Geochemical Transactions

Research article

Calculation of the relative metastabilities of proteins using the CHNOSZ software package

Jeffrey M Dick*[1]

Department of Earth and Planetary Science, University of California, Berkeley, CA 94720, USA

E-mail: Jeffrey M Dick* - jedick@berkeley.edu;
*Corresponding author

Abstract

Background: Proteins of various compositions are required by organisms inhabiting different environments. The energetic demands for protein formation are a function of the compositions of proteins as well as geochemical variables including temperature, pressure, oxygen fugacity and pH. The purpose of this study was to explore the dependence of metastable equilibrium states of protein systems on changes in the geochemical variables.

Results: A software package called CHNOSZ implementing the revised Helgeson-Kirkham-Flowers (HKF) equations of state and group additivity for ionized unfolded aqueous proteins was developed. The program can be used to calculate standard molal Gibbs energies and other thermodynamic properties of reactions and to make chemical speciation and predominance diagrams that represent the metastable equilibrium distributions of proteins. The approach takes account of the chemical affinities of reactions in open systems characterized by the chemical potentials of basis species. The thermodynamic database included with the package permits application of the software to mineral and other inorganic systems as well as systems of proteins or other biomolecules.

Conclusion: Metastable equilibrium activity diagrams were generated for model cell-surface proteins from archaea and bacteria adapted to growth in environments that differ in temperature and chemical conditions. The predicted metastable equilibrium distributions of the proteins can be compared with the optimal growth temperatures of the organisms and with geochemical variables. The results suggest that a thermodynamic assessment of protein metastability may be useful for integrating bio- and geochemical observations.

Background

Owing to the growing body of compositional data for microbial proteins and the exploration of environments that are extreme from the human standpoint, it has become possible in recent years to draw correlations between the compositions of proteins and environmental parameters such as temperature [1]. Accounting for the underlying causes of the observed correlations between environmental parameters and protein composition is an ongoing challenge. Biochemical approaches are based in part on the notion that proteins from thermophilic and hyperthermophilic organisms should have greater structural stabilities than their mesophilic counterparts [2]. Compositional features of thermophilic proteins that may enhance their structural stabilities include increased numbers of hydrophobic residues, stronger charge interactions on the protein surfaces, and other properties of the amino acid

sequence [3]. However, it has also been suggested that, at least for sulfur, the elemental makeup of proteins is correlated with the chemical compositions of the environment [4]. This study was motivated by the desire to explore a possible thermodynamic explanation for the relationship between protein composition and the extracellular environment, which is shaped in part by geochemical constraints.

A thermodynamic assessment of protein metastability provides a framework for describing the relationship between geochemistry and protein composition that until now has received relatively little attention. The geochemical literature abounds with examples of theoretical calculation of the compositions of stable and/or metastable equilibrium reference states as a way to predict the distributions of, and reaction pathways among, minerals and inorganic or organic aqueous species [5, 6]. In recent years, the calculation [7-11] and experimental investigation [12-14] of metastable equilibrium states in biogeochemical systems has gained traction. The primary advantage of extending a framework of this type to proteins and other biomacromolecules is that it places biochemical reactions in the same context as observations on the inorganic systems to which microbial metabolic pathways are coupled. Temperature, pressure, oxidation state and pH are just some of the variables that are commonly measured in geochemical studies that also appear explicitly in the thermodynamic representation of protein metastability reactions.

This study was undertaken in order to explore the thermodynamic relationships between geochemical variables and protein composition for model proteins from a number of organisms adapted to different environments. The cell-surface glycoproteins in archaea and the surface-layer proteins in bacteria [15, 16] were chosen for this purpose because they are intimately associated with the extracellular aquatic and mineralogical setting.

Because experimental values of the standard molal Gibbs energies of the model proteins were not available, they were calculated using previously reported group additivity and equations of state algorithms that are referenced to ionized unfolded aqueous proteins [17, 18]. These values are requisite for calculating the composition of the metastable equilibrium state in an open system described by chemical potentials of basis species, or perfectly mobile components [19-22]. The predicted chemical activities of species can then be displayed on chemical predominance and/or speciation diagrams whose axes correspond to intensive chemical variables. Because of the lack of integration of algorithms for calculating thermodynamic properties of proteins in

available geochemical equilibrium software packages, the task of calculating and graphically representing the metastable equilibrium distributions of the proteins was managed through development of the CHNOSZ software package, which is introduced in this study.

The implementation of the thermodynamic algorithms and data into the package is described first below. The results of the calculations for the model system of proteins are then described and are displayed primarily in the form of diagrams depicting the calculated metastable equilibrium distributions of the proteins. The graphical depictions shown below are only limited portrayals of the metastable equilibrium states of systems of proteins, which are in fact multidimensional functions of thermodynamic variables. The predicted response of at least one of the metastability reactions between proteins from hyperthermophilic and mesophilic organisms appears to be aligned with the differences in temperature, pressure and oxidation state between their environments. However, more tests in other systems will be required to assess the generality of the approach. Some potential implications of the findings are addressed briefly in the concluding remarks, and the paper is finished with a section devoted to the methods adopted for writing protein metastability reactions and computing their thermodynamic properties.

Implementation

The CHNOSZ software package consists of source code, data files, and documentation. It is written for the cross-platform R software environment [23]. The package can be freely downloaded from the project website at http://www.chnosz.net. The features of the package, its basic program structure, and the thermodynamic database are summarized in the following paragraphs.

Features

CHNOSZ was developed in order to ease calculations of 1) the standard molal thermodynamic properties of chemical species and reactions as a function of temperature and pressure, 2) the standard molal thermodynamic properties and equations of state parameters of neutral and ionized proteins using group additivity algorithms, 3) the chemical affinities of formation reactions of species of interest from basis species describing the system, and to assist in 4) generating metastable equilibrium activity diagrams for systems of biomolecules and/or other species.

The functions provided in CHNOSZ are suitable for either interactive use or scripted operation. The diagrams that are produced can be viewed on screen or saved as postscript files. Because the thermodynamic database

includes the chemical formulas of species in addition to their standard molal thermodynamic properties, functions operating on user-input chemical reactions have the option to check, and possibly automatically correct, the mass balance of the reactions. This feature can speed up user interaction with the program and the writing of program scripts. The program has been designed with features in mind and is not presently optimized for speed. Most of the diagrams shown below can be produced in under a minute, but temperature-pressure diagrams of the same resolution require substantially more computational time, owing to the number of times the equations of state subroutines are called.

The package was developed with the goal of analyzing protein reactions, but the range of systems that can be studied using the software is limited only by the species available in the thermodynamic database, to which the user can make either temporary or persistent additions or updates. Complete documentation of the functions, including examples derived from the geochemical literature and this study, is provided with the package. Usage of the major functions in CHNOSZ is summarized below.

Standard molal properties

The relationships among the primary functions provided in CHNOSZ and some of the accessory functions are depicted in the flowchart shown in Fig. 1. Calculation of the standard molal thermodynamic properties of species and chemical reactions as a function of temperature and pressure is implemented in the primary function subcrt. The name of this function is a variation of the name of the SUPCRT92 software package [24]. The temperature and pressure ranges of calculations possible using subcrt are the same as those for SUPCRT92.

The accessory function water implements two computational options for calculating the thermodynamic and electrostatic properties of liquid H_2O as a function of temperature and pressure. The first of these options provides an interface to the FORTRAN subroutine named H2O92D.F that was distributed with SUCPRT92 [24] and that is included in the CHNOSZ source package. The calculation of the properties of liquid H_2O in this case is consistent with data and equations from Refs. [25-27] and others (see Ref. [24]). The stated temperature and pressure limits of applicability for these calculations, described in Ref. [24], are from $0.01°C$ and P_{SAT} (*i.e.*, 1 bar at temperatures below $100°C$ and the saturation vapor pressure of H_2O at higher temperatures) to $2250°C$ and 30000 bar. However, electrostatic properties of the solvent, which are required by the revised Helgeson-Kirkham-Flowers (HKF) equations of state for aqueous
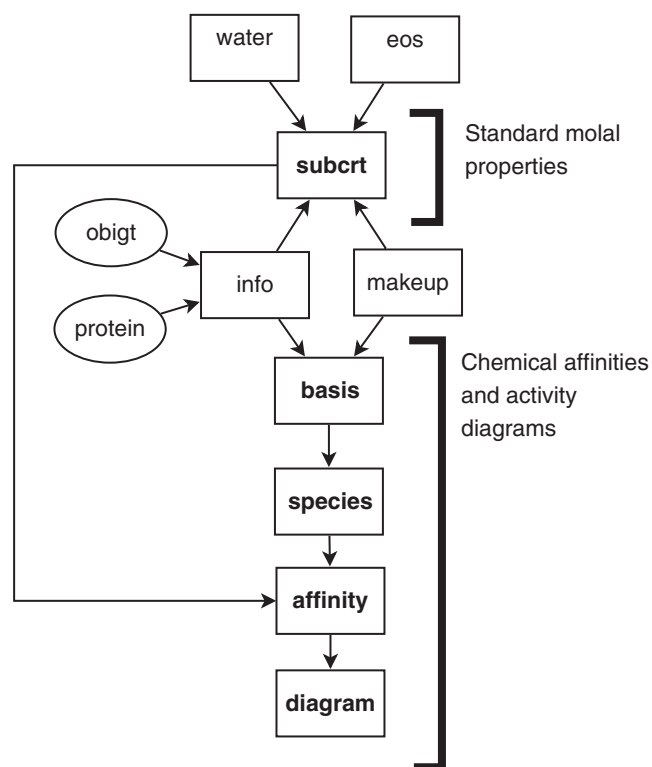

Figure 1
Functions and data flow in the CHNOSZ program. Data sources are represented by ellipses, and functions by boxes. Computations in CHNOSZ are initiated by the user accessing the primary functions, shown in bold font. The accessory functions, shown in normal font, perform many of the underlying calculations.

species, can not be computed above $1000°C$ and 5000 bar. An alternative computational option for the properties of liquid H_2O corresponds to the IAPWS-95 formulation for thermodynamic properties [28] coupled with equations for electrostatic properties taken from Ref. [29].

The functions denoted by eos in Fig. 1 actually consist of two functions, hkf, for calculating as a function of temperature and pressure the standard molal thermodynamic properties of aqueous species using the revised HKF equations of state [30-33], and cgl, for calculating the properties of crystalline, gaseous and liquid (except H_2O) species. The heat capacity equation implemented in CHNOSZ for these species contains up to six terms, as used in Ref. [34]; the first three terms are those in the Maier-Kelley equation [35, 36] which is used in the SUPCRT92 package.

The accessory function info provides a bridge between the thermodynamic and protein databases and the other

functions. The function known as makeup is concerned with conversion between various computer- and human-readable representations of the chemical compositions of species. Its primary purpose is to transform the chemical formulas of species contained in the thermo-dynamic database (*e.g.*, 'C4H6NO4-' for aspartate) into dataframe objects (which in R are similar to matrices with named columns and rows) so that other functions or makeup itself can perform further calculations on the stoichiometries of species. This function is also responsible for transforming a compositional dataframe back into a one-line chemical formula, and for calculating the reaction coefficients of basis species in formation reactions of the species of interest. It is with the aid of this function that subcrt checks whether a user-input chemical reaction is balanced with respect to mass and charge and automatically corrects the reaction if the necessary basis species have been defined.

Examples of the usage of the info and subcrt functions are shown in the program transcript in Fig. 2. The standard molal thermodynamic properties at 25°C and 1 bar and the equations of state parameters of chicken lysozyme (LYSC_CHICK, accession no. P00698 in the Swiss-Prot database [37]) can be retrieved using the code shown in Fig. 2a. The properties and parameters whose values appear in the example are standard molal Gibbs energy ($\Delta G°$) and enthalpy ($\Delta H°$) of formation from the elements (cal mol^{-1}), standard molal entropy ($S°$), heat capacity ($C_P°$) and c_1 (cal K^{-1} mol^{-1}), standard molar volume ($V°$) (cm^3 mol^{-1}), a_1 (cal bar^{-1} mol^{-1}), a_2 and ϖ (cal mol^{-1}), a_3 (cal K bar^{-1} mol^{-1}), and a_4 and c_2 (cal K mol^{-1}). The parameters a_1, a_2, a_3, a_4, c_1, c_2 and ϖ are species-dependent coefficients in the revised HKF equations of state. Note that the properties and parameters of proteins returned by info are those of nonionized proteins; the ionization contributions to thermodynamic properties of proteins are calculated using a separate function. Sample code for calculating the standard molal thermodynamic properties of LYSC_CHICK as a function of temperature at P_{SAT} is shown in Fig. 2b, where the units are °C (T), bar (P), g cm^{-3} (ρ, density of water) and those listed above for the standard molal properties. The reaction-balancing feature of subcrt is demonstrated in Fig. 2c for Reaction 1 (below). In this mode, all the user has to do is identify the basis species in the system and the reaction coefficients of the proteins, and the program finds the correct quantities of basis species to add to the reaction.

Chemical affinities and metastability diagrams
The primary function subcrt and the related accessory functions permit calculation of the standard molal Gibbs energies of protein formation reactions and correspond-ing values of the equilibrium constants (K_r in Eqn. M7).

Calculation of the activity products and chemical affinities of reactions (Q_r and A_r in Eqn. M7) is implemented in the sequence of primary functions basis, species, affinity that is depicted in Fig. 1.

Two conditions are required of a valid set of basis species in CHNOSZ: 1) the number of basis species is equal to the number of elements (and charge, if present). 2) The stoichiometric matrix denoting the elemental composi-tion (and charge if present) of the basis species, which is square according to condition (1), is non-singular and has a real inverse. These two conditions ensure that a formation reaction for any species of interest in the system can be written using only positive or negative real numbers as reaction coefficients on the basis species. The basis species themselves can be any species that are present in the thermodynamic database, including nonionized proteins. The function basis also permits redefining the physical states of basis species (if a corresponding species in that state is present in the thermodynamic database) and/or setting the activities (a) or fugacities (f) of the basis species to be used in the following calculations. These values have default settings given by log a = -3 for aqueous species, log f = 0 for gases and log a = 0 for other species. The function basis can also be used to assign a buffer to one or more basis species so that the activities or fugacities of those basis species are taken from the buffer system.

After defining the basis species, the user can select any number of species of interest using the primary function species. The user may also call species to remove species or to alter the chemical activities or fugacities of the species of interest to be used in the calculations of chemical affinity. These values default to log a = -3 for aqueous species, log f = 0 for gases and log a = 0 for other species.

The function affinity permits calculation of log Q_r and A_r of formation reactions (such as those represented generically by Reaction M1) using Eqn. (M7) taking into account the activities and/or fugacities of the basis species and the species of interest. The contributions of the Q_r and K_r terms to the calculation are denoted conceptually in Fig. 1 by the two arrows, from the top and left, respectively, pointing toward the box labeled affinity. The calculations of chemical affinity can be carried out at a single point in temperature, pressure, chemical activity space, or as a function of one or two of T, P and logarithms of chemical activity or fugacity of the basis species. The accessory function buffer is invoked by affinity if one or more basis species were previously associated with a buffer system; the activities or fugacities of the basis species constrained in this way are then used by the program to calculate log Q_r using Eqn. (M5).

(*a*)
```
> info('LYSC_CHICK')
protein: found LYSC_CHICK (C613H959N193O185S10, 129 residues).
info: 3021 refers to LYSC_CHICK, C613H959N193O185S10 aq (BBA+03).
> info(3021)
          name abbrv              formula state source1 source2 date        G        H
3021 LYSC_CHICK <NA> C613H959N193O185S10   aq  BBA+03    <NA>  <NA> -4206050 -10369700
          S      Cp        V      a1    a2    a3       a4     c1      c2   omega Z
3021 4175.86 6415.553 10420.89 251.258 34588 450.87 -4095000 7768.7 -7015000 -794000 0
```

(*b*)
```
> subcrt('LYSC_CHICK')
subcrt: 1 species at 15 values of T and P (wet).
$species
           name              formula state ispecies
3021 LYSC_CHICK C613H959N193O185S10    aq     3021
$out
$out$LYSC_CHICK
         T        P      rho     logK        G         H        S        V       Cp
1    0.01 1.000000 0.9998289 3286.749 -4108100 -10510350 3684.406 10049.21 4409.319
2   25.00 1.000000 0.9970614 3083.063 -4206050 -10369700 4175.860 10420.95 6415.518
3   50.00 1.000000 0.9880295 2919.759 -4317263 -10199867 4722.506 10600.23 7073.981
4   75.00 1.000000 0.9748643 2788.465 -4442106 -10018826 5261.992 10708.15 7376.581
5  100.00 1.013220 0.9583926 2682.511 -4580176  -9832092 5779.892 10782.93 7548.444
6  125.00 2.320144 0.9390726 2596.618 -4730054  -9641537 6273.262 10840.94 7665.198
7  150.00 4.757169 0.9170577 2526.937 -4892670  -9448098 6742.892 10891.15 7760.881
8  175.00 8.918049 0.8923427 2470.400 -5065799  -9251881 7190.897 10940.05 7856.945
9  200.00 15.536499 0.8647434 2424.600 -5249238  -9052464 7620.111 10994.29 7974.628
10 225.00 25.478603 0.8338733 2387.629 -5442322  -8848798 8034.140 11063.40 8146.453
11 250.00 39.736493 0.7990719 2357.064 -5644451  -8638774 8438.039 11165.02 8439.891
12 275.00 59.431251 0.7592362 2334.423 -5855127  -8417956 8840.380 11338.89 9030.924
13 300.00 85.837843 0.7124075 2316.077 -6074053  -8175407 9259.928 11694.88 10513.444
14 325.00 120.457572 0.6545772 2302.355 -6301438  -7875355 9754.686 12644.61 15882.325
15 350.00 165.211289 0.5746875 2293.455 -6539432  -7308661 10654.664 17027.35 60917.017
```

(*c*)
```
> basis('CHNOS')
    C H N O S ispecies logact state
CO2 1 0 0 2 0       69     -3    aq
H2O 0 2 0 1 0        1      0   liq
NH3 0 3 1 0 0       68     -4    aq
H2S 0 2 0 0 1       70     -7    aq
O2  0 0 0 2 0     2469    -80   gas
> subcrt(c('CSG_METVO','CSG_METJA'),c(-1/553,1/530),T=c(25,50,100))
protein: found CSG_METVO (C2575H4097N645O884S11, 553 residues).
protein: found CSG_METJA (C2555H4032N640O865S14, 530 residues).
subcrt: 2 species at 3 values of T and P (wet).
subcrt: reaction is not balanced; it is missing this composition:
            C         H         N         O          S
  -0.1643352 -0.1988672 -0.04118189 -0.03352213 -0.006523593
subcrt: adding missing composition from basis definition and restarting...
subcrt: 7 species at 3 values of T and P (wet).
$reaction
           coeff      name       formula state ispecies
3022 -0.001808318 CSG_METVO C2575H4097N645O884S11    aq     3022
3023  0.001886792 CSG_METJA C2555H4032N640O865S14    aq     3023
69   -0.164335187      CO2                 CO2    aq       69
1    -0.031137193    water                 H2O   liq        1
68   -0.041181890      NH3                 NH3    aq       68
70   -0.006523593      H2S                 H2S    aq       70
2469  0.163142721   oxygen                  O2   gas     2469
$out
    T       P      logK        G        H        S         V        Cp
1  25 1.00000 -12.09114 16495.26 17332.75 2.824065 -4.911481 -7.379423
2  50 1.00000 -11.11283 16431.84 17160.39 2.268477 -5.111446 -6.544202
3 100 1.01322  -9.57098 16341.70 16848.21 1.369472 -5.355989 -6.080541
```

Figure 2
Transcript of CHNOSZ session to calculate thermodynamic properties of proteins and reactions. Commands at the prompt (>) were entered to calculate (*a*) the standard molal thermodynamic properties at 25°C and 1 bar and equations of state parameters of nonionized chicken lysozyme (LYSC_CHICK), (*b*) the standard molal thermodynamic properties of lysozyme as a function of temperature at P_{SAT} and (*c*) the standard molal properties of the nonionized counterpart to Reaction 1 as a function of temperature at P_{SAT}.

The results of the calculations performed by affinity are accepted as input by diagram, which produces the diagrams using plotting functions provided in the R distribution. Many options are available for adding labels and legends and otherwise customizing the plot style.

Thermodynamic database

The database of thermodynamic properties packaged with CHNOSZ is contained in a file named OBIGT.csv. Work on this database was motivated by a software project developed by H. C. Helgeson and coworkers, named OrganoBioGeoTherm, that provides a Windows interface to the SUPCRT92 program (J. J. Donovan, personal communication).

The thermodynamic data file has records for over 2500 inorganic, organic and biochemical crystalline, gaseous, liquid and aqueous species. The thermodynamic data were originally taken from the data file distributed with the SUPCRT92 package. Updates since that time were taken from the SLOP98 data file downloaded from http://geopig.asu.edu and from recent reports of thermodynamic data and revised HKF equations of state parameters for aqueous inorganic and organic species, as well as proteins and other species of biogeochemical interest [[38-40], and others]. The records in the data file include the names, states and chemical formulas of the species, up to two literature citations, and values of the standard molal thermodynamic properties at 25°C and 1 bar and equations of state parameters. The comma-separated-value (.csv) file format permits rapid reading of the data file by the CHNOSZ program or other software as well as addition to or modification of the file contents by the user. The CHNOSZ package also provides utility functions that can be used to export or import thermodynamic data to or from the SUPCRT92 data file format.

The data file protein.csv of amino acid compositions of proteins has records for over 200 proteins including those referred to in the present study. The user can add the composition of a protein to CHNOSZ by modifying this file, or at run time by inputting the amino acid composition of the protein at the command line or requesting a search of the online Swiss-Prot database http://www.expasy.org[37] through the function called protein.
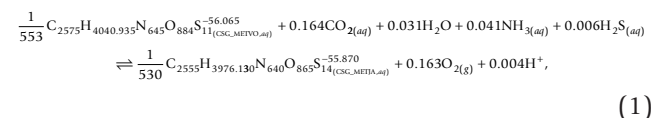
Results

The model cell-surface proteins used in this study are listed in Table 1. The selected organisms were chosen to represent diverse geochemical environments. It can be seen from the optimal growth temperatures given in Table 1 that three of the organisms (*M. jannaschii*, *M. sociabilis* and *M. fervidus*) are hyperthermophilic, others

such as *M. voltae* are mesophilic, and one organism (*M. burtonii*) is psychrotolerant. The chemical formulas and standard molal Gibbs energies of the proteins shown in Table 1 are those calculated for the nonionized aqueous proteins. Although the real proteins form crystalline or paracrystalline lattices on the cell surface [41], we are restricted at this time to using an aqueous group additivity model for lack of a crystalline analog. The present formulation is also restricted to the polypeptide molecules of proteins and does not take account of the presence of the carbohydrate chains in the glycoproteins. The standard molal Gibbs energies of ionized proteins were calculated in the present study by combining those of the nonionized proteins with ionization contributions (see Ref. [18] and the Methods).

The relative metastabilities of the model proteins were calculated as a function of temperature, pressure and chemical activities or fugacities of basis species. Results of the calculations are presented below primarily on metastable equilibrium activity diagrams depicting either the predominant protein species as a function of two intensive variables, or on speciation diagrams showing the metastable equilibrium chemical activities of proteins as a function of a single variable. The computations were carried out using the CHNOSZ software package together with a program script for use with the package that is provided in Additional File 1.

Predominance diagrams

To assess the relative metastabilities of surface-layer proteins from different organisms as a function of temperature, pressure and oxidation state, we can first write a reaction between the cell-surface proteins from *M. voltae* and *M. jannaschii* as

$$\frac{1}{553}C_{2575}H_{4040.935}N_{645}O_{884}S_{11(CSG\_METVO,aq)}^{-56.065} + 0.164CO_{2(aq)} + 0.031H_2O + 0.041NH_{3(aq)} + 0.006H_2S_{(aq)}$$

$$\rightleftharpoons \frac{1}{530}C_{2555}H_{3976.130}N_{640}O_{865}S_{14(CSG\_METJA,aq)}^{-55.870} + 0.163O_{2(g)} + 0.004H^+,$$

$$(1)$$

which is a specific statement of Reaction M2 for the ionized proteins. The coefficient in front of each of the protein formulas is the reciprocal of the number of amino acid residues in the corresponding protein. Hence, protein length is conserved in Reaction 1. Let us now write a specific statement of Eqn. (M8) for Reaction 1 as

$$\log K_1 = A_1/2.303RT + \log \frac{a_{CSG\_METJA}^{1/530}}{a_{CSG\_METVO}^{1/553}} + \log \frac{f_{O_2(g)}^{0.163}a_{H^+}^{0.004}}{a_{CO_2(aq)}^{0.164}a_{H_2O}^{0.031}a_{NH_3(aq)}^{0.041}a_{H_2S}^{0.006}},$$

$$(2)$$

where R stands for the gas constant and $\log K_1$ and A_1 denote, respectively, the logarithm of the equilibrium constant and the chemical affinity of Reaction 1.

Table 1: Model proteins used in the present study.

Organism	T_{opt} [a]	Protein[b]	ID	Length	Formula[c]	ΔG_f° [c]
Methanothermus sociabilis	88 [62]	CSG_METSC	P27374	571	$C_{2812}H_{4405}N_{747}O_{872}S_{16}$	-21875
Methanocaldococcus jannaschii	85 [43]	CSG_METJA	Q58232	530	$C_{2555}H_{4032}N_{640}O_{865}S_{14}$	-24236
Methanothermus fervidus	83 [63]	CSG_METFE	P27373	571	$C_{2815}H_{4411}N_{747}O_{872}S_{14}$	-21796
Haloarcula japonica	42 [64]	CSG_HALJP	Q9C4B4	828	$C_{3669}H_{5647}N_{971}O_{1488}$	-43458
Methanococcus voltae	38 [42]	CSG_METVO	Q50833	553	$C_{2575}H_{4097}N_{645}O_{884}S_{11}$	-24881
Methanococcoides burtonii	23 [65]	CSG_METBU	Q12YZ7	278	$C_{1362}H_{2111}N_{355}O_{442}S_4$	-11677
Acetogenium kivui	66 [66]	SLAP_ACEKI	P22258	736	$C_{3584}H_{5648}N_{926}O_{1138}S$	-29331
Bacillus stearothermophilus	65 [67]	SLAP_BACST	P35825	1198	$C_{5676}H_{9113}N_{1489}O_{1863}S_3$	-48792
Bacillus licheniformis	50 [68]	SLAP_BACLI	P49052	844	$C_{3977}H_{6396}N_{1068}O_{1286}S_2$	-33598
Aeromonas salmonicida	23 [69]	SLAP_AERSA	P35823	481	$C_{2250}H_{3580}N_{618}O_{716}S_2$	-18233

a. T_{opt} stands for laboratory optimal growth temperatures in °C, which were taken from the indicated sources. **b**. Compositions of cell-surface glycoproteins (CSG) from archaea and surface-layer proteins (SLAP) from bacteria were taken from the Swiss-Prot/UniProt database [37] using the noted accession numbers (IDs). Signal sequences were excluded where annotated, and names of the proteins except CSG_METBU correspond to those used in the Swiss-Prot database. **c**. Chemical formulas and standard molal Gibbs energies of formation from the elements at 25°C and 1 bar (ΔG_f° , in kcal mol^{-1}) were calculated for nonionized aqueous proteins.

The equal-activity boundary shown in Fig. 3a between CSG_METVO and CSG_METJA is consistent with metastable equilibrium between the proteins, or $A_1 = 0$. The location of the boundary can be calculated by combining Eqn. (2) with $A_1 = 0$, the equilibrium constant of the reaction, and the reference activities of the basis species and proteins. In this study, the reference activities of the proteins were set to 10^{-3} and those of the basis species set to the values listed in the Methods.

In Reaction 1 it can be noted that $O_{2(g)}$ appears on the same side of the reaction as $C_{2555}H_{3976.130}N_{640}O_{865}S_{14(CSG\_METJA,aq)}^{-55.870}$; hence, the metastability of this protein is increased relative to that of $C_{2575}H_{4040.935}N_{645}O_{884}S_{11(CSG\_METVO,aq)}^{-56.065}$ by decreasing log $f_{O_{2(g)}}$, which can be seen in Fig. 3a. It is also apparent from Fig. 3a that at pH 7, the formation of CSG_METJA is predicted to be favored by increasing pH. However, at pHs less than ~6, increasing pH favors formation of CSG_METVO. This observation is consistent with the variation in the charges of the proteins as a function of pH, which are shown normalized to the lengths of the proteins in Fig. 4a. For example, at pH 2, the charge per residue of CSG_METJA is greater than that of CSG_METVO, and a statement of Reaction 1 written for the proteins in their calculated ionization states at this pH would have H^+ as a reactant instead of a product. The standard molal Gibbs energies of the ionized proteins which were used to calculate log K_1 are depicted in Fig. 4b per residue of protein.

Figure 3a was generated in CHNOSZ using a sequence of commands similar to the following. The complete program script for this and the other figures is provided in Additional File 1:

```
basis("CHNOS + ")
species(c("CSG_METSC","CSG_METJA", CSG_METFE",
   "CSG_HALJP","CSG_METVO","CSG_METBU","SLAP_ACEKI",
   "SLAP_BACST","SLAP_BACLI","SLAP_AERSA"))
a < −affinity (pH = c (0,14), O2 = c (−85,−60))
diagram(a)
```

$$(3)$$

Execution of the first command shown in Example 3 defines the basis species characterizing the chemical system. Here, 'CHNOS+' is a keyword that identifies the basis species used in this paper and that appear in Reaction 1. The second command defines the species of interest, corresponding to the proteins listed in Table 1. With the third command, the chemical affinities of the formation reactions of each of the proteins are calculated on a two-dimensional grid as a function of pH and log $f_{O_{2(g)}}$ and the results assigned to a temporary object. Finally, the fourth command instructs the program to produce a metastable equilibrium activity diagram for the system, which in this case is a predominance diagram as a function of pH and log $f_{O_{2(g)}}$. The reference temperature and pressure and activities of the basis species and proteins are not explicitly specified in Example 3, and are set to default values by the program that correspond to those described in the Methods.

The approach used in CHNOSZ to make predominance diagrams does not rely on writing metastability reactions as represented by Reaction 1 but instead on using formation reactions for the proteins. For example, a specific statement of Reaction M1 for CSG_METJA in its computed ionization state at 25°C, 1 bar and pH 7 is
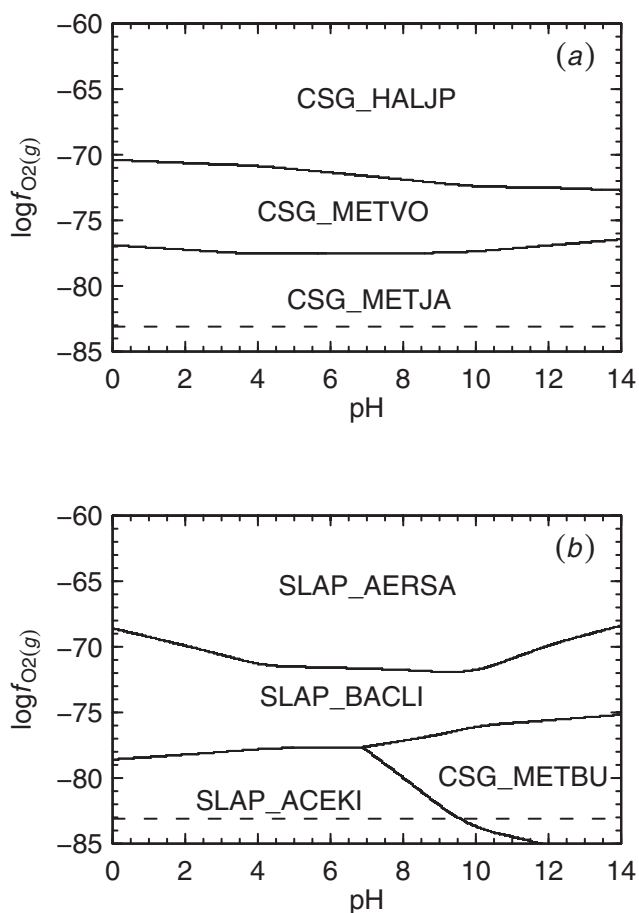
Figure 3
Relative metastabilities of proteins. log $f_{O_{2(g)}}$ -pH diagrams at 25°C and 1 bar were constructed using activities of the basis species given in the Methods. Predominance field boundaries correspond to metastable equilibrium activities of proteins equal to 10^{-3}. The diagrams were made for (*a*) all of the proteins listed in Table 1 and (*b*) the proteins listed in Table 1 except for those appearing in the first diagram. The dashed line appearing in each diagram represents the lower (reducing) stability limit of H_2O.
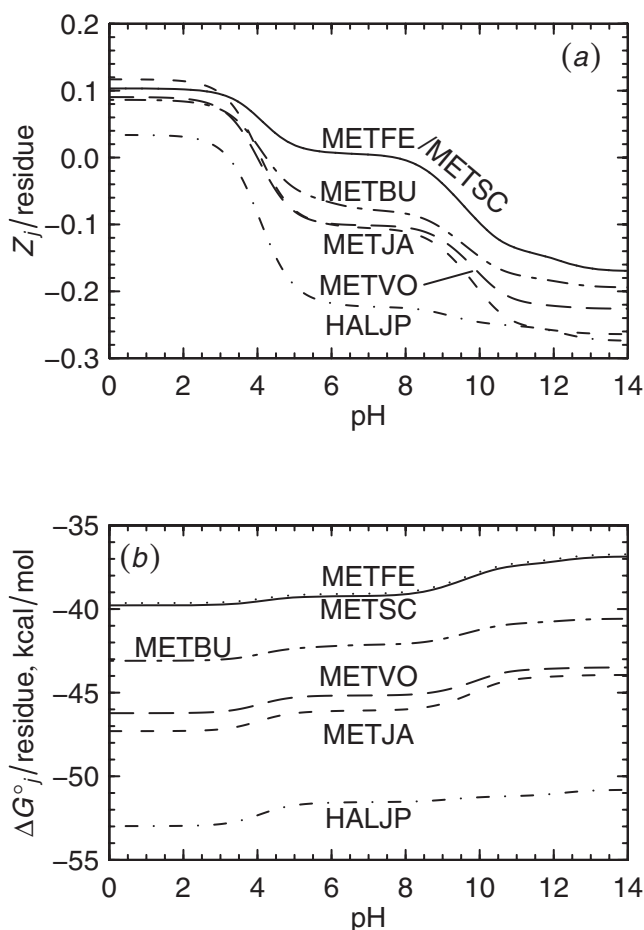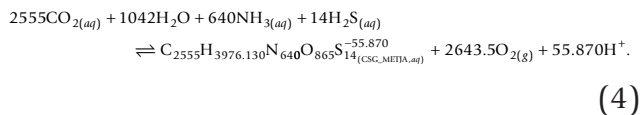

Figure 4
Properties of archaeal surface-layer proteins. Shown are calculated values of the net charge per residue (*a*) and standard molal Gibbs energy of formation from the elements (*b*) at 25°C and 1 bar for surface-layer proteins from archaeal species listed in Table 1. The computed charges per residue of CSG_METFE and CSG_METSC are indistinguishable from one another in (*a*).

$$2555CO_{2(aq)} + 1042H_2O + 640NH_{3(aq)} + 14H_2S_{(aq)}$$
$$\rightleftharpoons C_{2555}H_{3976.130}N_{640}O_{865}S_{14}^{-55.870}{}_{(CSG\_METJA.aq)} + 2643.5O_{2(g)} + 55.870H^+.$$

$$(4)$$

Using CHNOSZ, the chemical affinities of Reaction 4 and its counterparts for any other specified proteins of interest are first computed using Eqn. (M7). The chemical affinities of the formation reactions are then compared with one another to determine the theoretically predominant protein given the input conditions, which is the one with the highest chemical affinity of formation per residue. In this way, it is possible to

generate predominance diagrams like those shown in Figs. 3a and 3b for any number of proteins. The diagram shown in Fig. 3a was produced using all ten proteins listed in Table 1, but only some of the proteins predominate at different points in the diagram. Removing these proteins from consideration leads to the results shown in Fig. 3b, where the metastability relationships among some of the less metastable proteins are depicted.
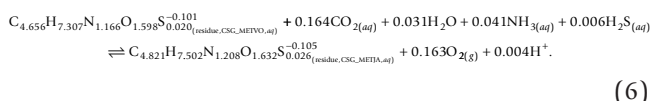
Chemical activity (speciation) diagrams
To calculate the chemical activities of proteins in metastable equilibrium, let us consider two ways of

writing the formulas of proteins in chemical reactions. The first is represented in Reaction 1 above, in which are entered the whole formulas of proteins. If the conditions are such that metastable equilibrium between the proteins in this reaction corresponds to activities of the proteins each equal to 10^{-3}, we have in Eqn. (2) $\log\left(a_{\text{CSG\_METJA}}^{1/530} / a_{\text{CSG\_METVO}}^{1/553}\right)$ = -0.0002. If we decrease $\log f_{\text{O}_{2(g)}}$ by a single unit, it follows from Eqn. (2) that $\log\left(a_{\text{CSG\_METJA}}^{1/530} / a_{\text{CSG\_METVO}}^{1/553}\right)$ = -0.0002 + 0.163 = 0.1628. Accordingly, supposing that $a_{\text{CSG\_METJA}}$ is held constant at 10^{-3}, the activity of CSG_METVO would be $\sim 10^{-93}$, a vanishingly small quantity. The relative metastabilities of proteins computed using this approach are shown graphically in Fig. 5a, where it can be seen that the logarithms of activities of the non-predominant proteins drop precipitously.

Let us propose to write the formulas of proteins in metastability reactions as residue equivalents instead of whole protein formulas. The chemical formula or any standard molal thermodynamic property of a residue equivalent of a protein is defined to be that of the protein divided by the length of the protein. In contrast, assuming activity coefficients of proteins and residue equivalents to be unity, the chemical activity of the residue equivalent of the jth protein ($a_{\text{residue}, j}$) is equal to the chemical activity of the protein (a_j) multiplied by the length of the protein (n_j):

$$a_{\text{residue}, j} = n_j \times a_j. \qquad (5)$$

We can rewrite Reaction 1 in terms of the residue equivalents of the proteins as

$$C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(\text{residue,CSG\_METVO},aq)}^{-0.101} + 0.164 CO_{2(aq)} + 0.031 H_2O + 0.041 NH_{3(aq)} + 0.006 H_2S_{(aq)}$$
$$\rightleftharpoons C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(\text{residue,CSG\_METJA},aq)}^{-0.105} + 0.163 O_{2(g)} + 0.004 H^+.$$
$$(6)$$

In Reaction 6, the coefficients on the reactant and product residue equivalents are both set to unity. Hence, in both Reactions 1 and 6 protein length is conserved. Using Eqn. (M8) we can write for Reaction 6,

$$\log K_6 = A_6 / 2.303RT + \log \frac{a_{\text{residue,CSG\_METJA}}}{a_{\text{residue,CSG\_METVO}}} + \log \frac{f_{\text{O}_{2(g)}}^{0.163} a_{\text{H}^+}^{0.004}}{a_{\text{CO}_{2(aq)}}^{0.164} a_{\text{H}_2\text{O}}^{0.031} a_{\text{NH}_{3(aq)}}^{0.041} a_{\text{H}_2\text{S}}^{0.006}}.$$
$$(7)$$

Let us now consider conditions such that the metastable equilibrium activities of the proteins are each equal to 10^{-3}. From Eqn. (5) we have $a_{\text{residue, CSG\_METJA}}$ = 0.530 and $a_{\text{residue, CSG\_METVO}}$ = 0.553, so $\log(a_{\text{residue,CSG\_METJA}}/a_{\text{residue, CSG\_METVO}})$ = - 0.018. Now, if $\log f_{\text{O}_{2(g)}}$ is decreased by one unit, it follows from Eqn. (7) that to maintain metastable equilibrium, $\log(a_{\text{residue,CSG\_METJA}}/a_{\text{residue, CSG\_METVO}})$ = -0.018 + 0.163 = 0.145. Supposing $a_{\text{residue,}}$
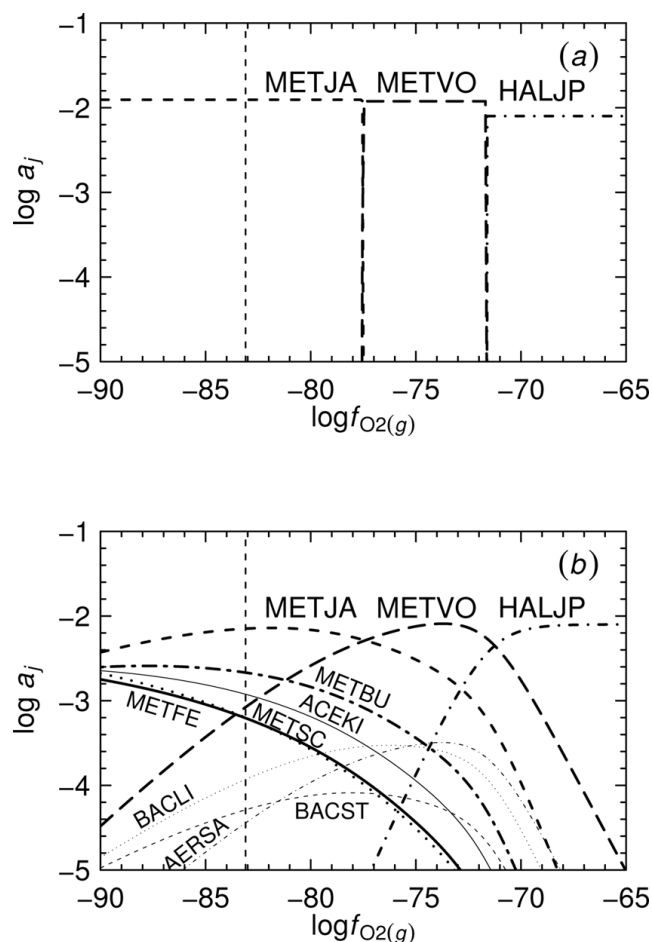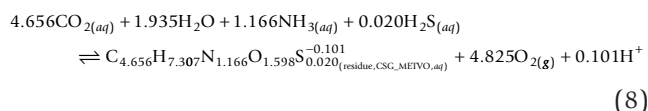

Figure 5
Metastable equilibrium chemical activities of proteins. Logarithms of chemical activities of the proteins listed in Table 1 were calculated at 25°C and 1 bar using reactions written for (*a*) whole protein formulas or (*b*) residue equivalents of the proteins. Activities of the basis species were set to the values given in the Methods, and initial activities of the proteins were set to 10^{-3}. The vertical dashed lines represent the lower stability limit of H_2O.
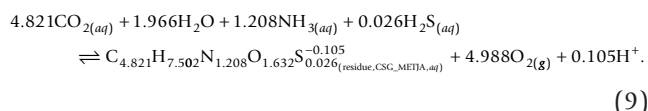
CSG_METJA to be held constant at 0.530 ($a_{\text{CSG\_METJA}} = 10^{-3}$), $a_{\text{residue,CSG\_METVO}}$ would be 0.380 ($a_{\text{CSG\_METVO}} = 10^{-3.16}$). This type of assessment leads to the results shown graphically in Fig. 5b, where it can be seen that the metastable equilibrium activities of the proteins as a function of $\log f_{\text{O}_{2(g)}}$ are within a few log units of each other, even for the non-predominant proteins.

The diagram shown in Fig. 5b was actually constructed using CHNOSZ by taking account of the formation reactions of residue equivalents of the proteins, instead of the metastability reaction represented by Reaction 6.

To demonstrate this procedure, let us write the formation reaction for the residue equivalent of CSG_METVO as

$$4.656CO_{2(aq)} + 1.935H_2O + 1.166NH_{3(aq)} + 0.020H_2S_{(aq)}$$
$$\rightleftharpoons C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(residue,CSG\_METVO,aq)}^{-0.101} + 4.825O_{2(g)} + 0.101H^+$$

(8)

and that for the residue equivalent of CSG_METJA as

$$4.821CO_{2(aq)} + 1.966H_2O + 1.208NH_{3(aq)} + 0.026H_2S_{(aq)}$$
$$\rightleftharpoons C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(residue,CSG\_METJA,aq)}^{-0.105} + 4.988O_{2(g)} + 0.105H^+.$$

(9)

Specific statements of Eqn. (M8) for Reactions 8 and 9 are, respectively,

$$A_8 / 2.303RT = \log K_8 - \log\left(a_{C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(residue,CSG\_METVO,aq)}^{-0.101}} f_{O_{2(g)}}^{4.825} a_{H^+}^{0.101}\right)$$
$$+ \log\left(a_{CO_{2(aq)}}^{4.656} a_{H_2O}^{1.935} a_{NH_{3(aq)}}^{1.166} a_{H_2S}^{0.020}\right)$$

(10)

and

$$A_9 / 2.303RT = \log K_9 - \log\left(a_{C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(residue,CSG\_METJA,aq)}^{-0.105}} f_{O_{2(g)}}^{4.988} a_{H^+}^{0.105}\right)$$
$$+ \log\left(a_{CO_{2(aq)}}^{4.821} a_{H_2O}^{1.966} a_{NH_{3(aq)}}^{1.208} a_{H_2S}^{0.026}\right).$$

(11)

At metastable equilibrium, $A_8 = A_9$, *i.e.* the chemical affinities of the formation reactions of the residue equivalents are equal. Values of log K_8 = -367.714 and log K_9 = -379.687 can be obtained using standard molal Gibbs energies at 25 °C and 1 bar of the basis species and of the ionized proteins at pH 7 (see Fig. 4b). Let us also substitute the reference activities of the basis species described in the Methods and log $f_{O_{2(g)}}$ = -80 to write

$$A_8 / 2.303RT = 0.189 - \log a_{C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(residue,CSG\_METVO,aq)}^{-0.101}}$$

(12)

and

$$A_9 / 2.303RT = 0.593 - \log a_{C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(residue,CSG\_METJA,aq)}^{-0.105}}.$$

(13)

There are three unknowns in Eqns. (12) and (13). Conservation of protein length leads to a third equation:

$$a_{C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(residue,CSG\_METVO,aq)}^{-0.101}} + a_{C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(residue,CSG\_METJA,aq)}^{-0.105}} = 1.083,$$

(14)

where the value on the right-hand side corresponds to initial activities of the proteins each equal to 10^{-3}. Solving Eqns. (12)–(14) gives $a_{C_{4.656}H_{7.307}N_{1.166}O_{1.598}S_{0.020(residue,CSG\_METVO,aq)}^{-0.101}}$ = 0.307, $a_{C_{4.821}H_{7.502}N_{1.208}O_{1.632}S_{0.026(residue,CSG\_METJA,aq)}^{-0.105}}$ = 0.776 and $A_{(8\ or\ 9)}/2.303RT$ = 0.703.

The addition of any protein to the system increases by one the number of unknowns in Eqn. (14) but also

provides another equation in the form of Eqns. (12) and (13). The procedure to set up and solve these equations has been encoded in a general form in CHNOSZ and was used to produce the diagrams shown in Fig. 5. The CHNOSZ program includes options to analyze the protein formation reactions using whole protein formulas or their residue equivalents, which were used to construct Figs. 5a and 5b, respectively. The logarithm of total activity of protein residues is 0.8211 in each of these figures, which corresponds to the sum of the activities of the residue equivalents of the ten model proteins whose starting activities are 10^{-3}.

Another way of representing the chemical speciation in a protein system is on a degree of formation diagram. The degree of formation of the kth protein (α_k) can be calculated from

$$\alpha_k = a_{\text{residue},k} / \sum_{j=1}^{j} a_{\text{residue},j},$$

(15)

where $\hat{j} = \hat{k}$ denotes the number of proteins in the system, $\sum_1^j a_{\text{residue},j}$ represents the total activity of protein residues, and $\sum_1^{\hat{k}} \alpha_k = 1$. The degrees of formation of the proteins corresponding to the logarithms of activities shown in Fig. 5b are depicted in the figure in Additional File 2. This degree of formation diagram aids in visualization of the computed relative abundances of the proteins on a non-logarithmic scale.

The residue-equivalent approach was used in this study only to produce the diagrams shown in Fig. 5b and Additional File 2. The predominance diagrams shown elsewhere were produced using whole protein formulas in the formation reactions. Extending the residue-equivalent method to these diagrams would subtly alter the positions of the predominance field boundaries, more so for reactions between proteins that differ significantly in length. The differences in the locations of the predominance field boundaries can be assessed in part by comparing the locations of the crossover between predominant proteins in Figs. 5a and 5b.

Temperature and pressure diagrams

The approach described above for constructing Fig. 3 in CHNOSZ was used to produce the diagrams shown in Figs. 6a and 6b. These diagrams portray the metastabilities among the predominant model proteins as a function of temperature or pressure and log $f_{O_{2(g)}}$. It is immediately apparent that log $f_{O_{2(g)}}$ and temperature have a close relationship along a reaction boundary. For

all of the reactions represented by the predominance field boundaries in Fig. 6a, increasing temperature is accompanied by an increase in the value of log $f_{O_{2(g)}}$. Hence, for a small positive increment in temperature at constant log $f_{O_{2(g)}}$, the metastability of CSG_METJA increases relative to that of CSG_METVO. If values of log $f_{O_{2(g)}}$ instead correspond as a function of temperature to the water stability limit (shown by the dashed line in Fig. 6a), increasing temperature would actually favor the formation of CSG_METVO relative to CSG_METJA.

We can recover nominal values of log $f_{O_{2(g)}}$ in the natural environments of *M. voltae* and *M. jannaschii* from geochemical data. The first of these organisms was originally isolated from the sediment of an estuary [42] and the other inhabits submarine hydrothermal vent

Figure 6
Temperature and pressure dependence of protein metastability. Predominance diagrams were constructed as a function of log $f_{O_{2(g)}}$ and (*a*) temperature at P_{SAT} or (*b*) pressure at 25°C. The dashed line in each diagram represents the lower stability limit of H_2O.

environments [43]. Values of $a_{H_{2(aq)}}$ (activity of dissolved hydrogen) were taken from [44] and converted to log $f_{O_{2(g)}}$ using the law of mass action for $H_2O \rightleftharpoons H_{2(aq)} + 0.5O_{2(g)}$ evaluated at 25°C and 1 bar to calculate a nominal range of log $f_{O_{2(g)}}$ for estuarine sediment of -73 to -70. Values of log $f_{O_{2(g)}}$ obtaining in mixed hydrothermal vent fluid and seawater at 100°C are in the range of -65 to -60 [45]. The first of these ranges would plot near the CSG_HALJP – CSG_METVO boundary in Fig. 6a at 25°C and the second one near the boundary between CSG_METVO and CSG_METJA at 100°C. This observation might support the notion that proteins from hyperthermophilic organisms like *M. jannaschii* are thermodynamically favored relative to those from mesophilic organisms by increasing temperature accompanied by changes in the geochemical oxidation state.

It appears in Fig. 6b that increasing pressure also generally favors those proteins in lower oxidation states, but that the dependence of equilibrium log $f_{O_{2(g)}}$ values on pressure is small relative to their dependence on temperature.
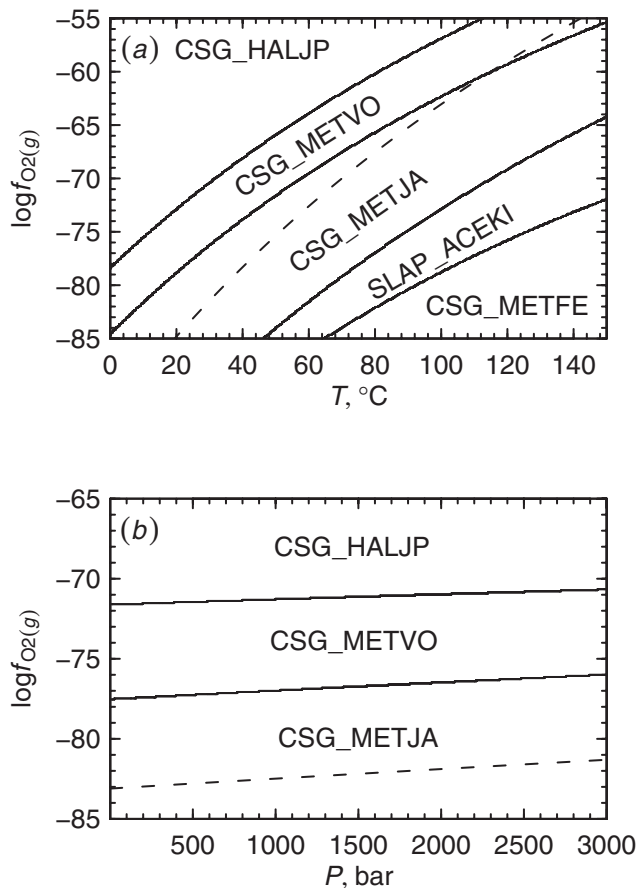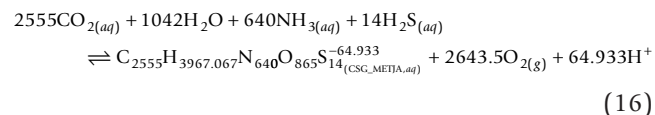
Proteins as chemical activity buffers
The chemical activities of basis species buffered by reacting protein assemblages correspond to the locations of the (pseudo)invariant points on metastable equilibrium predominance diagrams. Equal activities of three proteins correspond to the triple point, which is a pseudoinvariant point, in the predominance diagram shown in Fig. 3b. The number of independent variables on the axes of this diagram is two; in an eight-dimensional predominance diagram (of temperature, pressure and six chemical activities) one could distinguish the true invariant points in this system where nine proteins coexist with equal metastable equilibrium activities.

Let us ask what are the activities of $CO_{2(aq)}$, H_2O, $NH_{3(aq)}$ and $H_2S_{(aq)}$ if they are buffered by a hypothetical metastable assemblage made up of the proteins from the METXX organisms listed in Table 1, at T = 100°C, P = 1000 bar, pH 7, log $f_{O_{2(g)}}$ = -58 and activities of proteins equal to 10^{-3}. At this temperature, pressure and pH the calculated charge of the cell-surface protein from *M. jannaschii* is -64.933. Consider then the formation reaction for this protein, which except for charge is equal to Reaction 4:

$$2555CO_{2(aq)} + 1042H_2O + 640NH_{3(aq)} + 14H_2S_{(aq)}$$
$$\rightleftharpoons C_{2555}H_{3967.067}N_{640}O_{865}S_{14(CSG\_METJA,aq)}^{-64.933} + 2643.5O_{2(g)} + 64.933H^+$$

$$(16)$$

A rearranged statement of Eqn. (M8) for this reaction can be written as

$$530\bar{A}_{16}/2.303RT - 2555\log a_{CO_{2(aq)}} - 640\log a_{NH_{3(aq)}} - 14\log a_{H_2S_{(aq)}} - 1042\log a_{H_2O}$$
$$= \log K_{16} - \log a_{C_{2494}H_{3967.067}N_{613}O_{841}S_{9(CSG\_METJA,aq)}^{-64.933}} - 2463.5\log f_{O_{2(g)}} - 64.933\log a_{H^+}$$

$$(17)$$

where $\bar{A}_{16} \equiv A_{16}/530$, and the right-hand side works out to -4772.316 at the conditions stated above. At metastable equilibrium, the values of $\bar{A}_{16}$ and its counterparts for the other proteins in the hypothetical assemblage are all equal. It follows that we can combine Eqn. (17) with its counterparts for the four other proteins, dropping the subscripts on $\bar{A}$, to write

$$\begin{bmatrix} 530 & 2555 & 1042 & 640 & 14 \\ 571 & 2812 & 1066 & 747 & 16 \\ 553 & 2575 & 1070 & 645 & 11 \\ 278 & 1362 & 519 & 355 & 4 \\ 571 & 2815 & 1071 & 747 & 14 \end{bmatrix} \times \begin{bmatrix} \bar{A}/2.303RT \\ -\log a_{CO_{2(aq)}} \\ -\log a_{H_2O} \\ -\log a_{NH_{3(aq)}} \\ -\log a_{H_2S_{(aq)}} \end{bmatrix} = \begin{bmatrix} -4772.316 \\ -5785.204 \\ -5021.307 \\ -2683.266 \\ -5825.691 \end{bmatrix}, \quad (18)$$

where the rows on the right-hand side and in the stoichiometric matrix on the left-hand side correspond to the proteins from the METXX organisms listed in Table 1. Solving Eqn. (18) gives $\bar{A}/2.303RT$ = -0.739, $\log a_{CO_{2(aq)}}$ = -8.44, $\log a_{H_2O}$ = 7.92, $\log a_{NH_{3(aq)}}$ = 27.92 and $\log a_{H_2S_{(aq)}}$ = -13.09. These values signify that the formation reactions of the proteins per residue are energetically unfavorable ($\bar{A}$ is negative) and that the hypothetical protein assemblage may not be metastably present (for example, the large positive values for a_{H_2O} and $a_{NH_{3(aq)}}$ differ from probable natural ranges). Unambiguous identification of a natural metastable protein assemblage may require more comprehensive calculations coupled with insight gained from experiments and observations in the field.

The pseudoinvariant point representing the buffer assemblage described above is shown in Figs. 7a and 7b. The same pseudoinvariant point is present in both figures, but different variables are projected onto each diagram. The temperature-pressure relationships appearing in Fig. 7a suggest that metastability of CSG_METJA increases relative to that of CSG_METVO with increasing temperature and/or pressure, but that the sensitivity to temperature is much greater than that to pressure. These relationships are also apparent in Figs. 6a and 6b. In the projection of Fig. 7a all the proteins at the pseudoinvariant point are not visible, but in Fig. 7b convergence of the five predominance fields is apparent. Note the similarity in Figs. 7b and 3a of the reaction boundary between CSG_METVO and CSG_METJA, as well as the nearly horizontal boundary between CSG_METSC and CSG_METFE, which would be expected from the

closeness of their ionization states as a function of pH (see Fig. 4a for the ionization states at 25°C).

Concluding remarks

A computer program called CHNOSZ was introduced in this paper for producing metastable equilibrium chemical activity diagrams for proteins. The methods used here were borrowed from geochemistry, and the program with the accompanying thermodynamic database is suitable for performing thermodynamic calculations in inorganic and mineral systems as well as organic and biochemical systems, or combinations thereof.
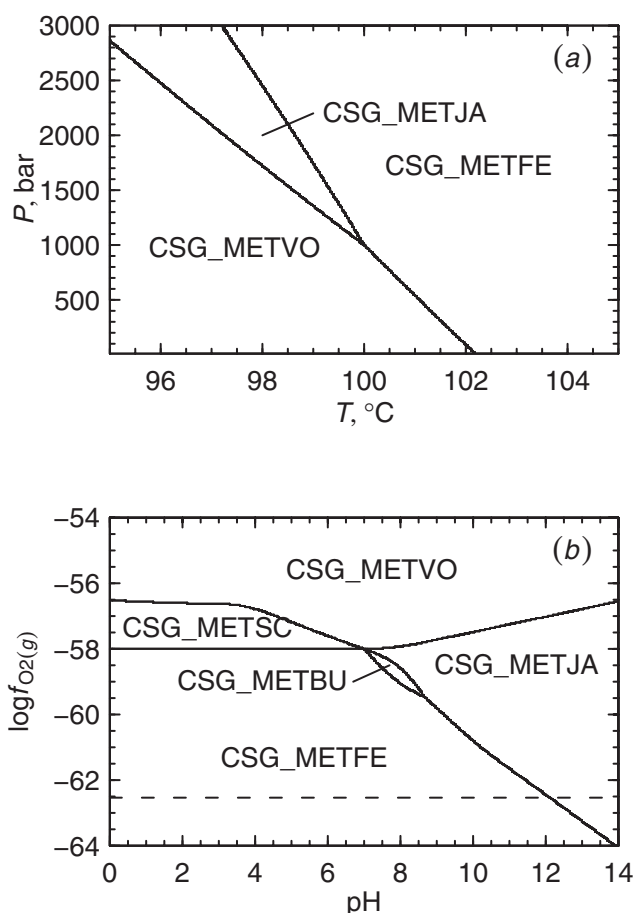
Figure 7
Metastabilities of proteins around a pseudoinvariant point. Activities of $CO_{2(aq)}$, $NH_{3(aq)}$, $H_2S_{(aq)}$ and H_2O were calculated using a buffer consisting of the proteins from the METXX organisms listed in Table 1. The variables set in the buffer calculation were activities of proteins equal to 10^{-3}, T = 100°C, P = 1000 bar, log $f_{O_{2(g)}}$ = -58 and pH 7. The diagrams show the variation of protein metastability as a function of (a) temperature and pressure and (b) log $f_{O_{2(g)}}$ and pH. The dashed line in (b) represents the lower stability limit of H_2O at this temperature and pressure.

To investigate the utility of the program for a geochemical description of protein reactions, metastability diagrams were produced for surface-layer proteins from a number of bacteria and archaea. The diagrams show either the metastably predominant proteins as a function of two intensive variables or the metastable equilibrium chemical activities of proteins as a function of one variable. The primary variables of interest in this study were log $f_{O_{2(g)}}$, pH, temperature and pressure. It was found that the predicted metastable equilibrium state of the system responded dramatically to changes in these variables. Representing the proteins in reactions by their residue equivalents instead of whole protein formulas gave rise to predicted equilibrium states in which many proteins coexist metastably with comparable chemical activities.

In the preceding sections we have considered the theoretical metastable equilibrium relationships among only a few model proteins. Because the software is now available to do so, a plethora of predictions concerning the energetically favorable outcomes of any number of overall protein mutation reactions is now within reach. Consideration of the results presented above, and of the wide range of model systems that could potentially be investigated in a similar manner, leads to the conclusion that the metastable equilibrium distribution of proteins in many cases does not mirror geobiochemical reality. Nevertheless, the ability to quantify the characteristics of metastable equilibrium reference states as a function of geochemical variables may be of utility in identifying specific pathways in evolution where the resulting proteins are relatively energetically favored. These particular outcomes may reflect a tendency for natural selection to increase the fit between phenotypes and their environments [46].

A thermodynamic and geochemical perspective on the relative metastabilities of proteins permits a quantitative integration of observations on the geosphere and biosphere. This study has only touched the surface of the myriad possible environments and organisms, the properties and chemical compositions of which are becoming more well constrained through experiment and observation. As these data grow in abundance, they will provide other opportunities where thermodynamic description of the chemical speciation of proteins can be tested and calibrated.

Methods
The thermodynamic conventions and relations used to compute the relative metastabilities of proteins in the present study are summarized below. The computational assessment depends first on the adoption of standard states for the species appearing in chemical reactions.
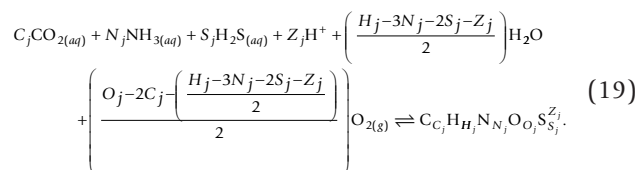
Standard state conventions
The standard state convention adopted for aqueous species other than H_2O corresponds to unit activity of a hypothetical one molal solution referenced to infinite dilution at any temperature and pressure [30, 47]. The conventional standard molal thermodynamic properties of both the aqueous electron and proton are taken to be zero at all temperatures and pressures [48]. For gases, the standard state convention is unit fugacity of the hypothetical pure ideal gas at 1 bar and any temperature. The standard state convention adopted for solids and liquids, including H_2O, corresponds to unit activity of the pure substance at any temperature and pressure.

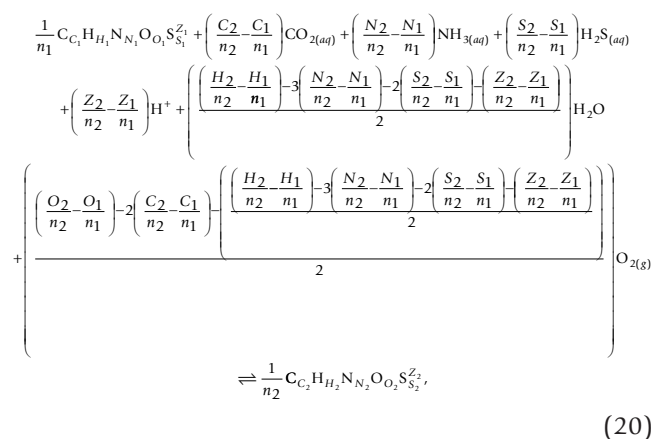### Protein formation and metastability reactions
The compositions of species of interest, such as proteins, are represented by linear combination of the compositions of basis species in a system (for an application in geochemical systems, see Ref. [49]). The number of basis species is the minimum required to write formation reactions for all possible species of interest. There are no thermodynamic restrictions on the actual identities of the basis species, and the basis species do not necessarily correspond to thermodynamic components in the system of interest [50]. Hence, the choice of basis species may be constrained by the chemical activities that can be measured in a system or that are thought to behave as perfectly mobile components [22]. The basis species used in the present study are $CO_{2(aq)}$, H_2O, $NH_{3(aq)}$, $H_2S_{(aq)}$, H^+ and $O_{2(g)}$.

Let a generic chemical formula for the jth ionized protein be written as $C_{C_j}H_{H_j}N_{N_j}O_{O_j}S_{S_j}^{Z_j}$, where C_j, H_j, N_j, O_j, S_j and Z_j denote the number of moles of the corresponding element (or charge) in one mole of protein. These coefficients can be non-integer and positive or negative (*e.g.*, Z_j usually is negative at some alkaline pHs). The formation reaction from basis species of one mole of the jth protein can be written as

$$C_j CO_{2(aq)} + N_j NH_{3(aq)} + S_j H_2S_{(aq)} + Z_j H^+ + \left(\frac{H_j - 3N_j - 2S_j - Z_j}{2} \right) H_2O$$

$$+ \left(\frac{O_j - 2C_j - \left(\frac{H_j - 3N_j - 2S_j - Z_j}{2} \right)}{2} \right) O_{2(g)} \rightleftharpoons C_{C_j}H_{H_j}N_{N_j}O_{O_j}S_{S_j}^{Z_j}. \quad (19)$$

The reaction coefficients on the basis species in Reaction M1 are completely determined by the chemical formulas of the protein and of the basis species. Depending on the sign of the coefficients in front of the basis species, they would appear in specific statements of Reaction M1 as reactants or products.

A generic metastability reaction between two proteins ($j = 1$ and $j = 2$) can be written as

$$\frac{1}{n_1}C_{C_1}H_{H_1}N_{N_1}O_{O_1}S_{S_1}^{Z_1} + \left(\frac{C_2}{n_2} - \frac{C_1}{n_1}\right)CO_{2(aq)} + \left(\frac{N_2}{n_2} - \frac{N_1}{n_1}\right)NH_{3(aq)} + \left(\frac{S_2}{n_2} - \frac{S_1}{n_1}\right)H_2S_{(aq)}$$

$$+ \left(\frac{Z_2}{n_2} - \frac{Z_1}{n_1}\right)H^+ + \left(\frac{\left(\frac{H_2}{n_2} - \frac{H_1}{n_1}\right) - 3\left(\frac{N_2}{n_2} - \frac{N_1}{n_1}\right) - 2\left(\frac{S_2}{n_2} - \frac{S_1}{n_1}\right) - \left(\frac{Z_2}{n_2} - \frac{Z_1}{n_1}\right)}{2}\right)H_2O$$

$$+ \left(\frac{\left(\frac{O_2}{n_2} - \frac{O_1}{n_1}\right) - 2\left(\frac{C_2}{n_2} - \frac{C_1}{n_1}\right) - \left(\frac{\left(\frac{H_2}{n_2} - \frac{H_1}{n_1}\right) - 3\left(\frac{N_2}{n_2} - \frac{N_1}{n_1}\right) - 2\left(\frac{S_2}{n_2} - \frac{S_1}{n_1}\right) - \left(\frac{Z_2}{n_2} - \frac{Z_1}{n_1}\right)}{2}\right)}{2}\right)O_{2(g)}$$

$$\rightleftharpoons \frac{1}{n_2}C_{C_2}H_{H_2}N_{N_2}O_{O_2}S_{S_2}^{Z_2},$$

(20)

which corresponds to the difference between specific statements of Reaction M1 for $j = 2$ and $j = 1$, divided by n_2 or n_1, respectively. Here, $1/n_1$ and $1/n_2$ denote the conservation coefficients for the corresponding proteins. Reaction M2 is balanced with respect to mass and charge for any values of n_1 and n_2. If $n_1 = n_2 = 1$, Reaction M2 denotes the mass balance constraints for the formation of one mole of product protein at the expense of one mole of reactant protein. Other values may be chosen for n_1 and n_2, depending on what is specified about the conservation constraints in the system. For example, if $n_1 = C_1$ and $n_2 = C_2$, the protein metastability reaction conserves carbon [18] (*i.e.*, the coefficient on $CO_{2(aq)}$ in Reaction M2 becomes zero). The protein metastability reactions considered in the present study are written for n_j equal to the length of the jth protein.

Relation of reaction energetics to activities of basis species

The standard Gibbs energy of the rth formation or metastability reaction (ΔG_r°) can be expressed as

$$\Delta G_r^\circ = \sum_r \hat{n}_{i,r}\Delta G_r^\circ,$$

(21)

where $\hat{n}_{i,r}$ and ΔG_i° denote, respectively, the stoichiometric reaction coefficient and standard molal Gibbs energy of formation from the elements of the ith basis species or protein in the reaction. For products in a reaction, $\hat{n}_{i,r} > 0$. The corresponding equilibrium constant of the reaction (K_r) is given by

$$\log K_r = -\Delta G_r^\circ / 2.303RT.$$

(22)

The equilibrium constant, like ΔG_r°, is a standard-state property which is is independent of composition and depends on temperature and pressure. The non-

standard-state counterpart to K_r is the activity product of the reaction (Q_r), which can be computed using

$$Q_r \equiv \prod_i a_i^{\hat{n}_{i,r}},$$

(23)

where a_i represents the chemical activity of the ith species in the reaction. For gaseous species, a_i in Eqn. (M5) is replaced by the fugacity of the species (f_i). Activity and fugacity coefficients are taken in a first approximation in this study to be unity.

The activity or fugacity of the ith aqueous or gaseous component is related to its chemical potential (μ_i) by [6]

$$\mu_i = \mu_i^\circ + RT\ln\frac{f_i}{f_i^\circ} = \mu_i^\circ + RT\ln a_i,$$

(24)

where μ_i° denotes the standard chemical potential of the ith species and f_i° stands for the fugacity of the species in its standard state, which is unity for gases.

The chemical affinities of reactions (A_r) can be computed from [51]

$$A_r = 2.303RT\log(K_r/Q_r),$$

(25)

which can be combined with Eqn. (M5) to write for Reaction M2

$$\log K_r = A_r / 2.303RT + \log\frac{a_{C_{C_2}H_{H_2}N_{N_2}O_{O_2}S_{S_2}^{Z_2}}^{1/n_2}}{a_{C_{C_1}H_{H_1}N_{N_1}O_{O_1}S_{S_1}^{Z_1}}^{1/n_1}} + \log\left(a_{CO_{2(aq)}}^{n_{CO_2}} a_{H_2O}^{n_{H_2O}} a_{NH_{3(aq)}}^{n_{NH_3}} f_{O_{2(aq)}}^{n_{O_2}} a_{H_2S_{(aq)}}^{n_{H_2S}} a_{H^+}^{n_{H^+}}\right).$$

(26)

In an equilibrium state, $A_r = 0$ for metastability reactions and Eqn. (M8) reduces to the logarithmic analog of the law of mass action equation for Reaction M2.

Reference activities of basis species and proteins

The reference temperature and pressure correspond to 25°C and 1 bar, respectively. The reference chemical activities of basis species used in this study are given by $\log a_{H_2O} = 0$, $\log a_{CO_{2(aq)}} = -3$, $\log a_{NH_{3(aq)}} = -4$, $\log a_{H_2S_{(aq)}} = -7$ and $\log a_{H^+} = -7$ (pH 7). The reference value for $\log a_{H_2O}$ corresponds to pure water, and the others are nominal values that generally fall within the compositional ranges of hydrothermal fluids and seawater [52]. The reference chemical activities of proteins are taken to be 10^{-3}, which is a nominal value that is similar to experimental concentrations used in protein unfolding studies [53].

Equations of state

The standard molal thermodynamic properties of aqueous species as a function of temperature and pressure

can be evaluated using the revised Helgeson-Kirkham-Flowers (HKF) equations of state [30-33, 54, 55]. The temperature dependence of the standard molal thermodynamic properties of crystalline, gaseous and liquid species other than H_2O are calculated using a standard equation for heat capacity [34, 35, 56]. For the basis species other than H^+ and e^-, values of the standard molal thermodynamic properties and of the equations of state parameters were taken from Refs. [55, 57] ($CO_{2(aq)}$, $NH_{3(aq)}$ and $H_2S_{(aq)}$) and [58, 59] ($O_{2(g)}$). The equations of state adopted for liquid H_2O in the present study are those used in the SUPCRT92 software package [24].

Group additivity algorithms for ionized proteins
The standard molal properties and revised HKF equations of state parameters of ionized proteins are calculated in the present study using group additivity algorithms and data taken from Ref. [18] and outlined briefly below. The standard molal Gibbs energy of the jth unfolded protein with net charge denoted by Z_j ($\Delta G^\circ_{UP_j^{Z_j}}$) can be written as

$$\Delta G^\circ_{UP_j^{Z_j}} = \Delta G^\circ_{UP_j^0} + \Delta G^\circ_{\text{ion},j}, \tag{27}$$

where $\Delta G^\circ_{UP_j^0}$ stands for the standard molal Gibbs energy of the completely neutral (nonionized) unfolded protein and $\Delta G^\circ_{\text{ion},j}$ stands for the contribution of ionization of sidechain and terminal groups to the standard molal Gibbs energy of the ionized protein. The latter term can be calculated by first writing

$$\Delta G^\circ_{\text{ion},j} = \sum_i n_{i,j} \alpha_i \Delta G^\circ_{\text{ion},i}, \tag{28}$$

where, for the ith type of ionizable sidechain or backbone group, $n_{i,j}$ represents the number of moles of the group in one mole of protein, α_i denotes the degree of ionization of the group ($0 < \alpha_i < 1$), and $\Delta G^\circ_{\text{ion},i}$ corresponds to the standard molal Gibbs energy of ionization of the group. Values of α_i and $\Delta G^\circ_{\text{ion},j}$ in Eqn. (M10) were taken in a simple approximation to be equal for all occurrences of a given ionizable group. It may be possible to refine this approach in the future by taking account of interactions of charged residues on the protein surfaces (Ref. [60] and others since).

Although and $\Delta G^\circ_{\text{ion},j}$ in Eqns. (M9) and (M10) are functions only of temperature and pressure for any protein in a defined charge state, α_i in Eqn. (M10) is a function of temperature, pressure, and solution pH [18]. Hence, $\Delta G^\circ_{\text{ion},j}$ and can be effectively computed for a given protein in different ionization states as a function of pH as well as of temperature and pressure. The net charge of the jth protein (Z_j) as a function of temperature, pressure and pH can be calculated using

$$Z_j = \sum_i n_{i,j} \alpha_i Z_i, \tag{29}$$

where Z_i denotes the charge (+1 or -1) of the ith ionized group and α_i (also in Eqn. M10) is given by

$$\alpha_i = \frac{1}{1 + 10^{Z_i(\text{pH} - pK_i)}}, \tag{30}$$

where pK_i represents the negative logarithm of the equilibrium constant for the deprotonation reaction of the ith ionizable group.

For a protein composed of a single polypeptide chain, the values of $\Delta G^\circ_{UP_j^0}$ in Eqn. (M9) can be calculated from the group additivity algorithm represented by [17, 18]

$$\Delta G^\circ_{UP_j^0} = \Delta G^\circ_{[\text{AABB}]} + (n_j - 1)\Delta G^\circ_{[\text{UPBB}]} + \sum_i^{20} n_{[SC]_i} \Delta G^\circ_{[SC]_i}, \tag{31}$$

where $\Delta G^\circ_{[\text{AABB}]}$, $\Delta G^\circ_{[\text{UPBB}]}$ and $\Delta G^\circ_{[SC]_i}$ denote the standard molal Gibbs energies of the amino acid backbone group, unfolded protein backbone group, and the ith type of amino acid sidechain group, respectively, $n_{[SC]_i}$ stands for the number of moles of the ith type of amino acid sidechain group in one mole of the protein, and

$$n_j = \sum_{i=1}^{20} n_{[SC]_i} \tag{32}$$

represents the total number of amino acid residues, or length of the protein. Values of $n_{[SC]_i}$ for the model proteins considered in the present study were retrieved from the Swiss-Prot/UniProt protein sequence database [37] (see Table 1).

The thermodynamic properties of unfolded aqueous proteins calculated using the above equations are taken in a first approximation to be representative of the proteins of interest, which may be folded and/or present in crystalline form in cells. Two observations lend support to the applicability of the unfolded protein reference state for the present calculations: 1) The standard molal Gibbs energies of protein folding would tend to cancel each other in metastability reactions, in which proteins appear on both sides of the reaction. 2) The Gibbs energy of unfolding for a small to average-sized protein is about two or three orders of magnitude smaller than the standard molal Gibbs energy for the unfolded protein itself. For

example, the Gibbs energy of unfolding of chicken lysozyme is ~14.5 kcal mol^{-1} at 25°C [61], but the standard molal Gibbs energy of this protein at 25°C and 1 bar is ~-4.2 × 10^3 kcal mol^{-1} (see Figs. 2a and 2b). The size of the unfolding property in this case is much smaller than the *ca.* ± 5% uncertainty ascribed to the group additivity algorithm [18]. It should be noted, however, that the compositional consequences of protein folding include changes in ionization state, and preferential surface exposure of charged residues [1], which would be manifested by changes in the reaction coefficients of basis species that might affect the outcome of metastability calculations to a greater extent than the differences in Gibbs free energy alone.

Competing interests

The authors declare that they have no competing interests.

Additional material

Additional file 1

Program script for generating figures. This text file contains the program script used to make the diagrams in Figs. 3, 4, 5, 6, 7. Use the commands listed at the top of the file to generate one or all of the figures on screen or in postscript format.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1467-4866-9-10-S1.txt]

Additional file 2

Degree of formation diagram. This file contains the degree of formation diagram related to Fig. 5b (see text) together with the program script used to make the figure. This additional material is the source of the graphical abstract for this paper.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1467-4866-9-10-S2.pdf]

References

1. Fukuchi S and Nishikawa K: **Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria.** *J Mol Biol* 2001, **309(4):**835–843.
2. Jaenicke R: **Protein stability and molecular adaptation to extreme conditions.** *Eur J Biochem* 1991, **202(3):**715–728.
3. Sanchez-Ruiz JM and Makhatadze GI: **To charge or not to charge?.** *Trends Biotechnol* 2001, **19(4):**132–135.
4. Mazel D and Marlière P: **Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins.** *Nature* 1989, **341(6239):**245–248.
5. Zen E-A: **Components, phases, and criteria of chemical equilibrium in rocks.** *Am J Sci* 1963, **261(10):**929–942.
6. Helgeson HC: **Evaluation of irreversible reactions in geochemical processes involving minerals and aqueous solutions. I. Thermodynamic relations.** *Geochim Cosmochim Acta* 1968, **32(8):**853–877.
7. Shock EL: **Organic acid metastability in sedimentary basins.** *Geology* 1988, **16(10):**886–890.
8. Shock EL: **Geochemical constraints on the origin of organic compounds in hydrothermal systems.** *Orig Life Evol Biosph* 1990, **20(3–4):**331–367.
9. Shock EL and Schulte MD: **Amino acid synthesis in carbonaceous meteorites by aqueous alteration of polycyclic aromatic hydrocarbons.** *Nature* 1990, **343(6260):**728–731.
10. Helgeson HC, Knox AM, Owens CE and Shock EL: **Petroleum, oil-field waters, and authigenic mineral assemblages: Are they in metastable equilibrium in hydrocarbon reservoirs?.** *Geochim Cosmochim Acta* 1993, **57(14):**3295–3339.
11. Helgeson HC and Amend JP: **Relative stabilities of biomolecules at high temperatures and pressures.** *Thermochim Acta* 1994, **245:**89–119.
12. Seewald JS: **Evidence for metastable equilibrium between hydrocarbons under hydrothermal conditions.** *Nature* 1994, **370(6487):**285–287.
13. Andersson E and Holm NG: **The stability of some selected amino acids under attempted redox constrained hydrothermal conditions.** *Orig Life Evol Biosph* 2000, **30:**9–23.
14. McCollom TM and Seewald JS: **Abiotic synthesis of organic compounds in deep-sea hydrothermal environments.** *Chem Rev* 2007, **107(2):**382–401.
15. Sleytr UB, Messner P, Pum D, Sára M and Eds: *Crystalline Bacterial Cell Surface Proteins* Austin, Texas: R. G. Landes Company; 1996.
16. Sára M and Sleytr UB: **S-layer proteins.** *J Bacteriol* 2000, **182 (4):**859–868.
17. Amend JP and Helgeson HC: **Calculation of the standard molal thermodynamic properties of aqueous biomolecules at elevated temperatures and pressures. II. Unfolded proteins.** *Biophys Chem* 2000, **84(2):**105–136.
18. Dick JM, LaRowe DE and Helgeson HC: **Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins.** *Biogeosciences* 2006, **3(3):**311–336.
19. Thompson JB: **The thermodynamic basis for the mineral facies concept.** *Am J Sci* 1955, **253(2):**65–103.
20. Korzhinskii DS: *Physicochemical Basis of the Analysis of the Paragenesis of Minerals* New York: Consultants Bureau, Inc; 1959.
21. Korzhinskii DS: **Thermodynamic potentials of open systems whose acidity and redox potential are determined by external conditions.** *Dokl Acad Sci USSR* 1963, **152(2):**175–177, [A.G.I. Translation]..
22. Korzhinskii DS: **The theory of systems with perfectly mobile components and processes of mineral formation.** *Am J Sci* 1965, **263(3):**193–205.
23. R Development Core Team: *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria; 2008 http://www.R-project.org.
24. Johnson JW, Oelkers EH and Helgeson HC: **SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C.** *Comp Geosci* 1992, **18(7):**899–947.
25. Levelt-Sengers JMH, Kamgarparsi B, Balfour FW and Sengers JV: **Thermodynamic properties of steam in the critical region.** *J Phys Chem Ref Data* 1983, **12:**1–28.
26. Haar L, Gallagher JS and Kell GS: *NBS/NRC Steam Tables* Washington, D. C.: Hemisphere; 1984.
27. Johnson JW and Norton D: **Critical phenomena in hydrothermal systems: state, thermodynamic, electrostatic, and transport properties of H_2O in the critical region.** *Am J Sci* 1991, **291(6):**541–648.
28. Wagner W and Pruss A: **The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use.** *J Phys Chem Ref Data* 2002, **31 (2):**387–535.
29. Archer DG and Wang PM: **The dielectric constant of water and Debye-Hückel limiting law slopes.** *J Phys Chem Ref Data* 1990, **19 (2):**371–411.
30. Helgeson HC, Kirkham DH and Flowers GC: **Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures. IV. Calculation of activity coefficients, osmotic coefficients,**

and apparent molal and standard and relative partial molal properties to 600°C and 5 Kb. *Am J Sci* 1981, **281** (10):1249–1516.

31. Tanger JC IV and Helgeson HC: **Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Revised equations of state for the standard partial molal properties of ions and electrolytes.** *Am J Sci* 1988, **288**:19–98.

32. Shock EL and Helgeson HC: **Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of organic species.** *Geochim Cosmochim Acta* 1990, **54** (4):915–945.

33. Shock EL, Oelkers EH, Johnson JW, Sverjensky DA and Helgeson HC: **Calculation of the thermodynamic properties of aqueous species at high pressures and temperatures: Effective electrostatic radii, dissociation constants, and standard partial molal properties to 1000°C and 5 kbar.** *J Chem Soc Faraday Trans* 1992, **88**(6):803–826.

34. Robie RA and Hemingway BS: *Thermodynamic Properties of Minerals and Related Substances at 298.15 K and 1 Bar (10.5Pascals) Pressure and at Higher Temperatures* Bulletin 2131, U. S. Geol. Surv; 1995.

35. Maier CG and Kelley KK: **An equation for the representation of high-temperature heat content data.** *J Am Chem Soc* 1932, **54** (8):3243–3246.

36. Richard L and Helgeson HC: **Calculation of the thermodynamic properties at elevated temperatures and pressures of saturated and aromatic high molecular weight solid and liquid hydrocarbons in kerogen, bitumen, petroleum, and other organic matter of biogeochemical interest.** *Geochim Cosmochim Acta* 1998, **62**(23–24):3591–3636.

37. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S and Schneider M: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.** *Nucleic Acids Res* 2003, **31**:365–370.

38. Amend JP and Plyasunov AV: **Carbohydrates in thermophile metabolism: Calculation of the standard molal thermodynamic properties of aqueous pentoses and hexoses at elevated temperatures and pressures.** *Geochim Cosmochim Acta* 2001, **65**(21):3901–3917.

39. LaRowe DE and Helgeson HC: **Biomolecules in hydrothermal systems: Calculation of the standard molal thermodynamic properties of nucleic-acid bases, nucleosides, and nucleotides at elevated temperatures and pressures.** *Geochim Cosmochim Acta* 2006, **70**(18):4680–4724.

40. LaRowe DE and Helgeson HC: **The energetics of metabolism in hydrothermal systems: Calculation of the standard molal thermodynamic properties of magnesium-complexed adenosine nucleotides and NAD and NADP at elevated temperatures and pressures.** *Thermochim Acta* 2006, **448** (2):82–106.

41. Nakamura S, Aono R, Mizutani S, Takashina T, Grant WD and Horikoshi K: **The cell-surface glycoprotein of *Haloarcula japonica* TR-1.** *Biosci Biotechnol Biochem* 1992, **56**(6):996–998.

42. Whitman WB, Ankwanda E and Wolfe RS: **Nutrition and carbon metabolism of *Methanococcus voltae*.** *J Bacteriol* 1982, **149** (3):852–863.

43. Jones WJ, Leigh JA, Mayer F, Woese CR and Wolfe RS: ***Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent.** *Arch Microbiol* 1983, **136**(4):254–261.

44. Hoehler TM, Alperin MJ, Albert DB and Martens CS: **Thermodynamic control on hydrogen concentrations in anoxic sediments.** *Geochim Cosmochim Acta* 1998, **62**(10):1745–1756.

45. McCollom TM and Shock EL: **Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems.** *Geochim Cosmochim Acta* 1997, **61**(20):4375–4391.

46. Waddington CH: *The Evolution of an Evolutionist* Edinburgh: Edinburgh University Press; 1975.

47. Lewis GN and Randall M: *Thermodynamics* New York: McGraw-Hill; 1961, [Rev. by K. S. Pitzer and L. Brewer]..

48. Drever JI: *The Geochemistry of Natural Waters* Upper Saddle River, New Jersey: Prentice Hall; 31997.

49. Lichtner PC: **Continuum formulation of multicomponent-multiphase reactive transport.** *Reactive Transport in Porous Media, Reviews in Mineralogy* Washington, D. C.: Mineralogical Society of America: Lichtner PC, Steefel CI, Oelkers EH 1996, **34**:1–81.

50. Helgeson HC: **Solution chemistry and metamorphism.** *Researches in Geochemistry* New York: Wiley: Abelson PH 1967, **2**:362–404.

51. Prigogine I and Defay R: *Chemical Thermodynamics* London: Longmans, Green and Co; 1954.

52. Amend JP and Shock EL: **Energetics of amino acid synthesis in hydrothermal ecosystems.** *Science* 1998, **281**(5383):1659–1662.

53. Pfeil W and Privalov PL: **Thermodynamic investigations of proteins. 1. Standard functions for proteins with lysozyme as an example.** *Biophys Chem* 1976, **4**:23–32.

54. Shock EL and Helgeson HC: **Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C.** *Geochim Cosmochim Acta* 1988, **52**(8):2009–2036.

55. Shock EL, Helgeson HC and Sverjensky DA: **Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species.** *Geochim Cosmochim Acta* 1989, **53**(9):2157–2183.

56. Helgeson HC, Owens CE, Knox AM and Richard L: **Calculation of the standard molal thermodynamic properties of crystalline, liquid, and gas organic molecules at high temperatures and pressures.** *Geochim Cosmochim Acta* 1998, **62**(6):985–1081.

57. Schulte MD, Shock EL and Wood RH: **The temperature dependence of the standard-state thermodynamic properties of aqueous nonelectrolytes.** *Geochim Cosmochim Acta* 2001, **65**(21):3919–3930.

58. Kelley KK: *Contributions to the Data in Theoretical Metallurgy XIII: High Temperature Heat Content, Heat Capacities and Entropy Data for the Elements and Inorganic Compounds* Bulletin 584, U. S. Bureau of Mines; 1960.

59. Wagman DD, Evans WH, Parker VB, Schumm RH, Halow I, Bailey SM, Churney KL and Nuttall RL: **The NBS tables of chemical thermodynamic properties. Selected values for inorganic and C_1 and C_2 organic substances in SI units.** *J Phys Chem Ref Data* 1982, **11**:1–392.

60. Linderstrøm-Lang KU: **On the ionization of proteins.** *CR Trav Lab Carlsberg* 1924, **15**(7):1–29.

61. Pfeil W and Privalov PL: **Thermodynamic investigations of proteins. 3. Thermodynamic description of lysozyme.** *Biophys Chem* 1976, **4**:41–50.

62. Lauerer G, Kristjansson JK, Langworthy TA, König H and Stetter KO: ***Methanothermus sociabilis* sp. nov., a second species within the *Methanothermaceae* growing at 97°C.** *Syst Appl Microbiol* 1986, **8**(1–2):100–105.

63. Stetter KO, Thomm M, Winter J, Wildgruber G, Huber H, Zillig W, Janécovic D, König H, Palm P and Wunderl S: ***Methanothermus fervidus*, sp. nov., a novel extremely thermophilic methanogen isolated from an Icelandic hot spring.** *Zentralbl Bakteriol Hyg Abt I Orig C* 1981, **2**(2):166–178.

64. Horikoshi K, Aono R and Nakamura S: **The triangular halophilic archaebacterium *Haloarcula japonica* strain TR-1.** *Experientia* 1993, **49**(6–7):497–502.

65. Franzmann PD, Springer N, Ludwig W, Demacario EC and Rohde M: **A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov.** *Syst Appl Microbiol* 1992, **15** (4):573–581.

66. Leigh JA and Wolfe RS: ***Acetogenium kivui* gen. nov., sp. nov., a thermophilic acetogenic bacterium.** *Int J Syst Bacteriol* 1983, **33** (4):886–886.

67. Wilson KS, Vorgias CE, Tanaka I, White SW and Kimura M: **The thermostability of DNA-binding protein HU from bacilli.** *Protein Eng* 1990, **4**:11–22.

68. Warth AD: **Relationship between the heat resistance of spores and the optimum and maximum growth temperatures of *Bacillus* species.** *J Bacteriol* 1978, **134**(3):699–705.

69. Guérin-Faublée V, Charles S, Chomarat M and Flandrois J-P: **Reappraisal of the effect of temperature on the growth kinetics of *Aeromonas salmonicida*.** *Lett Appl Microbiol* 1997, **25** (5):363–366.