The B₁₂-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila*: An evolutionary solution to the ribonucleotide reductase conundrum

(post-genomic science/evolution)

ANDREAS TAUER* AND STEVEN A. BENNER*†‡

*Department of Chemistry, Eidgenössische Technische Hochschule Zürich, Universitätstrasse 16, CH-8092 Zürich, Switzerland; and †Department of Chemistry, University of Florida, Gainesville, FL 32611

ABSTRACT   A coenzyme B₁₂-dependent ribonucleotide reductase was purified from the archaebacterium *Thermoplasma acidophila* and partially sequenced. Using probes derived from the sequence, the corresponding gene was cloned, completely sequenced, and expressed in *Escherichia coli*. The deduced amino acid sequence shows that the catalytic domain of the B₁₂-dependent enzyme from *T. acidophila*, some 400 amino acids, is related by common ancestry to the diferric tyrosine radical iron(III)-dependent ribonucleotide reductase from *E. coli*, yeast, mammalian viruses, and man. The critical cysteine residues in the catalytic domain that participate in the thiyl radical-dependent reaction have been conserved even though the cofactor that generates the radical is not. Evolutionary bridges created by the *T. acidophila* sequence and that of a B₁₂-dependent reductase from *Mycobacterium tuberculosis* establish homology between the Fe-dependent enzymes and the catalytic domain of the *Lactobacillus leichmannii* B₁₂-dependent enzyme as well. These bridges are confirmed by a predicted secondary structure for the *Lactobacillus* enzyme. Sequence similarities show that the N-terminal domain of the *T. acidophila* ribonucleotide reductase is also homologous to the anaerobic ribonucleotide reductase from *E. coli*, which uses neither B₁₂ nor Fe cofactors. A predicted secondary structure of the N-terminal domain suggests that it is predominantly helical, as is the domain in the aerobic *E. coli* enzyme depending on Fe, extending the homologous family of proteins to include anaerobic ribonucleotide reductases, B₁₂ ribonucleotide reductases, and Fe-dependent aerobic ribonucleotide reductases. A model for the evolution of the ribonucleotide reductase family is presented; in this model, the thiyl radical-based reaction mechanism is conserved, but the cofactor is chosen to best adapt the host organism to its environment. This analysis illustrates how secondary structure predictions can assist evolutionary analyses, each important in "post-genomic" biochemistry.

The deoxyribonucleotides required for DNA replication are biosynthesized via a pathway involving reduction of the corresponding ribonucleotides, a step catalyzed by the enzyme ribonucleotide reductase (RNR) (for reviews, see refs. 1 and 2). Ribonucleotide reduction presents to the biochemist one of the more interesting evolutionary conundrums in central metabolism. The reduction pathway is almost certainly as old as the divergence of archaebacteria, eubacteria, and eukaryotes (3), as it is found in all modern organisms studied to date (1–4). It is unlikely that this reflects convergent evolution,

however. Direct removal of oxygen from a ribonucleotide has none of the chemical inevitability that would make different organisms invent it independently (4). Further, it would seem to be strategically acceptable to prepare deoxyribose derivatives more simply by an aldol condensation between acetaldehyde and the appropriate derivatives of glyceraldehyde (4, 5).

The catalysts for the reduction, the RNRs themselves, are not obviously homologous in various organisms, however. Although they all generate a 3′-radical on their nucleotide substrate (6), RNRs from different organisms differ greatly in how they initiate the radical reaction (1–4). At least four radical generation schemes are known. The RNR from humans generates the radical using a non-heme iron in a separate protein cofactor. Certain RNRs from eubacteria use non-heme iron as well. Other eubacterial RNRs use vitamin B₁₂, however, whereas still others have a poorly defined manganese cofactor or employ glycyl radicals in the reaction (1–4).

Different classes of RNRs have intriguing sequence "motifs" involving cysteines that appear to be important for the catalysis (in *Escherichia coli*, Cys-439, the putative radical site, and Cys-225 and Cys-462, which delivers two electrons and a proton) (7–9). These motifs offer tantalizing suggestions that all RNRs are related by common ancestry but underwent divergent evolution so massive that only traces of evidence for homology remain in the sequences themselves (1, 2, 6–8). These motifs are inadequate to provide a statistically significant case for homology, however, and motifs are notoriously inadequate for confirming homology in general (10).

To better understand the evolutionary history of the RNR complex of proteins, we have examined RNRs from archaebacteria. We report here that the RNRs from three organisms in two branches of the Archaea kingdom, *Halobacterium cutirubrum*, *Haloferax volcanii*, and *Thermoplasma acidophila*, all require B₁₂ as a cofactor. Further, we have purified the RNR from *T. acidophila*, cloned and sequenced the gene, and expressed the archaebacterial RNR in active form in *E. coli*. Through these analyses and the detection of sequence "bridges" connecting various classes of RNRs, we have convincingly confirmed the suggestions of Stubbe (1) and Reichard (2) that the core catalytic subunits of many (and perhaps all) RNRs are homologous. Recently developed tools for predicting secondary structure from sequence data (11) also confirmed suggestions of long-distance homology (12).

MATERIALS AND METHODS

RNR assays were performed with 2, 8-³H-ADP [NEN, 10 μM, 0.5 μCi (1 Ci = 37 GBq)] as substrate in 50 μl of Tris·HCl

Abbreviation: RNR, ribonucleotide reductase.
‡To whom reprint requests should be addressed at: Department of Chemistry, University of Florida, Gainesville, FL 32611.

buffer (100 mM Tris·HCl 1 mM dGTP 10 mM MgSO$_4$ 10 mM DTT 100 μM B$_{12}$). Reactions were started by adding a mixture of nucleotides, effectors, and coenzymes to protein ($\approx$0.5 μg); incubated (1 h, 55°C); and terminated (10–15% conversion) by boiling in a water bath (10 min). The mixtures were cooled to room temperature and diluted with 1 volume of Tris·OH buffer (100 mM, pH 10.0). Reactants and products were dephosphorylated (alkaline phosphatase, 1 unit, 1 h, 37°C), diluted with cold 2′-deoxyadenosine (50 nmol), separated by silica gel thin layer chromatography (MeOH saturated with potassium borate tetrahydrate), and counted.

Cells ($\approx$2 g/10 liters) of *T. acidophila* (DSM 1728) were grown at 55°C to late log phase (13), harvested by centrifugation (10 min at 2500 $\times$ g), washed [25 ml of 100 mM (NH$_4$)$_2$SO$_4$/25 mM KH$_2$PO$_4$/10 mM MgSO$_4$/5 mM CaCl$_2$, pH 1.5)], and stored at -80°C. To isolate RNR, cells (2.5 g) were freeze-thawed (two times, 5 min per cycle) in Tris (25 ml of 100 mM Tris, pH 8.0, with HCl), and sonicated (5 min). The supernatant (by centrifugation for 10 min at 30,000 $\times$ g) was treated (10 min at 4°C) with a 20% volume of cold polyethylene glycol 8000 (200 g/kg) in Tris buffer. The pellet (resulting from centrifugation for 10 min at 4°C at 30,000 $\times$ g) was dissolved in Tris. The solution was applied to a Q-Sepharose column (2.5 $\times$ 30 cm, pre-equilibrated with Tris·HCl buffer, 100 mM, pH 8.0), and the column was eluted with a linear gradient of NaCl (0–1 M, 250 ml/h) in Tris. Fractions with highest activity were treated at 3°C with (NH$_4$)$_2$SO$_4$ (saturated at 4°C, pH 7.0, with NaOH) to 30% saturation and incubated (10 min), and the pellet (centrifugation for 10 min at 30,000 $\times$ g) was redissolved in Tris (5 ml). This solution was applied to a hydroxyapatite column (2.5 $\times$ 5 cm, Bio-Rad, pre-equilibrated with Tris). The column was washed (Tris, 5 column volumes) and eluted with a K$_2$HPO$_4$ gradient (0–250 mM in Tris, RNR elution at $\approx$100 mM phosphate). Fractions containing enzyme were treated with a saturated (NH$_4$)$_2$SO$_4$ solution (to 55% saturation) and incubated overnight. The protein was recovered by centrifugation (10 min at 30,000 $\times$ g), redissolved in Tris (10 mM), and fractionated by HPLC gel filtration (LKB GlasPak TSK-G3000SW, 8 $\times$ 300 mm) at room temperature (10 mM Tris, pH 8.0, 0.5 ml/min). Fractions with activity were concentrated (ultrafiltration, Centricon 10, for 30 min at 3500 $\times$ g) to 0.75 ml. The RNR was essentially pure by gel electrophoresis (90-fold purification). An additional 1.1- to 1.3-fold purification was possible with HPLC-DEAE anion exchange chromatography (Biogel DEAE TSK 5PW, Bio-Rad), and the RNR was eluted (at room temperature, 0.75 ml/min) with a gradient of (NH$_4$)$_2$SO$_4$ (0–350 mM over 20 min, Tris) (Table 1).

Peptide were obtained from purified RNR by CNBr digestion. Protein following the hydroxyapatite column was precipitated [(NH$_4$)$_2$SO$_4$], the precipitate was dissolved in trifluoroacetic acid (TFA; 100 μl, 70% in water), and the mixture was treated with CNBr (few grains) and incubated (12 h at room temperature in the dark). The mixture was diluted (water, 0.6 ml) and dried

(vacuum). The residue was dissolved in TFA (0.1 ml, 50%), and the peptides were resolved by C18-HPLC chromatography (Vydac 218TP54, A = 0.1% TFA in water, B = 0.05% TFA in CH$_3$CN, 0–60% B over 60 min, then 60–100% B over 15 min, detection at 215 nm) (14) and then sequenced (each 2 μg) by automated Edman degradation. An N-terminal decapeptide and an interior sequence of 23 amino acids was obtained.

A subgenomic library for *T. acidophila* was constructed in pUC18 (15). A nondegenerate 62-mer hybridization probe (RnrP1), based on the internal peptide sequence, was used to Southern blot the library, yielding a 2.8-kbp *Hin*dIII fragment containing the N-terminal region of the RNR gene. A second probe (RnrP3) was designed from the DNA sequence of this fragment and used to recover a 3.8-kbp *Hin*dIII fragment containing the C-terminal end of the gene. Chromosomal DNA from *T. acidophila* was digested separately with *Eco*RI, *Hin*dIII, or *Bam*HI. Fragments were resolved by agarose electrophoresis (1%). Long nondegenerate probes were designed from peptide sequences obtained by Edman degradation using codons chosen based on an analysis of the existing data base of *T. acidophila* encoded sequence, and used in Southern blots. Probe RnpP1 (5′-AAGATAGATGGCAAGACGTTCACGGAT-GTGGCGAAGTCGTACATACTGTACAGGGAGAAGAG, from the N-terminal sequence, hybridized to 10-, 2.8-, and 1.7-kbp fragments in the three digests, respectively). The 1.7-kbp *Bam*HI fragment was extracted from the gel, mixed with pUC18 digested with *Bam*HI, ligated, and used to transform JM105 cells, which were screened by colony hybridization with RnrP1 at 50°C. Ten plasmids were isolated from the positive clones and restriction-mapped, identifying eight with the 1.7-kbp segment inserted in both orientations. The insert was sequenced and compared with the peptide sequences obtained above.

The RNR from *T. acidophila* was heterologously expressed in *E. coli*. Chromosomal DNA (3 ng) from *T. acidophila* containing the RNR gene was amplified (PCR, Amplitaq) (16). The DNA was digested with *Not*I and *Nde*I, the fragments were resolved by agarose (1%) gel electrophoresis, and the 2.7-kbp band was ligated into a pET23b vector, similarly digested, behind its lac promoter. The product was used to transform JM105 cells. Plasmid was isolated from positive colonies, and the RNR gene was sequenced. The RNR gene was induced (isopropyl β-[sca]d-thiogalactoside) at 37°C (4 h, with time points) and purified as above.

Secondary structure predictions were made using the heuristics developed in these laboratories (17) and implemented in the DARWIN server (http://cbrg.inf.ethz.ch/).

RESULTS

Based on size exclusion chromatography, active RNR isolated from *T. acidophila* was estimated to have a molecular weight of 100,000. SDS/PAGE gave a similar a weight, indicating that

Table 1.   Purification of RNR from *T. acidophila*

Step	Activity, nmol/h	Volume, ml	Total protein, mg	Specific activity, nmol/h·mg	Yield, %	Purification factor
Crude extract	2050	22	246	8.3	100	1
PEG precipitation	1590	10	115	13.8	78	1.7
Q-Sepharose	2310	40	41	56	113	6.7
(NH$_4$)$_2$SO$_4$ precipitate	2160	5	11.3	191	105	23
BioGel HTP	2060	0.5	4.48	460	100	55
HPLC sizing	255	2.1	0.33	773	12.4	93
HPLC-DEAE	109	2.75	0.12	908	5.3	110

From a 10-liter culture of *T. acidophila* in Tris·HCl buffer (100 mM, pH 8.0, 3°C). Sonicated extracts from 2 g of cells were fractionated (4-8%, wt/vol) with polyethylene glycol (PEG) 8000. The enriched protein was loaded onto a 150-ml Q-Sepharose column and eluted with a linear gradient of NaCl (0-1 M). After fractionation [30% saturation (NH$_4$)$_2$SO$_4$], protein was loaded onto a 25-ml BioGel HTP column and eluted with a linear gradient of potassium phosphate (0-0.25 M, in 100 mM Tris·HCl, pH 8.0). RNR was precipitated [30-55% saturated (NH$_4$)$_2$SO$_4$], and the concentrated protein was purified by HPLC size exclusion chromatography (LKB TSK-G3000, 8 $\times$ 300 mm, room temperature).

Biochemistry: Tauer and Benner

Proc. Natl. Acad. Sci. USA 94 (1997)   55

the active protein is a monomer, as is the B_{12}-dependent RNRs from *Lactobacillus* (7), *Thermus aquaticus* X-1 (18), and *Anabaena* sp. (7119) (19). However, like the Fe-dependent RNR from *E. coli* and mammals, the preferred substrates are nucleoside diphosphates.

The temperature optimum for the reduction is 55°C, corresponding to the habitat temperature of *T. acidophila*. Reduction of ADP follows pseudo Michaelis–Menten kinetics (a K_m for ADP of 64 μmol at 100 μM coenzyme B_{12} and 1 mM 2′-dGTP). Reduction of ADP depends on added DTT as a reducing agent and dGTP as an effector; TTP is a less potent effector, whereas dATP, ATP, and dCTP show no effector potency (data not shown). The enzyme requires added 5′-deoxyadenosyl cobalamin, and catalyzes release of tritium from 5′-[^{3}H]deoxyadenosyl cobalamin to water, establishing that the enzyme uses B_{12} (20). Hydroxyurea (50 mM), an inhibitor of iron-dependent RNRs, does not inhibit the RNR. These data suggest that the *Thermoplasma* RNR is not iron-dependent.

Because the RNRs from *Halobacterium cutirubrum* and *Haloferax volcanii* were instable at [NaCl] lower than those found physiologically, we were unable to obtain these proteins homogenous in active form. Assays on partially purified fractions suggested that these RNRs also use B_{12} as a cofactor. Coenzyme B_{12} (but not vitamin B_{12}) stimulated the activity of both proteins, and neither was inhibited by hydroxyurea (data not shown). TTP increased the rate of nucleotide reduction in both cases as well.

The recombinant RNR expressed in *E. coli* had a specific activity (1100 nmol/h·mg) slightly higher than the highest obtained with the protein isolated directly from *T. acidophila* cells. This specific activity is rather low; it is not clear whether this indicates a correspondingly low rate *in vivo* or whether the *in vitro* assay conditions fail to reproduce *in vivo* conditions adequately.

The N-terminal untranslated region contains an A+T rich putative box A 24 bases upstream from a putative box B (not shown), fitting the consensus archaebacterial box A and box B sequences (21). A ribosome binding site complementary to the 16S rRNA from *T. acidophila* (22) lies close to the initiating ATG. The archaebacterial protein was readily expressed by *E. coli* (23). The open reading frame encodes a protein with a calculated isoelectric point of 7.05 and molecular weight of 97,000, in good agreement with the biochemical experiments.

Most remarkable is the sequence of the *T. acidophila* RNR protein (Fig. 1). The protein displays sequence similarity that indisputably establishes homology of one segment to the glycyl radical-dependent anaerobic RNR (residues 1–150) from *E. coli*, and of another segment to the iron-dependent RNR from *E. coli* and mammals (residues 250–680). The second homology is supported by 25% sequence identity in the catalytic domain that includes the redox active cysteines of the Fe-dependent enzyme (Cys-225 and Cys-462), and Cys-439, the putative thiyl radical (6). Homology was also indicated by the correspondence between the predicted secondary structure of the RNR with the experimental structure of the Fe-dependent RNR from *E. coli* (24). Further, the regions of sequence similarity corresponded to domains observed in the crystal structure of the Fe-dependent RNR. Thus, the entire alpha–beta domain of the *E. coli* RNR is found in homologous form in the RNR from *T. acidophila*, except for the final two helix-strand elements, which may dock the radical-generating domain of the heterodimer in the *E. coli* enzyme (24). The *T. acidophila* RNR presumably has its radical-generating domain fused in one peptide chain (see below) and therefore does not need docking elements.

Pursuing this observation, the database was probed using the sequence from the *Lactobacillus* RNR (B_{12}-dependent). This sequence is similar (30.9% identical, similarity score 957, PAM 126, over 700 amino acids) to an open reading frame from mycobacteriophage L5, which infects the Gram-positive bacterium *Mycobacterium tuberculosis*. This similarity indisputably establishes homology between *Lactobacillus* and *Mycobacterium* RNRs. The *Mycobacterium* sequence was successfully aligned as

well with the *T. acidophila* RNR sequence (Fig. 1). While no similarity exists to adequately establish homology between the B_{12}-dependent *Lactobacillus* sequence and any of the Fe-dependent RNR sequences, the statistically significant connections of the *Lactobacillus* sequence via the *Mycobacterium* sequence to the *T. acidophila* sequence permits a "bridge" (27) to be made between the catalytic domain of the B_{12}-dependent *Lactobacillus* RNR and the catalytic domains of the Fe-dependent *E. coli* (Fig. 1). This bridge was confirmed by the correspondence between the predicted secondary structure obtained from the *Lactobacillus–Mycobacterium* pair with the secondary structure determined by crystallography in the core domain of the *E. coli* RNR (24). Thus, the additional sequences unify the catalytic domains of all RNRs except the anaerobic enzyme from *E. coli* (dependent on the glycyl radical). This implies in turn that the preference for substrate (NDP versus NTP for the *Lactobacillus* RNR), which is presumably specified in the catalytic domain, is not evolutionarily conserved but rather diverged during the divergence of the catalytic domain from a common ancestor.

Anaerobic RNR from *E. coli* (using a glycyl radical) does not stand in evolutionary isolation, however. The N-terminal domain of the anaerobic RNR was significantly similar in sequence to the N-terminal region of *T. acidophila* RNR, and the Fe-dependent RNRs also displayed significant similarity to the last part of this domain (*T. acidophila* residues 110–150). Further, a predicted secondary structure of the N-terminal domain suggests that it is predominantly helical, as is the domain in the *E. coli* enzyme (24). This prediction therefore suggests an evolutionary bridge between the glycyl radical, dinuclear iron, and *T. acidophila* B_{12}-dependent RNRs, at least in the N-terminal domain.

A search of the sequence data base left only two regions of the *T. acidophila* RNR without obvious homologs. The first is a 100-aa segment (residues 150–247) connecting the first (presumably regulatory) domain to the catalytic domain. DARWIN failed to find any sequence in the data base alignable to this segment. This region lies in a region of the fold that should tolerate insertions, and its presence does not compromise the conclusion that the *T. acidophila* B_{12}-dependent and Fe-dependent RNRs are homologous.

DARWIN also found no statistically significant similarities between any protein in the data base and residues 650–857. Based on analogy with other RNRs, the region was tentatively assigned as a B_{12}-binding domain. No significant alignment could be obtained with this domain in the sequence of the B_{12}-dependent RNR from *Lactobacillus*, however. Nor was statistically significant sequence similarity found to B_{12}-binding domains from other B_{12}-dependent proteins. The region contained, however, two signature sequences (DXHXXG and SXL, but not GG) found in the B_{12}-binding domain of an entirely different B_{12}-dependent enzyme, the methionine synthase from *E. coli*, proposed from the crystal structure to be important for cofactor binding (29, 30).

DISCUSSION

There has been much discussion concerning biochemical markers that might be used to define distant evolutionary relationships between macromolecules, especially after the similarities in sequences that might provide a statistically significant statement of homology have vanished. It is generally accepted that the tertiary structure of a protein is conserved after sequence similarity has vanished, and can be used as an indicator of homology (31). Unfortunately, the tertiary structure is a difficult feature of a biomolecule to determine experimentally. Therefore, other properties, including mechanism, substrate specificity, stereospecificity, quaternary structure, codon selection, regulatory patterns, disulfide connectivity, the conservation of short sequence motifs, and the order of presumed catalytic residues in the polypeptide chain, have
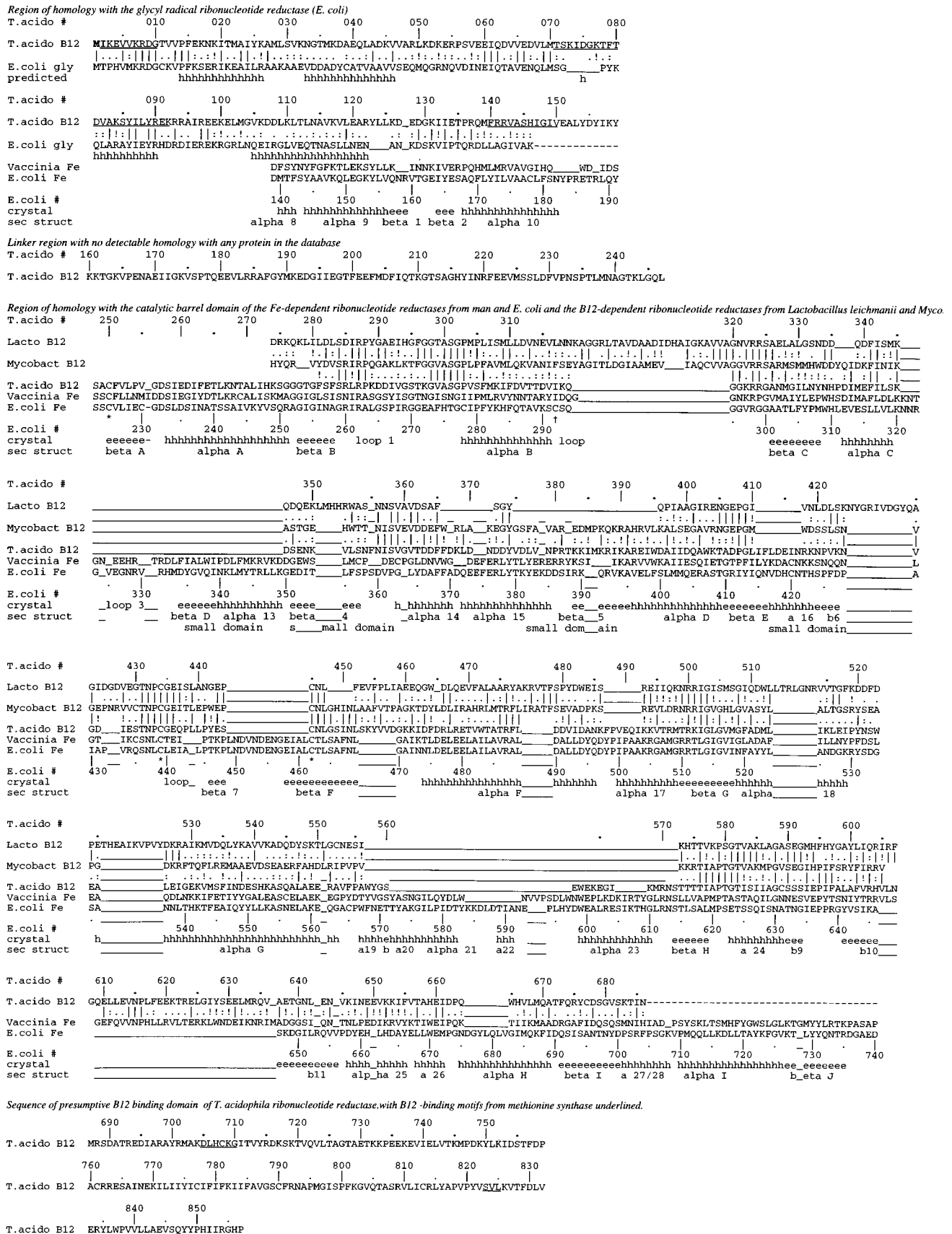
Region of homology with the glycyl radical ribonucleotide reductase (E. coli)

```
T.acido #        010       020       030       040       050       060       070       080
                  |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12  MIKEVVKRDGTVVPFEKNKITMAIYKAMLSVKNGTMKDAEQLADKVVARLKDKERPSVEEIQDVVEDVLMTSKIDGKTFT
             |...:|||||..|||::.:!..||.!.::::.....:.:.|.|.::...:!:||||.:||::.||:.  .!.:
E.coli gly   MTPHVMKRDGCKVPFKSERIKEAILRAAKAAEVDDADYCATVAAVVSEQMQGRNQVDINEIQTAVENQLMSG____PYK
predicted             hhhhhhhhhhhhh     hhhhhhhhhhhhhhh                              h

T.acido #          090       100       110       120       130       140       150
                    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12  DVAKSYILYREKRRAIREEKELMGVKDDLKLTLNAVKVLEARYLLKD_EDGKIIETPRQMFRRVASHIGIVEALYDYIKY
             ::|!:||.||..|.  ||:!.!..:  .....   .:.:|!.||.:!:..!::.
E.coli gly   QLARAYIEYRHDRDIEREKRGRLNQEIRGLVEQTNASLLNEN___AN_KDSKVIPTQRDLLAGIVAK------------
             hhhhhhhhhhh        hhhhhhhhhhhhhhhhhhh
Vaccinia Fe                      DFSYNYFGFKTLEKSYLLK__INNKIVERPQHMLMRVAVGIHQ____WD_IDS
E.coli Fe                        DMTFSYAAVKQLEGKYLVQNRVTGEIYESAQFLYILVAACLFSNYPRETRLQY

E.coli #                          |    .    |    .    |    .    |    .    |    .    |
crystal                          140       150       160       170       180       190
sec struct                    hhh hhhhhhhhhhhhhheee  eee hhhhhhhhhhhhhh
                          alpha 8     alpha 9 beta 1 beta 2   alpha 10
```

Linker region with no detectable homology with any protein in the database

```
T.acido # 160       170       180       190       200       210       220       230       240
           |    .    |    .    |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12 KKTGKVPENAEIIGKVSPTQEEVLRRAFGYMKEDGIIEGTFEEFMDFIQTKGTSAGHYINRFEEVMSSLDFVPNSPTLMNAGTKLGQL
```

Region of homology with the catalytic barrel domain of the Fe-dependent ribonucleotide reductases from man and E. coli and the B12-dependent ribonucleotide reductases from Lactobacillus leichmanii and Myco.

```
T.acido #     250       260       270       280       290       300       310                320       330       340
               |    .    |    .    |    .    |    .    |    .    |    .    |                   |    .    |    .    |
Lacto B12                          DRKQKLILDLSDIRPYGAEIHGFGGTASGPMPLISMLLDVNEVLNNKAGGRLTAVDAADIDHAIGKAVVAGNVRRSAELALGSNDD__QDFISMK__
                                   ..::.|.|:|.|||.|.:!:..||||||!!..:||.:.|:.:.:||..|.:!.:!!     |.||.||.||.:.:!      :|.   :.|:!|
Mycobact B12                       HYQR__VYDVSRIRPQGAKLKTFGGVASGPLPFAVMLQKVANIFSEYAGITLDGIAAMEV___IAQCVVAGGVRRSARMSMMHWDDYQIDKFINIK__
                                   !..!|!:..........||||||:.::.:!::..::::....                    !.||.|.|.!::::.. :| :!|.:|
T.acido B12  SACFVLPV_GDSIEDIPETLKNTALIHKSGGGTGFSFSRLRPKDDIVGSTKGVASGPVSFMKIFDVTTDVIKQ_____GGKRRGANMGILNYNHPDIMEFILSK__
Vaccinia Fe  SSCFLLNMIDDSIEGIYDTLKRCALISKMAGGIGLSISNIRASGSYISGTNGISNGIIPMLRVYNNTARYIDQG_____GNKRPGVMAIYLEPWHSDIMAFLDLKKNT
E.coli Fe    SSCVLIEC-GDSLDSINATSSAIVKYVSQRAGIGINAGRIRALGSPIRGGEAFHTGCIPFYKHFQTAVKSCSQ_____GGVRGGAATLFYPMWHLEVESLLVLKNNR

E.coli #                 *        |    .    |    .    |    .    |    .    |    t                          |    .    |    .    |
crystal                 230       240       250       260       270       280       290                     300       310       320
sec struct     eeeeee-  hhhhhhhhhhhhhhhhhhh eeeeee   loop 1        hhhhhhhhhhhhhh loop                       eeeeeeee   hhhhhhhh
               beta A      alpha A          beta B                   alpha B                                  beta C     alpha C
```

```
T.acido #               350       360       370       380       390       400       410       420
                         |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
Lacto B12     _____QDQEKLMHHRWAS_NNSVAVDSAF_____SGY_____QPIAAGIRENGEPGI_____VNLDLSKNYGRIVDGYQA
                       .....!  |.||..|      _  __  :!:|.||       |.|!|||||               :.!::.||
Mycobact B12  _____ASTGE____HWTT_NISVEVDDEFW_RLA__KEGYGSFA_VAR_EDMPKQKRAHRVLKALSEGAVRNGEPGM_____WDSSSLSN_____V
                       .|..:_____   .|.|||!!! !|...:!|..:|._.| .:._|!!.|.|!|               :.!::|  _____|
T.acido B12   _____DSENK____VLSNFNISVGVTDDFFDKLD__NDDYVDLV_NPRTKKIMKRIKAREIWDAIIDQAWKTADPGLIFLDEINRKNPVKN_____V
Vaccinia Fe   GN_EEHR___TRDLFIALWIPDLFMKRVKDDGEWS____LMCP__DECPGLDNVWG__DEFERLYTLYERERRYKSI___IKARVVWKAIIESQIETGPPFILYKDACNKKSNQQN_____L
E.coli Fe     G_VEGNRV__RHMDYGVQINKLMYTRLLKGEDIT____LFSPSDVPG_LYDAFFADQEEFERLYTKYEKDDSIRK__QRVKAVELFSLMMQERASTGRIYIQNVDHCNTHSPFDP_____A

E.coli #                |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
crystal                330       340       350       360       370       380       390       400       410       420
sec struct    _          __   eeeeeehhhhhhhhhh eeee____eee   h_hhhhhh hhhhhhhhhhhhhh   ee__eeeeehhhhhhhhhhhhhhhheeeeeeehhhhhhheee  _____
              _              beta D alpha 13 beta___4          _alpha 14 alpha 15      beta__5            alpha D   beta E   a 16  b6
                        small domain     s____mall domain                           small dom__ain                         small domain_____
```

```
T.acido #     430       440             450       460       470       480       490       500       510       520
               |    .    |    .          |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
Lacto B12     GIDGDVEGTNPCGEISLANGEP_____CNL___FEVFPLIAEEQGW_DLQEVFALAARYAKRVTFSPYDWEIS_____REIIQKNRRIGISMSGIQDWLLTRLGNRVVTGFKDDFD
              |...|..|||||||:|..  |               |:|   .|!:.|.:.|.|.:.::.|::|.!           !!|:|!||||||  .||.|   ..:|.|.:
Mycobact B12  GEPNRVVCTNPCGEITLEPWEP_____CNLGHINLAAFVTPAGKTDYLDLIRAHRLMTRFLIRATFSEVADPKS_____REVLDRNRRIGVGHLGVASYL_____ALTGSRYSEA
              |!  !..!|||||.:|! !|:.               :|. ..||.!:...!|.|.:...::|.::.:||.!|::|:           :|:!|||||.|   .||.!
T.acido B12   GD___IESTNPCGEQPLLPYES_____CNLGGSINLSKYVVDGKKIDFDRLRETVWTATRFL____DDVIDANKFPVEQIKKVTRMTRKIGLGVMGFADML_____IKLEIPYNSW
Vaccinia Fe   GT___IKCSNLCTEI___PTKPLNDVNDENGEIALCTLSAFNL_____GAIKTLDELEELAILAVRAL____DALLDYQDYPIPAAKRGAMGRRTLGIGVIGLADAF_____ILLNYPFDSL
E.coli Fe     IAP__VRQSNLCLEIA_LPTKPLNDVNDENGEIALCTLSAFNL_____GAINNLDELEELAILAVRAL____DALLDYQDYPIPAAKRGAMGRRTLGIGVINFAYYL_____ANDGKRYSDG

E.coli #      |    _    |     _  .   |    .    |   *    .    |    .    |    .    |    .    |    .    |    .    |    .    |
crystal      430       440          450       460       470       480       490       500       510       520       530
sec struct    _____   loop_ eee   eeeeeeeeeeee_____   hhhhhhhhhhhhhhhhhh____hhhhhhhh   hhhhhhhhhheeeeeeeeeehhhhhh_____hhhhh
                          beta 7         beta F   _____            alpha F_____             alpha 17  beta G  alpha_____ 18
```

```
T.acido #     530       540       550       560                570       580       590       600
               |    .    |    .    |    .    |                   |    .    |    .    |    .    |
Lacto B12     PETHEAIKVPVYDKRAIKMVDQLYKAVVKADQDYSKTLGCNESI_____KHTTVKPSGTVAKLAGASEGMHFHYGAYLIQRIRF
              |._____.|||.::::.|.:....|..:.:|               ||.|.!:.|||||!| |.|||..:.!...|:
Mycobact B12  PG_____DKRFTQFLREMAAEVDSEAERFAHDLRIPVPV_____KKRTIAPTGTVAKMPGVSEGIHPIFSRYFIRRV
              ..:.....|!.!.....|....|.::|               :.!.!.|||||.|.| :.:.|.::|.||
T.acido B12   EA_____LEIGEKVMSFINDESHKASQALAEE_RAVFPAWYGS_____EWEKEGI_____KMRNSTTTTIAPTGTISIIAGCSSSIEPIPALAFVRHVLN
Vaccinia Fe   EA_____QDLNKKIFETIYYGALEASCELAEK_EGPYDTVVGSYASNGILQYDLW_____NVVPSDLWNWEPLKDKIRTYGLRNSLLVAPMPTASTAQILGNNESVEPYTSNIYTRRVLS
E.coli Fe     SA_____NNLTHKTFEAIQYYLLKASNELAKE_QGACPWFNETTYAKGILPIDTYKKDLDTIANE___PLHYDWEALRESIKTHGLRNSTLSALMPSETSSQISNATNGIEPPRGYVSIKA___

E.coli #      |    .    |    .    |    .    |    .    |    .    |    .    |    _    .    |    .    |    .    |    .    |
crystal      530       540       550       560       570       580       590       600       610       620       630       640
sec struct    h_____hhhhhhhhhhhhhhhhhhhhhhhhhhh_hh   hhhhhhhhhhhhhh    hhh ___    hhhhhhhhhhhh  eeeeee  hhhhhhhhhee      eeeeee___
                          alpha G               _        a19 b a20  alpha 21  a22       alpha 23     beta H   a 24       b9
```

```
T.acido #    610       620       630       640       650       660       670       680
              |    .    |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12   GQELLEVNPLFEEKTRELGIYSEELMRQV_AETGNL_EN_VKINEEVKKIFVTAHEIDPQ_____WHVLMQATFQRYCDSGVSKTIN-------------------------------
              :|:.|||.:  ....|..!!:!!|..:! |!.|::._::.|.!!!!!|.|.|..                :.!.|.  ..:.!..|.:|.|:
Vaccinia Fe   GEFQVVNPHLLRVLTERKLWNDEIKNRIMADGGSI_QN_TNLPEDIKRVYKTIWEIPQK_____TIIKMAADRGAFIDQSQSMNIHIAD_PSYSKLTSMHFYGWSLGLKTGMYYLRTKPASAP
E.coli Fe     _____SKDGILRQVVPDYEH_LHDAYELLWEMPGNDGYLQLVGIMQKFIDQSISANTNYDPSRFPSGKVPMQQLLKDLLTAYKFGVKT_LYYQNTRDGAED

E.coli #                               |    .    |    .    |    .    |    .    |    .    |    .    |    .    |    _    .    |    .    |
crystal                               650       660       670       680       690       700       710       720       730       740
sec struct    _____eeeeeeeeee hhhh_hhhhh hhhhh   hhhhhhhhhhhhhhhh eeeeeeeeehhhhhhhhh hhhhhhhhhhhhhhhhhhee_eeeeeee
                                        b11        alp_hha 25  a 26      alpha H      beta I  a 27/28    alpha I         b_eta J
```

Sequence of presumptive B12 binding domain of T. acidophila ribonucleotide reductase, with B12 -binding motifs from methionine synthase underlined.

```
T.acido #    690       700       710       720       730       740       750
              |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12   MRSDATREDIARAYRMAKDLHCKGITVYRDKSKTVQVLTAGTAETKKPEEKEVIELVTKMPDKYLKIDSTFDP

             760       770       780       790       800       810       820       830
              |    .    |    .    |    .    |    .    |    .    |    .    |    .    |    .    |
T.acido B12   ACRRESAINEKILIIYICIFIFKIIFAVGSCFRNAPMGISPFKGVQTASRVLICRLYAPVPYVSVLKVTFDLV

             840       850
              |    .    |    .    |
T.acido B12   ERYLWPVVLLAEVSQYYPHIIRGHP
```

FIG. 1.   Sequence of the B12-dependent ribonucleotide reductase (RNR, EC 1.17.4.2) from *T. acidophila* aligned with other proteins. T.acido #, sequential numbering of sequence of B12-dependent ribonucleotide reductase from *T. acidophila*; T.acido B12, sequence of the B12-dependent RNR from *T. acidophila* (GenBank accession no. U73619); *E. coli* gly, sequence of the glycyl radical-dependent anaerobic RNR from *E. coli* (GenBank accession no. P289093); Vaccinia Fe, sequence of the catalytic subunit of the Fe-dependent RNR from vaccinia virus (strain Copenhagen; GenBank accession no. P20503; Lacto B12, sequence of the B12-dependent RNR from *Lactobacillus leichmannii* (7); Mycobact B12, sequence of

Biochemistry: Tauer and Benner

Proc. Natl. Acad. Sci. USA 94 (1997)   57

been used to define family relationships among proteins for which three-dimensional models do not exist (32). We have argued against (4) and others have argued for (2, 6) evolutionary homology within the ribonucleotide reductase family based on these and other nonstructural grounds.

Similar discussions surround the assignment of "primitive" biomolecular behaviors to a common ancestor in a protein family. The most direct approach requires reconstructing the ancestral protein in the laboratory and studying its behavior (33). This is, unfortunately, also experimentally challenging, and many have suggested alternative criteria for inferring primitive behaviors. With RNRs, some have argued that glycyl radical-dependent RNRs are primitive (2), as these operate anaerobically, and life presumably originated in an anaerobic environment before oxygen emerged 2.5 billion years ago (4). Others suggest that the Fe-dependent enzymes are primitive, because ancient organisms lived on an earth that was warmer, and Fe-dependent enzymes avoid unstable cofactors (34). B_{12} might also be viewed as a cofactor that contains elements of structure that might have arisen prebiotically (35), implying perhaps that B_{12}-dependent RNRs are ancient (4).

Some time ago, we systematically examined biochemical data to learn whether we could find examples of nonstructural behaviors of proteins that could be shown to have diverged within a pair of proteins that was indisputably homologous by sequence analysis (32). Many examples established that most, if not all, of these nonstructural features can diverge in a biomolecular family well before sequence similarity ceases to provide a reliable indicator of homology. This implies, in turn, that nonstructural features will not generally be reliable indicators of homology (or nonhomology) in proteins that have no statistically significant sequence similarity.

The reason for this is simple enough to state in its general form. Most nonstructural properties of a protein are intricately tied to its selected function, and diverge (or are conserved) in response to changes in (or conservation of) the environmental parameters to which natural selection responds. Nonstructural features are unreliable as evolutionary markers for the same reason that the molecular clock often fluctuates episodically (36), metabolic behaviors of microorganisms are poor indicator of phylogeny (37), and functional macroscopic features of an animal are not as useful as nonfunctional features for deducing patterns of ancestry (38). Natural selection perturbs selected behaviors in a fashion that is difficult to describe using analytically simple models.

The ribonucleotide reductase provides a fascinating illustration of these problems. The choice of cofactor to generate a radical in a RNR proves to be an inadequate criterion for assessing homology. Instead, the radical generator appears to have diverged to reflect the availability of the cofactor in the environment. When vitamin B_{12} is scarce, selection pressure has evidently favored its elimination as a cofactor. Accordingly, plants, which lack B_{12}, have a non-B_{12} RNR (39), as do mammals, which do not biosynthesize B_{12}. Both rely on dinuclear iron as the ultimate source of the radical in the RNR reaction, which in turn requires oxygen, which is present in the environments of plants and animals. Under anaerobic conditions, *E. coli*, which also lacks B_{12} biosynthesis (and which has alternative pathways to biosynthesize essential biomolecules such as methionine if B_{12} is unavailable), dinuclear iron is not an option as a cofactor; hence, the use of a glycyl radical RNR. In contrast, in a metabolic context where vitamin B_{12} is plentiful, selection pressure evidently favor its use. Archaebacteria are rich in B_{12}; they biosynthesize it and use it to biosynthesize components

of their membranes. Thus, a purely adaptive model explains the choice of cofactor in the RNR family, which in turn makes the choice of cofactor a poor indicator of homology.

At least two of the three critical thiols (Cys-439 and Cys-462) are conserved and embedded in a homologous protein scaffolding in all of the RNRs discussed here. These cysteines are believed to participate in a reaction mechanism involving a thiyl radical at position 439 (2, 6). This suggests that a thiyl mechanism was the initial solution chosen by nature, and that this solution was sufficiently efficient and superior to others that nature might have tried to have persisted throughout the period of natural history that separates archaebacteria, eubacteria, and eukaryotes. The alternative explanation, that nature tried no other mechanistic alternatives, cannot be ruled out, however. Thus, it will be interesting to learn whether a similar evolutionary connection might be found that will join the catalytic domains of the anaerobic *E. coli* RNR and the manganese-dependent enzymes into the evolutionary family defined by the RNR from *T. acidophila* and its homologs.

Cys-225, which participates in the delivery of the two electrons and a proton to carbon-2 of the substrate, remains paradoxical. It is found in a region of acceptable sequence similarity between the *T. acidophila* enzyme and the *E. coli* aerobic enzyme. It is not aligned, however, in the *Lactobacillus* and *Mycobacterium* phage sequences (Fig. 1). DARWIN will exclude a motif from an alignment if it does not meet rigorous standards of significance. Even by eye, however, no cysteine residue is found 20–40 residues upstream of the start of the global alignment (*E. coli* position 249 in Fig. 1) in either the phage or the *Lactobacillus* sequence. The first conserved cysteine appears 108 residues (for the *Lactobacillus* sequence) and 126 residues (for the phage sequence) before the start of the global alignment (sequences not shown). Both cysteines are embedded in a segment indicative of an active site. If the universal mechanism is to hold, these cysteines must be the equivalent of Cys-225 in the *E. coli* RNR, with an additional segment of approximately 100 aa predicted to be spliced into the folded structure between these cysteines and the remainder of the barrel.

In this context, it is important to note that although the *Mycobacterium* phage RNR is presumably B_{12}-dependent based on its sequence, this has not been established experimentally. Further, an RNR from *M. tuberculosis*, a presumed host of the phage, is an iron-dependent RNR (40). This itself is intriguing, as viruses generally share metabolic similarities with their hosts. Still more intriguing is the fact that the *M. tuberculosis* RNR is most closely related to the nrdEF gene from *Salmonella typhimurium* (41), which in turn is unusual as an organism for having two RNRs belonging to the same biochemical class (42).

A structural model for the divergence of cofactor usage is also apparent from this model. The final two alpha–beta units in the *Thermoplasma* RNR, the putative sites for docking of the Fe-subunit in the *E. coli* enzyme (24), are missing from the catalytic domain of the *Thermoplasma* enzyme. Replacing it is the putative B_{12}-binding domain fused as part of a single polypeptide chain. The Fe-radical is some distance from the critical thiol residues at the active center in the *E. coli* RNR, implying that long range electron transfer is necessary for the reaction to proceed (43). Tyr-730 and Tyr-731, conserved in the Fe-dependent enzymes, have been proposed to be involved in this process. These tyrosines are not found in the *T. acidophila* RNR. His-707, the presumptive ligand to cobalt in the B_{12} cofactor, comes close to the place where these tyrosines might be aligned, however. It is intriguing to

---

the presumed B_{12}-dependent RNR (Gene 50 protein) from mycobacteriophage L5 (GenBank accession no. Q05262); *E. coli* **#**, sequential numbering of the sequence of the Fe-dependent ribonucleotide reductase from *E. coli*, corresponding to numbering in ref. 24; pred, secondary structure (h = helix; e = strand) predicted by method in refs. 17, 25, and 26; crystal secondary structure, secondary structure (h = α-helix; e = β-strand) determined by crystallography (24) for the *E. coli* enzyme, with designation of the secondary structure below. Multiple alignment was from DARWIN (12, 27). Gaps are not moved to reflect experimental or predicted structures. A dash indicates insignificant sequence similarity, and an underscore indicates deletion. Underlined sequences in *T. acidophila* RNR are peptide sequences of CNBr peptide fragments. Cysteines involved in catalysis (Cys-225, Cys-462, and Cys-439) are marked with an asterisk. The cysteine in the regulatory site is marked by a dagger (28).

speculate that the distance over which the electron is transferred in the *T. acidophila* enzyme is not as long as in the Fe-dependent enzymes, because the radical generating unit, here the B$_{12}$, can approach the active center of the catalytic domain more closely.

Two tools used in this evolutionary analysis should be applicable generally. First, while experimental determination of folded structure remains difficult, prediction of folded structure has become almost routine when more than a single homologous sequence is available (11, 25, 26). Analyses of patterns of variation and conservation among homologous sequences have proven to provide reliable predictions of secondary structure, as demonstrated by bona fide predictions made and announced before an experimental structure is known (11, 25, 26). Although predictions with smaller numbers of sequences are less reliable (and, normally, insufficiently reliable to serve as a starting point for building tertiary structural models), they can indicate whether two proteins are homologous following the loss of clear sequence similarity (12). In the RNR family, predictions of secondary structure suggest that at least part of the helical first domain of the *Thermoplasma* and glycyl radical RNR is homologous to analogous domain of the Fe-dependent RNRs. Other predictions confirm the proposal that the *Lactobacillus* enzyme is homologous to the Fe-dependent RNRs in the central barrel of the catalytic domain. The second prediction finds close analogy as a prediction problem to the prediction of phospho-β-D-galactosidase barrel (44), where prediction was possible for the conserved, core elements of the beta barrel but not for the external elements. Predicted secondary structures should become important tools for detecting long distance homology, especially as genome sequencing projects make sequence data more abundant.

The second evolutionary tool illustrated by this analysis is the use of sequence bridges to connect distant sequences into "connected" components (12). Here, the sequences from *Thermoplasma* and *Mycobacterium* establish an evolutionary bridge that argues for homology between the B$_{12}$-dependent enzyme from *Lactobacillus* and the Fe-dependent enzyme from *Escherichia*. A simple comparison of the sequences of the enzymes from *Lactobacillus* and *Escherichia* individually is unable to detect statistically significant sequence similarity. The theoretical reasons for such bridges remain unclear, even as their value in an evolutionary analysis is obvious. As genomic data bases become completed, bridges should become easier to find.

The B$_{12}$ binding site of the *T. acidophila* RNR is not, however, securely identified. Sequence motifs (such as the DXHXXG sequence) are generally unreliable in identifying long distance homology (10). They become reliable only when embedded in a secondary structure prediction that shows that the same motif in two protein families is flanked by the same secondary structural elements. The best prediction assigns an alpha–beta structure in this region (as is seen in the B$_{12}$-binding domain of methionine synthase), but with a large segment (some 70 aa) inserted between the critical motifs, and little correspondence between predicted and experimental secondary structure elements elsewhere. A prediction based on a single sequence is, however, little more than an educated guess (45). Therefore, further sequence data will be necessary to learn whether the B$_{12}$-binding domain of the *T. acidophila* RNR is homologous to the same domain in methionine synthase. With the completion of archaebacterial genomes imminent, these additional data should soon be available.

1. Stubbe, J. (1990) *J. Biol. Chem.* **266,** 5329–5332.
2. Reichard, P. (1993) *Science* **260,** 1773–1777.
3. Harder, J. (1993) *FEMS Microbiol. Rev.* **12,** 273–292.
4. Benner, S. A., Ellington, A. D. & Tauer, A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7054–7058.
5. Hoffee, O. M., Rosen, O. M. & Horecker, B. L. (1965) *J. Biol. Chem.* **240,** 1512–1516.
6. Mao, S. S., Holler, T. P., Yu, G. X., Bollinger, J. M. J., Booker, S., Johnston, M. I. & Stubbe, J. (1992) *Biochemistry* **31,** 9733–9743.
7. Booker, S. & Stubbe, J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8352–8356.
8. Lin, A.-N., Ashley, G. W. & Stubbe, J. (1987) *Biochemistry* **26,** 6905–6909.
9. Sun, X., Harder, J., Krook, M., Jörnvall, H. & Sjöberg, B.-M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 577–581.
10. Bork, P. (1992) *Curr. Opin. Struct. Biol.* **2,** 413–421.
11. Benner, S. A., Gerloff, D. L. & Jenny, T. F. (1994) *Science* **265,** 1642–1644.
12. Benner, S. A., Cohen, M. A., Gonnet, G. H., Berkowitz, D. B. & Johnsson, K. (1993) in *The RNA World*, eds. Gesteland, R. & Atkins, J. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 27–70.
13. Darland, G. & Brock, T. D. (1970) *Science* **170,** 1416–1418.
14. Matsudaira, P. (1990) *Methods Enzymol.* **182,** 602–613.
15. Yanish-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
16. Innis, M. A. & Gelfand, D. H. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic, New York).
17. Benner, S. A., Badcoe, I., Cohen, M. A. & Gerloff, D. L. (1994) *J. Mol. Biol.* **235,** 926–958.
18. Sando, G. N. & Hogenkamp, H. P. C. (1973) *Biochemistry* **12,** 3316–3322.
19. Gleason, F. K. & Frick, T. D. (1980) *J. Biol. Chem.* **255,** 7728–7733.
20. Abeles, R. H. & Beck, W. S. (1967) *J. Biol. Chem.* **242,** 3589–3593.
21. Hain, J., Reiter, W.-D. & Zillig, W. (1992) *Nucleic Acids Res.* **20,** 5423–5428.
22. Ree, H. K., Cao, K., Thurlow, D. L. & Zimmermann, R. A. (1989) *Can. J. Microbiol.* **35,** 124–133.
23. Sutherland, K. J., Henneke, C. M., Towner, P., Hough, D. W. & Danson, M. J. (1990) *Eur. J. Biochem.* **194,** 839–844.
24. Uhlin, U. & Eklund, H. (1994) *Nature (London)* **370,** 533–539.
25. Benner, S. A. (1996) in *Protein Engineering: A Guide to Design and Production*, eds. Craik, C. S. & Cleland, J. (Wiley–Liss, New York), pp. 71–99.
26. Benner, S. A., Chelvanayagam, G. & Turcotte, M. (1996) *Chem. Rev.*, in press.
27. Gonnet, G. H., Cohen, M. A. & Benner, S. A. (1992) *Science* **256,** 1443–1445.
28. Eriksson, S., Sjöberg, B.-M., Jörnvall, H. & Carlquist, M. (1986) *J. Biol. Chem.* **261,** 1878–1882.
29. Drennan, C. L., Matthews, R. G. & Ludwig, M. L. (1994) *Curr. Opin. Struct. Biol.* **4,** 919–929.
30. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. (1994) *Science* **266,** 1669–1674.
31. Chothia, C. & Lesk, A. (1986) *EMBO J.* **5,** 823–826.
32. Benner, S. A. & Ellington, A. D. (1990) *Bioorg. Chem. Front.* **1,** 1–70.
33. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. (1995) *Nature (London)* **374,** 57–59.
34. Daniel, R. M. & Danson, M. J. (1995) *J. Mol. Evol.* **40,** 559–563.
35. Eschenmoser, A. (1988) *Angew. Chem. Int. Ed. Engl.* **27,** 5–39.
36. Kreitman, M. & Akashi, H. (1995) *Annu. Rev. Ecol. Syst.* **26,** 403–422.
37. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
38. Gould, S. J. (1980) *The Panda's Thumb* (Norton, New York).
39. Philipps, G., Clément, B. & Gigot, C. (1995) *FEBS Lett.* **358,** 67–70.
40. Yang, F., Lu, G. & Rubin, H. (1994) *J. Bacteriol.* **176,** 6738–6743.
41. Jordan, A., Gibert, I. & Barbé, J. (1994) *J. Bacteriol.* **176,** 3420–3427.
42. Jordan, A., Gibert, I. & Barbé, J. (1995) *Gene* **167,** 75–79.
43. Sjöberg, B.-M. (1994) *Structure (London)* **2,** 793–796.
44. Gerloff, D. L., Chelvanayagam, G. & Benner, S. A. (1995) *Proteins Struct. Funct. Genet.* **23,** 446–453.
45. Benner, S. A. (1989) *Adv. Enzyme Regul.* **28,** 219–236.