

Предсказание аминокислот по нерибосомным кодам.

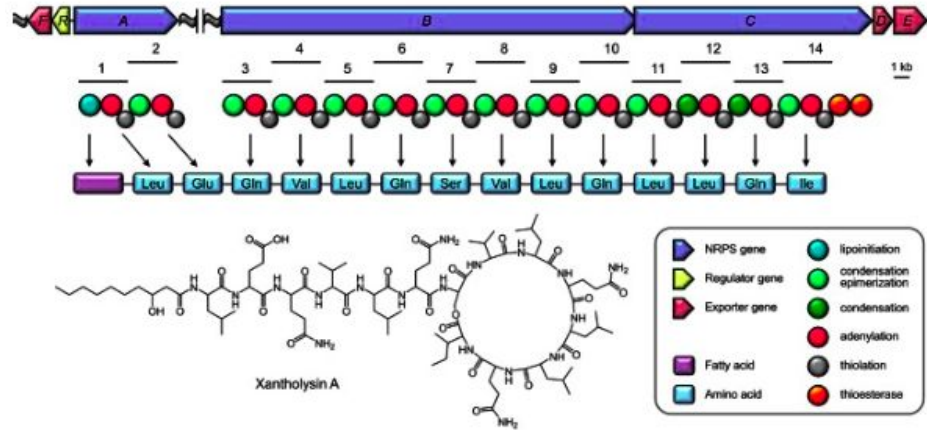
Исаев Василий

Гуревич Алексей, Кунявская Ольга

Нерибосомные пептиды (Nonribosomal peptide, NRP) — природные соединения, которые синтезируются многими микроорганизмами и имеют важные лекарственные свойства.

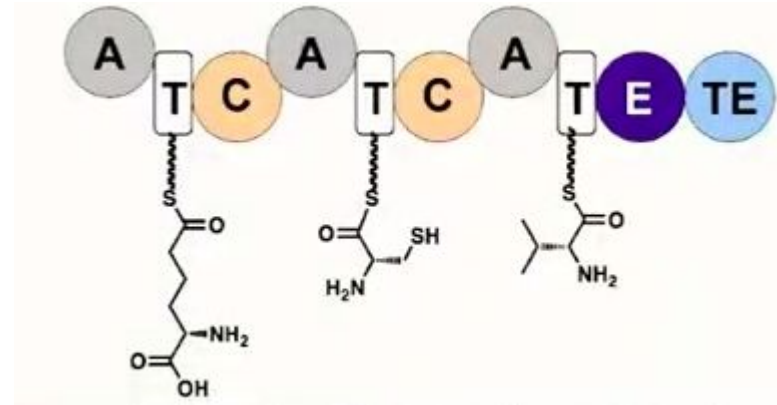


Нерибосомная синтетаза(NRPS) -- белковый комплекс, который производит NRP. Состоит из повторяющихся доменов.



А-домен отвечает за привлечение нужной аминокислоты.

Специфическая последовательность аминокислот в А-домене — **нерибосомный код**.



Аналоги: NRPsPredictor2, AntiSMASH

Алгоритм nrpspredictor2

Вход: 10-АА код, база данных aminos {amino: 10-АА}

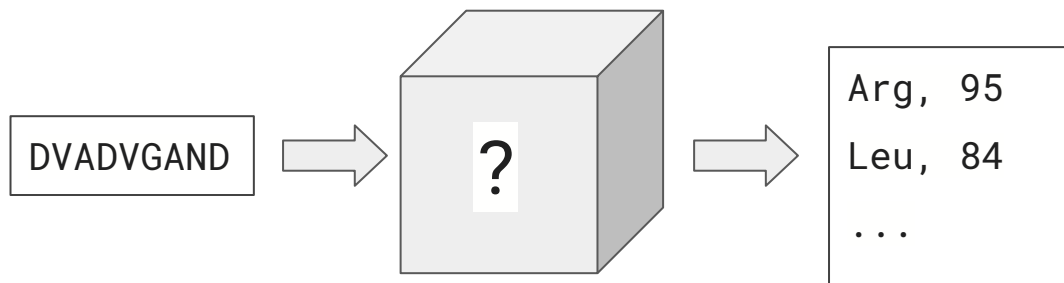
Находит код с самым большим количеством совпадений аминокислот

Скор: количество совпавших АА

Проблемы:

- Для двух аминокислот могут быть из одинаковые скоры, не понятно, как выбрать максимум
- Плохо предсказываются аминокислоты, мало представленные в базе данных
- **AntiSMASH5** не выдает предсказания

Цель: предсказать аминокислоты по нерибосомным кодам



Задачи:

1 - Преобразовать вывод antiSMASH5 к виду antiSMASH3

2 - Реализовать алгоритм как в Nrp predictor

```
ctg1_orf00029_A1      arg
arg(70.0);bmt(70.0);phe(60.0);
trp(50.0);bmt(50.0);...
```

```
"large_cluster_pred": ["phe", "trp", "phg", "tyr", "bht"]
```

3. Определение для одинаковых скоров

4. Улучшение текущего алгоритма

Алгоритм со взвешенным скором

Вход: 10-AA код, база данных aminos {amino: 10-AA}

Идея: Для каждой позиции в 10-AA коде считаем вероятность конкретной аминокислоты, возвращаем сумму вероятностей для каждой буквы

база данных - {"cys" : "AA", "cys" : "AB"}

считаем для cys вероятность букв на каждой позиции

1: { A : 1.0, B : 0.0 }

2: { A : 0.5, B : 0.5 }

для "BB" определяем аминокислоту:

$$0.0 + 0.5 = 0.5$$

Результаты

1. Реализован алгоритм из NRPsPredictor2
2. Реализован алгоритм перевода результатов antiSMASH5 к виду antiSMASH3
3. Реализован взвешенный алгоритм
4. Реализован алгоритм с гибридным скором

Результаты

aa-id	nrps	my_precision
kyn	phe(70.0);bmt(70.0);tyr(60.0)	kyn:(90.0);bmt:(55.0);vol:(50.0);
ile	val(90.0);ile(70.0);leu(60.0)	ile:(52.7);dpg:(46.0);val:(44.9)
dhb	thr(100.0);dht(90.0);allothr(90.0);	dhab:(90.0);dht:(90.0);thr:(69.0);

Репозиторий

<https://github.com/Vasiliy566/NERPA>