

## **Supporting Information**

### **Unveiling the multipath biosynthesis mechanism of 2-phenylethanol in *Proteus mirabilis***

Jinbin Liu<sup>a</sup>, Yajun Bai<sup>b</sup>, Tai-ping Fan<sup>c</sup>, Xiaohui Zheng<sup>b\*</sup>, Yujie Cai<sup>a\*</sup>

<sup>a</sup> The Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology,

Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, China

<sup>b</sup> College of Life Sciences, Northwest University, Xi'an, Shanxi 710069, China

<sup>c</sup> Department of Pharmacology, University of Cambridge, Cambridge CB2 1T, UK

First author: Jinbin Liu

<sup>a\*</sup> Corresponding authors: Yujie Cai

The Key Laboratory of Industrial Biotechnology, Ministry of Education, School of  
Biotechnology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, China

Tel.: +86-18961727911

Fax: +86-0551-85327725

E-mail: [yjcai@jiangnan.edu.cn](mailto:yjcai@jiangnan.edu.cn)

Address: Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, China

<sup>b\*</sup> Xiaohui Zheng

E-mail: [zhengxh@nwu.edu.cn](mailto:zhengxh@nwu.edu.cn)

Address: College of Life Sciences, Northwest University, Xi'an, Shanxi 710069,  
China

## Methods

### Construction of the recombinant strain

The *aaat-1*, *aaat-2*, *gsa*, *acoat*, and *dans* genes were amplified by Prime STAR HS DNA polymerase. The pCold II plasmids were extracted by TaKaRa Mini BEST Plasmid Purification Kit Ver.4.0 and digested with restriction endonucleases *SacI/SalI*. The genes and digested plasmids were ligated with In-Fusion HD Cloning Kit. The constructed plasmids pCold-II-*aaat-1*, pCold-II-*aaat-2*, pCold-II-*gsat*, pCold-II-*acoat*, and pCold-II-*dans* were transformed into *E. coli* JM109 competent cells. The verified recombinant plasmid was transformed into *E. coli* BL21 competent cells for expression. The primers used in this study are listed in Table S3.

### Overexpression and purification of recombinant enzyme

The recombinant proteins were overexpressed and collected for lysing following the method suggested by the reference.<sup>1</sup> Then, the recombinant proteins were purified using an AKTA Avant system following the method suggested by the reference.<sup>2</sup>

### Enzyme activity assays

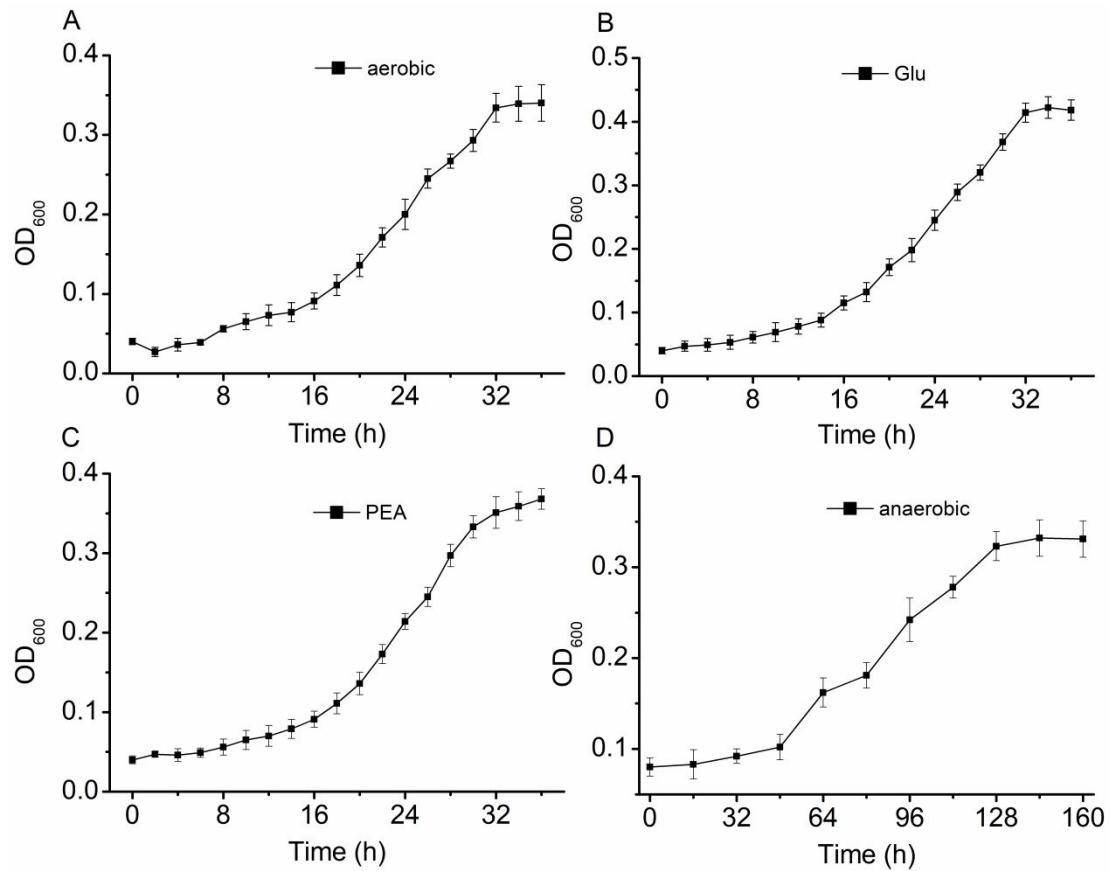
The optimum temperature of AAATs were determined using L-Phe as substrate and  $\alpha$ -ketoglutarate (2-KG) as co-substrate at temperatures ranging from 20°C to 60°C. The optimum pH of AAATs were determined using L-Phe as substrate and  $\alpha$ -ketoglutarate as co-substrate in 20 mM sodium phosphate-sodium citrate (pH 6.0 – 8.0) and 20 mM sodium carbonate-sodium bicarbonate (8.0 –10.0).

The relative activity of AAATs toward L-Phe and 2-PEA was determined using an assay mixture containing 5 mM L-Phe, or 5 mM 2-PEA, 0.1 mM PLP, 20 mM sodium phosphate buffers (pH = 8), co-substrate ( $\alpha$ -ketoglutarate, 2-KG; oxaloacetate, OAA; pyruvate, PA) 5 mM, purified enzyme 0.25 mg/mL in a total of 1 mL. The mixture was incubated at 40°C for 2 h. One mU enzyme activity was defined as the amount of enzyme required to catalyze the conversion of 1 nmol substrate. Experiments were performed in triplicate

## References

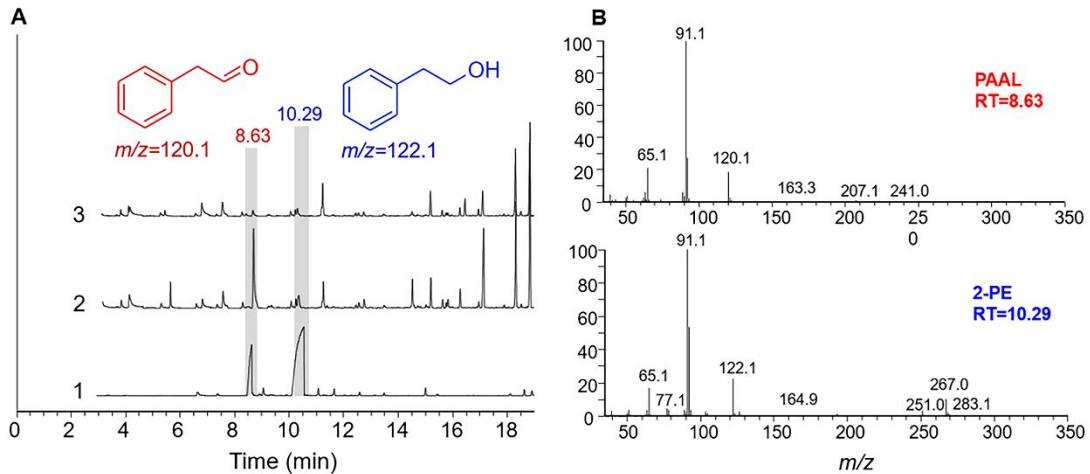
1. Liu, J.; Jiang, J.; Bai, Y.; Fan, T. P.; Zhao, Y.; Zheng, X.; Cai, Y. Mimicking a new 2-phenylethanol production pathway from *Proteus mirabilis* JN458 in *Escherichia coli*. *J. Agric. Food Chem.* **2018**, *66* (13), 3498-3504.
2. Wang, B.; Bai, Y.; Fan, T.; Zheng, X.; Cai, Y. Characterisation of a thiamine diphosphate-dependent alpha-keto acid decarboxylase from *Proteus mirabilis* JN458. *Food Chem.* **2017**, *232*, 19-24.

## Figures

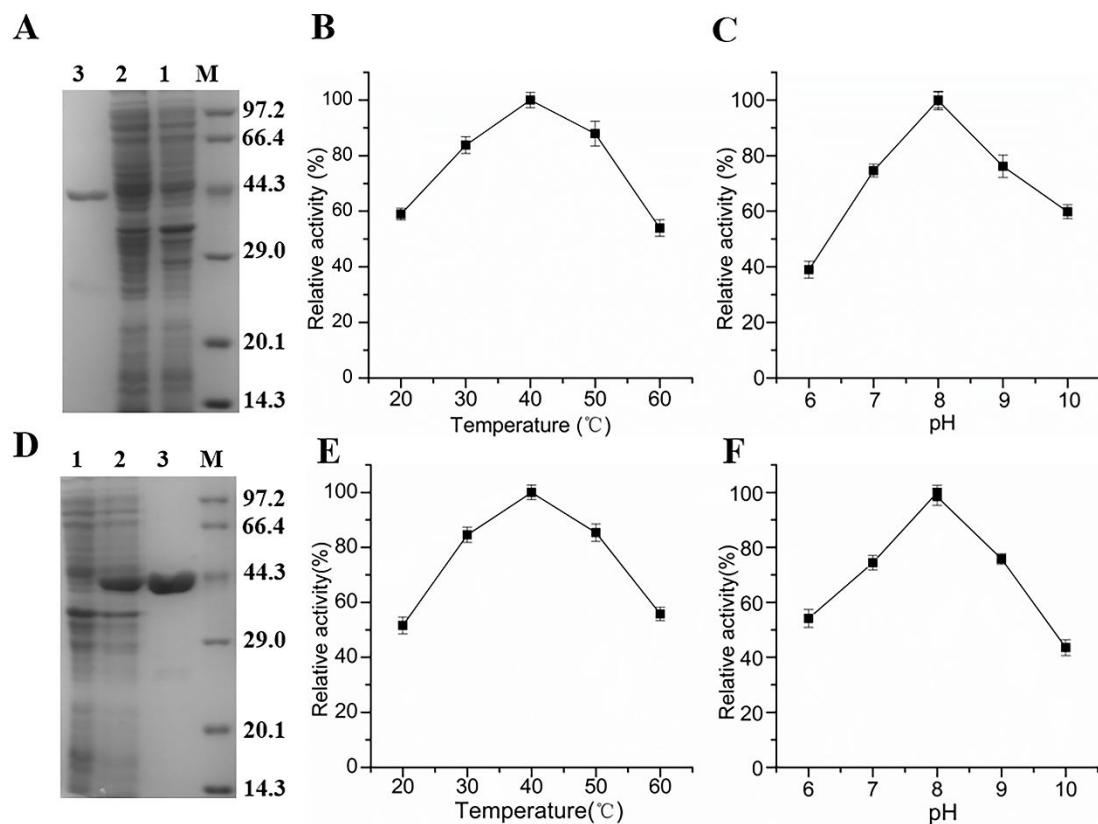


**Figure S1.** Growth kinetics of *P. mirabilis* in modified MSM

A) *P. mirabilis* cells were inoculated in Phe-MSM under aerobic conditions. B) *P. mirabilis* cells were inoculated Glu-MSM under aerobic conditions. C) *P. mirabilis* cells were inoculated in PEA-MSM under aerobic conditions. D) *P. mirabilis* cells were inoculated in Phe-AMSM under anaerobic conditions. Data are mean  $\pm$  SD of three independent experiments.

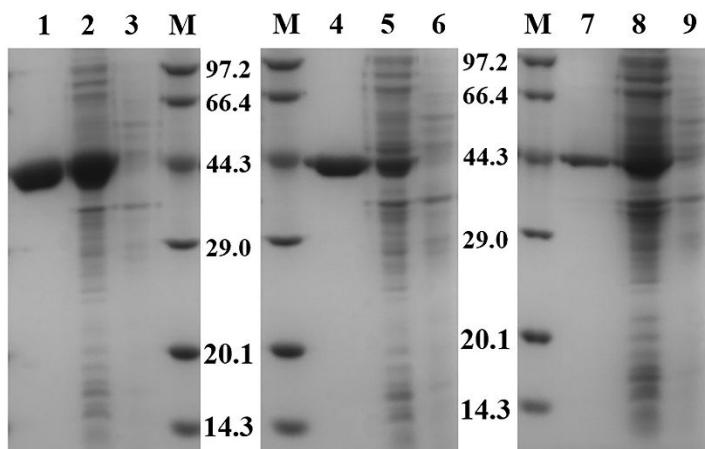


**Figure S2.** Metabolites identification in the supernatant of PPA-MSM and PEA-MSM. A. GC identification of metabolites resulting from PPA and 2-PEA metabolism. 1) the standard, RT = 8.63 PAAL, RT = 10.29 2-PE. 2) Sample of the supernatant of PPA-MSM 3) Sample of the supernatant of PEA-MSM. B. Mass spectrum identification of the metabolites.



**Figure S3.** Purification and characterization of AAATs

A. SDS-PAGE analysis of the recombinant strain and AAAT-1. M, protein marker; 1, *E. coli* BL21 with empty pCold- II plasmid; 2, *E. coli* BL21 with pCold- II -*aaat-1*; 3, the purified AAAT-1. B. Effect of temperature on the enzyme activity of AAAT-1. C. Effect of pH on the enzyme activity of AAAT-1. D. SDS-PAGE analysis of the recombinant strain and AAAT-2. M, protein marker; 1, *E. coli* BL21 with empty pCold- II plasmid; 2, *E. coli* BL21 with pCold- II -*aaat-2*; 3, the purified AAAT-2. B. Effect of temperature on the enzyme activity of AAAT-2. C. Effect of pH on the enzyme activity of AAAT-2. Date are mean ± SD of three independent experiments.



**Figure S4.** SDS-PAGE analysis of the recombinant strain and  $\omega$ -transferases

M, protein marker; 1, the purified GSA; 2, *E. coli* BL21 with pCold- II -*gsa*; 3, *E. coli* BL21 with empty pCold- II plasmid. 4, the purified ACOAT; 5, *E. coli* BL21 with pCold- II -*acoat*; 6, *E. coli* BL21 with empty pCold- II plasmid. 7, the purified DANS; 8, *E. coli* BL21 with pCold- II -*dans*; 9, *E. coli* BL21 with empty pCold- II plasmid.

## Tables

**Table S1 Primers for RT-qPCR used in this study**

Gene	Primes (5'→3')	
<i>16S rRNA</i>	TCATAAAGTCTGTCGTAGT	TATTCACCGTAGCATTCT
<i>pma</i>	CACTTAGATGCGGTATTCT	CCTCAGAGATAATTGGTAAT
<i>pm1</i>	AAGATGATGTTGTTGTAAT	GAGGTGATGTAATAAGAACG
<i>aaat-1</i>	TATAATGAGCGTGTAGGT	CGGTGGATTAGAGTAGTT
<i>aaat-2</i>	CGATTGTATGTGATAACG	GCAGGACTTGAATAGATA
<i>disa</i>	TGCCTTACCAATAGATGAA	CTAACACGCCAATAATACC
<i>kdc</i>	AAGCAGTATCAGCAATCA	AACATAGTGGCATAAGGTAA
<i>adh-1</i>	TAAGATTGCCGATTATGA	CGATAAGATTGGTGTAC
<i>adh-2</i>	TTGGTCTTGGTTGATGAG	CAGTAATTGCTGGTGGTA
<i>adh-3</i>	TAAGATTGCCGATTATGA	CGATAAGATTGGTGTAC
<i>adh-4</i>	CCTATAATGGACAACCTATT	GTAAGCAGACTTCTTCAT
<i>adh-5</i>	TTATGGTCGTGAGATGTT	ATGAGTAATAATCGGTGAGA

**Table S2 Specific activities of AAAT toward L-Phe and 2-PEA**

Enzymes	Substrates	Co-substrates	Relative activity (mU·mg <sup>-1</sup> )
AAAT-1	L-Phe	2-KG	47760.65 ± 1680.32
		OAA	30240.75 ± 1440.14
		PA	6000.01 ± 1080.84
AAAT-2	L-Phe	2-KG	29160.19 ± 1920.20
		OAA	22800.39 ± 720.26
		PA	3000.14 ± 360.84
AAAT-1	2-PEA	2-KG	34.47 ± 2.32
		OAA	16.83 ± 2.18
		PA	4.02 ± 1.67
AAAT-2	2-PEA	2-KG	2.08 ± 0.94
		OAA	6.71 ± 0.86
		PA	7.85 ± 1.04

**Table S3 Primers for PCR used in this study**

Gene	Primers (5'~3')
<i>aaat-1</i>	F: TCATCATAT <u>GGAGCTCGT</u> GTTCAACAGGTTGAAGCGTTG R: TAGACTGCAG <u>GTGACT</u> TAACTGACAGCTGCAAAGGCTTG
<i>aaat-2</i>	F: TCATCATAT <u>GGAGCTCATG</u> TTGAGAAAATCATGCTGCACCAG R:TAGACTGCAG <u>GTGACT</u> AAAGTACAGCGACAATGGCTTCACACA
<i>dasns</i>	F:TCATCATAT <u>GGAGCTCATG</u> ACACCCGATGATATTGCTTTG R:TAGACTGCAG <u>GTGACT</u> TAATTGTAAGATTAACTGCTTTCT
<i>acoat</i>	F:TCATCATAT <u>GGAGCTCATG</u> ATAAAATCAACGTATTAATAGGGC R:TAGACTGCAG <u>GTGACT</u> CACTCTGCTTTACAAACTCAGCAAG
<i>gsa</i>	F:TCATCATAT <u>GGAGCTCATG</u> GCAAGTCTGAAACGCTTATAA R:TAGACTGCAG <u>GTGACT</u> TACGCTTCATTTGCTAAAGGCATA

### The *P* values used for statistical analysis

The variances between 12 h and 24 h with growth in L-Phe

Gene	<i>P</i> value	Mean1	Mean2	Difference	t ratio
<i>aaat-1</i>	7.29E-31	7.45	151.24	-143.787	29.8878
<i>aaat-2</i>	6.71531E-16	5.47	64.90	-59.43	12.3532
<i>pma</i>	2.68121E-17	6.10	71.26	-65.1533	13.5429
<i>pm1</i>	0.0360034	3.70	14.11	-10.4067	2.16315
<i>aadc</i>	2.59869E-07	2.85	32.09	-29.2367	6.0772
<i>kdc</i>	7.3173E-33	5.86	166.22	-160.36	33.3328
<i>adh-1</i>	0.00530373	4.29	18.40	-14.1133	2.93363
<i>adh-2</i>	1.41E-10	6.44	46.46	-40.02	8.31864
<i>adh-3</i>	0.00114254	5.46	22.20	-16.7433	3.48031
<i>adh-4</i>	6.93E-19	8.47	80.52	-72.0533	14.9772
<i>adh-5</i>	1.00E-08	5.17	39.03	-33.8667	7.0396

The variances between growth in Glu and L-Phe

Gene	P value	Mean1	Mean2	Difference	t ratio
<i>aaat-1</i>	1.69E-06	2.61	142	-139.387	43.3481
<i>aaat-2</i>	4.09E-06	1.80	78.2	-76.3967	34.7546
<i>pma</i>	4.55E-05	8.40	76.26	-67.86	18.9708
<i>pm1</i>	0.014728	10.68	19.14	-8.46333	4.11038
<i>aadc</i>	0.000797	2.02	27.65	-25.6267	9.13381
<i>kdc</i>	1.37E-05	4.48	134.76	-130.277	25.6434
<i>adh-1</i>	0.001906	5.09	13.25	-8.16667	7.26539
<i>adh-2</i>	0.000248	1.48	41.97	-40.4833	12.3352
<i>adh-3</i>	0.000514	5.25	21.32	-16.0733	10.2325
<i>adh-4</i>	4.38E-06	15.14	73.19	-58.0567	34.1575
<i>adh-5</i>	0.000675	6.45	26.65	-20.1933	9.53702

The variances between 12 h and 24 h with growth in 2-PEA

Gene	P value	Mean1	Mean2	Difference	t ratio
<i>aaat-1</i>	2.41821E-20	5.41	119.72	-114.31	24.2622
<i>aaat-2</i>	5.4049E-07	4.10	34.52	-30.42	6.45663
<i>adh-1</i>	0.0473518	2.14	11.91	-9.77333	2.07439
<i>adh-2</i>	1.2019E-06	4.88	33.89	-29.0067	6.15665
<i>adh-3</i>	0.0358694	4.24	14.63	-10.3867	2.20457
<i>adh-4</i>	1.89633E-14	6.00	73.71	-67.7067	14.3707
<i>adh-5</i>	0.000341418	7.02	26.23	-19.21	4.07731

The variances between 48 h and 96 h with growth in L-Phe

Gene	P value	Mean1	Mean2	Difference	t ratio
<i>aaat-1</i>	1.51101E-06	5.27	166.39	-161.12	70.4039
<i>aaat-2</i>	2.41563E-35	7.77	95.02	-87.2433	38.1223
<i>pma</i>	0.643446	0.29	1.36	-1.06667	0.466096
<i>pm1</i>	0.650704	0.39	1.43	-1.04333	0.455901
<i>aadc</i>	1.57554E-09	4.06	21.43	-17.37	7.59009
<i>kdc</i>	5.34516E-07	6.06	172.69	-166.627	72.8101
<i>adh-1</i>	0.0155569	2.51	8.27	-5.76	2.51692
<i>adh-2</i>	9.46037E-27	5.56	60.00	-54.4333	23.7855
<i>adh-3</i>	0.0101654	3.43	9.58	-6.14667	2.68588
<i>adh-4</i>	8.11234E-34	6.16	86.50	-80.34	35.1058
<i>adh-5</i>	0.000556525	3.27	11.79	-8.52	3.72295

## DNA and protein sequences of the related enzymes

L-aromatic amino acid transaminase(AAAT-1)

Coding DNA sequence:

GTGTTCAACAGGTTGAAGCGTTGCGGGAGATCCGATTCTCTCATTAATGGATGTCT  
ATAATAAAGATCCACGTCAAGATAAAATTAACCTAAGTATTGGTCTTATTATGATGA  
AGAAGGCAAAACCCCCATTCTAGGCACTGTGTCTGGCGCGTCAGCAGCTTAATGC  
GATGACACCTACCGCAACACTGTATTGCCAATGGAAGGACTCGCCCTATGCCAC  
GAAGTACAGACTCTCCTTTGGTGTGATAACCCGCTATTGCAGACAAAAAGATTG  
CTACCATAACAAACATTAGGTGGCTCCGGAGCGCTAAAAGTCGGCGCAGATTCTAC  
ATCGCTATTCCAAGCTCTGAAGTTGGATCAGTGATCCCACCTGGGACAATCATGC  
GTCTATTGCGGGCTCCGGTTAAAGTAAATTATTATCCTTATTGATCCGAAA  
CCAAAGGGTCAAATCGATGCATTAATAGATTGCTTAAAAAACTCCTGAAAAAA  
GCATCGTACTGATGCACCCATGTTGCCATAATCCAACAGGATCCGATCTCACCAAAAGC  
ACAATGGGATCAGGTATTGAGATCCTCAAAGCGCGTCAAGCAATCCCTTCCTGAT  
ATTGCTTATCAGGGATTGCTGAAAGTTAGATGATGATGCTTATGCGGTCGCGCTA  
TGGCAAAAGCGGGTTACCTGTATTAGTCAGTAACTCATTCTCTAAAATCTCGGGAT  
TTATGGGAACGGCGGGTGGCTTATCGATTGTGATAACCGCAAAGAGAGTGTGA  
GCATGTATTAGGGCAATTAAAAGCGGGCGCACGTCGTATCTATTCAAGTCCTGCTAAC  
TATGGCGACAACGGTAAATCAAGTTATCAGATCATGTATTAACCGCACAATGGC  
AAAAAGAAGTGGCTATATCGTGACCCGATCAAAGAGATGCGTGTAACTTAGTTA  
ATGCCTTAAAGAAGCATTACCTGAGAAAAACTTGATCATTATTAACACAACGCG  
GTATGTTCACTGATTACCGGTTTCACCAAGAGCAAGTAGATAGATTACGTGAAGAGGT  
TGGTATTACCTGATTGGTACAGGTCGTGTTGTATGGCGGGTGTCAATAACAATAAT  
GTGCAGCGATTGCACAAGCCTTGCAGCTGTCAGTTAA

Protein sequence:

VFQQVVEAFAGDPILSLMDVYNKDRQDKINLSIGLYYDEEGKTPILTVSVARQLNAMT  
PTATLYLPMEGLAPYRHEVQTLLFGADNPLIADKKIATIQLGGSGALKVGADFLHRYFPS  
SEWWISDPTWDNHASIFAGSGFKVNYYPYFDPETKGVKFDALIDCFKKLPEKSIVLMHPC  
CHNPTGSDLTKAQWDQVIEILKARQAIPFLDIAYQGFAESLDDAYAVRAMAKAGLPVL  
VSNSFSKIFGIYGERAGGLSIVCDNAKECEHVLGQLKAGARRIYSSPANYGAQLVNQVLS  
DHVLTAQWQKEVAHMRDRIKEMRVTLVNALKEALPEKNFDHLLTQRGMFSYTGSPEQ  
VDRRLREEVGIYLIGTGRVCMAVGNNNNVQRIAQFAAVS\*

L-aromatic amino acid transaminase(AAAT-2)

Coding DNA sequence:

ATGTTGAGAAAATCATTGCTGCACCAGCCGATCCTATTCTGGGTTAGCTGATAGTT  
TCCGTTCTGATACTCGTAAAACAAAATTAACCTAGGGATTGGTGTATAAGGATGA

AACAGGTAAAACACCTGACTCACCAACCGTTAAAAAAGCAGAAAATATTATTAGA  
AAACGAATCCACTAAAAATTATCTCCTATTAGCGGTATTCCCTGAGTTGGTGCACGT  
ACTCAAGCACTGTTGGTGAACAACACCCATCATCACAGAAAACGTGCACGT  
ACAGCACAAGCTCCGGTGGTACTGGGCATTACGTATTGCTGCCACTTATTGCTC  
AGCAAACATAATGCTAAACGTGTATGGATAAGTAACCCGACTTGGCCAACCATAATA  
ATATTTCAGACTGCAGGCTTAGAGATTGCCAATATGATTATTGATGCCGAAAG  
CCACGGTTAGATTGAAGGTATGCTAGCCAGCCTACAAATGCACAAGCAGGTGA  
TGTGCTGCTGTCACGGATGTTGCCATAACCCAAAGTGGTATTGACCCAACGCTAGAG  
CAATGGCGTCAACTGGCTGCCTATCCGCAGAGAAAGGTTGGTACCAAGTATTGACT  
TTGCTTACCAAGGTTGCTAATGGCTTGAAAGAACGCGCAAGGCTACGCCATT  
TGCTGAAAGCAATCCTGAGCTGATTGTCAGCTCATATTCAAAAAACTTGGCCTT  
TATAATGAGCGTGTAGGTGCTGCACATTGTCGAAAAGAGAGCGATACTGCTGAG  
AAAGCATTAGCCAAGCCAAAGCGATTATCGCGAAACTACTCTAATCCACCGCA  
CATGGTGCATCTGTGGTGACTIONTATTCTAACCCAGAGTAAAGAAGAGTGG  
TTGAAGAGCTACCACAATGCGTGAGCGCATTAGCGTATGCGCCAATTGCTGGTGA  
CAACGTTACAAGAAAAAGCGCAAAACAAGATTCAAGCTTATTATTGACCAAAATG  
GTATGTTCTCATTTAGTGGCTTAATAAAGAGCAAGTTGAACGTCTACGTGCTGAATA  
TGGTATTATATTGTCGGCTCTGGCGTATTAATGTCGAGGCTAACACTGGAAAAT  
ATGGTCCATTGTGTGAAGCCATTGTCGCTGACTTAA

Protein sequence:

MFEKIIAAPADPILGLADSFRSDTRENKINLGIVYKDETGKTPVLTVKKAEKYLLENES  
TKNYLPISGIPEFGAVTQALLFGEQHPIITEKRARTAQAPGGT GALRIAADFIAQQTNAKRV  
WISNPTWPNNHNNIFQTAGLEICQYDYYDAESHGLDFEGMLASLQNAQAGDVVLFHGCCH  
NPSPIDPTLEQWRQLAALSAEKWLPVFDFAYQGFANGLEEDAQGLRLFAESNPELVAS  
SYSKNFGLYNERVGACTIVAKESDTAEKAFSQAKAIIRANYSNPPAHGASVVTILSNPEL  
KEEWIEELTTMRERIQRMRQLLVTLQEKGAKQDFSFIDQNGMFSFSGLNKEQVERLRA  
EYGIYIVGSGRINVAGLTMVPLCEAIVAVL\*

L-amino acid deaminase (Pm1)

Coding DNA sequence:

ATGGCAATAAGTAGAAGAAAATTATTCTTGGTGGCACAGTGGTGTGCTGCTGCAG  
GCGCTGGGATTITAACACCTATGTTAACCGCGAGAAGGGCGTTTGTCCCTGGTACGCC  
GAGACATGGTTTGTGAGGGAACCTGGCGGTCCATTACCGAAACAAGATGATGTTGT  
TGTAATTGGTGCAGGTATTAGGTATTATGACCGCGATTAACCTGCTGAGCGTGGC  
TTATCTGTCACAATCGTGAAAAGGAAATTGCCGGCGAACATCATCGCGATTCT  
ATGGTCAAGCTATTAGCTATAAAATGCCAGATGAAACCTCTTATTACATCACCTCGG  
GAAGCACCCTGGCGTGAATGAACGCTAAAGTTGGTATTGATACCACTATCGTAC  
ACAAGGTCGTAGAAGTCCTTAGATGAAGAAGATTAGAAAACGTAAGAAAATG  
GATTGATGCTAAAGCAAAGATGTTGGCTCAGACATTCCATTAGAACAAAAATGAT

TGAAGGCGCTGAGTTAAAACAACGTTACGTGGCGTACCACTGATTGGAAAATTGC  
TGGTTCGAAGAAGACTCAGGAAGCTCGATCCTGAAGTTGCGACTTTGTGATGGCA  
GAATATGCCAAAAAAATGGGTATCAAATTTCACAAACTGTGCAGCCCGTGGTTA  
GAAACGCAAGCTGGTGTATTCTGATGTTGTAACAGAAAAAGGACCAATTAAAACC  
TCTCGTGTGTCGCCGGTGGTGTACGTTATTATGCAGAACCTAAATGT  
TGATGTACCAACATTACCTGCTTATCAATCACAGCAATTAAATTAGCGCAGCACCAAT  
GCGCCAGGTGGAAACGTTGCTTACCCGGCGGAATTTCTCCGTGAGCAAGCGGAT  
GGAACGTATGCAACTCTCCTCGTGTATTGCTCCGGTAGTAAAAGAACATCATT  
CTTACGGCTATAAATATTACCTCTGCTGGCTTACCTGATTCCCAGTACATATT  
TTAAATGAGCAGTTGATTAATTCTTATGCAATCAACACATTGGGATCTTAATGAAG  
AGTCGCCATTGAAAAATACGTGATATGACCGCTCGCCTGATCTGCCAGAATTAAA  
TGCCTCACTGGAAAAACTGAAAAAGAGTTCCCAGCATTTAAAGAACATCAACGTTAAT  
TGATCAGTGGAGTGGTGCATGGCGATTGCACCAGATGAAAACCAATTATCTCTGA  
TGTAAAGAGTATCCAGGCCTAGTTATTAAACTGCAACAGGTTGGGAATGACCGA  
AAGCCCTGTATCAGCAGAAATTACAGCAGATTATTAGGCAAAAAACCTGTATT  
AGATGCCAAACCATTAGTCTGTATCGTTCTAA

Protein sequence:

MAISRRKFILEGGTVVAVAAGAGILTPMLTREGRFVPGTPRHGFVEGTGGPLPKQDDVVVI  
GAGILGIMTAINLAERGLSVTIVEKGNIAGEQSSRFYQQAISYKMPDETFLHHLGKHRWR  
EMNAKVGIDTYRTQGRVEVPLDEEDLENVRKWIDAKSKDVGSDIPFRKMIEGAELKQ  
RLRGATTDWKIAGFEEDSGSFDPPEVATFVMAEYAKKMGIKIFTNCAARGLETQAGVISDV  
VTEKGIKTSRVVVAAGGVWSRLFMQNLNVDPVTPAYQSQQLISAAPNAPGGNVALPGG  
IFFREQADGTYATSPRVIVAPVVKESFTYGYKYLPLLALPDFPVHISLNEQLINSFMQSTH  
WDLNEESPFEKYRDMTALPDLPNELNASLEKLKKEPPAFKESTLIDQWSGAMAIAPDENPII  
SDVKEYPGLVINTATGWGMTESPVAEITADLLGKKPVLDAKPFSLYRF\*

L-aromatic amino acid decarboxylase (AADC)

Coding DNA sequence:

ATGGCAAATCACAAAGAAACCTGTCTGAGTGCTGCAGAGCGTGTGGTGTACCCGA  
ACCGTGGATACCTAACATGTTATCATCGAGAGATGGAAAGATGTTGATGGATAGT  
GAGTCCTTTATTCTAGTCCAATAACTCACAGATAAAACTAATGAGAGATTAGAGA  
ACAAAATAAAAGCAATACCATGAAAATAAGCTGTGAAGGGATCTTATGGTATA  
TTTAAATAATTAAACCAGCAGCTAATGAATAATAACCTGAGATTGAGCCTTATT  
CTCACATTACAACCTCCCGTCATGCGATTGGCAAATTAGGCTACCTTCTGCACTG  
GCCTATAATGCTGATAATGTTTAACAAACATTACCCATACAAAAACAGCAAGAG  
CAAAAATAACTCATGATTGTAAACTAACATCGGTTGAACCAGATAAAGCATTG  
GGCATATTACCAACCAGTCATGTTTGCTGTTATGAAATTGTGGGTTGAGAAAT  
TTAAAAACACTGCCGATGGCGATTGCACGTACCAAAATCAAGAGATCTGGTTGCA  
GATAAAAAGCATTTGAGTTATTAAATATGTCTACTGATATTATCCATCAGTG

AACAGCTATATGAGCGAGATATTTGATGATGTCAGCCATTAACTTGTGCGTGGCAC  
CGGCATGACAAGAACGATGCATTAGGTAATTACTAGTACCGATATACCGCTTCGA  
ATTTGGAAGAAGGCGATGGATATCTTGGGGTAGGTTATGATAATCTCATTGCCTTA  
CCAATAGATGAAGAGTTAAAACAGATATAGCTAAAACGCGTCATATTGTGCTTCGC  
TTAATAGAGCAGGGAGAGCCGATACTCGGTATTATTGGCGTGTAGGCTCATCGACA  
TACGGTAGCATTGACAAACTTAAGCCTTTCAATTACGTGAAGAGTGTGAAAAAA  
CAACATAACAACTCATTCTATATTCACATTGATGCATCCCAATTAGGCTATATGAAAAA  
GCCTGTTCTGACGAAAATAATAGCATCATCCCTATAACATATTGTGCGATAAAACT  
TAAGCAAGAACGACCTTACTTGAGTTAACCTCAGATATTATGATAGTTTACTCAG  
CTATCAAGAGCTGATTCTGTTCTTGAGCCATTCAAGCGGGATTTCCTTATCC  
TACGGGGATTGTTGTATAAAAGATGCTCGATTAAGTCGTTCTCGCTAACCGCCA  
AAACTGTGCCAGAACAGAGGGAAATTCCGATGAAATTGATGGGATCCAATCAGCA  
CCTGCGGTTGCATCAATGTGGGGATCCATCAGCTATATGCTTTAACATTGGGT  
ATGGGCAATGCGCTCAAATCAATGGGAAGCTAGAGTAAAACGAAACAACAAATT  
AAACAAGAGAACGAATTGAAAAATCGGGTATCTCTTATGCTATTCAATTGGCCAG  
TTTCAGATTAAATAAAACTCACTTTGCCATTGCTCTAAAGGCAACCAATGTTAAAA  
ATACAAAATCAGCTAAATAATGCTATTATAAAATTAGAAATAAGAGCCAATAT  
AAGCCAGATCTCACCTATTAGTGCTGTGCTCGTAATAGCGATGATACCTGCTC  
AGTTTGTCAACAATGTCATTAGTAAAGATGAATGGAAGCAAGTTAACCACTTAA  
GTCTTTACAATTAAACGATAAAAACACTGATAATGGAGTGAAGAGCATATTAAAG  
CAGGGTACGACTATATTAAAAAGTTGTCCTATCTGCATTAAATAG

Protein sequence:

MANHKETCLSAEAEVFDFRTVDTLTCYHREMVKMLMDSESFYSSPNNSQINTNERLENK  
IKSNTIENKAVKGYSYGDILNNLNQQLMNNKPEIEPYFSHYNSPVMRLAKLGYSALAYNA  
DNVFNKHSPIQKQQEQKITHDLCKLIGFEPDKAFGHITSHVFACYEILWALRNLKTPMA  
IARHPKSRLDVADKKAFELFNMSITDILSISEQLYERDIFDDVSHLTCRGTGMTRTMHLGK  
LLVPIYRFEFWKKAMDILGLGYDNLIALPIDEEFKTDIAKTRHIVRLIEQGEPILGIVLGS  
STYGSIDKLKPLFELREECEKQHNNSFYIHIDASQLGYMKSLFLDENNSIIPYNILCDKLKQ  
EAPLLELTSDIYDSFTQLSRADSVSFEPFQAGFSPYPTGIVCIKDARLSRFLANPPKLCPEEG  
NIPDEIDGIQSAPAVASMWGIHQLYAFNHFGYQCAQNQWEARVKLEQQLQENEFEKS  
GISYAIHILPVSDFNKLTFAIALKGNQCLKIQNQLNNAIYKNLEIKSQYKPDLNLLVLCARN  
SDDIPAQFCQQCHLGKDEWKQVNHLSSLQTIKNTDKWSEEHIKAGYDYIKKVVLALK\*

adenosylmethionine-8-amino-7-oxononanoate transaminase (DANS)

Coding DNA sequence:

ATGACACCCGATGATATTGCTTGATTACGCCATTGGCACCCCTATACCTCGAT  
GAGTAACCCATTACCGGCTTATCCAATCGTAAGTGCAGAAAGGGTAGACTCACTTA  
GCGAATGGCAAACAATTAAATTGATGGTATGTCTCATGGTGGCTGCGATCCACGGCT  
ATAACCACCTGAACCTCAATACGGCGGTAACAGAACAACTAGCAGCAATGTCTCATG  
TGATGTTGGTGGTATCACTCATCCCCAGCGGTTGCCATTGTCGAAAACATTGGC

TATCACTCCCACCCCACTAGAGTGTATTCCTCGCTGATTCAAGGCTCTGTAGCGGTTG  
AAGTGGCGATAAAATGGCATTGCAATATTGGCAGGCAAAAGGTAAAAACGCCAG  
CGCATTGTGGCATTAAAACGCAGGTTATCATGGTGATACATTGGCGCGATGTCAGTGT  
GTGATCCTGACAATTGATGCATAGTCTACAAAGGCTACCTGCCAATCATCTGTT  
TGTCGAAGCGCCAAAACCGGATTATCAACCCTGGGACGCTACTGATATAGACGC  
TTTACGCACCACATTAGCACACATCATCAACATATTGCCGCCGTGATGCTAGAGCCA  
ATCGTCCAAGGAGCTGGTGGTATGCGCATATACCACCCAGAGTATTAAACACAAGCT  
CGAGCGTTATGTGATGAATTAAATGTGTTATTAAATTGCCGATGAAATGCCACTGGAT  
TTGGCAGAACCGTAAATTATTGCTTGTGAACACGCCGGATCTCCCCGATATTAT  
GTGTGTAGGTAAAGCATTAAACAGGTGGCTATATGACGCTATCCGCCACGTTAACTAC  
ACGCCATATTGCCGATACAATCAGCAAGGTGATGCAGGTTGCTTATGCATGCCCT  
ACCTATATGGGAATCCTCTGCCGTGCGGTAGCGAATGCTAGTTATCACTATTAG  
AGCAAGGTCACTGGGTTAATCAAGTCGCCAAATTGAAGATCAACTAAAACCGAAT  
TGTTACCCCTAAAGCAAGCAAAAAGCGTTAAAGATGTGCCGGTACTTGGTGTATTG  
GTGTCGTAGAAATGGTTGAGCCAGTCAATATGGCTAAACTACAGAAATACTTGTCA  
ATGAAGGGTTGGATACGACCTTGGTCAACTGATTATATTATGCCCTTATATT  
ATATCACCTGAAAAGCTCACTAAACTCACTCAAGCAATTGAGAAAGCAGTTAATCTT  
ACCAATTAA

Protein sequence:

MTPDDIAFDLRIWHPYTSMSNPLPAYPIVSAKGVELTLANGQLIDGMSSWWAAIHY  
NHPELNTAVTEQLAAMSHVMFGGITHPPAVALCRKLLAITPTPLECIFLADSGSVAVEAI  
KMALQYWQAKGEKRQRIVALKRGYHGDTFGAMSVCDPNSMHSLYKGYLPNHLFVEA  
PKTGFYQPWDATDIDALRTTAAQHHQHIAAVMLEPIVQGAGGMRIYHPEYLTQARALCD  
EFNVLLIADEIATGFGRGKLFACEHAGISPDIMCVGKALTGGYMTLSATLTTREHIADTISQ  
GDAGCFMHGPTYMGNPLACAVANASLSLEQGHWVNQVAQIEDQLKTELLPLKQAKSV  
KDVRVLGAIGVVEMVEPVNMAKLQKYFVNEGWWIRPGQLIYIMPPYIISPEKLTKLTQAI  
EKAVNLTN\*

### *N*-acetylornithine aminotransferase (ACOAT)

Coding DNA sequence:

ATGATAAAATCAACGTATTAATAGGGCAACTTACGATCAGGTAATGTTGCCAATTATT  
CACCAAGCAGATTTATTCCCTGTTAAAGGGCAAGGTAGCCGTGTTGGGATCAACAGG  
GAAAATCCTATATTGATTTGCTGGTGGTATTGCTGTTCTAGCGCTTGGGATGCGCA  
TCCACTATTGAATAAACGCACTGCAAGAGCAAAGTGAGAAGCTTGGCATGTAAGTAA  
TGTGTTACTAATGAGCCGGCATTATGTTAGCTCAGAAATTAAATCCAACATACTTT  
GCTGAGCGTGCTTTTGCCAATTCTGGGGCAGAAGCGAATGAAGCCGCACTTAAAT  
TAGCACGACATTACGCTATTACCCGTATAATCCTTATAAAACGAAGATAATTGCCCT  
CCACCATGCCTTCATGGTAGAACATTTCAGTCTGTAGGGGGACAACCTAA  
TATGCGGATGGTTGGCCTAACCTGCGGATATTACATGTCCCTTAACTCTCT

TGATGCTGTAAAAGCCGTATTGACTCTCATACTTGTGCCGTAAACTGGAGCCTATC  
CAAGGTGAGGGCGGAGTGACACCGCGACAGCGGAGTTATGCATGGCTTACGTACA  
CTATGTGATAAGCATCAAGCGCTATTAAATTCTTGATGAAGTACAAACGGGTATGGGG  
CGAACCGGCAAACATTGCTTATATGCACTATGATATACAGCCCATTATTACCA  
GTGCCAAAGCATTAGGTAGTGGTTTCCTCTCAGTGCATGTTAACACCAGAGGATAT  
CGCTTCGGTAATGGGGTAGGAACACATGGTACGACTTATGGGGGAACCCATTAGC  
CTGTGCAGTAGGCAATGCTGCTTGATTAATCAACACTCAGGCAGTATTAGAAGGT  
GTCAGTGAGCGTCATCGTGGATAGTGGATGAATTGCAGGCAATAACCAAACCTAT  
GCTGTATTTCCAAATTGTTGACAAGGCTTATTAATTGGCGCGCAGTTAGCACAC  
TGTATGCAGGTAAAGCAAAAGAGTTACTGCCTAGGCCTCTAACATGGCTTAATGA  
TGTTAAATGCCGGGCTGATGTAATACGCTTACACCATCGCTGATTATTACCTCGA  
AGAGTTACAAGAGGGCATGAATGCGTTACGAAAGCGCTTGAGTTGTAAAAGC  
AGAGTGA

Protein sequence:

MINQRINRATYDQVMLPIYSPADFIPVKQGSRVWDQQGKSYIDFAGGIAVLALGHAPL  
LNKALQESEKLWHVSNVFTNEPALCLAQKLIQHTFAERVFFANSGAEANEALKLARH  
YAITRHNPYKTKIIAFHHAFHGRTFFTJVGGQPKYADGFGPKPADIHPVFNSLDAVKAV  
IDSHTCAVILEPIQGEGGVTPATAEFMHGLRTLKDQHQALLILDEVQTGMGRGKLFAYM  
HYDIQPDIITSAKALGSGFPLSAMLTEDIASVMGVGTHGTTYGGNPLACAVGNAAFDLI  
NTQAVLEGVSERHRWIVDELQAINQTYAVFSQIRGQGLIGAQLAPLYAGKAKELLALAS  
KHGLMMLNAGADVISRFTPSLIITLEELQEGMNALRKALAEFVKAE\*

glutamate-1-semialdehyde 2,1-aminomutase(GSA)

Coding DNA sequence:

ATGAGCAAGTCTGAAACGCTTATAATCTGCACAAACAAGTGATAACCTGGTGGCGTC  
AATTCAACGGTTAGAGCATTAAACGGTGTGGCACACCGCTTTTATTGAACGTG  
CTAATGGTGCATATTGATGCCATGGACGTGCTTACTTGACTATGTCGGTTC  
ATGGGGACCGATGGTATTAGGCCACAATCACCCGCTATTGCCACGCTGTAAGTGAT  
GCGGTGAAAAAGGCCTAACGCTTGGTGCCTACGGCTGCTGAAGTGGAAATGGCA  
AATTAGTCACTGAGTTAGTGCCTCGATGGATATGGTCGTATGGTGAACTCAGGTA  
CAGAACGCCACCATGAGTGCCTCGCTGGCTACACGGGCGTGACAAAAA  
TCATTAATTAAGGCTGTTACCATGGACATGCAGACTGCTTATTAGTCAGCAGG  
ATCTGGTCATTGACGATGGACAACCAAACCTCCAGGCGTCTGCTGATTTGTT  
AAACACACTTAACTGTACTTACAACGATTAAATTCAAGTACGCTAACAGCATTGAAA  
ATTATCAGGAAGAAATCGCCTGTATTATTGTTGAACCCGTTGCCGGTAATATGAAC  
TGTTCACCTAAAGCCGATTTCTACCGGATTACGTGCATTATGTGATGAATTGGT  
GCATTACTGATTATTGATGAAGTAATGACAGGTTTCGTGCGCATTAGGCAGCGC  
AAGCTTATTACGATGTTGATCCTGACTTGACCTGTTAGTAAATTATAGGTGGTGG  
CATGCCCTGTAGGCGCCTCGCGGACATAAGAAGTGATGTCTCAATTAGCCCCTATC  
GGTCCTGTCTACCAAGCGGGAACATTATCGGTAACCCATTGCGATGGCGCGGGT

CTAGCTTGTGCAAGAAGTTCACAAACCGGGAGTCACCAAACACTGGATGAACCTTA  
CTACTATGTTAGCGGATGGCTTACTTGAAAAAGCACAACAAGCAGGTATCCCTATGG  
TAGTAAACCATGTAGGCGGTATGTTGGTTATTCTTACCGATGCCAAAGAAGTGAC  
TTGCTATCAAGACGTAATGAACACTGCGATGTCGAACGCTTAAACACAGTTCTTCACTTA  
ATGTTAGAAAAACGTATTACCTCGCTCCGTCGGCTTTGAAGCAGGCTTATGTCTA  
TCGCTCATAGCAAAGAAGATATTCAACGTACTATCGATGCGGCTGAATATGCCTTAG  
CAAAATGAAAGCGTAA

Protein sequence:

MSKSETLYNLAQQVIPGGVNSPVRAFNGVGGTPLFIERANGAYIYDADGRAYLDYVGSW  
GPMVLGHNHPAIRHAVTDQVKGLSGAPTAEEVEMANLVTTELVPSMDMVRMVNSGT  
EATMSAIRLARGYTGRDKIIFEGCYHGHADCLLVKAGSGALTMGQPNSPGVPADFVKH  
TLTCTYNDLNSVRQAFENYQEEIACIIVEPVAGNMNCVPPKADFLPGLRALCDEFGALLII  
DEVMTGFRVALGGAQYYDVDPDLTCLGKIIGGGMPVGAFFGHKEVMSQLAPIGPVYQ  
AGTLSGNPIAMAAGLACLQEVSQPGVHQTLDELTTMLADGLLEKAQQAGIPMVNVHVG  
GMFGLFFTDAKEVTCYQDVMNCDVERFKQFFHLMLEKRIYLAPSAFEAGFMSIAHSKEDI  
QRTIDAAEYAFSKMKA\*