Testing the Hypothesis that the Nylonase NylB Protein Arose *de novo* via a Frameshift Mutation

Salvador Cordova and John Sanford*

FMS Foundation, 7160 Stone Hill Rd., Livonia, NY 14487 USA.
John Sanford, President. Salvador Cordova, Research Associate.
* Correspondence: John Sanford

ABSTRACT

In 1984, Susumu Ohno hypothesized that the nylon-degrading enzyme NylB arose *de novo* via a frameshift mutation within a hypothetical precursor protein (PR.C). However, Ohno never tested his hypothesis or provided supporting biological evidence. For decades, Ohno's famous frame-shift hypothesis has been uncritically accepted as the correct explanation for the origin of NylB and has been used to illustrate how simple it is for a totally new enzyme to arise spontaneously. In this paper we test Ohno's hypothesis in light of data not available in 1984.

We searched multiple protein databases and found that the NylB protein is widely occurring, has thousands of homologs, and is found in diverse organisms and diverse habitats. Conserved domain searches showed that the NylB sequence is homologous to beta lactamases - a family of highly conserved enzymes. However, our searches showed that there is no evidence for the existence of Ohno's hypothetical PR.C protein, nor any credible homolog. Our results effectively falsify Ohno's frameshift hypothesis.

We extended this analysis to other nylonases and found all the nylonases we examined had large numbers of homologs throughout the biosphere. This falsifies the long-held assumption that all nylonases evolved after the invention of nylon in 1935.

Keywords: Bioinformatics, Proteomics, Protein Evolution

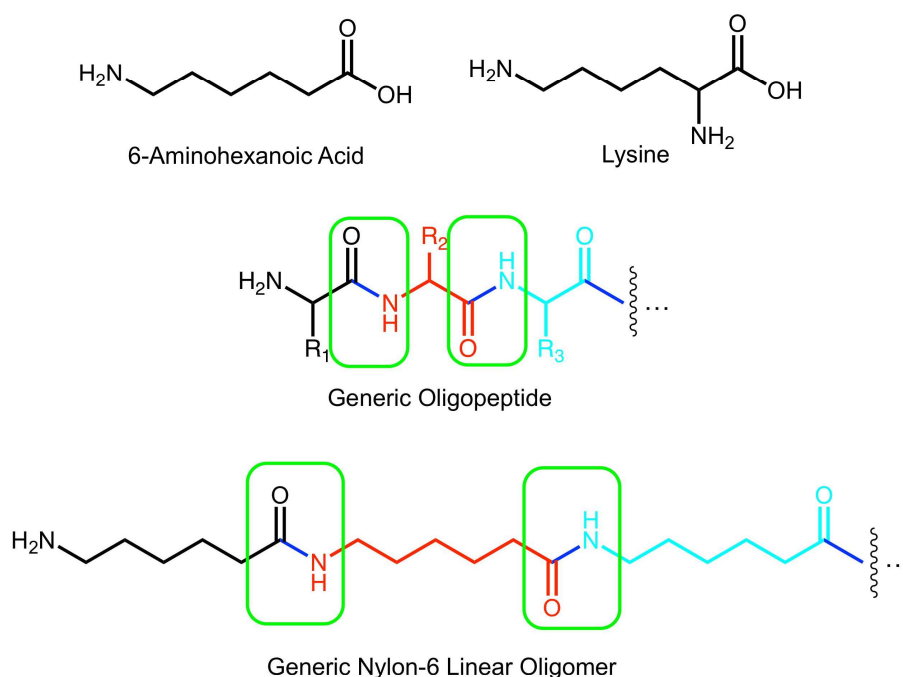
INTRODUCTION

Nylon is a synthetic polymer that was invented in 1935. From the mid-1950's onward, a variety of enzymes have been discovered that can degrade nylon-6 oligomers into their monomer components.[1,2] Such enzymes have been colloquially referred to as "nylonases". It has been widely assumed that all nylonases have evolved since 1935. However, there are now many reasons to doubt this assumption.

Below we review various hypotheses that have been developed to understand the origin of nylonase enzymes – in particular the origin of NylB. The three major hypotheses for the origin of NylB are: 1) Okada *et al.*'s post-1935 gene duplication and mutation hypothesis;[3] 2) Ohno's post-1935 frame shift hypothesis;[4] 3) Yomo *et al.*'s pre-1935 NylB homologs hypothesis.[5] It should be noted Okada was co-author of Yomo *et al.*

In 1957, Ebata and Morita discovered the first enzyme that could breakdown nylon.[6] They found that Trypsin, a widely conserved enzyme in mammals, was shown to

degrade nylon-6 oligomers. This capability obviously existed in Trypsin prior to the invention of nylon in 1935. Trypsin is a protease, and nylon has some protein-like molecular features (Figure 1). Therefore, it should not be surprising that Trypsin might degrade nylon. It is important to note that many enzymes that existed long before the invention of nylon might still manifest "nylonase" activity. This does not necessarily imply a newly evolved enzyme function. Hence application of the term "nylonase" can be ambiguous. We will use the term *nylonase* to refer to all enzymes with measureable or predicted nylonase activity.

Figure 1

Structures of 6-aminohexanoic acid, lysine, and amide bonds in peptides and nylon-6 linear oligomers. Top, structures for 6-aminohexanoic acid and lysine are shown. Middle, a generic oligopeptide is shown with amide bonds highlighted in blue with green boxes. Bottom, a generic, linear nylon-6 oligomer is shown with amide bonds highlighted in blue with green boxes. For both oligomer structures, the individual monomers are shown in black (first), red (second), and light blue (third). Line and dots at right indicate additional monomers.

In 1966, Fukumura first discovered that a bacterium (*Corynebacterium aurantiacum B-2)* could metabolize nylon[7] and he isolated two of the enzymes involved.[8]

From the mid 1970's to early 1980's Kinoshita, Okada and others published a series of papers on isolation of two nylonase enzymes (eventually named NylA and NylB) from the *Achromobacter gutatus KI72* (renamed *Flavobacteria KI72*).[1,3] The corresponding genes were on the plasmid pOAD in KI72.[9] A paralog of NylB named NylB′ was also discovered which had substantially lower nylonase capability than NylB.[2] In 1993 one more nylonase called NylC in the same bacterium was discovered on the same plasmid.[10] The natural ability of the KI72 strain to metabolize nylon is apparently due to the coordinated action of this set of four linked complementary nylonase genes.[2] Yet Kinoshita claimed that all these genes were "newly evolved" since 1935.[11]

Okada *et al.* was the first to present a hypothesis regarding the origin of NylB. In 1983, Okada argued NylB was a paralog that arose via a gene duplication event from a linked gene coding NylB′. He assumed this must have occurred sometime after 1935.[3] His model requires that the duplicate gene, NylB, would need to acquire 47 residue substitutions via point mutations in just a few decades. Although the paralogous nature of NylB and NylB′ suggests a gene duplication event, there was no direct evidence that it happened post 1935, and he gave no reason why NylB′ might not have arisen from NylB instead.

In his 1984 paper, Susumu Ohno offered a second major hypothesis for the origin of NylB. Ohno criticized Okada's 1983 hypothesis because it required too many point mutations to effect so many amino acid substitutions in so little time. Ohno said, "so extensive an amino acid divergence is not expected to occur in so short a time span."

Ohno took Okada's published sequence known as RS-IIA (which encoded NylB) and constructed a hypothetical sequence he called PR.C by simply deleting a single nucleotide from the RS-IIA sequence and relabeling it as PR.C. Ohno claimed PR.C was the ancestral sequence of NylB. He claimed that shortly after 1935, a single nucleotide insertion in the gene encoding his hypothetical PR.C protein yielded the present-day RS-IIA sequence that now encodes NylB. Ohno criticized Okada's hypothesis as being unrealistic because it required so many point mutations, yet Ohno's hypothesis required an essentially random amino acid sequence arising from a frameshift to instantly form a stable, functional, and specific enzyme. Ohno had no direct evidence that the hypothetical PR.C protein even existed, and his frameshift mutation was purely hypothetical. Yet Ohno put forward his hypothesis so forcefully that readers accepted his model as if it were history, and his paper continues to be cited as if the hypothetical frameshift mutation was actually an observed

fact.[12,13,14,15,16,17]

Just 5 years after Ohno published his frameshift hypothesis, Kanagawa *et al.* discovered another NylB enzyme in another bacterium, *Pseudomonas* NK87, which also had the ability to degrade nylon-6.[18] This effectively falsified Ohno's claim that NylB was unique. This new NylB gene sequence was highly divergent, having only 53% DNA similarity,[19] and only 35% protein sequence similarity compared to Kinoshita's NylB in KI72 (the one Ohno claimed was truly unique). Kanagawa designated this newly discovered NylB as p-NylB and re-named the previously discovered NylB and NylB′ proteins as f-NylB and f-NylB′, respectively.

In 1991, Kato *et al.* attempted to explore Okada's hypothesis by experimentally mutating the 47 amino acids in NylB′ that were divergent from NylB.[20] They discovered that only two of the 47 amino acids were required to enhance nylonase activity in NylB′ up to the level of NylB. The two linked genes coding for NylB and NylB′ were substantially divergent (making duplication and divergence in just a few decades very unlikely), yet they were also substantially homologous (ruling out a single frame shift for the origin of two proteins simultaneously).[4]

In 1992, in response to Kanagawa's discovery of p-NylB, Yomo *et al.* co-authored a paper with Urabe and Okada, to put forward a third competing hypothesis regarding the origin of NylB. Yomo *et al.* argued that Kinoshita's f-NylB and Kanagawa's p-NylB homologs descended from a common ancestor that existed about 140 million years ago.[5] Yomo *et al.* wrote: "The distance between P-nylB and F-nylB (or F-NylB′) is much larger than between F-nylB and F-NylB′. The time divergence of F-nylB and P-nylB is estimated to be at least 1.4×10^8 years… Therefore, most of the

amino acid substitutions from the ancestor of the *nylB* gene family to its descendants of today might have occurred before the beginning of nylon manufacture."

In 1995, experiments by Prijambada, Negoro, Yomo, and Urabe, showed that strains of the bacteria *Pseudomonas aeruginosa* PAO1 which initially lacked activity toward nylon-6 linear and cyclic dimers could be selectively evolved into a strain that could digest these dimers.[21] The evolved descendant from the ancestral PAO1 that had nylon digesting capability was designated PAO5502. However, Prijambada *et al.* point out, "a molecular basis for the emergence of nylon oligomer metabolism in PAO5502 is still unknown." We will explore in this paper some of the sequences in PAO1 that may have bearing on Prijambada's experiment, its relation to Kato's experiment, Ohno's hypothesis, and the evolution of NylB nylonases in general.

In 2007, Sudhakar demonstrated that strains of *Bacilus cereus* found in the Indian Ocean could digest nylon-6,[22] and so we also searched for evidence of NylB in *Bacilus cereus* and its possible bearing on the question of Ohno's hypothesis. In this paper we test Ohno's famous frameshift hypothesis. We will show that multiple lines of evidence falsify Ohno's hypothesis, but are consistent with Yomo's model. Ohno had three primary claims: a) he claimed the NylB protein never existed until sometime after 1935; b) he claimed NylB arose as a *de novo* protein as the result of a frameshift mutation in a precursor protein; c) he claimed he knew the exact sequence of his hypothetical precursor protein. Since the sequences of the NylB protein and Ohno's hypothetical protein are both known, Ohno's hypothesis is readily testable using protein databases.

If Ohno's hypothesis were correct, then a protein database search should reveal evidence that Ohno's hypothetical precursor protein actually existed, had a history, and so should have many protein homologs. At the same time there should be clear evidence that the NylB protein really is a unique protein, with no history, and no protein homologs.

Conversely, if Ohno's hypothesis were wrong, then a protein database search should reveal evidence that the hypothetical precursor protein never existed, had no history, and should have few if any homologs. At the same time there should be evidence that the NylB protein is *not* unique, has a history, and has numerous homologs.

RESULTS

A search for the number of organisms with various explicit nylonase names along with Ohno's PR.C sequence was made in UNIPROT. A search for potential remote homologs was made using psi-BLAST and SPARCLE. The SPARCLE numbers were not adjusted for redundancies but were provided to give an idea of the degree of representation of the homologs in the databases which SPARCLE surveys. The results are shown in Table 1.

Table 1
Search results for homologs of NylB, NylB′, NylA, NylC, 6-aminohexanoate hydrolase, and Ohno's hypothetical PR.C.

Protein Name	Organisms with Protein in UNIPROT	psi-BLAST Potential Homologs	NIH SPARCLE Potential Homolog Entries

NylB	201	20,000+	68,000+
NylB′	134	20,000+	68,000+
NylA	193	20,000+	53,000+
NylC	7	6,936	44
6-Aminohexanoate Hydrolase	2115	20,000+	68,000+
Ohno's Hypothetical PR.C Ancestor of NylB	0	0	0

From the lists of predicted nylonases generated through UNIPROT, data was gathered in the Conserved Domain Database (CDD) and then tallied to see the most common domain family for each nylonase (Table 2). The data used to construct Table 1 and 2 can be found in the in Supplementary Tables S1, S2, S3, S4, and S5.

Table 2

Homology of various nylonases to known enzyme families such as beta lactamases, amidases, and peptidases.

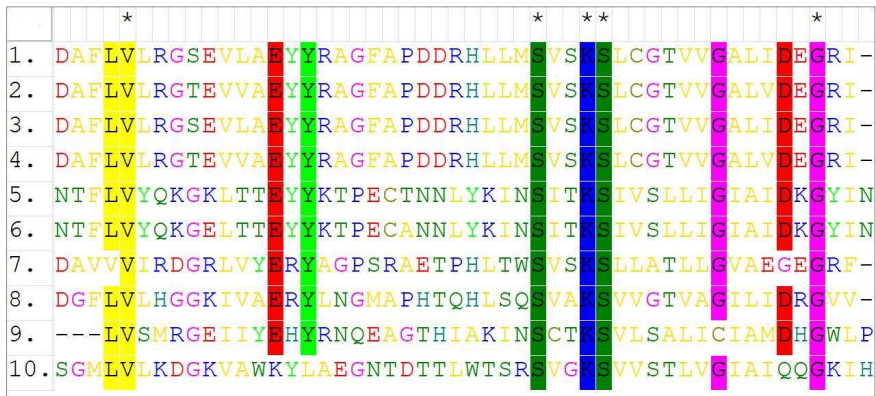
Protein	Most Representative Conserved Domain Family	Conserved Domain Accession Numbers
NylB	beta lactamase	COG1680, pfam00144
NylB′	beta lactamase	COG1680, pfam00144
NylA	amidase	COG0154, cl18951, pfam01425,PRK07869, PRK06061, pfam07501, PRK06529
NylC	peptidase	cd00123, cl00603, cd02252

Several proteins labeled as NylB were also labeled as beta lactamases. Several proteins labeled as NylA were also labeled as amidases. The listing of these proteins with their accession numbers can be found in the Supplementary Tables S6 and S7.

CDD analysis (as provided by GenBank) scored the similarity of the NylB that Ohno studied and COG1680 beta lactamase at 130 bits, which implies the probability that a random amino acid polymer would achieve that level of similarity to the archetypal COG1680 beta lactamase is one chance in 2^{130}.

These proteins were also aligned using the MUSCLE alignment algorithm to show some of the conserved features of the NylB homologs, particularly the Serine-X-X-Lysine motif. This Serine-X-X-Lysine motif agrees with previous X-ray crystallography studies of NylB in *Arthrobacter* KI72.[23] The first 8 of 10 proteins in the alignment (figure 2) were proteins from organisms that had experimental evidence of nylonase NylB activity, and the last two were provided for comparison as they are remote homologs with only predicted NylBs (as of this writing).

We searched the databases for a NylB homolog in *Pseudomonas aeruginosa* PAO1 and found a sequence (accession AAG07735.1) that had 100% identity (96% coverage) to a predicted NylB in another strain of *Pseudomonas aeruginosa* (accession CKI08964.1). This was used as one of the proteins featured in the MUSCLE alignment.

Figure 2

MEGA 6.0 MUSCLE Alignment of critical segment of selected NylB and NylB' homologs showing conserved motifs, particularly the Serine-X-X-Lysine motif confirmed by X-ray crystallography. Using the GenBank Accession numbers for each row: (1) NylB Flavobacteria KI72 CAA24927.1, (2) NylB' Flavobacteria KI72 CAA26616.1, (3) NylB Agromyces KYR5 BAE97621.1, (4) NylB' Agromyces KYR5 BAE97630.1, (5) NylB Bacillus cereus SME41718.1, (6) NylB' Bacillus cereus SME45458.1, (7) NylB Pseudomonas aeruginosa PAO1 AAG07735.1, (8) NylB Pseudomonas NK87 BAA01524.1, (9) NylB' Paenibacillus sp. AD87 OAX47860.1, (10) Beta-Lactamase/NylB Escherichia coli SCQ13749.1.

The MUSCLE alignment (Figure 2) provided an opportunity to compare the segments which Kato modified in his 1991 experiments[20] with the corresponding segments in PAO1 referenced in Prijambada *et al.*'s 1995 experiment[25] and in NK87 referenced in Kanagawa *et. al's* 1989 experiment.[18]

Kato *et al.* mutated a glycine amino acid to an aspartic amino acid within the sequence segment "QTHGRSA" in NylB′ of KI62. The MUSCLE alignment relates that sequence in KI72 to the sequence "VAMLYTR" in PAO1 and "AQLDVAS" in NK87. Similarly Kato mutated histidine to asparagine within a sequence identified as "GFAHGGV" which corresponds to "FVGSSYV" in PAO1 and "ALGDGGF" in NK87.

For completeness we did a distance computation using MUSCLE aligned sequences and found that the NylB and NylB′ paralogs in it had 11% sequence divergence in KI72 and similarly the NylB and NylB′ paralogs in *Bacilus cereus* had 12% sequence divergence, yet the distance between the NylB in KI72 and NylB in *Bacilus cereus* was around 75%. The distance matrix can be also found in Supplementary Table S8.

MATERIALS AND METHODS

The GenBank accession numbers for the database searches for NylB were established by typing in the sequence from Ohno's original 1984 paper into the NCBI BLAST tool. The sequences in Ohno's paper were derived from Okada's 1983 paper that reported the sequences of several nylonases in Flavobacteria KI72. Subsequent papers discovered more nylonases on the same plasmid, and all of the K172 nylonases were designated NylA, NylB, NylB′, and NylC.[3,10] These sequences were used to define the searches for nylonase homologs under those names, specifically in terms of GenBank accession numbers BAA05090.1 for NylA, CAA24927.1 for NylB, CAA26616.1 for NylB′, and BAA01528.1 for NylC.

The gene coding for the NylB protein was contained in a segment of DNA Okada *et al.* called RS-IIA.[3] It is worth mentioning that it appears Ohno mislabeled Okada's RS-IIA as R-IIA in his paper.[3,4] Also, it appears Ohno made either a typo in transcription or failed to clearly account for the creation of a premature stop codon in construction of his PR.C from the RS-IIA sequence. Okada's paper and GenBank indicate that the end of Ohno's PR.C (derived from RS -IIA) should be "GCGGCGTGA," not "GCGGCTGA" as was the case in Ohno's paper. Given that Okada's paper was the source of the actual sequence data, with Ohno's work deriving from that paper, the error must be Ohno's and not Okada's.

The UNIPROT searches were easily conducted by simply going to the UNIRPOT.org website and typing search terms such as "NylB", "NylB′", "NylA", "NylC", and "6-aminohexanoate hydrolase." Lists of proteins for each of these nylonases were created by using a simple Java program to filter out duplicate experimental entries. Afterward, manual review of the filtered lists was also conducted to remove spurious search results. An example of such a spurious result was the result induced by organisms like Nylandria that happened to have the "Nyla" string in them. These lists were used as the foundation for tallying CDD data found in the Supplementary Tables.

Gathering of the CDD results for each protein can be illustrated by the example of NylB in KI72. The gene accession number for the *nylB* gene in KI72 was determined to be X00046.1, and for the NylB protein CAA24927.1. We then went to Protein Database at the URL https://www.ncbi.nlm.nih.gov/protein and entered CAA24927.1 as the search term. This brought up a page that provided a hyperlink titled "identify conserved domains" which links to a Conserved Domain Database (CDD) page. The CDD page showed similarity of this NylB homolog to a beta lactamase domain with the CDD accession number of COG1680. Also, on this page is the probability bit score of similarity of this NylB homolog to the COG1680 beta lactamase domain.

The CDD page also provided a hyperlink titled "domain architecture ID 10004149", which links to the SPARCLE viewer that gives the statistics for the variety of entries

in the databases that conform to the NylB architecture (ID 10004149) associated with Ohno/Okada's NylB homolog. Because of the dynamic nature of databases constantly being updated, the exact numbers provided by the SPARCLE viewer change from day to day. A snapshot of the statistics reported by SPARCLE were then recorded in the results section, but these numbers should not be expected to be the same over time given the ever-expanding size of the databases which SPARCLE is surveying.

Similar procedures were used to tally the conserved domains and SPARCLE numbers for all the other proteins in the lists generated by UNIPROT for NylB, NylB′, NylA, and NylC. The CDD results for NylB, NylB′, NylA, and NylC were listed in Supplementary Tables S1, S2, S3, S4. No CDD analysis was conducted for the general 6-aminohexanoate hydrolases since they are a collection of several enzymes that include NylB, NylB′, NylA, and NylC plus many others.

The psi-BLAST searches were accomplished through the NCBI protein BLAST website. The psi-BLAST algorithm option was enabled by checking the appropriate box in the program selection section. Under the algorithm parameters section, Max Target sequences was reset from the default of 500 to 20,000. All other algorithm parameters were at the default settings which were: Automatically adjust parameters for short input sequences, Expect threshold = 10, word size = 3, Max matches in a query range = 0, Matrix = BLOSUM62, Gap Cost = Existence 11 Extension 1, Compositional adjustments = Conditional compositional score matrix adjustment, no filters or masks, psi-blast threshold = .005, pseudocount =0. Only 1 psi-BLAST iteration was run for each sequence.

The homologs chosen to search for NylB, NylB′, NylA, NylC were from *Flavobaceria* KI72 since these were the first named homologs in Ohno's paper and subsequent papers studying nylonases in *Flavobaceria* KI72. Since NylA, NylB, NylC were all 6-aminohexanoate hydrolases, the union of all the entries with homologs to these three nylonases would be at least as large as the largest set, namely that for NylB, thus the figure of 20,000+ in the psi-BLAST column and 68,000+ in the SPARCLE column for 6-aminohexanoate hydrolases was entered in Table 1.

To verify the possible conflict in assigning functions to protein predictions, a manual search through the lists of several thousand potential NylB and NylA homologs created by psi-BLAST was painstakingly reviewed and a listing of some of the notable conflicts were listed in Supplementary Tables S6 and S7.

As shown in Supplementary Table S9, we constructed a set of 10 protein sequences in FASTA format that was composed of a mix of NylB homologs from organisms that had been mentioned in literature as having credible evidence of nylonase activity and for comparison included in the mix one NylB having a weak e-value compared to NylB from KI72 as well as a NylB homolog from a well-known organism, *E. coli*. In the case of *Pseudomonas aeruginosa* PAO1, although the sequence (accession AAG07735.1) is not listed as NylB, it is 100% identical (96% coverage) to a sequence in another strain of *Pseudomonas aeruginosa* in GenBank that goes by the name of NylB (accession CKI08964.1) . For *E. Coli*, the gene was called *nylB*, even though the predicted expressed protein is called a beta lactamase (accession SCQ13749.1). Data in Supplementary Table S9 was the basis of the MUSCLE alignment.

The sequences in Table S9 were then put in MEGA 6.0 to generate MUSCLE alignments. We confirmed by inspection that the Serine-X-X-Lysine motif that appears in the MUSCLE alignment (Figure 2) agreed with Negoro's X-ray crystallography of NylB.[20,23] The amino acid sequences in PAO1 and NK87 which corresponded to the sequence in KI72 in Kato *et al.* experiment were determined by inspection of the MUSCLE alignment. The distance matrix was generated also by MEGA with the parameters stated in Supplementary Table S8.

DISCUSSION

Ohno's famous 1984 paper claimed to show how a frameshift mutation might have given rise to a totally novel protein such as a nylonase enzyme. Ever since that time it has been widely accepted that this was the correct explanation for the origin of the NylB enzyme.[12,13,14,15,16,17] It has been widely assumed that this happened in an extremely short timeframe, soon after the invention of nylon in 1935. Some people, by extension, have assumed that Ohno's frameshift claim might help explain other nylonases such as NylA and NylC. Most broadly, Ohno's frameshift paper is considered by many to be the best proof of the rapid evolution of a *de novo* gene/protein.

Many readers have not realized that Ohno's 1984 claims were not supported by any type of evidence – his model was entirely speculative. Ohno presented his assertions very forcefully, as if they were facts. It seems that many readers of that paper got the impression that Ohno actually had observational evidence for the existence of his specified precursor protein and his specified frameshift mutation.

Experiments by Kato *et al.* in 1991 and Prijambada *et al.* in 1995 failed to confirm that NylB evolved via frameshift mutation, and in fact argues against Ohno's hypothesis. In the case of Kato's experiment, it suggests NylB nylonase evolution is feasible by as little as two amino acid changes in an ancestral homolog rather than a frameshift mutation affecting 400+ amino acids (as in Ohno's hypothesis).

In the case of Prijambada *et al.*'s experiment where nylon digesting ability was evolved via directed evolution in the lab, the presence of a NylB homolog in strains of *Pseudomonas aeruginosa* PAO1 suggests the lab-based directed evolution of nylon digestion from PAO1 involved point mutations of a pre-existing NylB homolog in PAO1, not a frameshift mutation. A question for future research is whether the mutations in Prijambada *et al.*'s experiment were of the same nature and located in the corresponding amino acid sequence segments as in the 1991 experiments by Kato *et al.* In the results section, we provided our estimate of the location of the corresponding segments of interest between NylB in KI72 and the NylB homologs in PAO1 and NK87. Based on the MUSCLE alignment, the presence of the aspartic acid in the NylB of NK87 at the corresponding location as the NylB in KI72 supports some of the experimental findings of Kato *et al.* 1991 and Negoro *et al.* 2005.[20,23]

Careful reading shows that Ohno's proposed precursor protein and his proposed frameshift mutation were only inferred. Therefore, at that time Ohno did not even have a testable hypothesis. Now, in the age of bioinformatics, we can do what Ohno could not do – we can test his model.

If Ohno's hypothesis were correct, then a protein database search should reveal evidence for the existence of Ohno's hypothetical precursor protein, which should have a history and should have protein homologs. On the flip side, there should be evidence that the NylB protein is a unique protein, with no history and no protein homologs.

Conversely, if Ohno's hypothesis were wrong, then a protein database search should reveal evidence that the hypothetical precursor protein never existed, has no history, and has few if any homologs. At the same time there should be evidence that the NylB protein is *not* unique, and so has a history and numerous homologs.

An objection might be raised that the psi-BLAST and SPARCLE results were artificially inflated by redundancies and spurious hits for NylB because of the search parameters we used. First, these parameters were the defaults set by the NIH, and second, the point of the comparison was to show that even under relaxed parameters, no remote potential homologs of PR.C could be detected. Because NylB is in the family of beta lactamases, and NylA in the family of amidases, NylB and NylA are clearly members of large protein families independent of the psi-BLAST and SPARCLE results.

Our results are very clear. Numerous protein databases show that there is no evidence that Ohno's hypothetical precursor protein ever existed. We found zero instances of the protein, and zero protein homologs.

This by itself is strong evidence against Ohno's hypothesis. But the most conclusive proof that Ohno's hypothesis is false is that the NylB gene is not at all unique – it is found in many organisms, in many habitats, and has a vast number of homologs.

Our search results indicate that homologs of NylB and various other 6-aminohexanoate hydrolases are very abundant. Some organisms with these homologous proteins have been experimentally shown to have the ability to digest nylon,[22,24] but most were not enzymatically tested. While sequence-based gene predictions cannot prove that all such NylB homologs can necessarily degrade nylon, such predictions point to a family of proteins that have very significant homology. 100% of the genes with the NylB designation from our UNIPROT-developed list which also had available CDD pages had beta lactamase domains (Supplementary Table S1). Beta lactamases are considered one of the most ancient proteins[25]. The divergence within the NylB class of enzymes was often very substantial. This precludes the possibility that all such enzymes arose from an isolated frameshift mutation that arose sometime after 1935. It should be obvious that a single frameshift mutation, in just a few decades, could not possibly have proliferated via horizontal gene transfer across a very large number of unrelated organisms found all around the world.

Ohno's hypothesis was based upon Kinoshita's NylB protein sequence. Based on the bits score assigned to this particular NylB gene by CDD (accession COG1680), we found these sequences were strikingly similar. According to CDD, the probability that this similarity to a COG1680 beta lactamase would arise by chance is 2^{-130}. Given the level of non-random similarity of NylBs to beta lactamase domains (Supplementary Table S1), there is clear homology of NylB with beta lactamases, and this is

illustrated by the fact that several entries in GenBank list the same protein as a NylB homolog and beta lactamase simultaneously (Supplementary Table S6).

There are many other problems with Ohno's claims which we do not have space to describe, but Ohno was clearly wrong on seven points.

1. Ohno implied that all nylonase enzymes evolved since 1935.
2. Ohno claimed that the NylB protein was new and unique.
3. Ohno's hypothetical precursor protein appears to have never existed.
4. Ohno's hypothetical frame-shift mutation appears to have never happened.
5. Ohno's claim that a random string of amino acids would give rise to a stable beta lactamase enzyme *in vivo* is not generally credible. This is especially clear in light of the fact that CDD indicates that that the probability of NylB being so similar to beta lactamase by chance is just 2^{-130}.
6. Ohno's claim that an entirely arbitrary amino acid sequence would instantly give rise to a specific and fully functional nylonase enzyme is not credible.
7. Ohno ignored the fact that NylB was not operating independently but was part of a catabolic chain, functioning in coordination with three other nylonases on the same plasmid (NylA, NylB′, and NylC). Indeed, NylB was shown to be co-regulated with NylC, sharing the very same promoter.[10]

Ironically, Ohno pointed out that the level of divergence of the paralogous pair of NylB and NylB′ in KI72 suggests that this paralogous pair existed prior to 1935.[4] A similar level of divergence exists in the paralogous NylB and NylB′ proteins in *Bacilus cereus* even though the NylB in *Bacilus cereus* is around 75% divergent from the NylB in KI72. These considerations cast some doubt on a post-1935 gene duplication hypothesis. Further, it appears such as in the case of *Pseudomonas* NK87 with a functioning nylonase NylB, that having a paralog is unnecessary for the evolution of NylB nylonase activity. But importantly, since Ohno's hypothesis only applies to KI72, it cannot account for the presence of the NylB paralogs in *Bacilus cereus* that are over 75% divergent from their counterparts in KI72 nor the NylB orthologs in the *Pseudomonas* strains.

Taken collectively, our findings very clearly refute Ohno's frame-shift hypothesis. However, our findings are consistent with Yomo *et al.*'s hypothesis that the NylB gene and its homologs have been around for a long time.

We extended our search to look for homologs of other nylonases such as NylB′, NylA, and NylC (all of which were assumed to have evolved since 1935). While Kinoshita did not detect physiological amidase activity for NylA,[1,9] our analysis clearly shows that NylA has amidase homology. Similarly, we found that NylC was homologous to a rare peptidase. We found several proteins had dual classifications such as beta lactamase and 6-aminohexanoate hydrolase (NylB), or amidase and 6-aminohexanoate cyclic hydrolase (NylA). In addition to experiments with a protease like Trypsin,[6] experiments have shown that even triacylglycerol lipases can act as nylonases.[26] Thus it appears that the term "nylonase" could be applied to members of the protease, beta lactamase, amidase, peptidase, and lipase enzyme families. This is in broad agreement with some of Yasuhira *et al.* and Negoro's findings that NylB and NylB′ are in the beta lactamase family, NylA is in the amidase family, and some nylonases have similarity to lipases.13.[23,27] In every case the proteins were found in various organisms and in various natural habitats - along with a great many homologs.

We conclude that all of these nylonases and their close homologs existed prior to 1935, although in some cases there may have been adaptive modifications after 1935. It appears that these various naturally occurring enzymes that happen to be able to degrade nylon have historically acted upon alternative nylon-like substrates.

SUMMARY/CONCLUSIONS –

The focus of this research has been to test Ohno's claim that sometime after 1935 the "nylonase" NylB arose *de novo* via a frameshift mutation in a precursor gene/protein. Ohno's hypothesis has been historically impactful - being considered a powerful proof that new genes and enzymes can instantly arise *de novo*. Although Ohno's model was largely speculative and was never actually tested, it has been uncritically accepted within the scientific community for several decades.

We have now been able to test Ohno's hypothesis, thanks to protein databases that were not available to Ohno. We have used these databases to unambiguously falsify Ohno's hypothesis. Ohno's hypothesis can be shown to be wrong on multiple levels.

More broadly, we have examined the widely-held assumption that there were no enzymes having nylonase activity prior to the invention of nylon in 1935. Ohno shared this assumption with most of the scientists of his day. However, the primary "nylonases" that have been studied (NylA, NylB, NylB′, and NylC), were all found on the same plasmid, functioning in coordination, suggesting that none of these genes/proteins could have arisen *de novo* in the very recent past. Our database searches show that all of these enzymes are widely distributed in the biosphere and have thousands of homologs. We also show that these enzymes belong to well characterized enzyme families that are ancient. It is clear that numerous enzymes existed prior to the invention of nylon, which were acting on other substrates - yet still had "nylonase-like" activity. In the future the term *nylonase* might be used with caution.

ACKNOWLEDGEMENTS
We thank Joe Deweese, Associate Professor in the Department of Pharmaceutical Sciences at Lipscomb University, for reviewing our manuscript, providing constructive feedback, helping with the chemical diagrams, and offering many valuable suggestions.

DATA AVAILABILITY STATEMENT

All relevant data will be made available including that provided in the supplementary section as well as in the public databases of UNIPROT, GenBank, SPARCLE, CDD as described in the materials and methods section.

CONTRIBUTIONS

SC did the bioinformatics searches and compiled results. SC and JS analyzed the results and wrote the main manuscript. Both authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

[1] Kinoshita, S., *et al*. 6-Aminohexanoic Acid Cyclic Dimer Hydrolase. A New Cyclic Amide Hydrolase Produced by Acromobacter guttatus KI72. *Eur. J. Biochem*. **80,** 489-495 (1977).

[2] Negoro, S.   Biodegradation of nylon oligomers. *Appl Micriobiol Biotechnol* (2000) **54**:461-466.

[3] Okada, H., Negoro, S., Kimura, H. & Nakamura, S. Evolutionary adaptation of plasmid-encoded enzymes for degrading nylon oligomers. *Nature*. Vol. **306,** 10 November 1983.

[4] 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.

[5] Yomo, T., Urabe, I. & Okada, H. No stop codons in the antisense strands of the genes for nylon oligomer degredation. *Proc. Natl. Acad. Sci. USA*. Vol **89,** pp. 3780-3784, May 1992. Evolution.

[6] Ebata, M. & Morita, K. Hydrolysis of ε-aminocaproyl compounds by trypsin. *The Journal of Biochemistry*, Vol. **46,** No 4, 1959.

[7] Fukumura, T. Hydrolysis of cyclic and linear oligomers of 6-aminocaproic acid by bacterial cell extract. *The Journal of Biochemistry*. Vol. **59,** No. 6, 1966.

[8] Fukumura, T. Two bacterial enzymes hydrolyzing oligomers of 6-aminocaproic acid. *The Journal of Biochemistry*. Vol. **59,** No. 6, 1966.

[9] Negoro, S., Taniguchi, T., Kanaoka, M., Kimura, H. & Okada, H. Plasmid-determined enzymatic degradation of nylon oligomers. *Journal of Bacteriology*. Vol. **155,** No 1, July 1983, p. 22-31.

[10] Negoro, S., Kakudo, S., Urabe, I. & Okada, H. A new nylon oligomer degradation gene (nylC) on plasmid pOAD2 from a *Flavobacterium* sp. *Journal of Bacteriology*. Vol. **174,** No 24, Dec. 1992, p. 7948-7953.

[11] Kinoshita, S., *et al*. Purification and Characterization of 6-aminohexanoic-acid-oligomer hydrolase of flavobacterium sp. KI72. *Eur J. Biochem*. **116,** 547-551 (1981).

[12] Sabath, N., Wagner, A., Karlin, D. Evolution of viral proteins originated *de novo* by overprinting. *Mol Biol Evol*. 2012 Dec; 29(12):3767-3780.

[13] Ruiz-Orera, J., *et al*. Origins of *de novo* genes in human and chimpanzee. *PLos Genet*. 2015 Dec; 11(12):e1005721.

[14] Dias, R., Austin, M., Kolaczkowski, O. & Kolaczkowski, B. Convergence of Domain Architecture, Structure, and Ligand Affinity in Animal and Plant RNA-Binding Proteins. *Mol Biol Evol*. 2017 Jun;34(6): 1429-1444.

[15] Fellner, L., *et al*. Evidence for the recent origin of a bacterial protein-coding, overlapping orphan gene by evolutionary overprinting. *BMC Evol Biol*. 2015; 15: 283.

[16] Neme, R. & Tautz, D. Phylogenetic patterns of emergence of new genes support a model of frequent *de novo* evolution. *BMC Genomics*. 2013; 14:117.

[17] Guogin, Y. & Stoltzfus, A. Population diversity of ORFan Genes in *Escherichia coli*. *Genome Biol Evol*. 2012; 4(11):1176-1187.

[18] Kanagawa, K., Negoro, S., Takada, N. & Okada, H. Plasmid dependence of *Pseudomonas* sp. Strain NK87 enzymes that degrade 6-aminohexanoate-cyclic dimer. *Journal of Bacteriology*. Vol. **171,** No. 6, June 1989, p. 3181-3186.

[19] Kanagawa, K., Mayumi, O., Negoro, S., Urabe, I. & Okada, H. Characterization of the 6-aminohexanoate-dimer hydrolase from *Pseudomonas* sp NK87. *Journal of General Microbiology*. (1993), **139,** 787-795.

[20] Kato, K., *et al*. 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).

[21] Prijambada, I.D, Negoro, S., Yomo, T. & Urabe, I. Emergence of nylon oligomer degradation enzymes in *Pseudomonas aeruginosa* PAO through experimental evolution. *Applied and Environmental Microbiology*. May 1995, p. 2020-2022.

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

[23] Negoro, S., *et al*. X-ray crystallographic analysis of 6-aminoohexanoate-dimer hydrolase. The *Journal of Biological Chemistry*. **Vol 280,** No. 47, pp. 39644-39652, November 25, 2005.

[24] Adeyemi, S.H., Ogunjobi, A.A. & Fagade, O.E. Hydrolytic degradation of nylon-6 by *Pseudomonas aeruginosa* HE858284 isolated from solid waste dumpsites in Lagos State, Nigeria. *Journal of Natural Sciences Research*. **Vol. 5,** No. 2, 2015.

[25] Hall, B.G. & Barlow M. Evolution of the serine beta-lactamases; past, present and future. *Proc Natl. Acad. Sci. USA*. **Vol 89,** pp. 3780-3784, May 1992.

[26] Kiumarsi, A & Parvinzadeh, M. Enzymatic hydrolysis of nylon 6 fiber using lipolytic enzyme. *The Journal of Biochemistry*. Vol **59,** No.6, 1996.

[27] Yasuhira, K., *et al*. X-ray Crystallographic analysis of the 6-aminohexanoate cyclic dimer hydrolase. *Journal of Biological Chemistry*. **285**(2): 1239-1248. 2010 Jan 8.