



Эволюционные домены

PFAM

29 апреля 2025

Спирин Сергей Александрович
Алешина Юлия Александровна

Как понять что белки гомологичны?

Как понять что белки гомологичны?

❖ Выравнивание

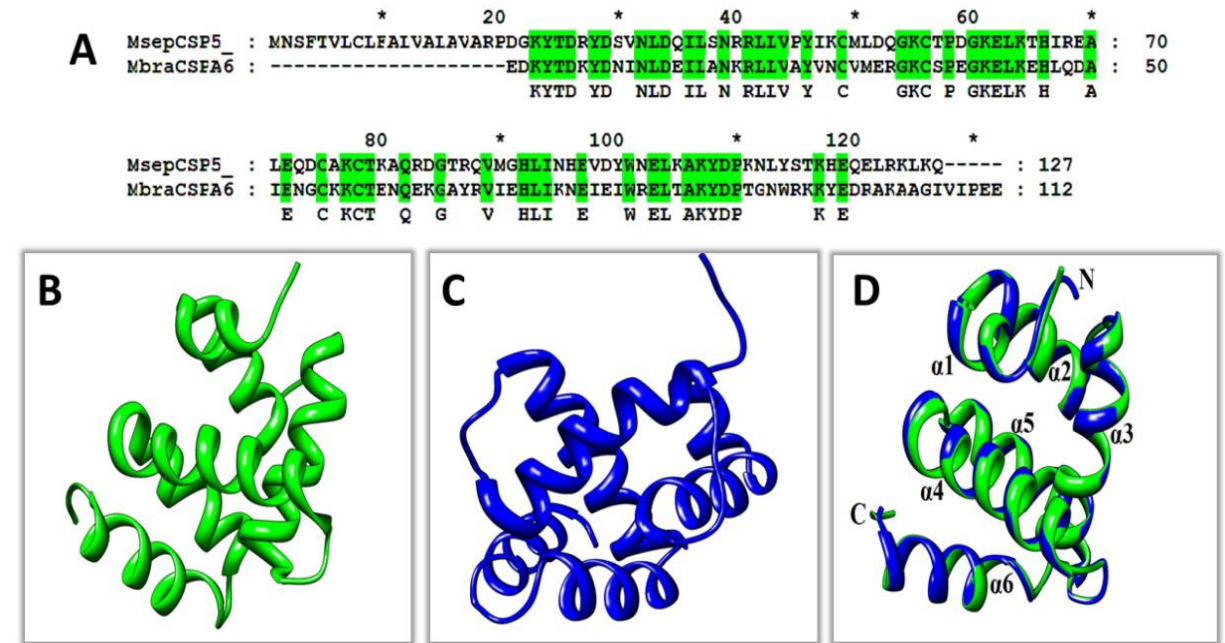
- ❖ Сходство (неслучайное!) последовательностей участков полипептидной цепи может свидетельствовать о гомологии белков
- ❖ В эволюционном выравнивании гомологичные аминокислоты стоят в одной колонке (кодоны происходят из одного и того же кодона общего предка белков)
- ❖ Напрямую проверить нельзя!
- ❖ Как находить блоки множественного выравнивания, построенного программой, которые соответствуют эволюционному?

	10										20																
K9Z0V0_CYAAP/42-80	L	A	Y	D	V	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	N	D	
K9SST0_9SYNE/42-80	L	A	Y	D	I	F	G	T	P	R	P	N	E	Y	T	N	D	R	Q	E	A	P	I	L	R	D	
K9SGD2_9CYAN/42-80	L	A	Y	D	V	F	G	T	P	R	P	N	E	Y	T	L	D	R	S	D	A	P	V	L	L	D	
K9Y0S0_STAC7/42-80	L	A	Y	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	L	E	L	P	I	N	D	
PSBE_RIPO1/42-80	L	A	Y	D	V	F	G	T	P	R	P	D	Q	Y	F	T	Q	E	R	Q	E	L	P	I	L	S	D
PSBE_SYNY3/42-80	L	A	Y	D	A	F	G	T	P	R	P	D	E	Y	F	T	Q	T	R	Q	E	L	P	I	L	Q	E
PSBE_GLOC7/42-80	L	A	Y	D	V	F	G	T	P	R	P	D	E	Y	T	Q	E	R	L	E	L	P	I	L	K	D	
PSBE_OSTTA/42-79	L	A	Y	D	V	F	G	T	P	R	P	N	E	Y	F	T	E	E	R	Q	E	L	P	L	I	S	D
K9W859_9CYAN/42-80	L	A	Y	D	V	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	V	E	A	P	I	L	S	D
PSBE_SYNE7/43-81	L	A	Y	D	A	F	G	T	P	R	P	N	E	Y	F	T	Q	D	R	T	E	V	P	I	V	S	D

А если белки накопили очень много различий?

А если белки накопили очень много различий?

- ❖ 3D-структуры белков консервативнее их последовательностей!
- ❖ Совмещение структур
 - ✓ C_{α} -атомы а.к.о. из одной колонки при совмещении 3D структур этих белков оказываются сближены



Совмещение структур белков

❖ Соответствие выравнивания совмещению 3D-структур можно проверить:

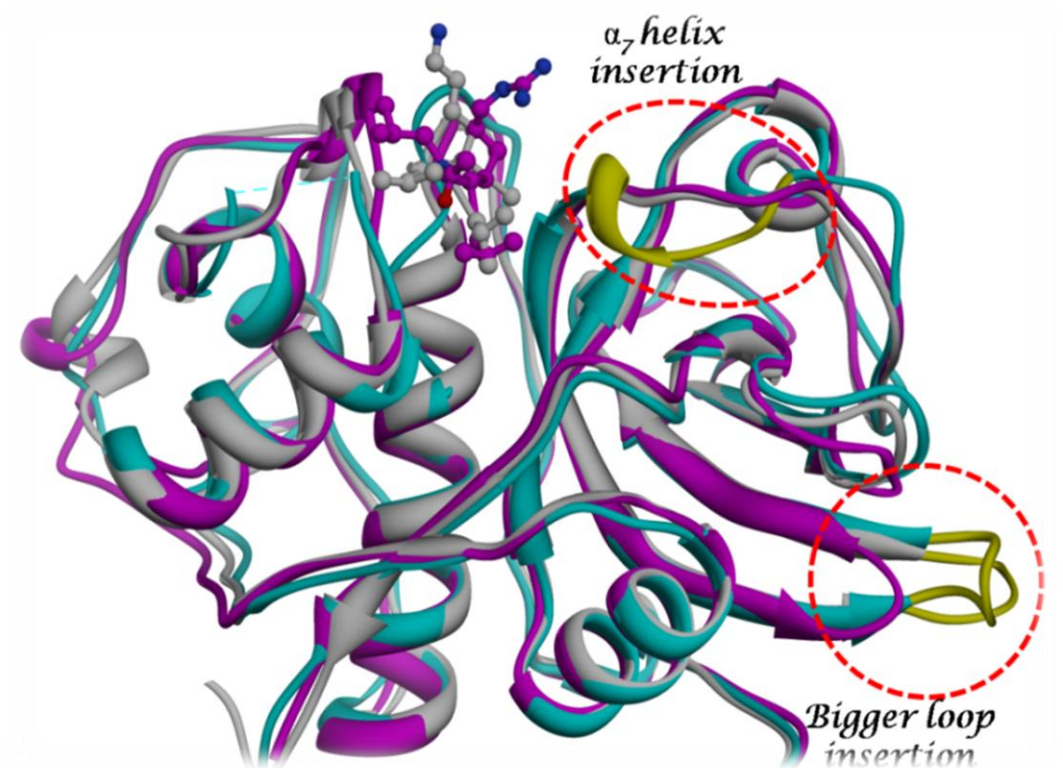
- ✓ Накопить денег и сделать PCA всех белков выравнивания
- ✓ **Проверить на белках с известной структурой**
- ✓ Предсказать 3D структуры всех белков и сравнить

Проблема предсказания: основой предсказания служит сходство фрагментов последовательности с фрагментами белков с известной 3D

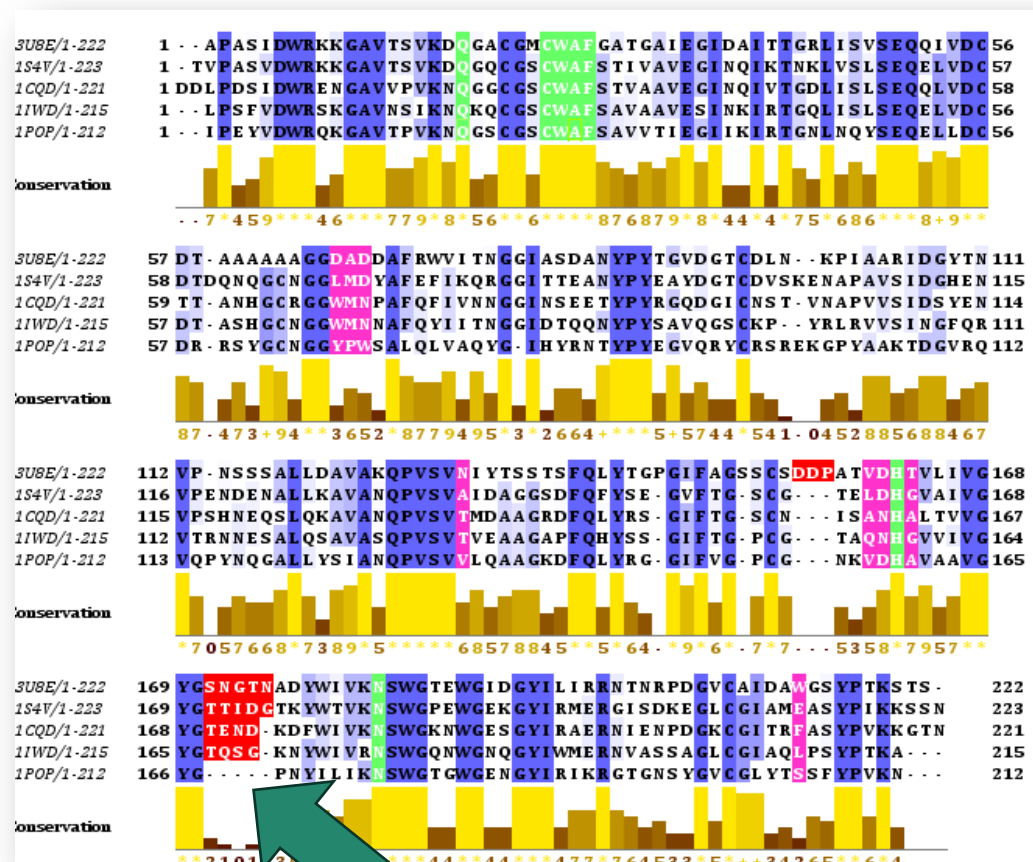
❖ Верно ли, что эволюция = сохранение хода полипептидной цепи в 3D?

Эволюция = сохранение хода полипептидной цепи в 3D?

- ❖ Отчасти, да – примеры совмещаемых структур со слабым сходством последовательностей
- ❖ Но сравнительно длинные вставки в одном и том же месте структуры могут произойти независимо и не быть гомологичными



Выравнивание по совмещению структур папаин-подобных цистеиновых протеаз растений



- Conserved
- Catalytic
- Active site
- 3U8E (*Crocus sativus*)
- 1S4V (*Ricinus communis*)
- 1CQD (*Zingiber officinale*)
- 1IWD (Ervatamin B)
- 1POP (*Carica papaya*)

В этих колонках мы не уверены, что амк. остатки гомологичны

Гомеобелки

Гомологичны ли эти белки?

[illegible]

Триптофансинтазы

		*	20	*	40	*	60	*	80	*				
TRPA_CHLTR	:	MSKLTQVFKQTK	----	LCIGYLTAGDGGTSYT	IEAAKALIQGGVD	ILELGFPFSDPVADNPE	IQVSHDRALAENLTSETL	LEIVEGIRAFNQEV	:	90				
TRPA_ECOLI	:	MERYESLFAQLKERKEGAFV	PFVT	LGDPGIEQSLKI	IDTLIEAGADALEL	GIPFSDPLADGPT	IQNATLRAFAAGVTPAQCF	EMLALIRQKHPTI	:	95				
TRPA_VIBCH	:	MNRYQALFQRLSAAQQGAFV	PFVT	IGDPNPEQSLAIMQ	TLIDAGADALEL	GMPFSDPLADGPT	IQGANLRALAAKTTPDIC	FELIAQIRARNPET	:	95				
TRP_COPCI	:	MEAIKKVFEQKKAQDATAFV	AFVTA	GYPKKEDTVPVLLAL	QAGGADI	IELGIPFSDPIADG	PTIQEANTVALKNDIDYPTV	LGQIREARQQGLTA	:	95				
TRP_NEUCR	:	MEGIKQTFQRCKAQNRALV	TYVT	AGFPHPETPDILL	AMEKGGADVIELGV	PFTDPIADGPT	IQNTANTIALQHGVTLQ	STLQMVVDARQRLKA	:	95				
A1C701_ASPCL	:	MEDIKSTFAKCKEQKRAALV	AYIT	AGYPTVEETVDILL	GLENGGADI	IELGIPFPTDPIADGPT	IQKANTKALANGVTVTIV	LQMVRTASRGLKA	:	95				
TRPB_CHLTR	:	-----							:	-				
TRPB_ECOLI	:	-----							:	-				
TRPB_VIBCH	:	-----							:	-				
		100	*	120	*	140	*	160	*	180	*			
TRPA_CHLTR	:	PLILYSYYPNLLQRDL	D-YLRLRK	DAGINGVCVIDL	PAPLSHGEKSP	FFEDLLAVGLDPI	LLISAGTTPERMSLI	QEYARGFLYYIP	PCQATR	DSE	: 184			
TRPA_ECOLI	:	PIGLLMYANLVFNKGID	EFYAQCEKVGVD	SVLVADVPE	-----	ESAPFRQAALRHNV	APIFCPPNADDDLLRQ	IASYGRGYTYLLS	RAGVTGAE	:	186			
TRPA_VIBCH	:	PIGLLMYANLVYARGID	DFYQRCQKAGVDS	VLIADVPTN	-----	ESQPFVAAA	EKFQIPIFIAPPTAS	DETLRAVAQLGKG	YTYLLS	RAGVTGAE	: 186			
TRP_COPCI	:	PVLLMGYYNPMLAYGED	KAIQDAAEAGANGF	IMVDLPPE	-----	EAIAFRQKCAASNLSY	VPLIAPSTTLKRI	QFLASIA	DSFIYVVSKMGT	TGSS	: 186			
TRP_NEUCR	:	PVMLMGYYNPLLSYGE	ERLLNDCKEAGVNG	FIIVDLPPE	-----	EAVSFRQLCTR	GGLSYVPLIAPATSD	ARMVLCQLAD	SEFIYVVS	RQVGTGAS	: 186			
A1C701_ASPCL	:	PVLLMGYYNPLLRYGE	ERMLKDCKEAGVNG	FIMVDLPPE	-----	EAVFRFDLCASNGLSY	VPLIAPATSESRM	KLLCKIAD	SEFIYVVS	SRMGVTGAT	: 186			
TRPB_CHLTR	:	-----									-			
TRPB_ECOLI	:	-----									-			
TRPB_VIBCH	:	-----									-			
		200	*	220	*	240	*	260	*	280				
TRPA_CHLTR	:	VG-----	IKEEF	RKVREHF-DLPI	VDRRDI	CDKKEAAHV	LNYS-DG-FIVK	TAFVHQTT--	MDSSV-----	ETLTALAQT	VIPG-----	: 253		
TRPA_ECOLI	:	NRAALP--	LNHLVAKLKEYN	-AAPPLQGF	GISAPDQVKAAID	DAGAAAG-AISG	SAIVKIEQHINEPE	-----	KMLAALKV	FVQPMKAATRS--	: 268			
TRPA_VIBCH	:	TKANMP--	VHALLERLQQFD	-APPALLGFG	ISEPAQVKQAI	EAGAAAG-AISG	SAVVKIETHLDNPA	-----	KQLTELAN	FTQAMKKATKI--	: 268			
TRP_COPCI	:	ANVAVNEEL	PTILSRIREYT	-HVLAVEF	GVATRDQFN	YVADAGADGVV	IGSRIVNAIKAAGE	GQVPQFVENY	CREVSGKG	-----	: 267			
TRP_NEUCR	:	GTLNAN--	LPELLARVKKYSGNKP	AAVGFGV	STHDFHTQVGAI	-ADG-VVVG	SMIITTIQKAAK	GEEVKAVQ	EYCSYLCGRN	FEQSA-----	HEL	: 272		
A1C701_ASPCL	:	GKLSSN--	LPELLKRVHQW	SGNVPAALGFGV	STREHFLDVQEL	-AEG-VVIG	SQIITTLAQAP	PAGQAQAE	EYLSVTGRRR	ERDAQGALTHEI	: 277			
TRPB_CHLTR	:	-----									-			
TRPB_ECOLI	:	-----									-			
TRPB_VIBCH	:	-----									-			
		*	300	*	320	*	340	*	360	*	380			
TRPA_CHLTR	:	-----									-			
TRPA_ECOLI	:	-----									-			
TRPA_VIBCH	:	-----									-			
TRP_COPCI	:	----	EPSRVRS	PGAAQRT	PSQLTPNAETA	-----	KGVENI	LPA	FGQFGGQYVPES	LVDA	LALEEAHKS	AIEDPAFWEEVRS	LY-T	: 344
TRP_NEUCR	:	----	NMGEALEAAKE	PPVGTAT	VDGVITEADIDA	-----	QLAALHGT	IPKRF	FGFEGGQYVPEAL	MDCLSELE	EFGFNKIKDD	PAFWEEYRS	YY-P	: 355
A1C701_ASPCL	:	----	NVLEAVEKVQ	TPAVSQPTD	-VITDADTPAG	PGLADQIEAL	NGAGNPAAQPS	RFGEFGGQYVPES	SLMDCLAE	LERGFQ	QALNDPKFWEE	FRSYY-P	: 370	
TRPB_CHLTR	:	-----												47
TRPB_ECOLI	:	-----												51
TRPB_VIBCH	:	-----												50

Продолжение выравнивания на следующем слайде

```

TRPA_CHLTR : -----*-----300-----*-----320-----*-----340-----*-----360-----*-----380-----: -
TRPA_ECOLI : -----: -
TRPA_VIBCH : -----: -
TRP_COPCI : ----EPSRVRSFGAAQRTPSQLTPNAETA-----KGVENILPARFGQFGGQYVPESLVDALAELEEAHKSIEDPAFWEEVRSLY-T : 344
TRP_NEUCR : NMGEALEAAKEPVGTTATVDGVITEADIDA-----QLAALHGTIPKRFGFEGGGQYVPEALMDCLELEEGFNKIKDDPAFWEEYRSYY-P : 355
A1C701_AS PCL : NVLEAVEKVQTPAVSQPTD-VITDADTPAGPLADQIEALNGAGNPAAQPSRFGEFGGQYVPESLMDCLAEELERGFQQALNDPKFWEEFRSYY-P : 370
TRPB_CHLTR : -----MFKKHK-PFGGAFLEPEELLAPITQNLKAWEILKTQQSFLSELDCILKN : 47
TRPB_ECOLI : -----MTTLLNPYFGEFGGMVYVQILMPALRQLEEAQFVSAQKDPEFQAQFNDLLKN : 51
TRPB_VIBCH : -----MAKLNAYFGEFGGQFVFPQILVPALEQLEQAFIDAQQDDAFRAEFMSLLQE : 50

TRPA_CHLTR : -----*-----400-----*-----420-----*-----440-----*-----460-----*-----: -
TRPA_ECOLI : -----: -
TRPA_VIBCH : -----: -
TRP_COPCI : YSNRPSNLYLAENLTKEAGGANIWLKREDLNHTGSHKINNALGQILLAKRIGKTRIIAETGAGQHG VATATVCAKFGLECVIYMGAEDVRRQALK : 439
TRP_NEUCR : WMGRPGQLHKAERLLEYAGGANIWLKREDLNHTGSHKINNALGQILLARRLGKKKIIAETGAGQHG VATATVCAKFGMECTVFMGAEDVRRQALN : 450
A1C701_AS PCL : YMGRPSSMHLANRLTEHVGGGANIWLKREDLNHTGSHKINNALGQILIAARLGKTRIIAETGAGQHG VATATVCAKFGMKCVVYMGAEDVRRQALN : 465
TRPB_CHLTR : YAGRQTPLETVKNFARAIDGPRVFLKREDLLHTGAHKLNNALGQCLLAKYLGKTRVVAETGAGQHG VATATACAYLGLDCVVYMGAKDVERQKPN : 142
TRPB_ECOLI : YAGRPTALTCKQNIITAGTN-TTLYLKREDLLHGGAHKTNOVLGQALLAKRMGKTEIIAETGAGQHG VASALASALLGLKCRIVMGAKDVERQSPN : 145
TRPB_VIBCH : YAGRPTALTTLQNIITKGTK-TKLYLKREDLLHGGAHKTNOVLGQALLAKRMGKHEIIAETGAGQHG VATALACALLGLKCRVYMGAKDVERQSPN : 144

TRPA_CHLTR : -----480-----*-----500-----*-----520-----*-----540-----*-----560-----*-----: -
TRPA_ECOLI : -----: -
TRPA_VIBCH : -----: -
TRP_COPCI : LFRIEMLGKAWVIPVHSGSCTLKDAVNEAMRDWVTNLSTTHYLVGSAIGPHFPPTIVRDFQKVIGEEIKAQLEKVRGKLPDVVVACVGGGSNAI : 534
TRP_NEUCR : VERMKLLGAK--VVAVEAGSRTLRLDAVNEALRYWVNLADTHYIIGSAIGPHFPPTIVRTFQSVIGNETKQOMLEKRGKLPDAVVACVGGGSNAV : 543
A1C701_AS PCL : VERMKLLGAS--VVAVDAGSRTLRLDAVNEALRAWVVDLDTHYIIGSAIGPHFPPTIVRTFQSVIGEEETKQOMKEAIGKLPDAVVACVGGGSNAV : 558
TRPB_CHLTR : VEKMRFLGAE--VVSVTKSGCGLKDAVNQALQDWATTHSFTHYCLGSALGPLPYDPDIVRFFQSVISAEVKEQIHAVAGRDPDILIAICIGGGSNAI : 235
TRPB_ECOLI : VERMRLMGAE--VIPVHSGSATLKDACNEALRDWSGSYETAHYMLGTAAGPHPYPTIVREFQRIIGEETKNQILAREGRLPDAVIACVGGGSNAI : 238
TRPB_VIBCH : VERMRLMGAT--VIPVHSGSATLKDACNEALRDWSASSETAHYLLGTAAGPHFPPTIVREFQRIIGEETKNQILAREGRLPDAVIACVGGGSNAI : 237

TRPA_CHLTR : -----580-----*-----600-----*-----620-----*-----640-----*-----660-----: -
TRPA_ECOLI : -----: -
TRPA_VIBCH : -----: -
TRP_COPCI : GTFYDFIPDKSVRLVGVEAGGEGIDGKHHSATLSMGQPGVLHGVRTYIILQDKAGQIIETHSISAGLDYPGVGPEHAWLKDSGRADYVVC TDE DAL : 629
TRP_NEUCR : GMFYPPFSNDPSVKLLGVEAGGDGVDTPRHSATLTAGSKGVLHGVRTYIILNQYQGQIEDTHSISAGLDYPGVGPESLSNWKDTERAKFVAATDAQAF : 638
A1C701_AS PCL : GMFYPPFAKDTSVKLVGVEAGGDGIDTNRHSATLSGGSKGVLHGVRTYVLDQEHGQISDTHSISAGLDYPGVGPESLSNWKDSDRAHFIAATDAQAL : 653
TRPB_CHLTR : GFFHHFIPNPKVQLIGVEGGGLGISSGKHAAHFATGRPGVFHGFYSYLLQDDQGVQLQTHSISAGLDYPSVGPDHAEMHESGRAFYTLATDEEAL : 330
TRPB_ECOLI : GMFADFINEITNVGLIGVEPGGHGIEGTGEHGAFLKHGRVGIYFGMKAPMMQTEDGQIEESYISAGLDFFSVGPQHAYLNSTGRADYVSI TDDEAL : 333
TRPB_VIBCH : GMFADFIEESVRLIGIEPAGKGIHTHQHGAFLKHGKTGIFFGMKAPLMQDEHQVQVEESYSVSAGLDFFSVGPQHAYLNAIGRAEYESI TDDEAL : 332

TRPA_CHLTR : -----*-----680-----*-----700-----*-----720-----*-----740-----: -
TRPA_ECOLI : -----: -
TRPA_VIBCH : -----: -
TRP_COPCI : RGFRLMTQKEGIIPALES SHAIWEGVKIAKSLP-KDKDIVICLSGRGDKDVEQISELPKWADKLDWHVSSNAIPSK : 705
TRP_NEUCR : EGFRLMSQLEGIIPALES SHGIWGALELAKTMK-PDEDVVICLSGRGDKDVQSVADELP IIGPKIGWDLRF-----: 708
A1C701_AS PCL : IGFRTLAQTEGIIPALES SHAVWGAMELAKSM--KTGDIVNLNLSGRGDKDVQSVADELPRLGPOIGWDLRF-----: 722
TRPB_CHLTR : RAFFLLTRNEGIIPALES SHALAHLVSIAPSLP-KEQIVIVNLNLSGRGDKDLPQIIRNRNGIYE-----: 392
TRPB_ECOLI : EAFKTLCLHEGIIPALES SHALAHALKMMRENPDKEQLLVNLSGRGDKDIFTVHDILKARGEI-----: 397
TRPB_VIBCH : DAFQALARNEGIIPALES SHALAHAIKMAAYEPDKEQLLVNLSGRGDKDIFTVHKLLEDKGAL-----: 396

```

Эволюция белков

Локальные изменения - результат небольших изменений в гене (замены аминокислот, делеции, вставки)

Большие изменения:

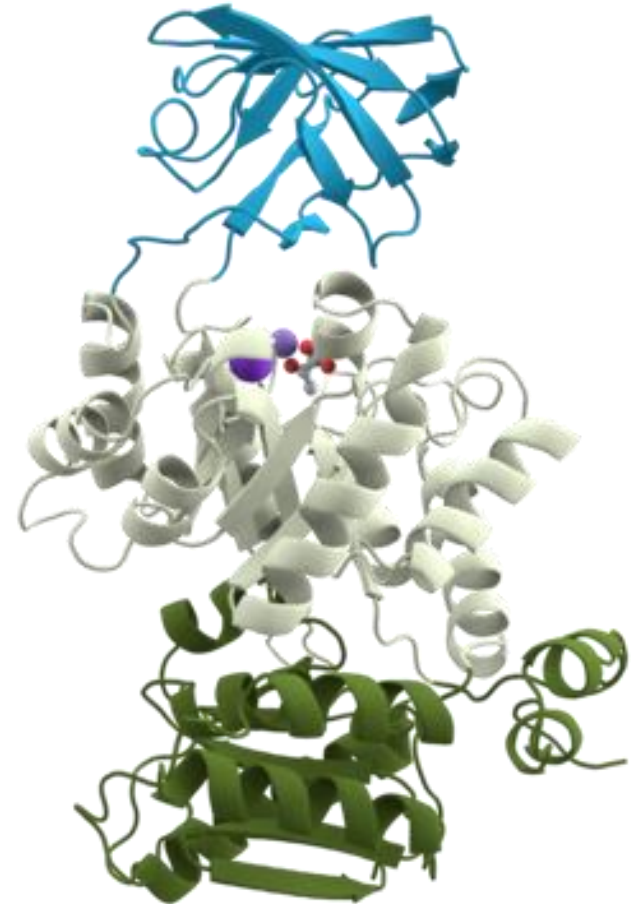
- ❖ Накопленные небольшие изменения
- ❖ Небольшие изменения гена, ведущие к большим изменениям белка
 - ✓ Мутация стоп кодона => удлинение последовательности белка
 - ✓ Мутация кодона на стоп кодон
 - Гибель белка = псевдогенизация
 - Укорочение последовательности белка
 - Программируемый сдвиг рамки считывания
 - ✓ Мутация в сайте инициации (начала) трансляции – приобретение нового старт-кодона
 - ✓ Делеция или вставка, не кратная 3 => сдвиг рамки считывания
- ❖ **Крупные перестройки генома, затрагивающие гены (инверсии, транслокации, большие делеции или вставки)**
- ❖ Закодированные в геноме перестройки для быстрого изменения последовательности белка (Ig, DGR)

Домены белков

- ❖ **ЭВОЛЮЦИОННЫЙ ДОМЕН** – достаточно длинный (более многих десятков а.к.о.) участок предкового белка, который эволюционировал только по типу локальных мутаций.
- ❖ При этом белки-потомки могли претерпевать крупные перестройки, не затрагивающие домен.
- ❖ В белке может быть один домен, два или много.
- ❖ **Доменная архитектура** – последовательность доменов в белке

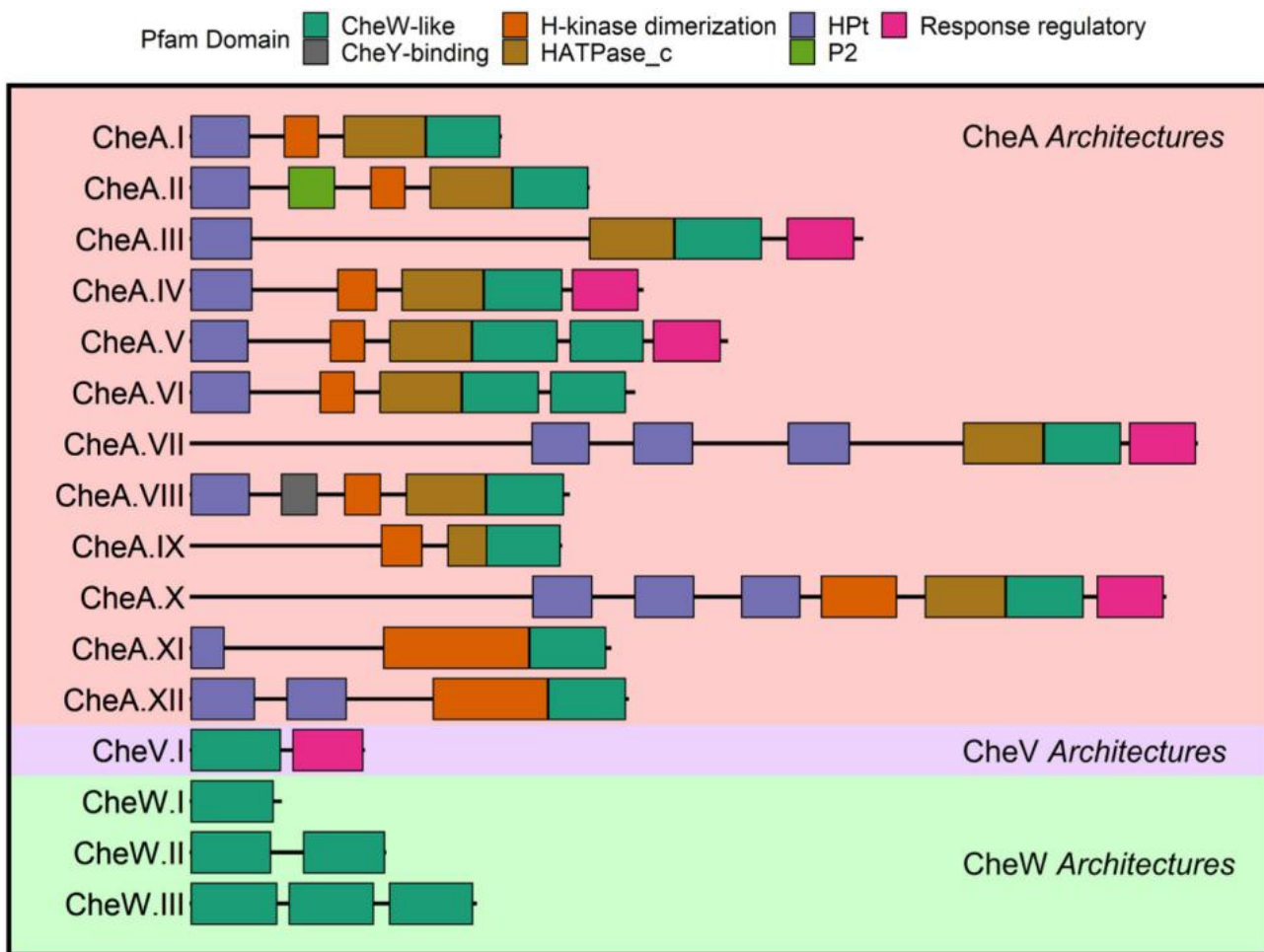
*Термин «домен белка» можно понимать по-разному

- ❖ Структурный домен – структурная единица белка, которая сворачивается независимо
- ❖ Функциональный домен – часть белка, ответственная за осуществление определенной функции

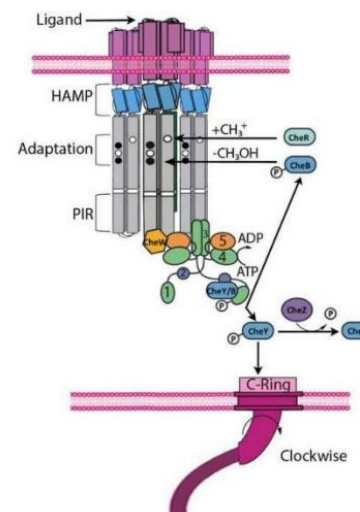


Пируваткиназа
(PDB: 1PKN)

Так изображают домены



Che белки участвуют в хемотаксисе бактерий



Доменные архитектуры белков, содержащих домен CheW-like

Многодоменные белки в эволюции образуются в результате последовательных единовременных крупных перестроек в генах белков

Домен белка – единица непрерывной эволюции белков

Семейства доменов

- ❖ ЭВОЛЮЦИОННЫЙ ДОМЕН – достаточно длинный (более многих десятков а.к.о.) участок предкового белка, который эволюционировал только по типу локальных мутаций.
- ❖ Домену дают название. Собирают представителей домена из всех белков, в которых их удаётся найти и строят выравнивание участков, соответствующих домену.
- ❖ Семейство доменов – совокупность таких участков.

Семейства доменов и базы данных

- ❖ Pfam
- ❖ PROSITE
- ❖ Prints
- ❖ SMART
- ❖ TIGERFAM



an **inte**grated **r**esource of **pro**tein families,
domains and functional sites

- В занятиях работаем с базой данных Pfam – protein families
- Pfam в 2023 году была поглощена консорциумом InterPro, посвященным семействам белков

Pfam – protein families



- ❖ Коллекция семейств доменов (~ 24 k записей)
- ❖ Состоит из двух частей
 - ✓ PfamA – курируемая часть
 - ✓ PfamB – семейства, сгенерированные автоматически

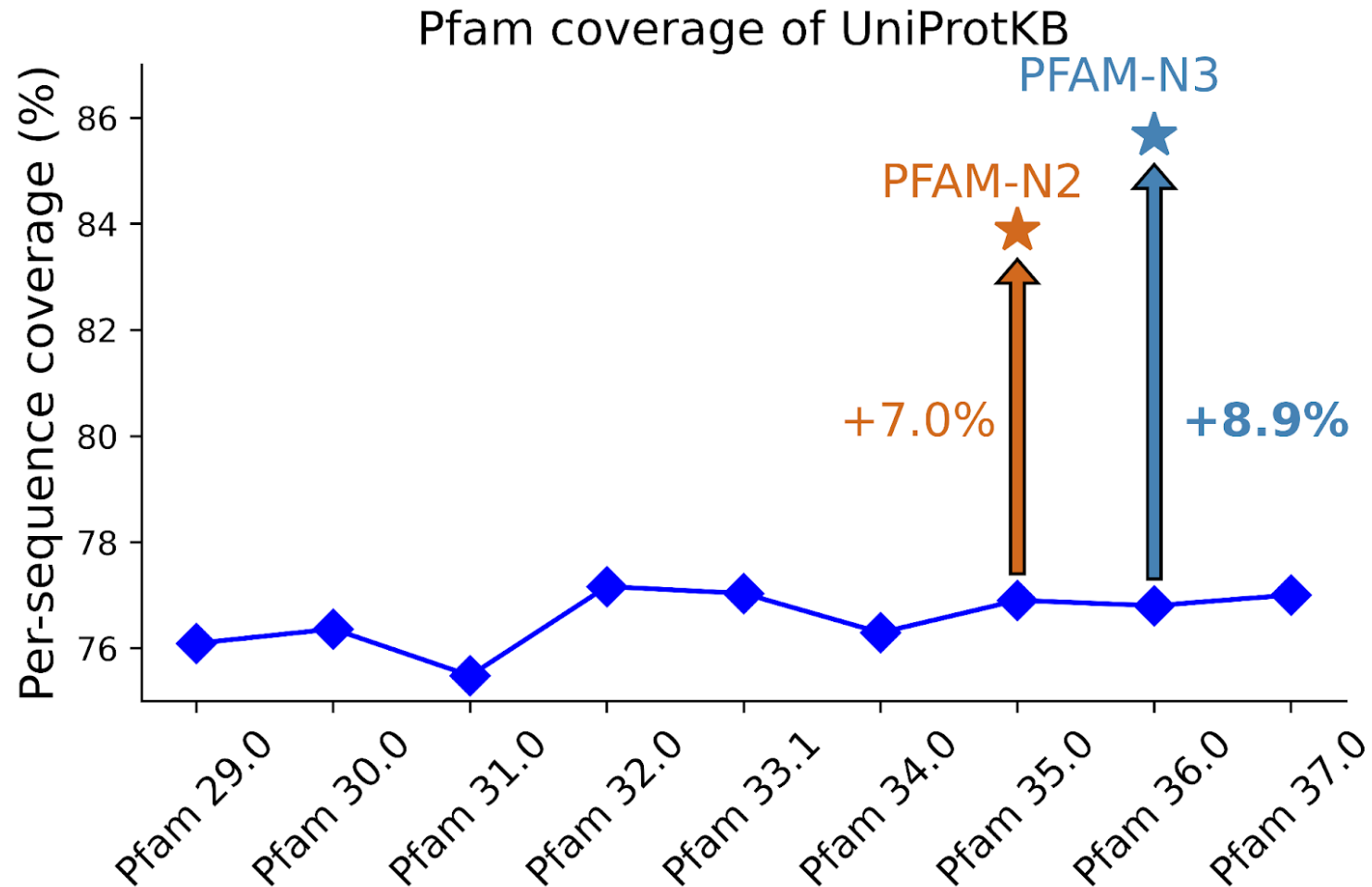
The Pfam protein families database: embracing AI/ML

Typhaine Paysan-Lafosse , Antonina Andreeva , Matthias Blum ,
Sara Rocio Chuguransky , Tiago Grego , Beatriz Lazaro Pinto , Gustavo A Salazar ,
Maxwell L Bileschi , Felipe Llinares-López , Laetitia Meng-Papaxanthos ... [Show more](#)

Nucleic Acids Research, Volume 53, Issue D1, 6 January 2025, Pages D523–D534,
<https://doi.org/10.1093/nar/gkae997>

Published: 14 November 2024 **Article history** ▼

Pfam-N – аннотации белковых доменов, полученные с помощью нейросети

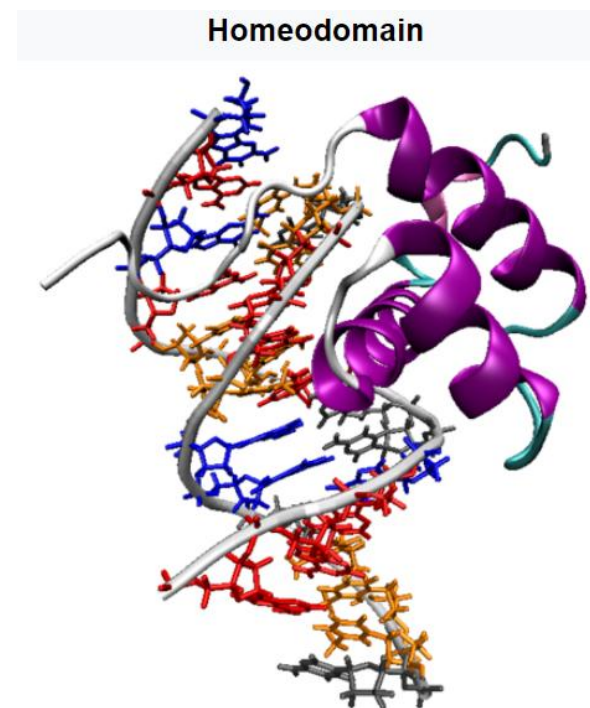


Эволюционные домены

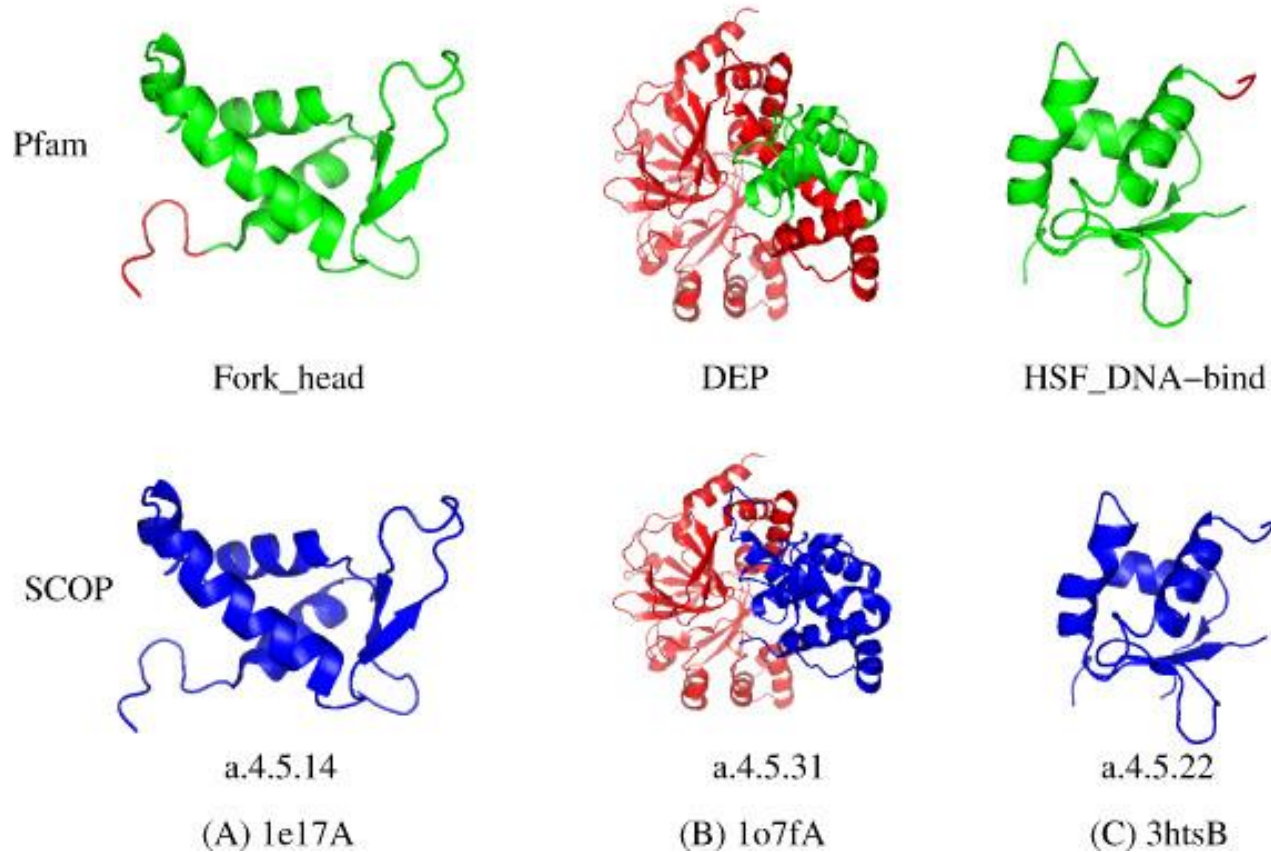
- ❖ Имеют определенную функцию (не всегда известна)
DUF – Domain of Unknown Function
- ❖ Часто совпадают со структурными доменами
(но не всегда)

Гомеодомен – ДНК СВЯЗЫВАЮЩИЙ ДОМЕН

Homeodomain proteins regulate gene expression and cell differentiation during early embryonic development, thus mutations in homeobox genes can cause developmental disorders.^[1]



Домены Pfam часто, но не всегда соответствуют структурным доменам SCOP



Такое соответствие
наблюдается не для
всех доменов Pfam

Domains in the chain

Secondary Structure

PDB: Chain C

Representative families

F PAX_fam - PAX family
PANTHER: PAIRED BOX PROTEIN PAX-6-REL

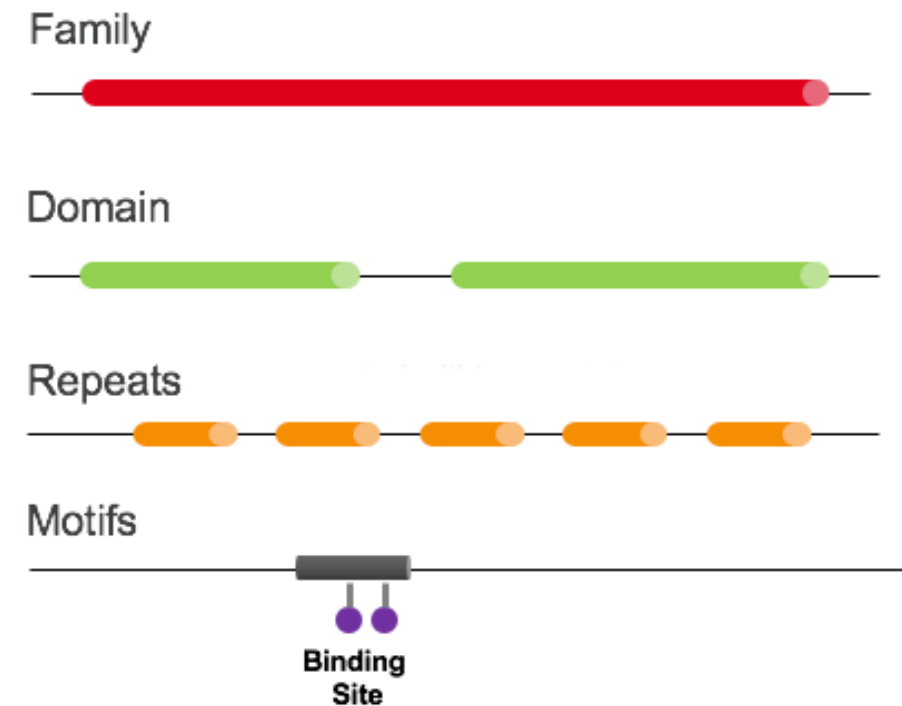
Representative domains

D Paired_dom - Paired domain
SMART: PAX - Paired Box domain
PROFILE: PAIRED_2 - Paired DNA-binding domain
PRINTS: PAIREDBOX - PAIREDBOX
PFAM: PAX - 'Paired box' domain
CDD: PAX - PAX

Pfam – protein families

❖ Типы записей в Pfam:

- ✓ Семейство – коллекция гомологичных участков белков.
- ✓ Домен – устойчивая структурная единица, или функциональный участок, встречающийся в различных белковых архитектурах.
- ✓ Повтор – короткая единица, нестабильная сама по себе, но образует стабильные структуры, если есть много копий.
- ✓ Мотив – короткий консервативный участок
- ✓ Клан – выделенная вручную группа похожих по последовательности и структуре доменов.



Pfam – protein families



- ❖ Для каждого семейства доменов есть
 - ✓ Название домена, ID AC, описание функции
 - ✓ Белки из UniProt, содержащие домен
 - ✓ Доменные архитектуры
 - ✓ Таксономическая распространенность
 - ✓ 3D структуры
 - ✓ **Множественное выравнивание**
 - **seed** - выравнивание семейства гомологичных фрагментов из разных белков, составленное экспертом для построения модели
 - **full** – выравнивание всех представителей семейства, хранящихся в Pfam (Interpro теперь позволяет скачать выравнивания PfamA целиком)
 - ✓ **НММ-профиль выравнивания**– модель семейства=математическое описание выравнивания по колонкам (Curation)

Pfam PF00046 Homeodomain

Pfam entry ⓘ

Overview

Proteins	263k
Domain Architectures	2k
Taxonomy	14k
Proteomes	3k
Structures	223
Profile HMM	
AlphaFold	170k
BFVD	8
Alignment	

Member database	Pfam ⓘ
Pfam type	domain
Short name	Homeodomain
Clan	HTH
Author	Eddy SR;0000-0001-6676-4706 ⓘ
Sequence Ontology	0000417

Description ⓘ Imported from IPR001356

This entry represents the homeodomain (HD), a protein domain of approximately 60 residues that usually binds DNA. It is encoded by the homeobox sequence [7, 6, 8], which was first identified in a number of *Drosophila* homeotic and segmentation proteins, but is now known to be well-conserved in many other animals, including vertebrates [1, 2], as well as plants [4], fungi [5] and some species of lower eukaryotes. Many members of this group are transcriptional regulators, some of which operate differential genetic programs along the anterior-posterior axis of animal bodies [3]. This domain folds into a globular structure with three α -helices connected by two short loops that harbour a hydrophobic core. The second and third form a helix-turn-helix (HTH) motif, which make intimate contacts with the DNA: while the first helix of this motif helps to stabilise the structure, the second helix binds to DNA via a number of hydrogen bonds and hydrophobic interactions, which occur between specific side chains and the exposed bases and thymine methyl groups within the major groove of the DNA. One particularity of the HTH motif in some of these proteins arises from the stereo-chemical requirement for glycine in the turn which is needed to avoid steric interference of the β -carbon with the main chain: for *cro* and repressor proteins the glycine appears to be mandatory, while for many of the homeotic and other DNA-binding proteins the requirement is relaxed.

Accession code
домена

Короткое название домена

Pfam PF00046 Homeodomain

Pfam entry ⓘ

Overview

Proteins 263k

Domain Architectures 2k

Taxonomy 14k

Proteomes 3k

Structures 223

Profile HMM

AlphaFold 170k

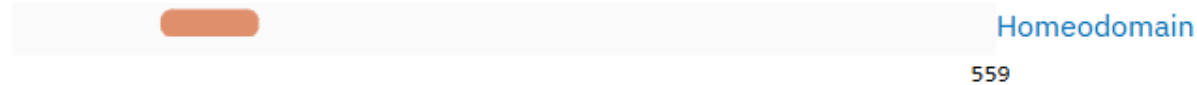
BFVD 8

Alignment

1881 domain architectures

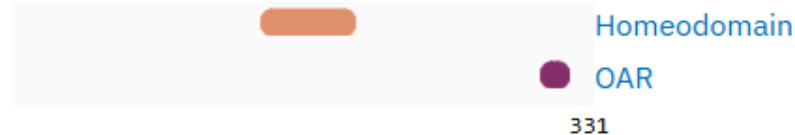
There are 136209 proteins with this architecture (represented by P07269):

PF00046



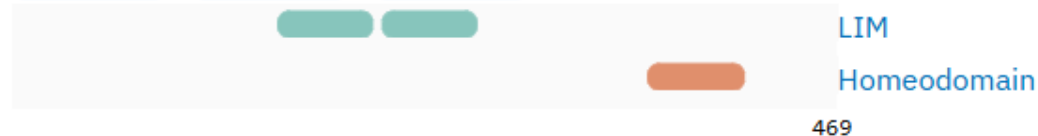
There are 14044 proteins with this architecture (represented by P70390):

PF00046 - PF03826



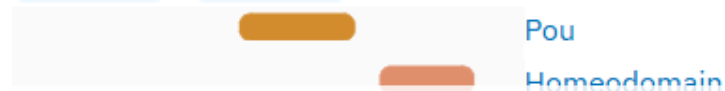
There are 13163 proteins with this architecture (represented by P29673):

PF00412 - PF00412 - PF00046



There are 11827 proteins with this architecture (represented by P10036):

PF00157 - PF00046



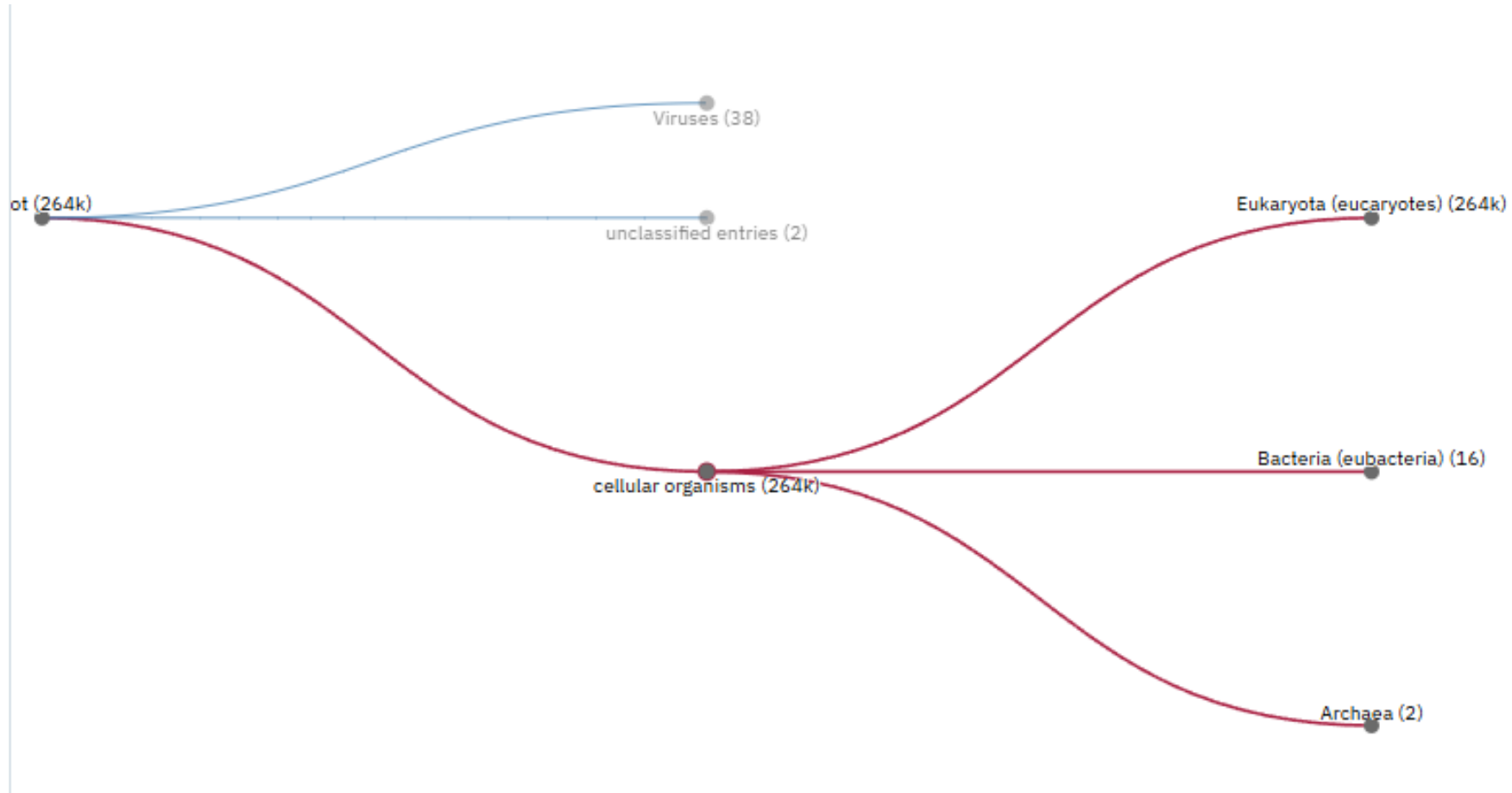
- Домены могут комбинироваться друг с другом – как модули
- Вариантов доменных архитектур может быть очень много

Таксономическая представленность домена



Диаграмма показывает, у **кого** встречается домен

Таксономическая представленность домена

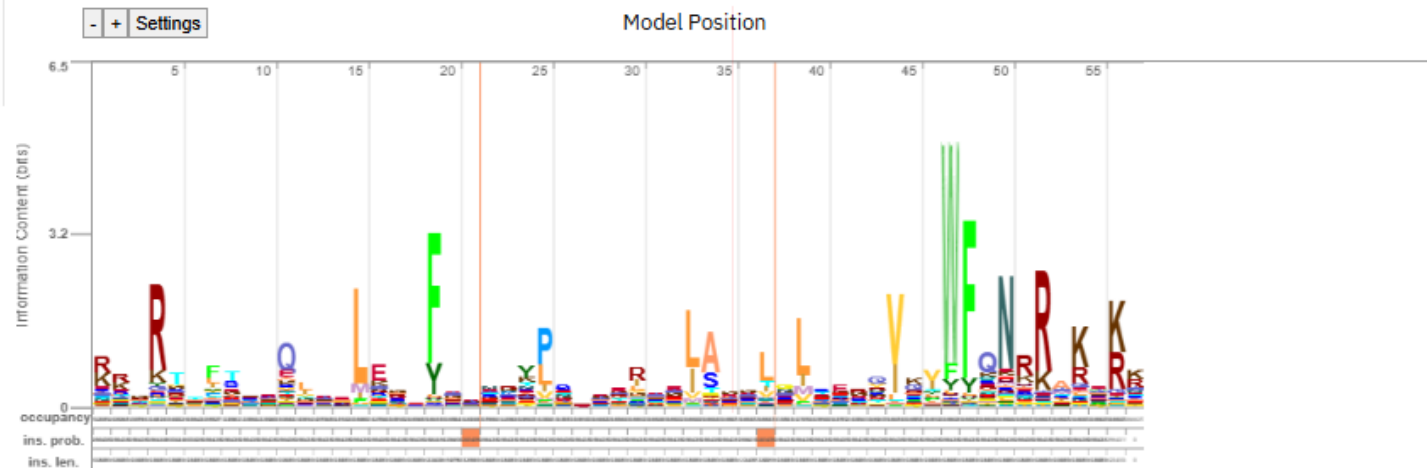


Overview	
Proteins	263k
Domain Architectures	2k
Taxonomy	14k
Proteomes	3k
Structures	223
Profile HMM	
AlphaFold	170k
BFVD	8
Alignment	

Profile HMM Information

HMM build commands	Build method: hmmbuild HMM.ann SEED.ann Search method: hmmsearch -Z 81514348 -E 1000 --cpu 4 HMM pfamseq
Gathering threshold	Sequence: 22.9 Domain: 22.9
Download	Download the raw HMM for this family

Profile HMM logo



- У каждого домена есть **НММ-профиль выравнивания** – модель семейства=математическое описание выравнивания по колонкам (Curation)
- НММ-профиль строится по выравниванию SEED
- Его можно использовать для поиска доменов данного семейства

ДВА ДОМЕНА гомеобелков: гомеодомен и OAR домен

SW: PMX1_CHICK/1	-----	20	40	60	80	100	120	140	1	80
SW: PMX2_HUMAN/1	-----									86
SW: PMX1_HUMAN/1	-----									80
SW: APX_BRARE/1	ISQAPQVSISRSKSYREN-APFSQS---	D-EGQSP	---EHMAQLVELST		---	LKFEEEDVVKKEACQDN	---	S		84
SW: APX_MOUSE/1	ISQAPQVSISRSKSYRENGAPFVPPPPALD-	ELSGPGCVAHPEERLSAASGPGSAPAAGCGTGAEDDEE	EELEDEDEEEEEELLEDDDEELLEDDARALLKEPRRCVSATTCTG	VAAAAA	AAAAAA	AAVATEGGEELSPEKEELLHPEDAEGKDGEDSVCLS				157
SW: AL_DROME/1-1	-----									72
SW: ALX4_MOUSE/1	-----									145
SW: ALX4_HUMAN/1	-----									157
SW: RX2_CHICK/1	-----									83
SW: RX2_BRARE/1	-----									92
SW: RX1_XENLA/1	-----									91
SW: RX_HUMAN/1-1	-----									97
SW: PIX2_BRARE/1	-----									66
SW: PIX2_HUMAN/1	-----									68
SW: PIX1_HUMAN/1	-----									64
SW: OTP_MOUSE/1	-----									90
SW: PMX1_CHICK/1	NDQLNSEE-----	180	200	220	240	260	280	300	*	213
SW: PMX2_HUMAN/1	CECPSPGRCG-----									221
SW: PMX1_HUMAN/1	NDQLNSEE-----									213
SW: APX_BRARE/1	ACGDSSEEC-----									230
SW: APX_MOUSE/1	ACGDSSEEC-----									303
SW: AL_DROME/1-1	DCEADEYA-----									212
SW: ALX4_MOUSE/1	EKADSESN-----									290
SW: ALX4_HUMAN/1	EKADSESN-----									302
SW: RX2_CHICK/1	KPSDEEQ-----									215
SW: RX2_BRARE/1	PDIPDEDQ-----									225
SW: RX1_XENLA/1	KLSDEEQ-----									224
SW: RX_HUMAN/1-1	KLSDEEQ-----									242
SW: PIX2_BRARE/1	SKNEDSN-----									212
SW: PIX2_HUMAN/1	GKNEDVGA-----									215
SW: PIX1_HUMAN/1	KCPEDSCAGCTCGCGADDPAKTKKQRRORTHTFSQQLQELBATFQNRVPMDSMBREIAVWNTLTPRVRVFRNRRAKWRRENNQQLDCKGCVYPQFSG									218
SW: OTP_MOUSE/1	NPSQACQQ-----									236
	k 4 4R Rt Fc QL eLE F 4 HYPD RE 6A L E R6qVfqnRRRAK544 e4									
SW: PMX1_CHICK/1	-----	320	340	360	380	400	420			245
SW: PMX2_HUMAN/1	-----									253
SW: PMX1_HUMAN/1	-----									245
SW: APX_BRARE/1	LGTLFCTAMFRHFAFISPTFGRFLSSMGLTSAATAALLRQTAPPVSPVQPSAALPEPPSSSSSTAADRASSIAALRLARHSA-QLTQLNILPGSTACKEVC									336
SW: APX_MOUSE/1	LSTFLGAAVFRHFAFISPAFGRFLSTMAPLTSAATAALLRQTAPPVGAVASGALADP									404
SW: AL_DROME/1-1	PPTPSPAGXAXPQLVIGIALTQQAASSLSPT									313
SW: ALX4_MOUSE/1	DFL									354
SW: ALX4_HUMAN/1	DFL									366
SW: RX2_CHICK/1	LPASYTPPPFL									290
SW: RX2_BRARE/1	LQPTYTAHPCFL									297
SW: RX1_XENLA/1	LPASYTPPPFI									296
SW: RX_HUMAN/1-1	LPASYTPPPPPFFL									319
SW: PIX2_BRARE/1	SISSMSMSSSMVPSAVTCVPGSSL									314
SW: PIX2_HUMAN/1	SISSMSMSSSMVPSAVTCVPGSSL									317
SW: PIX1_HUMAN/1	SISSMTMPSSMCPGAVGMPNLSCL									314
SW: OTP_MOUSE/1	SQCSLAACPPPNMCLNSLACSGNAGLQ									325
	S6A LR Ka 2h									

ДВА ДОМЕНА гомеобелков: гомеодомен и OAR домен

[illegible]

- ❖ Часто белки с одинаковой доменной архитектурой имеют сходство последовательностей в границах доменов, а вне них сходство не детектируется
- ❖ Последовательности междоменных участков не имеют признаков гомологии, что происходило в эволюции – неизвестно

Доменная архитектура

[X1WJ92_ACYPI](#) [Acyrtosiphon pisum (Pea aphid)]
Uncharacterized protein (408 residues)



There are 14044 sequences with the following architecture: **Homeodomain, OAR**

Гомеодомен является доменом

There are 25976 sequences with the following architecture: X2JL88 DROME [Drosophila melanogaster (Fruit fly)] Uncharacterized

[Show all sequences with this architecture](#)

There are 2311 sequences with the following architecture

Show all sequences with this architecture.

There are 2108 sequences with the following architecture

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

There are 1903 sequences with the following architecture

 Show all sequences with this architecture.

There are 1836 sequences with the following architecture: **X1WJ92_ACYP1** [*Acyrtosiphon pisum* (Pea aphid)] Uncharacterized p

A diagram of a chromosome with a green oval and a purple rectangle.

мен (зелёный)

Notes (continued)

лен и в негoмo

ВК следует из

як слідує з

АЗИЯ ДОМЕННЬ

ур с гомеодом

Доказательство - выравнивание

G4VR6E_SCHEMA/203-259
 HME2B_DANIRE/173-229
 HMIN_BOHMO/373-429
 T1EES_HELRO/41-97
 HM05_CAEL/36-92
 G13B4X_CRIGR/25-81
 F10RFX_DANIRE/136-192
 B8A59N_DANIRE/135-191
 Q91967_CHKCK/77-133
 DLX38_DANIRE/126-182
 HM43_CAEL/103-155
 HM23_CAEL/212-268
 MSX3_MOUSE/88-144
 HM30_CAEL/96-152
 BARH2_RAT/230-286
 BARH1_DROME/300-356
 BARX2_MOUSE/138-194
 BSH_DROME/275-331
 H2XU6_CIOIN/470-524
 HM11_CAEL/95-151
 SLOU_DROME/546-602
 F6VNQ6_XENTR/112-168
 TIN_DROME/302-358
 NKX25_RAT/138-194
 H0XR12_OTOGA/100-156
 HM09_CAEL/71-127
 H2VEX2_TAKRU/13-69
 U3K517_FICAL/95-115
 TLX3_CHKCK/173-229
 U3JZQ6_FICAL/136-188
 LBX1_MOUSE/126-182
 G4VG64_SCHEMA/38-94
 BCD_DROME/98-153
 BCD_DROME/98-153 (SS)
 VENTX_HUMAN/92-148
 VENT1_XENTR/128-184
 Q8Q49C_XENTR/190-246
 K48821_SOLLC/24-79
 PHO2_YEAST/77-134
 WOX9_ARATH/52-113
 WOX9_ORYS3/11-72
 WOX2_ORYS3/24-85
 WOX4_ARATH/87-148
 WOX1_ARATH/73-134
 WOX2_ARATH/11-72
 WOX5_ORYS3/41-102
 WUS_SOLLC/25-85
 WOX6_ARATH/58-119
 YHP1_YEAST/174-230
 YOK1_YEAST/177-233
 WARA_DICD1/163-219
 PHX1_SCHPO/169-223
 CUT_DROME/1746-1802
 CXU2_MOUSE/1114-1170
 CXU1_MOUSE/1240-1296
 QX281H_CAEL/212-268
 HBX2_DICD1/486-542
 B8Y1_XENTR/234-293

[illegible]

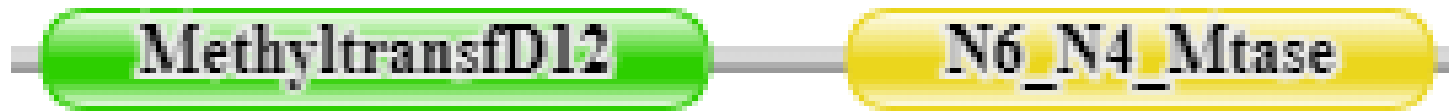
Гомеодомен (зелёный)
представлен и в негомологичных
белках, как следует из
разнообразия доменных
архитектур с гомеодоменом

Как выровнять эти две последовательности?

There are 9 sequences with the following architecture:

MethyltransfD12, N6_N4_Mtase

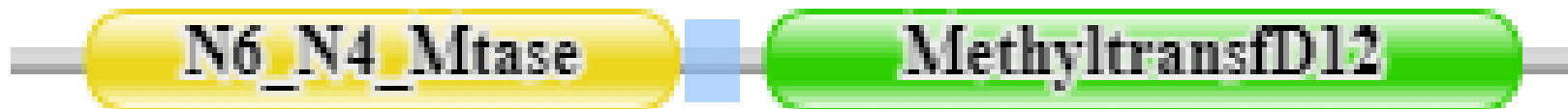
[A0A2Z5QVW5_9MICC](#) [**D12-N6_N4**]



There are 5 sequences with the following architecture:

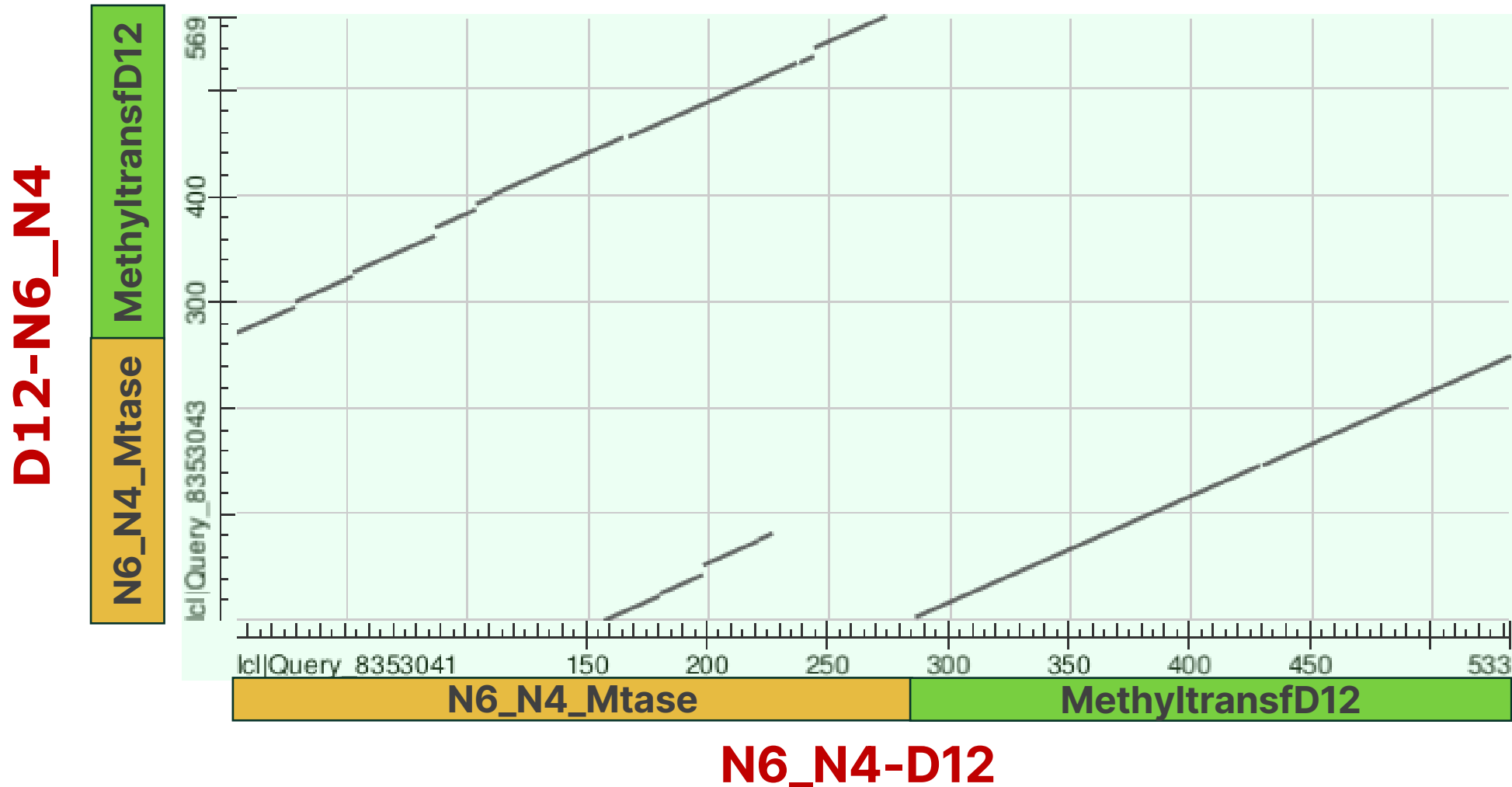
N6_N4_Mtase, MethyltransfD12

[A0A1I7GYG0_9CLOT](#) [**N6_N4-D12**]

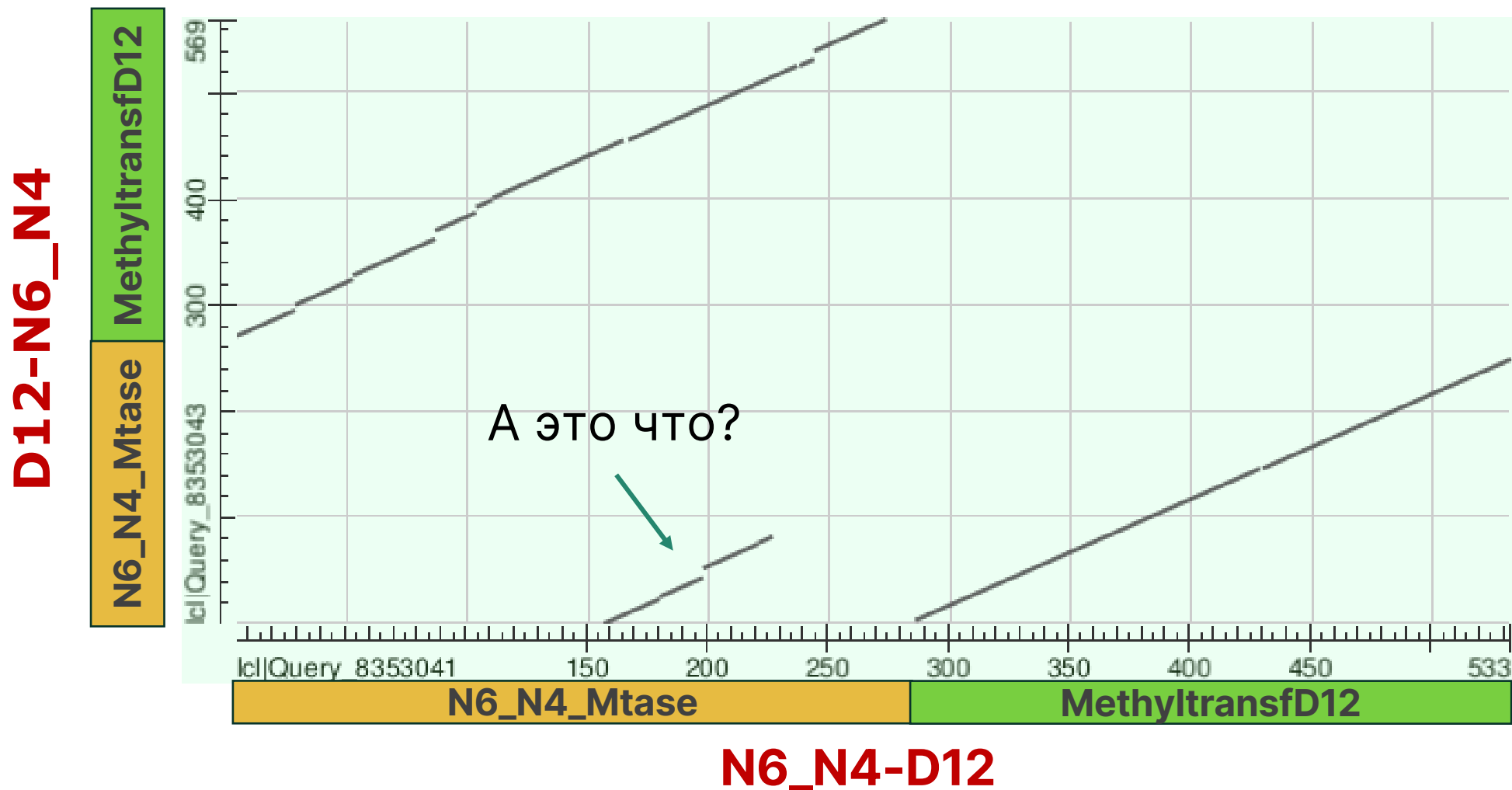


Как такое может возникнуть?

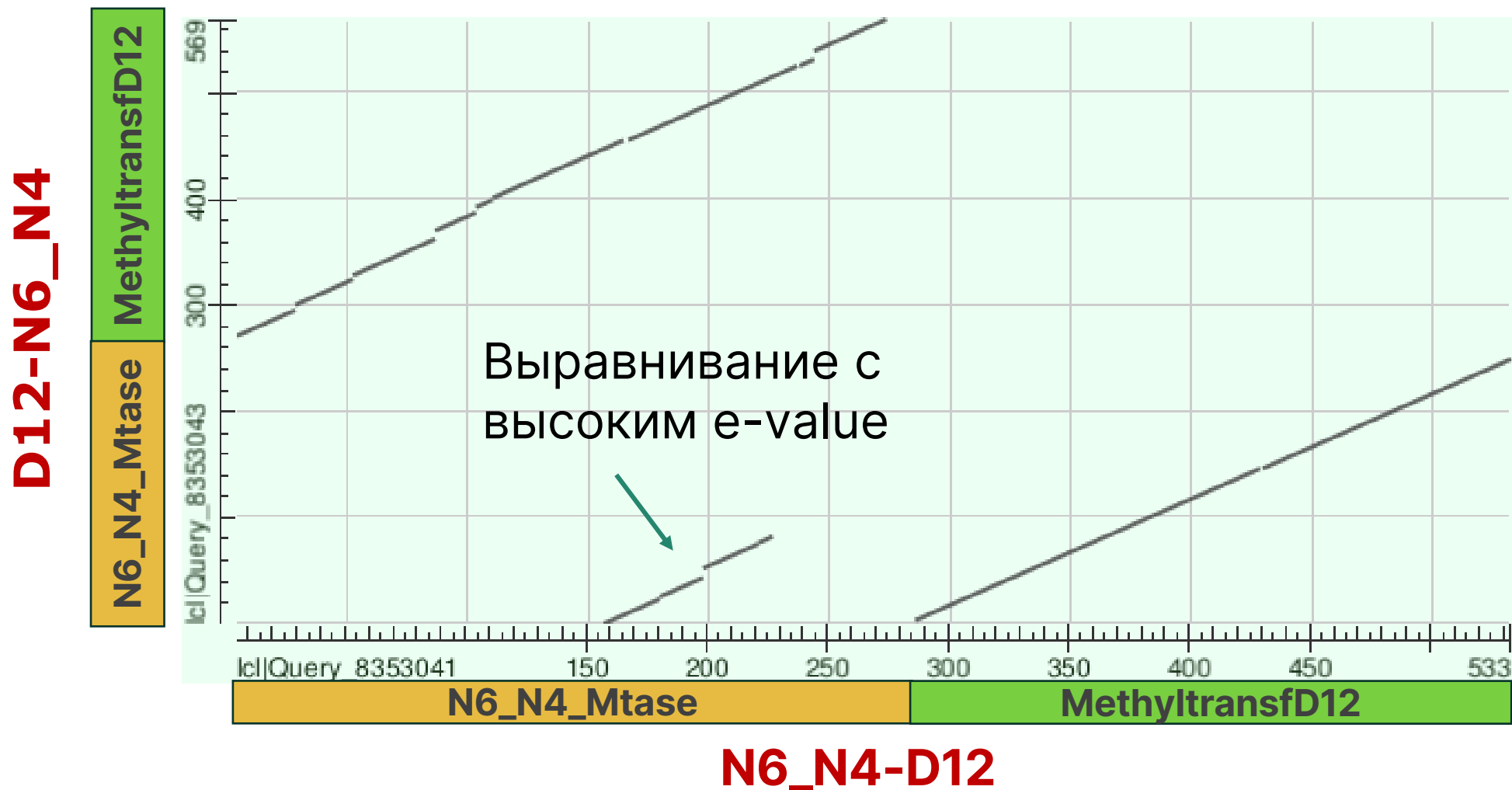
Карта локального сходства – множественные локальные выравнивания



Карта локального сходства – множественные локальные выравнивания



Карта локального сходства – множественные локальные выравнивания



PFAM

- ❖ Заходим на сайт PFAM (InterPro)
- ❖ Находим домен PF00145
- ❖ Что можем узнать о домене из записи?

Заполняем форму для самостоятельной работы
(ссылка на странице семестра)



Спасибо за внимание!