The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates

ABSTRACT

UDGs form a single protein superfamily with a distinct structural fold and a common evolutionary origin. Differences in the catalytic mechanism of the different families combined with the construction of the catalytic pocket have, however, resulted in extreme sequence divergence of these enzymes.

INTRODUCTION

Background Mutagenic uracil appears in DNA opposite to guanine as a result of misincorporation or of deamination of cytosine. Similarly, the deamination process generates thymine opposite guanine in those organisms that undergo cytosine methylation. DNA is safeguarded from the consequences of these events by the activity of uracil DNA glycosylases (UDGs), which remove uracil (and sometimes thymine) from the sugar backbone of DNA without breaking the phosphodiester bonds in the backbone. There are different types of these enzymes in the three superkingdoms of life. The best studied family of UDGs, typified by the Escherichia coli Ung protein, is largely specific for uracil and is present in a variety of bacteria, eukaryotes and large eukaryotic DNA viruses. The mismatch-specific uracil DNA glycosylases (MUGs) have been identified in eukaryotes and several bacteria and, unlike the Ung-family enzymes, are additionally active on G:T mismatches. Comparison of the crystal structures of these two enzymes has shown that they are structurally very similar, despite the low sequence similarity. Subsequently, two other classes of UDGs have been characterized, one from thermophilic archaea and several bacteria (hereinafter called AUDG) and the other from vertebrates (SMUG). The latter enzyme has a high specificity for uracil and for single-stranded substrates. The single-strand-specific UDGs (ssUDGs) are believed to be functionally similar to the UNGs and MUGs because they possess motifs similar to the catalytic motifs of the latter enzymes despite supposedly lacking significant sequence similarity to them. In contrast, the structural and evolutionary affinities of the AUDGs are uncertain. Thus, considerable structural diversity appears to exist among the UDGs, their generally similar catalytic activities notwithstanding. Here, using sequence profile searches, multiple alignment analysis and structural comparisons, we unify all known UDGs into a single protein superfamily and predict a common α/β fold for them. We additionally detect several new probable UDGs that are distinct from the already characterized families, and explore the evolutionary scenarios that could have resulted in the observed phyletic distribution of these enzymes.

CONCLUSION

Conclusions Using sequence profile searches, multiple alignment analysis and protein structure comparisons, we have shown that all known UDGs form a single protein superfamily with a distinct structural fold and a common evolutionary origin. The extreme sequence divergence of different families of UDGs is probably due to differences in their biochemistry, with only the general shape of the protein molecule and the binding pocket being essential for the DNA glycosylase reaction per se. Although the UDG superfamily is nearly ubiquitous among cellular life forms, the individual families show limited and distinct phyletic distributions. The emerging evolutionary scenario for the UDGs involves multiple events of lateral gene

transfer and lineage-specific gene loss. In addition, we predict two previously undetected families of UDGs; the experimental investigation of their functions is expected to broaden the current perspective on these critical repair enzymes.