



Научно-технологический
университет

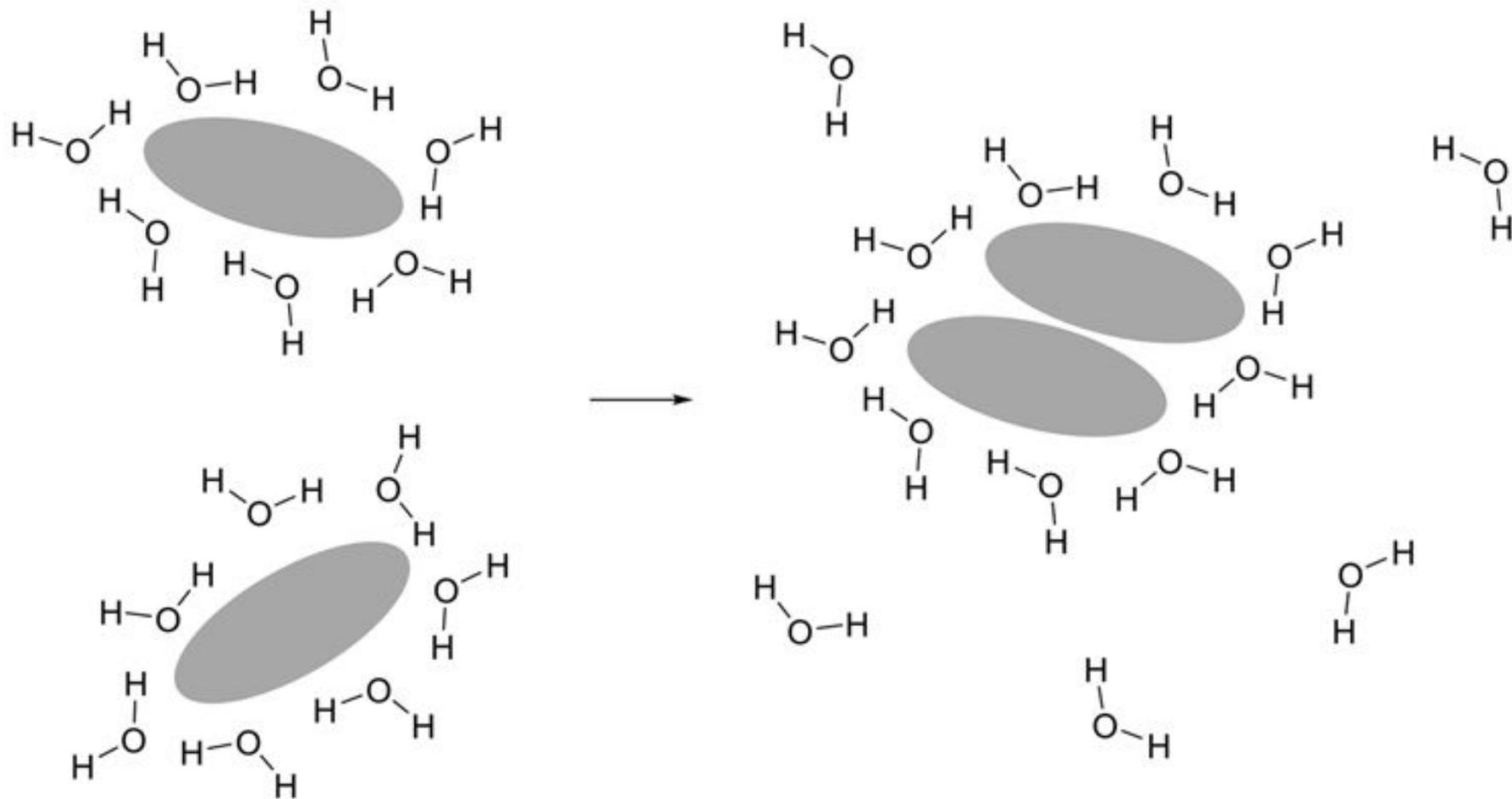
Сириус

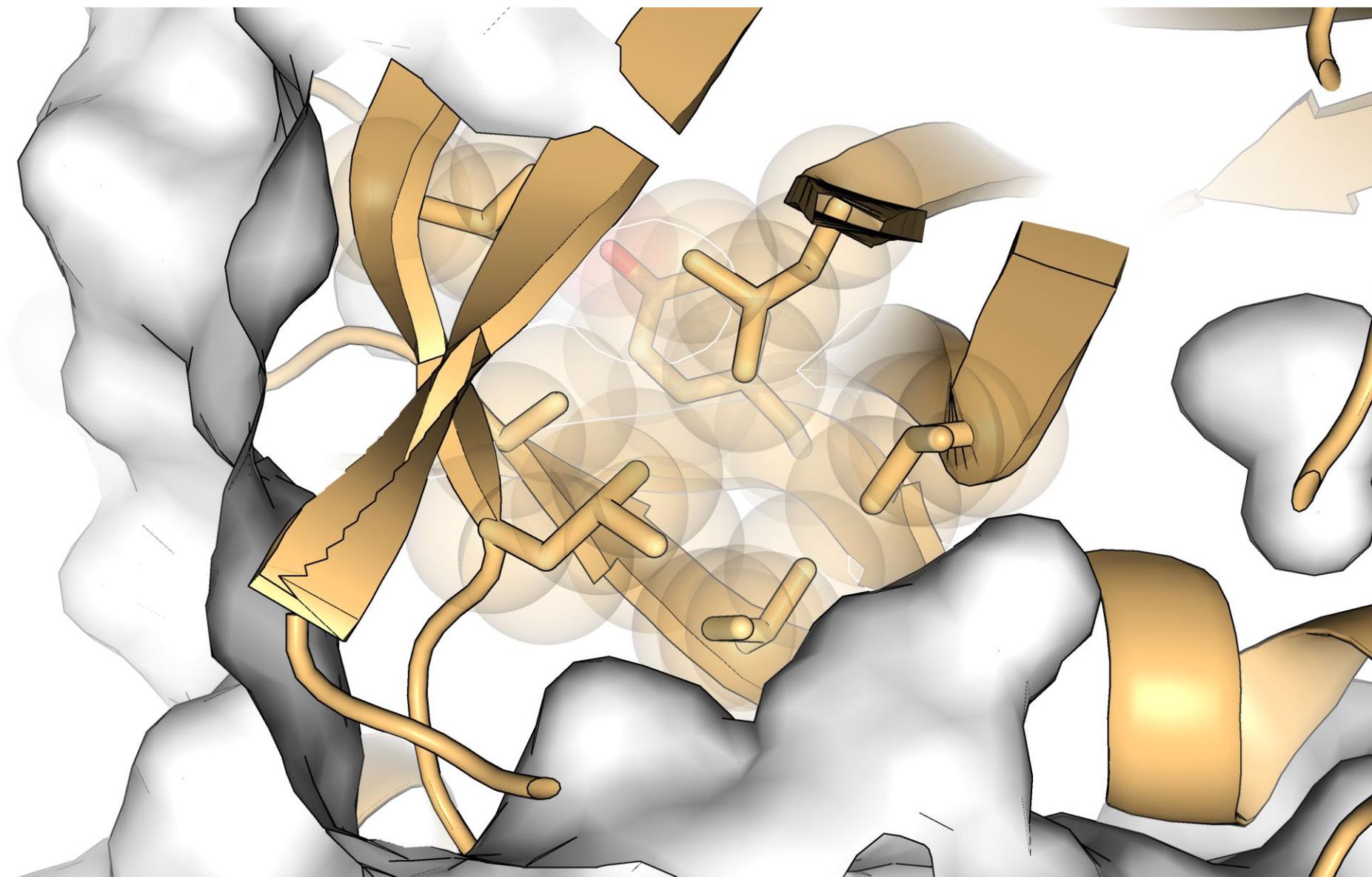
Структурная биоинформатика | Лекция 11

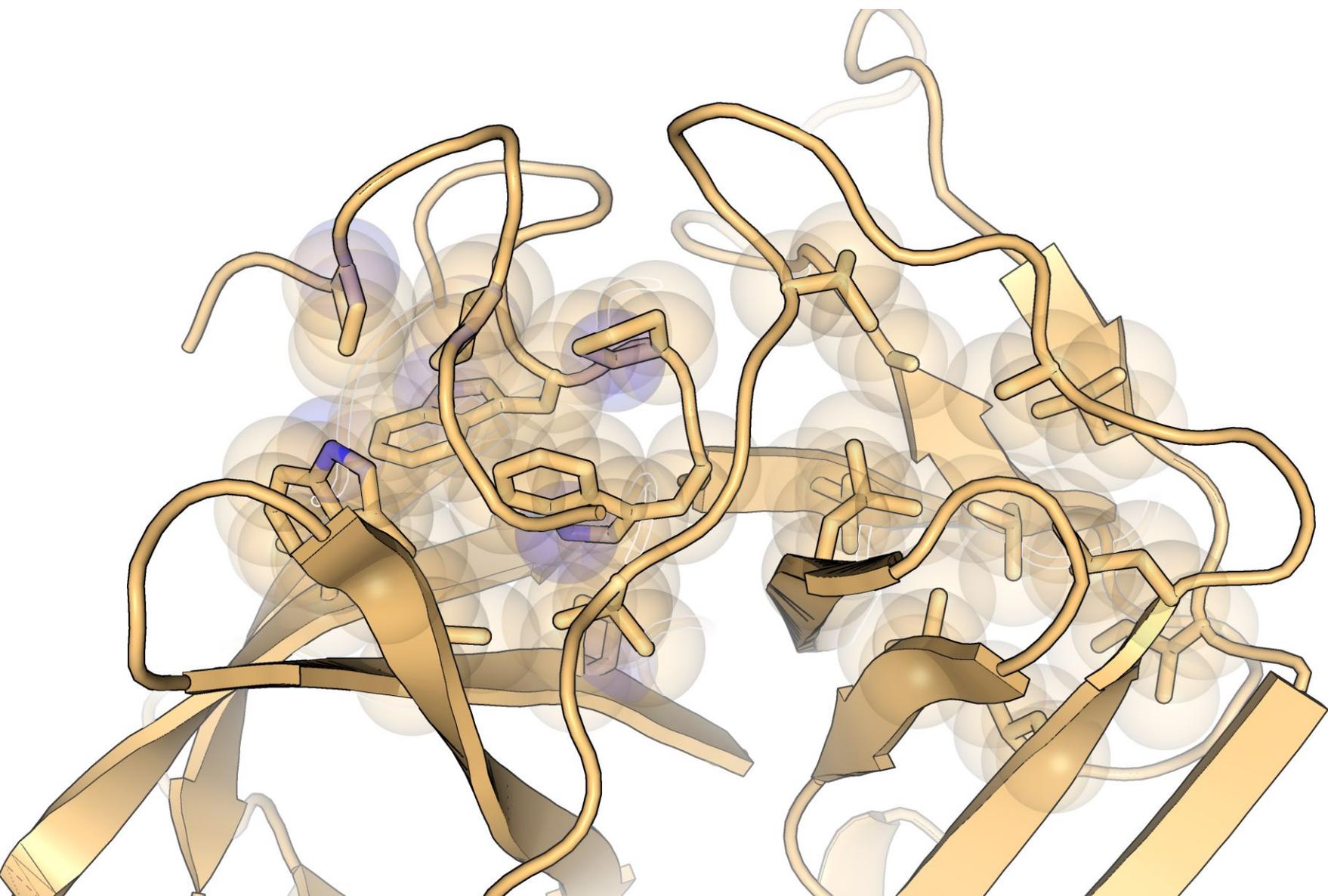
Гидрофобное ядро. Структурные домены

Александр Злобин

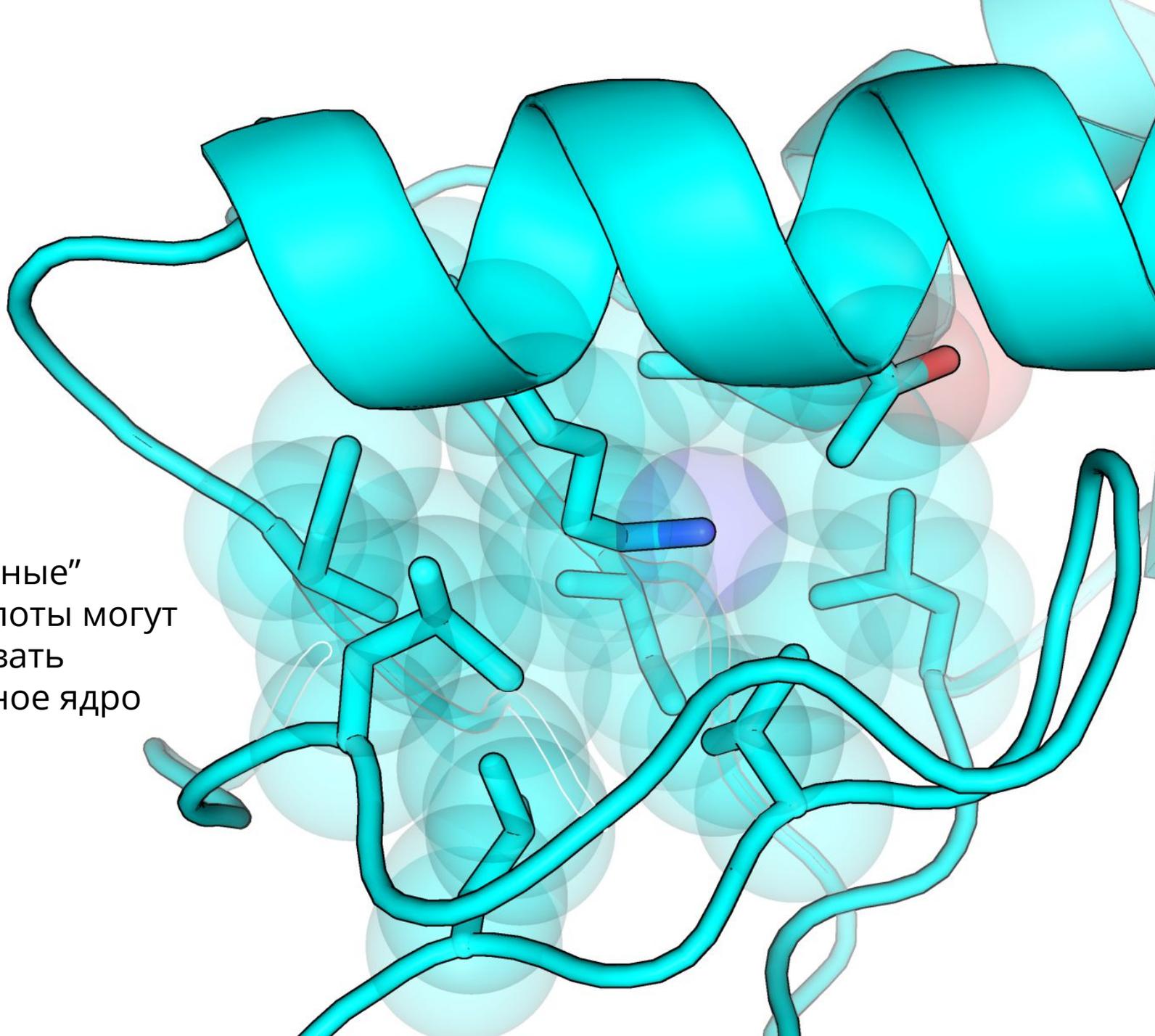
Гидрофобный эффект







Не только
"гидрофобные"
аминокислоты могут
образовывать
гидрофобное ядро



- Т.н. гидрофобное ядро дает существенный вклад в стабильность глобулы большинства белков
- Гидрофобные “ядрышки” могут служить зародышами в процессе правильной укладки полипептидной цепи
- Гидрофобный эффект важен для белок-белкового взаимодействия, взаимодействия белок-ДНК и других межмолекулярных взаимодействий
- Поиск в PubMed по “hydrophobic core” дает ~10000 статей, из них 281 содержат это словосочетание в названии.

Состав гидрофобного ядра

Некоторые позиции внутри гидрофобного ядра важнее для процесса фолдинга, чем остальные

**nature structural
& molecular biology**

Letter | [Published: 14 January 2002](#)

Hydrophobic core packing in the SH3 domain folding transition state

[Julian G.B. Northey, Ariel A. Di Nardo & Alan R. Davidson](#) 

Nature Structural Biology **9**, 126–130 (2002) | [Download Citation](#) 

Состав гидрофобного ядра

Гидрофобное ядро - подвижная структура, способная менять форму и адаптироваться в результате взаимодействий. Длинные остатки повышают подвижность, короткие и разветвленные - снижают

**nature
communications**

Article | [Open Access](#) | Published: 26 September 2019

Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk

Julia C. Heiby, Benedikt Goretzki, Christopher M. Johnson, Ute A. Hellmich  & Hannes Neuweiler 

Nature Communications **10**, Article number: 4378 (2019) | [Cite this article](#)

Подвижность гидрофобного ядра

... the present study suggests a universal mechanism for dynamically driven allosteric activation of kinases mediated by coordinated signal transmission through ordered motifs in their hydrophobic cores

RESEARCH ARTICLE | BIOCHEMISTRY

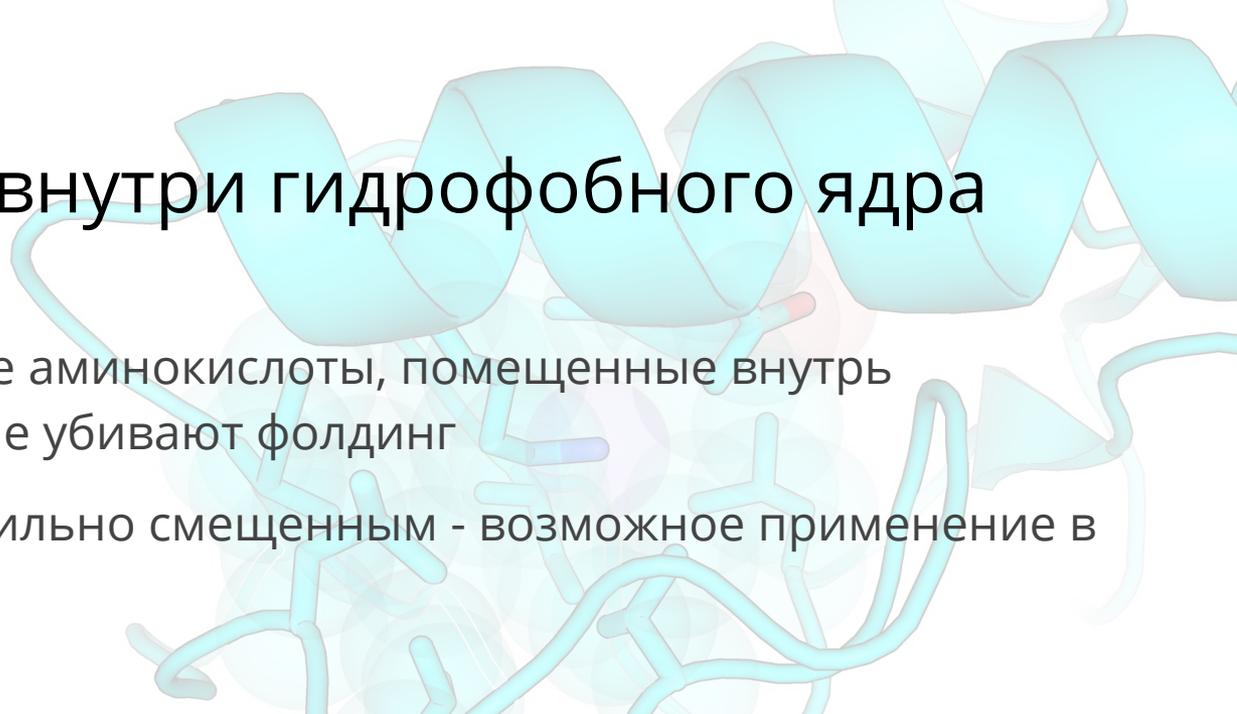
A dynamic hydrophobic core orchestrates allostery in protein kinases

Jonggul Kim^{1,2,*}, Lalima G. Ahuja³, Fa-An Chao^{2,†}, Youlin Xia², Christopher L. McClendon^{3,4,‡}, Alexandr P. Kornev³, Susan ...

+ See all authors and affiliations

Science Advances 07 Apr 2017:
Vol. 3, no. 4, e1600663
DOI: 10.1126/sciadv.1600663

Негидрофобика внутри гидрофобного ядра



- Одиночные полярные аминокислоты, помещенные внутрь гидрофобного ядра, не убивают фолдинг
- Их pKa оказывается сильно смещенным - возможное применение в катализе?

High tolerance for ionizable residues in the hydrophobic interior of proteins

Daniel G. Isom, Brian R. Cannon, Carlos A. Castañeda, Aaron Robinson, and Bertrand García-Moreno E.

PNAS November 18, 2008 105 (46) 17784-17788; <https://doi.org/10.1073/pnas.0805113105>

Как находить гидрофобные ядра?

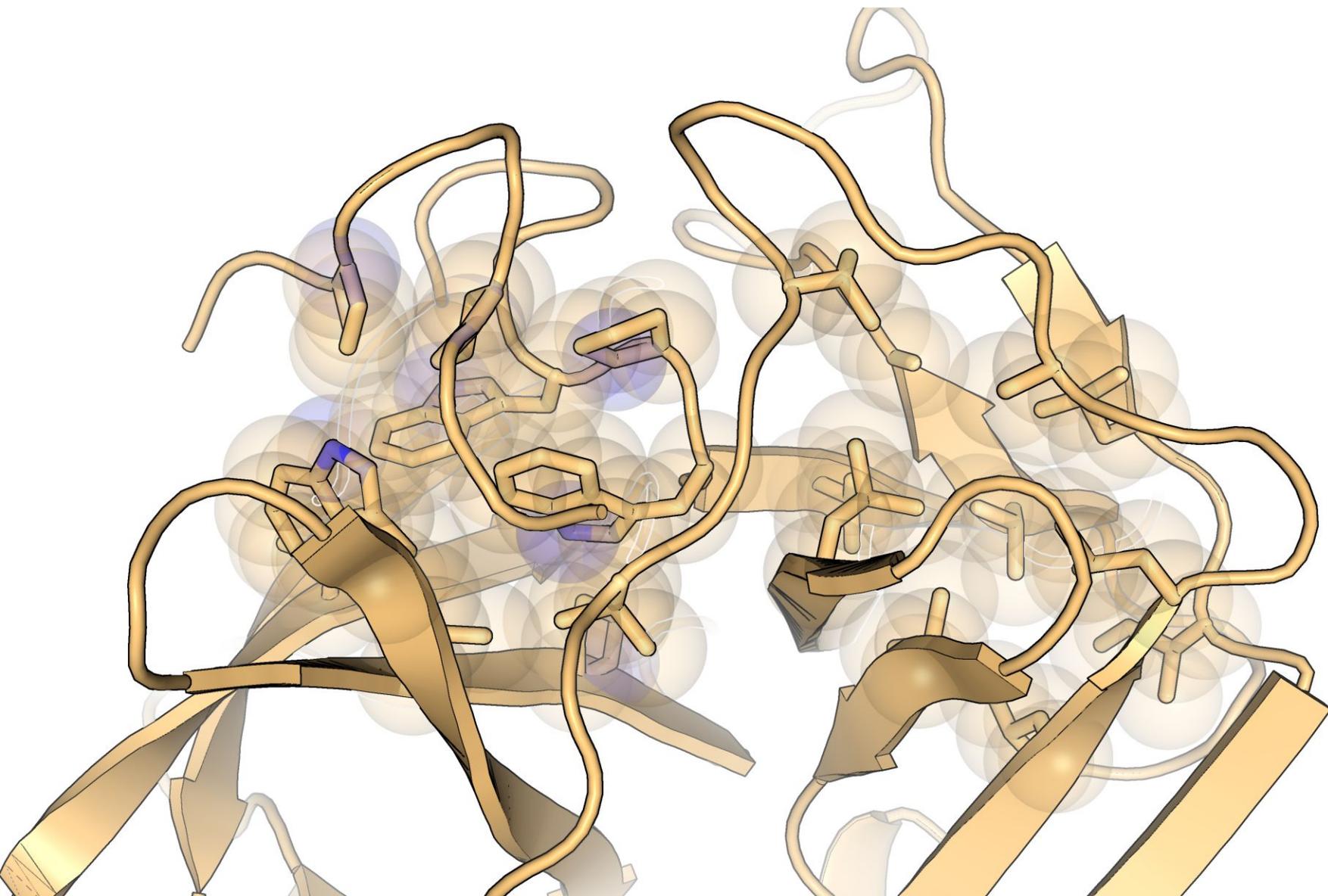
- Kannan & Vishveshwara, 1999
- Tsai & Nussinov, 1997
- Swindells, 1995
- Zehfus, 1995
- Heringa & Argos, 1991
- Plochocka et al., 1988
- Alexeevski et al., 2003

Swindells: группировка гидрофобно взаимодействующих неэкспонированных остатков

Отбираются остатки, которые

- Слабо экспонированы (<7%)
- Принадлежат спиральям или тяжам
- Более 75% контактов их атомов с другими атомами классифицируются как гидрофобные
- Контактom считается сближение “тяжелых” атомов на сумму ван-дер-ваальсовых радиусов + 1Å
- Гидрофобным контактом считается контакт атомов углерода

- Принадлежат спиральям или тягам



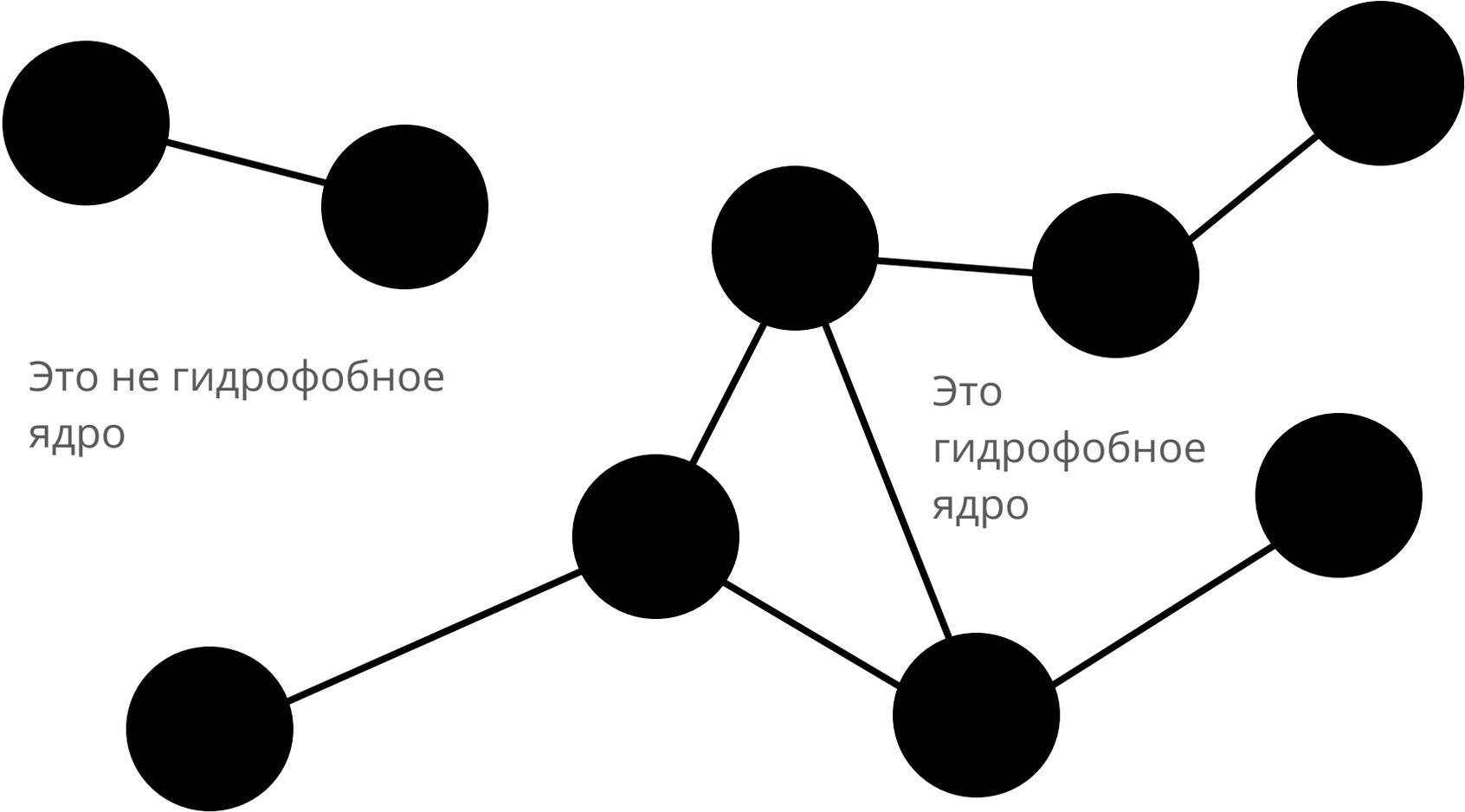
Swindells

Два остатка из отобранных считаются взаимодействующими гидрофобно, если число гидрофобных межатомных контактов превосходит число иных межатомных контактов.

Строится граф:

- Вершина – отобранный остаток
- Ребро соединяет вершины, если соответствующие остатки гидрофобно взаимодействуют
- Связные компоненты графа, содержащие 5 или более остатков, называются гидрофобными ядрами

Swindells



Alexeevski et al.: кластеры неполярных групп

- Элементарной единицей служат неполярные атомные группы (CH₃ и т.п.), а не аминокислотные остатки целиком
- Алгоритм основан на делении целого, а не на наращивании из элементов
- Цель – найти области пространства, заполненные преимущественно неполярными группами:

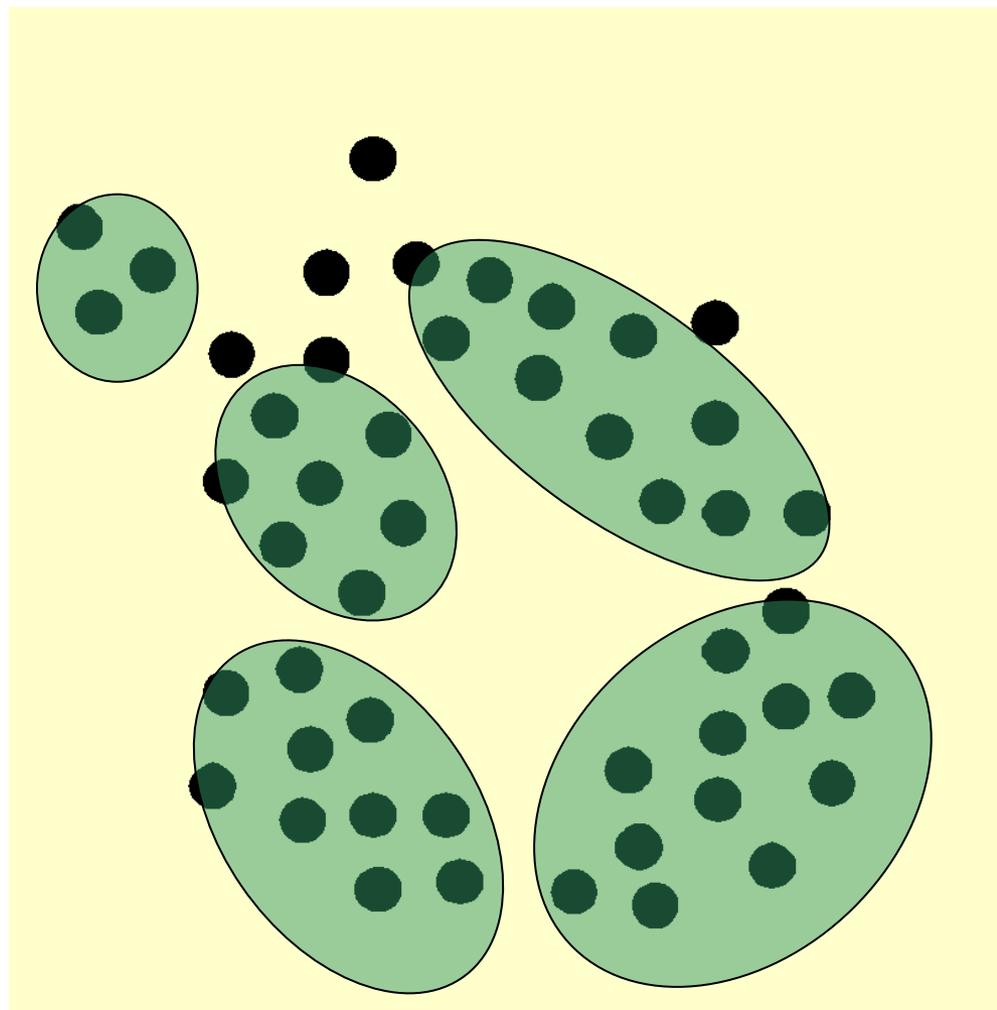


(при этом эти группы не должны быть связаны ковалентно с O или N)

Alexeevski et al.

“Гидрофобный кластер”:

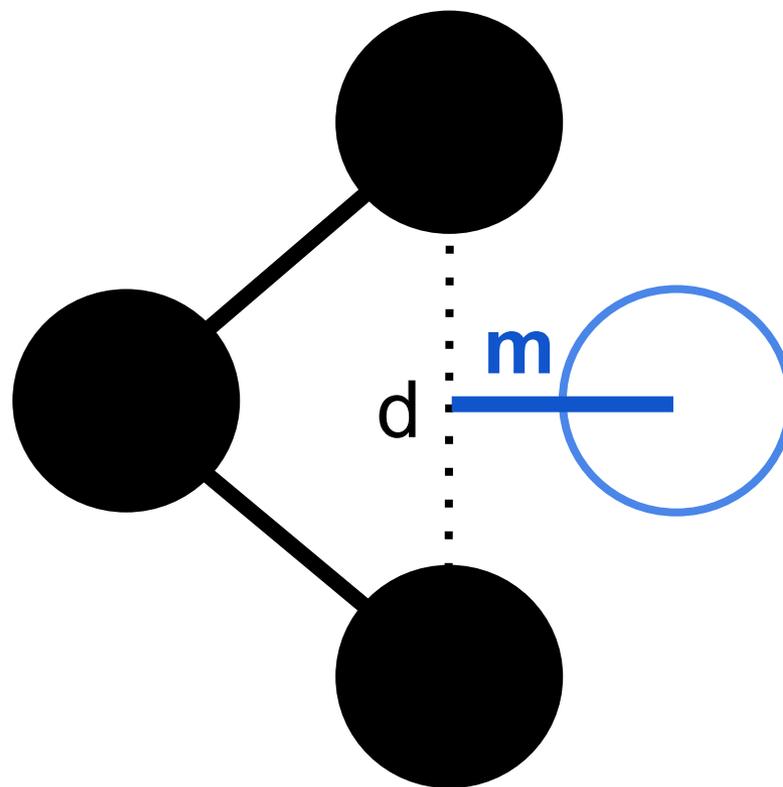
- состоит из неполярных групп
- каждая неполярная группа имеет несколько контактов с другими неполярными группами кластера
- взаимодействие между разными кластерами слабое



Alexeevski et al.

Шаги алгоритма

- Построить граф контактов неполярных атомных групп
Вершина - группа атомов
Два типа контактов:
ковалентные (сплошная линия)
и нековалентные (пунктир)
- Найти “(k, l)-разрезы” графа
k и l – параметры алгоритма, например,
 $k = 2, l = 1$
- Удалить все (k, l)-разрезы
- Связные компоненты получившегося графа объявляются гидрофобными кластерами



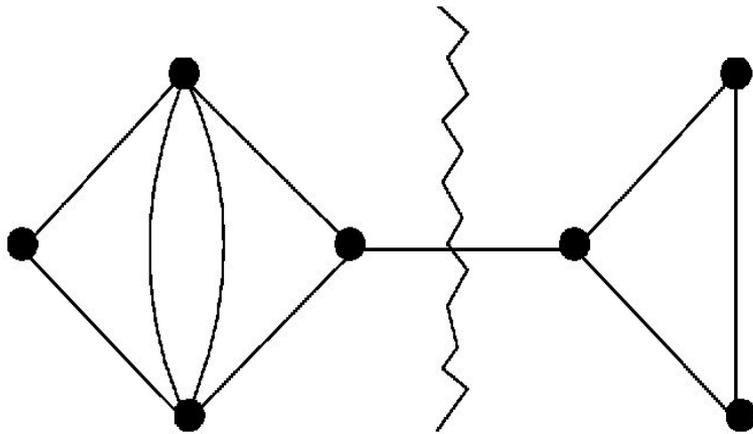
$d \leq d_0$, (d_0 – порог
расстояния, 3.5–5.4Å)

$m \geq m_0$ ($m_0 = d_0/2$)

Alexeevski et al.

Разрез и (k, l) -разрез графа

В теории графов k -разрезом называется подмножество k ребер такое, что граф, полученный удалением ребер подмножества, распадется на две или более связных компонент

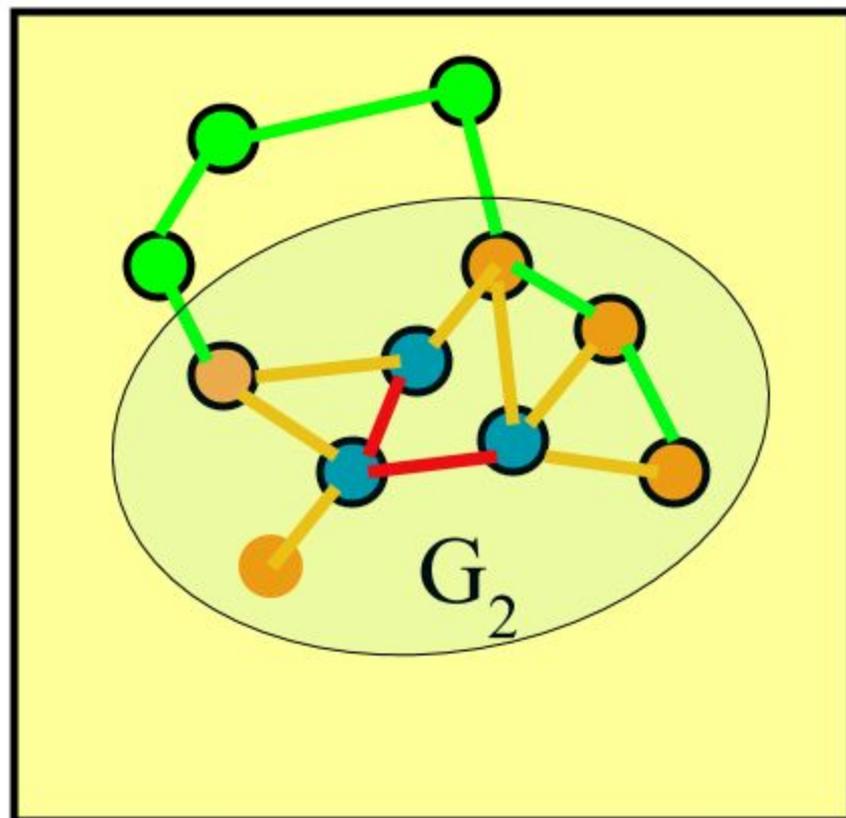
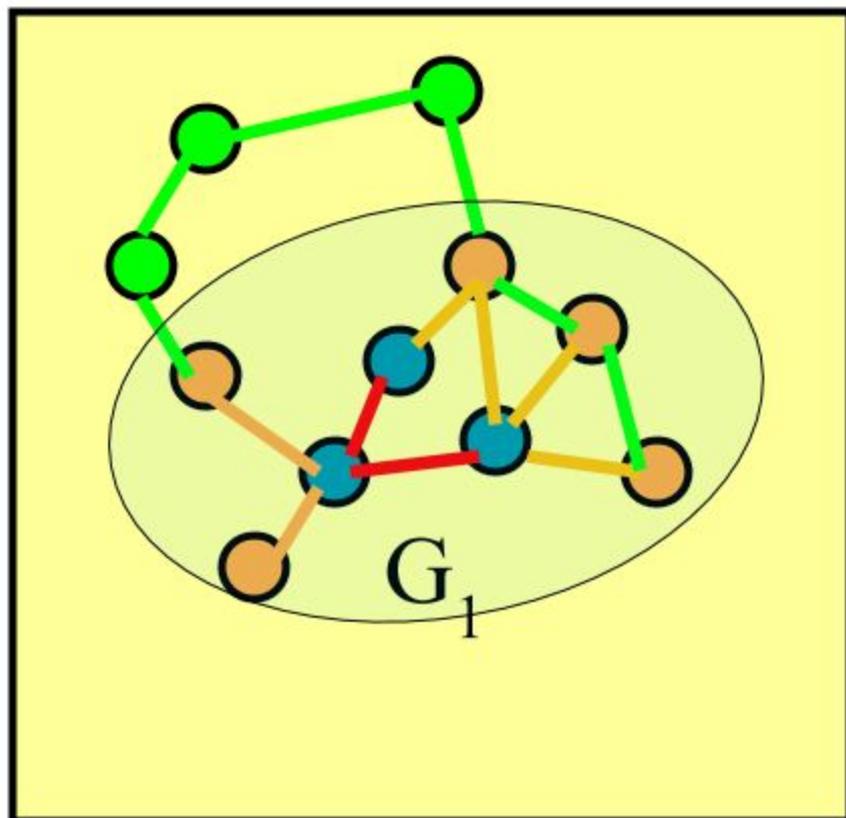


1-разрез графа

(k, l) -разрезом графа назовем связный подграф G с k ребрами такой, что l -реберная окрестность подграфа G после удаления ребер G распадается на две или более связных компоненты

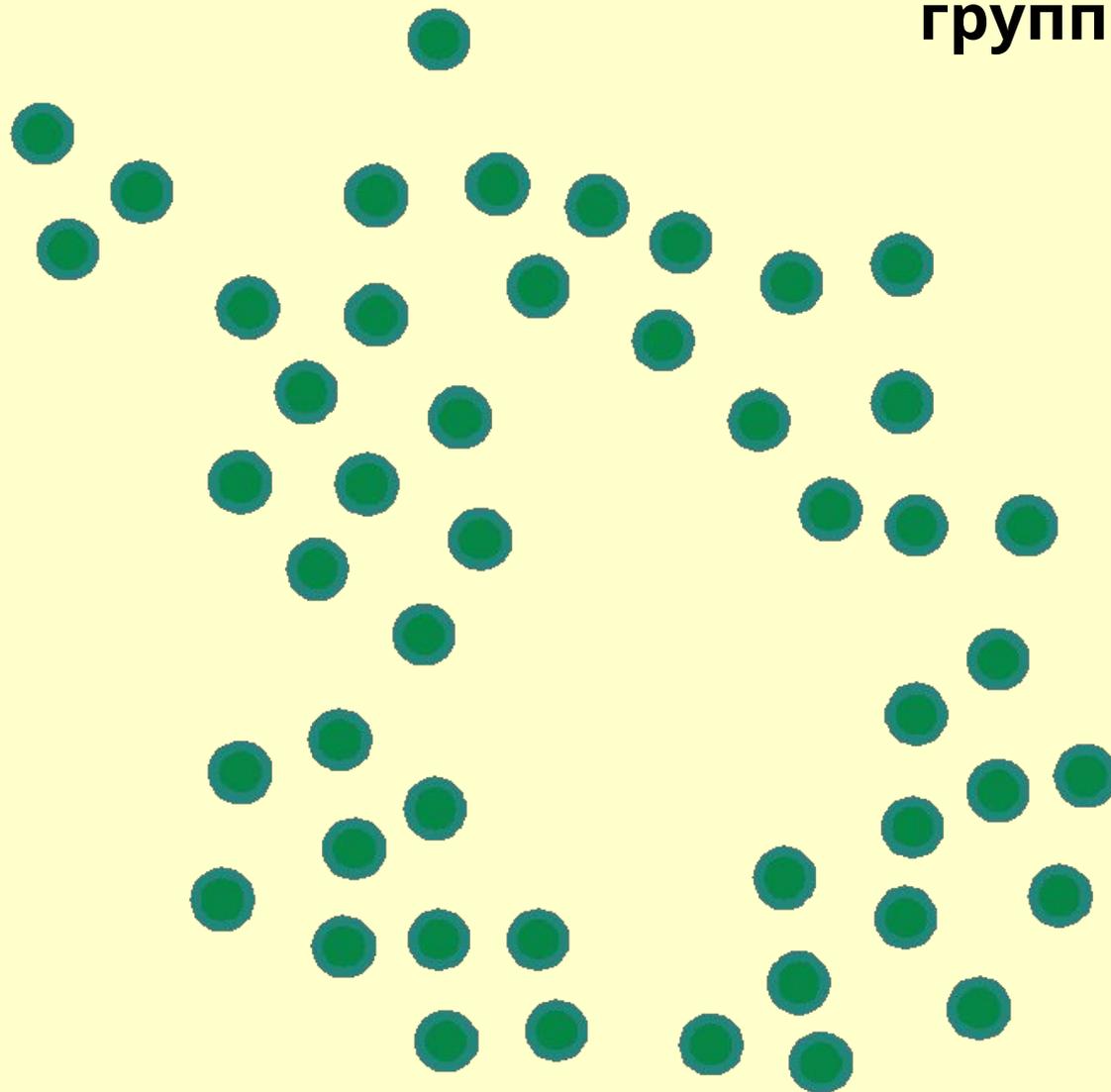
Подграф G_1 (красные
ребра) является $(2,1)$ -
разрезом

Подграф G_2 (красные
ребра) **не является**
 $(2,1)$ -разрезом

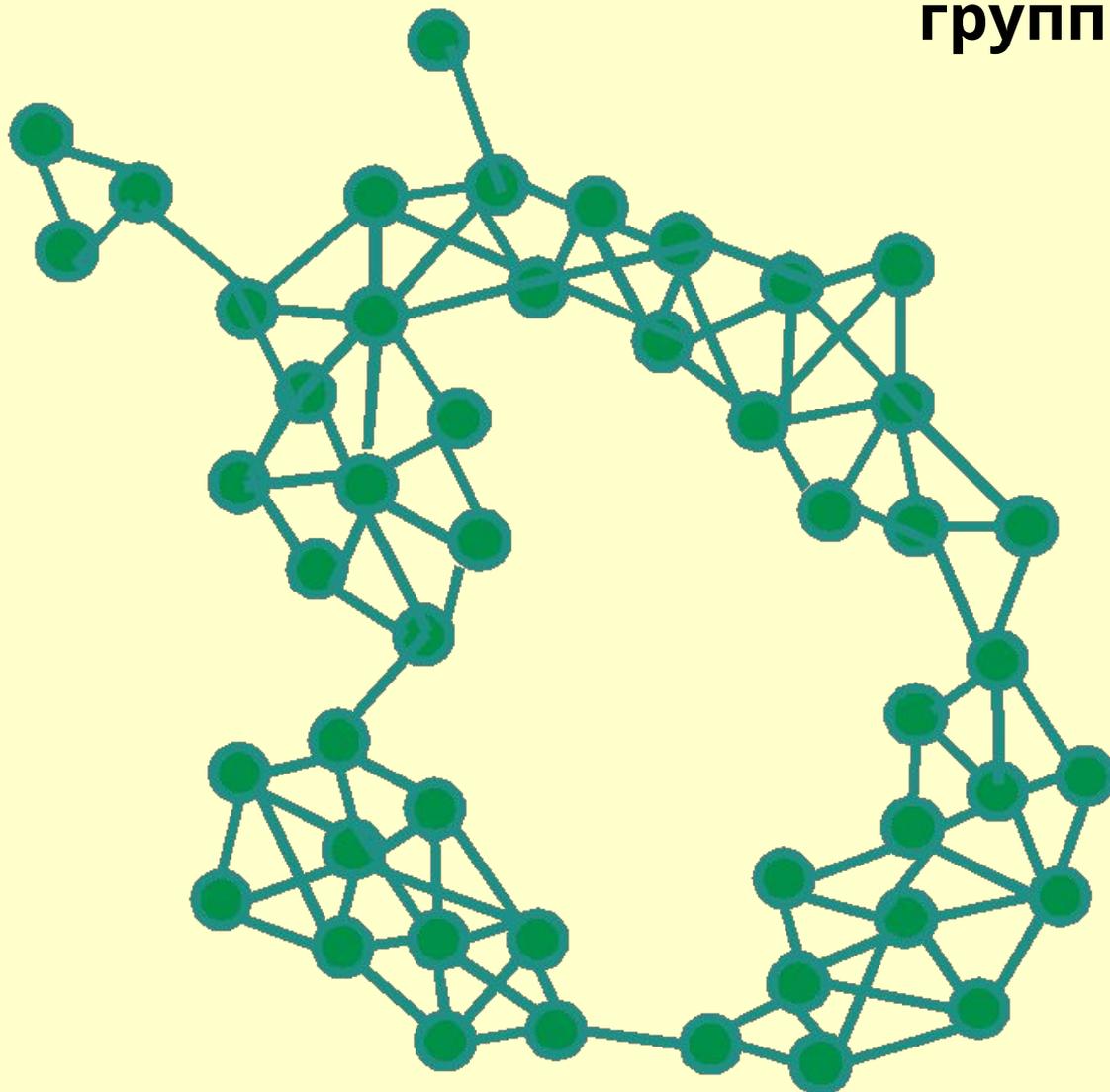


1-окрестности графов G_1 и G_2 обведены овалами

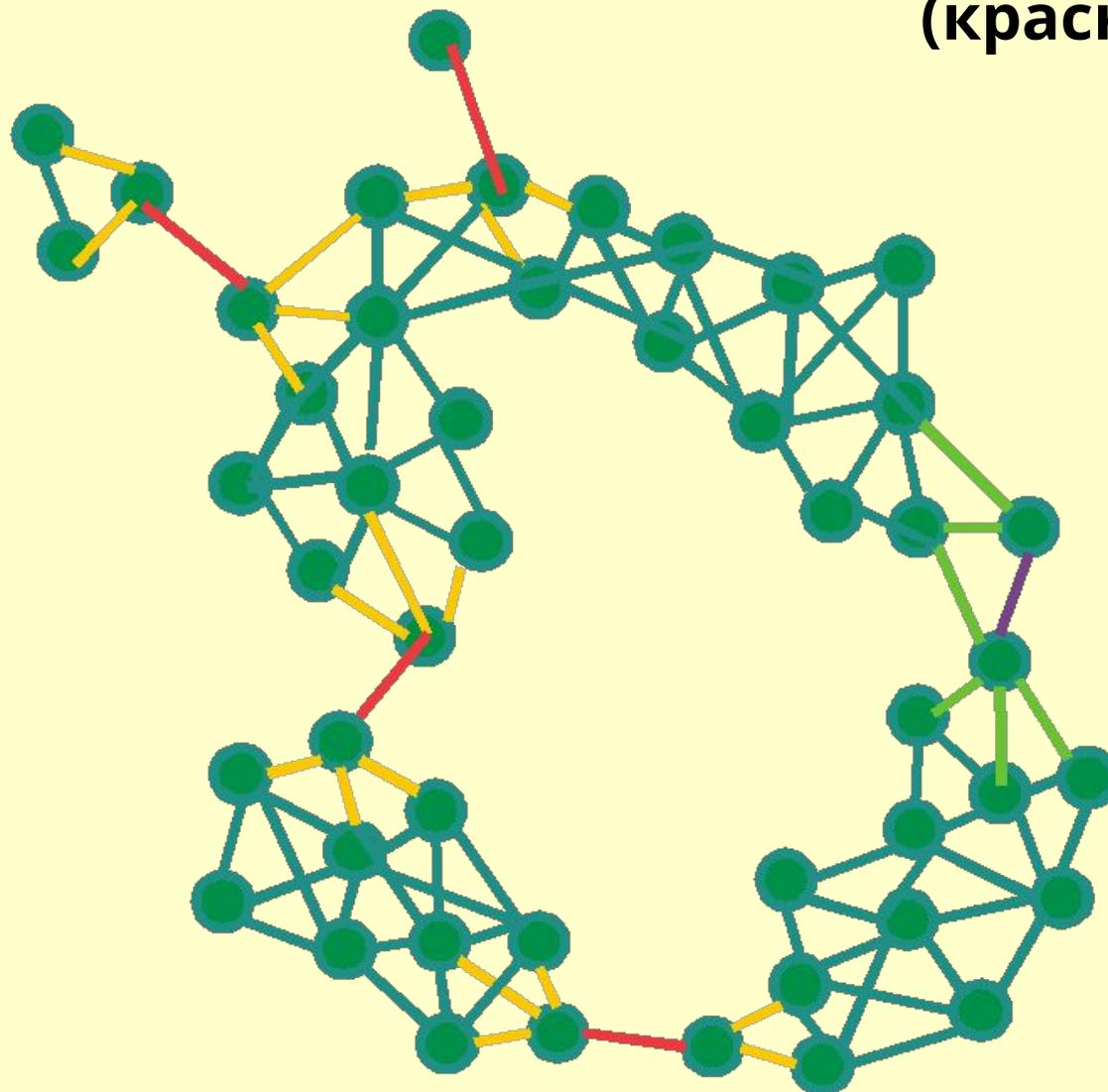
Неполярные группы



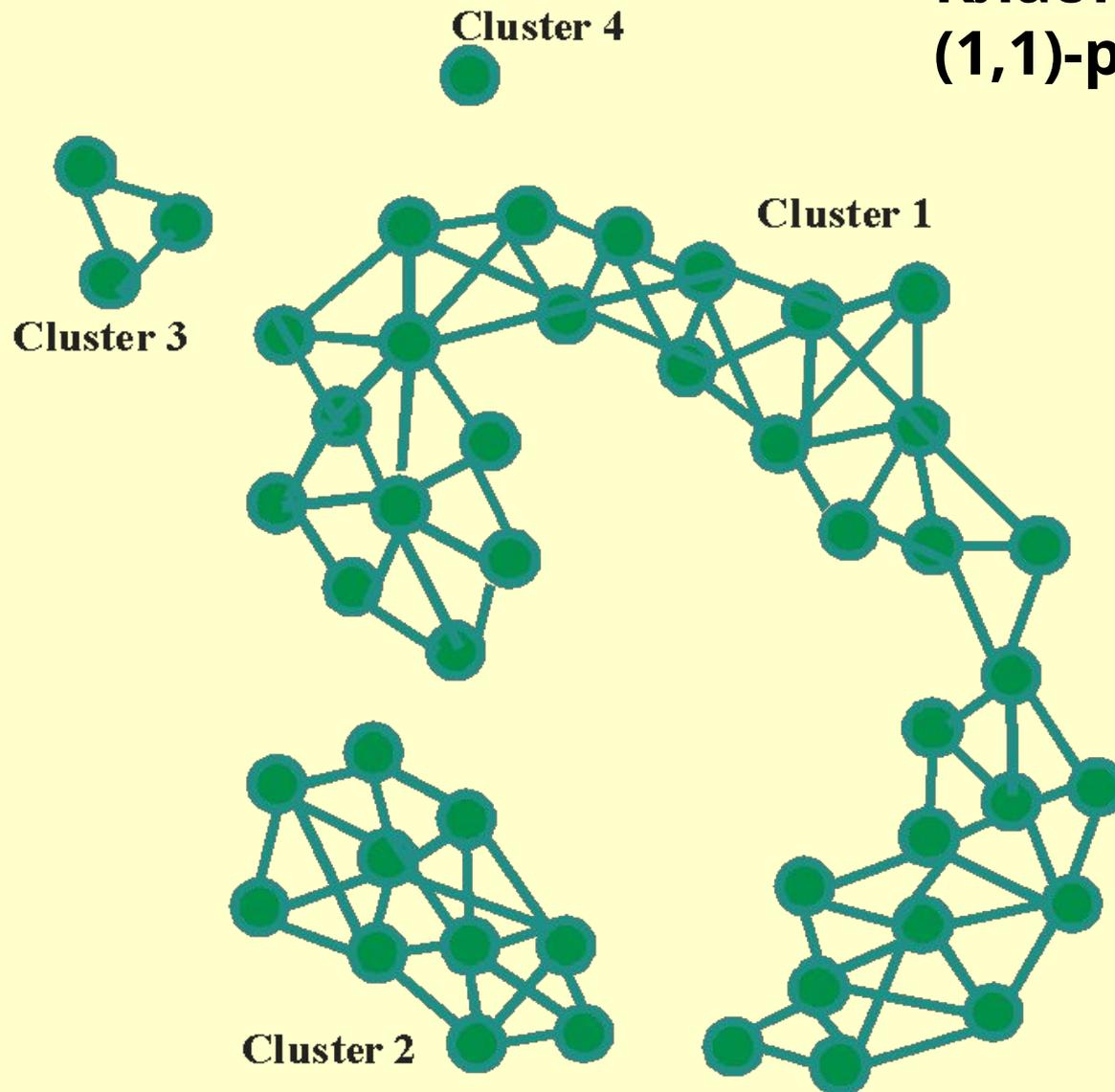
Граф неполярных групп



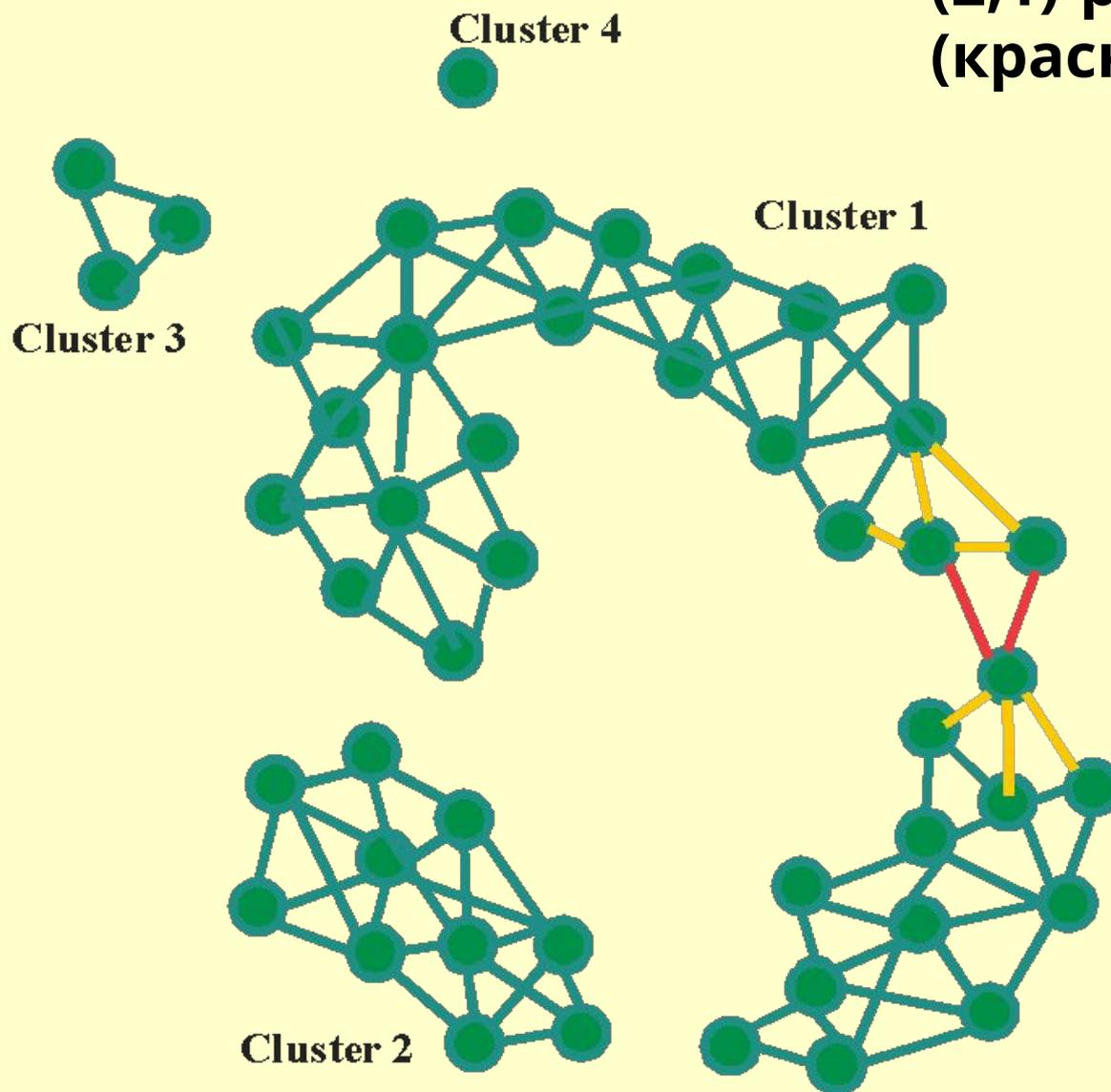
**(1,1)-разрезы
(красные ребра)**



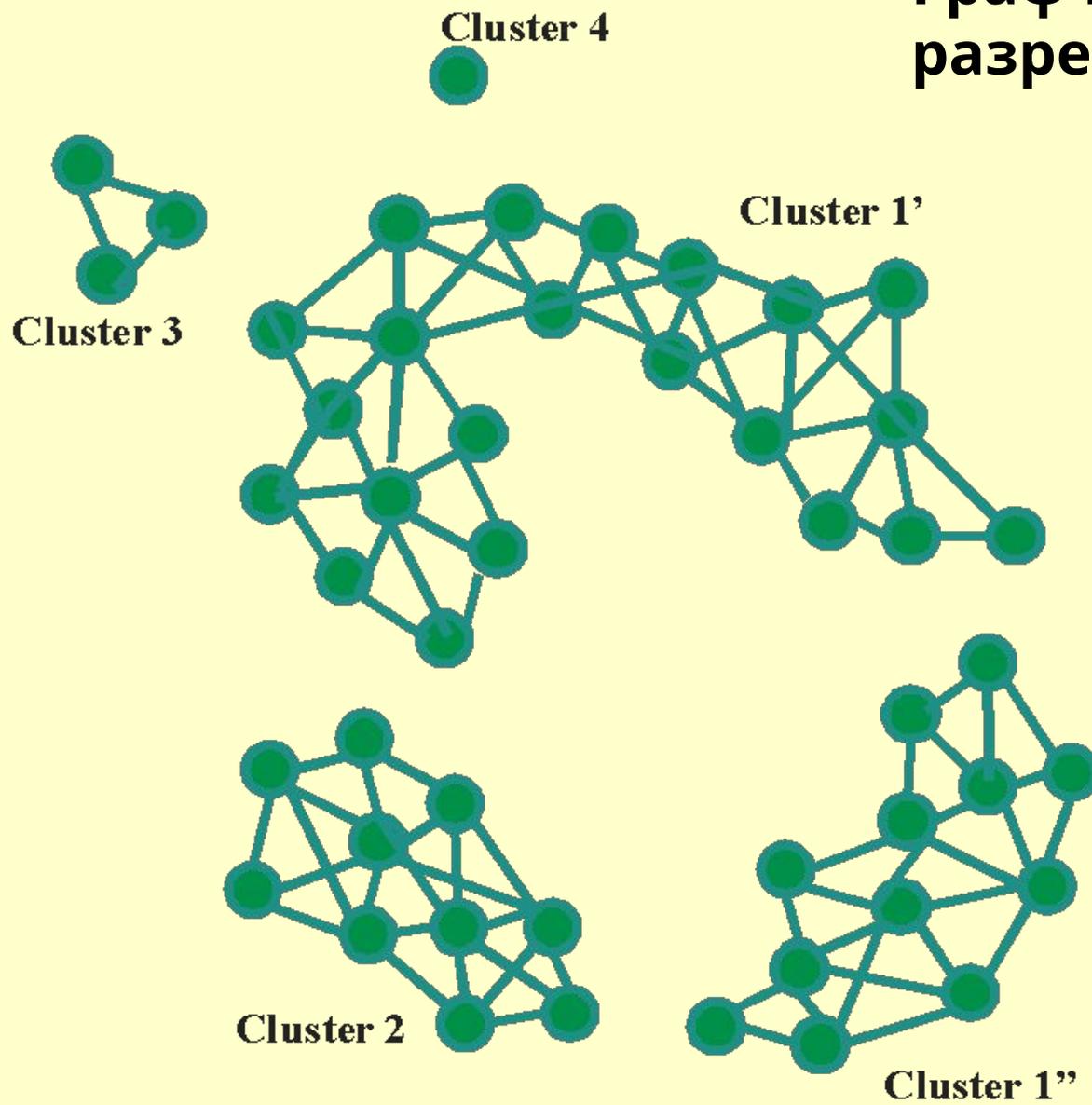
Кластеры после (1,1)-разрезов



(2,1)-разрез (красные ребра)



Граф после (2,1)-разреза



Hydrophobic Clusters in 3D structures

[NPIDB Home](#)

This form lets you to produce a list of [hydrophobic clusters](#) in a PDB file specified by user.

Enter a 4-letter PDB code,
e.g.: 1apl:

Search clusters: whole structure
 between chains

Enter your e-mail here
(optional):

Parameters:

Distance threshold

[List of atoms](#)

Chains filter

(1-20:A)

Show clusters not smaller than atoms

User data file in PDB format

No file selected.

<http://mouse.belozersky.msu.ru/npidb/cgi-bin/hftri.pl>

Домены

Что такое домен?

Три определения:

- По функции (функциональный домен)
- По сравнению последовательностей (эволюционный домен)
- По структуре (структурный домен)

Функциональный домен (биохимия/биоинженерия)

Минимальная часть полипептидной цепи, которая:

- сохраняет (in vitro) как минимум одну из активностей полного белка
- как правило, может автономно свернуться в правильную, нативную структуру

Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution

Victoria Derbyshire, David W. Wood, Wei Wu, John T. Dansereau, Jacob Z. Dalgaard, and Marlene Belfort

PNAS October 14, 1997 94 (21) 11466-11471; <https://doi.org/10.1073/pnas.94.21.11466>

Рекомбиназа А из *Mycobacterium tuberculosis* (790 а.о.) содержит интеин (440 а.о.), белок, обладающий способностью автономно вырезаться из полипептидной цепи белка-предшественника (явление белкового сплайсинга). Это – первая активность интеина.

There are 83 sequences with the following architecture: DnaB, DnaB_C, Intein_splicing, LAGLIDADG_3

[DNAB_MYCTU](#) [*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)] Replicative DNA helicase
EC=3.6.4.12 EC=3.1.-.- (874 residues)

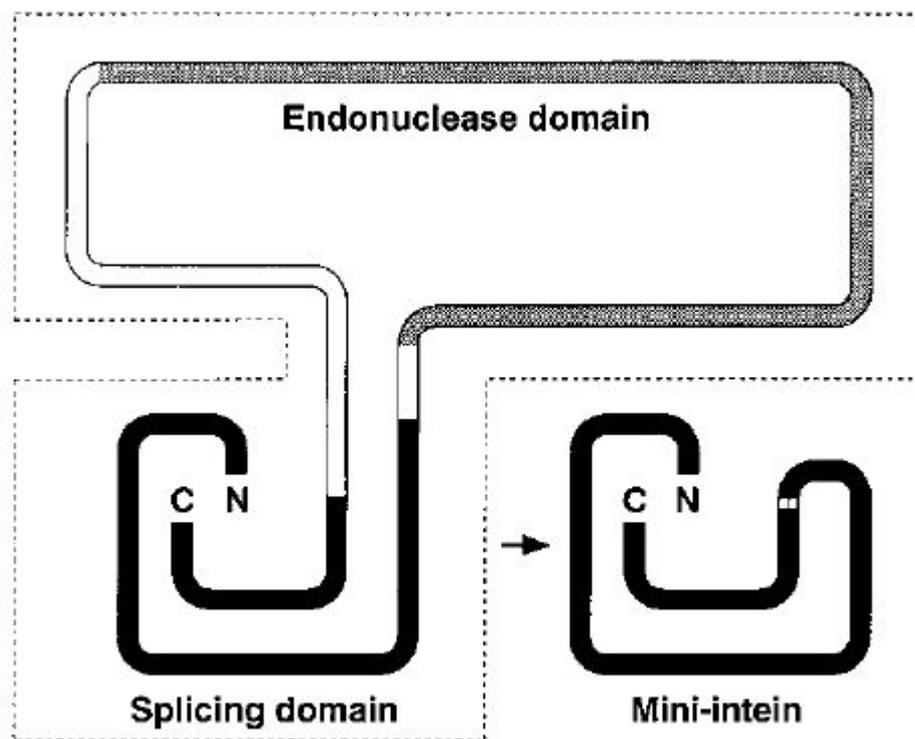


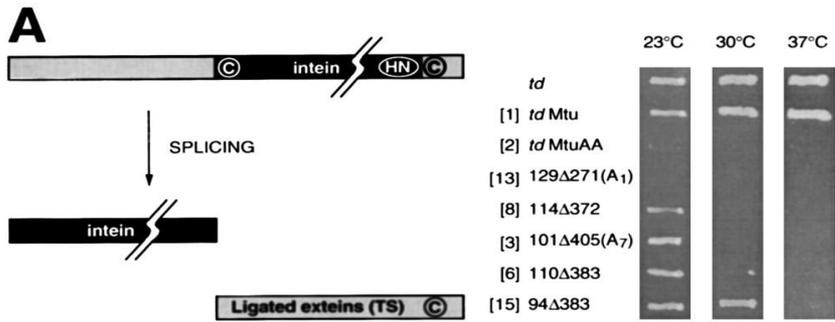
[Show](#) all sequences with this architecture.

Этот интеин обладает также эндонуклеазной активностью (вторая активность).

По сходству последовательности этого белка с последовательностями других, более изученных интеинов, в т.ч. интеина с расшифрованной РСА структурой (1VDE), была высказана гипотеза о том, что за две разные активности отвечают разные домены.

При этом за белковый сплайсинг отвечает домен, который составлен из N-концевого и C-концевого участков полипептидной цепи





Для проверки гипотезы авторы создали 21 конструктор генов интеина, в которых удалены разные внутренние участки полипептидной цепи.

B

	Intein size (aa)	TS phenotype	MIC splicing
1	440	+5	+5
2	440	-	-
3	137	+2	0*
4	147	+1	+1
5	159	+1	+1
6	168	+2	+3
7	171	+3	+2
8	183	+1	+2
9	192	+1	+1
10	165	+1	+1*
11	187	+1	+1
12	198	+1	+1
13	299	-	+1
14	154	+3	+2
15	152	+3/+4	+2
16	149	-	-
17	149	+2	+1
18	170	+1	+1
19	160	-	-
20	142	+2	+1
21	137	+1	+1

endonuclease/spacer

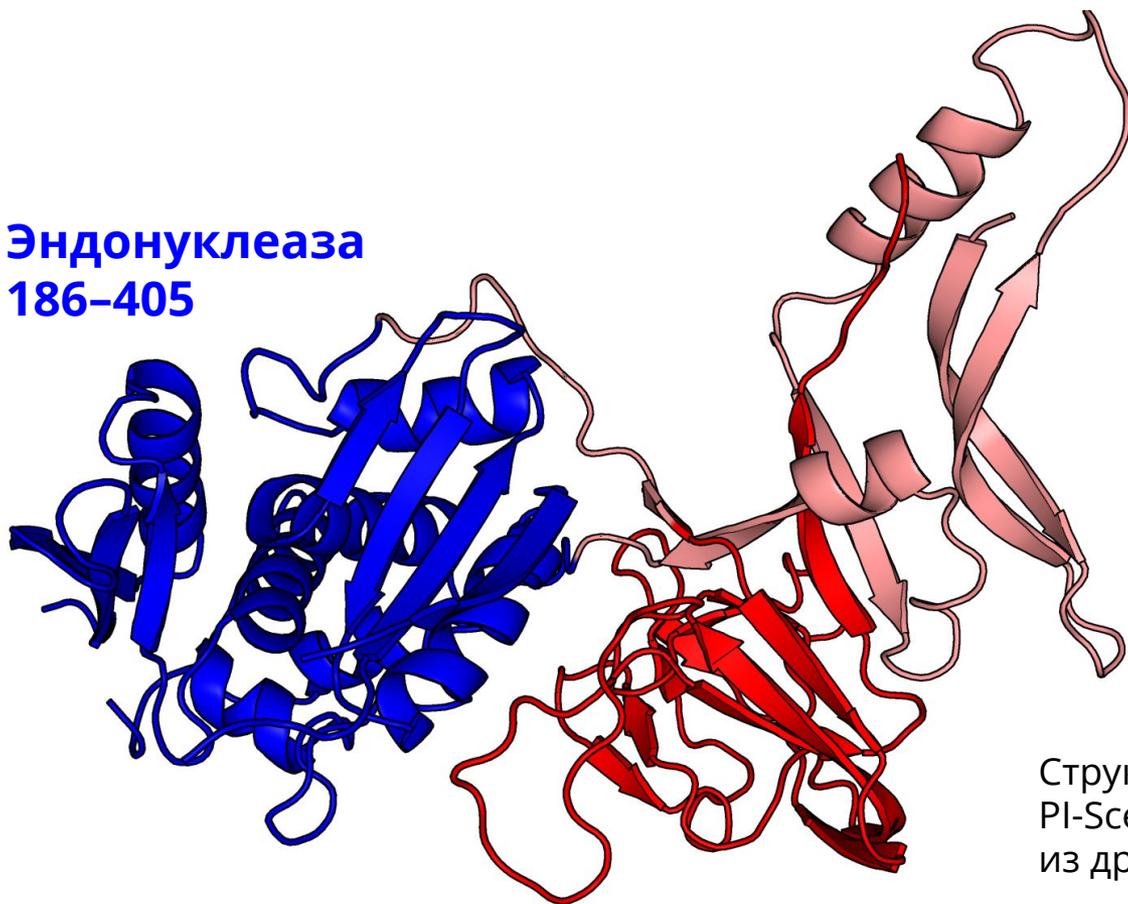
Конструкторы были встроены в ген другого белка (тимидилатсинтазы, TS) и экспрессировались в *E.coli*

Активность проверялась по наличию нативного белка TS (без вставки интеина)

Результат: белковый сплайсинг сохранялся в тех случаях, когда удаленный участок не затрагивал первые 96 и последние 35 а.о.

Вывод авторов: функциональный домен автономного белкового сплайсинга состоит из остатков 1–96 и 406–440 (всего 131 из полных 440)

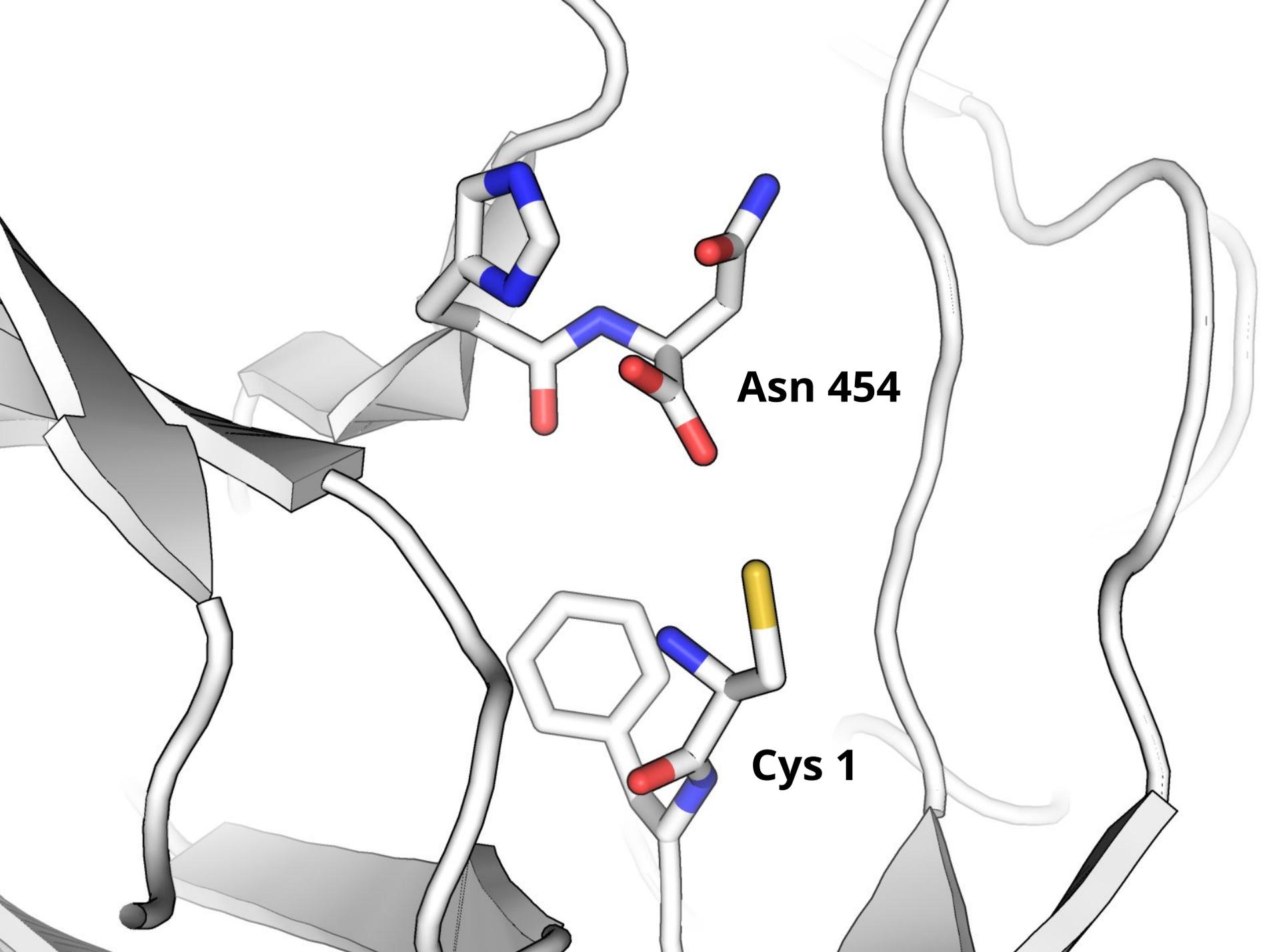
Эндонуклеаза
186–405



Интеин
1–181,
406–454

Минимально
функциональный
интеин
1–96,
419–454

Структура гомологичного белка
PI-SceI – хоминг эндонуклеазы
из дрожжей (PDB код 1VDE)



Последовательность интеина консервативна

Об этом свидетельствуют доменные архитектуры трех белков из разных грибов, описанные в Pfam



Доменная структура полноразмерного белка PI-Scel



Доменная структура белка VMA1



Доменная структура белка TFP1 (аннотирован по сходству)

Эволюционный домен (биоинформатика: последовательности)

Достаточно длинный участок полипептидной цепи, который:

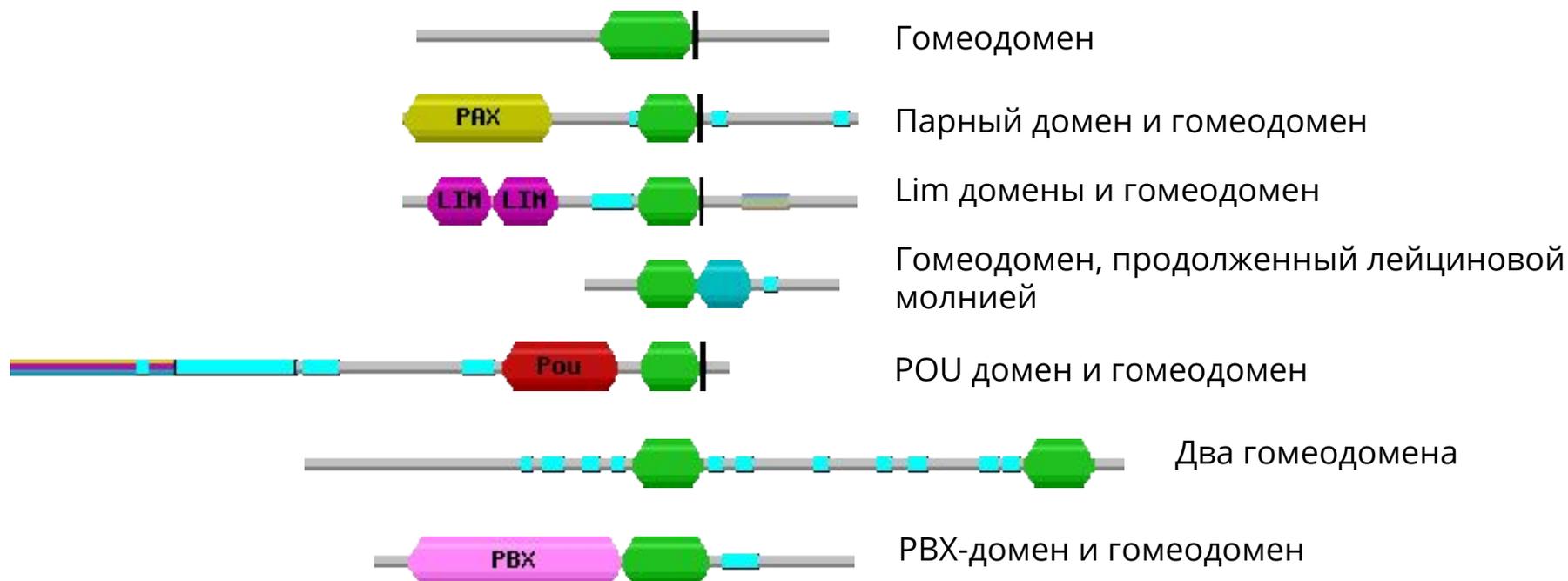
- эволюционно консервативен — существуют достоверно сходные участки в других белках
- Обычно замечен в перемешивании доменов (domain shuffling), то есть имеются примеры белков, где есть достоверно сходные с ним участки, но есть также несходные между собой (но эволюционно консервативные) участки

Белки, содержащие два эволюционных домена: гомеодомен и OAR домен (N-концевые участки не показаны)

SW: PDK1_CHICK/1	-----	20	-----	40	-----	60	-----	80	-----	100	-----	120	-----	140	-----	1	-----	80	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	86	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	80	
SW: RX2_BRARE/1	ISQAPQV	SISRSKSYREN	-APFSQS	-----	D	EQGSP	-----	EHMAQELVELST	-----	LKFEEDV	VVKEACGDN	-----	S	-----	-----	-----	-----	84	
SW: RX2_MOUSE/1	ISQAPQV	SISRSKSYRENG	APFVPPPPALD	-----	ELSCPG	CVVAHP	-----	FEERLSAAS	-----	CGPSAPA	-----	AACGGT	-----	CAEDDE	-----	EELEDE	-----	EEDE	84
SW: AL_DROME/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	72	
SW: ALX4_MOUSE/1	-TFLS	SACAGCG	FGDAKSR	-----	RYRAC	QODLAA	-----	PLESSSG	-----	ARGCS	-----	FMKFP	-----	QPPPT	-----	QP	-----	145	
SW: ALX4_HUMAN/1	-TFLS	AAAAK	AGCGFGDAKSR	-----	RYRAC	QODLAT	-----	PLESCAG	-----	ARGCS	-----	FMKFP	-----	QPPPT	-----	QP	-----	157	
SW: RX2_CHICK/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	83	
SW: RX2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	92	
SW: RX1_XENLA/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	91	
SW: RX_HUMAN/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	97	
SW: PDK2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	66	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	68	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	64	
SW: OTP_MOUSE/1-	-----		-----		-----		-----		-----		-----		-----		-----		-----	90	
SW: PDK1_CHICK/1	-----	60	-----	180	-----	200	-----	220	-----	240	-----	260	-----	280	-----	300	-----	213	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	221	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	213	
SW: RX2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	230	
SW: RX2_MOUSE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	303	
SW: AL_DROME/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	212	
SW: ALX4_MOUSE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	290	
SW: ALX4_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	302	
SW: RX2_CHICK/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	215	
SW: RX2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	225	
SW: RX1_XENLA/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	224	
SW: RX_HUMAN/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	242	
SW: PDK2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	212	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	215	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	218	
SW: OTP_MOUSE/1-	-----		-----		-----		-----		-----		-----		-----		-----		-----	236	
SW: PDK1_CHICK/1	-----	320	-----	340	-----	360	-----	380	-----	400	-----	420	-----					245	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	253	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	245	
SW: RX2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	336	
SW: RX2_MOUSE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	404	
SW: AL_DROME/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	313	
SW: ALX4_MOUSE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	354	
SW: ALX4_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	366	
SW: RX2_CHICK/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	290	
SW: RX2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	297	
SW: RX1_XENLA/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	296	
SW: RX_HUMAN/1-1	-----		-----		-----		-----		-----		-----		-----		-----		-----	319	
SW: PDK2_BRARE/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	314	
SW: PDK2_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	317	
SW: PDK1_HUMAN/1	-----		-----		-----		-----		-----		-----		-----		-----		-----	314	
SW: OTP_MOUSE/1-	-----		-----		-----		-----		-----		-----		-----		-----		-----	325	

Гомеодомены активно перемешивались в эволюции

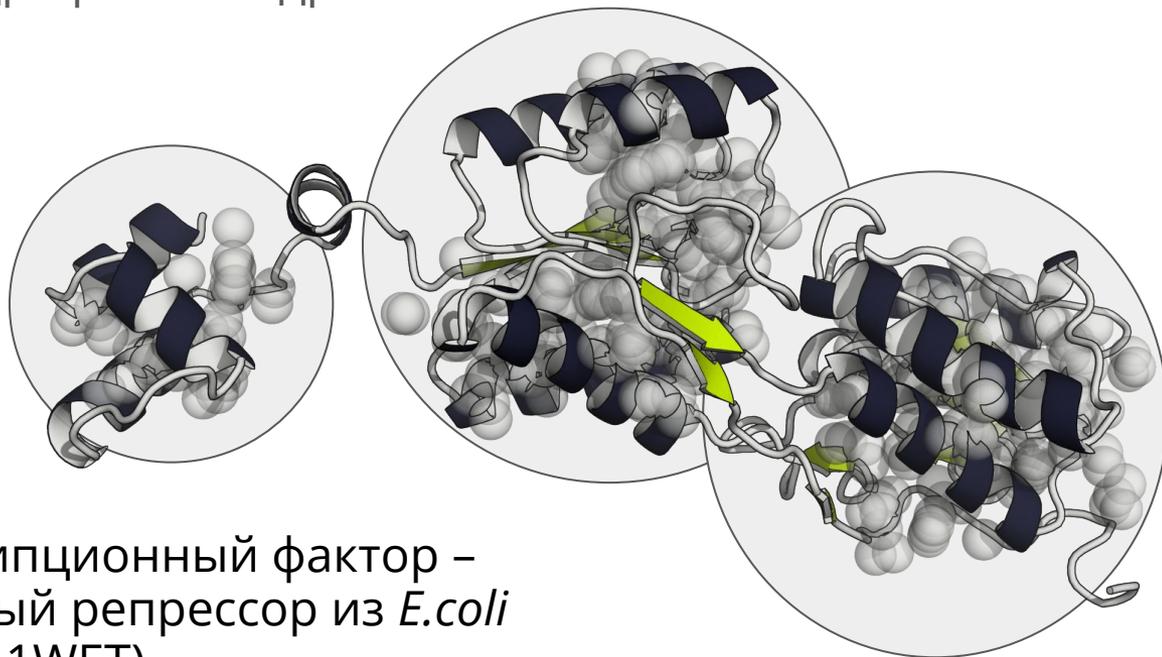
Об этом можно судить по 609 ! (2018 год) различным доменным архитектурам гомеобелков (Pfam)



Структурный домен (биоинформатика 3D структур)

Обособленная в пространстве часть белка,
его структурная единица, имеющая:

- сравнительно мало контактов с другими частями белка
- собственное гидрофобное ядро



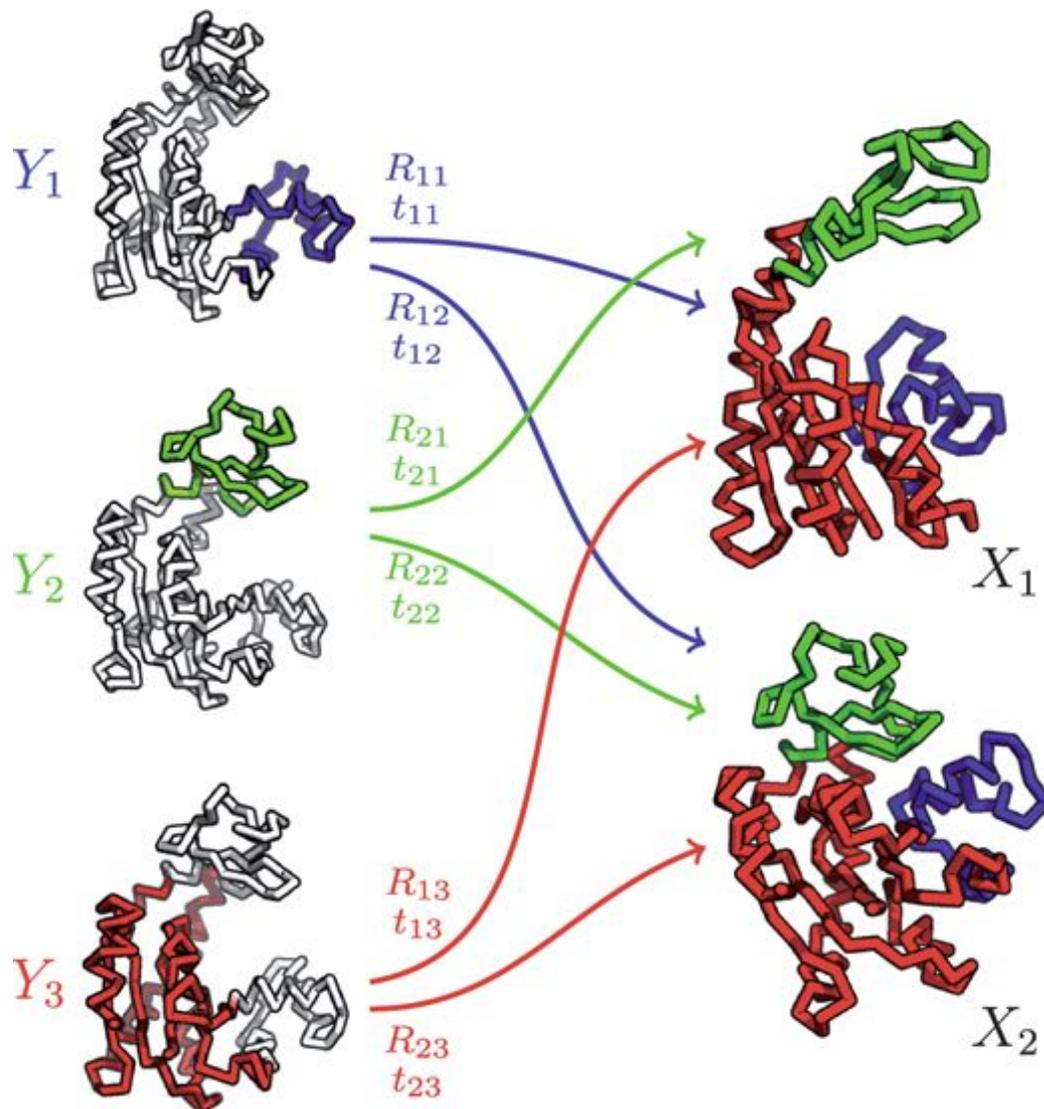
Транскрипционный фактор –
пуриновый репрессор из *E.coli*
(PDB код 1WET)

Структурный домен (биофизика)

Часть молекулы
белка, движущаяся
как твердое тело

A probabilistic model for detecting rigid domains in protein structures

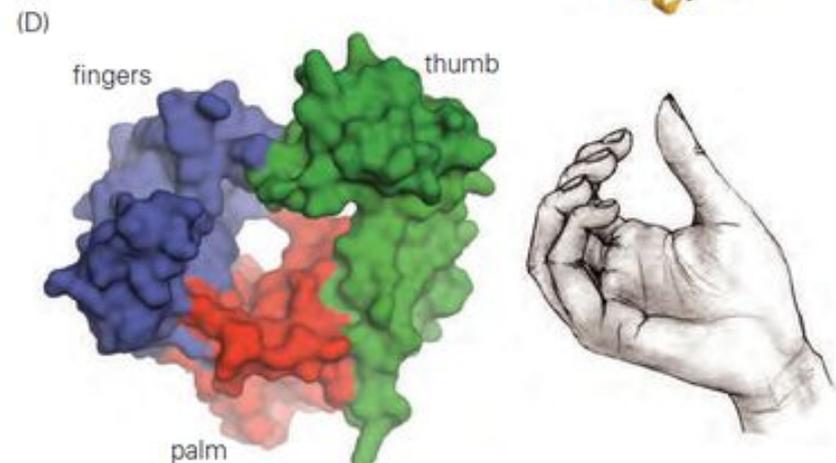
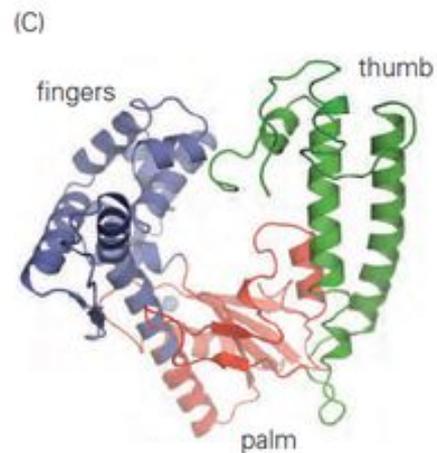
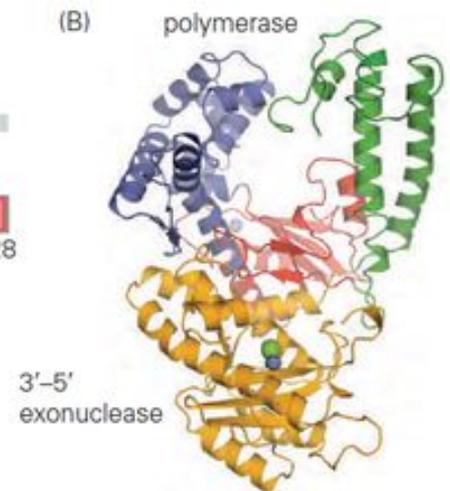
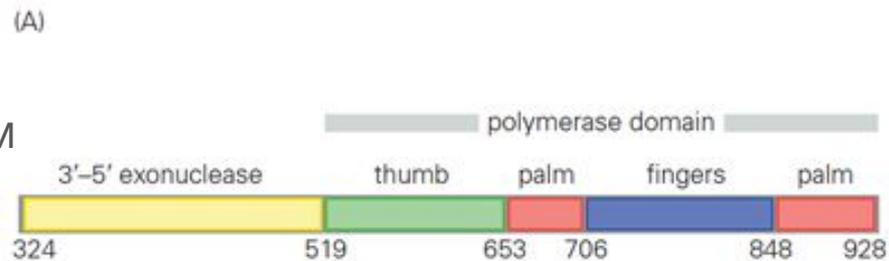
Thach Nguyen, Michael Habeck
Bioinformatics, Volume 32, Issue 17, 1
September 2016



Домен белка XXX (жизнь)

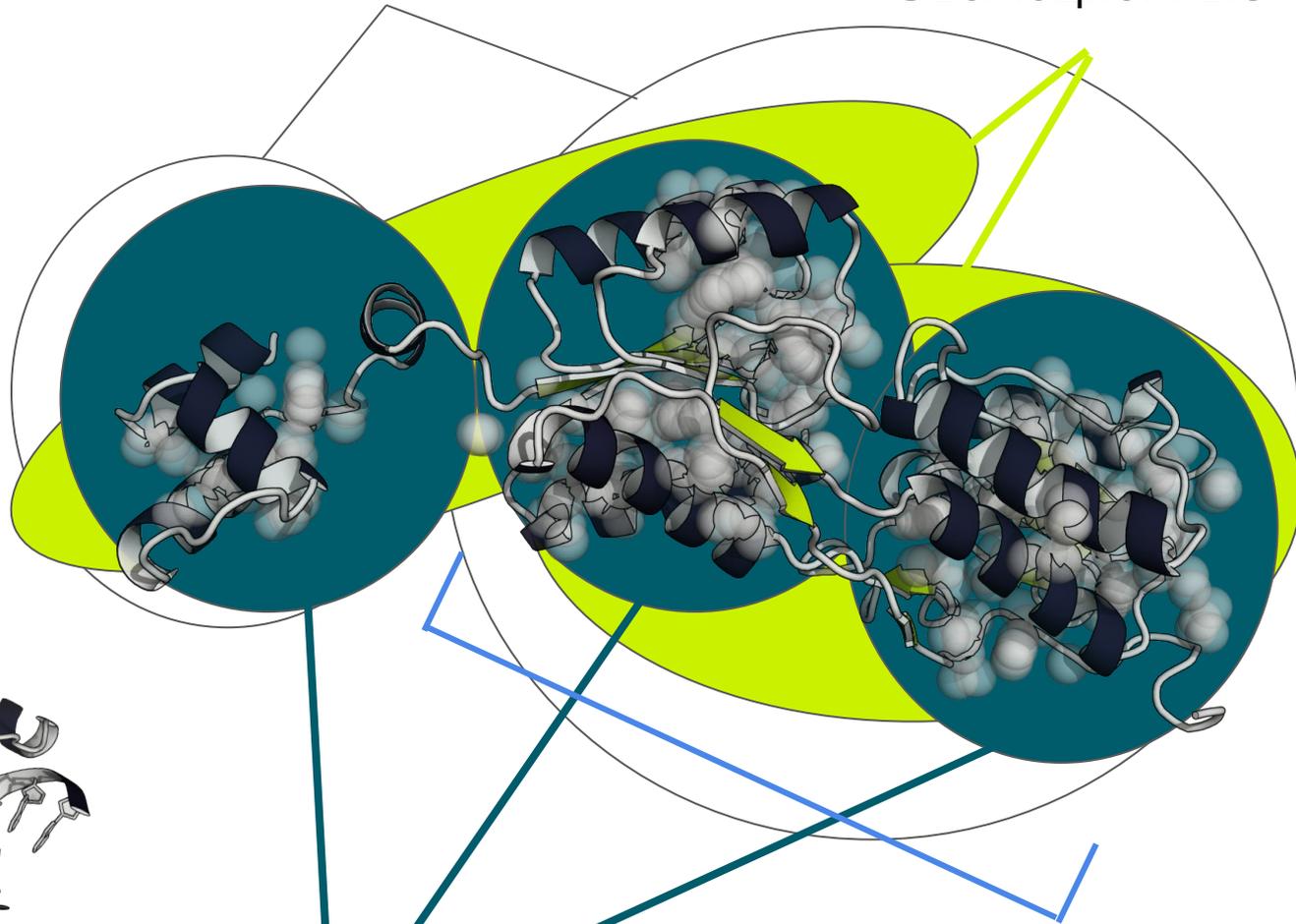
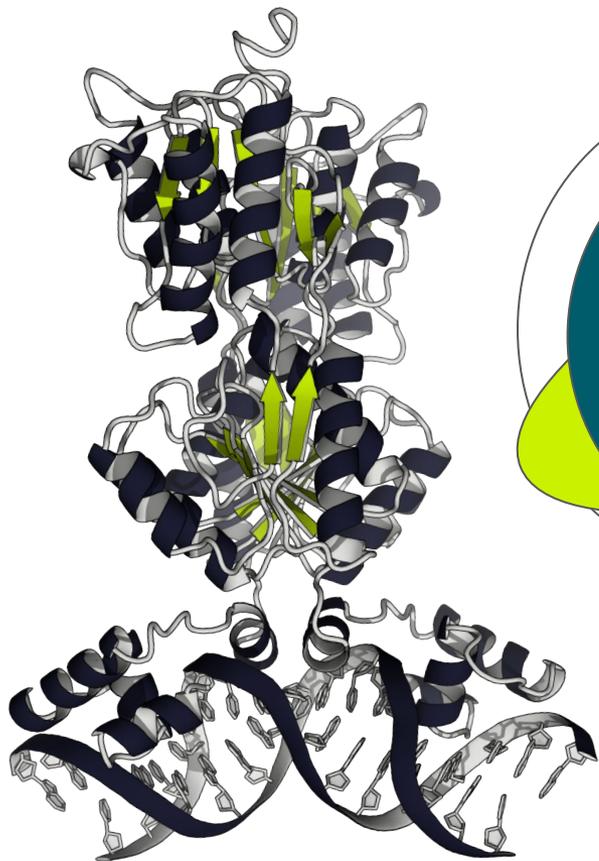
Часть белка, названная доменом

- Субъективизм
- Образность
- Традиция



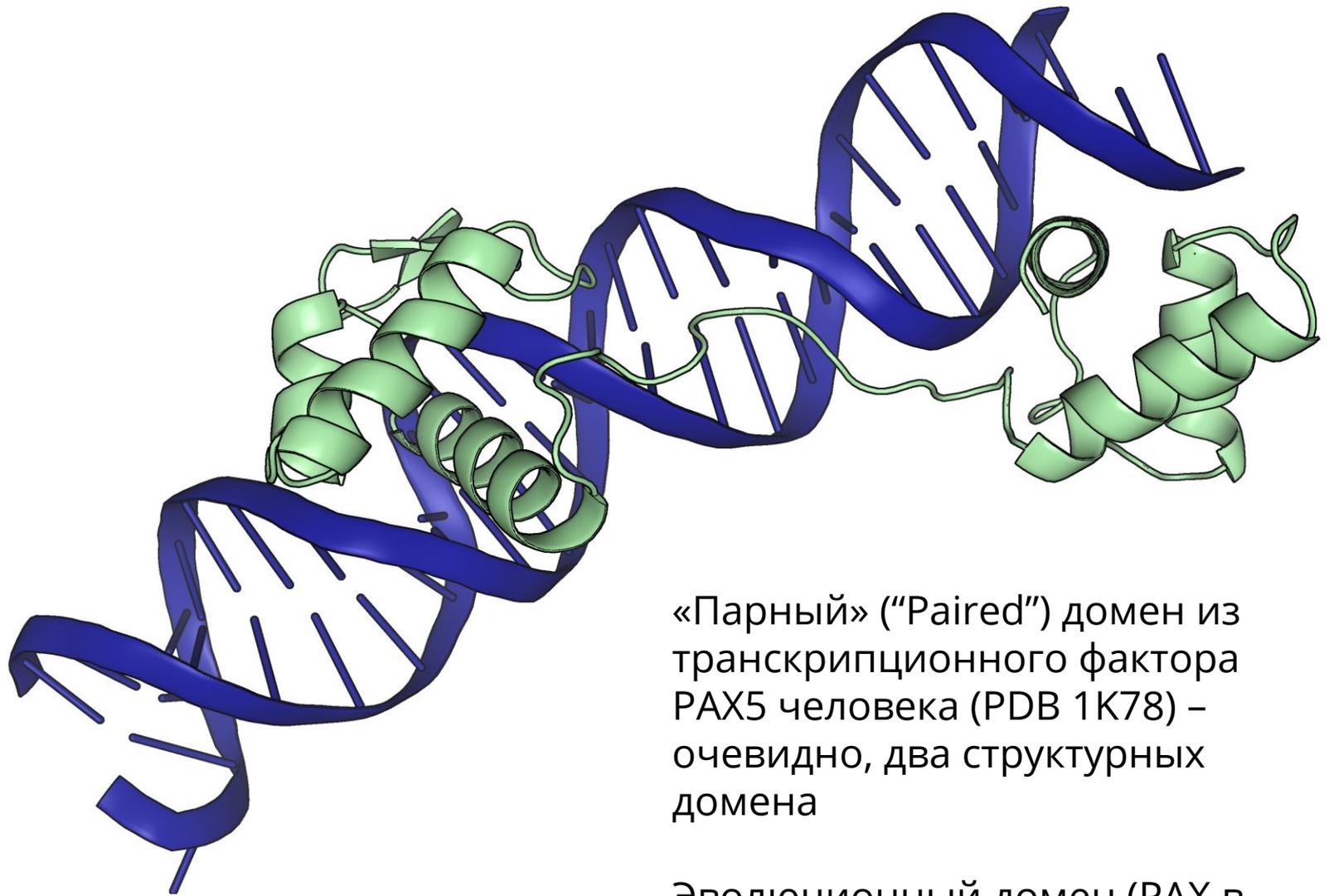
Функциональные

Эволюционные



Структурные

Классифицируется
(БД SCOP) как один
структурный домен



«Парный» ("Paired") домен из транскрипционного фактора PAX5 человека (PDB 1K78) – очевидно, два структурных домена

Эволюционный домен (PAX в Pfam) включает оба структурных домена (126 а.о.)

Последовательности PAX/prd доменов консервативны по всей длине

```

                *           20           *           40           *           60
PAX6_BRARE : SHSGVNQLGGVVFVNGRPLPDSTRQKIVELAHSGARPCDISRIHQVSNCGCVSKILGRYYETGSIRP : 65
PAX2_HUMAN : GHGGVNQLGGVVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKP : 65
PAX1_MOUSE : TYGEVNQLGGVVFVNGRPLPNAIRLRIVELAQLGIRPCDISRQLRVSHGCVSKILARYNETGSILP : 65
PAX3_HUMAN : GQGRVNQLGGVVFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLRVSHGCVSKILCRYQETGSIRP : 65
GSBN_DROME : GQGRVNQLGGVVFINGRPLPNHIRLKI VEMAASGVRPCVISRQLRVSHGCVSKILNRYQETGSIRP : 65
GSB_DROME/ : GQGRVNQLGGVVFINGRPLPNHIRRQIVEMAAAGVRPCVISRQLRVSHGCVSKILNRFQETGSIRP : 65
g VNQLGGVVF6NGRPLP1 R IVE6A G RPC ISRqLrVShGCVSKIL R5 ETGSI P

```

```

                *           80           *           100           *           120
PAX6_BRARE : RAIGGSKPR-VATPEVVGKIAQYKRECPSIFAWWEIRDRLISEGVCTNDNIPSVSSINRVLR : 125
PAX2_HUMAN : GVIGGSKPK-VATEKVVVDKIAEYKRQNPTMFAWEIRDRLIAEGICDNDTVPSVSSINRIIR : 125
PAX1_MOUSE : GAIGGSKPR-VTTPNVVKHIRDYKQGDPIFAWEIRDRLIADGVCDKYNVPSVSSISRILR : 125
PAX3_HUMAN : GAIGGSKPKQVTTDPVEKKIEEYKRENPGMFSWEIRDKLLKDAVCDRNTVPSVSSISRILR : 126
GSBN_DROME : GVIGGSKPK-VTSPEIETRIDELRKENPSIFSWEIREKLIKEGFAD---PPSTSSISRLLR : 122
GSB_DROME/ : GVIGGSKPR-VATPDIESRIEELKQSQPGIFSWEIRAKLIEAGVCDKQNAPSVSSISRLLR : 125
g IGGSKP4 V 3P 6 I 4 P 6F WEIR 4L6 g cd PSvSSI R66R

```



N-концевой структурный домен
парного домена
хорошо совмещается с C-концевым

Синий – N-концевой

Зеленый – C-концевой

Но достоверного сходства
последовательностей не наблюдается

Домены

Алгоритмы

3D-координаты



Подход сверху-вниз

Подход снизу-вверх

Подразделение цепи
на меньшие элементы

Объединение элементов
вторичной структуры



Оценка каждого домена



Принятие/не принятие такого присвоения



Предсказанные домены

- Макс. гидрофобное ядро элемента
- Макс. компактность
- Мин. площадь интерфейса
- Наличие точек перегиба между элементами

- Мин. размер элемента
- Макс. глобулярность
- Мин. разрезание через вторичную структуру
- Макс. число непоследовательных фрагментов внутри домена

На чем основаны методы

- Домен имеет собственное гидрофобное ядро (пример: алгоритм DETECTIVE Swindells, 1995)
- Домен – это часть белка, внутри которой много контактов аминокислотных остатков, а между доменами – мало контактов (пример: алгоритм ДОМАК, Siddiqui&Barton, 1995)

(Эти алгоритмы мы разберем только для исторической справки, сейчас их никто не применяет. Однако общие идеи в области остались от них)

Siddiqui&Barton, 1995: ДОМАК

Сверху – вниз, от целого – к части

- Предпосылки: домен состоит из одного или двух непрерывных участков полипептидной цепи
- Число контактов между остатками внутри домена больше, чем число междоменных контактов

Siddiqui&Barton, 1995: DOMAK

Формализация

- Два остатка контактируют, если расстояние между ними меньше 5Å
- Если белок разбит на две части, A и B, то определяется индекс разделенности:

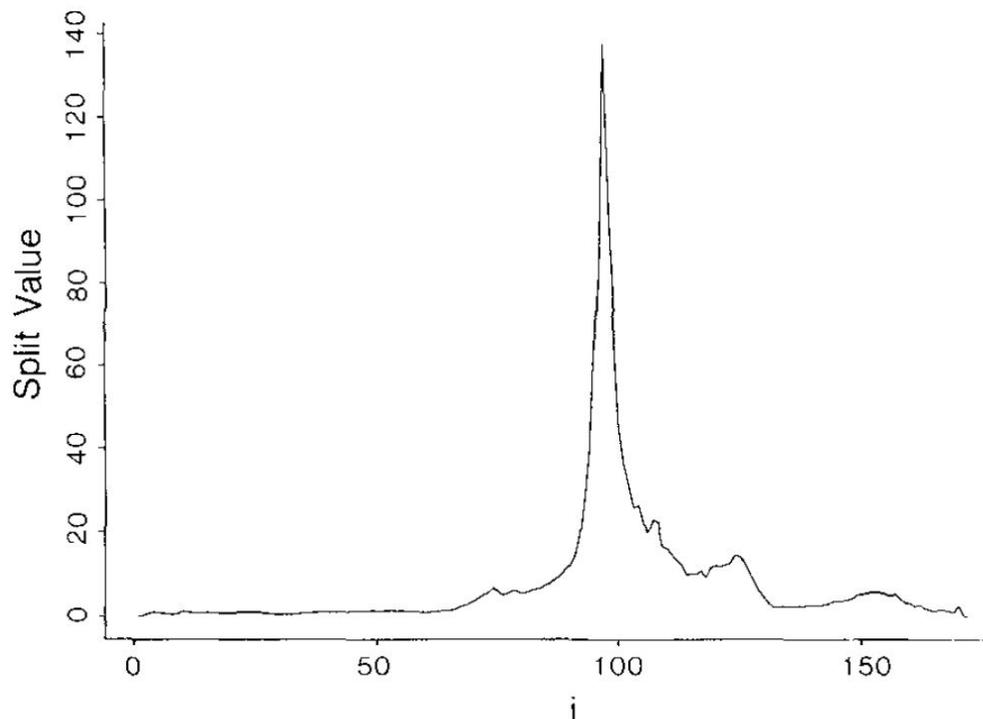
$$\mathit{SplitValue} = (\mathit{int}_A / \mathit{ext}_{AB}) \cdot (\mathit{int}_B / \mathit{ext}_{AB})$$

int_A – число пар контактирующих остатков из A

int_B – число пар контактирующих остатков из B

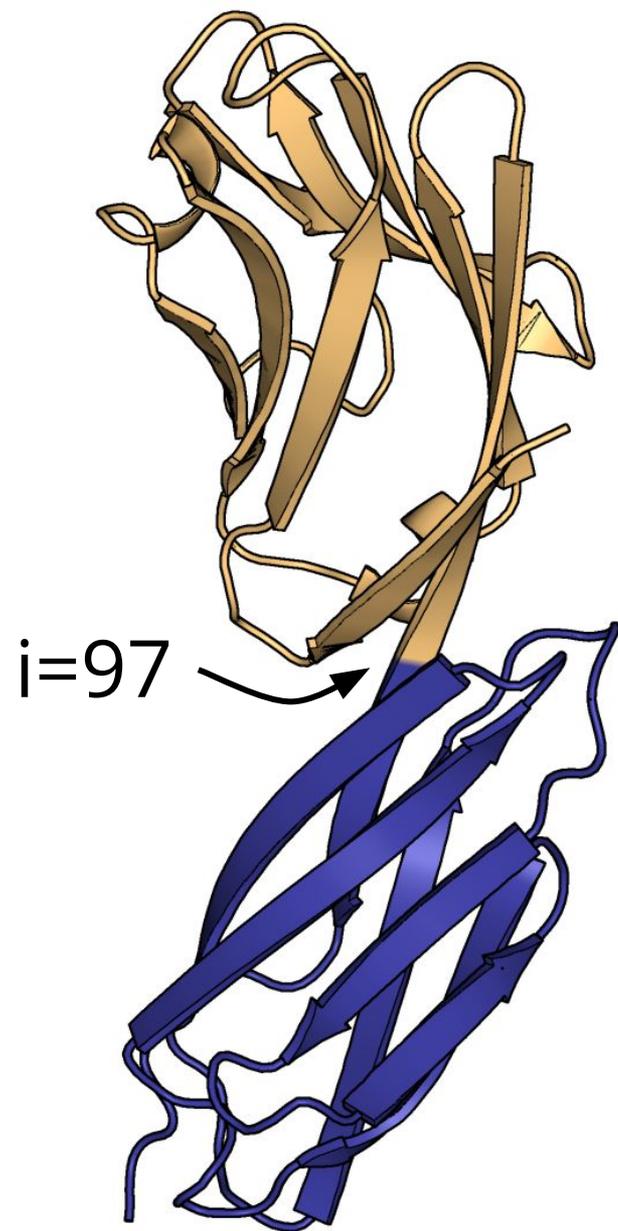
ext_{AB} – число пар контактирующих остатков, один из A, а другой – из B

Siddiqui&Barton, 1995: DOMAK

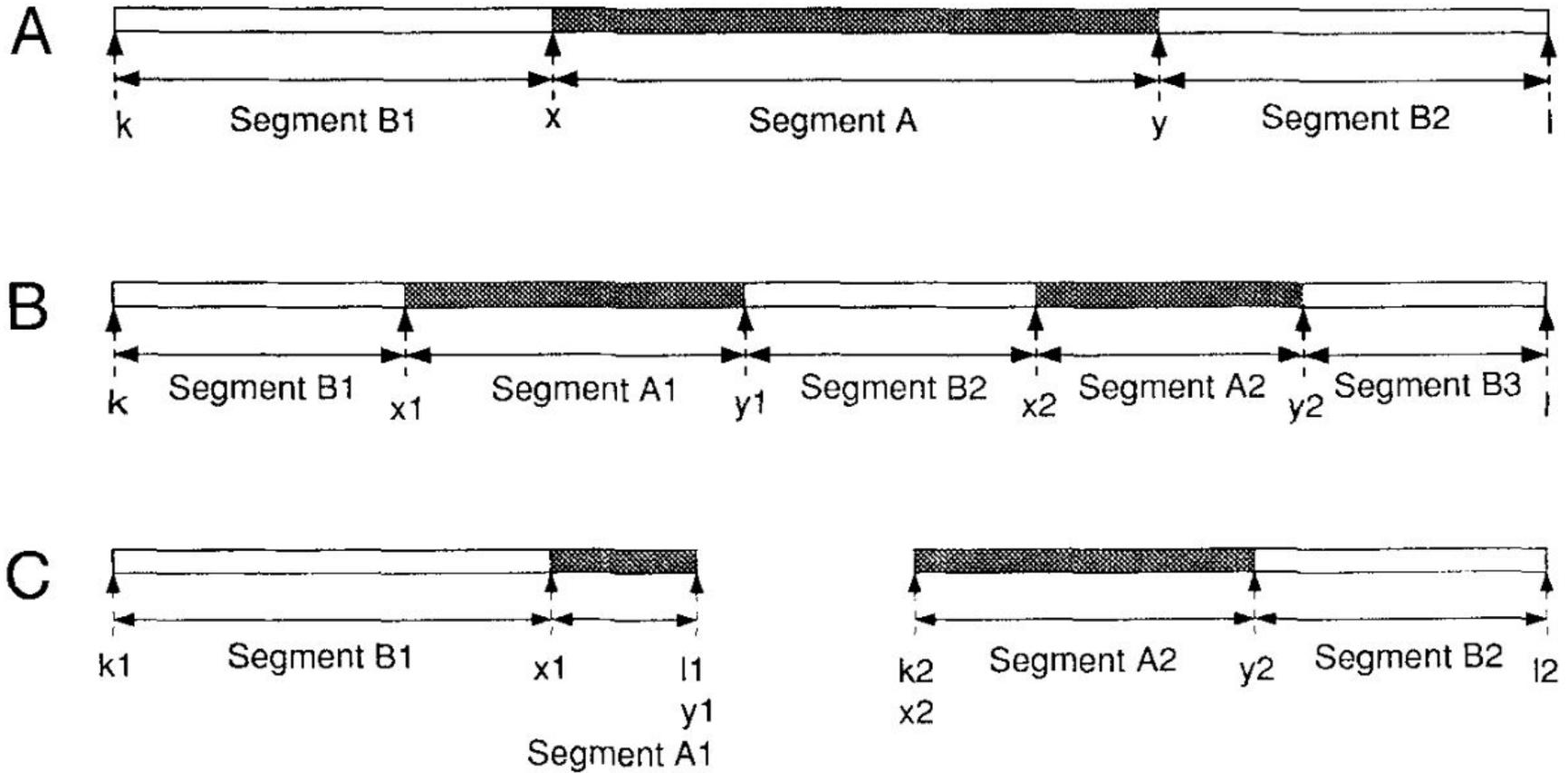


PDB ID 1CDH

Часть А: N-конец полипептидной цепи
до остатка i ;
часть В – от $(i+1)$ до C-конца



Siddiqui&Barton, 1995: DOMAK



Siddiqui&Barton, 1995: DOMAK

- К полной цепи применяются методы **A** и **B**. Выбирается разделение с лучшим индексом
- К полученным двум доменам применяется та же процедура. В случае, когда домен состоит из двух сегментов, применяется также метод **C**.
- Алгоритм останавливается в зависимости от пороговых значений:
MDS – минимальный размер домена (в числе остатков)
MSS – минимальный размер сегмента
- Отдельная процедура предусмотрена для сегментов, длина которых между **MDS** и **MSS**
- Найденные домены проверяются на “компактность”; некомпактные – сливаются в один

Домен объявляется некомпактным, если его радиус гирации сильно отличается от такового для белков с данным числом остатков

Swindells, 1995. DETECTIVE

Снизу – вверх, наращивание частей

Предпосылка: каждый домен имеет свое гидрофобное ядро.

Этапы:

1. выявление гидрофобных ядер в структуре
2. «натягивание» доменов на гидрофобные ядра

Гидрофобные ядра – еще не домены. Для получения доменов применяется многоходовая процедура чистки-слияния

Методы выделения доменов (из обзора Veretnik & Shindyalov, 2005)

Method	Year Generation	Strategy	Type of domains generated	Approaches/models used
Rossmann and Liljas (Rossmann and Liljas, 1974)	1974 first generation	top down	contiguous	Distance plots of the structure against itself using C- α distances, search for strong interactions close to diagonal
Crippen (Crippen, 1978)	1978 first generation	bottom-up	contiguous and non-contiguous	. Clustering of small structural units
Rose (Rose, 1979)	1979 first generation	top down	contiguous	Cutting the projection of 3D structure onto 2D domain disclosing plain
Wodak and Janin (Wodak and Janin, 1981)	1981 first generation	top down	contiguous	Finding minimum in the interface between two domains
PUU (Holm and Sander, 1994)	1994 second generation	top down	contiguous and non-contiguous	Rendering of the contact matrix, constructed using rigid body oscillation
DETECTIVE (Swindells, 1995a)	1995 second generation	bottom-up	Contiguous and non-contiguous	Building of the hydrophobic core
Islam et al. (Islam et al., 1995)	1995 second generation	top down	contiguous and non-contiguous	Finding minima in the inter-domain contact density
DOMAK (Siddiqui and Barton, 1995)	1995	top down	contiguous and	Splitting structure by maximizing

Большинство методов основано на принципах, близких к DOMAK

	second generation		non-contiguous	intra-domain/inter-domain contacts
Sowdhamini and Blundell (Sowdhamini and Blundell, 1995)	1995 second generation	bottom-up	contiguous and non-contiguous	Clustering of secondary structures
Taylor (Taylor, 1999)	1999 second generation	bottom-up	contiguous and non-contiguous	Clustering of residues in spatial proximity using Ising model
STRUDEL (Wernisch et al., 1999)	1999 second generation	top down	contiguous and non-contiguous	Finding minimum inter-domain contacts using Kernighan-Lin graph heuristics
DomainParser (Xu et al., 2000)	2000 second generation	top down	contiguous and non-contiguous	Finds minimum inter-domain contacts using graph theoretical approach with maximum flow/minimum cut using Ford-Fulkerson algorithm
(Xuan et al., 2000)	2000 Second generation	bottom-up	contiguous and non-contiguous	Assemble domains from rudimentary fragments using fuzzy clustering
PDP (Alexandrov and Shindyalov, 2003)	2003 second generation	top down	contiguous and non-contiguous	Finding partitioning with minimal number of contacts between domains
HVdWD (Hierarchy of Van der Waals Domains) (Berezovsky, 2003)	2003 second generation	bottom-up	contiguous and non-contiguous	Clustering of short segments. Both initial segments and the clustering threshold are based primarily on Van der Waals interactions among atoms.
(Kundu et al., 2004)	2004 Second generation	Top down	contiguous and non-contiguous	Decomposition of the structure using Gaussian Network Model; assumes semi-independent motion of domains.

Table 1. Summary of domain decomposition methods.



Partitioning Protein Structures into Domains: Why Is it so Difficult?

Timothy A. Holland¹, Stella Veretnik²  , Ilya N. Shindyalov², Philip E. Bourne^{2,3}

 [Show more](#)

<https://doi.org/10.1016/j.jmb.2006.05.060>

[Get rights and content](#)

**315 вручную
присвоенных
разбиения для
бенчмарков**

Сейчас применяются:

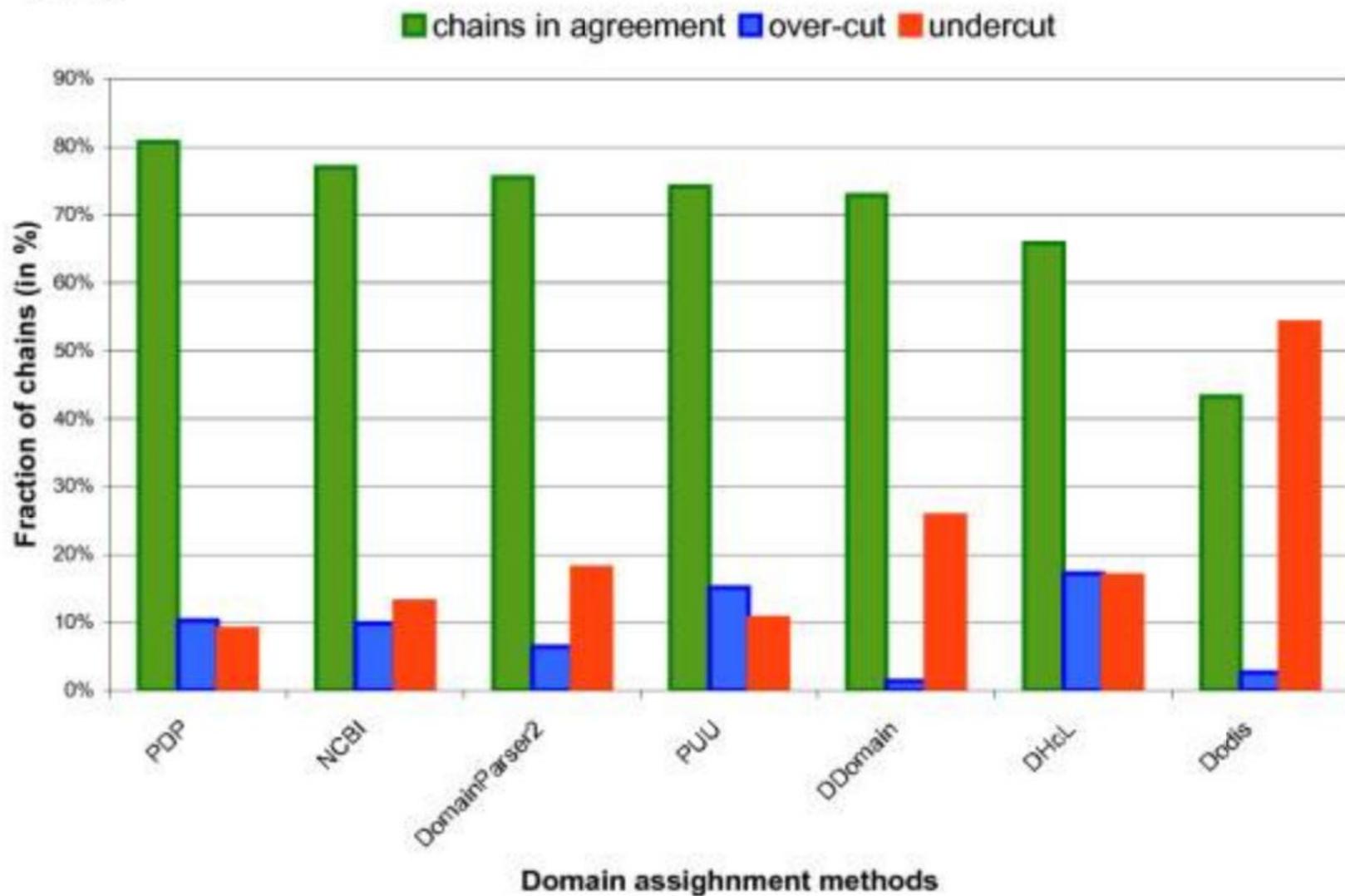
- Protein Domain Parser (PDP)
- DomainParser2
- PUU
- DDomain
- DHcL
- Dodis

dConsensus: там, где есть много
мнений, можно сделать голосование

<http://pdomains.sdsc.edu/dConsensus>

Alden et al., BMC Bioinformatics, 2010

A.



An ambiguity principle for assigning protein structural domains

Guillaume Postic^{1,2,3,4,*}, Yassine Ghouzam^{1,2,3,4}, Romain Chebrek^{1,2,3,4} and Jean-Christophe Gelly^{1,2,3,4,*}

+ See all authors and affiliations

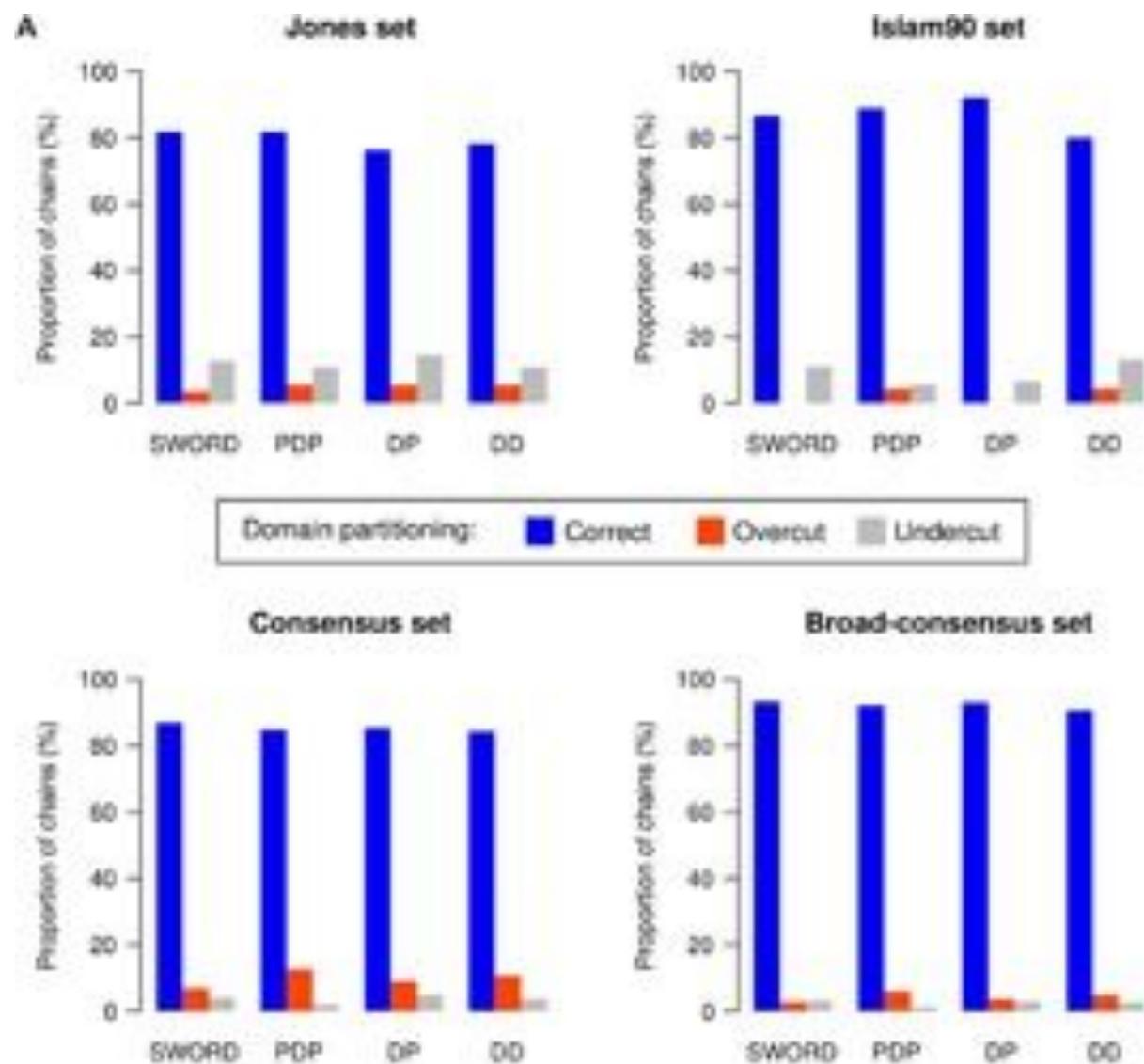
Science Advances 13 Jan 2017:
Vol. 3, no. 1, e1600552
DOI: 10.1126/sciadv.1600552

‘Protein Peeling’: an approach for splitting a 3D protein structure into compact fragments

Jean-Christophe Gelly, Alexandre G. de Brevern , Serge Hazout [Author Notes](#)

Bioinformatics, Volume 22, Issue 2, 15 January 2006, Pages 129–133, <https://doi.org/10.1093/bioinformatics/bti773>

Published: 14 November 2005 **Article history** ▼



Классификации структурных доменов

SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>) (остановлен 2009)

- ручная детекция доменов
- 4 основных уровня классификации (класс, укладка, суперсемейство, семейство)
- SCOP2 - более автоматизированная версия SCOP
- SCOPe - extended
(работа продолжена другой группой, актуально на 2019)

CATH (<http://www.cathdb.info/>)

- полуавтоматическая детекция доменов
- 4 основных уровня классификации (класс, архитектура, топология, суперсемейство)