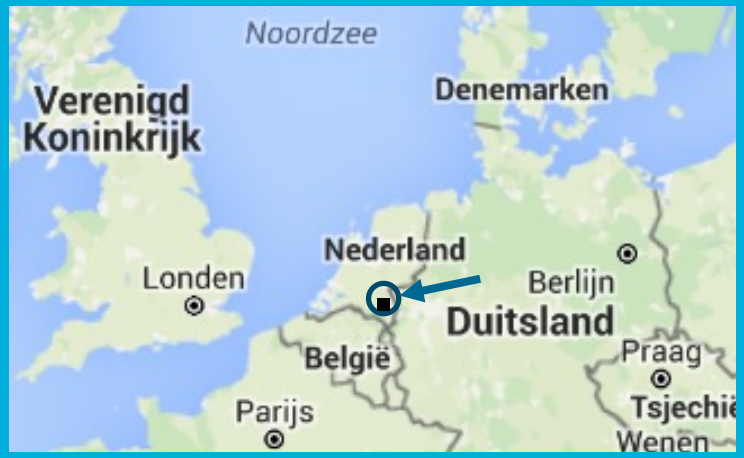
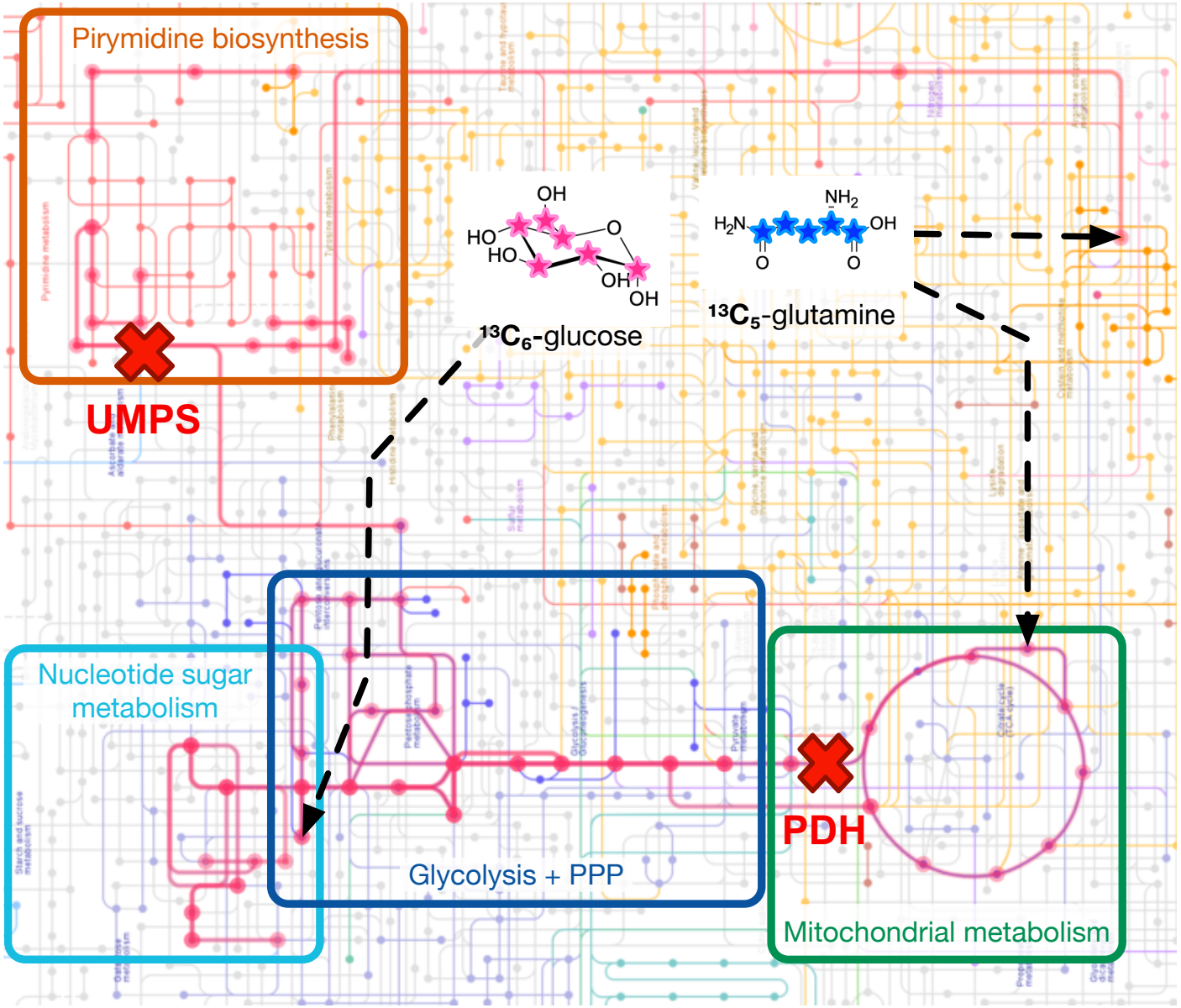
Using RDKit to build analytical assays for diagnosis of metabolic disorders

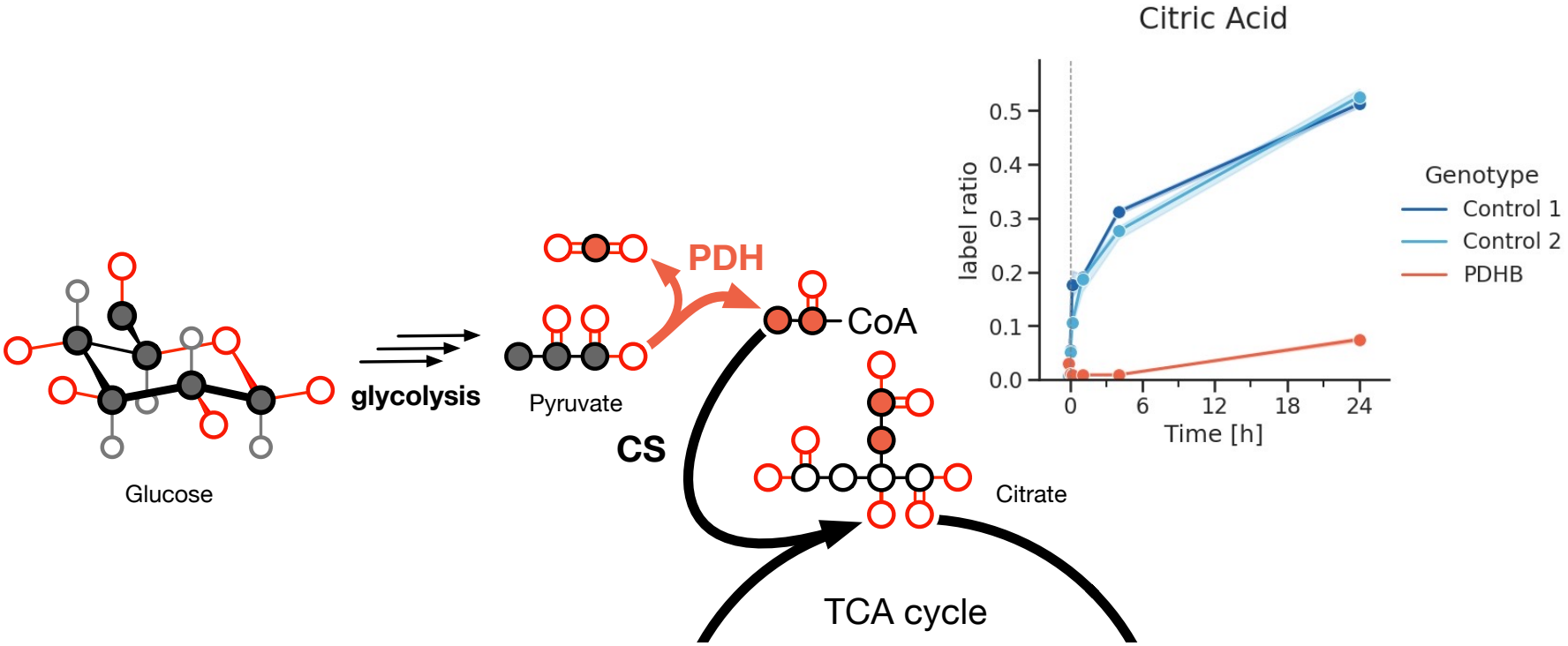


Marek Noga
Translational Metabolic Laboratory

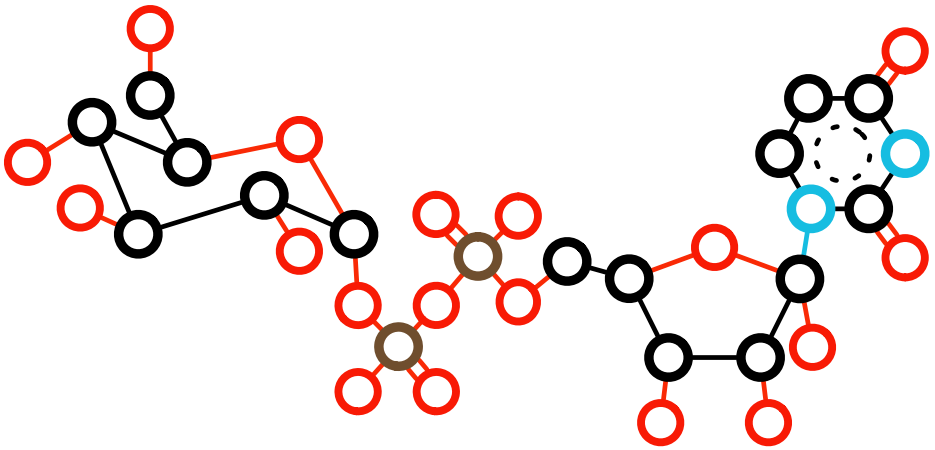
Radboudumc
university medical center


Pirymidine biosynthesis
UMPS
$^{13}C_6$-glucose
$^{13}C_5$-glutamine
Nucleotide sugar metabolism
Glycolysis + PPP
PDH
Mitochondrial metabolism

Using stable isotope labeling and mass spectrometry to diagnose PDH deficiency



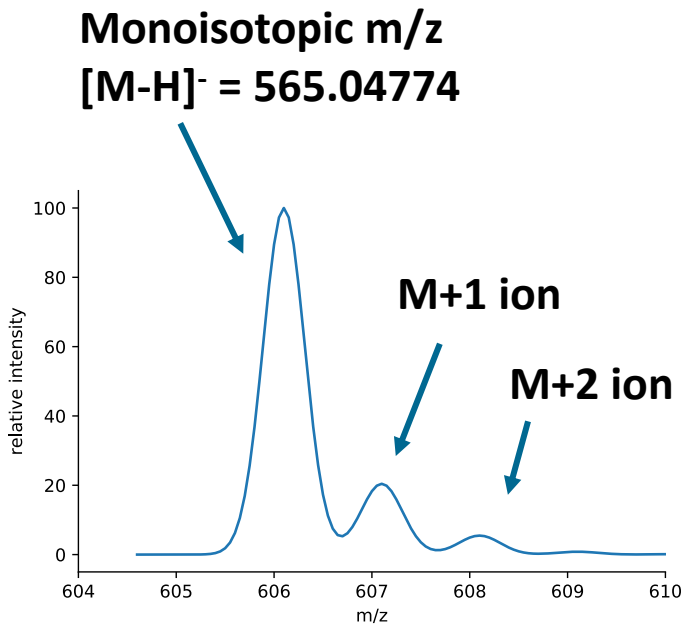
How to measure the mass of a molecule?



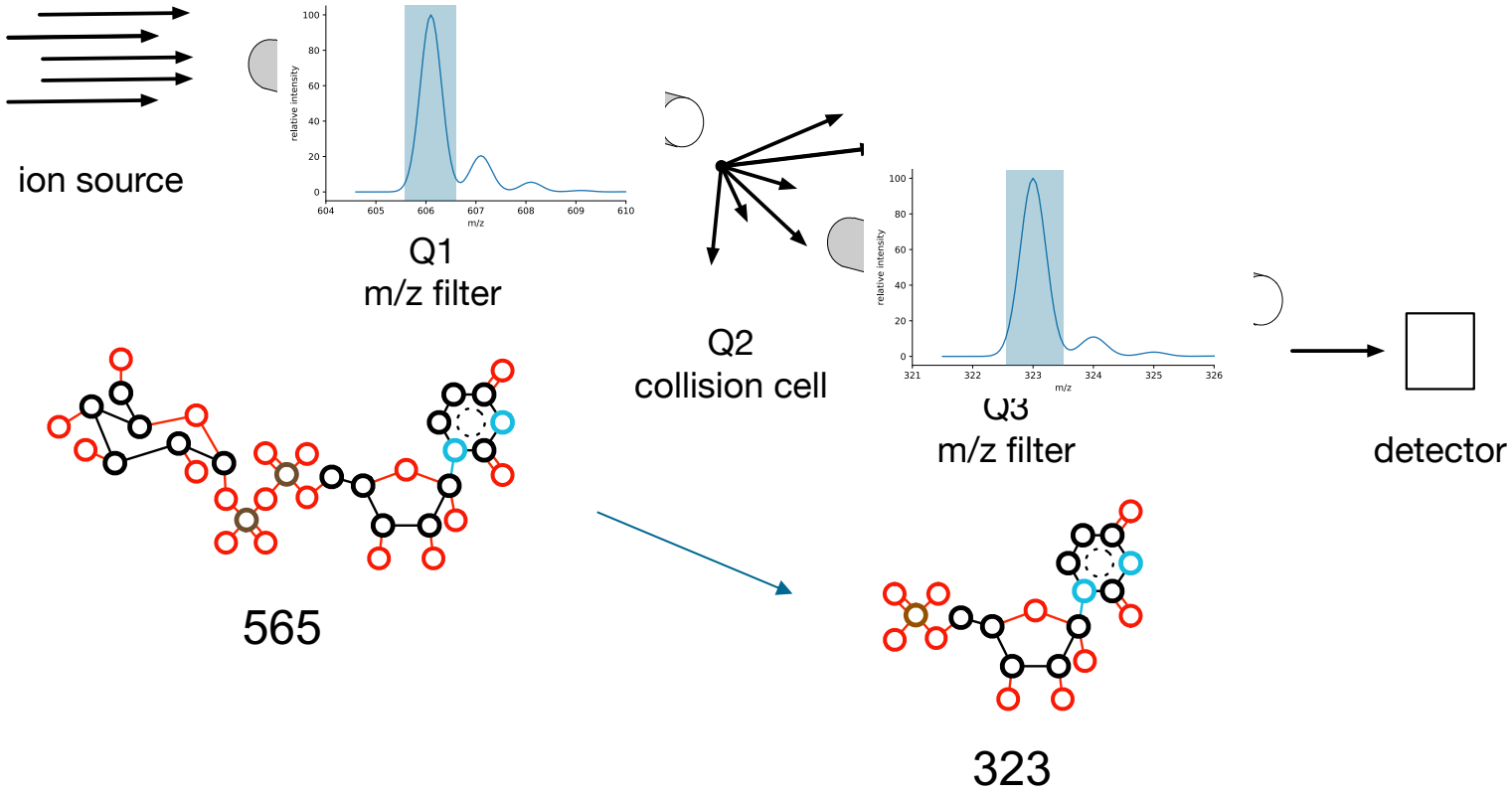
UDP-glucose

$C_{15}H_{24}N_2O_{17}P_2$

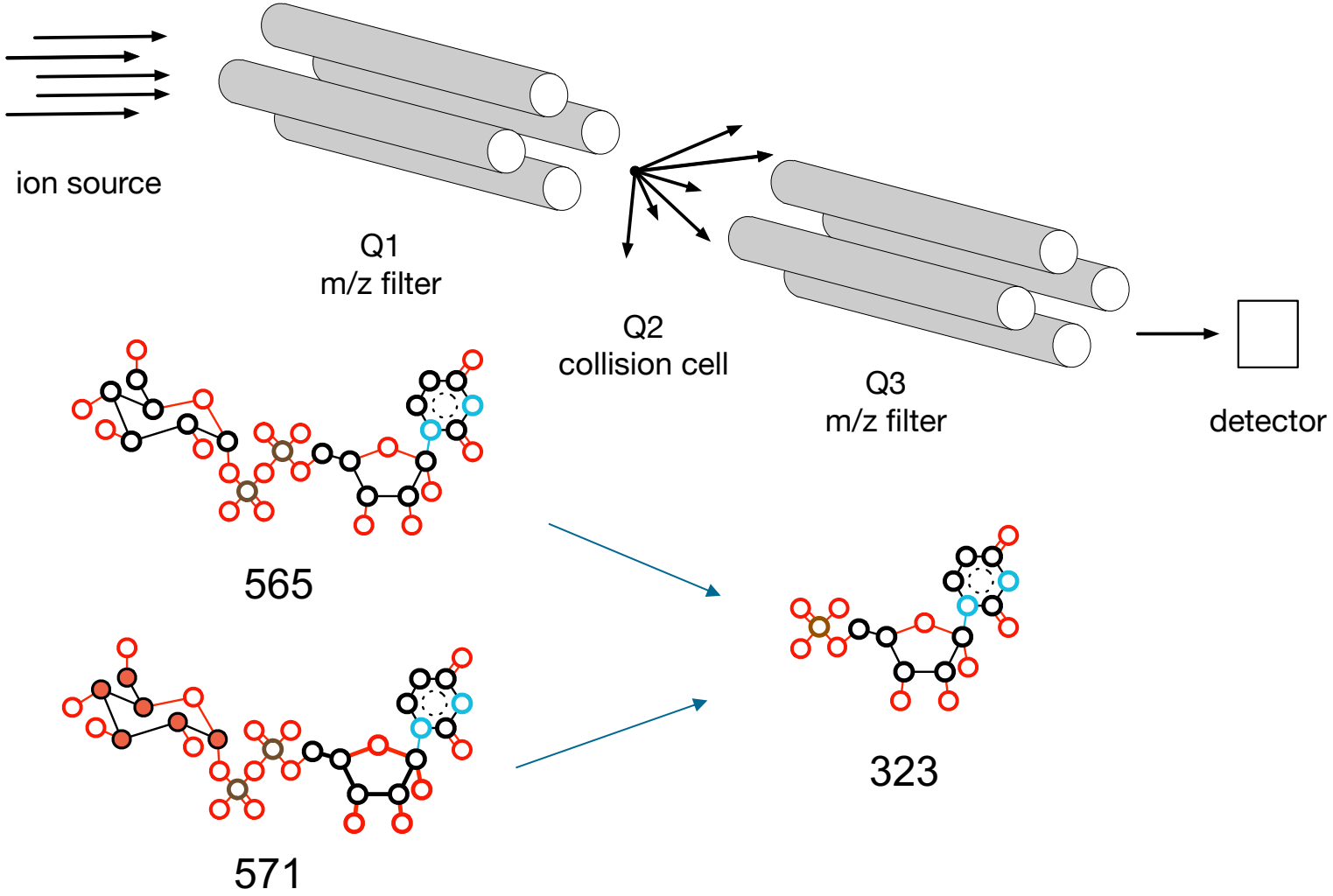
MW = 566.302 g/mol



Triple Quadrupole MS (QQQ)

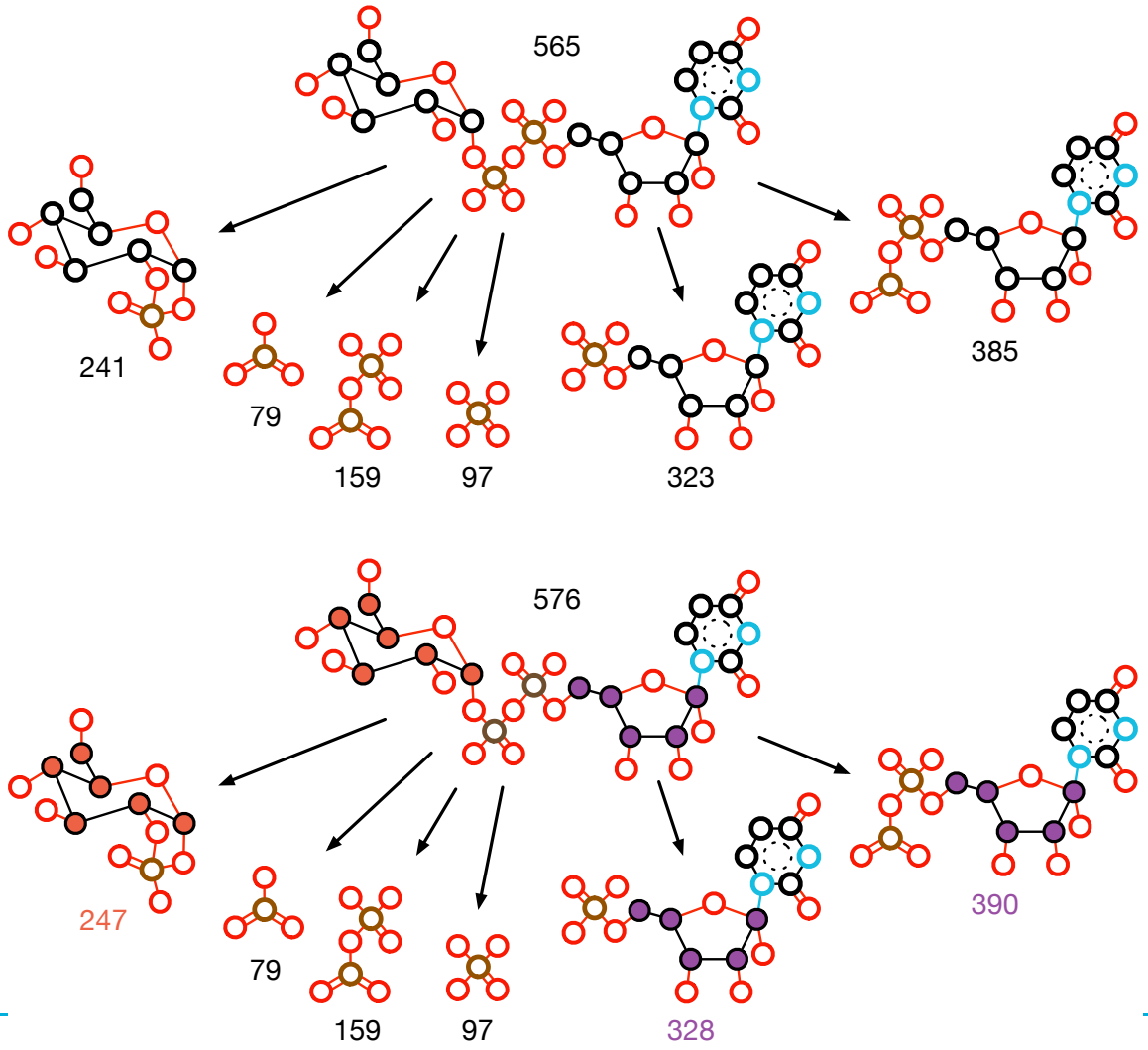


565 -> 323 transition allows for selective detection of UDP-glucose



$^{13}C_6$-UDP-glucose requires different transition!

Fragmentation of UDP-glucose



This looks like a lot of fragments…

- Nucleotide sugar assay
  - 24 compounds
  - 113 transitions

Adding labels:

Sugar moiety
Ribose in nucleotide
Acetyl (UDP-HexNAc, etc)
Pyruvate (CMP-Neu5Ac)
Nucleobase

After adding labels:

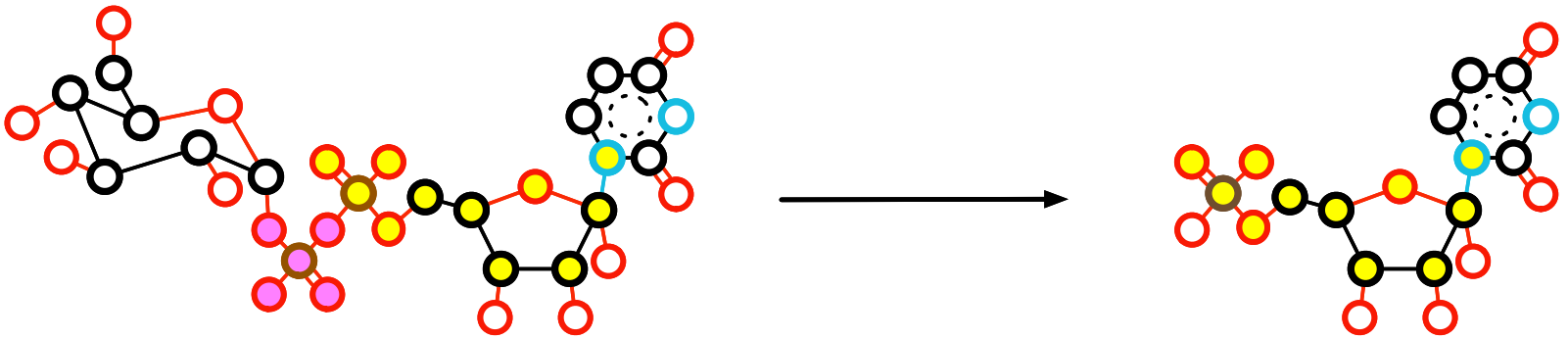
$^{13}C_6$-glucose: 375

^{15}N-glutamine: 296

$^{13}C_6$-galactose + $^{13}C_3$-glucose double labeling: 655

And there are 107 additional compounds waiting on *the shortlist*

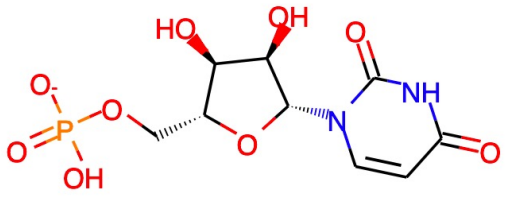
Fragmentation can be modeled



[9]:
```
nmp_ejection = ('[C,$(P(O)(=O)O)]O' # oxygen linked to phosphate group or carbon
                '[$(P(O)(=O)(O)O):2]' # phosphorus of phosphate group
                '[$(O(P(O)(=O))CC1CCC(n)O1):3]'  # oxygen linked to ribose ring , should work on any nucleotide
                '>>'
                '[O-][P:2][O:3]' # retain oxygen 3 (with ribose ring) and phosphorus
                )

ms.test_reaction(nmp_ejection, upd_glucose)
ms.test_reaction(nmp_ejection, CMP_Neu5Ac)
_ = ms.test_reaction(nmp_ejection, cmp['GDP-Glucose'])
```

```
[C,$(P(O)(=O)O)]O[$(P(O)(=O)(O)O):2][$(O(P(O)(=O))CC1CCC(n)O1):3]>>[O-][P:2][O:3]
323.0286
```



RDKit for the rescue!

- Input:
  - Molecule list with structures SMILES
- All fragmentation reactions encoded by reaction SMARTS:
  - All transitions in nucleotide sugar assay generated using just 9 reactions
  - (written manually, thank you Greg for the recursive SMARTS tutorial)
- Algorithmic merge of degenrate transitions
  - Isomeric compounds
  - Undistiguishable labels
  - For $^{13}C_6$-galactose + $^{13}C_3$-glucose double labeling: 1099 -> 655
- Output:
  - Instrument method in vendor format
    - Agilent, Waters, Sciex
  - Data processing method
    - Skyline from MacCoss Lab, open source (https://skyline.ms/)

Unforseen advantages

- Method generator works as expected when fed with stuctures synthetic analogs, like 2-fluoro-fucose

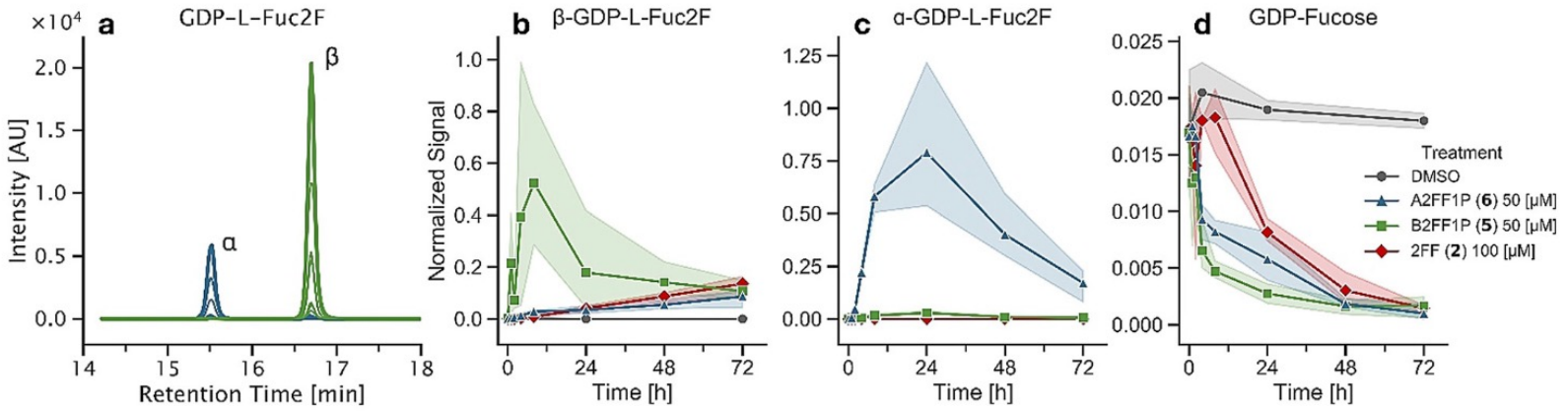
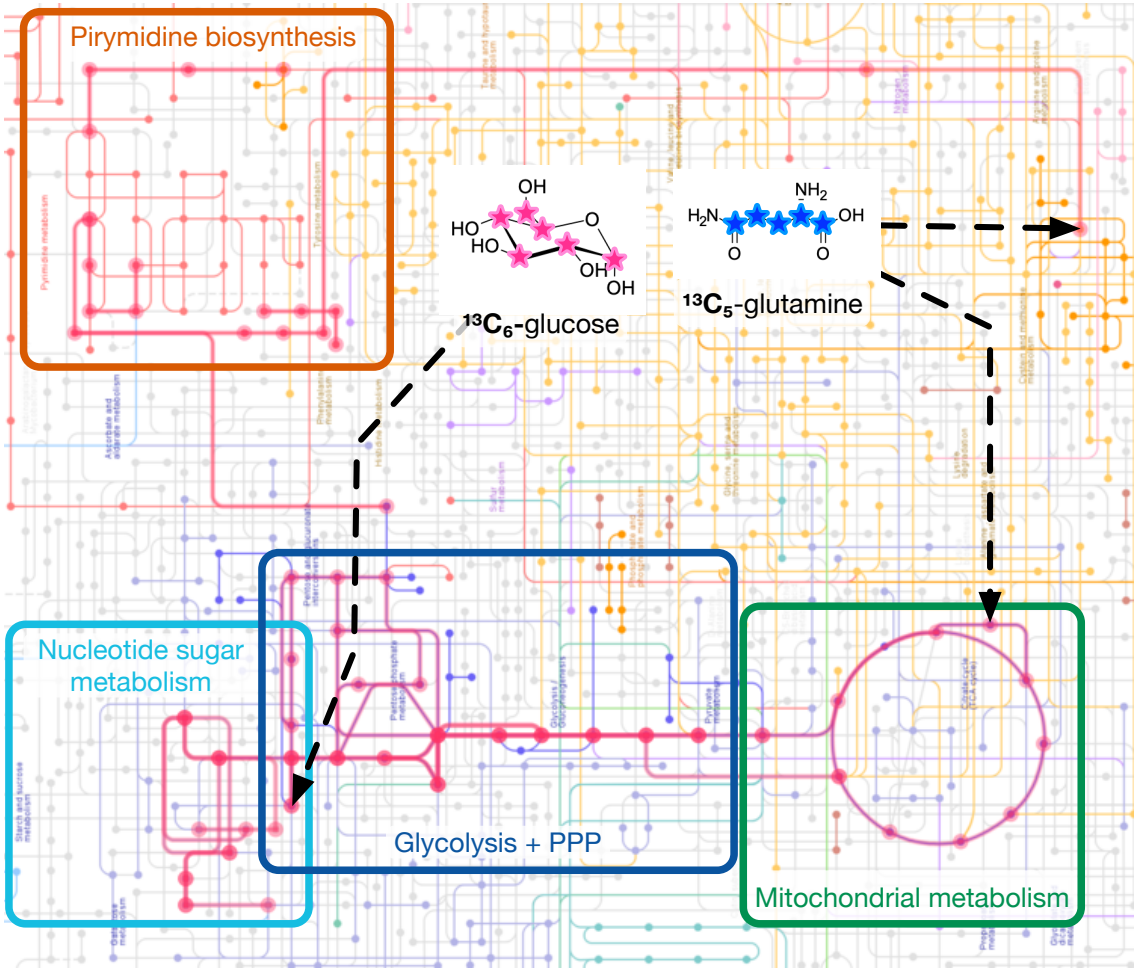


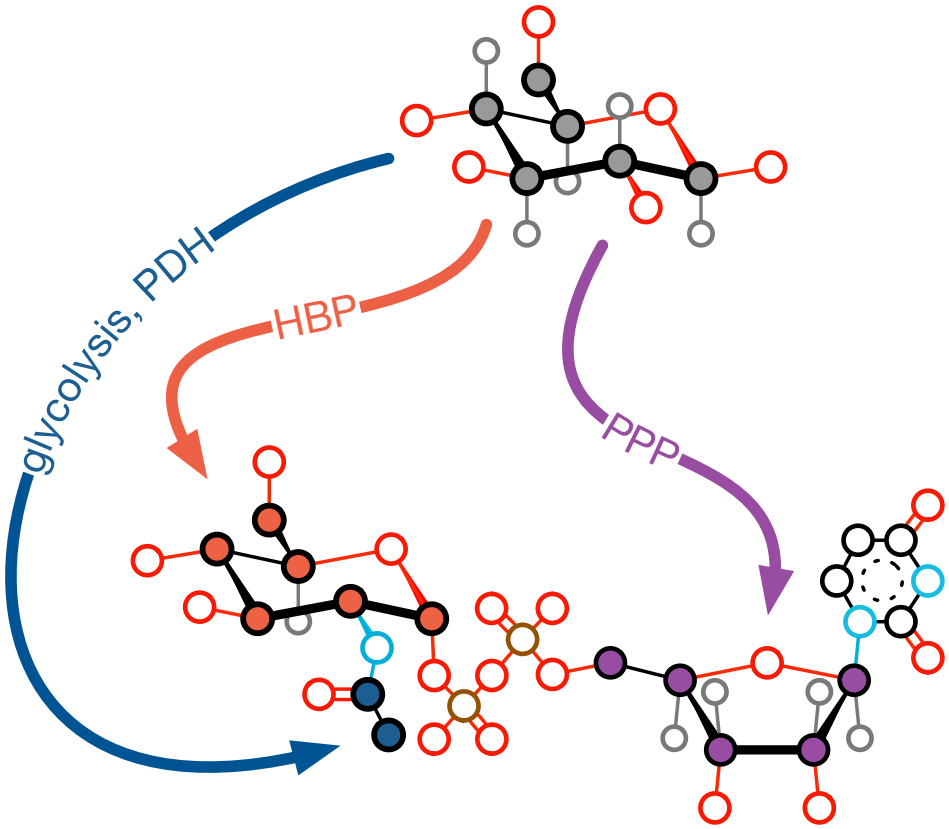
Figure 3. Nucleotide sugar analysis. THP1 cells were incubated for indicated time points with DMSO control, 50 μM A2FF1P, 50 μM B2FF1P or 100 μM 2FF. (a) Relative retention times of α- and β-GDP-L-Fuc2F by incubation with A2FF1P and B2FF1P for 4 hours. (b–d) After sample preparation, the β-GDP-L-Fuc2F (b), α-GDP-L-Fuc2F (c), β-GDP-L-Fucose (d) and other nucleotide levels (Figure S4) were analyzed using reverse-phase ion pairing chromatography coupled to a triple quadrupole mass spectrometer operating in negative ion mode and presented as their abundance in the nucleotide sugar pool ($n = 3$).

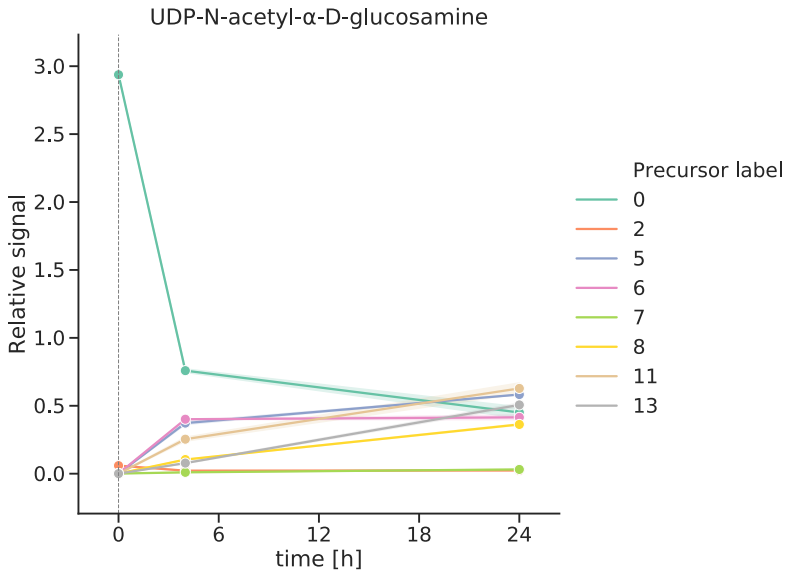
Pijnenborg, J. F. A. *et al.* Cellular Fucosylation Inhibitors Based on Fluorinated Fucose-1-phosphates**. *Chem European J* **27**, 4022–4027 (2021).

Dealing with complexity



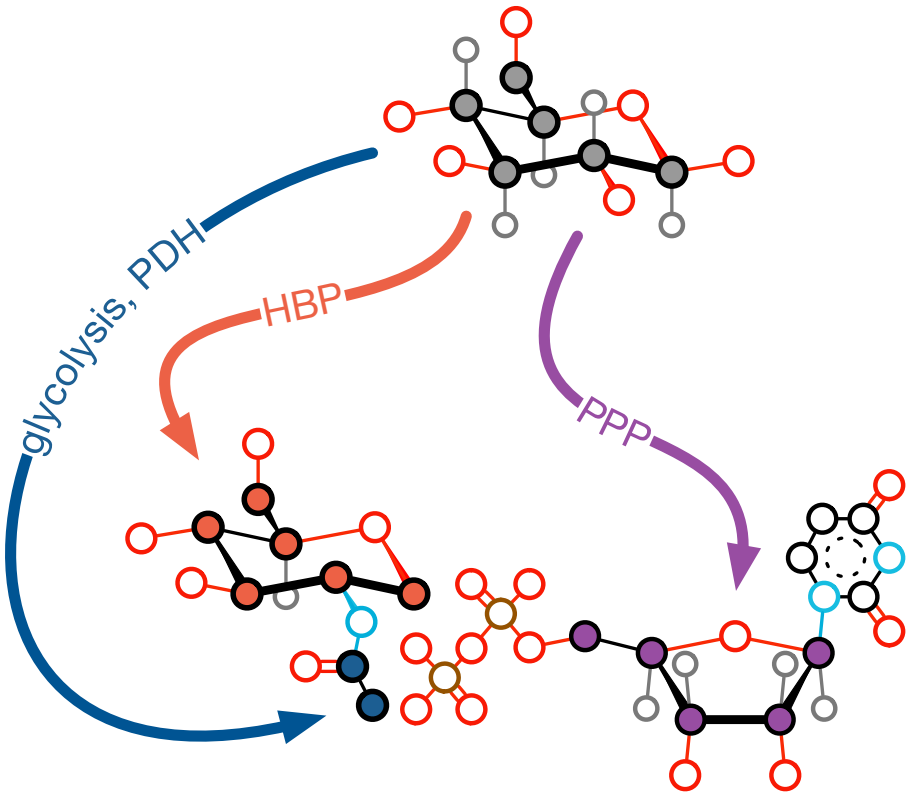
Biosynthesis of UDP-GlcNAc

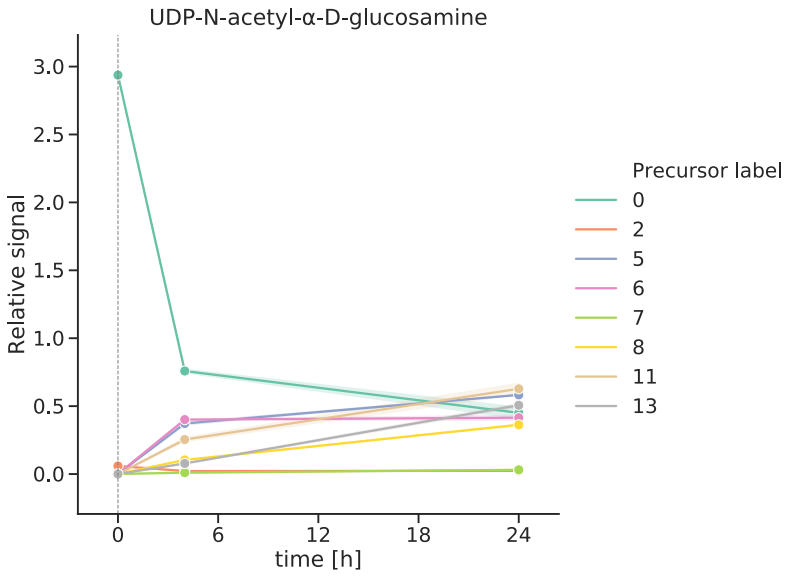




Radboudumc

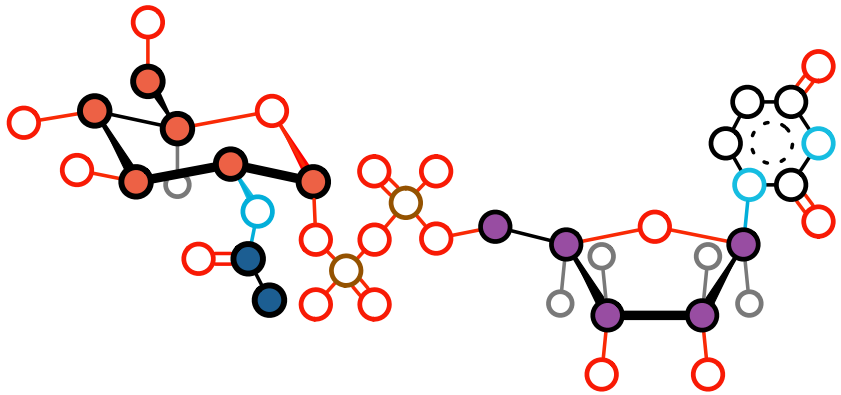
Disentangling UDP-GlcNAc synthesis



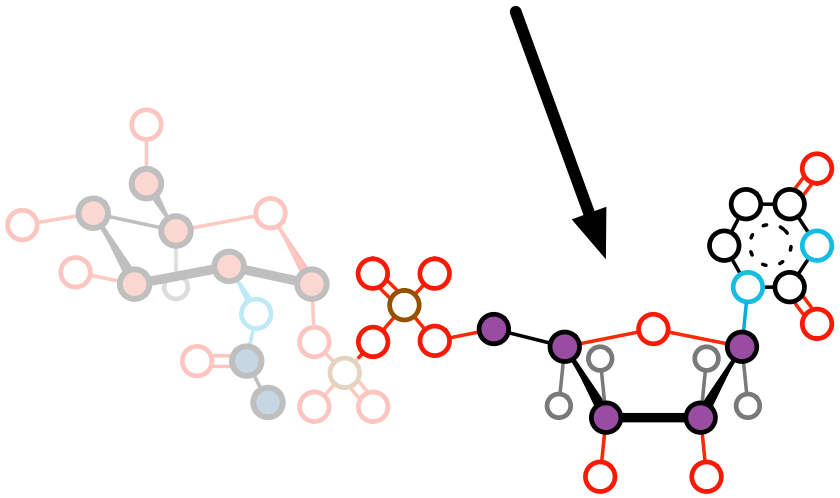


Radboudumc

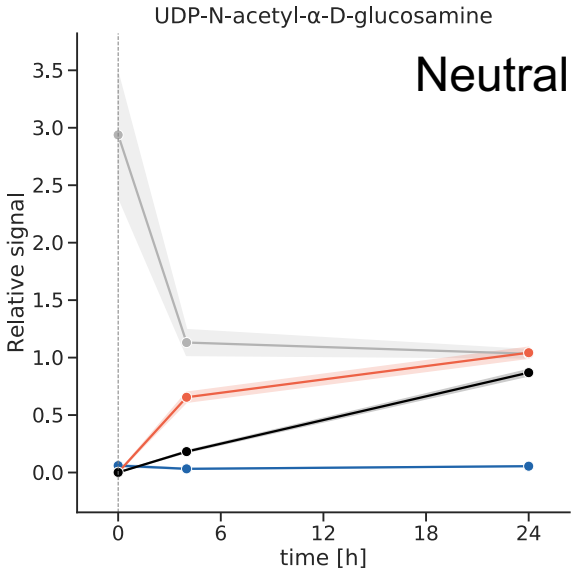
Disentangling UDP-GlcNAc synthesis

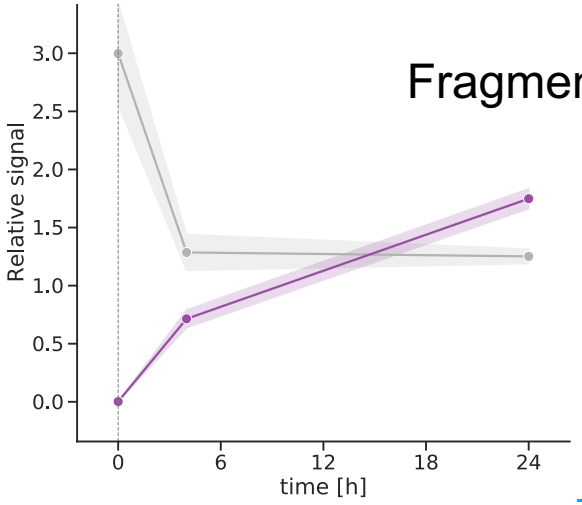


Precursor Ion

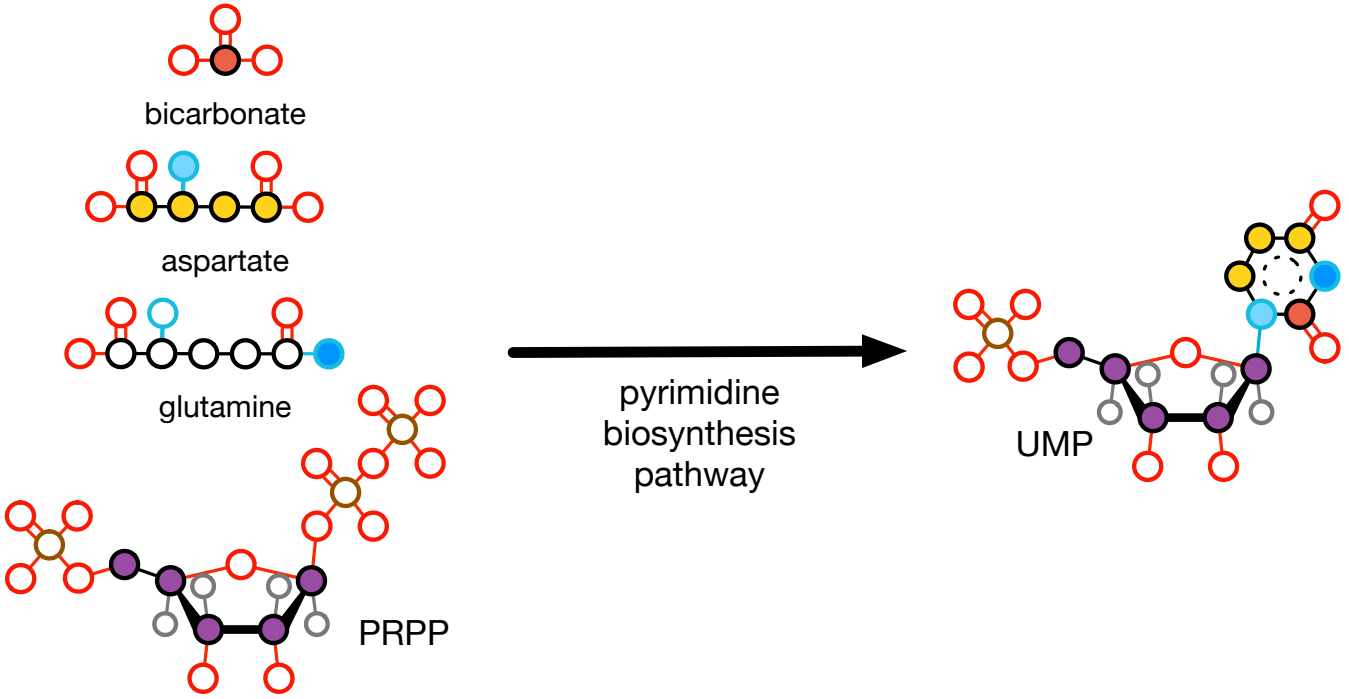


Neutral Loss Fragment Ion





Using SMARTS for labeling selectivity

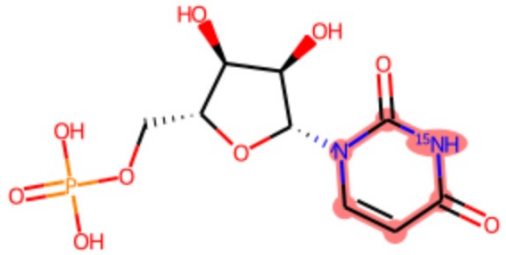


Let's say we want to generate UMP labeled with ^{15}N originating from ^{15}N-glutamine

Using SMART for labeling selectivity

```
[12]: ms.label_by_substructure_match(
          ump,
          "[#7]1@[$([#6]=0)]@[#7]@[$([#6]=0)]@[#6]@[#6]1",
          element="N",
          isotope=15,
          atom_indexes=(3,),
      )
```

[12]:



- It is convenient to use IUPAC atom numbering to define label positions
- RDKit-provided atom indexes use different system
- Solution: write specific SMARTS with match order equivant to IUPAC atom numbering
- Is there a SMARTS extension for match named match reference, equivalent of `(?P<name>...)` in regular expressions?

Furute prospects

Interation with genome-scale metabolic models

- Curently labeling is only semi-automated, SMARTS-based
- Structures, reactions, stoichometry is already available in pathway and model databases
- Atom mappings available, but likely it would be better to re-write whole metabolism in reaction SMARTS
- Some unexpected limitations by stereochemistry handling in reactions in RDKit:
  - Aconitase (TCA cycle):
    - Citrate --> (D-threo-)Isocitrate
    - Citrate is pro-chiral, becomes chiral with certain labels
    - Only one isotopomer of isocitrate should be generated

Furute prospects (long term)

Build a general, structure-based MS fragmentation spectra predictor

- It would greatly support identificaiton of unknown molecules
- Reaction SMARTS build manually, based guided by experience and experimental data
  - One transformation = one stable, detectable ion
- Often product ions results from multiple elementary reactions, including re-arrangements
- Likely it is possible define a set of elementary reactions (as reaction SMARTS) and rules connecting them to predict MS spectrum of an arbitrary structure
- Some systems exist:
  - Proteins and peptides, using sequence input
  - Lipids or glycans, using dedicated notations
  - Current systems for metabolites are based on bond-breaking models and machnie learning, likely not chemistry-aware enough

Acknowledgments

- Nucleotide biosynthesis (MUMC+)
  - Marit v/d Wiel
  - Sandra Coenen
  - Jörgen Bureau
  - Laura Steinbusch
- Glycosylation disorders (Radboudumc)
  - Raisa Veizaj
  - Hanneke Kwast
  - Federica Conte
  - Dirk Lefeber
- Mitochondrial disorders (Radboudumc)
  - Nick Zomer
  - Liesbeth Wintjes
  - Arno van Rooij
  - Frans v/d Brandt
  - Richard Rodenburg
- Radboudumc
  - Alain van Gool
- Inspiration and prior work
  - Greg Bokinsky (TU Delft)
  - Luc Patiny (EPFL)