ISSN 0233-7657. Біополімери і клітина. 2004. Т. 20. № 1-2

Molecular cloning, sequencing and sequence analysis of *Thermus thermophilus* tyrosyl-tRNA synthetase

A. D. Yaremchuk[1, 2], **O. P. Kovalenko**[1], **O. I. Gudzera**[1], **M. A. Tukalo**[1, 2]

[1] Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine
150 Acad. Zabolotnoho str., Kyiv, 03143, Ukraine

[2] EMBL Grenoble Outstation
156X, 38042 Grenoble cedex 9, France

The gene encoding tyrosyl-tRNA synthetase (TyrRS) from the extreme thermophilic eubacterium T. thermophilus HB27 has been cloned and sequenced. The open reading frame encodes a polypeptide chain of 432 amino acid residues in length (molecular mass 48717 Da). Comparison of the amino acid sequence of the T. thermophilus TyrRS (TyrRSTT) with those of TyrRS from various organisms shows that T. thermophilus enzyme shares a branch in the philogenetic tree of eubacterial TyrRSs with the enzymes from Aquifex aeolicus, Deinococcus radiodurans, Haemophilus influenzae and Helicobacter pyroly (40—57 % amino acid identity), distinct from the branch containing Esherichia coli, Chlamydia trachomatis and Bacillus stearothermophilus, for example (24—28 % amino acid identity). The TyrRS active site domain is highly conserved, whereas a C-terminal tRNA binding domain contains only few conserved residues. But even in the active site exists one very important difference between the two groups of bacterial TyrRSs: Lys-41 in TyrRSTT (and in TyrRS from many human pathogenic bacteria) is conserved as a tyrosine in another group of bacterial TyrRSs and eukaryotic sequences including human. This knowledge could be exploited in the design of new antibiotics.

Introduction. The aminoacyl-tRNA synthetases (ARSs) are highly diversified enzyme family that catalyze the ligation of cognate amino acids to their cognate tRNAs. For most ARSs this reaction proceeds via a two-step process. In the first step of the aminoacylation reaction, the amino acid is activated by ATP to form an enzyme-bound aminoacyl adenylate intermediate. Then, in the second step, the aminoacyl moiety is transferred to the 3'-terminal adenosine of the cognate tRNA.

Generally, but with some exceptions, all cells or organelles in which there is protein biosynthesis have a complement of 20 enzymes. These enzymes are divided into two quite distinct structural classes on the basis of primary structure and the fold of the catalytic domain [1, 2].

The class I synthetases possess a catalytic domain that is the Rossman dinucleotide-binding fold domain which contains the signature sequences «HIGH» and «KMSKS».

The class II enzymes have a catalytic domain consisting of seven anti-parallel β-strands and contains the three class II-defining motifs. Tyrosyl-tRNA synthetase (TyrRS) is a homodimeric class I aminoacyl-tRNA synthetases. This enzyme is unique among all aminoacyl-tRNA synthetases in having two types of $tRNA^{Tyr}$: with a long variable loop for prokaryotes and eukaryotic organelles and with a short variable loop for archaea and eukaryotes. Also, the acceptor stem of $tRNA^{Tyr}$ of prokaryotes, mitochondria and chloroplasts have the G1-C72 base pair found in most tRNAs while the first base pair of $tRNA^{Tyr}$ of eukaryotic cytoplasm and archaea is C1-G72 [3].

Eukaryote cytoplasmic and prokaryote tyrosyl-

tRNA synthetases can not cross-aminoacylate their respective $tRNAs^{Tyr}$. It has been shown that interchange of the first base pair is sufficient for the species-specific aminoacylation [4]. Knowledges of the structural basis for such kind of co-adoptation of a synthetases to tRNAs is important for understanding of the origin of the genetic code and specificity of synthetase-tRNA recognition and also can be used for drug discovery. Therefore we cloned the *tyrS* gene of *T. thermophilus* as part of structural study of TyrRSTT and its complexes with substrates. Here we report the cloning, sequencing and sequence analysis of *T. thermophilus* tyrosyl-tRNA synthetase.

Materials and Methods. Restriction endonucleases, T4 DNA ligase, bulk *Escherichia coli* tRNA, lysozyme, the digoxygenin DNA labeling and detection kit were from «Boehringer» (GFR), Tub DNA polymerase and [^{35}S]-dATP[S] from «Amersham» (Great Britain), sequence version 2.0 DNA sequencing kit and Tag cycle sequencing kit from US Biochemical Corp. pCR2.1-TOPO vector was from «Invitrogen» (USA).

TyrRS was purified from *T. thermophilus* HB27 cells as described [5]. Genomic DNA from *T. thermophilus* cells was purified by the method of Marmur [6]. The amino acid sequences of the N-terminal peptide and three internal peptides of the purified TyrRS were determined by the Protein and the Peptide group at EMBL, Heidelberg, by microsequencing. Appropriate oligodeoxyribonucleotides were purchased from Genosys. The polymerase chain reaction (PCR) was carried out for 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 50 °C and 1 min elongation at 72 °C in 100 μl reaction buffer containing 50 mM Tris-HCl, pH 9.0, 1.5 mM $MgCl_2$, 20 mM ammonium sulfate, 1 μl genomic DNA from *T. thermophilus* HB27, 0.2 mM dNTP, 40 pmol N-terminal primer, 40 pmol internal primer and 2.5 U Tub DNA polymerase. Both strands of the *tyrS* gene were sequenced by the dideoxynucleotide chain-termination method [7] using [^{35}S]-dATP[S] and the Sequence version 2.0 DNA sequencing kit. To overcome the problems associated with the high G-C content of DNA, the ΔTaq cycle sequencing kit was used.

Results and Discussion. *Cloning and sequencing of the T. thermophilus tyrS gene.* The purified TyrRSTT provided several short peptide sequences: an N-terminal sequence of 20 amino acid residues and several internal tryptic peptides, which were determined at EMBL, Heidelberg, by the Protein and the Peptide group. Using sequence information from an N-terminal sequence (AGTGHTPEEALALLKRGAEE) and one internal tryptic peptide (YEAGIPISLLVELLYPFAQ) two PCR primers (5′-GCSGGSACGGSCACACSCCSGAGGA-3′ and 5′-GATSGGRATSCCSGCCTCGTA-3′) were designed taking into account the preferential codon usage of *T. thermophilus*, with the third base of each codon being G or C. With these two primers, a partial gene fragment (526 bp) of TyrRSTT was amplified by polymerase chain reaction. That this fragment corresponded to a putative *tyrS* gene was verified by cloning into pCR2.1-TOPO vector and DNA sequencing. The sequence analysis clearly indicates that this fragmant is a 5′ part of the *tyrS* gene. Furthermore, the translated open reading frame shows significant sequence similarities with tyrosyl-tRNA synthetases from other sources.

The PCR fragment was labelled with digoxygenin and used for Southern blot hybridization to *T. thermophilus* genomic DNA digested with several restriction enzymes. The 1350 bp *XmaI* fragment was hybridized to the probe. The fragment was cloned into the appropriate sites of plasmid *pUC19*, and genomic sublibrary was constructed in *E. coli* XL1-Blue MRFB. The positive clones were screened from the genomic sublibrary by plaque hybridization with the same probe. The 1350 bp *XmaI* fragment was sequenced and found to contain a full length DNA of the *T. thermophilus tyrS* gene. The open reading frame of the *tyrS* gene is composed of 1296 bp, from which the sequence of 432 amino acid residues comprising one subunit of *T. thermophilus* TyrRS was deduced (fig. 1). The calculated relative molecular mass per subunit (48717 Da) is in agreement with that estimated by SDS-polyacrylamide gel electrophoresis (50000 Da) for the purified TyrRS from *T. thermophilus* cells [5]. From amino acid composition, the isoelectric point of 6.07 and a molar extinction coefficient at 280 nm (ε) of 32550 $M^{-1}cm^{-1}$ ($E^{mg/ml}$ = = 0.67 $ml \cdot mg^{-1} \cdot cm^{-1}$) were determined for the subunits.

Sequence analysis of TyrRS. Comparison of the amino acid sequence of the *T. thermophilus* TyrRS with those of homologous enzymes from various organisms shows that *T. thermophilus* TyrRS shares a branch in the phylogenetic tree of eubacterial TyrRS with the enzymes from *Aquifex aeolicus*, *Deinococcus radiodurans*, *Helicobacter pylori* and *Haemophilus influenzae*, distinct from the branch containing *E. coli*, *Bacillus stearothermophilus*, *B. subtilis*

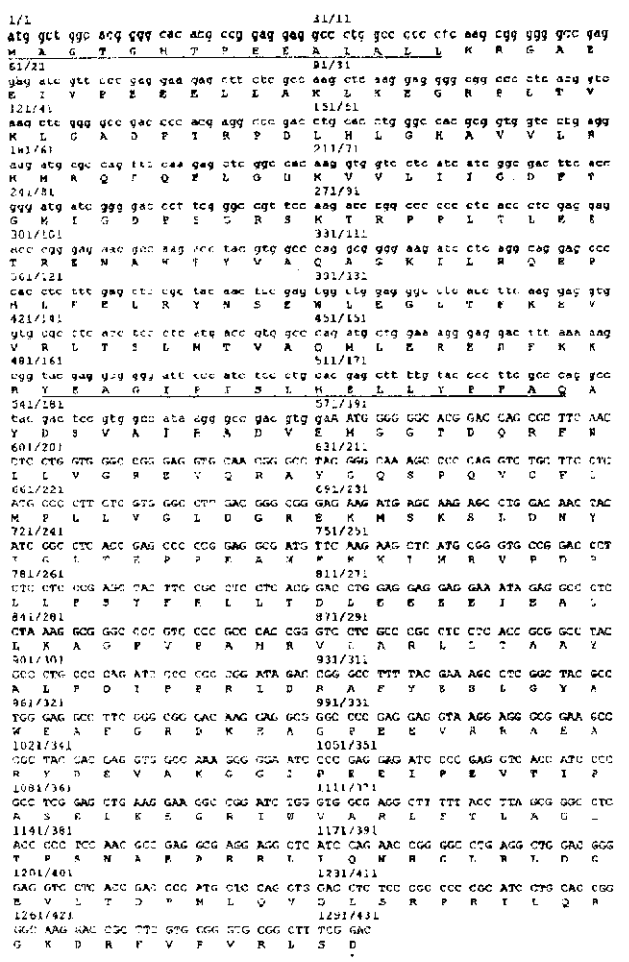


Fig. 1. Nucleotide sequence of the *T. thermophilus tyrS* gene and the deduced amino acid sequence of tyrosyl-tRNA synthetase. The amino acids underlined correspond to the peptide sequences determined by protein sequencing

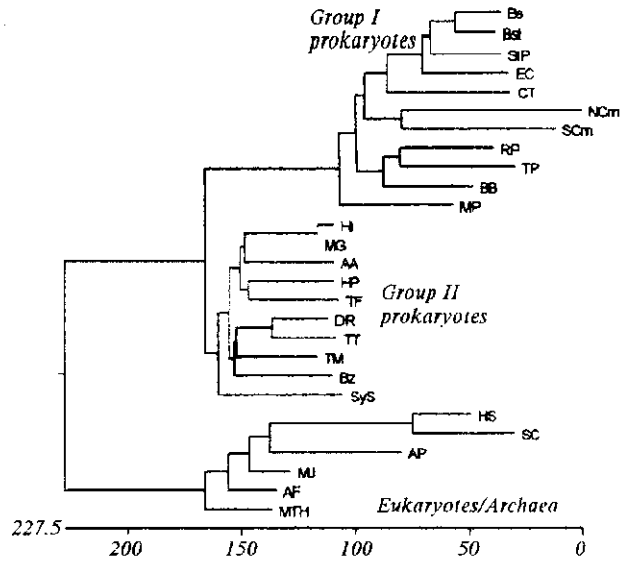
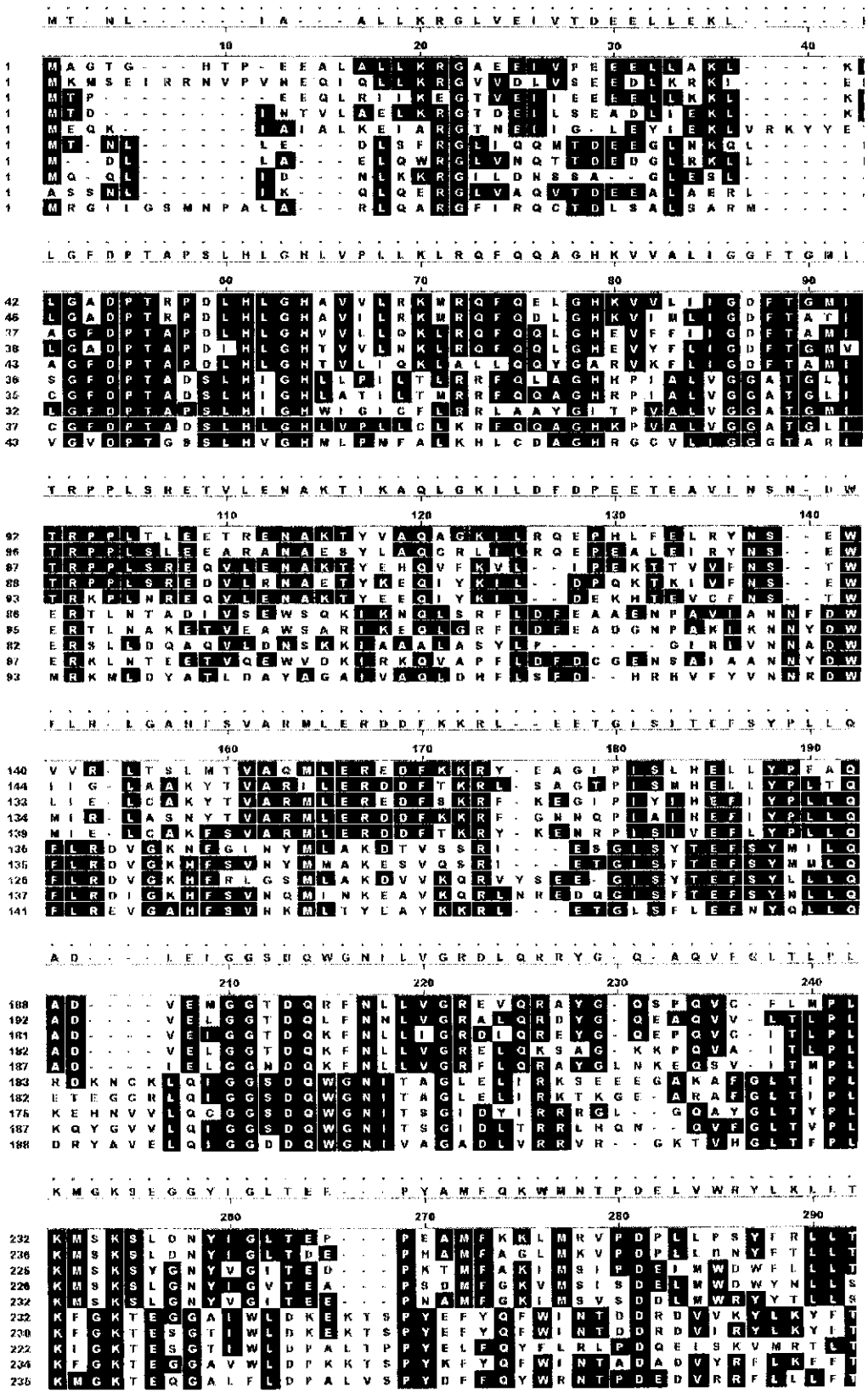


Fig. 2. Phylogenetic tree of TyrRS sequences. The tree has been rooted between the Bacteria including mitochondrial TyrRS and the *Archaea* plus *Eukarya*. Abbreviations: *AA — Aquifex aeolicus*; *AP — Aeropyrum pernix*; *AF — Archaeoglobus fulgidus*; *BB — Borrelia burgdorferi*; *Bs — Bacillus subtilis, tyrS* gene; *Bz — Bacillus subtilis, tyrZ* gene; *Bst — Bacillus stearothermophilus*; *CT — Chlamydia trachomatis*; *DR — Deinococcus radiodurans*; *EC — Escherichia coli*; *HI — Haemophilus influenzae*; *HP — Helicobacter pylori*; *HS — Homo sapiens*; *MTH — Methanobacterium thermoautotrophicum*; *MJ — Methanococcus jannaschii*; *MG — Mycoplasma genitalium*; *MP — Mycoplasma pneumoniae*; *NCm — Neurospora crassa*, mitochondrial; *RP — Rickettsia prowazekii*; *SC — Saccharomyces cerevisiae*, cytoplasmic; *SCm — Saccharomyces cerevisiae*, mitochondrial; *StP — Streptococcus pyogenes*; *SyS — Synechocystis* species; *TF — Thiobacillus ferrooxidans*; *TM — Thermotoga maritime*; *TP — Treponema pallidum*; *TT — Thermus thermophilus*. The tree was generated using MegAlign with version 5.1 of DNASTAR package programs. The length of each pair of branches represents the distance between sequence pairs, while the units at the bottom of the tree indicate the number of substitution events

and eukaryotic mitochondrial TyrRS, for example (fig. 2).

The sequence identity between *T. thermophilus* TyrRS and *E. coli*, *B. stearothermophilus* or *B. subtilis* enzymes is relatively low (24—28 %) if compare to that of the enzymes from *H. pylori*, *H. influenzae*, *A. aeolicus* and *D. radiodurans* (40—57 %). Alignment of bacterial tyrosyl-tRNA synthetases shows important sequence identity (about 60 %) in the catalytic domain including the «HIGH» and «KMSKS» motifs (Fig. 3). The α-helical and C-terminal domains which have crucial role in the recognition of class II type $tRNA^{Tyr}$ [8] are less well conserved among all bacterial and mitochondrial tyrosyl-tRNA synthetases (data not shown). The most phylogenetically conserved residues in the two groups of bacterial TyrRSs are located at the junction of the KMSKS loop (residues 190—244 in TyrRSTT). Two lysines (Lys-232 and Lys-234 in TyrRSTT) in the KMSKS motif and the first histidine and glycine (His-52 and Gly-54 in TyrRSTT) in the HIGH motif are strongly conserved in both groups of eubacterial TyrRSs and are important for the binding of ATP [8]. On the other hand, the serine is generally conserved at the third position of KMSKS motif in the TyrRS in members of the same phylogenetic branch as *T. thermophilus*, whereas glycine is found at this position in the second group of the bacterial TyrRSs.



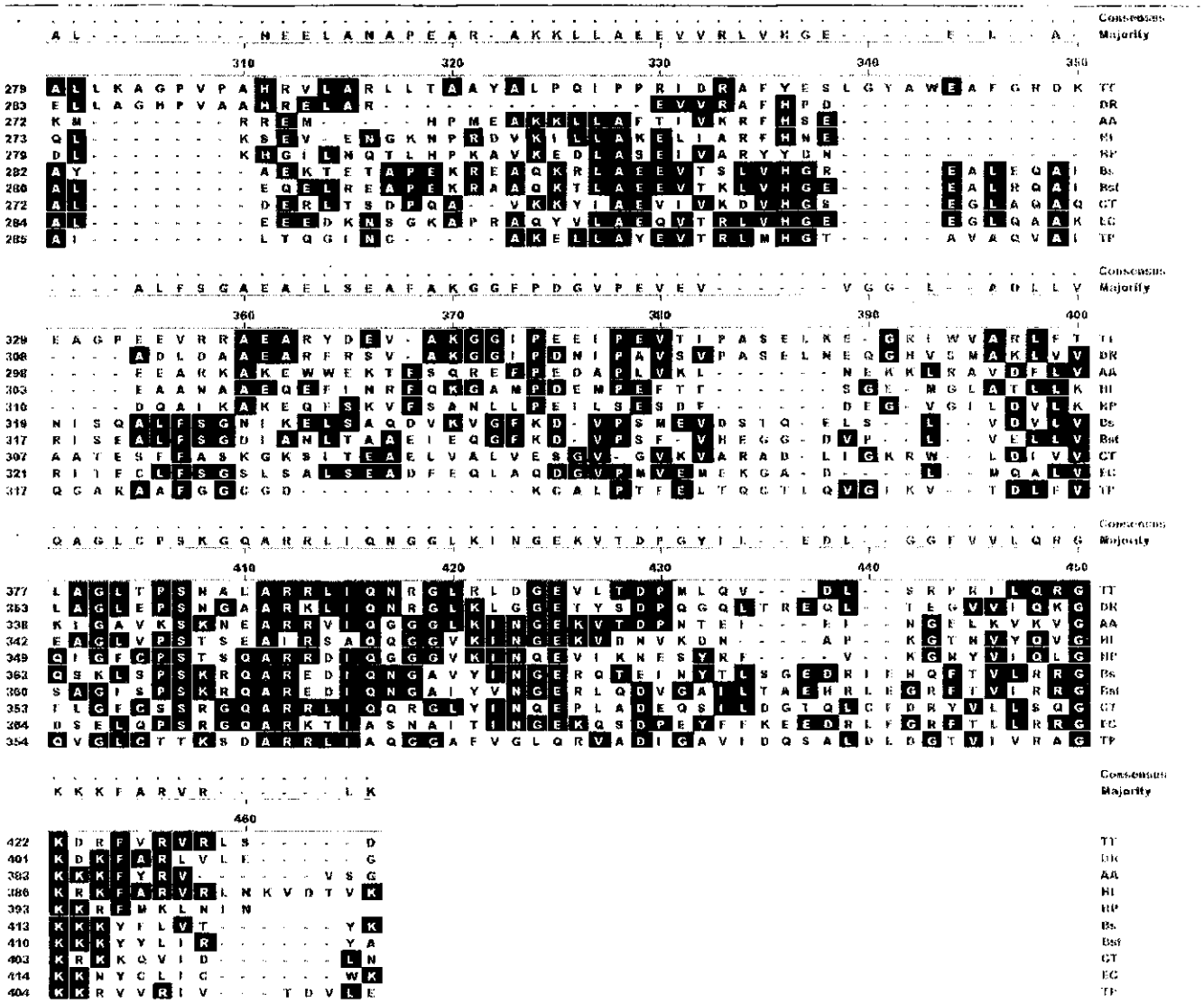


Fig. 3. Alignment of the sequenses of tyrosyl-tRNA synthetases from various organisms. Abbreviations are as in fig. 2. Protein sequences were aligned by the Clustal W program with version 5.1 of DNASTAR package programs

Also, Lys-41 (10 residues before the HIGH motif) is important for the tyrosine binding in TyrRSTT (our unpublished data) and is absolutely phylogenetically conserved. This residue is conserved as a tyrosine in the second group of bacterial TyrRSs (fig. 3) and also in the most archael and eukaryotic sequences including *Homo sapiens* (data not shown). Among organisms of this group prokaryotic TyrRS there are human pathogenic bacteria as *H. influenzae*, *H. pylori*, *Mycoplasma genitalium* and *Vibrio cholerae*. Knowledge of such differences in the catalytically important residues could be exploited for synthesis the compounds that inhibit bacterial TyrRS specifically and could become potent antibacterial drugs.

Bacterial resistance to established antibiotics continues to pose an increasing problem in clinical practice. In this regard, aminoacyl-tRNA synthetases, and in particular tyrosyl-tRNA synthetase, provide a promising platform to develop novel antibiotics that show no cross-resistance to other classical antibiotics [9, 10].

Г. Д. Яремчук, О. П. Коваленко, О. Й. Гудзера, М. А. Тукало

Клонування, визначення та аналіз нуклеотидної послідовності гена тирозил-тРНК синтетази із *Thermus thermophilus*

Резюме

Клоновано та визначено нуклеотидну послідовність гена, що кодує тирозил-тРНК синтетазу (TyrRS) із екстремально

термофільної еубактерії T. thermophilus HB27 (TyrRSTT). Відкрита рамка зчитування кодує поліпептидний ланцюг довжиною 432 амінокислотних залишки (молекулярна маса 48717 Да). Порівняння амінокислотної послідовності TyrRSTT з відповідними послідовностями інших організмів виявило, що фермент із T. thermophilus належить до тієї ж гілки філогенетичного дерева еубактеріальних TyrRS, що й ферменти із Aquifex aeolicus, Deinococcus radiodurans, Haemophilus influenzae і Helicobacter pyroly (ідентичність 40—57 %), але не до тієї, до якої належать, наприклад, Escherichia coli, Chlamydia trachomatis і Bacillus stearothermophilus (24—28 % ідентичності). Амінокислотна послідовність каталітичного домену висококонсервативна, тоді як тРНК-зв'язувальний С-кінцевий домен містить лише невелику кількість консервативних залишків. Але навіть в активному центрі існує важлива відмінність між двома групами еубактеріальних TyrRS: залишок Lys-41 в TyrRSTT (і в TyrRS із багатьох патогенних бактерій людини) представлений консервативним залишком тирозину в бактеріальних TyrRS іншої групи, а також еукаріотичних TyrRS, включаючи людину. Ця відмінність може бути використана при створенні нових антибіотиків.

А. Д. Яремчук, О. П. Коваленко, О. И. Гудзера, М. А. Тукало

Клонирование, определение и анализ последовательности гена тирозил-тРНК синтетазы из *Thermus thermophilus*

Резюме

Клонирован ген, кодирующий тирозил-тРНК синтетазу (TyrRS) из экстремально термофильной эубактерии T. thermophilus HB27 (TyrRSTT), и определена его нуклеотидная последовательность. Открытая рамка считывания кодирует полипептидную цепь длиной 432 аминокислотных остатка (молекулярная масса 448717 Да). Сравнение аминокислотных последовательностей TyrRSTT с соответствующими последовательностями из других организмов выявило, что фермент из T. thermophilus относится к той же ветви филогенетического древа эубактериальных TyrRS, что и ферменты из Aquifex aeolicus, Deinococcus radiodurans, Haemophilus influenzae и Helicobacter pyroly (идентичность 40—57 %), но не к той, к которой принадлежат, например, Escherichia coli, Chlamydia trachomatis и Bacillus stearothermophilus (24—28 % идентичности). Аминокислотная последовательность каталитического домена высококонсервативна, в то время как тРНК-связывающий С-концевой домен содержит лишь несколько консервативных остатков. Однако даже в последовательности активного центра отмечено важное различие между двумя группами эубактериальных TyrRS: остаток Lys-41 в TyrRSTT (и в TyrRS из многих патогенных бактерий человека) представлен консервативным остатком тирозина в бактериальных TyrRS другой группы, а также в TyrRS эукариот, включая человека. Это отличие может быть использовано при создании новых антибиотиков.

REFERENCES

1. *Eriani G., Delarue M., Poch O., Gangloff J., Moras D.* Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs // Nature.—1990.—**347**, N 6289.—P. 203—206.
2. *Cusack S., Berthet-Colominas C., Hartlein M., Nassar N., Leberman R.* A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å // Nature.—1990.—**347**, N 6290.—P. 249—255.
3. *Steinberg S., Misch A., Sprinzl M.* Compilation of tRNA sequences and sequences of tRNA genes // Nucl. Acids Res.—1993.—**21**, N 13.—P. 3011—3015.
4. *Quinn C. L., Tao N., Shimmel P.* Species-specific microhelix aminoacylation by eukaryotic pathogen tRNA synthetase dependent on a single base-pair // Biochemistry.—1995.—**34**, N 39.—P. 12489—12495.
5. *Яремчук А. Д., Тукало М. А., Егорова С. П., Коноваленко А. В., Мацука Г. Х.* Выделение тирозил-тРНК синтетазы из *Thermus thermophilus* HB-27 // Укр. біохім. журн.—1990.—**62**, № 2.—С. 97—99.
6. *Marmur J.* A procedure for the isolation of deoxyribonucleic acid from microorganisms // J. Mol. Biol.—1961.—3.—P. 208—218.
7. *Sanger F., Nicklen S., Coulson A. R.* DNA sequencing with chain-terminating inhibitors // Proc. Nat. Acad. Sci. USA.—1977.—**74**, N 12.—P. 5463—5467.
8. *Yaremchuk A., Kriklivyi I., Tukalo M., Cusack S.* Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition // EMBO J.—2002.—**21**, N 14.—P. 3829—3840.
9. *Qiu X., Janson C. A., Smith W. W., Green S. M., McDevitt P., Johanson K., Carter P., Hibbs M., Lewis C., Chalker A., Fosberry A., Lalonde J., Berge J., Brown P., Houge-Frydrych C. S., Jarvest R. L.* Crystal structure of *Staphylococcus aureus* tyrosyl-tRNA synthetase in complex with a class of potent and specific inhibitors // Protein Sci.—2001.—**10**, N 10.—P. 2008—2016.
10. *Kim S., Lee S. W., Choi E.-C., Choi S. Y.* Aminoacyl-tRNA synthetases and their inhibitors as a novel family of antibiotics // Appl. Microbiol. Biotechnol.—2003.—**61**, N 4.—P. 278—288.

УДК 577.21:577.217.32
Надійшла до редакції 28.03.03