Chemical Engineering Research Bulletin 24 (2024) 1-12
Available online at http://banglajol.info/index.php/CERB
DOI: https://doi.org/10.3329/cerb.v24i1.86724

A Bioinformatic Approach to the Bioprospecting of Plastic Waste Degrading Enzymes

Joseph Maskery[1] and M. Ahsanul Islam[1]*

[1]Department of Chemical Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK

*Correspondance: m.islam@lboro.ac.uk

Abstract

Plastic waste pollution poses a major threat to environmental sustainability, and the biochemical degradation of plastics through enzymatic processes offers an attractive and viable solution to this menacing problem. However, identifying suitable plastic degrading enzymes is a significant challenge. Traditional bioprospecting methods such as high-throughput enzyme screening, although identified several experimentally validated enzymes, are largely unsuccessful, time consuming, and expensive. Bioinformatics can offer a cheaper, quicker, and more successful approach to bioprospecting of novel plastic degrading enzymes. This study describes the development of an automated bioinformatics pipeline to identify proteins homologous to a query protein using a range of criteria, including sequence similarity, evolutionary relationship, and structural alignment. Using the sequences of known plastic degrading enzymes as input, the pipeline identified four homologous proteins with a high potential for plastic degrading functionality: Lipase, Lipase 1, and two uncharacterized esterases, XCC2094 and ATU5261. These proteins and the microorganisms which produce them can be tested *in vitro* to confirm their plastic degrading abilities, with the aim of identifying microorganisms capable of degrading all the seven types of plastics, or simply providing better plastic degrading capabilities.

Keywords: Plastic waste, polyethylene terephthalate, polyurethane, bioprospecting, bioinformatics, sequence similarity, phylogenetic analysis, structural alignment.

1. Introduction

Plastic pollution is one of the biggest environmental problems currently facing the modern society. The specific properties of plastics, including durability, light weight, and chemically non-reactive, make them the perfect material for thousands of everyday uses; however, these same properties also make them incredibly resistant to the natural process of degradation, which can take thousands of years. In 2015, around 6.3 billion tonnes of total plastic waste were generated, with 79% of that ending up in landfill sites. By 2050, it is predicted that 12 billion tonnes of plastic waste will be in landfill sites or in the natural environment (Geyer, et al., 2017). Not only does the plastic waste take up a lot of space in landfill sites; if leaked into the natural environment, it can disrupt ecosystems, harm wildlife, and contaminate the food chain, which ultimately ends up in humans (Seltenrich, 2016).

Current methods of dealing with plastic waste come with significant limitations: only certain plastics can be recycled, incineration releases hazardous ash and chemicals (e.g., greenhouse gasses) into the environment, and there is only so much space available for landfill (Gourmelon, 2015). Microorganisms capable of degrading plastics have provided a promising and attractive alternative to plastic waste disposal. Certain microorganisms produce enzymes

capable of breaking down plastic polymers. For example, Ideonella sakaiensis is a microorganism that produces the enzyme PET hydrolase, capable of degrading the plastic polyethylene terephthalate (PET) (Han, et al., 2017). Whilst this shows a great promise for enzymatic degradation of plastics, there is currently little knowledge about the existence of other types of plastic degrading enzymes. So far, only two of the seven main types of plastic by production (PET and polyurethane) are known to be degraded by known, experimentally validated plastic degrading enzymes. However, no enzyme has been identified yet to degrade the other five types of plastics: polyethylene, polypropylene, polyphthalamide, polyvinyl chloride, and polystyrene. In addition to this, methods of scaling-up the degradation of plastics via microorganisms to tackle the millions of tonnes of plastics being produced each year are currently non-existent (Ghosh, et al., 2013). For the microbial enzymatic degradation of plastics to be a viable option, enzymes must be able to degrade all types of plastics at a suitable rate, so that the process can be scaled up to depolymerise the millions of tonnes of plastics being produced each year. Hence, the need for the discovery of new plastic degrading enzymes and microorganisms is clear.

The search and evaluation of biological materials for use in valuable products and applications is known as bioprospecting (Artuso, 2002). Traditional methods for bioprospecting of novel enzymes are generally very time consuming, expensive, and resource-heavy, with a low success rate. These methods include the screening of natural samples such as soil and industrial waste, random mutagenesis to create a preferred enzyme mutant, or the sequencing of entire genomes in attempt to find a desired enzyme gene sequence (Singh, et al., 2019). Bioinformatics, on the other hand, has presented researchers with multiple new methods of bioprospecting for novel enzymes (Greenbaum, et al., 2001). Tools such as sequence similarity searches can be used to search biological databases for proteins with similar amino acid sequences. Therefore, given a query protein with a known, biological function, it is possible to find a range of candidate proteins that are highly likely to share a similar, if not the same, function. Not only is this process much more efficient than conventional screening methods, but it also requires far less resources and financial commitment whilst providing a higher success rate (Singh, et al., 2019). Moreover, several individual bioinformatic programs can be combined to create an analysis workflow or pipeline, which can be further automated to obtain the desired, final output from a single user input quickly and efficiently.

The primary aim of this project is to identify novel enzymes, having a similar functionality to the experimentally validated, known plastic degrading enzymes through developing an automated bioinformatic workflow or pipeline. The identification of novel enzymes will help identify their corresponding microorganisms, with the aim of finding microorganisms capable of degrading the remaining five types of plastics. The bioinformatic pipeline consists of an automated sequence similarity analysis workflow and an additional protein structure similarity analysis. A sequence similarity network will be created in addition to a phylogenetic tree, displaying a wider, more general look into the relationships of the protein families around the query proteins. The novel, potentially plastic degrading enzymes and microorganisms identified in this study would benefit from further in vitro analysis to confirm their plastic degrading abilities.

2. Materials and methods

Fig. 1 shows the overall bioinformatic pipeline developed in this study and used for identifying novel plastic degrading enzymes. As seen in Fig 1, the workflow is split into 3 parts: an automated pipeline to analyse sequence similarity (green), the protein structure analysis (red), and the sequence similarity network generation (orange). All parts of the workflow start with the input of a query protein sequence. For the purpose of this study to find novel plastic degrading enzymes and microorganisms, the query proteins used were known, experimentally validated plastic degrading enzymes found via the search of literature.

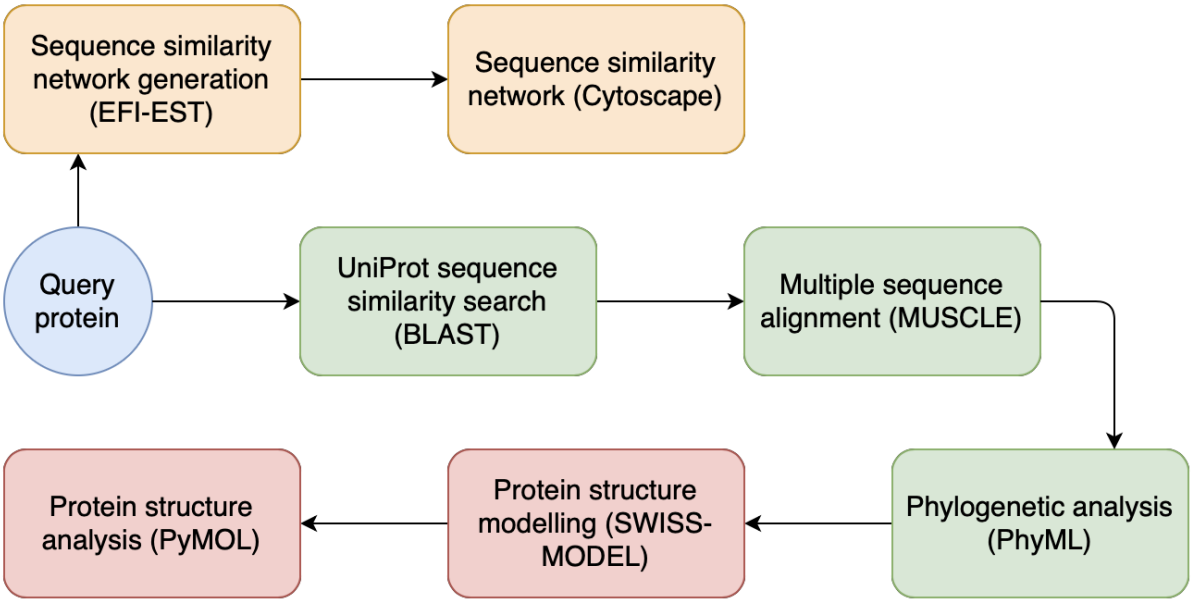

Figure 1: The bioinformatic pipeline developed and used in this study for identifying novel plastic degrading enzymes.

Sequence similarity analysis pipeline

The sequence similarity analysis pipeline was automated using the coding language Python. Python, along with Python modules such as BioPython (Cock, et al., 2009), was used to link the command line tools BLAST, MUSCLE, and PhyML, so that the output of one programme was the input to the subsequent programmes. First, NCBI BLAST (Coordinators, 2018) was used to search the UniProt database by using the query protein accession number, with results limited to 20 hits, a maximum e-value of 2.5, and a minimum length of 70% of the query protein sequence. These parameters ensure that no significant hits will be missed, whilst not overloading the analysis with more irrelevant protein hits. MUSCLE (Edgar, 2004) was then used for a multiple sequence alignment (MSA) of the proteins returned by BLAST, with all default settings being used. Finally, PhyML (Gascuel, et al., 2010) was used to run a phylogenetic analysis from the results of the MSA using all default settings. Python and BioPython were used for parsing of all results and file format conversion between the external programmes, as well as for general functionality and usability of the pipeline, and the Python module Matplotlib was used to draw the phylogenetic tree. Homologous proteins were selected from the phylogenetic tree to undergo a further structural analysis. The accession numbers, lengths, and alignment e-values of these proteins were taken from the BLAST results and recorded. The e-value is a number used by BLAST to quantify the quality of a protein alignment, and so is a good indication of sequence similarity.

Protein structure analysis

The structures of the selected proteins were obtained from the SWISS-MODEL Repository (Bienert, et al., 2017), a database of annotated 3D protein structures, where possible. If the protein structure was not present in the SWISS-MODEL repository, the structure was modelled from its protein sequence using SWISS-MODEL (Waterhouse, et al., 2018). The protein structures were loaded into PyMOL (DeLano, 2002), where they were superimposed using the 'super' command. The RMSD value (including all atoms) calculated by the superposition of the proteins was recorded for each protein pair, which quantifies the difference in atom pair positions on each protein and hence the differences in overall structure.

Generation of the sequence similarity network (SSN)

The SSN was generated using the web-tool EFI-EST (Zallot, et al., 2018) via the input of the query protein sequence. The initial BLAST search was limited to 50 hits and was set to exclude protein fragments, which are smaller fragments of existing, full-length protein sequences. When finalizing the SSN, the edge cut-off alignment score was set to correspond to a protein similarity of 60% (this was calculated using the % similarity vs alignment score graph provided by the EFI-EST web-tool). Therefore, edges will only be drawn between proteins which are at least 60% similar. Cytoscape (Shannon, et al., 2003) was used to view and reformat the generated SSN. Node colours were set to correspond to the class of the proteins and edge thickness was set to correspond to the % identity, so that the edges were thicker between proteins with higher % identity. The nodes were organised into clusters depending on their class and their similarity to each other.

3. Results and Discussion

Through a search of literature, four biochemically characterised and experimentally validated plastic degrading enzymes were identified: PET hydrolase, MHETase, Polyurethanase A, and Polyurethanase B (Danso, Chow, & Streit, 2019). These enzymes are reported to degrade the plastics, PET and polyurethane. The amino acid sequences of these enzymes were used as input to the developed bioinformatic analysis workflow, leading to the identification of six homologous proteins with potential plastic degrading functions. To further investigate if these proteins have the similar plastic degrading functions as their homologous query proteins, the structures of these proteins were obtained and analysed against the structures of the query proteins to which they relate to. Table 1 displays the results of this analysis. As can be seen from Table 1, four of the six proteins identified to have a high likelihood of plastic-degrading functionality, and hence further in vitro investigation into their function is warranted and recommended.

To decide whether further investigation of a candidate protein is worthwhile, three main factors were considered: the evolutionary relationship of the candidate to the query protein, the quality of the pairwise sequence alignment of the candidate to the query protein, and the quality of the structural alignment of the candidate to the query protein. The name of a protein can also be a good indication of its function, and so this factor is also briefly considered during the analysis.

Table 1: List of homologous proteins returned by the analysis pipeline

Query protein (accession number)	Candidate protein (accession number)	Host microorganism	Sequence alignment e-value	RMSD value (Å)	Plastic degradation likelihood
PET hydrolase (A0A0K8P6T7)	Lipase 1 (P19833)	*Moraxella sp.*	3.58E-66	2.731	High
	Non-heme chloroperoxidase (P25026)	*Burkholderia pyrrocinia* *Pseudomonas pyrrocinia*	0.104	19.281	Low
MHETase (A0A0K8P8E7)	Uncharacterised esterase XCC2094 (Q8UK62)	*Xanthomans capastries pv. campestries*	1.28E-56	4.463	High
	Uncharacterised esterase ATU5261 (Q8UK62)	*Agrobacterium fabrum* *Agrobacterium tumefaciens*	3.99E-47	5.542	High
Polyurethanase A (Q4KBS6)	Lipase (P41773)	*Pseudomonas fluorescens*	0	1.411	High
Polyurethanase B (Q4KBS3)	Lipase (P41773)	*Pseudomonas fluorescens*	7.29E-115	3.004	High
	Nodulation protein O (P15728)	*Rhizobium leguminosarum bv. viciae*	0.000634	16.928	Low

Evolutionary relationship

The evolutionary relationship was determined from the phylogenetic trees produced by the pipeline, and candidate proteins were only selected for further analysis if they were deemed homologous to the query protein (i.e., they share a common ancestor). Proteins were identified as homologous from the phylogenetic tree analysis if they branched/ split off from the same single node as the query protein. This shows that the query and candidate proteins are closely related to a single common ancestor and are therefore likely to share similar properties to each other. All six homologous proteins are displayed in Table 1, next to the query protein to which they relate to: PET hydrolase was found to be homologous to Lipase 1 and Non-heme chloroperoxidase, MHETase was found to be homologous to Uncharacterized esterases XCC2094 and ATU5261, Polyurethanase A was found to be homologous to Lipase, and Polyurethanase B was found to be homologous to Lipase and Nodulation Protein O.

Quality of pairwise sequence alignment

The pairwise sequence alignment was performed in the pipeline using BLAST. A BLAST pairwise sequence alignment is quantified by the e-value, which represents how many hits of the same quality can be expected to be returned from the database by chance. Therefore, the

lower the e-value, the more significant is the alignment and the corresponding hit. E-values below 10 are generally considered to be biologically significant, and e-values in the range of 0-0.001 are considered to be very high-quality alignments (Wheeler & Bhagwat, 2007).

Five of the seven candidate proteins fall in the 0-0.001 range. Lipase 1, Lipase, Uncharacterized esterase XCC2094, and Uncharacterized esterase ATU5261 all have e-values well below the 0.001 mark. This shows that the matches are high quality and are biologically significant, as it is extremely unlikely that they have been matched by chance. Nodulation protein O matched with Polyurethanase B at an e-value of 0.000634, which is also below the 0.001 mark. It can therefore also be considered as a high-quality match, however not at the same certainty as the previous 4 proteins. Non-heme chloroperoxidase matched with PET hydrolase with an e-value of 0.104, and so whilst it is still biologically significant, it cannot be considered a high-quality match.

Quality of structural alignment

The structure of each homologous protein was analysed by the superposition of its three-dimensional (3D) structure against the query protein's structure. The quality of a structural alignment can generally be observed by eye; however, RMSD values were used in order to quantify the alignment quality, which is especially important in cases where structural differences were not so clear. RMSD is the root-mean-square deviation of pairs of atoms between two 3D protein structures (Carugo & Pongor, 2001). In general, if the RMSD value between two protein structures is less than 3 Å, the alignment can be considered high quality and the structures can be considered very similar. An RMSD of 6 Å or higher suggests large differences between the structures, and hence, they cannot be considered similar. The length of the proteins must also be considered because the shorter the proteins are, the more significant a higher RMSD value becomes (Reva, et al., 1998).

The highest quality alignment was obtained between the structures of Lipase to Polyurethanase A at 1.411 Å (Table 1). From both visual observations and the RMSD value, it is clear that the structures of these two proteins are very similar. The structural alignments of Lipase to Polyurethanase B, and Lipase 1 to PET hydrolase returned RMSD values of 3.004 Å and 2.731 Å, respectively (Table 1), which along with visually strong alignments confirm that the structures are very similar. The alignments of MHETase to the uncharacterized esterases XCC2094 and ATU5261 returned RMSD values of 4.463 Å and 5.542 Å, respectively. These values are a little higher than the 3 Å mark, however these proteins are very long at 576 and 553 amino acids, respectively. These proteins produced visually strong alignments, and taking their length into consideration, it is clear that these proteins have a significant structural similarity to MHETase. The alignments of Non-heme chloroperoxidase to PET hydrolase, and Nodulation protein O to Polyurethanase B returned RMSD values of 19.281 Å and 16.928 Å, respectively. These values are well beyond 3 Å mark, and along with clear visual differences, it is certain that the two sets of proteins are not structurally similar to each other. Since protein function is directly dependent on the 3D structure, a similar structure indicates a similar functionality between the compared proteins.

Protein name

Considering the three factors, it is clear that Non-heme chloroperoxidase and Nodulation protein O are very unlikely to have plastic degrading functionality, and so it is not feasible to take them further for an in vitro investigation. Although deemed as homologous to their respective query proteins, these results are unsurprising. The name Non-heme

chloroperoxidase suggests that the function of the protein involves the chlorination of organic molecules, which has very little to do with the hydrolysis of polymers. Nodulation suggests that Nodulation protein O is involved in the nodulation process occurring in plant roots, which again has very little to do with the hydrolysis of polymers. It is likely that these proteins were identified as similar by the BLAST search due to sharing short sequence segments of high similarity; however, their overall sequence is still significantly different. BLAST results such as these can occur purely by chance and shows why further analysis is required.

The promising results for Lipase, Lipase 1, and the uncharacterized esterases XCC2094 and ATU5261 are also expected. Lipases and esterases are part of a class of enzymes known as hydrolases, which catalyse the hydrolysis (splitting) of large molecules into smaller ones. This is a good sign as plastic degradation is essentially the breakdown of long polymer chains into monomers. For example, esterases are responsible for the splitting of ester bonds in polyester polymers. Examples of polyester plastics are PET and polybutylene terephthalate (PBT). Although lipases are generally thought to break down fat molecules, there is experimental evidence suggesting a role in the degradation of plastics such as PET (Danso, Chow, & Streit, 2019).

Case study for homologous proteins of PET hydrolase

A case study for identifying the candidate proteins homologous to the PET hydrolase enzyme is described in this section to explain the full analysis process carried out in this study for each experimentally characterised query enzyme.

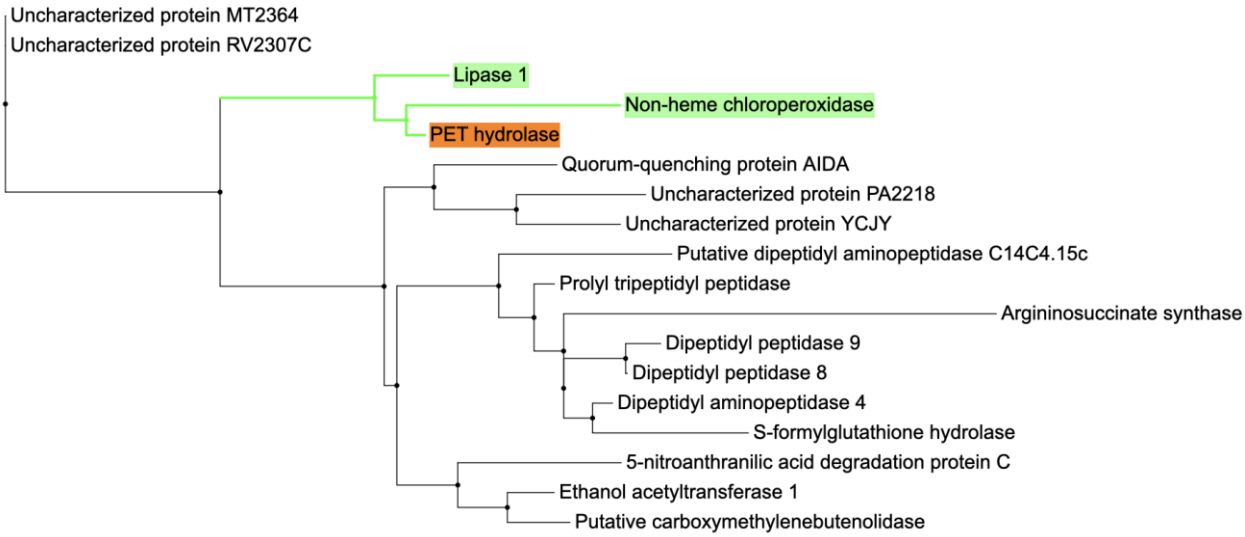

Figure 2: A phylogenetic tree of PET hydrolase

Fig. 2 displays the phylogenetic tree of the query protein PET hydrolase (highlighted in orange) produced by the automated sequence similarity pipeline. PET hydrolase and the green nodes, Non-heme chloroperoxidase and Lipase 1 have all split off from the same original branch and so, it is clear that they share the same common ancestor. The evolutionary distance (signified by the branch length) is very small between the three proteins. Thus, it is likely that they still share a similar original structure and function as the common ancestor because there has not been a large amount of time spent for them to change significantly and acquire diverse functions through evolution. The BLAST results for these two proteins were analysed in more detail to find the e-value of the matches. As previously discussed, Lipase 1 was found to be a high-quality alignment, whereas Non-heme chloroperoxidase was found to be a lower quality match. The e-values provide more evidence as to whether a protein may have plastic degrading

functionality; however, it is not conclusive on its own and so, both proteins were taken forward to the final stage of the analysis.

Fig 3 displays the sequence similarity network (SSN) created for PET hydrolase. It displays the relationships between protein families related to PET hydrolase. SSNs allow large-scale visualization and analysis of sequence–function space in the context of entire protein families. SSNs are used to visualize and analyse the evolutionary and functional relationships of families, or clusters, or proteins. An SSN links together all the results of a BLAST search of the query protein. Single proteins are represented by nodes and the edges (lines) joining them represent their relatedness based on their BLAST alignment score (e-value) (Gerlt, 2016).

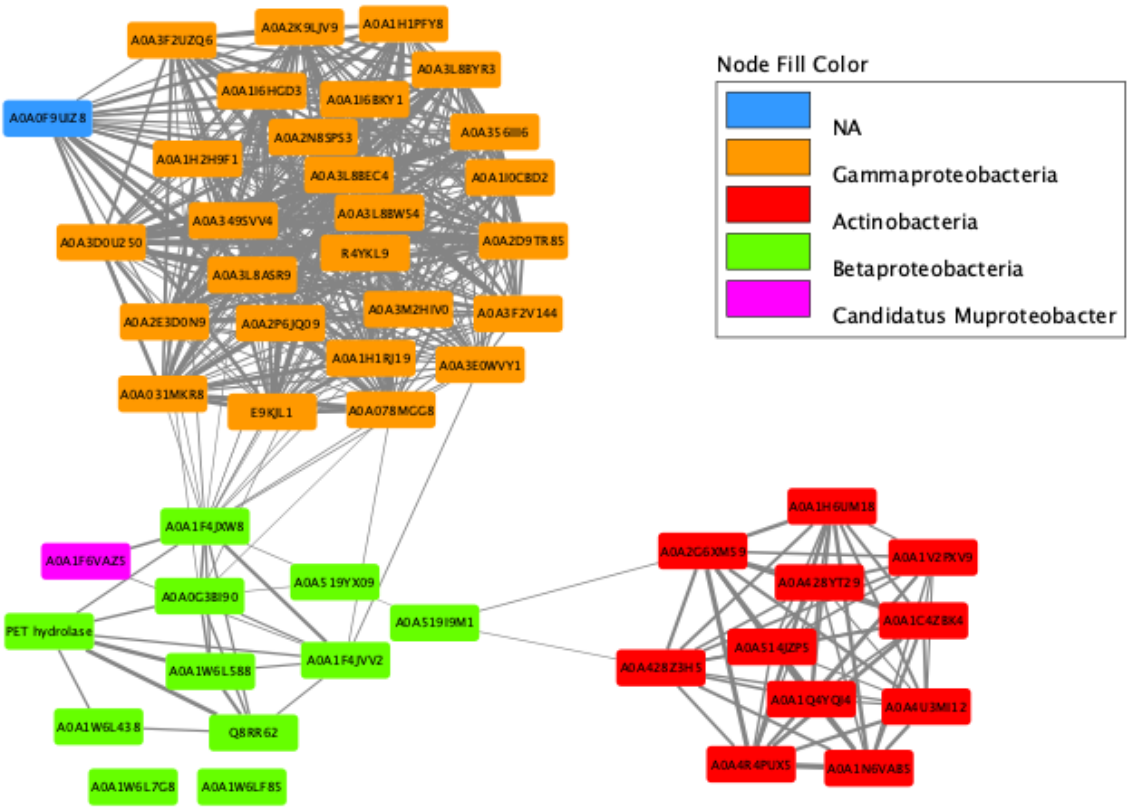

Figure 3: The SSN of PET hydrolase.

In Fig. 3, an edge value cut-off was used, so that an edge was only drawn between two proteins if they passed a minimum similarity value of 60%. This has grouped families of proteins together in clusters of high similarity, and therefore high potential of having the same/ similar functions. The edge thickness has been varied depending on the similarity percentage between the proteins; so, the thicker the edges are, the more similar are the proteins. Thus, it is clear that PET hydrolase is most similar to the proteins with accession numbers A0A1W6L588 and Q8RR62 (each 82% similar); these proteins are a DLH domain-containing protein from the organism Rhizobacter gummiphilus and a PBS (A) depolymerase from the organism Acidovorax delafieldii, respectively. The homologous relationship of proteins with at least 60% similarity is generally considered to be strong (Pearson & Sierk, 2005), and so there is a high potential of similar functionality of all proteins belonging to the cluster.

The nodes in this SSN have been coloured depending on the class of the microorganism that produces them. The largest class is Gammaproteobacteria, representing 26 of the 50 total nodes,

with 11 coming from Betaproteobacteria, and 11 coming from Actinobacteria. Whilst Gammaproteobacteria produce the majority of proteins similar to PET hydrolase, the proteins most similar to PET hydrolase are produced by the Betaproteobacteria class. Therefore, further research into this class of microorganism would be most beneficial, as it is most likely to result in the discovery of a novel plastic degrading microorganism or protein. Sequence similarity networks are not as rigorous as phylogenetic trees, and so the SSN was only created in order to show a larger, more general protein family relationship for each query protein.

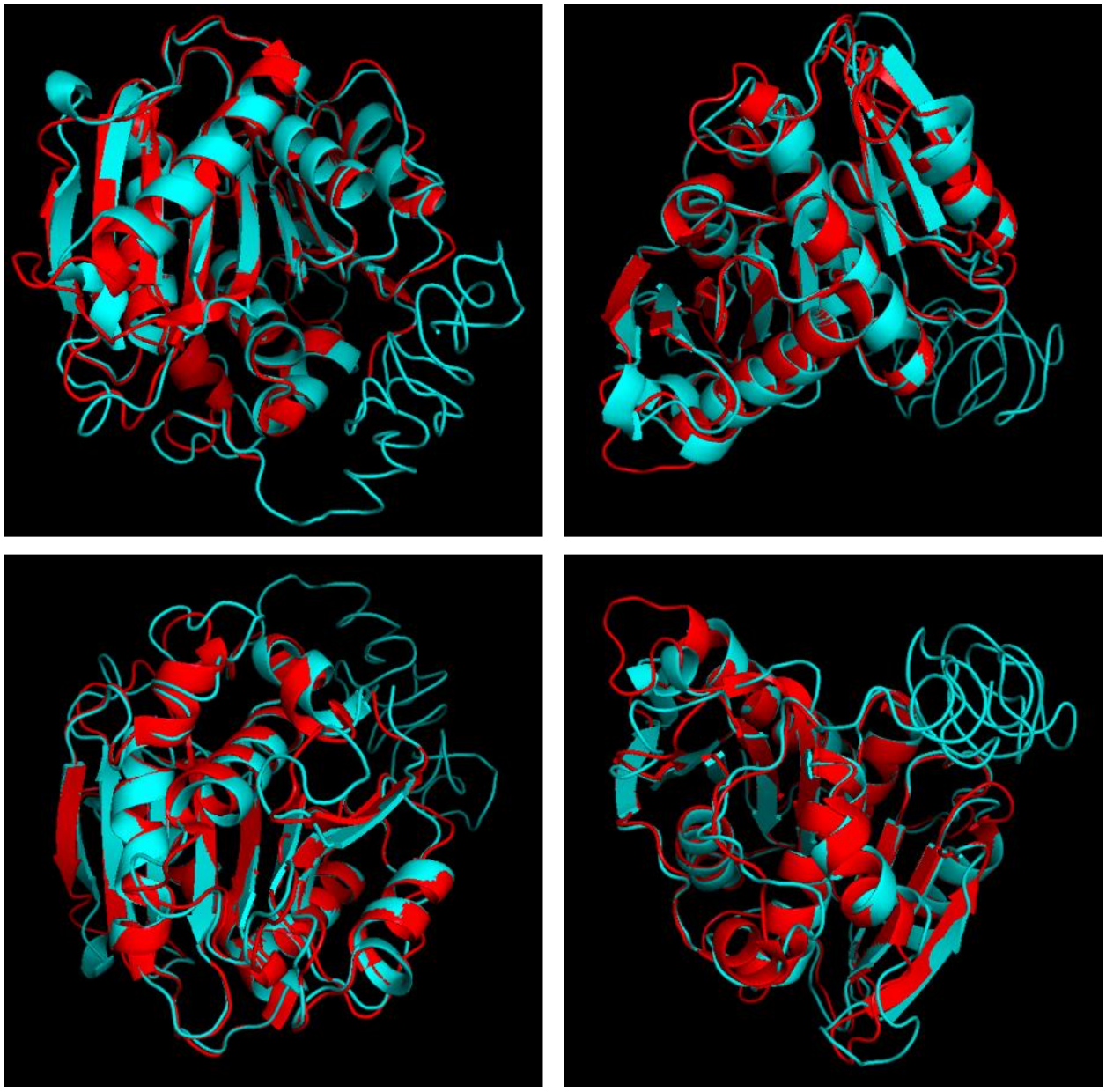

Figure 4: Superposition of 3D structures of PET hydrolase (red) and Lipase 1 (blue)

Fig 4 shows the superposition of the 3D structures of PET hydrolase (red) and Lipase 1 (blue), each rotated 90º around the x-axis and displayed in ribbon diagrams. Whilst analysing protein structures, the most important features to look at the are arrows and spirals, representing beta-pleated sheets and alpha-helices, respectively, as this information constitutes the most important part of the protein structure. A visual observation of the alignments shows that the structure of Lipase 1 is very similar to that of PET hydrolase, with arrows and spirals of very similar shapes are found in very similar positions. The RMSD value for this alignment

(calculated by Pymol) was 2.731 Å, confirming that the proteins are structurally and hence, possibly functionally very similar.

Fig 5 shows the superposition of the 3D structures of PET hydrolase (red) and Non-heme chloroperoxidase (green). The 3D structure of Non-heme chloroperoxidase (NHC) consists of three chains, whereas PET hydrolase consists of only 1. Chains A and C of NHC are largely irrelevant to PET hydrolase, and so were hidden to make the alignment clearer. Even after this modification, a good alignment was not achieved, as there are clear differences in the positions and shapes of the arrows and spirals. The RMSD value for this alignment was 19.281 Å, confirming that these two proteins do not have a similar structure and hence, are highly unlikely to have similar functions.

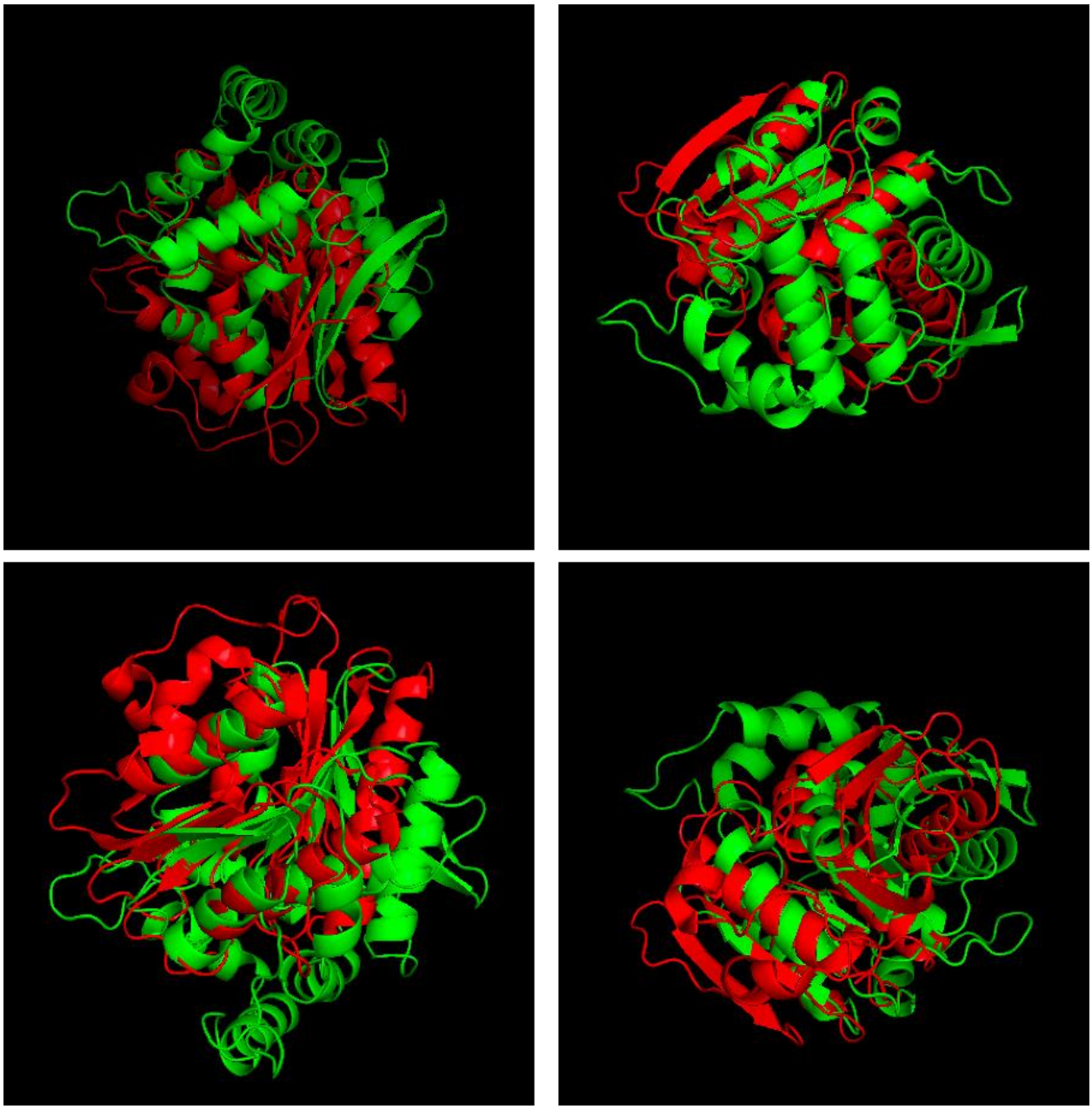

Figure 5: Superposition of PET hydrolase (red) and Non-heme chloroperoxidase (green).

Thus, based on the aforementioned analyses performed using the three criteria, homology, quality of pairwise sequence alignment, and quality of structural alignment, it can be concluded

that Lipase 1 is a strong candidate to be experimentally tested for plastic-degradation activities, whereas Non-heme chloroperoxidase is not

4. Conclusions

The developed bioinformatics pipeline has successfully identified six proteins which are homologous to known plastic degrading enzymes. After undergoing further rigorous and systematic analysis, these six proteins were narrowed down to four candidate proteins with the high potential of having a plastic degrading functionality. Although a rigorous and thorough bioinformatics analysis has identified these potential plastic degrading enzymes, their functionality must be tested experimentally by using appropriate in vitro enzyme assay experiments to be confirmed about their predicted functions. This is because bioinformatics analyses are only computational analyses that can predict the function of a protein sequence with a high degree of certainty but cannot confirm the predicted functions, as the analysis methods primarily rely on the sequence homology information provided by the experimentally characterised proteins. However, such analyses are extremely valuable because they can significantly expedite the discovery of novel plastic degrading enzymes with better functional and catalytic characteristics by narrowing down the potential candidates from hundreds and thousands of amino acid sequences present in the protein databases. An alternative use of the developed pipeline is to apply it directly to certain metagenomes, rather than searching entire protein databases. Most plastic-degrading microorganisms and proteins were identified by the traditional screening of environmental samples such as soil samples taken from outside of plastic recycling plants. Through the development of metagenome sequencing, it is now possible to use bioinformatic tools such as this pipeline to search entire metagenome sequences from environmental samples, rather than general protein databases. It is important that traditional methods of bioprospecting for novel enzymes should continue. This pipeline is capable of quickly, cheaply, and efficiently producing a list of potential candidate proteins with high similarity to a query protein which can be taken for further experimental and functional analysis. However, the results of the pipeline are only as good as the quality of the initial query protein. Therefore, high-quality, experimentally validated plastic degrading enzymes are required by the pipeline in order for it to produce high-quality results.

5. References

Artuso, A. (2002). Bioprospecting, Benefit Sharing, and Biotechnological Capacity Building. World Development, 30(8), 1355-1368.

Bienert, S. et al. (2017). The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Research, Volume 45, 313-319.

Carugo, O. & Pongor, S. (2001). A normalized root-mean-square distance for comparing protein three-dimensional structures. Protein Science, 10(7), 1470-1473.

Cock, P. J. A. et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422-1423.

Coordinators, N. R., 2018. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Research, 46(1).

Danso, D., Chow, J. & Streit, W. R. (2019). Plastics: environmental and biotechnological perspectives on microbial degradation. Applied and Environmental Microbiology, 85(19).

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography, Volume 40, 82-92.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792-1797.

Gascuel, O. et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307-321.

Gerlt, J. A., 2016. Tools and strategies for discovering novel enzymes and metabolic pathways. Perspectives in Science, Volume 9, 24-32.

Geyer, R., Jambeck, J. R. & Law, K. L. (2017). Production, use, and fate of all plastics ever made. Science Advances, 3(7).

Ghosh, S. K., Pal, S. & Ray, S. (2013). Study of microbes having potentiality for biodegradation of plastics. Environmental Science and Pollution Research, Volume 20, 4339-4355.

Gourmelon, G. (2015). Global Plastic Production Rises, Recycling Lags. Worldwatch Institute.

Greenbaum, D., Gerstein, M. & Luscombe, N. M., 2001. What is Bioinformatics? A proposed definition and overview of the field. Methods of Information in Medicine, 40(4), 346-358.

Han, X. et al. (2017). Structural insight into catalytic mechanism of PET hydrolase. Nature Communications, Volume 8.

Pearson, W. R. & Sierk, M. L. (2005). The limits of protein sequence comparison? Current Opinion in Structural Biology, 15(3), 254-260.

Reva, B. A., Finkelstein, A. V. & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? Folding and Design, 3(2), 141-147.

Seltenrich, N. (2016). New link in the food chain? Marine plastic pollution and seafood safety. Environmental Health Perspectives, 124(7).

Shannon, P. et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498-2504.

Singh, H., Srinivasan, S. & Kamble, A. (2019). In-silico bioprospecting: finding better enzymes. Molecular Biotechnology, Volume 61, 53-59.

Waterhouse, A. et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, Volume 46, 296-303.

Wheeler, D. & Bhagwat, M. (2007). BLAST QuickStart. In: N. Bergman, ed. Comparative Genomics. Totowa: Humana Press.

Zallot, R., Oberg, N. & Gerlt, J. A. (2018). "Democratized" genomic enzymology web tools for functional assignment. Current Opinion in Chemical Biology, 47(1), 77-85.