

Interactive comment on “Uncovering chemical signatures of salinity gradients through compositional analysis of protein sequences” by Jeffrey M. Dick et al.

Jeffrey M. Dick et al.

jeff@chnosz.net

Received and published: 15 October 2020

Point-by-point Response to Anonymous Referee #2

In general, I found the paper to be another interesting read from the primary author. However, as a microbiologist that is interested in understanding how energy availability and demand affect the distribution of microorganisms and their evolution, I would appreciate seeing a more robust effort to link the thermodynamics way of thinking (as presented here) to physiological process or mechanism that could then be used to gauge why such patterns may exist. More or

C1

less, I think this is a missed opportunity that, if executed effectively, could elevate the utility of this paper and this way of thinking. Thus, I strongly suggest the authors attempt to explain their observations at a level that makes sense to the more biologically oriented reader. As I was reading this, I could not help but think to myself how any one or several observations made sense from the level of phenotype and natural selection. The authors might consider asking themselves this same question and then speculating where possible to make this body of work a greater utility for the community.

Thank you for your detailed attention to the concepts and analysis in our paper and your suggestions for improving the work. We respond to the main critiques below:

1) a more robust effort to link the thermodynamics way of thinking (as presented here) to physiological process or mechanism

This is an ongoing challenge. An obstacle (which could also be seen as a “missed opportunity”) is that the thermodynamic way of thinking deals with energetic differences between two states of a system; without further (i.e. extra-thermodynamic) constraints, it is not possible to explicitly deal with underlying mechanisms in a thermodynamic model. This paper does not attempt to build such a thermodynamic model, but uses thermodynamics as a guiding concept. A major application of thermodynamics in geochemistry is to describe and predict compositional changes in a system, e.g. the distribution of aqueous species and mineral phases with different chemical formulas. The aim of this paper is to develop a framework for describing compositional changes in geobiochemical systems, and one of the first challenges is to recognize that the most appropriate descriptive variables are probably different from inorganic geochemical systems. We present our conceptual arguments that oxidation and hydration state should be considered as primary variables, develop metrics that quantify them, and use the metagenomic data to explore how these metrics respond to environmental gradients of salinity and redox conditions. Clearly, this is far from the sophisticated applications of thermodynamics in geochemistry, but it serves as a step toward

C2

a broader appreciation that compositional changes are not random, but are aligned with environmental conditions. That should motivate the development of more rigorous thermodynamic models in future studies.

As a partial response to the request for a more mechanistic understanding, it can be noted that Fig. 7 has been redrawn to place time on the horizontal axis. With this change, it should be more apparent that the chemical composition of the differentially expressed proteins changes dynamically in laboratory experiments.

2) I strongly suggest the authors attempt to explain their observations at a level that makes sense to the more biologically oriented reader.

The paper uses some technical language from physical chemistry and thermodynamics by necessity, and these technical terms are defined when introduced. These concepts are used to quantitatively analyze metagenomic datasets that are chosen to represent well-known regional gradients. The analysis of laboratory data includes protein expression in response to salt and osmotic shock. Therefore, the core of the paper is concerned with biological phenomena in an environmental context. The mixing of biological data and physicochemical metrics is what makes this paper unique; removing the quantitative language would eliminate its main contribution.

We note that the entire section on “Conceptual background” was added in a previous revision (before submission to this journal) to make the paper more accessible to biologists. The paragraphs here deal with issues about intracellular conditions, amino acid composition, distinction with polymerization reactions, selection for structural stability of proteins, other variables like temperature and pH, and relation of the basis species to biosynthetic mechanisms. However, our intention is not to write a theoretical paper but rather to present a coherent set of data analyses to convince the reader that compositional differences of proteins have a basic significance in geobiochemical systems.

3) I could not help but think to myself how any one or several observations made sense from the level of phenotype and natural selection. The authors might con-

C3

sider asking themselves this same question and then speculating where possible to make this body of work a greater utility for the community.

We believe that the analysis of laboratory experiments of protein expression in salt and osmotic conditions does provide basic information about the effects of the environment on the observable characteristics of cells. Admittedly, this is only one aspect of the phenotype, and other types of experiments could be considered, like gene expression, metabolomes, and metabolic fluxes, but analysis of those types of data is out of the scope of this paper.

A relevant finding from a paper in preparation is that the stoichiometric hydration state of differentially expressed proteins is strongly decreased in 3D (tissue-like) compared to 2D (monolayer) culture conditions of eukaryotic cells (Dick, 2020). The lower n_{H_2O} in 3D culture has some similarity to the observation in this study that metagenome-inferred proteins in particles tend to have lower hydration state compared to free-living fractions. These responses could plausibly be associated with lower water accessibility in the interiors of particles in environmental samples and in spheroids in 3D cell culture.

Regarding the evolutionary implications, another paper is in early preparation that shows the hydration and oxidation state computed for whole proteomes of phylogenetic groups predicted from the RefSeq database. This tree-like view of the chemical composition no doubt would help solidify the relevance of the physicochemical concepts used here to biological systems. That is being developed for a separate paper with its own set of data analysis of microbial community composition and it is too early to cite the results in this paper.

The following list of minor comments is meant to further improve this work: Line 1: For the average reader – what is the connection between thermodynamics and environmental variation. Lead in with this first.

Added text in Abstract: “Prediction of the direction of change of a system under specified environmental conditions is one reason for the widespread utility of thermodynamic

C4

models in geochemistry.”

Line 8: Replace “behave” with something more valid. The metric does not correlate for XXX in hypersaline environments. . .

Changed “behave” to “respond”.

Line 15: Communities do not adapt, populations of individuals do.

Changed “communities” to “populations”.

Line 26: I would not call this complementary but rather an interrelated approach since selection (imposed as an argument in previous paragraph) can and should act on the energetic demand of protein synthesis.

Changed “complementary” to “interrelated”.

Line 39-40: What about the authors own work on the communities inhabiting the out flow channel at Bison Pool, Yellowstone?

Added references and reworded the sentence for better context: “The oxidation state of proteins as well as lipids has been shown to be associated with oxidation-reduction (redox) gradients in a hot spring (Dick and Shock, 2011; Boyer et al., 2020), but so far energetic models have not been broadly adopted as a tool for relating metagenomic and geochemical data.”

Line 44-45: While I don’t disagree with this assumption, at least as a first order constraint, it would be useful to relate to the reader why this assumption is made. Perhaps to avoid this confusion, the authors move this statement to below where they describe and justify their approach.

This sentence has been moved down to the second point in the “Conceptual background” section, following the reference about missing hydrogen and oxidation state in stoichiometric models (Karl and Grabowski, 2017).

C5

Line 58-62: This paragraph seems out of place. I suggest moving the discussion of what you did previously up in the introduction and add the last sentence of this paragraph to the end of the preceding paragraph.

The statement of previous work and what’s new in this study has been moved up to the position of the former Lines 44-45 mentioned in the previous comment. The long-term research goal has been removed, because it doesn’t seem to fit anywhere now.

Line 67: alternatives to what?

Each area of concern is summarized here as “X or Y”, which seems consistent with the dictionary’s definition of an alternative as “a choice between two things”. To avoid confusion, this has been reworded as “six areas of concern summarized as: 1) . . . 2)”

Line 305-310: I don’t understand the reasoning here? Why did eukaryotes start to become important in these systems? Are there actually eukaryotes in these systems? The authors have the data to evaluate this and should evaluate it to see if the logic makes sense.

This has been removed in the revision. The comparison of the average stoichiometric hydration state of human proteins with *E. coli* and the metagenomic data analyzed in this study provided preliminary support for the concept of a lower n_{H_2O} in eukaryotes, but a more targeted data analysis is needed to strengthen this claim. Also note that the human and *E. coli* proteomes have been removed in the revised description of the choice of basis species (Fig. 1).

Line 315: Why would heterotroph proteomes have a lower hydration state?

There might be something basically different about their metabolic pathways in terms of water requirements at the biochemical level. Apart from *E. coli*, there probably are not many existing metabolic models that could be used to test this speculation. Added sentence: “A better understanding of these trends would require more extensive phyloge-

C6

netically resolved comparisons of the compositional differences as well as biochemical (or computational) analyses of water fluxes in metabolic pathways.”

Line 315-317: is there an argument to be made about why a major evolutionary transition favors a shift from higher to lower dehydration state? i.e., is this an adaptive feature that allows the latter to compete with the former from an evolutionary perspective?

This is certainly a valid question, but we are unable to provide a convincing mechanistic reason for why lower hydration state might offer a selective advantage. Perhaps it should be considered not as adaptation but as physical constraint, similar in a way to Gould and Lewontin (1979)’s spandrels. Structures that are physically durable, such as macromolecular complexes in organelles or larger assemblages like tissues, might be those that are relatively dry. Physical dryness (i.e. lower water content) could be a selective force for lower stoichiometric hydration state of biomolecules, but the latter by itself may have no fitness advantage.

If lower n_{H_2O} turned out to characterize some evolutionary transitions, it would seem to be consistent with the postulate that “ontogeny recapitulates phylogeny” and the observation that progressive loss of water occurs in animal development through the stages of embryo, fetus, birth and growth (Moulton, 1923).

[These ideas are rather speculative, and don’t specifically deal with the (non-eukaryotic) metagenomes that are analyzed here, so haven’t been added to the text.]

Line 325: is it possible that diffusion limitation makes H₂O less available to cells living nearer to a particle surface? Again, an explanation for what the observations might mean is warranted.

Particles likely provide opportunities for some amount of physical separation from the bulk aqueous phase; it’s harder to pin down the molecular mechanisms. Added: “Together with the lower n_{H_2O} for proteins inferred from metagenomes and metatranscrip-

C7

tomes in the larger size fractions from Baltic Sea samples, this could reflect a lower availability of H₂O to organisms living near the particle surface due to physical separation from the bulk aqueous phase and associated diffusion limitation or lower water activity (Wang et al., 2003).”

Line 350: proteins in metagenomes

Changed “plume metagenomes” to “proteins in plume metagenomes”.

Line 360: Could this be due to aquaculture and introduction of more organic compounds/waste and its selection of heterotrophic taxa, that as stated earlier in the paper, tend to host proteomes with a lower hydration state

This seems very reasonable. Added: “The microbial communities in the aquaculture ponds may not be responding as they would in a typical natural system that is less nutrient-rich. As noted above for putative heterotroph-rich zones in other systems, the lower stoichiometric hydration state could be associated with the enrichment of heterotrophic taxa, in this case due to the addition of organic compounds to the aquaculture ponds.”

See also the response to Referee #1 and the revised discussion of the analysis of differentially expressed proteins: “The negative shift of Δn_{H_2O} associated with most organic solutes compared to NaCl lends support to the notion that high organic loading could contribute to the relatively low n_{H_2O} of protein sequences from metagenomes of freshwater aquaculture systems.”

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2020-146>, 2020.

C8