

Разработка среды для построения структур белковых алгоритмов

Рак Алексей

23 мая 2018 г.

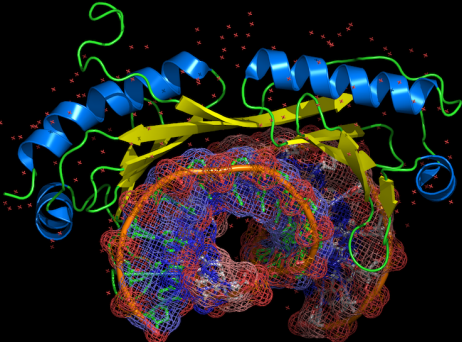
- 1 Среда
- 2 Алгоритм выравнивания
- 3 Результаты

PyMOL

```

ScenePNG: wrote 1280x952 pixel image to file "out8.png".
PyMOL>ray 2440, 1132
Ray: render time: 3.81 sec. = 945.7 frames/hour (6.85 sec. accum.).
PyMOL>ray 2334, 1132
Ray: render time: 3.74 sec. = 961.6 frames/hour (10.59 sec. accum.).
PyMOL>ray 2292, 1132
Ray: render time: 2.50 sec. = 1437.5 frames/hour (13.09 sec. accum.).
PyMOL>ray 2292, 1132
Ray: render time: 4.57 sec. = 787.6 frames/hour (17.67 sec. accum.).
PyMOL>ray 2264, 1132
Ray: render time: 4.44 sec. = 810.1 frames/hour (22.11 sec. accum.).
  
```

PyMOL>



Reset Zoom Orient Draw/Ray

Unpick Deselect Rock Get View

|< < Stop Play > >| MClear

Builder Properties Rebuild

	A	S	H	L
all				
igne 1/1				
cc 1/1				
cd 1/1				

Mouse Mode 3-Button Viewing

Buttons L M R Wheel

Keys Rotate Move MoveZ Slab

Shift +Box +Box Clip MovS

Ctrl Move PKAt PK1 MovZ

Ctrl Sele Orig Clip MovZ

SingleClick +/- Cent Menu

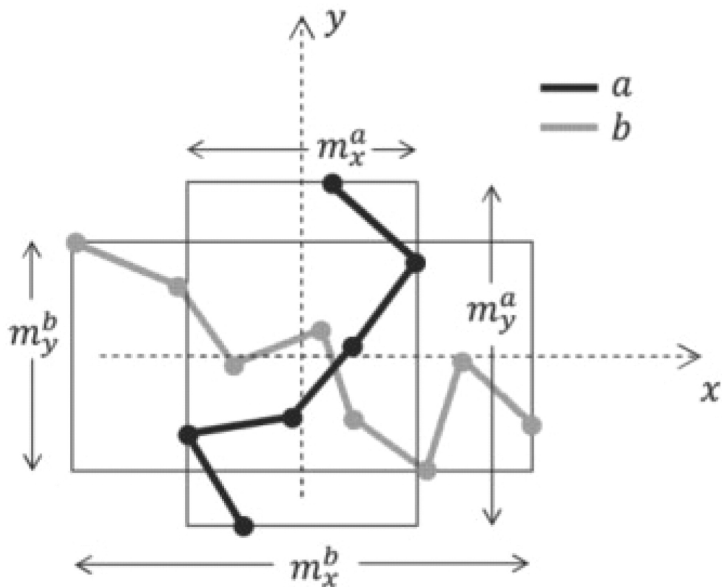
DoubleClick Menu PKAt

Selecting Residues

State 1/ 1

PyMOL>

Структурное выравнивание



- $CA \leq \sigma$

- $CA \leq \sigma$
- GDT_TS

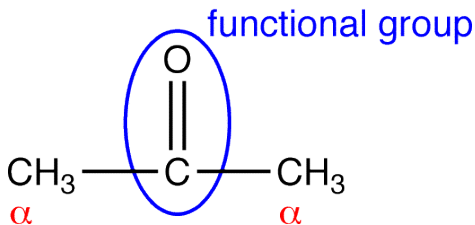
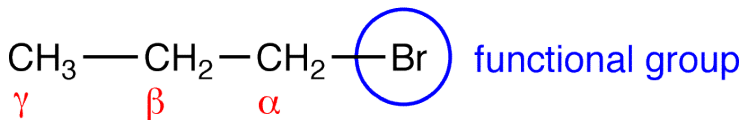
- $CA \leq \sigma$
- GDT_TS
- ALo

- $CA \leq \sigma$
- GDT_TS
- ALo
- MaxSub

- $CA \leq \sigma$
- GDT_TS
- ALo
- MaxSub
- TM-score

- $CA \leq \sigma$
- GDT_TS
- ALo
- MaxSub
- TM-score
- Q-score

Альфа-карбонь



Пространство допустимых решений

$$t = (\alpha, \beta, \gamma, u, v, w)$$

$$I = [0, 2\pi] \times [0, \pi] \times [0, 2\pi] \times [-M_x, m_x] \times [-M_y, M_y] \times [-M_z, M_z]$$

$$M_x = \frac{m_x^a + m_x^b}{2}$$

$$M_y = \frac{m_y^a + m_y^b}{2}$$

$$M_z = \frac{m_z^a + m_z^b}{2}$$

$$r_{\alpha,\gamma} = \frac{6\sqrt{2}\pi R_b}{\varepsilon}$$

$$r_\beta = \frac{3\sqrt{2}\pi R_b}{\varepsilon}$$

$$s_x = 1 + \frac{2\sqrt{3}M_x}{\varepsilon}$$

$$s_y = 1 + \frac{2\sqrt{3}M_y}{\varepsilon}$$

$$s_z = 1 + \frac{2\sqrt{3}M_z}{\varepsilon}$$

$$d_r = \frac{\varepsilon}{3\sqrt{2}R_b}$$

$$d_t = \frac{\varepsilon}{\sqrt{3}}$$

В ε -сети находится $O(\frac{n^6}{\varepsilon^6})$ ячеек.

Для подсчёта результата метрики требуется $O(n^2)$ времени.

Общая сложность алгоритма $O(\frac{n^8}{\varepsilon^6})$.

Поиск оптимального решения

```
1:  $\varepsilon \leftarrow 1$ 
2:  $t_\varepsilon^\sigma \leftarrow \text{EPSILON} - \text{OPTIMAL}(a, b, \sigma, \varepsilon)$ 
3:  $t_{eps}^{\sigma-\varepsilon} \leftarrow \text{EPSILON} - \text{OPTIMAL}(a, b, \sigma - \varepsilon, \varepsilon)$ 
4:  $\varepsilon \leftarrow \frac{\varepsilon}{2}$ 
5: while  $|S(a, t_\varepsilon^\sigma(b), \sigma + \varepsilon)| - |S(a, t_\varepsilon^{\sigma-\varepsilon}(b), \sigma)| > 0$  do
6:    $t_\varepsilon^\sigma \leftarrow \text{EPSILON} - \text{OPTIMAL}(a, b, \sigma, \varepsilon)$ 
7:    $t_{eps}^{\sigma-\varepsilon} \leftarrow \text{EPSILON} - \text{OPTIMAL}(a, b, \sigma - \varepsilon, \varepsilon)$ 
8:    $\varepsilon \leftarrow \frac{\varepsilon}{2}$ 
9: end while
10: return  $t_\varepsilon^{\sigma-\varepsilon}$ 
```

Набор для тестирования содержит 195 пар белков связанных на различных условиях согласно структурной классификации SCOP: 57 family пар, 75 superfamily пар, и 63 fold-пары.

Таблица: Общее число пар в тестовом наборе, которые могут быть наложены на расстояние не превосходящем 3 Å

	MAX-PAIRS	LGA	TM-align	Mammoth	Mustang
Family	4689	4585	4460	4264	4231
Superfamily	4378	4247	4140	3713	3319
Flod	2870	2720	2634	2100	1834

Набор для тестирования содержит 195 пар белков связанных на различных условиях согласно структурной классификации SCOP: 57 family пар, 75 superfamily пар, и 63 fold-пары.

Таблица: Общее число пар в тестовом наборе, которые могут быть наложены на расстояние не превосходящем 5 Å

	MAX-PAIRS	LGA	TM-align	Mammoth	Mustang
Family	5261	5130	5059	5019	4983
Superfamily	5240	5033	4928	4702	4532
Flod	3575	3409	3279	2842	2816