Structure, function and evolution of the XPD family of iron–sulfur-containing 5′→3′ DNA helicases

Malcolm F. White

Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, U.K.

Abstract

The XPD (xeroderma pigmentosum complementation group D) helicase family comprises a number of superfamily 2 DNA helicases with members found in all three domains of life. The founding member, the XPD helicase, is conserved in archaea and eukaryotes, whereas the closest homologue in bacteria is the DinG (damage-inducible G) helicase. Three XPD paralogues, Fanq (Fanconi’s anaemia complementation group J), RTLE (regular of telomere length) and Chl1, have evolved in eukaryotes and function in a variety of DNA recombination and repair pathways. All family members are believed to be 5′→3′ DNA helicases with a structure that includes an essential iron–sulfur-cluster-binding domain. Recent structural, mutational and biophysical studies have provided a molecular framework for the mechanism of the XPD helicase and help to explain the phenotypes of a considerable number of mutations in the XPD gene that cause three different genetic conditions: xeroderma pigmentosum, trichothiodystrophy and Cockayne’s syndrome. Crystal structures of XPD from three archaeal organisms reveal a four-domain structure with two canonical motor domains and two unique domains, termed the Arch and iron–sulfur-cluster-binding domains. The latter two domains probably collaborate to separate duplex DNA during helicase action. The role of the iron–sulfur cluster and the evolution of the XPD helicase family are discussed.

Eukaryotic XPD (xeroderma pigmentosum complementation group D)

The XPD/Rad3 helicase family comprises a group of related superfamily 2 DNA helicases with a 5′→3′ directionality. Family members are found in all domains of life, including bacteria, archaea and eukaryotes. In eukaryotes, XPD (Rad3 in Saccharomyces cerevisiae) functions as part of the TFIIF complex. The ten-subunit TFIIF complex has a dual role in transcription initiation and NER (nucleotide excision repair). In the former, TFIIF opens the DNA around the promoter site, whereas, in the latter, it opens the DNA around a site of DNA damage. XPD is an essential structural component of TFIIF and acts as a bridging subunit between the core TFIIF subunits and the CAK (CDK (cyclin-dependent kinase)-activating kinase) complex [1]. The helicase activity of XPD is essential for NER, but is not required for transcription initiation. A second helicase in TFIIF, XPB (xeroderma pigmentosum complementation group B), has the opposite polarity to TFIIF. Although XPB is a very weak helicase, its ATPase activity is essential for both NER and transcription initiation [1,2].

Key words: archaeon, Fanconi’s anaemia complementation group J (Fanq), helicase, iron–sulfur-cluster-binding domain, molecular evolution, xeroderma pigmentosum complementation group D (XPD).

Abbreviations used: BRCA1, breast cancer 1 early-onset; CS, Cockayne’s syndrome; DinG, damage-inducible G; Fanq, Fanconi’s anaemia complementation group J; NER, nucleotide excision repair; RTLE, regular of telomere length; ssDNA, single-stranded DNA; TFIIF, transcription factor II H; TTD, trichothiodystrophy; XP, xeroderma pigmentosum; XPD, xeroderma pigmentosum complementation group B; XPG, xeroderma pigmentosum complementation group D.

Mutations in the XPD gene give rise to three different genetic conditions in humans: XP (xeroderma pigmentosum), TTD (trichothiodystrophy) and CS (Cockayne’s syndrome). XP is characterized by extreme light sensitivity, inability to remove UV photoproducts from DNA and highly elevated rates of skin cancer (reviewed in [3]). These characteristics are consistent with a loss of the helicase activity of XPD and therefore lack of NER activity. In contrast, TTD mutations cause developmental and growth abnormalities, as well as reduced DNA-repair activity, but typically do not result in elevated cancer rates. These symptoms are thought to arise due to defects in both NER and transcription arising from the relevant XPD mutations [3]. Defective transcription probably prevents cells becoming cancerous, explaining the apparent discrepancy between XP and TTD. CS is a rare disease where both transcription and repair are defective and is characterized by segmental progeria (premature aging) [3].

Archaeal XPD: discovery of the iron–sulfur cluster

Most archaea contain a clear homologue of eukaryotic XPD. The function of this protein in archaea is still unclear, although a role in an NER-type pathway has been predicted [4]. The protein is not part of a TFIIF-type complex and apparently functions as a monomer. When the XPD homologue from Sulfolobus acidocaldarius was cloned and overexpressed in Escherichia coli, it was immediately apparent that the protein was an iron–sulfur protein owing...
to its yellow–green colour [5]. The likely iron–sulfur-cluster-binding site was identified as a set of four conserved cysteine residues close to the Walker B motif, and this was confirmed by site-directed mutagenesis. It was shown that the iron–sulfur domain was not essential for stability or the enzyme’s ability to bind to ssDNA (single-stranded DNA) or its ATPase activity, but was essential for the helicase activity of the protein. It also became clear that the four cysteine residues were conserved in eukaryotic XPD, suggesting that the iron–sulfur domain was present. This was confirmed by genetic experiments in S. cerevisiae, which showed that the abolition of iron–sulfur cluster binding by mutagenesis of a conserved cysteine resulted in a severe UV-sensitive phenotype in yeast, consistent with a loss of activity of Rad3 in vivo [5]. Furthermore, the four cysteine residues were also conserved in the other eukaryotic XPD paralogues: FancJ (Fanconi’s anaemia complementation group J), RTEL (regulator of telomere length) and Chl1, suggesting a general role for the cluster in this broad family of helicases.

**DinG (damage-inducible G): the bacterial XPD**

Most bacteria also contain a homologue of XPD. This was originally identified as a gene induced in the SOS response in E. coli: DinG [6]. However, the function of DinG in bacteria is not well defined, as knockouts have little discernible phenotype. In vitro, the DinG enzyme has activity very like that of archaeal XPD, and the protein is not thought to exist in a complex [7]. When the iron–sulfur cluster was identified in archaeal XPD, it was noted that the four conserved cysteine ligands were also present in most DinG sequences at a similar location. This evidence was taken to suggest that DinG was also an iron–sulfur-cluster-binding protein [5]. Subsequent biochemical data confirmed this prediction [7].

Although most bacterial DinG sequences have four conserved cysteine residues, a minority have no cysteine residues in this region (e.g. Staphylococcus aureus DinG) and therefore presumably cannot bind an iron–sulfur cluster [8]. Furthermore, some DinG sequences have an N-terminal exonuclease 1 domain, suggesting a fusion of a helicase activity with a 3′→5′ ssDNA exonuclease. This could result in a protein that unwinds DNA and degrades the displaced strand, similar to the Werner protein in eukaryotes [9]. However, there are currently no biochemical data to support this idea.

**Eukaryotic paralogues of the XPD helicase**

In eukaryotic genomes up to three paralogues of XPD have been identified. All share the four conserved cysteine residues and are therefore likely to be iron–sulfur proteins. The best characterized is the helicase FancJ [also known as BACH1/BRIP1 (BRCA1 breast cancer 1 early-onset)-interacting protein C-terminal helicase 1)] which plays a role in a DNA cross-link repair pathway that is mutated in Fanconi’s anaemia patients (reviewed in [10]). FancJ interacts with the breast cancer-susceptibility protein BRCA1 and may also have a role in double-strand break repair [11]. Recently, a role for FancJ in the unwinding of G4 quadruplex structures during DNA replication has been proposed [12,13]. One mutation of human FancJ that results in Fanconi’s anaemia is targeted to Ala349, which is mutated to a proline residue. This residue is not conserved in the XPD family, but is positioned next to the fourth conserved cysteine ligand in human FancJ. Mutation of the equivalent residue in archaeal XPD (F136P) resulted in the destabilization of the iron–sulfur cluster and consequently a loss of helicase activity [5]. This suggests that the A349P mutation of FancJ in humans probably causes Fanconi’s anaemia by disrupting the iron–sulfur cluster-binding domain and thus inactivating the FancJ helicase [5]. A second paralogue is RTEL. Rtel-knockout mice die during gestation with multiple developmental defects, genomic instability and telomere loss [14]. RTEL has also been suggested to be a functional equivalent of yeast Srs2 which antagonizes homologous recombination by dissociating D-loops in metazoan [15]. Thirdly, the Chl1 protein is found in yeast and some metazoan. It also contains the four conserved cysteine residues and is therefore probably an iron–sulfur-dependent helicase. In yeast, Chl1 is important for chromosome segregation [16]. Human Chlr1 interacts with the cohesin complex, and deletion of the mouse gene is lethal and has defects in chromosome segregation [17,18].

**Crystal structures of XPD**

Crystal structures of archaeal XPDs have been reported recently by three groups [8,19,20]. All three structures reveal a four-domain organization, with the iron–sulfur-cluster-binding domain and an Arch domain arising from the first of two canonical helicase motor domains. Together, the Arch and iron–sulfur domains form a channel through which ssDNA is dragged by the action of the motor domains in a cyclical ATP-dependent reaction (Figure 1). This is consistent with a role for the iron–sulfur domain in breaking the DNA duplex, in other words acting as the ‘ploughshare’ seen in many DNA helicases [21], and data from fluorescently labelled DNA species support the idea that the DNA duplex is broken near the iron–sulfur-binding domain [22]. However, it is not yet clear which specific part of the protein is utilized to separate the DNA strands, and it remains a formal possibility that the DNA helix is broken at the other side of the protein on motor domain 2, as suggested by Wolski et al. [19]. Further work will be required to resolve these issues.

The crystal structures and accompanying biochemical data shed light on the in vivo consequences of the mutations seen in human XP, TTD and CS. XP mutations are clustered in two main areas: around the ATP-binding site, and along the path of the ssDNA along the top of the two motor domains. XP residues tend to be highly conserved from archaea through to humans. These observations suggest a role in DNA binding and catalysis. Mutations causing XP are thought to inactivate XPD helicase without perturbing the overall structure of the enzyme. This has the consequence of knocking out NER activity while preserving the role of TFIIH in transcription initiation where the helicase activity is not required [23].
The role of the iron–sulfur cluster

The discovery of an iron–sulfur cluster domain in XPD was unexpected, and the role of the cluster has been the subject of considerable speculation. Iron–sulfur clusters are comparatively rare in DNA-repair proteins; the only other example being the cluster found in certain members of the DNA glycosylase family [24]. Here, the cluster has been postulated to have a structural role, but also potentially a role in the detection of DNA damage. The Barton group has suggested that glycosylases with an iron–sulfur cluster might pass an electron on to DNA as a mechanism to test for the presence of DNA damage [25]. Since XPD has also been predicted to play a role in the detection of DNA damage during NER, this has prompted speculation that the iron–sulfur cluster might play an active role in the detection of DNA damage, such as photoproducts. However, although this is superficially an attractive hypothesis, the observation that four human helicases with different functions in the cell each have an iron–sulfur cluster leads to the obvious conclusion that the cluster cannot have such a specific function.

A second possibility is that the cluster confers a redox-sensing function on the protein. In this scenario, the activity of an enzyme such as XPD could be controlled by the oxidation state of the iron–sulfur cluster, specifically by inactivating the protein under conditions of oxidative stress by decomposing the cluster. However, it seems counterintuitive that a DNA-repair enzyme required for the removal of DNA lesions caused by oxidative stress should be deactivated when it is most needed.

Instead, it is more likely that the iron–sulfur cluster has a purely structural role in stabilizing the small domain that, with the Arch domain, separates the DNA duplex during helicase function. Iron–sulfur clusters have recently been discovered in several other proteins, including DNA primase [26], RNA polymerase [27], and the bacterial AddAB and eukaryotic Dna2 enzymes [28]. It is possible that iron–sulfur clusters are a type of ‘molecular fossil’ that predates the advent of aerobic lifestyles. Iron–sulfur clusters may have been replaced by redox-insensitive zinc domains in most situations, but have persisted in some proteins owing to the quirks of evolution. Bear in mind that the machinery for iron–sulfur cluster biosynthesis, although complex, is essential anyway for the assembly of redox proteins. There is therefore presumably no great cost in maintaining iron–sulfur clusters as a structural feature in proteins where the cluster is relatively redox insensitive. For both AddAB and DinG, some bacterial lineages where iron is particularly toxic or in short supply appear to have overcome the requirement for an iron–sulfur-cluster-binding domain [28]. Again, this suggests a structural rather than functional/sensing role for the iron–sulfur cluster in these proteins.

Evolution of the XPD family

The ubiquitous nature of the XPD family across all three domains of life suggests that the ancestral helicase including the iron–sulfur-binding domain is an ancient protein that was present in the last common ancestor. The closer similarity of eukaryotic and archaean XPD compared with the much more distant bacterial DinG protein mirrors the situation for

---

**Figure 1 | Structure of XPD from Thermoplasma acidophilum**

Left-hand panel: cartoon of the XPD structure from *T. acidophilum* [19] with the four major domains labelled and the iron–sulfur-cluster-binding domain highlighted in blue. Right-hand panel: the protein is shown as a space-filling representation, emphasizing the narrow pore formed between the Arch and iron–sulfur cluster domains that is wide enough to accommodate ssDNA. A schematic representation of the likely path of DNA emphasizes the possibility that the DNA duplex is separated at the constriction formed by these two domains, and that the single strand is dragged across the top of the motor domains from right to left in this depiction.
many proteins that are involved in information processing. In particular, proteins involved in DNA replication and recombination, transcription and translation are all more highly similar in eukaryotes/archaea and more divergent or even unrelated in bacteria [29,30]. The observation that archaeal XPD is most similar to the eukaryotic XPD (rather than any of the paralogues) and the ubiquitous nature of XPD across all of the eukaryotic phyla suggest that XPD is the founding member of the family. The same argument suggests that archaeal XPD may have a role to play in an archaeal NER-like pathway, though this remains to be proved.

The XPD paralogues FancJ, RTEL and Chl1 probably arose during the evolution and diversification of the eukaryotic domain as new DNA-repair pathways evolved. In bacteria, the role of DinG is also still unclear, but it is also likely to function in a repair pathway of some description. The lack of phenotype of a DinG knockout may be explained by the existence of a second repair pathway with overlapping specificity. As shown in Figure 2, DinG in Gram-positive bacteria acquired a further function with the fusion of an exonuclease III-type domain at the N-terminus. This nuclease–helicase fusion is reminiscent of the Werner's syndrome nuclease III-type domain at the N-terminus. This nuclease–helicase fusion is reminiscent of the Werner's syndrome nuclease III-type domain at the N-terminus. This nuclease–helicase fusion is reminiscent of the Werner's syndrome nuclease III-type domain at the N-terminus.

## Acknowledgements

Thanks to Jana Rudolf, Anne-Marie McRobbie, Christophe Rouillon and all members of the Scottish Structural Proteomics Facility for their contributions to this research.

## Funding

This work was funded by Cancer Research UK.

## References


Received 9 January 2009
doi:10.1042/BST0370547