Selection of Viral RNA-Derived tRNA-Like Structures with Improved Valylation Activities†

Jens Wientges,‡,§ Joern Pütz,|| Richard Giegé,|| Catherine Florentz,|| and Andreas Schwienhorst*,‡

Abteilung Molekulare Genetik und Praeparative Molekularbiologie, Institut fuer Mikrobiologie und Genetik, Grisebachstrasse 8, 37077 Göttingen, Germany, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 2, 37077 Göttingen, Germany, and Unité Propre de Recherche 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg-Cedex, France

Received December 13, 1999; Revised Manuscript Received February 28, 2000

ABSTRACT: The tRNA-like structure (TLS) of turnip yellow mosaic virus (TYMV) RNA was previously shown to be efficiently charged by yeast valyl-tRNA synthetase (ValRS). This RNA has a noncanonical structure at its 3′-terminus but mimics a tRNA L-shaped fold, including an anticodon loop containing the major identity nucleotides for valylation, and a pseudoknotted amino acid accepting domain. Here we describe an in vitro selection experiment aimed (i) to verify the completeness of the valine identity set, (ii) to elucidate the impact of the pseudoknot on valylation, and (iii) to investigate whether functional communication exists between the two distal anticodon and amino acid accepting domains. Valylatable variants were selected from a pool of 2×10^{13} RNA molecules derived from the TYMV TLS randomized in the anticodon loop nucleotides and in the length (1−6 nucleotides) and sequence of the pseudoknot loop L1. After nine rounds of selection by aminoacylation, 42 have been isolated. Among them, 17 RNAs could be efficiently charged by yeast ValRS. Their sequence revealed strong conservation of the second and the third anticodon triplet positions (A_{56}, C_{55}) and the very 3′-end loop nucleotide C_{53}. A large variability of the other nucleotides of the loop was observed and no wild-type sequence was recovered. The selected molecules presented pseudoknot domains with loop L1 varying in size from 3−6 nucleotides and some sequence conservation, but did neither reveal the wild-type combination. All selected variants are 5−50 times more efficiently valylated than the wild-type TLS, suggesting that the natural viral sequence has emerged from a combination of evolutionary pressures among which aminoacylation was not predominant. This is in line with the role of the TLS in viral replication.

Turnip yellow mosaic virus (TYMV)[1] belongs to a class of positive-strand RNA plant viruses whose genomes contain 3′-terminal structural elements that resemble the L-shape of transfer RNAs (Figure 1). In addition, these tRNA-like structures (TLSs) share several functional characteristics with tRNA such as interaction with aminoacyl-tRNA synthetases, translational elongation factors EF-Tu and EF-1α, nucleotidyl-transferase, and RNase P (reviewed in refs *1*−*4*). The aminoacylation properties of these RNAs gained much interest and are presently best understood. Three major aminoacylation specificities have been reported so far. Tymoviral RNAs (*5*−*7*), including TYMV RNA as type member, as well as furoviruses (*8*) are charged with valine, RNAs from bromoviruses (*9*), cucumoviruses (*10*), and hordeiviruses (*11*) are charged with tyrosine, and RNAs from tobamoviruses (with tobacco mosaic virus as type member) are substrates for histidyl-tRNA synthetases (*12*−*14*). It has also been observed that TLSs from all viral groups are substrates, although with different efficiencies, of histidyl-tRNA synthetase (*15*, *16*) and that an engineered TYMV TLS can become an efficient substrate of methionyl-tRNA synthetase (*17*).

The genomic RNA of TYMV can be valylated to high levels in vitro with ValRS from different origins including bacteria, yeast, animal, and plants (*5*, *18*−*20*). Aminoacylation occurs also in vivo as detected after microinjection into *Xenopus* oocytes (*21*) and in infected Chinese cabbage tissue (*22*). Valylation of TYMV RNA by the yeast enzyme proved to be efficient, with a k_{cat}/K_m ratio only about 25-fold lower than aminoacylation of the cognate yeast $tRNA^{Val}$ (*19*). Elements responsible for specific aminoacylation, the so-called identity elements, have been identified both for valylation by the yeast (*23*) and the wheat germ (*24*) enzymes, on 3′-end fragments including at least 82−86 of the last nucleotides of the viral RNA. The yeast enzyme is strongly dependent on the central anticodon nucleotide (A_{56}), and to a low extent to residues A_{54} (in the anticodon loop), U_{43} (in the hinge region between the acceptor domain and the anticodon domain), and A_4 at the very 3′-end of the molecule. The wheat germ enzyme has also a very strong requirement for nucleotide A_{56} as well as for the neighboring

† This research was supported by grants of the Deutsche Forschungsgemeinschaft to A.S., by the French-German Procope program to A.S. and C.F. and was partly supported by a grant of the European Community (BIO4-CT98-0189).

* To whom correspondence should be addressed. Phone: (+49) 551 393822. Fax: (+49) 551 393805. E-mail: aschwie1@gwdg.de.

‡ Abteilung Molekulare Genetik und Praeparative Molekularbiologie.

§ Max-Planck-Institute for Biophysical Chemistry.

|| Unité Propre de Recherche 9002 du CNRS.

[1] Abbreviations: TYMV, turnip yellow mosaic virus; TLS, tRNA-like structure; K_m, Michaelis constant; k_{cat}, apparent rate constant; ValRS, valyl-tRNA synthetase; tRNA, transfer RNA.

10.1021/bi992852l CCC: $19.00
Published on Web 04/26/2000

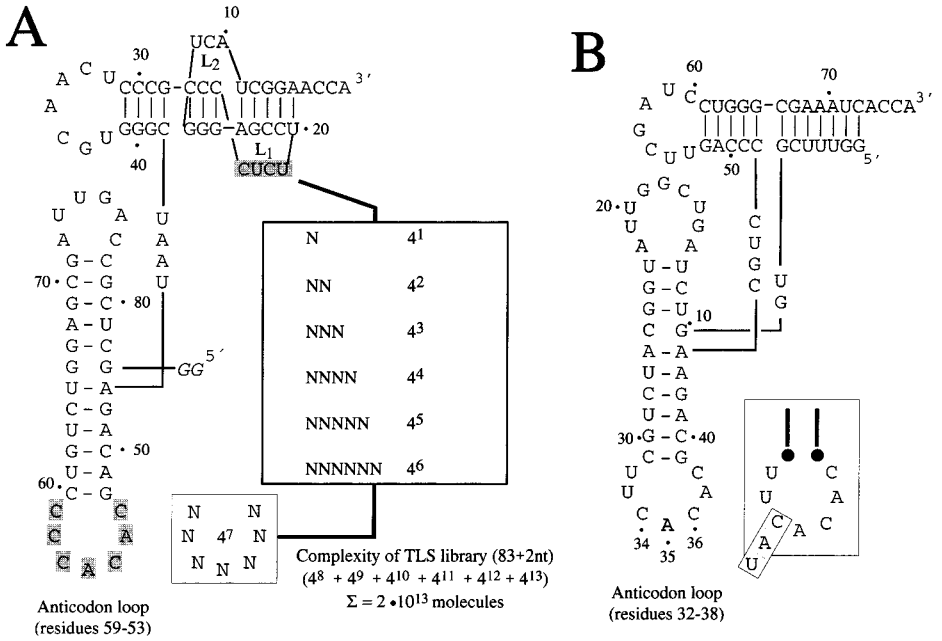

FIGURE 1: L-shape folded sequences of valine accepting RNAs. (A) Sequence of the tRNA-like domain found at the 3′ end of TYMV RNA (*61*). For convenience, numbering of the tRNA-like structure starts at the 3′-end, at variance with the convention in nucleic acids. Two G residues at position +1 and +2 at the 5′ end (given in italics) were added to enhance in vitro transcription by T7 RNA polymerase. Positions of the anticodon loop (7 nucleotides) and the pseudoknot loop L1 (1−6 nucleotides) which were randomized are shaded. The complexity of the TLS library is indicated. (B) Sequence of unmodified yeast $tRNA^{Val}$ (*62*). Nucleotides are numbered according to (*46*). A consensus anticodon loop sequence derived from 18 yeast $tRNA^{Val}$ genes (*45*), is shown in the boxed panel. Nucleotides are strictly conserved except for position 34 as highlighted by the three squares.

nucleotide C_{55} and the anticodon loop nucleotide closest to the 3′-end, C_{53} (*24*). These important anticodon nucleotides are conserved in all valylatable viral TLS (*6*, *8*). In addition, the equivalent nucleotides have been reported as identity elements for *Escherichia coli* $tRNA^{Val}$ (*25*−*28*). The pseudoknot present at the 3′-end of the TYMV TLS, and allowing for a mimicry of an acceptor stem, does not contain identity elements. However, the existence of the pseudoknot is by itself a critical element for aminoacylation activity of the TLS, and mutations affecting its stability are detrimental (*29*).

In canonical tRNAs, identity is commonly defined by unique sets of elements including nucleotides, chemical groups, or structural elements, and well-defined sets have been defined for a very large number of prokaryotic and eukaryotic systems (for reviews, see ref *30*−*32*). Moreover, a number of discoveries have both refined and complicated the understanding of tRNA aminoacylation identity and opened for new and alternate solutions to the problem. Thus, cryptic elements (*33*) as well as permissive elements (*34*) were discovered. Also, evidences for the existence of alternate identity sets for a given specificity were reported. Thus, an inactive *E. coli* $tRNA^{Ala}$ in which the major identity determinant, base-pair G_3-U_{70} in the acceptor stem, has been replaced by a G_3-C_{70} pair, undergoes functional compensation in vivo by second site mutation at a distal base pair (*35*). Studies on arginylation of different tRNAs by yeast ArgRS (*36*) demonstrated that this enzyme recognizes two different sets of anticodon loop nucleotides for the specific aminoacylation of $tRNA^{Arg}$ and $tRNA^{Asp}$ in vitro transcripts. Footprinting as well as kinetic analysis of mutants (*37*) support the existence of alternate mechanistic routes by which each identity set triggers the activation of the same synthetase.

Interestingly, in the case of the valylatable TLSs, a large variability in anticodon loop nucleotides and in length and sequence of the pseudoknot loop is observed (*6*, *8*). In addition, second-site suppressor mutations have also been observed for the TYMV TLS, both in vitro (*24*) and in vivo (*38*), that compensate for an initial loss in aminoacylation efficiency due to mutation at position 55, one of the major identity elements for the wheat germ enzyme. These results suggest that alternative specific recognition sets may also exist in tRNA-like structures. Thus, in this study, the sequence requirements for efficient valylation of TYMV RNA by yeast ValRS were investigated. A major objective was to analyze the variability of the identity set as well as the structural plasticity of the anticodon loop and the pseudoknot loop L1, tolerated by the synthetase. Using a novel method for the in vitro selection of aminoacylatable RNAs (*39*), a large ensemble of variants was generated and subjected to functional selection. The initial pool was designed to have the anticodon loop and the L1 loop (length and sequence) of the pseudoknot in the acceptor stem randomized. Stringency of selection was adjusted to reveal the most active molecules. After nine rounds of selection, only variants that exhibited a significantly increased valine accepting activity as compared to that of the wild-type TLS were captured. These variants show a strong conservation of the major valine identity elements in the anticodon loop and variability in the other nucleotides of the loop. They tolerate length variations in the pseudoknot loop L1 with, however, a strong bias toward three- and four-membered loops and the presence of a conserved U residue.

MATERIALS AND METHODS

Chemicals and Enzymes. Yeast valyl-tRNA synthetase (ValRS) and $tRNA^{Val}$ were purified as described (*40*). T7

RNA polymerase was purified from *E. coli* strain BL21 harboring the plasmid pAR1219 as described (*41*). AMV reverse transcriptase and dNTPs were obtained from Boehringer Mannheim (Germany). NHS-Biotin and NTPs were from Sigma (Germany), sulfo-NHS-Biotin was from Pierce. Restriction enzymes *Sma*I and *Eco*RV, polynucleotide kinase, and T4 DNA ligase were from Stratagene. *Taq* DNA polymerase was from Perkin-Elmer, L-[^{3}H]valine (46.2 Ci/mmol) from NEN, streptavidine-conjugated magnetic beads from Deutsche Dynal (Germany), and synthetic oligonucleotides were chemically synthesized by NAPS Göttingen GmbH (Germany) using phosphoramidite chemistry. Randomization of desired positions in synthetic oligonucleotides was carried out by using equimolar mixtures of all four phosphoramidites freshly prepared prior to synthesis. Purification of oligonucleotides was performed either by denaturing polyacrylamide gel electrophoresis (PAGE) or by size-exclusion HPLC (Progel-TSK G 3000 SWXL column and 0.2 M NaOAc pH 6.0, 1% MeOH as elution buffer) and subsequent EtOH precipitation.

Preparation of Transcription Template and RNA. The initial template DNA consisted of a synthetic DNA oligonucleotide corresponding to the T7 RNA polymerase promoter directly connected to the downstream tRNA-like gene with a 7 nt randomized anticodon loop (position 32−38) and a randomized pseudoknot loop L1 differing in length by 1−6 nucleotides (N_1−N_6) (Figure 1A). Conversion of the single-stranded initial DNA pool to double-stranded (ds) transcription templates was performed by PCR and utilized elongation of primers complementary to the 5′- and 3′- fixed regions by the *Taq* DNA polymerase. The 5′ primer TY-B-5′ (5′-GCT-AAT-ACG-ACT-CAC-TAT-AGG-GCT-CGC−CAG-TTA-GCG-AGG-TCT-3′; T7 promoter underlined) and the 3′-primer TY-Val-3′ (5′-TGG-TTC−CGA-TGA-CCC-TCG-GA-3′) were either annealed to the wild-type oligonucleotide from the TYMV RNA 3′-end (5′GGG-CTC-GCC-AGT-TAG-CGA-GGT-CTG-TCC-CCA-CAC-GAC-AGA-TAA-TCG-GGT-GCA-ACT-CCC-GCC-CCT-CTT-CCG-AGG-GTC-ATC-GGA-ACC-A-3′) or to the DNA library template oligonucleotide LibB (5′-GGG-CTC-GCC-AGT-TAG-CGA-GGT-CTG-TCN-NNN-NNN-GAC-AGA-TAA-TCG-GGT-GCA-ACT-CCC-GCC−C(N_1-N_6)T-CTT-CCG-AGG-GTC-ATC-GGA-ACC-A-3′; randomized nucleotides underlined). PCR amplification was carried out in a 100 μL reaction volume containing 100 ng of template oligonucleotide, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, 10 mM dithiotreitol, 200 μM of each dNTP, 0,3 μM of each of the primers, and 2.5 U *Taq* DNA polymerase. Using the following PCR cycle, 94 °C (3 min), 62 °C (2 min), 72 °C (30 s), sufficient double-stranded DNA was obtained after 20−30 cycles. PCR products were phenolized and ethanol precipitated. Populations of RNA variants were obtained by in vitro transcription of PCR-amplified template DNA containing the TYMV 3′-end variant genes under the control of the T7 RNA polymerase promoter. Site-directed mutagenesis was performed by PCR with mutagenic primers using either the wild-type oligonucleotide derived from TYMV RNA 3′-end or the corresponding genes of the tRNA-like variants. RNA was obtained by transcription of 5 μg of PCR amplified DNA by T7 RNA polymerase. Transcription reactions (250 μL) were performed in 40 mM Tris-HCl, pH 8.1, 22 mM $MgCl_2$, 1 mM spermidine, 5 mM dithiothreitol, 0.01% Triton X-100, 4 mM each nucleoside triphosphate (NTP), 16 mM GMP, 40 units of RNasin (Promega, Madison, WI), and the appropriate amount of T7 RNA polymerase. After incubation at 37 °C for 3 h, the reaction is treated with 20 units of RNase-free DNase at 37 °C for 15 min and EDTA (pH 8.0) was added to a final concentration of 50 mM.

Full-length pool transcripts were systematically separated from unincorporated nucleotides and abortive transcription products using size-exclusion HPLC (Progel-TSK G 3000 SWXL column and 0.2 M NaOAc, pH 6.0, 1% MeOH as elution buffer). Since some preparations of T7 RNA polymerase were described to produce RNAs with considerable 3′-end heterogeneity (*24*), ^{32}P-labeled transcripts from wild-type TLS and corresponding variants were analyzed by denaturing 20% PAGE. It was found that all RNA variants with valylation activity terminated predominantly with CCA 3′-end sequences.

Selection of Valylatable TYMV RNA Variants. Selection of aminoacylatable RNAs was performed as described (*38*). The RNA pool was aminoacylated with yeast ValRS in 50 mM Tris-HCl (pH 7.8), 30 mM KCl, 25 mM $MgCl_2$, 10 mM ATP, 4 mM glutathion, 5 μM valine, and 1 μM RNA transcript. Synthetase concentrations (20−0.5 nM) and incubation times (15 to 1 min) varied as indicated. The aminoacylation reaction was stopped by phenol extraction at pH 4.0. RNAs were then purified by size-exclusion HPLC (Progel-TSK G 3000 SWXL column and 0.2 M NaOAc, pH 6.0, 1% MeOH as elution buffer) and subsequent EtOH precipitation. Further steps for the selection of aminoacylated RNAs were carried out as described (*39*). They include biotinylation of aminoacylated RNA and subsequent trapping on streptavidin-coated magnetic beads. The captured molecules were converted into DNA by RT-PCR. The PCR products were transcribed to begin a new round of selection and the pool RNA was further used for aminoacylation studies (plateau charging). Alternatively, PCR products were cloned to allow sequence analysis and production of individual RNA species on the basis of individual plasmid sequences.

Cloning and Sequence Analysis of Selected RNA Species. Solid-phase captured RNAs were reverse transcribed and amplified by PCR. The resulting double-stranded DNAs were directly ligated into pCR 2.1 plasmid using the TA-cloning method (Invitrogen) according to the suppliers instructions. Plasmid DNA from single clones was isolated using the QIAprep 8 Miniprep Kit on QIAvac 6S (Qiagen, Germany). Purified plasmid DNA was analyzed by dye-terminator sequencing with Thermo Sequenase (Amersham, England). The sequencing protocol included 30 PCR cycles (30 s at 98 °C, 30 s at 55 °C, and 4 min at 60 °C) as described previously (*42*).

Valylation of tRNA-Like Variants. Valylation properties of pools and TYMV 3′ end variants were characterized by charging activity (plateau) and valylation kinetics (k_{cat}/K_m). Standard plateaus were determined using 20 nM ValRS and 1 μM transcripts. Reaction mixtures contained 50 mM Tris-HCl, pH 7.8, 30 mM KCl, 25 mM $MgCl_2$, 10 mM ATP, 4 mM glutathion, and 5 μM valine. Valylation was monitored over 12 min. Alternatively, less stringent conditions for inactive variants were used comprising 200 nM of enzyme and 2 h of incubation time. For kinetic parameter determina-

tions, initial rates were measured with 2−20 nM yeast ValRS and with RNA concentrations between 150 and 2400 nM. Data were analyzed on Eady-Hofstee plots; (k_{cat}/K_m) values for replicate experiments varied at most by 20%.

Aminoacylation properties of the variants were compared to those of a wild-type reference molecule formed by the 83 natural 3′-end nucleotides of the viral RNA as well as of two additional nonviral G residues at the 5′-end. This fragment is 59-fold less efficiently valylatable than a yeast $tRNA^{Val}$ in vitro transcript (see Results), which is in agreement (within a 2-fold range) with the efficiency of the full-length viral genomic RNAs (*19*). Aminoacylation properties of viral TLS are known to be sensitive to experimental conditions (for example refs *7* and *16*) so that special care was taken in the present work to perform all comparative experiments under strictly the same conditions.

Site-Directed Mutagenesis of Cloned tRNA-Like Sequences. In general, site-directed mutagenesis of cloned tRNA-like variants was performed by PCR using mutagenized primers. In this way, the wild-type L1 sequence (CUCU) was introduced into the sequence context of variants 8-8 and 8-39 using primers TY-B-5′ and TY-L1-wt (5′-TGG-TTC−CGA-TGA-CCC-TCG-GAA-GAG-GGG-CGG-GA-3′). Similarly, the L1 sequence (UCG) of variant 8-8 was introduced into the wild-type tRNA-like sequence context using primers TY-B-5′ (see above) and TY-L1−3 (5′-TGG-TTC−CGA-TGA-CCC-TCG-GAC-GAG-GGC-GGG-A-3′), respectively. Variant sequences of wild-type RNA and variant 8-39 with a single C residue replacing the L1 region were generated by PCR using primers TY-B-5′ and TY-L1-C (5′-TGG-TTC-CGA-TGA-CCC-TCG-GAG-GGG-CGG-GA-3′).

Sequences of mutagenized variants were confirmed by dye-terminator sequencing after directly ligating PCR products into pCR 2.1 plasmid using the TA-cloning method (Invitrogen) according to the suppliers instructions.

RNA Secondary Structure Prediction. For the prediction of RNA secondary structures, we used the APL-programmed genetic algorithm (*43*) implemented into the program STAR (*44*).

RESULTS

Construction of the TYMV RNA 3′-End Library. A library of DNA-oligonucleotides corresponding to the T7 RNA polymerase promoter directly connected to two additional guanosine residues (at position +1 and +2, to enhance transcription by T7 RNA polymerase) and the partially randomized 83 nt sequence of the TYMV RNA 3′-end (TLS) (Figure 1A) was used as template for PCR amplification. After transcription of the double-stranded template DNA, in vitro transcripts were purified by size-exclusion chromatography generating the initial RNA-library (LibB). In a similar way a control RNA containing the wild-type sequence was prepared.

The LibB library contained molecules randomized both in the anticodon loop as well as in the sequence and the length of pseudoknot loop L1 in the acceptor domain. The library was obtained using a template oligonucleotide mixture composed of equimolar amounts of six different oligonucleotides. The pseudoknot loop L1, 1−6 nucleotides in length, was completely randomized (Figure 1A). Although each family of molecules containing variants of the same size is

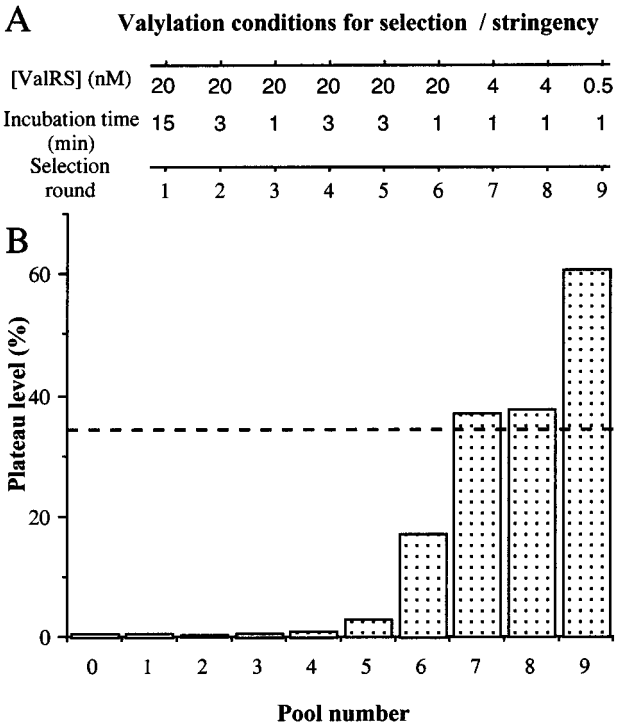

FIGURE 2: (A) Stringency of selection conditions in terms of ValRS concentration and incubation time. (B) Overall charging level of RNA pools selected for valine accepting activity. Pool 0 constitutes the initial RNA library. Determination of plateaus for each pool was done under identical conditions, with 1 μM pool RNA and at 20 nM ValRS. The broken line represents the plateau of the wild-type TLS (34%).

equally populated, a single molecular species with a 1-membered L1-loop is present with approximately 1000 times more copies in the initial library than a single molecular species with a six-membered L1-loop. In 1 μg (35 pmol; 2×10^{13} molecules) of RNA comprising the initial library, however, each of the different possible 8.9×10^7 molecular species is present in at least several thousand copies assuming a binomial distribution. Thus, one expects a complete coverage of the mutant space. The efficient randomization in the initial library was verified by cloning and sequencing 20 variants of the initial pool. As anticipated, the L1 domain in the variants encompassed sequences of length from 1−6 nucleotides. Both mutated regions were efficiently randomized and did not show any sequence bias (data not shown).

Selection of Valylatable RNAs and Pool Valylation Experiments. Valine accepting RNAs were enriched over nine selection rounds according to a method based on derivatization of the charged amino acid (*39*). The stringency of selection was increased stepwise by varying enzyme concentration and aminoacylation times (Figure 2A). Starting with 20 nM ValRS and an aminoacylation time of 15 min, incubation was further reduced to 1 min. In round 7, the enzyme concentration was decreased to 4 nM and in the final round it was even reduced to 0.5 nM.

To monitor the progressive enrichment of valine accepting tRNA-like variants during the selection procedure, plateau levels of enriched pools were determined by incubating 1 μM pool RNA with 20 nM enzyme for 12 min at 30 °C (Figure 2B). No significant aminoacylation was observed before the fourth round of selection. Starting with the pool from round 7, the valylation capacity of each subsequent

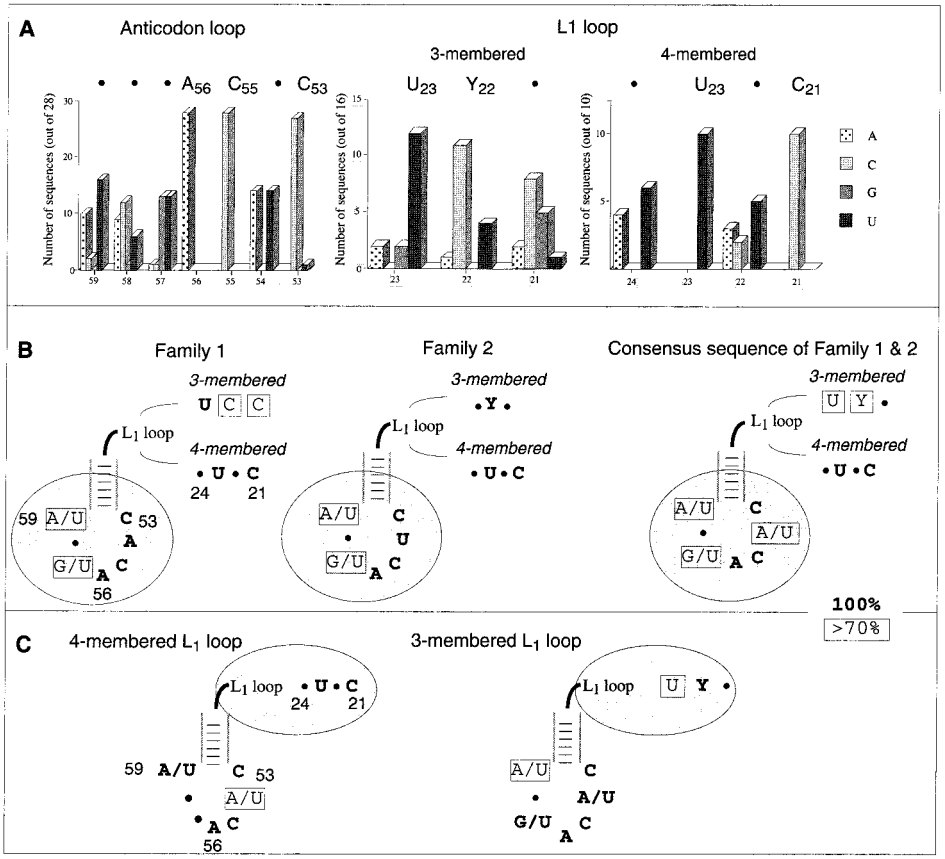

FIGURE 3: Sequence analysis of active variants selected under stringent aminoacylation conditions. (A) Distribution of nucleotides within the anticodon loops of 28 variants and within the pseudoknot loop L1 of variants sharing a three-membered loop (16 variants) or a four-membered loop (10 variants). Consensus sequences of the considered domains are given on top of the histograms. (B) Consensus nucleotides within active variants, based on a classification on the distribution of anticodon loop nucleotides into two families. (C) Consensus sequences of active variants according to a classification according to the length of the pseudoknot loop L1. Y stands for pyrimidine.

pool was better than that of pure wild-type TLS, as reflected by increased plateau levels (up to 1.7-fold of the wild-type level) that were also reached much more quickly than in the case of wild-type (data not shown). Note that between rounds 7 and 8, the plateau level does not change significantly. After increasing the stringency, i.e., decreasing the enzyme concentration, however, the plateau level (round 9) increased again, indicating that there is a correlation between stringency and plateau level.

First Evaluation of Individual RNA Species. RNAs from pools 8 and 9 were analyzed in detail. A total of 51 clones was sequenced. From these, nine clones showed alterations outside the anticodon and pseudoknot regions, most being conservative mutations in base-paired helical domains (e.g., base-pair flipping or G−C to G−U transitions) or mutations in the single-stranded hinge region (e.g., nucleotides 46−43) of the TLS. Variants carrying these mutations were not considered for further analysis here. A second set, formed by 14 clones, did not show any trace of consensus sequence in the anticodon loop and had a large variability in length of loop L1. Finally, a third set of clones contained 28 sequences with conserved anticodon nucleotides A_{56} and C_{55}. All variants of selection rounds 8 and 9 without additional mutations outside of randomized regions (42 RNAs), were subjected to an estimation of their global valylation properties by measurement of aminoacylation plateau levels. None of the variants from the second set, for example, without conserved nucleotides in the anticodon loop, showed significant valylation under standard conditions (20 nM ValRS; 1 μM RNA; 12 min incubation). At less stringent conditions (200 nM enzyme; 3 h incubation time), however, variants with an A residue at position 56 were chargeable to 40−50% (not shown), whereas other variants showed plateau levels of <1%. Valylation experiments on the third set of molecules, for example, those with conserved anticodon loop nucleotides, resulted in the finding that all molecules are excellent substrates for yeast ValRS with charging levels varying from 40 to 100%. It is noteworthy that all the analyzed variants that belong to this family reach a significantly higher plateau as compared to the wild-type TLS. Furthermore, plateaus are reached much faster, already indicating that k_{cat} of the selected variants is increased as compared to that of wild-type TLS.

Sequence Analysis of Active RNAs. Sequence comparison of the 28 active variants can be done either on the basis of conservation in the anticodon loop, on the basis of the pseudoknot length and sequence, or on combination of both. Figure 3A gives the statistical distribution of nucleotides within all anticodon loops and within the two main types of pseudoknot loops L1, those with three residues and those with four residues. Since an absolute conservation in all 28 sequences of anticodon nucleotides $A_{56}C_{55}$, which are known to be major valine identity elements, emerges obviously from the first glance analysis of the sequences, a classification based on the anticodon loop nucleotides is presented in Table 1.

In addition to the absolute conservation of the above-mentioned two nucleotides, all but one sequence have residue

Table 1: Kinetic Parameters for Valylation of Active[a] TLS Variant Transcripts with Yeast ValRS

Family	RNA	Sequences: anticodon loop	Sequences: pseudoknot loop L1	k_{cat} (s^{-1})	K_m (μM)	Gain[d] (x-fold)
		59 58 57 56 55 54 53	a b 24 23 22 21			
	wt TLS RNA	C C C **A C A C**	C U C U	0.24	2.44	1
		32 33 34 35 36 37 38				
	wt tRNA^{Val}[b]	U U C **A C A C**		0.75	0.13	59
	variants	59 58 57 56 55 54 53	a b 24 23 22 21			
	8-48	U A U **A C A C**	U U U C	3.00	0.86	36
	9-18	U A U **A C A C**	U U C C	nd[c]	nd	nd
1.a	8-8	U A U **A C A C**	U C G	2.67	0.93	29
	9-5	U A U **A C A C**	U C C	1.82	1.24	15
	8-6	A C U **A C A C**	U C U U G	2.64	1.22	22
1.b	8-40	A C U **A C A C**	U U U C	4.18	0.81	53
	9-16	A C U **A C A C**	U C C	1.89	0.82	24
1.c	9-15	U A G **A C A C**	U A C	nd	nd	nd
	9-36	U A G **A C A C**	U C C	1.89	0.90	29
	9-31	A U C **A C A C**	U U A C	nd	nd	nd
	9-30	A U U **A C A C**	U U A C	nd	nd	nd
1.d	8-17	U U G **A C A C**	A U U C	nd	nd	nd
	9-7	A U G **A C A C**	U C C	1.89	0.90	29
	9-28	A G U **A C A C**	U U U	nd	nd	nd
	8-44	U C G **A C U C**	A U U C	nd	nd	nd
2.a	9-2	U C G **A C U C**	A U A C	nd	nd	nd
	9-23	U C G **A C U C**	U U G	2.83	1.00	29
2.b	8-39	A C G **A C U C**	U C G	3.32	0.68	50
	9-19	A C G **A C U C**	G U A	nd	nd	nd
2.c	8-21	U A G **A C U C**	U U U U C C	1.95	0.84	24
	9-13	U A G **A C U C**	G U G	1.93	1.06	19
	8-18	U U U **A C U C**	U C C	nd	nd	nd
	9-45	U C U **A C U C**	U C A	1.71	1.05	17
	8-24	A C G **A C U C**	U C G	nd	nd	nd
2.d	9-3	C U U **A C U C**	A C C	0.89	1.68	5
	8-3	A A A **A C G C**	A U U C	1.21	0.89	12
	9-32	A U C U **A C U C**	U U C C	1.21	0.89	14
	8-7	U C G **A C** U U	A C C	0.96	0.85	12

[a] All analyzed variants show aminoacylation plateaus of at least 40%. [b] Major species in yeast. [c] Not determined. [d] Gain is defined as $(k_{cat}/K_m)_{rel} = (k_{cat}/K_m)_{mutant}/(k_{cat}/K_m)_{wild-type}$.

C_{53}. For further analysis, the variants could be divided into two families, each with 14 variants, corresponding to two types of sequence homologies within nucleotides 56−53. In family 1, all 14 variants share the wild-type tetranucleotide $A_{56}C_{55}A_{54}C_{53}$. In family 2, A_{54} is predominantly replaced by U (in 13 of 14 molecules) and C_{53} is only replaced once by U_{53}. Thus, most molecules have $A_{56}C_{55}U_{54}C_{53}$, one has $A_{56}C_{55}G_{54}C_{53}$, and another one has $A_{56}C_{55}U_{54}U_{53}$. Interestingly, one variant (9-32, Table 1), has an eight-membered anticodon loop. Further ranking within each family was done according (i) to sequence homology of the 5′ end of the anticodon loop nucleotides $N_{59}N_{58}N_{57}$ (subfamilies) and (ii) to the number of pseudoknot loop L1 nucleotides. Subfamilies 1.d and 2.d (Table 1) contain variants without any sequence homology in the $N_{59}N_{58}N_{57}$ triplet. Except for an $U_{59}A_{58}G_{57}$ conservation in families 1.c and 2.c, none of the sequence combinations for these three nucleotides found in one family is present also in the other family. It is interesting to notice that among the given sequences, none of them presents the canonical wild-type TLS anticodon loop sequence. Further, none of these molecules possesses a canonical tRNA^{Val} anticodon loop sequence. Concerning the pseudoknot, the length of loop L1 varied between 3 and 6 nucleotides with a preference for 3 residues (16 variants) and 4 residues (10 variants). Only one five-membered loop and one six-membered loop were found (Table 1). No variant with less than 3 nucleotides in loop L1 was found.

In both families, consensus sequences of the anticodon loop nucleotides 57−59, as well as in the pseudoknot loops, have been tentatively established, considering a conservation of more than 70% as significant (Figure 3B). In family 1, positions 59 and 57 were equivalently occupied by A/U or G/U, respectively, and position 58 was of random type. In family 2, exactly the same conservations were observed. Thus, a consensus of both sequence families leads to a domination of position 59 by U and A (27 of 28 variants), a random nucleotide at position 58, and G or U at position 57 (26 of 28 variants). Sequence comparisons within the three-membered L1 loops in family 1 (7 variants) revealed the presence of a conserved U at the 5′-end of the loop and strong bias toward C residues at the central and 3′-end positions. Sequence analysis of three-membered L1 loops in family 2 (9 variants) revealed no conservation at the 5′- and 3′-terminal positions and the presence of pyrimidines at the central position of the loop. Altogether, the variants with a three-membered L1 loop, share a preference for a pyrimidine nucleotide at the central position and an U at the 5′-end. Comparison of the four-membered loop L1 sequences in both families (6 variants in family 1 and 4 variants in family 2), leads to the same consensus sequence 5′-$(U/A)_{24}U_{23}(U/A/C)_{22}C_{21}$-3′. Comparison of this consensus with the four-membered loop L1 wild-type TLS reveals that U_{23} is a strictly conserved nucleotide. Interestingly, the complete wild-type sequence of loop L1 was not found within the selected variants.

An alternative classification of all variants, based primarily on the length of the pseudoknot loop and secondarily on the anticodon loop sequence, leads to new consensus combinations in the anticodon loop (Figure 3C). Variants with 4 nucleotides in loop L1 share a strict conservation for U or A residues at position 59, no conservation at positions 57 and 58, C_{56}, A_{55}, a very strong tendency for U or A residues at position 54, and C_{53}. At opposite, those variants with 3 nucleotides in loop L1, have a strict conservation of U or A residues at position 54 and a strict conservation of U or G residues at position 57 as well as a strong preference for A or U residues at position 59. There is no conservation at position 58.

Kinetic Analysis of Selected Variants. The Michaelis−Menten parameters k_{cat} and K_m of a representative set of 17 of the 28 selected variants were established and their valylation efficiency calculated (Table 1). Efficiency corresponds to $(k_{cat}/K_m)_{rel} = (k_{cat}/K_m)_{mutant}/(k_{cat}/K_m)_{wild-type}$ and is expressed here as a "Gain" in activity.

Without exception, all variants exhibited a decreased K_m value, compared to $K_m = 2.44$ μM for the wild-type TLS. The distribution of K_m values of the variants was remarkably narrow with a mean value of 0.97 ± 0.23 μM. Variant 9-3 with a $K_m = 1.68$ μM showed by far the largest deviation from mean. Again without exception, all variants also exhibited an increase in k_{cat} as compared to that of the wild-type TLS (0.24 s^{-1}). Here, however, values between individual variants varied from 0.89 to 4.18 s^{-1}. As a consequence, the relative gain of chargeability is largely attributed to differences in k_{cat}. All selected variants exhibit increased valylation efficiency as compared to the wild-type TLS with gains between 5- and 53-fold. Thus, some variants reach the

efficiency of valylation of the natural substrate of the enzyme, yeast $tRNA^{Val}$, which is about 59-fold higher than that of the wild-type TLS (Table 1).

Variants 8-40 and 8-39 were the most active variants with gains of 53- and 50-fold, respectively. The relative contributions of K_m and k_{cat} to the increased valylation efficiency, however, are rather different to the situation in $tRNA^{Val}$. Both variants showed higher k_{cat} (4−6-fold) and higher K_m (5−6-fold) as compared to $tRNA^{Val}$, which compensate each other and lead to a similar global efficiency. Variant 8-40 has an $A_{59}CUACAC_{53}$ anticodon loop and a four-membered, U-rich, L1 loop ($U_{24}UUC_{21}$), and variant 8-39 has an $A_{59}CGACUC_{53}$ anticodon loop and a three-membered, UCG, L1 loop. Interestingly, even a variant displaying an eight-membered anticodon loop was an excellent substrate for ValRS showing a 14-fold increased aminoacylation efficiency compared to that of the reference transcript. Among selected variants, in general, sequences deviating considerably from consensus sequences were less active substrates for ValRS, the least active being variant 9-3 (5-fold gain), which shows the lowest degree of similarity as compared to consensus sequences.

Importance of Anticodon and Acceptor Domain Nucleotides in Active Variants. A number of variants for which the kinetic parameters have been established, have the same nucleotide composition of loop L1 (Figure 4A). This allows, in turn, a refined analysis of the contribution of anticodon loop nucleotides to aminoacylation efficiency. Two variants have a common $U_{24}UUC_{21}$ sequence in loop L1 and $U_{57}ACAC_{53}$ in the anticodon loop, two others share an UCG sequence of loop L1 and $A_{56}C_{55}$, C_{53} in the anticodon loop, and four others share an UCC sequence in loop L1 and $A_{56}CAC_{53}$ in the anticodon loop. In each of these groups, the efficiency in aminoacylation varies from 1.5−1.9-fold, suggesting that the 3 nucleotides at the 5′-end of the anticodon loop, as well as nucleotide 54, contribute only moderately to aminoacylation efficiency.

To further estimate the influence of the anticodon loop versus L1 loop size/sequence on specific valylation, new variants displaying a same anticodon sequence but different L1 loop size/sequence were designed, constructed, and analyzed (Figure 4B). Thus, the L1 UCG sequence common to two rather active variants (8-8 and 8-39) has been introduced into the sequence context of wild-type TLS (variant B), and the wild-type sequence of loop L1 ($C_{24}UCU_{21}$) has been introduced into the two previous variants, thus leading to variants C and E. In addition, deletion mutants of the wild-type and the variant 8-39, where the L1 loop has been replaced by a single C, have been generated (variants G and H, respectively). The Michaelis−Menten parameters k_{cat} and K_m of these variants have been measured and the relative efficiencies established (Figure 4B). A "vertical" analysis of the properties of the variants displayed in Figure 4B allows again investigation of the anticodon nucleotides effect on aminoacylation. It appears that any deviation from the $C_{59}CC_{57}$ wild-type sequence improves aminoacylation by factors ranging from 5.7 to 32, suggesting that nucleotides 59−57 are not random in the very active variants picked out during the selection process.

Another type of analysis of the valylation properties of these variants, based on sequence variations in loop L1 for a given anticodon loop, gives an additional insight ("horizontal" analysis of Figure 4B). Thus, replacement of the wild-type CUCU sequence of loop L1 by a UCG sequence, affects aminoacylation efficiency by a same factor of 4.4−6.4, in each of the three anticodon contexts. In other words, the exchange of the wild-type L1 loop by a sequence originating from a selected variant has a similar effect within the context of three different anticodon loop sequences. The two variants having the L1 loop replaced by a single C are only poorly valylated, and their relative specificity constants are decreased by factors of 50 (variant H) and 200 (variant G) as compared to the wild-type TLS. Although the 4-fold stronger effect observed for variant G may reveal a small differential effect of shortening the L1 loop according to the sequence of the anticodon loop, it cannot be excluded that this variation is within the higher error range of measuring low activities and is thus not significant.

Relationship between Anticodon and Acceptor Domains. An accurate estimation of the relationship between the anticodon and the pseudoknot domains of the TLS, can be deduced from comparisons between experimental and theoretical values of the aminoacylation efficiency (*45*). Variant D can be considered as derived from the wild-type TLS by a combination of two major mutations, namely replacement of the anticodon domain (as in variant C) and replacement of the pseudoknot domain (as in variant B). If the two structural domains contribute independently to aminoacylation efficiency of variant D, the theoretical gain calculated on the basis of the individual gains of the single mutated variants B and C should lead to the same value as the experimentally determined gain. This is indeed the case with an experimental gain of 29-fold and a theoretical gain of $5.7 \times 4.4 = 25$-fold. Along the same line, variant F can be considered as a combination of mutational steps going from wild-type TLS to variant E in regard of the anticodon loop and to variant B in regard of the pseudoknot. The theoretical gain calculated for the double variant F, $7.8 \times 4.4 = 34$, is very close to the experimental gain of 50 (ratio of 1.47 between the two values). Similarly, comparison between theoretical and experimental efficiencies in valylation of variant F (originating from C via E and D, experimental gain of $50/5.7 = 8.7$ versus theoretical gain of $1.4 \times 5.1 = 7.2$), variant H (originating from the wild-type via E and G, 0.16 versus $0.05 \times 7.8 = 0.39$), and variant H (originating from variant B via variants F and G, 0.036 versus theoretical gain of 0.012) revealed similar values, with a largest ratio of 3 between the two values. Thus, within the limits of errors, it can be considered that the anticodon and the pseudoknot domains contribute mainly independently to aminoacylation.

To test whether the selection against the wild-type anticodon loop trinucleotide $C_{59}CC_{57}$ was dependent on the $G_{85}GG_{83}$-stretch at the 5′-end of the variants (Figure 1A), we also analyzed a wild-type version of the TYMV 3′-end RNA sequence and an engineered version of the selected variant 8-8 carrying a single guanine at the 5′-end (G_{83}) instead of the GGG-stretch. However, no significant changes in valylation efficiency were observed (data not shown).

Theoretical Structural Analysis of the TYMV TLS Library. The genetic algorithm implemented as part of the STAR program (*44*) and which allows prediction of pseudoknots was used both to analyze the potential folding of the selected variants and to have an insight about the distribution of secondary structures of RNAs from the initial library with

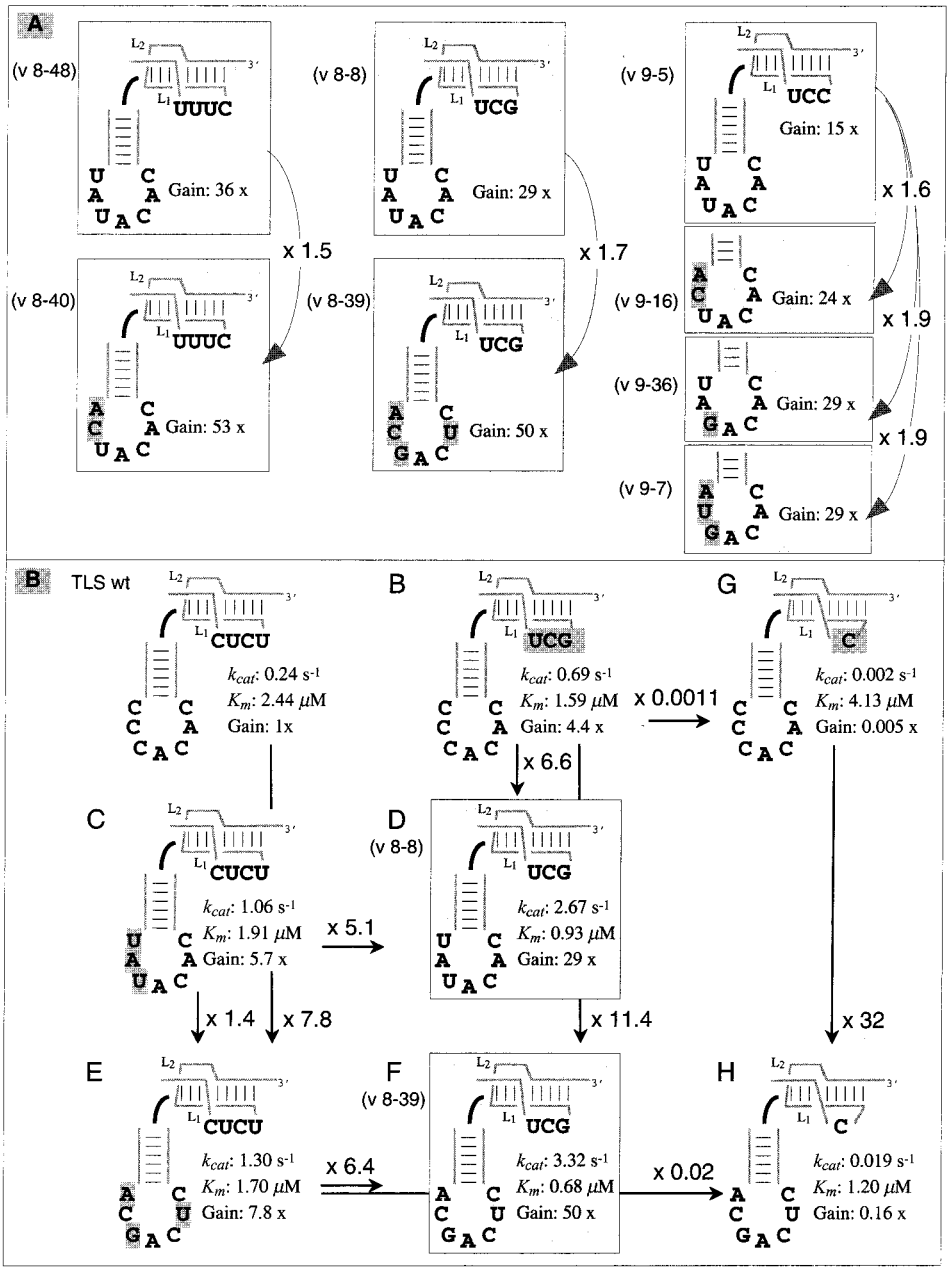

FIGURE 4: Contribution of anticodon loop and pseudoknot loop L1 nucleotides of variant TYMV TLS to efficient valylation by yeast ValRS. The TLS are represented in a schematic way with only sequences of importance explicitly indicated. Variants which have emerged at either round 8 or 9 of the selection procedure appear in a gray background, those which correspond to the wild-type sequence or which have been engineered appear without background. Sequence differences between investigated molecules and reference molecules are dashed. (A) Comparative analysis of selected variants sharing a same pseudoknot sequence but with different anticodon loop nucleotides at positions 59, 58, 57, and 54. Names of variants are given on the left sides of the panels and are according to Table 1. Arrows indicate the gain of valylation efficiency between 2 variants. (B) Comparative analysis of selected and engineered TLS. A vertical analysis allows further analysis of the contribution of anticodon nucleotides 59, 58, 57, and 54 to valylation, whereas a horizontal comparison allows an investigation of the contribution of pseudoknot nucleotides. Moreover, since some variants can be considered as global "double mutants" of others, theoretical efficiencies can be estimated on the basis of additivity of the two domains, and compared to those determined experimentally. Thus, variants D, F, and H can be considered as double variants of the wild-type TLS via variants B/C, B/E, and G/E, respectively.

special attention to the influence of pseudoknot loop L1 sequences.

All except one of the 28 active variants summarized in Table 1 were predicted to fold into the wild-type secondary structure as depicted in Figure 1A. Interestingly, the only active variant predicted to fold into an alternative structure was the less active one among all the functional variants (9-3). In contrast, only five of 16 inactive variants were predicted to fold into the wild-type secondary structure, suggesting that correct folding is a necessary but not sufficient prerequisite to high level valylation.

To analyze more rigorously the effect of loop L1 sequence on the global folding of TYMV RNA variants, the secondary structure of all possible variants of the wild-type sequence with the L1 loop sequence replaced by any of the possible tri- and tetranucleotide sequences (a total of 320 sequences) was also predicted. Approximately 73% of the tetranucleotide loops and 78% of the trinucleotide loops were predicted to fold into the wild-type secondary structure and, thus allow for pseudoknot formation. This means that 27 and 22%,

respectively, of the molecules do have sequences in loop L1 which hinder correct folding. Interestingly, none of the molecules that contain a trinucleotide loop with either an A or a U at the 5′-end (which is the case for 14 out of 16 of the in vitro selected variants) were predicted to fold into alternative structures as compared to the wild-type secondary structure. In contrast, 44% of the molecules with trinucleotide loops with a G or a C residue at their 5′-end were predicted to fold into alternative structures. In regard of tetranucleotide L1 loops, only 4% of the molecules that contain a U residue at position 23 (which was strictly conserved in the selected variants) were predicted to fold into alternative structures. In contrast, depending on the nature of the nucleotide, between 38 and 53% of the tetranucleotide loops with nucleotides other than U at position 23, were predicted to fold into alternative structures. No such correlation was found for position 21, which also was strictly conserved in four-membered L1 loops.

DISCUSSION

Key Features for Efficient Valylation of TLS by Yeast ValRS. This study was intended (i) to search for alternative valylation identity elements in the anticodon loop for valylation of the TYMV tRNA-like structure and (ii) to search for alternative structures/sequences in the pseudoknot domain. Accordingly, a library of TYMV RNA variants was designed with mutations concentrated in these two domains. Starting with a pool of 2×10^{13} variants that completely covers the mutant space, nine rounds of selection were carried out, using a method for the selection of specifically aminoacylatable RNAs in vitro (*39*). Combined kinetic and sequence analyzes of 51 selected individual RNAs lead to the following major conclusions: (i) all selected active variants are more efficient substrates for ValRS than the wild-type TYMV TLS, (ii) selection of variants as active as the cognate substrate of the enzyme, yeast $tRNA^{Val}$, has occurred, (iii) all variants do fold into the pseudoknotted tRNA-like structure, (iv) anticodon loop and pseudoknot domains act independently, (v) all active variants possess the central and 3′-end anticodon-triplet nucleotides A_{56} and C_{55}, (vi) valylation tolerates variability in size and sequence of loop L1 from the pseudoknot.

Independence of Anticodon Loop and Pseudoknot Domains. The variability of selected active variants both in sequence of the complete anticodon loop and in sequence and size of the pseudoknot loop L1 confirms that both structural domains of the TLS contribute to efficient recognition and aminoacylation by yeast ValRS. To verify whether these domains contribute independently or in a cooperative manner (*45*), the efficiency of a number of mutants bearing exchanged anticodon or pseudoknot domains were calculated on a theoretical basis and compared to the experimental data. Only a limited number of variants which were picked up during the in vitro selection process were useful for those comparisons, so that a series of rationally engineered additional variants were created (Figure 4B). Experimental and theoretically estimated aminoacylation efficiencies on the basis of additivity of the anticodon and pseudoknot domains of engineered variants revealed that they vary at most by a factor 3. Hence, based on the set of data obtained so far, the contributions of both domains to the overall valylation activity are likely additive. Thus, the anticodon loop and the pseudoknot loop L1 can be considered as separable and independent entities. This knowledge allows now for easier interpretation of data to estimate the role of individual nucleotides in valine identity of the TLS variants.

Refined View of Anticodon Loop Recognition. Approximately 65% of variants from a random sample of RNAs from rounds 8 and 9 (with mutations in the randomized domains) revealed a substrate activity toward ValRS that was significantly increased as compared to the wild-type. All these molecules contained nucleotides A_{56} and C_{55}. The remaining 35% of variants showed plateau levels of charging of less than 1% using standard conditions. At less stringent conditions, however, sequences containing an A residue at position 56 could be charged to levels of up to 40%. Since no consensus motifs could be traced within the group of less active variants, we assume that these mutants were co-selected by some unspecific adsorption to magnetic beads rather than through some specific function. Consequently, these variants were excluded from further analysis. However, the very low activity of these variants at opposite to the spectacular activity of those RNAs bearing the two nucleotides A_{56} and C_{55} demonstrates the importance of both nucleotides for valylation specificity. Similar results were obtained for C_{53} which was substituted only once (out of 28 sequences) by a U residue. Thus, A_{56}, C_{55}, and C_{53} emerge as primary identity elements required for valylation of the TLS by yeast ValRS, whatever the size and sequence of loop L1 of the pseudoknot. This extends previous mutational studies which showed that A_{56}, C_{55}, and C_{53} are major identity elements for wheat germ valylation (*23*, *24*) to the yeast enzyme. These nucleotides are found in the consensus sequences of both valylatable plant viral tRNA-like structures (*6*, *8*) and yeast $tRNA^{Val}$ (*46*) (Figure 1B). It is thus concluded that A_{56}, C_{55}, and C_{53} are necessary and sufficient valylation identity elements. Under the stringent conditions used in the present work, no alternate combination of nucleotides at these positions, efficiently recognized by yeast ValRS, could be selected.

At position 54, variants showed predominantly the wild-type A residue (50% of variants) or a U residue (46% of variants) which lead to the general classification into two families according to the consensus sequences $A_{56}CAC_{53}$ or $A_{56}CUC_{53}$ of the 3′-end of the loop (Table 1). Although A_{54} is strictly conserved in plant viral valine-specific tRNA-like structures, previous mutational studies on TYMV RNA indicated that position 54 had a moderate contribution to valylation efficiency (*23*, *24*) with C being the less favorable substitution. In line with these data, none of the selected molecules contained a C residue at this site. Interestingly, all canonical tRNAs (*47*) possess a purine at the equivalent position. The selection of valylatable variants with a U at this position suggests that the presence of a purine in canonical tRNAs is linked to other functions than aminoacylation. Alternatively, tolerance of a pyrimidine in the TYMV TLS may be linked to new structural features of the anticodon loop linked to nucleotides from the 5′-side of the loop.

Nucleotides at positions 57−59 forming the 5′-end of the anticodon loop, although variable, are not random in the selected valylatable variants. In particular, position 57 and 59 showed absolute preferences for G/U and A/U, respectively (Figure 3, panels B and C). Similar preferences for

purines at these positions can be seen in the TLS of furoviruses (*8*) as well as in TLSs from various tymoviruses (*6*). Surprisingly, none of the wild-type cytosines was found at these positions. By varying the sequence of the very 5′-end of the complete TLS, it could be demonstrated that the selection against the wild-type $C_{59}CC_{57}$ sequence at these positions was not due to the nonviral G-rich sequences at the 5′-end of the transcripts that potentially could favor the formation of an (inactive) alternative structure (*48*). It is noteworthy that in all 18 $tRNA^{Val}$ genes from yeast, position 32, which corresponds to position 59 in TYMV RNA, is exclusively occupied by an U residue, which is also the predominant nucleotide found in the selected variants.

Canonical tRNAs possess a strictly conserved U residue at position 33 (*47*), equivalent to position 58 in the TLS. The role of this nucleotide is of structural nature and allows a turn of the RNA chain, so that tRNA nucleotides 34−38 (equivalent to 57−53 in TLS, see Figure 1A) are stacked on the anticodon stem helix (reviewed in ref *49*). In the selected TYMV TLS, no conservation of a U at this position was found. This likely is again related to the selection pressure applied, which involves only recognition by ValRS, and not binding to the ribosome and messenger RNA, as was the case during evolution of natural tRNAs. The fact that natural viral tRNA-like sequences are not involved in protein synthesis (*2*) can likely explain the general absence of an U at position 58. Residue 57, which corresponds to the wobble nucleotide of the anticodon triplet, has a strong preference for U/G, whereas in the natural $tRNA^{Val}$ species, it presents a large variability (*46*). Here again, selection pressure did not take into account other functions than valylation and especially did not take into account codon reading.

Comparison of aminoacylation efficiencies of selected and engineered variants sharing a same pseudoknot structure/sequence but different anticodon loop nucleotides allows the search for a role of nucleotides 59, 58, and 57. Interestingly, as soon as the sequence deviates by at least two positions from the wild-type sequence CCC, there is a strong gain in aminoacylation efficiency (by factors varying from 5.7 to 32). However, provided two C's are mutated, any tested combination of nucleotides lead to the same effect. Thus, it appears that these nucleotides contribute in a significant way to the aminoacylation properties of the RNAs. Since all non-C combinations are more favorable for valylation and are also present in the consensus yeast $tRNA^{Val}$ (especially U_{32}, equivalent to U_{57}), it is concluded that the three C residues occurring in the wild-type TLS introduce negative effects for recognition by yeast ValRS. It is worth mentioning that the three-C stretch is absent from any valylatable TLS with the exception of KYMV in addition to TYMV (*6*). Thus, the sensitivity of plant ValRS to these nucleotides may be somewhat different from that of the yeast enzyme.

In conclusion, yeast ValRS is able to charge with high efficiency a series of substrates which do have non canonical anticodon loops. In these anticodon loops, two very conserved elements are absent, namely U_{33} and a purine at position 37. Crystallographic structures of tRNAs complexed to their cognate aminoacyl-tRNA synthetase [e.g., in the glutamic and aspartic systems (*50*−*53*)] highlight indeed that the anticodon loop structure is completely changed as compared to that of a noncomplexed tRNA. This situation likely occurs also for tRNAs interacting with ValRSs. Thus, the native structure of the loop, typically linked to the presence of residue U_{33}, is likely not important for aminoacylation but for the other functions of the tRNA, such as binding to the ribosome and to the messenger RNA.

Contribution of the Pseudoknot Loop L1 to Valylation. In the course of the present selection experiments with yeast ValRS, those variants were enriched that contain a three- or a four-membered L1 loop, indicating that these loop sizes are more favorable for specific valylation. Variants with loop L1 shorter than 3 nucleotides vanished completely, whereas there still remained some variants with five- and six-membered loop L1. This is in agreement with the phylogenetic comparison of different tymoviral and furoviral RNAs (*6*, *8*), in which L1 loops may adopt various sizes and sequences and still keep excellent substrate properties to ValRS. In addition, previous studies with wheat germ ValRS had revealed that mutants remained excellent substrates for ValRS as long as the overall A-type RNA helix geometry and the pseudoknot stability are maintained (*29*). Here, no simple correlation could be established between loop L1 size and aminoacylation efficiency. Indeed, in some cases, mutants with larger L1 loops performed better than the concomitant variants with smaller L1 loops and vice versa.

Interestingly, for either of the two major classes of L1 loop sizes, consensus sequences were observed, which allowed for an alternative classification of active variants (Figure 3C). The class of variants with a four-membered L1 loop contained two remarkably conserved nucleotides, C_{21} and U_{23}. The latter is also conserved with respect to wild-type TYMV RNA. Similarly, in the three-membered L1 loops, there is a conserved pyrimidine residue.

It is likely that these sequences have been selected for structural reasons rather than for contacts with the synthetase. This assumption is based on two main considerations. First, RNA secondary structure prediction of variants of the wild-type TLS with all possible four-membered L1 loop sequences revealed that alternative folding of molecules can be effectively suppressed by keeping a U conserved at position 23. As a result of the statistical survey, 96% of all possible four-membered L1 loop variants with an U_{23} are predicted to fold into the wild-type secondary structure. In agreement with these findings, 96% of the family of active sequences from selection rounds 8 and 9, but only 30% of the inactive molecules, were predicted to fold into the wild-type secondary structure. Second, the recently solved crystal structure (*54*) and NMR structure (*55*) of pseudoknots, and especially of the TYMV TLS acceptor stem (*55*), allow for new considerations on the relative importance on individual nucleotides in folding and/or stabilization of the pseudoknot. Interestingly, the conserved residue U_{23} in our selected active four-membered L1 loop variants has a particular fully extended conformation so that to cross the major groove. Probably, the unusual conformation of base 23, necessary to cross the major groove, determines the nature of the nucleotide at this position. At position 21, throughout all valylatable tymoviral RNAs (except CcTMV RNA), a pyrimidine is conserved. In contrast to U_{21} of wild-type TYMV RNA, selected sequences contain a C which may also stack on the helix.

The class of active variants with three-membered L1 loops contains a strongly conserved pyrimidine in the central position of the loop and has a strong preference of U at the 5′-end of the loop. RNA secondary structure prediction of variants of the wild type TLS transcript with all possible three-membered L1 loop sequences revealed that alternative folding of molecules can be totally suppressed by keeping a U (or an A) conserved at position 23.

Finally, class I synthetases, the family to which belongs ValRS, bind to the minor groove side of the acceptor helix, in the vicinity of the extended 3′-end (reviewed in ref *32*). Footprinting experiments between ValRS and the TYMV TLS (*56*), as well as comparison of the TLS NMR structure with the X-ray structure of $tRNA^{Glu}$ complexed with its cognate synthetase, a class I enzyme (*50*, *51*), reveals that both the L1 and the L2 loop are at the side opposite from where the enzyme binds (*55*). Thus, loop L1 residues may not directly interact with the synthetase, and conserved residues more likely allow the formation of the pseudoknot structure and stabilize it. Further analysis has to confirm, whether structural requirements, i.e., structural flexibility facilitating partial refolding imposed by the synthetase upon substrate binding, rather than direct recognition, are the basis of conservation of sequence elements within the L1 loop of selected TYMV RNA variants.

No variant with a single nucleotide in loop L1 was selected under the stringent conditions used here. However, a deletion of loop L1 down to a single C surprisingly resulted in a 1.2-fold increase in aminoacylation relative to wild-type TYMV RNA when tested with wheat germ ValRS (*29*). Analysis of two engineered variants constructed in the present work and with a single nucleotide in loop L1 leads to 50−200-fold losses when tested with the yeast enzyme. Thus, the two enzymes may have different levels of tolerance to variability in the pseudoknot structure. However, since the aminoacylation efficiencies are very sensitive to reaction conditions (*7*), it is also possible that the differences in aminoacylation properties observed reflect this sensitivity.

General Conclusions and Perspectives. The structure−function relationship in tRNAs for their aminoacylation by synthetases has been conceptualized by the paradigm according to which the tRNA structure is a scaffolding allowing optimal presentation of the identity elements to the enzyme (*32*, *57*). This view is supported by the functionality of numerous tRNAs engineered in their three-dimensional structure. It is also supported by the existence in nature of mimics such as the TLS in plant RNA viruses. In the present work, we intended to explore TLS variants recognized by yeast ValRS. Even if a great structural variability already exists when comparing different natural tymoviral or furoviral TLSs, it was interesting to know whether additional structural alternatives could exist, compatible with efficient valylation. To address this question, we chose a priori to modify the anticodon loop and part of the pseudoknot within the TYMV TLS, two structural elements known to be important for valine identity and folding of the amino accepting end, respectively. Surprisingly, none of the active variants which were selected did share the wild-type TYMV sequence except for the identity nucleotides in the anticodon loop, despite the very strong selection conditions used. Even more, these variants were more efficiently charged than the control wild-type TLS. Altogether, analysis of the variants further supports the paradigm: (i) no canonical anticodon loop structure, as required for protein synthesis, is a prerequisite for efficient recognition by ValRS, (ii) a large structural variability is allowed in loop L1 of the pseudoknot, but (iii) the identity elements in the anticodon must be present.

Since the TLSs present at the 3′-end of viral RNAs are of biological importance, especially in initiation of minus-strand replication (reviewed in ref *4*), one can question about the activity of the newly selected variants in the TYMV life cycle. Since this activity implies recognition of the TLS by the plant synthetase (*58*), elongation factor EF-1α (*17*), the nucleotidyl-transferase (CCAse) (*59*), the viral replicase (*60*), and presumably by other protein factors such as the coat protein (*61*), it would be interesting to elucidate whether optimal sequences for yeast ValRS, both in the L1 loop and in the anticodon loop are "permissive elements" (*34*) for these proteins. Thus, it will be informative to analyze RNAs selected for optimal valylation in vitro for their "viability" in terms of replication in vitro as well as in vivo.

ACKNOWLEDGMENT

We thank C. Luge for technical assistance. The authors greatly appreciated interesting discussions with M. Sissler and would also like to thank G. Keith for providing a compilation of the $tRNA^{Val}$ genes from yeast.

REFERENCES

1. Hall, T. C. (1979) *Int. Rev. Cytol. 60*, 1−26.
2. Haenni, A.-L., Joshi, S., and Chapeville, F. (1982) *Prog. Nucleic Acid Res. Mol. Biol. 27*, 85−104.
3. Mans, M. W., Pleij, C. W. A., and Bosch, L. (1991) *Eur. J. Biochem. 201*, 303−324.
4. Florentz, C., and Giegé, R. (1995) in *tRNA: Structure, Biosynthesis, and Function.*, (Söll, D., and RajBhandary, U. L., Eds) pp 141−163, Am. Soc. Microbiol. Press, Washington DC.
5. Pinck, M., Yot, P., Chapeville, F., and Duranton, H. (1970) *Nature 226*, 954−956.
6. Van Belkum, A., Bingkun, J., Rietveld, K., Pleij, C. W. A., and Bosch, L. (1987) *Biochemistry 26*, 1144−1151.
7. Dreher, T. W., and Goodwin, J. B. (1998) *Nucleic Acids Res. 26*, 4356−4364.
8. Goodwin, J. B., and Dreher, T. W. (1998) *Virology 246*, 170−178.
9. Hall, T. C., Shih, D. S., and Kaesberg, P. (1972) *Biochem. J. 129*, 969−976.
10. Kohl, R. J., and Hall, T. C. (1974) *J. Gen. Virol. 25*, 257−261.
11. Agranovsky, A. A., Dolya, V. V., Gorbulev, V. G., Kozlov, Y. V., and Atabekov, J. G. (1981) *Virology 113*, 174−187.
12. Oberg, B., and Philipson, L. (1972) *Biochem. Biophys. Res. Commun. 48*, 927−932.
13. Carriquiry, E., and Litvak, S. (1974) *FEBS Lett. 38*, 287−291.
14. Felden, B., Florentz, C., McPherson, A., and Giegé, R. (1994) *Nucleic Acids Res. 22*, 2882−2886.
15. Rudinger, J., Florentz, C., and Giegé, R. (1994) *Nucleic Acids Res. 22*, 5031−5037.
16. Felden, B., Florentz, C., Westhof, E., and Giegé, R. (1998) *Biochem. Biophys. Res. Commun. 243*, 426−434.
17. Dreher, T. W., Tsai, C.-H., and Skuzeski, J. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12212−12216.
18. Yot, P., Pinck, M., Haenni, A. L., Duranton, H., and Chapeville, F. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 1345−1352.
19. Giegé, R., Briand, J.-P., Mengual, R., Ebel, J.-P., and Hirth, L. (1978) *Eur. J. Biochem. 84*, 251−256.

20. Dreher, T. W., Florentz, C., and Giegé, R. (1988) *Biochimie 70*, 1719−1727.
21. Joshi, S., Haenni, A.-L., Hubert, E., Huez, G., and Marbaix, G. (1978) *Nature 275*, 339−341.
22. Joshi, S., Chapeville, F., and Haenni, A.-L. (1982) *EMBO J. 1*, 935−938.
23. Florentz, C., Dreher, T. W., Rudinger, J., and Giegé, R. (1991) *Eur. J. Biochem. 195*, 229−234.
24. Dreher, T. W., Tsai, C.-H., Florentz, C., and Giegé, R. (1992) *Biochemistry 31*, 9183−9189.
25. Chu, W.-C., and Horowitz, J. (1991) *Biochemistry 30*, 1655−1663.
26. Pallanck, L., and Schulman, L. H. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3872−3876.
27. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1991) *Biochem. Biophys. Res. Commun. 177*, 619−623.
28. Horowitz, J., Chu, W. C., Derrick, W. B., Liu, J. C., Liu, M., and Yue, D. (1999) *Biochemistry 38*, 7737−7746.
29. Mans, R. M. W., Van Steeg, M. H., Verlaan, P. W. G., Pleij, C. W. A., and Bosch, L. (1992) *J. Mol. Biol. 223*, 221−232.
30. McClain, W. H. (1993) *FASEB J. 7*, 72−78.
31. Saks, M. E., Sampson, J. R., and Abelson, J. N. (1994) *Science 263*, 191−197.
32. Giegé, R., Sissler, M., and Florentz, C. (1998) *Nucleic Acids Res. 26*, 5017−5035.
33. Saks, M. E., and Sampson, J. R. (1996) *EMBO J. 15*, 2843−2849.
34. Frugier, M., Helm, M., Felden, B., Giegé, R., and Florentz, C. (1998) *J. Biol. Chem. 273*, 11605−11610.
35. Hou, Y. M., and Schimmel, P. (1992) *Biochemistry 31*, 10310−10314.
36. Sissler, M., Giegé, R., and Florentz, C. (1996) *EMBO J. 15*, 5069−5076.
37. Sissler, M., Giegé, R., and Florentz, C. (1998) *RNA 4*, 647−657.
38. Tsai, C. H., and Dreher, T. W. (1992) *J. Virol. 66*, 5190−5199.
39. Pütz, J., Wientges, J., Sissler, M., Giegé, R., Florentz, C., and Schwienhorst, A. (1997) *Nucleic Acids Res. 25*, 1862−1863.
40. Kern, D., Giegé, R., Robbe-Saul, S., Boulanger, Y., and Ebel, J.-P. (1975) *Biochimie 57*, 1167−1176.
41. Becker, H. D., Giegé, R., and Kern, D. (1996) *Biochemistry 35*, 7447−7458.
42. Freiberg, C., Perret, X., Broughton, W. J., and Rosenthal, A. (1996) *Genome Res. 6*, 590−600.
43. Van Batenburg, F. H. D., Gultyaev, A. P., and Pleij, C. W. A. (1995) *J. Theor. Biol. 174*, 269−280.
44. Abrahams, J. P., Van der Berg, M., Van Batenburg, E., and Pleij, C. W. A. (1990) *Nucleic Acids Res. 18*, 3035−3044.
45. Pütz, J., Puglisi, J. D., Florentz, C., and Giegé, R. (1993) *EMBO J. 12*, 2949−2957.
46. Hani, J., and Feldmann, H. (1998) *Nucleic Acids Res. 26*, 689−696.
47. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
48. Mans, R. M. W., Verlaan, P. W. G., Pleij, C. W. A., and Bosch, L. (1990) *Biochim. Biophys. Acta 1050*, 186−192.
49. Dirheimer, G., Keith, G., Dumas, P., and Westhof, E. (1995) in *tRNA: Structure, Biosynthesis, and Function* (Söll, D., and RajBhandary, U. L., Eds.) pp 93−126, Am. Soc. Microbiol. Press, Washington DC.
50. Rould, M. A., Perona, J. J., Söll, D., and Steitz, T. A. (1989) *Science 246*, 1135−1142.
51. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) *Nature 352*, 213-218.
52. Ruff, M., Boeglin, M., Cavarelli, J., Krishnaswamy, S., Mitschler, A., Podjarny, A., Poterszman, A., Rees, B., Thierry, J.-C., and Moras, D. (1992) in *Structure and Function* (Sarma, R. H., and Sarma, M. H., Eds.) pp 159−168, Adenine Press.
53. Cavarelli, J., Rees, B., Eriani, G., Ruff, M., Boeglin, M., Gangloff, J., Thierry, J.-C., and Moras, D. (1994) *EMBO J. 13*, 327−337.
54. Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999) *Nat. Struct. Biol. 6*, 285−292.
55. Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W., Heus, H. A., and Hilbers, C. (1998) *Science 280*, 434−438.
56. Florentz, C., and Giegé, R. (1986) *J. Mol. Biol. 191*, 117−130.
57. Giegé, R., Frugier, M., and Rudinger, J. (1998) *Curr. Opin. Struct. Biol. 8*, 286−293.
58. Tsai, C. H., and Dreher, T. W. (1991) *J. Virol. 65*, 3060−3067.
59. Rao, A. L. N., Dreher, T. W., Marsh, L. E., and Hall, T. C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5335−5339.
60. Deiman, B. A. L. M., Koenen, A. K., Verlaan, P. W. G., and Pleij, C. W. A. (1998) *J. Virol. 72*, 3965−3972.
61. Rohozinski, J., and Hancock, J. M. (1996) *J. Gen. Virol. 77*, 163−165.
62. Rietveld, K., Van Poelgeest, R., Pleij, C. W. A., Van Boom, J. H., and Bosch, L. (1982) *Nucleic Acids Res. 10*, 1929−1946.
63. Bonnet, J., Ebel, J.-P., Dirheimer, G., Shershneva, L. P., Krutilina, A. I., Venkstern, T. V., and Bayev, A. A. (1974) *Biochimie 54*, 1211−1213.

BI992852L