

CHNOSZ

Reaction path calculations using the transfer function

1 Purpose

The transfer function, introduced in CHNOSZ-0.8, is meant to model the reaction paths of minerals and proteins. A reaction path portrays the progressive formation and destruction of minerals or proteins as the aqueous solution chemistry changes in response to incremental dissolution of the starting material. This function can be used to investigate the sequence and abundance of minerals formed in weathering processes and possibly of proteins formed in the cell cycle or evolution.

The PDF version of this document can be found [here](#)¹.

2 Principle

Reaction paths are often visualized on chemical activity diagrams. First, a stability diagram is constructed where the axes correspond to activities of aqueous species that can be altered by the dissolution of a reactant mineral or protein. The dissolution of the reactant might initially lead to the formation of other products because the starting composition of the fluid lies outside the stability field of the reactant mineral or protein.

The function takes a stepwise approach to the determination of a reaction path. It is based on the hypothesis of partial equilibrium of the products with the aqueous solution (Helgeson et al., 1969²) [1]. The most stable product mineral or protein with respect to the current solution chemistry is identified. A small amount of the primary reactant is provisionally introduced to the system. A conservation constraint is used to calculate the amount of the product that could be formed from this addition of reactant. (Al^{+3} and PBB (protein backbone) are the conserved components in the mineral and protein examples below.) If the change in solution chemistry as a consequence of the transformation of reactant to product is within a limit defined by the user, the step is allowed to proceed. Otherwise, the step fails and in the next step a smaller amount of primary reactant is provisionally introduced.

What happens when the solution chemistry changes so that a different product from that originally formed becomes stable? In an open system, the previously formed product is no longer available to react, so the formation of the new product simply proceeds according to whatever conservation constraint is in place. In a closed system, the previously formed product is now a reactant (referred to here as the secondary reactant). The primary and secondary reactants now both transform into the product, and the stoichiometry of this overall transformation is governed by a coupling constraint in addition to the primary conservation constraint.

3 Limitations

The current calculations do not take account of the speciation of the aqueous components. For example, if $NH_{3(aq)}$ is among the basis species, that particular stoichiometry will appear in the dissolution reactions even if the pH is below a value where NH_4^+ is actually more stable.

A major motivation for writing this code was to generate reaction path diagrams for proteins. Certain terminology borrowed from the geochemical world may not be entirely consistent with biochemical reality. An example is the mention of the dissolution of reactants. In cells, do proteins really dissolve to form

¹<http://www.chnosz.net/transfer/transfer.pdf>

²[http://dx.doi.org/10.1016/0016-7037\(69\)90127-6](http://dx.doi.org/10.1016/0016-7037(69)90127-6)

CO₂, H₂O, NH₃ and so on? I suspect that the mechanism of protein formation and destruction in cells is more likely to be bound to amino acids than these chemical components. However, the *process* of protein transformation (the coupled dynamic changes in cellular chemistry and the population of proteins) rather than its *mechanism* might be amenable to our geochemically derived model.

4 Definitions

Conservation Constraint The rule by which an amount of product is calculated from a provisional amount of primary reactant. In the examples below, this “primary conservant” is Al⁺³ or PBB (protein backbone).

Coupling Constraint In a closed system, the rule that determines the amount of secondary reactant that is introduced along with an amount of primary reactant. In the mineral examples below, this “secondary conservant” is H₄SiO₄ at the gibbsite-kaolinite boundary and K⁺ at the kaolinite-muscovite boundary.

Destruction Exponent (alpha) An exponent that determines the fraction of primary reactant that is destroyed at each step of the simulation. With each step of the simulation, the program may adjust this exponent so that the change in solution chemistry (quantified by logarithms of activities of aqueous solutes) does not surpass the maximum set by the user (i.e., devmax).

Primary Reactant The mineral(s) or protein(s) that are defined by the user to be initially available to react.

Secondary Reactant In a closed system simulation, the mineral(s) or protein(s) previously formed that are destroyed when the solution chemistry enters a different stability field.

5 Arguments to function

nsteps How many steps to run the simulation.

dmode Destruction mode (can be used to specify whether a system is open or closed):

none Don't react secondary reactants (open system).

coupled React secondary reactant by coupling to primary reactant (closed system).

devmax The maximum change in logarithm of activity of any basis species that is permitted in any step.

plot The numeric identification of basis species whose logarithms of activities should be plotted on the diagram. These are also the basis species that can be the secondary conservant.

ibalance The numeric identification of a single basis species whose amount is conserved in each step (the primary conservant).

fmode Formation mode, describes the composition of the product:

one Only form the single most stable species.

many Form various species in equilibrium amounts.

buffers List of basis species whose activities are affected by a buffer system in addition to reactions in the system.

alphamax Maximum value for the destruction exponent.

alphastart Starting value for the destruction exponent.

T Temperature.

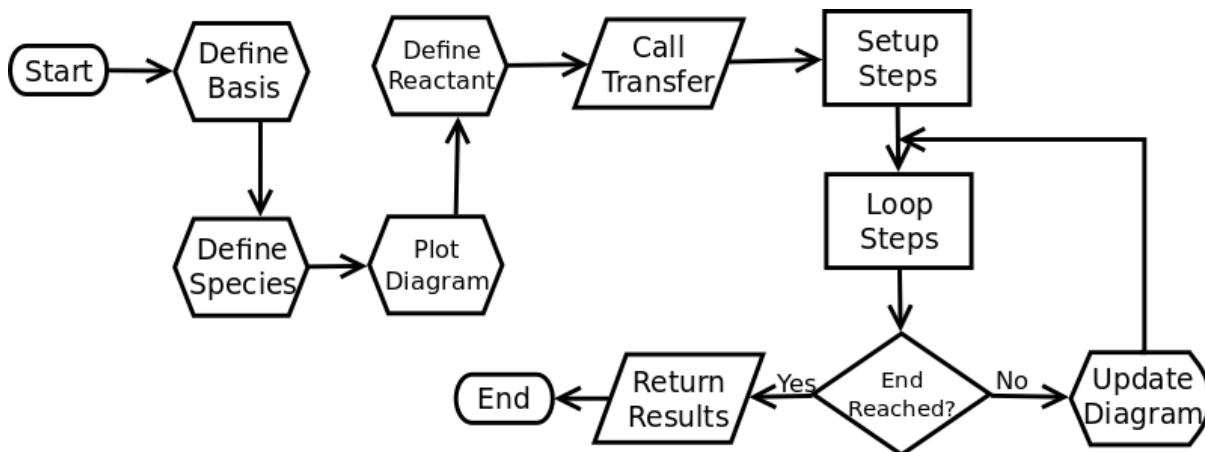
P Pressure.

6 Algorithms

6.1 Preparation

In preparation for the reaction path simulation, first define the basis species and the species of interest (minerals or proteins) using the `basis` and `species` functions of CHNOSZ. Use `affinity` and `diagram` to construct a predominance diagram; the axes correspond to basis species that are candidates for the secondary conservants. Use `species` again to define the reactant, by giving one or more reacting species a non-zero activity (e.g. logarithm of activity equal to zero) and the remaining species an activity close to zero (e.g. logarithm of activity equal to -999). Call `transfer` with desired arguments. The function performs some setup steps, then initiates the major loop. At each step the function prints diagnostic messages to the screen; at each successful step it adds a point to the diagram indicating the current chemical conditions. When the end is reached (by hitting the maximum number of steps defined by the user) the function returns the results, including the numbers of moles of each species formed at each step.

A flowchart representing the preparatory steps and the function call and return is shown below.



6.2 Setup steps

Here the function initializes counters and constants, then processes the argument list. If `ibalance` (primary conservant) is `PBB`, the function stops with an error if all species are not in fact proteins, otherwise adds `PBB` rows to the basis and species stoichiometric matrices. If `ibalance` is `NULL`, an outside function (`which.balance`) is called to find the first basis species that is present in all species; this becomes the primary conservant. Everything is ready for the major loop.

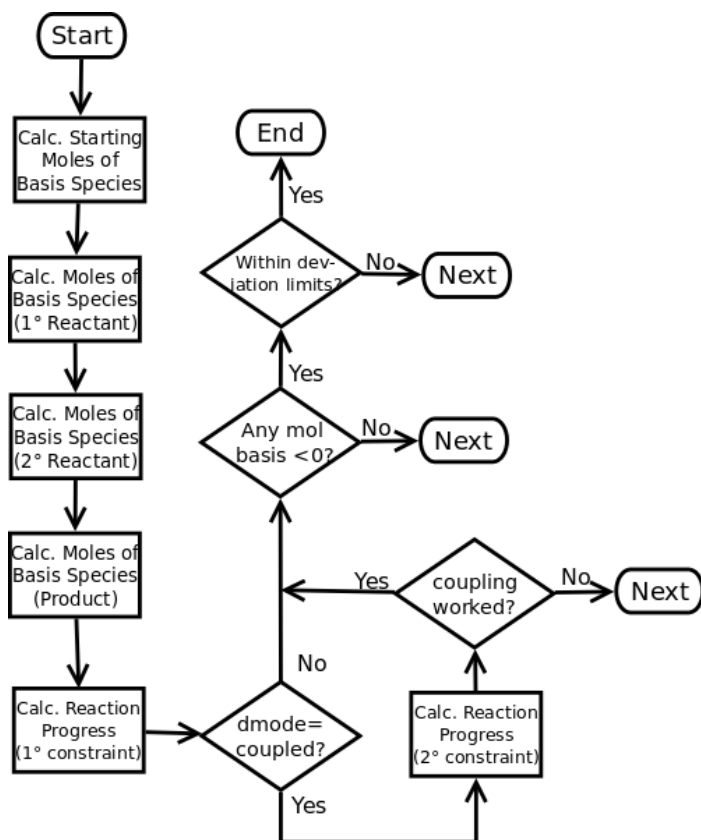
6.3 Loop steps: reaction stoichiometry and solution chemistry

At the beginning of each loop, the current value of alpha (destruction exponent) is calculated. If the previous step worked, the value of alpha is increased by one (we try to react more of the reactant in this step) but if the previous step failed, the value of alpha is decreased by one (we react less of the reactant) in an attempt to stay within the deviation limits of solution chemistry set by the user.

Then the function calculates the current numbers of moles of each basis species, and the changes that accompany dissolution of the primary reactant and dissolution of the secondary reactant (if any). The most stable product (or products if fmode is many) is also identified. Then, for an open system (dmode == none), the transformation of the reactant to product is calculated by observing the primary conservation constraint. For a closed system (dmode == coupled), the secondary conservation constraint is also invoked. For coupling to work, the secondary conservant must be present in the transformation reactions of the primary reactant to the product and of the secondary reactant to the product.

The calculation of reaction progress for either the open or closed system case takes account of the destruction exponent (alpha) which determines the amount of primary reactant that is available at that step. If the transformation of primary and secondary reactants to products causes the number of moles of any basis species to become negative, or the change in logarithm of activity of any basis species to exceed the deviation limit (devmax) the step fails and the next loop is initiated.

The figure below shows the major computations for each step. Failed steps lead to the “next” exits. If successful, the “end” of the step is reached and the reaction has progressed! The ensuing changes in the amounts of basis species and of reactant and product species are logged, a point is made on the activity diagram, and the following step is initiated.



7 Mineral Examples

The following examples examine the dissolution of K-feldspar in closed and open systems. The results are similar to the spreadsheet calculations for these systems described in Steinmann et al., 1994³ [2]. Each example below consists of a brief description of the strategy involved, the source code, and output in the form of a movie and a figure.

7.1 Feldspar dissolution – closed system

This example shows the dissolution of K-feldspar in a closed system (previously formed minerals are allowed to react). Here is the code to produce the diagrams, followed by a description of what each line does. The example may also be run in CHNOSZ with `feldspar('closed')`.

```
basis(c('Al+3', 'H4SiO4', 'K+', 'H2O', 'H+', 'O2'), c(0, -6, -6, 0, 0, 0))
species(c('k-feldspar', 'muscovite', 'pyrophyllite', 'kaolinite', 'gibbsite'))
a <- affinity(H4SiO4=c(-6, -2), 'K+'=c(-3, 8))
diagram(a)
basis('pH', 4)
species(1:5, c(-4, rep(-999, 4)))
t <- transfer(550, dmode='coupled', plot=c(2, 3), devmax=0.2)
plot.transfer(t)
```

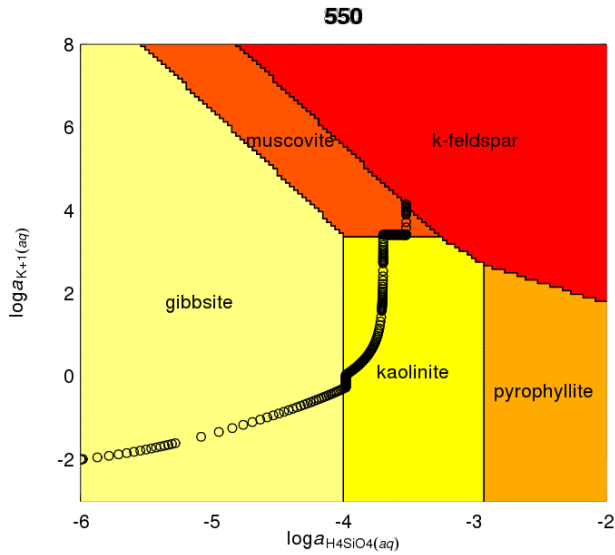
1. Identify the basis species and set their initial logarithms of activities. We use a unit activity for H₂O, and the values of -6 for H₄SiO₄ and K⁺ are consistent with the initial conditions used in Ref. [2]. So that the diagram we produce is equivalent to one with $\log(a_{K^+}/a_{H^+})$ on the y-axis, the pH here is initially set to 0. The values for Al⁺³ and O_{2(g)} don't affect the results, because the first is the primary conservant and the second does not enter any possible transformation reactions (they are not redox reactions).
2. Identify the species of interest (minerals) in the system. The default in CHNOSZ is to assign all mineral species unit activities, so that the diagram that follows is an equal-activity diagram.
3. Calculate the affinities of the formation reactions of the minerals on a two-dimensional grid.
4. Plot a predominance (equal-activity) diagram. The diagram function balances the reactions on the first basis species that is present in all the species of interest; in this case it is Al⁺³. As noted above, the y-axis which here is labeled as $\log a_{K^+}$ is numerically equivalent to $\log(a_{K^+}/a_{H^+})$ because we have set $\log a_{H^+}$ to zero.
5. Set the initial pH to 4 for the simulation.
6. Set the logarithms of activities of the minerals: we will react K-feldspar but start with essentially nothing (10^{-999} moles) of anything else.
7. Perform the reaction path simulation. 550 is the maximum number of steps for the simulation, the destruction mode is coupled (for a closed system), the basis species whose logarithms of activities we wish to plot on the equal-activity diagram are the 2nd and 3rd (H₄SiO₄ and K⁺), and the maximum

³<http://ccm.geoscienceworld.org/cgi/content/abstract/42/2/197>

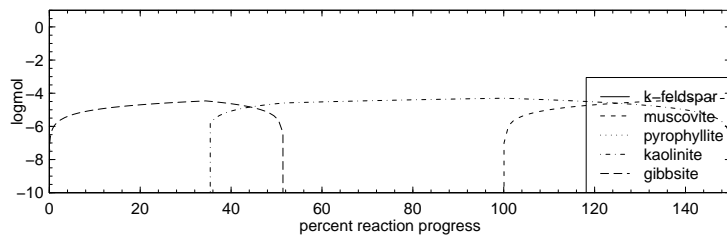
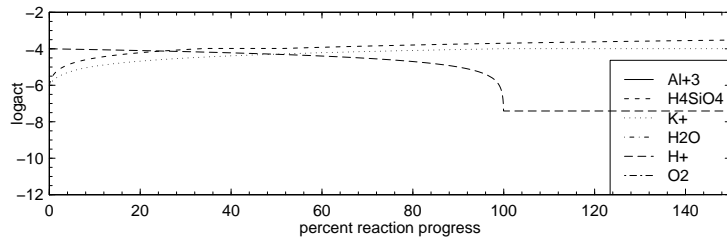
deviation of logarithm of activity of any basis species that is permitted at any step is 0.2. (The default value is 0.1, but this leads to a significantly longer simulation time.)

- The results of the transfer calculation are plotted to show the logarithms of activities of basis species and the amount of minerals produced with reaction progress.

The movie (Shockwave Flash) is here⁴. The final appearance of the reaction path on the equal-activity diagram is shown below:



Below is the plot showing the change of solution chemistry and minerals produced with reaction progress. At 100% reaction progress, 10^{-4} moles of K-feldspar have reacted (this is right around step 403, as the system enters the muscovite stability field).



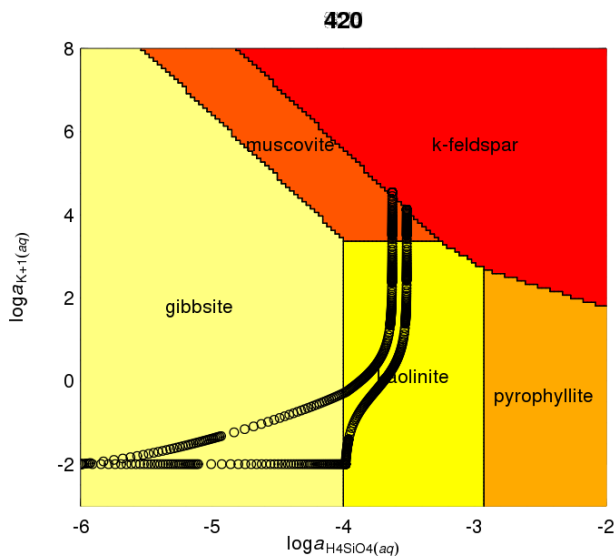
⁴<http://www.chnosz.net/transfer/feldspar-closed.htm>

7.2 Feldspar dissolution – open system

The sequence of commands below starts after the first six shown above. The three calls to transfer represent different starting constraints. For the first we react potassium feldspar, as above but instead in an open system; the result is the top curve in the diagram. Then we ask what happens when kaolinite is reacted instead of K-feldspar; this takes us horizontally from the starting point to the gibbsite-kaolinite boundary. Finally, we use that value of $\log a_{\text{H}_4\text{SiO}_4(\text{aq})}$ as a starting point for again reacting K-feldspar, which gives us the lower curved line that starts in the kaolinite field and stops at the muscovite - K-feldspar boundary. These segmented reaction paths are discussed in more detail by in Ref. [2]. This simulation can also be run in CHNOSZ using `feldspar('open')`.

```
t <- transfer(450,dmode='none',plot=c(2,3))
species(c('k-feldspar','kaolinite'),c(-999,-4))
t <- transfer(150,dmode='none',plot=c(2,3))
species(c('k-feldspar','kaolinite'),c(-4,-999))
basis('H4SiO4',t$basis[rownames(t$basis)=='H4SiO4',ncol(t$basis)])
t <- transfer(420,dmode='none',plot=c(2,3))
```

The movie is here⁵. The reaction path is also shown in the figure below.



8 Protein Examples

The examples below show computational results for reaction paths among some proteins that make up the anaphase promoting complex in yeast. We ask what proteins might form if we add to the system the composition of one of the proteins (APC2) in the complex.

First we consider a path on a $\log a_{\text{H}_2(\text{aq})}$ - $\log a_{\text{CO}_2(\text{aq})}$ diagram. Hydrogen activity is a way of quantifying oxidation-reduction (or redox) state and CO_2 activity may be thought of as measuring the carbonation potential. Note that we use hydrogen activity instead of oxygen fugacity as a measure of oxidation-reduction

⁵<http://www.chnosz.net/transfer/feldspar-open.htm>

state. Although the latter has been used in a previous publication⁶ [3] it is not suitable as a process variable in reaction path simulations. This is because the stepwise mass transfer calculations simulate finite changes of the concentrations of aqueous solutes in a reservoir of fluid, and an oxygen activity of 10^{-80} to 10^{-70} (which is very approximately the range that corresponds to the range of oxidation potentials in terms of $\log a_{\text{H}_2(\text{aq})}$ used here) would require *very* small steps and make the simulation intractable.

8.1 Anaphase-promoting complex, closed system

Notice how in the feldspar diagram above the reaction path only had to cross boundaries that were horizontal and vertical, so the secondary conservant in these cases was just either one of the basis species that are represented by the axes. In the protein diagram the reaction boundaries have slopes between zero and infinity so the secondary conservation constraint is a linear combination of the basis species shown on the axes.

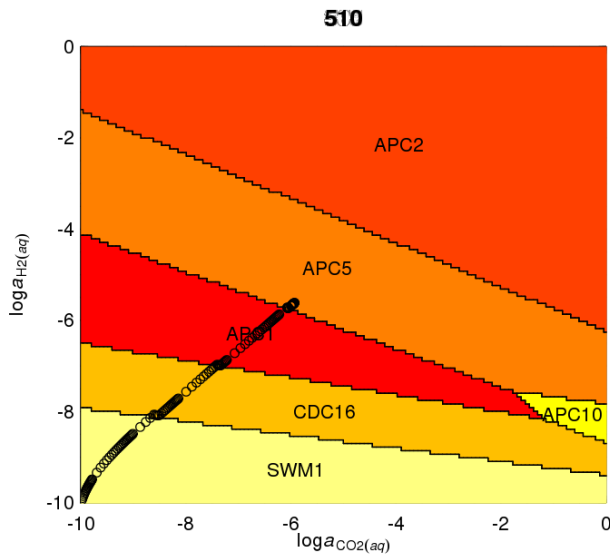
This reaction path calculation can also be performed in CHNOSZ using `apc('closed')`.

```
basis(c("CO2", "H2O", "NH3", "H2", "H2S"), c(-10, 0, -4, -10, -7))
basis("H2", "aq")
species(c("APC1", "APC2", "APC5", "CDC16", "CDC23", "APC10", "SWM1"), "YEAST")
a <- affinity(CO2=c(-10, 0), H2=c(-10, 0))
diagram(a, residue=TRUE, as.residue=TRUE)
species(1:nrow(species()), -999)
species("APC2_YEAST", 0)
t <- transfer(510, ibalance="PBB", plot=c(1, 4), dmode="coupled", devmax=0.15)
plot.transfer(t, ylim=c(-22, -6), logprogress=TRUE)
```

1. Define the basis species including the starting values for $\log a_{\text{H}_2(\text{g})}$ and $\log a_{\text{CO}_2(\text{aq})}$. For now we will be using uncharged proteins (the stepwise calculations used by `transfer` are not yet speed-optimized for ionized proteins) so the proton is not in the basis definition.
2. Because CHNOSZ likes to make H_2 and O_2 gases in the basis definition, set the state of H_2 to aqueous.
3. Load the proteins.
4. Calculate the affinities of the formation reactions of the proteins on a 2-dimensional grid.
5. In making the predominance diagram we consider uncharged proteins and the relative stability fields of the residues of the proteins.
6. We are only going to react a single protein composition so make the activities of the proteins very low.
7. Make the activity of the protein we are reacting much higher.
8. Perform the transfer calculation. The primary conservant is the protein backbone group so in using `transfer` we are effectively looking at reactions among protein residues (that's why we did `as.residue=TRUE` in the arguments to `diagram` above). We set up a closed-system reaction path (`dmode=coupled`).
9. Plot the calculated logarithms of activities of basis species and of proteins, setting an option to plot against the logarithm of the reaction progress variable.

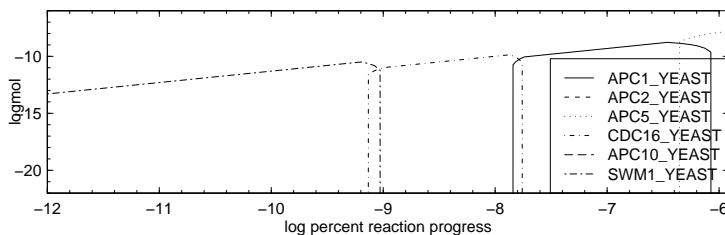
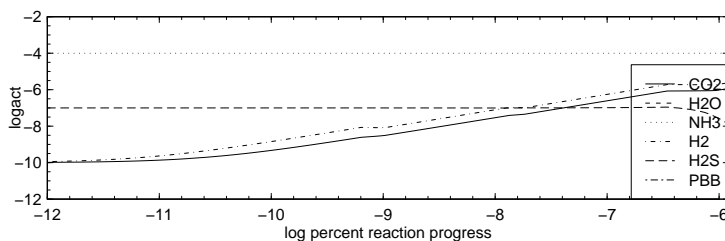
⁶<http://www.geochemicaltransactions.com/content/9/1/10>

The movie is here⁷. The reaction path diagram is shown below.



The solution chemistry never gets to the APC2 metastability field even though we are adding its composition to the system. In the figure below it can be discerned that this is a consequence of a significant drop in the availability of sulfur (as H_2S). Because the transformation of APC2 (moles sulfur per residue = 0.035) to APC5 (moles sulfur per residue = 0.044) is a sulfur-consuming reaction, it can not proceed significantly once the concentration of the sulfur-containing basis species, in this case hydrogen sulfide, drops too low. The situation would be different if we started with a greater amount of H_2S or used the “buffer” argument to transfer to bleed sulfur into the system during the computer experiment (try out `apc('buffer')` in CHNOSZ).

The values of percent reaction progress are shown in logarithmic units here. The entire percent reaction progress scale would increase by a factor of 10 for each unit decrease in the logarithm of activity of the starting protein (note that we started with log activity of zero for the reactant APC2). Here you can see the places where $\log a_{\text{H}_2(\text{aq})}$ momentarily drops as a reaction boundary is crossed.



⁷<http://www.chnosz.net/transfer/apc-closed-co2.htm>

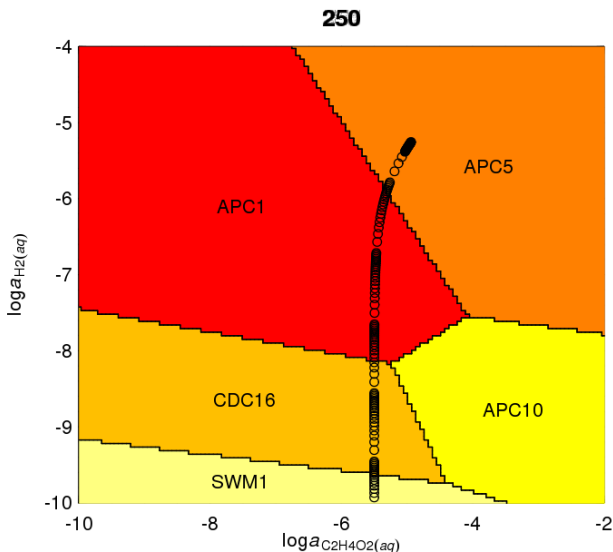
8.2 Anaphase-promoting complex, open system, different basis species

The reaction boundaries above all have negative slopes because we used CO_2 to stand for the carbon-containing basis species. It is so highly oxidized (nominal oxidation state of carbon equal to +4) that any increase in $\log a_{\text{CO}_2(\text{aq})}$ in a system that maintains its metastable population of biomacromolecules must be offset by a decrease in $\log a_{\text{H}_2(\text{aq})}$. (To understand why, consider for example the idealized reaction $2\text{H}_2 + \text{CO}_2 = \text{C} + 2\text{H}_2\text{O}$ where C stands for carbon in organic molecules like proteins). Let us try substituting a neutral (in terms of oxidation state) species for CO_2 . One candidate is acetic acid: one can see by its formula $\text{C}_2\text{H}_4\text{O}_2$ that the nominal oxidation state of carbon in this species is zero.

The relevant commands are printed below. We also throw another twist in the equation by changing the formation mode with the `fmode=many` argument. This tells the function to form many proteins in amounts, consistent with equilibrium among the products, that are calculated using the diagram function in CHNOSZ (Dick, 2008⁸ [3]). This simulation can also be performed in CHNOSZ using `apc('many', basis='acetic')`.

```
basis(c("acetic acid", "H2O", "NH3", "H2", "H2S"), c(-5.5, 0, -4, -10, -7))
basis("H2", "aq")
species(c("APC1", "APC2", "APC5", "CDC16", "APC10", "SWM1"), "YEAST")
a <- affinity(C2H4O2=c(-10, -2), H2=c(-10, -4))
diagram(a, residue=TRUE, as.residue=TRUE)
species(1:nrow(species()), -999)
species("APC2_YEAST", 0)
t <- transfer(250, ibalance="PBB", plot=c(1, 4), dmode="none",
  devmax=0.15, fmode="many")
```

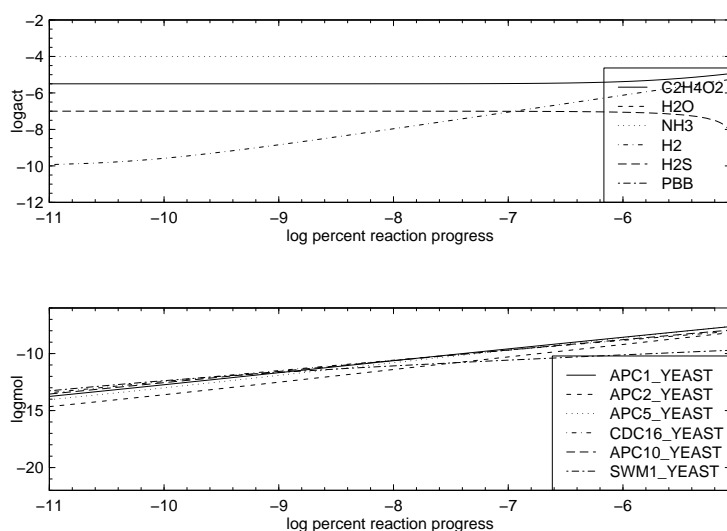
The movie is here⁹. The reaction path diagram is shown below.



The logarithms of activities of basis species and the amounts of proteins formed are shown in the figure below.

⁸<http://www.geochemicaltransactions.com/content/9/1/10>

⁹<http://www.chnosz.net/transfer/apc-open-c2h4o2.htm>



Do the relative abundances of proteins predicted using a metastable equilibrium reaction path model have any relationship to the relative abundances of proteins in cells? The table below lists the abundances of the proteins (in relative units) found by Ghaemmaghami et al., 2003¹⁰ [4] (the YeastGFP protein expression study). The two most abundant proteins in the table show up as the two most abundant proteins in the figure above at log percent reaction progress between ca. -9 and -8.5. So, the data give some sign of an energy minimizing process, but equilibrium is far from being attained: Our reactant protein always shows up with a lower relative abundance in the model that what is observed in the cell.

Protein	ORF	Abundance
APC1	YNL172W	178
APC2	YLR127C	432
APC5	YOR249C	259
CDC16	YKL022C	2750
APC10	YHR166C	1360
SWM1	YGL240W	NA

9 Discussion

The mineral reaction paths shown above are obviously meaningful to geochemistry students and researchers; such things have been published and commented on many times in the literature. But do the protein reaction paths mean anything? Perhaps the anaphase-promoting complex is not the greatest example since it is hard to imagine why the cell would form each of the proteins in the complex under differing conditions when they are all needed together for the complex to assemble. But maybe the proteins are formed at different times and kept around for the final assembly of the complex; that situation might permit more control by the cell and allow it to respond to unexpected environmental changes.

Another reservation I have about this model is that I have chosen to titrate in a particular protein, because (by analogy with K-feldspar) it shows up on the high-activity corner of a predominance diagram. Perhaps there is a particular stoichiometry, that of a proto-protein, that never actually forms but is the source of many different proteins in the cell by virtue of modulation of redox state, carbon, nitrogen and sulfur availability, and pH, hydration state and ionic strength, in the vicinity of protein synthesis pathways.

A good place to go with this might be to consider proteins that form in a well-understood sequence in

¹⁰<http://dx.doi.org/10.1038/nature02026>

the cell, including cell-division control proteins other than the anaphase-promoting complex. The relationship between subcellular chemistry and the sequential formation of organelles (Dick, 2009¹¹) [5] is another potential target.

10 Other Resources

There are many software tools that can be used to simulate reaction paths. Just a few are listed below. There is no reason (in theory) that any of these could not also be used to study protein reaction paths.

- PHREEQC - An example of the feldspar dissolution calculation is provided with the documentation¹².
- Geochemist's Workbench - This software provides a module known as react¹³.
- GEMS-PSI - Gibbs energy minimization code that has a process simulator for mass transfer calculations¹⁴.

References

- [1] H. C. Helgeson, R. M. Garrels, and F. T. Mackenzie. Evaluation of irreversible reactions in geochemical processes involving minerals and aqueous solutions. II. Applications. *Geochim. Cosmochim. Acta*, 33(4):455 – 481, 1969.
- [2] P. Steinmann, P. C. Lichtner, and W. Shotyk. Reaction path approach to mineral weathering reactions. *Clay Clay Min.*, 42(2):197 – 206, 1994.
- [3] J. M. Dick. Calculation of the relative metastabilities of proteins using the CHNOSZ software package. *Geochem. Trans.*, 9:10, 2008.
- [4] S. Ghaemmaghami, W. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737 – 741, 2003.
- [5] J. M. Dick. Calculation of the relative metastabilities of proteins in subcellular compartments of *Saccharomyces cerevisiae*. *BMC Syst. Biol.*, 2009. submitted for publication.

Updated 2009-04-22 by Jeffrey M. Dick

¹¹<http://www.arxiv.org/abs/0812.0191>

¹²http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/html/final-75.html

¹³http://crustal.usgs.gov/projects/aqueous_geochemistry/geochemists_workbench.html

¹⁴<http://gems.web.psi.ch/screenshots/ssgal14.html>