



The responsibility of bringing forward new ideas in science is shared by authors, editors and the whole scientific community. A practical case.

SEAN RA DEVENISH¹, PIETRO GATTI-LAFRANCONI¹

1. University of Cambridge

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

pg356@cam.ac.uk

DATE RECEIVED:

June 11, 2015

KEYWORDS:

publishing, evolution, peer-review, proteins, redox

© Devenish et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



READING SOMETHING NEW

Every once in a while, it is healthy for established concepts and standard procedures to be challenged by ideas that come from outside the discipline's specialist field. They might be incomplete, often they are naïve, but can nevertheless trigger a reassessment of conventions and suggest new lines of study.

The paper '*Average oxidation state of carbon in proteins*' by Jeffrey M. Dick (Dick, 2014) uses molecular formula analysis of proteins to address molecular evolution (and co-evolution) in biological systems. It is suggested that measuring the average formal oxidation state of carbon of proteins and DNA from their chemical formulae can illuminate the evolutionary relationships between genetic code and proteins, and explain differences in subcellular localisation and environmental adaptation. The extreme naivety of an approach that deliberately ignores the complexity of biological systems (folding, function, metabolism, energy costs, and transport, to mention a few) is intriguing, but the paper fails to deliver as a consequence of flawed methods and an excessively superficial approach to biology.

A NEW METHOD, BUT A GOOD ONE PLEASE

The author uses the average oxidation state of carbon (ZC) as a quick (and admittedly superficial) metric to study large dataset of proteins. ZC is defined as:

$$Z_C = \frac{-n_H + 3n_N + 2n_O + 2n_S + Z}{n_C},$$

where Z is the charge on the molecule, and nC, nH, nN, nO and nS are the numbers of the subscripted elements in the chemical formula of the molecule.

First of all, it is important to note that calculation of the oxidation state of carbon is a formalism that is useful solely for monitoring changes of oxidation state of individual carbon atoms, and the absolute values, let alone the average value over an entire molecule, does not have particular meaning. The example provided is the Zc score for hen egg white lysozyme (chemical formula C613H959N193O185S10) that equals 0.016. The formula ignores factors as charge (offset by the inclusion of the Z value in the equation), folding, post-translational modifications, solubility, stability etc, that are thus deemed irrelevant for understanding the biological function of individual proteins and the evolution and complexity of entire organisms.

In addition, as the algorithm counts also the (invariable) main chain atoms, which contribute heavily due to the carbonyl group, so that small residues automatically end up with high Zc scores, even if the side chain is completely reduced. On the other hand, large residues that are not extensively oxidised in the side chain to counteract their bulk receive low ZC scores, implying that they are highly reduced. This includes residues such as phenylalanine (-0.44), tyrosine (-0.22) and tryptophan (-0.18) despite these residues being more oxidised in the side chain than for example alanine (0), serine (0.67) or glycine (1). Instead of reporting solely on the oxidation-reduction state of a chemical compound, the ZC score correlates with the chemical nature and bulkiness of amino acids' side chains:

Table 1: Amino acid properties and ZC scores

group	Amino acids	Average ZC score	Edited ZC score
hydrophobic	A,C,G,I,L,M,F,P,W,V	-0.255	-0.394 (no G)
polar	N,Q,S,T,Y	0.370	
basic	R,H,K	0.110	0.500 (no K)
acid	D,E	0.700	

Given the above, it is perhaps unsurprising that there is a weak correlation between ZC and hydrophobicity. This correlation is significantly improved if the side chain only is used to calculate the ZC score (Figure 1), but this is due simply to the fact that amino acids that have oxidised carbon atoms in their side-chains are polar.

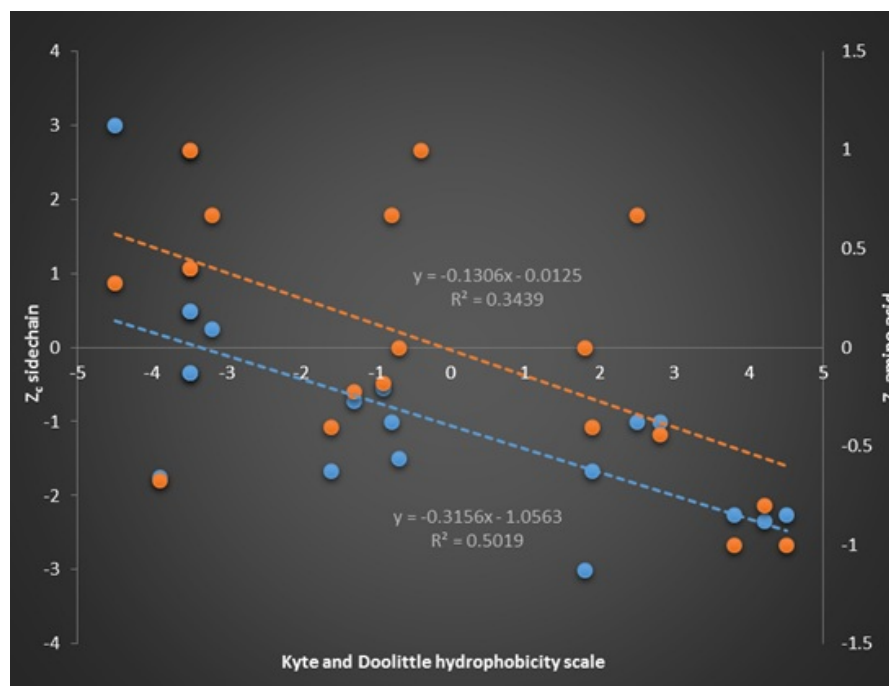


Figure 1: Hydrophobicity values for amino acids (according to the hydrophobicity scale of Kyte and Doolittle (Kyte & Doolittle, 1982)) plotted against the ZC score calculated on either full amino acid structures or side chains only (orange and blue data, respectively).

Finally, a note on the suitability of ZC to study *proteins*. The degree of oxidation of a given amino acid, reflected in its ZC score, is a function of the biosynthesis of that amino acid. This implies that the ZC of a whole protein only reflects (at best) the flux of the amino acid biosynthetic pathways of a cell and is not correlated with the function, nature, location, concentration or stability of proteins (let alone their evolution).

BIOLOGY CAN BE RE-INTERPRETED, BUT MUST BE UNDERSTOOD FIRST.

To kick off the analysis with an easy task, ZC is used to probe an evolutionary relationship between genetic code and individual amino acids. To this aim, the ZC of base 'duplets' (of RNA, and not DNA, oddly) is compared with that of encoded amino acids (Figure 1). Ignoring errors in the figure itself, the correlation is weak, with a R^2 value of 0.36 (which can also be obtained by plotting the molecular weight of the duplex vs ZC). This is not the place to reassess the evolution of the genetic code (and address misunderstandings of key references), but one might wonder what the idea is behind highlighting the existence of a correlation (hence implying an evolutionary driving force) between the redox state of codons and those of amino acids. ZC's bias on the bulkiness of chemical structures implies that its value for DNA bases reflects the Purines/Pyrimidines structural classification, with T/C having lower ZC scores than A/G. This classification thus ignores codon/anticodon interaction strengths. The degrees of redundancy of different amino acids also is ignored, and indeed ZC shows no correlation with the observed frequency of each amino acid (Figure 2).

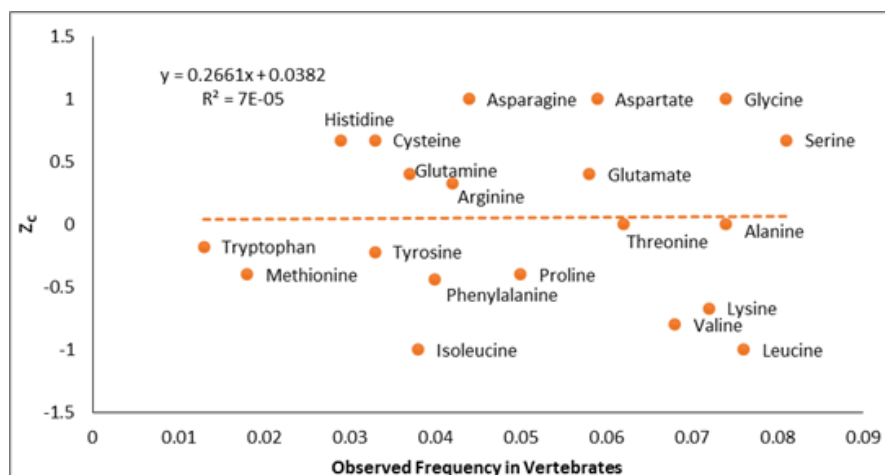


Figure 2: Correlation between the ZC score of individual amino acids and their observed frequency in vertebrates (values from (King & Jukes, 1969)).

The second example suffers from both methodological and conceptual mistakes. Human membrane proteins are compared with the whole proteome and differences found in the average ZC score explained as being due to the redox state of the lipid environment in which membrane proteins are located. There is a significant difference between the composition of *membrane proteins* and that of *transmembrane domains*, the latter limiting the analysis to those (often minority) protein stretches effectively immersed in the lipid bilayer. Also, the correlation, albeit weak, between ZC and hydrophobicity of residues discussed above (and highlighted in the published paper as the first finding) causes an intrinsic bias in the method used. The observed correlation results from ZC being a proxy for side chain hydrophobicity, which is itself correlated with occurrence in transmembrane domains. These methodological flaws together with omission of the hydrophobic character of both lipids and membrane-spanning amino acids lead to statements like:

"Thus, the proteins located in the membranes are, on average, more reduced than other proteins in humans. A possible implication is that the coexistence of relatively reduced proteins with other relatively reduced biomolecules (lipids) reflects a compositional similarity that would contribute to energy optimization if metabolic pathways for proteins and lipids were operating under common redox potential conditions."

It is not clear why the redox potential for the synthesis of lipids and proteins should be more relevant than the chemical compatibility that is required to generate a lipophilic environment (i.e. to exclude water, hence generate a *compartment*). Acknowledging the bias in ZC calculation, and following a rather intuitive logic to explain the preferential distribution of amino acids in transmembrane domains

(supported by evidence(Ulmschneider & Sansom, 2001)), allows rewriting the previous sentence as one which would be considered general knowledge:

Thus, the proteins located in the membranes are, on average, more hydrophobic than other proteins in humans. A possible implication is that the coexistence of relatively hydrophobic proteins with other relatively hydrophobic biomolecules (lipids) reflects a compositional similarity that would contribute to structural stability.

Similarly, the whole analysis of subcellular localisation is flawed. It is already not clear why subcellular localisation should be linked to proteins' ZC scores. Not only is redox state defined at the level of amino acid (and not protein) biosynthesis, but proteins are synthesized essentially in two compartments only, and are then transported to their destination. Hence, any correlation between the ZC score of a protein and the redox state of its final compartment is an artefact. More likely, and again a result of the bias in the ZC score formula, the observed correlation mirrors the increase in the polarity of secreted proteins to ensure their unassisted solubility in the extracellular environment.

It is suggested that the environment (temperature and redox potential) might play a role in altering the average oxidation state of fixed carbon by providing an easily-accessible source of reducing power. However, these energy calculations all assume starting materials of CO₂, H₂O, NH₃, H₂S and H⁺ at fixed concentrations, and are poorly representative of real biosynthetic processes where for most organisms biosynthesis actually begins from moderately oxidised sugar molecules. Inconsistencies in the choice of the organisms are common: *M. burtonii* is a methylotroph, living on methylamines and methanol, not CO₂. *B. japonicum* is a nitrogen fixing root symbiont, so uses N₂ and as a carbon source consumes sugars from the plant, not CO₂. *T. ferrooxidans* (since 2000 renamed as an acidithiobacillus) lives on iron and sulfur and does fix carbon, but also N₂ and likes pH 4.5 - 1.3, not 7.

ON IDEAS, METHODS AND PUBLISHING

New ideas that contribute to scientific discussion should always be supported and welcomed. In this case, it could be interesting to look at the role, if any, minimisation of energetic cost has played throughout evolution, but it would be critical for such an analysis to be properly parameterized (i.e. accurately modelling the external (including energy inputs and feed sources) and internal environments of the cell). This would be a significant challenge, and would have to take into consideration that evolution can only change in a stepwise fashion from a defined starting point and also account for some degree of stochasticity, which is pervasive in life. As it should be evident by now (pending further discussion and criticism to this letter), the reported analysis fails to deliver a usable and biologically meaningful metric to do so. Also, none of the essential elements mentioned above seem to feature in Dick's analysis.

Moving beyond the specific assessment of methods and results, the publication of a paper like the one here discussed raises a few additional questions, which are listed below and left unanswered for open discussion.

1. Is a reductionist approach that uses a chemical quantity to measure (molecular) evolution of general applicability?
2. To which extent should a new idea be stretched before it becomes un-defendable? What is the most suitable platform for proposing new ideas in science?
3. How does a journal set the balance between fostering new ideas and vouching for scientific soundness? How does this responsibility partition between editors and anonymous peer-reviewers?
4. How much time should be devoted to a disappointing paper before it makes more sense to let it survive unharmed in the scientific literature?

Authors, editors, reviewers and non-specialist readers are invited to discuss these points.

REFERENCES

Dick, J. M. (2014). Average oxidation state of carbon in proteins. *J. R. Soc. Interface*, 20131095. doi:10.1098/rsif.2013.1095

King, J. L., & Jukes, T. H. (1969). Non-Darwinian Evolution. *Science*, 788-798. doi:10.1126/science.164.3881.788

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol.*, 105-32. doi:10.1016/0022-2836(82)90515-0

Ulmschneider, M. B., & Sansom, M. S. (2001). Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1-14. doi:10.1016/S0005-2736(01)00299-1