



HAL
open science

Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families

Didier Flament, Tristan Barbeyron, Murielle Jam, Philippe Potin, Mirjam Czjzek, Bernard Kloareg, Gurvan Michel

► **To cite this version:**

Didier Flament, Tristan Barbeyron, Murielle Jam, Philippe Potin, Mirjam Czjzek, et al.. Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families. *Applied and Environmental Microbiology*, 2007, 73 (14), pp.4691-4694. 10.1128/AEM.00496-07. hal-00616725

HAL Id: hal-00616725

<https://hal.univ-brest.fr/hal-00616725>

Submitted on 5 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Short communication**

2

3 **Alpha-agarases define a new family of glycoside hydrolases, distinct from**
4 **beta-agarase families**

5

6 Didier FLAMENT[§], Tristan BARBEYRON, Murielle JAM, Philippe POTIN, Mirjam

7

CZJZEK, Bernard KLOAREG and Gurban MICHEL*

8

9 Centre National de la Recherche Scientifique; Université Pierre et Marie Curie-Paris6, Unité

10 Mixte de Recherche 7139 "Marine Plants and Biomolecules", Station Biologique, F-29682

11 Roscoff Cedex, Bretagne, France

12

13 Keywords: marine bacteria, agar, glycoside hydrolase, CBM6, calcium-binding module

14 Running title: Alpha-agarases define a new family of glycoside hydrolases

15

16 *Corresponding author: Tel (33) 298 29 23 30, Fax (33) 298 29 23 24, E-mail:

17 gurban@sb-roscoff.fr

18 § Current address: UMR 6197 (IFREMER / CNRS), Laboratoire de Microbiologie et

19 Environnements Extrêmes, Plouzane, Bretagne, France

20

21 Abbreviations: aa, amino acid(s); bp, base pair(s); kb, kilobase(s); nt, nucleotide, kDa,

22 kiloDalton; CBM(s), carbohydrate-binding module(s); GH(s), glycoside hydrolase(s); SDS-

23 PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, ORF, open reading frame;

24 TSP3, thrombospondin type 3 repeats.

1 **ABSTRACT**

2 The gene of the α -agarase from “*Alteromonas agarilytica*” has been cloned and
3 sequenced. The gene-product (154 kDa) is unrelated to β -agarases and instead belongs to a
4 new family of glycoside hydrolases (GH96). The α -agarase also displays a complex
5 modularity, with the presence of five TSP3 repeats and three CBM6s.

1 Agars are the main cell wall components of numerous red macroalgae. These polymers
2 consist of 3,6-anhydro-L-galactoses and D-galactoses alternatively linked by α -(1,3) and β -
3 (1,4) linkages (10). Agars constitute a crucial carbon source for a number of marine bacteria
4 which secrete agarolytic enzymes, mainly β -agarases (EC 3.2.1.81) which hydrolyze the β -
5 (1,4) linkages (See for a review, 12). They are found in three distinct families of glycoside
6 hydrolases, families GH16, GH50 and GH86 (CAZY database, 6). Structural data are only
7 available for the GH16 β -agarases *ZgAgaA* and *ZgAgaB* from *Zobellia galactanivorans* (1, 2,
8 9). In contrast, “*Alteromonas agarilytica*” secretes an α -agarase (EC 3.2.1.158) which
9 randomly hydrolyzes the α -(1,3) linkages in agars, releasing agarotetraose as its main end
10 product (13, 15, 18). In this context, we sought to determine whether the agarases of the α -
11 type share any structural relationships with β -agarases. We report here the cloning of the α -
12 agarase gene *agaA* from “*A. agarilytica*” and we demonstrate that α -agarases define a new
13 GH family.

14 **Cloning of the *agaA* gene.**

15 The α -agarase from “*A. agarilytica*” (*AaAgaA*) was purified as previously described
16 (13) and N-terminal and internal peptide sequences were determined (Pasteur Institute,
17 France): ETLELQAESFANSGG (A) and QPRVYNPNEHIVAEIQGPAT (B), respectively. With
18 degenerated oligonucleotides derived from these microsequences (CARGCIGARTCITYGCI AA
19 and TAYAAAYCCNAAYGARCA YAT, respectively), a 2.5 kb DNA fragment was amplified by PCR
20 using *A. agarilytica* genomic DNA and labeled with [α -³²P]dCTP. This radiolabeled probe
21 was used to screen an “*A. agarilytica*” genomic library, prepared as previously described (5).
22 Among the ~5000 recombinant clones, two positive clones (pAA1 and pAA2) were identified
23 with inserts of 7.4 kb and 17.9 kb in length. Southern blot analysis and plasmid mapping
24 indicated that both inserts encompassed the same gene which is present in only one copy in
25 the genome. Plasmid pAA1 was sequenced on both strands over 4651 bp and a single ORF

1 was identified and referred to as *agaA* (4287 bp). Potential -35 and -10 promoter regions
2 (TTGAtc and TAcAca) and a Shine Dalgarno sequence (GGAG) were identified upstream of
3 the start codon. A possible transcription termination codon was found downstream of the
4 TAA stop codon. The deduced gene-product is a preprotein of 1429 residues (154 kDa) which
5 includes the peptides A and B. A signal peptide cleaved between A26 and E27 is predicted by
6 SIGNALP (4), consistent with the N-terminal sequencing of the purified extracellular α -
7 agarase.

8 Several attempts were made to overexpress *AaAgaA* and its isolated modules in *E. coli*
9 with pET or pGEX vectors in various conditions. Unfortunately, the constructs always yielded
10 inclusion bodies. However, within one week of culture at 22°C on Zd agar broth (3), the *E.*
11 *coli* clones harboring the plasmid pAA1 dug a hole in the substratum, indicating agar
12 degradation. Therefore, under the control of its own promoter, *agaA* was successfully
13 translated into an active, recombinant enzyme, confirming that this gene indeed encodes the
14 α -agarase.

15 ***AaAgaA* is a complex, modular protein with a catalytic domain defining a new GH** 16 **family**

17 Only the N-terminal region of *AaAgaA* displays significant sequence similarity with
18 proteins in the UniProt database. Based on InterProScan (14), eight distinct modules were
19 identified in this region (Fig. 1A), five thrombospondin type 3 repeats (TSP3-1, D171-G203;
20 TSP3-2, D360-L392; TSP3-3, D393-A425; TSP3-4, D426-L458; TSP3-5, D459-G491) and
21 three carbohydrate-binding modules from the family 6 (CBM6-1, E27-R159; CBM6-2, E209-
22 T343; CBM6-3, S659-L792).

23 The closest characterized protein matching the TSP3 repeats of *AaAgaA* is the cellulase
24 CelG from *Pseudoalteromonas haloplanktis*. In CelG, the TSP3 repeats constitute an
25 extended linker connecting the GH5 catalytic module and a C-terminal CBM5 (17). The TSP3

1 repeats of *AaAgaA* present about 50% sequence identity with their counterparts in CelG.
2 They also display ~30% sequence identity with the “true” type 3 repeats found in human
3 thrombospondin, whose crystal structure has been solved (11). These modules lack secondary
4 structures and are organized around a core of calcium ions coordinated by conserved
5 aspartates (DxDxDGxx[D/N]xxDxC motif). The conserved cysteine is involved in a
6 disulphide bridge linking adjacent TSP3 repeats, strengthening their stability (11). This motif
7 is strictly conserved in each of the TSP3 repeats of *AaAgaA* (Fig. 1B), indicating that these
8 modules adopt similar structure and likely bind calcium ions. This is consistent with the
9 observations that α -agarase activity is stabilized by the presence of calcium ions (13, 18).

10 In BLASTp searches with the three CBM6s from *AaAgaA*, the highest E-values are
11 always obtained for the CBM6 sequences tethered to β -agarases, while they significantly
12 decrease with CBM6 linked to non-agarolytic enzymes. Only the CBM6s attached to the β -
13 agarases *SdAga16B* and *SdAga86E* from *Saccharophagus degradans* were shown to actually
14 bind agarose (8). A pairwise comparison indicates a strong sequence identity (51%) of
15 CBM6-1 with the CBM6 from *SdAga16B* (Fig. 1C), while CBM6-2 and CBM6-3 are more
16 divergent (28% and 26%, respectively). The crystal structure of the CBM6 from *SdAga16B*
17 revealed that five residues are critical for the recognition of the non-reducing end of the
18 agarose chain: Asn39, Tyr40, Trp97, Trp127 and Asn130 (8). Four of these residues are
19 strictly conserved in CBM6-1, while Tyr40 is substituted by a similar aromatic amino acid,
20 Phe64 (Fig. 1C). Altogether these results strongly suggests that CBM6-1 is an agar-binding
21 module and likely displays selectivity towards the non reducing termini of agarose chains. In
22 contrast, the five critical residues are only partially conserved in CBM6-2 and CBM6-3.
23 Therefore the specificity of these latter CBM6 is less certain.

24 Finally, a BLASTp search on the patent databank at the NCBI identified two proteins
25 from a marine bacterium with a strong sequence identity with the C-terminal region of

1 *AaAgaA* (49% and 77% respectively, Fig. 2). These proteins, also described as α -agarases
2 (16), encompass three modules, two N-terminal CBM6s and the C-terminal module conserved
3 with *AaAgaA*. Since the N-terminal region of *AaAgaA* encompasses only additional, non-
4 catalytic modules, its conserved C-terminal region (Asn809-His1429) likely contains its
5 active site. Together these three catalytic modules (~620 residues) constitute a new family of
6 glycoside hydrolases, referred to as family GH96 (CAZY database). Therefore α -agarases are
7 structurally unrelated to the β -agarases from the families GH16, GH50 and GH86.

8 **ACCESSION NUMBERS**

9 The nucleotide sequence of the α -agarase from “*Alteromonas agarilytica*” has been
10 deposited in Genbank with the accession number AF121273. Its amino acid sequence is
11 available in Swiss-Prot under the accession number Q9LAP7.

12 **ACKNOWLEDGMENTS**

13 The authors gratefully acknowledge the help of the Conseil Régional de Bretagne,
14 which supported this research through a Ph.D. fellowship to D. F. and a research grant (PRIR
15 n° 98C389).

1 REFERENCES

- 2 1. **Allouch, J., W. Helbert, B. Henrissat, and M. Czjzek.** 2004. Parallel substrate
3 binding sites in a beta-agarase suggest a novel mode of action on double-helical
4 agarose. *Structure* **12**:623-632.
- 5 2. **Allouch, J., M. Jam, W. Helbert, T. Barbeyron, B. Kloareg, B. Henrissat, and M.**
6 **Czjzek.** 2003. The three-dimensional structures of two beta-agarases. *J. Biol. Chem.*
7 **278**:47171-47180.
- 8 3. **Barbeyron, T., A. Gerard, P. Potin, B. Henrissat, and B. Kloareg.** 1998. The
9 kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and
10 phylogenetic relationships within family-16 glycoside hydrolases. *Mol. Biol. Evol.*
11 **15**:528-537.
- 12 4. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved
13 prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783-795.
- 14 5. **Colin, S., E. Deniaud, M. Jam, V. Descamps, Y. Chevolot, N. Kervarec, J. C.**
15 **Yvin, T. Barbeyron, G. Michel, and B. Kloareg.** 2006. Cloning and biochemical
16 characterization of the fucanase FcnA: definition of a novel glycoside hydrolase
17 family specific for sulfated fucans. *Glycobiology* **16**:1021-1032.
- 18 6. **Coutinho, P. M., and B. Henrissat.** 1999. Carbohydrate-active enzymes: an
19 integrated database approach., p. 3-12. *In* H. J. Gilbert, G. Davies, B. Henrissat, and B.
20 Svensson (ed.), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society
21 of Chemistry, Cambridge.
- 22 7. **Gouet, P., X. Robert, and E. Courcelle.** 2003. ESPript/ENDscript: Extracting and
23 rendering sequence and 3D information from atomic structures of proteins. *Nucleic*
24 *Acids Res.* **31**:3320-3.

- 1 8. **Henshaw, J., A. Horne-Bitschy, A. L. van Bueren, V. A. Money, D. N. Bolam, M.**
2 **Czjzek, N. A. Ekborg, R. M. Weiner, S. W. Hutcheson, G. J. Davies, A. B.**
3 **Boraston, and H. J. Gilbert.** 2006. Family 6 carbohydrate binding modules in beta-
4 agarases display exquisite selectivity for the non-reducing termini of agarose chains. *J.*
5 *Biol. Chem.* **281**:17099-17107.
- 6 9. **Jam, M., D. Flament, J. Allouch, P. Potin, L. Thion, B. Kloareg, M. Czjzek, W.**
7 **Helbert, G. Michel, and T. Barbeyron.** 2005. The endo-beta-agarases AgaA and
8 AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes
9 with different molecular organizations and catalytic behaviours. *Biochem. J.* **385**:703-
10 713.
- 11 10. **Kloareg, B., and R. Quatrano.** 1988. Structure of the cell walls of marine algae and
12 ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol. Ann.*
13 *Rev.* **26**:259-315.
- 14 11. **Kvansakul, M., J. C. Adams, and E. Hohenester.** 2004. Structure of a
15 thrombospondin C-terminal fragment reveals a novel calcium core in the type 3
16 repeats. *EMBO J.* **23**:1223-1233.
- 17 12. **Michel, G., P. Nyval-Collen, T. Barbeyron, M. Czjzek, and W. Helbert.** 2006.
18 Bioconversion of red seaweed galactans: a focus on bacterial agarases and
19 carrageenases. *Appl. Microbiol. Biotechnol.* **71**:23-33.
- 20 13. **Potin, P., C. Richard, C. Rochas, and B. Kloareg.** 1993. Purification and
21 characterization of the alpha-agarase from *Alteromonas agarlyticus* (Cataldi) comb.
22 nov., strain GJ1B. *Eur. J. Biochem.* **214**:599-607.
- 23 14. **Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R.**
24 **Lopez.** 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**:116-
25 120.

- 1 15. **Rochas, C., P. Potin, and B. Kloareg.** 1994. NMR spectroscopic investigation of
2 agarose oligomers produced by an alpha-agarase. *Carbohydr. Res.* **253**:69-77.
- 3 16. **Tomono, J., Y. Nomura, H. Sagawa, T. Sakai, and I. Kato.** 2003. Alpha-agarase
4 and process for producing the same. Patent US 6599729.
- 5 17. **Violot, S., N. Aghajari, M. Czjzek, G. Feller, G. K. Sonan, P. Gouet, C. Gerday,**
6 **R. Haser, and V. Receveur-Brechot.** 2005. Structure of a full length psychrophilic
7 cellulase from *Pseudoalteromonas haloplanktis* revealed by X-ray diffraction and
8 small angle X-ray scattering. *J. Mol. Biol.* **348**:1211-1224.
- 9 18. **Young, K. S., S. S. Bhattacharjee, and W. Yaphe.** 1978. Enzymic cleavage of the
10 alpha-linkages in agarose to yield agarooligosaccharides. *Carbohydr. Res.* **66**:207-212.
11

1 **FIGURE LEGENDS**

2 **Figure 1:** **A.** Modular architecture of the α -agarase *AaAgaA*. CBM6 and GH96 refer to
3 carbohydrate-binding modules of the family 6 and glycoside hydrolase module of the family
4 96, respectively. The grey boxes correspond to the thrombospondin type 3 repeats (TSP3). **B.**
5 Structure-based alignment of the TSP3 repeats of the human thrombospondin (PDB code
6 1UX6), of the α -agarase *AaAgaA* and of the cellulase CelG from *Pseudoalteromonas*
7 *haloplanktis* (trEMBL code: O86099). **C.** Structure-based alignment of the three CBM6s from
8 the α -agarase *AaAgaA* and of the agar-specific CBM6s tethered to the β -agarases *SdAga16B*
9 (Genpept: ABD80437) and *SdAga86E* (Genpept: ABD81915) from *Saccharophagus*
10 *degradans*. These modules are compared to the secondary structures of the CBM6 appended
11 to *SdAga16B* (PDB code 2CDO). Alpha helices and beta strands are represented as helices
12 and arrows, respectively, and beta turns are marked with TT. The black triangles mark the
13 residues involved in the recognition of the non-reducing end of agarose chains. Figure 1B, C
14 and Figure 2 were prepared using the program ESPrit (7) and use the same colour codes.

15 **Figure 2:** Sequence alignment of the catalytic module of AgaA from *A. agarilytica* and
16 of Aga14 and Aga15 (Patent US 6599729). *Dark shaded boxes* enclose conserved positions.

A

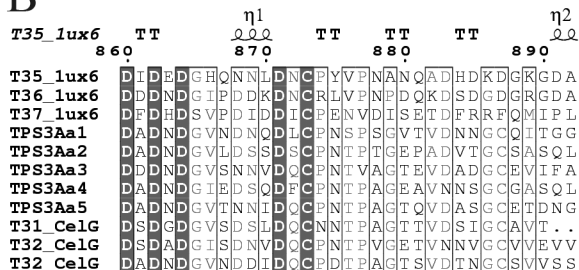
CBM6-1

CBM6-2

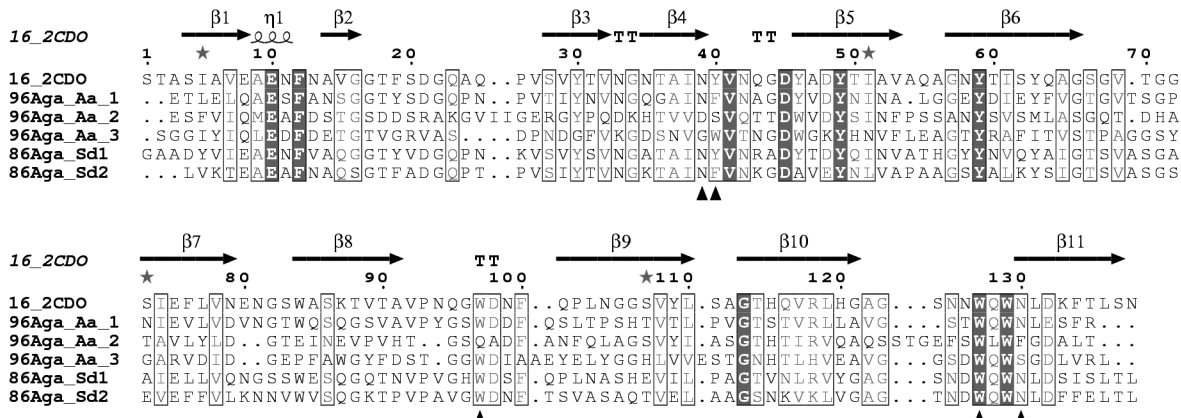
CBM6-3

GH96

B



C



820 830 840 850 860 870 880 890 900
A-aga Aa. HIVAEIQGPATGLQYLKTEVEIPLANKVVKSDVWYTYPQNRNLVVVDGDTTPYADFGATGAFWGHPEEHDFFYDDTVIMDVAVNVVDDFQSEG
A-aga15 HTVTEIEGPPVGLPFLKKPVQVPTANRLLKSDVWYTYPQNNEL...QGFDNFEGATGSAFWGHPEPENFYDDTTIIDW.TQLVQNYQIGIG
A-aga14 HISSDDTNGGSNQAMHLEFPYAIPEESRKIKKSSVWYTYPQNSNL...AGFSDFCATGAFWGHMPEEDLYDSGVLSNW.VNQVQGYRNG

910 920 930 940 950 960 970 980 990
A-aga Aa. FEYTAGREFDWGYGWFTEFTTNPOPHYVQTLDRNVNRMFMGYLSHDCYNNWLSNHSFAFVPFMKSOVDQILKANPDKLMFDTQTNSTR
A-aga15 IEYTAGREFDWGFRWVTEYLTNPEEHYVVKTLDRNVNRMFMGYLSYNGYNNWLSNHSFAFVPHMKSOVDQILRANPDKLMFDTQTNSTR
A-aga14 LDYVVGREGFDWGRWFTEYVGDPEESHWARTLDDPILMSFMGYHEHNGLYNGWLSNHSPTFVDFPKSOVDALLSANVSHIMFDSQTSSTR

1000 1010 1020 1030 1040 1050 1060 1070 1080
A-aga Aa. STDLMRTFGGDFSPYAMENFRVWLKPKYSNAQLVSMGINDITSFEDYGAYLRAQGIHTHDWSNAGDITSGNIFMMEDFIYFNRDVWNOKFAE
A-aga15 STDLMRTFGGDFNDYAMANFRVWLDPKYSSELSAMGINDIATFNRYDFLLARGVHTHSFSNAADITSGDVELEDFIYFNRDVWNOKFAE
A-aga14 STDLMGTFGGDFSTWSMDAFREYMRDKYTTABLNRNKGITNNAFNRYRNLRSRGNVTHASYSMAANKITSGTELFDFIYFNRAVLEBKMAE

1090 1100 1110 1120 1130 1140 1150 1160 1170
A-aga Aa. VLEIYIRQQRPNIEIGASTHLFESRCYIFNENITFLSGELNLGARTSISELPTNIIIVHLKGAQAVDKTLAYFPYPWFEDLRIQNAERFGR
A-aga15 VLDYIRMQRPNIEIGASTHLFESRCYIFNENITFLSGELNLGARTTIAELPTNIIIVHLKGAQAVDKKPLAYFPYPWFEBLIRDQNAERFGR
A-aga14 VLDYIRSIDADIEIGASTALTEARGYIFDKDITFLAGELAMGSAVA.DEMPIPIIISHLKSAEAVDKTLVYFPYPWFVKDLYDRNSPQMAR

1180 1190 1200 1210 1220 1230 1240 1250
A-aga Aa. GWVVAQAYAYGGLFSIPANVWVGGEV.FTWSPGADNYRDIYQFVRAQANLLDGYTSYAKAGYVHAMFSSMKAGFIDGGNQVQSSVKILTE
A-aga15 GWVVAQAYAYGGLFSIPANVWVGNTGENTWSPGADNYRDIYQFVRAQSNLFDNYTSYAKVGLVHAMYSMKAGFIDGGNQIQSSVKILTE
A-aga14 TWIIAQSAYAMGAI FSIPANVWIGDA...GVWSPGADNYRDIYQFASDNSALLLDGYDAFSKVGIVSPMMSALDITTWIDGNSRLQTSIRYLI

1260 1270 1280 1290 1300 1310 1320 1330 1340
A-aga Aa. DNNINFDMLVFGDAGYPVVPVRQADFDFKFEYIFVYDGLNLYLTAEQQAVLDAQGSKVKKHIGQRGTIAGL...QINVSINGSVSNEIVSAV
A-aga15 DNNINFDLLVFGDEGYPPVPRTEDEFNQFAHIFVYDGLLSYLTAEQQAVLDAQGSKVKKHIGQRGTITGI...DINVSINGSLSNEIVSAV
A-aga14 NNNINFDLLVFGDGPVVPVPTQAQLSALDAIIVVSDRKYLYLDAQNALLDANNQKVLDDLNSADTAAINALKATNISVITIGNAAADDITITAL

1350 1360 1370 1380 1390 1400 1410 1420
A-aga Aa. SRTHETDSTAPYVVHLLNRPFF..AGGVTPILNNVVEVAIPASVYFPQGVTSAKLHLPLDGSSTIVAVSTNANGDTPVVSVSNLEVWGMILELAH
A-aga15 SRTHETNTNAPYVVHLLNRPFF..SGGVTPILSGVEVAIPQGYFPEDVTSATLHLPLDGTSTNLVSTNNSNGDAVITVNNLEVWGMILELAH
A-aga14 SRVHESNNNAPYVVIQLNRPVNPANGVTPVLSNVKIAIPQGYFPEGITQATVHKPAGGSVNNANITSNNGDYPVITVNNLEVWGMILELAH