The Sabatier principle as a tool for discovery and engineering of industrial enzymes

Kari, Jeppe; Schaller, Kay S.; Molina, Gustavo A; Borch, Kim; Westh, Peter

Published in:
Current Opinion in Biotechnology

DOI:
10.1016/j.copbio.2022.102843

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Kari, J., Schaller, K. S., Molina, G. A., Borch, K., & Westh, P. (2022). The Sabatier principle as a tool for discovery and engineering of industrial enzymes. *Current Opinion in Biotechnology*, *78*, [102843]. https://doi.org/10.1016/j.copbio.2022.102843

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

The Sabatier principle as a tool for discovery and engineering of industrial enzymes

Jeppe Kari[1], Kay Schaller[2,3], Gustavo A Molina[2,4], Kim Borch[5] and Peter Westh[2]

The recent breakthrough in all-atom, protein structure prediction opens new avenues for a range of computational approaches in enzyme design. These new approaches could become instrumental for the development of technical biocatalysts, and hence our transition toward more sustainable industries. Here, we discuss one approach, which is well-known within inorganic catalysis, but essentially unexploited in biotechnology. Specifically, we review examples of linear free-energy relationships (LFERs) for enzyme reactions and discuss how LFERs and the associated Sabatier Principle may be implemented in algorithms that estimate kinetic parameters and enzyme performance based on model structures.

Addresses
[1] Roskilde University, Dept. Science and Environment, Universitetsvej 1, DK-4000 Roskilde, Denmark
[2] Technical University of Denmark, Dept. of Biotechnology and Biomedicine, Sølvtofts Plads 224, DK-2800, Kgs. Lyngby, Denmark
[3] University of Copenhagen, Dept. of Drug Design and Pharmacology, Universitetsparken 2, DK-2100 Copenhagen, Denmark
[4] Technical University of Denmark, The Novo Nordisk Foundation Center for Biosustainability, Build. 220, Kemitorvet, DK-2800, Kgs. Lyngby, Denmark
[5] Novozymes A/S, Biologiens vej 2, DK-2800, Kgs. Lyngby, Denmark

Corresponding author: Peter Westh (petwe@dtu.dk)

Abbreviations: LFER, Linear free energy relationship; SP, Sabatier Principle

Current Opinion in Biotechnology 2022, **78**:102843

This review comes from a themed issue on **Chemical Biotechnology**

Edited by **Bernd Nidetzky** and **Byung-Gee Kim**

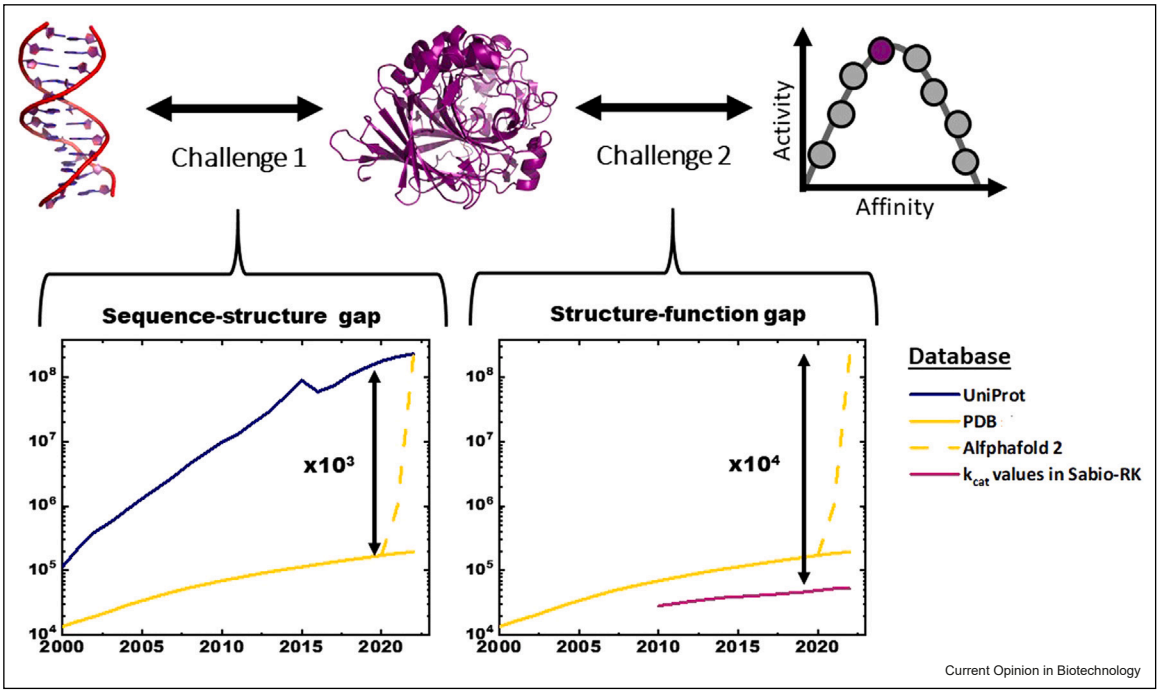
https://doi.org/10.1016/j.copbio.2022.102843

Introduction

Extended use of biocatalysis is one of the main avenues toward more sustainable industries. This may involve the use of free- or immobilized enzymes produced separately or the application of whole-cell catalysts. In either case, one major challenge is to find enzymes with the stability, specificity (or promiscuity), and catalytic rate that are required under industrial conditions. The challenge of finding adequate enzymes is inescapably rooted in the astronomical sequence space of protein molecules, but recent advances hold promise for faster and more effective methods to search for biocatalysts. One major leap forward comes from the success of AlphaFold [1] and other approaches [2] in rapidly generating reliable model structures from protein sequences. These tools can bring our knowledge on protein tertiary structure on par with the sequence information, and hence help closing the so-called sequence-structure gap, which has increased to around 10^3 known sequences for every known structure (Figure 1).

The emerging affluence of structural data will undoubtedly become vital for future enzyme design, but routes to harness this potential are not trivial. Thus, a static 3D structure only provides overall clues about function, mechanism, and kinetics [3]. Other innovative inputs will be required to effectively exploit structural information for *in silico*-guided enzyme design. In essence, the challenge (labeled '2' in Figure 1) is now to close the gap between structure on one hand and specificity, mechanism, and kinetics on the other. One approach to this is classical structure-based molecular modeling [4•], which obviously becomes much more powerful with an almost endless supply of all-atom model structures. Perhaps more importantly, the surge in structural information can spur progress in enzyme design supported by statistical profiling and machine learning (ML) [5–9]. The efficiency of these approaches depends on the way the structural data are incorporated into the design algorithm, and a wide range of principles for this are currently being tested [7•]. One principle, which was recently applied successfully for the design of improved plastic-degrading enzymes [10••], assesses typical features of the local microenvironments of amino acids based on training against thousands of known structures. Once trained, the algorithm can identify positions that have an unbalanced local environment and hence appear as good targets for engineering. Other workers have used more physics-based principles, including accessible surface area, logP, pK_a, and nucleophilicity of catalytic residues in design algorithms [11]. Most work in this area has targeted enzyme selectivity

Figure 1



Two main challenges for computer-assisted discovery and engineering of industrial enzymes. The first involves fast and reliable modeling of enzyme structure based on the sequence. The second is the computation of phenotypical traits from the (model) structure. For decades, the sequence libraries have outnumbered structure libraries and the retio is now a factor of around 10^3, but recent progress in modeling all-atom structures is about to close this gap (left panel). This opened new opportunities for coming work that addresses the even larger structure-function gap (right panel).

and the range of substrates that can be modified, while less attention has gone into improving catalytic performance [12•]. However, improvements in catalytic performance under industrial conditions appear important as technical gauges such as catalytic space–time yield, are typically low for biocatalysts compared with inorganic catalysis [12•].

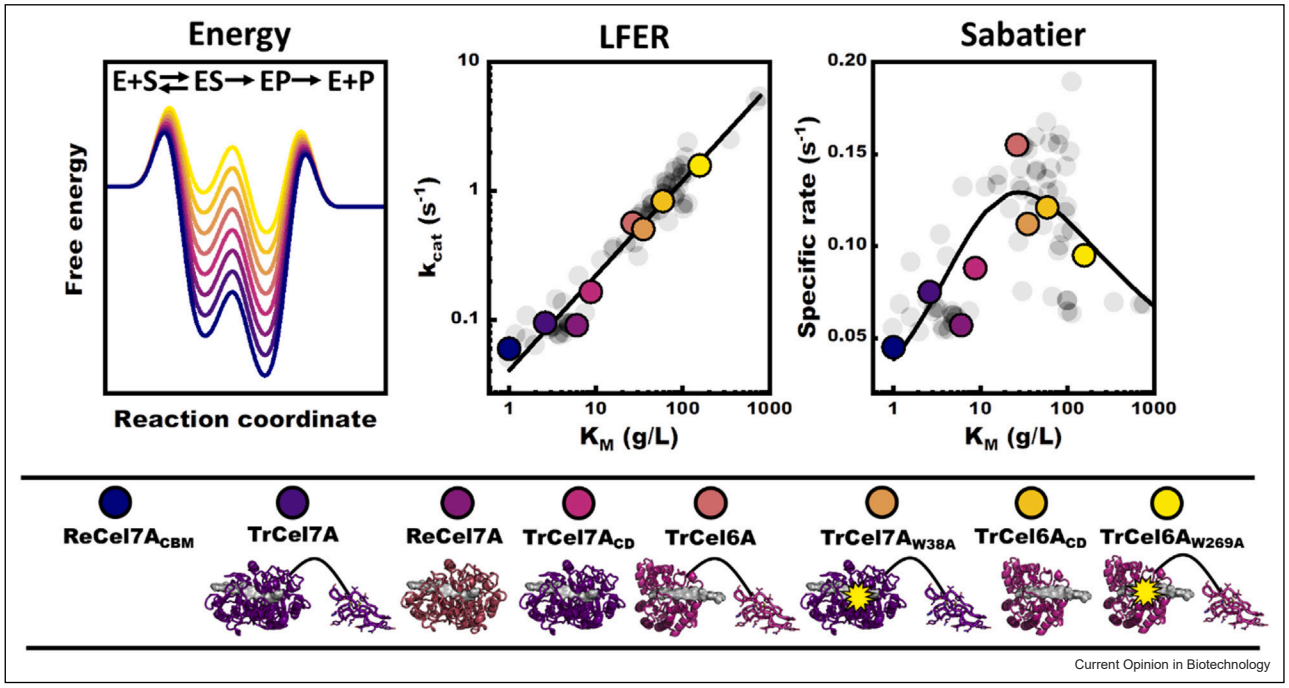
Here, we discuss a physics-based principle, which may be implemented for *in silico* screening of enzyme kinetic parameters. Specifically, we review literature that links substrate-binding strength and turnover rates for isofunctional enzymes. This type of linkage occurs when there is scaling between the free energy of binding and the activation barrier (Figure 2), and it holds a potential to significantly limit the search space. Scaling of binding- and activation-free energies is reflected qualitatively in the Sabatier Principle (SP) and quantitatively in so-called linear free-energy relationships (LFERs). Both SP and LFERs constitute cornerstones within basic and applied (nonbiochemical) catalysis, but are only considered sporadically in biochemistry and biotechnology. We argue that the current state of affairs with dramatic increments in the number and quality of protein model structures provides an excellent foundation for the implementation of SP and LFER-based algorithms in the design and discovery of industrial enzymes.

Enzyme reactions and the Sabatier Principle

Enzymatic turnover rates are notoriously difficult to assess *in silico*. Rigorous approaches to this problem involve high-level theory and extended computation time [13], and hence remain unrealistic for larger sequence spaces. Machining learning-based approaches to turnover numbers have emerged [14], but it has proven challenging to identify patterns in catalytic rates based on sequence- or structure data. This latter difficulty reflects at least in part that the catalytic rate is globally encoded and hence sensitive to the entire sequence [5], but it could also result from functional effects of cofactors, reaction conditions, conformation of polymeric substrates, and other factors that are not represented directly in the sequence [6]. In light of this complexity, incorporation of basic physical constraints could be useful in structure-kinetics algorithms.

The century-old SP states that catalysis is most effective when catalyst-reactant interactions are of intermediate strength. This is because weak binding (yellow lines in Figure 2) leads to an insufficient population of

Figure 2

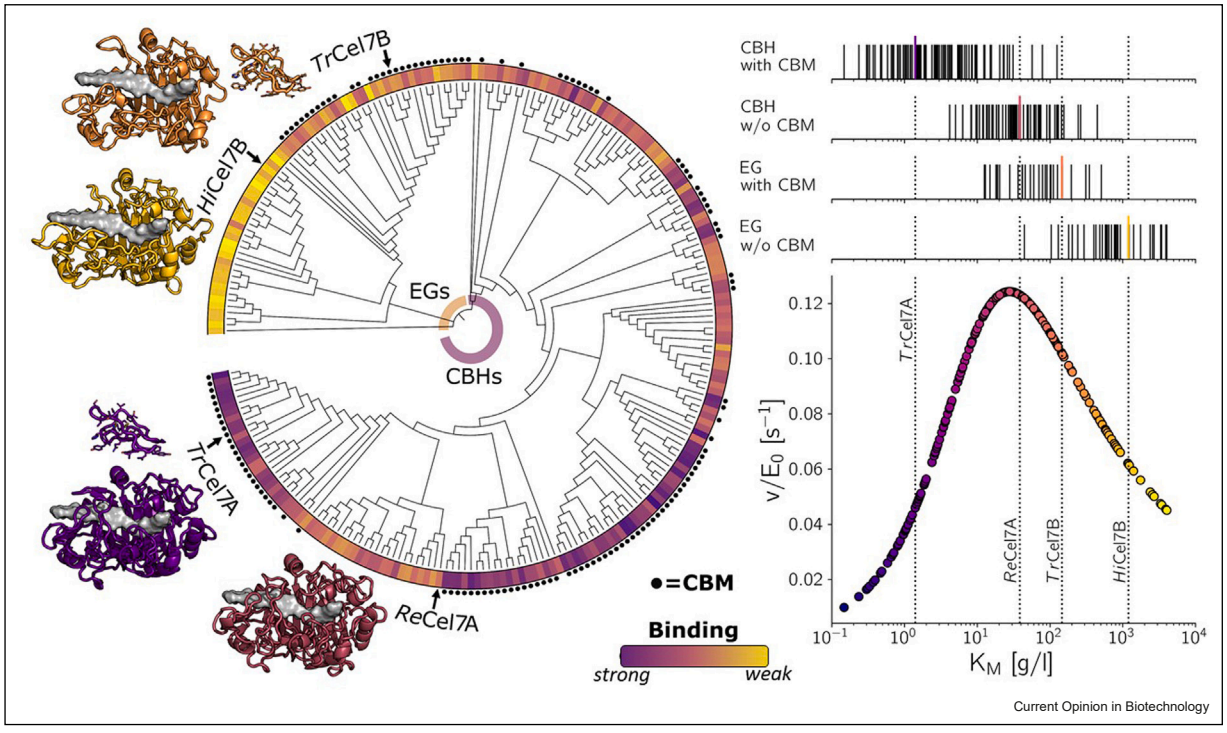


Schematic illustration of relationships between energy diagrams (left), scaling relationships (middle), and the SP illustrated in a volcano plot (right). This example is based on data for eight fungal cellulases (wild types and variants) identified below the plots. (Re: Rasamsonia emersonii, Tr: Trichoderma reesei, CD: catalytic domain, i.e. enzyme variants where the binding domain has been truncated). The strength of substrate interaction decreases from blue to yellow, and this is illustrated in the schematic energy diagrams in the left panel. Here, strong binders have deep energy wells for the ES and EP intermediates, and hence experience high activation barriers for EP dissociation. The diagrams show that the energy of different intermediates and transition states is not independent, and this is illustrated more directly by the LFER in the middle panel. This plot presents experimental data for the enzymes considered here (colored symbols) and many other cellulases (gray symbols), and it identifies a linear relationship between binding energy (approximated by $\ln K_M$) and activation energy (approximated by $\ln k_{cat}$). The substrate in these measurements was microcrystalline cellulose. Enzymes to the bottom/left of the LFER have tight binding and slow turnover at saturation, while those to the top/right have weak binding and fast turnover at saturation. One direct corollary of this is that the specific activity at a fixed substrate load (subsaturation) has a biphasic behavior as shown in the right panel. This "volcano" reflects accumulation of intermediates and hence slow reaction of strong binders to left side of the apex. Conversely, weak binders to the right of the apex perform poorly inspite of their high k_{cat} because too little complex accumulates. This is a restatement of the SP.

intermediate, while strong interactions (blue lines in left panel of Figure 2) delay the catalytic cycle through accumulation of stable intermediates in deep energy wells. A similar line of thinking underpins fundamental descriptions of enzyme function [15], but is rarely linked explicitly to the SP. Practical application of the SP often involves plots with data from a family of related catalysts. The abscissa is some measure of catalyst-substrate interaction strength and the ordinate represents the catalytic rate under a fixed set of experimental conditions. According to the SP, such plots will have a maximum at an intermediate interaction strength, and they have hence been dubbed 'volcano plots' (Figure 2, right panel). A volcano curve arises if the free energies of intermediates and transition states are not independent, but shift commensurately for different catalysts [16] as exemplified schematically in Figure 2 (left panel). This mutual interdependence of energy levels can also be expressed as a linear scaling between binding- and activation energies, that is, an LFER. This is illustrated in the middle panel of Figure 2, and discussed in some more detail in the figure legend. Some influential examples of LFERs include Brønsted- and Hammet plots, which are widely used to elucidate rates, mechanisms, and regioselectivity in organic chemistry [17]. Within inorganic catalysis, LFERs have proven useful both in attempts to rationalize comparative data and as a tool to engineer improved catalysts. Success in this area has been ascribed to the remarkable simplification, which occurs when only one or a few intermediates need to be assessed computationally [18].

Direct experimental evidence for LFERs or SP behavior for enzyme reactions remains sparse. The most prominent example is still the seminal work by Fersht et al., who used experimental LFERs for a series of tyrosyl–tRNA-synthetase variants to elucidate interactions in the transition state [19,20]. A similar approach has been used to study the nature of the transition state, when one enzyme acts on different, chemically related

Figure 3



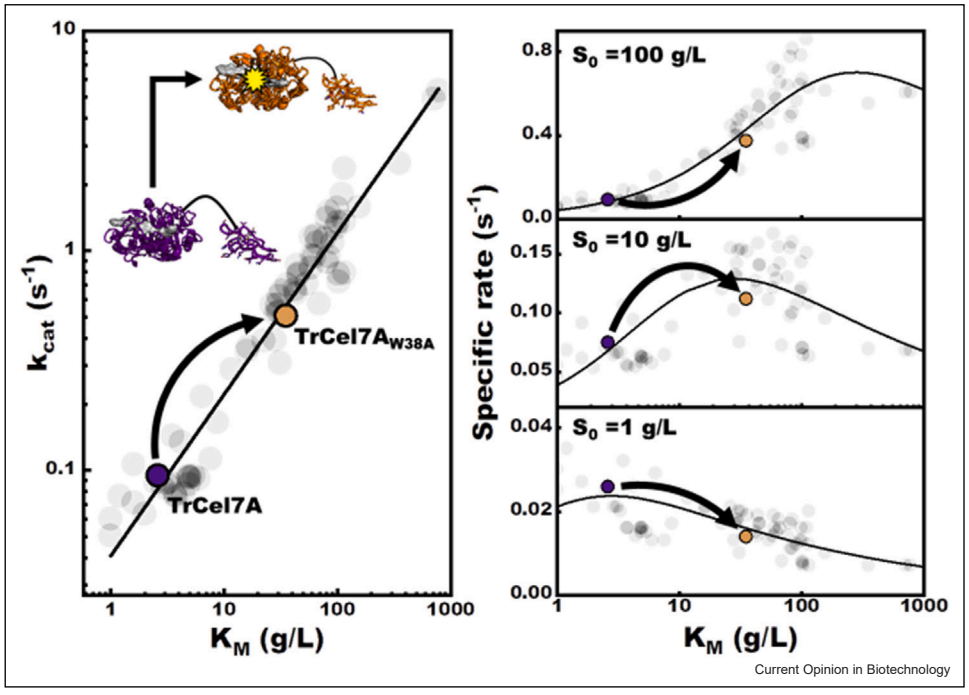
SP as a tool for enzyme discovery. Substrate-binding strengths were computed for all cellulases listed in GH7 and indicated by the color code in the rim of the phylogenetic tree [29]. The investigated enzymes are either endoglucanases (EG) (EC 3.2.1.4) or cellobiohydrolases (CBH) (EC 3.2.1.91) as specified in the core of the tree. Enzymes that have a CBM (see Figure 2) are identified by a black dot at the outer rim. The rug plots ('bar codes') in the upper right reveal relationships between binding strength, mode of action (EG or CBH), and the presence of a CBM. The computed binding strengths were first converted to kinetic parameters using an experimental LFER [30]. Second, these parameters were used to estimate hydrolytic rates for these previously uncharacterized enzymes on 10 g/l microcrystalline cellulose (lower-right panel). The volcano identifies a number of enzymes (e.g. Cel7A from *R. emersonii*) near the apex that are likely candidates for good performance under these conditions. We note that the location of the apex depends on the experimental conditions, and it will, for example, shift toward higher optimal K_M values for higher substrate loads (c.f. Figure 4).

substrates [21]. More recently, LFERs and volcano plots have been used to rationalize the turnover of a synthetic DNA-enzyme nanostructure [22], as well as the hydrolysis of insoluble substrates such as cellulose and polyester [23,24]. In addition, some studies have reported kinetic data for the modification of different insoluble substrates, which appear consistent with an LFER although this was not explicitly pointed out by the authors [25–27].

Sabatier Principle in enzyme discovery and -engineering

The power of the SP lies in its potential to simplify the selection of enzymes with desired kinetic properties. Thus, if a characteristic binding strength is indicative of good turnover at a given set of conditions, the immense computational endeavor of assessing energies of intermediates and activation barriers can be replaced by calculations of substrate-binding strength. While computational determination of ligand-binding strength is far from trivial, numerous fast methods have been developed primarily for drug discovery [28], and some of these can probably be applied to enzyme-substrate interactions with minor adaptations. This was illustrated in a study on cellulases [29], which used a previously established LFER [30] between substrate-binding strength and maximal turnover to guide enzyme discovery. Specifically, kinetic parameters for all enzymes belonging to Glycoside Hydrolase Family 7 (GH7), were estimated from their sequences in two computational steps [29]. The first step involved determination of a model structure (i.e. challenge 1 in Figure 1), and the second was estimation of interaction strength based on very short molecular dynamics simulations and a Linear Free Energy [31] (LIE) method, which was adapted to account for the multidomain structure of the enzymes and the solid nature of the substrate. With the experimental scaling curve [30] and computed binding strengths [29] at hand, the maximal turnover for uncharacterized enzymes could be read off the LFER (c.f. middle panel in Figure 2). Hence, the LFER served to resolve challenge 2 in Figure 1, and the catalytic performance of the enzymes could be estimated as illustrated in Figure 3. In this example, the throughput was limited by the speed of computing enzyme-substrate-binding strength, and one way to speed up this part is to use machine learning (ML) instead of explicit simulations. This can be done by training an ML algorithm against

Figure 4



SP as a tool in enzyme engineering. The left panel shows an experimental LFER for cellulases, which highlights the effect of the mutation W38A in TrCel7A. It appears that this point mutation changes the kinetics to an extent that corresponds to almost half the functional breadth found for a large group of structurally and mechanistically different enzymes. Interestingly, the mutant does not show independent changes in kinetic parameters, but shifts along the LFER. Such changes in catalytic properties, which relies on small alterations in the sequence, are readily described by LFER-based analyses and this may be used in engineering campaigns. It appears, for example, from the volcano plots (right panels) that this particular mutation might benefit enzyme performance at intermediate and high substrate loads while it causes adverse effects on activity at low loads.

experimental data or rigorously computed interaction strengths. An initial attempt of using this strategy appeared recently and showed promising results with respect to throughput and reliability [32].

Algorithms that are supported by LFERs may also have potential within enzyme engineering. As in enzyme discovery, the advantage lies in prediction and sorting of variant kinetics. This could be implemented with a strategy akin to the one in Figure 3, using computed binding strength or interaction parameters derived by ML for putative variants instead of wild-type sequences. An example of this idea is illustrated in Figure 4.

Concluding remarks

LFERs and the associated SP are well established within organic chemistry and nonbiochemical catalysis, and have proven useful both for fundamental and applied problems within these areas. Further use within computational biocatalysis appears promising but will require much better insights into the nature and prevalence of LFERs for enzyme reactions. Some results (discussed above) support the incidence of LFERs for different hydrolases acting on insoluble substrates. Examples of LFERs for soluble substrates have also appeared, but broader approaches such as meta-analyses of kinetic parameters from databases have not found clear relationships, for example, between K_M (a descriptor of interaction strength) and k_{cat} (a descriptor of activation energy) for enzymes acting on their native, soluble substrate [33,34]. This could be because K_M is a poor descriptor of substrate affinity or scaling only occurs within groups of related, isofunctional enzymes, but there are also challenges related to noise in databases and poor lab-to-lab reproducibility of kinetic parameters [14,35,36]. This calls for further systematic and self-consistent biochemical work on enzyme libraries, which are expressed, purified, and characterized under the same conditions. While this line of work may appear archaic in the view of some, it is our opinion that self-consistent biochemical data are in short supply and that this makes up a bottleneck for progress within computational design of industrial enzymes. This is particularly the case for insoluble substrates, which are common in industry but not covered at all by current databases. Fortunately, recent progress in the automation of gene manipulation, heterologous expression, protein purification, and activity assays has paved the way for comparative biochemistry at a 'medium-throughput' level, and if indeed LFERs can be established from this type of work, they can be applied in different ways. One obvious, but also quite elaborate, approach is to analyze computed substrate-

binding strengths against an experimental LFER as in the example in Figure 2. A noteworthy aspect of this method is that the optimal binding strength shifts with experimental conditions such as temperature and the load- and physical properties of an insoluble substrate. This dependence obviously comes with the drawback that a new experimental LFER must be established for every condition of interest. On the other hand, the sensitivity to both catalyst and conditions provides the advantage that the approach may be applied within both enzyme- and process engineering. We also envision that linear scaling can be incorporated in structure-kinetics algorithms, in cases where an experimental LFER is not available as a 'standard curve'. Thus, if a reaction is indeed limited by a linear scaling, it can be used for a coarse sorting of phenotypes, even when slope and intercept are not known. This less work-intensive strategy can be used as an early, *in silico* phase of engineering campaigns that would serve to unveil relative kinetic traits within a library of enzyme variants. This would be valuable, for example, in defining an initial group of enzymes with a suitable phenotypical breadth for subsequent engineering, and we note that this is not readily done on the basis of sequence alone. This is emphasized in Figure 4, which shows how a point mutation quite far from the catalytic residues in a cellulase may evoke large phenotypical changes. In fact, the change corresponded to about half of the functional breath defined by k_{cat} and K_M for over 80 other structurally and functionally diverse cellulases. Interestingly, k_{cat} and K_M did not change independently, but shifted along the LFER, and a similar behavior with shifts up or down the LFER has been found for several other enzyme variants [30]. We posit that LFER-based algorithms are suited to detect and rationalize kinetic changes in enzyme variants that are caused by small alterations in the sequence, and that incorporation of scaling principles in ML algorithms such as those discussed above, could present a powerful tool in virtual bioprospecting of industrial enzymes.

CRediT authorship contribution statement

Jeppe Kari: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing; **Kay Schaller:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing; **Gustavo Molina:** Conceptualization, Data curation, Investigation. Methodology, Writing – review & editing; **Kim Borch:** Conceptualization, Funding acquisition, Supervision; **Peter Westh:** Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – original draft.

Conflict of interest statement

Kim Borch works for Novozymes, a major provider of industrial enzymes. The other authors have no conflicts of interest.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the Novo Nordisk Foundation (grant NNFSA170028392 to PW), Independent Research Fund Denmark (grant 8022–00165B to PW), and Innovation Fund Denmark (grant 0224–00033A to PW).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest.

1. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583-589.

2. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.*: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871-876.

3. Ourmazd A, Moffat K, Lattman EE: **Structural biology is solved — now what?** *Nat Methods* 2022, **19**:24-26.

4. • Roda S, Robles-Martín A, Xiang R, Kazemi M, Guallar V: **Structural-based modeling in protein engineering. A must do**. *J Phys Chem B* 2021, **125**:6491-6500.
A clear and up-to-date overview of molecular modeling applied to industrial biocatalysts.

5. Yan B, Ran X, Jiang Y, Torrence SK, Yuan L, Shao Q, Yang ZJ: **Rate-perturbing single amino acid mutation for hydrolases: a statistical profiling**. *J Phys Chem B* 2021, **125**:10682-10691.

6. Mazurenko S, Prokop Z, Damborsky J: **Machine learning in enzyme engineering**. *ACS Catal* 2020, **10**:1210-1223.

7. • Frappier V, Keating AE: **Data-driven computational protein design**. *Curr Opin Struct Biol* 2021, **69**:63-69.
An accessible discussion on the challenges and possibilities of incoporating diverse kinds of data into computational protein design.

8. Vanacek P, Sebestova E, Babkova P, Bidmanova S, Daniel L, Dvorak P, Stepankova V, Chaloupkova R, Brezovsky J, Prokop Z, *et al.*: **Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization**. *ACS Catal* 2018, **8**:2402-2412.

9. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, Wicky BIM, Courbet A, de Haas RJ, Bethel N, *et al.*: **Robust deep learning-based protein sequence design using ProteinMPNN**. *Scienc* 2022, **378**:49-56, https://doi.org/10.1126/science.add2187(0)

10. •• Lu H, Diaz DJ, Czarnecki NJ, Zhu C, Kim W, Shroff R, Acosta DJ, Alexander BR, Cole HO, Zhang Y, *et al.*: **Machine learning-aided engineering of hydrolases for PET depolymerization**. *Nature* 2022, **604**:662-667.
A clear example of succesful ML-based design of an industrial enzyme. Mutations identified by a structure-based algorithm were tested in different enzyme scaffolds and shown to improve performance.

11. Yang M, Fehl C, Lees KV, Lim E-K, Offen WA, Davies GJ, Bowles DJ, Davidson MG, Roberts SJ, Davis BG: **Functional and informatics analysis enables glycosyltransferase activity prediction**. *Nat Chem Biol* 2018, **14**:1109-1117.

12. • Hauer B: **Embracing nature's catalysts: a viewpoint on the future of biocatalysis**. *ACS Catal* 2020, **10**:8418-8427.
A particularly broad and insightful analysis of the potential and bottlenecks for further implementation of industrial biocatalysis.

13. Jerves C, Neves RPP, Ramos MJ, da Silva S, Fernandes PA: **Reaction mechanism of the PET degrading enzyme PETase**

studied with DFT/MM molecular dynamics simulations. *ACS Catal* 2021, **11**:11626-11638.

14. Heckmann D, Lloyd CJ, Mih N, Ha Y, Zielinski DC, Haiman ZB, Desouki AA, Lercher MJ, Palsson BO: **Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models**. *Nat Comm* 2018, **9**:5252.

15. Fersht AR: **Catalysis, binding and enzyme-substrate complementarity**. *Proc R Soc Lond B Biol Sci* 1974, **187**:397-407.

16. •• Anand M, Rohr B, Statt MJ, Nørskov JK: **Scaling relationships and volcano plots in homogeneous catalysis**. *J Phys Chem Lett* 2020, **11**:8518-8526.
A visionary account on how LIE relationships and the SP can be applied as a tool for the design of inorganic catalysts. The paper is accessible for non-specialists.

17. Anslyn EV, Dougherty DA: **Modern Physical Organic Chemistry**. University Science Books; 2004.

18. Norskov JK, Bligaard T, Rossmeisl J, Christensen CH: **Towards the computational design of solid catalysts**. *Nat Chem* 2009, **1**:37-46.

19. Fersht AR, Wells TNC: **Linear free-energy relationships in enzyme binding interactions studied by protein engineering**. *Prot Eng* 1991, **4**:229-231.

20. Fersht AR, Leatherbarrow RJ, Wells TNC: **Structure-activity-relationships in engineered proteins - analysis of use of binding-energy by linear free-energy relationships**. *Biochemistry* 1987, **26**:6030-6038.

21. Toney MD, Kirsch JF: **Direct bronsted analysis of the restoration of activity to a mutant enzyme by exogenous amines**. *Science* 1989, **243**:1485-1488.

22. Lin JL, Wheeldon I: **Kinetic enhancements in DNA-enzyme nanostructures mimic the Sabatier Principle**. *ACS Catal* 2013, **3**:560-564.

23. Kari J, Olsen JP, Jensen K, Badino SF, Krogh KBRM, Borch K, Westh P: **Sabatier principle for interfacial (heterogeneous) enzyme catalysis**. *ACS Catal* 2018, **8**:11966-11972.

24. Bååth JA, Jensen K, Borch K, Westh P, Kari J: **Sabatier principle for rationalizing enzymatic hydrolysis of a synthetic polyester**. *JACS Au* 2022, **2**:1223.

25. Feller BE, Kellis JT, Cascão-Pereira LG, Robertson CR, Frank CW: **Interfacial biocatalysis on charged and immobilized substrates: the roles of enzyme and substrate surface charge**. *Langmuir* 2011, **27**:250-263.

26. Feller BE, Kellis JT, Cascão-Pereira LG, Robertson CR, Frank CW: **The role of electrostatic interactions in protease surface diffusion and the consequence for interfacial biocatalysis**. *Langmuir* 2010, **26**:18916-18925.

27. Hardesty JO: **Biocatalysis at the solid-liquid interface: subtilisin on surface-bound polypeptides and amylase variants on starch granules (PhD thesis)**. Department of Chemical Engineering. Stanford University; 2008.

28. Morris CJ, Corte DD: **Using molecular docking and molecular dynamics to investigate protein-ligand interactions**. *Mod Phys Lett B* 2021, **35**:2130002.

29. Schaller KS, Molina GA, Kari J, Schiano-di-Cola C, Sørensen TH, Borch K, Peters GNH, Westh P: **Virtual bioprospecting of interfacial enzymes: relating sequence and kinetics**. *ACS Catal* 2022, **12**:7427-7435.

30. Kari J, Molina GA, Schaller KS, Schiano-di-Cola C, Christensen SJ, Badino SF, Sørensen TH, Røjel NS, Keller MB, Sørensen NR, *et al*.: **Physical constraints and functional plasticity of cellulases**. *Nat Comm* 2021, **12**:3847.

31. Gutiérrez-de-Terán H, Aqvist J: **Linear interaction energy: method and applications in drug design**. *Methods Mol Biol* 2012, **819**:305-523.

32. Schaller KS, Kari J, Borch K, Peters GH: **Binding prediction of multi-domain cellulases with a dual-CNN**. *arXiv Prepr arXiv* 2022, https://doi.org/10.48550/arXiv.2207.02698

33. Sousa SF, Ramos MJ, Lim C, Fernandes PA: **Relationship between enzyme/substrate properties and enzyme efficiency in hydrolases**. *ACS Catal* 2015, **5**:5877-5887.

34. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R: **The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters**. *Biochemistry* 2011, **50**:4402-4410.

35. Halling P, Fitzpatrick PF, Raushel FM, Rohwer J, Schnell S, Wittig U, Wohlgemuth R, Kettner C: **An empirical analysis of enzyme function reporting for experimental reproducibility: Missing/incomplete information in published papers**. *Biophys Chem* 2018, **242**:22-27.

36. Rich RL, Papalia GA, Flynn PJ, Furneisen J, Quinn J, Klein JS, Katsamba PS, Waddell MB, Scott M, Thompson J, *et al*.: **A global benchmark study using affinity-based biosensors**. *Anal Biochem* 2009, **386**:194-216.