

на правах рукописи

УДК 577.322.23: 577.112.7

АНАШКИНА АНАСТАСИЯ АНДРЕЕВНА

Стерическая избирательность белок-белковых и
белок-нуклеиновых контактов

03.00.02. – биофизика

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва 2008

Работа выполнена в Учреждении Российской академии наук
Институте молекулярной биологии им. В. А. Энгельгардта РАН и
на кафедре Молекулярной биофизики Факультета молекулярной и
биологической физики Московского физико-технического
института (государственного университета).

Научный руководитель:

доктор физико-математических наук, профессор

Туманян Владимир Гайевич

Официальные оппоненты:

доктор физико-математических наук

Намиот Владимир Абрамович

доктор физико-математических наук

Нечипуренко Юрий Дмитриевич

Ведущая организация:

Институт теоретической и экспериментальной биофизики РАН, г.
Пушино

Защита состоится _____ декабря 2008 г. в _____ час. _____ мин. на заседании
диссертационного совета Д 501.001.96 при Московском
государственном университете им. М.В. Ломоносова, Москва
(Россия 119991, Москва, Ленинские горы 1, корп. 12, МГУ,
Биологический факультет, кафедра биофизики).

С диссертацией можно ознакомиться в библиотеке Московского
государственного университета им. М.В. Ломоносова.

Автореферат разослан «__» _____ ноября _____ 2008 г.

Ученый секретарь диссертационного Совета

доктор биологических наук, профессор

Кренделёва Т.Е.

Общая характеристика работы

Введение. Актуальность проблемы

Общепринято, что физико-химические свойства макромолекул определяются последовательностью их химических единиц. Знание принципов взаимодействия биологических макромолекул, понимание связи между последовательностью аминокислот, строением и взаимодействием белков дает возможность для развития фундаментальных подходов, а также служит основой для компьютерного моделирования взаимодействий белковых молекул, что, в свою очередь, может существенно изменить процесс создания новых лекарственных средств. Адекватный способ моделирования взаимодействий между белковыми молекулами позволит существенно сократить временные и финансовые затраты, связанные с исследованием свойств новых препаратов.

Белок-белковые взаимодействия являются решающими практически во всех основных биологических процессах, таких как клеточная регуляция, пути биосинтеза и распада, передача сигнала, инициация репликации ДНК, транскрипция и трансляция, образование олигомеров и мультимолекулярных комплексов, упаковка вирусов и иммунный ответ. Благодаря своей важности взаимодействия белков - предмет многих исследований. Однако расшифровка природы взаимодействий белков является достаточно трудной задачей. С физико-химической точки зрения имеет место сложный баланс относительного вклада различных взаимодействий, и большое разнообразие мотивов, составляющих область взаимодействия.

Проблема определения контактов между аминокислотными остатками является довольно сложной и требует усовершенствованных методов анализа. В большинстве случаев эффекты зависят от параметров, используемых при вычислении контактов. Поэтому желательно, если это возможно, применение непараметрического метода. Разбиение Вороного-Делоне – классический непараметрический метод, позволяющий определять прямые контакты между атомами, и, следовательно, между аминокислотами. Разбиение Вороного-Делоне уже давно положительно зарекомендовало себя в структурной биологии. С помощью этого разбиения исследовали плотность упаковки атомов в белках, средние объемы аминокислот, объемы атомов на поверхности

белков, а также средние объемы атомов в пространственных структурах большого числа неорганических соединений и белков.

Предсказывать возможные области взаимодействия между двумя данными белками, а также конкретные аспекты таких взаимодействий в настоящее время возможно лишь с ограниченной точностью и большими временными затратами. Структуры белков состоят из большого числа атомов, остатки аминокислот зачастую довольно подвижны, и имеют много степеней свободы, а потому процесс реконструкции взаимодействия сопряжен с большой трудоемкостью в вычислительном смысле. Даже возрастающие с каждым годом мощности суперкомпьютеров и кластеров не позволяют провести такой анализ для всего банка белковых структур. В связи с этим встает задача разработки новых подходов к моделированию взаимодействий белковых цепей и конструированию областей взаимодействия, а для этого необходимо знать характерные особенности таких взаимодействий.

Вопрос о том, какие взаимодействия играют ключевую роль в формировании белковых комплексов, в том числе ван-дер-ваальсовы взаимодействия, солевые мостики, водородные связи, является самым существенным вопросом в понимании белок-белковых ассоциаций. И хотя этот вопрос поднимается в большом числе исследований, окончательный ответ еще не ясен. Ответ на этот вопрос должен не только пролить свет на механизм белкового узнавания, но и позволит создать базу для конструирования эффективно связывающихся лигандов.

Цель и задачи исследования

Главной целью проведённой работы было установление статистических закономерностей, характерных для контактов аминокислот во взаимодействиях между белковыми молекулами и между молекулами белка и ДНК. В задачи работы входило:

- разработка метода оценки поверхностей контакта белок-белковых и белок-нуклеиновых комплексов на основе разбиения Вороного-Делоне и создание комплекса программ, позволяющего выявлять и анализировать контакты внутри биологических макромолекул и между ними, начиная с атомного уровня организации;
- создание невырожденных выборок белок-белковых интерфейсов, домен-доменных интерфейсов и белковых цепей, а также белок-нуклеиновых комплексов;

- построение разбиения Вороного-Делоне для каждого белка/комплекса выборки и определение контактов, начиная с атомного уровня;
- разработка модели случайно взаимодействующих аминокислотных остатков (остатков и нуклеотидов) на поверхности области взаимодействия и установление свойств модели;
- определение пар аминокислотных остатков (остатков и нуклеотидов), взаимодействующих чаще или реже, чем можно ожидать из модели случайных взаимодействий.

Научная новизна

Впервые для определения контактирующих атомов использовано разбиение Вороного-Делоне. Предложен и использован алгоритм, время работы которого пропорционально числу рассматриваемых атомов, что позволяет применить построение Вороного-Делоне к структурам белковых и белок-нуклеиновых комплексов. Показана устойчивость метода к достаточно большим вариациям в координатах атомов, т.е. адекватность результатов для белков, определенных с достаточно низким разрешением РСА.

Впервые проанализированы выборки максимально возможного объема различных типов белок-белковых и белок-нуклеиновых интерфейсов. Контакты определены на атомарном уровне, на основе этих данных установлены контакты между химическими единицами.

Построены модели случайного и специфического контактирования аминокислот на поверхности белок-белкового или белок-нуклеинового взаимодействия. На основе этих моделей определены предпочтения для взаимодействия аминокислотных остатков на поверхности контактов в белок-белковых или белок-нуклеиновых комплексах.

Впервые показано, что наибольшая предпочтительность наблюдается для взаимодействий между двумя цистеинами, как на поверхности взаимодействия двух идентичных белковых цепей (гомодимеров), так и на поверхности взаимодействия разных белковых цепей (гетерокомплексов), а также для взаимодействий между идентичными и разными белковыми доменами.

Показано, что для всех типов белок-белкового взаимодействия, а также для взаимодействия между аминокислотами в рамках одной белковой цепи важную роль играют противоположно заряженные пары аминокислот, демонстрирующие высокий уровень предпочтения.

Установлены специфические (предпочтительные) взаимодействия между аминокислотами и нуклеотидами. Показано, что с цитозином специфически связываются отрицательно заряженные аминокислоты – аспарагиновая и глутаминовая кислоты.

Практическое значение работы

Полученные в результате проведенных исследований данные по аминокислотным предпочтениям могут представлять ценность при создании системы предсказания областей белок-белкового и белок-нуклеинового узнавания и связывания, исследовании пространственных структур фибриллярных и глобулярных белков, при разработке новых биологически активных веществ с заданными свойствами. Кроме того, они вносят вклад в понимание механизмов формирования структуры белковых молекул. Использовать такой материал можно будет в различных прикладных областях, в том числе в области медицины и фармакологии.

Апробация работы и публикации

Основные результаты диссертации были представлены на XLIV научной конференции МФТИ (2001 г.), на 13-й и 14-й Международных зимних молодежных научных школах «Перспективные направления физико-химической биологии и биотехнологии», на VIII международной конференции: образование, экология, экономика, информатика (Астрахань, 2003), на III съезде биофизиков России (Воронеж, 2004), на XII международной конференции "Математика. Компьютер. Образование" (Пушино, 2005), на Международной московской конференции по вычислительной молекулярной биологии (МССМВ'05), на семинаре Института проблем управления им. В.А.Трапезникова "Экспертные оценки и анализ данных" 24 мая 2006 г, на совместном собрании 52-й Конференции американского биофизического общества и 16-го Международного биофизического конгресса (Лонг Бич, 2008).

По материалам диссертации опубликовано 11 печатных работ, в том числе 2 статьи в реферируемых журналах из списка ВАК, 3 в материалах международных и российских конференций, 6 в тезисах международных и российских конференций.

Объём и структура диссертации

Диссертация изложена на 95 страницах, иллюстрирована 22 рисунками и содержит 27 таблиц, список литературы включает 145

ссылок. Диссертация состоит из введения, шести глав, включая обзор литературы, выводов, списка цитированной литературы и приложения.

Краткое содержание работы

Введение

Введение содержит обоснование актуальности темы диссертации, ее научной новизны и практической значимости, приведены положения, выносимые на защиту.

Глава 1. Обзор литературы

В обзоре литературы рассматриваются:

- классификация белок-белковых комплексов и областей взаимодействия; роль гидрофобных и гидрофильных взаимодействий в процессе сворачивания белка и белок-белковом узнавании; результаты изучения интерфейсов белок-белковых комплексов и аминокислотных предпочтений другими авторами;
- специфическое и неспецифическое связывание нуклеиновых кислот белком; вклад различных факторов во взаимодействия между белком и ДНК: водородные связи, ионные пары, Ван дер Ваальсовы взаимодействия; опосредованные водой контакты; гидрофобный эффект; взаимные конформационные перестройки.

В обзоре литературы подробно анализируются работы по исследованию пространственной структуры сайтов связывания и взаимодействию между аминокислотами на поверхности сайтов связывания, а также аминокислотному составу сайтов связывания. Работы по определению аминокислотного состава и предпочтениям сделаны, как правило, на небольших выборках, от 70 до 170 белок-белковых комплексов. Определения предпочтения сделаны на разных типах мест связывания или внутренних местах связывания, и сравнивать или комбинировать результаты таких исследований оказывается затруднительным. Следует отметить, что во всех без исключения рассмотренных работах контактирующие атомы определялись по критерию расстояния, т.е. если на расстоянии, например, 5 Å от данного атома находится атом из другой белковой цепи, данный атом считается участвующим во взаимодействии с ним.

Многочисленные исследователи, изучая группу гомологичных белков, определяют важные для распознавания и связывания аминокислоты в рамках данной группы, однако общих, универсальных

правил или «кода» распознавания для всех белков по-прежнему не найдено.

Делается вывод о необходимости развития новых подходов к исследованию предпочтений взаимодействия аминокислот на поверхности белок-белковых комплексов и белок-нуклеиновых комплексов.

Глава 2. Методы

В первой части главы содержится описание разбиения Вороного-Делоне, обсуждаются возможности метода, дается сравнение с методом определения контактов по расстоянию, а также показана устойчивость результатов к варьированию координат. Вторая часть главы посвящена моделям случайного и неслучайного типов контактов на поверхности белок-белковых взаимодействий. В третьей части главы описываются использованные в данной работе статистические методы и понятия, в числе последних - предлагаемый индекс представленности контактов.

Разбиение Вороного-Делоне

Подробно описывается разработанная методика построения разбиения Вороного-Делоне для белков и нуклеиновых кислот. Приводится алгоритм программы, обсуждаются особенности и тонкости построения разбиения. Достаточно подробно описано применение метода для определения параметров атомов и аминокислот, таких как объем и площадь поверхности. В этом разделе также приводятся значения определяемых величин из литературных источников и их сравнение с данными, полученными в этом исследовании. Проведено сравнение разбиения Вороного-Делоне и метода определения контактов по расстоянию. Подробно описано исследование устойчивости к варьированию координат.

В качестве исходной информации используются пространственные координаты всех атомов белковой молекулы. В разбиении Вороного-Делоне пространство внутри белковой глобулы распределяется между всеми ее атомами по следующему принципу: разделяющая плоскость проводится между двумя соседними атомами через середину отрезка, соединяющего эти атомы и перпендикулярно ему. Такие плоскости образуют вокруг каждого атома выпуклый многогранник произвольного вида, называемый полиэдром Вороного. Область внутри многогранника лежит ближе к данному атому, чем к любому другому. Таким образом, контакт между двумя атомами существует, если у этих атомов есть

общая грань полиэдра Вороного с площадью, отличной от нуля. Следовательно, контакт между двумя аминокислотами определяется как совокупность общих граней полиэдров Вороного составляющих их атомов.

Применение разбиения Вороного-Делоне к структурам белковых и белок-нуклеиновых комплексов стало возможным благодаря принципиально новому алгоритму вычислений. Все ранее применяемые методы использовали алгоритмы, в которых зависимость времени работы от числа рассматриваемых точек N была пропорциональна N^3 , в лучших алгоритмах $N^2 \ln N$. В нашей работе предложен и использован алгоритм, время работы которого пропорционально числу точек N в рассматриваемом ансамбле.

С помощью вычислений демонстрируется замечательная устойчивость разбиения Вороного-Делоне. По этому свойству метод Вороного-Делоне значительно превосходит существующие методы определения контактов по расстоянию. Варьирование каждой координаты атомов возможно до 1 Å в любом направлении без потери значимых контактов. Кроме того, в рамках данного метода можно оценивать значимость каждого контакта по размеру площади и расстоянию между атомами.

Статистические модели случайно и специфически контактирующих аминокислот

Для полноценной интерпретации полученных данных, для того, чтобы оценить и выявить отклонения от случайных процессов, была разработана математическая статистическая модель контактирующих аминокислот: два круга бросают на некоторую область случайным образом, и каждый раз фиксируют площадь перекрывания. Выразим плотность вероятности $\frac{dP}{dS}$ случайных пересечений как зависимость от площади пересечения S в параметрическом виде:

$$\left\{ \begin{array}{l} \frac{dP}{dS} = \frac{\pi L}{r(R-2r)^2 \sqrt{1 - \frac{L^2}{4r^2}}}, \\ S(L) = 2(r^2 \arcsin(\sqrt{1 - \frac{L^2}{4r^2}}) - \frac{Lr}{2} \sqrt{1 - \frac{L^2}{4r^2}}). \end{array} \right. \quad (1)$$

В случае специфических взаимодействий можно предположить, что контакты обладают некоторой, отличной от нуля средней площадью контакта, обусловленной физико-химической природой взаимодействия

остатков. В этом случае распределение расстояний между центрами кругов подчиняется нормальному распределению. Выразим распределение площадей неслучайных контактов также в параметрическом виде:

$$\begin{cases} f(L) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(L-a)^2}{2\sigma^2}}, \\ S(L) = 2(r^2 \arcsin(\sqrt{1-\frac{L^2}{4r^2}}) - \frac{Lr}{2} \sqrt{1-\frac{L^2}{4r^2}}). \end{cases} \quad (2)$$

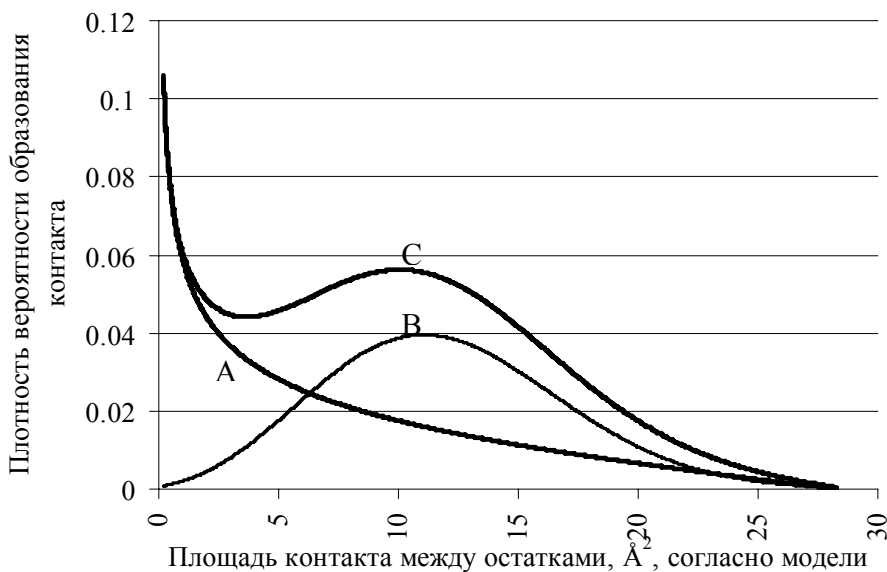


Рисунок 1. Графики, отражающие системы (1) и (2), моделируют распределения площадей случайных (А) и неслучайных (В) контактов. (А) График системы (1) в параметрической форме. График отражает распределение площадей случайных контактов. (В) График системы (2) в параметрической форме. График отражает распределение площадей неслучайных контактов. (С) Сумма графиков (А) и (В).

Зависимость (1) показана на Рис.1, кривая А. По мере увеличения площади контакта число контактов резко уменьшается. Другими словами, распределение для случайных контактов показывает наличие большого числа малых по площади контактов. Зависимость (2) показана ниже на Рис.1, кривая В. Кривая имеет колоколообразную форму, несимметричную, с некоторым, существенно отличным от нуля средним значением. Распределение для неслучайных контактов отражает существование некоторой характерной площади контакта.

Статистические методы

Рассматриваются использованные в данной работе статистические понятия и методы, в том числе методы оценки статистической

значимости, влияние размера выборки, сравнение наборов данных, коэффициент корреляции наборов данных, метод оценки ошибок и индекс представленности контактов. Оценивается минимальный размер выборки, необходимый для того, чтобы оцениваемый параметр находился в доверительном интервале заданной длины с требуемой точностью. Для оценки аминокислотного состава интерфейсов с точностью 2% в доверительном интервале 95% минимальное количество исследованных интерфейсов должно быть около 460.

Индекс представленности контакта

В качестве величины, отражающей соотношение между числом наблюдаемых и ожидаемых контактов, был выбран индекс представленности контакта g_{ij} , как отношение числа наблюдаемых контактов c_{ij} (вычисленных по разбиению Вороного; два остатка считаются взаимодействующими, если два атома имеют общую грань с ненулевой площадью поверхности) и ожидаемых значений, определенных для таблиц сопряженности:

$$g_{ij} = \frac{c_{ij}}{p_{ij} * N}, \quad (3)$$

где $p_{ij} * N$ - ожидаемые частоты для таблиц контактов.

Вероятность случайного контакта p_{ij} между остатками i и j типов:

$$p_{ij} = \frac{n_i}{N} * \frac{m_j}{N}, \quad (4)$$

где общее число контактов $N = \sum_i \sum_j c_{ij}$, а $n_i = \sum_j c_{ij}$, $m_j = \sum_i c_{ij}$.

Эта величина (3) отражает соотношение между наблюдаемым числом контактов и ожидаемым. Если ожидаемое число контактов близко к наблюдаемому, g_{ij} принимает значение, близкое 1. g_{ij} меньше 1, если ожидаемое число контактов больше наблюдаемого, и больше 1, если ожидаемое число контактов меньше наблюдаемого.

Глава 3. Исследование белок-белковых взаимодействий

Представленные в этой Главе результаты относятся к изучению контактов между аминокислотами в белок-белковых интерфейсах гомодимеров и гетерокомплексов.

Выборка белок-белковых комплексов

Для изучения закономерностей, существующих в белок-белковых интерфейсах, использовалась репрезентативная выборка из 4602 белок-

белковых интерфейсов из работы (Mintz, Shulman-Peleg et al. 2005). Данная выборка содержит 3067 интерфейсов, образованных идентичными цепями (гомодимеры) и 1535 интерфейсов, образованных различными цепями (гетерокомплексы).

Статистика атом-атомных контактов

Основываясь на списке контактов между атомами, было построено распределение расстояний между контактирующими атомами, распределение площадей межатомных контактов и корреляция между расстоянием и площадью контакта взаимодействующих атомов. Всего в данной выборке исследовано 2 057 304 межатомных контактов.

Аминокислотный состав белок-белковых интерфейсов

В этом разделе приводятся полученные данные по аминокислотному составу интерфейсов, а также подробное сравнение с результатами, полученными другими авторами.

Контакты между аминокислотными остатками

Полный список исследованных взаимодействий содержит 4602 интерфейса, 421 956 контактов между аминокислотами и 2 057 304 контакта между атомами.

Также в данном разделе рассматривается распределение площадей контакта между аминокислотами, взаимосвязь между площадью интерфейса и количеством участвующих в его формировании аминокислот, а также исследование площадей интерфейсов. В данном разделе также приводятся таблицы чисел контактов между остатками типов i и j для гомодимеров и гетерокомплексов. В обоих случаях очевидна «склонность» к образованию контактов между противоположно заряженными аминокислотами (около 24% всех контактов) и между гидрофобными остатками (Leu, Val, Ile, Phe, Ala) (около 33% всех контактов).

Предпочтения в контактах между остатками

В этом разделе подробно обсуждаются индексы представленности (3), вычисленные отдельно для гомодимеров и гетерокомплексов, сравнение их между собой и с литературными данными.

Получены следующие результаты:

- статистически значимы индексы представленности для контактов между двумя цистеинами, как для гомодимеров, так и для

гетерокомплексов. Для гомодимеров это значение g_{ij} составляет 4.85, а для гетерокомплексов 3.48, что свидетельствует о максимальной предпочтительности такого контакта;

- следующими по величине индекса представленности g_{ij} идут контакты между остатками противоположного знака. Это характерно как для гомодимеров, так и для гетерокомплексов; отклонение от 1 также статистически значимо.

- гидрофобные остатки образуют парные контакты со значениями индекса представленности, хотя и больше 1, но не имеющими строгой статистической значимости;

- самые низкие значения индексов представленности наблюдаются для взаимодействий одинаково заряженных аминокислот;

- индексы представленности для контактов между остатками одного типа в случае взаимодействия идентичных цепей выше, чем для гетерокомплексов. Этот эффект замечен и другими авторами. Следует отметить, что свойства интерфейсов, образованных двумя идентичными белковыми цепями, отличаются от свойств интерфейсов гетерокомплексов. В частности, первые содержат контакты, образованные симметричными парами аминокислот. В центральной области взаимодействия происходит пространственное сближение остатков, имеющих одинаковые номера в полипептидной цепи, тем самым сближаются остатки одного типа. Обнаружено 6600 таких контактов среди 167000 контактов между остатками в интерфейсах гомодимеров. В литературе также встречаются упоминания, что интерфейсы гомодимеров содержат больше контактов между аминокислотами одного типа благодаря оси симметрии второго порядка;

- приведенная гистограмма расстояний между атомами серы контактирующих остатков цистеина демонстрирует специфические взаимодействия между остатками цистеина. В области 2 Å наблюдается четко выраженный пик, представляющий собой межцепочечные дисульфидные мостики. Сравнение гистограмм площади для всех контактов остатков цистеина и для дисульфидных связей подтверждает гипотезу о том, что полное распределение состоит из распределения площадей случайных контактов, когда контакт происходит вследствие сближения, и распределения площадей специфических контактов, обусловленных физико-химическими взаимодействиями. Специфические контакты S-S, обусловленные формированием дисульфидной связи, имеют среднюю площадь взаимодействия 8,23 Å²;

- проведена оценка ошибок для величин g_{ij} с использованием метода «складного ножа». Диапазон значений ошибки от 0,03% до 0,18%, среднее значение 0,06%. Максимальная ошибка составляет 0,18% для контактов Gly-Pro (0,0027), а минимальная 0,03% для контактов Arg-Asp (0,0005) и Arg-Leu (0,0002).

Глава 4. Исследование междоменных взаимодействий в глобулярных белках

Данная глава посвящена исследованию взаимодействий доменов, определенных в базе данных доменов CATH. Рассматриваются контакты между тремя типами домен-доменных интерфейсов: образованных доменами, расположенными в разных белковых цепях и имеющими разную классификацию по CATH; интерфейсов, образованных одинаковыми по классификации CATH доменами, принадлежащими разным белковым цепям; интерфейсов, образованных доменами, расположенными в одной белковой цепи

Невырожденная подборка интерфейсов составила: 342 интерфейса, образованных разными доменами, расположенными в разных белковых цепях; 557 интерфейсов, образованных одинаковыми доменами, принадлежащими разным белковым цепям; 377 интерфейсов, образованных доменами, расположенными в одной белковой цепи.

Всего рассмотрено 274 656 атомных контактов, составляющих 56 843 контактов между аминокислотными остатками. Построено распределение расстояний между контактирующими атомами и площади межатомных контактов, корреляция между площадью контакта и расстоянием, взаимосвязь между площадью интерфейса и количеством участвующих в его формировании атомов.

Аминокислотный состав домен-доменных интерфейсов

Определен аминокислотный состав интерфейсов во взаимодействиях домен-домен. Заряженные аминокислоты Arg, Asp, Glu, Lys на поверхности домен-доменных интерфейсов составляют 24% от общего количества аминокислот. Гидрофобные аминокислоты Ala, Ile, Leu, Val и Phe составляют 32%. Ser и Thr 11%. Мало His (3%), Met(2%), Trp(2%) и Cys(1%).

Контакты между аминокислотными остатками

Построена гистограмма площади контакта между аминокислотами на поверхности домен-доменных интерфейсов, вычислена корреляция

между площадью области связывания и числом аминокислот. Среднее значение площади интерфейса получилось 400 \AA^2 , среднее число контактов между аминокислотными остатками на интерфейс 44.5, а средняя площадь контакта 8.6 \AA^2 .

Предпочтения в контактах между остатками

В этом разделе подробно обсуждаются числа контактов и индексы представленности, вычисленные отдельно для: 1) взаимодействий между доменами одной белковой цепи, 2) между разными доменами разных цепей и 3) между одинаковыми доменами, принадлежащими разным белковым цепям и сравнение их между собой.

Получены следующие результаты:

- во всех указанных выше случаях самые большие числа контактов наблюдаются для взаимодействий между противоположно заряженными аминокислотами (Arg, Lys – Asp, Glu) и между гидрофобными аминокислотами (Leu, Ala, Ile, Phe, Thr, Val). Однако, если говорить о сравнении межцепочечных и внутрицепочечных взаимодействий между доменами, то (если сравнивать относительное число контактов) взаимодействия между доменами, расположенными в разных белковых цепях, более богаты контактами между заряженными остатками, такими как Arg, Asp, Glu, Lys, а также Gln и Asn, и менее богаты гидрофобными взаимодействиями, такими как Leu, Ile, Phe, Val, Trp;
- следует отметить увеличение индекса представленности на главной диагонали матрицы индексов для контактов между идентичными доменами, расположенными в разных белковых цепях. Это усиление диагонали напоминает таковое для контактов между гомодимерами. Все рассуждения относительно контактов, образованных вследствие сближения остатков с одинаковыми номерами, и, следовательно, одного типа, применимы и в данном случае;
- значение индекса представленности контакта Cys-Cys для взаимодействия между одинаковыми доменами составляет 5,8, что является самым большим значением. Отметим, что для взаимодействий между разными доменами значение индекса составляет только 1,24 для одной цепи и 1,17 для доменов, расположенных в разных белковых цепях;
- статистически значимыми также можно считать отклонения для контактов между противоположно заряженными остатками: Arg-Asp, Arg-Glu, Lys-Asp, Lys-Glu для всех трех типов рассматриваемых интерфейсов. Также в эту категорию попадают контакты His-Asp и His-Glu для взаимодействий между доменами в одной белковой цепи;

- все индексы представленности для контактов между гидрофобными аминокислотными остатками лежат в «серой зоне»- величины от 1,20 до 1,58. Значение р-уровня для этих величин составляет около 0,28, что не позволяет сделать вывод о значимости отклонений;
- контакты между гидрофобными и заряженными остатками, за редкими исключениями, имеют величины индексов меньше единицы: от 0,53 до 0,94. Это свидетельствует о том, что контакты такого типа реализуются реже, чем можно было бы предположить из случайной модели;
- одинаково заряженные остатки также имеют величины индексов меньше 1, от 0,29 до 0,99. Это, также как и в случае контактов между гидрофобными и заряженными остатками, свидетельствует о том, что контакты такого типа реализуются реже, чем можно было бы предположить из случайной модели;
- следует отметить, что статистически значимыми являются индексы представленности контактов Cys-Lys (1,88) и Cys-Ser (1,73) для взаимодействий между разными доменами, принадлежащими разным белковым цепям. Также, на фоне усиленной центральной диагонали, отметим контакты His-His (1,74) и Met-Met (2,15) для взаимодействий идентичных доменов, расположенных в разных белковых цепях.
- проведена оценка ошибок для величин g_{ij} с использованием метода «складного ножа». Диапазон значений ошибки составил от 0,12% до 0,78%, среднее значение 0,28%. Максимальная ошибка составляет 0,78% для контактов Cys-Asn (0,0057), а минимальная 0,12% для контактов Leu-Val (0,0017).

Глава 5. Исследование взаимодействий белок-ДНК

В этой главе изложены результаты исследования взаимодействий белок-ДНК с помощью разбиения Вороного-Делоне.

Получены следующие результаты:

- построено распределение площадей контактов между аминокислотами и нуклеотидами ДНК;
- определен аминокислотный и нуклеотидный состав интерфейсов исследованных белок-нуклеиновых комплексов;
- получены таблицы чисел контактов между атомами аминокислот и нуклеотидов и суммарно на уровне аминокислота – нуклеотид (Таблица 1);
- приблизительно треть всех контактов составляют контакты с положительно заряженными аминокислотами Arg и Lys, 32,3%. Ser и

Thr дают следующий по величине вклад – 15%. Asp образует 6 % контактов, также как и Gly;

Таблица 1. Контакты и индексы представленности для взаимодействий аминокислотных остатков с нуклеотидами ДНК.

	Число контактов, c_{ij}					Индексы представленности, g_{ij}			
	A	T	G	C	Σ	A	T	G	C
ALA	1283	1456	1313	1268	5320	0,95	0,98	0,97	1,13
ARG	6092	6192	6463	4698	23445	1,02	0,94	1,08	0,95
ASN	2013	2284	1746	1499	7542	1,05	1,08	0,91	0,94
ASP	660	476	1162	804	3102	0,84	0,55	1,47	1,23
CYS	149	183	165	155	652	0,90	1,00	0,99	1,13
GLN	1561	1564	1349	1486	5960	1,03	0,94	0,89	1,18
GLU	869	893	1036	1091	3889	0,88	0,82	1,05	1,33
GLY	1933	2235	1772	1552	7492	1,02	1,06	0,93	0,98
HIS	953	1379	1014	627	3973	0,94	1,24	1,00	0,75
ILE	1090	1215	1011	827	4143	1,04	1,05	0,96	0,95
LEU	1069	1196	882	829	3976	1,06	1,07	0,87	0,99
LYS	4527	4654	4291	3630	17102	1,04	0,97	0,98	1,01
MET	383	599	545	359	1886	0,80	1,13	1,13	0,90
PHE	852	1144	953	645	3594	0,93	1,13	1,04	0,85
PRO	961	1016	712	701	3390	1,12	1,07	0,82	0,98
SER	2298	2667	2527	1862	9354	0,97	1,02	1,06	0,95
THR	2432	2831	2244	2003	9510	1,01	1,06	0,93	1,00
TRP	276	330	294	401	1301	0,83	0,90	0,89	1,46
TYR	1287	1480	1505	1157	5429	0,93	0,97	1,09	1,01
VAL	1256	1461	1053	863	4633	1,07	1,12	0,89	0,89

- получена таблица индексов представленности (Табл.1) контактов между аминокислотами и нуклеотидами ДНК. Согласно индексам представленности, можно выделить следующие предпочтения: для аденина – Pro; для тимина – His, а также Met, Phe, Val; для гуанина – Asp, с чуть меньшим предпочтением Met; для цитозина – значительный пик для Trp, Glu, Asp, а также для Gln, Cys, Ala.

- проведена оценка ошибок для величин g_{ij} с использованием метода «складного ножа» теста. Величина ошибки не превышает 0,3%. Диапазон значений ошибки составил от 0,0004 до 0,0034, среднее значение 0,0012. В процентном отношении средняя ошибка составляет

0,12%. Максимальная ошибка составляет 0,32% для контактов Cys-A (0,0029), а минимальная 0,04% для контактов Arg-T (0,0004).

Далее рассмотрено изменение индекса представленности при последовательном исключении малых контактов, имеющих площадь от 1 до 9 Å (Таблица 2).

Таблица 2. Основные результаты изменения индекса представленности при последовательном исключении малых по площади контактов.

	Увеличение индекса	Максимальное значение
A	Val, Pro	Val
T	His, Ala, Leu	His
G	Asp, Arg, Ile	Asp
C	Glu, Cys, Trp	Glu

Далее рассмотрены взаимодействия между белком и ДНК на уровне атомарных контактов. Представлены числа контактов между атомами белка и атомами ДНК. Все атомы белка рассортированы на две группы – атомы основной цепи и атомы остатков, а все атомы ДНК разделены на четыре группы – атомы сахара, фосфатной группы и нуклеинового основания, выходящие в малую или большую бороздку двойной спирали.

Получены следующие результаты:

Arg дает наибольший вклад – около 22% всех контактов наблюдаются между атомами аргинина и атомами нуклеотидов. Lys дает 13% всех атом-атомных контактов, тогда как Thr – 7%, Ser – 6,6%, Asn – 5,9%, Tyr – 5,8%. Вклад каждой из остальных аминокислот не превышает 5%.

Следует отметить, что контакты атомов бокового радикала Arg с атомами ДНК составляют более 90% всех контактов атомов Arg, из них 60% это контакты атомов бокового радикала Arg с сахарофосфатным остовом ДНК, и около 30% составляют контакты между атомами бокового радикала и атомами нуклеинового основания. Контакты атомов Arg с атомами нуклеотидов осуществляются как по малой, так и по большой бороздке спирали ДНК. Заметим, что с А взаимодействие происходит чуть чаще по малой бороздке, а с Т, G, С – преимущественно по большой бороздке. Lys, также как и Arg, преимущественно взаимодействует с сахарофосфатным остовом ДНК. Также наблюдаются контакты между атомами бокового радикала Lys и атомами большой бороздки нуклеинового основания ДНК, однако с G

их наблюдается в два раза больше, чем с А или Т, а с С приблизительно в два раза меньше. Взаимодействия Ser, Thr и Tyr главным образом касаются сахарофосфатного остова ДНК, однако также наблюдаются контакты между атомами этих аминокислот и атомами большой бороздки нуклеинового основания Т. Asn, как и Gln, образует контакты с сахарофосфатным остовом. Они также образуют контакты между атомами бокового радикала и атомами нуклеиновых оснований ДНК, причем контактов по большой бороздке приблизительно втрое больше. Следует отметить, что Phe взаимодействует как с сахарофосфатным остовом ДНК, так и с нуклеиновыми основаниями. Контакты с нуклеиновыми основаниями А и Т проходят преимущественно по малой бороздке, тогда как контактов Phe с G и C по большой бороздке наблюдается больше, чем по малой.

Теперь рассмотрим, какие атомные контакты характерны для взаимодействий из таблицы 2. Для Glu с С боковой радикал глутаминовой кислоты образует контакты с атомами нуклеинового основания, выходящими в большую бороздку, а также с атомами сахара. Для атомов бокового радикала Asp с G, в порядке убывания, характерны контакты с сахаром, фосфатной группой, большой бороздкой, малой бороздкой. Для His с Т – с сахаром, большой бороздкой, фосфатной группой, малой бороздкой. Для Val с А характерны контакты атомов бокового радикала с сахаром, малой бороздкой, фосфатной группой, большой бороздкой.

Далее рассматриваются примеры из литературы, когда отрицательно заряженные аминокислотные остатки важны для ДНК-белкового узнавания и связывания. Можно предположить, что связывание полианионной ДНК зависит от кластеров положительно заряженных аминокислот в сближенной с ДНК областью ДНК-связывающего белка. Однако такие кластеры положительно заряженных аминокислот были бы электростатически невыгодными без стабилизирующих взаимодействий с соответствующими отрицательно заряженными остатками. Вместе с тем отрицательно заряженные остатки создают возможность для последующей диссоциации белка и ДНК.

Глава 6. Обсуждение результатов

Выявление характерных аминокислотных контактов и особенностей взаимодействия аминокислот на поверхности белок-белковых сайтов связывания – это необходимые шаги на пути к

пониманию пространственной структуры белковых комплексов, а значит и к созданию системы компьютерного моделирования таких взаимодействий.

В этом исследовании продемонстрирована обоснованность применения разбиения Вороного-Делоне. Этот непараметрический метод позволяет по атомным координатам определить контакты между атомами и аминокислотами однозначно и математически строго. Математическая процедура была протестирована на небольших отклонениях в значениях координат атомов, и продемонстрировала стабильность результатов. Разбиение Вороного-Делоне показало замечательную устойчивость к ошибкам, даже при больших отклонениях в координатах (до 1 Å). Кроме того, данный метод позволил оценивать значимость каждого контакта по размеру площади и расстоянию.

Разработана статистическая математическая модель контактирующих аминокислот в следующем приближении: два круга бросают на некоторую область случайным образом или прицельно. Аналитически определена плотность вероятности случайных и неслучайных пересечений как зависимость от площади пересечения. На примере общего распределения площадей контактов и распределении площадей контактов между двумя остатками цистеина показана адекватность модели. В рамках модели можно оценить предпочтительность каждого типа контактов. Сравнение распределения площадей контактов и модельных распределений позволяет сделать вывод о случайном или неслучайном взаимодействии, а также о соотношении случайных и неслучайных контактов. Несколькими статистическими методами показано, что таблицы контактов отражают неслучайный характер образования контактов

Проведенный анализ контактов аминокислот на поверхности взаимодействия различных типов белковых комплексов позволил выявить статистически значимые отклонения от случайной модели. Так, наиболее предпочтительными являются контакты между остатками цистеина, а также между остатками с противоположными зарядами.

Если говорить о сравнении межцепочечных и внутрицепочечных взаимодействий между доменами, то (если сравнивать относительное число контактов) взаимодействия между доменами, расположенными в разных белковых цепях более богаты контактами между заряженными остатками, такими как Arg, Asp, Glu, Lys, а также Gln и Asn, и менее

богаты гидрофобными взаимодействиями, такими как Leu, Ile, Phe, Val, Trp.

Важное отличие интерфейсов, образованных разными и одинаковыми белковыми единицами, заключается во встречаемости контактов между остатками одного типа. Остатки с одинаковыми номерами (и, как следствие, одного типа) пространственно сближаются в центральной области интерфейса. Этот эффект становится очевидным при сравнении диагональных элементов таблиц индексов представленности.

На основании проведенных исследований специфичности взаимодействий остатков и нуклеотидов на интерфейсах белок-ДНК показано, что значимыми являются отклонения для контактов Asp-G, Trp-C, Glu-C, Asp-C и His-T. На атомном уровне наибольший вклад дает Arg – около 22% всех контактов наблюдаются между атомами Arg и атомами нуклеотидов. Lys дает 13% всех атом-атомных контактов, тогда как Thr – 7%, Ser – 6.6%, Asn – 5.9%, Tyr – 5.8%. Вклад каждой из остальных аминокислот не превышает 5%.

Выводы

1. Для определения контактирующих атомов и оценки поверхности взаимодействия белок-белок или белок-ДНК разработан метод на основе разбиения Вороного-Делоне. В рамках метода предложен способ оценки случайности образования контактов, который включает статистическую модель случайных и неслучайных контактов, позволяющую оценивать статистическую значимость обнаруженных контактов.
2. Предложен алгоритм, время работы которого пропорционально числу атомов. Разработано программное обеспечение, включающее модули для разбиения Вороного-Делоне, последующей обработки информации о контактах, а также визуализации пространственных построений.
3. Составлены невырожденные выборки интерфейсов белок-белковых и белок-нуклеиновых взаимодействий. С помощью разработанного комплекса программ изучена поверхность взаимодействия в больших выборках белок-белковых и белок-нуклеиновых комплексов.
4. Показано, что для всех рассмотренных типов белок-белковых взаимодействий статистически значимыми отклонениями от

случайной модели являются контакты между остатками цистеина, а также между остатками с противоположными зарядами. Установлено, что для белковых комплексов, образованных двумя идентичными субъединицами, характерно увеличение доли контактов между аминокислотными остатками одного типа.

5. На основании проведенных исследований специфичности взаимодействий аминокислотных остатков и нуклеотидов на интерфейсах белок-ДНК сделан вывод, что значимыми являются контакты типа Asp-G, Trp-C, Glu-C, Asp-C и His-T. Предложен механизм участия отрицательно заряженных остатков в специфических взаимодействиях белок-ДНК.

Список работ, опубликованных по теме диссертации

1. Анашкина А.А., Туманян В.Г., Кузнецов Е.Н., Галкин А.В., Есипова Н.Г. *Геометрический анализ ДНК-белковых взаимодействий на основе метода Вороного-Делоне*. Биофизика, 2008, т. 53, с. 402-406.
2. Anashkina A.A., Kuznetsov E., Esipova N., Tumanyan V. *Protein-Protein Interfaces: Amino Acids Bias for Heterocomplexes and Homodimers*. Joint Meeting of the Biophysical Society 52nd Annual Meeting and 16th International Biophysics Congress 2-6 February, 2008, Long Beach, CA, USA
3. Anashkina A.A., Kuznetsov E.N., Esipova N.G., Tumanyan V.G. *Comprehensive Statistical Analysis of Residues Interaction Specificity at Protein-Protein Interfaces*. Proteins, 2007, v. 67, p. 1060-1077
4. Анашкина А.А., Кузнецов Е.Н. и др. *Статистический анализ распределения контактов в белок-белковых комплексах*. Третья международная конференция по проблемам управления (20 – 22 июня 2006 года). Пленарные доклады и избранные труды. -М.: Институт проблем управления. 2006.
5. Anashkina A.A., Tumanyan V.G. *Voronoi-Delaune tessellation for protein-protein complexes modeling*. Proceedings of the international Moscow conference on computational molecular biology. Moscow, Russia, July 18-21, 2005, pp. 36-37.
6. Anashkina A. A., Tumanyan V. G. *Voronoi–Delaune tessellation for protein-protein complexes modeling*. Russia, Moscow. "Математика. Компьютер. Образование". Сб. трудов XII международной конференции. Под общей редакцией Г.Ю.

- Ризниченко. Ижевск: Научно-издательский центр "Регулярная и хаотическая динамика", 2005. Vol. 3, pp. 877-889.
7. Анашкина А.А., Туманян В.Г. *Применение построения Вороного-Делоне для исследования пространственных отношений в белковой глобуле*. III съезд биофизиков России, 24-29 июня 2004 г, Воронеж, стр. 751.
 8. Анашкина А.А., Туманян В.Г. *Исследование доменной структуры глобулярных белков с помощью построения Вороного-Делоне*. VIII международная конференция: образование, экология, экономика, информатика. Астрахань, Россия, сентябрь 15-20, 2003.
 9. Анашкина А.А., Туманян В.Г. *Исследование доменной структуры глобулярных белков с помощью построения Вороного-Делоне*. 14-я Международная зимняя молодежная научная школа «Перспективные направления физико-химической биологии и биотехнологии», 11-15 февраля 2002 г.
 10. Анашкина А.А., Туманян В.Г. *Применение построения Вороного-Делоне для исследования доменной структуры глобулярных белков*. XLIV научная конференция МФТИ, 2001
 11. Анашкина А.А., Березовский И.Н., Есипова Н.Г., Туманян В.Г. *Анализ доменной структуры некоторых глобулярных белков и их комплексов*. 13-я Международная зимняя молодежная научная школа «Перспективные направления физико-химической биологии и биотехнологии», февраль 2001 г.