

Крупноблочные модели белка: Решетчатые модели

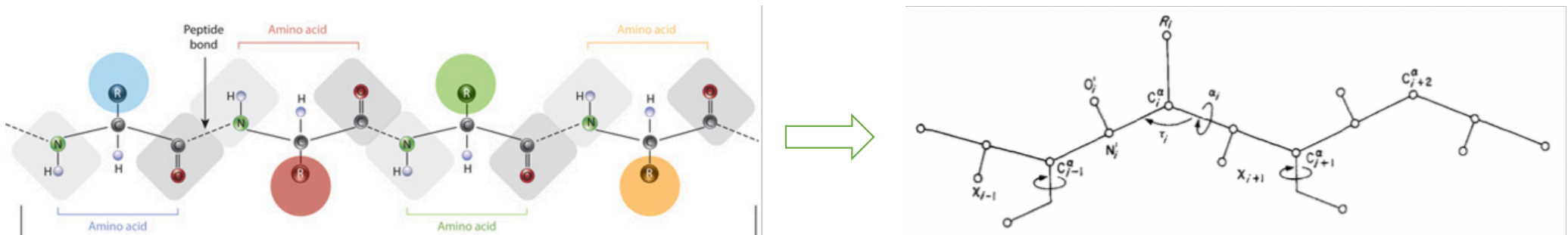
Игнатов Андрей Дмитриевич

aignatov@hse.ru

t.me/a_whynot

Крупноблочные модели белковой геометрии

- Геометрия белка очень сложна, поскольку подразумевает большое число степеней свободы.
- Для упрощения геометрии предлагаются крупноблочные модели, которые подразумевают обобщение групп атомов:



- Обобщение может происходить на разных уровнях.

Крупноблочные модели белковой геометрии

- Варианты обобщения:
 - Вся боковая цепочка становится одним обобщенным атомом
 - Вся пептидная группа становится обобщенным атомом
 - ...
 - Весь **остаток** становится обобщенным атомом!
 - См. DEMO 😊

Крупноблочные модели белковой геометрии

- Вдобавок к сильному упрощению геометрии скажем, что мономеры могут быть только двух видов – гидрофобные (H) или полярные (P).
 - Получившиеся мономеры соединены жесткими связями единичной длины вдоль цепочки.

Крупноблочные модели белковой геометрии

- Вдобавок к сильному упрощению геометрии скажем, что мономеры могут быть только двух видов – гидрофобные (H) или полярные (P).
 - Получившиеся мономеры соединены жесткими связями единичной длины вдоль цепочки.
 - **Вопрос для обсуждения:** а может ли такая модель иметь хоть что-то общее с реальным миром? 🤔

Крупноблочные модели белковой геометрии

- Вдобавок к сильному упрощению геометрии скажем, что мономеры могут быть только двух видов – гидрофобные (Н) или полярные (Р).
 - Получившиеся мономеры соединены жесткими связями единичной длины вдоль цепочки.
 - **Вопрос для обсуждения:** а может ли такая модель иметь хоть что-то общее с реальным миром? 🤔
Может и имеет! Гидропатия – один из главных движущих механизмов фолдинга белка.

Крупноблочные модели белковой геометрии

- Упростим модель еще больше!

- Пусть теперь задана некоторая дискретная решетка с фиксированным шагом.

- Все мономеры могут находиться только в узлах этой решетки.

- Если говорить формально:

- **Решетка L** – счетное множество точек такое, что $u + v \in L$ для любых $u, v \in L$.

- Простой способ задания решетки – минимальный набор векторов N_L , с помощью которого можно выразить любой узел решетки. Любой узел u решетки – линейная комбинация векторов из N_L с неотрицательными целыми коэффициентами:

$$L = \{u \in \mathbb{R}^2 \mid u = \sum_{v \in N_L} c_v \cdot v, c_v \in \mathbb{Z}^+\}$$

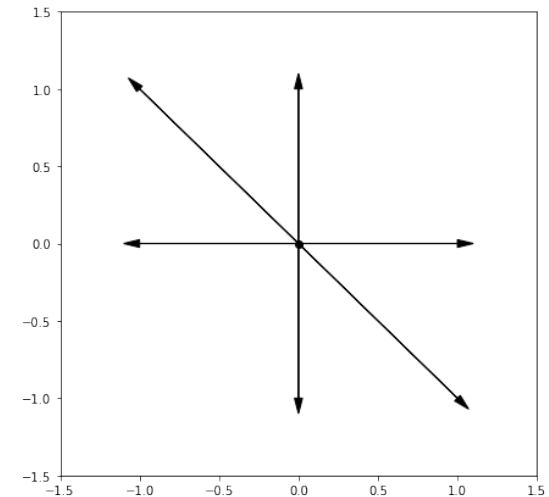


Table 1. N_L sets for squared and triangular 2D lattices

Lattice	Square	Triangular
N_L vectors	(0,1), (0,-1), (1,0), (-1,0)	(1,0), (-1,0), (0,1), (0,-1), (-1,1), (1,-1)

Крупноблочные модели белковой геометрии

• Упростим модель еще больше!

Это
HP-модель!

- Пусть теперь задана некоторая дискретная решетка с фиксированным шагом.
- Все мономеры могут находиться только в узлах этой решетки.
- Если говорить формально:
 - **Решетка L** – счетное множество точек такое, что $u + v \in L$ для любых $u, v \in L$.
 - Простой способ задания решетки – минимальный набор векторов N_L , с помощью которого можно выразить любой узел решетки. Любой узел u решетки – линейная комбинация векторов из N_L с неотрицательными целыми коэффициентами:

$$L = \{u \in \mathbb{R}^2 \mid u = \sum_{v \in N_L} c_v \cdot v, c_v \in \mathbb{Z}^+\}$$

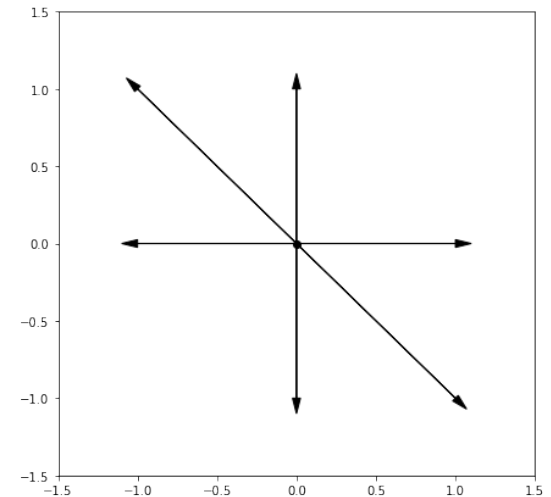


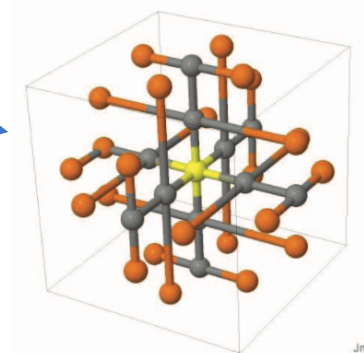
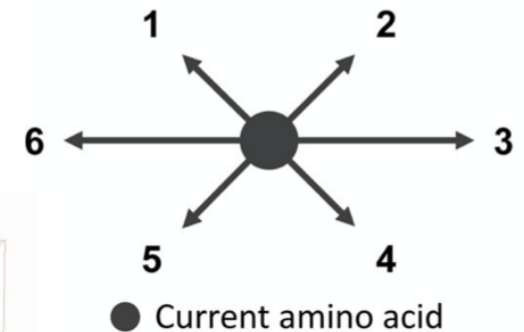
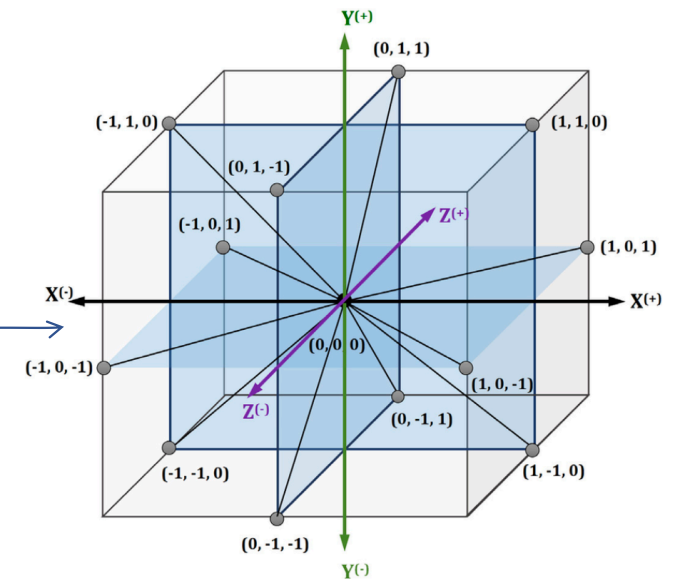
Table 1. N_L sets for squared and triangular 2D lattices

Lattice	Square	Triangular
N_L vectors	(0,1), (0,-1), (1,0), (-1,0)	(1,0), (-1,0), (0,1), (0,-1), (-1,1), (1,-1)

А какие бывают решетки?

- Lattices:

- Cubic;
- Face-centered-cubic (FCC);
- 2D-Triangular
- 2D-Square
- Knight's Walk (KW)
- Diamond.
- ...



А какие бывают решетки?

- Lattices:

- Cubic;

- Face-centered-cubic (FCC);

- 2D-Triangular

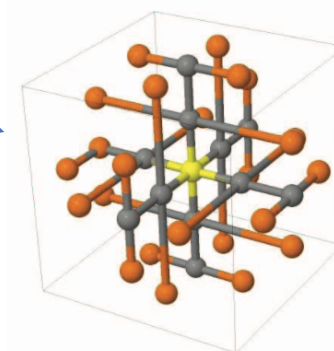
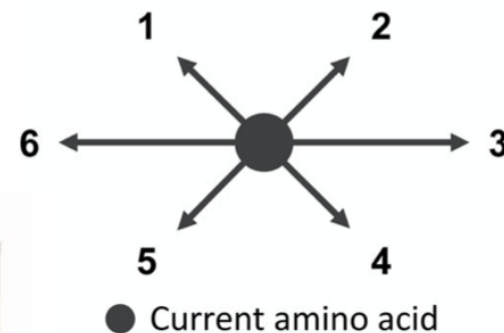
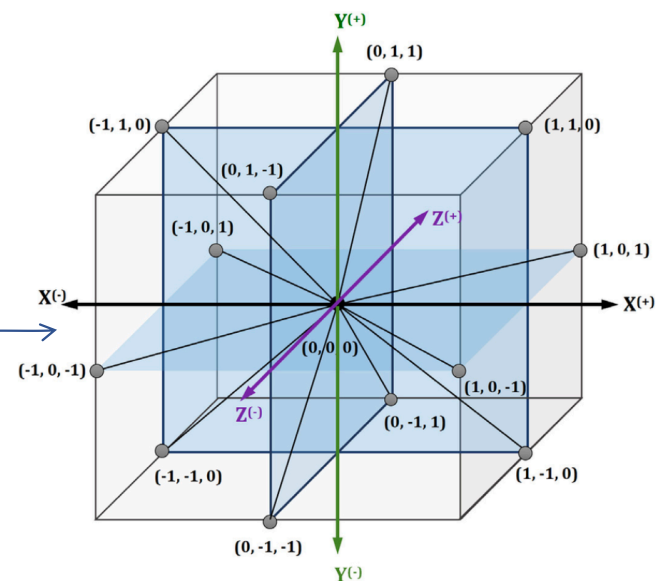
- 2D-Square

- Knight's Walk (KW)

- Diamond.

- ...

- **Вопрос для обсуждения:**
а все ли решетки хороши?



А какие бывают решетки?

- Lattices:

- Cubic;

- Face-centered-cubic (FCC);

- 2D-Triangular

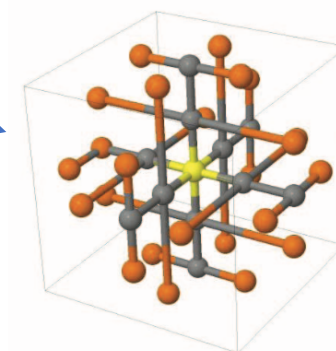
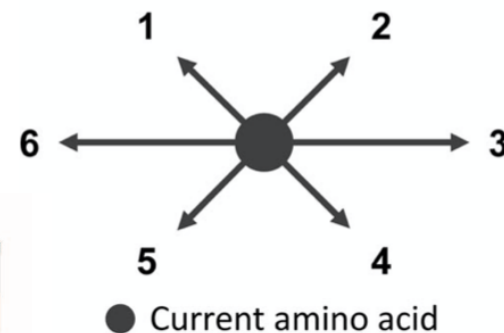
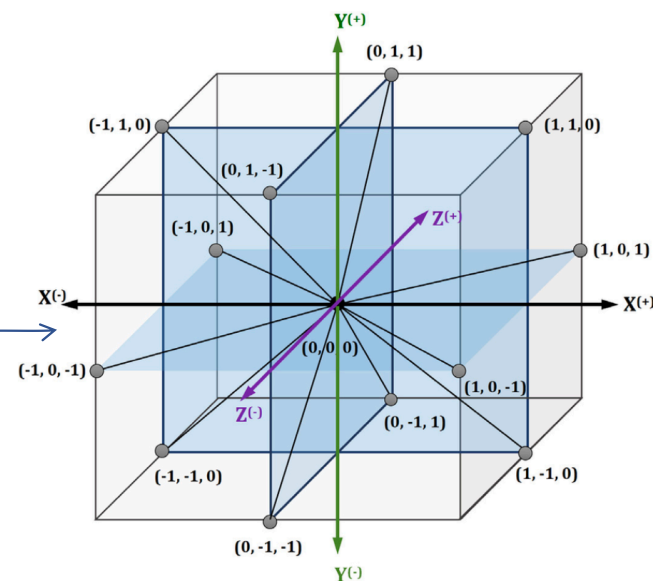
- 2D-Square

- Knight's Walk (KW)

- Diamond.

- ...

- **Вопрос для обсуждения:**
а все ли решетки хороши?
Нет!



Постановка задачи: *HP*-модель (2D)

- Пусть $M = \{M_1, \dots, M_{n_S}\}$ – **размещение (mapping)**, в котором $M_i = (x_i, y_i)$ – позиция мономера S_i . Тогда целевая функция:

- $E_S(M) = \sum_{1 \leq i+1 < j \leq n_S} \Delta(M_i, M_j) \cdot e(S_i, S_j)$, где:

- $\Delta(M_i, M_j) = \begin{cases} 1, & \text{if } M_i - M_j \in N_L \\ 0, & \text{иначе} \end{cases}$
- $e(S_i, S_j) = \begin{cases} -1, & \text{if } S_i = \text{'H'} \text{ и } S_j = \text{'H'} \\ 0, & \text{иначе} \end{cases}$

- Тогда задача:

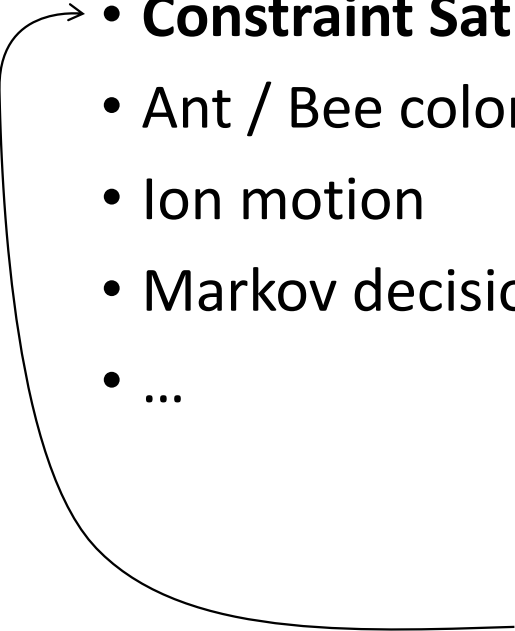
$$\begin{cases} E_S(M) \rightarrow \min \\ M_i - M_{i-1} \in N_L, i = 2, \dots, n_S \\ M_i \neq M_j, i, j = 1, \dots, n_S, i \neq j \end{cases}$$

Как решать такую задачу?

- Constraint Satisfaction (удовлетворение ограничений)
- Ant / Bee colony
- Ion motion
- Markov decision process
- ...

Как решать такую задачу?

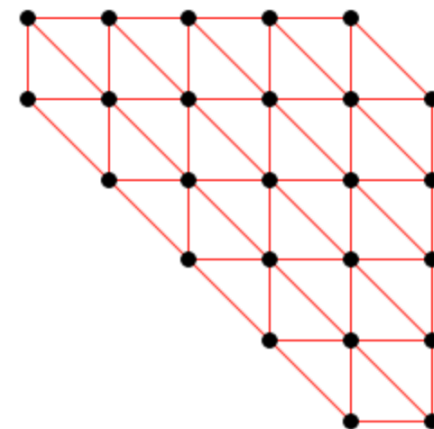
- **Constraint Satisfaction (удовлетворение ограничений)**
- Ant / Bee colony
- Ion motion
- Markov decision process
- ...



Будем работать с этим! Такой подход – единственный **точный** из всех перечисленных.
N.B.: Что значит точный? **С гарантией оптимальности решения!**

Стандартный точный подход¹

- Стандартный точный подход состоит из двух шагов:
 - H -мономеры собираются в максимально плотное H -ядро. P -момеры игнорируются.

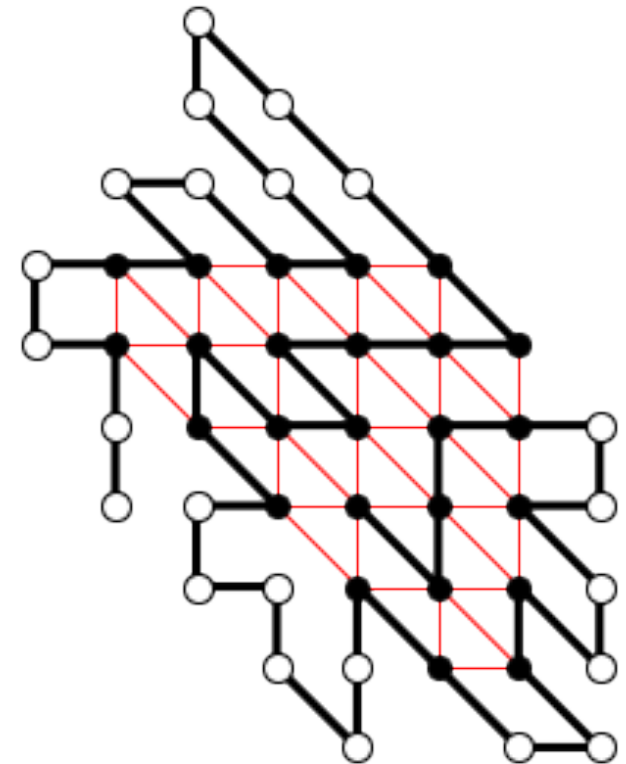


¹ Mann M., Backofen R. Exact methods for lattice protein models
//Bio-Algorithms and Med-Systems. – 2014. – Т. 10. – №. 4. – С. 213-225.

Стандартный точный подход¹

- Стандартный точный подход состоит из двух шагов:

- *H*-мономеры собираются в максимально плотное *H*-ядро. *P*-мономеры игнорируются.
- Производится попытка разместить весь белок в решетке с учетом *H*-ядра.
 - Если попытка успешна, конформация найдена.
 - Иначе проверяется другое *H*-ядро с таким же числом контактов.
 - Если никакое из максимально плотных ядер не подошло, то рассматриваются ядра с уменьшенным числом контактов.



¹ Mann M., Backofen R. Exact methods for lattice protein models
//Bio-Algorithms and Med-Systems. – 2014. – Т. 10. – №. 4. – С. 213-225.

Стандартный точный подход¹

- А имеет ли такой подход смысл?

Да. Утверждается, что у 95% белков есть выраженное гидрофобное ядро.

Н.В.: У некоторых белков гидрофобное ядро может быть не одно! Оно может быть поделено на несколько локальных доменов².

- Какова точность модели?

Как и у всех крупноблочных моделей, точность *HP*-модели ограничена уровнем упрощения геометрии. Но расположение мономеров в узлах решетки вносит дополнительный негативный эффект!

- Зачем же тогда нужна такая модель?

Для создания хорошего начального приближения для применения впоследствии более точных методов.

² Wu, L. C., & Kim, P. S. A specific hydrophobic core in the α -lactalbumin molten globule. *Journal of molecular biology*, 280(1), 175-182. (1998)