# КЛАССИФИКАЦИЯ СТРУКТУР РОДОПСИНОВОГО РЕЦЕПТОРА СОВРЕМЕННЫМИ МЕТОДАМИ СТРУКТУРНОЙ БИОИНФОРМАТИКИ<sup>\*</sup>

© 2012 г. Г.В. Новиков<sup>1</sup>, В.С. Сивожелезов<sup>1\*\*</sup>, А.С. Шебанова<sup>2</sup>, К.В. Шайтан<sup>2</sup>

<sup>1</sup> Институт биофизики клетки РАН, 142290 Пущино Московской обл., ул. Институтская, 3; факс: (4967)330-509, электронная почта: vsivo00@gmail.com, vsivo@icb.psn.ru <sup>2</sup> Биологический факультет МГУ им. М.В. Ломоносова, 119991 Москва

Поступила в редакцию 20.12.11 После доработки 19.01.12

Создана классификация кристаллографических структур родопсинов быка и кальмара, соответствующих различным стадиям фотоцикла рецептора. С помощью сервера ProCKSI проведено сравнение выбранных пространственных структур, на основании которого построены классификационные схемы (дендрограммы). Для сравнения пространственных структур трансмембранных белков подобрано оптимальное сочетание методов (консенсусный метод), реализуемых на ProCKSI. Кластеризация структур также осуществлена методом главных компонент, с помощью которого получено хорошее соответствие классификации, построенной на основе данных консенсусного метода ProCKSI. Анализ полученных результатов позволил определить основные движения отдельных трансмембранных доменов данных белков, которые удалось связать с различными стадиями фотоактивации родопсина. Найденная комбинация методов может быть использована в качестве современного аналитического инструмента для изучения конформационной динамики мембранных рецепторов.

КЛЮЧЕВЫЕ СЛОВА: рецептор, родопсин, активация, конформация, классификация.

Рецепторы, сопряженные с G-белком (GPCR) относятся к наиболее обширному классу трансмембранных рецепторов [1]. На данный момент уже расшифровано ~2000 генов, кодирующих ~900 различных GPCR. Известно, что данный класс мембранных рецепторов играет важнейшую роль в самых различных биологических процессах, опосредуя отклик на такие внеклеточные сигналы, как гормоны и нейромедиаторы, пахучие вещества и световые стимулы. В связи с этим очевидна терапевтическая значимость GPCR, являющихся основными мишенями для большинства лекарственных средств [2–4].

Известно, что GPCR – белки с выраженной структурной динамикой, которые способны принимать различные конформации в зависимости от ряда внешних условий [5]. При этом молекула рецептора обладает способностью взаимодействовать с различными лигандами, что в итоге может приводить к изменению ее конформации, а также к запуску специфических каскадов биохимических реакций [6]. Зрительный рецептор родопсин млекопитающих (рис. 1) – один из наиболее изученных семиспиральных трансмембранных белков сегодня. Известно, что при поглощении кванта света происходит изомеризация хромофора 11-цис-ретиналя в транс-форму. Считается, что данная реакция – основная предпосылка небольших конформационных изменений в отдельных трансмембранных (TM) спиралях рецептора [7]. С помощью ряда биофизических методов было показано также, что полный цикл фотоактивации родопсина сопровождается значительным смещением внутриклеточного конца шестой спирали рецептора на 5-6 Å по направлению от плоскости α-спирального пучка. Кроме того, были

Принятые сокращения: GPCR – G-protein coupled receptors, TM – трансмембранный, PDB – банк данных белковых структур, ProCKSI – сервер «Protein (Structure) Comparison, Knowledge, Similarity and Information», RMSD – среднеквадратическое отклонение, PCA – метод главных компонент.

<sup>\*</sup> Первоначально английский вариант рукописи был опубликован на сайте «Biochemistry» (Moscow), Papers in Press, BM 11-325, 00.00.2012.

<sup>\*\*</sup> Адресат для корреспонденции.



**Рис. 1.** Общая топология семитрансмембранного рецептора на примере родопсина млекопитающих (pdb id 3pqr). Градациями серого цвета показан ход цепи от внеклеточного *N*-конца (светлым), до внутриклеточного *C*-конца (темным). Лиганд *транс*-ретиналь представлен в виде шаровой модели

обнаружены синхронные движения пятой, а также небольшое вращение седьмой спиралей рецептора (рис. 2). В итоге совокупность данных конформационных изменений в ТМ-фрагменте рецептора приводила к открытию щели в цитоплазматической области данной молекулы, в которую впоследствии мог входить G-белок.

На сегодня методом рентгенокристаллографии уже расшифровано несколько структур родопсина млекопитающих, которые соответствуют различным стадиям фотоактивации данного рецептора. В результате сравнительного анализа данных экспериментальных структур можно сформировать целостное представление о процессе фотоактивации родопсина, несмотря на возможную нехватку промежуточных структур в данном цикле. Таким образом, исследуя конформационное разнообразие мембранных рецепторов, представленное, например, в существующих экспериментальных структурах, можно в итоге попытаться оценить взаимосвязь динамики их структуры (движения отдельных α-спиралей) и функции (передачи сигнала).

БИОХИМИЯ том 77 вып. 5 2012

С помощью современных методов структурной биоинформатики нам удалось создать структурно-функциональную классификацию экспериментальных структур рецептора родопсина. Отметим, что ранее используемые в данной работе методы были проверены только на глобулярных (водорастворимых) белках. Например, для ряда семейств протеинкиназ уже была получена похожая классификация [8]. Известно, что водорастворимые белки обладают в целом более разнообразными мотивами укладки полипептидной цепи по сравнению с ТМ-доменом мембранных белков. В последнем случае гидрофобным окружением мембраны существенно ограничивается общее число возможных вариантов упаковки α-спиралей в пучке ТМ-домена. Таким образом, нам предстояла проверка методов структурного сравнения на топологически идентичных структурах, незначительно различающихся лишь пространственным расположением отдельных α-спиралей.

С помощью строго определенной комбинации вышеупомянутых методов нам удалось выя-



Рис. 2. Пространственная суперпозиция неактивного (темнового) и активированного (MetaII\*) состояния родопсина млекопитающих. Структуры показаны со стороны клеточного пространства. Светлым цветом показана неактивная (pdb id lgzm), а темным – активная (pdb id 3pqr) конформации рецептора. Цифрами обозначены номера трансмембранных спиралей, а стрелками – направление структурных изменений, сопровождающих переход из одной формы в другую. Наибольшие структурные изменения отмечаются в пятой и в шестой спиралях рецептора. При этом седьмая спираль рецептора также претерпевает небольшие изменения, вращаясь против часовой стрелки

вить основные закономерности распределения активных и неактивных форм рецептора родопсина. При этом наиболее достоверные результаты были получены в результате применения комплексного подхода к анализу полученных результатов. Данная методика заключалась, вопервых, в визуальном анализе полученной классификации экспериментальных структур, вовторых, в поиске ответов на вопросы, связанные с той или иной кристаллографической структурой, в соответствующей литературе (в дальнейшем исходную кристаллографическую работу с данной пространственной структурой будем называть первоисточником), и в-третьих, в собственной проверке результатов первоисточника. Последняя проводилась путем сравнения полученных численных значений среднеквадратичного отклонения (RMSD), определяющего попарные различия между исследуемыми структурами, а также исследования этих структур с помощью программы-визуализатора.

#### МЕТОДЫ ИССЛЕДОВАНИЯ

Исходный набор данных. Использованные в данной работе пространственные структуры мембранных рецепторов были получены из белкового банка данных (http://www.rcsb.org/pdb). Нами были исследованы все известные на момент проведения данной работы кристаллографические структуры рецепторов родопсина.

Сервер Protein (Structure) Comparison, Knowledge, Similarity and Information (ProCKSI). Основная часть работы, включая пространственное выравнивание исследуемых структур, построение и анализ дистанционных матриц, а также кластеризацию результатов, была проведена на внешнем сервере ProCKSI. Сервер ProCKSI представляет собой систему, предназначенную для сравнения и анализа пространственных структур белков с помощью ряда интегрированных алгоритмов [9]. Отметим, что в рамках данной работы мы остановились на изучении именно алгоритмов пространственного выравнивания, исключив методы выравнивания по аминокислотным последовательностям. Данное решение было связано с тем, что различные семейства мембранных рецепторов при низкой гомологии по аминокислотной последовательности (типичное значение составляет 20-30%) обладают крайне высоким сходством в пространственной организации семиспирального трансмембранного домена. В результате были проверены все методы, представленные в ProCKSI, в первую очередь основанные на попарном измерении RMSD для исследуемых структур. В этот ряд вошли такие методы, как DaliLite, основанный на пространственном выравнивании дистанционных матриц, TM-align, основанный на динамическом программировании, а также давно известный метод комбинаторного перебора СЕ и, наконец, методика Vorolign, основанная на контактах Вороного. Важно отметить, что каждый из представленных методов позволяет проводить как классическое измерение RMSD, так и специфические оценки подобия структур. Таким образом, в рамках конкретного исследования существует возможность определения наиболее подходящих методов для конкретного типа исследуемых структур. В итоге в качестве выходных данных с помощью ProCKSI можно получить классификационное распределение исследуемых структур в форме, схожей с филогенетическими деревьями (дендрограммами).

Оценка структурного сходства белков. Кратко описаны отдельные алгоритмы, интегрированные в ProCKSI, а также их специфические критерии для оценки пространственного сходства исследуемых структур.

Наиболее традиционная мера, используемая для сравнения пространственных структур белков, – RMSD [10, 11]. Данная величина представляет собой среднее расстояние между атомами (обычно главной цепи) выравниваемых белков. Считается, что, изучая любую динамическую систему, необходимо учитывать совокупность флуктуаций в рассматриваемой окрестности. При этом данный показатель может описывать и среднеквадратичные флуктуации исследуемой системы:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}, \qquad (1)$$

где  $\delta_i$  — расстояние в каждой из N пар одинаковых атомов главной цепи белка (обычно *Сб*).

Отметим, что наряду с оценкой среднеквадратического отклонения в каждом методе ProCKSI есть более специфические критерии для оценки сходства исследуемымх структур. Например, авторами метода TM-align [12] была также предложена специфическая оценка, отражающая общую степень сходства общей топологии между исследуемыми структурами – Тороlogical Score (TM-score):

TM-score = max [ 
$$1/L_{\text{target}} \sum_{l}^{L_{\text{aligned}}} \frac{1}{1 + (d_i/d_0(L_{\text{target}}))}$$
]. (2)

Здесь L<sub>target</sub> и L<sub>aligned</sub> – длины последовательностей белка-мишени и выравненного участка

соответственно;  $d_i$  — расстояние между *i*-ми парами оснований,  $d_0(L_{\text{target}})$  — нормирующий коэффициент шкалы расстояний.

Данный критерий характеризует сходство между двумя структурами с помощью значения, заключенного в интервале (0,1], где 1 соответствует двум структурам с абсолютно идентичной пространственной топологией. При этом значение ниже 0,2 соответствует случайно выбранным структурам, в то время как показатель >0,5может свидетельствовать о том, что данные структуры имеют общую укладку полипептидной цепи в достаточно грубом приближении.

Еще одним примером может служить другой специфический критерий – V-score, используемый в методе Vorolign и являющийся нормализованной мерой оценки сходства между исследуемыми структурами в рамках данного метода [13]. В вычислительной геометрии диаграмма Вороного представляет собой особый способ разложения метрического пространства, определяемый расстояниями по отношению к выделенному дискретному набору объектов данного пространства. В простейшем случае имеется совокупность точек S на плоскости, называемых сайтами Вороного. При этом каждый сайт S может быть заключен в специфический объем V(S), называемый ячейкой Вороного. В данном объеме, в свою очередь, можно также рассмотреть совокупность точек, находящихся в окрестности заданной стартовой точки S. Для оценки степени структурного сходства между исследуемыми белками используются Сβ-атомы отдельных аминокислотных остатков (для глицина - $C\alpha$ -атомы), в качестве исходного набора точек евклидового пространства. Последующее преобразование данного пространства в форму выпуклого многогранника в результате позволяет рассматривать исходный набор данных в виде совокупности ячеек Вороного. Например, для двух сравниваемых белков  $X = x_1 x_2 \dots x_p$  и  $Y = y_1 y_2 \dots y_q$ , набор ближайших соседей заданного основания  $x_i$  определяется как  $N(x_i) = (x_{i_1}, x_{i_2}, ..., x_{i_n})$ . Таким образом, для оценки сходства между двумя основаниями  $x_i$  и  $y_i$  в конечном счете необходимо рассчитать сходство между соседними наборами (ячейками) оснований, которые задаются соответствующими функциями  $N(x_i)$  и  $N(y_i)$ . Сходство между двумя отдельными основаниями  $x_{i_k}$  и  $y_{i_l}$ , расположенными в соседних ячейках  $k = \hat{N}(x_i)$  и  $l = N(y_i)$  определяется как

Sim  $(x_{i_k}, y_{j_l}) = \omega 1$  AA  $(x_{i_k}, y_{j_l}) + \omega 2$  SSE  $(x_{i_k}, y_{j_l})$ . (3)

В данном выражении АА отражает сходство между отдельными аминокислотными остатками, а SSE — показатель сходства элементов вто-

БИОХИМИЯ том 77 вып. 5 2012

ричной структуры. При этом обе данных величины определяются в процессе подсчета значений соответствующих матриц. Наконец, сходство между двумя соседними наборами ячеек оценивается методом динамического программирования с учетом равенства (3), а также при добавлении коэффициента  $p_u$  для учета всех невыравненных пар оснований. При этом величина S(k,l) динамической матрицы S задается так:

$$S(k,l) = \begin{cases} S(k-1, l-1) + Sim(x_{i_k}, y_{j_l}) \\ S(k-1, l) - p_u \\ S(k, j-l) - p_u \end{cases}.$$
 (4)

В итоге Sim( $N(x_i)$ ,  $N(y_j)$ ), соответствующая элементу последнего ряда и столбца динамической матрицы S, используется для расчета сходства между двумя окружениями оснований  $x_i$  и  $y_j$ .

Алгоритм построения консенсуса ProCKSI. Основная функция ProCKSI заключается в сравнении исследуемых пространственных структур с помощью ряда интегрированных алгоритмов [7]. В итоге полученные результаты преобразуются в форму, наиболее удобную для последующего анализа. Отметим, что с помощью ProCKSI можно осуществить объединение полученных различными методами результатов в консенсус для заданного набора данных. Объединение проводится посредством усреднения результатов структурного выравнивания. Таким образом, с помощью данного механизма можно существенно расширить первоначальный набор данных, отражающих сходство исследуемых структур, получая в итоге различные консенсусы. Посредством добавления либо исключения отдельных компонентов можно обнаружить, какой именно консенсус будет давать наиболее правдоподобное распределение исследуемых структур. Отметим, что данная методика уже была проверена экспериментально в результате структурного сравнения различных семейств протеинкиназ [8]. В цитируемой работе было показано, что наилучшее распределение данных белков было получено именно с помощью консенсусного метода ProCKSI, а не с помощью отдельных методов, входящих в данный консенсус. Наконец, результаты данной работы сравнили с данными, приведенными в уже существующей структурной классификации, представленной в базе данных SCOP (Structural Classification of Proteins, http://scop.mrc-lmb. cam.ac.uk/scop/). Исходя из результатов данного эксперимента, мы решили провести собственную проверку всех методов, представленных в ProCKSI, чтобы выяснить, какой из алгоритмов наиболее подходит для решения нашей задачи. В первую очередь, необходимо было понять, имеет ли смысл применение данных методов по отношению к исследуемым нами структурам (мембранным рецепторам), поскольку все данные алгоритмы изначально были разработаны для глобулярных белков.

На первом этапе нашего исследования был проведен контрольный эксперимент, в рамках которого наряду со структурами GPCR были рассмотрены различные белки бактерий и архей (например, бактериородопсины и сенсорные родопсины архей). Выбор этих структур был сделан в первую очередь на основании структурного сходства в пространственной организации TM-домена данных белков и GPCR. В случае адекватности применяемых методов ожидалось получение классификации, в которой все структуры рецепторов GPCR должны были быть сгруппированы отдельно от структур прочих бактериальных белков. В итоге нами были получены ожидаемые результаты.

На следующем этапе необходимо было провести более тонкую подстройку используемых классификационных методов в соответствии с поставленными задачами. Важно отметить, что данные методы обладали высокой точностью по отношению к тем задачам (а именно, конкретным белкам), для которых они были изначально разработаны. Например, ранее уже была показана высокая точность метода TM-align по отношению к водорастворимым (глобулярным) белкам [12, 14]. В этих работах авторами было показано, что наиболее оптимальной мерой, определяющей пространственное сходство между исследуемыми структурами, была именно ТМscore. Аналогичные результаты были получены и для метода Vorolign [13] (исследователями была показана более высокая точность критерия Vscore по сравнению с классическим измерением RMSD). В рамках нашего исследования эмпирически было установлено, что измерения, основанные на RMSD, давали в конечном счете более правильное распределение экспериментальных структур мембранных белков. Наконец, наилучшие результаты были получены при использовании комбинации результатов, полученных методами TM-align и Vorolign.

Исследование конформационной динамики родопсинов млекопитающих методом главных компонент. Метод главных компонент (PCA) — один из основных способов уменьшения размерности исходных данных с наименьшей потерей значимой информации [15]. Расчет главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных. Наряду с этим PCA интенсивно используется в биоинформатике для сокращения размерности исследуемых структур, выделения наиболее значимой информации, а также при визуализации полученных результатов [15–18].

В случае ансамбля биологических пространственных структур данный метод представляет собой ортогональное линейное преобразование исходных данных из декартовой системы координат в новую систему (так называемых коллективных координат). Основная цель РСА заключается в упрощении исходного структурного разнообразия в исследуемом наборе данных путем определения преобладающих направлений структурных изменений. В итоге в рамках полученной системы коллективных координат основная дисперсия исходных данных располагается вдоль первых главных осей. Обычно РСА применяется к ансамблю различных конформеров исследуемого белка, определенных, например, в присутствии различных субстратов. Другой пример – анализ различных моделей в рамках исследуемой ЯМР-структуры. Наконец, в качестве исходных данных можно использовать ансамбль снимков молекулярной динамики. Существует ряд экспериментальных доказательств достоверности результатов, полученных с помощью РСА. Например, пригодность данного метода в рамках исследований функциональной динамики была недавно продемонстрирована результатами работ, проведенных на структурах убиквитина [19]. В этом случае была найдена уникальная компонента движений, ответственная за лиганд-распознающую способность данного белка.

Рассмотрим траекторию из N атомов в момент времени  $F(X_i, Y_i, Z_i)$ , где i = 1, 2, ..., N. При анализе ансамблей экспериментальных структур исследуемого белка на первом шаге осуществляется построение ковариационной матрицы C, состоящей из  $3N \times 3N$  элементов, Для расчета ковариационной матрицы C используется выражение

$$\left\langle \Delta R \Delta R^{T} \right\rangle = m^{-1} \sum_{A} \left[ \Delta R^{(A)} \Delta R^{(A)T} \right],$$
 (5)

где R — набор координат какой-либо точки траектории, суммирование осуществляется по всем *m* исследуемым структурам, а  $\Delta R$  — показатель отклонения конкретной структуры *A* от усредненного ансамбля < R >. Индекс *T* обозначает обращение исходной матрицы. Последующее разложение матрицы *C* с помощью преобразования

$$C = \sum_{i=1}^{m} \sigma_{i} p_{i} p_{i}^{T}$$
(6)

приводит к вычислению главных компонент системы (собственных векторов)  $p_i$ , а также соответствующих им значений дисперсии (собственных значений)  $\sigma_i$ . Здесь,  $\sigma_i$  — наибольшая дисперсия системы, а  $p_i$  (3*N*-мерный вектор) описывает смещение *N*-оснований, представленных в этом случае их С $\alpha$ -атомами, по данной наиболее изменчивой моде, часто называемой первой главной модой. Наконец, среднеквадратичное отклонение для исследуемых структур может быть найдено в результате вычисления следа tr матрицы *C* 

$$\langle \text{RMSD} \rangle = [\text{tr}(C)/N]^{1/2}$$
(7)

Из выражений (6) и (7) следует, что именно первая главная компонента ( $p_1$ ), которой соответствует наибольшее значение дисперсии  $\sigma_1$ , вносит основной вклад в усредненное среднеквадратичное отклонение <RMSD>.

С помощью РСА было проведено исследование конформационной динамики рецептора родопсина млекопитающих. В ходе анализа был построен ансамбль из всех известных кристаллографических структур родопсина, отражающих конформационное поведение данного рецептора. На первом шаге было проведено пространственное выравнивание исследуемых экспериментальных структур относительно структуры 1gzm [20]. Эта структура представляет собой инактивированную форму рецептора, полученную в тригональной кристальной форме с разрешением 2,85 Е. На следующем этапе был проведен расчет ковариационной матрицы, а также была получена проекция ансамбля кристаллографических структур на плоскость главных координат. На заключительном этапе была проведена визуализация полученных данных. Совокупность данных операций была выполнена с помощью программы ProDy [21] в компьютерной среде Python (http://www.csb.pitt.edu/prody/).

Визуализация полученных классификационных деревьев (дендрограмм) исследуемых структур осуществлялась с помощью пакета программ Newick Utilities [22], работающего в среде Linux, а пространственная суперпозиция отдельных структур – с помощью визуализатора РуМоl.

#### РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Классификация экспериментальных структур родопсина по версии ProCKSI. С помощью консенсусного метода ProCKSI была создана классификация всех известных экспериментальных структур рецепторов родопсина быка и кальмара. Этот выбор был обусловлен в первую очередь

БИОХИМИЯ том 77 вып. 5 2012

тем, что именно для родопсина млекопитающих к настоящему времени уже получен ряд экспериментальных структур, отражающих различные стадии процесса фотоактивации данного рецептора [23–25]. Для сравнения для β-адренергических рецепторов при методом кристаллографии уже получено сопоставимое число пространственных структур. Тем не менее в последнем случае подавляющее число таких структур представляет собой неактивную форму данного рецептора, соответствующую основному состоянию родопсина. Лишь в 2011 г. появились две новые кристаллографические структуры, соответствующие истинно активной форме β-2адренергического рецептора [26]. При этом какие-либо промежуточные формы данного рецептора, соответствующие известным интермедиатам родопсина, закристаллизовать так и не удалось. В случае всех прочих GPCR на момент проведения нашей работы существовало лишь незначительное количество экспериментальных структур данных рецепторов, представленных, как правило, только неактивными формами. Таким образом, именно совокупность экспериментальных структур родопсина быка была единственной статистическизначимой выборкой.

Общий характер полученного распределения кристаллографических структур родопсина представлен на дендрограмме (рис. 3). В результате анализа полученной классификации в первую очередь можно сделать вывод об относительной адекватности полученного распределения. На рис. 3 видно, что рецепторы обоих родопсиновых семейств группируются раздельно. Это свидетельствует о существовании структурных различий между ними [24, 27]. Данное наблюдение подтверждается рядом первоисточников, в которых были подробно описаны особенности исследуемых структур. Например, в работах [24, 28] были описаны основные структурные различия двух данных типов рецепторов. В первую очередь авторами этих работ отмечались различия в степени углубления пятой и шестой спиралей рецептора в цитоплазматическое пространство (рис. 4). Кроме того, в структуре родопсина кальмара была обнаружена дополнительная девятая спираль на С-конце, расположенная после короткой восьмой спирали, присутствующей и в родопсине млекопитающих. Наконец, существует принципиальная разница между путями сигнализации обоих рецепторов [27]. Известно, что в случае родопсина быка передача сигнала осуществляется через специфический белок трансдуцин (Gt), взимодействующий с фосфодиэстеразой. В случае же родопсина кальмара передача сигнала протекает посредством другого специфического G-белка (Gq),

действующего через фосфолипазу С. Таким образом, функциональные особенности обоих родопсиновых рецепторов прямо отражаются на их пространственных структурах. Важно, что данные тонкие структурные отличия в рамках одной и той же пространственной топологии были корректно распознаны с помощью консенсусного метода ProCKSI и отражены на дендрограммах.

При анализе результатов полученной классификации родопсина млекопитающих (рис. 3) можно отметить равномерное распределение



**Рис. 3.** Классификация экспериментальных структур двух семейств родопсина, полученная с помощью консенсусного метода ProCKSi. Темно-серые кружки соответствуют неактивным (ground, batho, lumi, metal) кристаллографическим структурам родопсина млекопитающих, а светлые — активным (Metall\* либо конститутивно активный опсин (Opsin\*)) формам данного рецептора. Черными кружками показаны две структуры родопсина кальмара, соответствующие его неактивному (ground) состоянию



**Рис. 4.** Сравнение кристаллографических структур родопсина млекопитающих (слева) и родопсина кальмара (справа) в основном (неактивном) состоянии (pdb ids 1gzm и 2z73 соответственно). Градациями серого цвета показан ход цепи, как на рис. 1. Для родопсина кальмара характерны большая степень углубления пятой и шестой спиралей рецептора в цитоплазматическое пространство, а также наличие дополнительной короткой восьмой спирали, лежащей перпендикулярно плоскости α-спирального пучка

кристаллографических структур, соответствующих различным интермедиатным формам данного рецептора. Из данного рисунка видно, что структуры, соответствующие неактивному (основному) состоянию родопсина (например, pdb ids 1f88, 1gzm), группируются отдельно от всех прочих структур данного рецептора. При более детальном рассмотрении дендрограммы можно наблюдать плавный переход из кластера неактивных структур по направлению к активным формам, соответствующим MetaII\*, либо конститутивноактивным (Opsin\*) формам рецептора (pdb ids 3dqb, 3pxo, 2x72, 3cap). Хотелось бы отметить, что одним из очевидных достоинств модели родопсина млекопитающих по сравнению с другими GPCR является то, что все известные интермедиатные формы данного рецептора достаточно детально описаны в литературе. Например, в одной из недавних работ, был показан целый ряд тонких структурных отличий, позволяющий отделить все инактивированные формы данного рецептора от его активных форм [25]. Таким образом, отслеживая динамику данных изменений, например сравнивая структуры, соответствующие промежуточным Batho- и Lumiформам, с активными MetaII\*-структурами,

БИОХИМИЯ том 77 вып. 5 2012

можно в итоге получить целостное представление о характере движения отдельных ТМ-доменов рецептора в ходе его активации.

Созданная нами классификация для родопсина млекопитающих находится в хорошем соответствии с данными, полученными также из первоисточников некоторых других экспериментальных структур этого рецептора [29, 30]. Например, структуры некоторых конститутивно активных опсинов (pdb ids 3dqb и 3cap), по мнению авторов, могли представлять собой экспериментальные модели состояния родопсина, наиболее приближенные к прошедшему полный цикл фотоактивации. В этих работах было показано, что по сравнению с инактивированной формой данного рецептора, полученной в комплексе с 11-цис-ретиналем, в апоформе данного белка (опсине) регистрировались заметные структурные изменения в участках, ответственных за связывание с G-белком. Такие изменения состояли из отклонения шестой спирали рецептора от плоскости ТМ-пучка и бокового смещения пятой спирали. Кроме этого была выявлена дестабилизация двух важных контактов: разрыв солевого мостика между третьей и шестой спиралями и разрыв ароматической связи между седьмой и восьмой спиралями рецептора. Наконец, было отмечено также возникновение двух новых стабилизирующих контактов между третьей и пятой, а также между пятой и шестой спиралями. Важно, что вышеописанная совокупность экспериментальных данных отражалась и на полученных нами дендрограммах в виде совместной группировки вышеупомянутых структур в единую подветвь с прочими активными структурами родопсина быка, соответствующими MetaII\*-состоянию (рис. 3). Наконец, отметим экспериментальные данные по сравнительно недавно закристаллизованным структурам бычьего родопсина (pdb id 30ax и 3pqr). Они соответствуют активированному MetaII\*-состоянию данного рецептора [23]. По аналогии с двумя вышеупомянутыми конститутивноактивными формами опсинов (pdb ids 2x72 и 3dqb) в данных структурах также регистрировались схожие конформационные изменения, сопровождающие фотоактивацию рецептора. Совокупность вышеизложенных экспериментальных данных также отражается на представленной дендрограмме (рис. 3).

Метод главных компонент. Проекция экспериментальных структур родопсина млекопитающих на две главные координаты. На следующем этапе мы решили проверить созданную нами классификацию родопсина млекопитающих с посощью РСА. Выбор этого метода был сделан в первую очередь исходя из литературных данных, свидетельствующих об успешном применении такой методики для исследования динамики белков [18]. В частности, в лаборатории И. Бахар ранее уже был проведен схожий эксперимент со всеми существующими на тот момент времени кристаллографическими структурами родопсина млекопитающих [16]. При этом результаты работы в целом находились в хорошем соответствии с распределением ProCKSI, полученным в рамках нашей работы. Мы решили еще раз проверить достоверность данного эксперимента, повторив его для большего числа экспериментальных структур родопсина млекопитающих и сравнив полученные результаты со структурными деревьями ProCKSI. Отметим, что мы располагали большим числом экспериментальных структур родопсина быка по сравнению с описанным в пионерской работе Бахар



**Рис. 5.** Результаты РСА-анализа 20 кристаллографических структур родопсина млекопитающих. Показана проекция ансамбля исследуемых структур на две главные компоненты. Согласно полученному распределению первая РСА-мода корректно кластеризует все неактивные (более широкое распределение выделено в рамке), а также активные (узкое распределение, в кружке) структуры MetaII\*, Opsin\* на две отдельные группы. По второй моде прослеживается дальнейшая дифференцировка исследуемых структур на уровне первого кластера. Таким образом, показана конформационная дисперсия (например, пространственные различия в организации петель)

[16]. В частности, у нас было несколько дополнительных структур активного рецептора, полученных в высоком разрешении.

Анализ полученных результатов (рис. 5) выявил хорошее соответствие полученной проекции кристаллографических структур родопсина результатам более ранней работы Бахар. В результате по первой главной компоненте была проведено правильное разбиение всех неактивных (темновых, Batho- и Lumi-форм) и активных структур родопсина (опсинов и форм MetaII\*) на две различные группы. По второй главной компоненте можно было наблюдать более тонкое распределение экспериментальных структур в рамках кластера всех инактивированных структур. Последнее распределение, в свою очередь, было наглядным представлением дисперсии данных структур, имеющих структурные различия в отдельных петлях, а также концевых участках рецептора. Замечательно, что полученное распределение по двум главным компонентам находилось в строгом соответствии со структурной классификацией рецепторов родопсина, полученной нами с помощью ProCKSI (рис. 3 и 5). Это, следовательно, могло еще раз свидетельствовать о правильности подобранного алгоритма классификации консенсусным методом ProCKSI. Так как мы оперировали с несколько большей выборкой экспериментальных структур по сравнению с таковой в работе Бахар, результаты проведенного РСА-анализа также немного различались. Например, полученное в работе Бахар относительно равномерное первоначальное распределение по двум главным координатам в нашем случае нарушалось структурой 2i37 (фотоактивированный депротонированный интермедиат), которая также группировалась между активными и неактивными формами данного рецептора на деревьях ProCKSI. В связи с этим очевидно, что использование быльших выборок экспериментальных структур дает в итоге более точные результаты РСА-анализа. Следовательно, имея более статистически значимую выборку различных кристаллографических структур изучаемого белка, можно в итоге сформировать истинное представление о конформационном поведении данного белка в ходе его фотоактивации.

Таким образом, нами была создана классификация кристаллографических структур рецепторов родопсина. В ходе анализа полученно-

го распределения удалось пролить свет на ряд структурно-функциональных особенностей данных белков. Например, анализируя совокупность структурных изменений, присутствующих в ряде интермедиатных структур рецептора родопсина млекопитающих, удалось связать различные стадии процесса фотоактивации с движениями отдельных трансмембранных доменов данных белков. В результате был разработан подход, позволяющий получать наглядные схемы, отражающие различные стадии активации мембранных рецепторов. Важно отметить, что данный подход может в дальнейшем применяться и к другим семиТМ-рецепторам, которые обладают большей терапевтической значимостью. Таким образом, совокупность полученных данных, характеризующаяся выраженным фундаментальным значением, может, например, содействовать разработке новых высокоэффективных лекарственных средств.

Отметим также, что полученное в рамках данной работы распределение экспериментальных структур GPCR находится в хорошем соответствии с первоисточниками. Важно, что различные структурно-функциональные особенности, замеченные в полученной классификации, легко можно объяснить исходя из соответствующих литературных данных. При этом наиболее четкая картина распределения была получена именно в результате применения консенсусного метода ProCSI. Наконец, полученное распределение экспериментальных структур родопсина подтвердилось результатами проведенного анализа с помощью РСА. В последнем случае была получена картина распределения указанных структур, идентичная полученной консенсусным методом ProCKSI. Это еще раз свидетельствует об адекватности использованных методов по отношению к исследуемым структурам мембранных рецепторов.

Следовательно, созданная классификация пространственных структур родопсинов позволила описать структурные изменения, происходящие при активации данных белков, а также ближе подойти и к описанию динамики наблюдаемых изменений. Более же детальное понимание динамического аспекта структурных изменений, сопровождающих активацию мембранных рецепторов, возможно либо при появлении новых экспериментальных структур данных белков, либо с помощью методов компьютерной симуляции.

#### НОВИКОВ и др.

#### СПИСОК ЛИТЕРАТУРЫ

- 1. Fredriksson, R., Lagerstrom, M.C., Lundin, L.G., and Schioth, H.B. (2003) Mol. Pharmacology, 63, 1256-1272.
- 2. Congreve, M., and Marshall, F. (2010) British J. Pharmacology, 159, 986-996.
- Hill, S.J., Williams, C., and May, L.T. (2010) British J. 3. Pharmacology, 161, 1266-1275.
- Kenakin, T. (2004) Trends Pharmacol. Sci., 25, 186-192. 4
- Kenakin, T., and Miller, L.J. (2010) Pharmacol. Rev., 62, 5. 265 - 304
- 6. Rajagopal, S., Rajagopal, K., and Lefkowitz, R.J. (2010)
- Najagopai, S., Kajagopai, K., and Eckowitz, K.J. (2010) Nature Rev. Drug Discovery, 9, 373–386. Nygaard, R., Frimurer, T.M., Holst, B., Rosenkilde, M.M., and Schwartz, T.W. (2009) Trends Pharmacol. Sci., 7. **30**, 249–259.
- Hanks, S.K., and Hunter, T. (1995) FASEB J., 9, 576-596.
- 9 Barthel, D., Hirst, J.D., Blazewicz, J., Burke, E.K., and Krasnogor, N. (2007) BMC Bioinformatics, 8, 416.
- 10. Damm, K.L., and Carlson, H.A. (2006) Biophys. J., 90, 4558 - 4573
- Maiorov, V.N., and Crippen, G.M. (1994) J. Mol. Biol., 11. 235, 625-634.
- Zhang, Y., and Skolnick, J. (2005) Nucl. Acids Res., 33, 12. 2302-2309.
- Birzele, F., Gewehr, J.E., Csaba, G., and Zimmer, R. 13. (2007) Bioinformatics, 23, e205-e211.
- Pandit, S.B., and Skolnick, J. (2008) BMC Bioinformatics, 14. 9, 531.
- 15. Ma, S., and Dai, Y. (2011) Brief. Bioinform., 12, 714-722
- Bahar, I. (2010) J. General Physiol., 135, 563-573. 16
- Maisuradze, G.G., Liwo, A., and Scheraga, H.A. (2009) J. 17. *Mol. Biol.*, **385**, 312–329. Yang, L.W., Eyal, E., Bahar, I., and Kitao, A. (2009)
- 18. Bioinformatics, 25, 606-614.

- 19. Lange, O.F., Lakomek, N.A., Fares, C., Schroder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B.L. (2008) Science, 320, 1471-1475.
- 20. Li, J., Edwards, P.C., Burghammer, M., Villa, C., and Schertler, G.F. (2004) *J. Mol. Biol.*, **343**, 1409–1438.
- Bakan, A., Meireles, L.M., and Bahar, I. (2011) 21. Bioinformatics, 27, 1575-1577.
- 22. Junier, T., and Zdobnov, E.M. (2010) Bioinformatics, 26, 1669-1670.
- Choe, H.W., Kim, Y.J., Park, J.H., Morizumi, T., Pai, E.F., Krauss, N., Hofmann, K.P., Scheerer, P., and Ernst, O.P. 23. (2011) Nature, 471, 651-655.
- 24. Kouyama, T., and Murakami, M. (2010) Photochem. Photobiol. Sci., 9, 1458-1465.
- Lodowski, D.T., Angel, T.E., and Palczewski, K. (2009) 25.
- *Photochem. Photobiol.*, **85**, 425–430. Rasmussen, S.G., Choi, H.J., Fung, J.J., Pardon, E., Casarosa, P., Chae, P.S., Devree, B.T., Rosenbaum, D.M., 26. Thian, F.S., Kobilka, T.S., Schnapp, A., Konetzki, I., Sunahara, R.K., Gellman, S.H, Pautsch, A., Steyaert, J., Weis, W.I., and Kobilka, B.K. (2011) Nature, 469, 175-180.
- Sugihara, M., Fujibuchi, W., and Suwa, M. (2011) *J. Phys. Chem. B*, **115**, 6172–6179. 27.
- Murakami, M., Kitahara, R., Gotoh, T., and Kouyama, T. 28. (2007) Acta Crystallographica. Section F, 63, 475–479.
- 29. Park, J.H., Scheerer, P., Hofmann, K.P., Choe, H.W., and Ernst, O.P. (2008) Nature, 454, 183-187.
- Scheerer, P., Park, J.H., Hildebrand, P.W., Kim, Y.J., Krauss, N., Choe, H.W., Hofmann, K.P., and Ernst, O.P. 30. (2008) Nature, 455, 497-502.

## CLASSIFICATION OF RHODOPSIN RECEPTOR **BY STATE-OF-THE-ART BIOINFORMATICS METHODS**

### G. V. Novikov<sup>1</sup>, V. S. Sivozhelezov<sup>1\*</sup>, A. S. Shebanova<sup>2</sup>, K. V. Shaitan<sup>2</sup>

<sup>1</sup>Institute of Cell Biophysics, Russian Academy of Sciences, ul. Institutskaya 3, Pushchino 142290, Moscow region, Russia; fax: (4967)330-509, E-mail: vsivo00@gmail. com, vsivo@icb.psn.ru

<sup>2</sup>Department of Biology, M. V. Lomonosov Moscow State University, Moscow 119991, Russia

> Received December 20, 2011 Revision received January 19, 2012

We report a classification of the crystallographic structures of bovine and squid rhodopsins corresponding to different stages of the receptor photocycle. Using the resource Protein (Structure) Comparison, Knowledge, Similarity, and Information server (ProCKSI), selected spatial structures were compared on the basis of classification schemes (dendrograms). To compare the spatial structures of transmembrane proteins, optimal consensus was developed from methods implemented in ProCKSI. Structures were clustered using principal component analysis (PCA), resulting in good agreement with the classification based the ProCKSI consensus method. Analysis of the results has revealed the basic movements of individual transmembrane domains of these proteins that we were able to relate to different stages of the photoactivation of rhodopsin. A combination of methods identified in this study can be used as an up-to-date analytical tool to study the conformational dynamics of membrane receptors.

Key words: receptor, rhodopsin, activation, conformation, classification