Hot-spring chemistry and model proteins from a metagenome

Jeffrey M. Dick and Everett L. Shock

August 17, 2011

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.

```
> library(CHNOSZ)
```

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] "reductsse"      "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 ($\overline{Z}_\mathrm{C}$) 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 $\overline{Z}_\mathrm{C}$. 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]), axes = FALSE)
> axis(1, at = 1:5)
> axis(2)
> box()
> col <- c("green", rep("black", 20))
> lwd <- c(3, rep(1, 20))
> clab <- c("hydrolase", "overall", "protease", "oxidoreductase",
      "transport", "membrane", "permease")
> ip <- protein(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)
 }
```
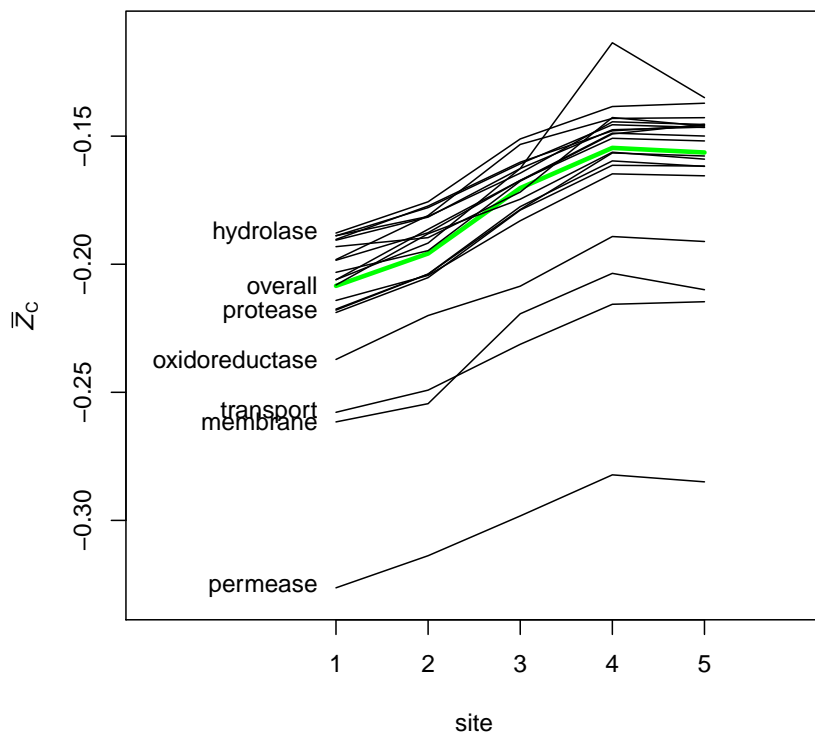
4 Chemical Setting

4.1 Basis Species

Use CHNOSZ's `basis` function to define the basis species. The basis species consist of HCO_3^-, H_2O, NH_3, HS^-, H_2 and H^+ (all aqueous species except for H_2O 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 °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(7.35, 7.678, 7.933, 7.995, 8.257)
```

The plot shows the temperature and pH as a function of distance. Use `splinefun`ctions to draw smooth lines between the points (a line formed by segments ending at the intervals in `xpoints`). Use the `mfrow` option to `par` 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))
> 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))
```
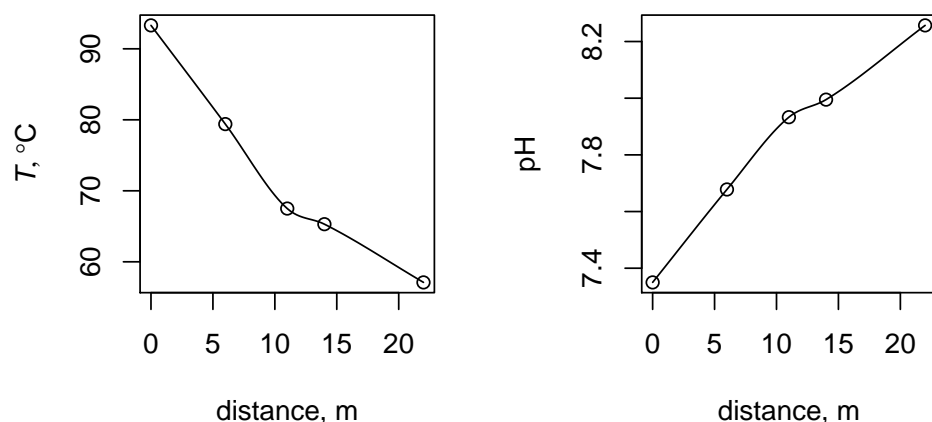
Besides temperature and pH, other chemical properties of the hot spring water change as it cools. Oxidation-reduction (redox) state of the water is left for exploration (see below), but the activities of the other basis species are set to values that approximate analytical observations. Set pH to the 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)

protein: found overall_bisonN (C1020.55H1612.76N270.30O288.87S5.69, 199.13 residues)
protein: found overall_bisonS (C935.79H1480.93N251.89O265.63S5.38, 184.347 residues)
protein: found overall_bisonR (C981.83H1550.71N271.40O279.29S5.36, 195.485 residues)
protein: found overall_bisonQ (C952.4400000000001H1506.47N266.32O274.31S5.83, 191.803 residues)
protein: found overall_bisonP (C941.7H1489.32N263.08O270.66S5.75, 189.402 residues)
     HCO3-      H2O    NH3  HS-       H2      H+ ispecies logact state          name
1 1020.55 -2772.78 270.30 5.69 2147.470 1026.24     2926     -3    aq overall_bisonN
2  935.79 -2541.74 251.89 5.38 1963.200  941.17     2927     -3    aq overall_bisonS
3  981.83 -2666.20 271.40 5.36 2047.265  987.19     2928     -3    aq overall_bisonR
4  952.44 -2583.01 266.32 5.83 1978.495  958.27     2929     -3    aq overall_bisonQ
5  941.70 -2554.44 263.08 5.75 1957.030  947.45     2930     -3    aq 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 <- protein("overall", sitenames)
> pl <- protein.length(-ip)
> mys <- species()
> mys[, 1:6]/pl

      HCO3-       H2O      NH3        HS-       H2       H+
1 5.125044 -13.92447 1.357405 0.02857430 10.78426 5.153618
2 5.076242 -13.78780 1.366391 0.02918409 10.64948 5.105426
3 5.022534 -13.63890 1.388342 0.02741898 10.47275 5.049953
4 4.965720 -13.46699 1.388508 0.03039577 10.31525 4.996116
5 4.971964 -13.48687 1.389003 0.03035871 10.33268 5.002323
```

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 increments from 323.15 to 373.15
energy.args: variable 2 is H2 at 128 increments from -7 to -4
affinity: loading ionizable protein groups
subcrt: 28 species at 128 values of T and P (wet)

> diagram(a, color = NULL, names = 1:5)

diagram: immobile component is protein backbone group
diagram: conservation coefficients are 199.13 184.347 195.485 191.803 189.402
diagram: using residue equivalents

> lines(Tlim, H2prot(Tlim), lty = 2)
> title(main = "overall")
```
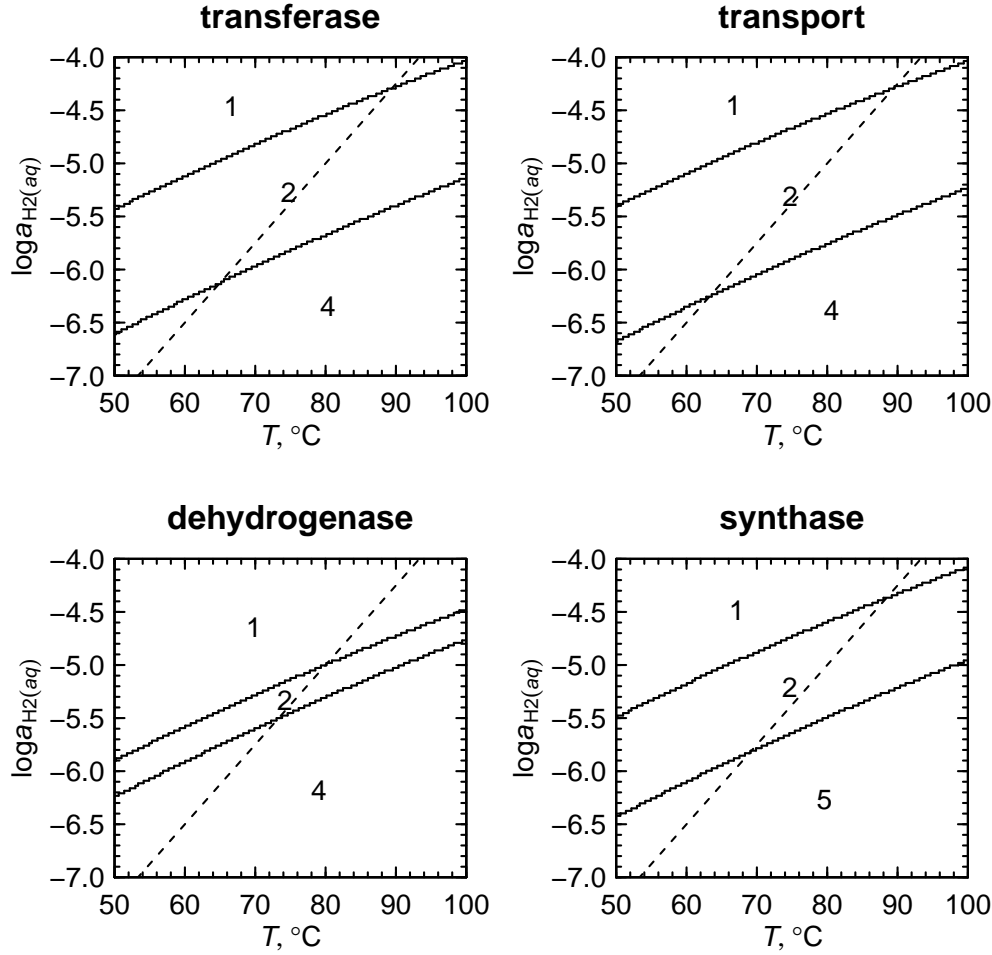
overall
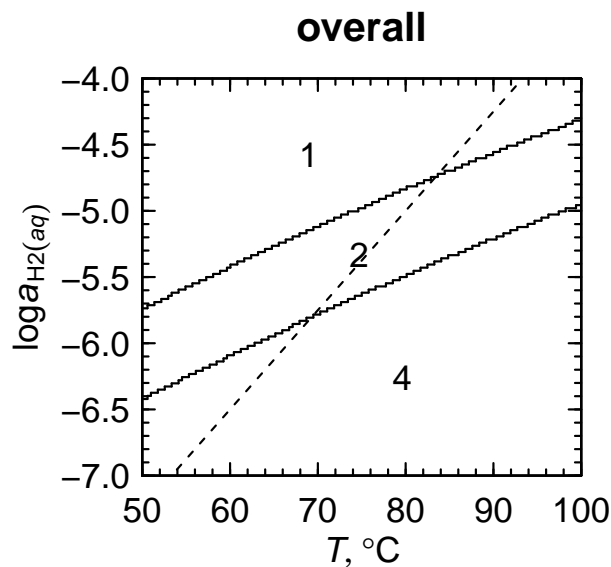
The stability fields for the proteins from the higher temperatures are at higher activities of hydrogen in the diagram. The dashed line enters the stability fields the model proteins at approximately the actual environmental temperatures.

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, color = NULL, names = 1:5)
    title(main = forclasses[i])
    lines(Tlim, H2prot(Tlim), lty = 2)
 }
> loadclass("overall")
```

transferase **transport**

dehydrogenase **synthase**

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 `pl`, to calculate the affinities per residue. After Table 5 in Ref. [1].

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

    HCO3-       H2O    NH3  HS-        H2       H+ ispecies logact state            name
1 1020.55 -2772.78 270.30 5.69 2147.470 1026.24     2926      0    aq overall_bisonN
2  935.79 -2541.74 251.89 5.38 1963.200  941.17     2927      0    aq overall_bisonS
3  981.83 -2666.20 271.40 5.36 2047.265  987.19     2928      0    aq overall_bisonR
4  952.44 -2583.01 266.32 5.83 1978.495  958.27     2929      0    aq overall_bisonQ
5  941.70 -2554.44 263.08 5.75 1957.030  947.45     2930      0    aq 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 increments from 330.25 to 366.45
energy.args: variable 2 is pH at 5 increments from 7.35 to 8.257
energy.args: variable 3 is H2 at 5 increments from -6.7175 to -4.0025
affinity: loading ionizable protein groups
subcrt: 28 species at 5 values of T and P (wet)
```

```
> a.res <- t(as.data.frame(a$values))/pl
> a.res

          [,1]      [,2]      [,3]      [,4]      [,5]
X2926 -18.72017 -27.89374 -35.38567 -36.82251 -42.26476
X2927 -18.84612 -27.91369 -35.31915 -36.73959 -42.12016
X2928 -19.12041 -28.05334 -35.34909 -36.74878 -42.05139
X2929 -19.26962 -28.08025 -35.27621 -36.65698 -41.88845
X2930 -19.25412 -28.07823 -35.28517 -36.66803 -41.90729
```

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]))

X2926 X2926 X2929 X2929 X2929
    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(pl)
> a.res

          [,1]      [,2]      [,3]      [,4]      [,5]
X2926 -21.01931 -30.19288 -37.68480 -39.12165 -44.56390
X2927 -21.11176 -30.17933 -37.58478 -39.00523 -44.38579
X2928 -21.41152 -30.34445 -37.64020 -39.03989 -44.34250
X2929 -21.55248 -30.36310 -37.55906 -38.93984 -44.17130
X2930 -21.53151 -30.35561 -37.56256 -38.94541 -44.18467

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

X2926 X2927 X2929 X2929 X2929
    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)
```

```
      HCO3-        H2O     NH3  HS-        H2        H+ ispecies logact state           name
1 1020.55 -2772.78 270.30 5.69 2147.470 1026.24     2926      -3    aq overall_bisonN
2  935.79 -2541.74 251.89 5.38 1963.200  941.17     2927      -3    aq overall_bisonS
3  981.83 -2666.20 271.40 5.36 2047.265  987.19     2928      -3    aq overall_bisonR
4  952.44 -2583.01 266.32 5.83 1978.495  958.27     2929      -3    aq overall_bisonQ
5  941.70 -2554.44 263.08 5.75 1957.030  947.45     2930      -3    aq 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 increments from 330.25 to 366.45
energy.args: variable 2 is pH at 89 increments from 7.35 to 8.257
energy.args: variable 3 is H2 at 89 increments from -6.7175 to -4.0025
affinity: loading ionizable protein groups
subcrt: 28 species at 89 values of T and P (wet)
```

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

```
> a$xname <- "distance, m"
> a$xlim <- c(range(xpoints), length(xpoints))
> a$xvals <- xpoints
> diagram(a, legend.x = NULL)
```
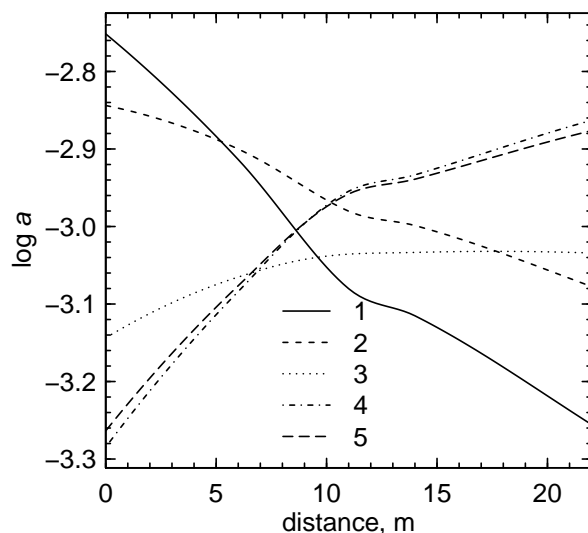
```
diagram: immobile component is protein backbone group
diagram: conservation coefficients are 199.13 184.347 195.485 191.803 189.402
diagram: using residue equivalents
diagram: log total activity of PBB (from species) is -0.01765322
```

```
> 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$xname <- "distance, m"
> a$xlim <- c(range(xpoints), length(xpoints))
> a$xvals <- xpoints
> col <- c("red", "orange", "yellow", "green", "blue")
> par(mfrow = c(1, 2))
> 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)
 }
```
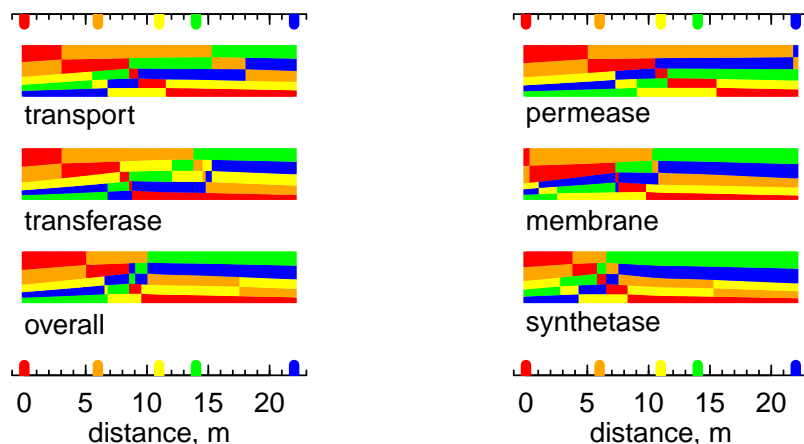

6   From Protein Stability to Redox Measurements

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. [4] for Ag/AgCl (1M KCl) is used, with temperature in degrees Celsius and potential in volts.

```
> E.AgAgCl <- function(T) {
    0.23737 - 0.00053783 * T - 2.3728e-06 * T^2 - 2.2671e-09 * (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.ORP <- c(93.9, 87.7, 75.7, 70.1, 66.4, 66.2)
> pH.ORP <- 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 $\log a_{H_{2(aq)}}$. First convert to Eh (in volts). Then convert Eh to pe. Then combine Eh and pH with the law of mass action for

$$2e^- + 2H^+ \rightleftharpoons H_{2(aq)}$$
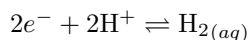
to calculate $\log a_{H_{2(aq)}}$. The law of mass action is the condition where the activity quotient (Q) of the reaction is equal to the equilibrium constant (K).

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

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

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

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

$$HS^- + 4H_2O \rightleftharpoons SO_4{}^{-2} + H^+ + 4H_{2(aq)}$$
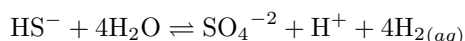
to calculate an effective activity of hydrogen. Sulfide was determined spectrophotometrically at the hot spring, and sulfate was determined using ion chromatography on water samples returned from the field.

```
> loga.HS <- log10(c(4.77e-06, 2.03e-06, 3.12e-07, 4.68e-07, 2.18e-07))
> loga.SO4 <- log10(c(0.00021, 0.000203, 0.000198, 0.000201, 0.000189))
> logK.S <- subcrt(c("HS-", "H2O", "SO4-2", "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.SO4 + 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

$$0.5\,O_{2(aq)} + H_{2(aq)} \rightleftharpoons H_2O\,.$$
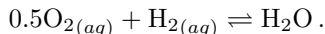
Convert the dissolved oxygen concentrations, in mg/L, to molarities (mol/L) and use these values to set the activity of $O_{2(aq)}$.

```
> 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.ORP, logaH2.ORP, pch = 15)
> lines(T.ORP, logaH2.ORP, 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", "ORP", "oxygen", "sulfur")
> text(c(72, 80, 72, 74), c(-4, -25, -34, -10.5), llab)
```
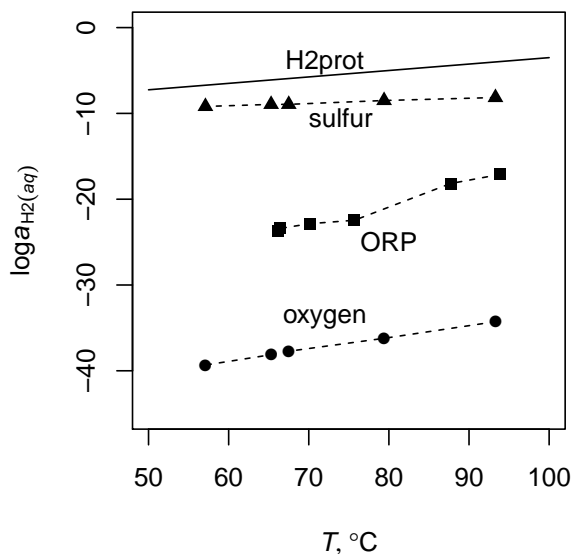
7 Conclusions

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, $\log a_{\mathrm{H}_{2(aq)}}$ and pH, and comparison of effective values of $\log a_{\mathrm{H}_{2(aq)}}$ from different redox reactions.

References

[1] J. M. Dick and E. 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] J. R. Havig, J. Raymond, D. Meyer-Dombard, N. Zolotova, and E. Shock. Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *J. Geophys. Res. - Biogeosciences*, 116:G01005, 2011. doi: 10.1029/2010JG001415.

[4] 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.