

Генеративные модели для структуры белков и их комплексов с другими молекулами.

Выполнил: Клыпа Р.С.

Научный руководитель: Грудинин С.В.

МФТИ

18 мая 2024 г.

Цели исследования

Цель работы

Генерация структуры молекулярных соединений (например, РНК) при взаимодействии с протеином - $P(\text{geom}_{RNA} | \text{geom}_{pr}, \text{seq}_{pr}, \text{seq}_{RNA})$.

Проблема

Существующие решения обладают рядом технических и теоретических недостатков.

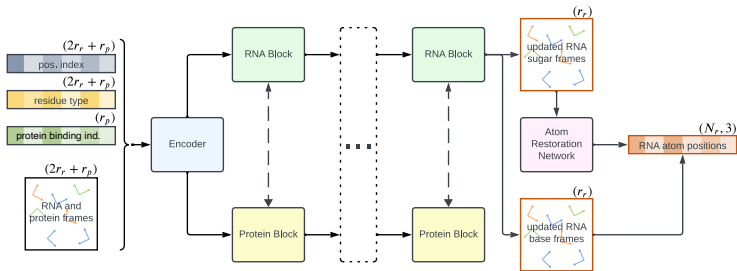
Метод решения

Дизайн архитектуры модели и генеративного процесса с учетом биологических и физико-химических особенностей задачи.

RoseTTaFoldNA

- ▶ Генеративная версия модели не позволяет использовать последовательности молекул как условие.
- ▶ Предиктивная не позволяет использовать структуру протеина как условие и не учитывает возможность существования различных итоговых структур.
- ▶ Модель зависит от Multiple Sequence Alignment, что делает ее неприменимой для значительного числа молекул РНК.
- ▶ Общая точность уступает таковой у SOTA в других областях структурной биологии.
- ▶ Генеративный процесс теоретически некорректен.

MolBindDif-ba и MolBindDif-bb



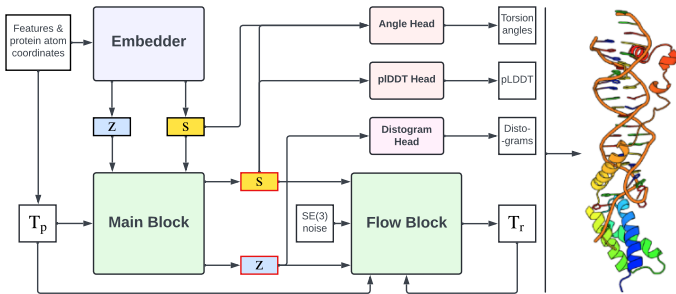
- ▶ Генеративный процесс - диффузия, Riemannian Score Matching.
- ▶ Вследствие независимости от MSA, имеют более широкое применение.
- ▶ Новое представление для РНК: нуклеотиды представлены как два твердых тела.

MolBindDif-ba и MolBindDif-bb

Model	rRMSD _{rr} , Å	rRMSD _{rp} , Å	IDDT _{rr}	IDDT _{rp}
MolBindDif-ba	10.2 ± 6.2	9.9 ± 7.4	0.18 ± 0.07	0.16 ± 0.07
MolBindDif-bb	10.7 ± 5.5	14.1 ± 7.8	0.17 ± 0.06	0.10 ± 0.05
RoseTTaFoldNA-short	13.3 ± 6.3	17.6 ± 7.8	0.19 ± 0.10	0.08 ± 0.04

- ▶ MolBindDif-ba и MolBindDif-bb существенно превосходят текущий SOTA (RoseTTaFoldNA) на ряде метрик.
- ▶ Новое представление РНК создает дополнительные ограничения.
- ▶ Генеративный процесс имеет теоретические недостатки.

MolBindFlow



- ▶ Генеративный процесс - flow matching на $SE(3)$.
- ▶ Имеет ряд технических преимуществ перед прошлыми моделями.
- ▶ Теоретический аппарат корректен и прост в использовании.

Полученные результаты

- ▶ Разработанные модели MolBindDif-ba и MolBindDif-bb имеют более широкое применение и результаты, превосходящие SOTA.
- ▶ Генеративные процессы в контексте группы SE(3) были адаптированы для молекулярно-структурной области.
- ▶ Модель MolBindFlow разрешает недостатки предыдущих моделей.

Дальнейшие исследования

- ▶ Усовершенствование модели MolBindFlow.
- ▶ Рассмотрение других генеративных задач молекулярной биологии.