

Validierungsbericht — Enzym-Metadaten

Werk: codex-vivum · **Version:** v1.0 · **Datum:** 2026-05-12 **Autor:** Thomas Schüller · ORCID 0009-0003-9799-6747 · CC BY 4.0

Anlass

Die bisherigen Prüfungen betrafen Schnittpositionen und Fragmentgrößen. Die Anwendung zeigt daneben Angaben, die in keine Rechnung eingehen, aber im Unterricht abgelesen werden: die Art des Überhangs und die Isoschizomere. Eine falsche Angabe wäre dort so schädlich wie eine falsche Position — sie fällt nur nicht auf, weil nichts abstürzt. Diese Prüfung schließt eine Lücke, die in den früheren Berichten als solche benannt war.

Anlage

Verglichen wurde gegen `Bio.Restriction` (Biopython), dessen Enzymdaten aus REBASE stammen. Die Angaben des Werkzeugs werden aus `src/lib/restriction.js` gelesen, nicht dupliziert.

Überhänge. `Bio.Restriction` führt das Attribut `ovhg`: negativ bedeutet 5'-Überhang, positiv 3'-Überhang, null ein glattes Ende. Das wird mit dem Feld `overhang` des Werkzeugs abgeglichen.

Isoschizomere. Geprüft wird nicht gegen die Isoschizomer-Liste der Referenz — diese ist umfangreicher als eine für die Lehre sinnvolle Auswahl und taugt nicht als Sollwert. Geprüft wird stattdessen die definierende Eigenschaft: Zwei Isoschizomere haben dieselbe Erkennungssequenz. Für jede Angabe im Werkzeug wird die Erkennungssequenz des genannten Partners bei der Referenz nachgeschlagen und verglichen.

Ergebnis

Prüfung	Umfang	Übereinstimmung
Art des Überhangs	43 Enzyme	43 (100 %)
Isoschizomer-Angaben	14 Angaben	14 (100 %) nach Korrektur
Zwei Isoschizomer-Angaben nennen Partner, die <code>Bio.Restriction</code> nicht führt; sie konnten nicht geprüft werden und sind als solche vermerkt.		

Befund — eine falsche Isoschizomer-Angabe

Der erste Durchlauf ergab einen Widerspruch: **PciI** (Erkennungssequenz ACATGT) war als Isoschizomer von **AflIII** geführt. AflIII erkennt jedoch ACRYGT — mit R für A oder G und Y für C oder T also vier Sequenzen: ACACGT, ACATGT, ACGCGT und ACGTGT. Die Sequenz von PciI ist davon nur eine.

Enzyme, deren Erkennungssequenz eine echte Teilmenge einer anderen ist, sind **keine** Isoschizomere. Isoschizomere haben dieselbe Erkennungssequenz; AflIII ist gegenüber PciI schlicht weniger spezifisch. Nach REBASE sind die Isoschizomere von PciI **BspLU11I** und **PscI**.

Die Angabe wurde entsprechend korrigiert. Im Testsatz hält seither eine Prüfung fest, dass eine Isoschizomer-Angabe nur zulässig ist, wenn die Erkennungssequenzen übereinstimmen — soweit der Partner selbst in der Bibliothek steht.

Warum das zählt. Wer im Unterricht liest, PciI und AflIII seien Isoschizomere, wird annehmen, dass beide dieselben Stellen schneiden. Auf einer Sequenz mit ACGCGT tun sie das nicht. Der Fehler hätte zu einem falsch geplanten Verdau geführt, ohne dass die Anwendung etwas Auffälliges angezeigt hätte.

Was dieser Bericht nicht zeigt

- Die Angaben zum **Herkunftsorganismus** sind nicht geprüft; Bio.Restriction führt dieses Feld nicht in vergleichbarer Form. Sie sind für die Rechnung ohne Belang, aber als ungeprüft zu betrachten.
- Zwei Isoschizomer-Angaben mit Partnern außerhalb der Referenz bleiben offen.
- Geprüft wurde die *Art* des Überhangs, nicht dessen Länge in Nukleotiden.

Reproduktion

```
pip install -r requirements.txt
python3 metadaten_validation.py
```

Rückgabewert 0 nur bei vollständiger Übereinstimmung. Ausgabe: `metadaten_output.txt`.