Nylon-eating bacteria—part 2: refuting Ohno's frame-shift theory

Royal Truman

Three decades ago Ohno proposed genes arose from simple oligomer repeats. He claimed that a 'T' insertion into gene PR.C consisting of multiple nt30 $(CGACGCCGCT)_3$ produced the ancestor of a new enzyme, E-II, which hydrolyzes Ald, a synthetic material of recent origin. However, our Blastp searches against the NCBI database found no proteins related to the product from PR.C. Using EMBOSS Needle it was not possible to align nt30 unambiguously to any of forty sliding windows 40 nt long from positions 1 through 79 in PR.C. These nt30 alignments were no better than thirty randomly generated sequences also 40 nt long having similar nucleotide distribution as PR.C. The optimal number of nt30 which align over the entire PR.C did not match as well as a simple chain of pure CG dinucleotides. Furthermore, the polypeptide chain derived from nt30 did not align with the PR.C's protein as well as a chain of pure arginines did. These and other tests reveal that the high C+G content of PR.C leads to statistical artefacts and no support for Ohno's frame-shift theory.

Ohno, a leading evolutionary theoretician, proposed several years ago that genes evolved from repetitive DNA sequences.[1] The probability of avoiding a *Stop* codon for *n* nucleotides (nt) is approximately $(61/64)^{n/3}$ assuming each codon is equiprobable. For a small, 100-residue, polypeptide this results in a probability of $(61/64)^{100} = 8.2 \times 10^{-3}$ and $(61/64)^{300} = 5.6 \times 10^{-7}$ for an average size protein.

Ohno attempts to circumvent these low probabilities, pointing out that

> "The situation is far more favorable with regard to repeats of base oligomers. Provided that the number of bases in the oligomeric unit is not a multiple of 3, three consecutive copies of it translated in three different reading frames constitutes the translation unit of such oligomeric repeats. Accordingly, $(61/64)^{3n/3}$ simplified to $(61/64)^n$, the fraction of the repeats of n-base-long randomly generated base oligomers, shall have not one, but all three, open reading frames which equal the total length of repeats—e.g. 59.14% of the monodecameric repeats shall have all three reading frames open for indefinite length."[2]

Although simple repeat genes are not known to code for anything functional and would not be preserved by natural selection, suppose original genes consisted of multiple consecutive copies of 10 nucleotides. For three linked copies in a particular reading frame the probability of avoiding a *Stop* is about $(61/64)^{3n/3} = (61/64)^{10} = 0.62$. Ohno had to assume that for unknown reasons the alternative ORF did not generate mutations over millions of years, which would produce *Stop* codons.

Multiple identical sequences will eventually repeat, so linking monodecameric units which lack a *Stop* would ensure a long uninterrupted Open Reading Frame (ORF). The length of the nucleotide sequence for each of the six reading frames, which would repeat endlessly, is easy to calculate.[3] It would therefore theoretically not be necessary to produce initial genes 100 to 1,000 nt long using random nucleotides in one step.

However, Ohno has avoided the low probability of generating average size genes of about 300 codons without a *Stop* codon at the price of introducing several other low probability requirements.

1. A particular 10 nt sequence is only one out of $4^{10} = 1.048.576$ alternatives. A large number of alternatives would have to be somehow generated, linked together and tested by natural selection in an ancestral life form to determine whether that nt30-based gene worked.
2. Ohno assumed only the same kind of oligomers were linked together and in a suitable number.
3. A functional gene, based on linked 30nt, would have to actually work, without a *Stop* signal at all, and then this *Stop* would have to be added later at a suitable position. It is not clear whether Ohno assumes the genetic code already existed or a preceding form of life based on a different kind of information processing.
4. Single or double nt indels able to produce a *Stop* codon must be avoided during evolution.
5. It is exceedingly unlikely complex protein-based molecular machines could be built from such simple polypeptides. No examples are known of protein-coding genes composed of nothing but identical 10 nt repeats.
6. To be of any evolutionary consequence, the majority of genes would have to be shown to have arisen from shorter repetitive elements.

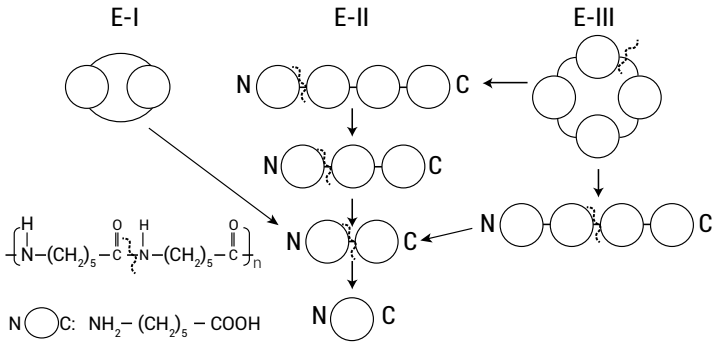

Figure 1. Degradation steps of nylon oligomers by enzymes E-I, E-II, and E-III.

The origin of enzymes E-I, E-II, and E-III

Ohno applied his theory to explain the origin of some genes used by bacteria to degrade side products from the manufacture of nylon-6.[2] Since these synthetic substances did not exist during the lifetimes of the ancestors, the origin of the degrading enzymes has been used extensively by evolutionists as proof that most proteins could have arisen by chance, a line of reasoning disputed in Part 1 of this series.[4]

The enzymes required for the nylon-degrading pathway, E-I, E-II and E-III, catalyze the reactions shown in figure 1.

Genes for enzymes F-E-I, F-E-II, F-E-II' and F-E-III found in *Flavobacterium* sp. 172 were located on the same plasmid, pOAD2. F-E-II and F-E-II' are 88% identical[5] although F-E-II' displayed <1% activity towards 6-aminohexanoate dimer (see figure 1). Neither showed any reactivity toward natural amide compounds tested. The sequences of F-E-II and F-E-III were completely unrelated.

Another bacterial strain, *Pseudomonas* sp. NK87, also possessed enzymes, P-E-I and P-E-II, but the genes were located on different plasmids. Although F-E-II and F-E-II' were also very similar, the homology between the amino acid sequences of F-E-II and P-E-II was only 37%.[6] The authors claimed using evolutionary assumptions that the genes for F-E-II and P-E-II shared a common ancestor at least 1.4 x 10^8 years ago.[7]

These observations show that enzymes able to hydrolyze amide bonds of synthetic materials can be sequentially similar or very different, and are sensitive to the chemical environment surrounding the amide bond being hydrolyzed, but say nothing about how they could have arisen since nylon began to be manufactured in the 1930s.

Yono also searched for long antisense strands lacking *Stop* codons, called Non Stop Frames (NSFs).[8] Remarkably, the antisense strands of the genes for E-II, E-II', and E-III enzymes also contained long NSFs (F-E-II: 392 codons; F-E-II': 392 codons; F-E-III: 394 codons: and P-E-II: 355 codons) and the regions coincide with those of the ORFs. The antisense sequences of the aligned genes for F-E-II, F-E-II', and P-II were even more different, in spite of supposedly having evolved from a common ancestral sequence.[7] They speculated that there is some special mechanism to protect these NSFs from mutations that generate *Stop* codons,[8] but we could not find any homologous proteins in the PIR and Swiss-Prot databases. Neither was Ohno able to identify homologous proteins for F-E-II, F-E-II', P-E-II, and F-E-III.

Putative origin of enzyme E-II

Understanding how the enzymes arose is fundamental to evaluating whether chance or pre-intended adaptability is more reasonable. Much of the debate has centred around Ohno's claim that one of the enzymes, *E-II*, arose from a chance point insertion, modifying the mRNA reading frame and leading to an entirely different protein.[9] This claim is often accepted as fact, and the significance of such a frame-shift argued[10–12] where the claim itself should have been examined more carefully, which we will now do. In Part 3 an alternative origin for E-II will be discussed, which specialists on the topic currently consider the better explanation. This reminds of so many cases in the evolutionary literature where interpretation of data became 'facts' or assumptions.[13]

Ohno's analysis had seemed compelling when published and interesting observations were brought to light. We decided to examine the data to judge whether a frame-shifted gene had arisen, considering that Ohno died in 2000 and thus can't defend his position.

Ohno's Theory for origin of E-II

Enzyme E-II catalyzes the last reaction shown in figure 1 and variants have been isolated from different bacteria.[2] Ohno believed that the E-II isozyme (from gene R-II_A), found on plasmid pOAD2 of *Flavobacterium Sp. K172*,[14] arose from a single T (Thymine) nucleotide insertion in a putative pre-existing gene he called PR.C ("PReexisted Coding sequence").[2] He observed that by removing a T from E-II a long stretch of nucleotides uninterrupted by *Stop* codons results, able to code for a 427-residue polypeptide.[15,16] Insertion of this T at nucleotide position 110 of PR.C (see figure 2) would have generated a *Stop* codon (TGA) in the original sequence and simultaneously produced a chain initiator codon (ATG) for the new 392 residue coding gene R-II_A.

Ohno believed PR.C derived from several repetitive copies 10 nt long $(CGACGCCGCT)_3$ ('30nt'), which would generate identical strings every 30 nt,[17,18] i.e. every ten codons.

This sequence codes for Arg-Arg-Arg-Ser-Thr-Pro-Leu-Asp-Ala-Ala[19] (AA-10), a theory widely accepted.[20] Ohno was convinced the cyclic dimer and linear oligomer hydrolases also originated through frame-shift mutations, and wrote:

> "I suggest that each of these unique genes for degradation of nylon by-products arose *de novo* independently from an alternative reading frame of the pre-existed, internally repetitious coding sequence. In particular, I suggest that the RS-IIA base sequence was originally a coding sequence for an arginine-rich polypeptide chain 427 or so residues long in its length and that the coding sequence for one of the two isozymic forms of 6-ALA LOH arose from its alternative open reading frame."[21]

Search in vain for the proposed PR.C sequence using Blastp

The theory assumes a copy of putative PR.C mutated to form E-II. If true, we should find similar sequences among extant organisms. We used Blastp software and the protein product from PR.C against the entire NCBI dataset but no candidate matches were found. We then limited the searches to 50 residue portions of PR.C.[22] Four of these portions found nothing remotely plausible, one matched poorly to a hypothetical protein, and the remaining three 50-aa sections of PR.C had poor overlap with sections of other proteins. To date no transcripts for PR.C have been reported.

Repeat sequences offered as evidence PR.C existed (see Appendix 2[23])

The existence of several short repeat polypeptide sequences in hypothetical gene PR.C was offered by Ohno as evidence for the repetitive nature of its ancestral gene (Appendix 1[24]).[2] We compared his repetitive patterns to a chain of $(AA\text{-}10)_3$: sequence 1 using the 1-letter abbreviations for amino acids for convenience, and confirmed that a double repeat of A-R-R-R in PR.C was present in a chain of $(AA\text{-}10)_3$. Furthermore, double copies of tetrapeptides R-R-S-G and R-R-R-R matched at three out of four residues, and so did three copies of R-A-A-R. However, in other cases double repeats of tetrapeptides in PR.C have no reasonable region of overlap in $(AA\text{-}10)_3$; for example, G-L-G-G, R-S-A-L, and R-A-A-A.

Ohno also drew attention to two similar decamers,[2] Dec1: R-A-D-R-R-G-A-H-R-S and Dec2: R-L-D-R-R-A-G-H-R-S, but visual inspection and using ClustalX shows insignificant overlap with a chain of linked AA-10, even using an unrealistic number of insertions or deletions: sequences 2 and 3. The repeats are scrambled all over PR.C, and Ohno did not discuss how these could have arisen from a $(AA\text{-}10)_n$ chain. Any mutational pathway would need to avoid creating *Stop* codons and produce functional proteins at every step. The PR.C gene and the resulting protein are *purely* hypothetical, and no similar sequences in other extant proteins were found by him or us.

Does the beginning of PR.C match an nt30? (see Appendix 2)

If PR.C evolved from a chain of nt30s it should be easy to find the first one near the beginning of PR.C unambiguously, as discussed in Appendix 2. We used EMBOSS Needle to align nt30 against sliding windows 40 nt long for PR.C positions 1–79. The forty comparisons had 15 to 23 perfect nt alignments, with an average of 19.9 nt and standard deviation $\sigma = 1.67$. All forty required many indels. None of the alignments used any of the first four nt positions of PR.C (i.e. the algorithm had to treated this region as an insertion).

We compared the quality of nt30 alignments for these 40 sequences using EMBOSS Needle with thirty randomly

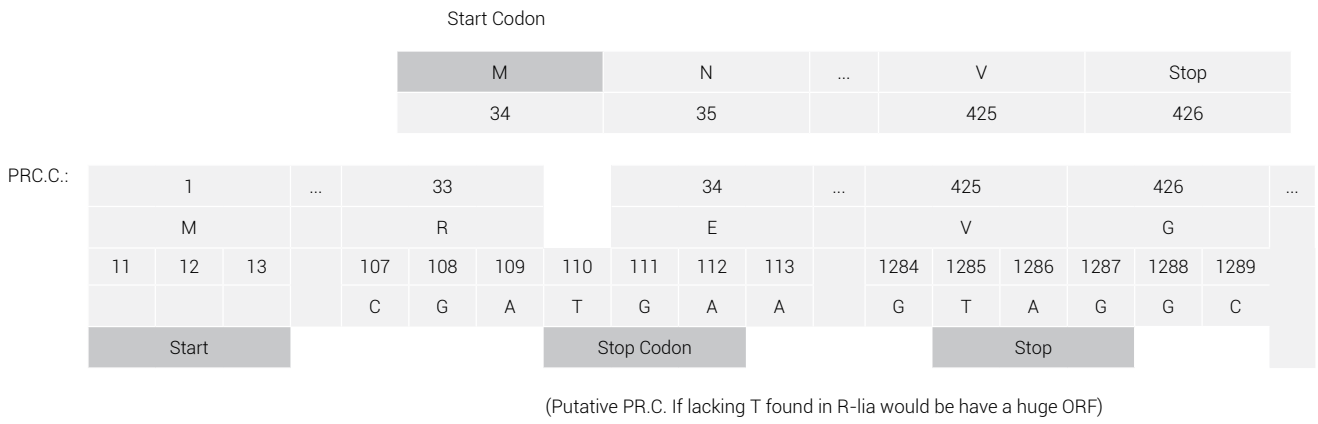

Figure 2. Gene $R\text{-}II_A$ was proposed to have arisen by a T nucleotide insertion in putative precursor gene PR.C, leading to a reading frame-shift. For the full sequence of R-IIA and PR.C see ref. 2.

```
(R-R-R-S-T-P-L-N-A-A)-(R-R-R-S-T-P-L-N-A-A) -(R-R-R-S-T-P-L-N-A-A)          (1)

AA-10: R R R S T P L D A A R R R S T P L D A A                              (2)
Dec1:  - - R - - - - - - - A D R R G A H R S - -

AA-10: R R R S T P L D A A R R R S T P L D A A                              (3)
Dec2:  - - R - - - L D - - - R R A G - - H R S.

PR.C (5-44)  CGA-GCCATGGGCTACATCGATCTCTCCGC-CCCCGTCGCG                      (4)
nt30         CGACGCC----GC----TCGA---CGCCGCTCGACGCCGCT)

Random #27   CCGGCACCGAGCCTGCCGGCACCG-TCG-CGCCGC-                          (5)
nt30         -CGACGCCG---CT--CGACGCCGCTCGACGCCGCT.

Random #11   CCACGCCGGGTGG--GCCGCTC--CTCCGGCT                              (6)
nt30         CGACGCC-GCTCGACGCCGCTCGACGCC-GCT.

Random #15:  CGGCGCCCGTTGGCGCCCGCCAGCT                                     (7)
nt30:        CGACGCCGCTCGACGCCCGCC-GCT.

PR.C (2-41)  CGCCGA-GCCATGGGCTACATCGATCTCTCCGCC-CCCGTCG-----               (8)
30nt         ---CGACGCC----GC----TCGA------CGCCGCTCGACGCCGCT

PR.C (18-57)  ACATCGATCTCTCCGCCCCCGTCG-CGATGATCGTCAGCGG--                  (9)
30nt           ----CGA---CGCCGC-----TCGACGCCGCTCGAC-GCCGCT.

PR.C_AA   RRGYTFGDTR ERTFHRPAPRQVS RSRGRGADTR                             (10)
10_AA(40) RRRSTPLDAA RR---RSTPLDAA RRRSTPLDAA.

PR.C_AA   RRAARRARDA APRSRAAARG DLHRRIPRAA                                 (11)
Rs        RRRRRRRRRR RRRRRRRRRR RRRRRRRRRR.
```

generated sequences having the same proportion of A, C, G, T nucleotides as PR.C. For the random set an average of 18.3 nt aligned with $\sigma = 2.20$.

Remarkably, the best alignments of nt30 occurred with cases from the random sequences. For example, for both data sets a maximum of 23 nt alignments were found but for the randomly generated data fewer indels were necessary: sequences 4 and 5.

Furthermore, notice that alignment 5 has a six-nt continuous region of overlap, whereas the maximum in sequence 4 is only 4 nt.

We also observed that the biologically most plausible best agreement with nt30, taking number of indels and perfect alignments into account, was with one of the randomly generated sequences: sequence 6.

We even find a seven-nt continuous region of overlap in 6.

Statistical coincidences in sequence data can deceive. This 22/32 perfect overlap could easily have been interpreted as evolutionary homology but is pure statistical coincidence.

As another example of the risk of being misled by statistic artefacts, we could assume one of the other randomly generated sequences had undergone a single six-nt deletion during its evolution (a multiple of three, and thus it would not create a frame-shift): sequence 7.

The impression of homology is overwhelming (19/25 perfect alignment with an eight-nt region of perfect alignment). Who would have suspected the top sequence had been generated randomly?

Conclusions from Appendix 2

There was no obvious location within the first 79 nt of PR.C for an initial 30nt sequence. As an example, the best alignment with 30nt for positions 2–41 and 18–57 are as follows: sequences 8 and 9.

Both align at 21 positions and have a comparable number of indels. For all the 40 sliding windows systematically examined, EMBOSS Needle was forced to insert many indels to produce the best alignment. In fact, throughout the entire PR.C sequences this was necessary to permit any semblance of a reasonable alignment.

The highly skewed distribution of nucleotides in 30nt and PR.C (C + G >70%) is responsible for the large number of possible alignments, apprently good fits unavoidable for statistical reasons only. Introduction of indels produced multiple alternative alignments of comparable quality.

Using the measured average number of alignment positions and standard deviation, we cannot reject the null hypothesis H_0 = 'The best alignment of nt30 within positions 1–79 of PR.C is due to chance'.

The analysis so far shows that good alignments between nt30 and PR.C is due primarily to the non-random distribution of nucleotides. We will show below that randomly generated sequences will align even better by optimizing the proportion of GC dinucleotides.

Observations from Appendix 3[25]

Analysis using nucleotides

Since the entire PR.C sequence is supposed to have derived from a chain of dozens of nt30 with no *Stop* codon, several fairly intact nt30 should be easy to find. We aligned varying numbers of 30nt copies systematically against PR.C using the default settings for EMBOSS Needle (figure 3), looking for the best fit.[26] The best candidate for an ancestral PR.C would have consisted of about forty 30nt copies. The optimal alignment had 53.3% nt identity and 30.6% gaps, but none of the forty regions could be interpreted as being highly conserved, far less a complete series of linked nt30.

The high content of nucleotides C and G in both sequences provides many opportunities for different alignments upon adding a few judicial indels. Therefore, we aligned the entire PR.C sequence with the optimal number of only GC doublets, found also by trial and error using EMBOSS Needle. The best overlap had 52.8% nt identity with 25.5% gaps.

Recalling that CG pairs are often used for regulatory purposes, we examined the original PR.C sequence and found a predominance of CG vs GC pairs (202 vs 159). We showed that this is unlikely to be coincidence. Thirty randomly generated sequences of the same length as PR.C (1,287 nt) having the same proportion of A, C, G, T as PR.C were optimally aligned to a long CG chain. As expected, the average number of resulting CG pairs (156.0, $\sigma = 8.24$) and GC pairs (158.9, $\sigma = 9.25$) were almost identical for this random data set.

We also examined the alignment of a chain of pure CGC triplets of optimal length and determined this also aligned well with PR.C: 50.8% identity, 28.0% gaps.

In another experiment we examined whether other good candidates besides pure CG and CGC could be hypothesized in addition to Ohno's nt30 sequence as an evolutionary starting point for PR.C. We reused some of the thirty randomly generated sequences mentioned above (i.e. having a similar proportion of nucleotides as PR.C), and replicated these to the length of the entire PR.C sequence. With little effort we already found one aligned example having 48.6% nt identity using 37.4% gaps with PR.C.

Analysis using amino acids

Ohno claimed the ancestral polypeptide arose from a chain of multiple RRRSTPLDAA. We used the protein

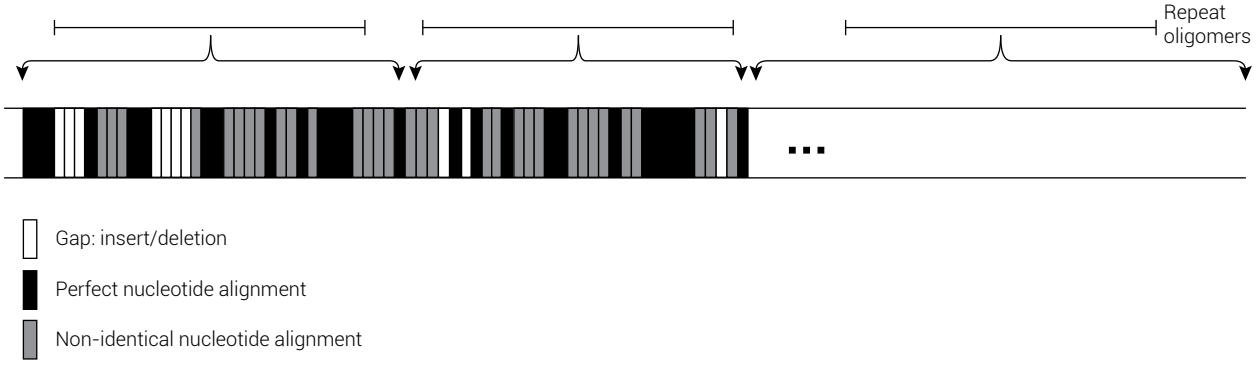

Figure 3. The optimal number of identical 30nt repeats (top bars in black) were optimally aligned throughout the entire PR.C sequence using EMBOSS Needle. The best sequence alignments are shown under the 30nt repeats. See Appendix 2 for details.

version of EMBOSS Needle and found that 40 decapeptides produced the best alignment against the 429 residues coded for by PR.C (PR_C_AA). To help visualize the results, here are the first three aligned blocks: sequence 10.

The most likely first block was not near the origin of PR.C, but began at residue position 24. Over the entire PR.C_AA only 129 of the 429 residues aligned, after introducing 107 indels (23.1% gaps). Over half of the aligned decamers required one or more residue indels. Although seven different amino acids were present in the decamers (R,S,T,P,L,D, and A), only two accounted for most of the alignments: R = 68/129 and A = 24/129.

Further tests demonstrated that a chain of pure arginine (R) aligns considerably better than Ohno's proposed 10-AA chains. To help visualize the results, here is a 30-residue alignment near the beginning (residue position 63) of PR_C_AA: sequence 11.

Sequence 11 was not the best thirty AA aligned region. Although a chain of R and 10-AA both aligned at about 28% positions with PR_C_AA, the alignment with Rs only used about half as many indels (12.6% gaps). A single gap covering positions 1–32 of PR_C_AA represented most of the indels and would be far easier to justify than the multiple gaps all over the alignment found when using 10-AA.

Conclusions from Appendix 3

Since the distribution of nucleotides in PR.C is significantly skewed (A = 15%; C = 37%; G = 33%; T = 15%) and thus 70.16% G + C[27], we expected and demonstrated that good alignments can be found with randomly generated sequences if the nucleotide proportions are similar. The expected number of GC and CG pairs is about 100 x $(0.7016/2)^2$ = 12.3% and for GC the content is as expected, 100 x (159/1286) = 12.4%, but for CG it is significantly higher: 100 x (202/1286) = 15.7%.

Not only does the 30nt oligomer proposed by Ohno have a similar distribution (A = 10%; C = 50%; G = 30%; T = 10) to PR.C, it benefitted from having a 60% to 40% proportion CG to GC pairs, which matches closely the proportion found in PR.C, unlike the randomly generated ones. In addition to the randomly generated blocks 30 nt long, there are countless others which would align better for every separate region within PR.C. Therefore, a repetitive structure for the ancestral PR.C has not been demonstrated since many alternative random sequences of the same length as PR.C would align better after taking the statistical factors into account.

Ohno did not offer any biological or chemical reason why the 10_AA decamer should be reasonable. We demonstrated that better starting point candidates could be proposed which are unlikely to provide any value to an organism, such as a pure CG chain. There is no doubt that many CG pairs are found in PR.C. It is known that these dinucleotides are often used to regulate genes in vertebrates, after cytosines in CpG dinucleotides have been methylated to form 5-methylcytosine CpG Islands.[28] Notice that the CpG section in the figure of this reference[25] resembles closely our results upon aligning a chain of CG against Pr.C.[29] Possibly a regulatory use for CpG remains to be discovered in prokaryotes.[25]

Pursuing the argument that reasonable-looking but biologically absurd alignments could be found, we demonstrated that a gene coding for pure arginine (R) aligns better with PR.C_AA than Ohno's decapeptide. This observation reflects the high C + G content of PR.C, since the codons for R are: (AGA, AGG, CGU, CGC, CGA, CGG). Thirty percent of 10-AA consists of R, which is very close to the proportion found in PR.C, so of course the EMBOSS could align both sequences after adding many indels.

Discussion

In this paper we examined Ohno's frame-shift hypothesis, which claims ancestral genes evolved from short repetitive coding sequences. First of all, we demonstrated that the protein coded for by the hypothetical PR.C gene from which E-II supposedly evolved was not found in the NCBI database.

Statistical artefacts permitted us to offer other sequences which align better with PR.C and PR.C_AA than Ohno's 30nt or AA-10, although these are even more biologically absurd.

It was found that the major part (42%) of Ohno's hypothetical PC.R gene consisted of only two amino acids, Arg (28.3%) and Ala (13.6%), a peculiar protein at best.[30] Six codons code for Arg (AGA, AGG, CGA, CGT, CGC, CGG) of which 13/18 are C or G. For the four Ala codons (GCA, GCC, GCG, GCT), 10/12 are C or G. Since 70.2% of the sequence under study is C + G and not 50% as expected for an equiprobable nucleotide proportion, much higher proportions of Arg and Ala are to be expected merely by chance in the putative PR.C gene.

Reasons for long ORF in PR.C

Note that the lower than expected number of *Stop* codons can be largely explained as coincidence. The three codons (TAA, TAG, TGA) consist of only 2/9 nucleotides C or G, and the start codon (ATG) is 1/3 C + G, and thus will be significantly underrepresented in a sequence of high C + G content. These factors contribute to longer theoretical ORFs being found in C + G-rich sequences.

Ohno calculated the probability of obtaining such long NSFs, taking the high G + C contents into account across all six reading frames for the four enzymes.[31] Assuming the

original gene possessed a 392 codon NSF, using evolutionary assumptions and the estimated time of divergence led to a probability of 0.007.[32] The authors speculated that some unknown mechanism must have been at play for tens of millions of years, which prevented the *Stop* codons from arising in an alternative reading frame, which then led to EII. However, a 0.7% probability is not prohibitive given the large number of organisms possessing high C + G gene contents.

Of course, this could more plausibly be interpreted as strong evidence that such genes have not been mutating for over a hundred million years. Note that the genes producing the enzymes which degrade synthetic substances were not essential for this purpose during the vast period of time, so accidental creation of *Stop* codons in the non-coding frame (the future EII) would not be selected against. Yomo recognizes the difficulty of somehow avoiding *Stop* codons in randomly mutating genes; for example, non-essential duplicate copies:

> "For example, when the homology of the sequences between the duplicated gene and its original one with 392 codons becomes 50%, the probability of the duplicated gene still occurring is only 6.8×10^{-6}. Therefore, we conclude that preparing new NSFs and keeping the NSFs open are basic mechanisms for the occurrence of new enzymes."[33]

For young earth creationists lack of a *Stop* codon might simply mean a few thousand years of mutations was not long enough to generate harmless *Stops* in alternative NSFs. Alternatively, it is possible that mutations able to produce a *Stop* codon in the other reading frame of EII's ancestor would destroy a necessary pattern needed for alternative codes. This line of reasoning is problematic for an evolutionist, since it implies very little tolerance to random mutations, the source of evolutionary opportunity.

Repetitive DNA sequences

We pointed out, based on the analysis presented in Appendix 2, that the evidence for a repetitive 30nt gene precursor is lacking, although some repetitive nt patterns may be present due to designed regulatory elements not related to protein coding.[34]

The chance of random frame-shifts leading to a properly expressed gene, coding for new enzymes like E-I, E-II and E-III, based on properly folded proteins, *which all happen to hydrolyze amides*, is infinitesimally small if by chance. The enzymes are structurally very different, and found on the same or different plasmids and even the bacterial chromosome. Subsequent to Ohno's publications, biodegradation of nylon-6 was reported by Negoro for many other micro-organisms, such as those present in intrauterine contraceptive devices *in vivo* and also biodegradation of nylon-6,6 by a lignin-degrading fungus.[35]

There has been no suggestion of frame-shifted genes nor repetitive structure for these new discoveries. In the case of the lignin-degrading fungus, the activity profile was identical to the reaction catalyzed by horseradish peroxidase, involving removing of a hydrogen radical from the methylene adjacent to the nitrogen. This mechanism is completely different than the breaking of the amide bond, key to the studies described above by Ohno. Degradation of synthetic materials used to manufacture nylon is apparently quite easy to catalyze and no single common ancestral gene needs to be postulated.

Perspective

Highly adaptable micro-organisms able to cope in many environments and perform various ecological services is what the creation model expects.[36] Activation of cryptic information or multiuse DNA coding is certainly sensible if carefully *designed* in advance, and supports Terborg's Baranome Hypothesis (or preformation-hypothesis or front loaded hypothesis).[37]

Creation scientists have used the existence of proteins based on different reading frames as evidence for design, since this happening by chance defies credibility, especially in higher organisms like humans with long generation times.[38,39] However, the genes under discussion here are unlikely to have originated in this manner.[40]

In a recent rebuttal to evolutionary claims in CreationWiki we read, "The frame-shift mutation did not add onto the existing DNA rather it only scrambled what was there and because it is in an environment to adapt to, it worked!"[41] However, a frame-shift is not how this enzyme arose, and Ohno's hypothesis should not be taken as a *fact* which requires an interpretation.

Other creation researchers are aware that the frame-shift theory, for which the evidence initially seemed compelling, is incorrect.[42] There is no reason information could not have been front-loaded,[43] able to be switched on instantly, *contra* the neo-Darwinian view of countless random and mostly non-productive mutations. However, Ohno's incorrect proposal for the origin of nylon-digesting enzymes was accepted uncritically as fact, especially his claim that a repetitive oligomer was the evolutionary starting point. This was spun into a dogmatic story which supposedly proved the overreaching claims by evolutionist theory such as Thwaite's three-decades-old statement that

> "… creationists, and others who should know better are dead wrong about the near-zero probability of new enzyme formation".[44]

References

1. Ohno, S. and Epplen, J.T., The primitive code and repeats of base oligomers as the primordial protein-encoding sequence, *Proc. Natl. Acad. Sci. USA* **80**: 3391–3395, 1983.
2. Ohno, S., Birth of a unique enzyme from an alternative reading frame of the pre-existed, internally repetitious coding sequence, *Proc. Natl. Acad. Sci. USA* **81**:2421–2425, 1984.
3. Suppose the oligomer has a length L. If is consists of a multiple of three nucleotides (exact codons), a series of linked oligomers will repeat every L nucleotides. If L consists of triplet(s) plus one nucleotide (nt), the oligomers will repeat every (L-1)*3 + 3 nt; if L consists of triplet(s) plus two nt, the oligomers will repeat every (L-2)*3 + 6 nt. See: creation.com/Oligomer-repeat-size.
4. Truman, R., Nylon-eating bacteria: part 1—discovery and significance, *J. Creation* **29**(1):95–102, 2015.
5. The putative indels generated to optimize the alignment were ignored. See Yomo, ref. 6, p. 3781.
6. Yomo, T., Urabe, I. and Okada, H., No stop codons in the antisense strands of the genes for nylon oligomer degradation, *Proc. Natl Sci. USA* **89**: 3780–3783, 1992.
7. Yomo *et al.*, ref. 6, p. 3781.
8. Yomo *et al.*, ref. 6, p. 3780.
9. Batten, D., The adaptation of bacteria to feeding on nylon waste, creation.com/the-adaptation-of-bacteria-to-feeding-on-nylon-waste.
10. Thomas, D., New Mexicans for Science and Reason presents Evolution and Information: The Nylon Bug, www.nmsr.org/nylon.htm.
11. Musgrave, I., Nylonase Enzymes, *The TalkOrigins Archive*, 20 April 2004, www.talkorigins.org/origins/postmonth/apr04.html.
12. Thwaites, W.M., New Proteins Without God's Help, *NCSE* **5**(2):1–3, 1985, ncse.com/cej/5/2/new-proteins-without-gods-help.
13. As examples of 'facts' which are actually evolutionary speculation, consider the notion of 'Junk DNA'. Recall statements like the following, made by Richard Dawkins: "Once again, creationists might spend some earnest time speculating on why the Creator should bother to litter genomes with untranslated pseudogenes and junk tandem repeat DNA." "Can we measure the information capacity of that portion of the genome which is actually used? We can at least estimate it. In the case of the human genome it is about 2%". Dawkins, R., The Information Challenge, www.skeptics.com.au/publications/articles/the-information-challenge/.
14. In older literature this strain of *Flavobacterium* had been classified as *Achromobacter guttatus* and was reclassified later as *Arthrobacter*.
15. Ohno, ref. 2, p. 2424 and see: creation.com/R-II-and-PR-C-Overview.
16. See: creation.com/R-II-and-PR-C.
17. Negoro, A., Shibata, N., Tanaka, Y., Yasuhira, K., Shibata, H., Hashimoto, H., Lee, Y-H., Oshima, S., Santa, R., Oshima, S., Mochiji, K., Goto, Y., Ikegami, T., Nagai, K., Kato, D., Takeo, M. and Higuchi., Y., Three-dimensional Structure of Nylon Hydrolase and Mechanism of Nylon-6-Hydrolysis, *J. Biol. Chem.* **287**:5079–5090, 2012.
18. Negoro *et al.*, ref. 17, p. 5080.
19. Ohno, ref. 2, p. 2425.
20. Okamura, K., Feuk, L., Marques-Bonet, T., Navarro, A. and Scherer, S.W., Frequent appearance of novel protein-coding sequences by frameshift translation, *Genomics* **88**:690–697, 2006.
21. Ohno, ref. 2, p. 2423.
22. See: creation.com/Find-similar-PR-C-proteins.
23. creation.com/Appendix2-nt30-oligomer-bootstrapping-from-beginning-of-PR-C.
24. creation.com/Appendix1-Repeat-sequences-in-putative-gene-PR-C.
25. creation.com/Appendix3-Optimal-number-and-alignment-of-30nt-copies-within-PR-C.
26. www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html.
27. See: creation.com/Proportion-of-nts.
28. Wikipedia, 'CpG site', en.wikipedia.org/wiki/CpG_site. CpG islands are generally associated with vertebrates, but Deaton and Bird point out that "In spite of their link with transcription, the functional significance of CGIs is only just beginning to emerge", in Deaton, A.M. and Bird, A., CpG islands and the regulation of transcription, *Genes & Dev.* **25**:1010–1022, 2011.
29. See: creation.com/Proportion-CG-and-GC-in-PR-C.
30. Ohno, ref. 2, p. 2424 and Appendix 2, ref. 23.
31. Yomo *et al.*, ref. 6, p. 3782.
32. Yomo *et al.*, ref. 6, p. 3783.
33. Yomo *et al.*, ref. 6, p. 3784.
34. genome.ucsc.edu/ENCODE/.
35. Negoro, S., Biodegradation of Nylon and other Synthetic Polyamides, *Biopolymers Online*, 2005, see chap. 18, www.wiley-vch.de/books/biopoly/pdf_v09/vol09_14.pdf.
36. Terborg, P., The 'VIGE-first hypothesis'—how easy it is to swap cause and effect', *J. Creation* **27**(3):105–112, 2013, creation.com/images/pdfs/tj/j27_3/j27_3_105-112.pdf.
37. Terborg, P., Evidence for the design of life: part 2—Baranomes, *J. Creation* **22**(3): 68–76, 2008, creation.com/baranomes-and-the-design-of-life.
38. Lian, H. and Landweber, L.F., A genome-wide study of dual coding regions in human alternatively spliced genes, *Genome Res.* **16**:190–196, 2006.
39. Tomkins, J., Gene Complexity Eludes a Simple Definition, www.icr.org/article/8129/.
40. Nylon-eating bacteria, en.wikipedia.org/wiki/Nylon-eating_bacteria.
41. Mutations don't add information, creationwiki.org/Mutations_don't_add_information_(Talk.Origins).
42. Purdom, G. and Anderson, K., A Creationist Perspective of Beneficial Mutations in Bacteria, answersingenesis.org/genetics/mutations/a-creationist-perspective-of-beneficial-mutations-in-bacteria/, 27 May 2009.
43. The chance of a fully random DNA sequence being able to code for a protein able to fold properly (and/or perform a biological service) is negligibly small. But the Creator could select coding sequences for the primary function which when read in another reading frame would also work. The sequences would have been very carefully selected and virtually impossible to recognize using existing statistical tests for randomness. However, if far more frame-shifted genes were found to produce functional proteins than random sequences with the same nucleotide proportion, that could be interpreted as evidence for informational front-loading.
44. Thwaites, W.M., New proteins without God's help, *NCSE* **5**(2):1–3, 1985, ncse.com/cej/5/2/new-proteins-without-gods-help.

Royal Truman *has bachelor's degrees in chemistry and in computer science from State University of New York; an MBA from the University of Michigan (Ann Arbor); a Ph.D. in organic chemistry from Michigan State University; and a two-year post-graduate 'Fortbildung' in bioinformatic from the Universities of Mannheim and Heidelberg. He works in Germany for a European-based multinational.*