and argininosuccinat lyase on proteome level
- Debottlenecking of last two reactions lowers intracellular concentrations of ornithine and citrulline.
- Overexpression of *argGH* results in increased extracellular arginine levels to >3.5 g /l.



Results & Discussion:

Pathway mapping of **Proteomics and Metabolomics** results: Debottlenecking by overexpression

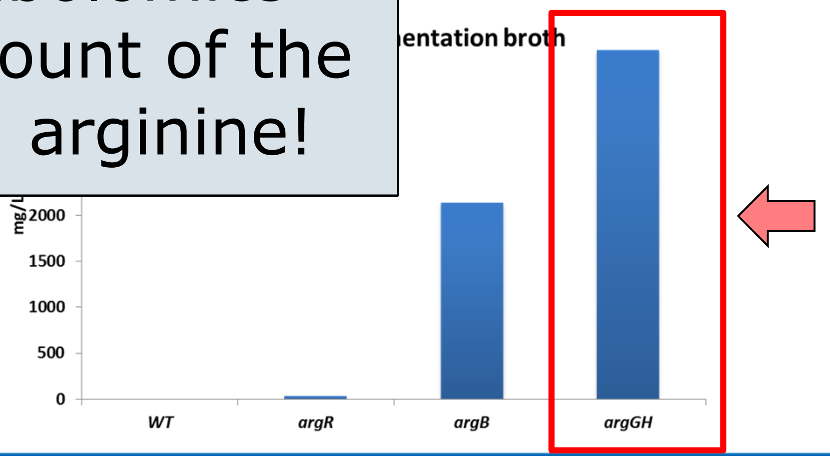
ΔargR argB^{fbr} **pZ8-1::*argGH* vs. WT**



Pathway designed with PathVisio

Debottlenecking by overexpression

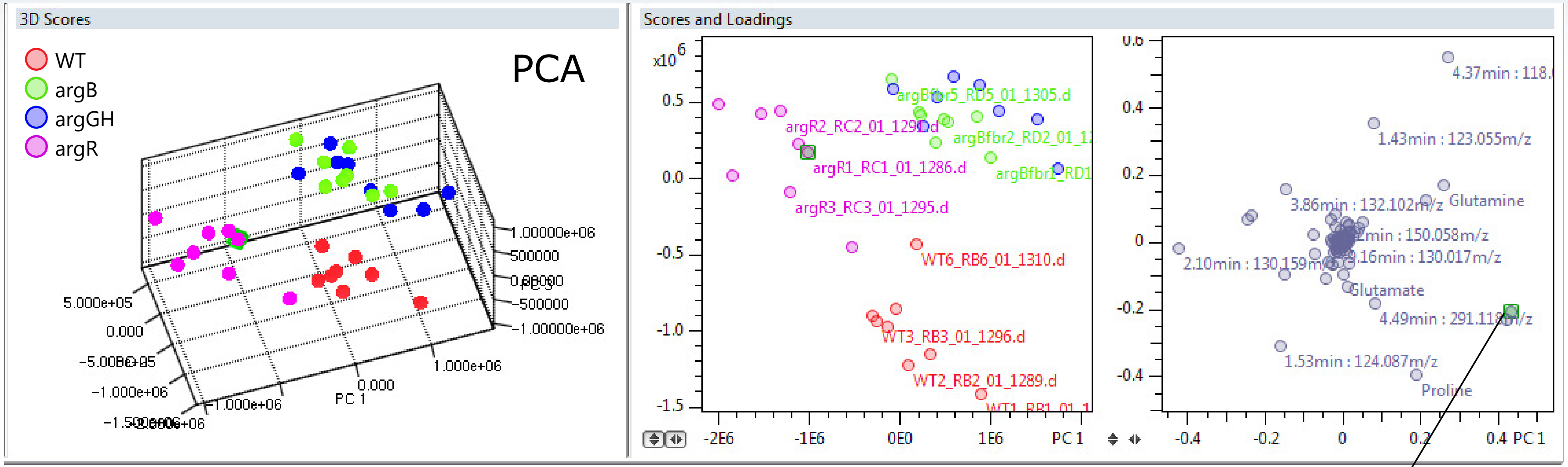
- Overexpression of *argGH* results in increased abundance of argininosuccinat synthetase and argininosuccinat lyase on proteome level
- Debottlenecking of last two reactions
  oncentrations of
  e.
  gGH results in
  ar arginine levels



This demonstrates that by applying rational strain design in combination with Proteomics and Metabolomics we could increase the amount of the desired target compound, arginine!

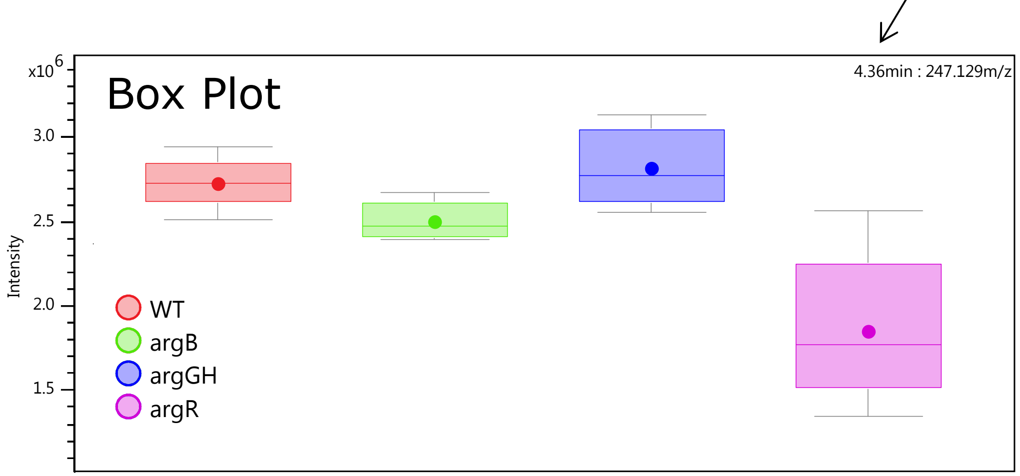
Results & Discussion:

Non-targeted Metabolomics data mining
Are there any “off target” changes we can detect in the metabolomics data?



PCA statistical analysis points to unknown compound lower abundant in *argR* mutant:

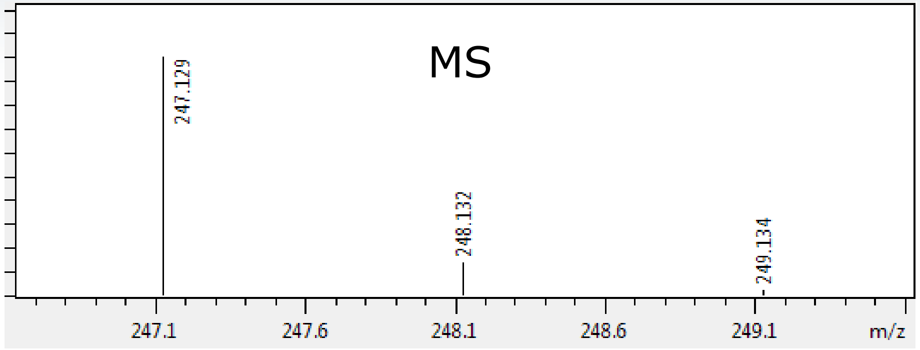
m/z 247.129
RT: 4.36 min

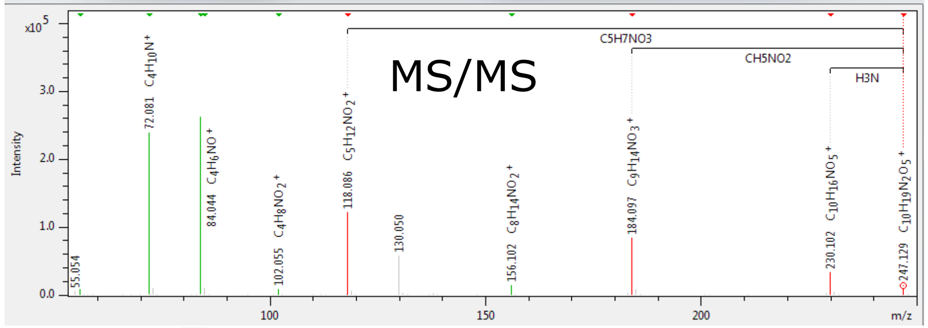


Results & Discussion:

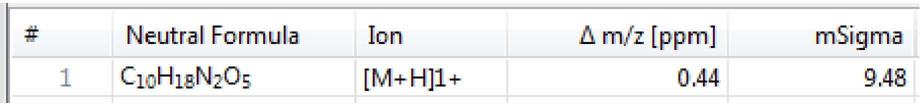
Identification of the target compound

Molecular Formula -> online Database query -> structure candidates -> *in-silico* fragmentation: gamma-Glu-Val





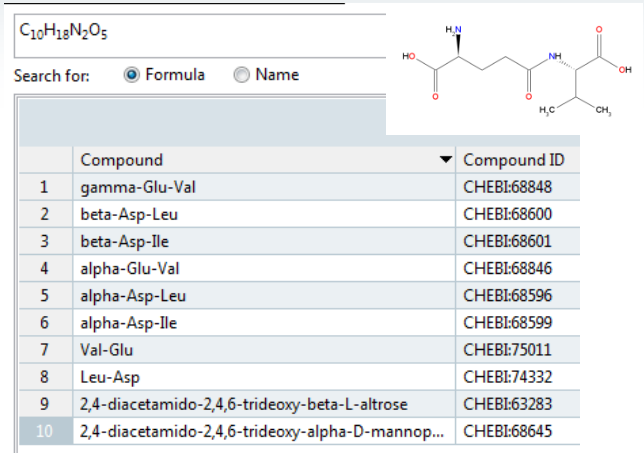
Formula generation



#	Neutral Formula	Ion	Δ m/z [ppm]	mSigma
1	$C_{10}H_{18}N_2O_5$	[M+H]1+	0.44	9.48

Unique formula generated by making use of accurate mass and isotopic pattern information in MS and MS/MS spectra.

Query for structure candidates



$C_{10}H_{18}N_2O_5$

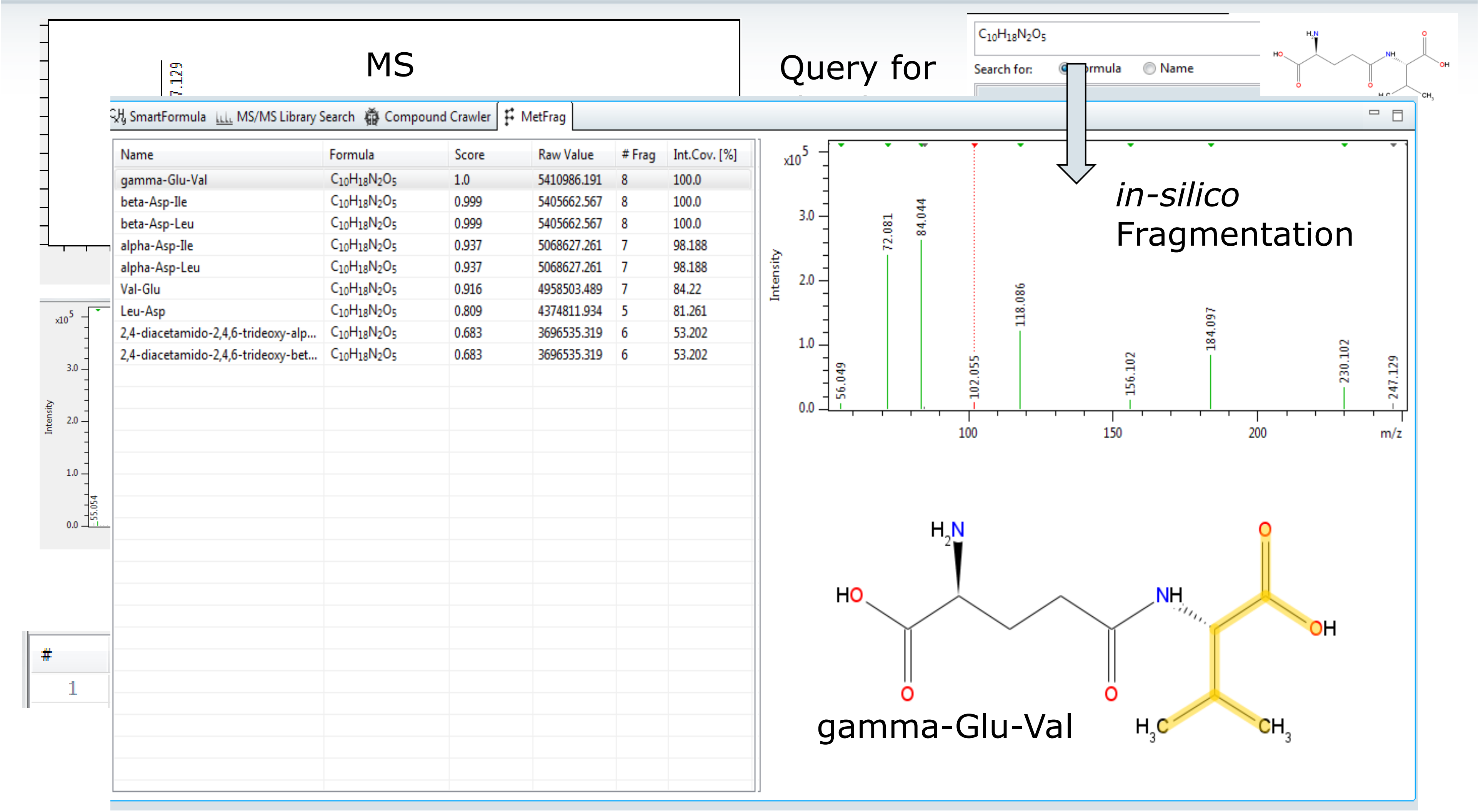
Search for: ● Formula ○ Name

	Compound	Compound ID
1	gamma-Glu-Val	CHEBI:68848
2	beta-Asp-Leu	CHEBI:68600
3	beta-Asp-Ile	CHEBI:68601
4	alpha-Glu-Val	CHEBI:68846
5	alpha-Asp-Leu	CHEBI:68596
6	alpha-Asp-Ile	CHEBI:68599
7	Val-Glu	CHEBI:75011
8	Leu-Asp	CHEBI:74332
9	2,4-diacetamido-2,4,6-trideoxy-beta-L-altrose	CHEBI:63283
10	2,4-diacetamido-2,4,6-trideoxy-alpha-D-mannop...	CHEBI:68645

Results & Discussion:

Identification of the target compound

Molecular Formula -> online Database query -> structure candidates -> *in-silico* fragmentation: gamma-Glu-Val



MS

Query for

$C_{10}H_{18}N_2O_5$

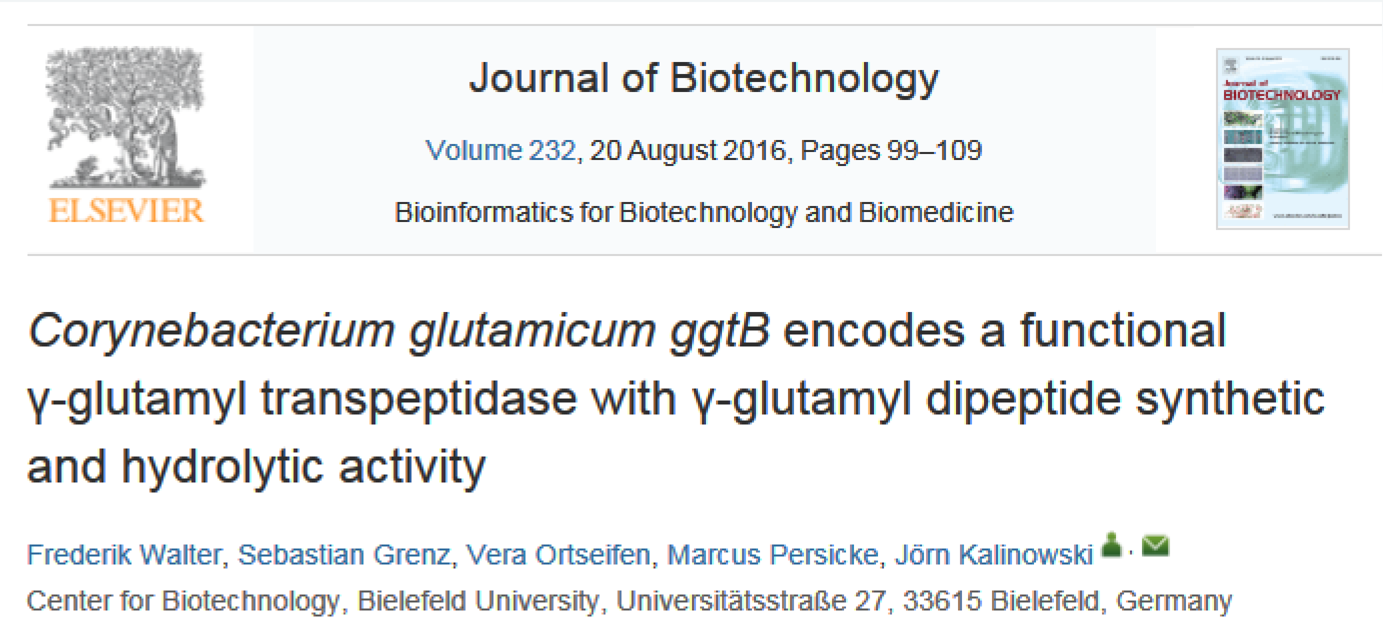
Search for: ● ...rmula ○ Name

Name	Formula	Score	Raw Value	# Frag	Int.Cov. [%]
gamma-Glu-Val	$C_{10}H_{18}N_2O_5$	1.0	5410986.191	8	100.0
beta-Asp-Ile	$C_{10}H_{18}N_2O_5$	0.999	5405662.567	8	100.0
beta-Asp-Leu	$C_{10}H_{18}N_2O_5$	0.999	5405662.567	8	100.0
alpha-Asp-Ile	$C_{10}H_{18}N_2O_5$	0.937	5068627.261	7	98.188
alpha-Asp-Leu	$C_{10}H_{18}N_2O_5$	0.937	5068627.261	7	98.188
Val-Glu	$C_{10}H_{18}N_2O_5$	0.916	4958503.489	7	84.22
Leu-Asp	$C_{10}H_{18}N_2O_5$	0.809	4374811.934	5	81.261
2,4-diacetamido-2,4,6-trideoxy-alp...	$C_{10}H_{18}N_2O_5$	0.683	3696535.319	6	53.202
2,4-diacetamido-2,4,6-trideoxy-bet...	$C_{10}H_{18}N_2O_5$	0.683	3696535.319	6	53.202

in-silico Fragmentation

gamma-Glu-Val

Results & Discussion:
gamma-glutamyl dipetides are known to be present in *C. glutamicum*



Journal of Biotechnology

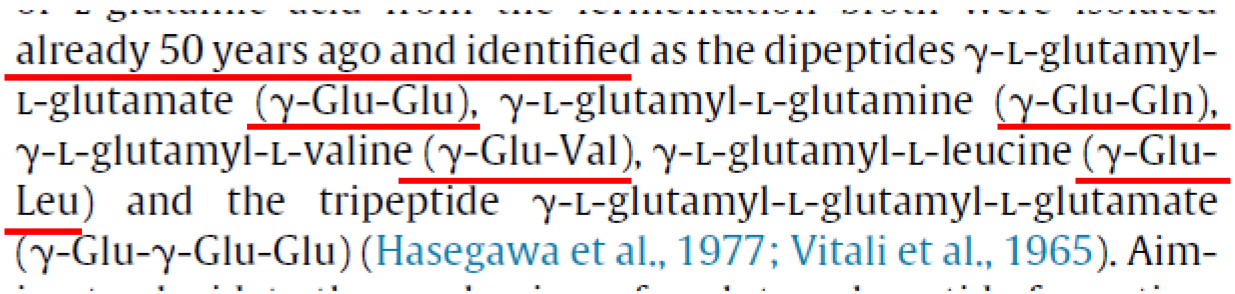
Volume 232, 20 August 2016, Pages 99–109

Bioinformatics for Biotechnology and Biomedicine

Corynebacterium glutamicum ggtB encodes a functional γ-glutamyl transpeptidase with γ-glutamyl dipeptide synthetic and hydrolytic activity

Frederik Walter, Sebastian Grenz, Vera Ortseifen, Marcus Persicke, Jörn Kalinowski

Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany



already 50 years ago and identified as the dipeptides γ-L-glutamyl-L-glutamate (γ-Glu-Glu), γ-L-glutamyl-L-glutamine (γ-Glu-Gln), γ-L-glutamyl-L-valine (γ-Glu-Val), γ-L-glutamyl-L-leucine (γ-Glu-Leu) and the tripeptide γ-L-glutamyl-L-glutamyl-L-glutamate (γ-Glu-γ-Glu-Glu) (Hasegawa et al., 1977; Vitali et al., 1965). Aim-

Results & Discussion:

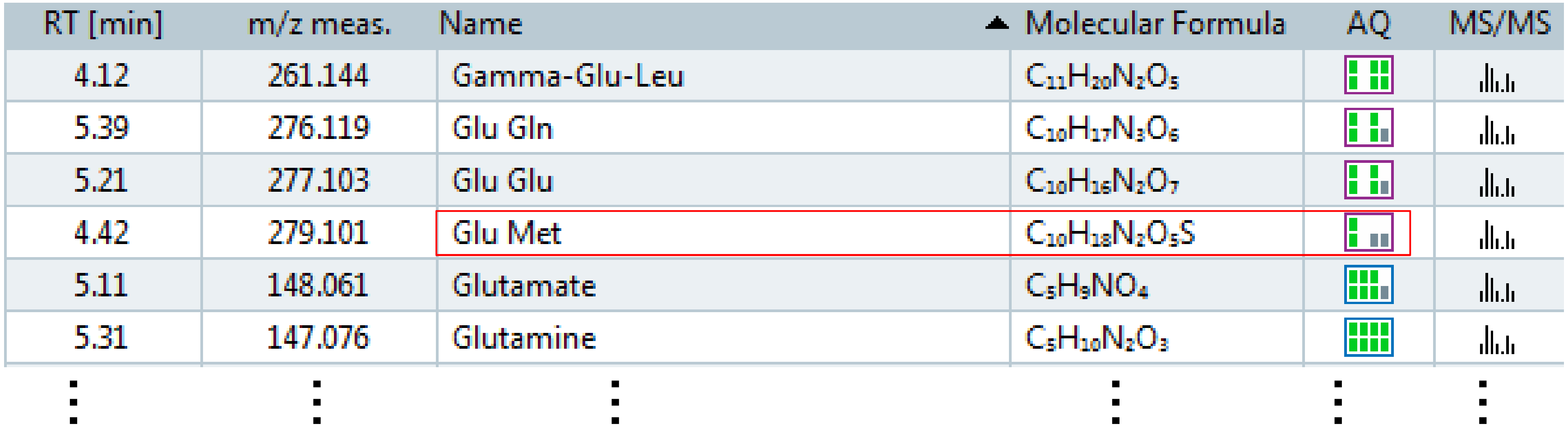
Can we find further di-peptides in the data set?

Automatic query of all buckets containing MS/MS spectra against Spectral libraries:

- HMDB Metabolite Library
- MetaboBASE Personal Library

HMDB Metabolite Library
BRUKER Daltonics

MetaboBASE Personal Library
BRUKER Daltonics

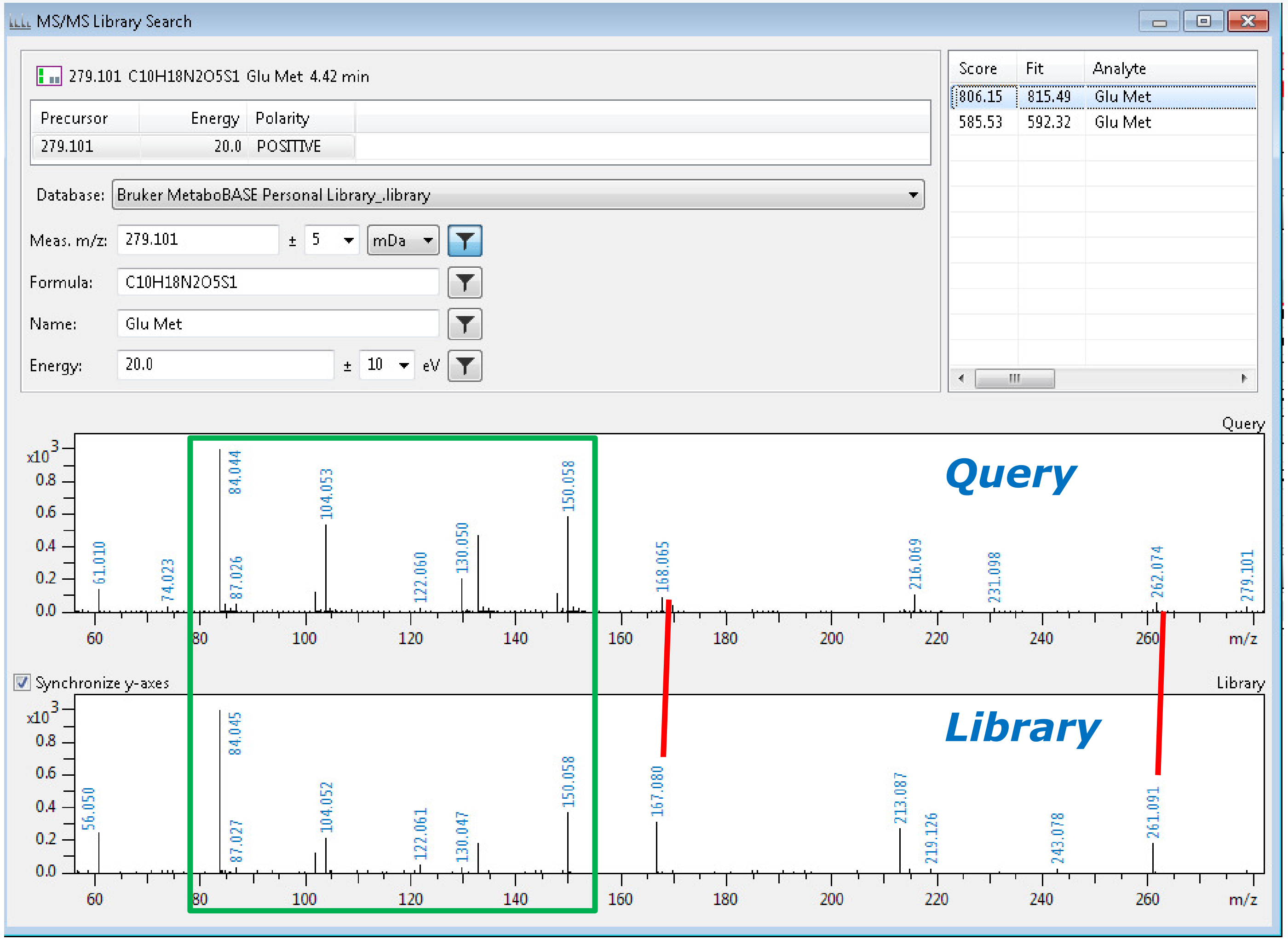


RT [min]	m/z meas.	Name ▲	Molecular Formula	AQ	MS/MS
4.12	261.144	Gamma-Glu-Leu	$C_{11}H_{20}N_2O_5$	🟩🟩🟩🟩🟩🟩	
5.39	276.119	Glu Gln	$C_{10}H_{17}N_3O_6$	🟩🟩🟩🟩🟩⬜	
5.21	277.103	Glu Glu	$C_{10}H_{16}N_2O_7$	🟩🟩🟩🟩🟩⬜	
4.42	279.101	Glu Met	$C_{10}H_{18}N_2O_5S$	🟩🟩⬜⬜	
5.11	148.061	Glutamate	$C_5H_9NO_4$	🟩🟩🟩🟩🟩🟩⬜	
5.31	147.076	Glutamine	$C_5H_{10}N_2O_3$	🟩🟩🟩🟩🟩🟩🟩🟩	
⋮	⋮	⋮	⋮	⋮	⋮

Spectral library query points to further di-peptides, but for several the MS/MS "AQ" indicated a lower confidence in the identification.
The bucket annotated "GluMet" (highlighted in red) was further investigated.

Results & Discussion:

Spectral library query points to *alpha*-Glu-Met
-> Manual inspection indicates the compound is actually a *gamma*-Glu-Met di-peptide


MS/MS Library Search

279.101 C10H18N2O5S1 Glu Met 4.42 min

Precursor	Energy	Polarity
279.101	20.0	POSITIVE

Database: Bruker MetaboBASE Personal Library_.library

Meas. m/z: 279.101 ± 5 mDa

Formula: C10H18N2O5S1

Name: Glu Met

Energy: 20.0 ± 10 eV

Score	Fit	Analyte
806.15	815.49	Glu Met
585.53	592.32	Glu Met

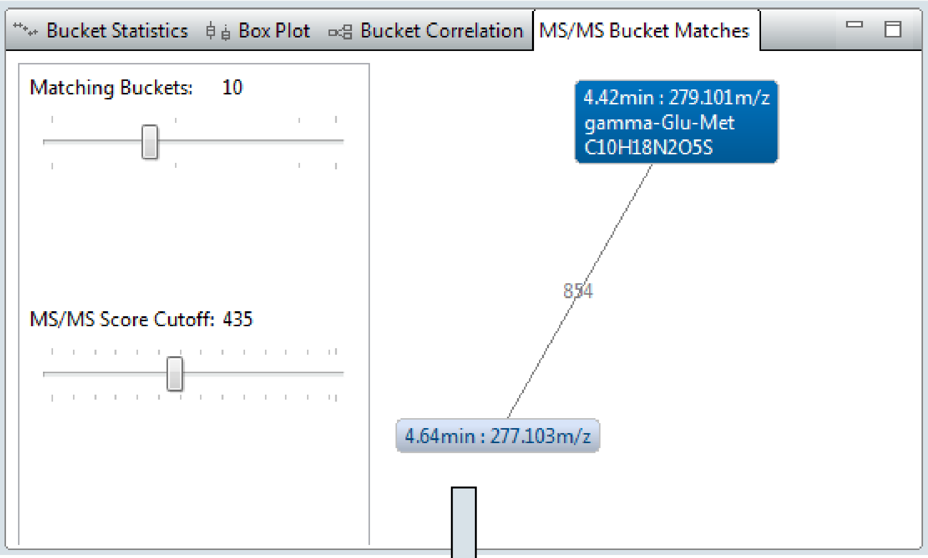
☑ Synchronize y-axes

Low m/z fragment ions match between Query and Library spectrum.

But different and characteristic fragment masses point to *gamma* instead of *alpha* linkage.

Results & Discussion:

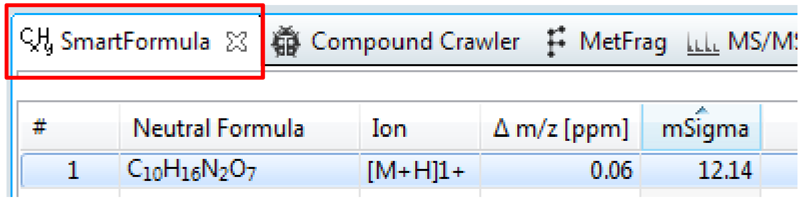
Following the identification of several gamma-glutamyl dipeptides an additional di-peptides in the dataset was identified by “MS/MS bucket matching”



Matching Buckets: 10

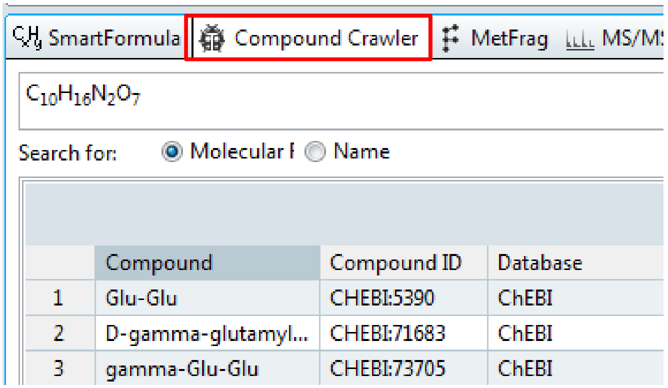
MS/MS Score Cutoff: 435

Formula generation



#	Neutral Formula	Ion	Δ m/z [ppm]	mSigma
1	$C_{10}H_{16}N_2O_7$	[M+H]1+	0.06	12.14

Structure search

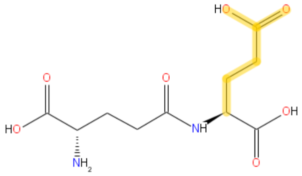


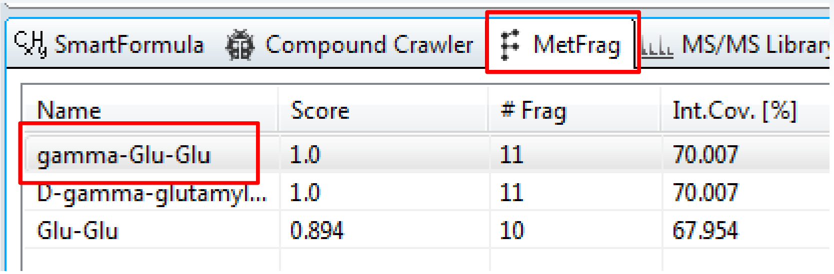
$C_{10}H_{16}N_2O_7$

Search for: ● Molecular F ○ Name

	Compound	Compound ID	Database
1	Glu-Glu	CHEBI:5390	ChEBI
2	D-gamma-glutamyl...	CHEBI:71683	ChEBI
3	gamma-Glu-Glu	CHEBI:73705	ChEBI

in-silico fragmentation

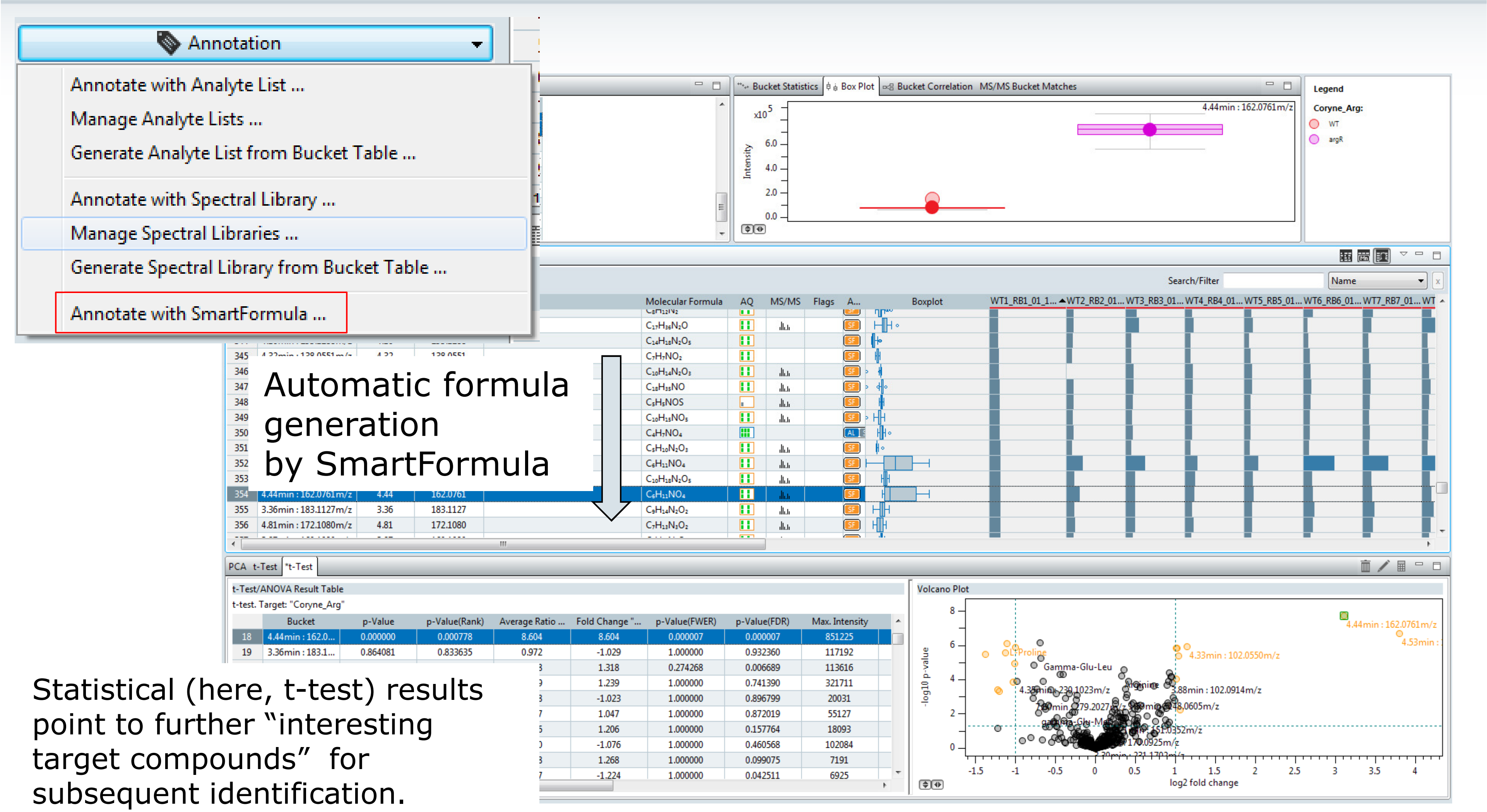




Name	Score	# Frag	Int.Cov. [%]
gamma-Glu-Glu	1.0	11	70.007
D-gamma-glutamyl...	1.0	11	70.007
Glu-Glu	0.894	10	67.954

Results & Discussion:

Finally, the most likely molecular formula for all buckets in the bucket table were automatically generated using “SmartFormula”



				Molecular Formula	AQ	MS/MS	Flags	A...
				$C_{17}H_{36}N_2O$				SF
				$C_{14}H_{18}N_2O_5$				SF
345	4.32min : 138.0551m/z	4.32	138.0551	$C_7H_7NO_2$				SF
346				$C_{10}H_{14}N_2O_3$				SF
347				$C_{18}H_{35}NO$				SF
348				C_5H_9NOS				SF
349				$C_{10}H_{15}NO_5$				SF
350				$C_4H_7NO_4$				AL
351				$C_5H_{10}N_2O_3$				SF
352				$C_6H_{11}NO_4$				SF
353				$C_{10}H_{18}N_2O_5$				SF
354	4.44min : 162.0761m/z	4.44	162.0761	$C_6H_{11}NO_4$				SF
355	3.36min : 183.1127m/z	3.36	183.1127	$C_9H_{14}N_2O_2$				SF
356	4.81min : 172.1080m/z	4.81	172.1080	$C_7H_{13}N_3O_2$				SF

Automatic formula generation by SmartFormula

t-Test/ANOVA Result Table

t-test. Target: "Coryne_Arg"

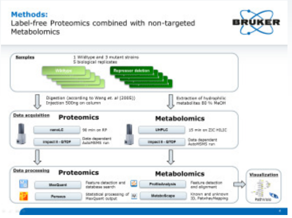
	Bucket	p-Value	p-Value(Rank)	Average Ratio ...	Fold Change "...	p-Value(FWER)	p-Value(FDR)	Max. Intensity
18	4.44min : 162.0...	0.000000	0.000778	8.604	8.604	0.000007	0.000007	851225
19	3.36min : 183.1...	0.864081	0.833635	0.972	-1.029	1.000000	0.932360	117192
					1.318	0.274268	0.006689	113616
					1.239	1.000000	0.741390	321711
					-1.023	1.000000	0.896799	20031
					1.047	1.000000	0.872019	55127
					1.206	1.000000	0.157764	18093
					-1.076	1.000000	0.460568	102084
					1.268	1.000000	0.099075	7191
					-1.224	1.000000	0.042511	6925

Statistical (here, t-test) results point to further “interesting target compounds” for subsequent identification.

Conclusions

Increasing arginine production in *C. glutamicum* by rational strain design in combination with metabolomics and proteomics

- Proteomics **and** metabolomics studies were conducted using one high resolution QTOF-MS/MS platform.



Methods:
Label-free Proteomics combined with non-targeted Metabolomics

Samples 1 Wildtype and 3 mutant strains
5 biological replicates

Wildtype

Repressor deletion

Digestion (according to Wang et. al (2005))
Injection 500ng on column

Extraction of hydrophilic metabolites 80 % MeOH

Data acquisition

Proteomics

nanoLC — 90 min on RP

impact II - QTOF — Data dependent AutoMSMS run

Metabolomics

UHPLC — 15 min on ZIC HILIC

impact II - QTOF — Data dependent AutoMSMS run

Data processing

Proteomics

MaxQuant — Feature detection and database search

Perseus — Statistical processing of MaxQuant output

Metabolomics

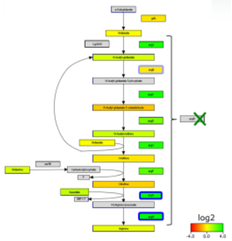
ProfileAnalysis — Feature detection and alignment

MetaboScape — Known and unknown ID, PathwayMapping

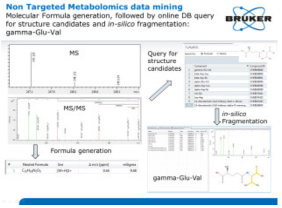
Visualization

Pathvisio

- Pathway mapping of proteomics data shows increased abundance of enzymes involved in arginine biosynthesis in mutant strains, but only in combination with metabolomics influence on arginine production could be determined and ultimately levels increased by rational strain design.



- Non-targeted metabolomics data evaluation enabled to identify compounds responsible for off-target metabolic changes.



Non Targeted Metabolomics data mining
Molecular Formula generation, followed by online DB query for structure candidates and *in-silico* fragmentation:
gamma-Glu-Val

MS

MS/MS

Formula generation

Query for structure candidates

in-silico Fragmentation

gamma-Glu-Val

Acknowledgements

University Bielefeld, Germany
- Marcus Persicke
- Frederik Walter
- Jörn Kalinowski

Bruker Daltonik GmbH
- Heiko Neuweger
- Nikolas Kessler
- Stephanie Kaspar-Schoenefeld

www.bruker.com

For research use only. Not for use in diagnostic procedures. Copyright © 2010 Bruker Daltonics. All rights reserved.