

New insights on the feature and function of Tail Tubular Protein B and Tail Fiber Protein of the Lytic Bacteriophage ϕ YeO3-12 specific for *Yersinia enterocolitica* Serotype O:3

Anna Pyra^{1*}, Karolina Filik², Bożena Szermer-Olearnik², Anna Czarny², Ewa Brzozowska^{2**}

¹ University of Wrocław, Faculty of Chemistry, 14 F. Joliot-Curie St, Wrocław, 50383, Poland

² Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, 12 R. Weigl St, Wrocław, 53114, Poland

Correspondence to:

* anna.pyra@chem.uni.wroc.pl

** ewa.brzozowska@hirszfeld.pl

BLAST amino acid sequence analysis [28] showing similarity of TTPBgp12 to other TPs.

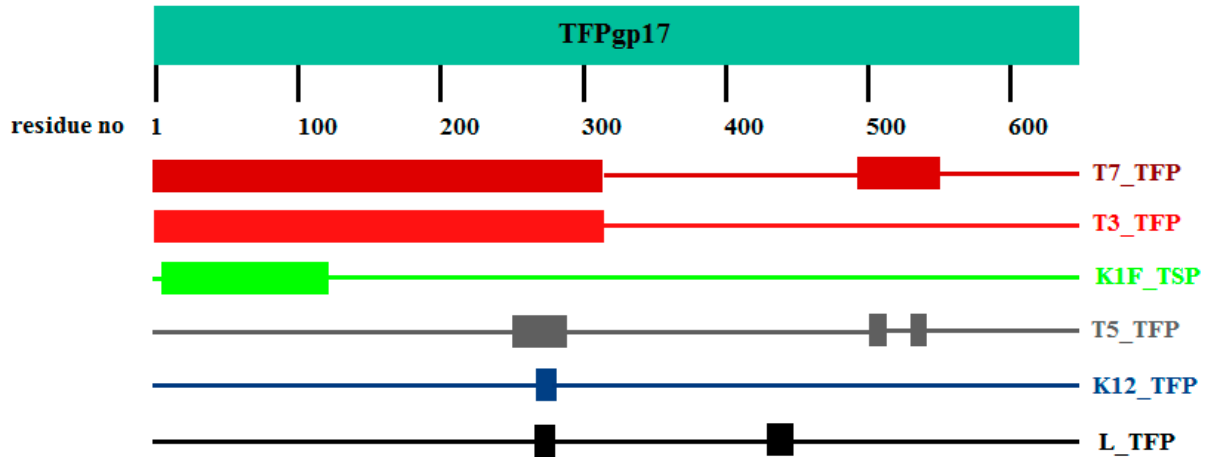
Phage Tail Protein (TP)	Phage of	% of amino acid sequence identity of TP
Tail Tubular Protein B (TTPB)	<i>Yersinia, Escherichia, Salmonella, Citrobacter, Serratia, Klebsiella, Leclercia, Enterobacteria</i>	97-99
	<i>Salmonella, Serratia, Escherichia, Enterobacteria, Erwinia, Yersinia, Pectobacterium</i>	65-70
	<i>Pectobacterium, Erwinia, Pantoea, Yersinia, Escherichia, Enterobacteria, Salmonella, Klebsiella</i>	62-65
Tail Fiber Protein (TFP)	<i>Enterobacter</i>	99-100
	<i>Escherichia, Stenotrophomonas, Klebsiella</i>	62-67

BLAST amino acid sequence analysis [29] showing similarity of TFPgp17 to other phage TFPs.

Phage of	% of amino acid sequence identity of TFP to TFPgp17
<i>Enterobacter, Yersinia, Serratia, Citrobacter, Shigella, and Escherichia</i>	99-100
<i>Dickeya</i>	92
<i>Yersinia</i>	81-84
<i>Pseudomonas and Klebsiella</i>	76 -78
<i>Dickeya</i>	68
<i>Klebsiella, Escherichia, Stenotrophomonas and Pseudomonas</i>	51-55
<i>Leclercia, Escherichia, Pectobacterium, Erwinia, Enterobacteria, Yersinia, Xylella, Klebsiella, Serratia, Kluyvera and Pseudomonas</i>	47-48
<i>Pseudomonas, Serratia, Citrobacter and Salmonella</i>	43-46
<i>Pseudomonas</i>	24

The BLAST amino acid sequence alignment scheme [28] showing similarity of TFPgp17 and TFP of *Enterobacteria* phage T7 (T7_TFP), TFP of *Enterobacteria* phage T3 (T3_TFP) of tail spike protein of *Enterobacteria* phage K1F (K1F_TSP), L-shaped TFP of *Escherichia* phage T5 (T5_TFP), side TFP of prophage K12 (K12_TFP) and TFP of *Enterobacteria* phage lambda (L_TFP).

Distribution of the top Blast Hits on 6 sequences



Sequence alignment of TTPBgp12 and T7_TTPgp12 from bacteriophage T7 (PDB code: 6r21) [13] using The Clustal Omega tool [24]. Identities between TTPBgp12 and T7_TTPgp12 are shown in red.

```

TTPBgp12      MALISQSIKLNKGGISQQPDILRFAEQGSVQINGWSSESEGLQKRPPMIHLKTLGPAGYV      60
T7_TTPgp12    MALISQSIKLNKGGISQQPDILRYPDQGSRQVNGWSSETEGLQKRPPLVFLNLTLDNGAL      60

TTPBgp12      GAQPYVHLINRDEFQYFVFTGEDIKVFDLDGKEYQVR--GDRSYVRANPREDLRMIT      118
T7_TTPgp12    GQAPYIHLINRDEHEQYAVFTGSGIRVFDLSGNEKQVRYPNGSNYIKTANPRNDRMVT      120

TTPBgp12      VADYTFVTNRKVVVQSNLQSVNLFQFQDQGDALINVRGGQYGRRLSIEFNGAERAQVLP      178
T7_TTPgp12    VADYTFIVNRNVVAQKNTKSVNLFNPNFNQDGLINVRGGQYGRELIVHINGKDVAKYKIP      180

TTPBgp12      DGSQFAHVNEVDGQAIAEKLAQLRNINLGNPNNDQDFNKWRFNVGPGFIHILAPNNDNVW      238
T7_TTPgp12    DGSQPEHVNTDAQWLAEELAKQMRNTNL-----SDWTVNVQGGFIHVVTAPSGQQID      231

TTPBgp12      GLQTKDGYADQLINPVTHYTQSFQKLPINAPDGYIVKIVGDTSKTADQYVRFDLNRKVV      298
T7_TTPgp12    SFTTKDGYADQLINPVTHYAQSFKLPFNAPNGYMKIVGDASKSADQYVRYDAERKVV      291

TTPBgp12      VETIGWNTRTHLYYHTMPWALVRASDGNDFKVLWEGARTVGGDDTNPYPSFTGQTINDI      358
T7_TTPgp12    TETLWNTEDQVLWETMPHALVRAADGNDFKVLWESPKSCGDVDNPFWPSFVGSINDV      351

TTPBgp12      FFFRNRLGFLSGENIILSRTSKYFNFPASVSNYSDDDPIDVAVSHNRVSTLKYAVPFSE      418
T7_TTPgp12    FFFRNRLGFLSGENIILSRTAKYFNFPASIANLSDDDPIDVAVSTNRIAILKYAVPFSE      411

TTPBgp12      ELLLWSDQAQFVLTAAGILSSRSVELNLTQFDVQDRARPHGVGRNVYFASPRASFTSIN      478
T7_TTPgp12    ELLIWSDEAQFVLTAAGTLTSKSVELNLTQFDVQDRARPHGIGRNVYFASPRASFTSIH      471

TTPBgp12      RYYAVQDVSSVKNAEDMTAHVPNYIPNGVFSISGTTAENFAAILTSGAPNRVYIYKFLYI      538
T7_TTPgp12    RYYAVQDVSSVKNAEDI TSHVPNYIPNGVFSICGSGTENFCSVLSHGDP SKIFMYKFLYL      531

TTPBgp12      DEEIRQQSWSHWDFGDNVTVFAAQVINSTMVLMGNEHAVVMGRHLHFTKN SIDIPGEPYR      598
T7_TTPgp12    NEE LRQQSWSHWDFGENVQVLAQCQSSSDMVVILRNEFNFTFLARISFTKNAIDLQGEPYR      591

TTPBgp12      LYIDAKRKYTIPTAGTYNDDTYQTSISLATIYGMNFTKGRVSVVFPDGKIEVDQFINGWS      658
T7_TTPgp12    AFMDMKIRYTIPTSGTYNDDTFTSIHIPTIYGANFGRGKITVLEPDGKITVFEQETAGWN      651

TTPBgp12      SDPVLRLDGNQEGQVVIYGFNIPFTYTFPSKFLIKKTAEDGSTATEDIGRLQLRRRAWVNYE      718
T7_TTPgp12    SDPVLRLSGNLEGRMVYIYGFNINPVYEFPSKFLIKQTADDGSTATEDIGRLQLRRRAWVNYE      711

TTPBgp12      DSGAFTIRVNNLSREFIYTMAGARLGSNLRVGRSNIGTGQYRFPVVGNAQTNLVITIEDS      778
T7_TTPgp12    NSGTFDIYVENQSSNWKYTMAGARLGSNLRVGRSNIGTGQYRFPVVGNAKFNFTVYILSD      771

TTPBgp12      ASTPLNIIGCGWEGNYLRRSSGI      801
T7_TTPgp12    ETPLNIIGCGWEGNYLRRSSGI      794

```

The results of HHPred analysis of TFPgp17 [31].

No	Homologous template protein	Source	PDB code	% of probability	Amino acid residues of template protein	Amino acid residues of TFPgp17	The length of polypeptide chain
1	Endo-N-acetylneuraminidase (chaperone)	<i>Enterobacteria</i> phage K1F	3GW6	99.4	10-275	419-643	208
2	Receptor recognition protein	<i>Salmonella</i> phage vB_SenMS16	6F45	99.3	78-193	524-645	113
3	L-shaped tail fiber protein	<i>Enterobacteria</i> phage T5	4UW8	99.2	232-426	412-645	156
4	Neck appendage protein (chaperone)	<i>Bacillus</i> phage GA-1	3GUD	99.1	1-123	539-641	93
5	phiAB6 tailspike	unidentified phage	5JSD	99.1	21-179	2-188	156
6	Tail spike protein	<i>Acinetobacter</i> phage vB_AbaP_AS12	6EU4	98.9	569-716	500-638	125

Multiple sequence alignment of TFPgp17 and TFP from phage T7 (T7_TFP, UniProt code: P03748), TFP from phage T3 (T3_TFP, UniProt code: P10308), tail spike protein from phage K1F (K1F_TSP, UniProt code: Q04830). Identities between TFPgp17 and all other selected proteins are shown in dark blue; those between TFPgp17, T7_TFP and T3_TFP are shown in red; those between TFPgp17, T7_TFP and K1F_TSP are shown in grey; those between TFPgp17 and T7_TFP are shown in turquoise; those between TFPgp17 and T3_TFP are shown in cyan and those between TFPgp17 and K1F_TSP are shown in orange.

```

TFPgp17  MATTIKIVMTYPLDGGSTDFNIPFEYLARKFVVRTLIGVDR----KELILNQOYRFATKT 56
T7_TFP   MANVIKIVLTYQLDGSNROFNIPFEYLARKFVVRTLIGVDR----KVLITINTDYRFATRT 56
T3_TFP   MANVIKIVLTYQLDGSNROFNIPFEYLARKFVVRTLIGVDR----KVLISINADYRFATRT 56
K1F_TSP  MS-TI---T-QFFSGNTQYRIEFDYLARTFVVVTLVMSNSPTLNRVLEVGROYRFLNPT 54

TFPgp17  TISTTRALGPFADGYTLIEIRRFSAIDRLVDFDGGILRAYDLNISQVQTLRVAAEARDL 116
T7_TFP   TISLTKANGPFADGYTTIELRRVTSITDRLVDFDGGILRAYDLNVAIQTMRVAAEARDL 116
T3_TFP   TISLTKANGPFADGYTTIELRRVTSITDQLVDFDGGILRAYDLNVAIQTMRVAAEARDL 116
K1F_TSP  MIEML---VDQSGFDIVRHKQF-CIDLVVDFRNGSVLTAEDLTTAEELCAIMIAKEGPDQ 110

TFPgp17  TADTIGVNDGHLN-----ARGRRIWNVADAQVVGDAINLQGI-----QRGNDSA 161
T7_TFP   TTDTIGVNDGHLN-----ARGRRIWNLANAVDDRDAVFFGQL-----KIDNQNS 161
T3_TFP   TADTIGVNDGHLN-----ARGRRIWNLANAVDDRDAVFFGQL-----KIDNQNS 161
K1F_TSP  TVDLAKEYADAAGSSAGNANDSEEARRIAESIRAA----GLIGYIETRRSFEKGYNVTT 165

TFPgp17  LNSANNKQEQADPATAANANDANNSANASASASASASAEALAKMWTSD-TVVEQDIESS 220
T7_TFP   WQARNEALQFNEAETFRQAEGFKNE-----SSTNATNKQRDET-KGYRDEAK-- 211
T3_TFP   WQARNEALQFNEAETFRQAETVFKNE-----SGTNATNKQRDEA-NGSRDEAE-- 211
K1F_TSP  WSE---VLLMEEDGYYRMDGTLFKNVFAGSTPETSQIGG-LGAWVSVGDAAALRQGISN- 220

TFPgp17  RTYALHSLYRNETKDSDRAAVSETHAK--ASECGAANSAAAKVSETAKA----- 271
T7_TFP   -RFONTAGQYATSAGNSAS-----AAH--QSEVNAENSATASANSALAEQ----- 254
T3_TFP   -QFONTAGQYATSAGNSAT-----TAT--QSEVNAENSATADANSALAEQ----- 254
K1F_TSP  ---PEGAILYFELHEAR---NLDKMDARGNGAKGDGVTDDTAAITSAINDTFVGGKING 273

TFPgp17  -SEERAITEASKLQGNDFAAAEISVTQNDVQKQAVVSSPGNITG--GGLVSTGAASIQK 328
T7_TFP   -QADRAREADKLEWYNGLAGAIDKVDGTVYKNGNIEANGRLYMTNDFDC-GQYQQFF 312
T3_TFP   -HADRLLEADKLGIFNGLAGRIDKGDGTVYKNGGIEANGRLYLTSDGDC-GQYQQFF 312
K1F_TSP  NGNTYKVTSLFDISRFINTRFVYERIPQQLFYASEEFVQGLFKITDTPY----- 324

TFPgp17  GALVGEDLIVGRDITAKQDMYSQR-----NIAVAGVYTAQGGIEQTLATNIYK 377
T7_TFP   CGVTHRYSVAGWGDENGCMKVVQRREWPTAIGGNIQLVNGQIITQCGMHTQGLK---- 367
T3_TFP   CGSACRYSVAGWGIKAWLKHVERRERTTAIVDNIQLVNGHIIAQGGDMHTGLK---- 367
K1F_TSP  -----YNANPQDKA-----FVYENVI----- 340

TFPgp17  LYRLHINNPQHVGRQGLHIGNESGSGESNFTINRAGSGGFVFRVNAENSVEVGRV 437
T7_TFP   -----LQNGHVQLLESAADKAIYILSKDGNRNNWYIGRSDN-----NN 406
T3_TFP   -----LQNGHALYLESAADKAIYILSKDGNRNNWYIGRSDN-----NN 406
K1F_TSP  -IAPINGSD---RNGVSRHVSQVKSQDDGQVWTFP-----EWLITDLRFDY--FTVNY 387

TFPgp17  DITGGQVIY-----ANNLQURSARIE----- 459
T7_TFP   DCTFHSHYH-----GTLTLKQDYAVV----- 428
T3_TFP   DCTFHSHYVY-----GTNLTLPDYAVV----- 428
K1F_TSP  HCSMSQVCRNELFAMIESTRLANQALTNCALNDRPHMSRLHLTGQITKAANQRYATIKVF 447

TFPgp17  GNNIVGQ-----NLYAGMSTMFEGNGLTGGINA---- 490
T7_TFP   NKH-----FHVQAVVAIDGNIQGT----- 449
T3_TFP   NGR-----FHVQAVVAIDGNIQGT----- 449
K1F_TSP  DRGLFVCDFVNFNSAVTGVSGDMTVATVIDKDNFTVLTFFNQTSDLNNAIGQWHDGTSF 507

TFPgp17  ---QNGNLMSGLNNSLFAKFFGGVQLFTAR-GGYLLEGRVDQI---AVGFRMFQSDR 541
T7_TFP   ---MGGKMLDAYLRDSFVAKSKANTQVMSGS-ACGGVSVTVSQDL--RFRNIWIKCAN 501
T3_TFP   ---MGRKMLDAYLMDTYKKTMAHTQVMAAA-SDSVMGGGSQ-----TDTLHRTCD 497
K1F_TSP  HKSPKRTDLGLIPS-----VTEVHSFATIDNNGFAMQYHQGVAPREVGLFYFPAF 560

TFPgp17  R----IKEDIKVVRSADDMNLIRSYIPVSYKYKQDASYTDNRGRTNTEIGKSRAGFIT 596
T7_TFP   N---SNMFFRTGPDG---T-----YFTASDGGNLRPQIHS-----NGLGFFGN 537
T3_TFP   A--TYGIRPFTTIGTSELVLIV-----STSPQPRRMLKFIHS-----NGRVFQ 541
K1F_TSP  NSFSNYVRRQIPSEYEPQASEPCIKYDGL-----YLITROT---RCDLGGSSLHR 609

TFPgp17  Q-CLIRINPEAVDVMSDGGQSPFNQIIQGLMLLVGNLDAARIQLEKDKT----- 645
T7_TFP   I-ADSRVFNALMIVENE----- 553
T3_TFP   I-ADRAATPTAIAVEDV----- 557
K1F_TSP  SRDIOQTKEKSLR-PPSDVNHNTLDFAVVCEDLIMFGSERAKNEWEAGAFDDRYKASTPR- 667

```

Multiple sequence alignment of TFPgp17 and L-shaped TFP from phage T5 (T5_TFP, UniProt code: P13390), side TFP from prophage K12 (K12_TFP, UniProt code: P76072) and TFP from phage lambda (L_TFP, UniProt code: P03764). Identities between TFPgp17 and all other selected proteins are shown in dark blue; those between TFPgp17, T5_TFP and K12_TFP are shown in turquoise; those between TFPgp17, K12_TFP and L_TFP are shown in red; those between TFPgp17, T5_TFP and L_TFP are shown in grey; those between TFPgp17 and T5_TFP are shown in pink; those between TFPgp17 and K12_TFP are shown in orange and those between TFPgp17 and L_TFP are shown in purple

TFPgp17	MATTIKTVMYPLDGSSTDFNIPFEYLAKF---VRVTLIGVDRK-EL-ILNQCYRFATK	55
T5_TFP	-----MAI---TK	5
K12_TFP	MAVKISGVVLDGDT-GKFPVQ-NCTIQLKAKRNSTVVVNTLASENPDZAGRYSDVEYGGY	58
L_TFP	MAVKISGVVLDGDT-GKFPVQ-NCTIQLKAKRNSTVVVNTVQSENPDZAGRYSDVEYGGY	58
TFPgp17	TTISTTRALGFADGYY--LIEIRRFPSATDRLVDFT---DGSILRAYDLNIQQVQLHV	109
T5_TFP	-IIL-----QQMVTMDQNSITAKYFKITVVLKNSIS---	36
K12_TFP	SVILLVEGFFPSHAGTITVYEDSQPOTLNDFLGAMTEDDARPEALRFRLELMVEEVARNAS	118
L_TFP	SVILLQVDFGFFPSHAGTITVYEDSQPOTLNDFLCAMTEDDARPEALRFRLELMVEEVARNAS	118
TFPgp17	AEEARDLADTIGVNDGNLDAFGRRIVNVAQAQVGDGAI--NLGQIQIRIND-----	159
T5_TFP	-----SITA-----ADVTSALSSKA---SGFAAKQS-----	60
K12_TFP	AVA--QNTAAAKSASDASTSAPE-AATHAADADARAASSTAGQAASASASASAGT	175
L_TFP	VVA--QSTADAKKSAGDASASAAQ-VAALVTDATDARAASSTAGQAASASASASAGT	175
TFPgp17	-----SALNSANRA--KQEAADRA	175
T5_TFP	---EINAKQSELNAKDSENE---AETISATSSQQSATQASASATASANAKAA	106
K12_TFP	ASTKATEAKSAAAASSSKSAATSAAGAATSETNASASLQSAATASATA--TTKASEA	232
L_TFP	ASAKATEAKSAAAASSSKSAATSAAGAATSETNAAASQQSAATASATA--ATKASEA	232
TFPgp17	TARANDANANASAS---SASSSSGSAELAKRMTSDTVVESDLESSRTYALHSHLYRN	232
T5_TFP	KTSETNANONKRAHSETNAASSASSAS-----S-----FAT	139
K12_TFP	ATSARDAAAASKEAASSETNASSASSASAA-----S-----SAT	265
L_TFP	ATSARDAAVAASKEAASSETNASSASSAGRAA-----S-----SAT	265
TFPgp17	ETKDSADRAAVSETNAKASEG-----	253
T5_TFP	AAENSARAARKTSETNAGNSAQAADASKTAAANSATAAK-----TSETNAK-----	184
K12_TFP	AAENSARAARKTSETNARSSEAAAGQASAAAAGSKTAAASASASAASSTAGQASASATAAGK	325
L_TFP	AAENSARAARKTSETNARSSEAAARSASAAAADAKTAAA-----	303
TFPgp17	-----GAANSAAAKVSETNAKASEERAITEASKLGDG---	286
T5_TFP	-----KSETAARKTSETNAKASEERAITEASKLGDG---	217
K12_TFP	SAESAASSASTATTKAGEATEQASAAARSAAAKTSETNAKASETSAESKTAASASSS	385
L_TFP	-----	303
TFPgp17	-----NDFAAAIESVTGNDV--KMGAVSSPGNITGGGLV-STGAASIQKCALV	332
T5_TFP	VT-----QYDW---FVGTNNNSVYVIAKLTDF-GAVS-CHLTL-----MITNGNY	259
K12_TFP	AASSASSASASAKDETRQAASAKSATTASTKATEAAGSATAAQSKSTAESAAATRAETA	445
L_TFP	-----GSASTASTKATEAAGSAVSASQSKSAARAAAPRONS	340
TFPgp17	GEDLIVGRDITAKQMYSQRNIAVAVTYAQGGIEQTLATINYNKLYRLHINSNQRVGG	392
T5_TFP	GSS---YGNL---DF---VEISARLNDAROVTS---ENIT-IFLSVRRLSF----	299
K12_TFP	AKR--AEDI-----ASVALEDAS-TTK---KGVV-QLSSATNSTSETLAAT	485
L_TFP	AKR--AEDI-----ASVALEDAD-TTR---KGVV-QLSSATNSTSETLAAT	380
TFPgp17	RQGLHIGH-NESG-----SGESNFTNNG---AGSGGFVFRVNAENSV	432
T5_TFP	---NLANDNQLRYGLV--EGDGYFEVWCYQRAI IKETRVAVLAQTGR--TELYIFGQV	351
K12_TFP	PKAVKSAVDNAEKRLQKQDQNGADIPOKGCFLNN-INAVSKTEFADNR-GRYVRYNAPAGA	544
L_TFP	PKAVKVVDETRKAPLDSFALT-----	403
TFPgp17	ETRV---DITGGVIY--ANHLQVRSGARIEG---NNNIVQQNLYAG-----MGST	476
T5_TFP	-SQDTQ-----PSGFIESLAARIYDQWKP--KADLGLNAMLVGFGLGGNGLYS	401
K12_TFP	TS-KYYPVVVRSAGSVSELSARVITATSTAGDFMNCCEPFGFMPGWN-----	595
L_TFP	-----GTPAPTALRGT-----	415
TFPgp17	MFEQGNLITGGIWAQGNLMSGLANNSLFAKPP-GVQLFTA-----FGGYILEGRVD	528
T5_TFP	SVQSNVDLILNKLKNGQYWRRAAESGARVDINDHSGGFYSBGGDTHAAINVQNTGIV-	460
K12_TFP	-----TDRGRYAYGMFWQYQNNERAIMS-----MSNKGDDL--SVFTVDCAA-	638
L_TFP	-----NNT-----QI--AN-----TA-	424
TFPgp17	GTAVGFRWFQSDRRKEDIKVVRSADDDMLNIR---SYIFVSYKYKDAYTDRNRRTN	583
T5_TFP	-----KVLATTORNL-----ASDIVYANTLY--GTANKPS---KSD	491
K12_TFP	-----F-----PVFAFIEDGLSISAPGADLVVNDTTYFGATNPATECIAAD	680
L_TFP	-----F-----VLAATAVIDA-SFDALNTINELAAALGN-----	453
TFPgp17	TIEGKRSRACFI-----TQ-----DLIRLMPFAVDVMSDGMQ-SFD	618
T5_TFP	VGLGNVNDQAQVQAGDVMSGDLDIRKETPSIRLKSSTQGNALMFPDNGGGERGVINSP	551
K12_TFP	VILDFFGRGFYESHSLIVNDNLSCKLFATDEIVARGGN---QIRMIGGEYGALWRND	736
L_TFP	-----DPDFATMTNALAGK---QFK-----N-	472
TFPgp17	FWQII-----GCLMLLV-----KNLARIQELEKQK-----	645
T5_TFP	NNGSLGEHIRAKTSDGTSTGDFIVRHDRIEAKDAKISYKISSARFSDNDNTAATN	611
K12_TFP	GAKTYLLL---TNQGVYGHNTLR-----FFALNATGELVIGTKLS---	776
L_TFP	---A-TL---T-----	476