TUTORIAL REVIEW

CSR
www.rsc.org/csr

Deoxyribonucleotides: the unusual chemistry and biochemistry of DNA precursors

Hartmut Follmann
Fachbereich Chemie-Biologie der Universität, 34109 Kassel, Germany.
E-mail: hfollman@uni-kassel.de

Received 29th October 2003
First published as an Advance Article on the web 31st March 2004

Deoxyribonucleotides, monomers of macromolecular DNA and the chemical matter of genes, have received surprisingly little attention among chemists and molecular biologists alike, although their origin, properties, and mechanism of enzyme-catalyzed formation bear unique chemical traits which are the basis of DNA replication. Apart from providing insights in bioorganic free radical chemistry, present interest in deoxyribonucleotides stems from the expected demand of hundreds of kilograms per year for DNA chips and antisense constructs used in gene therapy, difficult to produce by conventional methods. A novel approach towards deoxyribonucleotide, and hence DNA formation in a putative primordial 'RNA world' has also recently emerged.

1 Introduction

Deoxyribonucleic acid, the giant macromolecule encoding the genetic information of all living cells, with its famous double-helical structure is one of the most spectacular and thoroughly analyzed chemical substances ever described. Indeed, it has become an icon of modern science. The existence of DNA, its replication, and nowadays, deliberate modification and manipulation of its genetic message *in vitro* appear to be of such overwhelming scientific and public interest that the origin and nature of deoxyribonucleotides, its building blocks, are simply neglected by most geneticists and other life scientists. It is considered that these monomers are available intracellularly at all times, and for *in vitro* experiments they come in the convenient deoxyribonucleotide kits found in any molecular biology laboratory.

The neglect of these fundamentals is puzzling. Obviously, DNA cannot replicate without a balanced supply of monomers. This restriction is easily recognized and experimentally exploited in cell cycle and cell proliferation research where deoxyribonucleotide formation is a prominent check point for onset of the S (DNA synthesis) phase.[1–3] The purine and pyrimidine ribonucleotides, found abundantly in energy-rich nucleotides (ATP, GTP), nucleotide coenzymes, and in RNA, are synthesized from ribose phosphate whereas in contrast their congeners, the 2-deoxyribose or 2-deoxyribose 5-phosphate components of DNA are not among the carbohydrate species of metabolic pathways. Moreover 2′-deoxyribonucleoside 5′-triphosphates (dNTPs), the substrates of DNA polymerases during DNA replication, are only present in minute amounts, if at all, in pools of soluble nucleotides extracted from microorganisms, or plant and animal tissues.[3,4]

This review outlines the specific chemistry and enzymology of deoxyribonucleotides. The future biotechnological potential of the universal, yet diversified and easily disturbed biosynthetic pathway will be assessed. Finally, biomimetic considerations that have led us to a model of abiotic deoxyribonucleotide formation and the emergence of DNA in earliest replicating cell-like systems are discussed.

2 2-Deoxy-D-ribose and its derivatives

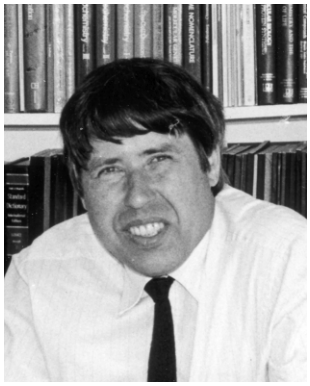
Unlike arabinose, ribose and xylose and their derivatives, 2-deoxyribose (2-deoxy-D-*erythro*-pentose) is a not an intermediate of hexose and pentose metabolism. It was isolated from hydrolysates of thymus DNA and identified as a previously unknown sugar ("thyminose") no less than 60 years after the discovery of DNA[5,6] because of its inherent instability during chemical hydrolysis which leaves only stable heterocyclic bases intact. In early syntheses[6,7] 2-deoxyribose was produced from D-ribose or D-arabinose by a glycal reaction in which water is added to the olefinic bond in D-arabinal, from D-erythrose by condensation with nitromethane (the Nef reaction) followed by reduction and hydrolysis of the ensuing 1-nitropentenose, or from D-glucose by alkaline degradation, generally in low yield. In a recent synthetic route to thymidine (*vide infra*) the deoxy sugar is obtained in seven steps with about 20% overall yield from D-glucose *via* 3-deoxy-D-glucose and Wohl degradation.

The only enzyme known to act on 2-deoxy-D-ribose (as its 5-phosphate) is a widely distributed lyase, deoxyribose phosphate aldolase (EC 4.1.2.4). It catalyses the reversible reaction [eqn. (1)]

2-deoxyribose-5-phosphate ⇆ acetaldehyde + D-glyceraldehyde-3-phosphate (1)

A bacterial enzyme preparation can conveniently be used for synthesis of deoxyribose-5-phosphate *in vitro*.[8] However, the *in vivo* function of deoxyribose phosphate aldolase is considered

Hartmut Follmann was born in 1936. He studied chemistry at the University of Marburg, Germany, and received his PhD with Karl Dimroth in 1964. From 1967 to 1970 he worked as a postdoctoral associate (with H.P.C. Hogenkamp) and instructor at the Biochemistry Department, Medical School, University of Iowa at Iowa City, USA. In 1972 he became professor of biochemistry in Marburg, and in 1988 moved to the newly established chair of biochemistry at the University of Kassel. He spent sabbatical years at the University of California at Berkeley (1979) and the University of Minnesota at Minneapolis (1991). His research activities center on the diversity and functions of the multiple redox proteins, thioredoxins and glutaredoxins, and the enzymology and metal requirement of ribonucleotide reduction for DNA precursor biosynthesis. Methylcytosine formation in plant DNA has been an additional research topic. Moreover, Hartmut Follmann has taken a strong interest in teaching and is the author of several laboratory course books.



DOI: 10.1039/a907361b

exclusively catabolic, *e.g.*, in the breakdown of DNA fragments and spent deoxyribonucleotides, as millimolar concentrations of free acetaldehyde would be required for the enzyme-catalyzed condensation reaction which do not occur intracellularly.

The significant and indispensable biochemical role of 2-deoxyribose lies in its geometrically *and* chemically favourable pentofuranose structure linking the heterocyclic bases, *via* *N*-glycosidic bonds, and the internucleotide 3′-5′ phosphodiester groups along a polynucleotide strand. The absence of a hydroxy group at deoxyribose atom C2, situated between the bases and the 3′-phosphate residues, renders DNA distinctly more stable towards chemical and enzymatic hydrolysis than RNA, where the C2′-OH facilitates chemical and enzymatic hydrolysis of phosphodiester bonds by neighbouring group participation. Given that two kinds of nucleic acids, of differing stabilities, were a major advantage for selection and functioning of high-fidelity genetic systems from the beginning, but that only ribose and ribonucleotides were likely constituents of primordial organic chemistry, a productive source of 2-deoxyribose must have been a key step of early molecular evolution in primitive, replicating cellular entities.

This condition appears difficult, if not impossible to meet by abiotic sugar chemistry. A wide variety of carbohydrates including ribose and other pentoses is formed by self-condensation of formaldehyde in the presence of basic catalysts in the long-known "formose reaction" , or in a benzoin-type condensation under thiazolium ion catalysis ("formoin reaction").[9] In contrast no trace of deoxyribose has been identified in these reaction mixtures. However, even *if* some 2-deoxyribose had been formed in an, as yet unexplored, specific abiotic reaction the sugar might not have yielded functional nucleic acids. Due to its –CHOH–CH_2–CHO structure deoxyribose is more reactive and less stable in solution at elevated temperatures than ribose. Acids induce dehydration and dismutation to levulinic acid, and amines catalyze dehydration to the α,β-unsaturated derivative, 2,3-dideoxy-2,3-dehydroribose or D-4,5-dihydroxy-2-pentenal (Fig. 1). Most importantly, facile addition of purines or pyrimidines to the double bond of the latter leads to the formation of 2,3-dideoxy-3-(9′-purino)- or 2,3-dideoxy-3-(1′-pyrimidino) pentoses in up to 90% product yields.[10] Isomerization of these artificial compounds to the isomeric natural deoxyribonucleosides carrying the bases at C1 does not occur. Even if formed in a "primordial broth", deoxyribose and its unsaturated derivative would thus have been scavenged by most other reactive components present. One may safely conclude that the free sugar is a most unlikely source of DNA monomers.

3 Reductive formation of deoxyribose from ribose nucleotides

It is the enzymology of its precursors that provides a convincing rationale for the existence of DNA. Deoxyribonucleotides are not built from deoxyribose phosphates, analogous to the common purine and pyrimidine anabolic pathways, but they originate through enzyme-catalyzed specific deoxygenation at C2′ of the abundantly available adenine, guanine, cytosine, and uracil ribonucleotides. Thymine, the pyrimidine base replacing uracil in DNA is formed only subsequently by methylation of deoxyuridylate, catalyzed by thymidylate synthase (EC 2.1.1.45) which is tightly coordinated with ribonucleotide reduction during the cell cycle. Once formed by reduction of a ribonucleotide, deoxyribose escapes the degradation reactions described above as the sugar is locked in stable *N*-glycosidic linkage at C1′. Nevertheless, the process poses a mechanistic problem in that direct displacement of a secondary sugar hydroxy group by hydrogen is not a trivial transformation. In fact a chemical counterpart of the enzyme-catalyzed reaction utilizing thiol functions for reducing equivalents (*vide infra*) in neutral aqueous media has been missing for decades.

3.1 The ribonucleotide reductase metalloenzyme family

All organisms rely on *de novo* biosynthesis of their DNA precursors. DNA fragments from decaying biomass, ubiquitously present in soil and water, or taken up in food, are not generally re-used. The exception is in some "salvage pathways" that enable rescue and re-incorporation of purine or pyrimidine bases and thymidine into nucleic acids in some bacteria and in mammalian cells.[11] Therefore, ribonucleotide reductases (RNR, EC 1.17.4.),[11–13] discovered by Peter Reichard in the 1950s,[14] are universal and essential enzymes which, together with the DNA polymerases and thymidylate synthase, are essential to DNA replication in all living cells, and also for viruses with large DNA genomes.

Ribonucleotide reductases catalyze reduction of the C2′ hydroxy group in ribonucleoside 5′-diphosphates or, more rarely, 5′-triphosphates [eqn. (2)]. Dicysteinyl proteins of the thioredoxin/glutaredoxin superfamily of small redox proteins serve as natural hydrogen donors, or dithiols of low molecular weight such as dithiothreitol or dihydrolipoate can function *in vitro*.

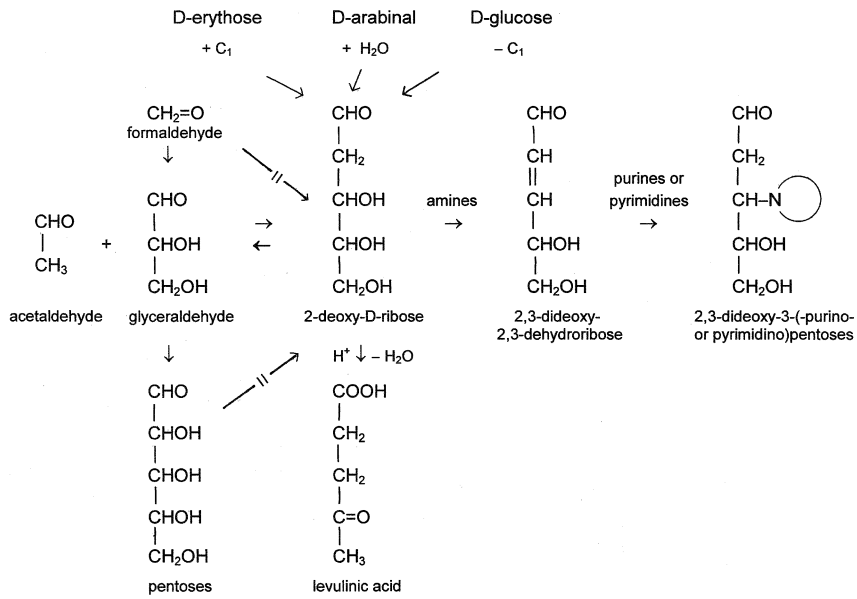


Fig. 1 Modes of formation and transformation of 2-deoxyribose in aqueous solution,[6] pertaining to chemical syntheses (top, *cf.* Fig. 7) and imaginable in prebiotic chemistry.

ribonucleotide(–OH) + $R(SH)_2$ → deoxyribonucleotide(–H) + RS_2 + H_2O (2)

The OH group is eliminated as water and is stereospecifically replaced by H under retention of configuration at C2′. The incoming hydrogen is ultimately derived from water due to rapid equilibration between the SH functions and H_2O, evidenced by formation of specifically deuterated or tritiated products when reduction is carried out in ^2H_2O or ^3H_2O, respectively.[15] Based on RNR-catalyzed deuterium incorporation from heavy water into newly formed DNA a non-toxic, non-mutagenic method has recently been designed for measuring DNA replication *in vivo* that is suitable for slow turnover cells and safe for use in humans.[16]

Reaction (2) is thermodynamically favorable ($\Delta G^{0\prime} = -70$ kJ mol^{-1}, at pH = 7, calculated from Gibbs free energies of formation), and it is irreversible since a disulfide is not able to act as hydroxylating agent in the reverse reaction. (Conversion of thymidine to thymine ribonucleoside, unique to fungal nucleotide metabolism, is catalyzed by a separate 2′-hydroxylase requiring oxygen, Fe^{2+} ions, and 2-oxoglutarate as cofactors.[17]) RNRs are highly specific for D-ribotides and do not tolerate methylation or other modifications of the sugar, nor do they act on D-arabinose- or D-xylose-derived nucleotides.[12,19] In contrast they are non-specific with respect to the base moiety of their substrates and will reduce all four common purine and pyrimidine ribonucleotides as well as base-modified nucleotides of natural or synthetic origin;[12] the latter capacity may not be of physiological significance. Kinetic analyses are, however, complicated by the enzymes' multiple allosteric sites which modulate reduction rates of the different substrates in the presence of deoxyribonucleotide products and ATP, that act as feedback activators or inhibitors.[2–4,24] This property (not dealt with here) is thought to guarantee a balanced supply of the four DNA polymerase substrates *in vivo*.

The mechanism of enzymatic ribonucleotide reduction has puzzled biochemists for decades and for a while was aptly characterized as "amazing and confusing".[18] An impressive set of differently substituted and isotope-labelled nucleotides were synthesized but failed to identify classical intermediates and substitution or elimination–addition pathways for the selective removal of the 2′-OH group (Fig. 2).[12] Thus, cytidine 2′-phosphate-5′-

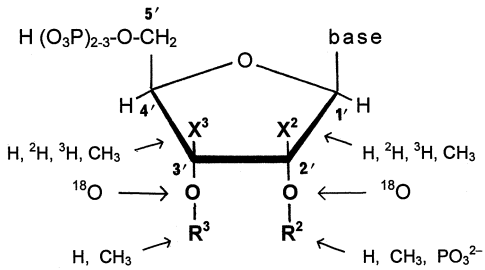


Fig. 2 Structures of 2′- and 3′-modified ribonucleotides (boldface) used in studies of RNR mechanism and specificity. Syntheses are described in[19,20,25] Base modifications are not included.

diphosphate, providing a better leaving group at C2′, is not an RNR substrate. Elimination of water and reduction of a 2′,3′ double bond, or dehydrogenation to a 2′-oxo function and its subsequent reduction can be excluded as only one labelled hydrogen is stereospecifically incorporated from the solvent in deuterated or tritiated water.[15] In a classic experiment ATP was site-specifically labeled with ^{18}O in the 2′ or 3′ OH groups and incubated with RNR with and without dithiols present.[19] Here the heavy oxygen was fully retained in the 3′ position, and was completely lost from C2′ in the presence but not in the absence of reductant, ruling out reductive opening of a possible 2′,3′ epoxide intermediate concomitant with oxygen migration as well as an S_N1-like cationic species at C2′. Equally important information has come from the study of substrates labeled with deuterium or tritium at C3′-H which exhibited up to 4-fold kinetic isotope effects, suggesting that cleavage of this bond is part of the mechanism.[20] The lack of substrate activity of 3′-*C*-methyl-substituted ribonucleotides[25] has confirmed that conclusion.

Ribonucleotide reductases are metalloenzymes or depend on metal cofactors containing iron, manganese, or cobalt. The transition metal centres do not directly participate in catalysis of the overall reaction (2) but serve to activate the side chains of tyrosine, glycine, and cysteine residues to radical species by hydrogen abstraction. Consequently, the enzymes are better classified by metal and organic free radical content than by gene or protein sequences. Although the chemistry of ribonucleotide reduction is the same throughout (*vide infra*) RNR proteins are unique among the universal, essential enzymes of life in that four biochemically and genetically distinct classes exist in different prokaryotes and in eukaryotic cells. (Their numbering has followed the order of discovery and characterisation.) This fascinating enzyme family has been the subject of numerous reviews,[11–13] and is only described here briefly. A list of the distribution among organisms can be found in a recent summary.[21]

Class I RNR, containing binuclear iron centres. DNA precursor formation *via* ribonucleotide reduction requires iron ions and oxygen in all eukaryotes (fungi, animals, and plants) and in many bacteria, in particular the Gram-negative proteobacteria. Class I RNRs are characteristic for most forms of aerobic life. The molecular structure of the enzyme of *Escherichia coli* is known in great detail.[21,23] It is considered the prototype of this class, including the mammalian enzymes. The proteins, encoded on *nrdAB* (class Ia) or, less frequently, *nrdEF* (class Ib) genes, usually assume an $\alpha_2\beta_2$ polypeptide composition but more highly aggregated states ($\alpha_6\beta_6$) of a mammalian enzyme have been observed in the presence of allosteric effectors.[24] The smaller R2 subunits (β_2) harbour an antiferromagnetically coupled μ-oxo-bridged di-iron-(III) centre and a stable tyrosyl (phenoxyl) free radical nearby.[22] Both these structures, essential for catalysis, are generated from Fe^{2+} ions, oxygen, and a specific tyrosine residue within the protein [eqn. (3)]:

$$3\,Fe^{2+} + Tyr\text{–}OH + O_2 + H^+ \rightarrow Fe^{3+}\text{–}O^{2-}\text{–}Fe^{3+} - Tyr\text{–}O^{\cdot} + H_2O + Fe^{3+} \quad (3)$$

The larger R1 subunits (α_2)[23] combine substrate and several effector nucleotide sites, a dicysteine domain that delivers reducing equivalents, and another essential cysteine residue providing a thiyl radical (*vide infra*).

Class II, coenzyme B12-requiring. Class II enzymes, encoded on *nrdJ*, have simpler monomeric or dimeric (α or α_2) structures. They are widely distributed among microorganisms, predominating in (facultative) anaerobes such as lactobacilli, cyanobacteria, and many archaebacteria. The monomeric RNR of *Lactobacillus leichmannii* is particularly well known.[26,27] The proteins are metal-free but have tight interaction with 5′-deoxyadenosyl cobalamin, the coenzyme form of vitamin B_{12}, inducing homolytic cleavage of its metalloorganic cobalt(III)-carbon bond to generate a Co(II) species and a 5′-deoxyadenosyl radical intermediate which initiate thiyl radical catalysis.[28] Oxygen is neither required by, nor does it adversely affect, class II enzymes.

Class III, FeS cluster/AdoMet-activated. Another anaerobic enzyme system found in bacteria and archaebacteria (*e.g.*, in anaerobically grown *E.coli*,[29] or in strictly anaerobic methanogens) combines a reductase (*nrdD* product), with an α or α_2 structure, and a small activating subunit (β_2, *nrdG* product) carrying an iron-sulfur cluster of the [4Fe-4S] type or, more likely,[30] of an unusual [4Fe-3S] structure plus a fourth still unidentified ligand. The cluster produces 5′-deoxyadenosyl radicals by reduction of *S*-adenosylmethionine (AdoMet) which in turn initiate a glycine radical-thiyl radical chain for substrate reduction. Other characteristics of class

III are rapid inactivation by oxygen and the utilization of formate, rather than dithiols, as the ultimate electron donor.[31]

Class IV, manganese-containing. A number of "high-GC" and "low-GC" Gram-positive aerobic bacteria (*e.g.*, members of the genera *Corynebacterium*, *Arthrobacter*, and *Bacillus*) specifically require Mn^{2+}, but not Fe^{2+} ions for ribonucleotide reduction when grown under physiological conditions, evidenced by incorporation of ^{54}Mn into the R2 subunit of RNR purified from wild-type *C. ammoniagenes* cells, and rapid reactivation of metal-depleted enzyme by manganese.[11,32] This protein, and the related enzyme from *C. glutamicum* have also been over-expressed in the original genetic background.[50] R2 was shown by metal analyses and EPR spectroscopy to contain tightly protein-bound, octahedrally coordinated divalent manganese ions and an organic free radical (likely tyrosyl), both correlated with enzyme activity.[32–34] The R1 and R2 subunits are encoded on the *nrdEF* genes like those of the class Ib proteins described in enterobacteria. Therefore, and because a diferric R2 subunit was obtained after cloning and over-expression of the *C. ammoniagenes nrdF* gene in *E.coli*[35] the nature of these bacterial RNRs as a distinct enzyme class IV has been disputed. The latter experiments, however, are inconclusive as the protein was produced in a heterologous expression system and acquired high activity only after treatment with iron ascorbate *in vitro*. It is well known that metal, including manganese, trafficking and uptake into metalloproteins are crucial among microorganisms and prone to artefacts in unphysiologic, heterologous situations.

The RNR classes are easily arranged in a "natural system"[11] which takes into account the anaerobic or aerobic environments and the differences in deoxyribonucleotide acquisition of present-day life-forms (Fig. 3). The surprising coexistence of four enzyme systems, three of which show little sequence or structural homology, that catalyze the same unique redox reaction has inspired many thoughts about the origins of ribonucleotide reduction and DNA monomers.[12,13,21,36,37] It is most sensible that RNR precursors arose polyphyletically. The anaerobic class II and III enzymes assembled by a modular combination of nucleotide-binding and cysteine-containing peptides together with different deoxyadenosyl radical-providing cofactors and reductants. With the rise of oxygen in the Earth's atmosphere precursors of aerobic class I and IV RNRs integrated still other transition metal centres suitable for oxidative thiyl radical generation.

3.2 The radical mechanism of enzymatic ribonucleotide reduction

A widespread involvement of organic free radicals, as stable species or transient intermediates, in redox enzyme mechanisms has been recognized in recent years.[39] This has arisen largely from the identification of catalytically competent amino acid residues by site-directed mutagenesis and application of rapid EPR techniques. However the presence of a radical chain in RNR, first inferred in *E. coli* 30 years ago, still represents one of the most complex cases.[40]

The mechanism of RNR catalysis has been analyzed proficiently in the laboratories of P. Reichard, H. P. C. Hogenkamp, B.-M. Sjöberg and J. Stubbe. The accepted sequence is summarized in Figs. 4 and 5. The key event in all ribonucleotide reductase proteins is the formation of a transient thiyl radical R–S$^{\cdot}$ at the substrate site through a long-distance radical transfer (Fig. 4).[38,40] The SH hydrogen of an essential cysteine residue that is situated in a conserved sequence environment[27] and exposed position in the large (R1, α) subunits of all RNRs, is abstracted by interaction with the primary tyrosyl, glycyl, or deoxyadenosyl[28] radicals. These organic radicals are in turn generated by the high-valent di-iron, iron-sulfur, or cobalamin metal cofactors located in the small subunits (R2, β) or bound to R1, respectively. Whereas the stable tyrosyl and glycyl radicals are well characterized by EPR spectroscopy, direct detection of the short-lived R–S$^{\cdot}$ species during catalysis has remained elusive. Only recently were EPR signals of thiyl radicals observed which had been generated chemically or photolytically in the R1 subunit of *E. coli* RNR in frozen solution[41] but their involvement in the catalytic cycle remains to be established.

At the active site (Fig. 5) a substrate ribonucleotide (**1**) is bound at its phosphates, and the two hydroxy groups of ribose are engaged in hydrogen bonds (not shown). Reduction is initiated by the thiyl radical abstracting the hydrogen atom from C3′ (**2**). Homolytic activation of this particular C–H bond, although probably thermodynamically unfavourable, is in accord with the isotope effects observed for 3′-deuterated and tritiated substrates[20] and the inactivity of *xylo*-ATP (with an OH group in the position of 3′-H)[19]

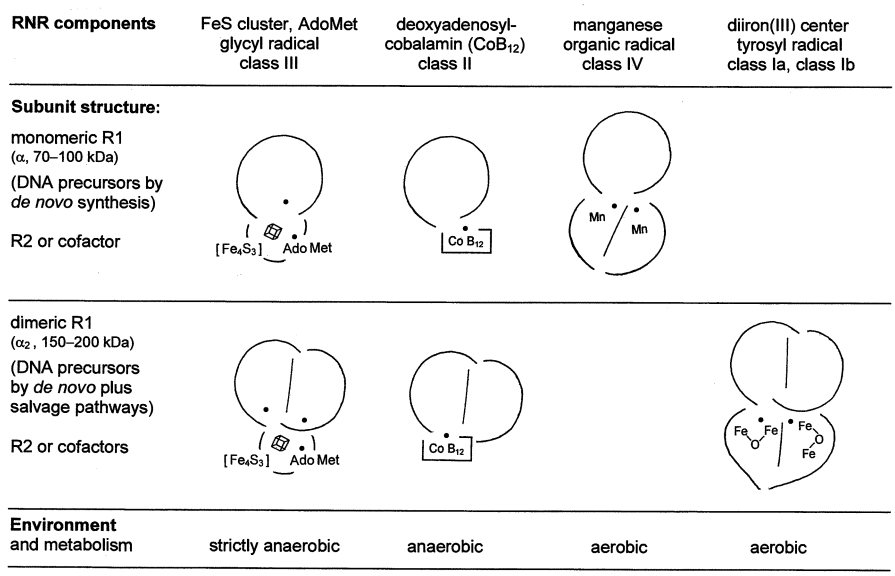


Fig. 3 A system of the currently known types of ribonucleotide-reducing enzymes. Structures are schematic (*cf.* text). Monomeric R1 appear to be limited to microorganisms that solely depend on *de novo* DNA precursor synthesis, necessitating simple feedback regulation of their RNR. Dimeric R1 are found in organisms capable of utilizing salvaged DNA components in addition to *de novo* synthesis and requiring more complex allosteric regulation of dNTP pools.[4,11] Occurrence in anaerobic and aerobic life, and in individual species is listed in.[21] AdoMet, *S*-adenosylmethionine; CoB_{12}, deoxyadenosylcobalamine; • organic (amino acid) radicals. The transient thiyl radicals within R1 proteins are not indicated.

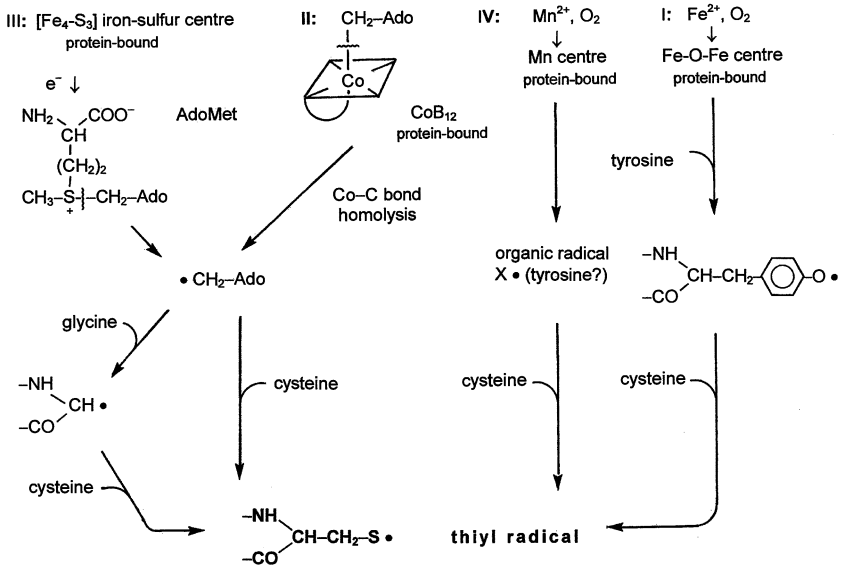


Fig. 4 Generation of the common thiyl radical (bottom) in radical transfer chains characteristic of (from left to right, anaerobic to aerobic) class III, II, IV, and class I RNRs.[28,30,40] A specific cysteine residue in class IV is implicated by analogy. AdoMet, *S*-adenosylmethionine; CoB_{12}, deoxyadenosylcobalamin; CH_2–Ado, 5′-deoxyadenosine.

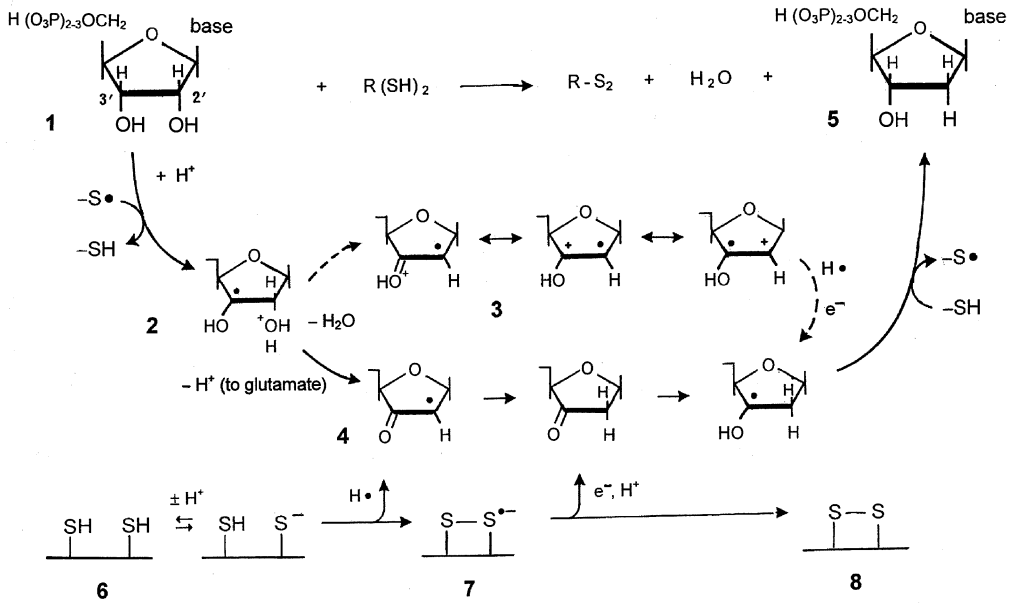


Fig. 5 The radical mechanism of RNR-catalyzed reduction of a ribonucleotide **1** to a 2′-deoxyribonucleotide **5** (top and center) concomitant with oxidation of a dithiol **6** to a disulfide **8** (bottom). *Cf.* text for intermediates **2**, **3**, **4** and **7**.

and 3′-*C*-methyl ribonucleotides.[25] It is also supported by theoretical considerations.[42] Following loss of water from C2′, the remaining pentofuranose structure may be viewed as a resonance-stabilized radical cation (**3**) or, if 3′-OH is deprotonated by a nearby glutamate, as neutral 3′-keto-2′-deoxynucleotide radical (**4**). Species **3** and **4** differ only in the protonation state. Not surprisingly, none of the carbon-centered radicals implicated in the transition state could as yet be observed by EPR spectroscopy. However, substrate analogues carrying 2′-azido or 2′-mercapto functions do produce EPR signals ascribed to nitrogen-centered or perthiyl radical species.[43] The latter are potent mechanism-based inhibitors that destroy RNR activity by interference with the catalytic radical chain. The substrate radicals (**3**, **4**) are reduced by hydrogen transfer to C2′ from an adjacent dicysteine site (**6**) in RNR class I and II, or by a single cysteine plus formate in class III. In this half-reaction an intermediate disulfide radical anion (**7**) could be detected by high-frequency time-domain EPR spectroscopy in a glutamate-to-glutamine mutant enzyme.[44] Final electron and hydrogen transfer returns H to position C3′ to form the 2′-deoxyribonucleotide product (**5**). The oxidized dicysteine (disulfide) domain (**8**) will be re-reduced by external dicysteine polypeptides (thioredoxin or glutaredoxin) *via* yet another conserved dicysteine sequence within the RNRs. Summarily, the two-electron (H^-) transfer mechanism operating in most dicysteine–disulfide oxidoreductases is broken up into two one-electron steps in ribonucleotide reduction in order to accomplish reduction of an alcohol carbon–oxygen bond.

3.3 Chemical deoxygenation reactions

Several routes have been explored for transforming ribonucleosides or ribonucleotides into their 2′-deoxyribose congeners as the former, carrying either natural or synthetic heterocyclic bases, are much more readily available. Most studies are aimed at expanding the repertory of base- or sugar-modified nucleoside analogues as potential enzyme inhibitors in nucleic acid and energy metabolism,

or as pharmacologically interesting antagonists (*vide infra*). It is obvious that in order to achieve selective removal of the 2′ secondary alcohol function a riboside has to be substituted with protecting groups at the other hydroxy groups. Furthermore, as direct displacement of the sugar OH function at C2′ is unfavourable sterically and electronically it has to be replaced by a better leaving group. All transformations of such highly derivatized educts require organic solvents and elevated temperature.

Early attempts at the conversion of adenosine or uridine (**9**) to the corresponding deoxyribosides (**15**) met with variable success[45–48] (Fig. 6). The adenine or uracil bases were modified for the activation of C2′ by conversion to cyclic 8,2′-thioanhydro or O^2,2′-anhydro arabinofuranosyl intermediates (**10**,**12**), but subsequent reductive desulfurization or dehalogenation resulted in poor overall yields. Cyclic orthoester (acyloxonium ion) derivatives bridging the ribose 2′ and 3′ positions (**11**) can serve as synthetic intermediates, with the disadvantage that their ring opening by halide ions (**13**) and catalytic hydrogenation produce mixtures of 2′- and 3′-deoxyribonucleosides. A generally applicable method of regiospecific 2′ deoxygenation became available with the adaptation of Barton–McCombie radical chemistry, originally conceived for reduction of hydroxy groups in steroids, to the nucleoside field.[45] In the most efficient, four-stage variant a ribonucleoside is converted to a 3′,5′-*O*-(tetraisopropyl)disiloxyl derivative (**14**) for blocking 3′- and 5′-hydroxys, and the 2′-OH group reacts with phenoxythiocarbonyl chloride to produce a phenyl thionocarbonate ester. *S*-Methyl xanthates are also suitable for 2′-O activation. Reduction then proceeds smoothly in refluxing toluene by treatment with tri-*n*-butyltin hydride plus 0.2 equivalents of azobis(isobutyronitrile) as a radical starter, followed by deprotection of the 3′ and 5′-OH groups. Natural and base-modified 2′-deoxyribonucleosides (**15**) have been produced on the gram scale in 60–75% overall yield. When tributyltin deuteride was used for hydrogenolysis, > 85% deuterium substitution in the original 2′-*ribo* configuration was observed. Obvious similarities exist between the chemical and enzyme-catalyzed deoxygenation mechanisms in the role of radicals and product stereoselectivity but "biomimetic relations"[45] are hard to reconcile with the highly diverging reaction media and substrate (sugar-modified nucleoside *vs*. natural nucleotide) structures.

4 Sources of deoxyribonucleotides for research and applied science

Deoxyribonucleotides, needed for research and medical applications, are in short supply. Unlike vitamins, coenzymes, ATP, *etc*. they cannot be extracted and purified from biological sources because they do not accumulate in cells and tissues. Deoxyribonucleoside triphosphates (dNTP) are produced by a few S phase-specific enzymes including RNR just in time for DNA replication and are immediately incorporated into the new DNA strands. Overproduction of dNTPs can be induced by inhibitors such as fluorouracil[3] but such dNTP pool imbalances are accompanied by "unbalanced cell growth" in microbial cultures and are not practical as a route to nucleotide production.

4.1 Present demand and production

The demand for purified deoxyribonucleosides and -nucleotides is high. DNA polymerase- and reverse transcriptase-catalyzed DNA syntheses *in vitro*, that are common in all areas of molecular biology, require kilograms of the four dNTP substrates worldwide. Much larger quantities are needed as starting material for drugs such as base-halogenated thymidine derivatives (idoxuridine, trifluridine) used against herpes virus and in tumor chemotherapy, or for sugar-modified compounds like 3′-azido thymidine (AZT, zidovidin) effective in HIV treatment. In addition, a rapidly growing demand has developed for deoxyribo oligonucleotides (20- to 100-mers) which are immobilized on DNA chips used in research and medical diagnosis, and required as antisense constructs in the field of gene therapy. It has been estimated by stock exchange experts that a total of 200 tons of deoxyribonucleosides (including 1,000–2,000 kg of oligonucleotides) might be required in future years, far exceeding the quantities currently available.

Three routes exist for large-scale production of deoxyribonucleotides: (I) total synthesis of deoxyribose, heterocyclic bases, and the *N*-glycosidic bond; (II) hydrolysis of DNA, the only natural product containing deoxyribonucleotides; and (III) biotechnological production by modified microorganisms or enzymes, yet to be established.

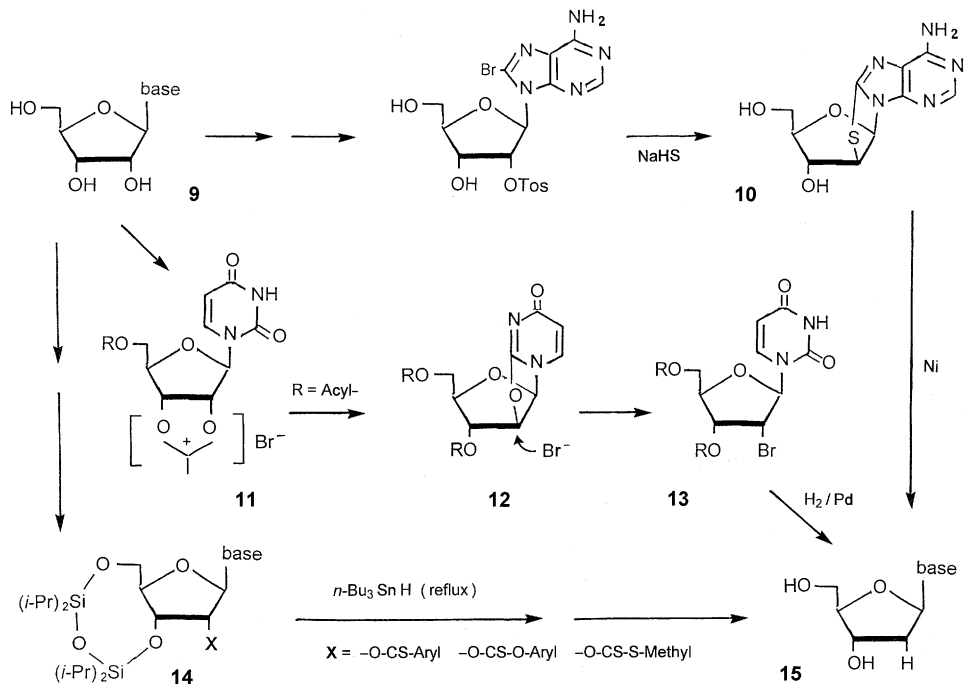


Fig. 6 Educts and conditions for chemical transformation of ribonucleosides (**9**) to 2′-deoxyribonucleosides (**15**).[45–47] The competing reaction producing a 3′-deoxyribonucleoside from **11** is not shown. Standard reaction steps for the introduction and removal of protecting groups are not specified.

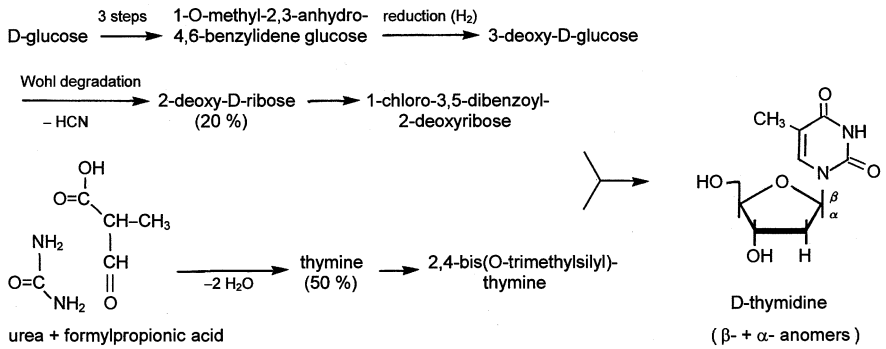


Fig. 7 Total synthesis of 2-deoxy-D-ribose, thymine and D-thymidine.[48] The mixture of thymidine β- and α-anomers can be resolved chromatographically.

Chemical synthesis of D-deoxyribose from D-glucose *via* 3-deoxy-D-glucose, followed by coupling with synthetic 2,4-bis-(trimethylsiloxyl)thymine by the Hilbert–Johnson method[48] provides more than 200 tons of thymidine annually (Fig. 7). This is mostly used for production of the anti-HIV drug AZT. The process includes 15 steps, requires large amounts of hazardous chemicals, produces a mixture of D-β-thymidine and the undesired α-anomer, and overall yields do not exceed 3%. A market price of about 500 US$ kg^{-1} thymidine reflects this situation.

It seems like an anachronism that 130 years after Friedrich Miescher purified DNA from salmon sperm at Basel on the Rhine, and about half a century after complex nucleic acid hydrolysates were effectively separated by ion exchange chromatography, extraction of salmon DNA and fractionation of its hydrolysates are still the source of highly priced and valued chemicals. The only difference is that today's process is carried out in 1000 L reactors and on huge chromatography columns (Fig. 8). Fish milt is very

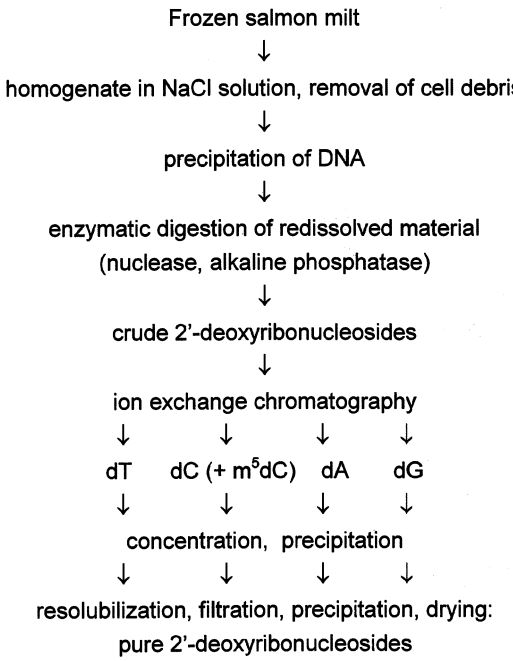


Fig. 8 Flow diagram for production of the four natural 2′-deoxyribonucleosides from salmon sperm DNA. m^5dC, 5-methyldeoxycytidine, present in vertebrate DNA in about 1%.

DNA-rich, and annual salmon catches of 200,000 tons from North American, Chilean, and Japanese waters currently supply about 3,000 tons of milt (3% by weight of the male salmon), which is frozen, shipped and processed in the US. One thousand kg (1 ton) of milt yields approximately 120 kg of wet DNA paste, that in turn yields 16 kg of purified deoxyribonucleosides. This corresponds to a recovery of about 65% based on the amount of DNA monomers present in the tissue. If milt of the total world salmon harvest could be collected, twice the amount (more than 100 tons of purified deoxyribonucleosides, or > 25 tons of each DNA monomer) would be obtained annually. For manufacture of oligodeoxyribonucleotides, the nucleosides have to be subjected to chemical phosphorylation or phosphitylation. Oligomers of defined sequence are then synthesized on solid supports (costly themselves) at an expense of several hundred $ per gram.

Although it is based on a supply of natural material and simple technology, the DNA hydrolysis route has its drawbacks. Annual salmon harvests cannot be guaranteed, and processing from milt to purified deoxyribonucleosides takes months. Therefore, orders for a kilogram of oligonucleotides used in clinical trials require long-term planning. The process is labour-intensive, and the unit costs of raw materials (and hence, the products) will not diminish with increased production volume. Most importantly, the method dictates that the yield of all four deoxyribonucleosides is about the same regardless of requirements, *e.g.* in the case of thymidine. Base-modified nucleotides cannot be obtained in this way at all. DNA fractionation may continue to provide a substantial portion of its four constituents but complementation by a more flexible, biotechnological production line would confer additional advantages.

4.2 Perspectives in biotechnology

Enzymatic formation of specific deoxyribonucleotides offers an obvious alternative route. Purine ribonucleotides such as inosine and guanosine 5′-monophosphate (IMP, GMP) are produced in large amounts (exceeding 1,000 tons per year) as flavour-enhancing food additives in the Asian fermentation industry; other monophosphates carrying modified pyrimidine and purine bases can also be obtained.[11] They are inexpensive chemicals but their reductive transformation into deoxyribonucleotides has never been seriously attempted. The complex nature of ribonucleotide reductases, in particular the well-known class I enzymes, makes them prone to inactivation by irreversible loss of radical content over hours of incubation, and hence they are not particularly attractive as catalysts for technology. Overproduction of the DNA monomers in genetically modified microorganisms is also difficult to envisage due to the mutagenic and growth-disturbing consequences of unbalanced DNA precursor pools.[1–3]

In a recent R&D project we have reasoned that the coenzyme B_{12}-dependent class II RNR and also the manganese-containing class IV enzymes found in coryneform bacteria (which are long established production strains in biotechnology) are kinetically simpler and potentially more suitable for long-term use in *in vitro* systems. Class II RNRs are able to reduce a wide range of base-modified ribonucleotides,[12] and are the only enzymes that could successfully be immobilized.[49] A partially purified, stable preparation of monomeric *Lactobacillus leichmannii* RNR, cloned and overexpressed in *Escherichia coli*[27] indeed continued to reduce ATP to dATP for 10 hours or more in the presence of very low (10

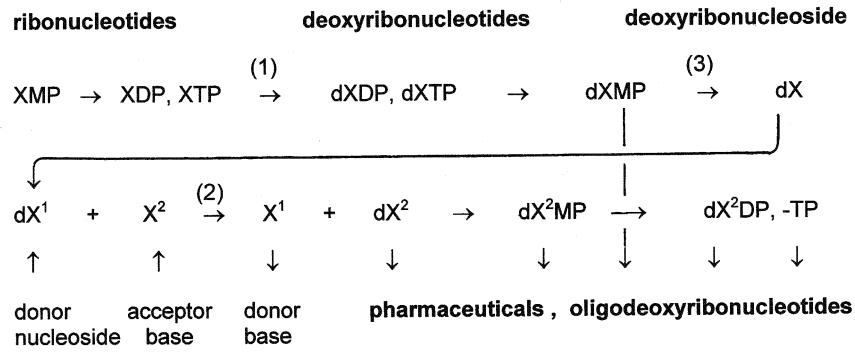


Fig. 9 A general scheme for enzyme-catalyzed production of natural and base-modified 2′-deoxyribonucleosides, -nucleotides and -oligonucleotides for medical and other applications. X denotes an unspecified heterocyclic base (purine, pyrimidine, or isomers and deaza derivatives thereof); -MP, -DP, -TP, 5′-mono-, di-, triphosphate. Enzymes are: (1), RNR; (2), *N*-deoxyribosyl transferase; (3) phosphatase. Acceptor bases can be derived from XMP, or be of synthetic origin. Liberated donor bases may be recycled.

μM) concentrations of coenzyme B_{12}. At a nucleotide ratio of AMP:ATP = 10, the monophosphate also served as a substrate. Purified 2′-deoxyadenylates were obtained in the laboratory on a 0.1–1 g scale using this system. Progress towards *in vitro* use has also been made with the dimeric manganese RNR of *Corynebacterium ammoniagenes* and *C. glutamicum* which do not require an expensive coenzyme for activity and tolerate high salt concentrations. Both subunits (nrdEF) could be cloned and overexpressed in a stable homologous expression system,[50] resulting in correct metal incorporation, good activity, and increased resistance towards RNR inhibitors. These strains are able to carry out ribonucleotide reduction in permeabilized cells, with the potential of isolating deoxyribotide products from the culture medium.

Coupling of biotechnologically optimized ribonucleotide reductases with the robust and readily available bacterial *N*-deoxyribosyl transferases (EC 2.4.2.6) capable of exchanging natural and modified, synthetic purine and pyrimidine bases[51] should eventually permit most flexible production of many different deoxyribonucleosides and -nucleotides (Fig. 9). Scaling-up, and industrial exploitation of this promising new route appear within reach.

5 DNA genomes *vs*. RNA world: A challenge for biomimetic chemistry

While the biochemistry of deoxyribonucleotides is fairly well understood, their abiotic source has remained a missing link between the popular concept of an "RNA world" at the origin of Life and the dominance of DNA genomes in all present-day organisms. If RNA did provide both information and catalysis in the earliest replicating molecular entities, why, and when, was 2-deoxyribose selected to constitute a second informational macromolecule, DNA?

The search for non-enzymatic, yet biomimetic analogues of RNR catalysis in aqueous systems might provide an answer.[9,12,52] A novel approach rests on an obvious similarity of enzymatic ribonucleotide reduction [eqn. (2)], and reduction of oxidized small organic molecules X, *e.g.* aldehydes or oxo acids, in the ferrous sulfide (FeS)/pyrite (FeS_2)-driven "iron-sulfur world" elaborated by Wächtershäuser[53] [eqn. (4)]. Substituting FeS–H_2S for the dithiol reductants of ribonucleotide reduction in eqn. (2) yields a hypothetical reaction scheme for abiotic deoxyribotide formation [eqn. (5)] amenable to experimental testing:

ribonucleotide(–OH) + $R(SH)_2$ → deoxyribonucleotide(–H) + RS_2 + H_2O (2)

$$FeS + H_2S + X \rightarrow FeS_2 + XH_2 \quad (4)$$

ribonucleotide(–OH) + FeS + H_2S → deoxyribonucleotide(–H) + FeS_2 + H_2O (5)

Such a system would mimic the action of anaerobic, FeS cluster-dependent class III RNR, with anionic nucleotides bound, and probably activated, at the positively polarized iron sulfide surface rather than at a protein site.

We have observed that adenosine 5′-phosphate (AMP, chosen because of the prevalence of adenine derivatives among prebiotic products and present-day biomolecules) indeed interacts with freshly precipitated FeS and hydrogen sulfide during anaerobic incubation.[52] Besides some hydrolysis, ^{1}H-NMR spectroscopy which allows to distinguish the chemical shifts and coupling of C1′-H and C2′-H (or C3′-H) signals in ribotides and 2′- (or 3′-) deoxyribotides revealed the formation of compounds bearing a deoxysugar, most likely representing 2′-deoxyadenylate with an admixture of the 3′-isomer. Notably, product formation required the addition of 0.2% selenium to the FeS–hydrogen sulfide medium. Selenium is a dehydrogenating agent in organic synthesis, and potentiates the catalytic power of cysteines in redox enzymes. Here, it is thought to promote initial activation of a nucleotide at the ribose moiety (*cf.* Fig. 5), thus completing the biomimetic model of RNR catalysis.

These exploratory studies support the notion that both RNA and DNA emerged almost in parallel. In a strongly reducing primordial world, with abiotic formation of ribonucleotides, RNA, and ribozyme-like polymerase activity, iron–sulfur chemistry could have supplied monomers for more hydrolysis-resistant DNA molecules. Alternatively, in a scenario in which RNA came along with amino acids and polypeptides, deoxyribonucleotides would have been produced by a modular ur-reductase integrating cysteine sequences and transition metal-derived cofactors. In any event a much delayed advent of DNA in molecular and cellular evolution seems unlikely. The overlap of organic and inorganic chemistry, conspicuous in DNA precursor biochemistry today, may have shaped Life from the beginning.

Acknowledgements

Our studies of deoxyribotide chemistry and enzymology have been supported by Deutsche Forschungsgemeinschaft. I wish to thank the many dedicated graduate students who have identified new reactions and enzymes in microorganisms and plants at the universities of Marburg and Kassel, and Alexandra Choubnikova for the exploratory biomimetic experiments. Special thanks are due to Drs. Harry Hogenkamp at the University of Minnesota, USA, and Georg Auling at the University of Hannover, Germany, for long-standing scientific collaboration and friendship. I am also grateful to Drs. Rolf Bald, Berlin, Germany, and Carol Brownson, London, UK, for their advice.

References

1 B. A. Kunz, S. E. Kohalmui, T. A. Kunkel, C. K. Mathews, E. M. McIntosh and J. A. Reidy, *Mutation Res.*, 1994, **318**, 1 and references therein.
2 V. Bianchi, S. Borella, C. Rampazzo, P. Ferraro, L. C. Bianchi, S. Skog and P. Reichard, *J. Biol. Chem.*, 1997, **272**, 16118.
3 W. Feller, G. Schimpff-Weiland and H. Follmann, *Eur. J. Biochem.*, 1980, **110**, 85.
4 P. Reichard, *Biochemistry*, 1987, **26**, 3245.
5 P. A. Levene and T. Mori, *J. Biol. Chem.*, 1929, **83**, 803.
6 W. G. Overend and M. Stacey, in *The Nucleic Acids, Chemistry and Biology*, ed. E. Chargaff, Academic Press, Inc., New York, 1955, **vol. 1**, pp. 48–60 and references therein.
7 E. Hardegger and G. N. Richards, *Methods in Carbohydrate Chemistry*, ed. R. L. Whistler, Academic Press, Inc., New York, 1962, **vol. 1**, 177.
8 W. E. Pricer and B. L. Horecker, *J. Biol. Chem.*, 1960, **235**, 1292.
9 H. Follmann, *Adv. Space Res.*, 1986, **6**, 33 and references therein.
10 G. L. Nelsestuen, *Biochemistry*, 1979, **18**, 2843.
11 G. Auling and H. Follmann, in *Metal Ions in Biological Systems*, ed. H. Sigel, Marcel Dekker, Inc., New York, Basel, 1994, **vol. 30**, pp. 131–161 and references therein.
12 M. Lammers and H. Follmann, in *Structure and Bonding*, Springer-Verlag Berlin, 1993, **vol. 54**, pp. 27–91 and references therein.
13 B.-M. Sjöberg, in *Structure and Bonding*, Springer-Verlag Berlin, 1997, **vol. 88**, pp. 139–173 and references therein.
14 E. Hammarsten, P. Reichard and E. Saluste, *J. Biol. Chem.*, 1950, **183**, 105.
15 T. J. Batterham, R. K. Ghambeer, R. L. Blakley and C. Brownson, *Biochemistry*, 1967, **6**, 1203.
16 R. A. Neese, L. M. Misell, S. Turner, A. Chu, J. Kim, D. Cesar, R. Hoh, F. Antelo, A. Strawford, J. M. McCune, M. Christiansen and M. K. Hellerstein, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 15345.
17 P. M. Schaffer, R. P. McCroskey and M. T. Abbott, *Biochim. Biophys. Acta*, 1972, **258**, 387.
18 J. Stubbe, *J. Biol. Chem.*, 1990, **265**, 5329.
19 H. Follmann and H. P. C. Hogenkamp, *Biochemistry*, 1969, **8**, 4372; H. Follmann and H. P. C. Hogenkamp, *Biochemistry*, 1971, **10**, 186.
20 J. Stubbe, M. Ator and T. Krenitzky, *J. Biol. Chem.*, 1983, **258**, 1625.
21 E. Torrents, P. Aloy, I. Gibert and F. Rodriguez-Trelles, *J. Mol. Evol.*, 2002, **55**, 138 and references therein.
22 P. Nordlund and H. Eklund, *J. Mol. Biol.*, 1993, **232**, 123.
23 U. Uhlin and H. Eklund, *Nature*, 1994, **370**, 533.
24 O. B. Kashlan, C. P. Scott, J. D. Lear and B. S. Cooperman, *Biochemistry*, 2002, **41**, 462.
25 S. P. Ong, L. S. Nelson, S. C. McFarlan and H. P. C. Hogenkamp, *Biochemistry*, 1992, **31**, 11210; S. P. Ong, L. S. Nelson, S. C. McFarlan and H. P. C. Hogenkamp, *Biochemistry*, 1993, **32**, 11397.
26 D. Panagou, M. D. Orr, J. R. Dunstone and R. L. Blakley, *Biochemistry*, 1972, **11**, 2378.
27 S. Booker and J. Stubbe, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 8352.
28 S. Licht, S. Booker and J. Stubbe, *Biochemistry*, 1999, **38**, 1221 and references therein.
29 D. T. Logan, J. Andersson, B.-M. Sjöberg and P. Nordlund, *Science*, 1999, **283**, 1499 and references therein.
30 J. Tamarit, C. Gerez, C. Meier, E. Mulliez, A. Trautwein and M. Fontecave, *J. Biol. Chem.*, 2000, **275**, 15669.
31 E. Mulliez, S. Ollaguier, M. Fontecave, R. Eliasson and P. Reichard, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 8759.
32 A. Willing, H. Follmann and G. Auling, *Eur. J. Biochem.*, 1988, **170**, 603.
33 U. Griepenburg, K. Blasczyk, R. Kappl, J. Hüttermann and G. Auling, *Biochemistry*, 1998, **37**, 7992.
34 U. Griepenburg, G. Lassmann and G. Auling, *Free Radical Res.*, 1996, **26**, 473.
35 Y. Huque, F. Fieschi, E. Torrents, I. Gibert, R. Eliasson, P. Reichard, M. Sahlin and B.-M. Sjöberg, *J. Biol. Chem.*, 2000, **275**, 25365 and references therein.
36 J. Harder, *FEMS Microbiol. Rev.*, 1993, **12**, 273.
37 A. M. Poole, D. T. Logan and B.-M. Sjöberg, *J. Mol. Evol.*, 2002, **55**, 180 and references therein.
38 S. Licht, G. J. Gerfen and J. Stubbe, *Science*, 1996, **271**, 477 and references therein.
39 J. Stubbe, *Chem. Commun.*, 2003, 2511.
40 A. Gräslund, *Methods Enzymol.*, 2002, **354**, 399 and references therein.
41 M. Kolberg, G. Bleifuss, B.-M. Sjöberg, A. Gräslund, W. Lubitz, F. Lendzian and G. Lassmann, *Arch. Biochem. Biophys.*, 2002, **397**, 57.
42 M. Mohr and H. Zipse, *Chem. Eur. J.*, 1999, **5**, 3046 and references therein.
43 J. Coves, L. H. de Fallois, L. Le Pape, J.-L. Decout and M. Fontecave, *Biochemistry*, 1996, **35**, 8595.
44 C. C. Lawrence, M. Bennatt, H. V. Obias, G. Bar, R. G. Griffin and J. Stubbe, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 8979.
45 M. J. Robins, S. F. Wnuk, A. E. Hernandez-Thirring and M. C. Samano, *J. Am. Chem. Soc.*, 1996, **118**, 1342 and references therein.
46 M. Ikehara, *Acc. Chem. Res.*, 1969, **2**, 47.
47 R. Marumoto and M. Honjo, *Chem. Pharm. Bull. (Tokyo)*, 1974, **22**, 128.
48 H. Vorbrüggen and C. Ruh-Polenz, in *Organic Reactions*, ed. L. A. Paquette, John Wiley & Sons, New York, 2000, **vol. 50**, pp. 12–108 and references therein.
49 P. Halicky, M. Kollarova, P. Kois and J. Zelinka, *Collect. Czech. Chem. Commun.*, 1989, **54**, 2528.
50 W. Oehlmann and G. Auling, *Microbiology*, 1999, **145**, 1595 and references therein.
51 J. Becker and M. Brendel, *Biol. Chem. Hoppe-Seyler*, 1996, **377**, 357 and references therein.
52 H. Follmann, *Biol. Chem.*, 2001, **382**, S20.
53 E. Blöchl, M. Keller, G. Wächtershäuser and K. O. Stetter, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 8117; D. Hafenbradl, M. Keller, G. Wächtershäuser and K. O. Stetter, *Tetrahedron Lett.*, 1995, **29**, 5179 and references therein.