База данных структурных мотивов белков: биологические и физико-химические свойства

Лицензия: CC BY-SA

Доступ: Доступен всем пользователям

Дата обновления: 02.02.2024

Теги:

ибмх, биоинформатика, аннотаця белков, промеотика, разработка лекарств

Описание

Набор данных создан с использованием технологий машинного обучения, нейронных сетей и содержит аннотации биологических и физико-химических свойств структурных мотивов белков. Аннотация включает параметры для: белка (название белка, название организма, код таксономической группы, биологический процесс и функция белка); структурного мотива белка (тип мотива, цепь в белке, локус и длина, аминокислотная последовательность, физико-химические свойства); эксперимента (источник структуры белка, разрешение структуры, экспериментальный метод и название эксперимента).

Набор данных предназначен для проведения структурного анализа белков, ассоциированных с развитием заболеваний, в том числе аберрантных форм белков, которые образованы вследствие аминокислотных замен, модифицирования после синтеза. Набор данных может применяться для решения прикладных медико-биологических задач, таких как разработка новых подходов к диагностике заболеваний, изучение молекулярных основ патогенеза, выявление мишеней белкового происхождения для лекарственных средств и проектирование миметиков (белков с заданными свойствами).

Набор данных содержит 3.96 млн аннотаций структурных мотивов в белковых структурах с указанием внутренних координат. Обновление версии набора данных выходит не реже одного раза в год. Набор доступен в формате CSV.

Характеристики датасета

Использование

Лицензия: CC BY-SA
Доступ: Доступен всем пользователям

Предоставление данных

Формат:
CSV

Характеристики набора

Количество наблюдений: 3960000

Обновления

Размещено: 02.02.2024
Обновлено: 02.02.2024
Частота обновления: Ежегодно
Дата следующего обновления: не указано

Происхождение

Источники данных: Институт биомедицинской химии имени В. Н. Ореховича

Структура набора данных

data_share (data_share)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
experiment_id	Идентификатор трехмерной структуры в Protein Data Bank	0			string
uniprot_id	Идентификатор белка в базе знаний UniProt	1330			string
protein_name	Название белка	101949			string
chain_id	Идентификатор белковой цепи в экспериментальной структуре	98257			string
resolution	Разрешение трехмерной структуры белка	2712626			numeric
molecular_function	Функция(и), которую выполняет белок в клетке/организме	2006228			string
experiment_type	Тип эксперимента, в котором получена трехмерная структура белка.	0			string
organism	Название организма, которому принадлежит белок (англ)	98257			string
biological_process	Биологический(е) процесс(ы), в которых участвует белок	2436633			string
fasta	Аминокислотная последовательность структурного мотива в формате однобуквенного кода	0			string
apfid	Уникальный идентификатор мотива, который содержит указание на экспериментальную структуру, позиции начала и конца мотива	0			string
motif_type	Тип структурного мотива	0			string
motif_len	Длина структурного мотива	0			integer
motif_start	Позиция начала структурного мотива в цепи белка	1567			integer
motif_end	Позиция конца структурного мотива в цепи белка	1567			integer

Описание

Набор данных создан с использованием технологий машинного обучения, нейронных сетей и содержит аннотации биологических и физико-химических свойств структурных мотивов белков. Аннотация включает параметры для: белка (название белка, название организма, код таксономической группы, биологический процесс и функция белка); структурного мотива белка (тип мотива, цепь в белке, локус и длина, аминокислотная последовательность, физико-химические свойства); эксперимента (источник структуры белка, разрешение структуры, экспериментальный метод и название эксперимента).