



УДК 57.012+57.016.4

АНАЛИЗ БЕЛОК-БЕЛКОВОГО ВЗАИМОДЕЙСТВИЯ CREBBP, НТТ и KMT2D МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

© 2025 г. И. И. Хегай*,[#], Ж. Гун*, В. М. Ефимов*

* ФИЦ “Институт цитологии и генетики” СО РАН, Россия, 630090 Новосибирск, просп. Акад. Лаврентьева, 10

Поступила в редакцию 29.04.2025 г.

После доработки 10.05.2025 г.

Принята к публикации 11.05.2025 г.

Транскрипционный коактиватор CREBBP (cyclic AMP response element binding protein-binding protein), нейроспецифический регулятор транскрипции НТТ (huntingtin) и гистон-лизинная метилтрансфераза KMT2D (lysine methyltransferase 2D) кооперативно участвуют в посттрансляционной модификации гистонов и регуляции дифференциальной экспрессии генов. Механизм межбелкового взаимодействия исследован слабо. Компьютерный анализ первичной структуры белков методом главных компонент выявил наличие максимально высокой корреляции между главными компонентами аминокислотных последовательностей в парах CREBBP–НТТ и CREBBP–KMT2D с глутамином. Траектория первой главной компоненты CREBBP практически совпадает с графиком позиционной частоты глутамин-содержания вдоль молекулы белка. Показано, что во вторичной структуре CREBBP существенную долю занимают элементы E-strand (extended strand) с открытой конформацией пептидной цепи. Полиглутаминовые тракты, локализованные на C-конце CREBBP, N-конце НТТ, N-, C-концах и в центре KMT2D также имеют открытую конформацию, облегчающую образование межмолекулярных водородных связей. Предполагается, что полиглутаминовые тракты C-конца CREBBP и N-конца НТТ непосредственно вовлечены в белок-белковый контакт CREBBP–НТТ. Аналогичная связь между полиглутаминовыми трактами C-конца CREBBP и центрального региона KMT2D фиксирует физическое взаимодействие в данной паре белков. Выявленные особенности исследованных белков могут быть использованы для конструирования новых фармакологических препаратов физико-химическими методами.

Ключевые слова: коактиватор CREBBP, нейроспецифический регулятор НТТ, метилтрансфераза KMT2D, полиглутаминовый тракт, открытая конформация, метод главных компонент

DOI: 10.7868/S1998286025060144

ВВЕДЕНИЕ

Белок-белковые взаимодействия опосредуют формирование и работу функционально активных комплексов сигнальных, транспортных и ферментативных белков в клетке и клеточном ядре. В зависимости от состава внутриядерных белков

трансактивационные комплексы оказывают кооперативные воздействия различной направленности на геномную экспрессию, регулируя транскрипцию множества генов метаболизма. Повсеместно экспрессируемый белок CREBBP выполняет функцию основного коактиватора транскрипционного

Сокращения: Ac-K – мотив с лизин-ацетилтрансферазной активностью (acetyl-lysine); AP-1 – активирующий белок-1, фактор транскрипции (activating protein-1); ASC-2 – сигнал-активирующий коинтегратор-2 (activating signal cointegrator-2); ASCOM – ASC-2-содержащий комплекс (ASC-2 containing complex); ATF – фактор, активирующий транскрипцию (activating transcription factor); BDNF – нейротрофический фактор мозга (brain-derived neurotrophic factor); CREB – цАМФ-чувствительный элемент-связывающий белок (cyclic AMP response element binding protein); CREBBP – CREB-связывающий белок (CREB-binding protein); E-strand – удлиненный филамент (extended strand); НТТ – хантингтин (huntingtin); KMT2D – лизин-метилтрансфераза гистона H3 (lysine methyltransferase 2D); MLL4 – белок 4 смешанно-клеточной лейкемии (mixed-lineage leukemia); PHD – гомеодомен растений (plant homeodomain); REST – фактор транскрипции, подавляющий репрессорный элемент-1 (repressor element-1 silencing transcription factor); SET – супрессор, энхансер, триторакс-домен (suppressor, enhancer, trithorax domain); SID – домен, взаимодействующий с SRC-1 (SRC-1 interacting domain); SOPMA – метод самооптимизированного прогнозирования с выравниванием (self-optimized prediction method with alignment); SRC-1 – коактиватор 1 стероидного рецептора (steroid receptor coactivator 1).

[#] Автор для связи: (тел.: +7 (383) 363-49-63-3503; эл. почта: khegay@bionet.nsc.ru).

фактора CREB [1, 2]. Белок CREBBP обладает лизин-ацетилтрансферазной активностью и участвует в модификации гистона H3. *N*-концевой участок гистона H3 не включается в состав нуклеосомного коры и выступает наружу, оставаясь доступным для модификации. Ацетилированные гистоны в промоторной области генов играют активирующую роль в регуляции транскрипции [3, 4]. Вместе с тем CREBBP – скаффолд для сборки ряда других транскрипционных факторов и ядерных белков в динамично регулируемые транскрипционные аппараты. Белки НТТ и KMT2D, локализующиеся в клеточном ядре, взаимодействуют с CREBBP в процессе модуляции структуры хроматина. НТТ экспрессируется в головном мозге и в ряде других органов экто- и энтодермального происхождения. Внутри клетки помимо ядра белок выявляется на внутриклеточных везикулах и микротрубочках, участвуя в транспорте из ядра в цитоплазму и далее в аксон [5]. Белок ядерного матрикса KMT2D входит в состав комплекса ASCOM, регулирующего экспрессию генов β -глобина гемоглобина, рецепторов эстрогена и тиреоидного гормона [6, 7]. Метилтрансферазная активность KMT2D воздействует на свободный *N*-концевой хвост гистона H3. Метилирование лизина гистона H3 также необходимо для активации транскрипции и представляет собой обратимый процесс [8]. Молекулярные механизмы белковых контактов преимущественно зависят от структуры боковых радикалов аминокислот

пептидной цепи. Физическая основа устойчивого контакта – формирование водородных связей между стерически совпадающими участками контактирующих молекул. Водородные связи стабилизируют возникающие при таком контакте кратковременные взаимодействия. Непосредственно в образовании водородных связей участвуют боковые радикалы полярных незаряженных аминокислот с серосодержащими, гидроксильными и амидными концевыми группами.

Цель данной работы – выявить в структуре CREBBP, НТТ и KMT2D ключевые элементы межмолекулярного взаимодействия белков в процессе посттрансляционной модификации гистонов, играющей важную роль в регуляции транскрипции генов в норме и патологических состояниях.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Первичные аминокислотные последовательности белков CREBBP, НТТ и KMT2D были обработаны согласно протоколу алгоритма Такенса и преобразованы в траекторные матрицы, на основании которых вычисляли главные компоненты. Данный алгоритм позволяет объединять и одновременно анализировать две и более различных аминокислотных последовательностей. На рис. 1 представлены коэффициенты корреляции главных компонент интегрированной первичной последовательности белков CREBBP и НТТ с отдельными аминокислотами. Наиболее высокий коэффициент корреляции с первой и второй глав-

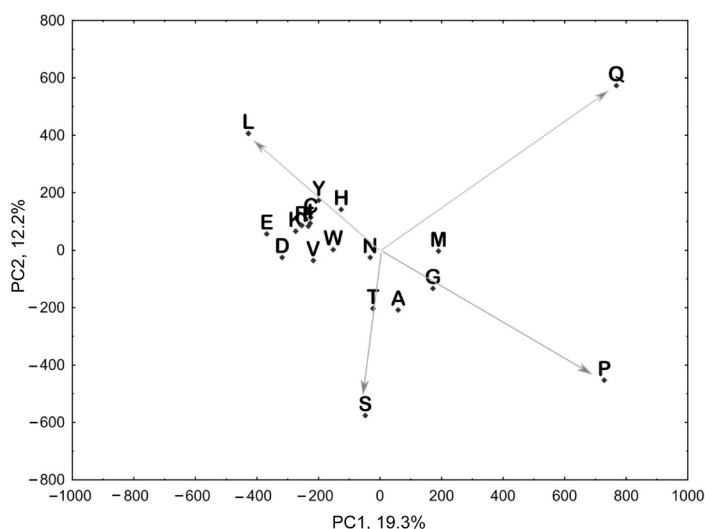


Рис. 1. Корреляция отдельных аминокислот с первой (PC1) и ортогональной второй (PC2) главной компонентой интегрированной последовательности белков CREBBP и НТТ. Коэффициенты корреляции приведены в формате $\times 1000$.

ными компонентами выявлен у глутамина. В координатах первой главной компоненты он составляет 0.743. Следующий по модулю коэффициент корреляции с первой главной компонентой определяется у пролина. Результаты аналогичного анализа, выполненного для белков CREBBP и KMT2D, показаны на рис. 2. Максимальный коэффициент корреляции с первой главной компонентой белков в данной паре также определяется у глутамина ($r = 0.924$); напротив, у пролина соответствующий показатель уходит в область отрицательных значений.

Белок CREBBP – скаффолд для НТТ и KMT2D. Анализ фазовой траектории первой главной компоненты аминокислотной последовательности непосредственно CREBBP и частоты глутамина в первичной структуре белка выявил практически полное совпадение динамики по всей длине молекулы (рис. 3). Коэффициент корреляции между параметрами составляет 0.795, что соответствует $p < 0.001$. Обе кривые имеют одинаковый максимум в C-концевой области. Совпадение максимумов свидетельствует о наличии в данной зоне белка структурных особенностей, связанных с глутамином. Анализ прогнозируемых элементов вторичной структуры белка CREBBP выполняли с использованием метода SOPMA. Показано, что большую часть полипептидной цепи занимает неупорядоченная укладка в форме случайных клубков (Random coil). При этом наряду с α -спиралями и β -цепями достаточно боль-

шую долю составляют слабо исследованные E-strand-элементы (Extended strand) (табл. 1). Еще одна важная особенность вторичной структуры CREBBP – представленность β -цепей исключительно в форме β -поворотов (Beta-turn). Компактные β -повороты изменяют направление главной цепи и обычно ограничивают с обеих сторон выступающие наружу компоненты вторичной структуры. Отсутствие β -мостов (Beta-bridge), служащих основанием для образования многоступенчатых β -листов, означает, что большая часть молекулы CREBBP вне α -спиралей представляет собой открытые конструкции с выпетливаниями в форме E-strand-элементов, фланкированные β -поворотами.

Известно, что ключевая аминокислота для образования β -поворотов – пролин [9]. Данные, представленные на рис. 1, свидетельствуют о том, что в паре CREBBP и НТТ пролиновые мотивы играют значительную роль. Можно предположить, что они участвуют в образовании множественных изгибов и выпетливаний, способных внедряться и действовать как якоря при наличии аналогичных выпетливаний на другой молекуле. Короткие петлевые участки, состоящие из ~14 а.о., в настоящее время рассматриваются как отдельные E-strand-элементы вторичной структуры [10]. Они не такие консервативные и регулярные по составу, как α -спирали и β -цепи, но вследствие ориентации наружу обладают свойством образовывать водородные связи между своими боковыми цепями

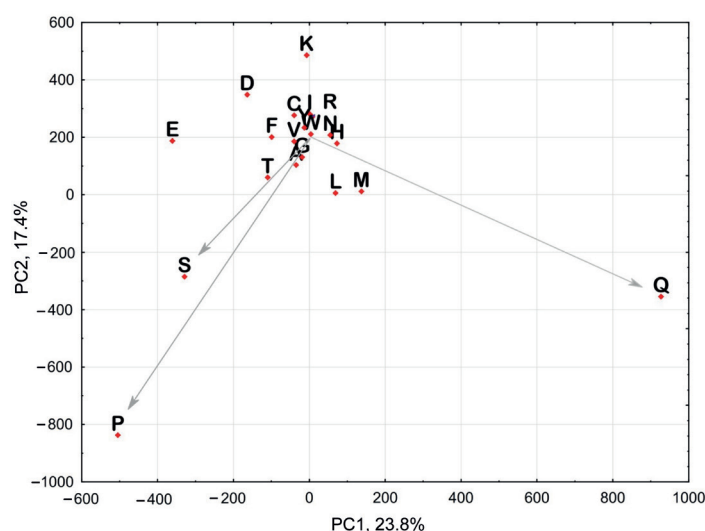


Рис. 2. Корреляция аминокислот с главными компонентами интегрированной последовательности белков CREBBP и KMT2D. Коэффициенты корреляции приведены в формате $\times 1000$, PC1 – первая главная компонента, PC2 – ортогональная вторая главная компонента.

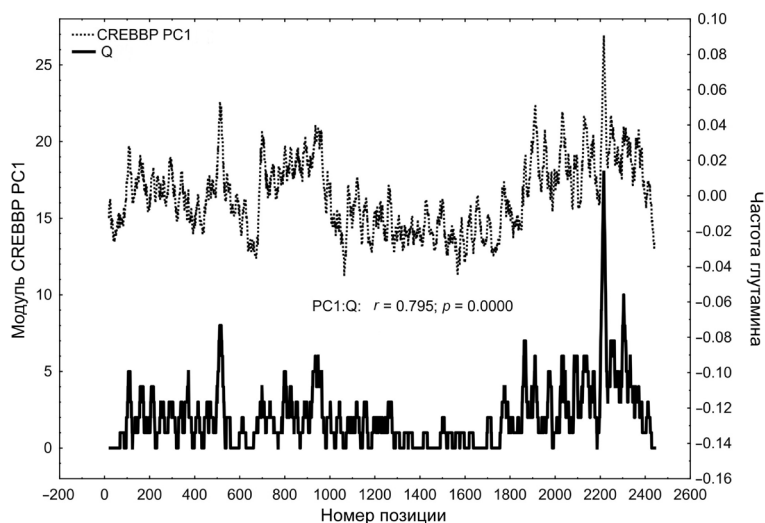


Рис. 3. Динамика позиционной изменчивости первой главной компоненты аминокислотной последовательности белка CREBBP и частоты глутамина.

Таблица 1. Прогнозируемый состав элементов вторичной структуры белка CREBBP

| Элемент вторичной структуры | Количество аминокислот | Доля от первичной структуры, % |
|-----------------------------|------------------------|--------------------------------|
| α -спираль | 744 | 30.47 |
| E-strand-элемент | 243 | 9.95 |
| β -мост | 0 | 0 |
| β -поворот | 157 | 6.43 |
| Случайный клубок | 1298 | 53.15 |

и амидными и карбонильными группами соседних молекул. Другая важная особенность E-strand-элементов заключается в том, что в выпрямленной конфигурации появляются доступные свободные атомы в карбонильных и амидных группах главной цепи, которые также могут образовывать встречные межмолекулярные водородные связи с радикалами полярных аминокислот соседних белков. Полярная аминокислота глутамин обладает длинной боковой цепью с концевым амидом. Длинные радикалы с повышенной вероятностью контактируют с внешними молекулами. Следовательно, присутствие глутамин в составе E-strand-элементов имеет особое значение для образования межмолекулярных водородных связей и стабилизации контактного взаимодействия белков. При совместном рассмотрении рис. 1 и 2 видно, что глутамин – наиболее важная аминокислота для функционирования CREBBP как скаффолд-белка для НТТ и KMT2D. Высокий коэффициент корреляции пролина на графике пары CREBBP–НТТ

может отображать факт наличия взаимодействующих E-strand-элементов в структуре CREBBP и НТТ. Напротив, низкий коэффициент корреляции пролина в паре CREBBP–KMT2D свидетельствует о существенно меньшем вкладе E-strand-элементов в характер взаимодействия белков, предположительно, вследствие значительной (более чем в 2.5 раза) разницы в длине пептидных цепей.

В структуре CREBBP можно выделить три функциональные зоны. N-концевая зона представляет собой трансактивационный домен, в котором расположены сайты взаимодействия с транскрипционным фактором CREB, активирующим фактором транскрипции ATF, гетеродимерным (c-Jun + c-Fos) транскрипционным фактором AP-1 и рядом других факторов, активирующих гены метаболизма. Среднюю треть молекулы занимает каталитический домен, где ключевым выступает Ас-К-мотив с лизин-ацетилтрансферазной активностью. Далее располагается C-концевой транс-

активационный домен, содержащий внутри себя домен SID, взаимодействующий с коактиватором ядерного стероидного рецептора 1 SRC-1 и тиреоидного рецептора [11]. На рис. 3 видно, что C-концевой домен характеризуется повышенным содержанием глутамина. Максимум соответствует полиглутаминовому тракту из 19 а.о. Полиглутаминовые тракты могут участвовать в образовании и фиксации белок-белковых контактов. Для белка в свободном состоянии характерны полиглутаминовые тракты, имеющие неактивную петлеобразную форму [12]. Свободный C-концевой полиглутаминовый тракт белка CREBBP представляет собой уникальный E-strand-элемент, обогащенный глутаминовыми радикалами и способный формировать множественные межмолекулярные водородные связи.

Биологическая функция белка НТТ в норме связана с регуляцией транскрипции в нервной ткани. НТТ стимулирует продукцию нейротрофического фактора BDNF, необходимого для нормального развития мозга. Хантингтин усиливает везикулярный транспорт BDNF вдоль микротубул аксона [13]. Непосредственный регулятор транскрипции нейрон-специфических генов – транскрипционный фактор REST [14]. REST взаимодействует с гистоновыми деацетилазами и подавляет экспрессию нейрональных генов. Белок НТТ связывается с REST и инактивирует репрессивное действие транскрипционного фактора. В самом начале первичной структуры НТТ расположен консервативный полиглутаминовый тракт. Размеры его в норме составляют 9–35 а.о. Мутантные формы НТТ характеризуются увеличенным размером полиглутаминового тракта – до 50 а.о. при умеренном развитии патологии и до 120 а.о. в случае максимально выраженной болезни Хантингтона [5]. В отличие от диффузной локализации нормального хантингтина, мутантные полиглутаминовые тракты имеют тенденцию образовывать в нервных клетках токсичные агрегаты. Скорость агрегации напрямую зависит от количества глутаминовых остатков [15]. Полиглутаминовые тракты формируют полярные застёжки-молнии, способствующие образованию межмолекулярных связей. В процессе нормального развития хантингтин взаимодействует с различными ядерными и моторными белками. Полиглутаминовый домен непосредственно участвует в образовании белок-белковой связи с транс-

крипционным коактиватором CREBBP. Со стороны CREBBP задействован аналогичный обогащенный глутамином C-конец белка [16]. Показано, что в основе патогенеза болезни Хантингтона лежит свойство мутантного НТТ связывать чрезмерное количество молекул CREBBP, создавая его дефицит и выводя коактиватор из ядра с образованием нейротоксичных мультимерных агрегатов.

Гистон-лизиновая метилтрансфераза KMT2D участвует в посттрансляционной модификации хроматина и относится к положительным регуляторам транскрипции генов. Ген *KMT2D* принадлежит к семейству генов смешанно-клеточной лейкемии, и у людей белок обозначается синонимом MLL4 [7]. В N-концевой области белка зафиксированы два тандема доменов PHD, выполняющих функцию распознавания сайтов метилирования гистона H3. Метилтрансферазная активность локализована в C-концевом SET-домене. Центральная часть молекулы KMT2D остается слабо исследованной [17]. В последовательности KMT2D человека выявлено 146 глутаминов, организованных в три полиглутаминовых тракта, расположенных, соответственно, в начале, середине и в C-концевой области белка [12]. Свободные полиглутаминовые тракты имеют открытую линейную структуру и могут образовывать ассоциации с различными ядерными белками и коактиваторами. В частности, KMT2D связывается с гистон-лизиновой ацетилтрансферазой CREBBP в процессе активации энхансеров и тканеспецифических транскрипционных факторов [18]. Каталитический домен с лизин-ацетилтрансферазной активностью занимает центр молекулы CREBBP. Соответственно, лизин-метилтрансферазный домен KMT2D расположен на C-конце. Физический контакт белков без нарушения функциональной активности ферментов возможен при взаимодействии C-концевого полиглутаминового тракта CREBBP с полиглутаминовым трактом средней части молекулы KMT2D. Показано, что свободные полиглутаминовые тракты способны объединять белки в комплексы, образуя взаимопроникающие спиральные выпетливания [19].

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Аминокислотная последовательность белка CREBBP (2442 а.о.) в формате FASTA была взята из GenBank NCBI (AAC51331.2). Аминокислот-

ные последовательности белков НТТ (3142 а.о.) и KMT2D (5537 а.о.) взяты из GenPept NCBI (код доступа NP_001375421.1 и NP_003473.3). Первичная структура белков исследована компьютерным методом главных компонент [20]. Выполнены три серии расчетов: только CREBBP, CREBBP в паре с НТТ и CREBBP в паре с KMT2D. На основании теоремы Такенса [21] исходную последовательность длины N нарежали на фрагменты скользящим окном ширины L . Позиция последней аминокислоты фрагмента обозначает номер фрагмента, первый фрагмент получает номер L , последний – N , всего, таким образом, $N - L + 1$ фрагментов. В каждой последующей серии окно сдвигается вправо на одну аминокислоту, и формируется новый ряд таблицы фрагментов. Совокупность фрагментов образует траекторную матрицу Такенса. Между всеми фрагментами вычисляется матрица квадратов евклидовых расстояний, которые однозначно определяют взаимное расположение фрагментов как точек в евклидовом пространстве. Методом главных координат (РСо) Гауэра [22] вычисляются координатные оси этого пространства. Проекция всех точек на эти оси образуют новые признаки, не коррелирующие между собой. Признаки упорядочиваются по убыванию дисперсии и представляют собой главные компоненты. Для их интерпретации используются коэффициенты корреляции главных компонент с частотами отдельных элементов последовательностей, в данном случае аминокислот. Оптимальная ширина скользящего окна для анализируемых последовательностей CREBBP, НТТ и KMT2D составила 18 а.о. В качестве расстояния между фрагментами выбрано хордовое расстояние Кавалли–Сфорца [23], по определению эквивалентное евклидовому.

Метод SOPMA использовали как инструмент для анализа вторичной структуры белка. Пакет программ SOPMA доступен на сайте <https://npsa-prabi.ibcp.fr/> в разделе “Методы”. На основе самооптимизированного прогнозирования с выравниванием первичной последовательности белка SOPMA рассчитывает его вторичную структуру [24].

Статистическую обработку полученных данных выполняли с использованием пакетов программ PAST 4 [25], Statistica 12 (StatSoft®), Jacobi 4 [26] и MS Excel 2016 (Microsoft®).

ЗАКЛЮЧЕНИЕ

Полиглутаминовые тракты – важные элементы белок-белковых связей между транскрипционными активаторами и коактиваторами. В основе образования устойчивых контактов лежит свойство бокового радикала глутамина образовывать водородные связи с электроотрицательными атомами других молекул. В случае полиглутаминовых трактов с открытой конформацией акцепторами водородной связи могут служить атомы главной цепи аминокислот. Полиглутаминовые тракты широко представлены в составе мультибелковых транскрипционных комплексов.

Результаты компьютерного анализа первичных последовательностей ядерных белков CREBBP, НТТ и KMT2D показали, что ключевая аминокислота межбелкового взаимодействия – глутамин. Взаимодействие транскрипционного коактиватора CREBBP с нейроспецифическим регулятором транскрипции НТТ происходит при непосредственном участии *C*-концевого полиглутаминового тракта CREBBP и *N*-концевой области НТТ, обогащенной глутамином. Аналогичная связь между CREBBP и гистон-лизиновой метилтрансферазой KMT2D может быть локализована по месту полиглутаминового тракта центрального региона KMT2D вследствие того, что *N*-концевая область KMT2D задействована в распознавании сайтов метилирования гистона H3, а *C*-концевая зона должна быть свободной для сохранения ферментативной активности. Полиглутаминовый тракт CREBBP представляет один из вариантов E-strand-элементов вторичной структуры, играющих важную роль в образовании межбелковых связей вследствие открытой конформации главной цепи.

Выявленные физико-химические особенности белков могут быть использованы для создания новых фармакологических препаратов, купирующих проявление нейродегенеративных заболеваний.

ФОНДОВАЯ ПОДДЕРЖКА

Работа выполнена при поддержке бюджетного проекта Института цитологии и генетики СО РАН FWNR-2022-0021, FWNR-2022-0019.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Статья не содержит описания исследований, выполненных с участием людей или использованием животных в качестве объектов. Информированное согласие не требовалось.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ВКЛАД АВТОРОВ

Автор ЖГ выполнила все компьютерные вычисления. Авторы ИИХ и ВМЕ в равной степени внесли вклад в статистический анализ и подготовку текста статьи.

ДОСТУПНОСТЬ ДАННЫХ

Данные, подтверждающие выводы настоящего исследования, можно получить у корреспондирующего автора по обоснованному запросу.

СПИСОК ЛИТЕРАТУРЫ

1. Goto N.K., Zor T., Martinez-Yamout M., Dyson H.J., Wright P.E. // J. Biol. Chem. 2002. V. 277. P. 43168–43174. <https://doi.org/10.1074/jbc.M207660200>
2. Zhu Y., Wang Z., Li Y., Peng H., Liu J., Zhang J., Xiao X. // Cancers. 2023. V. 15. P. 1219. <https://doi.org/10.3390/cancers15041219>
3. Vorobyeva N.E., Mazina M.Y. // Russ. J. Genetics. 2021. V. 57. P. 751–763. <https://doi.org/10.1134/S1022795421060144>
4. Liu R., Wu J., Guo H., Yao W., Li S., Lu Y., Jia Y., Liang X., Tang J., Zhang H. // MedComm. 2020. 2023. V. 4. P. e292. <https://doi.org/10.1002/mco2.292>
5. Schulte J., Littleton J.T. // Curr. Trends Neurol. 2011. V. 5. P. 65–78. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3237673/>
6. Lee J., Kim D.-H., Lee S., Yang Q.-H., Lee D.K., Lee S.-K., Roeder R.G., Lee J.W. // PNAS. 2009. V. 106. P. 8513–8518. <https://doi.org/10.1073/pnas.0902873106>
7. Blawski R., Vokshi B.H., Guo X., Kittane S., Salaku M., Chen W., Gjyzari M., Cheung T., Zhang Y., Simpkins C., Zhou W., Kulick A., Zhao P., Wei M., Shivashankar P., Prioleau T., Razavi P., Koche R., Rebecca V.W., de Stanchina E., Castel P., Chan H.M., Scaltriti M., Cocco E., Ji H., Luo M., Toska E. // Cell Rep. 2024. V. 43. P. 114174. <https://doi.org/10.1016/j.celrep.2024.114174>
8. Knorre D.G., Kudryashova N.V., Godovikova T.S. // Acta Naturae. 2009. V. 1. P. 29–51. <https://doi.org/10.32607/20758251-2009-1-3-29-51>
9. Marcelino A.M.C., Gierasch L.M. // Biopolymers. 2008. V. 89. P. 380–391. <https://doi.org/10.1002/bip.20960>
10. Eswar N., Ramakrishnan C., Srinivasan N. // Protein Engineering, Design and Selection. 2003. V. 16. P. 331–339. <https://doi.org/10.1093/protein/gzg046>
11. Valor L.M., Viosca J., Lopez-Atalaya J.P., Barco A. // Curr. Pharm. Des. 2013. V. 19. P. 5051–5064. <https://doi.org/10.2174/13816128113199990382>
12. Vaglietti S., Ferdinando Fiumara F. // NAR Genomics and Bioinformatics. 2021. V. 3. P. lqab032. <https://doi.org/10.1093/nargab/lqab032>
13. Gauthier L.R., Charrin B.C., Borrell-Pagès M., Dompierre J.P., Rangone H., Cordelières F.P., Mey J.D., MacDonald M.E., Lessmann V., Humbert S., Saudou F. // Cell. 2004. V. 118. P. 127–138. <https://doi.org/10.1016/j.cell.2004.06.018>
14. Butler-Ryan R., Wood I.C. // Metab. Brain Dis. 2021. V. 36. P. 1135–1150. <https://doi.org/10.1007/s11011-021-00719-2>
15. Hatters D.M. // IUBMB Life. 2008. V. 60. P. 724–728. <https://doi.org/10.1002/iub.111>
16. Nucifora F.C., Sasaki M., Peters M.F., Huang H., Cooper J.K., Yamada M., Takahashi H., Tsuji S., Troncoso J., Dawson V.L., Dawson T.M., Ross C.A. // Science. 2001. V. 291. P. 2423–2428. <https://doi.org/10.1126/science.1056784>
17. Froimchuk E., Jang Y., Ge K. // Gene. 2017. V. 627. P. 337–342. <https://doi.org/10.1016/j.gene.2017.06.056>
18. Lai B., Lee J.E., Jang Y., Wang L., Peng W., Ge K. // Nucleic Acids Res. 2017. V. 45. P. 6388–6403. <https://doi.org/10.1093/nar/gkx234>
19. Schaefer M.H., Wanker E.E., Andrade-Navarro M.A. // Nucleic Acids Res. 2012. V. 40. P. 4273–4287. <https://doi.org/10.1093/nar/gks011>
20. Ефимов В.М., Ефимов К.В., Ковалева В.Ю. // Вавиловский журн. генетики и селекции. 2019. Т. 23. С. 1032–1036. <https://doi.org/10.18699/VJ19.584>
21. Takens F. // Dynam. System. Turbulence, Lecture Notes Mathematic. 1981. V. 898. P. 366–381. <https://doi.org/10.1007/BFb0091924>
22. Gower J.C. // Biometrika. 1966. V. 53. P. 325–338. <https://doi.org/10.1093/biomet/53.3.4.325>
23. Cavalli-Sforza L.L., Menozzi P., Piazza A. // J. Asian Studies. 1995. V. 54. P. 2173–2219. <https://doi.org/10.2307/2058750>
24. Geourjon C., Deléage G. // Comput. Appl. Biosci. 1995. V. 11. P. 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>
25. Hammer Ø., Harper D.A., Ryan P.D. // Palaeontologia Electronica. 2001. V. 4. P. 1–9. https://palaeo-electronica.org/2001_1/past/issue1_01.htm
26. Polunin D., Shtager I., Efimov V. // bioRxiv. 2019. P. 803684. <https://doi.org/10.1101/803684>

Analysis of the Protein-Protein Interaction of CREBBP, HTT, and KMT2D by Principal Components Method

I. I. Khagai^{*,#}, R. Gong^{*}, and V. M. Efimov^{*}

[#]*E-mail: khagai@bionet.nsc.ru*

^{*} *Institute of Cytology and Genetics Federal Research Center, Siberian Branch, Russian Academy of Sciences, prosp. Acad. Lavrentieva 10, Novosibirsk, 630090 Russia*

The transcriptional coactivator CREBBP (cyclic AMP response element binding protein-binding protein), neuron-specific transcriptional regulator HTT (huntingtin) and histone-lysine methyltransferase KMT2D (lysine methyltransferase 2D) cooperatively participate in post-translational modification of histones and regulation of differential gene expression. The mechanism of protein-protein interactions has been poorly studied. Computer analysis of the primary structure of proteins using the principal components method revealed the presence of the highest possible correlation between the main components of amino acid sequences in the CREBBP–HTT and CREBBP–KMT2D pairs with glutamine. The trajectory of the first principal component of CREBBP practically coincides with the graph of the positional frequency of glutamine along the protein molecule. It is shown that in the secondary structure of CREBBP a significant share is occupied by E-strand (extended strand) elements with an open conformation of the peptide chain. Polyglutamine tracts localized at the C-terminus of CREBBP, N'-terminus of HTT, N-, C-termini and in the center of KMT2D also have an open conformation facilitating the formation of intermolecular hydrogen bonds. It is assumed that the polyglutamine tracts of the C-terminus of CREBBP and the N-terminus of HTT are directly involved in the protein-protein contact of CREBBP–HTT. A similar connection between the polyglutamine tracts of the C-terminus of CREBBP and the central region of KMT2D fixes the physical interaction in this pair of proteins. The identified features of the studied proteins can be used to design new pharmacological drugs using physicochemical methods.

Keywords: coactivator CREBBP, neuron-specific regulator HTT, methyltransferase KMT2D, polyglutamine tract, open conformation, principal components method