

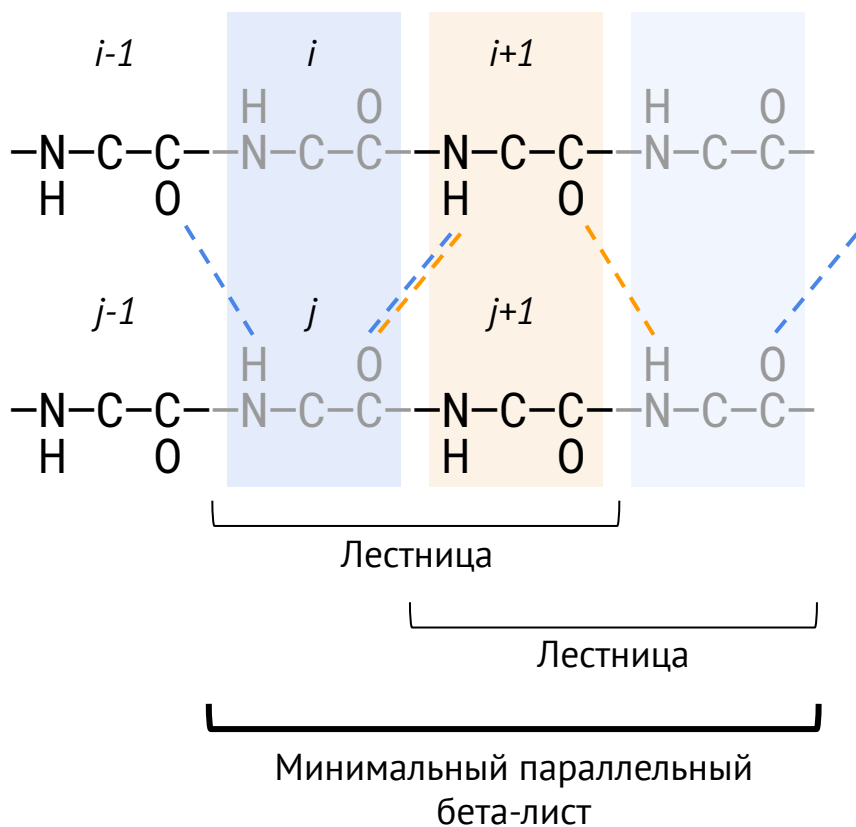
Факультет биоинженерии и биоинформатики,  
Московский государственный университет имени М. В. Ломоносова



# Структурные домены. Пространство укладки

Лекция 5, биоинформатика, 4 курс ФББ МГУ, осенний семестр  
Злобин А. С., [alexander.zlobin@fbb.msu.ru](mailto:alexander.zlobin@fbb.msu.ru)

# Мостики и лестницы

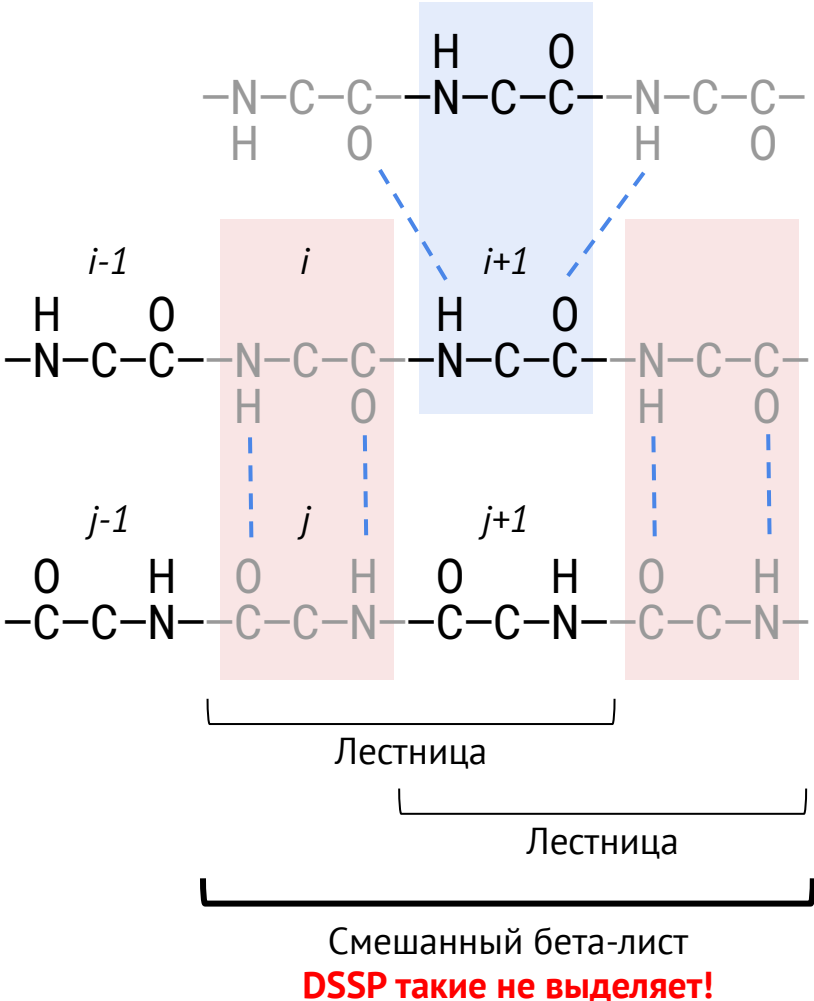
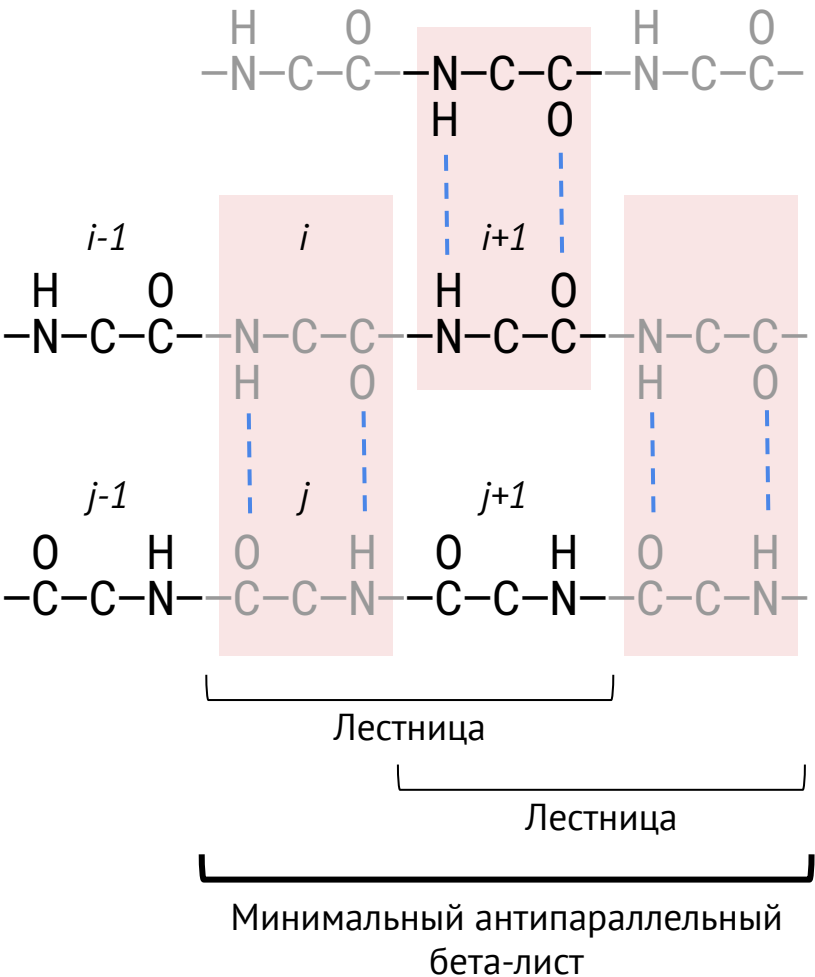


Лестница = один или более мостиков **одного типа**, идущих подряд (по одному тяжу не более 1 остатка между лестницами, по другому не более 4)

Две лестницы связаны, если имеют общий остаток

Бета-лист: максимальный связный набор лестниц

# Мостики и лестницы



# Укладки

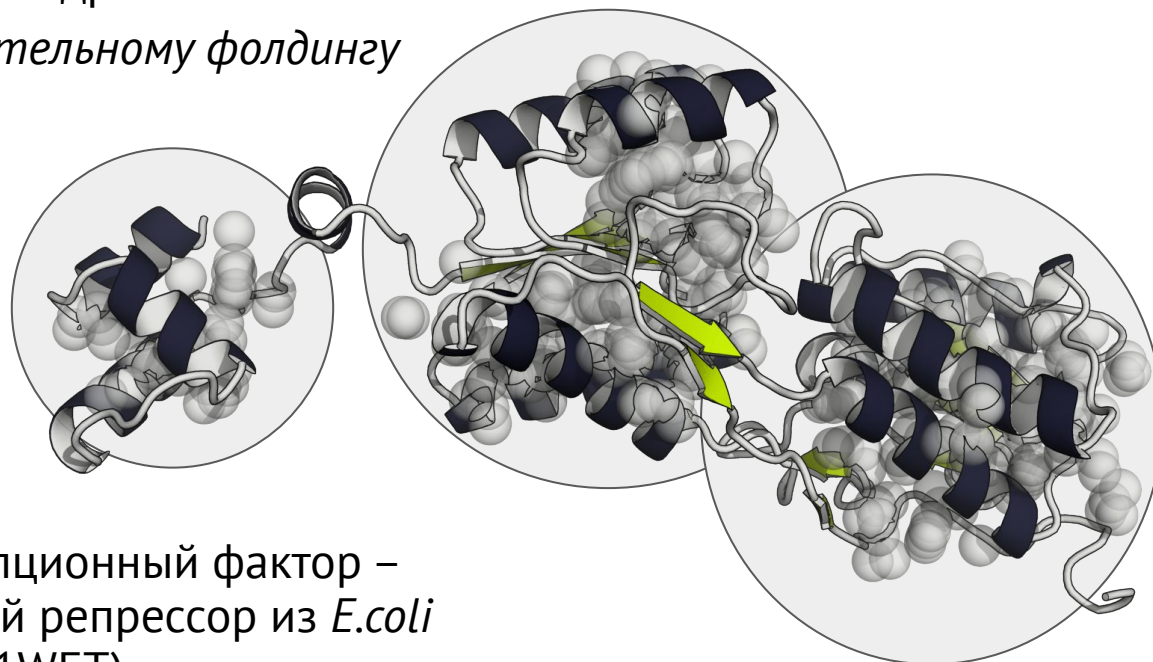
- Функция определяется на уровне структуры
- Структуры консервативнее последовательностей
- Хотим узнать функцию нашего белка и знаем его структуру – сначала стоит посмотреть на уже аннотированные похожие структуры
- Подспорье в изучении структурно-функциональных взаимоотношений – базы данных расклассифицированных укладок **структурных доменов**

Но что такое **структурный домен**?

# Структурный домен (биоинформатика 3D структур)

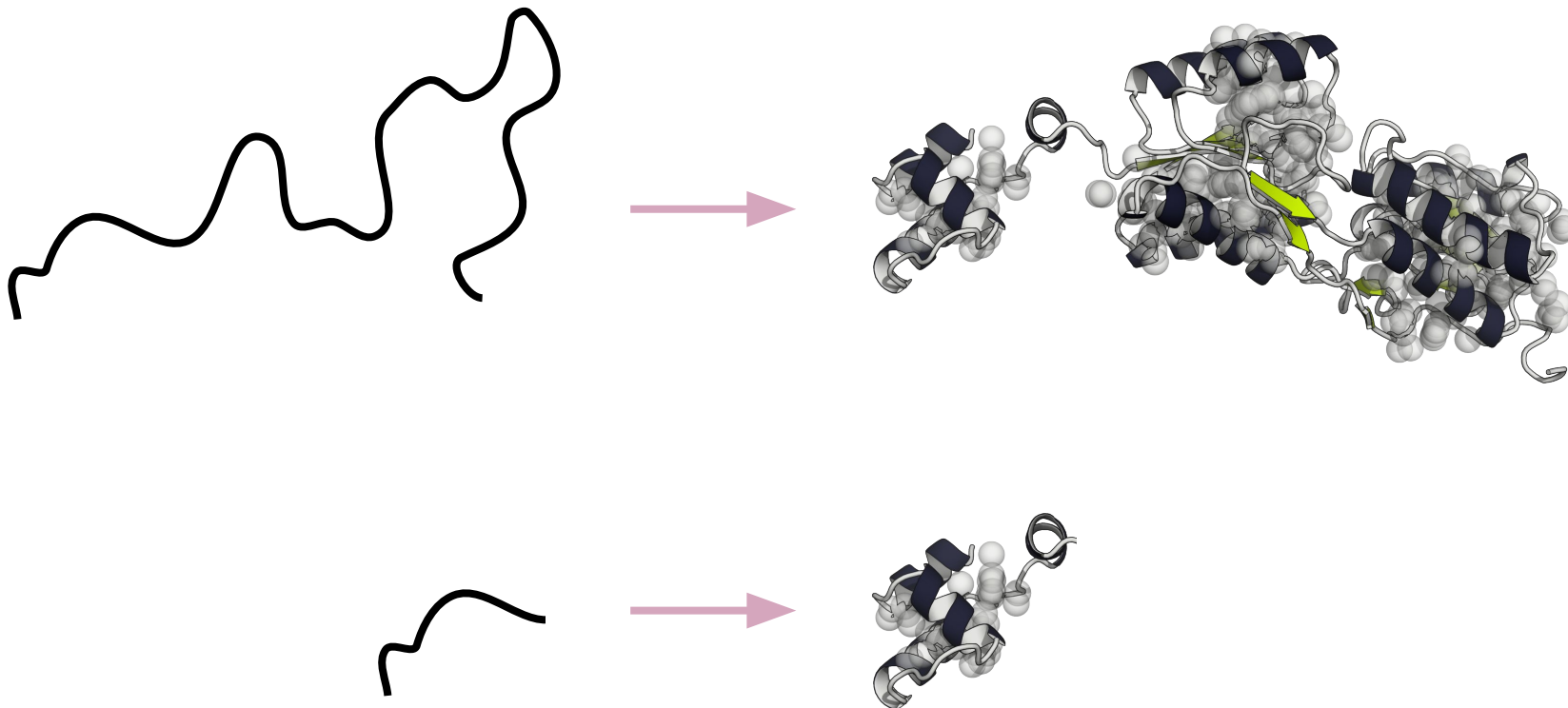
Обособленная в пространстве часть белка,  
его структурная единица, имеющая:

- сравнительно мало контактов с другими частями белка
- собственное гидрофобное ядро
- *возможность к самостоятельному фолдингу*



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

# Возможность к самостоятельному фолдингу?



# 3D-координаты



## Подход сверху-вниз

Подразделение цепи на  
меньшие элементы

## Подход снизу-вверх

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

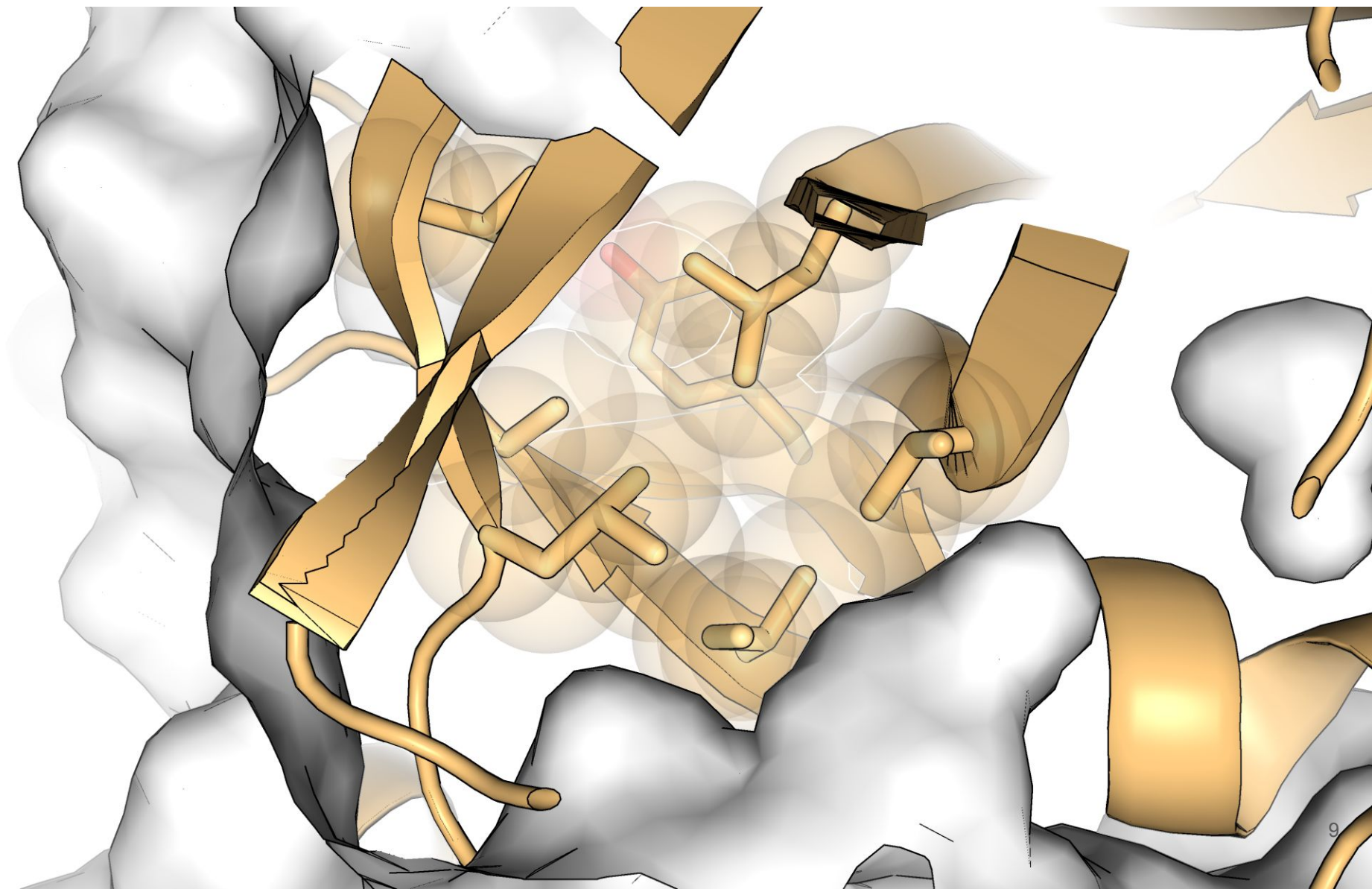
## На чем основаны методы

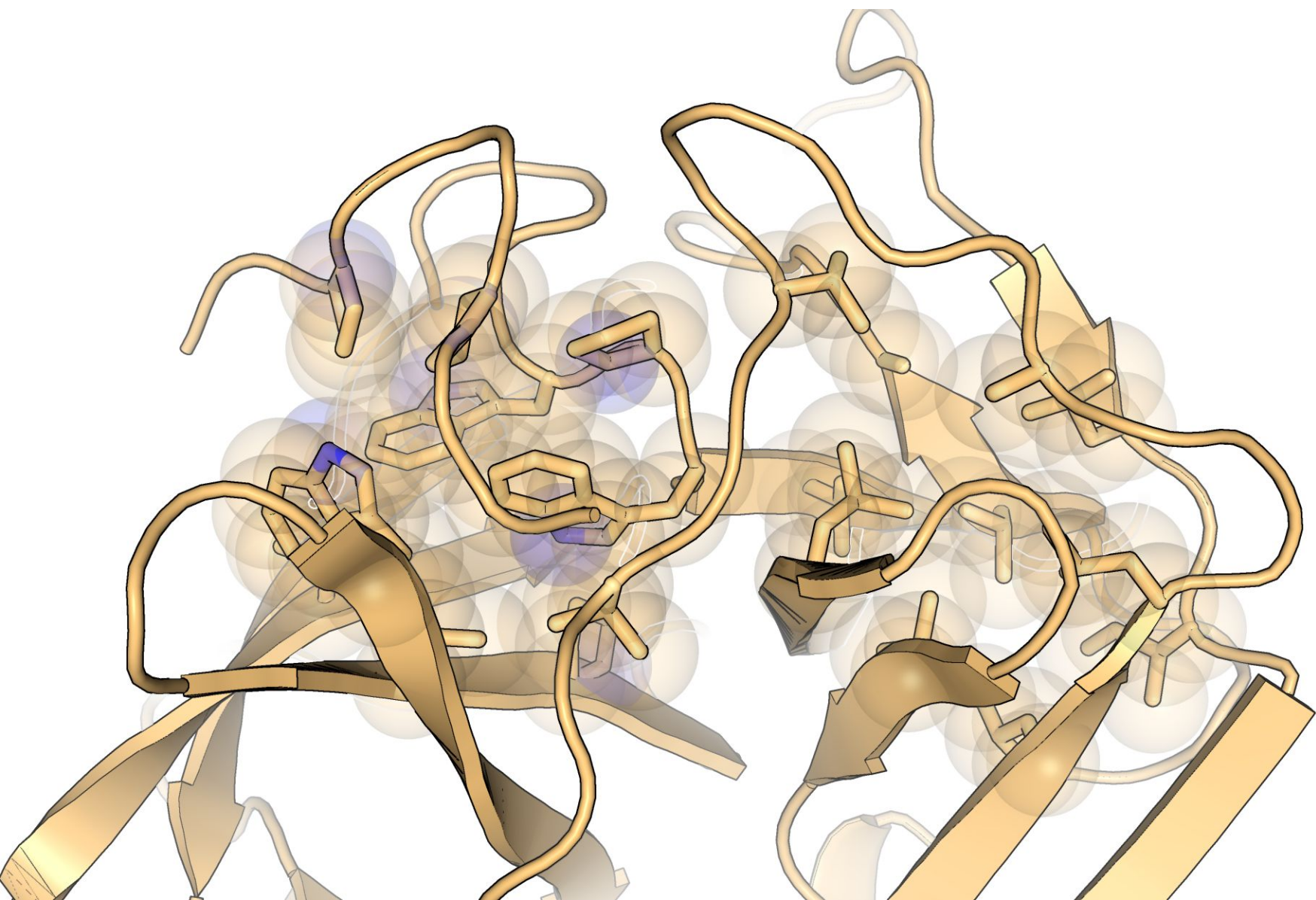
- Домен имеет собственное гидрофобное ядро (пример: алгоритм DETECTIVE Swindells, 1995)
- Домен – это часть белка, внутри которой много контактов аминокислотных остатков, а между доменами – мало контактов (пример: алгоритм ДОМАК, Siddiqui & Barton, 1995)

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

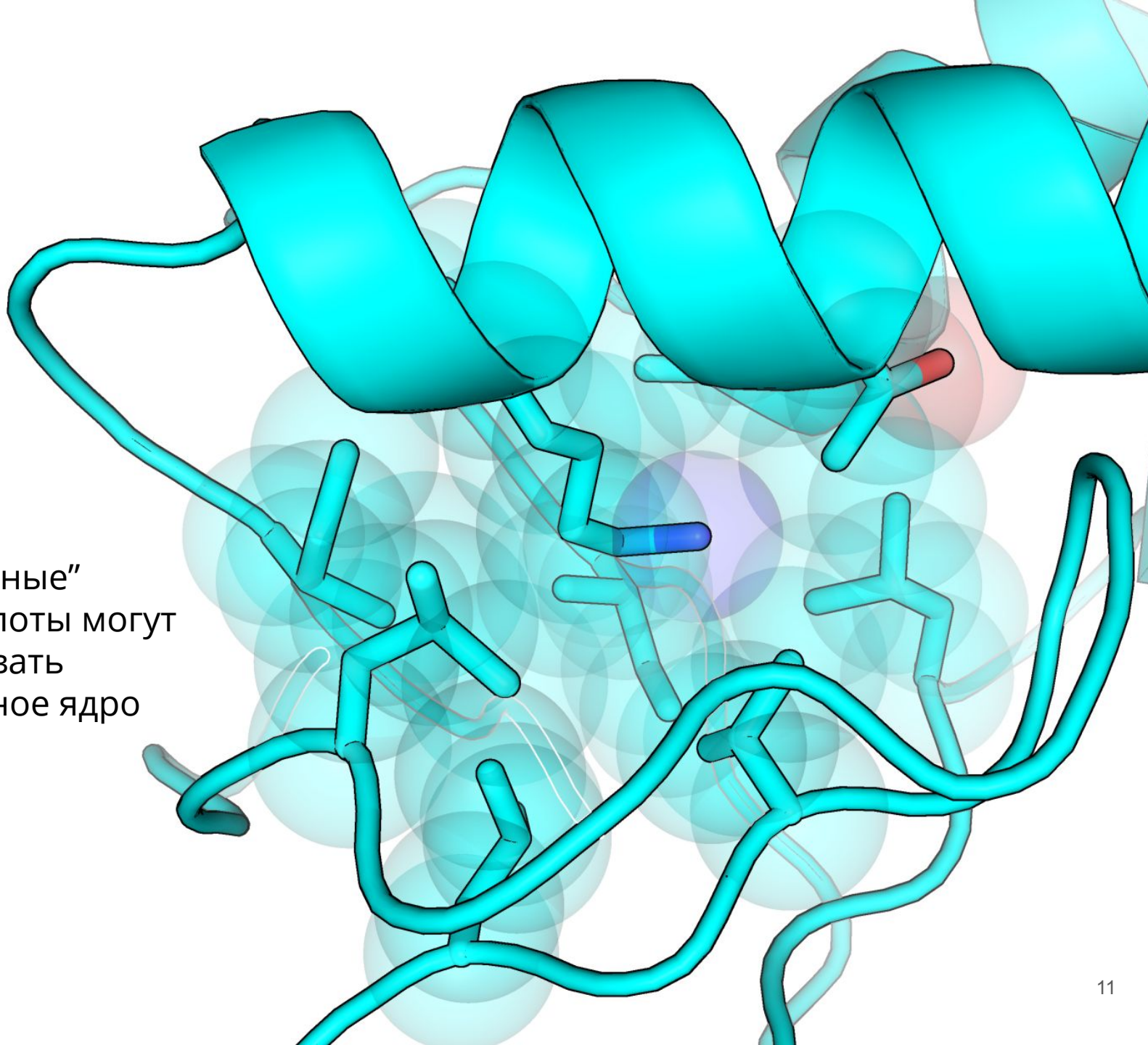


# Гидрофобные ядра





Не только  
“гидрофобные”  
аминокислоты могут  
образовывать  
гидрофобное ядро



# 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

# 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<sup>1,2,\*</sup>, Lalima G. Ahuja<sup>3</sup>, Fa-An Chao<sup>2,†</sup>, Youlin Xia<sup>2</sup>, Christopher L. McClendon<sup>3,4,‡</sup>, Alexandr P. Kornev<sup>3</sup>, 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?

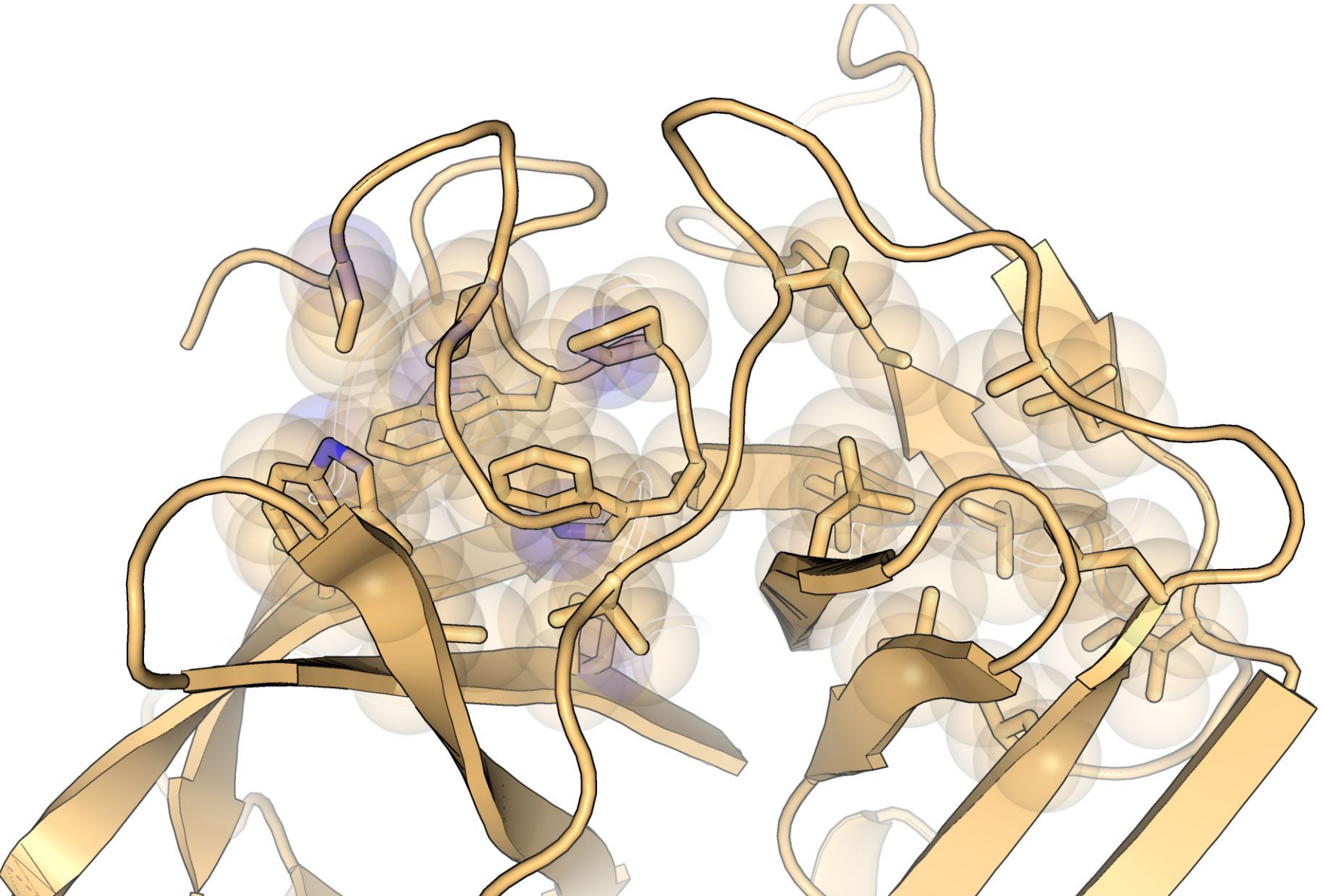
- Kannan & Vishveshwara, 1999
- Tsai & Nussinov, 1997
- Swindells, 1995
- Zehfus, 1995
- Heringa & Argos, 1991
- Plochocka et al., 1988
- Alexeevski et al., 2003

# 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 спиралям или тягам



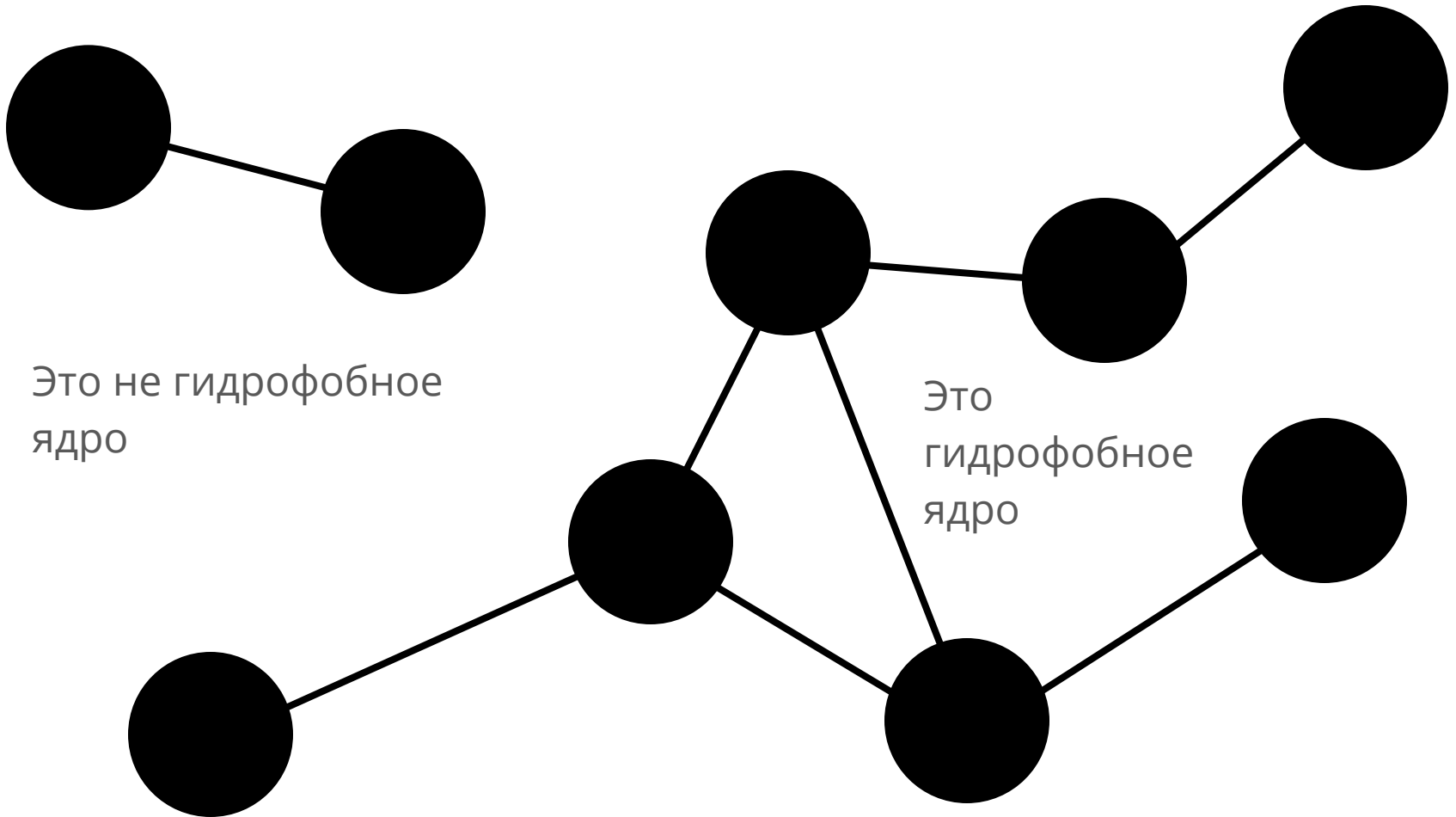
# Swindells

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

Строится граф:

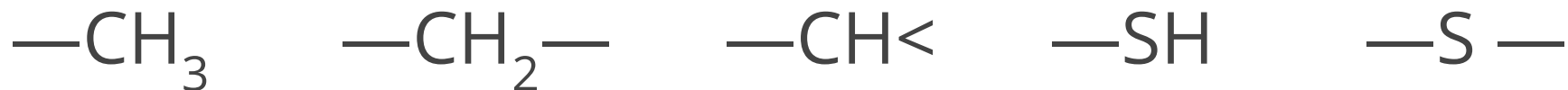
- Вершина – отобранный остаток
- Ребро соединяет вершины, если соответствующие остатки гидрофобно взаимодействуют
- Связные компоненты графа, содержащие 5 или более остатков, называются гидрофобными ядрами

# Swindells



# Alexeevski et al.: кластеры неполярных групп

- Элементарной единицей служат неполярные атомные группы (CH<sub>3</sub> 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:

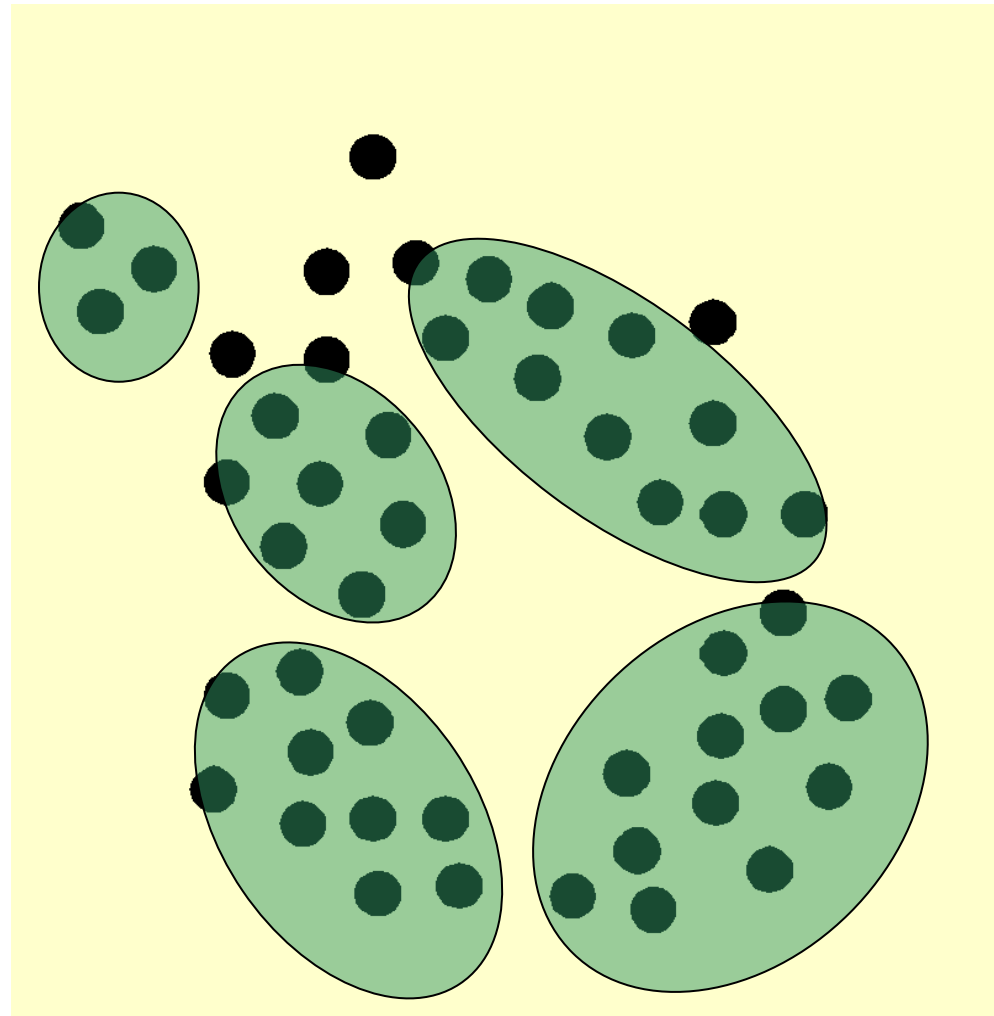


(при этом эти группы не должны быть связаны ковалентно с O или N)

# Alexeevski et al.

“Гидрофобный кластер”:

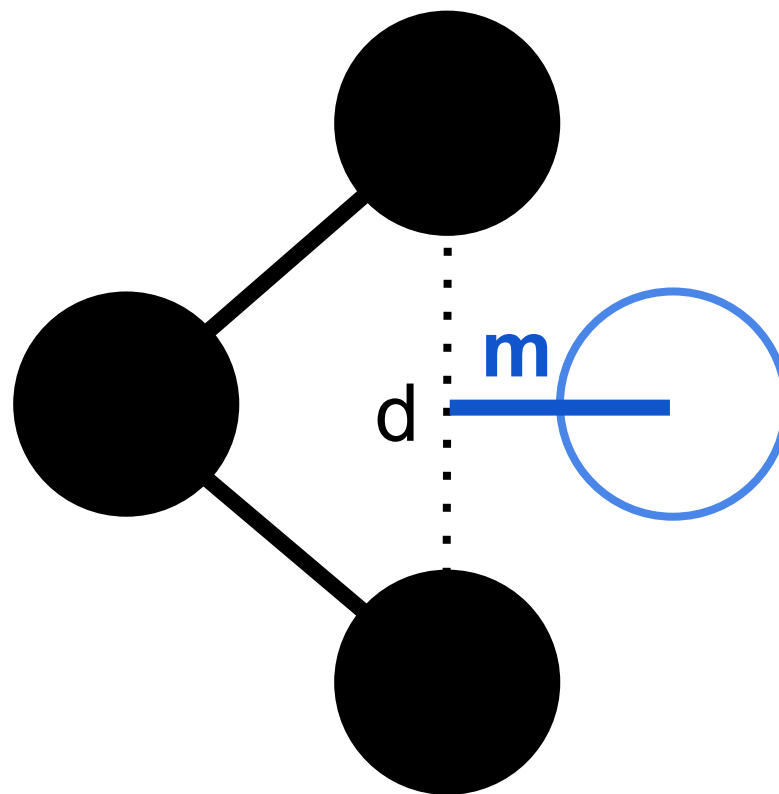
- состоит из неполярных групп
- каждая неполярная группа имеет несколько контактов с другими неполярными группами кластера
- взаимодействие между разными кластерами слабое



# 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)

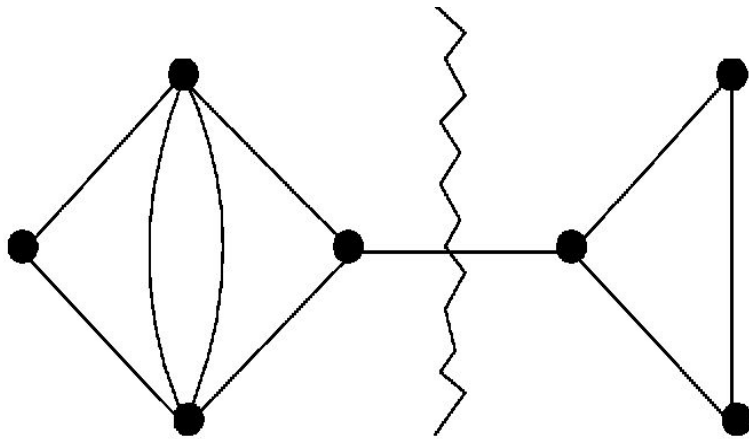
$m \geq m_0$  ( $m_0 = d_0/2$ )



# 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

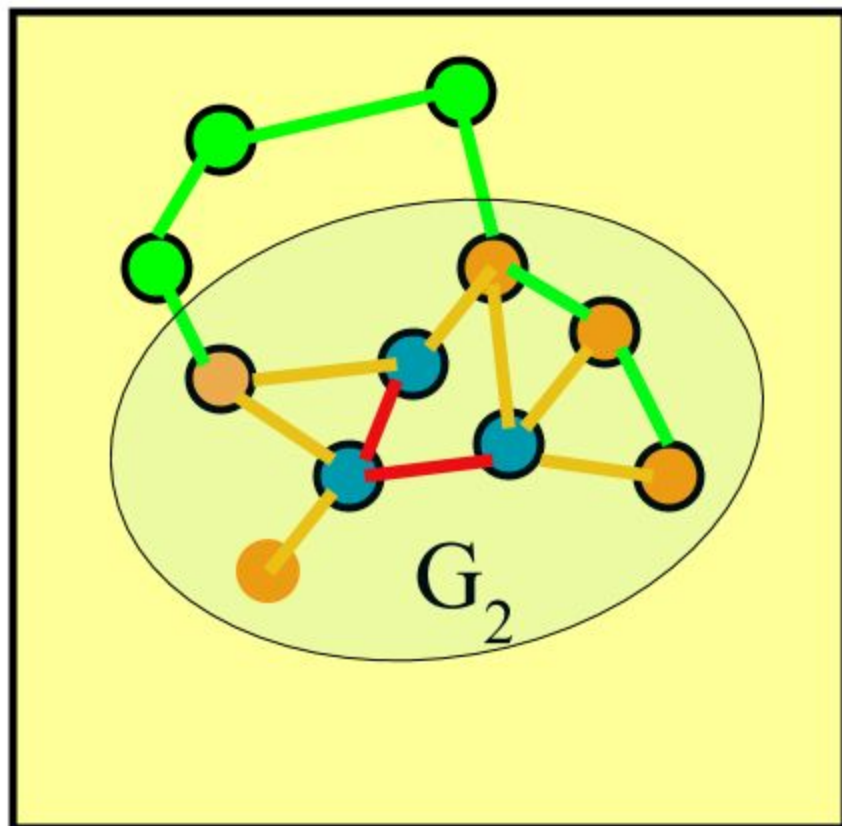
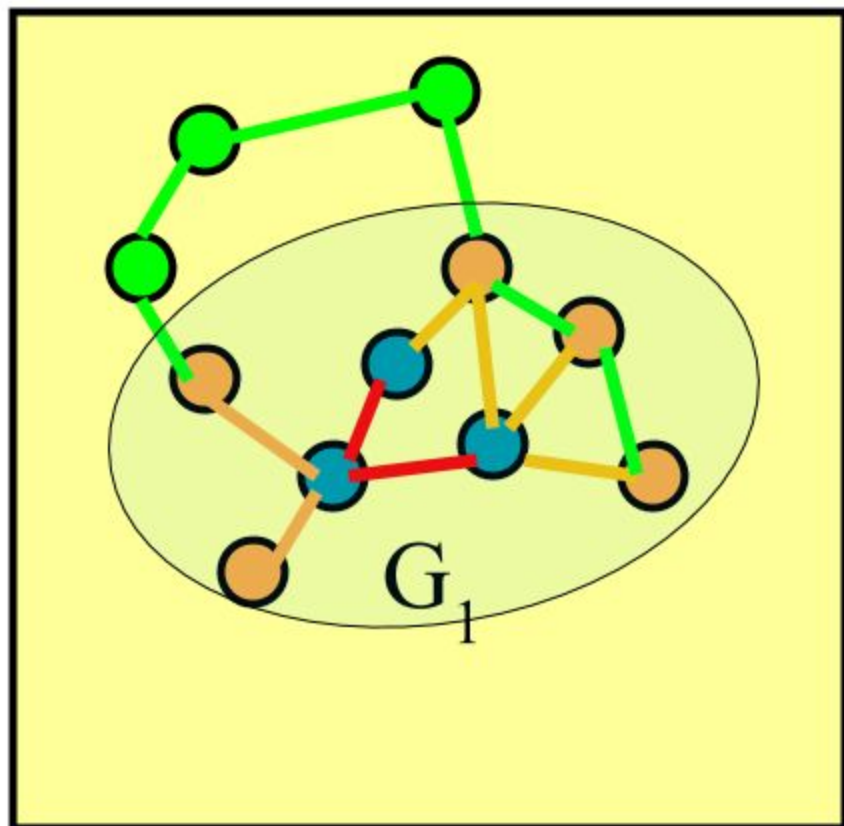


1-разрез графа

$(k,l)$ -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

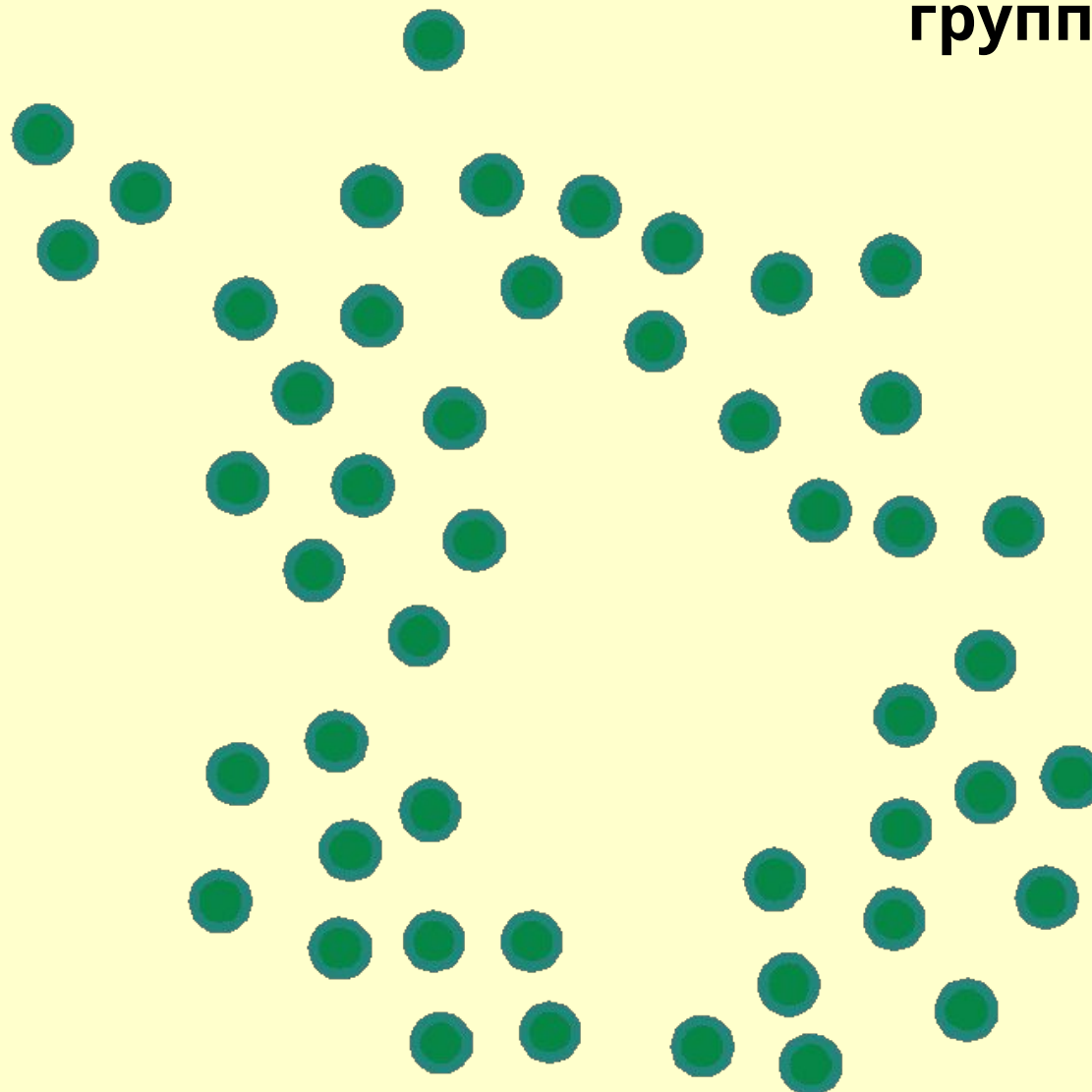
Подграф  $G_1$  (красные  
ребра) является  $(2,1)$ -  
разрезом

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

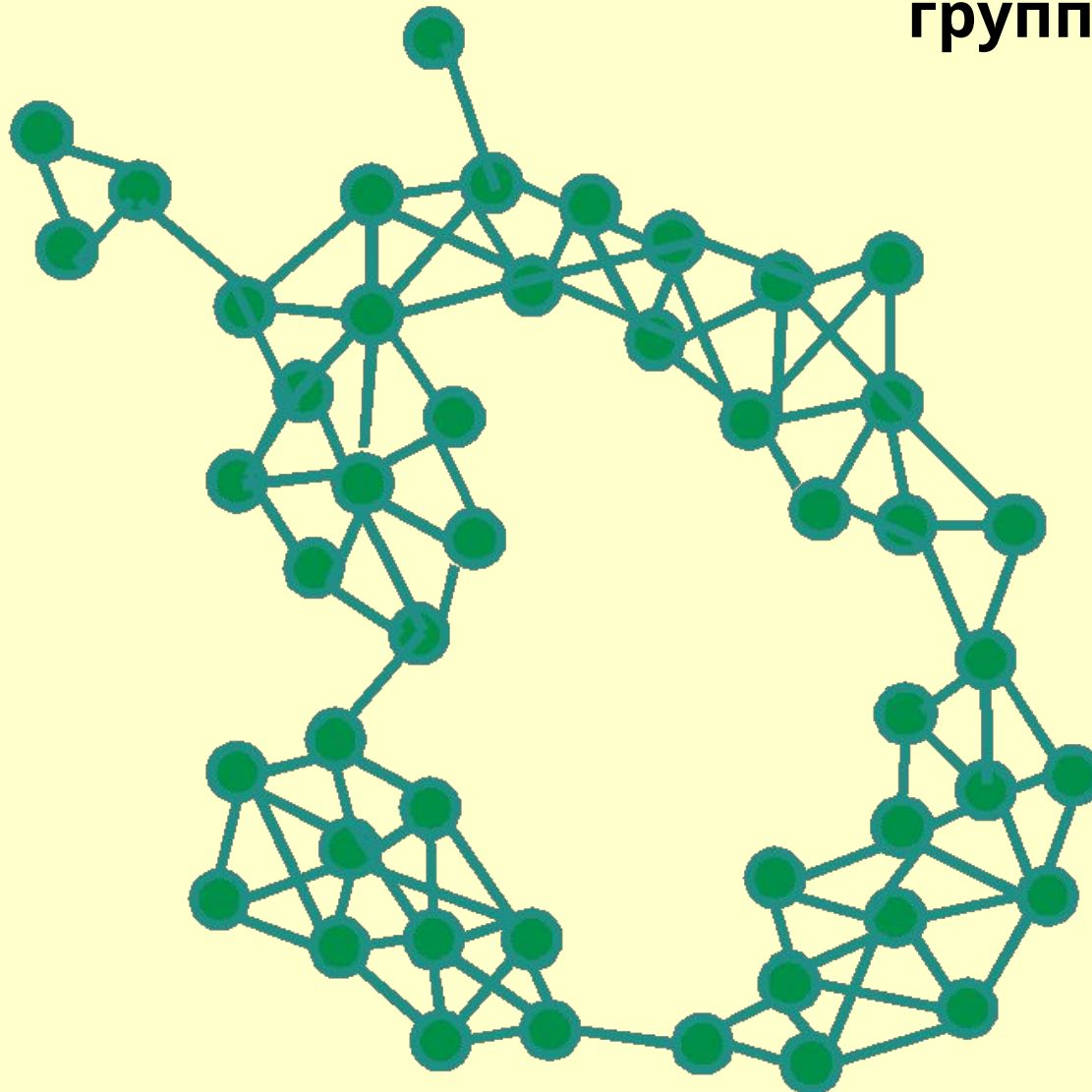


1-окрестности графов  $G_1$  и  $G_2$  обведены овалами

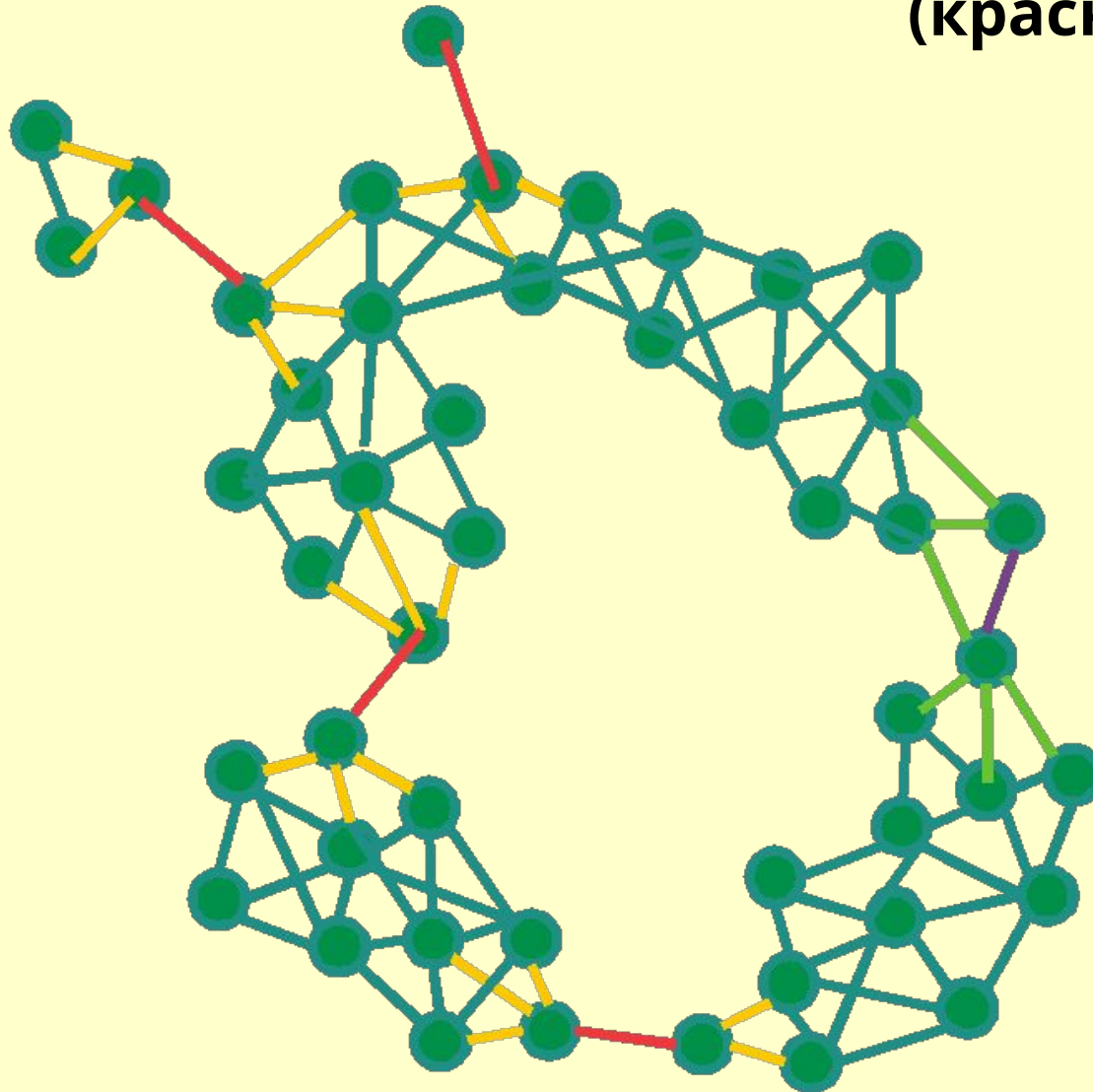
# Неполярные группы



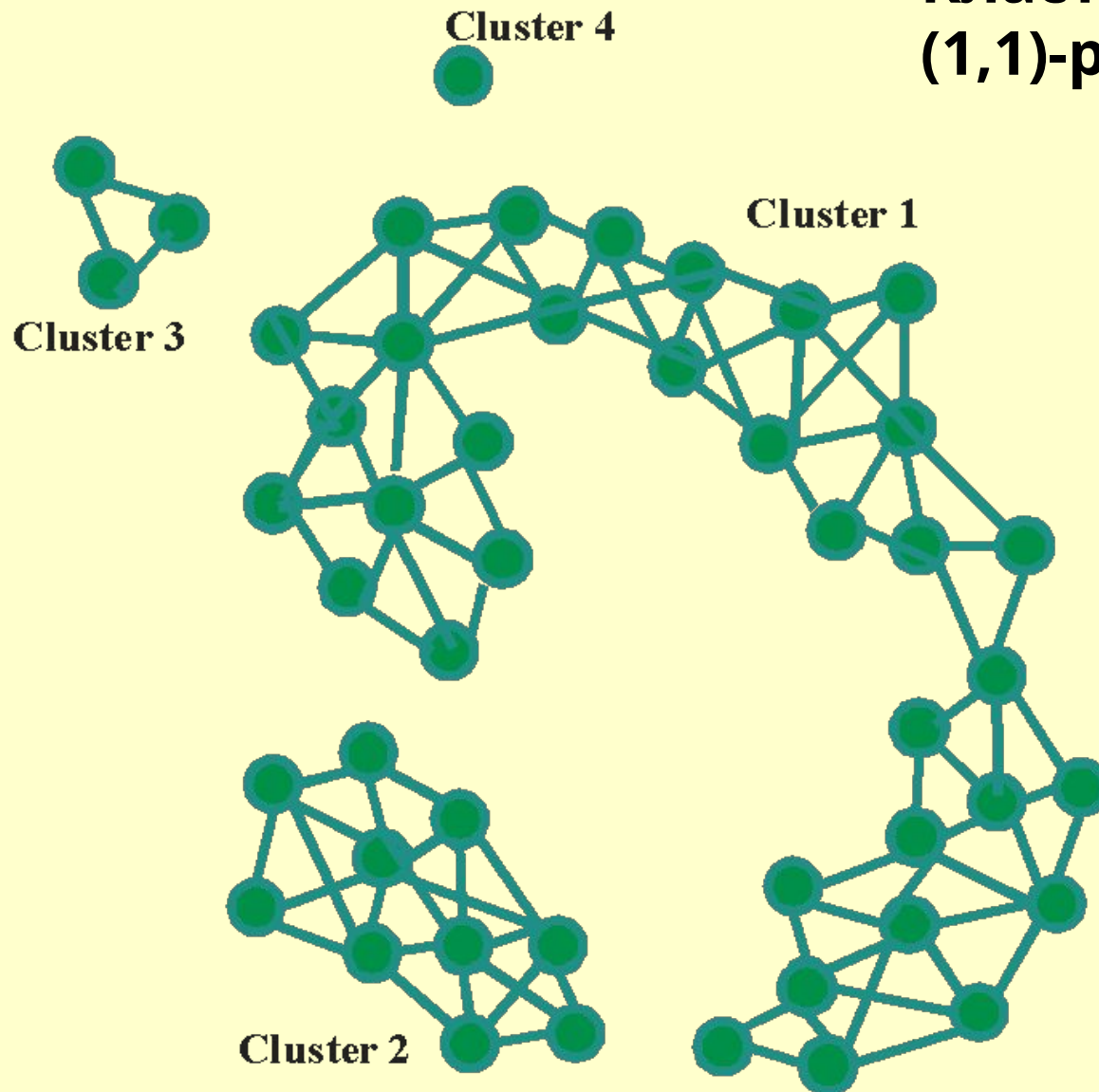
# Граф неполярных групп



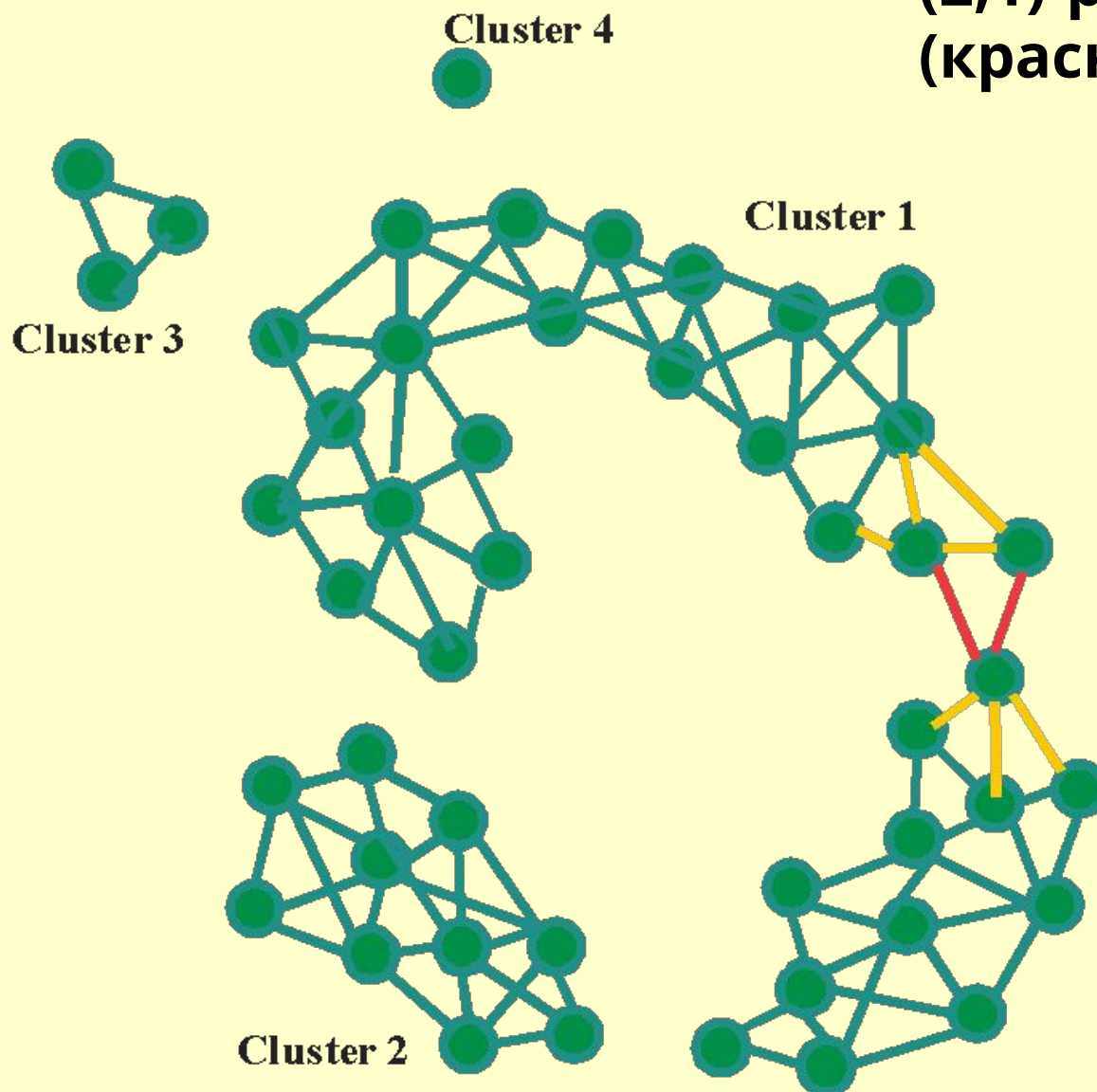
**(1,1)-разрезы  
(красные ребра)**



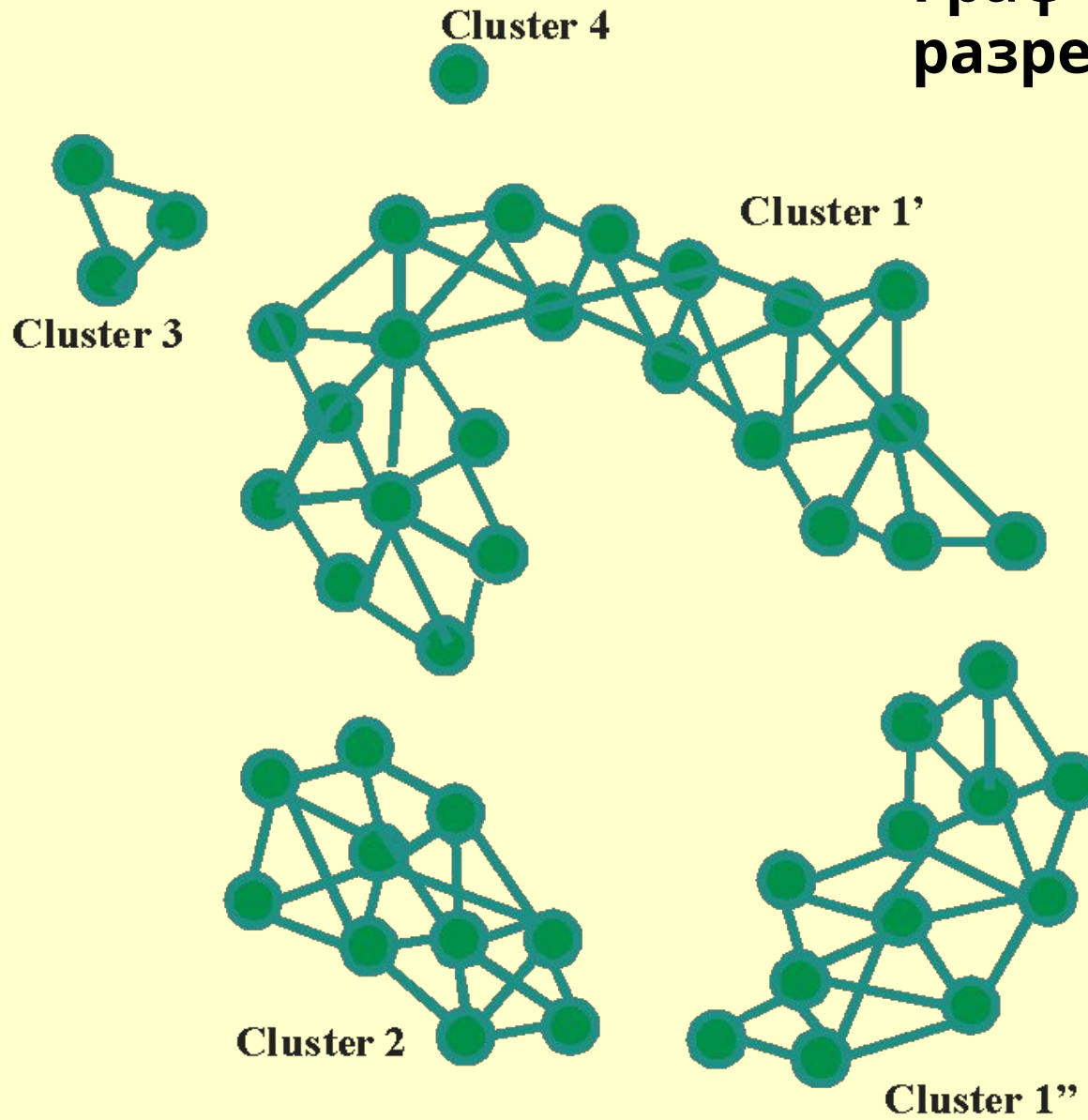
# Кластеры после (1,1)-разрезов



# (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>

# Возвращаемся к доменам

# Swindells, 1995. DETECTIVE

Снизу – вверх, наращивание частей

Предпосылка: каждый домен имеет свое гидрофобное ядро.

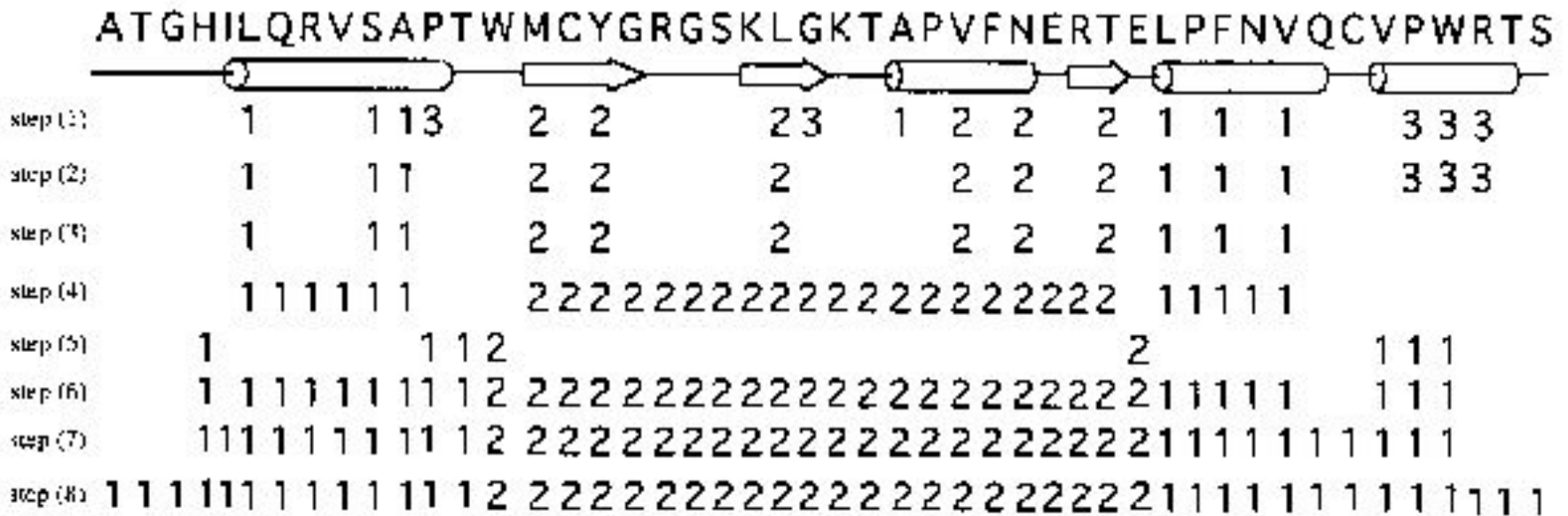
Этапы:

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

Алгоритм демонстрируется на примере (см. рис.)

- (1) найдено 3 кластера – 1-й, 2-й и 3-й
- (2) остатки, окруженные “чужими” вычищаются
- (3) кластеры, включающие меньше пяти остатков, вычищаются
- (4) заливка некластеризованных остатков
- (5–6) оставшиеся некластеризованные остатки присоединяются по контактам к кластерам предыдущего шага
- (7–8) опять прочистка, заливка и присоединение хвостов



# 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:

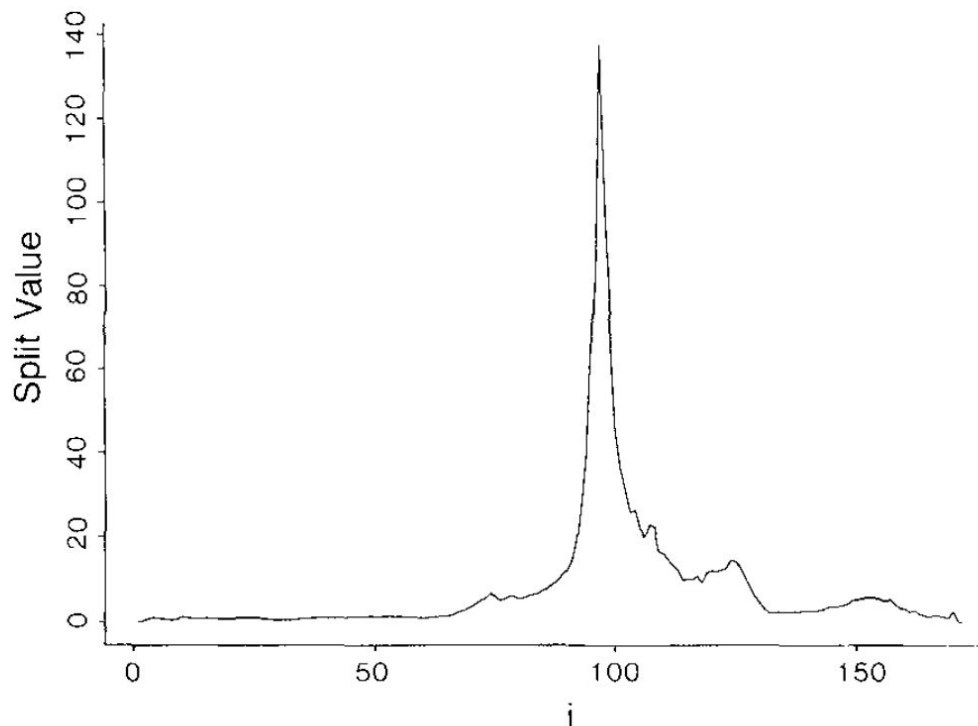
$$\mathit{SplitValue} = (\mathit{int}_A / \mathit{ext}_{AB}) \cdot (\mathit{int}_B / \mathit{ext}_{AB})$$

$\mathit{int}_A$  – число пар контактирующих остатков из A

$\mathit{int}_B$  – число пар контактирующих остатков из B

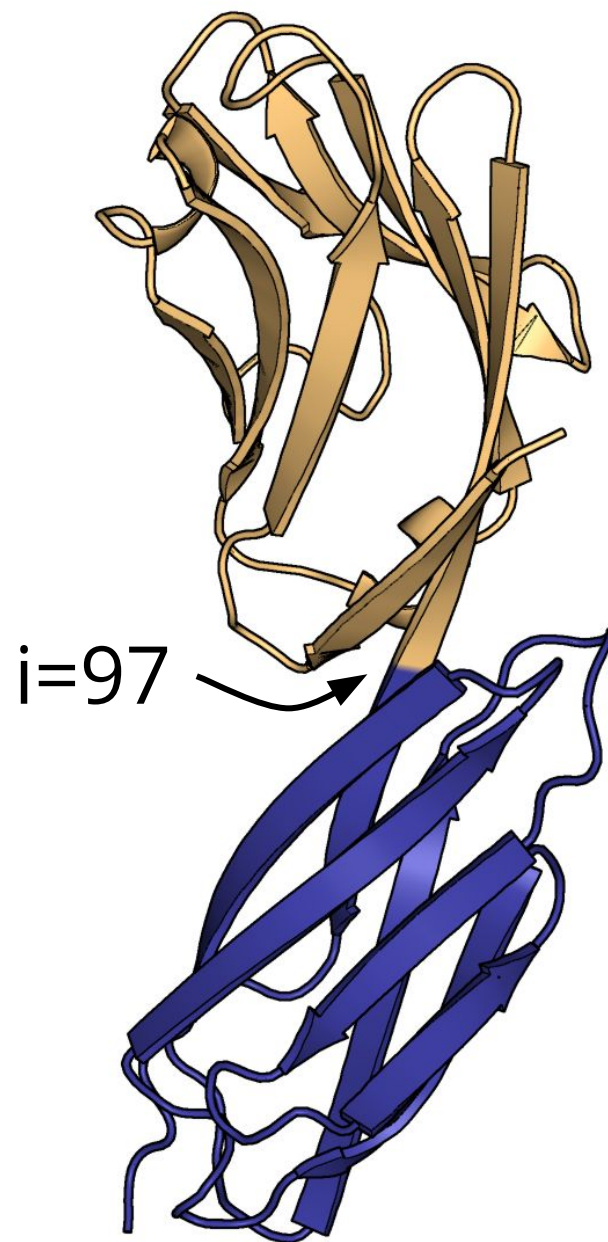
$\mathit{ext}_{AB}$  – число пар контактирующих остатков, один из A, а другой – из B

# Siddiqui & Barton, 1995: DOMAK

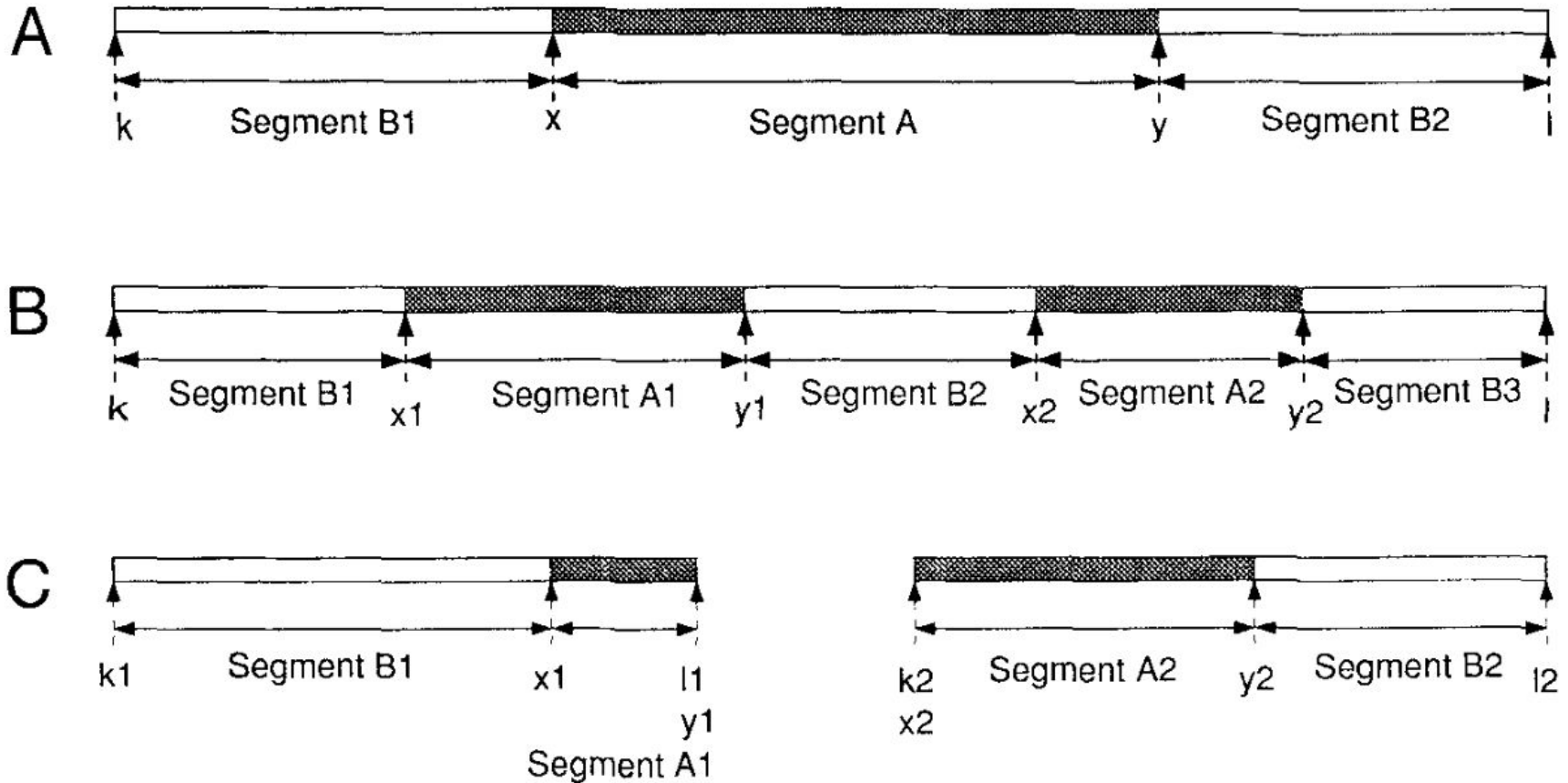


## PDB ID 1CDH

Часть A: N-конец полипептидной цепи  
до остатка  $i$ ;  
часть B – 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

# 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- $\alpha$ 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
STRU DL (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. Holland<sup>1</sup>, Stella Veretnik<sup>2</sup>  , Ilya N. Shindyalov<sup>2</sup>, Philip E. Bourne<sup>2,3</sup>

 [Show more](#)

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

[Get rights and content](#)

**315 вручную  
присвоенных  
разбиения для  
бенчмарков**

Сейчас применяются:

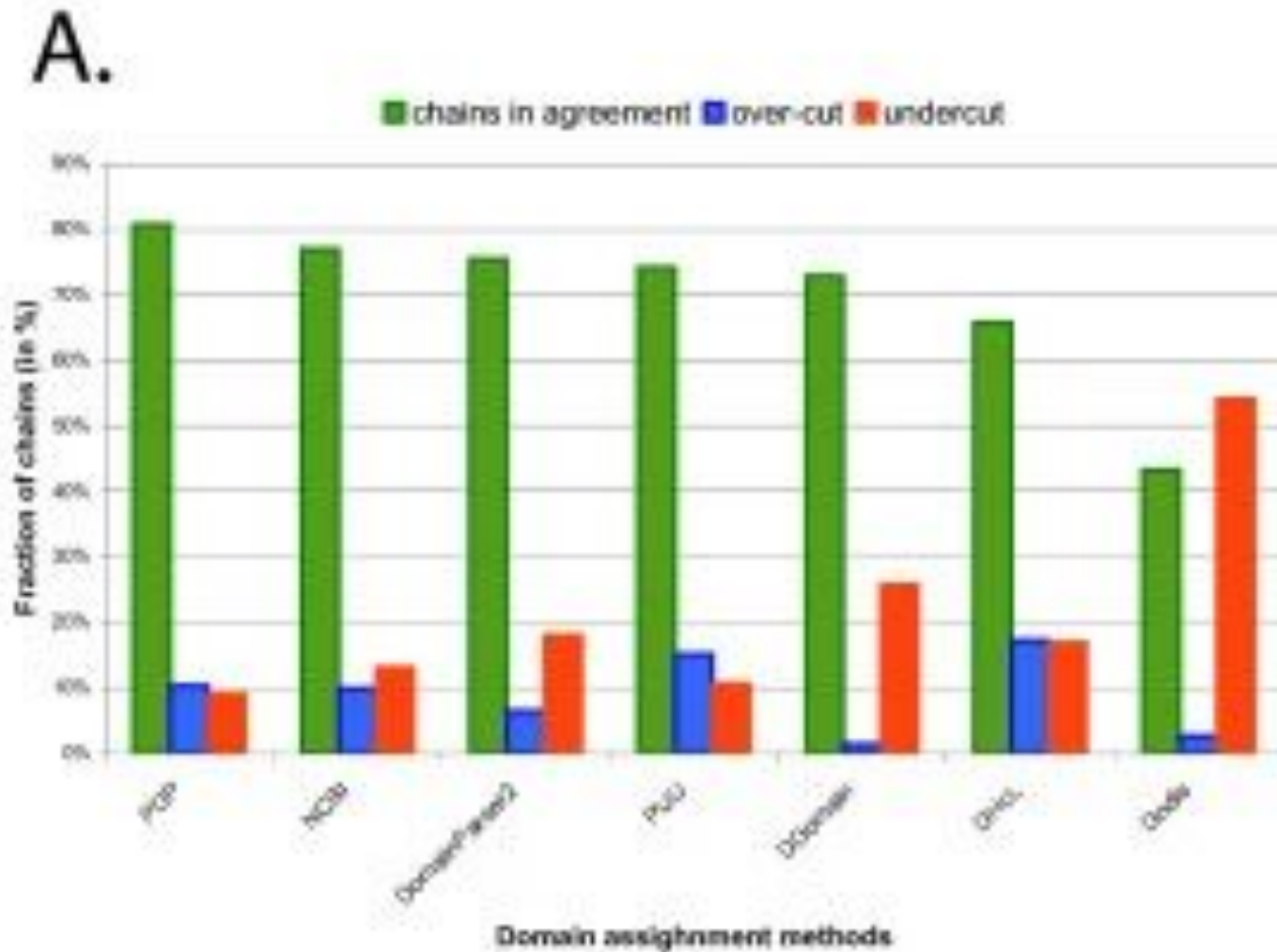
- Protein Domain Parser (PDP)
- DomainParser2
- PUU
- DDomain
- DHcL
- Dodis

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

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

*Alden et al., BMC Bioinformatics, 2010*

**dConsensus:** как не стоит делать иллюстрации к своей статье  
На 2020 год сервис не работает



# An ambiguity principle for assigning protein structural domains

Guillaume Postic<sup>1,2,3,4,\*</sup>, Yassine Ghouzam<sup>1,2,3,4</sup>, Romain Chebrek<sup>1,2,3,4</sup> and Jean-Christophe Gelly<sup>1,2,3,4,\*</sup>

+ See all authors and affiliations

*Science Advances* 13 Jan 2017:

Vol. 3, no. 1, e1600552

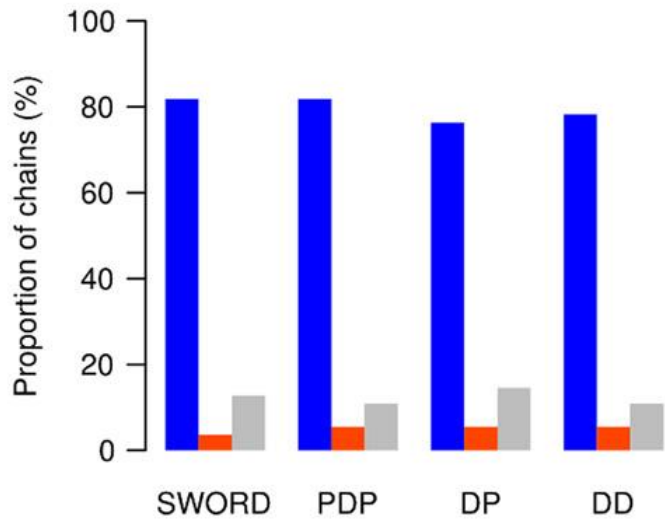
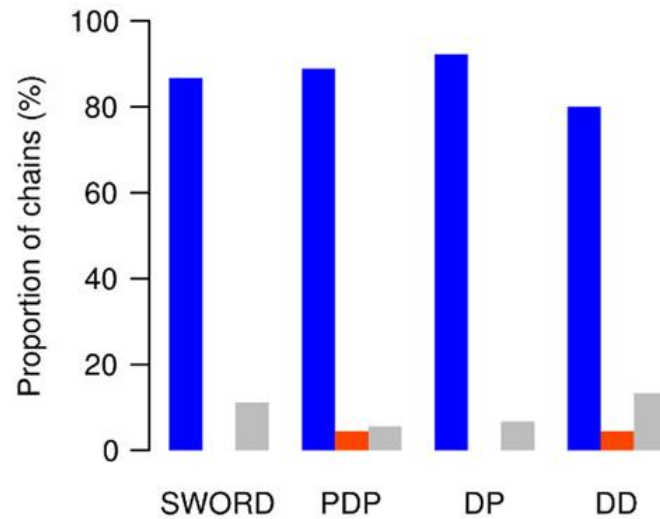
DOI: 10.1126/sciadv.1600552

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

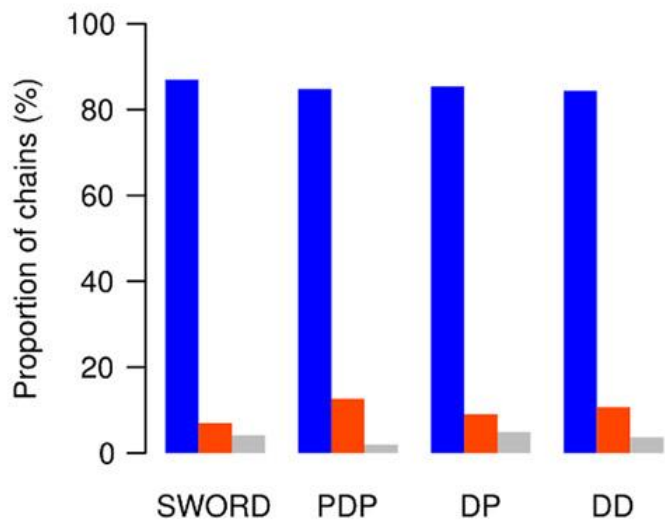
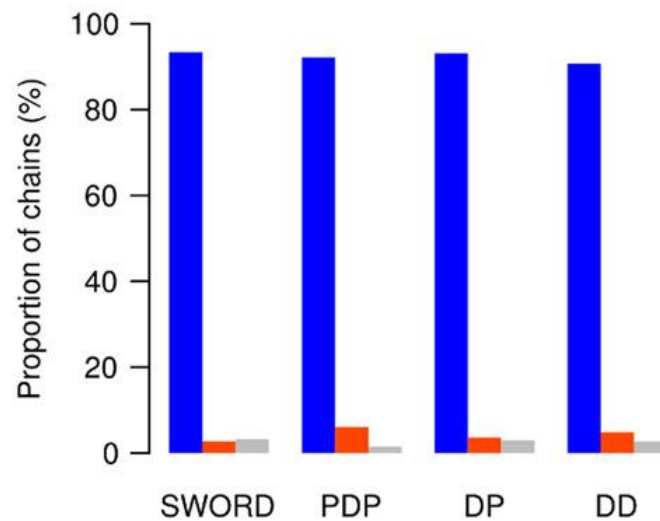
Jean-Christophe Gelly, Alexandre G. de Brevern , Serge Hazout [Author Notes](#)

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

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

**A****Jones set****Islam90 set**

Domain partitioning: ■ Correct ■ Overcut ■ Undercut

**Consensus set****Broad-consensus set**

# Классификации структурных доменов

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

- ручная детекция доменов
- 4 основных уровня классификации (класс, укладка, суперсемейство, семейство), также группирует по физико-химическому типу
- нестрогая иерархичность
- эволюционный контекст

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

- полуавтоматическая детекция доменов
- 4 основных уровня классификации (класс, архитектура, топология, суперсемейство)
- строгая иерархичность
- без контекста