

title: "Browser-clientseitige Werkzeuge für datenschutzkonforme Sequenz-Analyse in der Lehre"

subtitle: "Architektur, rechtliche Einordnung und Validierung"

slug: "codex-vivum"

author: "Thomas Schüller"

orcid: "0009-0003-9799-6747"

email: "forschung@tschueller.com"

affiliation: "unabhängige Forschung, Wien (ohne institutionelle Affiliation, privat finanziert)"

date: "2026-05-12"

version: "v1.0"

language: "de"

license: "CC BY 4.0 (Text) · AGPL-3.0-or-later (Quellcode)"

keywords:

- Bioinformatik-Lehre
- clientseitige Werkzeuge
- local-first
- PCR-Primer-Design
- Schmelztemperatur
- Restriktionsenzyme
- DSGVO
- Open Science

abstract: |

Werkzeuge der Molekularbiologie und Bioinformatik sind in Lehrkontexten ungleich zugänglich. Kommerzielle Software ist kostenpflichtig, freie Web-Werkzeuge übertragen Sequenzdaten an Server Dritter. Diese Arbeit dokumentiert eine browser-clientseitige Werkzeugfamilie für drei Routine-Aufgaben (Restriktionsenzym-Site-Mapping, PCR-Primer-Design, FASTA-Verarbeitung), bei der alle Berechnungen im Browser stattfinden. Drei Beiträge: erstens die Einordnung in einen seit etwa 2019 gewachsenen Feldstand clientseitiger Bioinformatik (WebAssembly-Suiten wie biowasm und ViralWasm, JavaScript-Bibliotheken wie SeqViz) mit Bestimmung der verbleibenden didaktischen Lücke; zweitens eine differenzierte rechtliche Einordnung, die clientseitige Verarbeitung als strukturelle, aber nicht absolute Risikominderung beschreibt und die Abgrenzung zur Medizinprodukte-Regulierung nach MDCG 2019-11 Rev. 1 (2025) vornimmt; drittens eine durchgeführte Validierung der T_m -Berechnung gegen eine unabhängige Implementierung derselben Methodik (120 Vergleiche, alle innerhalb ± 2 °C, konstanter Versatz von $-0,89$ °C, zurückführbar auf die Konzentrationskonvention C_T gegenüber $C_{T/4}$). Limitationen werden benannt: eine Validierung gegen experimentelle Schmelztemperaturen steht aus, ebenso die Prüfung der Restriktions-Site-Erkennung.

1. Einleitung

1.1 Problemstellung

Der Zugang zu Werkzeugen der Molekularbiologie und Bioinformatik ist in Lehrkontexten ungleich verteilt. Forschungsstarke Universitäten unterhalten Lizenzen für kommerzielle Pakete; Schulen, Pädagogische Hochschulen und Einrichtungen mit knapper Ausstattung tun dies regelmäßig nicht.

Diese Ungleichheit ist empirisch belegt. Die NIBLSE-Erhebung unter 1.260 Lehrenden der Life Sciences (Williams, Drew, Galindo-Gonzalez et al. 2019) fand als häufigste Barriere fehlende Fachkenntnis und Fortbildung der Lehrenden; daneben wurden überfüllte Curricula, unzureichende mathematisch-informatische Vorbereitung der Studierenden sowie begrenzter Zugang zu Ressourcen einschließlich Hardware, Software und geprüfter Lehrmaterialien genannt. Bemerkenswert ist die Diskrepanz zwischen Zustimmung und Praxis: Eine große Mehrheit der Befragten hielt die Integration von Bioinformatik für geboten, doch nur etwa ein Drittel unterrichtete entsprechende Inhalte tatsächlich. Die Folgerhebung von 2022 unter 509 Lehrenden (Frontiers in Education 2023) bestätigte die Lage; Lehrende an Minority-Serving Institutions und aus unterrepräsentierten Gruppen berichteten Barrieren häufiger. Die Zugangsfrage ist damit zugleich eine Gerechtigkeitsfrage.

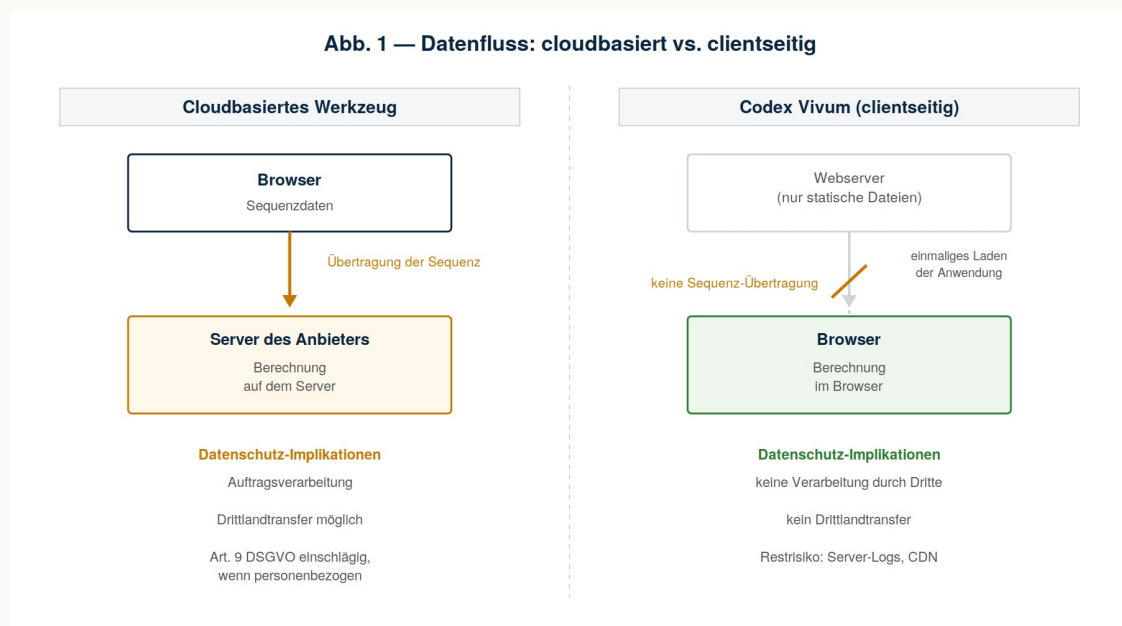
Für den Schulbereich zeigen Arbeiten wie das 4273 π -Projekt (Barker et al. 2022), dass Bioinformatik-Unterricht an Schulen machbar und curricular anschlussfähig ist, sofern die technische Einstiegshürde niedrig bleibt.

Hinzu tritt eine datenschutzrechtliche Lage. Cloudbasierte Analysewerkzeuge übertragen Sequenzen an Server Dritter. Sofern ein Personenbezug besteht oder herstellbar ist, sind genetische Daten besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO. Auch unabhängig davon entsteht regelmäßig eine Auftragsverarbeitungslage, bei Anbietern außerhalb der Europäischen Union zusätzlich eine Transferfrage.

1.2 Ansatz

Codex Vivum (lat. *lebendiges Verzeichnis*) verfolgt einen bewusst engen Zuschnitt: drei fokussierte Werkzeuge für Routine-Aufgaben, vollständig im Browser ausgeführt, ohne Installation, ohne Konto, ohne Übertragung der Eingabedaten.

Die Architektur ist keine bloße technische Präferenz, sondern eine strukturelle Antwort auf die beschriebene Lage: Was nicht übertragen wird, muss weder vertraglich abgesichert noch in ein Drittland transferiert werden (Abb. 1).



1.3 Beitrag

Die Arbeit leistet dreierlei: (1) Einordnung in den Stand der Technik clientseitiger Bioinformatik und

Bestimmung der verbleibenden Lücke; (2) differenzierte rechtliche Einordnung einschließlich der Grenzen des Ansatzes und der Abgrenzung zur Medizinprodukte-Regulierung; (3) Dokumentation einer durchgeführten Validierung der Schmelztemperatur-Berechnung mit offengelegter Methodik und benannten Rest-Lücken.

2. Stand der Technik

2.1 Clientseitige Werkzeuge

Das Feld hat sich seit der Verabschiedung von WebAssembly Core 1.0 als W3C-Empfehlung im Dezember 2019 deutlich entwickelt. Drei Gruppen lassen sich unterscheiden.

JavaScript-Bibliotheken. SeqViz (Lattice Automation, MIT-Lizenz) stellt einen Sequenzviewer für DNA, RNA und Protein bereit, mit linearer und zirkulärer Darstellung, Annotationen, Primern, Restriktionsschnittstellen und Übersetzung; unterstützt werden unter anderem FASTA, GenBank, SBOL und SnapGene. Der Open Vector Editor (TeselaGen) bietet Plasmid-Editierung, Schnittstellen-Inspektion und virtuelle Verdau; das zugehörige Repository ist inzwischen als *deprecated* markiert, was bei einer Nachnutzung als Abhängigkeit zu berücksichtigen ist.

WebAssembly-Suiten. biowasm kompiliert etablierte C/C++-Werkzeuge (samtools, bcftools, minimap2, seqtk, bedtools) nach WebAssembly und macht sie über eine JavaScript-Schnittstelle im Browser ausführbar. Darauf aufbauend beschreibt ViralWasm (Ji, Aboukhalil & Moshiri 2024) eine clientseitige Werkzeugsuite für Virengenomik. Die Arbeit begründet die Architektur ausdrücklich damit, dass webbasierte Plattformen andernfalls potenziell sensible Rohsequenzdaten an entfernte Server senden müssten — dieselbe Erwägung, die dem hier beschriebenen Ansatz zugrunde liegt. Dokumentierte Grenzen betreffen die Kompilierung (32-bit, eingeschränkter SIMD-Befehlssatz), fehlende JVM-Unterstützung und Speichergrenzen des Browsers.

Serverbasierte Web-Werkzeuge. Davon abzugrenzen sind Werkzeuge, die zwar über den Browser bedient werden, die Berechnung aber serverseitig ausführen; hierzu zählen PlasMapper 3.0 (Wishart et al. 2023), CHOPCHOP (Montague et al. 2014; Labun et al. 2016) und SequenceServer. Für die datenschutzrechtliche Bewertung ist diese Unterscheidung entscheidend und wird in der Rezeption häufig übersehen.

2.2 Verbleibende Lücke

Die vorhandenen clientseitigen Angebote sind entweder Komponenten zur Integration in eigene Anwendungen (SeqViz, Open Vector Editor) oder Suiten mit Forschungsfokus (biowasm, ViralWasm). Ein Satz didaktisch reduzierter, unmittelbar benutzbarer Einzelwerkzeuge, die ohne Installation, ohne Konto und ohne Serverkomponente in einer Unterrichtsstunde einsetzbar sind, ist unterbesetzt. Genau dort setzt die hier beschriebene Werkzeugfamilie an.

Offene Lehrmaterial-Infrastrukturen wie das Galaxy Training Network und das EMBL-EBI-Trainingsportal (Swan et al. 2024) adressieren dieselbe Zugangsfrage, arbeiten jedoch serverbasiert und sind eher auf Fortgeschrittene ausgerichtet.

3. Methode und Architektur

3.1 Werkzeuge

Restriktionsenzym-Site-Mapping. Analyse gegen eine Bibliothek von 40 Enzymen; die Spezifitäten folgen REBASE (Roberts et al. 2015). Funktionen: Site-Detektion, Kartierung in linearer und zirkulärer DNA, Fragmenttabelle für Einzel- und Mehrfachverdaue, Agarosegel-Simulation mit logarithmischer Migrations-Approximation, CSV-Export. Die Expansion mehrdeutiger Erkennungssequenzen folgt der IUPAC-Nomenklatur (IUPAC-IUB 1985); geprüft wurden die kritischen Klassen *AvaI* (CYCGRG), *ApoI* (RAATTY), *BstYI* (RGATCY) und *NlaIV* (GGNNCC).

PCR-Primer-Design. Schmelztemperatur nach dem Nearest-Neighbor-Modell (SantaLucia 1998) mit Salzkorrektur nach Owczarzy et al. (2004). Geprüft werden Self-Dimer, Haarnadelstrukturen, Paar-Dimere, GC-Clamp am 3'-Ende und Homopolymer-Läufe; Ausgabe als bewertete Paarliste mit CSV-Export.

FASTA-Verarbeitung. Parsing, GC-Gehalt, Basenkomposition, Molekulargewicht, Reverse Complement, Transkription, Sechs-Frame-Translation nach NCBI Translation Table 1, ORF-Suche oberhalb einer konfigurierbaren Mindestlänge.

3.2 Architektur-Entscheidungen

Alle Berechnungen laufen in JavaScript im Browser; der Webserver liefert ausschließlich statische Dateien. Die Werkzeuge sind als eigenständige Komponenten ausgelegt und einzeln deploybar. Bibliotheken werden zur Bauzeit gebündelt; zur Laufzeit bestehen keine Verbindungen zu Dritten. Eingaben verlassen den Browser nicht; es werden weder Cookies noch Local-Storage-Einträge zu Eingabezwecken angelegt. Das erzeugte statische Bündel umfasst etwa 282 kB JavaScript (81 kB gzip) und 45 kB CSS (8 kB gzip).

4. Rechtliche Einordnung

4.1 Datenschutz

Genetische Daten sind besondere Kategorien nach Art. 9 DSGVO nur, soweit sie personenbezogen sind. Die in der Lehre üblichen Sequenzen — Plasmide, Referenzgene von Modellorganismen, synthetische Beispiele — sind das regelmäßig nicht. Erst wenn Lernende eigene, potenziell rückführbare Sequenzen einbrächten, griffe der erhöhte Schutz; bei Minderjährigen zusätzlich Art. 8 DSGVO. Die clientseitige Architektur reduziert dieses Risiko strukturell, weil solche Daten das Endgerät nicht verlassen.

Der Europäische Datenschutzausschuss hat 2026 Leitlinien zur Anonymisierung sowie zur Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken zur Konsultation gestellt. Beide befinden sich im Entwurfsstadium und sind mit entsprechendem Vorbehalt heranzuziehen.

Für Transfers in Drittländer besteht seit dem Angemessenheitsbeschluss zum EU-US Data Privacy Framework (2023/1795) eine Grundlage; das Gericht der Europäischen Union wies eine Klage dagegen 2025 ab, ausdrücklich bezogen auf die Rechts- und Faktenlage bei Erlass. Ein Rechtsmittel ist anhängig. Für cloudbasierte Werkzeuge bleibt damit strukturelle Unsicherheit, der eine clientseitige Architektur schlicht ausweicht.

4.2 Grenzen des Ansatzes

Die Aussage „keine Datenübertragung“ gilt für die Eingabedaten, nicht für den Betrieb insgesamt. Drei Restrisiken bleiben und sollten offen kommuniziert werden:

- **Server-Logs.** Der ausliefernde Webserver verarbeitet IP-Adressen, die nach der Rechtsprechung personenbezogen sein können. Rechtsgrundlage ist regelmäßig Art. 6 Abs. 1 lit. f DSGVO.
- **Content Delivery Networks.** Werden Bibliotheken oder Schriften von Dritt-CDNs geladen, entsteht bereits beim Seitenaufruf ein Datenabfluss. Gegenmaßnahme: vollständige Selbst-Auslieferung aller Ressourcen.
- **Telemetrie Dritter.** Analysewerkzeuge im Seiten-Kontext unterlaufen den Ansatz. Gegenmaßnahme: Verzicht.

Subresource Integrity sichert die Integrität eingebundener Ressourcen, ist aber keine Datenschutzmaßnahme.

4.3 Abgrenzung zur Medizinprodukte-Regulierung

Nach der Leitlinie MDCG 2019-11 in der Fassung von 2025 qualifiziert Software nur bei medizinischer Zweckbestimmung als Medizinprodukte-Software. Werkzeuge ohne diagnostische oder therapeutische Zweckbestimmung fallen weder unter die Verordnung (EU) 2017/745 noch unter die Verordnung (EU) 2017/746. Da die Zweckbestimmung das entscheidende Kriterium ist, wird sie in der Anwendung ausdrücklich benannt: ausschließlich Lehre und Forschung, keine Diagnostik, keine klinische Entscheidungsunterstützung. Der Hinweis erscheint beim ersten Aufruf und dauerhaft im Fußbereich.

5. Validierung der Schmelztemperatur-Berechnung

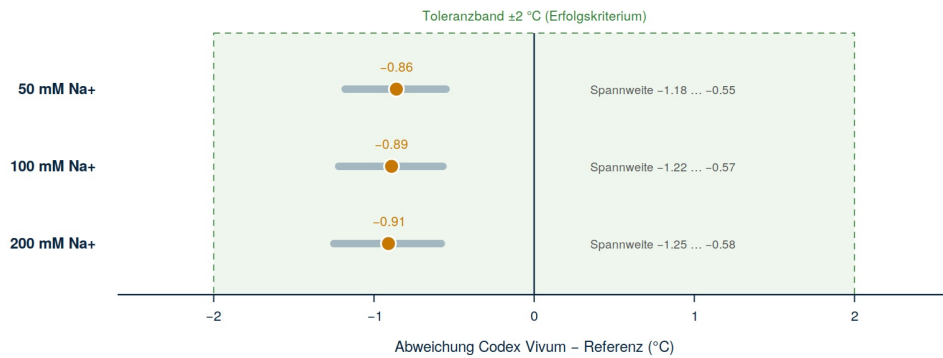
5.1 Anlage

Die Berechnung wurde aus der Anwendung nach Python portiert und gegen Biopython `MeltingTemp.Tm_NN` verglichen — eine unabhängige Implementierung derselben Methodik (SantaLucia 1998 mit Owczarzy-Salzkorrektur). Testmenge: 40 systematisch erzeugte Primer über Längen von 16 bis 30 Nukleotiden und GC-Gehalte von 30 bis 70 Prozent, jeweils bei 50, 100 und 200 mM Na⁺ und 250 nM Primerkonzentration; insgesamt 120 Vergleichspunkte. Das Skript ist deterministisch und liegt der Arbeit bei.

5.2 Ergebnis

Abb. 2 — Tm-Validierung: Abweichung gegenüber Referenz-Implementierung

120 Vergleiche · 40 Primer (16–30 nt, 30–70 % GC) · 250 nM · Referenz: Biopython Tm_NN (SantaLucia 1998 + Owczarzy 2004)



Alle 120 Werte innerhalb ± 2 °C. Der konstante Versatz ist auf die Konzentrationskonvention (C_T vs. C_T/4) zurückzuführen, nicht auf abweichende Parameter.

Bedingung	Mittlere Abweichung	Standardabweichung	Max. Betrag	innerhalb ± 2 °C
50 mM Na ⁺	-0,86 °C	0,16 °C	1,18 °C	100 %
100 mM Na ⁺	-0,89 °C	0,17 °C	1,22 °C	100 %
200 mM Na ⁺	-0,91 °C	0,17 °C	1,25 °C	100 %
gesamt (n = 120)	-0,89 °C	0,17 °C	1,25 °C	100 %

Das vorab festgelegte Kriterium (höchstens ± 2 °C bei mindestens 80 Prozent der Fälle) ist erfüllt.

5.3 Deutung des Versatzes

Der Versatz von $-0,89$ °C ist konstant über alle Bedingungen. Ein erster Durchlauf ergab scheinbar $-1,78$ °C und nur 73 Prozent innerhalb der Toleranz. Die Ursachenanalyse zeigte keinen Rechenfehler, sondern eine unterschiedliche Konzentrationskonvention: Das Nearest-Neighbor-Modell definiert die Schmelztemperatur über die Gesamtstrangkonzentration C_T , wobei für nicht-selbstkomplementäre Duplexe der Faktor $C_T/4$ gilt. Biopython bildet dies über zwei Konzentrationsparameter ab; die Dokumentation weist ausdrücklich darauf hin, dass andere Werkzeuge $C_T/4$ verwenden und die Parameter entsprechend zu setzen sind. Nach Angleichung der Konvention sank der Versatz auf $-0,89$ °C; bei einer exakt passend parametrisierten Einzelsequenz betrug die Differenz $0,02$ °C.

Dieser Befund ist nicht nur ein methodisches Detail. Die Uneinheitlichkeit der Konventionen zwischen Werkzeugen ist seit langem dokumentiert (Panjkovich & Melo 2005; Chavali et al. 2005) und eine wesentliche Quelle scheinbarer Widersprüche. Nutzerinnen und Nutzer, die Werte verschiedener Werkzeuge vergleichen, müssen die zugrunde gelegte Konvention kennen. Die Anwendung weist sie daher aus.

5.4 Was diese Validierung nicht zeigt

Verglichen wurde gegen eine Implementierung derselben Methodik, nicht gegen Messwerte. Die absolute Genauigkeit gegenüber experimentell bestimmten Schmelztemperaturen ist damit nicht geprüft. Für diesen Schritt stehen öffentliche Datensätze bereit: die Duplex-Messreihen von Owczarzy et al. (2004, 2008), die Referenzmenge von 348 Sequenzen aus Panjkovich & Melo (2005) sowie die Nearest Neighbor Database der SantaLucia-Gruppe.

5.5 Einordnung der Größenordnung

Eine Abweichung von 0,89 °C ist nur im Verhältnis zu der Streuung zu beurteilen, die zwischen den veröffentlichten Methoden selbst besteht. Auf denselben 20 Primern ergeben vier publizierte Parametersätze im Mittel: Breslauer 1986 61,23 °C, Sugimoto 1996 56,41 °C, SantaLucia 1998 52,75 °C, SantaLucia & Hicks 2004 52,69 °C. Die Streuung je Sequenz über diese vier Sätze beträgt im Mittel **8,67 °C**, im Extremfall 14,63 °C.

Die eigene Abweichung liegt damit bei etwa einem Zehntel dessen, was allein die Wahl des Parametersatzes ausmacht. Daraus folgt zweierlei. Die Umsetzung von SantaLucia 1998 ist genau im Rahmen des überhaupt Bestimmbaren. Und, für den Lehrzweck wichtiger: Ein T_m -Wert ohne Angabe der Methode ist wenig aussagekräftig — wer Werte zweier Werkzeuge vergleicht, vergleicht womöglich Breslauer gegen SantaLucia und sieht fast neun Grad Unterschied, ohne dass eines falsch rechnet. Genau dies berichten Panjkovich und Melo (2005) als Ergebnis ihres Methodenvergleichs. Die Anwendung weist die verwendete Methode deshalb aus.

Der Unterschied zwischen SantaLucia 1998 und der neueren Zusammenfassung von SantaLucia und Hicks (2004) beträgt auf derselben Testmenge 0,06 °C im Mittel; die Wahl des älteren, breiter dokumentierten Satzes ist für ein Lehrwerkzeug damit unbedenklich.

6. Prüfung der Restriktionsstellen-Erkennung

6.1 Anlage und Ergebnis

Die zweite vor-registrierte Prüfung betrifft die Erkennung der Schnittstellen. Verglichen wurde gegen Bio.Restriction aus Biopython [cock2009], dessen Enzymdaten aus REBASE stammen. Testmenge: 19 Sequenzen — zwölf erzeugte Sequenzen über die Längen 1 bis 10 kb bei GC-Gehalten von 35 bis 65 Prozent, fünf konstruierte Randfälle (Stelle am Sequenzanfang, am Ende, über die Naht zirkulärer DNA, dicht aufeinanderfolgende und mehrdeutige Muster) sowie die Klonierungsvektoren pUC19 (2.686 bp, GenBank L09137) und pBR322 (4.361 bp, J01749.1). Geprüft wurden alle 43 Enzyme der Bibliothek, je für lineare und zirkuläre DNA: 1.634 Positionsvergleiche und 178 Vergleiche der Fragmentgrößen.

Hinzu tritt ein Abgleich gegen eine dritte Quelle, die von der Referenzimplementierung unabhängig ist: Für pUC19 ist publiziert, dass die zehn Polylinker-Enzyme EcoRI, SacI, KpnI, SmaI, BamHI, XbaI, SalI, PstI, SphI und HindIII je genau einmal schneiden. Das Werkzeug findet für alle zehn genau eine Schnittstelle.

Nach Behebung des in Abschnitt 6.2 beschriebenen Fehlers stimmen alle 1.634 Positionen und alle 178 Fragmentgrößen mit der Referenz überein. Die drei Enzyme mit mehrdeutiger Erkennungssequenz — AvaI, BstEII, HinfI — stimmen auf allen Testsequenzen überein.

6.2 Ein gefundener Fehler

Der erste Durchlauf ergab 99,18 Prozent Übereinstimmung. Alle Abweichungen betrafen NlaIV mit der Erkennungssequenz GGNCC; das Werkzeug fand systematisch weniger Stellen als die Referenz. Die Prüfung der abweichenden Positionen zeigte, dass es sich um gültige Treffer handelte: An Stellen wie GGGTCCC überlappen zwei Erkennungssequenzen, und der reguläre Ausdruck setzte nach einem Treffer hinter dessen Ende fort statt eine Position weiter. Der zweite Treffer ging verloren.

Der Fehler äußerte sich nicht als Fehlfunktion, sondern als stilles Fehlen einzelner Schnittstellen — mit

entsprechend falschen Fragmentlängen. Ein systematischer Test ergab, dass sich sieben der 43 Muster selbst überlappen können (ClaI, NheI, NotI, NsiI, PvuI, SphI, NlaIV); bei den sechs eindeutigen Mustern ist die Überlappung in Zufallssequenzen selten genug, dass sie in der Testmenge nicht auftrat. Der Fehler war dort latent vorhanden.

Behoben wurde er, indem der Suchausdruck in einen nichts verbrauchenden Lookahead gefasst wird und der Suchlauf um eine Position statt um die Trefferlänge vorrückt.

Die Fortführung der Prüfung ergab zwei weitere Mängel. Erstens führte das Prüfskript eine von Hand übertragene Zweitfassung der Enzymtabelle; der Abgleich mit der Quelle zeigte neun Abweichungen. Geprüft wurden teilweise Enzyme, die es im Werkzeug nicht gibt, während neun vorhandene nie erfasst wurden — der zunächst berichtete Wert bezog sich damit auf eine Bibliothek, die so nicht existiert. Das Skript liest die Tabelle nun aus der Quelle und bricht ab, wenn sie fehlt.

Zweitens ergab der Lauf über die tatsächliche Bibliothek 99,04 Prozent, mit Abweichungen ausschließlich bei BsaI: Jede Erkennungsstelle im Gegenstrang lag um genau vier Nukleotide daneben. BsaI ist ein Typ-IIS-Enzym und schneidet außerhalb seiner Erkennungssequenz, nach REBASE GGTCTC(1/5) — ein Nukleotid hinter der Stelle im oberen, fünf im unteren Strang. Das Datenmodell kannte nur einen Schnittversatz; für Stellen im Gegenstrang wird der zweite gebraucht. Behoben durch ein zusätzliches Feld im Enzymeintrag und eine gesonderte Rechnung für Typ-IIS-Enzyme.

Diese drei Befunde sind der eigentliche Ertrag der Prüfung. Keiner wäre ohne den Vergleich gegen eine unabhängige Implementierung aufgefallen; der zweite betraf nicht das Werkzeug, sondern die Prüfung selbst, und der dritte wäre ohne seine Behebung verborgen geblieben, weil BsaI in der abgetippten Liste fehlte. Sie werden hier dokumentiert, weil eine Validierung, die nur Bestätigungen berichtet, ihren Zweck verfehlt.

6.3 Grenzen dieser Prüfung

Die Referenz bezieht ihre Enzymdaten aus derselben Quelle wie das Werkzeug. Übereinstimmung belegt die korrekte Umsetzung dieser Daten, nicht deren Richtigkeit. Geprüft wurden Positionen und Fragmentlängen, nicht die Gel-Darstellung und nicht die Angaben zu Überhängen. Getestet wurde auf erzeugten Sequenzen; ein Abgleich an annotierten Vektoren wie pUC19 gegen NEBcutter bleibt sinnvoll. Methylierungsabhängigkeit, etwa bei DpnI, wird vom Werkzeug nicht modelliert.

6.4 Prüfung der Enzym-Metadaten

Die Anwendung zeigt neben den Positionen zwei Angaben, die in keine Rechnung eingehen, aber abgelesen werden: die Art des Überhangs und die Isoschizomere. Beide wurden gegen dieselbe Referenz geprüft. Die Art des Überhangs stimmt bei allen 43 Enzymen. Bei den Isoschizomeren ergab sich ein Widerspruch: PciI (ACATGT) war als Isoschizomer von AflIII geführt, das jedoch ACRYGT erkennt — vier Sequenzen, von denen ACATGT nur eine ist. Enzyme, deren Erkennungssequenz eine echte Teilmenge einer anderen bildet, sind keine Isoschizomere; die Isoschizomere von PciI sind nach REBASE BspLU11I und PscI. Die Angabe wurde korrigiert.

Der Fall ist lehrreich, weil er sich der bisherigen Prüflogik entzogen hätte: Er verändert keine berechnete Position, sondern nur eine angezeigte Beziehung. Wer daraus schlösse, beide Enzyme schnitten dieselben Stellen, plante auf einer Sequenz mit ACGCGT einen Verdau falsch — ohne dass die Anwendung etwas Auffälliges

gezeigt hätte. Nicht geprüft sind die Angaben zum Herkunftsorganismus; die Referenz führt dieses Feld nicht in vergleichbarer Form.

7. Limitationen

- Keine Validierung gegen experimentelle Schmelztemperaturen (Abschnitt 5.4).
- Der Abgleich der Restriktionsstellen erfolgte gegen eine unabhängige Implementierung und gegen die publizierte Karte von pUC19 (Abschnitt 6), nicht gegen NEBcutter selbst.
- Kein Sequenzabgleich gegen Genomdatenbanken: Solche Suchen setzen serverseitige Indizes voraus und sind mit einer rein clientseitigen Architektur nicht abbildbar. Verwiesen wird auf NCBI BLAST und Ensembl.
- Verweise auf Primärliteratur werden manuell gepflegt; eine automatisierte Anbindung besteht nicht.
- Nutzung setzt einen aktuellen Browser voraus — im Lehrkontext gerade dort relevant, wo die Ausstattung knapp ist.
- Die Parameter folgen SantaLucia (1998); spätere Zusammenfassungen (SantaLucia & Hicks 2004) wurden nicht eingearbeitet. Für den Lehrzweck ist das eine bewusste Vereinfachung, keine Genauigkeitsaussage.

8. Einordnung

8.1 Local-first als Position

Kleppmann et al. (2019) formulieren mit dem Begriff *local-first software* eine Reihe von Zielen, darunter Offline-Fähigkeit, Datenhoheit der Nutzenden und Langlebigkeit unabhängig vom Fortbestand eines Anbieters. Die hier beschriebene Architektur ist ein Anwendungsfall dieser Position im Bildungsbereich: Sie macht Datenschutz nicht zur Frage eines Vertrags, sondern zur Eigenschaft des Aufbaus.

8.2 Offene Lehrwerkzeuge als Allmende

Hess und Ostrom (2007) beschreiben Wissen als Allmende; das daran anschließende Rahmenwerk zur Governance von Wissensallmenden (Frischmann, Madison & Strandburg 2014) überträgt Ostroms Analyseansatz auf Datenteilung und Software-Zusammenarbeit. Der für dieses Vorhaben belastbare Kern lautet: Das Wissen selbst ist nicht-rival, die *Infrastruktur* seiner Pflege — Hosting, Kuratierung, Dokumentation, Lizenzpflege — ist es nicht. Offene Lehrsoftware ist damit nicht durch ihre Lizenz nachhaltig, sondern durch geregelte Pflege. Das ist weniger eine Metapher als eine Anforderung an das weitere Vorgehen.

8.3 Zur Grenze evolutionärer Analogien

Analogien zwischen Software-Ökosystemen und ökologischen oder evolutionären Mechanismen (Nischenbildung, Selektion, Redundanz) liegen nahe und werden häufig verwendet. Eine Prüfung der Literatur ergibt jedoch, dass sie dort überwiegend illustrativ bleiben und selten empirisch operationalisiert sind. Sie werden hier deshalb nicht als tragendes Argument geführt. Der commons-theoretische Rahmen ist für die Fragestellung belastbarer.

9. Veröffentlichung und Reproduzierbarkeit

Lizenzen. Text und Dokumentation stehen unter CC BY 4.0, der Quellcode unter AGPL-3.0-or-later. Die Netzwerk-Klausel der AGPL verpflichtet, wer die Software als Dienst anbietet, zur Offenlegung des Quellcodes einschließlich Änderungen. Als alleiniger Rechteinhaber kann der Autor daneben abweichende Lizenzen vergeben. Sollten Beiträge Dritter aufgenommen werden, ist dafür eine vorherige Beitragsvereinbarung erforderlich; ohne eine solche ist eine spätere abweichende Lizenzierung praktisch ausgeschlossen.

Die naheliegende Alternative — eine Lizenz nach dem Muster „für private Nutzung frei, für gewerbliche entgeltlich“ — wurde geprüft und verworfen. Drei Gründe. Erstens verbietet Punkt 6 der Open Source Definition Einschränkungen nach Einsatzfeld; eine solche Lizenz wäre nicht Open Source, und software-spezifische Fachzeitschriften verlangen durchgängig eine anerkannte Lizenz. Zweitens ist der Begriff „nicht-kommerziell“ nachweislich unscharf: Die von Creative Commons beauftragte Untersuchung [ccnc2009] fand außerhalb von Werbung und direktem Gelderwerb mehr Unsicherheit als Klarheit, und ein deutsches Berufungsgericht wertete diese Unklarheit ausdrücklich zulasten des Lizenzgebers (OLG Köln, Urteil vom 31.10.2014, Az. 6 U 60/14). Drittens greift der Schutzzweck hier nicht: Lizenzen wie die Business Source License wehren ab, dass ein fremder Anbieter Server-Software als eigenen Dienst vermarktet — bei einer Anwendung ohne Server-Backend besteht diese Gefahr nicht.

Daraus folgt ein Modell in drei Stufen: private und institutionelle Nutzung kostenfrei über die AGPL; entgeltlich nur der *Betrieb* als Dienst und die Befreiung von den Offenlegungspflichten. Der Vorteil ist nicht bloß rechtlicher Natur — es muss nie entschieden werden, ob eine Nutzung „kommerziell“ ist. Die Frage, an der Lizenzen mit Nicht-Kommerz-Klausel scheitern, stellt sich nicht.

Zitierbarkeit. Der Arbeit liegen eine `CITATION.cff` nach dem Citation File Format sowie Zenodo-Metadaten bei. Für die Archivierung ist die Trennung von versionsbezogener und konzeptbezogener DOI vorgesehen, wie sie Zenodo bereitstellt. Die Ausrichtung folgt den FAIR-Prinzipien für Forschungssoftware (Chue Hong et al. 2022; Barker et al. 2022) und den Software Citation Principles (Smith et al. 2016).

Veröffentlichungsweg. Für eine Einreichung beim Journal of Open Source Software ist zu beachten, dass web-basierte Werkzeuge dort nur eingeschränkt im Themenbereich liegen; verlangt werden insbesondere eine von der Oberfläche getrennte, testbare Kernbibliothek und automatisierte Tests. Beides liegt inzwischen vor (siehe Abschnitt Reproduzierbarkeit). Alternativen mit anderem Zuschnitt sind BMC Bioinformatics (Software Article) und F1000Research.

Reproduzierbarkeit. Die Validierungsskripte sind deterministisch (feste Zufallsstartwerte) und liegen mit Abhängigkeitsangabe bei. Die Anwendung selbst verarbeitet keine personenbezogenen oder sensiblen Daten.

Aufbau des Quelltextes. Die Rechenlogik liegt getrennt von der Oberfläche in eigenen Modulen (`src/lib/restriction.js`, `src/lib/thermo.js`) und besteht aus reinen Funktionen ohne Bezug zum Anzeigeteil. Darauf setzt ein Testsatz mit 44 Prüfungen auf, der ohne Zusatzsoftware läuft (`node --test`). Drei dieser Prüfungen halten die in Abschnitt 6 beschriebenen Fehler fest, damit sie nicht unbemerkt zurückkehren. Diese Trennung ist zugleich die Voraussetzung dafür, dass ein Tool-Beitrag beim Journal of Open Source Software im Themenbereich liegt.

10. Schluss

Der Beitrag dieser Arbeit liegt weniger im Funktionsumfang als im Zuschnitt: kein Server, kein Konto, keine Laufzeit-Abhängigkeit von Dritten, dafür offengelegte Methodik und benannte Grenzen. Die Validierung zeigt, dass die Berechnung mit einer unabhängigen Implementierung derselben Methodik übereinstimmt, und zugleich, wie leicht ein Konventionsunterschied als Fehler erscheinen kann.

Die nächsten Schritte sind bestimmt: Validierung gegen experimentelle Messwerte und Trennung von Kernbibliothek und Oberfläche. Die Prüfung der Restriktionsstellen ist erfolgt und hat einen Fehler zutage gefördert, der ohne sie unbemerkt geblieben wäre. Bis dahin gilt die hier dokumentierte Aussage in genau dem Umfang, in dem sie belegt ist — und nicht darüber hinaus.

Anhang A — Betrieb

Das Bündel enthält ein Bau-Setup (Vite) sowie ein vorab erzeugtes statisches Verzeichnis, das ohne lokale Werkzeugkette unmittelbar auf einen Webserver übertragen werden kann. Der Basispfad ist in der Konfiguration einstellbar. Empfohlen wird die Selbst-Auslieferung aller Ressourcen ohne Dritt-CDN.

Anhang B — Hinweis zur Zweckbestimmung

Diese Anwendung ist kein Medizinprodukt im Sinne der Verordnung (EU) 2017/745 und kein In-vitro-Diagnostikum im Sinne der Verordnung (EU) 2017/746. Sie dient ausschließlich Lehr- und Forschungszwecken. Keine medizinische Diagnose, keine Therapieempfehlung, keine klinische Entscheidungsunterstützung. Für Laborarbeit gilt das jeweils anwendbare Recht in eigener Verantwortung. Haftung im gesetzlich zulässigen Rahmen ausgeschlossen.

Anhang C — Offenlegung

Entstehung. Die Werkzeuge, die Prüfskripte, die Abbildungen und der Text dieser Arbeit sind unter Verwendung eines KI-Assistenzsystems (Claude, Anthropic) entstanden. Konzeption, inhaltliche Entscheidungen, Prüfung und Verantwortung liegen beim Autor.

Diese Angabe steht hier, weil sie für die Einordnung erheblich ist: Ein Text, der von einem System mitverfasst wurde, das Plausibles hervorbringen kann, ohne es geprüft zu haben, muss strengeren Belegansprüchen genügen als einer aus geprüftem Fachwissen. Genau darum ist jede Aussage dieser Arbeit an eine nachvollziehbare Quelle oder an ein ausführbares Prüfskript gebunden, und jede Einschränkung ist benannt.

Prüfbarkeit. Die drei Validierungen liegen als Skripte bei, sind über feste Zufallsstartwerte reproduzierbar und enden nur bei vollständiger Übereinstimmung mit Rückgabewert null. Der Testsatz umfasst 46 Prüfungen ohne Fremdadhängigkeiten. Wer die Aussagen dieser Arbeit anzweifelt, kann sie ausführen.

Vier gefundene Fehler. Die Prüfungen haben vier Mängel zutage gefördert: verlorene überlappende Erkennungsstellen, eine im Prüfskript doppelt geführte Enzymtabelle, einen falschen Schnittversatz bei Typ-IIS-Enzymen im Gegenstrang und eine falsche Isoschizomer-Angabe. Alle sind behoben und durch Prüfungen abgesichert. Sie werden genannt, weil ihre Existenz etwas über die Verlässlichkeit des Vorgehens aussagt — und weil eine Arbeit, die nur Bestätigungen berichtet, weniger Vertrauen verdient, nicht mehr.

Vollständige Offenlegung einschließlich der Aufstellung dessen, was ausdrücklich nicht belegt ist, im begleitenden Dokument `OFFENLEGUNG.md`. —

Literatur

1. Barker D., Alderson R.G., McDonagh J.L., Plaisier H., Comrie M.M., Duncan L., Muirhead G.T.P., Sweeney S.D. (2022). Bringing bioinformatics to schools with the 4273 π project. *PLOS Computational Biology* 18(1): e1009705. DOI: 10.1371/journal.pcbi.1009705

2. Barker M., Chue Hong N.P., Katz D.S. et al. (2022). Introducing the FAIR Principles for research software. *Scientific Data* 9: 622. DOI: 10.1038/s41597-022-01710-x
3. Chavali S., Mahajan A., Tabassum R., Maiti S., Bharadwaj D. (2005). Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics* 21(20): 3918–3925. DOI: 10.1093/bioinformatics/bti633
4. Chue Hong N.P. et al. (2022). FAIR Principles for Research Software (FAIR4RS Principles) 1.0. Research Data Alliance. DOI: 10.15497/RDA00068
5. Frischmann B.M., Madison M.J., Strandburg K.J. (Hrsg.) (2014). *Governing Knowledge Commons*. Oxford University Press.
6. Hess C., Ostrom E. (Hrsg.) (2007). *Understanding Knowledge as a Commons: From Theory to Practice*. MIT Press.
7. IUPAC-IUB Joint Commission on Biochemical Nomenclature (1985). Nomenclature for incompletely specified bases in nucleic acid sequences. *Nucleic Acids Research* 13(9): 3021–3030. DOI: 10.1093/nar/13.9.3021
8. Ji D., Aboukhalil R., Moshiri N. (2024). ViralWasm: a client-side user-friendly web application suite for viral genomics. *Bioinformatics* 40(1): btae018. DOI: 10.1093/bioinformatics/btae018
9. Kleppmann M., Wiggins A., van Hardenberg P., McGranaghan M. (2019). Local-first software: you own your data, in spite of the cloud. *Proceedings of the 2019 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward!)*, 154–178.
10. Labun K., Montague T.G., Gagnon J.A., Thyme S.B., Valen E. (2016). CHOPCHOP v2. *Nucleic Acids Research* 44(W1): W272–W276. DOI: 10.1093/nar/gkw398
11. Medical Device Coordination Group (2025). MDCG 2019-11 Rev. 1 — Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 and Regulation (EU) 2017/746.
12. Montague T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E. (2014). CHOPCHOP. *Nucleic Acids Research* 42(W1): W401–W407. DOI: 10.1093/nar/gku410
13. Owczarzy R., You Y., Moreira B.G., Manthey J.A., Huang L., Behlke M.A., Walder J.A. (2004). Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry* 43(12): 3537–3554. DOI: 10.1021/bi034621r
14. Owczarzy R., Moreira B.G., You Y., Behlke M.A., Walder J.A. (2008). Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* 47(19): 5336–5353. DOI: 10.1021/bi702363u
15. Panjkovich A., Melo F. (2005). Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics* 21(6): 711–722. DOI: 10.1093/bioinformatics/bti066
16. Panjkovich A., Norambuena T., Melo F. (2005). dnaMATE: a consensus melting temperature prediction server for short DNA sequences. *Nucleic Acids Research* 33: W570–W572. DOI: 10.1093/nar/gki379
17. Roberts R.J., Vincze T., Posfai J., Macelis D. (2015). REBASE — a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* 43: D298–D299. DOI: 10.1093/nar/gku1046

18. SantaLucia J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *PNAS* 95(4): 1460–1465. DOI: 10.1073/pnas.95.4.1460
19. SantaLucia J., Hicks D. (2004). The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure* 33: 415–440. DOI: 10.1146/annurev.biophys.32.110601.141800
20. Smith A.M., Katz D.S., Niemeyer K.E., FORCE11 Software Citation Working Group (2016). Software citation principles. *PeerJ Computer Science* 2: e86. DOI: 10.7717/peerj-cs.86
21. Swan A., Broadbent J., Gurwitz K. et al. (2024). Making bioinformatics training FAIR: the EMBL-EBI training portal. *Frontiers in Bioinformatics* 4: 1347168. DOI: 10.3389/fbinf.2024.1347168
22. Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung). ABl. EU L 119/1.
23. Williams J.J., Drew J.C., Galindo-Gonzalez S. et al. (2019). Barriers to integration of bioinformatics into undergraduate life sciences education. *PLOS One* 14(11): e0224288. DOI: 10.1371/journal.pone.0224288
24. Williams J.J. et al. (2023). Revisiting barriers to implementation of bioinformatics into life sciences education. *Frontiers in Education* 8: 1317191. DOI: 10.3389/educ.2023.1317191
25. Wishart D.S. et al. (2023). PlasMapper 3.0 — a web server for generating, editing, annotating and visualizing publication quality plasmid maps. *Nucleic Acids Research* 51(W1): W459–W467. DOI: 10.1093/nar/gkad276
26. Creative Commons / Netpop Research (2009). *Defining Noncommercial: A Study of How the Online Population Understands Noncommercial Use*. Creative Commons Corporation.
27. Open Source Initiative. *The Open Source Definition*, Punkt 6 (No Discrimination Against Fields of Endeavor).
28. Oberlandesgericht Köln, Urteil vom 31.10.2014, Az. 6 U 60/14.

Korrespondenz: forschung@tschueller.com · **ORCID:** 0009-0003-9799-6747 **Affiliation:** unabhängige Forschung, Wien (privat finanziert) **Lizenz:** CC BY 4.0 (Text) · AGPL-3.0-or-later (Quellcode) · **Version:** v1.0 · 2026-05-12