

## Разработка математических методов и программных средств выявления функциональных сайтов белков

Шутова И.В.\* , Голубович В.П.

*Институт биоорганической химии НАН Беларуси, Минск*

**Аннотация.** Разработан программный комплекс, предназначенный для анализа структуры поверхностей белков с целью выявления функциональных сайтов связывания белков. Комплекс позволяет осуществлять построение поверхностей белков, поиск структурных особенностей построенных поверхностей (карманов, каналов и полостей) и исследование найденных структурных особенностей. С помощью разработанного комплекса определен один из функциональных сайтов гранулоцитарного колониестимулирующего фактора человека и спрогнозирован его низкомолекулярный пептидный функциональный аналог.

**Ключевые слова:** функциональный сайт белка, поверхность белка, структурные особенности поверхности белка, карманы, каналы, полости.

### ВВЕДЕНИЕ

Поиск потенциальных сайтов связывания белков – актуальнейшая задача биохимии, решение которой лежит в ее новой, быстро развивающейся области – биоинформатике. Успешное предсказание участков связывания белков с их лигандами является предпосылкой для решения ряда других не менее актуальных задач биохимии, к которым можно отнести прогноз белок-белковых и белок-ДНК взаимодействий, выявление функциональных сайтов белков с известными функциями и предсказание функциональных свойств новых белков, структуры которых получены, в частности, в ходе выполнения геномных проектов. Каждая из перечисленных задач лежит в основе решения разнообразных прикладных и практически значимых подзадач, связанных, в первую очередь, с целенаправленным конструированием соединений с заданной биологической активностью. Такой рациональный подход к созданию соединений с заданными свойствами позволяет существенно снизить как временные, так и материальные затраты на создание, в частности, новых лекарственных препаратов.

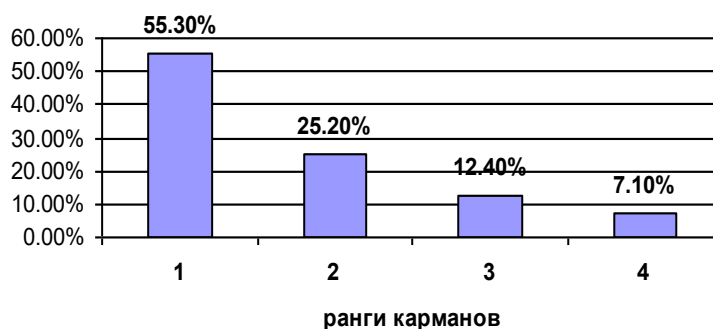
Так как функциональные свойства белков определяются их конформацией, то есть расположением полипептидной цепи в пространстве [1], то поиск сайтов связывания белков логично основывать на анализе и исследовании их пространственной структуры. Развитие физико-химических методов, позволяющих экспериментально определять пространственную организацию макромолекул, привело к появлению банков и баз с биохимическими данными и размещению в них расшифрованных пространственных структур как белков, так и их комплексов с лигандами различной природы. Накопление большого объема этих биохимических данных дало импульс к

---

\* shutova2001@mail.ru

разработке теоретических методов выявления функциональных сайтов белков, в основу которых положены анализ и исследование пространственных структур белков.

В настоящее время для теоретического предсказания функциональных сайтов белков применяются различные подходы, наиболее распространенным из которых является исследование поверхностей белков на наличие структурных особенностей разных типов – карманов, каналов и полостей. К сайтам связывания белков при таком подходе обычно относят структурные особенности с наивысшими рангами, то есть с наибольшей площадью поверхности и/или занимаемым объемом [2–5]. Действительно, согласно исследованию [6], чуть более половины (55.3%) сайтов связывания глобулярных белков размещаются в карманах первого ранга, а карманы первого и второго рангов совместно охватывают более 80% исследуемых функциональных сайтов (рис. 1). Поэтому наиболее распространенным подходом к исследованию найденных карманов является их ранжирование, то есть упорядочивание карманов по размеру. Как мы уже упоминали, чаще всего ранжирование карманов осуществляют по площади их поверхности либо занимаемому ими объему, однако в некоторых случаях применяют ранжирование карманов по убыванию количества их атомов, доступных для растворителя [7]. С содержательной точки зрения эти способы ранжирования эквивалентны.



**Рис. 1.** Распределение сайтов связывания белков по рангам карманов (данные взяты из работы [6]). Исследование проводилось на выборке из 226 лиганд-связывающих карманов. Карманы, соответствующие сайтам связывания ранжировались по объему. Карману с наибольшим объемом присваивался ранг один, следующему за ним по объему карману – ранг два и т.д.

Другими способами исследования найденных карманов являются анализ распределения электростатического потенциала, нахождение гидрофобных участков, анализ консервативности фрагментов, соответствующих найденным карманам [6].



**Рис. 2.** Поиск сайтов связывания белков с лигандами [8].

Таким образом, решение задачи по определению специфических сайтов связывания белков включает два основных этапа: во-первых, построение поверхности белка, а во-вторых – ее исследование, которое обычно подразумевает определение структурных особенностей поверхности белка, выявление гидрофобных областей и анализ распределения заряда на поверхности, а также определение консервативных областей на поверхности (рис. 2).

Таким образом, целью работы являлась разработка математических методов и программных средств определения функциональных сайтов белков на основе исследования структурных особенностей молекулярных поверхностей белков.

## КОМПЛЕКС ДЛЯ ПРЕДСКАЗАНИЯ ФУНКЦИОНАЛЬНЫХ САЙТОВ БЕЛКОВ

В данной работе мы представляем программный комплекс, предназначенный для выявления функциональных сайтов белков. Для определения потенциальных участков связывания белков с лигандами мы предлагаем осуществлять поиск карманов на поверхностях белков и исследование их гидрофобных свойств и распределения заряда на построенных поверхностях. Разработанный нами программный комплекс состоит из трех модулей, которые позволяют строить молекулярную поверхность белка (модуль 1); определять карманы на построенной поверхности (модуль 2) и осуществлять анализ распределения заряда и гидрофобных областей в карманах и на поверхности белка (модуль 3) (рис. 3).

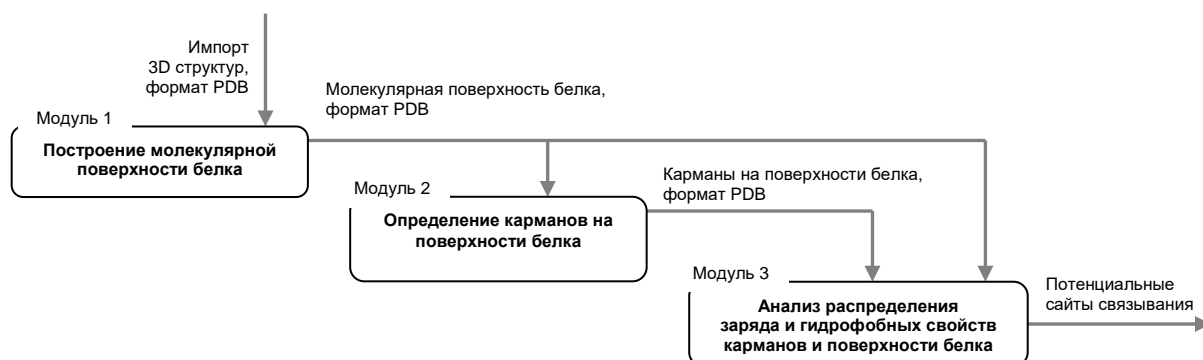


Рис. 3. Схема программного комплекса.

Комплекс был разработан и реализован нами на языке программирования C++. Исходными данными для всех трех модулей служат координаты атомов аминокислотных остатков, импортируемые из файлов в формате PDB. Выходные данные модуля 1 (поверхность белка) и модуля 2 (карманы) также сохраняются в формате PDB и могут служить входными данными для следующих модулей. Потенциальные сайты связывания белков (выходные данные модуля 3) сохраняются в виде списка формирующих их аминокислотных остатков в формате txt.

## ПОСТРОЕНИЕ ПОВЕРХНОСТЕЙ БЕЛКОВ

Первым этапом определения функциональных сайтов белков является построение их молекулярных поверхностей. Отметим, что поверхность белковых молекул имеет весьма сложную, специфическую форму, которая во многом определяет функции конкретной макромолекулы в организме. Наличие поверхности в качестве отдельного объекта во многих случаях позволяет оценить и предсказать взаимодействие белка с другими объектами, выявить аминокислотные остатки, экспонированные на поверхности белка и доступные для воды, ионов металлов и других лигандов, осуществить компьютерное конструирование сайтов связывания белков.

## 1. Типы поверхностей белков и методы их построения

Существующие в настоящее время алгоритмы построения и исследования поверхности макромолекул можно разделить на две основные группы – аналитические методы [9–12] и представление поверхности точками [13–14].

Аналитические методы описывают поверхность как совокупность простейших геометрических объектов, например, пересекающихся сфер, кубов, треугольников или плоскостей, каждый из которых можно задать в виде функции. Наибольшее распространение получили модели, в которых атомы представлены в виде сфер с радиусами, соответствующими атомным ван-дер-ваальсовым радиусам, а поверхность молекулы представляет собой пересечение этих сфер. Построенную таким образом поверхность называют ван-дер-ваальсовой. В случае если радиус сферы каждого атома увеличить на значение, равное радиусу молекулы растворителя, то получится поверхность, доступная для растворителя [15] (рис. 4-а). Молекулярная поверхность образуется при «прокатке» по поверхности белка пробной молекулы растворителя [16] (рис. 4-б).

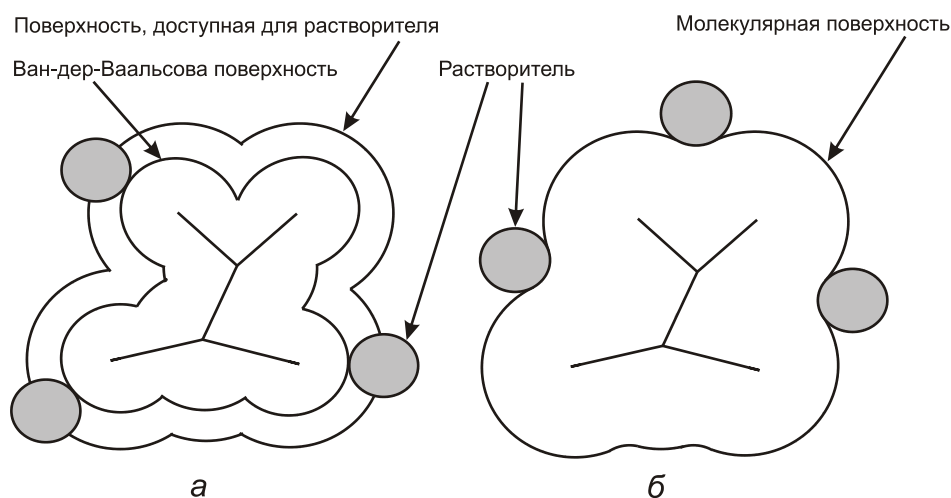


Рис. 4. Типы поверхностей молекул (адаптировано из [17]).

При аналитическом методе построения поверхности белковой молекулы для каждого атома область поверхности задается в виде функции, которая может быть проинтегрирована. Таким образом, одновременно с построением поверхности макромолекулы, решаются задачи по вычислению ее площади и нахождению объема белковой глобулы.

Метод представления поверхности точками основан на выборке через регулярные интервалы значений с исследуемых поверхностей. В качестве последних, как правило, выступают сферы вокруг атомов макромолекулы.

Сравнивая эти две группы методов построения поверхности макромолекул, необходимо отметить, что узким местом в аналитических методах является построение пересечения поверхностей атомов. В частности, если поверхность атомов представляется в виде сфер (самый распространенный подход), то для каждой пары пересекающихся сфер необходимо построить окружности их пересечения. Таким образом, при построении поверхностей белков возникает нетривиальная задача нахождения пересечения трех и более сфер, для решения которой необходимо разрабатывать специальные методы. Поэтому большинство программных продуктов, представляющих поверхность макромолекул аналитическими методами, реализуют алгоритмы *визуализации*, то есть прорисовки видимой части поверхности молекулы. Разница между алгоритмами визуализации и построения весьма существенная, так как *только при построении* мы можем говорить о поверхности как об отдельном объекте и

осуществлять ее дальнейшее исследование – например, изучать распределение на поверхности электростатического потенциала и гидрофобных участков, вычислять площадь контактной поверхности белков, определять локальные экстремумы, формирующие выпуклости и карманы на поверхности. Программы же *визуализации* поверхностей макромолекул позволяют лишь получить общее визуальное представление о видимой части поверхности, что малопригодно для дальнейшего исследования.

Методы, основанные на представлении поверхностей макромолекул в виде точек, в отличие от аналитических методов, могут быть описаны достаточно простыми математическими моделями, которые, кроме того, хорошо алгоритмируются. Причем поверхности, построенные на основе таких моделей, могут быть представлены в виде отдельного объекта, доступного для дальнейшего анализа.

Для построения поверхностей белков мы выбрали метод представления поверхности точками, так как предполагалось ее дальнейшее исследование с целью предсказания функциональных сайтов белков.

## 2. Алгоритм построения молекулярной поверхности

Введем следующие обозначения:

1. Пусть  $A$  – атом, для которого проверяем условие его вхождения в поверхность белка, тогда  $S_A$  – сфера, описанная вокруг атома  $A$  с радиусом  $R_A$ , соответствующим ван-дер-ваальсову радиусу атома  $A$ .

2. Пусть  $N_i$  – любой другой атом, кроме атома, для которого проверяем условие его вхождения в поверхность белка, тогда  $S_{N_i}$  – сфера, описанная вокруг атома  $N_i$  с радиусом  $R_{N_i}$ , соответствующим ван-дер-ваальсову радиусу атома  $N_i$ ,  $i = \overline{1, m}$ ; где  $m$  – количество атомов поверхности белка.

3. Пусть  $B$  – точка на поверхности сферы  $S_A$ , тогда  $S_{AB}$  – сфера, касающаяся сферы  $S_A$  в точке  $B$ , с радиусом  $R_S$ , соответствующим радиусу молекулы растворителя (рис. 5).

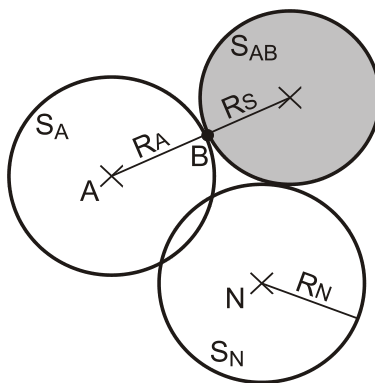


Рис. 5. Представление поверхности точками.  $A$  – атом, для которого проверяем условие его вхождения в поверхность белка.

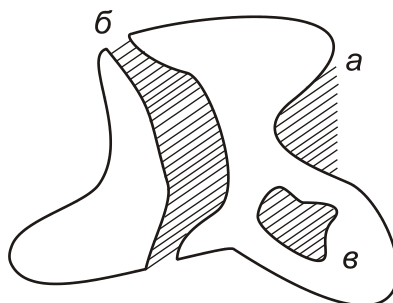
Для построения молекулярной поверхности вокруг каждого атома исследуемого белка строили сферу  $S_A$ . Поверхность каждой сферы равномерно покрывали точками. В каждой из этих точек строили сферу  $S_{AB}$ . Точку  $B$  считали внешней, если соответствующая ей сфера  $S_{AB}$  не пересекалась ни с одной сферой  $S_{N_i}$  для всех атомов  $N_i$  белка. Атом  $A$  включали в молекулярную поверхность белка, если поверхность его сферы  $S_A$  не менее чем на 20% состояла из внешних точек.

Исходными данными для построения поверхности служили координаты атомов белков, импортируемые из файлов формата PDB. Ван-дер-ваальсовы радиусы химических элементов принимали равными значениям, как в работе [18]. В качестве растворителя брали воду, и принимали радиус молекулы воды равным 1.4 Å.

Количество точек на поверхности сферы  $S_A$  изменяли от 6 до 1074. Построенную в результате работы программы поверхность представляли списком формирующих ее атомов аминокислотных остатков и сохраняли в виде файла формата PDB.

## ПОИСК КАРМАНОВ НА ПОВЕРХНОСТИ БЕЛКОВ

Поиск структурных особенностей поверхностей белков является следующим после построения поверхности этапом определения их сайтов связывания. Несмотря на то, что к настоящему времени не разработана четкая систематизация структурных особенностей поверхностей макромолекул, обычно к ним относят объекты трех типов – карманы, каналы и полости (рис. 6). Они отличаются друг от друга топологическим строением поверхностей, и, соответственно, возможными способами выявления. Различные подходы позволяют осуществлять поиск структурных особенностей как всех трех типов, так и некоторых из них. Наибольшее развитие получили методы определения карманов, так как считается, что именно в них вероятнее всего размещаются функциональные сайты белков [19]. Однако каналы и полости также играют важную роль в организации пространственной структуры белков и проявлении их функциональных свойств. Поэтому мы считаем целесообразным рассматривать три типа структурных особенностей поверхностей белков – карманы, каналы и полости.



**Рис. 6.** Топологическое строение структурных особенностей поверхностей: а – карман, б – канал, в – полость.

При исследовании структурных особенностей поверхностей макромолекул применяют два различных подхода – энергетический и геометрический. Методы, основанные на энергетическом подходе, оценивают энергию взаимодействия белка и пробной молекулы, размещенной в заданной точке и, по сути, совмещают в себе и методы докинга [20–22]. В качестве пробной молекулы, как правило, выступают функциональные группы (например, метильные, гидроксильные, амидные) и ионы металлов. Второй подход основан на геометрической интерпретации структурных особенностей поверхности белка. Одним из наиболее популярных геометрических алгоритмов поиска карманов на поверхностях белков является SurfNet [23].

### 1. Алгоритм SurfNet

Строится молекулярная точечная поверхность белка, в которой атомы белка представляются в виде сфер с радиусами, равными ван-дер-ваальсовым радиусам соответствующих элементов. Для каждой пары атомов поверхности белка строится пробная сфера между ними (рис. 7). Если она пересекается с атомами белка, то ее радиус уменьшается до тех пор, пока она не будет пересекаться ни с одним атомом белка. Если конечный радиус пробной сферы больше либо равен некоторому заранее указанному значению (обычно от 1 до 3 Å), то исходная пара атомов считается входящей в карман, канал или полость. Процесс осуществляется для всех пар атомов поверхности белка.

Алгоритм SurfNet позволяет осуществлять поиск всех типов структурных особенностей поверхности белка – карманов, каналов и полостей с заданным радиусом пробной сферы, который зависит от размера предполагаемого лиганда. Кроме того, алгоритм дает возможность осуществлять дополнительные исследования найденных структурных особенностей, такие как вычисление их площади и объема.

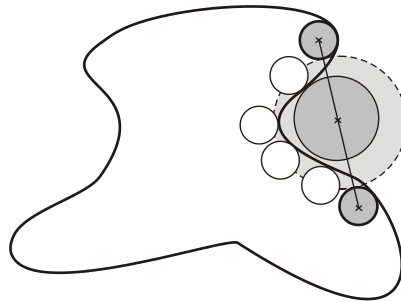


Рис. 7. Алгоритм SurfNet.

Основным недостатком алгоритма SurfNet является большое количество ложноположительных и ложноотрицательных определений карманов, то есть ошибок первого и второго рода. Ошибки первого рода в большинстве случаев возникают в случаях причисления к карманам плоских участков поверхности белка с большим радиусом кривизны (рис. 8-а). Такие структурные особенности – псевдо-карманы – часто имеют достаточно большую площадь поверхности и объем, в силу чего ошибочно распознаются как карманы первого ранга и рассматриваются как потенциальные сайты связывания белков.

Кроме того, из-за особенности алгоритма, SurfNet идентифицирует карманы только правильной формы, и не определяет структурные особенности в случаях, если поверхность кармана имеет неправильную форму с выступами внутри кармана (рис. 8-б). Как видно на рисунке, атом N образует в кармане внутренний выступ, с которым пересекается сфера  $S_{\text{проб}}$ . Поэтому атомы А и В не определяются как формирующие карман, что на самом деле не так. Если сдвинуть сферу  $S_{\text{проб}}$  вдоль отрезка АВ в направлении атома А, то становится понятно, что между атомами А и В можно вписать сферу  $S_{\text{проб}}$  с радиусом, не меньшим чем  $R_{\text{min}}$ , которая не будет пересекаться ни с одним другим атомом белка. Этот недостаток алгоритма является причиной ошибок второго рода, так как карманы на поверхностях белков, как правило, имеют неправильную форму.

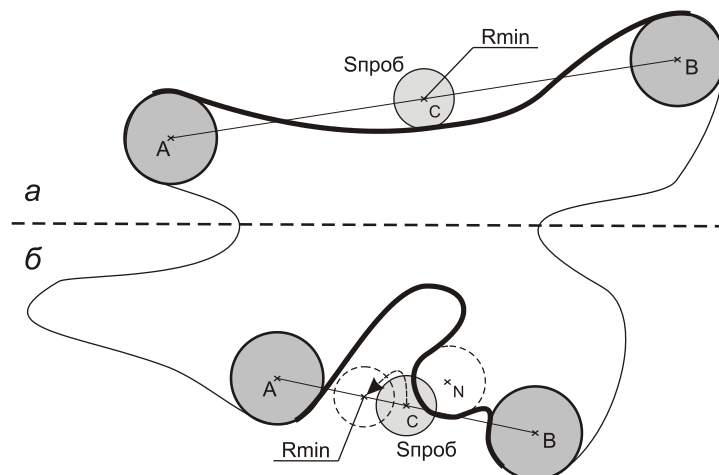


Рис. 8. Типы карманов: а – псевдо-карман; б – карман неправильной формы с выступом внутри.  $R_{\text{min}}$  – минимально допустимый радиус пробной сферы  $S_{\text{проб}}$ .



Мы выбрали алгоритм SurfNet в качестве базового, так как, несмотря на описанные недостатки, он отличается высоким качеством определения карманов правильной формы разного размера и позволяет осуществлять дальнейшее исследование найденных структурных особенностей поверхностей белков.

## 2. Модификации алгоритма SurfNet

Нами разработаны две модификации алгоритма SurfNet, в которых устранены его недостатки, описанные выше. Введем следующие обозначения:

1. Пусть  $A$  и  $B$  – атомы поверхности белка, для которых проверяем условие их вхождения в карман; тогда  $S_A$  и  $S_B$  – сферы, описанные вокруг них, с радиусами  $R_A$  и  $R_B$ , равными ван-дер-ваальсовым радиусам атомов  $A$  и  $B$  соответственно.

2. Пусть  $N_i$  – любой другой атом поверхности белка, кроме атомов  $A$  и  $B$ ; тогда  $S_{N_i}$  – сфера, описанная вокруг атома  $N_i$  с радиусом  $R_{N_i}$ , равным ван-дер-ваальсову радиусу атома  $N_i$ ,  $i = 1, m$ ; где  $m$  – количество атомов поверхности белка.

3. Пусть  $AB$  – отрезок, соединяющий атомы  $A$  и  $B$ ,  $L$  – длина этого отрезка; тогда  $R(A,B) = (L - R_A - R_B)$  – расстояние между сферами  $S_A$  и  $S_B$ .

4.  $R_{\min}$  и  $R_{\max}$  – минимальный и максимальный радиусы определяемого кармана, соответственно.

5. Пусть  $C$  – центр отрезка  $AB$ ; тогда  $S_{\text{проб}}$  – пробная сфера с центром в точке  $C$  с радиусом  $R_{\text{проб}}$ , простроенная между атомами  $A$  и  $B$  (рис. 9-а).

6. Пусть  $D$  – точка на отрезке  $AB$ , находящаяся на расстоянии  $d = (R_A + R_{\min})$  от точки  $A$ ; тогда  $S_{\min}$  – сфера с центром в точке  $D$  с радиусом  $R_{\min}$ , простроенная между атомами  $A$  и  $B$  (рис. 9-б).

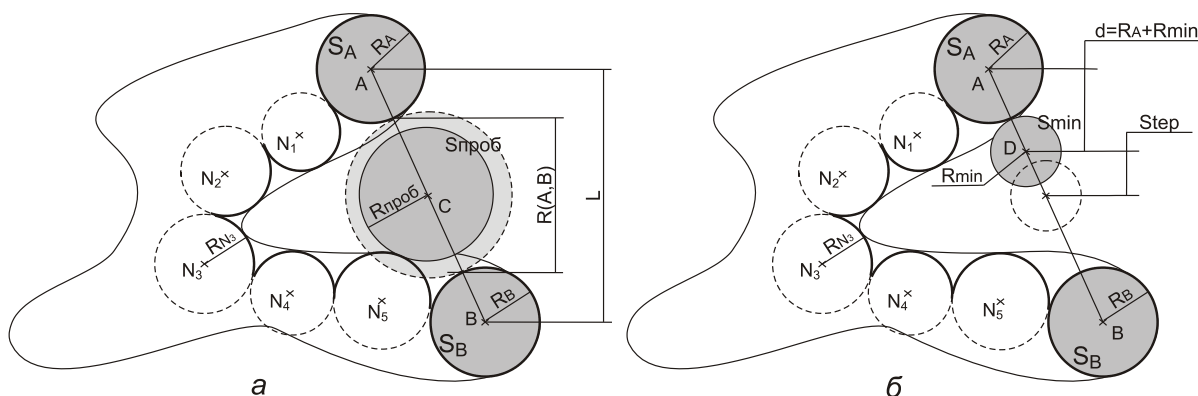


Рис. 9. Определение кармана на поверхности белка; а – алгоритм 1; б – алгоритм 2.

**Алгоритм 1.** Между каждой парой атомов  $A$  и  $B$  поверхности белка находили расстояние  $R(A,B)$ . Если выполнялись условия:

$$R(A,B) \leq R_{\max}, \quad (1)$$

$$R(A,B) \geq R_{\min}, \quad (2)$$

то между атомами  $A$  и  $B$  строили пробную сферу  $S_{\text{проб}}$  с центром на середине отрезка  $AB$  и с радиусом  $R_{\text{проб}} = (1/2)R(A,B)$ . Радиус пробной сферы  $R_{\text{проб}}$  уменьшали до тех пор, пока она не пресекалась ни с одной другой сферой  $S_{N_i}$  для всех атомов  $N_i$  поверхности белка (рис. 9-а). Атомы  $A$  и  $B$  считали входящими в карман, если выполнялось условие:

$$R_{\text{проб}} \geq R_{\min}, \quad (3)$$

где  $R_{\min}$  – минимальный радиус определяемого кармана.



**Алгоритм 2.** Между каждой парой атомов А и В поверхности белка находили расстояние  $R(A,B)$ . Если выполнялись условия (1) и (2), то между атомами А и В строили отрезок АВ, на котором в точке D располагали центр сферы  $S_{\min}$  (рис. 9-б). Радиус сферы принимали равным минимальному радиусу определяемого кармана  $R_{\min}$ .

Если сфера  $S_{\min}$  не пересекалась ни с одной другой сферой  $S_{N_i}$  для всех атомов  $N_i$  поверхности белка, то есть:

$$S_{\min} \cap S_{N_i} = \emptyset, \quad i = \overline{1, m}, \quad (4)$$

то атомы А и В считали входящими в карман, и переходили к следующей паре атомов поверхности белка.

Если условие (4) не выполнялось, то центр сферы  $S_{\min}$  сдвигали от точки D вдоль отрезка АВ в направлении атома В на величину шага  $Step$  и опять проверяли условие (4).

Итерация для атомов А и В заканчивалась в том случае, если (i) выполнялось условие (4) или (ii) сфера  $S_{\min}$  пересекалась со сферой  $S_B$ . Во втором случае считали, что атомы А и В не входят в карман и переходили к следующей паре атомов поверхности белка. Процесс осуществляли для всех пар атомов поверхности белка.

Поиск каналов и полостей по алгоритму 1 и алгоритму 2 осуществляется таким же образом, как и описанный выше поиск карманов.

Исходными данными для описанных алгоритмов служили координаты атомов поверхности белка, импортируемые из файла формата PDB. Ван-дер-ваальсовы радиусы для химических элементов принимали равными значениям, как в работе [18]. Переменные  $R_{\min}$  и  $R_{\max}$  соответствуют минимальному и максимальному радиусу определяемых карманов и обусловлены размерами потенциального лиганда. Их значения изменяли для  $R_{\min}$  от 1 до 1.6 Å, для  $R_{\max}$  – от 3 до 8 Å. Значение шага  $Step$ , на который сдвигали пробную сферу в алгоритме 2, изменяли от 1 до 3 Å. Найденные в результате работы программы структурные особенности представляли в виде списка формирующих их атомов и сохраняли в файле формата PDB. Кроме того, найденные карманы, каналы и полости сохраняли в виде простого списка с указанием формирующих их аминокислотных остатков в файле формата txt.

## ИССЛЕДОВАНИЕ СТРУКТУРНЫХ ОСОБЕННОСТЕЙ ПОВЕРХНОСТЕЙ БЕЛКОВ

Исследование найденных структурных особенностей – карманов, каналов и полостей – мы предлагаем производить двумя способами. Первый способ заключается в ранжировании найденных карманов. В качестве способа ранжирования мы выбрали упорядочивание найденных структурных особенностей по убыванию количества их атомов, доступных для растворителя. Выбор этого способа обусловлен простотой и, следовательно, эффективностью его реализации. Количество атомов, доступных для растворителя, подсчитывалось для каждого найденного кармана на этапе его обнаружения. Упорядочивание карманов по убыванию количества их атомов осуществлялось сортировкой простыми обменов. Затем карману (либо карманам) с максимальным количеством атомов присваивался ранг один. Карманам с меньшим количеством атомов последовательно присваивались ранги два, три и т.д. по мере убывания этого количества.

Вторым способом исследования является анализ распределения заряда и гидрофобных областей в карманах и на поверхностях белков. Для этого на множестве всех аминокислот мы определили три подмножества. В подмножество (AA+) включили аминокислоты, которые имеют положительно заряженные боковые группы; в подмножество (AA-) отнесли аминокислоты с отрицательно заряженными боковыми

группами; к подмножеству (ААН) отнесли аминокислоты, проявляющие гидрофобные свойства.

Определим входной список для модуля 3 как список атомов аминокислот в PDB формате, который импортируется из модуля 1 или модуля 2 (рис. 3). В первом случае из модуля 1 импортируется список атомов молекулярной поверхности белка и проводится поиск гидрофобных и заряженных областей на всей поверхности белка. Во втором случае из модуля 2 импортируются списки атомов, формирующих карманы на поверхности белка, и исследуются заряд и гидрофобные свойства карманов. Каждый список в этом случае соответствует отдельному карману.

Введем следующие обозначения:

1. Пусть  $x$  – аминокислота из входного списка, которая принадлежит одному из подмножеств (АА+), (АА–), (ААН), то есть выполняется хотя бы одно из условий:

$$x \in (\text{АА}+), \quad (5)$$

$$x \in (\text{АА}-), \quad (6)$$

$$x \in (\text{ААН}). \quad (7)$$

2. Пусть  $y$  – аминокислота из того же входного списка, что и  $x$ , которая принадлежит одному из подмножеств (АА+), (АА–), (ААН), то есть выполняется хотя бы одно из условий:

$$y \in (\text{АА}+), \quad (8)$$

$$y \in (\text{АА}-), \quad (9)$$

$$y \in (\text{ААН}). \quad (10)$$

3. Пусть  $A_x$  и  $A_y$  – атомы аминокислотных остатков  $x$  и  $y$  соответственно,  $R(A_x, A_y)$  – расстояние между атомами  $A_x$  и  $A_y$ .

Для определения заряженных и гидрофобных областей просматривали все пары атомов входного списка. Для определения положительно заряженных областей для каждого атома  $A_x$  такого, что выполнялось условие (5), находили все атомы  $A_y$  такие, что выполнялось условие (8) и

$$R(A_x, A_y) < H, \quad (11)$$

где переменная  $H$  соответствует максимальному межатомному расстоянию, на котором могут возникать нековалентные взаимодействия.

Для определения отрицательно заряженных областей для каждого атома  $A_x$  такого, что выполнялось условие (6), находили все атомы  $A_y$  такие, что выполнялись условия (9) и (11). Для определения гидрофобных областей для каждого атома  $A_x$  такого, что выполнялось условие (7), находили все атомы  $A_y$  такие, что выполнялись условия (10) и (11). Атомы  $A_x$  и  $A_y$  из одного входного списка.

Описанный алгоритм определения гидрофобных и заряженных областей реализован на языке C++. Значение переменной  $H$  изменяли от 3 до 6 Å. Определенные в результате работы программы заряженные и гидрофобные области представляли в виде списка формирующих их атомов и сохраняли в файле формата PDB, а также в виде простого списка с указанием формирующих их аминокислотных остатков в файле формата txt.

Предложенные способы исследования структурных особенностей поверхностей белков можно применять как совместно (последовательно друг за другом в любом порядке), так и отдельно друг от друга.

## ВЫЯВЛЕНИЕ ФУНКЦИОНАЛЬНЫХ САЙТОВ БЕЛКОВ

Разработанный программный комплекс был применен нами для исследования некоторых белков и выявления их функциональных сайтов. Функциональные сайты белков определяли следующим образом. Строили молекулярную поверхность исследуемого белка. Далее определяли структурные особенности построенной

поверхности – карманы, каналы и полости разного размера. Найденные структурные особенности ранжировали по количеству их атомов, доступных для растворителя. На следующем этапе анализировали распределение заряда и гидрофобных областей в карманах и на поверхности исследуемого белка. Карманы старших рангов, образующие заряженные участки или гидрофобные области, рассматривали как потенциальные сайты связывания исследуемого белка с лигандом.

С помощью разработанного комплекса нами были успешно предсказаны функциональные сайты тиропероксидазы человека [24], онокстатина М и некоторых других факторов роста [25], каспаз 7 и 9 [26]. В данной статье в качестве примера применения разработанного комплекса мы приводим выявление функционального сайта гранулоцитарного колониестимулирующего фактора человека.

## 1. Выявление функционального сайта гранулоцитарного колониестимулирующего фактора человека

Природный чГ-КСФ является кислым гликопротеидом и существует в двух формах: белок типа А состоит из 177 аминокислотных остатков, белок типа В, более активный и представляющий собой основную форму – из 174 аминокислотных остатков [27]. В банке данных ProteinDataBank (PDB; <http://www.rcsb.org/pdb/>) мы осуществили поиск пространственных структур чГ-КСФ, в результате которого было найдено две структуры – 1RHG (кристаллическая структура чГ-КСФ, тип В, получена методом рентгеноструктурного анализа, разрешение 2.2 Å) и 1GNC (кристаллическая структура чГ-КСФ, тип А, получена методом ядерно-магнитного резонанса, 10 конформационных вариантов мономера чГ-КСФ). Для дальнейшей работы мы выбрали структуру 1RHG, так как она соответствует более активной форме белка чГ-КСФ.

Далее, с помощью модуля 1 разработанного комплекса мы построили поверхность чГ-КСФ (PDB ID 1RHG). Поиск структурных особенностей поверхности чГ-КСФ осуществляли с помощью модуля 2 разработанного комплекса по алгоритму 2 со следующими параметрами:  $R_{\max} = 6\text{Å}$ ,  $R_{\min} = 1.2\text{Å}$ ,  $Step = 2\text{Å}$ . Найденные карманы были упорядочены по количеству их атомов, доступных для растворителя. С помощью модуля 3 мы осуществили исследование найденных карманов для выявления гидрофобных кластеров и заряженных областей. Мы обнаружили, что карман первого ранга, формируемый остатками Glu19, Gln20, Lys23, Asp109 и Asp112 (рис. 10), образует на поверхности белка заряженную область и потенциально может соответствовать сайту связывания чГ-КСФ с его рецептором.

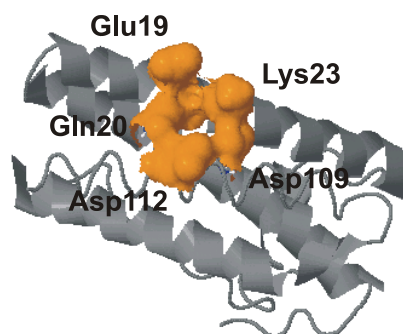
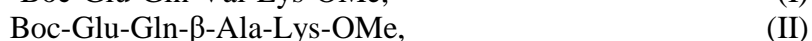


Рис. 10. Карман первого ранга чГ-КСФ (PDB ID 1RHG).

Полученные результаты были сопоставлены с литературными данными по исследованию Г-КСФ и его рецептора. Так, в работе [28] показано, что наиболее значимыми аминокислотными остатками в связывании Г-КСФ с рецептором являются Glu19, Asp109 и Asp112, что полностью согласуется с нашими результатами.

На основе полученных данных нами предложено два тетрапептида, соответствующих карману первого ранга  $\alpha$ Г-КСФ:



которые, по нашему мнению, могут моделировать биологическую функцию  $\alpha$ Г-КСФ. Предложенные соединения были синтезированы в Институте биоорганической химии НАН Беларуси. В центральной научно-исследовательской лаборатории Белорусского государственного медицинского университета была проведена *in vitro* оценка влияния спрогнозированных и синтезированных соединений на рост и дифференцировку линейных эмбриональных стволовых клеток (ЭСК). Исследование показало, что соединение (I) в концентрации 0.18 мкМ проявляет высокую активность по отношению к дифференцировке клеточных элементов в направлении эпителиальных предшественников [29].

Отметим, что полученный биологический ответ является одним из видов активности  $\alpha$ Г-КСФ. Известно, локальное применение этого белка, в частности, при поверхностной травме кожи, стимулирует процесс эпителизации, компенсаторные механизмы и обеспечивает скорейшее заживление раневого дефекта [30, 31]. Рекомбинантный  $\alpha$ Г-КСФ человека широко применяется при повреждении кожи и мягких тканей с тяжелыми морфофункциональными и косметическими последствиями [32]. Таким образом, можно сделать вывод, что с помощью разработанного комплекса мы определили один из функциональных сайтов  $\alpha$ Г-КСФ, а спрогнозированный нами тетрапептид Woc-Glu-Gln-Val-Lys-OMe является функциональным аналогом этого белка.

## ЗАКЛЮЧЕНИЕ

Нами разработан программный комплекс для определения функциональных сайтов белков на основе исследования структурных особенностей молекулярных поверхностей белков. Для построения поверхностей белков был выбран метод представления поверхности точками, который наилучшим образом подходит для их дальнейшего исследования и поиска структурных особенностей. Построенные поверхности могут быть экспортированы в формате PDB, который является фактическим стандартом представления пространственной структуры макромолекул.

Для определения карманов на поверхности белков нами разработана модификация алгоритма SurfNet и осуществлена ее программная реализация на языке C++. Введенное ограничение максимально возможного радиуса искомым карманов позволило отсеять псевдо-карманы – плоские участки поверхностей белков с большим радиусом кривизны. Разработанный алгоритм позволяет определять все карманы с заданными значениями максимально и минимально допустимых радиусов, в том числе и сложные по конфигурации карманы с выступами и с неправильной формой поверхности, не определяемые другими алгоритмами. Предложенные модификации алгоритма SurfNet уменьшают количество ошибок первого и второго рода, что оказывает существенное влияние на качество поиска структурных особенностей поверхностей белков. Также предложен способ исследования найденных карманов, который позволяет исследовать гидрофобные свойства и распределение заряда в карманах и на поверхностях белков.

Применение разработанного комплекса позволило определить один из функциональных сайтов гранулоцитарного колониестимулирующего фактора человека и спрогнозировать его низкомолекулярный пептидный функциональный аналог.

## СПИСОК ЛИТЕРАТУРЫ

1. Северин Е.С., Алейникова Т.Л., Осипов Е.В., Силаева С.А. *Биологическая химия*. МИА, 2008.
2. DesJarlais R.L., Sheridan R.P., Seibel G.L., Dixon J.S., Kuntz I.D. and Venkataraghavan R. *J. Med. Chem.* 1988. V. 31. № 4. P. 722–729.
3. Hendlich M., Rippmann F., Barnickel G. *J. Mol. Graph. Model.* 1997. V. 15. P. 359–363, 389.
4. Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R. and Ferrin T.E. *J. Mol. Biol.* 1982. V. 161. P. 269–88.
5. Laskowski R.A., Luscombe N.M., Swindells M.B. and Thornton J.M. *Protein Sci.* 1996. V. 5. P. 2438–2452.
6. Glaser F., Morris R.J., Najmanovich R.J., Laskowski R.A. and Thornton J.M. *Proteins*. 2006. V. 62. № 2. P. 479–488.
7. Yu J., Zhou Y., Tanaka I. and Yao M. *Bioinformatics*. 2010. V. 26. № 1. P. 46–52.
8. Zvelebil M.J., Baum J.O. *Understanding Bioinformatics*. Garland Science, 2008.
9. Richmond T.J. *J. Mol. Biol.* 1984. V. 178. P. 63–89.
10. Perrot G., Cheng B., Gibson K.D., Vila J., Palmer K.A., Nayeem A., Maigret B. and Scheraga H.A. *J. Comp. Chem.* 1992. V. 13. P. 1–11.
11. Gogonea V., Osawa E. *J. Mol. Struct. (THEOCHEM)*. 1994. V. 311. P. 305–324.
12. Sridharan S., Nicholls A., Sharp K.A. *J. Comp. Chem.* 1994. V. 16. № 8. P.1038.
13. Connolly M.L. *J. Appl. Crystallogr.* 1983. V. 16. P. 548–558.
14. Connolly M.L. *Science*. 1983. V. 221. P. 709–713.
15. Lee B., Richards F.M. *J. Mol. Biol.* 1971. V. 55. P. 379–400.
16. Richards F.M. *Ann. Rev. Biophys. Bioeng.* 1977. V. 6. P. 151–176.
17. Michael L. Connolly. Molecular Surfaces: A Review. URL: <http://www.netsci.org/Science/Compchem/feature14.html> (дата обращения: 05.12.2010).
18. Зефирова Ю.В., Зоркий П.М. *Журн. структур. химии*. 1974. Т. 15. № 1. С.118–122.
19. Laurie A.T.R., Jackson R.M. *Current Protein and Peptide Science*. 2006. V. 7. P. 395–406.
20. Laurie A.T., Jackson R.M. *Bioinformatics*. 2005. V. 21. P. 1908–1916.
21. Ruppert J., Welch W., Jain A.N. *Protein Sci.* 1997. V. 6. P. 524–533.
22. An J., Totrov M., Abagyan R. *Mol. Cell Proteom.* 2005. V. 4. P. 752–761.
23. Laskowski R.A. *J. Mol. Graph.* 1995. V. 13. P. 323–330, 307–308.
24. Шутова И.В., Грибовская О.В., Цыганова О.В., Голубович В.П. *Доклады НАН Беларуси*. 2009. Т. 53. № 1. С. 79–84.
25. Шутова И.В., Голубович В.П. В: *Биорегуляторы: исследование и применение. Сборник научных статей*. Под ред. академика Ф.А. Лахвича. Минск: Белорусская наука. Вып. 2. 2008. С. 186–191.
26. Martinovich V., Gribovskaya O., Shutova I., Golubovich V. *Journal of Peptide Science*. 2010. V. 16. № S2. P. 165.
27. Lu H.S., Boone T.C., Souza L.M., Lai P.H. *Arch. Biochem. Biophys.* 1989. V. 268. P. 81.
28. Young D.C. *Protein Science*. 1997. V. 6. P. 1228–1236.
29. Голубович В.П., Горанов В.А., Коваленко А.П., Мельник О.В., Шутова И.В. *Пептид, обладающий ростостимулирующей активностью гранулоцитарного колониестимулирующего фактора человека*. Патент № 2385875RU. 2010.
30. Welte K., Gabilove J., Bronchud M.H., Platzer E. and Morstyn G. *Blood*. 1996. V. 88. P. 1907.
31. Давыдовский А.Г., Климкович В.А. В: *Актуальные проблемы гематологии и трансфузиологии*. Материалы научных трудов VI съезда гематологов и трансфузиологов Республики Беларусь. Минск. 2007. С. 214–216.

32. Давыдовский А.Г., Климович В.А., Москва В.В., Дмитриев В.В. В: *Актуальные проблемы гематологии и трансфузиологии*. Материалы научных трудов VI съезда гематологов и трансфузиологов Республики Беларусь. Минск. 2007. С. 76–77.

Материал поступил в редакцию 03.12.2010, опубликован 14.12.2010.