Nylonase Genes and Proteins – Distribution, Conservation, and Possible Origins.

S.T. Cordova[1] and J.C. Sanford[2*]

[1]FMS Foundation, 877 Marshall Rd, Waterloo, NY 13165
[2]Horticulture Section, School of Integrative Plant Science, Hedrick Hall, NYSAES, Cornell University
*Corresponding Author – *jcs21@cornell.edu*

ABSTRACT

Nylon comprises a family of man-made substances that were first manufactured in 1935. Nylonases are biological enzymes that can break down nylon oligomers. Although the most prominent nylonases are within the family of enzymes classified as 6-aminohexanoate hydrolases, some enzymes not formally classified as 6-aminohexanoate hydrolases also have the ability to breakdown nylons, and so can also be classified as nylonases.

Organisms that encode a nylonase enzyme do not necessarily have the ability to actually survive on a nylon substrate as their sole carbon source. Among the first documented organisms that did have this ability was the soil bacterium *Arthrobacter* KI72.

It has long been thought that nylonase genes and proteins were essentially absent from the biosphere prior to 1935. This belief led to the widespread assumption that any nylonase gene observed in the present must have emerged since 1935. Several authors developed hypothetical models of how a specific nylonase gene (the *nylB* gene found within *Arthrobacter* KI72), might have arisen very recently as a *de novo* gene.

In this paper we show that the widely-held assumption that all nylonase genes must have evolved very recently is no longer credible. This is in light of the wide-spread distribution of diverse nylonases throughout the biosphere. Likewise, we show that the early speculations regarding the possible *de novo* origin of the *nylB* nylonase gene are no longer credible.

Our review of the literature shows that a variety of nylonase-digesting bacteria have been found in extremely diverse natural environments – far removed from any synthetic nylon sources. In addition, we show there are over 1800 organisms with computationally predicted (provisional*) 6-aminohexanoate hydrolase/nylonase* genes in the NIH-funded UNIPROT database. These 1800 predicted nylonases are not yet experimentally confirmed to cleave nylons, but have significant homology to the experimentally confirmed nylonases.

In addition to 6-aminohexanoate hydrolases, proteases like trypsin and certain lipases have been experimentally demonstrated to have nylonase activity. If we include all proteases and lipases that may have nylonase activity, the number of organisms with nylonase activity may exceed ten thousand. The widespread distribution of nylonases and their homologs strongly suggests that nylonases were already widespread prior to 1935.

Lastly, we have carefully examined the claims that the *nylB* gene arose as a *de novo* gene very recently. The theories of Ohno (the frame shift hypothesis) and Okada (the gene duplication hypothesis), were speculative in nature, and yet were uncritically accepted. In light of new data, these early speculations no longer appear tenable.

INTRODUCTION AND SURVEY OF NYLONASES

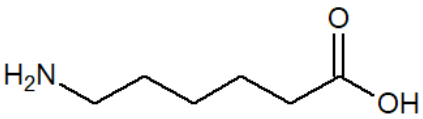
Nylons exist in a variety of chemical forms, which can cause confusion in terms of defining nylonase activity. The nylon of primary interest in this review is nylon-6, and unless otherwise stated the term "nylon" will refer to short nylon-6 oligomers that are generally a hexamer or less in length. Nylon-6's are polymers of de-hydrated lysine-like molecules known as 6-aminohexanoic acids (figure 1).

The 6-aminohexanoic acids have the chemical formula $C_6H_{13}NO_2$, and the dehydrated form of the 6-aminohexanoic acid that constitutes the nylon-6 monomer has the chemical formula $C_6H_{11}NO$. 6-aminohexanoates are the conjugate base of 6-aminohexanoic acids. Terms like 6-aminocaproic acid or ε-polycaproamide are also used to describe nylons. Commercially viable nylon-6's are composed of polymers connecting at least 100 dehydrated 6-aminohexanoic acid monomers.[1] The nylon-6 linear oligomers are terminated by a lysine-like monomer (figure 2). There also exist nylon-6 cyclic oligomers.

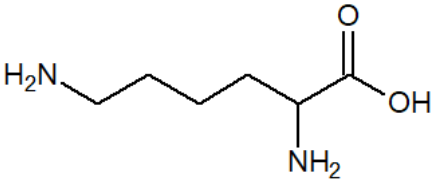
Because this paper may be of interest to a fairly wide range of scientists, we wish to clarify that when we are specifying a nylonase gene we use italicized letters, beginning with lower case (i.e., *nylB*), but when we are specifying nylonase proteins we use non-italicized letters, beginning with upper case (i.e., NylB). This is the standard convention used in this application.

Many biological nylonases (such as the NylB enzyme, which is encoded by the *nylB* gene, which is found within the bacterium *Arthrobacter* KI72), can only degrade nylon dimers and short oligomers[2] where the lysine-like component constitutes a noticeable fraction of the nylon oligomer. Once the oligomer is larger than a hexamer, it causes the lysine-like terminus to occupy a smaller proportion of the total oligomer, and effectively precludes enzymatic action.[3] This suggests the less a nylon oligomer resembles a naturally occurring amino acid such as lysine (and/or the lysine-like terminus is physically prevented access to the enzyme's active site by numerous non-lysine-like nylon monomers as would be the case with a large nylon polymer), the less effective a nylonase will be in cleaving the nylon.

Because most biological nylonases can only degrade short nylon polymers, it is worth emphasizing that the term "nylonase" can be misleading. The term nylonase clouds the fact that most biological nylonases cannot actually digest what are normally considered nylons (commercially viable nylons are polymers of 100 or more monomers).[4] Most "nylonases" can only digest waste product nylons that have already been degraded into short oligomers.

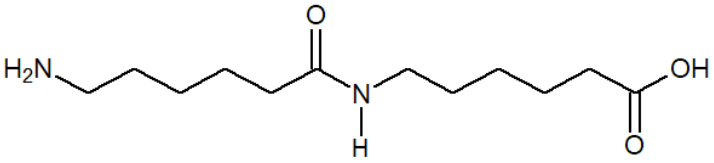


6-aminohexanoic acid

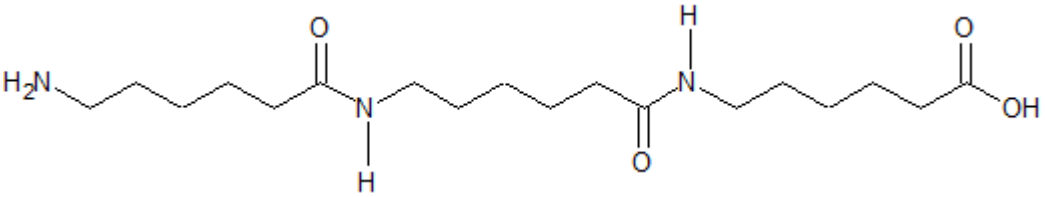


lysine

Figure 1. A man-made 6-aminohexanoic acid molecule is depicted (at the top), which is an analog of a naturally occurring biological lysine molecule (at the bottom). The molecules are shown in a neutral charge configuration. Some references may show a slightly different configuration in an ionized state.



nylon-6 linear dimer



nylon-6 linear trimer

Figure 2. A nylon-6 dimer is depicted at the top and nylon-6 trimer at the bottom. The lysine-analogous 6-aminohexanoate group is on the right side for each molecule. The lysine-like termination group will also be present for higher order nylon oligomers and polymers as well. The molecules are shown in the neutral charge configuration. Some references may show the molecules in a slightly different configuration in an ionized state.

Bonds joining the nylon-6 monomer units are similar to the peptide bonds joining amino acids in a protein. One might expect that the similarity of these bonds might enable a protease to be able to break the bonds in a short nylon oligomer, and indeed this appears to be the case. The earliest report of a nylonase appeared in 1959, when Ebata and Morita experimentally demonstrated that the protease known as trypsin can act as a nylonase.[5,6] Trypsin is a very common protein-cleaving enzyme that appears in vertebrate eukaryotes and various prokaryotes. Unfortunately, Ebata and Morita did not specify what the organismal source was of the trypsin they used in their experiments.

While Ebata and Morita demonstrated proteases could act a nylonases, it was also found that triacylglycerol lipases could act as nylonases. In 2009, Kiumarsi and Parvinzadeh reported hydrolyzing nylon-6 using a lipase known as Lipex50T commercially available from Novozyme (formlerly Novo Nordisk).[7] Unfortunately, data on the degree, if any, that Lipex50T diverges from natural lipases was not readily available. Lipex is a commercially marketed lipase that originated from the LIP gene of

Thermomyces lanuginosus (formerly *Humicola lanuginosa*), by way of directed evolution in the laboratory.

It is important to point out that an organism having nylonase activity may still not be able to metabolize nylon as its sole carbon source. Furthermore, in some cases the nylonase-mediated breakdown of nylon creates 6-aminohexanoic acids, which can be toxic to the organism. Nevertheless, there are some organisms that not only have a nylonase, but can also live on a nylon substrate. In 1966, Fukumura reported such an organism, *Corynebacteria aurantiacum B-2*.[8] Since that time, the inventory of bacteria that are known to digest nylon has steadily increased. One of the most recent examples of this was published in 2016, where Mahdi *et. al*, reported on a thermophilic bacteria with nylonase activity, which was found in the soils of Iraq.[9]

Although various instances of nylon-digesting bacteria have been found, the best known instance of this is *Arthrobacteria* KI72 (formerly *Flavobacteria* sp. KI72, formerly *Acromobacter gutatus* KI72).[10] This bacterium was extracted from soil samples and then tested for ability to grow on nylon waste by-products.[11] KI72's nylon-eating ability was reported by Kinoshita in 1973 and 1975.[12] Even though Kinoshita acknowledged Fukumura's earlier 1996 work, and Fukumura acknowledged Ebata's 1959 work, the Kinoshita papers have been touted as the first discovery of nylonase activity in the biological world. Kinoshita's discovery led to the identification of the first nylonase gene.[13] It appears that Kinoshita believed his nylonase gene was a newly evolved gene that arose due to the introduction of man-made nylon.[14,15] In his 1977 paper, Kinoshita described a particular nylonase[16] (eventually named NylA in a later paper by Okada[17]), as being "exceedingly rare" because other strains of the same bacteria did not have the nylonase activity. Yet by 1981 he suggested that evolution of nylonases "is not a very rare phenomenon in nature."[18]

Some popular literature has claimed that KI72 was extracted from the nylon-rich environment of the waste water of a nylon factory.[19,20,21] However, it should be pointed out the most important original papers that described nylonase in KI72 did not make that claim, but rather claimed KI72 came from soil in an unspecified geographic location.[22] Furthermore, in 2007, Sudhakar *et al.* reported various marine bacteria such as *Bacillus cereus*, *Vibrio furnisii*, and *Brevundimonas vesicularis* in the Indian Ocean, which could digest nylon.[23] In all these cases it would seem difficult to argue that such nylon digesting ability evolved in response to a nylon rich environment. We will show that other environments that are far removed from human activity (i.e., arctic soils), can harbor organisms such as *Cryobacterium arcticum*, which also seem to possess nylonases as predicted by UNIPROT.[24,25]

After Kinoshita's initial discovery of *nylA*, three other nylonase genes were discovered in the very same KI72 strain, namely *nylB*, *nylB'*, and *nylC*.[26] The *nylA*, *nylB*, and *nylC* genes were unrelated and non-homologous. The *nylB'* gene was partially homologous to *nylB*. Given three entirely unrelated nylonase genes in the same organism, the claim that nylonases were necessarily very new and very rare should have been challenged at that time. It was further discovered the four nylonase genes were found on the same plasmid, and that two of them were part of the same operon – sharing the same promoter.[27] In 2007, three of these four nylonase genes (*nylB*, *nylB'*, and *nylC)* were also reported to be present on the chromosome of a different genus, *Agromyces KYR5*.[28]

As we will show, we have identified 162 organisms that contain computationally-predicted *nylA* genes, refuting Kinoshita's early claim that *nylA* is exceedingly rare. We have identified 193 organisms that contain computationally-predicted *nylB* genes and we have identified over 1800 organisms that contain

computationally-predicted 6-aminohexanoate hydrolase/nylonase genes. The predicted *nylB* genes are frequently seen to be non-homologous on the DNA level, even though the predicted proteins they code for are structurally homologous or functionally convergent. The existence of structurally similar and functional convergent NylB enzymes coded from non-homologous *nylB* genes precludes large scale post-1935 horizontal gene transfer as the sole explanation for the widespread distribution of *nylB*.

There is evidence that nylonase function can be acquired or enhanced through adaptive processes taking place in pre-existing nylonase homologs. The process is similar to the process of bacteria evolving antibiotic resistance through point mutations of existing genes. For example, Prijambada reported in 1995 that nylon-digesting ability could be evolved in the laboratory from bacterial strains that initially lacked such ability.[29] It was shown that the human pathogenic bacteria *Pseudomonas aeruginosa* evolved a nylon digesting strain in just 9 days by restricting the kinds of nutrients available to the colony. Even though Prijambada was not able to determine the exact mutations and genes responsible for the acquired ability to degrade nylon, the rapidity of the change suggests nylonase activity arises very easily.

The ease of adapting pre-existing genes via point-mutation for nylonase function was further demonstrated in 1991 when Kato reported that a mere 2-residue change out of the 392-residues of the pre-existing homolog of NylB, known as NylB', enabled it to break down nylon 1000 times more effectively than before the change.[30]

In summary, the literature clearly indicates that nylonase activity is widely observed in nature, and that it is also very easily derived. This strongly suggests that nylonase activity is very low hanging fruit – it does not seem to require a highly constrained or specific active site, and is may be readily derived from other enzymes having similar activities.

METHODS/RESULTS – DISCOVERING NEW NYLONASES AND HOMOLOGS

Enzyme homologies and functional convergences of nylonases can be inferred in a variety of ways. The most direct method is a simple search for DNA sequences or protein sequences that are similar to known nylonases. This is not as comprehensive as structural searches since structurally similar proteins may have only 12% sequence similarity.[31,32] We initially began with the *nylB* gene, doing BLASTN and BLASTP searches, and then contrasted those BLAST results with searches using UNIPROT.

We first did a BLASTN search of the NCBI database, looking for homologs of the *nylB* gene in *Arthrobacter* KI72. We found only one substantially different type of bacterium, *Agromyces sp KYR5*, that contained DNA homologous to the *nylB* gene.[33] However, when an amino acid sequence search for the NylB protein was made using BLASTP, many types of bacteria were found to have homologous proteins. For example, 11 different species had E-values of effectively zero, and 25 species had an E-value in the range of e^{-100}. Because such databases contain only a tiny fraction of all protein sequences in the biosphere, there must be a great number of species having proteins homolgous to NylB. Therefore, there is no rational basis to presume a recent *de novo* origin of the *nylB* gene.

When the search for the *nylB* gene was expanded for *nylB* genes that coded for structurally and functionally similar proteins using the computational gene predictions from the UNIPROT database of proteins, around 193 different strains and species emerged that contained a predicted gene called *nylB*

or variations of that theme such as *nylB_1*. Such a search is easily accomplished by going to the UNIPROT website and simply typing in "nylB" in the search box.

One such organism listed from the search for *nylB* genes was the *nylB_1* gene of *Bacillus cereus.*[34] As mentioned earlier, the marine bacteria *Bacillus cereus* from the Indian Ocean has been experimentally confirmed to digest nylon - validating the UNIPROT prediction of a *nylB* gene in this species.

Given the observed nylonase activity in this bacterium, and the fact that it has a NylB homolog, it is reasonable to assume that the predicted *Bacillus cereus nylB* genes are active. It is interesting to point out, that the predicted the *Bacillus cereus nylB_1* gene has 94.4% sequence identity with the predicted *nylB* gene of a strain of the human pathogenic bacteria *Streptococcous pneumoniae.*[35]

Remarkably, *Streptococcous pneumoniae* also has UNIPROT-predicted homologs of the two other well studied nylonase genes, *nylB'* and *nylA*. These results raise a profound question: "Why are bacteria from diverse environments around the globe coding for diverse families of nylonase proteins that are structurally and functionally homologous, but are encoded by non-homologs DNA sequences?" This seems to be happening in soil bacteria, marine bacteria, and human pathogenic bacteria.

Further search for organisms with predicted or confirmed *nylB'* homologs yielded approximately 125 organisms. A search for organisms with a predicted or confirmed *nylA* homolog yielded approximately 162 organisms. A search for organisms with a predicted or confirmed *nylC* homolog yielded approximately 9 organisms. These numbers are not exact, because some UNIPROT entries are for the same organism under different names and the UNIPROT database is ever-expanding.

The names used for the predicted proteins of *nylB* and *nylB'* were mostly "6-aminohexanoate dimer hydrolases", but a few were named simply "hydrolases" or "beta-lactamases." NylA entries for predicted proteins were mostly named "amidases" and the rest "6-aminohexanoate cyclic dimer hydrolases." NylC predicted proteins were named "6-aminohexanoate dimer hydrolases", "endotype6-aminohexanoate-oligomer hydoralses" and "NylC-like proteins."

When the search for nylonases was expanded to "6-aminohexanoate hydrolase", 1827 organisms were found with predicted proteins for such nylonases, but the majority were not listed with genes names such as *nylA, nylB, nylB'* or *nylC*. This search is very easy to carry out by simply going to the UNIPROT website and entering "6-aminohexanoate hydrolase" in the search box. It is worth noting "nylonase" is a colloquial term, so UNIPROT will not return any hits if "nylonase" is used as a search term.

The UNIPROT list of predicted 6-aminohexanoate hydrolases is at least tentatively supported by exact or related organisms confirmed to digest nylon such as those mentioned earlier such as *Bacillus cereus, Vibrio furnisii, Corynebacterium aurantiacum B-2, Pseudomonas aeruginosa*. It seems conservation of the protein motifs of confirmed nylonases is beyond dispute even without complete experimental demonstration of nylonase activity for every nylonase gene reported by UNIPROT.

Detailed output from the UNIPROT analyses, as well as details about the BLASTP searches, are available under "Supplemental materials."

DISCUSSION

We began this work hoping to better understanding the various claims regarding the *de novo* origin of certain nylonase genes. The idea that nylonases would have arisen very recently, *de novo*, was based upon the widely-held assumption that nylonases would have been essentially non-existent prior to the artificial manufacture of nylon. This basic assumption would not be justified if there were any nylon-like polymers in nature, or if nylonase activity required very low specificity, such that enzymes with other functions might also possess or acquire nylonase activity.

Our analyses indicate that nylonase genes are abundant, come in many diverse forms, are found in a great number of organisms, and these organisms are found within a great number of natural environments. We also show that nylonase activity is easily acquired through mutation of other enzymes, which strongly suggests that nylonase activity has very low specificity of the active site.

These findings refute the widely held assumption that nylonases were essentially non-existent before 1935. In this light, there is no reason to believe that any nylonase emerged since 1935, and so there is no solid basis for invoking any *de novo* nylonase genes. Therefore, it seems only reasonable to re-examine the earlier claims of *de novo* genes.

Various Models Claiming Newly Evolved Enzymes

In 1977 Kinoshita reported on a nylonase enzyme that would eventually be named NylA. He hypothesized that "the enzyme has evolved by adaptation to a new synthetic substance which is a waste product of nylon-6 production."[36] Kinoshita did limited testing of the enzyme, and as best as he could determine, it appeared the enzyme had lost all ability to catalyze reactions involving naturally occurring biological substances. He suggested that the enzyme was "an evolved enzyme which originally had an activity on a physiological substrate but lost it by the evolutionary mutation." By 1981, Kinoshita was referring to the NylA and NylB enzymes as "the two newly evolved enzymes" without specifying how the enzymes evolved.

The phrase "newly evolved enzymes" came to mean different things to different theorists. Okada in 1983 indicated that as little as 1 residue change could create a nylonase from a pre-existing gene, but then in the same paper he argued for a gene duplication of a pre-existing gene (*nylB'*) followed by enough point mutations to create a 47 residue changes resulting in the NylB protein consisting of 392 residues.[37]

In 1984, Ohno was critical of Okada's gene duplication hypothesis, and instead suggested NylB was the result of a single frame shift mutation resulting in over 400 simultaneous residue changes in an open reading frame that would then have coded for 427 residues. He declared that this hypothetical transformation was the "Birth of a Unique Enzyme", and he specifically used the phrase "*de novo*" to describe the transformation[38] (as best as can be determined, Kinoshita and Okada do not actually use the phrase "*de novo*"). In fact, Ohno hinted that the identical hypothetical frame shift event as happened in *nylB* also happened independently in *nylB'* (even though *nylB'* codes for a different protein with 47 divergent amino acids).

The Kinoshita narrative about nylonase evolution led to the Okada narrative, which led to the Ohno narrative. The evolution of the narrative quickly progressed to the point that popular literature characterized Ohno's frame shift model as a direct observation and an established fact. By 2016 the popular narrative included claims that x-ray crystallography proved that the NylB protein folds were all

de novo, and could be directly attributed to the hypothetical frame shift event.[39] The popular narrative also implied that the hypothetical frame shift mutation was directly observed under reproducible laboratory conditions.[40]

The independent hypotheses of Ohno and Okada about the origin of the *nylB* gene appear to have been accepted uncritically, despite their speculative nature and the absence of any direct support. In retrospect, Okada and Ohno's speculations seem especially *ad hoc*, given that *nylB* was found physically linked to two other non-homologous nylonase genes on the very same plasmid (i.e., *nylA* and *nylC*). No explanation for the evolution of these other nylonases was ever attempted, nor was there any attempt to explain how all three nylonase genes could have become linked.

Strangely, no one seems to have asked the obvious question: "if the *nylB* gene evolved *de novo* very recently, then where did the linked *nylA* and *nylC* nylonase genes come from?" Nor has anyone commented on the fact that *nylB* and *nylC* were found tightly coupled, being found in the same polycistron and sharing the very same promoter. Logically, if *nylB* required a *de novo* explanation, certainly so did *nylA* and *nylC*. Furthermore, if these three *de novo* nylonase genes arose independently, how could we explain how they fortuitously landed on the same plasmid, where they were found to be co-regulated and were working in concert as part of a catabolic cascade? Indeed, during the last four decades why has no one commented on the relevance of the many newly emerging nylonase genes and related proteins listed in the literature review above? In light of all these issues, it seems remarkable that for decades no one challenged the widely accepted premise that *nylB* arose *de novo* in the very recent past.

Problems with Ohno's Frame-shift Hypothesis

In the introduction of Ohno's paper he describes his real purpose – he wished to support his broader thesis that frame-shift mutations play a major role in the evolution of new proteins. Ohno was not primarily interested in understanding the origin of the nylonase gene – his primary interest was in supporting this broader thesis. He designed his model for the origin of NylB nylonase as a hypothetical example that might illustrate how frame-shifts could routinely create new proteins instantaneously.

<u>The speculative nature of Ohno's paper</u> – Ohno speculated that an unknown precursor gene gave rise to NylB, via a frame shift mutation. He speculated that a specific hypothetical frameshift mutation in the hypothetical precursor gene gave rise to what is now the start codon of *nylB*. Ohno's hypothesis required that *nylB* arose from a DNA sequence encoding an unknown precursor protein consisting of a specific string of 427 amino acids. He speculated that this DNA sequence underwent a frame-shift mutation which fortuitously and instantly established an entirely new gene, encoding a totally new sequence of 392 amino acids. He hypothesized that by chance that new protein had *de novo* nylonase activity that enabled a nylonase-based metabolism.

Ohno asserted his hypothesis so forcefully that it appears that many readers thought he actually had evidence for a hypothetical frame shift event. Remarkably, Ohno's hypothesis was not critically examined, and was never rigorously tested. His hypothesis involved a series of assumptions: 1) he presumed a precursor gene/protein existed having a function totally unrelated to nylonase activity; 2) he presumed that a very specific frame-shift mutation turned a serendipitous long open reading frame into a random string of amino acids; 3) he presumed that this amino acid string serendipitously and instantly encoded a stable and fully functional nylonase enzyme.

In his paper Ohno showed the sequence that he imagined was the precursor gene. Ohno hypothesized the single nucleotide frame shift insertion shown below. The top line represents the relevant section of his imagined ancestor to the *nylB* gene, which he calls PR.C, while the bottom line is the relevant section of the actual *nylB* gene, showing the presumed insertion of a thymine residue, which would have created the start codon of the *nylB* gene:

….TCGGAGACACTCGA-GAACGCACGTTCCACC…. (Ohno hypothetical PR.C sequence)
….TCGGAGACACTCGATGAACGCACGTTCCACC…. (actual *nylB* gene seen in KI72)

By Ohno's own admission, this was speculation, but somehow his hypothesis came to be generally accepted in popular literature as if it were a documented historical event.[41] Ohno's simple thought experiment became widely accepted as an empirical proof that functional proteins can routinely and instantly emerge from what are essentially random DNA sequences.[42]

The problem of stop codons – Ohno's broader theory is now largely forgotten, but unfortunately his hypothetical illustration (the *nylB* frame-shift hypothesis) has come to be mistaken as an historical event. The reason that Ohno's broader thesis is now forgotten is because geneticists understand that frame-shift mutations consistently result in multiple stop codons, which result in the premature truncation of any hoped-for new enzymes.

In the particular example of the *nylB* gene, Ohno had speculated that there just happened to be an alternative long open reading frame in the region of the presumed precursor gene. Statistically, this would have been very improbable. Without the fortuitous pre-existence of the long alternative open reading frame, Ohno would have had no rational basis for his frame-shift model. For this reason, Ohno needed to explain the fortuitous alternative ORF. He did this by claiming that the NylB protein sequence just happens to have a great deal of internal repeats, which perhaps tended to exclude stop codons. He tried to support this idea in Figure 2 of his paper.[43] He shows a handful of very short amino acid strings that occur more than once within the NylB protein sequence. Ohno's claim of extreme internal redundancy is not persuasive - the actual amount of internal homology does not appear to be greater than expected, and Ohno did no statistical analysis that would support the idea that internal similarities were more abundant than expected by chance. In this light, Ohno's hypothesis requires that the long alternative reading frame (that would fortuitously and instantly give rise to an active nylonase enzyme), arose by extremely good luck - against long odds. Given this limitation, even if his NylB model were correct, it would not help support his broader thesis that new genes typically arise by frame shift mutations.

The apparent non-existence of the reputed precursor gene - If the *nylB* gene arose just a few decades ago, due to a simple frame shift mutation in a previously existing gene, it would be reasonable to expect evidence for the original precursor gene/protein (which should still be found in many places in the biosphere). Ohno predicted the exact precursor DNA sequence, and the exact protein sequence of the hypothetical precursor gene. We have looked for Ohno's hypothetical precursor gene using BLASTN and BLASTP searches and have failed to detect clear evidence for Ohno's proposed precursor DNA or protein sequences. This is direct evidence against Ohno's hypothesis.

It is important to note that alternate reading frames do exist in biology, whereby different proteins can be alternately read within the same genetic locus via frame-shift *reading* (not *mutation*). In systems where such frame-shifted *reads* are needed by the organism, such overlapping genetic elements would

be even less tolerant to frame-shift *mutations* than normal. So the existence of functional alternate reading frames does not help Ohno's frame-shift mutation hypothesis.

Given that BLASTP returned 11 organisms with predicted proteins similar to NylB and UNIPROT found 193 organisms with provisional proteins similar to NylB, it is striking that the same databases gave no proteins similar to Ohno's hypothetical PR.C sequence. The most reasonable explanation for the lack of evidence for Ohno's hypothetical ancestral PR.C sequence is that it never existed, and the post-1935 frame-shift mutation never happened.

The problem of creating a stable and functional protein from a random sequence - When Ohno published his frame shift hypothesis it appears he was not aware of the extreme improbability of establishing a random *de novo* protein sequence that can fold into a stable form and can effectively perform any specific function. It is widely understood that a random amino acid sequence is not likely to be stable in the cellular environment, and that even if it were to be stable, it would have essentially zero chance of having a specific function directly relevant to the needs of the organism. Since the time of Ohno we have learned a great deal about the nature of biological information and the daunting probabilistic constraints that limit the creation of meaningful biological information from random strings of characters.[44,45]

The difficulty of explaining of all the other *nylB* genes and homologs - If a frame-shift mutation gave rise to the original *nylB* gene, this raises the problem of where the *nylB'* gene, and the many other *nylB* homologs came from, especially the ones having mostly structural homology in the coded proteins rather than sequence homology in the genes. The *nylB'* gene is only 88% identical to the *nylB* gene. It is not reasonable to invoke the identical frame shift mutation in both genes.

In regard to this problem, Ohno's thinking seems ambiguous:

> "It is of interest to note here that this stretch of base sequence is duplicated elsewhere within the pOAD2 genome roughly 90˚ away and that the coding sequence for the second isozymic form of 6-AHA LOH is found in this duplicated stretch. Thus, a pair of isozymic preexisted coding sequences might have given rise independently to the coding sequences for two isozymic forms of 6-AHA LOH."

In regard to this same problem, Thwaites of the NCSE has interpreted Ohno's comments as indicating that two identical frame shifts happened independently on the same plasmid, creating two different but homolgous nylonases.[46]

A similar problem arises for the *nylB* and *nylB'* genes found in the chromosome of *Agromyces KYR5* which are also each 1179 bases in length. Given this additional problem, one must invoke even more unreasonable assumptions. There has to be a very specific frame shift mutation followed by horizontal gene transfer of both new genes to another type of bacteria. Alternatively, Ohno's frame shift mutation must have happened identically but independently four times in four separate reading frames, in two different genera.

The problem of non-trivial errors - Lastly, it appears Ohno made some non-trivial errors in his PNAS publication. For example, he states in the abstract that the hypothetical precursor (PR.C) was 472 amino acids long, but in his figures caption he says (correctly), it was 427 amino acids long. A more serious error arises in the final part of his PR.C sequence, which inadvertently results in a pre-mature stop codon. He makes no mention of why he deleted a guanine from the tail end of his PR.C sequence,

thus creating a pre-mature stop codon for his hypothetical protein. The tail end of his hypothetical PR.C sequence reads "GCGGCTGA" but the Genbank entries of the RS-II_A sequence of KI72 and Okada's paper records the proper sequence as "GCGGCGTGA". (See "Supplement G" that describes in detail Ohno's errors.) On many levels, the paper by Ohno seems poorly done, which seems surprising for a publication of the National Academy of Science.

Problems with Okada's 1983 Gene Duplication Hypothesis

Ironically, Ohno's hypothesis emerged in part because of problems he perceived in Okada's hypothesis. The gene duplication hypothesis of Okada *et al.* was proposed a year before Ohno's paper in 1983.[47] Okada *et al.* assumed, as did Ohno, that the *nylB* evolved sometime after 1935. Okada et al. proposed that *nylB* arose from the duplication of the linked *nylB'*. But Okada's model now seems very unlikely for several reasons.

Firstly, Negoro reported in 1992 that *nylB* shares a promoter with *nylC*,[48] and this strongly suggests a pre-1935 coordinated function involving both *nylB* and *nylC*. This leaves *nylC* entirely unexplained, and strongly argues that *nylB* and *nylC* emerged as a functional unit (a polycistron), much earlier than 1935.

If there had been a gene duplication, it would be much more reasonable to invoke the hypothesis that *nylB* gave rise to *nylB'*, rather than vice versa. Otherwise a random duplicate copy of *nylB* would have to arise and fortuitously land in the *nylC* cistron at precisely the right position for the co-regulation of the active nylonase NylC and the not-yet-nylonase NylB. Then the not-yet-nylonase NylB would have to accumulate a very large number of mutations on its way to becoming the active NylB that is co-regulated with NylC. The genes *nylB* and *nylB'* differ by 140 point mutations (out of 1179 bases).[49] This very precise insertion point for the duplication event, followed by so many mutations, would have to happen in just a few decades, which seems extremely unlikely - as pointed out by Ohno.[50]

A diagram in S. Negoro's 2000 paper[51] suggests that the NylA and NylC nylonases work together with the NylB nylonase. This apparently involves a catabolic pathway wherein NylA and NylC degrade complex cyclic nylon oligomers to nylon linear dimers, which are then degraded by the NylB nylonase. This sort of coordinated specialization between three non-homologous nylonase genes, two of which share the same promoter, is remarkable. If this nylon degrading cascade composed of NylA, NylB and NylC evolved post-1935, it would require simultaneous evolutionary convergence and coordination from three non-homologous genes. How could all this come together in a few decades? Given the shared promoter and the 3-way cooperation of three unrelated nylonases, it is not credible that NylB arose very recently via random genetic duplication. This sophisticated genetic network must have existed long before 1935, for the purpose of catabolizing an unknown natural substance.

In addition to the above problems, Ohno's frame shift hypothesis and Okada's recent duplication scenario are now falsified because of the many organisms that we have now shown have either *nylB* genes or homologs (see literature review above). Such genes are not found in nylon-rich environments but are found in pathogens, soils and oceans. UNIPROT listed 193 organisms with predicted *nylB-like* genes and 125 organisms with predicted *nylB'-like* genes. It seems clear that we do not yet know the origin of either *nylB* or *nylB'* - just as we do not know the origin of *nylA* or *nylC*. There is no reason to think any of these genes arose *de novo* in the recent past.

CONCLUSION

We began this work hoping to better understanding the various claims of the *de novo* origin of certain nylonase genes. Kinoshita claimed the nylonases he studied were newly evolved enzymes that did not exist prior to 1935. Okada claimed the "newly evolved" NylB enzyme arose via gene duplication of NylB′ followed by 47 residue substitutions. Ohno claimed the "*de novo*" NylB and NylB′ enzymes arose via two independent single frame shift mutations in each of the corresponding genes which instantly created two functional *de novo* proteins.

Our analyses indicate that nylonase genes are actually abundant, come in many diverse forms, are found is a great number of organisms, and such organisms are found in a great number of natural environments. We show there is no reason to think that any of these nylonases emerged since 1935, and so there is no basis for invoking any *de novo* genes arising since 1935. Furthermore, there are numerous glaring problems with the specific *de novo* speculations of Okada and Ohno. The early claims of *de novo* nylonase genes were unsupported and speculative, and in light of new data these hypotheses now appear to be unwarranted and essentially falsified.

The discovery of numerous naturally-occurring genes having nylonase activity, along with a multitude of homologous genes and proteins that provisionally have similar activities, opens the door to further exploration of nylonases and their functions.

ENDNOTES

[1]Negoro, S. Biodegradation of Nylon Oligomers. *Appl Microbiol Biotechnol* (2000) 54:461-466. Springer-Verlag 2000.

[2] Negoro, S. Biodegradation of Nylon Oligomers. *Appl Microbiol Biotechnol* (2000) 54:461-466. Springer-Verlag 2000.

[3] Kinoshita, S. Purification and Characterization of 6-Aminohexanoic-Acid-Oiligomer Hydrolase of *Flavobacterium* sp. KI72. *Eur. J. Biochem*. 116, 547-551 (1981).

[4] Kinoshita, S. Purification and Characterization of 6-Aminohexanoic-Acid-Oiligomer Hydrolase of *Flavobacterium* sp. KI72. *Eur. J. Biochem*. 116, 547-551 (1981). The paper says: "As regular nylon fiber is higher than dictamer, this enzyme can not hydrolyze ordinary nylon."

[5] Ebata, M. Morita, K. Hydrolysis of ε-Aminocaproyl Compounds by Trypsin. *The Journal of Biochemsitry*, Vol 46, No. 4, 1959.

[6] Fukumura, T. Hydrolysis of Cyclic and Linear Oligomers of 6-Aminocaproic Acid by a Bacterial Cell Extract. *The Journal of Biochemistry*, Vol. 59, No. 6, 1966.

[7] Kiumarsi, A. Parvinzadeh, M. "Enzymatic Hydrolysis of Nylon 6 Fiber Using Lipolytic Enzyme." *Journal of Applied Polymer Science* DOI 10.1002/app. 4 June 2009.

[8] Fukumura, T. "Hydrolysis of Cyclic and Linear Oligomers of 6-Aminocaproic Acid by a Bacterial Cell Extract". *The Journal of Biochemistry*, Vol. 59, No. 6, 1966.

[9] Mahdi, MS. Ameen, R. Ibrahim, H. "Study on Degradation of Nylon 6 by thermophyilic bacteria Anoxybacillus rupiensis Ir3 (JQ912241)." *International Journal of Advanced Research in Biological Sciences*. ISSN:2348-8069. DOI: 10.22192/ijarbs. Coden:IJARQG(USA) Volume 3, Issue 9- 2016. www.ijarbs.com

[10] The re-naming history is complex. The first name change was noted by Ohno in 1984 in "Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious codng sequence. Proc Natl. Acad. Sci. USA Vol 81, pp. 2421-2425, April 1984. Evolution.". The second name change was noted in "Genetic Organization of Nylon-Oligomer-Degrading Enzymes from Alkalophilic Bacteria, *Agromyces* sp. KYR5. Journal of Bioscience and Bioengineering. Vol. 104, No. 6, 521-524, 2007. DOI: 10.1263/jbb.104.521."

[11] Kinoshita, S. Kageyama, S. Iba, K. Yamada, Y. Okada, H. "Utilization of a Cyclic Dimer and Linear Oligomers of ε-Aminocaproic Acid by Achromobacer guttatus KI 72." Agr. Biol. Chem., 39 (6), 1219-1223, 1975.

[12] The reference for 1973 discovery was in the end note of Kinoshita's 1975 "Utilization of Cyclic Dimer and Linear Oligomers of ε-Aminocaproic Acid by Achromobacer guttatus KI 72 paper. The end note said "Kinoshita, S. Kobayashi, E. Okada, H. *J Ferment. Technol.*, 51, 719(1973)." That Journal appears to be out of print and is not readily available.

[13]Kinoshita, S. Kageyama, S. Iba, K. Yamada, Y. Okada, H. "Utilization of a Cyclic Dimer and Linear Oligomers of ε-Aminocaproic Acide by Achromobacer guttatus KI 72." Agr. Biol. Chem., 39 (6), 1219-1223, 1975.

[14] Kinoshita, S. Negoro, S. Muramatsu, M. Bisaria, V. Sawada, S. Okada, H. "6-Aminohexanoic Acid Cyclic Dimer Hydrolase. A New Cyclic Amide Hydrolased Produced by Acromobacter guttatus KI72." *Eur. J. Biochem*. 80,489-495 (1977).

[15] Kinoshita, S. Terada, T. Taniguchi, T. Takene, Y. Masuda, S. "Purification and Characterization of 6-Aminohexanoic-Acid-Oligomer Hydrolase of *Flavobacterium* sp. KI72.

[16] Kinoshita, S. Negoro, S. Muramatsu, M. Bisaria, V. Sawada, S. Okada, H. "6-Aminohexanoic Acid Cyclic Dimer Hydrolase. A New Cyclic Amide Hydrolased Produced by Acromobacter guttatus KI72." *Eur. J. Biochem*. 80,489-495 (1977).

[17] Okada, H. Negoro, S. Kimura, H. Nakamura, S. "Evolutionary adaptation of plasmid-encoded enzymes for degrading nylon oligomers." *Nature* Vol. 306 November 1983.

[18] Kinoshita, S. Purification and Characterization of 6-Aminohexanoic-Acid-Oiligomer Hydrolase of *Flavobacterium* sp. KI72. *Eur. J. Biochem*. 116, 547-551 (1981).

[19] http://biologos.org/blogs/dennis-venema-letters-to-the-duchess/intelligent-design-and-nylon-eating-bacteria/

[20] https://ncse.com/cej/5/2/new-proteins-without-gods-help

[21] Miller, K. Only a Theory: Evolution and the Battle for America's Soul. Penguin June 12, 2008.

[22] Negoro, S. Shinagawa, H. Nakata, A. Kinoshita, S. Hatozaki, T. Okada H. "Plasmid Control of 6-Aminohexanoic Acid Cyclic Dimer Degradation of Enzymes of Flavobacterium sp KI72." *Journal of Bacteriology*, July 1980, p. 238-245. 0021-9193/80/00-238/08$02.00/0.

[23] Sudhakar, M. Priyadarshini, C. Doble, M. Murthy, P. Venkatesan, R. "Marine bacteria mediated degradation of nylon 66 and 6." *International Biodeterioration & Biodegradation*. 60 (2007) 144-151.

[24] The UNIPROT number for the 6-aminohexanoate hydrolase in *Cryobaceria arcticum*, an organism found in Arctic is A0A1B1BHD0.

[25] Bajerski, F. Ganzer, L. Mangelsdorf, K. Lipski, A. Wagner, D. "Cryobacterium arcticum sp. nov., a psychrotolerant bacterium from an Arctic soil". 01 August 2011, *International Journal of Systematic and Evolutionary Microbiology*. Microbiology 61:1849-1853, doi:10.1099/ijs.0.027128-0.

[26]Negoro, S. Kakudo, S. Urabe, I. Okada, H. "A New Nylon Oligomer Degradation Gene (nylC) on Plasmid pOAD2 from *Flavobacterium* sp." *Journal of Bacteriology*, Dec. 1992, p 7948-7953 0021-9193/92/247948-06$02.00/0.

[27] Negoro, S. Kakudo, S. Urabe, I. Okada, H. "A New Nylon Oligomer Degradation Gene (nylC) on Plasmid pOAD2 from *Flavobacterium* sp." *Journal of Bacteriology*, Dec. 1992, p 7948-7953 0021-9193/92/247948-06$02.00/0.

[28]Negoro, S. Kakudo,S. Urabe,I. Okada, H. "Genetic Organization of Nylon-Oligomer-Degrading Enzymes from Alkalophilic Bacteria, *Agromyces* sp. KYR5." *Journal of Bioscience and Bioengineering*. Vol. 104, No. 6, 521-524, 2007. DOI: 10.1263/jbb.104.521.

[29]Prijambada, I. Negoro, S. Yomo, T. Urabae, I. "Emergence of Nylon Oligomer Degradation Enzymes in *Pseudomonas aeruginosa* PAO through Experimental Evolution." *Applied and Environmental Microbiology*, May 1995, p 2020-2022. 0099-2240/95/$04.00+0.

[30]Kato, K. Fujiyama, K. Hatanaka, H. Priyambada, I. Negoro, S. Urabe, I. "Amino acid alterations essential for increasing the catalytic activity of the nylon-oligomer-degradation enzyme of *Flavobacterium* sp." Eur. J. Biochem. 200, 165-169(1991). 0014295691000522N.

[31]Rost, B. "Twilight zone of protein sequence alignments." *Protein Engineering* vol. 12 no.2 pp85-94, 1999.

[32]Rost, B. "Protein structures sustain evolutionary drift." *Fold Des.* 1997;2(3):219-24.

[33]The GenBank accession number is D10686 for *nylB Flavobacteria* sp. Plasmid pOAD2. The *Flavobaceria* strain was renamed to *Arthrobacteria*, and was formerly named *Acromobacter guttatus*

[34]The UNIPROT entry for nylB_1 for *Bacillus cereus* is A0A0K6IYF0

[35]The UNIPROT entry for nylB for *Streptococcus pneumoniae* is A0A0u0EFW2.

[36] Kinoshita, S. Negoro, S. Muramatsu, M. Bisaria, V. Sawada, S. Okada, H. "6-Aminohexanoic Acid Cyclic Dimer Hydrolase. A New Cyclic Amide Hydrolased Produced by Acromobacter guttatus KI72." *Eur. J. Biochem*. 80,489-495 (1977).

[37] Okada, H. Negoro, S. Kimura, H. Nakamura, S. "Evolutionary adaptation of plasmid-encoded enzymes for degrading nylon oligomers." *Nature* Vol. 306 November 1983.

[38] Ohno, S. "Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence." *Proc Natl. Acad. Sci. USA*. Vol. 81, pp. 2421-2425, April 1984. Evolution.

[39] http://biologos.org/blogs/dennis-venema-letters-to-the-duchess/intelligent-design-and-nylon-eating-bacteria/

[40] Miller, K. Only a Theory: Evolution and the Battle for America's Soul. Penguin June 12, 2008.

[41]http://biologos.org/blogs/dennis-venema-letters-to-the-duchess/intelligent-design-and-nylon-eating-bacteria/

[42]https://ncse.com/cej/5/2/new-proteins-without-gods-help

[43]Ohno, S. "Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence." *Proc Natl. Acad. Sci. USA*. Vol. 81, pp. 2421-2425, April 1984. Evolution.

[44]Marks, R. Dembski, W. Sanford, J. Behe, M. Gordon, M. *Biological Information: New Perspectives*. http://www.biologicalinformationnewperspectives.org/

[45]Axe, D. "Estimating the Prevalence of Protein Sequences Adopting Functional Enzyme Folds" (Axe, J Mol Biol 341, 1295-1315, 2004).

[46] https://ncse.com/cej/5/2/new-proteins-without-gods-help

[47]Okada, H. Negoro, S. Kimura, H. Nakamura, S. "Evolutionary adaptation of plasmid-encoded enzymes for degrading nylon oligomers." *Nature* Vol. 306 November 1983.

[48]Negoro, S. Kakudo, S. Urabe, I. Okada, H. "A New Nylon Oligomer Degradation Gene (nylC) on Plasmid pOAD2 from *Flavobacterium* sp." *Journal of Bacteriology*, Dec. 1992, p 7948-7953 0021-9193/92/247948-06$02.00/0.

[49] Okada, H. Negoro, S. Kimura, H. Nakamura, S. "Evolutionary adaptation of plasmid-encoded enzymes for degrading nylon oligomers." *Nature* Vol. 306 November 1983.

[50] Ohno, S. "Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence." *Proc Natl. Acad. Sci. USA*. Vol. 81, pp. 2421-2425, April 1984. Evolution.

[51]Negoro, S. Biodegradation of Nylon Oligomers. *Appl Microbiol Biotechnol* (2000) 54:461-466. Springer-Verlag 2000.