Ministerium für Wirtschaft,
Industrie, Klimaschutz und Energie
des Landes Nordrhein-Westfalen

Trend-Monitoring
NRW 2040: Trendstudie Proteinstrukturvorhersage

www.wirtschaft.de

Inhaltsverzeichnis

Inhaltsverzeichnis 2

Einordnung 4

Vorgehensweise 5

1. Einführung in das Thema: Proteine und Proteinstrukturvorhersage 7

1.1. Was sind Proteine? 7

1.2. Welche Funktionen üben Proteine in Organismen aus? 8

1.3. Wirtschaftliche Bedeutung von Proteinen - Vision einer Bioökonomie 8

1.4. Räumliche Proteinstruktur ist entscheidend für biologische Funktion der Proteine 9

1.5. Strukturebenen von Proteinen 9

1.6. Von der Fragestellung der Proteinfaltung zur Proteinstrukturvorhersage 10

2. Proteinstrukturvorhersage in den Lebenswissenschaften 11

2.1. Einschätzungen zur Zukunft der Strukturbiologie 12

2.2. Forschung an den noch bestehenden Einschränkungen bei der Proteinstrukturvorhersage 12

2.3. Zukunftsvision: Molekulardynamische Simulation ganzer Zellen 13

2.4. Neue KI-Anwendungen in den Lebenswissenschaften 13

3. Protein-Engineering 16

3.1. Kohlenstoff-Fixierung 16

3.2. Bioremediation von Öl 16

4. Proteindesign 18

4.1. Mit ChatGPT vergleichbare Programme für das Proteindesign 18

4.2. Proteindesign bestimmter Proteinklassen 19

5. Biokatalyse – Enzyme 21

5.1. Biokatalyse von PET und anderen Kunststoffen 22

5.2. Bioremediation von Organohaliden 23

5.3. Datenbank der Enzym-Mechanismen 23

5.4. Proteindesign für Rekombinasen zur Genome-Editierung 23

6. Bioproduktion 24

6.1. Biosensoren 24

6.2. Beschleunigung von Forschung und Entwicklung in der Bioindustrie 24

6.3. Automatisierte Entwicklung von Bioprozessen 25

7. Gesundheit - Medikamentenentwicklung 26

7.1. Zukünftige Pandemien 26

7.2. Vernachlässigte Krankheiten 27

7.3. Malaria 27

7.4. Antibiotika-Resistenz 27

7.5. Tuberkulose 28

7.6. Krebs aus Sicht der Strukturbiologie 28

8. Tiergesundheit 29
8.1. Vetinformatik: ein Forschungskonzept für die Veterinärmedizin 29
9. Pflanzengesundheit 30
9.1. Schutz vor Pilzerkrankungen 30
9.2. Schutz vor Viruserkrankungen 30
10. Weitere Anwendungen 31
10.1. Proteine bei der Eisbildung in Wolken 31
10.2. Verbesserte Nutzung von Keratin aus Geflügelfedern 31
11. NRW-Perspektive 32
11.1. Proteinstrukturvorhersage in den Lebenswissenschaften 32
11.2. Protein-Engineering und Proteindesign 39
11.3. Biokatalyse 40
11.4. Bioproduktion 43
11.5. Gesundheit - Medikamentenentwicklung 46
11.6. Pflanzengesundheit - NRW-Akteure 51
12. SWOT-Analyse NRW und Fazit 53
12.1. SWOT-Analyse NRW 53
12.2. Fazit 56

Einordnung

Das Ministerium für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen (MWIKE) hat ein Trend-Monitoring beauftragt, um NRW-spezifische Trends mit zeitlicher Perspektive bis 2040 abzuleiten, die für die wirtschaftspolitische Gestaltung in NRW von besonderer Bedeutung sind. Dies ist bspw. dann gegeben, wenn Trends in erheblichem Ausmaß zur Erreichung der Ziele und Strategien der Landesregierung (z. B. Industriepolitisches Leitbild, Digitalisierungsstrategie, Innovationsstrategie) beitragen könnten. Über die Laufzeit des Monitorings soll eine Reihe von Trendstudien zu identifizierten Themen ausgearbeitet und präsentiert werden und so Impulse für den wirtschaftspolitischen Zukunftsdiskurs in NRW geben.

Das vorliegende Dokument enthält die erste Trendstudie mit dem Thema Anwendungen der Proteinstrukturvorhersage.

Ziel des Dokumentes

Diese Trendstudie soll der allgemeinen Öffentlichkeit in NRW eine Informationsgrundlage bereitstellen, um einen wirtschaftspolitischen Zukunftsdiskurs in NRW anzustoßen. Gegenstand des Diskurses ist die Frage, wie das Land NRW auf lange Sicht – mit dem Zeithorizont 2040 – von den Fortschritten bei der Proteinstrukturvorhersage profitieren kann und welche wirtschaftlichen und gesellschaftlichen Anwendungspotenziale in NRW von besonderem Interesse sind.

Die Veröffentlichung der Trendstudie wird durch bis zu zwei Präsentationen kommunikativ begleitet. Diese Präsentationen fungieren als Impulse für den Zukunftsdiskurs in NRW über die bearbeiteten Trendthemen. Dabei können die beteiligten NRW-Akteure ihre eigenen Einschätzungen zu den Ergebnissen aus dieser Trendstudie artikulieren und eigene Bewertungen zu den globalen Chancen und Risiken sowie zu den Stärken und Schwächen in NRW in Bezug auf die betrachteten Trends einbringen.

Vorgehensweise

Die Studienausarbeitung beruht auf der folgenden Grundidee: Die im Zuge der Themen-Identifikation und Erstbewertung im Monitoring-Prozess gefundenen NRW-Bezüge werden bei der Ausarbeitung der Trendstudien näher analysiert, ergänzt und schließlich zu einer SWOT-Analyse[1] verdichtet. Zunächst werden mögliche weltweit absehbare Trendverläufe und -ausprägungen erhoben und dargestellt, die soweit möglich auch die wirtschaftlichen Anwendungspotenziale umfassen. Die zugehörigen Ergebnisse finden sich in den Kapiteln 3 bis 10. Im zweiten Schritt werden in Kapitel 11 dazu – z. T. beispielhaft – jeweilige Akteure und Aktivitäten in NRW vor- und gegenübergestellt. Durch den Abgleich der globalen und NRW-spezifischen Perspektiven ergeben sich die vergleichenden Informationen für die SWOT-Analyse in Kapitel 12. Es werden so wirtschaftspolitische Chancen und Herausforderungen deutlich und zugehörige Handlungsoptionen können als Impuls für einen gesellschaftlichen und wirtschaftspolitischen Zukunftsdiskurs abgeleitet werden. Umsetzungsreife Handlungsempfehlungen werden nicht angestrebt.

Literaturanalyse als Informationsgrundlage

Eine Literaturanalyse eignet sich im Allgemeinen dazu, den Stand des Wissens in betrachteten Themenbereichen strukturiert zu erheben und inhaltlich zu umreißen. Dabei ist es lohnend, in der Literaturanalyse Zukunftssichten verschiedener Autoren zu sammeln und gewichtet gegenüberzustellen (Einzelheiten - vgl. Methodenstudie zum Trendmonitoring NRW 2040).

Beim vorliegenden Thema der Proteinstrukturvorhersage ging die Literaturanalyse von der folgenden zentralen Quelle aus:

> Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Es wurden wissenschaftliche Publikationen ausgewertet, die auf diese zentrale Quelle verweisen. Die Literaturanalyse wurde dadurch recht herausfordernd, weil das Feld eine hohe Dynamik aufweist:

> Jun. 2023 > 7.000 Zitationen
>
> Okt. 2023 > 9.000 Zitationen
>
> Dez. 2023 > 11.000 Zitationen
>
> Jun. 2024 > 14.000 Zitationen

Die Sichtung wurde im ersten Durchgang daher auf die Titel beschränkt und nachfolgend durch eine Suche nach zukunftsorientierten Schlagworten weiter fokussiert (futur*, trend, perspective). Zudem wurden Arbeiten gesichtet, die explizit Anwendungsbezüge adressieren, sowie Review-Artikel.

Nach Sichtung der Literatur und dem Ablesen einer Gliederung anhand der hauptsächlichen Anwendungspotenziale wurde eine ergänzende Recherche nach Start-up-Unternehmen in der Datenbank Crunchbase mit Suchbegriffen entlang dieser Anwendungspotenziale durchgeführt (u. a. „protein structure“, „protein folding“, „biocatalys*“, „protein design“, „protein engineering“, „enzyme design“, „enzyme engineering“)

Mit Blick auf NRW-Unternehmen wurde in der Datenbank Crunchbase zusätzlich nach u. a. den Suchbegriffen „proteins“, „enzymes“, „peptides“ bei Einschränkung auf Unternehmen mit Sitz in NRW gesucht sowie in den Unternehmenslisten des Verbands der forschenden Pharmaunternehmen, vfa,[2] sowie dem Unternehmensverzeichnis von BIO.NRW.[3]

[1] SWOT – „strengths, weaknesses, opportunities, threats“; Stärken, Schwächen, Chancen, Risiken.
[2] Vgl. vfa, Interaktive Standortkarte: https://www.vfa.de/de/verband-mitglieder.
[3] Vgl. https://bio.nrw.de/project/spot-on-biotechnology-business-2023/.

Analyse anhand der STEEP-Dimension und Bewertungskriterien

Die im Zuge der Literaturanalyse identifizierten Dokumente wurden nachfolgend gesichtet und analysiert. Dazu wurden die wichtigsten Aussagen anhand der STEEP-Dimensionen[4] geordnet. Soweit vorhanden wurden Gemeinsamkeiten und Unterschiede herausgearbeitet. Anhand der Häufigkeit von Themen und der Schlüssigkeit von Argumentationen wurde eine Ordnung nach Akzeptanz und Verbreitung von Zukunftssichten aufgestellt. Dabei ergab sich eine Konzentration auf die wissenschaftliche und technische Dimension, die dem Reifegrad des Themas und insofern auch dem betrachteten Zeithorizont 2040 entspricht. Ein Bezug zu den sozialen, wirtschaftlichen, ökologischen und politischen Dimensionen kann allerdings oft indirekt über die identifizierten Anwendungspotenziale hergestellt werden.

SWOT-Analyse NRW

Die SWOT-Analyse wurde als etabliertes Instrument im Rahmen der Ausarbeitung dieser Trendstudie verwendet und dabei die identifizierten trendbezogenen Stärken und Schwächen des Landes NRW als „interne Faktoren“ betrachtet. Diese wurden externen Faktoren in der Form von globalen Chancen und Risiken gegenübergestellt, die etwa vom trendbezogenen globalen Marktumfeld ausgehen. Auf diesem Wege wird ein Abgleich der globalen und NRW-spezifischen Trend-Perspektiven geleistet und zu einer SWOT-Analyse verdichtet. Die Ergebnisse sind in Kapitel 12 dargestellt.

[4] STEEP – „social, technical, economic, environmental, political“; sozial, technisch, wirtschaftlich, umweltbezogen, politisch.

1. Einführung in das Thema: Proteine und Proteinstrukturvorhersage

1.1. Was sind Proteine?

Proteine werden umgangssprachlich auch Eiweiße genannt.[5] Es handelt sich dabei um lange Ketten, die in der Natur aus 20 verschiedenen Aminosäuren aufgebaut werden. Kürzere Ketten werden als Peptide bezeichnet. Von einem Protein spricht man typischerweise dann, wenn die Aminosäurekette aus mehr als 100 Aminosäuren besteht. Proteine können aber auch viel größer sein und mehrere Tausend von Aminosäuren enthalten. Die einzelnen Aminosäuren haben eine einheitliche Grundstruktur. Sie bestehen aus einem zentralen Kohlenstoffatom C (in Abb. 1.1 schwarz) mit einer Aminogruppe (chemische Formel NH_2, der Stickstoff N grau und der Wasserstoff H in Abb. 1.1 blau) und einer Carbonsäuregruppe (chemische Formel COOH, Sauerstoff O in Abb. 1.1 rot). Die 20 Aminosäuren unterscheiden sich nur durch ihre sogenannten Restgruppen, in Abb. 1.1 als Kästchen dargestellt und mit R_1 bzw. R_2 bezeichnet.

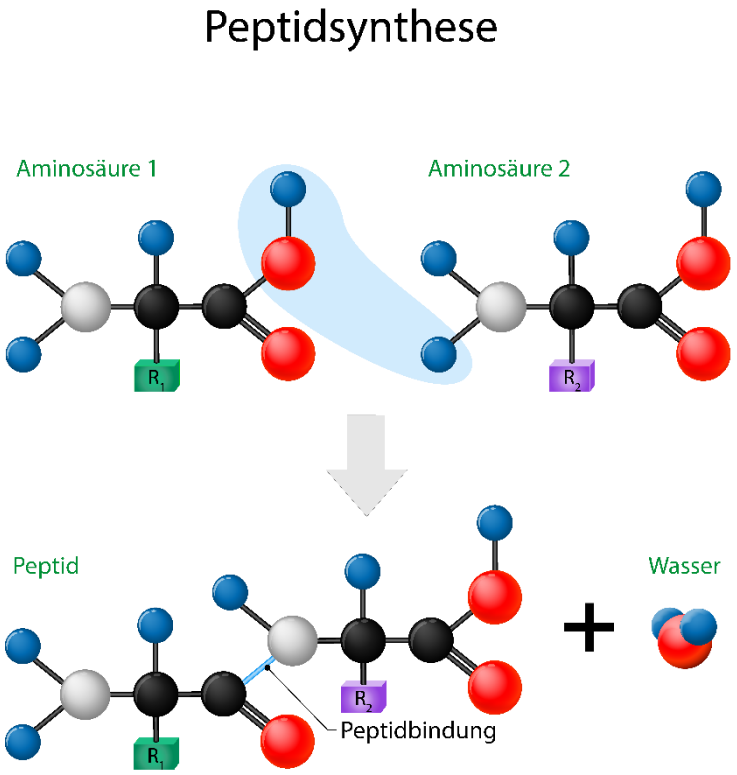

Abbildung 1.1: Peptidbindung - Bildung von Aminosäureketten (Quelle: Adobe Stockphoto).

Die Aminogruppe einer Aminosäure kann mit der Carbonsäuregruppe einer anderen Aminosäure reagieren. Dabei entsteht eine Bindung zwischen den beiden Aminosäuren, die als Peptidbindung bezeichnet wird, außerdem wird ein Wassermolekül frei (chemische Formel H_2O). Das Besondere an dieser Reaktion ist, dass das entstandene Molekül, das auch als Dipeptid bezeichnet wird, ebenfalls an einem Ende eine Aminogruppe und am anderen Ende eine Carbonsäuregruppe aufweist. Daher ist es möglich, dass sich nun Schritt für Schritt weitere Aminosäuren anlagern können und so lange Molekülketten entstehen, die in spezifischer räumlicher Anordnung das fertige Protein bilden. Diese Molekülketten sind durch die Abfolge der Aminosäuren charakterisiert, die man auch als Aminosäuresequenz bezeichnet. Die Aminosäuresequenz der natürlichen Proteine lässt sich bereits aus der Erbinformation ableiten, die im Wesentlichen die Baupläne für alle Proteine enthält.[6]

[5] Eine anschauliche Darstellung der nachfolgend beschriebenen Sachverhalte bietet das Lehrvideo „Proteine“: https://studyflix.de/chemie/proteine-3582.
[6] Die genauen Abläufe werden in dem Lehrvideo „Proteinbiosynthese“ erläutert: https://studyflix.de/biologie/proteinbiosynthese-2288.

1.2. Welche Funktionen üben Proteine in Organismen aus?

Proteine sind zum einen wesentlich für den Zellaufbau und bilden die Grundlage der Zellstruktur und damit letztlich auch für die äußere Form von Pflanzen, Tieren und Menschen. Proteine mit dieser Funktion werden Strukturproteine genannt. Andere Proteine, die sogenannten Enzyme, bewerkstelligen und beschleunigen den Ablauf bestimmter biochemischer Reaktionen. Sie sind wesentlich für den Stoffwechsel und damit bspw. für die Versorgung des Organismus mit Energie, wodurch die Körperfunktionen aufrechterhalten werden. Es gibt Proteine, die eine große Rolle für Muskeln und Bewegung spielen, diese werden Motorproteine genannt. In der Immunabwehr haben Proteine ebenfalls wichtige Funktionen etwa in der Form von Antikörpern, die körperfremde Krankheitserreger erkennen und helfen, sie zu bekämpfen. Proteine dienen in der Form von Hormonen als Signal- oder Botenstoffe, die bestimmte Anpassungsreaktionen des Körpers und Stoffwechselprozesse auslösen. Daneben gibt es Proteine, die andere Stoffe im Körper transportieren, sog. Transportproteine. Besonders bekannt ist hierbei das Hämoglobin, das für den Sauerstofftransport im Körper sorgt. Andere Transportproteine ermöglichen und regulieren den Transport von Stoffen durch die Zellhülle. Schließlich gibt es noch Speicherproteine zur Sammlung von Aminosäuren zum späteren Gebrauch.

1.3. Wirtschaftliche Bedeutung von Proteinen - Vision einer Bioökonomie

Angesichts dieser vielfältigen und zentralen Funktionen von Proteinen in Lebensprozessen aller Art, haben Proteine und das Wissen über Proteine - zumindest indirekt - auch eine große, wirtschaftliche Bedeutung. Das hängt damit zusammen, dass einerseits schon heute viele Wirtschaftszweige biobasierte Ausgangsstoffe nutzen bzw. biologische Grundlagen haben wie z.B. natürlich die Landwirtschaft und Ernährungsindustrie, aber auch die Forstwirtschaft mit der Holz- und Papierindustrie, die Leder- und Textilindustrie, Umweltwirtschaft und Medizin. In all diesen Wirtschaftszweigen kann ein besseres Verständnis von Lebensprozessen, das auch von einem genaueren Wissen über Proteine abhängt, ein wesentlicher Faktor für Effizienzsteigerungen sein. Hinzu kommt andererseits die überragende Rolle, die biobasierte Prozesse in einer zukünftigen, industriellen Bioökonomie spielen können und sollen, die wie folgt definiert ist:

> *Die industrielle Bioökonomie ermöglicht es, fossil basierte Produkte und Prozesse durch biobasierte zu ersetzen und zudem vollkommen neue Produkte, Produktionsverfahren und Geschäftsmodelle zu schaffen, indem biologisches Wissen mit technologischen Lösungen kombiniert und die natürlichen Eigenschaften biogener Rohstoffe hinsichtlich ihrer Kreislauffähigkeit, Erneuerbarkeit und Anpassungsfähigkeit genutzt werden. Bioökonomische Innovationen basieren auf der nachhaltigen Nutzung von Ressourcen, der Substitution von fossilen durch biologische Rohstoffe, der Nutzung von Reststoffen oder Nebenproduktströmen und dem Einsatz von biotechnologischen Prozessen.*[7]

Durch die industrielle Bioökonomie wird die wirtschaftliche Bedeutung von Proteinen und dem Wissen über Proteine auf die chemische Industrie, die Energiewirtschaft und die Grundstoffindustrie ausgeweitet und verstärkt. Das Bundeswirtschaftsministerium[8] zitiert Zahlen aus dem Bioökonomie-Monitoring der EU:[9] Demnach sind ca. 17,5 Mio. Beschäftigte in der EU im Bereich der Bioökonomie tätig, in Deutschland sind es 3,1–3,6 Mio. Das entspricht in der EU und in Deutschland jeweils ca. 9 % aller Beschäftigten. Sie erwirtschaften bezogen auf die EU ca. 614 Mrd. Euro und in Deutschland zwischen 165 und 265 Mrd. Euro bzw. zwischen 5 und 6 % des deutschen BIP. Mit Blick auf das Jahr 2050 wird geschätzt, dass dieser Anteil auf 9 bis 10 % steigen wird.

Im Jahr 2022 betrug der Umsatz der deutschen Biotechnologie-Branche 25 Mrd. Euro bei einer Beschäftigungszahl von 47.400.[10] In Bezug auf die unmittelbaren wirtschaftlichen Potenziale im Zusammenhang mit „technischen" Proteinen stehen die medizinischen Anwendungen im Vordergrund in der sog. roten oder medizinischen Biotechnologie. So erreichte der Umsatz mit Biopharmazeutika im Jahr 2022 einen Wert von 17,8 Mrd. Euro.[11] In

[7] Awenius, M., Bronsema, V., Grimm, L. et al. Leitmarkt Bioökonomie: Transformation und Skalierung für Nachhaltigkeit und Ressourcenwende. Positionspapier der Dialogplattform Industrielle Bioökonomie (2023). https://www.bmwk.de/Redaktion/DE/Downloads/I/Industrielle-Biooekonomie/positionspapier-der-dialogplattform-industrielle-biookonomie-leitmarkt-biookonomie.pdf.
[8] BMWK, Industrielle Bioökonomie: Wachstum und Innovation für den Standort Deutschland (2024). https://www.bmwk.de/Redaktion/DE/Dossier/industrielle-biooekonomie-wachstum-und-innovation.html.
[9] Knowledge Centre for Bioeconomy: https://knowledge4policy.ec.europa.eu/bioeconomy/monitoring_en.
[10] Bauer, M.K.A. Deutscher Biotechnologie-Report 2023. EY (2023). https://www.ey.com/de_de/forms/download-forms/2023/06/deutscher-biotechnologie-report-2023.
[11] Lücke, J., Bädeker, M. Medizinische Biotechnologie in Deutschland 2023 - Biopharmazeutika: Wirtschaftliche Kennziffern für den Biotech-Standort Deutschland (2023). BCG, vfa (Hrsg.). https://www.vfa.de/download/biotech-report-kompakt-2023.pdf.

der weißen oder industriellen Biotechnologie spielen die Enzyme eine herausragende Rolle. Insgesamt gibt es etwa 250 Enzyme als kommerzielle Produkte. Zu den Enzymen mit den größten Produktionsvolumina zählen die sog. Hydrolasen – etwa Proteasen oder Amylasen, die in Waschmitteln oder in der Lebensmittelherstellung eingesetzt und im Tonnenmaßstab produziert werden. Der Weltmarkt für Enzyme wird für das Jahr 2021 auf 6,4 Mrd. US-Dollar geschätzt.[12]

1.4. Räumliche Proteinstruktur ist entscheidend für biologische Funktion der Proteine

Wenn man sich die oben beschriebenen Funktionsweisen von Strukturproteinen, Motorproteinen, Transport- und Speicherproteinen vorstellt, so wird deutlich, dass hierfür neben chemischen Eigenschaften ganz unmittelbar auch die räumliche Struktur der Proteine entscheidend ist. Aber auch bei Enzymen, Hormonen und Antikörpern ist es die Kombination aus chemischen Eigenschaften (an der Oberfläche eines Proteins) und der räumlichen Struktur, die für die biologische Funktion ausschlaggebend ist. Man spricht in diesem Zusammenhang auch vom Schlüssel-Schloss-Prinzip. Dabei müssen meist zwei - manchmal aber auch mehrere - Komponenten räumlich zueinander passen (wie ein Schlüssel räumlich zu einem Schloss), damit es zu einer bestimmten biochemischen Reaktion kommen kann.

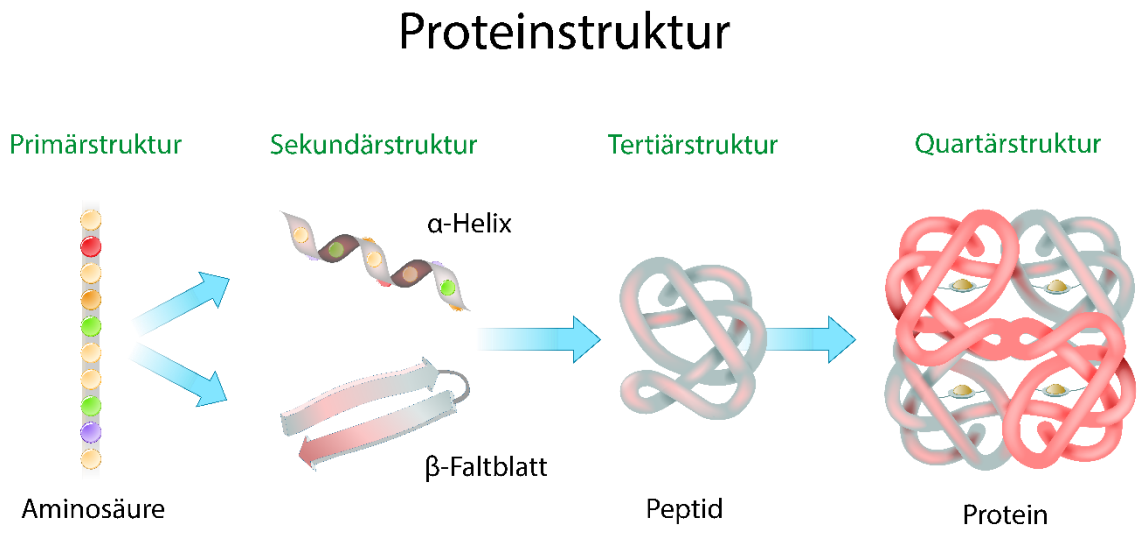

Abbildung 1.2: Räumliche Proteinstrukturen (Quelle: Adobe Stockphoto).

1.5. Strukturebenen von Proteinen

Bei der räumlichen Struktur von Proteinen werden mehrere Ebenen unterschieden, dargestellt in Abb. 1.2. Das Zustandekommen der verschiedenen Strukturebenen kann man sich durch Fäden oder Gummibänder verbildlichen, die verdrillt und zu einem Knäuel gewickelt werden. Die lineare Abfolge der einzelnen Aminosäuren eines Proteins, die Aminosäuresequenz, wird als Primärstruktur bezeichnet, die man sich wie in der Abbildung 1.2 als eine Perlenkette vorstellen kann. Die Perlen entsprechen in diesem Bild den Aminosäuren und der Faden zwischen den einzelnen Perlen den Peptidbindungen. Die Sekundärstruktur der Proteine beruht nicht auf molekularen Bindungen, sondern kommt durch sog. Wasserstoffbrückenbindungen[13] zustande, die bewirken, dass sich die lineare Kette faltet. Um im Bild der Perlenkette zu bleiben, kann man sich vorstellen, dass die Perlen kleine Magnete sind, die leicht aneinanderhaften. Die häufigsten Strukturen, die sich bei diesen Faltungen bilden, sind eine Schraubenform, die als α-Helix bezeichnet wird und eine flächige Form, die man β-Faltblatt nennt, bei der die Abschnitte der Kette nebeneinanderliegen. Mit der Tertiärstruktur wird die vollständige räumliche Struktur eines einzelnen Proteins bezeichnet, die sich aus der Sekundärstruktur aufgrund weiterer zwischenmolekularer Kräfte

[12] Die Angaben zu Enzymen nach: Graf, P. Enzyme – die Supertalente der Bioindustrie. Vgl. Abschnitt „Wirtschaftliches Potenzial“ (2023). https://biooekonomie.de/themen/dossiers/enzyme-die-supertalente-der-bioindustrie.
[13] Für eine anschauliche Erläuterung siehe das Lehrvideo „Wasserstoffbrückenbindungen“: https://studyflix.de/chemie/wasserstoffbruckenbindung-1668.

und durch sog. Disulfidbrücken[14] ausbildet. Für eine bildliche Vorstellung kann man an eine Telefonschnur bei älteren Telefonmodellen denken, die sich nach längerem Gebrauch in schwer zu entwirrende Knäuel verdrehen können. Lagern sich mehrere Einzelproteine zu einem Komplex zusammen, so spricht man von einer Quartärstruktur. Das oben angesprochene Hämoglobin ist eine Quartärstruktur und besteht beispielsweise aus vier Untereinheiten.

1.6. Von der Fragestellung der Proteinfaltung zur Proteinstrukturvorhersage

Der Forscher Christian Anfinsen[15] konnte 1961 zeigen, dass die räumliche Struktur von Proteinen, die sich durch die Proteinfaltung ergibt, nur von der Sequenz von Aminosäuren abhängt, aus der das Protein gebildet wird, und dass keine weitere genetische Information in den Faltungsprozess einfließt.

Aus dieser Beobachtung ergibt sich die Folgerung, dass sich im Prinzip die Struktur eines Proteins allein aus seiner Aminosäuresequenz ableitet. In welche Struktur sich ein Protein mit gegebener Aminosäuresequenz faltet, ist Gegenstand der wissenschaftlichen Aufgabenstellung der Proteinfaltung. Zur Einordnung der Schwierigkeit dieser Fragestellung weisen etwa Baker und Atzori darauf hin, dass ein typisches Protein 10^{300} mögliche Konformationen aufweisen kann.[16] Welche Konformation aus dieser riesigen Zahl ist nun diejenige, die das Protein tatsächlich einnimmt? Die Berechnung und der Vergleich der Eigenschaften aller möglichen Konformationen eines Proteins würden länger dauern als das Alter des Universums (für nur ein Protein).

Trotz dieser großen theoretischen Schwierigkeiten wurde über Jahrzehnte an der Fragestellung der Proteinfaltung geforscht. Seit 1994 gibt es den Wettbewerb „Critical Assessment of Protein Structure Prediction" (CASP), bei dem Vorhersagemodelle auf experimentell bekannte, aber noch nicht publizierte Proteinstrukturen angewendet und die Genauigkeit der Vorhersagen verglichen werden. Über die Jahre gab es dabei inkrementelle Fortschritte. Im Jahr 2016 erreichte das beste Modell einen Wert von 41 (auf einer Skala von 0 bis 100) für das Maß zur Vorhersagepräzision beim Vergleich zwischen den vorhergesagten und den experimentellen Strukturen. Im Jahr 2020 erreichte dann das Modell AlphaFold einen Durchschnittswert von 87 und konnte zwei Drittel der vorgelegten Proteinstrukturen mit einer Genauigkeitskennziffer oberhalb von 90 vorhersagen, was als gleichbedeutend mit der Genauigkeit der experimentellen Strukturbestimmung angesehen wird. Mit diesem Ergebnis wurde von vielen Beobachtern die Fragestellung der Proteinfaltung als gelöst angesehen, auch wenn es nach wie vor Limitierungen gibt (vgl. Kapitel 2).[16]

Seit Juli 2021 ist die AlphaFold-Software unter einer Open-Source-Lizenz auch für kommerzielle Unternehmen nutzbar. Die Funktionsweise wurde in einem wissenschaftlichen Fachjournal veröffentlicht[17] und die Strukturvorhersagen sind in öffentlich zugänglichen Datenbanken verfügbar, darunter die „AlphaFold Protein Structure Database"[18], die (Stand 2023) 200 Millionen Proteinstrukturvorhersagen enthält. Zum Vergleich: Die Datenbank der experimentell erhobenen Proteinstrukturen, ohne deren öffentlich zugänglichen Informationen der Erfolg von AlphaFold und der Weiterentwicklung AlphaFold2 unmöglich gewesen wäre, enthält (Stand Nov. 2023) knapp 230.000 Einträge.[19] AlphaFold beruht auf einer Kooperation des KI-Unternehmens DeepMind, einer Tochter von Google, und dem „European Bioinformatics Institute" (EBI; deutsch: Europäisches Institut für Bioinformatik), das Teil des Europäischen Laboratoriums für Molekularbiologie (EMBL) ist. Beide Organisationen haben ihren Sitz in Großbritannien.[27]

---

[14] In einer bestimmten Aminosäure (Cystein) kommt Schwefel vor. Zwischen den Schwefelatomen zweier räumlich benachbarter Cystein-Moleküle können kovalente Bindungen entstehen, die als Disulfid-Brücken bezeichnet werden. Vgl. Wikipedia-Eintrag „Disulfidbrücke"; https://de.wikipedia.org/wiki/Disulfidbrücke.
[15] Nobelstiftung: Darstellung zur wissenschaftlichen Arbeit von Christian Anfinsen: https://www.nobelprize.org/prizes/chemistry/1972/anfinsen/facts/.
[16] Baker, C.M. & Atzori, A. AlphaFold: Deep Learning, Drug Discovery and the Protein Structure Revolution. Chimia 76, 364 (2022). https://doi.org/10.2533/chimia.2022.364.
[17] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2.
[18] Varadi, M., Anyango, S. Deshpande, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research 50, D1, D439–D444 (2022). https://doi.org/10.1093/nar/gkab1061.
[19] Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. Nat Struct Mol Biol 10, 980 (2003). https://doi.org/10.1038/nsb1203-980.

AlphaFold war ursprünglich der Name des DeepMind-Teams, das den Wettbewerb CASP13 im Dezember 2018 gewonnen hat.[20] Die wesentliche Weiterentwicklung, die im November 2020 am darauffolgenden Wettbewerb CASP14 teilnahm und ebenfalls den ersten Platz belegte, trägt den Namen AlphaFold2.[17] Den aktuellsten Wettbewerb hat AlphaFold2 ohne eigene Teilnahme stark beeinflusst, weil der Quellcode von AlphaFold2 seit 2021 öffentlich ist und die erfolgreichsten der eingereichten Beiträge AlphaFold2 in der einen oder anderen Form integriert haben.[21] Weite Verbreitung hat auch die Anwendung RoseTTAFold gefunden, mit der eine Proteinstrukturvorhersage auf einem leistungsfähigen, handelsüblichen Computer („Gaming PC") in teilweise nur zehn Minuten berechnen kann.[22,23] Die Anwendung ColabFold zielt ebenfalls auf schnelle Berechenbarkeit der Proteinstrukturvorhersagen und erreicht dabei bis zu 1.000 Vorhersagen pro Tag auf einem Rechner mit nur einer GPU („graphic processing unit").[24] Das Tool OpenFold zielt auf eine Implementierung von AlphaFold2 mit vergleichbar hoher Genauigkeit, die es zusätzlich erlaubt, neue Fragestellungen zu untersuchen sowie Modelle auf selbst gewählten Datensätzen zu trainieren und dadurch besser zu verstehen, wie der Lernfortschritt genau zustande kommt.[25] Im Folgenden wird entweder allgemein von Proteinstrukturvorhersage gesprochen oder bei spezifischen Anwendungen von dem jeweils eingesetzten Tool zur Proteinstrukturvorhersage.

Es gibt Einschätzungen,[26] wonach die durch AlphaFold2 ausgelösten Fortschritte größere und schnellere Auswirkungen haben könnten als das Humangenomprojekt, das weithin als Revolution der Lebenswissenschaften anerkannt wird. Denn die Proteinstrukturvorhersagen sind im Vergleich zur Frühphase der Gensequenzierung kostengünstig, sodass Vorhersagen in absehbarer Zeit für alle Proteine zur Verfügung stehen könnten, während die Techniken zur Sequenzierung zunächst noch Jahre der Weiterentwicklung brauchten, bevor sie breiter eingesetzt werden konnten. Zudem bieten Proteinstrukturen unmittelbarere biologische Erkenntnisse, indem die Struktur eines Proteins direkt mit seiner Funktion verbunden ist, während die Auswirkungen von Gensequenzen zumindest zu Beginn der Genomik-Ära weitgehend undurchschaubar waren.

Insofern könnte die Revolution im Wissen zur Proteinfaltung schneller Früchte tragen. Gleichwohl wird erwartet, dass es ein oder zwei Jahrzehnte dauert, bis die vollen Auswirkungen dieser Entwicklung (wissenschaftlich, medizinisch und wirtschaftlich) richtig eingeschätzt werden können. Fachleute gehen schon heute davon aus, dass auf Basis dieses neuen Wissens Anwendungen in der Biotechnologie, Medizin, Landwirtschaft, Lebensmittelwissenschaft und dem Bioengineering ermöglicht werden.[27]

In den nachfolgenden Kapiteln 2 bis 10 werden Hinweise aus der wissenschaftlichen Literatur zu diesen zukünftigen Anwendungen und erste heutige Beispiele dargestellt.

2. Proteinstrukturvorhersage in den Lebenswissenschaften

Die unmittelbarsten Anwendungen von AlphaFold2 und verwandten Algorithmen zur Proteinstrukturvorhersage liegen in der Wissenschaft. Dadurch entstehen weitere wissenschaftliche Erkenntnisse und damit indirekt - gerade auf lange Sicht bis 2040 - auch neue wirtschaftlich relevante Anwendungspotenziale. In diesem Kapitel werden Aussagen zur direkten Wirkung der Proteinstrukturvorhersage auf die Lebenswissenschaften umrissen und soweit möglich Perspektiven für Anwendungen aufgezeigt.

[20] Senior, A.W., Evans, R., Jumper, J. et al. Improved protein structure prediction using potentials from deep learning. Nature 577, 706–710 (2020). https://doi.org/10.1038/s41586-019-1923-7.
[21] Schreiner, M. CASP15: AlphaFold's success spurs new challenges in protein-structure prediction. The Decoder 14 (2022). https://the-decoder.com/casp15-alphafolds-success-brings-new-challenges/.
[22] University of Washington, Institute for Protein Design. RoseTTAFold: Accurate protein structure prediction accessible to all. Pressemeldung, 15. Jul. 2021. https://www.ipd.uw.edu/2021/07/rosettafold-accurate-protein-structure-prediction-accessible-to-all/.
[23] Baek, M., Dimaio, F., Anishchenko, I. et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science 373, 871–876 (2021). https://doi.org/10.1126/science.abj8754.
[24] Mirdita, M., Schütze, K., Moriwaki, Y. et al. ColabFold: making protein folding accessible to all. Nat Methods 19, 679–682 (2022). https://doi.org/10.1038/s41592-022-01488-1.
[25] Ahdritz, G., Bouatta, N., Kadyan, S. et al. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. bioRxiv 2022.11.20.517210 (2022). https://doi.org/10.1101/2022.11.20.517210.
[26] AlQuraishi, M. Protein-structure prediction revolutionized. Nature 596, 487–488 (2021). https://doi.org/10.1038/d41586-021-02265-4.
[27] Cusack, S., Eustermann, S., Kleywegt, G. et al. Great expectations – the potential impacts of AlphaFold. DB, EMBL News, 21. Jul. 2021. https://www.ebi.ac.uk/about/news/perspectives/alphafold-potential-impacts.

2.1. Einschätzungen zur Zukunft der Strukturbiologie

Eine gemeinschaftlich erarbeitete Einschätzung („community assessment“) der innerwissenschaftlichen Anwendungspotenziale von AlphaFold2 kommt zu dem Schluss, dass „advances are likely to have a transformative impact in structural biology and broader life-science research“[28] und ähnlich Varadi (2023):[29] „The arrival of the new generation of accurate protein structure prediction tools is a transformative time for structural biology, structural bioinformatics, drug discovery and many other fields of the life sciences.”

Robinson (2021) entwickelt eine Zukunftsvision mit einem Zeithorizont von 15 Jahren[30] und konstatiert: „The potential applications of AlphaFold2 are only limited by one’s imagination“. Für die unmittelbare Zukunft verweist Robinson konkret auf den Sprung von der Vorhersage von Proteinstrukturen zur Vorhersage von Proteinfunktionen mit primären Anwendungen in der Entwicklung und dem Design von Medikamenten.[31]

2.2. Forschung an den noch bestehenden Einschränkungen bei der Proteinstrukturvorhersage

Bei aller Begeisterung über die Fortschritte bei der Vorhersage von Proteinstrukturen, die sich auch in der wissenschaftlichen Literatur widerspiegelt, bestehen noch einige Einschränkungen.[32,33] Diese werden umgekehrt als Chancen für künftige Fortschritte betrachtet, vgl. dazu auch (Robinson, 2021):

- Viele Proteine existieren als Komplexe mit anderen Proteinen, Nukleinsäuren (DNA oder RNA) oder Liganden.[34] AlphaFold2 sagt(e) (zunächst) keine 3D-Strukturen für Protein-Protein- oder Protein-DNA/RNA/Ligand-Komplexe voraus.[35] Im Oktober 2021 veröffentlichte DeepMind jedoch ein AlphaFold-Modell namens AlphaFold-Multimer, das speziell auf Proteinkomplexe (multimere Proteine bekannter Stöchiometrie) trainiert wurde. Das Team wendete es auf Tausende von Komplexen in der „Protein Data Bank“ (PDB) an und erreichte eine Vorhersage von rund 70 % der bekannten Protein-Protein-Wechselwirkungen und behielt gleichzeitig eine hohe Genauigkeit innerhalb der einzelnen Aminosäure-Ketten bei.[36,37] Es besteht aber noch weiterer Forschungsbedarf.
- Proteine sind keine statischen, sondern dynamische Systeme. D. h. sie nehmen in Abhängigkeit von ihrer Umgebung oder ihrem Zustand unterschiedliche Strukturen an, sogenannte Konformationen. AlphaFold2 erzeugt von einem gegebenen Protein nur eine Form, auch wenn von dem Protein mehrere Konformationen bekannt sind. Dies lässt viele Fragen über die Dynamik von Proteinen offen, die für das Verständnis ihrer biologischen Funktion wesentlich sind.
- Es gibt Regionen innerhalb von Proteinen, die von Natur aus ungeordnet oder in isolierter Form unstrukturiert sind. Für solche Regionen liefert AlphaFold2 meist eine Vorhersage mit einem niedrigen Wert für das Maß zur Vorhersagepräzision. AlphaFold2 kann allerdings hilfreich sein, um solche Regionen zu identifizieren.

---

[28] Akdel, M., Pires, D.E.V., Pardo, E.P. et al. A structural biology community assessment of AlphaFold2 applications. Nat Struct Mol Biol 29, 1056–1067 (2022). https://doi.org/10.1038/s41594-022-00849-w.

[29] Varadi, M., Bordin, N., Orengo, C. et al. The opportunities and challenges posed by the new generation of deep learning-based protein structure predictors. Current Opinion in Structural Biology 79, 102543 (2023). https://doi.org/10.1016/j.sbi.2023.102543.

[30] Robinson, S.L. Artificial intelligence for microbial biotechnology: beyond the hype. Vgl. Abschnitt “Looking to the future: the next 15 years”. Microbial Biotechnology 15, 1 (2021). https://doi.org/10.1111/1751-7915.13943

[31] Zur längerfristigen Perspektive der Zukunftsvision, vgl. Abschnitt 2.4.

[32] Vgl. (Cusack, 2021) FN 9; Abschnitt „Current limitations of the prediction method“. Bei dieser und insgesamt allen Aussagen zum Stand der Forschung ist der jeweilige Zeitpunkt der Aussage (und der zitierten Quelle) zu berücksichtigen. Bei dieser Auflistung an Einschränkungen ist dies Juli 2021. Zitierter Wissensstand meist zwischen Juli 2021 und Ende 2023.

[33] Tourlet, S., Radjasandirane, R., Diharce, J. et al. AlphaFold2 Update and Perspectives. Vgl. Abschnitt 4. BioMedInformatics 3, 378–390 (2023). https://doi.org/10.3390/biomedinformatics3020025.

[34] „Als Ligand wird in der Biochemie und in verwandten Wissenschaften ein Stoff bezeichnet, der an ein Zielprotein, beispielsweise einen Rezeptor, spezifisch binden kann.“ Wikipedia-Eintrag zu „Ligand“; https://de.wikipedia.org/wiki/Ligand_(Biochemie). Dabei wird als „Rezeptor (von lateinisch recipere ‚aufnehmen‘ bzw. ‚empfangen‘) […] in der Biochemie ein Protein oder ein Proteinkomplex bezeichnet, sofern daran Signalmoleküle binden können, die dadurch Signalprozesse im Zellinneren auszulösen vermögen.“ Wikipedia-Eintrag zu „Rezeptor“; https://de.wikipedia.org/wiki/Rezeptor_(Biochemie).

[35] Mit Blick auf Abbildung 1.2 und die dort erläuterten Begriffe könnte man sagen, dass AlphaFold2 Tertiärstrukturen vorhersagt nicht aber Quartärstrukturen.

[36] Evans, R., O’Neill, M., Pritzel, A. et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034 (2021). https://doi.org/10.1101/2021.10.04.463034.

[37] Callaway, E. What’s next for AlphaFold and the AI protein-folding revolution. Nature 604, 234–238 (2022). https://doi.org/10.1038/d41586-022-00997-5.

- AlphaFold2 erhebt nicht den Anspruch, den "Faltungspfad" vorherzusagen, d. h. es wird die finale Struktur des Proteins vorhergesagt, nicht aber welche Zwischenstrukturen auf dem Weg dahin von dem Protein durchlaufen werden.
- Training und Validierung von AlphaFold2 waren nicht auf die Auswirkungen von Mutationen ausgerichtet. Insbesondere deswegen wird nicht erwartet, dass die Auswirkungen von Punktmutationen erfasst werden, die ein Protein destabilisieren können.
- Viele virale Proteine fehlen noch.[38]
- AlphaFold kann keine posttranslationalen Modifikationen vorhersagen. Es gibt einige Dutzend solcher Modifikationen, die erst nach der Umwandlung der genetischen Information in die Aminosäuresequenz erfolgen und unterschiedliche Funktionen haben. Typische Beispiele sind Modifikationen zur Regulierung der Aktivität von Enzymen oder zur Abspaltung von Teilen der Aminosäurekette, die die Funktion haben, den Bestimmungsort oder den Transportweg eines Proteins festzulegen, und die nach der Ankunft des Proteins am Bestimmungsort nicht mehr benötigt werden. Fehler bei posttranslationalen Modifikationen können aber auch die Ursache von Erkrankungen sein, etwa wenn Proteine dadurch ihre eigentliche Funktion verlieren.[39]

2.3. Zukunftsvision: Molekulardynamische Simulation ganzer Zellen

Stevens et al. (2023) entwerfen eine Zukunftsvision für die Lebenswissenschaften ausgehend von der folgenden Überlegung.[40] Das ultimative, ideale Mikroskop, das auf eine Zelle gerichtet wird, müsste die Dynamik aller Bestandteile der Zelle mit atomarer Auflösung sichtbar machen können. Tatsächliche Mikroskope sind durch physikalische Schranken limitiert. Im Gegensatz dazu steht die Computermikroskopie laut der zitierten Arbeit kurz davor, die beschriebene Herausforderung zu meistern. Es wird ein integrativer Ansatz vorgestellt, um eine bestimmte Minimalzelle in voller Komplexität zu modellieren. In diesem Ansatz sind Proteinstrukturvorhersagen von AlphaFold2 enthalten. Die Autoren sehen dadurch die Möglichkeit eröffnet, die räumlich-zeitliche Entwicklung von ganzen Zellen im Computer zu untersuchen und in Zukunft auf andere Zelltypen auszuweiten.

2.4. Neue KI-Anwendungen in den Lebenswissenschaften

In der Literaturanalyse wurde immer wieder deutlich, dass der Erfolg von AlphaFold eine Ermutigung und Inspiration darstellt, KI-Methoden auch in anderen Kontexten und mit anderen Zielen innerhalb der Lebenswissenschaften einzusetzen. Ewan Birney benennt als Direktor von EMBL-EBI drei Erfolgsfaktoren:[41]

1. KI-Vision und Talent

2. Umfassende, wahrhaft offene Daten, die von einer Gemeinschaft bereitgestellt und gut gepflegt werden

3. Gut durchdachter und durchgeführter Wettbewerb mit offenen Regeln und aussagekräftigen Erfolgskennzahlen

Birney betont, dass es aus seiner Sicht zukünftig noch viele andere biologische und medizinische Fragestellungen gibt, die mit einer ähnlichen KI-Vision und dem Zugang zu großen Mengen an Daten hoher Qualität gelöst werden können. Er weist explizit auf einige Datensätze hin, die am EMBL-EBI gepflegt werden: Ensembl mit Genom-Daten, Reactome mit Daten zu biochemischen Stoffwechsel- und Signalwegen auf zellulärer Ebene, ChEMBL mit Daten zu bioaktiven Molekülen mit medikamentenähnlichen Eigenschaften.

Die Zukunftsvision von Robinson (2021) mit dem Zeithorizont 2040 geht in die gleiche Richtung. Gerade weil biologische Experimente mit lebenden Organismen sich nur selten exakt so verhalten wie erwartet, wird Potenzial gesehen, KI-Modelle in einer Abfolge der Schritte „design-build-test-learn“ zu trainieren, um ein besseres Verständnis und bessere Vorhersagen zu erhalten. Auf diesem Wege soll KI eine verbesserte Reproduzierbarkeit und Vorhersagbarkeit in biologischen Experimenten ermöglichen.

---

[38] Bassani, D. & Moro, S. Past, Present, and Future Perspectives on Computer-Aided Drug Design Methodologies. Molecules 28, 3906 (2023). https://doi.org/10.3390/molecules28093906.
[39] Vgl. Lehrtext “Posttranslationale Proteinmodifikationen” unter https://www.lecturio.de/artikel/medizin/posttranslationale-proteinmodifikationen.
[40] Stevens, J.A., Grünewald, F., van Tilburg, P.A.M. et al. Molecular dynamics simulation of an entire cell. Front Chem 11, 1106495 (2023). https://doi.org/10.3389/fchem.2023.1106495.
[41] Birney, E. AlphaFold: where will the next AI breakthrough come from? (2022). https://www.ebi.ac.uk/about/news/perspectives/alphafold-where-will-the-next-ai-breakthrough-come-from.

Weitere Beispiele aus der Literaturanalyse beziehen sich auf die Zuordnung biologischer Funktionen zu Proteinen (Annotation),[42] die Identifizierung von Protein-Protein-Interaktionen (PPI) im Zusammenhang mit humanen Viren,[43] Deep-Learning-Anwendungen[44] in mehreren biowissenschaftlichen Bereichen (Proteinfunktionsvorhersage, Genom-Engineering, Systembiologie und Datenintegration sowie der Erforschung von Abstammungsbeziehungen),[45] Stammzellenforschung,[46] synthetische Biologie[47] und jenseits der Lebenswissenschaften bspw. das Feld der Polymerforschung.[48]

2.4.1. Paradigmenwechsel für die angewandte Forschung aller Branchen absehbar

Allgemein in der analysierten Literatur und besonders in den obigen Beispielen wird ein klarer Trend erkennbar, der eine Abkehr vom Prinzip des Reduktionismus in der Wissenschaft bedeutet. Die Fragestellung der Proteinfaltung stellt damit ein Musterbeispiel für eine Aufgabenstellung dar, die trotz intensiver Bemühungen über Jahrzehnte hinweg durch Anwendung der zugrundeliegenden chemischen und physikalischen Naturgesetze nicht gelöst werden konnte. Der Erfolg von AlphaFold beruht vielmehr auf einer Mustererkennung (im Abgleich multipler Übereinstimmungen von Aminosäuresequenzen (MSA - „Multiple Sequence Alignment“ mit Regionen hoher oder geringer Ähnlichkeit zu experimentell ermittelten Proteinstrukturen), ohne den Rückgriff auf die relevanten Naturgesetze.[49] In der Literaturanalyse wird deutlich, dass an vielen Stellen versucht wird, von diesem Erfolg zu lernen und ihn auf andere Aufgabenstellungen zu übertragen.

Dieser sich nun abzeichnende Paradigmenwechsel kann insbesondere für die anwendungsorientierte Forschung und die Unternehmensforschung bedeutsam werden, da dort die zu untersuchenden Systeme durch die jeweilige Anwendung definiert und häufig komplexer als gewünscht sind. Das sich abzeichnende KI-basierte Paradigma eröffnet einen neuen Zugang zu diesen anwendungsbezogenen Forschungsfragen mit großem Potenzial.[50] Den Anforderungen an die Vertraulichkeit der Daten, die in der Industrie oft höher sind als in der akademischen Forschung, kann dabei möglicherweise durch das Konzept des Föderalen Lernens Rechnung getragen werden. Dabei werden Algorithmen an verschiedene Datenzentren geschickt und dort jeweils lokal trainiert.[51]

Die hier vorgestellten Überlegungen kulminieren in der Zukunftsvision autonomer Labore,[52] die auch als automatisierte oder selbstgesteuerte Labore (engl. “self driving labs” SDL) bzw. Robotik-Labore bezeichnet werden[53,54] und die Zukunft der Wissenschaft darstellen könnten.[52] So sollen Robotik-Labore einfache Formen der wissenschaftlichen Forschung automatisiert durchführen, indem sie selbstständig neue Hypothesen aufstellen, Experimente zur Prüfung dieser Hypothesen festlegen und dann Laborroboter steuern und programmieren, um die Experimente physisch durchzuführen. Das Robotik-Labor kann die Ergebnisse dieser Experimente untersuchen, analysieren und Wahrscheinlichkeiten für die Richtigkeit der Hypothesen auf der Grundlage der Experimente

[42] de Crécy-lagard, V., de Hegedus, R.A., Arighi, C. et al. A roadmap for the functional annotation of protein families: a community perspective. Database (2022). https://doi.org/10.1093/database/baac062.
[43] Yang, X., Yang, S., Ren, P. et al. Deep Learning-Powered Prediction of Human-Virus Protein-Protein Interactions. Frontiers in Microbiology 13 (2022). https://doi.org/10.3389/fmicb.2022.842976.
[44] Deep Learning (DL), dt. „tiefes Lernen“ als Zweig der KI und besonders des maschinellen Lernens, vgl. Wikipedia-Eintrag „Deep Learning“; https://de.wikipedia.org/wiki/Deep_Learning.
[45] Sapoval, N., Aghazadeh, A., Nute, M.G. et al. Current progress and open challenges for applying deep learning across the biosciences. Nat Commun 13, 1728 (2022). https://doi.org/10.1038/s41467-022-29268-7.
[46] Ouyang, J.F., Chothani, S. & Rackham, O.J.L. Deep learning models will shape the future of stem cell research. Stem Cell Reports 18, 1, P6–12 (2023). https://doi.org/10.1016/j.stemcr.2022.11.007.
[47] Synthetische Biologie bezeichnet ein interdisziplinäres Feld mit dem Ziel, künstliche biologische Systeme zu erzeugen und technisch zu optimieren (vgl. Wikipedia-Eintrag „Synthetische Biologie“; https://de.wikipedia.org/wiki/Synthetische_Biologie); zur Anwendung von KI in dem Kontext, vgl. Beardall, W.A.V., Stan, G.-B. & Dunlop, M.J. Deep Learning Concepts and Applications for Synthetic Biology. GEN Biotechnology 1, 4, 360–371 (2022). https://www.liebertpub.com/doi/abs/10.1089/genbio.2022.0017.
[48] Martin, T.B. & Audus, D.J. Emerging Trends in Machine Learning: A Polymer Perspective. ACS Polymers Au 3, 3, 239–258 (2023). https://doi.org/10.1021/acspolymersau.2c00053.
[49] Ebenso ist es bei den aktuell so populären KI-Sprachmodellen, deren Erfolge nicht auf der Kenntnis und Implementierung grammatischer Regeln, sondern auf Mustererkennung beruhen.
[50] Messeri & Crockett (2024) warnen allerdings vor Übertreibungen, so dass es wohl auf die richtige Balance und Kombination von Methoden ankommt. Messeri, L. & Crockett, M.J. Artificial intelligence and illusions of understanding in scientific research. Nature 627, 49–58 (2024). https://doi.org/10.1038/s41586-024-07146-0.
[51] OECD. Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. Paris: OECD Publishing (2023). https://doi.org/10.1787/a8d820bd-en, S. 172f.
[52] Ibid. S. 114f.
[53] Abolhasani, M. & Kumacheva, E. The rise of self-driving labs in chemical and materials sciences. Nat Synth 2, 483–492 (2023). https://doi.org/10.1038/s44160-022-00231-0.
[54] Martin, H.G., Radivojevic, T., Zucker, J. et al. Perspectives for self-driving labs in synthetic biology. Current Opinion in Biotechnology 79, 102881 (2023). https://doi.org/10.1016/j.copbio.2022.102881.

anpassen.[52] Solche Labore können gerade in der angewandten Forschung der Zukunft eine zunehmende Rolle spielen, weil sie rund um die Uhr arbeiten können und in ihren Aufgaben schneller, präziser und kostengünstiger sind als Menschen und ohne Ermüdung arbeiten können.[52]

3. Protein-Engineering

Beim „Protein-Engineering“ [55] werden Proteine gezielt für spezifische Anwendungen entworfen und gebaut. Beispiele für heutige Anwendungen sind optimierte Amylasen zur Aufspaltung von Stärke in Zucker in Waschmitteln zur Fleckenentfernung. Die Fortschritte bei der Proteinstrukturvorhersage könnten das Protein-Engineering schneller, günstiger und präziser machen.

Zudem ermöglicht die Proteinstrukturvorhersage auch viel grundlegender das Design neuer Proteine zu ganz spezifischen, neuen Zwecken, wie etwa der Erzeugung von neuen Oberflächeneigenschaften in Beschichtungen oder dem Abbau von Kunststoffen (s.a. Kapitel 4).

Dabei spielen Ansätze, die den Bedarf an experimentellen Daten minimieren, eine wichtige Rolle, wie bspw. integrierte Design-Build-Test-Learn-Zyklen.[56] Diese Ansätze werden nach Einschätzung von Freschlin et al. (2022) dazu beitragen, die Grenzen des Protein-Engineerings auszuweiten. Dies soll es ermöglichen, Proteine mit Funktionen zu entwickeln, die der Leistung und Komplexität natürlicher Systeme gleichkommen, und dafür sorgen, dass das Protein-Engineering Beiträge zur Lösung großer, globaler Herausforderungen in den Bereichen menschlicher Gesundheit, Landwirtschaft, Umwelt und Energie leisten wird.[57] Im ähnlichen Sinne argumentieren auch Kouba et al. (2022) und heben die Bedeutung der Sammlung und Kuratierung von offenen Daten („open data“) als Erfolgsfaktoren hervor, zusammen mit Vergleichswettbewerben und neuen anspruchsvollen Daten zur Proteindynamik, die völlig neue Wege im Protein-Engineering eröffnen und somit die Biotechnologie der Zukunft prägen könnten.[58]

3.1. Kohlenstoff-Fixierung

Sogenannte C1-Gase, darunter Kohlendioxid (CO_2) und Kohlenmonoxid (CO), spielen eine wesentliche Rolle in der Klimakrise. Daher stellt sich die Frage, wie sich C1-Gase binden und recyceln lassen. Bae et al. (2022) diskutieren einen vielversprechenden Ansatz in einer Übersichtsarbeit über die Verwendung von Mikroorganismen als Biokatalysatoren zur Umwandlung von C1-Gasen in werthaltige Chemikalien.[59] Acetogene Bakterien (Acetogene) haben dabei besondere Beachtung gefunden, denn in ihnen ist der sog. Wood-Ljungdahl-Weg (WL) aktiv, der nicht nur CO_2 sondern auch CO bindet. Die Umwandlung von C1-Gasen zur Produktion verschiedener Biochemikalien durch technische Acetogene wird zwar in industriellem Maßstab aktuell bspw. durch das Unternehmen LanzaTech in sechs kommerziellen Anlagen realisiert.[60] Mit dadurch bislang eingesparten CO_2-Emissionen von ca. 400.000 Tonnen bestehen aber noch große Potenziale. In der zitierten Arbeit (Bae, 2022) werden u. a. synthetische Ansätze diskutiert. Im Abschnitt zu den zukünftigen Perspektiven wird auf diese Möglichkeit Bezug genommen und aufgezeigt, dass die Anwendung von Tools zur Vorhersage von Proteinstrukturen (wie AlphaFold2 oder RoseTTAFold) auch Schritte zur Aufdeckung der Funktionen und Mutationen vieler Proteine im WL-Weg erlaubt. Insofern erscheint es Bae et al. möglich, dass Proteine, die am WL-Weg beteiligt sind, manipuliert oder ein neuartiger C1-Fixierungsweg mit Hilfe von Deep-Learning-Tools entworfen und konstruiert werden könnte.

3.2. Bioremediation von Öl

Enzyme, die Alkane oxidieren, spielen eine große Rolle im globalen Kohlenstoffkreislauf sowie der Bioremediation von Öl, die auch mit dem Begriff der biologischen Sanierung von Öl-Verschmutzungen bezeichnet wird. Die Alkan-Monooxygenase (AlkB) oxidiert die meisten der Alkane mittlerer Länge in der Umwelt. Williams und Austin (2022)

[55] Unter Protein-Engineering wird die Entwicklung und Herstellung von unnatürlichen Polypeptiden verstanden, häufig durch Veränderung von Aminosäuresequenzen, die in der Natur vorkommen; vgl. https://www.nature.com/subjects/protein-engineering. Unter Proteindesign (vgl. Kapitel 4) wird in diesem Dokument Forschung und Entwicklung an künstlichen Proteinen „de novo“ (von Grund auf) verstanden, also an Proteinen, die weiter von natürlich vorkommenden Proteinen entfernt sind. Die Begriffsabgrenzung ist allerdings nicht immer eindeutig und es finden sich in der Literatur auch Beispiele, in denen die Begriffe Protein-Engineering und Proteindesign mit anderen Bedeutungen verwendet werden.
[56] Vgl. auch Kapitel 12, Abschnitt „Weitere Ebene der Digitalisierung im Innovations-Modus der Bioökonomie“
[57] Freschlin, C.R., Fahlberg, S.A. & Romero, P.A. Machine learning to navigate fitness landscapes for protein engineering. Current Opinion in Biotechnology 75, 102713 (2022). https://doi.org/10.1016/j.copbio.2022.102713.
[58] Kouba, P., Kohout, P., Haddadi, F. et al. Machine Learning-Guided Protein Engineering. ACS Catal 13, 21, 13863–13895 (2023). https://doi.org/10.1021/acscatal.3c02743.
[59] Lee, H., Bae, J., Jin, S. et al. Engineering Acetogenic Bacteria for Efficient One-Carbon Utilization. Front Microbiol 13, 865168 (2022). https://doi.org/10.3389/fmicb.2022.865168.
[60] Vgl. Webseite des Unternehmens https://ir.lanzatech.com/.

weisen darauf hin, dass ein Interesse besteht, synthetische "autarke" Monooxygenasen herzustellen, wofür ein Verständnis der relevanten Proteinstrukturen wesentlich sein könnte. Diese synthetischen Fusionsproteine könnten u. a. hochspezialisierte Funktionen und einen leichteren biotechnologischen Einsatz haben.[61]

In einer Struktur-Untersuchung von *Fontimonas thermophila* AlkB (FtAlkB) (Guo, 2023), in die auch AlphaFold-Ergebnisse eingeflossen sind, wird die Architektur dieser Familie wichtiger Enzyme entschlüsselt. Es stellt sich heraus, dass AlkB dabei eine Proteinfaltung annimmt, die zuvor experimentell nicht beobachtet wurde. Insofern stellen diese Ergebnisse eine Grundlage für zukünftige Bemühungen, AlkB für biotechnologische Anwendungen im Sinne des Protein-Engineerings zu modifizieren. Die selektive Aktivierung von C-H-Bindungen gilt als heiliger Gral der Chemie (Guo, 2023). Daher ist die Art und Weise, wie Mitglieder der betrachteten Superfamilie die C-H-Aktivierung ermöglichen, seit langem von großem Interesse. Die angesprochene Strukturuntersuchung verschafft hierzu wertvolle Einblicke.[62]

[61] Williams, S.C. & Austin, R.N. An Overview of the Electron-Transfer Proteins That Activate Alkane Monooxygenase (AlkB). Front Microbiol 13, 845551 (2022). https://doi.org/10.3389/fmicb.2022.845551.
[62] Guo, X., Zhang, J., Han, L. et al. Structure and mechanism of the alkane-oxidizing enzyme AlkB. Nat Commun 14, 2180 (2023). https://doi.org/10.1038/s41467-023-37869-z.

4. Proteindesign

Die Anwendung der Proteinstrukturvorhersage wird mit wachsender Genauigkeit der Vorhersagen immer interessanter für das Proteindesign - also, um neue Proteine zu entwickeln, die bestimmte gewünschte Funktionen aufweisen (Bordin, 2022).[63] In diesem Sinne wurden erste Deep-Learning-Ansätze entwickelt, die neue Proteine "halluzinieren"[64] und bei denen Modelle zur Proteinstrukturvorhersage bestätigen, dass sie sich zu plausiblen Strukturen falten können. Solche Tools können demnach innerhalb von Millisekunden neue Proteinsequenzen komplett „de novo“ oder, ähnlich wie bei einer Textvervollständigung, abhängig von einer vorgegebenen Sequenzeingabe, erzeugen. Das bedeutet, dass solche Tools zusammen mit Modellen zur Proteinstrukturvorhersage Millionen von potenziell faltbaren, ML-generierten Sequenzen[65] zuverlässig *in silico* vorfiltern können. Dies spart Energie, Zeit und Ressourcen, so dass In-vitro- und In-vivo-Experimente nur auf besonders aussichtsreiche Fälle konzentriert werden können. Zwar funktionieren in diesem Ansatz längst nicht alle Designs, aber in der Literatur werden erste Erfolge benannt.[66]

Bordin et al. äußern die Zukunftserwartung, dass Ansätze, die Proteinsequenzen ausgehend von einer 3D-Struktur generieren,[67] in der Post-AlphaFold2-Ära an Bedeutung gewinnen werden. Dies entspricht in bestimmter Hinsicht der entgegengesetzten Richtung der ursprünglichen Faltungsfragestellung. Diese Ansätze ermöglichen es, beim Proteindesign anders gestellte Aufgaben zu untersuchen. Beispielsweise kann die Suche nach Proteinen, die an bestimmte (Stellen von) Zielproteinen binden sollen, von der geometrischen Struktur des Zielproteins ausgehen.[68] Das Proteindesign beginnt dann mit dem geometrischen Design einer Struktur, die zum betrachteten Zielprotein passt bzw. dazu komplementär ist. Danach folgt die Betrachtung der inversen Faltung als Suche nach einer Aminosäuresequenz, die sich in die Zielgeometrie falten kann. Dies kann etwa für Enzyme oder Therapeutika von Interesse sein.[69]

In einem ähnlichen Sinne artikuliert Rudden (2022), dass *Deep Learning* bereits gegenwärtig die Fähigkeiten im Proteindesign ausweitet und auch dazu beiträgt, dass das maschinelle Lernen selbst kontinuierliche Verbesserungen erfährt.[70] Vor diesem Hintergrund wird die Erwartung geäußert, dass das Design von Proteinen künftig mit erheblicher Flexibilität an Konformationen und somit verbesserter funktioneller Kapazität möglich sein wird.[71]

4.1. Mit ChatGPT vergleichbare Programme für das Proteindesign

Im Bereich der KI hat seit spätestens November 2022 mit der Veröffentlichung von ChatGPT die sog. generative KI für große, öffentliche Aufmerksamkeit gesorgt. Sie wird „generativ“ genannt, weil durch knappe Eingaben (sog. Prompts) ein gewünschtes Ergebnis generiert wird. Dies kann bei den sog. Sprachmodellen - wie eben ChatGPT - die Ausgabe von realistischen Texten sein, während bei bestimmten der sog. Diffusionsmodelle die Ausgabe von realistischen Bildern erfolgt.[72]

Es zeigt sich nun, dass sich die dabei zugrundeliegenden Prinzipien auch im Kontext von Aminosäuresequenzen und Proteinen erfolgreich einsetzen lassen. Denn mit dem Verständnis von Proteinen als Abfolge von Aminosäuren lässt sich eine gewisse formale Ähnlichkeit zu Wörtern als einer Abfolge von Buchstaben bzw. zu Texten als einer Abfolge von Wörtern ableiten. In einem Übersichtsartikel kommen Ferruz und Höcker (2022) für Sprachmodelle zu

[63] Bordin, N., Dallago, C., Heinzinger, M. et al. Novel machine learning approaches revolutionize protein knowledge. Trends in Biochemical Sciences 48, 4, 345–359 (2023). https://doi.org/10.1016/j.tibs.2022.11.001.
[64] Callaway, E. Scientists are using AI to dream up revolutionary new proteins. Nature 609, 661–662 (2022). https://doi.org/10.1038/d41586-022-02947-7.
[65] ML – Maschinelles Lernen.
[66] Hie, B.L., Shanker, V.R., Xu, D. et al. Efficient evolution of human antibodies from general protein language models. Nat Biotechnol (2023). https://doi.org/10.1038/s41587-023-01763-2.
[67] Diese Ansätze werden in der Literatur als „inverse folding“ oder „protein sequence design“ bezeichnet.
[68] Bryant, P. & Elofsson, A. Peptide binder design with inverse folding and protein structure prediction. Commun Chem 6, 229 (2023). https://doi.org/10.1038/s42004-023-01029-7.
[69] Amani, K. What Is Inverse Folding & How To Practically Apply It. Blog des Unternehmens Neuosnap (2023). https://neurosnap.ai/blog/post/65908e76104e7841a40c3187.
[70] Eine laufend aktualisierte Literaturliste bietet: List of papers about Proteins Design using Deep Learning. Verfügbar unter: https://github.com/Peldom/papers_for_protein_design_using_DL/tree/main.
[71] Rudden, L.S.P., Hijazi, M. & Barth, P. Deep learning approaches for conformational flexibility and switching properties in protein design. Front Mol Biosci 9, 928534 (2022). https://doi.org/10.3389/fmolb.2022.928534.
[72] Die bekanntesten Beispiele dafür sind etwa Stable Diffusion, DALL-E und Midjourney.

Proteinen zu dem Schluss, dass sie das Feld des Proteindesigns revolutionieren und neue Lösungen für viele aktuelle und zukünftige gesellschaftliche Herausforderungen bieten werden. Mit dem Artikel sollen sowohl die Forschungsgemeinschaften der Künstlichen Intelligenz als auch der Biochemie erreicht werden, um die Anwendung von NLP-Methoden in der Proteinforschung zu fördern.[73] Ein konkretes Beispiel für diese Art stellen Ferruz et al. (2022) mit der Beschreibung des Modells ProtGPT2 vor.[74]

Gleiche Überlegungen werden auch für Diffusionsmodelle angestellt: Die Arbeitsgruppe von David Baker an der Universität von Washington hat etwa untersucht, ob sich auf Basis der verbesserten Proteinstrukturvorhersage mit Diffusionsmodellen das Proteindesign verbessern lässt. Hiervon ausgehend wurde das Modell RFdiffusion entwickelt.[75] Ein Nature-Feature[76] erläutert zu diesem und ähnlichen KI-Modellen, dass sie auf denselben Prinzipien beruhen wie die oben angesprochenen KI-Programme, die auf eine Texteingabe hin realistische Bilder erzeugen. Solche "Diffusions"-Netzwerke werden mit Daten trainiert, seien es Bilder oder eben Proteinstrukturen, die dann immer stärker durch Rauschen überlagert werden und schließlich keine Ähnlichkeit mehr mit dem Ausgangsbild oder der Ausgangsstruktur haben. Das Netz lernt dann, die Daten zu "entrauschen". Netze wie RFdiffusion wurden mit Zehntausenden von Proteinstrukturen aus der Protein Data Bank (PDB) trainiert. Soll das Netzwerk ein neues Protein gestalten, wird mit einer zufälligen Auswahl von Aminosäuren begonnen. Darauf folgen einige Zyklen der „Entrauschung“, wodurch etwas entsteht, das einem realen Protein ähnelt. Die so generierten Proteine könnten die Basis für Impfstoffe, Therapeutika oder Biomaterialien bilden.[76]

4.2. Proteindesign bestimmter Proteinklassen

4.2.1. Peptid-Design

Generative KI-Modelle können Daten erzeugen, die über den Datenpool hinausgehen, in dem die zugrundeliegenden Trainingsdaten enthalten waren.[77] Sie sind in diesem Sinne ein effizientes und schnelles Werkzeug für die Erkundung des riesigen Suchraums hochdimensionaler Daten wie DNA-/Proteinsequenzen und sie erleichtern das Design von Biomolekülen mit gewünschten Funktionen. Wan et al. (2022) geben einen Überblick über das aufstrebende Feld der generativen Modelle in ihrer Anwendung auf Peptide. Insbesondere werden mehrere weit verbreitete generative Modelle sowie ihre Anwendungen für das Design von Peptiden mit bestimmten Eigenschaften vorgestellt: z. B. antimikrobiell, krebshemmend oder zelldurchdringend für den Wirkstofftransport. Peptid-Design wird zukünftig auch im Hinblick auf Biomaterialien und deren Anwendungen, bspw. im biomedizinischen Kontext oder zum Frostschutz, von Interesse sein.[78]

4.2.2. Metalloproteine

Fortschritte in den Bereichen der computergestützten Berechnung, der Gensynthese, der Strukturbiologie und der Evolution haben die Möglichkeiten, Proteine *de novo* zu entwerfen, erheblich verbessert. Dabei können laut Chalkley et al. (2022) die Funktionen der Klasse der Metalloproteine, dies sind Proteine, die ein oder mehrere Metalle enthalten, als ein wichtiger Vergleichsmaßstab dienen, wie weit das Verständnis der zugrunde liegenden Prinzipien zum Struktur-Funktion-Zusammenhang von Proteinen fortgeschritten ist. Es wird die Zukunftserwartung geäußert, dass weitere wissenschaftliche-technische Fortschritte das De-novo-Design funktioneller Metalloproteine zunehmend zur Routine machen werden, wodurch Anwendungen in einer Vielzahl von Bereichen eröffnet werden, wie der grünen Katalyse, der Sensorik und Therapeutik.[79] Mattocks et al. (2023) untersuchen etwa das Protein Lanmodulin, das selektiv an bestimmte Metalle aus der Gruppe der seltenen Erden bindet und erwarten auf längere

[73] Ferruz, N. & Höcker, B. Controllable protein design with language models. Nature Machine Intelligence (2022). https://doi.org/10.1038/s42256-022-00499-z.
[74] Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. Nature Communications (2022). https://doi.org/10.1038/s41467-022-32007-7.
[75] Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion, Nature (2023). https://dx.doi.org/10.1038/s41586-023-06415-8.
[76] Callaway, E. AI tools are designing entirely new proteins that could transform medicine. Nature 619, 236–238 (2023). https://doi.org/10.1038/d41586-023-02227-y.
[77] Wan, F., Kontogiorgos-Heintz, D. & de la Fuente-Nunez, C. Deep generative models for peptide design. Digital Discovery 1, 195 (2022). https://pubs.rsc.org/en/content/articlelanding/2022/DD/D1DD00024A.
[78] Wang, Y., Stebe, K.J., de la Fuente-Nunez, C. et al. Computational Design of Peptides for Biomaterials Applications. ACS Appl Bio Mater (2023). https://doi.org/10.1021/acsabm.2c01023.
[79] Chalkley, M.J., Mann, S.I. & DeGrado, W.F. De novo metalloprotein design. Nat Rev Chem 6, 31–50 (2022). https://doi.org/10.1038/s41570-021-00339-5.

Sicht das gezielte Design[80] von Liganden, die in Trennprozessen für seltene Erden genutzt werden könnten, die häufig Bestandteil von Elektronikschrott sind.[81]

4.2.3. Optoproteine

Månsson et al. (2022) befassen sich in einer Übersichtsarbeit mit den Grundprinzipien und Anwendungen von OptoGelen.[82] Darunter verstehen sie Hydrogele mit lichtprogrammierbaren Eigenschaften, die mit natürlich vorkommenden, photoschaltbaren Proteinen ("Optoproteinen") ausgestattet sind. Licht wird als primäre Energiequelle der Erde beschrieben und erläutert, dass die Evolution deshalb hochgradig elaborierte Funktionen wie den Phototropismus und die Steuerung des Tag-Nacht-Rhythmus entwickelt hat. Diese Funktionen werden durch Optoproteine vermittelt. Es gibt viele verschiedene Optoproteine, die auf jeweils spezifische Wellenlängen aus dem gesamten Spektrum des sichtbaren Lichts von ultraviolett bis infrarot reagieren, indem sie ihre Struktur verändern, wodurch Signale innerhalb von Zellen übertragen werden. Durch den Einsatz von Optoproteinen in Strukturen außerhalb biologischer Zellen können diese Strukturen durch Licht geschaltet oder gesteuert werden.[83] Auch wenn es andere Wege gibt, Materialeigenschaften dynamisch zu steuern, haben lichtempfindliche Proteine mehrere klare Vorteile, darunter die präzise räumlich-zeitliche Steuerbarkeit, die Reversibilität, die biologische Abbaubarkeit und die Biokompatibilität. Es existiert ein großer Raum an Gestaltungsmöglichkeiten, der laut Månsson et al. bislang nur wenig ausgeschöpft wurde. Zusätzlich wird erwartet, dass die Fortschritte in der Bioinformatik und der Proteinstrukturvorhersage in Zukunft die schnelle Entdeckung neuer Optoproteine ermöglichen werden. Anwendungen werden z. B. in der Mechanobiologie und im Engineering von Organoiden gesehen,[82] worunter künstliche organähnliche Mikrostrukturen verstanden werden. Eine Anwendung von Optoproteinen in Zellen wird in Abschnitt 11.2.2 dargestellt.

[80] Choi, T.S. & Tezcan, F.A. Overcoming universal restrictions on metal selectivity by protein design. Nature 603, 522–527 (2022). https://doi.org/10.1038/s41586-022-04469-8.
[81] Mattocks, J.A., Jung, J.J. & Lin, C.Y. et al. Enhanced rare-earth separation with a metal-sensitive lanmodulin dimer. Nature 618, 87–93 (2023). https://doi.org/10.1038/s41586-023-05945-5.
[82] Månsson, L.K., Pitenis, A.A. & Wilson, M.Z. Extracellular Optogenetics at the Interface of Synthetic Biology and Materials Science. Frontiers in Bioengineering and Biotechnology 10 (2022). https://doi.org/10.3389/fbioe.2022.903982.
[83] Als Modell für eine solche Struktur wird ein Netz aus Kollagen betrachtet, das mit Optoproteinen ausgestattet ist und sich unter dem Einfluss von blauem Licht versteift und im Dunkeln weich wird. Es wird erwähnt, dass diese Struktur als Schnittstellenmaterial in der Prothetik angewendet werden könnte, vgl. Hopkins, E., Valois, E., Stull, A. et al. An Optogenetic Platform to Dynamically Control the Stiffness of Collagen Hydrogels. ACS Biomater Sci Eng 7, 2, 408–414 (2020). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8544164/.

5. Biokatalyse – Enzyme

Enzyme bewerkstelligen und beschleunigen den Ablauf bestimmter biochemischer Reaktionen und sind wesentlich für den Stoffwechsel von Organismen, vgl. Kapitel 1. Weil sie Biomoleküle sind und durch die jeweiligen Reaktionen selbst nicht verändert werden, werden Enzyme auch als Biokatalysatoren bezeichnet.[84]

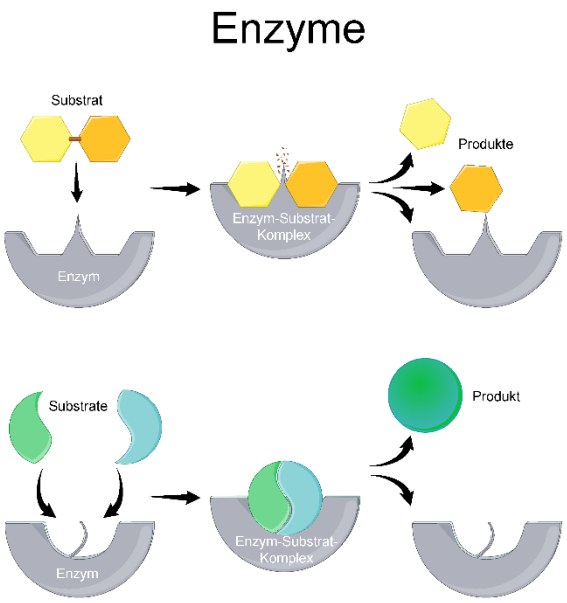

Abbildung 5.1: Schematische Darstellung der Funktion von Enzymen (Quelle: Adobe Stockphoto).

Diejenigen Stoffe, die durch Enzyme umgesetzt werden, werden als Substrat bezeichnet. Die Stoffe, die am Ende einer enzymatischen Reaktion vorliegen, werden als Produkt bezeichnet. Es gibt unterschiedliche Enzyme für die Aufspaltung von Substraten in mehrere Produkte und für die Synthese von einem Produkt aus mehreren Substraten. Enzyme können nur jeweils ein Substrat umsetzen und bei dem jeweiligen Substrat auch nur eine bestimmte Wirkung erzielen - auch wenn dasselbe Substrat bspw. durch ein anderes Enzym auf eine andere Weise umgesetzt werden kann. Abbildung 5.1 zeigt den schematischen Ablauf enzymatischer Reaktionen nach dem Schlüssel-Schloss-Prinzip. Die katalytische Wirkung des Enzyms besteht darin, dass die sog. Aktivierungsenergie für die jeweilige Reaktion herabgesetzt wird durch die Bildung eines sog. Enzym-Substrat-Komplexes als Übergangszustand im Reaktionsablauf. In praktischer Hinsicht bedeutet dies, dass die katalysierte Reaktion bei niedrigerer Temperatur ablaufen kann. Hierin und in der Spezifität der Reaktionen liegt häufig die höhere Nachhaltigkeit von enzymatischen Reaktionen begründet, die auch für industrielle Anwendungen angestrebt wird.

Die Entdeckung neuer Enzyme zusammen mit dem Bestreben, chemische Prozesse nachhaltiger zu gestalten, haben in diesem Sinne zu einem gesteigerten Interesse der Industrie an biokatalytischen Verfahren zur Herstellung hochwertiger sowie chiraler chemischer Grundstoffe geführt.[85] Die Verbesserung von Methoden des Protein-Engineerings sowie von rechnerbasierten Werkzeugen hat zu erheblichen Fortschritten bei der Entdeckung und Herstellung von Enzymen für biokatalytische Prozesse geführt. Dennoch wird das Wissen über die Beziehung zwischen Enzymstruktur und -funktion als sehr lückenhaft beschrieben.[85] Faktoren für neue Erkenntnisse über Enzyme waren und sind: Molekularbiologie mit einfacheren oder automatisierbaren Prozessen wie Klonierung und Mutagenese; strukturelle Informationen aus den Methoden der Röntgenkristallographie, der NMR und der Kryo-EM bis zur aktuellen Möglichkeit, Strukturen von bisher nicht charakterisierten oder sogar nicht kristallisierbaren Enzymen vorherzusagen. Dies kann die Kombination von Enzymentdeckung und Protein-Engineering wesentlich verbessern, so dass Ding et al. (2022) die Zukunftserwartung äußern, dass Entwicklungen im Bereich des Protein-Engineering wahrscheinlich den größten Einfluss auf die industrielle Biokatalyse haben werden. Das Engineering

[84] Vgl. Lehrvideo „Enzyme“: https://studyflix.de/biologie/enzyme-2662.
[85] Ding, Y., Perez-Ortiz, G., Peate, J. et al. Redesigning Enzymes for Biocatalysis: Exploiting Structural Understanding for Improved Selectivity. Front Mol Biosci 9, 908285 (2022). https://doi.org/10.3389/fmolb.2022.908285.

von Enzymen kann dabei durch rationales Design erfolgen oder durch nicht-rationale Methoden wie die gerichtete Evolution. In einem ähnlichen Sinne äußern sich Xie und Warshel (2023), wonach die Durchbrüche bei der Vorhersage von Proteinstrukturen wirksame Werkzeuge für die Übersetzung von Proteinsequenzen in Strukturen bereitgestellt und damit generative Modelle für das De-novo-Design von Proteinen ermöglicht haben. Generative Diffusionsmodelle (s. o. Abschnitt 4.1) erlauben das Design von Proteinen einschließlich des Gerüsts der aktiven Zentren von Enzymen. Es wird das Fazit gezogen, dass die Konstruktion effizienter Enzyme *de novo* schwierig bleibt, da die Enzymkatalyse noch viele Rätsel aufgibt und dass physikbasierte Modelle sowohl verbessert als auch in generative Designs integriert werden müssen.[86]

5.1. Biokatalyse von PET und anderen Kunststoffen

In einer Übersichtsarbeit stellen Magalhães et al. (2023) den aktuellen Wissensstand zum Thema Biokonversion von PET und zukünftige Perspektiven dar:[87] Kunststoffe sind demnach sehr langlebige und weit verbreitete Materialien. Die gebräuchlichen Methoden des Abbaus, der Beseitigung und des Recyclings von Kunststoffen sind verbesserungsbedürftig. Inzwischen hat sich der biologische Abbau (durch Verwendung von Mikroorganismen für das Materialrecycling) als ernstzunehmende Alternative zu den bisher verwendeten Methoden etabliert.[87]

Es sind mittlerweile mehrere PET-abbauende Enzyme entdeckt worden. Basierend auf umfangreichen Arbeiten zur Strukturaufklärung und -charakterisierung sehen Magalhães et al. als sich abzeichnenden Trend für die Zukunft auf diesem Gebiet das Engineering und die Verbesserung dieser bisher gefundenen Enzyme. Da rationales Proteindesign eine genaue Kenntnis der Proteinfunktion und -struktur voraussetzt, wird der Zunahme hochaufgelöster Proteinstrukturen in der PDB-Datenbank und der Verbesserung von Strukturvorhersagetools wie AlphaFold eine große Bedeutung zugemessen und festgestellt, dass sie eine atomare/molekulare Perspektive auf die Struktur und Funktion der Enzyme zulassen. Verbesserungspotenziale werden u. a. in Bezug auf die thermische Stabilität, den enzymatischen Umsatz und die Effizienz gesehen. Obwohl mehrere der identifizierten Enzyme als sehr vielversprechend angesehen werden, wird (im Erscheinungsjahr des Artikels 2021) konstatiert, dass ihre Aktivitäten und thermischen Stabilitäten noch weit von den Schwellenwerten für eine effiziente und großtechnische Anwendbarkeit entfernt sind.[87] Das französische Unternehmen Carbios ist nach eigenen Angaben jedoch bereits heute das weltweit erste Unternehmen, das das enzymatische Recycling mit hoher Effektivität auf Plastik - und hier speziell auf PET[88] - anwendet.[89] In Pressemeldungen wird von der Eröffnung einer Demonstrationsanlage für das PET-Biorecycling aus Polyester-Textilien im Jahr 2023 berichtet[90] und die Eröffnung der weltweit ersten PET-Biorecycling-Fabrik für das Jahr 2025 angekündigt.[91]

In weiteren Arbeiten wird der Bogen zur allgemeineren Frage der Zirkularität von Plastik und dem biokatalytischen Abbau weiterer Kunststoffe gespannt. Dabei betonen Orlando et al. (2023) die Bedeutung von Spezialdatenbanken (Plastics Microbial Biodegradation Database, PlasticDB und PAZy) zu Sequenzen, Strukturen und Funktionen und heben das Potenzial von spezialisierten Strukturvorhersagetools hervor.[92] Herbert et al. (2022) ergänzen die Forderung, die Suche auf neue Kandidaten-Enzyme zu konzentrieren, die das Potenzial für sprunghafte

[86] Xie, W.J. & Warshel, A. Harnessing Generative AI to Decode Enzyme Catalysis and Evolution for Enhanced Engineering. bioRxiv (2023). https://doi.org/10.1101/2023.10.10.561808.
[87] Magalhães, R.P., Cunha, J.M. & Sousa, S.F. Perspectives on the Role of Enzymatic Biocatalysis for the Degradation of Plastic PET. Int J Mol Sci 22, 11257 (2021). https://doi.org/10.3390/ijms222011257.
[88] Grégory, A., Anglade, J., Gavalda, S. et al. Assessment of Four Engineered PET Degrading Enzymes Considering Large-Scale Industrial Applications. ACS Catalysis 13, 20, 13156–13166 (2023). https://doi.org/10.1021/acscatal.3c02922 sowie Unternehmenspressemeldung zum Erscheinen des Artikels; https://www.carbios.com/newsroom/en/carbios-demonstrates-superior-performance-of-its-enzyme-in-world-renowned-scientific-publication.
[89] Webseite des Unternehmens Carbios „Leverage the exceptional specificity of enzymes“: https://www.carbios.com/en/enzymes.
[90] Carbios. „Carbios unveils major innovation in polyester recycling with new textile preparation line, in the presence of Roland Lescure, French Minister of Industry.“ Pressemeldung, 2. Okt. 2023. https://www.carbios.com/newsroom/en/carbios-unveils-major-innovation-in-polyester-recycling-with-new-textile-preparation-line-in-the-presence-of-roland-lescure-french-minister-of-industry/#.
[91] Carbios. „CARBIOS and De Smet Engineers & Contractors enter engineering partnership for construction of world’s first PET biorecycling plant.“ Pressemeldung, 15. Feb. 2024. https://www.carbios.com/newsroom/en/carbios-and-de-smet-engineers--contractors-enter-engineering-partnership-for-construction-of-worlds-first-pet-biorecycling-plant/#.
[92] Orlando, M., Molla, G., Castellani, P. et al. Microbial Enzyme Biotechnology to Reach Plastic Waste Circularity: Current Status, Problems and Perspectives. Int J Mol Sci 24, 3877 (2023). https://doi.org/10.3390/ijms24043877.

Verbesserungen gegenüber existierenden Enzymen haben. Nur auf diese Kandidaten sollte sich die Forschungsarbeit im Hinblick auf Protein-Engineering und industrielle Skalierung konzentrieren.[93]

5.2. Bioremediation von Organohaliden

Halogen-organische Chemikalien wie polychlorierte Biphenyle (PCB), polybromierte Diphenylether (PBDE) sowie Per- und polyfluorierte Alkylverbindungen (PFAS) sind weit verbreitete Schadstoffe in der Umwelt. Xu et al. (2023) weisen darauf hin, dass bestimmte Mikroben diese Schadstoffe abschwächen können, so dass das Potenzial besteht, Techniken der Bioremediation auf Basis funktioneller Mikroorganismen zu entwickeln. Maschinelles Lernen, Quantenchemie und biomolekulare Modellierung werden als zukunftsträchtige Instrumente benannt, die zur Optimierung des Systems der Bioremediation, zur Vorhersage des Verbleibs von Chemikalien in der Umwelt und zur Erforschung funktioneller Enzyme dienen können. Potenzial wird auch dem Einsatz von AlphaFold2 zugesprochen, um die Wirkmechanismen von Enzymen sowie die Docking-Mechanismen von Organohaliden und den jeweiligen Enzymen zu untersuchen und so neue, optimierte Enzyme zu entdecken.[94]

5.3. Datenbank der Enzym-Mechanismen

Im Laufe der Jahre wurden Hunderte von Reaktionsmechanismen von Enzymen untersucht. Auf Basis dieser Literatur zur biologischen Katalyse stellen Ribeiro et al. (2023) EzMechanism als ein neues wissensbasiertes, frei zugängliches Werkzeug vor, das in der Lage ist, automatisch mechanistische Pfade für ein gegebenes dreidimensionales aktives Zentrum zusammen mit der Enzymreaktion abzuleiten. Dieses basiert auf einem Satz katalytischer Regeln, die aus der Datenbank „Mechanism and Catalytic Site Atlas“ zusammengestellt wurden. In einer Zukunftsvision der Autoren sind Computer in der Lage, die Funktionsweise von Enzymen und die von ihnen katalysierten Reaktionen *ab initio* allein auf der Grundlage ihrer Sequenzen vorherzusagen. Dies ist ein ehrgeiziges Ziel, das jedoch als immer dringlicher angesehen wird, je mehr Sequenzen und Strukturen verfügbar sind. Während sich andere Berechnungsmethoden in den Bereichen Bioinformatik und computergestützte Chemie beständig weiterentwickeln (AlphaFold, Simulation von Makromolekülen, Molekulardynamik und molekularem Docking), sehen die Autoren aktuelle und zukünftige Version von EzMechanism als ein entscheidendes Element dieser idealen Zukunftsvision.[95]

5.4. Proteindesign für Rekombinasen zur Genome-Editierung

Bestimmte Rekombinasen[96] vom Tyrosin-Typ sind wirksame Werkzeuge für das Genom-Engineering, von denen die ersten technischen Varianten therapeutisches Potenzial gezeigt haben. Bisher wurden Designer-Rekombinasen an neue DNA-Zielstellen hauptsächlich durch gerichtete molekulare Evolution angepasst. Dies ist zwar effektiv, aber mühsam und zeitaufwändig. Schmitt et al. (2022) stellen das Tool RecGen („Recombinase Generator“) vor, einen Algorithmus für die intelligente Erzeugung von Designer-Rekombinasen. Dieses Tool basiert auf Trainingsdaten von mehr als einer Millionen Rekombinase-Aminosäuresequenzen mit spezifischer Aktivität für 89 verschiedene Zielstellen. Die experimentelle Validierung des Tools zeigt, dass der Algorithmus Rekombinase-Sequenzen mit Aktivität an neuen Zielstellen vorhersagen kann, die nicht Teil der Trainingsdaten waren.[97]

Schmitt et al. betonen die bemerkenswerten Ergebnisse, die mit Transformatormodellen erzielt wurden, und schließen auf das Potenzial solcher Modelle auch für Rekombinase-Sequenzdaten, insbesondere um Modellvorhersagen auf funktionell nützliche Proteine zu filtern.[97]

[93] Herbert, J., Beckett, A.H. & Robson, S.C. A Review of Cross-Disciplinary Approaches for the Identification of Novel Industrially Relevant Plastic-Degrading Enzymes. Sustainability 14, 15898 (2022). https://doi.org/10.3390/su142315898.
[94] Xu, G., Zhao, S., Liu, J. et al. Bioremediation of organohalide pollutants: progress, microbial ecology, and emerging computational tools. Current Opinion in Environmental Science & Health 32, 100452 (2023). https://doi.org/10.1016/j.coesh.2023.100452, vgl. S. 7 linke Spalte.
[95] Ribeiro, A.J.M., Riziotis, I.G., Tyzack, J.D. et al. EzMechanism: an automated tool to propose catalytic mechanisms of enzyme reactions. Nat Methods (2023). https://doi.org/10.1038/s41592-023-02006-7.
[96] Rekombinasen sind Enzyme, die die genetische Rekombination katalysieren.
[97] Schmitt, L.T., Paszkowski-Rogacz, M., Jug, F. et al. Prediction of designer-recombinases for DNA editing with generative deep learning. Nat Commun 13, 7966 (2022). https://doi.org/10.1038/s41467-022-35614-6.

6. Bioproduktion

Auf dem Weg zu einer breiten Nutzung der industriellen Biotechnologie bis schließlich zur Bioökonomie sind auch Innovationen im Bereich der biotechnologischen Produktions- und Prozesstechnik nötig. In diesem Zusammenhang hat sich das Konzept von Biofoundries entwickelt. Hierunter versteht man Infrastrukturen, die die schnelle Entwicklung, Konstruktion und Erprobung genetisch veränderter Organismen für biotechnologische Anwendungen und die Forschung ermöglichen.[98] Auch spielen Fragen der Automatisierung biologischer Experimente[99] oder der beschleunigten Entwicklung von Enzymen, Zelllinien und Bioprozessen[100] eine Rolle. Es gibt inzwischen auch Biofoundries auf Unternehmensebene mit Möglichkeiten zur Kooperation mit Kunden und externen Forschenden.[101] In diesem Kapitel werden Kontexte aufgezeigt, in denen die Proteinstrukturvorhersage - möglicherweise im Rahmen von Biofoundries - zur Etablierung der Bioökonomie beitragen kann.

Pathwaydesign (oder auch „metabolic engineering")[102] ist in diesem Sinne eine vielversprechende Strategie für die Defossilisierung von Kunststoffen und Kraftstoffen, bei der die chemische Massenproduktion durch biosynthetische Verfahren ersetzt wird. Nach Carruthers und Lee (2022) hat die synthetische Biologie im letzten Jahrzehnt eine regelrechte Revolution erlebt. Diese zeichnete sich dadurch aus, dass Mikroben für die Bioproduktion auf der Basis von Design-Build-Test-Learn-Zyklen in hoher Konzentration relativ einfach modifiziert und optimiert werden konnten, wodurch in der Summe die Fähigkeit zur Entwicklung komplexer Biosynthesewege völlig verändert wurde.[103] Es wird hervorgehoben, dass ein Bündel wissenschaftlich-technischer Fortschritte hier zusammenspielt: Genom-Editierung [CRISPR; CRISPRi], adaptive Laborevolution, fortschrittliche Technologien zum Bibliotheks-Screening. Zu diesem Bündel gehört als wesentlicher Bestandteil auch die Proteinstrukturvorhersage. Anwendungs- und Umsetzungspotenziale bestehen in Bezug auf: Erzeugung von Nahrungsergänzungsmitteln, Biokraftstoffen für Treibstoffmischungen [Limonen, Isoprenol, Isobutanol], Biopolymere und chemische Grundstoffe [z. B. Adipinsäure, Milchsäure, L-Lysin-3-Hydroxypropionsäure].[103]

6.1. Biosensoren

Allosterische Transkriptionsfaktoren (aTF) sind Biosensoren zum Nachweis einer breiten Palette von Molekülen (wie Ionen, Zucker, Medikamente, Hormone), die für ökologische-, medizinische und industrielle Anwendungen eingesetzt werden können. Tellechea-Luzardo et al. (2023) argumentieren, dass aTF nahezu idealtypische Biosensoren sein könnten, wenn es nur bessere rechnergestützte Werkzeuge für ihr Design gäbe. Sie vertreten die Ansicht, dass es mit den neuen Methoden zur Proteinstrukturvorhersage möglich sein könnte, das computergestützte aTF-Design vom Niveau einer bloßen Anleitung für sehr aufwändige Mutagenese-Experimente auf bestenfalls halbquantitativer Basis zu einer Vorhersage hoher Zuverlässigkeit zu machen.[104]

6.2. Beschleunigung von Forschung und Entwicklung in der Bioindustrie

Smanski et al. (2022) verweisen darauf, dass die Bioökonomie ein rasantes Wachstum erlebt. Die Verbesserungen und Kostensenkungen bei der DNA-Synthese und DNA-Sequenzierung ermöglichen demnach neue Ansätze in den frühen Phasen von Forschung und Entwicklung (FuE). Sie erlauben ein massiv paralleles Design einschließlich Entwicklung und Evaluierung von Hunderten bis Tausenden von mutmaßlichen Produktionsstämmen. In dem

[98] Hillson, N., Caddick, M., Cai, Y. et al. Building a global alliance of biofoundries. Nat Commun 10, 2040 (2019). https://doi.org/10.1038/s41467-019-10079-2.
[99] MaxGENESYS Biofoundry; https://www.mpi-marburg.mpg.de/927693/maxgenesys.
[100] Jülich Biofoundry; https://www.fz-juelich.de/en/ibg/ibg-1/juelich-biofoundry-en.
[101] Campus des französischen Unternehmens Lesaffre in Lille, der nach eigenen Angaben die größte Biofoudry Europas beheimatet, Pressemeldung „Lesaffre inaugurates its Campus and announces its R&D ambitions to accelerate innovation". 13 Okt. 2022. https://www.lesaffre.com/press-room/lesaffre-inaugurates-its-campus-and-announces-its-rd-ambitions-to-accelerate-innovation/ sowie https://www.lesaffre.com/trends-mag/biofoundries-speed-up-the-discovery-of-novel-yeasts-and-bacteria-of-interest/ und https://www.lesaffre.com/press-room/lesaffre-and-laboratory-of-biology-and-modelling-of-the-cell-lbmc-of-the-cnrs1-and-the-ens2-of-lyon-a-new-partnership-to-improve-innovation-in-the-fermentation-field/.
[102] Pathwaydesign (auch metabolic engineering) ist eine Methode der Biochemie, bei der die Stoffwechselwege durch Proteindesign geändert werden. Das Pathwaydesign ist eine Form des rationalen Designs. Vgl. Wikipedia-Eintrag „Phatwaydesign"; https://de.wikipedia.org/wiki/Pathwaydesign.
[103] Carruthers, D.N. & Lee, T.S. Translating advances in microbial bioproduction to sustainable biotechnology. Front Bioeng Biotechnol 10, 968437 (2022). https://doi.org/10.3389/fbioe.2022.968437.
[104] Tellechea-Luzardo, J., Stiebritz, M.T. & Carbonell, P. Transcription factor based biosensors for screening and dynamic regulation. Front Bioeng Biotechnol 11, 1118702 (2023). https://doi.org/10.3389/fbioe.2023.1118702.

Maße, in dem sich die Möglichkeiten für ein rationales genetisches Design verbessern, können sich diese Bibliotheken auch mithilfe von ML zunehmend auf Varianten hoher Leistungsfähigkeit konzentrieren. Auch in diesem Kontext der bioindustriellen Fertigung wird ein Zusammenspiel von Fortschritten und Werkzeugen konstatiert: Dazu zählen CRISPR-Cas9, Sequenzierungsprojekte zur Entdeckung von Enzymen sowie die neuen Tools zur Proteinstrukturvorhersage usw. Diese rasanten Fortschritte bei den Werkzeugen ermöglichen es nach Einschätzung der Autoren sowohl kleinen als auch großen Unternehmen, Organismen für die bioindustrielle Fertigung mit höherer Präzision und Geschwindigkeit zu entwickeln. Während diese Innovationen die Entwicklung von Ideen bis zu einem Reifegrad des Konzeptnachweises beschleunigen können, halten sie weitere Innovationen im Pilot-, Hochlauf- und Industrie-Maßstab für erforderlich.[105]

6.3. Automatisierte Entwicklung von Bioprozessen

Duong-Trung et al. (2023) listen die Erfolge des maschinellen Lernens in unterschiedlichen Bereichen der Biologie auf: Up-Downstream-Prozesse, Bioprozesse für die Herstellung chemischer und biologischer Produkte, Enzyme und Zellwachstum, Expressionssysteme für Zellkulturen und eben die Proteinstrukturvorhersage. Sie weisen aber darauf hin, dass maschinelles Lernen bislang noch nicht umfassend für die Bioprozessentwicklung eingesetzt wird. Vor diesem Hintergrund entwickeln sie die Zukunftsvision einer Automatisierung in der experimentellen Biotechnologie, sowohl in einzelnen Labors als auch in laborübergreifend, in Verbindung mit einer robusten Entscheidungsfindung durch verbesserte Modelle des maschinellen Lernens. Dies könnte ihrer Ansicht zufolge in die Entwicklung vollautomatischer, kontinuierlicher Bioprozesstechniken münden, die nicht von menschlichen Eingriffen abhängen oder diese einschränken.[106]

[105] Smanski, M.J., Aristidou, A., Carruth, R. et al. Bioindustrial manufacturing readiness levels (BioMRLs) as a shared framework for measuring and communicating the maturity of bioproduct manufacturing processes. Journal of Industrial Microbiology and Biotechnology 49, 5 (2022). https://doi.org/10.1093/jimb/kuac022.

[106] Duong-Trung, N., Born, S., Kim, J.W. et al. When bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. Biochemical Engineering Journal 190, 108764 (2023). https://doi.org/10.1016/j.bej.2022.108764.

7. Gesundheit - Medikamentenentwicklung

So wie Proteine vielfältige Funktionen im menschlichen Körper haben, nehmen sie auch bei vielen Krankheiten eine Schlüsselrolle ein. Aus diesem Grunde sind Proteine häufig sog. Zielmoleküle (engl. „targets"), auf die Medikamente einwirken sollen, um ihre therapeutische Wirkung zu entfalten.[107] Mögliche Proteine als Zielmoleküle sind dabei bspw. körpereigene Enzyme, Rezeptoren[34] oder Ionenkanäle, die die Durchlässigkeit von Biomembranen (wie z.B. der Zellhülle) für bestimmte Stoffe steuern. In diesem Fall ist das genaue Verständnis dieser Zielmoleküle auch für die Entwicklung wirksamer Medikamente wesentlich. Umgekehrt werden Proteine (bzw. die kleineren Peptide) als Medikamente immer wichtiger. Ebrahimi und Samanta (2023) weisen darauf hin, dass nachdem erstmals 1982 ein Protein-basiertes Therapeutikum als Medikament zugelassen wurde, der Weltmarkt für Protein-basierte Medikamente heute auf die 400-Milliarden-USD-Marke zusteuert und etwa die Hälfte der zehn meistverkauften Medikamente inzwischen Protein-basiert sind.[108] Bei diesen Medikamenten gibt es die folgenden wichtigsten Klassen: Antikörper, Enzyme, Blutgerinnungsfaktoren, Hormone und Zytokine (regulieren das Zellwachstum).[109]

Angesichts dieses immensen wirtschaftlichen Potenzials und der Tatsache, dass die biologische und damit auch medizinische Funktion von Proteinen von ihrer Form, d.h. ihrer dreidimensionalen Struktur abhängt, liegt es nahe, dass die neuen Werkzeuge zur Proteinstrukturvorhersage zukünftig auch im Feld der Medikamentenentwicklung Anwendung finden werden. Wie groß dieser Einfluss heute schon ist bzw. in Zukunft sein kann, ist in der Wissenschaft durchaus umstritten.[110,111,112] Es ist durchaus denkbar, dass zur Entfaltung des vollen Potenzials zunächst noch weitere Fortschritte - bspw. hinsichtlich der räumlichen Genauigkeit der Strukturvorhersagen oder der Anwendbarkeit auf Protein-Komplexe bzw. Protein-Ligand-Komplexe[113] (vgl. Abschnitt 2.2) – notwendig sind, von denen allerdings unklar ist, ob, inwieweit und wann diese öffentlich gemacht werden.[114]

Die folgenden Abschnitte enthalten aktuelle Zukunftseinschätzungen zu konkreten Anwendungsfeldern der Proteinstrukturvorhersage im Feld der Medikamentenentwicklung.

7.1. Zukünftige Pandemien

Pavan und Moro (2023) verweisen auf eine aktuelle Studie, der zufolge sich die Wahrscheinlichkeit für eine Epidemie, ausgelöst durch eine hochansteckende Krankheit, in den nächsten Jahrzehnten vordoppeln könnte, und deuten dies als Beleg dafür, dass die erfolgreichen rechnergestützten Strategien aus der Biologie, die schon gegen COVID-19 eingesetzt wurden, dann wieder nützlich sein werden.[115]

Bei der COVID-19-Pandemie und der Entwicklung von SARS-CoV-2-Impfstoffen hat die Kenntnis der Spike-Proteinstruktur zum Verständnis der Oberflächentopologie des Virus und seiner Antigenität beigetragen. In diesem Zusammenhang weisen Thornton et al. (2021) darauf hin, dass am 3. September 2021 bereits 1.491 Strukturen von SARS-CoV-2-Proteinen in der wwPDB vorhanden waren, die von Labors aus der ganzen Welt zusammengetragen wurden. Sie erwarten, dass die Möglichkeit, virale Spike-Proteine genau vorherzusagen, im

[107] Vgl. Wikipedia-Eintrag „Target (Biologie)"; https://de.wikipedia.org/wiki/Target_(Biologie).
[108] Ebrahimi, S.B. & Samanta, D. Engineering protein-based therapeutics through structural and chemical design. Nat Commun 14, 1, 2411 (2023). https://doi.org/10.1038/s41467-023-38039-x.
[109] Ibid., vgl. Box 1.
[110] Callaway, E. AlphaFold found thousands of possible psychedelics. Will its predictions help drug discovery? Nature 626, 14–15 (2024). https://doi.org/10.1038/d41586-024-00130-8.
[111] Baker, C.M. & Atzori, A. AlphaFold: Deep Learning, Drug Discovery and the Protein Structure Revolution. Chimia 76, 364 (2022). https://doi.org/10.2533/chimia.2022.364.
[112] OECD. Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. Paris: OECD Publishing (2023). https://doi.org/10.1787/a8d820bd-en, zur Rolle von KI in der Wirkstoffforschung siehe S. 160ff.
[113] Krishna, R., Wang, J., Ahern, W. et al. Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom. bioRxiv 2023.10.09.561603 (2023). https://doi.org/10.1101/2023.10.09.561603.
[114] Isomorphic Labs, Pressemeldung „Isomorphic Labs kicks off 2024 with two pharmaceutical collaborations", 7. Jan. 2024: „Isomorphic Labs announces strategic collaborations with two of the world's leading pharmaceutical companies, Eli Lilly and Company and Novartis. These partnerships have the potential to be worth nearly $3 billion to Isomorphic Labs, excluding any royalties that may result from future drug sales, and are testament to Isomorphic's approach to drug design and the progress we are making. Both collaborations are multi-target, focused on small molecules, and include a combination of upfront and milestone payments." Vorauszahlungen in Höhe von insgesamt knapp 80 Mio. $ wurden vereinbart, weitere Zahlungen sind erfolgsabhängig. https://www.isomorphiclabs.com/articles/isomorphic-labs-kicks-off-2024-with-two-pharmaceutical-collaborations.
[115] Pavan, M. & Moro, S. Lessons Learnt from COVID-19: Computational Strategies for Facing Present and Future Pandemics. Int J Mol Sci 24, 4401 (2023). https://doi.org/ 10.3390/ijms24054401.

Vergleich zur experimentellen Strukturbestimmung neu auftretender Viren bei zukünftigen Pandemien eine sehr schnelle Analyse ermöglichen wird.[116]

7.2. Vernachlässigte Krankheiten

Baker und Atzori (2022) nennen einen Bereich der Arzneimittelentdeckung, in dem AlphaFold wahrscheinlich die Methode der Wahl zur Gewinnung von Strukturinformationen werden wird, und das sind Medikamente für vernachlässigte Krankheiten. Krankheiten, von denen überwiegend arme Bevölkerungsgruppen in Entwicklungsländern betroffen sind. Diese Länder haben oft keinen Zugang zu teuren Versuchsanlagen und es handelt sich oft um Krankheiten, bei denen die relevanten Proteinstrukturen durch die PDB-Strukturen nicht gut abgedeckt werden. Vor diesem Hintergrund hat DeepMind bereits eine Partnerschaft mit der „Drugs for Neglected Diseases Initiative“ angekündigt.[117]

Als ein Beispiel sei auf folgende Krankheit hingewiesen: Mehr als 100 Millionen Menschen sind mit dem karzinogenen Blutegel Schistosoma haematobium infiziert, dem Erreger der urogenitalen Schistosomiasis (auch Bilharziose). Trotz dieser großen Bedeutung ist nur wenig über diesen Blutegel und seine Interaktionen mit den Zwischenwirten Mensch und Schnecke sowie der Pathogenese auf molekularer und biochemischer Ebene bekannt. Stroehlein et al. (2022) berichten über das erste Genom auf Chromosomenebene und deutlich verbesserte Genmodelle für S. haematobium sowie über Vorhersagen der Strukturen aller sieben Omega-1-Homologe[118] von S. haematobium mithilfe von AlphaFold.[119] Omega-1 ist ein Protein, das im Krankheitsverlauf Leberschädigungen auslöst.[120]

7.3. Malaria

Stofberg et al. (2022) berichten, dass Malaria nach wie vor zu den tödlichsten parasitären Krankheiten der Tropen zählt und eine Bedrohung für die öffentliche Gesundheit darstellt. Zudem führt die Entwicklung von Resistenzen gegen Malariamedikamente dazu, dass die erzielten Erfolge bei der Bekämpfung der Krankheit zunichtegemacht werden. Der Parasit verfügt über einen komplexen Lebenszyklus, wodurch er fast alle bisher bekannten Malariamedikamente überlisten kann. Stofberg et al. stellen fest, dass die Verfügbarkeit von qualitativ hochwertigen dreidimensionalen Strukturmodellen bestimmter Proteine, die von AlphaFold gelöst wurden, in diesem Zusammenhang die Entwicklung von selektiven Inhibitoren (Hemmstoffen) beschleunigen könnte. Dies verleiht den dreidimensionalen Strukturen, die für das In-silico-Design von Inhibitoren dringend benötigt werden, eine neue Tragweite, die möglicherweise ein exponentielles Wachstum bei der Entwicklung von Arzneimitteln gegen Plasmodien auslösen werden.[121]

7.4. Antibiotika-Resistenz

Angesichts der raschen Entwicklung der Multi-omic-Methoden und entsprechender Datenbanken ist der Beitrag von Studien der Bioinformatik bei den Forschungsbemühungen zur Bekämpfung multiresistenter Keime allgegenwärtig. Durch Analyse des bakteriellen Genoms oder Proteoms, bspw. in Form von In-silico-Ansätzen wie der Protein-Interaktom-Analyse, können neue wichtige Proteine in den gefährlichen Bakterien entdeckt werden, die potenziell als Impfstoffkandidaten in Frage kommen. Mit der zunehmenden Anzahl von Einträgen in experimentell validierten 3D-Strukturdatenbanken, wie der Protein Data Bank (PDB), kann die Stärke der Wechselwirkungen zwischen kleinen Molekülen und Proteinen sowie der Protein-Protein-Wechselwirkungen durch In-silico-Ansätze vorhergesagt werden, wozu auch Ergebnisse von AlphaFold beitragen. Eine solche Technik wird auch in der

[116] Thornton, J.M., Laskowski, R.A. & Borkakoti, N. AlphaFold heralds a data-driven revolution in biology and medicine. Nat Med 27, 1666–1669 (2021). https://doi.org/10.1038/s41591-021-01533-0.
[117] Baker, C.M. & Atzori, A. AlphaFold: Deep Learning, Drug Discovery and the Protein Structure Revolution. Chimia 76, 364 (2022). https://doi.org/10.2533/chimia.2022.364.
[118] Proteine werden als „homolog“ bezeichnet, wenn sie eine große Ähnlichkeit der Aminosäuresequenzen aufweisen, was als Evidenz für eine gemeinsame Abstammung gewertet wird; vgl. Wikipedia-Eintrag „Sequence homology“; https://en.wikipedia.org/wiki/Sequence_homology.
[119] Stroehlein, A.J., Korhonen, P.K., Lee, V.V. et al. Chromosome-level genome of Schistosoma haematobium underpins genome-wide explorations of molecular variation. PLoS Pathog 18, 2, e1010288 (2022). https://doi.org/10.1371/journal.ppat.1010288.
[120] Fitzsimmons, C.M., Schramm, G., Jones, F.M. et al. Molecular characterization of omega-1: A hepatotoxic ribonuclease from Schistosoma mansoni eggs. Molecular and Biochemical Parasitology 144, 1, 123–127 (2005). https://doi.org/10.1016/j.molbiopara.2005.08.003.
[121] Stofberg, M.L., Caillet, C., de Villiers, M. et al. Inhibitors of the Plasmodium falciparum Hsp90 towards Selective Antimalarial Drug Design: The Past, Present and Future. Cells 10, 2849 (2021). https://doi.org/ 10.3390/cells10112849.

Impfstoffforschung eingesetzt, indem die Peptidbindung an Immunzellrezeptoren rechnerisch abgeleitet wird. Ähnliche Methoden werden in der In-silico-Wirkstoffforschung angewandt. So kann beispielsweise ein virtuelles Screening großer chemischer Datenbanken für ein bestimmtes Wirkstoff-Target neue antibakterielle Wirkstoffe hervorbringen, die neu eingesetzt oder weiterentwickelt werden können.[122]

7.5. Tuberkulose

Die derzeitigen Anti-Tuberkulose-Therapien enthalten Medikamente, die die replizierende Form der Bazillen abtöten, aber nicht gegen ruhende Populationen von Bazillen wirken. Die Identifizierung von Arzneimitteln, die arzneimitteltolerante ruhende Populationen bekämpfen können, ist insofern wichtig für die Entwicklung wirksamer Vorgehensweisen zur Verkürzung der Behandlungsdauer von TB. Mit diesem Ziel haben Sharma et al. (2023) mit Hilfe von In-silico-Ansätzen Untersuchungen unter Verwendung von AlphaFold durchgeführt, um die arzneimittelwirksamen Ziele bestimmter Phytomoleküle in ruhenden Bazillen zu verstehen und so die Arzneientwicklung zu unterstützen.[123]

7.6. Krebs aus Sicht der Strukturbiologie

In einer Übersichtsarbeit beschreiben Ozdemir und Nussinov (2023) den Stand der Technik bei Protein-Protein-Interaktionen (PPIs) zwischen Wirt und Pathogenen aus strukturbiologischer Sicht. Es werden rechnergestützte Aspekte der Vorhersage dieser PPIs erörtert, einschließlich des maschinellen Lernens und der künstlichen Intelligenz. Anhand von Beispielen wird gezeigt, wie theoretische Berechnungsansätze zu einem therapeutischen Wirkstoff führen können. Die Autoren erwarten, dass KI-gesteuerte Ansätze in Zukunft die Strukturmodellierung im Bereich der Wirt-Pathogen-PPIs dominieren werden, trotz der bekannten Schwächen von Modellen wie AlphaFold hinsichtlich der Proteindynamik.[124]

[122] Tan, Y.C. & Lahiri, C. Promising Acinetobacter baumannii Vaccine Candidates and Drug Targets in Recent Years. Front Immunol 13, 900509 (2022). https://doi.org/10.3389/fimmu.2022.900509.
[123] Sharma, S., Chikhale, R., Shinde, N. et al. Targeting dormant phenotype acquired mycobacteria using natural products by exploring its important targets: In vitro and in silico studies. Front Cell Infect Microbiol 13, 1111997 (2023). https://doi.org/10.3389/fcimb.2023.1111997.
[124] Ozdemir, E.S. & Nussinov, R. Pathogen-driven cancers from a structural perspective: Targeting host-pathogen protein-protein interactions. Front Oncol 13, 1061595 (2023). https://doi.org/10.3389/fonc.2023.1061595.

8. Tiergesundheit

8.1. Vetinformatik: ein Forschungskonzept für die Veterinärmedizin

Weil sie eine wichtige Rolle in der Geschichte der Menschheit gespielt haben und weiterhin spielen, werden Nutztiere als integraler Bestandteil der Gesellschaft angesehen. Die Industrialisierung - auch der Tierhaltung - hat jedoch natürliche Ressourcen erschöpft und den Klimawandel weltweit verschärft. Dadurch wurde das Auftreten bestimmter Krankheiten begünstigt und die Produktivität von Viehbeständen verringert. Gleichzeitig benötigt die wachsende Weltbevölkerung ausreichend Nahrungsmittel, was auch zu einer steigenden Nachfrage nach tierischen Erzeugnissen führen würde. Nach Darstellung dieser Grundüberlegungen kommen Pathak und Kim (2022) in ihrer Arbeit zu der Forderung nach Innovationen in der Veterinärwissenschaft, um deren Produktivität sowohl in quantitativer als auch qualitativer Hinsicht zu steigern. Sie verweisen auf Durchbrüche bei Multi-omics-Plattformen, die eine Fülle von genetischen und genomischen Daten für Nutztiere hervorgebracht haben und die nun in nutzbares Wissen für Zucht, Krankheitsvorbeugung und -management, Produktivität und Nachhaltigkeit transferiert werden müssen. Die Vetinformatik wird als neues bioinformatisches Forschungskonzept bzw. -ansatz vorgestellt, die die Veterinärwissenschaft revolutioniert, indem sie einen interdisziplinären Ansatz verwendet, um die komplexen molekularen Mechanismen von Tiersystemen zu verstehen. Sie verfolgt das Ziel, die veterinärmedizinische Forschung zu beschleunigen und die Lebensmittel- und Ernährungssicherheit zu gewährleisten.[125] Die Modellierung, Visualisierung und Validierung von Proteinstrukturen soll dabei eine zentrale Rolle spielen. So weisen Pathak und Kim darauf hin, dass die Entwicklung von AlphaFold die Modellierung von Proteinstrukturen wesentlich erleichtert hat. Sie argumentieren, dass, wenn die Struktur des Zielproteins nicht durch experimentelle Techniken aufgeklärt werden kann, rechnerische Ansätze von immenser Bedeutung sind. Bei fehlenden experimentellen Daten kann die Proteinstruktur für die Verwendung beim Wirkstoffscreening rechnerisch vorhersagt werden. Auch könnten rechnerische Ansätze zur Vorhersage von Protein-Protein-Wechselwirkungen, zum Strukturvergleich und zum Alignment verwendet werden. Es wird auch auf die Entwicklung mehrerer Tools für die Visualisierung, Verfeinerung und Validierung der vorhergesagten 3D-Proteinmodelle hingewiesen (PyMOL, Swiss PDB Viewer, Rampage, PROCHECK usw.). Mit dieser Argumentation kommen sie zu dem Schluss, dass Akteure der Veterinärwissenschaften besser darüber informiert werden müssen, welche Vorteile die Anwendung der Vetinformatik zum Verständnis von Tierhaltungssystemen einschließlich der Verbesserung des Tierwohls und die Entdeckung von Medikamenten bieten kann.[125]

[125] Pathak, R.K. & Kim, J.-M. Vetinformatics from functional genomics to drug discovery: Insights into decoding complex molecular mechanisms of livestock systems in veterinary science. Front Vet Sci 9, 1008728 (2022). https://doi.org/10.3389/fvets.2022.1008728.

9. Pflanzengesundheit

9.1. Schutz vor Pilzerkrankungen

Roth et al. (2023) verweisen darauf, dass Pilzproteine nur einen winzigen Bruchteil im bekannten und erforschten Teil des Protein-Universums darstellen, was ihrer Einschätzung nach dazu führt, dass die Werkzeuge zur Proteinstrukturvorhersage in der Pilzforschungsgemeinschaft bislang nur begrenzt eingesetzt werden.[126] Sie beobachten jedoch, dass seit der Veröffentlichung von AlphaFold, dessen Potenzial für eine Revolutionierung der pharmazeutischen Behandlung von menschlichen Pilzerregern diskutiert wird, und dass darüber hinaus AlphaFold auch bei der Analyse eines breiten Spektrums pilzlicher Pflanzenpathogene eingesetzt wird. Insgesamt kommen Roth et al. zu der Einschätzung, dass AlphaFold ein vielversprechendes Potenzial besitzt, um durch rationales Proteindesign wertvolle und/oder neuartige enzymatische Aktivitäten in Pilzen zu erzielen.

9.2. Schutz vor Viruserkrankungen

Pflanzeninfizierende Viren beeinträchtigen nicht nur den Ertrag der infizierten Pflanzen, sondern auch deren Nährstoffgehalt erheblich. Hohe Ernteverluste aufgrund von Ausbrüchen von Pflanzenviren sind für die weltweite Ernährungssicherheit von entscheidender Bedeutung und erfordern daher dringend die Umsetzung intelligenter Managementmaßnahmen, so führen Ghosh et al. in ihrer Arbeit aus.[127] Sie betonen, dass die Funktion eines Proteins am besten durch die Visualisierung seiner Struktur verständlich wird, und weisen auf das Hemmnis hin, dass bisher nur ein sehr kleiner Teil der pflanzlichen Proteine, die am Signalweg der pflanzlichen Immunantwort beteiligt sind, strukturell charakterisiert wurde. Zudem seien die Strukturen der pflanzlichen Virusproteine ebenfalls weitgehend ungeklärt, sodass arbeitsintensive Methoden der Proteinkristallisation den größten Engpass darstellen. Sie sehen in der Entwicklung von AlphaFold2 eine Revolution der Vorhersage von Proteinstrukturen. Schließlich kommen Ghosh et al. zu der Zukunftserwartung, dass die Strukturvorhersage von Pflanzenvirusproteinen und die Vorhersage ihrer chemischen Inhibitoren sowie deren anschließende erfolgreiche Verabreichung eine neuartige und wirksame Strategie zur Bekämpfung von Viruserkrankungen von Pflanzen sein wird.

[126] Roth, M.G., Westrick, N.M. & Baldwin, T.T. Fungal biotechnology: From yesterday to tomorrow. Front Fungal Biol 4, 1135263 (2023). https://doi.org/10.3389/ffunb.2023.1135263.
[127] Ghosh, D., Chakraborty, S., Kodamana, H. et al. Application of machine learning in understanding plant virus pathogenesis: trends and perspectives on emergence, diagnosis, host-virus interplay and management. Virol J 19, 42 (2022). https://doi.org/10.1186/s12985-022-01767-5.

10. Weitere Anwendungen

10.1. Proteine bei der Eisbildung in Wolken

Mikrobiell produzierte Eiskeimbildungsproteine (INpro) sind einzigartige molekulare Strukturen mit der höchsten bekannten katalytischen Effizienz für die Eisbildung, wie Hartmann et al. (2022) ausführen. Sie spielen eine Rolle bei der Regenbildung und bei Frostschäden an Nutzpflanzen und finden vielfältige kommerzielle Anwendung, so dass ein gründliches Verständnis der inhärenten Mechanismen zur Eiskeimbildung von Nutzen ist. Hartmann et al. haben AlphaFold verwendet, um das erste Ab-initio-Strukturmodell eines bakteriellen INpro zu entwickeln. Mithilfe einer Messung mit Synchrotronstrahlung wurde ein Teil des Modells experimentell validiert. Durch die Kombination von Funktionsstudien von gereinigtem rekombinantem INpro, Elektronenmikroskopie und Modellierung zeigen Hartmann et al. weiter, dass die Bildung von Dimeren und Oligomeren höherer Ordnung der Schlüssel zur INpro-Aktivität ist. Sie sehen in ihrer Arbeit einen großen Fortschritt im Verständnis der molekularen Grundlagen der bakteriellen Eiskeimbildung und die Grundlage für die Untersuchung der mechanistischen Rolle der INpro-induzierten Eisbildung in der Atmosphäre. Schließlich halten sie das kommerzielle Design und die Produktion von Eiskeim-bildenden Partikeln für industrielle Anwendungen in der Zukunft für möglich.[128]

10.2. Verbesserte Nutzung von Keratin aus Geflügelfedern

Die Entsorgung von Keratin-Abfällen, die in verschiedenen Branchen anfallen, ist weltweit zu einer großen Herausforderung geworden. Jährlich gelangen fast eine Milliarde Tonnen Keratin-Abfälle in die Umwelt, wie von Tuly et al. (2023) ausgeführt wird.[129] Diese Quelle legt nahe, dass das Verständnis der strukturellen Eigenschaften von Keratin[130] von großer Bedeutung für die Bewältigung von Keratin-Abfällen und ihre potenzielle Anwendung in Keratin-inspirierten Biomaterialien ist. In der Arbeit wurde die molekulare Struktur von Hühnerfederkeratin durch AlphaFold2 und quantenchemische Berechnungen charakterisiert.

[128] Hartmann, S., Meilee, L., Lasse, S.A. et al. Structure and Protein-Protein Interactions of Ice Nucleation Proteins Drive Their Activity. Frontiers in Microbiology 13 (2022). https://doi.org/10.3389/fmicb.2022.872306.
[129] Tuly, J.A., Ma, H., Lee, H.-J. et al. Insights of keratin geometry from agro-industrial wastes: A comparative computational and experimental assessment. Food Chemistry 418, 135854 (2023). https://doi.org/10.1016/j.foodchem.2023.135854.
[130] Keratine sind der Hauptbestandteil von Haaren, Nägeln, Krallen, Klauen, Hufen, Hörnern, Schnäbeln und Federn, vgl. Wikipedia-Eintrag “Keratine”; https://de.wikipedia.org/wiki/Keratine.

11. NRW-Perspektive

In diesem Kapitel werden die Anwendungspotenziale der Proteinstrukturvorhersage aus der NRW-Perspektive betrachtet. Zunächst stellt die folgende Tabelle die Felder der Innovationsstrategie des Landes NRW und einige der oben genannten Anwendungspotenziale gegenüber.

Felder der NRW-Innovationsstrategie	Anwendungspotenziale Proteinstrukturvorhersage
Werkstoffe	Biomaterialien (Abschnitte 4.1, 4.2.1, 10.2)
Produktion	Biokatalyse (Kap. 5); Bioproduktion (Kap. 6)
Vernetzte Mobilität und Logistik	Biotreibstoffe (Kap. 5)
Umweltwirtschaft und Circular Economy	Bioremediation (Abschnitt 3.2 und 5.2); Biokatalyse (Kap. 5); Bioproduktion (Kap. 6)
Energie	Biotreibstoffe (auch Kap. 5)
Bauen	Nicht direkt belegt - Biobasierte Baumaterialien
Medizin, Gesundheit und Life Science	Arzneimittelentwicklung (Kap. 7), Tier- und Pflanzengesundheit (Kap. 8 und 9)
Kultur, Medien- und Kreativwirtschaft und innovative Dienstleistungen	Nicht direkt belegt. Im Pharmasektor innovative Bioinformatik-Dienstleistungen möglich.[114]
Schlüsseltechnologien der Zukunft, IKT	Proteinstrukturvorhersage als neues Thema aus dem Bereich der Life Science (gesamtes Thema); Robotik-Labore (Abschnitt 2.4 als Zusammenspiel von Innovationen aus Robotik, Analytik, KI und Software); Längerfristig: Molekulare Kommunikation mit Proteinen;[131,132] Proteine in der Datenverarbeitung[133]

Nachfolgend werden einzelne Anwendungspotenziale und Beispiele einschlägiger Akteure (mit Schwerpunkt Wissenschaft und soweit möglich aus der Wirtschaft) aus NRW aufgeführt. Die NRW-Akteure werden in Kurzprofilen charakterisiert. Diese enthalten eine Einordnung, bei der Inhalte aus den Selbstbeschreibungen der Akteure stichwortartig mit den Inhalten aus den Anwendungsbeispielen in den obigen Kapiteln 2 bis 10 in Bezug gesetzt werden. Die Reihenfolge der Abschnitte 11.1 bis 11.6 entspricht der Abfolge der Kapitel 2 bis 10.

11.1. Proteinstrukturvorhersage in den Lebenswissenschaften

11.1.1. OpenFold

Wie in Kapitel 1 ausgeführt, zielt OpenFold auf eine Implementierung von AlphaFold2 mit vergleichbarer Genauigkeit, die es darüber hinaus erlaubt, neue Fragestellungen zu untersuchen sowie Modelle auf selbstgewählten (und auch proprietären) Datensätzen zu trainieren, um besser zu verstehen, wie der Lernfortschritt genau zustande kommt.[134]

[131] Cevallos, Y., Nakano, T., Tello-Oquendo, L. et al. Modeling Gene Expression and Protein Delivery as an End-to-End Digital Communication System. Open Bioinformatics Journal 14, 21–35 (2021). https://doi.org/10.2174/1875036202114010021.
[132] National Academies of Sciences, Engineering, and Medicine. Communication and Information Transmission Using Biotechnology: Abridged Version. Washington: The National Academies Press (2022). https://doi.org/10.17226/26579.
[133] Wang, J., Qian, F., Huang, S. et al. Recent Progress of Protein-Based Data Storage and Neuromorphic Devices. Adv Intell Syst 3, 2000180 (2021). https://doi.org/10.1002/aisy.202000180.
[134] Ahdritz, G., Bouatta, N., Kadyan, S. et al. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. bioRxiv 2022.11.20.517210 (2022). https://doi.org/10.1101/2022.11.20.517210.

In einer Pressemeldung wird 2024 darauf hingewiesen, dass OpenFold zwei neue Tools bereitstellt: SoloSeq als Large Language Model (LLM) für Proteine, das direkt in OpenFold integriert ist und OpenFold-Multimer speziell für Protein-Protein-Komplexe.[135]

Ein direkter NRW-Bezug besteht darin, dass die Bayer AG mit Sitz in NRW sowie UCB mit einer FuE-Niederlassung in NRW Kooperationspartner von OpenFold sind.

11.1.2. NRW-Akteure

Beispiele für NRW-Akteure - Lebenswissenschaften
Unternehmen
Bayer AG siehe Abschnitte 11.5.3 sowie 11.6.; hier als Kooperationspartner im OpenFold-Konsortium, weil OpenFold als offene Software für die Proteinstrukturvorhersage auch im Forschungskontext relevant ist.

Cube Biotech GmbH, Monheim „Unleashing the power of proteins for research and progress, Cube Biotech supports the pharmaceutical and biotechnical community with all its scientific expertise. By leveraging the latest technologies and techniques, we are able to provide a comprehensive range of products and services that address the full spectrum of protein research, from expression and purification to characterization and stabilization.“ https://cube-biotech.com/about-us/about-cube/ „little is still known about many membrane protein structures and their underlying working principles, due to their challenging solubilization criteria. Modern medicine would greatly benefit from a deeper understanding of this protein type. Recently though, this old problem is picking up speed again, as various AI systems like DeepMind's AlphaFold2 join the fight. They are capable of predicting a 3D protein structure directly from its amino acid sequence with ever-increasing accuracies. […] In the future, the intertwinement of AI and experimental methods like these will introduce a new era of biological and medical research, shaping fields like digital biology or biomedical computing. Novel approaches will renew our drug discovery processes, accelerating them immensely. Whatever path we pursue, the membrane proteome will play a vital part in it.“ https://cube-biotech.com/knowledge/membrane-protein-stabilization/membrane-proteins/ **Einordnung:** Unternehmen, das Produkte und Dienstleistung zur Unterstützung der Proteinforschung anbietet. Verweist explizit auf Fortschritte in der Proteinstrukturvorhersage. Bezug zu Kapitel 7 Gesundheit - Medikamentenentwicklung, s. o.: „Modern medicine would greatly benefit […]“
EducatedGuess.ai GbR, Neunkirchen im Siegerland „Gemeinsam Künstliche Intelligenz für die Lebenswissenschaften nutzbar machen“ „[…] wissenschaftliche […] Anwendung von Ansätzen der künstlichen Intelligenz, insbesondere im Bereich der Life Sciences. Hierbei arbeiten wir an verschiedenen Projekten in Zusammenarbeit, u. a. mit Forschern des Max Planck Instituts für Pflanzenzuchtforschung, sowie der Carnegie Institution for Science.“ **Einordnung:** Beispiel zu Abschnitt 2.4 Neue KI-Anwendungen in den Lebenswissenschaften

[135] OpenFold verlinkt unter https://openfold.io/ auf Business Wire, OpenFold Biotech AI Research Consortium releases SoloSeq and Multimer. An integrated protein Large Language Model with 3D structure generation. 19. Feb. 2024. https://www.businesswire.com/news/home/20221122005163/en/%C2%A0OpenFold-AI-Research-Consortium-Welcomes-4-New-Members-Bayer-Dassault-CHARM-Therapeutics-and-BaseCamp-Research.

Forschungseinrichtungen
RWTH Aachen Physikalisches Institut, AG Biophysik „Die Forschungsaktivitäten […] der AG Biophysik haben zum Ziel ein fundamentales physikbasiertes Verständnis zur Funktionsweise biologischer makromolekularer Systeme zu erlangen. In diesem Zusammenhang sind für uns Proteine (proteinbasierte) makromolekulare Komplexe, auch im Wechselspiel innerhalb ihres zellulären Kontextes, von besonderem Interesse. Wir wenden im Wesentlichen fluoreszenzbasierte Techniken an, welche die Möglichkeiten bieten Proteineigenschaften und Interaktionen mit hoher Empfindlichkeit und Selektivität zu untersuchen. Durch die Entwicklung sehr leuchtstarker und photo-stabiler Fluoreszenzfarbstoffe und durch die enormen methodischen Fortschritte in der hochauflösenden Fluoreszenzmikroskopie/Spektroskopie sind heutzutage Studien an Proteinen bis hin zum Einzelmolekülniveau möglich.“ https://www.institut-1a.physik.rwth-aachen.de/cms/INSTITUT-1A/Forschung/~iecj/Biophysik/ **Einordnung:** Beispiel für Abschnitt 2.2 Forschung an den noch bestehenden Einschränkungen bei der Proteinstrukturvorhersage; hier: Protein-Komplexe, Dynamik von Faltungsvorgängen[136]
Uniklinik RWTH Aachen Comprehensive Diagnostic Center Aachen „Ziel des CDCA ist es, in Zusammenarbeit zwischen Instituten und Kliniken der Uniklinik RWTH Aachen und anderer Fakultäten der RWTH geeignete Werkzeuge zu entwickeln, zu testen und zu etablieren, um die sehr unterschiedlichen diagnostischen Daten automatisiert auszuwerten, zusammenzuführen und somit für das Wohl des Patienten effizient nutzbar zu machen. Klinische Datenbanken sollen aufgebaut und die Qualität und Robustheit der diagnostischen Daten und Abläufe sollen sorgfältig überprüft werden. Es sollen übergreifende Datenformate generiert werden, die es dem Computer erlauben, Informationen aus der Histopathologie, OMICs und Bildgebung effizient zusammenzuführen und gemeinsam auszuwerten. Auch die Automatisierung der Datenanalyse erfordert eine enge Kooperation zwischen Ärzten, Informatikern, Mathematikern, Ingenieuren und naturwissenschaftlichen Grundlagenforschern. Um die Daten in krankheitsmechanistische Zusammenhänge zu bringen und hierdurch für die Entwicklung von medikamentösen und nicht-medikamentösen Therapiestrategien in vollem Umfang nutzbar zu machen, bedarf es ferner Systembiologen mit profunder Expertise in der mathematischen Modellentwicklung und Simulation. Neben dem Zugriff auf umfangreiche Datenbanken ist demnach eine Vielzahl von strukturellen Maßnahmen notwendig, die mit der Entwicklung leistungsfähiger digitaler Infrastruktur unter **Nutzung künstlicher Intelligenz und von Maschinenlernen** einhergehen.“ https://www.ukaachen.de/kliniken-institute/comprehensive-diagnostic-center-aachen-cdca/zentrum/ziele-aufgaben/ **Einordung:** Beispiel für Abschnitt 2.4 Neue KI-Anwendungen in den Lebenswissenschaften; bspw. auch Verweis auf Erfolgsprinzipien der Proteinstrukturvorhersage.[137]
Uniklinik RWTH Aachen Institut für Humangenetik und Genommedizin „Die Humangenetik hat sich durch den Einsatz neuer molekularer Techniken wie der Hochdurchsatzsequenzierung (Next-Generation-Sequencing) und der fortschreitenden Bioinformatik rasant entwickelt und bietet Ihnen damit modernste Möglichkeiten zur Diagnostik genetisch-bedingter Erkrankungen.“ https://www.ukaachen.de/kliniken-institute/institut-fuer-humangenetik-und-genommedizin/institut/ **Einordnung:** Beispiel für Kapitel 2 Proteinstrukturvorhersage in den Lebenswissenschaften; Nutzung der Proteinstrukturvorhersage[138,139]

[136] Bustorff, N. & Fitter, J. Features of Protein Unfolding Transitions and Their Relation to Domain Topology Probed by Single-Molecule FRET. Biomolecules 13, 1280 (2023). https://doi.org/10.3390/biom13091280.
[137] Han, T., Kather, J.N., Pedersoli, F. et al. Image prediction of disease progression for osteoarthritis by style-based manifold extrapolation. Nat Mach Intell 4, 1029–1039 (2022). https://doi.org/10.1038/s42256-022-00560-x.
[138] Mattern, L., Begemann, M., Delbrück, H. et al. Variant of the catalytic cysteine of UFSP2 leads to spondyloepimetaphyseal dysplasia type Di Rocco. Bone Reports 18, 101683 (2023). https://doi.org/10.1016/j.bonr.2023.101683.
[139] Mensah, M.A., Niskanen, H., Magalhaes, A.P. et al. Aberrant phase separation and nucleolar dysfunction in rare genetic diseases. Nature 614, 564–571 (2023). https://doi.org/10.1038/s41586-022-05682-1.

Universität Bielefeld Fakultät für Chemie, Arbeitsbereich Biochemie, AG Strukturbiochemie, Prof. Dr. Hartmut Niemann „Die Kenntnis der dreidimensionalen Struktur von Molekülen ist essentiell zum Verständnis ihrer Funktion. Wir bestimmen die Raumstruktur von Proteinen und makromolekularen Komplexen mit Hilfe der Röntgenkristallographie. Zur funktionellen Charakterisierung verwenden wir biochemische und zellbiologische Methoden. Unser Hauptinteresse gilt bakteriellen Virulenzfaktoren und deren Wechselwirkung mit dem Wirt.“ https://www.uni-bielefeld.de/fakultaeten/chemie/ag/bc4-niemann/index.xml **Einordnung:** Nutzung von Proteinstrukturvorhersagen[140] - Abschnitt 2.2 Forschung an den noch bestehenden Einschränkungen bei der Proteinstrukturvorhersage.
Universität Bielefeld Technische Fakultät BIBI - Bielefelder Institut für Bioinformatik-Infrastruktur „The Bielefeld Institute for Bioinformatics Infrastructure (BIBI) was founded in summer 2019 and slightly restructured in spring 2021. It pools the bioinformatics service and training activities at the Faculty of Technology, Faculty of Biology and Center for Biotechnology (CeBiTec) at Bielefeld University. The goal is to create a powerful institute offering research infrastructure for all areas in the digital life sciences - together with ZB MED, Information Center for Life Sciences in Köln and Bonn, and in cooperation with the German Network for Bioinformatics Infrastructure (de.NBI), ELIXIR Germany and the Institute of Bio- and Geosciences (IBG-5) at Forschungszentrum Jülich.” https://www.uni-bielefeld.de/fakultaeten/technische-fakultaet/forschung/forschungszentren/bibi/index.xml **Einordnung:** Zur Nutzung und Weiterentwicklung von Proteinstrukturvorhersagen in Wissenschaft (Kapitel 2) und Wirtschaft sind auch leistungsfähige Bioinformatik-Infrastrukturen erforderlich.
Ruhr-Universität Bochum Biochemie I - Rezeptorbiochemie Erforschung von Glutamatrezeptoren https://www.ruhr-uni-bochum.de/bc1/forschung/index.html **Einordnung:** Nutzung von AlphaFold.[141] Bezug zu Abschnitt 2.2.
Ruhr-Universität Bochum Biochemie II Molekulare Biochemie „Zellen von Pflanzen, Tieren und Menschen verfügen über ein komplexes System verschiedener Membranen, die eine intrazelluläre Kompartimentierung ermöglichen. Jede Membran besteht aus einer Lipiddoppelschicht und darin eingebetteten Proteinen, die zahlreiche, essentielle Funktionen erfüllen. Kanäle und Transporter kontrollieren den Transport von Ionen, Lipiden und Metaboliten. An der Signalweiterleitung sind Rezeptoren beteiligt, die Veränderungen in der zellulären Umgebung detektieren und zellspezifische Antworten initiieren. Störungen in den zellulären Transport und Signalsystemen verursachen oft schwere Erkrankungen beim Menschen. Ziel unserer Forschungsarbeiten ist es, die molekularen Mechanismen der Transportsysteme und Signaltransduktion aufzuklären.“ https://www.ruhr-uni-bochum.de/bc2/forschung/index.html **Einordnung:** Proteine erfüllen wie oben angeführt wesentliche Funktionen für Membrane, die von deren Struktur abhängen. Nutzung von AlphaFold.[142] Bezug zu Abschnitt 2.2.
Ruhr-Universität Bochum Fakultät für Biologie und Biotechnologie, Forschungsbereich Proteinforschung und Biotechnologie „Im Forschungsbereich Proteinforschung und Biotechnologie wird die Rolle von Proteinen und RNA-Molekülen bei der Steuerung biologischer Prozesse und die Möglichkeiten ihrer Nutzung für die Biokatalyse untersucht. Dabei setzen die beteiligten Lehrstühle und Arbeitsgruppen ein breites Spektrum von bioinformatischen, molekularbiologischen, biochemischen und biophysikalischen Methoden ein, um Fragen von der

[140] Gilzer, D., Schreiner, M. & Niemann, H.H. Direct interaction of a chaperone-bound type III secretion substrate with the export gate. Nat Commun 13, 2858 (2022). https://doi.org/10.1038/s41467-022-30487-1.

[141] „In the absence of solved structures […] predicted structures […] were used which have been mapped using the predictions of DeepMind's new artificial intelligence called AlphaFold“ vgl. Amina, B., Leifeld, J., Shaukat, J. et al. Tetraspanins as Potential Modulators of Glutamatergic Synaptic Function. Frontiers in Molecular Neuroscience 14 (2022). https://doi.org/10.3389/fnmol.2021.801882.

[142] „Currently, no comparable structural data are available for plastidic Alb3 proteins. We therefore used the structure prediction from the EMBL-EBI-AlphaFold“ vgl. Ackermann, B., Dünschede, B., Pietzenuk B. et al. Chloroplast Ribosomes Interact With the Insertase Alb3 in the Thylakoid Membrane. Frontiers in Plant Science 12 (2021). https://doi.org/10.3389/fpls.2021.781857.

Grundlagenforschung bis hin zur Translation in die Medizin oder die Biotechnologie zu untersuchen. Erkenntnisse aus Untersuchungen auf atomarer und molekularer Ebene werden auf die Ebene von Zellen und Organismen übertragen, um komplexe physiologische Vorgänge zu verstehen.“ https://www.biologie.ruhr-uni-bochum.de/biodek/fakultaet/ls_ag/index.html.de#Proteinforschung%20und%20Bioengineering **Einordung:** Bezug inhaltlich offensichtlich aufgrund des Schwerpunkts Proteinforschung und der Anwendung bioinformatischer Methoden, daher zu Abschnitt 2.2 sowie Kapitel 5 Biokatalyse - Enzyme und Kapitel 7 Gesundheit - Medikamentenentwicklung. Auch Nutzung von AlphaFold.**Fehler! Textmarke nicht definiert.**
Universität Bonn Institut für Strukturbiologie „We are interested in the structure–function relationship of biological macromolecules.“ http://www.isb.uni-bonn.de/ AG Membranproteine „In humans, 25% of all protein coding genes encode membrane proteins. Many of them are important drug targets.“ http://www.isb.uni-bonn.de/research/regulation-of-transcription-5 **Einordnung:** Schwerpunkt Strukturbiologie und Nutzung von AlphaFold.[143] Bezug zu Abschnitt 2.2 sowie Kapitel 7 Gesundheit - Medikamentenentwicklung
Max-Planck-Institut für molekulare Physiologie, Dortmund Abteilung Mechanistische Zellbiologie „Das Wunder des Wachstums wird durch die Wiederholung eines Prozesses möglich gemacht, der jedes Mal fehlerfrei ablaufen muss: die Zellteilung, bei der aus einer Zelle zwei identische Tochterzellen entstehen. In unserer Abteilung untersuchen wir die unterschiedlichen molekularen Mechanismen, die gewährleisten, dass das Erbmaterial korrekt auf die beiden Tochterzellen verteilt wird. Dazu erforschen wir die Struktur und Funktion einer Vielzahl von Proteinen, die daran beteiligt sind.“ Abteilung Strukturbiochemie „Zellen stehen im ständigen Austausch mit ihrer Umgebung. Eine maßgebliche Rolle dabei spielen Proteine, die in die Membran eingebettet sind. Ziel unserer Forschung ist es zum einen, die Funktionsweise solcher Membranproteine zu verstehen. Dafür ist es unabdingbar, zunächst ihre Struktur aufzuklären. Eine Technik, die wir dazu verwenden, ist die Elektronenkristallographie von zweidimensionalen Kristallen. Zum anderen untersuchen wir größere Komplexe von Proteinen, molekulare Maschinen, die lebenswichtige Funktionen ausführen. Da größere Protein-Komplexe jedoch nur schwer zu kristallisieren sind, setzten wir zur Strukturanalyse die Kryo-Elektronenmikroskopie ein.“ https://www.mpi-dortmund.mpg.de/forschung/abteilungen **Einordnung:** Bezugnahme auf den Zusammenhang von Struktur und Funktion von Proteinen, Betrachtung von Protein-Komplexen zus. mit Nutzung von AlphaFold.[144,145,146,147] Passend zu Abschnitt 2.2.

[143] Peter, M.F., Ruland, J.A., Depping, P. et al. Structural and mechanistic analysis of a tripartite ATP-independent periplasmic TRAP transporter. Nat Commun 13, 4471 (2022). https://doi.org/10.1038/s41467-022-31907-y.
[144] Altmannova, V., Firlej, M., Müller, F. et al. Biochemical characterisation of Mer3 helicase interactions and the protection of meiotic recombination intermediates. Nucleic Acids Research 51, 9, 4363–4384 (2023). https://doi.org/10.1093/nar/gkad175.
[145] Wang, Z., Grange, M., Pospich, S. et al., Structures from intact myofibrils reveal mechanism of thin filament regulation through nebulin. Science 375, 6582, eabn1934 (2022). https://doi.org/10.1126/science.abn1934.
[146] Chen, C., Piano, V., Alex, A. et al. The structural flexibility of MAD1 facilitates the assembly of the Mitotic Checkpoint Complex. Nat Commun 14, 1529 (2023). https://doi.org/10.1038/s41467-023-37235-z.
[147] Rai, A., Singh, A.K., Bleimling, N. et al. Rep15 interacts with several Rab GTPases and has a distinct fold for a Rab effector. Nat Commun 13, 4262 (2022). https://doi.org/10.1038/s41467-022-31831-1.

Heinrich-Heine-Universität Düsseldorf Center for Structural Studies (CSS) „Das CSS verfolgt wissenschaftliche Aufgaben und Ziele im Bereich der strukturenliefernden Verfahren, die räumliche Auflösungen im atomaren Bereich ermöglichen […]“ Forschungsbereich Computergestützte Methoden - Modeling und Simulation „Mit den unterschiedlichsten computergestützten Methoden ist es möglich ein breites Spektrum an Fragestellungen zu beantworten, die von rein statischen Modellen bis hin zum Einfluss von Proteinmutationen auf die Dynamik eines Proteins oder Komplexes reichen.“ https://www.css.hhu.de/dienstleistungen/modeling-und-simulation **Einordnung:** Beispiel für die Verbindung von experimentellen Methoden mit der Proteinstrukturvorhersage - Bezug zu Abschnitt 2.2, gesprochen wird auch vom „virtuellem Screening neuer Leitstrukturen zur Wirkstoffentwicklung“ - Bezug zu Kapitel 7 Gesundheit - Medikamentenentwicklung.
Forschungszentrum Jülich Institut für Bio- und Geowissenschaften (IBG) Biotechnology (IBG-1) mit Teilbereich, “Molecular Enzyme Technology” „The topic of ‚Molecular Enzyme Technology” headed by Prof. Jaeger deals with the identification, isolation and characterization of novel bacterial enzymes.“ Bioinformatics (IBG-4) „Research in bioinformatics at Forschungszentrum Jülich plays a leading role on national and international level […] in the area of structure-based bioinformatics focusing on understanding, predicting and modulating of the dynamics and interactions of biomolecules.“ https://www.fz-juelich.de/en/ibg **Einordnung:** IBG-1 mit Nutzung von AlphaFold,[148] s.a. Eintrag zum Institut Molekulare Enzymtechnologie (IMET) in Abschnitt 11.3.3. IBG-4 aus obiger Selbstbeschreibung. Passt zu Abschnitt 2.2.
Forschungszentrum Jülich Institute for Biological Information Processing „Our goal is to quantitatively and qualitatively understand the cellular and molecular mechanisms of biological information processing down to the atomic level. The focus here is on information processing in and between nerve cells.“ https://www.fz-juelich.de/en/ibi Institutsbereich Molecular and Cellular Physiology (IBI-I-1) „Ion channels, ion transporters and receptors are critical for the generation, the propagation and the transmission of cellular signals. We use a combination of experimental and computational approaches to study the function of these proteins at the molecular level and to understand their contributions to cell and organ functions.“ https://www.fz-juelich.de/en/ibi/ibi-1 Institutsbereich Structural Biochemistry (IBI-I-7) „The function of every cell and organism depends critically on the dynamic interactions between biological macromolecules and on their correct three-dimensional structure. Faulty interactions and misfolded structures lead to disease and aging. In IBI-7 - Structural Biochemistry, research aims to understand these interactions and to determine the three-dimensional structure of protein complexes involved in crucial cellular processes.“ https://www.fz-juelich.de/en/ibi/ibi-7 Institutsbereich Cellular Structural Biology (IBI-6) verlinkt auf (ER-C-3) s. nächster Eintrag **Einordnung:** Ausgewählte Beschreibungen inhaltlich einschlägig. Nutzung von AlphaFold.[149] Passend zu Abschnitt 2.2.
Forschungszentrum Jülich Institut Ernst Ruska-Centrum für Mikroskopie und Spektroskopie mit Elektronen (ER-C) Institutsbereich Strukturbiologie (ER-C-3)

[148] Krüger, A. & Frunzke, J. A pseudokinase version of the histidine kinase ChrS promotes high heme tolerance of Corynebacterium glutamicum. Frontiers in Microbiology 13 (2022). https://doi.org/10.3389/fmicb.2022.997448.
[149] Guzman, R.E., Sierra-Marquez, J., Bungert-Plümke, S. et al. Functional Characterization of CLCN4 Variants Associated With X-Linked Intellectual Disability and Epilepsy. Frontiers in Molecular Neuroscience 15 (2022). https://doi.org/10.3389/fnmol.2022.872407.

„Am ER-C-3 verwenden wir einen umfassenden elektronenmikroskopischen Ansatz, um die biologischen Strukturen von membranassoziierten Proteinkomplexen zu untersuchen. Unsere wichtigsten Untersuchungsmethoden sind die Einzelteilchen-Elektronenkryomikroskopie (Kryo-EM) sowie die Elektronenkryo-Tomographie (Kryo-ET), die wir auch entwickeln, um bestehende Bildgebungstechnologien in Richtung hochauflösender Strukturbiologie zu verbessern. Weitere Informationen erhalten Sie auf unseren englischsprachigen Seiten."
https://www.fz-juelich.de/de/er-c/er-c-3
Einordnung: Proteinstrukturvorhersagen lt. Literatur teils nützlich als Modelle zur Interpretation der experimentellen Daten. Nutzung von AlphaFold.**Fehler! Textmarke nicht definiert.**,[150] Bezug zu Abschnitt 2.2.

Forschungszentrum Jülich
Institute for Advanced Simulation (IAS) - Jülich Supercomputing Centre (JSC)
Institutsbereich Computational Biomedicine (IAS-5 / INM-9)
„The institute develops and uses molecular simulation, machine learning and bioinformatics tools to investigate neurotransmission in the human brain."
https://www.fz-juelich.de/en/inm/inm-9
Einordnung: Inhaltliche Bezüge maschinelles Lernen und molekulare Simulation. Nutzung von AlphaFold.[151] Passt zu den Abschnitten 2.2 sowie 2.4.

Universität zu Köln
Institut für Genetik
„Das 1962 von Max Delbrück gegründete Institut für Genetik umfasst derzeit 16 Lehrstühle und mehrere unabhängige Nachwuchsgruppen, die Spitzenleistungen in Forschung und Lehre fördern. Die Forschung am Institut für Genetik deckt ein breites Spektrum an Themen ab, das von der Molekular- und Zellbiologie bis hin zur Computational Biology reicht. Molekulare Mechanismen auf allen Ebenen werden in Modellorganismen, Zellkulturen und in silico untersucht. Das Institut bietet Zugang zu gemeinsamer Infrastruktur und zentralen Serviceeinheiten, die es jungen Forschern ermöglichen, in einem exzellenten internationalen Forschungsumfeld an seinen beiden Standorten, dem Biozentrum und dem CECAD-Forschungszentrum für altersassoziierte Krankheiten, zu arbeiten."
https://genetik.uni-koeln.de
Forschungsgruppe Hofmann - Protein Evolution
„The overarching research topic of our group is the evolutionary history of present-day biochemical pathways and their protein components: How did proteins evolve and diversify their functionality?"
http://www.uni-koeln.de/math-nat-fak/genetik/groups/Hofmann/index.html
Einordnung: Nutzung von AlphaFold[152] und AlphaFold-Multimer.[153] Passt zu Abschnitt 2.2.

Deutsche Zentralbibliothek für Medizin (ZB MED) - Informationszentrum Lebenswissenschaften, Köln
„ZB MED – Informationszentrum Lebenswissenschaften ist das nationale Infrastruktur- und Forschungszentrum für lebenswissenschaftliche Daten und Informationen."
https://www.zbmed.de/
Einordnung: Zur Nutzung und Weiterentwicklung von Proteinstrukturvorhersagen sind auch leistungsfähige Infrastrukturen erforderlich, um den Zugang zu relevanten lebenswissenschaftlichen Daten zu gewährleisten. .

Universität Münster
Institut für Medizinische Physik und Biophysik
AG KryoEM komplexer Nanosysteme
Forschungsthema: Peroxisomale Biogenese

[150] Gewehr, L., Junglas, B., Jilly, R. et al. SynDLP is a dynamin-like protein of Synechocystis sp. PCC 6803 with eukaryotic features. Nat Commun 14, 2156 (2023). https://doi.org/10.1038/s41467-023-37746-9.
[151] Mueller, N.P.F., Carloni, P. & Alfonso-Prieto, M. Molecular determinants of acrylamide neurotoxicity through covalent docking. Front Pharmacol 14, 1125871 (2023). https://doi.org/10.3389/fphar.2023.1125871.
[152] Grønbæk-Thygesen, M., Kampmeyer, C., Hofmann, K. et al. The moonlighting of RAD23 in DNA repair and protein degradation. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1866, 2, 194925 (2023). https://doi.org/10.1016/j.bbagrm.2023.194925.
[153] Cerullo, F., Filbeck, S., Patil, P.R. et al. Bacterial ribosome collision sensing by a MutS DNA repair ATPase paralogue. Nature 603, 509–514 (2022). https://doi.org/10.1038/s41586-022-04487-6.

„Unsere Forschung konzentriert sich auf das molekulare Verständnis der peroxisomalen Biogenese. Peroxisomen sind dynamische kleine Organellen, die in fast allen Eukaryoten vorkommen. […] Peroxisomale Dysfunktion und gestörter peroxisomaler Import führen zu verheerenden, angeborenen Stoffwechselstörungen, was die Bedeutung des Verständnisses dieser grundlegenden biologischen Fragen weiter unterstreicht."
https://www.medizin.uni-muenster.de/impb/das-institut/cryoem/forschung.html
Einordnung: Nutzung von AlphaFold[154] - Bezug zu Abschnitt 2.2.

11.2. Protein-Engineering und Proteindesign

11.2.1. Neuartige Protein-basierte Zuckerersatzstoffe - Novel Sweets

Das Fraunhofer IME erforscht zusammen mit zwei Industriepartnern im Rahmen des Projektes Novel Sweets mit Förderung des Bundesministeriums für Ernährung und Landwirtschaft neue Süßungsmittel auf Proteinbasis.[155] Hintergrund ist die Tatsache, dass mit dem steigenden Zuckerkonsum der letzten Jahrzehnte auch bestimmte chronische Erkrankungen weltweit stark zugenommen haben. Genannt werden Karies, Bluthochdruck, Schlaganfall, Herzinfarkt, Diabetes Typ 2, Übergewicht und Fettleibigkeit bei Kindern und Erwachsenen, die durch den langfristigen Verzehr von Zucker verursacht bzw. begünstigt werden. Daher werden in dem Projekt süß schmeckende Proteine untersucht, um gesündere Alternativen zu Zucker zu entwickeln und den genannten Krankheiten entgegenzuwirken. Stefan Rasche, Abteilungsleiter am Fraunhofer IME, erklärt im Interview, dass in dem Projekt nach einem Weg gesucht wird, die natürlich vorkommenden süß schmeckenden Proteine biotechnologisch nachzubauen. Ausgangspunkt dafür seien die 3-D-Strukturen der fünf Proteine. Dann würden mit Hilfe von Computerverfahren die Bereiche ausfindig gemacht, mit denen die Proteine an die menschlichen Rezeptoren für Süßgeschmack binden, mit der Absicht, diese Bereiche zu modifizieren, um die Süßkraft und Stabilität der untersuchten Proteine zu verbessern.[156]

11.2.2. Biokonversion von Kohlendioxid

Das EU-Forschungsprojekt BIOCON-CO2 lief in der Zeit von Jan. 2018 bis Juni 2022 und wurde mit 6.9 Mio. € gefördert,[157] um biologische Prozesse zu entwickeln und zu validieren, mit denen Kohlendioxid aus der Eisen-, Stahl- und Zementindustrie sowie aus der Stromerzeugung für die Herstellung von Chemikalien und Kunststoffen genutzt werden kann. Mit dem Fraunhofer IME war eine Projektpartner aus NRW beteiligt.[158]

Eine der Aufgaben bestand darin, eine Klasse von Bakterien mit dem Namen Clostridia auf ihre Eignung zur Verwendung in den angestrebten, biologischen Prozessen zu untersuchen. Darüber hinaus sollten geeignete Bakterienstämme für eine verbesserte Produktion bestimmter Alkoholverbindungen und mit einer erhöhten Sauerstofftoleranz entwickelt und unter industriell relevanten Bedingungen getestet werden.

Im Rahmen des Projektes wurden auch Ansätze des Protein-Engineering zur Optimierung von Enzymen diskutiert.[159]

11.2.3. NRW Akteure

Beispiele für NRW-Akteure - Protein-Engineering / Proteindesign
Unternehmen

[154] Rüttermann, M. & Gatsogiannis, C. Good things come to those who bait: the peroxisomal docking complex. Biological Chemistry 404, 2–3, 107–119 (2023). https://doi.org/10.1515/hsz-2022-0161.
[155] Vgl. Projektdarstellung: Neuartige Protein-basierte Zuckerersatzstoffe mit verbesserten gustatorischen und prozesstechnischen Eigenschaften - Novel Sweets, Fraunhofer IME https://www.ime.fraunhofer.de/de/Forschungsbereiche/MB/Pflanzenbiotechnologie/Projekte/novelsweets.html.
[156] Interview mit Stefan Rasche, Wie süß können Proteine sein? Magazin Forschungsfragen 2, 34 (2024). https://www.ime.fraunhofer.de/content/dam/ime/de/documents/MB/forschungsfelder-2-2023_NovelSweets_Interview_Rasche.pdf.
[157] Projektwebseite BIOCON-CO2, Übersicht: https://biocon-co2.eu/about/project-overview/.
[158] Projektwebseite BIOCON-CO2, Konsortium: https://biocon-co2.eu/about/consortium/.
[159] Projektwebseite BIOCON-CO2, Workshop-Ankündigung: https://biocon-co2.eu/new-year-new-workshop/.

SeSaM-Biotech GmbH, Aachen - Profil s. Abschnitt 11.3.3
Enzymaster Deutschland GmbH, Düsseldorf oder Hilden - Profil s. Abschnitt 11.3.3
Forschungseinrichtungen
Fraunhofer IME - Profil s. Abschnitt 11.4.2
RWTH Aachen Lehrstuhl für Biotechnologie AG Schwaneberg - Protein Engineering mit evolutiven und rationalen Methoden „Die Forschung ist in fünf Abteilungen gegliedert mit dem Ziel die Methoden der Gelenkten Evolution zu verbessern, neue Designprinzipien von Proteinen aufzufinden und die Translationsforschung für interaktive Materialien (Pflanzengesundheit; Medizin), Biokatalyse und Bioökonomie zu befördern.“ https://www.biotec.rwth-aachen.de/go/id/imne **Einordnung:** Stichwort Protein-Engineering sowie Designprinzipien von Proteinen, daher passend zu Kapitel 3 bzw. Kapitel 4 Proteindesign.
DWI – Leibniz-Institut für Interaktive Materialien, Aachen Abteilung Biohybrid Systems „Die Abteilung Biohybrid Systems entwickelt und optimiert Proteine, Enzyme und Peptide für den Einsatz in Biohybridmaterialien mit maßgeschneiderten Eigenschaften und Funktionalitäten. Die Schlüsseltechnologieplattform besteht aus > 50 Peptiden, welche in Wasser bei Raumtemperatur an vielfältige Oberflächen binden. Wir haben uns das Ziel gesetzt, die Interaktionen zwischen den zur Immobilisierung eingesetzten Ankerpeptiden und den Oberflächen auf molekularer Ebene zu verstehen.“ https://www.dwi.rwth-aachen.de **Einordnung:** „entwickelt und optimiert Proteine, Enzyme und Peptide“ passt zu Kapitel 3 und 4.
Heinrich-Heine-Universität Düsseldorf Arbeitsgruppe Synthetic Membrane Systems „Membrane protein folding: „We employ biophysical and structural methods to elucidate the structure and dynamics of the complex along the membrane protein biogenesis.“ Crowding in biological membranes: „development of de novo proteins“ https://www.synmemsys.hhu.de/en/ **Einordnung:** Nutzung von AlphaFold[160] - Stichworte „de novo Proteine“ und synthetische Membran-Systeme. Bezug zu Abschnitt 2.2 sowie Kapitel 4 Proteindesign

11.3. Biokatalyse

11.3.1. Projekt FutureEnzyme

Im Rahmen des EU-finanzierten Projekts FuturEnzyme[161] sollen die Entdeckung, Konzeption, Optimierung und Formulierung kostengünstiger Enzyme für wirtschaftlich tragfähige Produkte in der Textil-, Waschmittel- und Kosmetikbranche verbessert werden.[162] Die angestrebte Entwicklungsplattform für Enzyme soll u. a. Big-Biodata-Mining aus öffentlichen und internen Datenbanken, maschinelles Lernen, Bioprospektion, Protein-Engineering, Nanobiotechnologie, Fermentation und Verarbeitungssysteme verbinden. Das Projekt hat eine Laufzeit vom 1. Juni 2021 bis zum 31. Mai 2025 und eine Fördersumme von 6 Mio. €. Von den 16 Partnern stammen 4 aus NRW, darunter zwei Industriepartner: Henkel und Evonik sowie das Institut für Molekulare Enzymtechnologie der Heinrich-

[160] Knospe, C.V., Kamel, M., Spitz, O. et al. The structure of MadC from Clostridium maddingley reveals new insights into class I lanthipeptide cyclases. Front Microbiol 13, 1057217 (2023). https://doi.org/10.3389/fmicb.2022.1057217 (zus. mit „Institute of Biochemistry“ und „Center for Structural Studies“, Heinrich-Heine-Universität Düsseldorf).
[161] Projekt-Webseite: https://www.futurenzyme.eu.
[162] Cordis-Datenbank, Projektdatenblatt FutureEnzyme, https://cordis.europa.eu/project/id/101000327.

Heine-Universität Düsseldorf und der Cluster Industrielle Biotechnologie. In einem Übersichtsartikel zum Projekt wird auf die Bedeutung der Proteinstrukturvorhersage für die Entwicklung neuer Enzyme hingewiesen.[163]

11.3.2. ESP – Enzyme-Substrat-Prognose

Ein Team unter Beteilung von Prof. Lercher an der Heinrich-Heine-Universität Düsseldorf hat ein allgemeines maschinelles Lernmodell zur Vorhersage von Enzym-Substrat-Paaren (genannt ESP) entwickelt, das eine Genauigkeit von über 91 % auf Grundlage unabhängiger und vielfältiger Testdaten erreicht.[164] Den Autoren zufolge kann ESP erfolgreich auf verschiedenste Enzyme und ein breites Spektrum an Metaboliten angewendet werden, die in den Trainingsdaten enthalten sind, und übertrifft bekannte Modelle, die für einzelne, gut untersuchte Enzymfamilien entwickelt wurden. ESP stellt Enzyme durch ein modifiziertes Transformatormodell dar und wird anhand von Daten trainiert, die mit zufällig ausgewählten kleinen Molekülen angereichert sind. Mit Blick auf zukünftige Entwicklungen bemerken die Autoren, dass AlphaFold und RoseTTAFold Proteinstrukturvorhersagen erleichtern und somit zukünftige Arbeiten auch Eingabemerkmale umfassen könnten, die aus solchen vorhergesagten Enzymstrukturen extrahiert wurden. Dies könnte nach Einschätzung der Autoren dazu beitragen, den experimentellen Aufwand – und die damit verbundenen Kosten – biochemischer Tests auf ein Niveau zu senken, das Labortests praktikabel macht.

11.3.3. NRW-Akteure

Beispiele für NRW-Akteure - Biokatalyse –Enzyme
Unternehmen
Henkel AG & Co. KGaA, Düsseldorf „Henkel holds leading positions in both industrial and consumer businesses: Our portfolio includes well-known hair care products, laundry detergents, fabric softeners as well as adhesives, sealants, and functional coatings.“ https://www.henkel.com **Einordnung:** Partner im Projekt FuturEnzyme, s. o. Abschnitt 11.3.1
Evonik Industries AG, Essen „Evonik ist ein weltweit führendes Unternehmen der Spezialchemie.“ https://corporate.evonik.com/de **Einordnung:** Partner im Projekt FuturEnzyme, s. o. Abschnitt 11.3.1
SeSaM-Biotech GmbH, Aachen „We love to contribute to your success with our expertise in Directed Enzyme Evolution covering the areas of Protein Engineering, Assay Development, Gene Library Design and more. Get in touch to learn more about our Patented Protein Engineering – Made in Germany.“ https://www.sesam-biotech.com/ **Einordnung:** Stichworte Enzyme, Protein-Engineering. Daher auch Bezug zu Kapitel 3 Protein-Engineering.
Enzymaster Deutschland GmbH, Düsseldorf oder Hilden (Tochter der Enzymaster (Ningbo) Bio-Engineering Co. Ltd., Ningbo (China)) „Enzymaster has developed the BioEngine®, a proprietary directed evolution platform with integrated computational enzyme engineering. The implementation of computational hot spot identification, in silico enzyme library screening, and in silico recombination into the directed evolution cycle allows us to shift most of the screening efforts from the laboratory into the computer and cover a large sequence space of the enzyme, all while accounting for the real chemical process conditions.“ https://enzymaster.de/technology/

[163] „Indeed, the use of bioinformatics, machine learning, accurate protein structure prediction and data-driven artificial intelligence (AI) techniques are essential to fully exploit the potential of sequencing data as a source of new enzymes […]“ vgl. Molina-Espeja, P., Sanz-Aparicio, J., Golyshin, P.N. et al. Enzymes for consumer products to achieve climate neutrality. Oxford Open Climate Change 3, 1, kgad003 (2023). https://doi.org/10.1093/oxfclm/kgad003.

[164] Kroll, A., Ranjan, S., Engqvist, M.K.M. et al. A general model to predict small molecule substrates of enzymes based on machine and deep learning. Nat Commun 14, 2787 (2023). https://doi.org/10.1038/s41467-023-38347-2 und https://esp.cs.hhu.de.

https://enzymaster.de/imprint/ https://enzymaster.de/who-we-are/ **Einordnung:** Enzym-Engineering. Insofern auch Bezug zu Kapitel 3 Protein-Engineering.
evoxx technologies GmbH, Monheim (Europa-Zentrale der AETL Gruppe, Indien) „evoxx technologies is an industrial biotech company focusing on the development and production of industrial enzymes […]“ https://evoxx.com/company/ https://evoxx.com/company/about-us/ „We make use of nature´s diversity to identify enzymes that have commercial potential. Over the last decade, large metagenomic libraries from different habitats for targeted screenings has been developed […] By focusing on specific catalytic reactions and other defined parameters, we identify suitable enzymes based on activity, sequences, structural properties and protein models.“ https://evoxx.com/services/enzyme-development/ **Einordnung:** Stichworte Entwicklung von Enzymen, Identifikation von Enzymen anhand von Sequenzen, Strukturen und Protein-Modellen.
Forschungseinrichtungen
Heinrich-Heine-Universität Düsseldorf Institut Molekulare Enzymtechnologie (IMET) „Das IMET als Institut der Heinrich-Heine-Universität Düsseldorf ist auf dem Gelände des Forschungszentrums Jülich (FZJ) in einen Zusammenschluss (IBG-1) mehrerer Institute eingebunden.“ https://www.iet.uni-duesseldorf.de **Einordnung:** Partner im Projekt FuturEnzyme, s. o. Abschnitt 11.3.1

Heinrich-Heine-Universität Düsseldorf AG Computational Cell Biology „Wir betrachten mechanistische Modelle als eine quantitative, umfassende Zusammenfassung unseres gegenwärtigen Verständnisses biologischer Systeme, aufbauend auf Physik und Chemie. In unseren Forschungsprojekten geht es um die folgenden Themen: • […] • Methoden der künstlichen Intelligenz zur Vorhersage von Eigenschaften von Enzymen und Transportern (benötigt zur Parameterisierung der Modelle).“ https://www.cs.hhu.de/lehrstuehle-und-arbeitsgruppen/computational-cell-biology **Einordnung:** AG, auf deren Arbeiten sich Abschnitt 11.3.2 bezieht. Mit zusätzlichem Bezug zu Abschnitt 2.4.
Cluster Industrielle Biotechnologie e.V. (CLIB) **Einordnung:** Partner im Projekt FuturEnzyme, s. o. Abschnitt 11.3.1 sowie Abschnitt 11.4.2.
Universität Bielefeld, Technische Fakultät, Centrum für Biotechnologie (CeBiTec), s. Abschnitt 11.4.2
Forschungszentrum Jülich Institutsbereich Bioinformatics (IBG-4) „Der Institutsbereich Bioinformatik (IBG-4) des Instituts für Bio- und Geowissenschaften bearbeitet und entwickelt Methoden und Algorithmen, um ein grundlegendes Verständnis von hochdimensionalen Daten und Prozessen in der Bioökonomie/in den Lebenswissenschaften zu erzielen und um die biologischen Funktionen von Molekülen/Enzymen anhand ihrer Strukturen zu verstehen, vorherzusagen und für industrielle und pharmazeutische Anwendungen zu optimieren.“ https://www.fz-juelich.de/de/ibg/ibg-4 **Einordnung:** Nutzung von ColabFold[165] in der Forschung an Enzymen. Bezug zu Kapitel 5 Biokatalyse - Enzyme sowie zu Kapitel 7 Gesundheit - Medikamentenentwicklung aufgrund der Nennung pharmazeutischer Anwendungen von Enzymen.

11.4. Bioproduktion

11.4.1. CLIB - Bio Scale-Up center NRW

Ein Autoren-Team vom CLIB-Cluster stellt einen allgemeinen Zusammenhang von AlphaFold2 zu Fortschritten in der Bioökonomie her.[166] Sie weisen darauf hin, dass in Nordrhein-Westfalen der Bau einer Anlage für bis zu TRL 9 in Planung sei, um die Schritte bis zur Kommerzialisierung besser unterstützen zu können und dass CLIB eine Machbarkeitsstudie für dieses Bio-Scale-Up-Zentrum NRW erstellt hat. Sie stellen einen Bezug zu AlphaFold dar und folgern in Analogie, dass Hochdurchsatztechnologien gepaart mit künstlicher Intelligenz zukünftig auch eine äußerst schnelle und effiziente Optimierung von Enzymen an die Bedingungen in industriellen biotechnologischen Prozessen erlauben könnten.

11.4.2. NRW-Akteure

Beispiele für NRW-Akteure - Bioproduktion
Unternehmen
ARTES Biotechnology GmbH, Langenfeld „Best way to get your recombinant microbial product! Cell Line and Process Development based on yeast and bacteria – global players in all markets manufacture their products made with ARTES.

[165] Marasco, R., Fusi, M., Coscolín, C. et al. Enzyme adaptation to habitat thermal legacy shapes the thermal plasticity of marine microbiomes. Nat Commun 14, 1045 (2023). https://doi.org/10.1038/s41467-023-36610-0.
[166] Kircher, M., Maurer, K.-H. & Herzberg, D. KBBE: The knowledge-based bioeconomy: Concept, status and future prospects. EFB Bioeconomy Journal 2, 100034 (2022). https://doi.org/10.1016/j.bioeco.2022.100034.

Recombinant proteins as reagents – global supply for your research.“
https://artes-biotechnology.biz/; https://artes-biotechnology.biz/enzymes/; https://artes-biotechnology.biz/biosimilars/
Einordnung: Produzent von Enzymen und Biosimilars (als Nachahmerprodukt eines Biopharmazeutikums) - insofern auch Bezüge zu Kapitel 5 Biokatalyse - Enzyme sowie Kapitel 7 Gesundheit - Medikamentenentwicklung.

LenioBio GmbH, Düsseldorf
„Meet ALiCE® - Radically simple biomanufacturing
No more cell-line development. No more upstream process development.
ALiCE® is cell-based biomanufacturing without the cells.
Explore the capabilities of our protein production platform technology.
Find out more about ALiCE
One protein expression platform, any protein“
https://www.leniobio.com/about
Einordnung: Anbieter einer zellfreien Plattform zur Protein-Produktion. Die Plattform ist synergistisch zu den neuen Möglichkeiten, die durch die Proteinstrukturvorhersage eröffnet werden, weil sich die Plattform für alle Proteine eignet - auch diejenigen, deren Design nun erst möglich wird.[167] Insofern auch Bezüge zu Kapitel 5 Biokatalyse - Enzyme sowie Kapitel 7 Gesundheit - Medikamentenentwicklung.

Numaferm GmbH, Düssedorf
„Unlock the power of peptides, pepteins and proteins
Don’t let cost stand in the way of innovation. Explore and harness the potential of affordable peptides, pepteins and proteins with our disruptive bioprocesses for discovery, optimization and production.“
https://numaferm.com/
Einordnung: Anbieter einer Peptid- und Protein-Produktionsplattform.

Forschungseinrichtungen

Center for Next Generation Processes and Products (NGP²), Aachen
„Die Realisierung regionaler, stofflicher und möglichst klimaneutraler Wertschöpfungsketten stellt hierbei ein Kernziel der Forschungsaktivitäten in NGP² dar.
Die Entwicklung dieser Wertschöpfungsketten erfordert aber nicht nur neue Prozesse und Produkte, sondern auch neue verfahrenstechnische Grundlagen, Methoden und Apparate. In NGP² soll dazu die NGP²-Bioraffinerie als Forschungsplattform sowohl entlang der Wertschöpfungskette als auch zur wissenschaftlichen Methodenentwicklung etabliert werden.“
https://www.avt.rwth-aachen.de/cms/avt/die-avt/~kmbz/ngp2/
Einordnung: eingebunden in das Bioeconomy Science Center (BioSC) (s.u.)

Fraunhofer-Institut für Molekularbiologie und Angewandte Oekologie IME, Aachen/Münster:
Forschungsthemen:

- Molekulare Biotechnologie
  - Bioprozessentwicklung
    - Rationales Design von Expressionsvektoren und Genen
    - Protein Produktion
  - Funktionelle und Angewandte Genomik
    - Pflanzliche Biopolymere; z. B. Forisome[168]
  - Industrielle Biotechnologie
    - Metabolic Engineering
  - Pflanzenbiotechnologie
    - Protein Engineering; z. B. Antikörpertechnologien und molekulare Enzymevolution
    - Produktion rekombinanter Proteine

[167] Persönliche Mitteilung.
[168] „Forisomes are unique […] protein complexes that facilitate sieve element occlusion after injury [of certain plants]“ und „By understanding the structure and function of forisomes and other structural phloem proteins, we can develop a toolbox for biotechnological applications in material science and medicine.“ Noll, G.A., Furch, A.C.U., Rose, J. et al. Guardians of the phloem – forisomes and beyond. New Phytol 236, 1245–1260 (2022). https://doi.org/10.1111/nph.18476.

- Verbesserte Pflanzeneigenschaften; z. B. Metabolic Engineering

https://www.ime.fraunhofer.de/de/Forschungsbereiche/MB.html

Einordnung: Thema von Kapitel 4 Protein-Engineering wird explizit genannt; Metabolic Engineering entsprechend Kapitel 6 Bioproduktion, dazu passend auch Produktion von Proteinen; Enzymevolution passend zu Kapitel 5 Biokatalyse - Enzyme.

Bioeconomy Science Center (BioSC)

Kompetenzzentrum aus: RWTH Aachen, die Heinrich-Heine-Universität Düsseldorf (HHU), die Friedrich-Wilhelms-Universität Bonn und das Forschungszentrum Jülich

„Zentrale Themen im BioSC sind die nachhaltige Produktion von Pflanzen als Nahrungs- und Futtermittel und Biomasse, neue bio-/chemokatalytische und biotechnologische Methoden und Prozesse zur Umsetzung von nachwachsenden Rohstoffen in Wertstoffe wie z.B. Feinchemikalien, Proteine, Enzyme, Biopolymere, Biokraftstoffe. Ebenso werden Fragestellungen der wirtschaftlichen Umsetzbarkeit der Bioökonomie und ihrer gesellschaftlichen Akzeptanz bearbeitet. Eine enge Kooperation mit der Industrie soll eine wirtschaftliche Umsetzung der Forschungsergebnisse beschleunigen. Das BioSC umfasst aktuell ca. 1.900 Mitarbeiter aus 67 Mitgliedsinstituten aus den Partnereinrichtungen."

https://www.biosc.de/bioeconomy_science_center

Partner Biochemische Pflanzenphysiologie, HHU Düsseldorf

„In unserer Arbeitsgruppe untersuchen wir verschiedene Proteine und Proteinkomplexe aus höheren Pflanzen mit dem Ziel, die physiologische Bedeutung und Regulation dieser Proteine zu verstehen."

https://www.biosc.de/biochemische_pflanzenphysiologie_hhu_duesseldorf

Partner Biochemie I, HHU Düsseldorf

„Die Struktur, Dynamik und Funktion von ABC Transportern ist der Forschungsschwerpunkt des Institutes für Biochemie. Hierzu verwenden wir die unterschiedlichsten biochemischen, biophysikalischen und strukturbiologischen Ansätze, die unter anderem Proteinexpression und –aufreinigung, Bestimmung von Protein-Ligand und Protein-Proteinwechselwirkungen und Röntgenkristallographie beinhalten."

https://www.biosc.de/biochemie_1_hhu_duesseldorf

Partner Strukturbiochemie - Physikalische Biologie, HHU Düsseldorf

„Die Funktion jeder Zelle und jedes Organismus hängt entscheidend von den dynamischen Interaktionen zwischen biologischen Makromolekülen und ihrer korrekten drei-dimensionalen Struktur ab. Fehlerhafte Interaktionen und fehlgefaltete Strukturen führen letztlich zu Krankheit und Alterung. Unser Ziel ist es, diese Interaktionen zu verstehen und die dreidimensionalen Strukturen der an entscheidenden zellulären Prozessen beteiligten Protein-Komplexe möglichst in atomarer Auflösung zu bestimmen."

https://www.biosc.de/willbold

Einordnung: Punktuelle Bezüge zu Proteinen, Proteinkomplexen mit deren Struktur und Funktion. Bezug zur Kapitel 6 Bioproduktion aufgrund der Zielsetzung der gesamten Einrichtung.

Universität Bielefeld, Technische Fakultät

Centrum für Biotechnologie (CeBiTec)

„The Center for Biotechnology – CeBiTec – is one of the largest faculty-spanning central academic institutions at Bielefeld University. Its purpose is to bundle the biotechnological activities and research projects at the university, to foster cross-linking of research approaches and technologies from different research fields and to develop innovative projects within its two main research areas:

- Large Scale Genomics and Big Data Bioinformatics and
- Metabolic Engineering of Unicellular Systems and Bioproduction."

https://www.cebitec.uni-bielefeld.de

Einordnung: Im Forschungsfeld „Metabolic Engineering of Unicellular Systems and Bioproduction" geht es u. a. um die Produktion von Proteinen und besonders Enzymen sowie die Produktion rekombinanter therapeutischer und diagnostischer Proteine. Insofern bestehen Bezüge zu Kapitel 5 Biokatalyse - Enzyme; Kapitel 6 Bioproduktion sowie Kapitel 7 Gesundheit - Medikamentenentwicklung.

11.5. Gesundheit - Medikamentenentwicklung

11.5.1. Nachweis pathogener Proteine

Prof. Dr. Klaus Gerwert (Ruhr-Universität Bochum, betaSENSE GmbH), wurde in der Kategorie „innovation“ mit dem NRW-Innovationspreis ausgezeichnet. Es hat mit dem „Immuno-Infrarot-Sensor“ einen Test entwickelt, mit dem sich frühzeitig Indikatoren neurodegenerativer Erkrankungen wie zum Beispiel Alzheimer nachweisen lassen. Es wird erwartet, dass auf Basis dieser frühen Diagnose betroffene Personen frühzeitig mit passgenauen Therapien versorgt werden können.[169]

Inhaltliche Querbezüge mit Arbeiten bzw. Tools zu fehlgefalteten Proteinen und insbesondere dem DeepMind-Tool AlphaMissense[170] sind naheliegend.

11.5.2. Optogenetik - Lichtgesteuerte Produktion von Biopharmazeutika

Forschende der Universität Bonn entwickeln zusammen mit dem Biotech-Start-up Ningaloo Biosystems GmbH (Köln) neue Technologien zur Herstellung von Biopharmazeutika. Beide Kooperationspartner werden bis Ende 2025 im Rahmen des Programms „Zentralen Innovationsprogramms Mittelstand“ (ZIM) des Bundesministeriums für Wirtschaft und Klimaschutz mit insgesamt 440.000 Euro gefördert.[171] Biopharmazeutika (oder auch Biologika) können zur Behandlung verschiedener Arten von Krebs und Autoimmunerkrankungen verwendet werden. Diese Klasse von Medikamenten stellt derzeit das wachstumsstärkste Segment der pharmazeutischen Industrie dar. Allerdings sind die Produktionskosten zum Teil sehr hoch. Ziel des Projektes ist es, ein erstes voll funktionsfähiges Produktionssystem im Labormaßstab auf Grundlage der Optogenetik zu entwickeln. Durch die Optogenetik kann zelluläre Aktivität mit Licht kontrolliert werden. Dies soll nun auf die biotechnologische Herstellung von Biologika in Zellen angewendet werden und die Produktionskosten substanziell senken.

11.5.3. NRW-Akteure

Beispiele für NRW-Akteure - Gesundheit - Medikamentenentwicklung
Unternehmen
Bayer AG, Leverkusen Pharmaceuticals Division „The convergence of science and technology has ushered in a new era of innovation in the life sciences. Bayer is driving this paradigm-shift by leveraging leading expertise in biology, chemistry, data science and artificial intelligence.“ https://www.bayer.com/de/pharma/innovation-forschung „AI in Drug Discovery: Mapping the chemical space The discovery of new drugs is an immensely complex and time-consuming endeavor. Identifying compounds that bind to specific disease-related targets requires the screening of extremely large numbers of small molecules. AI is revolutionizing the way potential compounds are being identified. For instance, AI is powering Bayer’s search for new drug targets – starting with heart failure. Using cutting edge technologies such as knowledge graph for patient subtyping, in-silico target identification and target ranking we aim to bring much needed treatments to patients – faster and with higher precision.“ https://www.bayer.com/en/pharma/artificial-intelligence#1

[169] https://www.wirtschaft.nrw/innovationspreis-2023.
[170] Cheng, J., Novati, G., Pan, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381, 6664 (2023). https://doi.org/10.1126/science.adg7492.
[171] Universität Bonn. Mit Licht neue Medikamente herstellen. Pressemitteilung, 31. Aug. 2023. https://www.uni-bonn.de/de/neues/mit-licht-neue-medikamente-herstellen sowie Ningallo, https://ningaloo.bio/2023/03/08/third-party-funding/.

Einordnung: Zitat aus Pressemeldung:[172] “At Bayer, we are driving innovation and transformation. We are excited to be partners with OpenFold, where we can work with leading experts across industries to develop the best methodologies for pharmaceutical and agricultural product design. By computationally simplifying our discovery process, we can increase the speed and quality of product design”
UCB S.A. „This week UCB joined the OpenFold Artificial Intelligence (AI) Research Consortium. Joining this consortium reinforces our commitment to innovation in drug discovery and an aim to address the unmet medical needs of patients.“ UCB-Pressemeldung, 19. Sep. 2023 https://www.ucb.com/Our-Science/magazine/detail/article/Committed-to-drug-discovery-innovation **Einordnung:** OpenFold als Tool zur Proteinstrukturvorhersage; UCB als forschendes Pharmaunternehmen (laut dem Verband der forschenden Pharmaunternehmen, vfa).[173]
Ningaloo Biosystems GmbH, Köln „The use of naturally occurring as well as recombinant photoswitches opens up completely new options to influence processes inside living cells. This novel methodology, known as optogenetics, is based on the use of proteins found in nature that enable organisms to sense light. The optogenetic photoswitches can either be used directly or fused with heterologous effector domains to trigger other desired cellular responses in a light-dependent manner. Recently, the panel of optogenetic tools has been greatly expanded. When such photoswitches are genetically grafted into cells that normally lack light sensitivity, processes as diverse as the control of membrane potentials, cellular signaling pathways, metabolic pathways, protein translocation, and gene expression can be controlled optically […]. We are committed to developing solutions containing both recombinant photoswitch vectors as well as illumination devices and software that enable our customers to take full advantage of optogenetics – by introducing entirely new process control modes exploiting unprecedented possibilities to optimize and steer bioprocessing in real-time. We will offer optogenetic systems that, among others, work with the most common cell culture formats used in the production of biopharmaceutics, e.g. for immunotherapy or gene therapy.“ https://ningaloo.bio/research/ **Einordnung:** Protein-Design basierend auf den Fortschritten der Proteinstrukturvorhersage könnten dazu genutzt werden, um Optoproteine mit neuen Funktionen zu entwickeln und das Spektrum an Steuerungsmöglichkeiten durch Licht noch zu erweitern.[174]
Cevec, Köln (Teil des Unternehmens Cytiva, wiederum Teil des US-Unternehmens Danaher) „Global life sciences leader Cytiva has acquired CEVEC Pharmaceuticals, a leading German provider of high-performance cell line development and viral vector manufacturing technologies. CEVEC will further strengthen Cytiva's leading position in biomanufacturing solutions.“ (Pressemeldung, 6. Okt. 2022) https://www.cytivalifesciences.com/en/de/news-center/cytiva-strengthens-cell-line-development-with-cevec-acquisition-10001 **Einordnung:** Zell-basierte Biotherapeutika-Herstellung. Darstellung zur Zukunft der KI in der Entdeckung von Medikamenten.[175]
NEUWAY Pharma GmbH, Bonn „NEUWAY Pharma is developing an entirely novel class of biotherapeutics built around its proprietary Engineered Protein Capsules (EnPC®), which can cross the blood-brain barrier and deliver transformative neuropharmaceuticals for the treatment of disorders of the central nervous system (CNS).“ https://www.neuway-pharma.com/ **Einordnung:** Verwendet Protein-Engineering für ein Produkt im Bereich der Wirkstofffreisetzung. Daher auch Bezug zu Kapitel 3 Protein-Engineering.
Cilian AG, Münster

[172] Business Wire. OpenFold AI Research Consortium Welcomes 4 New Members: Bayer, Dassault, CHARM Therapeutics and BaseCamp Research. 22. Nov. 2022. https://www.businesswire.com/news/home/20221122005163/en/%C2%A0OpenFold-AI-Research-Consortium-Welcomes-4-New-Members-Bayer-Dassault-CHARM-Therapeutics-and-BaseCamp-Research
[173] vfa, Interaktive Standortkarte. https://www.vfa.de/de/verband-mitglieder.
[174] Persönliche Mitteilung.
[175] Cytiva. Revolutionizing drug discovery: The emergence of AI. 26. Jan. 2024. https://www.cytivalifesciences.com/en/us/news-center/ai-in-drug-discovery-10001.

„Die Cilian AG ist ein innovatives Biotechnologieunternehmen, das die einzigartigen Fähigkeiten von Ciliaten für die industrielle Produktion von therapeutischen Proteinen nutzt.“
„Experten für virale Impfstoffe, Verdauungsenzyme und monoklonale Antikörper für die Krebsimmuntherapie mit modernster Produktions- und Entwicklungstechnik.“
https://www.cilian.de/
Einordnung: Zellbasierte Produktion von therapeutischen Proteinen.

DISCO Pharmaceuticals GmbH, Köln
„At DISCO we are able to unlock the entire cancer cell surface at scale by identifying surface-bound proteins and uncovering protein communities. This approach thus permits unravelling comprehensive target information that neither standard proteomics, nor genome or transcriptome sequencing can provide.“
https://discopharma.de/#about
Einordnung: Plattform zur Identifizierung von Proteinen auf der Oberfläche bestimmter Krebszellen als Verfahren zur Identifizierung von „drug targets“.

Forschungseinrichtungen

Universität Bielefeld, Technische Fakultät, Centrum für Biotechnologie (CeBiTec), s. Abschnitt 11.4.2

Ruhr-Universität Bochum - Schwerpunkt Proteinforschung
mit folgenden Einrichtungen
Medizinisches Proteom-Center
„Das Medizinische Proteom-Center […] ist eines der führenden Institute für die Proteom- und Proteinanalytik. Mit Hilfe von gel- und chromatographie-basierten Trenntechniken und nachfolgender Massenspektrometrie, sowie mittels Proteinbiochips und innovativer Bioinformatik werden hier Fragestellungen aus allen Bereichen der Medizin, wie auch aus dem naturwissenschaftlichen Grundlagenbereich bearbeitet.“
https://www.mpc.ruhr-uni-bochum.de und
PRODI - Zentrum für Proteindiagnostik
„Methodenentwicklung zur (Früh-)Diagnostik und Therapieprädiktion
Um die Herausforderungen einer alternden Gesellschaft zu meistern, ist das vorrangige Ziel im Zentrum für Proteindiagnostik (PRODI) die Entwicklung neuer, minimalinvasiver Methoden zur Diagnostik und Therapieprädiktion. Im Fokus stehen dabei neurologische und onkologische Erkrankungen. Insbesondere werden innovative Methoden entwickelt, die bereits in frühen Krankheitsstadien präzise Informationen über Erkrankungen und individuelles Therapieansprechen liefern.“
http://www.prodi.rub.de und
PURE - Protein research Unit Ruhr within Europe
„Das Europäische Proteinforschungskonsortium PURE wurde 2010 an der Ruhr-Universität Bochum mit Unterstützung des Ministeriums für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen gegründet. International renommierte Proteinforscher der Ruhr-Universität Bochum und der Universität Duisburg-Essen und ausgewiesene klinische Forscher bündeln hier ihre Forschungsaktivitäten zur Früherkennung von Krankheiten wie Krebs, Parkinson, Multiple Sklerose oder Alzheimer.“
http://www.pure.ruhr-uni-bochum.de/wissenschaftler.html.de
sowie
Protein Research Department
„Proteine haben vielseitige Funktionen, welche notwendig für das zelluläre Überleben sind. Hierzu gehören z. B. verschiedene Stoffwechsel-, Transport-, und Signaltransduktionswege. Fehlerhafte Proteine, die ihre Aufgabe nicht mehr erfüllen können, verursachen schwerwiegende Erkrankungen. Um die Struktur, Funktionsweise und Netzwerke von Proteinen zu untersuchen, haben sich über 30 Wissenschaftler: Innen der Fakultäten Biologie und Biotechnologie, Chemie und Biochemie, Physik und Astronomie sowie Medizin im Rahmen des Protein Research Departments (PRD) zusammengeschlossen. Die Forschung im PRD befasst sich z. B. mit Fragen, wie einzelne Proteine das Überleben der Zelle beeinflussen, wie Sie sich die Struktur und Funktion von Proteinen entschlüsseln lässt oder wie sich Proteininteraktionen nutzen lassen, um neue Medikamente gegen Erkrankungen zu entwickeln.“
http://www.rd.ruhr-uni-bochum.de/protein/
Einordnung: Organisatorische Bezüge und Vernetzung der verschiedenen Einrichtungen nicht ohne weiteres ersichtlich. Insbesondere Aussage zum PRD zur Entschlüsselung von Struktur und Funktion von Proteinen

einschlägig. Zumindest punktuell Nutzung von AlphaFold.[176] Bezug zu Kap. 7 sowie Abschnitt 2.2. Ebenfalls Bezug erkennbar[177] zu Abschnitt 2.4 Neue KI-Anwendungen in den Lebenswissenschaften
Ruhr-Universität Bochum Biochemie II Biomolekulare NMR-Spektroskopie „Biochemistry seeks to understand life at a molecular level by examining the relationship between the structure and function of biomolecules, such as proteins, nucleic acids, and lipids. To achieve this, we use a variety of techniques, particularly biomolecular NMR spectroscopy, in order to determine the three-dimensional structures and dynamics of biomolecules as well as how they interact with each other and other molecules in solution at near-physiological conditions. […] Our research focuses in the main on medically-relevant proteins, specifically those involved in the development of cancerous tumours, to be able to better understand the causes of the condition and propose more effective treatment strategies.“ https://www.ruhr-uni-bochum.de/bionmr/index.html **Einordnung:** Verbindung von Funktion und Struktur von Proteinen wird explizit angesprochen. Fokus auf medizinisch relevante Proteine mit Bezug zu Krebserkrankungen. Nutzung von AlphaFold.[178]
Universität Bonn, Medizinische Fakultät Institut für Biochemie und Molekularbiologie „Die Arbeitsgruppen des Institutes befassen sich mit der Funktion und Biogenese von Mitochondrien und Lysosomen sowie mit Selenoproteinen. Das breite wissenschaftliche Spektrum der bearbeiteten Projekte der verschiedenen Arbeitsgruppen reicht von der Grundlagenforschung bis hin zum Verständnis der Pathomechanismen von Erkrankungen.“ https://www.ibmb.uni-bonn.de/de/forschung **Einordnung:** Nutzung von AlphaFold.**Fehler! Textmarke nicht definiert.** Nennung „Pathomechanismen von E rkrankungen“ - insofern Bezug zu Kapitel 7 Gesundheit - Medikamentenentwicklung.
Universitätsklinikum Bonn Institut für Klinische Chemie und Klinische Pharmakologie Arbeitsgruppe Schmid-Burgk „Angeborene Immunantworten können lebensrettend oder lebensbedrohlich sein. Unsere Gruppe untersucht Immunsignalwege auf molekularer und zellulärer Ebene, für die wir neue Technologien wie CRISPaint und In-Cellulo-Sequenzierung entwickeln. Darüber hinaus arbeiten wir an Algorithmen zur Suche nach unbekannten Diversitäten in Sequenzierungsdaten, um selbstdiversifizierende Systeme zu charakterisieren.“ https://www.ukbonn.de/ikckp/forschung/ **Einordung:** Nutzung von AlphaFold im medizinischen Kontext[179] - Bezug zu Kapitel 7.
Heinrich-Heine-Universität Düsseldorf, Mathematisch-Naturwissenschaftliche Fakultät Institut für Pharmazeutische und Medizinische Chemie https://www.pharmazie.hhu.de/institut-fuer-pharmazeutische-und-medizinische-chemie AG Gohlke Computational Pharmaceutical Chemistry and Molecular Bioinformatics „Our current research focusses on four main areas, for which we develop and apply techniques grounded in computational biophysical chemistry, molecular bioinformatics, and computational pharmaceutical chemistry. I. Structure, Function, and Molecular Recognition Investigated by Molecular Dynamics Simulations II. Rigidity Theory for Biomolecules III. Structure Prediciton by Comparative Modeling and Using (Co-)Evolutionary Information IV. Modulating Biomolecular Interactions“

[176] „We matched these cross-links either to homology models based on available PDB structures […] or to the AlphaFold Protein Structure Database“ vgl. Singh, J., Elhabashy, H., Muthukottiappan, P. et al. Cross-linking of the endolysosomal system reveals potential flotillin structures and cargo. Nat Commun 13, 6212 (2022). https://doi.org/10.1038/s41467-022-33951-0.
[177] Vgl. Künstliche Intelligenz in der Proteindiagnostik: http://www.prodi.rub.de/bioinformatik/.
[178] „However, CSP mapping to the surface of the predicted Otoraplin structure by DeepMind's AlphaFold identified the interaction site to be at the disulfide and distal loop[21]“. Vgl. Arnolds, O. & Stoll, R. Characterization of a fold in TANGO1 evolved from SH3 domains for the export of bulky cargos. Nat Commun 14, 2273 (2023). https://doi.org/10.1038/s41467-023-37705-4.
[179] Rouillon, C., Schneberger, N., Chi, H. et al. Antiviral signalling by a cyclic nucleotide activated CRISPR protease. Nature 614, 168–174 (2023). https://doi.org/10.1038/s41586-022-05571-7 (zus. mit Universität Bonn, Institut für Strukturbiologie; Max-Planck-Institute for Neurobiology of Behavior—caesar, Bonn, Germany).

https://cpclab.uni-duesseldorf.de/index.php/Overview
Einordnung: Stichworte „computational pharmaceutical chemistry“, „structure prediction“, Nutzung von ColabFold.[180]

[180] Marasco, R., Fusi, M., Coscolín, C. et al. Enzyme adaptation to habitat thermal legacy shapes the thermal plasticity of marine microbiomes. Nat Commun 14, 1045 (2023). https://doi.org/10.1038/s41467-023-36610-0.

11.6. Pflanzengesundheit - NRW-Akteure

Beispiele für NRW-Akteure – Pflanzengesundheit
Unternehmen
Bayer AG Crop Science Divison „In Monheim am Rhein befindet sich der Sitz der Bayer AG, Crop Science Division. Die Crop Science Division erforscht und entwickelt hier neue Pflanzenschutzmittel und steuert von hier aus das weltweite Marketing seiner Produkte.“ https://www.bayer.com/de/de/monheim-startseite **Einordnung:** Nutzung von AlphaFold in wissenschaftlichen Publikationen (Ralf Nauen).[181,182] Bezug zu Pflanzengesundheit entsprechend der Selbstbeschreibung. Passend zu Kapitel 9 Pflanzengesundheit. In einer Nachricht zu Openfold stellt eine Person aus der Bayer Crop Science Division Bezug zu Anwendungen im Bereich der Landwirtschaft her.[183]
Forschungseinrichtungen
Cluster of Excellence on Plant Sciences (CEPLAS) „Der Cluster of Excellence on Plant Sciences (CEPLAS) ist eine gemeinsame Initiative der Heinrich-Heine-Universität Düsseldorf (HHU), der Universität zu Köln (UzK), des Max-Planck-Instituts für Pflanzenzüchtungsforschung Köln (MPIPZ) und des Forschungszentrums Jülich (FZJ). Im Cluster CEPLAS entwickeln Wissenschaftler/innen gemeinsam innovative Strategien für eine nachhaltige Pflanzenproduktion.“ https://www.ceplas.eu/de/startseite **Einordung:** exemplarisch - Nutzung von AlphaFold im Kontext der Untersuchung von Klima-angepassten Pflanzen.[184]
Universität Bielefeld Fakultät für Chemie, Arbeitsbereich Physikalische Chemie, AG Atmospheric & Physical Chemistry, Prof. Dr. Thomas Koop „In the biosphere, many organisms have to cope with subzero temperatures, either temporary or permanently. Therefore, fishes, insects and plants have developed strategies to protect their living cell membranes from the destructive growth of ice crystals by producing antifreeze proteins (AFP).“ https://www.uni-bielefeld.de/fakultaeten/chemie/ag/pc2-koop/research/ Abschnitt „Ice nucleation and growth“. **Einordnung:** Nutzung von Proteinstrukturvorhersagen im Zusammenhang mit Frostschäden an Pflanzen[185] - Kapitel 9 Pflanzengesundheit.

Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln „Lässt sich Pflanzenzüchtung in eine rationale, prädiktive Wissenschaft verwandeln?“ Diese Frage motiviert alle Forschungsprogramme am Institut. Wir möchten herausfinden, wie ein detailliertes Verständnis der molekularen

[181] Duarte, A., Pym, A., Garrood, W.T. et al. (2022). P450 gene duplication and divergence led to the evolution of dual novel functions and insecticide cross-resistance in the brown planthopper Nilaparvata lugens. PLoS Genet 18, 6, e1010279. https://doi.org/10.1371/journal.pgen.1010279.

[182] Haas, J., Beck, E., Troczka, B.J. et al. A conserved hymenopteran-specific family of cytochrome P450s protects bee pollinators from toxic nectar alkaloids. Sci Adv 9, eadg0885 (2023). https://doi.org/10.1126/sciadv.adg0885

[183] Business Wire. OpenFold Biotech AI Research Consortium releases SoloSeq and Multimer, an integrated protein Large Language Model with 3D structure generation. 19. Feb. 2024. https://www.businesswire.com/news/home/20221122005163/en/%C2%A0OpenFold-AI-Research-Consortium-Welcomes-4-New-Members-Bayer-Dassault-CHARM-Therapeutics-and-BaseCamp-Research.

[184] „In light of climate change and human population growth one of the most challenging tasks is to generate plants that are Fe-efficient, resilient to low Fe supply and Fe-biofortified. For such endeavors, it is crucial to understand the regulation of Fe acquisition and allocation in plants.” Vgl. Lichtblau, D.M., Schwarz, B., Baby, D. et al. The Iron Deficiency-Regulated Small Protein Effector FEP3/IRON MAN1 Modulates Interaction of BRUTUS-LIKE1 With bHLH Subgroup IVc and POPEYE Transcription Factors. Front Plant Sci 13, 930049 (2022). https://doi.org/10.3389/fpls.2022.930049.

[185] „Bacterial ice nucleation proteins (INPs) can cause frost damage to plants by nucleating ice formation at high sub-zero temperatures.“ Vgl. Forbes, J., Bissoyi, A., Eickhoff, L. et al. Water-organizing motif continuity is critical for potent ice nucleation protein activity. Nat Commun 13, 5019 (2022). https://doi.org/10.1038/s41467-022-32469-9.

Mechanismen, die in Modellpflanzenarten definiert sind, verwendet werden kann, um ausgewählte Merkmale in Kulturpflanzen vorherzusagen.“ https://www.mpipz.mpg.de/ **Einordnung:** Nutzung von AlphaFold[186] mit Bezug zu Pflanzen insofern passend zu Kapitel 9 Pflanzengesundheit.
Universität Münster Institut für Biologie und Biotechnologie der Pflanzen - IBBP AG Functional Biopolymers and Molecular Phytopathology „Dabei interessieren wir uns weniger für die reinen Materialeigenschaften als vielmehr für die biologischen Aktivitäten der Biopolymere, insbesondere von komplexen Zuckern […]. Wir wollen verstehen, wie die biologischen Funktionen der komplexen Zucker von ihren biochemischen Strukturen und ihren physikochemischen Eigenschaften bestimmt werden. Dazu identifizieren, charakterisieren und optimieren wir Polysaccharid-modifizierende Enzyme […], also die biologischen Werkzeuge derjenigen Organismen, die diese Polysaccharide produzieren oder abbauen. Wir entwickeln Methoden, um die mit solchen Enzymen modifizierten Polysaccharide oder aus ihnen enzymatisch hergestellten Oligosaccharide bio- und physikochemisch zu analysieren und ihre biologischen Aktivitäten zu erkennen, zu messen und zu verbessern. Letztlich können wir so Produkte herstellen, die für bestimmte Aufgaben optimiert sind, z.B. für einen biologischen Pflanzenschutz oder für neuartige Wundverbände, unter denen auch chronische Wunden zuverlässig heilen können.“ https://www.uni-muenster.de/Biologie.IBBP/agmoerschbacher/forschung/index.html **Einordnung:** Nutzung von AlphaFold.[187,188] Bezug zu Pflanzenschutz, Enzymen. Passend zu Kapitel 9 sowie Kapitel 5 Biokatalyse - Enzyme.

[186] Gao, H., Song, W., Severing, E. et al. PIF4 enhances DNA binding of CDF2 to co-regulate target gene expression and promote Arabidopsis hypocotyl cell elongation. Nat Plants 8, 1082–1093 (2022). https://doi.org/10.1038/s41477-022-01213-y.
[187] Cord-Landwehr, S. & Moerschbacher, B.M. Deciphering the ChitoCode: fungal chitins and chitosans as functional biopolymers. Fungal Biol Biotechnol 8, 19 (2021). https://doi.org/10.1186/s40694-021-00127-2.
[188] Weyer, R., Hellmann, M.J., Hamer-Timmermann, S.N. et al. Customized chitooligosaccharide production—controlling their length via engineering of rhizobial chitin synthases and the choice of expression system. Frontiers in Bioengineering and Biotechnology 10 (2022). https://doi.org/10.3389/fbioe.2022.1073447.

12. SWOT-Analyse NRW und Fazit

12.1. SWOT-Analyse NRW

In diesem Kapitel werden die bisher dargestellten Ergebnisse zu einer SWOT-Analyse[189] für das Land NRW im Feld der Proteinstrukturvorhersage (PSV) verdichtet.

12.1.1. Chancen

Die gesamte Trendstudie ist motiviert als Hinweis auf die vielfältigen wirtschaftlichen Chancen, die sich aus dem wissenschaftlichen Durchbruch bei der Proteinstrukturvorhersage ergeben.

Wie in Kapitel 1 ausgeführt, sind Proteine Biomoleküle mit mannigfachen Funktionen. Als Strukturproteine bestimmen sie den Zellaufbau und die Zellstruktur und damit letztlich auch für die äußere Form von Pflanzen, Tieren und Menschen. Auch sind die meisten Enzyme Proteine. Diese bewerkstelligen und beschleunigen den Ablauf von biochemischen Reaktionen und sind damit bspw. wesentlich für die Versorgung eines Organismus mit Energie. Andere Proteine spielen eine wichtige Rolle für Muskeln und Bewegung, die sog. Motorproteine. Darüber hinaus haben Proteine wichtige Funktionen in der Immunabwehr, etwa in der Form von Antikörpern, die Krankheitserreger erkennen und helfen, sie zu bekämpfen. Proteine können auch Hormone sein und als Signal- oder Botenstoffe dienen, um Anpassungsreaktionen des Körpers und Stoffwechselprozesse auszulösen. Daneben gibt es Proteine, die andere Stoffe im Körper transportieren, sog. Transportproteine, wie zum Beispiel das Hämoglobin, das für den Sauerstofftransport im Körper verantwortlich ist.

Angesichts dieser vielfältigen und zentralen Funktionen von Proteinen in Lebensprozessen aller Art, haben Proteine und das Wissen über Proteine - zumindest indirekt - auch eine große, wirtschaftliche Bedeutung. Dies hängt damit zusammen, dass einerseits bereits heute viele Wirtschaftszweige biobasierte Ausgangsstoffe nutzen bzw. biologische Grundlagen haben, wie natürlich die Landwirtschaft und Ernährungsindustrie, aber auch die Forstwirtschaft mit der Holz- und Papierindustrie, die Leder- und Textilindustrie, die Umweltwirtschaft und die Gesundheitswirtschaft. In all diesen Wirtschaftszweigen kann ein besseres Verständnis von Lebensprozessen, das auch von einem genaueren Wissen über Proteine abhängt, ein wesentlicher Faktor für Effizienzsteigerungen und Innovationen sein. Hinzu kommt andererseits die herausragende Rolle, die biobasierte Prozesse in einer zukünftigen industriellen Bioökonomie spielen können. Durch die industrielle Bioökonomie wird die wirtschaftliche Bedeutung von Proteinen und dem Wissen über Proteine auf die chemische Industrie, die Energiewirtschaft und die Grundstoffindustrie ausgeweitet und verstärkt.

Dabei ist zu berücksichtigen, dass die räumliche Struktur der Proteine in Kombination mit den chemischen Eigenschaften an ihrer Oberfläche für die biologische Funktion entscheidend ist. Man spricht in diesem Zusammenhang auch vom Schlüssel-Schloss-Prinzip. Dabei müssen meist zwei Komponenten räumlich zueinander passen (wie ein Schlüssel räumlich zum Schloss), damit eine bestimmte biochemische Reaktion ablaufen kann. Daraus erklären sich die großen Chancen, die aus den aktuellen, immensen Fortschritten im Bereich der Proteinstrukturvorhersage zunächst in wissenschaftlicher (vgl. Kapitel 2), aber langfristig auch in wirtschaftlicher Hinsicht erwachsen, wie in den Kapiteln 3 bis 10 näher ausgeführt wird. Sämtliche absehbare Chancen sind dabei im allgemeinen auch Chancen für NRW.

Um die Stärken und Schwächen in NRW im Vergleich zur globalen Situation herauszuarbeiten, werden in der folgenden Tabelle 12.1 (S. 54) verschiedene Informationen und Rechercheergebnisse kompakt gegenübergestellt. Die Tabelle enthält in der linken Spalte die übergeordneten Anwendungspotenziale der Proteinstrukturvorhersage, die in den entsprechend bezeichneten Kapiteln 2 bis 10 oben näher beschrieben wurden. Kapitel 11 „NRW-Perspektive“ stellt die identifizierten Trends aus NRW-Sicht dar mit NRW-Anwendungsbeispielen und NRW-Akteuren. Die Spalte ganz rechts enthält in komprimierter Form eine Zusammenfassung der NRW-Akteure mit indirektem Bezug zur Proteinstrukturvorhersage (PSV) als einfache Zählung nach Zugehörigkeit zu einer der drei Gruppen Großunternehmen (GU), kleine und mittelständische Unternehmen (KMU) bzw. Start-ups sowie Forschungseinrichtungen.[190] Indirekter Bezug bedeutet dabei, dass zwar durch die Akteure selbst keine Proteinstrukturvorhersagen erstellt werden, die Akteure aber von der steigenden Qualität der Vorhersagen in

[189] Engl. Abkürzung für *Strengths* (Stärken), *Weaknesses* (Schwächen), *Opportunities* (Chancen) und *Threats* (Risiken).
[190] Auch bei den Forschungseinrichtungen gibt es Größenunterschiede von einzelnen, kleinen Arbeitsgruppen bis zu großen Institutsverbünden mit teilweise mehr als 1.000 Forschenden.

unterschiedlicher Form profitieren können. In der mittleren Spalte der Tabelle („Weltweite Perspektive - Start-ups mit direktem PSV-Bezug“) ist die Anzahl der Start-ups aufgeführt, deren Geschäftsmodell direkt auf Dienstleistungen im Zusammenhang mit der PSV beruht. Diese Start-ups wurden anhand einer Crunchbase-Recherche weltweit identifiziert. Aus Gründen der Übersichtlichkeit werden die Namen der betreffenden Start-ups in Fußnoten angegeben. In der Spalte „NRW-Perspektive Direkter PSV-Bezug“ finden sich Akteure aus NRW mit direktem Bezug zur PSV. Die leere Spalte „Weltweit - Indirekter PSV-Bezug (nicht erhoben)“ ist enthalten, um augenscheinlich zu machen, dass Informationen zu weltweiten Akteuren mit indirektem PSV-Bezug aus Ressourcengründen nicht erhoben werden konnten.[191]

Anwendungs-potenziale	Weltweit - Indirekter PSV-Bezug (nicht erhoben)	Weltweite Perspektive Start-ups mit direktem PSV-Bezug	NRW-Perspektive Direkter PSV-Bezug	NRW-Perspektive Indirekter PSV-Bezug
2. Lebens-wissenschaften		Speziell 2.4 KI in den Lebenswissenschaften: --> 2 Start-ups[192]	Wissenschaftliche Arbeiten	(1 GU) 2 KMU/Start-ups 18 Forschungseinrichtungen
3. Protein-Engineering u. 4. Proteindesign		--> 4 Start-ups[193] (für diverse Märkte)	-	2 KMU/Start-ups 4 Forschungseinrichtungen
5. Biokatalyse – Enzyme		--> 1 Start-up[194]	-	2 GU 3 KMU/Start-ups 5 Forschungseinrichtungen
6. Bioproduktion		-	-	3 KMU/Start-ups 4 Forschungseinrichtungen
7. Gesundheit - Medikamenten-entwicklung		--> 13 Start-ups[195] In Verbindung mit Protein-Engineering/Design --> 6 Start-ups[196]	Openfold - Partner Bayer und UCB (mit FuE in NRW)	2 GU 5 KMU/Start-ups 6 Forschungseinrichtungen
8. Tiergesundheit		-	.	-
9. Pflanzengesundheit		1 Start-up[197]	Openfold - Partner Bayer	1 GU 3 Forschungseinrichtungen

Tabelle 12.1: Übersichtsdarstellung von Akteuren mit direktem und indirektem Bezug zur Proteinstrukturvorhersage.

[191] Methodische Einordnung: Bei jedem geprüften Akteur erfolgte eine Einschätzung zum fachlichen Bezug zur Proteinstrukturvorhersage (in Kap. 11 jeweils im Abschnitt „Einordnung“). Die erreichbare bzw. erreichte Vollständigkeit der Identifikation von Akteuren und die erreichbare bzw. erreichte Genauigkeit der Zuordnung (auch zu Anwendungsfeldern) sind Funktionen der verfügbaren Ressourcen und damit begrenzt. Außerdem ist zu bedenken, dass wie in der Vorgehensweise dargelegt, bei Unternehmen aus NRW auch Akteure mit indirektem Bezug zur PSV vorgestellt werden, um möglichst viele mögliche Akteure für den angedachten Zukunftsdiskurs zu erfassen. Dies führt dazu, dass die Zahlen der identifizierten Unternehmen nicht direkt vergleichbar sind, konkret weil in NRW bspw. auch Protein-Produzenten enthalten sind, nach denen aus Ressourcengründen nicht auch weltweit gesucht wurde. Insofern sollten die Inhalte der Tabelle entsprechend zurückhaltend interpretiert werden.

[192] Zilixir Biotech, www.zelixir.com; Trident Bioscience, www.trident.bio.

[193] Arzeda, www.arzeda.com; Profluent Bio, www.profluent.bio; Monod Bio, www.monod.bio; Levinthal, www.levinthal.bio.

[194] Medium Biosciences, www.medium.bio.

[195] Isomorphic Labs, www.isomorphiclabs.com; Relay Therapy, www.relaytx.com; Seismic Therapeutic, www.seismictx.com; Synthekine, www.synthekine.com; Atomic AI, www.atomic.ai; LabGenius, www.labgeni.us;
Outpace Bio, www.outpacebio.com; Black Diamond Therapeutics, www.blackdiamondtherapeutics.com; Interline Therapeutics, www.interlinetx.com; Antiverse, www.antiverse.io; Kactus Biosystems, www.kactusbio.com; PharmCADD, www.pharmcadd.com; DeepSeq.AI, www.deepseq.ai.

[196] Evozyne, www.evozyne.com; GRO Biosciences, www.grobio.com; Neoleukin Therapeutics, www.neoleukin.com; Diffuse Bio, www.diffuse.bio; MoleculeMind, www.moleculemind.com; Generate:Biomedicines, www.generatebiomedicines.com.

[197] Protabit (u. a. für Monsanto), www.protabit.com.

12.1.2. Risiken

Der Erfolg des Unternehmens DeepMind mit AlphaFold ist auch ein Erfolg von Investitionen des Großunternehmens Google/Alphabet aus der Digitalbranche in die Pharma- und Biotechnologie-Branche. In der öffentlichen und wissenschaftlichen Wahrnehmung steht aktuell der Nutzen von AlphaFold als Open-Source-Software und der Datenbank der AlphaFold-Proteinstrukturvorhersagen als Open-Data im Vordergrund. Gleichwohl ist mit der Investition zweifellos auch das Ziel verbunden, für Google/Alphabet ein wirtschaftliches Standbein außerhalb der Digitalwirtschaft im engeren Sinne und ohne Abhängigkeit von Werbung als Einnahmequelle aufzubauen, was neben Google/Alphabet auch für andere große digitale Plattformen von Interesse ist. Der DeepMind-Ausgründung Isomorphic Labs ist es - wie bereits erwähnt – gelungen, zwei Großunternehmen (Eli Lilly and Company und Novartis) als Kunden zu gewinnen, die im Erfolgsfall Umsätze in Milliardenhöhe zuzüglich Lizenzgebühren für mögliche, zukünftige Medikamente generieren könnten. Dieses erste Beispiel verdeutlicht, dass trotz des Zugangs zu einer neuen Branche ein typisches Geschäftsmodell der Digitalbranche beibehalten wird, bei dem Lizenzeinnahmen im Vordergrund stehen und Produkthaftungsrisiken begrenzt werden. Damit ist für die etablierten Akteure aus dem Pharmasektor das konkrete Risiko verbunden, dass sich die Gewinnmargen reduzieren könnten.

Mit der Publikation von AlphaFold 3 (am 8. Mai 2024)[198] wird ein weiteres Muster aus der Digitalbranche sichtbar, dass darin besteht, ein digitales Ökosystem aufzubauen, das in diesem Fall überspitzt als „Betriebssystem für Proteine“ bezeichnet werden könnte. In der betreffenden Publikation heißt es: „AlphaFold 3 will be available as a non-commercial usage only server at https://www.alphafoldserver.com, with restrictions on allowed ligands and covalent modifications.“ Damit wird eine kostenlose Entwicklungsplattform bereitgestellt, die Forschende und Entwickelnde an AlphaFold bindet und zugleich DeepMind ermöglicht, aus deren Nutzungsverhalten zu lernen. Über wissenschaftliche Nachwuchskräfte, die nach der Nutzung von AlphaFold aus dem akademischen Bereich in die Industrie wechseln, kann zudem langfristig Einfluss auf die industrielle FuE gewonnen werden, wodurch sich für die etablierten Akteure der Pharmabranche das Risiko ergibt, an Einfluss sowie Teile der Wertschöpfung zu verlieren.

Ein weiterer Erfolgsfaktor für AlphaFold ist neben der finanziellen Investition die Technologieführerschaft von DeepMind im Bereich KI bzw. maschinelles Lernen, die sich bereits in früheren Erfolgen von DeepMind wie AlphaGo, AlphaGoZero und AlphaZero gezeigt hat.[199] In diesem Zusammenhang ist zu berücksichtigen, dass die Technologieführerschaft im Bereich der Sprachmodelle und generativen KI ebenfalls bei US-amerikanischen Großunternehmen gesehen wird, die wie Google/Alphabet über ausreichende Ressourcen verfügen, um mit dieser Technologieführerschaft und geeigneten Investitionen eine Hebelwirkung im Bereich der Gesundheitswirtschaft bzw. der Pharmaindustrie zu erzielen. So hat sich zumindest zeitweilig auch das Unternehmen Meta in dem Feld der Proteinstrukturvorhersage engagiert.[200] Zudem ist bekannt, dass einige dieser großen, amerikanischen Plattformen umfangreiche Sammlungen von Gesundheitsdaten anlegen, so dass hier jederzeit Überraschungen möglich sind. Des Weiteren zeigt die Arbeit von DeepMind am Tool AlphaMissense, dass auch Marktpotenziale im Zusammenhang mit der Fehlfaltung von Proteinen gesehen werden, an denen in NRW ebenfalls intensiv und erfolgreich gearbeitet wird. Diese exemplarischen Beobachtungen verdeutlichen das Risiko für die etablierten Akteure der Pharmazie und Biotechnologie, dass neue Wettbewerber aus der Digitalbranche erwachsen, deren Wettbewerbsfähigkeit möglicherweise übersehen oder falsch eingeschätzt wird.

12.1.3. Schwächen NRW

Im Gegensatz dazu ist zu beobachten, dass bezogen auf NRW praktisch keine Bezüge in Richtung Proteinstrukturvorhersage bei Unternehmen der Digitalwirtschaft oder bereits länger aktiven KI-Unternehmen aus NRW identifiziert werden konnten.[201] Dies gilt auch im Hinblick auf Forschungseinrichtungen aus dem Bereich der allgemeinen KI-Forschung, während es einige KI-Bezüge bei dezidierten Forschungseinrichtungen der Bioinformatik gibt. Darin ist eine Schwäche in NRW zu sehen. Aufgrund der immensen Bedeutung, die KI und maschinellem Lernen im gesamten Themenfeld der Lebenswissenschaften beigemessen wird, kann gefolgert

[198] Abramson, J., Adler, J., Dunger, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w.
[199] Dabei ging es um das stufenweise schnellere Erlernen von Brettspielen wie etwa Go und Schach, wobei es gelang das Leistungsvermögen der besten menschlichen Spieler zu übertreffen; vgl. etwa den Wikipedia-Eintrag zu „DeepMind“; https://de.wikipedia.org/wiki/DeepMind#AlphaGo.
[200] Lin, Z., Akin, H., Rao, R. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379, 1123–1130 (2023). https://doi.org/10.1126/science.ade2574.
[201] Zudem gilt für NRW - wie auch für die meisten Weltregionen - dass es an schlagkräftigen Plattform-Unternehmen fehlt.

werden, dass eine Frage für den Zukunftsdialog darin bestehen sollte, ob und wenn ja, wie bestehende Stärken von NRW in den Lebenswissenschaften mit den ebenfalls erkennbaren Stärken von NRW im Bereich KI noch besser verzahnt werden können.

Aus Tabelle 12.1 wird zudem unmittelbar eine Schwäche in NRW deutlich: Es gibt in NRW (bislang) keine Start-ups im Bereich der Proteinstrukturvorhersage. Zur Einordnung dieser offensichtlichen Schwäche wäre aber auch die Frage zu bewerten, inwieweit sich die identifizierten, globalen Start-ups gegenüber dem führenden Unternehmen Isomorphic Labs (als DeepMind-Ausgründung) als wettbewerbsfähig erweisen werden.

12.1.4. Stärken NRW

Tabelle 12.1 zeigt auch zwei Stärken in NRW und zwar erstens im Bereich „Gesundheit - Medikamentenentwicklung“ und zweitens im Bereich „Biokatalyse – Enzyme“. Diese beiden Bereiche zeichnen sich dadurch aus, dass es in NRW sowohl Großunternehmen (GU) als auch KMU/Start-ups sowie (größere) Forschungseinrichtungen gibt. In dieser ausgewogenen Konstellation kann ein Standortvorteil gesehen werden. Zudem ist es von Vorteil, dass es Großunternehmen im Bereich der pharmazeutischen und der chemischen Industrie gibt mit teils inhaltlich klar erkennbaren Bezügen zur Proteinstrukturvorhersage. Solche Unternehmen könnten prinzipiell als regional relevante Netzwerkknoten, Kunden, Kooperationspartner oder Investoren in Betracht kommen, selbst wenn ihr Blick zumeist weltweit ausgerichtet ist.

12.2. Fazit

Zusammenfassend lassen sich die folgenden Beobachtungen festhalten:

- Aufgrund der immensen Fortschritte im Feld der Proteinstrukturvorhersage ergeben sich vielfältige, wirtschaftliche Chancen und Anwendungspotenziale auch für Akteure in NRW. Diese bestehen selbst dann, wenn in NRW keine führende Position in der Proteinstrukturvorhersage im engeren Sinne erreicht werden sollte.
- Für etablierte Akteure in der Pharma- und Biotechnologiebranche ergeben sich Marktrisiken aufgrund neuer Wettbewerber mit Hintergrund in der Digitalbranche.
- Es ist zu diskutieren, ob und ggf. wie die bestehenden Stärken in NRW in den Lebenswissenschaften mit den ebenfalls erkennbaren Stärken in NRW im Bereich KI noch besser verzahnt werden können.
- In den Bereichen „Gesundheit - Medikamentenentwicklung“ und „Biokatalyse – Enzyme“ erscheint NRW vergleichsweise gut aufgestellt, um von den Fortschritten bei der Proteinstrukturvorhersage zu profitieren.

AlphaFold und verwandte Modelle der Proteinstrukturvorhersage stehen allerdings nicht für sich allein und werden nicht allein „die Welt verändern“. Sie gehören jedoch zu einer Reihe lebenswissenschaftlicher Durchbrüche wie Gensequenzierung, omics-Methoden und Gen-Editierung. Im Zusammenspiel dieser wissenschaftlichen Durchbrüche, erscheinen manche langgehegten Visionen inzwischen deutlich realistischer - auch für eine Umsetzung in NRW.

Längerfristig ist zu erwarten, dass der Durchbruch bei der Fragestellung der Proteinfaltung die Nutzung von KI und maschinellem Lernen für viele andere biologische, biotechnologische und medizinische Fragestellungen inspirieren und dadurch zahlreiche, weitere Anwendungspotenziale eröffnen wird.

Impressum

Herausgeber
Ministerium für Wirtschaft,
Industrie, Klimaschutz und Energie
des Landes Nordrhein-Westfalen
Tel.: + 49 (0) 211/61772-0
Fax: + 49 (0) 211/61772-777

Internet: www.wirtschaft.nrw
E-Mail: poststelle@mwike.nrw.de

Abteilung 3
Wirtschaftspolitik

Diese Studie wurde von der VDI Technologiezentrum GmbH im Auftrag des Ministeriums für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen erstellt.

Bildnachweise
© MWIKE NRW/Csaba Mester - Foto Berger Allee
Bildquelle: © iStock.com/bjdlzx

Redaktion/Redaktionsteam
Dr. Dirk Holtmannspötter, VDI Technologiezentrum GmbH

Die Publikation ist auf der Homepage des Ministeriums für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen unter www.wirtschaft.nrw/broschuerenservice als PDF-Dokument abrufbar.

Hinweis

Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit der Landesregierung Nordrhein-Westfalen herausgegeben. Sie darf weder von Parteien noch von Wahlbewerberinnen und -bewerbern oder Wahlhelferinnen und -helfern während eines Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden.

Dies gilt auch für Landtags-, Bundestags- und Kommunalwahlen sowie für die Wahl der Mitglieder des Europäischen Parlaments.

Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung.

Eine Verwendung dieser Druckschrift durch Parteien oder sie unterstützende Organisationen ausschließlich zur Unterrichtung ihrer eigenen Mitglieder bleibt hiervon unberührt. Unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Schrift der Empfängerin oder dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Landesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte.

© Dezember 2024 / MWIKE24-040

Ministerium für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen Berger Allee 25, 40213 Düsseldorf
www.wirtschaft.nrw