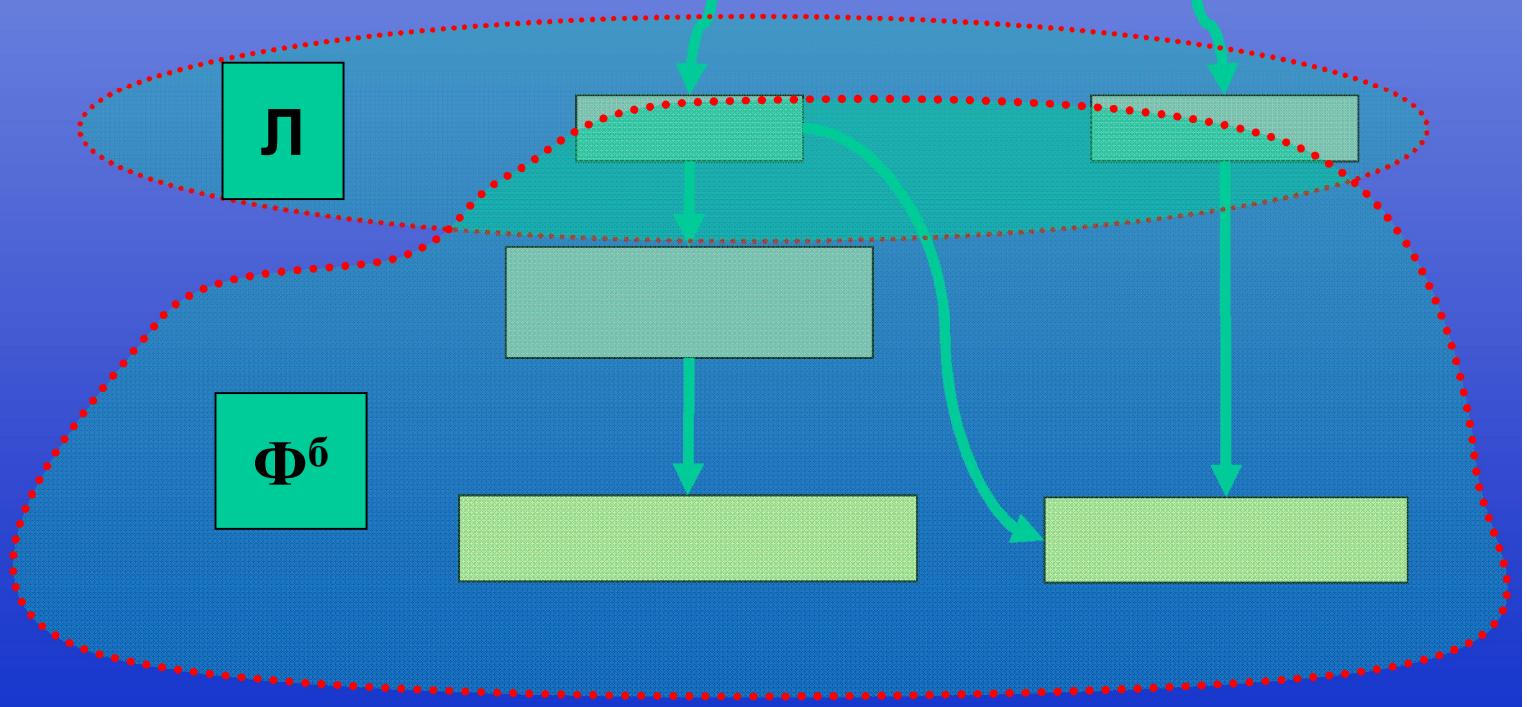
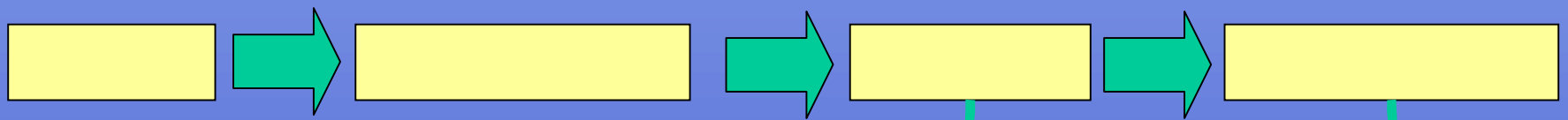


...





Экосистема

Популяция

Организм

Системы органов

Органы

Ткани

Клетки

Молекулярные каскады

РНК

ДНК

Атомы

Молекулы

Атомы

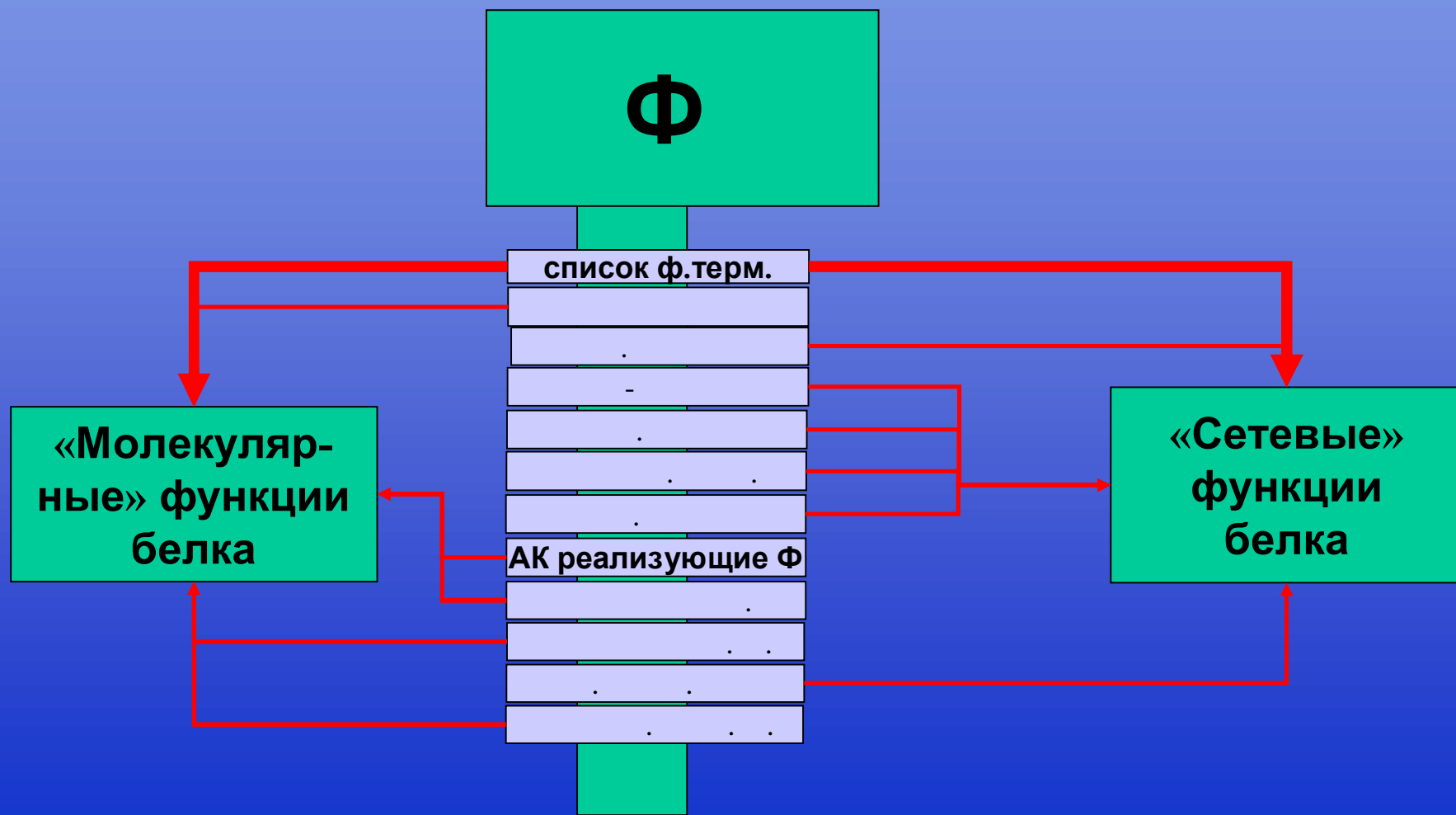
Молекулы

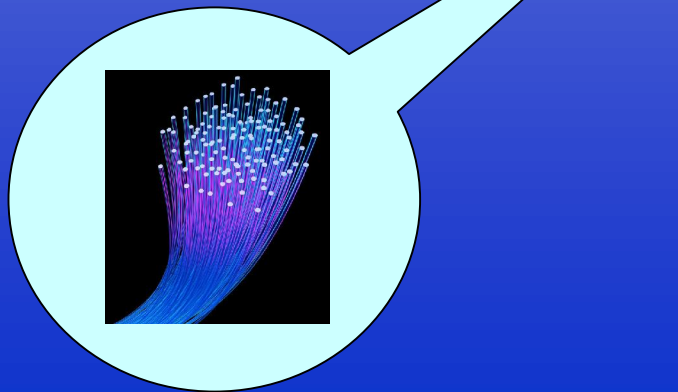
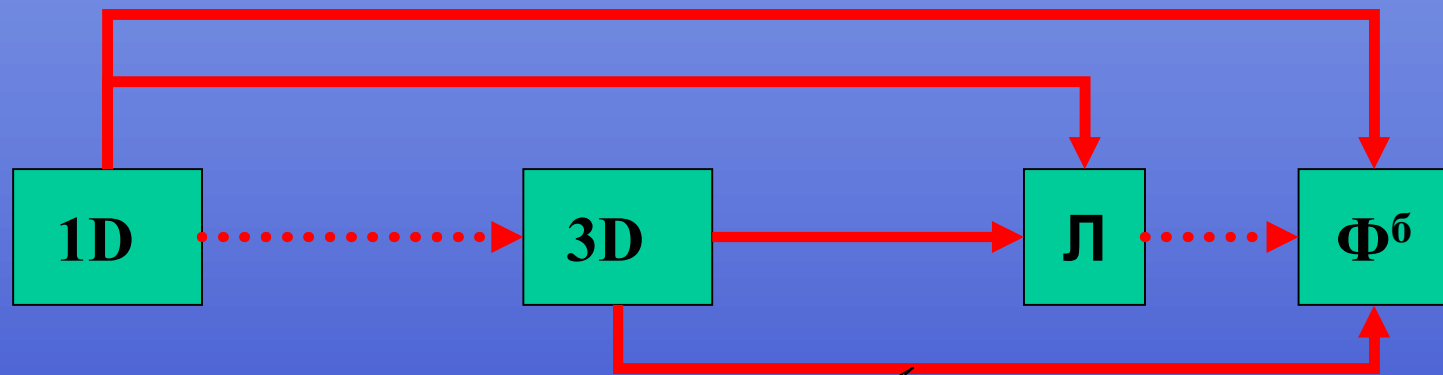
3D

Ф

Л

1D⁶





3D>3D 3D>

Общая методология исследования:

0. *Формулировка биофизических гипотез*

I. (,)

-

II.

III.

, /) , (

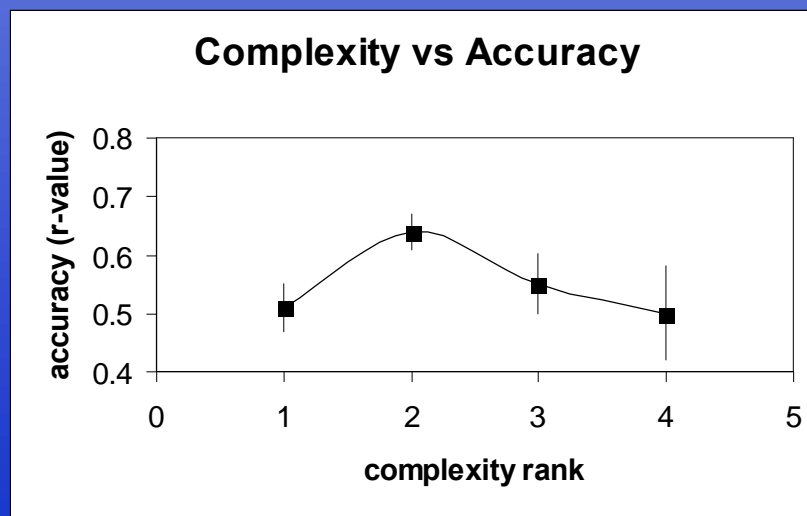
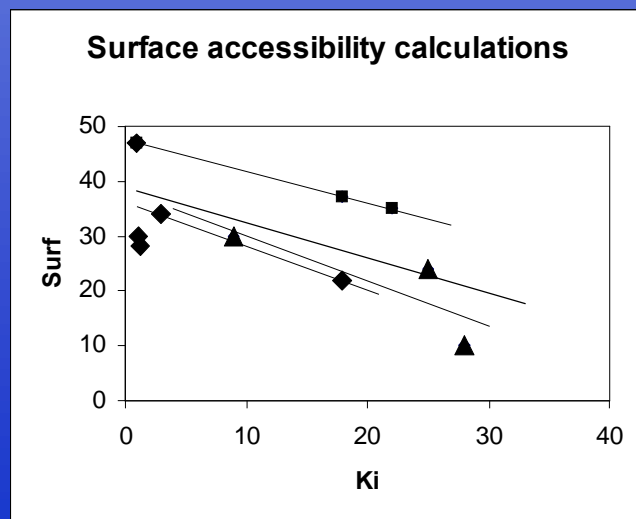
()

- **Пример:**

-10+

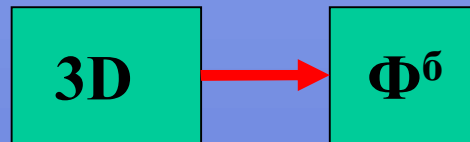
- **Четыре модели различной информационной сложности:**

- 1. - $< 4\text{\AA}$
- 2. (surface accessibility)
- 3. /
- 4. (UFF, AMBER)



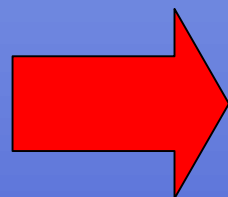
(Torshin, 2006)

Структурные компоненты				
Функция (на основе 3D)	()		« »	-
Функция (на основе 1D)			RGD	-
Фолдинг	+,- ()			
Стабильность				

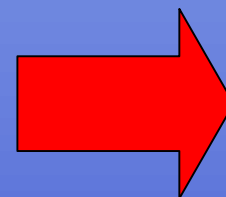


- **Функция (на основе 3D)**
 - **Метод центра заряда**
 - **Метод редких конформаций**
 - **Метод «карманов» и планарных участков**
 - **Метод молекулярно-энергетических профилей**

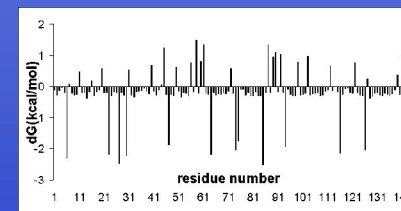
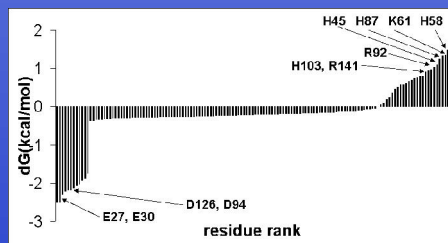
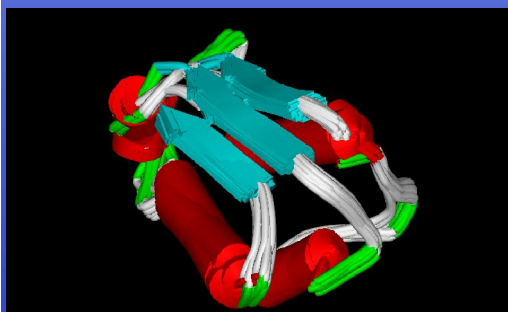
3D



Биофизические
«особенности»
3D



Проекция на
1D



•
•

—

•••

Структурные компоненты	4-27 25-50 70-88 92-110	(1 - , 2 -)	(a/b, PDB 1b27_A)	:
Функция (на основе 3D)	T26 K27 S28 I51 D54 I55 F56 R72 E73 Y90	73- 87 () 35- 45 ()	33-45 (<250A ²)	/ K27 D54 R59 E60 (K66 R69) E73 R83 R87 H102
Функция (на основе 1D, 2D)	3 16- 18 TYN 47-49 PGK 64-66 PGK 43-50 ATP_BND	RGD	51/109 идентичных остатков	/ K27 K49 D54 R59 E60 (K66 R69) E73 R83 R87 H102
Фолдинг	+,- F7 V10 L20 Y24 I51 G52 G53 I55 F56 W71 E73 A74 D75 I76 R87 I88 L89 Y90 Y97 Y103	10-24 29- 37 40-53 79-89 91- 96,99-103,106-110	26-28 () 35-38 () 56-59 ()	7-
Стабильность		54-62,73- 87,35-45	102 c ; A37 S38 F56 R59 E60 R83 N84 S85 Y103; « » E73- Y103 D101-T105	. ddG ddG (r=-0.67, P<0.01, /)

A1*	Стаб денат МД	
Q2	Стаб денат МД	
V3		
I4		I4V $\neq 1$
N5	Стаб взаимод комп мод 4-27/70-88	N5A $\neq 1$
T6	Стаб денат МД	
F7	+/- ЦЗ	
D8	Стаб денат МД	
G9	Нукл комп мод 4-27, стаб денат МД	
V10*	Нукл комп мод 4-27, +/- ЦЗ	
A11	Нукл комп мод 4-27	
D12		
Y13*	Нукл комп мод 4-27, стаб денат МД	
L14*	Нукл комп мод 4-27, стаб денат МД	
Q15	Стаб денат МД	
T16	Редк Злепт Нукл комп мод 4-27	T16S $\neq 0$
Y17	Редк Злепт Нукл комп мод 4-27	
H18	Редк Злепт	H18Q $\neq 0$
K19	Стаб денат МД	
L20*	+/- ЦЗ, стаб денат МД	
P21*	Стаб денат МД	
D22*	Стаб денат МД	
N23	Стаб комп мод 4-27	N23A $\neq 1$
Y24*	+/- ЦЗ, стаб денат МД	Адро АСомана
I25	Стаб денат МД	I25V $\neq 1$
T26*	ОЦЗ кластер	T26A $\neq 1$
K27*	ОЦЗ, наиболее стабилизирующий	Активный центр
S28	ОЦЗ кластер	
E29	Стаб денат МД	E29G $\neq 1$
A30*	Нукл комп 25-30, стаб денат МД	
Q31	Нукл комп 25-30, стаб денат МД Полость	Полость акт центра
A32	Нукл комп 25-30, стаб денат МД	
L33	Нукл комп 25-30, стаб денат МД	L33Q $\neq 1$
G34*	Нукл комп 25-30, стаб денат МД	
W35*	Нукл комп мод 25-30 Полость	Полость акт центра
V36	Нукл комп мод 25-30	V36T $\neq 1$
A37	Пептид редк конф	интерфейс барназа-барстар
S38	Пептид редк конф	интерфейс барназа-барстар
K39	Пептид редк конф, стаб денат МД	
G40	Пептид редк конф, стаб денат МД	
N41*	Стаб комп мод 25-30, стаб денат МД	N41D $\neq 1$
L42*	Пептид редк конф, стаб денат МД	
A43	Пептид редк конф, ATR_BND	
D44	Пептид редк конф, ATR_BND, стаб МД	
V45*	Пептид редк конф, ATR_BND, стаб МД	V45A $\neq 1$
A46	ATR_BND, стаб денат МД	
P47*	ATR_BND, Редк Злепт, стаб денат МД	
G48*	ATR_BND, Редк Злепт, стаб денат МД	
K49*	Взаимод 4-27/70-88, Редк Злепт, стаб МД	Активный центр
S50	ATR_BND	
I51*	ОЦЗ кластер, +/- ЦЗ кластер	I51V $\neq 1$
G52*	+/- центры заряда	
G53*	+/- центры заряда	
D54*	ОЦЗ, наиб стабилизирующий, стаб МД	Активный центр, D54N
I55	ОЦЗ, +/- ЦЗ, стаб денат МД	
F56*	ОЦЗ кластер, +/- центры заряда	интерфейс барназа-барстар
S57	Полость	Полость акт центра
N58*	Полость	Полость акт центра N58A $\neq 0$

()

R59*	Наиболее дестабилизирующий	Акт центр, интерфейс барназа-барстар
E60*	Наиболее стабилизирующий, стаб МД	Акт центр, интерфейс барназа-барстар
G61		
K62	Полость	Полость акт центра K62R $\Phi \approx 0$
L63*	Нуклеация комп мод 55-70	
P64*	Нукл комп мод 55-70, редк 3пепт, стаб МД	
G65	Нуклеация комп мод 55-70, редк 3пепт	
K66	Наиболее дестабилизирующий, редк 3пепт	Конст ассоц
S67	Нуклеация комп мод 55-70	
G68	Нуклеация комп мод 55-70	
R69*	Взаимод мод 70-88/92-110, наиб дестаб	Конст ассоц
T70		
W71*	+, - центры заряда, стаб денат МД	
R72	ОЦЗ кластер, стаб денат МД	
E73*	+, - ЦЗ, ОЦЗ, взаимод 70-88/92-110, наиб стаб	Активный центр
A74*	Пептид редк конф, +, - ЦЗ	Ядро Н-обмена
D75*	Стаб комп мод 70-88, +, - ЦЗ Полость	Полость акт центра
I76	Нукл комп мод 70-88, +, - ЦЗ	I76V $\Phi \approx 1$
N77*	Нукл комп мод 70-88	N77A $\Phi \approx 1$
Y78*	Нукл комп мод 70-88	Y78F $\Phi \approx 1$
T79	Нукл комп мод 70-88, стаб денат МД	
S80	Нукл комп мод 70-88	
G81*	Нукл комп мод 70-88, стаб денат МД	
F82	Нукл комп мод 70-88 Полость	Полость акт центра
R83*	Пептид редк конф, наиб дестаб	Акт центр, интерфейс барназа-барстар
N84	Стаб комп мод 70-88, стаб денат МД	интерфейс барназа-барстар
S85	Пептид редк конф 73-87	интерфейс барназа-барстар
D86*	Пептид редк конф 73-87, стаб денат МД	
R87*	Пептид редк конф, наиб дестаб, +, - ЦЗ	Активный центр, ядро Н-обмена
I88*	+, - центры заряда	Ядро Н-обмена I88V $\Phi \approx 0$
L89*	+, - центры заряда	Ядро Н-обмена
Y90*	ОЦЗ +, - центры заряда	Ядро Н-обмена
S91*	Стаб денат МД +, - центры заряда	S91A $\Phi \approx 0$
S92	+, - центры заряда	S92A $\Phi \approx 0$
D93*	Взаимод комп мод 70-88/92-110	
W94	Нуклеация комп мод 92-110	
L95*	Нуклеация комп мод 92-110	
I96	Нуклеация комп мод 92-110	
Y97*	Нуклеация комп мод 92-110 +, - ЦЗ	Ядро Н-обмена
K98	+, - центры заряда	
T99*	Стаб денат МД	
T100	Стаб денат МД	
D101*	Стаб комп мод 92-110, стаб денат МД	
H102*	Наиболее дестабилизирующий, стаб МД	Акт центр интерфейс барназа-барстар
Y103*	Взаимод комп мод 70-88/92-110, +, - ЦЗ	интерфейс барназа-барстар
Q104	Полость	Полость акт центра
T105	Стаб комп мод 92-110	
F106*	Стаб денат МД Полость	Полость акт центра
T107	Стаб денат МД	
K108	Стаб денат МД	
I109		

- **Физические закономерности есть, но распознавание на их основе не всегда эффективно**
- **Значительное число эмпирических параметров,**

«

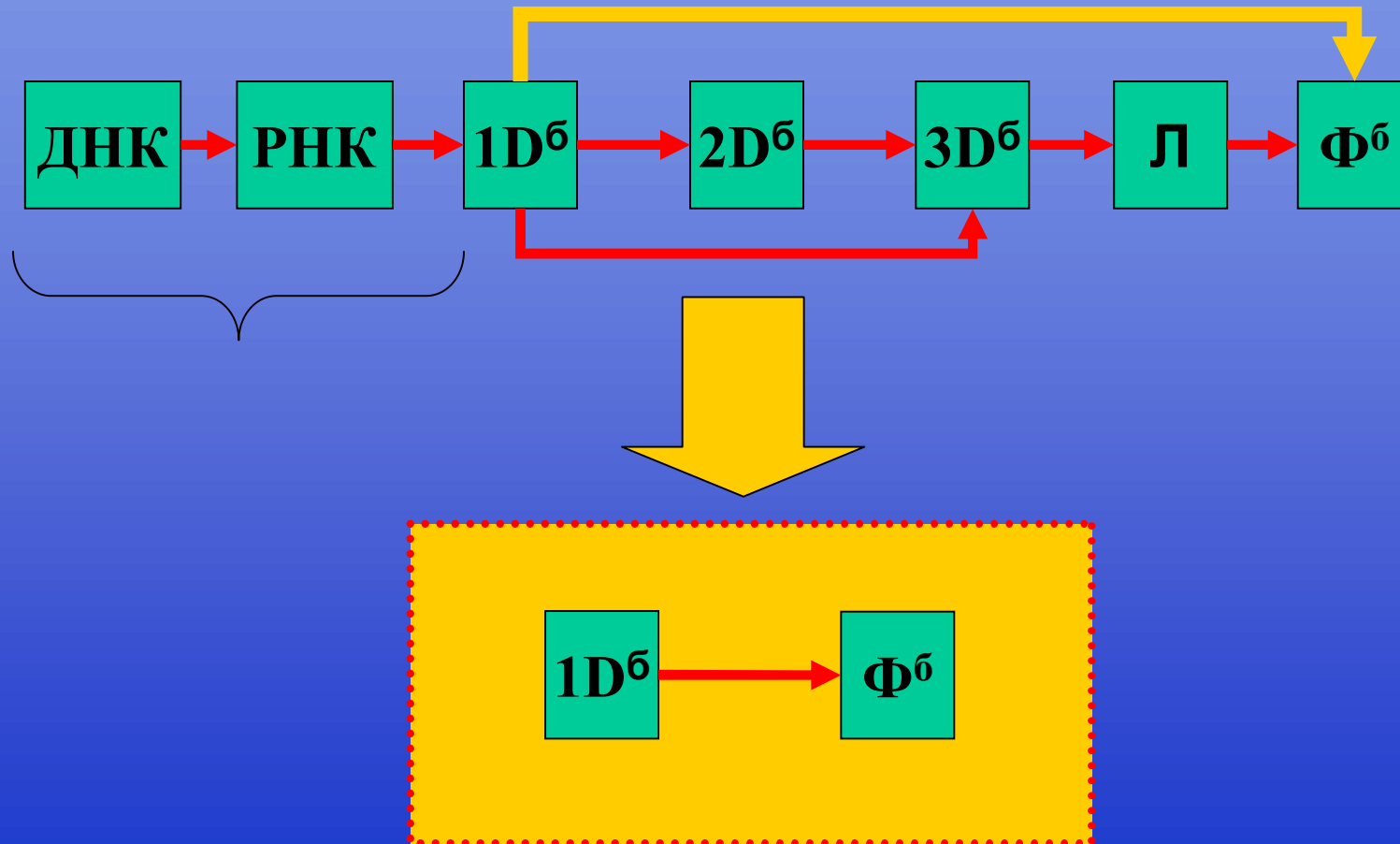
»

•

«

»

,



VHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVYPWTQRFFESFGDLSTPDAV
MGNPKVKAHGKKVLGAFSDGLAHLNKGTFATLSELHCDKLHVDPENFRLLGN
VLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

- GO:0006810 transport
- GO:0015671 oxygen transport
- GO:0005344 oxygen transporter activity
- GO:0030185 nitric oxide transport
- GO:0046872 metal ion binding
- GO:0005506 iron ion binding
- GO:0020037 heme binding
- GO:0030492 hemoglobin binding
- GO:0005833 hemoglobin complex
- GO:0008217 regulation of blood pressure
- GO:0050880 regulation of blood vessel size

алфавит A

$$A = \{ a_1, a_2, \dots, a_n \}, n > 0$$

множество белков A^*

$$A^* = \bigcup_{l=1}^{\infty} A^l$$

белок P

$$P^l - \text{белок длины } l \\ P^l \in A^*$$

терминологический словарь T

$$T = \{ t_1, t_2, \dots, t_m \} \quad t_j$$

множество аннотаций T^*

$$T^* = \bigcup_{l=1}^{\infty} T^l$$

аннотация белка P

$$t = \{ t_1, t_2, \dots, t_k \} \\ t \subset T^*$$

множество
прецедентов

$$A^* \times T^*$$

набор прецедентов Pr

$$Pr \subseteq A^* \times T^*$$

i -ый прецедент
 $(P_i, t^i) \in Pr$

Аннотирующая функция

$$F : A^* \rightarrow T^*$$

корректность

$$\forall_{Pr} (P, t) : F(P) = t$$

Задача аннотации $Z^{ID>\Phi}(Pr)$

разрешимость (**Th.1**)

$$\forall_{Pr} (P_i, t^i), (P_j, t^j), i \neq j : (P_i = P_j) \Rightarrow (t^i = t^j)$$

регулярность

$$\forall_{Pr} (P_i, t^i), (P_j, t^j), i \neq j \Rightarrow (P_i \neq P_j)$$

био-логично!

терминологический словарь

$$T = \{t_1, t_2, \dots, t_m\}$$

t-классы белков $C^{t_i} \subset A^*$

$$C^{t_1}, C^{t_2}, \dots, C^{t_m}$$

аннотированный белок $P_i \in Pr$

k_i классов

$$C^{t_1}, C^{t_2}, \dots, C^{t_k}$$

$$c^{t_1}, c^{t_2}, \dots, c^{t_k}, c^t \subset Pr$$

перекрывание классов

$$\forall i, j = 1..m : C^{t_i} \cap C^{t_j} \neq \emptyset$$

оценка – сравнение мощностей c^t

сцепление терминов

$$s_{ij} = \text{sign} \left(|c^{t_i}| - |c^{t_j}| \right) \cdot \frac{|c^{t_i} \cap c^{t_j}|}{\min(|c^{t_i}|, |c^{t_j}|)}$$

$$|s_{ij}| = 1$$

$$|s_{ij}| = 0$$

$$s_{ij} \in [-1..1]$$

инверсия класса C^{t_i}

$$C^{t_i} \cap \overline{C^{t_i}} = \emptyset$$

термин $t_j \in T$

разбиение Pr

$$c^{t_j} \subseteq C^{t_j} \quad \overline{c^{t_j}} \subseteq \overline{C^{t_j}}$$

$$c^{t_j} \cup \overline{c^{t_j}} = Pr$$

$$c^{t_j} \cap \overline{c^{t_j}} = \emptyset$$

вектор аннотации \vec{t}^i

$$\vec{t}^i = \{t_1(P_i), \dots, t_j(P_i), \dots, t_m(P_i)\}$$

$$P_i \in C^{t_j} \quad t_j(P_i) = 1$$

t-классификатор $f_{t_j}(P_i)$

корректность

$$\forall P \in Pr : f_{t_j}(P) = \begin{cases} 1, & \text{если } P \in c^{t_j} \\ 0, & \text{если } P \notin c^{t_j}, P \in \overline{c^{t_j}} \\ \Delta - \text{отказ от распознавания} \end{cases}$$

Th 2. Корректность

каждого t-классификатора – необходимое и

достаточное условие

корректности функции F.

Сайты (от англ. site, «местоположение»)

$$S_j(P_i) = \{i_1^i, \dots, i_k^i\}$$

$$|S_j(P_i)| = k \quad [S_j(P_i)] = (i_k^i - i_1^i + 1)$$

$$S_t = \bigcup_{k=1}^{|M|} \hat{m}_k = M_\Sigma(M)$$

Корректная локальная функция

$$\forall_{Pr} t \quad \forall_{Pr} P \quad \exists_{k=L(M)+1}^{|P|-R(M)} : f_t^{loc} (\eta(k, M_\Sigma, P)) = t(P)$$

... , $t(P)=1$

Th3. локальный t -классификатор f_t^{loc} существует тогда и только тогда, когда для f_t^{loc} выполняется условие разрешимости.

Критерий локальной разрешимости

$$\forall_{Pr} (P_1, t(P_1)), (P_2, t(P_2)) \quad \forall_{i,k \in N} (i,k) : \eta(i, \hat{M}_\Sigma(M), P_1) = \eta(k, \hat{M}_\Sigma(M), P_2) \Rightarrow t(P_1) = t(P_2)$$

Критерий локальной разрешимости на масках

$$\forall_{Pr} j \quad \forall_{j=1}^m (P_1, t_j(P_1)), (P_2, t_j(P_2)) \quad \forall_{i,k \in N} (i,k) \left(\forall_{l=1}^{|M|} \hat{m}_l : \eta(i, \hat{m}_l, P_1) = \eta(k, \hat{m}_l, P_2) \right) \Rightarrow t_j(P_1) = t_j(P_2)$$

- $Z(Pr, M), Pr -$
 M

—
—
—

(1,1'') не монотонен M

$M (M)$

0-тупиковая M

условие (1'') выполнено для M , но не выполнено $\forall M' \subset M \quad M_\Sigma(M') \subset M_\Sigma(M)$

тупиковая M

условие (1'') выполнено для M , но не выполнено $\forall M' \subset M$

Свойство ядерности M

$$\bigvee_{i=1}^N \exists \mu : \bigwedge_{j=1}^{N, i \neq j} (\mu \notin \hat{m}_j)$$