

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ им. М.В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

РЕФЕРАТ

на тему

ПРЕДСКАЗАНИЕ БЕЛКОВ, ВЗАИМОДЕЙСТВУЮЩИХ С  
НУКЛЕИНОВЫМИ КИСЛОТАМИ

(основано на статье *Shazman S, et al, 2007*)

Реферат по структурной биоинформатике  
студентки 4 курса

Моисеевой Евгении Витальевны

Преподаватели:  
А. В. Алексеевский  
С. А. Спирин  
В. Ю. Лунин  
Е. А. Аксянов

Москва  
2014

# Оглавление

|  |          |
|--|----------|
| Аннотация.....   | 2        |
| Содержательная часть.....  | 3        |
| <i>Зачем искать самый большой участок положительного заряда на поверхности белка?.....</i> | <i>3</i> |
| <i>Аналоги.....</i>  | <i>3</i> |
| <i>PatchFinderPlus.....</i>  | <i>4</i> |
| <i>Примеры работы PFplus из статьи.....</i>  | <i>6</i> |
| <i>Другие примеры.....</i>   | <i>8</i> |
| Заключение.....  | 11       |
| Список литературы.....   | 11       |

## Аннотация

---

Представленный реферат основан на статье [1] и посвящен ее краткому обсуждению. Статья опубликована в 2007 году в журнале NAR (Nucleic Acids Research) группой израильских ученых. В данной статье авторы описывают алгоритм работы программы **PatchFinderPlus (PFplus)** и веб-сервиса, основанного на ней. Ссылка на веб-сервис <http://pfp.technion.ac.il>, который применим для определения и отображения самого большого положительного заряда на поверхности белка.

# Содержательная часть

---

## *Зачем искать самый большой участок положительного заряда на поверхности белка?*

Конечно, никто не станет спорить, что «устройство» поверхности белка в большой степени определяет характер и тип его взаимодействия с другими молекулами. Таким образом, знание о положении самого большого участка положительного заряда на поверхности белка может помочь теоретически предсказывать положение сайтов таких взаимодействий как ДНК-белок, белок-белок, белок-отрицательно\_заряженная\_мембрана, рецептор-лиганд (в т.ч. ингибитор или субстрат). Конечно, такое предсказание не будет являться непреложной истиной расположения сайта, так как существует много белков, которые не вовлечены в такие взаимодействия, однако это наверняка как-то согласуется с их функцией или локализацией в клетке.

## *Аналоги*

Для того, чтобы выявить преимущества и недостатки описанной в статье программы PFplus, стоит сказать пару слов о ее аналогах, так как очевидно, что она такая не одна уникальная. Я рассмотрю только те аналоги, которые приводят авторы статьи в своем вступлении.

Например, сервер **PreDs** [2], который предназначен для поиска участков связывания двуцепочечной ДНК на поверхности белка. Предсказание делается на основе оценки электростатического потенциала разных точек вблизи поверхности белка, а так же характера (степени кривизны и наличия углублений) этой поверхности. Сервис должен быть доступен по ссылке <http://pre-s.protein.osaka-u.ac.jp/~preds/>, однако при попытке перехода на нее браузер выдает сообщение о недоступности страницы.

Другим аналогом PFplus является программа для графической презентации и анализа поверхности макромолекул GRASP (Graphical Representation and Analysis of Surface Properties) [3] и веб-сервис, основанный на ней – GRASS. Существует очень подробное руководство пользователя, где описано все, что умеет считать данная программа [4]. В этом списке – расчет электростатического потенциала и доступности поверхности. Электростатический потенциал рассчитывается аналогично PFplus – с использованием finite-difference Poisson–Boltzmann equation (FDPB) (об этом немного позже). Однако, как пишут авторы, эта программа не предназначена для поиска изолированных участков положительного заряда на поверхности белка. Ссылка, указанная в статье - [http://honiqlab.cpmc.columbia.edu/cgi-bin/GRASS/surfserv\\_enter.cgi](http://honiqlab.cpmc.columbia.edu/cgi-bin/GRASS/surfserv_enter.cgi), тоже не работает.

Еще авторы приводят в качестве примера программу Qgrid - <http://www.netasa.org/qgrid/index.html> [5]. Этот сервер способен обнаруживать заряженные и гидрофобные кластеры в белке. Программа Qgrid рассчитывает распределение зарядов и гидрофобных участков по всему белку и при помощи алгоритма иерархической кластеризации выдает кластеры атомов остатков, разделенные по заряду или гидрофобности. Однако, этот сервис тоже больше не существует.

Все эти примеры были опубликованы раньше 2004 года, наверняка за 10 лет было разработано что-то работающее (хотя бы открывающиеся страницы), но можно потратить очень много времени на обсуждение всего, поэтому я перейду к рассмотрению собственно описанной в статье программы.

Вот так выглядит интерфейс веб-сервиса (Рисунок 1).

**Patch Finder Plus**  
A web-server for extracting and displaying electrostatic patches on a protein surface

Home | Overview | Manual | Versions | Contact | Version 2.3

**PDB Selection** ⓘ

Enter a PDB ID:

Or, upload a PDB file:

**Chain Selection** ⓘ

Calculate all protein chains (each chain is calculated separately)

Or, select a protein chain identifier:

Send a link to the results by E-mail. E-mail address:

Рисунок 1. Интерфейс веб-сервиса PatchFinderPlus <http://pfp.technion.ac.il/>.

На вход программе подается 4х-буквенный PDB-код или PDB-файл координат атомов. Обязательные строки в файле – АТОМ, такая строка должна быть для всех атомов остатка. При обработке все небелковые цепи игнорируются.

В статье сказано, что одним из выходных файлов будет PDB-файл, где указаны только выбранные остатки с заголовком из их номеров (в сегодняшней версии его уже нет), и не совсем понятно, что там за файл такой, его существование не очень логично. Кроме того, выдается список остатков, относящихся к найденному самому большому положительно заряженному участку поверхности, который визуализируется в Jmol (Рисунок 2). В статье авторы обещали сделать число продолжительных (больше какого-то размера) участков – опцией, однако, PFPplus так и считает один самый большой.

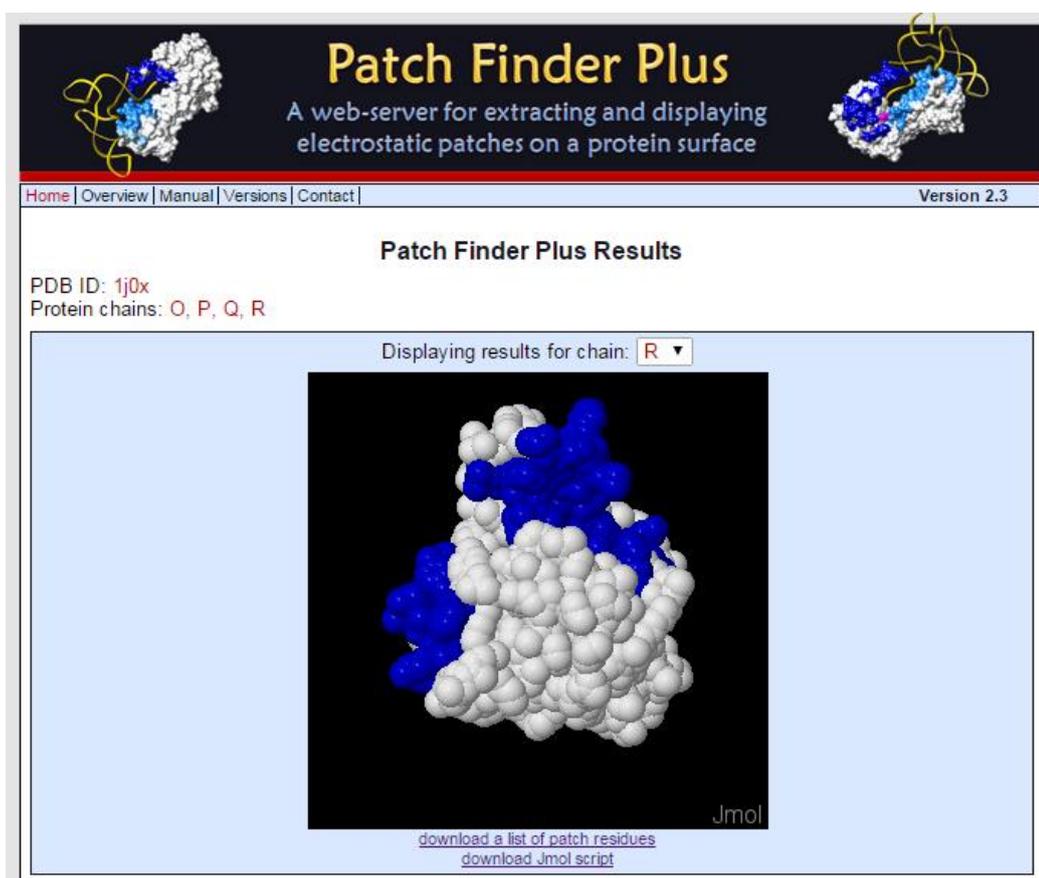


Рисунок 2. Результаты работы программы над записью PDB 1J0X (соматическая ГАФД из кролика) цепь R. Визуализация данных производится для всех белковых цепей отдельно.

#### Алгоритм работы

Ниже приведу краткое описание этого алгоритма. Он весьма незамысловат, поэтому не требует особых пояснений.

1. Расчет электростатического потенциала Ван-дер-Ваальсовой поверхности белка с использованием уравнения Пуассона-Больцмана.
  - Добавление атомов водорода - программа HBPLUS [6].
  - Решение уравнения Пуассона-Больцмана – программа от University of Houston Brownian Dynamics (UHBD) [7].

В этой версии программы все параметры расчетов зафиксированы (авторы обещают сделать их настраиваемыми, однако все еще ничего не поменялось).

Table 1. Parameters used for electrostatics calculations

| Parameter                           | Default values |
|-------------------------------------|----------------|
| Parameter set                       | OPLS           |
| Probe radii                         | 1.4 Å          |
| Stern layer                         | 2.0 Å          |
| Temperature                         | 298 K          |
| Ionic strength                      | 150 mM         |
| Dielectric constant protein/solvent | 2.0/78.0       |
| Grid dimension                      | 65 × 65 × 65   |
| Grid spacing                        | 1.5 Å          |

Наверное, стоит без лишних математических и физических подробностей пояснить, что такое уравнение Пуассона-Больцмана. Уравнение Пуассона – уравнение в частных производных, применяемое в электростатике для описания поля потенциальной энергии при известном распределении заряда. Уравнение Пуассона-Больцмана – модифицированное уравнение Пуассона, в котором учтены эффекты сольватации ионов в электростатическом поле вокруг молекулы. Оно описывает электромеханический потенциал ионов в диффузном слое.

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = -\frac{\rho_e}{\epsilon\epsilon_0}$$

2. Построение молекулярной поверхности – программа DMS: «прокатывание» шарика радиуса  $r$  по Ван-дер-Ваальсовой поверхности молекулы [8].
3. Отбор точек, посчитанных программой UNBD, которые находятся наиболее близко к поверхности.
4. Все точки фильтруются по порогу  $2kT/e$  (те, у которых потенциал больше этого, оставляются).
5. Выбор одного самого большого положительно заряженного участка поверхности.
6. Соотнесение остатков белка с этой поверхностью (сеткой), при этом, даже если один атом попал выше порога – весь остаток записывается.

#### *Преимущества*

- Сервер существует и работает.
- Производится поиск точек равномерной сетки, потенциал в которых выше заданного порога, следовательно нет никаких эвристических суждений (в отличие от метода кластеризации).
- Детекция самого большого положительно заряженного участка поверхности.
- Графическое изображение результата.
- Позволяет работать и с ЯМР структурами (берет на вход первую модель), т.е. не требуется высокое разрешение, но в остатке должны присутствовать все атомы.

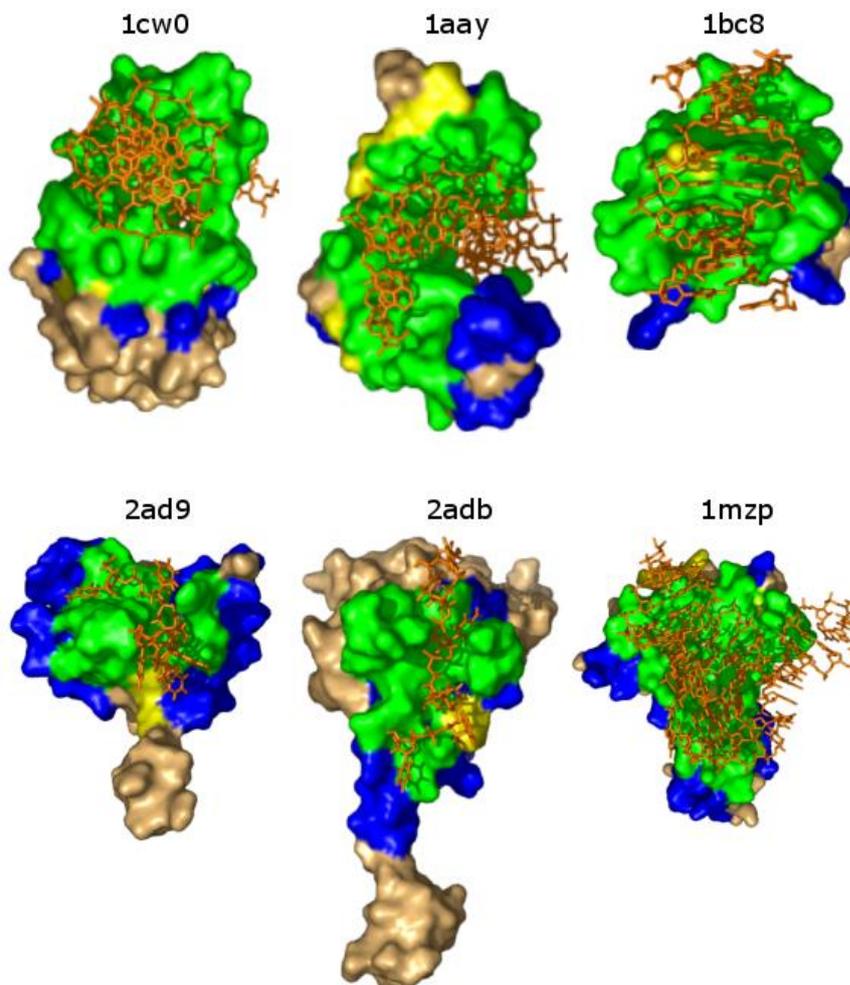
#### *Примеры работы PFplus из статьи*

На *Рисунок 3* приведены примеры работы программы на некоторых НК-связывающих белках (*Таблица 1*). Можно сказать, что такое заметное перекрытие предсказанных и реально определенных сайтов связывания ДНК положительно характеризует работу программы.

*Таблица 1. Структуры PDB, использованные в качестве примера.*

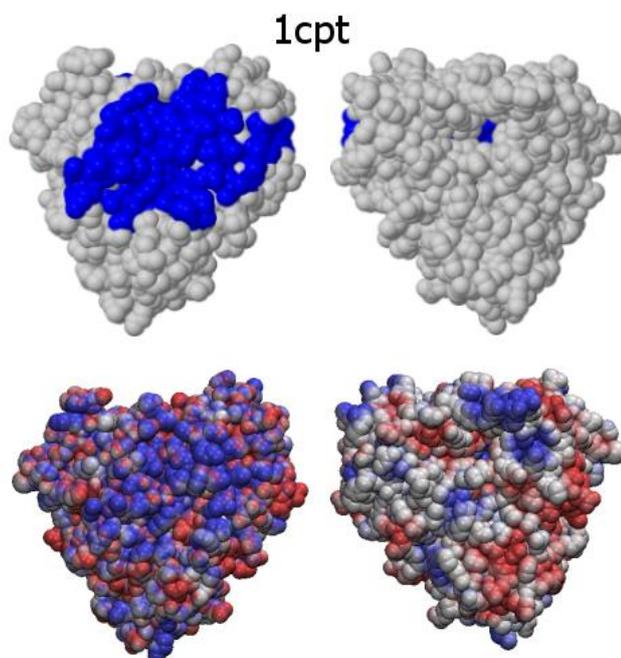
| PDB-код     | Описание   | Метод определения структуры |
|-------------|--|-----------------------------|
| <b>1cw0</b> | very short patch repair (VSR) endonuclease   | X-RAY DIFFRACTION           |
| <b>1aay</b> | ZIF268 zinc_finger-DNA complex   | X-RAY DIFFRACTION           |
| <b>1bc8</b> | SAP-1 and C-FOS promoter complex (SAP-1 is a transcription factor, which cooperates with SRF protein to activate transcription of the c- | X-RAY DIFFRACTION           |

|             |   |                   |
|-------------|---|-------------------|
|             | fos protooncogene)  |                   |
| <b>2ad9</b> | Polypyrimidine Tract Binding protein (PTB) RBD1 complexed with CUCUCU RNA   | SOLUTION NMR      |
| <b>2adb</b> | Polypyrimidine Tract Binding protein (PTB) RBD2 complexed with CUCUCU RNA   | SOLUTION NMR      |
| <b>1mzp</b> | L1 protuberance in the ribosome (L1 protuberance of the 50S ribosomal subunit is implicated in the release/disposal of deacylated tRNA from the E site) | X-RAY DIFFRACTION |



*Рисунок 3. Перекрывание (зеленый) вычисленного участка поверхности (синий) с экспериментально определенной поверхностью (желтый) взаимодействия с НК*

На *Рисунок 4* показано сравнение самого большого положительно заряженного участка поверхности белка, определенного программой PFplus, с раскраской "по заряду" из PyMol для структуры 1cрт – цитохром P450terp. Это демонстрирует, что программа правильно определяет самый большой положительно заряженный участок поверхности, хотя, конечно он может и не быть связан с НК-связывающей функцией. Однако, программа включает весь остаток целиком, даже если порог потенциала перешагивает только один (или несколько) из его атомов. Это не очень правильно с точки зрения последующего моделирования взаимодействия, то есть надо всегда помнить, что в поддержание структуры комплекса белок-лиганд участвуют только какие-то определенные атомы остатка.



*Рисунок 4. Сравнение самого большого положительно заряженного участка поверхности белка, определенного программой PFplus, с раскраской "по заряду" из PyMol для структуры 1cpt – цитохром P450terp.*

И наконец, авторы приводят пример для ТАТА-связывающего белка для демонстрации влияния параметров расчетов на результат (*Рисунок 5*). Очевидно, что чем выше порог на потенциал, тем меньше будет финальный участок, и чем больше шаг сетки тем он будет больше. Параметры по умолчанию для PFplus сервера – 1.5 Å и 2кТ/е. Выбор такого порога на потенциал обусловлен ранее полученными данными о комплексах белок-ДНК.

### *Другие примеры*

Я решила тоже проверить работу программы на других белках: соматическая ГАФД кролика (1J0X, гомотетрамер) (*Рисунок 6*) и ревертаза вируса HIV-1 (1HYS, комплекс с полипуриновым трактом) (*Рисунок 7*). Дегидрогеназа не ДНК-связывающий белок, однако нашелся большой положительно заряженный участок (думаю, он найдется в любом белке), который соответствует сайту связывания глицеральдегид-3-фосфата, который является субстратом ГАФД и имеет отрицательно заряженную фосфатную группу. Для ревертазы же, которая является гетеродимером для одной цепи (А) достаточно хорошо была определена поверхность, вероятно принимающая участие в связывании нуклеиновой кислоты. В данной структуре белок закристаллизован помимо полипуринового тракта с антителом, которое как раз и связывается с цепью В (на рисунке антитела нет, зато есть дырка в поверхности – место его связывания). Цепь В не обладает функцией связывания НК, однако положительный участок тоже есть, как раз с ним связывается антитело.

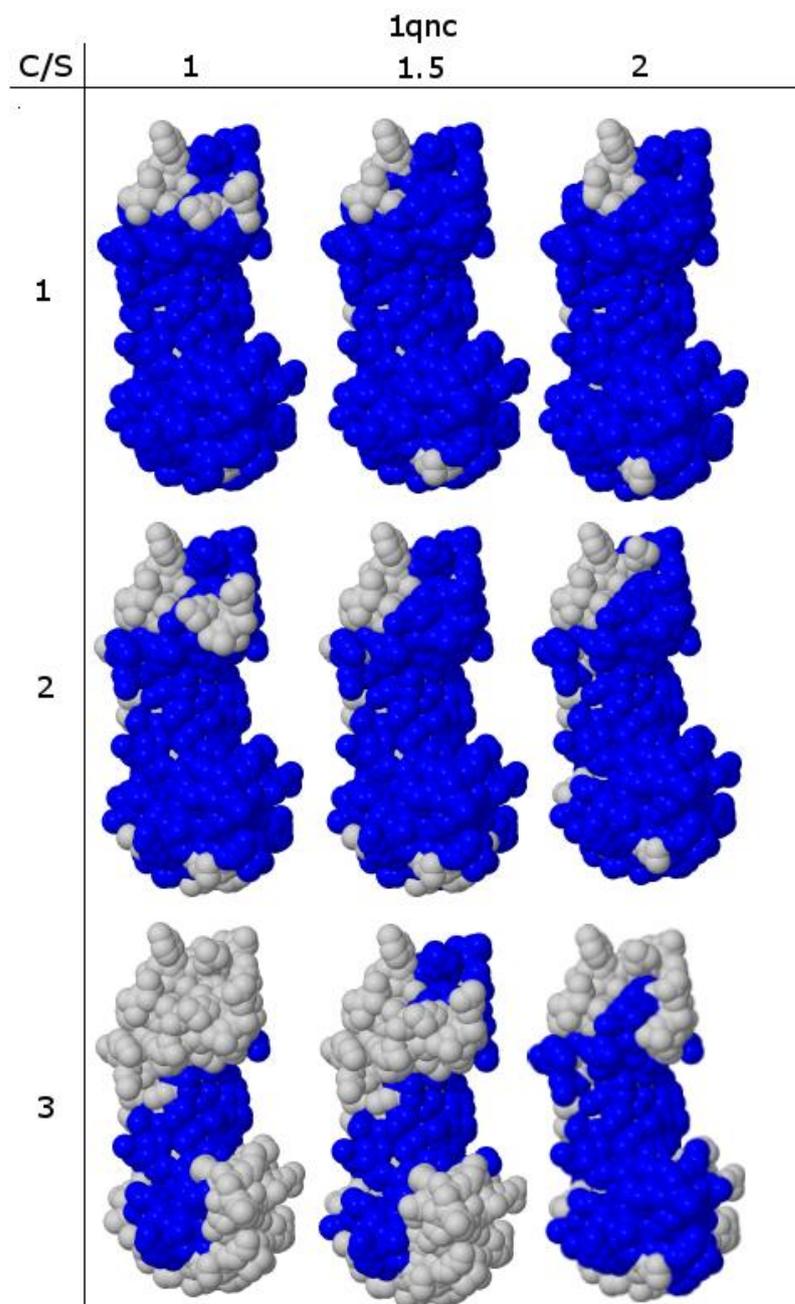
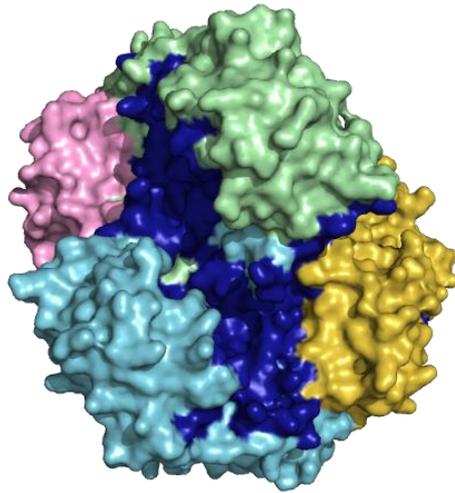
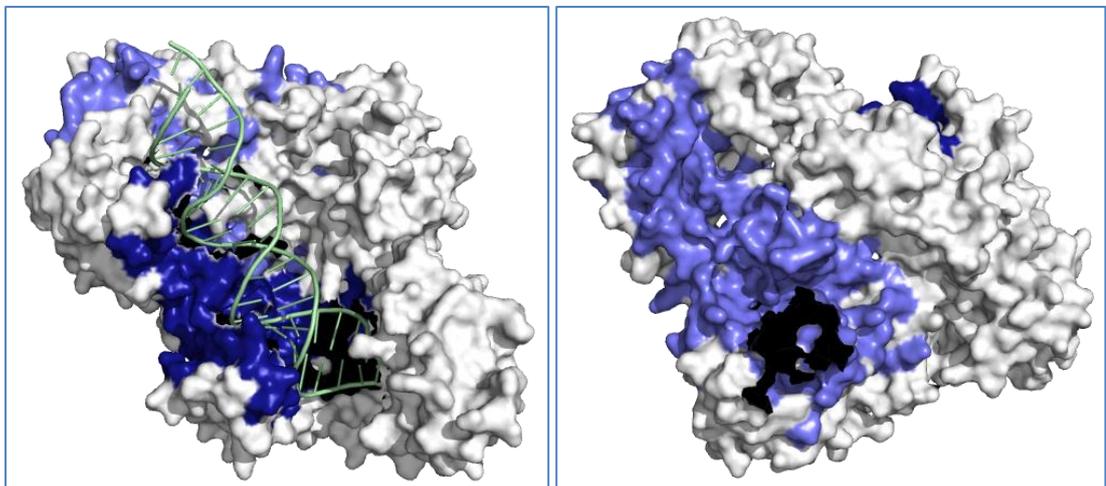


Рисунок 5. ТАТА-связывающий белок 1qnc. Колонки – разный шаг сетки (для решения электростатических уравнений) - 1, 1.5, 2 . Строки – разные пороги на потенциал – от 1 до 3 \*kT/e).



*Рисунок 6. Структура соматической ГАФД кролика 1J0X.*



*Рисунок 7. Структура ревертазы вируса HIV-1 в комплексе с полипуриновым трактом.*

# Заключение

---

В заключение можно сказать, что программа работает достаточно верно, о чем свидетельствуют примеры для ДНК-связывающих белков. То есть теоретически использовать ее для предсказания участков связывания можно. Однако, стоит относиться с вниманием что не все атомы такого участка в действительности заряжены (имеют вблизи своей поверхности потенциал выше определенного порога) и не могут принимать участия в стабилизирующих комплекс взаимодействиях. Программа работает достаточно быстро, что тоже является несомненным плюсом.

Данная статья была процитирована за 7 лет всего 16 раз, что отражает невысокую применимость данного сервиса.

# Список литературы

---

1. Shazman S, Celniker G, Haber O, Glaser F, Mandel-Gutfreund Y. Patch Finder Plus (PFplus): A web server for extracting and displaying positive electrostatic patches on protein surfaces. *Nucleic Acids Research* 2007;35 (Web Server issue):W526-W530. doi:10.1093/nar/gkm401.
2. Yuko Tsuchiya, Kengo Kinoshita, and Haruki Nakamura. PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics* (2005) 21 (8): 1721-1723.
3. Nayal M, Hitz BC, Honig B. GRASS: a server for the graphical representation and analysis of structures. *Protein Science : A Publication of the Protein Society* 1999;8(3):676-679.
4. <http://www.msg.ucsf.edu/local/programs/grasp/html/>
5. Ahmad,S. and Sarai,A. (2004) Qgrid: clustering tool for detecting charged and hydrophobic regions in proteins. *Nucleic Acids Res.*, 32, W104–W107.
6. McDonald,I.K. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, 238, 777–793.
7. Davis,M.E, Mudura,J.D., Luty,B.A. and McCammon,J.A. (1991) Electrostatics and diffusion of molecules in solution. *Comp. Phys.Commun.*, 62, 187–197.
8. Richards,F.M. (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.*, 82, 1–14.