



Randomized Sampling



LEARNING OBJECTIVES

Each student should be able to:

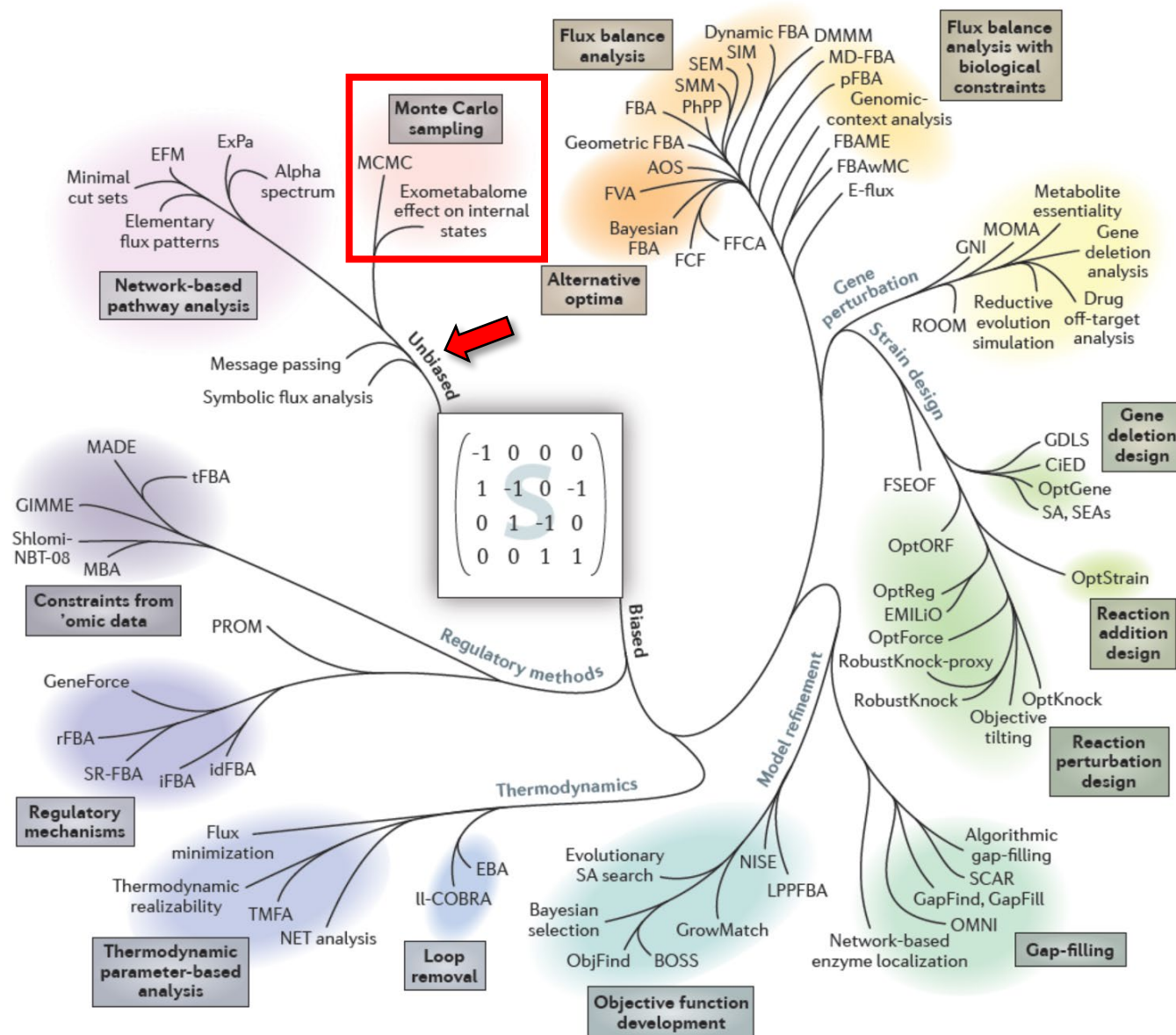
- Explain randomized sampling
- Explain solution space
- Demonstrate the ability to use the COBRApy tool for sampling
- Explain the role of a scatter matrix



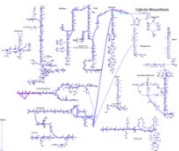
Lesson Outline

- Randomized Sampling
- Randomized Sampling Examples
- Scatter Matrix

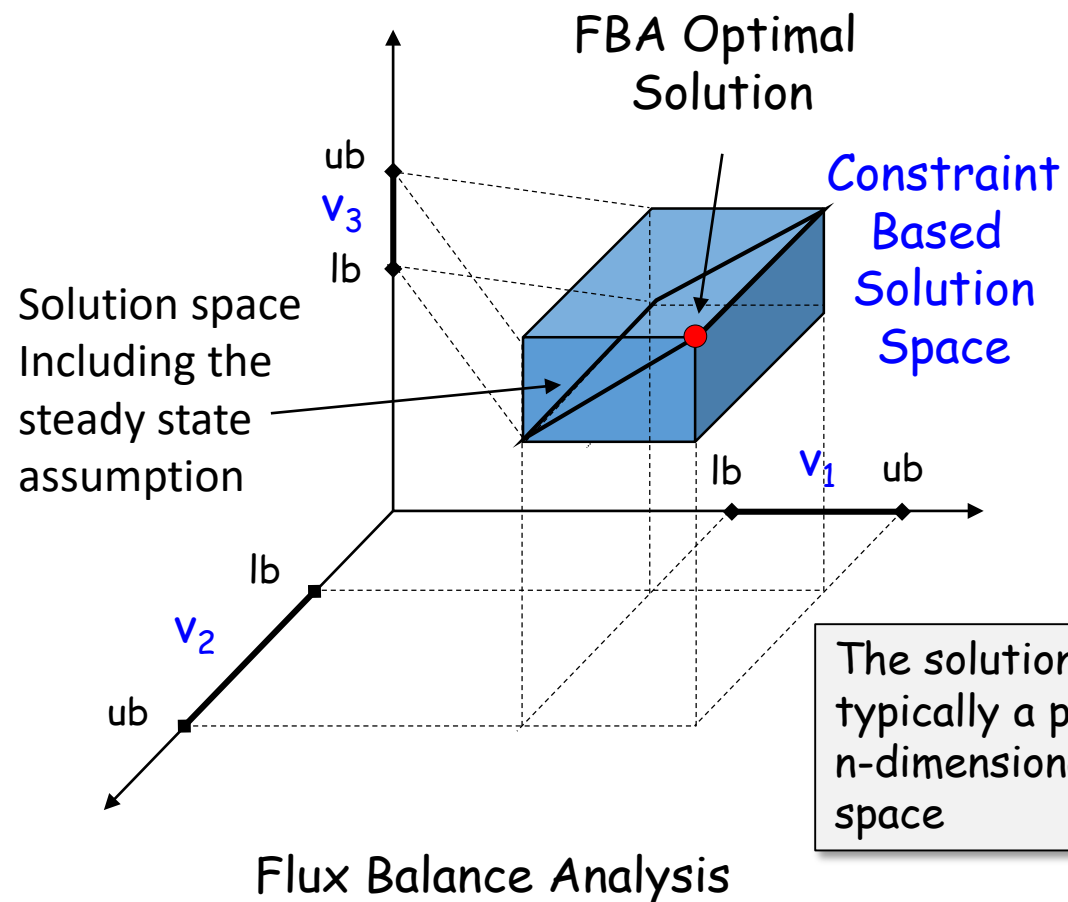
The 'Phylogeny' of Constraint-based Modeling Methods



Lewis, N. E., H. Nagarajan, et al. (2012). "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods." Nature reviews. Microbiology 10(4): 291-305.



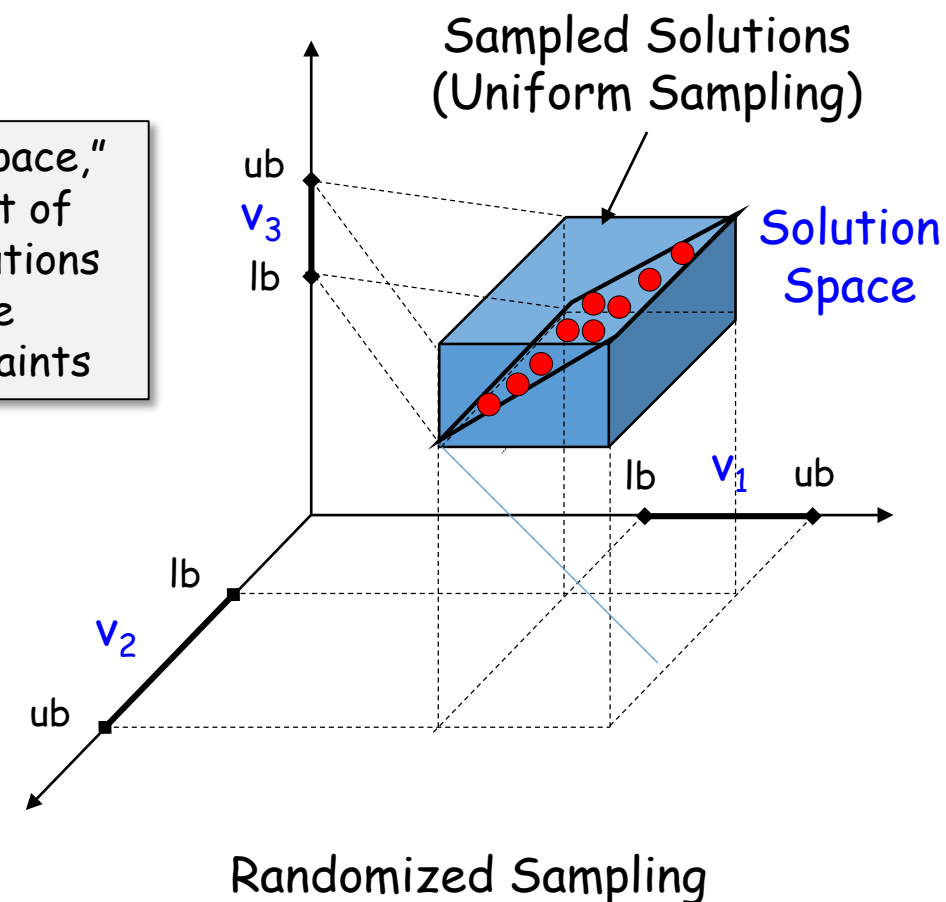
Solution Space



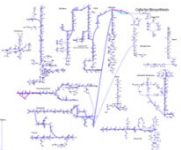
Identifies one solution from all solution space

The "solution space," contains the set of all feasible solutions that satisfy the imposed constraints

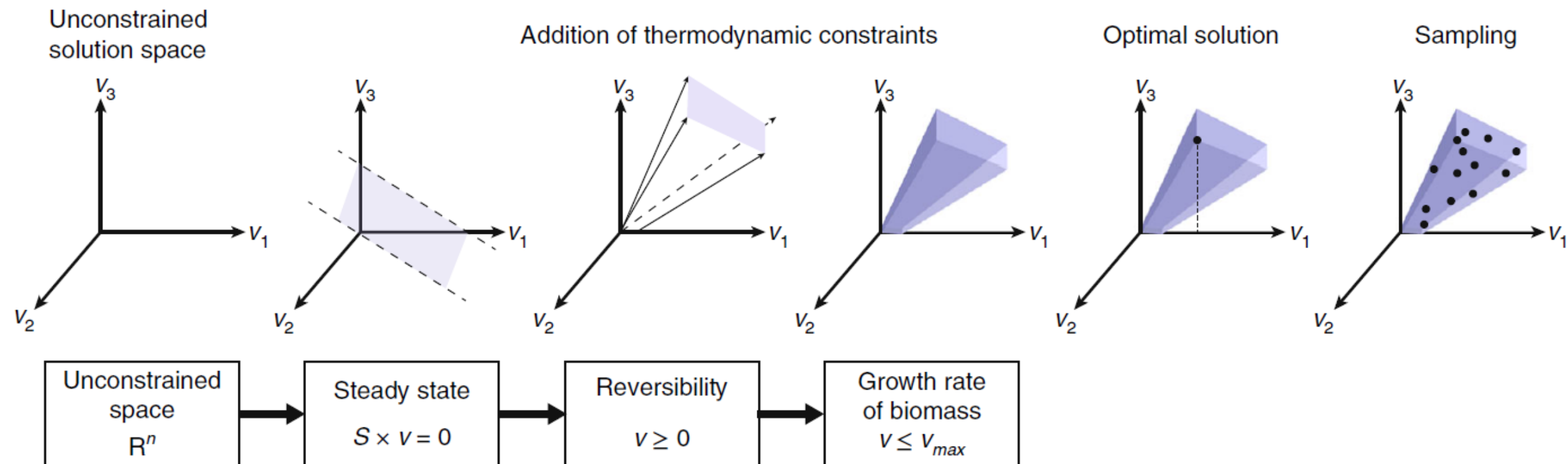
The solution space is typically a polytope in n-dimensional solution space



Multiple solutions representing all solution space



Solution Spaces

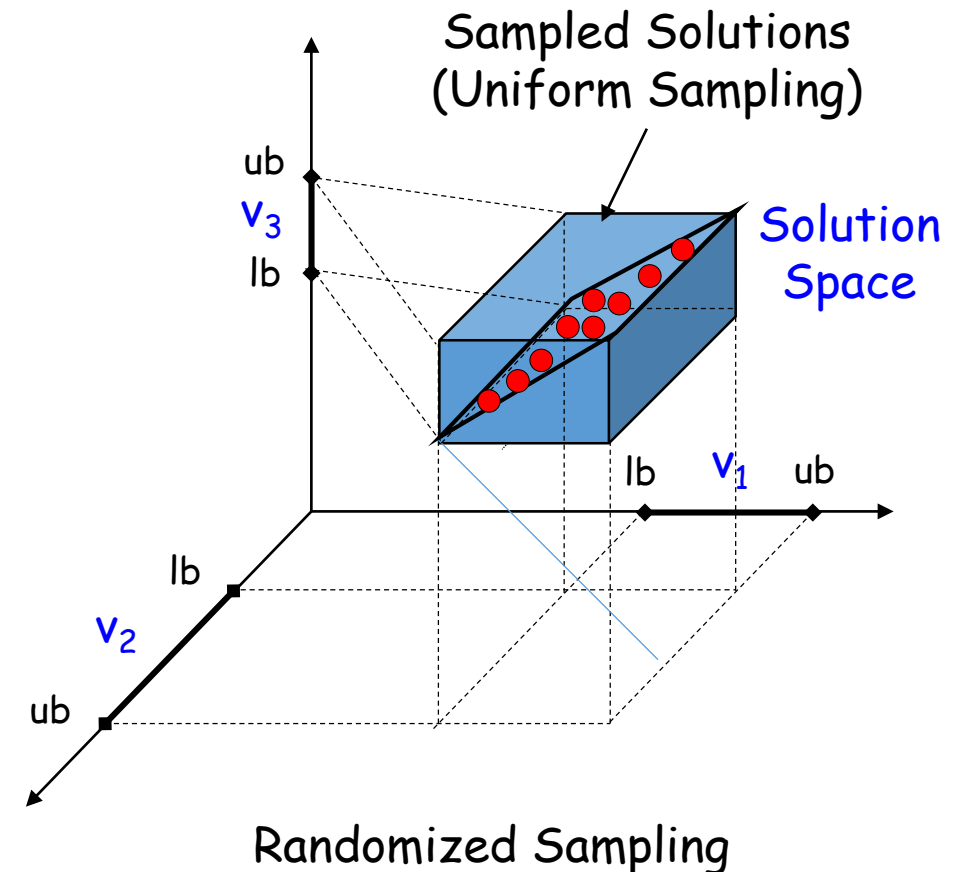


Laurent Heirendt et al, Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0, Nature Protocols, volume 14, pages 639-702, 2019

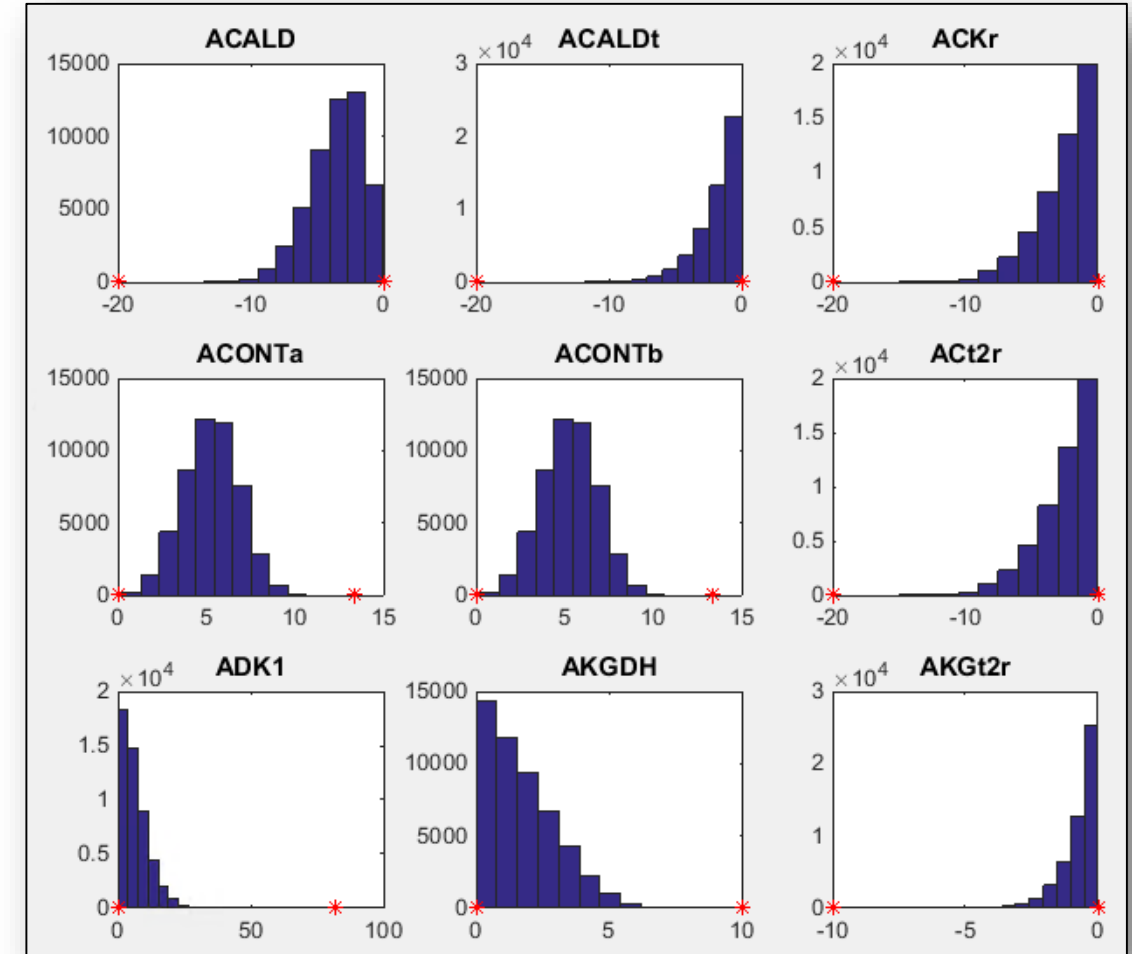
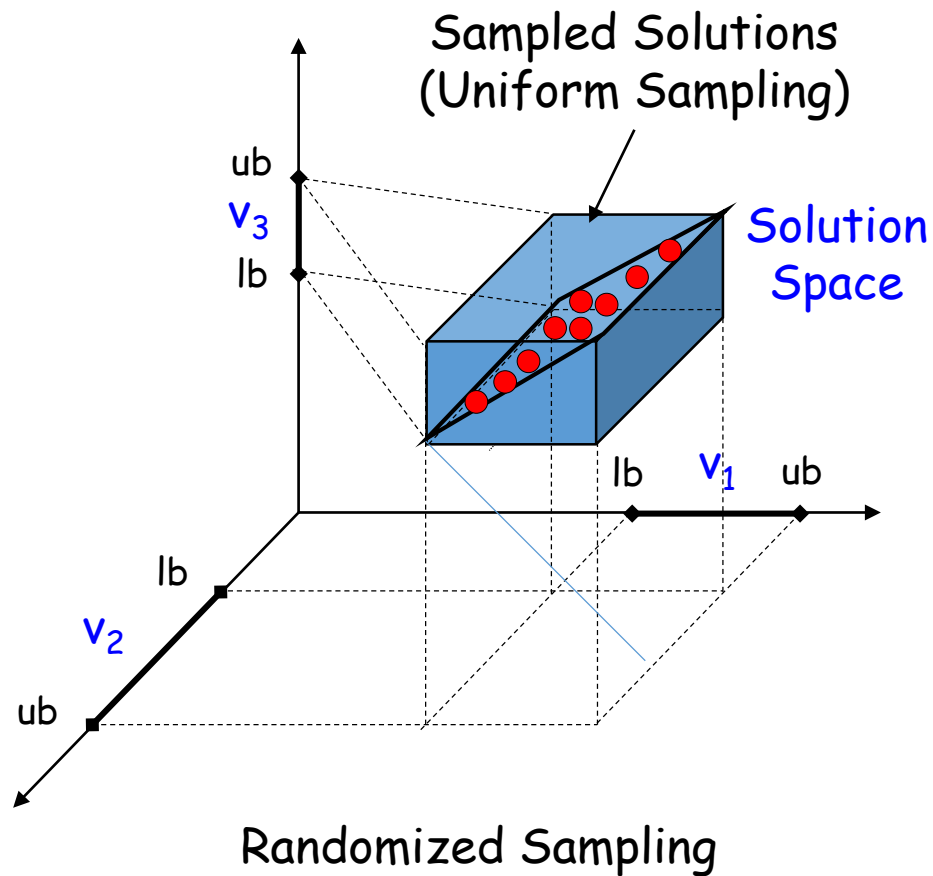


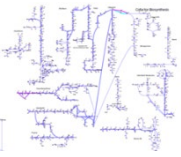
Uniform Random Sampling

- An alternative approach to characterizing the contents of a network's solution space is uniform random sampling.
- This approach involves obtaining a statistically meaningful number of solutions that have been uniformly distributed through the entire solution space.
- Randomized sampling of candidate network states throughout an entire solution space gives an unbiased assessment of its properties.
- The process of obtaining a uniform set of candidate solutions includes:
 1. Defining the space to be sampled based on the imposed constraints
 2. Randomly sampling it based on uniform statistical criteria
 3. Further segmenting the solution space based on additional post-sampling criteria as necessary.
- An objective function is not needed in sampling

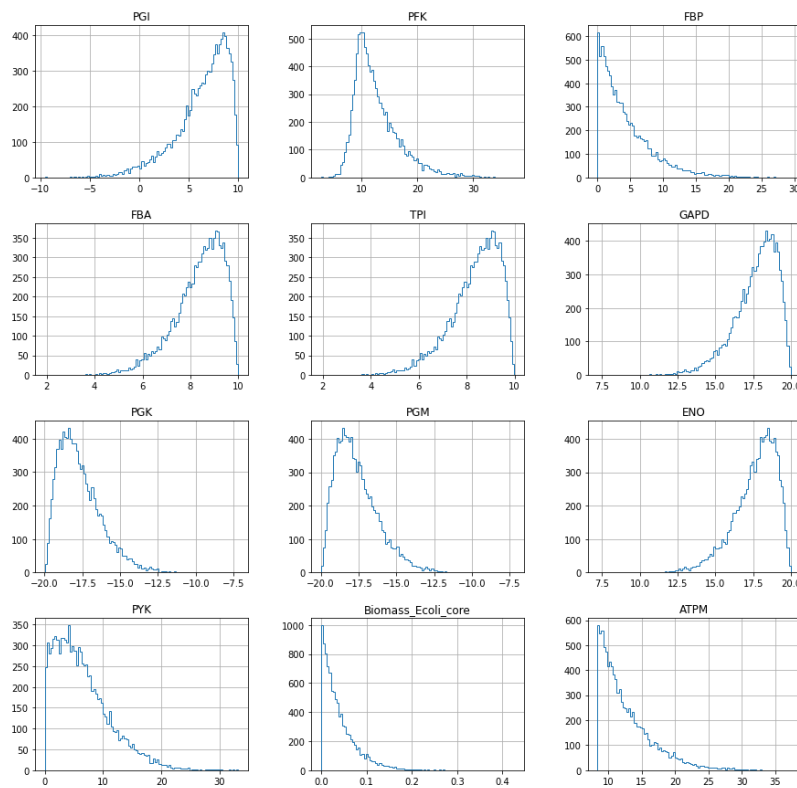


Sampled Flux Distributions

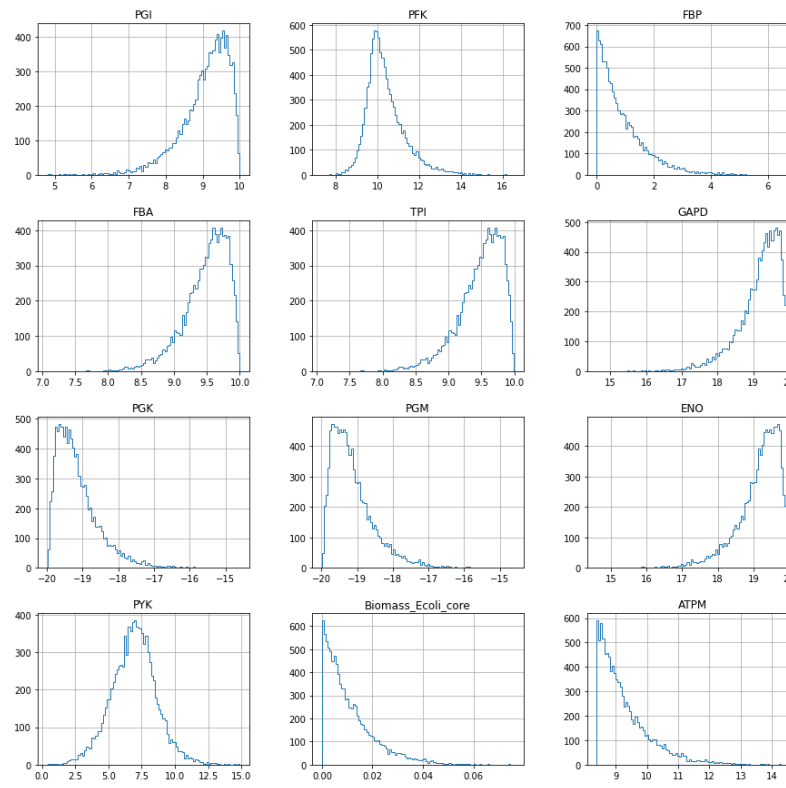




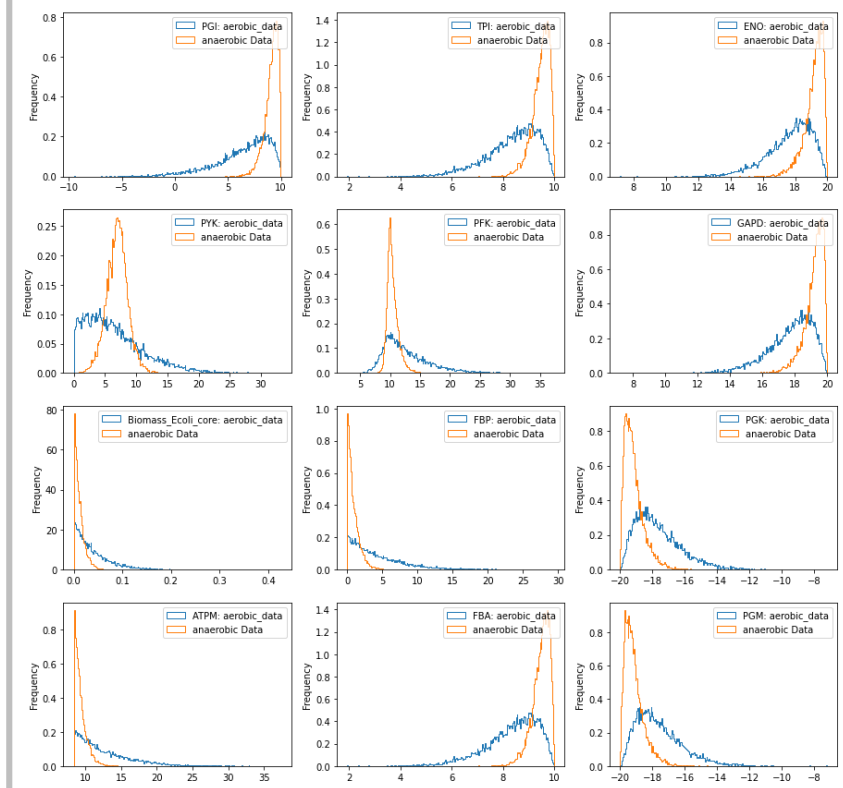
Sampling Images



Aerobic Samples



Anaerobic Samples

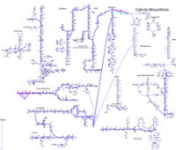


Combined Samples



Lesson Outline

- Randomized Sampling
- • Randomized Sampling Examples
- Scatter Matrix



Sampling Examples: Aerobic vs. Anaerobic

Sampling_aerobic_vs_anaerobic.ipynb

Sampling: Aerobic vs. Anaerobic

Loading the appropriate Python and COBRApy packages

```
In [1]: import cobra.test
from cobra.test import create_test_model
import numpy as np
import pandas as pd
from cobra.sampling import sample
import matplotlib.pyplot as plt
pd.set_option('display.max_rows', 500)
```

Set model environmental conditions

```
In [2]: model = create_test_model("textbook")
#model = cobra.io.load_json_model('./iJO1366.json') # Model must be in the same directory
model.reactions.EX_o2_e.lower_bound = -20
model.reactions.EX_glc__D_e.lower_bound = -10
```

Academic license - for non-commercial use only - expires 2022-10-10
Using license file C:\Users\hinton\gurobi.lic

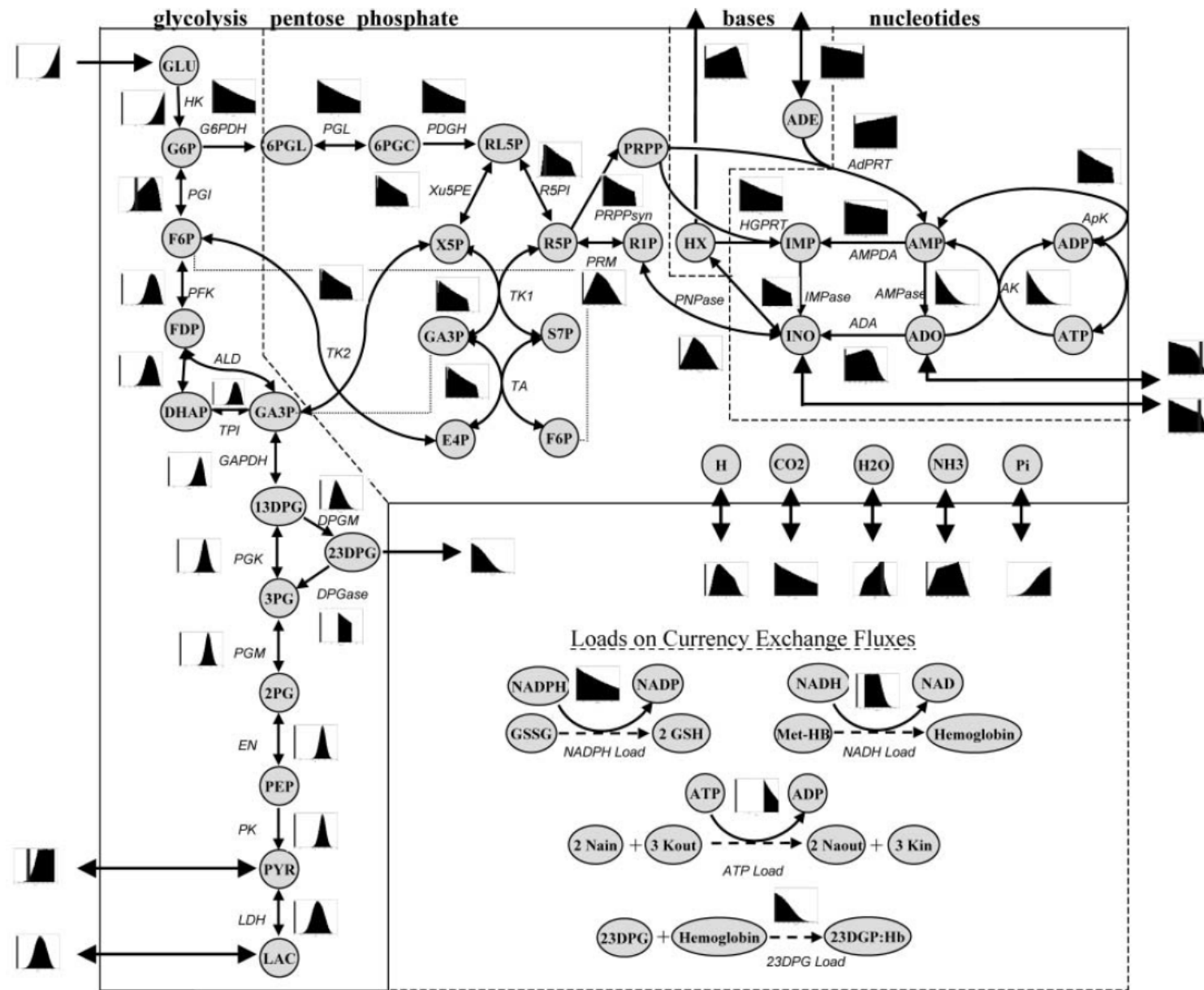
Sample aerobic model

```
In [3]: aerobic = sample(model, 10000, processes=8) #Set this to 1000 for faster, less accurate performance
aerobic
```

Probability Flux Distributions for Human Red Blood Cells

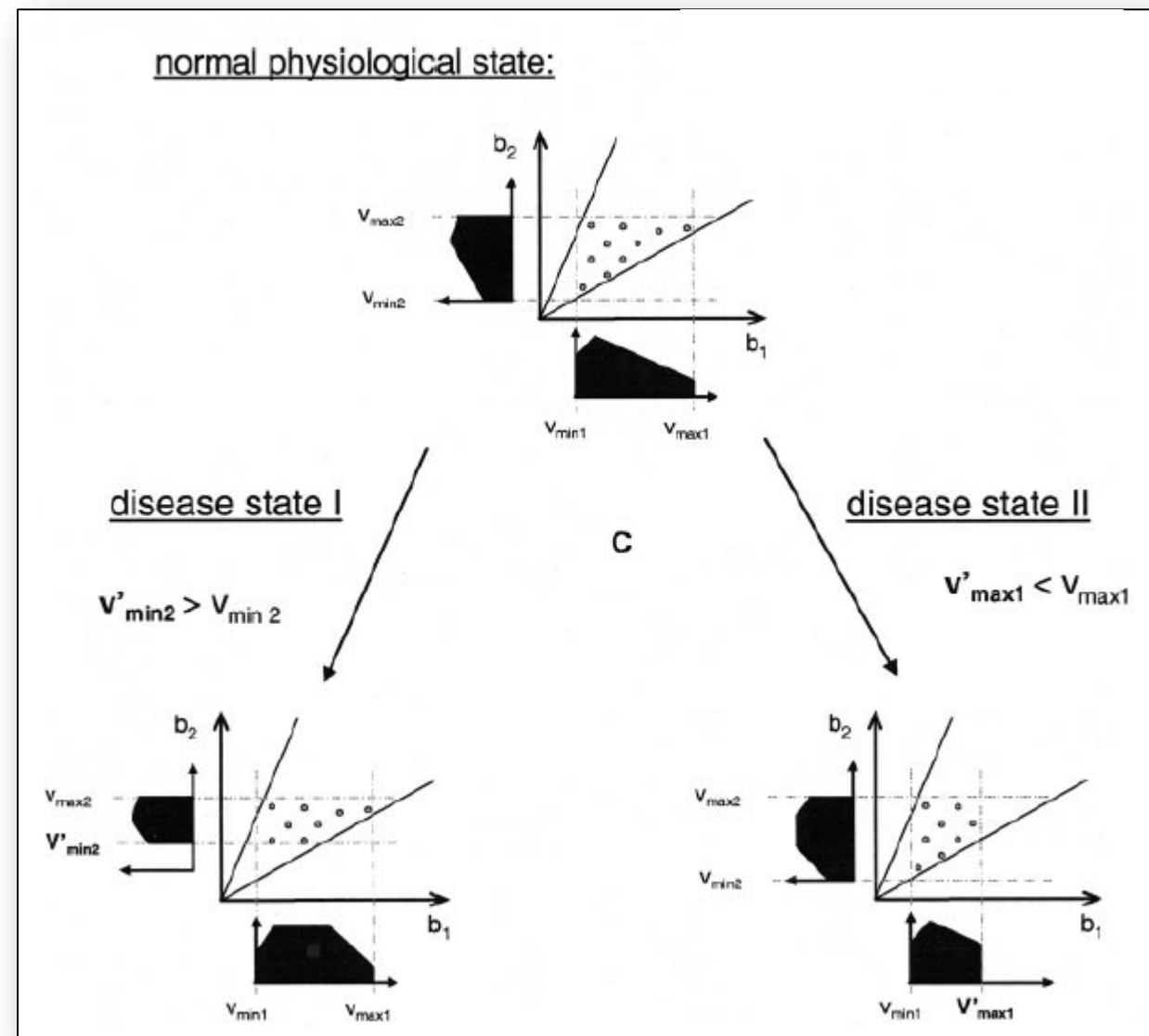
- The red blood cell model with imposed maximum and minimum constraints on each flux was sampled using the *in silico* algorithm.
- The histograms next to each reaction represent the distribution of solutions with respect to each reaction flux. The vertical shaded line on each plot indicates where the zero flux line is.
- Due to the convexity of the solution space, no distribution can have more than one peak.
- The flux distribution shape gives information about the sensitivity of the solution space to each constraint.
- If a flux distribution has a right peak, decreasing a maximum constraint will eliminate many solutions from the valid solution space.
- Reactions that are part of the same pathway with no intermediate branch points (PGM, EM, PK) all have the same flux distributions.
- Distributions shown are based on 500,000 uniformly distributed points in the steady-state flux space.

Price, N. D., J. Schellenberger, et al. (2004). "Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies." *Biophysical journal* 87(4): 2172-2186.



Metabolic Network States Under Normal And Disease Conditions

By applying constraints (V_{min} , V_{max}) to reaction, uptake, and secretion rates based on experimental data, the range of allowable steady states of the metabolic network consistent with the experimental data can be generated.



Thiele, I., N. D. Price, et al. (2005). "Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet." The Journal of biological chemistry 280(12): 11683-11695.



Red Blood Cell Sampling

RBC_sampling.ipynb

Red Blood Cell Sampling: Drugs

Set the environment

```
In [1]: import cobra.test
import pandas as pd
import numpy as np
import pandas as pd
import escher
from escher import Builder
from cobra.sampling import sample
import matplotlib.pyplot as plt
from cobrapy_bigg_client import client
pd.set_option('display.max_rows', 1000)
pd.set_option('display.width', 1000)
pd.set_option('display.max_colwidth', None)
```

Load the red blood cell model iAB_RBC_283.json

```
In [2]: model = client.download_model('iAB_RBC_283', save=False) # Loading the model to the simulation
```

Academic license - for non-commercial use only - expires 2022-10-10
Using license file C:\Users\hinton\gurobi.lic

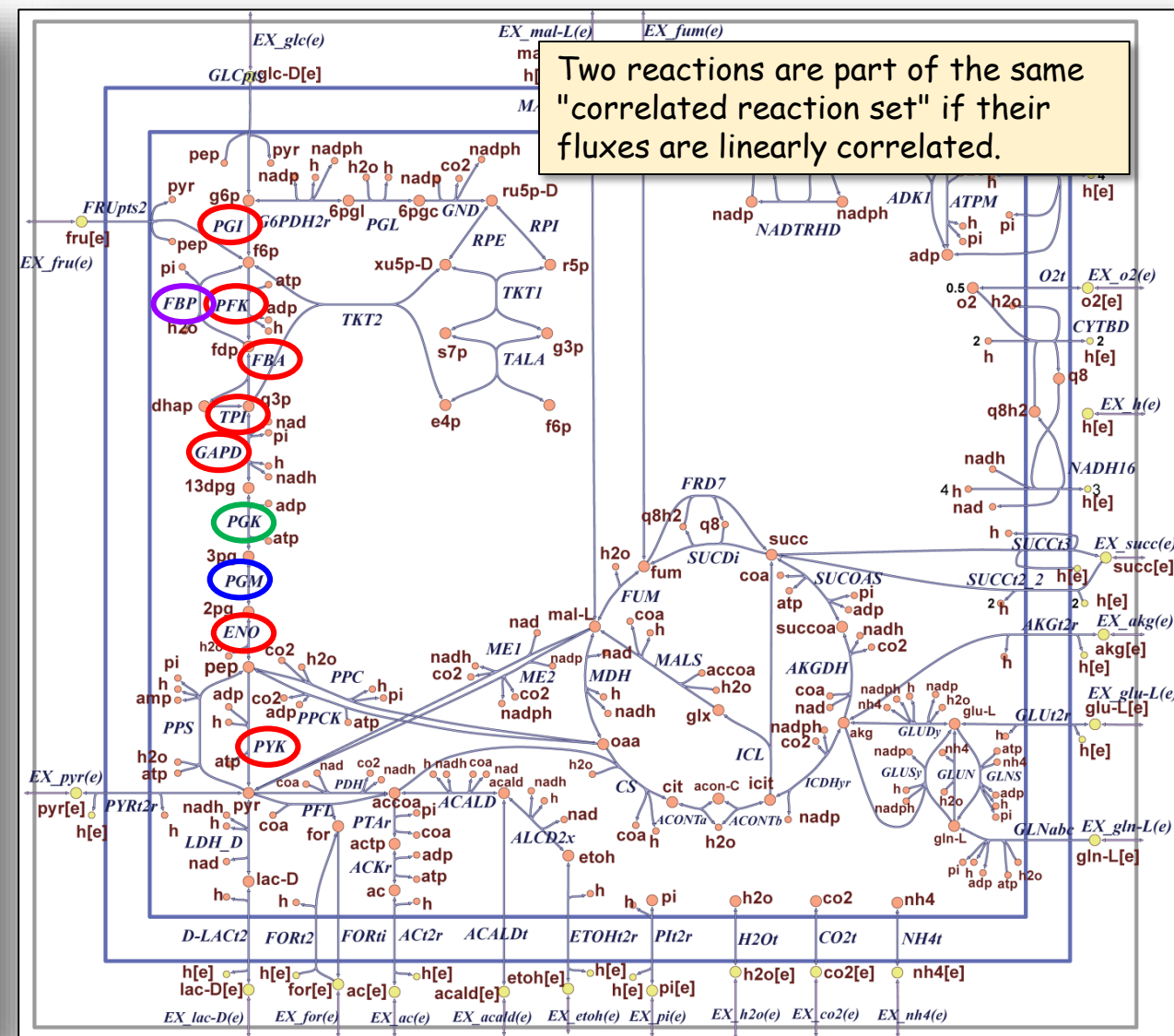
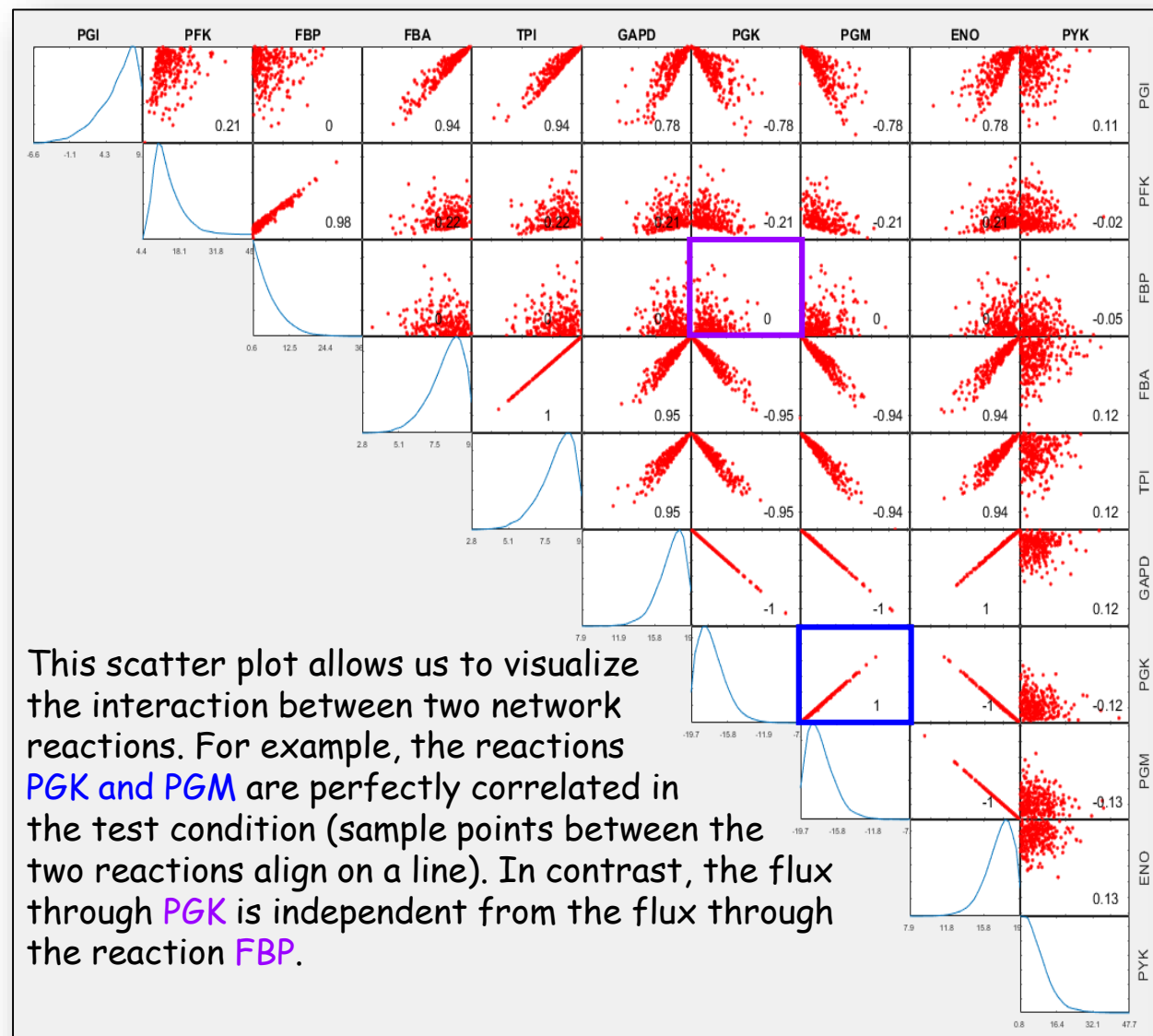
The model summary



Lesson Outline

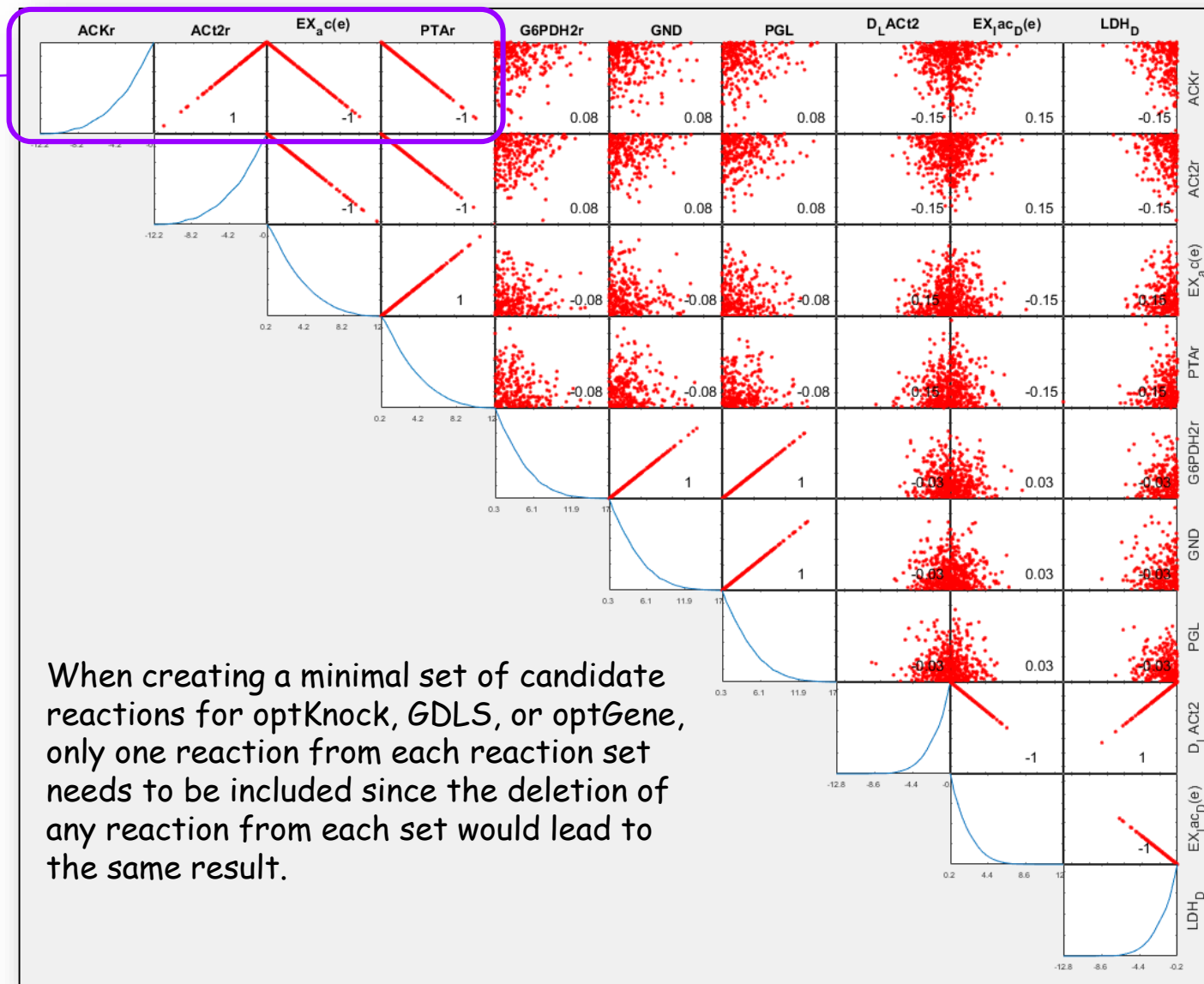
- Randomized Sampling
- Randomized Sampling Examples
- • Scatter Matrix

Scatter Matrices



Correlated Reaction Sets Scatter Matrix

Set#	Reactions		
[1]	'ACKr'	'ACT2r'	'EX_ac(e)'
[2]	'G6PDH2r'	'GND'	'PGL'
[3]	'D_LACT2'	'EX_lac_D(e)'	'LDH_D'
[4]	'CYTBD'	'EX_o2(e)'	'O2t'
[5]	'Biomass'	'EX_pi(e)'	'PIt2r'
[6]	'ALCD2x'	'ETOHt2r'	'EX_etoh(e)'
[7]	'ACONTa'	'ACONTb'	'CS'
[8]	'TALA'	'TKT1'	
[9]	'ICL'	'MALS'	
[10]	'GAPD'	'PGK'	
[11]	'FBA'	'TPI'	
[12]	'EX_pyr(e)'	'PYRt2r'	
[13]	'EX_nh4(e)'	'NH4t'	
[14]	'EX_h2o(e)'	'H2Ot'	
[15]	'EX_glu_L(e)'	'GLUt2r'	
[16]	'EX_glc(e)'	'GLCpts'	
[17]	'EX_for(e)'	'PFL'	
[18]	'ENO'	'PGM'	
[19]	'CO2t'	'EX_co2(e)'	
[20]	'AKGt2r'	'EX_akg(e)'	
[21]	'AKGDH'	'SUCOAS'	
[22]	'ADK1'	'PPS'	
[23]	'ACALDt'	'EX_acald(e)'	

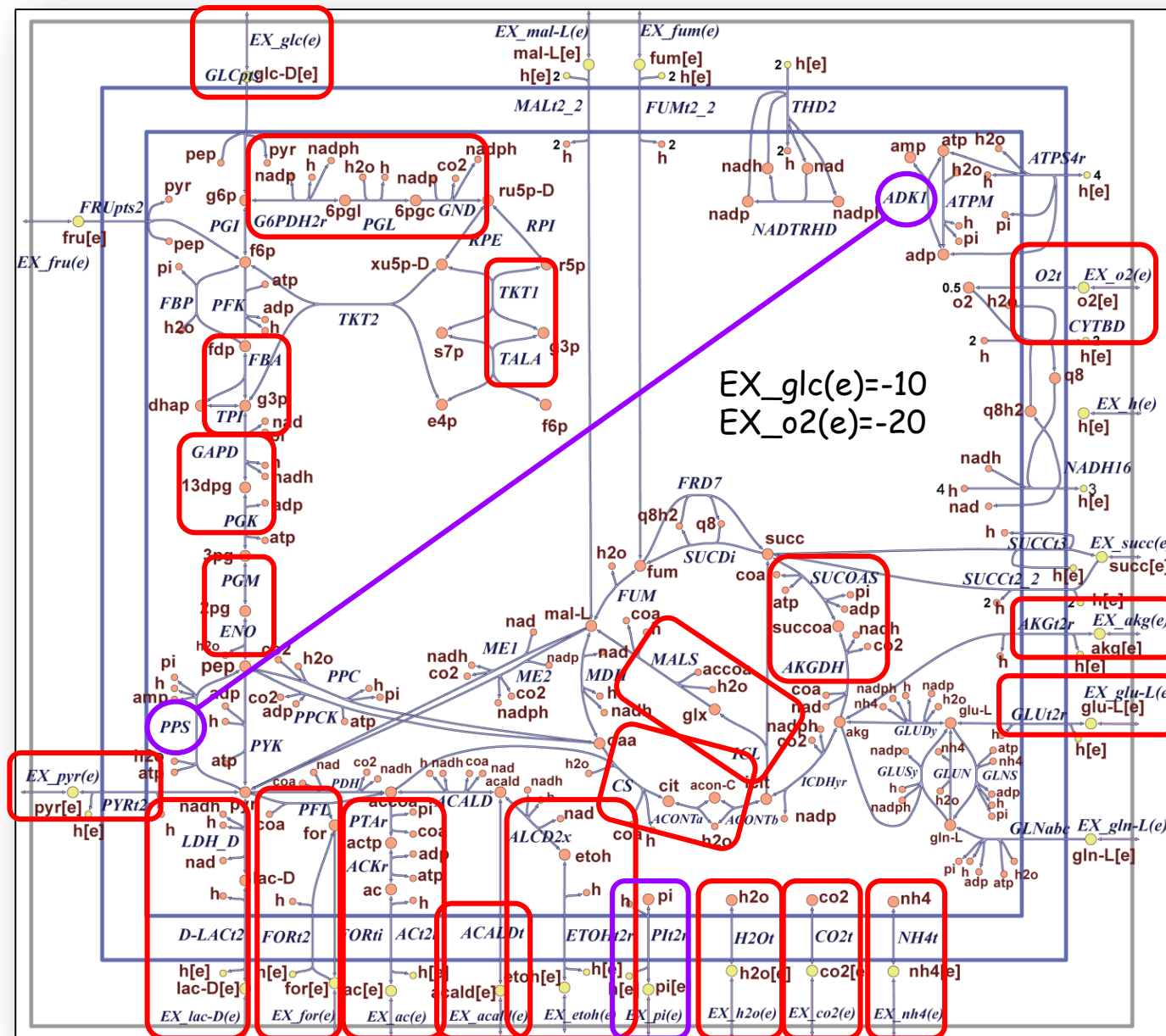


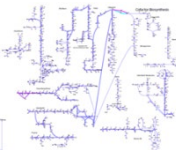
Mapped Correlated Reaction Sets

Set# Reactions

[1]	'ACKr'	'ACt2r'	'EX_ac(e)'	'PTAr'
[2]	'G6PDH2r'	'GND'	'PGL'	
[3]	'D_LACt2'	'EX_lac_D(e)'	'LDH_D'	
[4]	'CYTBD'	'EX_o2(e)'	'O2t'	
[5]	'Biomass'	'EX_pi(e)'	'PIt2r'	
[6]	'ALCD2x'	'ETOHt2r'	'EX_etoh(e)'	
[7]	'ACONTa'	'ACONTb'	'CS'	
[8]	'TALA'	'TKT1'		
[9]	'ICL'	'MALS'		
[10]	'GAPD'	'PGK'		
[11]	'FBA'	'TPI'		
[12]	'EX_pyr(e)'	'PYRt2r'		
[13]	'EX_nh4(e)'	'NH4t'		
[14]	'EX_h2o(e)'	'H2Ot'		
[15]	'EX_glu_L(e)'	'GLUt2r'		
[16]	'EX_glc(e)'	'GLCpts'		
[17]	'EX_for(e)'	'PFL'		
[18]	'ENO'	'PGM'		
[19]	'CO2t'	'EX_co2(e)'		
[20]	'AKGt2r'	'EX_akg(e)'		
[21]	'AKGDH'	'SUCOAS'		
[22]	'ADK1'	'PPS'		
[23]	'ACALDt'	'EX_acald(e)'		

IdentifyingCorrelSets.m





Scatter Matrices

Sampling_scatter_matrix.ipynb

Sampling: Scatter Matrix

Loading the appropriate Python and COBRApy packages

```
In [1]: import cobra.test
        from cobra.test import create_test_model
        import numpy as np
        import pandas as pd
        from cobra.sampling import sample
        import matplotlib.pyplot as plt
        pd.set_option('display.max_rows', 500)
```

Set model environmental conditions

```
In [2]: model = create_test_model("textbook")
        #model = cobra.io.load_json_model('./iJO1366.json') # Model must be in the same directory
        model.reactions.EX_o2_e.lower_bound = -20
        model.reactions.EX_glc__D_e.lower_bound = -10
```

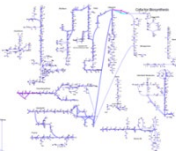
Sample aerobic model

```
In [3]: aerobic = sample(model, 10000, processes=8) #Set this to 1000 for faster, less accurate performance
        aerobic
```



Lesson Outline

- Randomized Sampling
- Randomized Sampling Examples
- Scatter Matrix



Reflective Questions

1. What are correlated reaction sets?
2. What is the solution space?
3. Is randomized sampling classified as biased or unbiased assessment?
4. What role does the objective function play in randomized sampling?
5. What role do reaction constraints play in the accuracy of the data generated by randomized sampling?
6. What is the difference between an optimal solution and a sampled solution?
7. How many peaks can each reactions distribution have?
8. What will happen if the maximum constraint is decreased for a reaction that has a peak on the right side of the distribution?
9. What is the relationship between the distributions of reactions that are part of the same pathway with no intermediate branch points?
10. What is the purpose of a scatter matrix?



References

1. Lewis, N. E., H. Nagarajan, et al. (2012). "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods." *Nature reviews. Microbiology* 10(4): 291-305.
2. Jan Schellenberger, PhD Dissertation, University of California, San Diego, 2010
3. Laurent Heirendt et al, Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0, *Nature Protocols*, volume 14, pages 639-702, 2019
4. B. O. Palsson, *Systems Biology: Constraint-based Reconstruction and Analysis*, Cambridge University Press, 2015
5. Price, N. D., J. Schellenberger, et al. (2004). "Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies." *Biophysical journal* 87(4): 2172-2186.
6. Thiele, I., N. D. Price, et al. (2005). "Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet." *The Journal of biological chemistry* 280(12): 11683-11695.

iAB_RBC_283 Red Blood Cell Attributes

iAB_RBC_Attributes.ipynb

Set the environment

```
In [1]: import cobra.test
import pandas as pd
import numpy as np
import pandas as pd
import escher
from escher import Builder
from cobra.sampling import sample
import matplotlib.pyplot as plt
from cobrapy_bigg_client import client
pd.set_option('display.max_rows', 1000)
pd.set_option('display.width', 1000)
pd.set_option('display.max_colwidth', None)
```

Load the red blood cell model iAB_RBC_283.json

```
In [2]: model = client.download_model('iAB_RBC_283', save=False) # Loading the model to the simulation
```

Set parameter Username

Academic license - for non-commercial use only - expires 2022-10-10

The model summary

```
In [3]: model
```

Mapped Correlated Reaction Sets

Set# Reactions

[1]	'ACKr'	'ACt2r'	'EX_ac(e)'	'PTAr'
[2]	'G6PDH2r'	'GND'	'PGL'	
[3]	'D_LACt2'	'EX_lac_D(e)'	'LDH_D'	
[4]	'CYTBD'	'EX_o2(e)'	'O2t'	
[5]	'Biomass'	'EX_pi(e)'	'PIt2r'	
[6]	'ALCD2x'	'ETOHt2r'	'EX_etoh(e)'	
[7]	'ACONTa'	'ACONTb'	'CS'	
[8]	'TALA'	'TKT1'		
[9]	'ICL'	'MALS'		
[10]	'GAPD'	'PGK'		
[11]	'FBA'	'TPI'		
[12]	'EX_pyr(e)'	'PYRt2r'		
[13]	'EX_nh4(e)'	'NH4t'		
[14]	'EX_h2o(e)'	'H2Ot'		
[15]	'EX_glu_L(e)'	'GLUt2r'		
[16]	'EX_glc(e)'	'GLCpts'		
[17]	'EX_for(e)'	'PFL'		
[18]	'ENO'	'PGM'		
[19]	'CO2t'	'EX_co2(e)'		
[20]	'AKGt2r'	'EX_akg(e)'		
[21]	'AKGDH'	'SUCOAS'		
[22]	'ADK1'	'PPS'		
[23]	'ACALDt'	'EX_acald(e)'		

