Almac Sciences' INSIGHT platform accelerates Genes-to-GMP implementation

Tom Moody[1,2], Derek J. Quinn[1], Jane Müller[1], Alexandra T. P. Carvalho[1]

[1]Almac Sciences Ltd., Department of Biocatalysis & Isotope Chemistry, 20 Seagoe Industrial Estate, Craigavon, UK;

[2] Arran Chemical Company Limited, Unit 1 Monksland Industrial Estate, Athlone, Co. Roscommon, Ireland;

Correspondence to:

tom.moody@almacgroup.com; derek.quinn@almacgroup.com; jane.mueller@almacgroup.com; alexandra.carvalho@almacgroup.com

Modern generation of Active Pharmaceutical Ingredients (APIs) for human and animal health involve increasingly complex molecular structures, predominantly exemplified by chiral small molecules, ever-longer peptides, semi-natural oligonucleotides, and therapeutic antibodies. Biotechnology has always held a key role in accessing high molecular weight complexity. Traditionally, complex structures like antibiotics, had to be produced by fermentation, whereas smaller chiral APIs were typically manufactured from a pool of chiral starting materials or classical resolution. This changed when modern asymmetric catalysis offered better productivity towards an ever-increasing range of accessible targets, resulting in APIs which were affordable and critically safe medicines for all.

Modern asymmetric catalysis utilises chemo-, organo-, and bio-catalysis. Biocatalysis stands out from this myriad of technologies for its unique ability to tap a practical infinity of natural diversity and its inbuilt ability to be easily modified or engineered in a purposeful and directed fashion. Supporting technologies for enzyme development include advancements in gene synthesis, high throughput screening methods and rational *in silico* technologies which have reduced timelines and cost for development, which in turn have accelerated their acceptance in chemical manufacture.

To support further deployment of enzyme chemistry, Almac has developed its own proprietary advanced enzyme discovery platform named INSIGHT (IN Silico Informed metaGenomic Harvesting Technology) (Figure 1) to further reduce timelines and contribute to the growing trend of the use of biotransformations for chemical manufacture towards APIs.

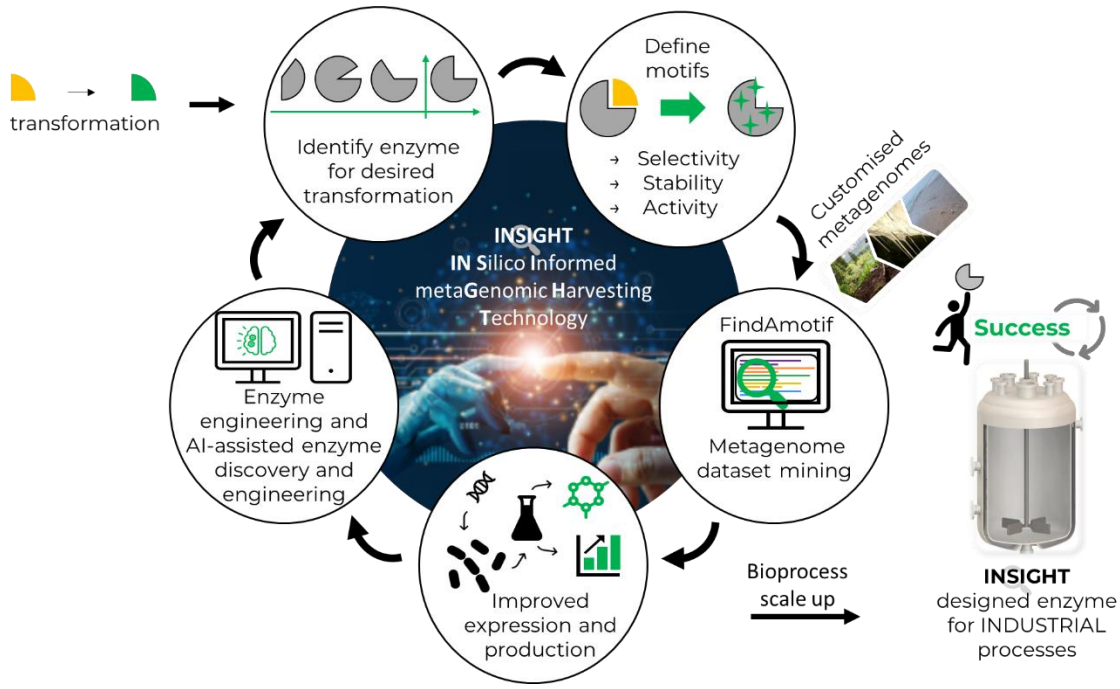


Figure 1: Almac's INSIGHT enzyme platform.

The INSIGHT enzyme platform was established with 4 key objectives in mind:

- Development of a new UK-based and freedom to operate metagenomic database unique to Almac.
- Establishment of a state-of-the-art computational facility for mapping enzymes with new substrates and interrogating the new metagenomic databases with computer-optimised enzyme sequences.
- Development of a cloning, expression and screening capability for these metagenomic derived smart enzyme libraries.
- Development and implementation of Artificial Intelligence (AI) algorithms to minimise enzyme discovery and engineering timelines.

By leveraging the INSIGHT platform, which involves metagenomic and bioinformatic enzyme discovery, high throughput screening, cloning and expression, *in silico* enzyme engineering and AI, Almac can identify and engineer new enzymes to support the efficient synthesis of chemical intermediates and APIs from gene to GMP manufacture. To augment the INSIGHT platform, Almac integrates these enzyme

libraries into its multi-level enzyme engineering platform to further shorten timelines and make engineering of enzymes affordable, resulting in the acceleration of "genes-to-GMP" implementation.

With these aims, the basic concept of INSIGHT is to improve the enzyme discovery and engineering process to allow more robust enzymes to be used for the displacement of multi-step chemical routes which are in many instances used to produce APIs, intermediates and/or building blocks by simpler, low cost and less polluting enzymatic processes. INSIGHT has been developed to use the power of *in silico* design and AI to aid in the structure identification of the best enzyme for a given biotransformation.

From this structural information, gene sequences can be generated containing unique hot spots which are required by the enzyme to have superior performance characteristics for a given biotransformation. These hot spots can then be identified from a unique metagenome and new enzymes identified for cloning, expression and testing. The Almac INSIGHT platform and approach results in the identification of a higher performing enzyme at project initiation and therefore can shorten timelines for the development of the final commercial enzyme (Figure 2).

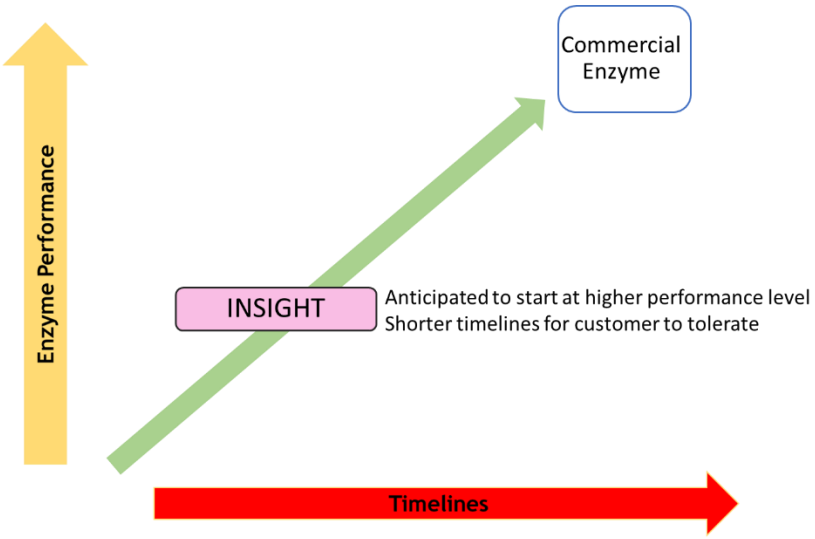


Figure 2: Almac's INSIGHT platform is used to identify higher performing enzymes at project initiation.

Unique metagenomic derived enzymes accessed using INSIGHT enzyme discovery

To ensure viable enzyme performance, which is a multi-step, multi-level process, it is crucial to correctly choose the initial enzyme. Often, nature has some suggestions towards a solution. This is particularly relevant in early clinical phase pharmaceutical development, where time and budget constraints exist. To serve these clients and projects, it is beneficial to have an extensive reservoir of library enzymes

from natural diversity and previous enzyme engineering projects. To add value to database-centric approaches, Almac uses INSIGHT to inform library design and access new natural enzymes to augment existing libraries where the selected enzymes already have specific catalytic hot spot motifs in their backbone resulting in superior enzymes from the outset.

INSIGHT smart enzyme selection from large data sets

Protein databases harbour more and more genomes and unannotated datasets and therefore the need for targeted enzyme discovery and smaller library sizes has increased. Almac has developed its FindAmotif algorithm to determine the enzymes' relevant region(s) for any desired biotransformation so that interrogation of any genome can be focused. Any information about the desired biotransformation, such as the relevant enzyme families, evolutionary information, enzyme kinetics and, mutational data, as well as the structure of relevant intermediates in the reaction mechanism and their conformational sampling obtained using molecular docking, Molecular Dynamics (MD), Quantum Mechanics/Molecular Mechanics (QM/MM) simulations. AI methods are then used to define enzymatic motifs relevant to the chosen biotransformation and are used to create the customized INSIGHT FindAmotif algorithm. With this powerful algorithm vast amounts of protein data from both public databases and Almac's own diverse metagenomic datasets can be filtered to create focused smart targeted enzyme libraries which reduce the screening burden significantly (see Figure 3).

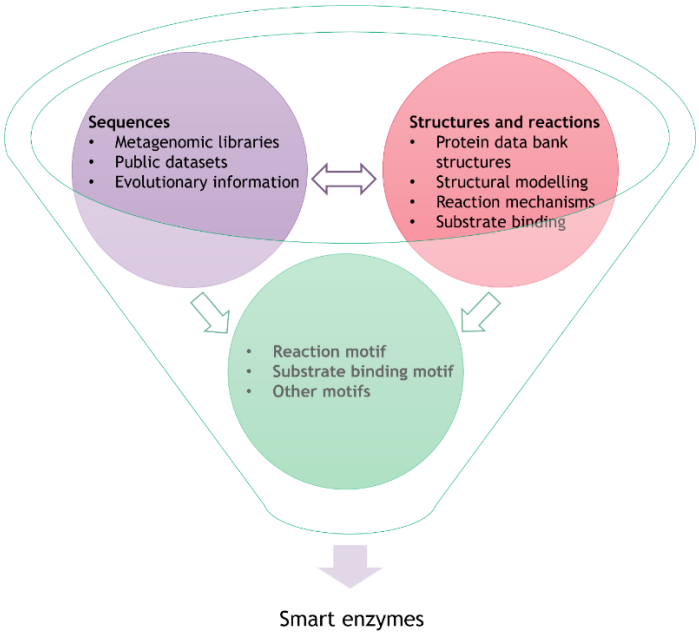


Figure 3: Oversight of INISGHT enzyme selection process leveraging sequence, structural and mechanistic information.

INSIGHT smart enzyme selection of carbonyl reductases for use with “bulky-bulky” substrates

Carbonyl reductases (CREDs) are an important class of industrial enzymes which reduce ketones into alcohols with excellent enantioselectivity. INSIGHT’s FindAmotif was used to identify important reaction and substrate-binding motifs to screen *in silico* metagenomic libraries. Using INSIGHT, a focussed library of candidate genes was identified and after gene synthesis, enzyme cloning, expression and testing, top hits were able to convert challenging substrates with conversions up to 80 % at screening scale (Figure 4).

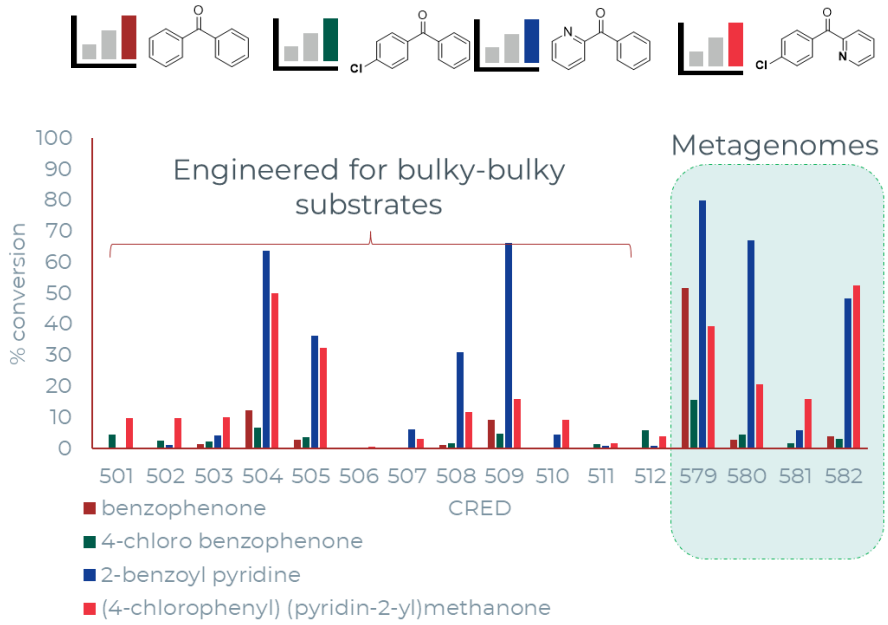


Figure 4: INSIGHT informed metagenomic library of bulky-bulky CRED enzymes.

The enzymes discovered through our INSIGHT methodology showed superior substrate promiscuity for bulky-bulky substrates compared to other engineered enzymes. This marks an improved starting point for subsequent enzyme engineering (if required).

Almac generates large amounts of data during enzyme screening and enzyme engineering projects. To fully leverage this data and continue to improve and expand enzyme panels and services, Almac are developing and utilising INSIGHT deep learning substrate-enzyme interaction models. These models are then used to predict further homologues and mutants of interest for specific substrates.

INSIGHT enzyme engineering and development for maximising biotransformation conditions

The developed reaction conditions for many biotransformations in a chemical reactor are often far from those observed for an enzyme in its natural biological environment. Due to this, enzymes may need to

be further engineered to increase stability, activity, etc. to be process ready. Typically, enzyme engineering is deployed as a one-size-fits-all solution, however, Almac uses a mixture of enzyme engineering, chemical reaction development and chemical engineering, which provides a more holistic approach. A one-track approach should never be used, and in many cases enzyme formulation, reaction development or deployment of chemical engineering tactics are the solution to deliver a fit-for-purpose process. However, if enzyme engineering is 100 % required, it is very important that any engineering of the enzyme is conducted so the overall conditions and reactor set-up is considered as part of the study.

For example, Transaminases (TAms) are important enzymes that catalyse the reversible transfer of an amino group from a donor to an acceptor, leading to the production of valuable chiral amines with high chemical and optical purity. However, these enzymes can suffer from a plethora of process inadequacies including poor tolerance to high concentrations of substrate, donor and by-products. Isopropylamine (IPAm) is a common inexpensive amine donor and should always be considered as a key material of any engineering programme to ascertain how the enzyme behaves in its presence. When using INSIGHT technology for the engineering of the TAm it is imperative that all models and algorithms take into account this key reaction material and no model is developed in the presence of water only. The differences in selection of mutational hotspots is profound from any study conducted using the INSIGHT methodology and would never be predicted unless a more defined process awareness was deployed. An example of an INSIGHT engineering TAm study resulting in an improved activity at high IPAm concentration is shown in Figure 5.

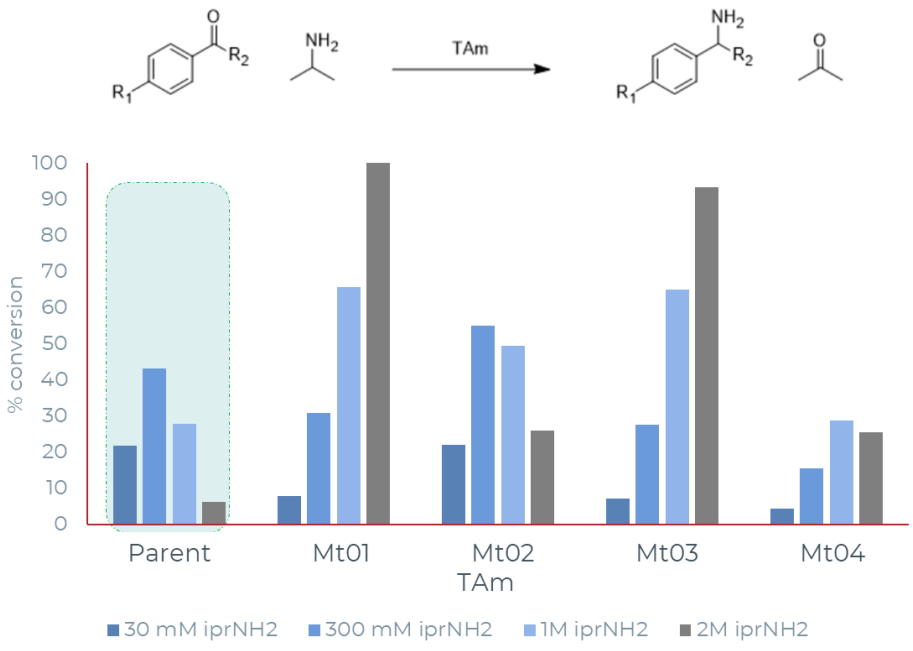


Figure 5: INSIGHT engineered TAm mutants able to tolerate high IPAm concentrations.

The INSIGHT engineered TAm had the ability to tolerate an increase in IPAm concentration which enabled the biotransformation to proceed towards reaction completion, whereas the starting TAm showed a decrease in product accumulation at high concentrations of IPAm amine donor.

INSIGHT *E. coli* expression platform to increase enzyme production

One of the key cost contributors for any biocatalytic process is the enzyme. If enzyme loading can be reduced (e.g. through increased activity with INSIGHT engineering), this reduces the economic burden and makes the process more economically feasible. However, another lever to pull is the cost for the enzyme production, which can be reduced from the overall expression levels per unit fermentation volume. More enzyme per litre of fermentation means a lower price per kg for the enzyme! Almac has developed its own INSIGHT *E. coli* protein expression platform to be deployed to all their inhouse and customer enzyme projects. This novel platform has been shown to produce on average 4-6-fold improvement in soluble expression of enzymes (Figure 6) when compared with the standard, baseline expression system. This results in the generation of higher yields of enzyme production in fed batch fermentation and lowers the cost per kg of enzyme.

A) B)

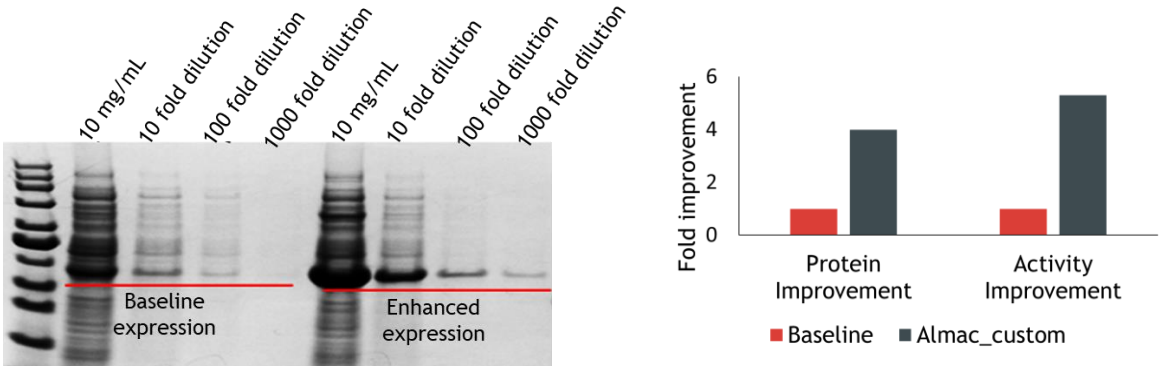


Figure 6: INSIGHT plasmid platform resulting in improved soluble (A) enzyme production and (B) resultant enzyme bulk activity.

INSIGHT specific eukaryotic expression system for difficult to express enzymes

INSIGHTs, eukaryotic secretory, high yielding, non-lipopolysaccharide producing GRAS organism for expression of enzymes is used to express enzymes which are not suitable for *E. coli* expression. This is essential when probing metagenomic space to deliver functional and active enzymes. With interchangeable genetic elements, promoters, integration sites and selection markers, this system is extendable to future improvements and multifunctional designs. As a proof of concept, animal derived lipases were inserted into this yeast expression system and using optimised growth conditions showed good expression and higher activity than that obtained by standard *E. coli* expression systems (Figure 7).

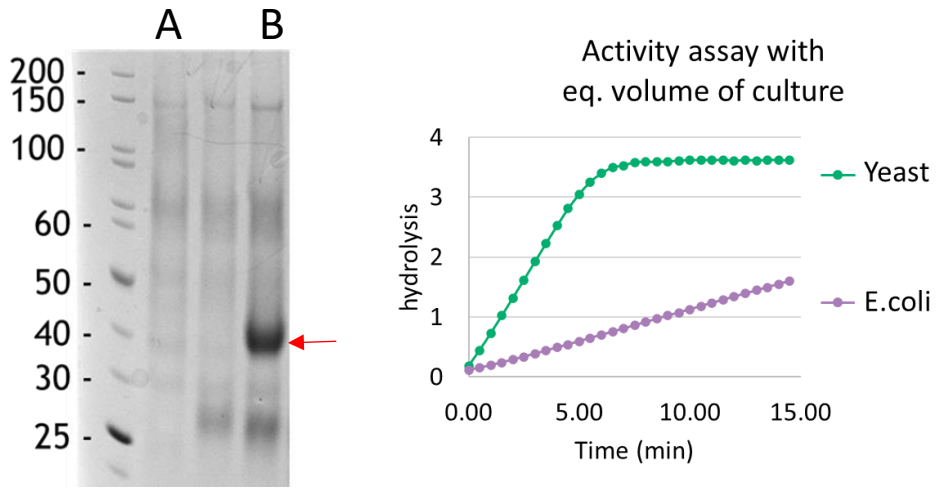


Figure 7: INSIGHT eukaryotic secretory expression of a selected animal derived lipase. A) No gene inserted and (B) gene inserted showing expression of the enzyme.

The SDS-PAGE analysis and activity test showed that this enzyme was expressed in a more active form in the INSIGHT eukaryotic secretory system compared to standard *E. coli*.

INSIGHT AI using machine learning and deep learning

Rapid prediction of enzyme properties for catalysing industrial substrates is crucial for selecting promising enzymes for further development and/or engineering. However, the search for suitable enzymes and optimisation of reaction conditions can be time-consuming and costly. Computational tools with AI algorithms such as classical machine learning (ML) and deep learning (DL) neural networks have emerged as valuable resources for predicting structures, generating protein sequences and predicting compound-protein interactions. Sustainable production utilising biological resources highlighted a pressing need to speed-up biocatalyst screening by integrating ML and AI techniques. Almac has developed its in-house INSIGHT ML and DL algorithms for prediction of engineered enzymes. The technology is being used to lower the screening burden through smart selection of genes and mutated genes at the discovery and engineering stages, respectively.

Conclusion

The proven ability of biocatalytic technology to produce hard cost savings for pre-existing processes or to provide economical access to new chemical entities in the pharmaceutical and chemical sector ensures increased investment year-on-year in this area. Key to the success of our INSIGHT gene-to-GMP concept is a deep understanding and experience of process development in combination with sophisticated *in silico* design methods and the valuable learning from existing data leveraged through ML and AI. The future is exciting for enzyme catalysis, and it is time to broaden your “INSIGHT” in this fast-evolving area.

INSIGHT and the future...

Moving forward, INSIGHT will continue to accelerate biocatalyst discovery, enzyme engineering and expression by bringing together computational, metagenomics, biology and chemistry technologies and methodologies. As the INSIGHT engineering, ML and AI process continues to improve, we are setting our sights on new pharmaceutical challenges and to truly to accelerate the gene to GMP concept. Enzyme catalysis is here and should be deployed. If you have not tried this technology, it is time to explore and widen your chemical armoury!

Biography

Prof. Tom Moody graduated from The Queen's University of Belfast with a 1st Class BSc (Hons) in chemistry in June 1998 before returning to gain a Ph.D. in Physical Organic chemistry in December 2001. He has also completed a master's in business graduating with distinction in July 2007 specializing in business strategy. His work has earned him numerous accolades and is co-author and author of >140 publications and patents. Tom is currently VP Technology Development and Commercialisation for Arran Chemical Company and Almac Sciences in Ireland and works in the field of chemical methodology and its application towards the synthesis of chiral molecules, metabolites and labelled compounds. He is also an honorary Professor at Queen's University of Belfast in Green Chemistry. E: tom.moody@almacgroup.com

Dr Derek Quinn has worked for over 20 years in the field of enzyme discovery, successfully elucidating and characterising enzymes from both prokaryotes and eukaryotes. He has extensive experience of expression and evolution of enzymes for their application in both biocatalysis and bioremediation. He leads a multi-disciplined biology group encompassing molecular biology, molecular modelling, rational mutagenesis and directed evolution. He also has experience in the exploitation of wild type microbes to carry out novel biotransformations of synthetic and natural compounds and translating these from research laboratory to industrial manufacturing. E: derek.quinn@almacgroup.com

Dr Jane Mueller is Technical Lead within Almac's Biocatalysis group. After obtaining a degree in Chemistry Jane finished her PhD in 2011 at the Technical University Berlin working on nonribosomal peptides for production of novel APIs. She is one of many international staff, enriching Almac's scientific community. Jane's career has encompassed many facets of Biocatalysis from metabolite ID, enzyme screening and engineering to scale up and production. She has carried out many projects, which went from initial screening hit to mg, g and kg scale since she joined Almac's Biocatalysis group in 2012 with expertise in Organic Chemistry as well as Analytics and Biotechnology. E: jane.mueller@almacgroup.com

Dr Alexandra Carvalho has a PhD in Computational Chemistry from Porto University. She did post-doctoral training at Girona University as a Beatriu de Pinós Marie Curie fellow and at Uppsala University as Senior researcher. In 2017, she founded the Rational Protein Engineering group at the Centre for Neurosciences and Cell Biology. Alexandra has expertise in computational chemistry methods applied to the study and engineering of enzymes, namely Quantum Mechanics/Molecular Mechanics simulations and enhanced sampling methods. E: alexandra.carvalho@almacgroup.com