J Mol Evol
DOI 10.1007/s00239-015-9678-8

LETTER TO THE EDITOR

The Ancient Heart of the Ribosomal Large Subunit: A Response to Caetano-Anolles

Anton S. Petrov[1] · **Loren Dean Williams**[1]

Received: 5 April 2015 / Accepted: 7 April 2015

Abstract Our recent Accretion Model of ribosomal evolution uses insertion fingerprints and a "trunk–branch" formalism to recapitulate the building up of common core rRNA of the Large Ribosomal Subunit. The Accretion Model is a conservative and natural extension of a method developed by Bokov and Steinberg (Nature 457:977–80, 2009), which confirms the correctness of lower resolution models by Fox and others. In each of these models, the LSU originates with the peptidyl transferase center (PTC), consistent with expectations that the ribosome is the source of defined-sequence functional proteins. In an adjacent note, Caetano-Anolles (J Mol Evol 80:162–165, 2015) disparages the Accretion Model, because it controverts the 'Growth Inferred by Genothermal Ordering' (GIGO) model. GIGO analyzes secondary structures, assigns the origin of the ribosome to a region outside of the PTC, and assumes or deduces that (i) large protein enzymes of defined amino acid sequence predate ribosomal synthesis of proteins, (ii) proteins directly replicate by non-ribosomal mechanisms, (iii) rRNA unfailingly increases in thermodynamic stability over time, and (iv) the Woese and Fox canonical tree of life is misrooted. Much of the specific GIGO critique of the Accretion Model is based on confusion about the three-dimensional nature of RNA and trunk–branch polymorphism; the Accretion Model incorporates several types of trunk–branch relationships.

✉ Loren Dean Williams
ldw@gatech.edu;
http://orcid.org/0000-0002-7215-4194

[1] School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA

The Ribosome

There is broad consensus about the centrality of translation in biological systems. It is accepted that the ribosome is the only source of defined-sequence protein in extant (Ogle and Ramakrishnan 2005) and ancestral (Woese 2000, 2001) biological systems. It is accepted that the lineage of the translation system maps out the canonical tree of life (Woese and Fox 1977). It is also generally accepted that the catalytic peptidyl transferase center (PTC) is the oldest part of the large ribosomal subunit (LSU) (Belousoff et al. 2010; Bokov and Steinberg 2009; Fox 2010; Hartman and Smith 2014; Hsiao et al. 2009; Krupkin et al. 2011; Mears et al. 2002; Petrov et al. 2014b; Smith et al. 2008; Wolf and Koonin 2007).

The Accretion Model

The recently described Accretion Model of ribosomal evolution recapitulates the building up of the common core of the LSU rRNA by stepwise additions of ancestral expansion segments (AESs) to a growing rRNA core (Petrov et al. 2014b). The expansions take place at sites marked by 'insertion fingerprints.'

The Accretion Model is a natural extension of an elegant and powerful method developed by Bokov and Steinberg (2009). The Steinberg method uses A-minor interactions (Cate et al. 1996; Nissen et al. 2001) to rank various elements within the ribosome by age. Steinberg made the important observation that, in a complex, the dependent element is the more recent addition. By analogy, the base of a pyramid must be older than the top of the pyramid because the top is dependent on the base.

Using molecular dependencies of A-minor interactions, Steinberg localized the ancestral core of the LSU to the

Published online: 16 April 2015

PTC, confirming the correctness of lower resolution models by Fox and Ashinikumar (2004). We integrated Steinberg's method with our insertion fingerprints and trunk–branch formalism to arrive at the Accretion Model. This integration increased the accuracy and resolution beyond previous models and drilled more deeply into the ancestral core of the ribosome.

Criticism of the Accretion Model

Caetano-Anolles has developed an opposing model (Caetano-Anolles 2002a, b, 2008, 2013; Harish and Caetano-Anolles 2012), which we call the GIGO.

The preceding manuscript in this issue (Caetano-Anolles 2015), in defense of GIGO, submits a series of perceived flaws in the Accretion Model and in the methods by which it was obtained. The author condemns a failure to test predictions of the Accretion Model independent of those of the Steinberg method, in particular, predictions that the PTC is the ancestral core of the LSU. However, the PTC origin of the ribosome, contrary to the claim of the author, was not an a priori hypothesis of Steinberg, but was one of the primary results of the method. Further, an independent test is impossible because the Accretion Model has subsumed the Steinberg method and is dependent on it. Moreover, one sees little gain in testing a theory that is generally accepted and well supported by a broad variety of other data (Belousoff et al. 2010; Bokov and Steinberg 2009; Fox 2010; Hartman and Smith 2014; Hsiao et al. 2009; Krupkin et al. 2011; Mears et al. 2002; Petrov et al. 2014b; Smith et al. 2008; Wolf and Koonin 2007). GIGO is the only opposition to the theory.

Caetano-Anolles (2015) has re-analyzed the insertion fingerprint data and claims errors that are said to support GIGO over the Accretion Model.

The author represents surprise at trunk–branch polymorphism and fails to note that supplementary materials of the Accretion Model show several types of trunk–branch relationships. A branch helix can be inserted into a trunk helix (into a stem) forming a Y, or into a loop, capping the helix and forming a T. Both of these are three-way junctions. There are several examples of helix capping within the common core, along with structurally documented examples within eukaryotic expansions. The specific junction questioned by the author (AES1/AES39) is a helix cap, which he confuses with a mis-ordered stem insertion. Other examples of helix caps are AES9/AES10&10a, AES 21/AES41, and ES30 of eukaryotes.

The author appears not to appreciate that Steinberg's method was incorporated into the Accretion Model to determine time directionality when trunk–branch relationships are ambiguous. The relative ages of some AES were inferred in part from A-minor interactions [which include G-minors (Xin et al. 2008)].

The author criticizes the input information used in the Accretion Model, stating, "In fact, many core insertion sites actually constitute well-characterized 3-way junctions typical of natural folding structures (Lescoute and Westhof 2006)." He does not inform his readers that the Accretion Model manuscript says the same thing and cites the same reference.

The author questions the utility of junctions and insertion fingerprints observed in three-dimensional structures, as used in formulating the Accretion Model. It is therefore ironic that GIGO uses cartoon-level secondary information and establishes rRNA fragments using faux junctions, without incorporating information from three-dimensional structures. GIGO uses growth boundaries that are inconsistent with real structures and assumes, for example, that a 7-way (!) junction in the LSU secondary structure corresponds to physical reality, to a 7-way junction in three dimensions.

The GIGO Method

Predictions of the Accretion Model, and other models of ribosomal origins and evolution, diverge sharply from those of the GIGO. Here, we examine GIGO and explain some of the sources of it's divergence from the consensus.

In GIGO, functional RNAs increase in thermodynamic stability and conformational order over time (Caetano-Anolles 2002b; Caetano-Anolles et al. 2008; Harish and Caetano-Anolles 2012). In GIGO, increasing stability is a primary driver of evolutionary processes. Old RNA is more stable and less polymorphic in conformation than young RNA.

> "During selection, sequence mutants optimize folding to fewer thermally accessible conformations…. This 'lock-in' process of structural canalization is autocatalytic and defines a general evolutionary trend of RNA molecules towards uniqueness, greater stability, and modularity… We here use… increased structural order as being ancestral…" (Harish and Caetano-Anolles 2012).

Histories and lineages in GIGO are based on statistics from secondary prediction of local RNA elements that are extracted from large RNAs. To obtain measures of uniqueness, stability, and modularity, GIGO uses prediction software [the Vienna RNA Package (Hofacker 2003)], which incorporates Turner 2004 nearest-neighbor parameters (Mathews et al. 2004) to model RNA secondary structures. Output from Vienna is used in GIGO to make phylogenetic trees of rRNA fragments. Important

parameters include Vienna predicted stabilities, lengths of helices, numbers of helices, and base-pairing frequencies. These parameters, the essential input of GIGO, are binned and converted to text strings; each RNA fragment is associated with a text string. The text strings are converted to phylograms by maximum parsimony. GIGO assumes that a given large rRNA is composed of small RNA elements that are all related by lineage. Long RNAs are (i) computationally fragmented and folded into local secondary structures, (ii) associated with text strings based on computed folding parameters, and (iii) organized in a system of lineage, in which more stable/more ordered fragments of rRNA are ancestors of less-stable/less-ordered fragments.

Evaluating GIGO

(1) *Internal Inconsistency* In GIGO, there is a general evolutionary trend in which rRNA elements increase in uniqueness, stability, and modularity over time (see above). If so, by conventional understanding of evolution, surviving progeny should be successively more stable than their ancestors. However, in the logic of GIGO, ancestral RNAs are of greater stability than progeny.

(2) *False Assumptions* Several foundational assumptions of GIGO appear to be questionable. One such assumption is that thermodynamic stabilities and/or conformation entropies of rRNA elements (i) change systematically over evolutionary time, (ii) at a rate that is uniform over the population of rRNA elements. Different rRNA elements are subjected to different evolutionary pressures. The stabilities of rRNAs and lengths of helices correlate with growth temperature (Wang et al. 2006), not with general evolutionary age. Stable secondary structures do not increase in frequency over time in a universal way in rRNAs or in other RNAs. GIGO cites Schultes as supporting these proposed changes over evolutionary time (Schultes et al. 1999). However, Schultes concluded that "the majority of the conformational order found in functional RNAs appears not to be the result of a long history of evolutionary modification but is inherent in the physiochemical interactions that drive RNA folding." GIGO predicts that continuous linear rRNA segments are older (more thermostable) than branched rRNA segments (less stable). This model is directly falsified by the youngest elements of rRNA, which are long linear GC rich helices in eukaryotes (ES27).

(3) *Familial Relationships and rRNA Elements* Another questionable assumption of GIGO is that local RNA sequences within long RNAs are related to each other by lineage. In GIGO, large rRNAs are composed exclusively of short RNA sequence elements that are related by mother–daughter relationships. A few short primordial RNA sequences are ancestral to all other sequence elements, which are widely dispersed in secondary structure and three-dimensional space.

(4) *Slicing and Dicing* Dubitable methods are used by GIGO to estimate local thermodynamics of folding. rRNAs, at the level of schematic cartoons, are sliced into fragments (amenable to modeling by Vienna) without respect for the actual structure of the ribosome. Paired strands are disjoined in the fragmenting. Therefore, the predicted secondary structures of GIGO fragments are not reconcilable with known rRNA secondary structures of the rRNA (Petrov et al. 2013; 2014a). In addition, GIGO artificially combines distinct helices (for example H41–H42) to artificially make longer and more "stable" elements.

(5) *Limitations of Thermodynamic Predictions* The thermodynamic parameters obtained from Vienna modeling of GIGO fragments have very little significance for real rRNA folding or stability. Vienna is a powerful platform for predicting secondary structures of simple RNAs, using nearest-neighbor analysis. However, Vienna is neither intended for nor capable of accurately predicting all of the local thermodynamic interactions within complex structures such as ribosomes. rRNAs contain non-canonical base pairs (Leontis et al. 2002), GNRA tetraloops (Hsiao et al. 2006; Mohan et al. 2010; Woese et al. 1990), base–backbone interactions (Lee and Gutell 2004; Leontis et al. 2002), A-minor interactions (Noller 2005), a variety of loops, pseudoknots, kink-turns (Leontis et al. 2006), tertiary interactions (Lescoute and Westhof 2006), coordinated magnesium ions (Hsiao and Williams 2009; Klein et al. 2004) as well as proteins. On average, Vienna accurately predicts 75 % of canonical base pairs (Mathews et al. 2004). Several of the helices in the core of the LSU (Helices 25A and 26A) are composed entirely of non-canonical base pairs (Leontis and Westhof 1998; Petrov et al. 2013) and cannot be predicted by Vienna. Even if the GIGO fragmenting process was accurately performed (see above), these local assemblies cannot be modeled with accuracy by Vienna.

(6) *Circular Argument* In GIGO, the path of evolution of rRNA is predetermined in an obvious way by the input data and the ungrounded theoretical approach. GIGO predicts on average that long helices (as arbitrarily defined in the initial computational fragmenting process) are old and so are ancestral to short

helices. GIGO demonstrates simply that Vienna assigns greater stability to longer helices than to shorter helices.

Grand Claims

Because of the flawed nature of the input data and assumptions, the output of GIGO is a model in which large folded proteins of defined amino acid sequence with complex catalytic functions predate ribosomal protein synthesis (Caetano-Anolles 2013). Proteins, unconstrained by the genetic code, are synthesized by unprecedented mechanisms. Proteins replicate without involvement of nucleic acids. The canonical tree of life is re-rooted within eukarya (Caetano-Anolles 2002b) or between archaea and eukarya (Caetano-Anolles et al. 2008).

Summary

Here, we have highlighted a subset of the weaknesses of GIGO. Even this sampling makes clear that GIGO does not present a credible challenge to the current consensus about the canonical tree of life, the centrality of translation in biological systems, or to the seminal roles of the peptidyl transferase center in ribosomal origins and evolution.

Acknowledgments The authors thank Chad Bernier, Jessica Bowman, Burak Gulen, Hyman Hartman, Nicholas Hud, Timothy Lenz and Roger Wartell for helpful discussions.

References

Belousoff MJ, Davidovich C, Zimmerman E, Caspi Y, Wekselman I, Rozenszajn L, Shapira T, Sade-Falk O, Taha L, Bashan A, Weiss MS, Yonath A (2010) Ancient machinery embedded in the contemporary ribosome. Biochem Soc Trans 38:422–427

Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. Nature 457:977–980

Caetano-Anolles G (2002a) Evolved RNA secondary structure and the rooting of the universal tree of life. J Mol Evol 54:333–345

Caetano-Anolles G (2002b) Tracing the evolution of RNA structure in ribosomes. Nucleic Acids Res 30:2575–2587

Caetano-Anolles G (2013) The phylogenomic roots of modern biochemistry, translation, and the genetic code. J Biomol Struct Dyn 31:14

Caetano-Anolles G (2015) Ancestral insertions and expansions of rRNA do not support an origin of the ribosome in its peptidyl transferase center. J Mol Evol 80:162–165

Caetano-Anolles G, Sun FJ, Wang M, Yafremava LS, Harish A, Kim HS, Knudsen V, Caetano-Anolles D, Mittenthal JE (2008) Origins and evolution of modern biochemistry: Insights from genomes and molecular structure. Front Biosci 13:5212–5240

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA (1996) RNA tertiary structure mediation by adenosine platforms. Science 273:1696–1699

Fox GE (2010) Origin and evolution of the ribosome. Cold Spring Harb Perspect Biol 2:a003483

Fox GE, Ashinikumar KN (2004) The evolutionary history of the translation machinery. In: de Pouplana LR (ed) The genetic code and the origin of life. Kluwer Academic/Plenum Publishers, New York, pp 92–105

Harish A, Caetano-Anolles G (2012) Ribosomal history reveals origins of modern protein synthesis. PLoS One 7:e32776

Hartman H, Smith T (2014) The evolution of the ribosome and the genetic code. Life 4:227–249

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431

Hsiao C, Williams LD (2009) A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. Nucleic Acids Res 37:3134–3142

Hsiao C, Mohan S, Hershkovitz E, Tannenbaum A, Williams LD (2006) Single nucleotide RNA choreography. Nucleic Acids Res 34:1481–1491

Hsiao C, Mohan S, Kalahar BK, Williams LD (2009) Peeling the onion: ribosomes are ancient molecular fossils. Mol Biol Evol 26:2415–2425

Klein DJ, Moore PB, Steitz TA (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. RNA 10:1366–1379

Krupkin M, Matzov D, Tang H, Metz M, Kalaora R, Belousoff MJ, Zimmerman E, Bashan A, Yonath A (2011) A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. Philos Trans R Soc Lond B Biol Sci 366:2972–2978

Lee JC, Gutell RR (2004) Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. J Mol Biol 344:1225–1249

Leontis NB, Westhof E (1998) A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. J Mol Biol 283:571–583

Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res 30:3497–3531

Leontis NB, Lescoute A, Westhof E (2006) The building blocks and motifs of RNA architecture. Curr Opin Struct Biol 16:279–287

Lescoute A, Westhof E (2006) The interaction networks of structured RNAs. Nucleic Acids Res 34:6587–6604

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101:7287–7292

Mears JA, Cannone JJ, Stagg SM, Gutell RR, Agrawal RK, Harvey SC (2002) Modeling a minimal ribosome based on comparative sequence analysis. J Mol Biol 321:215–234

Mohan S, Hsiao C, Bowman JC, Wartell R, Williams LD (2010) RNA tetraloop folding reveals tension between backbone restraints and molecular interactions. J Am Chem Soc 132:12679–12689

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA (2001) RNA tertiary interactions in the large ribosomal subunit: the a-minor motif. Proc Natl Acad Sci USA 98:4899–4903

Noller HF (2005) RNA structure: reading the ribosome. Science 309:1508–1514

Ogle JM, Ramakrishnan V (2005) Structural insights into translational fidelity. Annu Rev Biochem 74:129–177

Petrov AS, Bernier CR, Hershkovitz E, Xue Y, Waterbury CC, Grover MA, Harvey SC, Hud NV, Wartell RM, Williams LD (2013) Secondary structure and domain architecture of the 23S rRNA. Nucleic Acids Res. 41:7522–7535

Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovitz E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, Williams LD (2014a) Secondary structures of rRNAs from all three domains of life. PLoS One 9:e88222

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD (2014b) Evolution of the ribosome at atomic resolution. Proc Natl Acad Sci USA 111:10251–10256

Schultes EA, Hraber PT, LaBean TH (1999) Estimating the contributions of selection and self-organization in RNA secondary structure. J Mol Evol 49:76–83

Smith TF, Lee JC, Gutell RR, Hartman H (2008) The origin and evolution of the ribosome. Biol Direct 3:16

Wang HC, Xia XH, Hickey D (2006) Thermal adaptation of the small subunit ribosomal RNA gene: a comparative study. J Mol Evol 63:120–126

Woese CR (2000) Interpreting the universal phylogenetic tree. Proc Natl Acad Sci USA 97:8392–8396

Woese CR (2001) Translation: in retrospect and prospect. RNA 7:1055–1067

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090

Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal-RNA–constraints on the sequence of tetra-loops. Proc Natl Acad Sci USA 87:8467–8471

Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. Biol Direct 2:14

Xin Y, Laing C, Leontis NB, Schlick T (2008) Annotation of tertiary interactions in RNA structures reveals variations and correlations. RNA 14:2465–2477