

УДК 539.26

ПРОБЛЕМА НЕОДНОЗНАЧНОСТИ РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ МАЛОУГЛОВОГО РАССЕЯНИЯ: ПОСЛЕДОВАТЕЛЬНЫЙ ПОДХОД НА ПРИМЕРЕ РЕЦЕПТОРА, ПОДОБНОГО РЕЦЕПТОРУ ИНСУЛИНА. МЕТОДЫ ИНТЕРПРЕТАЦИИ ДАННЫХ МУРР

© 2021 г. М. В. Петухов^{a, *}, П. В. Конарев^a, В. В. Волков^a,
А. А. Можяев^{a, b}, Э. В. Штыкова^a

^aИнститут кристаллографии им. А.В. Шубникова,
ФНИЦ “Кристаллография и фотоника” РАН,
Москва, 119333 Россия

^bИнститут биоорганической химии
им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН,
Москва, 117997 Россия

*e-mail: maxim@embl-hamburg.de

Поступила в редакцию 16.03.2021 г.

После доработки 25.03.2021 г.

Принята к публикации 25.03.2021 г.

Построение трехмерных моделей белковых макромолекул представляет собой серьезную проблему из-за возможной неоднозначности решения обратной задачи восстановления трехмерной структуры по одномерному профилю малоуглового рассеяния. Целевая функция этой задачи может иметь несколько локальных минимумов, что приводит к зависимости решения от стартовых значений параметров моделей и от метода поиска глобального минимума. Проблема создания структурных моделей усложняется также усреднением картины рассеяния по всем ориентациям частиц в пространстве, а при наличии полидисперсности и/или полиморфизма – распределением по размерам и по форме рассеивающих объектов. В данной работе проблема неоднозначности решения обратных задач и восстановления трехмерной структуры белка рассматривается на примере исследования строения эктодомена рецептора, подобного рецептору инсулина (ectoIRR) в растворе. В работе представлен последовательный подход к решению данной проблемы, начиная от определения общих структурных параметров и *ab initio* восстановления формы до моделирования жесткими телами (методом молекулярной тектоники), гибридными методами и анализ кривых рассеяния с помощью разложения на сингулярные векторы.

Ключевые слова: малоугловое рентгеновское рассеяние, структурное моделирование, рецепторные тирозинкиназы, рецептор подобный рецептору инсулина

DOI: 10.31857/S0233475521040095

ВВЕДЕНИЕ

Благодаря широкому использованию источников синхротронного излучения, новых типов регистрирующих устройств и развитию новых подходов к обработке и интерпретации данных эксперимента метод малоуглового рассеяния (МУР) превратился в один из самых мощных и универсальных инструментов для исследования надатомной структуры вещества, который применяется при анализе структуры дисперсных систем, в молекулярной биологии, биофизике и других областях науки и техники [1–5].

Уровень структурной информации, заключенной в данных МУР, определяется природой рассеивающих объектов. Наиболее плодотворно тео-

рия малоуглового рассеяния и подходы к решению структурных задач разработаны для исследования монодисперсных систем идентичных частиц. Такие системы исследуются, например, при анализе высокоочищенных растворов биологических макромолекул. В этом случае интенсивность рассеяния непосредственно связана со структурой частиц, что дает уникальную возможность определения их формы и размера, а в ряде случаев и внутреннего строения с разрешением порядка 1 нм.

Общие параметры исследуемых образцов, так называемые инварианты МУР, однозначно определяются непосредственно по кривым рассеяния, не требуя структурного моделирования [1]. К ин-

вариантам МУР относятся максимальный размер частиц (D_{\max}), радиус инерции (R_g), объем Порода (V_p) и молекулярная масса (MM) [1, 6]. Построение же трехмерных моделей представляет собой значительно более сложную проблему из-за плохой обусловленности обратной задачи восстановления трехмерной структуры по одномерному профилю рассеяния. Целевая функция этой задачи может иметь несколько локальных минимумов, что приводит к зависимости решения от стартовых значений параметров моделей и от метода поиска глобального минимума. Проблема создания структурных моделей усложняется также усреднением картины рассеяния по всем ориентациям частиц в пространстве, а при наличии полидисперсности и/или полиморфизма распределением по размерам и по форме рассеивающих объектов. Таким образом, однозначное решение обратной задачи рассеяния, т.е. восстановление трехмерной структуры объекта высокого разрешения по экспериментальным данным МУР, в общем случае невозможно. Тем не менее, учитывая значения инвариантов, рассчитанных по экспериментальным кривым, налагая определенные ограничения на искомую структуру, например, вводя условие симметричности частицы или ее однородности, и используя структурную информацию, полученную комплементарными методами, коридор возможных решений удается значительно сузить. Кроме того, современные методы обработки и интерпретации данных малоуглового рассеяния [5, 7] позволяют за сравнительно короткое время, используя разные подходы и многократный запуск соответствующих программ, восстановить структуру исследуемого образца, что дает возможность сравнения, усреднения, кластеризации и последующего анализа набора полученных решений [8, 9]. В результате удается получить достаточно подробную структуру рассеивающего объекта.

В данной работе проблема неоднозначности решения обратных задач и восстановления структуры по данным малоуглового рентгеновского рассеяния (МУРР) рассматривается на примере исследования строения эктодомена рецептора, подобного рецептору инсулина (ectoIRR) в растворе. Рецептор, подобный инсулиновому рецептору (IRR), принадлежит к суперсемейству трансмембранных рецепторных тирозинкиназ, которые регулируют различные жизненно важные процессы в организме путем передачи сигнала об изменении условий окружающей среды во внутриклеточное пространство, активируя тем самым ответ клетки на возникшие изменения. IRR был открыт в 1989 году [10], но его структура и функция долгое время не были известны. Только недавно впервые было показано, что этот белок активизируется исключительно при изменении кислотно-щелочного баланса, т.е. для передачи

сигнала во внутриклеточную среду этому белку не требуется присоединение к нему лигандов [11, 12]. Необычные функциональные свойства этого рецептора привлекли внимание исследователей всего мира, однако структура полноразмерного белка неизвестна до сих пор, и только в 2019 году методами МУРР и атомно-силовой микроскопии впервые была определена конформация его эктодомена при разных pH [13]. Тем не менее дальнейшее изучение свойств ectoIRR показывает, что из-за сложности его структурной организации вопрос об однозначности полученной в 2019 году структуры остается открытым. Именно поэтому этот белок был выбран в качестве объекта для выработки оптимальной стратегии минимизации неоднозначности решения обратных задач МУРР при исследовании структур сложных биологических макромолекул в растворе.

В работе представлен последовательный подход к решению данной проблемы, начиная от определения общих структурных параметров и *ab initio* восстановления формы до моделирования жесткими телами (методом молекулярной тектоники) и гибридными методами.

МАТЕРИАЛЫ И МЕТОДЫ

Получение рекомбинантного ectoIRR. Рекомбинантный эктодомен рецептора, подобного рецептору инсулина (ectoIRR), был экспрессирован, очищен и охарактеризован, как описано в [13].

Анализ данных малоуглового рентгеновского рассеяния: первичная обработка данных. Малоугловой эксперимент с IRR был проведен на станции BioSAXS P12 в EMBL (DESY, Гамбург, Германия) [14] совместно с эксклюзионной гель-хроматографией (ЭГХ) с целью разделить возможные олигомеры и для обеспечения его монодисперсности [15].

Обработка полученных экспериментальных данных и их дальнейшая интерпретация проводилась с помощью программного комплекса ATSAS [16], который представляет собой один из самых надежных и популярных инструментов анализа, визуализации и моделирования данных МУР. Все представленные ниже программы и программные блоки, использованные в данном исследовании, являются частью последней модификации ATSAS.

Первичная обработка экспериментальных данных состоит в ряде манипуляций с набором экспериментальных кривых рассеяния, тщательность выполнения которых определяет количество и качество полученной структурной информации. Программный многофункциональный блок PRIMUS [17], разработанный для выполнения предварительных действий с исходными экспериментальными данными МУРР, обеспечивает

необходимый и достаточный базис для дальнейшего структурного анализа.

Процесс первичной обработки включает в себя статистический анализ набора кривых малоуглового рассеяния для данного образца за период измерения, их усреднение и шкалирование. С помощью программного блока PRIMUS проводится вычитание рассеяния фоном (растворителем), сшивка данных, измеренных на разных угловых интервалах, экстраполяция к нулевой концентрации образца, вычисление инвариантов с использованием графиков Гинье и Порода и многое другое.

Следует отметить, что в случае использования онлайн ЭГХ для обеспечения монодисперсности образца усреднение кривых рассеяния, а также вычитание рассеяния буфером проводится с помощью программы CHROMIXS [18].

Радиус инерции R_g определяется из аппроксимации Гинье [19] при очень малых углах, где интенсивность рассеяния для идентичных не взаимодействующих частиц определяется как $I(s) = I_0 \exp[-(sR_g)^2/3]$. При этом интенсивность рассеяния в нулевой угол I_0 пропорциональна молекулярной массе. Объем Порода V_p рассеивающего объекта вычисляют с использованием нормированного инварианта Порода Q [20].

Для оценки свернутости макромолекулы используется график в координатах Кратки: $s^2 I(s)$ от s . Интенсивность рассеяния структурированных, плотно свернутых частиц демонстрирует хорошо выраженный колоколообразный максимум на графике Кратки [20]. Разупорядоченные, развернутые рассеивающие объекты подчиняются закономерности s^{-2} , образуя статистически свернутую Гауссову цепь (клубок) [21].

Модифицированный вариант программного блока PRIMUS, PRIMUS/QT [7], содержит инструмент “Molecular weight”, который дает Байесовскую оценку молекулярной массы, основываясь на нескольких параметрах [22].

Встроенная в PRIMUS утилита SASPLOT [7] позволяет оценить качество разностных кривых рассеяния благодаря возможности детальной визуализации их отдельных участков. Определение геометрических и весовых характеристик и ряда других структурных параметров также требует тщательного визуального контроля, и хотя в настоящее время первичную обработку данных малоуглового рассеяния можно проводить в полностью автоматическом режиме [23], в случае исследования сложных объектов и систем, обладающих полиморфизмом и/или полидисперсностью, эту процедуру следует выполнять вручную, а графический интерфейс комплекса PRIMUS позволяет оценить качество выполняемых манипуляций.

Программа SHANUM [24] используется для оценки полезного участка кривой рассеяния, где высокоугловые точки с сильным шумом исключаются из рассмотрения как неинформативные. Программа DATCLASS [25] используется для предсказания типа частицы по кривой рассеяния (компактная, вытянутая, плоская, кольцевая, полая, случайная цепь). Оценка неоднозначности определения формы для конкретных образцов проводится с помощью программы AMBIMETER [26], которая сравнивает экспериментальную кривую рассеяния с “картой неоднозначности”, рассчитанной по 14000 топологий, и дает количественный прогноз для *ab initio* восстановления формы. Индекс неоднозначности *a-score* является логарифмом числа соседних кривых на карте неоднозначности для данной экспериментальной кривой. Величина *a-score* < 1.5 практически гарантирует однозначное восстановление. При значениях *a-score* > 2.5 восстановление имеет множество, во всех остальных случаях восстановление потенциально неоднозначно.

Анализ данных малоуглового рентгеновского рассеяния: определение общей формы частиц по данным МУРР. Для разбавленных растворов со случайной ориентацией монодисперсных, не взаимодействующих частиц рассеяние является изотропным, и его интенсивность $I(s)$ выражается как функция модуля вектора рассеяния s ($s = |\mathbf{s}| = 4\pi \sin(\theta)/\lambda$), где λ – длина волны падающего излучения и 2θ – угол рассеяния:

$$I(s) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(sr)}{sr} dr, \quad (1)$$

здесь функция распределения по расстояниям $p(r)$ связана преобразованием Фурье с интенсивностью рассеяния:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} s^2 I(s) \frac{\sin(sr)}{sr} ds. \quad (2)$$

Таким образом, функции распределения по расстояниям $p(r)$ может быть найдена с помощью уравнений (1) и (2). Эта функция содержит информацию о форме и структуре частицы и позволяет оценивать ее максимальный размер D_{\max} из условия $p(r) = 0$ при $r > D_{\max}$. В практических исследованиях $p(r)$ рассчитывают не прямым интегрированием интенсивности (что приводит к сильным эффектам обрыва), а с помощью так называемого косвенного Фурье-преобразования и программы GNOM пакета ATSAS [27].

Реконструкция трехмерной модели объекта по его одномерной кривой рассеяния является не вполне корректной задачей, поскольку множественные структуры гипотетически могут обеспечить один и тот же профиль рассеяния. Упрощение, уменьшающее неоднозначность такой ре-

конструкции, состоит в представлении частицы с низким разрешением как однородного тела. Форма частицы может быть описана как набор конечных объемных элементов (шариков), расположенных на плотной гексагональной упаковке. Наиболее распространенный подход для определения формы с низким разрешением использует минимизацию целевой функции с помощью имитации отжига для создания компактного взаимосвязанного ансамбля шариков в объеме поиска (обычно, сфера с диаметром D_{\max}), которая соответствует экспериментальным данным. Этот подход реализован в программе DAMMIN [28]. Данный подход не требует наличия априорной структурной информации и, таким образом, моделирование ведется *ab initio*.

В процессе моделирования выполняется поиск оптимальной модели, где минимизируется отклонение между экспериментальными данными $I_{\text{exp}}(s)$ и профилем рассеяния, вычисленным из модели $I_{\text{calc}}(s)$:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{\text{exp}}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2, \quad (3)$$

где c – шкалирующий коэффициент, N – количество экспериментальных точек, а σ обозначает ошибки эксперимента.

Для получения структурных моделей, отвечающих надатомному разрешению, на искомую модель накладываются условия связности и компактности. Поэтому целевая функция, минимизируемая в процессе восстановления структуры, имеет вид: $f(X) = \chi^2 + \alpha P(X)$, где $P(X)$ – штраф за некомпактность модели ($\alpha > 0$ – его вес, задаваемый пользователем).

Программу DAMMIN для *ab initio* моделирования можно применять для любых монодисперсных образцов, а для восстановления доменной структуры белков была разработана программа GASBOR [29]. Виртуальные остатки в этой программе формируют белковоподобную цепь, свернутую так, чтобы рассеяние от полученной структуры соответствовало рассеянию от исследуемой белковой макромолекулы в растворе с разрешением порядка 0.5 нм.

Расчет кривой МУРР от моделей атомного разрешения. Более детальная интерпретация данных МУРР возможна, если атомные модели высокого разрешения всей макромолекулы или ее отдельных фрагментов доступны из кристаллографических данных или ЯМР. В этом случае МУРР позволяет обнаружить структурное различие кристалл–раствор, определить биологически активную конформацию или олигомерное состояние. Следует отметить, что точное построение кривых рассеяния в растворе по атомным координатам рассеивающих объектов не является три-

виальной задачей. На профили рассеяния влияют эффекты растворителя, в частности наличие гидратной оболочки со средней плотностью, примерно на 10% превышающей плотность основного растворителя. Тем не менее расчет рассеяния макромолекулой в растворе и приближение экспериментальной кривой может быть произведено программой CRY SOL [30].

Восстановление структуры по данным МУРР методом молекулярной тектоники. В тех случаях, когда рассеивающие частицы представляют собой макромолекулярные комплексы, чрезвычайно важным подходом к интерпретации данных МУРР является моделирование жесткими телами, или метод молекулярной тектоники [31]. Часто сложный биологический комплекс не может быть закристаллизован, и расположение субъединиц в нем остается неизвестным, но удается закристаллизовать и определить структуру отдельных составляющих комплекса с атомным разрешением. Тогда взаимное расположение субъединиц в комплексе можно получить методом молекулярной тектоники из данных МУРР. То есть метод используется в том случае, когда известны структуры отдельных частей, составляющих комплекс. Кроме того, этим методом можно анализировать внутреннюю структурную гибкость и подвижность, присущую субъединицам в комплексе или доменам в границах одной макромолекулы. Моделирование жесткими телами пространственной структуры комплексов заключается в перемещении и вращении отдельных элементов структуры и получении минимального отклонения χ^2 между модельной и экспериментальными кривыми, что осуществляется с помощью программы SASREF [31]. Процедура моделирования методом молекулярной тектоники легко обобщается на случай произвольного числа жестких тел K и в общем случае может быть описана с помощью $6(K-1)$ параметров. При этом амплитуды рассеяния отдельных субъединиц рассчитываются с помощью программы CRY SOL.

Поскольку в белковых макромолекулах часто жесткие домены соединены между собой гибкими, разупорядоченными, подвижными линкерами, анализ структуры таких белков проводится с помощью дальнейшего развития метода молекулярной тектоники, что реализовано, например, в программе CORAL [8]. Программа совершает гибридное моделирование, используя молекулярную тектонику для позиционирования доменов и *ab initio* подход для представления линкеров в виде цепей, составленных из виртуальных остатков. Принцип гибридного моделирования биологических объектов описан ниже.

Гибридные методы восстановления структуры по данным МУРР. Сочетание разных подходов, например молекулярной тектоники и других струк-

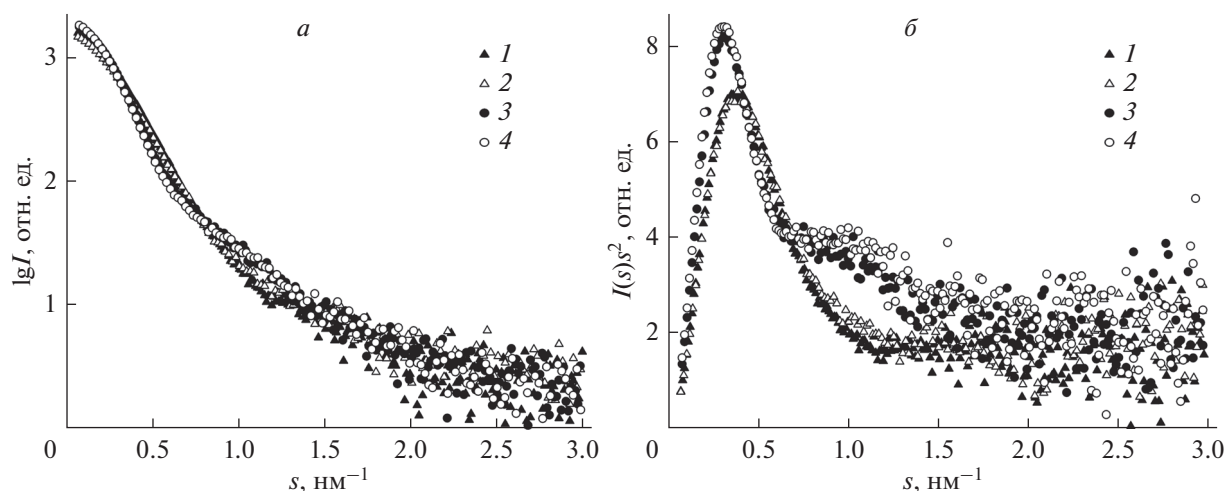


Рис. 1. Экспериментальные кривые малоуглового рассеяния (а) и графики Кратки (б) для эктодомена рецептора, подобного рецептору инсулина (ectoIRR), в растворе при pH 7.0 и 9.0: кривые 1 (pH 7.0) и 2 (pH 9.0) – измерения, проведенные до применения онлайн хроматографии; кривые 3 (pH 7.0) и 4 (pH 9.0) – измерения с использованием эксклюзионной гель-хроматографии.

турных методов, позволяет создать кластер методов, которые называются гибридными. Одним из наиболее важных подходов к изучению макромолекулярных комплексов гибридными методами является сочетание методов *ab initio* и молекулярной тектоники. Среди программ, реализующих этот подход, следует особо отметить программу EOM, которая представляет собой метод оптимизации ансамбля частиц (доменов, субъединиц) и, помимо функции *ab initio* восстановления цельной формы, позволяет анализировать гибкость биологических макромолекул, их развернутость и подвижность отдельных частей относительно друг друга, т.е. эта программа представляет собой инструмент для описания распределения по форме и по размеру частично или полностью развернутой макромолекулы в растворе [32].

Модели, полученные различными способами, совмещаются с помощью программы SUPCOMB [33], чтобы определить разницу в их пространственной организации и указать нормализованное пространственное отклонение (NSD). Как правило, значение NSD, близкое к 1.0, показывает, что модели, полученные разными методами, схожи.

Анализ полидисперсных и полиморфных растворов. В случае полидисперсных и полиморфных систем (разбавленные растворы) результирующая интенсивность рассеяния представляет собой линейную комбинацию отдельных компонент и описывается уравнением

$$I(s) = \sum_{k=1}^K (v_k I_k(s)), \quad (4)$$

где K – число компонент системы, v_k – объемные фракции и $I_k(s)$ – интенсивности рассеяния от каждой компоненты.

Для количественного анализа объемной доли различных компонентов в растворе (в случае, когда число компонент и их интенсивности рассеяния известны) применяется программа OLIGOMER [17]. Программа использует алгоритм неотрицательных линейных наименьших квадратов для минимизации расхождения χ^2 между предсказанной кривой рассеяния от смеси и экспериментальными данными МУРР.

Одним из эффективных модельно-независимых методов оценки числа компонент является разложение на сингулярные векторы. Когда имеется набор данных МУРР, записанный при изменяющихся условиях, программа SVDPLOT [17] применяется для вычисления сингулярных векторов и связанных с ними сингулярных значений. Число неслучайно осцилируемых сингулярных векторов со значительными сингулярными значениями позволяет оценить минимальное количество независимых кривых, необходимых для представления всего набора данных, т.е. количество значимых компонент в смеси.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Практическое применение методов интерпретации данных МУРР на примере кривых рассеяния от ectoIRR. Экспериментальные кривые малоуглового рассеяния ectoIRR при pH 7.0 и 9.0 показаны на рис. 1а.

Как видно из рис. 1а, до применения эксклюзионной гель-хроматографии кривые рассеяния ectoIRR характерны скорее для полидисперсных

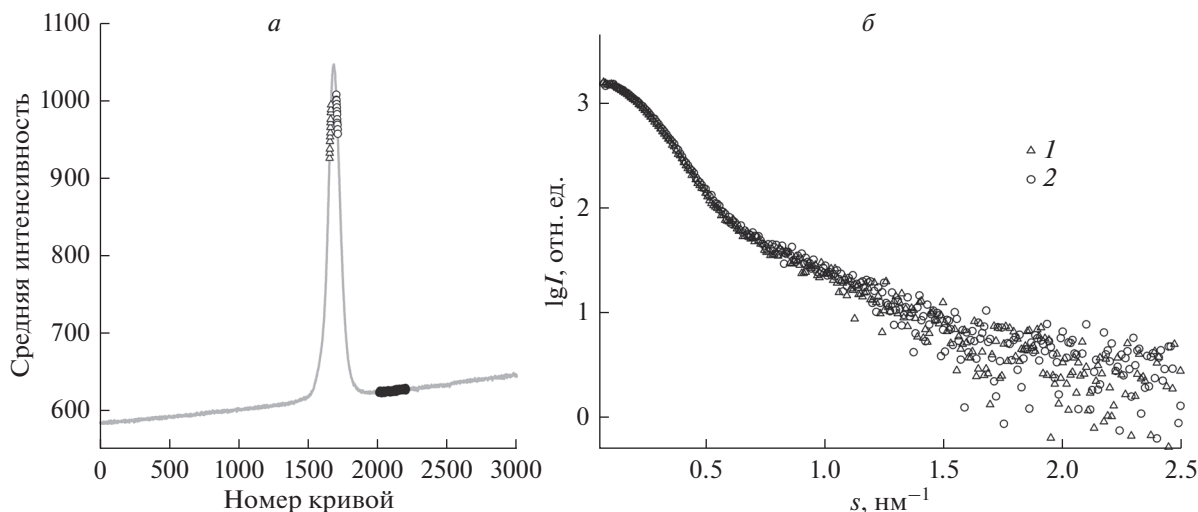


Рис. 2. *а* – Хроматограмма estoIRR при pH 9. Сплошной линией показаны усредненные интенсивности рассеяния в зависимости от номера кривой при элюировании образца. Треугольниками показаны кривые с левой стороны, кружочками – с правой. Черным цветом отмечены точки, соответствующие буферу. *б* – Усредненные кривые малоуглового рассеяния белка, взятые слева (1) и справа (2) от пика хроматограммы.

соединений со слабо выраженным форм-фактором. Особенностью этих кривых является также то, что интенсивность рассеяния в самых малых углах несколько ниже, чем для кривых после хроматографии. Более ярко указанные отличия выявляются на графиках в координатах Кратки (рис. 1б). Можно предположить, что образцы estoIRR до прохождения через хроматографическую колонку, будучи полидисперсными, могут содержать некоторое количество низкомолекулярных примесей, что выражается в уменьшении рассеяния в нулевой угол и уменьшении амплитуды максимума на графике Кратки в интервале угловых векторов $0.25\text{--}0.75\text{ нм}^{-1}$. Поэтому для структурного анализа далее использовались кривые рассеяния, дополнительно очищенные с помощью ЭГХ.

Поскольку IRR активизируется исключительно при изменении кислотно-щелочного баланса, следовало ожидать заметного изменения профиля МУРР при изменении pH, однако кривые рассеяния при защелачивании раствора практически не отличаются друг от друга, как это уже наблюдалось нами ранее [13], и тогда был сделан вывод, что структура estoIRR на макроуровне не изменяется при изменении pH, а конформационные перестройки ограничиваются локальным вращением доменов белка за счет гибких линкеров между ними. Аналогичные выводы были сделаны также для одного из гомологов IRR – для эктодомена рецептора инсулиноподобного фактора роста в работе [34].

Для выяснения того, какие именно структурные характеристики и на каком структурном уровне изменяются при активации estoIRR, в

данной работе был проведен тщательный анализ кривых малоуглового рассеяния и рассчитаны соответствующие структурные параметры.

Первичная обработка данных МУРР и общие структурные параметры. Поскольку хроматограммы estoIRR при pH 7.0 и pH 9.0 практически совпадают, на рис. 2а для примера представлена хроматограмма эктодомена при pH 9.0, представляющая собой узкий и на первый взгляд симметричный пик, характерный для монодисперсного вещества. Несколько кривых рассеяния элюатом были взяты по отдельности с правой и левой сторон пика для усреднения и последующего сравнения, т.е. для того, чтобы убедиться в симметричности пика хроматограммы и исключить присутствие дополнительных фракций estoIRR. Из отдельно усредненных кривых МУРР с левой и правой сторон пика было вычтено рассеяние буфером (соответствующие кривые помечены черными точками на хроматограмме) в программе CHROMIXS. Полученные таким образом профили рассеяния представлены на рис. 2б.

Как видно из рис. 2б, усредненные кривые рассеяния элюата estoIRR, взятые слева и справа от пика хроматограммы, полностью совпадают. Отсюда следует, что после хроматографической колонки раствор белка становится монодисперсным и, следовательно, рассеяние от него может быть использовано для детального количественного анализа структуры estoIRR.

Согласно программе SHANUM, полезный диапазон данных заканчивается при значении $s \approx 2.5\text{--}3.0\text{ нм}^{-1}$ (примерно 15 шенноновских каналов). Однако при расчете максимальных размеров и функций распределения по расстояниям

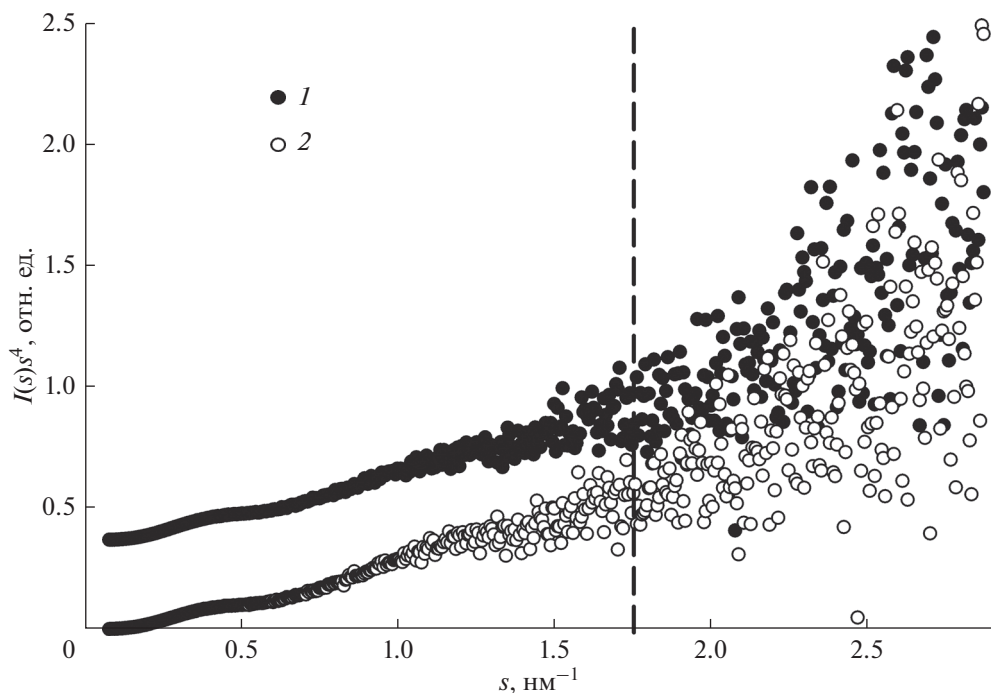


Рис. 3. Графики Порода для рН 7.0 (вверху) и рН 9.0 (внизу) кривых рассеяния *ectoIRR*. Вертикальная пунктирная линия ограничивает участок векторов рассеяния, соответствующего рассеянию однородной частицей.

$\rho(r)$ программой GNOM используемый диапазон данных был ограничен сверху 8–10 шенноновскими каналами, где ширина одного шенноновского канала равна π/D_{\max} . Это соответствует выходу на плато графика Порода $I(s)s^4$ от s для s в области примерно 1.5–1.8 nm^{-1} (рис. 3). Этот участок соответствует рассеянию однородной частицей и может быть использован в дальнейшем для *ab initio* восстановления формы *ectoIRR* в растворе.

Классификация типа структуры с помощью программы DATCLASS указала на компактную форму доменов молекулы *ectoIRR*, что хорошо согласуется с формой кривых рассеяния на графике Кратки (рис. 1б). Характерный колоколообразный вид графиков Кратки как для рН 7.0, так и для рН 9.0 свидетельствует о том, что *ectoIRR* в основном состоит из упорядоченных, свернутых, т.е. компактных, доменов. Тем не менее, для обеих кривых в результате сравнения с “картой неоднозначности” программой AMBIMETER было получено среднее значение параметра неоднозначности 2.1, что предполагает потенциально неоднозначное восстановление формы рассеивающего объекта [26]. Поскольку выше мы показали, что раствор *ectoIRR* при разных рН является монодисперсным, то такая неоднозначность восстановления структуры может быть вызвана потенциальным полиморфизмом белка, состоящего из множества отдельных доменов, соединенных гибкими линкерами.

Большое значение при решении обратных задач МУРР имеет тщательное определение интегральных макромолекулярных характеристик рассеивающих объектов. Радиусы инерции, посчитанные с помощью формулы Гинье, и другие инварианты МУРР, рассчитанные по кривым МУРР, приведены в табл. 1.

Прежде всего, в табл. 1 обращают на себя внимание более низкие значения инвариантов для *ectoIRR*, рассчитанных по кривым рассеяния без применения эксклюзионной гель-хроматографии. Совокупность полученных данных свидетельствует о полидисперсности: низкие значения MM и R_g говорят о присутствии в растворе низкомолекулярных примесей, в то время как увеличенное значение D_{\max} подразумевает наличие некоторого количества агрегатов белка. В целом, это еще раз подчеркивает необходимость использования онлайн ЭГХ для детального структурного анализа биологических объектов.

Также важен выбор правильного интервала угловых векторов. Породовый объем для *ectoIRR* при рН 7.0 и 9.0 (табл. 1), посчитанный в диапазоне данных до восьми шенноновских каналов, имеет явно завышенное значение, и если использовать эмпирическое соотношение $MM_{\text{Porod}} = V_{\text{Porod}}/1.65$, то значения молекулярных масс будут в 2 раза превышать соответствующие теоретические значения. Это значение для глобулярных белков было получено ранее [8]. Эмпирический

Таблица 1. Общие структурные параметры ectoIRR при pH 7.0 и 9.0

Параметр	Без ЭГХ		После ЭГХ		
	pH 7.0	pH 9.0	pH 7.0	pH 9.0	
				левая часть ЭГХ	правая часть ЭГХ
R_g , нм	5.4 ± 0.1	5.3 ± 0.1	5.7 ± 0.1	5.8 ± 0.1	5.8 ± 0.1
D_{\max} , нм	19.0 ± 1.0	18.0 ± 1.0	17.0 ± 1.0	17.5 ± 1.0	17.5 ± 1.0
V_{Porod} , нм ³ (8 шенноновских каналов)	455 ± 30	440 ± 30	650 ± 30	700 ± 30	700 ± 30
V_{Porod} , нм ³ (15 шенноновских каналов)	455 ± 30	440 ± 30	540 ± 30	580 ± 30	570 ± 30
MM_{Porod} , кДа (15 шенноновских каналов)	190 ± 10	180 ± 10	225 ± 20	240 ± 20	240 ± 20
MM_{IRdimer} , кДа теоретическое значение	230	230	230	230	230
Исключенный объем из <i>ab initio</i> восстановления	500 ± 20	500 ± 20	510 ± 20	520 ± 20	520 ± 20
V_{DAM} , нм ³					
MM_{DAM} , кДа	210 ± 20	210 ± 20	210 ± 20	220 ± 20	220 ± 20
$MM_{\text{Байес}}$, кДа	240	240	320	320	320

коэффициент 1.65 определялся для PDB структур, не содержащих гетероатомы. В данном случае молекула IRR содержит гликаны, что может существенно изменить соотношение. Для того чтобы оценить коэффициент для данного конкретного случая, использовалась гомологичная модель рецептора инсулина с гликанами, взятая из малоугловой базы данных SASBDB с кодом SASDHF2. Для этой модели, имеющей $MM = 251$ кДа, рассчитанный Породовский объем составил 600 нм^3 . Таким образом, было получено значение коэффициента для вычисления молекулярных масс, равное 2.4. Это значение и было использовано для оценки MM в настоящей работе в диапазоне угловых векторов до 15 шенноновских каналов. В этом случае молекулярные массы ectoIRR однозначно соответствуют димеру, что отвечает известным литературным данным, свидетельствующим, что рецептор, подобный рецептору инсулина, изначально существуют в мембране в виде димера, где соответствующие субъединицы связаны дисульфидными мостиками [35]. Показанные в табл. 1 значения MM хорошо согласуются с молекулярной массой, вычисленной по рассеянию в нулевой угол: 220 ± 20 кДа для ectoIRR после ЭГХ.

Завышенное значение Породовского объема в диапазоне данных до 8 шенноновских каналов можно объяснить неоднородностью макромолекул ectoIRR из-за наличия гликанов в его составе. Молекулярная масса эктодомена, посчитанная

по аминокислотной последовательности белка, равна примерно 190 кДа, т.е. на долю гликанов приходится почти 20%, и такой рассеивающий объект нельзя рассматривать как однородный. Однако начальный участок кривой рассеяния до 8 шенноновских каналов следует использовать для грубого *ab initio* восстановления формы ectoIRR в растворе. В этом случае мы используем упрощение, уменьшающее неоднозначность восстановления структуры, представляя форму частицы низкого разрешения в качестве однородного тела.

Значения молекулярных масс растворенных частиц были также рассчитаны с помощью инструмента “Molecular Weight” в программе PRIMUS/QT. Молекулярный вес, посчитанный с помощью этой программы, независимо от того, какой диапазон данных использовался (полноразмерный или же до восьми шенноновских каналов) был одинаков в пределах ошибки. Полученное значение в 320 кДа с доверительным интервалом [221–372] кДа соответствует массе димера ectoIRR в присутствии гликанов. Кроме того, следует еще раз подчеркнуть, что узкий пик на гель-хроматограмме свидетельствует о наличии только одного олигомерного состояния, и, соответственно, возможен только полиморфизм, но не полидисперсность.

Ab initio моделирование формы IRR. В результате трехмерного моделирования методом конечных объемных элементов в программе DAMMIN

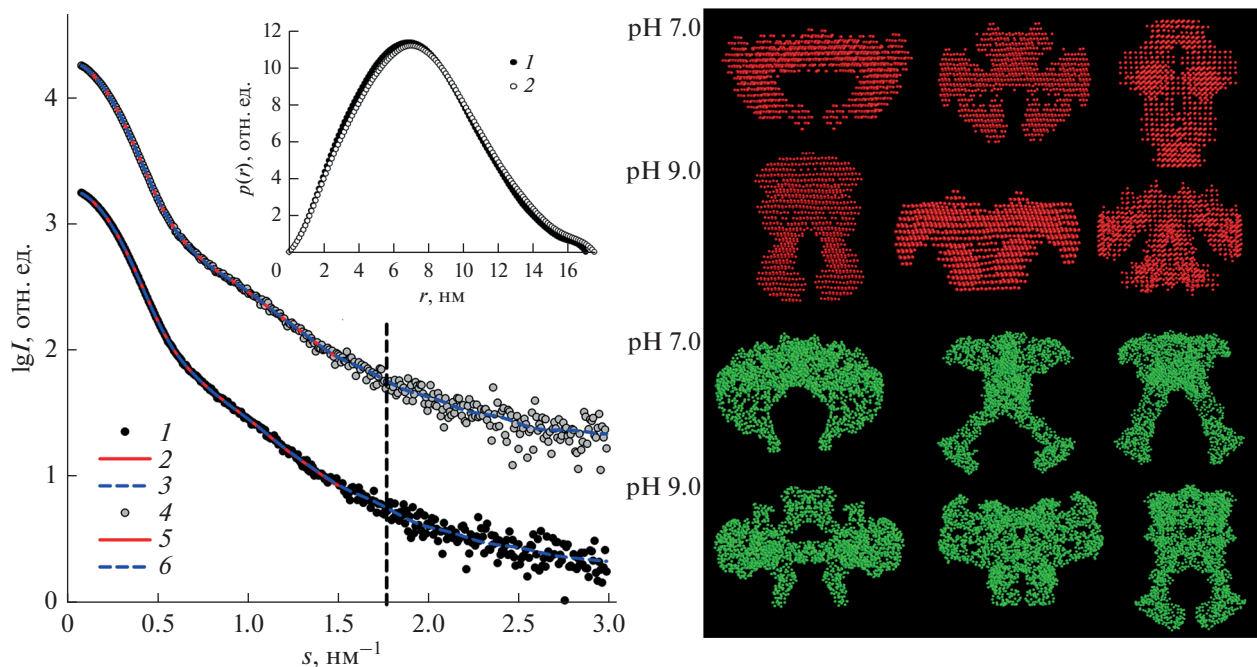


Рис. 4. *Ab initio* восстановление (программами DAMMIN и GASBOR) структуры ectoIRR с приближениями на разных интервалах кривых рассеяния (вертикальная пунктирная линия ограничивает область порядка 8 шенноновских каналов): 1 и 4 – экспериментальные данные МУПП от ectoIRR для pH 7.0 и 9.0 соответственно; 2 и 5 – расчетные кривые от моделей, полученные программой DAMMIN; 3 и 6 – расчетные кривые от моделей, полученные программой GASBOR. Справа – характерные *ab initio* формы ectoIRR: красные получены с помощью программы DAMMIN, зеленые – с помощью программы GASBOR. Вставка: функции распределения по расстояниям $p(r)$: 1 – pH 7.0; 2 – pH 9.0.

экспериментальные данные ectoIRR были приближены в диапазоне s до 8 шенноновских каналов, на котором частица может быть представлена как однородное тело (графики Порода, рис. 3). Это упрощение для такого белка, как ectoIRR, содержащего гетероатомы, необходимо для уменьшения степени неоднозначности решения обратных задач. Типичное восстановление формы молекулы IRR приведено на рис. 4. Объем полученных *ab initio* моделей соответствовал димерному состоянию белка (табл. 1).

Ab initio моделирование программами DAMMIN и GASBOR дали во всех случаях хорошее приближение к экспериментальным кривым рассеяния с $\chi^2 = 1.1–1.3$. Полученные формы (рис. 4, справа), хотя и имеют одни и те же значения инвариантов рассеяния, представляли собой сильно различающиеся по своей конформации структуры, что подтверждает предположение о полиморфизме ectoIRR в растворе, возникающее из-за большого количества гибких линкеров, соединяющих отдельные домены белка. Эти формы присутствовали как при восстановлении в растворе ectoIRR при pH 7.0, так и при pH 9.0 в диапазоне до 8 шенноновских каналов. На более длинном участке восстановленные с помощью программы DAMMIN формы были рыхлые и не имели определенной структуры. Поэтому одним из выводов данного

раздела является рекомендация использовать для *ab initio* моделирования структуры белков с большим количеством разупорядоченных фрагментов более короткие участки экспериментального профиля в соответствии с шенноновской оценкой. Для программы GASBOR такого ограничения нет, поскольку для этого метода возможно восстановление доменной структуры белка. Поэтому структурные модели рассчитывались в диапазоне до 15 шенноновских каналов. Таким образом, двумя различными методами восстановления структуры низкого разрешения были получены сходные формы с разнообразными конформациями. В целом, на этапе *ab initio* восстановления структуры по данным МУПП было показано, что решение обратной задачи в данном случае усложняется наличием полиморфизма ectoIRR.

Приближение данных МУПП атомной моделью эктодомена рецептора инсулина (ectoIR) – гомолога ectoIRR. Нами была произведена попытка приблизить экспериментальные данные МУПП от ectoIRR доступной атомной моделью гомологичного белка IR. Модель ectoIR и соответствующее приближение программой CRY SOL приведены на рис. 5. В данном случае приближался весь диапазон экспериментальных данных, поскольку использовалась неоднородная модель (в отличие от *ab initio* моделирования программой DAMMIN).

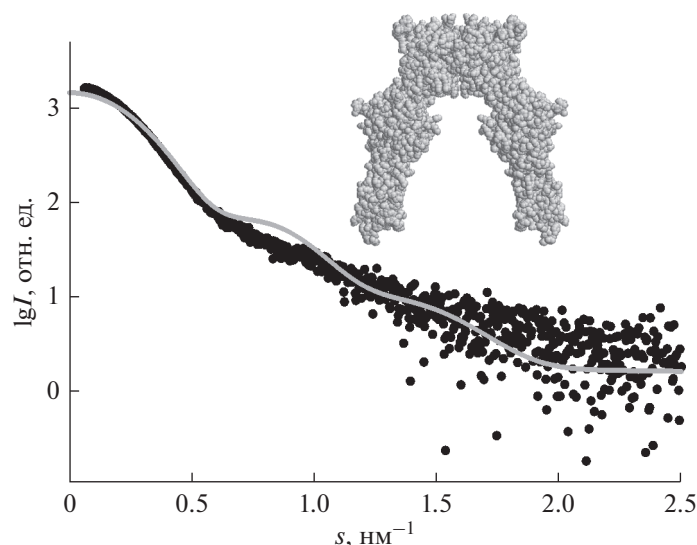


Рис. 5. Приближение кривой рассеяния ectoIRR атомной моделью гомологичного белка ectoIR. Экспериментальные данные МУПП от ectoIRR (черные точки) и приближение моделью ectoIR (серая линия). Вставка – атомная модель димера ectoIR.

Как видно из рисунка, модель плохо соответствует экспериментальной кривой от ectoIRR, по-видимому, вследствие другого конформационного состояния эктодомена инсулинового рецептора IR.

Молекулярная тектоника для анализа четвертичной структуры ectoIRR

В настоящей работе четвертичная структура тетрамера IRR была восстановлена на основании структуры доменов гомологичного инсулинового рецептора. Программой CORAL моделировалось взаимное расположение индивидуальных доменов в асимметрической части молекулы ectoIRR и конформация гибких линкеров в виде цепей из виртуальных остатков, соединяющих соответствующие домены. Симметрично связанная структура второго мономера генерировалась автоматически. В качестве структурного шаблона, также содержащего гликаны для гибридного моделирования ectoIRR, эктодомен рецептора инсулина (ectoIR) был взят из банка биологических данных малоуглового рассеяния, SASBDB (www.sasbdb.org), код доступа SASDHF2 [34]. В отдельных многократных запусках программы CORAL димер ectoIRR был представлен в виде двух полипептидных цепей, связанных между собой осью второго порядка (симметрия P2). Каждый мономер был разделен на следующие четыре фрагмента: аминокислоты от 1 до 308, 312–592, 596–655 и 755–909. Последовательные фрагменты были соединены линкерами из виртуальных остатков соответствующей длины. Третий и четвертый фрагменты в ходе моделирования смешались согласованным образом (одинаковые вращения, одинаковые сдвиги), чтобы сохранить

структурную целостность двух последних фибронектиновых доменов на С-конце молекулы.

Типичные реконструкции даны на рис. 6 вместе с соответствующими приближениями данных МУПП. Как видно из рисунка, модели не являются идентичными вследствие полиморфизма. При многократном восстановлении получается множество различных конформаций, которые тем не менее имеют общие черты, и их можно разделить на группы по схожести. В целом, полученные модели в основном отражают структуры низкого разрешения *ab initio* восстановления формы, т.е. структуры со сближенными фибронектиновыми доменами (закрытые формы) и структуры, где эти домены пространственно разделены (открытые формы).

Несмотря на различие конформационных форм все модели, как и в случае *ab initio* восстановления, показали хорошее приближение к экспериментальным кривым рассеяния с $\chi^2 = 1.0\text{--}1.2$.

Анализ подвижности доменов белка IRR. Степень подвижности доменов белка ectoIRR была оценена с помощью программы EOM (Ensemble Optimization Method). Как и при моделировании с помощью метода молекулярной тектоники, использовалась структура гомологичного белка IR, при этом мономер ectoIR был разбит на два домена (содержащих остатки 1–465 и 472–910 соответственно), разделенных гибким звеном из 7 остатков. Программа EOM создавала ансамбль из 10000 моделей со случайно выбранными взаимными ориентациями доменов, а затем, используя генетический минимизационный алгоритм, выбирала из этого ансамбля оптимизированный набор моделей, комбинация которых наилучшим

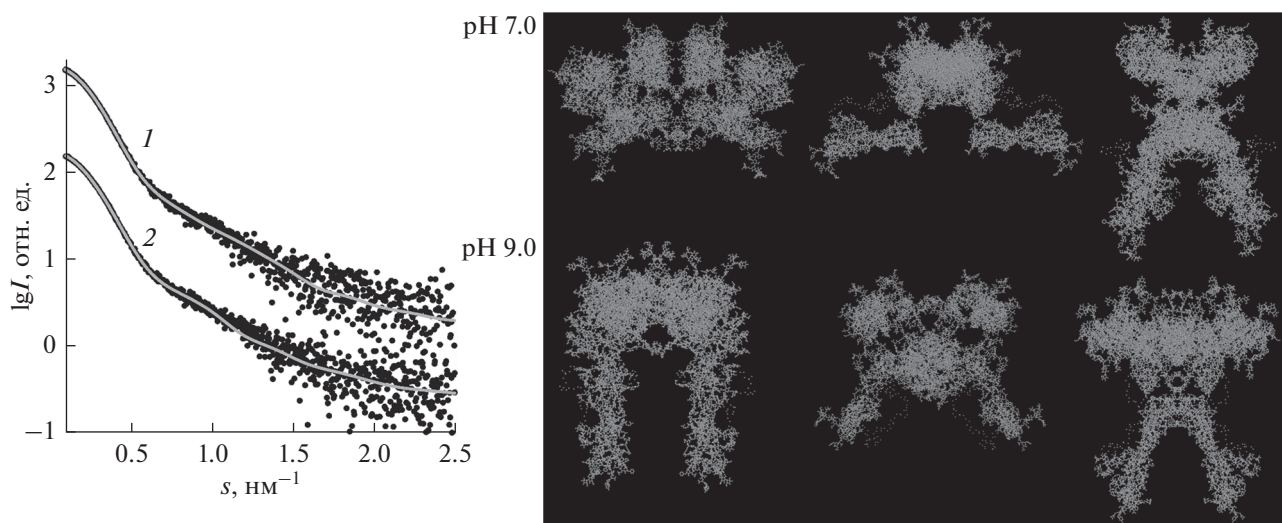


Рис. 6. Приближение данных МУРР методом молекулярной тектоники для рассеяния от ectoIRR при pH 7.0 (1) и pH 9.0 (2). Экспериментальные данные показаны черными точками. Приближение с помощью моделей даны серыми линиями. Справа – типичные модели тетрамера ectoIRR.

образом приближала малоугловые данные. Исходный ансамбль состоял из димерных молекул ectoIR, половина из которых являлась гомодимерами (т.е. обладала P2-симметрией), вторая половина – гетеродимерами (симметрия P1). Однако при таком выборе доменов адекватно приблизить экспериментальную кривую не удалось: по-видимому, число степеней свободы при создании случайного ансамбля оказалось недостаточным. После этого каждый мономер ectoIR был разбит на пять частей (содержащих остатки 1–290, 299–450, 459–571, 580–756, 790–908, соответственно, и соединенных между собой гибкими звеньями) и проведено моделирование с помощью программы ЕОМ. В этом случае число степеней свободы при создании ансамбля оказалось достаточным, и экспериментальные данные IRR удалось хорошо приблизить $\chi^2 = 2.3\text{--}2.8$ (рис. 7). При этом из сравнения распределения радиусов инерции моделей исходного и оптимизированного ансамблей видно, что были выбраны достаточно компактные структуры с радиусами инерций в диапазоне от 4.5 до 6.0 нм, степень подвижности доменов можно отнести к умеренной, что подтверждается оценками параметрами подвижности Rflex, если для исходного случайного ансамбля этот параметр составлял 87.3%, то для оптимизированного набора моделей он составил 61.6%. Таким образом, одним из возможных состояний системы IRR может являться ансамбль умеренно подвижных димеров, имеющих не менее 3–5 гибких звеньев, разделяющих домены белка.

Модельно-независимый анализ минимального числа компонент (конформаций) белка IRR в растворе. Минимальное число независимых и значимых компонент, необходимых для описания экс-

периментального набора данных МУРР в сочетании с эксклюзионной гель-хроматографией (ЭГХ-МУРР), можно оценить с помощью сингулярного разложения. Для этого набор кривых ЭГХ-МУРР можно представить в виде матрицы, взяв экспериментальные кривые, соответствующие хроматографическому пику (номера кривых между 1600 и 1800, рис. 2). Далее этот набор используется в качестве входных данных для программы SVDPLOT [17], которая проводит его разложение на сингулярные векторы и определяет соответствующие сингулярные значения. На рис. 8 представлены результаты такого разложения, где сингулярные значения расположены в порядке убывания (рис. 8а и 8б), а соответствующие сингулярные векторы для лучшей визуализации смещены вдоль вертикальной оси друг относительно друга. Из рис. 8в видно, что сингулярные значения выходят на слабо меняющееся плато в районе 10–15 компонент, а первые 12 сингулярных векторов (расположенных сверху вниз) имеют неслучайные осцилляции вокруг горизонтальной линии. Это позволяет сделать вывод о наличии в системе минимум 10–15 независимых конформаций белка, что хорошо согласуется с выводом о подвижности доменов ectoIRR белка при моделировании методом оптимизированного ансамбля (ЕОМ), а также многообразием форм, полученных при *ab initio* моделировании.

Таким образом, проведенный модельно-независимый анализ и демонстрация наличия в растворе ectoIRR 10–15 конформаций белка подтверждают результаты моделирования структуры этого домена *ab initio* и гибридными методами. Поэтому для белка, который участвует в регуляции кислотно-щелочного баланса в организме и

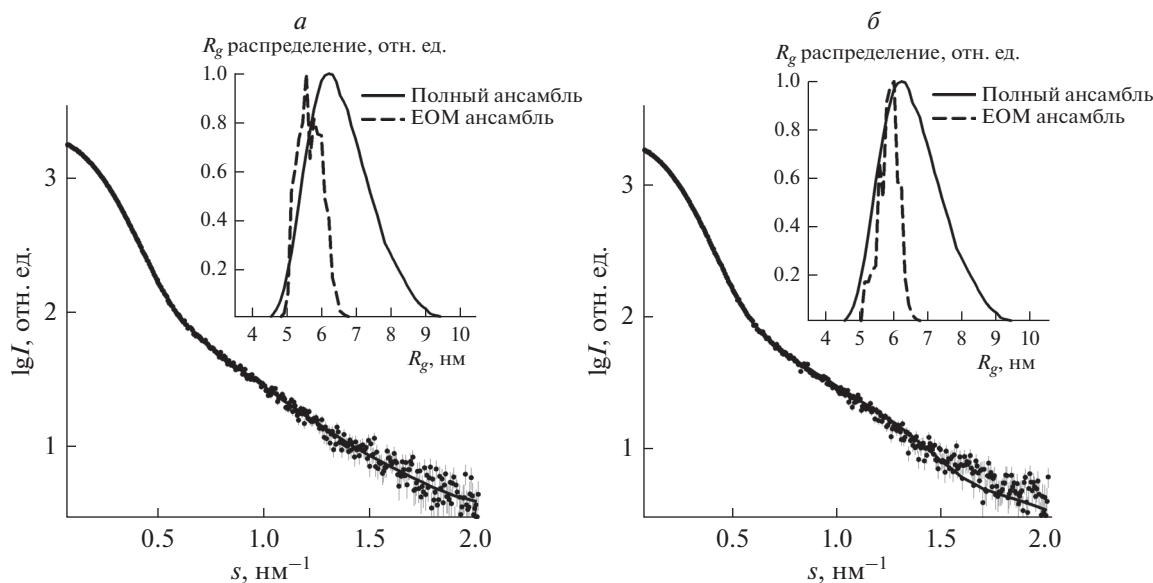


Рис. 7 Анализ разупорядоченности ectoIRR при pH 7 (а) и при pH 9 (б). Экспериментальные данные МУРР показаны черными точками, приближение с помощью разупорядоченных димеров, построенных программой EOM – сплошной линией. На вставке – распределение R_g в исходном наборе (сплошная линия) и в оптимизированном ансамбле (штриховая линия).

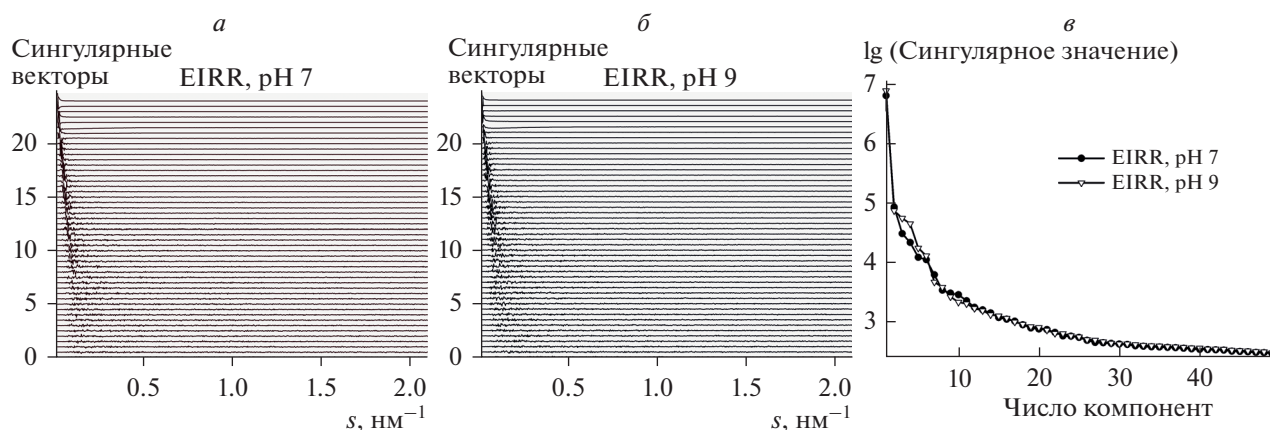


Рис. 8. Разложение набора данных ЭГХ-МУРР для ectoIRR на сингулярные векторы и определение соответствующих сингулярных значений.

структура которого должна определяться pH раствора, необходимо также провести количественный анализ содержания тех или иных конформаций в зависимости от кислотности среды. Для этого использовалась программа OLIGOMER [17]. Для определения объемных фракций v_k каждой компоненты смеси программа нашла линейную комбинацию интенсивностей рассеяния от каждой компоненты $I_k(s)$ в смеси конформаций. Для расчетов были использованы наиболее часто встречающихся формы белка, полученные при моделировании гибридным методом программой CORAL. Эти формы можно объединить в две ос-

новные группы: закрытая и открытая конформация. В результате было показано, что для pH 7.0 характерна открытая (релаксированная) конформация со значением $v_1 = 0.84$ и присутствует фракция закрытой конформации с $v_2 = 0.16$. Для pH 9.0 фракция с открытой конформацией $v_1 = 0.15$ и в основном в растворе присутствует активная, закрытая конформация со сближенными фибронектиновыми доменами с $v_2 = 0.85$. Полученные приближения соответствовали экспериментальным данным с $\chi^2 = 2.8$ для pH 7.0 и $\chi^2 = 1.5$ для pH 9.0 (рис. 9).

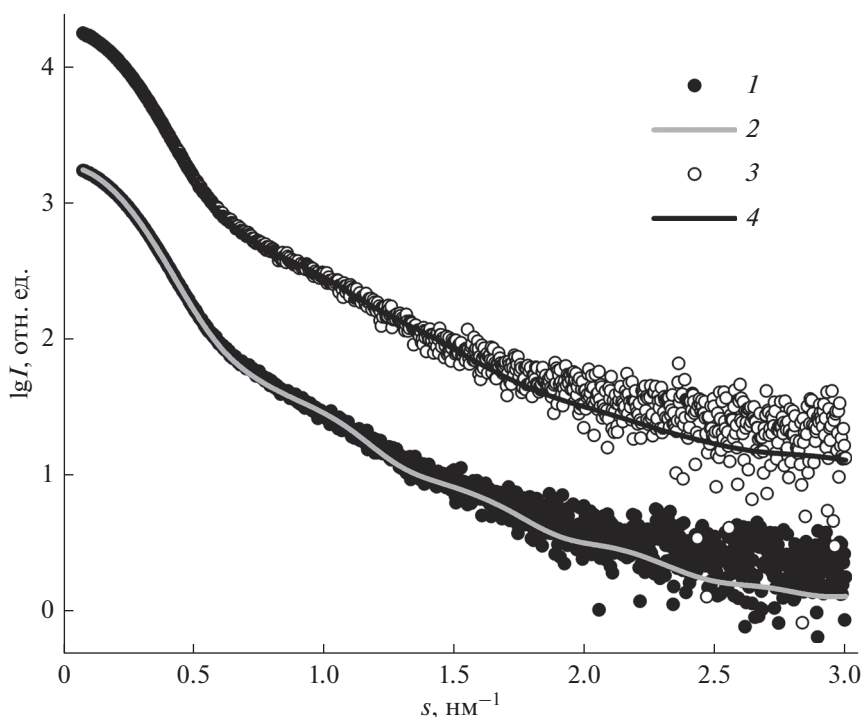


Рис. 9. Результат моделирования программой OLIGOMER смесей различных конформаций: 1 и 3 — экспериментальные кривые МУРР ectoIRR при pH 7.0 и 9.0, соответственно; 2 и 4 — теоретическое рассеяние от смеси конформаций при pH 7.0 и 9.0, соответственно.

ЗАКЛЮЧЕНИЕ

В данной работе проблема неоднозначности решения обратных задач и восстановления трехмерной структуры белка по данным МУРР рассматривалась на примере исследования строения эктодомена рецептора, подобного рецептору инсулина (ectoIRR), в растворе. Мы показали, что этот белок, имея множество разупорядоченных гибких фрагментов, существует в растворе в виде набора конформаций. Был сделан вывод, что конформационные перестройки происходят не только на локальном уровне, как предполагалось раньше [13], но значительно меняется общая форма эктодомена. Полученные данные соответствуют выводам недавно опубликованной работы, где методом атомно-силовой микроскопии было показано разнообразие форм полноразмерного рецептора, подобного рецептору инсулина [36].

Мы продемонстрировали, что последовательный подход к исследованию белков, обладающих полиморфизмом, на примере эктодомена IRR, начиная от определения их общих структурных параметров и *ab initio* восстановления формы с последующим моделированием жесткими телами (методом молекулярной тектоники) и гибридными методами, приводит к достаточно полному описанию структуры белка. Для эктодомена IRR это было особенно важно, так как предыдущие исследования не позволили нам показать измене-

ние конформации белка при изменении pH, т.е. при выполнении рецептором своих биологических функций.

Второй задачей нашего исследования была демонстрация на достаточно сложном примере современных возможностей малоуглового рассеяния. Мы показали, что при решении обратных задач МУРР с принципиальной невозможностью получения однозначного структурного ответа, тем не менее, можно достаточно полно и количественно описать конформационное состояние рассеивающего объекта. При этом, анализируя данные малоуглового рассеяния, стоит критично оценивать полученные результаты и при возможности использовать комплементарные методы. В нашей работе, где мы поставили цель показать современные подходы и возможности МУРР, мы принципиально не прибегали к данным других структурных методов, хотя именно их использование позволяет значительно сузить коридор возможных решений обратной задачи малоуглового рассеяния.

Большое значение также имеет выбор правильного диапазона данных при первичной оценке общих геометрических параметров, таких как исключенный объем, радиус инерции, максимальный размер, а также при моделировании структуры разными методами. При *ab initio* моделировании формы следует использовать только

первые восемь-десять шенноновских каналов, а дальнейшее увеличение интервала угловых векторов приводит к неадекватному восстановлению структуры с мнимо подробными деталями. С другой стороны, для гибридных методов диапазон углов должен быть максимально широким.

Очень важно использование онлайн эксклюзионной гель-хроматографии: восстановление структуры с разрешением порядка 1 нм в малоугловом рассеянии возможно только для мономерных растворов белка. Однако даже для такого сложного объекта исследования, как эктодомен рецептора, подобного рецептору инсулина, с его гибкостью и полиморфизмом удалось построить структурные модели, отражающие его биологические свойства. Мы надеемся, что наша работа полезна не только в плане изучения конкретно структуры ectoIRR, но также описывает ряд значимых методик МУРП и сценарий, которого следует придерживаться в процессе исследований структурно сложных биологических объектов и комплексов.

Авторы приносят благодарность С.М. Jeffries и Д.И. Свергуну за помощь в проведении малоуглового эксперимента на станции P12 BioSAXS (EMBL, Гамбург) и ценную дискуссию.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Источники финансирования. Работа выполнена при поддержке Министерства науки и высшего образования в рамках Государственного задания ФНИЦ “Кристаллография и фотоника” РАН и Российского фонда фундаментальных исследований (проекты № 17-00-00487 КОМФИ, № 17-00-00488 КОМФИ и № 20-04-00959а).

Соответствие принципам этики. Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Feigin L.A., Svergun D.I. 1987. *Structure analysis by small-angle x-ray and neutron scattering*. New York: Plenum Press. 335 p.
2. Svergun D.I., Shtykova E.V., Volkov V.V., Feigin L.A. 2011. Small-angle X-ray scattering, synchrotron radiation, and the structure of bio- and nanosystems. *Crystallography Reports*. **56** (5), 725–750.
3. Graewert M.A., Svergun D.I. 2013. Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* **23** (5), 748–754.
4. Blanchet C.E., Svergun D.I. 2013. Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu. Rev. Phys. Chem.* **64**, 37–54.
5. Schroer M.A., Blanchet C.E., Gruzinov A.Y., Grawert M.A., Brennich M.E., Hajizadeh N.R., Jeffries C.M., Svergun D.I. 2018. Smaller capillaries improve the small-angle X-ray scattering signal and sample consumption for biomacromolecular solutions. *J. Synchrotron Radiation*. **25**, 1113–1122.
6. Svergun D.I., Koch M.H., Timmins P.A., May R.P. 2013. *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. London: Oxford University Press. 358 p.
7. Franke D., Petoukhov M.V., Konarev P.V., Panjkovich A., Tuukkanen A., Mertens H.D.T., Kikhney A.G., Hajizadeh N.R., Franklin J.M., Jeffries C.M., Svergun D.I. 2017. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50** (Pt 4), 1212–1225.
8. Petoukhov M.V., Franke D., Shkumatov A.V., Tria G., Kikhney A.G., Gajda M., Gorba C., Mertens H.D.T., Konarev P.V., Svergun D.I. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystal.* **45** (2), 342–350.
9. Volkov V.V., Svergun D.I. 2003. Uniqueness of ab initio shape determination in small angle scattering. *J. Appl. Crystal.* **36**, 860–864.
10. Shier P., Watt V.M. 1989. Primary structure of a putative receptor for a ligand of the insulin family. *J. Biol. Chem.* **264** (25), 14605–14608.
11. Deev I.E., Vasilenko K.P., Kurmangaliev, E.S., O.V., Popova N.V., Galagan Y.S., Burova E.B., Zozulya S.A., Nikol'skii N.N., Petrenko A.G. 2006. Effect of changes in ambient pH on phosphorylation of cellular proteins. *Dokl. Biochem. Biophys.* **408**, 184–187.
12. Deyev I.E., Sohet F., Vassilenko K.P., Serova O.V., Popova N.V., Zozulya S.A., Burova E.B., Houillier P., Rzhovsky D.I., Berchatova A.A., Murashev A.N., Chugunov A.O., Efremov R.G., Nikol'sky N.N., Bertelli E., Eladari D., Petrenko A.G. 2011. Insulin receptor-related receptor as an extracellular alkali sensor. *Cell Metab.* **13**, 679–689.
13. Shtykova E.V., Petoukhov M.V., Mozhaev A.A., Deyev I.E., Dadinova L.A., Loshkarev N.A., Goryashchenko A.S., Bocharov E.V., Jeffries C.M., Svergun D.I., Batishchev O.V., Petrenko A.G. 2019. The dimeric ectodomain of the alkali-sensing insulin receptor-related receptor (ectoIRR) has a droplike shape. *J. Biol. Chem.* **294** (47), 17790–17798.
14. Blanchet C.E., Spilotros A., Schwemmer F., Graewert M.A., Kikhney A., Jeffries C.M., Franke D., Mark D., Zengerle R., Cipriani F., Fiedler S., Roessle M., Svergun D.I. 2015. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystal.* **48** (Pt 2), 431–443.
15. Jeffries C.M., Graewert M.A., Blanchet C.E., Langley D.B., Whitten A.E., Svergun D.I. 2016. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nature Protocols*. **11** (11), 2122–2153.

16. Manalastas-Cantos K., Konarev P.V., Hajizadeh N.R., Kikhney A.G., Petoukhov M.V., Molodenskiy D.S., Panjkovich A., Mertens H.D.T., Gruzinov A., Borges C., Jeffries C.M., Svergun D.I., Franke D. 2021. ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Cryst.* **54**, 343–355.
17. Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J., Svergun D.I. 2003. PRIMUS – a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystal.* **36**, 1277–1282.
18. Panjkovich A., Svergun D.I. 2018. CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics.* **34** (11), 1944–1946.
19. Guinier A. 1939. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris).* **12**, 161–237.
20. Glatter O., Kratky O. 1982. *Small angle X-ray scattering*. London: Academic Press. 515 p.
21. Debye P. 1947. Molecular-weight determination by light scattering. *J. Phys. Colloid Chem.* **51** (1), 18–32.
22. Hajizadeh N.R., Franke D., Jeffries C.M., Svergun D.I. 2018. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci. Rep.* **8** (1), 7204.
23. Graewert M.A., Franke D., Jeffries C.M., Blanchet C.E., Ruskule D., Kuhle K., Flieger A., Schafer B., Tartsch B., Meijers R., Svergun D.I. 2015. Automated pipeline for purification, biophysical and X-Ray analysis of biomacromolecular solutions. *Sci. Rep.* **5**, 10734.
24. Konarev P.V., Svergun D.I. 2015. A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCRJ.* **2**, 352–360.
25. Franke D., Jeffries C.M., Svergun D.I. 2018. Machine learning methods for X-Ray scattering data analysis from biomacromolecular solutions. *Biophys. J.* **114** (11), 2485–2492.
26. Petoukhov M.V., Svergun D.I. 2015. Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr. Section D. Biol. Crystallogr.* **71**, 1051–1058.
27. Svergun D.I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503.
28. Svergun D.I. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76** (6), 2879–2886.
29. Svergun D.I., Petoukhov M.V., Koch M.H.J. 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80** (6), 2946–2953.
30. Svergun D.I., Barberato C., Koch M.H.J. 1995. CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
31. Petoukhov M.V., Svergun D.I. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89** (2), 1237–1250.
32. Bernado P., Mylonas E., Petoukhov M.V., Blackledge M., Svergun D.I. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129** (17), 5656–5664.
33. Kozin M.B., Svergun D.I. 2001. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.
34. Whitten A.E., Smith B.J., Menting J.G., Margetts M.B., McKern N.M., Lovrecz G.O., Adams T.E., Richards K., Bentley J.D., Trewella J., Ward C.W., Lawrence M.C. 2009. Solution structure of ectodomains of the insulin receptor family: The ectodomain of the type 1 insulin-like growth factor receptor displays asymmetry of ligand binding accompanied by limited conformational change. *J. Mol. Biol.* **394** (5), 878–892.
35. De Meyts P. 2008. The insulin receptor: A prototype for dimeric, allosteric membrane receptors? *Trends Biochem. Sci.* **33** (8), 376–384.
36. Batishchev O.V., Kuzmina N.V., Mozhaev A.A., Goryashchenko A.S., Mileshina E.D., Orsa A.N., Bocharov E.V., Deyev I.E., Petrenko A.G. 2021. Activity-dependent conformational transitions of the insulin receptor-related receptor. *J. Biol. Chem.* <https://doi.org/10.1016/j.jbc.2021.100534>

The Ambiguity Problem in Solving Inverse Problems of Small-angle Scattering: A Consistent Approach by the Example of Insulin Receptor-related Receptor. Methods for SAXS Data Interpretation

M. V. Petoukhov^{1,*}, P. V. Konarev¹, V. V. Volkov¹, A. A. Mozhaev^{1,2}, E. V. Shtykova¹

¹*Shubnikov Institute of Crystallography of FSRC “Crystallography and Photonics”,
Russian Academy of Sciences, Moscow, 119333 Russia*

²*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,
Moscow, 117997 Russia*

*e-mail: maxim@embl-hamburg.de

The construction of three-dimensional models of protein macromolecules is a challenging problem due to the poor convergence of the inverse problem of reconstructing a three-dimensional structure from a one-dimensional small-angle scattering profile. The target function can have several local minima, which leads to the

dependence of the solution on the starting values of the model parameters and on the method for finding the global minimum. The problem of creating structural models is also complicated by averaging the scattering pattern over all orientations of particles in space, and in the presence of polydispersity and/or polymorphism, by the size and shape distribution of scattering objects. In this work, the problem of ambiguity in solving inverse problem and restoring the three-dimensional structure of a protein is considered by the example of studying the structure of the ecto-domain of the insulin receptor-related receptor (ectoIRR) in solution. The work presents a consistent approach to solving this problem, starting from the determination of general structural parameters and *ab initio* shape reconstruction to modeling by rigid bodies (by the method of molecular tectonics), hybrid methods and analysis of scattering profiles by singular vector decomposition.

Keywords: small-angle X-ray scattering (SAXS), structural modeling, receptor tyrosine kinases, insulin receptor-related receptor