



Richard Dickerson, Molecular Clocks, and Rates of Protein Evolution

David Alvarez-Ponce¹

Received: 4 October 2020 / Accepted: 7 November 2020 / Published online: 18 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Proteins approximately behave as molecular clocks, accumulating amino acid replacements at a more or less constant rate. Nonetheless, each protein displays a characteristic rate of evolution: whereas some proteins remain largely unaltered over large periods of time, others can rapidly accumulate amino acid replacements. An article by Richard Dickerson, published in the first issue of the *Journal of Molecular Evolution* (J Mol Evol 1:26–45, 1971), described the first analysis in which the rates of evolution of many proteins were compared, and the differences were interpreted in the light of their function. When comparing the sequences of fibrinopeptides, hemoglobin, and cytochrome *c* of different species, he observed a linear relationship between the number of amino acid replacements and divergence time. Remarkably, fibrinopeptides had evolved fast, cytochrome *c* had evolved slowly, and hemoglobin exhibited an intermediate rate of evolution. As the *Journal of Molecular Evolution* celebrates its 50th anniversary, I highlight this landmark article and reflect on its impact on the field of Molecular Evolution.

Keywords Richard Dickerson · Molecular clock · Evolutionary rate

Introduction

The field of Molecular Evolution came into existence in the 1960s, when scientists started to gather the first sets of protein sequences and structures from different organisms, enabling comparative studies. In 1971, the *Journal of Molecular Evolution* was created to serve the community of scientists working on that emerging field of inquiry. To commemorate the 50th anniversary of the Journal, each associate editor has been invited to highlight one classical paper from the Journal and comment on its Significance and subsequent impact.

The paper that I have chosen, entitled “The structure of cytochrome *c* and the rates of molecular evolution,” was published in the very first issue of the Journal (March 1971). In this article, Richard Dickerson compares the sequences of cytochrome *c*, hemoglobin, and fibrinopeptides (as well as other proteins) from different species to infer their rates of evolution and proposes explanations for why some of these

proteins evolve fast whereas other remain largely unaltered during long evolutionary periods (Dickerson 1971).

The Author

Richard E. Dickerson (born 1931) obtained his Bachelor's degree in Chemistry from the Carnegie Institute of Technology (now Carnegie Mellon University) in 1953 and his PhD in Physical Chemistry from the University of Minnesota in 1957. He then was a postdoctoral researcher at Leeds University and Cambridge University. Subsequently, he was a faculty member at the University of Illinois (1959–1963), the California Institute of Technology (1963–1981), and the University of California, Los Angeles (1981–2004). He was elected as a member of the National Academy of Sciences and the American Academy of Arts and Sciences in 1985.

He made major contributions to the area of Structural Biology. Under the supervision of John C. Kendrew, he determined the first atomic structure of a protein (myoglobin). During his time at the California Institute of Technology, he studied the structure of cytochrome *c* (the paper highlighted here is from that time). At the University of California, Los Angeles, he shifted his focus to DNA, determining the first atomic structure of the B form of DNA. Since

Handling editor: Aaron Goldman

✉ David Alvarez-Ponce
dap@unr.edu

¹ Department of Biology, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557, USA

his retirement in 2004, he writes about the history of his discipline.

Context: Molecular Clocks

Zuckerkandl and Pauling (1962) proposed that proteins from the same family should evolve at a more or less constant rate and used this assumption to date the origin of globins. The following year, Margoliash made the more formal statement that “it appears that the number of residue differences between cytochrome *c* of any two species is mostly conditioned by the time elapsed since the lines of evolution leading to these two species originally diverged.” (Margoliash 1963), and conducted a first test of the molecular clock hypothesis.

One year later, Doolittle and Blombäck (1964) compared a few mammalian fibrinopeptide sequences. For each pair of species, they computed the percent sequence identity, and they obtained the divergence time from the literature. They then represented these numbers in a graph: each data point corresponded to a species pair, the *x*-axis represented divergence times, and the *y*-axis represented the percent of sequence identity. They found a negative correlation between both variables, in support of the molecular clock hypothesis. The relationship appeared to be curved, with sequence identity approaching a plateau, consistent with mutational saturation of nonsynonymous sites.

In the subsequent years, it was debated whether the molecular clock hypothesis was indeed correct (for review, see Morgan 1998; Kumar 2005). Even though many proteins do not evolve under a strict molecular clock (protein evolution can accelerate in certain lineages or slow down in others), proteins tend to evolve at a more or less constant rate. Thus, molecular clocks have proven to be a useful tool to estimate divergence times or rates of sequence evolution.

The Paper

By 1971, from comparison of the sequences of a few proteins in a number of species, it had become apparent that proteins evolved at different rates (see Zuckerkandl and Pauling 1965). However, Dickerson’s work represented the first exhaustive analysis that compared the rates of evolution of multiple proteins, and tried to explain the reasons for the differences.

In his paper, Dickerson elaborated a graph similar to the one by Doolittle and Blombäck (1964), but with some differences (I have reproduced Dickerson’s graph in Fig. 1). First, he used percent differences rather than percent identities. Second, percent differences were converted into percent changes by correcting for multiple amino acid replacements

on the same position. To that end, he used the formula $m/100 = -\ln(1 - n/100)$, where *m* is the number of changes that occurred per 100 residues and *n* is the number of differences observed per 100 residues. Third, a higher number of species were included. Last, and perhaps most importantly, Dickerson analyzed the evolution of not one, but three proteins simultaneously: cytochrome *c*, hemoglobin, and fibrinopeptides. An early version of the graph (with the axes inverted) had been included in a book 2 years before (Dickerson and Geis 1969).

For each protein, he found a linear, positive relationship between divergence time and the percent of sequence differences, in support of the molecular clock hypothesis. In addition, the slope of the regression lines was markedly different for the three proteins: the slope was weak for cytochrome *c* (indicating a slow pace of evolution), steep for fibrinopeptides (indicating fast evolution), and intermediate for hemoglobin (indicating an intermediate evolutionary rate). Following Nolan and Margoliash (1968), he estimated the Unit Evolutionary Period (i.e., the time required for the accumulation of a 1% difference at the amino acid level) to be 20.0 MY, 5.8 MY, and 1.1 MY for cytochrome *c*, hemoglobin, and fibrinopeptides, respectively (i.e., according to his calculations, fibrinopeptides evolved ~18 times faster than cytochrome *c*).

The graph represents an excellent tool to illustrate the molecular clock concept and how different proteins evolve at different rates. Thus, not surprisingly, the graph has been reproduced, sometimes with some modifications, in many textbooks (e.g., Baum et al. 2013; Hamilton 2009; Pevsner 2009; Ruse and Travis 2009; Russell 2003). The divergence times estimates available at the time have since been improved. Nonetheless, a recent reassessment of Dickerson’s work using current divergence time estimates reached the same conclusions (Robinson et al. 2016).

Dickerson attempted to explain the reasons for the different rates of evolution of the different proteins in the light of their functions (see Fig. 2, also borrowed from his paper). He proposed that the low rate of evolution of cytochrome *c* (the main focus of the paper) may be due to the fact that the protein must interact with its reductase and oxidase complexes (these are large complexes, and thus, a large fraction of cytochrome *c*’s surface is used in the interaction; see Fig. 2). In addition, he interpreted the different degree of conservation of the different parts of the protein in the light of the three-dimensional structure of the protein in different species, which he and his colleagues had recently determined (Dickerson et al. 1971, 1972). He also attributed the high rate of evolution of fibrinopeptides to the fact that they are not part of mature fibrin (they are domains of fibrinogen that are excised as fibrinogen is converted into fibrin; Fig. 2), and thus, their amino acid sequences are expected to be under weaker

Fig. 1 Linear relationship between divergence time (x-axis) and percent of amino acid changes (y-axis) for fibrinopeptides, hemoglobin, and cytochrome *c*. The figure illustrates the molecular clock concept and the fact that these three proteins evolve at different rates. Quantities next to the name of each protein correspond to Unit Evolutionary Period estimates. (Figure taken from Dickerson 1971)

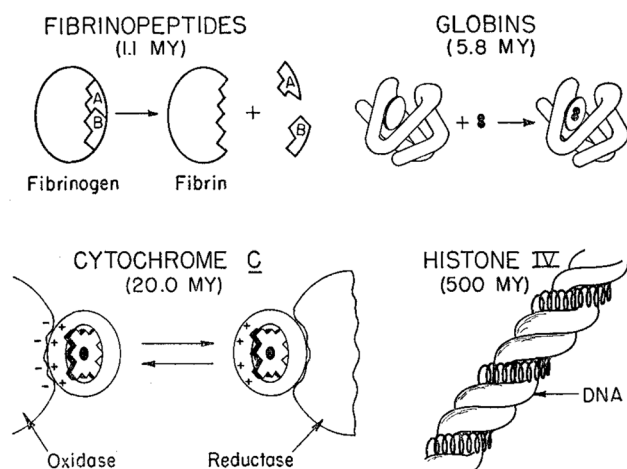
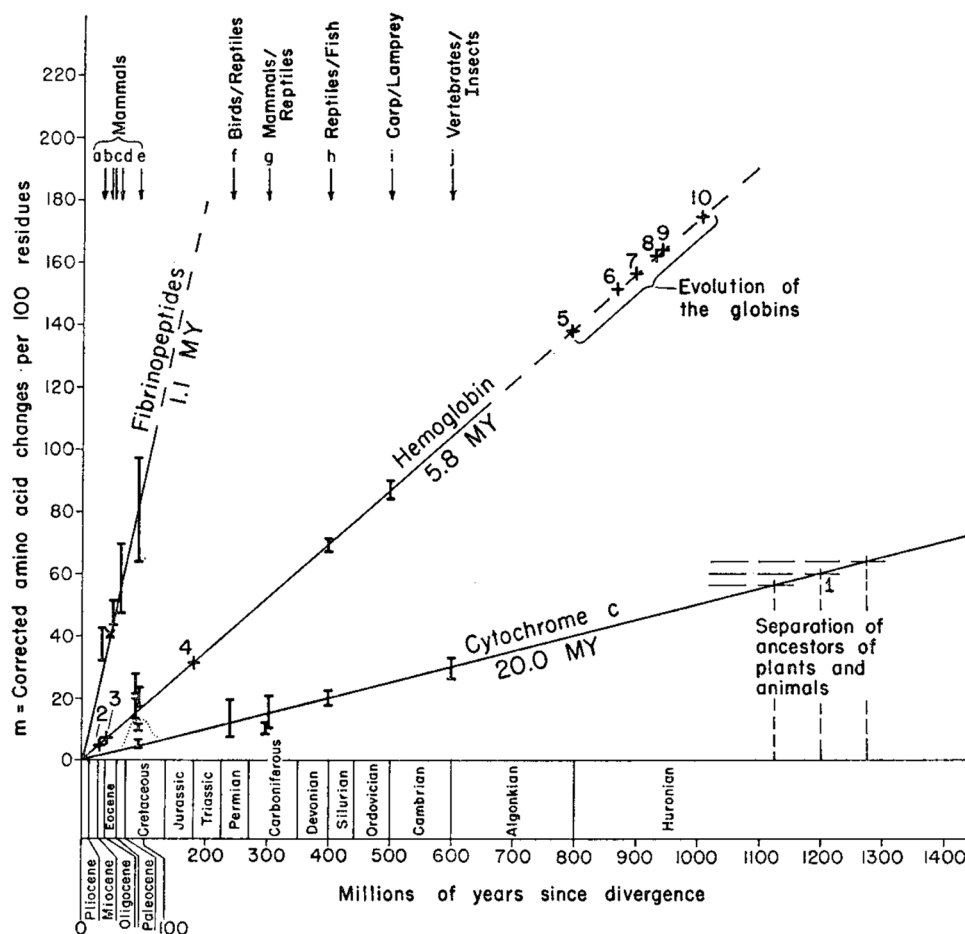


Fig. 2 Mechanisms of action of four proteins. The purpose of the figure was to try to explain the different rates of evolution of each protein. Fibrinopeptides are not part of mature fibrin. Globins use only a small part of their surface to interact with oxygen. Cytochrome *c* uses a large fraction of its surface to interact with its oxidase and reductase complexes. Histone H4 is tightly packaged with DNA. Quantities next to the name of each protein correspond to Unit Evolutionary Period estimates. (Figure taken from Dickerson 1971)

selective constraints. The intermediate rate of evolution of hemoglobin was attributed to the fact that it interacts with O_2 and CO_2 molecules, which are much smaller than the cytochrome *c* reductase and oxidase complexes (Fig. 2).

He also commented on the rates of evolution of a number of other proteins for which much less sequence data were available at the time (thus, he considered his estimates as preliminary). For instance, he estimated the Unit Evolutionary Period of histone H4 to be 500 MY (i.e., according to his calculations, histone H4 evolved 25 times slower than cytochrome *c* and ~ 450 times slower than fibrinopeptides), which he attributed to histone H4's interaction with DNA (Fig. 2). He also noted that the insulin peptide C (which is excised during insulin maturation and thus expected to be under weaker selective constraints at the sequence level) evolves much faster than peptides A and B (which conform the mature protein).

Dickerson also noted that the internal, often hydrophobic parts of proteins tend to be more conserved than the external, often hydrophilic parts. He thus predicted that large proteins, by virtue of their lower surface-to-volume ratio, would tend to exhibit a low overall rate of evolution.

Our Understanding of Rates of Protein Evolution 50 Years Later

In the last 50 years, advances in sequencing techniques have dramatically increased the wealth of protein sequence data, and now many studies of rates of protein evolution encompass thousands of proteins. As a result, we now know that rates of protein evolution vary by orders of magnitude. In addition, the availability of other “omics” datasets has allowed scientists to identify a substantial list of factors that have an impact on rates of protein evolution. Due to space constraints, here I will only comment on some of these factors. For a more comprehensive review, see e Pál et al. (2006), Alvarez-Ponce (2014), and Zhang and Yang (2015).

Many of Dickerson’s intuitions have been confirmed. For instance, we now know that protein-buried residues tend to evolve much slower than those at the surface (e.g., Goldman et al. 1998), and that protein–protein interactions indeed constrain protein evolution (Fraser et al. 2002; Kim et al. 2006; Alvarez-Ponce et al. 2017). However, the relationship between protein lengths and their rates of evolution appears to be more complex than predicted by Dickerson, with some studies finding a positive correlation, others finding a negative correlation, and yet others finding no significant correlation (for review, see Alvarez-Ponce 2014).

Scientists have also identified trends that were hard to foresee 50 years ago. Of note, we now know that a major determinant of rates of protein evolution is gene expression: highly expressed proteins tend to evolve slowly compared with lowly expressed ones (Pál et al. 2001). The leading hypotheses to explain this trend propose that highly expressed genes may be under increased selection to encode proteins that are unlikely to misfold (Drummond et al. 2005) and to misinteract with other molecules (Levy et al. 2012; Yang et al. 2012), and to encode highly stable mRNAs (Park et al. 2013). In addition, in multicellular organisms, another major determinant is expression breadth: genes expressed in many tissues/organs tend to be more conserved throughout evolution than those expressed in fewer tissues/organs (Duret and Mouchiroud 2000).

Other factors affecting rates of protein evolution include gene essentiality (essential genes tend to evolve more slowly than nonessential ones; Hurst and Smith 1999; Alvarez-Ponce et al. 2016), gene duplication (immediately after gene duplication, one of the copies tends to undergo accelerated evolution for a short period of time; Jordan et al. 2004; Pegueroles et al. 2013), chaperone dependency (proteins that interact with chaperones tend to evolve fast, which has been attributed to the fact that chaperones can compensate for mutations that would be otherwise

deleterious; Bogumil and Dagan 2012, Alvarez-Ponce et al. 2019), subcellular compartment (rates of protein evolution are, on average, highest for extracellular proteins, high for membrane proteins, low for cytosolic proteins and lowest for nuclear proteins; e.g., Julenius and Pedersen 2006), protein function (certain categories tend to evolve faster than others; e.g. Greenberg et al. 2008), and position in molecular networks (e.g., Fraser et al. 2002; Alvarez-Ponce 2012). Acceleration of protein evolution can occur by either relaxation of purifying selection or by positive selection. Genes often found to be under positive selection include secreted and cell membrane proteins, and those involved in immunity, host-pathogen interaction, reproduction, and sensory perception (Biswas and Akey 2006; van der Lee et al. 2017).

Different aspects of the structure of proteins have also been linked to their rates of evolution. In general, highly “designable” proteins (those for which many protein sequences are compatible with the function of the protein) are expected to evolve fast. Consistently, proteins with a high contact density, with a high stability, or with disulfide bonds, tend to evolve fast (Bloom et al. 2006a, b; Feyertag and Alvarez-Ponce 2017). Within a protein, amino acids involved in many interactions (intramolecular or intermolecular) tend to evolve slowly (Toft and Fares 2010), whereas intrinsically disordered regions tend to evolve fast (Brown et al. 2002).

Dickerson’s landmark paper greatly advanced our understanding of the fact that proteins evolve at different rates, and of the reasons behind these different rates of evolution. Not surprisingly the paper has received a significant number of citations (as of November 2020, it has been cited over 620 times according to Google Scholar and over 480 times according to the Web of Science). It can be argued that the paper started an important line of inquiry that is still generating important results today.

Acknowledgements Research in my lab is supported by grant MCB 1818288 from the National Science Foundation.

References

- Alvarez-Ponce D (2012) The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution. *BMC Evol Biol* 12(1):192
- Alvarez-Ponce D (2014) Why proteins evolve at different rates: the determinants of proteins’ rates of evolution. In: Fares M (ed) *Natural selection: methods and applications*. CRC Press, London, pp 126–178
- Alvarez-Ponce D, Sabater-Muñoz B, Toft C, Ruiz-González MX, Fares MA (2016) Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in *Escherichia coli*. *Genome Biol Evol* 8(9):2914–2927
- Alvarez-Ponce D, Feyertag F, Chakraborty S (2017) Position matters: network centrality considerably impacts rates of protein evolution

- in the human protein–protein interaction network. *Genome Biol Evol* 9(6):1742–1756
- Alvarez-Ponce D, Aguilar-Rodríguez J, Fares MA (2019) Molecular chaperones accelerate the evolution of their protein clients in yeast. *Genome Biol Evol* 11(8):2360–2375
- Baum DA, Futuyama DJ, Hoekstra HE et al (2013) The Princeton guide to evolution. Princeton University Press, Princeton
- Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends Genet* 22(8):437–446
- Bloom JD, Drummond DA, Arnold FH, Wilke CO (2006a) Structural determinants of the rate of protein evolution in yeast. *Mol Biol Evol* 23:1751–1761
- Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006b) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103:5869–5874
- Bogumil D, Dagan T (2012) Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* 51:9941–9953
- Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* 55:104–110
- Dickerson RE (1971) The structure of cytochrome *c* and the rates of molecular evolution. *J Mol Evol* 1:26–45
- Dickerson RE, Geis I (1969) The structure and action of proteins. Harper & Row, New York
- Dickerson RE, Takano T, Eisenberg D, Kallai OB, Samson L, Cooper A, Margoliash E (1971) Ferricytochrome *c*: I. general features of the horse and bonito proteins at 2.8 Å resolution. *J Biol Chem* 246(5):1511–1535
- Dickerson RE, Takano T, Kallai OB, Samson L (1972) Ferricytochrome *c*: II. Chain flexibility and a possible reduction mechanism. In: Åkeson Å, Ehrenberg A (eds) Structure and function of oxidation–reduction enzymes. Pergamon, Stockholm, pp 69–83
- Doolittle RF, Blombäck B (1964) Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* 202(4928):147–152
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 102:14338–14343
- Duret L, Mouchiroud D (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol* 17:68–74
- Feyertag F, Alvarez-Ponce D (2017) Disulfide bonds enable accelerated protein evolution. *Mol Biol Evol* 34(8):1833–1837
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. *Science* 296:750–752
- Goldman N, Thorne JL, Jones DT (1998) Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* 149:445–458
- Greenberg AJ, Stockwell SR, Clark AG (2008) Evolutionary constraint and adaptation in the metabolic network of *Drosophila*. *Mol Biol Evol* 25:2537–2546
- Hamilton MB (2009) Population genetics. Wiley-Blackwell, Hoboken
- Hurst LD, Smith NG (1999) Do essential genes evolve slowly? *Curr Biol* 9:747–750
- Jordan IK, Wolf YI, Koonin EV (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol* 4:22
- Julenius K, Pedersen AG (2006) Protein evolution is faster outside the cell. *Mol Biol Evol* 23:2039–2048
- Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314(5807):1938–1941
- Kumar S (2005) Molecular clocks: four decades of evolution. *Nat Rev Genet* 6(8):654–662
- Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci U S A* 109:20461–20466
- Margoliash E (1963) Primary structure and evolution of cytochrome *c*. *Proc Natl Acad Sci U S A* 50:672–679
- Morgan GJ (1998) Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *J Hist Biol* 1:155–178
- Nolan C, Margoliash E (1968) Comparative aspects of primary structures of proteins. *Annu Rev Biochem* 37(1):727–791
- Pál C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158:927–931
- Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. *Nat Rev Genet* 7:337–348
- Park C, Chen X, Yang JR, Zhang J (2013) Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 110:E678–E686
- Pegueroles C, Laurie S, Albà MM (2013) Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. *Mol Biol Evol* 30(8):1830–1842
- Pevsner J (2009) Bioinformatics and functional genomics. Wiley Blackwell, Hoboken
- Robinson LM, Boland JR, Braverman JM (2016) Revisiting a classic study of the molecular clock. *J Mol Evol* 82(2–3):110–116
- Ruse M, Travis J (2009) Evolution: the first four billion years. Belknap, Cambridge
- Russell PJ (2003) Essential iGenetics. Benjamin Cummings, San Francisco
- Toft C, Fares MA (2010) Structural calibration of the rates of amino acid evolution in a search for Darwin in drifting biological systems. *Mol Biol Evol* 27:2375–2385
- van der Lee R, Wiel L, van Dam T, Huynen MA (2017) Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res* 45(18):10634–10648
- Yang JR, Liao BY, Zhuang SM, Zhang J (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci U S A* 109:E831–E840
- Zhang J, Yang JR (2015) Determinants of the rate of protein sequence evolution. *Nat Rev Genet* 16:409–420
- Zuckerkandl E, Pauling LB (1962) Molecular disease, evolution, and genetic heterogeneity. In: Kasha M, Pullman B (eds) Horizons in biochemistry. Academic Press, New York, pp 189–225
- Zuckerkandl E, Pauling LB (1965) Evolutionary divergence and convergence in proteins. *Evol Genes Proteins* 97:166