marine creatures living in those environments.

or burning it.

The more labor intensive solution is to burn

the plastic to help regain some of the energy lost from its production.

This comes with the

harmful fumes which can wreak havoc amongst various organs and body systems [1].

Placing

it into the ocean and landfill

are both quick and easy,

yet have vast ecological

impacts for

to placing it into a landfill

Bioremediation is a more viable solution for plastic waste management as it converts

harmful

polutants into harmless byrpoducts.

Bioremediation can be helpful

because it is

cheaper and can provide a more long-term low-tech solution [7].

While it is true that biore-

mediation is slower and incredibly plastic specific, it’s important to remember that enzyme

design could help improve upon the specificity by designing enzymes to degrade various poly-

1

6].

Jonah Nichols

May 20, 2022

1

Introduction

Plastic is an incredibly versatile and cheap material.

It is used in nearly every industry,

from packaging to consumable goods to medicine to research [4].

Estimates for global plastic

production range from 335 MT to 380 MT [3,

4,

Applications of Machine Learning in Enzyme Design

This quantity is high due to its high

versatility and durability.

After its usage though,

plastic lingers in range from a few years

to decades.

Polyethylene terephthalate,

for example,

takes one year to degrade one µm of

plastic from the outside in [2]

Plastic waste management can be done in a few different ways, including but not limited

Given inexperience in the subject,

a one-hot encoding of

the sequence was built to create a suitable

output for the reaction.

Each amino acid was represented by an integer.

These sequences

were then converted into PyTorch tensors for input into the machine learning algorithm

paired up with its corresponding reaction fingerprint.

These pairs were split randomly into

a ratio of 75% training and 25% test datasets

2.2

Neural Network Design

For every sequence,

the neural

network was designed to a very low degree

of

confidence.

The input layer was the reaction fingerprint previously generated.

Three

repetitions of linear layers followed by ReLU activation layers expanded the fingerprint out

to 7192 bits, corresponding to the longest sequence found.

A final linear layer presented the

output of the model.

2

reactions from UniprotKB.

Not only could this bioremediate plastic-contaminated samples,

it could also handle

other chemical compounds [5].

2

Methods

2.1

Data Acquisition

RheaDB was utilized to find ester hydrolysis reactions according to the Enzyme Classifica-

tion (EC) standards.

From selected reactions, Rhea’s cross references were utilized to obtain

sequences for the various chemical

mers.

For every sequence-reaction

pair,

the Morgan Fingerprint Difference of the reaction was created using RDKit’s built-in

fingerprinting function.

Due to a lack of known directional

data,

unknown directional

en-

zymes were assumed to catalyse the left-to-right version of the reaction.

Next the fingerprint

was converted into a PyTorch tensor for import into the machine learning algorithm.

of

encoded sequence in number form, to be converted back to a proper sequence.

Unfortunately

the output received was a sequence of numbers between 12-14 and then a gradual

descent

to 7’s, presumably the ”pad” character created to sync

4

Discussion

Future work on the subject is key to enhancing our ability to create and use enzymes to

catalyze reactions.

The most arbitrary extension of this work is in rewriting the featurization

The output given was to be a one-hot

the enzymes to more resemble other machine learning projects.

Rather than represent

each amino acid as a number,

represent it as a 21-bit tensor.

Each bit would represent a

different amino acid being turned on.

Another more laborious route for the project is to move towards a 3D-based approach

opposed to the current sequence-based approach.

This route is more difficult due to the

3

the model’s learning capabilities and pattern recognition were

Training and Analysis

For training, the inputs and outputs were batched to 64 samples and shuffled to limit over-

fitting.

The L1 Loss function was used to determine the error of

the model.

Stochastic

gradient descent was utilized to optimize the model.

The training sequence was run for 100

epochs to get an optimized model.

Due to time constraints,

2.3

not determined.

3

Results

Of

the 9890 sequences queried from UniprotKB with the enzyme class of

3.1.x.x,

19084

reaction-sequence pairs were generated and used for training in the machine learning model.

In a visual examination of the output the neural network gave, it became apparent that

the neural network had not learned much of the data.

Chandrakant S Karigar and Shwetha S Rao.

Ali Chamas et al. “Degradation rates of plastics in the environment”. In: ACS Sustain-

able Chemistry & Engineering 8.9 (2020), pp. 3494–3511.

[3]

Oliver Drzyzga and Auxiliadora Prieto.

“Plastic waste management,

a matter for the

‘community’”. In: Microbial

biotechnology 12.1 (2019), p. 66.

[4]

Roland Geyer, Jenna R Jambeck, and Kara Lavender Law. “Production, use, and fate

of all plastics ever made”. In: Science advances 3.7 (2017), e1700782.

[5]

[2]

“Role of microbial

enzymes in the biore-

mediation of pollutants: a review”. In: Enzyme research 2011 (2011).

[6]

Walter Leal Filho et al. “An overview of the problems posed by plastic products and the

role of extended producer responsibility in Europe”.

In:

Journal

of cleaner production

214 (2019), pp. 550–558.

4

transition state using density-functional theory (DFT) and building up the enzyme around

input for a neural

network.

Before the

fasta-approach branch it was intended as a structure-based approach but processing times

were too long for a limited time scale.

Another another pathway to take would be returning the project to a more typical enzyme

design process.

This would involve completely refactoring all

of the work done so far,

but

may offer better end results.

The process, described briefly, is beginning with the simulated

conversion from a structure to a one-dimensional

said transition state.

Regardless, continuing work on enzyme design is irrefutably important.

Creating enzymes

from scratch has possibilities ranging from new antibiotics to ecological benefits.

References

´

[1]

Nagy

Agnes and KUTI Rajmund. “The environmental impact of plastic waste incinera-

tion”. In: AARMS–Academic and Applied Research in Military and Public Management

Science 15.3 (2016), pp. 231–237.

[7]

Mn Vidali. “Bioremediation. an overview”. In: Pure and applied chemistry 73.7 (2001),

pp. 1163–1172.

5