OPINION NOTE

Thermostable and promiscuous Precambrian proteins

Valeria A. Risso,[1] Jose A. Gavira[2] and Jose M. Sanchez-Ruiz[1]*

[1] *Departamento de Quimica Fisica, Facultad de Ciencias, Universidad de Granada, 18071-Granada, Spain.*

[2] *Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra (Consejo Superior de Investigaciones Científicas – Universidad de Granada), Avenida de la Palmeras 4, Armilla, 18100-Granada, Spain.*

* To whom correspondence and proofs must be addressed:
Jose M. Sanchez-Ruiz
Departamento de Quimica Fisica
Facultad de Ciencias
Campus Fuentenueva s/n
18071-Granada, Spain
TEL: 34-958243189
FAX: 34-958272879
E-MAIL: sanchezr@ugr.es

Running title: Precambrian proteins

Abstract: We discuss the unique properties found and the new and exciting possibilities implicit in recent laboratory resurrections of Precambrian proteins.

Pauling and Zuckerkandl noted fifty years ago that the sequences of a given protein in several extant (i.e., modern) organisms could potentially be used to derive a reasonable approximation to the sequence of the same protein in their common ancestor (Pauling and Zukerkandl, 1963). The field of "chemical paleogenetics" they envisioned was not immediately recognized by the scientific community. Certainly, the scarcity of known primary structures made the reconstruction of ancestral sequences a far-fetched goal in the sixties. It has also been noted (Harms and Thornton, 2013) that molecular approaches to evolution were originally proposed by chemists (the likes of Linus Pauling and Margaret Dayhoff) and were not initially well received by evolutionary biologists, who at the time focused mainly on organisms and populations (Aronson, 2002). Starting in the nineties of the past century, changes in attitude towards molecular evolution studies, developments in Bioinformatics and the genomics-era availability of increasing numbers of protein sequences, have contributed to make ancestral sequence reconstruction an accepted evolutionary tool (Liberles, 2010; Mirceta *et al.*, 2013).

Several widely available programs can currently compute estimates of the sequences at the internal nodes of a phylogenetic tree from the extant sequences upon which the tree is based. The simplest approach used is called "maximum parsimony" and selects as the ancestral amino acid at each given position that which minimizes the number of mutational changes required to account for the amino acid residues present at the same position in the extant protein sequences used for phylogenetic analysis. More sophisticated approaches model molecular evolutionary change in terms of a continuous-time Markov chain, employ an empirical rate matrix (derived from the previous analysis of a large sequence data

set) for the probabilities of amino acid substitutions and typically select the ancestral state at each position on the basis of the maximum likelihood criterion (i.e., the selected ancestral amino acid maximizes the probability of observing the amino acids present in the extant proteins). Actually, these probabilistic approaches to sequence reconstruction do not return just a single ancestral amino acid for each given position, but provide a vector with the "probabilities of being the ancestral state" for the 20 natural amino acids; the set of most probable amino acids thus define what is known as the "most probabilistic" ancestral sequence. The reader is referred to the published literature (Liberles, 2010) for more detailed descriptions of ancestral sequence reconstruction algorithms and for discussions on relevant related issues (Bayesian inference versus maximum likelihood in ancestral reconstruction, procedures to account for rate variations among sites or parts of the phylogenetic tree, etc.). At a very basic level, however, an intuitive feeling about the overall process is gained by recalling that an amino acid sequence can be regarded as a word written using an alphabet of twenty letters. Ancestral protein sequence reconstruction is then akin to the reconstruction analyses commonly used in historical linguistics to obtain approximations to words in an ancient extinct language (a proto-language) from the corresponding words in several descendant modern languages (for a recent example, see Bouchard-Côte *et al.*, 2013). There is, nevertheless, a fundamental difference between protein sequences and language words in this context: while reconstructions of extinct languages will necessarily remain untested hypotheses, Molecular Biology methodologies allow the actual preparation in the laboratory of the proteins encoded by the reconstructed sequences, what is generally known in the field as "resurrecting" the ancestral proteins. In this way, a first validation of

the sequence reconstruction process becomes possible, since the "laboratory-resurrected" proteins are expected to display reasonable properties in terms of structure, stability and biological function. In fact, it is the set of relevant protein proteins (the so-called protein phenotype) what is targeted in experimental ancestral resurrection studies. For this reason, several reconstructed sequences for a given ancestral node (arising from different phylogenetic trees showing variation below the node, from different random samplings of the posterior probability distribution produced by the reconstruction program used, etc.) are customarily subjected to laboratory resurrection in order to test phenotypic robustness against the (unavoidable) uncertainties in sequence reconstruction (for recent examples see: Risso *et al.*, 2013; Akanuma *et al.*, 2013).

In the last twenty years or so, laboratory resurrection of ancestral proteins has been used to address many important issues in evolution, including the adaptation of proteins to varying environments over planetary time scales (Gaucher *et al.*, 2008; Perez-Jimenez *et al.*, 2011; Risso *et al.*, 2013), the mechanisms of generation of new protein functions (Voordeckers *et al.*, 2012) and the origin of complexity in biomolecular machines (Finnigan *et al.*, 2012). The reader is referred to recent reviews (Benner *et al.*, 2007; Harms and Thornton, 2010) for readable accounts of these and other interesting applications of ancestral protein resurrection. Here, we will only emphasize two important points:

First, the properties determined for laboratory resurrections of ancestral proteins often lead to coherent evolutionary narratives that correlate the molecular and paleontological records of life and reveal adaptations to environmental changes over planetary time scales. One well-known paradigmatic example will suffice to illustrate the point. Ruminant digestion arose about 40

million years ago probably linked to the lowering of temperatures at the end of the Eocene. This climate change likely led to the widespread emergence of grasses, a source of food for which ruminant digestion is advantageous. Digestion of the fermenting microorganisms in the foregut of ruminants generates large amounts of RNA, thus creating the need for abundant and efficient digestive ribonucleases (Barnard, 1969). Laboratory resurrection of the ancestral ribonucleases for the artiodactyl lineage was reported by Benner and coworkers in the nineties (Jermann *et al.*, 1995). They found the resurrected enzymes to display the properties expected for digestive ribonucleases (resistance to proteolysis, high activity towards single-stranded RNA and small RNA substrates, low activity towards double-stranded RNA), but only up to the last common ancestor of ruminants. Actually, enzymes corresponding to older evolutionary nodes displayed a 5-fold increase in activity towards double-stranded RNA (a non-digestive substrate), a 5-fold decrease in activity towards single-stranded RNA and small RNA (the digestive substrates), an enhanced binding to double-stranded DNA and a decreased resistance to proteolysis, supporting that digestive ribonucleases arose from a non-digestive ancestor that was recruited for ruminant digestion about 40 million years ago (Benner *et al.,* 2007; Jermann *et al.*, 1995). Actually, the ancestral resurrection exercise of Benner and coworkers anticipated the existence of non-digestive ribonucleases with interesting and "unusual" activities, of which may examples are currently known (Pizzo and D'Alessio, 2007). Certainly, a "direct" validation of these reconstruction/resurrection efforts is scarcely possible, as preservation of usable millions-years-old samples of protein or DNA is expected to be highly uncommon (Schweitzer, 2011; Penney *et al.*, 2013). Nevertheless, the study of Benner and coworkers on ribonucleases (Jermann *et al.* , 1995) and other

similar experimental analyses (Benner *et al.*, 2007; Harms and Thornton, 2010) show that, although the reconstruction of ancestral sequences is unavoidably uncertain to some extent, it is to some significant degree validated at the phenotypic level by the fact that the properties of the proteins resurrected in the laboratory are typically robust and consistent with the ancestral properties expected from coherent evolutionary narratives.

Second, starting with the work of Gaucher and co-workers on elongation factors published in 2008 (Gaucher *et al*. 2008), several ancestral resurrections studies (Perez-Jimenez *et al.,* 2011; Risso *et al.*, 2013; Ingles-Prieto *et al.*, 2013; Akanuma *et al.*, 2013) have successfully targeted billions-years-old Precambrian nodes (1 billion years = 1 thousand million years). Figure 1 shows a simple representation of the geologic time scale. The huge diversity of animal life we can see almost everywhere around us originated in the Cambrian explosion of life about 540 million years ago (Gould, 1989) and most fossils displayed in museums actually belong to the last few hundred million years. On the other hand, comparatively little is known about Precambrian life, although remains of it (stromatolites, microfossils, molecular fossils and chemical isotopic signatures in rocks) certainly exist (Knoll, 2003). Laboratory resurrection of Precambrian proteins provides an excellent opportunity to obtain information about the environment surrounding Precambrian life. Furthermore, reconstructed Precambrian sequences differ extensively from the sequences of the corresponding modern proteins and, therefore, we may expect the encoded proteins (the resurrected proteins) to display unusual and unique properties that depart substantially from those of modern proteins. In the paragraphs below we

briefly describe the unique properties found and the new and exciting possibilities implicit in recent Precambrian protein resurrection studies.

Laboratory resurrection targeting Precambrian nodes has been found to lead to well folded proteins and, in fact, 3D-structure determination by X-ray crystallography has shown that laboratory resurrections of Precambrian thioredoxins (Ingles-Prieto *et al.*, 2013), β-lactamases (Risso *et al.*, 2013) and nucleoside diphosphate kinases (Akanuma *et al.* , 2013) share the canonical fold of the corresponding extant proteins (Fig. 1). These results provide direct evidence of the often-assumed slow evolution of protein structures (as compared with the much faster evolutionary change of amino acid sequences). From a more general viewpoint, it emerges that Precambrian protein resurrection combined with 3D-structue determination may be used to address the many unresolved issues in our understanding of the evolution of protein structures (Ingles-Prieto *et al.*, 2013), such as the origin and age of the different folds or the molecular mechanisms behind the convergent evolution of structures and the transitions between different folds.

For four different protein systems (elongation factors, thioredoxins, β-lactamases and nucleoside diphosphate kinases), increases in denaturation temperature of 30-40 degrees Celsius have been found upon "traveling back in time" several billion years (Gaucher *et al.*, 2008; Perez-Jimenez *et al.,* 2011; Risso *et al.*, 2013; Akanuma *et al.*, 2013) (Fig. 1). Such stability enhancements (with respect to the corresponding modern mesophilic proteins) are enormous, much larger than those typically obtained in engineering/design studies aimed at protein stabilization (Wijma *et al.*, 2013). Clearly, these results are suggestive of protein adaptation to a high temperature environment and support that Precambrian life

was thermophilic, a possibility consistent with several scenarios, including that ancestral oceans were hot (Gaucher *et al.*, 2008), that ancient life thrived in hot spots, such as hydrothermal systems (Lane and Martin, 2013), or perhaps that only tough, thermophilic organisms survived bombardment events in the young planet (Sleep, 2010). High stability is not the only evidence of adaptation to the environment found in Precambrian resurrection studies. Laboratory resurrections of ~4 billion years old thioredoxins were found to display high activity at acidic pH (Perez-Jimenez *et al.*, 2011), plausibly an adaptation to the acidic character of the primitive oceans (presumably linked to high levels of CO_2 in the primitive atmosphere, since CO_2 dissolves in water to give carbonic acid).

Promiscuity, roughly speaking the capability to perform a variety of molecular tasks (Nobeli *et al.*, 2009), may be a common feature in laboratory resurrections of Precambrian proteins. Ancestral resurrection work on β-lactamases (Risso *et al.*, 2013), the enzymes responsible for the primary mechanism of resistance towards β-lactam antibiotics, provides an excellent example. A modern, non-clinical lactamase, such as TEM-1, is a penicillin specialist with high catalytic efficiency for degradation of penicillin antibiotics and rather low efficiency for degradation of, for instance, third generation antibiotics. Laboratory resurrections of Precambrian lactamases, on the other hand, show similar levels of catalytic efficiency for a variety of antibiotics (Fig. 1), i.e., they display substrate promiscuity. Actually, the ancestral catalytic efficiency levels (Risso *et al.*, 2013) are similar to those of an average modern enzyme (Bar-Even *et al.*, 2011) and 2-3 billion years old lactamases can be considered as efficient promiscuous enzymes. This ancestral efficient promiscuity can be explained if ancient bacteria produced a variety of ß-lactam antibiotics as, for instance, a device

to achieve nutrients by killing competitors (Hall and Barlow, 2004) and β-lactamases arose as a mechanism of defense. Another possibility is that the promiscuity of resurrected ancients lactamases simply reflects the generalist properties expected in the early stages in the evolution of enzymes. That is, the laboratory resurrected ancestral lactamases (Risso *et al.*, 2013) provides evidence of the evolutionary conversion of generalists (promiscuous enzymes) into specialists proposed by Jensen many years ago (Jensen, 1976).

Beyond the evolutionary narratives summarized above, it must be noted that hyperstability with promiscuity is a winning combination from a protein-engineering point of view, since both features contribute to high evolvability (Risso *et al.*, 2013). High stability may allow the introduction of functionally useful but destabilizing mutations without impairing proper folding. Attempts to generate new enzyme functions of biotechnological interest through laboratory evolution will, therefore, benefit from employing a hyperstable protein as the starting scaffold (Bloom *et al.*, 2006). Furthermore, significant promiscuous levels may provide essential seeds for the laboratory-evolution of high levels of a targeted function (Nobeli *et al.*, 2009). Clearly, to the extent that the described protein hyperstability and promiscuity are robust ancestral features, laboratory resurrection of Precambrian proteins will have a strong impact in protein engineering and protein biotechnology.

Finally, resurrected Precambrian enzymes appear to be able to work within modern organisms, as shown by the fact that the laboratory resurrections of Precambrian lactamases endow a *E. Coli* strain with resistance towards a variety of antibiotics (Risso *et al.*, 2013) (Fig. 1). This result implies, not only that the ancestral lactamases can degrade antibiotics *in vivo*, but also that they undergo all

the processes and interactions required for their efficient export to the periplasm. For instance, export through the bacterial Sec system (Chatzi *et al.*, 2012) involves interaction with chaperones that keep the preprotein in a non-native state and target it to the translocase machine, as well as, upon translocation, cleavage of the signal peptide by a protease and folding of the mature protein in the periplasm. Certainly, the capability of resurrected Precambrian proteins to function *in vivo* is far from being a trivial result. It immediately suggests, in fact, a number of very interesting possibilities, such as the realization of laboratory replays of the molecular tape of protein evolution (Sanchez-Ruiz, 2012) or the synthetic biology exploitation of the unique properties of Precambrian proteins. Furthermore, it would appear at least conceivable that in the near future ancestral resurrection studies will be extended beyond the one-protein level and that functional networks of interacting ancestral proteins will be successfully integrated in modern organisms. Such efforts could be viewed as the first steps towards the laboratory resurrection of reconstructed ancestral microorganisms.

Acknowledgements

Work in the authors' labs is supported by grants BIO2012-34937, CSD2009-00088 (to J.M.S.-R.), BIO2010-16800, "Factoría Española de Cristalización" Consolider-Ingenio 2010 (to J.A.G.) from the Spanish Ministry of Economy and Competitiveness, and Feder Funds (J.M.S.-R. and J.A.G.). The authors declare no conflict of interest.

References

Akanuma, S., Nakajima, Y., Yokobori, S., Kimura, M., Nemoto, N., Mase, T., Miyazono, K., Tanokura, M. and Yamagishi, A. (2013) Experimental evidence for the thermophilicity of ancestral life. Proc Nat Acad Sci USA. 110:11067-11072.

Aronson, J.D. (2002) Molecules and monkeys: George Gaylord Simpson and the challenge of molecular evolution. Hist Phil Life Sci 24:441-465.

Barnard, E.A. (1969) Biological function of pancreatic ribonuclease. Nature 221:340-344.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davivi, D., Tawfik, D.S, and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50:4402-4410.

Benner, S.A., Sassi, S.O., and Gaucher, E.A.(2007) Molecular paleoscience: systems biology from the past. Adv Enzymol Relat Areas Mol Biol 75:1-132.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006) Protein stability promotes evolvability. Proc Natl Acad Sci USA 103:5869-5874.

Bouchard-Côte, A., Hall, D., Griffiths, T.L., and Klein, D. (2013). Automated reconstruction of ancient languages using probabilistic models of sound change. Proc Natl Acad Sci USA 110:4224-4229.

Chatzi, K.E., Sardis, M.F., Karamanou, S., and Economou, A. (2012) Breaking on through to the other side: protein export through the bacterial Sec system. Biochem J 449:25-37.

Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W. (2012). Evolution of increased complexity in a molecular machine. Nature 481:360-364.

Gaucher, E.A., Govindarajan, S., and Ganesh, O.K. (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451:704-707.

Gould, S.J. (1989) Wonderful Life. The Burgess shale and the nature of history. New York, USA: Norton.

Hall, B.G., and Barlow, M. (2004) Evolution of the serine beta-lactamases: past, present and future. Drug Resist Updat 7:111-123.

Harms, M.J., and Thornton, J.W. (2010) Analyzing protein structure and function using ancestral gene resurrection. Curr Opin Struct Biol 20:260-366.

Harms, M.J., and Thornton, J.W. (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat Rev Genetics 14:559-571.

Ingles-Prieto, A., Ibarra-Molero, B., Delgado-Delgado, A., Perez-Jimenez, R., Fernandez, J.M., Gaucher, E.A., Sanchez-Ruiz, J.M., and Gavira, J.A. (2013). Conservation of protein structure over four billion years. Structure 21:1-8.

Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. Ann Rev Microbiol 30:409-425.

Jermann, T.M., Opotz, J.G., Stackhouse, J., and Benner, S.A. (1995) Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature 374:57-59.

Knoll, A.H. (2003) Life on a young planet: the first three billion years of evolution on Earth. Princeton, New Jersey, USA: Princeton University Press.

Lane, N., and Martin, W.F. (2013) The origin of membrane energetics. Cell 151:1406-1416.

Liberles, D.E. (2010) Ancestral sequence reconstruction. Oxford, UK: University Press.

Mirceta, S., Signore, A.V., Burns, J.M., Cossins, A.R., Campbell, K.L., and Berenbrink, M. (2013) Evolution of mammalian diving capacity traced by myoglobin net surface charge. Science 340:124192.

Nobeli, I., Favia, A.D., and Thornton, J.M. (2009) Protein promiscuity and its implications for biotechnology. Nat Biotechnol 27:157-167.

Pauling, L., and Zuckerkandl, E. (1963) Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life. Acta Chem Scan 17:S9-S16.

Penney, D., Wadsworth, C., Fox, G., Kennedy, S.L., Preziosi, R.F., and Brown, T.A. (2013) Absence of ancient DNA in sub-fossil insect inclusions preserved in 'Anthropocene' Colombian copal. PLoS ONE 8(9):e73150.

Perez-Jimenez, R., Ingles-Prieto, A., Zhao, Z.M., Sanchez-Romero. I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T.J., Tanokura, M., Holmgren, A., Sanchez-Ruiz, J.M., Gaucher, E.A., and Fernandez J.M. (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat Struct Mol Biol 8:592-596.

Pizzo, E., and D'Alessio, G. (2007) The success of the RNase scaffold in the advance of biosciences and in evolution. Gene 406:8-12.

Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. J Am Chem Soc 35:2899-2902.

Sanchez-Ruiz, J.M. (2012) On promiscuity, changing environments and the possibility of replaying the molecular tape of life. Biochem J 445:e1-e3.

Schweitzer, M.H. (2011) Soft tissue preservation in terrestrial Mesozoic vertebrates. Annu Rev Earth Planet Sci 39:187-216.

Sleep, N.H. (2010) The Hadean-Archaean environment. Cold Spring Harb Perpesct Biol 2:a002527.

Voordeckers, K., Brown, C.A., Vanneste, K., van der Zande, E., Voet, A., Maere, S., and Verstrepen, K.J. (2012). Reconstruction of ancestral metabolic enzymes reveals

molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS Biology 10, e1001446.

Wijma, H.J., Floor, R.J., and Janssen, D.B. (2013). Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. Curr Opin Struct Biol 23:588-594.

Legends to the figures

Figure 1.- Changes in structure, stability and function of β-lactamases over a planetary time scale as inferred from ancestral resurrection targeting Precambrian nodes (Risso *et al.* , 2013). The laboratory-resurrected proteins correspond to the last common ancestors of enterobacteria (ENCA), gammaproteobacteria (GPBCA), various gram negative bacteria (GNCA) and various gram-positive and gram-negative bacteria (PNCA). Data for the extant (modern) TEM-1 β-lactamase are also included. *In vitro* function is given in terms of the Michaelis-Menten parameters for antibiotic degradation while *in vivo* function was assessed from the antibiotic minimum inhibitory concentrations (MIC) for a *E. coli* strain transformed with a plasmid containing the ancestral proteins. Denaturation temperatures were obtaineda by scanning calorimetry applying overpressure to increase the boiling point of water (since some of the ancestral denaturaturaion temperatures are above 100 ºC). 3D structures were determined by X-ray crystallography. See Risso *et al.* for details.