

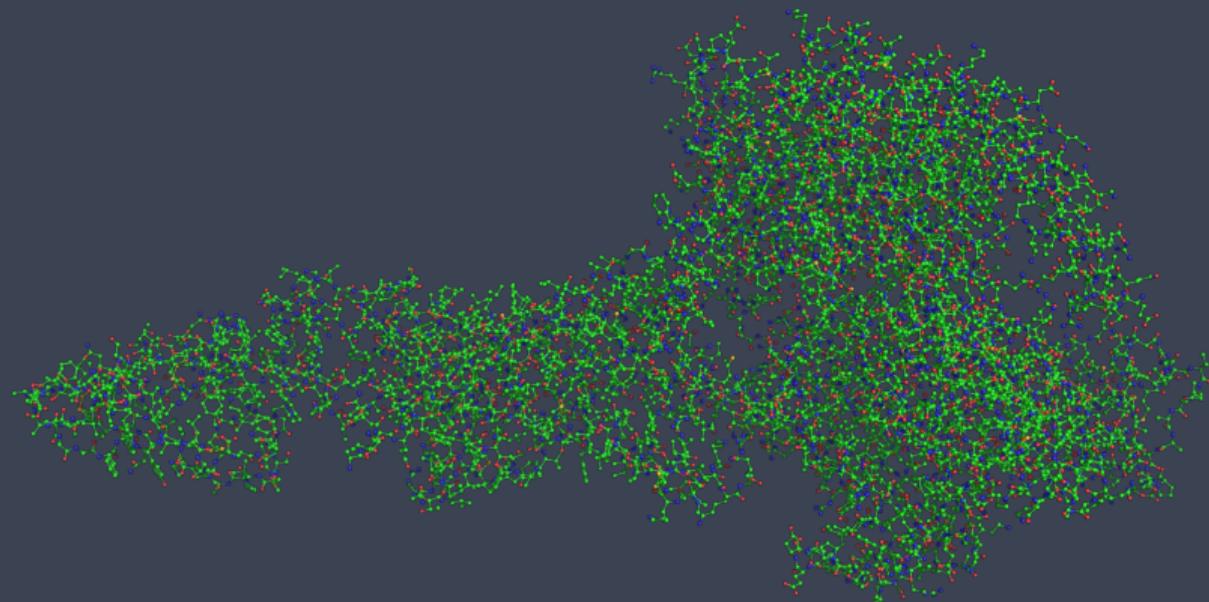
Лекция 1. Введение в структуре белка, визуализация

Курс: Методы машинного обучения в дизайне белков

by Головин А.В.¹ (¹МГУ им М.В. Ломоносова, Факультет Биоинженерии и
Биоинформатики)

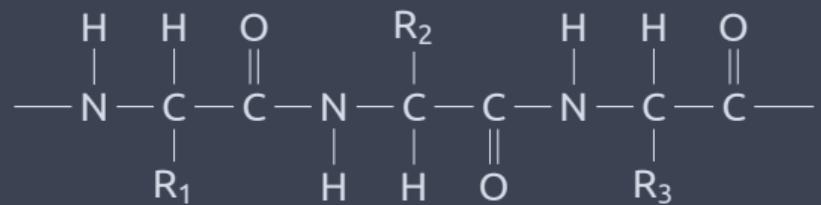
on Москва, 2023

» Структура рецептора



» Что такое белок?

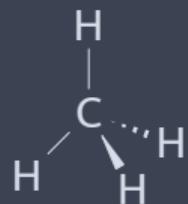
Белки — высокомолекулярные органические вещества, состоящие из соединённых в цепочку пептидной связью альфа-аминокислот.(wikipedia)



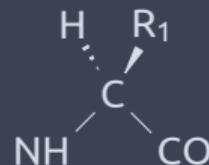
Или: белок это линейный полярный полимер, где мономерами является выборка из примерно 20 L-альфа-аминокислот.



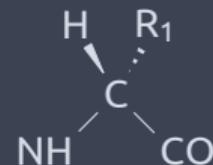
» Что такое L-альфа-аминокислота?



атом углерода в sp^3 гибридизации имеет тетраэдрическое окружение



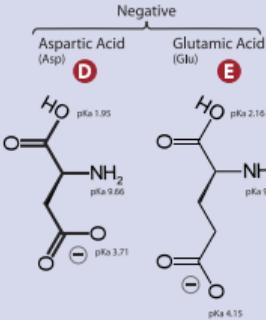
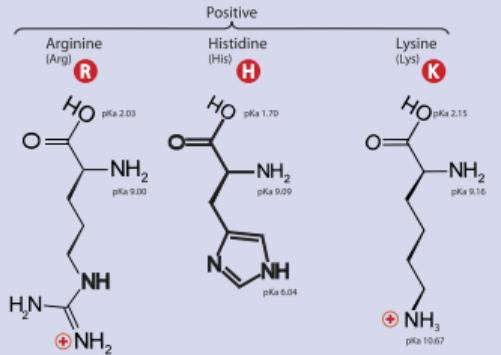
L-аминокислота



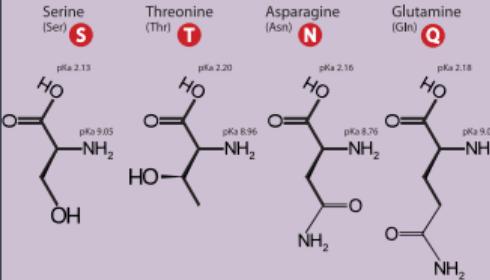
D-аминокислота

» АМИНОКИСЛОТЫ

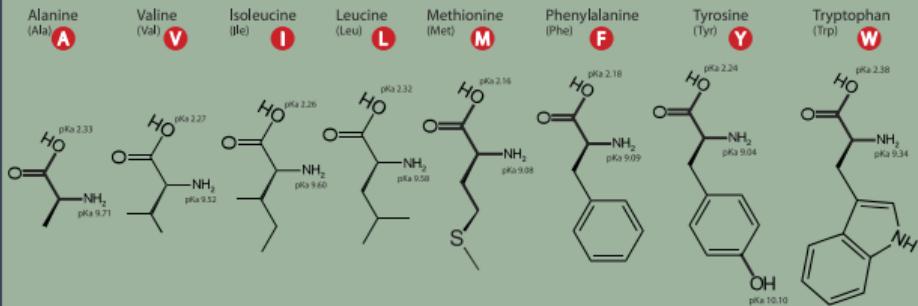
A. Amino Acids with Electrically Charged Side Chains



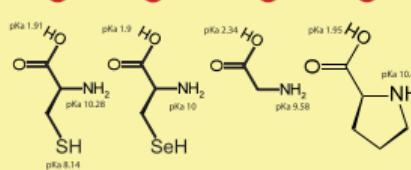
B. Amino Acids with Polar Uncharged Side Chains



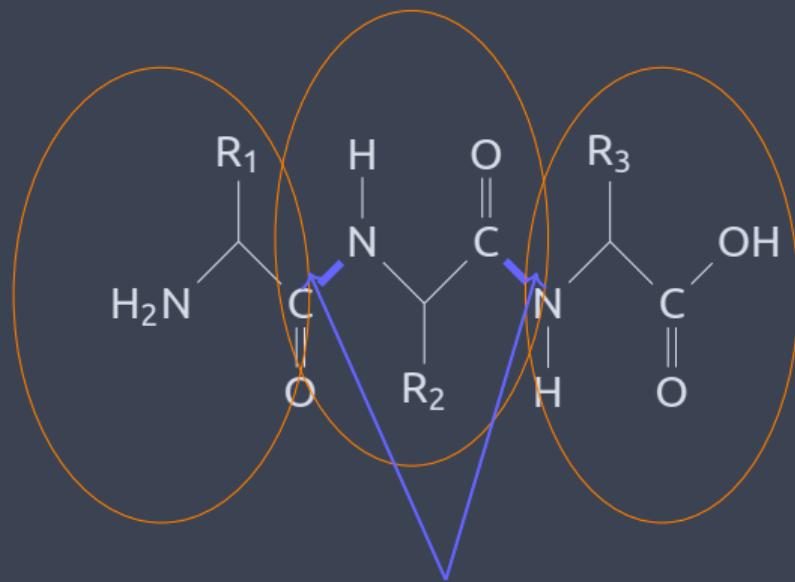
C. Special Cases



C. Special Cases

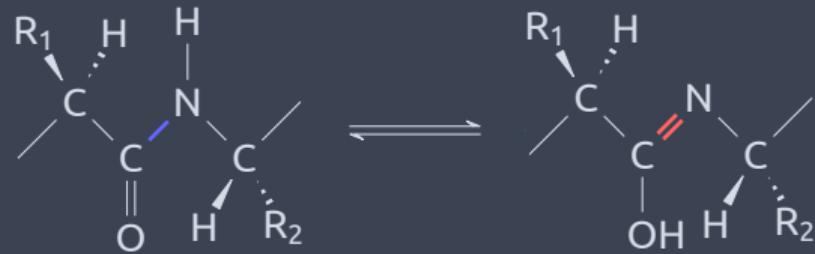


» Пептидная связь



Пептидные связи

» Пептидная связь, таутомерия

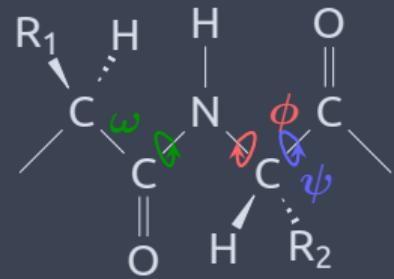


» Пептидная связь, свойства

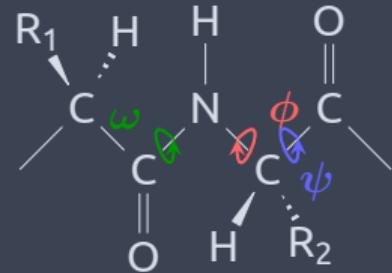
- * Пептидная связь прочнее, чем другие амиды
- * Атомы пептидного звена (C_{α} -C-N- C_{α}) лежат в одной плоскости
- * Валентные углы у атомов С и N примерно равны 120°
- * Вращение вокруг связи C-N затруднено
- * Возможны cis- и trans-конфигурации; в белках преобладают trans
- * Карбонильный кислород – хороший акцептор водорода
- * Амидный азот – хороший донор водорода



» Вращения вокруг связей в остеове белка



» Вращения вокруг связей в остеове белка



ϕ
 ψ

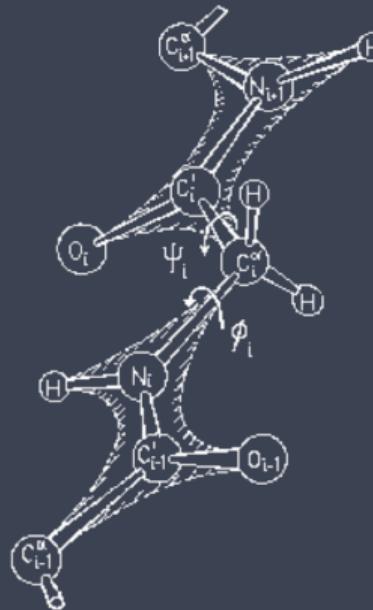
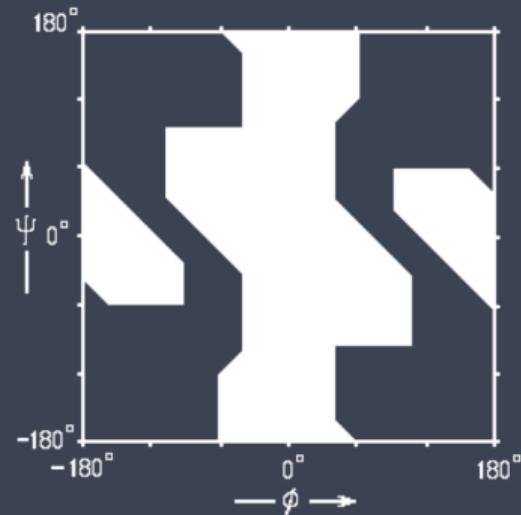
} теоретически могут быть: от -180^0 до $+180^0$

а ω ?



» Карта Рамачандрана

даже в полиглициновой цепи существуют стерические ограничения



Карта запрещенных (■) конформаций глицина



» Уровни организации структуры белка

- * Первичная структура
- * Вторичная структура
- * Укладка (*fold*)
- * Третичная структура
- * Четвертичная структура



» Первичная структура

Первичная структура – это аминокислотная последовательность:

Met-Ala-Gly-Trp-Ala-Val-Asp ...



» Вторичная структура

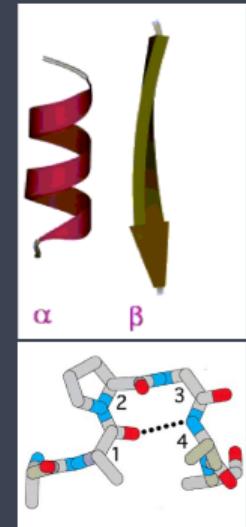
Вторичная

структурой белка - это упорядоченные расположения атомов основной цепи полипептида, безотносительно к типам боковых цепей (групп) и их конформациям.

Если упорядоченность

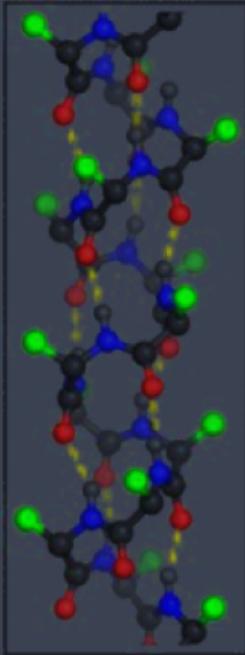
такова, что двугранные углы одинаковы у всех остатков, то говорят о регулярной вторичной структуре. Регулярными вторичными структурами являются спирали и β -структуры.

Пример нерегулярной вторичной структуры β -поворот (β -изгиб, реверсивный поворот).

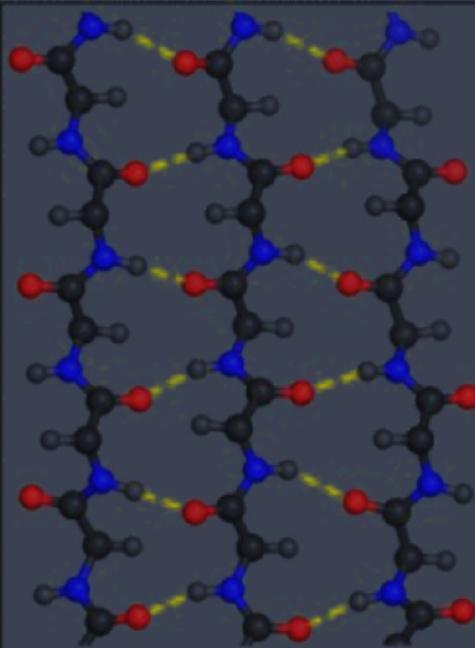


» Регулярные вторичные структуры

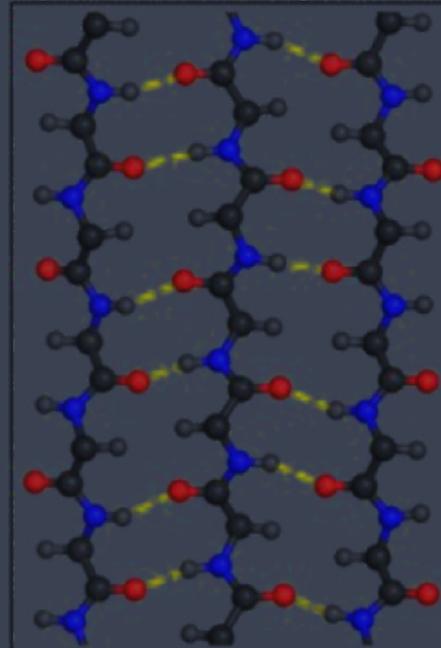
a) α -Helix



b) Parallel β -Sheet



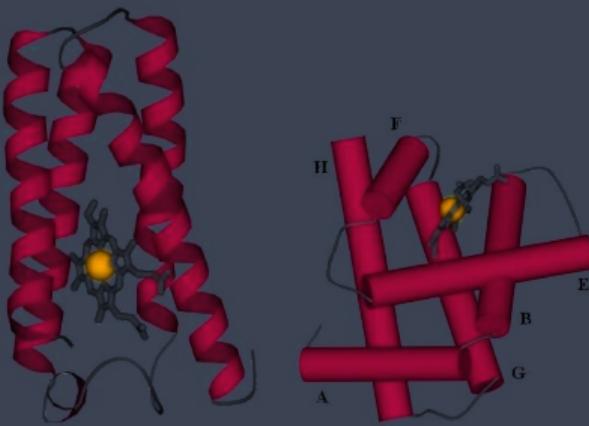
c) Anti-Parallel β -Sheet



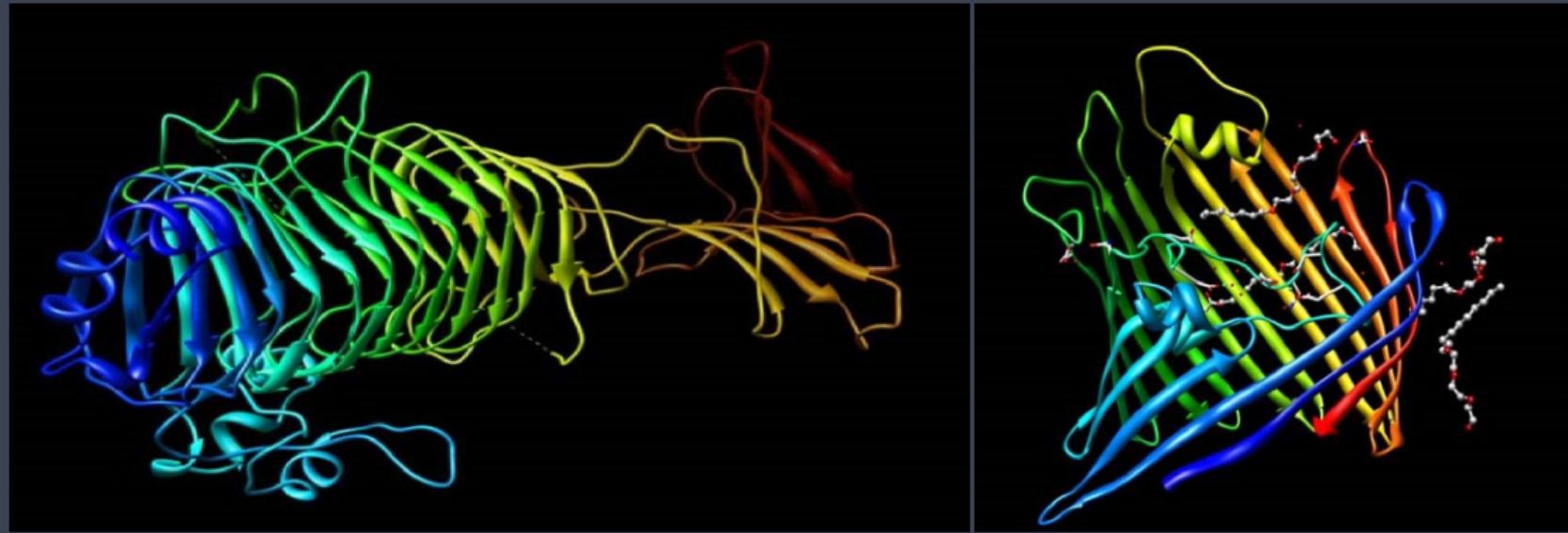
» Укладка (fold)

Укладкой называют организацию в пространстве элементов регулярной вторичной структуры.

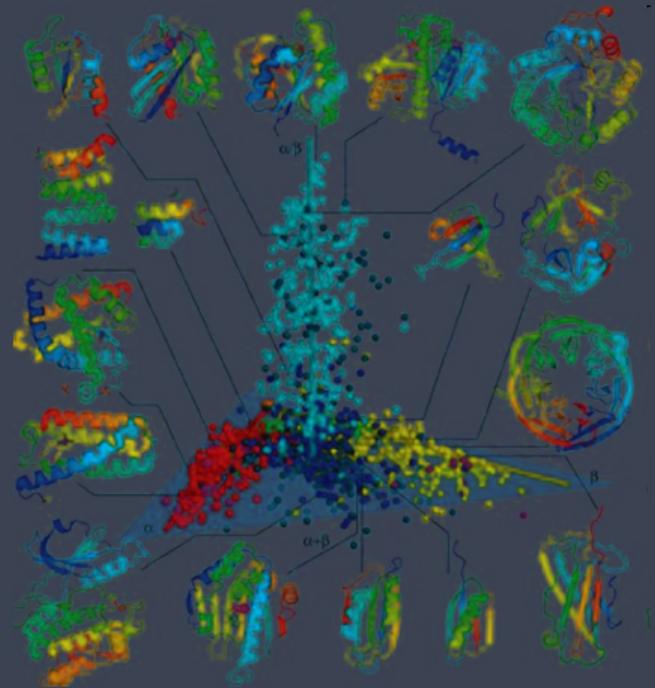
Пример: α -спиральные белки



» β структурные белки



» Распределение в природе



» Третичная структура

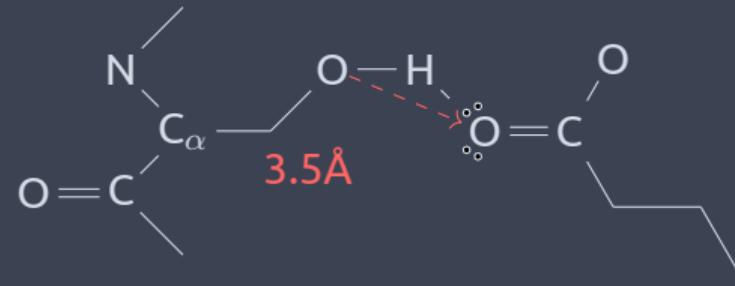
Третичной структурой называют расположение в пространстве всех атомов одной полипептидной цепи.

Т.е. описание третичной структуры включает в себя:

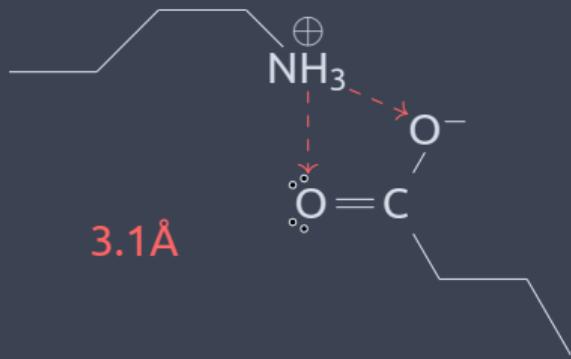
- * описание элементов вторичной структуры,
- * описание типа укладки,
- * описание структуры петель,
- * описание конформаций боковых групп всех аминокислотных остатков.



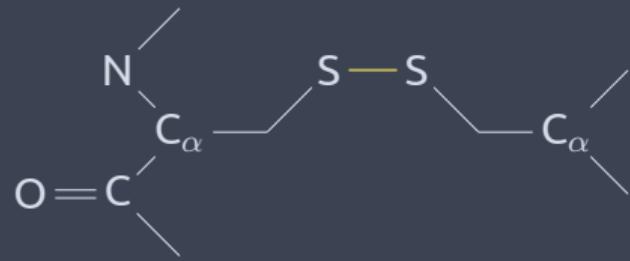
» Вспомогательные взаимодействия: водородные связи



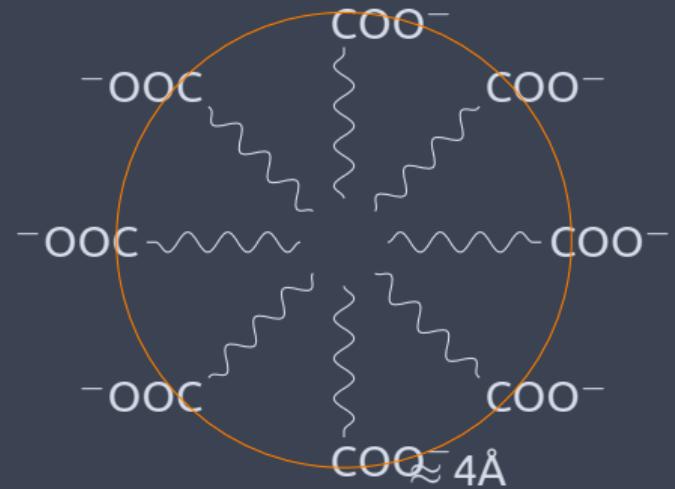
» Ионные пары



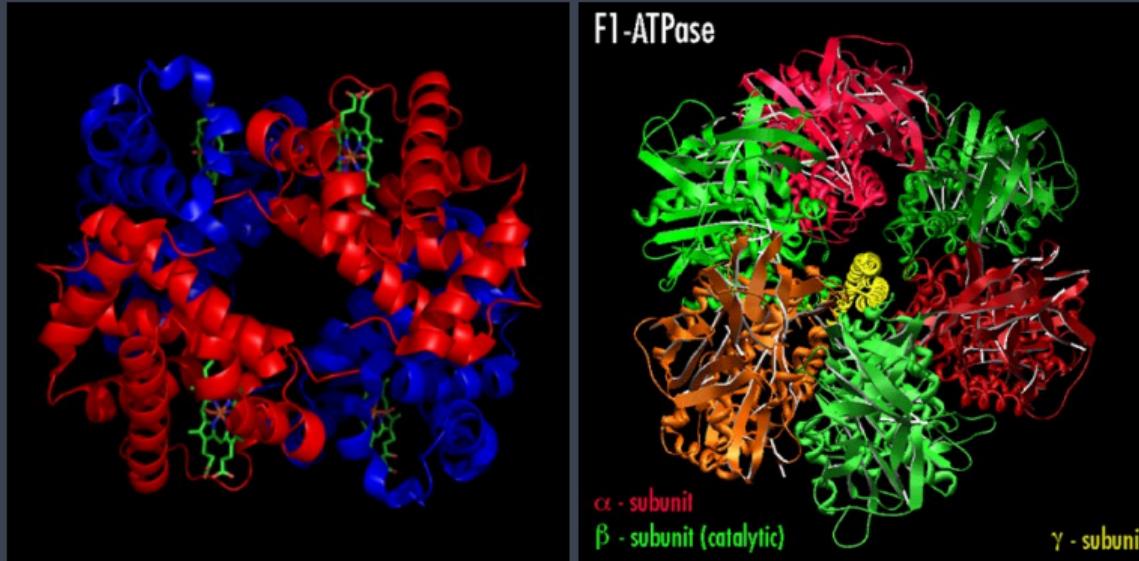
» Дисульфидные мостики характерны для секретируемых белков



» Гидрофобные взаимодействия – главный фактор, заставляющий глобулу свертываться



» От четвертичной структуры к молекулярным машинам



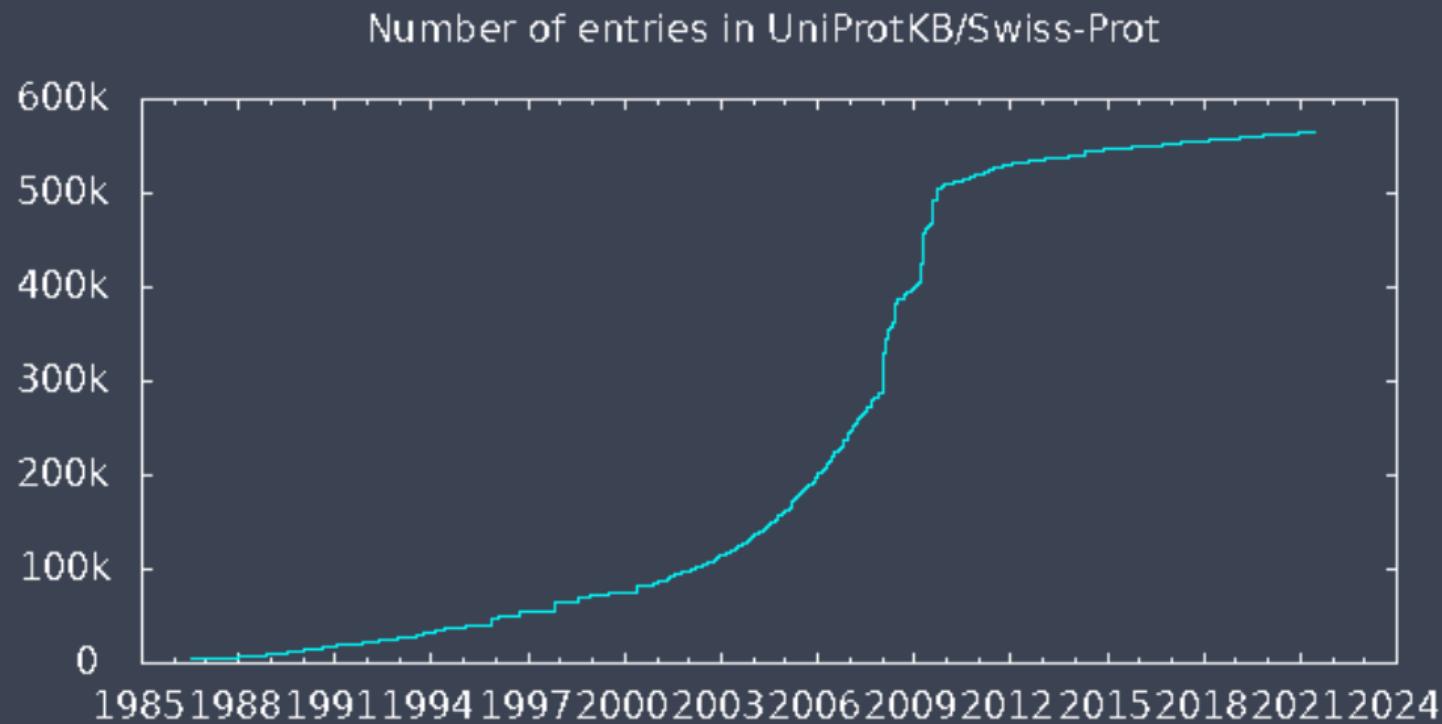
» UniProt Knowledgebase

UniProtKB – две базы аннотированных белковых последовательностей с общим форматом записей.

- * **TrEMBL** (от Translated EMBL) – автоматическая база данных, содержащая, в основном, формальные трансляции открытых рамок считывания, предсказанных в нуклеотидных последовательностях.
- * **Swiss-Prot** (раньше была отдельным банком) – курируемая база данных. Кураторы выбирают записи из TrEMBL, проверяют и дополняют их, переносят в Swiss-Prot.



» Рост числа записей в Swiss-Prot



» Одна запись UniProtKB

- * Одна запись – все продукты одного гена из организмов одного вида. Известные изоформы, полиморфизмы и т.д. указывают в аннотации записей.
- * Изоформы указаны в полях CC (подраздел "Alternative products") и FT (конкретные участки различий), полиморфизмы указывают в поле FT.
- * Правило не строгое, из него есть исключения. Например, если для гена известно множество изоформ, сильно отличающихся по последовательности и функциям, то для них создадут несколько записей



» Таблица локальных особенностей (Feature table, FT)

Имеет строгий формат, список и описание всех возможных ключей доступно на сайте UniProt.

```
FT CHAIN 1..188
FT /note="Isochorismatase family protein YecD"
FT /id="PRO_0000201831"
FT HELIX 6..8
FT /evidence="ECO:0000244|PDB:1J2R"
FT STRAND 9..14
FT /evidence="ECO:0000244|PDB:1J2R"
FT REGION 5..34
FT /note="Interaction with RNase E"
FT ACT_SITE 209
FT /note="Proton donor"
FT /evidence="ECO:0000255|HAMAP-Rule:MF_00318,
FT ECO:0000269|PubMed:15003462"
FT METAL 246
FT /note="Magnesium"
FT /evidence="ECO:0000269|PubMed:11676541,
FT ECO:0000269|PubMed:16516921"
FT BINDING 159
FT /note="Substrate"
FT /evidence="ECO:0000255|HAMAP-Rule:MF_00318"
FT MOD_RES 257
FT /note="N6-acetyllysine"
FT /evidence="ECO:0000269|PubMed:18723842"
FT CONFLICT 180..188
FT /note="SVEEILNAL -> TWKRSSTRYDLHRSTAMVAS (in Ref. 1)"
```

вторичная структура

сайты связывания

модифицированные остатки

разнотчения в последовательности

И Т.Д.



» Выравнивание

DNAK_THEAC	82	KFKVFDKEFTPQQISAFILQKIKKDA-EAFLGEPVN	EAVITVPAYFNDNQR	131		
DNAK_PICTO	82	KYKIFGKEYTPQQISAFILQKIKRDA-EAFLGEPVT	DAVITVPAYFNDNQR	131		
HSCA_ACIF2	116	RRLRTVAGEKSPVEVSAEILRVLKERAVETLGGE	P-EGAVITVPAYFDEAQ	165		
HSCA_ACIF5	116	RRLRTVAGEKSPVEVSAEILRVLKERAVETLGGE	P-EGAVITVPAYFDEAQ	165		
DNAK_THEAC	132	QATKDAGTIAGFDVKRIINEP	TAAALAYGVDKSGKSEKILVF	DLGGGTLDV	182	
DNAK_PICTO	132	QATKDAGAIAGLNVRRIINEP	TAAACLAYGIDKLNQT	LKIVIYDLGGGTLDV	182	
HSCA_ACIF2	166	QATKDAARLAGLNVLRLLAEPT	AAAVAYGLDGKGSEGI-FAI	YDLGGGTFDI	215	
HSCA_ACIF5	166	QATKDAARLAGLNVLRLLAEPT	AAAVAYGLDGKGSEGI-FAI	YDLGGGTFDI	215	
DNAK_THEAC	183	TIMDFGDGVFQVLSTSGDTRL	GGTDMDEAIVNYIADD	FQKKEGIDLRKDRS	233	
DNAK_PICTO	183	TIMDFGQGVFQVLSTSGDT	HLGGTDMDEAIVNFLADNF	QRENGIDLRKDH	S	233
HSCA_ACIF2	216	SILRLQAGVFEVLATAGDS	ALGGDDMDHALAEWLMQE	---EGGDASDPLW	262	
HSCA_ACIF5	216	SILRLQAGVFEVLATAGDS	ALGGDDMDHALAEWLMQE	---EGGDASDPLW	262	
DNAK_THEAC	234	AYIRLRDAAEKAKIELSTTL	STDIDL PYITV	TNSGPKHIKMTL	TRAKLEEL	284
DNAK_PICTO	234	AYIRLRDAAEKAKIELSTV	LETIEINLPYITAT	QDGPKHLQYT	LTRAKFEEL	284
HSCA_ACIF2	263	RRQVLQQ-ARTAKEALSAVA	ET-----MIVL	TPSGRAAREIKL	SRGRLES	307
HSCA_ACIF5	263	RRQVLQQ-ARTAKEALSAVA	ET-----MIVL	TPSGRAAREIKL	SRGRLES	307
DNAK_THEAC	285	ISPIVERVKGPIDKALEGA	KLKKTEITKLLFVG	GPTRIPYVRKYVEDYLG	I	335
DNAK_PICTO	285	IAPIVDRSKVPLDTALEG	AKLKKGDIDKII	LGPGTRIPYVRKYVEDYF	G	335
HSCA_ACIF2	308	IQPVIQRSLPACRRALRDAG	LKLDEIEGVVLVGGATRV	PAVRAMVEEFFRQ	358	
HSCA_ACIF5	308	IQPVIQRSLPACRRALRDAG	LKLDEIEGVVLVGGATRV	PAVRAMVEEFFRQ	358	



» Как устроено вещество?

- * Первые шаги к пониманию того, что вещество состоит из маленьких элементов сделал Лукреций, давно.
- * Первые эксперименты по установлению структуры были проведены только в начале 20 века
- * Появились специальные молекулярные наборы из шариков и палочек
- * Правильное использование химической информации позволило строить первые модели, очень похожие на результаты РСА.



» Как это работает?



» Для чего нужны модели?

- * Упрощение сложного объекта до анализа только той части которая предполагаемо является объектом интереса.
- * Модель является иллюстрацией для дедуктивного анализа очень сложных или многочисленных явлений.
- * Часто модели отражают реальность не полностью, но моделируемой точности бывает достаточно для понимания рассматриваемой системы.



» Финальный этап: Дизайн

Моделирование структуры, определение свойств это шаги к самому
важному этапу :

дизайну или проектированию нового с заданными свойствами.



» Масштабы в моделировании



» Компьютер

- * Формально для решения задач моделирования компьютер не является обязательным элементом
- * Быстрый компьютер значительно увеличивает точность и широту исследования, и следовательно достоверность моделирования.
- * Количество вычислений отражает степень исследования конформационного пространства



» Компьютер и программы

Those programs **always provide a result**, the evaluation of which is at liberty of the user. The programs **tend stubbornly** to calculate **every absurd application** and present a result-not only a number, but also a graph and represent a further instrument of seduction for the uncritical use of algorithms.



» Визуализация с PyMol



» Для чего нужен PyMol

- * Визуализация pdb и прочих файлов с координатами атомов
- * Изготовление высококачественных изображений
- * Начальное редактирование структур



» Системные требования

Компьютер: чем мощнее процессор и чем больше памяти, тем лучше
3D монитор не обязателен, но поддерживается

Операционная система: любая, под Linux проще установить, и он лучше работает с памятью.



» Как установить?

- * Компиляция из исходников: <http://pymol.svn.sourceforge.net/>
- * Установка бинарных пакетов в Ubuntu Linux: sudo apt-get install pymol
- * Установка с Conda: conda install -c schrodinger pymol
или conda install -c conda-forge pymol-open-source



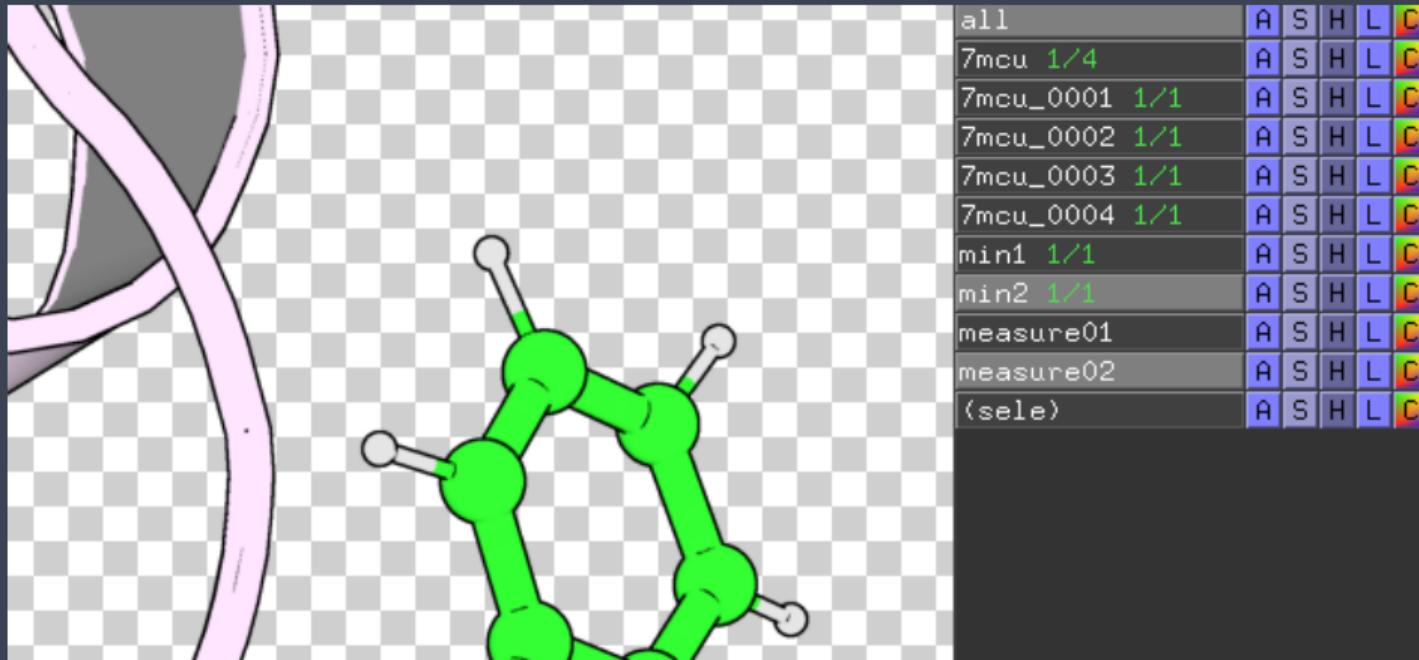
» PyMol - это GPL программа?

Да, PyMol это GPL-программа;

- * исходный код доступен на sourceforge.net
- * Бинарные пакеты для windows стоят денег и продаются:
<http://pymol.org/academic.html>
- * Бинарные пакеты для Linux собираются майтенерами



» Меню объекта/выборки



» Выборки

- * Можно задать с помощью кликов мыши, удерживая SHIFT
- * Удобнее писать выражения в командной строке

Например: *Select backbone, name cat+cn*



» Операторы множеств

- * Логические операторы AND, OR, NOT

Операция OR может быть записана как ”.

Упражнение: Документ PDB содержит описание структуры, состоящей из белка, фрагмента ДНК и молекул воды. Что получится, если задать следующие команды ?

select protein or dna

select protein and dna

select not water

- * Оператор WITHIN(...)

select all within 3.5 of resi 20

select s1, (byres n. ca) within 3.5 of resn LIG



» Help selections

Длинное

```
name <atom names>           n. <atom names>
resn <residue names>         r. <residue names>
resi <residue identifiers>   i. <residue identifiers>
chain <chain ID>            c. <chain identifiers>
id <original-index>
hydrogen                      h.
all                            *
visible                        v.
hetatm
byres <selection>           br. <selection>
byobj <selection>            bo. <selection>
around <distance>            a. <distance>
expand <distance>             e. <distance>
in <selection>
like <selection>              l. <selection>

<selection> within <distance> of <selection>
<selection> w. <distance> of <selection>
```

Короткое

Актуальное:

https://pymolwiki.org/index.php/Selection_Algebra



» Примеры выборок

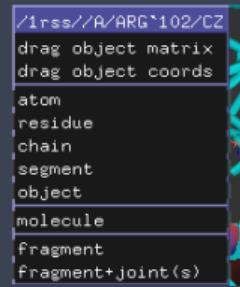
`sel=select`

- * `sel s1, n. ca and c. A` : все атомы CA в цепи A
- * `sel s2, n. ca and (c. A or c. B)` : атомы CA цепей A и B
- * `sel s3, resn GLU and resi 100` : остаток 100 если он GLU
- * `sel s4, resi 100-120+130` : атомы остатков 100-120 и 130
- * `sel s5, byres(name CG)` : атомы остатков где есть CG



» Иерархическое определение выборки

Легко увидеть иерархию правым кликом по атому



sel s1, a/102/cz : атом cz в остатке 102

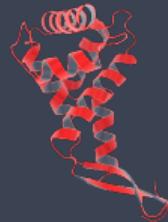
sel s2, 100-120/N and c. A : атомы N в остатках 100-120 цепи a

sel s3, a/100+120/ : все атомы остатков 100 и 120 в цепи A



» Трассировка лучей, команда ray

Подробно: <http://www.pymolwiki.org/index.php/Ray>



No ray



ray_trace_mode,0



ray_trace_mode,1

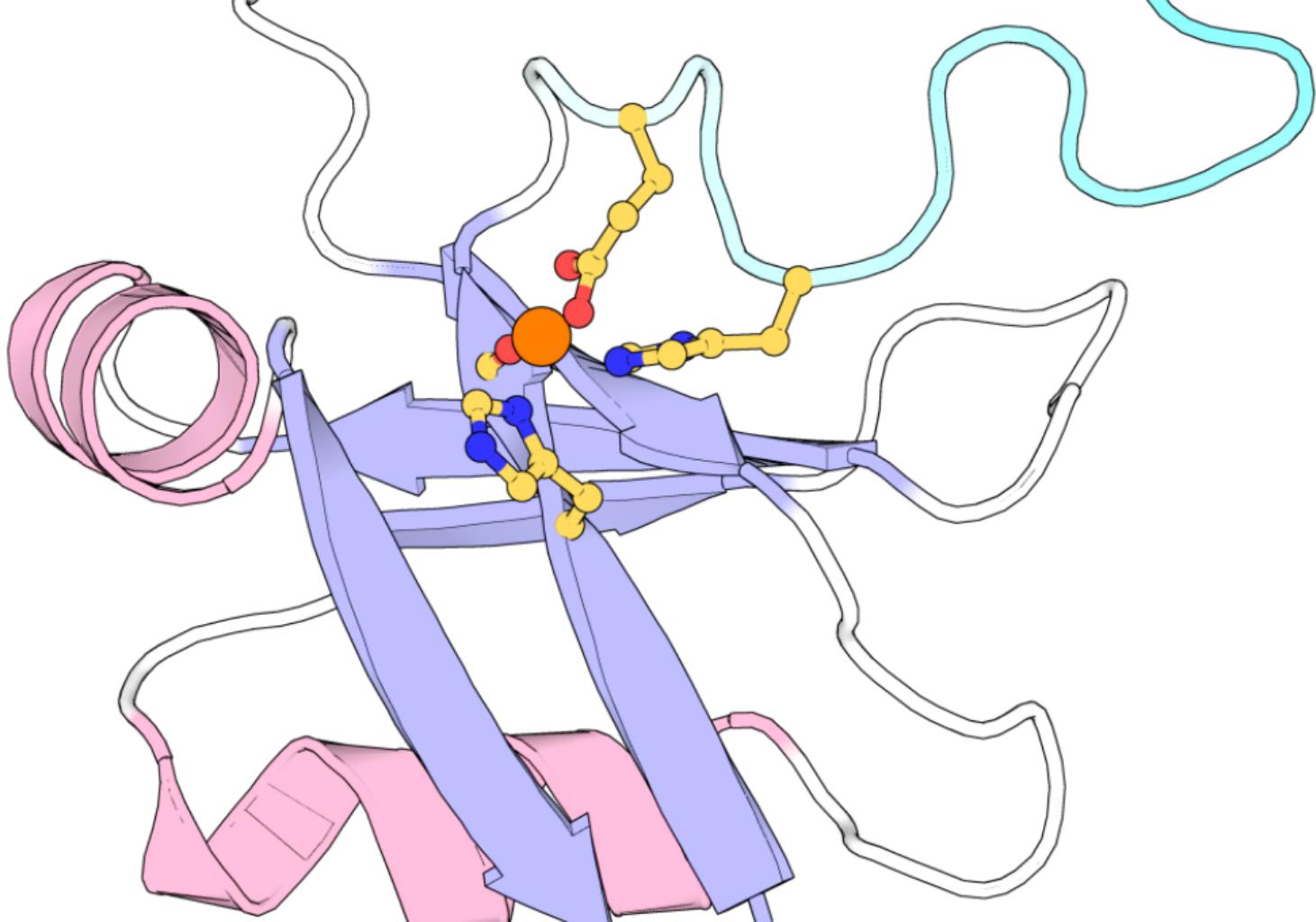


ray_trace_mode,2



ray_trace_mode,3





» Настройки изображения

<http://www.pymolwiki.org/index.php/Category:Settings>

- * PyMol содержит порядка 600 настроек
- * Не все документированы
- * Большинство интуитивно понятны
- * Настройки доступны через меню или в командной строке набрать:
set первые буквы имени опции и клавиша tab для достроения



» Примеры

```
#initial setup
viewport 600, 600 — размер графического окна
set auto_zoom, off — не приближать новые объекты
set auto_show_lines, off — не показывать линии автоматически
set auto_show_selections, off — не показывать выборку автоматически
#cartoon parameters
set cartoon_fancy_helices,1 — изменение вида спиралей
set cartoon_highlight_color, grey60 — цвет внутренней стороны спиралей
set cartoon_dumbbell_length,1.0 —ширина ленты в спирали
set cartoon_rect_length,1.40000 — ширина ленты в бета
set cartoon_loop_radius,0.3 — толщина неструктур. участка
set cartoon_smooth_loops=0 — без сглаживания
```



» \$HOME/. pymolrc

```
set sphere_scale ,0.2
set async_builds , 1
set ribbon_width , 8
set antialias , 2
```



» \$HOME/.pymolrc.py

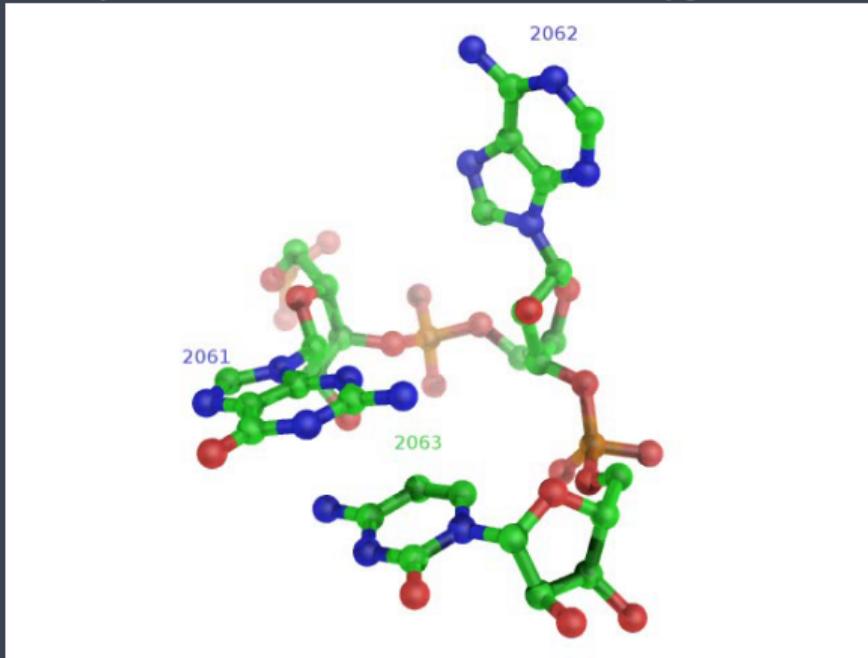
```
from pymol import cmd
from pymol import stored

@cmd.extend
def superall(target):
    seleobjs = cmd.get_object_list('all')
    for o in seleobjs:
        if o != target:
            cmd.super(o, target)
            print(o)
```



» Анимация в PyMol

Если структура содержит более чем одну модель, то в PyMol можно анимировать движение молекулы переходом от одной модели к другой

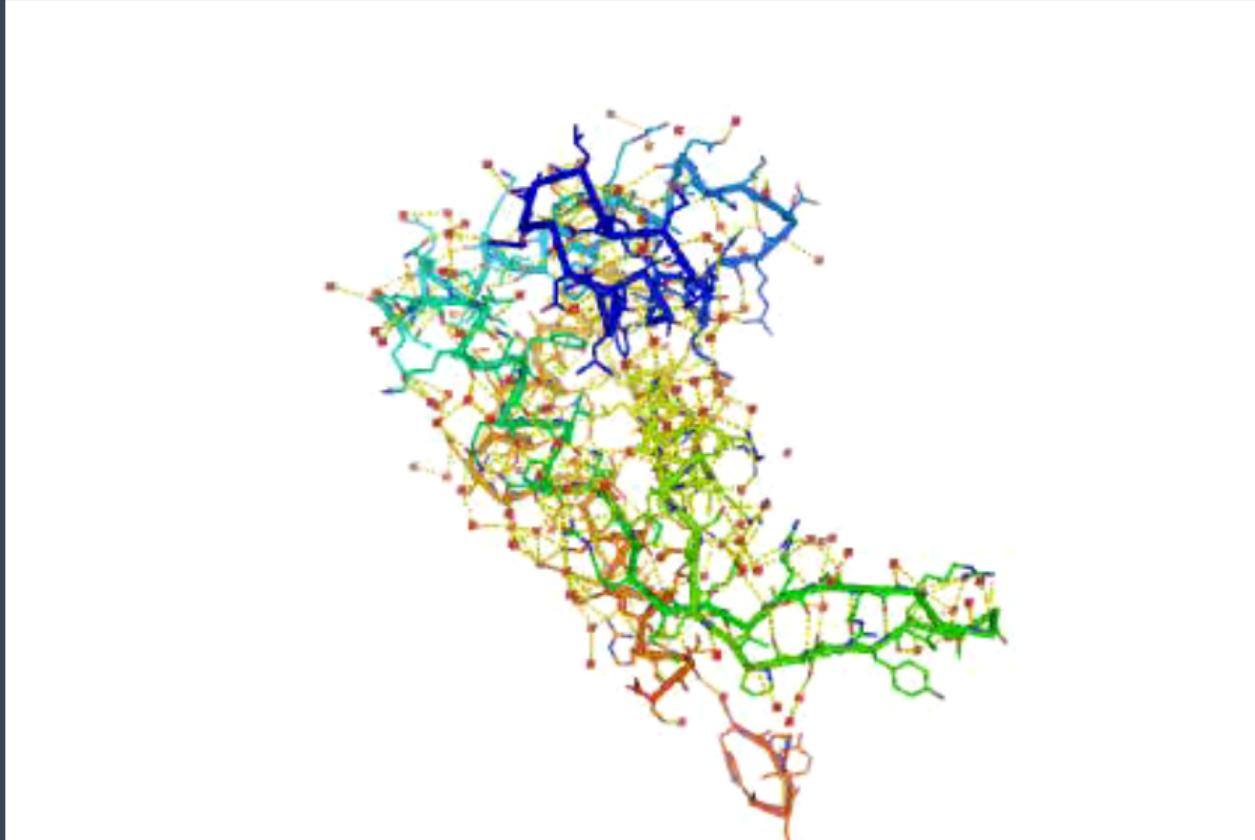


» Пример

- * Action->Preset->Technical (viewer gui)
- * Scene->Store->F1
- * zoom i. 90 # увеличение остатка 90
- * Scene->Store->F2
- * Movie->Program->Scene Loop->Y-Rock->4 Seconds Each
- * File-> Save movie



» Результат



» Анимация, терминология

- * Объект и выборка : смотри выше
- * states: конформация или набор координат
- * scene: позиция камеры и отображение
- * frames: это кадры в анимации, содержит state и scene



» Анимация, команды

mset 1 -55 : задать анимацию от 1 до 55 state на 55 кадров (frames)

mset 1 x90 : задать анимацию первого state от 1 до 90 кадров

mset 1 x30 1 -15 15 x30 15 -1 : первые 30 кадров state 1, следующие 15 кадров это состояния 1-15, следующие 30 кадров состояние 15, следующие 15 кадров состояния от 15 до 1



» Анимация, команды

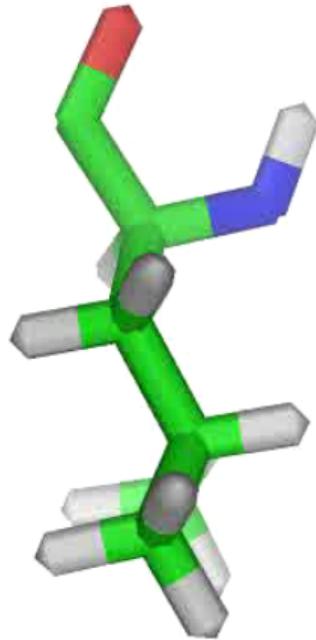
mview : команда для создания ключевых точек

Пример :

- * mset 1 x100
- * frag leu # создаём LEU
- * orient # ориентируем его
- * mview store # запоминаем ключевую точку
- * frame 100 # переходим в кадр 100
- * zoom ID 10 # увеличиваем атом №10
- * mview store # запоминаем ключевую точку
- * mview reinterpolate # делаем интерполяцию



» Результат mview



» Дополнительные команды

- * `mmatrix` : устанавливает вид для первого кадра
- * `util.mrock` : покачивание сцены на определённый угол
- * `util.mrock(start, finish, angle, phase, loop-flag)`
- * `util.mroll` : вращение вокруг оси Y
- * `util.mroll(start, finish, loop-flag)`
- * `mdo` : (устарело) запуск какой-либо команды в заданном кадре

Актуальная информация:

<https://pymolwiki.org/index.php/MovieSchool>



» Сохранение анимации

Старый путь:

```
set ray_trace_frames,1  
mpng mymovie
```

Нужны программы avidemux, Virtual Dub, mencoder для того, чтобы собрать ролик с нужным сжатием (кодек)

Новый путь: File->Save movie ; есть недостаток, старый офис понимает только avi с определённым кодеком



» Моделирование и редактирование в PyMol

- * Можно перемещать объекты и сохранять их новые координаты
- * Можно рассчитать вторичную структуру
- * Можно менять координаты отдельных атомов
- * Можно вносить мутации в белок (но не НК)
- * Можно конвертировать L->D аминокислоты
- * Можно добавлять протоны
- * Можно выравнивать в пространстве молекулы
- * Можно добавлять некоторые фрагменты из библиотеки и собственные



» Перемещение объектов

Рекомендуемый порядок действий:

- * `set retain_order #` надо сохранить порядок атомов
- * `create newobj, sele #` создаём новый объект, страховка
- * `translate [0,10,0], newobj #` перемещаем
- * `rotate x,90,newobj #` вращаем
- * `save newfile.pdb, newobj`

Операции по перемещению и вращению можно делать мышкой в режиме `editing`



» Изменение координат отдельных атомов и объектов

alter_state 1,(pdb1cse),x=x-10.0

Или translate [0,10,0], A/100/NZ



» Удаление связей, но не атомов

- * Выберите первый атом, `ctrl+middle click`, выберите второй атом, `ctrl+middle click`
- * И `unbond` или `ctrl+D`

Внимание! Координаты атомов не меняются, только исчезает изображение связи



» Мутация аминокислот

- * Запустите wizard->mutagenesis
- * Выберите аминокислоту для мутации
- * Справа выберите, на что мутировать
- * Выберите ротамер с помощью управления movie
- * Закончите процедуру с Apply



» Добавление протонов

Работает с молекулами, т.е. объектами

`create gln, A/101/`

`h_add gln`

Или через меню `action` объекта.

Есть вероятность, что протоны будут добавлены неверно, если PyMol неправильно угадал валентность тяжёлых атомов.



» Суперпозиция в пространстве

Задача достаточно нетривиальная, и есть разные пути:

Белки:

Другое:

pair_fit

Желательно указывать родственные атомы в молекулах

pair_fit (tRNA10 and resid 10:15 and name P), (ref4 and resid 10:15 and name P)



» Добавление органических фрагментов или а.к.

- * С помощью `ctrl+middle click` выделите шариком атом, к которому будет присоединяться фрагмент
- * В меню Build выберите нужный фрагмент
- * С помощью `ctrl+left click` выберите торсионный угол
Или
- * Создайте свою молекулу (ChemSketch)
- * Сохраните как pkl в `<pymol_path>/data/chempy/fragments`
- * `editor.attach_fragment('pk1','my_fragment_name',11,0)`
11 - это номер атома в фрагменте для связи



» Sculpting, что ЭТО?

Это похоже на real-time оптимизатор геометрии, но это алгоритм, который старается сохранить значения длины связей, углов, торсионных углов при изменении координат.



» Как запустить sculpting?

У вас достаточно мощный компьютер? Тогда:

- * Переводим мышь в режим редактирования
- * Выбираем "auto-sculpting" из меню Sculpting
- * Выбираем Sculpting из меню Wizard
- * Выбираем центральный атом для модификаций Ctrl-middle-click
- * Тянем атом в любую сторону ctrl-left-click-and-drag



» Скриптование в PyMol

Возможны как скрипты из команд, так и скрипты на Python

Запуск скриптов из команд:

@ myfile.pml

Запуск скриптов на питоне:

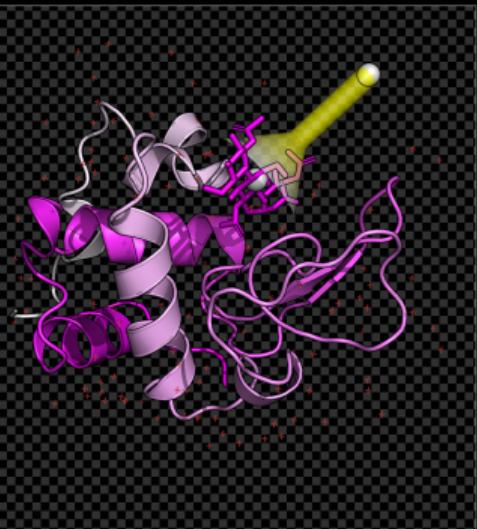
run myfile.py



» Пример

```
fetch 1cll, async=0
as lines, n. C+O+N+CA
zoom i. 4+5
mset 1 x1440
mview store
python
for x in range(0,144):
    cmd.frame((10*x)+1)
    cmd.zoom( "n. CA and i. "+ str(x) + "+"+ str(x+1))
    cmd.mview("store")
python end
frame 288
mview store
mview reinterpolate
```



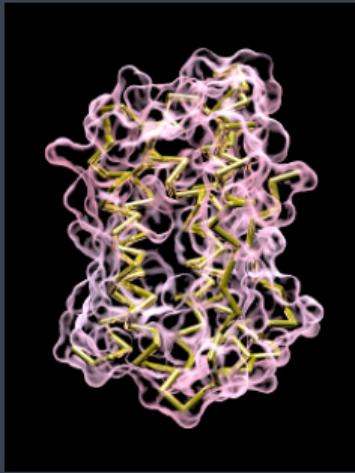
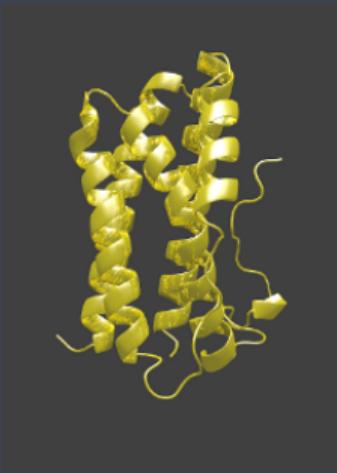
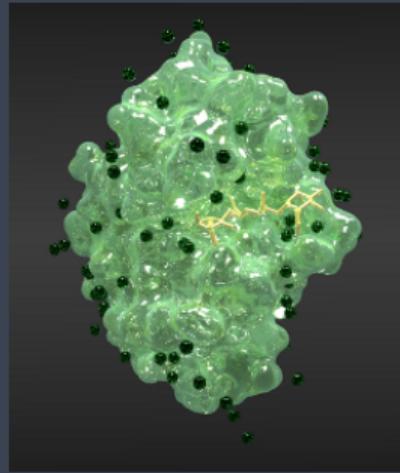


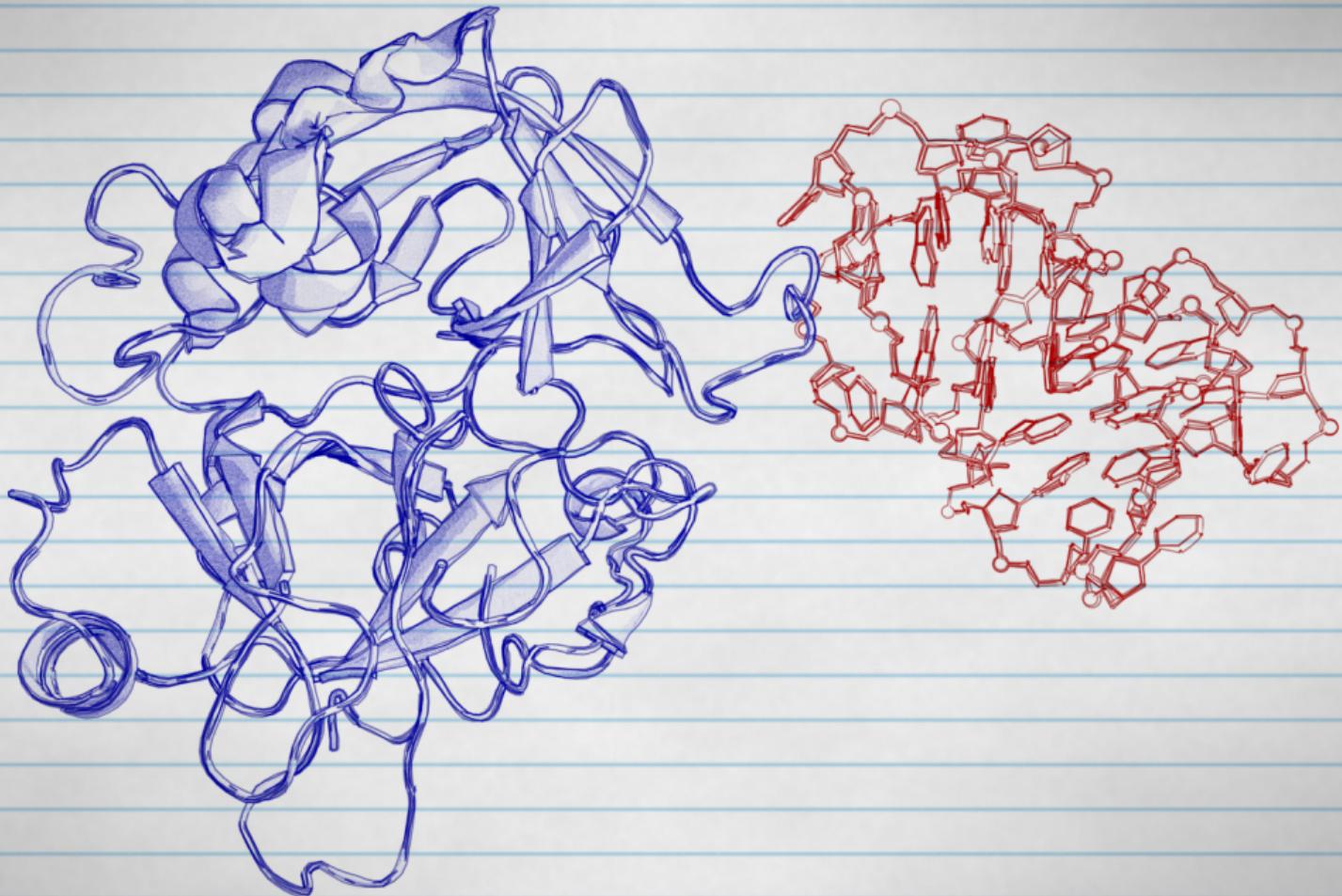
A screenshot of the Mac OS 9 Control Panel. The window title is "Buttons". Inside, there are two rows of buttons labeled "Left" and "Right". Each row has four buttons: "None", "Normal", "Hand", and "Size". The "Normal" button in the left column is highlighted with a green border.



Какая-то муть

» Объекты из PyMol можно использовать в разных 3D программах







Анимация структуры в Blender