Proteomic indicators of oxidation and hydration state in colorectal cancer

Jeffrey M. Dick[*1]

[1]**Language Institute, Chiang Mai University, Chiang Mai, Thailand**

ABSTRACT

New integrative approaches are needed to harness the potential of rapidly growing data sets of gene and protein expression and microbial taxonomic identification in colorectal cancer (CRC). Using reported data from human proteomic and microbial taxonomic studies, this study compares the average oxidation state of carbon (Z_C) and water demand per residue ($\bar{n}_{H_2O}$) of proteins from tumors to those in healthy tissues or less advanced cancer stages. The major compositional trends are lower Z_C and, to a lesser extent, higher $\bar{n}_{H_2O}$, in tumor vs normal groups, or carcinoma vs adenoma groups. Comparison of microbial protein compositions shows a small shift toward lower Z_C in bacteria enriched in fecal samples from cancer patients. Thermodynamic calculations of the relative chemical stabilities of proteins show that the cancer-related proteins tend to be stabilized by higher chemical activity of H_2O (more hydrating) and/or lower fugacity of O_2 (more reducing). The compositional exploration of molecular data suggests that a systematic chemical transformation is a central aspect of the cancer proteome. The thermodynamic calculations highlight the potential for interpreting proteomic data within a biochemical context, and may contribute to a better understanding of the microenvironmental requirements for cancer initiation and progression.

Keywords: colorectal cancer, proteomics, gut microbiome; redox potential; chemical thermodynamics

1 INTRODUCTION

Colorectal cancer (CRC) is one of the most common and well-studied human cancers, and is a model for the prevailing theory of genetic transformation as the primary driver of cancer progression (Kinzler and Vogelstein, 1996). Differential gene expression and proteomic data sets are now available to characterize numerous specific experimental and clinical situations. Recently, many studies have generated proteomic data for candidate biomarkers for early detection of colorectal cancer with a focus on the membrane fractions (Chen et al., 2010; Kume et al., 2014; Sethi et al., 2015), while others have focused on subcellular responses (e.g. chromatin; Knol et al., 2014) or stage-specific effects (Uzozie et al., 2014; Sethi et al., 2015). Likewise, microbial associations with CRC are increasingly well characterized using 16S rRNA and metagenomic data (Wang et al., 2012; Zeller et al., 2014; Candela et al., 2014).

[*]j3ffdick@gmail.com

The broad issue that motivates this study is how can molecular data be used to understand biological processes underlying oncogenesis? Many approaches have been utilized in the interpretation of gene- and protein-expression data; e.g. heat maps and cluster algorithms (such as principal-components analysis or PCA), network analysis integrated with large-scale protein and gene interaction networks, and functional classification using the Gene Ontology (GO) or other sources of functional information. These techniques often have as goals the identification of significantly altered genes or proteins for further interrogation as biomarkers, or functional annotations that inform conceptual and mathematical models of molecular interactions. Nevertheless, despite the widespread use of bioinformatic tools, there is growing concern that the full potential of the vast molecular data sets has not been realized (e.g. Keating and Cambrosio, 2012).

Compared to interpretations focused on molecular interactions and mechanisms of gene regulation, high-level descriptions and analysis of molecular data have other benefits. For example, the information-theoretic concept of maximum entropy has been used to identify gene expression signatures in carcinogenesis (Zadran et al., 2013). Notably, the information-theoretic based maximum entropy calculations are analogous to thermodynamic equilibrium; these calculations are performed without explicit accounting for the underlying biomolecular interactions, but still allow for quantification of a potential energy "attractor space", or stable or steady state of the system. However, information-theoretic models are limited in their assessment of environmental influences such as oxidation and hydration potential. Theoretical calculations of the relative stabilities of populations of proteins in an environmental context are possible in a chemical thermodynamic framework. This study proposes an application of the concept of chemical affinity as a measure of stability (i.e. propensity for formation) which is grounded in energetic calculations that account for differences in amino acid composition (Dick et al., 2006; LaRowe and Dick, 2012). The chemical thermodynamic descriptions have the advantage of being expressed in terms of variables that can be compared with independent biochemical measurements.

One of the characteristic features of tumors is varying degrees of hypoxia (Höckel and Vaupel, 2001). The sensitivity of tumors to oxygen is exploited by hyperbaric oxygen treatment to enhance the effects of radiotherapy (Bertout et al., 2008). Hypoxic conditions also promote the subcellular (mitochondrial) generation of reactive oxygen species (ROS) (Murphy, 2009). In addition to the molecular responses associated with hypoxia, cancer cells and tissues exhibit changes in oxidation-reduction (redox) state. Redox potential (Eh) monitored *in vivo* in a fibrosarcoma cell line is altered compared to normal fibroblasts (Hutter et al., 1997), and it has been suggested that cancer cells have suppressed ranges of redox oscillations during the cell cycle (Hoffman et al., 2008).

The hydration state of cancer cells and tissues may also vary considerably from their normal counterparts. Microwave detection of differences in dielectric constant resulting from greater water content in malignant tissue is being explored for medical imaging of breast cancer (Lazebnik et al., 2007; Grzegorczyk et al., 2012). IR and Raman spectroscopic techniques also provide evidence for a greater hydration state of cancerous breast tissue, resulting from interaction of water molecules with hydrophilic cellular structures of cancer cells but negligible association with the hydrophobic molecules (triglycerides, fatty acid derivatives) more common in normal cells

(Abramczyk et al., 2014). Increased hydration levels may also be associated with increased hyaluronan in the extracellular matrix (ECM) of migrating and metastatic cells (Toole, 2002).

Besides these specific responses associated with oxidation and hydration state, other considerations point to a general importance of oxidation and hydration state for metabolic, cellular and physiological processes. For example, the differential reactions and transport of lactate and pyruvate are tightly interlinked with carbon oxidation state in metabolic pathways (Brooks, 2009), and have implications for cancer metabolism (Semenza, 2008). It has been hypothesized that the increased hydration of cancer cells underlies a reversion to a more embryonic state (McIntyre, 2006). Therefore, these two variables – redox and hydration potential – have been selected as the focus of the descriptions and explorations in this study.

The compositional and thermodynamic relationships are described in the Methods. Because reaction coefficients on O_2 and H_2O are to some extent interdependent, an appropriate choice of basis species is needed in order to separate the effects of the variables. The first part of the Results shows the comparisons for human and microbial proteins (Sections 3.1–3.2) using molecular data taken from the proteomic studies cited in the first paragraph above, as well as from one gene-expression study (Hlubek et al., 2007). There is a major decrease in the average oxidation state of carbon in chromatin-binding proteins in carcinoma and in proteins in some membrane fractions of tumor samples (data from Knol et al., 2014; Chen et al., 2010; Kume et al., 2014), and an increase in the water demand of proteins in precancerous adenomas (Uzozie et al., 2014), and other membrane fractions (Sethi et al., 2015). The protein compositions of bacteria that are more abundant in cancer-derived samples are also generally shifted toward lower oxidation state of carbon. These systematic patterns in the chemical compositions of proteins are aligned with the physiological hypoxic and hydrated conditions of many cancer cells.

To better understand the biochemical limits of these changes, the second part of the Results turns to a thermodynamic description in terms of intensive chemical variables. By using a chemical affinity (negative Gibbs energy) calculation, a chemical thermodynamic prediction of the most stable molecules (akin to a potential "attractor space") can be generated (Sections 3.3–3.5). The stability fields of the cancer-associated proteins are found more often at lower oxygen fugacity and/or higher water activity. The transitions between stability regions of cancer and healthy-related proteins in terms of oxidation and hydration potential vary among data sets, suggesting that specific cell-biological and physiological effects can be identified.

Not only multistep genetic changes, but microenvironmental dynamics are interrelated determinants of cancer progression (Schedin and Elias, 2004). Here, biochemical patterns in proteomic data sets were explored using descriptions grounded in chemical thermodynamics. Mapping the independently calculated chemical stabilities allows for a step forward from a purely compositional description, and contributes toward understanding the subcellular and microenvironmental chemical conditions that are required for cancer progression.

2 METHODS

2.1 Data sources

This section briefly describes the data sources: literature reference, cell or tissue type, experimental conditions, analytical methods (e.g. proteomic analysis using label-free MS or iTRAQ label techniques), statistical analysis of highly differentially abundant proteins in the samples, and additional data selection steps applied in this study. The heading of each subsection is used in the tables and plot titles to identify the different data sets; the words separated by a slash refer to the proteins with higher differential expression in "cancer" (or more advanced cancer stage) / "control" (healthy or less advanced cancer stage) groups.

Names or IDs of genes or proteins given in the papers were searched in UniProt, and corresponding UniProt accessions are provided in the Supporting Information. Amino acid sequences of human proteins were taken from the UniProt reference proteome (file UP000005640_9606.fasta.gz, dated 2015-04-29, downloaded from ftp.uniprot.org on 2015-05-20). Entire sequences were used; i.e., signal and propeptides were not removed when calculating the amino acid compositions.

2.1.1 Chromatin-binding in carcinoma/adenoma

In the study of Knol et al. (2014), a combined laboratory/bioinformatic approach was used to identify chromatin-bound (CB) proteins in colon adenomas and carcinomas. Differential biochemical extraction was used to isolate the chromatin-binding fraction, which was then analyzed using mass spectrometry (LC-MS/MS) for protein identification and label-free quantification. Sequences were compared with seven different databases to assign gene annotations. Differentially abundant proteins were identified using beta-binomial statistics (significance threshold 0.05); of these, sequences having three or more database hits to a nuclear annotation were kept in the final list. Of the 169 proteins listed by Knol et al. (2014), the 27 found only in adenomas, and the 26 found only in carcinomas, are used here for comparison (see file `KWA+14.csv` in the Supporting Information).

2.1.2 Membrane fraction tumor/normal

In the study of Chen et al. (2010), solvent extraction and gel-assisted digestion were used to prepare proteins from the membrane fraction of samples of eight tumor tissues from CRC patients and eight matched normal tissues. Based on proteomic analysis using iTRAQ LC-MS/MS, 42 proteins exhibited at least two-fold differential expression (34 upregulated in tumor samples, 8 down-regulated) at a 95% confidence level (Student's t-test $p < 0.05$). Data including UniProt accessions of these proteins were taken from Supplementary Table S2 of Chen et al. (2010) and stored in file `CCF+10.csv` in the Supporting Information (this study).

2.1.3 Membrane fraction tumor/polyp

Kume et al. (2014) used phase-transfer surfactant (PTS) to extract proteins from the membrane fractions of six matched tissue samples of polyps and cancer with or without metastasis. iTRAQ labeling and LC-MS/MS were used for proteomic analysis and initial assessment of expression level differences, and SRM/MRM was used to better quantify the expression levels of biomarker candidates. At an expression ratio <0.5 or

>2.0 and p-value <0.1, 66 proteins were found to increase, and 13 found to decrease, between polyps and cancer without metastasis. Data including UniProt accessions of these proteins were taken from Table IIIA and IIIC of Kume et al. (2014) and stored in file `KMK+14.csv` in the Supporting Information. Only proteins recorded as having GO cellular components "membrane" and not "extracellular" were used in this study (45 higher in cancer / 6 higher in polyps).

2.1.4 Membrane enriched tumor/normal

In the study of Sethi et al. (2015), samples of tumor and adjacent normal tissues from eight patients were characterized by label-free nanoLC-MS/MS of the membrane-enriched proteome. 184 proteins were differentially expressed with a fold change > 1.5 and p-value < 0.05. Protein identifiers and fold changes were taken from Supporting Table 2 of Sethi et al. (2015) (69 up-regulated and 115 down-regulated in tumors). The identifiers were manually searched in UniProt and the corresponding accessions stored in file `STK+15.csv` in the Supporting Information.

2.1.5 Epithelial cell signature adenoma/normal

Uzozie et al. (2014) analyzed 30 samples of colorectal adenomas and paired normal mucosa using iTRAQ labeling, OFFGEL electrophoresis and LC-MS/MS. Proteins with expression fold changes ($\log_2$) at least +/- 0.5 and statistical significance threshold $q <$ 0.02 were then compared with protein expression in colorectal cancer cells and HCEC cells (normal control). The proteins that were also quantified in cell-line experiments were classified as "epithelial cell signature proteins" (111), with the remainder (101) interpreted as having a stromal origin. Information on the proteins, including UniProt accessions, was taken from Table III of Uzozie et al. (2014) and stored in file `UNS+14.csv` in the Supporting Information. Because some proteins have multiple UniProt accessions in the source table, the actual number of epithelial proteins here is 121, with 64 higher in adenoma and 57 higher in normal tissue.

2.1.6 Stromal cell signature adenoma/normal

This is the set of non-epithelial proteins taken from Table III of Uzozie et al. (2014) (UniProt accessions: 18 higher in adenoma, 108 higher in normal tissue) and stored in file `UNS+14.csv` in the Supporting Information.

2.1.7 Wnt targets tumor/invasion front

Hlubek et al. (2007) identified Wnt target genes that are up- and down-expressed in cells in the interior and invasion front of human colon tumors. Here, proteins for the 27 genes reported by Hlubek et al. (2007) to be significantly up-regulated in the interior (27) or invasion front (9) of tumors are used in the comparisons.

2.1.8 Serum biomarkers abundance up/down

Jimenez et al. (2010) compiled a list of candidate serum biomarkers from a meta-analysis of the literature. In the meta-analysis, both up- and down-expressed proteins were associated with some number of studies; the data set used in this study is based on those proteins identified in at least 3 studies (up-expressed: 26 proteins; down-expressed: 19 proteins).

2.2 Average oxidation state of carbon; Basis set I

For a protein with formula $C_cH_hN_nO_oS_s$, the average oxidation state of carbon (Z_C) is

$$Z_C = \frac{3n + 2o + 2s - h}{c} \tag{1}$$

This calculation permits a simple comparison of the relative degrees of oxidation of protein molecules (Dick, 2014).

To proceed to more sophisticated chemical thermodynamic calculations, an important choice must be made regarding the *basis species* used to describe the system. The basis species, like thermodynamic components, are a minimum number of chemical formula units that can be linearly combined to generate the composition of any chemical species in the system of interest. Within these constraints, any specific choice of a basis set is permitted in theory. In making the choice of components, convenience (Gibbs, 1875), ease of interpretation and relationship with measurable variables, as well as availability of thermodynamic data (e.g. Helgeson, 1970), and kinetic favorability (May et al., 2001) are other useful considerations. Once the basis species are chosen, the stoichiometric coefficients in the formation reaction for any species are determined.

Following previous studies (e.g. Dick, 2008), the basis species initially chosen here are CO_2, H_2O, NH_3, H_2S and O_2 (basis set I). For a protein with formula $C_cH_hN_nO_oS_s$, the reaction representing the overall formation of the protein from these basis species is

$$c\mathrm{CO_2} + n\mathrm{NH_3} + s\mathrm{H_2S} + n_{\mathrm{H_2O}}\mathrm{H_2O} + n_{\mathrm{O_2}}\mathrm{O_2} \rightleftharpoons \mathrm{C}_c\mathrm{H}_h\mathrm{N}_n\mathrm{O}_o\mathrm{S}_s \tag{R1}$$

where $n_{H_2O} = (h - 3n - 2s)/2$ and $n_{O_2} = (o - 2c - n_{H_2O})/2$. Dividing n_{H_2O} by the length of the protein gives the water demand per residue ($\overline{n}_{H_2O}$), which is useful for comparing proteins of different size.

These or similar sets of inorganic species (such as H_2 instead of O_2) are often used in studying reaction energetics in geobiochemistry (e.g. Shock and Canovas, 2010). However, because of the stoichiometric *interdependence* that links redox and hydration potentials, it is possible that a different set of basis species is more appropriate for convenient description of subcellular processes.

2.3 Basis set II; Thermodynamic calculations

As seen in Fig. 1a and b, there is a tight correlation between Z_C of protein molecules and n_{H_2O} in the reactions to form the proteins from basis set I. This correlation is expected based on stoichiometric requirements. The equation for n_{H_2O} (after Reaction R1) can be combined with Eq. (1) to write

$$Z_C = \frac{2}{c}(o - n_{\mathrm{H_2O}}) \tag{2}$$

For given c and o, a linear relationship holds between Z_C and n_{H_2O}, but for a population of proteins, variations in the chemical composition cause a dispersion around a linear trend, which is apparent in Fig. 1.

The appearance of this correlation raises a troubling issue. Although the basis species, like thermodynamic components, are necessarily independent compositional variables, in the systems of proteins considered here basis set I leads to a strong

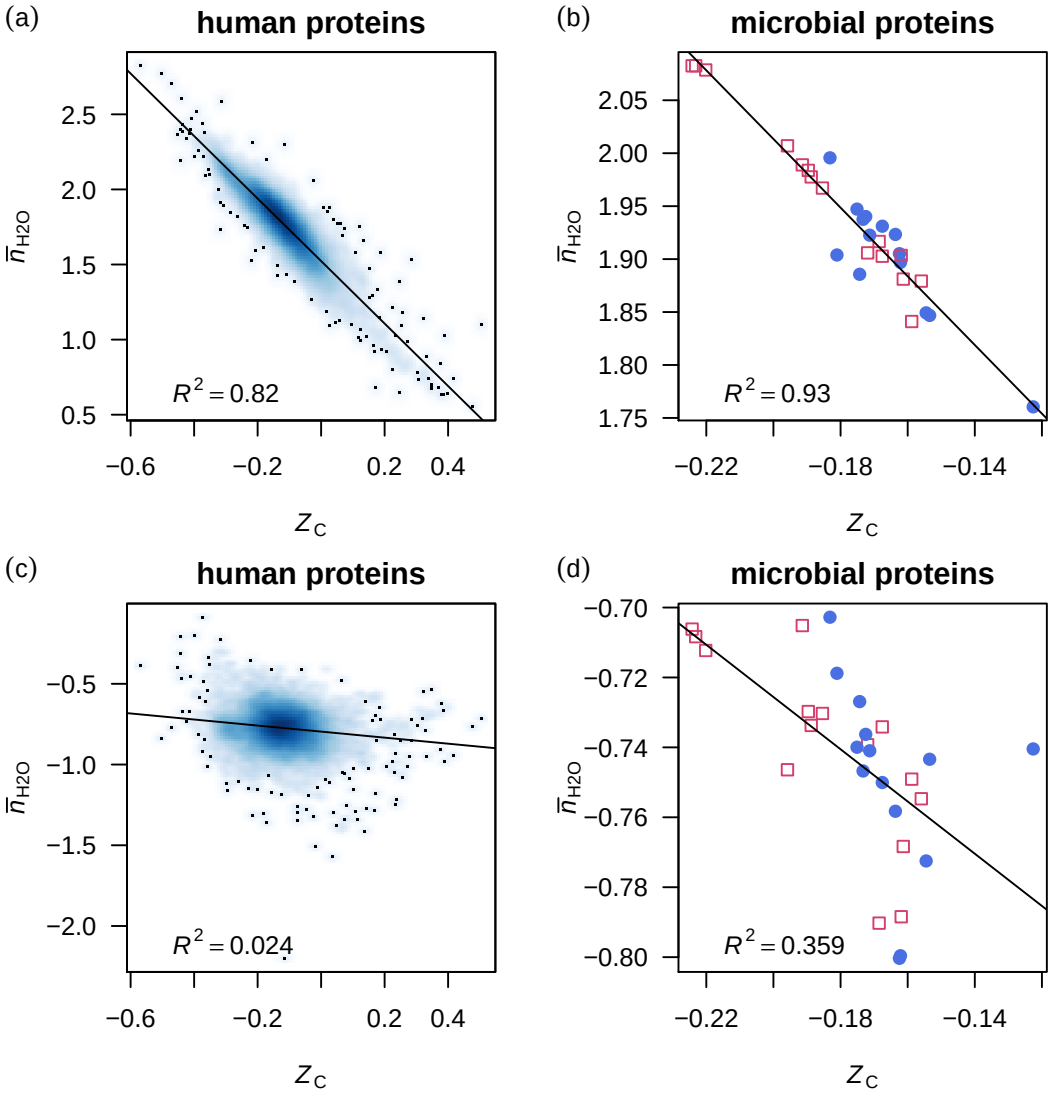

Figure 1. Scatterplots of average oxidation state of carbon (Z_C) and water demand per residue ($\bar{n}_{H_2O}$) for (a,c) individual human proteins and (b,d) overall (mean) composition of proteins from microbial genomes, using (a,b) basis set I (e.g. Reaction R1) or (c,d) basis set II (e.g. Reaction R2). Linear least-squares fits and R^2 values are shown. In (a) and (c), the intensity of shading corresponds to density of points, produced using the `smoothScatter()` function of R graphics (R Core Team, 2015).

dependence between the two variables we are primarily interested in. Intuition suggests that some combinations of basis species will alter, and in some cases, decrease, the correlation between individual compositional variables such as Z_C and n_{H_2O}.

If we "knew" the conditions in the cellular subsystem(s) of interest, any choice of basis species would be equally convenient, because we could simultaneously vary all of their chemical potentials appropriately. However, since we must begin in a more exploratory mode, we restrict attention here to one or two variables, with the implication that the others are held constant. In a subcellular setting, assuming constant values of chemical potentials of species such as CO_2, NH_3 and H_2S may be less

appropriate than postulating constant (or "buffered") potentials of more complex metabolites. Perhaps, in modeling systems of proteins, constant chemical activities of amino acids would be a reasonable starting assumption.

Although 1140 3-way combinations can be made of the 20 common proteinogenic amino acids, only 324 of the combinations contain cysteine and/or methionine (one of these is required to provide sulfur), and of these only 300, when combined with O_2 and H_2O, are compositionally independent. The slope, intercept and R^2 of the least-squares linear fits between Z_C and $\overline{n}_{H_2O}$ using each of these basis sets are listed in the Supporting Information (`AAbasis.csv`). Many of these sets have lower R^2 and lower slopes than found for basis set I (Fig. 1a, b), indicating a decreased correlation, which is desirable for our purposes. From those with a lower correlation (but not the lowest), the basis set including cysteine, glutamic acid, glutamine and O_2 and H_2O (basis set II) has been selected for use in this study. The scatter plots and fits between Z_C and $\overline{n}_{H_2O}$ using basis set II are shown in Fig. 1c and d.

A secondary consideration in choosing this set (rather than others with even lower R^2) is based on the centrality of glutamine and glutamic acid in many metabolic pathways (e.g. DeBerardinis and Cheng, 2010). Accordingly, these amino acids may be kinetically more reactive than others in mechanisms underlying protein synthesis and degradation. The presence of side chains derived from cysteine and glutamic acid in the abundant glutathione molecule (GSH), associated with redox homeostasis, is also suggestive of a central metabolic requirement for these amino acids. Again, it must be stressed that the current choice of basis species is not an absolute, uniquely determined result; greater experience with thermodynamic modeling and better biochemical intuition will likely provide reasons to iterate these calculations using different basis sets (perhaps including metabolites other than amino acids).

A general formation reaction using basis set II is

$$n_{Cys}C_3H_7NO_2S + n_{Glu}C_5H_9NO_4 + n_{Gln}C_5H_{10}N_2O_3 + n_{H_2O}H_2O + n_{O_2}O_2 \rightarrow C_cH_hN_nO_oS_s \quad \text{(R2)}$$

where the reaction coefficients (n_{Cys}, n_{Glu}, n_{Gln}, n_{H_2O} and n_{O_2}) can be obtained by solving

$$\begin{bmatrix} 3 & 5 & 5 & 0 & 0 \\ 7 & 9 & 10 & 2 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 2 & 4 & 3 & 1 & 2 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} n_{Cys} \\ n_{Glu} \\ n_{Gln} \\ n_{H_2O} \\ n_{O_2} \end{bmatrix} = \begin{bmatrix} c \\ h \\ n \\ o \\ s \end{bmatrix} \quad (3)$$

Although the definition of basis species requires that they are themselves stoichiometrically non-degenerate, the matrix equation emphasizes the interdependence of the reaction coefficients with each other. A consequence of this multiple dependence is that *single* variables such as n_{H_2O} are not *simple* variables, but are influenced by both the inherent chemical makeup of the protein and the choice of basis species used to describe the system.

Regarding the actual mechanism of synthesis, we are definitely not saying that proteins are synthesized by combining the specific molecules shown in Reaction R2;

reactions such as this are simply a way to account for mass-conservation requirements. This allows us to generate an internally consistent thermodynamic description of the effects of changing the local environment (i.e. chemical potentials of O_2 and H_2O) on the overall potential for formation of different proteins.

As an example of a specific calculation, consider the following reaction:

$$7C_3H_7NO_2S + 535.6C_5H_9NO_4 + 515.2C_5H_{10}N_2O_3 \rightarrow C_{5275}H_{8231}N_{1573}O_{1762}S_7 + 895.2H_2O + 522.4O_2 \quad \text{(R3)}$$

Here, the the reaction is written for the formation of one mole of the protein MUC1, a chromatin-binding protein that is highly up-expressed in CRC cells (Knol et al., 2014). Using Eq. (1), the average oxidation state of carbon (Z_C) in this protein is 0.005. According to this reaction, water is released in the overall formation of the protein, so the *water demand* (n_{H_2O}) is negative. The length of this protein is 1255 amino acid residues, giving the water demand per residue, $\overline{n}_{H_2O} = -895.2/1255 = 0.71$. The value of Z_C indicates that MUC1 is a relatively highly oxidized protein, while its $\overline{n}_{H_2O}$ places it near the median water demand for cancer-associated proteins in this data set (see Table 1 and Fig. 2 below).

2.4 Thermodynamic calculations

Thermodynamic properties for the amino acids and of unfolded proteins estimated using amino acid group additivity were calculated as described by Dick et al. (2006), taking account of updated values for the methionine sidechain group (LaRowe and Dick, 2012). All calculations were carried out at 37 °C and 1 bar. The temperature dependence of standard Gibbs energies was calculated using the revised Helgeson-Kirkham-Flowers (HKF) equations of state (Helgeson et al., 1981; Tanger and Helgeson, 1988). Thermodynamic properties for O_2 (gas) were calculated using data from Wagman et al. (1982) together with coefficients for the Maier-Kelley heat capacity function (Kelley, 1960), and those for H_2O (liquid) using data and extrapolations coded in Fortran subroutines from the SUPCRT92 package (Johnson et al., 1992).

Chemical affinities of reactions were calculated using activities of amino acids set to 10^{-4}, and activities of proteins equal to 1/(protein length) (i.e., unit activity of residues). Continuing with the example of Reaction R3, an estimate of the standard Gibbs energy (ΔG_f°) of the protein (Dick et al., 2006) at 37 °C is -40974 kcal/mol; combined with the standard Gibbs energies of the basis species, this give a standard Gibbs energy of reaction (ΔG_r°) equal to 66889 kcal/mol. At $\log a_{H_2O} = 0$ and $\log f_{O_2} = -65$, with activities of the amino acid basis species equal to 10^{-4}, the overall Gibbs energy (ΔG_r) is 24701 kcal/mol. The negative of this value is the chemical affinity (A) of the reaction. The per-residue chemical affinity (used in order to compare the relative stabilities of proteins of different sizes) for formation of protein MUC1 in the stated conditions is -19.7 kcal/mol.

The amino acid compositions of different proteins would yield different values of the per-residue chemical affinity. The chemical affinities of formation of all proteins are also sensitive to the environmental conditions represented by temperature (T), pressure (P) and the chemical potentials of basis species. Without making any statement about mechanism, proteins with higher (more positive) chemical affinity are relatively

energetically stabilized, and theoretically have a higher propensity to be formed. Therefore, the differences in affinities reflect not only the oxidation and hydration state of the protein molecules but also the potential for local environmental conditions to constrain the degree of formation and relative abundances of proteins. Here, the relative stabilities as a function of $\log a_{H_2O}$ and $\log f_{O_2}$ are mapped on to diagrams using a rank-difference summary of the group-wise (i.e. cancer- vs healthy-associated proteins) chemical affinities.

2.5 Rank calculations

Consider a hypothetical system composed of 4 cancer (C) and 5 healthy (H) proteins. Suppose that under one set of conditions (i.e. specified $\log a_{H_2O}$ and $\log f_{O_2}$), the per-residue affinities of the proteins give the following ranking in ascending order (I):

C	C	C	C	H	H	H	H	H
1	2	3	4	5	6	7	8	9

This gives as the sum of ranks for cancer proteins $\sum r_C = 10$, and for healthy proteins $\sum r_H = 35$. The difference in sum of ranks then is $\Delta r_{C-H} = -25$; the negative value is associated with a higher rank sum for the healthy proteins, indicating that these as a group are more stable than the cancer proteins. In a second set of conditions, we might have (II):

H	H	H	H	H	C	C	C	C
1	2	3	4	5	6	7	8	9

Here, the difference in rank sums is $\Delta r_{C-H} = 30 - 15 = 15$.

For systems where the numbers of cancer and healthy proteins are equal, the maximum possible differences in rank sums would have equal absolute values, but that is not the case in this and other systems having unequal numbers of cancer and healthy proteins. To characterize those data sets, the weighted rank-sum difference can be calculated using

$$\Delta\bar{r} = 2\left(\frac{n_H}{n}\sum r_C - \frac{n_C}{n}\sum r_H\right) \tag{4}$$

where n_H, n_C and n are the numbers of healthy, cancer, and total proteins in the comparison. In the example here, we have $n_H/n = 5/9$ and $n_C/n = 4/9$. Eq. (4) then gives $\Delta\bar{r} = -20$ and $\Delta\bar{r} = 20$, respectively, for conditions (I) and (II) above, showing equal weighted rank-sum differences for the two extreme rankings.

We can also consider a situation where the ranks of the cancer and healthy proteins are evenly distributed:

H	C	H	C	H	C	H	C	H
1	2	3	4	5	6	7	8	9

Here the absolute difference in rank sums is $\Delta r_{C-H} = 5$, but the weighted rank-sum difference is $\Delta\bar{r} = 0$. The zero value for an even distribution, and the opposite values for the two extremes, demonstrate the acceptability of this weighting scheme.

3 RESULTS

3.1 Compositional descriptions of human proteins

For proteins in the selected data sets (Section 2.1), Fig. 2 shows values of *average oxidation state of carbon* calculated from the chemical formulas of the proteins (Z_C; Eq. 1) and *water demand per residue* calculated from the overall formation reaction in terms of basis set II ($\bar{n}_{H_2O}$; Reaction R2).

Fig. 2a reveals that Z_C of chromatin-binding proteins in carcinoma reported by Knol et al. (2014) are shifted to lower (more negative) values compared to proteins in adenoma, except for three outliers at high Z_C (MUC1, LRP1 and PTGES3). A moderate shift toward higher $\bar{n}_{H_2O}$ is exhibited by chromatin-binding proteins in carcinoma. The Z_C of proteins in membrane fractions of tumor samples identified by Chen et al. (2010) are lower than paired normal samples (Fig. 2b), and the Z_C of proteins in membrane fractions of tumor samples identified by Kume et al. (2014) are lower than paired polyp samples (Fig. 2c). These observations are summarized in Table 1, giving the median values of Z_C and $\bar{n}_{H_2O}$ and p-values calculated using the Mann-Whitney test (a non-parametric test that does not assume a particular distribution). Although Z_C is consistently lower for tumor and carcinoma-derived proteins in these data sets, the lower degree of statistical significance (represented by the p-values) precludes making a definitive statement about relative water demand for these proteins.

In the next row of Fig. 2, compositions of membrane-enriched proteins in the study of Sethi et al. (2015) are seen to be offset toward higher $\bar{n}_{H_2O}$ for tumor relative to normal samples (Fig. 2d). The adenoma-derived proteins reported by Uzozie et al. (2014) as identified with "epithelial cell signature" have a strongly elevated relative water demand per residue compared to proteins from normal tissue, while adenoma-derived proteins more likely to have a stromal origin show a somewhat smaller increase in $\bar{n}_{H_2O}$ relative to normal tissue (Fig. 2e,f).

Overall, the compositions of the proteins in carcinoma and tumor-derived proteins relative to their adenoma or normal counterparts manifest either a strongly lower median Z_C with variable sign of the median $\bar{n}_{H_2O}$ difference, or a higher median $\bar{n}_{H_2O}$ with smaller Z_C difference (Fig 2i). These trends are also apparent, although with lower statistical significance, for proteins of Wnt target genes in the tumor center compared to the invasion front (lower Z_C, Fig. 2g) and for candidate serum biomarkers with increased relative abundance compared to those with decreased relative abundance in CRC patients (higher $\bar{n}_{H_2O}$, Fig. 2h). The possible significance of these findings in relation to biochemical and physiological observations of tumor oxidation and hydration state is discussed further below.

3.2 Compositional descriptions of microbial proteins

Healthy individuals maintain a thriving enteric microbial population. The microbiota support important functions including digestion (Turnbaugh et al., 2009) and immunological regulation and inflammatory responses, which are strongly tied to oxidation-reduction conditions (Koboziev et al., 2014). The healthy gut can be characterized by a "core microbiome" composed of a common set of genes (or functions) that are present regardless of differences in taxonomic composition (Turnbaugh et al., 2009). Because of the diverse responses of microbial populations to

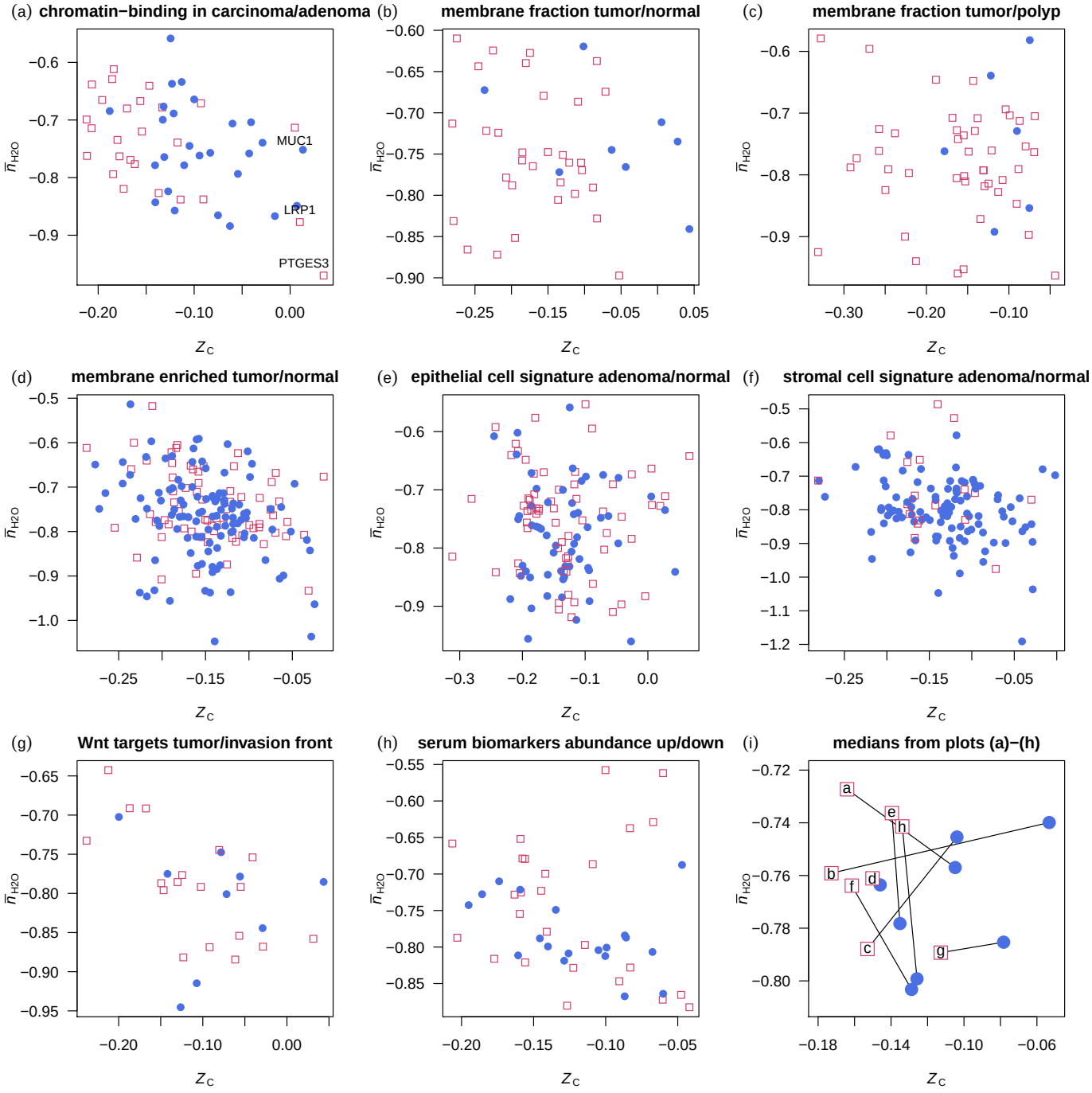

Figure 2. Average oxidation state of carbon (Z_C) and water demand per residue ($\bar{n}_{H_2O}$) for proteins differentially expressed in cancerous tissues or samples (see Sections 2.1.1–2.1.8 for data sources and Table 1 for summary statistics of the comparisons). Open red squares represent the tumor-associated proteins or more advanced cancer stages, and filled blue circles represent proteins associated with normal tissue or less advanced cancer stages. Plot (i) shows the median values in each of the data sets.

Table 1. Summary of compositional comparisons: median values of Z_C and $\overline{n}_{H_2O}$ in human proteins and overall (mean) protein compositions of microbes, and Wilcoxon rank sum test (Mann-Whitney) statistics and p-values. Depending on the data sets (see footnotes and Section 2.1), "control" refers to normal tissue or less advanced cancer stage (e.g. adenoma), and "cancer" refers to cancer tissue or more advanced cancer stage (e.g. carcinoma).

	Z_C medians			$\overline{n}_{H_2O}$ medians		
comparison	control	cancer	p-value	control	cancer	p-value
Human proteins						
Chromatin-binding in carcinoma/adenoma [a]	-0.105	-0.164	0.0002	-0.757	-0.727	0.5
Membrane fraction tumor/normal [b]	-0.053	-0.173	0.004	-0.740	-0.759	0.5
Membrane fraction tumor/polyp [c]	-0.104	-0.153	0.05	-0.745	-0.788	0.5
Membrane enriched tumor/normal [d]	-0.146	-0.150	0.7	-0.764	-0.761	0.2
Epithelial cell signature adenoma/normal [e]	-0.135	-0.140	0.9	-0.778	-0.736	0.06
Stromal cell signature adenoma/normal [e]	-0.129	-0.162	0.4	-0.803	-0.764	0.09
Wnt targets tumor/invasion front [f]	-0.078	-0.113	0.5	-0.785	-0.789	0.7
Serum biomarkers abundance up/down [g]	-0.126	-0.134	0.8	-0.799	-0.741	0.3
Microbial proteins						
fecal 16S rRNA [h]	-0.163	-0.177	0.3	-0.765	-0.741	0.2
fecal metagenome [i]	-0.174	-0.190	0.6	-0.738	-0.730	0.5
co-abundance groups [j]	-0.138	-0.196	0.2	-0.742	-0.746	0.8

a. Knol et al. (2014); **b**. Chen et al. (2010); **c**. Kume et al. (2014); **d**. Sethi et al. (2015); **e**. Uzozie et al. (2014); **f**. Hlubek et al. (2007); **g**. Jimenez et al. (2010); **h**. Wang et al. (2012); **i**. Zeller et al. (2014); **j**. Candela et al. (2014).

the cancer state (Sears and Garrett, 2014), summary data on microbial populations from three studies were selected for inclusion in the descriptions of chemical compositions generated in this study. First, in a study of 16S RNA of fecal microbiota, Wang et al. (2012) reported genera that are significantly increased and decreased in CRC compared to healthy patients. In order to compare the chemical composition of the microbial population, single species with sequenced genomes were chosen to represent each of these genera (see Table 2). Where possible, the species selected are those mentioned by Wang et al. (2012) as being significantly altered, or are species reported in other studies to be present in healthy or cancer states (see Table 2).

In the second study considered (Zeller et al., 2014), changes in the metagenomic abundance of fecal microbiota associated with CRC were analyzed for their potential as a biosignature for cancer detection. The species shown in Fig. 1A of Zeller et al. (2014) with a log odds ratio greater than 0.15 were selected for comparison, and are listed in Table 3. Here, *Fusobacterium* is identified as a pro-carcinogenic group, as has also been reported elsewhere (Kostic et al., 2012; Castellarin et al., 2012). Finally, Candela et al. (2014) report the findings of a network analysis that identified 5 microbial "co-abundance groups" at the genus level. As before, single representative species were selected in this study, and are listed in Table 2. Except for the presence of *Fusobacterium*, the co-abundance groups show little genus-level overlap with profiles

Table 2. Microbial species selected as models for genera and co-abundance groups that differ between CRC and healthy patients.

Phylum	Species	Abbrv.	Bioproject	Refs.
Model species for genera significantly higher in healthy patients[a]				
Bacteroidetes	*Bacteroides vulgatus* ATCC 8482	Bvu	PRJNA13378	**c**
Bacteroidetes	*Bacteroides uniformis* ATCC 8492	Bun	PRJNA18195	**c**
Firmicutes	*Roseburia intestinalis* L1-82 (DSM 14610)	Rin	PRJNA30005	**d**
Bacteroidetes	*Alistipes indistinctus* YIT 12060	Ain	PRJNA46373	**c**
Firmicutes	*Eubacterium rectale* ATCC 33656	Ere	PRJNA29071	**e**
Proteobacteria	*Parasutterella excrementihominis* YIT 11859	Pex	PRJNA48497	**f**
Model species for genera significantly higher in CRC patients[a]				
Bacteroidetes	*Porphyromonas gingivalis* W83	Pgi	PRJNA48	**g**
Proteobacteria	*Escherichia coli* NC101	Eco	PRJNA47121	**c**,**h**
Firmicutes	*Enterococcus faecalis* V583	Efa	PRJNA57669	**c**
Firmicutes	*Streptococcus infantarius* ATCC BAA-102	Sin	PRJNA20527	**i**
Firmicutes	*Peptostreptococcus stomatis* DSM 17678	Pst	PRJNA34073	**j**
Bacteroidetes	*Bacteroides fragilis* YCH46	Bfr	PRJNA58195	**g**
Model species for protective co-abundance groups[b]				
Actinobacteria	*Bifidobacterium longum* NCC2705	Blo	PRJNA57939	**g**,**k**
Firmicutes	*Faecalibacterium prausnitzii* SL3/3	Fpr	PRJNA39151	**e**,**l**
Model species for pro-carcinogenic co-abundance groups[b]				
Fusobacteria	*Fusobacterium nucleatum* ATCC 23726	Fnu	PRJNA49043	**m**,**n**
Bacteroidetes	*Prevotella copri* DSM 18205	Pco	PRJNA30025	**k**,**o**
Firmicutes	*Coprobacillus* sp. D7	Csp	PRJNA32495	**h**

a. Genus identification from Table 2 of Wang et al. (2012). Based on comments in Wang et al. (2012), *Bacteroides* is represented here by two species (*B. vulgatus* and *B. uniformis*) in healthy patients, and one species (*B. fragilis*) in CRC patients. **b**. Genus-level definition of co-abundance groups from Candela et al. (2014). **c**. Wang et al. (2012); species closely related to 16S rRNA-derived operational taxonomic units (OTUs; Figure 2 of Wang et al., 2012) or otherwise mentioned by those authors (*E. faecalis*). **d**. Duncan et al. (2002). **e**. Louis and Flint (2007). **f**. Nagai et al. (2009). **g**. Chen et al. (2012). **h**. Candela et al. (2014). **i**. Biarc et al. (2004). **j**. Zeller et al. (2014). **k**. Weir et al. (2013). **l**. Sokol et al. (2008). **m**. Castellarin et al. (2012). **n**. Kostic et al. (2012). **o**. cf. Chen et al. (2012); Candela et al. (2014) (more abundant in CRC patients); Weir et al. (2013) (more abundant in healthy subjects).

derived from the previous two studies.

For each of the microbial species listed in Tables 2 and 3, an overall protein composition was calculated from the NCBI Bioproject genomes as average compositions of protein sequences, without any weighting for transcript or actual protein abundance in organisms, and exclude any post-translational modifications. The water demand per residue ($\overline{n}_{H_2O}$) vs. oxidation state of carbon (Z_C) in proteins from all of the microbial species considered here are plotted in Fig. 1b and d.

The plots of $\overline{n}_{H_2O}$ vs. Z_C for overall microbial proteins in each data set are shown in Fig. 3. It is apparent that the proteins in the microbes from cancer patients generally have lower Z_C than the healthy patients in the same study (see also Table 1). The two *Fusobacterium* species identified by Zeller et al. (2014) have the lowest Z_C of any microbial species considered here. The overall human protein composition is also plotted in Fig. 3, showing higher Z_C than any of the microbial proteomes except for

Table 3. Species from a consensus microbial signature for CRC classification of fecal metagenomes (Zeller et al., 2014). Only species reported as having a log odds ratio bigger than ± 0.15 are listed here, together with strains and Bioproject ID's used as models in the present study.

Species	Strain	Abbrv.	Bioproject
Higher in CRC patients			
Fusobacterium nucleatum subsp. vincentii	ATCC 49256	Fnv	PRJNA1419
Fusobacterium nucleatum subsp. animalis	D11	Fna	PRJNA32501
Peptostreptococcus stomatis	DSM 17678	Pst	PRJNA34073
Porphyromonas asaccharolytica	DSM 20707	Pas	PRJNA51745
Clostridium symbiosum	ATCC 14940	Csy	PRJNA18183
Clostridium hylemonae	DSM 15053	Chy	PRJNA30369
Lactobacillus salivarius	ATCC 11741	Lsa	PRJNA31503
Higher in healthy patients			
Clostridium scindens	ATCC 35704	Csc	PRJNA18175
Eubacterium eligens	ATCC 27750	Eel	PRJNA29073
Methanosphaera stadtmanae	DSM 3091	Mst	PRJNA15579
Phascolarctobacterium succinatutens	YIT 12067	Psu	PRJNA48505
unclassified *Ruminococcus* sp.	ATCC 29149(*)	Rsp	PRJNA18179
Streptococcus salivarius	SK126	Ssa	PRJNA34091

* *R. gnavus*

Bifidobacterium longum.

The proteins from cancer-associated bacteria do not have a consistently different median $\overline{n}_{H_2O}$ than those from healthy-associated bacteria, but the comparisons show that the overall changes in microbial populations associated with CRC favor species with proteins that are more reduced than the healthy populations.

3.3 Thermodynamic descriptions: background

Many workers have investigated the oxidation-reduction (redox) state of cells and biomolecules and redox-dependent signaling processes in both healthy and cancerous cells. A simplified categorization of redox molecules and processes might include the following. 1) Reactive oxygen species, earlier thought to be generally harmful in cells, are now considered to be important signaling molecules (e.g. Sarsour et al., 2009). 2) Low-molecular weight thiols are also involved in molecular signaling, and have been proposed as key components of cell-cycle progression (Menon et al., 2003). The concentrations of reduced thiols and oxidized disulfides, especially of reduced and oxidized glutathione, are often used to quantify redox potential (Eh). 3) Oxygen itself is a major contributor to redox reactions. Tumor hypoxia is common in cancer (Höckel and Vaupel, 2001), and regulatory response to oxygen is important for oncogenesis (Harris, 2002). 4) The covalent structure of organic molecules determines the oxidation state of carbon; differences in molecular composition imply that energetics of biosynthetic processes generally have a strong dependence on changing redox potential. The different oxidation states of metabolites including lactate and pyruvate are essential in accounting for their utilization in metabolic processes (e.g. Brooks, 2009). The

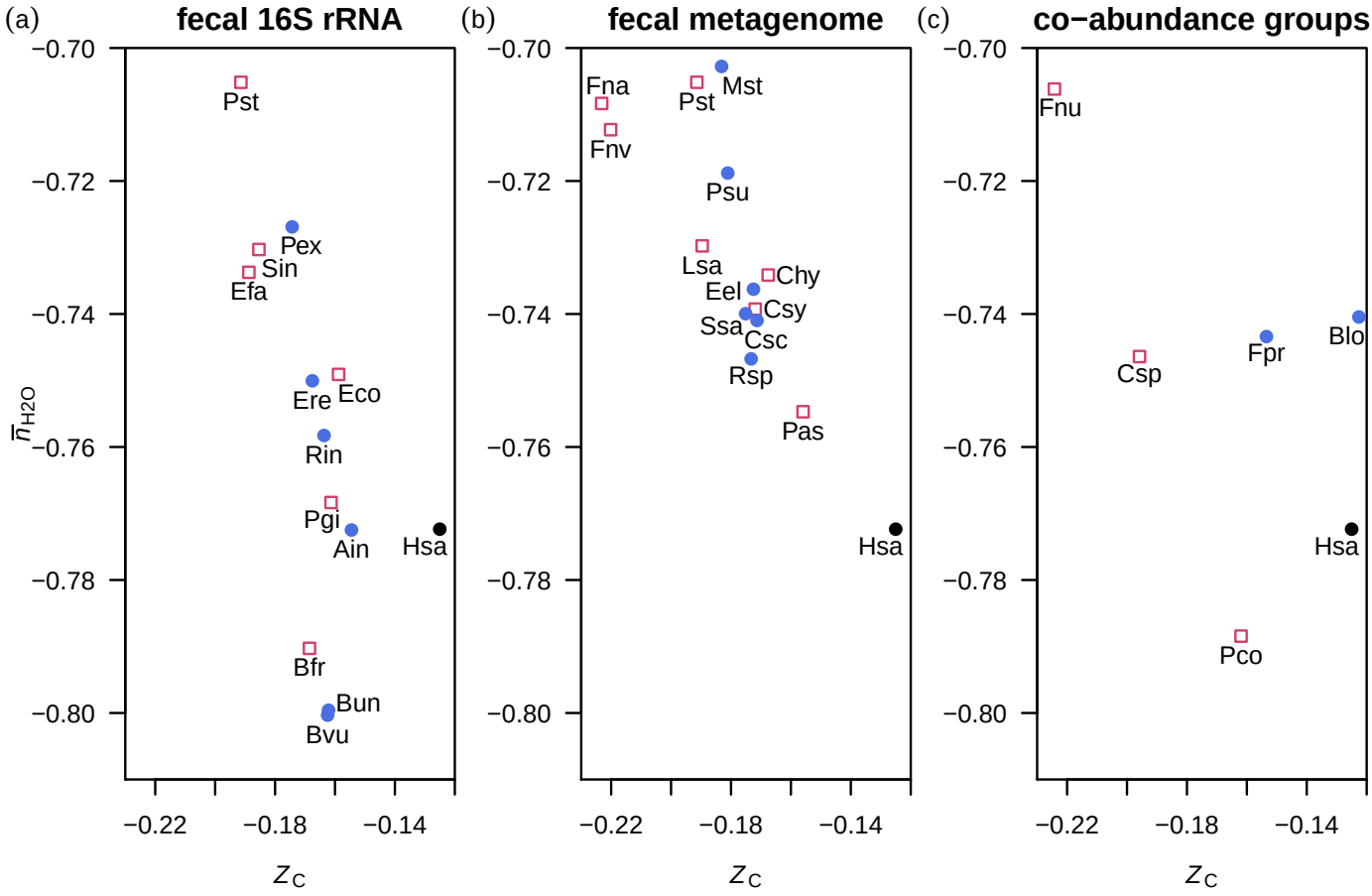

Figure 3. Average oxidation state of carbon (Z_C) and water demand per residue ($\bar{n}_{H_2O}$) for overall amino acid compositions of proteins in genomes of healthy- and cancer-associated microbes. Data are shown for representative species for (a) microbial genera identified in fecal 16s RNA (Wang et al., 2012; Table 2 top), (b) microbial signatures in fecal metagenomes (Zeller et al., 2014; Table 3), and (c) microbial co-abundance groups (Candela et al., 2014; Table 2 bottom).

subcellular distribution of proteins with different oxidation states of carbon can also be used to suggest that the synthesis and degradation of all proteins in cells depend on and influence subcellular redox conditions (Dick, 2014).

Higher chemical affinities correspond with greater relative stability, so the relative stabilities of the proteins are a function of both amino acid composition and environmental conditions represented by the activities of the basis species, including not only O_2 and H_2O but the others. When the highest chemical affinities are mapped onto 2-dimensional spaces (here, $\log f_{O_2}$ and $\log a_{H_2O}$), a diagram with fields results, which may be called a predominance diagram, referring to the limiting equilibrium case, or more generally, a maximal relative stability diagram (or, more simply, stability diagram). However, even in actual situations in which equilibrium is not achieved, the calculations add a layer of information about relative energetics of reactions. The environmentally dependent reaction energetics support a richer interpretation than comparison of chemical compositions alone.

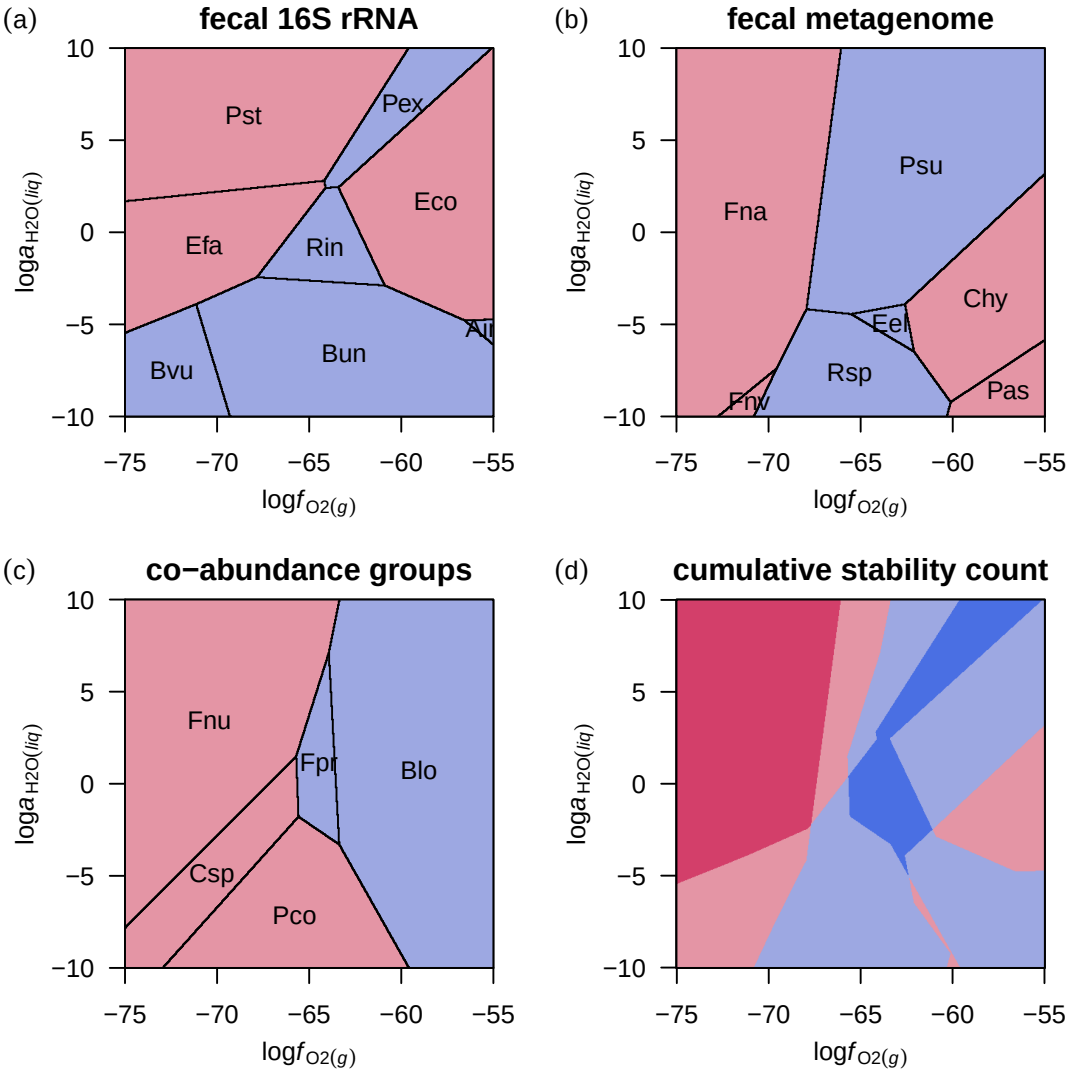

Figure 4. Maximal relative stability diagrams for overall microbial protein compositions, showing the range of oxygen fugacity and water activity (in log units: $\log f_{O_2}$ and $\log a_{H_2O}$) where the protein composition from the labeled species has a higher affinity (lower Gibbs energy) of formation than the others. Blue and red shading designate microbes associated with healthy and cancer samples, respectively. Plot (d) is a composite figure in which the intensity of shading corresponds to the number of overlapping cancer- or healthy-related proteins in the preceding diagrams.

3.4 Relative stability fields for microbial populations

Stability diagrams are shown in Fig. 4a–c for the three sets of microbial proteins described above. The first diagram, representing significantly changed genera detected in fecal 16S rRNA (Wang et al., 2012; first part of Table 2), shows maximal stability fields for proteins from 5 species associated with healthy patients, and 3 species associated with CRC patients, in the range of $\log f_{O_2}$ and $\log a_{H_2O}$ shown. The other 4 proteins in the system are not more stable than the others over this range of conditions.

The relative positions of the stability fields in Fig. 4a are roughly aligned with the

values of Z_C and $\bar{n}_{H_2O}$ of the proteins; note for example the high-log f_{O_2} positions of the fields for the relatively high-Z_C *Escherichia coli* and *Alistipes indistinctus*, and the high-log a_{H_2O} position of the field for the high-$\bar{n}_{H_2O}$ *Peptostreptococcus stomatis*. Aside from *E. coli*, the proteins from the species associated with CRC in this dataset occupy the lower log f_{O_2} (reducing) and higher log a_{H_2O} zones of this diagram.

In thermodynamic calculations for proteins from bacteria detected in fecal metagenomes (Zeller et al., 2014; Table 3), 3 of 6 healthy-associated microbes, and 4 of 7 cancer-associated microbes exhibit maximal relative stability fields (Fig. 4b). Here, the cancer-associated proteins occupy the more reducing (*Fusobacterium nucleatum* subsp. vincentii and subsp. animalis) or more oxidizing (*Clostridium hylemonae*, *Porphyromonas asaccharolytica*) regions, while the proteins from bacteria more abundant in healthy individuals are relatively stable at moderate oxidation-reduction conditions. For the proteins representing microbial co-abundance groups (Candela et al., 2014; second part of Table 2), all of the 5 bacterial species show up on the diagram. Here, the cancer-associated proteins are more stable at reducing conditions and the healthy-related proteins are stabilized by oxidizing conditions. These patterns in relative stability are in accord with the distribution of Z_C of the proteins (Fig. 2b, c).

In Fig. 4d is a composite representation of the calculations, in which greater cumulative counts of maximal stability of healthy- and cancer-related proteins in the three studies are represented by deeper blue and red shading, respectively. The region predicted to be most energetically favorable for the formation of proteins in bacteria enriched in CRC is found at low log f_{O_2}; proteins from bacteria that are abundant in healthy patients tend to be stabilized by moderate values of log f_{O_2}. These oxidation-potential distinctions between the two groups are most apparent at high log a_{H_2O}. Thus, despite the differences in experimental design and specific taxonomic identity among the three data sets considered here, the thermodynamic calculations reveal a shared pattern of relative stabilities in terms of oxidation state.

3.5 Relative stability fields for human proteins

Compared to the overall microbial protein compositions used here (Tables 2, 3), there is a greater number of individual proteins identified in the data sets for healthy and cancer proteins in humans. A plot of the relative affinities of formation of individual proteins in Fig. 5a shows a sensitivity to theoretical redox potential, with lower log f_{O_2} tending to provide more favorable energetic conditions for the formation of up-expressed proteins in chromatin-binding fractions in carcinoma compared to adenoma. How can this visual impression be summarized quantitatively, in order to explore the data in multiple dimensions (both log a_{H_2O} and log f_{O_2})? Unlike the small number of overall microbial proteins, the higher numbers of proteins in these human comparisons diminish the utility of comparisons using maximal stability diagrams.

In Fig. 5b, the difference in mean values of chemical affinity per residue of carcinoma and adenoma-associated proteins is plotted, appearing as a straight line as a function of log f_{O_2}. This linear behavior would translate to evenly-spaced contours on a log f_{O_2}-log a_{H_2O} diagram. The rank-difference calculation of affinities (see Methods), shown by the jagged curving line Fig. 5b, provides a richer response function to changing chemical conditions. It emphasizes the transition zone between groups of proteins, and gives rise to two-dimensional stability diagrams with curved and diversely

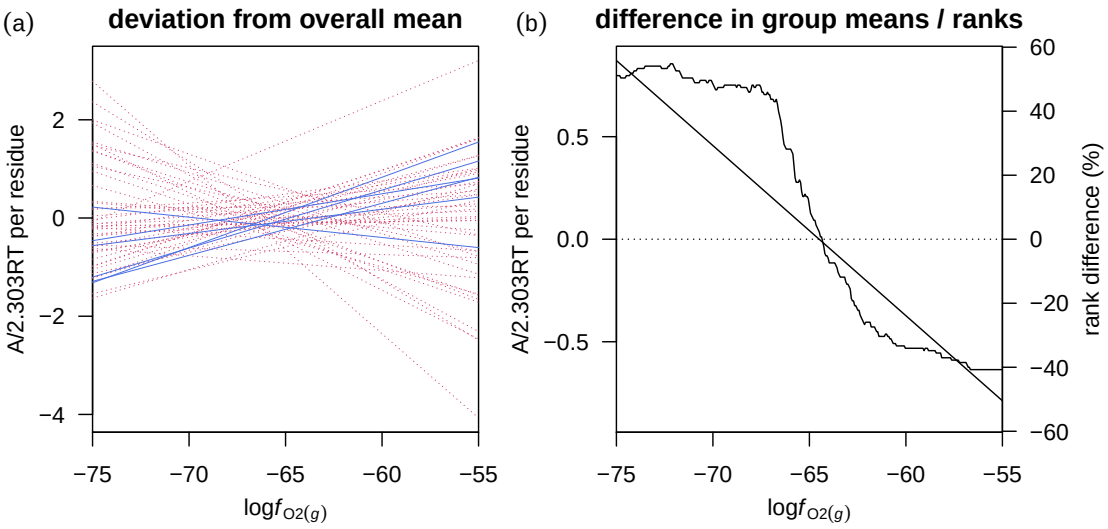

Figure 5. Calculated chemical affinities per residue of proteins in the "membrane fraction tumor/polyp" data set. Values for individual proteins as a function of $\log f_{O_2}$ at $\log a_{H_2O} = 0$ are shown in plot (a) as deviations from the mean value for all proteins. Healthy- and cancer-related proteins are indicated by solid blue and dashed red lines, respectively. Plot (b) shows the difference in mean value between healthy and cancer proteins (straight line and left-hand y-axis) and the weighted difference in sums of ranks of affinities as a percentage of maximum possible rank-sum difference (jagged line and right-side y-axis). Positive values of affinity or rank-sum difference in plot (b) correspond to relatively greater stability of the cancer-related proteins.

spaced contours.

The differences between sums of ranks of healthy- and cancer-associated proteins are depicted in Fig. 6 by the intensity of shading, with darker blue areas corresponding to more a negative difference (i.e. healthy-related proteins more stable) and darker red areas to a more positive difference (i.e. cancer-related proteins more stable). Under the interpretation of the energies as an indicator of propensity of formation of the proteins, the blue zones can be considered to be a prediction of the chemical conditions most conducive for a healthy state.

The proteins in the first three data sets are distinguished by large changes in the average oxidation state of carbon (Z_C) (Table 1). Correspondingly, the three affinity rank-difference plots (Fig. 6a-c) show sub-vertical equal-rank lines, with a higher ranking of stabilities of cancer-related proteins toward the left-hand side, or lower $\log f_{O_2}$. In the "membrane fraction tumor/polyp" data set, both Z_C and $\overline{n}_{H_2O}$ vary considerably, and the equal-rank line in Fig. 6c accordingly has a more diagonal slope. The second row of Fig. 6 shows plots for data sets in which the compositional differences are more strongly determined by changes in $\overline{n}_{H_2O}$. In each of these comparisons, the groups of cancer proteins are energetically favored by increasing $\log a_{H_2O}$.

The final two stability plots in Fig. 6 show the calculated relations for data sets that are derived from patterns of gene expression for a specific signaling pathway ("Wnt

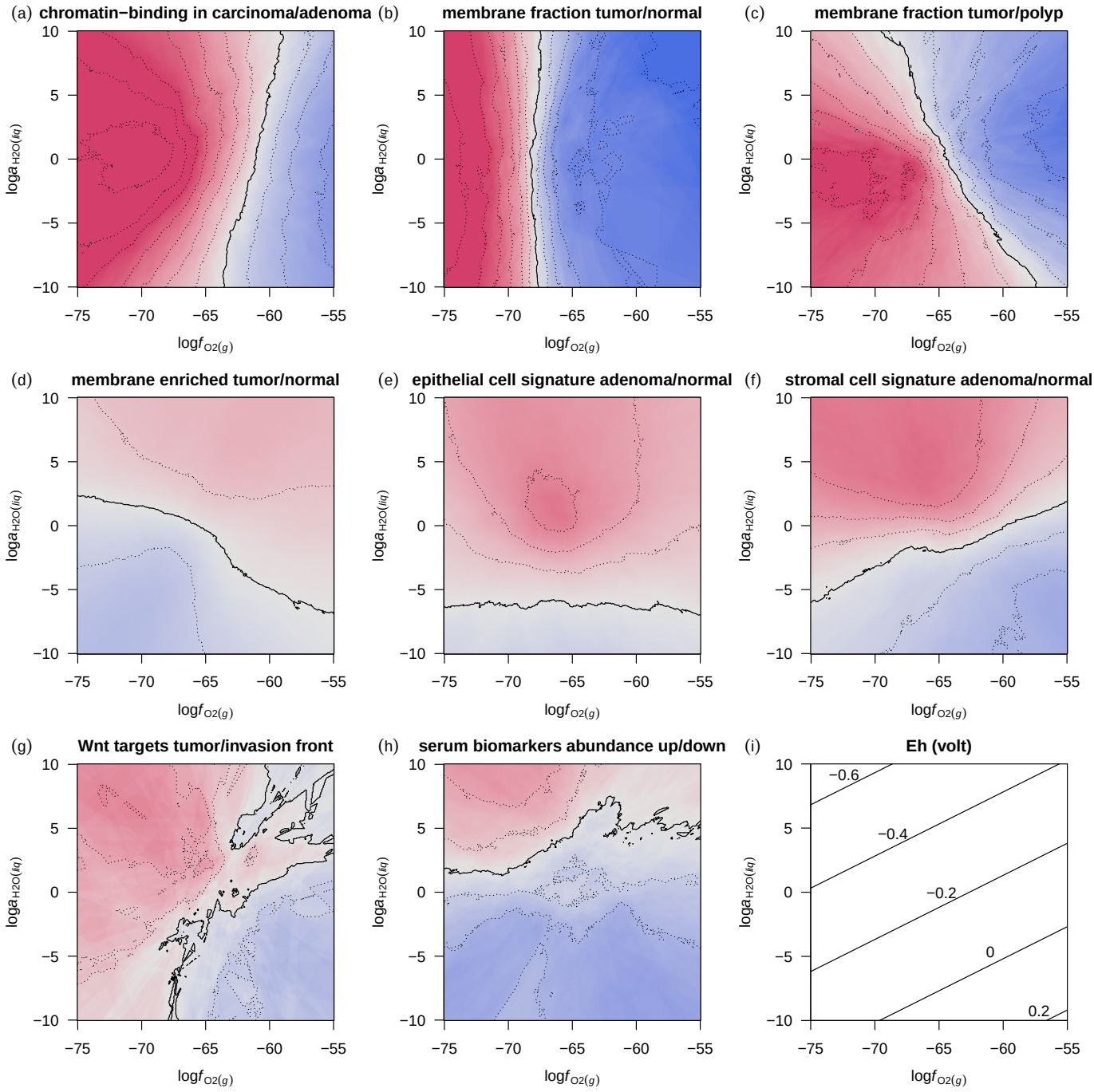

Figure 6. Weighted rank-sum comparisons of chemical affinities of formation of human proteins as a function of $\log f_{O_2}$ and $\log a_{H_2O}$. The solid lines indicate equal ranking of proteins in the cancer and healthy groups, and dotted contours are drawn at 10% increments of the maximum possible rank-sum difference. Blue and red areas correspond to higher ranking of cancer- and healthy-related proteins, respectively, with the intensity of the shading increasing up to 50% the maximum possible rank-sum difference. (For readers without a color copy: the cancer stability fields lie to the left of (a,b,c,g) or above (d,e,f,h) the stability fields for healthy proteins.) Panel (i) shows calculated values of Eh over the same range of $\log f_{O_2}$ and $\log a_{H_2O}$ (cf. Reaction R4).

targets in tumor/invasion front") or from a meta-analysis of extra-cellular proteins ("serum biomarkers abundance up/down"). Some features of the amino acid compositions of proteins in these data sets lead to stability diagrams that have a more complex topography (Fig. 6g-h). Nevertheless, the relative stabilization of tumor interior proteins and of up-expressed serum biomarker candidates follows the trends seen above of a stronger association of cancer with more reducing or more hydrating conditions.

The interpretation of $\log f_{O_2}$ and $\log a_{H_2O}$ as simple expressions of physical quantities (i.e. concentrations) is not straightforward (see Discussion). Nevertheless, effective values of oxidation-reduction potential (Eh) can be calculated by considering the water oxidation reaction, i.e.

$$H_2O \rightleftharpoons \tfrac{1}{2}O_2 + 2H^+ + 2e^- \qquad \text{(R4)}$$

If one assumes that $\log a_{H_2O} = 0$ (unit water activity, as in an infinitely dilute solution), this reaction can be used to interconvert $\log f_{O_2}$, pH and pe (or, in conjunction with the Nernst equation, Eh) (e.g. Garrels and Christ, 1965, p. 176; Anderson, 2005, p. 363). However, in the approach proposed here for metastable equilibrium among proteins in a subcellular metabolic context, no such assumptions are made on the *operational* value of $\log a_{H_2O}$, used as an internal indicator, not necessarily externally buffered by an aqueous solution. Consequently, the *effective* Eh is considered to be a function of variable $\log f_{O_2}$ and $\log a_{H_2O}$, as shown in Fig. 6i for pH = 7.4 and T = 37 °C. This comparison gives some perspective on operationally reasonable ranges of $\log f_{O_2}$ and $\log a_{H_2O}$.

The physiological reduction potential monitored by the reduced glutathione (GSH) / oxidized glutathione disulfide (GSSG) couple ranges from ca. -260 mV for proliferating cells to ca. -170 mV for apoptotic cells (Schafer and Buettner, 2001), lying toward the middle part of the range of conditions shown in Fig. 6 (e.g. Eh = -0.2 V corresponds to $\log f_{O_2}$ = -62.8 at unit activity of H_2O). Notably, this implies that the chemical energetic conditions most favorable for formation of cancer-related proteins (lower $\log f_{O_2}$ or higher $\log a_{H_2O}$) may be associated with redox potentials that characterize cell proliferation.

4 DISCUSSION

Here, proteomic data were analyzed within a chemical thermodynamic framework that supports placing chemical compositions of proteins in a microenvironmental context. To identify some of the implications of these findings, first consider some aspects of hypoxia and redox potential in cancer cells and tissues.

Hypoxia, or low oxygen concentration (Höckel and Vaupel, 2001), is often described as a key feature of the tumor microenvironment. A key cellular response to hypoxia is "oxidative stress", leading to a greater production of reactive oxygen species (ROS) (Murphy, 2009) and disruption of redox-sensitive signaling networks (e.g. Wei et al., 2013). In contrast to oxygenation level, oxidation-reduction potential (redox potential or Eh) can be monitored by the concentrations of reduced and oxidized glutathione (GSH/GSSG) and other redox couples. Cellular growth state can be influenced by hypoxia and/or redox potential; hypoxia is known to maintain

undifferentiated states of stem cells (Mohyeldin et al., 2010), and proliferation is generally associated with lower intracellular Eh values (Hutter et al., 1997; Schafer and Buettner, 2001), while a rise in intracellular Eh accompanies differentiation (Nkabyo et al., 2002).

The results here show that the chemical makeup of some populations of proteins is relatively reduced in cancer cells (the chromatin-binding and membrane proteins, and also the Wnt targets). The reduced chemical composition suggests the possibility of molecular adaptation to the biochemical constraints associated with lower oxygen and/or redox potential. These comparisons therefore identify a high-level cellular response to microenvironmental conditions that must coexist, but is not necessarily coincident, with the mechanistic basis for signaling in the HIF-1 pathway.

A thermodynamic perspective may open up other possibilities for control. For example, lactate buildup is usually regarded as a primary factor for lowered pH in the tumor microenvironment, promoting tumor growth (Gatenby and Gillies, 2004). However, tumors are not homogeneous. Often, catabolic products of stromal cells feed the anabolic requirements of epithelial cancer cells (Martinez-Outschoorn et al., 2014). The metabolic consequences of the "lactate shuttle" that operates in this scenario are tied to differences in oxidation state of lactate (Z_C=0) and pyruvate (Z_C=0.667) (Brooks, 2009). Interestingly, the decrease in Z_C of proteins in cancer relative to healthy samples is greater for the significantly changed proteins in stromal cells than in epithelial cells (see Table 1). The accumulation of the relatively reduced lactate molecule and the decrease in oxidation state of the proteome may both be associated with reducing conditions in specific locations within tumors.

Whether specific prevention and treatment strategies (for example, antioxidants) can prevent or reverse the redox-linked cellular adaptations in cancer is a matter of continuing debate. In general there is little evidence for a tight link between redox potential and hypoxia in tumors or healthy tissues, but one study has reported a correlation between redox potential and oxygenation in tumors on the basis of magnetic resonance imaging (MRI) using redox and oxygen-sensitive MRI probes (Hyodo et al., 2012). From a thermodynamic perspective, the cancer-related shifts toward a reduced protein composition constitute a cellular response that decrease the energy for biomolecular synthesis in a reducing environment. Therefore, it may be worthwhile to consider the implication of the current results that a return to normal proteomic state entails a net oxidation of the proteome.

In addition to oxidation state, the importance of hydration state can not be neglected. Increased hydration state in cells has been identified as a signal for protein synthesis and cell proliferation (Häussinger, 1996). A possible consequence of increasing water activity is enhanced rates of enzymatic hydrolysis (Cohen and Wolfenden, 1971), which could lead to DNA-damaging mutation through hydrolytic deamination of methylated C residues (Kinzler and Vogelstein, 1996). In some of the data sets considered here (membrane enriched, epithelial and stromal cell signature, and serum biomarkers), the cancer proteome is stabilized by a higher hydration potential. Therefore, it seems possible (but by no means conclusive) that whole-cell and secreted proteins may be driven to more hydrated states, in contrast with certain subcellular systems (chromatin-binding proteins, membrane fractions) which are more strongly shifted toward reducing conditions.

Regarding microbial populations associated with healthy and cancer tissues in the colon, the growth of bacteria that produce butyrate is associated with a healthy status and decreased levels of inflammation (Hamer et al., 2008; Candela et al., 2014). One of the beneficial outcomes of increased fiber in the diet is greater production of butyrate and other short-chain fatty acids (SCFA). The increased SCFA abundance in turn enhances microbial growth and protein synthesis, also depleting the nitrogen availability in the colon, measured as NH_3 (Cummings, 1981). The results of the present study show that proteins in cancer-associated bacteria are relatively reduced. Beyond the recognition of anaerobic conditions in the gut, little information can be found in the literature about the effects on the gut microbiome of changes in oxidation conditions. The combination of changes in oxidation state and nitrogen demand suggest the need for a multidimensional analysis to understand the chemical conditions that may enhance or inhibit the growth of pro-inflammatory microbes.

Some words should be added on the limitations of thermodynamic variables as indicators of the oxidation and hydration state of the system. Here, we selected $\log a_{H_2O}$ and $\log f_{O_2}$ as primary variables of interest. However, the conceptual basis of these variables as *indicators* of the hydration and oxidation state of the system (Anderson, 2005) does not support a direct interpretation in terms of measurable concentrations. There are astronomical differences between theoretical values of oxygen fugacity and actual concentrations or partial pressures of oxygen (e.g. Anderson, 2005, p. 364–365). Partial pressures of oxygen in human arterial blood are around 90-100 mmHg, and some thresholds for hypoxic conditions include 10 mmHg for energy metabolism, 0.5 mmHg for mitochondrial oxidative phosphorylation, and 0.02 mmHg for full oxidation of cytochromes (Höckel and Vaupel, 2001). Taking 1 mmHg = 1/760 atm = 1/750 bar and assuming ideal mixing, the equivalent range of oxygen fugacities indicated by these measurements is $\log f_{O_2}$ = -4.57 to -0.88, higher by far than the theoretical values that distinguish the relative stabilities of cancer and healthy-associated proteins computed here.

Likewise, the range of $\log a_{H_2O}$ shown here deviates tremendously from laboratory-based determination of water activity or hydration levels. For example, experimental determination of water activity in saturated protein solutions is as low as 0.5 (Knezic et al., 2004), and recent experiments and extrapolations show a likely limit in the range of ca. 0.600 to 0.650 for growth of various xerophilic and halophilic eukaryotes and prokaryotes (Stevenson et al., 2015), but cytoplasmic water activity is probably not greatly different from aqueous growth media, at 0.95 to 1 (Cayley et al., 2000). The theoretically computed transitions in relative stabilities between proteins from cancer and healthy tissues occur at much lower values of a_{H_2O} (ca. 10^{-6}; Fig. 6e) or at values approaching 1, depending on the oxygen fugacity (Figs. 6d,f).

Despite the difficulties in a quantitative interpretation, theoretical predictions of stabilization of cancer-related proteins by a decrease in $\log f_{O_2}$ (Fig. 6a-c,g) or increase in $\log a_{H_2O}$ (Fig. 6d-f,h) can both be interpreted qualitatively as corresponding with a decrease in redox potential (Fig. 6i).

5 CONCLUSION

Conventional methods for interpreting proteomic and microbial data sets emphasize protein function and mechanistic descriptions of the normal operation of cells as well as abnormal signaling and growth in cancer. However, an integrated picture of proteomic remodeling in cancer may benefit from accounting for the stoichiometric and energetic requirements of protein formation. This study has identified a strong shift toward more reduced proteins in colorectal cancer. Importantly, this pattern is identified across multiple data sets, increasing confidence in its systematic nature. In some other data sets, a smaller but still systematic change can be identified indicating greater water demand of human proteins in cancer compared to normal tissue.

The findings from this exploration of data unmask biochemical patterns that are consistent with observations of hypoxia (possibly associated with reduced oxidation state) and higher hydration state in cancer cells and tumors. As noted by others, such chemical shifts may favor cellular proliferation and reversion to an embryological mode of growth. The proteomic data can be explicitly linked to microenvironmental conditions using thermodynamic models, which give estimates of the oxidation- and hydration-potential limits for relative stability of groups of proteins. These calculations outline a path connecting the dynamic compositions of proteomes to biochemical measurements such as Eh. In conjunction with growing and more targeted proteomic data sets, future studies can take advantage of this approach in order to characterize and understand in more detail the microenvironmental requirements for the initiation, progression and metastasis of CRC and other cancers.

REFERENCES

Abramczyk, H., Brozek-Pluska, B., Krzesniak, M., Kopec, M., and Morawiec-Sztandera, A. (2014). The cellular environment of cancerous human tissue. Interfacial and dangling water as a "hydration fingerprint". *Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy*, 129(0):609–623. DOI 10.1016/j.saa.2014.03.103

Anderson, G. M. (2005). *Thermodynamics of Natural Systems*. Cambridge University Press, Cambridge, 2nd edition.

Bertout, J. A., Patel, S. A., and Simon, M. C. (2008). The impact of O_2 availability on human cancer. *Nature Reviews. Cancer*, 8(12):967–975. DOI 10.1038/nrc2540

Biarc, J., Nguyen, I. S., Pini, A., Gossé, F., Richert, S., et al. (2004). Carcinogenic properties of proteins with pro-inflammatory activity from *Streptococcus infantarius* (formerly *S.bovis*). *Carcinogenesis*, 25(8):1477–1484. DOI 10.1093/carcin/bgh091

Brooks, G. A. (2009). Cell–cell and intracellular lactate shuttles. *Journal of Physiology*, 587(23):5591–5600. DOI 10.1113/jphysiol.2009.178350

Candela, M., Turroni, S., Biagi, E., Carbonero, F., Rampelli, S., et al. (2014). Inflammation and colorectal cancer, when microbiota-host mutualism breaks. *World Journal of Gastroenterology*, 20(4):908–922. DOI 10.3748/wjg.v20.i4.908

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., et al. (2012). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Research*, 22(2):299–306. DOI 10.1101/gr.126516.111

Cayley, D. S., Guttman, H. J., and Record, Jr., M. T. (2000). Biophysical characterization of changes in amounts and activity of *Escherichia coli* cell and compartment water and turgor pressure in response to osmotic stress. *Biophysical Journal*, 78(4):1748–1764. DOI 10.1016/S0006-3495(00)76726-9

Chen, J.-S., Chen, K.-T., Fan, C.-W., Han, C.-L., Chen, Y.-J., et al. (2010). Comparison of membrane fraction proteomic profiles of normal and cancerous human colorectal tissues with gel-assisted digestion and iTRAQ labeling mass spectrometry. *FEBS Journal*, 277(14):3028–3038. DOI 10.1111/j.1742-4658.2010.07712.x

Chen, W., Liu, F., Ling, Z., Tong, X., and Xiang, C. (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, 7(6):e39743. DOI 10.1371/journal.pone.0039743

Cohen, R. M. and Wolfenden, R. (1971). Cytidine deaminase from *Escherichia coli*: purification, properties, and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *Journal of Biological Chemistry*, 246(24):7561–7565

Cummings, J. H. (1981). Short chain fatty acids in the human colon. *Gut*, 22(9):763–779. DOI 10.1136/gut.22.9.763

DeBerardinis, R. J. and Cheng, T. (2010). Q's next: The diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*, 29(3):313–324. DOI 10.1038/onc.2009.358

Dick, J. M. (2008). Calculation of the relative metastabilities of proteins using the CHNOSZ software package. *Geochemical Transactions*, 9:10. DOI 10.1186/1467-4866-9-10

Dick, J. M. (2014). Average oxidation state of carbon in proteins. *Journal of the Royal Society Interface*, 11:20131095. DOI 10.1098/rsif.2013.1095

Dick, J. M., LaRowe, D. E., and Helgeson, H. C. (2006). Temperature, pressure, and electrochemical constraints on protein speciation: Group additivity calculation of the standard molal thermodynamic properties of ionized unfolded proteins. *Biogeosciences*, 3(3):311–336. DOI 10.5194/bg-3-311-2006

Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S., and Flint, H. J. (2002). *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, 52(5):1615–1620. DOI 10.1099/ijs.0.02143-0

Garrels, R. M. and Christ, C. L. (1965). *Solutions, Minerals, and Equilibria*. Harper & Row, New York.

Gatenby, R. A. and Gillies, R. J. (2004). Why do cancers have high aerobic glycolysis? *Nature Reviews. Cancer*, 4(11):891–899. DOI 10.1038/nrc1478

Gibbs, J. W. (1875). On the equilibrium of heterogeneous substances (first part). *Transactions of the Connecticut Academy of Arts and Sciences*, 3:108–248

Grzegorczyk, T. M., Meaney, P. M., Kaufman, P. A., di Florio-Alexander, R. M., and Paulsen, K. D. (2012). Fast 3-D tomographic microwave imaging for breast cancer detection. *IEEE Transactions on Medical Imaging*, 31(8):1584–1592. DOI 10.1109/TMI.2012.2197218

Hamer, H. M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F. J., et al. (2008). Review article: The role of butyrate on colonic function. *Alimentary Pharmacology & Therapeutics*, 27(2):104–119. DOI 10.1111/j.1365-2036.2007.03562.x

Harris, A. L. (2002). Hypoxia – A key regulatory factor in tumour growth. *Nature Reviews. Cancer*, 2(1):38–47. DOI 10.1038/nrc704

Häussinger, D. (1996). The role of cellular hydration in the regulation of cell function. *Biochemical Journal*, 313:697–710. DOI 10.1042/bj3130697

Helgeson, H. C. (1970). Description and interpretation of phase relations in geochemical processes involving aqueous solutions. *American Journal of Science*, 268(5):415–438. DOI 10.2475/ajs.268.5.415

Helgeson, H. C., Kirkham, D. H., and Flowers, G. C. (1981). Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: IV. Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600°C and 5 Kb. *American Journal of Science*, 281(10):1249–1516. DOI 10.2475/ajs.281.10.1249

Hlubek, F., Brabletz, T., Budczies, J., Pfeiffer, S., Jung, A., et al. (2007). Heterogeneous expression of Wnt/β-catenin target genes within colorectal cancer. *International Journal of Cancer*, 121(9):1941–1948. DOI 10.1002/ijc.22916

Höckel, M. and Vaupel, P. (2001). Tumor hypoxia: Definitions and current clinical, biologic, and molecular aspects. *Journal of the National Cancer Institute*, 93(4):266–276. DOI 10.1093/jnci/93.4.266

Hoffman, A., Spetner, L. M., and Burke, M. (2008). Ramifications of a redox switch within a normal cell: Its absence in a cancer cell. *Free Radical Biology and Medicine*, 45(3):265–268. DOI 10.1016/j.freeradbiomed.2008.03.025

Hutter, D. E., Till, B. G., and Greene, J. J. (1997). Redox state changes in density-dependent regulation of proliferation. *Experimental Cell Research*, 232(2):435–438. DOI 10.1006/excr.1997.3527

Hyodo, F., Davis, R. M., Hyodo, E., Matsumoto, S., Krishna, M. C., et al. (2012). The relationship between tissue oxygenation and redox status using magnetic resonance imaging. *International Journal of Oncology*, 41(6):2103–2108. DOI 10.3892/ijo.2012.1638

Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. A. (2010). Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *Journal of Proteomics*, 73(10):1873–1895. DOI 10.1016/j.jprot.2010.06.004

Johnson, J. W., Oelkers, E. H., and Helgeson, H. C. (1992). SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. *Computers & Geosciences*, 18(7):899–947. DOI 10.1016/0098-3004(92)90029-Q

Keating, P. and Cambrosio, A. (2012). Too many numbers: Microarrays in clinical cancer research. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1):37–51. DOI 10.1016/j.shpsc.2011.10.004

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

Kinzler, K. W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell*, 87(2):159–170. DOI 10.1016/S0092-8674(00)81333-1

Knezic, D., Zaccaro, J., and Myerson, A. S. (2004). Thermodynamic properties of supersaturated protein solutions. *Crystal Growth & Design*, 4(1):199–208. DOI 10.1021/cg034072o

Knol, J. C., de Wit, M., Albrethsen, J., Piersma, S. R., Pham, T. V., et al. (2014). Proteomics of differential extraction fractions enriched for chromatin-binding proteins from colon adenoma and carcinoma tissues. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844(5):1034–1043. DOI 10.1016/j.bbapap.2013.12.006

Koboziev, I., Webb, C. R., Furr, K. L., and Grisham, M. B. (2014). Role of the enteric microbiota in intestinal homeostasis and inflammation. *Free Radical Biology and Medicine*, 68(0):122–133. DOI 10.1016/j.freeradbiomed.2013.11.008

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., et al. (2012). Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Research*, 22(2):292–298. DOI 10.1101/gr.126573.111

Kume, H., Muraoka, S., Kuga, T., Adachi, J., Narumi, R., et al. (2014). Discovery of colorectal cancer biomarker candidates by membrane proteomic analysis and subsequent verification using selected reaction monitoring (SRM) and tissue microarray (TMA) analysis. *Molecular & Cellular Proteomics*, 13(6):1471–1484. DOI 10.1074/mcp.M113.037093

LaRowe, D. E. and Dick, J. M. (2012). Calculation of the standard molal thermodynamic properties of crystalline peptides. *Geochimica et Cosmochimica Acta*, 80:70–91. DOI 10.1016/j.gca.2011.11.041

Lazebnik, M., Popovic, D., McCartney, L., Watkins, C. B., Lindstrom, M. J., et al. (2007). A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. *Physics in Medicine and Biology*, 52(20):6093–6115. DOI 10.1088/0031-9155/52/20/002

Louis, P. and Flint, H. J. (2007). Development of a semiquantitative degenerate real-time PCR-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. *Applied and Environmental Microbiology*, 73(6):2009–2012. DOI 10.1128/AEM.02561-06

Martinez-Outschoorn, U. E., Lisanti, M. P., and Sotgia, F. (2014). Catabolic cancer-associated fibroblasts transfer energy and biomass to anabolic cancer cells, fueling tumor growth. *Seminars in Cancer Biology*, 25:47–60. DOI 10.1016/j.semcancer.2014.01.005

May, P. M., , and Murray, K. (2001). Database of chemical reactions designed to achieve thermodynamic consistency automatically. *Journal of Chemical & Engineering Data*, 46(5):1035–1040. DOI 10.1021/je000246j

McIntyre, G. I. (2006). Cell hydration as the primary factor in carcinogenesis: A unifying concept. *Medical Hypotheses*, 66(3):518–526. DOI 10.1016/j.mehy.2005.09.022

Menon, S. G., Sarsour, E. H., Spitz, D. R., Higashikubo, R., Sturm, M., et al. (2003). Redox regulation of the G(1) to S phase transition in the mouse embryo fibroblast cell cycle. *Cancer Research*, 63(9):2109–2117

Mohyeldin, A., Garzón-Muvdi, T., and Quiñones Hinojosa, A. (2010). Oxygen in stem cell biology: A critical component of the stem cell niche. *Cell Stem Cell*, 7(2):150–161. DOI 10.1016/j.stem.2010.07.007

Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. *Biochemical Journal*, 417:1–13. DOI 10.1042/bj20081386

Nagai, F., Morotomi, M., Sakon, H., and Tanaka, R. (2009). *Parasutterella excrementihominis* gen. nov., sp. nov., a member of the family *Alcaligenaceae* isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, 59(7):1793–1797. DOI 10.1099/ijs.0.002519-0

Nkabyo, Y. S., Ziegler, T. R., Gu, L. H., Watson, W. H., and Jones, D. P. (2002). Glutathione and thioredoxin redox during differentiation in human colon epithelial (Caco-2) cells. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 283(6):G1352–G1359. DOI 10.1152/ajpgi.00183.2002

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Sarsour, E. H., Kumar, M. G., Chaudhuri, L., Kalen, A. L., and Goswami, P. C. (2009). Redox control of the cell cycle in health and disease. *Antioxidants & Redox Signaling*, 11(12):2985–3011. DOI 10.1089/ars.2009.2513

Schafer, F. Q. and Buettner, G. R. (2001). Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radical Biology and Medicine*, 30(11):1191–1212. DOI 10.1016/S0891-5849(01)00480-4

Schedin, P. and Elias, A. (2004). Multistep tumorigenesis and the microenvironment. *Breast Cancer Research*, 6(2):93–101. DOI 10.1186/bcr772

Sears, C. L. and Garrett, W. S. (2014). Microbes, microbiota, and colon cancer. *Cell Host & Microbe*, 15(3):317–328. DOI 10.1016/j.chom.2014.02.007

Semenza, G. L. (2008). Tumor metabolism: Cancer cells give and take lactate. *The Journal of Clinical Investigation*, 118(12):3835–3837. DOI 10.1172/JCI37373

Sethi, M. K., Thaysen-Andersen, M., Kim, H., Park, C. K., Baker, M. S., et al. (2015). Quantitative proteomic analysis of paired colorectal cancer and non-tumorigenic tissues reveals signature proteins and perturbed pathways involved in CRC progression and metastasis. *Journal of Proteomics*, 126:54–67. DOI 10.1016/j.jprot.2015.05.037

Shock, E. and Canovas, P. (2010). The potential for abiotic organic synthesis and biosynthesis at seafloor hydrothermal systems. *Geofluids*, 10(1-2):161–192. DOI 10.1111/j.1468-8123.2010.00277.x

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., et al. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43):16731–16736. DOI 10.1073/pnas.0804812105

Stevenson, A., Cray, J. A., Williams, J. P., Santos, R., Sahay, R., et al. (2015). Is there a common water-activity limit for the three domains of life? *ISME Journal*, 9(6):1333–1351. DOI 10.1038/ismej.2014.219

Tanger, J. C. IV. and Helgeson, H. C. (1988). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Revised equations of state for the standard partial molal properties of ions and electrolytes. *American Journal of Science*, 288(1):19–98. DOI 10.2475/ajs.288.1.19

Toole, B. P. (2002). Hyaluronan promotes the malignant phenotype. *Glycobiology*, 12(3):37R–42R. DOI 10.1093/glycob/12.3.37R

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., et al. (2009). A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484. DOI 10.1038/nature07540

Uzozie, A., Nanni, P., Staiano, T., Grossmann, J., Barkow-Oesterreicher, S., et al. (2014). Sorbitol dehydrogenase overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: A quantitative proteomics study. *Molecular & Cellular Proteomics*, 13(5):1198–1218. DOI 10.1074/mcp.M113.035105

Wagman, D. D., Evans, W. H., Parker, V. B., Schumm, R. H., Halow, I., et al. (1982). The NBS tables of chemical thermodynamic properties. Selected values for inorganic and C_1 and C_2 organic substances in SI units. *Journal of Physical and Chemical Reference Data*, 11:1–392. DOI 10.1063/1.555661

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., et al. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME Journal*, 6(2):320–329. DOI 10.1038/ismej.2011.109

Wei, W., Shi, Q., Remacle, F., Qin, L., Shackelford, D. B., et al. (2013). Hypoxia induces a phase transition within a kinase signaling network in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15):E1352–E1360. DOI 10.1073/pnas.1303060110

Weir, T. L., Manter, D. K., Sheflin, A. M., Barnett, B. A., Heuberger, A. L., et al. (2013). Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS ONE*, 8(8):e70803. DOI 10.1371/journal.pone.0070803

Zadran, S., Remacle, F., and Levine, R. D. (2013). miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. *Proceedings of the National Academy of Sciences of the United States of America*, 110(47):19160–19165. DOI 10.1073/pnas.1316991110

Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology*, 10(11):766. DOI 10.15252/msb.20145645