Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 475–478, January 1997
Biochemistry

Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: A critical enzyme in the evolution of DNA genomes?

(adenosylcobalamin/RNA world/hyperthermophile/Archaea/intein)

JOAN RIERA*, FRANK T. ROBB†, ROBERT WEISS‡, AND MARC FONTECAVE*§

*Laboratoire d'Etudes Dynamiques et Structurales de la Sélectivité, Unité Mixte de Recherches Centre National de la Recherche Scientifique 5616, Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France; †Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, MD 21202; and ‡Department of Genetics, University of Utah, Salt Lake City, UT 84112

Communicated by Peter Reichard, Karolinska Institute, Stockholm, Sweden, November 7, 1996 (received for review October 3, 1996)

ABSTRACT **Ribonucleotide reductase (RNR), the enzyme responsible for deoxyribonucleotide synthesis, has been isolated from *Pyrococcus furiosus*, a deeply branching hyperthermophilic, strictly anaerobic archaeon. Its gene has been cloned, sequenced, and shown to harbor two insertions encoding inteins. The purified enzyme absolutely requires adenosylcobalamin for activity, a trait that defines it as a member of class II (adenosylcobalamin-dependent) prokaryotic RNRs. On the other hand, the archaeal RNR has significant amino acid sequence homology with class I (aerobic non-heme iron-dependent) and class III (anaerobic iron–sulfur-dependent) RNRs present in eukaryotes and bacteria, respectively. It is proposed that this enzyme may be the closest possible relative of the original RNR, which allowed the key "RNA world" to "DNA world" transition, and that the different classes of present-day RNRs are the products of divergent evolution.**

It is believed that life was first based on RNA. The conversion from RNA to DNA genomes required the emergence of a biosynthetic pathway supplying deoxyribonucleotides, the DNA precursors. In all present-day organisms studied so far (bacteria and eukaryotes), deoxyribonucleotides are produced exclusively by reduction of the corresponding ribonucleotides, a reaction catalyzed by the radical enzyme ribonucleotide reductase (RNR; ref. 1). It is thus conceived that the evolution of ribonucleotide reduction was the key event that allowed the "RNA world" to "DNA world" conversion. However, the great diversity of RNRs, in contemporary metabolism, contradicted the concept of a divergent evolution of this class of enzymes. Three classes of RNRs (class I in certain aerobic bacteria and in all higher eukaryotes; class II in bacteria, both aerobes and anaerobes; and class III in anaerobically growing facultative anaerobes) with no distinct amino acid sequence similarity between them and with very different cofactors have been described to date (1). We now report the characterization of an archaeal RNR. The enzyme, from *Pyrococcus furiosus*, a deeply branching hyperthermophilic, strictly anaerobic archaeon that grows optimally at 100°C, has both eukaryal and bacterial features and homology with all extant classes of RNR. Our data support the hypothesis that the three classes of present-day RNRs derive from a common ancestor, similar to the archaeal RNR, which may have allowed the key RNA to DNA transition in anaerobic, and probably hot, environments.

MATERIALS AND METHODS

Purification of the Archaeal Reductase. All manipulations were carried out aerobically at 4°C, except for the dATP-Sepharose chromatography step, which was run at room temperature. The starting material was anaerobically grown *P. furiosus* cells, kindly provided by P. Forterre (University Paris XI, Orsay, France). In a typical experiment, 13 g of tightly packed cells was sonicated in 20 ml of 50 mM Tris·HCl, pH 8/4 mM EDTA/5 mM DTT (buffer A) in the presence of 1 mM phenylmethylsulfonylfluoride and centrifuged at 180,000 × *g* for 90 min. Soluble extracts (1.02 g of protein) were first treated with 3% streptomycin sulfate and centrifuged at 40,000 × *g* for 15 min, and then most of the activity was precipitated with 50% saturated ammonium sulfate. After overnight dialysis against buffer A at 4°C, the active solution, containing 0.38 g protein, was loaded onto a 25-ml phenyl Sepharose (Pharmacia) column and equilibrated with buffer A containing 0.5 M KCl. The column was eluted with buffer A plus 0.5 M KCl (0.1 ml/min), then 0 M KCl (0.5 ml/min). When the absorption at 280 nm was below 0.05, the activity was eluted with water (1 ml/min). A significant amount of enzyme was retained on the column and was eluted with 30% ethylene glycol. The two last fractions were combined and exchanged with buffer A.

The solution (58 mg) was then loaded onto a 2-ml dATP-Sepharose column, prepared according to ref. 2 and equilibrated with buffer A plus 10% glycerol (buffer B) containing 0.5 M KCl. The column was washed with the equilibration buffer (0.2 ml/min). The reductase was then eluted with 2.5 mM ATP in buffer B. After overnight dialysis against buffer B and concentration by centrifugation in Centricon 30 microconcentrators (Amicon), aliquots of the enzyme solution were stored in liquid nitrogen.

Assay for Enzyme Activity. In a standard assay, the protein sample was incubated at 80°C for 30 min in the presence of 5 μM adenosylcobalamin (AdoCbl), 0.5 M sodium acetate, 40 mM DTT, and 1 mM [^{3}H]CDP in 50 μl of 50 mM Tris·HCl (pH 7.5). The reaction was started by addition of the enzyme. The amount of dCDP formed was determined as previously described (3). One unit of activity is defined as 1 nmol of dCDP formed per min. Specific activity is given in units per mg of protein (see Table 1).

Preparation of the DNA Library. The *P. furiosus* (DSM 3638) strain was a gift from Michael Adams at the University of Georgia (Athens, GA). The strain was grown at 95°C as previously described (4).

The library of *P. furiosus* genomic DNA consisted of a mixture of partially digested DNA as described below. Freshly isolated genomic DNA (4 μg) was digested with each of the four restriction enzymes *Dra*I, *Eco*RV, *Hinc*II, and *Pvu*II. Fragments, in the 6–8 kbp size range, were then pooled and methylated at *Eco*RI sites by incubation at 37°C for 2 hr with 600 units of *Eco*RI methylase (New England Biolabs) in 1×

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA
0027-8424/97/94475-4$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: RNR, ribonucleotide reductase; AdoCbl, adenosylcobalamin.

Data deposition: The sequence reported in this paper has been deposited in the dbEST data base (accession no. U78098).

§To whom reprint requests should be addressed. e-mail: Marc.Fontecave@ujf-grenoble.fr.

methylase buffer, containing 100 μg/ml BSA and 80 μM S-adenosylmethionine in a 230-μl reaction volume. The reaction was stopped by incubation at 65°C for 20 min. Phosphorylated 8-mer *Eco*RI linkers (15 μg; Amersham) were ligated to the pooled digested DNA (16 μg) by preincubation at 37°C for 10 min in 1× ligase buffer, followed by addition of 180 Weiss units of T4 DNA ligase (Amersham) and incubation at 16°C for 18 hr. The ligated genomic DNA–linker mixture was digested with *Eco*RI (900 units) at 37°C for 2 hr to create *Eco*RI ends on the insert DNA. Excess linkers were removed by filtration through Centricon 100 spin columns (Amicon). Insert DNA was ligated with Lambda Zap II *Eco*RI-digested vector (Stratagene). All libraries were packaged using Gigapack II packaging extracts (Stratagene). The amplified library was converted to pBluescript II plasmids using R408 helper phage (Stratagene), transfected into *Escherichia coli* XL1 blue cells (Stratagene), and plated on Luria broth (LB) agar containing ampicillin (50 μg/ml), 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal), and isopropyl-1-thio-D-galactoside (IPTG).

DNA Sequencing. Plasmid DNA (3–5 μg) prepared using the Magic Minipreps kit (Promega) was sequenced by the dideoxy nucleotide chain termination method using the Sequenase version 2.0 kit (United States Biochemical) with [α-^{35}S]dATP (Amersham). In later studies, the Vistra (Amersham) automated template preparation system was used to prepare template DNA. Single run sequences were obtained using T3 and T7 promoter primers (Promega). Continuous DNA sequences were prepared using a multiplex sequencing regime (5). The DNA sequences have been submitted to the dbEST data base [National Center for Biotechnology Information (NCBI), National Institutes of Health].

Data Base Comparison. The sequence data bases Swiss-Prot, GenBank, and the nr (non-redundant) data base at NCBI were searched using BLASTX and BLASTN (6) with the BLOSUM62 substitution matrix and default parameters.

RESULTS

Purification and Characterization of the Archaeal RNR. The archaeal RNR was first detected in soluble cell-free extracts of *P. furiosus*, which catalyzed CDP reduction in the presence of DTT. Activity could be detected with the addition of AdoCbl to the reaction mixture, strongly suggesting the presence of a class II enzyme (1). Furthermore, the enzyme required extremely high temperatures, as expected for a hyperthermophilic organism. Both EDTA and acetate stimulated the activity and were included in the assay. Soluble extracts and subsequent purified fractions could be manipulated aerobically at room temperature, also reflecting the extreme stability of the enzyme. As shown in in Table 1, the activity was purified 3000-fold in three steps (ammonium sulfate precipitation, phenyl Sepharose, and dATP-Sepharose chromatography), and a pure monomeric enzyme with a molecular mass of ≈90 kDa was isolated. The key purification step was affinity chromatography on dATP-Sepharose (400-fold purification), a method used successfully for the purification of several other RNRs (7, 8). Binding to this ligand also indicated that the archaeal RNR is allosterically regulated by nucleoside triphosphates, a typical characteristic of all of the RNRs studied to date.

The main characteristics of the enzyme are the following: (*i*) enzyme substrates are ribonucleoside diphosphates (K_m for CDP was 0.07 mM) and not triphosphates; (*ii*) the reaction has an optimal temperature of 80°C; (*iii*) as for class II enzymes, the archaeal enzyme consists of only one oxygen-insensitive polypeptide chain, as shown by gel filtration of the pure protein on Ultrogel AcA34 (Sepracor, IBF, France), and the catalyzed reaction requires AdoCbl (K_m = 1 μM) as a radical generator and DTT (K_m = 22 mM) as a reducing agent; (*iv*) however, the N-terminal region, determined by Edman degradation of the pure protein, shows convincing identity with those of the large subunits of class I and class III enzymes, but not with that of class II enzymes (Fig. 1*a*); and (*v*) furthermore, unlike class II RNRs and more like eukaryotic RNRs, the archaeal enzyme is inhibited most strongly by dATP, an allosteric effector (data not shown).

Table 1. Purification of the RNR from *P. furiosus*

Step	Protein, mg	Specific activity	Total units	Purification factor
Extract	1020	0.17	173	1
Ammonium sulfate	380	0.35	133	2
Phenyl Sepharose	58	1.3	77	7.6
dATP-Sepharose	0.1	510	51	3000

One unit of activity is defined as 1 nmol of dCDP formed per min. Specific activity is given in units per mg of protein.

Sequence Analysis of the Archaeal RNR. The RNR gene was discovered during a genomic survey of the genome of *P. furiosus* (4). The genomic DNA library described in *Materials and Methods* was used as a source of 1450 templates for random sequencing, using both manual and automated approaches. A total of 301 kb of raw sequence data was analyzed by the BLASTX algorithm (6), and several significant hits to the RNR gene were obtained (4). The sequence from these hits was then used to search the data base of the genomic sequence project of the hyperthermophic organism *Pyrococcus furiosus* (R.W.), which is in progress, using multiplex sequencing methods (5). Surprisingly, the open reading frame, commencing with the 5′ end corresponding exactly to the purified protein as shown from the identity of the first 15 aa of the N terminus, was 4.5 kb long. The deduced protein sequence contained 1740 aa with a molecular mass of 200 kDa, whereas the native protein had a molecular mass of 90 kDa. This discrepancy arises from the presence of two novel inteins (containing 454 and 382 residues, respectively) within the 200-kDa protein, which presumably are spliced out to generate the 90-kDa mature enzyme (Fig. 1*b*). The splicing sites, in particular the conserved short sequence His-Asn-Cys/Thr found at the intein-C–extein border, and the characteristic LAGLI-DADG motif, are clearly identified by comparison with previously described inteins (9). Two inteins have been discovered within the *Thermococcus litoralis* DNA polymerase (10), and a single intein has been identified in the *P. furiosus* type III topoisomerase homolog (J. DiRuggiero, K. M. Borges, F.T.R., and A. Bogert, unpublished work). There is apparently a significant predisposition for inteins to insert within genes encoding DNA and nucleotide modifying enzymes of Archaea. This is a weak trend in other organisms (see ref. 11 for review).

While no significant sequence homology could be found between the three classes of RNRs, except for very short sequence fragments (12, 13), the deduced archaeal protein sequence aligned with that of the large subunit of eukaryal RNR (for example, 28% identity and 48% similarity with murine RNR; Fig. 1*b*), but also, albeit at a slightly lower significance level, with that of the large subunits of class I (22% identity and 47% similarity with *E. coli*) and class III (23% identity and 46% similarity with *E. coli*) bacterial RNRs and with that of class II (22% identity and 48% similarity with *Lactobacillus leichmannii*) bacterial enzymes. In class I and class II RNRs, two essential cysteines are directly involved in nucleotide reduction (13, 14). A third cysteine is proposed to be the transient protein radical required for substrate activation (15, 16). It is interesting to note that the archaeal RNR contains these cysteines correctly placed within regions with strong similarity to both class I and class II RNRs, suggesting similar enzyme mechanisms (Fig. 2).

DISCUSSION

The absence of sequence similarities between the three classes of RNRs has previously led to uncertainty about whether RNRs have evolved independently on more than one occasion and converged from different precursors guided by the necessities of the chemistry used by these enzymes (17), or, instead, all three classes evolved from a common ancestor—i.e., divergent evolution. The second hypothesis is favored by Reichard

FIG. 1. (*a*) Comparison of the N-terminal sequences of *P. furiosus* RNR, the large subunit of class I (mouse) RNR, and the large subunit of class III (anaerobic *E. coli*) RNR. (*b*) Alignment of the predicted amino acid sequences of *P. furiosus* (upper sequence) and mouse (large subunit) RNRs using the BLASTX program. The two inteins of the archaeal RNR are double-underlined. The conserved LAGLI-DADG motif characteristic of protein spacer regions is indicated with the symbol * (7). Critical amino acids at the intein–extein junctions are indicated with the symbol #. The entire RNR sequence was determined by dideoxy sequencing of a 9-kb plasmid subclone from a *P. furiosus* genomic library (R.W., D. M. Dunn, J. L. Cherry, M. D. Stump, P. Cartwright, *et al.* unpublished work).

(1), essentially on the basis of strong analogies between the three classes, as far as their allosteric regulation is concerned. As mentioned previously, the three classes of RNR enzymes are all subject to feedback control by nucleoside triphosphates, and they respond in a similar fashion to different effector nucleotides (1, 18).

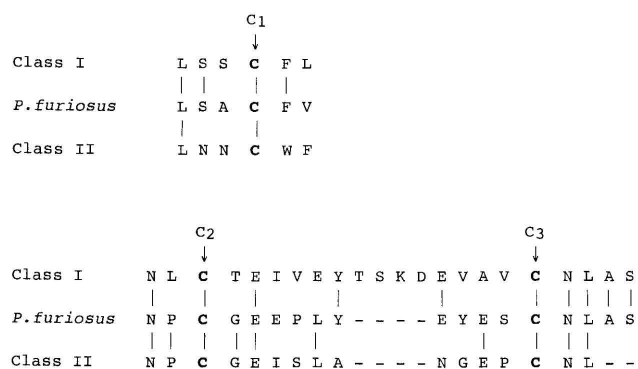

FIG. 2. The essential cysteines in the large subunit of class I (mouse) RNR and in class II (*L. leichmannii*) RNR are present in *P. furiosus* RNR. C_1 and C_3 designate the cysteines that directly reduce the ribonucleotide. C_2 is the cysteinyl radical (13).

The newly discovered RNR isolated from a hyperthermophilic archaebacterium, *P. furiosus*, now provides a convincing link between the different classes and strongly supports the notion of a divergent evolution of the RNRs. A recent study of the AdoCbl-dependent enzyme from a thermophilic archaebacterium, *Thermoplasma acidophila*, led to a similar conclusion (19). We propose that the three classes differ in the ways that they address the requirement for an essential protein radical: in all class I RNRs, a small subunit containing a non-heme iron center generates an essential tyrosyl radical in the presence of oxygen (20, 21); in class II, AdoCbl serves as a source of deoxyadenosyl and, subsequently, thiyl radicals (15); and in class III RNRs, a small subunit containing an iron–sulfur center generates a glycyl radical in the presence of *S*-adenosylmethionine (22–24). We now suggest that homologous proteins may have diverged from a common ancestor similar to the *P. furiosus* RNR to accommodate either a non-heme iron protein, an iron–sulfur protein, or an AdoCbl cofactor to control the radical chemistry. On the other hand, it appears that, evolutionarily, redox active cysteines and enzyme mechanisms have been conserved.

Concerning the nature of the ancestral RNR, we agree with the general opinion that class I enzymes are the latest versions of RNR, an enzyme that evolved when oxygen accumulated at the surface of the earth (1). Iron and oxygen provided a new and efficient mechanism for generating protein radicals.

In spite of the complexity of its protein structure, the oxygen-sensitive class III enzyme has certain features we would expect of a putative primitive RNR. First, it uses an iron–sulfur center and *S*-adenosylmethionine for generating a radical. Iron–sulfur clusters are modern versions of pyrite, a form of iron widely present at the surface of the early earth. Furthermore, *S*-adenosylmethionine is an RNA cofactor and may be considered as a vestige of the RNA world, where, as a 5′ deoxyadenosyl radical source, it played the role of a much simpler AdoCbl-like cofactor (25). Second, class III enzymes do not require a complex electron transfer chain for the reduction of ribonucleotides, as do class I and class II enzymes, but instead they use formate, one of the simplest organic reductants and, in all probability, an important product of prebiotic chemistry (26). That, as in class I and class II enzymes, redox-active cysteines are present for mediating electron and radical transfers is a likely possibility (27); however, it remains to be shown. Third, the methane-producing archaeon, *Methanococcus jannaschii*, whose genome has been recently sequenced, contains a gene which can be assigned to a class III RNR, on the basis of sequence similarities (28). However, the RNR from this organism remains to be characterized at the protein level.

The AdoCbl enzymes have the versatility that we expect in the ancestral enzyme, since they function both anaerobically and aerobically. They consist of a single protein which, by controlling the cobalt–carbon bond homolysis of the cofactor, provides the simplest mechanism to generate free radicals (29). They are widely present in bacteria (30). Now we and others (19) show that this enzyme is also present in thermophilic and hyperthermophilic, anaerobic archaea. The detection of a generalized RNR in *P. furiosus*, one of the deepest evolutionary branches of the Archaea, is significant. It strongly suggests that the original RNR, thermophilic or not, might have been a class II enzyme, which may have diverged quite early in evolution to the present class I and class III enzymes.

We thank E. Mulliez (Université J. Fourier, Grenoble, France) and P. Forterre (Université Paris XI) for discussions and assistance in some experiments and J. Gagnon (Institute de Biologie Structurale, Grenoble) for performing the N-terminal amino acid analysis. This work was supported by a grant from Spanish Ministerio de Educacion y Ciencia to J.R. by an Advanced Technology Program grant from the U.S. Department of Commerce to F.T.R., and by a grant from the Microbial Genomes Program, the U.S. Department of Energy, to R.W.

1. Reichard, P. (1993) *Science* **260,** 1773–1777.
2. Berglund, O. & Eckstein, F. (1972) *Eur. J. Biochem.* **28,** 492–496.
3. Thelander, L., Sjöberg, B.-M. & Eriksson, S. (1978) *Method Enzymol.* **51,** 227–237.
4. Borges, K. M., Brummet, S. R., Bogert, A. P., Davis, M. C., Hujer, K. M., Domke, S. T. Szasz, J., Ravel, J., DiRuggiero, J., Fuller, C., Chase, J. W. & Robb, F. T. (1996) *Genome Sci. Technol.* **1,** 37–46.
5. Cherry, J. L., Young, H., DiSera, L. J., Ferguson, F. M., Kimball, A. W., Dunn, D. M., Gesteland, R. F. & Weiss, R. B. (1994) *Genomics* **20,** 68–74.
6. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
7. Eliasson, R., Pontis, E., Fontecave, M., Gerez, C., Harder, J., Jörnvall, H., Krook, M. & Reichard, P. (1992) *J. Biol. Chem.* **267,** 25541–25547.
8. Engström, Y., Eriksson, S., Thelander, L. & Akerman, M. (1979) *Biochemistry* **18,** 2941–2948.
9. Fsihi, H., Vincent, V. & Cole, S. T. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3410–3415.
10. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. (1992) *Nucleic Acids Res.* **22,** 1125–1127.
11. Cooper, A. A. & Stevens, T. H. (1995) *Trends Biochem. Sci.* **20,** 351–356.
12. Booker, S. & Stubbe, J. (1993) *Proc. Natl. Acad. Sci, USA* **90,** 8352–8356.
13. Mao, S. S., Holler, T. P., Yu, G. X., Bollinger, J. M., Booker, S., Johnston, M. I. & Stubbe, J. (1992) *Biochemistry* **31,** 9733–9743.
14. Booker, S., Licht, S., Broderick, J. & Stubbe, J. (1994) *Biochemistry* **33,** 12676–12685.
15. Licht, S., Gerfen, G. J. & Stubbe, J. (1996) *Science* **271,** 477–481.
16. Mao, S. S., Yu, G. X., Chalfoun, D. & Stubbe, J. (1992) *Biochemistry* **31,** 9752–9759.
17. Doolittle, R. F. (1994) *Trends Biochem. Sci.* **19,** 15–18.
18. Eliasson, R., Pontis, E., Sun, X. & Reichard, P. (1994) *J. Biol. Chem.* **269,** 26052–26057.
19. Tauer, A. & Benner, S. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 53–58.
20. Ochiai, E.-I., Mann, G. J., Gräslund, A. & Thelander, L. (1990) *J. Biol. Chem.* **265,** 15758–15761.
21. Bollinger, J. M., Edmondson, D. E., Huynh, B. H., Filley, J., Norton, J. R. & Stubbe, J. (1991) *Science* **253,** 292–298.
22. Ollagnier, S., Mulliez, E., Gaillard, J., Eliasson, R., Fontecave, M. & Reichard, P. (1996) *J. Biol. Chem.* **271,** 9410–9416.
23. Sun, X., Ollagnier, S., Schmidt, P. P., Atta, M., Mulliez, E., Lepape, L., Eliasson, R., Gräslund, A., Fontecave, M., Reichard, P. & Sjöberg, B.-M. (1996) *J. Biol. Chem.* **271,** 6827–6831.
24. Sun, X., Harder, J., Krook, M., Jörnvall, H., Sjöberg, B.-M. & Reichard, P. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 577–581.
25. Frey, P. A. (1993) *FASEB J.* **7,** 662–670.
26. Mulliez, E., Ollagnier, S., Fontecave, M., Eliasson, R. & Reichard, P. (1995) *Proc. Natl. Acad. Sci.* **92,** 8759–8762.
27. Eliasson, R., Reichard, P., Mulliez, M., Ollagnier, S., Fontecave, M., Liepinsh, E. & Otting, G. (1995) *Biochem. Biophys. Res. Commun.* **214,** 28–35.
28. Bult, C. J., White, O., Olsen, B. J., Zhou, L., Fleishmann, R. D., *et al.* (1996) *Science* **273,** 1058–1073.
29. Halpern, J. (1985) *Science* **227,** 869–874.
30. Harder, J. (1993) *FEMS Microbiol. Lett.* **12,** 273–292.