Title Page

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

Authors: Salvador Cordova[1] and John Sanford[2]
[1]FMS Foundation, 75 Lafayette Center, Lafayette Avenue, Canandaigua, NY 14424, USA, corresponding author;
[2]FMS Foundation, 75 Lafayette Center, Lafayette Avenue, Canandaigua, NY 14424, USA.

Salvador Cordova email – acordov4@alumni.jh.edu
John Sanford email – jsanford@fmsfound.org

Corresponding author details:
Salvador Cordova
FMS Foundation
Phone: 1-703-283-2497

Abstract

Background
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, the frameshift hypothesis was never actually tested, and there was no actual supporting biological evidence supporting the hypothesis. For decades, the frameshift hypothesis has been uncritically accepted as the correct explanation for the origin of NylB. In this paper we have surveyed the literature relevant to the frameshift hypothesis as well as the various alternative models that have been published regarding the origin of NylB. We have employed bioinformatic methods and leveraged databases not available when the frameshift hypothesis was first put forward.

Results
We searched multiple protein databases to determine the distribution of NylB and any possible homologs. We then determined the distribution of other known nylonases and their possible homologs. We also determined the distribution of Ohno's hypothetical PR.C. protein and any possible homologs. Lastly, we determined what protein families the various nylonases belong to.

We found that the NylB protein is widely occurring, has thousands of homologs, and is found in diverse organisms and diverse habitats. It is not a new or unique protein. Likewise, we found that the other known nylonases are also widely occurring, have thousands of homologs, and are found in diverse organisms and diverse habitats. However, the hypothetical PR.C protein does not show up in any of the same databases, and there is no evidence of any homologs. Conserved domain searches showed that NylB is a member of the beta lactamase protein family – a highly conserved family of enzymes. Likewise, the other known nylonases belong to well-characterized enzyme families.

Conclusions
Our results very effectively falsify the NylB frameshift hypothesis, while they strongly support an alternative hypothesis by Yomo. Like NylB, none of the other nylonases that we examined were substantively new or unique. All had thousands of homologs, and they were found in diverse organisms and diverse habitats. Our findings not only falsify the NylB frameshift hypothesis, they also falsify the long-held assumption that all nylonases evolved after the invention of nylon in 1935.

Keywords

Nylonase, NylB, NylA, NylC, Beta lactamase, 6-aminohexanoate hydrolase, Frameshift

Background

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 primary 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] and 3) Yomo *et al.*'s pre-1935 NylB homologs hypothesis. [5]

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 measurable or predicted nylonase activity.
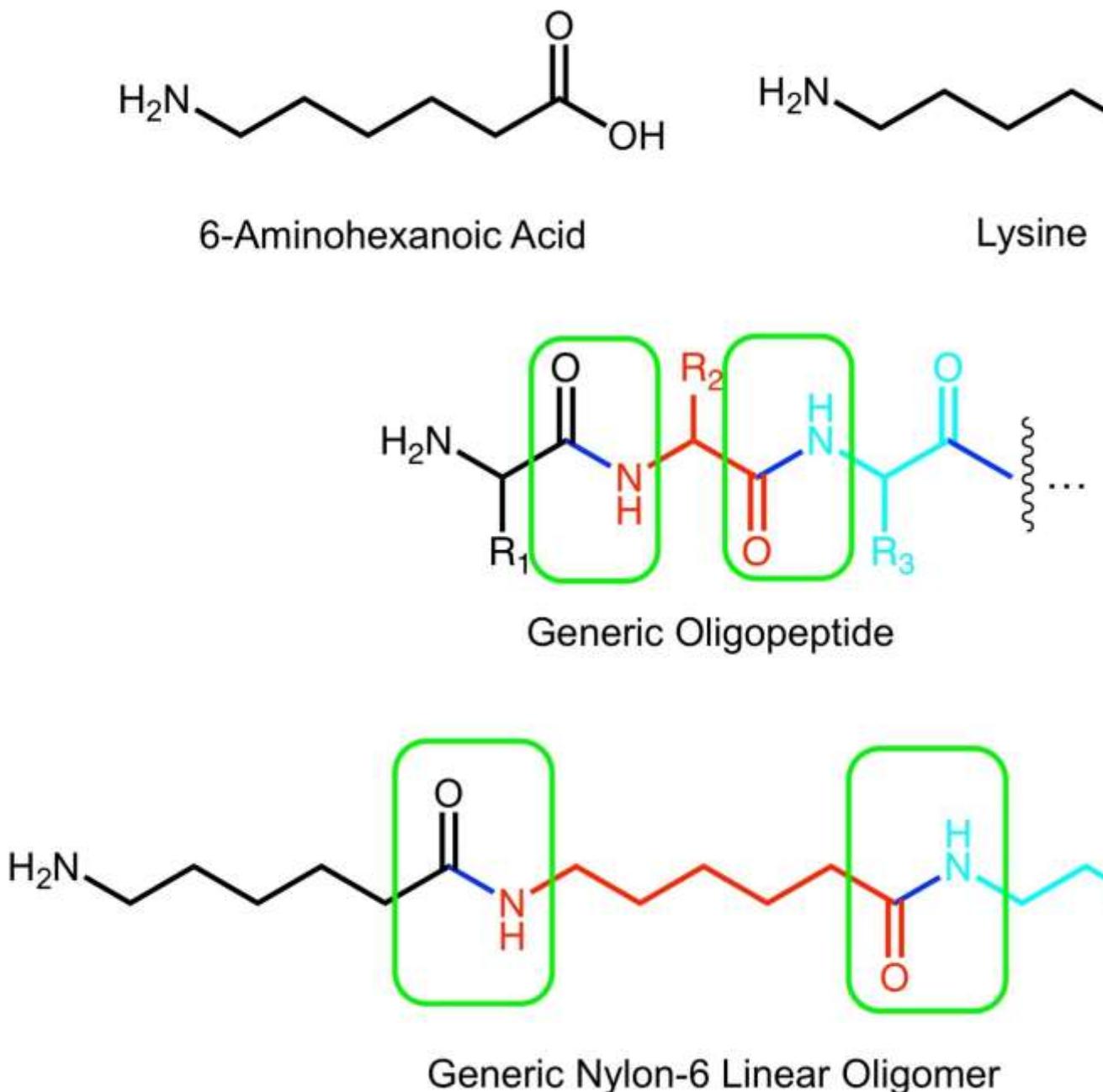
Figure 1

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). The squiggly line and dots at the 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 *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, was discovered on the same plasmid in the same bacterium. [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 specific 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 required 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 prior to 1935, and he gave no reason why NylB' might not have arisen from NylB instead of the reverse.

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 then 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 the Nylb frameshift hypothesis required an essentially random amino acid sequence to arise from a frameshift and to instantly form a stable, functional, and specific enzyme. Ohno had no direct evidence that the hypothetical PR.C protein even existed because his frameshift mutation was purely hypothetical. Yet the NylB frameshift hypothesis was stated 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 the NylB frameshift hypothesis was put forward, 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 had 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 still sufficiently homologous to rule 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."

In 2007, Sudhakar demonstrated that strains of *Bacilus cereus* found in the Indian Ocean could digest nylon-6. [22] To confirm this, we searched protein data bases for evidence of NylB in *Bacilus cereus.*

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 now readily testable using protein databases.

If the NylB frameshift hypotheses 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 the NylB frameshift hypotheses 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

We used a spectrum of search strategies with differing criteria to detect homologs or exact sequences.   On one end of the spectrum were searches that were strongly constrained and thus expected to have few if any false positives but may admit a number of false negatives (rejected sequences), and on the other end of the spectrum searches that had relaxed parameters and thus expected to have fewer false negatives but admit a number of false positives. The searches are shown in Table 1, where the most constrained searches results are toward the left, and the relaxed searches toward the right.   There was almost a complete absence of the predicted PR.C across the entire spectrum of search strategies. Reruns of the BLASTP search often resulted in numbers ranging from 0 to 9 hits when no limit was imposed for e-values, hence the number was reported as "unstable."

Since PR.C corresponds to an exact gene sequence with a specific insertion mutation, a highly relaxed BLASTN would flood the results with false positives of unshifted nylonase homologs. Hence, the more relaxed search strategies were conducted using BLASTP on the predicted amino acid sequence of PR.C rather than BLASTN on the nucleotide sequence. However, the left column shows the search result of a constrained BLASTN search where the number of hits for the PR.C DNA sequence was adjusted for hits missing the critical segment (DNA sequence coordinates 99-100) where the supposed insertion mutation that created a new start codon was claimed by Ohno to occur after 1935.

The SPARCLE numbers were not adjusted for redundancies and hence yielded many false positives but were provided to give an idea of the degree of representation of the homologs in the databases which SPARCLE surveys. The number "44" for NylC in the rightmost column strongly contrasts with the relatively large numbers for the other nylonases.   The figure of 10,250, 46,000, and 68,000 for 6-AH Hydrolase were composite scores produced by adding the BLASTP and SPARCLE hits for NylA, NylB, NylC, respectively (treating numbers for NylB and NylB' as redundant).

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

Protein Name	BLASTN on PR.C and KI72 Sequences	Named Entries in UniProt	BLASTP on PR.C and KI72 Sequences e-value < e^{-50}	BLASTP on PR.C and KI72 Sequences any e-value	NIH SPARCLE Potential Homologs
NylB	6	201	5000+	20,000+	68,000+
NylB′	6	134	5000+	20,000+	68,000+
NylA	8	193	5000+	20,000+	53,000+
NylC	7	7	250+	6,000+	44
6-AH Hydrolase	21	2115	10,250+ (composite)	46,000+ (composite)	68,000+ (composite)
Ohno's PR.C	0 (adjusted)	0	0	0-9 (unstable)	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).

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

The data used to construct Table 1 and 2 can be found in the in Supplementary Tables S1, S2, S3, S4, and S5.

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}.

A representative sample of these proteins (mostly those reported with experimentally verified nylonase activity) 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 has been confirmed by X-ray crystallography 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). AAG07735.1 was used as one of the proteins featured in the MUSCLE alignment.
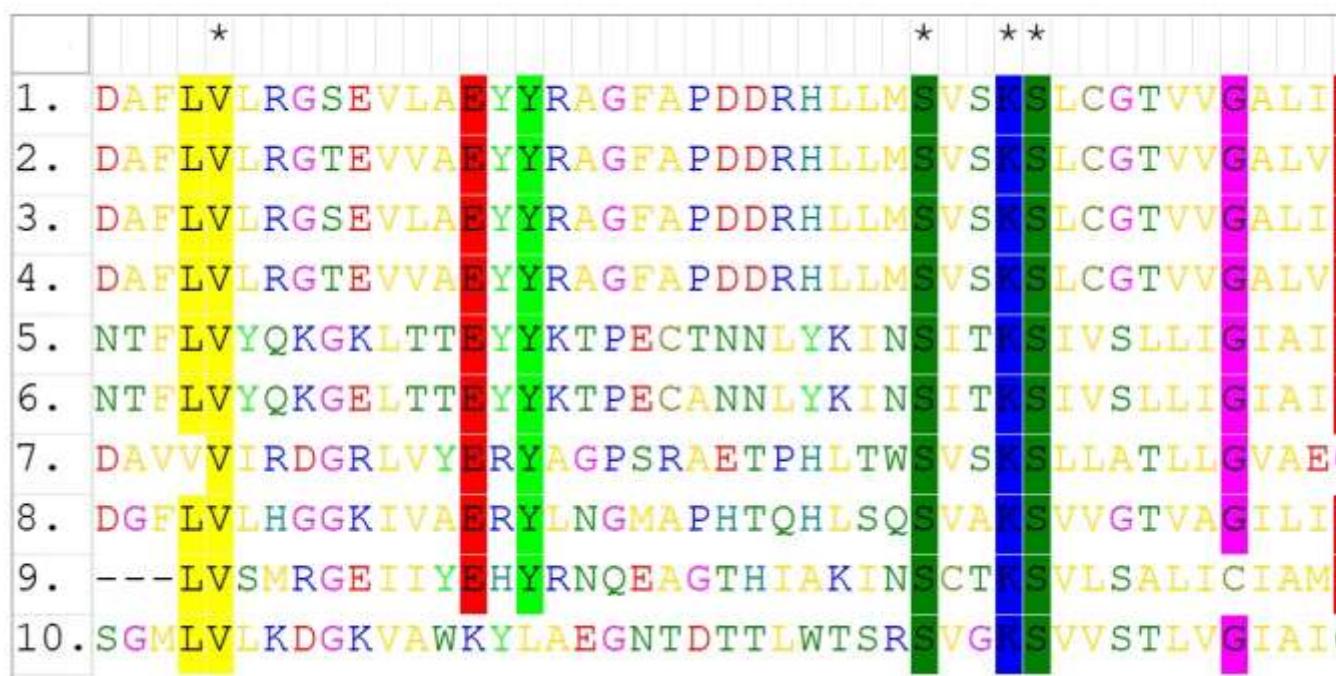
Figure 2

MEGA 6.0 MUSCLE Alignment of a 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 complete MUSCLE alignment is available in Supplementary Table S8.

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.

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. By extension, it might be erroneously assumed that the NylB 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 they argue 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.

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. In addition to the experimental evidence that could reasonably be deemed sufficient to refute Ohno's hypothesis, now, in the age of bioinformatics, we can do what Ohno could not do – we can further test his model using bioinformatic tools.

If the NylB frameshift hypotheses 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 the NylB frameshift 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.

Table 2 shows NylB is in the family of beta lactamases, and NylA is in the family of amidases. Therefore, NylB and NylA are both clearly members of very well-known protein families (independent of the BLAST and SPARCLE results).

Although it could be argued that the absence of PR.C in the databases might be due to the fact it might exist but simply has not yet been found, the most conclusive proof that the NylB frameshift 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.

The searches for PR.C were conducted across a broad spectrum of strategies that would result in varying levels of false negatives and false positives.   Even using the most generous search strategies, there was no credible evidence of the PR.C sequence.

Our BLASTP searches for PR.C included extremely relaxed constraints, such that false positive hits for PR.C would be expected. Furthermore, it would be expected that if PR.C existed long before 1935, while NylB only emerged after 1935, then PR.C should have many more homologs than NylB. The relative representation of BLASTP hits for PR.C versus NylB hits argues strongly against PR.C's actual existence.   Thus, the unstable numbers of BLASTP hits (ranging from 0 ot 9) for PR.C with weak e-values can be reasonably regarded as both spurious and unconvincing of PR.C existence.

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. All 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 frame shift 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. We computed the sequence similarity of NylB to the architecture of the beta lactamase protein family. 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 degree of non-random similarity of NylBs to beta lactamase domains (Supplementary Table S1), there is no doubt of the 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).

The NylB frameshift hypothesis was premised upon numerous assumptions that we now know are incorrect, and so his hypothesis is falsified on several levels:

1. The widely held assumption that all nylonase enzymes evolved since 1935 was incorrect.
2. Ohno's assumption that the NylB protein was a new and unique protein was incorrect.
3. Ohno assumed a hypothetical but specific precursor protein that now appears to have never existed, and thus the hypothetical frame-shift mutation appears to have never happened.
4. Ohno claimed that a random string of amino acids could reasonably be expected to give rise to a specific, functional, beneficial, and stable enzyme. Having all these things happen by chance is so incredibly unlikely that it is hard to imagine. This is especially clear in light of the fact that CDD database indicates that that the probability of NylB would be so similar to beta lactamase by chance would be essentially impossible (2^{-130}).
5. Lastly, it was established 8 years after the frameshift hypothesis was published that NylB was not operating independently, but was clearly 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] Ohno had no access to this data at the time he published his hypothesis.

Ironically, Ohno pointed out that the level of divergence of paralogous pair of NylB and NylB' in KI72 suggests that this paralogous pair must have 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 themselves cast serious doubt on any post-1935 gene duplication hypothesis. Furthermore, it appears that in the case of *Pseudomonas NK87* (with a functioning nylonase NylB), that having a paralog is unnecessary for the evolution of NylB nylonase activity. Importantly, since the frameshift 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.

Overlapping reading frames are known to exist in biology, [26,27] and Okamura [28] has speculated that several such human genes may have originated via frame shift mutations. However, the larger question of *de novo* origination of genes and proteins in general and the role of frameshifts specifically in creation of *de novo* proteins is beyond the scope of this paper.   Our present focus is specifically on whether a frameshift mutation after 1935 was the mechanism that created NylB.

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 proteases like Trypsin,[6] experiments have shown that even triacylglycerol lipases can act as nylonases.[29] 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 share some passing similarities to lipases. [13,23,30] 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.

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*. While the frameshift model was speculative and was never actually tested, it has been uncritically accepted within the scientific community for several decades.

In the last three decades, there have been some authors who have questioned the frameshift hypothesis and have proposed alternative explanations of how NylB might have arisen. While that work did suggest the frameshift hypothesis might be wrong, these authors did not rigorously falsify the frameshift hypothesis, and so the frameshift model has continued to be cited as a fact – even up to the time of this writing.

Thanks to protein databases and bioinformatic tools that were not available until quite recently, we have been able to unambiguously falsify Ohno's hypothesis. We have shown that Ohno's hypothesis can be falsified 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 had previously been acting on other substrates, but also happened to have "nylonase-like" activity. In the future the term *nylonase* might be used with more caution.

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 uniprot.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 ID 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 BLAST searches were accomplished through the NCBI protein BLAST website. The 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 = 6, 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.

The homologs chosen to search for NylB, NylB', NylA, NylC were from Flavobacteria KI72 since these were the first named homologs in Ohno's paper as well as subsequent papers studying nylonases in Flavobacteria KI72.

In Table 1, The figure of 10,250, 46,000, and 68,000 for 6-AH Hydrolase were composite scores produced by adding the BLASTP and SPARCLE hits for NylA, NylB, NylC, respectively (treating numbers for NylB and NylB' as redundant).

To verify the possible conflict in assigning functions to protein predictions, a manual search through several thousand potential NylB and NylA homologs created by 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*, 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.

Supplementary Table S10 contains our reconstruction of the PR.C amino acid sequence from Ohno's paper and the Genebank data with the original X00046.1 DNA sequence and the relevant verbatim segment of X00046.1 with capitalization and spacing to highlight areas Ohno modified to create PR.C.

List of Abbreviations
BLAST – Basic Local Alignment Search Tool
CDD – Conserved Domain Database
MEGA – Molecular Evolutionary Genetics Analysis
MUSCLE – Multiple Sequence Comparison by Log-Expectation
NCBI – National Center for Biotechnology Information
NIH – National Institutes of Health
SPARCLE – Subfamily Protein Architecture Labeling Engine
UniProt – Universal Protein Resource

Declarations

Ethics approval and consent to participate

Not applicable.
Consent for publication

Not applicable.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files which were derived from public databases of UNIPROT, GenBank, SPARCLE, CDD as described in the materials and methods section.

Competing interests
The authors declare they have no competing interests.

Funding

Not applicable.

Authors' 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.

Acknowledgements
Not applicable.

Authors' information
JS is a retired research professor, Cornell University and President of FMS Foundation, Inc. SC is a research associate for FMS Foundation, Inc.

References

1. Kinoshita S, Negoro S, Muramatsu M, Bisaria VS, Sawada S, Okada H. 6-Aminohexanoic Acid Cyclic Dimer Hydrolase A New Cyclic Amide Hydrolase Produced by Acromobacter guttatus KI72. Eur. J. Biochem. 1977;80:489-495.

2. Negoro S.   Biodegradation of nylon oligomers. Appl Micriobiol Biotechnol 2000 October;54:461-466.

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

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, Evolution. 1984 April;81:2421-2425.

5. Yomo T, Urabe I, Okada H. No stop codons in the antisense strands of the genes for nylon oligomer degradation. Proc. Natl. Acad. Sci. USA, Evolution. 1999 May;89:3780-3784.

6. Ebata M, Morita K. Hydrolysis of ε-aminocaproyl compounds by trypsin. The Journal of Biochemistry. 1959;46**(**4).

7. Fukumura T.   Hydrolysis of cyclic and linear oligomers of 6-aminocaproic acid by bacterial cell extract. The Journal of Biochemistry.1966;59(6).

8. Fukumura, T. Two bacterial enzymes hydrolyzing oligomers of 6-aminocaproic acid. The Journal of Biochemistry. 1966;59(6).

9. Negoro S, Taniguchi T, Kanaoka M, Kimura H, Okada H. Plasmid-determined enzymatic degradation of nylon oligomers. Journal of Bacteriology. 1983 July;155(1): 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. 1992 Dec;174(24):7948-7953.

11. Kinoshita S, Terada T, Taniguchi T, Takene Y, Masuda S, Matsunaga N, Okada H. Purification and Characterization of 6-aminohexanoic-acid-oligomer hydrolase of flavobacterium sp. Ki72. Eur J. Biochem. 1981;116:547-551.

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, Hernandez-Rodriguez J, Chiva C, Sabido E, Kondova I, Bontrop R, *et al*. Origins of *de novo* genes in human and chimpanzee. PLos Genet. 2015 Dec 31; 11(12):e1005721. doi: 10.1371/journal.pgen.1005721

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, Simon S, Scherling C, Witting M, Schober S, Polte C, *et al*. Evidence for the recent origin of a bacterial protein-coding, overlapping orphan gene by evolutionary overprinting. BMC Evol Biol. 2015 Dec 18;15: 283. doi: 10.1186/s12862-015-0558-z

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. 1989 June;171(6):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, Fujiyama K, Hatanaka HS, Priyambada ID, Negoro S, Urabe I, Ikada H. Amino acid alterations essential for increasing the catalytic activity of the nylon-oligomer-degradation enzyme of *Flavobacterium* sp. Eur J. Biochem. 1991 Aug 15;200(1):165-169.

21. Prijambada ID, Negoro S, Yomo T, Urabe I. Emergence of nylon oligomer degradation enzymes in *Pseudomonas aeruginosa* PAO through experimental evolution. Applied and Environmental Microbiology. 1995 May;61(5):2020-2022.

22. Sudhakar M, Priyadarshini C, Doble M, Murthy PS, Venkatesan R. Marine bacteria mediated degradation of nylon 66 and 6. International Biodeterioration & Biodegradation. 2007;60(3):144-151.

23. Negoro S, Ohki T, Shibata N, Mizuno N, Wakitani Y, Tsurukame J, *et al*. X-ray crystallographic analysis of 6-aminoohexanoate-dimer hydrolase. The Journal of Biological Chemistry. 2005 Nov 25;280(47):39644-39652.

24. Adeyemi SH, Ogunjobi AA, Fagade OE. Hydrolytic degradation of nylon-6 by *Pseudomonas aeruginosa*HE858284 isolated from solid waste dumpsites in Lagos State, Nigeria. Journal of Natural Sciences Research. 2015;5(2).

25. Hall BG, Barlow M. Evolution of the serine beta-lactamases; past, present and future. Proc Natl. Acad. Sci. USA. 1992 May, 89:3780-3784.

26. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. J Virol. 2009 Oct;83(20):10719-36. Doi: 10.1128/JVI.00595-09.

27. Sherr CJ. Divorcing ARF and p53: an unsettled case. Nat Rev Cancer. 2006 Sep;6(9):663-73.

28. Okamura K, Feuk L, Marques-Bonet T, Navarro A, Scherer SW. Frequent appearance of novel protein-coding sequences by frameshift translation.   Genomics. 2006 December;88(6):600-697.

29. Kiumarsi A, Parvinzadeh M. Enzymatic hydrolysis of nylon 6 fiber using lipolytic enzyme. The Journal of Biochemistry. 1996;59(6).

30. Yasuhira K, Shibata N, Mongami G, Uedo Y, Atsumi Y, Kawashima Y, Hibino A, *et al.* X-ray Crystallographic analysis of the 6-aminohexanoate cyclic dimer hydrolase. Journal of Biological Chemistry. 2010 Jan 8;285(2):1239-1248.