

Hypothesis

Open Access

## Calculation of the relative metastabilities of proteins in subcellular compartments of *Saccharomyces cerevisiae*

Jeffrey M Dick<sup>1,2</sup>

Address: <sup>1</sup>Department of Earth and Planetary Science, University of California, 307 McCone Hall, #4767 Berkeley, CA 94720-4767, USA and <sup>2</sup>Current address : School of Earth and Space Exploration, Arizona State University, P.O. Box 871404, Tempe, AZ 85287-1404, USA

Email: Jeffrey M Dick - jedick@berkeley.edu

Published: 18 July 2009

Received: 31 January 2009

BMC Systems Biology 2009, 3:75 doi:10.1186/1752-0509-3-75

Accepted: 18 July 2009

This article is available from: <http://www.biomedcentral.com/1752-0509/3/75>

© 2009 Dick; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Protein subcellular localization and differences in oxidation state between subcellular compartments are two well-studied features of the cellular organization of *S. cerevisiae* (yeast). Theories about the origin of subcellular organization are assisted by computational models that can integrate data from observations of compositional and chemical properties of the system.

**Presentation and implications of the hypothesis:** I adopt the hypothesis that the state of yeast subcellular organization is in a local energy minimum. This hypothesis implies that equilibrium thermodynamic models can yield predictions about the interdependence between populations of proteins and their subcellular chemical environments.

**Testing the hypothesis:** Three types of tests are proposed. First, there should be correlations between modeled and observed oxidation states for different compartments. Second, there should be a correspondence between the energy requirements of protein formation and the order the appearance of organelles during cellular development. Third, there should be correlations between the predicted and observed relative abundances of interacting proteins within compartments.

**Results:** The relative metastability fields of subcellular homologs of glutaredoxin and thioredoxin indicate a trend from less to more oxidizing as mitochondrion – cytoplasm – nucleus. Representing the overall amino acid compositions of proteins in 23 different compartments each with a single reference model protein suggests that the formation reactions for proteins in the vacuole (in relatively oxidizing conditions), ER and early Golgi (in relatively reducing conditions) are relatively highly favored, while that for the microtubule is the most costly. The relative abundances of model proteins for each compartment inferred from experimental data were found in some cases to correlate with the predicted abundances, and both positive and negative correlations were found for some assemblages of proteins in known complexes.

**Conclusion:** The results of these calculations and tests suggest that a tendency toward a metastable energy minimum could underlie some organizational links between the chemical thermodynamic properties of proteins and subcellular chemical environments. Future models of this kind will benefit from consideration of additional thermodynamic variables together with more detailed subcellular observations.

## Background

A complex interplay of chemical and biological forces is responsible for subcellular structure. There exist in eukaryotic cells gradients between subcellular compartments of chemical properties such as pH, oxidation-reduction (or redox) state and chemical activity of water, among others [1-5]. Different population of proteins are localized within each subcellular compartment [6-8]. Within compartments, the relative abundances or levels of different proteins are not equal [9], and different proteins predominate in the various subcellular populations depending on growth state of the cell and exposure to environmental stress [10]. Physical separation of key enzymes is thought to be essential in the cytoskeletal network and in regulation of metabolic pathways and other cellular functions [11,12]. The patterns of subcellular structure persist even though populations of proteins turnover through continual degradation and synthesis in cells [13].

The biosynthesis and transport of proteins in an energy-demanding process [14]. If cells have evolved to minimize their energy expenditure in the maintenance of biological function, it may be reasonable to expect to find signals of energy minimization in cellular organization. One such example is the finding that the relative abundances of amino acids in proteins correlate inversely with the metabolic cost of amino acid synthesis [15,16], and that this is a temperature-dependent function [17]. This observation is consistent with the notion that not all proteins are equal in energetic terms. In thermodynamic calculations of chemical affinity [18], the energy demands of protein formation (including synthesis and transport) are also a function of the local physical chemical environment, which includes variables such as oxidation-reduction potential [19]. It follows that subcellular structures that are characterized by differences in the amino acid composition of proteins and in chemical potentials have distinct energetic consequences.

For the purposes of this study, the hypothesis is made that cellular organization is in a local energy minimum. Energy minimization in biological operations is not a new hypothesis, especially in the context of fitness and adaptation to the environment [20-22]. However, the implications of this hypothesis for subcellular organization have not been investigated from the standpoint of equilibrium chemical thermodynamics. Algorithms for computing the requisite standard molal Gibbs energies of proteins [23] and the relative abundances of proteins in metastable equilibrium [19] have recently been reported. The goal of this study is to perform these types of calculations for model systems representative of various levels of subcellular organization and to compare the results with observations and measurements reported in the literature. If successful, this exercise may lead to an enhanced aware-

ness about the chemical forces that shape cellular structure.

The theoretical approach adopted here is based on the description of a chemical system in terms of intensive variables. These variables include temperature, pressure and the chemical potentials of the system. It is convenient to denote the chemical potentials by the chemical activities or fugacities of basis species, for example the activity of  $H^+$  (which defines pH) or the fugacity of oxygen. This permits comparison of the parameters of the model with reference systems described in experimental and other theoretical biochemical studies. In the following calculations, temperature and pressure were set to 25°C and 1 bar, respectively, and the logarithm of oxygen fugacity is the primary variable of interest. Below, oxidation-reduction potential and oxygen fugacity are used synonymously, and redox refers specifically to Eh. The oxidation-reduction potential of a system can be expressed in terms of Eh using an equation given in the Methods.

*S. cerevisiae* was chosen as a model system for the current investigation because there is abundant information about the subcellular distribution of proteins as well as independent measurements of the pH and oxidation state of some subcellular compartments. Also, the cellular development of yeast is extensively documented, which can yield other comparisons for some of the results of the model calculations.

There are two major parts to this paper. In the first part, the reactions corresponding to intercompartmental interactions between subcellular homologs (or isoforms) of particular enzymes and between reference model proteins for different compartments are quantified by calculating the oxygen fugacities for equal chemical activities of the reacting proteins in metastable equilibrium. The relative metastabilities of the reference model proteins are compared with some observations from the literature about reaction progress during the cell cycle. Specific known interactions between compartments are considered in order to derive values of the oxygen fugacity within compartments that best metastabilize the proteins contained within them. In the second part of this paper, the relative abundances of model proteins in metastable equilibrium are calculated and compared with measured abundances. The range of protein abundances in a metastable equilibrium population often approaches that seen in experiments over a narrow window of oxygen fugacity. Positive and negative correlations between the calculated and experimental relative abundances are found in some cases. The paper concludes with a summary of the findings and other implications of the hypothesis.

### Presentation of the hypothesis

Part of a cell's expenditure of metabolic fuel is directed toward the formation of proteins, including their synthesis and transport to other compartments. Even when it is normalized to the lengths of the proteins, the energy required for protein formation is not a constant, but depends on the composition and environment of the protein. If these energy differences are quantified, the relative abundances of model proteins in metastable equilibrium can be calculated. The compositions of these metastable assemblages depend on local environmental variables such as oxygen fugacity, which is a scale for oxidation-reduction potential in a system. The major hypothesis adopted for this investigation is that energy minimization is a force contributing to the organization of cells; this implies the possibility of an evolutionary convergence between biomolecular composition and the chemical properties of subcellular compartments. In a first set of tests of this hypothesis, chemical reactions among model proteins in known intercompartmental interactions were used to obtain values of oxygen fugacity for subcellular compartments that can be compared with measured redox values. A second set of calculations presented here shows that the relative abundances of proteins within compartments and of those that form complexes can be correlated in some cases with metastable equilibrium assemblages.

These results provide theoretical constraints on the spontaneous generation of order in the distributions of proteins within cells and imply that work done by maintaining oxidation-reduction gradients can selectively alter the degrees of formation of assemblages of proteins.

### Testing the hypothesis

#### Relative metastabilities of subcellular homologs of redoxins

Yeast cells have cytoplasmic, nuclear and mitochondrial homologs of glutaredoxin [24-26] and cytoplasmic and mitochondrial homologs of thioredoxin and thioredoxin reductase [27,28]. The names and chemical formulas of these proteins are listed in Table 1, together with some computed properties. The average nominal oxidation state of carbon ( $\bar{Z}_C$ ) is a function of the relative proportions of the elements in the chemical formula (see Methods). In Table 1 the proteins with the lowest values of  $\bar{Z}_C$  are the mitochondrial homologs and those with the highest values of  $\bar{Z}_C$  are the nuclear homologs. Accordingly, the formation of the mitochondrial and nuclear proteins are energetically favored by relatively reducing and oxidizing conditions, respectively.

**Table 1: Subcellular isoforms of glutaredoxin, thioredoxin and thioredoxin reductase in yeast<sup>a</sup>.**

Protein	SWISS-PROT	Location	Length	Formula	$\Delta G_f^\circ$	Z	$\bar{Z}_C$
Glutaredoxin							
GLRX1	<a href="#">P25373</a>	Cytoplasm	110	$C_{549}H_{886}N_{146}O_{170}S_4$	-4565	-5.8	-0.182
GLRX2	<a href="#">P17695</a>	Mitochondrion	143	$C_{715}H_{1161}N_{181}O_{213}S_5$	-5617	0.1	-0.255
GLRX3	<a href="#">Q03835</a>	Nucleus	285	$C_{1444}H_{2195}N_{371}O_{463}S_{10}$	-12031	-24.5	-0.094
GLRX4	<a href="#">P32642</a>	Nucleus	244	$C_{1226}H_{1910}N_{316}O_{389}S_6$	-10276	-17.8	-0.140
GLRX5	<a href="#">Q02784</a>	Mitochondrion	150	$C_{762}H_{1200}N_{196}O_{227}S_6$	-5841	-6.1	-0.192
Thioredoxin							
TRX1	<a href="#">P22217</a>	Cytoplasm	102	$C_{502}H_{785}N_{123}O_{150}S_5$	-3969	-3.1	-0.211
TRX2	<a href="#">P22803</a>	Cytoplasm	103	$C_{497}H_{780}N_{122}O_{153}S_5$	-4056	-3.1	-0.197
TRXB1	<a href="#">P29509</a>	Cytoplasm	318	$C_{1509}H_{2412}N_{402}O_{471}S_{12}$	-12330	-4.7	-0.159
TRX3	<a href="#">P25372</a>	Mitochondrion	127	$C_{651}H_{1049}N_{167}O_{181}S_{10}$	-4617	4.9	-0.255
TRXB2	<a href="#">P38816</a>	Mitochondrion	342	$C_{1640}H_{2615}N_{449}O_{501}S_{14}$	-12841	-1.5	-0.145

a. Amino acid compositions of subcellular isoforms of glutaredoxin (GLRX), thioredoxin (TRX) and thioredoxin reductase (TRXB) in *S. cerevisiae* were taken from the SWISS-PROT database [64] (accession numbers shown in the table). Chemical formulas of nonionized proteins, and calculated standard molal Gibbs energy of formation from the elements ( $\Delta G_f^\circ$ , in kcal mol<sup>-1</sup> at 25°C and 1 bar) and net ionization state (Z) at pH = 7 of charged proteins are listed. Average nominal oxidation state of carbon ( $\bar{Z}_C$ ) was calculated using Eqn. (6).

To quantify the metastability limits of the proteins in terms of the chemical environment, one can assess the energetics of formation reactions for the proteins. At metastable equilibrium, the predominant protein in a population is the one with the highest chemical activity (in this communication, activity refers to chemical activity rather than enzymatic activity; activity is equivalent to concentration for ideal systems, where activity coefficients are unity). This statement implies that the overall formation reaction from basis species (see Methods) for the predominant protein has a lower Gibbs energy (or higher chemical affinity) than any of the others. The consequences of these relationships can be portrayed on chemical activity diagrams using a previously described procedure that is encoded in the CHNOSZ software package [19], which was used to perform the calculations reported below (see Methods). Additional File 1 includes the program script and data files that were used to carry out these calculations and generate the tables and figures. So that the results described in this section can be reconstructed at pH = 7, the standard molal Gibbs energies ( $\Delta G_f^\circ$ ) and net charges of ionized proteins at this pH are listed in Table 1.

In Figs. 1a and 1b the metastable equilibrium predominance limits of ionized proteins in the glutaredoxin and thioredoxin/thioredoxin reductase model systems are shown as a function of the logarithm of oxygen fugacity and pH. The computation of the relative metastabilities of the proteins included all five model proteins in the glutaredoxin system as candidates, but note in Fig. 1a that only two of the five proteins appear on the diagram. Those that do not appear are less metastable, or have greater energy requirements for their formation over the range of conditions represented in Fig. 1a than either of the proteins appearing in the figure.

The equal-activity lines in these pH diagrams are curved because the ionization states of the proteins depend on pH. The observation apparent in Fig. 1a that increasing  $\log f_{O_2(g)}$  favors formation of the cytoplasmic protein homolog relative to its mitochondrial counterpart is also true for the thioredoxin/thioredoxin reductase system shown in Fig. 1b. In comparing Figs. 1a and 1b note that in the latter figure, predominance fields for a greater number of candidate proteins appear, and that the predominance field boundary between mitochondrial and cytoplasmic proteins occurs at a lower oxidation-reduction potential. The dashed lines shown in each diagram of

Fig. 1 are reference lines denoting the reduction stability limit of H<sub>2</sub>O ( $\log f_{O_2(g)} \approx -83.1$  at 25°C and 1 bar [29]).

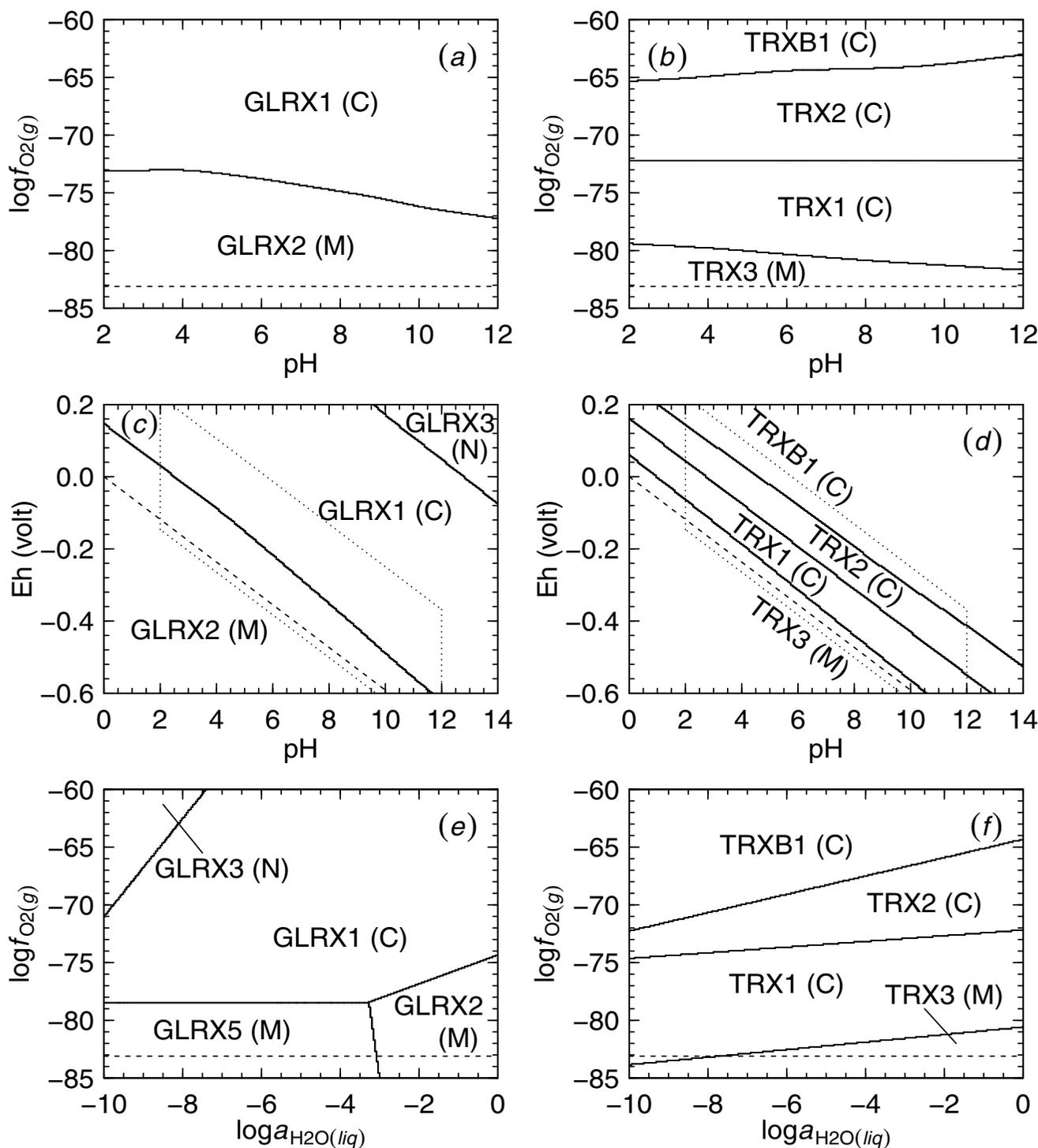
Predominance diagrams as a function of Eh and pH for the glutaredoxin and thioredoxin/thioredoxin reductase systems are shown in Figs. 1c and 1d. Like  $\log f_{O_2(g)}$ , Eh and pH together are a measure of the oxidation-reduction potential of the system; the different scales can be converted using Eqn. (5) in the Methods. The trapezoidal areas bounded by dotted lines in Figs. 1c and 1d show the ranges of Eh and pH corresponding to the limits of the  $\log f_{O_2(g)}$ -pH diagrams of Figs. 1a and 1b. It can be deduced from these diagrams that if the upper  $\log f_{O_2(g)}$  limit of Fig. 1a were extended upward, this diagram would include a portion of the predominance field for the nuclear protein GLRX3.

It appears from Figs. 1a–b that increasing increasing  $\log f_{O_2(g)}$  at constant pH, or increasing pH at constant oxidation-reduction potential have similar consequences for the relative metastabilities of the cytoplasmic and mitochondrial homologs. In this analysis, however, pH does not appear to be a very descriptive variable; the magnitude of the effect of changing oxygen fugacity over several log units is greater than the effect of changing pH by several units. In further calculations described below pH was set to 7.

In Figs. 1e and 1f the logarithm of activity of water ( $\log a_{H_2O}$ ) appears as a variable. In Fig. 1e it can be seen that the formation of a nuclear homolog of glutaredoxin is favored relative to the cytoplasmic homologs by decreasing activity of water and/or increasing oxygen fugacity, and that increasing relative metastabilities of the mitochondrial proteins are consistent with lower oxidation-reduction potentials. In Fig. 1f it appears that the formation of the thioredoxin reductase relative to thioredoxin is favored by increasing  $f_{O_2(g)}$ , and that for the thioredoxin the relative metastabilities of the mitochondrial proteins increase with decreasing  $f_{O_2(g)}$ .

#### Comparison with subcellular redox measurements

Let us compare the positions of the predominance fields in Fig. 1 with measured subcellular redox states. The values of Eh derived from the concentrations of oxidized and



**Figure 1**  
**Relative metastabilities of homologs of glutaredoxin and thioredoxin/thioredoxin reductase.** Predominance diagrams were generated for subcellular isoforms of (a, c, e) glutaredoxin (GLRX) and of (b, d, f) thioredoxin (TRX) and thioredoxin reductase (TRXB) in *S. cerevisiae*. The letters in parentheses following the labels indicate the subcellular compartment to which the protein is localized (C – cytoplasm; M – mitochondrion; N – nucleus). Calculations were performed for ionized proteins at 25°C and 1 bar and for reference activities of basis species noted in the Methods. Reduction stability limits of  $H_2O$  are shown by dashed lines; the dotted lines in (c) and (d) correspond to the plot limits of (a) and (b).

**Table 2: Nominal electrochemical characteristics of subcellular environments in eukaryotes. Values refer to yeast cells unless noted otherwise.**

Environment	Eh, volt	pH	$\log f_{O_2(g)}^m$
Extracellular (intestine)	-0.137 to -0.80 <sup>a</sup>	3 <sup>g</sup>	-83.3 to -79.4
Cytoplasm	-0.235 to -0.222 <sup>b</sup>	6.5 <sup>h</sup>	-75.9 to -75.0
Nucleus	- <sup>c</sup>	7.7 <sup>i</sup>	- <sup>c</sup>
Mitochondrion	-0.360 <sup>d</sup>	8 <sup>j</sup>	-78.3
Endoplasmic reticulum	-0.185 to -0.133 <sup>e</sup>	7.2 <sup>k</sup>	-69.7 to -66.2
Vacuole	> +0.769 <sup>f</sup>	6.2 <sup>l</sup>	> -9.2

a. [30] (*Homo sapiens*). b. The lower and upper values are taken from [32] and [31], respectively. c. The state of the GSSG/GSH couple in the nucleus is thought to be more reduced than in the cytoplasm [40] (see text). d. [33] (*Homo sapiens* HeLa [34] cells). e. [2] (*Mus musculus*: mouse hybridoma cells [35]). f. Calculated by combining the law of mass action for  $Fe^{+3} + e^- \rightleftharpoons Fe^{+2}$  using standard molal Gibbs energies taken from Ref. [65] with  $a_{Fe^{+3}} = a_{Fe^{+2}}$ . g. [36] (*Homo sapiens*). h. [37] (yeast). i. [66] (organism unspecified). j. [38] (HeLa). k. [67]. l. [1]. m. Values of Eh and pH listed here were combined with Eqn. (5) at  $T = 25^\circ C$ ,  $P = 1$  bar and  $a_{H_2O} = 1$  to generate the values of  $\log f_{O_2(g)}$ .

reduced glutathione (GSSG and GSH, respectively) [2,30-32] and fluorescent probes [33] in extra- and subcellular environments reported in various studies were converted to corresponding values of  $\log f_{O_2(g)}$  using Eqn. (5) in the Methods and are listed in Table 2. In order to fill in the table as completely as possible, it was necessary to consider measurements performed on eukaryotic cells other than *S. cerevisiae* (e.g., HeLa [34] and mouse hybridoma [35] cells). The values of pH required for conversion of Eh to  $\log f_{O_2(g)}$  were also retrieved from the literature [36-38]. The computation of  $\log f_{O_2(g)}$  from Eh was performed at  $25^\circ C$  and 1 bar and with  $\log a_{H_2O} = 0$ . No measurements of vacuolar Eh were found, but it has been noted that  $Fe^{+3}$  predominates over  $Fe^{+2}$  in this compartment [39]. Hence, a nominal (and relatively very oxidizing) value of Eh for the vacuole was calculated that corresponds to equal activities of  $Fe^{+3}$  and  $Fe^{+2}$ .

The current understanding of the major trends of redox states in compartments of eukaryotic cells can be summarized as, from most reducing to most oxidizing, mitochondrion – nucleus – cytoplasm – endoplasmic reticulum (ER) – extracellular [40]. Strong redox gradients within the mitochondrion are essential to its function [41], which is not captured by the single values listed in Table 2. Comparison nevertheless with the computational

results shown in Fig. 1 indicates that a relatively reducing environment does favor the mitochondrial homologs over the others shown in the diagram.

Measurements of GSH/GSSG concentrations point to a lower redox state in the nucleus than in the cytoplasm, but the present model has the nuclear proteins favored by relatively oxidizing conditions. Studies using nuclear magnetic resonance (NMR) showing that the hydration state of the nucleus is higher than the cytoplasm [42,3] also seem to contradict the trend in Fig. 1e that the formation of the nuclear proteins is favored relative to their cytoplasmic counterparts by decreasing activity of water. Finally, mitochondrial pH is somewhat higher than that of the cytoplasm [37,38], but in Figs. 1a and 1b it appears that the predicted energetic constraints favor the cytoplasmic proteins at higher pHs. These comparisons indicate that the investigated metastable equilibrium constraints are not entirely responsible for the spatial distribution of the isoforms of redoxins in the cell.

#### Relative metastabilities of reference model proteins

The reference model proteins used in this study represent the overall amino acid compositions of the proteins in individual compartments. The amino acid compositions of reference model proteins for 23 subcellular compartments were calculated as described in the Methods and are listed in Additional File 2; the chemical formulas and standard molal Gibbs energies are listed in Table 3. The predominance diagrams in Fig. 2 depicting the relative metastabilities of the reference model proteins as a function of  $\log f_{O_2(g)}$  and  $\log a_{H_2O}$  were generated in sequential order. The first diagram in this figure corresponds to a system in which all 23 reference model proteins were considered. Subsequent diagrams in Fig. 2 were generated by eliminating from consideration some or all of the reference model proteins represented by predominance fields in the immediately preceding diagram. It can be seen in Fig. 2a that consideration of 23 reference model proteins resulted in predicted predominance fields for four proteins over the ranges of  $\log f_{O_2(g)}$  and  $\log a_{H_2O}$  shown in the diagram. The reference model proteins appearing in successive diagrams in Fig. 2 are characterized by increasingly higher predicted energy requirements for their formation. Hence, the mitochondrial, nuclear and cytoplasmic reference model proteins appearing in Fig. 2b-d are relatively less metastable compared to those of early Golgi and ER appearing in Fig. 2a.

Can the relative metastabilities of proteins be linked to the order of their appearance in the cell cycle? It is note-

**Table 3: Chemical formulas, standard molal Gibbs energies and other properties for reference model proteins for compartments in yeast cells<sup>a</sup>.**

Location	Number	Length	Formula	$\Delta G_f^\circ$	Z	$\bar{Z}_C$	$\log f_{O_2(g)}$
actin	22	469.41	C <sub>2316</sub> H <sub>3636</sub> N <sub>632</sub> O <sub>721</sub> S <sub>10</sub>	-18506	-5.3	-0.119	-74.7
ambiguous	123	572.02	C <sub>2816</sub> H <sub>4465</sub> N <sub>759</sub> O <sub>870</sub> S <sub>17</sub>	-22616	-9.6	-0.146	NA
bud	57	462.56	C <sub>2366</sub> H <sub>3668</sub> N <sub>614</sub> O <sub>681</sub> S <sub>20</sub>	-16641	5.4	-0.179	-75.4
bud.neck	11	905.26	C <sub>4543</sub> H <sub>7203</sub> N <sub>1250</sub> O <sub>1443</sub> S <sub>26</sub>	-38103	-16.8	-0.113	-69.2
cell.periphery	38	826.07	C <sub>4178</sub> H <sub>6505</sub> N <sub>1098</sub> O <sub>1229</sub> S <sub>33</sub>	-30641	1.7	-0.164	-79.4
cytoplasm	746	436.12	C <sub>2164</sub> H <sub>3440</sub> N <sub>590</sub> O <sub>659</sub> S <sub>12</sub>	-17065	-2.7	-0.149	-73.5
early.Golgi	9	622.83	C <sub>3198</sub> H <sub>5068</sub> N <sub>821</sub> O <sub>972</sub> S <sub>21</sub>	-25441	-19.7	-0.193	-78.0
endosome	30	484.06	C <sub>2441</sub> H <sub>3871</sub> N <sub>661</sub> O <sub>767</sub> S <sub>14</sub>	-20265	-12.1	-0.133	-75.8
ER	197	245.57	C <sub>1206</sub> H <sub>1897</sub> N <sub>314</sub> O <sub>365</sub> S <sub>8</sub>	-9276	-3.0	-0.173	-77.8
ER.to.Golgi	5	595.02	C <sub>2951</sub> H <sub>4601</sub> N <sub>790</sub> O <sub>907</sub> S <sub>18</sub>	-22861	-13.3	-0.127	NA
Golgi	14	481.43	C <sub>2502</sub> H <sub>3880</sub> N <sub>651</sub> O <sub>740</sub> S <sub>13</sub>	-18516	-5.4	-0.167	-77.6
late.Golgi	29	787.94	C <sub>4015</sub> H <sub>6312</sub> N <sub>1044</sub> O <sub>1217</sub> S <sub>22</sub>	-31255	-24.7	-0.174	-77.4
lipid.particle	17	501.83	C <sub>2573</sub> H <sub>3985</sub> N <sub>672</sub> O <sub>751</sub> S <sub>17</sub>	-18687	-4.1	-0.167	-75.0
microtubule	10	497.13	C <sub>2509</sub> H <sub>3969</sub> N <sub>690</sub> O <sub>774</sub> S <sub>17</sub>	-20031	-4.0	-0.125	-75.0
mitochondrion	426	402.95	C <sub>1987</sub> H <sub>3166</sub> N <sub>542</sub> O <sub>596</sub> S <sub>13</sub>	-15221	3.3	-0.160	-75.9
nuclear.periphery	46	815.58	C <sub>4111</sub> H <sub>6516</sub> N <sub>1092</sub> O <sub>1272</sub> S <sub>20</sub>	-33146	-11.5	-0.159	-77.0
nucleolus	60	605.43	C <sub>2990</sub> H <sub>4768</sub> N <sub>820</sub> O <sub>957</sub> S <sub>14</sub>	-25444	-10.4	-0.121	-75.6
nucleus	453	339.30	C <sub>1683</sub> H <sub>2686</sub> N <sub>472</sub> O <sub>517</sub> S <sub>8</sub>	-13342	3.4	-0.129	-75.0
peroxisome	18	422.30	C <sub>2117</sub> H <sub>3334</sub> N <sub>568</sub> O <sub>641</sub> S <sub>13</sub>	-16400	-2.0	-0.150	-74.8
punctate.composite	61	467.64	C <sub>2320</sub> H <sub>3662</sub> N <sub>633</sub> O <sub>751</sub> S <sub>10</sub>	-19985	-22.3	-0.102	NA
spindle.pole	30	398.53	C <sub>1996</sub> H <sub>3176</sub> N <sub>555</sub> O <sub>642</sub> S <sub>12</sub>	-17251	-13.5	-0.100	-79.5
vacuolar.membrane	45	709.78	C <sub>3532</sub> H <sub>5555</sub> N <sub>943</sub> O <sub>1075</sub> S <sub>23</sub>	-27439	-15.3	-0.150	-73.4
vacuole	67	428.85	C <sub>2078</sub> H <sub>3186</sub> N <sub>542</sub> O <sub>668</sub> S <sub>14</sub>	-17065	-18.1	-0.093	-73.2

a. Chemical formulas of nonionized reference model proteins and standard molal Gibbs energy of formation from the elements ( $\Delta G_f^\circ$  in kcal mol<sup>-1</sup>, at 25°C and 1 bar) and net ionization state (Z) at pH = 7 of ionized reference model proteins were calculated using the amino acid compositions given in Additional File 2. Values of the nominal oxidation state of carbon ( $\bar{Z}_C$ ) were calculated using Eqn. (6). The model  $\log f_{O_2(g)}$  values for the compartments were obtained from the metastability limits of subcellular interactions listed in Table 4 (see text).

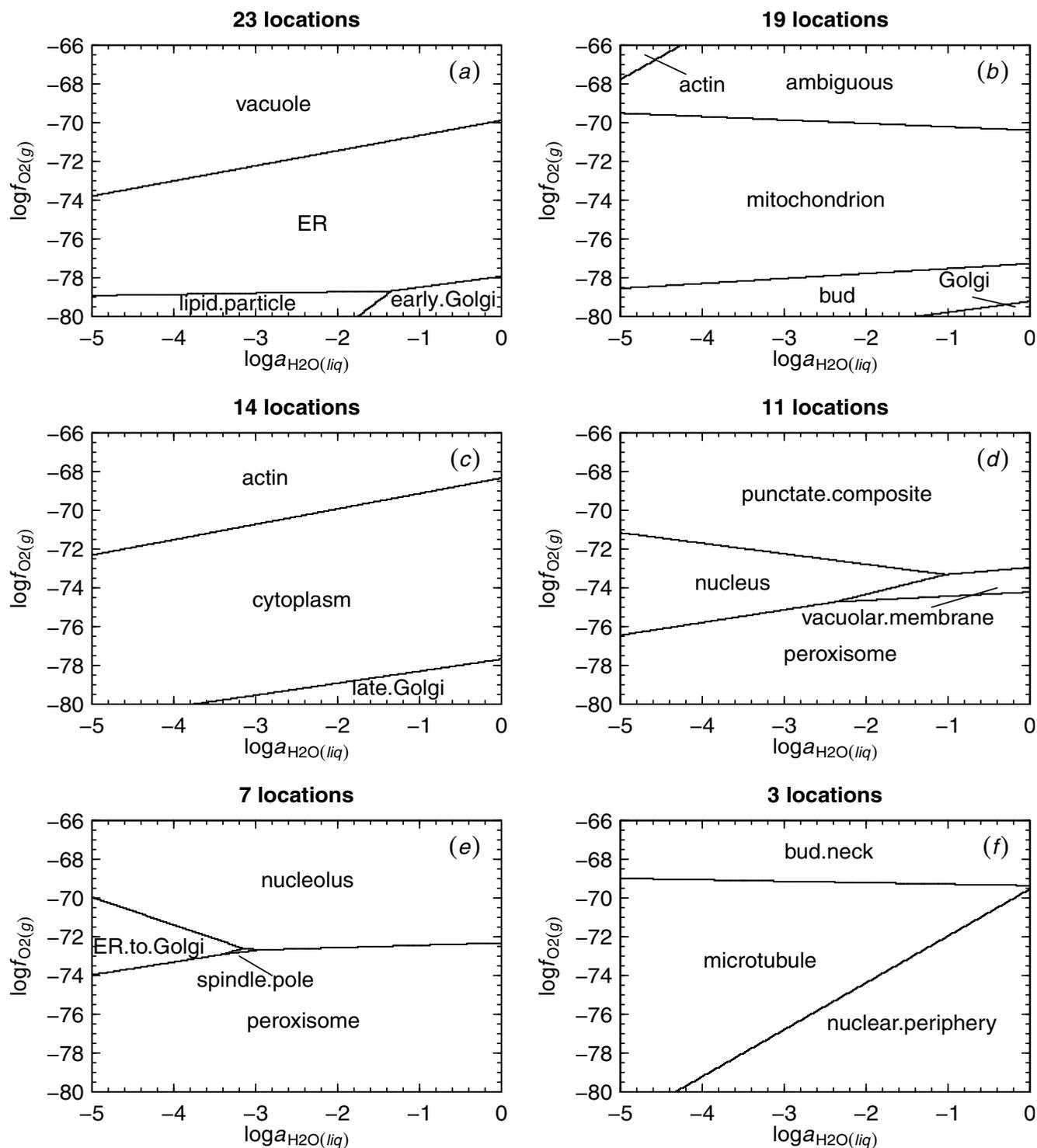
worthy that the reference model proteins representing the two cytoskeletal systems in yeast cells, actin and microtubule, appear near opposite ends of the energy spectrum. This outcome may be consistent with the observation that actin in different forms appears to be present at most stages of the cell cycle [43], but that the microtubule cytoskeleton grows during anaphase (i.e., the stage of the cell cycle characterized by physical separation of the chromosomes [44]) and is degraded during other stages of the cell cycle [43,44]. The outcome of the mitotic cycle in *S. cerevisiae* is the growth of a new cell in the form of a bud [44]. Not all structures in the bud form simultaneously. Instead, it has been observed that [45]"the endoplasmic reticulum, Golgi, mitochondria, and vacuoles all begin to populate the bud well before anaphase and that their segregation into the bud does not require microtubules". From Fig. 2 it is apparent that the proteins in the vacuole, ER, mitochondria and Golgi are all energetically less costly than many of their counterparts in other subcellular locations. The proteins in actin and lipid particles are also relatively metastable, which could imply that they too have a primary position in the formation of new cells.

These and other potential consequences of energetic differences between the biomacromolecules in subcellular compartments have not been fully explored.

**Intercompartmental protein interactions**

The diagrams in Fig. 2 show the metastability limits for interactions between predominant reference model proteins for different subcellular compartments. However, many subcellular interactions may in fact be meta-metastable with respect to the reaction boundaries shown in Fig. 2. For example, interactions occur between proteins in the cytoplasm and nucleus [46], but the reference model proteins for these compartments do not share a reaction boundary in Fig. 2. Below, known intercompartmental interactions are combined with the oxygen fugacity requirements for equal activities of the reference model proteins to characterize compartmental oxidation-reduction potentials.

To assess the biochemical evidence for specific interactions between proteins in different compartments in yeast cells, a series of review papers was surveyed



**Figure 2**

**Relative metastabilities of reference model proteins for compartments.** Predominance diagrams were generated as a function of  $\log f_{O_2(g)}$  and  $\log a_{H_2O}$  at 25°C and 1 bar for the reference model proteins listed in Table 3. The diagram in (a) represents 23 reference model proteins; diagrams in panels (b)-(f) represent systems with successively fewer reference model proteins as candidates.

**Table 4: Major intercompartmental protein interactions in yeast<sup>a</sup>.**

Interaction	$\Delta n_{O_2}$	$\log f_{O_{2(g)}}$	Interaction	$\Delta n_{O_2}$	$\log f_{O_{2(g)}}$
<b>actin-bud</b>	0.262	-74.9	<b>vacuole-bud</b>	0.384	-75.3
<b>actin-bud.neck</b>	0.078	-83.5	<b>vacuole-cell periphery</b>	0.305	-75.7
<b>actin-cell periphery</b>	0.183	-75.3	<b>vacuole-cytoplasm</b>	0.188	-73.4
<b>actin-endosome</b>	0.129	-75.6	<b>vacuole-endosome</b>	0.251	-75.9
<b>actin-vacuolar membrane</b>	0.081	-73.3	<b>vacuole-late Golgi</b>	0.358	-75.4
<b>actin-mitochondrion</b>	0.049	-64.2	<b>nucleus-actin</b>	-0.039	-74.3
<b>actin-microtubule</b>	0.123	-78.3	<b>nucleus-microtubule</b>	0.084	-80.1
<b>microtubule-bud</b>	0.139	-71.8	<b>nucleus-spindle pole</b>	0.014	-82.3
<b>microtubule-bud neck</b>	-0.045	-69.3	<b>nucleus-bud</b>	0.223	-74.9
<b>microtubule-cell periphery</b>	0.060	-69.3	<b>nucleus-bud neck</b>	0.039	-92.5
<b>microtubule-cytoplasm</b>	-0.057	-89.8	<b>nucleus-cytoplasm</b>	0.027	-59.7
<b>microtubule-spindle pole</b>	-0.070	-79.7	<b>nucleus-nucleolus</b>	-0.031	-70.0
<b>spindle.pole-cytoplasm</b>	0.013	-35.1	<b>nuclear periphery-bud neck</b>	-0.081	-69.4
<b>spindle.pole-nuclear periphery</b>	0.106	-76.3	<b>nuclear periphery-cytoplasm</b>	-0.093	-82.0
<b>ER-cell.periphery</b>	0.142	-82.3	<b>nuclear periphery-nucleus</b>	-0.120	-77.0
<b>ER-cytoplasm</b>	0.025	-96.5	<b>nuclear periphery-nucleolus</b>	-0.152	-75.5
<b>ER-early Golgi</b>	0.259	-78.0	<b>peroxisome-cell periphery</b>	0.064	-78.7
<b>ER-nuclear.periphery</b>	0.118	-85.0	<b>peroxisome-cytoplasm</b>	-0.053	-80.0
<b>ER-peroxisome</b>	0.078	-85.3	<b>peroxisome-lipid particle</b>	0.140	-74.9
<b>Golgi-endosome</b>	-0.205	-75.9	<b>peroxisome-mitochondrion</b>	-0.071	-80.0
<b>Golgi-vacuole</b>	-0.456	-75.9	<b>mitochondrion-cell periphery</b>	0.135	-79.4
<b>Golgi-late Golgi</b>	-0.097	-77.5	<b>mitochondrion-cytoplasm</b>	0.017	-79.9
<b>Golgi-early Golgi</b>	-0.034	-89.0	<b>mitochondrion-nucleus</b>	-0.010	-23.6

a. Interactions between proteins in different subcellular locations in *S. cerevisiae* were identified in the literature. The calculated reaction coefficients on  $O_{2(g)}$  and the metastable equilibrium value of  $\log f_{O_{2(g)}}$  were calculated for each reaction between reference model proteins. Names of locations shown in bold indicate that the model value of  $\log f_{O_{2(g)}}$  for this compartment (Table 3) lies in the metastability range for the reference model protein in the particular reaction.

[43,47,48,46,49,50]. The identified source statements are listed in Additional File 3, and simplified pairwise representations of the interactions are summarized in Table 4. Of 190 possible combinations between any two of the 20 subcellular compartments (this count excludes the ambiguous location and ER to Golgi and punctate composite, which did not appear in the literature survey), 46 interactions were identified through this survey.

Chemical reactions corresponding to each of the interactions listed in Table 4 are listed in Additional File 4. The values of  $\Delta n_{O_{2(g)}}$  (coefficient on  $O_{2(g)}$  in the reactions) are listed in Table 4 together with the values of  $\log f_{O_{2(g)}}$  calculated for equal chemical activities of the two reference model proteins in each reaction. Note that there are some reactions where the absolute value of  $\Delta n_{O_{2(g)}}$  is substantially smaller than the others; these include spindle pole-cytoplasm and mitochondrion-nucleus. Because of the small value of  $\Delta n_{O_{2(g)}}$  in these reactions, the values of  $\log$

$f_{O_{2(g)}}$  for equal activities of these proteins tend to be more extreme than for other reactions. The sign of  $\Delta n_{O_{2(g)}}$  denotes the thermodynamically favored direction of the reaction as  $\log f_{O_{2(g)}}$  is changed from its equal-activity value; for example, at  $\log f_{O_{2(g)}} = -74.9$ , the reference model proteins of actin and bud can metastably coexist with equal chemical activities, but at higher values that of actin predominates in a metastable assemblage.

The interactions listed in Table 4 were used to obtain model values of the oxygen fugacity in each compartment that are listed in Table 3. The model value of the oxygen fugacity for each compartment was selected so that in as many cases as possible the reactions listed in Table 4 favor the formation of the reference model protein for this compartment relative to those of interacting compartments. For example, the  $\log f_{O_{2(g)}}$  listed for the actin compartment is -74.7, which allows this reference model protein

to be metastable relative to its interacting partners in the first four reactions listed in Table 4. The limits of the model values of  $\log f_{O_2(g)}$  were set to between ca. -70 and -80, so that some of the more extreme values listed in Table 4 were not considered in the analysis.

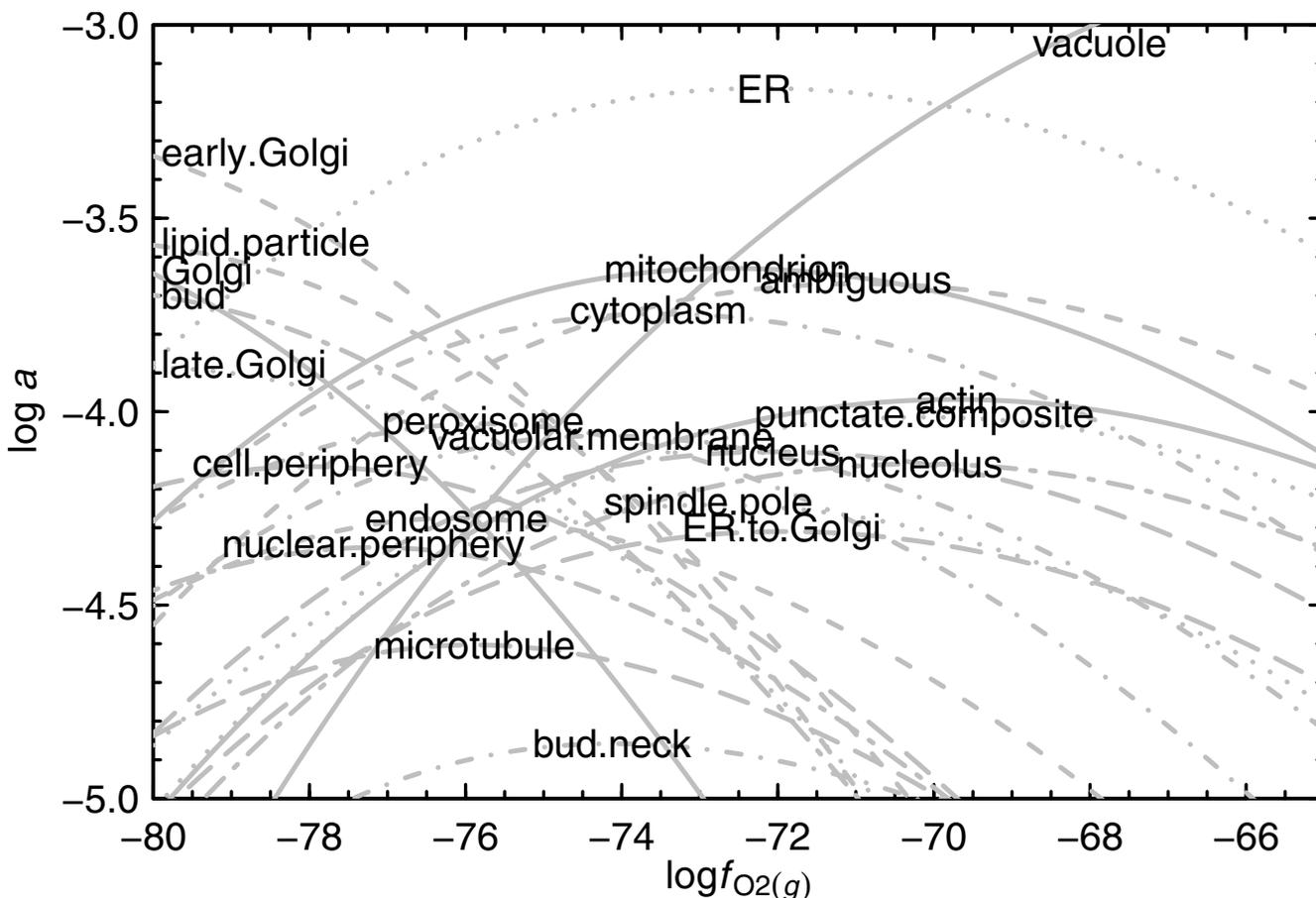
It is notable that at  $\log f_{O_2(g)} = -75$ , the reference model protein for microtubule is not metastable with respect to any of its interacting partners except for bud neck. The reference model protein for microtubule only becomes relatively metastable at high oxygen fugacities (w.r.t. bud and cell periphery) or at low oxygen fugacities (w.r.t. cytoplasm and spindle pole). Hence, the value of  $\log f_{O_2(g)} - 75$  taken here for the microtubule compartment is differ-

ent from the values for other compartments, in that this represents conditions where the formation of its reference model protein is more unfavorable than that of most of its interacting partners.

**Calculation of relative abundances of proteins**

Above, the interactions between subcellular homologs of enzymes and reference model proteins for subcellular compartments were used to derive oxygen fugacity limits for metastable reactions of proteins in different compartments. In the second part of this study, attention is focused on the relative abundances and intracompartamental interactions of proteins.

The logarithms of activities consistent with metastable equilibrium among all 23 reference model proteins are plotted in Fig. 3 as a function of  $\log f_{O_2(g)}$ . The relative



**Figure 3**  
**Metastable equilibrium abundances of reference model proteins and proteins as a function of oxygen fugacity.**  
 The relative abundances of the reference model proteins identified in Table 1 were calculated as a function of  $\log f_{O_2(g)}$  at 25°C and 1 bar and with total activity of amino acid residues equal to unity.

**Table 5: Oxygen fugacities, root mean square deviations and correlation coefficients in comparisons of intracompartamental protein interactions<sup>a</sup>.**

Location	Most abundant proteins				Model complexes				
	<i>n</i>	$\log f_{O_2(g)}$	RMSD	$\rho$	Complex	<i>n</i>	$\log f_{O_2(g)}$	RMSD	$\rho$
actin	22	-75.5	0.61	0.19	1	5	-77.0	0.49	-0.10
ambiguous	50	-73.5	0.90	0.42	2	7	-76.5	0.52	0.50
bud	50	-72.5	1.17	-0.02	3	5	-73.5	0.53	-0.30
bud neck	11	-75.5	0.73	0.02	4	6	-78.5	0.56	0.66
cell periphery	38	-74.5	0.63	0.42	5	4	-74.5	0.45	0.20
cytoplasm	50	-78.0	1.09	0.19	6	7	-78.5	0.64	-0.82
early Golgi	9	-74.0	0.72	0.45	7	4	-76.0	0.66	-0.80
endosome	30	-75.5	0.86	0.28	8	4	-76.5	0.76	-0.80
ER	49	-76.0	0.97	0.03	9	3	-77.0	0.07	1.00
ER to Golgi	5	-78.0	0.40	0.40	10	4	-76.0	0.45	1.00
Golgi	14	-76.0	0.88	-0.54	11	10	-74.5	0.62	-0.04
late Golgi	29	-76.0	0.73	0.17	12	5	-75.0	1.15	0.60
lipid particle	17	-78.0	0.92	0.22	13	12	-76.0	0.93	-0.19
microtubule	10	-75.0	0.61	0.36	14	7	-74.5	1.04	-0.75
mitochondrion	50	-75.0	0.53	0.46	15	17	-77.5	0.44	0.30
nuclear periphery	46	-76.0	0.62	0.32	16	23	-76.0	0.43	0.52
nucleolus	50	-74.0	0.72	0.18	17	6	-77.5	0.26	1.00
nucleus	50	-75.0	0.80	-0.02	18	5	-78.5	0.24	0.90
peroxisome	18	-75.5	0.55	0.56	19	8	-75.5	0.65	0.57
punctate composite	49	-74.0	0.78	0.19	20	15	-74.0	0.68	0.64
spindle pole	30	-75.5	1.02	0.12	21	5	-72.0	0.78	0.80
vacuolar membrane	45	-74.0	0.92	0.48	22	15	-74.0	0.94	0.50
vacuole	50	-74.5	1.42	0.23	23	9	-75.0	0.71	0.52

a. Values of  $\log f_{O_2(g)}$  in each location were obtained by comparing calculated and experimental logarithms of activities of the most abundant proteins in different subcellular compartments and of model complexes for each location (Additional File 6). *n* denotes the number of model proteins used in the calculations. RMSD values were calculated using Eqn. (7), and  $\rho$  denotes the Spearman rank correlation coefficient, calculated using Eqn. (8).

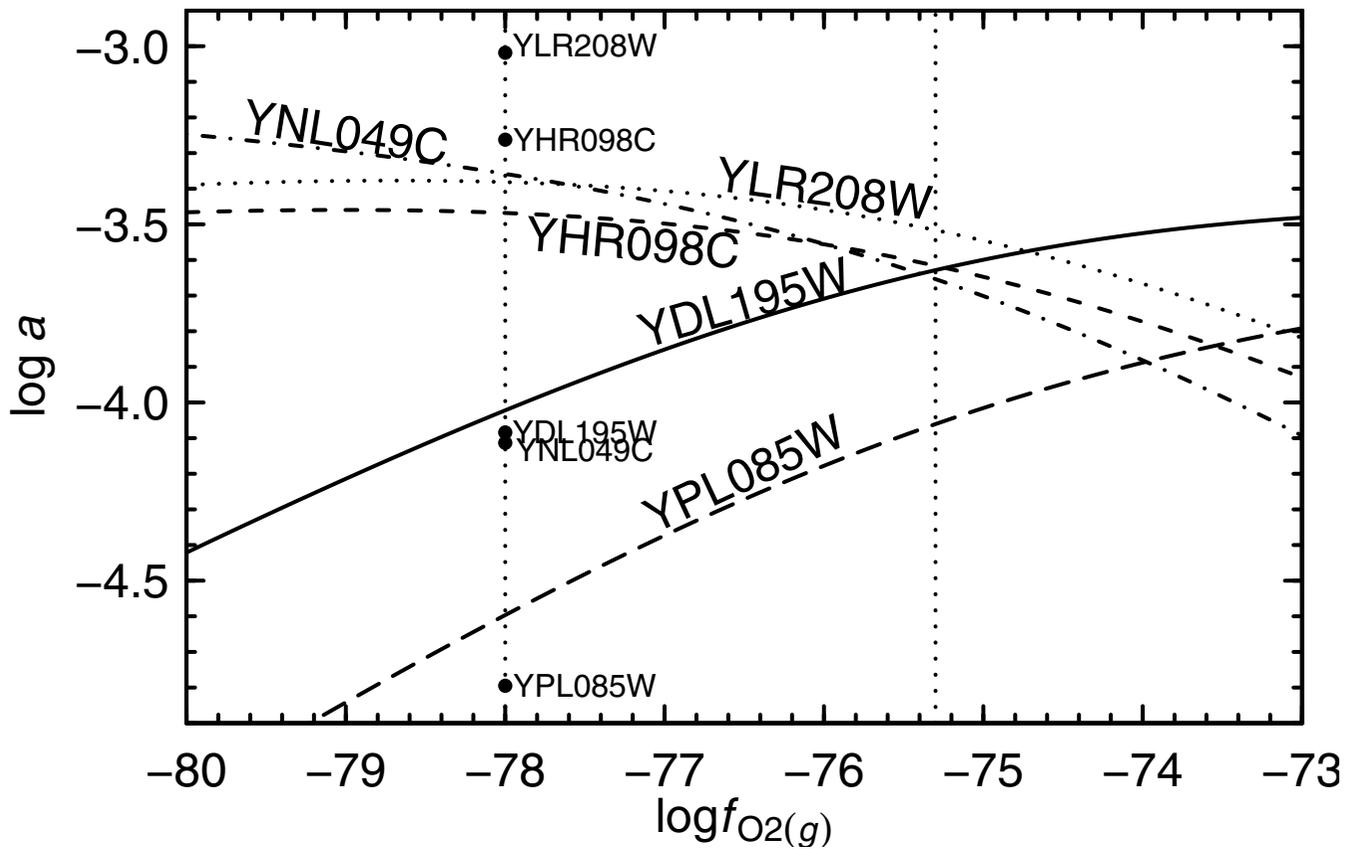
abundances of the proteins were calculated as described in Ref. [19] for reactions that conserve amino acid residues; a specific example of this type of calculation is described in the Methods. Fig. 3 presents in a different form the relationships shown in Fig. 2a at  $\log a_{H_2O} = 0$ . Note that the same proteins predominate at the extremes of oxygen fugacity represented in Fig. 3 and in Fig. 2a (reducing – early Golgi; oxidizing – vacuole) and that the reference model protein of microtubule appears with low relative abundance. Also note that there is a minimum in the range of calculated activities of the reference model proteins around  $\log f_{O_2(g)} = -75$  to  $-76$ ; changing oxidation-reduction potential alters not only the identity of the predominant protein in a metastably interacting population but also the relative abundances of all the others. There is probably not a single value of  $\log f_{O_2(g)}$  where the calculated relative abundances of the reference model proteins shown in Fig. 3 reflect the composition of the cell. Let us

therefore look more closely at the relative abundances of proteins within compartments.

**Relative abundances of proteins within compartments**

To model each of the compartments, up to 50 experimentally most abundant proteins were identified using data from Ref. [9]. The proteins that were selected were localized exclusively to each compartment, except for those of the bud. The numbers of proteins used to model each compartment are listed in Table 5, and the names of the proteins together with computational results given in Additional File 5.

In Fig. 4 the relative abundances of the five model proteins localized exclusively to ER to Golgi are shown as a function of  $\log f_{O_2(g)}$ . A worked-out example of the calculations leading to this figure is described in the Methods. The results of the calculations described there correspond to the dotted line at  $\log f_{O_2(g)} = -75.3$  in Fig. 4. At this oxygen fugacity, the rank order of abundances of the model



**Figure 4**

**Metastable equilibrium abundances of reference model proteins and proteins as a function of oxygen fugacity.**

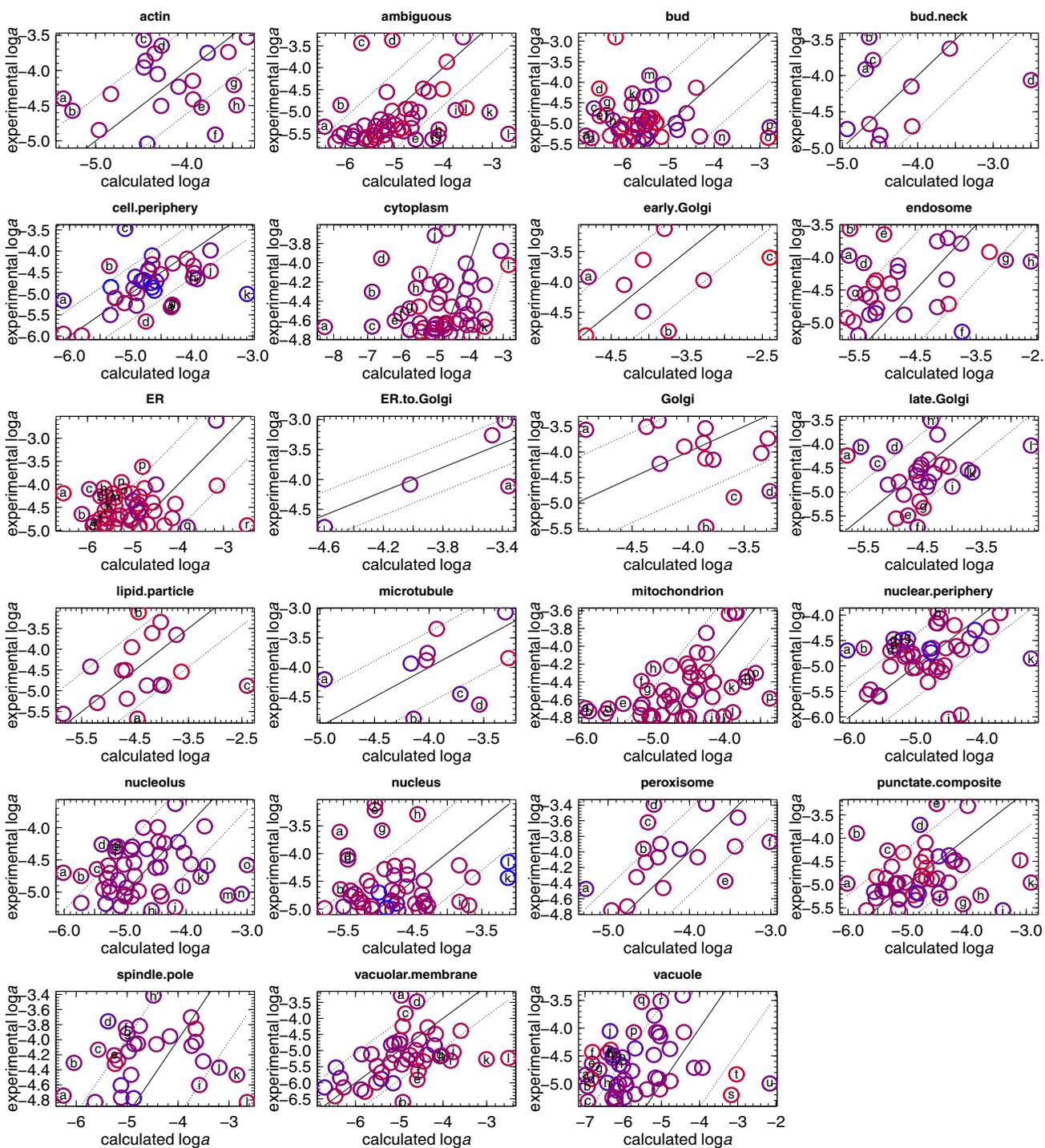
The relative abundances of five proteins localized to ER to Golgi whose experimental abundances were reported in [9] were calculated as a function of  $\log f_{O_2(g)}$  at 25°C and 1 bar and with total activity of amino acid residues equal to unity. The rightmost dotted line indicates conditions where the calculated abundance ranking of the proteins is identical to that found in the experiments, and the leftmost dotted line where the calculated logarithms of activities have a lower overall deviation from experimental ones, which are indicated by the points. This value of  $\log f_{O_2(g)}$  (-78) was used to construct the corresponding diagram in Fig. 5.

proteins in metastable equilibrium is identical to the rank order of experimental abundances. The figure was generated in whole by carrying out the calculation for different reference values of  $\log f_{O_2(g)}$ . There is a narrow range on either side of  $\log f_{O_2(g)} = -75.3$  (ca.  $\pm 0.05$ ) where the relative abundances of the proteins in metastable equilibrium occur in the same rank order. Beyond these limits, changing  $f_{O_2(g)}$  drives the composition of the metastable equilibrium assemblage to other states that do not overlap as closely with the experimental rankings.

The experimental abundances of the proteins reported by [9] are 21400, 12200, 1840, 1720 and 358, respectively,

in relative units. These abundances were scaled to the same total activity of amino acid residues (unity) used in the calculations to generate the experimental relative abundances plotted at the dashed line in Fig. 4 at  $\log f_{O_2(g)} = -78$ . Under these conditions, the metastable equilibrium abundances of the proteins do not occur in exactly the same rank order as the experimental ones, but there is a greater overall correspondence with the experimental relative abundances.

Similar calculations were repeated for each of the other compartments identified in Ref. [8]. The relative abundances of the proteins were calculated at 0.5 log unit



**Figure 5**

**Comparison of experimental and calculated logarithms of activities of proteins in compartments.** Red and blue colors denote, respectively, low and high average nominal carbon oxidation states ( $\bar{Z}_C$ ) of the protein. Dotted lines are positioned at one RMSD above and below one-to-one correspondence, which is denoted by the solid lines. Outlying points are labeled with letters that are keyed to the proteins in Additional File 5. The values of  $\log f_{O_{2(g)}}$  used in the calculations are listed in Table 5.

increments from  $\log f_{O_{2(g)}} = -82$  to  $-70.5$ . Scatterplots of the experimental vs. calculated relative abundances are shown in Additional File 6. These comparisons were assessed to obtain values of  $\log f_{O_{2(g)}}$ , listed in Table 5, that yield the best fit between calculated and experimental relative abundances. The best-fit calculated relative abundances are listed together with the experimental ones in Additional File 5, and the corresponding best-fit scatterplots for each set of model proteins are shown in Fig. 5.

The retrieval of optimal values of  $\log f_{O_{2(g)}}$  was aided by calculating the root mean square deviation (RMSD) of logarithms of activities using Eqn. (7) and the Spearman rank correlation coefficient ( $\rho$ ; Eqn. 8) between experimental and calculated logarithms of activities. The dotted lines in Fig. 5 were drawn at one RMSD on either side of the one-to-one correspondence, denoted by the solid lines in this figure. The RMSD values were used to identify outliers that are identified in Fig. 5 by letters that are listed in Additional File 5. To aid in distinguishing the points, they were assigned colors on a red (reduced) – blue (oxidized) scale that reflects the average nominal oxidation state of carbon of the protein (Eqn. 6).

There is a considerable degree of scatter apparent in many of the plots shown in Fig. 5, so a low degree of certainty may be associated with the  $\log f_{O_{2(g)}}$  values regressed from these comparisons. In specific cases such as peroxisome and nuclear periphery a lower overall deviation is apparent and a positive correlation appears between the calculated and experimental relative abundances. Because they were regressed from intracompartamental protein abundance data, the values of  $\log f_{O_{2(g)}}$  listed in Table 5 might not be as representative of subcellular oxidation-reduction conditions as those listed in Table 3, which have the additional benefit of being based partly on known subcellular interactions (see above).

The comparisons depicted in Fig. 5 and in Additional File 6 are also significant because they reveal that the range of protein abundances observed in cells is accessible in a metastable equilibrium assemblage at some values of  $\log f_{O_{2(g)}}$ . For example, the range of experimental abundances of the model proteins in actin covers about 1.6 orders of magnitude, while the calculated abundances vary over about 2.2 orders of magnitude. Extreme values

**Table 6: List of selected complexes<sup>a</sup>.**

#	Location	Complex	References
1	actin	Arp2/3 complex	[68,69] (423)
2	ambiguous	cyclin-dependent protein kinase complex	(343)
3	bud	actin-associated motor protein complex 2	[70] (49)
4	bud.neck	septin complex	[71] (333)
5	cell.periphery	exocyst complex	(120)
6	cytoplasm	translation initiation factor eIF3	(45)
7	early.Golgi	SNARE complex	[72] (113)
8	endosome	ESCRT I & II complexes	[73,74]
9	ER	signal recognition complex	(52)
10	ER.to.Golgi	coatamer COPII complex	(340)
11	Golgi	Golgi transport complex	(293)
12	late.Golgi	retrograde protein complex	[75] (114)
13	lipid.particle	sterol biosynthesis enzymes	[53]
14	microtubule	DASH complex	[76]
15	mitochondrion	mitochondrial ribosome small subunit	(9)
16	nuclear.periphery	nuclear pore complex	[77]
17	nucleolus	small subunit processome	[78] (70)
18	nucleus	RNA polymerase I	(30)
19	peroxisome	integral to peroxisomal membrane	(GO:0005779)
20	punctate.composite	proteins localized here and early.Golgi	
21	spindle.pole	spindle-pole body complex	[79] (219)
22	vacuolar.membrane	VO vacuolar ATPase complex	(14)
23	vacuole	vacuolar proteases and other canonical proteins	[52]

a. Numbers in parentheses refer to the ID number of the complex, if available, from <http://yeast-complexes.embl.de/>. embl.de [80]. Compositions and localizations of complexes were also taken from references listed in square brackets.

of  $\log f_{O_2(g)}$  tend to weaken this correspondence. The lowest degree of correspondence occurs for the cytoplasmic proteins, where  $\sim 5$  orders of magnitude separate the predicted relative abundances of the top 50 most abundant proteins, which in the experimental measurements have a dynamic range spanning about 1.2 orders of magnitude. The great degree of scatter apparent in many of the comparisons in Fig. 5 could be partly a consequence of including in the comparisons model proteins that do not actually interact with each other, despite their high relative abundances. To address this concern, a more directed approach was adopted below that takes account of fewer numbers of proteins that are known to interact through the formation of complexes.

#### Relative abundances of proteins in complexes

The correspondence between the calculated and experimental relative abundances of the five model proteins in ER to Golgi raises the question of what characteristics of the proteins might be responsible for this result. Scanning the functional annotations of these proteins reveals that they are part of the COPII coat complex [51]. The results for this model system suggested that focusing on specific complexes in other compartments could yield interesting results. Because the interactions of proteins to form complexes is essential in cellular structure and regulating the functions of enzymes [12], factors that affect the relative abundances of the complexing proteins may be fundamental to the control of metabolic processes.

The model complexes used in this study are identified in Table 6 and the individual proteins in each complex are listed in Additional File 7. Each complex was nominally associated with a subcellular compartment based on the names and descriptions of the complexes available in the literature. Some exceptions are the cyclin-dependent protein kinase complex, the proteins of which are largely cytoplasmic and nuclear [8], but here is placed in the slot for the ambiguous location because no definitely ambiguously localized complexes could be identified. The proteins listed in Additional File 7 under punctate composite are not part of a named complex but were chosen because they are localized to early Golgi and have the punctate composite characterization [8]. The other exceptions are the vacuolar model proteins (proteases and other canonical vacuolar proteins [52]), enzymes of the ergosterol biosynthetic pathway, some of which are associated with the lipid particle [53], and proteins integral to the peroxisomal membrane, which were identified using the Gene Ontology (GO) annotations in the SGD [51].

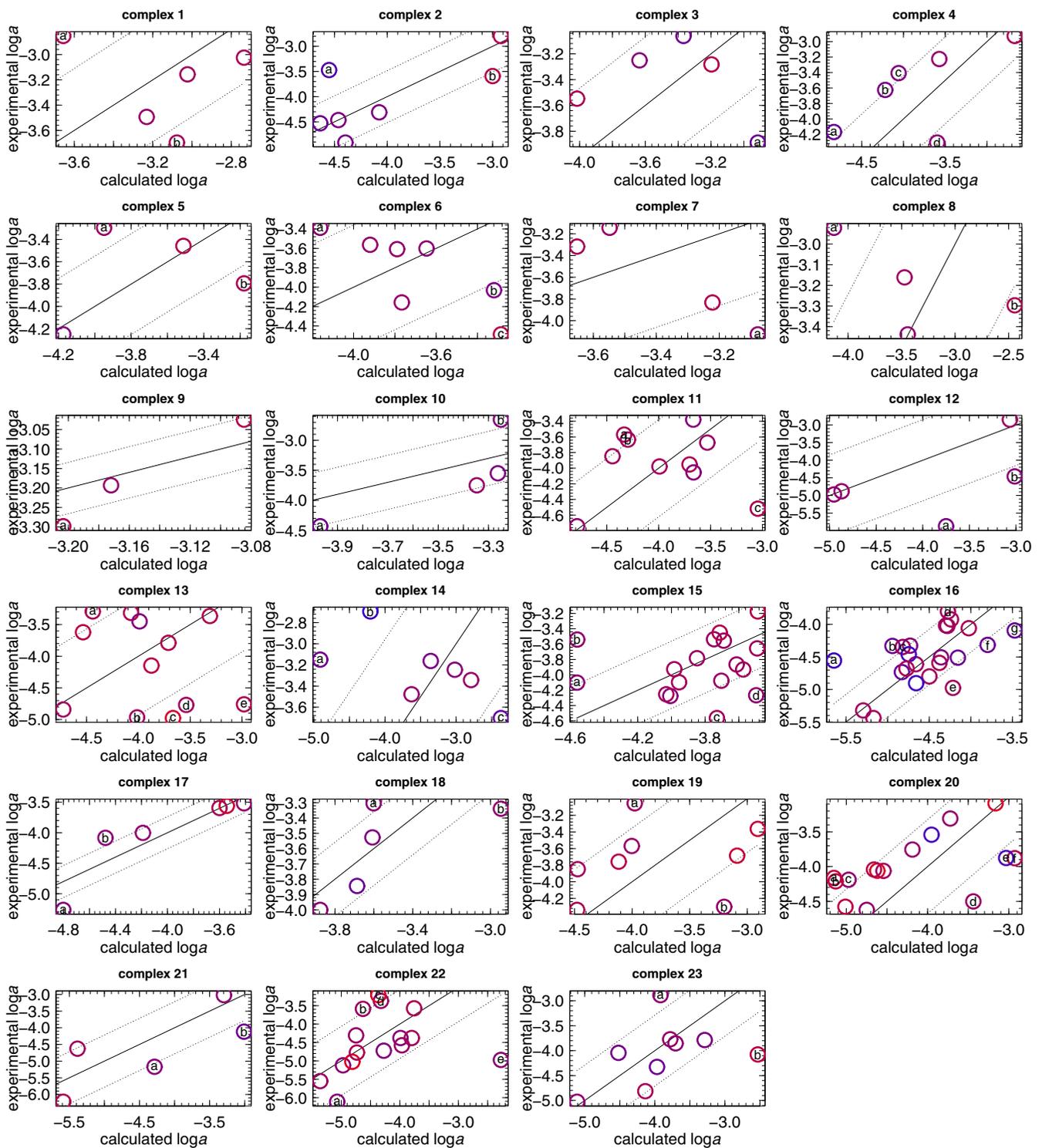
The calculated metastable equilibrium logarithms of activities of the proteins in each complex are shown as a

function of  $\log f_{O_2(g)}$  in Additional File 8. The calculated logarithms of activities of the proteins were compared with experimental ones by constructing scatterplots at 0.5 log unit intervals from  $\log f_{O_2(g)} = -82$  to  $-70.5$ , which are shown in Additional File 6. As described above, visual assessment of fit was used in combination with the RMSD and Spearman rank correlation coefficients to obtain values of  $\log f_{O_2(g)}$  that maximize the correspondence with experimental relative abundances. The resulting calculated relative abundances are listed together with the experimental ones in Additional File 9.

The number of model proteins in the complexes is less than the number of most abundant proteins in the compartments considered in the preceding section. Some of the model complexes represented in Fig. 6 exhibit an apparent positive correlation between calculated and experimental logarithms of activities; these include nuclear pore complex and small subunit processome. A negative correlation between calculated and experimental logarithms of activities is apparent for proteins in the ESCRT I & II complexes and DASH complex. A few of the other complexes (Golgi transport complex, sterol biosynthesis enzymes) exhibit very little overall correspondence between calculated and experimental logarithms of activities.

The results in Fig. 6 permit an interpretation of the relative energetic requirements for formation of different groups of interacting proteins. Take for example complex 14, which is the DASH complex that associates with the microtubule. A negative correlation between the experimental and calculated relative abundances is apparent for this complex in Fig. 6. The RMSD between calculated and experimental logarithms of activities of proteins is 1.04, which is among the highest listed in Table 5. Note from Eqn. (11) that a  $\sim 1$  log unit change in the chemical activity of a chemical species corresponds to a Gibbs energy difference equal to  $2.303RT$ . An average difference of  $\sim 1$  between calculated and experimental logarithms of activity indicates that the formation of the proteins requires  $2.303RT = 1364 \text{ cal mol}^{-1}$  beyond what would be needed if the proteins formed in metastable equilibrium relative abundances. On the other hand, the formation in specific oxidation-reduction conditions of proteins making up other assemblages where cellular abundances positively correlate with and span the same range as the metastable equilibrium distribution can proceed close to a local minimum energy required for protein formation.

Because of their relatively high energy demands, proteins in complexes such as the DASH complex and the spindle



**Figure 6**  
**Comparison of experimental and calculated logarithms of activities of model proteins for complexes.** Symbols are as in Fig. 5; the model proteins and the outliers (identified by letters) are listed in Additional File 9.

pole body are likely to be more dynamic in the cell. (Note that although a positive rank correlation coefficient for the latter complex is reported in Table 5, at a lower oxygen fugacity ( $\log f_{O_{2(g)}} = -77.5$ ) an inverse correlation results between experimental abundances and calculated metastable equilibrium relative abundances of the proteins in this complex; see Additional File 6). The finding made elsewhere of some inverse relationships between relative abundance of proteins and corresponding mRNA levels was interpreted as evidence for additional effort on the part of the cell [54]. An inverse relationship that opposes equilibrium may be favored in evolution because of the strategic advantage of incorporating otherwise costly (rare) amino acids that increase enzymatic diversity [55].

The differences in the numbers of proteins considered in each of the comparisons implies that the values of the correlation coefficients are not directly comparable. The p-values for each of the correlations listed in Table 5 were calculated and are reported in Additional File 10. The p-value is the probability that the value of the observed correlation coefficient can be met or exceeded by a random configuration of the system. The present calculations suggest that the lowest p-values are associated with the collections of greater than ca. 40 proteins listed in Table 5 that have a Spearman rank correlation coefficient greater than ca. 0.4. Using the p-value as a criterion, the most convincing demonstrations of the existence of correlations appear in the most abundant proteins of the vacuolar membrane and cell periphery. By comparison, the smaller systems of proteins making up complexes, which in some cases have higher correlation coefficients, also have relatively high p-values, indicating a greater probability that the same result can be obtained in a random configuration.

#### **Implications of the hypothesis**

Tests of several specific predictions of the hypothesis were discussed in the preceding sections. The major results of these calculations and comparisons are listed below.

1. Subcellular homologs of glutaredoxin and thioredoxin are metastable at different  $\log f_{O_{2(g)}}$  ranges. The mitochondrial homolog appears to be more reduced, and the nuclear one most oxidized. Reactions within the glutaredoxin system also exhibit sensitivity to hydration state.
2. Reference model proteins for 23 subcellular locations also have metastability limits in  $\log f_{O_{2(g)}}$  space. The relationships are consistent with a relatively oxi-

dized nuclear reference model protein, but that for the mitochondrion is intermediate between the nucleus and the cytoplasm. Among the reference model proteins predicted to be most stable (Fig. 2a–b), four are known to be involved in the early stages of formation of the bud. Golgi is predicted to be a reduced compartment while actin and the vacuole are relatively oxidized.

3. The least stable reference model proteins are those for the microtubule and bud neck (Fig. 2f and Fig. 3). This observation suggests that the proteins in the microtubule are very reactive with other cellular components, and/or that they have a relatively high turnover rate.

4. Observed trends in the relative abundances of the most abundant proteins in some compartments can be correlated with the relative abundances of proteins predicted using a metastable equilibrium model. Correlations between observed and predicted relative abundances for smaller numbers of proteins that make up complexes can also be documented. In some cases negative correlations may be supported, such as for the DASH complex (microtubule), translation initiation factor, and the early Golgi SNARE complex. The maintenance of these complexes might entail a higher energy demand than for others in the cell.

If the hypothesis adopted for this study was true, it would imply that there are processes that impart an energetic bias on the appearance of proteins in specific compartments. The thermodynamic model described above by itself gives no information about the possible nature of processes involved. Two processes that could be important are the work against diffusional gradients required for active protein transport and the turnover rates of subcellular populations of proteins. Regarding the former, it is the gradient of chemical potential (not concentration) of the biomacromolecule, that appears in statements such as Fick's Law. Differences in the chemical conditions between compartments would be expected to differentially contribute to the activity coefficients of proteins, so that the cost of transport to various compartment is not equal. Although the activity coefficients of proteins were not considered in this study, their values might depend on oxidation or hydration potential, so the the current results could be implicitly influenced by nonideality in the subcellular system. Regarding the latter process, one may expect that the turnover rates of proteins are tied to the local chemical environment. If, for example, the turnover rate of a population of proteins minimizes at a specific oxygen fugacity, then any deviations away from this oxidation potential would increase the turnover rate and

cause the cell to expend more energy in maintaining this population.

If chemical energy minimization by the cell results in an increase in fitness, the energetic effects of the physical-chemical processes outlined above may constrain the overall process of natural selection. Therefore, the hypothesis also implies that an expected outcome of evolution is the formation of biomacromolecules with lower energy demands compared with other possible, and otherwise equal, products. How does this connect with the mechanism of protein transport and trafficking, i.e. that from their place of synthesis (ribosomes) proteins are transported to different subcellular locations, often under the influence of specific signal sequences? At one extreme, it is possible that the energetic differences between compartments are not influential in the evolution of the mechanism of protein sorting and trafficking. However, if the signal sequences themselves are chemically reactive to varying degrees depending on their subcellular environment, then selection for mutations in them might be tuned to both function and chemistry. It would not be surprising then to find evidence for the chemical adaptation of signal sequences to specific compartments.

These results and observations support the notion that changing oxidation-reduction potential can selectively alter the potential for reactions leading to formation of proteins and their complexes. Chemical selectivity in the dynamic formation in the cell of high-energy proteins could lead to transient formation of complexes that function only under certain conditions. Because of the different stability limits of the reference model proteins in log  $f_{O_2(s)}$  space, these results also in principle support the notion that "a fundamental redox attractor underpins ... core cellular processes"[56]. In reality, many chemical properties vary spatially in cells, including the hydration state, pH, activities of  $CO_2$  and  $H_2S$ , and temperature and pressure in the extracellular environment. These all factor into the Gibbs energy changes accompanying the chemical transformations between proteins, as do the thermodynamic properties of protein folding reactions and nonideality in protein solutions. Because of its energetic basis, the model used here can be extended in the future to incorporate the effects of these variables. Building these relationships into a multidimensional thermodynamic assessment is a promising avenue for predicting the chemical features of proteomic adaptation in the context of the cellular environment.

## Methods

The essential steps in the calculations reported here are 1) defining standard states, 2) identifying model proteins for systems of interest, 3) assessing the relative abundances of model proteins in metastable equilibrium, 4) visualizing the results of the calculations on chemical diagrams and 5) comparing the computational results with experimental biochemical and proteomic data.

### Standard states and chemical activities

The activity of a species is related to the chemical potential of the species by

$$\mu = \mu^\circ + RT \ln a,$$

where  $R$  and  $T$  represent, respectively, the gas constant and the temperature,  $\mu$  and  $\mu^\circ$  stand for the chemical potential and standard chemical potential, respectively, and  $a$  denotes activity. No provision for activity coefficients of proteins or other species was used in this study; under this approximation, the activity of an aqueous species is equal to its concentration (molality).

The standard state for aqueous species including proteins specifies unit activity of the aqueous species in hypothetical one molal solution referenced to infinite dilution. The standard molal Gibbs energies of the proteins were calculated with the CHNOSZ software package [19] using group additivity properties and parameters taken from Ref. [23].

### Reference model proteins for amino acid compositions

The overall amino acid compositions of proteins in 23 subcellular locations in *S. cerevisiae* were calculated by combining localization [8] and abundance [9] data for proteins measured in the YeastGFP project with amino acid compositions of proteins downloaded from the *Saccharomyces* Genome Database (SGD) [51]. Of 4155 ORF names listed in the YeastGFP dataset, all but 12 are present in SGD (the missing ones are YAR044W, YBR100W, YDR474C, YFL006W, YFR024C, YGL046W, YGR272C, YJL012C-A, YJL017W, YJL018W, YJL021C and YPR090W).

To generate reference model proteins that are most representative of each compartment, proteins that were annotated in the YeastGFP study as being localized to more than one compartment were excluded from this analysis (except for bud; see below), as were those for which no abundance was reported. The names of the open reading frames (ORFs) corresponding to the proteins in the YeastGFP data set were matched against the SGD's protein\_properties.tab file downloaded on 2008-08-04. This search yielded a number of model proteins for each compartment, ranging from 5 (ER to Golgi) to 746 (cyto-

plasm); see Table 3. The names of the compartments used throughout the tables and figures in this paper correspond to the notation used in the YeastGFP data files.

It was found that no proteins with reported abundances and localized to the bud were exclusive to that compartment, hence all of the proteins localized there (which also have localizations in other compartments) were taken as models for the bud reference model protein. The amino acid composition of the reference model protein for each compartment was calculated by taking the sum of the compositions of each model protein for a compartment in proportion to its fractional abundance in the total model protein population of the compartment. The resulting amino acid compositions are listed in Additional File 2. The corresponding chemical formulas of the nonionized reference model proteins and the calculated standard molal Gibbs energies of formation from the elements at 25°C and 1 bar of the ionized reference model proteins are shown in Table 3.

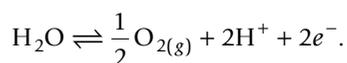
#### Metastability calculations

Diagrams showing the predominant proteins and the relative abundances of proteins in metastable equilibrium were generated using the CHNOSZ software package [19]. These calculations take account of formation reactions of the proteins written for their residue equivalents [19]. An example of this approach is described further below for a specific model system.

The basis species appearing in the formation reactions studied here are  $\text{CO}_{2(aq)}$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_{3(aq)}$ ,  $\text{O}_{2(g)}$ ,  $\text{H}_2\text{S}_{(aq)}$  and  $\text{H}^+$ . The reference activities used for the basis species were  $10^{-3}$ ,  $10^0$ ,  $10^{-4}$ ,  $10^{-7}$  and  $10^{-7}$ , respectively, for  $\text{CO}_{2(aq)}$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_{3(aq)}$ ,  $\text{H}_2\text{S}_{(aq)}$  and  $\text{H}^+$ . In the case of diagrams showing Eh as a variable, the aqueous electron ( $e^-$ ) was substituted for  $\text{O}_{2(g)}$  in the basis species. Reference values for  $a_{e^-}$  or  $f_{\text{O}_{2(g)}}$  are not listed here because one or the other is used as an independent variable in each of the calculations described above.

#### Conversion between scales of oxidation-reduction potential

Conversion between the  $\log f_{\text{O}_{2(g)}}$  and Eh scales of oxidation-reduction potential can be made by first writing the half-cell reaction for the dissociation of  $\text{H}_2\text{O}$  as



Taking  $\text{pH} = -\log a_{\text{H}^+}$  and  $\text{pe} = -\log a_{e^-}$ , the logarithmic analog of the law of mass action for Reaction 2 can be written as:

$$\log K_2 = \frac{1}{2} \log f_{\text{O}_{2(g)}} - 2\text{pH} - 2\text{pe} - \log a_{\text{H}_2\text{O}},$$

where  $\log K_2$  stands for the logarithm of the equilibrium constant of Reaction 2 as a function of temperature and pressure. Eh is related to pe by [57]

$$\text{pe} = \frac{F}{2.303RT} \text{Eh},$$

where  $F$  and  $R$  denote the Faraday constant and the gas constant, respectively. Combining Eqns. (3) and (4) yields the following expression for Eh as a function of  $\log f_{\text{O}_{2(g)}}$  and other variables:

$$\text{Eh} = \frac{2.303RT}{2F} \left( \frac{1}{2} \log f_{\text{O}_{2(g)}} - 2\text{pH} - \log a_{\text{H}_2\text{O}} - \log K_2 \right).$$

At 25°C and 1 bar,  $F/2.303RT = 16.903 \text{ volt}^{-1}$  and  $\log K_2 = -41.55$ ; for  $\text{pH} = 7$  and  $\log a_{\text{H}_2\text{O}} = 0$ , a value of  $\text{Eh} = 0 \text{ V}$  corresponds to  $\log f_{\text{O}_{2(g)}} = -55$ . Eqn. (5) permits the conversion between Eh and  $\log f_{\text{O}_{2(g)}}$  as well at other temperatures, pHs, and activities of  $\text{H}_2\text{O}$ .

#### Average nominal oxidation state of carbon

Let us write the chemical formula of a species of interest as  $\text{C}_{n_C}\text{H}_{n_H}\text{N}_{n_N}\text{O}_{n_O}\text{S}_{n_S}^Z$  where  $Z$  denotes the net charge. The average nominal oxidation state of carbon ( $\bar{Z}_C$ ) of this species is given by

$$\bar{Z}_C = \frac{Z - n_H + 2(n_O + n_S) + 3n_N}{n_C}.$$

Eqn. (6) is consistent with the electronegativity rules described in [58] and is compatible with the equation for average oxidation number of carbon used in [59]. For example, Eqn. (6) can be used to calculate the average nominal oxidation states of carbon in  $\text{CO}_2$  and  $\text{CH}_4$ , which are +4 and -4, respectively. Note that the proportions of oxygen and other covalently bonded heteroatoms contribute to the value of  $\bar{Z}_C$  of a protein or other molecule, but that proton ionization does not alter the nominal carbon oxidation state, because of the opposite

contributions from  $Z$  and  $n_H$  in Eqn. (6). In the 4143 proteins identified in the YeastGFP subcellular localization study and found in the *Saccharomyces* Genome Database, the minimum and maximum of  $\bar{Z}_C$  are -0.414 and 0.390, respectively. Of the proteins in this dataset, six have  $\bar{Z}_C < -0.35$  (YDR193W, YDR276C, YEL017C-A, YJL097W, YML007C-A, YMR292W) and six have  $\bar{Z}_C > 0.15$  (YCL028W, YHR053C, YHR055C, YKR092C, YMR173W, YPL223C). The points in the scatterplots in this paper (Figs. 5 and 6 and Additional File 6) are colored on a continuous red-blue scale according to the value of  $\bar{Z}_C$  of the proteins, where maximum red occurs at  $\bar{Z}_C = -0.35$  and maximum blue occurs at  $\bar{Z}_C = 0.15$ .

**Comparison with experimental relative abundances**

The root mean square deviation between calculated and experimental logarithms of activities was calculated using

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (X_{calc,i} - X_{expt,i})^2}{n}}$$

where  $X_{calc,i}$  and  $X_{expt,i}$  denote the calculated and experimental logarithms of activities and  $n$  stands for the number of proteins. In the calculations described above, experimental abundances of proteins in each model system were scaled so that the total chemical activity of amino acid residues was equal to unity.

The Spearman rank correlation coefficient ( $\rho$ ) was calculated using

$$\rho = 1 - \frac{6d}{n(n^2 - 1)}$$

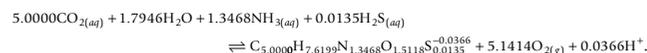
where  $d = \sum_{i=1}^n (x_{calc,i} - x_{expt,i})^2$  and  $X_{calc,i}$  and  $X_{expt,i}$  stand for the ranks of the corresponding logarithms of activities.

**Calculating relative abundances of proteins in metastable equilibrium**

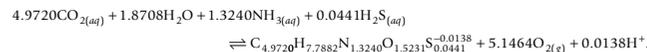
The following example demonstrates the procedure used to calculate the relative abundances of proteins in metastable equilibrium. The model proteins for ER to Golgi, in order of decreasing abundance in the cell reported by [9], are YLR208W, YHR098C, YDL195W, YNL049C and YPL085W. (For simplicity, the proteins are identified here by the names of the open reading frames (ORF).) The formula of the uncharged form of the first protein, YLR208W, is  $C_{1485}H_{2274}N_{400}O_{449}S_4$ , and its amino acid

sequence length is 297 residues. The standard molal Gibbs energy of formation from the elements ( $\Delta G_f^\circ$ ) of this protein at 25°C and 1 bar calculated using group additivity [23] is -10670 kcal mol<sup>-1</sup>. At this temperature and pressure and at pH = 7, group additivity can also be used to calculate the charge of the protein (-10.8832) and the standard molal Gibbs energy of formation from the elements of the charged protein (-10880 kcal mol<sup>-1</sup>). The formula of the protein in this ionization state is  $C_{1485}H_{2263.1168}N_{400}O_{449}S_4^{-10.8832}$ . Dividing by the length of the protein, we find that the formula and standard molal Gibbs energy of formation from the elements of the residue equivalent of YLR208W are  $C_{5.0000}H_{7.6199}N_{1.3468}O_{1.5118}S_{0.0135}^{-0.0366}$  and -36.633 kcal mol<sup>-1</sup>, respectively.

The formation from basis species of the residue equivalent of YLR208W is consistent with



Similar reasoning can be applied to write the formation reaction of the residue equivalent of YHR098C as



At 929 residues, YHR098C is over 3 times as long as YLR208W, but in the formation reactions from the basis species of the residue equivalents of the two proteins, the coefficients on the basis species are similar. The difference between the coefficients of the same basis species in the reactions signifies the response of the metastable equilibrium assemblage to changes in the corresponding chemical activity or fugacity. For example, because  $v_{CO_2,9} < v_{CO_2,10}$ ,  $v_{NH_3,9} < v_{NH_3,10}$  and  $v_{O_2,9} < v_{O_2,10}$  increasing  $a_{CO_2(aq)}$ ,  $a_{NH_3(aq)}$  or  $f_{O_2(g)}$  at constant  $T, P$  and chemical activities of the other basis species shifts the metastable equilibrium in favor of YLR208W at the expense of YHR098C. Here,  $v_i$  denotes the reaction coefficient of the  $i$ th basis species or protein, which is negative for reactants and positive for products as written. Conversely, because  $v_{H_2O,9} > v_{H_2O,10}$ ,  $v_{H_2S,9} > v_{H_2S,10}$  and  $v_{H^+,9} > v_{H^+,10}$  increasing  $a_{H_2O}$ ,  $a_{H_2S(aq)}$  or  $a_{H^+}$  (decreasing pH) at constant  $T, P$  and chemical activities of the other basis species shifts the metastable equilibrium in favor of YHR098C at the expense of YLR208W. The magnitude of the effect is proportional to the size of the differ-

ence between the coefficients of the basis species in the reactions, and it can be quantified for a specific model system using the following calculations.

To assess the relative abundances of the proteins in metastable equilibrium, we proceed by calculating the chemical affinities of each of the formation reactions. The chemical affinity ( $A$ ) is calculated by combining the equilibrium constant ( $K$ ) with the reaction activity product ( $Q$ ) according to [60]

$$A / 2.303RT = \log(K / Q) = \log \left( \frac{-\Delta G_r^\circ / 2.303RT}{\prod a_i^{v_i}} \right),$$

where 2.303 is the natural logarithm of 10,  $R$  stands for the gas constant,  $T$  is temperature in degrees Kelvin,  $\Delta G_r^\circ$  is the standard molal Gibbs energy of the reaction, and  $a_i$  and  $v_i$  represent the chemical activity and reaction coefficient of the  $i$ th basis species or species of interest (i.e., residue equivalent of the protein) in the reaction. Let us calculate  $\Delta G_r^\circ$  (in kcal mol<sup>-1</sup>) of Reaction 9 by writing

$$\begin{aligned} \Delta G_9^\circ &= 1 \times -36.633 + 5.1414 \times 0 + 0.0366 \times 0 \\ &\quad - 5.0000 \times -92.250 - 1.7946 \times -56.688 \\ &\quad - 1.3468 \times -6.383 - 0.0135 \times -6.673 \\ &= 535.036. \end{aligned}$$

In Eqn. (12) the values of  $\Delta G_f^\circ$  of O<sub>2(g)</sub> and H<sup>+</sup> are both zero, which are consistent with the standard state conventions for gases and the hydrogen ion convention used in solution chemistry. The values of  $\Delta G_f^\circ$  of the other basis species are taken from the literature [61-63]. The value of  $\log K_9$ , consistent with Eqn. (12) is -392.19.

We now calculate the activity product of the reaction using

$$\begin{aligned} \log Q_9 &= 1 \times 0 + 5.1414 \times -75.3 + 0.0366 \times -7 \\ &\quad - 5.0000 \times -3 - 1.7946 \times 0 - 1.3468 \times -4 - 0.0135 \times -7 \\ &= -366.92. \end{aligned}$$

The values of  $a_i$  used to write Eqn. (13) are the reference values listed in the Methods for  $a_{\text{CO}_2(\text{aq})}$ ,  $a_{\text{H}_2\text{O}}$ ,  $a_{\text{NH}_3(\text{aq})}$ ,  $a_{\text{H}_2\text{S}(\text{aq})}$  and  $a_{\text{H}^+}$ . The value of  $f_{\text{O}_2(\text{g})}$  used in Eqn. (13) ( $\log f_{\text{O}_2(\text{g})} = -75.3$ ) is also a reference value that, it will be shown, characterizes a metastable equilibrium distribution of proteins that is rank-

identical to the measured relative abundances of the proteins. Finally, the value of  $a$  of the residue equivalent of the protein in Eqn. (13) is set to a reference value of unity ( $\log a = 0$ ). If we are only concerned with the relative abundances of the proteins in metastable equilibrium, the actual value used here does not matter so long as it is the same in the analogous calculations for the other proteins.

Combining Eqns. (11)–(13) yields  $A_9/2.303RT = -25.25$  (this is a non-dimensional number). Following the same procedure for the other four proteins (YHR098C, YDL195W, YNL049C and YPL085W) results in  $A/2.303RT$  equal to -24.86, -24.74, -24.93 and -24.94, respectively. Now let us turn to the relative abundances of the proteins in metastable equilibrium, which can be expressed in a manner analogous to a Maxwell-Boltzmann distribution:

$$\frac{a_i}{a_t} = \frac{e^{A_i / RT}}{\sum_{i=1}^n e^{A_i / RT}},$$

where  $a_t$  denotes the total activity of residue equivalents in the system and  $n$  stands for the number of proteins in the system. Note regarding the left-hand side of Eqn. (14) that because we are taking activity coefficients of unity, the ratio  $a_i/a_t$  is equal to the ratio of concentrations of residue equivalents in the system. No negative sign appears in front of  $A/RT$  in the exponents Eqn. (14) because the chemical affinity is the negative of Gibbs energy change of the reaction. Note in addition that the values of  $A/2.303RT$  given above must be multiplied by  $\ln 10 = 2.303$  before being substituted in Eqn. (14). By taking  $a_i = 1$ , we can combine Eqn. (14) with  $A/RT$  of each of the formation reactions to calculate chemical activities of the residue equivalents of the proteins equal to 0.0905, 0.2248, 0.2994, 0.1944 and 0.1909, respectively. The lengths of the proteins are 297, 929, 1273, 876 and 2195, so the corresponding logarithms of activities of the proteins are e.g.  $\log(0.0905/297) = -3.52$  for YLR208W, and -3.61, -3.63, -3.65 and -4.06 for the remaining proteins, respectively.

### Competing interests

The author declares that he has no competing interests.

### Authors' contributions

JMD conceived the study and wrote the manuscript.

## Additional material

### Additional file 1

**Program script and data files for generating figures.** This program script and supporting files were used to generate the figures shown above. To generate the figures, the contents of the zip file can be placed into the R working directory before loading the CHNOSZ package (version 0.8). Then read in the script with `source('plot.R')`. More details on the operation are provided at the top of the script file.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S1.zip>]

### Additional file 2

**Amino acid compositions of reference model proteins.** Overall amino acid compositions of proteins in subcellular locations of *S. cerevisiae* were calculated from YeastGFP localization [8] and abundance [9] data downloaded from <http://yeastgfp.ucsf.edu/> combined with protein compositions downloaded from the Saccharomyces Genome Database <http://www.yeastgenome.org/>. The amino acid compositions of the reference model proteins were used to calculate the properties listed in Table 3.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S2.csv>]

### Additional file 3

**Interactions between subcellular compartments in yeast.** This file lists statements from Refs. [43,47,48,46,49,50] used to identify the interactions between proteins in different compartments of *Saccharomyces cerevisiae* that are listed in Table 4.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S3.pdf>]

### Additional file 4

**Intercompartmental protein reactions.** This table lists chemical reactions between residue equivalents of reference model proteins for interactions identified above. The charges of the reference model proteins were calculated at 25°C, 1 bar and pH = 7.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S4.txt>]

### Additional file 5

**Abundance data for model proteins for compartments.** For the up to 50 most abundant model proteins in each compartment are listed the ORF name, sequence length, average nominal oxidation state of carbon (Eqn. 6), computed standard molal Gibbs energy at 25°C and 1 bar of the ionized protein and charge at pH = 7 and calculated and experimental logarithm of activity. This file also identifies the outlying points labeled with letters in Fig. 3.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S5.csv>]

### Additional file 6

**Abundance comparison for model proteins for compartments and complexes.** Scatterplots of experimental vs. calculated logarithm of activity of model proteins in subcellular compartments were generated for a range of logarithm of oxygen fugacity from -82 to -70.5. The legend of each diagram indicates the logarithm of oxygen fugacity (" $\log f_{O_2}$ " in the legend), root mean square deviation (" $r_{msd}$ " in the legend; RMSD in Eqn. 7) and the Spearman rank correlation coefficient (" $r_r$ " in the legend;  $\rho$  in Eqn. 8).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S6.pdf>]

### Additional file 7

**Identities of proteins in selected complexes.** Lists the proteins in the selected model complexes and whether their abundances are reported in the YeastGFP dataset. Proteins without experimental abundance data were not used in the comparisons discussed in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S7.pdf>]

### Additional file 8

**Plots of relative abundances of model proteins for complexes.** The calculated relative abundances of model proteins in selected complexes are shown as a function of  $\log f_{O_2(s)}$ .

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S8.pdf>]

### Additional file 9

**Abundance data for model proteins for complexes.** For model proteins in selected complexes (see Additional File 7) are listed the ORF name, sequence length, average nominal oxidation state of carbon (Eqn. 6), computed standard molal Gibbs energy at 25°C and 1 bar of the ionized protein and charge at pH = 7 and calculated and experimental logarithm of activity.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S9.csv>]

### Additional file 10

**Calculation of p-values for abundance rank correlations.** This file lists calculated p-values for the Spearman rank correlation coefficients and describes the steps used in the calculations.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-75-S10.pdf>]

## Acknowledgements

The comments of two anonymous reviewers helped to improve this paper. This material is based upon work supported by the National Science Foundation under grant EAR-0309829 and the Department of Energy under grant DE-FG02-03ER151418, both awarded to the author's Ph.D. advisor, Professor Harold C. Helgeson. His enthusiasm for thermodynamics made this project possible.

## References

- Preston RA, Murphy RF, Jones EW: **Assay of vacuolar pH in yeast and identification of acidification-defective mutants.** *Proc Natl Acad Sci USA* 1989, **86(18)**:7027-7031.
- Hwang C, Sinskey AJ, Lodish HF: **Oxidized redox state of glutathione in the endoplasmic reticulum.** *Science* 1992, **257(5076)**:1496-1502.
- Morrill GA, Kostellow AB, Osterlow K, Gupta RK: **Differences in hydration state of nucleus and cytoplasm of the amphibian oocyte.** *J Membrane Biol* 1996, **153**:45-51.
- Al-Habori M: **Microcompartmentation, metabolic channelling and carbohydrate metabolism.** *Int J Biochem Cell Biol* 1995, **27(2)**:123-132.
- Aw TY: **Intracellular compartmentation of organelles and gradients of low molecular weight species.** In *Microcompartmentation and Phase Separation in Cytoplasm*, Int. Rev. Cytol Volume 192. Edited by: Walter H, Brooks DE, Srere PA. San Diego: Academic Press; 2000:223-253.
- Cedano J, Aloy P, Pérez-Pons JA, Querol E: **Relation between amino acid composition and cellular location of proteins.** *J Mol Biol* 1997, **266**:594-600.
- Andrade MA, O'Donoghue SI, Rost B: **Adaptation of protein surfaces to subcellular location.** *J Mol Biol* 1998, **276(2)**:517-525.
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast.** *Nature* 2003, **425(6959)**:686-691.
- Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS: **Global analysis of protein expression in yeast.** *Nature* 2003, **425(6959)**:737-741.
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol Biol Cell* 2000, **11(12)**:4241-4257.
- Schekman R: **Protein localization and membrane traffic in yeast.** *Annu Rev Cell Biol* 1985, **1**:115-143.
- Doxsey S, McCollum D, Theurkauf W: **Centrosomes in cellular regulation.** *Annu Rev Cell Dev Biol* 2005, **21**:411-434.
- Halvorson H: **Intracellular protein and nucleic acid turnover in resting yeast cells.** *Biochim Biophys Acta* 1958, **27(2)**:255-266.
- Morowitz HJ: **Foundations of Bioenergetics.** New York: Academic Press; 1978.
- Seligmann H: **Cost-minimization of amino acid usage.** *J Mol Evol* 2003, **56(2)**:151-161.
- Swire J: **Selection on synthesis cost affects interprotein amino acid usage in all three domains of life.** *J Mol Evol* 2007, **64(5)**:558-571.
- Berezovsky IN, Zeldovich KB, Shakhnovich EI: **Positive and negative design in stability and thermal adaptation of natural proteins.** *PLoS Comput Biol* 2007, **3(3)**:498-507.
- Kondepudi DK, Prigogine I: *Modern Thermodynamics: From Heat Engines to Dissipative Structures* New York: John Wiley & Sons; 1998.
- Dick JM: **Calculation of the relative metastabilities of proteins using the CHNOSZ software package.** *Geochem Trans* 2008, **9**:10.
- Wicken JS: **A thermodynamic theory of evolution.** *J Theor Biol* 1980, **87**:9-23.
- Aita T, Husimi Y: **Fitness spectrum among random mutants on Mt. Fuji-type fitness landscape.** *J Theor Biol* 1996, **182(4)**:469-485.
- Demetrius L, Ziehe M: **Darwinian fitness.** *Theor Popul Biol* 2007, **72(3)**:323-345.
- Dick JM, LaRowe DE, 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.
- Pedrajas JR, Porras P, Martínez-Galisteo E, Padilla CA, Miranda-Vizuete A, Bärceña JA: **Two isoforms of *Saccharomyces cerevisiae* glutaredoxin 2 are expressed in vivo and localize to different subcellular compartments.** *Biochem J* 2002, **364**:617-623.
- Molina MM, Belli G, de la Torre MA, Rodríguez-Manzanique MT, Herrero E: **Nuclear monothiol glutaredoxins of *Saccharomyces cerevisiae* can function as mitochondrial glutaredoxins.** *J Biol Chem* 2004, **279**:51923-51930.
- Herrero E, Ros J, Tamarit J, Belli G: **Glutaredoxins in fungi.** *Photosynth Res* 2006, **89**:127-140.
- Pedrajas JR, Kosmidou E, Miranda-Vizuete A, Gustafsson JA, Wright APH, Spyrou G: **Identification and functional characterization of a novel mitochondrial thioredoxin system in *Saccharomyces cerevisiae*.** *J Biol Chem* 1999, **274(10)**:6366-6373.
- Trotter EW, Grant CM: **Overlapping roles of the cytoplasmic and mitochondrial redox regulatory systems in the yeast *Saccharomyces cerevisiae*.** *Eukaryot Cell* 2005, **4**:392-400.
- Garrels RM: *Mineral Equilibria* New York: Harper & Brothers; 1960.
- Dahm LJ, Jones DP: **Rat jejunum controls luminal thiol-disulfide redox.** *J Nutr* 2000, **130(11)**:2739-2745.
- Trotter EW, Grant CM: **Non-reciprocal regulation of the redox state of the glutathione-glutaredoxin and thioredoxin systems.** *EMBO Rep* 2003, **4**:184-188.
- Drakulic T, Temple MD, Guido R, Jarolim S, Breitenbach M, Attfield PV, Dawes IW: **Involvement of oxidative stress response genes in redox homeostasis, the level of reactive oxygen species, and ageing in *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2005, **5(12)**:1215-1228.
- Hanson GT, Aggeler R, Oglesbee D, Cannon M, Capaldi RA, Tsien RY, Remington SJ: **Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators.** *J Biol Chem* 2004, **279(13)**:13044-13053.
- Macville M, Schröck E, Padilla-Nash H, Keck C, Ghadimi BM, Zimonjic D, Popescu N, Ried T: **Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping.** *Cancer Res* 1999, **59**:141-150.
- ATCC: **The Global Bioresource Center: Product Description.** CRL-1606 2008 [<http://www.atcc.org/tabid/452/Default.aspx?ATCCNum=CRL-1606&Template=cellBiology>].
- Mojaverian P: **Evaluation of gastrointestinal pH and gastric residence time via the Heidelberg Radiotelemetry Capsule: Pharmaceutical application.** *Drug Dev Res* 1996, **38(2)**:73-85.
- Imai T, Ohno T: **Measurement of yeast intracellular pH by image processing and the change it undergoes during growth phase.** *J Biotech* 1995, **38(2)**:165-172.
- Llopis J, McCaffery JM, Miyawaki A, Farquhar MG, Tsien RY: **Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins.** *Proc Natl Acad Sci USA* 1998, **95(12)**:6803-6808.
- Singh A, Kaur N, Kosman DJ: **The metalloredoxase Fre6p in *Fe*-Efflux from the yeast vacuole.** *J Biol Chem* 2007, **282(39)**:28619-28626.
- Hansen JM, Go YM, Jones DP: **Nuclear and mitochondrial compartmentation of oxidative stress and redox signaling.** *Annu Rev Pharmacol Toxicol* 2006, **46**:215-234.
- Go YM, Jones DP: **Redox compartmentalization in eukaryotic cells.** *Biochim Biophys Acta-Gen Subj* 2008, **1780(11)**:1271-1290.
- Päuser S, Zschunke A, Khuen A, Keller K: **Estimation of water content and water mobility in the nucleus and cytoplasm of *Xenopus laevis* oocytes by NMR spectroscopy.** *Magn Reson Imaging* 1995, **13(2)**:269-276.
- Botstein D, Amberg D, Mulholland J, Huffaker T, Adams A, Drubin D, Stearns T: **The yeast cytoskeleton.** In *The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:1-90.
- Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD: *Molecular Biology of the Cell* 2nd edition. New York: Garland Publishing, Inc; 1989.
- Lew DJ, Weinert T, Pringle JR: **Cell cycle control in *Saccharomyces cerevisiae*.** In *The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:607-695.
- Wente SR, Gasser SM, Caplan AJ: **The nucleus and nucleocytoplasmic transport in *Saccharomyces cerevisiae*.** In *The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:471-546.
- Kaiser CA, Gimeno RE, Shaywitz DA: **Protein secretion, membrane biogenesis, and endocytosis.** In *The Molecular and Cellular Biology of the Yeast *Saccharomyces*: Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:91-227.
- Jones EW, Webb GC, Hiller MA: **Biogenesis and function of the yeast vacuole.** In *The Molecular and Cellular Biology of the Yeast *Sac**

- charomyces: *Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:363-470.
49. Lazarow PB, Kunau W: **Peroxisomes**. In *The Molecular and Cellular Biology of the Yeast Saccharomyces: Cell Cycle and Cell Biology* Edited by: Pringle JR, Broach JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1997:547-605.
  50. Pon L, Schatz G: **Biogenesis of yeast mitochondria**. In *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics* Edited by: Broach JR, Pringle JR, Jones EW. New York: Cold Spring Harbor Laboratory Press; 1991:333-406.
  51. SGD Project: **Saccharomyces Genome Database**. 2007 [<http://www.yeastgenome.org>].
  52. Sarry JE, Chen S, Collum RP, Liang S, Peng M, Lang A, Naumann B, Dziarszinski F, Yuan CX, Hippler M, Rea PA: **Analysis of the vacuolar luminal proteome of *Saccharomyces cerevisiae***. *FEBS J* 2007, **274(16)**:4287-4305.
  53. Mo CQ, Bard M: **Erg28p is a key protein in the yeast sterol biosynthetic enzyme complex**. *J Lipid Res* 2005, **46(9)**:1991-1998.
  54. Tuller T, Kupiec M, Ruppin E: **Determinants of protein abundance and translation efficiency in *S. cerevisiae***. *PLoS Comput Biol* 2007, **3(12)**:2510-2519.
  55. Wicken JS: *Evolution, Thermodynamics, and Information* Oxford University Press; 1987.
  56. Murray DB: **On the temporal self-organisation of *Saccharomyces cerevisiae***. *Curr Genomics* 2004, **5(8)**:665-671.
  57. Drever JL: *The Geochemistry of Natural Waters* 3rd edition. Upper Saddle River, New Jersey: Prentice Hall; 1997.
  58. Hendrickson JB, Cram DJ, Hammond GS: *Organic Chemistry* 3rd edition. New York: McGraw-Hill; 1970.
  59. Buvet R: **General criteria for the fulfillment of redox reactions**. In *Bioelectrochemistry I: Biological Redox Reactions, of Ettore Majorana International Science Series Volume 11*. Edited by: Milazzo G, Blank M. New York: Plenum Press; 1983:15-50.
  60. Prigogine I, Defay R: *Chemical Thermodynamics* London: Longmans, Green and Co; 1954.
  61. Helgeson HC, Kirkham DH: **Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: I. Summary of the thermodynamic/electrostatic properties of the solvent**. *Am J Sci* 1974, **274(10)**:1089-1198.
  62. Wagman DD, Evans WH, Parker VB, Schumm RH, Halow I, Bailey SM, Churney KL, Nuttall RL: **The NBS tables of chemical thermodynamic properties. Selected values for inorganic and C<sub>1</sub> and C<sub>2</sub> organic substances in SI units**. *J Phys Chem Ref Data* 1982, **11**:1-392.
  63. Shock EL, Helgeson HC, 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.
  64. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003**. *Nucleic Acids Res* 2003, **31**:365-370.
  65. Shock EL, Sassani DC, Willis M, Sverjensky DA: **Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes**. *Geochim Cosmochim Acta* 1997, **61(5)**:907-950.
  66. Chan P, Lovrić J, Warwicker J: **Subcellular pH and predicted pH-dependent features of proteins**. *Proteomics* 2006, **6(12)**:3494-3501.
  67. Wu MM, Llopis J, Adams S, McCaffery JM, Kulomaa MS, Machen TE, Moore HPH, Tsien RY: **Organelle pH studies using targeted avidin and fluorescein-biotin**. *Chem Biol* 2000, **7(3)**:197-209.
  68. Welch MD, Iwamatsu A, Mitchison TJ: **Actin polymerization is induced by Arp2/3 protein complex at the surface of *Listeria monocytogenes***. *Nature* 1997, **385(6613)**:265-269.
  69. Mullins RD, Pollard TD: **Structure and function of the Arp2/3 complex**. *Curr Opin Struct Biol* 1999, **9(2)**:244-249.
  70. Schmid M, Jaedicke A, Du TG, Jansen RP: **Coordination of endoplasmic reticulum and mRNA localization to the yeast bud**. *Curr Biol* 2006, **16(15)**:1538-1543.
  71. Frazier JA, Wong ML, Longtine MS, Pringle JR, Mann M, Mitchison TJ, Field C: **Polymerization of purified yeast septins: Evidence that organized filament arrays may not be required for septin function**. *J Cell Biol* 1998, **143(3)**:737-749.
  72. Burri L, Lithgow T: **A complete set of SNAREs in yeast**. *Traffic* 2004, **5**:45-52.
  73. Kostelansky MS, Schluter C, Tam YYC, Lee S, Ghirlando R, Beach B, Conibear E, Hurley JH: **Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer**. *Cell* 2007, **129(3)**:485-498.
  74. Hierro A, Sun J, Rusnak AS, Kim J, Prag G, Emr SD, Hurley JH: **Structure of the ESCRT-II endosomal trafficking complex**. *Nature* 2004, **431(7005)**:221-225.
  75. Conibear E, Cleck JN, Stevens TH: **Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p**. *Mol Biol Cell* 2003, **14(4)**:1610-1623.
  76. Miranda JLL, De Wulf P, Sorger PK, Harrison SC: **The yeast DASH complex forms closed rings on microtubules**. *Nat Struct Mol Biol* 2005, **12(2)**:138-143.
  77. Rout MP, Aitchison JD, Suprpto A, Hjertaas K, Zhao YM, Chait BT: **The yeast nuclear pore complex: Composition, architecture, and transport mechanism**. *J Cell Biol* 2000, **148(4)**:635-651.
  78. Bernstein KA, Gallagher JEG, Mitchell BM, Granneman S, Baserga SJ: **The small-subunit processome is a ribosome assembly intermediate**. *Eukaryot Cell* 2004, **3(6)**:1619-1626.
  79. Vinh DBN, Kern JW, Hancock WO, Howard J, Davis TN: **Reconstitution and characterization of budding yeast  $\gamma$ -tubulin complex**. *Mol Biol Cell* 2002, **13(4)**:1144-1157.
  80. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffmann V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G: **Proteome survey reveals modularity of the yeast cell machinery**. *Nature* 2006, **440(7084)**:631-636.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

