

*Внутридоменные перестройки
последовательности белка: РНК-зависимые
РНК полимеразы.*

Введение

РНК-зависимая РНК-полимераза (RdRs) — фермент, способный синтезировать РНК на матрице РНК. RdRs играют ключевую роль в жизненном цикле РНК-содержащих вирусов. RdRs является одним из самых консервативных вирусных ферментов и состоит из 3 доменов: N-концевого, центрального и С-концевого (рис. 1) [1]. Центральный домен отвечает за полимеразную активность и состоит из 3 субдоменов: *fingers*, *palm* и *thumb* (рис. 1,2).

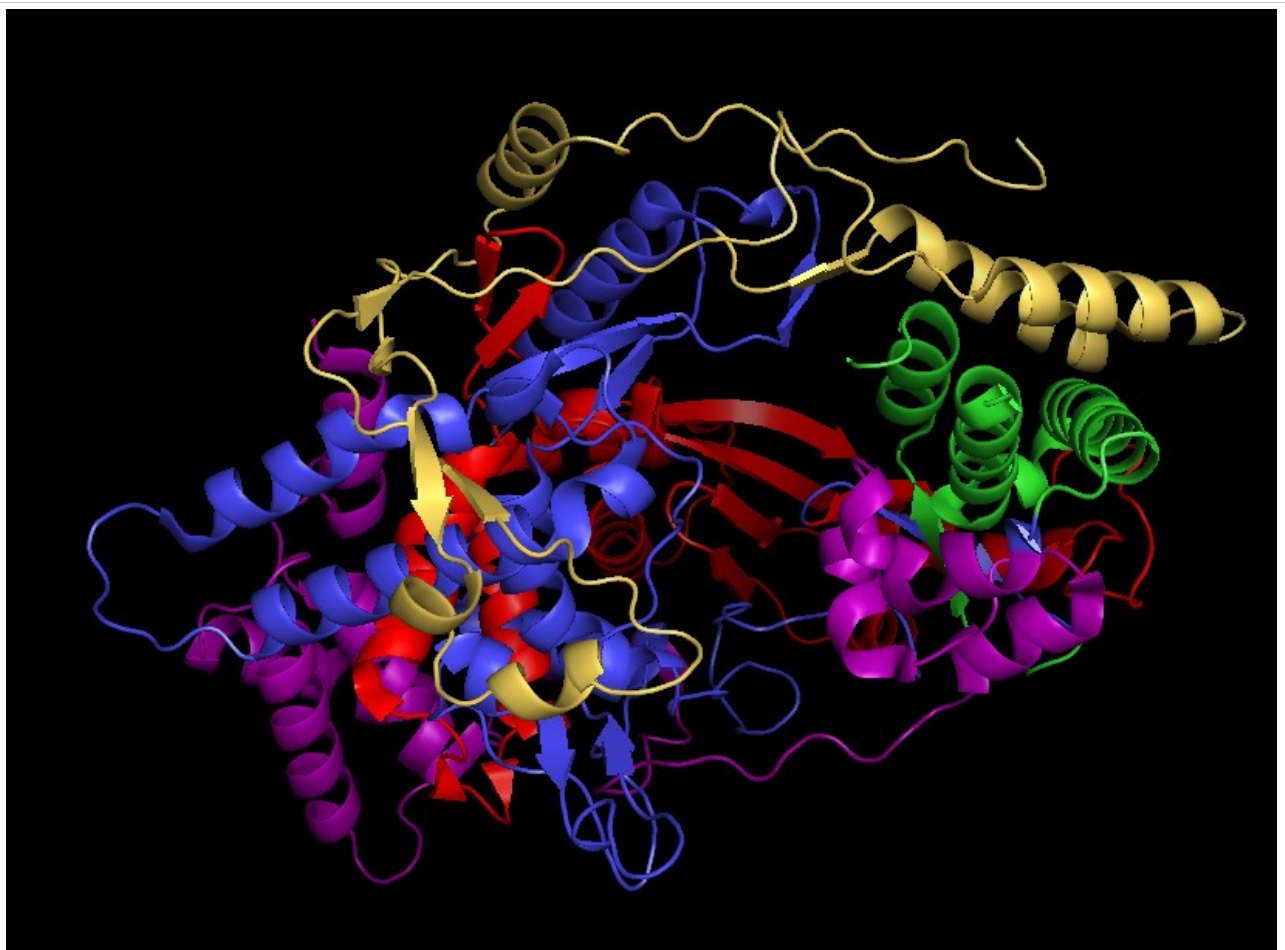


Рисунок 1. РНК-зависимая РНК-полимераза бирнавируса Infectious Bursal DiseaseVirus (IBDV), PDB ID 2PGG, раскрашенная по доменам, согласно [3]: N-концевой - желтый, С-концевой — фиолетовый, центральный домен размечен по субдоменам *fingers* - синий, *palm* — красный и *thumb* — зеленый.

В 2002 году Gorbalenya et al. [2] обнаружили внутридоменную перестройку внутри субдомена *palm*. В субдомene *palm* выделяют 3 консервативных мотива, расположенных в порядке А-> В->С в направлении от N- к С-концу. Мотивы А и С участвуют в связывании ионов Mg^{2+} и Mn^{2+} , необходимых для катализа, а мотив В отвечает за различие между нуклеотидами и дезоксикулеотидами, т. е. определяет, что будет синтезироваться, ДНК или РНК. Однако, у вирусов семейства *Birnaviridae* и некоторых вирусов семейства *Tetraviridae*, вместо канонической последовательности А-> В->С эти мотивы расположены в порядке С->А-> В (рис. 3). Удивительно, что перестановка ~ 22 остатков мотива С примерно на 100 аминокислотных остатков ближе к N-концу, в upstream мотива А, не приводит в потере активности. Авторы предположили, что такая перестановка практически не изменяет

третичную структуру. Пример того, как при изменении последовательности, т. е. при изменении топологии может сохраняться третичная структура, представлен на рис. 4 а и б.

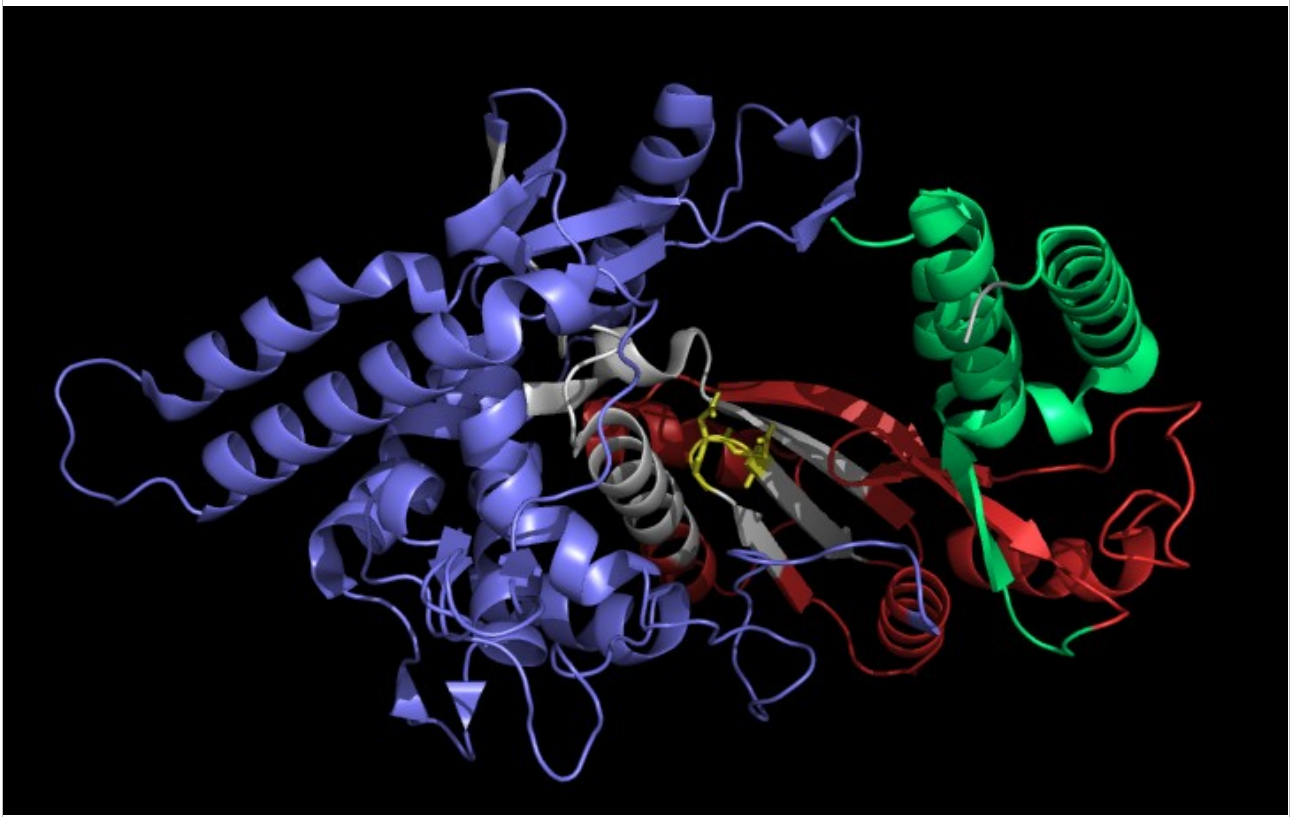


Рисунок 2. Центральный домен RdRs бирnavируса IBDV, PDB ID 2PGG, размечен по субдоменам, согласно [3]: *fingers* - синий, *palm* — красный и *thumb* — зеленый. Белым показаны мотивы А, В и С, желтым — консервативная триада -ADN- мотива С (у большинства вирусов, в том числе у тетравирусов, -GDD-), которая участвует в связывании необходимых для катализа ионов металлов.

Как видно из рисунка, разрывы происходят в трех точках полипептидной цепи, соответствующих петлям, а элементы вторичной структуры сохраняют свою целостность. Такая циклическая перестановка могла произойти в результате одновременной дупликации

Motifs	--G----	--F----	-----A-----	-----B-----	-----C-----	----
PV3L	TSAGYFY39KDELRSKTKVE	QGKSLI32AVGCDPDL-FWSKIPVLM	-----	0--EKLFAYTYGYDA44GGMPSCSG	SIFNSMINLI15DHLKMIAYGD	VIASYPHE33VT
HpeV1	TSAGYSF48KDELRLKDKIA-SGKTFCI32AVGINPYK-DWFHMINALN	-----	0--DYNYEMYSQYDG46GGMPSSPC	TVLNSLCLNLI13-DCLPIVYGD	VILLSLDKE34VE	
InFV	TSAGYPL49KDELRPSEKLRRFGCTVF31AIGLNPNSGDWRRLRDTLLE	---KG-	0--PCLLQMYKNSD54NGVLACHFPM	SVVNSVNLII19-LTYIIVMGD	VVISLQPK40FE	
DCV	TSAGFFY56KDERRDIKVN-VGKTAVF31AVGTNVYSSDWERIAKRLKT	---KG-	0--SHVIAGDFGNFDG59HSQPSGNPF	VVIINCLYNSI24EYVALITYGD	NVLNIDAK38IF	
PYFV	TSECYPL56KDERRALDKIYEKPKTRLF31MVGINPHSREWSDLAISLAS	---FS-	0--PYGFNGDFANFDG48GGGSPCFM	VVIFNSFTINLF25KYVRACVYGD	NIVAIKME41FD	
RTSV	TSAGYFY61KDERRKLAKIYEKPATR31QVGINPESMEWSDLINSLLR	---VN-	0--TKGFAGYSKFDG50QGMPSCFAM	VVIFNSFVNYI25YCKIVAYGD	NVVSVHE40MS	
CPSMV	TSECFFH55KDELPLRKIYEKPKTRCF31QVGINPYGMEWSRLAMSLE	---KG-	0--NNILCCYSFDFG50CGIPSCFEL	VVICNSIFNEL24TFVKMVTYGD	NLLSVQSA39CD	
TobRV	TSECYFY57KDELLVERKVLQKLGTRNF31QVGINVQGREWGLLMDRLAA	---KN-	066SVAYNCYSKFDG64GGIRRG-AL	IVLLNSVFNEI19TFVTVLVYGD	NLIAVDP39LD	
FCVF9	TSCY--PY44KDELRPVEKVS-EGKRPMI31QVINMDSPPSVEALFQRIKS	---A-	066AKVFAVYSKWD547SGLPSMPL	SVINSINHCL22STFFIMTYGD	GVYMPFIM36VV	
LORDV	TSSGHPH43KDELVKTKIYKGIKRL31RVGMNNE-DGPIIFERHSR	---Y-	066KYHYDAYSRWDS47EGLPSGVPC	SQWNSIAHWLI18ANSLSFSGD	EIVSTDIN35LT	
APV	TSAG--PY63KVENIEKEKAQ-KGKVRIL31AIGQNPYT-TSTIWHRFST	---KQ-	066GKIINTKFAFDK52NGNESGTFV	TLLNCVSVHI27SRSELAAILGD	KTQVSKD31IN	
TVMV	AAVC--AL41KAELRPFEKLI-ENKTRVF31SVGMTKFGGWNLGKLP	---DG-	066WVYCDAYSQFDS51KGNNSQPS	VVDNMLVLL18-VCKFFANGD	LIIAISPE30LW	
BaMMV	TSGG--EL42KAELRSSQKIL-ERKTRVF31TVGINFNGWQNLKLNK	---PG-	066WTHGSGSRFDS49CGNNSQPS	VVDNMLLMT19ENFVFCNGD	NKFAMSP31NP	
HCoV	KSAGWPL39KYAISGKERA-----	TV30VIGTTKFGWQNLKLNLMADVDDP	0--KLMGWYPKCDR53GGTSSDAT	AYANSVFNI53KHFSSMILSD	SVVCYKKT40HE	
BEV	SSMFINI38KFAISAKARA-----	TV33LIGVSKYGLKFKSKFLKDKYGAIEGF	0--DVPFSGYTKCDR49GGTSSDAT	AHSTTFYNYM46NTYFLNFLSD	SFIFSKPE34EE	
GAV	KSSGIGI39KISIQPVDKAL-----	SI31LIGFKDTHCGINKLINGIKAGFNPK	0GKSKWISQYKFDT52HGVSSGDGA	AIKNSHCNSV49NISRCATLSD	TLIAINPD31YE	
EAV	SQAGFNG35KRCYCSKYI-----	SI32YLKSKFDPIPADK-----	0--YCLETLLESCDR46GGLSSCDPI	SISNTIYSLV33RYVRVYYSDD	VVLTTPN33-S	
	ts G p Ke * k r f g * * *		D d g sG T * nt* *		ygDd *	
TaV	ASAGPY46KTDREVESL--GEKTRPY33AGFSWINGGAEDLAIWARQAGEAGN	20GKLYRICPFKQMDG58DGLMTGVVG	TLEDTVKSAI17KPPRIACVGD	TDIIVYRKDE38QK		
EeV	ASAGPY46KTDREVEKEL--GEKTRPY33AGFSWINGGAEDLVIWARGAGEVGR	20GRLYRICPFKQMDG58DGLMTGVVG	TLEDTVKSAI17EPPRIACVGD	TDIIVYRQKK38QK		
IPNVJ	SSAGLY54KGERYQVDW--DKKTRNI33LGGSPFHGMDRIMTIIRD	S--LDN18N--TWYSILEKGEA58YQGS	ENAFFLNNHMLTI21DEDLVMI	AIN6--YILQDE49LD		
IBDV	SSSGLPY59KAERYDKSTW--LTKTRNI34LKKNPFRCGLNRIWEI	----LAE18N--TWYSILEKGEA57YQGM	ENAFFINNHLLTTL21EPPKALV	AIN6--YIVHSE46LD		

Рисунок 3. Выравнивание субдоменов *palm* для канонических полимераз пикорнавирусов (сверху) и для полимераз вирусов, у которых обнаружена перестановка (*Thosea asigna* virus (TaV), *Euprosterina elaeasa* virus (EeV), infectious bursal disease virus (IBDV), infectious pancreatic necrosis virus strain Jasper (IPNVJ)). Для удобства восприятия стрелкой показана обратная перестановка мотива С [2].

фрагмента А-В-С, делеции N-концевого фрагмента АВ и С-концевого фрагмента С (Рис. 5). Промежуточные варианты ABCАВС, ABCАВ или САВС, скорее всего, будут неактивны, поэтому эти три достаточно маловероятных события должны произойти одновременно.

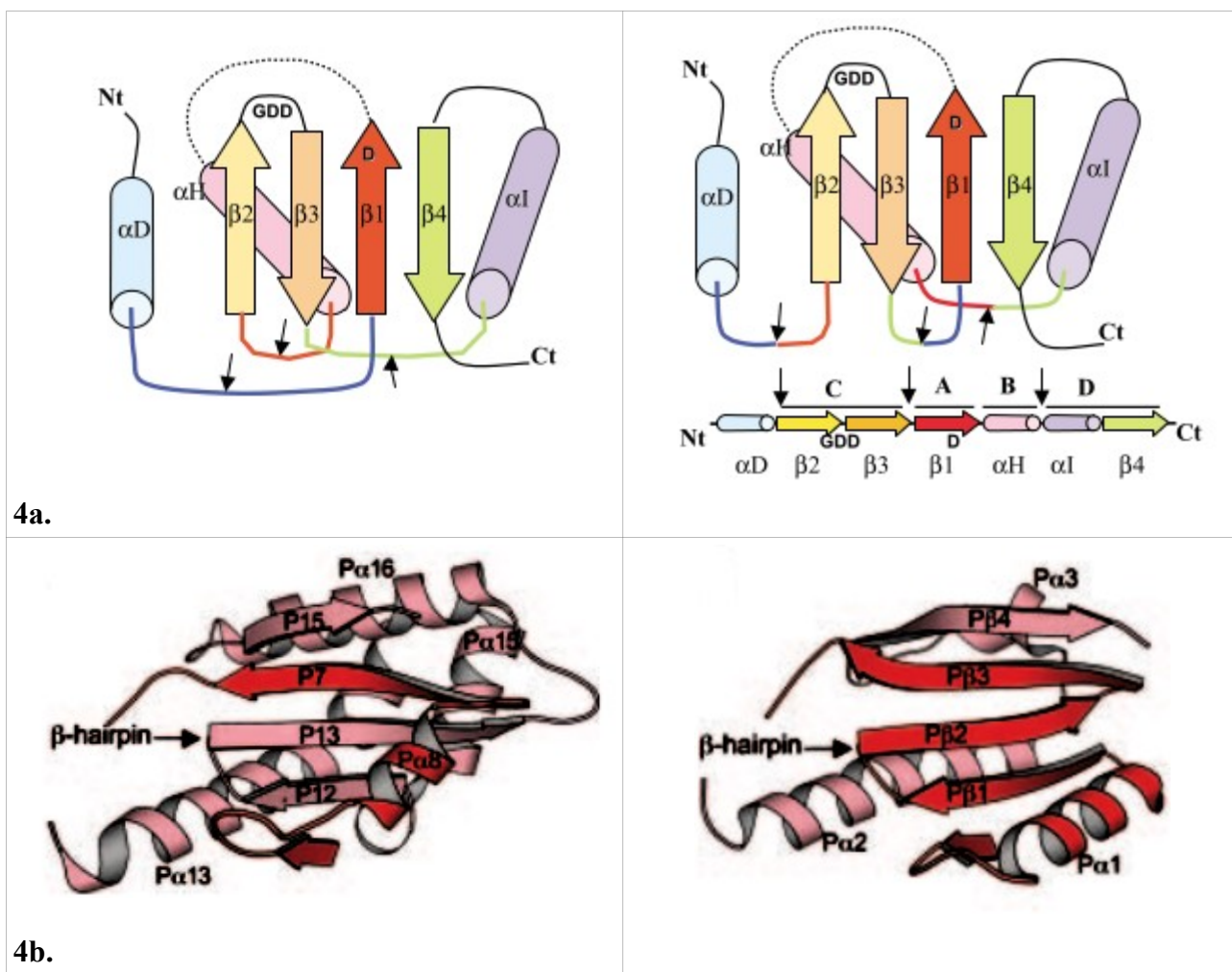


Рисунок 4а. Схема организации третичной структуры канонического (слева) и неканонического (справа) субдоменов. Места разрывов цепи показаны стрелками. Как видно, они приходятся на петли и повороты и не затрагивают элементы вторичной структуры [2]. Элементы вторичной структуры размечены согласно канонической RdRp 1RDR из банка PDB.

Рисунок 4б. Субдомены *palm* реовируса (канонический, слева) и IBDV (неканонический, справа) [3].

Казалось бы, такое редкое событие, вероятнее всего, произошло однажды, у общего предка *Tetraviridae* и *Birnaviridae*. Но *Tetraviridae* относятся к ss+ РНК-содержащим вирусам (т. е. в вирусной частице одноцепочная +РНК), а *Birnaviridae* — к ds, т. е. к двуцепочным (рис. 6). Получается, что:

(1) В ходе эволюции организация наследственного аппарата некоторых групп вирусов изменялась несколько раз. Сначала у общего предка тетра- и бирнавирусов возникла внутридоменная перестановка в RdRp, затем организация наследственного аппарата поменялась с ds на ss+ или наоборот. В этом случае необходимо пересмотреть текущую классификацию вирусов.

(2) Либо такая перестановка все-таки возникала неоднократно, тогда нужно оценить вероятность подобного события и определить вес циклической перестановки при подсчете веса выравнивания.

В 2002 году, когда перестановка была впервые обнаружена, Gorbalenya et al. склонялись к первому варианту, но в 2009 году Sabanadzovich et al. [5] обнаружили точно такую же перестановку в RdRp вируса виноградной лозы *Grapevine virus Q* (GVQ) и его близкого гомолога *Grapevine Syrah virus 1* (GSV1).

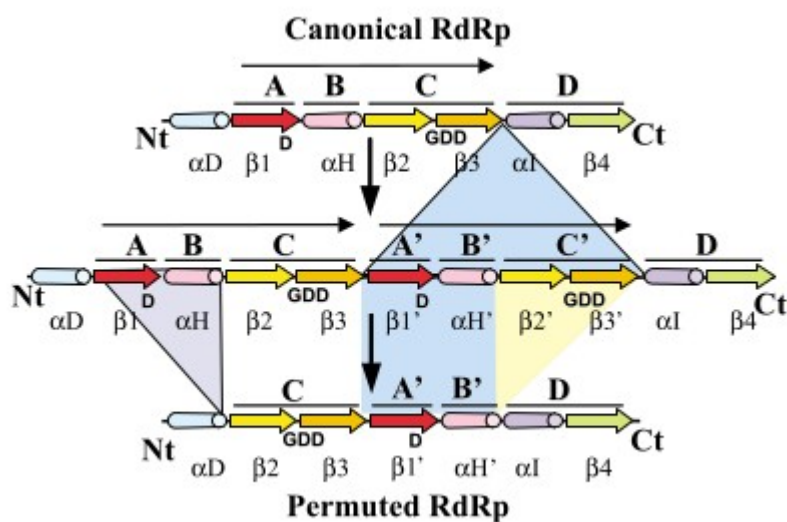


Рисунок 5. Предположительный механизм возникновения циклической перестановки [2].

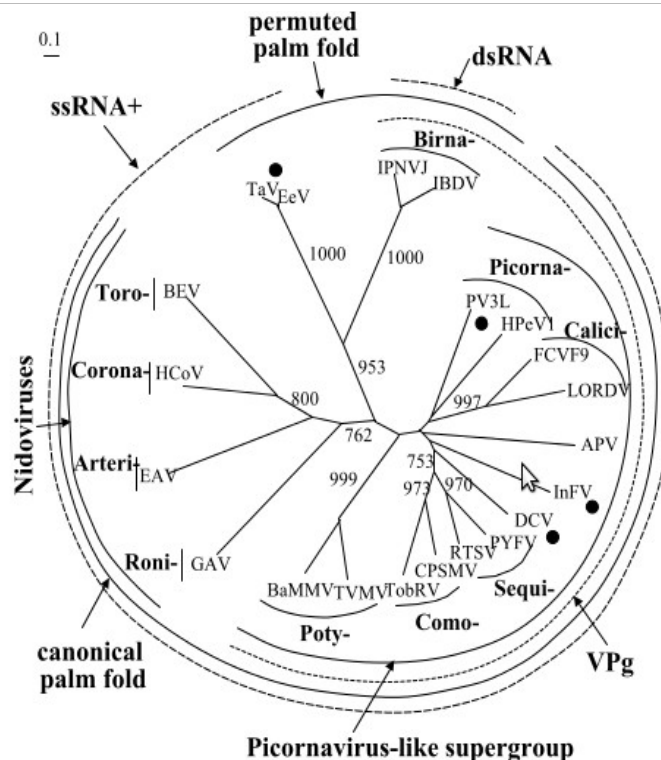


Рисунок 6. Филогенетическое дерево, построенное на основе выравнивания на рис.3 методом neighbour-joining [2]. Видно, что Tav/EeV и IBDV/IPNVJ образуют обособленные кластеры.

В данном случае, целый ряд фактов указывает на то, что эта перестановка возникла второй раз независимо:

- Известно много очень близких гомологов RdRp внутри семейства *Tymoviridae*, к которому принадлежат GVQ и GSV1. Если поискать с помощью PSI-BLAST гомологов последовательности, соответствующей субдомену *palm* из GVO, то будут обнаружены последовательности, подобные изображенной на рис. 7. Если сделать аналогичный поиск для TaV или EeV, то будут найдены TaV, EeV, их близкий гомолог из семейства *Tetravirida Drosophila melanogaster tetravirus SW-2009a* (DtV) и еще один вирус, к *Tetraviridae* не относящиеся - *Drosophila A virus* (DAV).

```
> ref|YP\_224293.1  RNA-dependent RNA polymerase [Citrus sudden death-associated virus]
Length=602

GENE ID: 3342435_CSDaVgp1 | replicase-associated polyprotein
[Citrus sudden death-associated virus] (10 or fewer PubMed links)

Score = 522 bits (1344), Expect = 2e-178, Method: Compositional matrix adjust.
Identities = 255/398 (64%), Positives = 295/398 (74%), Gaps = 21/398 (5%)

Query 12  FNPERRDDVFVEAAVVCVCGDSSNAPQVSTHFLPETRRPLHFDLPSAKPEFAAHEAPAPLT 71
          P+ + DV + A +V G G NAPQVS+HFLPETRRPLH+D+PSA PE A T
Sbjct 18  LKPDFQGDVVISAPIVLGSGELNAPQVSSHFLPETRRPLHWDIPSAIPESATRPDSTEPT 77

Query 72  DTFIEPVYPGETFENIAAHFLPAHDPEVKEILFKDQRSNQFPFIDQPFHVGAQPASLCAA 131
          + EPVYPGETFEN+AAHFLPAHD +EI ++ Q SNQFP +D+ FH+ AQP SL AA
Sbjct 78  TSHPEPVYPGETFENLAAHFLPAHDPTDREIYWQQLSNQFPHMDKEFHLLAAQPMSSLAA 137

Query 132 VHHSKKDPTLLAASIEKRLRFRASDAPYQITAKDEILGSILFEAHCRAMRRDPNVRVPFD 191
          VH K+DPTLL ASI+KRLRFR SD PYQIT KDEILG +LFE CRA R P FD
Sbjct 138 VHQEKGQDPTLLPASIQKRLRFRPSDKPYQITPKDEILGQLLFEGLCRAYHRSPFHTAETF 197

Query 192 EALFAECIALNEFAQLTSKTQAVIMANHERSDPDWRYTAVRIFAKNQHKVNSGSLFGPWK 251
          LFAECI LNEFAQL+SKTQA IM N RSDPDWR++AVRIF+K QHKVN GS+F WK
Sbjct 198 PVLFAECINLNEFAQLSSKTQATIMGNARRSDPDWRWSAVRIFSKTQHKVNEGSIFRSWK 257

Query 252 ACQTLALMHDVAVILLFGPVKKYQLIHDERDRPEHIFIYAGRTPQEMSEWCQKFLTPRSAS 311
          ACQTLALMHDVAV+L+ GPKVKKYQ + D+RDRP H++I+AG TP +MS WCQ+ LT
Sbjct 258 ACQTLALMHDVAVLILGPKVKKYQVFDQRDRPRHLYIHAGNTPSQMSNWCQQLHTTAVK- 316

Query 312 SPVPVMVSGDDSLIGCHPHFVANDYTAFDQSQHGEEAVLERLKMERNIPEWLITLHIMI 371
          +ANDYTAFDQSQHGEEA VLER KMER++IP+ LI LHI +
Sbjct 317 -----LANDYTAFDQSQHGEEAVLERKKMERLSIPQALIDLHIL 356

Query 372 KTHITTQFGPLTCMRLTGEPGTYDNSDYNLAVIFLEY 409
          KTH++TQFGPLTCMRLTGEPGTY DNSDYNLAV+ EY
Sbjct 357 KTHVSTQFGPLTCMRLTGEPGTYDNSDYNLAVVNCEY 394
```

Рисунок 7. Фрагмент полипротеина из *Citrus sudden death-associated virus*, близкий гомолог (64% identity) GVQ найденный с помощью BLAST. В этой RdRp нет перестановки, поэтому в выравнивании напротив мотивов С будут длинные серии гэпов. Однако, несмотря на это, процент identity достаточно большой и такое выравнивание будет иметь большой score.

- GVQ и GSV1 — ss+ РНК-содержащие вирусы растений, а бирнавирусы и тетравирусы заражают животных
- РНК-полимераза *Tymoviridae* синтезируется в составе полипротеинового комплекса, содержащего также хеликазу, метилтрансферазу и энодопептидазу
- РНК-полимераза *Tymoviridae*, в отличие от полимераз тетра- и бирнавирусов не способна к белковому праймингу

Получается, что подобные перестройки случаются гораздо чаще, чем предполагалось изначально. Современные алгоритмы построения выравниваний такие перестановки никак не учитывают. Gorbalenya et al. предложили следующий алгоритм поиска перестановок [2]:

- 1). По 58 последовательностям RdRp из 13 семейств ss+ РНК-содержащих вирусов было построено множественное выравнивание.
- 2). По этому выравниванию строится НММ-профиль.
- 3). Подготавливается база данных для поиска с помощью профиля. Для этого в исходной последовательности делаются всевозможные обратные перестановки. Выбираются 3 числа i , j и L , такие что:

$$1 \leq i < j < L \leq N+1,$$

i — номер позиции начала перестановки

j — номер позиции конца перестановки

L — номер позиции вставки

N — количество аминокислотных остатков в последовательности

- 4). Проводится поиск с помощью профиля и строятся графики зависимости отношения E_0/E_s от i и от L , где E_0 — E-value исходной последовательности, E_s — E-value последовательности с обратной перестановкой.

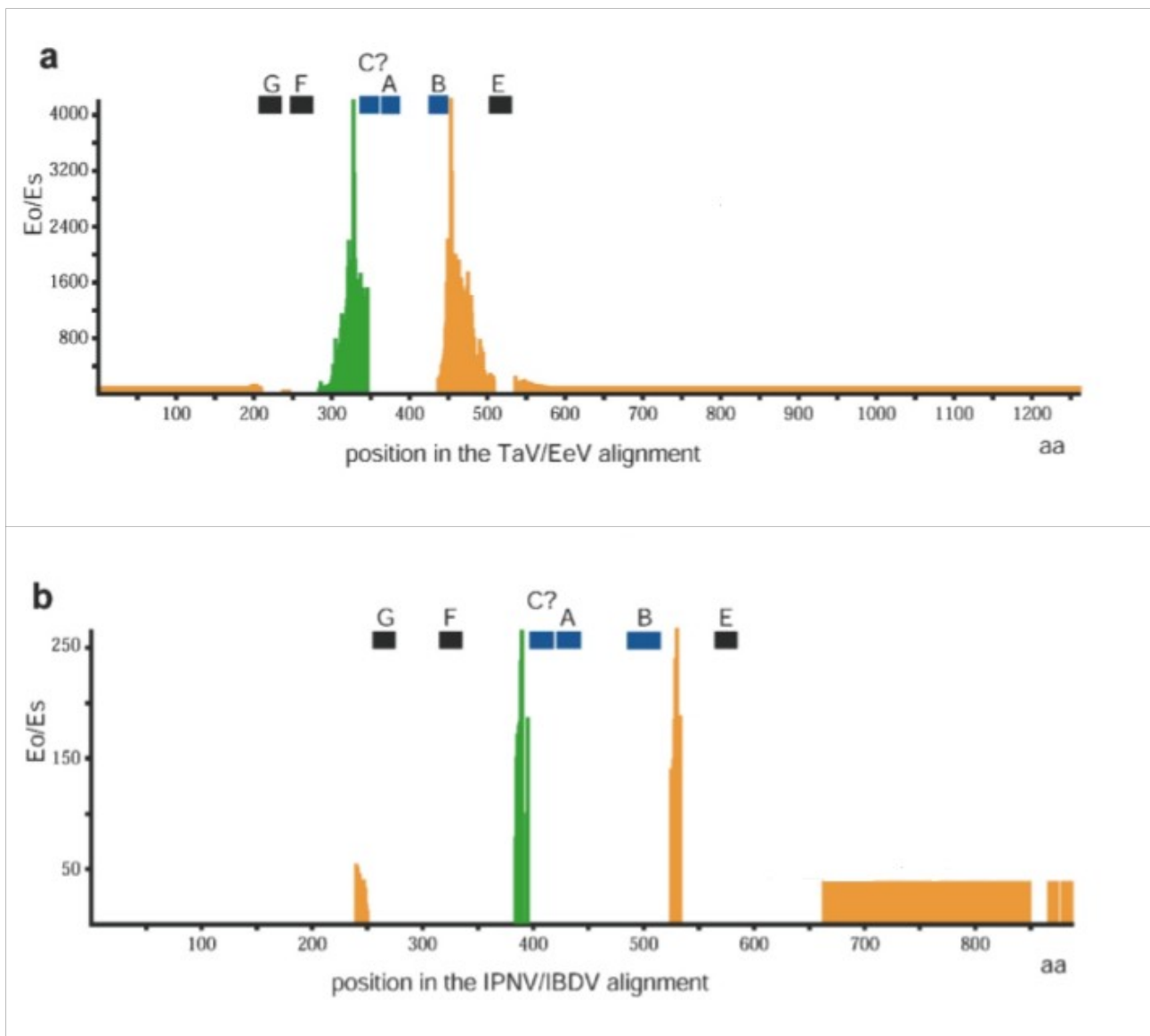


Рисунок 8. Поиск по базам данных, построенных по последовательностям TaV/EeV и IBDV/IPNVJ. Зеленые пики соответствуют распределению $E_0/E_s(i)$, а оранжевые - $E_0/E_s(L)$.

Алгоритм был проверен на последовательностях полимераз TaV, EeV, IBDV и IPNVJ. Максимумы пиков соответствуют позициям *i* и *L* перестановки (рис. 8). Размер базы для поиска перестановок внутри последовательности длиной 1000 а.о. составляет $\sim 10^8$ последовательностей, поэтому, чтобы уменьшить размер базы данных рассматривались все перестановки размером 9 а.о., что не повлияло на результат.

В заключение нужно сказать, что сегодня выделяют новую группу *Permutotetraviridae* [4], включающую в себя TaV, EeV и DtV. Они отличаются от *Tetraviridae* не только наличием перестановки в RdRp, но и особенностями устройства генома [2].

Поиск неканонических RdRp среди вирусных белков

Приведенный протокол позволяет найти любую перестановку в выбранной последовательности RdRp. Однако, он не подходит для поиска неканонических полимераз среди большого количества вирусных белков, т. к. даже при переборе всех перестановок длиной 9 внутри 1 последовательности длиной ~ 1000 , размер базы данных составляет $\sim 10^5$. Для этого удобно будет создать профиль, соответствующий именно неканоническим полимеразам. (Обычный поиск BLAST тоже не подходит, т. к. находятся только самые близкие гомологи. Например, по последовательности IBDV можно найти все RdRs бирнавирусов, но не получается найти GVQ. Либо, наоборот, находятся очень близкие последовательности, в основном отличающиеся только отсутствием перестановки.)

На сегодняшний день известны 2 неканонические полимеразы в семействе *Tymoviridae* (GVQ и GSV1), 3 — в семействе *Permutotetraviridae* (TaV, EeV и DtV). Кроме того, предположительно все RdRp из семейства *Birnaviridae* тоже являются неканоническими (поиск BLAST белков, гомологичных канонической полимеразе *Human poliovirus 1* (PDB ID 3OL7) среди белков *Birnaviridae* значимых находок не дал).

По последовательностям, соответствующим доменам *palm* из полимераз вирусов TaV, EeV, DtV, DAV, IBDV, IPNV, BshV (*Blotched snakehead virus*, *Birnaviridae*) и GVQ с помощью сервиса COBALT (<http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi>) было построено множественное выравнивание (рис. 9).

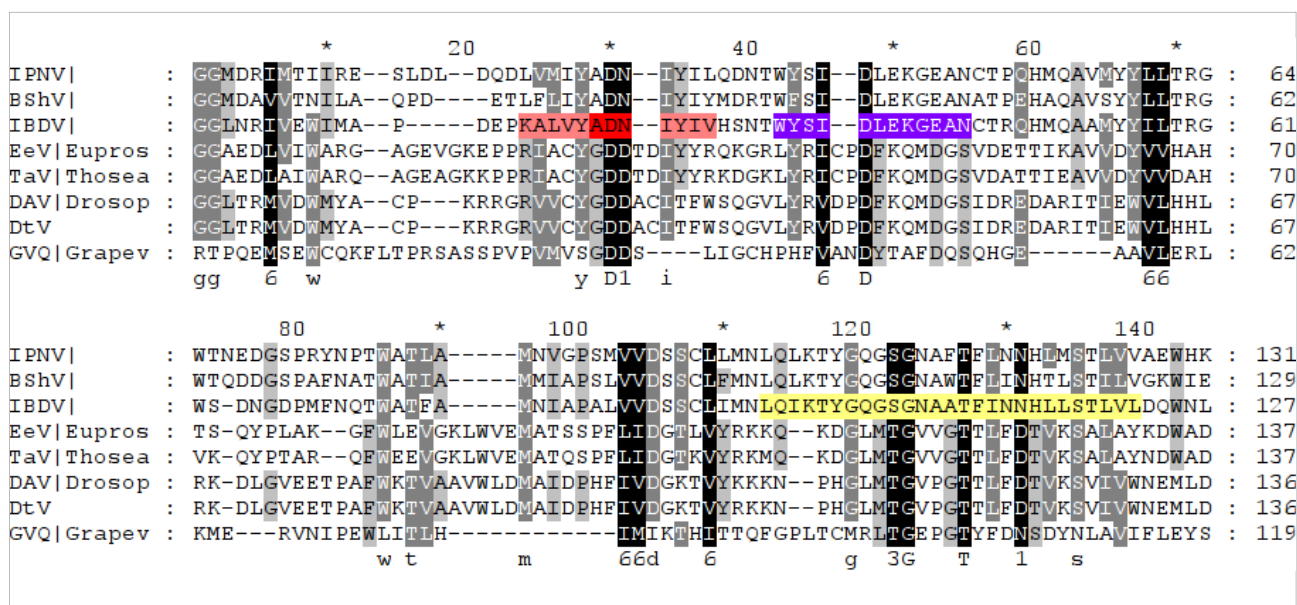


Рисунок 9. Множественное выравнивание последовательностей, соответствующих доменам *palm* из полимераз вирусов TaV, EeV, DtV, DAV, IBDV, IPNV, BshV (*Blotched snakehead virus, Birnaviridae*) и GVQ построенное с помощью сервиса COBALT.

По этому выравниванию с помощью rftools был построен профиль.

Для наиболее консервативных позиций в мотивах А (-DX₄₋₅[DE]-), В (-GX₂₋₃TX₃[DN]-) и С (-[GA]D[DN]-) веса были скорректированы вручную.

Чтобы оценить эффективность полученного профиля, проводился поиск по базе, содержащей последовательности всех вирусных полимераз, взятых из RNA Virus Database (<http://tree.bio.ed.ac.uk/mavirusdb/index.php>) и полимераз, использованных для построения профиля (т.к. не все они присутствовали в этой базе). Результаты поиска представлены в таблице 1.

Score	длина, а.о.	организм	Семейство	Перестановка
23.120	136	TaV	Tetraviridae	есть
22.620	136	EeV	Tetraviridae	есть
22.520	135	DtV	Tetraviridae	есть
22.450	135	DAV	Unclassified	есть
21.330	130	IPNV	Bimaviridae	есть
20.910	126	IBDV	Bimaviridae	есть
20.630	130	Marine birnavirus	Bimaviridae	есть
20.460	130	Paralichthys olivaceus birnavirus	Bimaviridae	есть
20.290	128	BshV	Bimaviridae	есть
10.940	118	GVQ	Bimaviridae	есть
10.230	110	Drosophila x virus	Bimaviridae	есть
1.290	37	Broad bean wilt virus	Secoviridae	нет
1.250	37	Human rotavirus G3	Reoviridae	нет
1.220	32	Rotavirus C - structural protein	Reoviridae	нет
1.150	35	Cocksfoot mild mosaic virus	Sobemovirus	нет
...

Таблица 1. Результаты поиска внутримономерных перестроек с помощью профиля среди последовательностей вирусных полимераз. Серым выделены полимеразы, у которых обнаружена внутримономерная перестройка.

Далее, поиск проводился по всем вирусным белкам из баз данных SwissProt/Uniprot и trEMBL с порогом 2. Результаты поиска представлены в таблице 2.

Анализ пространственных структур субдоменов *palm* канонической и неканонической RdRp

1).

Выравнивание структур центральных доменов канонической (3OL7:A, *Picornaviridae; Human enterovirus C*) и неканонической полимеразы (2PGG, *Birnaviridae; Infectious bursal disease virus*) было выполнено в Rymol с помощью команд align и super. Результат выравнивания представлен на рисунке 10.

Как видно из рисунка, первое выравнивание не очень хорошее, его RMSD составляет 15.84 Å. Второе выравнивание гораздо лучше - RMSD 3.88 Å. Это, вероятно, связано с тем,

Score	длина, а.о.	ID	организм	Семейство	Перестановка
23.120	136	Q6A562_9VIRU	<i>Thosea asigna virus</i>	Tetraviridae	есть
22.620	136	Q8QYM6_9VIRU	<i>Euprostepma elaeasa virus</i>	Tetraviridae	есть
22.520	135	D0U497_9VIRU	<i>Drosophila melanogaster tetravir</i>	Tetraviridae	есть
22.450	135	C6FGJ5_9VIRU	<i>Drosophila A virus</i>	Unclassified	есть
21.330	130	Q8JJE9_9VIRU	<i>Infectious pancreatic necrosis vir</i>	Bimaviridae	есть
20.920	126	Q8B540_IBDV	<i>Avian infectious bursal disease v</i>	Bimaviridae	есть
20.830	130	B2ZJ00_9VIRU	<i>Aquabimavirus TAB98</i>	Bimaviridae	есть
20.640	130	Q8BAD0_9VIRU	<i>Marine bimavirus</i>	Bimaviridae	есть
20.630	130	Q8BAD4_9VIRU	<i>Yellowtail ascites virus</i>	Bimaviridae	есть
20.460	130	A8IFB1_9VIRU	<i>Paralichthys olivaceus bimavirus</i>	Bimaviridae	есть
20.290	128	RDRP_BSNV_Q8	<i>Blotched snakehead virus (BSNV</i>	Bimaviridae	есть
13.250	129	RDRP_DXV96_Q	<i>Drosophila x virus (isolate Chung</i>	Bimaviridae	есть
12.950	129	G9I7F8_9VIRU	<i>Espirito Santo virus</i>	Bimaviridae	есть
12.160	74	H2EIJ1_9VIRU	<i>Aquabimavirus Garra/F73-11/IRL</i>	Bimaviridae	есть
10.940	118	C8YZ75_9VIRU	<i>Grapevine virus Q</i>	Tymoviridae	есть
10.560	118	C1J9J7_9VIRU	<i>Grapevine Syrah virus 1</i>	Tymoviridae	есть
10.230	131	Q2PBR4_9VIRU	<i>Tellina virus 1</i>	Unclassified	есть, bima-like
10.200	144	D0U496_9VIRU	<i>Drosophila melanogaster bimavin</i>	Bimaviridae	есть
7.920	139	D3GMG3_9VIRU	<i>Rotifer bimavirus strain Palavas</i>	Bimaviridae	есть
3.800	77	Q0ZAR7_9VIRU	<i>Nemesia ring necrosis virus</i>	Tymoviridae	нет
3.770	133	Q6AW71_9VIRU	<i>Bombyx mori Macula-like virus</i>	Tymoviridae	нет
3.730	116	B7U7P6_9VIRU	<i>Grapevine fleck virus</i>	Tymoviridae	нет
...

Таблица 2. Поиск по всем вирусным белкам из баз данных SwissProt/Uniprot и trEMBL с порогом 2. Серым выделены полимеразы, в которых есть перестановка. Жирным выделены полимеразы, по которым строился паттерн.

что align берет за основу выравнивание последовательностей, построенное без учета вторичной структуры. Алгоритм super строит пространственное выравнивание, никак не учитывая выравнивание последовательностей.

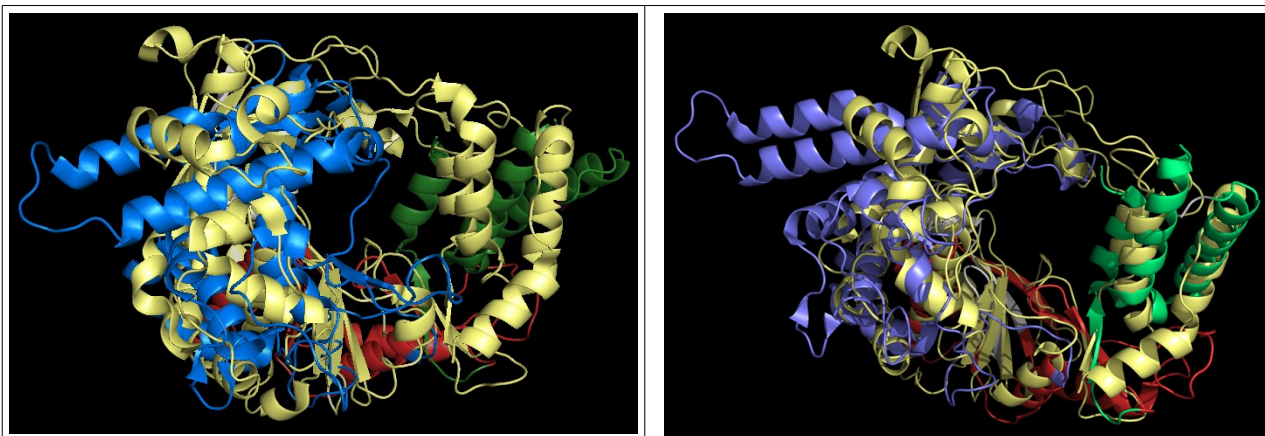


Рисунок 10. Выравнивание структур центральных доменов канонической (3OL7:A, Picornaviridae; Human enterovirus C) и неканонической полимераз (2PGG, Birnaviridae; Infectious bursal disease virus) с помощью команд align (слева) и super (справа).

2).

Поиск в базе данных PDB структур, похожих на неканоническую полимеразу 2PGG проводился с помощью алгоритма SSM, реализованного в сервисе PDBeFold (<http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver>).

Алгоритм запускался со следующими параметрами:

PDB	2PGG
Chains	A
Lowest acceptable match (%)	30
Source	Whole PDB archive
match individual chains	yes
match connectivity*	
best matches only	yes
unique matches only	yes
if no matches within limits of acceptability are found, show close ones	yes
Precision	normal
Sort by	RMSD

* результат поиска почему-то не зависел от того, стояла ли галочка в чекбоксе

Таблица 3. Параметры запуска поиска структур в PDBeFold.

Всего было обнаружено 348 находок, большинство из которых были полимеразами IBDV, IPNV или полиовирусов. Среди первых 4 уникальных находок было 2 канонические полимеразы и 2 неканонические (Таблица 4).

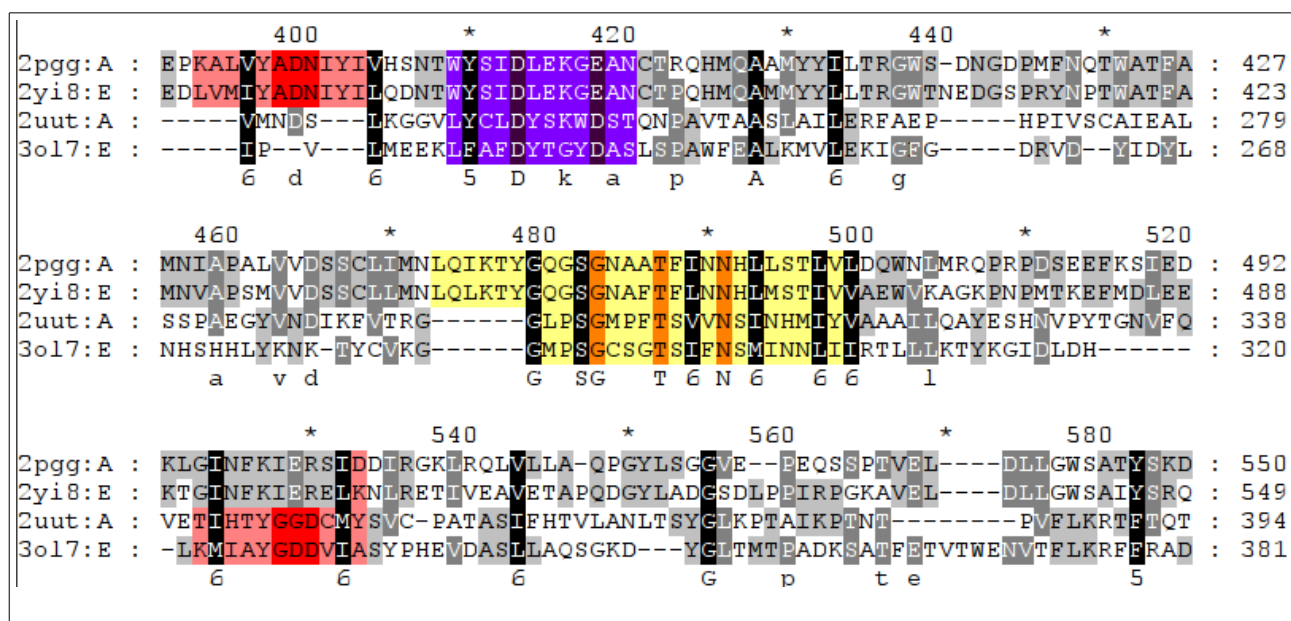


Рисунок 11. Выравнивание мотивов A, B и C субдомена palm. Красным выделен мотив A, фиолетовым — B, желтым C. Последовательность 2uut:A вместо -GDD- содержит -GGD-, т. к. это искусственно полученный мутант.

PDB ID	организм	перестановка	RMSD	Q-score
2PGG:A	IBDV	есть	0.0	
2YI8:E	IPNV	есть	0.798	0.177
2UUT:A	<i>Sapporo virus</i>	нет	3.013	0.148
3OL7:E	<i>Human enterovirus C</i>	нет	3.392	0.139

Таблица 4. Первые 4 уникальные находки PDBeFold.

Выравнивание субдоменов *palm* изображено на рисунке 11.

На рисунках 12 - 14 изображены структурные выравнивания, полученные с помощью PDBeFold.

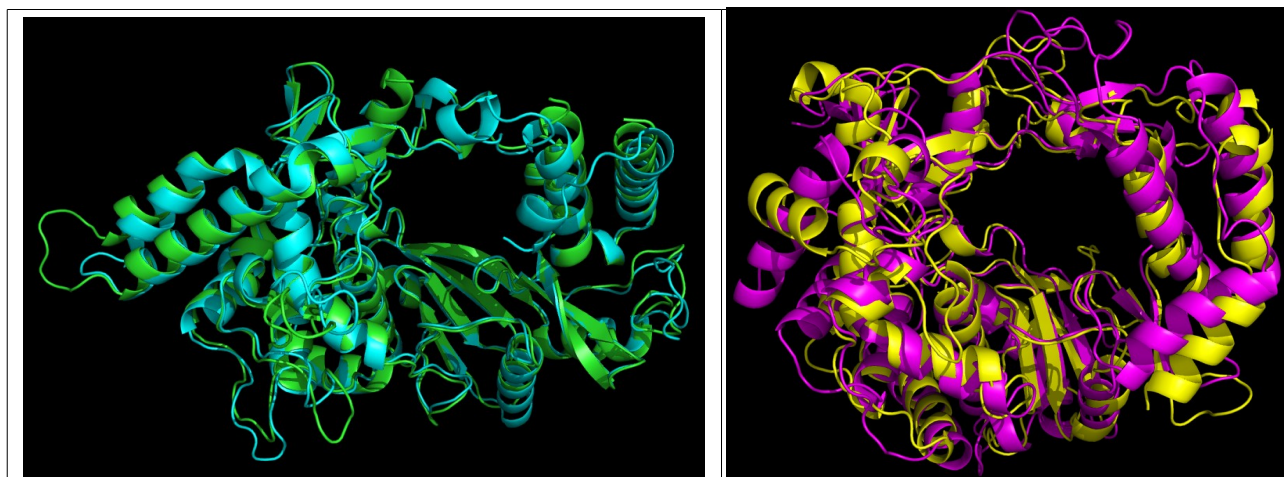


Рисунок 12. Структурное выравнивание неканонических (слева) и канонических (справа) центральных доменов. 2PGG:A — зеленый, 2YI8:E — голубой, 2UUT:A — фиолетовый, 3OL7:E — желтый.

Канонические и неканонические домены довольно похожи между собой. RMSD для пары 2PGG:A/2YI8:E составляет 0.798 Å, а для 2UUT/3OL7 — гораздо больше, 3.194 Å.

Неканонический домен имеет более вытянутую форму, чем канонический, поэтому выравнивание получается чуть хуже — RMSD 3.392 Å (рис. 13). Как и предполагали Gorbalenya et al., трипептиды GDD и ADN находятся в одинаковых положениях в пространственной структуре, хотя по последовательности, расстояние между ними больше 100 а.о. (рис. 14).

К сожалению, структура [4] неканонической RdRp из *Permutotetraviridae* до сих пор в PDB не выложена.



Рисунок 13. Структурное центральные доменов выравнивание неканонической (2PGG:A, отмечена зеленым) и канонической (3OL7:E, отмечена желтым) полимераз.

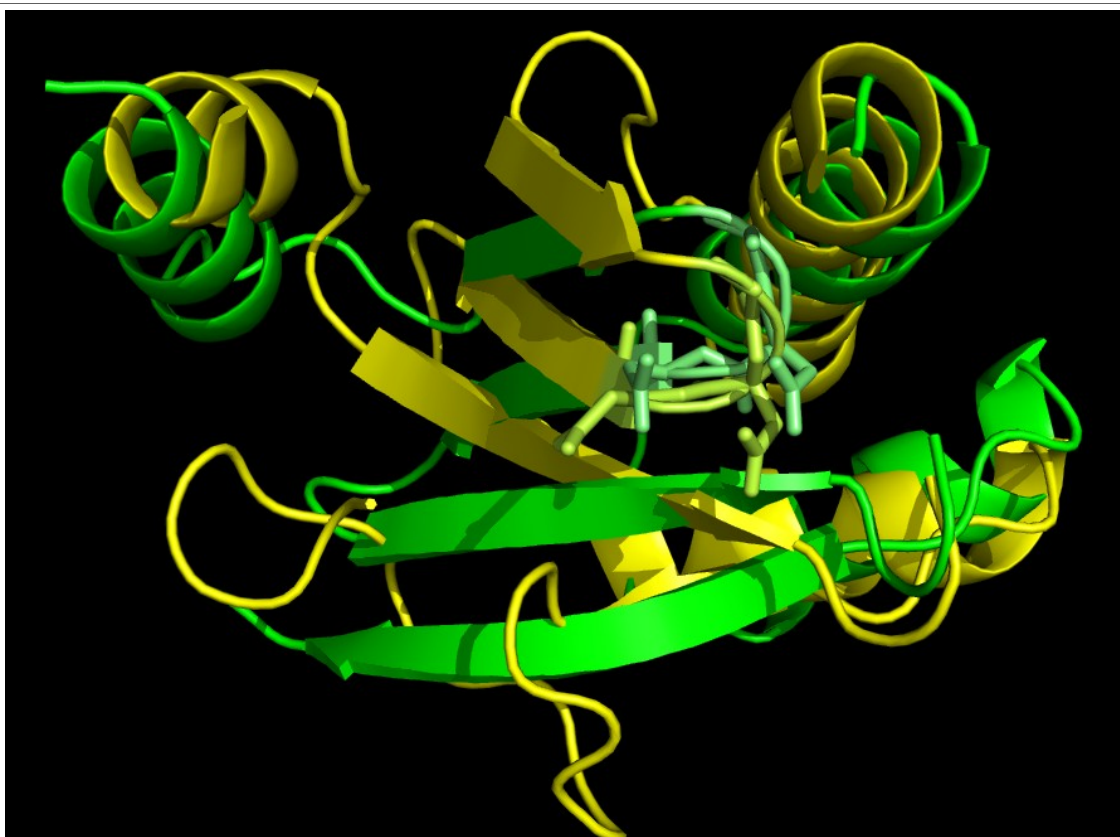


Рисунок 14. Выравнивание субдоменов *palm* из структур 2PGG (зеленый) и 3OL7 (желтый). Остов представлен в модели cartoons, а трипептиды GDD и ADN — в проволочной модели.

Список литературы

- [1] Koonin EV // The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. // *J Gen Virol.* 1991 Sep;72 (Pt 9):2197-206.
- [2] Gorbalenya AE, Pringle FM, Zeddam JL, Luke BT, Cameron CE, Kalmakoff J, Hanzlik TN, Gordon KH, Ward VK // The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. // *J Mol Biol.* 2002 Nov 15;324(1):47-62.
- [3] Pan J, Vakharia VN, Tao YJ // The structure of a birnavirus polymerase reveals a distinct active site topology // *Proc Natl Acad Sci U S A.* 2007 May 1;104(18):7385-90. Epub 2007 Apr 24.
- [4] Ferrero D, Buxaderas M, Rodriguez JF, Verdaguer N // Purification, crystallization and preliminary X-ray diffraction analysis of the RNA-dependent RNA polymerase from *Thosea asigna* virus // *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2012 Oct 1;68(Pt 10):1263-6. doi: 10.1107/S1744309112037529
- [5] Sabanadzovic S, Ghanem-Sabanadzovic NA, Gorbalenya AE // Permutation of the active site of putative RNA-dependent RNA polymerase in a newly identified species of plant alpha-like virus. // *Virology.* 2009 Nov 10;394(1):1-7. doi: 10.1016/j.virol.2009.08.006. Epub 2009 Sep 29.
- [6] Garriga D, Navarro A, Querol-Audí J, Abaitua F, Rodríguez JF, Verdaguer N // Activation mechanism of a noncanonical RNA-dependent RNA polymerase. // *Proc Natl Acad Sci U S A.* 2007 Dec 18;104(51):20540-5. Epub 2007 Dec 11.