

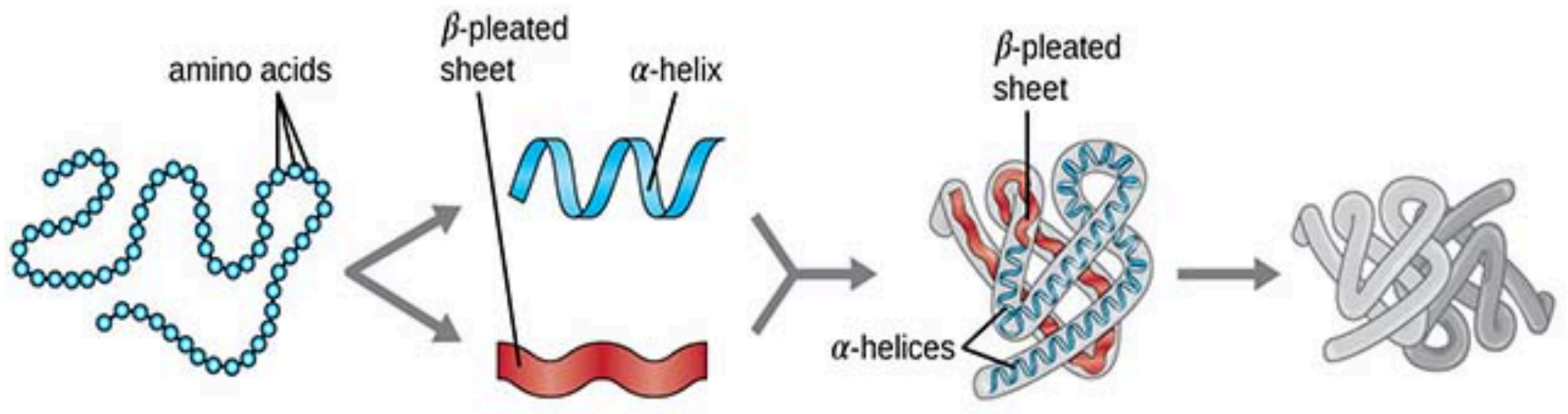
Геометрия белковых молекул

Игнатов Андрей Дмитриевич

aignatov@hse.ru

t.me/a_whynot

Типы геометрических структур белка



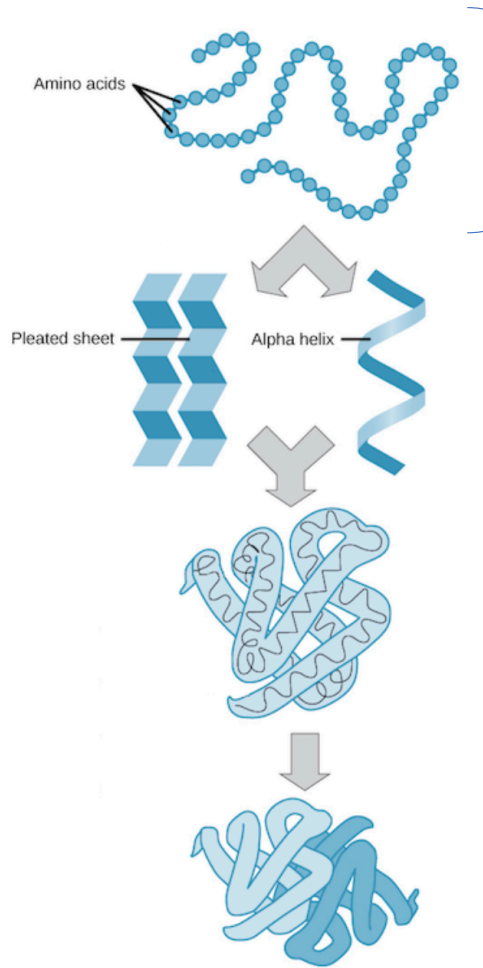
Первичная структура
(последовательность
аминокислот)

Вторичная структура
(регулярные структуры
с характерными
геометрическими
особенностями)

Третичная структура
(свернутая белковая
последовательность).
**Белок в классическом
понимании**

Четвертичная структура
(две совмещенные
полипептидные цепи)

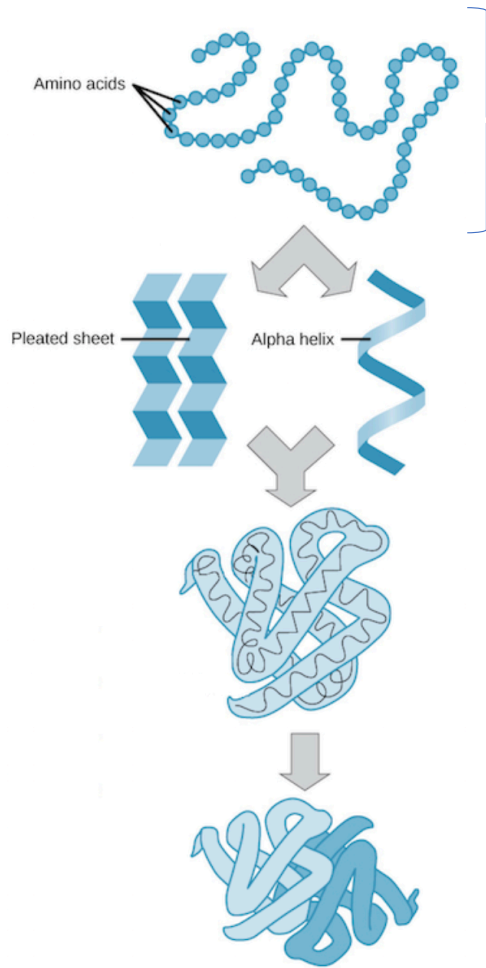
Типы геометрических структур белка



- Первичная структура:

- Состоит из аминокислотных остатков, соединенных пептидными связями.

Типы геометрических структур белка

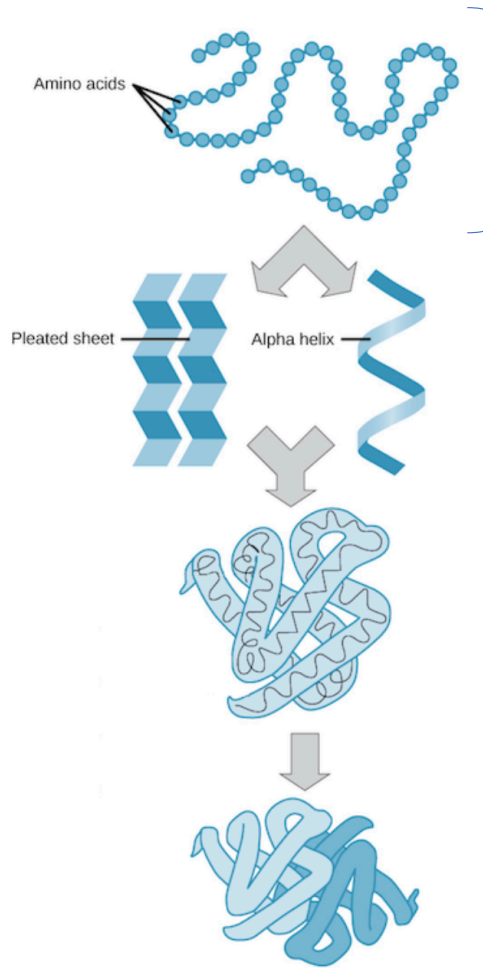


- Первичная структура:

- Состоит из аминокислотных остатков, соединенных пептидными связями.

// Вопрос: сколько пептидных связей в белке длиной N ?

Типы геометрических структур белка



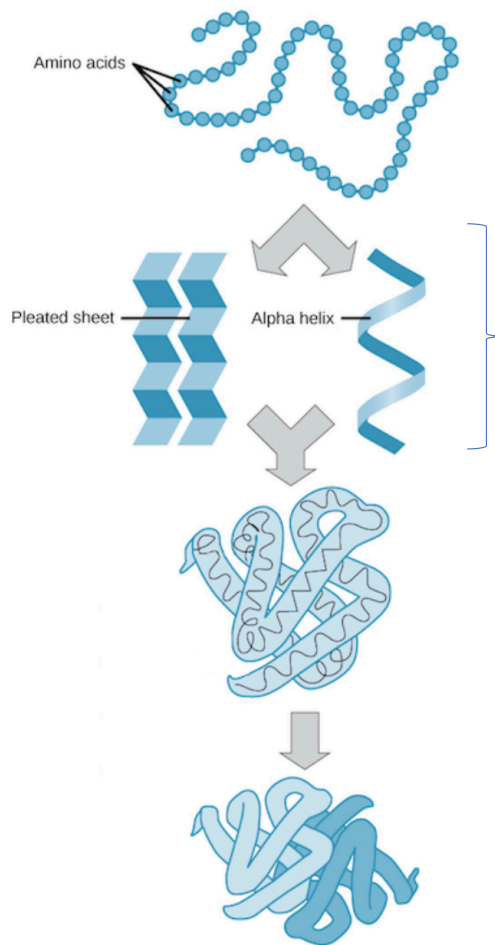
- Первичная структура:

- Состоит из аминокислотных остатков, соединенных пептидными связями.

// Вопрос: сколько пептидных связей в белке длиной N ?

- Так как белок никогда не существует в виде развернутой цепи, первичной структурой можно считать его буквенный код.

Типы геометрических структур белка



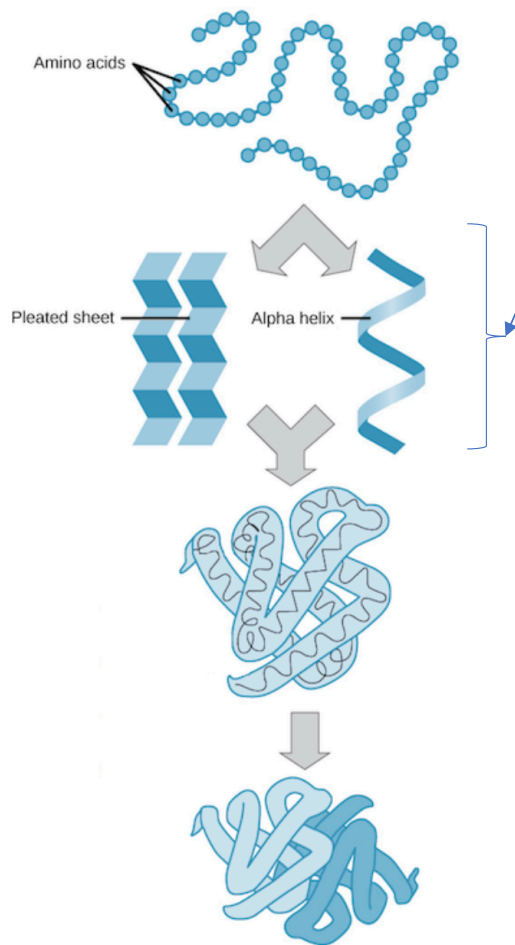
- Вторичная структура:

- Набор регулярных (с повторяющейся геометрией) и нерегулярных структур, в которые объединяются аминокислотные остатки при сворачивании белка.
- Укрепляются водородными связями между определенными атомами каркаса и боковых цепочек.
- Существует 8 типов вторичных структур (по классификации DSSP¹). Наиболее распространенные – α -спираль и β -лист (на рисунке слева).
- DSSP – алгоритм оценки водородных связей.

// Комментарий: DSSP не предсказывает водородные связи, а находит вероятные связи в свернутом белке.

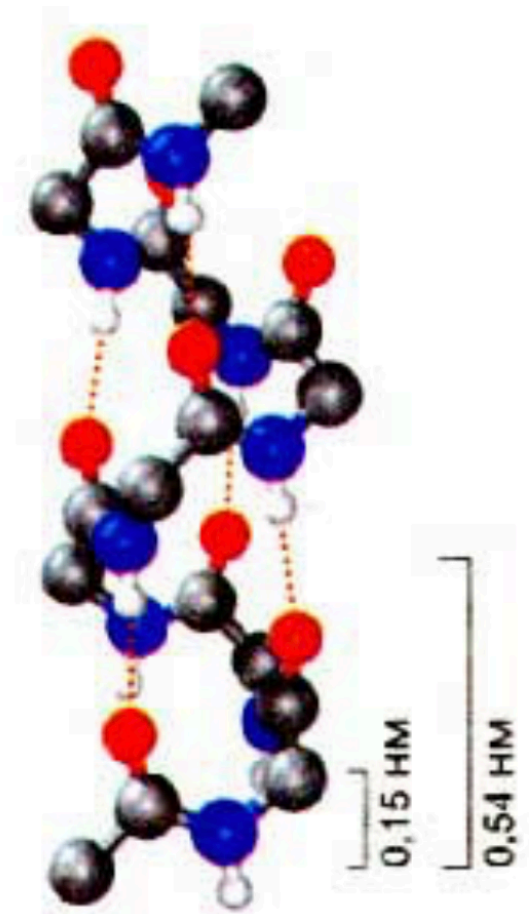
¹ Kabsch, W. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features [текст] / W. Kabsch, C. Sander // Biopolymers: Original Research on Biomolecules. — 1983. — т. 22, No 12. — с. 2577—2637.

Типы геометрических структур белка

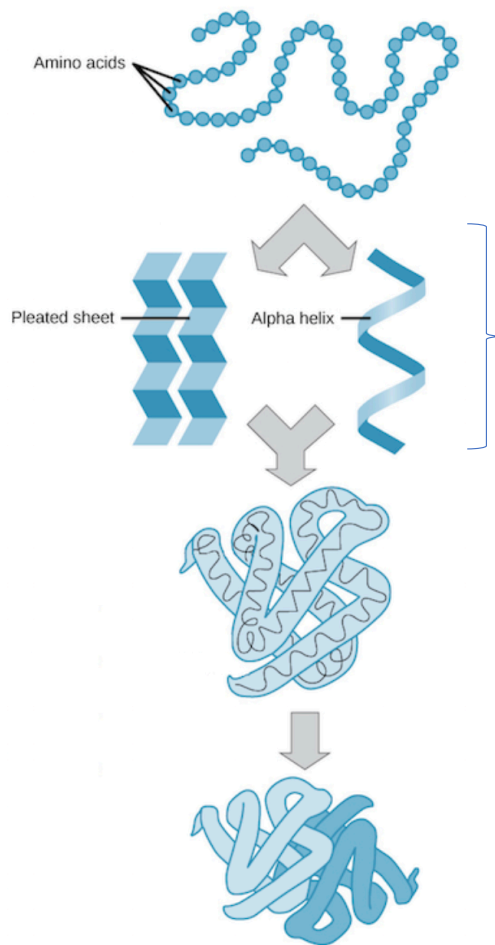


- Типы вторичных структур по DSSP:

- Н (4-12 α -спираль). Правозакрученная, редко левозакрученная спираль; укрепляется водородной связью группы NH i -ого остатка и группой CO ($i + 4$)-ого остатка. Средний диаметр спирали равен 12\AA , что и дало ей название 4-12.

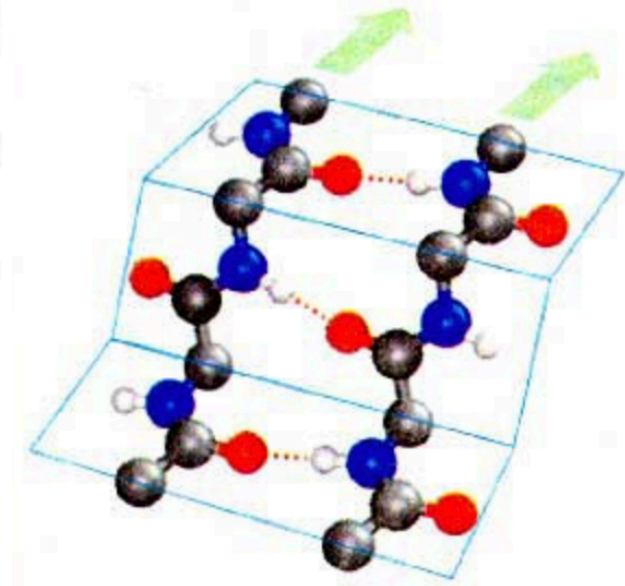
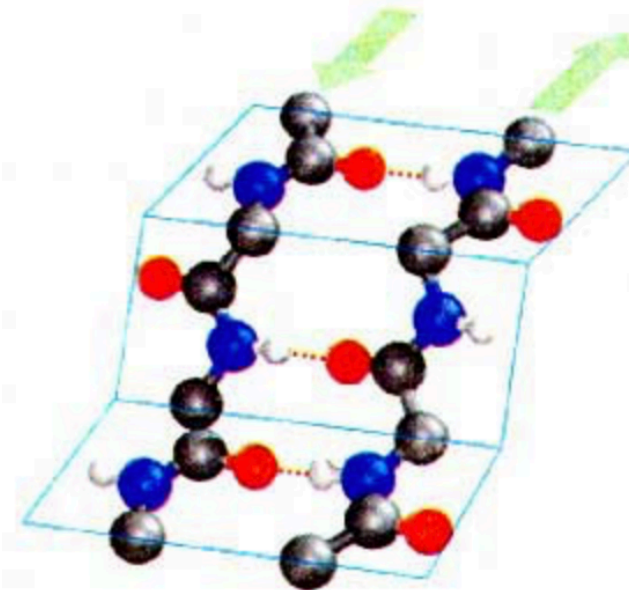


Типы геометрических структур белка

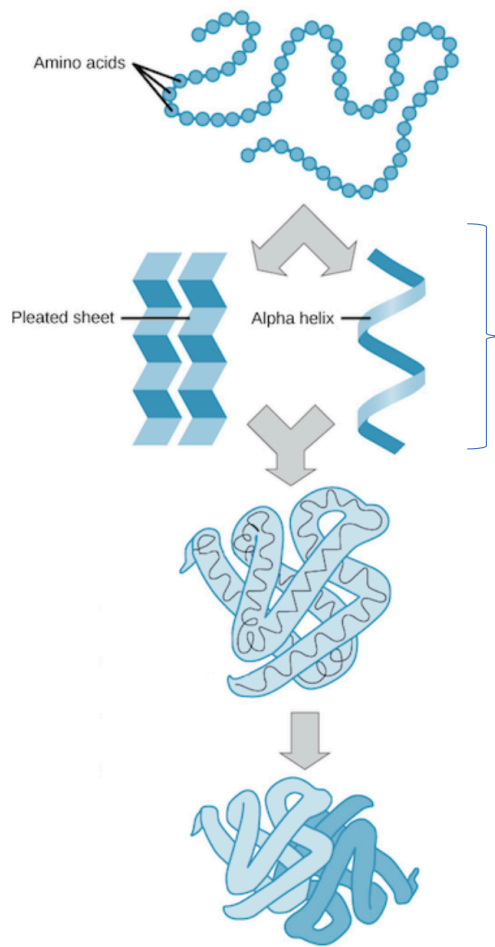


- Типы вторичных структур по DSSP:

- E (Beta-strand). β -цепь; несколько таких цепей, соединенных боковыми водородными связями, образуют β -лист.



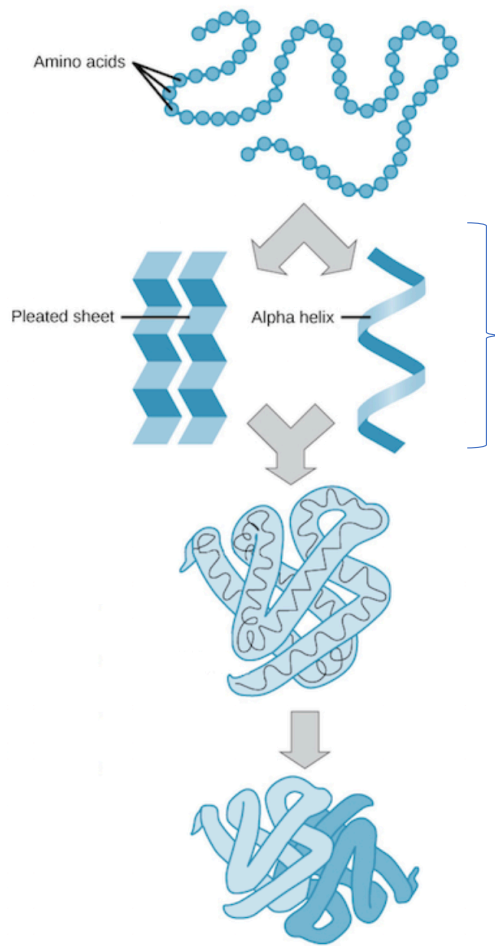
Типы геометрических структур белка



- Типы вторичных структур по DSSP:

- G (3-10 спираль). Геометрия такой спирали аналогична α -спирали, но разница индексов между связанными остатками составляет 3.
- I (π -спираль). Геометрия такой спирали аналогична α -спирали, но разница индексов между связанными остатками составляет 5.
- T (Turn). одиночный разворот цепи, закреплённый водородной связью. Закрепляющая водородная связь может быть через 3, 4 или 5 остатков.

Типы геометрических структур белка

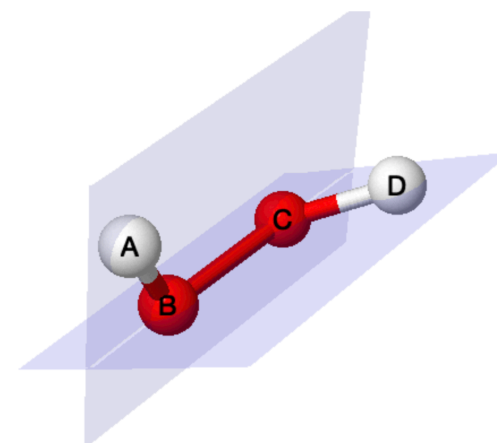
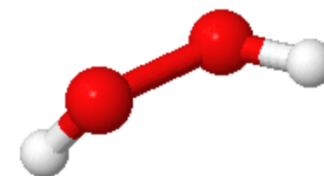


- Типы вторичных структур по DSSP:

- B (Isolated beta-bridge residue) – одиночный остаток, образующий водородную связь с другим аналогичным остатком. Это изолированный аминокислотный остаток, не входящий ни в какие другие регулярные вторичные структуры, и такой тип встречается редко.
- S (Bend) – искривление. Искривление определяется для среднего из пяти остатков, если $\angle\{(C_{\alpha}^i - C_{\alpha}^{i-2}), (C_{\alpha}^{i+2} - C_{\alpha}^i)\} > 70^{\circ}$.
// Комментарий: Это единственный тип вторичной структуры, который не зависит от водородных связей, а является следствием белковой геометрии.

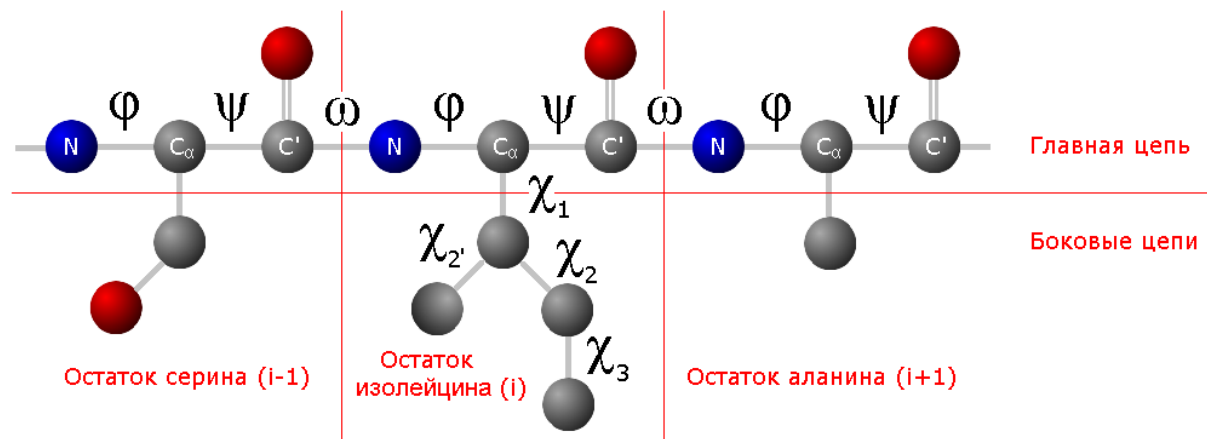
Образование вторичных структур

- Образование вторичных структур происходит при вращении каркаса остатка вокруг своих осей.
- Осями в этом случае являются связи атомов каркаса:
 $(N - C_\alpha), (C_\alpha - C), (C - N)$.
- Однако вращение вокруг таких осей нельзя описать планарными углами: модель получается слишком ограниченной.
- Вращение описывается с помощью двугранных углов (углов между двумя плоскостями).



Двугранные углы каркаса аминокислоты

- Три основных двугранных угла каркаса аминокислотного остатка i :
- $\omega_i = \angle(C_{\alpha}^{i-1}, C^{i-1}, N^i, C_{\alpha}^i)$;
- $\phi_i = \angle(C^{i-1}, N^i, C_{\alpha}^i, C^i)$;
- $\psi_i = \angle(N^i, C_{\alpha}^i, C^i, N^{i+1})$.
- На практике имеет смысл рассматривать только ϕ, ψ .

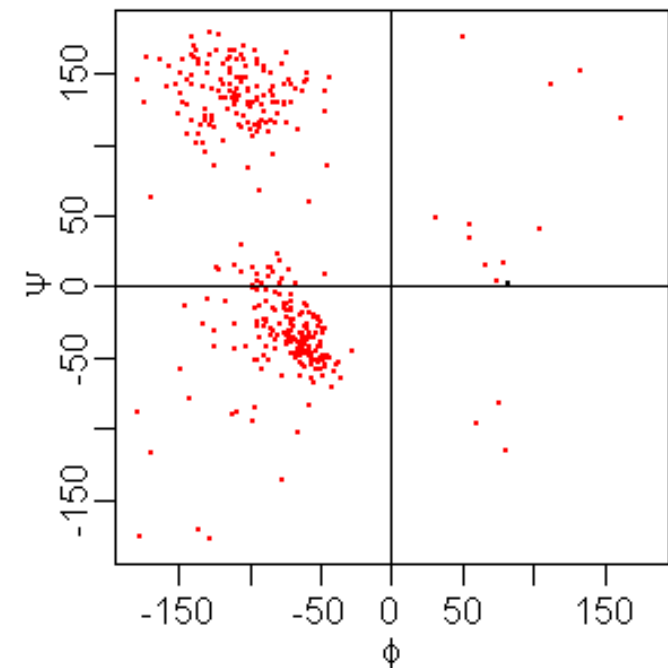


Двугранные углы каркаса аминокислоты

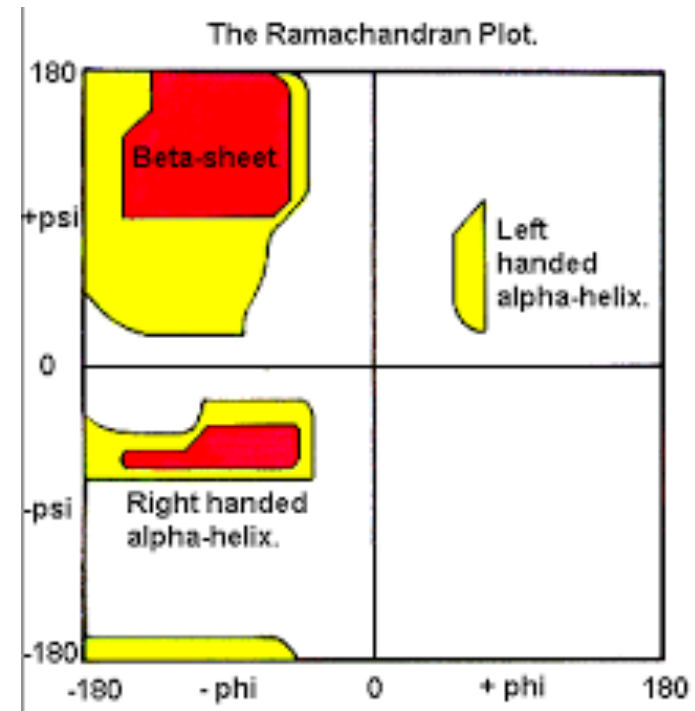
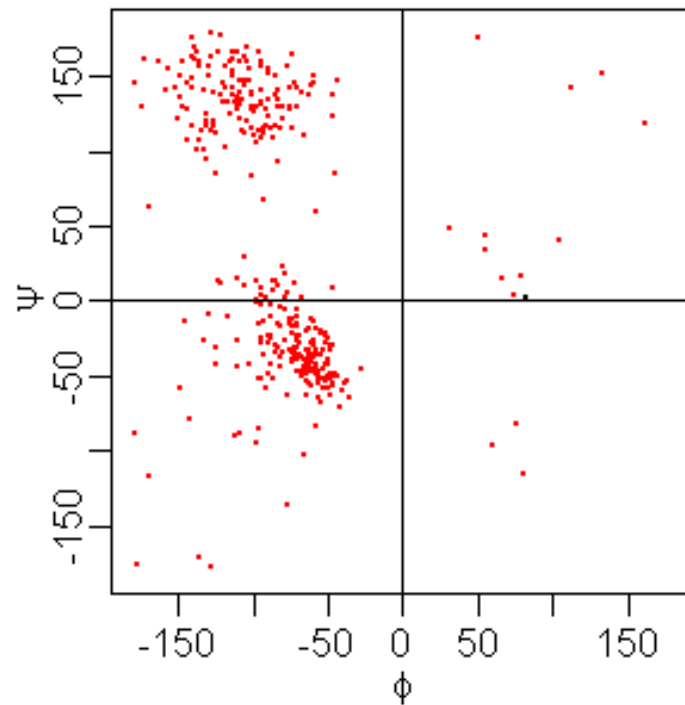
- // Вопрос: сколько всего углов ω , ϕ , ψ есть в белке длиной N ?

Двугранные углы каркаса аминокислоты

- // Вопрос: сколько всего углов ω , ϕ , ψ есть в белке длиной N ?
- В силу геометрической совместности ϕ , ψ не рассматриваются по отдельности, только парами.
- Все пары значений (ϕ, ψ) можно отобразить на диаграмме, называемой **картой Рамачандрана**.
- На ней выделяется несколько характерных областей, которые соответствуют определенным **вторичным структурам**.

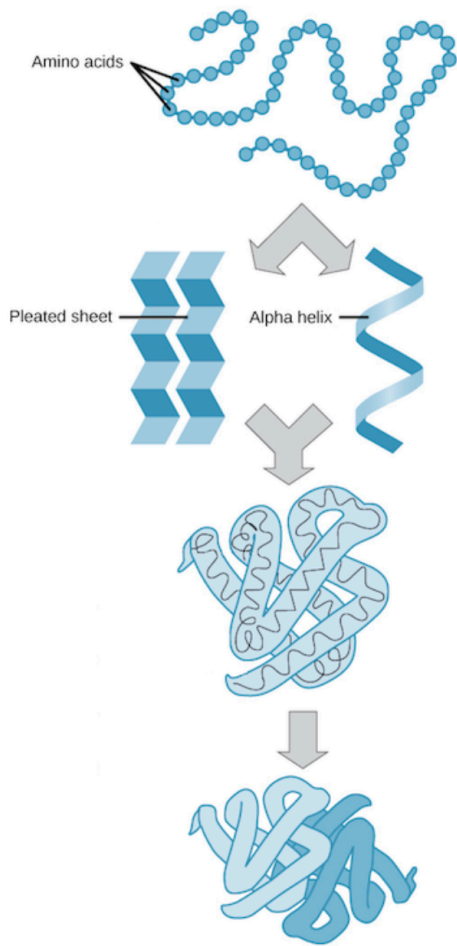


Двугранные углы каркаса аминокислоты



<http://bioinformatics.org/molvis/phipsi/index.htm>

Типы геометрических структур белка

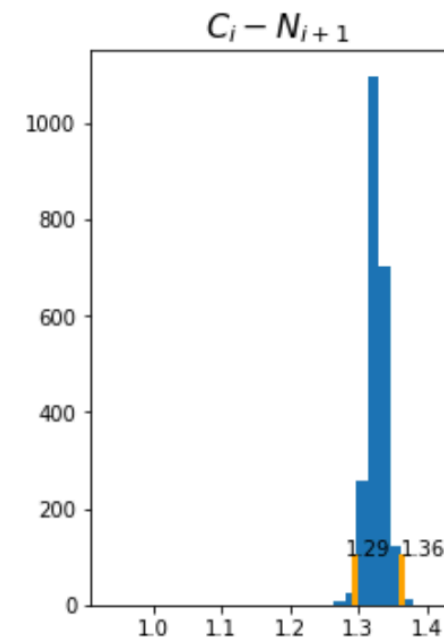
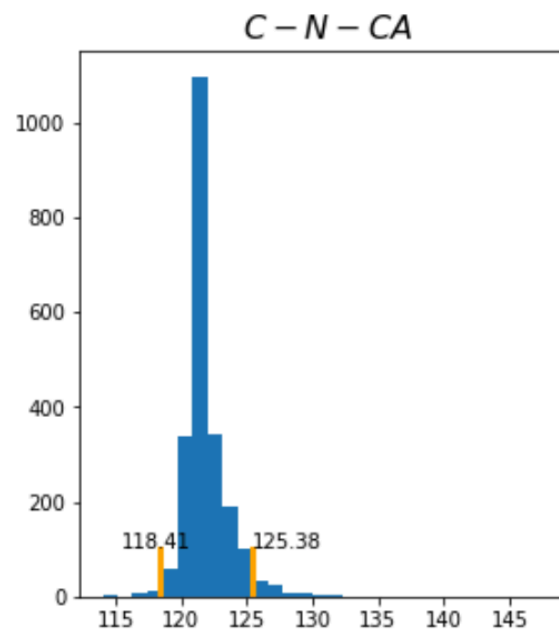


- Третичная структура:

- Представляет собой взаимное расположение вторичных структур цепочки.
- Это расположение определяется:
 - Водородными связями между боковыми цепочками;
 - Гидрофобностью / гидрофильностью боковых цепочек;
 - Сульфидными мостиками.

Планарные углы и длины связей

- Значения всех планарных углов и длин связей кластеризуются около одного значения.
- Однако разброс значений может быть довольно большим, в результате чего малое отклонение может привести к большой потере точности.



Оценка качества предсказания

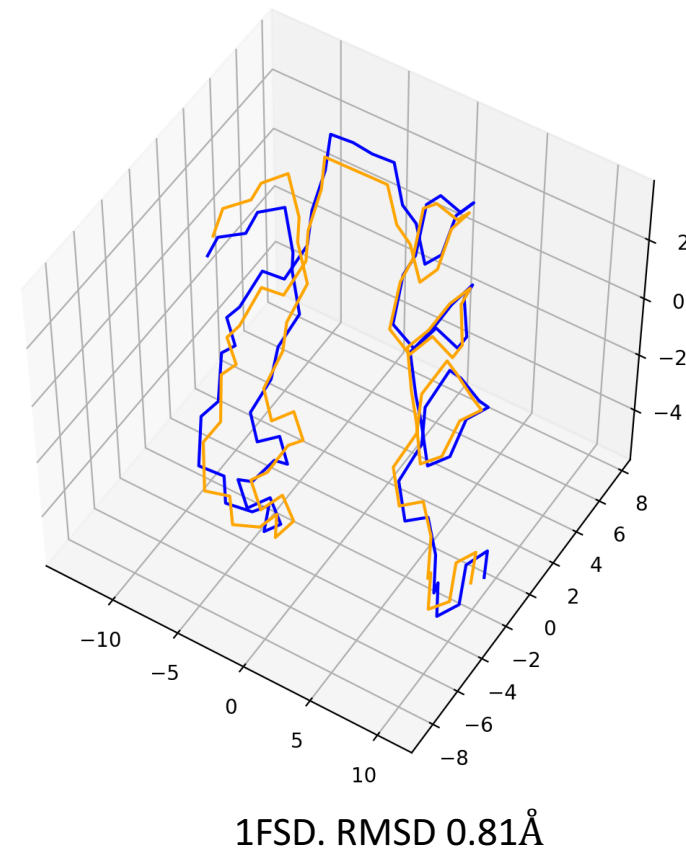
- Для оценки качества предсказания чаще всего используется метрика RMSD (Root Mean Squared Distance):

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \|c_i^{pred} - c_i^{orig}\|^2}$$

- Для разных задач может быть выбран разный порог $RMSD$, при котором предсказание считается 'хорошим'. Обычно это 2 – 4Å.
// Комментарий: для сравнения порядка: диаметр атома углерода равен 1.5Å.

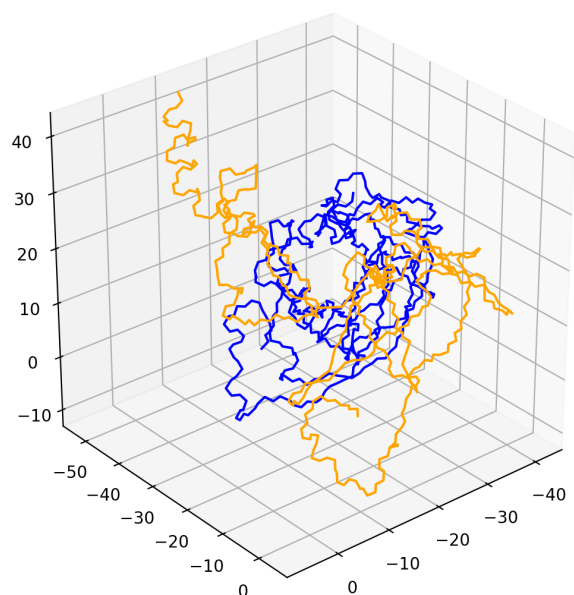
Оценка качества восстановленной геометрии

- Проведем эксперимент для белка:
 - Получим оригинальные значения ϕ, ψ ;
 - Всем длинам связей и планарным углам присвоим среднее значение по соответствующей выборке;
 - Восстановим координаты всех атомов белка;
 - Сопоставим координаты оригинальной и предсказанной структур.

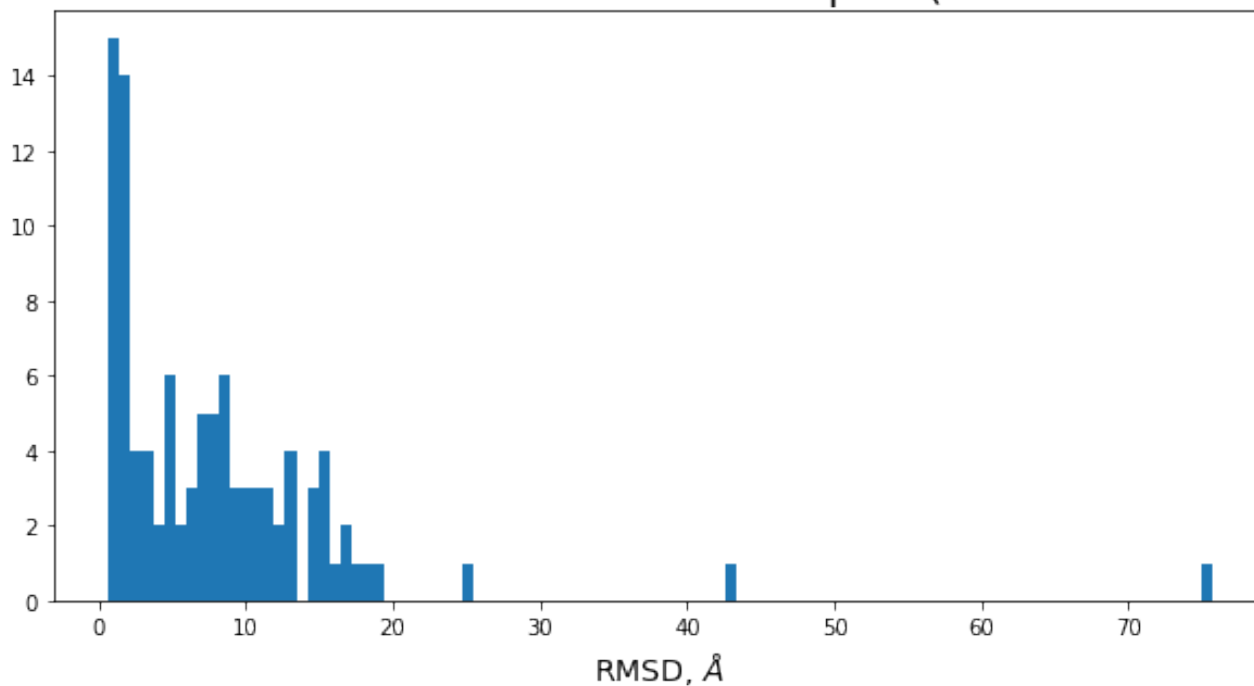


Оценка качества восстановленной геометрии

- Но для более длинных белков (> 100) все может быть гораздо хуже:



Точность восстановленной геометрии (по 100 белкам)



Проверка понимания 🤔

- Что такое первичная структура белка?
- Может ли вторичная структура состоять из ровно одного остатка?
- Могут ли все значения ϕ, ψ в белке быть равными 180° ?
- Влияют ли боковые цепочки на образование вторичных структур?
- Мы обсудили метрику RMSD. А имеет ли смысл оценивать предсказание по точности предсказания значений ϕ, ψ ?
- // Доп. вопрос: предположите, почему двугранный угол ω убирается из предсказания геометрии?