

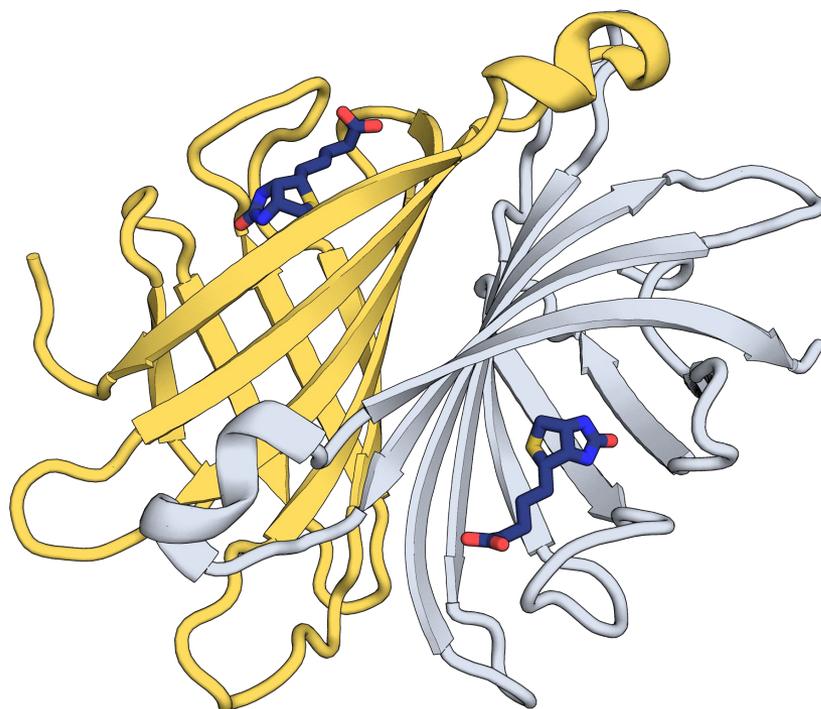
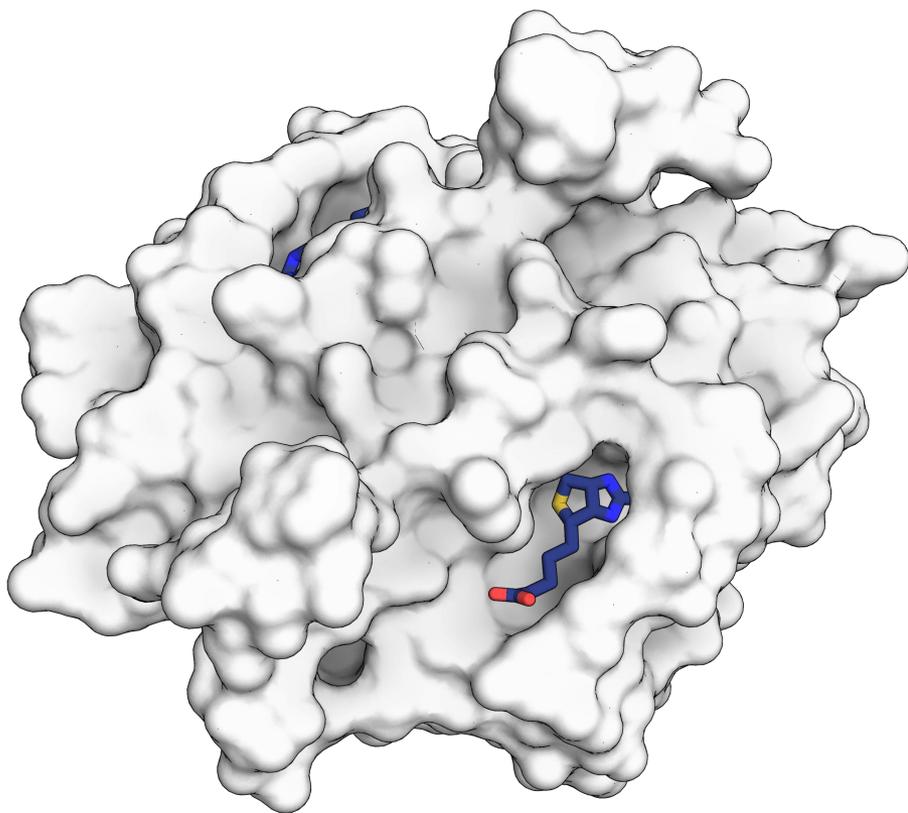
Факультет биоинженерии и биоинформатики,
Московский государственный университет имени М. В. Ломоносова



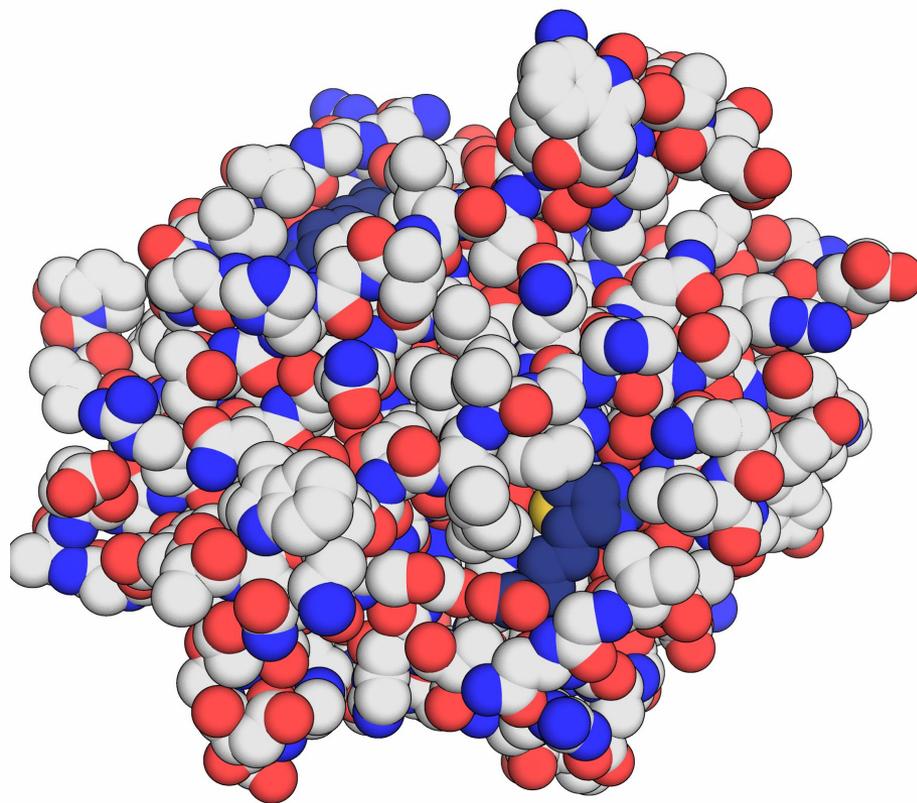
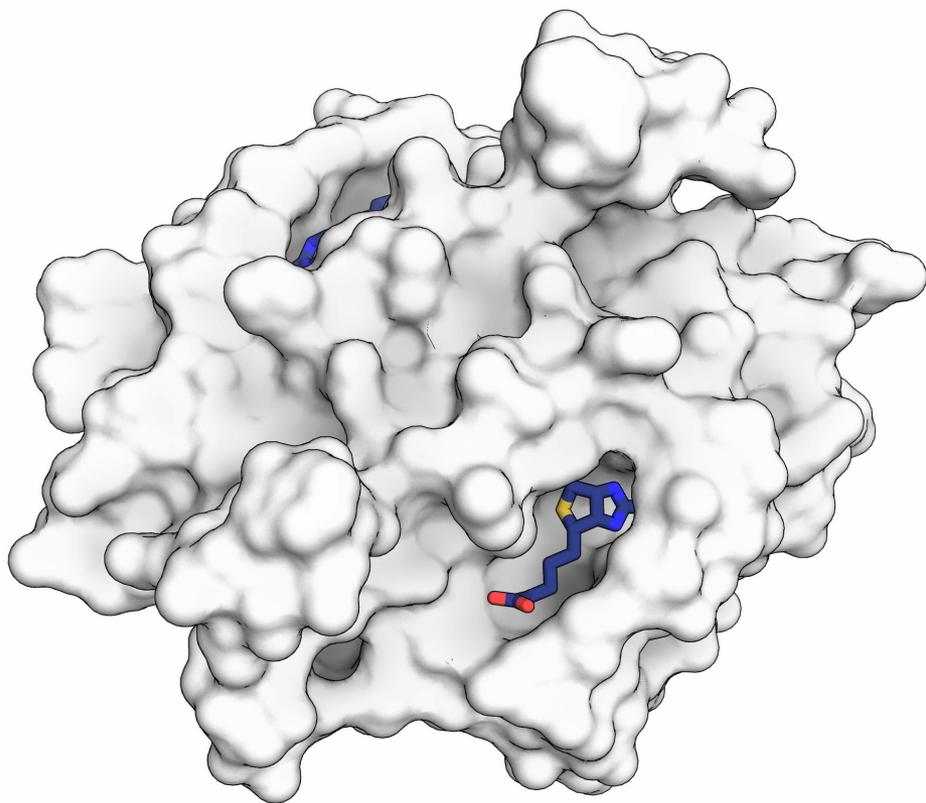
Поверхности макромолекул. Полости и каналы, карманы связывания. рКа

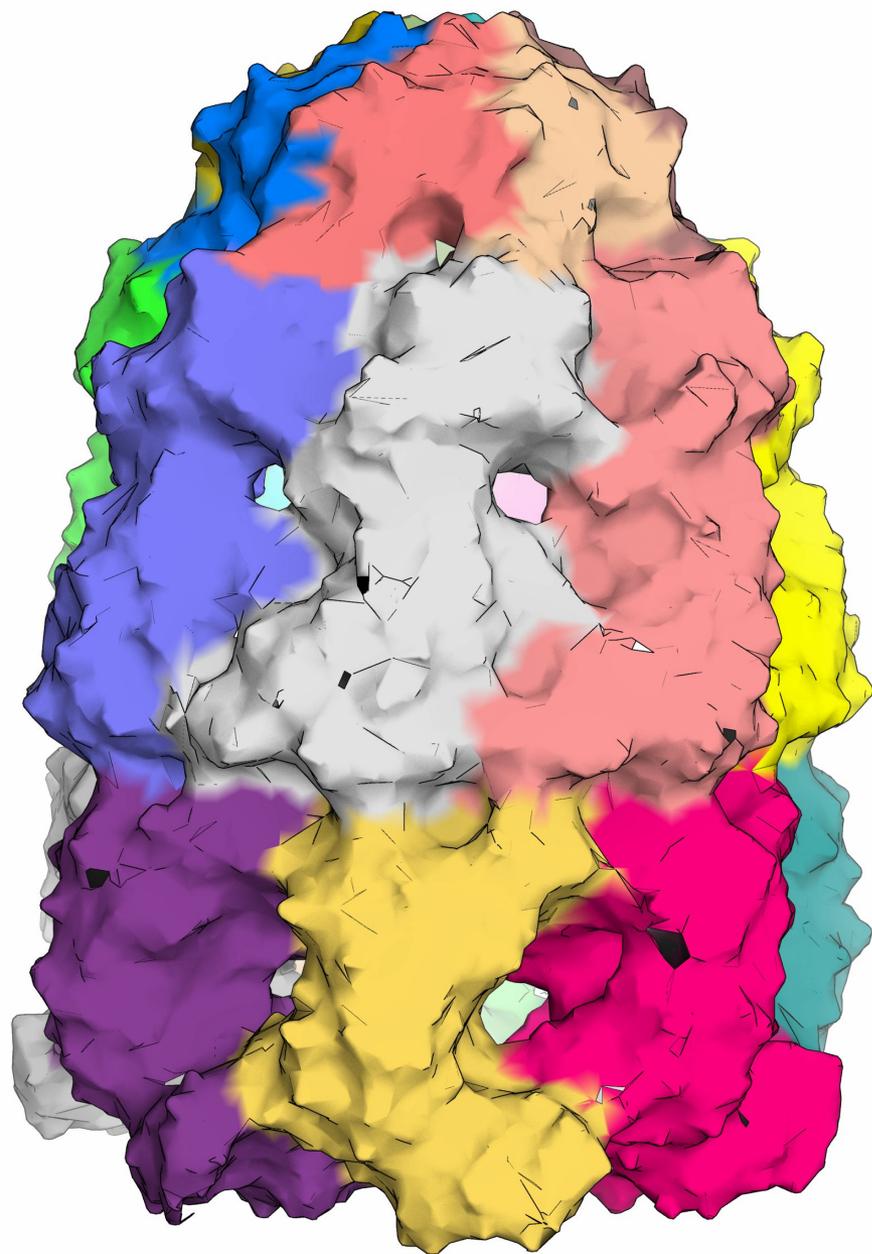
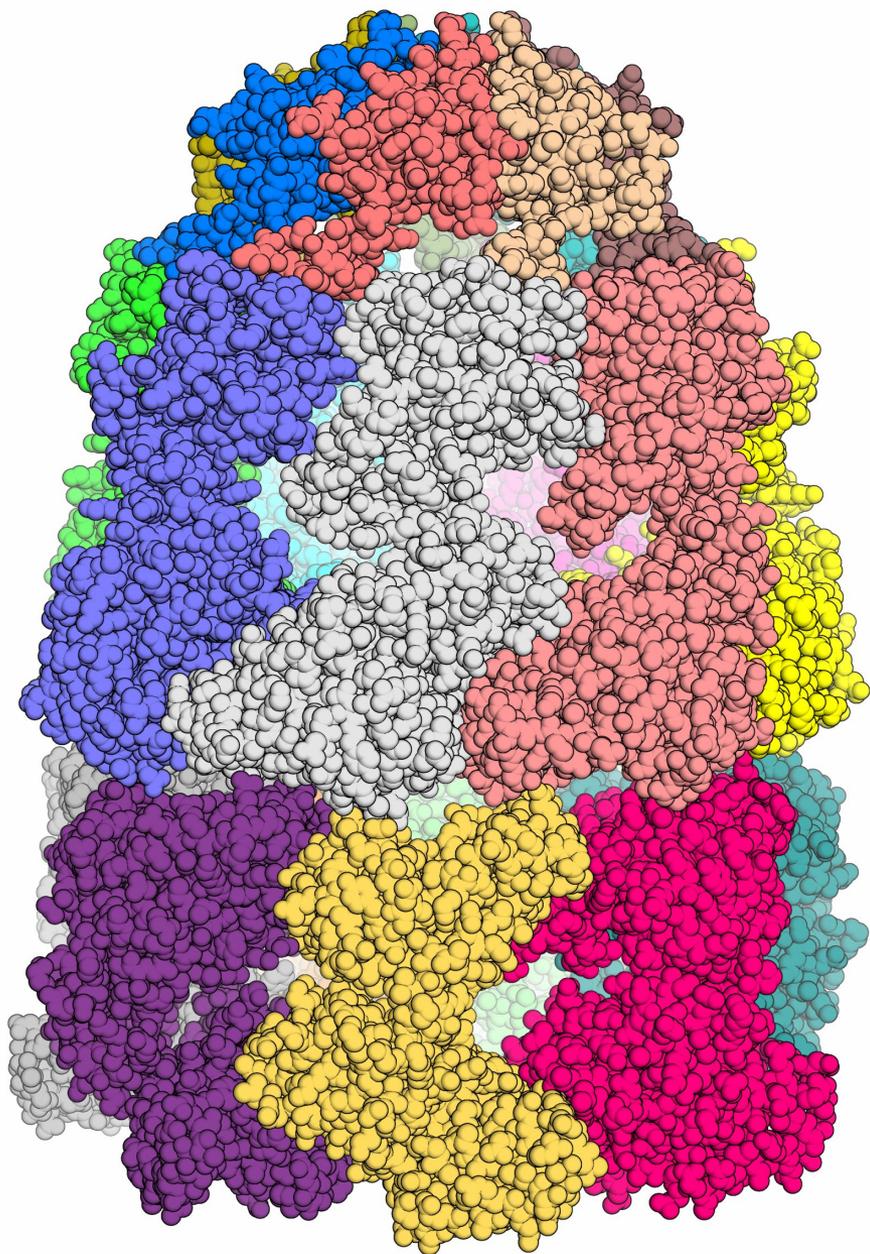
Лекция 3, биоинформатика, 4 курс ФББ МГУ, осенний семестр
Злобин А. С., alexander.zlobin@fbb.msu.ru

Поверхность – что это?



Поверхность – что это?

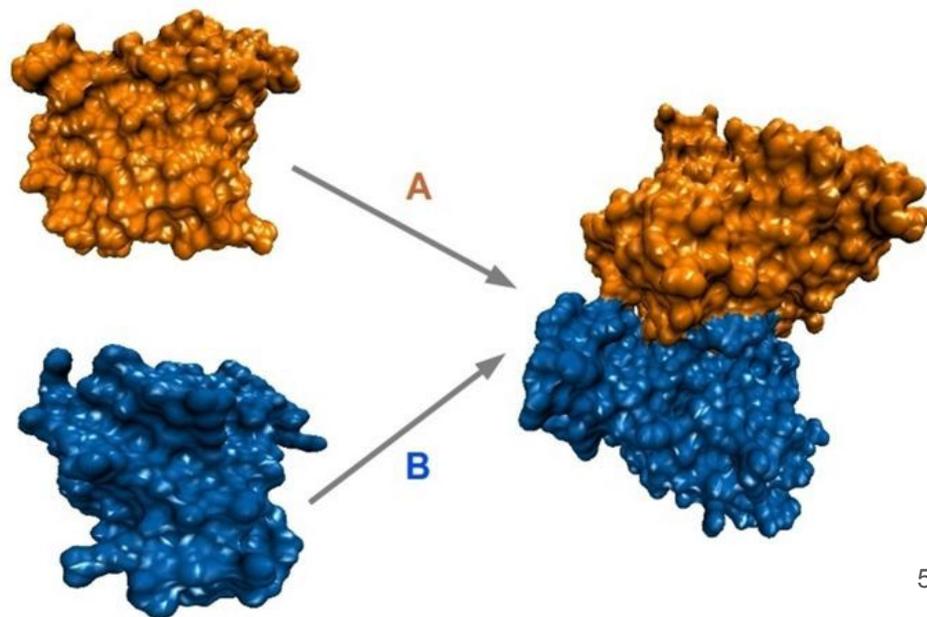




Зачем нужна поверхность как отдельный объект?

Вычисление площади поверхности

- Важнейший параметр в фолдинге
Livingstone, Spolar, Record, **Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area**, *Biochemistry*, 1991
- Shape complementarity – важный концепт в изучении связывания. Аффинность пропорциональна площади поверхности контакта (если точнее, то компонент энергии взаимодействия, обусловленный гидрофобным эффектом)



MMGB(PB)SA

Один из способов оценки энергии связывания из статичных моделей.

$$\Delta G_{\text{bind,aq}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{bind,solv}} - T\Delta S, \quad (1)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{covalent}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdW}}, \quad (2)$$

$$\Delta E_{\text{covalent}} = \Delta E_{\text{bond}} + \Delta E_{\text{angle}} + \Delta E_{\text{torsion}}, \quad (3)$$

$$\Delta G_{\text{bind,solv}} = \Delta G_{\text{polar}} + \Delta G_{\text{non-polar}}, \quad (4)$$

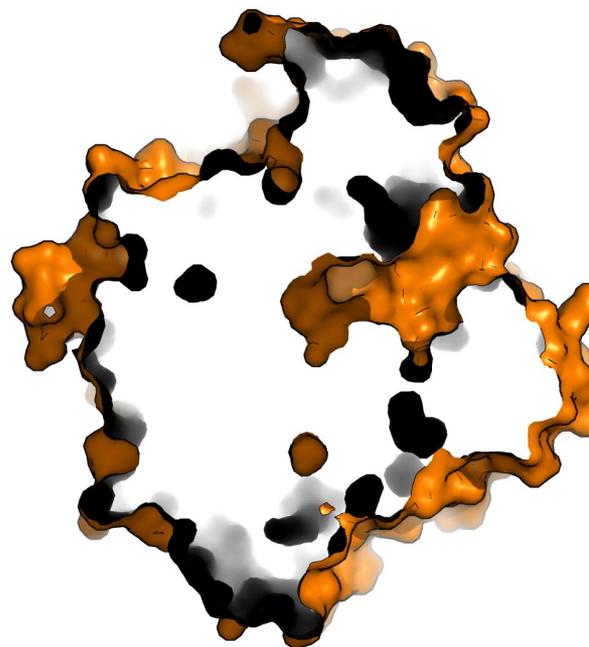
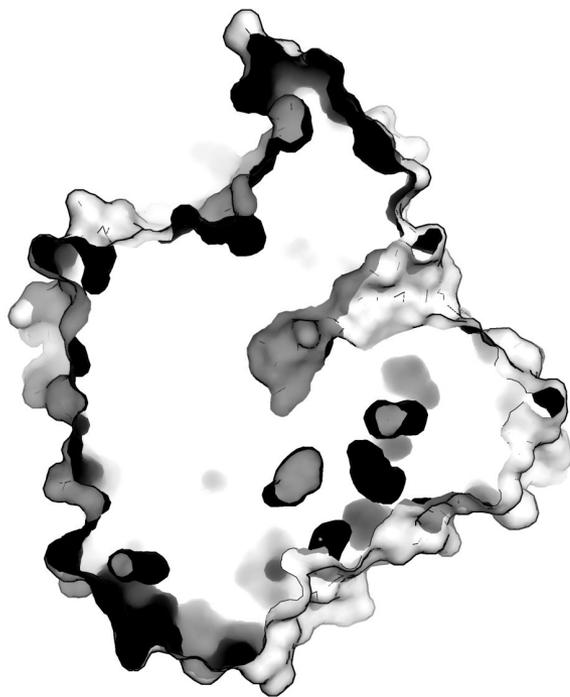
$$\Delta G_{\text{non-polar}} = \gamma^* \text{SASA} + b.$$

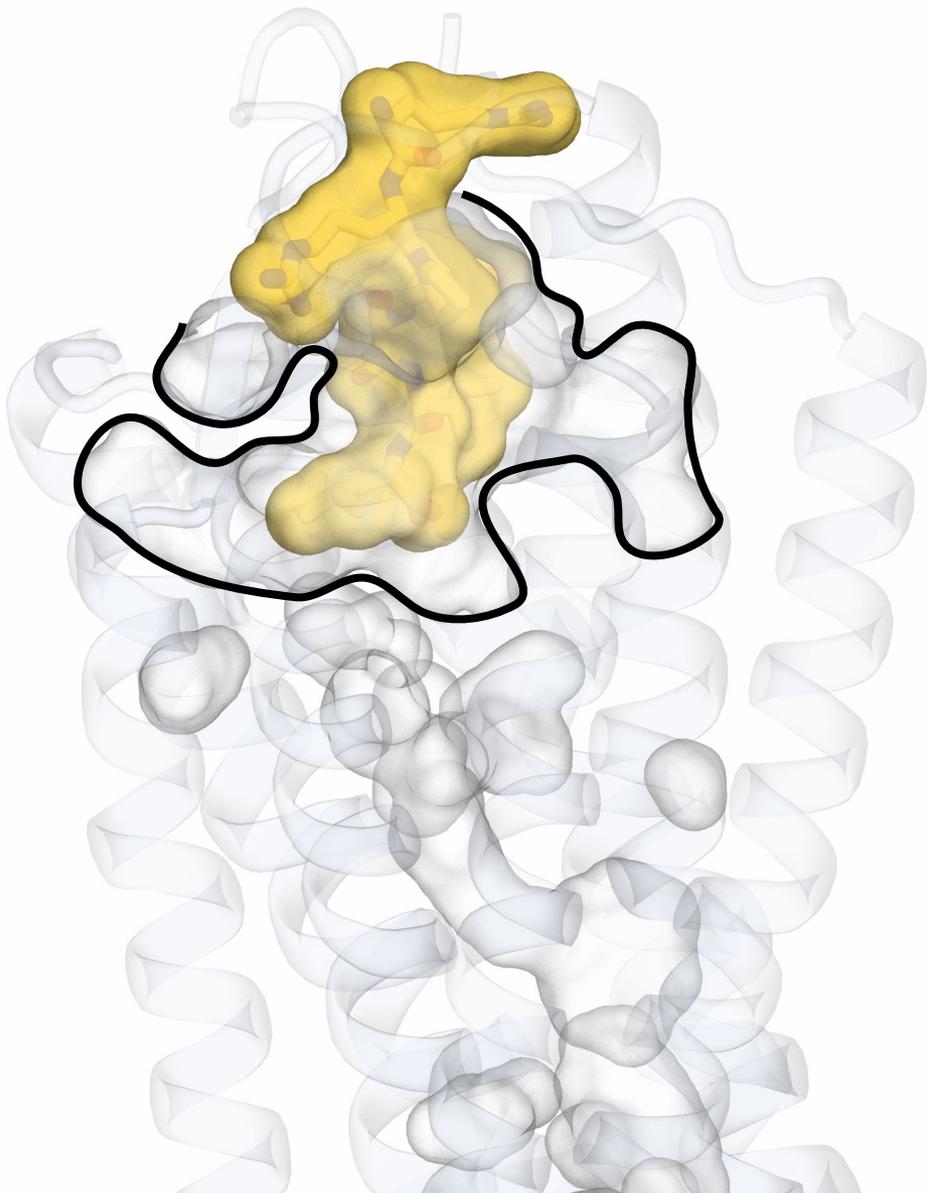
$$\begin{aligned} \Delta G_{\text{non-polar}} &= \Delta G_{\text{dispersion}} + \Delta G_{\text{cavity}} \\ &= \Delta G_{\text{dispersion}} + \gamma^* \text{SAV} + b. \end{aligned}$$

Зачем нужна поверхность как отдельный объект?

Визуальное определение полостей и каналов

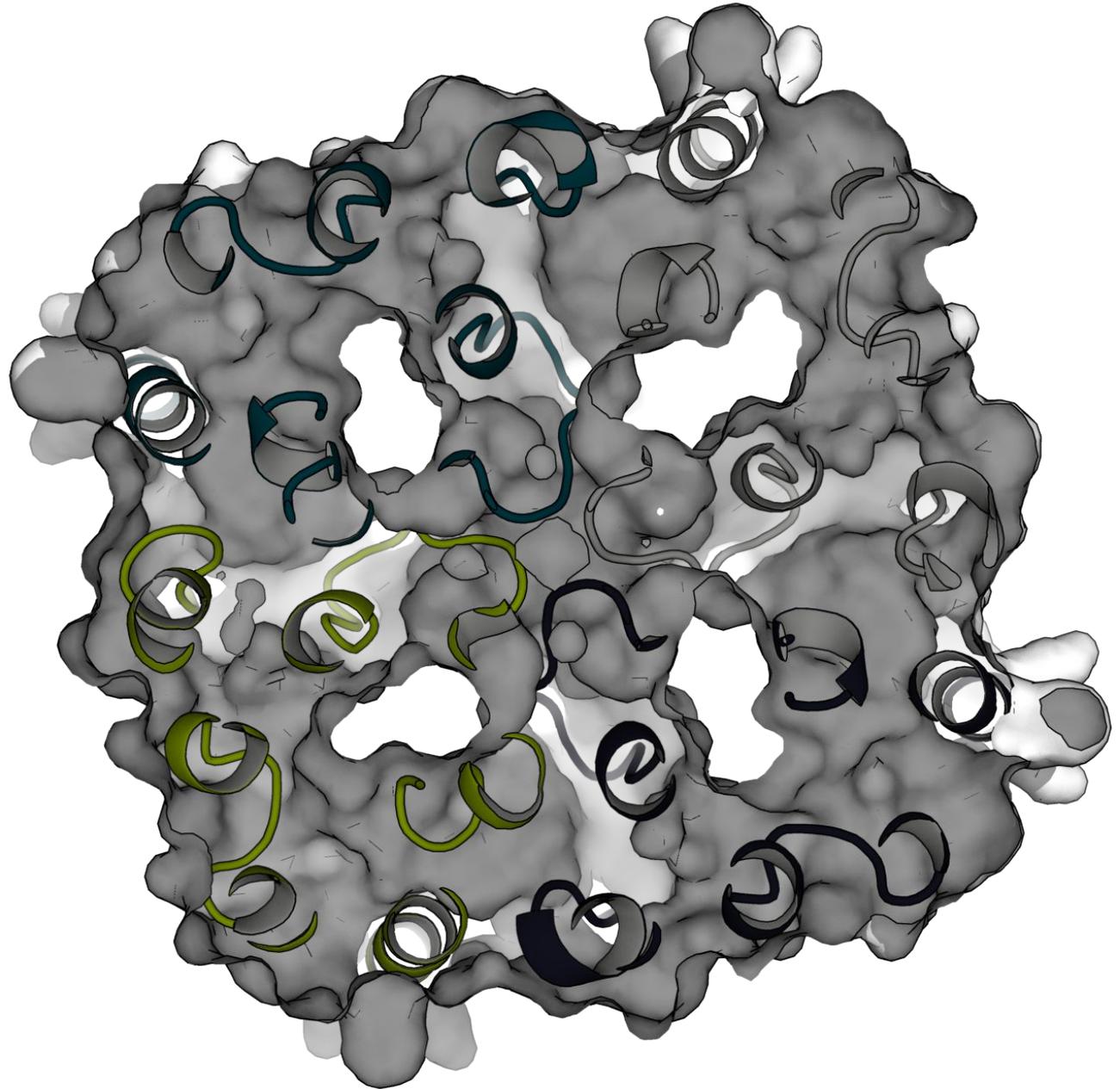
- В полости может связываться лиганд
- По каналу он может транспортироваться





Аквапорин

Переносит через мембрану воду



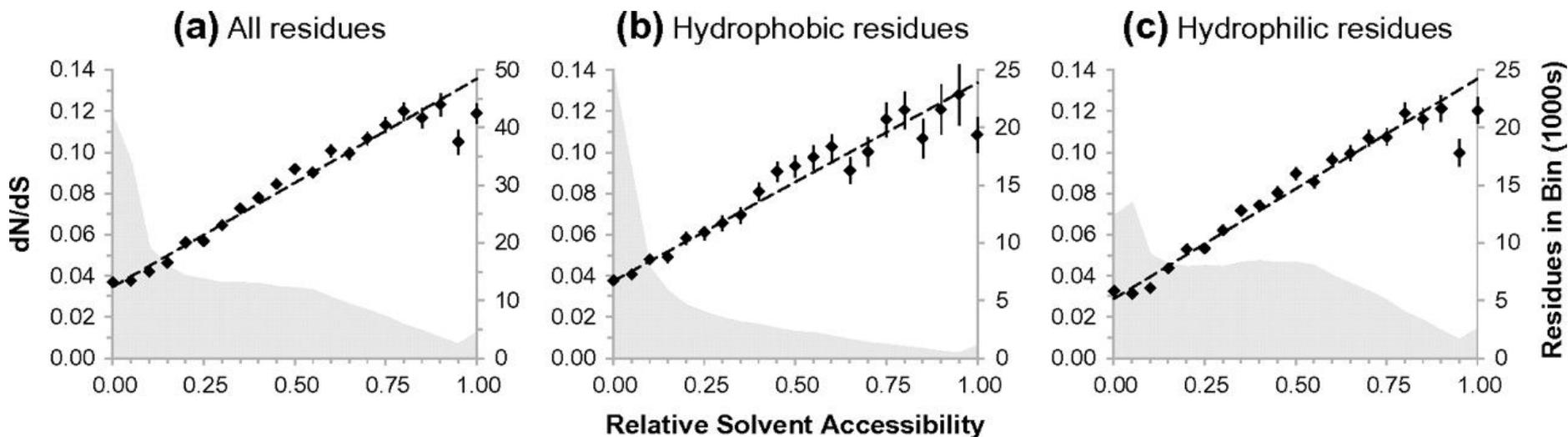
Зачем нужна поверхность как отдельный объект?

Расположен ли остаток на поверхности?

(и, следовательно, является ли он доступным растворителю)

- Одинаково ли эволюционирует поверхность и все остальное?
- Одинаково ли эволюционирует внутренняя и внешняя поверхности?
- Есть ли связь вторичной структуры остатка и его экспонированности?
- Что может значить аномальная экспонированность остатка?
- Сколько полярных остатков на поверхности – связь с растворимостью

Сила отбора и расположение на поверхности



Чем более погружен остаток, тем с меньшей вероятностью в соответствующем кодоне возникает несинонимическая замена.

Eric A. Franzosa, Yu Xia, **Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level**, *Molecular Biology and Evolution*, Volume 26, Issue 10, October 2009

Что может значить аномальная экспонированность остатка?

- В среднем Glu, Asp хорошо экспонированы
Glu, Asp без контактов с растворителем – часть каталитического сайта
- В среднем Phe, Trp, Ala, Leu, Ile, Val плохо экспонированы.
Если видим их на поверхности, вероятно, наш белок в реальности в этом месте контактирует с другим белком

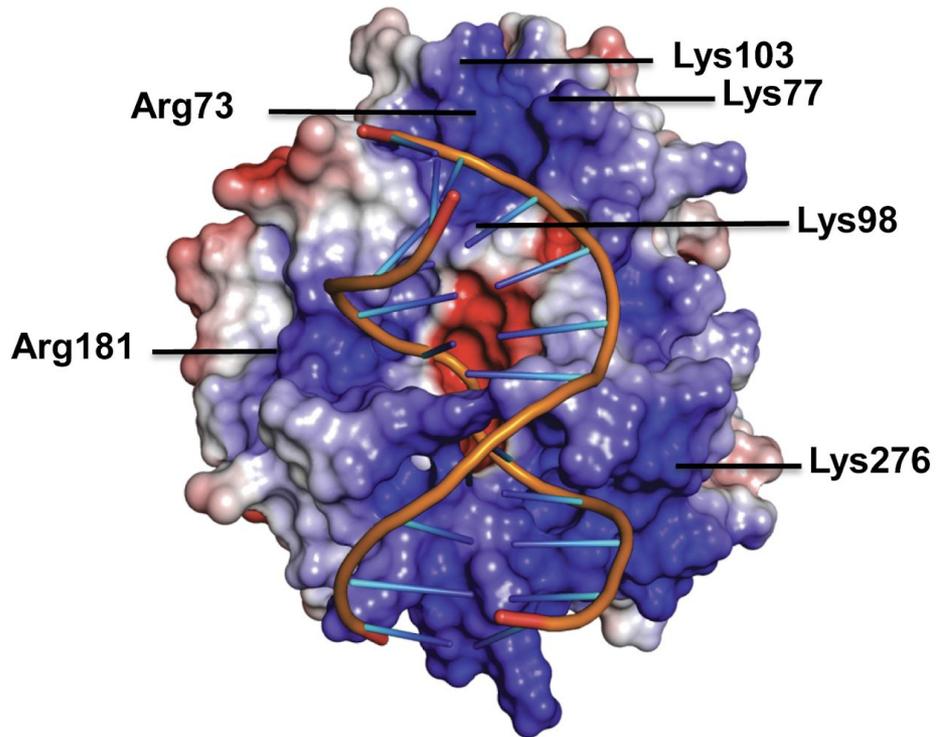
Зачем нужна поверхность как отдельный объект?

Можем нанести физико-химические свойства на поверхность и визуализировать, что помогает делать выводы быстро

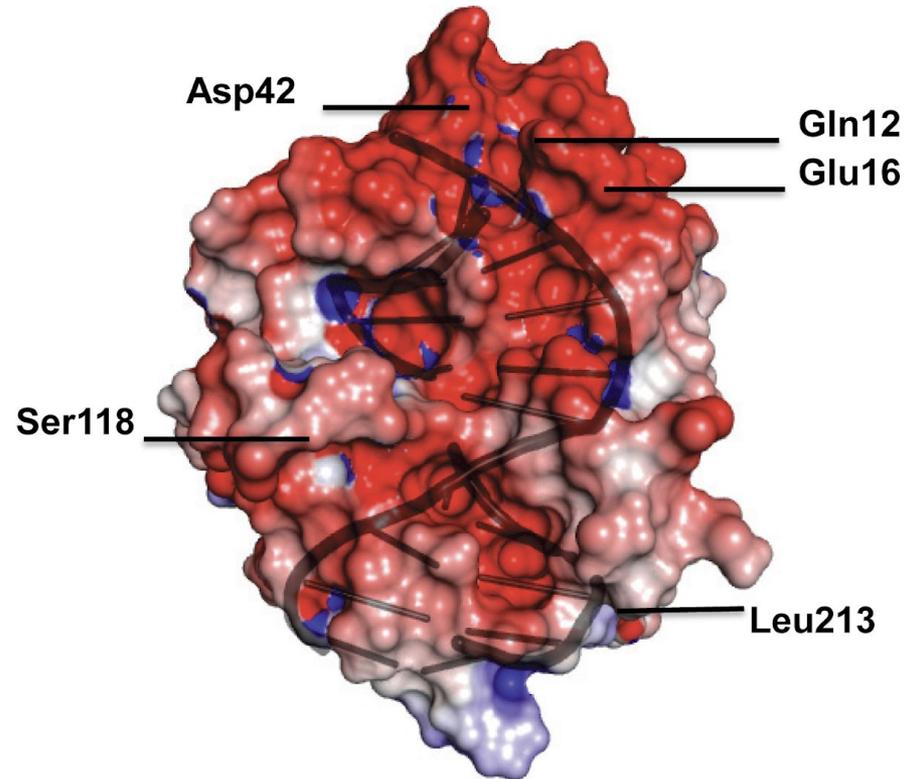
- Визуализация электростатического потенциала
- Визуализация гидрофильных/гидрофобных областей

Поверхностный электростатический потенциал

A



B



Tetyana Milojevic, Irina Grishkovskaya, Elisabeth Sonnleitner, Kristina Djinovic-Carugo, Udo Bläsi, **The *Pseudomonas aeruginosa* Catabolite Repression Control Protein Crc Is Devoid of RNA Binding Activity**, *Plos One*, 2013

Поверхностный электростатический потенциал

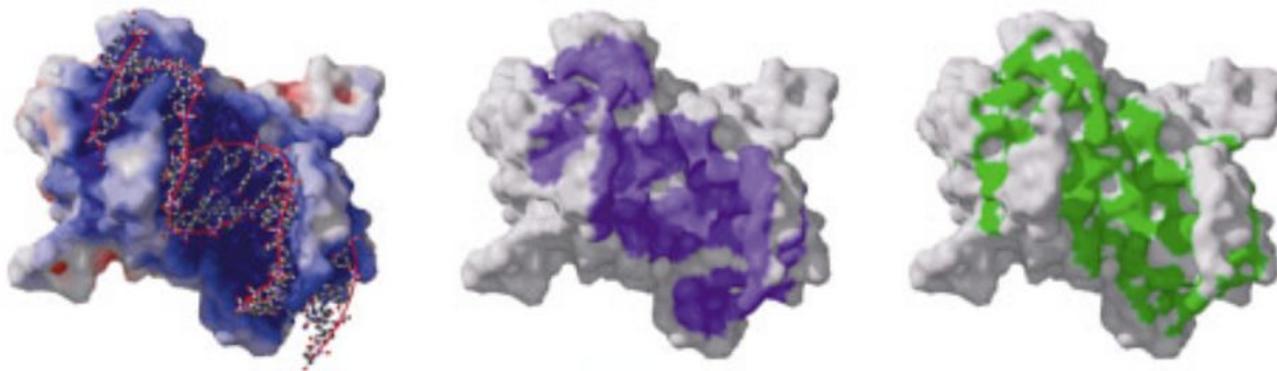
Может быть использован для предсказания мест связывания анионов – например, нуклеиновых кислот

Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. Jones S, Shanahan HP, Berman HM, Thornton JM. *Nucleic Acids Res.* 2003

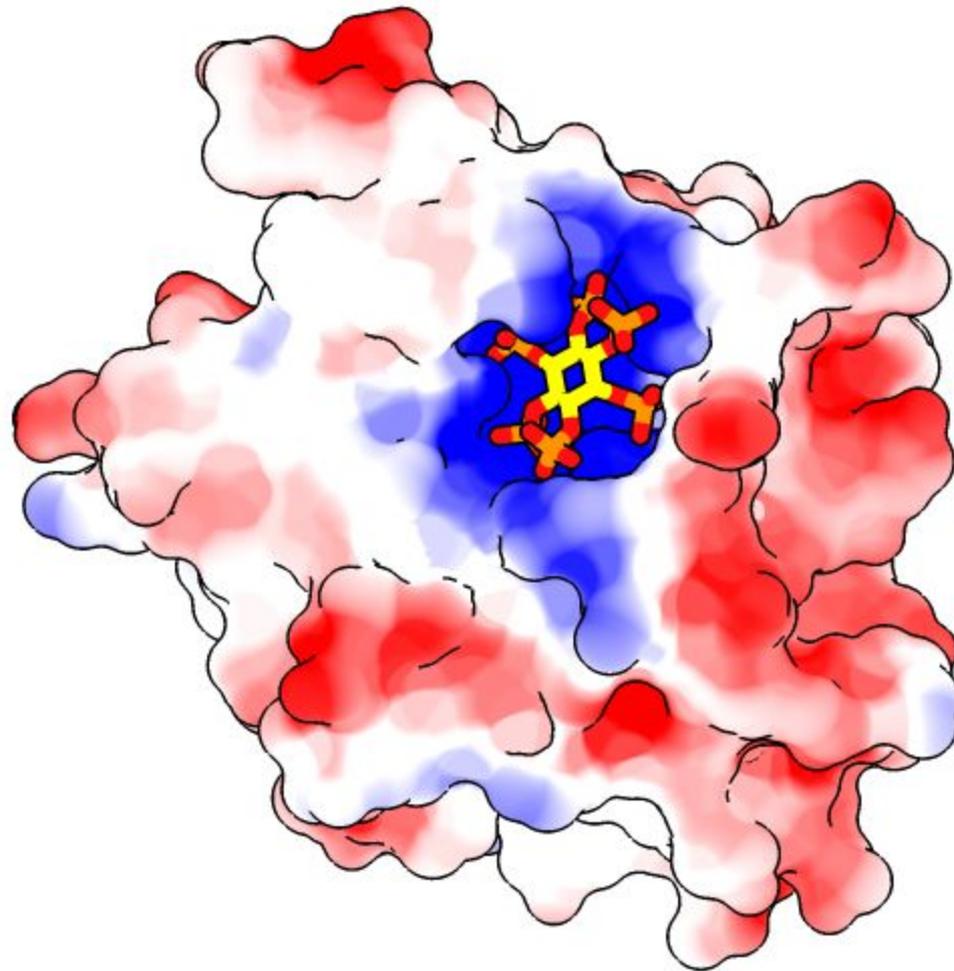
Structure-based prediction of DNA-binding sites on proteins Using the empirical preference of electrostatic potential and the shape of molecular surfaces. Yuko Tsuchiya, Kengo Kinoshita, Haruki Nakamura, *Proteins*, 2004

(A) 1b3t (nuclear protein EBNA1)

P_{score}	0.42
CC_V	0.64

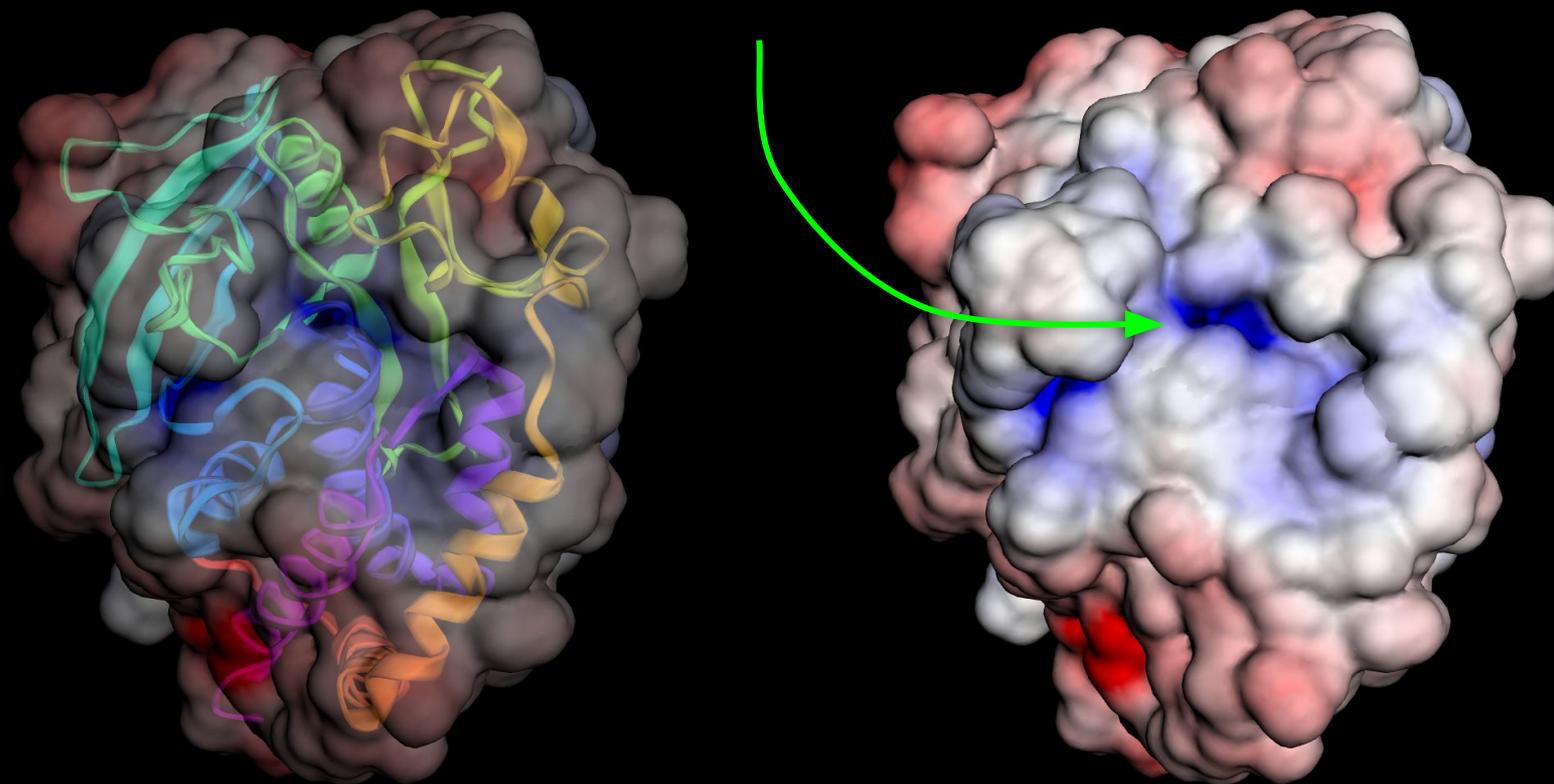


* Сейчас лучше справляются комбинированные модели, например Guy Nimrod, András Szilágyi, Christina Leslie, Nir Ben-Tal, **Identification of DNA-binding Proteins Using Structural, Electrostatic and Evolutionary Features**, *Journal of Molecular Biology*, 2009



Кейс: тирозин фосфатаза. Сервер [APBS](#)

Самое положительно заряженное место –
карман связывания фосфата



Гидрофобный потенциал

- Условные единицы, присваиваемые определенному типу аминокислотного остатка
- Существует порядка 100 различных шкал гидрофобности
- Исходные данные – физико-химические свойства пептидов из 2-50 остатков

$$t_i = a_0 + \sum_{s \in AAs} n_s^i a_s$$

t_i – свойство конкретного i -го пептида, например, время удержания на гидрофобной колонке

n_s^i – количество остатков типа s в пептиде i

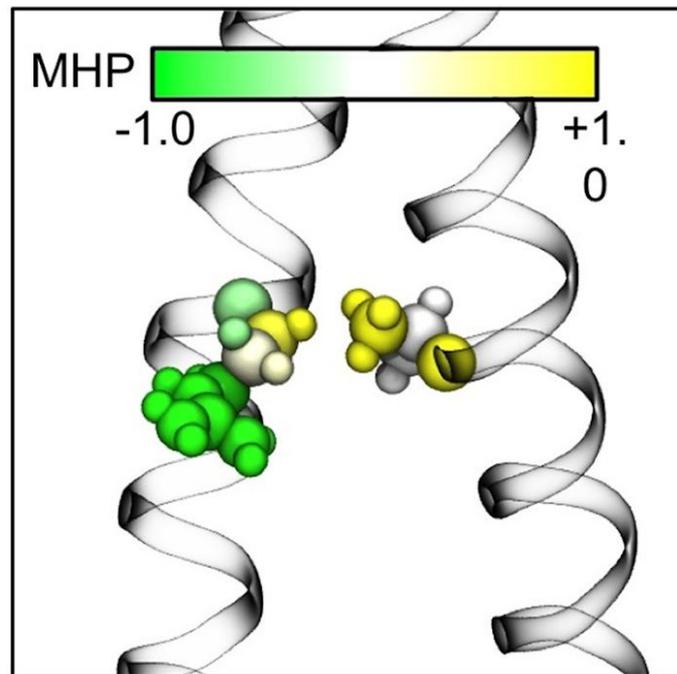
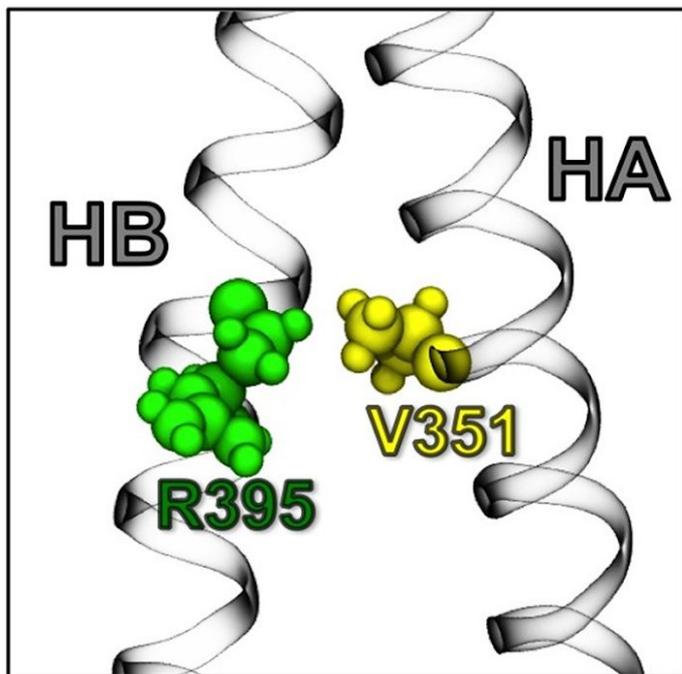
Подбираются коэффициенты a_0 и a_s для всех s

Шкалы гидрофобности

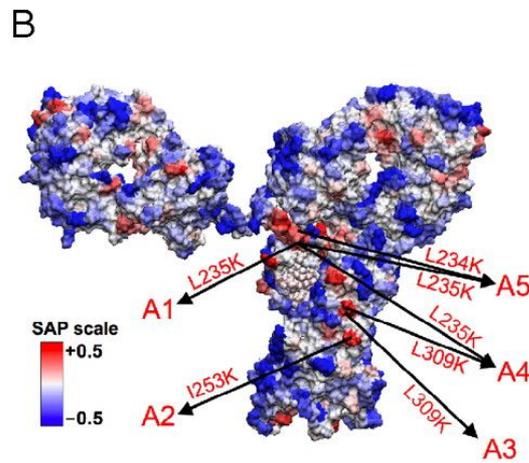
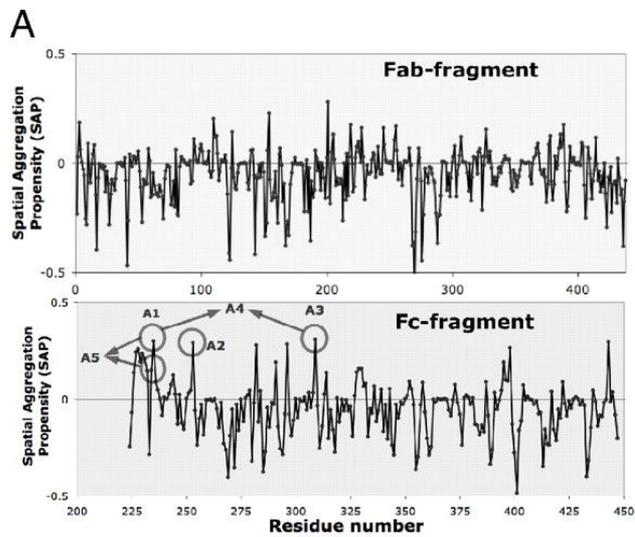
aa	aa	Kyte-Doolittle	Hopp-Woods	Cornette	Eisenberg	Rose	Janin	Engelman (GES)
A	Alanine	1.80	-0.50	0.20	0.62	0.74	0.30	1.60
C	Cysteine	2.50	-1.00	4.10	0.29	0.91	0.90	2.00
D	Aspartic acid	-3.50	3.00	-3.10	-0.90	0.62	-0.60	-9.20
E	Glutamic acid	-3.50	3.00	-1.80	-0.74	0.62	-0.70	-8.20
F	Phenylalanine	2.80	-2.50	4.40	1.19	0.88	0.50	3.70
G	Glycine	-0.40	0.00	0.00	0.48	0.72	0.30	1.00
H	Histidine	-3.20	-0.50	0.50	-0.40	0.78	-0.10	-3.00
I	Isoleucine	4.50	-1.80	4.80	1.38	0.88	0.70	3.10
K	Lysine	-3.90	3.00	-3.10	-1.50	0.52	-1.80	-8.80
L	Leucine	3.80	-1.80	5.70	1.06	0.85	0.50	2.80
M	Methionine	1.90	-1.30	4.20	0.64	0.85	0.40	3.40
N	Asparagine	-3.50	0.20	-0.50	-0.78	0.63	-0.50	-4.80
P	Proline	-1.60	0.00	-2.20	0.12	0.64	-0.30	-0.20
Q	Glutamine	-3.50	0.20	-2.80	-0.85	0.62	-0.70	-4.10
R	Arginine	-4.50	3.00	1.40	-2.53	0.64	-1.40	-12.3
S	Serine	-0.80	0.30	-0.50	-0.18	0.66	-0.10	0.60
T	Threonine	-0.70	-0.40	-1.90	-0.05	0.70	-0.20	1.20
V	Valine	4.20	-1.50	4.70	1.08	0.86	0.60	2.60
W	Tryptophan	-0.90	-3.40	1.00	0.81	0.85	0.30	1.90
Y	Tyrosine	-1.30	-2.30	3.20	0.26	0.76	-0.40	-0.70

Шкалы гидрофобности – спускаемся до атомов

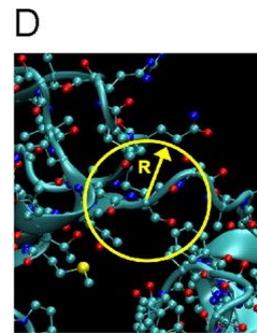
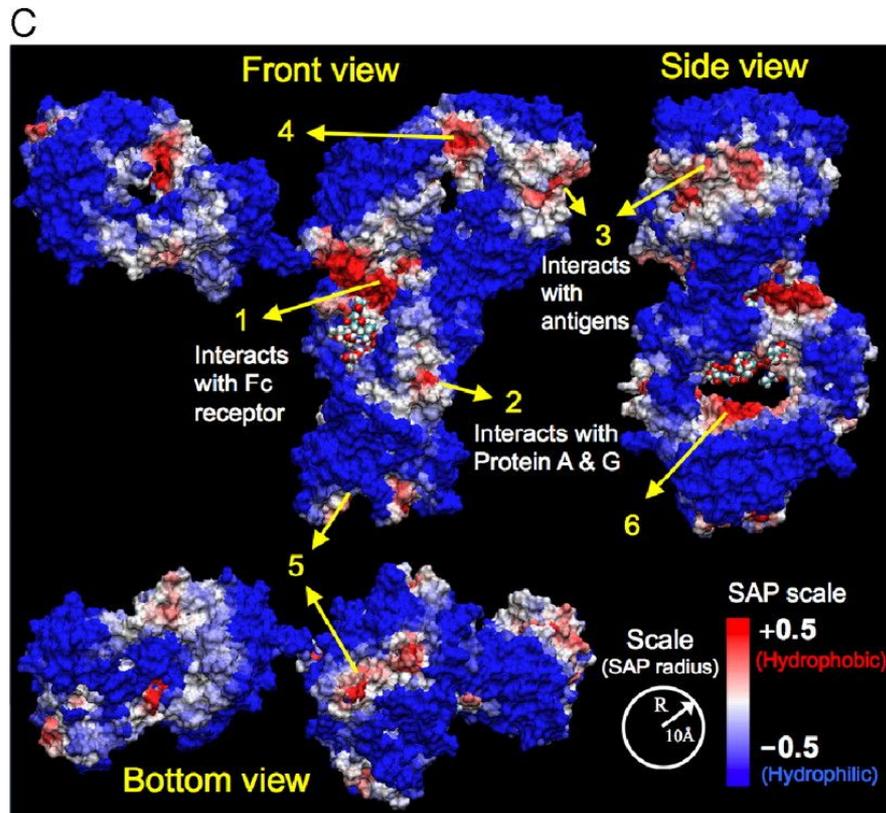
(A)



Fine mapping of hydrophobic contacts reassesses the organization of the first three dystrophin coiled-coil repeats. Dominique Mias-Lucquin, Angélique Chéron, Elisabeth Le Rumeur, Jean-François Hubert, Olivier Delalande. *Protein Science*, 2018



Гидрофобные регионы на поверхности склонны к агрегации



Идентификация таких регионов и мутирование на полярные аминокислоты понизило ненужную агрегацию терапевтических антител.

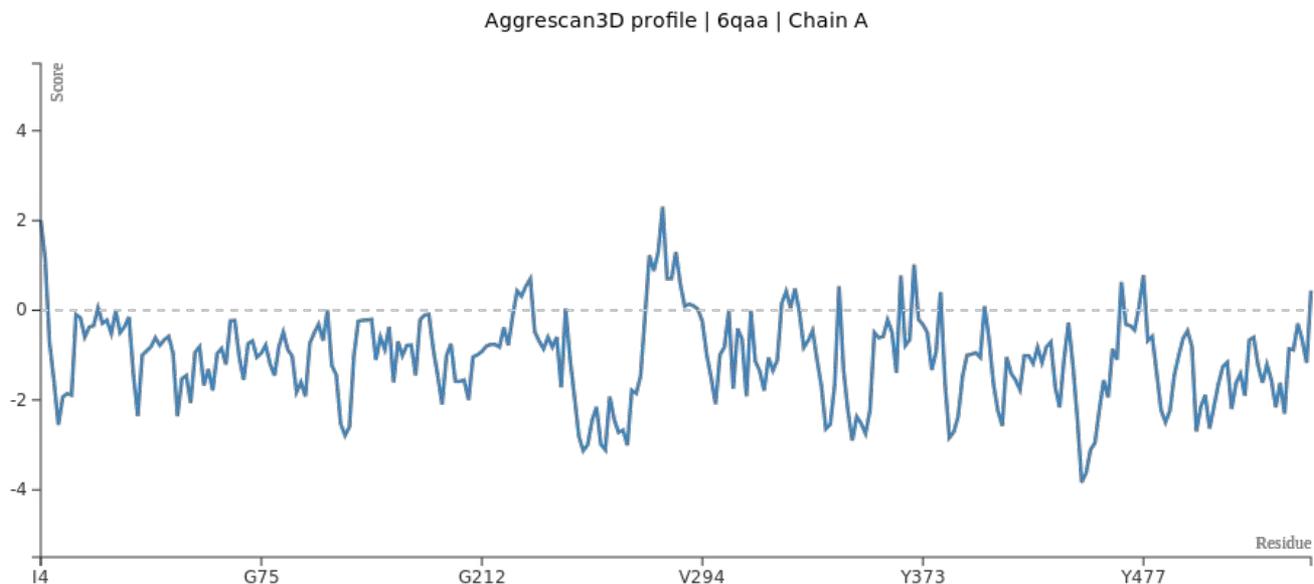
Design of therapeutic proteins with enhanced stability. Naresh Chennamsetty, Vladimir Voynov, Veysel Kayser, Bernhard Helk, and Bernhardt L. Trout. *PNAS* July 21, 2009

Инструменты

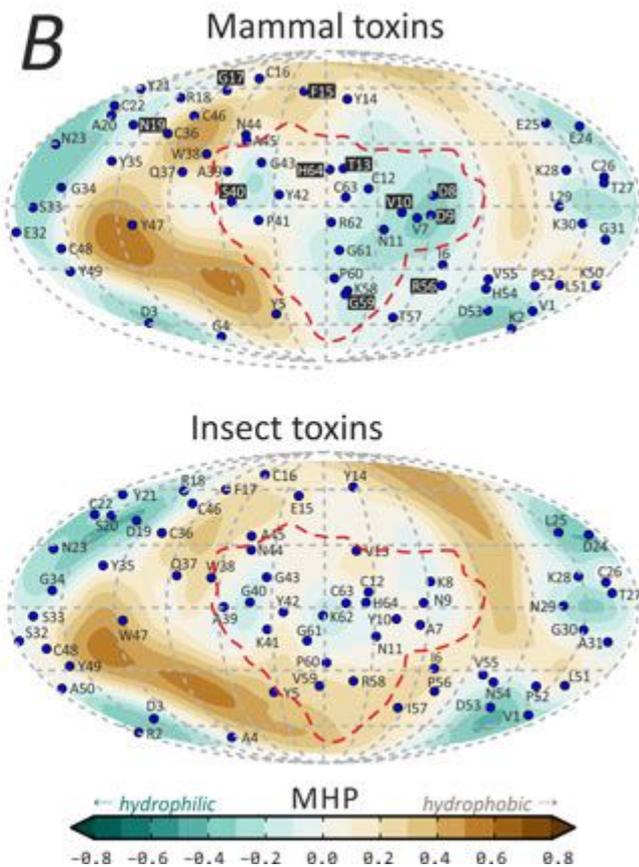
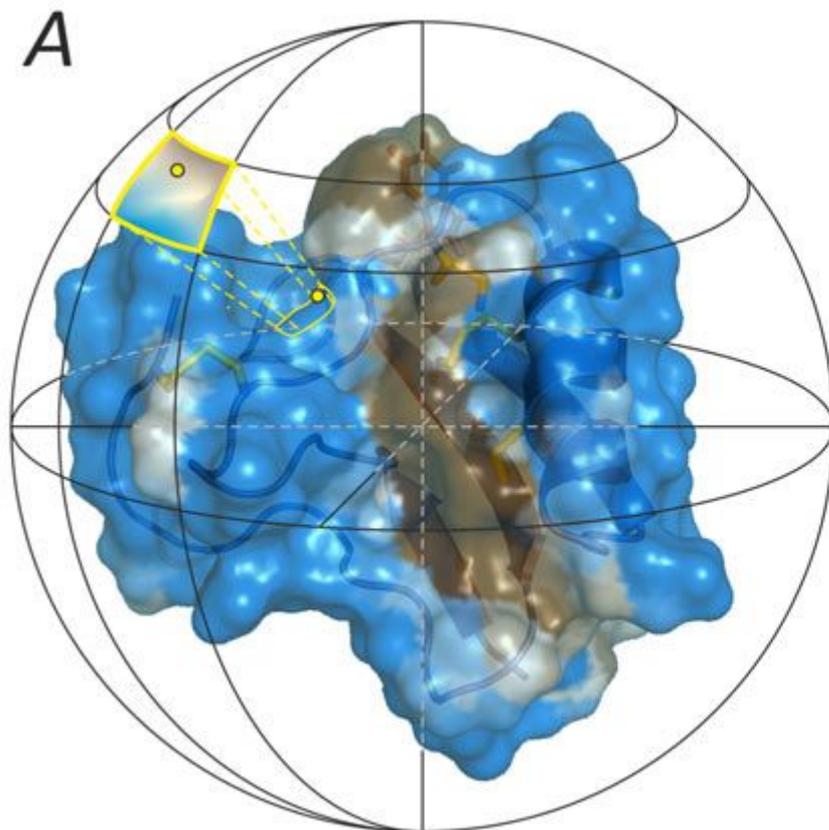
Aggrescan3D

CamSol

Снижение вероятности гидрофобной агрегации коррелирует с повышением растворимости белка. Для многих промышленных и терапевтических применений это важная задача.

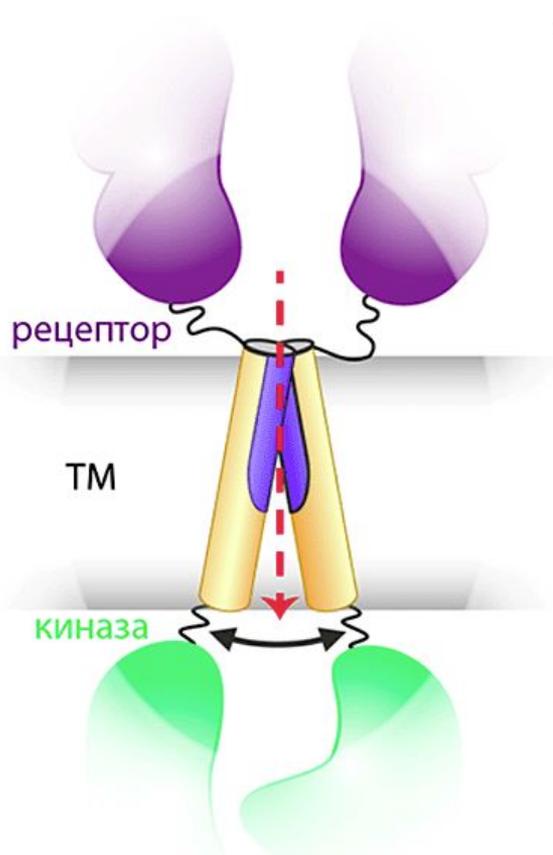


От поверхности к Image Recognition

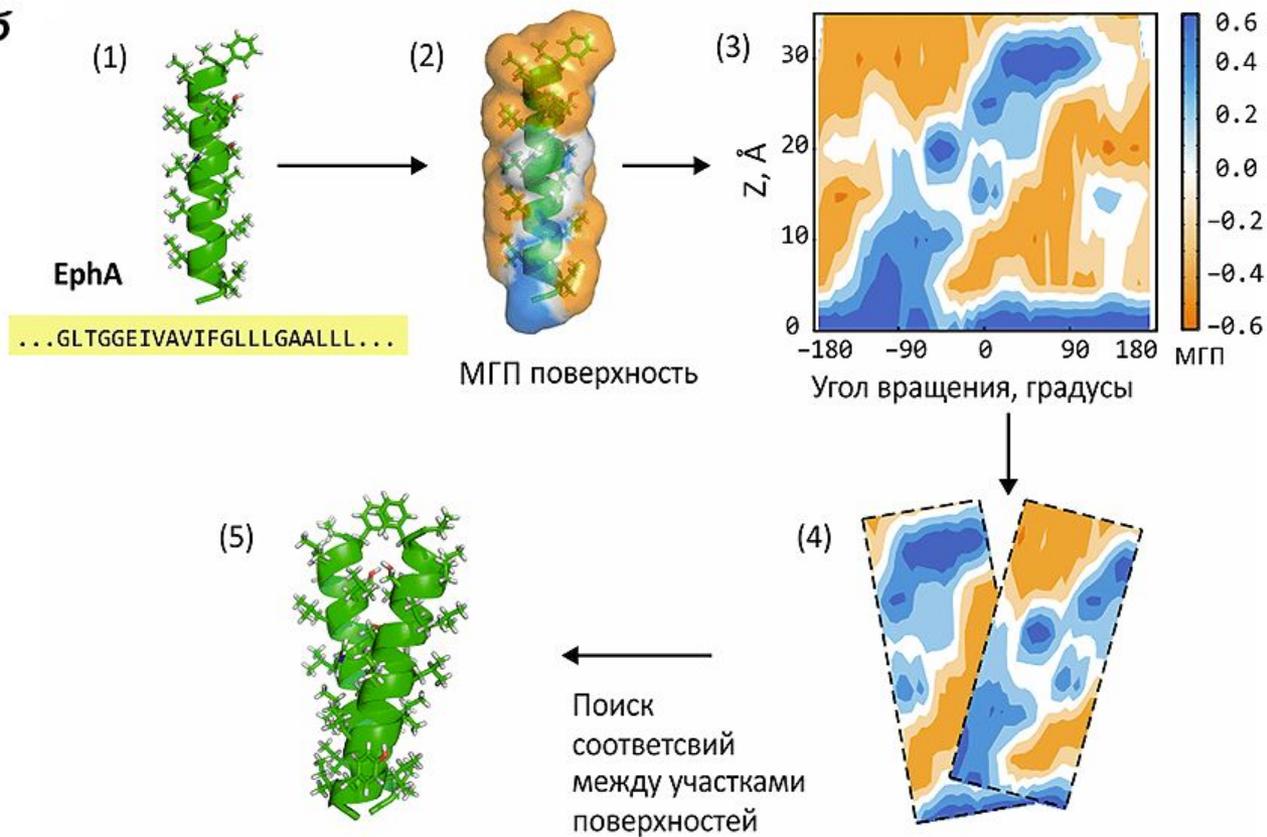


Предсказание ориентации спиралей при димеризации

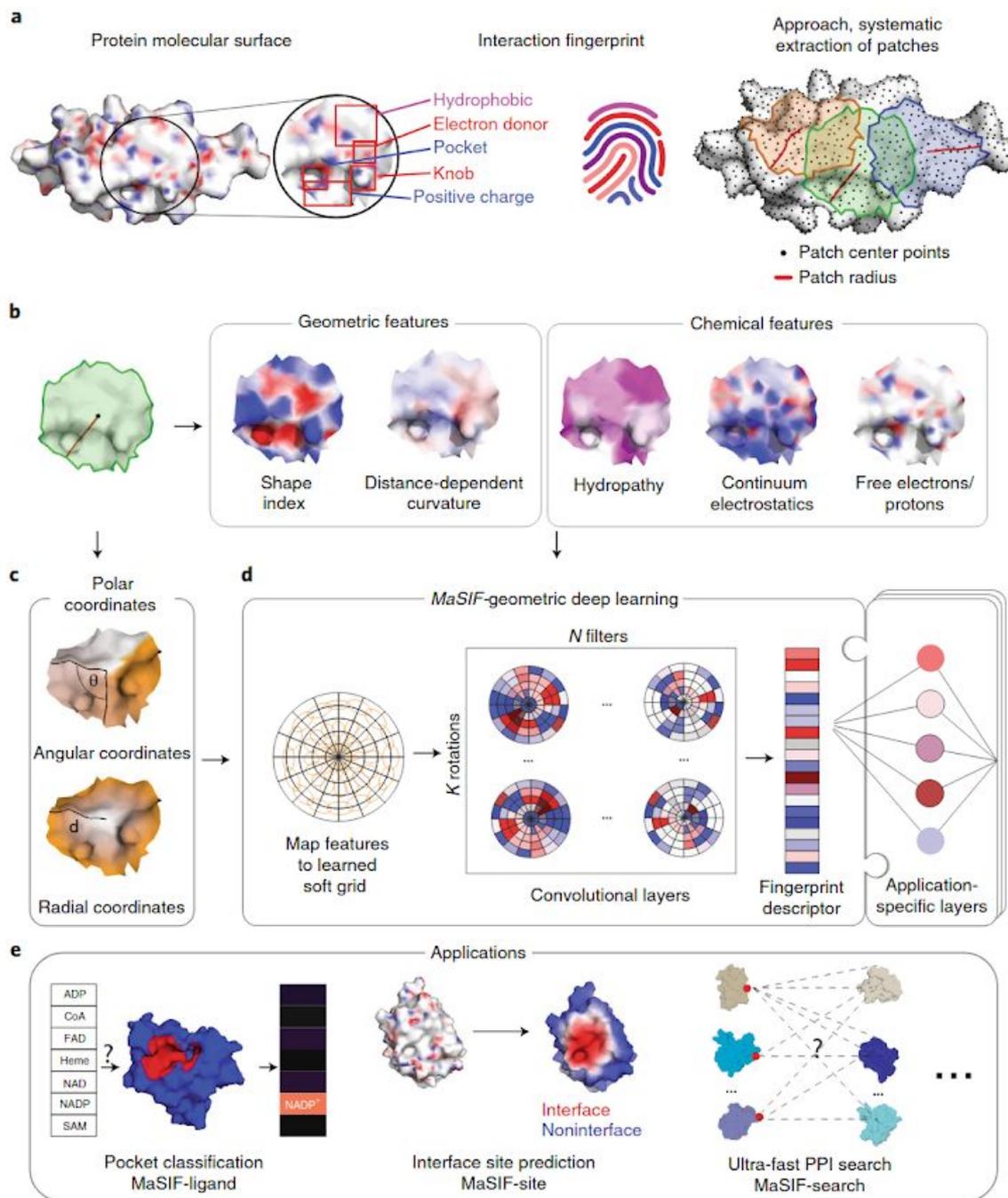
a



б



MaSIF



Так а все же, что такое сама поверхность?

Поверхность – граница между объектом и внешней средой

Но в микромире абсолютно “непрерывных” тел не бывает. Нужно указывать для каких частиц непроницаема область пространства – нейтрино? фотонов? электронов? протонов? других молекул (каких)?

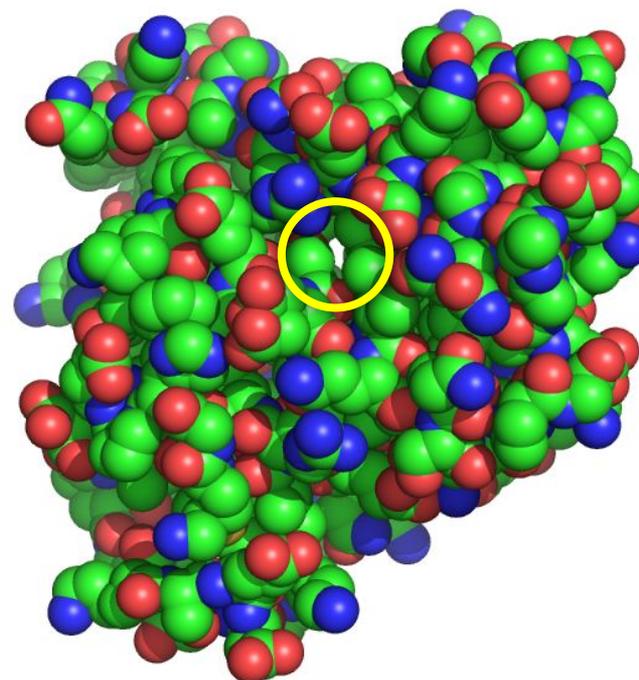
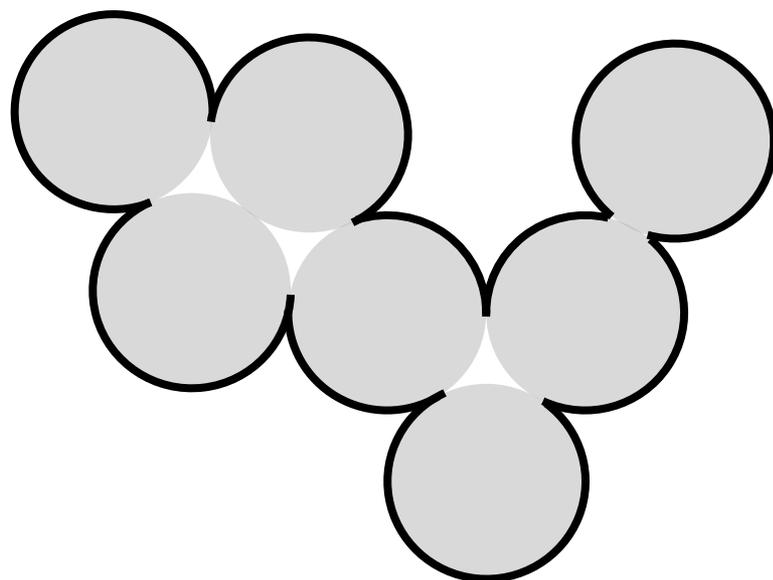
Так а все же, что такое сама поверхность?

Вообще говоря, выделяют три вида поверхностей:

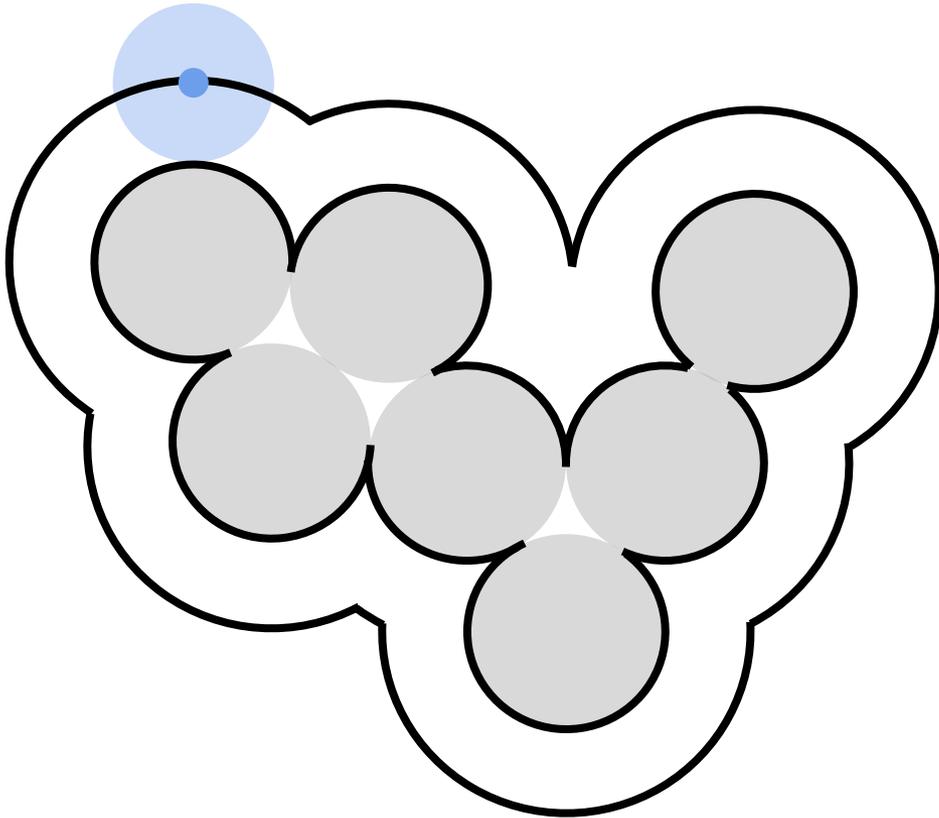
- Ван-дер-Ваальсова поверхность
- Поверхность, доступная растворителю
- Молекулярная поверхность (поверхность Конноли)

} Lee & Richards, 1971, JMB

VdW поверхность:



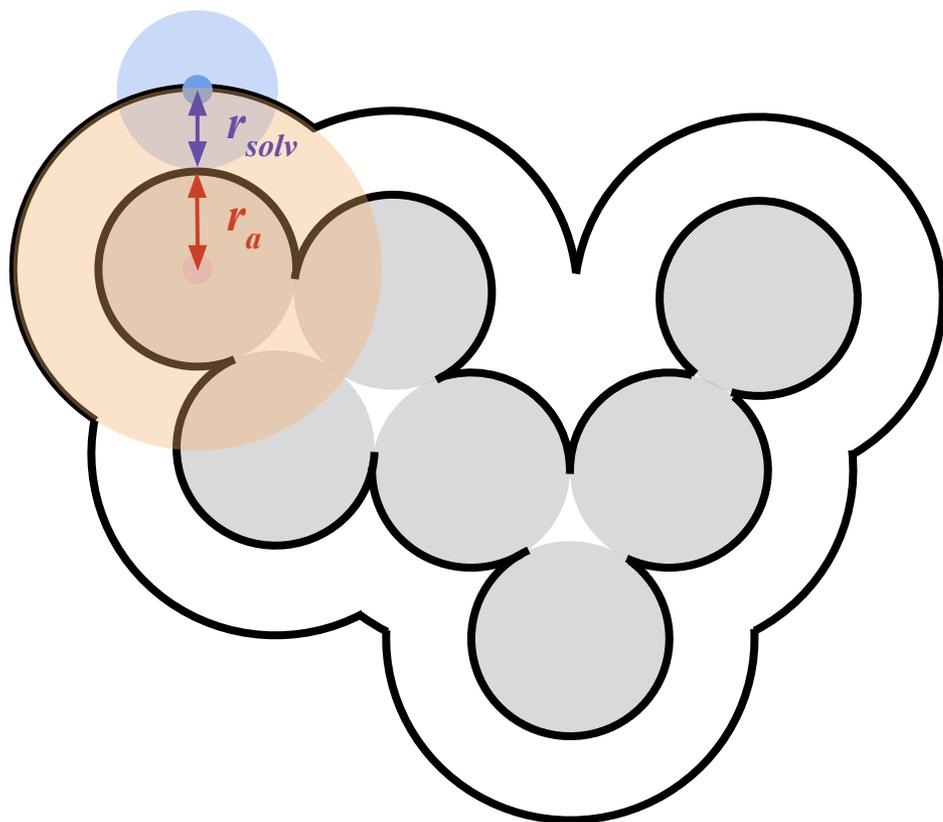
Поверхность, доступная растворителю (SAS)



Это множество наиболее близких к атомам белка точек, в которых может находиться центр молекулы растворителя.

Растворитель описывается простой сферой. Для “воды” её радиус обычно 1.4\AA .

Поверхность, доступная растворителю (SAS)



Это множество наиболее близких к атомам белка точек, в которых может находиться центр молекулы растворителя.

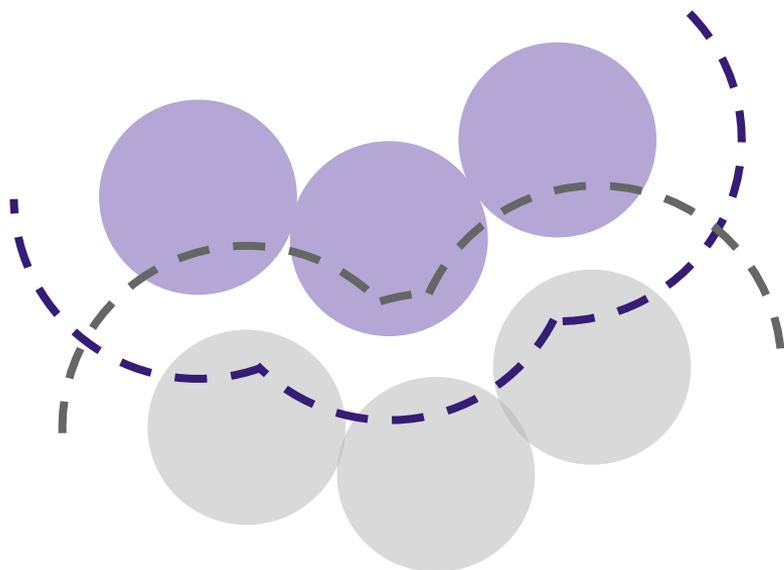
Растворитель описывается простой сферой. Для “воды” её радиус обычно 1.4Å.

С точки зрения вычислений построить SAS – то же самое, что и VdW поверхность, только за VdW радиус каждого атома a белка берется $r_a^{SAS} = r_a + r_{solv}$

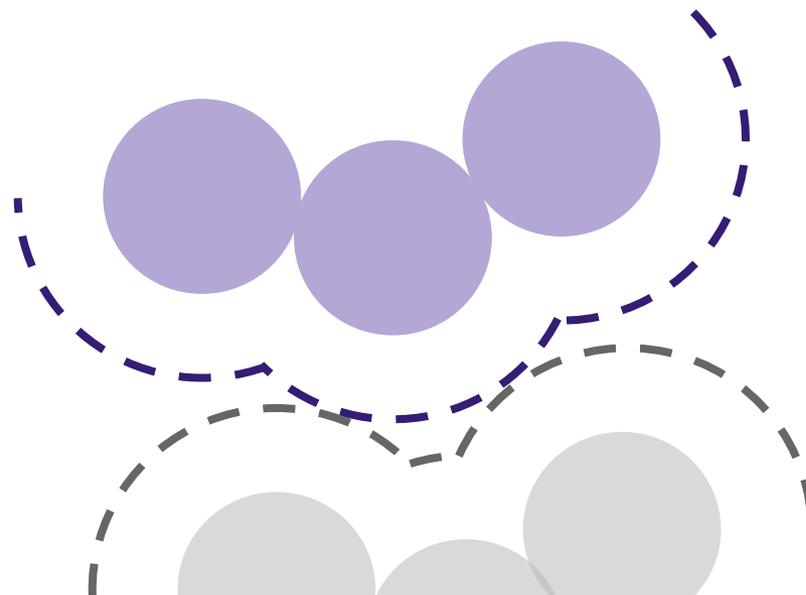
Применение SAS

Поверхность, доступная для воды, используется, например, для того, чтобы показать, какие аминокислотные остатки чаще экспонированы – доступны для воды.

SAS не всегда применима, так как «раздувает» молекулу. Например, при контакте двух белков их SAS пересекаются:



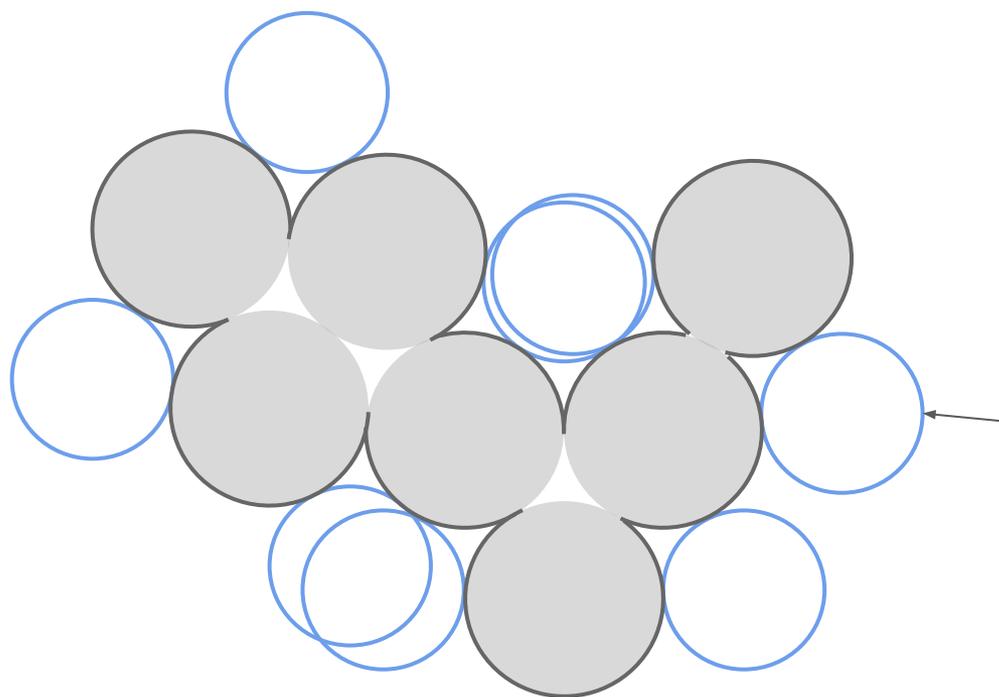
Молекулы действительно в контакте, но никакой shape complementarity!



Shape complementarity, но, получается, между белками поместится целый слой молекул воды – контакт сомнительной силы получается

Молекулярная поверхность (MS)

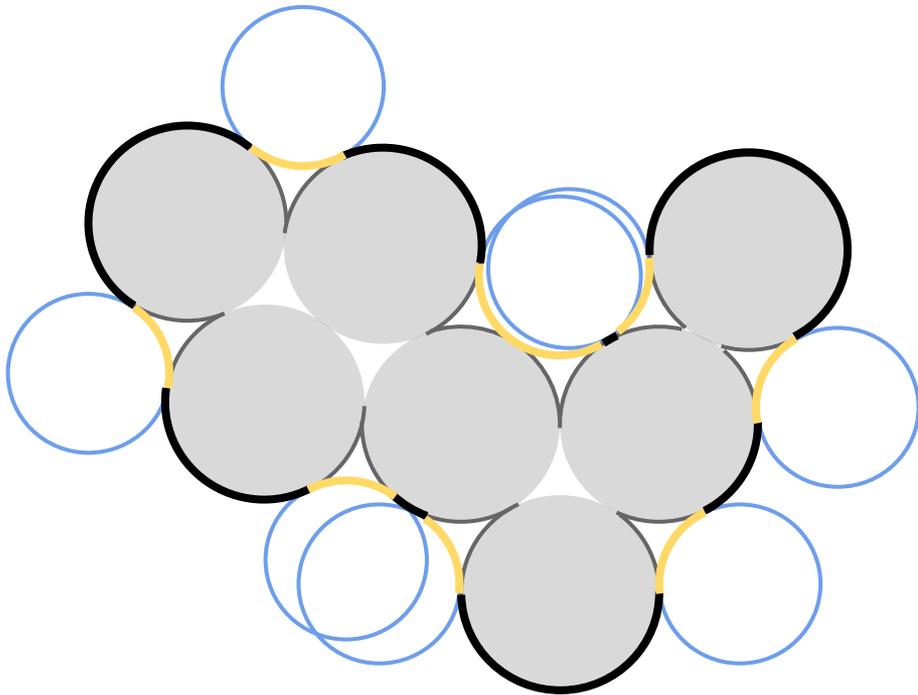
Она же поверхность Конноли



Молекулы воды, касающиеся белка в 2 или 3 точках (в 3D)

Молекулярная поверхность (MS)

Она же поверхность Конноли

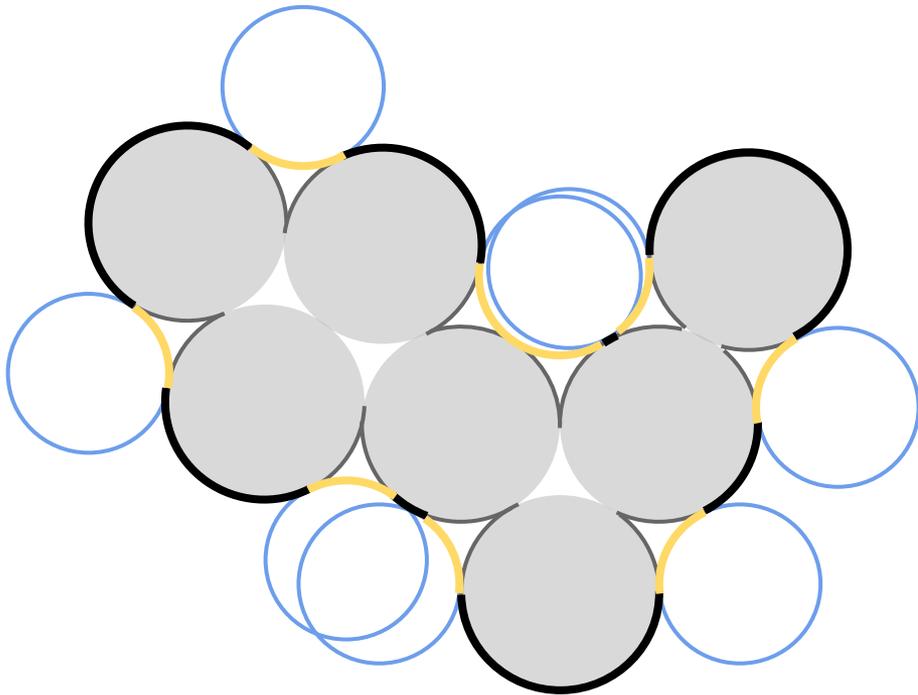


Состоится из:

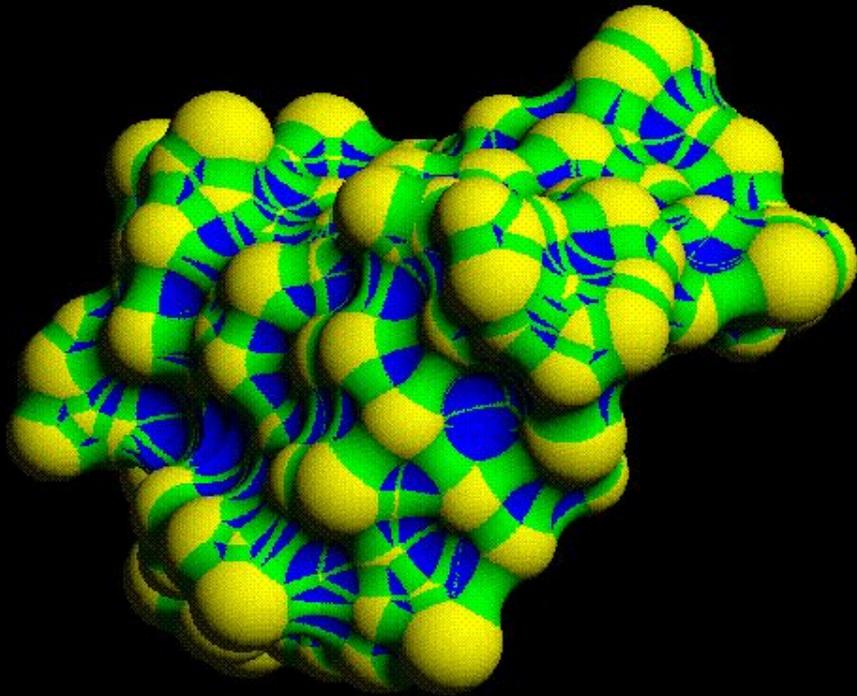
- Поверхности контакта (contact surface) – **черной**
- Дополнительной поверхности (reentrant surface) – **желтой**

Молекулярная поверхность (MS)

Она же поверхность Конноли



- **Поверхность контакта**
образована точками ван-дер-ваальсовых сфер атомов белка, которых может коснуться ван-дер-ваальсова сфера молекулы воды
- **Дополнительная поверхность**
образована поверхностью молекул воды, касающихся белка в двух или трех точках



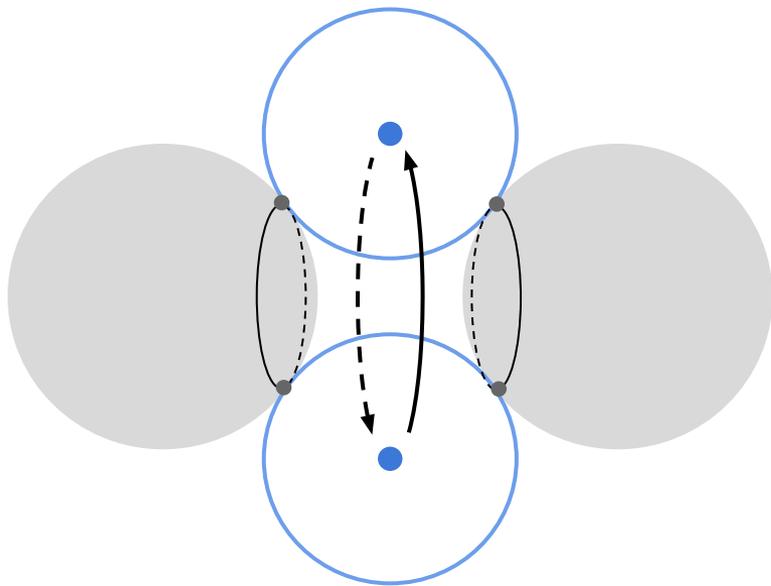
**Молекулярная поверхность
состоит из кусков трёх видов:**

кусок “выпуклой” сферы (жёлтая)

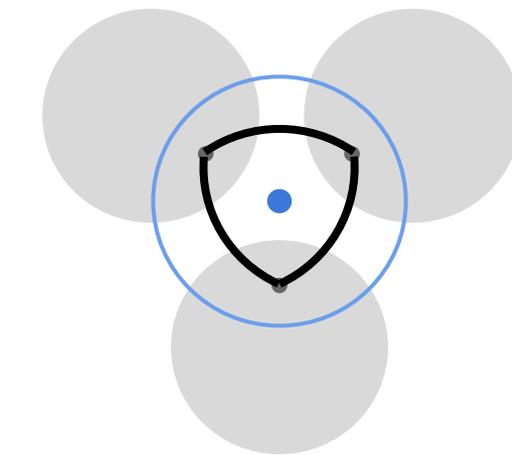
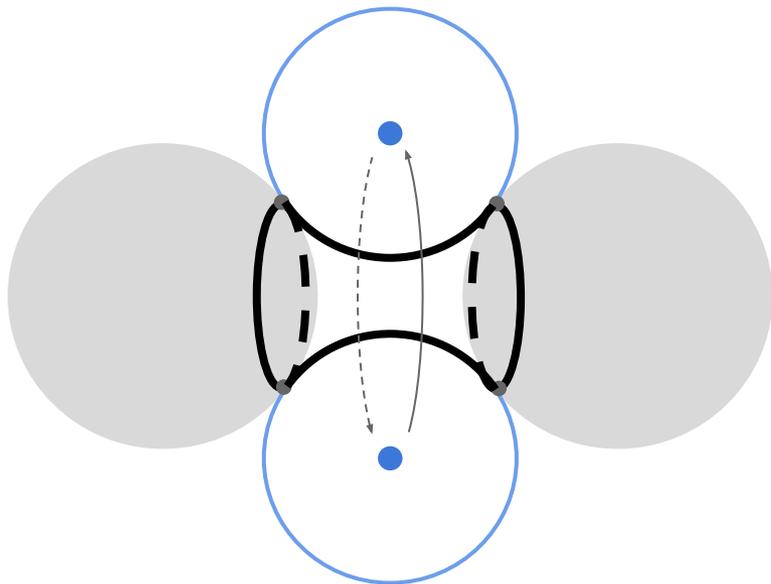
кусок “вогнутой” сферы (синяя)

тороидальная часть (зеленая)

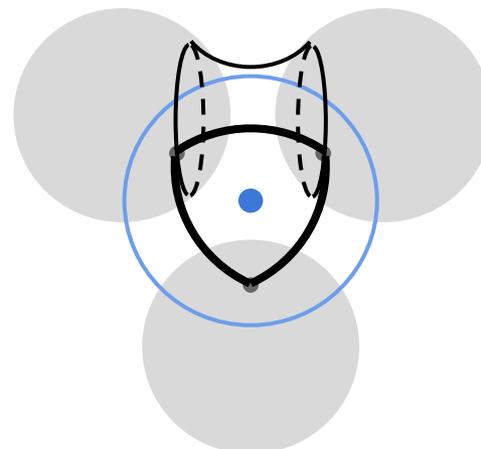
Все куски соединяются
гладким образом – без углов

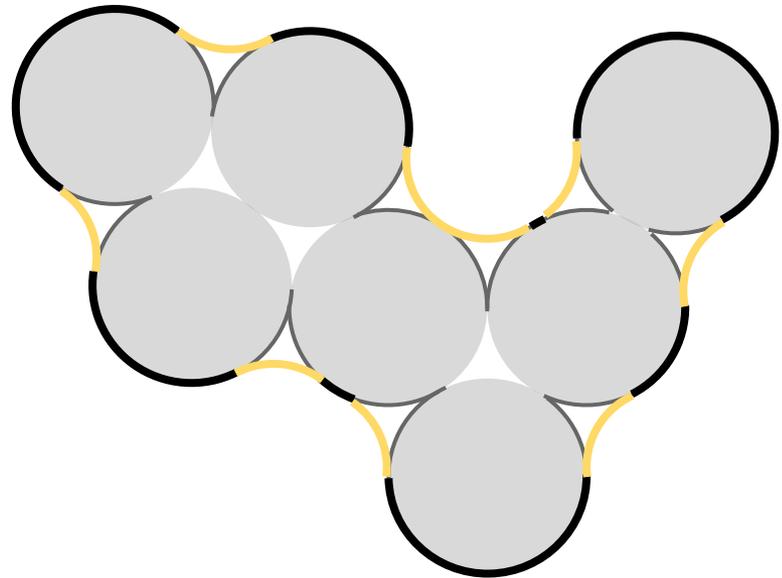
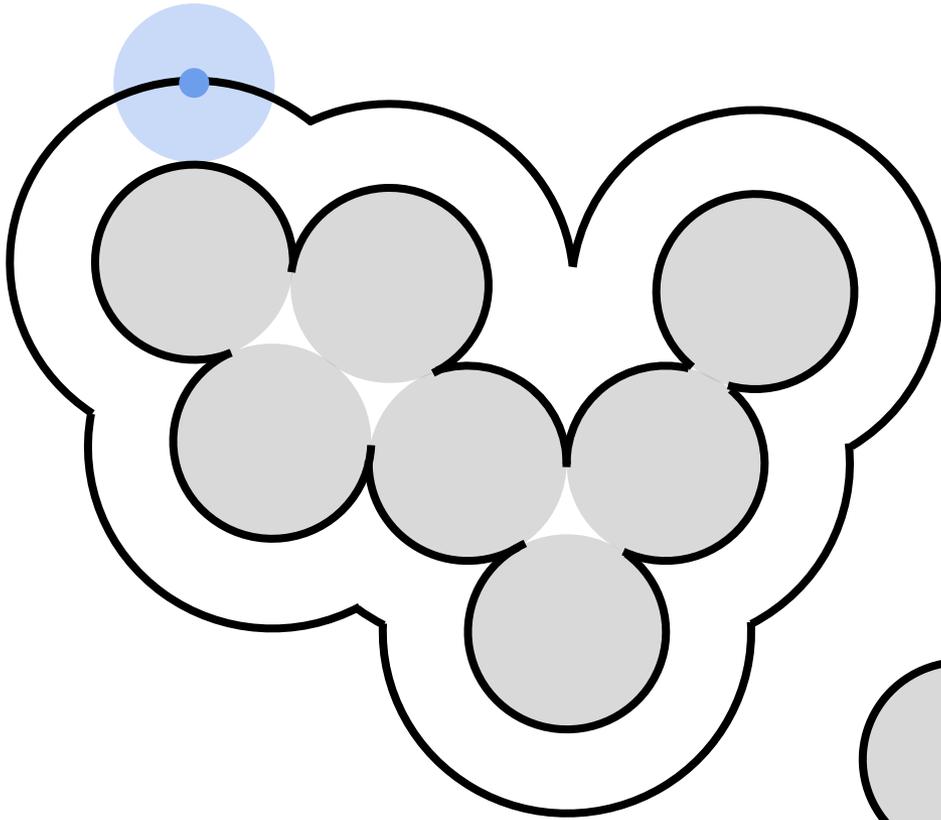


Тороидальная поверхность заматывается подвижным шариком (H_2O), который вращается между двумя фиксированными шарами (CH_3), все время касаясь обоих



Вогнутая сфера получается в том случае, когда шар H_2O касается одновременно трёх атомов белка





Вычисление площади поверхности

- Приближённые аналитические методы (Richards&Lee, 1971; Wodak and Janin, 1980)
- Представление поверхности точками (Shrake&Rupley, 1973; Connolly, 1983)
- Точные аналитические методы (Gibson&Scheraga, 1987; Richmond, 1984)

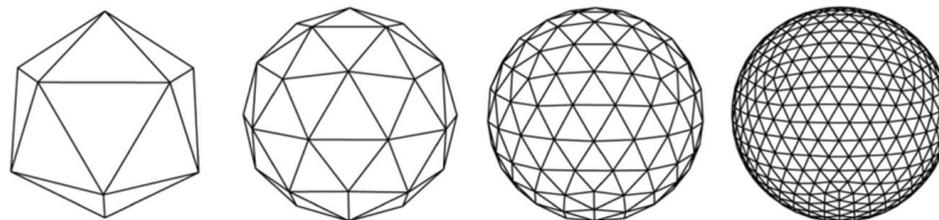
Метод срезов Ли – Ричардса для вычисления площади SAS

1. Структура режется на «ломтики» фиксированной толщины
2. Для каждого «ломтика» находятся круги от «срезов» атомов
3. Вычисляется длина границы
4. Умножается на толщину дольки
5. Берется сумма по всем срезам

Имплементирован, например, в программе [FreeSASA](#)

Mitternacht S. (2016) **FreeSASA: an open source C library for solvent accessible surface area calculations.** *F1000Research*

Shrake – Rupley

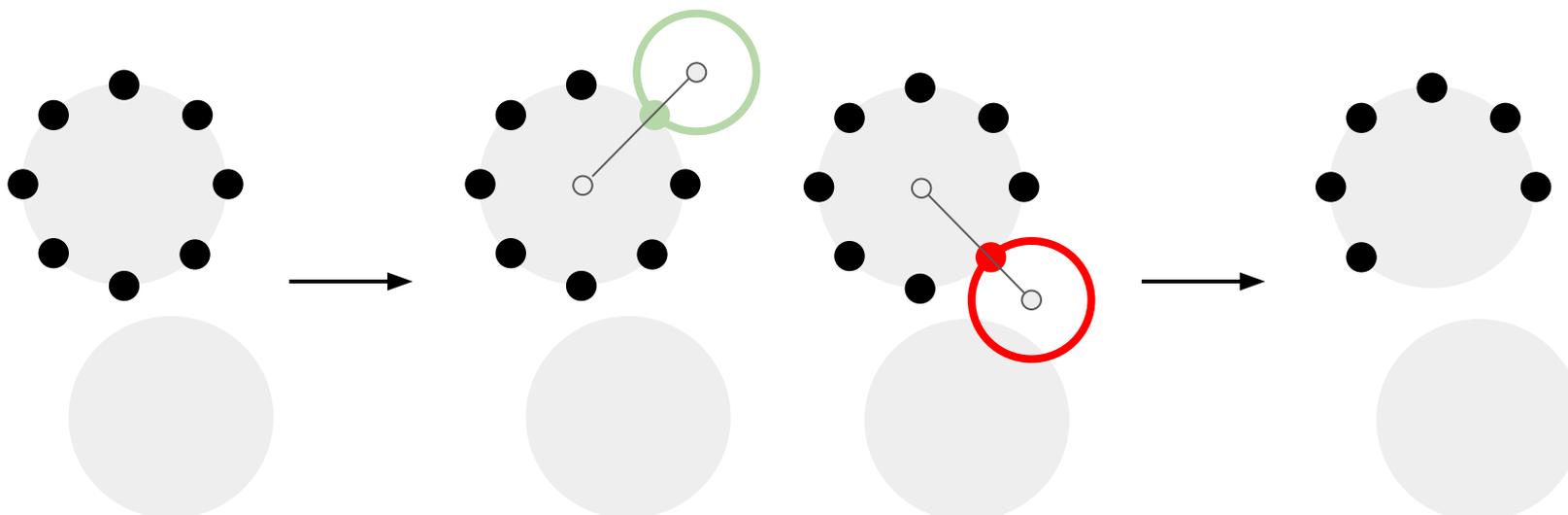


Контактная поверхность

На поверхности каждой VdW сферы атома белка строится равномерная сеть точек;

Для каждой точки проверяется, что молекула воды, касающаяся этой точки, не пересекается с белком;

Если пересекается, то точка **удаляется**.

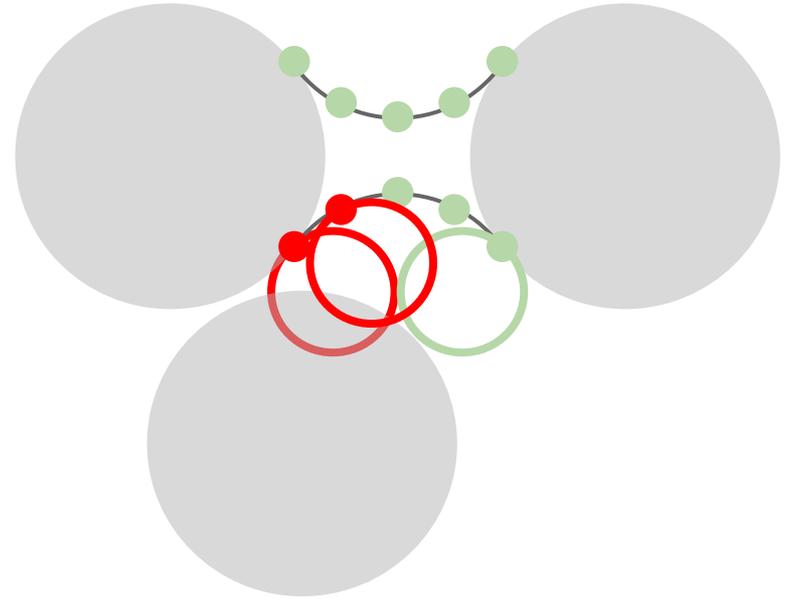


Shrake – Rupley

Дополнительная поверхность – тороидальная

Каждая пара соседних атомов определяет тороидальную поверхность между ними

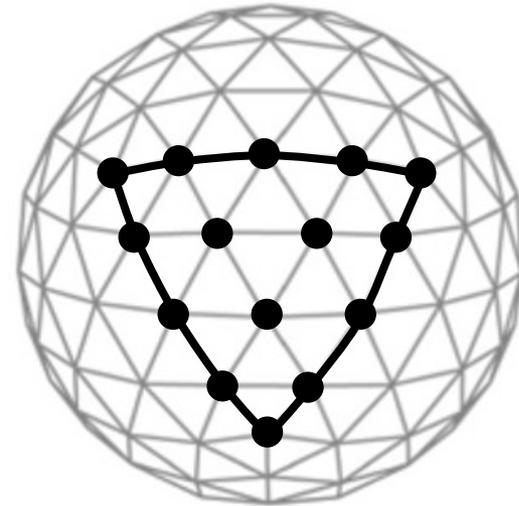
На этой поверхности строится равномерная сеть точек. Далее – как для контактной поверхности.



Дополнительная поверхность – сферическая

Каждая тройка соседних атомов определяет сферическую дополнительную поверхность – ван-дер-ваальсову поверхность молекулы воды, касающейся этих атомов.

Если эта молекула воды не пересекается с белком, то на подходящей части этой поверхности строится равномерная сеть точек.



Shrake – Rupley

Оставшиеся точки представляют поверхность молекулы белка.

Их число пропорционально площади поверхности. На этих точках может быть построена триангуляция поверхности для визуализации (или более точного подсчета площади).

Shrake-Rupley имплементирован в PyMol (**get_area**)

Например, для одной контактной сферы:

$$Area = 4\pi r^2 \frac{N_{left}}{N_{all}}$$

Или

$$Area = \sum_{t \in triangles} S_t$$

Потенциальные артефакты при подсчете SAS

Присутствие лигандов (в том числе биомолекулярной природы)

Отсутствие протонов

Отсутствие атомов радикалов

Наличие alt loc'ов

Выбор радиуса растворителя

Выбор алгоритма (тип и уровень приближения может существенно сказываться)

Экспонированность остатка

Для каждого остатка считается площадь, выходящая на молекулярную поверхность (дополнительная площадь делится между соседями).

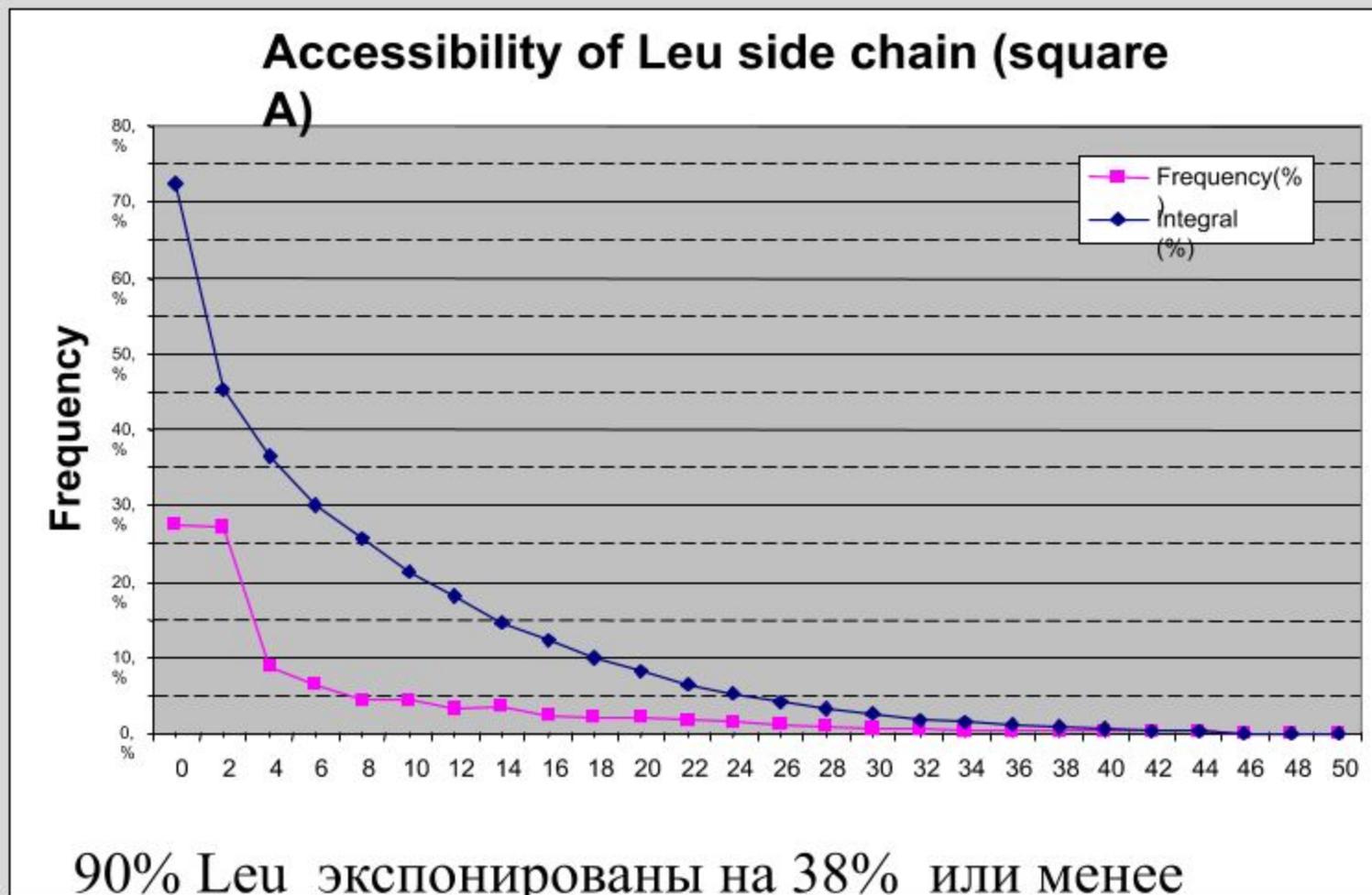
Эта площадь сравнивается с максимально возможной – при полностью раскрытой боковой цепи остатка того же типа в составе трипептида Gly – X – Gly.

Вычисляется отношение – это и есть **экспонированность**.

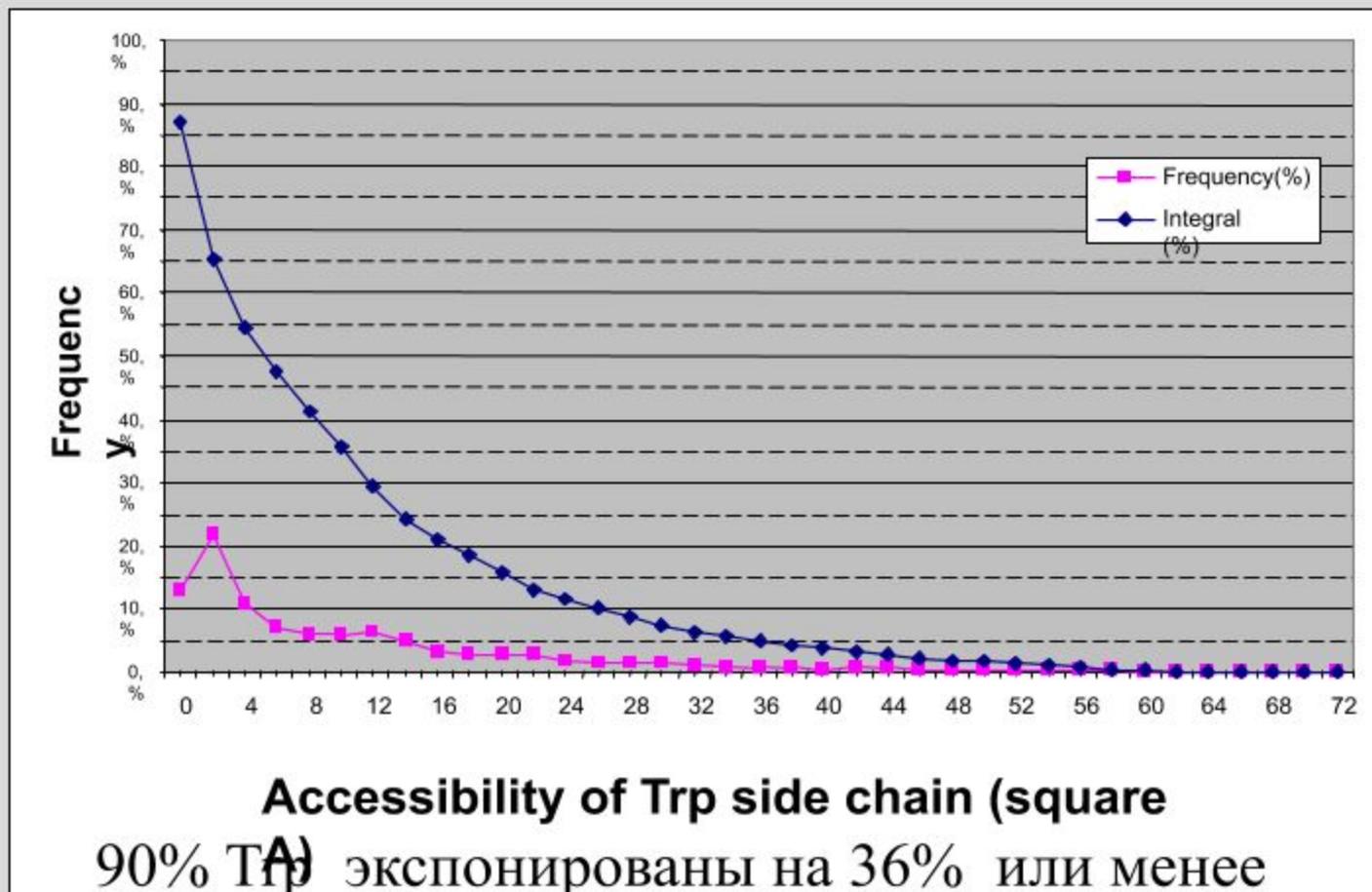
Иногда используют **погруженность** = 1 - экспонированность

Экспонированность боковой цепи Leu (похожие графики у Val, Ile, Met)

Max=48Å²

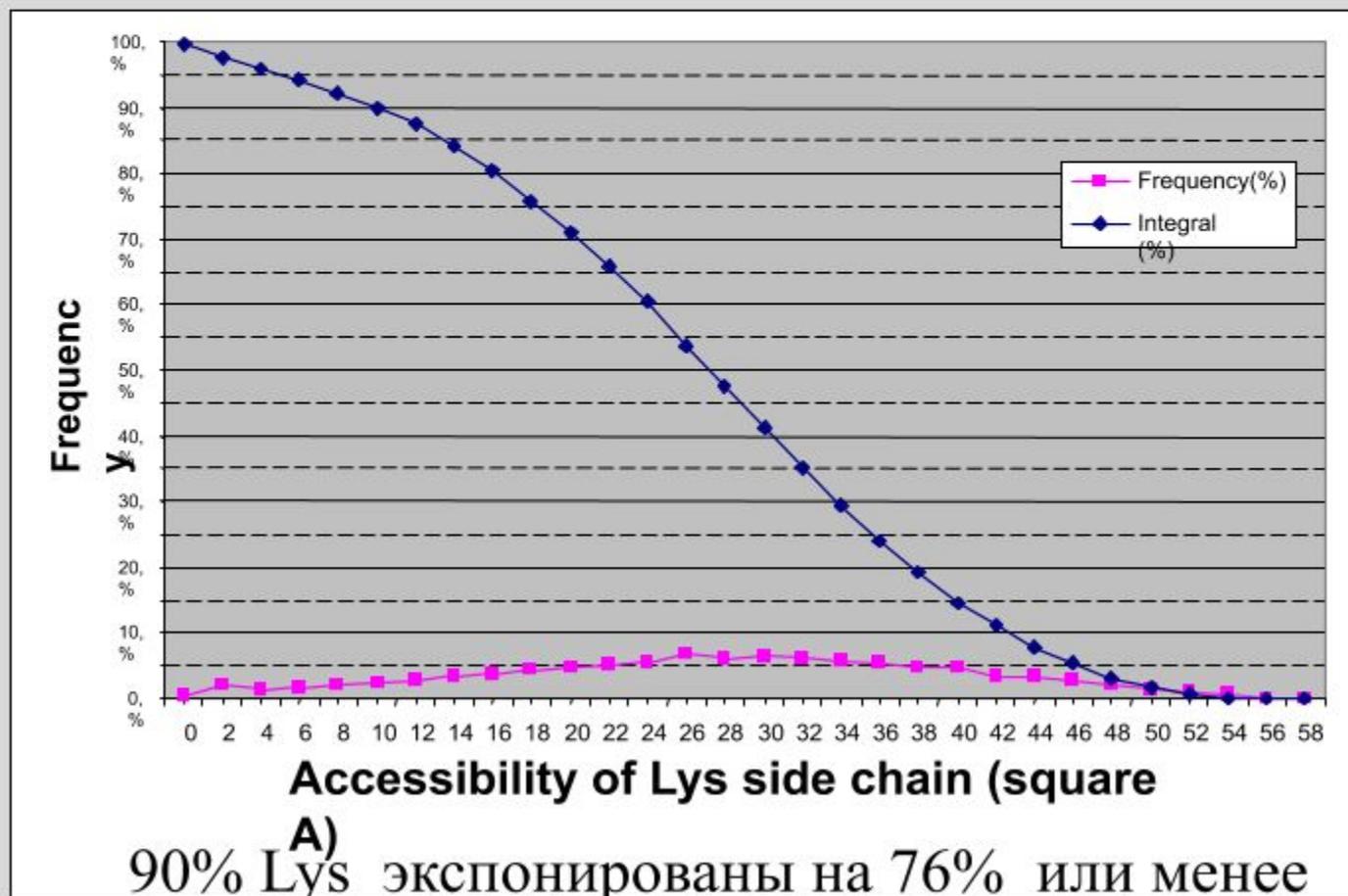


Экспонированность боковой цепи Trp (похожие графики у Tyr, His, Phe, Pro)



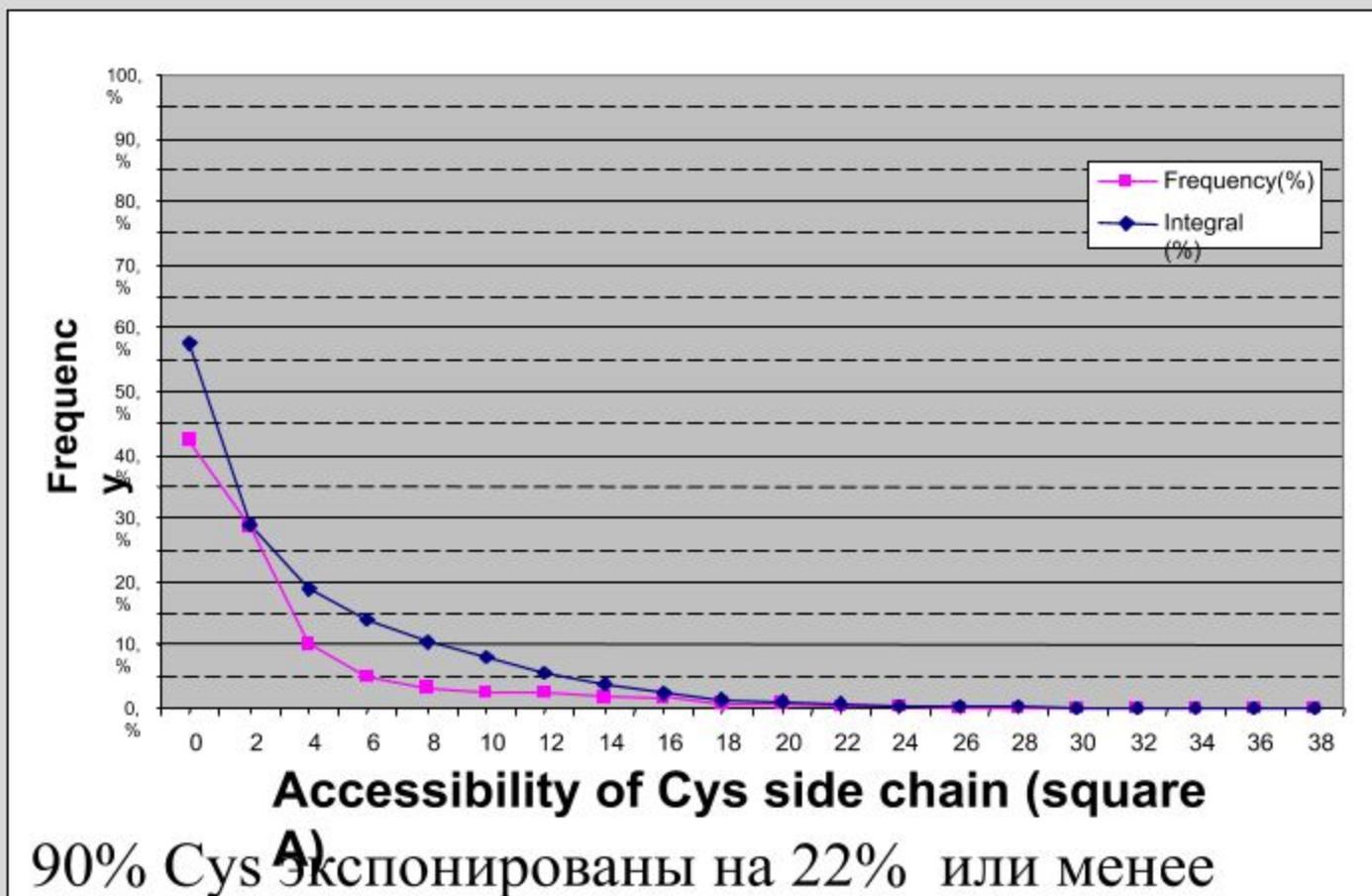
Max=72Å²

Экспонированность боковой цепи Lys (похожие графики у Arg, Gln, Glu, Asn, Asp)

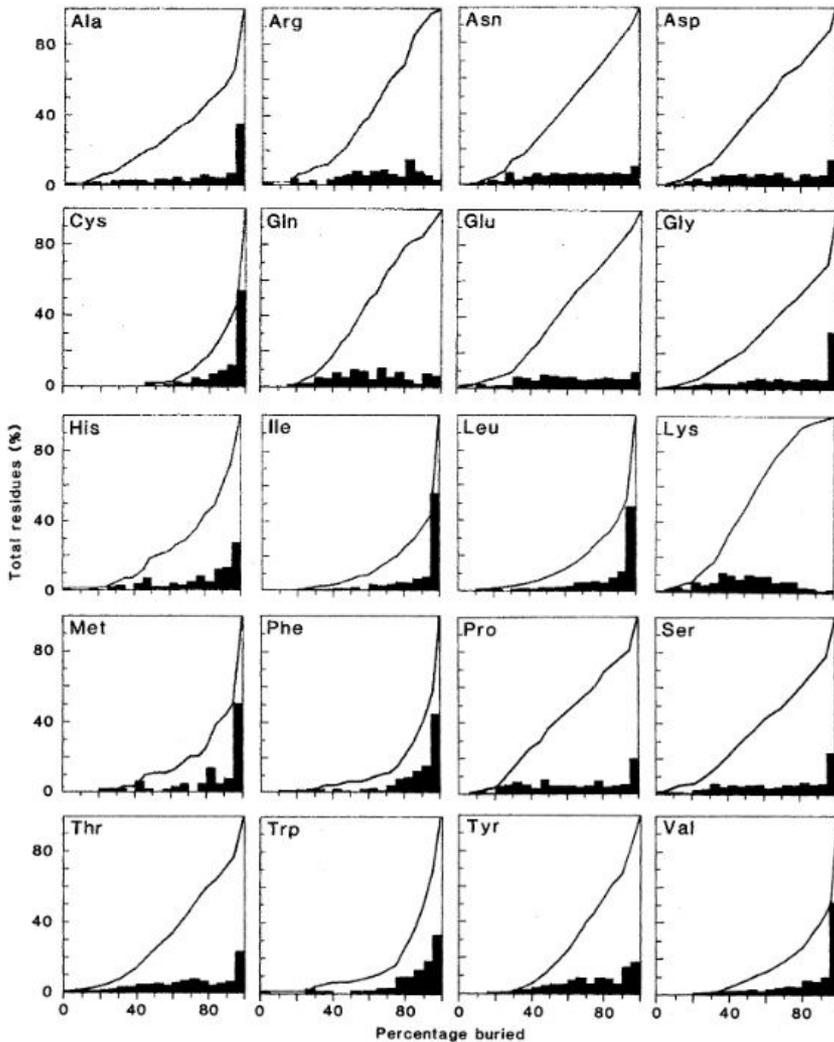


Экспонированность боковой цепи Cys

Max=37Å²



Погруженность



Hydrophobicity of Amino Acid Residues in Globular Proteins. George D. Rose, Ari R. Geselowitz, Glenn J. Lesser, Richard H. Lee and Micheal H. Zehfus. *Science*, 1985

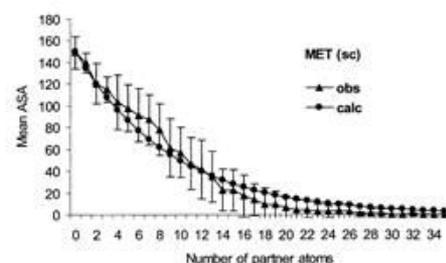
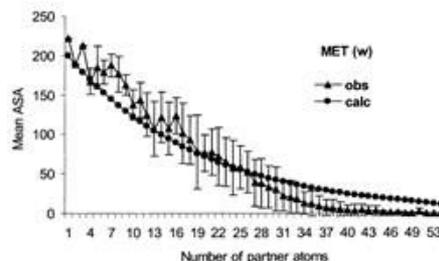
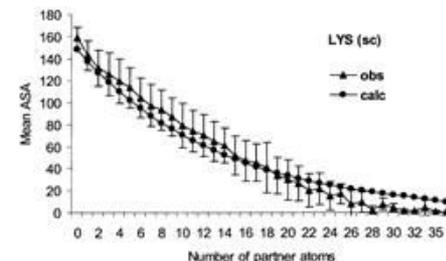
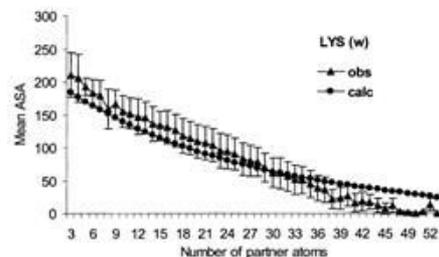
Погруженность, но ленивая

Как альтернативу отношению площадей считают число тяжелых атомов белка на каком-то расстоянии от данного (предлагается использовать 4.5Å).

Выясняется, что такая величина неплохо коррелирует с экспонированностью, и зависимость может быть описана как

$$y = a_1 * \exp\left(-\frac{x}{a_2}\right)$$

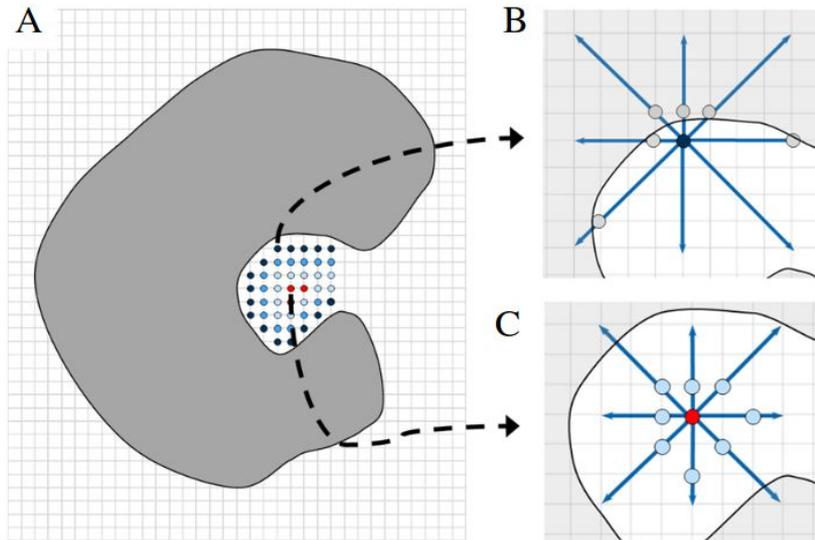
где y это SASA, а x это число атомов-соседей.



Карманы СВЯЗЫВАНИЯ

Способы определения карманов связывания

Алгоритмы на гридах (сетках). CAVIAR



Инициализируется сетка с определенным шагом

Для каждой вершины по каждому направлению вычисляется, упирается ли луч из этой вершины в какой-либо атом белка, расположенный не дальше порогового расстояния

Если таких направлений больше, чем другое пороговое значение, то точка называется “внутренней”

Для новых итераций атомами белка считаются так же вершины сетки, обработанные ранее

Способы определения карманов связывания

Алгоритмы на гридах (сетках).

Примеры программ, реализующих данный метод:

- POCKET
- LIGSITE
- McVol
- VICE
- VolSite
- CAVIAR

Способы определения карманов связывания

Алгоритмы на гридах (сетках) с оценкой взаимодействия с пробой

Инициализируется сетка с определенным шагом

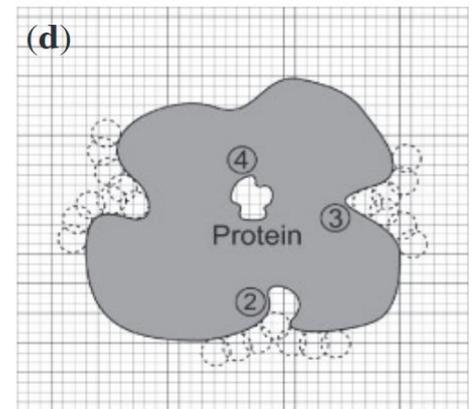
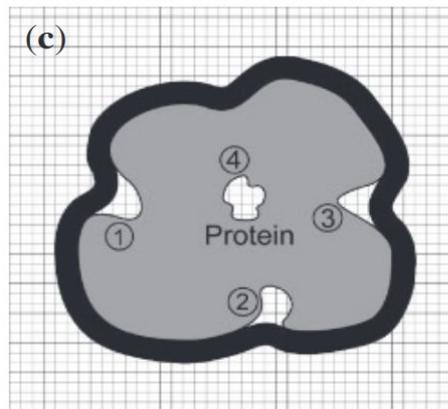
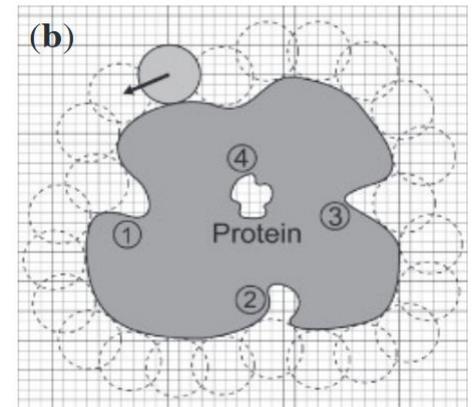
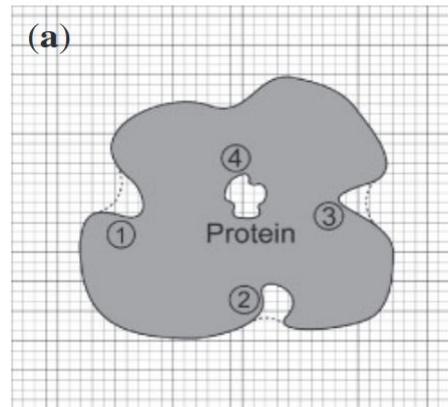
В каждую вершину помещается пробная частица

Для этой частицы вычисляются “скоры” взаимодействий с белком (электростатические, VdW)

Способы определения карманов связывания

Алгоритмы, замощающие пространство сферами. POCASA

- Схоже с построением SAS
- Варьирование радиуса пробы приводит к различным результатам



Способы определения карманов связывания

Алгоритмы, замощающие пространство сферами.

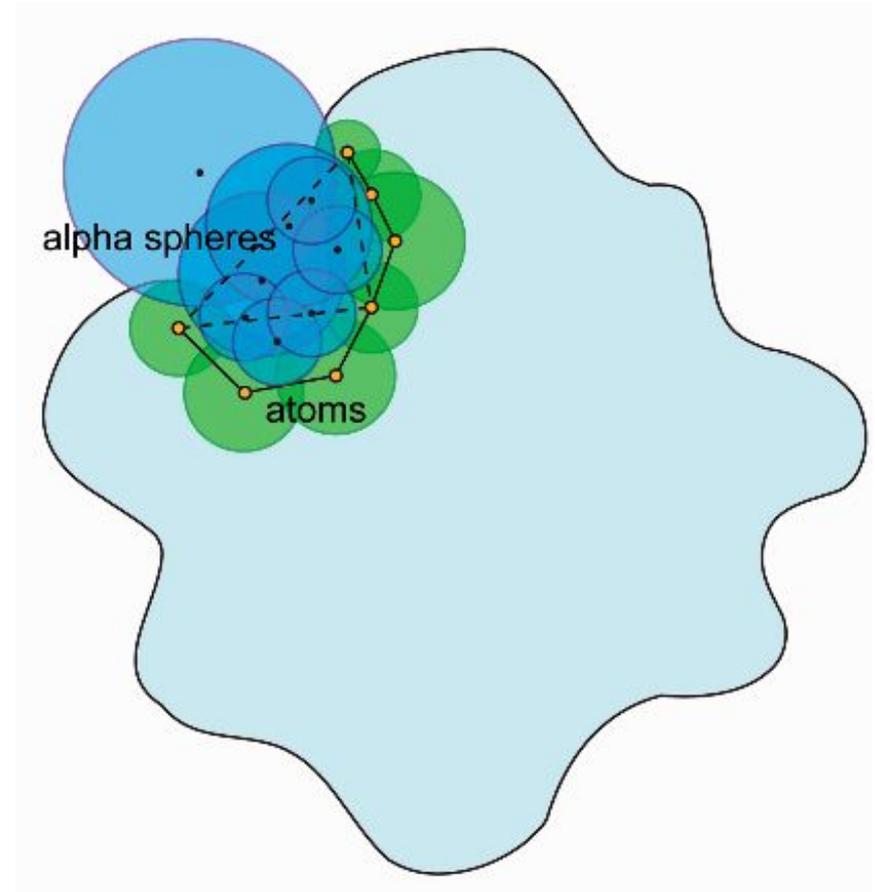
Примеры программ, реализующих данный метод:

- SURFNET
- PASS
- PHECOM
- KVFinder
- GHECOM
- SCREEN
- POCASA

Способы определения карманов связывания

Алгоритмы на основе альфа-сфер. Focket

- Альфа-сфера касается белка в 4 точках и характеризует локальную вогнутость
- Выполняется построение всех альфа-сфер и фильтрация – нас не интересуют очень маленькие и очень большие сферы
- Кластеризация сфер



Способы определения карманов связывания

Алгоритмы на основе альфа-сфер

Примеры программ, реализующих данный метод:

- FindSurf
- CAST
- APROPOS
- Fpocket
- SiteFinder

Способы определения карманов связывания

Алгоритмы на основе глубокого обучения

Kalasanty

Обучающая выборка – набор 3Д изображений, воксель имеет значение 1, если находится внутри кармана, 0, если снаружи.

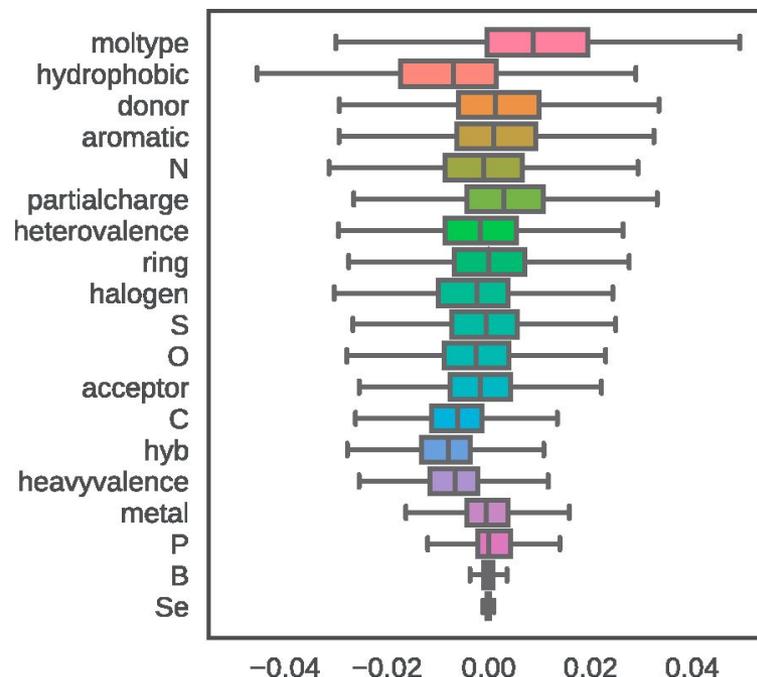
На входе – набор из 18 3Д каналов, содержащих информацию о различных физико-химических скорях (например, канал углерода содержит 1 в ячейке, куда попадает центр какого-либо атома углерода).

Похожий подход использует DeepSite

Способы определения карманов связывания

Features для описания атома:

- 1) 1-9 features (типы атомов)
B, C, N, O, P, S, Se, halogen, metal
значения: 0 или 1
- 2) 10 feature (*hyb*)
значения: 1, 2, 3
- 3) 11 feature (*heavyvalence*)
значения: 1 целое число, количество связей с другими тяжелыми атомами
- 4) 12 feature (*heterovalence*)
значения: 1 целое число, количество связей с другими типами гетероатомов
- 5) 13-17 features
hydrophobic, aromatic, acceptor, donor, ring
значения: 0 или 1
- 6) 18 feature (*partialcharge*)
значения: 1 число с плавающей точкой
- 7) 19 feature (*moltype*)
значения: -1 (для лиганда) или 1 (для белка)



Валидация определенных полостей

Данные, которые использовались для валидации в статьях:

- MetaPocket ([Zhang et al., 2011](#))
- DUD-E ([Mysinger et al., 2012](#))
- База данных scPDB ([Desaphy et al., 2015](#))
- PDBbind ([Liu et al., 2015](#))
- GPCR ([Marchand et al., 2020](#))

Все перечисленные датасеты доступны в [GitHub-репозитории](#)

Возможные критерии успеха при идентификации полостей:

- Наличие атомов лиганда в пределах 4 Å от геометрического центра определенной полости
- Непосредственное сравнение с данными, приведенными в литературе (подходит для небольшого набора данных)

Опыт практического применения

prankweb - не работает

deepsite - просто выдает сайт. И все.

cavityplus - медленно, непрозрачно, нет объемов

castp - неправильные полости

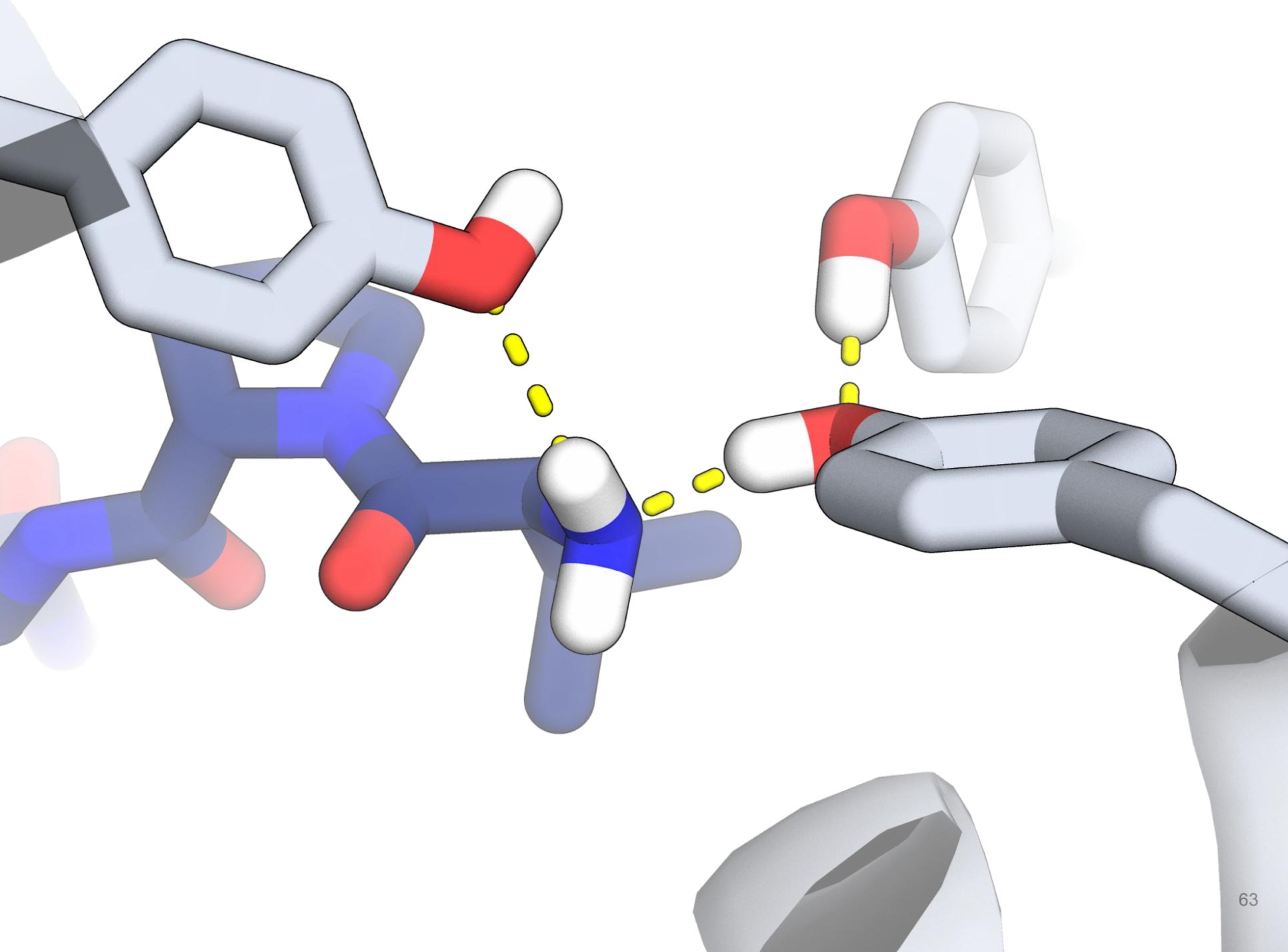
depth - неправильные полости

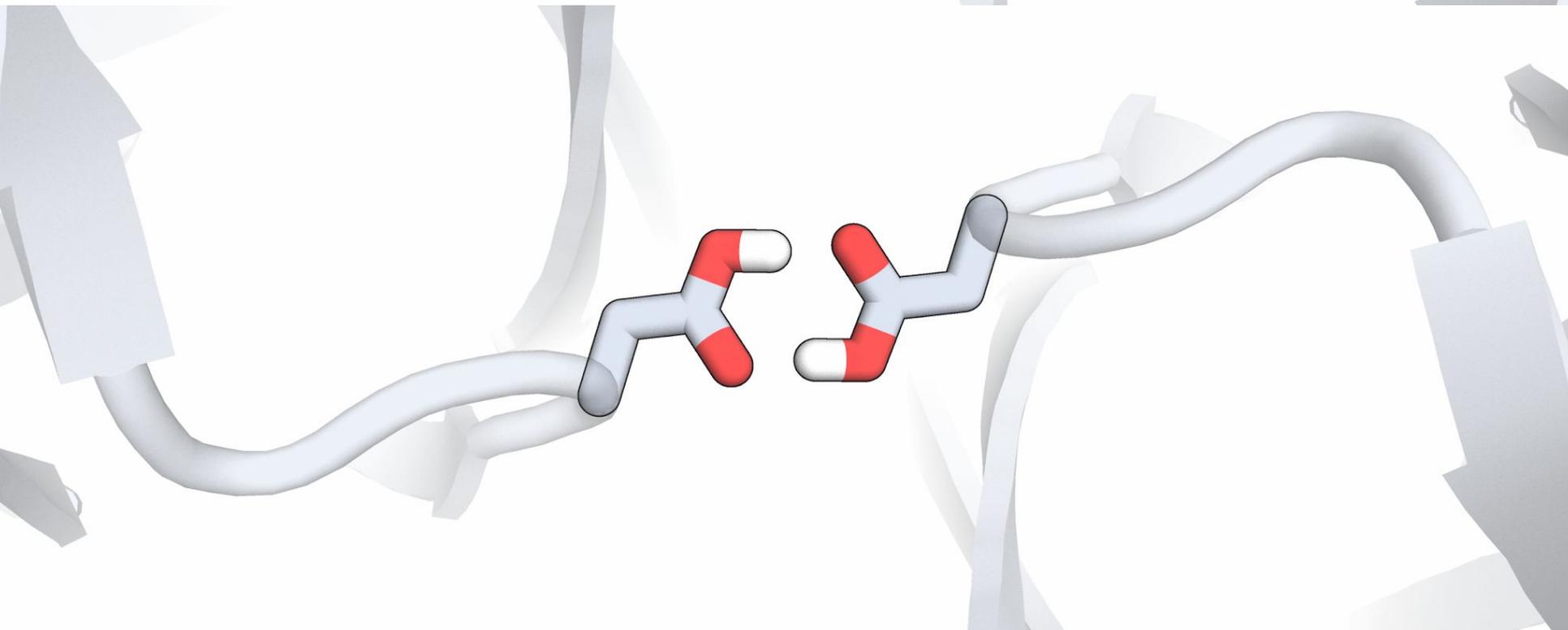
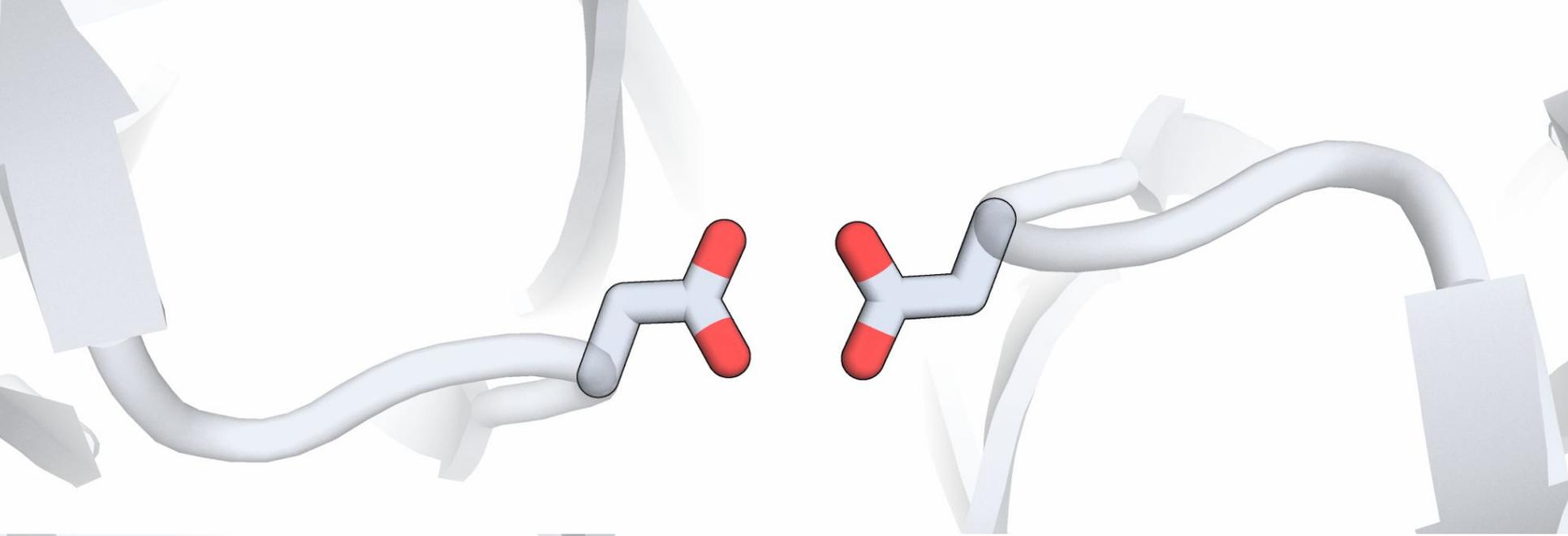
По удобности использования и содержанию выдачи на 2020 год все еще непревзойденным остается fprocket (2009).

Предрасчитанные карманы: Castp, НКPocket

Бенчмарки: CavBench

Протомеры





Предсказание pKa в белках

Мы знаем pKa титрабельных групп в растворе:

Arg 12.48

Lys 10.53

Tyr 10.07

Cys 8.18

Asp 3.65

Glu 4.25

His 6.00

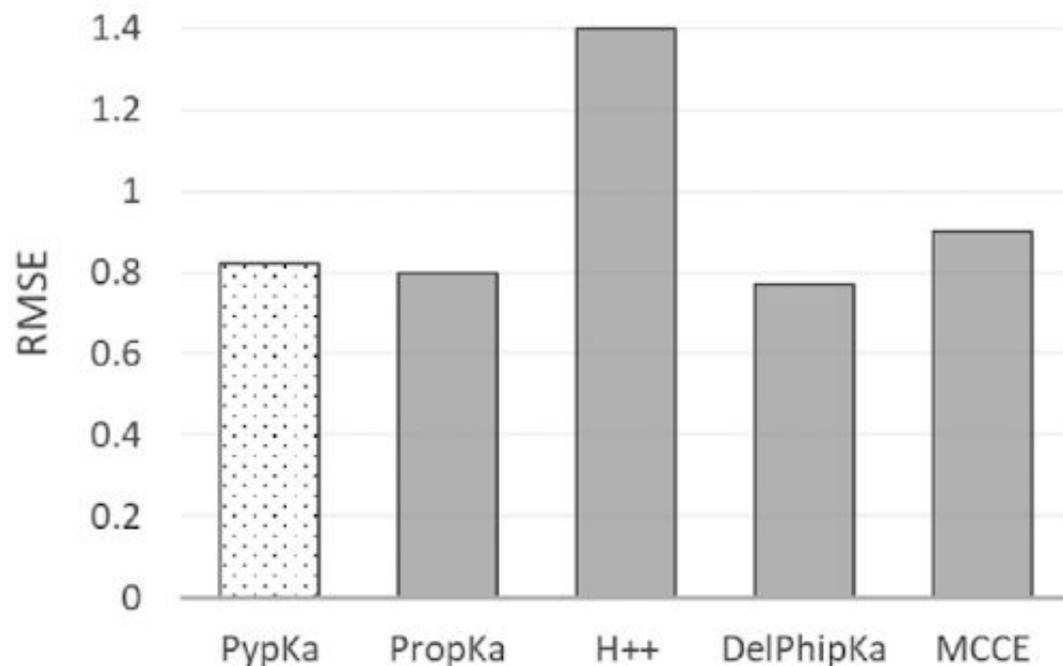
А насколько они вообще могут отличаться внутри белка? [PKAD DB](#)

Методы предсказания pKa остатков в белке

Основанные на уравнении Пуассона-Больцмана (pypka, delphipka)

Основанные на молекулярном моделировании (CpHMD)

Эмпирические (propka)

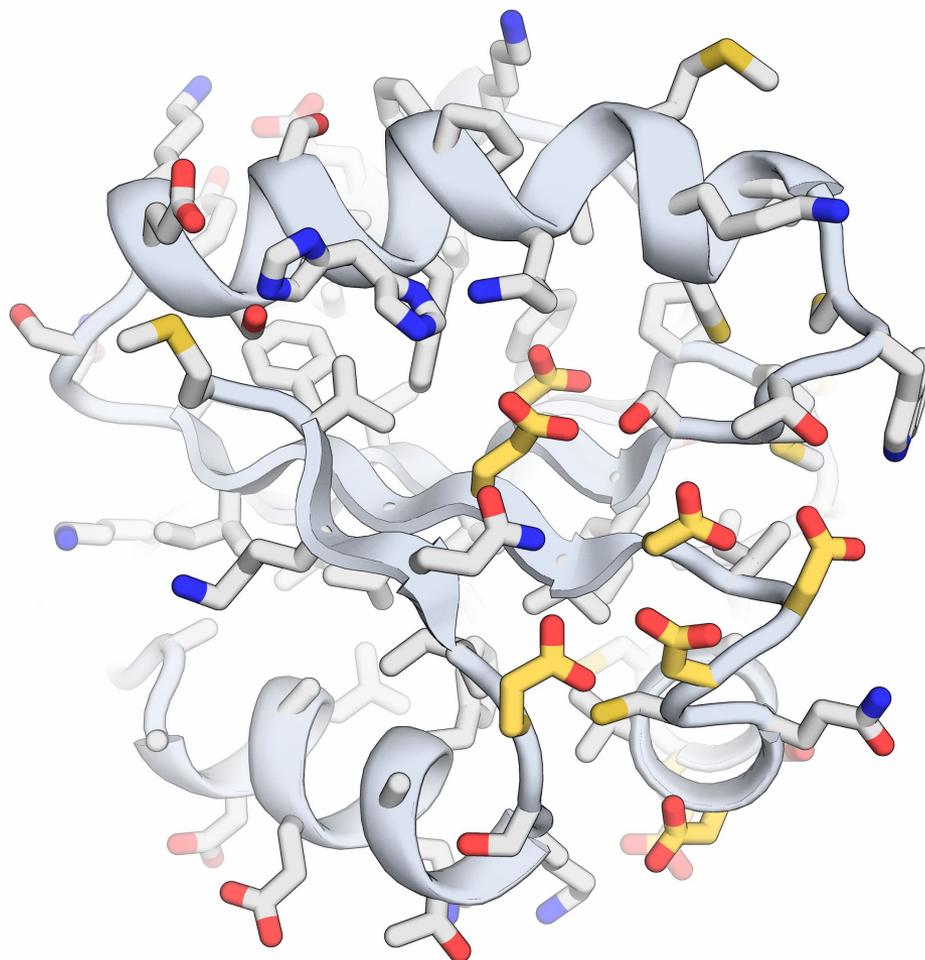


PKAD DB

Около 1500 остатков с экспериментально измеренными рКа (в основном ЯМР небольших белков)

Большинство значений не сильно отличается от референсных для раствора

Для большинства представленных белков протомер остатка не влияет на структуру и функцию



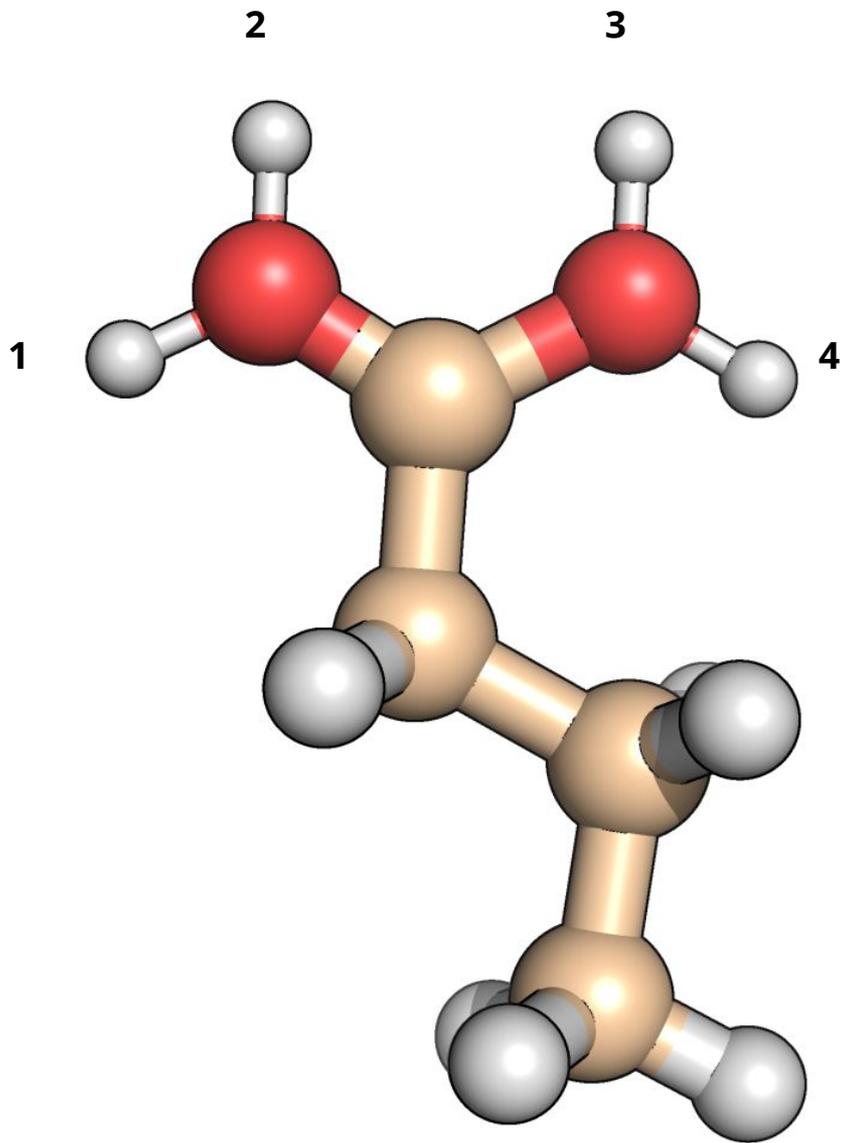
Методы предсказания pKa остатков в белке

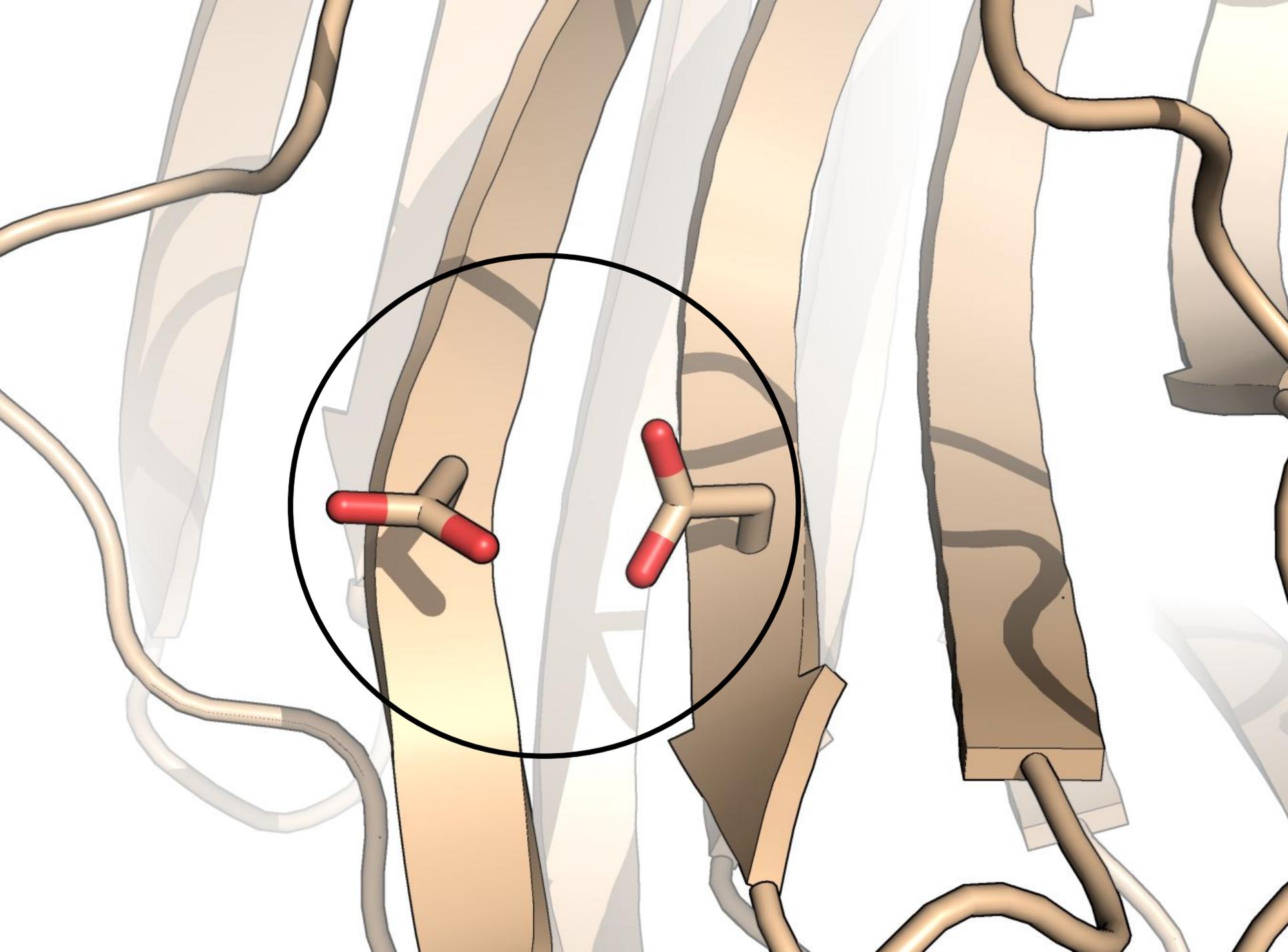
Имеют смысл только в ситуации, когда остаток действительно имеет возможность динамически протонироваться/депротонироваться в своем положении.

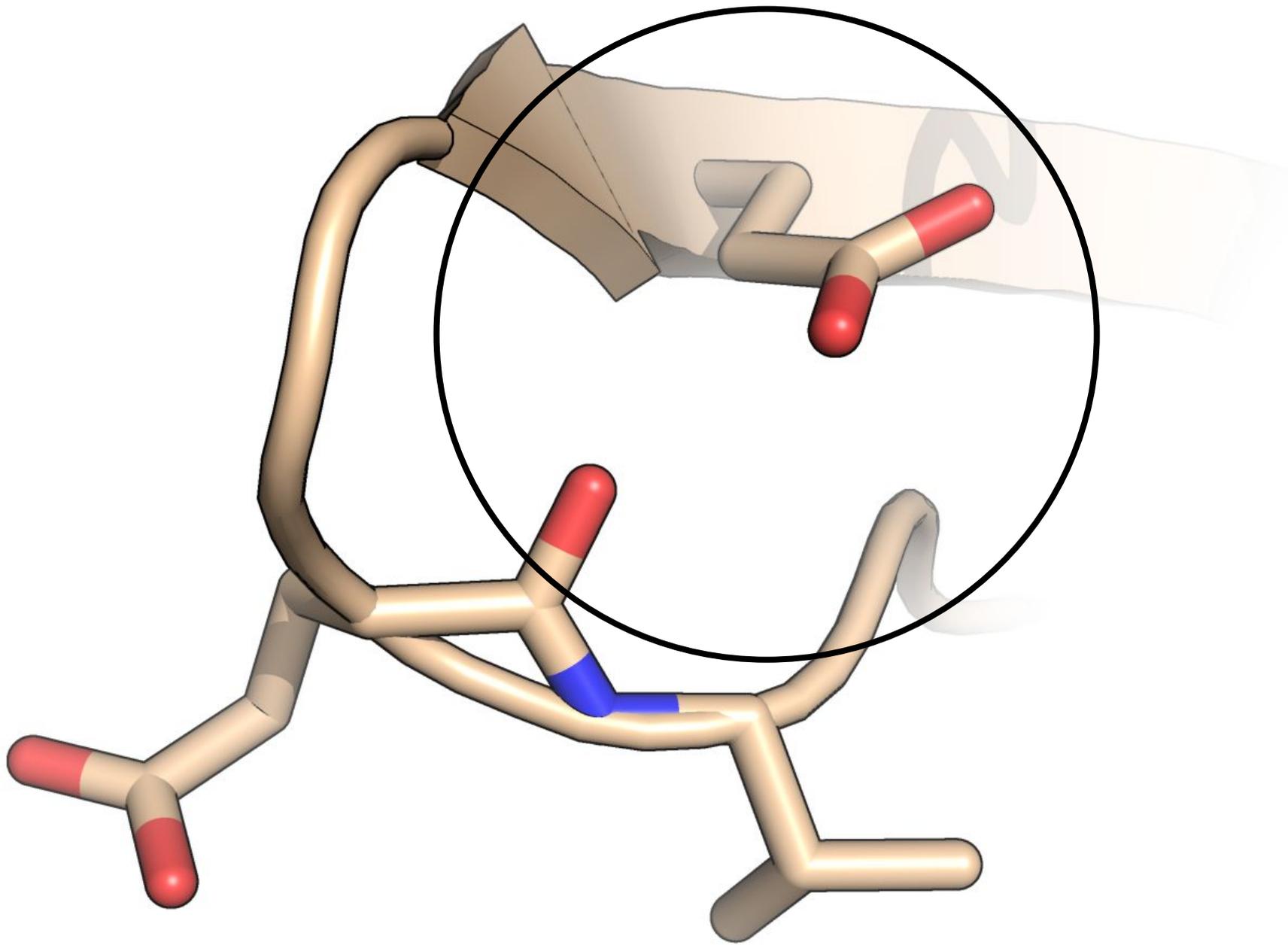
Используйте **propka** или аналоги.

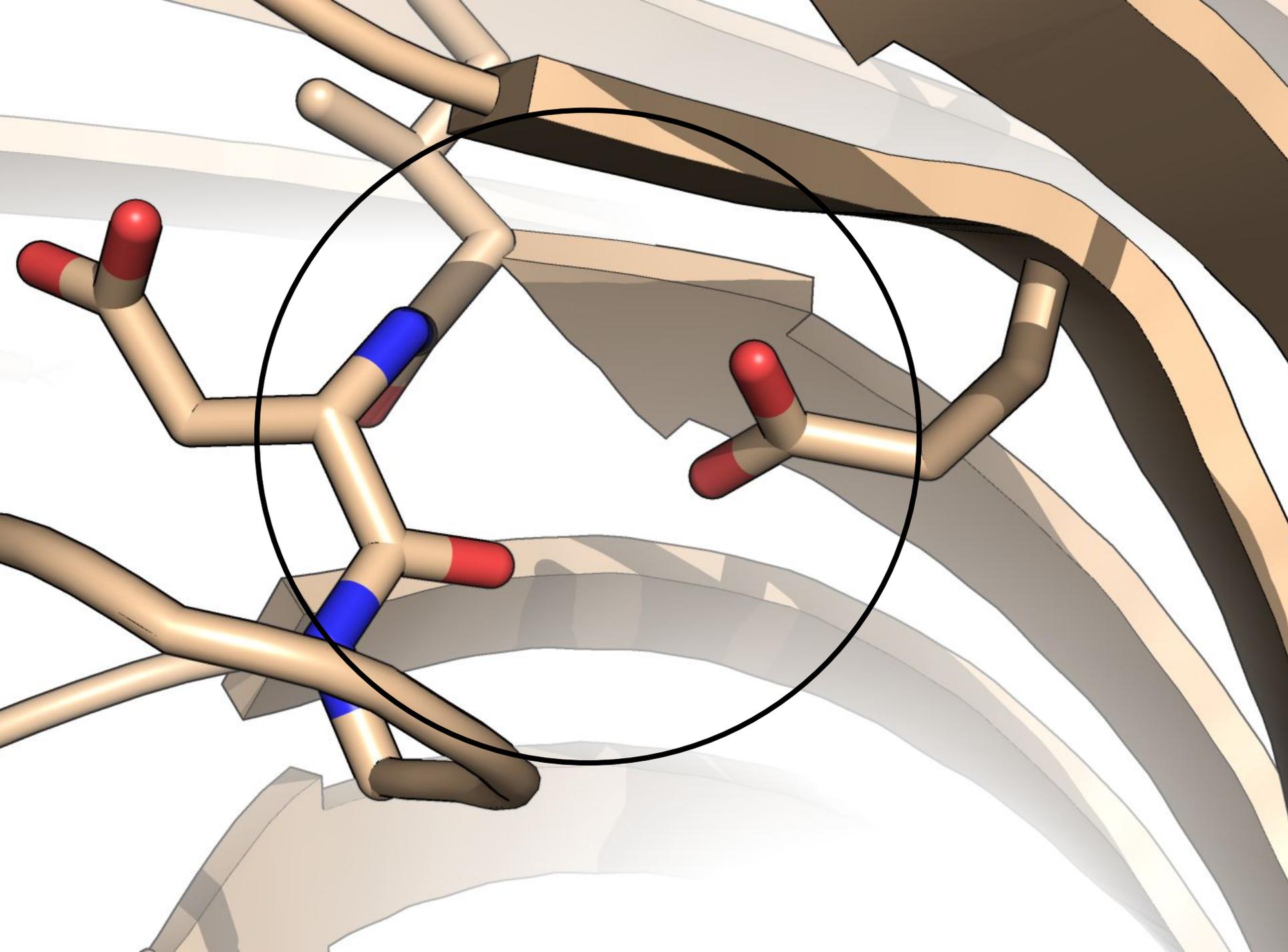
В противном случае мы решаем другую задачу – оценку правдоподобия каждого протомера при условии данных координат. Стоит использовать **molprobity** или аналоги.

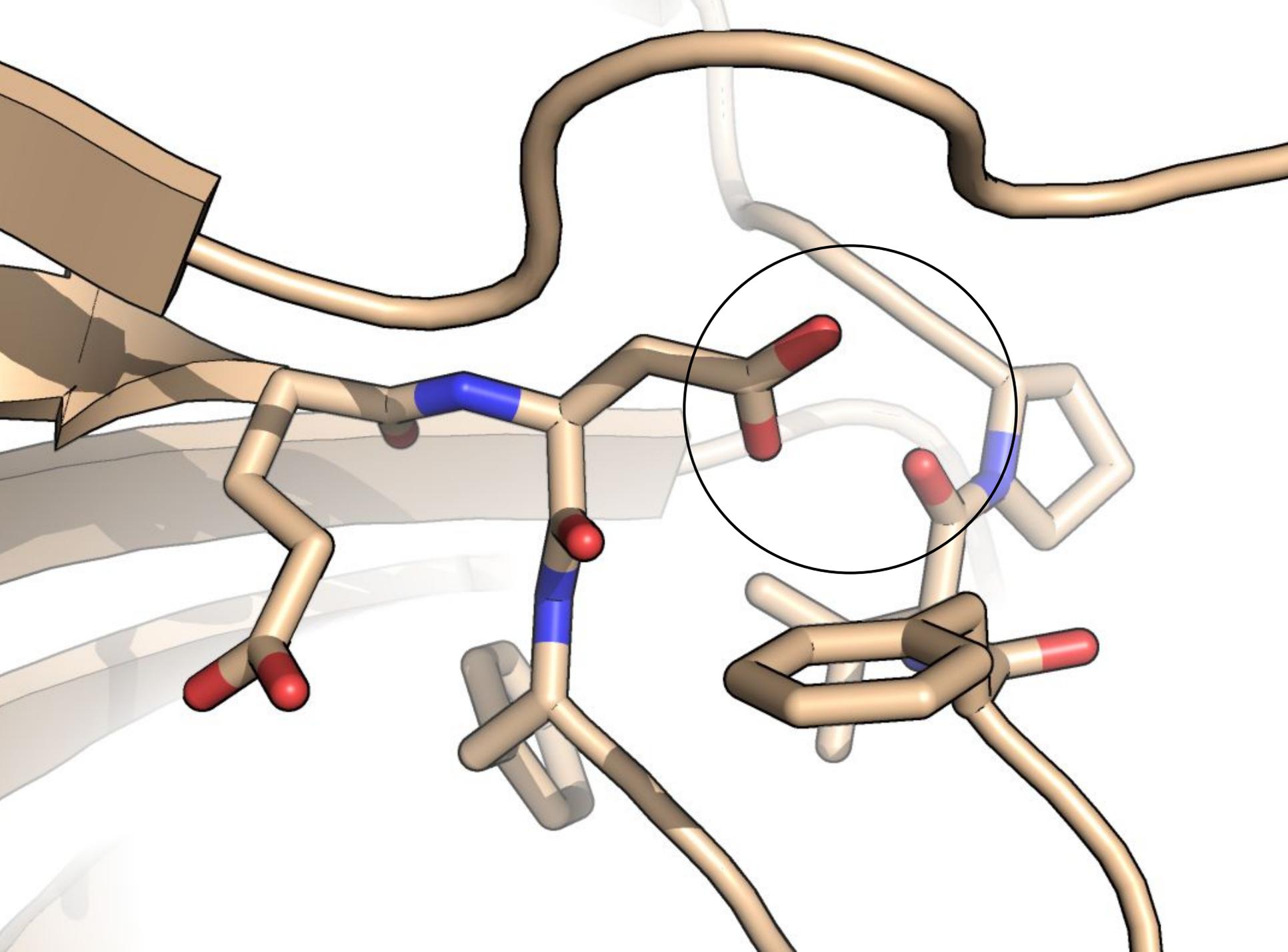
Всегда проверяйте все предсказания вручную!











> [Chembiochem](#). 2017 Apr 4;18(7):666-675. doi: 10.1002/cbic.201600646. Epub 2017 Feb 23.

On the Influence of the Protonation States of Active Site Residues on AChE Reactivation: A QM/MM Approach

Thomas Driant ¹, Florian Nachon ², Cyril Ollivier ¹, Pierre-Yves Renard ³, Etienne Derat ¹

Propka

$$pK_{a,i}^{protein} = pK_{a,i}^{water} + \Delta pK_{a,i}^{water \rightarrow protein}$$

$$\Delta pK_{a,i} = \Delta pK_{a,i}^{Self} + \Delta pK_{a,i}^{QQ}$$

$$\Delta pK_{a,i}^{Self} = \Delta pK_{a,i}^{Desolv} + \Delta pK_{a,i}^{Q\mu}$$

$$\Delta p K_{a,i}^{QQ}$$

Вообще говоря, электростатические взаимодействия стоит рассматривать

а) не между точечными зарядами, а между регионами “зарядовой плотности”

б) с переменной диэлектрической проницаемостью

Эту задачу решает уравнение Пуассона-Больцмана

В **propka** а) игнорируется, а б) аппроксимируется:

$$\varepsilon = \varepsilon_{\text{surface}} - (\varepsilon_{\text{surface}} - \varepsilon_{\text{buried}}) \cdot w_{\text{pair}}(N)$$

$$w(N) = \begin{cases} 0 & \text{if } N \leq N_{\min} \\ \frac{N - N_{\min}}{N_{\max} - N_{\min}} & \text{if } N_{\min} < N < N_{\max} \\ 1 & \text{if } N \geq N_{\max} \end{cases}$$

N – число тяжелых атомов в радиусе 15

$$\Delta p K_{a,i}^{Q\mu} \simeq \Delta p K_{a,i}^{\text{HB}} + \Delta p K_{a,i}^{\text{RE}}$$