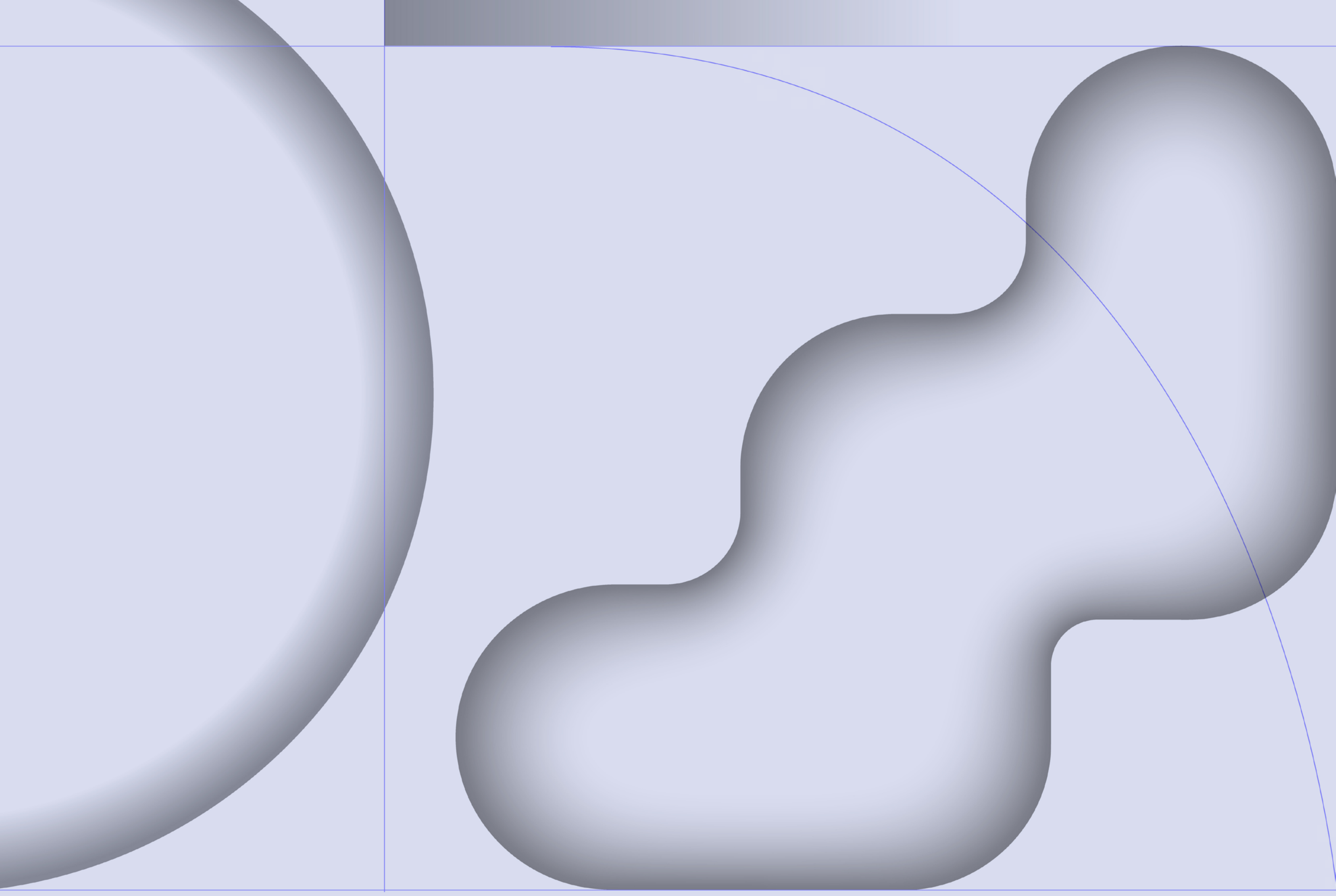


Rapid Discovery and Validation of Substrate-Specific Ketoreductases for API synthesis

Application note

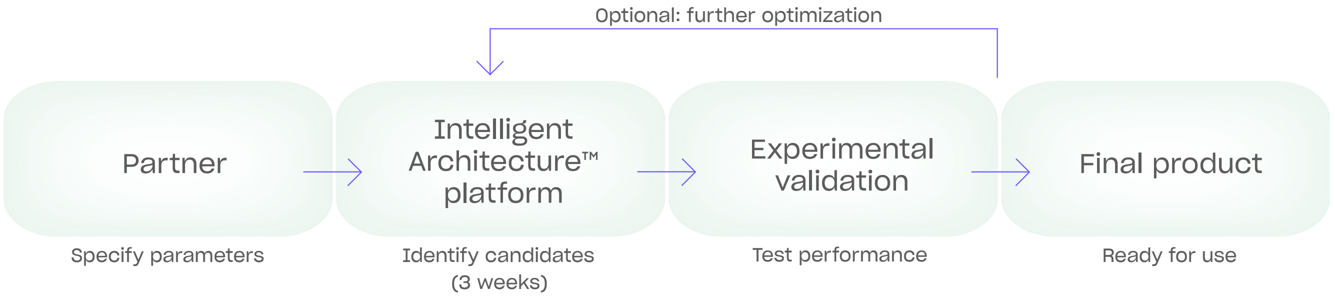
BM-AN0002-RevA

biomatter

Rapid Discovery and Validation of Substrate-Specific Ketoreductases for API synthesis

Summary

- Computational enzyme identification can provide significant time savings for custom enzyme development compared to traditional methods.
- The Intelligent Architecture™ platform is capable of screening millions of sequences in a matter of weeks and selecting a small number of candidates with a high success rate.
- The platform can identify novel enzymes with highly specific functions, yet diverse sequences and structures for an expanded range of secondary properties.
- Applying the Intelligent Architecture™ platform to the discovery of novel substrate-specific ketoreductases resulted in the identification of 9 candidates and validation of 4 active enzymes in just 3 weeks.


Optional: further optimization
Partner
Specify parameters
Intelligent Architecture™ platform
Identify candidates (3 weeks)
Experimental validation
Test performance
Final product
Ready for use

I. Enzyme discovery for API manufacturing

Enzymes are sophisticated biological machines with applications that span many areas, from food and beverage to the manufacturing of active pharmaceutical ingredients (APIs)[1,2,3]. With an estimated market value of $2.4 billion in 2022[4], enzymes are indispensable to the pharmaceutical industry. They provide unrivaled advantages to chemical catalysts, such as the ability to enable shorter and more efficient API synthesis routes, activity under mild reaction conditions, high stereoselectivity, and reduction of waste in the production process[5]. Ketoreductases are among the most widely used enzymes in API synthesis[6,7]. These biocatalysts reduce ketones to stereospecific secondary alcohols, a structural molecule class present in many bioactive molecules, fine chemicals, and pharmaceuticals, including blockbuster drugs such as montelukast, atorvastatin or ezetimibe[8]. As such, they are an active area of R&D for pharmaceutical companies looking to develop novel efficient API synthesis routes and secure their position within a competitive IP landscape.

The big challenge in the industry is identification of enzymes with the right substrate specificity, stereoselectivity, desired co-factor preference, significant activity, and stability under process conditions. Obtaining an enzyme with the desired characteristics can be done by either searching for an existing variant that checks all the requirements or engineering a custom enzyme. Regardless of the approach, the process can take many months. Searching through genomic and metagenomic libraries can provide a good starting point. However, comprehensive collections are not readily available and require a lot of effort to collect, annotate, and maintain. Initial identification of enzyme candidates requires the resource-intensive development of high-throughput screening methods or selection strategies to narrow down the pool of candidates. Specific enzyme features can also be obtained through experimental approaches, such as directed evolution or rational engineering, but both methods require significant time, resources, and deep expertise in enzyme engineering.

Our goal is to simplify the process of custom enzyme development by narrowing down the candidate pool through advanced computational strategies.

The Intelligent Architecture™ platform was created to address the limitations of existing enzyme engineering approaches and enable faster discovery, reduce the number of variants to screen, and ensure a high hit rate during validation. By using multi-stage sequence, structure, and function-based filtering, the platform can search through over a million variants in a matter of weeks and reliably identify enzymes with specified target characteristics. To demonstrate the capabilities of the platform, we have used it to identify a ketoreductase capable of reducing the Wieland-Miescher ketone (WMK), a common building block of API product routes. Our *in silico* approach resulted in the identification of 9 candidates and validation of 4 active enzymes with the desired substrate specificity in **only 3 weeks**, demonstrating the power of the Intelligent Architecture™ platform for enzyme discovery applications.

II. Computational identification of enzyme candidates

The Wieland-Miescher ketone is a common building block for terpenes and steroids which has been employed in the total synthesis of more than 50 natural products and identified as an intermediate in the retrosynthesis of promising APIs with anti-cancer, antimicrobial, antiviral, anti-neurodegenerative, and immunomodulatory activities[9]. To identify ketoreductases capable of reducing the WMK substrate to 5-hydroxy-4a-methyl-4,4a,5,6,7,8-hexahydronaphthalen-2(3H)-one (**Figure 1**), we employed the Intelligent Architecture™ platform. We started with a database of 50 million sequences and used domain profile and fold-level searches to narrow down the enzyme search space to 1.5 million enzymes (**Figure 2a**). The sequences were grouped by similarity and colored based on the probability of belonging to a particular EC reaction class, with the unassigned (gray) proteins falling into the vast "unannotated" category[10]. Using the current state-of-the-art methods, it is difficult to tease out the function or properties of such sequences; however, with the Intelligent Architecture™ platform we are able to consider enzyme variants that are outside of the predicted enzyme reaction class, significantly broadening the diversity of identified candidates.

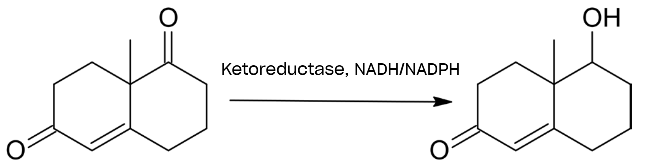


Figure 1. Schematic representation of the target enzyme reaction.

During the subsequent stages of the filtering process (Figure 2b) we applied multiple levels of structural filtering to reduce the number of candidates from 20,000 sequences to approximately 800. Typically, such a deep analysis may take days to perform for each individual enzyme variant, but the Intelligent Architecture™ platform can do this on millions of sequences in just a few weeks. At the final selection stage, a small number of sequences is carefully chosen for experimental testing based on the scores provided by the platform.

As a result of this effort, we identified 9 enzymes potentially capable of reducing WMK. Among the selected variants, the sequence similarity ranged from 12.7% to 29.6%. The variants also displayed different structural motifs, including TIM barrel and Rossmann fold (Figure 2c). The high sequence diversity among the selected candidates improves the ability to find variants with suitable organism expression, solubility, and co-factor preference or select other desirable secondary characteristics (such as susceptibility to inhibitors, stability, and resistance to organic solvents). The identified sequences were all from the unannotated category, underscoring the importance of expanding the search space beyond the specified enzyme reaction class. The nearest annotated homologs, sharing 36-45% sequence identity, belonged to xylose reductase (EC:1.1.1.307) or oxoacyl reductase (EC 1.1.1.100) classes (**Table 1**). This highlights the

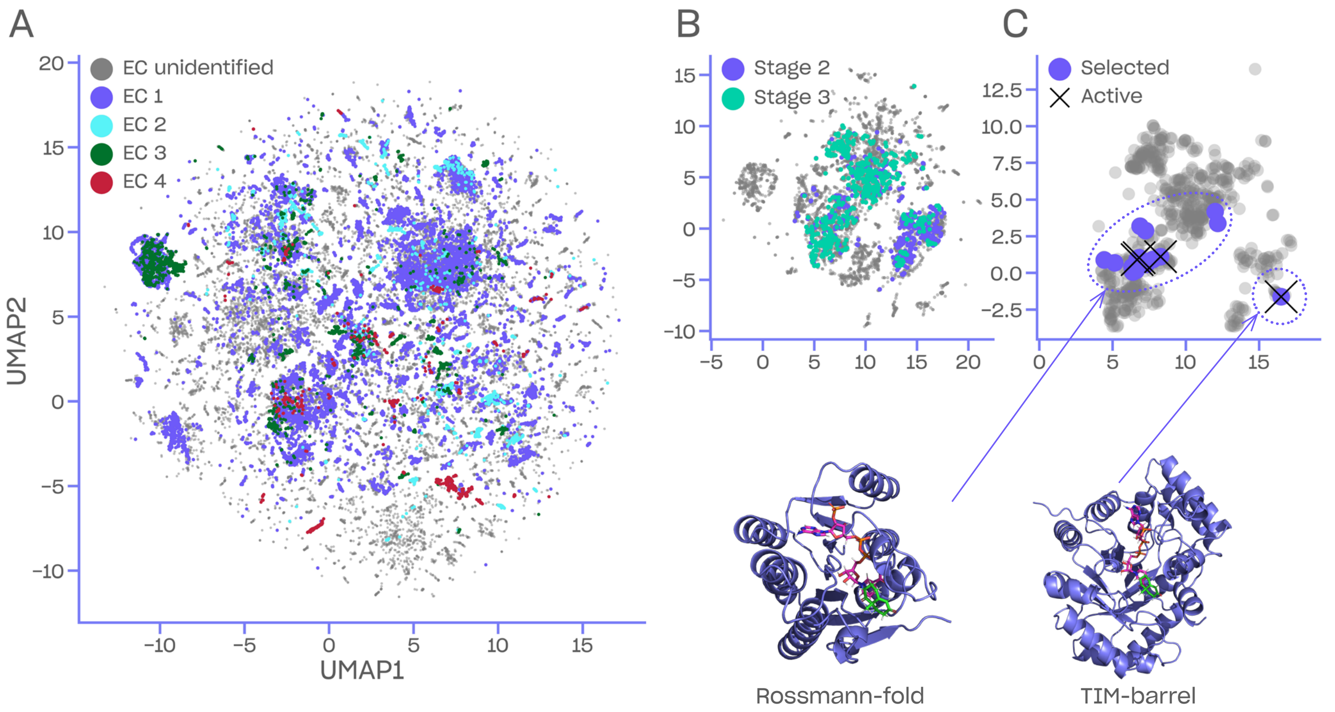


Figure 2. he Intelligent Architecture™ computational search process: (**A**) The initial sequence space of 50 million sequences is narrowed down to 1.5 million using domain profile and fold-level searches. Selected enzyme candidates are grouped by sequence similarity and colored according to EC class. (**B**) Further refinement of the search space down to ~20,000 sequences using structure-level filters. (**C**) The final 9 sequences selected for experimental testing were chosen from a group of 800 candidates based on the platform's scoring capabilities. The panel below shows the different structural motifs displayed by the selected finalists. All data points were plotted using UMAP projections of protein sequence ESM2 embeddings at different search stages.

advantage of the Intelligent Architecture™ platform over traditional search methods which may overlook potentially viable candidates that fall outside of the specified reaction class.

III. Experimental validation of selected candidates

Following the identification of 9 enzyme candidates via computational search, those enzyme sequences were synthesized and expressed in *E. coli*. All proteins were successfully purified using his-tags in sufficient quantities for experimental testing, although two had solubility lower than 50%. To test the enzyme activity, purified proteins were incubated for 16 hours with racemic WMK substrate and NADH or NADPH co-factors in the reaction buffer. Ketone reduction was monitored by measuring absorbance at 340 nm. Proteins able to generate OD reduction of at least 1 unit were identified as active (**Figure 3a**). This stringent filter was used to exclude artifacts and enzymes with minimal activity.

The initial test identified 4 active proteins displaying ketoreductase activity. Enzymes KR-3 and KR-4 had a strong preference for NADH as a co-factor, KR-6 strongly preferred NADPH, and KR-1 showed flexible co-factor utilization, with a preference for NADPH. The specific activity for active proteins was measured in triplicate using the preferred co-factor

Enzyme	Species	Fold	Nearest annotated homolog EC class
KR-1	Candida sphaerica	TIM barrel	EC:1.1.1.307 D-xylose reductase [NAD(P)H]
KR-3	Mycobacterium sp.	Rossmann fold	EC:1.1.1.100 3-oxoacyl-[acyl-carrier-protein] reductase
KR-4	Firmicutes sp.	Rossmann fold	EC:1.1.1.100 3-oxoacyl-[acyl-carrier-protein] reductase
KR-6	Sorangiineae sp.	Rossmann fold	EC:1.1.1.100 3-oxoacyl-[acyl-carrier-protein] reductase

Table 1. Identified active enzymes and their attributes.

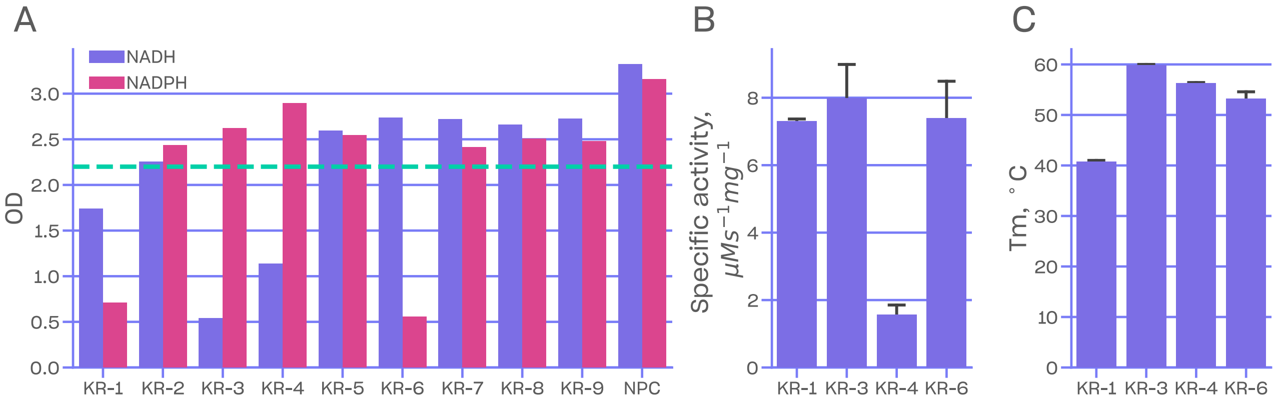


Figure 3. Experimental testing of computationally identified enzyme candidates: (**A**) Activity assay using racemic Wieland-Miescher ketone (WMK) as substrate and NADH or NADPH as co-factors. The reaction was monitored at 340 nm over 16 hours. Enzymes that decreased the OD by at least 1 unit were selected as active. NPC – no protein negative control. (**B**) Specific activity of active enzymes for the WMK substrate, using the preferred co-factor. Specific activities were determined based on triplicate measurements. (**C**) Protein melting temperature of identified active enzymes.

for each screened enzyme. The reaction was monitored by measuring absorbance at 340 nm over time. Specific activity was calculated from the linear range when no more than 10% of the substrate has been consumed (**Figure 3b**).

In order to ensure that the enzymes are stable for downstream processing and applications, enzyme stability was also evaluated for the selected variants. The protein melting point for active proteins was measured in duplicate using GloMelt™ Thermal Shift Protein Stability Kit (**Figure 3c**). Three of the four proteins had a melting temperature above 50°C, which typically correlates with good enzyme stability in practical applications.

IV. Discussion

The example of identifying novel substrate-specific and highly active ketoreductases presented here demonstrates the effectiveness of the Intelligent Architecture™ platform as a computational enzyme discovery tool. The platform is capable of analyzing millions of protein sequences from public and proprietary databases much faster than traditional computational approaches. The iterative filtering process allows us to significantly narrow down the search space, while preserving high sequence diversity, by using structure- and function-based filter parameters such as active site shape, enzyme size, target substrate and product structures, reaction mechanism, co-factor preference, as well as other features. Typically, analyzing at this depth and resolution may take days for each enzyme variant, but the Intelligent Architecture™ platform performs it on millions of variants within a few weeks.

Leveraging the respective strength of data-driven AI and physics-based approaches enables us to carefully select a very small number of candidates for experimental testing with a high probability of having the intended function, which can save months of development time. In this case, we were able to narrow down the candidate pool to just 9 enzyme sequences and find 4 active enzymes that showed specific activity towards the substrate. These enzymes were highly diverse in terms of sequences and structures, providing a wide range of secondary properties such as co-factor preference and stability. One of the notable features of the Intelligent Architecture™ platform is that it allows for the identification of enzyme sequences that had not been previously characterized and are difficult to predict i*n silico*. This translates into potential implications for staking the competitive enzyme IP space.

Overall, our computational enzyme discovery approach can help enzyme developers save significant time and testing resources. The discovery segment of the Intelligent Architecture™ platform can be additionally combined with optimization and design segments to fine-tune the desired enzyme properties. Finally, the *in silico* enzyme development can be supported by experimental validation at our laboratories. With enzyme identification made easy, Biomatter eliminates a significant bottleneck in biopharma process development, facilitating API discovery and manufacturing.

References

1. From enzyme discovery to special applications. Biotechnol. Adv. 40, 107520 (2020).

2. Adams, J. P., Brown, M. J. B., Diaz-Rodriguez, A., Lloyd, R. C. & Roiban, G.-D. Biocatalysis: A Pharma Perspective. Adv. Synth. Catal. 361, 2421–2432 (2019).

3. Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. Nat. Rev. Chem. 2, 409–421 (2018).

4. MarketsandMarkets Research, Ltd. Enzymes Market worth $16.9 billion by 2027. GlobeNewswire News Room: https://www.globenewswire.com/en/news-release/2022/10/11/2531530/0/en/Enzymes-Market-worth-16-9-billion-by-2027.html (2022).

5. Simić, S. et al. Shortening Synthetic Routes to Small Molecule Active Pharmaceutical Ingredients Employing Biocatalytic Methods. Chem. Rev. 122, 1052–1126 (2022).

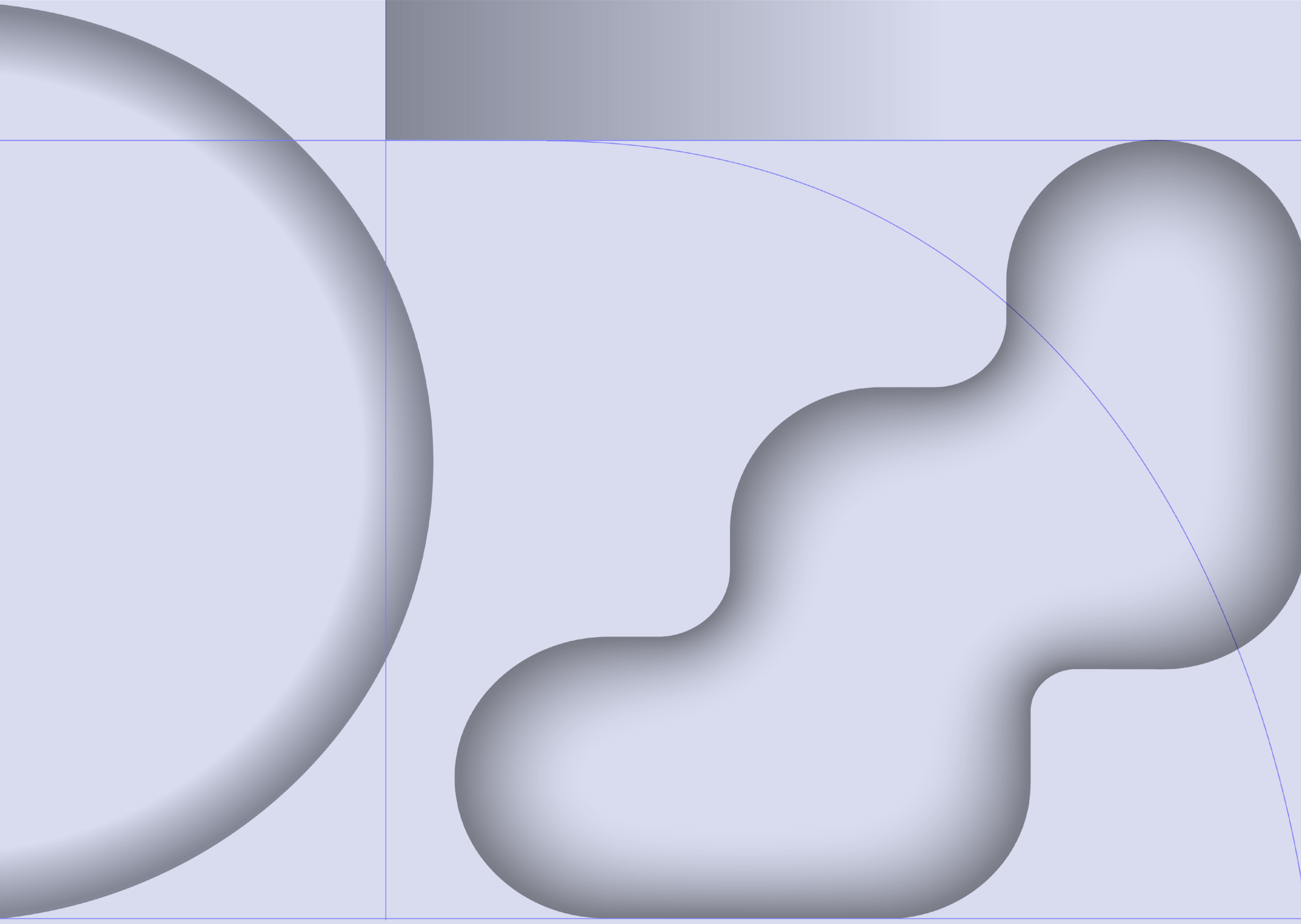
6. Lalonde, J. Highly engineered biocatalysts for efficient small molecule pharmaceutical synthesis. Curr. Opin. Biotechnol. 42, 152–158 (2016).

7. Zheng, Y.-G. et al. Recent advances in biotechnological applications of alcohol dehydrogenases. Appl. Microbiol. Biotechnol. 101, 987–1001 (2017).

8. Huisman, G. W. & Collier, S. J. On the development of new biocatalytic processes for practical pharmaceutical synthesis. Curr. Opin. Chem. Biol. 17, 284–292 (2013).

9. Bradshaw, B. & Bonjoch, J. The Wieland-Miescher Ketone: A Journey from Organocatalysis to Natural Product Synthesis. Synlett 2012, 337–356 (2012).

10. de Crécy-lagard, V. et al. A roadmap for the functional annotation of protein families: a community perspective. Database 2022, baac062 (2022).



Thank you for reading our application note.

BM-AN0002-RevA

Let’s explore partnership opportunities!

partnering@biomatter.ai

biomatter