Genomics & Informatics Vol. 8(4) 170-176, December 2010

Comparative Analysis of the Three Classes of Archaeal and Bacterial Ribonucleotide Reductase from Evolutionary Perspective

Meenal G. Pangare and Sathees B. Chandra*

Department of Biological, Chemical and Physical Sciences, Roosevelt University, Chicago, USA

Abstract

The Ribonucleotide reductases (RNR) are essential enzymes that catalyze the conversion of nucleotides to deoxynucleotides in DNA replication and repair in all living organisms. The RNRs operate by a free radical mechanism but differ in the composition of subunit, cofactor required and regulation by allostery. Based on these differences the RNRs are classified into three classes- class I, class II and class III which depend on oxygen, adenosylcobalamin and S-adenosylmethionine with an iron sulfur cluster respectively for radical generation. In this article thirty seven sequences belonging to each of the three classes of RNR were analyzed by using various tools of bioinformatics. Phylogenetic analysis, dot-plot comparisons and motif analysis was done to identify a number of differences in the three classes of RNRs. In this research article, we have attempted to decipher evolutionary relationship between the three classes of RNR by using bioinformatics approach.

Keywords: classes, evolution, Ribonucleotide reductase

Introduction

The Ribonucleotide reductases are enzymes with a complex structure which are present in all cellular organisms: bacteria, viruses, eukaryotes and archaea. The Ribonucleotide reductase catalyses the reduction of ribonucleotide diphosphate to deoxyribonucleoside, which are precursors for the essential steps for DNA repair and replication (Jordan and Reichard, 1998). The ribonucleotide reductases are divided into 3 classes- I, II and III. All the 3 classes carry out the reduction reaction by free radical chemistry and involve proteins with free radical amino acids. However the metallocofactor required for initiation of the reduction reaction and requirement for oxygen varies (Sjöberg B-M, 1997).

Class I RNRs are aerobic and function strictly in the presence of oxygen for the production of tyrosyl radical by a di-iron center. *Escheriachia Coli* was the microorganism from which the first reductase was characterized and has become the prototype of class I (Fontecave *et al.*, 1992). The *E. coli* class I reductase is a hetero tetramer ($\alpha 2 \beta 2$) made up of two homodimer protein called NrdA ($\alpha 2$) and NrdB ($\beta 2$) (Nordlund and Reichard, 2006; Torrents *et al.*, 2007). The $\beta 2$ polypeptide contains the tyrosyl radical. The class I RNR is further subdivided into Ia and Ib class on the basis of allostery and identity of sequence. Class Ia is encoded by *nrdAB* gene and class Ib is encoded by the *nrdEF gene*. Class II RNR is best characterized from *Lactobacillus leichmannii* (Blakley, 1978; Booker & Stubbe, 1993; Panagou *et al.*, 1972). The class II RNRs consists of a single subunit (mostly $\alpha 2$ homodimer) encoded by *nrdJ* genes and require adenosylcobalamin (AdoCbl) for radical generation. This process does not require oxygen. Hence class II RNR can work under aerobic as well as anaerobic conditions (Nordlund and Reichard, 2006; Torrents *et al.*, 2007). The class III RNR generate glycyl radical by using S-adenosylmethionine (SAM) as a cofactor and NrdG protein as an activator. It consists of $\alpha 2 \beta 2$ heterodimer which is encoded by the *nrdDG* genes. The class III is represented by the anaerobic RNR of *E. coli* (Ollagnier *et al.*, 1996). The class I RNR genes are encoded by most Eukaryotic organisms whereas a few bacteria and archaea have genes encoding two or all three RNRs (Jordan *et al.*, 1999). All three RNRs operate by a radical mechanism. Also in all three RNRs each of the four ribonucleotides are reduced by a single enzyme. The substrate specificity is determined by allosteric effect due to binding of nucleoside triphosphate to a particular site which is different from the catalytic site. This specific site binding is a property is unique to RNRs (Jordan and Reichard, 1998). The class Ia RNRs also has an activity site which controls the overall activity of the enzyme using ATP regardless of the specificity. Considering the different metallocofactors required by different classes of RNRs it seems that the three classes of RNRs evolved independently.

Despite these differences the similar catalytic mechanism of all the three classes of RNR and presence of

*Corresponding author: E-mail schandra@roosevelt.edu
Tel +1-847-619-7968, Fax +1-847-619-8555
Accepted 1 November 2010

ribonucleotide seems to suggest that they evolved from a common ancestor (Reichard, 1997; Stubbe *et al.*, 2001). Also the evolution of different enzymes at different times brings forward a possibility that the original RNR is a ribozyme and not a protein (Benner *et al.*, 1989).

The evolution of the three classes of Ribonucleotide reductase has always been an enigma with a few suggesting convergent evolution whereas a few suggesting divergent evolution. In this article, we analyze the sequences of the three classes of RNRs, to answer this question, using a number of bioinformatics tools.

Methods

The RNR sequences for analysis were obtained from the NCBI website (http://www.ncbi.nlm.nih.gov/). The class of each of RNR sequence was reconfirmed by using RNR db a specialized database for RNR, which can be accessed at http://rnrdb.molbio.su.se. Out of several available RNR sequences, we selected only the bacterial and archaeal sequences and excluded viruses and eukaryotes. A total of 37 sequences representing the core enzyme of the three classes of RNRs were meticulously chosen based on their size, genetic composition, radical chemistry, cofactor requirement and organism origin. We obtained fourteen sequences of bacteria belonging to Class I- seven sequences representing class Ia and seven sequences representing class Ib. We obtained ten sequences of class II which had both bacterial as well as archaeal sequences. Class III was represented by thirteen bacterial sequences. Until now, various hypotheses regarding the evolution of three classes of RNR have been proposed. Here, for the first time we tried to analyze the evolution of the 3 classes of RNRs using bioinformatics.

MEME Motif discovery tool was used to identify the similar motifs in each of the sequences. All the settings were set to default, except for the maximum number of Motifs which was increased from three to ten (Bailey *et al.*, 2006). The phylogenetic tree and the motif analysis were then used to construct dot plots. The position of

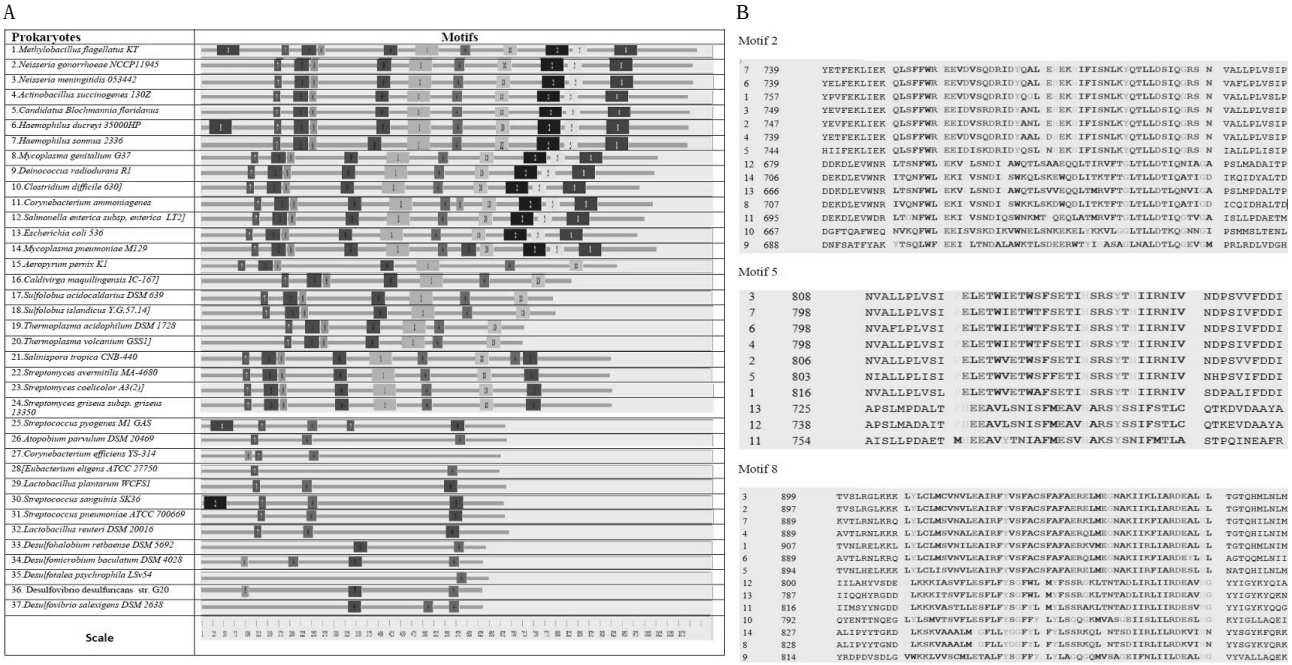

Fig. 1. (A) The motifs present in all 37 sequences of all the three classes of RNRs are shown in the table above. To the left are the names of all archaea and bacteria whose RNR sequences were taken for analysis. Except for sequences 15 through 20 which are obtained from archaea rest all belong to bacteria. Sequences 1 through 7 represent class Ia, 8 through 14 class Ib, 15 through 24 class II and 25 through 37 class III. The numbers corresponding to the archaeal and bacterial sequences in above table are also used in analysis in (B). (B) The sequences of motifs 2, 5 and 8 are shown in the figure above. Beginning from left the first column shows the number corresponding to the name and number of the bacteria/archaea in table of (A). The second number represents the sequence length in RNR enzyme and the third column shows the sequence. So in Motif 2 the first number is 7 which represents bacterium *Haemophilus somnus* 2336 with a sequence length of 739 amino acids which is shown in the sequence.

a specific amino acid motif in the selected protein sequence was found by dot plots.

In order to compare the similarity as well as difference in the sequences of each class of RNR the dot matcher program was used to construct dot plots. The similarity in the protein sequences can be easily assessed from dot plots simply by seeing a diagonal fragment in between the X and Y axis of a graph, which is constructed by using data matrix, distance matrix and chi squared analysis (Landes *et al.*, 1998). Thus similar sequence show a diagonal line whereas this line is absent or highly fragmented in dissimilar sequences. We first constructed dot plots by using sequences belonging to the same class of RNR and then by using each sequence from a different class, using different combinations of class each time. The parameters of the program were mostly set at default except for window size of 10 and a threshold of 23 (Rice and Longden, 2000).

The selected sequences were obtained in FASTA format and then aligned by using Clustal X (Thompson *et al.*, 1997). Neighbour joining method was used to construct the phylogenetic tree from the sequences which were aligned using PHYLIP (Felsenstein, 1989). The phylogenetic tree was then bootstrapped in order to see how well the sequences related to each other. Finally treeview was used to see their position in each clade and study if the RNR sequences were related by evolution (Page *et al.*, 1996).

Results

Analysis of RNR motifs in class I, class II and Class III of RNR reveals the distinct features in all the three classes of RNRs (Fig. 1). The RNR of class Ia and Ib share Motif 2 and Motif 8. These two motifs are present in sequences 1 (*Methylobacillus flagellatus KT*) through 14 (*Mycoplasma pneumoniae M12*). Motif 5 is present in all bacteria of only class Ia and three bacterial sequences of class Ib:11 (*Corynebacterium ammoniagenes*), 12 (*Salmonella enterica subsp. enterica serovar Typhimurium str. LT2*) and 13 (*Escherichia coli 536*). Motif 1, 3, 6 and 10 are present in sequences 1 (*Methylobacillus flagellatus KT*) through 24 (*Streptomyces griseus subsp. griseus NBRC 13350*) which incorporates all the sequences from class I and class II. Motif 7 and Motif 9 are present in sequences belonging to class I, class II and a few sequences of class III. Motif 7 is present in sequences from 1 (*Methylobacillus flagellatus KT*) through 32 (*Lactobacillus reuteri DSM 20016*) with the exception of sequence 27 (*Corynebacterium efficiens YS-314*) whereas Motif 9 is seen in sequences 1 through 25 (*Streptococcus pyogenes M1 GAS*) except sequence 14, sequences 30 (*Streptococcus sanguinis SK36*) and sequence 31 (*Streptococcus pneumoniae ATCC 700669*). However, there is no motif that is common to all members belonging to class III RNR.

The degree of similarity between the protein sequences of the three classes of RNR's can be easily seen by comparing the dot plots in Fig. 2. The dot plots

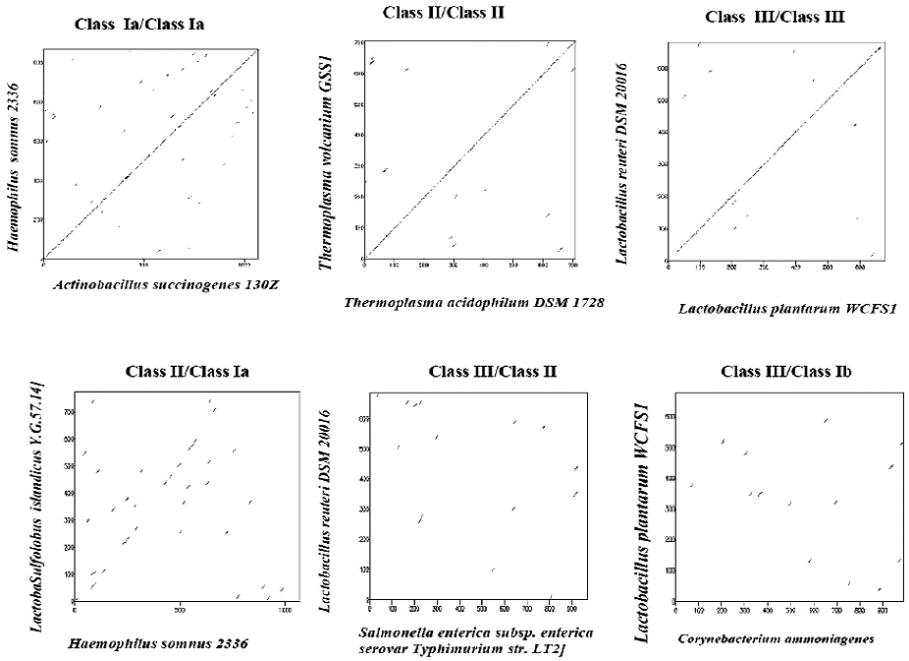

Fig. 2. The dot plot comparison of sequences belonging to the same class resulted in a collinear diagonal fragment. The dot plots are within same class of RNRs such as class Ia /class Ia, class II /class II and class III/ class III. Dot plot comparison using RNRs from different classes such as class II/class Ia, class III/classII and class III/classIb resulted in a plot with numerous non collinear fragments.

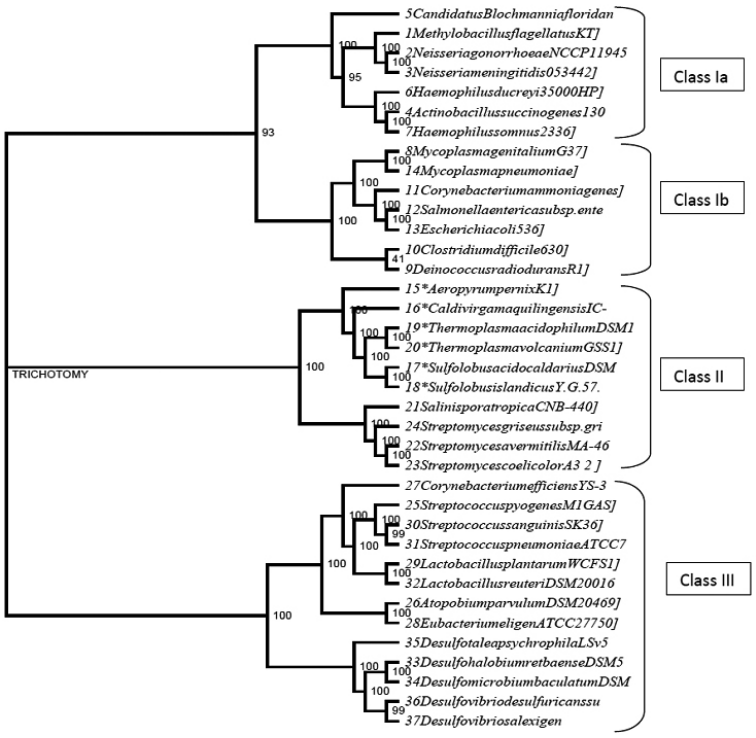

Fig. 3. A Phylogenetic tree constructed by using amino acid sequences belonging to all the three classes of RNRs. The name of the bacteria/archaea and the class to which it belongs is seen on the right side of the cladogram. The scores seen on the tree show the sequence similarity between the sequences of each class. The sequences with asterisk belong to archaea whereas the sequences without asterisk belong to bacteria. The names of bacteria/archaea along with accession numbers in parentheses are as follows. The numbers from 1 through 7 represent class Ia:1] *Methylobacillus flagellatus KT* (YP_545332.1), 2] *Neisseria gonorrhoeae* NCCP11945 (YP_002001927.1), 3] *Neisseria meningitidis 053442* (YP_001599335.1), 4] *Actinobacillus succinogenes 130Z* (YP_001344760.1), 5] *Candidatus Blochmannia floridanus* (NP_878761.1), 6] *Haemophilus ducreyi 35000HP* (NP_874098.1), 7] *Haemophilus somnus 2336* (YP_001783715.1). Class Ib RNR are represented by Arabic numerals 8 through 14: 8] *Mycoplasma genitalium G3* (NP_072897.1), 9] *Deinococcus radiodurans R1* (NP_051640.1), 10] *Clostridium difficile 630* (YP_001089513.1), 11] *Corynebacterium ammoniagenes* (GenBank: CAA70765.1),12] *Salmonella enterica subsp. enterica serovar Typhimurium str. LT2* (NP_461733.1), 13] *Escherichia coli 536* (YP_670530.1), 14] *Mycoplasma pneumoniae M129* (NP_110012.1). The numbers 15 through 24 show RNR class II sequences. 15] **Aeropyrum pernix K1* (NP_148354.2), 16] **Caldivirga maquilingensis IC-167* (YP_001540469.1), 17] **Sulfolobus acidocaldarius DSM639* (YP_255981.1), 18] **SulfolobusislandicusY.G.57.14* (YP_002837466.1), 19] **Thermoplasma acidophilum DSM 1728* (NP_394926.1), 20] **Thermoplasma volcanium GSS1]* (NP_110611.1), 21] *Salinispora tropica CNB-440* (YP_001158297.1), 22] *Streptomyces avermitilis MA-4680* (NP_823637.1), 23] *Streptomyces coelicolor A3(2)* (NP_629929.1), and 24] *Streptomyces griseus subsp. griseus NBRC 13350* (YP_001823227.1). Class III RNR are represented from number 25 through 37. 25] *Streptococcus pyogenes M1 GAS* (NP_270034.1), 26] *Atopobium parvulum DSM 20469* (YP_003180143.1), 27] *Corynebacterium efficiens YS-314* (NP_738973.1) 28] *Eubacterium eligens ATCC 27750* (YP_002931010.1), 29] *Lactobacillus plantarum WCFS1* (NP_786274.1), 30] *Streptococcus sanguinis SK36* (YP_001036145.1), 31] *Streptococcus pneumoniae ATCC 700669* (YP_002510259.1), 32] *Lactobacillus reuteri DSM 20016* (YP_001271911.1), 33] *Desulfohalobium retbaense DSM 5692* (YP_003196948.1), 34] *Desulfomicrobium baculatum DSM 4028* (YP_003159885.1), 35] *Desulfotalea psychrophila LSv54* (YP_063800.1), 36] *Desulfovibrio desulfuricans subsp. desulfuricans str. G20* (YP_386773.2), 37] *Desulfovibrio salexigens DSM 2638* (YP_002993099.1).

constructed with sequences belonging to same class shows a linear graph whereas dot plots constructed in between two classes show a high degree of dissimilarity. Dot plots constructed even within two subclasses of the same class: class Ia and class Ib resulted in non-collinear multiple fragments. These findings match with the motifs discovered which are shared only between a particular class.

As seen in Fig. 3 below phylogenetic analysis of the three classes of RNRs resulted in the formation of a tree with three distinct clades for class I, class II and class III RNRs. Class I is further subdivided into two clades:

class Ia and class Ib. The bootstrap value for class I, class II and class III RNRs are mostly above ninety and from the tree we can infer that all the three classes of RNR share a common ancestor. As seen in Fig. 3, the lower bootstrap score is only seen between *Clostridium difficile 630* and *Deinococcus radiodurans* R1 both belonging to RNR class Ib. *Clostridium difficile 630* by its spore forming ability is highly resistant whereas *Deinococcus radiodurans R1* is radio resistant. These bacteria have adapted themselves to new environments which other bacteria and archaea were not able to propagate and this fact is reflected in the RNR sequence of these bacteria and the evolutionary distance in the phylogenetic tree.

Discussion

The three classes of RNR share a common ancestor and each class evolved independently from each other before the tree of life diversified (Torrents *et al.*, 2002). The multiple sequence alignment which was performed on 37 bacterial and archaeal sequences representing all 3 classes of RNRs gives sufficient evidence to conclude that they all have a common evolutionary origin. The phylogenetic tree shows three distinct clades for each class of RNR, thus providing a definitive evidence of common ancestral origin of all classes of RNRs with divergent evolution. All the three classes of RNRs synthesize deoxyribonucleotides by catalyzing the disruption of 2' carbon-hydroxyl bond to form 2'carbon hydrogen bond. The solvent provides the hydrogen, which replaces the hydroxyl while retaining the overall configuration (Licht *et al.*, 1999).Using the prototype for class I (*Escheriachia Coli*) and Class II (*Lactobacillus leichmannii*) enzymes the radical mechanism of ribonucleotide reduction can be described in brief as follows ((Blakley, 1978; Booker & Stubbe, 1993; Fontecave *et al.*, 1992; Panagou *et al.*, 1972). The Cys439 of the *E. coli* or Cys408 of the *L. leichmannii* has a thiyl radical. The thiyl radical creates an active substrate radical by removing hydrogen from C-3'. Two redox active cysteines (Cys 225 and Cys462 in *E. coli* and Cys119 and Cys 419 in *L. leichmannii*) reduce the active substrate radical (Reichard, 1997; Stubbe *et al.*, 2001). The thiyl radical is generated from the tyrosyl radical in presence of oxygen linked di iron center in class I RNR and from adenosylcobalamin in class II RNR. Class III RNR has glycyl radical as stable radical on the large subunit and has 4Fe-4S cluster and requires S-adenosylmethionine for radical generation (Ollagnier *et al.*, 1996). Indirect evidence suggesting participation of cysteines and thiyl radical in reduction of the ribose ring in class III RNR also occurs (Uhlin *et al.*, 1994). So similar catalytic mechanism suggests a common ancestory. The level of deoxyribonucleotides is maintained by regulation of RNRs by allosteric as well as transcription regulatory mechanism. Allosteric control is achieved by binding of nucleoside triphosphates to a specificity site on the substrate which is distinct from the active site along with binding of ATP or dATP to active site causing stimulation or inhibition respectively, and thus controlling the overall activity of the enzyme (Nordlund and Reichard, 2006).

In the past few years NrdR, which was first described in *Streptomyces coelicolor* (Borovak *et al.*, 2004) is being suggested as a novel global regulator controlling the expression of *nrd* genes in all three classes of RNRs. It has been reported that in a variety of microorganisms a putative NrdR box is present upstream of nrd genes from all three classes of RNRs (Rodionov *et al.*, 2005). NrdR has an ATP cone which is similar to the RNR class I and III allosteric activity site; suggesting intracellular ATP/dATP pools acting as a stimulus for NrdR for regulation of *nrd* expression (Aravind *et al.*, 2000). The analysis of the different motifs show that class I has unique motifs 2 and 8 which are absent in the two other classes of RNRs. Class I and II do share Motifs 1, 3, 6 and 10 providing evidence that class I and II are more closely related than class III. Also, a unique motif present in only class III RNR sequences is non- existent. Motif 5 is peculiar as it is present in all bacteria and archaea belonging to class Ia. However it is present in only sequences 11 through 13 of class Ib and excludes sequences 8, 9, 10 and 14. *Deinococcus radiodurans R1* and *Clostridium difficile 630* are highly resistant organisms and probably evolved and adapted to certain environments which other bacteria and archaea were not able to propagate and hence lack motif 5. The same fact is further highlighted by low bootstrap scores of these two bacteria in phylogenetic analysis. *Mycoplasma genitalium G3* and *Mycoplasma pneumoniae M129* have *nrdF* sequences that have a Val, Pro and Lys instead of Glu98, Glu157 and Glu192 respectively (Eriksson *et al.*, 1998). This explains absence of Motif 5 in these two bacteria.

The dot-plots which were used as a comparative tool between two sequences showed a high degree of similarity within members of the same class. However dot plots between members of different classes just resulted in multiple fragments without solid collinear lines suggesting no similarity between sequences of different classes. These results are consistent even with Motif analysis. In conclusion, we can say that the three classes of RNRs have a common ancestor yet they evolved separately. In this study by comparing the organization of motifs, the protein sequences, dot plots

an ancient enzyme. *J. Mol. Evol.* 55, 138-152.

Torrents, E., Grinberg, I., Gorovitz-Harris, B., Lundstrom, H., Borovok, I., Aharonowitz, Y., Sjöberg, B.M., and Cohen, G. (2007). NrdR controls differential expression of the Escherichia coli ribonucleotide reductase genes. *J. Bacteriol.* 189, 5012-5021.

Uhlin, U., and Eklund, H. (1994). Structure of ribonucleotide reductase protein R1. *Nature* 370, 533-539.