Volume 17 Number 11 1989 Nucleic Acids Research

Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences

Richard G.H.Cotton* and R.Duncan Campbell[1]

Murdoch Institute, Royal Children's Hospital, Melbourne 3052, Australia and [1]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Received March 21, 1989; Revised and Accepted May 10, 1989

ABSTRACT

The chemical reactivity of matched T and C bases to osmium tetroxide and hydroxylamine near mismatched and unmatched bases in a heteroduplex between two strands of DNA with multiple differences was examined. Data was available for matched bases one or two positions away from 24 mismatches. Reactive bases were found near 16 of the mismatches and were usually one or two bases away. This reactivity is consistant with structural studies indicating perturbation of the duplex around mismatches and will allow another mode of study of the effect of mismatches. The reactivity of these bases was found not to be strongly correlated with mismatch type or GC basepair content of the basepairs around the mismatches. Extra reactivity may have been promoted by the presence of either T or C in the mismatch allowing increased reactivity of nearby T or C. The utility of the phenomenon for the detection of mutations is discussed. Unmatched bases in the heteroduplex also gives rise to reactive matched bases nearby.

INTRODUCTION

Chemical reagents have been used to probe a variety of forms of DNA and RNA to detect and investigate unusual or single stranded structures as well as their widespread use for sequencing (1). These structures include Z DNA (2), Hoogsteen basepairing (3), $(AT)_n$ sequences (4), triple helical DNA (5) and Holliday junctions (6). These structures are usually sequence dependent and are generated by torsional strain on the duplex. The reagents have also been used for structural studies in tRNA (7) and 18S rRNA (8) for example.

We have recently shown that mismatches in DNA can be detected by chemical reactivity using osmium tetroxide (detecting mismatched T) and hydroxylamine (detecting mismatched C) (9). Mismatched A & G bases can be detected as mismatched T & C bases respectively by use of probes of opposite sense. In the course of these studies it was noticed that not only are the

mismatched T & C bases reactive but also a lesser but marked reactivity of T & C bases near mismatches, however this phenomenon was not examined in detail.

Structural studies of mismatches and surrounding bases has been intense in recent years (reviewed in 10 and 11). The interest in their structure stems from the mismatch being the target of repair enzymes in living cells (12). The chemical study of mismatches (9) complement these structural studies but as the structural studies also indicated perturbation of surrounding bases it was thought important to examine the chemical reactivity of such bases. Perturbation of bases around drug binding sites has also been reported (13-14). Detection of mutations by cleavage of bases near mismatches not including C or T (in the case of osmium tetroxide or hydroxylamine) where the test nucleic acid is single stranded, such as in mRNA is also an important objective.

This study examines the extent of reactivity of C and T bases near and in some cases more distant from various mismatches and loops in a complex heteroduplex formed from two pieces of DNA with multiple differences.

MATERIALS AND METHODS

The source of DNA has been described (9). All studies were made using the 357 bp MspI/NcoI fragment of the 21 hydroxylase B gene (15) for the production of probes of both senses by end labelling (9). Heteroduplexes were formed with MspI/NcoI digests of the 21 hydroxylase A pseudogene (9). The 363bp region studied is shown in Fig. 1. The 21 hydroxylase A gene differs in 35 positions (including 6 base insertions) from the 21 hydroxylase B gene in this region. Each mutation is named according to the position of the base change from the first difference called A1. This region contains mutations A17-A51.

Partial cleavage of the heteroduplexes was essentially as described (9) and particular conditions are given in the figure legends. Cleavage was assayed by electrophoresis on 8, 10 and 20% denaturing urea gels (9).

Analysis of cleavage was by densitometry of the bands on the autoradiograph produced from the dry gel.

Maxam and Gilbert sequencing of the labelled probes was performed (1,9) and samples were run next to the analysis to position the T and C residues being studied.

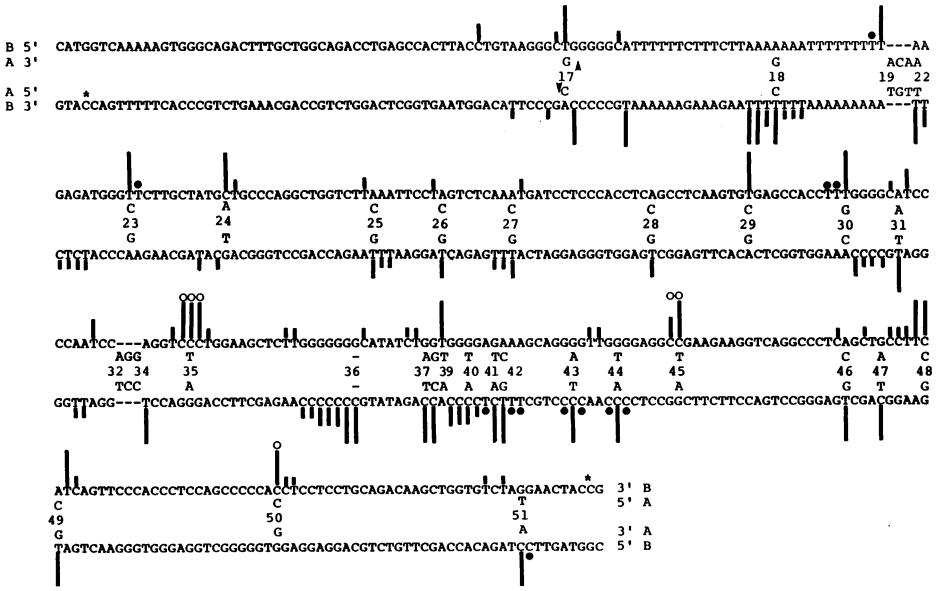
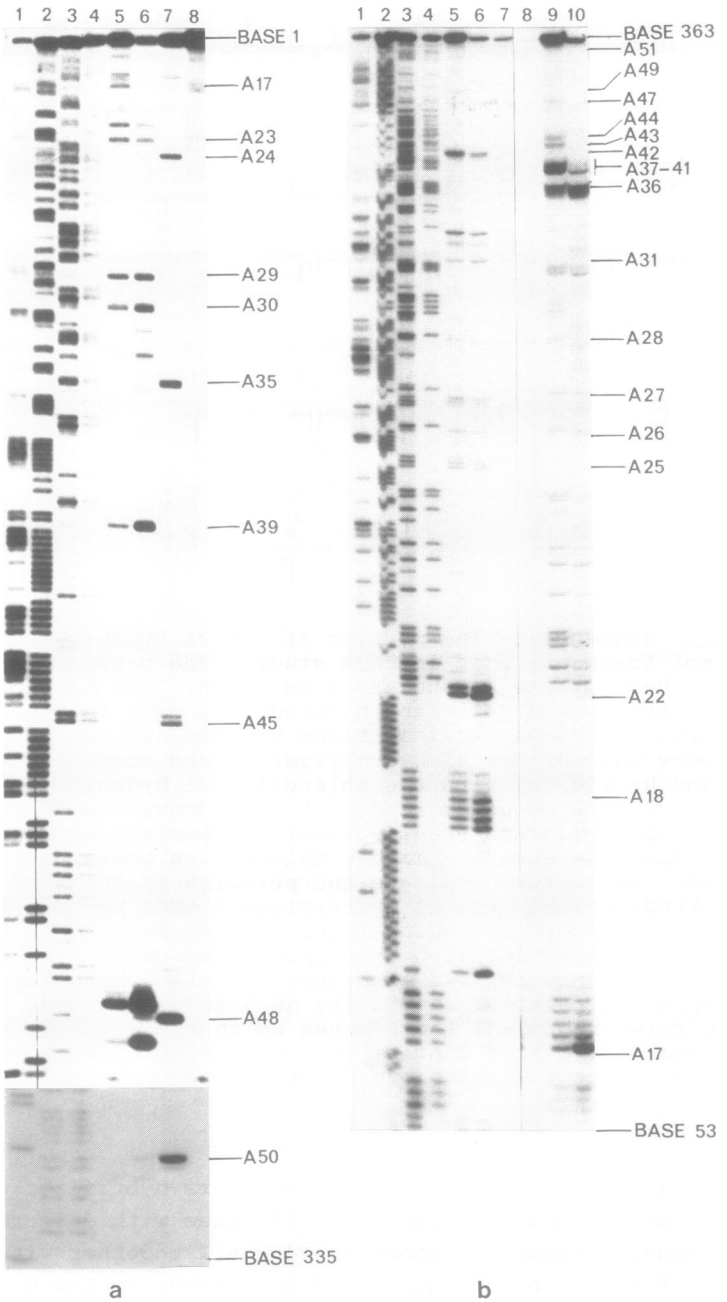


Figure 1. Sequence of the portion of the 21 hydroxylase B gene (NcoI/MspI fragment) used in this study. The bases which differ in the 21 hydroxylase A gene are also shown. The sequences are paired as they would be in the heteroduplex such that the actual mismatches, deletions and insertions created are indicated. Probes were made of the 21 hydroxylase B gene strands for both senses and hybridized with the unlabelled 21 hydroxylase A gene. Numbers above the sequence represent 21 hydroxylase A gene mutation number (Cotton et al., 1988). Asterisks show the base used to label the strand. Each complete line contains 100 base positions. The arrows indicate the position of the MspI cut sites introduced into the 21 hydroxylase A gene by mutation A17. Vertical lines above and below T and C bases represents the reactivity of that base with the appropriate chemical when the particular sequence is used as probe. The lengths of these lines represent relative reactivity as judged from Fig. 2. Closed circles represent those bases which are probably reactive but the analysis cannot confirm it. Open circles represent C bases clearly reactive with osmium tetroxide.

RESULTS

End labelled probes of both senses were hybridized with the A gene to create heteroduplexes of 363 bases with differences at 35 positions. These are shown in Figure I together with the position of the label. Samples of each heteroduplex were treated with osmium tetroxide or hydroxylamine and then piperidine and the fragments separated on denaturing gel


1 2 3 4 5 6 7 8
BASE 1
A17
A23
A24
A29
A30
A35
A39
A45
A48
A50
BASE 335
a
1 2 3 4 5 6 7 8 9 10
BASE 363
A51
A49
A47
A44
A43
A42
A37–41
A36
A31
A28
A27
A26
A25
A22
A18
A17
BASE 53
b

electrophoresis (Figure 2). Autoradiograms were scanned to quantitate the radioactivity in each band (not shown). Only those showing convincing reactivity were scored positive after inspection of gels and traces. To be scored as positive reactivity (Table I, Fig. I) the band must be more reactive than a nearby base of the same type. (This reactivity is considerably less than that of a mismatched base). This data was translated to bars as shown on Fig. I. This potentially allows examination of bases near all the mismatches or other perturbations. However due to lack of resolution in the higher molecular weight area A17-23 top (sense probe heteroduplex) (Fig. 2a) and A41-51 (antisense probe heteroduplex) (Fig. 2b) analysis cannot be complete in these regions. Fig. I gives an overview of the reactivity of T and C bases in the two complementary heteroduplexes.

There are 30 isolated (mismatch at least 4 bases from another mismatch or unmatched base) mismatches which can be analysed for reactive bases near the mismatch (Table I).

The results show that in the 30 isolated mismatched bases in the probes tested at least 16 show reactivity of C or T bases up to 3 bases away but usually 1 or 2 bases away. In the 13 cases of unreactive C or T bases next or next but one to a mismatched base in the probe, 4 were T bases.

Figure 2. Analysis of fragments produced by partial cleavage of heteroduplexes of end labelled 21 hydroxylase B and unlabelled 21 hydroxylase A gene using osmium tetroxide and hydroxylamine.

(a) Labelled sense strand of the 21 hydroxylase B gene hybridized with the 21 hydroxylase A gene. Lanes 1-4 Maxam and Gilbert sequencing tracks G, C + A, T + C and C respectively. Lane 5 and 6 osmium tetroxide reaction for 1 and 5 min. respectively. Lane 7 hydroxylamine reaction for 10 min., lane 8 homoduplex control incubated for 5 min. with osmium tetroxide. Gels were 10% (top) or 20% (bottom) denaturing gels.

(b) Labelled antisense strand of the 21 hydroxylase B gene hybridised with the 21 hydroxylase A gene. Lanes 1-4 Maxam and Gilbert sequencing tracks G, G + A, T + C and C respectively. Lanes 5 and 6 osmium tetroxide reaction with heteroduplex for 1 and 5 min. respectively. Lane 7 osmium tetroxide reaction with homoduplex for 5 min. Lane 8 hydroxylamine reaction with homoduplex for 60 min. Lane 9 and 10 hydroxylamine reaction with heteroduplex for 10 and 60 min. respectively. Gel was an 8% denaturing gel.

TABLE 1

Reactivity of matched C and T bases one and two positions from mismatches at least 4 bases distant from another mismatch or unmatched base. Reaction in the third position is in brackets. In some cases where T bases are reactive on each side result is given as T1,T1. In describing the mismatch the base in the probe is given first.

Mismatch		Sense Probe			Antisense Probe		GC pairs
No.	Mismatch	Reactive	Nonreactive	Mismatch	Reactive	Nonreactive	around base†
A18	A/G			T/C	T1+,T2+,T1+, T2+(T3+,T3+)		0
A23	T/C	?	C2	A/G		C1,C2	4
A24	C/A	T1		G/T	C1,(T3+)	C2	4
A25	A/C	T1		T/G	T1+,T2+		1
A26	A/C	T1	C2,T2	T/G		C1	4
A27	A/C	T1		T/G	T1+,T2+	C2	2
A28	A/C		C1,C2,T2	T/G		C1	6
A29	T/C		T2	A/G		C1,C2,T2	5
A30	T/G	?		A/C	C1+,C2+(C3*)		6
A31	A/A	C1,T1	C2	T/T	C2(C3*)		7
A35	C/T	C1+,C1+, T2+,T2+		G/A			5
A43	G/A	T2		C/T	?		5
A44	G/T	T2+		C/A	?		5
A45	C/T	C1+		G/A		C1,C2	6
A50	C/C	C1+,T2	C2	G/G			3

*Common to mutation A30, A31.
+Reactive base the same as one of the two in the mismatch
†The number of GC pairs out of 8 around mismatch base (19)
? Reactivity of nearby matched bases not able to be ascertained.

Where mismatches are close (less than 4 bases between each or an insertion and hence not included in the analysis i.e. A21-22, A37-42 and A46-49) there is as expected reactivity of nearby matched T and C bases. However the distance of reactive matched bases from the mismatch is not noteably increased compared with those near the isolated mismatches.

Reactivity also occurs around insertions (A19-21 and A32-34). The T base next to A19 is very reactive and the reactivity of T and C bases up to 6 bases from A21 may be added to by the mismatch at A22. The T base next to A34 is also very

reactive as are T bases 3, 4 and 5 bases from A32. It is notable that two C bases next to A32 are not reactive.

The bulge C or G (A36) causes extensive reactivity particularly on one side where T bases 7,8 bases from the bulge G and C bases 1-6 bases from the bulge C are reactive. On the other side a C base next to the G bulge is reactive.

The only other feature which causes reactivity of surrounding matched T and C bases is the strand break at A17 due to the creation of an Msp-1 site by this mutation. It can be seen that reactivity of bases up to 12 positions from the break are observed.

It should be noted that while not the subject of this work, all additional isolated mismatched T and C bases studied in this work and not reported before (9) are reactive. These are 7 T mismatches (1 T/C; 5 T/G; 1 T/T) and 4 C mismatches (2 C/T;, 2 C/A). No mismatched C or T Bases are unreactive.

DISCUSSION

This study has established that matched bases near mismatches are often reactive (Table I). As the conditions used in this study were only partial cleavage conditions it is highly likely that if individual mismatches were the only ones in a duplex of otherwise matched bases e.g. 500 bp long,and if longer reaction was applied, further reactivity of nearby matched bases would be revealed. These findings are consistent with solution NMR studies (16) which have indicated that mismatched bases perturb the surrounding duplex. However this phenomenon has not been detected in x-ray studies of crystals (reviewed in 10 and 11). They are also consistent with studies on destabilization of duplexes by mismatches as measured by dissociation (17) and by helix formation rates (18).

Bases around mismatches may be reactive in the present system for two main reasons (a) the destabilizing effect of the mismatches as described by others in solution studies (17,18) or (b) bases in the mismatch may react and perturb the mismatch so nearby matched T bases become reactive. Thus a T in a mismatched pair, when reacted with osmium tetroxide, may have even less affinity for the other base in the pair thus further destabilizing the duplex making nearby matched T bases more reactive than they would be near an unmodified T mismatch. The extent of their reactivity in either case may be dependent on

position and number of GC basepairs in relation to AT pairs around the area of the mismatch or the bases involved in the mismatch (see below).

Another phenomenon described by Fazakerley et al (19) which may explain the reactivity of matched bases near mismatches is the formation of loops of unmatched bases caused by one of the two bases of the mismatched pair forming a normally matched pair with a nearby complementary base. For example at mutation A18 with the antisense probe the mismatched T instead of mispairing with the C might correctly pair with one or other of the two A bases each side of the C on the other strand. This would lead to a reactive T bulge in the probe.

Relatively marked reactivity occurs (to at least the 3rd position) when runs of 3 or more C or T bases occur in the same strand, one of the run usually being present in the mismatch. Examples are seen at A18 antisense strand, A25 and A27 antisense strand, A30 antisense strand, A35 sense strand and A36 antisense strand (bulge). It is thought that this may be because reaction of the chemical at the mismatched base may further perturb the area making it easier for nearby matched bases to react i.e. a "chain reaction" ensues. If this is true and such mismatches are eliminated from Table 1 (marked +), it can be seen that many matched bases near mismatches not containing the reporter base (T or C) still show as reactive. It has been shown that osmium tetroxide reacts with C bases at about one fortieth the rate of reaction with T base (discussed in 9). If this reagent is perturbing mismatches which include C (as well as those containing T) it can be seen that eliminating from consideration mismatches containing C when osmium tetroxide is used as the probe, still leaves 4 mismatches leading to reactivity of matched bases next to mismatches [i.e. A31, A43 (sense probe) and A24,A31 (antisense probe)]. Others (18) have reported however a run of T's at a mismatch can be very destabilizing and this is consistant with our results.

The reactivities listed in Table 1 were analysed to see if they were predictable on the basis of conclusions from other studies. Duplexes containing G mismatches have been found to be the most stable (17) or easily formed (18) however while there is a trend, reactivity is not consistant with this. In fact 5 bases in the 1 or 2 position are not reactive around A/C mismatches which are found to be one of the least stable (17)

and has been suggested to perturb the duplex over several basepairs (20). Other workers have found that repair was dependent on the proportion of GC pairs present in the 8 pairs surrounding a mismatch (20), but while there was a trend, numbers were also too small to show a relationship to chemical reactivity (Table I). Werntges et al (20) from theoretical studies suggested permanant loop formation by particular mismatches after theoretical treatment of their data. Pyrimidine: pyrimidine mismatches particularly showed this possibility and our data shows that reactivity of nearby matched bases is high in this type of mismatch. Numbers were too small to allow comparison of reactivity around mismatches with GC each side compared with AT each side where the reported reactive base was not in these mismatched basepairs but it is notable that at A28 C2 and T2 are negative and at A29 T2 is negative (sense strand probe) and at A29 T2 is negative and at A45 C2 is negative (antisense strand probe). Thus it is possible and perhaps expected (21) that GC pairs would oppose the disrupting forces of a mismatch. Thus it is possible that all these features may play a role in the reactivity of surrounding matched bases i.e. mismatch type and number of GC basepairs each side of the mismatch and in the area of the mismatch. However model experiments are necessary to test this hypothesis.

It is clear from Fig. 1 that deletions e.g. A36 would be readily detectable directly or indirectly. Insertions (A19-21 and A32-34) would be detectable indirectly by reactivity of bases in the probe paired with bases near the insert.

The reactivity of matched bases near a mismatch is likely to have important practical consequences, regardless of the mechanism and mode of enhancement of this reaction. If one is using a particular probe the position of a mismatch and hence a mutation may be signalled by reaction of a nearby matched C or T where there is a mismatched A or G in the probe. (Of course use of probe of opposite sense would pinpoint the mutation by detecting the unmatched T or C respectively). In the case of the examples given here (Table 1) under conditions of partial cleavage, 8 of 13 of the mutations would be detected indirectly where A and G appear in the probe. When more complete cleavage is applied all may be detected. However to be complete, one would (when using DNA) use probes of both senses. However the single sense probe becomes more important when examining single

stranded RNA for mutations (mRNA or viral RNA). A recent example with collagen mRNA has detected the mutation by cleavage of a matched base near a mismatch (22). Also in a situation where there are multiple differences between probe and unlabelled nucleic acid, one may not need to detect all differences as a pattern of difference may suffice as has recently been described in single stranded viral RNA (Cotton and Wright, submitted). In this case it was possible to obtain a fingerprint of the difference between two virus strains. Such a fingerprint may be useful in epidemiological studies.

ACKNOWLEDGEMENTS

Dr. N. Rodrigues is thanked for DNA samples and Mrs. W. Russell for technical assistance.

This work was supported by the National Health and Medical Research Council of Australia and the Medical Research Council of Great Britain.

[I]Abbreviations : T, thymine; C, cytosine; A, adenine; G, guanine

*To whom correspondence should be addressed

REFERENCES

1. Maxam, A.M. & Gilbert, W. (1980) Methods Enzymol. **65**, 499-560.
2. Johnson, B.H. & Rich, A. (1985) Cell **42**, 713-724.
3. Mendel, D. and Dervan, P.B. (1987) Proc. Natl. Acad. Sci. USA **84**, 910-914.
4. McClellan, Palecek, E. and Lilley, D.M.J. (1986) Nucl. Acids Res. **14**, 9291-9309.
5. Johnson, B.H. (1988) Science **241**, 1800-1804.
6. Duckett, D.R., Murchie A.I.H., Diekmann, S., von Kitzing, E., Kemper, B. and Lilley, D.M.J. (1988) Cell **55**, 79-89.
7. Cramer, F. (1971) Prog. Nucleic Acid Res. Mol. Biol. **11**, 391-421.
8. Rairkar, A., Rubino, H.M. and Lockard, R.E. (1988) Biochemistry **27**, 582-592.
9. Cotton, R.G.H., Rodrigues, N.R. & Campbell, R.D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 4397-4401.
10. Kennard, O. (1988) In Sarma, R.H. and Sarma, M.H. (eds.) Structure and Expression: DNA and Its Drug Complexes, Adenine Press, Vol. 2. 1-25.
11. Patel, D.J., Shapiro, L. and Hare, D. (1987) Ann. Rev. Biophys, Biophys. Chem. **16**, 423-454.
12. Kramer, B., Kramer, W. and Fritz, H-J. (1984) Cell **38**, 879-887.
13. Jeppesen, C. and Nielsen, P.E. (1988) FEBS Letters **231**, 172-176.
14. Portugal, J., Fox, K.R., McLean, M.J., Richenberg, J.L. and Waring, M.J. (1988) Nucl. Acids Res. **16**, 3655-3670.

15. Rodrigues, N.R., Dunham, I., Yu, C.Y., Carroll, M.C., Porter, R.R. & Campbell, R.D. (1987) EMBO J. **6**, 1653-1661.
16. Gao, X. and Patel, D.J. (1987) J. Biol. Chem. **262**, 16973-16984.
17. Ikuta, S., Takagi, K., Bruce Wallace, R. and Itakura, K. (1987) Nucl. Acids Res. **15**, 797-811.
18. Aboul-ela, F, Koh, D. and Tinoco, Jr., I. (1985) Nucl. Acids Res. **13**, 4811-4824.
19. Fazakerley, G.V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G.A., van Boom, J.H., Jones, M. and Radman, M. (1986) EMBO J. **5**, 3697-3703.
20. Werntges, H., Steger, G., Riesner, D. and Fritz, H-J. (1986) Nucl. Acids Res. **14**, 3773-3790.
21. Tinoco, Jnr. I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature **246**, 40-41.
22. Bateman, J.F., Lamande, S.R., Dahl, H.-H., Chan, D., Mascara, T. and Cole, W.G. J. Biol. Chem. (in press).