

Enzymtechnologie

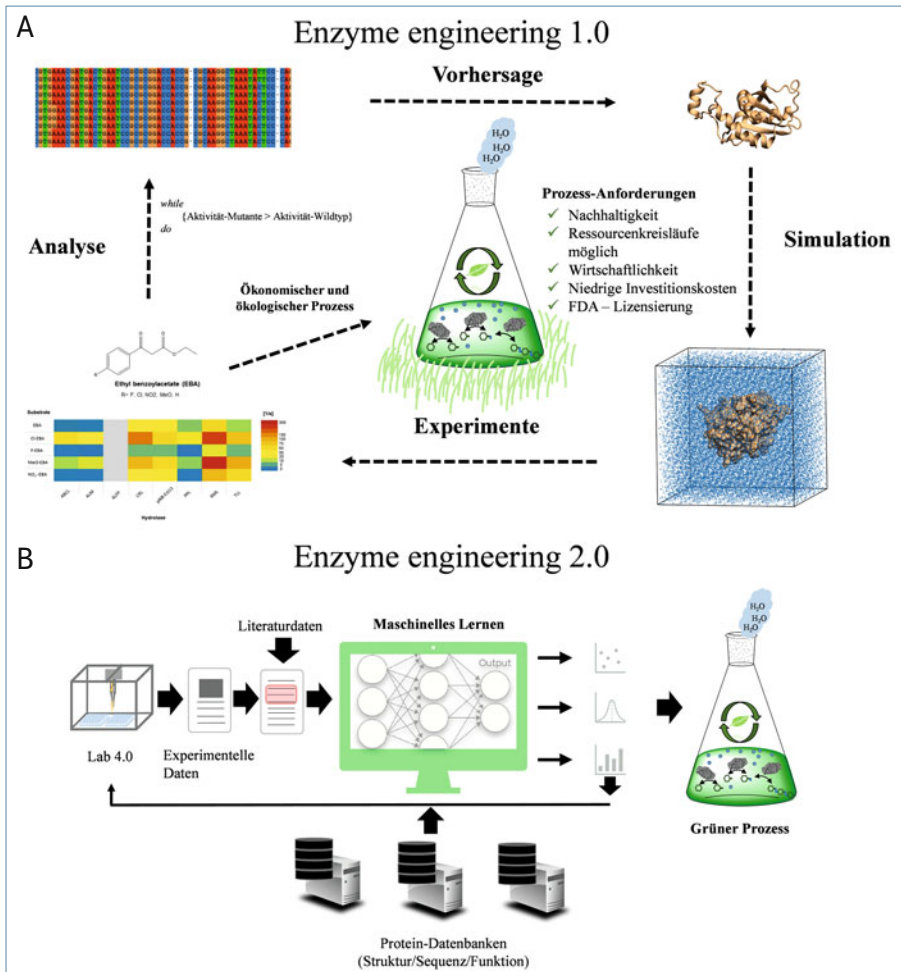
Neue *in silico*-Methoden für die Etablierung einer Grünen Chemie

SVEN JAGER¹, OLIVER BUB²¹ COMPUTATIONAL BIOLOGY AND SIMULATION, TU DARMSTADT² INSTITUT FÜR BIO- UND LEBENSMITTELTECHNOLOGIE, BEREICH II – TECHNISCHE BIOLOGIE, KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT)

Many fine chemicals have to meet high demands in the chemical and pharmaceutical industries in terms of optical purity, which can lead to high costs in the production and, in addition, to high amounts of waste. Catalyst optimization is necessary here because the observed substrates are mostly „man-made materials“ and evolution has not yet had time to evolve biocatalysts for this purpose. This is where advanced computational concepts such as simulation or machine learning help to improve such designs in terms of efficiency and sustainability.

DOI: 10.1007/s12268-018-0892-y
© Springer-Verlag 2018

Wir befinden uns an der Schwelle einer neuen Industrie, die nicht nur die digitale Revolution einbezieht (Industrie 4.0), sondern auch die Ressourcenkreisläufe unserer Wirtschaft einschließt. Die lineare Ökonomie setzte seit der industriellen Revolution voraus, dass Produkte auf einer Deponie landen, in die Umwelt gelangen oder verbrannt werden. Eine neue ökonomische Strategie, Abfälle zu vermeiden oder wieder in Wertschöpfungsketten einzuleiten, wird unter dem Begriff der Kreislaufwirtschaft verstanden, der auch ein Schlüsselement zur Bekämpfung des Klimawandels sein könnte und unsere Gesellschaft unabhängiger von Konfliktmaterialien machen könnte. Daher sind die Effizienz, mit der auch Chemikalien hergestellt werden, und der ökologische Fußabdruck betroffen. Viele Wirkstoffe und Feinchemikalien müssen hohen Ansprüchen in der Chemie und Pharmazie in Bezug auf optische Reinheit entsprechen, was zu hohen Kosten bei ihrer Herstellung führen kann und zudem zu hohen Abfallmengen führt. Ein Beispiel für eine neue, effiziente und chirale enzymatische Grüne Chemie ist das Diabetes-mellitus-Typ-2-Medikament Sitagliptin, welches als chirales Arzneimittel für die Hemmung der Dipeptidylpeptidase 4 zuständig ist. Darüber hinaus sind als weitere Beispiele die Pharmazeutika Montelukast, Abacavir, (S)-Ibuprofen und L-Ephedrin bekannt, die darlegen, dass Enzyme an Bedeutung für die Herstellung von chiralen Wirkstoffen gewonnen haben, insbesondere auch, um die Prinzipien der Grünen Chemie zu erfüllen [1, 2]. Gerade Projekte im Rahmen von Pharmazeutika mit hohen Produktpreisen könnten Beispiele mit



◀ **Abb. 1:** Schemata zum modernen *Enzyme Engineering* des 21. Jahrhunderts. **A, Enzyme Engineering 1.0:** Die klassischen Verfahren setzen auf die Kombination von *in silico*-Vorhersagen, MD(*molecular dynamics*)-Simulationen und notwendigen Laborexperimenten. **B, Enzyme Engineering 2.0:** Das etablierte Konzept wird um die Möglichkeit, ML(*machine learning*)- und DL(*deep learning*)-Anwendungen zu integrieren, erweitert. Zudem können Automatisierungen den experimentellen Ablauf beschleunigen.

Symbolcharakter für die neuen Möglichkeiten der modernen Biokatalyse werden, die sich nicht mehr nur auf klassische Lipasen beschränken, sondern auch andere Enzymklassen umfassen. Lange Zeit galt die Synthese chiraler Pharmazeutika als ein Steckenpferd der klassischen Chemie, die bestenfalls Enzyme nutzte, um ein Racemat zu spalten. Die Enzyme, die bis heute wirtschaftliche Bedeutung erlangten, waren hauptsächlich Lipasen, die besonders in der Waschmittelchemie Anwendung fanden und stellenweise zur Racematspaltung von racemischen Estern verwendet werden. Neben der industriellen Anwendung im Bereich der Pharmazie sind vor allem die Lebensmittelindustrie, Biotechnologieforschung sowie Gewerbe und Haushalte Abnehmer von Enzymprodukten. Der Schwerpunkt der industriellen Enzympräparate liegt nach wie vor auf der Enzymklasse der Hydrolasen. Die einzige Ausnahme bildet hierbei die Glukoseisomerase zur Herstellung von *high fructose corn*-Sirup im Produktionsmaßstab von 10^7 Tonnen pro Jahr [3]. Neben der Anwendung der Hydrolasen fanden lange Zeit keine weiteren Enzymklassen Anklang, obwohl die enzymatische Vielfalt schier endlos erscheint und entgegen der weitverbreiteten Meinung die Kosten für industriell verwendete Enzymprodukte zwischen 50 und 500 US-Dollar pro Kilogramm sowie die durchschnittlichen anteiligen Produktkosten zwischen 0,1 und 10 US-Dollar pro Kilogramm liegen [3]. Es gibt sogar Beispiele für den Umsatz von Großmengen an Enzym durch private Haushalte und Kleinunternehmen, so z. B. die Transglutaminase zur Erzeugung von quervernetzten Proteinmixturen, die zur Erzeugung von Fleischanaloga verwendet wird. Die Transglutaminase ist derzeit beim Großhändler Amazon für rund 13 Euro pro 100 Gramm Enzymprodukt erhältlich. Das erwartete Marktvolumen für Enzyme soll im Jahr 2018 gar auf 4,3 Milliarden Euro steigen [4]. Ein Grund, warum der Anwendungsbereich dennoch beschränkt ist, liegt in der Natur der beschriebenen Enzyme, die zumeist nur an die Bedingungen ihrer Wirtorganismen adaptiert sind. Eine Anwendung finden aber vor allem robuste Enzympräparate, wie z. B. die weitverbreitete immobilisierte Lipase B aus *Candida antarctica* (bekannt als Novozym 435 oder N435), welche auch in organische Chemielabore Einzug gehalten hat. Lipasen sind zwar noch die wichtigste Klasse an Biokatalysatoren, aber es hat ein Wandel begonnen, der als „dritte Welle der Biokatalyse“ verstanden wird, die

sich vor allem darin begründet, verbesserte und angepasste Enzyme zu erzeugen [5].

Grüne Biokatalyse

Die größte Aufmerksamkeit erlangte die Firma Codexis mit der Erzeugung einer ω -Transaminase zur effizienten Darstellung von Sitagliptin. Die unter dem Namen ATA-117 bekannte ω -Transaminase wurde durch Kombination von gelenkter Evolution und Protein-Engineering gezielt an die Prozessbedingungen zur Synthese von Sitagliptin angepasst. Das Ergebnis war ein sowohl ökonomisch als auch ökologisch verbesserter Sitagliptin-Prozess, der eine zehn bis 13 Prozent höhere Produktausbeute lieferte, eine um 56 Prozent gesteigerte Produktivität aufzeigte und die Abfallmenge um 19 Prozent reduzierte [6]. Dieses Leuchtturmprojekt aus dem Jahr 2010 könnte das Signal für die chemische Industrie sein, eine Grüne enzymbasierte Chemie des 21. Jahrhunderts zu entwickeln und zu etablieren. Hierbei können Enzyme als alternative Katalysatoren für Prozesse in Betracht gezogen werden, um Abfallströme zu reduzieren, Energiekosten zu sparen und Anlagenkosten für Prozesse zu vermeiden. Es müssen jedoch prozessstabile Enzyme designt werden. Das gezielte Enzymdesign hat vor allem das Ziel, die Biokatalysatoreigenschaften zu verbessern, um die Stabilität in Lösungsmitteln, die Temperatur- und Langzeitstabilität sowie die Akzeptanz von hohen Substrat- und Produktkonzentrationen durch den Biokatalysator zu optimieren. Die gewählte Strategie von Codexis zum gezielten ω -Transaminase-Design könnte mit dem heutigem Wissenstand wesentlich schneller und effizienter durchgeführt werden, da zum einen die Anzahl an charakterisierten ω -Transaminasen in der Zeit bis heute zugenommen hat, zum anderen könnten neue *in silico*-Ansätze den Umfang enzymatischer Screeningverfahren maßgeblich reduzieren. Rechenbasierte Verfahren zur gezielten Optimierung der Enzymstabilität und zur Veränderung der Substratselektivität sind bereits seit mehr als zehn Jahren in der Anwendung, aber das größte Hemmnis bleibt bis heute das fehlende Verständnis der Relation zwischen Proteinstruktur und -funktion.

Modernes Enzym-Engineering

Ein vielversprechender Ansatz ist die Methode des maschinenbasierten Lernens (ML), welche zu einem besseren Verständnis zwischen Sequenz und Funktion führen kann, wie die Forschungsgruppe um Francis Arnold

eindrucksvoll am Beispiel von designten Membranproteinen zeigte [7]. Diese Art der Optimierung ist hier notwendig, da die betrachteten Substrate meist synthetisch sind (*man-made materials*) und die evolutionäre Zeitskala zu kurz für die Evolvierung von Biokatalysatoren ist. Neben randomisierten Screenings werden Simulationsmethoden wie *molecular dynamics* (MD) immer häufiger für das Design von (Bio-)Molekülen zu Hilfe genommen (**Abb. 1**). Solche Methoden berechnen die Bewegung von molekularen Strukturen auf sehr kleinen Zeitskalen (10^{-15} Sekunden) und bieten somit wichtige zeitaufgelöste Informationen anstelle von statischen Kristallstrukturen. Somit können mikroskopische Prozesse jenseits der Auflösungsgrenze betrachtet werden. Das abgeleitete Wissen kann genutzt werden, um molekulare Designhypothesen zu bilden und reduziert im besten Fall den Lösungsraum eines Screeningansatzes [8]. Durch die Optimierung der zugrunde liegenden Algorithmen in den letzten zehn Jahren sowie die Beschleunigung der Berechnung von Kräften durch die Nutzung von Grafikkarten können für die Simulation relevantere Zeitskalen erreicht werden [9]. Unter Betrachtung dieser Entwicklung werden Simulationsmethoden zunehmend im Bereich des molekularen Designs und der Wirkstoffforschung herangezogen. Trotz ausgeklügelter Simulations- und Vorhersagemethoden bleibt bis heute das größte Problem das fehlende Verständnis über die Relation zwischen Proteinsequenz und -funktion. Die besagte Methode des ML konnte schon an einfachen Beispielen wie dem Design von peptidbasierten Wirkstoffen gezeigt werden [10]. Das Potenzial ist hier aber noch lange nicht ausgeschöpft, wie die Forscher um Francis Arnold eindrucksvoll am Beispiel von designten Membranproteinen aufzeigten [7]. ML kann aber auch mit MD-Simulationen kombiniert werden. Solche Modelle benötigen in einigen Fällen nicht einmal die Sequenz, sondern lediglich biophysikalische Parameter, welche direkt aus der Sequenz berechnet werden, um ein erstes Modell von Sequenz und Funktion zu erstellen. Leider benötigt ML immer eine Vielzahl an Daten zum Trimmen des Modells. Wenn aber ein ML-Modell auf der Grundlage von schnell durchführbaren Experimenten, wie dem *Next Generation Sequencing*, trainiert wird, ist das Problem der Datenerhebung schnell gelöst. Gerade im RNA/Aptamer-Design wird diese Kombination immer häufiger angewendet, um HTS-Prozesse, wie z. B. *systematic evolution of*

ligands by exponential enrichment (SELEX), zu optimieren [11]. Eine Weiterentwicklung und Ergänzung zum ML stellt das *deep learning* (DL) dar, das künstliche neuronale Netze nachahmt und zur Lösung von Gesichts- und Spracherkennung bereits verwendet wird. Die Forschungsgruppe von Lana Garmire nutzte beispielsweise DL erfolgreich zur Differenzierung von Brustkrebstypen auf der Grundlage von Stoffwechseldaten [12].

Vierte Welle des Enzym-Engineerings

Im Rahmen einer Grünen Chemie helfen die genannten Verfahren, schneller Enzyme für den Abbau von z. B. synthetischen Materialien zu finden und dabei den experimentellen Aufwand zu verringern. Gerade bei groß angelegten Screenings kann eine vorherige Betrachtung oder das Einschränken des möglichen Sequenzraums einen entscheidenden Beitrag liefern, die Erfolgsquote zu maximieren sowie den gesamten Zeitaufwand des Screenings zu minimieren. Der US-amerikanische IT-Spezialist IBM könnte mit seinem kognitiven und intelligenten Rechenansatz „Watson“ auch in der Wissenschaft und Forschung den klassischen Forschungsbetrieb verändern und so durch die Hilfe von groß angelegten Datenbanken Lösungsvorschläge erzeugen und Screeningverfahren drastisch verkürzen. Watson kann verschiedene Datenquellen verknüpfen, wie z. B. Daten-, Literatur-, Patente-, Protein/DNA-Sequenzen-Datenbanken, sowie chemische/pharmakologische Daten integrieren und dem Forscher Modelle vorschlagen. Es wurde bereits demonstriert, dass Watson neue Wirkstoffe oder Wirkstoff-

ziele aus Datenbankwissen generieren kann, warum sollte dies für ein gezieltes Enzymdesign keine erfolgreiche Strategie werden [13]? Der Biokatalyse steht eine Revolution bevor, die den wissenschaftlichen Betrieb in Forschungsinstituten aufwirbeln und zugleich Enzyme in einer neuen Industrie etablieren wird. Dazu bedarf es nun mutigen Forschern, die Enzyme-Engineering mit DL-Prozessen kombinieren, um gezielt die Eigenschaften von Enzymen vorherzusagen zu können. ■

Literatur

- [1] Garske AL, Kapp G, McAuliffe J (2017) Industrial Enzymes and Biocatalysis. In: Kent JA, Bommaraju TV, Barnicki SD (Hrsg) Handbook of Industrial Chemistry and Biotechnology. Springer International Publishing, Basel, S. 1625
- [2] Anastas PT, Warner JC (1998) Green Chemistry: Theory and Practice. Oxford University Press, New York
- [3] DiCosimo R, McAuliffe J, Poulou JA et al. (2013) Industrial use of immobilized enzymes. *Chem Soc Rev* 42:6437–6474
- [4] MarketsandMarkets (2014) Specialty enzymes market by source, type, application & geography global trends & forecasts to 2018. <https://www.marketsandmarkets.com/Market-Reports/specialty-enzymes-market-21682828.html>
- [5] Daiha KdeG, Angeli R, de Oliveira SD et al. (2015) Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *PLoS One* 10:e0131624
- [6] Jordan A, Gathergood N (2013) Designing safer and greener antibiotics. *Antibiotics* 2:419–438
- [7] Bedbrook CN, Yang KK, Rice AJ et al. (2017) Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization. *PLoS Comput Biol* 13:e1005786
- [8] He D, Huang L, Xu Y et al. (2017) Molecular dynamics directed rational design and fluorescence binding assay of phosphopeptide ligands for PLK polo-box domain. *Mol Simul* 43:176–182
- [9] Kutzner C, Páll S, Fechner M et al. (2015) Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J Comput Chem* 36:1990–2008
- [10] Giguère S, Laviolette F, Marchand M et al. (2015) Machine learning assisted design of highly active peptides for drug discovery. *PLoS Comput Biol* 11:e1004074
- [11] Quang NN, Perret G, Ducongé F (2016) Applications of high-throughput sequencing for *in vitro* selection and characterization of aptamers. *Pharmaceuticals* 9:76

[12] Alkawaa F, Chaudhary K, Garmire LX (2018) Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J Proteome Res* 17:337–347

[13] Chen Y, Ellenee Argentinis JD, Weber G (2016) IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther* 38:688–701

Korrespondenzadressen:



Sven Jäger

Computational Biology and Simulation
TU Darmstadt
Schnittspahnstraße 2
D-64287 Darmstadt
Tel.: 06151-16-20372
jager@bio.tu-darmstadt.de



Oliver Buß

Institut für Bio- und Lebensmitteltechnologie
Bereich II – Technische Biologie
Karlsruher Institut für Technologie
Engler-Bunte-Ring 3
D-76131 Karlsruhe
Tel.: 0721-608-467-39
Fax: 0721-608-448-81
oliver.buss@kit.edu