

Модификация библиотеки DAFNA для предсказания олигонуклеотидов-трансформеров

Слынько И.А., Юрушкин М.В., Бачурин С.С.

Южный федеральный университет, Институт ММиКН им. И. И. Воровича

Контакты: islynko@sfedu.ru;

GitHub: github.com/myurushkin/dfalib

Введение в неканонические структуры ДНК

Что такое НС ДНК?

Неканонические структуры ДНК (НС ДНК) — это структуры, отличные от классической двойной спирали. К ним относятся G-квадруплексы, i-мотивы, триплексы, шпильки и участки с альтернативными водородными связями.

Условия формирования

Формирование НС ДНК зависит от последовательности и условий среды. Например, G-квадруплексы требуют наличия ионов K^+ , а i-мотивы формируются при низком pH и наличии C-богатых последовательностей.

Биологическая роль

НС ДНК играют важную роль в регуляции клеточных процессов, включая репликацию, транскрипцию, поддержание теломер и контроль экспрессии генов. Их стабильность и динамика могут влиять на генетическую стабильность и адаптацию клетки.

Олигонуклеотиды-трансформеры



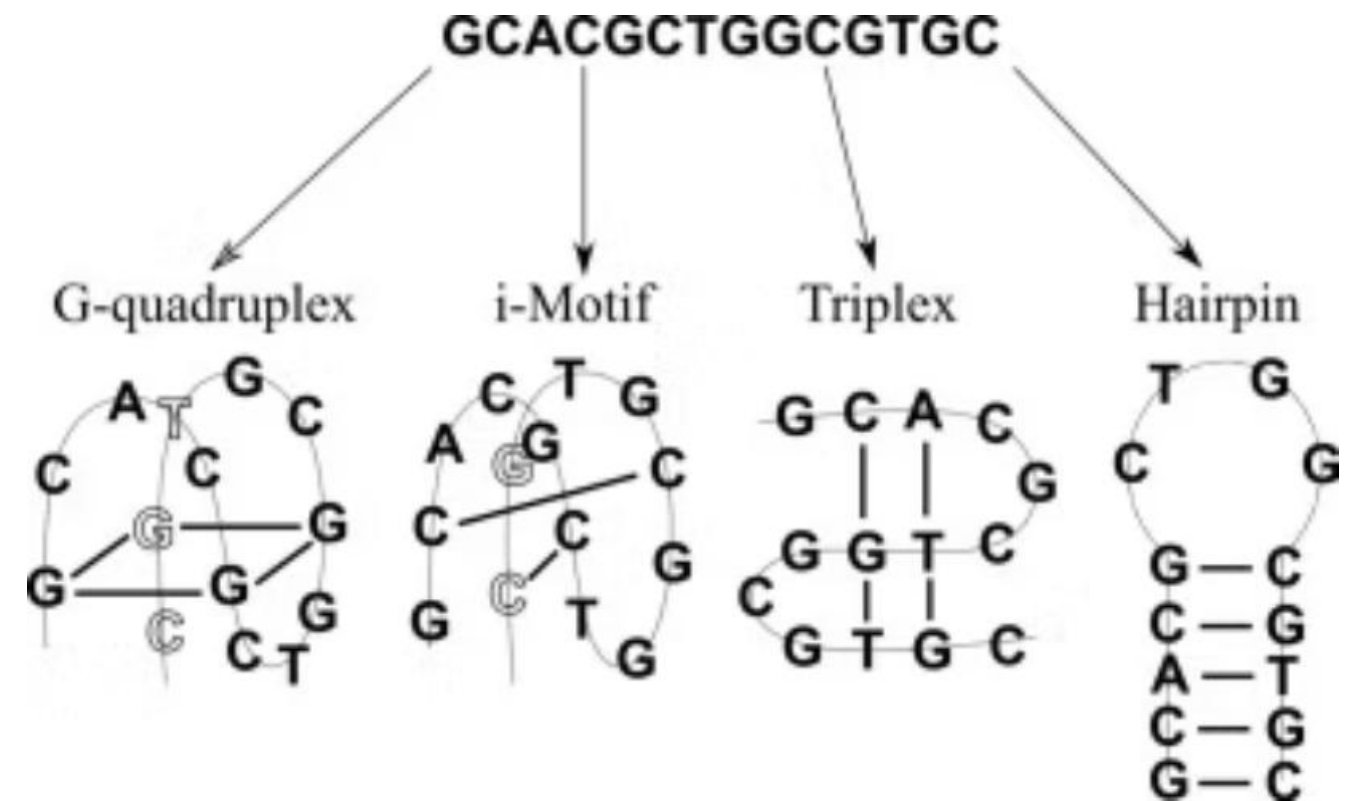
Идея трансформера

Последовательность, формирующая разные НС ДНК в зависимости от условий.



Значение

Универсальные молекулярные переключатели для терапии и наноинженерии.



Язык паттернов $\sqsupset$ AFNA

Общая идея: Описание паттернов неканонических структур через регулярные выражения

$GQD = X^* g\{m\} X\{1\} X^* g\{m\} X\{1\} X^* g\{m\} X\{1\} X^* g\{m\} X^*, m=[2:20]$

$IMT = X^* c\{a\} X\{1\} X^* c\{b\} X\{2\} X^* c\{a\} X\{1\} X^* c\{b\} X^*, a=[2:20], b=[1:20]$

$HRP = X^* a\{a\} X^* t\{t\} X^* c\{c\} X^* g\{g\} X\{3\} X^* c\{g\} X^* g\{c\} X^* a\{t\} X^* t\{a\} X^*, a=[1:], t=[1:], c=[1:], g=[1:]$

$TRP = X^* (g\{g\} a\{a\})\{m\} X\{3\} X^* (c\{g\} t\{a\})\{m\} X\{3\} X^* (t\{a\} c\{g\})\{m\}, a=[1:20], g=[0:20], m=[1:20]$

$X = (a|c|g|t)$

$result = INTERSECT(GQD, IMT, HRP, TRP)$

Архитектура алгоритма

Шаг 1

Преобразование паттерна в недетерминированный конечный автомат (НКА).

Шаг 2

Преобразование НКА в детерминированный конечный автомат (ДКА).

Шаг 3

Пересечение и минимизация автоматов для оптимизации.

Реализация и производительность

2.3

Время поиска (сек)

Минимальные строки,
соответствующие паттернам

145000

Общее количество строк

Найдено в эксперименте

200

Кандидаты отобраны

С максимальными значениями сил

Примеры найденных «трансформеров»

Последовательность	GQD	IMT	Шпилька	Триплекс
tcaccggaccggccggc cgg	2	1.5	2	4
tggaccaaccggccggc cgg	2	2	2	3
tctccggaccggccggc cgg	2	1.5	2	4
acaccggtccggccggc cgg	2	1.5	2	4

Поиск в геноме человека

- Результаты применены к обратной задаче — поиску последовательностей в геноме человека.
- Полная расшифровка генома завершена в 2022 году.
- Такие последовательности не найдены, что вызывает научный интерес.

Перспективы

Синтез in vitro

Проверка образования нужных структур в лабораторных условиях и оценка их стабильности.

Квантово-химическое моделирование

Детальная оценка термодинамики и кинетики переходов между формами «трансформера».

Молекулярные переключатели

Разработка DNA-устройств, изменяющих конформацию под воздействием ионов, pH или лиганда.

Нанотехнологии и материалы

Включение трансформеров в ДНК-наборы для сборки динамических наноструктур.

Биосенсоры и диагностические системы

Создание сенсоров, реагирующих на изменение среды через конформационные переходы.

Синтетическая биология

Инструменты для программируемой регуляции генов и создания «умных» наноагентов.

Заключение

DAFNA и паттерны

Новый формализм позволяет точно описывать и искать «трансформеры» — последовательности, образующие четыре неканонические структуры ДНК.

Результаты поиска

Найдено более 145 000 кандидатов, из них отобрано 200 с наивысшими «силами» структур.

Геном человека

В полном человеческом геноме таких последовательностей не встретилось, что может отражать эволюционные или биохимические ограничения.

Дальнейшие шаги

Синтез *in vitro* и квантово-химические исследования, оценка биомедицинского потенциала и возможностей нанотехнологий.