FuturEnzyme

Technologies of the FUTURe for low-cost ENZYMEs for environment-friendly products

WP 2
12 months meeting

Project funded by the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No [101000327]

Original image: Illustration by Ainhoa Quirós

Work package number [9]	WP2	Lead beneficiary [10]	2 - BSC
Work package title	Machine learning enzyme bio-prospecting integrated into an industrial context		
Start month	1	End month	48

Participation per Partner

Partner number and short name	WP2 effort
1 - CSIC	3.00
2 - BSC	32.00
3 - BANGOR	2.00
4 - UHAM	6.00
5 - UDUS	1.00
13 - SCHOELLER	4.00
14 - HENKEL	2.00
15 - EVO	1.00
Total	51.00

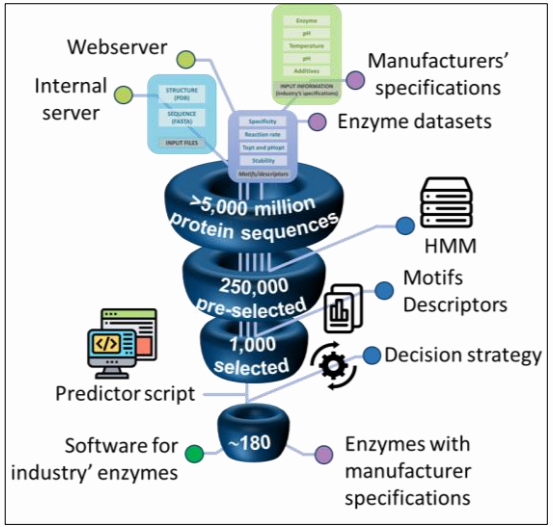
Objectives

Task 2.1. Compile the on-demand manufacturers' needs and specifications M1-M6
Task Lead Partner – CSIC
Participants: SCHOELLER, HENKEL, EVO

Task 2.2 Pre-selecting candidate sequences through extensive homology search M1-M48
Task Lead Partner – UHAM
Participants: BSC, CSIC, UDUS, BANGOR

Task 2.3 Motif buildup for massive and smart search of enzymes fitting manufacturers' needs M1-M42
Task Lead Partner – BSC
Participants: CSIC

Task 2.4 Iterative and decision-making hierarchical procedure for speed up enzyme discovery M3-M48
Lead partner – BSC
Participants: CSIC, UDUS, BANGOR, UHAM



*HMM: Hidden Markov Model

Summary

- Real-life substrates have been provided by HENKEL, EVONIK and SCHOELLER.
- The enzymes to focus on and the process and products specificiations have been provided by HENKEL, EVONIK and SCHOELLER.
- An exhaustive bibliographic (scientific and patent) search has been completed.
- A draft HMM and computational pipeline was implemented for sequence-based enzyme discovery.
- First enzyme candidates have been selected and datasets obtained to be further integrated into the predictive tool.
- Selected enzymes have been tested through both computational (BSC) and experimental (CSIC) methodologies.

Task 2.1 Compile the on-demand manufacturers' needs and specificiations M1-M6

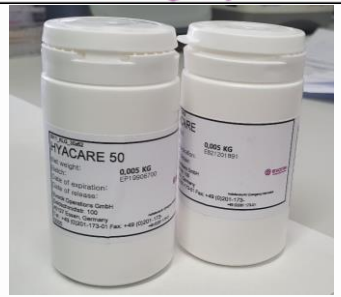
- **HENKEL, EVONIK and SCHOELLER** have already supplied the partners with the real samples to work with, namely, real-life hyaluronic acid (added to real-file cosmetics), liquid detergents and textiles/fabrics; this is key because we started working with real samples very early on.
- **HENKEL, EVONIK and SCHOELLER** have defined and shared with partners the needs and specifications, including enzymes of interest, substrates to work with, conditions at which enzymes should work, etc.
- **CSIC** has performed an exhaustive patent and bibliographic search for the:
  - Production of hyaluronic acid for cosmetics
  - Use of enzymes in detergent compositions
  - Use of enzymes in textile industry

Materials from industrial partners

Henkel

CSIC · Prifysgol Bangor University · IST-ID Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento · hhu Heinrich Heine Universität Düsseldorf · UH

EVONIK Leading Beyond Chemistry



CSIC · Prifysgol Bangor University · IST-ID Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento · hhu Heinrich Heine Universität Düsseldorf · UH · CNR IRBIM

schoeller® Switzerland

CSIC · Prifysgol Bangor University · IST-ID Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento · hhu Heinrich Heine Universität Düsseldorf · UH · BioChemSolutions

FUTUR ENZYME

Materials from industrial partners

Material	Sent by Henkel	Received by partners
1x Canister detergent_A gel without enzymes, 2,5 % gap to be filled with enzymes/water	August, 2021	August-september 2021
2x Bottles detergent_A gel including all enzymes, no gap	August, 2021	August-september 2021

Material	Received by CSIC	Sent to partners
Hyaluronic acid, Hyacare	5g, august 2021	0.8 g, 08.09.2021
Hyaluronic acid, Hyacare 50	5g, august 2021	0.8 g, 08.09.2021

Materials from industrial partners

schoeller® Switzerland

Sample	Sample quality	Finished goods artikel Nr.	Quality ROH=Raw VORB= Pre-treated	Components / Weight	Pre-treatment steps	Received by CSIC	Sent to partners
1	1-a 1-b	61488	61488Z ROH 61488Z VORB	92% PA, 8% EL 180g/m^2	Chemical cleaning and washing	2m piece, 27.08.2021 2m piece, 27.08.2021	15 cm piece, 08.09.2021 15 cm piece, 08.09.2021
2	2-a 2-b	61988	61988F1 ROH 61988F1 VORB	92% PA, 8% EL 280g/m^2	Chemical cleaning and washing	2m piece, 27.08.2021 2m piece, 27.08.2021	15 cm piece, 08.09.2021 15 cm piece, 08.09.2021
3	3-a 3-b	67007	67007 ROH 67007 VORB	88% PA,12% EL 135g/m^2	Washing	2m piece, 27.08.2021 2m piece, 27.08.2021	15 cm piece, 08.09.2021 15 cm piece, 08.09.2021
4	4-a 4-b	3X58	2X34G ROH 3X58 VORB	100% PES 100g/m^2	Alkaline boiling	2m piece, 27.08.2021 2m piece, 27.08.2021	15 cm piece, 08.09.2021 15 cm piece, 08.09.2021
5	5-a 5-b	66299	5237/00 ROH 5237/00 VORB	92% CO, 8% EL 240g/m^2	Desizing, washing out, bleaching, washing	2m piece, 14.09.2021 2m piece, 14.09.2021	- -
6	6-a 6-b	E03130	E03130 ROH E03130 VORB	80%PA6 , 20%EL	Chemical cleaning	2m piece, 27.08.2021 2m piece, 27.08.2021	15 cm piece, 08.09.2021 15 cm piece, 08.09.2021

Manufacturers' needs and specifications

	LIQUID/DOSE CAP DETERGENT
Products to be made	Laundry & Home Care (LHC)'s leading premium liquid detergent and/or unit dose caps products with enzymes.
Request	Enzymes for removing fatty oil stains.
Innovation	Innovation will come because the use of enzymes will improve removal of stubborn stains at low temperatures while decreasing chemical usage. A central point is to lower the amount of surfactant in the detergent formulation as much as possible by adding enzymes.
Priority enzymes to be targeted	Among all enzyme classes discussed in the proposal, priority target will be enzymes for removing specific fatty oil stains, that will include: • True lipases (EC 3.1.1.3) • Esterases (EC 3.1.1.1) • Cutinases (EC 3.1.1.74) and related fatty-oil degrading hydrolases
Non-priority enzymes to be targeted	Aside the priority classes, other enzyme classes relevant to detergents are also considered, that include: • Proteases/peptidases, suitable for protein-based stain removal (i.e. blood, milk, grass) at low temperature, e.g., type family S08 (alcalase), type papain (EC 3.4.22.2), type savinase-esperase (EC 3.4.21.14), type subtilisin-alcalase (EC 3.4.21.62), type trypsin and protease inhibitor. • Amylase (EC 3.2.1.1) and other glycoside hydrolases • Peroxidases and related enzymes (EC 1.1.3.-, EC 1.11.1.- or EC 1.10.3.2), very specific in the potential use case (to be discussed in more detail in case they become relevant).
Specifications that enzymes should meet	The enzymes should be active and stable under conditions relevant to the wash cycle and to storage. Below, the specifications are summarized: • The enzymes should be stable for at least 2 to 3 months at 30˚C in the liquid detergent formulation. Note: This stability refers to the stability of the enzymes in the detergent formulation. • The enzymes should be effective and stable at a washing temperature between 20 and 40˚C and at pH 7.0-8.5, at least during an operation time of a common wash cycle (120 min). Note: This stability and activity refer to that of the enzymes in a wash liquor mimicking the detergent-water mixture in a washing machine; this wash liquor consists in about 50 g liquid detergent per 20 liter of water.

Manufacturers' needs and specifications

	COSMETIC FORMULATIONS
Products to be made	EVO's leading cosmetics integrating ingredients produced by enzymes
Request	Enzymes for degrading hyaluronic acid to products of defined size to be integrated into cosmetics
Innovation	Hyaluronic acid is widely used for cosmetic applications where it mainly acts as natural moisturizer and as anti-aging active. Specially, the biological anti-aging activity is limited by the enormous molecular size of hyaluronic acid that can reach up to 2,000 kDa and interferes with its penetration into the skin. Fragmentation of large hyaluronic acid polymers can markedly improve its penetration abilities. Nevertheless, pro-inflammatory responses have been reported for very small hyaluronic acid fragments (5-15 kDa) which are recognized by special receptors of the immune-system; therefore, size matters, and has to be above or below a specific threshold. In this case it should be below 5 kDA, prefered 1-2 kDa, so that the new molecule will better penetrate into the skin, making the cosmetic more effective, and the production process more sustainable.
Priority enzymes to be targeted	Priority targets will be enzymes degrading hyaluronic acid: • Heparanase (EC 3.2.1.166) • Hyaluronate lyase (cd01083 - EC 4.2.2.1) • Hyaluronidase (EC 3.2.1.35, EC3.2.1.36, pfam03662, pfam01630).
Specifications that enzymes should meet	Hyaluronic acid is actually produced by fermentation of Bacillus subtilis (non-pathogenic) and an environmentally friendly, solvent free recovery process. Existing technologies like thermal degradation are unsuitable for achieving the targeted molecular weight and polydispersity. We can envision two options for producing small hyaluronic acid with 1-2 kDA molecular weight: • An enzyme that can be added during the fermentation to prevent additional process steps to make the small hyaluronic acid. The possibility that the new enzyme can be integrated into the Bacillus subtilis that produce the hyaluronic acid may be also evaluated. • An enzyme that can be added after fermentation in the current solvent free process, which should improve the LCA. The fermentation conditions and the thermal denaturation conditions cannot be provided by Evonik. CSIC will start a large bibliographic and patent search to find most common conditions for such processes.

Manufacturers' needs and specifications

Priority	1	2	3
Possible applications/scope	Cleaning/pretreatment of synthetic fibres	Chalk marks	Replacement of the bleaching processes
Substrate	Polyester fibres (PES) / polyamide fibres (PA) containing elastane (polyether-polyurea copolymer)	Cotton (CO), polyester fibres (PES), polyamide fibres (PA)	Cotton (CO)
Desired effect/change	Fully removal of spinning additives (see details below*)	Solving the problem of writing on the finished textile	Decoloring of natural fibres and cotton hasks
State of the art	Solvent cleaning or insufficient washing, which creates problems in the subsequent processing	F-based marks for hydrophobic materials	Chemical bleaching (Chlorid or Peroxid)
Impact to Schoeller	Huge	Huge	Low
Impact to other textile producers	Huge	Huge	High
Priority High-Med-Low	High	High	High to Low
Lab application possible?	Yes	Yes	Yes
Test method	Analytical extraction	Physical, observational	Chemical test tensile, degree of whiteness and DP (degree of average polimerization)
Effect/result proof	Reducing dyeing, finishing problems and second quality products	With less chemicals, similar effects	Achieving maximum whiteness and reducing dye stuff
How to quantify	1. Avoiding solvents 2. Bulk trial dyeing comparison	Calculating the sparing amounts of chalkmarks	Saving on chemicals
Reducing reworks and off-quality	Yes	Yes, sparing quite a lot of money through the whole textile processing chain	To some extent
Comments	-	-	-
Priority enzymes to be targeted	Lipases, cutinases, poliuretanases, amidases	Lipases, esterases, poliuretanases, amidases, cellulases	Bleaching enzymes (oxidoreductases)
Conditions for process/product	See details below*		See details below*
Screening method for enzymes	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.

Manufacturers' needs and specifications

Priority	4	5	6
Possible applications/scope	Surface functionalization/modification	Improved hydrophilicity	Improved hydrophobicity
Substrate	Polyester fibres (PES), modification and plasma treatment	Polyester fibres (PES) / polyamide fibres (PA) containing elastane (polyether-polyurea copolymer)	Polyester fibres (PES) / polyamide fibres (PA) containing elastane (polyether-polyurea copolymer)
Desired effect/change	Generating functional groups/layers	Higher absorbency (by pre-processing) and better humidity management (finishing)	Better water /soil repellency with less chemicals, removal of residual substrates
State of the art	Heating (natriumhydroxide) and atmospheric plasma	Solvent cleaning	Higher amounts of chemicals
Impact to Schoeller	Medium	Huge	Huge
Impact to other textile producers	Medium	Huge	Huge
Priority High-Med-Low	Low	High	High
Lab application possible?	Yes	Yes	Yes
Test method	Physical testing (permanent treatments)	Physical testing- absorbency	Physical testing
Effect/result proof	Bonding strenghts and higher washability	Improved dyeing process, moisture management	Improved water and soil repellency with less chemicals
How to quantify	Managable	Hydrophil tests for uniform hydrophilicity	Reduction of used chemicals
Reducing reworks and off-quality	No	Yes	Yes
Comments	-	-	-
Priority enzymes to be targeted	Lipases, cutinases, esterases	Lipases, cutinases, poliuretanases, amidases, proteases (subtilisin, bromelain type)	Lipases, cutinases, poliuretanases, amidases, proteases (papain)
Conditions for process/product	See details below*	See details below*	See details below*
Screening method for enzymes	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2

Manufacturers' needs and specifications

Priority	7	8	9
Possible applications/scope	Improved fixation of PA dyeing (amino multiplier?)	Fewer water consumption in the dyeing process	Higher effectiveness of existing enzyme treatments on natural and synthetic fibres
Substrate	Polyamide fibres (PA)	Polyester fibres (PES), cotton (CO)	Cellulosic fibre
Desired effect/change	Better fixation with fewer color consumption	Still large amounts of water is consumed in dyeing process; yet to be defined whether reduction is possible by enzyme treatment	Desizing, bleaching, bio-polishing
State of the art	Chemicals treatment	Extensive rinsing process a high water and time consuming process	Chemicals
Impact to Schoeller	High	High, technical feasibility with enzymes hard to realise	Too Low
Impact to other textile producers	High	High	Relevant
Priority High-Med-Low	Medium	High - see comments	Low
Lab application possible?	Yes	Yes	-
Test method	Fastness, dye consumption tests	-	-
Effect/result proof	Less dye materials and improved fastness	-	-
How to quantify	Dye stuff consumption and fastness	Water energy saving	Quite time-consuming compared to the existing processes
Reducing reworks and off-quality	Yes, especially reducing chemicals	-	-
Comments	-	-	Schoeller is using amylases for desizing of cellulosic frequently
Priority enzymes to be targeted	Amidases, proteases (alcalase, subtilisin), lipases, esterases	Lipases, cutinases, cellulases	Cellulases and amylases
Conditions for process/product	See details below*	See details below*	See details below*
Screening method for enzymes	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.	The methods for screening and characterizing the enzymes need to be adapted by partners, as detailed in Deliverable 3.2.

Manufacturers' needs and specifications

The substrate generally used for bio-processing includes paraffin, mineral oil, silicon oil, acrylic acids, and ester oils, and those chemicals need to be eliminated at the end of the processing procedure by the action of enzymes to avoid extensive water consumption.

Chemistry used for polyamide (PA)/polyethylene terephthalate(PET)/polyester (PES) fibres, would be:

- Thermostable ester oils as lubricants.
- Various fatty alcohol, fatty acid or fatty acid amide derivatives, ethoxylated or ethoxylated / propoxylated as emulsifier / wetting agent / cohesion component.
- Phosphoric acid esters, phosphonic acid derivatives as antistatic agents.
- Small amounts of antioxidants, corrosion protection agents and in some cases in-can preservatives.

Chemistry used for polyurethane (PUE) filaments would be:

- Low-viscosity silicone oils (PDMS) as lubricants.
- Low-viscosity mineral oils as lubricants.
- Magnesium stearate as a release agent.

Regarding texturing preparation, as a rule, 2 preparations are applied.

1. First, spin preparation during the spinning of the partially orientated yarn (POY) filament (layer approx. 0.4 percent by weight): ethylene oxide (EO) / propylene oxide (PO) copolymers as lubricants, fatty alcohol alkoxylates as wetting / spreading agents. Possibly small amounts of fatty acid ethoxylate as wetting / spreading agent or cohesive component. Smallest amounts of phosphoric acid ester as an antistatic agent.
2. During texturing, before winding, a winding oil (application approx. 1.5 - 3 percent by weight): mineral oil as a lubricant, fatty alcohol / fatty acid ethoxylate as an emulsifier.

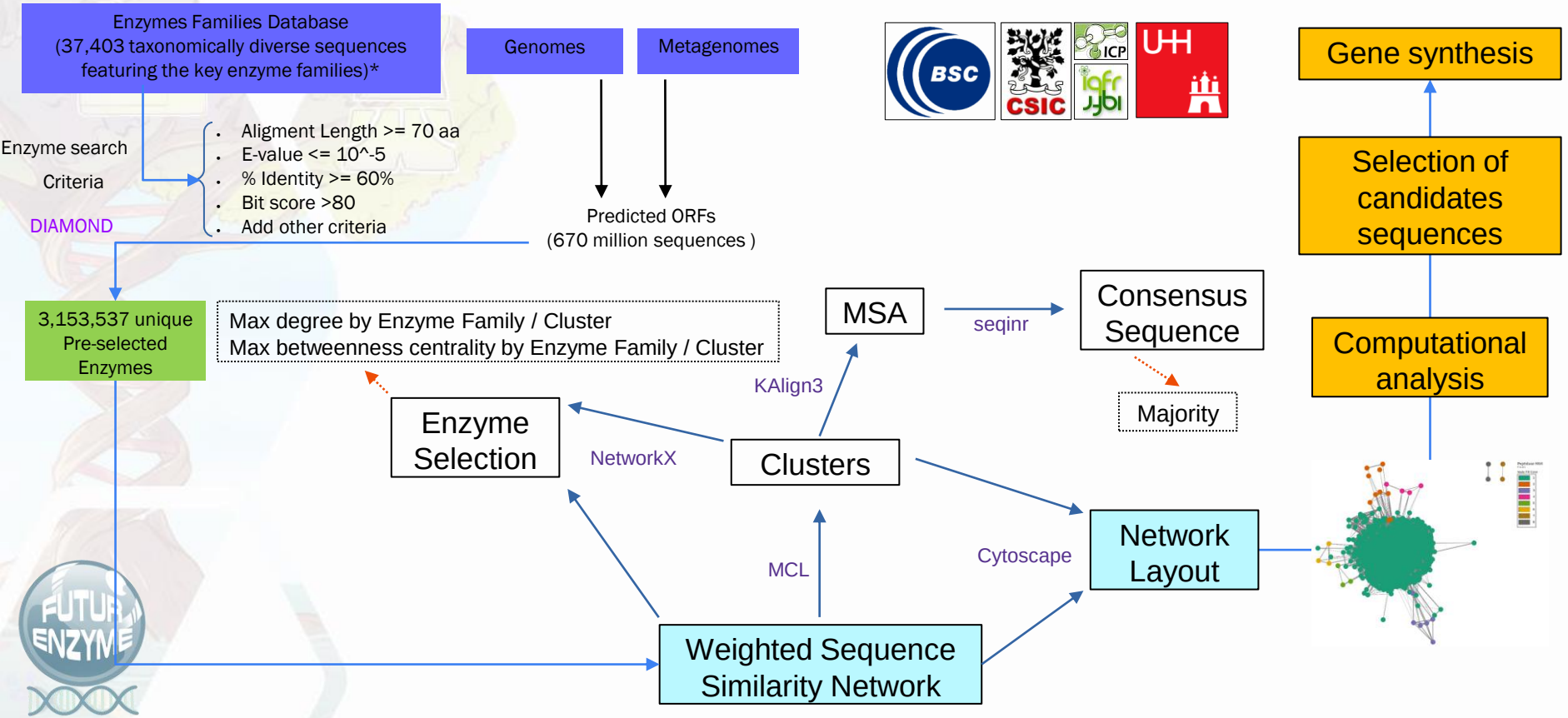
In Europe in particular, there are always discussions in connection with emissions on the stenter caused by spool oil, and mineral oil in particular is held responsible for this. That is why there are also more thermally stable winding oils, but they are correspondingly more expensive and therefore not very common. There the mineral oil gets through replaces thermostable ester oils or carbonic acid esters (Bozetto technology).

Task 2.2 Pre-selecting candidate sequences through extensive homology search M1-M48

- **CSIC** designed and used a reference a manually curated database with 37,403 diverse protein sequences featuring enzyme families relevant to the project to screen a total of 670 million sequences from 13 public and internal metagenomes, and 48 genomes
- A total of 3,153,537 sequences were selected by running Diamond after screening
- Network analysis was performed, and 481 clusters identified.
- One enzyme per cluster was selected and using computational analysis 108 were found to encode full length proteins with catalytic residues and domains.
- Actually, 47 genes have been subjected to gene synthesis.

Note: in next slide the in silico pipeline for enzyme search and selection is detailed

Pipeline for *in silico* search for enzymes



*Enzymes from patents, bibliography and one representative per taxonomic class

Pipeline for *in silico* search for enzymes

List of public and internal sequence repositories and genomes screened

# Public databases*	Details	CDS
CAZyDB.07312020.dmnd`	http://bcb.unl.edu/dbCAN2/download/	1716043
mardb_proteins_V6.dmnd	https://public.sfb.uit.no/MarDB/; BLAST/proteins/mardb_proteins_V6.faa	46739080
marfunV3_proteins.dmnd	https://public.sfb.uit.no/MarFun/; BLAST/proteins/marfunV3_proteins.faa	71374
marref_proteins_V6.dmnd	https://public.sfb.uit.no/MarRef/; BLAST/proteins/marref_proteins_V6.faa	4726614
nr.dmnd	ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz	371327556
uniprot_sprot.dmnd	https://ftp.uniprot.org/pub/databases/uniprot/ current_release/knowledgebase/complete/uniprot_sprot.fasta.gz	564638
uniprot_trembl.dmnd	https://ftp.uniprot.org/pub/databases/uniprot/ current_release/knowledgebase/complete/uniprot_trembl.fasta.gz	214406399
IGC.dmnd	-	9878647
MAGProts.dmnd	-	208832
totalProtsMetaProBone.dmnd	-	10402509
Irish and Mediterranean.dmnd	https://bangoroffice365-my.sharepoint.com/personal/chsa18_bangor_ac_uk/	449245
Human microbiome	https://db.cngb.org/microbiome/genecatalog/genecatalog_human/)	10000000

Pipeline for *in silico* search for enzymes

List of public and internal sequence repositories and genomes screened

# Additional genomes*	Details	CDS
HF571520-HF571521	Halorhabdus tiamatea SARL4B	3023
JFHS00000000.1	Psebau_v14	7839
LGTE00000000.1	ASM126341v1	3097
NC_015151.1	ASM19031v1	2320
NZ_AROI00000000.1 / NZ_NWMT00000000.1	Pseudomonas pelagia CL-AP6	4112
NZ_FOGN01000016	Pseudomonas bauzanensis	3241
NZ_LT629748.1	Pseudomonas litoralis	3717
NZ_NBYK00000000.1	Pseudomonas aestusnigri	3510
NZ_PPSK00000000.1	Pseudomonas oceani	3757
PRJEB12275	Cuniculiplasma divulgatum, C. divulgatum PM4	1816
PRJEB12276	Cuniculiplasma divulgatum (ASM90008351v1) Thermosinus carboxydivorans Nor1, ASM16915v1 (AAWL00000000.1)	2750
ABXP00000000.2	Caldanaerobacter subterraneus subsp. pacificus DSM 12653 (ASM15627v2)	2511
ATYG00000000.1	Carboxydothermus ferrireducens DSM 11255, ASM42756v1	2492
BDJL00000000.1	Carboxydothermus islandicus, ASM195032v1	2480
CP000141.1	Carboxydothermus hydrogenoformans Z-2901, ASM1286v1	2620
CP001463.1	Thermococcus sibiricus MM 739, ASM2254v1	2036
CP002952.1	Thermococcus sp. AM4, ASM15120v2	2222
CP003321.1	Desulfurococcus amylolyticus DSM 16532, ASM23101v3	1421
CP003423.1	Fervidicoccus fontis Kam940, ASM25842v1	1385
CP003531.1	Thermogladius calderae 1633, ASM26449v1	1414
CP003557.1	Melioribacter roseus P3M-2, ASM27914v1	2840
CP006646.1	Thermofilum adornatum, ASM44601v1)	1896
CP007493.1	Thermofilum adornatus 1505, ASM81324v1	1924

Pipeline for *in silico* search for enzymes

List of public and internal sequence repositories and genomes screened

# Additional genomes*	Details	CDS
CP009552.1	Geoglobus acetivorans, ASM78925v1	2218
CP009961.1	Thermofilum uzonense, ASM99380v1	1455
CP013050.1	Thermococcus barophilus, ASM143345v1	2634
CP018099	Caldithrix abyssi DSM 13497, ASM188681v1	4214
GCA_001306115.1	Ornatilinea apprima, ASM130611v1	3347
CP028858.1	Haloarculaceae archaeon HArcel1, ASM305836v1	2532
LJCQ00000000.1	Acidiplasma aeolicum, ASM139969v1	1722
LKBG00000000.1	Acidiplasma aeolicum, ASM140294v1	1696
NC_008260.1	Alcanivorax borkumensis SK2, ASM936v1	2755
CP005996.1, CP006601.1 (plasmid)	Cycloclasticus zancles 78-ME, ASM44259v1	2584
CP008874.1, CP008875.1 (plasmid)	Halanaeroarchaeum sulfurireducens, ASM101111v1	2228
CP011564.1, CP011565.1 (plasmid)	Halanaeroarchaeum sulfurireducens, ASM130565v1	2270
CP016804.1	Halodesulfurarchaeum formicicum, ASM188695v1	2100
CP016070.1	Halodesulfurarchaeum formicicum, ASM176731v1	2023
CP044129.1, CP044130.1 (plasmid)	Halomicrobium sp. LC1Hm, ASM961799v1	3447
CP025066.1	Halalkaliarchaeum desulfuricum, ASM295277v1	3232
CP064789.1, CP064790.1 (plasmid)	Haloarculaceae archaeon HSR-Bgl, ASM1709444v1	3117
CP064791.1, CP064792.1 (plasmid)	Haloarculaceae archaeon HSR-Est, ASM1709446v1	2859
CP064787.1	Haloarculaceae archaeon HSR12-1, ASM1709450v1	3055
CP064788.1	Haloarculaceae archaeon HSR12-2, ASM1709452v1	3024
CP040089.1	DPANN group archaeon LC1Nh, ASM961797v1	1162
CP064786.1	Halobacteriaceae archaeon AArc-S, ASM1709448v1	3120
CP024047.1, CP024045.1 (pla1); CP024046.1 (pla2)	Natrarchaeobaculum sulfurireducens, ASM343082v1	3708
CP027033.1, CP027032.1 (plasmid)	Natrarchaeobaculum sulfurireducens, ASM343080v1	3737
GCF_000023945.1	Halorhabdus utahensis DSM 12940	3048
GCF_000470655.1	Halorhabdus tiamatea SARL4B	3175
TOTAL		670625822

Pipeline for *in silico* search for enzymes

List of public and internal sequence repositories and genomes screened

# Additional genomes*	Details	CDS
CP064786.1	Halobacteriaceae archaeon AArc-S, ASM1709448v1	3120
CP024047.1, CP024045.1 (pla1); CP024046.1 (pla2)	Natrarchaeobaculum sulfurireducens, ASM343082v1	3708
CP027033.1, CP027032.1 (plasmid)	Natrarchaeobaculum sulfurireducens, ASM343080v1	3737
GCF_000023945.1	Halorhabdus utahensis DSM 12940	3048
GCF_000470655.1	Halorhabdus tiamatea SARL4B	3175
TOTAL		670625822

Pipeline for *in silico* search for enzymes

List of selected BLAST-hit candidates per each of the reference enzyme classes

# Class of enzymes	Sequences in the reference fasta	Sequences identified by BLAST
Amidase (FuturEnzyme - textile).fas	1	194
Amylase (COG0366 - FuturEnzyme - detergent).fas	21092	1048575
Amylase (EC3.2.1.1 - FuturEnzyme - detergent).fas	4	679
Lactonase (COG1735 - FuturEnzyme - detergent).fas	1069	119142
Lactonase (EC3.1.1.25 - FuturEnzyme - detergent).fas	24	2682
Cutinases (EC3.1.1.74 - detergent & textile).fas	76	546
Cutinases (pfam01083 - detergent & textile).fas	70	824
Lipase-Esterase (FuturEnzyme - detergent).fas	76	546
PLA, PCL, Impranil DNL hydrolases (detergent & textile).fas	26	3022
Poly(ethylene terephthalate) hydrolases (detergent & textile).fas	38	4615
Polyurethanase (1) (detergent & textile).fas	50	4605
Polyurethanase (2) -lipase class 3 (detergent & textile).fas	370	28415
Polyurethane degrading urease (EC3.5.1.5 - textiles).fas	828	152894
Heparanase (EC 3.2.1.166 - cosmetic).fas	4	386
Hyaluronate lyase (cd01083 - EC4.2.2.1 - cosmetic).fas	355	41852
Hyaluronidase (EC3.2.1.36 - cosmetic).fas	2	95
Hyaluronidase (EC4.2.2.1-cosmetic).fas	292	36725
Hyaluronidase (pfam03662 - cosmetic).fas	65	6701
Hyaluronidases (EC3.2.1.35 - cosmetic).fas	4317	380042
Hyaluronidases (pfam01630 - cosmetic).fas	5	2219
Peptidase type Bromelain (EC3.4.22.32 - textile).fas	2	179
Peptidase type family M04 (detergent & textile).fas	225	32971
Peptidase type family S08 (alcalase - detergent & textile).fas	1116	199971
Peptidase type Papain (EC3.4.22.2 - detergent & textile).fas	41	5459
Peptidase type savinase-esperase (EC3.4.21.14 - detergent & textile).fas	8	1515
Peptidase type subtilisin-alcalase (EC3.4.21.62 - detergent & textile).fas	4703	804058
Trypsin and protease inhibitor (detergent).fas	3	136
Peroxidases (detergent).fas	159	16189
TOTAL	37403	3153537

Pipeline for *in silico* search for enzymes

List of selected BLAST-hit candidates per each of the reference enzyme classes

# Enzyme class	Sequences identified by BLAST	Nr of clusters
Amidase (textile).fas	194	22
Amylase (COG0366 - detergent).fas	1048575	21
Amylase (EC3.2.1.1 - detergent).fas	679	
Cutinases (EC3.1.1.74 - detergent & textile).fas	255991	3
Cutinases (pfam01083 - detergent & textile).fas	2175	9
Heparanase (EC 3.2.1.166 - cosmetic).fas	386	1
Hyaluronate lyase (cd01083 - EC4.2.2.1 - cosmetic).fas	41852	87
Hyaluronidase (EC3.2.1.36 - cosmetic).fas	95	22
Hyaluronidase (EC4.2.2.1-cosmetic).fas	36725	38
Hyaluronidase (pfam03662 - cosmetic).fas	6701	-
Hyaluronidases (EC3.2.1.35 - cosmetic).fas	380042	14
Hyaluronidases (pfam01630 - cosmetic).fas	2219	4
Lactonase (COG1735 - detergent).fas	119142	-
Lactonase (EC3.1.1.25 - detergent).fas	2682	-
Lipase-Esterase (detergent).fas	680	112
Mono(ethylene terephthalate) hydrolases (EC 3.1.1.102 - detergent & textile).fas	824	13
Peptidase type Bromelain (EC3.4.22.32 - textile).fas	179	5
Peptidase type family M04 (detergent & textile).fas	32971	8
Peptidase type family S08 (alcalase - detergent & textile).fas	199971	55
Peptidase type Papain (EC3.4.22.2 - detergent & textile).fas	5459	1
Peptidase type savinase-esperase (EC3.4.21.14 - detergent & textile).fas	1515	34
Peptidase type subtilisin-alcalase (EC3.4.21.62 - detergent & textile).fas	804058	7
Peroxidases (detergent).fas	16189	-
PLA, PCL, Impranil DNL hydrolases (detergent & textile).fas	3022	19
Poly(ethylene terephthalate) hydrolases (detergent & textile).fas	4615	4
Polyurethanase (1) (detergent & textile).fas	4605	1
Polyurethanase (2) -lipase class 3 (detergent & textile).fas	28415	1
Polyurethane degrading urease (EC3.5.1.5 - textiles).fas	152894	-
Trypsin and protease inhibitor (detergent).fas	136	-
TOTAL	3153537	

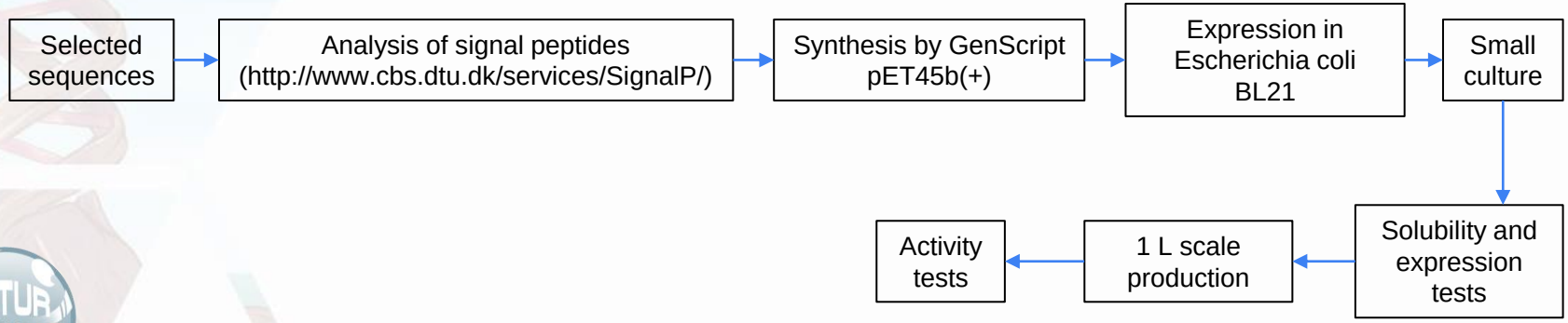
Pipeline for *in silico* search for enzymes

Image representing the different clusters within enzyme classes

Pipeline for *in silico* search for enzymes

List of selected BLAST-hit candidates per each of the reference enzyme classes for gene synthesis

# Enzyme Class	Number
Amidase	6
Amylase	4
Hyaluronidase	11
Hydrolase (esterase, lipase, plastic-degrading)	25
Peptidase	1
TOTAL	47



Task 2.3 Motif building for massive and smart search of enzymes fitting manufacturers' needs M1-M42

Up to now CSIC have compiled the characteristics (substrate specificity, activity in conditions relevant to the project, T_{opt}, pH_{opt} and T_d) of 84 enzymes (lipases, hyaluronidases) relevant for detergents, textiles and hyaluronic acid. This information will be further integrated into the predictive tool. *Note: for characteristics see WP4.*

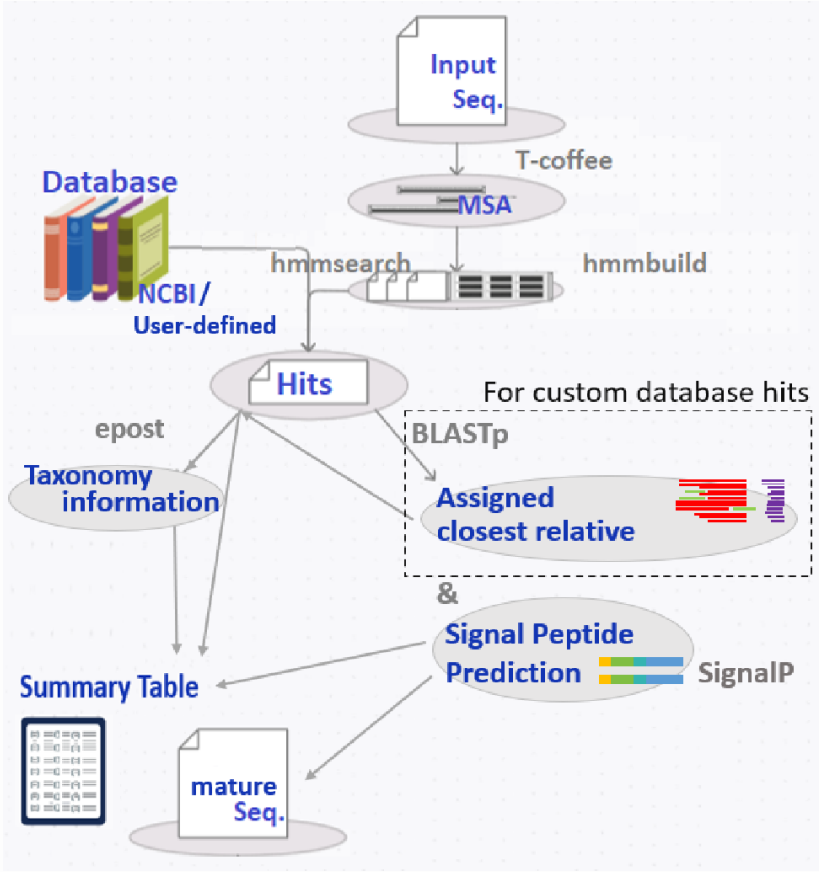
Task 2.4 Iterative and decision-making hierarchical procedure for speed up enzyme discovery

- Construction of Hidden Markov Models (HMMs) for searching both global databases and private datasets

- Development of the AHA-Tool: an **A**utomatic **H**MM Search and **A**nalysis Tool

AHA-Tool v.1.0 functions:

- Multiple Sequence Alignment
- Data & DB formatting
- HMM build
- HMM search
- Sequence extraction
- BLAST
- Taxonomy gathering
- Secretion peptide prediction



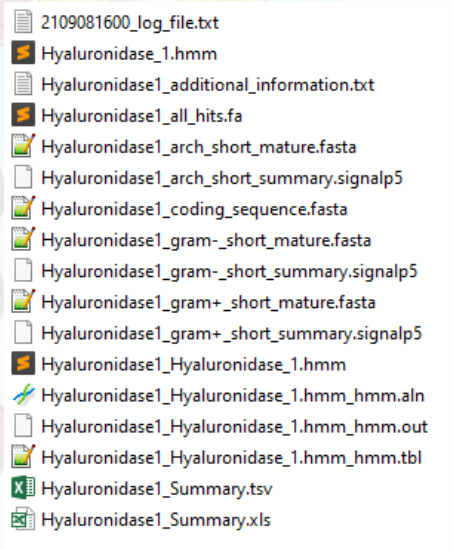
AHA-Tool v.1.0 functions:

- Multiple Sequence Alignment
- Data & DB formatting
- HMM build
- HMM search
- Sequence extraction
- BLAST
- Taxonomy gathering
- Secretion peptide prediction

v.2.0 updates:

- Automatic detection of input file (fasta or hmm)
- Automatic update of NCBI's non-redundant database
- Connectivity check with NCBI server prior to BLAST
- Computing of a maximum likelihood tree with RAxML
- Macro plug-in for automatic taxonomy graphs in Excel
- New output table
- New folder structure
- Code simplified and minor bugs corrected
- Possibility to concatenate jobs

```
mibi_hh01@mibihh01-Precision-5820:/media/mibi_hh01/4TB_SSD/PABLO/AHATool$ bash AHATool.sh -h
USAGE: AHATool.sh [flags] args
flags:
  -p,--prefix:  The prefix the tool will use for produced files. (default: '2205201755')
  -i,--input:  the input file (fasta, aln or hmm). (default: 'sequences.fasta')
  -d,--database:  database options: 1. nr_db; 2. custom_db (default: 'nr.fa')
  -u,--update:  database update if possible? yes/no? (default: 'yes')
  -c,--cladogram:  Prepare tree file for cladogram? yes/no? (default: 'yes')
  -e,--evalue:  e-value (recommended: 1e-10). (default: 0.0000000001)
  -t,--threads:  processor options: 1, 2, 4 (default: 2)
  -h,--help:  show this help (default: false)
```




```
nanopore@nanopore-OptiPlex-7050:/media/nanopore/248d77f7-21d3-4d0e-84bf-9a1d36e43deb1/PABLO/BScNele/AHATool$ bash AHATool.sh -p Hyaluronidase1 -i Hyaluronidase_1.hmm -d nr.fa -u no -c no -t 4

======================================================================
Welcome to AHATool: an Automatic HMM and Analysis Tool.
V.2
Microbiology and Biotechnology - Streit's lab
University of Hamburg (D)
Developed by Nele Schulte (and P.Pérez-García)
======================================================================
Executed on 08.09.2021 04:00:38  by nanopore
AHATool will be executed with the following parameters:
Directory to work in: /media/nanopore/248d77f7-21d3-4d0e-84bf-9a1d36e43deb1/PABLO/BScNele/AHATool
Database: nr.fa
Number of threads: 4
Input file: Hyaluronidase_1.hmm
E-value: 0.0000000001
Output prefix: Hyaluronidase1
RAxML tree: no
Checked for online connection to NCBI.
======================================================================
Checking for needed software:
--------------------------------
 All needed packages are installed.
======================================================================
Checking the input folder...
--------------------------------
Files found: 12
FASTA files found: 2
======================================================================
Checking for needed files:
--------------------------------
 Database (nr.fa) exists.
 Index file (nr.fa.ssi) exists.
======================================================================
The files created within this run will be identifiable by the prefix
"Hyaluronidase1".
They will be saved in the folder /Project_Results
and the subfolder "2109081600".
======================================================================
Searching profile HMM against given database...
```

Outlook

- Apply tool for identifying candidates from other enzyme classes
- Input sequences from partners needed!
- Hyaluronidases, solvent stable enzymes etc.

From the previous selected 108 sequences, a characterization followed by PELE simulations were done.

- Enzyme characterization (catalytic residues, domains, etc)
- PELE simulations for protein and ligand systems

Initial Protein Characterization

PFAM

Complete domains:

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Alpha-amylase	Alpha amylase, catalytic domain	Domain	CL0058	55	406	55	401	1	335	337	393.0	1.5e-117	n/a	Show
Malt_amylase_C	Maltogenic Amylase, C-terminal domain	Domain	CL0369	506	580	506	579	1	74	75	39.2	5.5e-10	n/a	Show

Insignificant Pfam-A Matches

Show or hide all alignments.

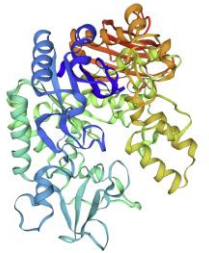
Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Alpha-amylase_C	Alpha amylase, C-terminal all-beta domai ...	Domain	CL0369	328	387	331	372	32	73	96	0.7	770	n/a	Show

Ensure protein family

SWISS-MODEL

Sort	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
1uok.1.A OLIGO-1,6-GLUCOSIDASE *CRYSTAL STRUCTURE OF B. CEREUS OLIGO-1,6-GLUCOSIDASE*							
		0.84	-	52.54	X-ray, 2.0Å	monomer ✓	None

PDB REFERENCE:
1UOK

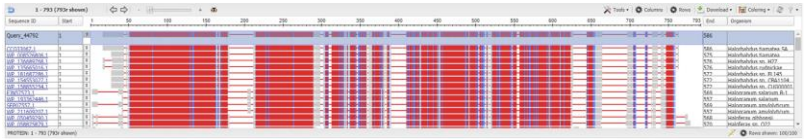


INTERPRO

Catalytic Active Site:

D354
E281
D225

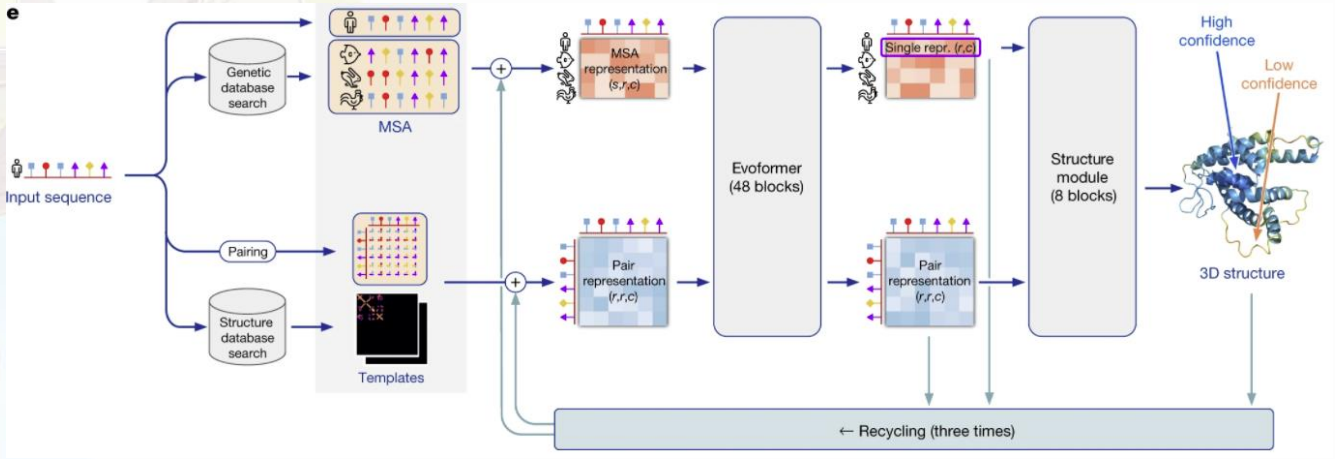
Multiple Sequence Alignment:



Check conservation
Check if patented (?)

BLASTP SUITE

Structure Models from Alphafold2.0



Jumper et al. (2021) Figure 1e. Alphafold Pipeline

Model 1.1.1 (Jumper et al. (2021)) with predicted TM-score (pTM) and aligned errors and full databases preset.

Protein Energy Landscape Exploration (PELE)

Metropolis Acceptance or Rejection

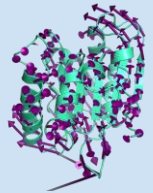
Initial Energy

Pele Step

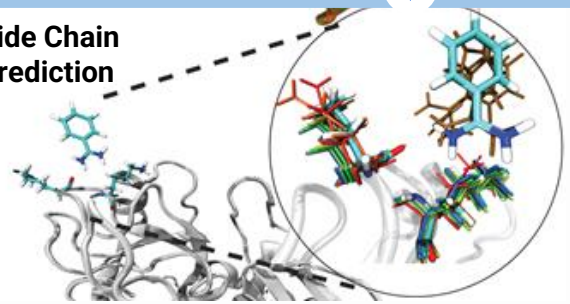
Perturbation

Ligand Perturbation

ANM (Minimization)

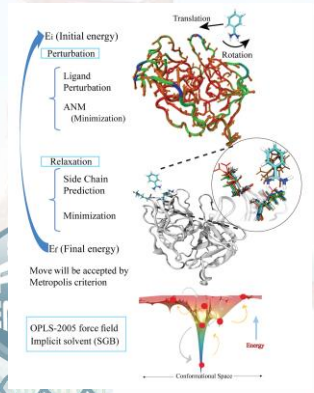


Side Chain Prediction

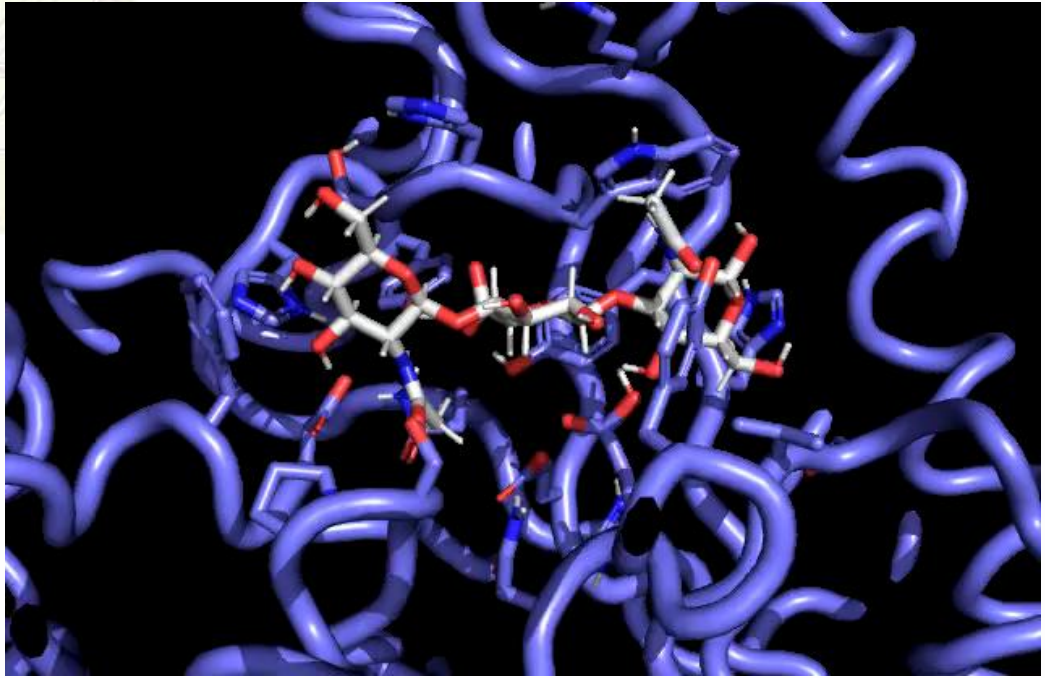


Relaxation Minimization

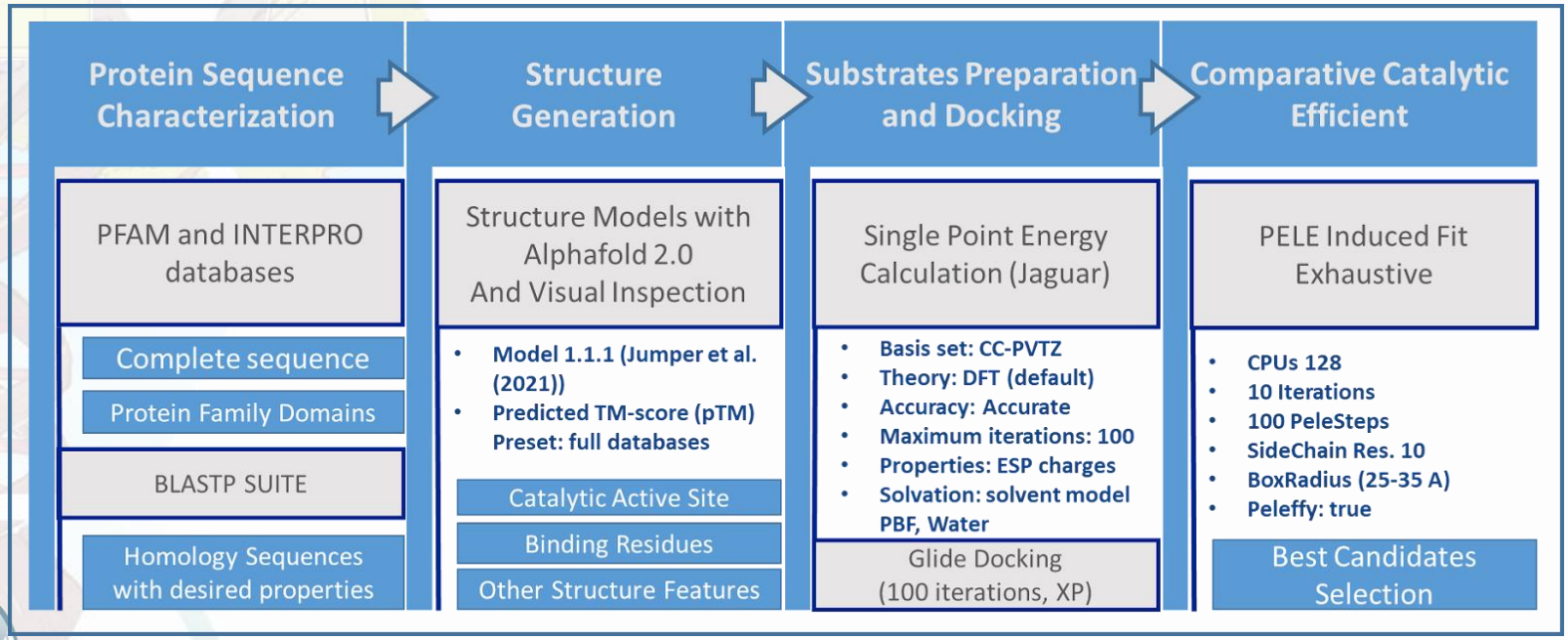
Final Energy



Protein Energy Landscape Exploration (PELE)

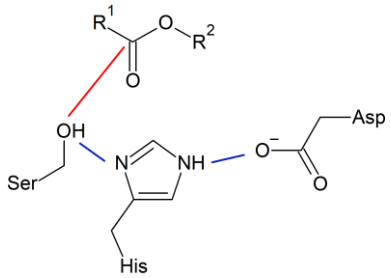


Workflow

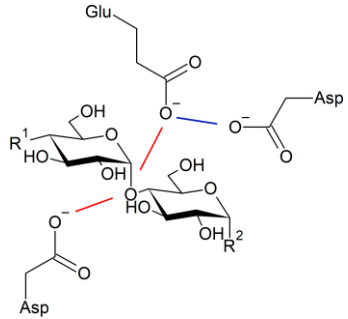

Protein Sequence Characterization
PFAM and INTERPRO databases
Complete sequence
Protein Family Domains
BLASTP SUITE
Homology Sequences with desired properties
Structure Generation
Structure Models with Alphafold 2.0 And Visual Inspection
Model 1.1.1 (Jumper et al. (2021))
Predicted TM-score (pTM)
Preset: full databases
Catalytic Active Site
Binding Residues
Other Structure Features
Substrates Preparation and Docking
Single Point Energy Calculation (Jaguar)
Basis set: CC-PVTZ
Theory: DFT (default)
Accuracy: Accurate
Maximum iterations: 100
Properties: ESP charges
Solvation: solvent model PBF, Water
Glide Docking (100 iterations, XP)
Comparative Catalytic Efficient
PELE Induced Fit Exhaustive
CPUs 128
10 Iterations
100 PeleSteps
SideChain Res. 10
BoxRadius (25-35 A)
Peleffy: true
Best Candidates Selection

Summary of simulations: Enzyme activities

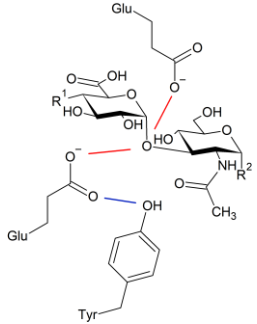
Esterase



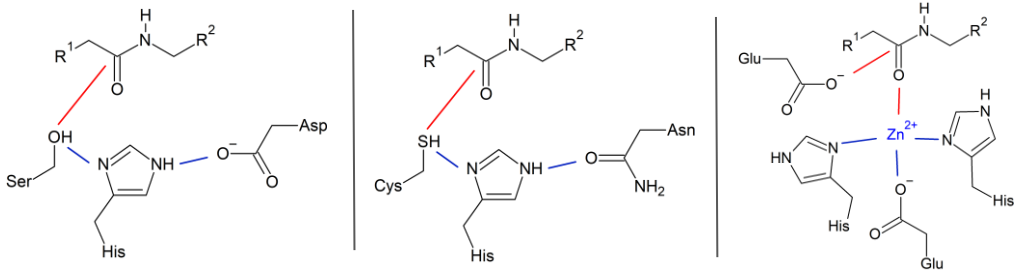
Amylase



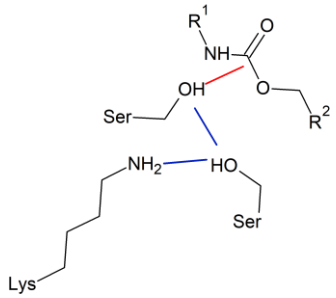
Hyaluronoglucoronidases



Proteases

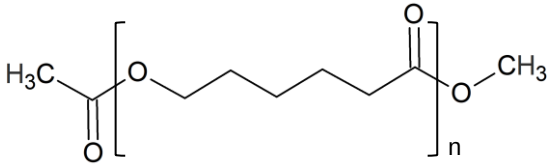


Amidase

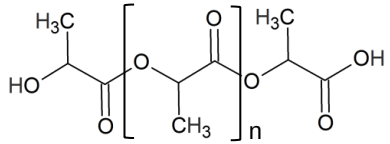


Summary of simulations: ligands

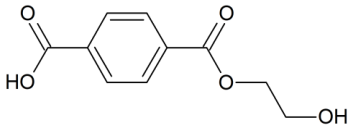
Esters



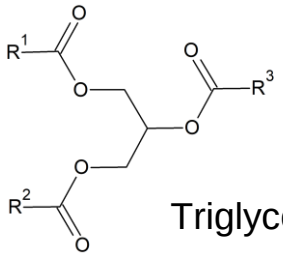
Polycaprolactone



Polylactic acid

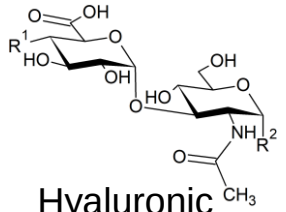


MHET

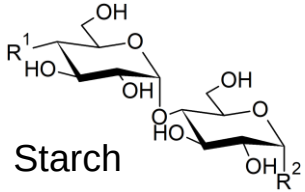


Triglyceryde

Glycans

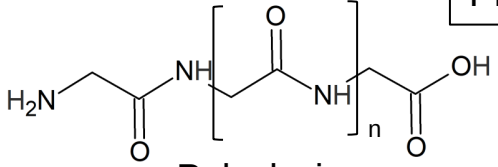


Hyaluronic

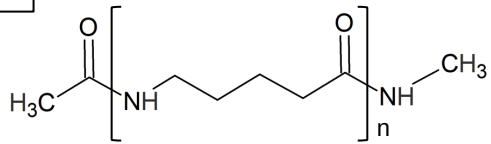


Starch

Proteins

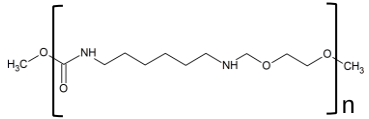


Polyglycine



Nylon

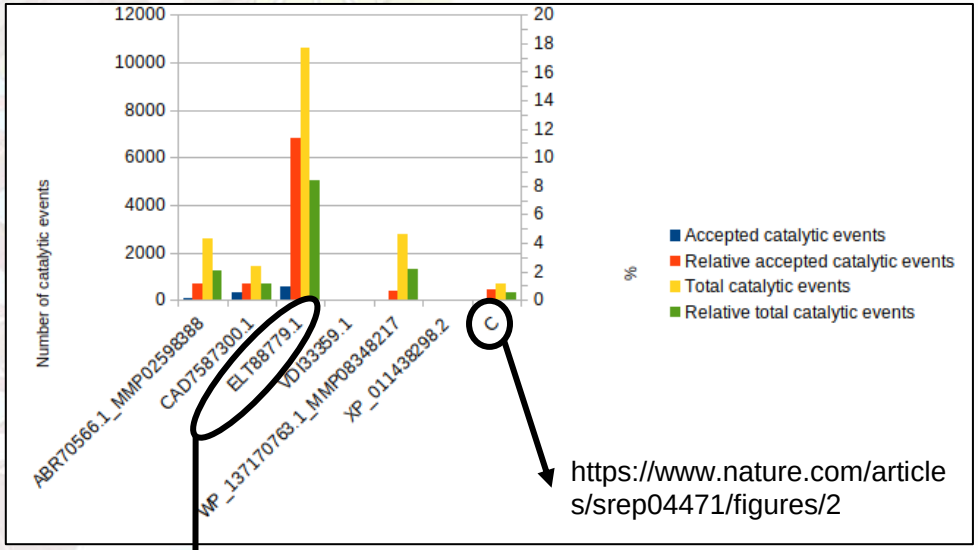
Amides



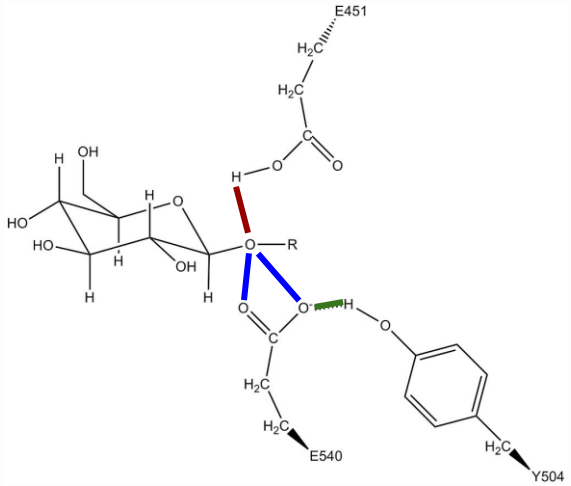
Aliphatic polyurethane

PELE results for hyaluronic acid

Hyaluronoglucoronidases (3.2.1.36/166)



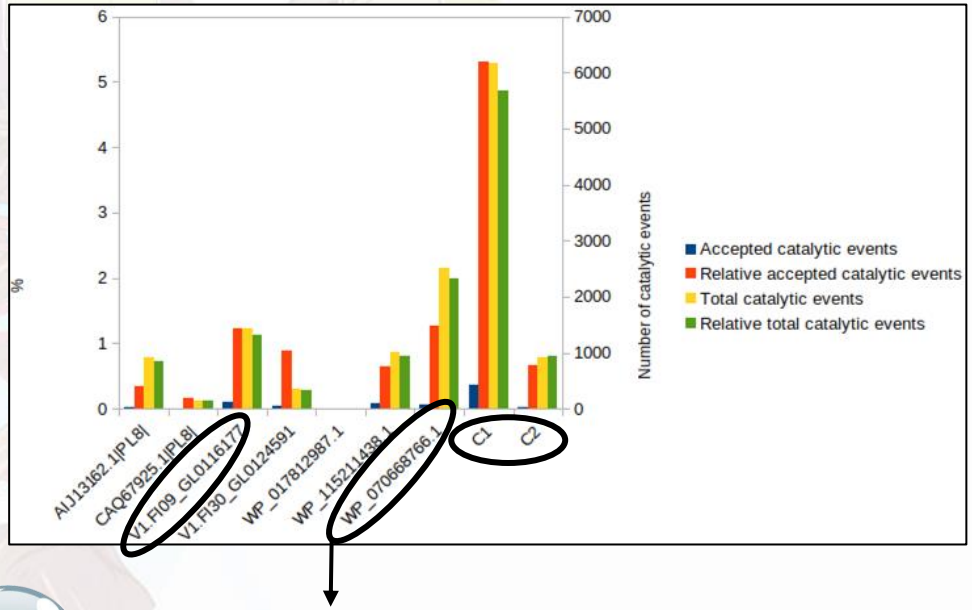
https://www.nature.com/articles/srep04471/figures/2



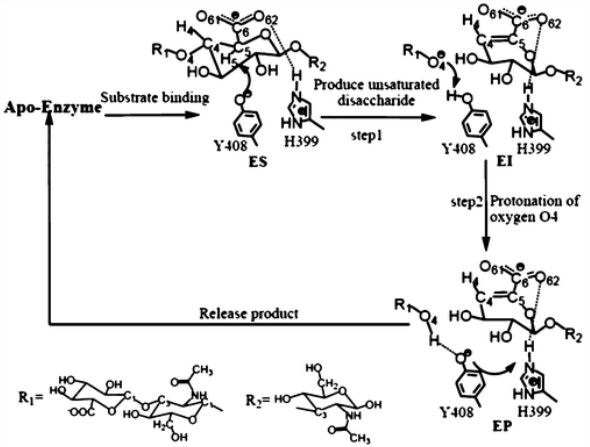
Hyaluronoglucoronidase (3.2.1.36) / Other sequences are heparanases (3.2.1.166)

PELE results for hyaluronic acid

Hyaluronate lyases (4.2.2.*)



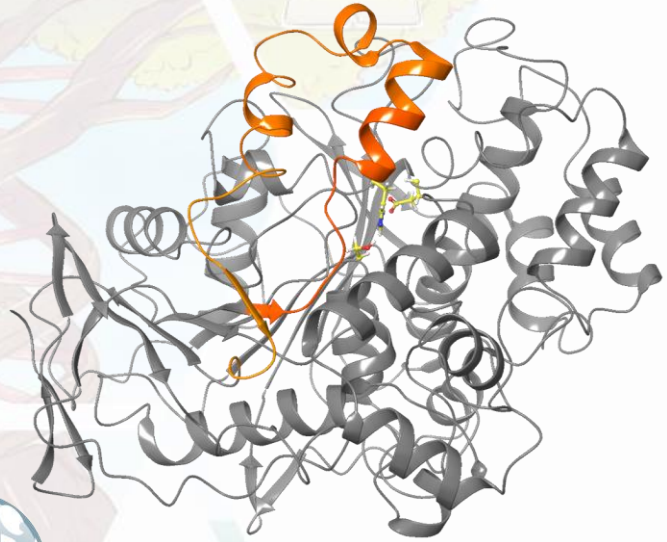
Glycoside Hydrolase Family 16



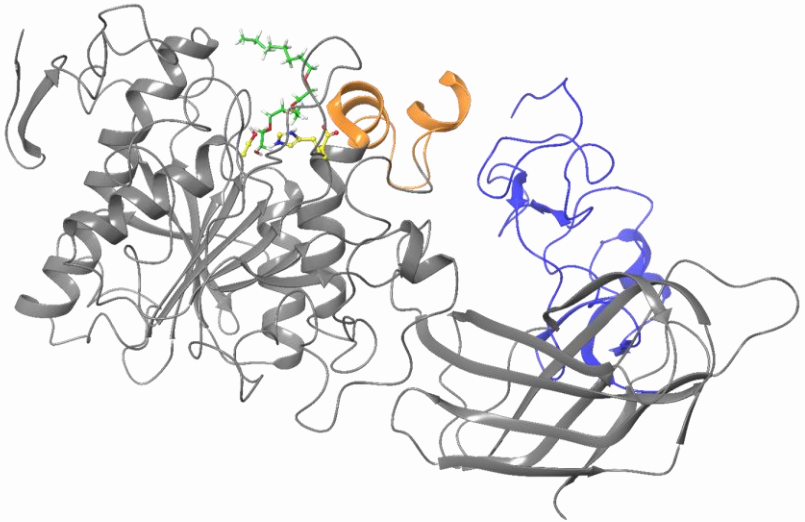
Lipases case

GC and human pancreatic lipases: LID domain and oil-water interface

Geotrichum candidum lipase A & B
(Unilever)



Human pancreatic lipase



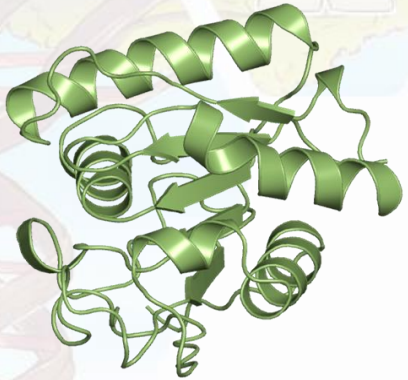
Active site

Lid domain

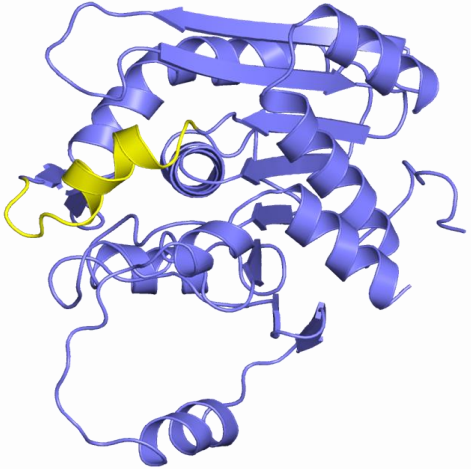
Colipase

Lipases case

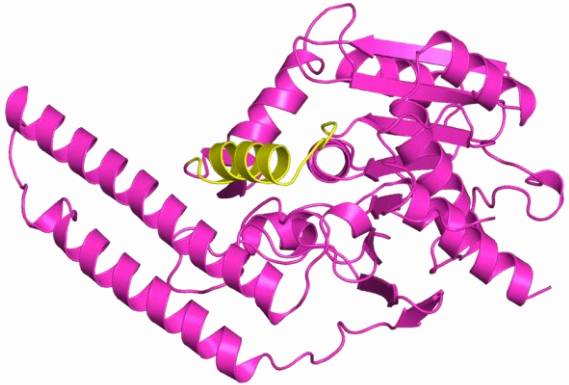
Lipase Structure Comparison among selected sequences



Felip9



Felip5-lid

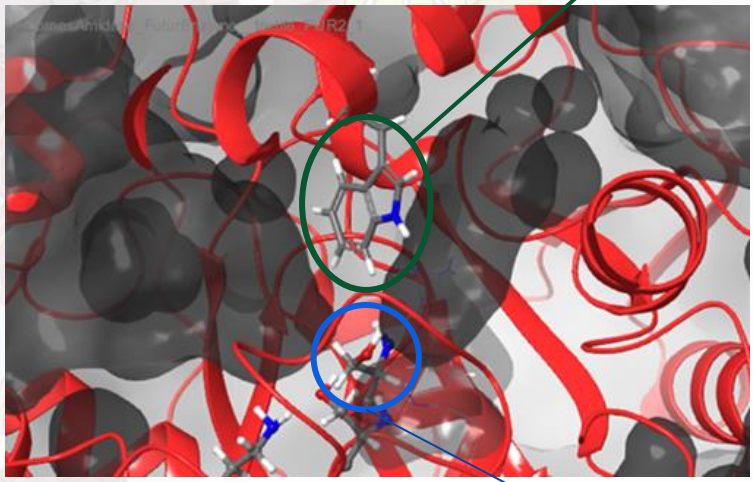


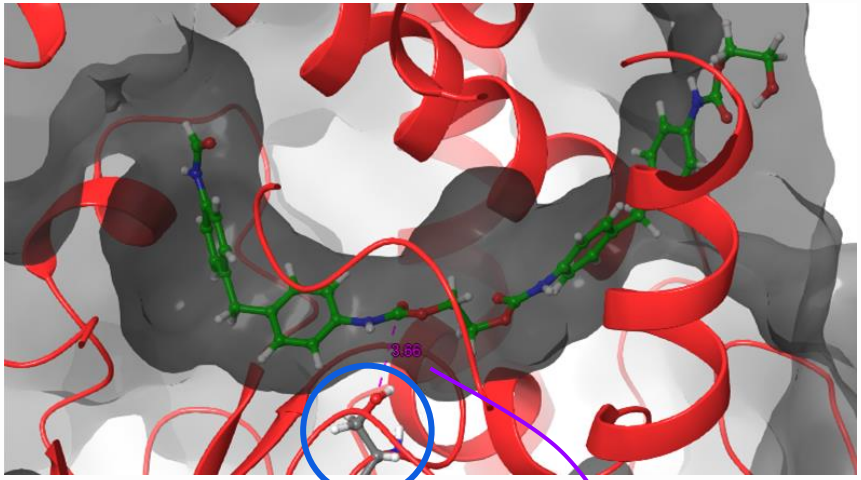
Node494-lip

The LID domains are different for all sequences

Amidases case

This residue controls
the flow of the ligand



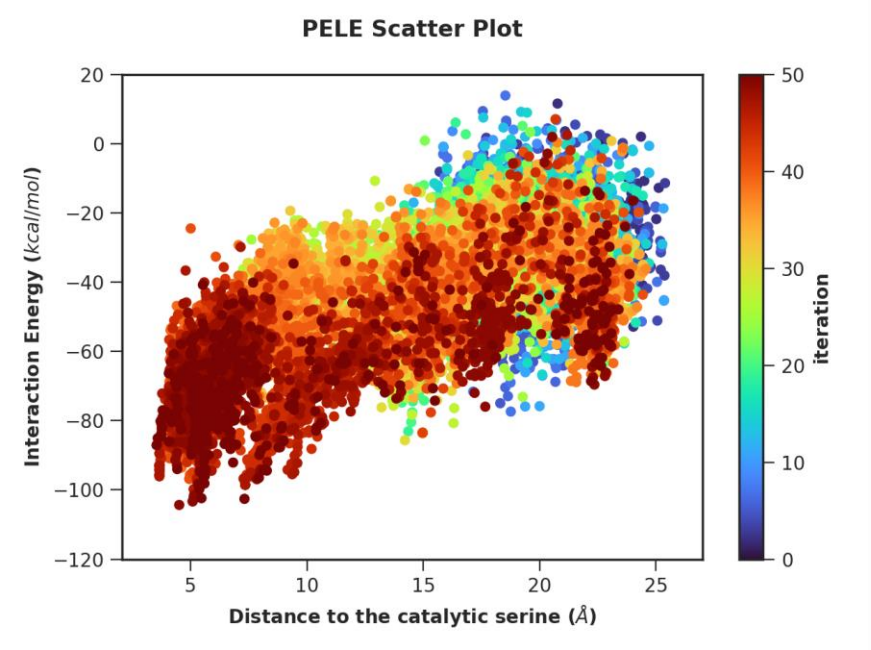


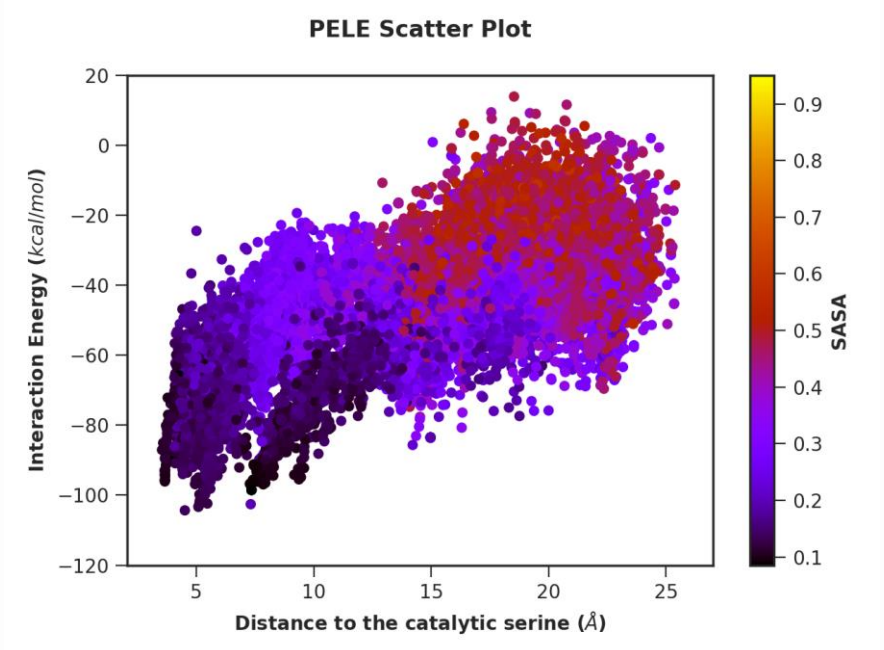
Catalytic serine

Catalytic distance

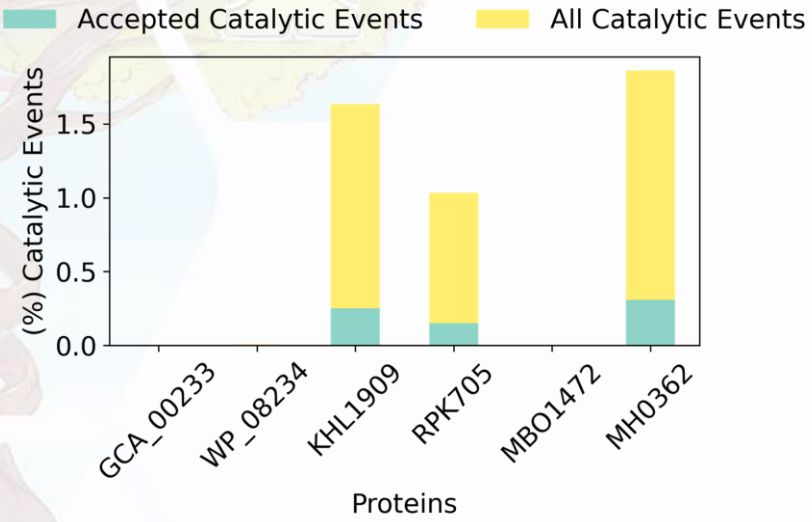
Amidases case

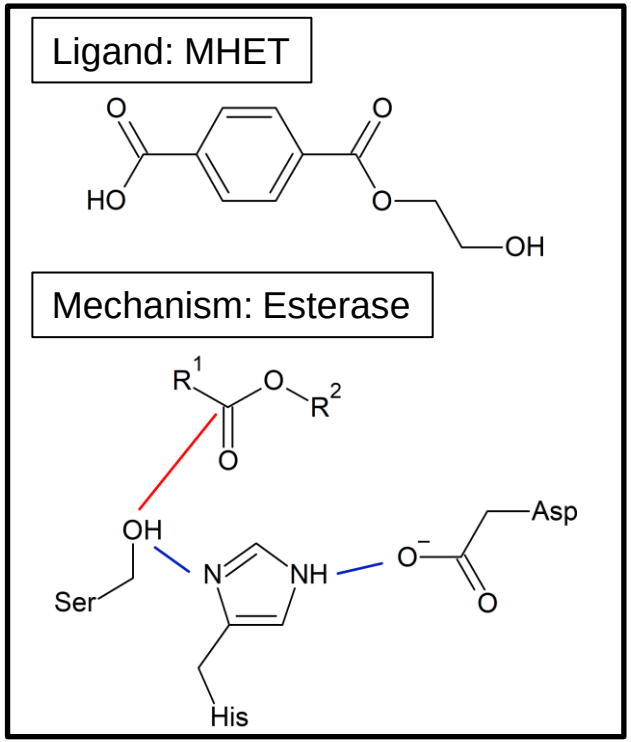
The simulation performed was an out-in, which enhances the poses that have shorter catalytic distances, in this case (left). When the ligand enters to the catalytic tunnel, SASA and interaction energy decrease (right).



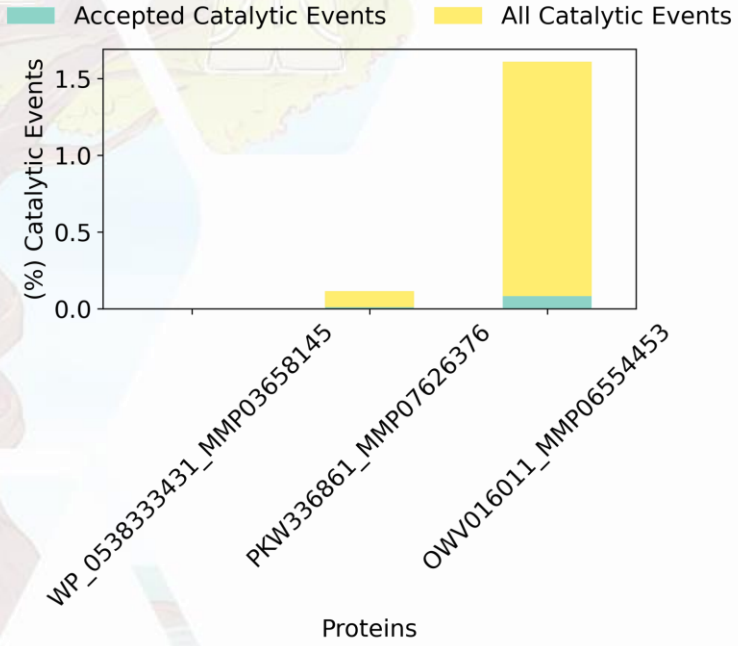


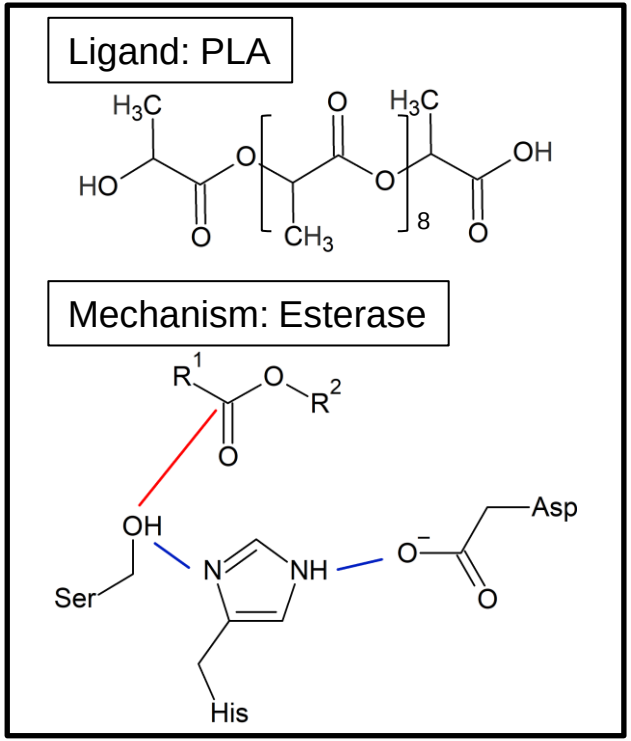
Catalytic events for the 6 selected MHETases



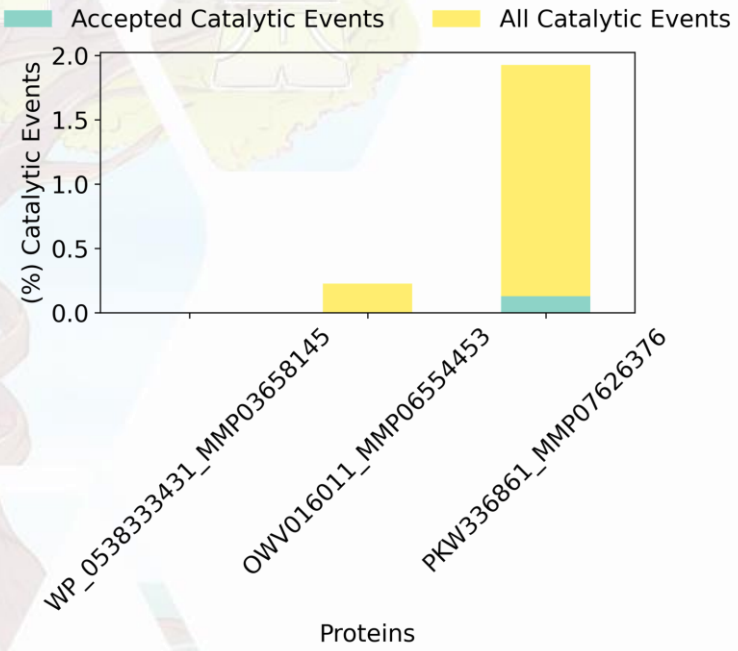


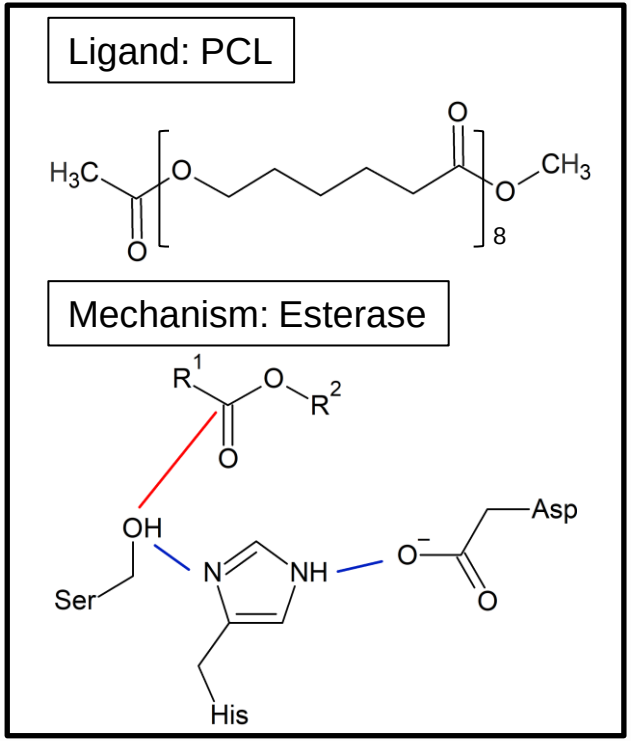
Catalytic events for the plastic degrading enzymes against PLA (polylactic acid)



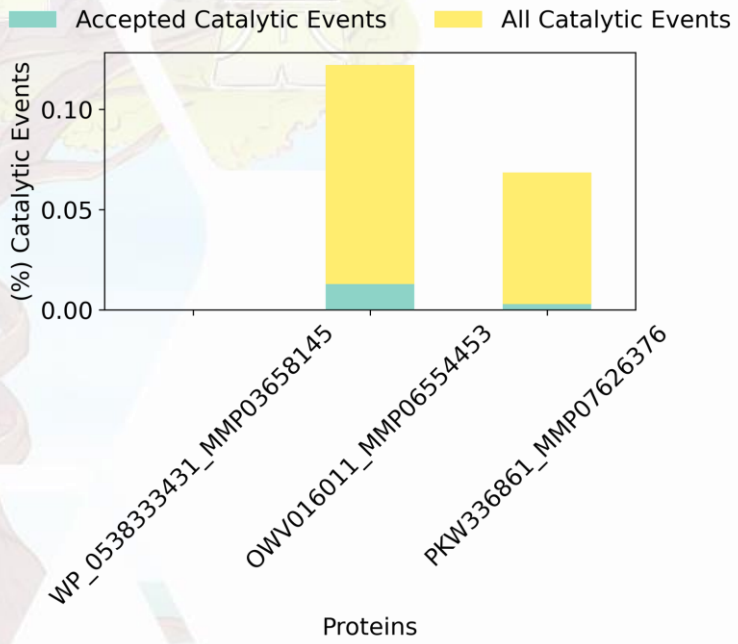


Catalytic events for the plastic degrading enzymes against PCL (polycaprolactone)

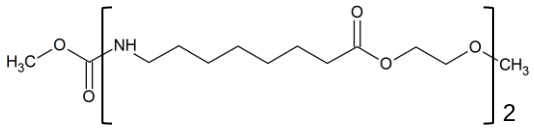




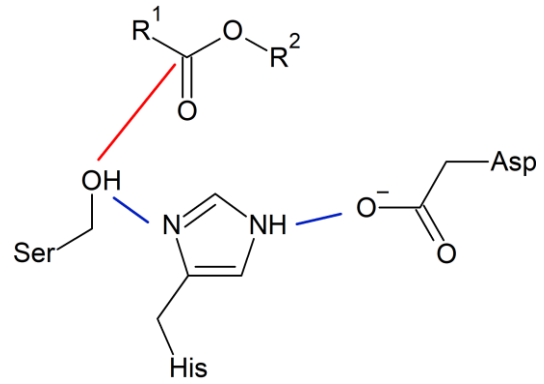
Catalytic events for the plastic degrading enzymes against aliphatic polyester-polyurethane



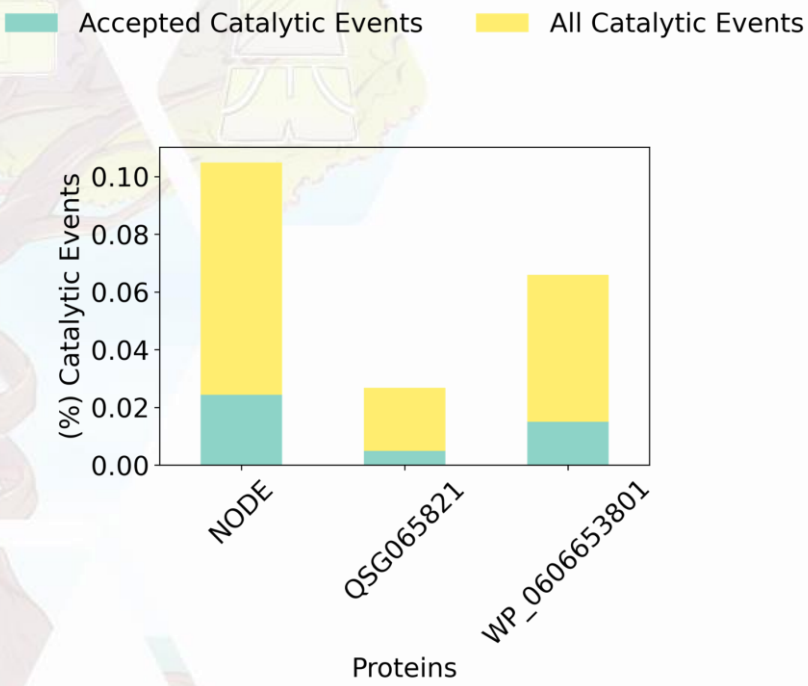
Ligand: Polyester-polyurethane



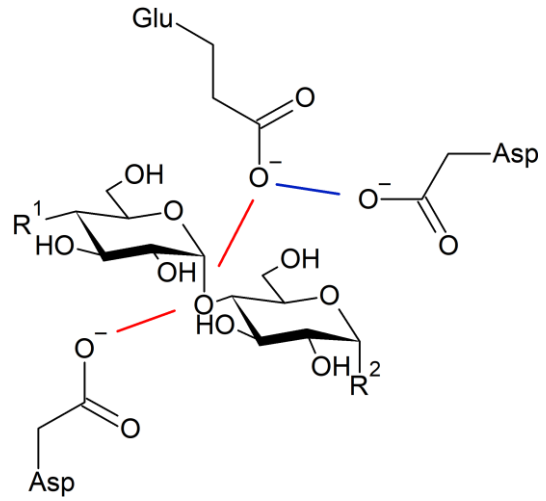
Mechanism: Esterase



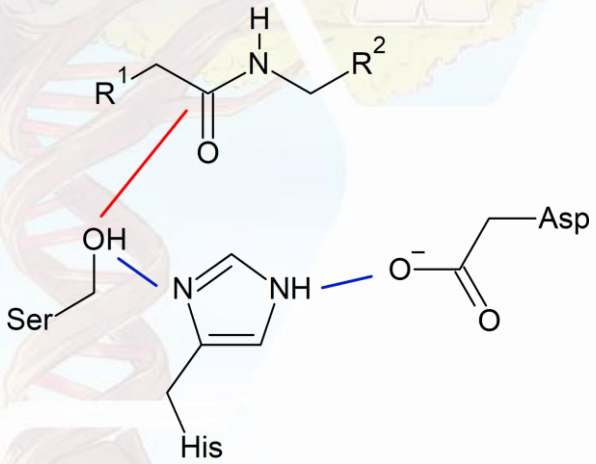
Catalytic events for the amylases



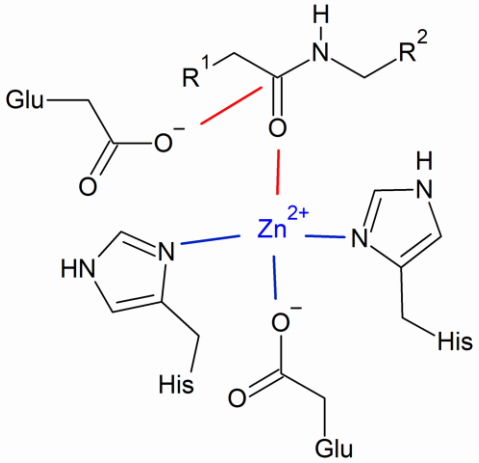
Ligand: Starch
Mechanism: Amylase



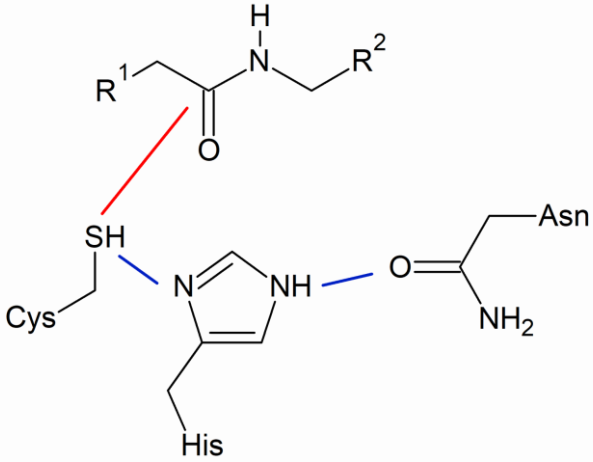
3 different mechanisms for proteases



Serine proteases



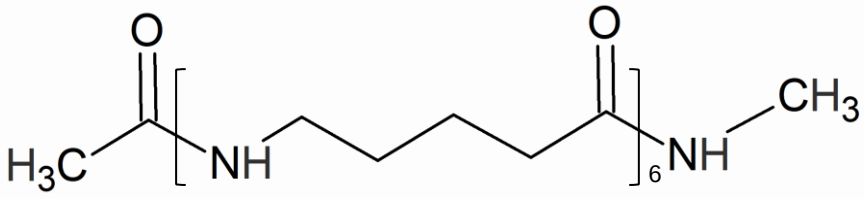
Metalloproteases



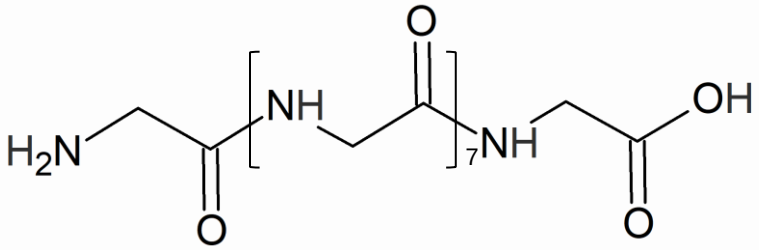
Cysteine proteases

Two different ligands for proteases

Nylon

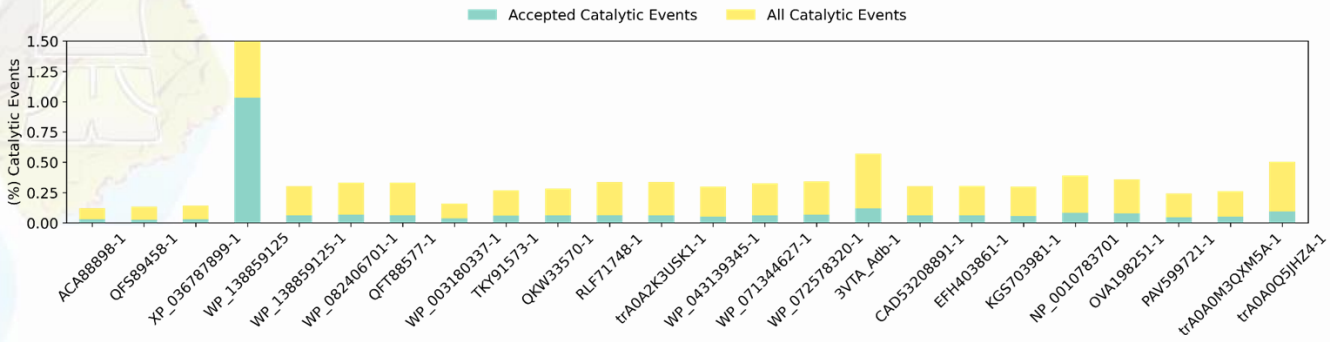


Polyglycine

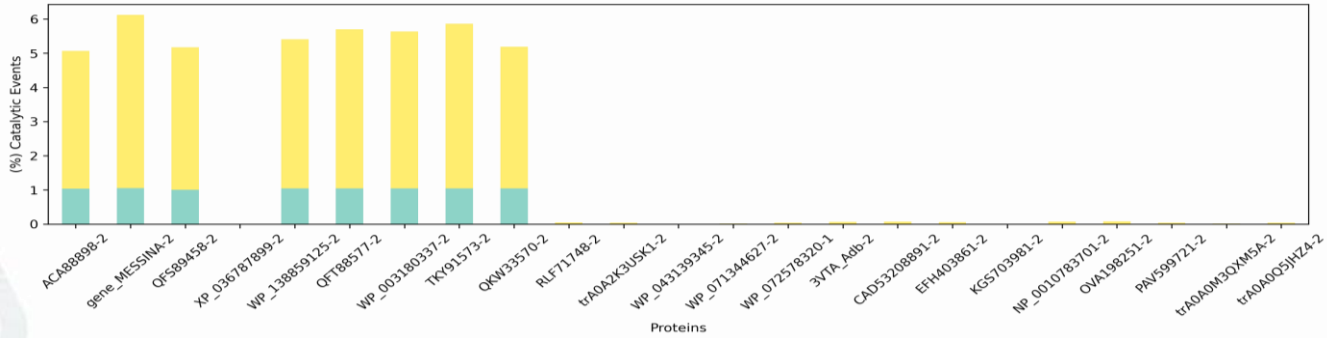


Catalytic events for the Serine proteases

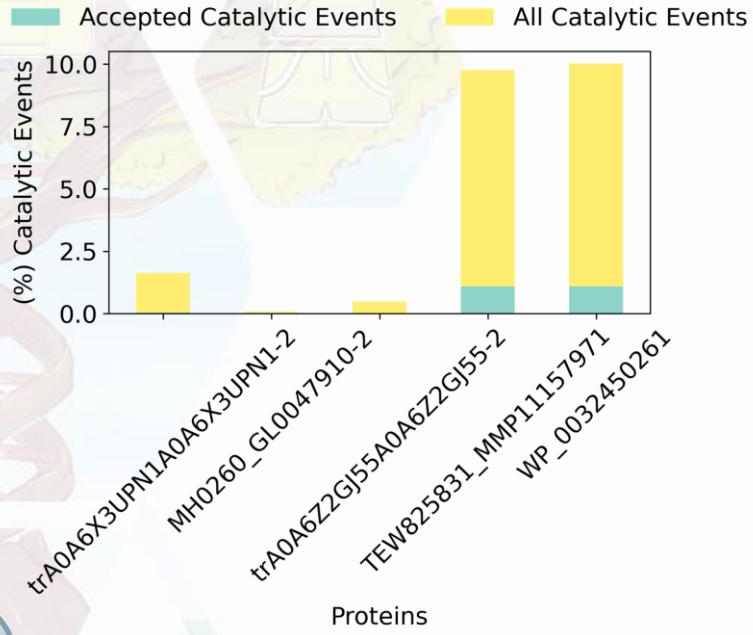
Nylon



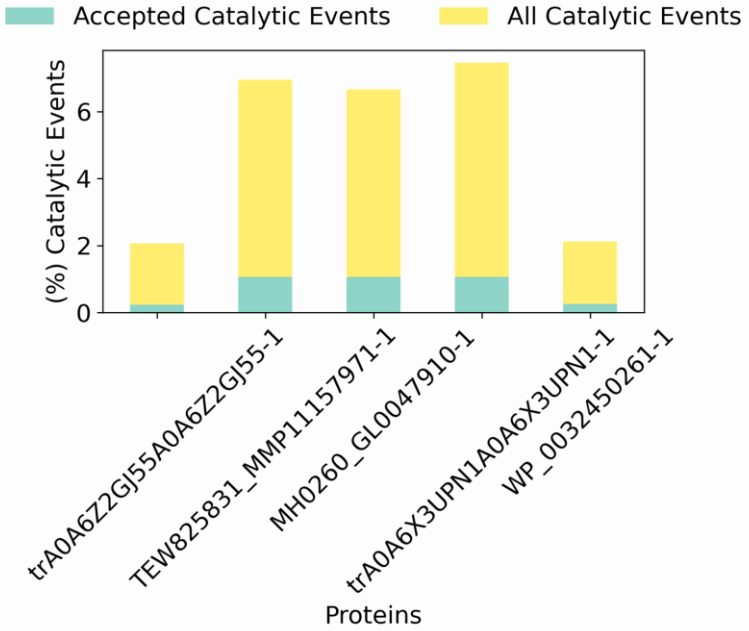
Polyglycine



Catalytic events for the metalloproteases

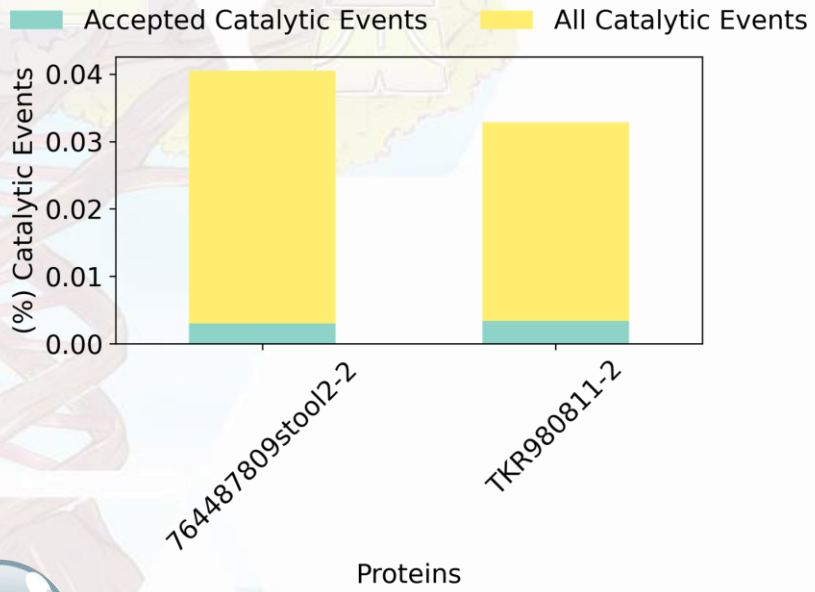


Nylon

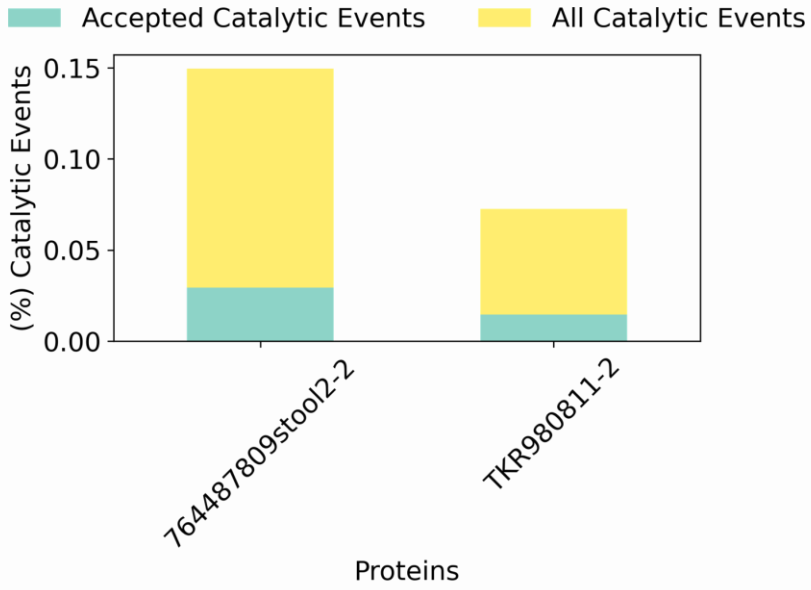


Polyglycine

Catalytic events for the cysteine proteases



Nylon



Polyglycine

Conclusions and Future Perspectives WorkPackage 2

- Combine experimental and computational results to refine the simulations
- Use the AHA-tool for finding new enzymes
- Do a second iteration of bioprospecting and filtering
- Extend to other activities that have not been tested yet
- Select the best sequences to create better mutants (WP-5)

FuturEnzyme

Technologies of the FUTURe for low-cost ENZYMEs for environment-friendly products

WP 2
12 months meeting

Project funded by the European Union's Horizon 2020
Research and Innovation Programme under grant agreement No [101000327]

Original image: Illustration by Ainhoa Quirós