The bio-macromolecular nucleic acids DNA and RNA are each built up from similar sets of four distinct building blocks called nucleotides. These nucleotides consist of a nucleobase that forms the coding entity, the (2'-deoxy-)ribose sugar and a phosphate group that allows for the connection of the individual nucleotides. Crucially, both DNA and RNA are post-synthetically modified by endogenous processes introducing a second layer of information into the nucleic acid, which is commonly referred to as epigenetic base modification. In mammalian DNA this occurs through methylation of 2'-deoxycytidine, forming 5-methyl-2'-deoxycytidine which has been linked to the crucially important gene expression for correct cellular function and when disregulated becomes a hallmark of disease. Unfortunately the DNA is also constantly modified by other events that damage the molecule carrying the all important genetic information; these events can be from both endogenous (e.g. reactive oxygen species) and exogenous (e.g. tobacco smoke) sources. This may lead to 'erasure' of genetic information by the formation of a so-called abasic site where the coding nucleobase has been removed, leaving an aldehyde residue at this site (Figure 1). Additionally, spontaneous 'erasure' of nucleobases does also occur making the abasic sites one of the most common occurrences of DNA damage. Surprisingly, it has so far remained elusive whether these events occur on specific sites on the genomic DNA. In other words, are certain parts of the genome more sensitive to this sort of damage?

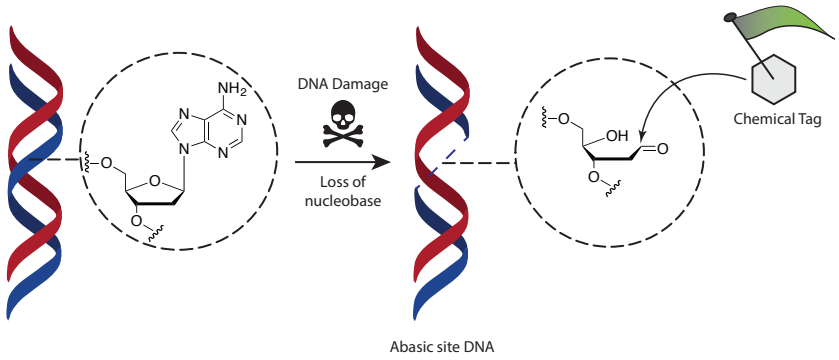

Figure 1. DNA nucleotides can form abasic sites through damage or spontaneous loss of the nucleobase. The resulting abasic site contains the reactive aldehyde group (red) which can be chemically tagged.

I have screened different chemical reactions to "tag" abasic site DNA damage in experiments to identify those genomic regions where such sites are present and to attempt to quantify them. Crucially, the selected chemistry should have the following properties: bio-orthogonal, selective, yield a stable intermediate, rapid and high yielding. To this end I screened hydrazines, oxy-amines and di-amines, some of which are commercially available and have been used in the past to quantify abasic sites on DNA. I have also explored the application of the so-called Pictet-Spengler (PS) reaction to tag abasic sites in DNA. By designing and synthesizing these PS probes and reacting them with short synthetic DNA it was found that this novel approach outcompetes traditional aldehyde reactive probes in terms of conversion. More importantly, the PS probe reacts at physiological pH and forms a very stable carbon-carbon bond between the tag and DNA.

Having identified an efficient new chemical procedure to tag abasic sites in DNA, we set out to develop a method to enrich abasic site DNA by means of a streptavidin pulldown. To this end a biotin affinity tag was added to our PS probe. This method will allow for the determination of those sites on genomic DNA where abasic sites are present at a resolution of approximately 200 – 500

bases. By selection of optimized conditions, we have now established that >100-fold enrichment of synthetic abasic sites containing DNA can be achieved using the PS-tag *in vitro.* This result, one of the key objectives, is very promising and efforts to translate the conditions established *in vitro* to genomic DNA are the subject of ongoing research. In this respect an important hurdle that has yet to be surmounted is the optimization of PCR amplification of the enriched 'tagged' abasic site-containing DNA, which is crucial order for genomic scale sequencing to finally be applied. Currently we observe polymerase stalling at our PS-tagged abasic sites on synthetic oligomers when amplification for sequencing is pursued directly after enrichment.

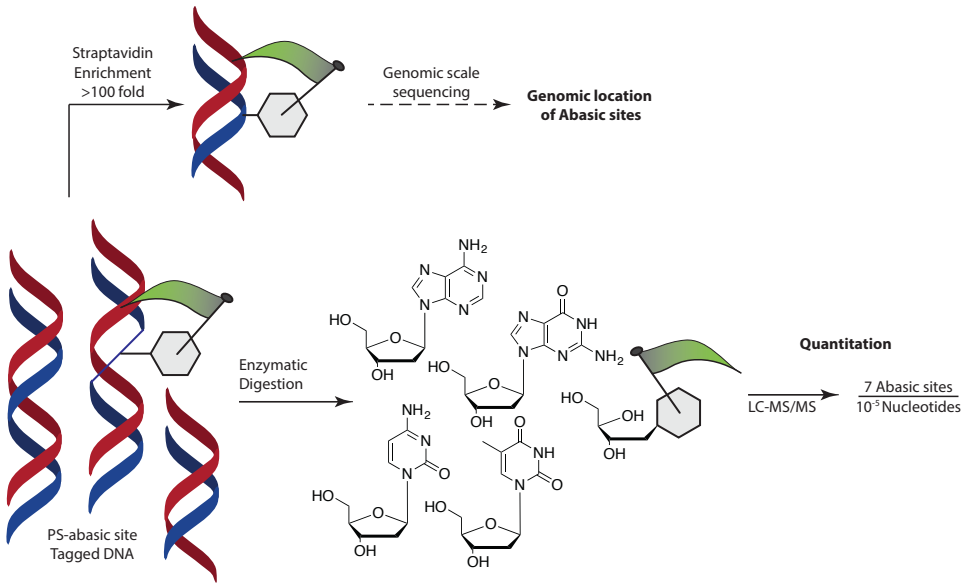

Figure 2. The Pictet-Spengler reaction was used on synthetic DNA to assess this chemistry to enrich abasic sites. Enrichment of >100-fold was achieved. The Pictet-Spengler reaction was then used to tag the DNA of HEK-293T cells *in vitro* to quantify abasic sites on genomic DNA.

Although sequencing by means of enrichment is a common technique and has been applied to several other modifications such as *N*6-methyl-2'deoxyadenosine, it does not provide information on the absolute quantities of a given DNA modification/level of damage. To this end, a method was developed to enzymatically digest 'tagged' abasic site DNA and provide subsequent quantification by state of the art liquid chromatography - tandem mass spectrometry. Preliminary data obtained from reacting the DNA of HEK-293T cells with the PS-probe, subsequent digestion and quantification, abasic site levels of are observed at an abundance of 7 sites per 10^{-5} genomic nucleotides.

In summary we have established novel chemistry to tag abasic sites. By using Pictet-Spengler chemistry we did not only improve reaction rates, but this chemistry can also be applied *in situ* due to its retained reactivity at physiological pH. Furthermore, the chemical adduct formed is hydrolytically stable, important for downstream analysis such as the performed LC-MS/MS quantitation. Finally, experiments revealed this chemistry can indeed be used to enrich abasic site-containing DNA *in vitro*. Towards genomic scale sequencing, final optimization of the PCR amplification of DNA enriched for abasic sites is now required.