2023

AMN Module 3

PATHWAY THERMODYNAMICS

TIMMY PAEZ WATSON

Contents

1 Introduction

What defines a living system? At this stage in your academic life one would expect that the answer to this question is simple. However, answers to this question have deep philosophical roots that go beyond the realms of life sciences. One such answer states that any system that actively extracts energy from the environment to avoid the natural state of the universe (chaos) is living. Living systems are the only systems that move towards organization as opposed to the disorder of entropy. How do cells do this? Simply by using the energy in its environment. It all boils down to thermodynamics.

In this module, we will deal with equilibrium thermodynamics to evaluate metabolic pathways. The tools we will learn in this module can be applied to any metabolic pathway and are useful to understand:

- Reversibility of metabolic reactions.
- Driving forces required for individual pathways.
- Rection bottlenecks, leading to metabolic points of control.
- Metabolic engineering targets.

Importantly, these results should always be critically examined. Equilibrium thermodynamics is based on the concept that any system in the universe must slowly settle towards an equilibrium characterized by a minimum energy, a maximum of entropy and remain like that forever. Living systems instead are able to generate mechanical or electrical power and (shockingly) self-replicate! This shows that living systems somehow manage to break this rule and stay far away from equilibrium.

For our analysis, we must revisit two fundamental laws of thermodynamics.

2 First law: Conservation of energy

Of importance to pathway thermodynamics, the first law states that when two initially isolated systems are combined into a new system, the total energy of the system will be equal to the sum of the internal energies of the two initial systems. Plainly put, given the following reversible reaction.

$$A + B \leftrightarrow C + D$$

The standard free energy of the system $\Delta_r G'^o$ is given by:

$$\Delta_r G'^o = -\Delta_f G_A'^o - \Delta_f G_B'^o + \Delta_f G_C'^o + \Delta_f G_D'^o$$

where $\Delta_f G_A'^o$ is the standard free energy of formation of metabolite A (or B, C, D, etc). This equation basically defines that the total energy in this system is conserved and equals to the individual contribution of each metabolite (with it's sign depending on the consumption/generation). The "*standard*" energy (denoted with ^{o}) is the formation energy of a compound in standard conditions without accounting for pH, ionic strength or any other factors (they are all taken as 0, hence the ^{o}). Below is a list of the different types of formation energies and what conditions they make reference to:

- $\Delta_f G_i^o$ is the standard energy of formation of *i* at 1 M, not considering pH, ionic strength, etc.
- $\Delta_f G_i'^o$ is the standard energy of formation of *i* at 1 M, given pH = 7 and ionic strength of 0.1 M.

- $\Delta_f G_i'^m$ is the energy of formation of *i* at 1 mM. This value is used quite a lot in biology since this is the expected concentration range of metabolites.

  Important! Water is an exception here, as it is kept at 1 M.

And similarly with the energy of the reactions:

- $\Delta_r G_j^o$ is the standard free energy of reaction *j* at 1 M for all species, not considering pH, ionic strength, etc.
- $\Delta_r G_j'^o$ is the standard free energy of reaction *i* at 1 M for all species, given pH of 7 and ionic strength of 0.1 M.
- $\Delta_r G_j'^m$ is the free energy of reaction *j* at 1 mM for all species (except water!). This value is used quite a lot in biology since this is the expected concentration range of metabolites.

2.1 ATP hydrolysis

Now that the nomenclature is clear, we can examine the free energy of a reaction. For example, given the following reaction of ATP hydrolysis:

$$ATP + H_2O \leftrightarrow ADP + Pi$$

The standard free energy of the system $\Delta_r G'^o$ is given by:

$$\Delta_r G'^o = -\Delta_f G'^o_{ATP} - \Delta_f G'^o_{H2O} + \Delta_f G'^o_{ADP} + \Delta_f G'^o_{Pi} = -29.6 \frac{kJ}{mol}$$

The hydrolysis of ATP thus generates approximately 30 kJ of energy per mol of ATP! Or probably even more given the conditions in the cell. Let's have a look by calculating the $\Delta_r G'^m$:

$$\Delta_r G'^m = \Delta_r G'^o + RT \ln(Q), \qquad Q = \frac{[ADP]*[Pi]}{[ATP]*[H2O]}$$

$$\Delta_r G'^m = -29.6 \frac{kJ}{mol} + 8.314 * 10^{-3} \frac{kJ}{mol * K} * 303.15\ K \ln\left(\frac{10^{-3}\text{M} * 10^{-3}\text{M}}{10^{-3}\text{M} * 1\text{M}}\right)$$

$$\Delta_r G'^m = -46.8 \frac{kJ}{mol}$$

Indeed, an additional 17 kJ/mol of energy are released from ATP hydrolysis when considering cellular-like concentrations (1 mM). Note that the concentration of H2O was kept at 1 M, being that the main responsible for the increase in $\Delta_r G'^m$. All this energy can be either dissipated as heat (*e.g.* some flowers give use the heat of ATP hydrolysis to make their fragrant molecules volatile) or used in other biochemical reactions. We will explore this combination of reactions in subsequent chapters.

3 Second law: Irreversibilities

The second law of thermodynamics states that when two systems interact with each other, they will reach mutual thermodynamic equilibrium. The initial state of these individual systems will dictate the directionality of energy, heat or mass transfer. In other words, and with special focus on biochemistry, when looking at a chemical reaction:

$$A + B \leftrightarrow C + D$$

The directionality of this reaction is dictated by the $\Delta_r G$. In principle, all chemical reactions are reversible (as stated in the thermodynamics law!), and the direction of the flux is dependent on the free energy of the reaction. Since this $\Delta_r G$ of any given reaction is influenced by the concentrations of substrates and products, we could calculate $\Delta_r G_{min}$ and $\Delta_r G_{max}$ influenced by the $[A]_{max}$, $[A]_{min}$, and subsequently for B, C and D. Thus, the directionality of any given reaction could be summarized as:

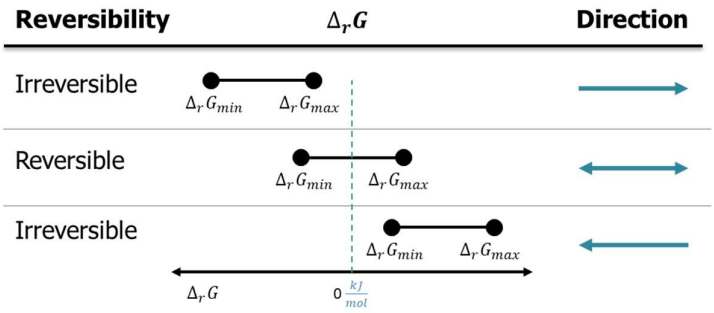

Figure 1. Schematic representation of the reversibility and directionality of chemical reactions as a consequence of their $\Delta_r G$.

The reversibility (or irreversibility) of the reactions is dependent on the chosen values for $[A]_{max}$, $[A]_{min}$, etc. For metabolic reactions, we can estimate these ranges based on the volume of a typical cell and experimentally measured ranges of metabolites. Commonly, ranges from 1 µM – 10 mM are used in pathway thermodynamic analysis.

Smaller concentrations are indeed possible, but they are limits. For example, concentrations of 1.6 nM in a typical bacterial cell (1 µm³) is equals to 1 molecule per cell. This is too low to be possible, unless there is a reaction consuming that metabolite with extremely high $-\Delta_r G$.

3.1 Acetyl CoA <-> Acetyl phosphate

Let us now explore a biochemical reaction that is close to equilibrium. Cells that can uptake acetate need to activate it with either CoA or phosphate, forming a thioester or ester bond respectively. These bonds are very strong to break, hence AcetylCoA or Acetyl_phosphate are analogous to ATP in the sense that freeing that additional group will release a lot of energy. Now, we will study the interconversion of these molecules:

$$Acetyl\ CoA + Pi \leftrightarrow Acetyl\ phosphate + CoA, \qquad \Delta_r G'^{o} = \Delta_r G'^{m} = 8.7\ \frac{kJ}{mol}$$

Now, we assess the directionality of this reaction by looking at the $\Delta_r G$. Say, for example, that the concentrations of phosphate in bacterial cells are allowed to change from 1 to 100 mM (accepted range for inorganic phosphate). This would lead to changes in the $\Delta_r G$ as seen in Figure 2.

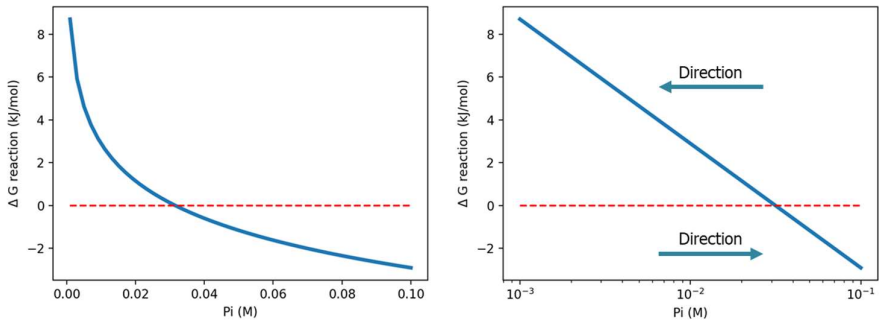

Figure 2. $\Delta_r G$ with varying phosphate concentrations. All other metabolites were kept at 1 mM. Plots on the left and right are the same, except that the *x-axis* has been transformed to logarithmic scale. Since the $\Delta_r G$ is affected by the ln(Q), this visualization is preferred in thermodynamics.

Similarly, the analysis can be done with changing the concentration of CoA (Figure 3). CoA concentrations in the cells range between 0.01 – 10 mM.

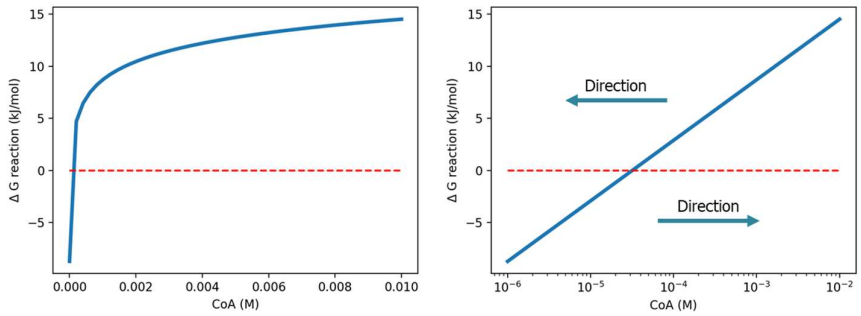

Figure 3. $\Delta_r G$ with varying CoA concentrations. All other metabolites were kept at 1 mM. Plots on the left and right are the same, except that the *x-axis* has been transformed to logarithmic scale. Since the $\Delta_r G$ is affected by the ln(Q), this visualization is preferred in thermodynamics.

And even a 2D analysis on the concentration of both CoA and Pi can be done (Figure 4).

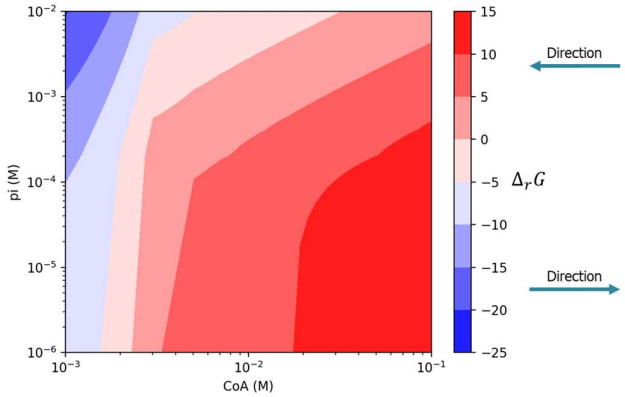

Figure 4. $\Delta_r G$ with varying CoA and Pi concentrations. All other metabolites were kept at 1 mM.

Work with the jupyter notebook file

You can work with this example in Python (Jupyter notebook format) on Brightspace (M4.1 Reaction thermodynamics).

```python
# Constants
R_value = 8.314e-3     # kJ/mol/K
T_value = 303.15       # K

# Standard dG of formation (from Equilibrator)
dG_AcCoA   = -1762.2  # KJ/mol
dG_Pi      = -1055.5  # KJ/mol
dG_AcetylP = -1099.3  # KJ/mol
dG_CoA     = -1709.7  # KJ/mol


phys_conc = 10**-3     # 1 mM

# Calculate DG of the reaction
dGr_0 = dG_AcetylP + dG_CoA -dG_AcCoA -dG_Pi   # Standard conditions
dGr_m = dGr_0 + R_value * T_value * np.log( (phys_conc*phys_conc)/(phys_conc*phys_conc) )
```

```python
# Loop over concentrations for Pi (1 - 100 mM)
pi_conc = np.linspace(10**-3, 10**-1, 50)
dGr_r = np.zeros(np.size(pi_conc))    # Empty list to append the results

for i, pi in enumerate(pi_conc):
    dGr_r[i] = dGr_0 + R_value * T_value * np.log( (phys_conc*phys_conc)/(pi*phys_conc) )

plt.figure( figsize = (12, 4))
plt.subplot(1,2,1)
plt.plot(pi_conc, dGr_r, linewidth = 3); plt.xlabel('Pi (M)'); plt.ylabel('$\Delta$ G reaction (kJ/mol)')
plt.plot(pi_conc, np.zeros(np.size(pi_conc)),'--r')
plt.subplot(1,2,2)
plt.plot(pi_conc, dGr_r, linewidth = 3); plt.xlabel('Pi (M)'); plt.ylabel('$\Delta$ G reaction (kJ/mol)')
plt.plot(pi_conc, np.zeros(np.size(pi_conc)),'--r')
plt.xscale('log')

plt.savefig('AcetylCoa_phosphate changes.png',dpi=200)
```

```python
# Loop over concentrations for CoA (0.01 - 10 mM)
coa_conc = np.linspace(10**-6, 10**-2, 50)
dGr_r = np.zeros(np.size(coa_conc))    # Empty list to append the results
for i, coa in enumerate(coa_conc):
    dGr_r[i] = dGr_0 + R_value * T_value * np.log( (coa*phys_conc)/(phys_conc*phys_conc) )

plt.figure( figsize = (12, 4))
plt.subplot(1,2,1)
plt.plot(coa_conc, dGr_r, linewidth = 3); plt.xlabel('CoA (M)'); plt.ylabel('$\Delta$ G reaction (kJ/mol)')
plt.plot(coa_conc, np.zeros(np.size(pi_conc)),'--r')
plt.subplot(1,2,2)
plt.plot(coa_conc, dGr_r, linewidth = 3); plt.xlabel('CoA (M)'); plt.ylabel('$\Delta$ G reaction (kJ/mol)')
plt.plot(coa_conc, np.zeros(np.size(pi_conc)),'--r')
plt.xscale('log')

plt.savefig('AcetylCoa_coa changes.png',dpi=200)
```

```python
# Now run the nested loop
dG_reaction = np.zeros([np.size(pi_conc),np.size(coa_conc)])       # Storage array

# Loop over concentrations for Pi
for i, pi in enumerate(pi_conc):
    # Loop over concentrations of CoA
    for j, coa in enumerate(coa_conc):
        dG = dGr_0 + R_value * T_value * np.log( (phys_conc*coa)/(pi*phys_conc) )
        dG_reaction[i,j] = dG

X, Y = np.meshgrid(pi_conc, coa_conc)
plt.contourf(X, Y, dG_reaction, cmap='bwr')
plt.xscale('log')
plt.yscale('log')
plt.ylabel('pi (M)'), plt.xlabel('CoA (M)')
plt.colorbar()
plt.savefig('AcetylCoA 2D changes.png',dpi=200)

plt.show()
```

4 Pathway thermodynamics

As outlined previously, thermodynamics link fundamental physical properties to metabolic pathways by means of individual reactions free energy calculations. By applying these methods, one can assume a range of metabolite concentrations and infer the direction and reversibility (or not) of individual reactions. However, metabolism does not consist of multiple independent reactions. Metabolic networks are strongly interwoven, with products of each reaction being the substrates for the subsequent reactions. What is more, several metabolites are shared with numerous reactions (*e.g.* ATP, NADH, CoA, etc.). Thus, a global analysis integrating metabolism is warranted.

Given the following pathway:

$$A \overset{1}{\leftrightarrow} B \overset{2}{\leftrightarrow} C \overset{3}{\leftrightarrow} D$$

One could calculate the standard free energy of each reaction as:

$$\Delta_r G_1'^o = \Delta_f G_B'^o - \Delta_f G_A'^o$$

$$\Delta_r G_2'^o = \Delta_f G_C'^o - \Delta_f G_B'^o$$

$$\Delta_r G_3'^o = \Delta_f G_D'^o - \Delta_f G_C'^o$$

We can calculate these values individually for small pathways, but what happens when dealing with pathways of hundreds of reactions? An easier way to perform this calculation is to use the stoichiometric matrix:

$$S = \begin{array}{c} \begin{matrix} v1 & v2 & v3 \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 \\ +1 & -1 & 0 \\ 0 & +1 & -1 \\ 0 & 0 & +1 \end{pmatrix} \end{array} \begin{matrix} \rightarrow A \\ \rightarrow B \\ \rightarrow C \\ \rightarrow D \end{matrix} \qquad \Delta_f G_i'^o = \begin{pmatrix} \Delta_f G_A'^o \\ \Delta_f G_B'^o \\ \Delta_f G_C'^o \end{pmatrix}$$

And the standard free energies of each reaction is given by:

$$\Delta_r G'^o = S^T \cdot \Delta_f G_i'^o$$

And following the same logic, the calculation of $\Delta_r G'$ for each reaction:

$$\Delta_r G_i' = \Delta_r G_i'^o + RT.\ln(Q_i)$$

Can be transformed to the calculation of $\Delta_r G'$ for the whole pathway:

$$\Delta_r G' = \Delta_r G'^o + RT \cdot S^T \cdot x$$

With 'x' being a vector containing the natural logarithm concentrations of each metabolite in the S matrix.

4.1 Pathway thermodynamics of glycolysis

Lets consider one of the most common pathways in biology: Embden-Meyerhof-Parnas (EMP) glycolysis. That is the *"typical"* glycolysis from textbooks. For this section, we will consider the following pathway (Figure 5) that includes all glycolytic reactions from glycogen up until pyruvate. Further this pathway consists of 4 reactions more to generate hydroxybutyrate.

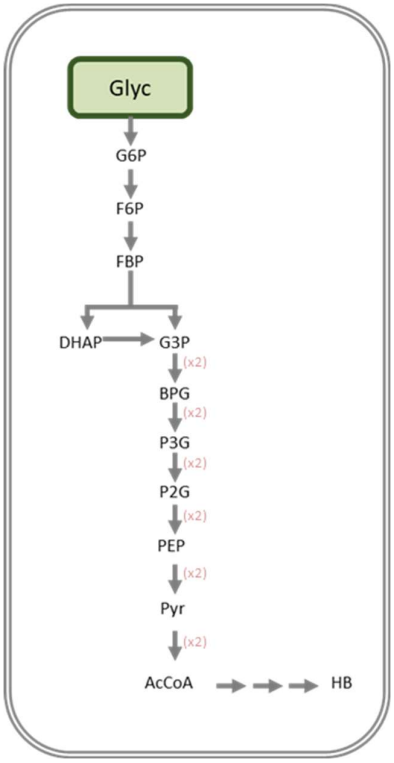

Figure 5. Schematic representation of EMP glycolysis from glycogen towards the production of hydroxybutyrate. Not that certain reactions have a (x2), which indicate two times the stoichiometry is considered in the S matrix.

Importantly, when generating the $\boldsymbol{S}$ matrix, all metabolites must be considered! That is, conserved moieties, H2O, amongst others. They all participate in the calculation of the free energy.

The formation Gibbs energy for each metabolite can be retrieved from databases. In this case, we have obtained them from an online tool called Equilibrator (https://equilibrator.weizmann.ac.il/). This tool estimates the formation Gibbs energy for each metabolite based on a method called component contribution. Performing the calculations described previously, we can obtain the $\Delta_r G'^o$ for each reaction in this network (Figure 6.A). Furthermore, we can plot the cumulative $\Delta_r G'$ of the pathway (Figure 6.B).

$\Delta_r G'^o$
of each reaction

```
glyc1 = -24.70 kJ/mol
emp1 = 3.20 kJ/mol
emp2 = -21.90 kJ/mol
emp3 = 21.50 kJ/mol
emp4 = 7.60 kJ/mol
emp5 = 11.60 kJ/mol
emp6 = -15.80 kJ/mol
emp7 = 11.80 kJ/mol
emp8 = -6.60 kJ/mol
emp9 = -58.40 kJ/mol
TCA1 = -59.80 kJ/mol
pha1 = 26.10 kJ/mol
pha2 = -6.50 kJ/mol
pha3 = -9.10 kJ/mol
```

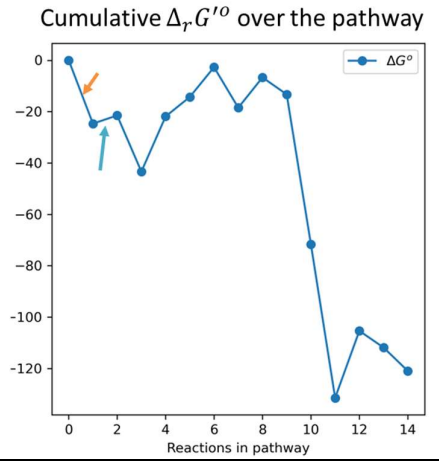

Figure 6. (A) $\Delta_r G'^o$ of each reaction involved in the pathway from Figure 5. (B) Cumulative $\Delta_r G'^o$ from the pathway, by adding the $\Delta_r G'^o$ of each reaction in the pathway. In orange and blue we have highlighted the $\Delta_r G'^o$ of the first two reactions (in A) and indicated their contribution to the cumulative $\Delta_r G'^o$ plot (in B).

From this initial analysis, we can learn quite some information about the pathway we are analysing. For example:

- The overall energy dissipation of the pathway ends in – 121 kJ/mol. This should be the same $\Delta_r G'^o$ of the global reaction (1 glucose + ….. <-> 1 HB + ….).
- The pathway's energy dissipation is negative. In global terms it should be feasible.
- Some reactions of the network present strong infeasibilities (slope of $\Delta_r G'^o$ positive). These are reactions 4 (emp3), 5 (emp4), 6 (emp5), 8 (emp7) and 11 (pha1).
- Some reactions of the network dissipate considerable amount of energy (slope of $\Delta_r G'^o$ negative). Especially, reactions *emp9* and *TCA1.*

From these observations, it seems like the pathway proposed should not occur under standard conditions. But wait a minute… glycolysis is the primarily method for many organisms (including us!) to degrade glucose and generate energy. How does it occur given these thermodynamic constrains?

The answer lies in the concepts. So far, we have calculated the **standard free energy** of the reactions. That is, we have assumed that all the metabolites are at concentrations of 1 M.

For this case, we will examine the reaction named *emp3*, which has a $\Delta_r G'^o$ = 21.5 kJ/mol. The reaction stoichiometry is: <<1 FBP ---> 1 G3P+ 1 DHAP>>. If we change the concentrations of these metabolites, the free energy of this reaction can become more or less feasible (Figure 7). However, since these metabolites are also involved in the previous or subsequent reactions, they too are affected by changes in their concentrations.

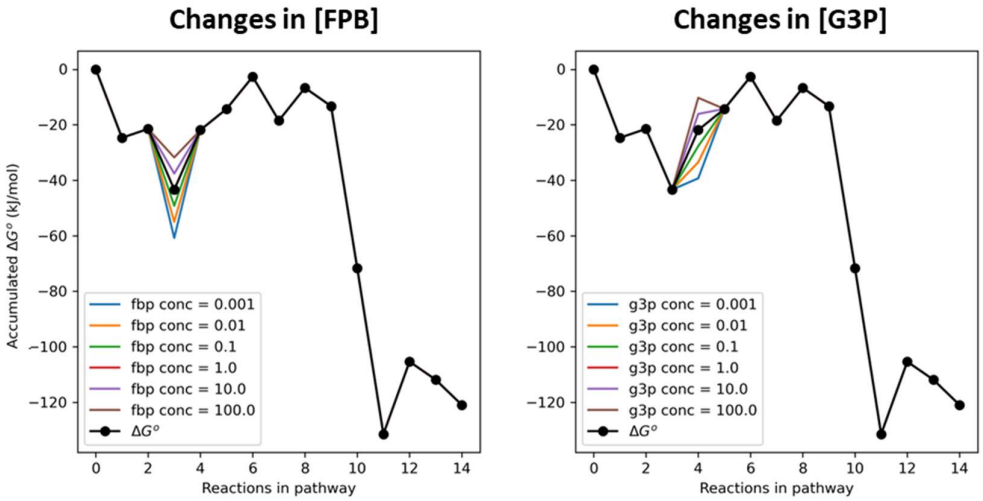

Figure 7. Cumulative free energy calculation from the pathway in Figure 5. In black, the $\Delta_r G'^{o}$ (standard) of the pathway (all metabolites at 1 M). Changes in the concentration of FBP (Left) or G3P (right) are shown in different colors.

This effect occurs throughout metabolism, with all metabolites at different concentrations. If we perform the same calculations using estimated concentrations of metabolites in *Escherichia coli* chemostats, we obtain the following distribution (Figure 8):

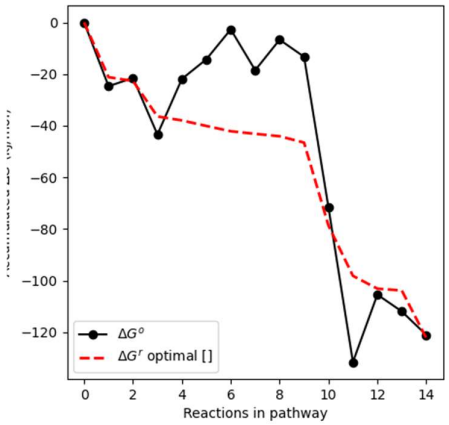

Figure 7. Cumulative free energy calculation from the pathway in Figure 5. In black, the $\Delta_r G'^{o}$ (standard) of the pathway (all metabolites at 1 M) and in red the $\Delta_r G'$ with specific metabolites' concentrations.

As you can see, glycolysis is feasible after all! But it is very dependent on the metabolite concentrations (context dependent!).

Work with the jupyter notebook file

You can work with this example in Python (Jupyter notebook format) on Brightspace (M4.2 Reaction thermodynamics).

5 Driving Force of a Pathway

Once the free energy of each reaction involved in a pathway has been calculated could you answer what is the pathway's driving force? Simply put, a pathway is limited by the reaction(s) with the highest $\Delta_r G'$; or as stated in literature (confusingly) the reaction with the lowest $-\Delta_r G'$. Since this reaction is limiting, the whole pathway will be dependent on this $\Delta_r G'$, and this value is named the Minimum Driving Force of the pathway.