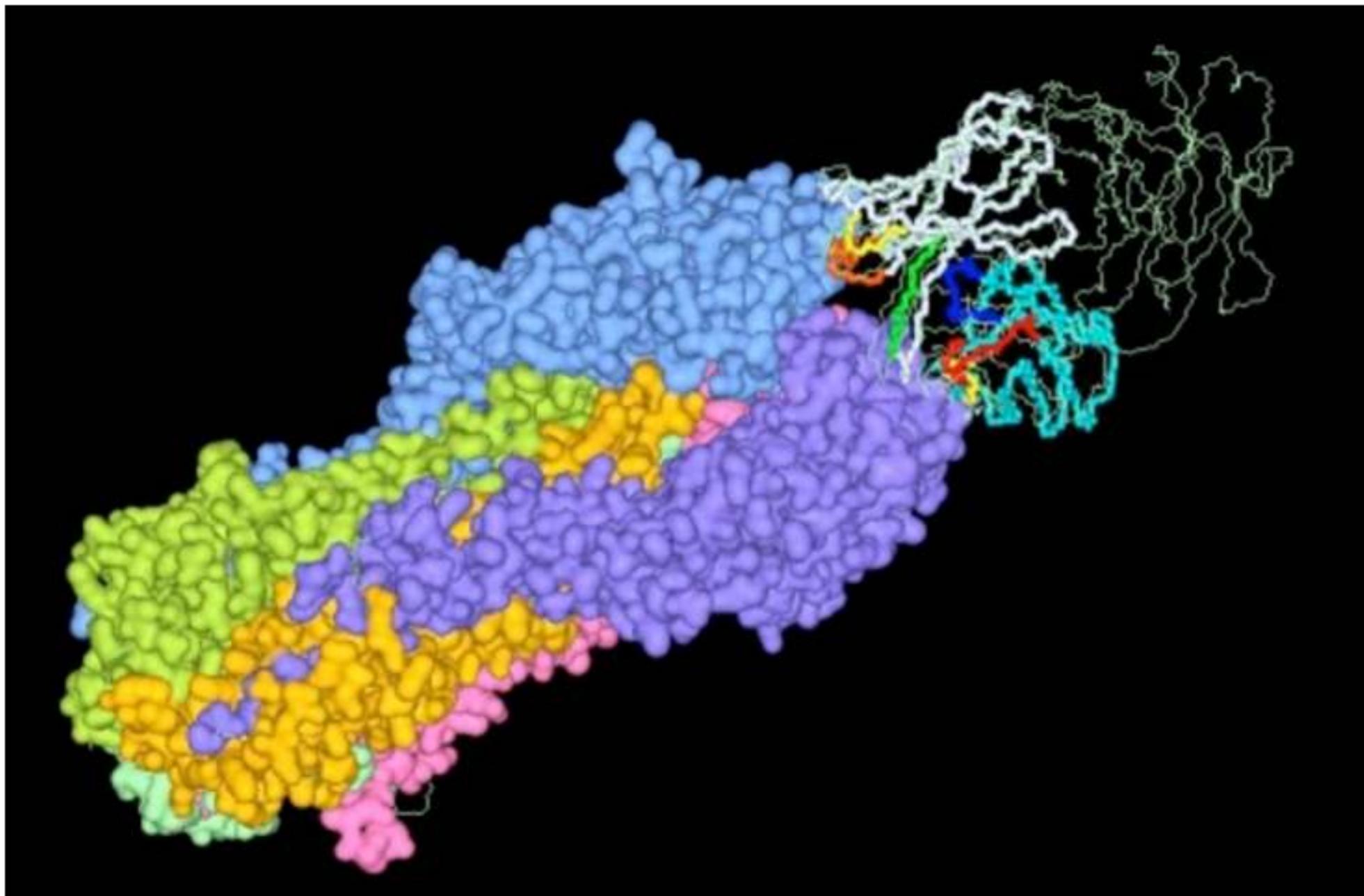


1. Молекулярный Докинг
2. Потенциалы для Молекулярных Докинга и Фолдинга
3. Анализ Нормальных Мод
4. Итеративные Методы Диагонализации

Сергей Грудинин, 06.05.2010

Что такое молекулярный докинг?



со страницы Дэвида Ритчи <http://hex.loria.fr/gallery/>

- молекулярный докинг = распознавание формы в 3D

Цель Докинга

Найти минимум функции потенциальной энергии как можно быстрее

$$E = \int \phi(\underline{r})\rho(\underline{r})dV$$

Для двух белков

$$\phi(\underline{r}) = \phi_A(\underline{r}) + \phi_B(\underline{r})$$

$$\rho(\underline{r}) = \rho_A(\underline{r}) + \rho_B(\underline{r})$$

И следовательно,

$$E = \int (\phi_A(\underline{r})\rho_B(\underline{r}) + \phi_B(\underline{r})\rho_A(\underline{r}))dV$$

- В случае приближения “жестких тел” у нас 6 степеней свободы
- Для больших белков нам потребуется около 1000 точек по каждому направлению
- Сложность прямого вычисления составит $\sim 10^9$ таких интегралов
- Современные алгоритмы обычно суммируют несколько таких членов

Методы Докинга

- Ab-Initio
 - быстрые Фурье-трансформы - FFT, Polar Fourier Correlations (~ минуты - часы)
 - МС/МД - подвижные боковые цепи и основная цепь (~ дни)
 - HEX, AutoDock, ZDock, RosettaDock...
- Улучшение качества предсказаний
 - потенциалы с учетом предыдущих решенных структур
 - ItScore
- Входные данные
 - биохимия - мутации оснований на поверхности раздела
 - ЯМР, метки, ...
 - анализ последовательности - “correlated mutations”
 - структурные базы данных (докинг по гомологии)
 - HadDock

Почему FFT?

- могут быть одновременно посчитаны сразу несколько членов функции потенциальной энергии
- поиск с низким разрешением или поиск с переменной разрешающей способностью выполняется уменьшением количества коэффициентов в разложении

FFT в Декартовой системе

$$f_{A,l,m,n} = \begin{cases} 1 & : \text{surface of molecule} \\ \rho & : \text{core of molecule} \\ 0 & : \text{open space} \end{cases}$$

$$f_{B,l,m,n} = \begin{cases} 1 & : \text{inside molecule} \\ 0 & : \text{open space} \end{cases}$$

$$f_{C_{\alpha,\beta,\gamma}} = \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N f_{A,l,m,n} \times f_{B,l+\alpha,m+\beta,n+\gamma}$$

α, β, γ - вектора смещения В относительно А

N - количество узлов сетки по каждому направлению

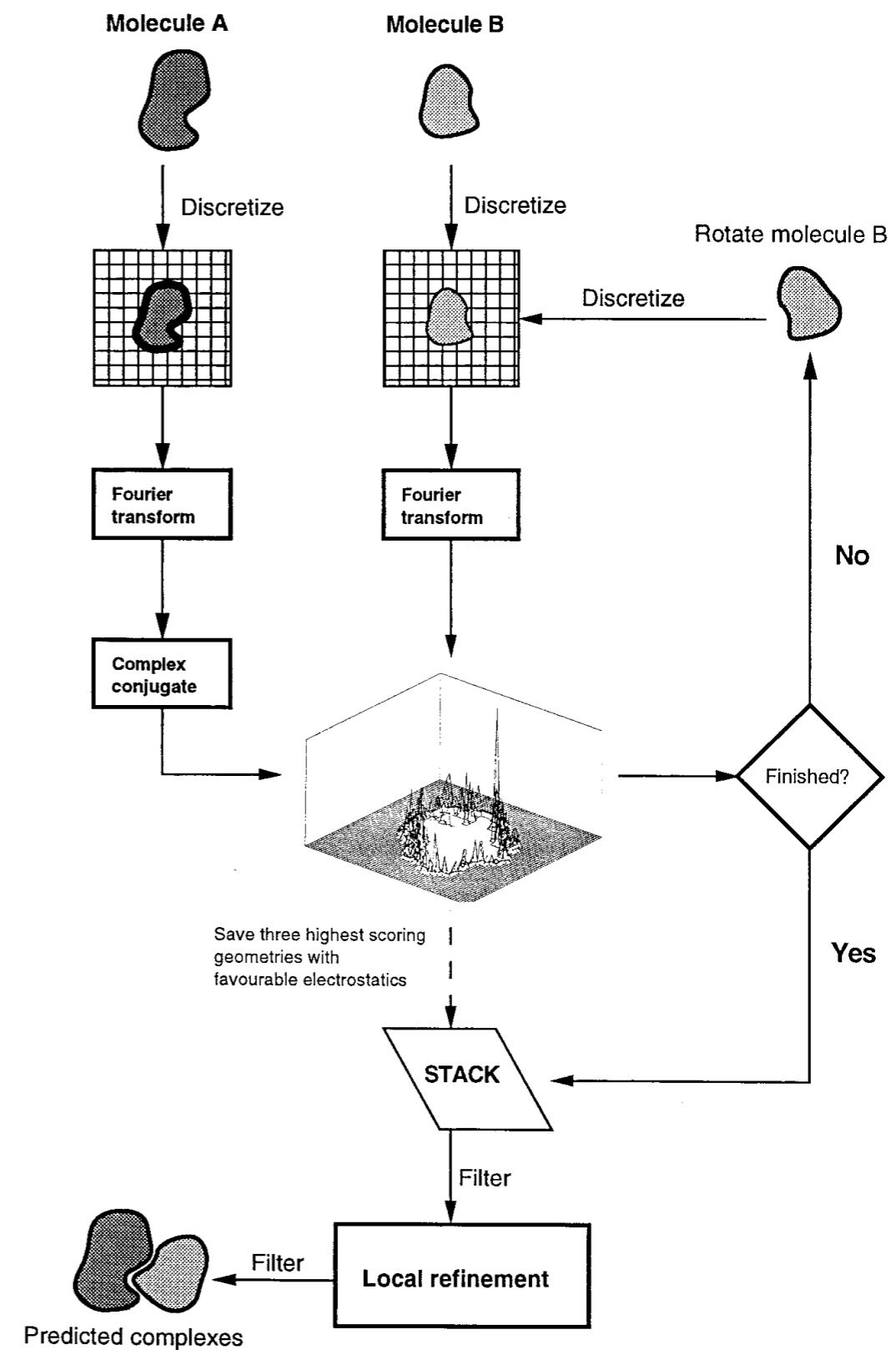
$$F_A = \text{DFT}(f_A)$$

$$F_B = \text{DFT}(f_B)$$

$$F_C = (F_A^*)(F_B)$$

$$f_C = \text{IFT}(F_C)$$

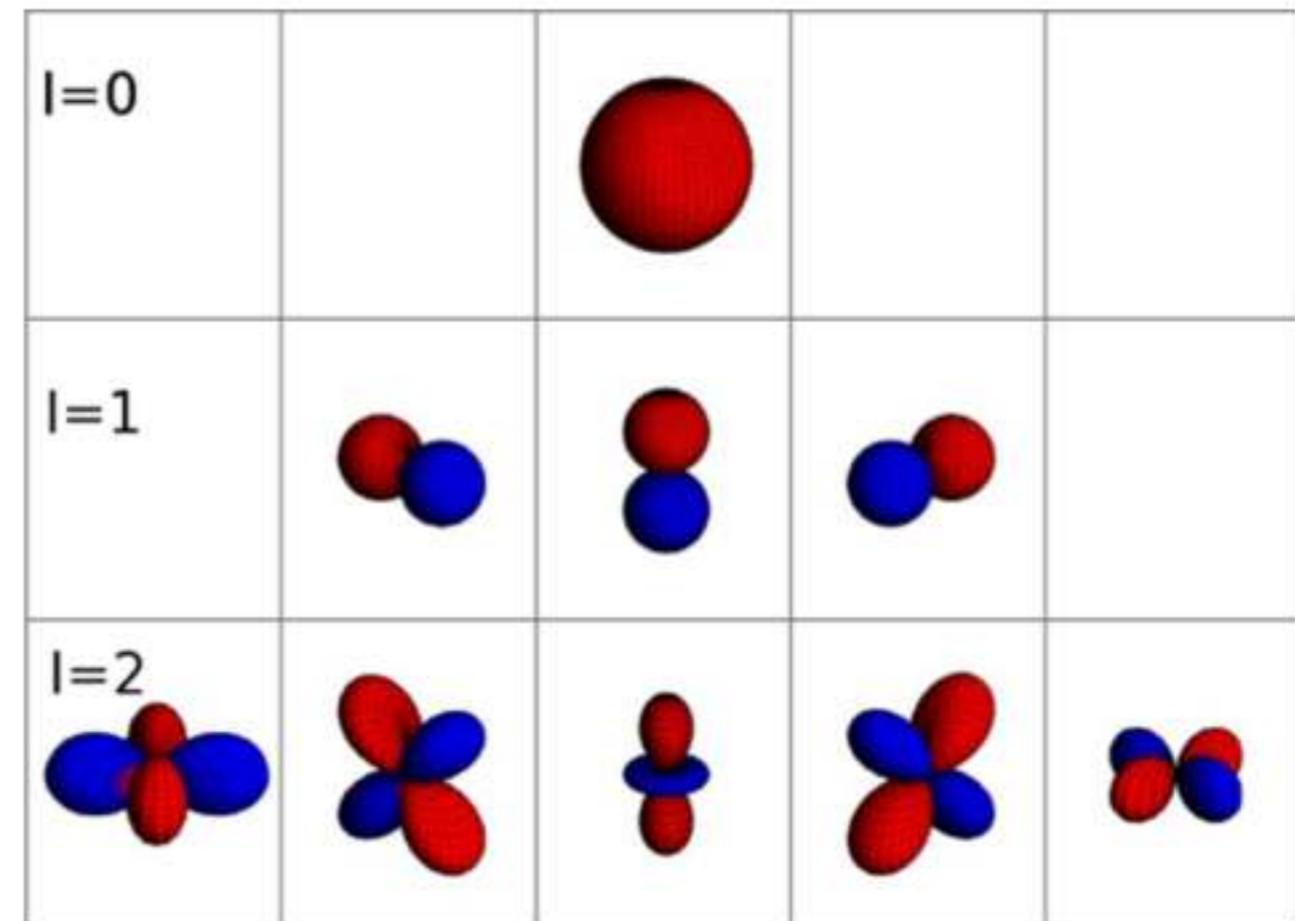
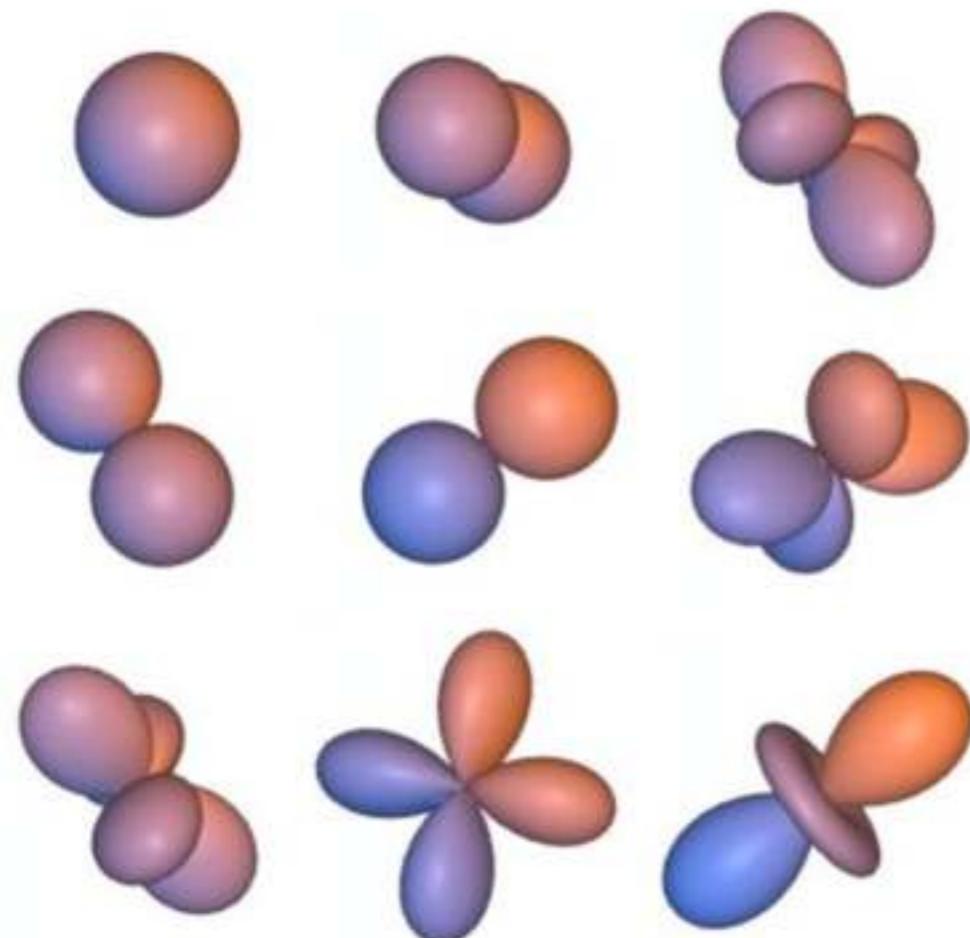
- Для каждой ориентации В требуется $O(N^6)$ вычислений функции корреляции прямым методом
- или $O(N^3 \log N^3)$ при помощи FFT



Почему не FFT в Декартовой системе?

- очень большая сетка при смещении одной молекулы относительно другой
- сетка коэффициентов должна быть пересчитана при повороте одной из молекул
- как правило, достаточно невысокое разрешение поиска $\sim 1\text{\AA}$

Сферические гармоники как базис



со страницы Дэвида Ритчи <http://hex.loria.fr/gallery/>

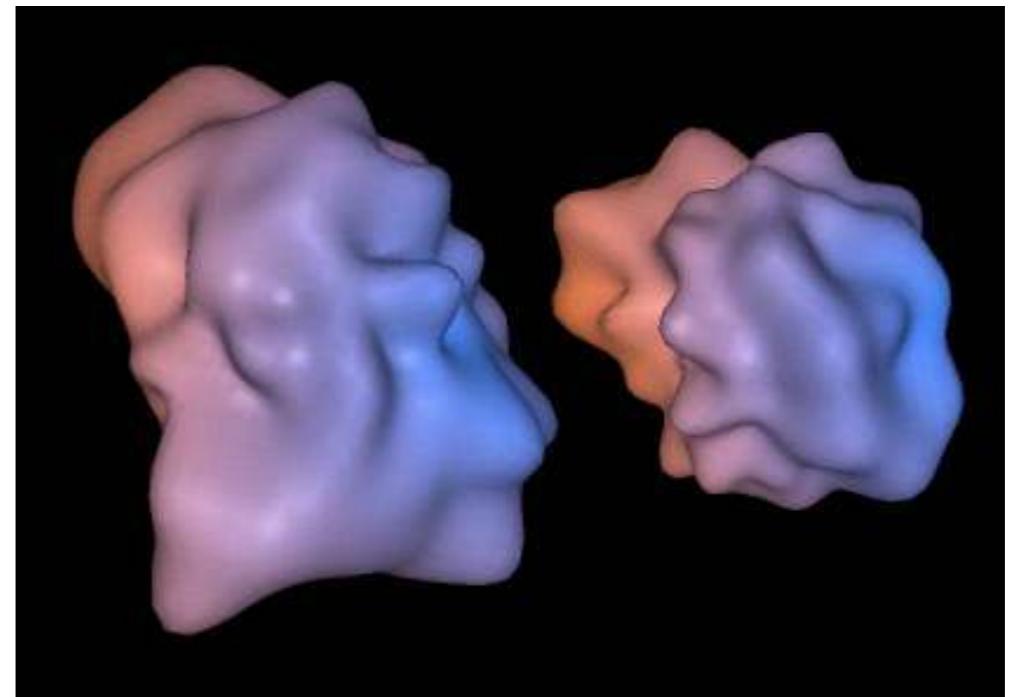
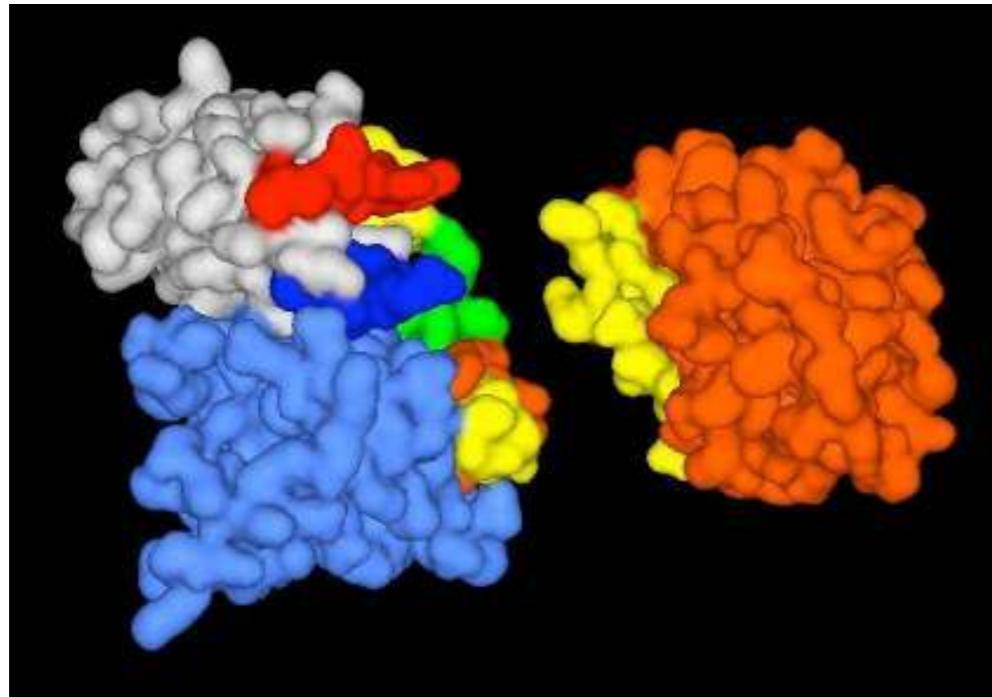
- ортогональность: $\int y_{lm}(\theta, \phi) y_{l'm'}(\theta, \phi) d\Omega = \delta_{ll'} \delta_{mm'}$

- поворот: $y_{lm}(\theta', \phi') = \sum_{m'=-l}^l R_{m'm}^{(l)}(\alpha, \beta, \gamma) y_{lm'}(\theta, \phi)$

Уравнения поверхности в 2D

Пример: 2D радиальные функции (256 коэффициентов)

$$r(\theta, \phi) = \sum_{l=0}^{15} \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \phi)$$



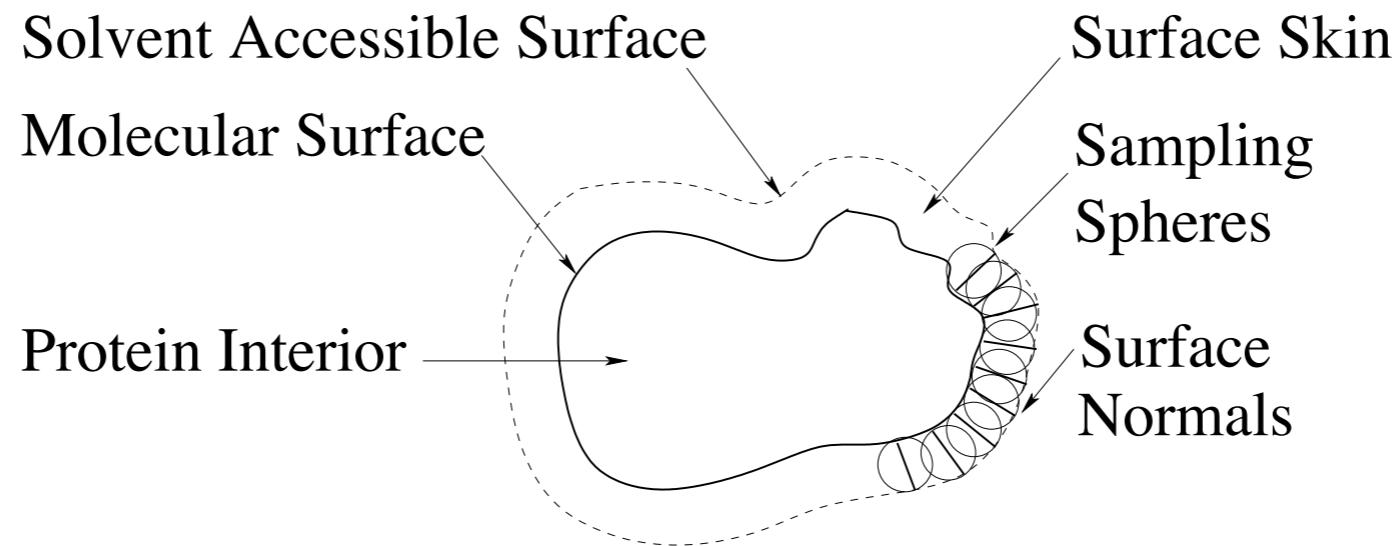
пример из презентаций Дэвида Ритчи <http://www.loria.fr/~ritchied/>

- ПОДХОДИТ ДЛЯ ПОИСКА СООТВЕТСТВИЙ В ПОХОЖИХ ПОВЕРХНОСТЯХ
- не подходит для докинга

Уравнения формы в 3D

из Ritchie & Kemp, Proteins 2000, 39, 178

- проецируем поверхность белка на сетку с шагом 0.75 Å ($\sim 10^6$ элементов)



поверхность: $\sigma(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{surface skin} \\ 0; & \text{otherwise} \end{cases}$

внутренность: $\tau(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{protein atom} \\ 0; & \text{otherwise} \end{cases}$

параметризуется как: $\sigma(\underline{r}) = \sum_{nlm}^N a_{nlm}^\sigma R_{nl}(r) y_{lm}(\theta, \phi)$

после интегрирования обеих частей: $a_{n'l'm'} = \int \sigma(\underline{r}) R_{n'l'}(r) y_{l'm'}(\theta, \phi) dV$.

коэффициенты вычисляются как: $a_{nlm}^\sigma \simeq \sum_c R_{nl}(r_c) y_{lm}(\theta_c, \phi_c) \Delta V$

- делаем это только один раз для каждого белка

Радиальные базисные функции

ГТО: форма

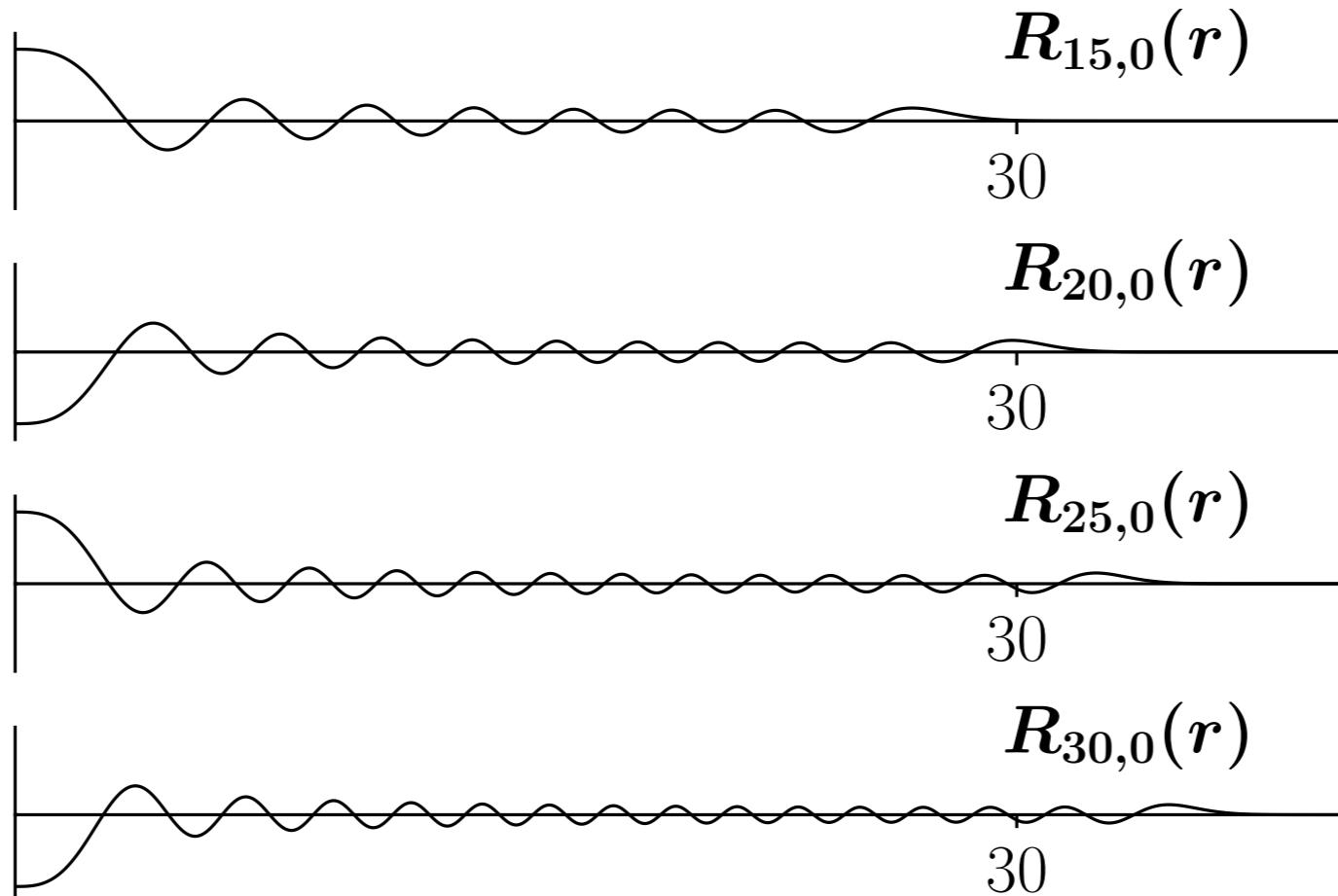
$$R_{nl}(r) = N_{nl}^{(q)} e^{-\rho/2} \rho^{l/2} L_{n-l-1}^{(l+1/2)}(\rho) \quad \rho = r^2/q, \quad q = 20$$

ETO: электростатика

$$R_{nl}(r) = N_{nl}^{(\Lambda)} e^{-\rho/2} \rho^l L_{n-l-1}^{(2l+2)}(\rho) \quad \rho = 2\Lambda r, \quad \Lambda = 1/2$$

Ортогональность:

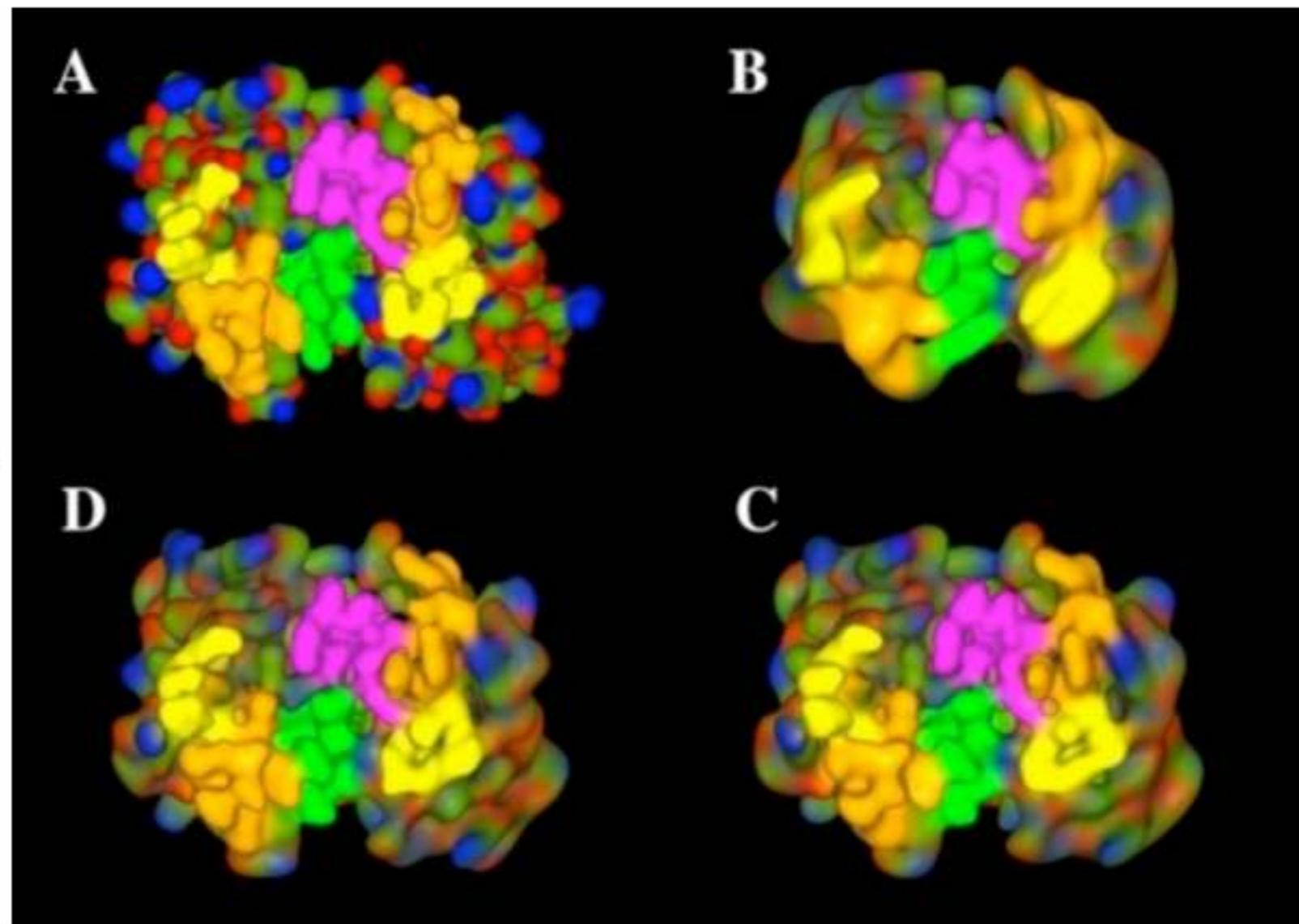
$$\int_0^\infty R_{nl}(r) R_{n'l}(r) r^2 dr = \delta_{nn'}$$



Реконструкция поверхности

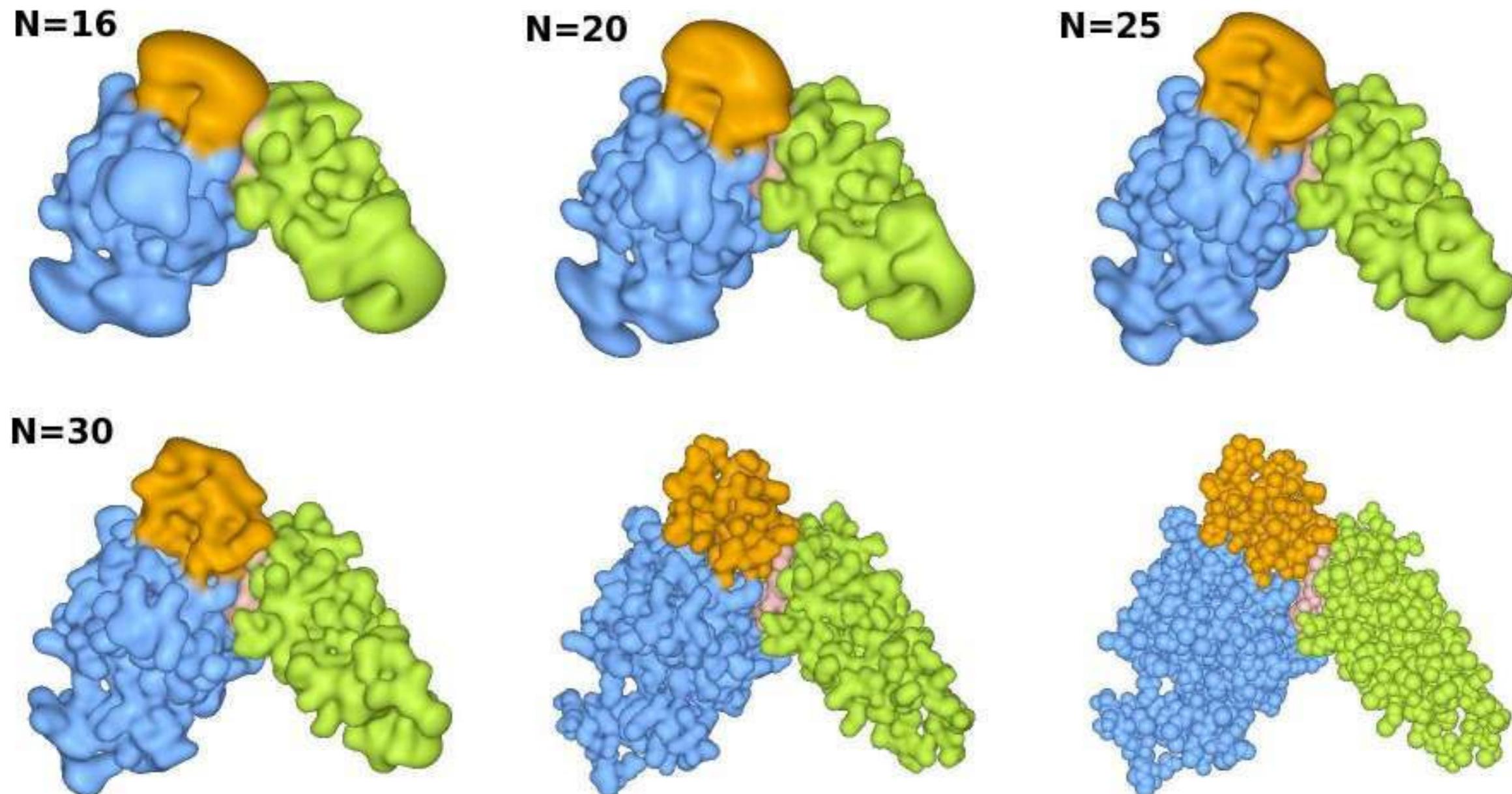
Рисунок Порядок Коэффициенты

A	Gaussians	-
B	N = 16	1,496
C	N = 25	5,525
D	N = 30	9,455



из DW Ritchie (2003) Proteins Struct. Funct. Bionf. 52 98–106

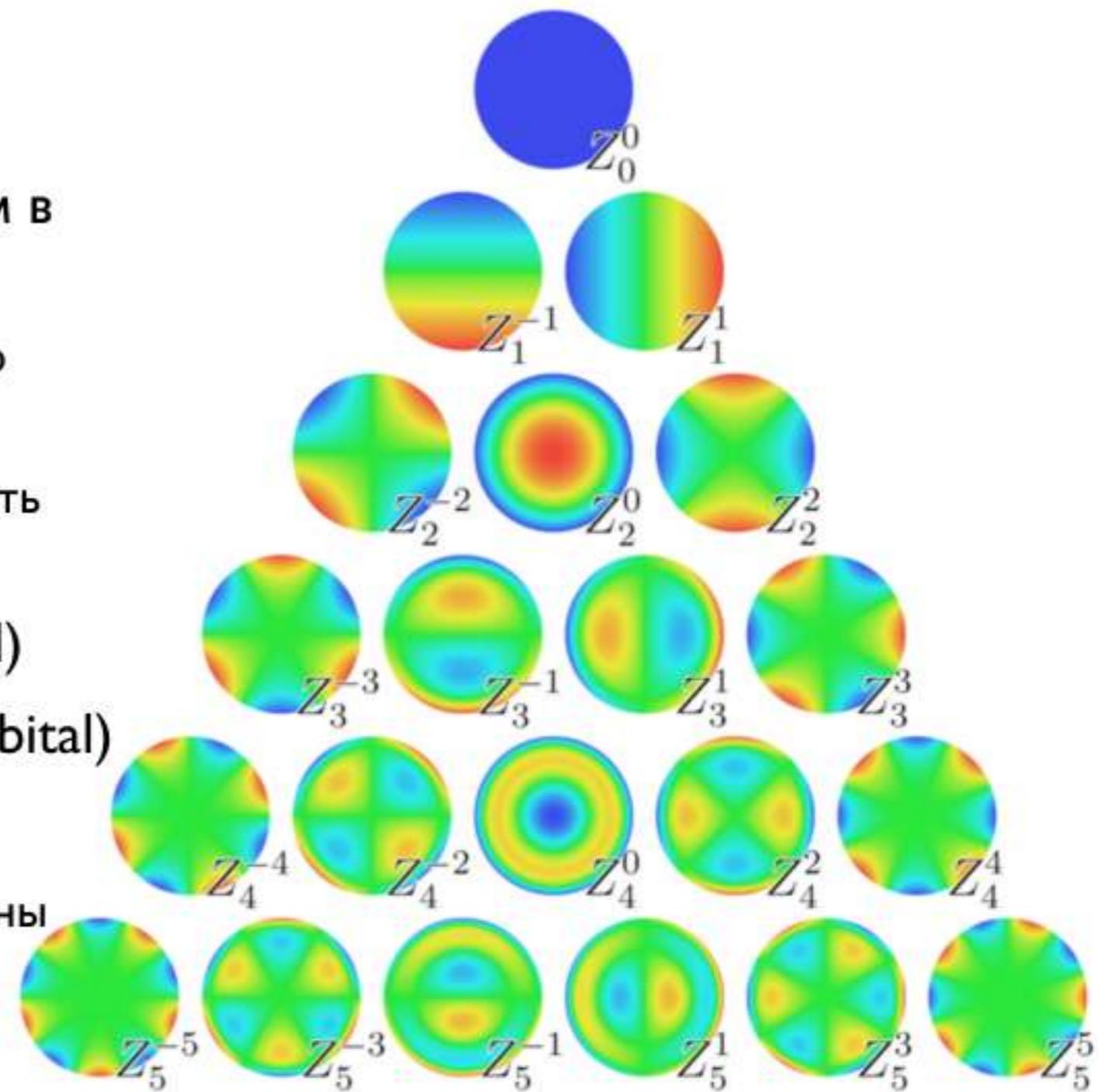
3D Реконструкция Формы – CAPRI T2I: Orc1/Sir1



из DW Ritchie (2008) Curr. Prot. Pep. Sci. 9(1) 1-15

Уравнения формы в 3D

- форма, электростатические и другие свойства молекулы представляются разложением в ортонормированном базисе
 - ориентация раскладывается по сферическим гармоникам
 - радиальные функции могут быть различны:
 - STO (Slater Type Orbital)
 - GTO (Gaussian Type Orbital)
 - ...
 - также используются многочлены Цернике (Zernike polynomials)



$$R_n^m(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k (n-k)!}{k! ((n+m)/2 - k)! ((n-m)/2 - k)!} \rho^{n-2k}$$

Корреляции

из DW Ritchie (2005) J. Appl. Cryst. 38 808–818

Поворот:

$$\hat{R}(\alpha, \beta, \gamma) \sigma_A(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\sigma'} R_{nl}(r) y_{lm}(\theta, \phi)$$

Повернутые коэффициенты:

$$a_{nlm}^{\sigma'} = \sum_{m'=-l}^l R_{mm'}^{(l)}(\alpha, \beta, \gamma) a_{nlm'}^{\sigma}$$

Сдвиг:

$$\hat{T}_z(R) \sigma_A(\underline{r}) = \sum_{nlm}^N a_{nlm}^{\sigma''} R_{nl}(r) y_{lm}(\theta, \phi)$$

Сдвинутые коэффициенты:

$$a_{nlm}^{\sigma''} = \sum_{n'l'}^N T_{nl,n'l'}^{(|m|)}(R) a_{n'l'm}^{\sigma}$$

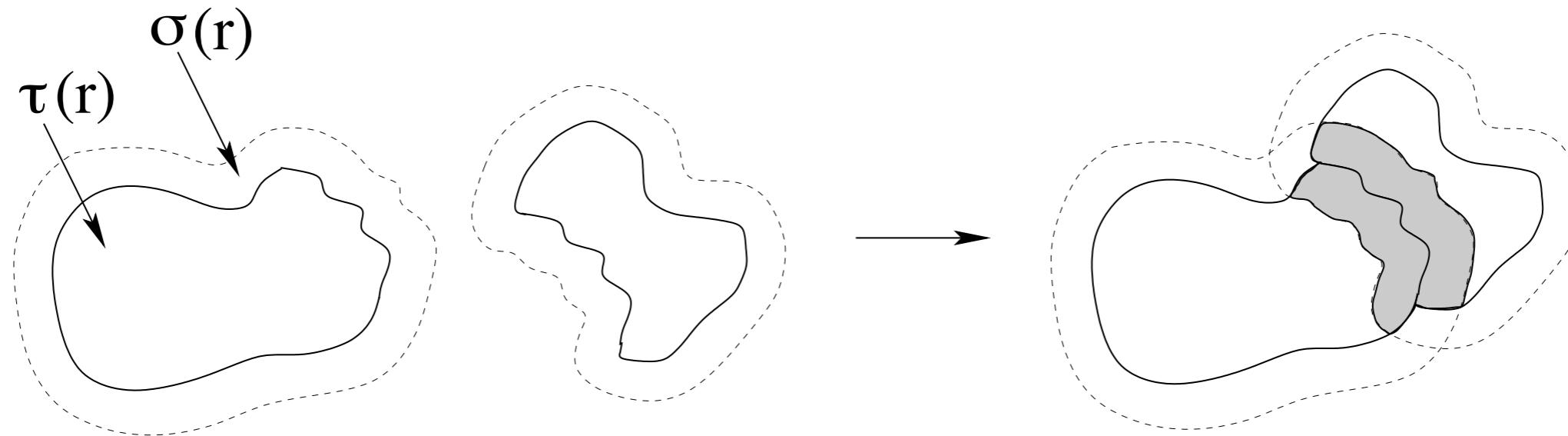
Итого:

$$\int \sigma'_A(\underline{r}) \tau''_B(\underline{r}) dV = \sum_{nlm}^N a_{nlm}^{\sigma'} b_{nlm}^{\tau''} \quad etc. \quad O(N^7)$$

Пространство поиска:

$\sim 10^9$ ориентаций ($\sim 10^6$ ориентаций/сек)

Докинг при помощи 3D полярных функций Фурье



плотности:

$$\sigma(\underline{r}) = \sum_{nlm}^N a_{nlm}^\sigma R_{nl}(r) y_{lm}(\theta, \phi) \quad \tau(\underline{r}) = \sum_{nlm}^N a_{nlm}^\tau R_{nl}(r) y_{lm}(\theta, \phi)$$

выгодные:

$$\int (\sigma_A(\underline{r}_A) \tau_B(\underline{r}_B) + \tau_A(\underline{r}_A) \sigma_B(\underline{r}_B)) dV$$

невыгодные:

$$\int \tau_A(\underline{r}_A) \tau_B(\underline{r}_B) dV$$

результат:

$$S_{AB} = \int (\sigma_A \tau_B + \tau_A \sigma_B - Q \tau_A \tau_B) dV \quad \text{Penalty Factor: } Q = 11$$

Матрицы сдвига

при помощи сферических трансформ Бесселя:

$$\tilde{R}_{nl}(\beta) = \sqrt{\frac{2}{\pi}} \int_0^\infty R_{nl}(r) j_l(\beta r) r^2 dr; \quad R_{nl}(r) = \sqrt{\frac{2}{\pi}} \int_0^\infty \tilde{R}_{nl}(\beta) j_l(\beta r) \beta^2 d\beta$$

можно показать, что:

$$T_{n'l',nl}^{(|m|)}(R) = \sum_{k=|l-l'|}^{l+l'} A_k^{(ll'|m|)} \int_0^\infty \tilde{R}_{nl}(\beta) \tilde{R}_{n'l'}(\beta) j_k(\beta R) \beta^2 d\beta$$

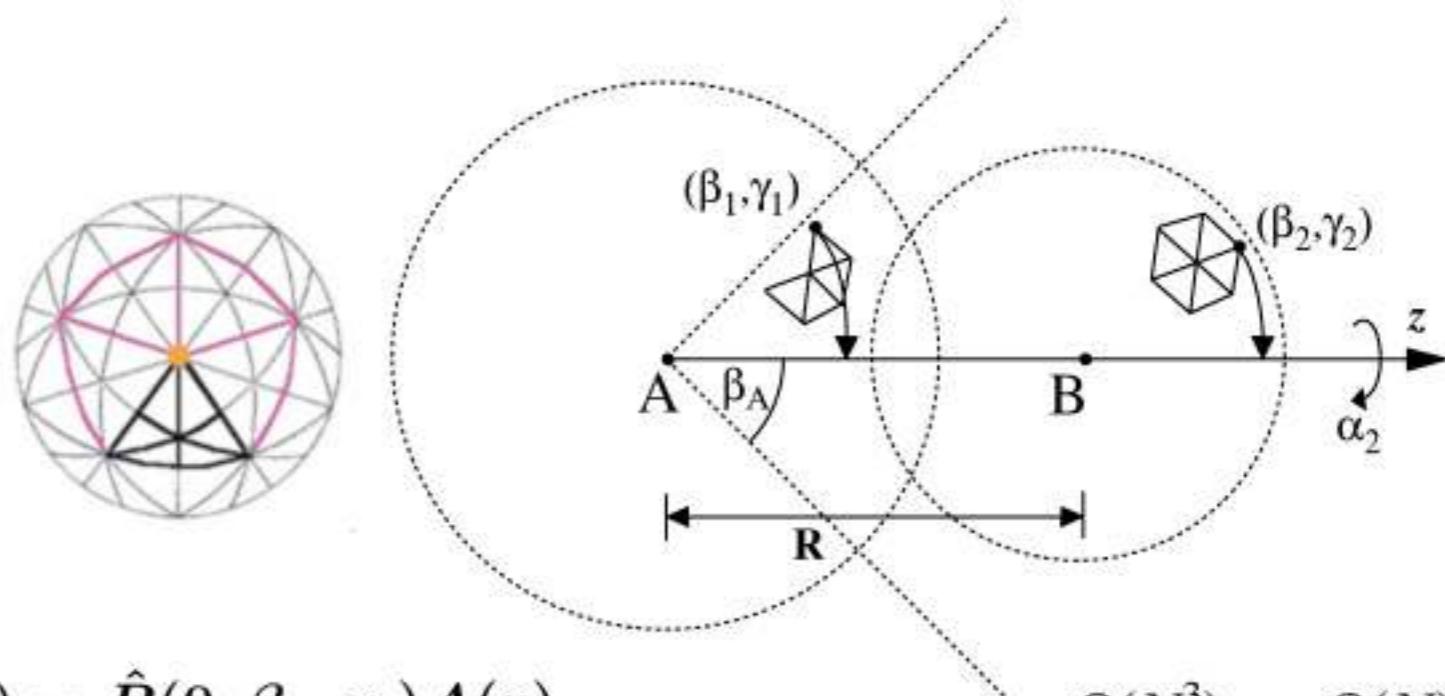
где:

$$A_k^{(ll'|m|)} = (-1)^{\frac{k+l'-l}{2}+m} (2k+1) [(2l+1)(2l'+1)]^{1/2} \begin{pmatrix} l & l' & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & k \\ m & \bar{m} & 0 \end{pmatrix}$$

- Можно вывести сдвиг аналитически для GTO и ETO базисных функций
- Требует математические библиотеки высокой точности
- Вычисляется один раз для $R = 1, 2, 3, \dots 50 \text{ \AA}$ и потом хранится на диске ($\sim 200 \text{ Mb}$)

6D ПОИСК В МОЛЕКУЛЯРНОМ ДОКИНГЕ

получаем 4 угла вращения из икосаэдра:



поворачиваем А ($\times 812 @ 7.5^\circ$): $A'(\underline{r}) = \hat{R}(0, \beta_1, \gamma_1)A(\underline{r})$

$O(N^3) \times O(N)$

сдвигаем А ($\times 50 @ 0.75\text{\AA}$): $A''(\underline{r}) = \hat{T}_z(-R)A'(\underline{r})$

$O(N^3) \times O(N^2)$

поворачиваем В ($\times 812 @ 7.5^\circ$): $B'(\underline{r}) = \hat{R}(0, \beta_2, \gamma_2)B(\underline{r})$

$O(N^3) \times O(N)$

прокручиваем В ($\times 64 @ 5.6^\circ$): $B''(\underline{r}) = \hat{R}(\alpha_2, 0, 0)B'(\underline{r})$

$O(N^3) \times O(N)$

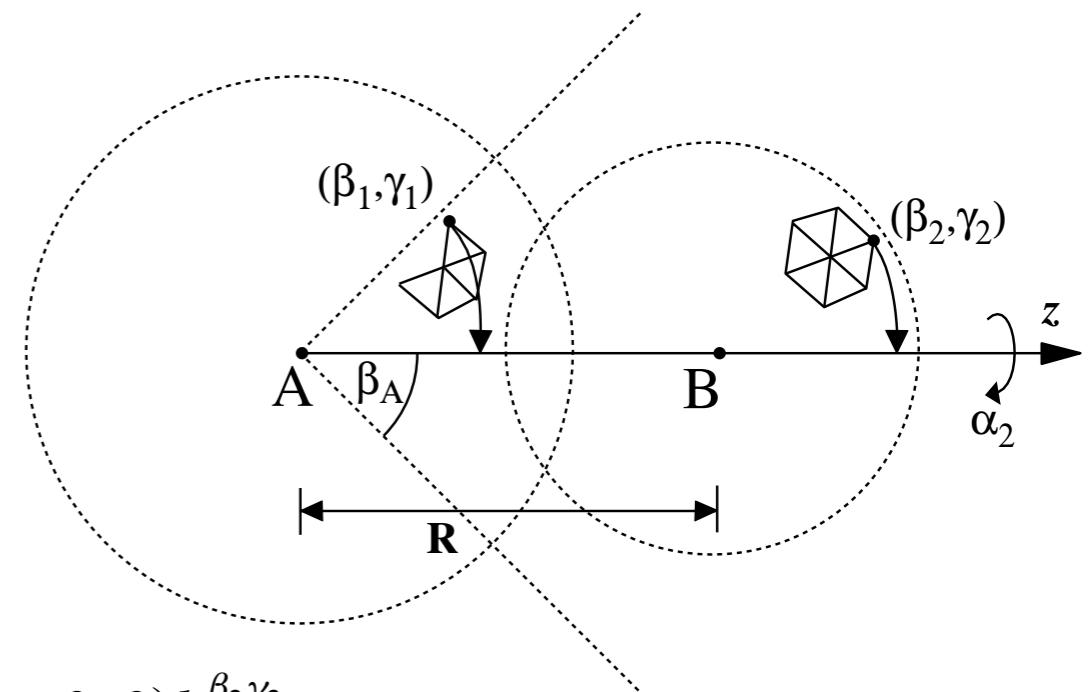
ID FFT:

$$S_{AB}(\alpha_2) = \sum_{m=1-N}^{N-1} P_m \cos m\alpha_2 + Q_m \sin m\alpha_2$$

пространство поиска:

$812 \times 50 \times 812 \times 64 \simeq 2 \times 10^9 (\sim 10^6/s \text{ on a 1GHz PIII Xeon})$

FFT корреляция



прокручиваем В ($\times 64 @ 5.6^\circ$):

$$b_{nlm}^{\alpha_2 \beta_2 \gamma_2} = \sum_{m'} R_{mm'}^{(l)}(\alpha_2, 0, 0) b_{nlm'}^{\beta_2 \gamma_2}$$

$$= \sum_{m'} b_{nlm'}^{\beta_2 \gamma_2} \cos m' \alpha_2 + b_{nlm'}^{\beta_2 \gamma_2} \sin m' \alpha_2. \quad O(N^3) \times O(N)$$

если посчитать корреляцию:

$$\int \sigma_A \tau_B dV = \sum_{nlm}^N a_{nlm}^\sigma b_{nlm}^{\tau} \quad O(N^3)$$

собираем коэффициенты:

$$P_m^{\beta_1 \gamma_1 R \beta_2 \gamma_2} = \sum_{nl} a_{nlm}^{R \beta_1 \gamma_1} b_{nlm}^{\beta_2 \gamma_2} \quad O(N) \times O(N^2)$$

$$Q_m^{\beta_1 \gamma_1 R \beta_2 \gamma_2} = \sum_{nl} a_{nlm}^{R \beta_1 \gamma_1} b_{nlm}^{\beta_2 \gamma_2} \quad O(N) \times O(N^2)$$

напрямую:

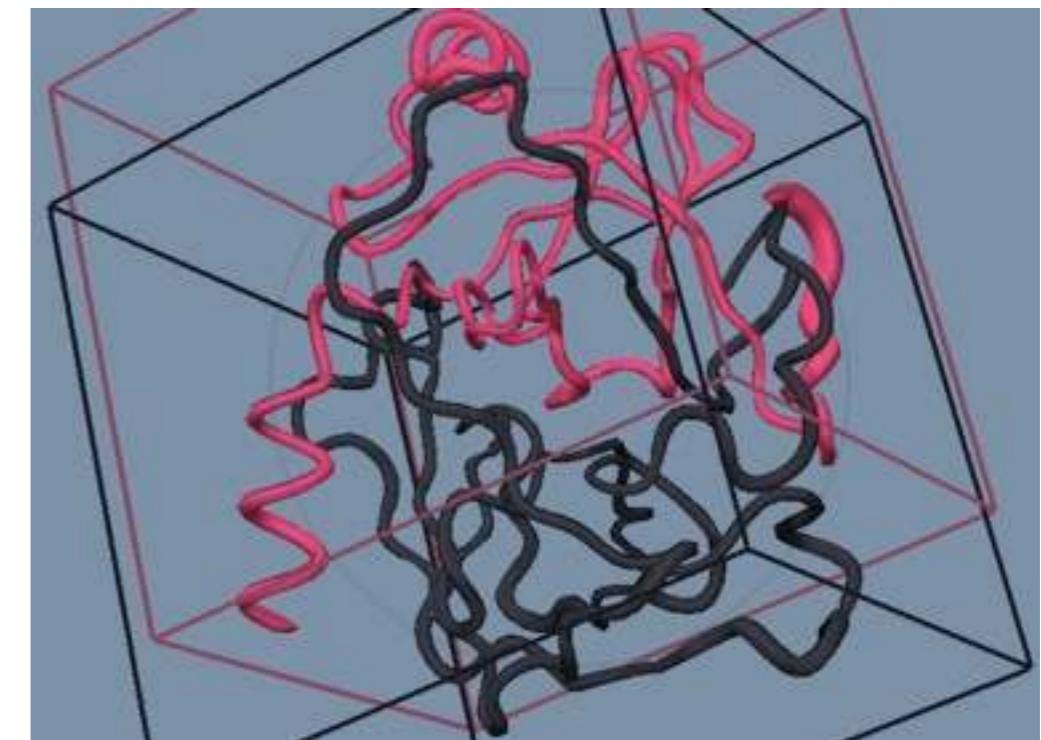
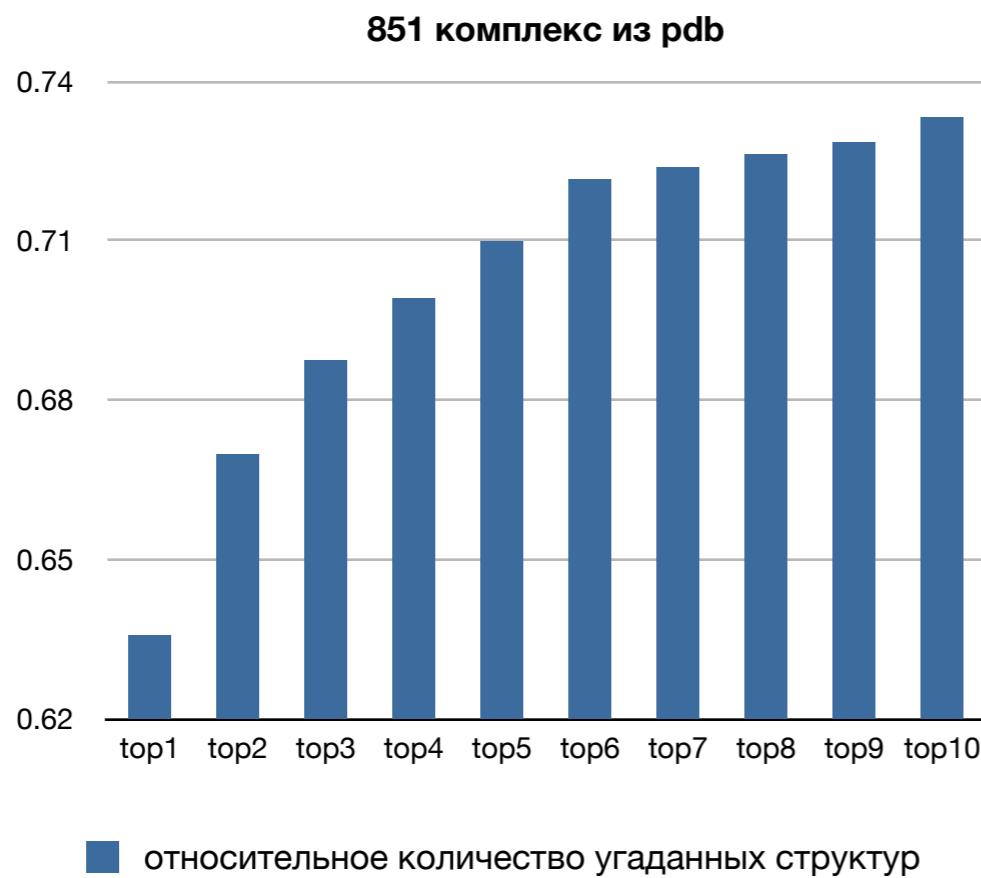
$$S_{AB}(\alpha_2) = \sum_{m=1-N}^{N-1} P_m \cos m \alpha_2 + Q_m \sin m \alpha_2 \quad O(NM)$$

ID FFT:

для M образцов:

$$O(M \log M)$$

пример для НЕХ

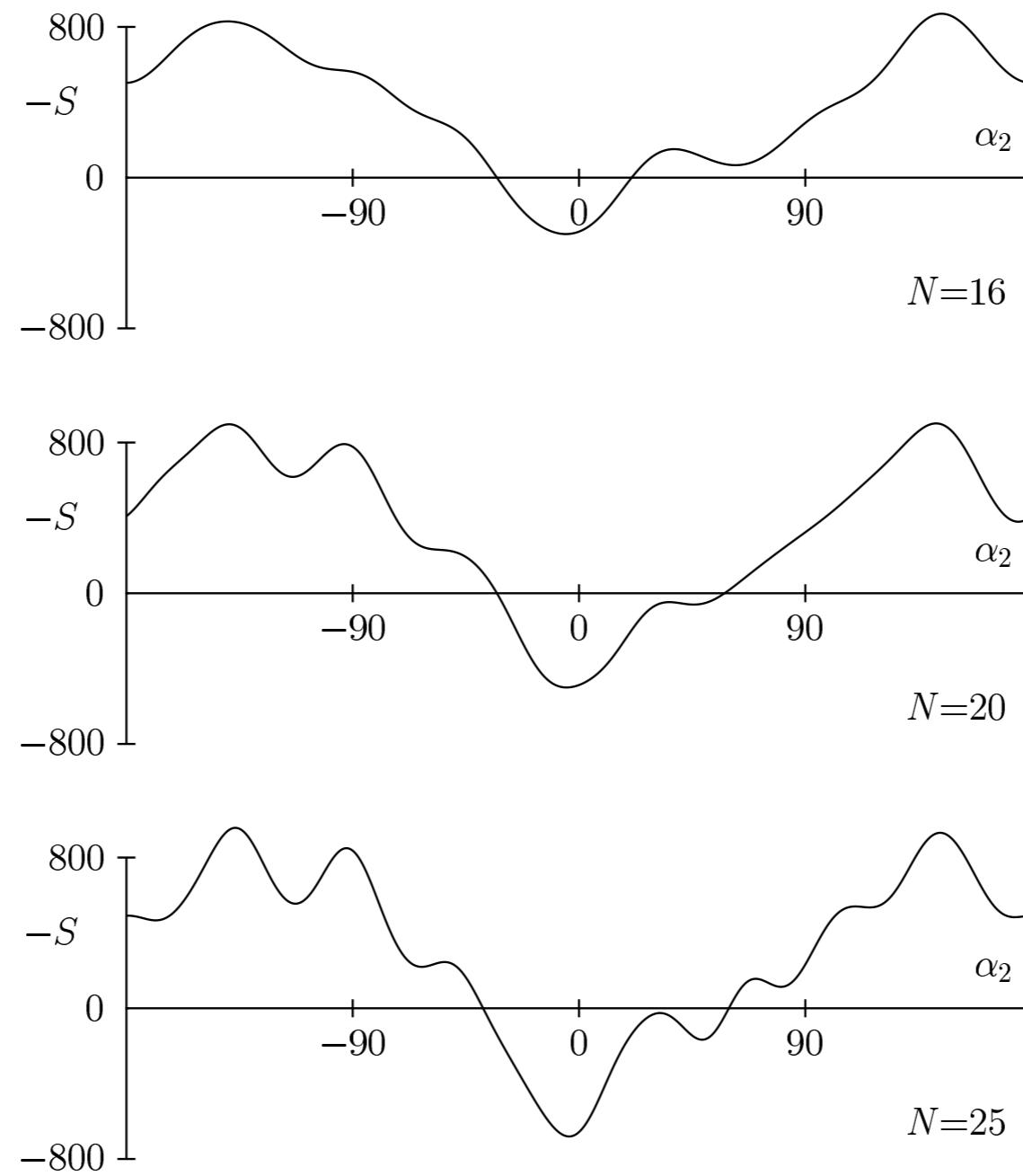


- 851 комплекс из PDB
- разрешение лучше чем 2.5 Å
- на границе раздела только стандартные основания
- каждая цепь по крайней мере 10 остатков
- гомология меньше 70 %
- шаг поиска 2.5 Å

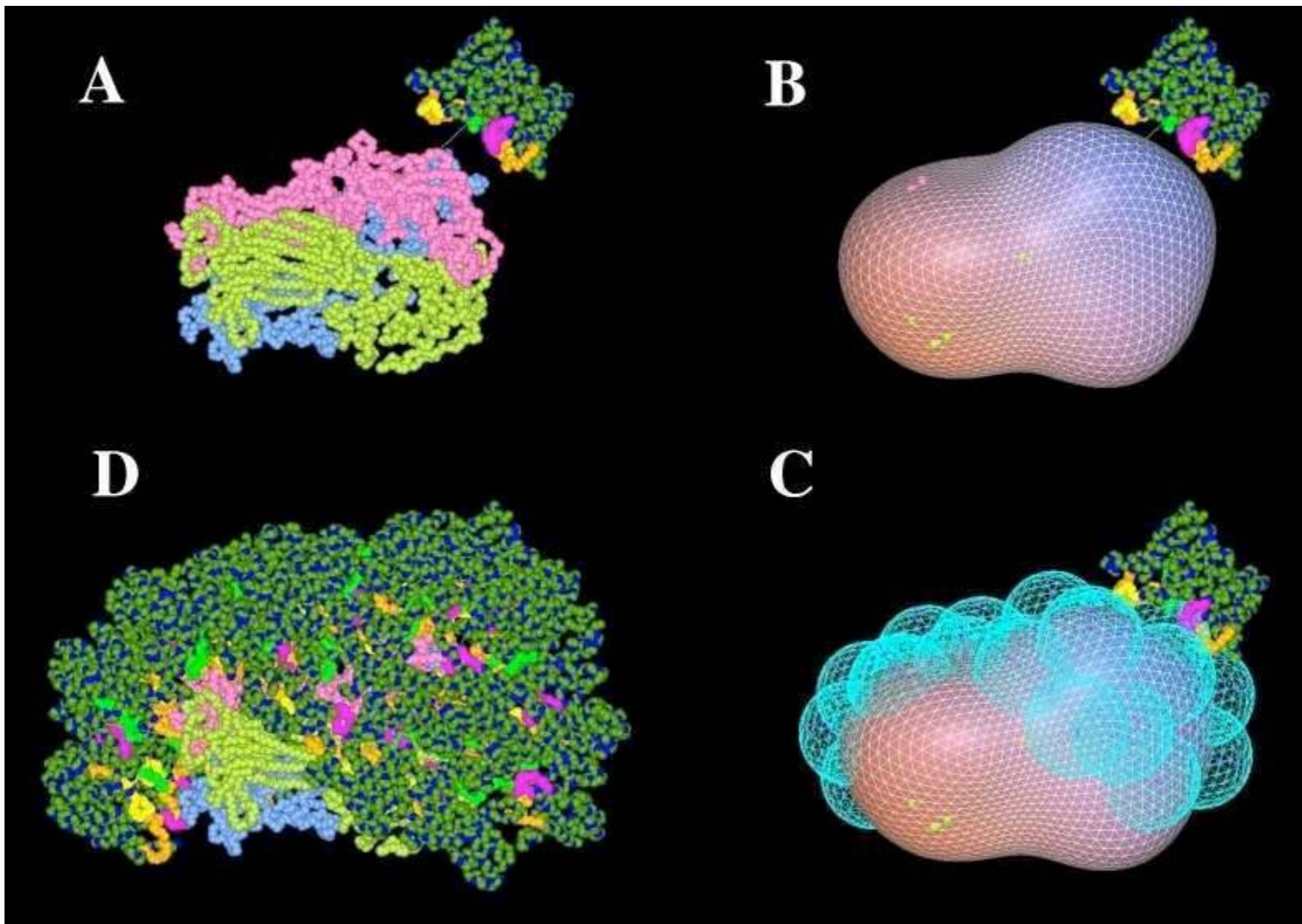
еще пример для НЕХ

Case	$N = 16$		$N = 20$		$N = 25$	
	Top	RMS	Top	RMS	Top	RMS
SIC	3,407	0.00	2	0.22	1	0.82
KAI	17	0.41	3	0.69	7	0.81
PTC	132	0.52	2	0.48	1	0.48
CGI	1	0.38	1	0.38	1	0.38
CHO	1	0.45	1	0.55	1	0.55
BGS	1	0.82	1	0.82	1	0.88
GGI	1	2.47	1	0.90	1	0.90
TET	5	1.48	1	1.16	1	1.03
FPT	102	1.04	1	0.42	1	0.42
IGF	3	0.71	1	0.77	1	0.77
JEL	4,867	0.81	1,060	0.81	2	0.81
BQL	524	1.85	12	0.96	1	0.39
HFL	318	1.01	5	1.00	1	1.00
HFM	7	2.19	27	1.09	10	1.09
VFB	8,344	1.49	216	0.20	9	0.20
MLC	1,401	0.00	116	0.00	187	0.84
MEL	9,898	1.03	27	1.03	3	1.03
JHL	385	0.62	8	0.38	1	1.08
FBI	14	1.09	1	1.09	1	0.38
NCA	68	1.53	1	0.32	1	0.32
NMB	160	2.43	1,630	1.39	1,009	1.39
NSN	19,992	1.11	716	0.75	1,130	2.29
IAI	1,381	1.48	111	0.37	20	1.39
DVF	11,145	0.00	88	1.38	49	0.44
KB5	140	0.34	1	0.34	78	1.38
IGC	1,328	1.74	269	0.81	1	0.34

Результат корреляции формы как функция угла вращения α_2 комплекс HyHel-5/Lysozyme

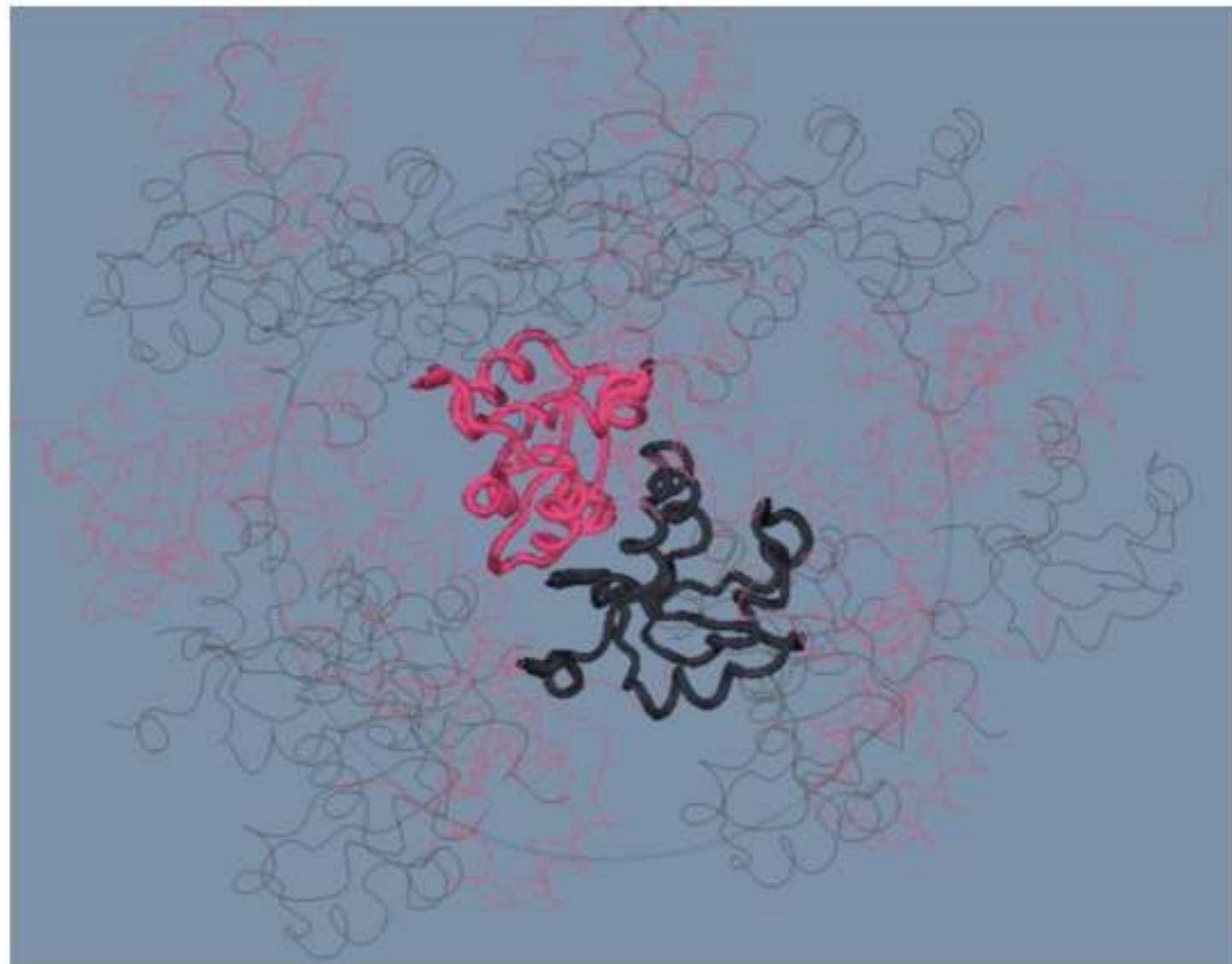
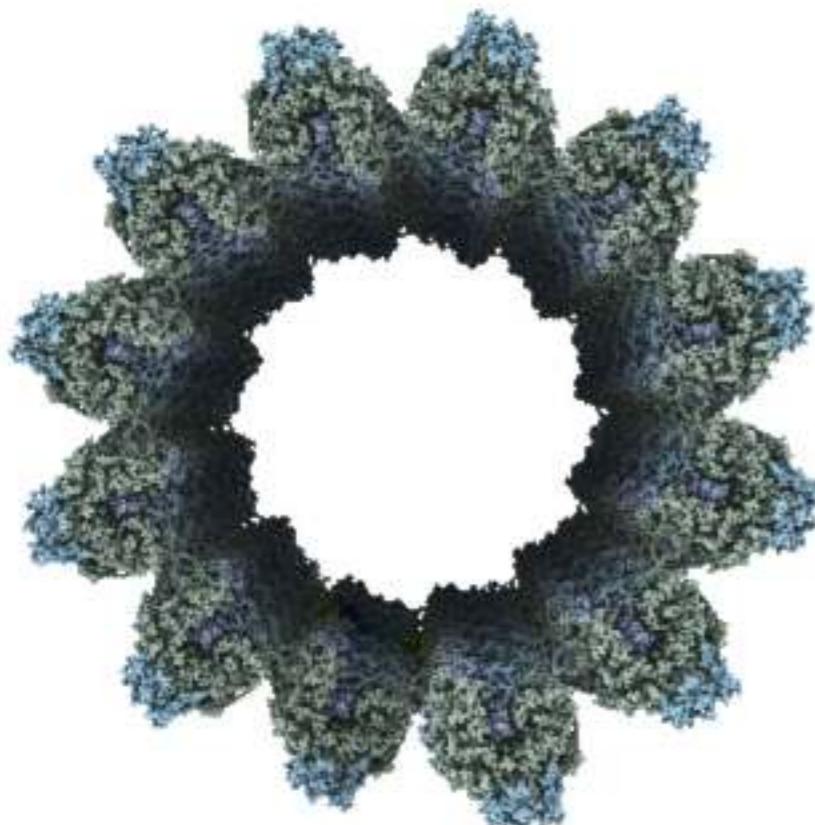


Макро-Докинг



- До сих пор нету эффективного метода, применимого ко всем большим структурам
- Обычно, большая структура разбивается на много маленьких, которые обсчитываются отдельно

Докинг с симметрией



- Нужен эффективный метод, позволяющий учитывать все типы симметрий при докинге
- На данный момент существуют алгоритмы, учитывающие только некоторые группы симметрий (некоторые из точечных групп)

CAPRI – Critical Assessment of Predicted Interactions

- начал в 2001 году вслед за CASP с 19ю группами и 7ю мишениями
- по крайней мере один белок в комплексе дан в несвязанной форме
- разрешен любой метод предсказания (гомология, литература,..)
- сейчас больше 60 групп (ни одной из России), раунд 21
- два последних раунда организованы Дэвидом Бейкером (Rosetta) для исследования механизмов предсказания
- 3 секции - предсказания, серверы и те, кто улучшают предсказания

Target	Receptor	Ligand	Type	Complex	Lab
1	HPr Kinase	HPr	U/U	Fieulaine et al.	Janin
2	Rotavirus VP6	MCV	U/B	Vaney et al.	Rey
3	Hemagglutinin	HC63	U/B	Barbey-Martin et al.	Knossow
4	α -Amylase	AMD10	U/B	Desmyter et al.	Cambillau
5	α -Amylase	AMB7	U/B	Desmyter et al.	Cambillau
6	α -Amylase	AMD9	U/B	Desmyter et al.	Cambillau
7	SpeA	TCR 14.3.D	U/U	Sundberg et al.	Mariuzza

результаты CAPRI (1-7 мишени)

Predictor	Software	Algorithm	T1	T2	T3	T4	T5	T6	T7
Abagyan	ICM	FF			**			***	**
Camacho	CHARMM	FF	*					***	***
Eisenstein	MolFit	FFT	*	*					***
Sternberg	FTDOCK	FFT		*				**	*
Ten Eyck	DOT	FFT	*	*				**	
Gray		MC						**	***
Ritchie	Hex	SPF			**			***	
Weng	ZDOCK	FFT		**					**
Wolfson	BUDDA/PPD	GH	*						***
Bates	Guided Docking	FF	-	-	-				***
Palma	BIGGER	GF	-	-	-			**	*
Gardiner	GAPDOCK	GA	*	*	-	-	-	-	-
Olson	Surfdock	SH	*			-	-	-	-
Valencia		ANN	*	-	-	-	-	-	-
Vakser	GRAMM	FFT	*		-	-	-	-	-

* low, ** medium, *** high accuracy prediction; — no prediction

результаты CAPRI (8-19 мишени)

CAPRI Results: Targets 8–19

Predictor	Software	T8	T9	T10	T11	T12	T13	T14	T18	T19
Abagyan	ICM	**		*	**	***	*	***	**	**
Wolfson	PatchDock	**	*	*	*	*	-	**	**	*
Weng	ZDOCK/RDOCK	**		*	***	***	***	***	**	**
Bates	FTDOCK	*		*	**	*		**	**	*
Baker	RosettaDock	-			**	***	**	***		***
Camacho	SmoothDock	**				***	***	**	**	*
Gray	RosettaDock	***	-	-	**	***				**
Bonvin	Haddock	-	-	**	**		***	***		
Comeau	ClusPro	**				***	*			*
Sternberg	3D-DOCK	**			*	*		**		*
Eisenstein	MolFit	***			*	***		**		
Ritchie	Hex				**	***	*	*		
Zhou		-	-		-	***	**	*	*	
Ten Eyck	DOT					***	***	**		
Zacharias	ATTRACT	**		-	-	-	-	***		**
Valencia		*			*	*	-			-
Vakser	GRAMM	-	-		-	-	-	**	**	
Umeyama					**	*				
Kaznessis		-	-			***				
Fano	Grid-Hex	-	-		*					

CAPRI: результаты раундов 13-19

- легкие мешени хорошо угадываются из-за маленьких конформационных изменений, такие как комплексы фермент-ингибитор
- иерархия
 - упрощенный поиск и/или поиск с твердыми телами
 - выбор интересных мест
 - доводка предсказаний
 - выбор лучших моделей
- Анализ Нормальных Мод для подвижности основной цепи
- биологическая информация иногда помогает
- докинг РНК - никто не предсказал несвязанную структуру, несколько групп предсказали связанную структуру

Потенциалы для молекулярных докинга и фолдинга

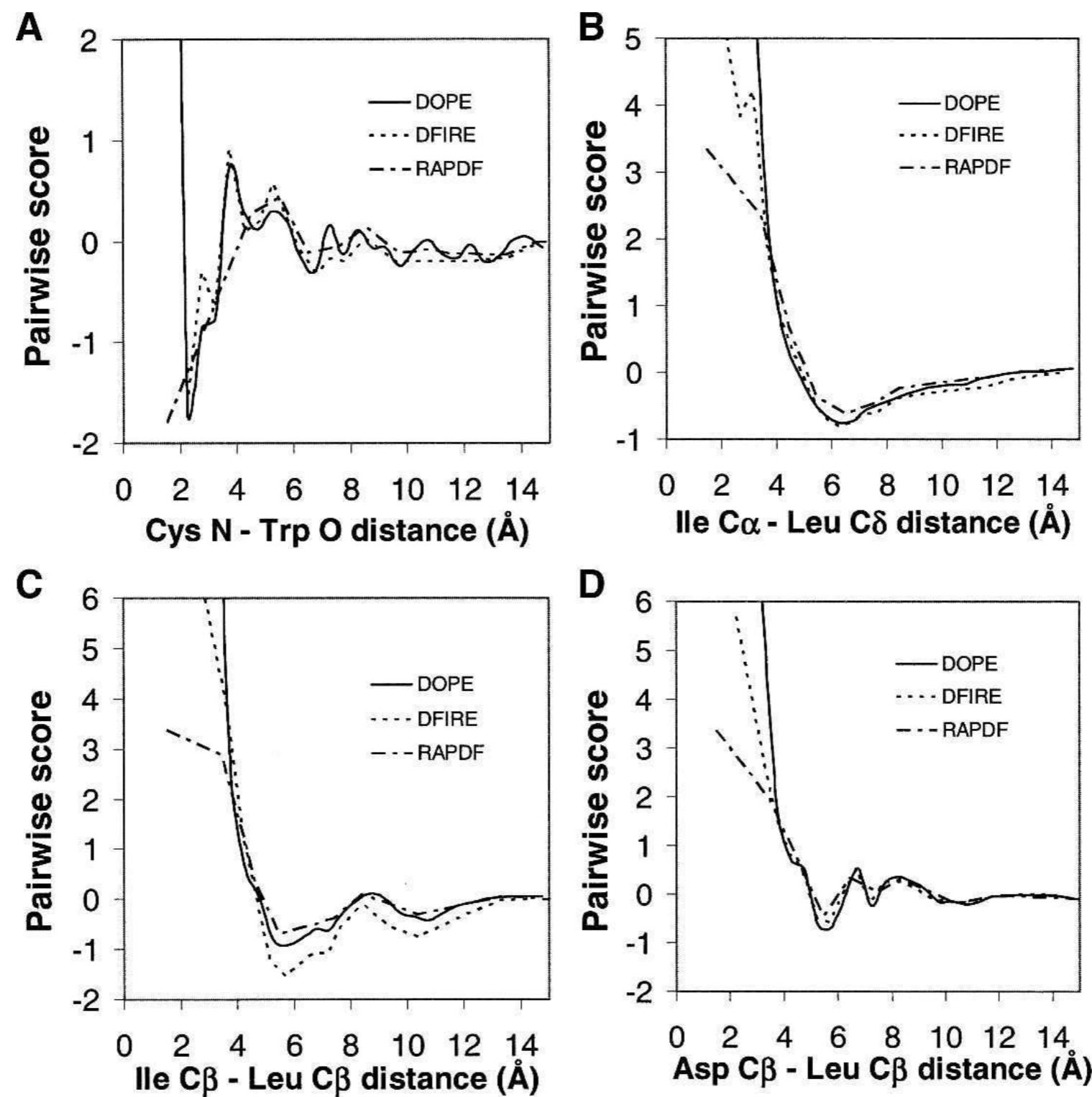
Потенциал DOPE (MODELLER)

$$G(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) \approx -k_B T \sum_{i \neq j}^N \ln g_{i,j}^{(2)}(r) \\ = \sum_{i \neq j}^N \bar{u}_{i,j}(r)$$

$$\bar{u}_{i,j}(r) = -k_B T \ln\left(\frac{p_{m,n}(r)}{p_{m,n}^{REF}(r)}\right) \approx -k_B T \ln\left(\frac{N_{m,n}^{OBS}(r)}{N_{m,n}^{REF}(r)}\right)$$

	DFIRE	Rosetta	ModPipe-Pair	ModPipe-Surf	ModPipe-Comb	DOPE
<i>4state-reduced</i>	NR	NR	NR	NR	NR	NR
1ctf	1	1	1	1	1	1
1r69	1	2	1	17	1	1
1sn3	1	1	1	7	1	1
2cro	1	5	1	103	1	1
3icb	4	6	15	33	8	1
4pti	1	1	1	71	1	1
4rxn	1	1	1	18	1	1
Correct	6	4	6	1	6	7
<i>fisa</i>						
1fc2	254	158	491	1	453	375
1hdd-C	1	90	293	18	135	1
2cro	1	26	11	146	19	1
4icb	1	1	196	2	167	1
Correct	3	1	0	1	0	3
<i>fisa_casp3</i>						
1bg8-A	1	1068	1	1180	282	1
1bl0	1	960	4	912	86	1

Пример для некоторых парных потенциалов



из M. Shen and A. Sali, Protein Sci 2006 15: 2507-2524.

Потенциал EMPIRE

$$\begin{aligned} E = & - 0.143 \times S_{\text{contact}} + 0.724 \times V_{\text{overlap}} + 1.72 \times E_{\text{hbond}} \\ & + 28.6 \times E_{\text{elec}} - 0.0467 \times \Delta S_{\text{pho}} + 0.0042 \times \Delta S_{\text{phi}} \\ & + 1.14 \times \Delta(F_{\text{phi}})^{30} + 7.95 \times V_{\text{exclusion}} - 0.919 \\ & \times \ln(f_1 \times f_2) - 4.3 \times N_{\text{ssbond}} - \Delta G_{\text{ref}} \end{aligned}$$

H-bond

$2.3 \text{ \AA} < R < 3.5 \text{ \AA}$

$\theta > 100^\circ$

$|\phi - 109.5^\circ| < 71.5^\circ, |\psi - 109.5^\circ| < 71.5^\circ$ for sp^3 acceptor

$\phi > 80^\circ, \psi > 80^\circ$ for sp^2 acceptor

Electrostatic

$$\Sigma\Sigma(q_i \times q_j)/r^2$$

$$r = R_{ij} \quad \text{if } 0.8 \times (r_i + r_j) \leq R_{ij} \leq 12$$

$$r = 0.8 \times (r_i + r_j) \quad \text{if } R_{ij} < 0.8 \times (r_i + r_j)$$

Solvation

$$(S_{\text{buried}}/S_{\text{total}})^{30}$$

Solvent-exclusion

$$\sum_i \sum_j \frac{2C_i V_j \exp(- (r_{ij}/\lambda)^2)}{4\pi \sqrt{\pi} \lambda r_{ij}^2}$$

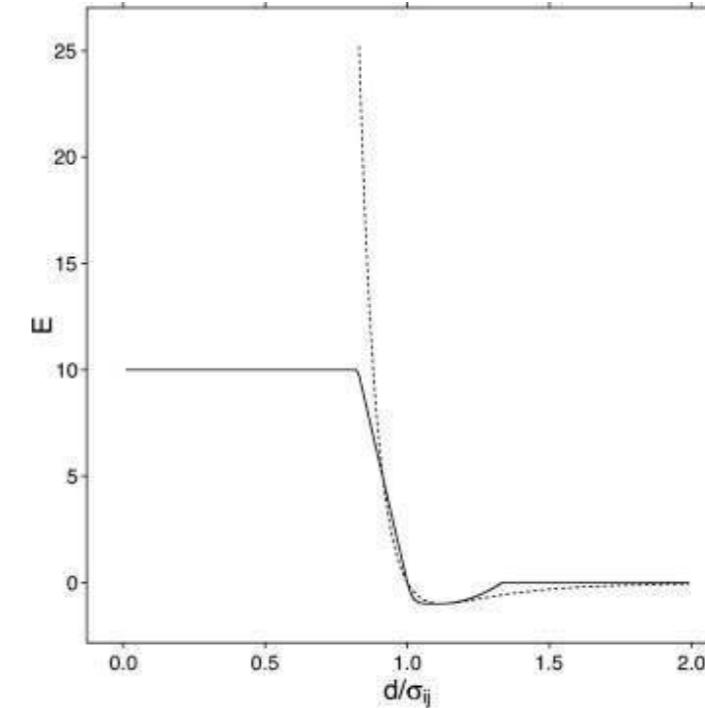
Потенциал AutoDock

$$\begin{aligned} E_{\text{dock}} &= E_{\text{vdw}} + E_{\text{H-bond}} + E_{\text{electrostatic}} + E_{\text{internal}} \\ &= \sum_{\text{protein}} \sum_{\text{ligand}} \left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^6} \right) + \sum_{\text{protein}} \sum_{\text{ligand}} E(t) \times \\ &\quad \left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}} \right) + \sum_{\text{protein}} \sum_{\text{ligand}} 332.0 \frac{q_i q_j}{\epsilon(d_{ij}) d_{ij}} + \\ &\quad \left\{ \sum_{\text{ligand}} \left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^6} \right) + \sum_{\text{ligand}} E(t) \left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}} \right) + \right. \\ &\quad \left. \sum_{\text{ligand}} 332.0 \frac{q_i q_j}{4 d_{ij} d_{ij}} \right\} \end{aligned}$$

Потенциал SCWRL4

VDW:

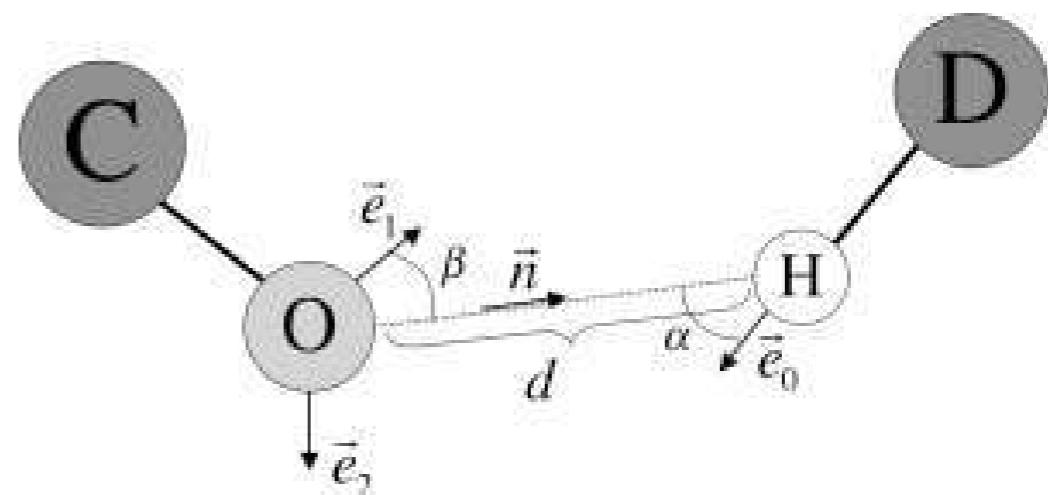
$$E_{vdw}(d) = \begin{cases} 10 & \text{if } \frac{d}{\sigma_{ij}} \leq 0.8254 \\ 57.273 \left(1 - \frac{d}{\sigma_{ij}}\right) & \text{if } 0.8254 \leq \frac{d}{\sigma_{ij}} \leq 1 \\ E_{ij} \left(10 - 9 \frac{d}{\sigma_{ij}}\right)^{\frac{57.273}{9E_{ij}}} - E_{ij} & \text{if } 1 < \frac{d}{\sigma_{ij}} < \frac{10}{9} \\ \frac{E_{ij}}{4} \left(9 \frac{d}{\sigma_{ij}} - 10\right)^2 - E_{ij} & \text{if } \frac{10}{9} \leq \frac{d}{\sigma_{ij}} < \frac{4}{3} \\ 0 & \text{if } \frac{d}{\sigma_{ij}} \geq \frac{4}{3} \end{cases}$$



H-bond

$$E_{[O,H]} = (1 - w)E_{vdw} + wBq_H q_O$$

$$w = \frac{\sqrt{(\sigma_d^2 - (d - d_0)^2)(\cos\alpha - \cos\alpha_{\max})(\cos\beta - \cos\beta_{\max})}}{\sigma_d \sqrt{(1 - \cos\alpha_{\max})(1 - \cos\beta_{\max})}}$$



$$E(\mathbf{r}) = \sum_{i=1}^N E_{\text{self}}(r_i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N E_{\text{pair}}(r_i, r_j) \quad E_{\text{self}}(r_i) = -k_i \log \frac{p(r_i)}{p(r_{\max})} + E_{\text{frame}}(r_i)$$

Потенциал ZRANK

$$\begin{aligned}Score = & w_{\text{vdW_a}} E_{\text{vdW_a}} + w_{\text{vdW_r}} E_{\text{vdW_r}} + w_{\text{elec_sra}} E_{\text{elec_sra}} \\& + w_{\text{elec_srr}} E_{\text{elec_srr}} + w_{\text{elec_lra}} E_{\text{elec_lra}} \\& + w_{\text{elec_lrr}} E_{\text{elec_lrr}} + w_{\text{ds}} E_{\text{ds}}\end{aligned}$$

- van der Waals attractive: 1.0
- van der Waals repulsive: 0.009
- Electrostatics short range attractive: 0.31
- Electrostatics short range repulsive: 0.34
- Electrostatics long range attractive: 0.44
- Electrostatics long range repulsive: 0.50
- Desolvation: 1.02

VDW:

$$r_{ij} < 8.0 \text{ \AA:}$$

$$E_{\text{vdW}}(i,j) = \varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$

для $r_{ij} < 0.6 \sigma_{ij}$ линеризация

Электростатика:

$$E_{\text{elec}}(i,j) = 332 \frac{q_i q_j}{r_{ij}^2}$$

Сольватация + Энтропия:

$$E_{\text{ds}}(i,j) = a_{ij}$$

Потенциал ROSETTA

backbone torsion angles

$$\sum_i -\ln [P(\phi_i, \psi_i | \text{aa}_i, \text{ss}_i)]$$

side index $i = \text{re}$
backbone torsion angles (36° bins)
amino acid type aa =
secondary structure type b

$$\text{dih}^{\text{tors}}_{\text{ccs}} = \frac{1}{12} \left[-2 \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \left\{ \begin{array}{l} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] \text{interatomic distance} \\ \text{geometric mean of atom well-depths}^d \\ 59.2 \left(\frac{d_{ij}}{r_{ij}} \right) + 5872.0 \left[e_{ij} / 87.0 \right] \text{summed van der Waals radii} \end{array} \right.$$

$$\text{hb}^f = \sum_j (-\ln [P(d_{ij} | h_j, \text{ss}_j)])$$

donor residue index $i = \text{dd}$
acceptor residue index $j = \text{dc}$
acceptor-proton interatomic distance
 $[P(\cos \theta_{ij} | d_{ij}, h_j, \text{ss}_j)]$ hybridization (sp^2/sp^3)
 $[P(\cos \psi_{ij} | d_{ij}, h_j, \text{ss}_j)]$ secondary structure type g

$$\text{solv} = \frac{1}{2} \left(\frac{2\Delta G_i^{\text{free}}}{4\pi^{3/2}\lambda_i r_{ij}^2} \right) \Delta G_{ij}^{\text{free}}$$

atom indices $i, j = \text{A}$
distance between atoms
immed van der Waals radii r_{ij}
correlation length $\lambda_i = c$
atomic volume $V = \varepsilon$
 ΔG^{free} = energy of a fully solvated

$$\text{pair interactions, s)} = -\ln \left[\frac{P(\text{aa}_i, \text{aa}_j | d_{ij})}{P(\text{aa}_i | d_{ij}) P(\text{aa}_j | d_{ij})} \right]$$

residue indices $i, j = \text{A}$
amino acid type aa =
distance between residues d_{ij}

$$\text{dun}^{\text{back}}_{\text{gy}} = \left[\frac{P(\text{rot}_i | \phi_i, \psi_i) P(\text{aa}_i | \phi_i, \psi_i)}{P(\text{aa}_i)} \right]$$

residue indices $i, j = \text{S}$
Dunbrack backbone-dependent
amer rot

$$\text{ref energy} = \sum_{\text{aa}} n_{\text{aa}}$$

amino acid type aa =
= backbone torsion angles
amino acid type aa =
number of residues = r

7, 12 Ramachar
preferer

15 Lennard-J
interacti

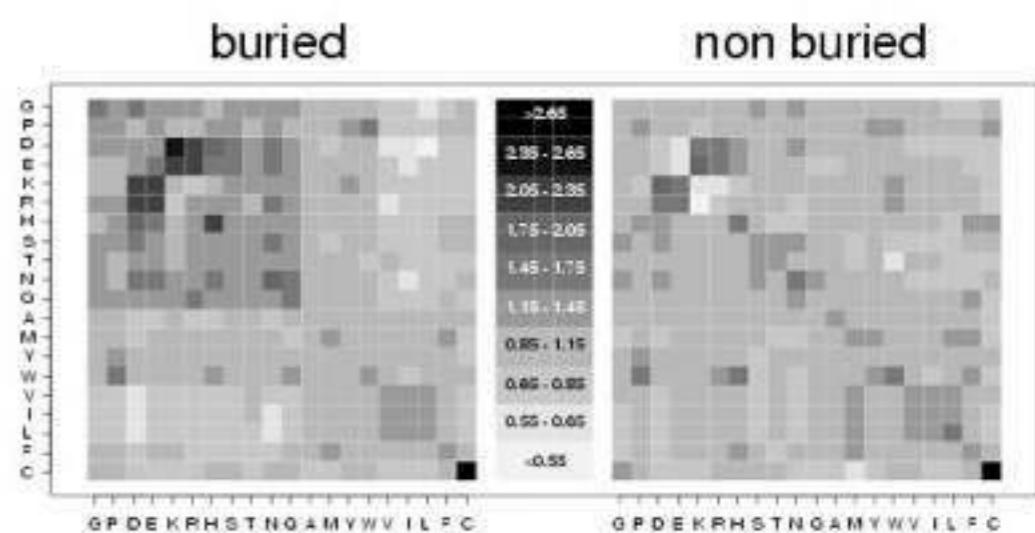
19 Hydrogen
bonding

15, 1 Solvation

15 Resid
inte
(ele
dis)

14, 1 Rotai
self

15 Unfol
ref



Анализ Нормальных Мод

→ отдельная презентация

Итеративные методы диагонализации

Степенной метод

- Пусть A - матрица с полным набором собственных значений и векторов (λ_i, q_i)
- Пусть собственные значения упорядочены так, что $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$
- Для любого вектора u , $A^k u = \gamma_1 \lambda_1^k q_1 + \sum_i \gamma_i \lambda_i^k q_i$
- При $k \rightarrow \infty$ член, содержащий λ_1 становится доминантным и $A^k u$ приближается к первому члену разложения

Алгоритм

- Выбираем $v^{(0)}$ так что $\|v^{(0)}\| = 1$. Затем для $k=1, 2, \dots$ мы повторяем

$$\begin{aligned} w &= Av^{(k-1)} \\ v^{(k)} &= \frac{w}{\|w\|} \\ \lambda^{(k)} &= (v^{(k)})^T A v^{(k)} \end{aligned}$$

- Скорость сходимости к собственным значениям и векторам

$$\begin{aligned} \|v^{(k)} - q_1\| &= O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \\ |\lambda^{(k)} - \lambda_1| &= O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right) \end{aligned}$$

- Предложен Мунтцем в 1913 году
- Стандартный метод до изобретения QR алгоритма

Обратный степенной метод

- Предположим μ НЕ собственное значение A
- Матрица $B = (A - \mu I)^{-1}$ имеет те же собственные вектора, что A , и собственные значения $(\lambda_j - \mu)^{-1}$
- Если выбрать μ близко к собственному значению, то $(\lambda_j - \mu)^{-1}$ будет огромным числом
- Мы можем применить степенной метод к матрице $(A - \mu I)^{-1}$ и мы получим собственные вектора q_j и собственные значения $(\lambda_j - \mu)^{-1}$

Алгоритм

- Выбираем $v^{(0)}$ так что $\|v^{(0)}\| = 1$. Затем для $k=1, 2, \dots$ мы повторяем

$$\begin{aligned} w &= (A - \mu I)^{-1} v^{(k-1)} \\ v^{(k)} &= \frac{w}{\|w\|} \\ \lambda^{(k)} &= (v^{(k)})^T A v^{(k)}. \end{aligned}$$

- Скорость сходимости к собственным значениям и векторам

$$\left\| v^{(k)} - q_j \right\| = O \left(\left| \frac{\mu - \lambda_j}{\mu - \lambda_k} \right|^k \right)$$

$$\left| \lambda^{(k)} - \lambda_j \right| = O \left(\left| \frac{\mu - \lambda_j}{\mu - \lambda_k} \right|^{2k} \right)$$

- λ_j - ближайшее собственное значение к μ и λ_k - второе ближайшее
- Предложен в 1944 году

Метод коэффициентов Релея

- Выбираем $v^{(0)}$ так что $\|v^{(0)}\| = 1$. На первом шаге $\lambda^{(0)} = (v^{(0)})^T A v^{(0)}$
Затем для $k=1, 2, \dots$ мы повторяем

$$\begin{aligned} w &= (A - \lambda^{(k-1)} I)^{-1} v^{(k-1)} \\ v^{(k)} &= \frac{w}{\|w\|} \\ \lambda^{(k)} &= (v^{(k)})^T A v^{(k)}. \end{aligned}$$

- Скорость сходимости к собственным значениям и векторам

$$\begin{aligned} \|v^{(k+1)} - q_j\| &= O\left(\left\|v^{(k)} - q_j\right\|^3\right) \\ |\lambda^{(k+1)} - \lambda_j| &= O\left(\left|\lambda^{(k)} - \lambda_j\right|^3\right) \end{aligned}$$

- Предложен в 1958 году

Метод Следа

$$\text{Tr}(\mathbf{A}^N) = \lambda_{dom}^N + \lambda_2^N + \lambda_3^N + \dots$$

$$\lambda_{dom} \approx [\text{Tr}(\mathbf{A}^N)]^{1/N}$$

- N нечетно, иначе теряется знак доминантного собственного значения
- возвведение в квадрат - не самая приятная матричная операция

QR алгоритм

- На шаге k матрица A_k разлагается на произведение ортогональной Q_k и верхней треугольной R_k :

$$A_k = Q_k R_k$$

- На следующем шаге:

$$A_{k+1} = R_k Q_k = Q_k^T Q_k R_k Q_k = Q_k^T A_k Q_k = Q_k^{-1} A_k Q_k,$$

- Матрицы A_k подобны и имеют одинаковые собственные значения

Метод понижения порядков

- При помощи степенного метода определяем максимальное собственное значение и его вектор для матрицы \mathbf{A}
- Умножаем максимальный собственный вектор на его же транспонированный и затем на максимальное собственное значение, получаем:

$$\mathbf{Z}^* = c \mathbf{X} \mathbf{X}^T$$

- Вычисляем новую матрицу:

$$\mathbf{A}^* = \mathbf{A} - \mathbf{Z}^* = \mathbf{A} - c \mathbf{X} \mathbf{X}^T$$

- При помощи степенного метода определяем максимальное собственное значение и его вектор для матрицы \mathbf{A}^*
- Матрица \mathbf{A}^* имеет те же собственные значения, что и \mathbf{A} , кроме доминантного

Литература

- Y. Saad. “Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms”. Halsted Press 1992 .