

Validation Report — Enzyme Metadata

Work: codex-vivum · **Version:** v1.0 · **Date:** 2026-05-12 **Author:** Thomas Schüller · ORCID 0009-0003-9799-6747 · CC BY 4.0

Motivation

The checks so far concerned cut positions and fragment sizes. The application also displays information that enters no calculation but is read in teaching: the type of overhang and the isoschizomers. A wrong entry there would be as harmful as a wrong position — it simply does not stand out, because nothing crashes. This check closes a gap that earlier reports named as such.

Design

The comparison was against `Bio.Restriction` (Biopython), whose enzyme data derive from REBASE. The tool's entries are read from `src/lib/restriction.js`, not duplicated.

Overhangs. `Bio.Restriction` carries the attribute `ovhg`: negative means a 5' overhang, positive a 3' overhang, zero a blunt end. This is compared with the tool's `overhang` field.

Isoschizomers. The check is not made against the reference's isoschizomer list — that list is more extensive than a selection sensible for teaching and does not serve as a target value. Instead the defining property is checked: two isoschizomers have the same recognition sequence. For every entry in the tool, the recognition sequence of the named partner is looked up in the reference and compared.

Result

Check	Scope	Agreement
Type of overhang	43 enzymes	43 (100 %)
Isoschizomer entries	14 entries	14 (100 %) after correction
Two isoschizomer entries name partners that <code>Bio.Restriction</code> does not carry; they could not be checked and are noted as such.		

Finding — one incorrect isoschizomer entry

The first run revealed a contradiction: **PciI** (recognition sequence ACATGT) was listed as an isoschizomer of **AflIII**. **AflIII**, however, recognises ACRYGT — with R for A or G and Y for C or T, that is four sequences: ACACGT, ACATGT, ACGCGT and ACGTGT. **PciI**'s sequence is only one of them.

Enzymes whose recognition sequence is a proper subset of another's are **not** isoschizomers. Isoschizomers have the same recognition sequence; **AflIII** is simply less specific than **PciI**. Per REBASE, the isoschizomers of **PciI** are **BspLU11I** and **PscI**.

The entry was corrected accordingly. A test now pins down that an isoschizomer entry is admissible only if the recognition sequences match, as far as the partner is itself in the library.

Why this matters. Anyone reading in class that PciI and AflIII are isoschizomers will assume both cut the same sites. On a sequence containing ACGCGT they do not. The error would have led to a wrongly planned digest without the application showing anything untoward.

What this report does not show

- The **source organism** entries are unchecked; Bio.Restriction does not carry this field in comparable form. They are irrelevant to the calculation but must be regarded as unverified.
- Two isoschizomer entries with partners outside the reference remain open.
- The *type* of overhang was checked, not its length in nucleotides.

Reproduction

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

Exit status 0 only on complete agreement. Output: `metadaten_output.txt` .