Slight sequence variations of a common fold explain the substrate specificities of tRNA-guanine transglycosylases from the three kingdoms

Christophe Romier, Joachim E.W. Meyer, Dietrich Suck*

EMBL, Meyerhofstrasse 1, 69017 Heidelberg, Germany

Received 11 September 1997

Abstract tRNA-guanine transglycosylases (TGTs) are the enzymes catalyzing the base exchange required for the synthesis of the modified bases derived from 7-deazaguanine in prokaryotic, archaebacterial, and eukaryotic tRNAs. Unlike the eukaryotic and archaebacterial enzymes, the prokaryotic TGTs have been clearly identified and highly characterized both biochemically and structurally. The recent occurrence in sequence databases of archaebacterial and eukaryotic proteins homologous to the prokaryotic TGTs reveals that all TGTs unexpectedly adopt a common fold. Observed sequence variations at the active site correlate well with their specificities for the various 7-deazaguanine derivatives and the total conservation of the catalytic residues strongly favors a common catalytic mechanism for all TGTs.

Key words: tRNA-guanine transglycosylase; Queuine; Archaeosine; Homology; Modeling; Catalytic mechanism

1. Introduction

Many modified bases are found in RNA and especially in tRNA [1] where they play various biochemical and structural roles [2]. Among them are the 7-deazaguanine derivatives queuine and archaeosine (Fig. 1). Queuine is found at the wobble position of prokaryotic and eukaryotic tRNAs specific for Asn, Asp, His and Tyr, with the exception of yeast and archaebacteria [3–5]. Several studies have highlighted the role of queuine at the translational level [6–12], although the free base may also play a role in the eukaryotic metabolism [13]. Additionally, queuine-modified tRNA levels have been shown to vary during development, differentiation, aging and cancer [14,15]. On the other hand, archaeosine is found exclusively at position 15 in the D-loop of most archaebacterial tRNAs where it is believed to play a structural role [16].

The synthesis of these derivatives is unique since, unlike the other modified bases, it starts outside the tRNA and requires a base exchange reaction catalyzed by enzymes known as tRNA-guanine transglycosylases (TGTs). In prokaryotes, queuine is synthesized de novo in a complex biosynthetic pathway [17]. The prokaryotic TGT is a 43 kDa zinc-containing enzyme [18–20] which replaces the encoded guanine at the wobble position with the queuine precursor 7-aminomethyl-7-deazaguanine ($preQ_1$; Fig. 1), although guanine and the queuine precursor 7-cyano-7-deazaguanine ($preQ_0$; Fig. 1), but not queuine, are also accepted as substrates in vitro [21,22]. Additionally, an anticodon stem-loop having a $U^{33}G^{34}U^{35}$ sequence has been shown to be a minimal substrate for *Escherichia coli* TGT [23–25]. We have recently solved the structure of *Zymomonas mobilis* TGT, alone and in complex with $preQ_1$ [20,26,27]. The enzyme adopts a non-canonical $(\beta/\alpha)_8$-barrel fold with the $preQ_1$-binding pocket located at the C-terminal face of the barrel. By band-shift assays and mutagenesis we have shown that the catalytic mechanism of prokaryotic TGT proceeds via the formation of a TGT/tRNA covalent intermediate, with Asp^{102} being the active site nucleophile of *Z. mobilis* TGT [28].

In eukaryotes, queuine is a nutrient and the replacement of the wobble guanine with queuine is carried out in a single enzymatic step, even though guanine, $preQ_0$ and $preQ_1$ are also substrates in vitro [29,30]. The $U^{33}G^{34}U^{35}$ sequence is also specifically recognized by the eukaryotic TGT [31]. However, contradictory data concerning the size and the oligomeric state of eukaryotic TGTs have been reported [29,30,32].

Recently, an archaebacterial TGT was isolated from *Haloferax volcanii* [33]. This enzyme is a 78 kDa protein which has been shown, through partial sequencing, to be sequence-related to the prokaryotic enzyme. However, the archaebacterial protein replaces the encoded guanine at position 15 in the dihydrouridine loop with $preQ_0$, guanine being also a substrate in vitro but not $preQ_1$ or queuine [33].

2. Materials and methods

The *Shigella flexneri* (SWISS-PROT accession code Q54177), *Haemophilus influenza* (SWISS-PROT accession code P44594), *Helicobacter pylori* (EMBL accession code Y12061), *Synechocystis* sp. (SWISS-PROT accession code Q55983), *Thermotoga maritima* (TIGR), *Caenorhabditis elegans* (EMBL accession code Z73899), mouse (EST), human (EST), *Methanoccocus jannaschii* (EMBL accession code U67495), *Archaeoglobus fulgidus* (TIGR) and *Methanobacterium thermoautotrophicum* (GTC) sequences were retrieved from the EMBL (European Molecular Biology Laboratory), SWISS-PROT, EST (Expressed Sequence Tags), TIGR (The Institute for Genomic Research, personal communication) and GTC (Genome Therapeutics Corporation, personal communication) databases. The *Zymomonas mobilis* (SWISS-PROT accession code P28720) and *Escherichia coli* (SWISS-PROT accession code P19675) sequences were used as starting search sequences within BLASTN [34]. When necessary, this search was repeated with full-length eukaryotic and archaebacterial sequences for full complementarity. The alignment of the sequences was done with CLUSTALW [35].

The mouse and human sequences were partially reconstructed from sufficiently long cDNA fragments stored in the EST databases. It should be noted that other eukaryotic proteins within the databases have been annotated as tRNA-guanine transglycosylases (SWISS-PROT accession codes P54578, P40826 and Q17361). These proteins are totally unrelated in sequence to the prokaryotic TGTs and the presence of homologs in the genome of *Saccharomyces cerevisiae* (accession name YFR010w) and *S. pombe* (EMBL accession code Z81317), organisms devoid of any 7-deazaguanine derivatives, strongly suggests that these proteins are not the eukaryotic TGTs.

Other partial sequences from *Vibrio cholerae*, *Deinococcus radiodurans*, *Plasmodium falciparum*, *Toxoplasma gondii*, rice and *Dro-*

*Corresponding author. Fax: (49) (6221) 387-306.
E-mail: suck@embl-heidelberg.de

PII S0014-5793(97)01175-7

sophila melanogaster were also retrieved from these databases. As these sequences are too short they are not shown but they support the conclusions drawn from the full-length sequences.

The *M. jannaschii* and *M. thermoautotrophicum* sequences have been obtained assuming a frameshift during sequencing which truncated the proteins at their N- and C-terminus, respectively. Use of different frames restored parts strongly homologous to the other archaebacterial sequences. Due to possible other sequencing errors, the exact N- and C-termini of the latter protein could not be precisely defined and a few extra residues may be present at both locations. The *T. maritima* and *A. fulgidus* sequences have been partially reconstructed from sufficiently long prokaryotic and archaebacterial homologous sequences found in their newly released genomes. Due to the unavailability of their DNA sequences, it is still impossible to confirm whether these sequences are contiguous on the genome.

The homology modeling of the three-dimensional structures of the *C. elegans* and the N-terminal part of the *M. jannaschii* proteins was carried out within TURBO-FRODO [36] using the X-ray structure of the *Z. mobilis* TGT/preQ$_1$ complex [27] as template. The active site vicinity of both proteins was refined stereochemically using TURBO-FRODO and X-PLOR [37].

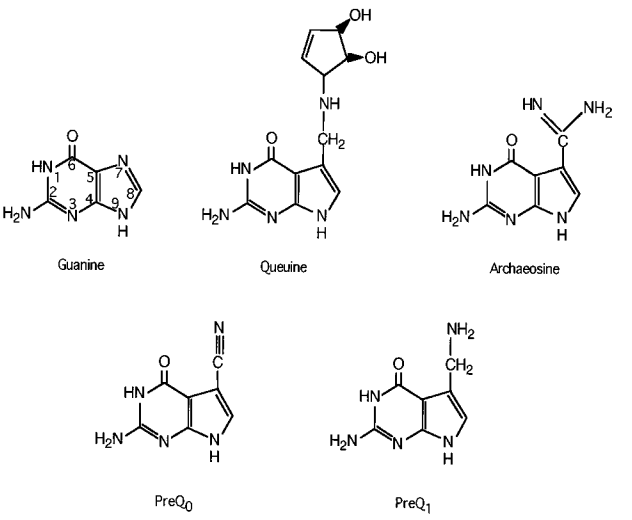


Fig. 1. Chemical structures of guanine and the different 7-deazaguanine derivatives.

3. Results

A search of sequence databases for protein sequences homologous to the prokaryotic TGTs yields, besides the various prokaryotic sequences, significant hits for eukaryotic and archaebacterial proteins (Fig. 2). Whereas the eukaryotic proteins have about the same length as the prokaryotic TGTs and are ~40% identical to them in sequence (Table 1; Fig. 2), the archaebacterial proteins are also clearly homologous to the prokaryotic TGTs but have about 300 additional residues at their C-terminus (Fig. 2). This C-terminal part is not as conserved as the N-terminal one and shows no homology to any other protein sequence. However, the zinc ligands and the nucleophilic aspartate are conserved in all the enzymes (Fig. 2).

The high homology between these proteins and knowledge of the three-dimensional structure of *Z. mobilis* TGT allowed the modeling of the *C. elegans* and the N-terminal part of the *M. jannaschii* proteins, in particular their active sites. Assuming that these proteins represent the eukaryotic and archaebacterial TGTs, it should be possible to explain their different substrate specificities on the basis of amino acid substitutions observed in the active site vicinity. For this purpose, the structure of the *Z. mobilis* TGT/preQ$_1$ complex was used as a template for modeling of the complexes between the proteins and the various modified bases. Especially, the position of the 7-deazaguanine moiety was kept as observed experimentally.

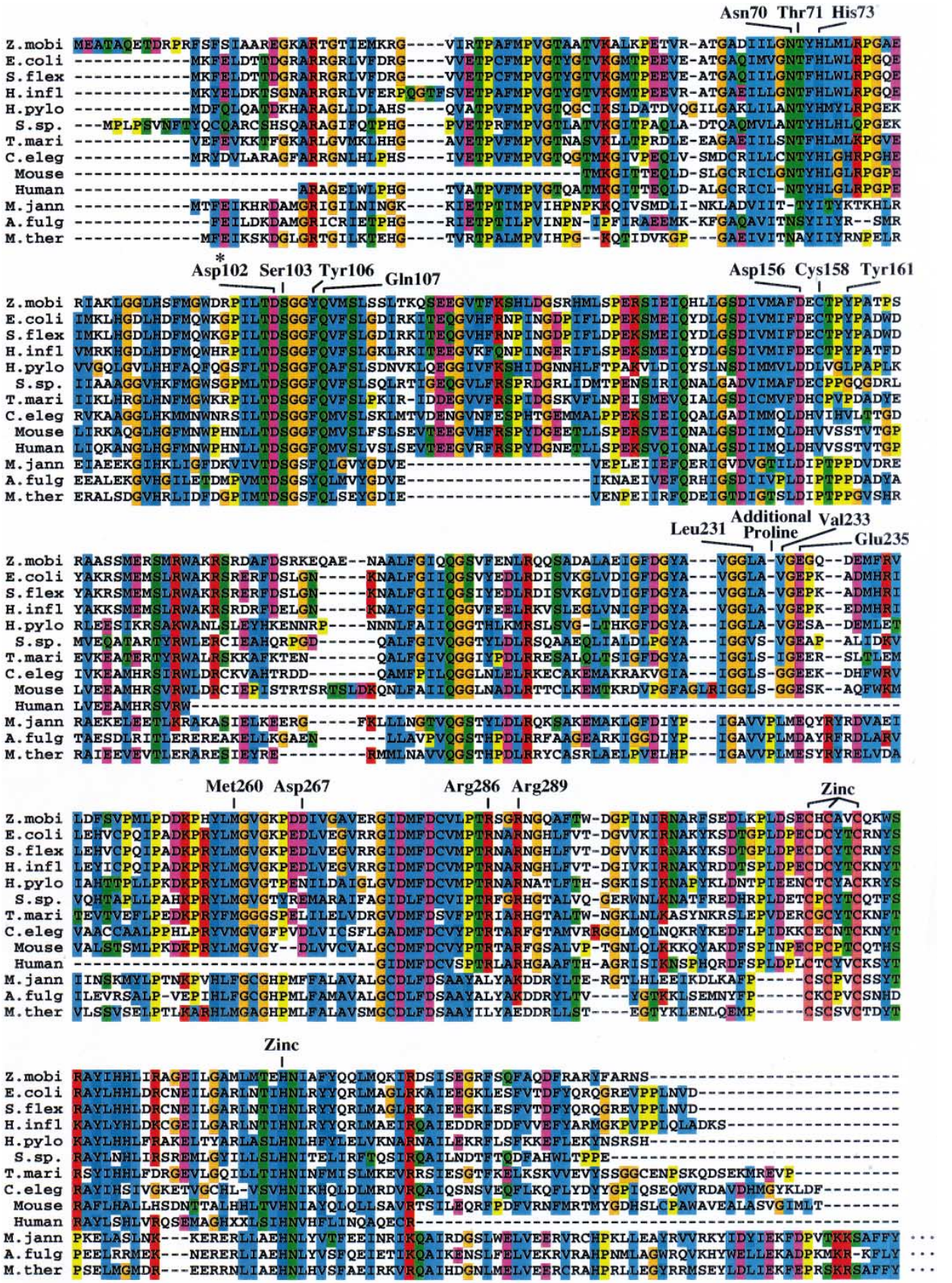
In the *Z. mobilis* TGT/preQ$_1$ structure, the base is found sandwiched between the side chains of Met^{260} on one side, and Tyr^{106} and Cys^{158} on the other side. The 7-deazaguanine moiety is specifically recognized through hydrogen bonding between the carboxylate of Asp^{156} and the 1-NH and 2-NH$_2$ groups of preQ$_1$, and between the amide nitrogen of Gly^{230} and the oxygen O6 of preQ$_1$ (Fig. 3A). It should be noted that this recognition pattern is also a guanine recognition motif present in many proteins binding a guanine moiety [38]. The additional amino group of preQ$_1$ contributes to recognition by forming a hydrogen bond with the carboxyl oxygen of Leu^{231} (Fig. 3A). A preQ$_0$ molecule could also be recognized in this site since the partial negative charge of the cyano nitrogen would be in hydrogen bonding distance to the amide nitrogen of Leu^{231} due to the linearity of the CCN group (data not shown). This recognition requires a slight tilting of the base, but preQ$_1$ is also able to adopt such an orientation when bound to TGT [27]. On the other hand, a queuine molecule does not fit into this pocket because of steric interference with Val^{233} which forms the ceiling of the pocket (Fig. 3A). These results are in agreement with the substrate specificity of *E. coli* TGT [21,22].

Very few changes are observed in the binding pocket of the eukaryotic enzyme. Cys^{158} is replaced by a valine, a mutation which should not prevent the hydrophobic interaction with the 7-aminomethyl moiety of the base. However, the smaller volume of a valine at this position combined with the replacement of Val^{233} by a glycine enlarges the binding pocket and allows a queuine molecule to bind (Fig. 3B). Actually, the interactions between the protein and the preQ$_1$ moiety are conserved, but an additional hydrophobic interaction between the cyclopentenediol moiety of queuine and Val^{158} is observed. As for the two hydroxyls, they are in hydrogen bonding distance to several carbonyl oxygens of the protein main chain. Since the preQ$_1$ binding specificity is not changed, guanine, preQ$_0$ and preQ$_1$ should also bind in this pocket, as it has been previously observed [29,30].

In the archaebacterial case, more sequence changes are observed. Here, Met^{260} is replaced by a phenylalanine, a mutation which should, however, not perturb the stacking of the base (Fig. 3C). More important are the other mutations located within or in the vicinity of the binding pocket. First, both Cys^{158} and Tyr^{161} are mutated into prolines. Although

→

Fig. 2. Alignment of the *Z. mobilis* (Z.mobi), *E. coli* (E.coli), *S. flexneri* (S.flex), *H. influenza* (H.infl), *H. pylori* (H.pylo), *Synechococcus* sp. (S.sp.), *T. maritima* (T.mari), *C. elegans* (C.eleg), mouse, human, *M. jannaschii* (M.jann), *A. fulgidus* (A.fulg) and *M. thermoautotrophicum* (M.ther) TGT sequences. The human and mouse sequences are not complete and gaps may represent missing data. Residues described in the text have been labeled with *Z. mobilis* numbers. Asp^{102}, marked with an asterisk, is the active site nucleophile of *Z. mobilis* TGT. The four zinc ligands are marked 'Zinc'. The amino acid marked 'Additional Proline' is the proline residue found exclusively in the archaebacterial sequences. The additional C-terminal residues of these latter sequences are not shown (marked as '...').


Asn70 Thr71 His73
Asp102 Ser103 Tyr106 Gln107
Asp156 Cys158 Tyr161
Additional Proline
Leu231 Val233 Glu235
Met260 Asp267
Arg286 Arg289
Zinc
Zinc
Z.mobi
E.coli
S.flex
H.infl
H.pylo
S.sp.
T.mari
C.eleg
Mouse
Human
M.jann
A.fulg
M.ther

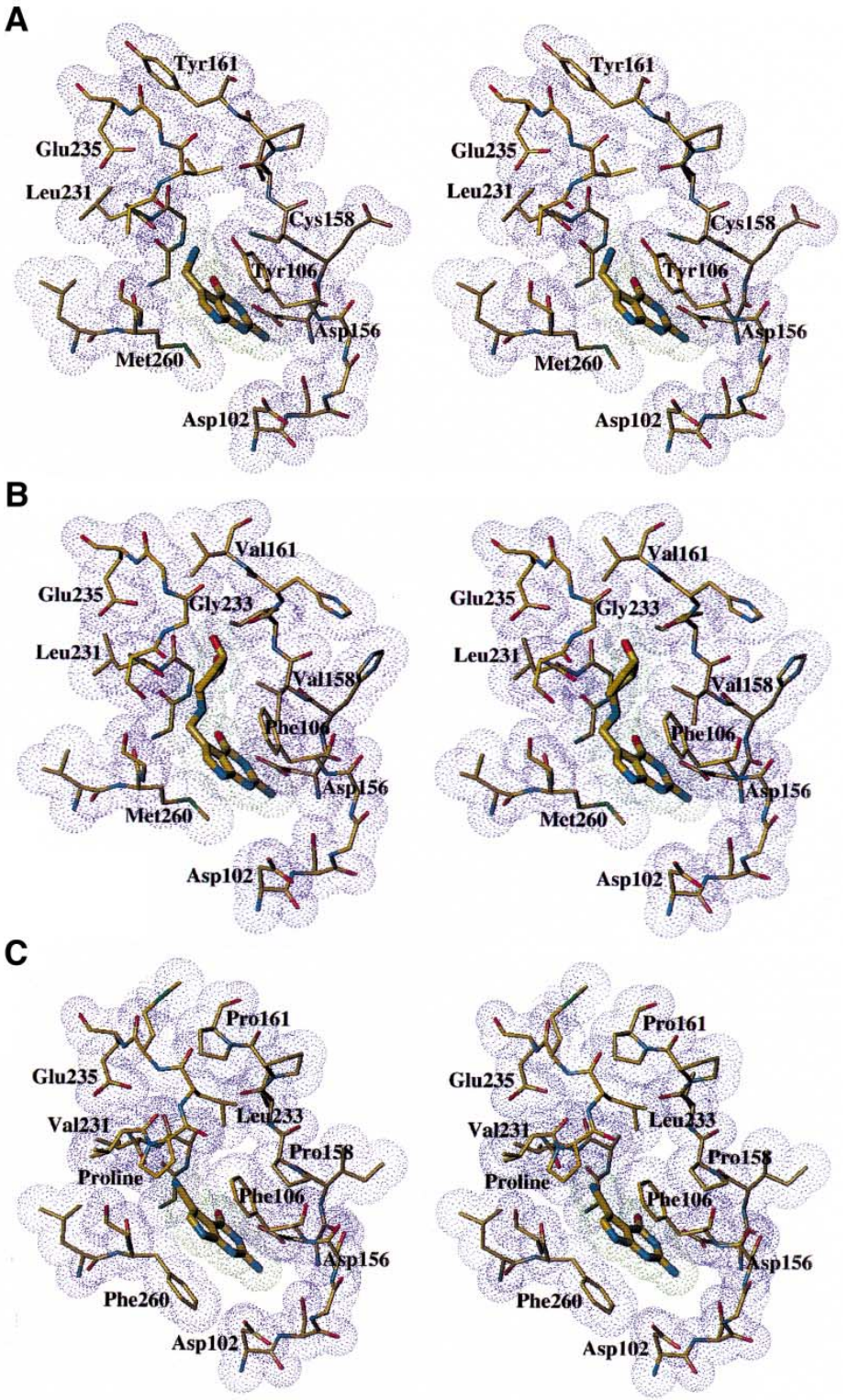


Fig. 3. Recognition of the cognate 7-deazaguanine derivatives by the various TGTs. For clarity the *Z. mobilis* numbering has been kept in all proteins. A: Observed recognition of $preQ_1$ by *Z. mobilis* TGT. B: Modeled recognition of queuine by *C. elegans* TGT. C: Modeled recognition of $preQ_0$ by *M. jannaschii* TGT. The residue marked Proline is the additional proline found exclusively in the archaebacterial sequences.

the former mutation does not change the hydrophobic character of the binding pocket at this location, the latter creates a stronger steric hindrance. Indeed, as the side chain of Tyr^{161} was pointing outwards, the proline is bulging out above the binding pocket (Fig. 3C). Facing this proline is a leucine which replaces the prokaryotic Val^{233}. This bulkier residue cannot be accommodated where the valine previously was because of the bulging proline opposite to it. Consequently, the loop formed by residues Val^{231} to Glu^{235} is pushed slightly downwards and towards the center of the binding pocket (Fig.

Table 1
Sequence identities between prokaryotic, eukaryotic and archaebacterial TGTs

	E.coli	S.flex	H.infl	H.pulo	S.sp.	T.mari	C.eleg	Mouse	Human	M.jann	A.fulg	M.ther
Z.mobi	54.9[a]	55.2	55.2	43.1	42.5	46.7	38.4	39.1	49.0	22.9[b]	21.0	22.6
E.coli		99.2	81.6	45.0	42.0	48.1	40.0	38.3	46.9	25.9	22.4	24.5
S.flex			81.6	45.6	42.3	48.9	40.3	38.6	47.3	26.1	22.4	24.5
H.infl				43.9	40.4	47.0	40.4	39.7	49.4	24.6	19.9	22.3
H.pylo					38.0	41.0	37.2	34.6	46.9	22.6	20.2	22.9
S.sp.						45.1	41.2	39.1	50.6	23.0	20.9	22.5
T.mari							41.6	36.9	49.0	25.8	22.3	25.0
C.eleg								50.3	54.8	19.0	17.4	18.1
Mouse									63.1	20.4	19.6	19.0
Human										19.9	19.9	21.2
M.jann											51.8	52.6
A.fulg												48.5

[a]All values are given as percentages.
[b]For the archaebacterial proteins, only the N-terminal part has been considered for calculations.

3C). Moreover, an additional proline – not present in the prokaryotic or eukaryotic proteins – is inserted immediately before the leucine (Fig. 2). Its role might be to rigidify and/or help to adjust the conformational change of the loop. Altogether, these modifications prevent the recognition of $preQ_1$ – and of course of archaeosine and queuine – by restricting the volume of the pocket in the region where its additional amino group binds. However, there is still enough space for binding a $preQ_0$ – or a guanine – molecule, in agreement with the observed substrate specificity of the archaebacterial TGT [33].

4. Discussion

Altogether, the results presented here strongly favor the hypothesis that the eukaryotic and archaebacterial proteins described are the tRNA-guanine transglycosylases of these two kingdoms and that all TGTs adopt a common fold. These findings are unexpected. First, no trace of an ancient queuine biosynthesis pathway has been found in eukaryotes, and the eukaryotic TGT recognizes directly queuine, a molecule much more bulky than $preQ_1$. Second, in archaebacteria the incorporation of $preQ_0$ is done at a totally different location of the tRNA and requires a completely different mode of tRNA recognition.

The prokaryotic and eukaryotic TGTs recognize the anticodon stem-loop of their cognate tRNAs, especially the specific $U^{33}G^{34}U^{35}$ sequence [23–25,31]. The high homology between these proteins explain why they display the same specificity towards tRNA. Especially, the two arginines 286 and 289 which have been postulated to play a major role in the binding of the anticodon stem-loop phosphate backbone and other residues (Asn^{70}, His^{73}, Thr^{71} and Asp^{267}) which have been implicated in the specific UGU sequence recognition [27] are, with the exception of Asp^{267}, totally conserved in the eukaryotic proteins (Fig. 2).

On the other hand, archaebacterial TGTs recognize the D-loop of archaebacterial tRNAs having a guanine at position 15 [16]. The additional 300 residues and/or the larger number of amino acid substitutions in the archaebacterial enzymes could be responsible for the difference in tRNA recognition. Furthermore, arginines 286 and 289, His^{73}, Asn^{70}, Thr^{71} and Asp^{267} are not highly conserved in these proteins (Fig. 2).

We have shown that prokaryotic TGTs catalyze their reaction via the formation of a protein/tRNA covalent intermediate, Asp^{102} being the active site nucleophile of *Z. mobilis* TGT, and have suggested that the wobble guanine is recognized within the $preQ_1$-binding pocket [27,28]. The total conservation of the nucleophile aspartate and of Gln^{107} – assumed to recognize the O2′ hydroxyl of the wobble guanine – together with the fact that all the enzymes are able to recognize guanine in their 7-deazaguanine-binding pockets are in agreement with a base exchange mechanism for the synthesis of the various 7-deazaguanine-modified tRNAs based upon a common protein fold and a similar catalytic mechanism, including mammals.

Finally, the case of *T. maritima*, an organism which has diverged quite early from the other prokaryotes, raises the question of the evolutionary relationship between the different TGTs. Indeed, *T. maritima* TGT, even though it is highly homologous to the other prokaryotic TGTs (Table 1 and Figs. 2 and 3), displays a few features characteristic for the archaebacterial TGTs, especially a proline is found at position 161 and an isoleucine at position 233. On the other hand, it does not have an additional proline within the loop formed by residues 231–235. It therefore remains to be determined whether this protein uses as substrate $preQ_1$ or $preQ_0$, this latter precursor being also found in prokaryotes [39]. If these hypotheses are relevant, they could indicate two possible evolutionary pathways. According to the first, the archaebacterial proteins could have evolved by protein fusion from an ancestor using $preQ_0$, and possibly close to the *T. maritima* TGT, in order to synthesize archaeosine in their tRNAs and therefore, through its structural role, to adapt to their challenging environments. Alternatively, the prokaryotic and eukaryotic proteins could have evolved from a common ancestor involved in archaeosine synthesis and used it to synthesize queuine in their tRNAs to cope with their more complicated translational events. In both cases, the eukaryotes would have discarded most of their queuine biosynthesis proteins to use queuine as a mere nutrient.

Acknowledgements: The authors thank The Institute for Genomic Research and Genomic Therapeutics Corporation for making available sequence data prior to publication.

References

[1] Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1994) Nucleic Acids Res. 22, 2183–2196.
[2] Björk, G.R. (1995) in: tRNA: Structure, Biosynthesis, and Func-

tion (Söll, D. and RajBhandary, U., Eds.), pp. 165–205, American Society for Microbiology, Washington, DC.
[3] Goodman, H.M., Abelson, J., Landy, A., Brenner, S. and Smith, J.D. (1968) Nature 217, 1019–1024.
[4] Harada, F. and Nishimura, S. (1983) Biochemistry 11, 302–308.
[5] Katze, J.R., Basile, B. and McCloskey, J.A. (1982) Science 216, 55–56.
[6] Grosjean, H.J., Henau, S. and Crothers, D.M. (1978) Proc. Natl. Acad. Sci. USA 15, 610–614.
[7] Meier, F., Suter, B., Grosjean, H., Keith, G. and Kubli, E. (1985) EMBO J. 4, 823–827.
[8] Bienz, M. and Kubli, E. (1981) Nature 294, 188–190.
[9] Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. and Gross, H.J. (1984) EMBO J. 3, 351–356.
[10] Beier, H., Barciszewska, M. and Sickinger, H.D. (1984) EMBO J. 3, 1091–1096.
[11] Hatfield, D. et al. (1989) Virology 173, 736–742.
[12] Durand, J.M. et al. (1994) J. Bacteriol. 176, 4627–4634.
[13] Langgut, W. (1996) Endocytobiosis Cell Res. 11, 233–238.
[14] White, B.N., Tener, G.M., Holden, J. and Suzuki, T. (1973) J. Mol. Biol. 74, 635–651.
[15] Dirheimer, G., Baranowski, W. and Keith, G. (1995) Biochimie 77, 99–103.
[16] Gregson, J.M. et al. (1993) J. Biol. Chem. 268, 10076–10086.
[17] Slany, R.K. and Kersten, H. (1994) Biochimie 76, 1178–1182.
[18] Reuter, K., Slany, R., Ullrich, F. and Kersten, H. (1991) J. Bacteriol. 173, 2256–2264.
[19] Chong, S., Curnow, A., Huston, T.J. and Garcia, G.A. (1995) Biochemistry 34, 3694–3701.
[20] Reuter, K. and Ficner, R. (1995) J. Bacteriol. 177, 5284–5288.
[21] Okada, N. and Nishimura, S. (1979) J. Biol. Chem. 254, 3061–3066.
[22] Okada, N. et al. (1979) J. Biol. Chem. 254, 3067–3073.
[23] Curnow, A.W., Kung, F.L., Koch, K.A. and Garcia, G.A. (1993) Biochemistry 32, 5239–5246.
[24] Nakanishi, S. et al. (1994) J. Biol. Chem. 269, 32221–32225.
[25] Curnow, A.W. and Garcia, G.A. (1995) J. Biol. Chem. 270, 17264–17267.
[26] Romier, C., Ficner, R., Reuter, K. and Suck, D. (1996) Proteins Struct. Funct. Genet. 24, 516–519.
[27] Romier, C., Reuter, K., Suck, D. and Ficner, R. (1996) EMBO J. 15, 2850–2857.
[28] Romier, C., Reuter, K., Suck, D. and Ficner, R. (1996) Biochemistry 35, 15734–15739.
[29] Shindo-Okada, N., Okada, N., Ohgi, T., Goto, T. and Nishimura, S. (1980) Biochemistry 19, 395–400.
[30] Slany, R.K. and Müller, S.O. (1995) Eur. J. Biochem. 230, 221–228.
[31] Carbon, P., Haumont, E., Fournier, M., de Henau, S. and Grosjean, H. (1983) EMBO J. 2, 1093–1097.
[32] Walden, T.L., Howes, N. and Farkas, W.R. (1982) J. Biol. Chem. 257, 13218–13222.
[33] Okada, N. et al. (1997) 17th International tRNA Workshop at Kazusa Akademia Center, Japan.
[34] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[35] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[36] Roussel, A. and Cambillau, C. (1992) TURBO-FRODO, AFMB-CNRS, Marseille.
[37] Brünger, A.T. (1992) X-PLOR Version 3.1 Manual, Yale University Press, New Haven, CT.
[38] Kjeldgaard, M., Nyborg, J. and Clarck, B.F.C. (1995) FASEB J. 10, 1347–1368.
[39] Noguchi, S. et al. (1978) Nucleic Acids Res. 5, 4215–4223.