

УДК 574/578

## Предсказание структуры комплексов белок-лиганд: от компьютерной модели к биологической функции

Ю. А. Косинский, Т. В. Пырков, С. В. Луценко, Р. Г. Ефремов

*ЮРИЙ АНЗЕЛЬМОВИЧ КОСИНСКИЙ* — младший научный сотрудник лаборатории структурной биологии Института биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН (ИБХ РАН). Область научных интересов: структурная биология, физика белка, биоинформатика.

*ТИМОФЕЙ ВЛАДИМИРОВИЧ ПЫРКОВ* — аспирант МФТИ. Область научных интересов: структурная биология, взаимодействия белок—липид.

*СВЕТЛАНА ВИКТОРОВНА ЛУЦЕНКО* — профессор Отделения биохимии и молекулярной биологии Орегонского университета, Портленд, США. Область научных интересов: биохимия, структурная биология.

*Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland OR, 97239, USA*

*РОМАН ГЕРБЕРТОВИЧ ЕФРЕМОВ* — доктор физико-математических наук, ведущий научный сотрудник лаборатории структурной биологии ИБХ РАН. Область научных интересов: биомембраны, структура и функции мембранных белков, биофизика.

*117997 Москва, В-437, ул. Миклухо-Маклая, 16/10, ИБХ РАН, тел. (495) 336-20-00, E-mail ykos@nmr.ru*

### Введение

Для понимания молекулярного механизма работы белков-ферментов необходим анализ их пространственной структуры атомного разрешения. Более того, часто нужна информация о пространственных структурах белка в разных функциональных состояниях: в комплексе с субстратами или их аналогами, ингибиторами и др. Между тем определение структуры белков с помощью экспериментальных методов во многих случаях представляет значительную проблему, требующую больших технических, временных и финансовых затрат. Наибольшие трудности возникают при структурных исследованиях мембранных белков, к которым относится множество важнейших ферментов, рецепторов и транспортеров, и только для нескольких типов мембранных белков экспериментально установлены пространственные структуры высокого разрешения. В то же время благодаря успехам геномных исследований для большинства биологически важных белков расшифрована аминокислотная последовательность.

В настоящее время широкое применение находит компьютерный метод построения моделей пространственных структур белков на основании гомологии с аминокислотными последовательностями родственных белков, для которых пространственная структура известна. Дальнейшее исследование трехмерных моделей белков с помощью методов молекулярного моделирования, таких как молекулярная динамика и докинг лигандов, позволяет эффективнее анализировать экспериментальные данные, осуществлять компьютерный дизайн точечных мутаций белков, а также выдвигать различные гипотезы о структуре и функции белка.

Одной из важных и актуальных задач в области структурной биологии является изучение мембранных белков из семейства АТФаз Р-типа [1]. Ферменты

этого обширного и широко распространенного в живых организмах семейства осуществляют активный транспорт специфичных для них катионов за счет энергии гидролиза АТФ. Для наиболее изученного представителя АТФаз Р-типа — саркоплазматической  $\text{Ca}^{2+}$ -АТФазы в последние годы получен ряд кристаллических структур высокого разрешения в различных функциональных состояниях, в том числе в комплексе с аналогами АТФ. Эти структуры используются в качестве основы («шаблона») для построения моделей пространственной структуры других АТФаз Р-типа на основании гомологии. Например, предложены трехмерные модели таких важнейших ферментов, как  $\text{Na}^+, \text{K}^+$ -АТФаза [2] и  $\text{H}^+$ -АТФаза плазматической мембраны [3].

Настоящая работа посвящена построению трехмерных моделей двух цитоплазматических нуклеотидсвязывающих доменов (так называемых N- и Р-доменов) медь-транспортирующей АТФазы человека и изучению их структуры, динамики и взаимодействий с АТФ с помощью методов молекулярного моделирования. Это исследование представляет не только фундаментальный интерес, но имеет также важное значение для биомедицинской практики. Мутации гена, кодирующего данный фермент, вызывают летальный гепато-неврологический синдром — болезнь Вильсона [4]. При этом заболевании многие и наиболее часто встречающиеся мутации гена локализируются в N- и Р-доменах и являются причиной нарушения связывания АТФ ферментом и/или процесса каталитического фосфорилирования. Выяснение молекулярного механизма проявления этих мутаций и выявление функционально важных аминокислотных остатков медь-транспортирующей АТФазы, отвечающих за связывание АТФ, необходимо для поиска путей лечения этого наследственного заболевания.

### Молекулярное моделирование структуры и динамики белков и комплексов белок-лиганд

Молекулярное моделирование можно определить как процесс описания физико-химических и биологических систем в рамках достаточно реалистичных атомных моделей, проводимый с целью понять и предсказать макроскопические свойства, основываясь на закономерностях поведения системы на молекулярном уровне. Молекулярное моделирование находит применение в различных сферах — это биомедицинские исследования, решение задач физической химии, разработка новых материалов и др. Быстрому развитию и совершенствованию этих методов способствует рост производительности и доступности вычислительной техники.

Выбор методов, используемых при моделировании, зависит от поставленной задачи, а также от возможности получить корректный результат на заданном уровне идеализации.

Как известно, наиболее точное описание молекулярной системы можно получить решением временного уравнения Шрёдингера. Однако это возможно лишь для систем, состоящих из нескольких атомов, а для практически всех молекулярных систем, интересных с биологической точки зрения, необходимо строить приближенные модели. При этом чем больше атомов в системе и чем больше характерное время исследуемых процессов, тем более упрощенные модели используют для ее описания. Как правило, при моделировании биомолекул (в особенности, биополимеров) квантовомеханический подход заменяют эмпирическим расчетом энергии атом-атомных взаимодействий — молекулярной механикой. В том случае, когда даже расчеты молекулярной механики неприменимы из-за сложности системы (например, до сих пор нереально моделировать процесс самоорганизации пространственной структуры для большинства белковых молекул), молекулярное моделирование осуществляют на основе сравнительного анализа доступной структурной и физико-химической информации. К этой категории принадлежат метод предсказания биологической активности низкомолекулярного вещества по его химической структуре (QSAR, Quantitative Structure-Activity Relationship) и метод предсказания пространственной структуры белков на основании гомологии.

Рассмотрим современные методы молекулярного моделирования, используемые в структурно-функциональных исследованиях биологических систем.

#### Метод молекулярной динамики

Функционирование белков определяется не только их структурой, но и динамическими свойствами. Для изучения конформационной подвижности белков в окружении растворителя применяется метод молекулярной динамики (МД). Кратко изложим основы этого метода.

Алгоритм метода молекулярной динамики производит решение уравнений движения классической механики Ньютона для системы из  $n$  взаимодействующих атомов:

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i(\vec{r}), \quad i = 1 \dots n$$

где  $m_i$  и  $\vec{r}_i$  — масса и радиус-вектор  $i$ -го атома, соответственно;  $t$  — время;  $\vec{F}_i$  — сила, действующая на  $i$ -й атом со стороны остальных частиц

$$\vec{F}_i = -\frac{\partial U(\vec{r})}{\partial \vec{r}_i}, \quad \text{где } \vec{r} = \{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n\};$$

$U(\vec{r})$  — классическое потенциальное поле, определяемое координатами всех атомов системы.

Решение уравнений движения производится численно. В ходе моделирования динамики в системе поддерживаются заданные значения температуры и/или давления, координаты атомов сохраняются в выходном файле через заданные интервалы времени. Траектория системы представляет собой значения координат атомов системы и их скорости как функции времени. После первоначальных изменений система обычно приходит в некоторое динамическое равновесное состояние (по определенному набору параметров). Используя принцип усреднения по равновесной траектории, многие макроскопические свойства системы (например, энергия взаимодействия молекул) и структурные характеристики можно оценить при обработке данных выходного файла.

Важно помнить об ограничениях метода молекулярной динамики в приложениях к биомакромолекулам. Следует иметь в виду, что способ определения и расчета потенциального поля, основанный на принципах молекулярной механики, может оказаться слишком грубым приближением для некоторых задач моделирования. Так, оценки энергий различных конформационных состояний молекулы (например, конформаций «спираль» и «клубок» для полипептидной цепи), полученные из экспериментальных данных и молекулярно-динамических расчетов, могут показать значительные расхождения. Кроме того, следует учитывать и еще одно существенное обстоятельство. Характерные времена многих важных для изучения динамических событий в белках имеют порядок микро- и даже миллисекунд, что обусловлено необходимостью преодолевать значительные энергетические барьеры в ходе динамических преобразований системы. В то же время из-за большого объема вычислений при моделировании молекулярной динамики системы белок-растворитель в большинстве случаев исследователь имеет возможность получить траекторию лишь на временах порядка  $10^4$ – $10^2$  нс.

Тем не менее, как показали многочисленные исследования, метод МД позволяет получить представление о свойственных изучаемому белку конформационных переходах. Как будет показано ниже, анализ взаимодействий белок-лиганд можно эффективно проводить на основе ансамбля структур белка, полученных с помощью метода МД. В большой степени успех вычислительного МД-эксперимента зависит от выбора потенциального поля (существуют различные аналитические выражения для потенциальной энергии и различные библиотеки параметров силового поля — GROMOS, CHARMM, OPLS и др. [5–7]), а также от выбора вариантов реализации и параметров алгоритма метода МД.

В нашей работе для расчета динамики цитоплазматических доменов АТФаз Р-типа мы использовали программу GROMACS [8] и потенциальное поле GROMOS87 [5]. Для учета фактора взаимодействия

белка с растворителем была выбрана модель явно заданного растворителя (хотя это многократно увеличивает объем вычислений). Молекулу белка помещали в прямоугольную ячейку, заполненную молекулами воды; поддерживали температуру на уровне ~300 К и давление ~1 атм, используя алгоритм Берендсена. Эффекты на границах ячейки компенсировали с помощью метода периодических граничных условий.

Наиболее полную информацию об алгоритмах и возможностях пакета GROMACS можно получить на интернет-сайте <http://www.gromacs.org>.

### Метод молекулярного докинга

В задачах по предсказанию структуры комплекса белок-лиганд молекулярный докинг является одним из наиболее востребованных методов молекулярного моделирования. Так, для выявления среди тысяч или даже миллионов соединений прототипов-лекарств, обладающих биологической активностью и действующих на определенный белок-мишень, используют автоматизированные процедуры высокопроизводительного скрининга, сокращающие число соединений-кандидатов на 3–4 порядка [9]. Резко возросшие в последнее десятилетие мощности компьютеров и появление новых вычислительных методов позволяют проводить виртуальный аналог скрининга — высокопроизводительный докинг, в ходе которого лиганды «встраивают» в активный центр белковой молекулы. Таким образом на компьютере (*in silico*) моделируется взаимодействие огромного числа лигандов с белком [10].

В алгоритмах докинга в качестве стартовой информации используют пространственную структуру белка-рецептора (как правило, конформационно неподвижную) и структуру лиганда, конформация которого и расположение его относительно активного центра рецептора оптимизируются в процессе докинга. Результатом докинга является набор конформаций лиганда, взаимодействующих с белковым сайтом связывания наилучшим образом с точки зрения оценочной функции докинга. Подобные эмпирические оценочные функции отражают различные аспекты межмолекулярных взаимодействий в комплексе белок-лиганд. Здесь можно провести аналогии между оценочной функцией и свободной энергией связывания лиганда белком. Существующие современные программы докинга (например, DOCK [11] и GOLD [12]) используют разные подходы для поиска конформаций лиганда (эвристический поиск с локальной минимизацией и генетический алгоритм, соответственно) и их проверки с помощью оценочных функций (модифицированное силовое поле AMBER и GoldScore/ChemScore, соответственно). Детальное сравнение этих программ на тестовом кристаллографическом наборе комплексов белок-лиганд приведено в обзоре [13].

Расчет значений оценочной функции должен занимать минимальное компьютерное время, так как при высокопроизводительном докинге исследуется взаимодействие до миллиона соединений с белком-мишенью. Для интенсификации процесса скрининга в оценочные функции, используемые в программах докинга, часто включают упрощенные эмпирические термы. Значения оценочных функций, как правило, слабо коррелируют с экспериментально определяемыми значениями свободной энергии связывания. Более

того, во многих случаях «правильные» комплексы белок-лиганд (т.е. согласующиеся с экспериментально установленными структурами), полученные методом докинга, характеризуются худшими значениями оценочной функции по сравнению с заведомо «неправильными» комплексами. В связи с этим при высокопроизводительном докинге, когда отбирается только один комплекс с лучшим значением оценочной функции, есть риск потерять перспективное соединение — прототип лекарства, так что необходимы дополнительные критерии и/или экспериментальная информация для выбора наиболее корректных результатов докинга.

В нашей работе для проведения процедуры докинга АТФ в структуры нуклеотид-связывающих доменов АТФаз Р-типа использовалась программа GOLD [12]. В качестве оценочной функции была выбрана GoldScore, которая включает термы, отражающие ван-дер-ваальсовы взаимодействия между атомами белковой молекулы и лиганда, а также между различными группами лиганда, и водородные связи, образующиеся в комплексе между белком и лигандом.

В заключение обратим внимание на необходимость критически анализировать результаты применения методов молекулярного моделирования в каждой конкретной задаче. Нужно убедиться, что используемые методы и протоколы приводят к корректным, не противоречащим экспериментальным результатам. В нашей работе, прежде чем исследовать модель пространственной структуры цитоплазматических доменов медь-транспортирующей АТФазы, мы проводили аналогичное моделирование кристаллических структур соответствующих доменов  $Ca^{2+}$ -АТФазы. Если результаты моделирования показывают согласие с экспериментальными данными для  $Ca^{2+}$ -АТФазы, то есть основания использовать эти подходы для предсказания свойств трехмерных моделей изучаемых белков. При этом, однако, надо помнить, что точность модели значительно меньше по сравнению с кристаллической структурой, и большое значение имеет сопоставление результатов моделирования с экспериментальными данными, например, с данными точечного мутагенеза.

### Исследование структурно-функциональных аспектов взаимодействия АТФаз Р-типа с АТФ с помощью методов молекулярного моделирования

*«Калибровочная задача»: от пространственной структуры  $Ca^{2+}$ -АТФазы к теоретической модели ее комплекса с АТФ*

В структурах АТФаз Р-типа (рис. 1) выделяют три цитоплазматических домена (домены А, N и Р) и трансмембранный (ТМ) домен, включающий от 7 до 10 (в зависимости от подтипа фермента) гидрофобных  $\alpha$ -спиралей, пронизывающих клеточную мембрану (см. обзор [1]). Из семейства АТФаз Р-типа наиболее изучена в структурном и функциональном плане  $Ca^{2+}$ -АТФаза саркоплазматического ретикула. Этот фермент обеспечивает откачку ионов  $Ca^{2+}$  из цитоплазмы мышечных клеток в полости саркоплазматического ретикула и тем самым выполняет важнейшую роль в осуществлении и регуляции мышечного сокращения.

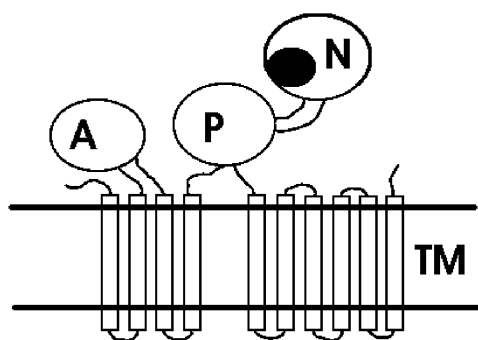


Рис. 1. Схема структурной организации АТФаз Р-типа.

N — нуклеотид-связывающий домен; P — фосфорилируемый домен; A — вспомогательный домен; ТМ — трансмембранный домен, состоящий из 10 α-спиралей. Черным кружком в N-доме показан участок связывания АТФ

В 2000 г. методом рентгеноструктурного анализа была впервые получена структура высокого разрешения этого белка в  $\text{Ca}^{2+}$ -связанной форме, обозначаемой как форма E1 [14]. На основе этой структуры было выяснено строение важнейших функциональных центров фермента: сайтов связывания АТФ и фосфорилирования, которые сформированы в цитоплазматических доменах (A, N, P), и сайта связывания ионов  $\text{Ca}^{2+}$  в ТМ-доме. В дальнейшем была получена кристаллическая структура  $\text{Ca}^{2+}$ -АТФазы в форме E2, в ТМ-доме которой отсутствуют ионы  $\text{Ca}^{2+}$  [15]. Сравнительный анализ структур E1 и E2 показывает, что в каталитическом цикле АТФаз Р-типа происходят весьма значительные перемещения цитоплазматических доменов, сопряженные с изменениями конформации белковой молекулы и взаимного расположения ТМ-сегментов. В связи с этим при переходе фермента из формы E1 в E2 (который сопровождается гидролизом АТФ и каталитическим фосфорилированием фермента по остатку Asp351) структура и свойства сайта

связывания в ТМ-доме изменяются, и сродство к ионам  $\text{Ca}^{2+}$  сильно снижается.

Как было уже сказано, функционирование белков определяется не только их структурой, но и динамическими свойствами. Известно, что в связывании АТФ  $\text{Ca}^{2+}$ -АТФазой одновременно участвуют остатки как N-домена (нуклеотид-связывающий сайт), так и P-домена (сайт фосфорилирования). В структурах  $\text{Ca}^{2+}$ -АТФазы в формах E1 и E2 характерное расстояние между указанными сайтами составляет около 25 и 17 Å, соответственно. Сближение сайтов при переходе E1 → E2 обусловлено доменным движением типа «закрывание», которое происходит в результате конформационных переходов в гибком («шарнирном») участке, соединяющем N- и P-домены. Обозначим конформации АТФ-связывающих доменов  $\text{Ca}^{2+}$ -АТФазы, отвечающие формам E1 и E2, как «открытая» и «закрытая», соответственно. Даже в «закрытой» конформации расстояние между доменами слишком велико, чтобы молекула АТФ могла взаимодействовать с сайтами обоих доменов одновременно. Полученные в 2004 г. кристаллические структуры  $\text{Ca}^{2+}$ -АТФазы в комплексе с аналогами АТФ [16] показали возможность дополнительного «закрывания» доменов, приводящего к необходимому сближению их сайтов.

В нашей работе расчеты молекулярной динамики изолированной пары N- и P-доменов были проведены на временном интервале ~10 нс. В качестве стартовых были взяты кристаллические структуры  $\text{Ca}^{2+}$ -АТФазы в формах E1 и E2 (их траектории обозначим МД(E1) и МД(E2), соответственно). Показано, что структура отдельных доменов остается достаточно стабильной в динамике, но их взаимное расположение значительно изменяется за счет конформационной подвижности «шарнирного» участка. Подобные структурные изменения удобно отражать с помощью междоменного угла  $\Theta$ , вершина которого расположена на «шарнирном» участке, а лучи направлены к центрам масс отдельных доменов (рис. 2А).

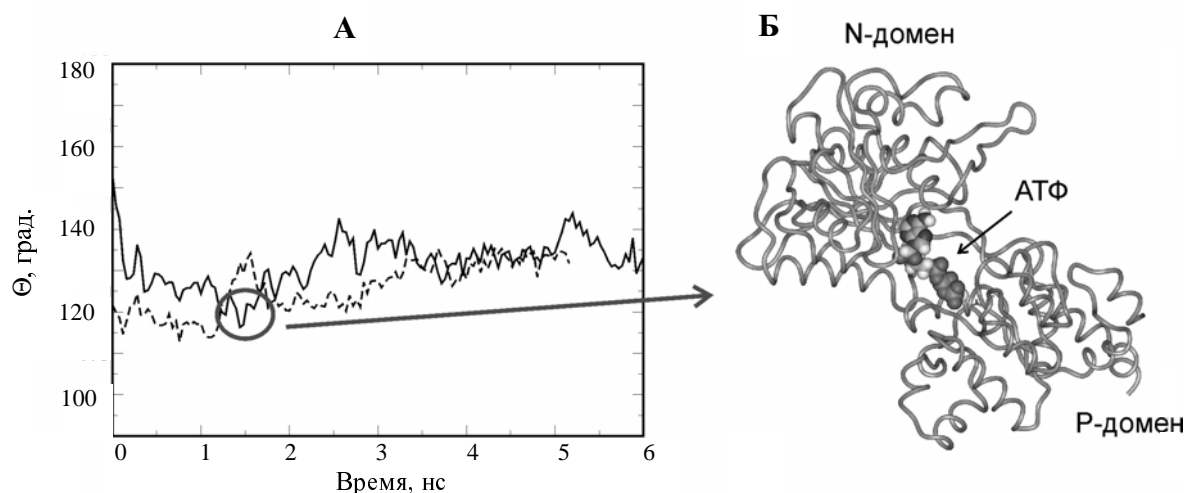


Рис. 2. Результаты модельного изучения связывания  $\text{Ca}^{2+}$ -АТФазой с АТФ:

(А) Изменение междоменного угла  $\Theta$  в ходе МД нуклеотид-связывающих доменов  $\text{Ca}^{2+}$ -АТФазы; зависимости для МД со стартовых структур E1 и E2 показаны сплошной и пунктирной линией, соответственно. (Б) Комплекс нуклеотид-связывающих доменов  $\text{Ca}^{2+}$ -АТФазы с АТФ, полученный в результате докинга в структуру из МД(E1), соответствующую одной из наиболее «закрытых» конформаций доменов ( $\Theta \sim 115^\circ$ ).

На траектории МД(E1) N- и P-домены сближаются в ходе движения типа «закрывание», при этом угол  $\Theta$  достигает минимального значения  $115^\circ$  в момент времени  $\sim 1,5$  нс. На траектории МД(E2), напротив, домены несколько удаляются друг от друга. В результате на каждой из МД-траекторий конформация белка достигает равновесия при небольших колебаниях значения угла  $\Theta$  около  $135^\circ$ . Выявленные доменные движения приводят к существенным изменениям позиции нуклеотид-связывающего сайта в N-доме относительно сайта фосфорилирования в P-доме (остаток Asp351), что вполне согласуется с экспериментальными структурными данными.

Исследование конформационного пространства молекулы  $\text{Ca}^{2+}$ -АТФазы методом молекулярного докинга дало возможность получить представление о связывании АТФ в различных равновесных состояниях. Процедура докинга АТФ была произведена для представительного набора конформеров, выбранных из траектории МД(E1), в которых наблюдаются различные взаимные ориентации N- и P-доменов (значение угла  $\Theta$  варьирует от  $115$  до  $145^\circ$ ). В результате были получены весьма разнообразные ориентации молекулы АТФ относительно белковой молекулы, но для анализа были выбраны только те комплексы, в которых концевая фосфатная группа лиганда (АТФ) удалена от остатка Asp351 в P-доме не более чем на  $10 \text{ \AA}$ , т.е. фосфатный «хвост» молекулы АТФ направлен в сторону сайта фосфорилирования. Критерием для отбора решений докинга служило также наличие в комплексе контактов АТФ с функционально важными остатками N-домена, установленными экспериментально [17].

Результаты докинга АТФ демонстрируют важную роль доменного движения типа «закрывание», которое ведет к сближению сайта связывания N-домена с фосфорилируемым остатком P-домена Asp351. Как видно на рис. 2Б, молекула АТФ, связанная своим

адениновым основанием в сайте N-домена, может вступать в непосредственный контакт с карбоксильной группой остатка Asp351 P-домена, если белковая молекула достигает «закрытой» конформации, соответствующей  $\Theta \approx 115^\circ$ .

В результате отбора были найдены комплексы белок-лиганд (рис. 3А), которые хорошо согласуются с кристаллической структурой комплекса  $\text{Ca}^{2+}$ -АТФазы с аналогом АТФ [16] (рис. 3Б). На рис. 3 видно, что взаимодействия остатков N-домена с АТФ в модельных комплексах и в комплексе, полученном экспериментально, сходны. Азотистое основание и остаток рибозы АТФ, встроенные в полость сайта N-домена, взаимодействуют с известными функционально важными остатками. Так, карбоксильная группа остатка Glu442 может формировать водородную связь с  $\text{NH}_2$ -группой аденинового основания. Боковые цепи остатков фенилаланина Phe487, лизина Lys515, треонина Thr441 и валина Val562 находятся в близком контакте с аденином, а остаток рибозы контактирует с боковыми цепями Lys492 и Val562. При этом наблюдается стэкинг-взаимодействие между бензольным кольцом фенилаланина Phe487 и ароматическим кольцом аденинового основания. Полученная структура комплекса демонстрирует важную роль остатка Arg560, который формирует до трех водородных связей с  $\alpha$ -,  $\beta$ - или  $\gamma$ - $\text{PO}_3$ -группами АТФ и определяет ориентацию фосфатного «хвоста».

На рис. 3 видны также различия между комплексами: 1) в кристаллической структуре N- и P-домены сближены в большей степени, чем в модельной структуре, что обеспечивает большую заглубленность аденинового основания АТФ в сайт N-домена; 2) в модельной структуре белка боковая цепь Arg489 удалена от сайта связывания в отличие от кристаллической структуры, в которой этот остаток стабилизирует фосфатный «хвост» АТФ за счет электростатических взаимодействий.

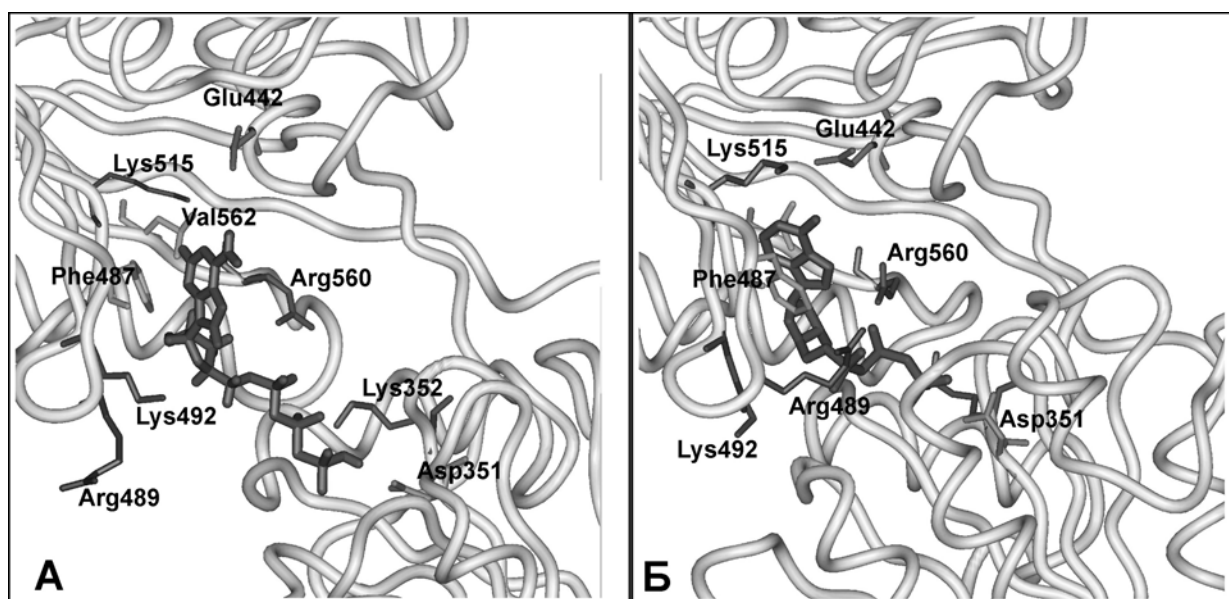


Рис. 3. Структура комплекса саркоплазматической  $\text{Ca}^{2+}$ -АТФазы с АТФ, полученная методом докинга (А), и кристаллическая структура комплекса  $\text{Ca}^{2+}$ -АТФазы с негидролизуемым аналогом АТФ [16] (Б)

Итак, комбинированное использование методов молекулярной динамики и молекулярного докинга позволило корректно предсказать структуру комплекса нуклеотид-связывающих доменов  $\text{Ca}^{2+}$ -АТФазы с АТФ. Проведение независимой процедуры докинга лиганда по отношению к каждому из набора МД-конформеров белка-мишени позволяет учесть его конформационную подвижность и выбрать наиболее достоверные структуры комплексов. В случае связывания АТФ АТФазой Р-типа важно учитывать как перемещения доменов относительно друг друга (изменение взаимного положения сайтов N- и Р-доменов), так и локальные изменения конформации петлевых участков и боковых цепей аминокислотных остатков, формирующих сайты связывания.

Апробированные на  $\text{Ca}^{2+}$ -АТФазе методы и протоколы молекулярного моделирования были использованы для анализа трехмерной модели АТФ-связывающих доменов медь-транспортирующей АТФазы человека.

*«Предсказательная задача»: от теоретической модели  $\text{Cu}^{2+}$ -АТФазы к пониманию функции фермента*

Структура медь-транспортирующей АТФазы человека, обозначаемой далее WNDP (Wilson Disease Protein), значительно отличается от структуры  $\text{Ca}^{2+}$ -АТФазы [18]. Кроме типичных для всех АТФаз Р-типа цитоплазматических доменов (А, N и Р), WNDP содержит дополнительный внутриклеточный сегмент, несущий шесть эквивалентных доменов, связывающих ионы меди, которые, вероятно, имеют регуляторную функцию. Кроме того, трансмембранный домен в WNDP включает 8 (а не 10, как в  $\text{Ca}^{2+}$ -АТФазе) сегментов. Выравнивание аминокислотных последовательностей WNDP и  $\text{Ca}^{2+}$ -АТФазы свидетельствует о гомологии только в области Р-домена и о ее отсутствии в области N-домена. Все это делает построение модели пространственной структуры WNDP по гомологии с  $\text{Ca}^{2+}$ -АТФазой задачей нетривиальной.

Известен ряд мутаций в гене, кодирующем WNDP, проявляющихся как фенотипический признак в избыточном накоплении меди в клетках ряда тканей, в особенности в печени и головном мозге [4]. Механизм влияния большинства мутаций на структуру, функцию и регуляцию фермента WNDP не изучен. Подробно исследована наиболее часто встречающаяся при болезни Вильсона мутация His1069Gln [19–21]. Установлено, что эта мутация нарушает каталитическое фосфорилирование в белке по остатку Asp1027. Остаток N-домена His1069, по-видимому, играет важную роль в координации АТФ в сайте связывания. Однако остается неизвестным, координирует ли остаток His1069 молекулу АТФ непосредственно или его влияние на связывание АТФ опосредовано другими остатками.

Основная цель нашего исследования — идентифицировать аминокислотные остатки WNDP, которые могут быть вовлечены в связывание нуклеотидов. Взаимодействие с АТФ и каталитическое фосфорилирование — важнейшие этапы в механизме транспорта ионов  $\text{Cu}^{2+}$  ферментом. В настоящее время доступной информации об этих процессах очень мало. Трудность заключается в том, что хорошо изученные остатки,

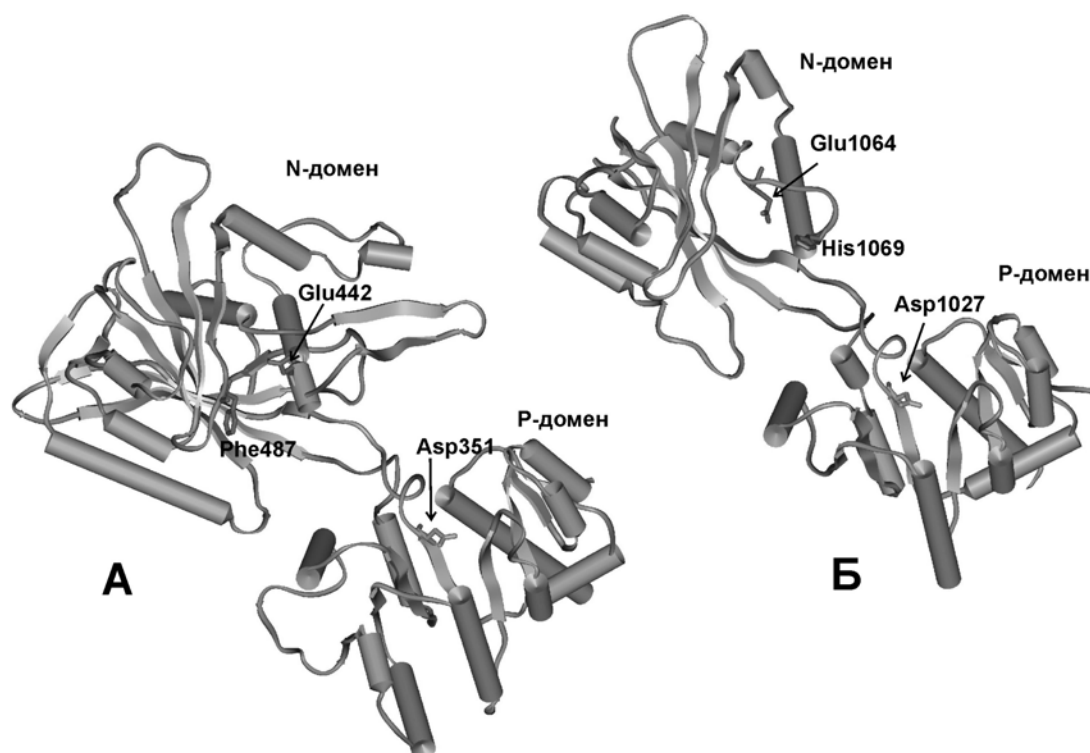
участвующие в связывании АТФ в  $\text{Ca}^{2+}$ -АТФазе, не консервативны в структуре WNDP. Это позволяет сделать вывод о том, что структуры сайтов связывания АТФ в указанных ферментах сильно различаются. Для того чтобы исследовать гипотетические структуры комплексов белка WNDP с АТФ и анализировать экспериментальные данные мутагенеза, необходимо построить модель пространственной структуры этого белка на основании гомологии.

Для построения трехмерной модели нуклеотид-связывающих доменов WNDP нужно выполнить выравнивание его аминокислотной последовательности с соответствующим регионом последовательности  $\text{Ca}^{2+}$ -АТФазы. В последовательности WNDP можно выделить ряд сегментов, для которых наблюдается высокая (с идентичностью ~35%) степень гомологии с  $\text{Ca}^{2+}$ -АТФазой. В структуре  $\text{Ca}^{2+}$ -АТФазы эти участки последовательности принадлежат Р-домену и «шарнирному» участку, соединяющему N- и Р-домены. Участок последовательности WNDP, соответствующий N-домену, напротив, не обнаруживает даже отдаленного сходства с соответствующим сегментом  $\text{Ca}^{2+}$ -АТФазы или какими-либо другими белками с известной пространственной структурой.

Для того чтобы установить возможный тип укладки основной цепи (фолда) в этом участке WNDP, был использован метод «протягивания» аминокислотной последовательности через набор структурных шаблонов [22]. Этот подход использует статистические данные по атом-атомным контактам в белках с известной пространственной структурой. Полученные результаты позволили заключить, что наиболее подходящим структурным шаблоном для модели WNDP является кристаллическая структура АТФ-связывающего фрагмента  $\text{Ca}^{2+}$ -АТФазы в состоянии E1. На основе результатов метода «протягивания» было получено выравнивание аминокислотных последовательностей N- и Р-доменов WNDP и  $\text{Ca}^{2+}$ -АТФазы (см. подробное описание в [23]).

Далее с помощью программы MODELLER [24] на основании кристаллических структур  $\text{Ca}^{2+}$ -АТФазы были получены модели пространственной структуры WNDP как в форме E1 (рис. 4), так и в форме E2. Алгоритм MODELLER формирует набор ограничений на межатомные расстояния в трехмерной модели, приняв за образец расстояния между гомологичными участками в структуре-шаблоне. Затем производится серия оптимизаций пространственной модели с тем, чтобы найти конформацию белка, наилучшим образом соответствующую этому набору ограничений.

Для исследования качества полученных моделей WNDP был применен подход «профиля окружения» остатков, предложенный Д. Айзенбергом [25] и реализованный в программе 3D\_Profiles. В этом методе рассчитываются характеристики микроокружения каждого остатка в трехмерной модели белка. Как и метод «протягивания», метод Айзенберга основан на использовании статистических данных по свойствам микроокружения различных видов аминокислотных остатков в известных белковых структурах. В случае заведомо «неправильной» трехмерной модели многие остатки оказываются в неестественном для данного вида аминокислотных остатков окружении, что отражается в низких значениях оценочной функции.



**Рис. 4.** Кристаллическая структура нуклеотид-связывающих доменов  $\text{Ca}^{2+}$ -АТФазы в состоянии E1 (А) и модель пространственной структуры нуклеотид-связывающих доменов WNDP (Б), построенная на ее основе.

На рис. 4А показаны консервативные остатки Phe487 и Glu442, отвечающие за связывание АТФ в N-домене, и фосфорилируемый остаток Asp351 P-домена; расстояние между остатками Glu442 и Asp351 составляет  $\sim 26$  Å. На рис. 4Б показаны остатки His1069 и Glu1064 N-домена и фосфорилируемый остаток Asp1027 P-домена; расстояние между Glu1064 (возможный аналог Glu442  $\text{Ca}^{2+}$ -АТФазы) и Asp1027 составляет  $\sim 28$  Å. Элементы вторичной структуры —  $\alpha$ -спирали и  $\beta$ -тяжи изображены в виде цилиндров и лент, соответственно.

Построенные нами модели нуклеотид-связывающих доменов WNDP характеризуются достаточно высокими значениями оценочной функции Айзенберга, что свидетельствует о хорошем качестве модели в целом. Как и ожидалось, качество модели по этому критерию для P-домена значительно лучше, чем для N-домена. Это может быть обусловлено отклонениями трехмерной модели N-домена от нативной структуры белка, причем главным образом в области петлевых участков.

Как видно из рис. 4, укладка P-доменов  $\text{Ca}^{2+}$ -АТФазы и трехмерной модели WNDP весьма сходна. В N-доменах этих белков наблюдаются значительные различия, обусловленные несколькими обширными делениями в аминокислотной последовательности WNDP относительно последовательности  $\text{Ca}^{2+}$ -АТФазы. Сопоставление данных мутагенеза [21] и модели нуклеотид-связывающих доменов WNDP показывает, что остатки, отвечающие за связывание АТФ — это прежде всего His1069 и Glu1064, в модельной системе расположены вблизи полости N-домена, аналогичной сайту связывания N-домена  $\text{Ca}^{2+}$ -АТФазы.

Для уточнения структуры сайта связывания АТФ, а также с целью установления взаимного расположения доменов для трехмерных моделей WNDP в формах E1 и E2 были выполнены расчеты молекулярной динамики в явно заданном растворителе. Наблюдаются дви-

жения доменов относительно друг друга, при этом отдельные домены WNDP в целом сохраняют свою структуру в ходе МД. Анализ изменений междоменного угла  $\Theta$  в МД-траекториях показал, что при старте из форм E1 и E2 получаются различные равновесные конформации доменов. Аналогично  $\text{Ca}^{2+}$ -АТФазе в траектории МД(E1) происходит доменное движение типа «закрывание», и  $\Theta$  достигает значения  $\sim 140^\circ$  (при стартовом значении  $\Theta = 160^\circ$ ). В траектории МД(E2) наблюдается дополнительное сближение доменов:  $\Theta$  стабилизируется около  $105^\circ$  (стартовое значение  $\Theta = 122^\circ$ ).

Протокол докинга, отработанный на примере  $\text{Ca}^{2+}$ -АТФазы, был применен для изучения взаимодействий АТФ с моделью АТФ-связывающих доменов WNDP. Докинг АТФ проводили для представительного набора конформеров WNDP, извлеченного из траекторий МД(E1) и МД(E2). Во всех комплексах, полученных в результате докинга, нуклеотид встраивается вблизи интерфейсного участка между N- и P-доменами, а фосфатные группы могут координироваться аминокислотными остатками из обоих доменов. Для анализа отбирали только те комплексы, в которых  $\gamma$ - $\text{PO}_3$ -группа АТФ была направлена в сторону сайта P-домена.

Фосфатный «хвост» молекулы АТФ, связанной в сайте N-домена посредством аденинового основания, может достигать фосфорилируемого остатка Asp1027 P-домена при условии  $\Theta \leq 105^\circ$ . Такие конформации наблюдали в траектории МД(E2). Результаты докинга показали воз-

возможность весьма разнообразных положений аденинового основания АТФ в полости сайта N-домена WNDP. По-видимому, такая «гетерогенность» комплексов вызвана относительно малой точностью предсказания пространственной структуры N-домена WNDP.

В результате докинга АТФ в МД(E1)-конформеры, характеризующие значениями угла  $\Theta > 125^\circ$ , были получены комплексы, в которых из-за слишком большого расстояния между сайтом связывания в N-доме и каталитическим остатком Asp1027 не может быть достигнут одновременный контакт лиганда с обоими доменами. Итоговые комплексы WNDP с АТФ можно разделить на две группы.

1. Адениновое основание и остаток рибозы АТФ встроены в полость N-домена (рис. 5), которую формируют преимущественно гидрофобные остатки. Для многих комплексов характерна водородная связь между  $\text{NH}_2$ -группой АТФ и карбоксильной группой остатка Glu1064 или Asp1164. Фосфатный «хвост» АТФ формирует одну или две водородных связи с боковой цепью Arg1151. Совместно с остатками Lys1028 и Asp1222 P-домена боковая цепь Arg1151 может фиксировать фосфатные группы АТФ и «направлять» их в сторону фосфорилируемого остатка Asp1027. При этом группа  $\gamma\text{-PO}_3$  лиганда остается удаленной от сайта фосфорилирования на расстояние более 8 Å.

2.  $\gamma\text{-PO}_3$ -группа АТФ находится в непосредственном контакте с остатком Asp1027, а адениновое основание располагается на интерфейсном участке между N- и P-доменами. В этом случае взаимодействия белок-лиганд определяются сеткой водородных связей между фосфатными группами АТФ и полярными остатками из P-домена.

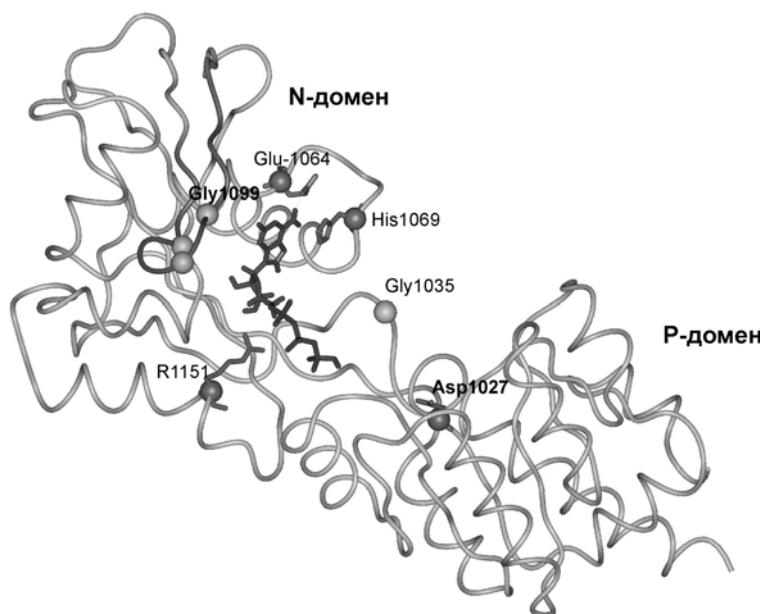


Рис. 5. Один из вариантов встраивания АТФ в структуру из МД E1-модели нуклеотид-связывающих доменов WNDP.

Расстояние между  $\gamma\text{-PO}_3$ -группой АТФ и фосфорилируемым остатком Asp1027 из P-домена составляет ~8 Å. Сферами обозначена локализация остатков, мутации по которым экспериментально изучены и связаны с болезнью Вильсона, а также фосфорилируемый остаток Asp1027. Выделены остатки Glu1064, His1069 и Arg1151 из N-домена, формирующие контакты с АТФ в предсказанной структуре комплекса

Таким образом, нуклеотид-связывающие домены WNDP содержат два мотива, отвечающих за связывание АТФ. Это 1) гидрофобная полость в N-доме, которая изолирует адениновое основание лиганда от молекул растворителя, и 2) кластер остатков в P-доме, фиксирующих фосфатный «хвост» АТФ за счет сетки водородных связей. Связывание АТФ с высокой аффинностью, по-видимому, достигается при взаимодействии молекулы с обоими аминокислотными мотивами одновременно. Необходимое для этого сближение связывающих мотивов происходит в результате доменного движения типа «закрывание», наблюдавшегося в расчете молекулярной динамики. По данным [21] связывание АТФ с изолированным N-домом WNDP характеризуется высокой аффинностью, близкой к показателю для нативного белка, т.е. вклад сайта N-домена в связывание АТФ является определяющим.

Из-за низкой гомологии аминокислотных последовательностей WNDP и структуры-«шаблона»  $\text{Ca}^{2+}$ -АТФазы трехмерная модель WNDP в области N-домена может значительно отличаться от реальной структуры. В связи с этим особенно актуальным становится сопоставление модели с экспериментальными данными мутагенеза. Наш анализ основан на результатах измерения параметра связывания (константы диссоциации) АТФ, проведенного для изолированного N-домена WNDP и его мутантов Glu1064Ala, His1069Gln, Arg1151His и Cys1104Phe [21]. Сделан вывод, что мутации Glu1064Ala и His1069Gln значительно снижают сродство белка к АТФ. Кроме того, мы использовали результаты исследований по мутагенезу, которые проводились для близкородственного WNDP белка — цинк-транспортующей АТФазы [26]. В этой работе исследовано влияние мутаций, аналогичных мутациям, связанных с болезнью Вильсона (Glu1064Ala, His1069Gln, Gly1099Ser, Gly1101Arg, Cys1104Phe и Gly1035Val), на взаимодействие с АТФ, фосфорилирование и транспорт катионов. Показано, что остатки, аналогичные Gly1099, Gly1101, Cys1104 и Gly1035, также важны для связывания АТФ наряду с Glu1064 и His1069.

На основании нашей модели (см. рис. 5) можно объяснить эффекты, измеренные в экспериментах по мутагенезу. Так, замена остатка Gly1035, расположенного на «шарнирном» участке между N- и P-доменами, на остаток валина (Val) может затруднять необходимые в каталитическом цикле доменные переходы и, следовательно, нарушать процесс фосфорилирования. Остаток Arg1151 находится на поверхности контакта между N- и P-доменами; площадь контакта между доменами значительно увеличивается при переходе из состояния E1 в E2. Таким образом, мутация Arg1151His может изменять соотношение энергий «открытого» и «закрытого» состояний и тем самым нарушать работу фермента.

В нашей модели остатки Glu1064 и His1069 находятся в контакте или хотя бы близки к молекуле АТФ, что может указывать на непосредственное их участие в



связывании лиганда. Замена Glu1064Ala в контексте нашей модели вызвана потерей водородной связи между карбоксильной группой остатка и NH<sub>2</sub>-группой аденинового основания АТФ. Сложнее интерпретировать влияние на связывание АТФ мутации His1069Gln, которая наиболее часто проявляется при синдроме Вильсона. В ряде комплексов, полученных методом докинга, остаток His1069 находится в контакте с фосфатным «хвостом» АТФ. На этом основании мы выдвигаем гипотезу, что эффект замены His1069Gln объясняется потерей положительного заряда протонированной формы боковой цепи остатка His, которая стабилизирует отрицательно заряженные фосфатные группы АТФ в сайте связывания.

Наконец, по данным [26] в стабилизации фосфатного «хвоста» АТФ должны участвовать NH-группы основной цепи остатков глицинового мотива (Gly<sup>1099</sup> - X- Gly<sup>1101</sup>) из петлевого участка в N-домене. В нашей модели WNDP этот петлевой участок расположен на некотором удалении от сайта связывания АТФ (см. рис. 5), что, по-видимому, является одной из неточностей модели.

Таким образом, на основании анализа экспериментальных данных можно заключить, что предложенная нами трехмерная модель АТФ-связывающих доменов WNDP и модель ее комплекса с АТФ являются в целом корректными. Эти модели могут служить стартовыми для последующих уточнений и помогут планировать и интерпретировать эксперименты по направленному мутагенезу нуклеотид-связывающих доменов WNDP и родственных ей АТФаз.

### Заключение

Сформулируем основные результаты представленной работы. С использованием широкого набора методов молекулярного моделирования проведено исследование структурно-функциональных аспектов взаимодействий медь-транспортирующей АТФазы человека с АТФ — в норме и при болезни Вильсона. Предварительно методы и протоколы молекулярного моделирования были успешно протестированы в «калибровочной» задаче предсказания структуры комплекса Ca<sup>2+</sup>-АТФазы с АТФ на основе кристаллических структур белка. С помощью метода молекулярной динамики показана целесообразность учета в докинге конформационной подвижности белка-мишени. Сделан вывод о том, что проведение независимой процедуры докинга лиганда (АТФ) для представительного набора МД-конформеров белка с последующей селекцией полученных комплексов значительно повышает предсказательную силу метода.

Принципиальная трудность в построении трехмерной модели Cu<sup>2+</sup>-АТФазы связана с низкой степенью гомологии аминокислотных последовательностей WNDP и Ca<sup>2+</sup>-АТФазы, используемой в качестве «шаблона», что предопределило неточности полученных моделей. Опыт показывает, что для построения трехмерных моделей белка, пригодных для достоверного предсказания структур комплексов белок-лиганд, требуется высокая степень гомологии (идентичности не менее 30+40%) аминокислотных последовательностей модели и белка-«шаблона». Соответственно, при моделировании структуры медь-транспортирующей АТФазы мы можем только приблизительно установить

укладку (фолд) белковой молекулы и локализацию функционально важных остатков. Тем не менее, анализируя данные направленного мутагенеза на основе полученной нами трехмерной модели WNDP, можно выдвигать гипотезы о роли тех или иных аминокислотных остатков при связывании АТФ и фосфорилировании. На основании модели возможно планирование новых экспериментов по направленному мутагенезу, результаты которых могут подтвердить или опровергнуть сделанные предсказания.

\* \* \*

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (гранты 04-04-48875-а, 05-04-49346-а, 06-04-49194-а), Федерального агентства по науке и инновациям РФ (Государственный контракт № 02.467.11.3003 от 20.04.2005 г., грант Президента РФ на поддержку ведущих научных школ № НШ-4728.2006.4), Российской академии наук (Программа «Физико-химическая биология»), Российского фонда содействия отечественной науке (грант Е.Р.Г.). Авторы признательны Межведомственному суперкомпьютерному центру РАН за предоставленные вычислительные ресурсы.

### ЛИТЕРАТУРА

1. Kuhlbrandt W. Nature, 2004, v. 5, p. 282–295.
2. Kaplan J.H. Annu. Rev. Biochem., 2002, v. 71, p. 511–535.
3. Palmgren M.G. Annu. Rev. Plant Physiol. Mol. Biol., 2001, v. 52, p. 817–845.
4. Bull P.C., Thomas G.R., Rommens J.M., Forbes J.R., Cox D.W. Nature Genet., 1993, v. 5, p. 327–337.
5. van Gunsteren W.F., Berendsen H.J.C. Gromos-87 manual; Biomos BV Nijenborgh 4, 9747 AG Groningen. 1987. The Netherlands.
6. Brooks B.R., Brucocoleri R.E., Olafson B.D., States D.J., Swaminathan S., Karplus M. J. Comp. Chem., 1993, v. 4, p. 187–217.
7. Jorgensen W.L., Tirado-Rives J. J. Am. Chem. Soc., 1988, v. 110, p. 1657–1666.
8. Berendsen H.J.C., van der Spoel D., van Drunen R. GRO-MACS Comp. Phys. Comm., 1995, v. 91, p. 43–56.
9. Valler M.J., Green D. Discov Today, 2000, v. 7, p. 286–293.
10. Lengauer T., Lemmen C., Rarey M., Zimmermann M. Ibid., 2004, v. 9, p. 27–34.
11. Wang J., Kollman P.A., Kuntz I.D. PROTEINS, 1999, v. 36, p. 1.
12. Jones G., Willett P., Glen R.C., Leach A.R., Taylor R. J. Mol. Biol., 1997, v. 267, p. 727–748.
13. Perola E., Walters W.P., Charifson P.S. PROTEINS, 2004, v. 56, p. 235–49.
14. Toyoshima C., Nakasako M., Nomura H., Ogawa H. Nature, 2000, v. 405, p. 647–655.
15. Toyoshima C., Nomura H. Ibid., 2002, v. 418, p. 605–611.
16. Toyoshima C., Mizutani T. Ibid., 2004, v. 430, p. 529–535.
17. Abu-Abed M., Mal T.K., Kainosho M., MacLennan D.H., Ikura M. Biochemistry, 2002, v. 41, p. 1156–1164.
18. Fatemi N., Sarkar B. J. Bioenerg. Biomembr., 2002, v. 34, p. 339–349.
19. Lutsenko S., Efremov R.G., Tsivkovskii R., Walker J.M. Ibid., 2002, v. 34, p. 351–362.
20. Tsivkovskii R., Efremov R.G., Lutsenko S. J. Biol. Chem., 2003, v. 278, p. 13302–13308.
21. Morgan C.T., Tsivkovskii R., Kosinsky Yu.A., Efremov R.G., Lutsenko S. Ibid., 2004, v. 279, p. 36363–36371.
22. Jones D.T. J. Mol. Biol., 1999, v. 287, p. 797–815.
23. Efremov R.G., Kosinsky Yu.A., Nolde D.E., e.a. Biochem. J., 2004, v. 382, Part I, p. 293–305.
24. Sali A., Blundell T.L. J. Mol. Biol., 1993, v. 234, p. 779–815.
25. Lüthy R., Bowie J.U., Eisenberg D. Nature, 1992, v. 356, p. 83–85.
26. Okkeri J., Laakkonen L., Haltia T. Biochem. J., 2004, v. 377, p. 95.