DNA repair and antigenic variation in Trypanosoma brucei

Ulbert, S.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 8

Conclusions and perspectives

Sebastian Ulbert
Chapter 8

Conclusions and perspectives

Biosynthesis of J

Base J replaces 1% of the thymine residues in the nuclear DNA of Trypanosoma brucei, mainly in repetitive telomeric sequences. Previous results led to a model in which J is synthesized in two steps via the intermediate 5-HmU (van Leeuwen et al., 1998). In Chapter 2, this two-step model was tested by excision of 5-HmU at the DNA level, mediated by expression of the human DNA glycosylase hSMUG1 in bloodstream form T. brucei. hSMUG1 was found to cause DNA damage due to massive and specific removal of 5-HmU. Trypanosomes expressing the enzyme showed a decrease in J level, indicating that 5-HmU in DNA is a precursor in J-biosynthesis (base J itself is not excised by hSMUG1). The fact that 5-HmU in DNA is freely accessible to a DNA glycosylase suggests that the two steps in J biosynthesis are separated events and that 5-HmU is a normal component of trypanosome DNA. Expression of hSMUG1 in insect form trypanosomes had no effect, except that it rendered the cells sensitive to incorporation of exogenous 5-HmU (Chapter 3). This shows that, like J, 5-HmU is normally absent in procyclic trypanosomes.

The DNA damage caused by hSMUG1 was sequence-specific, only sequences that were J-modified were fragmented. We conclude that 5-HmU colocalizes with J, is present in the telomeric repeats but is absent in chromosome-internal genes. Together with the previous observation that random incorporation of exogenous 5-HmU leads to J distributed over the whole genome (van Leeuwen et al., 1998), the results strongly support the idea that the formation of 5-HmU normally occurs in a sequence-dependent manner and restricts the bulk part of J to repetitive telomeric DNA. It should be noted that 5-HmU can also arise through oxidative DNA damage. However, this is likely to happen independently from sequence context and hence we believe that the 5-HmU generated by oxidative attack on thymine does not contribute significantly to the level of J in the genome. Taken together, the results obtained with hSMUG1 in trypanosomes confirm the two step model for J-biosynthesis, with 5-HmU as an intermediate.

Chapter 2 presents evidence that the formation of 5-HmU can be inhibited by incorporation of the nucleoside analog BrdU, as trypanosomes fed with BrdU became less sensitive to expression of hSMUG1, although the DNA glycosylase remained functional (S.U. and P.B., unpublished). Inhibition of 5-HmU synthesis also explains that the decrease in J level upon BrdU-incorporation is higher than expected by loss of thymine (as substrate for J) from DNA (van Leeuwen et al., 104
Conclusions and perspectives

1998). This inhibition could be due to binding of the enzyme which cannot convert the substrate to the product 5-HmU but remains bound to the base analog, similar to the inhibition of the mammalian DNA methyltransferase by 5-azacytosine (Juttermann et al., 1994). Thus, BrdU might be used as a tool to find the putative thymine hydroxylase by incubating a BrdU-containing oligonucleotide in trypanosome extracts and analyzing the proteins binding to it.

J and BER

BER is a system to remove damaged bases from DNA and its major components are DNA glycosylases that excise a variety of modified bases. Although J is a bulky base modification, it is not recognized to a significant extent by any of the DNA glycosylases from various organisms (Chapter 5). This indicates that evolutionarily conserved DNA glycosylases are highly specific for the base modifications they encounter in a cell. A base that is never seen in DNA and does not significantly harm DNA integrity does not fit into the recognition pattern of the BER machinery. Similarly, BrdU does not seem to be a target for BER (G.W. Teebor, personal communication). This suggests that J alone is not a reason for T. brucei to lack evolutionarily conserved BER factors. Indeed, we found that trypanosomes have a BER system similar to other organisms as they contain genes for the major conserved DNA glycosylases and have BER activities detectable by in vitro assays.

The situation is less clear for 5-HmU, the precursor of J. We found weak activities of the DNA glycosylases AlkA, Mug and its human homolog hTDG, against 5-HmU paired to A. Mug and hTDG showed a much higher activity against 5-HmU when paired to G, but such a mispair is unlikely to occur in trypanosome DNA. Only hSMUG1 excised 5-HmU efficiently, independent of the base pair. The expression of hSMUG1 was lethal to trypanosomes resulting from excision of 5-HmU (Chapter 2), suggesting that a BER activity against this base would be disadvantageous for T. brucei. Taking into account the phylogenetic distribution of the DNA glycosylases removing 5-HmU, the lack of homologs for AlkA and hSMUG1 in the (incomplete) trypanosome genome databases does not necessarily represent an adaptation to 5-HmU. Nevertheless, a sequence similar to hTDG and its bacterial homolog Mug was found in Leishmania. The lack of a detectable activity against ethenocytosine in trypanosomes, another TDG substrate, might suggest that the kinetoplastid enzyme has evolved differently from its homologues in other organisms in order to tolerate 5-HmU. However, no in vivo excision of 5-HmU has yet been found in nature, except in higher eukaryotes containing SMUG1 (Boorstein et al., 1987 and 2001), hence the role of TDG in this context remains unclear, and further work is required to address this question.
Chapter 8

Chapter 5 shows that the presence of J is unlikely to result in the absence of certain DNA glycosylases. Nevertheless, it still might be the reason for having extra DNA glycosylases. In higher eukaryotes that methylate cytosine, deamination of 5-MeC creates a T:G mispair, which can be repaired by the excision of mispaired T by BER, an activity absent in other organisms. Similarly, trypanosomes might contain enzymes that specifically repair lesions arising through a damaged J. It would be interesting to synthesize an oxidatively damaged J and to test whether it is excised by trypanosome extracts, and whether DNA glycosylases from other sources recognize it.

The negative results on J and BER raise the question whether J might have led to adaptations in other DNA repair systems or in nuclear processes such as transcription or DNA replication.

BER is the major pathway responsible for removing damaged bases from DNA, but other pathways exist, for example bulky base lesions that interfere with DNA structure are removed by NER (Friedberg, 2001). There is no evidence, however, that J distorts the DNA helix. Earlier work has suggested that 5-HmU is not detected by the NER system, as cells with an active NER system, but lacking a DNA glycosylase able to remove 5-HmU, do not touch this base (Mi et al., 1997). The addition of a glucose does not distort the DNA structure either, as models of fully glucosylated d(J-A)$_n$ and d(J)$_n$d(A)$_n$ duplexes have shown that the glucose can be accommodated in the major groove of B-DNA without steric hindrance (Gao et al., 1997). Hence, we consider it unlikely that NER would recognize and remove base J. DNA repair enzymes acting by direct damage reversal are highly specific and those analyzed to date act on methyl groups arising through alkylation of bases (Falnes et al., 2002; Rydberg et al., 1990), making it unlikely that they would recognize J. Trypanosomes are also endowed with MMR and double strand break repair systems very homologous to other organisms (Bell and McCulloch, 2002; McCulloch and Barry, 1999; Robinson et al., 2002; Conway et al., 2002), and our finding that *T. brucei* has a functional BER fits into the overall picture that mechanisms repairing DNA damage are as conserved as the DNA itself.

It is not yet known whether J interferes with transcription or DNA replication in other organisms. It is not likely to do so in trypanosomes, as it would be very disadvantageous if the replication machinery was stalled at every J residue, and experimental evidence argues against a direct role of J in transcriptional silencing (van Leeuwen et al., 1998). Furthermore, several modifications at the 5-position of pyrimidines in DNA, such as 5-MeC and 5-HmU (but not 5-formyluracil, which is target for BER), do not interfere with replication or basepairing properties (Heinemann and Hahn, 1992; Zhang et al., 1999; Bjelland et al., 2001), in line with structural data on J-containing DNA (Gao et al., 1997). However, it would be interesting to test whether prokaryotic or other
eukaryotic RNA and DNA polymerases recognize J as a normal thymine residue. If not, this would indicate specific adaptations of these proteins in trypanosomes.

To conclude, J might add epigenetic information to the primary DNA sequence, which is used by the trypanosome in a way that is not yet understood. It is interesting to ask how this base could evolve in the presence of DNA repair. The available data on the (lack of) interaction between J and BER suggest that evolution has generated a base modification that is not recognized by DNA repair. This was favored over adapting a conserved pathway such as BER to a modified base, which might even have been impossible: a novel modified base representing a target to an intrinsic DNA glycosylase would have been excised before it could have given any advantage to an organism, maybe even causing DNA damage due to excessive BER. This evolutionary model is analogous to methylated cytosine and adenine in other organisms, which are also not removed by DNA repair. In essence, DNA repair can be seen as a limiting factor for the evolution of enzymatically modified bases, and those known to date seem to have evolved without disturbing DNA repair.

Alternatively, J might have been present before the invention of some of the conserved DNA glycosylases, hence the overall BER system evolved with a genome that contained J and therefore does not recognize it as damage. We find this unlikely and favor the model that the evolution of J represents a specific phenomenon of a common ancestor of kinetoplastids, euglenoids and Diplonema (see also Chapter 1). However, the phylogenetic distribution of J might yet be wider than expected, and the situation might be similar to cytosine methylation in Drosophila, that was only discovered after finding the genes for DNA methyl transferases in its genome (Lyko, 2001). Once the genes for enzymes making J are identified, it will be of interest to search for similar sequences in other organisms to get more insight into the origin and evolution of this modified base.

Activation of VSG gene expression sites

To survive the exposure to the host immune system, T. brucei has developed a sophisticated system of antigenic variation. Chapter 6 investigates the key feature in this system, the control of VSG expression sites. These highly homologous polycistronic transcription units harbor VSG genes and display allelic exclusion, i.e. only one of them is fully transcribed at a time. The trypanosome is able to change the VSG expressed by replacing the VSG gene in the active expression site or by switching off one expression site and activating another one (in situ switch). In a previous study, a putative intermediate of the in situ switch was identified that was rapidly
switching between two expression sites marked with drug resistance genes (Chaves et al., 1999). This has led to the model of a pre-active expression site, which differs from a silent site in its ability to be readily activated. Chapter 6 further investigates the pre-active state, after we generated trypanosomes with three marked expression sites, and the results show that expression sites are not regulated independently, but that there is a form of cross-talk between them, coupling activation of one expression site to inactivation of the previously active expression site (Chaves et al., 1999; Borst and Ulbert, 2001).

The newly tagged expression site entered the rapid switching phenotype at the same frequency as the expression sites previously analyzed, suggesting that the pre-active state is a general feature of an intermediate in expression site switching. The results showed that maximally two expression sites entered the pre-active state, whereas the third one did not participate in the rapid switching and remained inactive. This finding fits with a model of limiting factors that are exclusively accessible for the active expression site. During the process of switching, a second expression site gets access to these factors and becomes pre-active. This happens at low frequency, and is most likely a stochastic event. The consequence is a short-lived competition of the two sites, which does not involve the other expression sites and rapidly results in only one being active (although this state can be trapped using drug selection). Although the limiting factors for expression site activation remain to be found, the model is compatible with a nuclear structure such as the expression site-associated body (ESB), identified by Navarro and Gull (2001, see also Chapter 1). Whether the two expression sites involved in the switching intermediate both localize to the ESB requires further study.

In the current model of allelic exclusion only one expression site is expressed at a given time. However, the inactive expression sites are not necessarily completely silent, as discussed in Chapter 6 (with similar observations also made by others; Navarro and Cross, 1998; Ansorge et al., 1999; Vanhamme et al., 2000). The degree of transcriptional silencing can be altered, leading to partially activated expression sites. This state affects sequences close to the promoter, is stable and erased by full activation of the site and subsequent shut-down after a VSG switch. Transcripts from promoter proximal sequences of several silent expression sites have been detected in the nucleus of wild type T. brucei (Vanhamme et al., 2000). However, they were not fully processed and exported into the cytoplasm. Hence, the partial activation described in Chapter 6 and by Navarro and Cross (1998) is a distinct phenomenon, as it leads to functional mRNA and appears at low frequencies. What distinguishes a partially activated expression site from a fully active one, besides the level of transcription, is not known as yet. If the elevated
transcription at a partially activated expression site in Chapter 6 is mediated by RNA Polymerase I (although this has not been tested), it could be that a silent expression site gets limited access to the ESB or is able to recruit RNA Polymerase I outside the ESB. Alternatively, the partially activated expression site could be located in the nucleolus, where the majority of RNA polymerase I is found in the cell. As possible ESB-specific transcription and RNA processing factors might be lacking outside the ESB, the level of expression would be low and limited to promoter proximal sequences.

Although silent expression sites are randomly distributed in the nucleus (Chaves, 2000; Navarro and Gull, 2001), the nuclear localization of a partially activated expression site has not yet been studied, and these models therefore require further analysis. The marker genes inserted in the expression sites of the HNPb cell line used in Chapter 5 are quite short in sequence (0.4 – 1 kb), and this makes it difficult to perform in situ hybridizations to detect these individual expression sites in the nucleus of *T. brucei* (S.U. and P.B., unpublished; Chaves, 2000). Nevertheless, as it is possible to insert much larger marker sequences close to expression sites and to perform antibody-staining of DNA sequences (Navarro and Gull, 2001), it would be feasible to localize expression sites in all stages identified so far (silent, active, pre-active and partially activated).

### References


Chapter 8


Conclusions and perspectives


