Volume 16 Number 20 1988 Nucleic Acids Research

Mispair formation in DNA can involve rare tautomeric forms in the template

Peter Strazewski+

Institut für organische Chemie, St Johanns-Ring 19, CH-4056 Basel, Switzerland

Received April 29, 1988; Revised August 19, 1988 and Accepted September 1, 1988

ABSTRACT
The formation of pyridine-pyrimidine- and pyrimidine-pyrimidine base pairs after *in vitro* DNA replication with the large fragment of *Escherichia coli* DNA polymerase I indicates that Watson-Crick-like base pairing between pyrimidine bases can occur in the enzyme due to the presence of the rare tautomers of deoxycytidylate and thymidylate in the template strand. The implications to mispair formation in DNA, such as the difference between the structures of the mispairs during and after replication, are discussed and the possible action of mutagenic DNA protonating and deprotonating agents *in vivo* is considered.

TEMPLATE-DIRECTED MUTAGENESIS *IN VITRO*

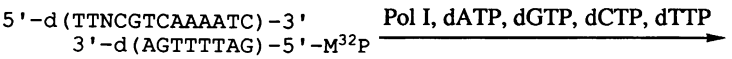
Introduction

In *E. coli* it is known that two enzymes - DNA polymerase I and DNA polymerase III (Pol I and Pol III) - are involved essentially in replication [1-4]. While Pol III does the main replication work *in vivo*, leaving one strand continuously and the other strand fragmentarily replicated, Pol I polymerizes 2'-deoxyribonucleoside-5'-triphosphates in the gaps between the *Okazaki* fragments in the 5'→3' direction including a 3'→5' exonuclease-proofreading (error rate ≅ 10^{-5}-10^{-6} *in vitro*, ≅ 10^{-8} *in vivo*) [4-5]. Additionally it is capable of excizing distorted mispairs in a duplex as a 5'→3' exonuclease with endonuclease activity (excision of up to 10 base pairs). Investigations on the high fidelity of polymerization and on the precision of proofreading led to the hypothesis that the recognition of the proper base to incorporate depended mainly on the matching of the hydrogen bonds between the template base and the nucleoside triphosphate.

Thus, the *in vitro* incorporation by Pol I of deoxynucleoside triphosphate analogues with reduced capability for hydrogen-bonding with a template should render information about the specificity of the enzyme. At the time it was known that purine analogues were poor substrates for Pol I and so we decided to synthesize three dCTP-**structure** analogues (dNTP: 4-deamino-2'-deoxycytidine-, 3-deaza-4-deamino-2'-deoxycytidine-, and 3-deaza-2'-deoxycytidine-5'-triphosphates). Apart from the natural nucleotides, we could not detect any polymerized dN in the prolonged DNA primer, whether with activated calf thymus DNA, nor with synthetic poly-d(GC) as a template, using Pol I, dATP, dGTP, dCTP, dTTP, and dNTP [6].

Enzymatic Synthesis of a Pyridine-Pyrimidine Base Pair

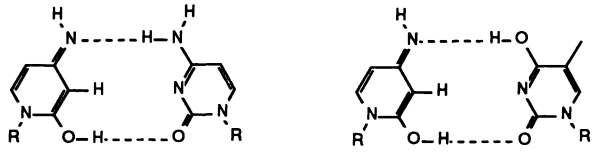
The Experiment. Then we switched to the opposite strategy [7] of using synthetic 13-mer template strands bearing at position 11 one analogue dN each, and letting Pol I (Klenow fragment, no 5'→3' exonuclease activity), dATP, dGTP, dCTP, and dTTP elongate a matched, labelled 8-mer primer strand (*scheme 1*) [8]. The specific incorporation of dC at position 3 in the primer



scheme 1

directed by dN = 3-deaza-2'-deoxycytidylate (d(3-deazaC)), as well as the complete and faultless prolongation of the chain could not be explained by means of a reduced capability of d(3-deazaC) to hydrogen-bond with the nucleoside triphosphates.

The Assumed Structure. The reason for the unusual pairing of a pyrimidine with a pyridine base was found in the special hydrogen substitution pattern of the base analogue. Since base-pairing with dG (and dT) is hindered by protruding interactions between the C(3)-hydrogen atom of the analogue and the imino protons of dG and dT, the only base pair that is bound by 2 H-bonds can be d(3-deazaC)*' (in its rare imino-enol form[1])·dC [9]. Alternatively, the hydrogen substitution pattern of d(3-deazaC)*' also matches to dT* (*scheme 2*).



scheme 2

There are possibilities of pairing d(3-deazaC) (in its normal tautomeric form) with other bases in two alternative wobble positions where the incoming nucleoside triphosphates are shifted either towards the minor or the major groove of the DNA. In these pairs the C(3)-hydrogen atom of the analogue would not prevent dGTP or dTTP to hydrogen-bond with d(3-deazaC) any more. Although it cannot be denied that **after polymerization**, a d(3-deazaC)*'·dC pair could tautomerize into the normal tautomeric form, accompanied by a rearrangement into a d(3-deazaC)·dC wobble pair, the lack of detectable amounts of dGMP or dAMP incorporated opposite d(3-deazaC) indicated that a wobble pair configuration was not possible **during polymerization**. It also indicated that no rare tautomeric forms of the nucleoside triphosphates such as dG*TP (lactim form of dGTP without the protruding imino proton) could be incorporated by Pol I, thus confirming other mutagenesis results (*vide infra*).

Admittedly, up to now no turnover of dCTP to dCMP or of any other nucleoside triphosphate was measured in this mutagenesis experiment, so no precise information about the proofreading of the pyridine-pyrimidine pair can be made. But, since the primer elongation was performed under standard conditions used for normal template-directed and Pol I promoted primer elongations and the elongation occurred quantitatively within that time, we conclude that the 3'→5' exonuclease proofreading activity of the Klenow fragment could be 'fooled' by a Watson-Crick-like geometry of the pyridine-pyrimidine base pair in the polymerase.

The possibility (as mentioned in [8]) that the analogue's tautomeric equilibrium could lie more on its imino-enol side in an aqueous solution due to the enhanced basicity of the O=C(2) [10-11] as compared with the parent compound dC was investigated [12]. ^{13}C-NMR spectroscopical data showed no substantial difference in the tautomeric equilibria of the nucleosides. Still, d(3-deazaC)*' must have been the only reactive species for structural reasons and, by overcoming the enzyme's proofreading activity, provided a specific and complete prolongation of the primer strand.

The Functional Analogy. As such d(3-deazaC)*' is a **functional** analogue of dG, because of its similar hydrogen donor/acceptor behaviour and its selectivity to bind dCTP in its normal tautomeric form. Like the previous incorporation experiments with pyrimidine ribonucleoside analogues [10,13], the negative results of the *Ames*-test and the "lambda-mutatest" on d(3-deazaC) (Th. Bickle, unpublished results) revealed that d(3-deazaC), though most probably phosphorylated to d(3-deazaC)TP *in vivo*, was not mutagenic (i.e. was not incorporated) and cannot be used as a C→G transversion mutagen.

The Verification Step. The peculiarity of d(3-deazaC) which as a template base can direct a specific incorporation, whereas d(3-deazaC)TP will not incorporate, can be rationalized by a substrate incorporation mechanism proposed by Ferrin & Mildvan [14] where d(3-deazaC)TP first is coordinated by Mg^{2+} of the Pol I-template-primer-complex at its γ-phosphate. This binding is only marginally dependent on the structure of the base. Prior to the polymerization step, a rate-limiting [15] β-phosphate coordination is required that cannot proceed unless the triphosphate base binds **in a Watson-Crick-like geometry** to the template base ("verification step"). d(3-deazaC)TP could not form a reasonable (more than one H-bond) Watson-Crick base pair to any of the four natural bases, d(3-deazaC)*TP could to dC, but its assumed rareness - if present in the enzyme at all - makes it a rather weak competitor to dGTP. As a template base it cannot be replaced and the polymerization will not proceed unless dCTP "meets" d(3-deazaC)*'.

THEORETICAL ASPECTS

Computer Modeling

The proposed structure of d(3-deazaC)*'·dC as depicted in *scheme 2* was simulated in a "modeling" computer programme for nucleosides and nucleotides (energy minimization performed with proprietary united-atom force field MOLOC developed by the CAMM group, Müller, K., Hoffmann-La Roche, CH - 4058 Basel). Apart from a slight shifting of the hydrogen binding axis

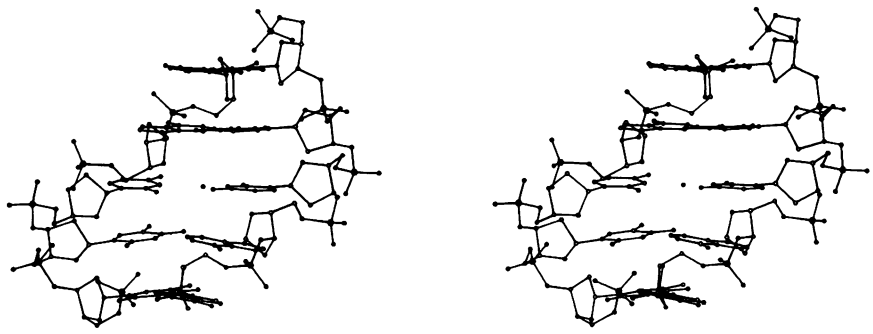


figure 1 **Stereo view of the pentamer double-helix on the minor groove. The dot stands for the O atom of H_2O.**

due to a third, longer hydrogen bond between N(3) and HO-C(2) of d(3-deazaC)*' and/or the repulsion between the non-bonding electron pair of N(3) and H-C(3)-electrons of d(3-deazaC)*', the bases paired as predicted. Next we incorporated d(3-deazaC)*' and dC in a canonical B-DNA double-stranded pentanucleotide [2]. The sequence was similar to a part of our 13-mer used in the Pol I experiment where dN_3 was d(3-deazaC)*':

$$5`\text{-d}(pT_5pT_4pN_3pG_2pC_1\ )\text{-}3`$$
$$3`\text{-d}(\ A_1pA_2pC_3pC_4pG_5p)\text{-}5`$$

Relaxation of the approximatively 200-atom system led to an almost unchanged structure where the crucial bases were paired by one hydrogen bond between the dC amino-function and the d(3-deazaC)*' imino-function (N-N-distance 3.0 Å). The other hydrogen bond which would have caused a major change in the sugar phosphate backbone, if bound directly with dC_3 (presumably a twisted bent), was instead directed by an ether oxygen from the ribose ring of the dG_2-unit in the same strand. No unusual strains appeared, all bases stacked as in normal B-DNA.

In order to fill the space between the two base oxygen atoms (OC(2)) in the minor groove, a water molecule was inserted with its oxygen atom in the middle between the others. After

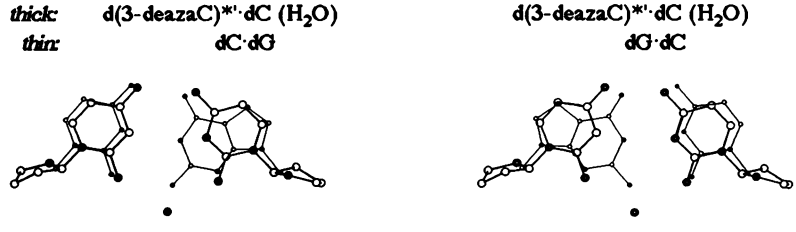


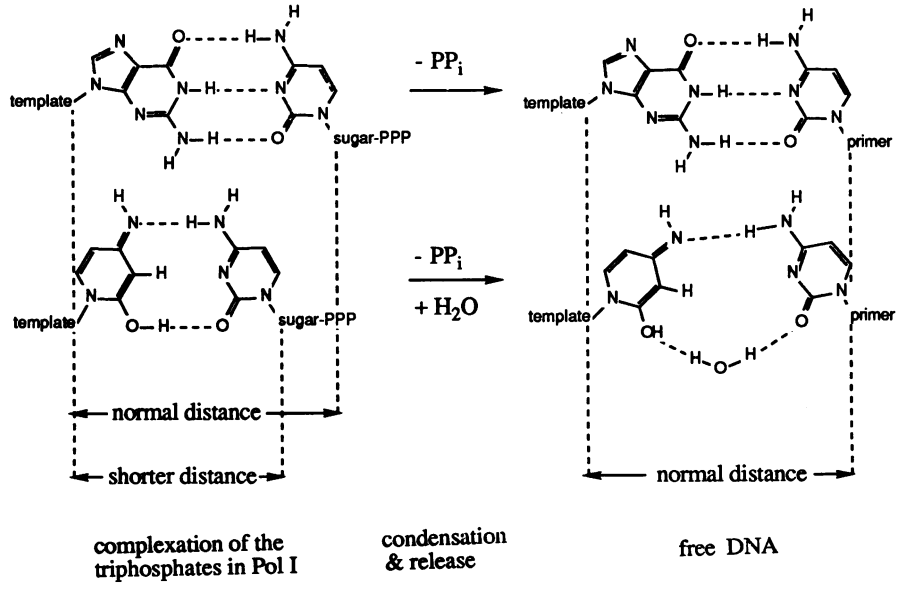
figure 2

complete relaxation of all internal degrees of freedom in the DNA, including the inserted water molecule, a local energy minimum was found resulting in a conformation that resembled qualitatively a normal B-DNA (*figure 1*). The water molecule stayed in the plane of the mismatched base pair and bridged the OC(2) atoms with the tightest H-bonds (O-O distances 2.9 Å, N-N distance 3.0 Å). The base stacking was still relatively close to planar and no tensions occured (compare with a modeled C·T-pair in a double strand [17]).

The highly ordered first hydration shell in the minor groove of B-DNA is thought to stabilize its conformation [18-19], so the additional water molecule may take part in the H-bond network along the minor groove, thus stabilizing the unusual base pair. *Figure 2* shows the isolated pyridine-pyrimidine pair from *figure 1* (C(1')-C(1') separation 10.3Å, $\lambda_{d(3\text{-deazaC})} = 49°$, $\lambda_{dC} = 46°$), once superimposed on C·G and once on G·C (taken from a canonical B-DNA double-helix, Brookhaven Protein Databank, entry code 1DNN). Note that the bridging water molecule lies - in this projection - in the proximity of the 2-amino group of dG, like the water molecules found to be important in hydrating A·T pairs in B-DNA [18-19].

Two Structures of One Mispair

The findings of the computer simulations cannot explain what the conformation of the 13-mer was like in the enzyme pocket, but the structures in *figures 1* and *2* show how unstable (partly mispaired) B-DNA could be stabilized after replication. The **structural inconspicuousness** of DNA is crucial for a mispair to last *in vivo*, otherwise a 5'→3' exonuclease is likely to excize the



scheme 3 incorporation of dCTP into DNA directed by dG and by d(3-deazaC)*'

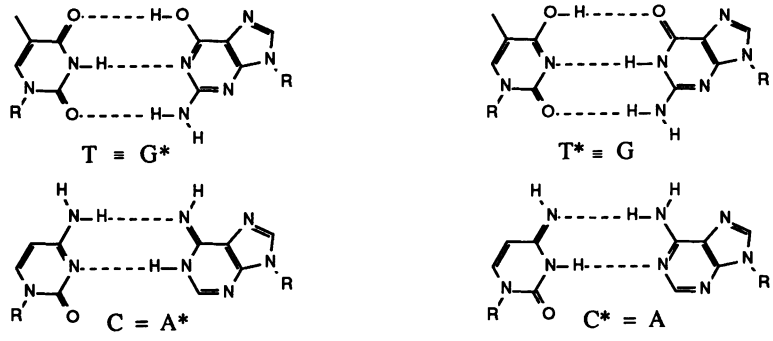
mispair. The findings also suggest that the hydration shell should not be neglected in conformation studies of DNA; its inclusion would re-adjust especially the position of the additional water molecule.

Conclusively, one can assume that the substitution pattern of the H-donor/acceptors of a template nucleotide base is more important for the specificity of Pol I-polymerization than the exact distance between the template and the triphosphate to be incorporated. The distance between the two strands in a free double-helix, however, is very important for the survival of a mispair (*scheme 3*).

MISMATCHES: FORMATION *versus* OCCURRENCE

Tautomeric Mispairs

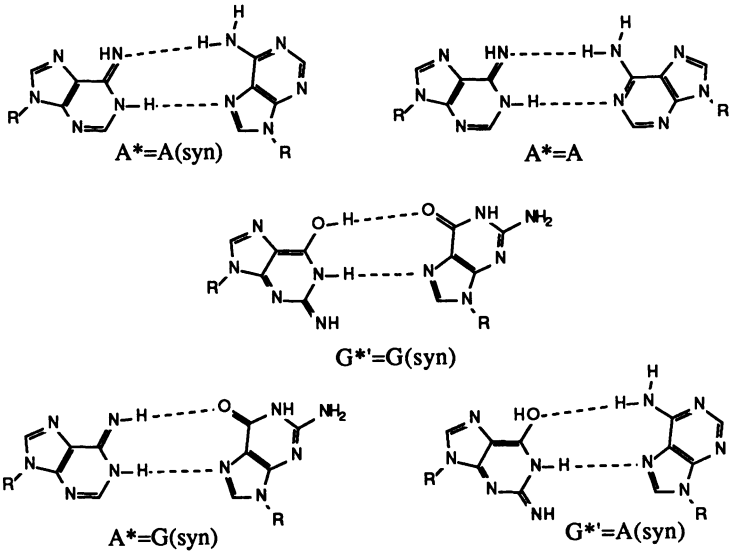
Pyrimidine-Purine Mispairs. As already proposed by Watson & Crick in 1953 [20] and later by Topal & Fresco [21], in principle, dG can base-pair with dT, and dA with dC, respectively. Tautomerization of one of the bases is necessary, where dG* and dC*, though less stable than dA* and dT*, are thought to be more probably formed [22] (*scheme 4*).



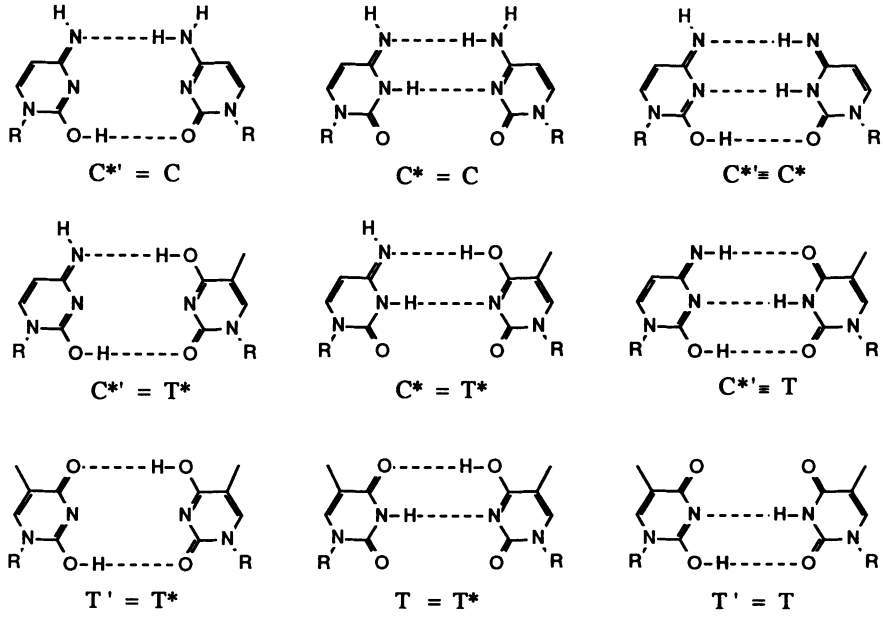
scheme 4

Purine-Purine Pairs. Topal & Fresco [21] proposed that tautomeric Hoogsteen (*anti-syn*) pairs were responsible for the formation of purine-purine pairs. The *anti-anti* pair A*=A has been considered by Drake & Baltz [23] (*scheme 5*).

Pyrimidine-Pyrimidine Pairs. Starting from the findings of our biochemical experiment with the mispair d(3-deazaC)*'·dC, analoguous pyrimidine-pyrimidine pairs can be built, provided that all tautomeric forms are allowed (*scheme 6*) [3]. The structures of the left column might be less favorable, because of the repulsion between the non-bonding electron pairs of N(3), still, the formation of a d(3-deazaC)*'·dC pair found in our experiment revealed that a pair like C*'=C is possible in the enzyme, assumed that the free electron pair of N(3) is comparable in dimension with a C-H bond. All the other possibilities should be quite stable, provided that the distance



scheme 5



scheme 6

between the corresponding sugar moieties can be easily narrowed. In addition, the homogeneous pairs C·C and T·T are equivalent with respect to an inter-base-pair exchange of acidic lactam and lactim protons (equivalent pairs: C*'≡C*⇔C*≡C*'; T=T*⇔T*=T; T'=T⇔ T=T'). A two-proton-transfer in C*'≡T results in C*≡T', a pair of a similar energy content, but if C*=T* is isomerized, probably the most stable pyrimidine-pyrimidine-pair is generated: C=T [3].

Alternatively, Watson-Crick-like mispairs could bear protonated bases as well. In the purine-pyrimidine and in the purine-purine series, protonation of a base pair does not suggest any further stabilization, compared to the merely tautomeric mispairs depicted in *scheme 4* and *5*. There may be some stabilization in protonating a C*'=C-, a C*'=T*-, and a T'=T*-pair (*scheme 6*, left column) due to a third (weaker) H-bond in $[C^{*'}\equiv C]^+$, $[C^{*'}\equiv T^*]^+$, and $[T'\equiv T^*]^+$. Yet, the sometimes preferred interpretation of a protonated species in favor of rare tautomers [25-30] which can explain some mispair formations [4] cannot be true for the d(3-deazaC)·dC-pair, because protonation would not additionally stabilize this pair.

Experimental Evidence for the Structures of Mispairs

Purine-Purine and Purine-Pyrimidine Mispairs. Mispairing between unmodified natural bases in synthetic deoxyribonucleotide poly- and oligomers were investigated by ^{1}H-NMR-spectroscopy [34-40] and some structures were shown by X-ray analysis to be wobble pairs [41-49], *anti-anti* [50] or Hoogsteen pairs [44,48-49]: G·T and A·G, neutral, A·C most probably protonated on N(1) of dA [27-28], all in their normal tautomeric forms.

Pyrimidine-Pyrimidine Pairs. None of the proposed structures of pyrimidine-pyrimidine pairs are yet proven for a DNA duplex. A low-temperature ^{1}H-NMR spectrum of d(CGCGAATTCTCG)$_2$ showed a resonance at 11.7 ppm from the C·T imino proton and led to the proposition of a base pair with Watson-Crick-like geometry lying in the normal tautomeric form and involving a bridging water molecule [39].

The observed pH-dependency of the formation of double helical poly-dC in antiparallel strands led to the proposition of a $C{\cdot}C^+$-wobble pair being protonated on one of the N(3) atoms, and both cytosines lying in the normal tautomeric form [51]. Since the conclusion on the structure of $C{\cdot}C^+$-pairs based on experiments either with poly-C, poly-d(CT), poly-dC or d($C_4A_4T_4C_4$)-oligomer, it need not necessarily be true for $C{\cdot}C^+$-pairs lying in a natural purine-pyrimidine base pair neighbourhood, where strains in the sugar-phosphate backbone are expected to extend the mispair.

Investigations in oligomers with hairpin loop structures that existed in equilibrium with their duplex form also revealed information about T·T-pairs. In these studies, the authors interpreted the ^{1}H-NMR signals of the imino protons allocated to T·T-pairs in the oligomers d(ATCCTATTAGGAT)$_2$ [52], d(ATCCTATTTAGGAT)$_2$ [53], and d(CGCGATTCGCG)$_2$ [54] as being derived from either wobble pairs or "open pairs" (no H-bonds). Very recently [55], the sequence d(CGCGCGTTTTCGCGCG) has been crystallized in a hairpin loop structure (2.1 Å resolution) which showed an unusual T·T-pairing, due to crystal packing forces. Since this intermolecular

mispair (between two hairpin loop residues) had a close Watson-Crick-like geometry (O^4-O^4-separation 2.5 Å, N(3)-N(3)-separation 2.6 Å), the authors concluded that the pair contained one thymine in its O^4-enol form (T·T*). This result is the first crystallographic evidence for the existence of a rare tautomeric form in a DNA base pair.

Structures of Mispairs Inside and Outside Replication Enzymes

The Wobble Pair / Tautomeric Pair Conflict. The structures of mispairs found in DNA crystals and solutions lead to a wobble pair/tautomeric pair conflict. The controversy seems to be nourished by the fact that the difference between the conditions during the generation of a mispair and after its formation was often ignored in the literature. As in our model with d(3-deazaC)*', it is very important to differentiate between the structure of a mispair in the polymerase and after its release into solution.

There are two experiments in the literature concerning the structure of mispairs in a polymerase giving evidence that, down to a certain minimal misincorporation frequency, no formation of wobble pairs in a polymerase was measurable (*i)* $\leq 4 \times 10^{-5}$ dGMP per dAMP incorporated by *E. coli* Pol I and directed by oligo-$rU_{54\pm11}$ as a template and primed with oligo-dA_{10} [14]; *ii)* no incorporation of dGMP or dAMP by *E. coli* Pol I directed by d(3-deazaC) in the template within the limits of detection of an autoradiogramme after Maxam-Gilbert sequencing [8]). On the other hand, a number of misincorporation experiments including all sorts of mipairs partly in astonishing high abundancy (*vide infra*) can neither disprove the above evidence, nor prove unequivocally the formation of wobble in favor of tautomeric mispairs during the polymerization.

This indicates that the formation of wobble pairs in a polymerase must be quite rare and therefore might be responsible for a mutational background of DNA, but cannot explain unusually high mispairing frequencies.

The Structure/Function Argument. Another argument is derived from selective constraints that are thought to be imposed on polymerases, in order to guarantee correct replications. Since the fidelity of replication depends strongly on the matching of H-bonds, and thus on the precise positioning of the triphosphate and the template bases towards each other, one does not expect a wobbling (relative to the N(3)-N(1)-axis of a Watson-Crick-pair) to be allowed in the nucleoside triphosphate binding site of an enzyme such as Pol I (*cf.* [14,56]), so the reason for the creation of a mispair cannot be derived *a priori* from the structures of the wobble pairs found in free, unbound DNA duplexes.

Wobble base pairs were first postulated to exist in the third base pair of the triplet codon in mRNA-tRNA complexes [24] allowing the recognition of synonymous codon triplets by the same tRNA which contributes to the degeneracy of the universal code. But there the situation is different because of the less rigid secondary structure of the anticodon loop in tRNA's needed for proper functioning.

Elucidation of Functional Analogies and Mechanisms of Mispairing

Structural Principles. Following the model picture depicted in *scheme 3*, the patterns of the

H-donor/acceptor substitution of pyrimidine bases required for all nucleotides to be incorporated, **in principle**, dC* appears as a, both, dT- and dG-functional analogue (C*·A in *scheme 4*; C*·C in *scheme 6*), dC*' as a dG- and a dA-functional analogue (C*'·C, C*'·T in *scheme 6*), dT* as a dC- and a dA-functional analogue (T*·G in *scheme4*; T*·T in *scheme 6*), and dT' as a dA-functional analogue (T'·T in *scheme 6*). As for the tautomeric purine bases: dA* appears as a dG-, dT- and a dC-functional analogue (A*·C in *scheme 4*; A*·A, A*·G in *scheme 5*), dG* as a dA-functional analogue (G*·T in *scheme 4*), and dG*' as a dC- and a dT-functional analogue (G*'·G, G*'·A in *scheme 5*).

Note that the term 'functional analogue' refers to bases with an altered hydrogen substitution pattern that can base-pair in the enzyme pocket with natural, normally hydrogen-substituted bases. (An additional way of forming a mispair where both bases are involved as rare tautomers will be presented further on.) **In reality**, i.e. after performing the appropriate experiments *in vitro* and *in vivo*, some of the functional analogies might have to be excluded or complemented with other possibilities.

Experiments with Structure Analogues of Rare Tautomers. In order to obtain experimental evidence for the structures depicted in the *schemes 4* , *5* and *6* being responsible for the formation of mispairs in the polymerase, not the crystallographic analysis of mispairs in free, double-helical DNA should be envisaged, but rather mutagenesis experiments with residues that are capable of mimicking rare tautomeric forms.

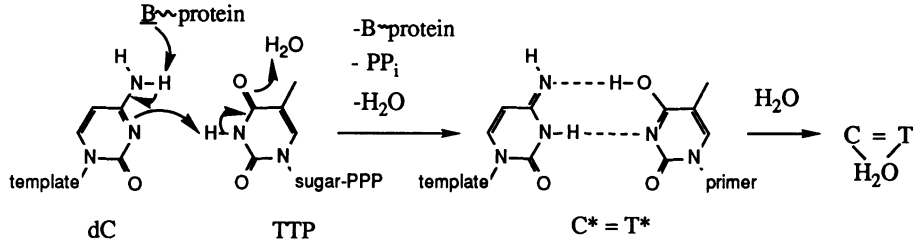
The behaviour of dC*' in a DNA template during replication can be understood by using d(3-deazaC) instead (*cf.* first section), because the only reactive form of d(3-deazaC), i.e. d(3-deazaC)*', can be regarded as a nucleoside model compound for dC*' in which the tautomeric form is fixed. This means that *in vitro* (and perhaps *in vivo*) dC*' in a template should bind preferentially dCTP during replication.

As for the alternative rare tautomeric form dC*, already in the early 1970's mutagenesis experiments showed a similar functional analogy derived from another modified residue. N^4-Hydroxycytidine (N^4-OH-C) in a RNA template was found to preferentially direct the incorporation of ATP [57-58]. Alternatively, d(N^4-OH-C)TP was incorporated on template-directed polymerization, both as dC and dT strongly dependent on the sequence [59]. Since N^4-OH-C lies wholly in the imino form [60] and therefore can be regarded as a model compound for C*, the experiments suggest that C* acts rather as an U- than a G-functional analogue and dC* is rather a dT- than a dG-functional analogue.

The behaviour of dT* [5], dT', dG*, dG*' and dG' on replication also remains to be investigated by means of *in vitro* site directed mutagenesis with similar nucleotide model compounds [6].

Possible Mechanisms of Formation Involving Rare Tautomers. There are situations imaginable where initially both, template and triphosphate bases, are temporarily in their rare tautomeric forms during replication. E.g.: While Pol I polymerizes nucleoside monophosphates between the *Okazaki* fragments, a DNA-binding protein would be capable of deprotonating with a basic

functional group a very exposed and yet unpaired dC-unit in the template strand. If the presence of the now protonated DNA-binding protein can prevent a reprotonation of the imino group, either N(3) or OC(2) can be protonated by water, leading to dC* or dC*', respectively. dC* is expected to bind preferentially to dATP and dC*' to dCTP. Yet, if the process described happened in the active site of Pol I (which seems to be possible when looking at the X-ray structure of the Klenow fragment [68]), the presence of unspecifically bound dTTP in the triphosphate binding site of Pol I may cause the protonation of N(3) of the anion of 4-imino-dC by itsself (i.e. by its own N(3)-proton), in the place of water, and would lead to a C*=T*-pair which could stabilize itsself by rearranging into a C=T-pair (*scheme 7*).



scheme 7

If the attacked amino group belonged to a dA moiety, dA* might result which in turn could bind to dCTP, dATP or dGTP. The binding preferencies of dA* are unknown. In the alternative case of a DNA-binding protein with an acidic functional group, OC(4) of a dT-unit could be protonated and would generate dT* (dT^{+}* $\rightarrow$ dT* + H^{+}). Again, whether dT* binds to dGTP or to dTTP is hard to say, because no unequivocal experimental evidence is available. If the proton release happened in Pol I, perhaps dCTP might be protonated and base-paired (analogous to *scheme 7*). DNA-binding proteins could generate dG' or dG*' and dT' leading to mispairs in a similar fashion.

According to such a mechanism one does not expect to favor the formation of pyrimidine-pyrimidine mispairs by biasing the concentrations of deoxynucleotide triphosphates during (*in vitro* and *in vivo*) replication [69-71], since the mispairing is thought to be provoked by an "anomalous behaviour" of the template (in agreement with Fersht *et al.* [79]). One does expect a certain difficulty in enhancing the formation of purine-purine mispairs due to their bulkiness and the limited space in Pol I [5], so the formation of purine-pyrimidine mispairs (i.e. of transition mutations) may predominate in high concentrations of the corresponding triphosphates. (The proposed mechanism for spontaneous transitions and transversions - mispairing of normal bases due to a reversible slippage of the template strand and looping out of one or several bases during replication [73-74] - are excluded [23], because no experimental evidence for such movements of a template in the active site of polymerases was found.)

A mechanism according to which deprotonated bases in the template strand lead directly to the formation of mispairs [75] will require - for the natural bases - the same conditions *in vivo* as described above, since the reprotonation of an anionic base (toward either the normal or the rare tautomeric form) should proceed rapidly at a physiological pH.

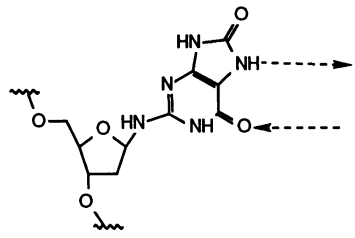
Replication and Incorporation Errors. In any case, neither **rare** tautomers (*cf.* first section) [6], nor anionic bases [75] misincorporate as monophosphates, yet, if the "unusual" tautomeric form is not rare, as in the case of d(N^4-OH-C)TP, misincorporations may occur [59]. The reason for this, possibly, is because polymerase-bound triphosphates occur only in their most probable form (which contrasts in a sense to the pathway depicted in *scheme 7*), meaning that **labile** functional analogues (rare tautomers and ionized bases) can only provoke replication errors and no incorporation errors.

Experimental Evidence for the Formation of Pyrimidine-Pyrimidine Pairs

Low But Measurable Abundance. In spite of pyrimidine-pyrimidine pairs being excluded from the theory of substitution mutations by Topal & Fresco 1976 [21], the spontaneous formation of C·T-pairs was first found in 1968 after replicating poly-dC templates with T4 DNA polymerase *in vitro* (T/G incorporation 10^{-5} to 10^{-6}) [76]. The formation of C·T-pairs occurred also on polymerizing poly-dA·poly-dT and poly-dG·poly-dC homopolymers with T4 DNA polymerase [77]. The measured turnovers of the uncomplementary nucleotides suggested, besides an unexpectedly favorable formation of C·T-pairs, that especially pyrimidine bases in the template provoke mispairing. More recently, the formation of C·T- and T·T-pairs after *in vitro* polymerization with polymerase α holoenzyme, due to misinsertion of deoxynucleoside triphosphates, was demonstrated using a natural template [78].

High Abundance. The formation of pyrimidine-pyrimidine mismatches *in vitro* was verified by a biochemical experiment similar to ours [79]. In this case, a modified guanylate residue was responsible for the yet unexplained mispairings generated in its nearest neighbourhood. The replication of two 2'-deoxy-8-hydroxyguanylate (d(8-OH-G)) containing 43- and 46-mer template strands and a 15-mer primer by Pol I (Klenow fragment) lead to erroneous replications. Several bases opposite d(8-OH-G) were found after the chain elongation indicating several (at least 2) isomeric forms of d(8-OH-G) being present in a thermodynamic equilibrium. The occurrence of dC paired with d(8-OH-G) indicated that d(8-OH-G) lay partly in its "trivial" N(9)-glycosidic *anti*-conformation; a shift of the *anti-syn*-equilibrium of d(8-OH-G), though, can be deduced from the presence of purine bases paired with d(8-OH-G). The formation of purine-purine (Hoogsteen?) pairs could yet only be observed in the strand where d(8-OH-G) was flanked by two pyrimidine bases. Finally, the presence of dT paired with d(8-OH-G) not only confirmed the assumption that d(8-OH-G) was mainly present in its 8-keto-7-NH-form [22] - the only H-donor/acceptor substitution pattern to base-pair with dT, but it also suggested alternatively to the N(9)-glycosidic *syn*-conformation that a N(9)→N^2-migration partly may have occurred,

facilitated by the 8-keto function, leading to 2'-deoxy-8-oxo-7*H*-neoguanylate (d(8-OH-neoG)) [80-82,23] (Proposed by Th. Bickle).



The finding of relatively highly abundant C·C-, C·T-, and T·T-pairs in the nearest neighbourhood of d(8-OH-G) suggests that pyrimidine-pyrimidine pairs play a crucial role in replication-dependent transversion mutations after all. dC showed, above all, to be prone to mispairing, while e.g. dA as a neighbour of d(8-OH-G) directed only matched base-pairing with dT.

Measured Acidity Indicates Tautomerizing Action. An explanation for such mispairs can be found in the different protolytic behaviour of d(8-OH-G). The enhanced acidity of d(8-OH-G) (pK(d(8-OH-G)-nucleoside) = <1.2; 8.5; 11.8 [7] / pK(dG-nucleoside) = 2.2; 9.3 [7] / d(8-OH-neoG) is expected to have a still lower pK [81]) - it was the most acidic base of the template - could have caused an eased tautomerization of a neighbouring dC-, and dT-unit, due to the protonation of OC(2) or N(3) of dC, and OC(2) or OC(4) of dT by d(8-OH-G), if not directly, then mediated by water molecules. With this assumption, not only the reason for the pyrimidine-pyrimidine-mispairing can be found, but also the reason for the pyrimidine-purine-mispairing C·A and T·G directed by a pyrimidine base in the nearest neighbourhood to d(8-OH-G). Additionally, only the formation of Watson-Crick-like mispairs that, in part, necessarily must have contained rare tautomeric forms (C·C and T·T) could explain the complete lack of proofreading of the nearest neighbour mispairs in this experiment.

Structure of Mispairs After Formation

Wobble and Hoogsteen Pairs? After being formed, proofread and released into solution the mispairs will change their structures due to the sudden absence of "constraints on substrate-template interactions" imposed by the enzyme [84]. So the mispairs may return to the normal tautomeric forms, and as the X-ray data suggest, some will be protonated and most of them will form wobble pairs. At present, the structures of mispairs after replication can only be proven by an X-ray analysis, since the interpretation of spectroscopic data of some mispairs in synthetic oligomers may be quite difficult. On the other hand, there is still some incertainty left about the crystal structures, since a double-helix may crystallize in another form than the major conformer is in solution. UV-spectroscopical data strongly suggest that G·T- and A·C-pairs can adopt a

Watson-Crick geometry and contain rare tautomeric forms (*scheme 4*) if stabilized appropriately [56].

Thermodynamic Stabilities. Evaluation of the relative stabilities of natural mispairs derived from the melting curves of synthetic oligomers led to a succession of mispairs, where G·T was unambiguously the most stable pair, followed by a "middle field" of G·G, G·A, C·T, T·T, and A·A within a wide range of tolerance, and finally C·C and, A·C as the weakest pairs [85-86]. The authors proceeded from the assumption that they were dealing with wobble pairs, intermediate or weak base pairs, and open base pairs, all in their normal tautomeric forms. In another study of mispair stabilities the succession was less clear due to strong salt and sequence dependencies [87].

Stability and Specifity of Incorporation or Recognition. However it may be, the biological relevance of the thermodynamic stability of mispairs in a free DNA double helix is questionable, because *in vitro* experiments with replication enzymes showed that little or no correlation to the stabilities could be observed, neither by misincorporation experiments with prokaryotic DNA polymerase [29,88] or eukaryotic DNA polymerase α [84], by the determination of misincorporation proofreading efficiencies of T4 DNA polymerase [89], nor by mispair correction experiments using DNA polymerase I, III [90-91], III-holoenzyme [70], and cell free extracts [92] which should simulate *in vivo* conditions best. Additionally, both, the proofreading efficiency measurements with T4 DNA polymerase [89] and kinetic misincorporation experiments with strongly biased triphosphate pools in different replication systems [93-94] showed a high template sequence-dependence.

Although the investigations in mispair repair efficiencies *in vivo* [95-98] must be compared with caution, since they were measured in different organisms, involved replicative and/or only post-replicative synthesis [95], and the efficiency depended on the neighbouring base pairs, a succession of mispairs could be deduced where C·C, C·T, and A·G were not, or only inefficiently repaired, while A·A, G·G, and in particular A·C, and G·T were efficiently repaired. The repair of T·T-mispairs varied from efficient [96] over intermediate [95,98] to inefficient [97]. Recent *in vitro* mispair specifity measurements of methyl-directed DNA mismatch corrections by the *mut H*, *mut L* and *mut S* gene products were consistent with the *in vivo* results [99].

Proofreading experiments with *Drosophila melanogaster* DNA polymerase-primase and its 3'→5' exonuclease subunit revealed a marked difference in excision selectivities of three mispairs (A·A, A·G and A·C) located at the 3'-terminus of the primer, dependent on whether the excision was performed in the presence or absence of concomitant polymerization [100]. In fact, for the first time a positive correlation of the excision selectivity of an (isolated) 3'→5' exonuclease and the thermodynamic stabilities of the mispairs in free DNA was found. Since in the presence of the polymerase-primase this selectivity was altered drastically, the authors concluded that the ability of the polymerase to extend a specific mispaired primer is a major factor that contributes to the overall efficiency of proofreading. Similar conclusions were drawn on sequence-dependent misincorporation experiments with the Klenow fragment of Pol I [101].

Kinetics of Polymerization

The kinetics of the polymerization reaction following a mispair which becomes rate-limiting must be dependent on the stability - and thus on the structure - of the mispair **in the polymerase**. All the above experiments showing the absence of correlation between the thermodynamic stabilities of mispairs and the misincorporation and proofreading experiments are further support for the idea that the structure of mispairs in free DNA (most of which are shown or assumed to be wobble pairs) cannot be the same as in the moment of their generation or proofreading in a polymerase. In view of the overall excision repair selectivities one must also conclude that the selectivities of most post-replicative repair systems do not depend on the stabilities of the mispairs in DNA, but rather on the degree of structural similarity to the normal Watson-Crick pairs in their neighbourhood. The precise parameters for the recognition of the differences between matched and mismatched pairs remain to be elucidated.

SUBSTITUTION MUTATIONS *IN VIVO*

Frequency of Transitions and Transversions

The Structural Inconspicuousness of Mispairs. Extracellular mutagens, mutagenic cell metabolites, defective (mutated) polymerases and repair systems, or mutator genes can eventually cause replication-dependent gene mutations. The "lifetime" of a G·T-hybrid, the critical parameter for a successful transition mutation, seems to be rather insufficient considering the *in vivo* mispair-repair experiments. The same goes for the other transition mispair A·C, while the transversion mispairs G·A and the pyrimidine-pyrimidine pairs seem to mimic a Watson-Crick-pair quite well.

Transitions and transversions were found in various *in vivo* experiments [102-106] or in homology studies [107-108], but, according to the inconspicuousness of their corresponding mispairs, only transversions seem to coincide with the theory: the majority of transversions found were G·C→T·A-, and A·T→C·G-transversions [102-103,107-108] which is plausible, because both corresponding mispairs (C·T and G·A) were shown to be inefficiently repaired *in vivo*. Less frequent were G·C→C·G- [102,104-105, 107-108], and A·T→T·A-transversions [106-108], while transitions were generally the most frequently found substitutions [107-108].

Neutral and Constrained Substitutions. One has to bear in mind that transitions are less selected against [109], because more transitions than transversions are silent (not amino acid replacing), so that in order to observe only substitutions imposed by mutation pressure without the effect of selective constraints, DNA sequences or substitution sites without any phenotypic effect have to be analyzed, or, at least the phenotypic effects must be kept small, e.g. by omitting any competition between mutant strains and by using short time periods for growth. The above experiments meet these criteria and still the transitions are generally the most abundant substitutions found.

Additionally, the formation of transitions can be explained by processes involving chemical transformations of a base in the DNA duplex **between** the replications. 2'-Deoxy-5-methylcyti-

dylate which is present in the whole genome in sufficient amounts (particularly in d(...pCpG...) sequences) could suffer from deamination to thymidylate [110-112] shortly before replication,and would generate a T·A-pair after replication. The reversed mutation, the A·T→G·C-transition could be initiated by the spontaneous C(6)-deamination of an adenylate moiety into inosinate shortly before replication which in term would pair preferentially with dC [113] and generate a G·C-pair after an excision repair or a second replication.

In the contrary case where a strong (positive) selection pressure is imposed on a sequence, as in a proto-oncogene, the major part of substitutions found were transversions, again in the same order of frequencies. 15 out of 21 transforming *ras*-oncogenes (including Human EJ/T24-Ha-*ras*1 [114-115]) resulted from at least one transversion in one of the crucial positions in the proto-oncogene. Again the predominant transversion was C·G→A·T (12 out of 17), followed by 3 G·C→C·G- and 2 A·T→T·A-transversions [116-118].

The Probability of Formation. Obviously the structural inconspicuousness of mispairs is not the only determinant for the frequencies of substitution mutations, but rather is overlapped by the ease or probability of formation during replication (polymerization and proofreading). Transition mispairs seem to be much more easily formed than purine-purine or pyrimidine-pyrimidine pairs.

The kinetic mechanism of mutagenesis [119] can explain the formation of mispairs due to varying K_m and (to a lesser extent) V_{max} values of nucleoside triphosphates in polymerase-primer-template-triphosphate complexes. The transition mispairs were shown to be more probably formed *in vitro* by polymerase α holoenzyme (no proofreading) due to lower dissociation rate constants of the triphosphates in purine-pyrimidine complexes as compared with those in pyrimidine-pyrimidine complexes [78]. Also *in vivo* mutagenesis experiments in prokaryotes showed that transitions were formed much more often than transversions in systems lacking proofreading [120] or mismatch repair activity [121], whereas in intact strains transitions did not predominate [122]. But while transitions were typically found equally distributed over all mutation sites, hot spots were often the preferred *loci* for transversions [120].

Sequence-Dependence of Substitution Mutations

Hot Spots. The occurrence of spontaneous mutational hot spots is not sufficiently explained by a model mechanism where only the misinsertion of nucleoside triphosphates (in their normal tautomeric forms) due to uncomplementary "base-pairing" [123] with the template bases (in their normal tautomeric forms) is assumed. It is unlikely that a locally preferred mispairing leading there to highly frequent substitution mutations could only be governed by special base stacking interactions with neighbouring base pairs [71,93,77] in a polymerase, or by the polymerase itsself, due to strong alterations of the stability of mispairs [78] in the active site (as compared to the stability of mispairs in free DNA) through the exchange of solvating water molecules by amino acid moieties derived from the polymerase.

Yet, a mispair containing a rare tautomeric form, thus bearing a more Watson-Crick-like geometry, would therefore enhance its probability of polymerization (passing the verification step, *cf.* first section), would prevent from being excized (passing the proofreading step) and would

accelerate the polymerization step of the following nucleotide [100]. A high sequence-dependence of such a mutation would be only partly derived from the stabilization of the mispair by base stacking interactions. The other part should be searched in the sequence specifity of protonating or deprotonating agents that could induce the tautomerization of a template base.

Sequence-Dependence of Mutagens. Sequence-dependent transversions and in part transitions require a very specific process for its beginning. It seems to be necessary that a mutagen, no matter if it is a cell metabolite or not, does not cause too severe a change in the nucleotide of the genome, so the defect can escape the repair system. "Chemically" the known mutagens often provoke striking changes leading to mutations *in vitro*.

The presence of oxygen radicals e.g., an unselective mutagen that causes among others transformations of dG-units into d(8-OH-G)-units in the genome, is suspected to lead to replication-dependent mutations in its nearest neighbourhood *in vivo* [79] - if it would not as well cause severe distortions in the B-DNA helix. But, provided that the repair system is efficient enough, it can be assumed that a d(8-OH-G) containing sequence preferentially would be excized together with its neighbouring mispairs by a 5'→3' exonuclease such as Pol I. Conclusions from the *in vitro* Pol I replication experiment with d(8-OH-G) containing templates to *in vivo* conditions should be drawn with caution, because single-stranded DNA-binding proteins that facilitate DNA replication [5,124-125] by accumulating cooperatively ahead of the polymerase, immobilize the strand by unspecific binding to the backbone and fix the bases by intercalation with aromatic amino acid moieties, are known to enhance the replication fidelity [5,126]. Sequence-dependent differences in the base-stacking interactions that are thought to contribute to the occurrence of mutational hot spots should be levelled *in vivo* by such protein-DNA interactions. Since rare tautomeric forms of the bases depend on their stabilization by favorable stacking interactions [56], possibly these proteins could influence the formation of nearest neighbour mutations.

Accessibility of Bases. Another requirement for a specific mutation is the structure of DNA which has to bear features that make it accessible to a "mild" mutagen. The gaps between the *Okazaki* fragments, if not sufficiently blocked by single-stranded DNA-binding proteins, could be more accessible to agents than usually (similar to the mutagenicity of N-methyl-N-nitrosourea towards double-stranded and replicating DNA [127]). During replication, the formation of secondary structures in the single-stranded sequences could also be taken into account, because in particular the bases in loops are strongly exposed to the solvent. As for duplexes, a Z-like conformation is known to bear its bases near to the hydration shell [128] so e.g. the 4-amino group of dC could be deprotonated easier, leading to dC* or dC*' on reprotonation of N(3) or OC(2), resp. (*cf.* third section). A longer sequence of non alternating A·T-pairs (starting from 4 base pairs) is known to cause a slight bent in the helix axis [129]. During the formation of a pyrimidine-pyrimidine-mispair a bent could be taken into account as well, in order to compensate the shortened distance between the sugar-phosphate backbones. Thus, the helix axis may perhaps influence the formation of mispairs, the more so as the 13-mer used as a template (*scheme 1*) bore a sequence of 4 non alternating A·T-pairs.

Tauterogens. Therefore, a specifically DNA protonating or deprotonating mutagen would serve to trigger substitution mutations. Although the equilibrium constants for the **spontaneous** tautomerization of the natural nucleoside bases are too low or too high as compared to the frequency of spontaneous mutations *in vivo* [73,130-131], and although tautomeric forms are not believed to be involved in substitution mutations because of the too high tautomerization energies [132], it should be possible to lower the activation barrier by e.g. protonation through a nearby acidic functional goup or by complexation with a *Lewis*-acid (e.g. Zn^{2+}) and subsequent deprotonation with a nearby basic functional group. As proposed in the third section, DNA-binding proteins (or some other smaller DNA-binding peptides) seem to be obvious candidates for DNA protonating or deprotonating, in the end tautomerizing mutagens, because they provide a variety of functionalities and may be very DNA sequence-specific.

The so formed tautomers could then be stabilized by solvatation and base pairing [55,131] and possibly by favorable base stacking [56,77], so that the **effective** tautomeric equilibrium would be markedly shifted during replication. Some evidence has already been found that rare tautomers of C-derivatives in favor of ionized (protonated) C-derivatives may be involved in mutagenesis of RNA, leading to C·G→U·A-transitions [133,58]. The proposed mechanism of formation and the extraordinary structural inconspicuousness of pyrimidine-pyrimidine-pairs might explain why transversions can readily happen at special positions in the genome.

CODA

Aim

By this study I would like to provoke discussions about the involvement of rare tautomeric forms in DNA leading to the **formation** of mispairs, in comparison to the occurrence and structure of mispairs after their formation, because, in spite of the available experimental results that support this theory, no experimental approach was yet presented that might hold out a prospect of proving that mutagenesis does or does not depend on such a mechanism *in vivo*.

The second aim, in turn, is to direct the researcher's attention to the possibility of the existence of "mild" mutagens that do not alter the structure of the DNA "chemically" (such as alkylating agents etc.), but merely change the hydrogen substitution pattern of DNA bases temporarily ("tauterogens") and therefore could have escaped any observation.

Acknowledgements

I gratefully thank Prof. Ch. Tamm (Institut für organische Chemie, Basel) for enabling and helping me to write this work, Prof. Th. Bickle for his cooperation (mutagenity tests, manuscript corrections), Dr. K. Müller for carrying out the computer modeling study and Prof. H. Sigel for the pK measurements and their advice. This investigation was supported by the Swiss National Science Foundation.

1) * stands for rare tautomeric form on C(4)-substituents of pyrimidines (pyridines) and C(6)-substituents of purines. ' stands for rare tautomeric form on C(2)-substituents.

2) The nucleotides 5'-d(pTpTpApGpC)-3' and 5'-d(pGpCpTpApA)-3' were taken from Brookhaven Protein Databank - entry code 1 DNN [16] and their sequences were modified.
3) Some of the structures were proposed by Drake & Baltz [23], C=T in analogy to C=U, proposed by Crick [24].
4) Such as the unexpected dC- (and partly dA- [31]) selectivity of interstrand-crosslinking 2'-deoxy-N^4,N^4-ethano-5-methylcytidylate (C^e) [31-32] and 2'-deoxy-N(6),N(6)-ethano-adenylate (A^e) [33] containing DNA strands, due to N(3)- and N(1)-protonation of C^e and A^e, respectively.
5) *In vivo* site directed mutagenesis induced by O^4-methylthymidylate containing templates led to the formation of O^4-methylT·G-pairs [61] which were proposed to base-pair in a Watson-Crick-like manner [62]. O^4-methylthymidine is not a very good model compound for dT*, because it lacks the O^4 proton that is "more needed" in T*=T than in T*≡G.
6) 6-Imino-1-methyladenine [63], (6-imino-)1-methyladenosine, O^6-methylinosine, O^4- ethyl-uridine, (4-imino-)3-methylcytidine [56], O^6-methyl-2'-deoxyguanosine [64-65], 5-aza-7-deaza-2'-deoxyguanosine (2-imino form in water) [66], 4-imino-N^4-hydroxy-cytidine, 4-imino-N^4-methoxy-cytidine [60] are known, so is 4-imino-1-methylcytosine as a platinum complex [67], while neither derivatives nor analogues of dG*' and dT' were to our knowledge yet characterized.
7) The d(8-OH-G)-nucleoside was synthesized by P.Strazewski, as described in the literature [83]. The pK values were measured by H. Sigel (unpublished results). For exact values and conditions: Prof. H.Sigel, Institut für anorganische Chemie, Spitalstrasse 51, 4056 Basel.

+Present address: University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, UK

REFERENCES

1. Kornberg, A. (1980) DNA Replication, Freeman W.H. & Co., San Francisco.
2. Kornberg, A. (1982) 1982 Supplement to DNA Replication, Freeman W.H. & Co.,San Francisco.
3. Kornberg, A. (1984) Adv. exp. Med. Biol. **179**, 3-16.
4. Hübscher, U. (1983) Experientia **39**, 1-126.
5. Loeb, L.A., Kunkel, Th.A. & Schaaper, R.M. (1980) In Alberts, B. & Fox, C.F. (eds), Mechanistic Studies of DNA Replication and Genetic Recombination, Academic Press, New York, pp.735-751.
6. Wachtl, M., Kohler, P. & Tamm Ch. (1980) Helv. Chim. Acta **63**, 2495-2502.
7. Cerletti, N.C. (1976) Dissertation, University of Basle.
8. Charczuk, R., Tamm, Ch., Suri, B. & Bickle, Th. (1986) Nucl. Acids Res. **14**, 9530.
9. Charczuk, R. (1986) Dissertation, University of Basle.
10. Bloch, A. & Dutschman G. (1973) J. Med. Chem. **16**, 294-297.
11. Barlin, G.B. & Pfleiderer, W. (1971) J. Chem. Soc. (**B**) Phys. Org. 1425-1432.
12. Strazewski, P.(1988) Nucl. Acids Res. **16**, 5191.
13. Bloch, A. & Wang, M.C. (1972) Biochem. Pharmacol. **21**, 1063-1073.
14. Ferrin, L.J. & Mildvan, A.S. (1986) Biochemistry **25**, 5131-5145.
15. Mizrahi, V., Henrie, R.N., Marlier, J.F., Johnson, K.A. & Benkovic, S.J. (1985) Biochemistry **24**, 4010-4018.
16. Sussman, J.L. & Trifonov, E.N. (1978) Proc. Natl. Acad. Sci. USA **75**, 103-107.
17. Keepers, J.W., Schmidt, P., James, Th.L. & Kollman, P.A. (1984) Biopolymers **23**, 2901-2929.
18. Drew, H.R. & Dickerson, R.E. (1981) J. Mol. Biol. **151**, 535-556.
19. Kopka, M.L., Fratini, A.V., Drew, H.R. & Dickerson, R.E. (1983) J. Mol. Biol. **163**, 129-146.
20. Watson, J.D. & Crick, F.H.C. (1953) Nature **171**, 964-967.
21. Topal, M.D. & Fresco, J.R. (1976) Nature **263**, 285-289.
22. Pullman, B. & Pullman A. (1971) Adv. Heterocycl. Chem. **13**, 77-159.
23. Drake, J.W. & Baltz, R.H. (1976) Ann. Rev. Biochem. **45**, 11-37.
24. Crick, F.H.C. (1966) J. Mol. Biol. **19**, 548-555.

25. Hunter, W.N., Brown, T., Anand, N.N. & Kennard, O. (1986) Nature **320**, 552-555.
26. Hunter, W.N., Brown, T. & Kennard, O. (1987) Nucl. Acids Res. **15**, 6589-6606.
27. Sowers, L.C., Fazakerley, G.V., Kim, H., Dalton, L. & Goodman, M.F. (1986) Biochemistry **25**, 3983-3988.
28. Kalnik, M.W., Kouchakdjian, M., Li, B.F.L., Swann, P.F. & Patel, D.J. (1988) Biochemistry **27**, 100-108.
29. Eritja, R., Kaplan, B.E., Mhaskar, D.N., Sowers, L.C., Petruska, J. & Goodman, M.F. (1986) Nucl. Acids Res. **14**, 5869-5884.
30. Sowers, L.C., Fazakerley, G.V., Eritja, R., Kaplan, B.E. & Goodman, M.F. (1986) Proc. Natl. Acad. Sci. USA **83**, 5434-5438.
31. Webb, Th.R. & Matteucci, M.D. (1986) Nucl. Acids Res. **14**, 7661-7674.
32. Webb, Th.R. & Matteucci, M.D. (1986) J. Am. Chem. Soc. **108**, 2764-2765.
33. Webb, Th.R. & Matteucci, M.D. (1987) Tetrahedron Lett. **28**, 2469-2472.
34. Early, Th.A., Olmsted, J., Kearns, D.R. & Lezius, A.G. (1978) Nucl. Acids Res. **5**, 1955-1970.
35. Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Dallas, J., Itakura, K. & Breslauer, K.J. (1982) Biochemistry **21**, 437-444.
36. Kan, L.S., Chandrasegaran, S., Pulford, S.M. & Miller, P.S. (1983) Proc. Natl. Acad. Sci. USA **80**, 4263-4265.
37. Patel, D.J., Kozlowski, S.A., Ikuta, S. & Itakura, K. (1984) Biochemistry **23**, 3207-3217.
38. Patel, D.J., Kozlowski, S.A., Ikuta, S. & Itakura, K. (1984) Biochemistry **23**, 3218-3226.
39. Patel, D.J., Kozlowski, S.A., Ikuta, S. & Itakura, K. (1984) Fed. Proc. Fed. Am. Soc. Exp. Biol. **43**, 2663-2670. For C·T see also [77].
40. Hare, D., Shapiro, L. & Patel, D.J. (1986) Biochemistry **25**, 7445-7456.
41. Brown, T., Kennard, O., Kneale, G. & Rabinovich, D. (1985) Nature **315**, 604-606.
42. Ho, P.S., Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H. & Rich, A. (1985) The EMBO Journal **4**, 3617-3623.
43. Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. (1985) J. Mol. Biol.**186**, 805-814.
44. Kennard, O. (1985) J. Biomol. Struct. Dynam. **3**, 205-226.
45. Hunter, W.N., Kneale, G., Brown, T., Rabinovich, D. & Kennard, O. (1986) J. Mol. Biol. **190**, 605-618.
46. Rabinovich, D., Haran, T., Eisenstein, M. & Shakked, Z. (1988) J. Mol. Biol. **200**, 151-161.
47. Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D. & Kennard, O. (1987) J. Biol. Chem. **262**, 9962-9970.
48. Brown, T., Hunter, W.N., Kneale, G. & Kennard, O. (1986) Proc. Natl. Acad. Sci. USA **83**, 2402-2406.
49. Hunter, W.N., Brown, T. & Kennard, O. (1986) J. Biomol. Struct. Dynam. **4**, 173-191.
50. Prive, G.G., Heinemann, U., Chandresegaran, S., Kan, L.-S., Kopka, M.L. & Dickerson, (1987) Science **238**, 498-504.
51. Gray, D.M. & Cui, T., Ratliff, R.L. (1984) Nucl. Acids Res. **12**, 7565-7580 + ref. therein.
52. Cornelis, A.G., Haasnoot, J.H.J., den Hartog, J.F., de Rooij, M., van Boom, J.H. & Cornelis, A. (1979) Nature **281**, 235-236.
53. Cornelis, A.G., Haasnoot, J.H.J., den Hartog, J.F., de Rooij, M., van Boom, J.H. & Cornelis, A. (1980) Nucl. Acids Res. **8**, 169-181.
54. Summers, M.F., Byrd, R.A., Gallo, K.A., Samson, C.J., Zon, G. & Egan, W. (1985) Nucl. Acids Res. **13**, 6375-6386.
55. Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. & Dickerson, R.E. (1988) Nature **334**, 175-179.
56. Fresco, J., Broitman, S. & Lane, A.E. (1980) In Alberts, B. & Fox, C.F. (eds), Mechanistic Studies of DNA Replication and Genetic Recombination, Academic Press, New York, pp. 753-768.
57. Singer, B. & Fraenkel-Conrat, H. (1970) Biochemistry **9**, 3694-3701.
58. Banks, G.R., Brown, D.M., Streeter, D.G. & Grossman,L. (1971) J. Mol. Biol. **60**, 425-439.

59. Topal, M.D., Hutchison III, C.A., Baker, M.S. & Harris, C. (1980) In Alberts, B. & Fox, C.F.(eds), Mechanistic Studies of DNA Replication and Genetic Recombination, Academic Press, New York, pp. 725-734.
60. Brown, D.M., Hewlins, M.J.E. & Schell, P. (1968) J. Chem. Soc. (C), 1925-1929.
61. Preston, B.D., Singer, B., Loeb, L.A. (1986) Proc. Natl. Acad. Sci. USA **83**, 8501-8505.
62. Kalnik, M.W., Kouchakdjian, M., Li, B.F.L., Swann, P.F. & Patel, D.J. (1988) Biochemistry **27**, 108-115.
63. Brookes, P. & Lawley, P.D. (1960) J. Chem. Soc. 539-545.
64. Topal, M.D. & Toorchen, D. (1983) Carcinogenesis **4**, 1591-1598.
65. Topal, D.M., Eadie, J.S. & Conrad, M. (1986) J. Biol. Chem. **261**, 9879-9885.
66. Rosemeyer, H. & Seela, F. (1987) J.Org.Chem. **52**, 5136-5143.
67. Lippert, B., Schöllhorn, H. & Thewalt, U. (1986) J. Am. Chem. Soc. **108**, 6616-6621.
68. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. & Steitz, T.A. (1985) Nature **313**, 762-766.
69. Sinha, N.K. & Haimes, M.D. (1981) J. Biol. Chem. **256**, 10671-10683.
70. Fersht, A.R. & Knill-Jones, J.W. (1981) Proc. Natl. Acad. Sci. USA **78**, 4251-4255.
71. Lai, M.-D. & Beattie, K.L. (1988) Biochemistry **27**, 1722-1728.
72. Fersht, A.R., Shi, J.-P. & Tsui, W.-C. (1983) J. Mol. Biol. **165**, 655-667.
73. Drake, J.W. (1970) Molecular Basis of Mutations, Holden-Day, San Francisco.
74. Kunkel, Th.A. & Alexander, P.S. (1986) J. Biol. Chem. **261**, 160-166.
75. Driggers, P.H. & Beattie, K.L. (1988) Biochemistry **27**, 1729-1735.
76. Hall, Z.W. & Lehman, I.R. (1968) J. Mol. Biol. **36**, 321-333.
77. Topal, M.D., DiGuiseppi, S.R. & Sinha, N.K. (1980) J. Biol. Chem. **255**, 11717-11724.
78. Boosalis, M.S., Petruska, J. & Goodman, M.F. (1987) J. Biol. Chem. **262**, 14689-14696.
79. Kuchino, Y., Mori, F., Kasai, H., Inoue, H., Iwai, S., Miura, K., Ohtsuka, E. & Nishimura, S. (1987) Nature **327**, 77-79.
80. Hemmens, W.F. (1963) Biochim. Biophys. Acta **68**, 284-288.
81. Shapiro, R. & Gordon, C.N. (1964) Biochem. Biophys. Res. Commun. **17**, 160-164.
82. Hemmens, W.F. (1964) Biochim. Biophys. Acta **91**, 332-334.
83. Lin, T.-S., Cheng, J.-Ch., Ishiguro, K. & Sartorelli, A.C. (1985) J. Med. Chem. **28**, 1194-1198.
84. Eritja, R., Horowitz, D.M., Walker, P.A., Ziehler-Martin, J.P., Boosalis, M.S., Goodman, M.F., Itakura, K. & Kaplan, B.E. (1986) Nucl. Acids Res. **14**, 8135-8153.
85. Aboul-ela, F., Koh, D. & Tinoco Jr., I. (1985) Nucl. Acids Res. **13**, 4811-4824.
86. Werntges, H., Steger, G., Riesner, D. & Fritz, H.-J. (1986) Nucl. Acids Res. **14**, 3773-3790.
87. Tibanyenda, N., de Bruin, S.H., Haasnoot, C.A.G., van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1984) Eur. J. Biochem. **139**, 19-27.
88. Mhaskar, D.N. & Goodman, M.F. (1984) J. Biol. Chem. **259**, 11713-11717.
89. Sinha, N.K. (1987) Proc. Natl. Acad. Sci. USA **84**, 915-919.
90. Fersht, A.R., Knill-Jones, J.W. & Tsui W.-C. (1982) J. Mol. Biol. **156**, 37-51.
91. Fersht, A.R. & Knill-Jones, J.W. (1983) J. Mol. Biol. **165**, 633-654.
92. Muster-Nassal, C. & Kolodner, R. (1986) Proc. Natl. Acad. Sci. USA **83**, 7618-7622.
93. Hillebrand, G.G., McCluskey, A.H., Abbott, K.A., Revich, G.G. & Beattie, K.L. (1984) Nucl. Acids Res. **12**, 3155-3171.
94. Hillebrand, G.G. & Beattie, K.L. (1985) J. Biol. Chem. **260**, 3116-3125.
95. White, J.H., Lusnak, K. & Fogel, S. (1985) Nature **315**, 350-352.
96. Dohet, C., Wagner, R. & Radman, M. (1985) Proc. Natl. Acad. Sci. USA **82**, 503-505.
97. Kramer, B., Kramer, W. & Fritz H.-J. (1984) Cell **38**, 879-887.
98. Claverys, J.-P., Méjean, V., Gasc, A.-M. & Sicard, A.M. (1983) Proc. Natl. Acad. Sci. USA **80**, 5956-5960.
99. Su, S.-S., Lahue, R.S., Au, K.G. & Modrich, P. (1988) J. Biol. Chem. **263**, 6829-6835.
100. Reyland, M.E., Lehman, I.R. & Loeb, L.A. (1988) J. Biol. Chem. **263**, 6518-6524.
101. Patten, J.E., So, A.G. & Downey, K.M. (1984) Biochemistry **23**, 1613-1618.
102. Ryan, J.P., Duncan, M.C., Bankaitis & Bassford Jr., P.J. (1986) J. Biol. Chem. **261**, 3389-3395.

103. Both, G.W., Shi, C.H. & Kilbourne, E.D. (1983) Proc. Natl. Acad. Sci. USA **80**, 6996-7000.
104. Dalbadie-McFarland, G., Neitzel, J.J. & Richards, H. (1986) Biochemistry **25**, 332-338.
105. Bateman, J.F., Chan, D., Walker, I.D., Rogers, J.G. & Cole W.G. (1987) J. Biol. Chem. **262**, 7021-7027.
106. Brooker, R.J. & Wilson, T.H. (1985) Proc. Natl. Acad. Sci. USA **82**, 3959-3963.
107. Jukes, Th. (1980) Science **210**, 973-978.
108. Gojobori, T., Li, W.-H. & Graur, D. (1982) J. Mol. Evol. **18**, 360-369.
109. Jukes, Th. (1987) J. Mol. Evol. **26**, 87-98 + ref. therein.
110. Radding, C.M. (1978) Ann. Rev. Biochem. **47**, 847-880.
111. Dunkan, B.K. & Miller, J.H. (1980) Nature **287**, 560-561.
112. Bird, A.P. (1980) Nucl. Acids Res. **8**, 1499-1504.
113. Martin, F.H. & Castro, M.M., Aboul-ela, F. & Tinoco Jr., I. (1985) Nucl. Acids Res. **13**, 8927-8938.
114. Tabin, C.J., Bradley, S.M., Bargmann, C.I., Weinberg, R.A., Papageorge, A.G., Scolnick, E.M., Dhar, R., Lowy, D.R. & Chang, E.H. (1982) Nature **300**, 143-149.
115. Reddy, E.P., Reynolds, R.K., Santos, E. & Barbacid, M. (1982) Nature **300**, 149-152.
116. Varmus, H.E. (1984) Ann. Rev. Genet. **18**, 553-612.
117. Bos, J.L., Toksoz, D., Marshall, C.J., Verlaan-de Vries, M., Veenemann, G.H.,van der Eb, A.J., van Boom, J.H., Janssen, J.W.G. & Steenvoorden, A.C.M. (1985) Nature **315**, 726-730.
118. Kozma, S.C., Bogaard, M.E., Buser, K., Saurer, S.M., Bos, J.L., Groner, B. & Hynes, N.E. (1987) Nucl. Acids Res. **15**, 5963-5971.
119. Goodman, M.F. & Branscomb, E.W. (1986) In Kirkwood, T.B.L., Rosenberger,R.F. & Galas, D.J.(eds), Accuracy in Molecular Processes, Chapman & Hall, London, New York, pp. 191-232.
120. Fowler, R.G., Schaaper, R.M. & Glickman, B.W. (1986) J. Bacteriol. **167**, 130-137.
121. Schaaper, R.M. & Dunn, R.L. (1987) Proc. Natl. Acad. Sci. USA **84**, 6220-6224.
122. Schaaper, R.M., Danforth, B.N. & Glickman, B.W. (1983) J. Mol. Biol. **189**, 273-284.
123. Hobza, P. & Sandorfy, C. (1987) J. Am. Chem. Soc. **109**, 1302-1307.
124. Alberts, B.M., Barry, J., Bedinger, P., Burke, R.L., Hibner, U., Liu, C.-C. & Sheridan, R. (1980) In Alberts, B. & Fox, C.F. (eds), Mechanistic Studies of DNA Replication and Genetic Recombination, Academic Press, New York, pp. 449-473.
125. Newport, J.W., Kowalczykowski, S.C., Lonberg, N., Paul, L.S. & Hippel, P.H. (1980) In Alberts, B. & Fox, C.F. (eds), Mechanistic Studies of DNA Replication and Genetic Recombination, Academic Press, New York, pp. 485-505.
126. Topal, M.D., Sinha, N.K. (1983) J. Biol. Chem. **258**, 12274-12279.
127. Codeiro-Stone, M., Topal, M.D. & Kaufman, D.G. (1982) Carcinogenesis **3**, 1119-1128.
128. Rich, A., Nordheim, A. & Wang, A.H.-J. (1984) Ann. Rev. Biochem. **53**, 791-846.
129. Koo, H.-S., Wu, H.-M. & Crothers, D.M. (1986) Nature **320**, 501-506.
130. Rein, R., Shibata, M., Garduno-Juarez, R. & Kieber-Emmons, T. (1983) In Clementi, E. & Sarma, R.H. (eds),Structure and Dynamics: Nucleic Acids and Proteins, Adenine Press, New York, pp. 269-288.
131. Scanlan, M.J. & Hillier, I.H. (1984) J. Am. Chem. Soc. **106**, 3737-3745.
132. Chuprina, V.P. & Poltev, V.I. (1985) Nucl. Acids Res. **13**, 141-154.
133. Brown, D.M. & Hewlins, M.J.E. (1969) Nature **221**, 656-657.