

На правах рукописи

РЫЧКОВ Георгий Николаевич

**МОДЕЛИРОВАНИЕ ГИДРАТАЦИИ БЕЛКОВ
НА ОСНОВЕ РАСЧЁТОВ ПЛОЩАДИ ПОВЕРХНОСТИ,
ДОСТУПНОЙ РАСТВОРИТЕЛЮ**

03.00.02 – биофизика

АВТОРЕФЕРАТ
диссертации
на соискание учёной степени
кандидата физико-математических наук

Санкт-Петербург – 2007

Работа выполнена в Лаборатории биофизики макромолекул Отделения молекулярной и радиационной биофизики Петербургского института ядерной физики им. Б.П. Константинова РАН (г. Гатчина).

Научный руководитель: кандидат физико-математических наук
Петухов Михаил Геннадьевич.

Официальные оппоненты: доктор физико-математических наук,
ведущий научный сотрудник
Тулуб Александр Александрович,
Санкт-Петербургский государственный
университет, Институт физиологии
им. А.А. Ухтомского.

доктор физико-математических наук,
ведущий научный сотрудник
Маслов Владимир Григорьевич,
Санкт-Петербургский государственный
университет информационных технологий,
механики и оптики.

Ведущая организация: Институт цитологии РАН,
Санкт-Петербург.

Защита состоится *30 октября 2007 г.* в __ часов на заседании диссертационного совета Д 212.229.25 при ГОУ ВПО «Санкт-Петербургский государственный политехнический университет» по адресу:

195251, Санкт-Петербург, ул. Хлопина, д. 5,
Факультет медицинской физики и биоинженерии, ауд. ____

С диссертацией можно ознакомиться в фундаментальной библиотеке ГОУ ВПО «Санкт-Петербургский государственный политехнический университет».

Автореферат диссертации разослан _____ *сентября 2007 г.*

Учёный секретарь
диссертационного совета
кандидат физико-математических наук



Власова О.Л.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

Многие современные теоретические исследования направлены на решение таких актуальных задач биофизики, как моделирование процессов сворачивания белков в биологически активную конформацию, предсказание их конформационной стабильности при внесении точечных мутаций, конструирование ферментов с заданными свойствами, анализ белок-белковых взаимодействий, а также докинг потенциальных лекарственных препаратов с гибкими рецепторами. Во всех этих процессах гидратация играет важную роль. Решения указанных задач предъявляют к моделированию гидратации ряд требований. В первую очередь необходимо корректно описывать изменения свободной энергии гидратации разных конформационных состояний одной молекулы – как при крупномасштабных конформационных переходах, так и при небольших локальных изменениях структуры. Кроме того, модель должна быть вычислительно эффективна, а используемый в ней потенциал гидратации должен достаточно просто встраиваться в существующие процедуры глобальной минимизации конформационной энергии исследуемой биологической системы. В частности, многообещающими представляются процедуры глобальной минимизации с использованием эффективных алгоритмов интервального анализа. Для успешного применения этих методов необходимо, чтобы энергетические потенциалы были определены с помощью простых дифференцируемых аналитических функций от координат атомов исследуемой молекулярной системы.

Из всех используемых моделей гидратации белков наиболее детализованной и точной является модель периодического водного бокса. Однако для каждого вычисления энергии гидратации необходимо термодинамически уравнивать водный бокс в течение длительного времени, а затем усреднять полученные данные по различным конформационным состояниям исследуемой молекулярной системы. Поэтому модель периодического водного бокса не может быть применена к исследованию процесса сворачивания белка, где из колоссального количества возможных конформаций белок выбирает всего одну уникальную конформацию, в которой он обладает биологической активностью. Для небольшого белка моделирование молекулярной динамики этого процесса в периодическом водном боксе даже на самых мощных кластерах ЭВМ заняло бы нереально долгое время.

Применение альтернативных моделей, описывающих воду непрерывной средой, позволяет значительно сократить время, требуемое для вычисления энергии гидратации. В наиболее вычислительно эффективных моделях гидратации для расчёта различных составляющих энергии гидратации широко используется площадь поверхности белка, доступная растворителю (ПДР). В частности, гидрофобная составляющая энергии гидратации считается линей-

но пропорциональной площади ПДР белка. Описан метод, позволяющий вычислять Ван-дер-Ваальсову энергию гидратации на основе интегрирования по ПДР белка. Однако до сих пор не было предложено достаточно точного способа вычисления энергии образования водородных связей между белком и водой на основе вычислений площади ПДР. При этом следует отметить, что для полярных и заряженных групп на поверхности белка наибольший вклад в общую энергию гидратации вносится именно водородными связями между белком и водой. Аналитическое выражение, связывающее площадь ПДР полярных и заряженных атомов белка с энергией образования ими водородной связи с водой, имело бы большое значение для усовершенствования моделей непрерывного растворителя.

В настоящее время имеются численные, а также аналитические методы точного и приближённого расчёта ПДР, каждый из которых имеет свои преимущества и недостатки. Точные аналитические методы расчёта площади ПДР сложны с математической точки зрения и не позволяют получать производные высших порядков по координатам атомов молекулы. Кроме того, описание конформации молекул с помощью математического аппарата интервального анализа включает применение точных аналитических методов для вычисления площади ПДР, где необходимо построение на поверхности молекулы точного расположения множества вершин, образуемых пересечениями многочисленных атомов.

Помимо точных аналитических методов в молекулярном моделировании часто применяются аналитические методы приближённого вычисления площади ПДР. Основными проблемами опубликованных на данный момент методов этого класса являются недостаточная точность вычислений и наличие большого количества эмпирических параметров, зависящих как от химической природы, так и от конформации исследуемой макромолекулы. Поэтому применение таких методов ограничивается лишь узким классом молекул в фиксированной конформации, на котором они были параметризованы. Создание нового аналитического метода приближённого вычисления площади ПДР, не использующего эмпирические параметры и обладающего достаточной точностью, позволит расширить границы применимости потенциалов гидратации, основанных на ПДР.

Цели и задачи исследования

Целью работы являлось создание простой и вычислительно эффективной модели, корректно описывающей взаимодействия белка с водой на основе расчётов его площади поверхности, доступной растворителю (ПДР).

Для достижения этой цели были поставлены следующие задачи:

1. Провести молекулярно динамическое исследование гидратации белка крамбин в периодическом водном боксе.
2. Проанализировав полученные молекулярно динамические траектории, установить связь между площадью ПДР полярных и заряженных атомов

белка и вероятностью образования ими водородных связей с молекулами воды.

3. На основе выявленных функциональных зависимостей получить аналитические выражения, описывающие свободную энергию образования водородных связей полярными и заряженными атомами белка с водой через площадь их ПДР.
4. Получить производные от свободной энергии гидратации по ПДР и сравнить их с традиционными атомарными параметрами гидратации (АПГ).
5. Разработать новый непараметрический метод приближённого вычисления площади ПДР для молекул любой химической природы, обладающий достаточной точностью вычислений.
6. Получить производные от площади ПДР по координатам атомов рассматриваемой молекулы.

Основные положения, выносимые на защиту

1. Для полярных и заряженных атомов белка получены аналитические выражения, определяющие зависимость атомарных параметров гидратации от их площади поверхности, доступной растворителю (ПДР). На основе полученных выражений показана некорректность широко используемых предположений о постоянстве атомарных параметров гидратации для описания взаимодействий белка с водой.
2. Получены аналитические выражения, позволяющие на основе площади ПДР полярных и заряженных атомов белка эффективно вычислять свободные энергии образования ими водородных связей с молекулами воды.
3. Разработан новый метод приближённого расчёта площади ПДР, и получены аналитические выражения для её первых производных по координатам атомов молекулы. Отсутствие в методе эмпирических параметров и достаточно высокая точность вычислений обеспечивают возможность его применения для расчёта площади ПДР любых химических соединений независимо от их конформации.

Научная новизна

- 1) Впервые получено аналитическое выражение, определяющее зависимость свободной энергии образования водородной связи молекулами воды с полярными и заряженными атомами белков от площади их ПДР.
- 2) Впервые показана некорректность использования традиционных АПГ, являющихся константами, для описания гидратации полярных и заряженных атомов белков. Взамен некорректного представления предложено аналитическое выражение, определяющее зависимость АПГ от площади ПДР.
- 3) Разработан новый метод приближённого вычисления площади ПДР и получены аналитические производные площади ПДР по координатам атомов рассматриваемой молекулы. В отличие от известных методов этого класса

в разработанном методе отсутствуют эмпирические параметры; разработанный метод обладает повышенной точностью, которая не зависит от химической природы и конформации исследуемой молекулы.

Теоретическая и практическая значимость

Полученные аналитические выражения, описывающие зависимость между площадью ПДР полярных и заряженных атомов белка и свободной энергией образования ими водородной связи с молекулами воды, могут быть успешно использованы во всех моделях растворителя непрерывного типа для создания новых более совершенных потенциалов, основанных на вычислении площади ПДР исследуемой молекулы. Разработка таких потенциалов с учётом существования водородных связей между белком и водой будет более полно отражать реальную физическую картину процессов гидратации.

Разработанный метод приближённого вычисления площади ПДР в силу своих преимуществ может быть с успехом применён в моделировании молекулярной динамики и в эффективных процедурах глобальной оптимизации конформационной энергии молекул различных классов.

Апробация работы

Материалы диссертации докладывались на IX Международной Пущинской школе-конференции «Биология – наука XXI века» (Пущино, 2005), на VI и VII конференциях молодых ученых Отделения молекулярной и радиационной биофизики ПИЯФ РАН (Гатчина, ПИЯФ РАН, 2005, 2006), на II научной конференции «Молекулярная генетика, биофизика и медицина» (Бреслеровские чтения-II) (Санкт-Петербург, СПбНЦ, 2006), на XIV Международной конференции «Ломоносов-2007» (Москва, МГУ, 2007) и на III Санкт-Петербургской конференции «Современные проблемы науки о полимерах» (Санкт-Петербург, ИВС, 2007).

Публикации

По материалам диссертации опубликовано 10 работ (стр. 18).

Объём и структура диссертации

Диссертационная работа изложена на 147 страницах машинописного текста и содержит следующие разделы: введение, обзор литературы, материалы и методы, результаты и обсуждение, выводы. Материал иллюстрирован 18 рисунками и 6 таблицами. Библиографический указатель включает 209 публикаций.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследования, приведены основные научные положения, показано научно-практическое значение исследования.

Глава 1. Обзор литературы

Рассмотрены две основные модели гидратации: модель периодических водных боксов и модель непрерывного растворителя. На основе термодинамичес-

кого цикла, описывающего процесс гидратации, учтены и проанализированы все физические вклады в потенциал взаимодействия белка с водой: электростатические, гидрофобные, Ван-дер-Ваальсовы взаимодействия и образование водородных связей. Рассмотрены различные методы расчёта этих энергетических составляющих. Отмечается, что энергии основных физических взаимодействий белка с водой в моделях непрерывного растворителя могут быть найдены с помощью расчётов площади поверхности белка, доступной растворителю (ПДР). Описаны известные методы расчёта площади ПДР.

Глава 2. Материалы и методы

2.1. Моделирование молекулярной динамики (МД). Расчёты проводились с помощью пакета программ для МД AMBER 4.1. В качестве объекта исследования был выбран белок крамбин из абиссинской капусты (*Crambe abyssinica*) (код PDB 1AB1), чья трёхмерная кристаллическая структура была получена с разрешением 0,87 Å. С целью устранения незначительных перекрытий Ван-дер-Ваальсовых оболочек атомов трёхмерная структура белка была регуляризована с помощью стандартной процедуры из пакета программ для молекулярного моделирования ICM. N-концевая аминокислотная группа и C-концевая карбоксильная группа белка считались нейтральными. Трёхмерная структура белка была размещена в октаэдрическом периодическом водном боксе. Молекулы воды в водном боксе моделировались потенциалом TIP3P. Энергии невалентных взаимодействий отсекались на расстояниях, превышающих 12 Å. Подготовка системы к МД заключалась в минимизации энергии молекул воды в водном боксе и последующем постепенном нагреве системы от 10 К до 293 К.

Моделирование МД проводилось при 293 К на протяжении 350 пс; декартовы координаты атомов системы сохранялись на жёсткий диск каждые 0,04 пс. Временной интервал молекулярно-динамической траектории был выбран равным 0,002 пс. На протяжении всего времени моделирования трёхмерная структура крамбина оставалась неподвижной. Чтобы сохранять в фиксированном состоянии длины всех ковалентных связей белка, применялся алгоритм SHAKE.

Анализ образования водородных связей проводился с помощью модуля CARNAL пакета программ AMBER 4.1. Вероятность образования водородной связи, P_{hb} , между молекулой воды и соответствующим атомом белка рассчитывалась как часть времени, в течение которого она существовала. P_{hb} вычислялась на последних 250 пс траектории МД, так как на протяжении первых 100 пс происходило уравнивание водного бокса. При анализе траектории МД предполагалось, что водородная связь образуется при условии, что расстояние между тяжёлыми атомами доноров и акцепторов $\leq 3,1$ Å для связи NH...O и $\leq 3,1$ Å для связи OH...O, при этом угол донор–H–акцептор не превышает 30°. В случае групп NH₂ и карбонильных/карбоксильных атомов кислорода аминокислот Asp, Asn,

Glu и Gln основной цепи, которые могут образовывать две симметрично расположенные водородные связи, накладывалось дополнительное условие, полученное статистическим анализом кристаллических структур из белковой базы данных: двугранный угол между молекулами воды должен быть больше 120° .

2.2. Статистический анализ белковой базы данных. Статистический анализ геометрии образования водородных связей между молекулами воды и соответствующими атомами белка проводился на репрезентативном наборе белков, состоящем из 42 кристаллических структур, полученных с разрешением менее $1,5 \text{ \AA}$. Гомология этих белков составляла менее 25% с R-фактором менее 0,19 (PDB-SELECT). Кристаллические структуры белков из репрезентативного набора были взяты из Brookhaven Protein Data Bank. Стерические условия образования водородной связи были аналогичны условиям, описанным в пункте 2.1.

2.3. Репрезентативный набор молекул для тестирования точности приближённых методов расчёта площади ПДР. Список молекул, использовавшийся при тестировании разработанного нами метода приближённого расчёта площади ПДР, включал в себя небольшие органические соединения, белки, ДНК, РНК и их комплексы, чьи трёхмерные структуры были взяты из Protein Data Bank (PDB). Чтобы ввести в трёхмерные структуры молекул атомы водорода, а также чтобы перевести белки в развёрнутое состояние, использовался пакет программ для молекулярного моделирования ICM. Перевод белков в развёрнутое состояние проводился по следующей схеме: двугранным углам ϕ , ψ и Ω были назначены значения 180° , после чего получившиеся структуры были энергетически минимизированы с помощью 200 шагов метода сопряженных градиентов.

2.4. Метод сокращения списка соседей. С целью увеличения скорости вычислений площади ПДР был использован метод сокращения списка соседей (NLR – neighbor list reduction), описанный в работе Weiser и сотр. Список соседей рассматриваемого атома включает в себя номера всех остальных атомов молекулы, имеющих с ним ненулевую площадь пересечения. NLR удаляет из списка соседей те атомы, которые не влияют на площадь ПДР рассматриваемого атома. При вычислениях площади ПДР список соседей некоторых атомов может сокращаться на 74%. В работе была использована 3NLR версия метода, где принимаются во внимание пересечения не более трёх сферических атомов: рассматриваемого, кандидата на удаление и одной из оставшихся соседних сфер.

ГЛАВА 3. Изучение образования водородных связей

молекулами воды с полярными и заряженными группами белка

Моделирование МД белка крамбин в периодическом водном боксе (см. параграф 2.1.) показало, что отдельные участки его поверхности различаются стабильностью водородных связей белка с водой в зависимости от их химической природы и доступности для молекул воды. По данным МД были получены зависимости вероятности образования водородной связи, P_{hb} ,

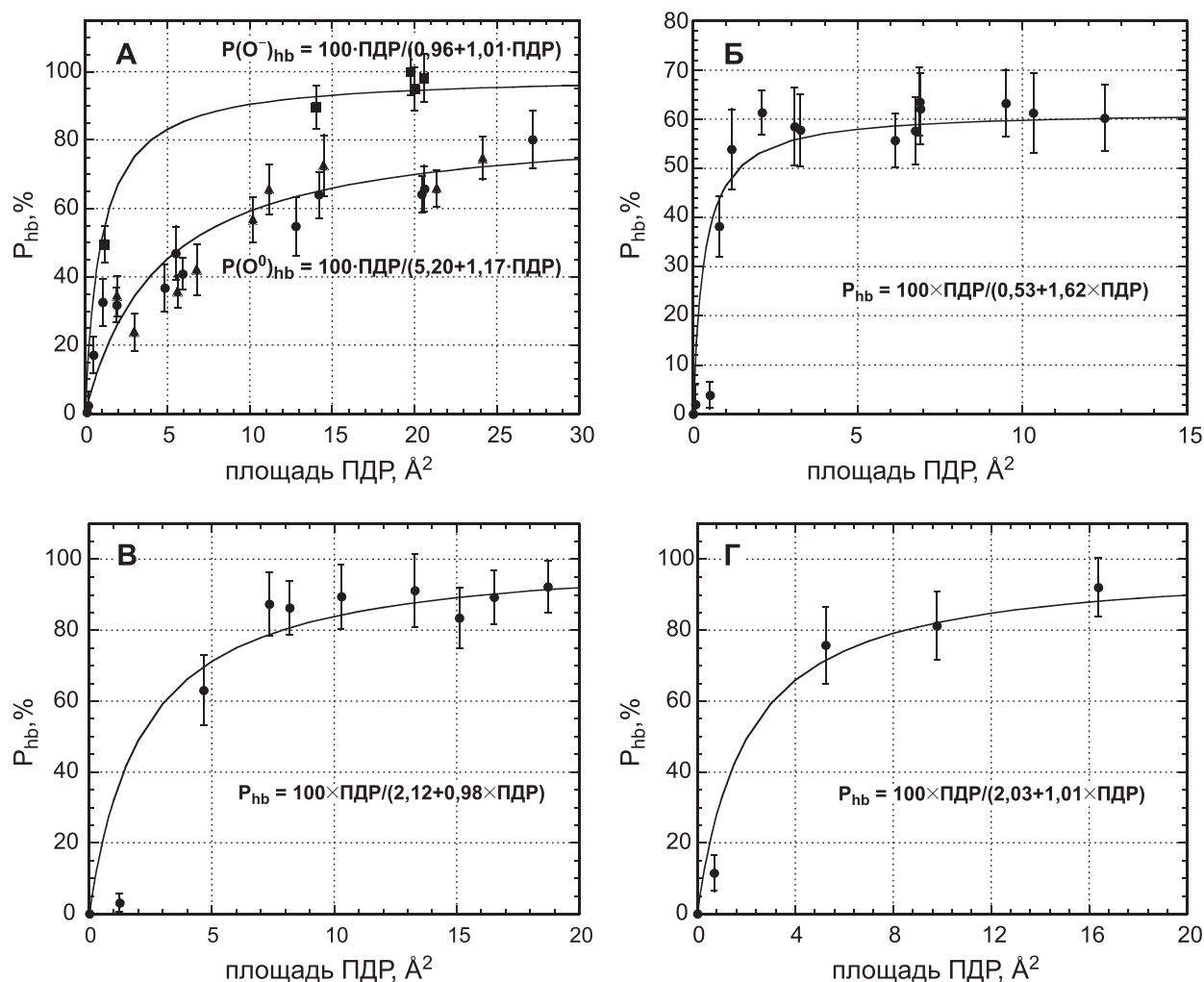


Рис. 1. Корреляционные графики площади ПДР акцепторных (А) и донорных (Б, В, Г) групп белка и вероятности образования водородных связей этими химическими группами. В секции (А) ● и ▲ — относятся к незаряженным карбонильным группам основной цепи белка и гидроксильным группам боковых цепей аминокислот соответственно; ■ — относятся к заряженным атомам кислорода карбоксильных групп боковых цепей Asp и Glu. (Б) Незаряженные группы NH и NH₂ в основной цепи белка и боковых ветвей аминокислот; (В) незаряженные группы OH в боковых цепях Ser и Thr; (Г) заряженные группы NH₂⁺ и NH₃⁺. Для групп-доноров водородной связи площадь ПДР относится к атомам водорода этих групп. На рисунке приводятся параметры аппроксимации кривой по формуле (2)

основными акцепторными и донорными группами белка с молекулами воды от площади ПДР этих групп (рис. 1). Необходимо отметить, что атомы белка, способные образовывать водные мостики, были исключены из исследования, так как для них не наблюдается простой зависимости величины P_{hb} от площади ПДР (это было показано ранее М.Г. Петуховым и сотр.).

По данным МД были предложены аналитические выражения, позволяющие на основе площади ПДР полярных и заряженных атомов белка вычислять свободные энергии образования ими водородных связей с молекулами воды. Был

проведён сравнительный анализ традиционных атомарных параметров гидратации (АПГ) и АПГ, полученных с помощью предложенного в работе аналитического выражения.

Также в главе приводятся результаты статистического анализа репрезентативного набора белков по таким параметрам, как: 1) величины двугранных углов между двумя молекулами воды, каждая из которых соединена водородной связью с интересующим (полярным либо заряженным) атомом белка; 2) величины площади ПДР полярных и заряженных атомов белков (см. параграф 2.2).

3.1. Случай, где атомы белка выступают в роли акцептора водородной связи.

Основными акцепторами водородной связи являются атомы кислорода в карбоксильной ($-\text{COO}^-$), карбонильной ($-\text{CO}$) и гидроксильной ($-\text{OH}$) химических группах. Атомы кислорода этих групп, выступая в качестве акцепторов, могут образовывать две водородные связи с молекулами воды.

Из графика на рис. 1А видно, что с увеличением площади ПДР атомов кислорода увеличивается и вероятность образования ими водородной связи с молекулами воды. При небольших значениях поверхности, доступной растворителю, стабильность водородной связи невысока. Зависимость P_{hb} от ПДР нелинейная и имеет гиперболический вид. Стабильность водородных связей, образованных заряженным атомом кислорода карбоксильных групп, выше, чем полярных атомов кислорода карбонильных и гидроксильных групп.

3.2. Случай, где атомы белка выступают в роли донора водородной связи.

В белках основными донорами протона при образовании водородной связи с водой являются амидо- ($-\text{NH}-$), amino- ($-\text{NH}_2$), и гидроксильная ($-\text{OH}$) химические группы. В зависимости от величины рН эти группы могут участвовать в образовании от одной до трёх водородных связей с водой.

На рис. 1В, Г и Д проиллюстрировано изменение вероятности образования водородной связи для групп белка, являющихся донором протона, в зависимости от их площади ПДР. Общая форма зависимости одинакова для всех рассматриваемых типов доноров и акцепторов, однако, максимальные значения P_{hb} существенно отличаются. Величины P_{hb} для гидроксильных групп, а также заряженных групп NH_3^+ находятся примерно на одном уровне ($\approx 90\%$), в то время как для незаряженных групп NH и NH_2 основной цепи белка и боковых цепей аминокислот достигают только 60%. Полученные результаты свидетельствуют о том, что незаряженные группы NH и NH_2 образуют относительно слабые водородные связи с водой, и их вклад в общую энергию гидратации белка заметно меньше, чем вклад, вносимый заряженными amino- и гидроксильными группами. Это хорошо согласуется с литературными данными, показывающими, что связь $\text{OH}\cdots\text{O}$ короче и сильнее в сравнении со связью $\text{NH}\cdots\text{O}$; при этом, водородная связь сильнее, если одна из образующих её химических групп обладает зарядом.

3.3. Образование водородных связей SH и S-CH₃ группами белка. Атомы серы представлены в боковых цепях аминокислот Cys и Met. К сожалению, в случае крамбина количества атомов серы было недостаточно, чтобы сделать достоверные выводы о корреляции их площади ПДР с вероятностью образования водородной связи. Следует отметить, что такие водородные связи вносят незначительный вклад в общую энергию гидратации белка.

3.4. Вклад водородных связей в энергию гидратации. В случае образования одной водородной связи между белком и молекулой воды свободная энергия образования этой связи, ΔG_{hb} , может быть вычислена с помощью классического соотношения между изменением свободной энергии одностадийной химической реакции и её константой равновесия:

$$\Delta G_{hb} = -RT \ln \left(\frac{P_{hb}}{1 - P_{hb}} \right), \quad (1)$$

где R – универсальная газовая постоянная, T – температура по шкале Кельвина, P_{hb} – вероятность образования водородной связи между белком и молекулой воды. Если выразить P_{hb} через площадь ПДР химической группы, с которой образована водородная связь, можно достаточно точно и быстро вычислить её энергетический вклад в энергию гидратации белка.

Из рис. 1 видно, что форма кривых зависимости P_{hb} от ПДР напоминает форму кривой насыщения Михаэлиса-Ментен, описывающей кинетику химической реакции. С использованием уравнений, задающих кинетику многоцентрового связывания рецептора (полярной либо заряженной группы белка) с лигандом (молекулой воды), была получена формула, связывающая поверхность, доступную растворителю, A_i , химической группы белка и вероятность образования ею водородной связи с молекулами воды:

$$P_{hb} = \frac{A_i}{A + B \times A_i}, \quad (2)$$

A и B – константы, зависящие от типа донора или акцептора водородной связи.

Благодаря простоте функций, используемых в предложенной модели, легко получить первую производную энергии образования водородных связей белка с водой по площади ПДР отдельных атомов молекулы:

$$\frac{d\Delta G_{hb}}{dA_i} = -RT \frac{1}{A_i \times \left(1 + \frac{B-1}{A} A_i \right)}. \quad (3)$$

Представлялось интересным провести сравнение производных по ПДР функции гидратации, полученной в нашей работе, с производными функций гидратации, полученных в других моделях. Из рис. 2 видно, что в нашем случае кривые зависимости производной от ПДР нелинейны и имеют участки быстрого возрастания (до 10 \AA^2) и участки насыщения (более 10 \AA^2).

В традиционной непрерывной модели гидратации предполагается, что свободная энергия гидратации атомов белка, ΔG_{solu} , линейно пропорциональна ПДР этих атомов: $\Delta G_{solu} = \text{АПГ} \times \text{ПДР}$. То есть в такой модели производная от ΔG_{solu} по ПДР является константой – АПГ (атомарным параметром гидратации, который выражает среднюю энергию гидратации, приходящуюся на 1 \AA^2 данного типа атома). Непрерывная модель гидратации была параметризована на экспериментальных данных по переносу из неполярного окружения в водное небольших органических молекул, сходных по химической структуре с боковыми цепями аминокислот. В таких молекулах атомы имеют максимальные значения площади ПДР, соответствующие району насыщения на наших графиках (рис. 2), где величину производной приблизительно можно считать постоянной. Отсюда понятно, почему авторы смогли успешно параметризовать непрерывные модели в предположении о постоянных значениях атомарных параметров гидратации.

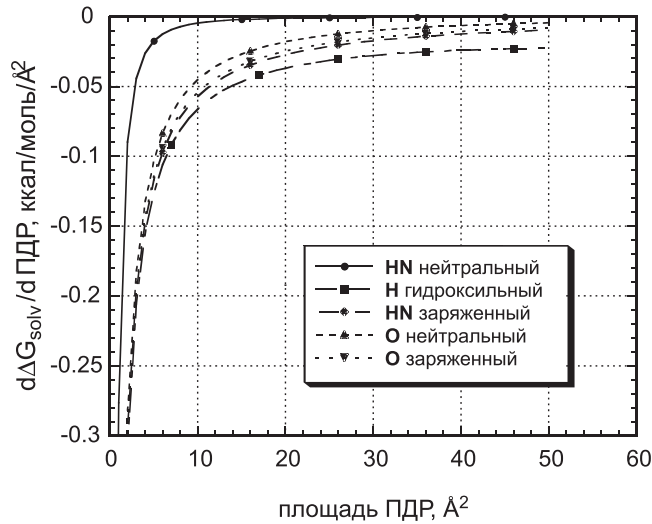


Рис. 2. Графики первой производной по площади ПДР функций свободной энергии образования водородных связей пятью основными полярными и заряженными группами белка с водой с использованием коэффициентов параметризации функций, указанных на рис. 1

В белках по сравнению с небольшими молекулами экспонированные на поверхность атомы имеют меньшую площадь ПДР. Наши расчёты показывают, что средние значения площади ПДР рассматриваемых типов атомов в белках распределены в большей степени в области от 0 \AA^2 до 10 \AA^2 и, следовательно, попадают в участок быстрого возрастания на графике, где АПГ нельзя считать постоянными. Таким образом, удалось объяснить, почему непрерывные модели с постоянными АПГ так неточны при вычислениях энергии гидратации белков.

ГЛАВА 4. Расчёт площади ПДР биополимеров с помощью объединения пар атомов (JNA)

В этой главе описывается новый метод приближённого вычисления площади ПДР. В отличие от известных методов этого класса в разработанном методе отсутствуют эмпирические параметры; метод обладает повышенной точностью, которая не зависит от химической природы и конформации исследуемой молекулы. Также глава содержит параграф, в котором подробно описана процедура получения первых производных площади ПДР по координатам атомов.

4.1. Алгоритм метода. В предложенном методе площадь ПДР отдельного атома рассматриваемой молекулы вычисляется с помощью рекурсивной процедуры, представленной на рис. 3. Общая молекулярная площадь ПДР является суммой площадей ПДР всех её атомов. Так как задача нахождения площади ПДР носит исключительно геометрический характер, то в дальнейшем словосочетание сферический атом, будет заменяться словом сфера.

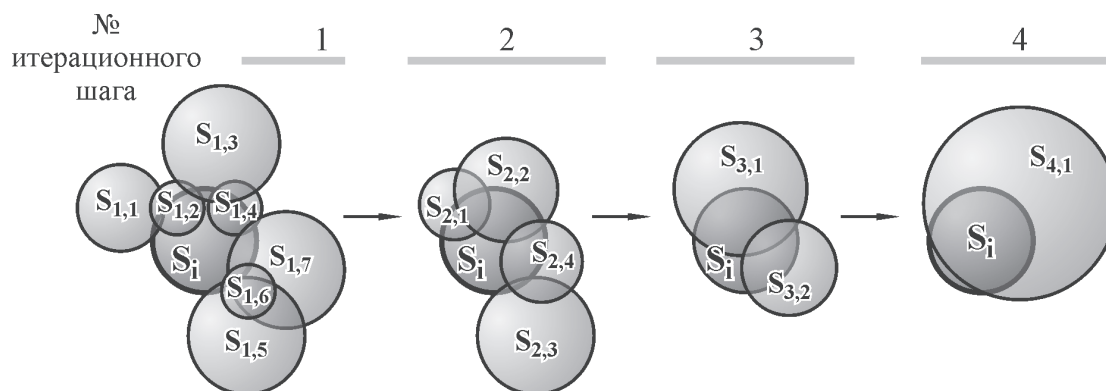


Рис. 3. Рекурсивный алгоритм объединения соседних пар сфер. Схема иллюстрирует алгоритм вычисления площади ПДР сферы S_i . $S_{l,m}$ – сферы из списка соседей сферы S_i ; индекс l указывает на номер итерационного шага, индекс m обозначает позицию сферы $S_{l,m}$ в списке соседей сферы S_i . Работа алгоритма показана в двумерном представлении на примере семи соседних сфер, распределённых вокруг центральной сферы S_i .

Пусть сфера S_i , для которой рассчитывается площадь ПДР, имеет семь соседних с ней сфер, обозначенных как $S_{1,1} - S_{1,7}$ и составляющих в совокупности список соседей сферы S_i (рис. 3). Индексы в обозначении соседних сфер указывают на номер текущего шага итерации и на позицию сферы в списке соседей на данном итерационном шаге. На каждом итерационном шаге сферы из списка соседей S_i по определённому правилу попарно объединяются в сферу большего радиуса. При этом результирующая сфера вырезает из поверхности S_i площадь, равную площади, суммарно вырезаемой двумя объединёнными сферами. В случае, когда число сфер в списке соседей является нечётным, непарная сфера перемещается на следующий итерационный шаг без изменений. Описанная итерационная процедура продолжается до тех пор, пока в списке соседей S_i останется либо одна сфера, либо несколько не пересекающихся друг с другом сфер. Таким образом, площадь поверхности рассматриваемой сферы, оказавшаяся внутри сфер, оставшихся в списке соседей (или внутри единственной оставшейся соседней сферы) на последнем итерационном шаге, будет эквивалентна площади, вырезаемой всеми исходными соседями из поверхности сферы S_i .

При объединении пары сфер вычисляются площади, вырезаемые из поверхности рассматриваемой сферы каждой из сфер по отдельности, и площадь, вырезаемая пересечением этих двух сфер.

В п. 4.1.1. диссертации приводится формула, описывающая площадь, вырезаемую сферой j из поверхности сферы i , A_{ij} :

$$A_{ij} = 2\pi r_i \left[r_i - \frac{d_{ij}}{2} - \frac{r_i^2 - r_j^2}{2d_{ij}} \right], \quad (4)$$

где r_i и r_j – радиусы сфер i и j соответственно, а d_{ij} – расстояние, разделяющее центры этих сфер.

В п. 4.1.2. анализируются все возможные варианты пресечения трёх сфер, которые условно делятся на «истинные» и «ложные» пересечения. Также описываются аналитические выражения для вычисления площади, вырезаемой из поверхности рассматриваемой сферы парой соседних сфер, $A_{l,jk}$, в случае «истинных» и «ложных» пересечений. Для «истинных» пересечений это выражение имеет вид

$$A_{l,jk} = 2r_i^2 \left\{ \tan^{-1} \left[\frac{d_{l,ik} \omega}{r_i q_2} (1 - \varepsilon_2) \right] + \tan^{-1} \left[\frac{d_{l,ij} \omega}{r_i q_3} (1 + \varepsilon_3) \right] \right\} - r_i \left\{ d_{l,ij} (1 + \varepsilon_3) \tan^{-1} \left[\frac{2\omega}{q_3} \right] + d_{l,ik} (1 - \varepsilon_2) \tan^{-1} \left[\frac{2\omega}{q_2} \right] \right\}. \quad (5)$$

В приведённой формуле ε_1 , ε_2 , ε_3 , q_1 , q_2 , q_3 и ω являются вспомогательными функциями, зависящими от радиусов трёх пересекающихся сфер и расстояний между их центрами.

В п. 4.1.3. и 4.1.4. объясняется критерий выбора сфер, являющихся кандидатами для объединения, подробно излагается процедура объединения отобранной из списка соседей пары сфер, и описываются условия завершения процесса объединения. Методика увеличения точности вычислений на основе фильтрации данных, полученных при вычислении площадей пересечения трёх сфер, обсуждается в п. 4.1.5.

4.2. Результаты работы алгоритма и их обсуждение. Для проверки точности метода JNA результаты вычисления приближённой площади ПДР сравнивали с результатами вычислений, выполненных точным аналитическим методом (см. табл.). В качестве тестового набора были выбраны молекулы разной химической природы, куда вошли небольшие органические молекулы, РНК, ДНК и белки в свёрнутой и в развёрнутой конформациях.

Можно заметить, что предложенный метод демонстрирует более высокую точность расчёта для многоатомных макромолекул по сравнению с небольшими органическими соединениями. Применение метода JNA к белкам в развёрнутой конформации показывает незначительное уменьшение точности по сравнению с их свёрнутыми формами. Присутствие атомов водорода в рассматриваемых молекулах существенно повышает точность вычислений.

Таблица. Результаты тестирования точности расчётов метода JNA, проведённого на трёх различных наборах молекул

Репрезентативный набор молекул	Средние атомарные отклонения, Å ²		Максимальн. атомарные отклонения, Å ²		Ошибка вычисления молекулярной площади, %	
	Мин. знач.	Макс. знач.	Мин. знач.	Макс. знач.	Мин. знач.	Макс. знач.
Небольшие органические молекулы, РНК, ДНК, белки и их комплексы	1,20	3,47	5,09	18,91	0,05	7,30
Свёрнутые белки*	0,78	0,97	10,61	14,86	3,0	8,5
Развёрнутые белки*	1,65	1,87	9,47	16,78	2,37	8,63

* Трёхмерная структура белков в полноатомном представлении.

На рис. 4 изображены графики корреляции между величинами *атомарной* площади ПДР, вычисленными на тестовом наборе молекул с помощью точного метода и метода JNA. Видно, что коэффициент корреляции между точными и приближёнными величинами атомарной площади ПДР во всех случаях очень высок и варьируется в пределах 0,96–0,98. Такой результат свидетельствует о том, что предложенный метод приближённого вычисления площади ПДР с равным успехом может быть применён как к молекулам в свёрнутых, так и в развёрнутых конформациях.

Графики на рис. 5 демонстрируют точность вычислений *молекулярной* площади ПДР методом JNA на трёх наборах молекул. Для всех наборов молекул коэффициент корреляции составляет 0,99. Следует отметить, что для свёрнутых и развёрнутых белков в полноатомном представлении предложенный метод несколько переоценивает суммарную площадь ПДР. Тем не менее, благодаря очень высокому коэффициенту корреляции, точность вычислений можно повысить, проведя простое линейное преобразование.

Кроме того, в п. 4.2. проводится сравнение точности результатов вычислений метода JNA с точностью двух альтернативных приближённых методов. Отмечается, что метод JNA рассчитывает средние атомарные величины с существенно меньшей ошибкой, производит меньшие максимальные атомарные отклонения, а суммарная площадь макромолекулярной ПДР вычисляется примерно одинаково точно всеми сравниваемыми методами.

4.3. Вычислительная эффективность метода JNA. Для оценки скорости расчётов разработанного метода было проведено сравнение с наиболее быстрым

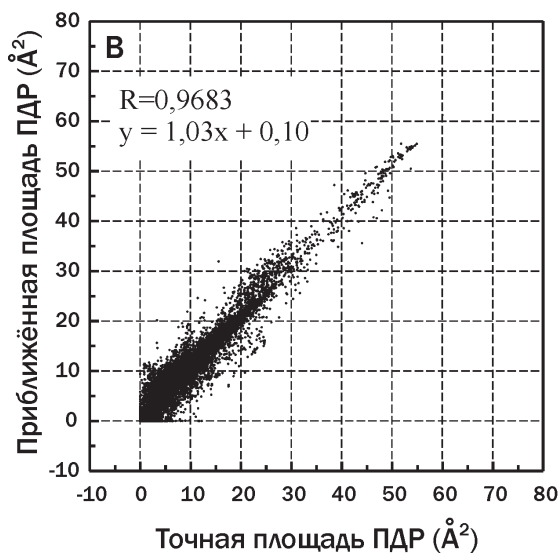
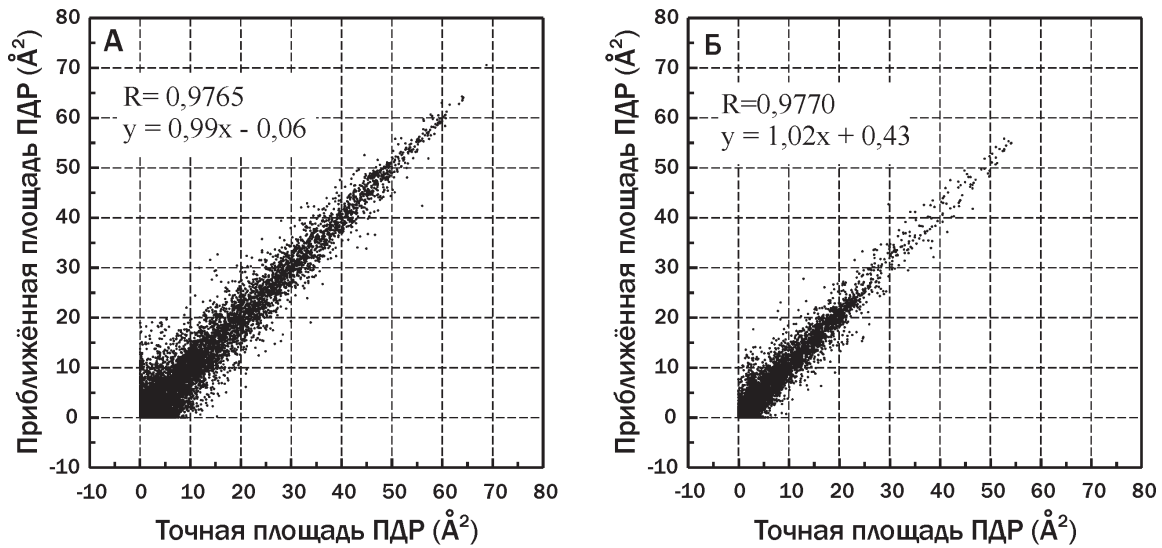


Рис. 4. Корреляционные графики атомарных площадей ПДР, вычисленных точно и приближённо с помощью метода JNA. На графиках приведены величины площади ПДР для каждого атома из двух тестовых наборов молекул: (А) – набор органических молекул различной химической природы; (Б) – набор небольших белков в полноатомном представлении в свёрнутой конформации и (В) – тот же набор белков, но в развернутой конформации. На графиках приводятся коэффициенты корреляции R и уравнения линии регрессии

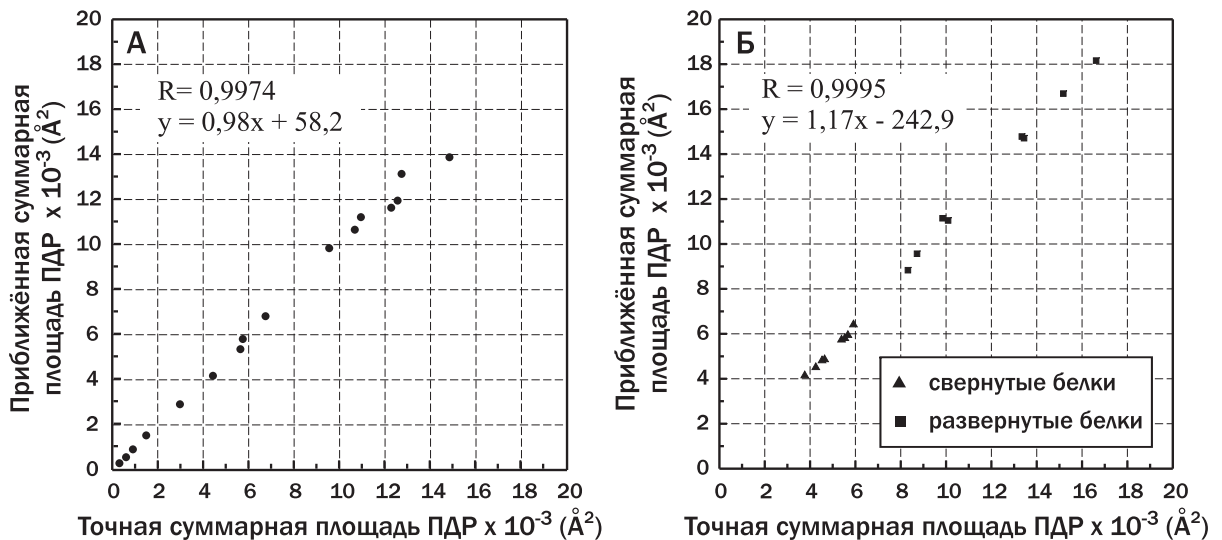


Рис. 5. Корреляционные графики суммарных площадей ПДР молекул из двух тестовых наборов, вычисленных точно и приближённо с помощью метода JNA. (А) – набор молекул различной химической природы; (Б) – набор небольших белков в свёрнутой и развернутой конформациях

из имеющихся на данный момент точных методов вычисления площади ПДР – GETAREA. Тестирование показало, что вычислительная скорость метода JNA сравнима со скоростью GETAREA. Вероятно, в дальнейшем производительность разработанного метода может быть существенно увеличена как за счёт оптимизации программного алгоритма, так и за счёт выбора упрощённых условий попарного объединения соседних сфер.

Необходимо отметить, что максимально высокая скорость вычислений не была главной целью разработки этого метода. В противоположность большинству методов вычисления площади ПДР был разработан непараметрический метод, который специфически можно использовать в процедурах глобальной оптимизации с применением эффективных алгоритмов интервального анализа.

4.4. Аналитические выражения для расчёта первых производных площади ПДР по координатам атомов. В параграфе подробно излагается алгоритм расчёта первых производных площади ПДР по координатам атомов.

Выводы

1. Для полярных и заряженных атомов белка получены аналитические выражения, определяющие зависимость атомарных параметров гидратации от их площади поверхности, доступной растворителю (ПДР). На основе полученных выражений показана некорректность широко используемых предположений о постоянстве атомарных параметров гидратации для описания взаимодействий белка с водой.
2. Получены аналитические выражения, позволяющие на основе площади ПДР полярных и заряженных атомов белка эффективно вычислять свободные энергии образования ими водородных связей с молекулами воды.
3. Разработан новый непараметрический метод приближённого расчёта площади ПДР, основанный на рекурсивной процедуре попарного объединения атомов, и показана его высокая точность, не зависящая от химической природы и конформации рассматриваемых соединений.
4. Получены аналитические выражения для первых производных площади ПДР по координатам атомов молекулы, позволяющие использовать разработанный метод в процедурах глобальной минимизации.

Работа была поддержана грантами: Санкт-Петербургского научного центра РАН (в период 2003-2006 гг.); Правительства Санкт-Петербурга для молодых учёных (в период 2004-2007 гг.); Российского министерства образования и науки (РНП 2.2.1.1.4663); РФФИ (07-04-00785-а).

Список работ, опубликованных по теме диссертации

1. Rychkov G. and Petukhov M. Joint neighbors approximation of macromolecular solvent accessible surface area // *J. Comp. Chem.*, – 2007. – Vol. 28, N. 12. – P. 1974-1989.
2. Рычков Г. и Петухов М. Модели гидратации биополимеров // Бреслеровские чтения: Молекулярная генетика, биофизика и медицина; под. ред. В.А. Ланцова, – 2007 – Выпуск 2. – *В печати.*
3. Petukhov M., Rychkov G., Firsov L. and Serrano L. H-bonding in protein hydration revisited // *Protein Science*, – 2004. – Vol. 13, N. 8. – P. 2120-2129.
4. Петухов М., Рычков Г. и Сerrано Л. Теоретическое моделирование гидратации белков // Бреслеровские чтения: Молекулярная генетика, биофизика и медицина сегодня; под. ред. В.А. Ланцова, – 2002. – Выпуск. 1. – СПб.: Издательство ПИЯФ РАН, – С. 148-168.
5. Рычков Г.Н. и Петухов М.Г. Расчёт площади поверхности биополимеров, доступной растворителю, с помощью объединения пар атомов // Материалы XIV Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов», – Москва. – 2007. – Том 1. – С. 51.
6. Рычков Г.Н. и Петухов М.Г. Новый непараметрический метод приближительного вычисления площади поверхности биополимеров, доступной растворителю (ПДР) // Конференция «Современные проблемы науки о полимерах», – Санкт-Петербург. – 2007. – С. 84.
7. Рычков Г.Н. и Петухов М.Г. Расчёт вероятностей образования водородных связей при гидратации белка на основе поверхности, доступной растворителю // «Биология – наука XXI века»: 9-я международная Пущинская школа-конференция молодых учёных. Сборник тезисов, – Пущино. – 2005. – С. 330.
8. Рычков Г.Н. и Петухов М.Г. Новый метод приближительного расчёта площади поверхности белка, доступной растворителю // «Биология – наука XXI века»: 8-я международная Пущинская школа-конференция молодых учёных. Сборник тезисов, – Пущино. – 2004. – С. 242.
9. Рычков Г.Н. и Петухов М.Г. Новый метод расчёта невалентных взаимодействий растворителя с внутренними атомами глобулярных белков // «Молодые учёные промышленности Северо-западного региона»: материалы семинаров политехнического симпозиума, – Санкт-Петербург. – 2004. – С. 72.
10. Rychkov G. and Petukhov M. Based on approximate accessible surface model of protein hydration // 13th European Students' Conference. Abstract book, – Humboldt University, Berlin. – 2002. – P. 21.