TD : Annotation d'une séquence de la famille Sigma-70

Rappel de Biologie: les facteurs sigma sont des protéines du complexe d'initiation de la transcription. Ce sont les facteurs sigma qui reconnaissent le promoteur et recrutent le "core enzyme" de la RNA polymérase composé des facteurs alpha, beta et gamma. Il y plusieurs sortes de facteurs sigma, qui reconnaissent des promoteurs différents grâce notamment à la région de fixation à l'ADN en position -35 qui est différente pour chaque facteur sigma (Figure 1).

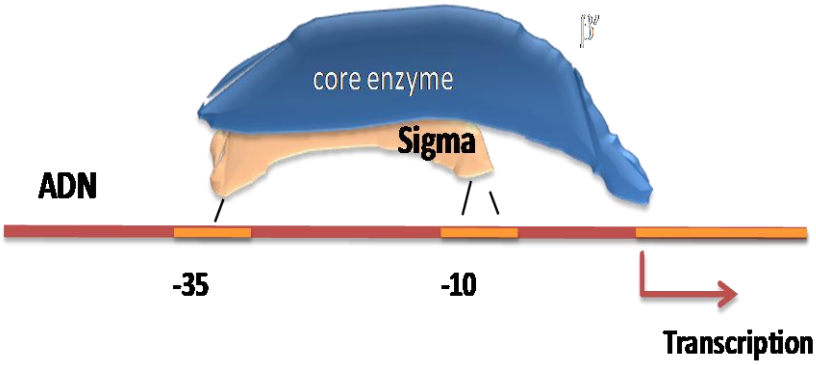

Figure 1: schéma de fonctionnement du Facteur sigma

On cherche à annoter un gène d'E. coli codant pour une protéine de 613 aa que nous appellerons protéine "unknown".

On analyse la constitution en domaines de "unknown" au moyen du site Interpro. On obtient le résultat suivant.

Banque	**Domaine**	**Nom**	**E-value**	**position**
PFAM	PF03979	Sigma70_r1_1	3.6e-30	[2-80]
PFAM	PF04542	Sigma70_r2	4.6e-23	[379-449]
PFAM	PF04539	Sigma70_r3	6.1e-29	[458-534]

Les informations trouvées dans PFAM sur chacun des 3 domaines sont les suivantes:

Family: Sigma70_r1_1 (PF03979)
[...] Region 1.1 is also involved in promoter binding

Family: Sigma70_r2 (PF04542)
Region 2 of sigma-70 is the most conserved region of the entire protein. All members of this class of sigma-factor contain region 2.
The high conservation is due to region 2 containing both the -10 promoter recognition helix and the primary core RNA polymerase binding determinant. The core binding helix, interacts with [...] the largest polymerase subunit, beta prime

Family: Sigma70_r3 (PF04539)
Region 3 forms a discrete compact three helical domain within the sigma-factor. [...] Region 3 primarily is involved in binding the core RNA polymerase in the holoenzyme.

Q1: Interprétez les résultats obtenus avec Interpro

On réalise un Blast de la séquence protéique contre Swissprot. On obtient 131 solutions de E-value <10, dont principalement des gènes des familles RpoS, RpoD, Rp32 et RpoF.

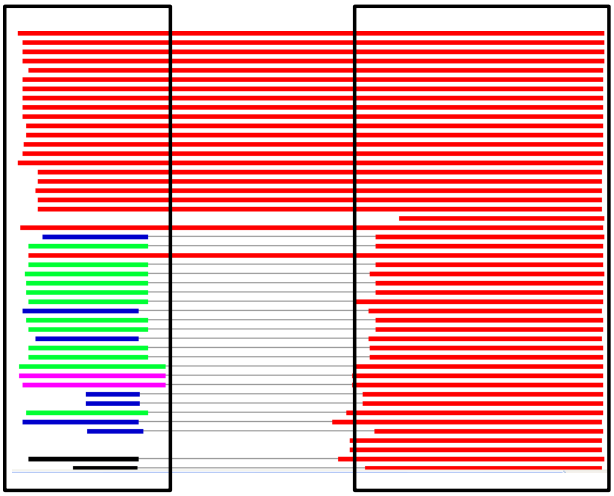

Figure 2. *schéma des alignements Blast (on ne voit ici que des alignements de E-value < 10e-10*:

Q2: A partir de la figure 2, sachant que le bloc conservé de droite est environ situé entre les positions 375 et 600 et le bloc moins conservé de gauche entre les positions 10 et 130, et en vous basant sur l'analyse des domaines, faites une hypothèse sur les fonctions des protéines trouvées par Blast.

On récupère les séquences d'une quinzaine de protéines ayant toutes une E-value Blast inférieure à 10e-10. Leurs noms sont les suivants :

```
RPOF_STRCO     RPOF_STRAU     RP32_PSEAE     RP32_PROMI     RP32_SERMA
RP32_ECOLI     RP32_CITFR     RPOD_LEPIN     RPOD_BUCBP     RPOD_BUCAI
RPOD_SALTY     RPOS_COXBU     RPOS_VIBCH     RPOS_SALTY     RPOS_ECOLI
```

Q3: Sans autre information que les E-values, les noms des protéines et la Figure 2, faites une hypothèse raisonnable quant à l'homologie/paralogie/orthologie des séquences sélectionnées.

On réalise un alignement multiple et un arbre phylogénétique des protéines sélectionnées à l'aide du serveur phylogeny.fr (options par défaut). On obtient l'alignement et l'arbre ci-dessous.

Q4: En analysant l'alignement multiple et en connaissant les différents domaines fonctionnels de la protéine, peut-on confirmer les hypothèses ci-dessus?

Q5: L'arbre phylogénétique confirme-t-il vos hypothèses? Comment?

```
CLUSTAL FORMAT: MUSCLE (3.7) multiple sequence alignment


RPOF_STRCO      ---------------------------------------MPASTAPQAPPAPPAQAQA--
RPOF_STRAU      ---------------------------------------MTVPASTAPQVPPQQDPQVPHPQ
RP32_PSEAE      ---------------------------------------MTTSLQPVHALVP----------
RP32_PROMI      ---------------------------------------MTQEMQSLALVPQ----------
RP32_SERMA      ---------------------------------------MTKEMQTLALVPQ----------
RP32_ECOLI      ---------------------------------------MTDKMQSLALAPV----------
RP32_CITFR      ---------------------------------------MTKEMQNLALAPV----------
RPOD_LEPIN      MENLQSMPEVQKIISLGKANGEVSYDDINEILPDKILNSEKIDDFFTLLH--EMGIEIVE
RPOD_BUCBP      -MEQNPQSQLKLLVTYGKEQGYLTYSEINDHLSDNIINSDQIEDIIQMIN--DMGIQVVE
RPOD_BUCAI      -MDQNPQSQLKLLVTHGKEQGYLTYSEVNDHLPEDIIDSEQIDDIIQMIN--DMGIPVVE
RPOD_SALTY      -MEQNPQSQLKLLVTRGKEQGYLTYAEVNDHLPEDIVDSDQIEDIIQMIN--DMGIQVME
unknown_EC      -MEQNPQSQLKLLVTRGKEQGYLTYAEVNDHLPEDIVDSDQIEDIIQMIN--DMGIQVME
RPOS_COXBU      MKTKTTKKTIKKAAKK-----------IKKPSKRKIKKTAKKSRPKKPIKASDHGLIFAK
RPOS_VIBCH      ------------------------------------MSVSNTVTKVEEFDFEDEALEVLE
RPOS_SALTY      ---------------------------------MSQNTLKVHDLNEDAEFDENGVEAFD
RPOS_ECOLI      ---------------------------------MSQNTLKVHDLNEDAEFDENGVEVFD


RPOF_STRCO      ---------------QAPAQAQEAPAPQRSRGADT----RALTQV--LFGELKGLAPGTP
RPOF_STRAU      EPREEP---------HEEPPSPPAAPRPQSRGADT----RALTQV--LFGQLKGLQPGTR
RP32_PSEAE      -------------------------------------GANLEA---YVHSVNSIPLLSP
RP32_PROMI      ------------------------------------GSIEA---YIRAANSYPMLTA
RP32_SERMA      ------------------------------------GSLEA---YIRAANAYPMLTA
RP32_ECOLI      ------------------------------------GNLDS---YIRAANAWPMLSA
RP32_CITFR      ------------------------------------GNLES---YIRAANAWPMLSA
RPOD_LEPIN      EYTRNTLEPASTL--VPKDDSKPARKKKESSASTS----GSEDPIKLYLREIGKVSLISG
RPOD_BUCBP      KAPDSDDLILHEIKRNNETDEDIVEATAQVLSTVESELGRTTDPVRMYMREMGTVELLTR
RPOD_BUCAI      EAPDADDLILNEI--NTDTDEDAVEAATQVLSSVESELGRTTDPVRMYMREMGTVELLTR
RPOD_SALTY      EAPDADDLLLAEN--TTSTDEDAEEAAAQVLSSVESEIGRTTDPVRMYMREMGTVELLTR
unknown_EC      EAPDADDLMLA----ENTADEDAAEAAAQVLSSVESEIGRTTDPVRMYMREMGTVELLTR
RPOS_COXBU      TKKET----------TEKEDAELANAKAKTKKRRET---RSSDPTQIYLRELGFQPLLNA
RPOS_VIBCH      TDAEL----------TSDEELVAVEGASEDVREEFDASAKSLDATQMYLSEIGFSPLLTA
RPOS_SALTY      EKALSE---------EEPSDNDLAEEELLSQGATQ----RVLDATQLYLGEIGYSPLLTA
RPOS_ECOLI      EKALVE---------QEPSDNDLAEEELLSQGATQ----RVLDATQLYLGEIGYSPLLTA
                                                        .        .

RPOF_STRCO      EHD-----------------------------------------------------------
RPOF_STRAU      EHE-----------------------------------------------------------
RP32_PSEAE      EQERELAERLFYQQ------------------------------------------------
RP32_PROMI      EEEKELAERLHYEG------------------------------------------------
RP32_SERMA      EEERELAERLHYQG------------------------------------------------
RP32_ECOLI      DEERALAEKLHYHG------------------------------------------------
RP32_CITFR      DEERALAEKLHYQG------------------------------------------------
RPOD_LEPIN      ETEVFLAKRIE-KGEKIIEETILSSSILRANYIKLLPKIRSKKIKVYDLIRVDKMYALNA
RPOD_BUCBP      EGEIDIAKRIE-DGINQVQCSVAEYPEAITHLLEQYNRVELGELRLSELIH--GFVDPNA
RPOD_BUCAI      EGEIDIAKRIE-EGINQVQCSVSEYPEAITYLLEQYDRVKTGQIRLSDIIT--GFVDPNA
RPOD_SALTY      EGEIDIAKRIE-DGINQVQCSVAEYPEAITYLLEQYDRVEAEEARLSDLIT--GFVDPNA
unknown_EC      EGEIDIAKRIE-DGINQVQCSVAEYPEAITYLLEQYDRVEAEEARLSDLIT--GFVDPNA
RPOS_COXBU      KEELKIARRVH-KG------------------------------------------------
RPOS_VIBCH      EEEVLYARRAL-RG------------------------------------------------
RPOS_SALTY      EEEVYFARRAL-RG------------------------------------------------
RPOS_ECOLI      EEEVYFARRAL-RG------------------------------------------------
                . :

RPOF_STRCO      -----------------------------------------------------------
RPOF_STRAU      -----------------------------------------------------------
RP32_PSEAE      ------------------------------------------DL-----------------
RP32_PROMI      ------------------------------------------DL-----------------
RP32_SERMA      ------------------------------------------DL-----------------
RP32_ECOLI      ------------------------------------------DL-----------------
RP32_CITFR      ------------------------------------------DL-----------------
RPOD_LEPIN      EEAHKLEELFFKNILVIQE----------------------QEKVLQEAVSKIRKYSET
RPOD_BUCBP      EAVHTSITNHINTNSEEHHHNDNTEENN--------DDDHLVDPELAREKFIALKNQYHI
RPOD_BUCAI      EEIFPPTAIHIGSELLDEQQNNEEDEENNQEDH---EDDHSIDPELANEKFSELRIQYNN
RPOD_SALTY      EEEMAPTATHVGSELSQEDLDDDEDEDEEDGDDDAADDDNSIDPELAREKFAELRAQYVV
unknown_EC      EEDLAPTATHVGSELSQEDLDDDEDEDEEDGDDDSADDDNSIDPELAREKFAELRAQYVV
RPOS_COXBU      ------------------------------------------DP-----------------
RPOS_VIBCH      ------------------------------------------DE-----------------
RPOS_SALTY      ------------------------------------------DV-----------------
RPOS_ECOLI      ------------------------------------------DV-----------------


RPOF_STRCO      -----------------------------------------------------------
RPOF_STRAU      -----------------------------------------------------------
RP32_PSEAE      -----------------------------------------------------------
RP32_PROMI      -----------------------------------------------------------
RP32_SERMA      -----------------------------------------------------------
RP32_ECOLI      -----------------------------------------------------------
RP32_CITFR      -----------------------------------------------------------
```

```
RPOD_LEPIN      SKKYKEFKEKIDASTEI----IHNAIRELGVSQKEIQKISQKIKSMVFRIKEIDRHFLKI
RPOD_BUCBP      TNYAIKTKNRNHQDTLIEINNLSEIFKQFRLVPKQFDHLVINMRKMMQKIRIQERNIAKL
RPOD_BUCAI      TNNTIKNKNRTHKDSLLEIYNLSEVFKQFRLVPKQFDHLVNNMRHMMERVRKQERIIIKL
RPOD_SALTY      TRDTIKAKGRSHAAAQEEILKLSEVFKQFRLVPKQFDYLVNSMRVMMDRVRTQERLIMKL
unknown_EC      TRDTIKAKGRSHATAQEEILKLSEVFKQFRLVPKQFDYLVNSMRVMMDRVRTQERLIMKL
RPOS_COXBU      ------------------------------------------------------------
RPOS_VIBCH      ------------------------------------------------------------
RPOS_SALTY      ------------------------------------------------------------
RPOS_ECOLI      ------------------------------------------------------------

RPOF_STRCO      ------------------------------------------------------------
RPOF_STRAU      ------------------------------------------------------------
RP32_PSEAE      ------------------------------------------------------------
RP32_PROMI      ------------------------------------------------------------
RP32_SERMA      ------------------------------------------------------------
RP32_ECOLI      ------------------------------------------------------------
RP32_CITFR      ------------------------------------------------------------
RPOD_LEPIN      KAQYGQDVRD--IKAFNRFIEKNEKLDDIEVKMGVNIDEVREVIKDIRNNERKLRRMEQE
RPOD_BUCBP      CIEYGNIPKKALLNFFSLQKTNQIWSKIVQNIKKPWIDKFNTIKPDIQMILKKLQKVEEE
RPOD_BUCAI      CVEICKMPKKNFIKIFPIKKINYLWFIREQNTNQPWSENLKKVKEDVFISVKKLIKIEEE
RPOD_SALTY      CVEQCKMPKKNFITLFTGNETSETWFNAAIAMNKPWSEKLHDVAEEVQRCLQKLRQIEEE
unknown_EC      CVEQCKMPKKNFITLFTGNETSDTWFNAAIAMNKPWSEKLHDVSEEVHRALQKLQQIEEE
RPOS_COXBU      ------------------------------------------------------------
RPOS_VIBCH      ------------------------------------------------------------
RPOS_SALTY      ------------------------------------------------------------
RPOS_ECOLI      ------------------------------------------------------------

RPOF_STRCO      ----------------------RVRAALIEANLPLVRYAAARFRSRNEPMEDVVQVGTI
RPOF_STRAU      ----------------------RVRGALIEANLPLVRYAAARFRSRNEPMEDVVQVGTI
RP32_PSEAE      ----------------------EAARQMVLAHLRFVVHIAKSYSGYGLAQADLIQEGNV
RP32_PROMI      ----------------------EAAKTLILSHLRFVIHVARSYSGYGLPQADLIQEGNI
RP32_SERMA      ----------------------DAAKQLILSHLRFVAHIARNYSGYGLPQADLIQEGNI
RP32_ECOLI      ----------------------EAAKTLILSHLRFVVHIARNYAGYGLPQADLIQEGNI
RP32_CITFR      ----------------------EAAKTLILSHLRFVVHVARNYAGYGLRQADLIQEGNI
RPOD_LEPIN      AGSTVQEIKDWGEKIIKGEREISQAKKELVKANLRLVVSIAKRYANRGMHFFDLIQEGNI
RPOD_BUCBP      TGLTIHQVKNINKRISLGEAKAKRAKKEMVEANLRLVISIAKKYTNRGLQFLDLIQEGNI
RPOD_BUCAI      TGLTIEQVKDINKRMSIGEAKARRAKKEMVEANLRLVISIAKKYTNRGLQFLDLIQEGNI
RPOD_SALTY      TGLTIEQVKDINRRMSIGEAKARRAKKEMVEANLRLVISIAKKYTNRGLQFLDLIQEGNI
unknown_EC      TGLTIEQVKDINRRMSIGEAKARRAKKEMVEANLRLVISIAKKYTNRGLQFLDLIQEGNI
RPOS_COXBU      ----------------------KARKQMIEANLRLVVKIARHYVNRGLPFLDLIEEGNL
RPOS_VIBCH      ----------------------AARKRMIESNLRLVVKISRRYSNRGLALLDLIEEGNL
RPOS_SALTY      ----------------------ASRRRMIESNLRLVVKIARRYGNRGLALLDLIEEGNL
RPOS_ECOLI      ----------------------ASRRRMIESNLRLVVKIARRYGNRGLALLDLIEEGNL
                                     :: ::* :*   :  : . .    *::: *.:

RPOF_STRCO      GLINAIDRFDPERGVQFPTFAMPTVVGEIKRYFRDNVRTVHVP-RRLHELWVQVNSATED
RPOF_STRAU      GLINAIDRFDPERGVQFPTFAMPTVVGEIKRYFRDNVRTVHVP-RRLHELWVQVNAATED
RP32_PSEAE      GLMKAVKRFNPEMGVRLVSFAVHWIKAEIHEFILRNWRIVKVATTKAQRKLFFNLRSQKK
RP32_PROMI      GLMKAVRHFNPEVGVRLVSFAVHWIKAEIHEYVLRNWRIVKVATTKSQRKLFFNLRKNKK
RP32_SERMA      GLMKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRNWRIVKVATTKAQRKLFFNLRKTKQ
RP32_ECOLI      GLMKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRNWRIVKVATTKAQRKLFFNLRKTKQ
RP32_CITFR      GLMKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRNWRIVKVATTKAQRKLFFNLRKTKQ
RPOD_LEPIN      GLIKAVDKFEYKKGYKFSTYATWWIRQAITRAISDQARTIRVP-VHMIEQVNKVIRETRL
RPOD_BUCBP      GLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIP-VHMIETINKLNRISRQ
RPOD_BUCAI      GLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIP-VHMIETINKLNRISRQ
RPOD_SALTY      GLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIP-VHMIETINKLNRISRQ
unknown_EC      GLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIP-VHMIETINKLNRISRQ
RPOS_COXBU      GLLTAVEKFDPERGFRFSTYATWWIRQTIERAIMNQSRTVRLP-IHVIKELNVYLRTAKK
RPOS_VIBCH      GLIRAVEKFDPERGFRFSTYATWWIRQTIERALMNQTRTIRLP-IHVVKELNIYLRTARE
RPOS_SALTY      GLIRAVEKFDPERGFRFSTYATWWIRQTIERAIMNQTRTIRLP-IHIVKELNVYLRTARE
RPOS_ECOLI      GLIRAVEKFDPERGFRFSTYATWWIRQTIERAIMNQTRTIRLP-IHIVKELNVYLRTARE
                **: *: .*:   * .: ::*   :   *   .  : * :.:.  .

RPOF_STRCO      LTTAFGRSPTTAEIAERLRITEEEVLSCIEAGRSYHATSLE-AAQEGDGLPGLLDRLGYE
RPOF_STRAU      LTTLHGRTPTTPEIAERLRISEDEVLSCIEAGRSYHATSLE-AAQEGDGMPGLLDRLGYE
RP32_PSEAE      RLAWLNNEEV-HRVAESLGVEPREV-REMESRLTGQDMAFD-PAADADDESAYQSPAHYL
RP32_PROMI      RLGWFNQDEV-ELVAKELGVSESDV-REMESRMSAQDMAFD-MSADDSDDPHPVAPVLFL
RP32_SERMA      RLGWFNQDEV-ELVARELGVTSKDV-REMESRMAAQDMTFDPTPDDEARDGQSMAPVLYL
RP32_ECOLI      RLGWFNQDEV-EMVARELGVTSKDV-REMESRMAAQDMTFD-LSSDDDSDSQPMAPVLYL
RP32_CITFR      RLGWFNQDEV-EMVARELGVSSKDV-REMESRMAAQDMTFD-MSSDDESDSQPMAPVLYL
RPOD_LEPIN      FIQEFGRDPNNEEIAERLGWPVQKV-KMVK-NVAREPISLE-IPVGSEEDSELGD--FIP
RPOD_BUCBP      ILQEIGREPTPEELSERMLIPEDKI-RKVL-KIAKEPISME-TPIGEDDDSHLGD--FLE
RPOD_BUCAI      MLQEIGREPTPEELSEKMLIPEDKI-RKVL-KIAKEPISME-TPIGDDDDSHLGD--FIE
RPOD_SALTY      MLQEMGREPTPEELAERMLMPEDKI-RKVL-KIAKEPISME-TPIGDDEDSHLGD--FIE
unknown_EC      MLQEMGREPTPEELAERMLMPEDKI-RKVL-KIAKEPISME-TPIGDDEDSHLGD--FIE
RPOS_COXBU      LTQEVDHEATPEDVAHLIDKPVQEI-RRIM-DLAPSATSID-VPISEDGQKSLVD--TLA
RPOS_VIBCH      LSQRLDHEPTPEEIALELDRPVDDV-TKML-RLNERISSVD-TPIGGDGDKALLD--ILP
RPOS_SALTY      LSHKLDHEPSAEEIAEQLDKPVDDV-SRML-RLNERITSVD-TPLGGDSEKALLD--ILA
RPOS_ECOLI      LSHKLDHEPSAEEIAEQLDKPVDDV-SRML-RLNERITSVD-TPLGGDSEKALLD--ILA
                     ..      ::  :     .:   :         :.:  .
```

```
RPOF_STRCO      DPA----LDGVEHRDL-------VRHLLVQLPEREQRILLLRY----YSNLTQSQISAEL
RPOF_STRAU      DPE----LAGVEHRDL-------VRHLLVQLPEREQRILLLRY----YNNLTQSQISAEL
RP32_PSEAE      EDHRYDPARQLEDADWSDSSSANLHEALEGLDERSRDILQQRW--LSEEKATLHDLAEKY
RP32_PROMI      EDKSSDFADGIEEDNWDNHAADRLTLAIKTLDERSQDIIRARW-LEDDNKSTLQELADKY
RP32_SERMA      QDKSSDFAEGIEEDNWESNAADKLAYALEGLDERSQHIIRARW-LDDDNKSTLQELADQY
RP32_ECOLI      QDKSSNFADGIEDDNWEEQAANRLTDAMQGLDERSQDIIRARW-LDEDNKSTLQELADRY
RP32_CITFR      QDKSSNFADGIEEDNWEDQAANKLTHAMEGLDERSQDIIRARW-LDEDNKSTLQELADRY
RPOD_LEPIN      DTEVETPVNAAASSILAEQ----IRQVLHTLPAREQKVIRMRFGLDDGYPQTLEEVGYQF
RPOD_BUCBP      DTNIELPLDSATSASLRSA----TKNVLSGLTTREAKVLRMRFGIDMNTDHTLEEVGKQF
RPOD_BUCAI      DTTLELPLDSATSESLRSA----THDVLSGLTAREAKVLRMRFGIDMNTDHTLEEVGKQF
RPOD_SALTY      DTTLELPLDSATTESLRAA----THDVLAGLTAREAKVLRMRFGIDMNTDHTLEEVGKQF
unknown_EC      DTTLELPLDSATTESLRAA----THDVLAGLTAREAKVLRMRFGIDMNTDYTLEEVGKQF
RPOS_COXBU      DDNNIDPARLIQNVDLQDH----IERWLAQLDERHREVVILRFGLHENEKGTLEAVGKAV
RPOS_VIBCH      DSHNADPEFSTQDDDIRES----LLNWLDELNPKQKEVLARRFGLLGYEPSTLEEVGREI
RPOS_SALTY      DEKENGPEDTTQDDDMKQS----IVKWLFELNAKQREVLARRFGLLGYEAATLEDVGREI
RPOS_ECOLI      DEKENGPEDTTQDDDMKQS----IVKWLFELNAKQREVLARRFGLLGYEAATLEDVGREI
                :                        :  *  .   ::  *:        *   :.

RPOF_STRCO      GVSQMHVSRLLARSFQRLRSANRIDA-------------
RPOF_STRAU      GVSQMHVSRLLARSFARLRSANRIEA-------------
RP32_PSEAE      NVSAERIRQLEKNAMSKLKGRILA----------------
RP32_PROMI      GVSAERVRQLEKNAMKKLRLAIED----------------
RP32_SERMA      GVSAERVRQLEKNAMKKLKMAIEA----------------
RP32_ECOLI      GVSAERVRQLEKNAMKKLRAAIEA----------------
RP32_CITFR      GVSAERVRQLEKNAMKKLRAAIEA----------------
RPOD_LEPIN      KVTRERIRQIEAKALRRLRHPSRSKKLK--DYIDG----
RPOD_BUCBP      DVTRERIRQIEAKALRKLRHPSRSEILR--SFLDD----
RPOD_BUCAI      DVTRERIRQIEAKALRKLRHPSRSEVLR--SFLDD----
RPOD_SALTY      DVTRERIRQIEAKALRKLRHPSRSEVLR--SFLDD----
unknown_EC      DVTRERIRQIEAKALRKLRHPSRSEVLR--SFLDD----
RPOS_COXBU      GLTRERVRQIQIDALQQLRHILEMEGVTGEEVED-----
RPOS_VIBCH      NLTRERVRQIQVEGLRRLREILVKQGLNMEALFNVEYDN
RPOS_SALTY      GLTRERVRQIQVEGLRRLREILQTQGLNIEALFRE----
RPOS_ECOLI      GLTRERVRQIQVEGLRRLREILQTQGLNIEALFRE----
                 ::  .: .:   .: .*.
```