

УДК 532.74; 577.15
© Коллектив авторов

ГИДРАТАЦИОННЫЙ МИКРОСКОП. ИНТЕРНЕТ-РЕСУРС ДЛЯ ВИЗУАЛИЗАЦИИ МИКРОСКОПИЧЕСКОЙ СТРУКТУРЫ ГИДРАТАЦИИ МАКРОМОЛЕКУЛ

Д. А. Тихонов¹, С. В. Павлышев², А. Н. Павлов³

¹Институт математических проблем биологии РАН,
142290 г. Пущино Московской области, Институтская 2, ИМПБ РАН
тел: 309-24-75, факс: 229-81-05, эл. почта: info@rismproteins.org

²LuckaSoft 46395 Bocholt, Germany

³ГНИИ ИТТ "Информика", г. Москва

Создан web-сервер, осуществляющий визуализацию данных по теоретическим расчетам гидратации макромолекул. Сервер предоставляет пользователю изображения атом-атомных корреляционных функций гидратации макромолекул. Корреляционные функции предварительно получены путем решения интегральных уравнений теории жидкости в приближении RISM (Reference Interaction Site Model). Дополнительная информация в виде таблиц парциальных атомных вкладов в термодинамические величины гидратации и объемная визуализация макромолекулы позволяет пользователю выбрать интересующие его функции в диалоговом режиме. В настоящее время сервер содержит данные о корреляционных функциях 926 пептидов с максимальным числом атомов 5124.

Ключевые слова: Интегральные уравнения теории жидкостей, приближение RISM, равновесные корреляционные функции, гидратация, пептиды.

ВВЕДЕНИЕ. Теоретический анализ водного окружения макромолекулы является важным элементом для понимания специфики биомолекулярных реакций. Одним из перспективных методов для изучения гидратации является метод интегральных уравнений теории жидкости в приближении RISM [1]. Метод позволяет получать атом-атомные корреляционные функции для всех атомов биомолекулы с атомами молекулы воды [2,3]. Сравнение корреляционных функций гидратации, полученных методом молекулярной динамики для пептида мет-энкефалина в работе [4], с теми же функциями, полученными методом RISM [5], выявляет хорошее согласование [4].

Вместе с тем метод молекулярной динамики подразумевает значительные затраты для вычисления корреляционных функций, поскольку для этого необходима обработка большого количества статистического материала, а, следовательно, и динамическая траектория при накоплении корреляционных функций. В методе RISM корреляционные функции получаются как решения системы $2 \cdot N_{at}$ нелинейных интегральных уравнений, которая может быть сведена к системе $2 \cdot N_{at} \cdot N$ нелинейных алгебраических уравнений, где N_{at} - число атомов в макромолекуле, N - число точек дискретизации корреляционной функции. Геометрия макромолекулы и параметры потенциала взаимодействия молекулы с атомами воды являются параметрами системы уравнений. Детали метода, математические выражения и алгоритм решения уравнений описаны в [2,3].

В отличие от метода молекулярной динамики геометрия молекулы предполагается жесткой, однако, как отмечается в работе [4], это лишь незначительно влияет на вид корреляционных функций. Исходя из корреляционных функций, можно получить все термодинамические величины, характеризующие гидратацию макромолекулы, в том числе и свободную энергию гидратации. Оценка энергии гидратации в рамках RISM подхода признана наиболее точной в ряду существующих эмпирических оценок, основанных на расчете доступной растворителю поверхности или методе масштабной частицы.

В физическом смысле равновесная корреляционная функция есть ни что иное как (с точностью до нормировок) плотность вероятности обнаружения некоей частицы на расстоянии r при условии, что некоторая другая частица находится в начале координат. В случае задачи по гидратации, корреляционные функции есть плотности вероятности обнаружения атома кислорода или водорода воды на заданном расстоянии от данного атома макромолекулы. Визуальный анализ корреляционных функций позволяет нам судить о степени гидратированности того или иного атома в макромолекуле.

На рисунке 1 приведены корреляционные функции атомов азота белкового остова с атомами кислорода молекул воды в зависимости от порядкового номера аминокислотного остатка для всех 40 аминокислотных остатков относительно небольшого пептида 1AML. (Обозначение по классификации, принятой в Protein Data Bank). Расчеты выполнены по методу RISM. Хорошо видно, что корреляционная функция сильно меняется как в зависимости от типа боковой группы аминокислоты, так и от конформации самого пептида. Действительно, для двух альфа-спиральных участков пептида в районе 14-24 и 32-35 аминокислот корреляционные функции сильно отличаются. Это говорит о том, что вторичная структура белка существенным образом влияет на гидратацию. Точно так же отличаются друг от друга корреляционные функции на других участках пептида, несмотря на то, что потенциал взаимодействия атомов азота с водой выбирался одинаковым для всех атомов азота полипептидной цепи. Существенное влияние на корреляционные функции оказывают боковые группы аминокислот, что отражает факт зависимости гидратации и от первичной структуры белка. Анализ корреляционных функций гидратации от всех атомов макромолекулы способен дать новую информацию о поведении белка в водном окружении.

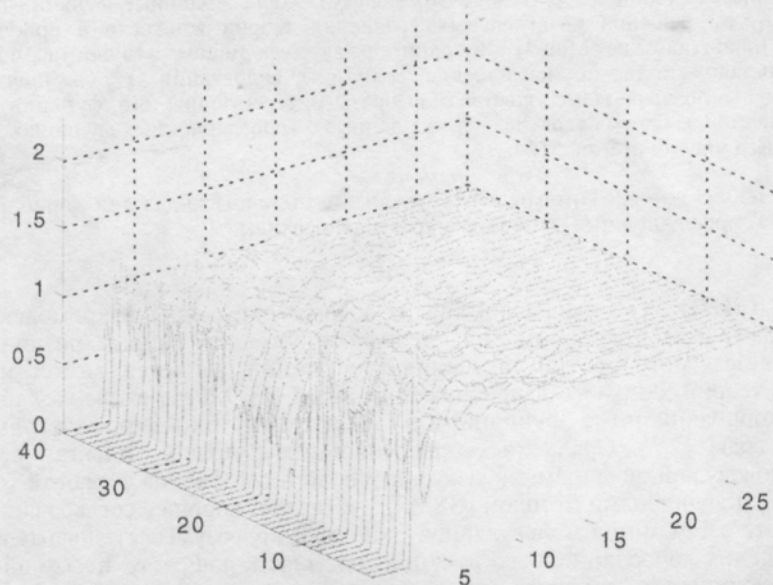


Рисунок 1.

Корреляционные функции для атомов азота белкового остова амилоидного пептида 1AML (Nat=601) с атомами кислорода воды.

Нами разработан универсальный подход для анализа термодинамических и структурных свойств гидратации биополимеров, а также создана и поддерживается база данных по расчетам гидратации [6]. Однако, вследствие того, что количество информации, получаемой в результате расчетов, очень велико, потребовалось изменение характера взаимодействия пользователя с базой данных. В частности, представляется нецелесообразным хранить в базе данных изображения всех корреляционных функций, поскольку это предъявляет нереальные требования к дисковому пространству и сильно увеличивает интернет-трафик. Поэтому в базе данных [6] отсутствует информация обо всех корреляционных функциях, вместо этого приведено краткое резюме лишь наиболее значимых.

Но предугадать заранее, какие из корреляционных функций заинтересуют пользователя, невозможно, поэтому доступ должен быть реализован ко всем этим функциям. Это тем более важно, если база данных используется в образовательных целях.

Для исключения противоречия между полнотой базы данных, с одной стороны, и быстротой взаимодействия пользователя с одновременным сохранением дискового пространства, с другой, мы предлагаем новый подход к представлению результатов, основанный на динамическом формировании страницы с использованием инструментов сетевого программирования. Нами создан веб-сервер <http://rismproteins.org>, при помощи которого пользователь получает доступ к изображениям всех интересующих его корреляционных функций гидратации выбранной макромолекулы. Физически сервер размещен в Европе, имеет малое время отклика, что позволяет осуществлять взаимодействие с ним в режиме реального времени. Далее мы приведем краткое описание работы базы данных и его графической оболочки взаимодействия с пользователем.

1. Описание работы сервера

Мы опишем только ту часть работы сервера, которая отвечает за отображение атом-атомных корреляционных функций пептидов. Остальные функции сервера являются подмножеством тех, которые описаны в работе [6].

Корреляционные функции всех представленных на сервере пептидов предварительно рассчитаны, результаты расчетов загружены на сервер в виде текстовых файлов, содержащих численные значения корреляционных функций. Численные значения представлены с точностью до одной сотой величины функции, чего вполне достаточно для создания качественного изображения ее графика. Каждому пептиду соответствует два файла, в которых построчно размещены значения корреляционных функций с атомом кислорода и с атомом водорода воды. Номер строки в файле соответствует номеру атома в пептиде. В целях экономии дискового пространства файлы сжаты до 10-15 процентов первоначального размера. Побочным положительным эффектом сжатия является более быстрое чтение файла, поскольку его загрузка требует меньше времени.

Для построения графика корреляционной функции разработана программа на базе Practical Extraction and Report Language (PERL). Отображение картинки осуществляется следующим образом: при выборе пользователем интересующего атома из таблицы атомных свойств, на вход программы передается номер строки, совпадающий с номером атома в молекуле, содержащей данные, необходимые для построения графика. Программа открывает 2 файла, содержащие значения корреляционных функций (для атома водорода и кислорода воды) из текущей директории, находит строки с затребованным номером и присваивает строковым переменным скрипта значения корреляционных функций, содержащиеся в найденных строках. Далее программа заполняет массивы данными строковых переменных (которые представляют собой N численных значений, разделенных пробелами), сортирует их для нахождения наибольших и наименьших значений переменной, рассчитывает на основе конфигурационных данных фактор масштабирования, создает массивы, содержащие целочисленные точки графика (номера пикселей), и на основании этих данных генерирует HTML-документ. Кроме того, на вход программы передается строка с дополнительной информацией об атоме, которая выводится на экран в качестве пояснения к графику и содержит классифицирующее название атома в рамках выбранного силового поля.

В настоящее время конфигурация программы (задание желаемых отступов от краев графиков, масштаб по осям, прочая дополнительная информация) производится путем редактирования конфигурационного файла. В перспективе мы планируем создание веб-интерфейса, который позволит пользователю самому настроить желаемый вид графика.

2. Описание интерфейса взаимодействия с пользователем

Здесь мы опишем основные страницы сервера, реализованные к настоящему моменту, а также дадим необходимые пояснения по взаимодействию пользователя с базой данных. Первая страница сервера содержит следующие элементы навигации: окно поиска пептида по его PDB-коду (простой и расширенный); ссылка на список всех пептидов содержащихся в базе, с дополнительными возможностями сортировки; ссылка на вводную часть, содержащую краткое описание ресурса, ссылка на основные математические выражения и числовые значения параметров использовавшихся при всех расчетах. Главная страница сервера содержит также инструмент обратной связи с возможными пользователями ресурса, который реализован в виде формы запроса, в которой пользователь может оставить свои комментарии о ресурсе, а также запрос на расчет нового пептида, если его нет в базе. Имеется также ссылка на контактную информацию разработчиков данного ресурса. Во вводной части имеются ссылки на сетевую страницу, содержащую параметры силового поля, используемого при расчетах, ссылки на программу, при помощи которой были предварительно оптимизированы структуры пептидов, краткое описание расчетного алгоритма, а также термодинамические параметры расчета.

Основной страницей меню базы, с которой пользователь попадает на страницу с информацией об интересующей его молекуле, является список пептидов. Список пептидов - динамически формируемая страница, содержащая пептиды, отвечающие заданному критерию поиска, или по непосредственной ссылке. Он представлен в виде таблицы, в которой приведены некоторые расчетные термодинамические характеристики молекул. По всем величинам допускается сортировка, для этого достаточно нажать на название свойства в верхней части таблицы. Первая колонка таблицы - порядковый номер пептида в базе, он имеет гиперссылку, которая открывает соответствующую страницу Protein Data Bank. Вторая колонка таблицы содержит PDB-код молекулы, он является ссылкой на страницу сервера с информацией о гидратации пептида. Далее мы рассмотрим одну из таких страниц сервера.

На рисунке 2 изображено окно Интернет-браузера с загруженной страницей для пептида 1A11 по классификации, принятой в Protein Data Bank. Страница состоит из 5 основных фреймов. Самый верхний фрейм содержит заголовок молекулы со ссылкой на Protein Data Bank, краткое название, а также навигацию для возврата к полному списку (таблице) пептидов. Левый верхний фрейм отображает трехмерную структуру молекулы. Он реализован аналогично [6]. Правый верхний фрейм - место для отображения графика корреляционных функций выбранного атома. Нижний левый фрейм - список всех аминокислот пептида, оформленный в виде таблицы парциальных термодинамических вкладов аминокислотных остатков в полную величину гидратации. Данный фрейм содержит навигационные элементы, которые позволяют перейти к детальной таблице атомных вкладов, расположенной в правом нижнем фрейме. Нижние фреймы - динамически формируемые страницы, они допускают сортировку всех величин в таблицах, аналогично тому, как это реализовано в списке пептидов.

Общая схема взаимодействия пользователя с сервером выглядит следующим образом: анализируя трехмерную структуру молекулы, пользователь решает, какой аминокислотный остаток пептида ему интересен. Затем в таблице аминокислотных остатков пользователь находит соответствующую аминокислоту и переходит к таблице атомарных вкладов. Таблица атомарных вкладов содержит ссылку, которая связана с номером атома в молекуле. При нажатии на этот номер происходит отображение графика корреляционной функции. Красная линия графика соответствует корреляционной функции атома водорода молекулы воды с данным атомом, синяя - атома кислорода.

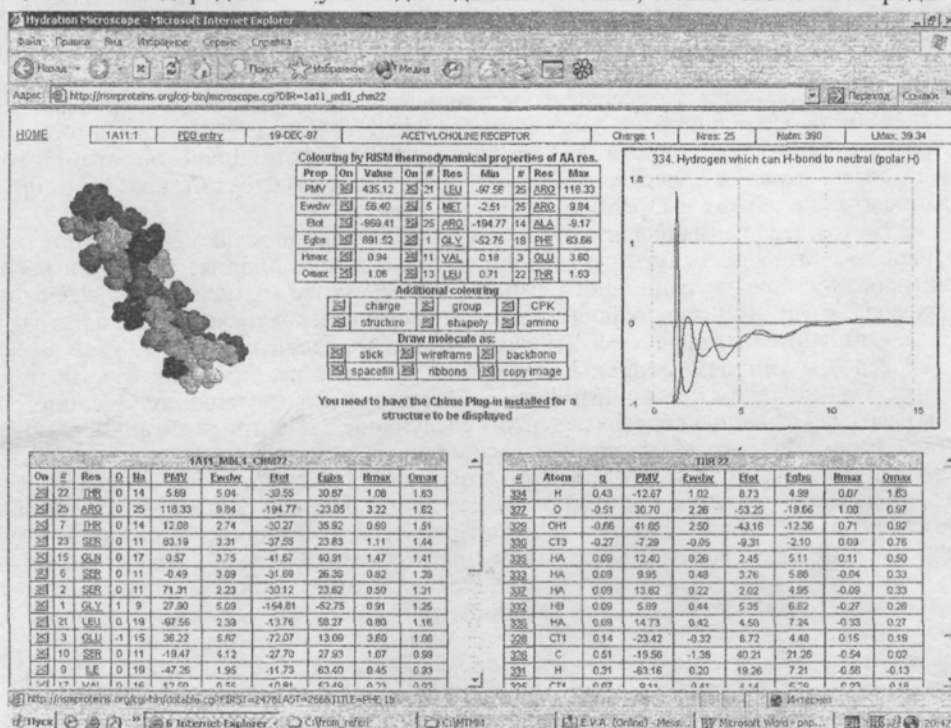


Рисунок 2.
Страница сервера <http://rismproteins.org> для пептида 1A11

ЗАКЛЮЧЕНИЕ. Данная работа является первой попыткой представить расчетную информацию о микроскопической структуре гидратации макромолекул на систематической основе. В большой степени этому способствует сам выбранный подход, поскольку в отличие от методов статистического моделирования, результаты расчетов в методе интегральных уравнений не зависят от длины траектории, по которой проводится усреднение, и выбранной начальной конфигурации, поэтому сравнения могут проводиться как в ряду различных белковых молекул, так и в ряду конформаций одной молекулы.

К настоящему моменту получены решения уравнений RISM для гидратации глобулярных белков с числом атомов в несколько тысяч. Посчитаны такие белки, как лизоцим, иммуноглобулин, рибонуклеаза. В перспективе планируется разработка и внедрение автоматического формирования базы данных по пользовательскому запросу, то есть предполагается, что пользователь сможет сам выполнить расчет корреляционных функций и термодинамических величин, переслав на сервер необходимые данные о геометрии, интересующей его молекулы и о типе силового поля. Сервер будет автоматически формировать соответствующую страницу результатов, делая ее доступной и для других исследователей. Разумеется, такой подход потребует организации вычислительного кластера, дальнейшего улучшения эффективности расчетной программы, поиска новых методов хранения и представления данных о расчете. Однако он позволит многократно увеличить объем информации о гидратации белков, что в конечном итоге, будет способствовать более глубокому пониманию поведения белковых макромолекул в растворе.

Авторы благодарны за предоставленный веб-хостинг фирме LuckaSoft GmbH. Авторы благодарны Масахиро Киношита за предоставленные данные, которые помогли в тестировании метода. Один из нас (Т.Д.А.) благодарен за финансовую поддержку исследований Е.Л. Никифорову.

ЛИТЕРАТУРА

1. Chandler D., Andersen H.C. (1972) J. Chem. Phys., **57**, 1930-1937.
2. Tikhonov, D. A. et al. (1998) J. Chem. Phys., **109**, 1528-1539.
3. Тихонов Д. А. (2002) В кн.: Компьютеры и суперкомпьютеры в биологии (В.Д.Лахно и М.Н.Устинина, ред.) Москва-Ижевск: Институт компьютерных исследований, с. 209-233
4. Dudowicz J., Freed K.F., Min-yi Shen (2003) J. Chem. Phys., **118**, 1989-1995
5. Kinoshita M., Okamoto Y., Hirata F. (1997) J. Chem. Phys., **107**, 1586-1599.
6. Тихонов Д.А., Павлов А.Н., Тюльбаева Г.Э., Устинин М.Н., Лахно В.Д. (2003) Интернет-конференция ИВТН <http://www.ivtn.ru/2-session/enter/paper.phtml?r=32>
База данных: <http://www.jcibi.ru/Tikhonov/index.shtml>

HYDRATION MICROSCOPE: AN INTERNET RESOURCE FOR MICROSCOPIC STRUCTURE VISUALIZATION OF THE HYDRATION OF MACROMOLECULES

D.A. Tikhonov¹, S.V. Pavlishev², A.N. Pavlov³

¹IMPB RAS, Institutskaya str. 2, Puschino, Moscow Region, Russia;
e-mail: info@rismproteins.org; tel.: +7 095 3092475; fax: +7 095 2298105

²LuckaSoft 46395 Bocholt, Germany

³GNII ITT "Informika"

A web-server has been created to visualize results of theoretical calculations of macromolecule hydration. The server displays to the user the graphs of atom-atom correlation functions of macromolecule hydration. The correlation functions are obtained by solving integral equations of the theory of liquids in RISM approximation. The 3D-visualization of the macromolecule and supplementary information, available as tables of partial atomic contributions to thermodynamic hydration values, allows the user to choose functions of interest in dialogue mode. By now the server stores the data on correlation functions of 926 peptides, with the maximum number of atoms of 5124.

Keywords: Integral equation theory of liquids, RISM approximation, equilibrium correlation function, hydration, peptides.