Задание 1

Определите, какие домены Pfam встречаются в последовательности вашего белка.

Был произведен поиск в Pfam и найдено пять семейств. В работе рассматривалось семейство

RNA\_pol\_Rpb1\_1 (PF04997)

Сохранен seed выравнивания семейства.

Задание 2

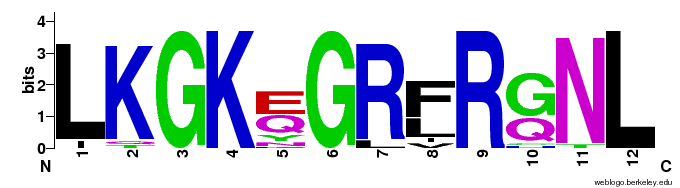
Была использована программа cons на сервере <http://emboss.bioinformatics.nl/>, был сохранен консенсус, а также был выбран блок 423-434 в полученном выравнивании.

Для блока 423-434 был построен LOGO

Консенсусная послдеовательность:

>EMBOSS\_001

LKGKEGRFRGNL



>Methanopyrus kandleri AV19, DNA-directed RNA polymerase subunit A'

MTLMEEPRVQDPPKYVKRILFGILPPDKIRKMSKVEVTSPETYD

EDGYPIEGGVMDTRMGVIDPGLRCRTCGQPAGRCPGHFGHIELARPVLHPRYAERIKD

VLRATCPECGRVKLPEDEIESYLREVREGIRPVKSVAAEVIDRAEKRESCPHCGAESR

KITYQKPTTIMMDDERLWPVDIRAWLEKIPDEDVEILGFHPERARPEWAVLTVLPVPP

VTMRPSIILETGERAEDDLTHKLVDIVRVNLRLKESLEAGAPEPIIEDQWDLLQYHVT

TYIDNEIGGVPVARHRTGGKPLKGIVQRLKGKEGRFRQNLAGKRVNFSARSVITPDPE

LDPFTLGVPEAIAKELTVPERVTEWNIDRLRKLVLNGPNKHPGANYVVKRDGSRIRIT

DENKEKLAEDLEPGDVVERHLMDGDPVMFNRQPTLHIQSELGFRVRVMPGKTFRLPAA

CYMGVGFNADFDGDEMNLHVPQSEEARAEVRELMGCYVYHLWPPKNGHEDVRNSPIHE

AIVGIHLLSRAWIPLREAYQLLEEVKKEGDWFEPTEVKLKPEAKYRRGFEEEIHELEF

VCGQPREGEPLVTGRQLLSVFLPDWLDRVELENAVCPDDEEHGKVVIEKGRIVRGFLD

EKTVHGIMTEIVHRYAAREVARGESYPWFKAAETTLRAMDKIGRLGLRFVTRYGFTLG

IDEFEDIYERFRDEVERLCDEAAKKAKEIIERGERRLQELEEHETCNRSRIERGEMLE

RNIESEVMAILNQPRVETERLLKKHRDLFNPAYIMPESGARGSITNIARMALIGGQQA

VMGERPHRAKHNTEHIMEKIVNYSYTDRVTPHFEKSLIPGVKEGGFVRSGFFEGLDMV

EYFLHAMSGREGLIDKGFRTADSGYLHRRYVTTALDLIVDSGGRVRDSANNVIQLTYG

EDGVDPAKKWGLNLDVEKALKVVKEVVLEEERRVRGGAG

Задание 4

Сильный паттерн - тот, с помощью которого, в идеале, находятся только гомологичные последовательности. Для его составления можно использовать много условий на колонки. Тем не менее, условия должны быть оправданы дополнительными соображениями - свойствами аминокислотных остатков, правильностью выравнивания, возможными удлиннениями гэпов и др. Плата за силу - возможный пропуск находок.

Слабый паттерн - тот, с помощью которого, в идеале, находятся все мотивы. Плата за слабость - возможность значительного числа ложных находок.

Сильный патерн: L-[KQT]-G-K-[EYQNK]-G-[RL]-[FLIV]-R-[GQH]-[NT]-L

Слабый патерн: L-[KQT]-G-K-X-G-[RL]-[FLIV]-R-X-[NT]-L

[442 hits (by 1 pattern) on 442 sequences]

На 12239 последовательности число находок не похоже.

L-X-G-K-X-G-X-[FLIV]-R-X-[NT]-L

[462 hits (by 1 pattern) on 462 sequences]

L-X-G-K-X-G-X-[FLIVM]-R-X-X-L

[470 hits (by 1 pattern) on 470 sequences]

Колонки консервативные, но встречаются не только эти буквы.

[LIV]-X-G-K-X-G-X-[FLIVM]-R-X-X-L

[893 hits (by 1 pattern) on 893 sequences]

[LIV]-X-G-K-X-G-X-[FLIVM]-R-X-X-[LIV]

[909 hits (by 1 pattern) on 909 sequences]

[LIV]-X-G-[KRH]-X-G-X-[FLIVM]-[RKH]-X-X-[LIV]

[1185 hits (by 1 pattern) on 1185 sequences]

[LIVFMA]-X-G-[KRH]-X-G-X-[FLIVM]-[RKH]-X-X-[LIVFMA]

[1573 hits (by 1 pattern) on 1564 sequences]

Далее писать слабый патерн я не видела смысла, так как схема была бы нарушена.

Сильный паттерн построен путем написания для каждой позиции всех возможных вариантов аминокислот, встречающихся в множественном выравнивании.

Слабый паттерн был написан при помощи такой схемы: если в позиции в разных последовательностях встречаются аминокислоты, обладающие определенным свойством - например алифатичные I и L, то множество дополняется до всех, обладающих этим свойством - V,I,L (алифатичные).

Если же в одной позиции встречаются аминокислоты, обладающие совершенно не сходными свойствами (P, V, R), то на эту позицию ставится X.

Мой белок является часто встречающимся и этим обусловлено многообразие находок.

Таким образом семейство консервативно, блок консервативен, встречается у всех секвенированных организмов, что объясняет большое количество находок в разных эукариотных организмах не родственных групп.