

# Hot-spring chemistry and model proteins from a metagenome

Jeffrey M. Dick and Everett L. Shock

January 12, 2014

## 1 Introduction

This vignette for the CHNOSZ package demonstrates calculations of the relative stabilities of proteins with amino acid compositions derived from metagenomic data in a hot spring. The calculations are described in more detail in a recent paper [1]. The purpose of this document is to present a minimal amount of code to show how most of the figures in the paper can be generated. The code actually used to make the figures in the paper is available online as supporting information for the paper.

“Bison Pool” is a hot spring in Sentinel Meadows, in the Lower Geyser Basin of Yellowstone National Park. Environmental DNA sequences from biofilm samples at five sites along the outflow are available from the Joint Genome Institute’s IMG/M system [2]. IMG/M also provides the sequences of proteins inferred from the metagenome, with predicted functional annotations. The sample IDs used in the field, and also in IMG/M, have the letters N, S, R, Q, P, which refer to sites 1, 2, 3, 4, 5, in that order; site 1 is the boiling source of the hot spring and site 5 is ~22 meters downstream where the water on some days was in the range of ~55-60 °C [3].

The amino acid compositions of the model proteins are derived from the metagenome. The “overall” model proteins have amino acid compositions that are the average of all protein sequences at any sampling site. There are also 20 different “classified” model proteins for subsets of the metagenome that share keywords in the annotations.

## 2 General Setup

Load CHNOSZ and the thermodynamic database.

```
> library(CHNOSZ)
> data(thermo)
```

The thermodynamic data for the methionine sidechain group used in the paper has been superseded in CHNOSZ\_0.9-9 [4], but the older dataset is still available using `add.obigt()`:

```
> add.obigt()
```

```
add.obigt: file has 303 rows; made 84 replacements, 219 additions, units = cal
add.obigt: file was /tmp/RtmppTadnh/Rinst722e511c7db7/CHNOSZ/extdata/thermo/OBIGT-2.csv
add.obigt: use data(thermo) to restore default database
```

After loading the package, the amino acid compositions of the model proteins are present in `thermo$protein`. Assemble the classification terms using one of the site names.

```
> tp <- thermo$protein
> classes <- tp$protein[tp$organism=="bisonN"]
> classes
```

```

[1] "overall"      "transferase"  "transport"    "dehydrogenase"
[5] "synthase"     "ATPase"      "kinase"       "synthetase"
[9] "membrane"     "permease"    "hydrolase"    "oxidoreductase"
[13] "transcription" "peptidase"   "ribosomal"    "protease"
[17] "reductase"    "transposase" "periplasmic"  "signal"
[21] "phosphatase"

```

Associate the sites with their letter codes.

```

> sites <- c("N", "S", "R", "Q", "P")
> sitenames <- paste("bison", sites, sep="")

```

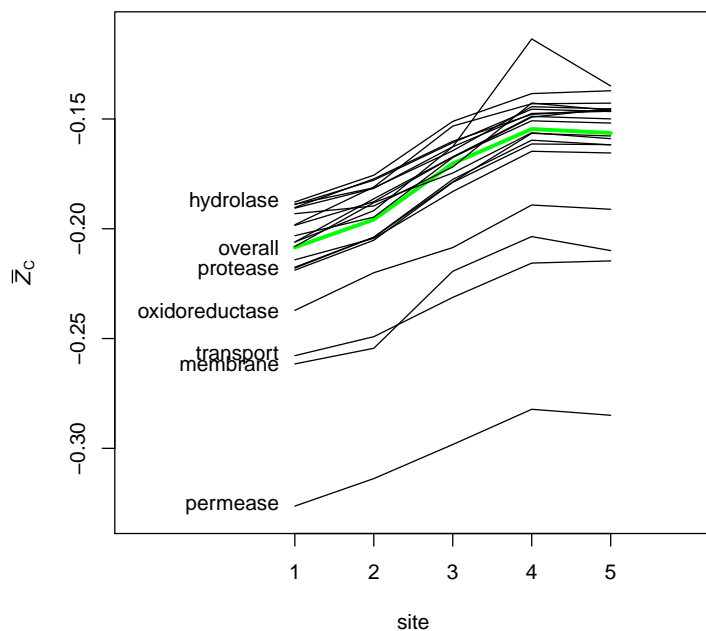
### 3 Average Oxidation State of Carbon in Model Proteins

Plot the average oxidation state of carbon ( $\bar{Z}_C$ ) of each of the classes of model proteins. Initialize the plot, labeling the  $x$ -axis at sites 1 to 5. Then set up colors for the lines (the “overall” model proteins in green) and identify the classes of proteins to label on the plot. For model proteins of all classes and sites, get the rownumbers of the proteins (ip), then extract the rows from the thermo\$protein data frame (p), then calculate the chemical formulas of the proteins (pf), and finally calculate values of  $\bar{Z}_C$ . Loop over the classes, plotting the lines and labels. After Fig. 4 in Ref. [1].

```

> plot(0, 0, xlim=c(-1, 6), ylim=c(-0.33, -0.11), xlab="site",
      ylab=expression(bar(italic(Z))[C]), xaxt="n")
> axis(1, at=1:5)
> col <- c("green", rep("black", 20))
> lwd <- c(3, rep(1, 20))
> clab <- c("hydrolase", "overall", "protease",
           "oxidoreductase", "transport", "membrane", "permease")
> ip <- iprotein(rep(classes, each=5), rep(sitenames, 20))
> p <- thermo$protein[ip, ]
> pf <- protein.formula(p)
> ZC.p <- ZC(pf)
> for(i in 1:length(classes)) {
  lines(1:5, ZC.p[(1:5)+5*(i-1)], col=col[i], lwd=lwd[i])
  if(classes[i] %in% clab) text(0.8, ZC.p[1+5*(i-1)], classes[i], adj=1)
}

```



## 4 Chemical Setting

### 4.1 Basis Species

Use CHNOSZ's basis function to define the basis species. The basis species consist of  $\text{HCO}_3^-$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_3$ ,  $\text{HS}^-$ ,  $\text{H}_2$  and  $\text{H}^+$  (all aqueous species except for  $\text{H}_2\text{O}$  liquid). Note: the functions `basis` and `species` are important for their side effect of setting up the thermodynamic system used by other functions in CHNOSZ.

```
> basis(c("HCO3-", "H2O", "NH3", "HS-", "H2", "H+"))
```

	C	H	N	O	S	Z	ispecies	logact	state
HCO3-	1	1	0	3	0	-1	13	0	aq
H2O	0	2	0	1	0	0	1	0	liq
NH3	0	3	1	0	0	0	68	0	aq
HS-	0	1	0	0	1	-1	22	0	aq
H2	0	2	0	0	0	0	66	0	aq
H+	0	1	0	0	0	1	3	0	aq

### 4.2 Hot-Spring Chemistry

To assess the relative stabilities of the proteins, parameters related to the chemistry of their environment are needed. The temperature (in  $^{\circ}\text{C}$ ) and pH measured in the field at the time the biofilm samples were collected for sequence analysis are defined below as a function of distance (in meters) from the source of the hot spring.

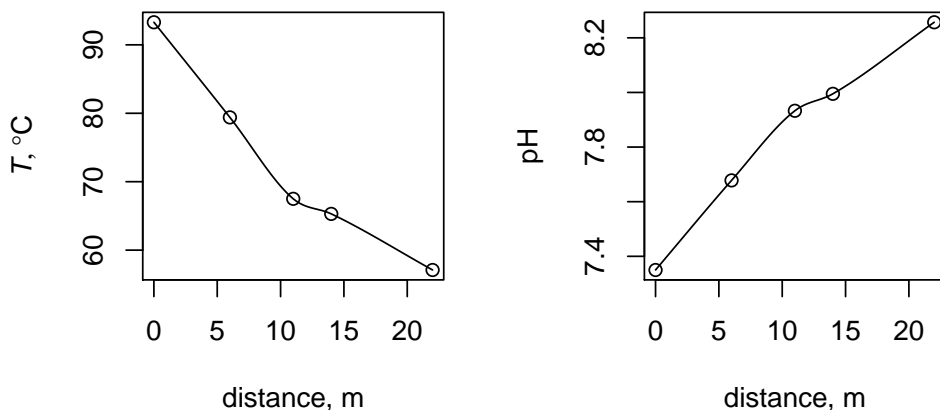
```
> distance <- c(0, 6, 11, 14, 22)
> T.bison <- c(93.3, 79.4, 67.5, 65.3, 57.1)
> pH.bison <- c(7.350, 7.678, 7.933, 7.995, 8.257)
```

The plot shows the temperature and pH as a function of distance. Use `splinefunctions` to draw smooth lines between the points (a line formed by segments ending at the intervals in `xpoints`). Use the `mfrow` option to put two plots next to each other, and CHNOSZ's `axis.label` function to make a formatted label for temperature. After Fig. 5a,c in Ref. [1].

```

> Tfun <- splinefun(distance, T.bison, method="mono")
> pHfun <- splinefun(distance, pH.bison, method="mono")
> xpoints <- seq(0, 22, 0.25)
> par(mfrow=c(1, 2), mar=c(4, 4, 3, 2))
> plot(distance, T.bison, xlab="distance, m", ylab=axis.label("T"))
> lines(xpoints, Tfun(xpoints))
> plot(distance, pH.bison, xlab="distance, m", ylab="pH")
> lines(xpoints, pHfun(xpoints))

```



Besides temperature and pH, other chemical properties of the hot spring water change as it cools. Oxidation-reduction (redox) state of the water is left for exploration (see below), but the activities of the other basis species are set to values that approximate analytical observations. Set pH to the value for site 3; the pH gradient will be considered later.

```

> basis(c("HCO3-", "NH3", "HS-", "H+"), c(-3, -4, -7, -7.933))

```

	C	H	N	O	S	Z	ispecies	logact	state
HCO3-	1	1	0	3	0	-1	13	-3.000	aq
H2O	0	2	0	1	0	0	1	0.000	liq
NH3	0	3	1	0	0	0	68	-4.000	aq
HS-	0	1	0	0	1	-1	22	-7.000	aq
H2	0	2	0	0	0	0	66	0.000	aq
H+	0	1	0	0	0	1	3	-7.933	aq

## 5 Relative Stabilities of Proteins

### 5.1 Formation Reactions of the Proteins

Use species to display the coefficients in formation reactions of the overall model proteins and set the species definition for CHNOSZ.

```

> species("overall", sitenames)

```

	HCO3-	H2O	NH3	HS-	H2	H+	ispecies	logact	state
1	1020.547	-2772.768	270.298	5.694	2147.462	1026.241	3590	-3	aq
2	935.793	-2541.750	251.890	5.381	1963.205	941.174	3591	-3	aq
3	981.834	-2666.209	271.402	5.362	2047.264	987.196	3592	-3	aq
4	952.439	-2583.005	266.317	5.834	1978.491	958.273	3593	-3	aq
5	941.702	-2554.448	263.077	5.750	1957.038	947.452	3594	-3	aq

name

1 overall\_bisonN

```

2 overall_bisonS
3 overall_bisonR
4 overall_bisonQ
5 overall_bisonP

```

After sequencing and assembly, many gene fragments remain, so the number of amino acid residues in the model proteins is the average length of mostly partial protein sequences and does not necessarily reflect the genetic sequence lengths. The first six columns of the output indicate the stoichiometries of the formation reactions. Although the reactions are written for non-ionized proteins,  $H^+$  has a non-zero coefficient because of the presence of other charged basis species.

The reactions can be divided by the lengths (number of amino acid residues) of the proteins to write per-residue formation reactions. First get the rownumbers of the overall model proteins in `thermo$protein(ip)`, then calculate the protein “length” (number of amino acid residues, not integers for the model proteins), then divide the protein formation reactions by the respective protein lengths.

```

> ip <- iprotein("overall", sitenames)
> pl <- protein.length(ip)
> mys <- species()
> mys[, 1:6]/pl

```

	HC03-	H2O	NH3	HS-	H2	H+
1	5.125029	-13.92441	1.357395	0.02859439	10.78422	5.153623
2	5.076258	-13.78786	1.366391	0.02918952	10.64951	5.105448
3	5.022554	-13.63894	1.388352	0.02742921	10.47274	5.049983
4	4.965715	-13.46697	1.388492	0.03041663	10.31522	4.996131
5	4.971975	-13.48691	1.388987	0.03035871	10.33272	5.002334

The above can be used to deduce for example that the mass-action effect of increasing the activity of  $H_2$  is to drive formation of the model protein for site 1 more than the others.

## 5.2 Predominances as a Function of Temperature and Hydrogen Activity

First set the temperature limits (`Tlim`) over which to perform the calculations. Then define a function `H2prot`, which gives the logarithm of activity of hydrogen ( $\log a_{H_2(aq)}$ ) as a linear function of temperature.

```

> Tlim <- c(50, 100)
> H2prot <- function(T) -11 + T * 3/40

```

The metastable equilibrium activity diagram shows predominance fields of the overall model proteins as a function of temperature and logarithm of activity of aqueous hydrogen. After Fig. 5b in Ref. [1].

```

> a <- affinity(T=Tlim, H2=c(-7, -4))

```

```

energy.args: pressure is Psat
energy.args: variable 1 is T at 128 values from 323.15 to 373.15 K
energy.args: variable 2 is log_a(H2) at 128 values from -7 to -4
subcrt: 11 species at 128 values of T and P (wet)
subcrt: 18 species at 128 values of T and P (wet)

```

```

> diagram(a, fill=NULL, names=1:5, normalize=TRUE)

```

```

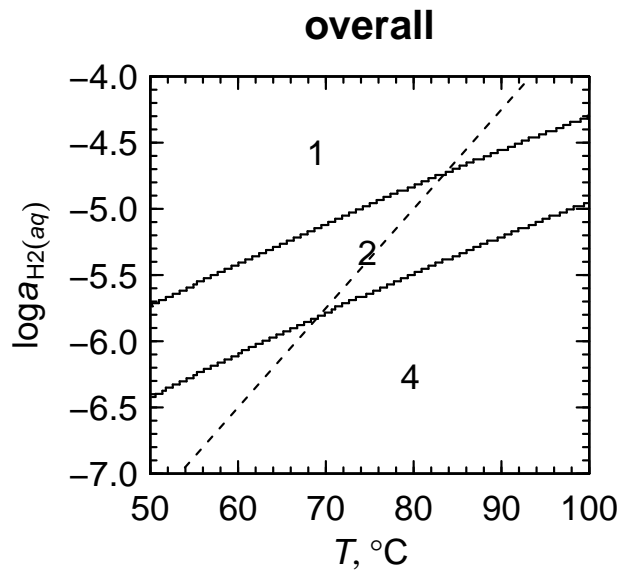
balance: coefficients are protein length
diagram: normalizing formulas in calculation of predominant species
diagram: plotting A/2.303RT from affinity(), divided by balancing coefficients

```

```

> lines(Tlim, H2prot(Tlim), lty=2)
> title(main="overall")

```



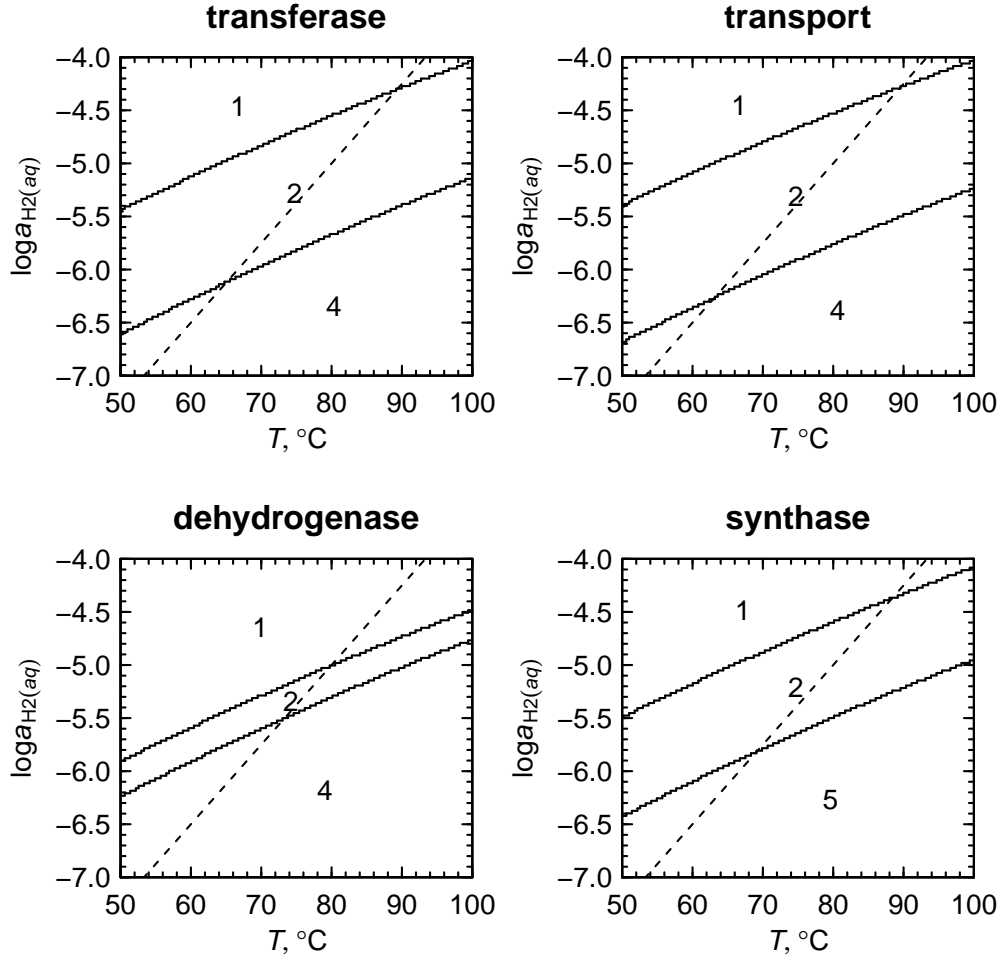
The stability fields for the proteins from the higher temperatures are at higher activities of hydrogen in the diagram. The dashed line enters the stability fields the model proteins at approximately the actual environmental temperatures.

Write a function to load one or more of the classes of model proteins.

```
> loadclass <- function(class) {
  species(delete=TRUE)
  species(rep(class, each=5), rep(sitenames, length(class)))
}
```

Then use this function to make equilibrium activity diagrams for a few different classes, afterwards restoring the species definition to the overall model proteins. To save space, the messages printed by CHNOSZ during the calculations are hidden. After Fig. 6 in Ref. [1].

```
> forclasses <- c("transferase", "transport", "dehydrogenase", "synthase")
> par(mfrow=c(2, 2))
> for(i in 1:4) {
  loadclass(forclasses[i])
  a <- affinity(T=Tlim, H2=c(-7, -4))
  diagram(a, fill=NULL, names=1:5, normalize=TRUE)
  title(main=forclasses[i])
  lines(Tlim, H2prot(Tlim), lty=2)
}
> loadclass("overall")
```



### 5.3 Chemical Affinities

Calculate the residue-normalized chemical affinities of the formation reactions of the overall model proteins. First set activities of the proteins equal to unity (logarithm of activity equal to zero). Then calculate affinities per mole of protein for the temperature, pH and  $\log a_{H_2(aq)}$  of each site. Use the lengths of the model proteins, defined above in `p1`, to calculate the affinities per residue. After Table 5 in Ref. [1].

```
> species(1:5,0)

      HC03-      H2O      NH3      HS-      H2      H+ ispecies logact state
1 1020.547 -2772.768 270.298 5.694 2147.462 1026.241    3590      0    aq
2  935.793 -2541.750 251.890 5.381 1963.205  941.174    3591      0    aq
3  981.834 -2666.209 271.402 5.362 2047.264  987.196    3592      0    aq
4  952.439 -2583.005 266.317 5.834 1978.491  958.273    3593      0    aq
5  941.702 -2554.448 263.077 5.750 1957.038  947.452    3594      0    aq

      name
1 overall_bisonN
2 overall_bisonS
3 overall_bisonR
4 overall_bisonQ
5 overall_bisonP

> a <- affinity(T=T.bison, pH=pH.bison, H2=H2prot(T.bison))
```

```

energy.args: pressure is Psat
energy.args: variable 1 is T at 5 values from 330.25 to 366.45 K
energy.args: variable 2 is pH at 5 values from 7.35 to 8.257
energy.args: variable 3 is log_a(H2) at 5 values from -6.7175 to -4.0025
subcrt: 11 species at 5 values of T and P (wet)
subcrt: 18 species at 5 values of T and P (wet)

> a.res <- t(as.data.frame(a$values))/p1
> a.res
      [,1]      [,2]      [,3]      [,4]      [,5]
X3590 -18.72091 -27.89448 -35.38639 -36.82324 -42.26549
X3591 -18.84592 -27.91351 -35.31897 -36.73942 -42.11999
X3592 -19.12097 -28.05392 -35.34968 -36.74937 -42.05199
X3593 -19.27012 -28.08074 -35.27670 -36.65748 -41.88895
X3594 -19.25362 -28.07774 -35.28470 -36.66756 -41.90683

```

The affinities are expressed as dimensionless values, i.e.  $A/2.303RT$  where  $A$ ,  $R$  and  $T$  stand for chemical affinity, gas constant, and temperature in Kelvin. The affinities are all negative, but highest at the conditions of sites 1 and 2 (columns 1 and 2) for the model proteins from site 1 (row 1), and highest at the conditions of sites 3 to 5 (columns 3 to 5) for the model protein from site 4 (row 4).

```

> sapply(1:5, function(x) which.max(a.res[, x]))

X3590 X3590 X3593 X3593 X3593
      1      1      4      4      4

```

This observation is mostly consistent with the predominance diagram shown before. However, the model protein for site 2 appears in the diagram but, per-residue, is not more stable than the others at the conditions of site 2. The disconnect is caused by the different numbers of residues in the model proteins, and the fact that the predominance diagram shows the relative stabilities of the proteins (based on per-residue formation reactions). Calculate size-adjusted per-residue affinities by subtracting the logarithm of the protein length from the per-residue affinities. (A related operation, for calculating activities of proteins, is described following Eq. 19 in Ref. [1].)

```

> a.res <- a.res - log10(p1)
> a.res
      [,1]      [,2]      [,3]      [,4]      [,5]
X3590 -21.02005 -30.19361 -37.68553 -39.12237 -44.56462
X3591 -21.11156 -30.17914 -37.58461 -39.00506 -44.38563
X3592 -21.41209 -30.34503 -37.64079 -39.04048 -44.34310
X3593 -21.55297 -30.36360 -37.55956 -38.94034 -44.17180
X3594 -21.53100 -30.35513 -37.56209 -38.94494 -44.18422

> sapply(1:5, function(x) which.max(a.res[, x]))

X3590 X3591 X3593 X3593 X3593
      1      2      4      4      4

```

Now there is a 1–2–4 progression in relative stabilities, which matches the stability fields in the equilibrium predominance diagram shown above.

## 5.4 Activities Along a Combined Chemical Gradient

The predominance diagrams above only shows the most stable protein at any temperature-log  $a_{H_2(aq)}$  combination. Combine the gradients of temperature, pH and hydrogen activity to plot the metastable equilibrium activities of the proteins. The total activity of residues is set by reference activities of the proteins equal to  $10^{-3}$ . Run the calculation for the distance intervals in xpoints.



```

> species(1:5, -3)

      HC03-      H2O      NH3      HS-      H2      H+ ispecies logact state
1 1020.547 -2772.768 270.298 5.694 2147.462 1026.241      3590      -3      aq
2  935.793 -2541.750 251.890 5.381 1963.205  941.174      3591      -3      aq
3  981.834 -2666.209 271.402 5.362 2047.264  987.196      3592      -3      aq
4  952.439 -2583.005 266.317 5.834 1978.491  958.273      3593      -3      aq
5  941.702 -2554.448 263.077 5.750 1957.038  947.452      3594      -3      aq
      name
1 overall_bisonN
2 overall_bisonS
3 overall_bisonR
4 overall_bisonQ
5 overall_bisonP

> xT <- Tfun(xpoints)
> xpH <- pHfun(xpoints)
> xH2 <- H2prot(xT)
> a <- affinity(T=xT, pH=xpH, H2=xH2)

energy.args: pressure is Psat
energy.args: variable 1 is T at 89 values from 330.25 to 366.45 K
energy.args: variable 2 is pH at 89 values from 7.35 to 8.257
energy.args: variable 3 is log_a(H2) at 89 values from -6.7175 to -4.0025
subcrt: 11 species at 89 values of T and P (wet)
subcrt: 18 species at 89 values of T and P (wet)

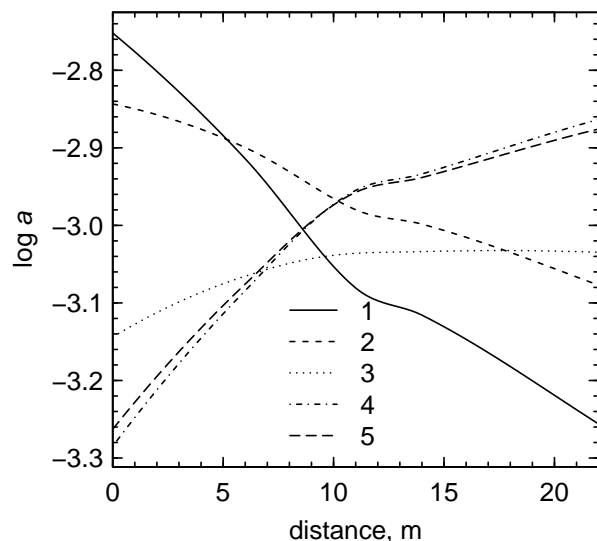
Label the x-axis "distance". To do so modify a couple of entries in the list returned by affinity (vars,
vals); otherwise the x-axis would represent temperature (the first variable in the argument list to affinity).
After Fig. 5f in Ref. [1].

> a$vars[1] <- "distance, m"
> a$vals[[1]] <- xpoints
> e <- equilibrate(a, normalize=TRUE)

balance: coefficients are protein length
equilibrate: balancing coefficients are 199.13 184.347 195.485 191.803 189.402
equilibrate: logarithm of total protein length is -0.0176532243866235
equilibrate: normalizing molar formulas by the balancing coefficients

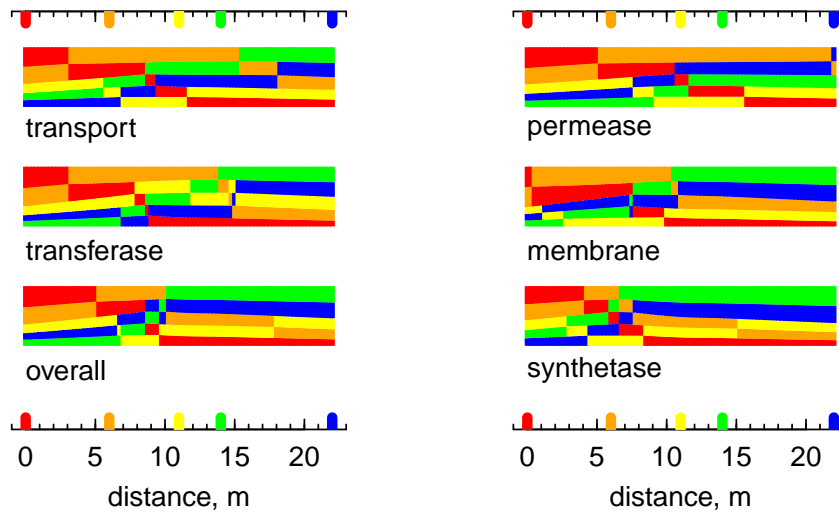
> diagram(e, legend.x=NULL)
> legend("bottom", lty=1:5, legend=paste(1:5, "      "), bty="n")

```



Plot the equilibrium degrees of formation as a function of distance for six different classes of proteins. Calculate the affinities for all of the proteins. For each group of three, make strip charts using the `strip` function in CHNOSZ. The heights of the bars represent the relative abundances of the model proteins. The five steps of the color code go from site 1 (red) to site 5 (blue). The messages printed by CHNOSZ during the calculations are not shown here.

```
> xclasses <- c("overall", "transferase", "transport", "synthetase", "membrane", "permease")
> loadclass(xclasses)
> a <- affinity(T=xT, pH=xpH, H2=xH2)
> a$vars[1] <- "distance, m"
> a$vals[[1]] <- xpoints
> col <- c("red", "orange", "yellow", "green", "blue")
> par(mfrow=c(1, 2), mar=c(4, 4, 1, 1))
> for(i in 1:2) {
  ispecies <- lapply((1:3)+(i-1)*3, function(x) {1:5+(x-1)*5} )
  names(ispecies) <- xclasses[(1:3)+(i-1)*3]
  strip(a = a, ispecies = ispecies, col = col, xticks = distance, cex.names = 1)
}
```



## 6 From Protein Stability to Redox Measurements

The temperature-hydrogen activity relationship depicted by the dashed line in the figures above was derived by considering the relative stabilities of the model proteins. How does it compare with measurements of redox chemistry in the hot spring? At least three types of field and analytical measurements are available for comparison: oxidation-reduction potential (ORP), dissolved oxygen, and sulfide/sulfate concentrations.

### 6.1 Oxidation-Reduction Potential

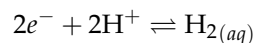
ORP, measured in the field using a portable probe and pH/voltmeter, can be converted to Eh by adding the potential of the reference electrode, in this case silver-silver chloride (Ag/AgCl) in saturated KCl. Lacking the high-temperature potentials of this electrode, the following equation from Ref. [5] for Ag/AgCl (1M KCl) is used, with temperature in degrees Celsius and potential in volts.

```
> E.AgAgCl <- function(T) {
  0.23737 - 5.3783e-4 * T - 2.3728e-6 * T^2 - 2.2671e-9 * (T+273)
}
```

Together with ORP, temperature and pH will all be needed. Data from Bison Pool and another hot spring shown in Fig. 9 of Ref. [1] are both included, but for simplicity here they are lumped together. The ORP values have units of mV.

```
> T.ORDP <- c(93.9, 87.7, 75.7, 70.1, 66.4, 66.2)
> pH.ORDP <- c(8.28, 8.31, 7.82, 7.96, 8.76, 8.06)
> ORP <- c(-258, -227, -55, -58, -98, -41)
```

Convert ORP to effective values of  $\log a_{\text{H}_2(\text{aq})}$ . First convert to Eh (in volts). Then convert Eh to pe. Then combine Eh and pH with the law of mass action for



to calculate  $\log a_{\text{H}_2(\text{aq})}$ . The law of mass action is the condition where the activity quotient (Q) of the reaction is equal to the equilibrium constant (K).

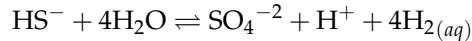
```
> Eh <- ORP/1000 + E.AgAgCl(T.ORDP)
> pe <- convert(Eh, "pe", T=convert(T.ORDP, "K"))
> logK.ORDP <- subcrt(c("e-", "H+", "H2"), c(-2, -2, 1), c("aq", "aq", "aq"), T=T.ORDP)$out$logK
```

```
subcrt: 3 species at 6 values of T and P (wet)
```

```
> logaH2.ORMP <- logK.ORMP - 2*pe - 2*pH.ORMP
```

## 6.2 Sulfide/Sulfate

For sulfide/sulfate, assign activities that are equal to concentrations (in molal units) measured in the field season of 2005, when the biofilm samples were collected for metagenomic sequencing, and use the law of mass action for



to calculate an effective activity of hydrogen. Sulfide was determined spectrophotometrically at the hot spring, and sulfate was determined using ion chromatography on water samples returned from the field.

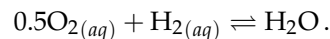
```
> loga.HS <- log10(c(4.77e-6, 2.03e-6, 3.12e-7, 4.68e-7, 2.18e-7))
> loga.S04 <- log10(c(2.10e-4, 2.03e-4, 1.98e-4, 2.01e-4, 1.89e-4))
> logK.S <- subcrt(c("HS-", "H2O", "S04-2", "H+", "H2"), c(-1, -4, 1, 1, 4),
  state=c("aq", "liq", "aq", "aq", "aq"), T=T.bison)$out$logK
```

```
subcrt: 5 species at 5 values of T and P (wet)
```

```
> logaH2.S <- (logK.S + pH.bison - loga.S04 + loga.HS) / 4
```

## 6.3 Dissolved Oxygen

Calculate the effective activity of hydrogen corresponding to the dissolved oxygen concentration using the law of mass action for



Convert the dissolved oxygen concentrations, in mg/L, to molarities (mol/L) and use these values to set the activity of  $\text{O}_{2(aq)}$ .

```
> DO <- c(0.173, 0.776, 0.9, 1.6, 2.8)
> logaO2 <- log10(DO/1000/32)
> logK <- subcrt(c("O2", "H2", "H2O"), c(-0.5, -1, 1), c("aq", "aq", "liq"), T=T.bison)$out$logK
```

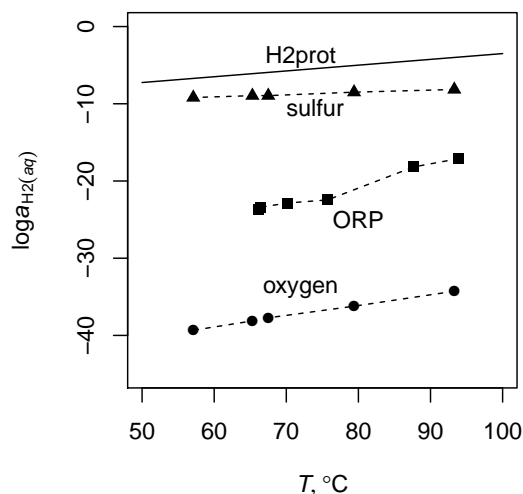
```
subcrt: 3 species at 5 values of T and P (wet)
```

```
> logaH2.O <- 0 - 0.5*logaO2 - logK
```

## 6.4 Comparison of Effective Hydrogen Activities

Make a plot showing the activities of hydrogen as a function of temperature derived from the relative stabilities of the proteins, and the ORP, sulfur and oxygen measurements. Connect the data points with dashed lines and add labels for the lines using specified coordinates. After Fig. 9 of Ref. [1].

```
> plot(Tlim, H2prot(Tlim), xlim=Tlim, ylim=c(-45,0),
  xlab=axis.label("T"), ylab=axis.label("H2"), type="l")
> points(T.ORMP, logaH2.ORMP, pch=15)
> lines(T.ORMP, logaH2.ORMP, lty=2)
> points(T.bison, logaH2.O, pch=16)
> lines(T.bison, logaH2.O, lty=2)
> points(T.bison, logaH2.S, pch=17)
> lines(T.bison, logaH2.S, lty=2)
> llab <- c("H2prot", "ORMP", "oxygen", "sulfur")
> text(c(72, 80, 72, 74), c(-4, -25, -34, -10.5), llab)
> # legend("bottomright", lty=c(1, NA, NA, NA), pch=c(NA, 15, 16, 17), legend=llab)
```



## 7 Conclusions

The calculations and figures presented here involve the average oxidation state of carbon in proteins, temperature and pH of the hot spring, relative stabilities of model proteins as a function of temperature,  $\log a_{H_2(aq)}$  and pH, and comparison of effective values of  $\log a_{H_2(aq)}$  from different redox reactions.

## 8 Document History

- 2011-08-23 Initial version (CHNOSZ\_0.9-7)
- 2012-01-07 Add `add.obigt()` for using superseded properties of methionine sidechain group.
- 2012-06-16 Use `iprotein()` instead of `protein()`.

## References

- [1] Jeffrey M. Dick and Everett L. Shock. Calculation of the relative chemical stabilities of proteins as a function of temperature and redox chemistry in a hot spring. *PLoS ONE*, 6(8):e22782, 2011. doi: 10.1371/journal.pone.0022782.
- [2] V. M. Markowitz, N. N. Ivanova, E. Szeto, K. Palaniappan, K. Chu, D. Dalevi, I. M. A. Chen, Y. Grechkin, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, P. Hugenholtz, and N. C. Kyrpides. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.*, 36:D534–D538, JAN 2008. ISSN 0305-1048. doi: 10.1093/nar/gkm869.
- [3] Jeff R. Havig, Jason Raymond, D’Arcy R. Meyer-Dombard, Natalya Zolotova, and Everett L. Shock. Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *J. Geophys. Res.*, 116:G01005, 2011. doi: 10.1029/2010JG001415.
- [4] Douglas E. LaRowe and Jeffrey M. Dick. Calculation of the standard molal thermodynamic properties of crystalline peptides. *Geochim. Cosmochim. Acta*, 80:70–91, 2012. doi: 10.1016/j.gca.2011.11.041.
- [5] A. J. Bard, R. Parsons, and J. Jordan. *Standard Potentials in Aqueous Solution*. M. Dekker, New York, 1985. URL <http://www.worldcat.org/oclc/12106344>.