

Common and specific amino acid residues in the prokaryotic polypeptide release factors RF1 and RF2: possible functional implications

Nina J. Oparina*, Olga V. Kalinina¹, Mikhail S. Gelfand^{2,3} and Lev L. Kisselev

Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Vavilov street 32, Moscow 119991, Russia, ¹Department of Bioengineering and Bioinformatics, Moscow State University, Vorob'evy Gory, 1-73, Moscow 119992, Russia, ²Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoi Karetnyi per., 19, Moscow 127994, Russia and ³State Scientific Centre GosNII Genetika, 1st Dorozhny pr. 1, Moscow, 113545, Russia

Received as resubmission June 20, 2005; Revised and Accepted August 30, 2005

ABSTRACT

Termination of protein synthesis is promoted in ribosomes by proper stop codon discrimination by class 1 polypeptide release factors (RFs). A large set of prokaryotic RFs differing in stop codon specificity, RF1 for UAG and UAA, and RF2 for UGA and UAA, was analyzed by means of a recently developed computational method allowing identification of the specificity-determining positions (SDPs) in families composed of proteins with similar but not identical function. Fifteen SDPs were identified within the RF1/2 superdomain II/IV known to be implicated in stop codon decoding. Three of these SDPs had particularly high scores. Five residues invariant for RF1 and RF2 [invariant amino acid residues (IRs)] were spatially clustered with the highest-scoring SDPs that in turn were located in two zones within the SDP/IR area. Zone 1 (domain II) included PxT and SPF motifs identified earlier by others as 'discriminator tripeptides'. We suggest that IRs in this zone take part in the recognition of U, the first base of all stop codons. Zone 2 (domain IV) possessed two SDPs with the highest scores not identified earlier. Presumably, they also take part in stop codon binding and discrimination. Elucidation of potential functional role(s) of the newly identified SDP/IR zones requires further experiments.

INTRODUCTION

Three triplets, UAA, UAG and UGA, located at the end of coding mRNA sequences, are signals for terminating

polypeptide synthesis on ribosomes. When one of these stop codons encounters the ribosomal A site while the P site is occupied by peptidyl-tRNA, a protein, called class 1 release factor (RF), binds to the A site [for recent reviews see (1–4)]. In prokaryotes RF1 decodes UAA and UAG stop codons, whereas RF2 is specific for UAA and UGA. The genetic code of mitochondria and mycoplasmas is modified and does not contain the UGA stop codon. Consequently, they possess only one RF that corresponds to bacterial RF1 (5). In eukaryotes, only one factor, named eRF1, decodes all three stop codons (6). Although it has been proved that eRF1 rather than the ribosome itself determines the specificity of stop codon decoding (7,8), the decoding mechanism remains enigmatic despite the fact that some amino acids essential for stop codon discrimination by eRF1 have already been identified (9–11).

In contrast to eukaryotes, for prokaryotic factors two 'protein anticodons' have been identified, PxT in RF1 and SPF in RF2 (12,13). Site-specific mutagenesis of these residues both *in vivo* and *in vitro* led to a switch in the stop codon specificity of the mutant *Escherichia coli* RFs. Presumably, these residues are involved in discrimination of the second and third purines of stop codons (12–14). However, it remains unknown whether the decoding specificity of RF1/RF2 is determined solely by these tripeptides or other amino acids also contribute to the discrimination process.

In an attempt to get better insight into stop codon recognition by bacterial RFs, we applied a newly developed algorithm SDPpred that automatically selects amino acid residues accounting for differences in functional specificity among homologous proteins (15,16). The prerequisite for applying this method is a common biochemical function for the selected family of proteins and different specificity in subfamilies. The RF1/RF2 case meets this requirement because of their well-known functional and structural similarity. This method

*To whom correspondence should be addressed. Tel: +7 095 1351419; Fax: +7 095 1351405; Email: nixie@eimb.ru

predicts specificity-determining positions (SDPs) that by definition are positions of multiple alignment which are conserved within subfamilies consisting of the proteins with the same specificity, but differ between these subfamilies. The major advantage of this approach stems from the fact that in contrast to some other previously suggested techniques, SDPpred directly takes into account the non-uniformity of amino acid substitution frequencies, automatically selects the fraction of alignment positions with the highest correlations with the specificity and finds differences that are least probable to arise by chance during evolution. Furthermore, SDPpred does not require information about the 3D protein structure, which instead can be used at the post-processing step to validate the predictions. There are also other approaches for comparative analysis of protein subfamilies. One of these methods, called DIVERGE (17) takes into account the differences in mutation rates in different parts of the phylogenetic tree of the protein family and identifies sites that had experienced an RS during their evolution.

In this work we compared the results of RF1/RF2 family analysis by means of SDPpred and DIVERGE. In addition we considered invariant amino acid residues (IRs) common for both subfamilies. All identified SDPs and IRs were mapped onto the 3D structure of both RFs. The bioinformatics approach predicts new functionally essential amino acid residues in RF family.

MATERIALS AND METHODS

Preparation of datasets

The RF1 and RF2 amino acid sequences of *E. coli* were used as the seeds for the BLAST (18) search. RF-like proteins (1334) were found in the current protein sequence databases (GenBank translations, PIR and others). We used the ClustalX software (<ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/>) (19) to align all these sequences. A maximum-likelihood species tree was reconstructed with the PROTML with the JTT amino acid replacement model from the PHYLIP package (<http://evolution.genetics.washington.edu/phylip.html>) (20). RF sequences from mycoplasma or mitochondrial genomes were not included in the training set, as they are deficient in RF2 and could therefore distort the pattern of RF1 specificity. Based on the multiple alignment and the phylogenetic tree we discarded all weak homologs, and all RFs containing large N- or C-terminal deletions. Then, among the remaining organisms only those with both RF1 and RF2 sequences were used for further analysis. Thus we obtained the multiple alignment of 234 RF1/RF2 orthologs from the genomes of various prokaryotes, consisting of 117 RF1s and 117 RF2s. The final alignment was manually controlled and edited using the BIOEDIT multiple alignment analysis package (<http://www.mbio.ncsu.edu/BioEdit/page2.html>). Selected sequences and alignments are available as Supplementary Data.

Selection of SDPs and RSPs

The multiple sequence alignment (MSA) of 234 RF sequences was used as the training set for SDP selection. SDPpred compares amino acid residues in paralogous protein subfamilies: it computes mutual information for each position of the

MSA; then, using random shuffling of each column, calculates the statistical significance of the observed values (*Z*-scores); and finally, using the Bernoulli estimator procedure, finds those columns of the MSA that have the highest *Z*-scores (SDPs) and have the minimal probability to arise by chance. SDPpred is available at <http://math.genebee.msu.su/~psn>. The set of obtained SDPs was tested as the profile for attribution of RF protein sequences not included into the training set.

The DIVERGE (17) algorithm utilizes two models of mutation rate variation among sites: the 'homogeneous gamma model', which assumes that the mutation rate at each site remains constant over the whole history of a protein family, and the 'non-homogeneous gamma model', which allows sites to mutate at different rates in various branches of the family tree (RS). DIVERGE calculates posterior probabilities of each position to follow the non-homogeneous model of functional divergence. If this probability equals 1 or is close to 1, this indicates positions, which experienced the RS. We refer to these positions as RSP positions. We used the cutoff of 0.99 for selection of RSPs.

Selection of IRs clustered with SDPs

Amino acid residues invariant in both RF1 and RF2 (IRs) were selected in the same MSA that was used for SDP and RSP prediction. For filtering of potential sequence errors we compared IRs list with 'highly conserved residues' (in this case we allowed 98% identity). The spatial localization of the selected IRs was further analyzed. We calculated the centroids (centres of mass of their side chains) for all SDPs and IRs using the SYBYL software (<http://www.tripos.com>). The crystal structures for both RF1 (21) and RF2 (22) were used for SDP and IR centroid identification. We calculated pairwise distances between these centroids (data not shown) for both crystal structures and selected the fraction of IRs that were the closest spatial neighbors of SDPs. For this fraction the IR-SDP distances varied from 3 to 7 Å. Solvent accessibility of SDPs and IRs was calculated using SYBYL (<http://www.tripos.com>) and FANTOM (http://www.scsb.utmb.edu/fantom/fm_home.html).

RESULTS

Comparative analysis of the RF1 and RF2 multiple alignment

The MSA of the RF1/RF2 training set was subjected to phylogenetic analysis. The phylogenetic tree, as expected, split into two branches formed separately by RF1s and RF2s. Thus we excluded the possibility of RF misannotation in the training set. The RF1 (UAG specific) and RF2 (UGA specific) parts of the training set were compared, and the SDPpred program selected 15 SDPs listed in Table 1. Three of these SDPs with the highest *Z*-scores mapped to domains II (position 205) and IV (positions 319 and 320). Other SDPs were located in the 143, 144, 201–217 and 319–334 regions of RF1/RF2 sequences (according to *E. coli* RF2 position numbering) (Table 1 and Figure 1). Although all SDPs could potentially be important for some RF1- or RF2-specific function(s) it seemed unlikely that all 15 amino acid residues could participate in stop codon decoding, the major RF1/RF2-specific

Table 1. Predicted SDPs in the RF subfamilies

Position of SDPs (<i>E.coli</i> RF2 numbering)	Amino acid residues		Z-score ^a	Human mRF1 ^b
	RF1	RF2		
143	Leu or Ile	Asp	27.83	Gln
144	Phe	Trp	27.98	Trp
201	Val	Leu	30.77	Val
202	Gln	Val	28.45	Gln
205 ^c	Pro	Ser	38.32	Pro
207 ^c	Thr	Phe	27.43	Val
208	Glu	Asp or Asn	22.90	Glu
213	Ile or Val	Arg	22.74	Ile
217	Ala or Thr	Phe	27.51	Thr
319	Arg	Trp	38.56	Gln
320	Ser	Gly	40.32	Ser
322	Arg	Gln	27.00	Arg
325	Thr	Ser	25.23	Thr
327	Asn	Val	25.85	Asn
334	Thr	Lys	23.54	Thr

^aCalculated as described by Kalinina *et al.* (15). The highest-scoring SDPs are highlighted.

^bThe amino acids of human mRF1 different from RF1 SDPs are shaded.

^cIdentified earlier (12) as parts of 'protein anticodon' that determines RF1/RF2 stop codon specificity towards purines.

function. For this reason, we further considered separately the subgroup of SDPs with the highest Z-scores (Table 1). These positions formed a distinct group preceding the first local minimum on the Bernoulli estimator plot (data not shown). The second SDP group consisted of middle-scoring SDPs (positions 143, 144, 201, 202, 207, 208, 213, 217, 322, 325, 327 and 334). These SDPs were clustered in space with the highest-scoring SDPs in superdomain II/IV (Figure 2) and probably were of secondary role in specific functions of RF1/RF2 family.

The generated SDP profile was tested at RF1 and RF2 sequences not included in the training set (as mentioned in Materials and Methods). All of them were attributed to RF1 or RF2 specificity group as anticipated from their annotations. We also compared the selected SDPs of RF1 with the vertebrate mitochondrial factor mRF1 amino acid sequences. As exemplified by human mRF1 (Table 1), we found differences in SDPs 143, 207, 217 and 319. Also, the PxT/SPF 'protein anticodon' of this mRF1 contained three amino acid insertions (data not shown). These data were consistent with known biases in the vertebrate mitochondrial genetic code, where it is believed that not only RF1-specific UAG and UAA stop codons serve as termination signals but in addition also the non-canonical AGA and AGG triplets (summarized in the NCBI Taxonomy Browser <http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c>) (5).

To identify other possible functionally important positions, we applied DIVERGE method (17) for selecting the RSPs. This method is based on comparison of the mutation rates between paralogous protein subfamilies (Materials and Methods). When setting the cutoff to 0.99, the DIVERGE program predicted 22 RSPs listed below in the order of diminishing 'probability scores': 206, 211, 310, 332, 350, 361, 131, 313, 190, 102, 356, 345, 121, 171, 348, 336, 221, 117, 150, 196, 213 and 333 (*E.coli* RF2 numbering). On the 3D structure some of these positions clustered with IRs and SDPs identified by the SDPpred approach (Figure 2).

Location of SDPs in the RF1 and RF2 3D structures

Initially, experiments on the prokaryotic RF1/RF2 stop codon decoding sites were based solely on the amino acid sequences (12) without knowing the 3D structures of RFs. Now the crystal structures became known for RF2 (22) and very recently for RF1 (21). The previously described 'protein anticodon' (12) or 'tripeptide discriminator' (13) motifs are located in domain II in the RF1/RF2 crystal structures. SDPs 201, 202, 205, 207, 208, 213 and 217 are located in the same region (Figure 2). The Ser205 residue (the first position of the PxT/SPF motif) belongs to the highest-scoring SDP group whereas the Phe207 residue (the third position of PxT/SPF motif) is one of the middle-scoring SDPs (Table 1). The structure of superdomain II/IV in both RFs appeared to be very similar if not identical (21,22). The location of SDPs on these structures is shown in Figure 3.

We compared the locations of the SDPs on the *E.coli* RF2 (22) and *Thermotoga maritima* RF1 (21) crystal structures, and on the cryo-EM-based model of the ribosome-bound *E.coli* RF2 (23,24). The main difference between the crystal and the ribosome-bound structures is associated with positioning of domain III that contains the universal GGQ tripeptide (25) against the other domains of RFs. Superdomain II/IV is similar in all three structures. There are no SDPs outside this superdomain (Table 1). In space, SDPs 143 and 144 are located proximal to the SPF-containing region. Therefore, these residues were attributed to the same spatial zone as SDPs in positions 201–217 (Figure 2).

SDPs of domain IV are located in two distinct regions. SDPs 322, 325, 327 and 334 clustered with domain II SDPs 201–217 (Figure 2). None of these SDPs are high-scoring (Table 1). SDPs 319 and 320 possess the highest Z-scores (Table 1) indicating their importance for RF1/RF2 specificity. They are located in the distal part of the domain IV of the RF 3D structure (Figure 3). Comparing the spatial location of the highest-scoring SDPs we observed two distinct zones at the 20–30 Å distance from each other. Zone 1 included position 205 while zone 2 contained amino acid residues 319 and 320 (Figure 2). The middle- and lower-scoring SDPs were distributed between these two zones indicating the probable involvement of the entire superdomain II/IV in the stop codon decoding.

An interesting feature of zone 2 is the profound amino acid differences between the RF1 and RF2 SDPs. RF1s contain the 'RS' motif in contrast to the 'WG' motif of RF2s (Table 1). Obviously, arginine and serine residues are chemically very different from tryptophan and glycine residues, respectively. Surprisingly, in the RF2 subfamily, the conserved hydrophobic tryptophan residue was located on the protein surface (Figure 3C).

Location of IRs clustered with SDPs in the RF1/RF2 3D structure

We analyzed residues common to the entire RF1/RF2 family. Amino acid residues (21) are invariant for all RFs of the training set, including: four arginine residues; seven polar and charged residues (asparagine, glutamine, glutamate, histidine and threonine), two aromatic residues (tyrosine) and eight 'structural' residues (glycine and proline) (Table 2). None of the IRs are aliphatic. The selected IRs were mapped

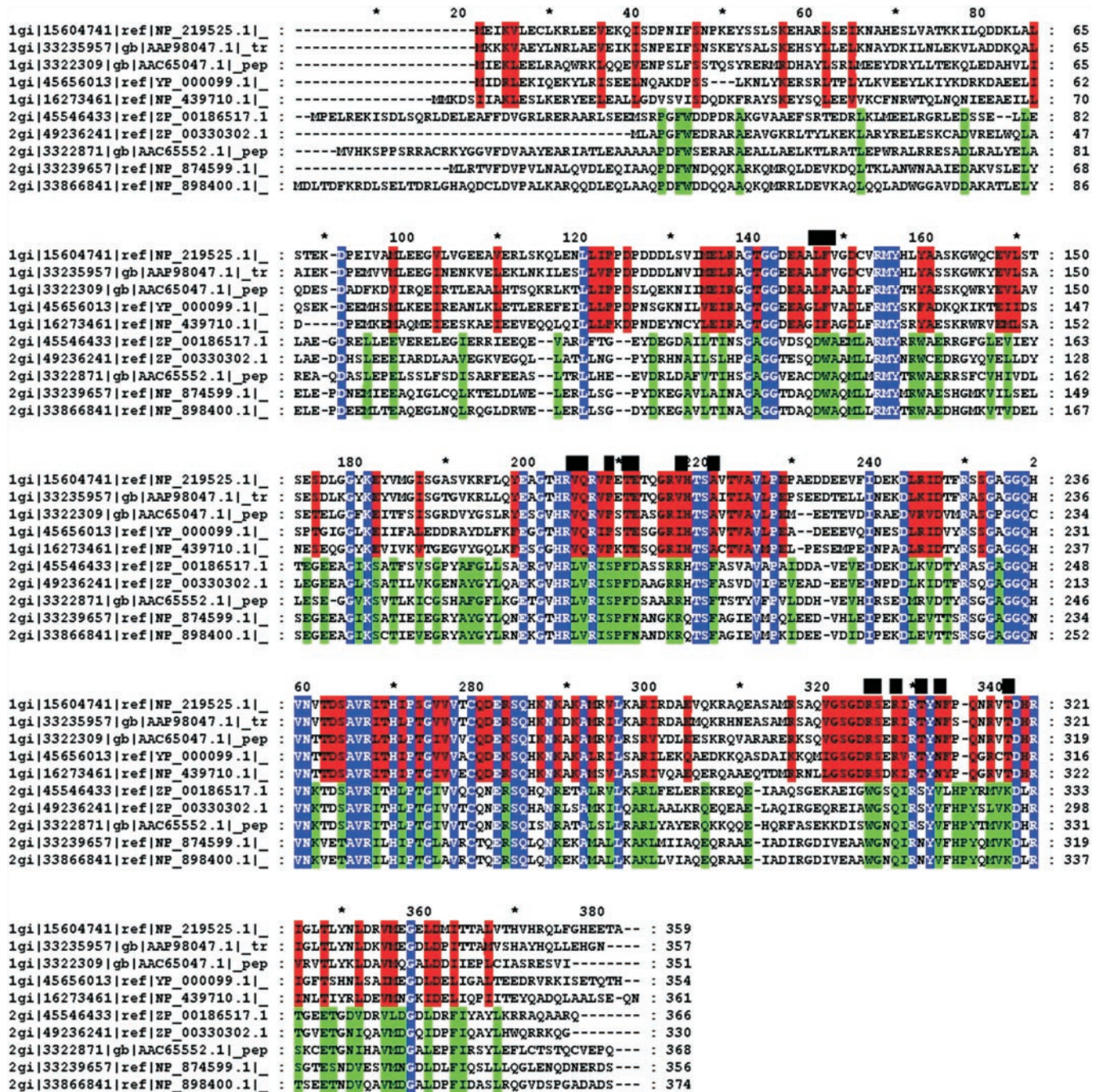


Figure 1. Multiple alignment of the RF1/RF2 family. Subset taken from MSA of RF1s and RF2s from eubacterial genomes is presented: gil15604741 (RF1, *Chlamydia trachomatis*), gil33235957 (RF1, *Chlamydomonas reinhardtii*), gil3322309 (RF1, *Treponema pallidum*), gil45656013 (RF1, *Leptospira interrogans*), gil16273461 (RF1, *Haemophilus influenzae*), gil45546433 (RF2, *Rubrobacter xylanophilus*), gil49236241 (RF2, *Moorella thermoacetica*), gil3322871 (RF2, *Treponema pallidum*), gil33239657 (RF2, *Prochlorococcus marinus*) and gil33866841 (RF2, *Synechococcus* sp.). The RF1s are shown as '1gi...' and RF2s as '2gi...'. Accession numbers and gi-identifiers are shown at the left. The alignment is highlighted according to the conservancy level of this MSA subset: RF1-conserved positions are red, RF2-conserved are green and RF1 + RF2 conserved are blue. SDPs are marked above the alignment by black boxes.

on RFs crystal structures. One IR cluster contained the functionally important GGQ loop (25–28) in domain III (Figures 2 and 3). For detecting IRs putatively involved in the stop codon recognition, we searched for IRs clustered with SDPs. Five IRs, namely Arg200, Arg203, Thr215, Arg324 and Tyr326, were located 3–7 Å from SDPs (according to centroids of

these amino acid residues) (Table 2). These residues could be involved in RNA–protein interactions via H-bond formation, electrostatic interactions and/or stacking in both RF1 (1r0x) and RF2 (1lqe) crystal structures (Figure 3). The arginine was the most frequent residue found in RNA–protein interacting sites (29,30). The presence of invariant arginines

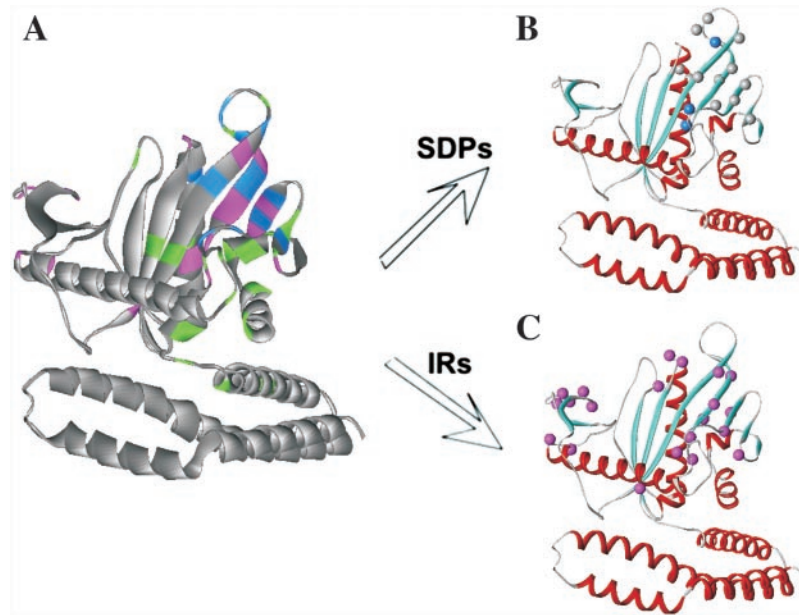


Figure 2. Location of putatively important amino acid residues in RF2 crystal structure. Ribbon models of RF2 3D structure. (A) SDPs (blue), IRs (violet) and RSPs (green). (B) The C α -atoms of the highest-scoring (blue) and middle-scoring (gray) SDPs. (C) The C α -atoms of IRs (violet).

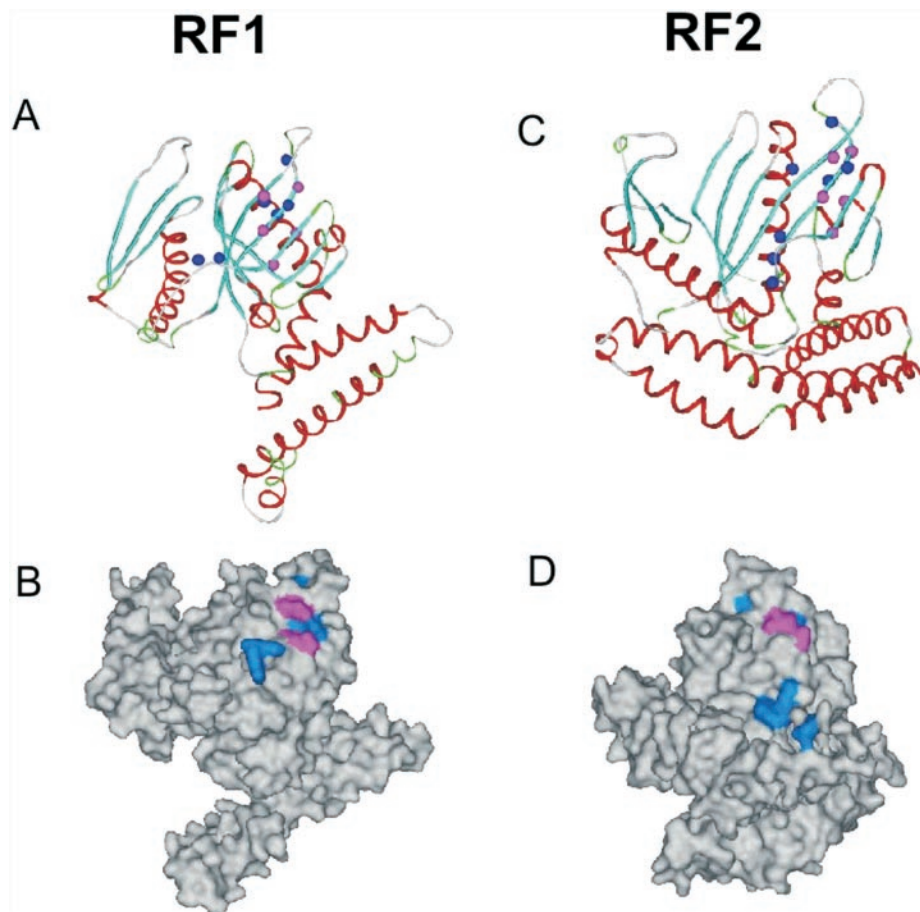


Figure 3. SDP/IR areas in the 3D structures of RF1 and RF2. Ribbon (A and C) and surface (B and D) models of RF1 (A and B) and RF2 (C and D) based on 1r0x and 1gqe crystal structures. (A and C) The C α -atoms of the highest-scoring SDPs (blue) and IRs clustered with SDPs (violet). The same color code is used for surface presentation (B and D).

Table 2. Invariant amino acids (IRs) in the RF family

Amino acid positions in <i>E.coli</i> RF2	Amino acid residues
137	Gly
138	Gly
152	Tyr
175	Gly
195	Glu
197	Gly
199	His
200	Arg
203	Arg
215	Thr
224	Pro
248	Gly
250	Gly
251	Gly
252	Gln
255	Asn
280	Gln
283	Asn
324	Arg
326	Tyr
337	Arg

IRs clustered with SDPs are highlighted.

in the vicinity of the SDPs may imply involvement of these residues in RNA–protein contacts. IRs were clustered only with SDPs from zone 1 whereas SDPs of zone 2 do not have any neighboring IRs.

DISCUSSION

Despite the growing body of data arising from genetic and biochemical analysis, crystal and ribosome-bound structures of bacterial RFs, the bioinformatics approaches have not been applied to assist in understanding of the stop codon decoding. The major problem is to identify all amino acid residues of RFs implicated in the stop codon recognition. Two tripeptides (PxT in RF1 and SPF in RF2) are involved in discrimination of the second and the third purine bases of the stop codons (12–14). It seems unlikely that high fidelity of translation termination (31) could be achieved by one-to-one interaction between a single amino acid residue and one of the stop codon nucleotides (29,30,32). Rather, more than one amino acid should be implicated in distinguishing between A and G (33). Therefore, it is reasonable to suppose that the ‘discriminator tripeptide’ (12–14) is not the only region involved in the stop codon decoding by RFs.

The current method of filtering the evolutionary signal among families of orthologous proteins by comparing paralogous protein subfamilies differing in ligand-binding specificity, SDPpred, has been developed and successfully applied to several protein subfamilies (15). This approach permits identification of RF1- and RF2-specific positions in the RF family because of their structural and functional resemblance.

Application of the SDPpred to RF protein subfamilies reveals that all SDPs (Table 1) are located in the superdomain II/IV forming an SDP-enriched area (Figure 2). We divided all SDPs into groups according to their Z-scores. Three SDPs with the highest Z-scores map to two distinct spatial zones in domains II and IV (in both RF1 and RF2 3D structures) (Figure 3). Zone 1 contains position 205 (numbering as in

E.coli RF2) identified earlier as the first residue of the ‘protein anticodon’. Position 207 from the same ‘protein anticodon’ is among middle-scoring SDPs (Table 1). As visualized by cryo-EM (23,24) this region is located in the decoding centre of the ribosome. These data are consistent with hydroxyl radical mapping of the *E.coli* RF2 SPF motif onto small ribosomal subunit (34). Although the SDPs are spread along two domains of RF1/RF2 (Table 1 and Figure 1), nevertheless, they are clustered in space (Figure 2). Properties of zone 1 are consistent with the suggestion that it forms an ‘extended’ stop codon recognition site which includes the ‘protein anticodon’ as an essential part. Probably, not all SDPs located in the vicinity of zone 1 are in direct contact with the second and/or third positions of the stop codon but we believe that all of them may contribute for high fidelity of stop codon discrimination (31).

The previously described PxT/SPF ‘tripeptide anticodon’ represents an example of differences in mutation rates for position 206 in RF1s (variable X) and RF2s (invariant P). This may imply that the selection pressure at this position is different in these subfamilies. This RS may indicate sites, in which functional constraints have changed in the course of evolution and differ in branches of the phylogenetic tree. Since the SDPpred method is not suitable to uncover motifs of this kind, we have applied the DIVERGE approach (17). In contrast to MSA-dependent cutoff of Z-scores, estimated by the SDPpred program, the DIVERGE program utilizes a user-defined threshold of ‘probability score’ for putatively important residues. We set this threshold to 0.99 and obtained a list of 22 potentially significant residues. These residues correspond to sites of the highest differences in mutation rates between RF1 and RF2 subfamilies. The best ‘probability score’ (1.00) was reached by only one amino acid residue and as anticipated, it was position 206. Positions with lower scores were spread over entire superdomain II/IV (Figure 2). Generally, SDPpred searched for positions that are conserved in each subfamily but differ between them. DIVERGE, in contrast, searches for positions that are, for example, conserved in one subfamily and variable in the other. It seems that the latter approach is somewhat less specific. Still, about a half of RSPs occur within or close to spatial zones defined by SDPs and IRs. Therefore, both methods point to the superdomain II/IV as the discriminator area in stop codon decoding.

The SDPs of zone 1 are within 3–7 Å distance from IRs Arg200, Arg203, Thr215, Arg325 and Tyr326 (Table 2 and Figure 2). Since SDPs of zone 1 are most probably responsible for the purine recognition in stop codons, these IRs may be implicated in the recognition of the first base of all stop codons, an invariant U. This hypothesis is consistent with high frequencies of arginine and threonine residues in known RNA-binding sites (29,30). Nothing has been proposed earlier in literature for prokaryotes concerning recognition of U in the stop codons. Other putative roles of these IRs could also be discussed, such as involvement in RF–16S rRNA interactions, stabilizing the ternary RF•ribosome•mRNA complex.

Zone 2 consists of SDPs with very high Z-scores, positions 319 and 320 (Table 1). These SDPs are not clustered in space with other high-scoring SDPs or IRs. Zone 2 has never been functionally identified in genetic or biochemical studies of bacterial RFs. What could be a functional role of this newly described SDP zone? Most probably, this zone along with zone 1, is implicated in the translation termination

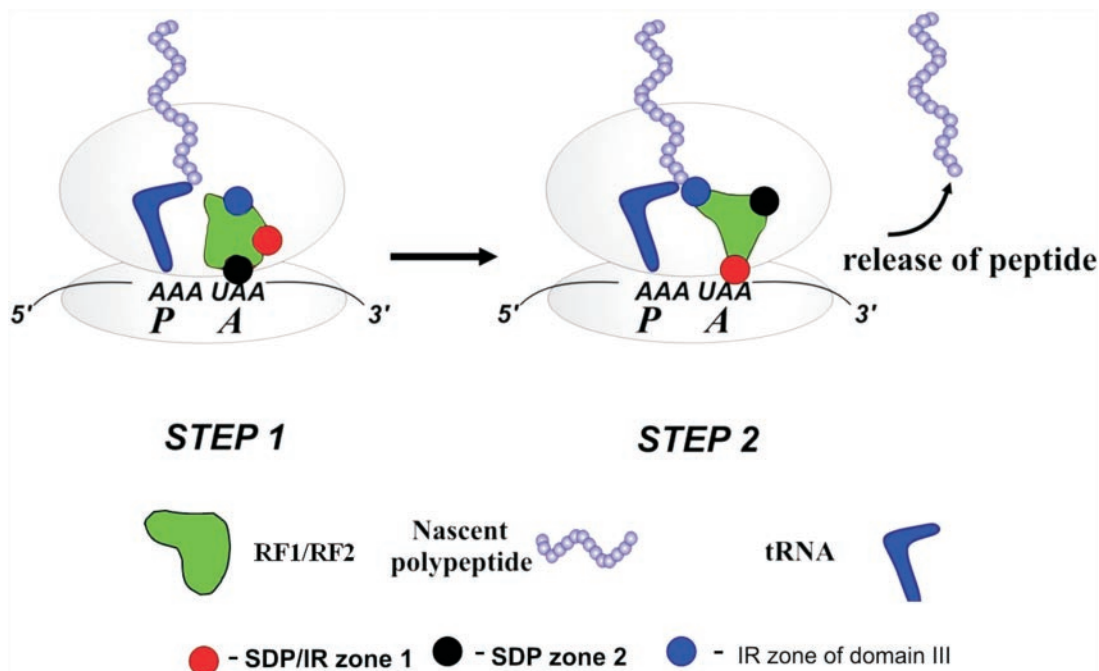


Figure 4. Hypothetical model of prokaryotic class 1 release factor stop codon decoding. Step 1: initial binding of the RF to the pretermination complex. The zone 2 amino acid residues are proposed to be essential for this step. Step 2: The conformational change of RF makes it possible for zone 1 residues to decode specifically the stop codon (UAA is given as an example). The specific binding provokes the signal transduction to the ribosomal peptidyl-transferase center and the release of the nascent polypeptide. Zone 1 (brown), zone 2 (black) and GGQ-containing IR cluster (blue) are schematically represented as circles on the RF structure.

machinery. Some data argue in favor of this assumption. Both zones are located in the same superdomain and oriented toward the decoding centre of the ribosome at the small ribosomal subunit (23,24). Most of the SDPs and IRs clustered with SDPs are located at the surface of the RF2 and RF1 crystal structures (Figure 3). The solvent accessibility of SDP/IR clusters reveals that, most probably, these amino acid residues are unshielded and accessible for interaction with stop codon and/or rRNA.

It has been proposed from genetic (35) and biochemical (36) data that RF1/2 binds to the bacterial ribosome in a two-step fashion; at the first step the binding is weak, whereas at the second, codon-specific step, it is strong. For eRF1, a two-step binding model is supported experimentally (37). If so, one may speculate that at the initial step of the RF1/2 binding to the ribosome, distal SDPs in domain IV play an essential role, probably interacting with nucleotides of rRNA or mRNA in the decoding site (Figure 4). The first step induces a conformational change in the RF1/2-ribosome complex. The only proved structural change in RF1/2 is the relocation of the domain III, shown by cryo-EM (23,24). It is not unlikely that after the zone 2 binding at the first step, some unfolding of the RF occurs. This could lead to direct specific contact of zone 1 with the stop signal in mRNA.

Experimental data available so far are consistent with the proposed functional role for two zones of the SDP/IR area. After random mutagenesis nine mutations of *Salmonella typhimurium* RF2 have been selected (35). These mutations alter the RF2 specificity towards stop codons *in vivo* and enable it to recognize the RF1-specific UAG codon. Four of nine point mutations were located in zones 1 and 2 of the SDP/IR area.

The essential role of selected SDPs in the stop codon recognition follows also from the data on mRF1s. Vertebrate mRF1s differ from both bacterial and eukaryotic RFs in their stop codon responses as mentioned in the Results (3,5,38). In parallel, SDPs in human mRF1 and bacterial RF1/RF2s differ in their patterns (Table 1). Unfortunately, at present, the set of mRF1 sequences from evolutionarily distant organisms is not sufficient for systematic analysis of this group of class 1 RFs by SDPpred with good statistical validity.

Domains I and III are implicated in functions common for both factors, e.g. interactions with RF3 and the ribosomal subparticles including the peptidyl-transferase centre. This explains why these domains contain no SDPs but exhibit several IRs.

Only five IRs are clustered with SDPs (Table 2) while others are located in different regions of RF with only one IR cluster in domain III including GGQ motif, shown to be functionally essential both for eukaryotes (25–27) and prokaryotes (28,39).

In conclusion, based on bioinformatics approach we have shown in accordance with earlier biochemical and genetic data that superdomain II/IV in RF family is implicated in specific stop codon discrimination. We excluded participation of domains I and III in this process. We predict that 'stop codon discriminator' site is 3D and composed of two distinct regions. We have predicted the involvement of invariant amino acids clustered with SDPs in recognition of the first base of all stop codons. We assume that stop codon discrimination is a two-step process which is accompanied by a conformational change in superdomain II/IV or alteration of mutual orientation of RF and the stop codon. Finally, we predict that positions 319 and 320 (numbering as in *E.coli*

RF2) in RF family are functionally essential. Obviously, our predictions have to be tested in future experiments.

ACKNOWLEDGEMENTS

This work was partially supported by the Presidium of the Russian Academy of Sciences (Programs 'Molecular and Cell Biology' and 'Origin and Evolution of the Biosphere'), by the Presidential program of supporting the leading scientific schools (via Ministry of Education and Science of the Russian Federation), the Russian Foundation for Basic Research (03-04-48943 and 05-04-49385), the Howard Hughes Medical Institute (55000309) and the Ludwig Institute for Cancer Research (CRDF RB0-1268). O.V.K. was supported by INTAS (04-83-3704). M.S.G. and N.J.O. were supported by the Russian Science Support Fund. Funding to pay the Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

- Kisselev, L., Ehrenberg, M. and Frolova, L. (2003) Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.*, **22**, 175–182.
- Nakamura, Y. and Ito, K. (2003) Making sense of mimic in translation termination. *Trends Biochem. Sci.*, **28**, 99–105.
- Poole, E.S., Askarian-Amiri, M.E., Major, L.L., McCaughan, K.K., Scarlett, D.J., Wilson, D.N. and Tate, W.P. (2003) Molecular mimicry in the decoding of translational stop signals. *Prog. Nucleic Acid Res. Mol. Biol.*, **74**, 83–121.
- Inge-Vechtomov, S., Zhouravleva, G. and Philippe, M. (2003) Eukaryotic release factors (eRFs) history. *Biol. Cell*, **95**, 195–209.
- Santos, M.A., Moura, G., Massey, S.E. and Tuite, M.F. (2004) Driving change: the evolution of alternative genetic codes. *Trends Genet.*, **20**, 95–102.
- Frolova, L., Le Goff, X., Rasmussen, H.H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.L., Celis, J.E., Philippe, M. *et al.* (1994) A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature*, **372**, 701–703.
- Kervestin, S., Frolova, L., Kisselev, L. and Jean-Jean, O. (2001) Stop codon recognition in ciliates: Euplates release factor does not respond to reassigned UGA codon. *EMBO Rep.*, **2**, 680–684.
- Ito, K., Frolova, L., Seit-Nebi, A., Karamyshev, A., Kisselev, L. and Nakamura, Y. (2002) Omnipotent decoding potential resides in eukaryotic translation termination factor eRF1 of variant-code organisms and is modulated by the interactions of amino acid sequences within domain 1. *Proc. Natl Acad. Sci. USA*, **99**, 8494–8499.
- Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G. and Stansfield, I. (2000) Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA*, **6**, 1236–1247.
- Frolova, L., Seit-Nebi, A. and Kisselev, L. (2002) Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA*, **8**, 129–136.
- Seit-Nebi, A., Frolova, L. and Kisselev, L. (2002) Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep.*, **3**, 881–886.
- Ito, K., Uno, M. and Nakamura, Y. (2000) A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature*, **403**, 680–684.
- Nakamura, Y. and Ito, K. (2002) A tripeptide discriminator for stop codon recognition. *FEBS Lett.*, **514**, 30–33.
- Nakamura, Y., Ito, K. and Ehrenberg, M. (2000) Mimicry grasps reality in translation termination. *Cell*, **101**, 349–352.
- Kalinina, O.V., Mironov, A.A., Gelfand, M.S. and Rakhmaninova, A.B. (2004) Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Sci.*, **13**, 443–456.
- Kalinina, O.V., Novichkov, P.S., Mironov, A.A., Gelfand, M.S. and Rakhmaninova, A.B. (2004) SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. *Nucleic Acids Res.*, **32**, W424–W428.
- Gu, X. and Vander Velden, K. (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics*, **18**, 500–501.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
- Felsenstein, J. (1996) Inferring phylogenies from protein sequence by parsimony, distance and likelihood methods. *Methods Enzymol.*, **266**, 418–427.
- Shin, D.H., Brandsen, J., Jancarik, J., Yokota, H., Kim, R. and Sung-Hou Kim, S.-H. (2004) Structural analyses of peptide release factor 1 from *Thermotoga maritima* reveal domain flexibility required for its interaction with the ribosome. *J. Mol. Biol.*, **341**, 227–239.
- Vestergaard, B., Van, L.B., Andersen, G.R., Nyborg, J., Buckingham, R.H. and Kjeldgaard, M. (2001) Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell*, **8**, 1375–1382.
- Rawat, U.B., Zavialov, A.V., Sengupta, J., Valle, M., Grassucci, R.A., Linde, J., Vestergaard, B., Ehrenberg, M. and Frank, J. (2003) A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature*, **421**, 87–90.
- Klaholz, B.P., Pape, T., Zavialov, A.V., Myasnikov, A.G., Orlova, E.V., Vestergaard, B., Ehrenberg, M. and van Heel, M. (2003) Structure of the *Escherichia coli* ribosomal termination complex with release factor 2. *Nature*, **421**, 90–94.
- Frolova, L.Y., Tsvikovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I. and Kisselev, L.L. (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*, **5**, 1014–1020.
- Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A. and Barford, D. (2000) The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell*, **100**, 311–321.
- Seit-Nebi, A., Frolova, L., Justesen, J. and Kisselev, L. (2001) Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucleic Acids Res.*, **29**, 3982–3987.
- Mora, L., Heurgue-Hamard, V., Champ, S., Ehrenberg, M., Kisselev, L.L. and Buckingham, R.H. (2003) The essential role of the invariant GGQ motif in the function and stability *in vivo* of bacterial release factors RF1 and RF2. *Mol. Microbiol.*, **47**, 267–275.
- Draper, D.E. (1999) Themes in RNA–protein recognition. *J. Mol. Biol.*, **293**, 255–270.
- Jones, S., Daley, D.T., Luscombe, N.M., Berman, H.M. and Thornton, J.M. (2001) Protein–RNA interactions: a structural analysis. *Nucleic Acids Res.*, **29**, 943–954.
- Freistoffer, D.V., Kwiatkowski, M., Buckingham, R.H. and Ehrenberg, M. (2000) The accuracy of codon recognition by polypeptide release factors. *Proc. Natl Acad. Sci. USA*, **97**, 2046–2051.
- Nobeli, J., Laskowski, R.A., Valdar, W.S. and Thornton, J.M. (2001) On the molecular discrimination between adenine and guanine by proteins. *Nucleic Acids Res.*, **29**, 4294–4309.
- Basu, G., Sivanesan, D., Kawabata, T. and Go, N. (2004) Electrostatic potential of nucleotide-free protein is sufficient for discrimination between adenine and guanine-specific binding sites. *J. Mol. Biol.*, **342**, 1053–1066.
- Scarlett, D.J., McCaughan, K.K., Wilson, D.J. and Tate, W.P. (2003) Mapping functionally important motifs SPF and GGQ of the decoding release factor RF2 to the *Escherichia coli* ribosome by hydroxyl radical footprinting. Implications for macromolecular mimicry and structural changes in RF2. *J. Biol. Chem.*, **278**, 15095–15104.
- Yoshimura, K., Ito, K. and Nakamura, Y. (1999) Amber (UAG) suppressors affected in UGA/UAA-specific polypeptide release factor 2 of bacteria: genetic prediction of initial binding to ribosome preceding stop codon recognition. *Genes Cells*, **4**, 253–266.

36. Wilson,D.N., Guevremont,D. and Tate,W.P. (2000) The ribosomal binding and peptidyl-tRNA hydrolysis functions of *Escherichia coli* release factor 2 are linked through residue 246. *RNA*, **6**, 1704–1713.
37. Chavatte,L., Frolova,L., Laugaa,P., Kisselev,L. and Favre,A. (2003) Stop codons and UGG promote efficient binding of the polypeptide release factor eRF1 to the ribosomal A site. *J. Mol. Biol.*, **331**, 745–758.
38. Askarian-Amiri,M.E., Pel,H.J., Cuevremont,D., McCaughan,K.K., Poole,E.S., Sumpter,V.G. and Tate,W.P. (2000) Functional characterization of yeast mitochondrial release factor 1. *J. Biol. Chem.*, **275**, 17241–17248.
39. Zavialov,A.V., Mora,L., Buckingham,R.H. and Ehrenberg,M. (2002) Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol. Cell*, **10**, 789–798.