Expansion and Exploration of Prebiotic Chemical Reaction Spaces through Rule Based Modeling

Nino Lauber

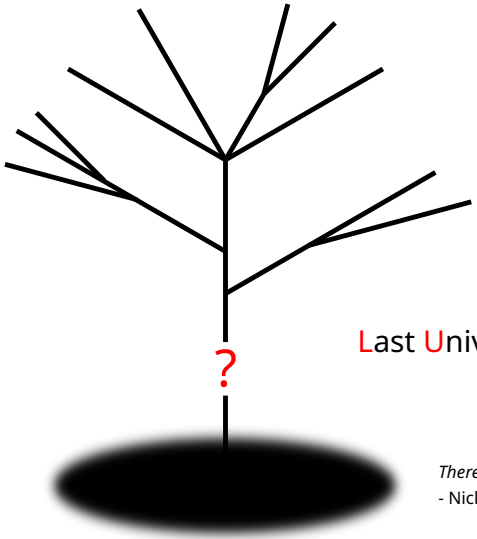
39th TBI Winterseminar

15th Februar 2024, Bled

Introduction

Motivation: Origins of Life (OL) Research

Motivation: Origins of Life (OL) Research



Last Universal Common Ancestor

LUCA

There is a black hole at the heart of biology.
- Nick Lane, The Vital Question

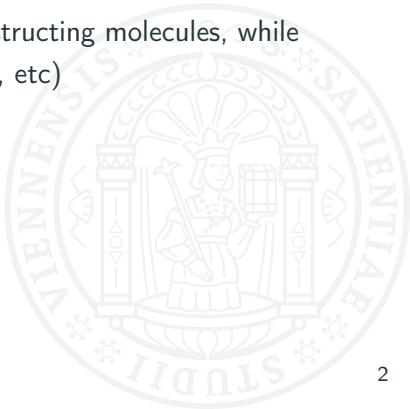
State of Research

State of Research

- **Substantial progress** has been made within prebiotic chemistry and origins of life (OL) research

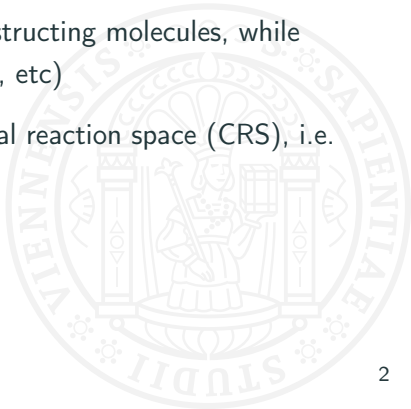
State of Research

- **Substantial progress** has been made within prebiotic chemistry and origins of life (OL) research
- Many lines of research focus on a **similar logic** of constructing molecules, while varying reaction conditions (catalysts, temperature, pH, etc)



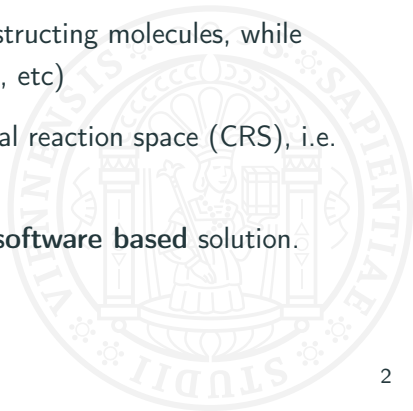
State of Research

- **Substantial progress** has been made within prebiotic chemistry and origins of life (OL) research
- Many lines of research focus on a **similar logic** of constructing molecules, while varying reaction conditions (catalysts, temperature, pH, etc)
- Need for a way to systematically **trace out** the chemical reaction space (CRS), i.e. the space of **possible reactions** of a given system.



State of Research

- **Substantial progress** has been made within prebiotic chemistry and origins of life (OL) research
- Many lines of research focus on a **similar logic** of constructing molecules, while varying reaction conditions (catalysts, temperature, pH, etc)
- Need for a way to systematically **trace out** the chemical reaction space (CRS), i.e. the space of **possible reactions** of a given system.
- Expansion and analysis of a CRS can be done using a **software based** solution.



Methods

Rule-Based Expansion of a Chemical Reaction Space

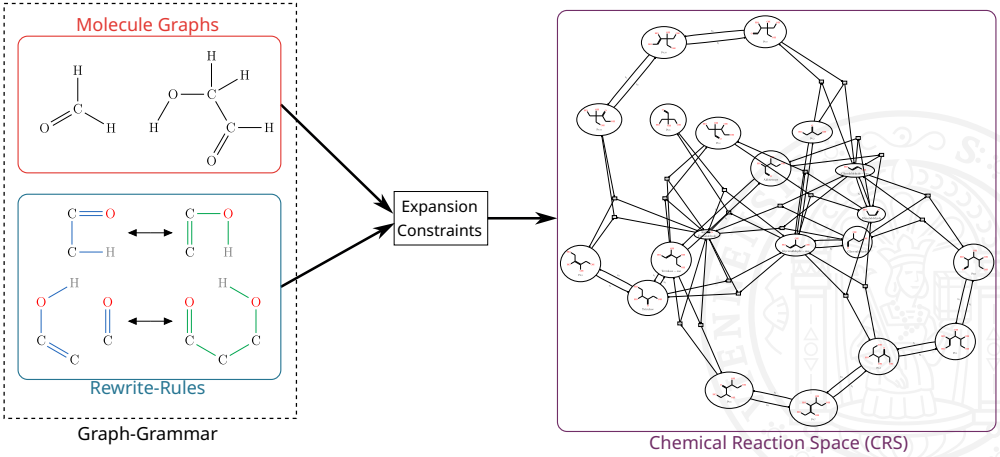
Rule-Based Expansion of a Chemical Reaction Space

Expansion of a CRS using the *MØD*[1] software package

[1] J. L. Andersen, C. Flamm, D. Merkle & P. F. Stadler; **LNCS** 9761:73–88; 2016.

Rule-Based Expansion of a Chemical Reaction Space

Expansion of a CRS using the *MØD*[1] software package



[1] J. L. Andersen, C. Flamm, D. Merkle & P. F. Stadler; **LNCS** 9761:73–88; 2016.

Search for Conserved Moieties

Search for Conserved Moieties

A CRS with m molecules and r reactions can be represented by it's **stoichiometric matrix** S ($m \times r$ matrix containing stoichiometric coefficients of each reaction)

Search for Conserved Moieties

A CRS with m molecules and r reactions can be represented by it's **stoichiometric matrix** S ($m \times r$ matrix containing stoichiometric coefficients of each reaction)

Any vector $c = (c_1, \ldots, c_m)$ with the property:

$$c \cdot S = 0$$

represents a **conserved moiety** within the CRS[1].

[1] S. Müller, C. Flamm & P. F. Stadler; **J. Cheminform.** 14(1):1–24; 2022

Search for Conserved Moieties

A CRS with m molecules and r reactions can be represented by it's **stoichiometric matrix** S ($m \times r$ matrix containing stoichiometric coefficients of each reaction)

Any vector $c = (c_1, \ldots, c_m)$ with the property:

$$c \cdot S = 0$$

represents a **conserved moiety** within the CRS[1].

Given N conserved moieties, any molecule in the system can be represented as a vector: $m = (m_1, \ldots, m_N)^T$, m_i amount of moiety i

[1] S. Müller, C. Flamm & P. F. Stadler; **J. Cheminform.** 14(1):1–24; 2022

Search for Autocatalytic (AC) Cycles

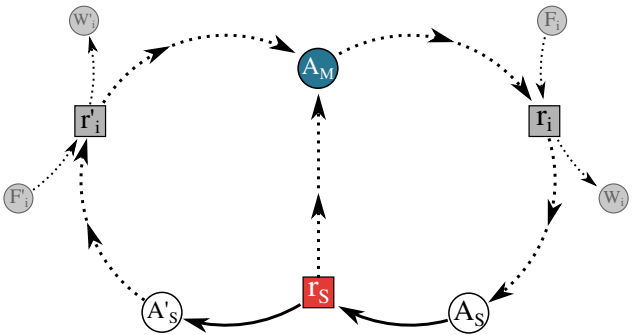
Search for Autocatalytic (AC) Cycles

Network topological definition of autocatalytic reaction cycles[1]

[1] A. Blokhuis, D. Lacoste & P. Nghe; **Proc. Natl. Acad. Sci.** 117(41): 25230–25236.

Search for Autocatalytic (AC) Cycles

Network topological definition of autocatalytic reaction cycles[1]

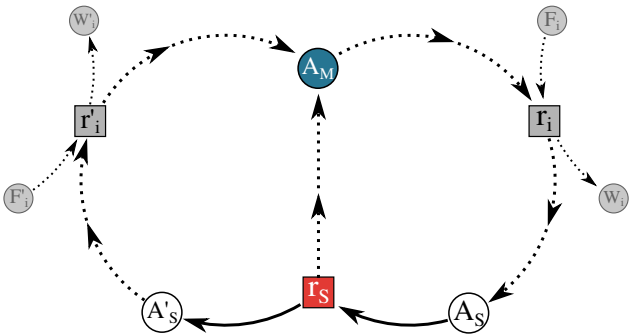


with permission from Phillip Honegger (publication forthcomming)

[1] A. Blokhuis, D. Lacoste & P. Nghe; **Proc. Natl. Acad. Sci.** 117(41): 25230–25236.

Search for Autocatalytic (AC) Cycles

Network topological definition of autocatalytic reaction cycles[1]



with permission from Phillip Honegger (publication forthcomming)

Search constraints:

- exclusion of molecules
- exclusion of reactions
- maximum cycle length

[1] A. Blokhuis, D. Lacoste & P. Nghe; **Proc. Natl. Acad. Sci.** 117(41): 25230–25236.

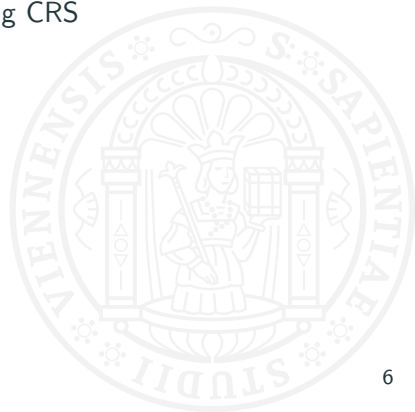
CRS Expansion-Analysis Workflow

CRS Expansion-Analysis Workflow

- Implement a given chemistry as a **graph-grammar**

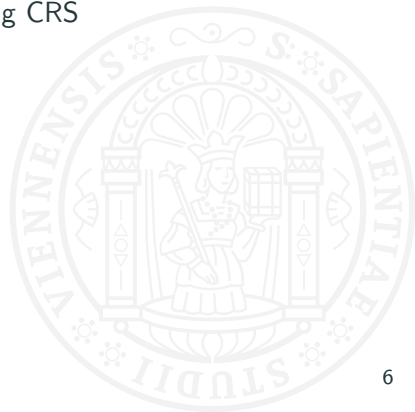
CRS Expansion-Analysis Workflow

- Implement a given chemistry as a **graph-grammar**
- Perform a **constrained expansion** of the corresponding CRS



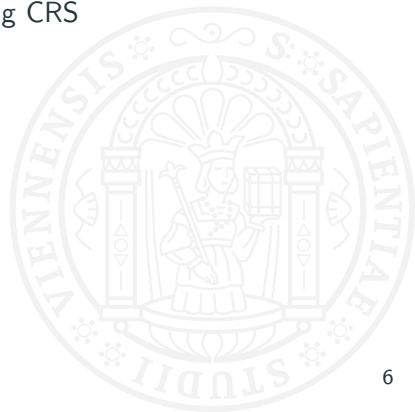
CRS Expansion-Analysis Workflow

- Implement a given chemistry as a **graph-grammar**
- Perform a **constrained expansion** of the corresponding CRS
- Analyze the **properties** of the CRS



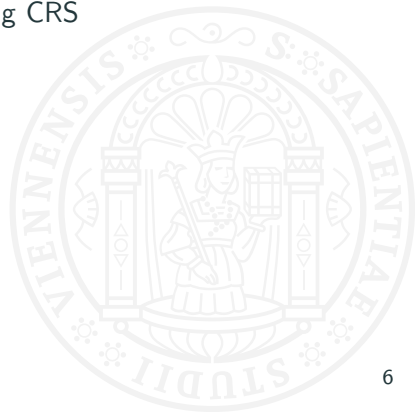
CRS Expansion-Analysis Workflow

- Implement a given chemistry as a **graph-grammar**
- Perform a **constrained expansion** of the corresponding CRS
- Analyze the **properties** of the CRS
- Search for certain **reaction pathways** within the CRS



CRS Expansion-Analysis Workflow

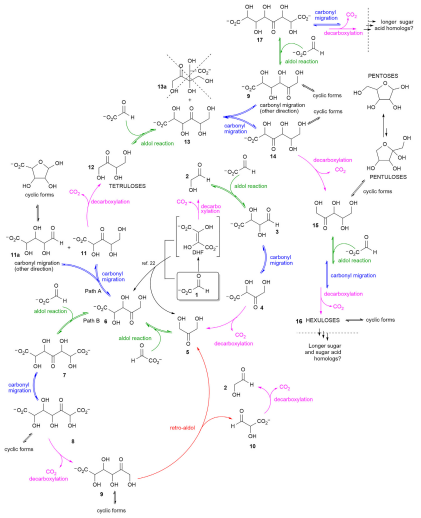
- Implement a given chemistry as a **graph-grammar**
- Perform a **constrained expansion** of the corresponding CRS
- Analyze the **properties** of the CRS
- Search for certain **reaction pathways** within the CRS
- Analyze the **properties** of the found pathways



Case Study: Glyoxylose Reaction

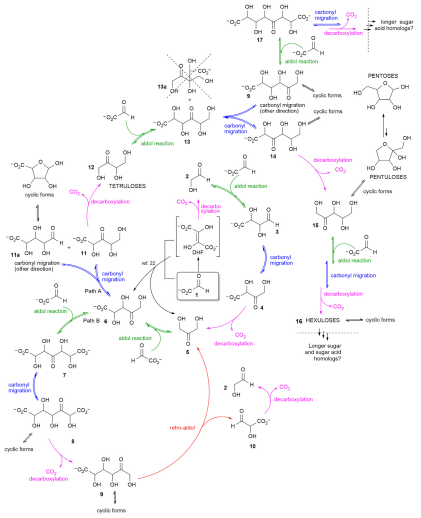
Reaction System

Reaction System



R. Krishnamurthy & C. L. Liotta; **Chem** 9(4): 784–797; 2023.

Reaction System



R. Krishnamurthy & C. L. Liotta; **Chem** 9(4): 784–797; 2023.

Initially present Molecules:

- Glyoxylate
- Glycolaldehyde

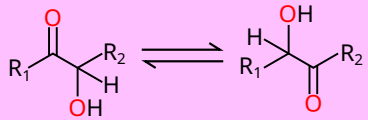
Four types of reactions:

- Carbonyl Migration
- (Retro)-Aldol
- Decarboxylation

Reaction Grammar

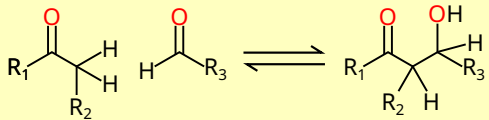
Reaction Grammar

1



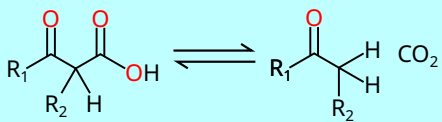
R_1: C, H
R_2: C, H

2



R_1: C, H
R_2: O, H
R_3: C, H

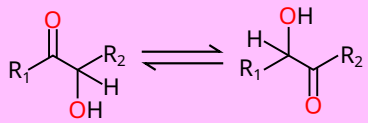
3



R_1: C, H
R_2: O, H

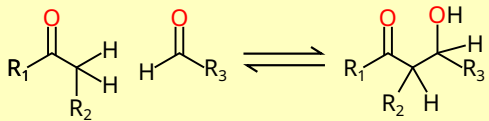
Reaction Grammar

1



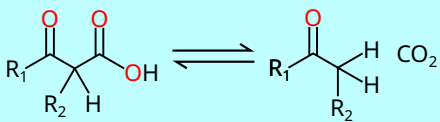
R_1: C, H
R_2: C, H

2



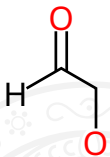
R_1: C, H
R_2: O, H
R_3: C, H

3

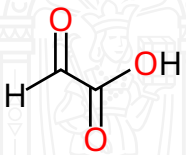


R_1: C, H
R_2: O, H

Glycolaldehyde (GLYC)



Glyoxylate (GLX)



Results

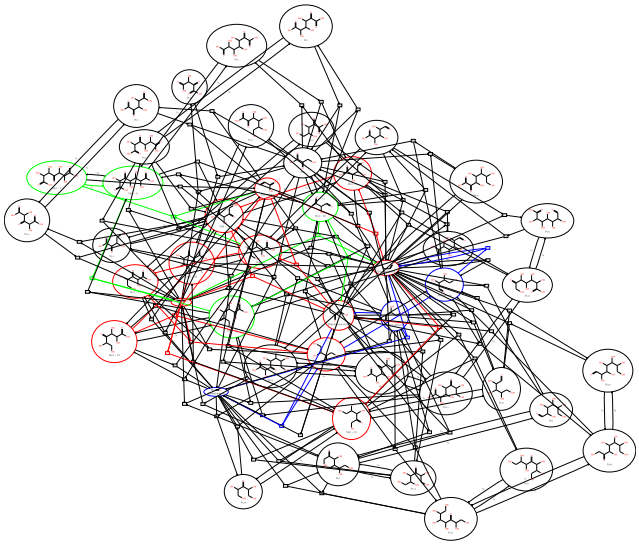
CRS Expansion

CRS Expansion

Expansion constraint:

- no molecules with **more than** 8 C-atoms

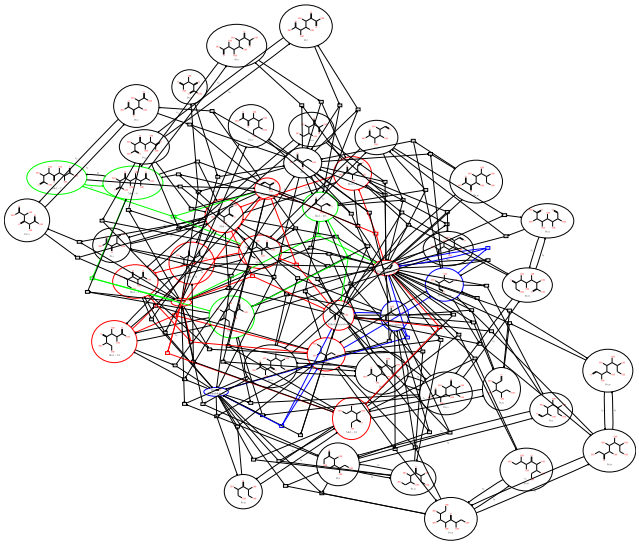
CRS Expansion



Expansion constraint:

- no molecules with **more than** 8 C-atoms

CRS Expansion



Expansion constraint:

- no molecules with **more than** 8 C-atoms

General CRS properties:

- 46 Molecules
- 156 Reactions

Conserved Moieties

Conserved Moieties

Two conserved moieties found in expanded CRS:

- H_2CO-moiety (carbonyl-, alcohol-group)
- CO_2-moiety (acid-group)

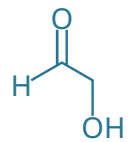
Any molecule in the CRS can be **decomposed** into a **combination** of them

Conserved Moieties

Two conserved moieties found in expanded CRS:

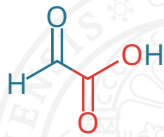
- H_2CO-moiety (carbonyl-, alcohol-group)
- CO_2-moiety (acid-group)

Any molecule in the CRS can be **decomposed** into a **combination** of them



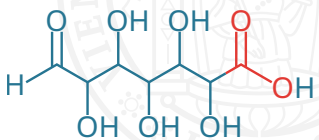
$\begin{pmatrix} 2 \\ 0 \end{pmatrix}$

Glyocolaldehyde (GYC)



$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Glyoxylate (GLX)



$\begin{pmatrix} 6 \\ 1 \end{pmatrix}$

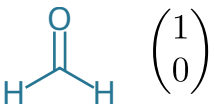
Conserved Moieties

Two conserved moieties found in expanded CRS:

- H_2CO-moiety (carbonyl-, alcohol-group)
- CO_2-moiety (acid-group)

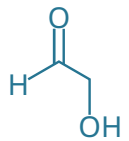
Any molecule in the CRS can be **decomposed** into a **combination** of them

Furthermore, define a set of **basis-species** (molecules with only one of each moiety):

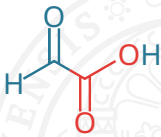


Formaldehyde (FORM) $\begin{pmatrix}1\\0\end{pmatrix}$

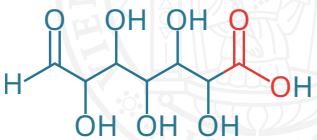
CO_2 $\begin{pmatrix}0\\1\end{pmatrix}$



Glyocolaldehyde (GYC) $\begin{pmatrix}2\\0\end{pmatrix}$



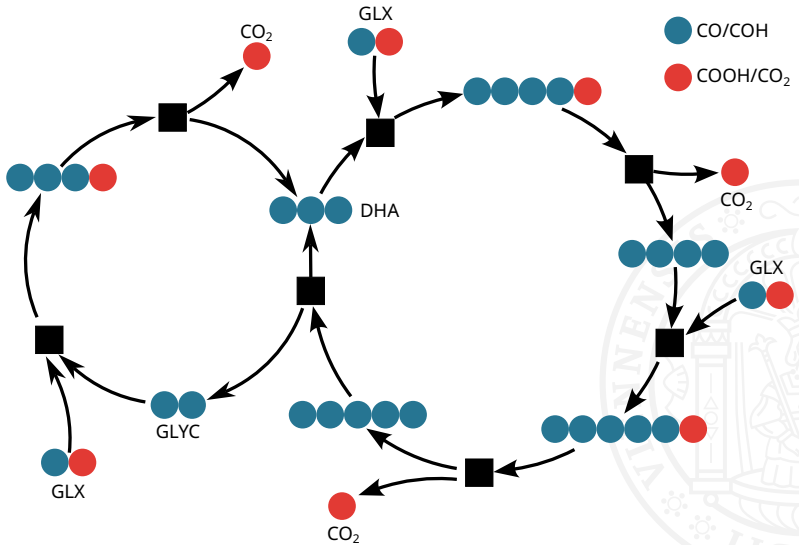
Glyoxylate (GLX) $\begin{pmatrix}1\\1\end{pmatrix}$



$\begin{pmatrix}6\\1\end{pmatrix}$

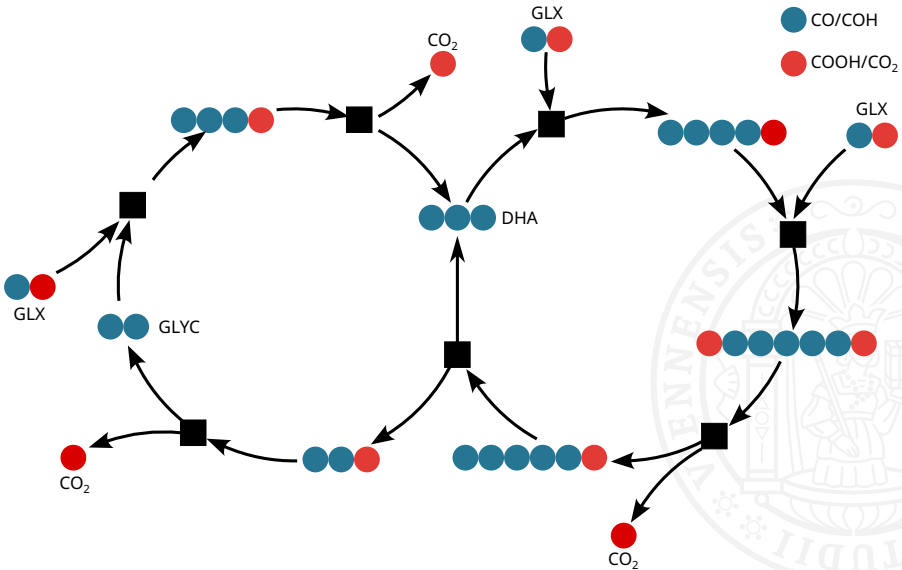
Conserved Moieties: Reaction Path A

Conserved Moieties: Reaction Path A



Conserved Moieties: Reaction Path B

Conserved Moieties: Reaction Path B



AC-Cycle Search

AC-Cycle Search

General constraints:

- GLX, FORM, CO_2 excluded as on-cycle species
- Carboxylation reactions excluded

AC-Cycle Search

General constraints:

- GLX, FORM, CO_2 excluded as on-cycle species
- Carboxylation reactions excluded

Two main search strategies on the CRS:

AC-Cycle Search

General constraints:

- GLX, FORM, CO_2 excluded as on-cycle species
- Carboxylation reactions excluded

Two main search strategies on the CRS:

1. "Strict" Search
   - only aldol-addition with GLX
   - maximum 7 reactions

$\Rightarrow$ 43 AC-Cycles

AC-Cycle Search

General constraints:

- GLX, FORM, CO_2 excluded as on-cycle species
- Carboxylation reactions excluded

Two main search strategies on the CRS:

1. "Strict" Search
   - only aldol-addition with GLX
   - maximum 7 reactions

$\Rightarrow$ 43 AC-Cycles

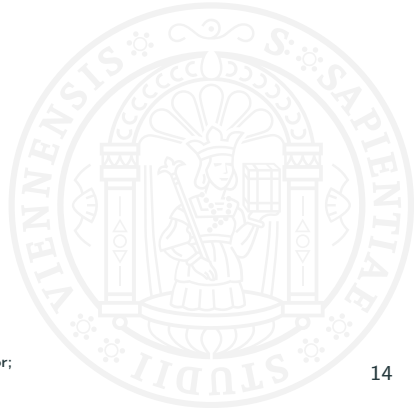
2. "Relaxed" Search
   - general aldol-additions
   - maximum 4 reactions

$\Rightarrow$ 3144 AC-Cycles

AC-Cycle Energies

AC-Cycle Energies

Calculate reaction energy $\Delta_r G^\circ$ using the *eQuilibrator*[1] software package

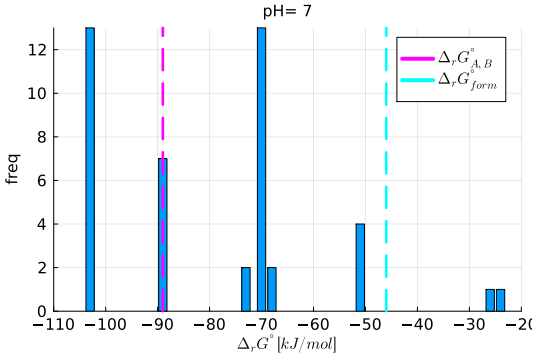


[1] M. E. Beber, M. G. Gollub, D. Mozaffari, K. M. Shebek, A. I. Flamholz, R. Milo & E. Noor; **Nucleic Acids Res.** 50(D1): D603–D609; 2022.

AC-Cycle Energies

Calculate reaction energy $\Delta_r G^\circ$ using the *eQuilibrator*[1] software package

"Strict Search" AC-Cycles:

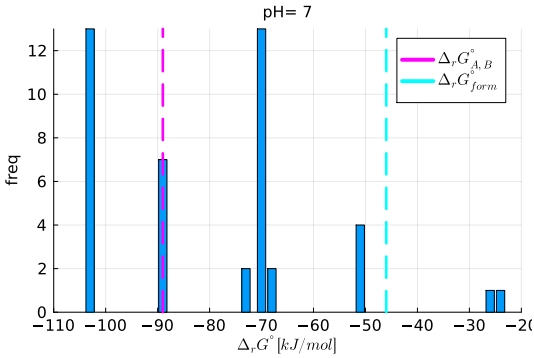


[1] M. E. Beber, M. G. Gollub, D. Mozaffari, K. M. Shebek, A. I. Flamholz, R. Milo & E. Noor; **Nucleic Acids Res.** 50(D1): D603–D609; 2022.

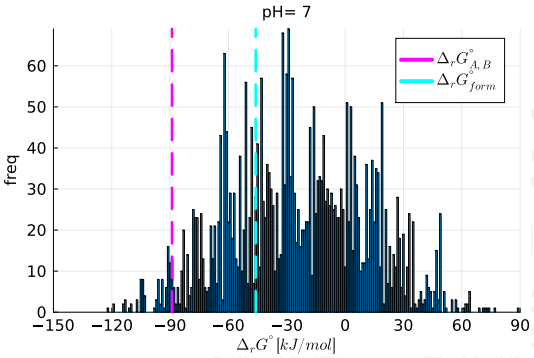
AC-Cycle Energies

Calculate reaction energy $\Delta_r G^\circ$ using the *eQuilibrator*[1] software package

"Strict Search" AC-Cycles:



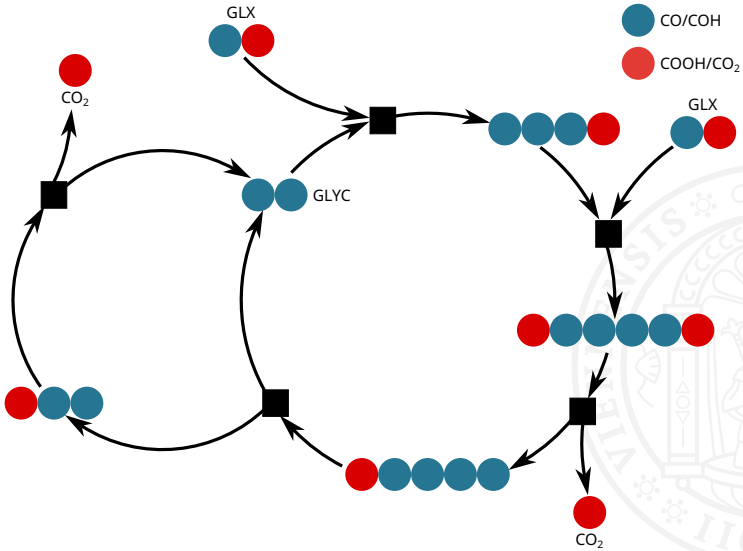
"Relaxed Search" AC-Cycles:



[1] M. E. Beber, M. G. Gollub, D. Mozaffari, K. M. Shebek, A. I. Flamholz, R. Milo & E. Noor; **Nucleic Acids Res.** 50(D1): D603–D609; 2022.

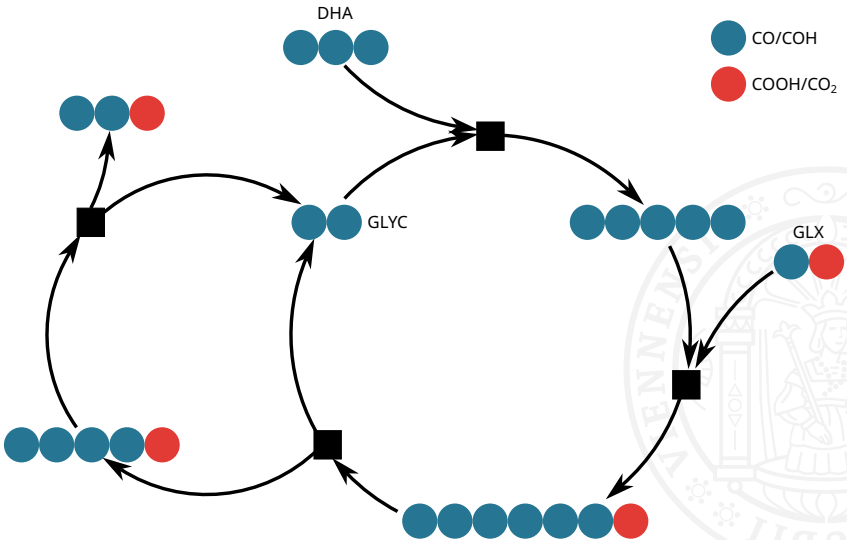
Types of AC-Cycles: "Build-Up" Cycle

Types of AC-Cycles: "Build-Up" Cycle



Types of AC-Cycles: "Recombination" Cycle

Types of AC-Cycles: "Recombination" Cycle



Outlook

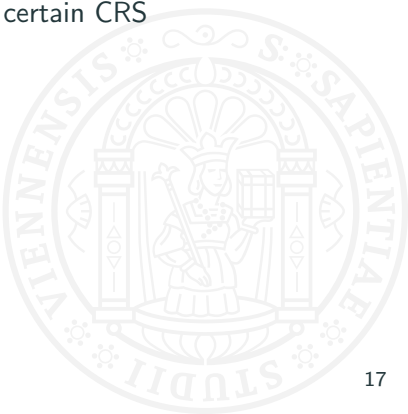
Outlook

Outlook

- Investigate and compare the CRSs associated to other types of chemistries

Outlook

- Investigate and compare the CRSs associated to other types of chemistries
- Find out how **specific** pathways could emerge out of a certain CRS



Outlook

- Investigate and compare the CRSs associated to other types of chemistries
- Find out how **specific** pathways could emerge out of a certain CRS
- Convert an expanded CRS into a **thermodynamic landscape** and investigate how it changes with different parameter choices

Outlook

- Investigate and compare the CRSs associated to other types of chemistries
- Find out how **specific** pathways could emerge out of a certain CRS
- Convert an expanded CRS into a **thermodynamic landscape** and investigate how it changes with different parameter choices
- Further contrast the role of "build-up" and "recombination" AC-cycles in OL and protometabolism

Thank you for your Attention!

Contact: nino.lauber@univie.ac.at

Thanks for the discussions:

Christoph Flamm,
Jakob L. Andersen,
Phillip Honegger,
Eric Smith,
Harrison B. Smith,
Ramanarayanan Krishnamurthy,
Charles L. Liotta

MATOMIC is funded by the Novo Nordisk Foundation, supported by grant NNF21OC0066551