Review

Cell PRESS

From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins

Gábor M. Harami, Máté Gyimesi, and Mihály Kovács

ELTE-MTA ''Momentum'' Motor Enzymology Research Group, Department of Biochemistry, Eötvös University, H-1117 Budapest, Pázmány P. stny. 1/C, Hungary

The winged helix domain (WHD) is a widespread nucleic-acid-binding protein structural element found in all kingdoms of life. Although the overall structure of the WHD is conserved, its functional properties and interaction profiles are extremely versatile. WHD-containing proteins can exploit nearly the full spectrum of nucleic acid structural features for recognition and even covalent modification or noncovalent rearrangement of target molecules. WHD functions range from sequence-recognizing keys in transcription factors and bulldozer-like strand-separating wedges in helicases to mediators of protein–protein interactions (PPIs). Further investigations are needed to understand the contribution of WHD structural dynamics to nucleic-acid-modifying enzymatic functions.

Architecture of the WHD

The helix–turn–helix (HTH) motif, comprising two helices connected by a sharp turn, is a widespread double-stranded (ds) DNA-binding element present in about one-third of DNA-binding proteins [1]. In canonical HTH motifs the turn defines an angle of about 120° between the two helices. The helix following the turn has been found to mediate sequence-specific dsDNA binding in transcription factors, and is thus termed the recognition helix. Based on the patterns of additional structural elements attached to the HTH motif, HTH-containing protein domains are classified into numerous subtypes [1].

In 1993 a new HTH-containing domain fold was discovered in the DNA-binding domain of the liver-specific transcription factor hepatocyte nuclear factor (HNF)3γ, also known as forkhead box (FOX)A3 [2]. This domain consists of three α-helices [H1–3; H2 and H3, together with a connecting turn (T), form an HTH motif]; three β-strands (S1–3); and two characteristic loops (the so-called 'wings', W1 and 2) in the order H1-S1-H2-T-H3-S2-W1-S3-W2. The appearance of W1 and W2 at the sides of H3 inspired the term 'winged HTH' or winged helix domain (Figure 1A). We note that the term wing is often used to refer to the β-strand–loop–β-strand structure following the HTH motif; for the purpose of this review, W1 and W2 refer to the loops alone.

Corresponding author: Kovács, M. (kovacsm@elte.hu).
Keywords: helix–turn–helix; RNA; Z-DNA; RecQ; protein–DNA interaction; helicase.

0968-0004/$ – see front matter
 http://dx.doi.org/10.1016/j.tibs.2013.04.006

The discovery of the WHD of HNF3γ was rapidly followed by the identification of various other WHD-containing proteins based on sequence and/or structural similarity [1,3]. Some WHD types lack the W2 loop and/or contain additional structural elements [4–6]. Although the overall structure of the WHD has been conserved throughout evolution, the following discussion highlights that its interaction profiles and functional properties in different proteins are extremely versatile (Table 1).

The key to recognition: diverse nucleic acid interaction profiles and effector functions

dsDNA sequence and/or modification recognition

The recognition of dsDNA sequence was the first identified biological role of WHDs. The sequence-specific dsDNA interaction of the HNF3γ WHD is brought about by interactions between H3 and the major groove of DNA (Figure 1B) [2]. This interaction profile has been conserved in other WHD-containing transcription factors including the FOX family [7], E2F transcription factor 4 - E2F dimerization partner 2 (E2F4-DP2) transcription factor [8], lambda excision A (LexA) repressor [9], catabolite gene activator protein (CAP) [5], and the E26 transformation-specific (ETS) family [10]. Thus, H3 can be identified as the recognition helix of the HTH motif contained within the WHD. Additional less-conserved interactions between W1 and the minor groove of DNA also contribute to dsDNA recognition by a subset of the above-mentioned proteins.

A recently discovered variant of an HNF3γ-like WHD–DNA interaction was described for *Streptococcus pneumoniae* R.DpnI restriction endonuclease (Figure 1C) [11]. R.DpnI recognizes GATC dsDNA sequences in which the adenine in at least one of the strands is methylated at its N^6 atom (m^6A), and cuts after the m^6A residue to generate blunt-ended DNA products. R.DpnI is proposed to protect *dam*$^-$ bacterial hosts from methylated phage DNA. The recognition helix of the WHD of R.DpnI fits into the major groove of fully methylated dsDNA in which the two methyl groups are in close proximity. The majority of WHD–DNA interactions are formed between five polar side chains located in the recognition helix and the DNA bases of the recognition sequence, including the m^6A methyl groups. Interestingly, the W1 loop of R.DpnI is short and does not interact with DNA. The DNA interaction of

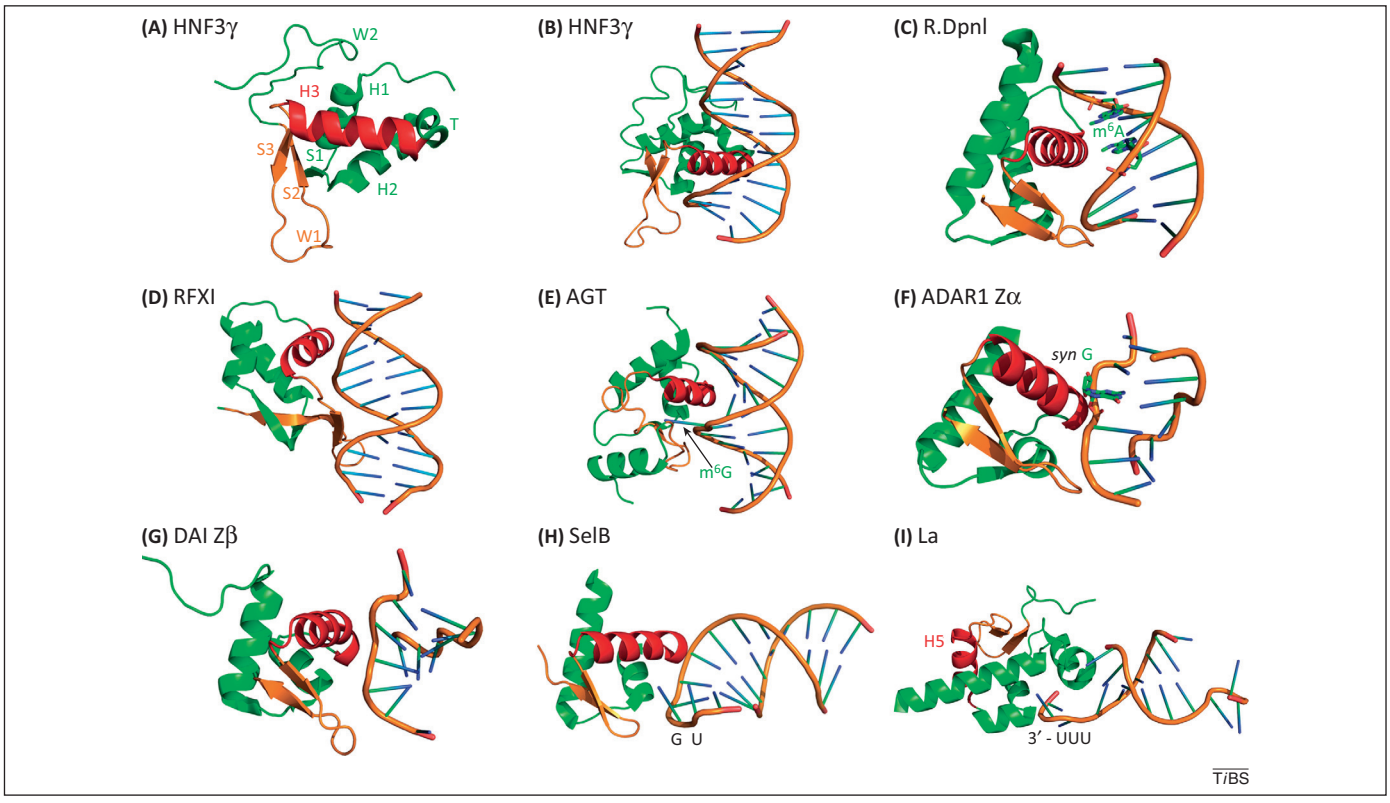

Figure 1. Winged helix domain (WHD) architecture and nucleic acid interaction profiles. **(A)** Cartoon representation of the WHD of human hepatocyte nuclear factor (HNF)3γ (PDB ID: 1VTN [2]) showing the arrangement of secondary structural elements. The recognition helix is colored red, W1 with adjacent β-strands is colored orange, and the rest of the WHD is colored green in all panels. Nucleic-acid-bound WHD structures of **(B)** human HNF3γ bound to double-stranded (ds) DNA (PDB ID: 1VTN [2]); **(C)** *Streptococcus pneumoniae* R.DpnI restriction endonuclease bound to methylated dsDNA (PDB ID: 4ESJ [11]), with methylated adenine bases shown as sticks); **(D)** human transcription regulatory factor (RFXI) bound to dsDNA (PDB ID: 1DP7 [4]); **(E)** human O6-alkylguanine-DNA methyltransferase (AGT) bound to methylated dsDNA (PDB ID: 1T38 [12]), with the methylated guanine flipped out of the DNA helix); **(F)** human dsRNA adenosine deaminase (ADAR1; Zα) bound to Z-DNA (PDB ID: 1QBJ [14]), with the recognized *syn* guanine shown as sticks); **(G)** human DNA dependent activator of IFN-regulatory factors (DAI; Zβ) bound to Z-DNA (PDB ID: 3EYI [23]); **(H)** *Moorella thermoacetica* selenocysteine tRNA-specific elongation factor (SelB) (WHD4) bound to a selenocysteine incorporation site (SECIS) mRNA hairpin (PDB ID: 1WSU [24]); **(I)** human La autoantigen (La domain) bound to RNA with a 3′-UUU overhang (PDB ID: 1YTY [6]).

R.DpnI exemplifies how both dsDNA sequence and base modification can be recognized by the WHD.

In contrast to the above 'canonical' (HNF3γ-like) WHD–DNA interaction profile, sequence-specific dsDNA recognition can also be mediated by interactions between W1 and the major groove of DNA, with H3 interacting with the minor groove. Such a binding profile was described for the transcriptional activator regulatory factor (RF)X1 (Figure 1D) [4].

Another type of WHD–dsDNA interaction was observed for human O_6-alkylguanine-DNA methyltransferase (AGT) (Figure 1E) [12]. AGT recognizes and removes alkyl groups from damaged DNA guanine bases in an irreversible and stoichiometric reaction. The amino acids that are crucial for the reaction are located in the recognition helix, which interacts weakly with the DNA minor groove. The majority of WHD–DNA interactions are formed between other WHD helices and the sugar-phosphate backbone of DNA. Besides this structural pattern, unique aspects of the DNA–AGT interaction include the covalent modification of the bound DNA by the WHD, and the recognition of the alkylation-induced distortion of the DNA double helix rather than a specific DNA sequence.

The above examples highlight how diverse WHD–dsDNA interaction patterns are utilized for dsDNA sequence or base modification recognition, with the recognition helix and the W1 wing of the WHD emerging as key elements mediating base-specific interactions.

Z-DNA recognition and stabilization

Z-DNA is a left-handed dsDNA conformer whose formation is facilitated by certain physicochemical conditions and sequence patterns. *In vivo*, Z-DNA has been shown to form transiently at specific transcription sites where it is proposed to provide relief of torsional strain and assist in transcription regulation [13]. The recognition and stabilization of the Z-DNA (or Z-RNA) conformation is an extreme functional adaptation of the WHD that has been shown for the WHDs (termed Z domains) of human dsRNA adenosine deaminase (ADAR1) [14,15], human DNA-dependent activator of interferon-regulatory factors (DAI) [16], vaccinia virus E3L protein [17], and zebrafish (*Danio rerio*) PKZ protein kinase [18].

Detailed structural and mechanistic information is available on the Z domains of ADAR1 and DAI. ADAR1 edits the informational content of mRNA by converting adenine bases to inosine via deamination [19]. The N-terminal deaminase domain of the protein is followed by two tandem Z domains (Zα and Zβ). The Zα domain recognizes and stabilizes left-handed Z-DNA structures that form during RNA transcription. The recognition helix and the W1 loop of Zα interact with the sugar-phosphate backbone of Z-DNA (Figure 1F). Within W1, two conserved

Table 1. Structural and functional properties of WHD–nucleic acid interactions

Recognized feature of nucleic acid	Activity of WHD-containing protein [a]	Biological process	Nucleic acid interaction type	Example	Refs
dsDNA sequence	Regulatory	Transcription	HNF3γ-like	HNF3γ (FOXA3)	[2,3]
				CAP (CRP)	[5]
				E2F4-DP2	[3,8]
				LexA	[9]
				Elk-4 (Sap1)	[10]
			Unique	RFX1	[3,4]
ssRNA sequence and terminus	Regulatory	RNA biogenesis	Unique	La (SS-B)	[6,26,27]
Modified base in dsDNA	Methyl transferase (resides within WHD)	DNA repair	Unique	AGT (recognizes m^6G)	[12]
	Endonuclease (linked)	Elimination of foreign DNA	HNF3γ-like	R.DpnI (recognizes Gm^6ATC)	[11]
Z-DNA conformation	Regulatory	Immune response	Z-DNA	DAI (DLM-1, ZBP1)	[16,21–23]
	Kinase (linked)	Immune response		PKZ	[18]
	Regulatory	Viral infection		E3L	[17]
	Deaminase (linked)	RNA editing		ADAR1	[3,14,19]
RNA hairpin	Regulatory, GTPase (linked)	Translation	Unique	SelB	[24,25]
DNA strand junctions and branches	Helicase (linked)	DNA recombination	HNF3γ-like (predicted)	RuvB	[49–51]
		Genome maintenance	Unique	RecQ helicases	[28,29]
None (PPI only)	Helicase (linked)	Genome maintenance	None	Hel308	[41,59]
				Hjm	[54]
	Endonuclease (linked)	Elimination of foreign DNA		FokI	[3,52]
	ssDNA binding	DNA metabolism		RPA	[53]
Unknown	Helicase (linked)	RNA splicing	Unknown	Brr2	[55]
				Prp22	[60]
				Prp43	[61]
		RNA degradation		Ski2	[56,57]
				Mtr4	[58]
		Translation		DHX29	[62]

[a]Linked activity: active site resides outside the WHD.

proline residues were proposed to be critical for this interaction pattern [14]. The only base-specific interaction of Zα is formed between a conserved tyrosine residue of the recognition helix and a Z-DNA guanine adopting the *syn* conformation. In contrast to Zα, the isolated Zβ domain has been found to be unable to bind Z-DNA due to the lack of the conserved tyrosine and other critical residues of the recognition helix [20].

DAI [also known as tumor stroma and activated macrophage protein DLM-1 or Z-DNA binding protein (ZBP)1] is a cytosolic DNA-sensing protein that activates the innate immune system by recognizing foreign cytosolic DNA and binding to regulatory factors [21]. Similarly to ADAR1, DAI possesses two Z domains. However, in the case of DAI, both Zα and Zβ are capable of binding Z-DNA [16,22,23]. The Z-DNA interaction profiles of the Zα domains of ADAR1 and DAI show high similarity, with minor differences in W1–DNA contacts. However, the Zβ domain of DAI interacts with Z-DNA in a markedly different fashion (Figure 1G). The recognition helix of Zβ adopts a kinked conformation with 3_{10}-helical and α-helical parts, unlike the straight α-helical recognition helix of Zα (Figure 1F,G). Moreover, the W1 loop of the DAI Zβ domain is distant from Z-DNA in the crystal structure (Figure 1G), although mutational analysis of this loop suggests that it contributes to DNA binding in an unknown manner [23]. This loop of Zβ lacks the proline residues of Zα that were found to interact with Z-DNA. However, the DAI Zβ domain forms a specific interaction with the bound Z-DNA via an arginine residue of the β-strand preceding the recognition helix. The significant differences between the DNA-binding interfaces of different Z domains sharing high sequence and structural similarity exemplifies that these interactions are subject to flexible adaptation. Comparison of the DNA interaction profiles of the WHDs of HNF3γ-like transcription factors (Figure 1B) with those of Z domains (Figure 1F,G) reveals that the recognition helix can interact either with the bases or the backbone of DNA, providing grounds for sequence-specific and conformation-specific recognition, respectively.

RNA sequence and conformation recognition

The WHD mediates RNA binding in the translation elongation factor SelB [24,25] and also in La proteins [6]. Selenocysteine tRNA-specific elongation factor SelB is essential for the incorporation of selenocysteine into nascent proteins. Ribosome-bound SelB recognizes a unique hairpin structure formed by selenocysteine incorporation sites (SECISs) on mRNA strands. SelB also binds selenocysteyl-tRNA. SelB proteins from *Moorella thermoacetica* (*Mt*) and *Escherichia coli* (*Ec*) contain four tandem WHDs (WHD1–4) in their C-terminal region. The crystal structure of the WHD1–4 region of *Mt* SelB complexed with SECIS-mRNA revealed that WHDs 2–4 make contacts to the RNA hairpin. Of these domains, WHD4 is responsible for recognition of the SECIS-mRNA-specific hairpin tip structure and exhibits a unique nucleic acid interaction profile (Figure 1H). This domain interacts with five consecutive phosphate groups of the mRNA backbone via H2 and the N-terminal part of H3 (the recognition helix). WHD4 forms only two base-specific interactions: it recognizes a

guanine and a uracil located at the tip of the RNA hairpin via a binding pocket formed by H2, H3, S2, W1, and S3 residues. Interestingly, the recognition helix of WHD3 interacts with the dsRNA major grove, similar to HNF3γ-like WHD–DNA interactions [25]. However, it only makes contacts with the RNA backbone.

La proteins are essential components of the eukaryotic translational system, and human La [also known as Sjögren's syndrome antigen B (SS-B)] has been identified as a rheumatic disease autoantigen [26]. La proteins specifically bind to the UUU-3′ end of RNA strands synthesized by RNA polymerase III, thereby protecting them from exonucleolytic degradation, facilitating tRNA maturation, and regulating translation. The extended WHD of La proteins (also known as the La motif) has the structure H1-H1′-H2-S1-H3-H4-H5-S3-W1-S2-W2, in which H5 corresponds to the recognition helix (Figure 1I) [6]. Structural and biochemical studies have revealed that the unique type of single-stranded (ss) RNA binding by this domain is mediated by the α-helical extensions (H1′, H2, and H4) of the WHD, with a lack of H5 involvement [6,27].

In summary, the recognition of target RNA molecules is mediated by various WHD structural elements, as observed for dsDNA-recognizing WHDs (Table 1). However, in parallel with the diverse biological functions of RNA, RNA-binding WHDs expand the array of recognized nucleic acid structural features.

A strand-separating wedge

Besides the above regulatory proteins and nucleic-acid-modifying enzymes, WHDs are also present in several helicases, which are enzymes that use NTP (nucleoside triphosphate) hydrolysis to fuel the unwinding of nucleic acid strands. Helicases possessing WHDs include the RecQ family [helicase superfamily (SF) 2], RuvB (SF6), the Ski2 family (SF2), and the DEAH box/RNA helicase A (RHA) family (SF2).

RecQ-family helicases are essential for DNA repair and genome maintenance both in bacteria and eukaryotes [28,29]. The human genome encodes five RecQ helicases termed RECQ1, Bloom's syndrome helicase (BLM), Werner's syndrome helicase (WRN), RECQ4 [Rothmund-Thomson syndrome helicase (RYS)] and RECQ5. Three of these paralogs are affected in genetic diseases: BLM in Bloom's syndrome, WRN in Werner's syndrome, and RECQ4 mainly in Rothmund–Thomson syndrome [28]. Besides their two tandem RecA-type motor domains, RecQ helicases possess various other structural elements including the family-specific RecQ C-terminal (RQC) region, which comprises a zinc binding domain (ZBD) and a WHD (Figure 2A–C). RecQ helicases exert their biological roles by processing complex DNA structures formed transiently during DNA replication and recombination [Holliday junctions (HJs), displacement loops (D-loops), replication fork-like structures, G-quadruplexes (G4)], generally in a non-sequence-specific manner.

Recent structural and biochemical studies support the involvement of the WHD in DNA binding by RecQ helicases [30–33]. Given the plasticity of the WHD–DNA interaction, it is not surprising that the WHDs of RecQ helicases interact with DNA in a manner markedly different from other WHD-containing proteins. However, RecQ helicase WHDs also have a unique functional adaptation: a subgroup of these enzymes has acquired a strikingly specific function as a DNA strand-separating wedge (Table 2). For instance, the crystal structure of the isolated WHD of human WRN complexed to a blunt-ended dsDNA substrate has shown that the W1 region forms a β-hairpin; an aromatic side chain located at the tip of the pin wedges between the last base pair of the substrate and flips out the base at the 5′ terminus (Figure 2D) [32]. This arrangement suggested a key role of W1 and the tip residue in dsDNA unwinding, which has been substantiated by mutagenesis studies [32,34]. The recognition helix does not interact with DNA in the WRN WHD crystal structure, albeit the NMR structure of the DNA-free WHD and mutational studies [35] indicate that it plays an indirect role in DNA binding. The DNA strand-separating role of the WHD was corroborated by studies in which a WRN construct comprising only the two RecA domains and the ZBD showed drastically weakened unwinding activity [36].

Structural and biochemical studies of another human RecQ helicase, RECQ1, revealed a similar involvement of its WHD in dsDNA unwinding [30,31]. RECQ1 processes DNA substrates containing ssDNA–dsDNA junctions ('tailed' dsDNA containing ssDNA overhangs) and/or branches (HJs and D-loops) (Table 2). The recognition helix of the RECQ1 WHD is buried and lies distant from the bound DNA substrate, whereas an aromatic side chain at the W1 tip wedges the junction point of the 3′- tailed DNA ligand (Figure 2E). Similar to the case of WRN, mutational analysis of RECQ1 has confirmed the importance of the WHD and the W1 tip residue in dsDNA unwinding [31].The above findings reveal that, in contrast to RFX1-like transcription factors, which utilize W1 for sequence-specific dsDNA recognition, RECQ1 and WRN use the same structural element for non-sequence-specific unwinding of DNA strands (Figures 1D, 2D–E). Interestingly, the W1 wing of the WHD of WRN and RECQ1 appears to be functionally equivalent to a DNA-separating β-hairpin located in the RecA (WHD-less) core of SF1 ({plasmid copy reduced (Pcr)A [37], UV resistance (Uvr)D [38], Rep [39], RecD2 [40]}), and SF2 (Hel308 [41]) helicases.

Contrary to expectation, the essential DNA strand-separating role of the WHD – and/or the W1 wing contained therein – does not appear to be a universal feature of RecQ helicases. In the DNA-free crystal structure of *Ec* RecQ helicase (Figure 2B), the W1 hairpin [13 amino acid residues (aa) including the β-strands] is shorter than that of WRN (16 aa) and RECQ1 (18 aa) (Figure 2F) [42]. Although it cannot be excluded that DNA binding may affect the conformation of this region, the nonessential nature of W1 in *Ec* RecQ is reflected in the finding that amino acid substitutions or deletions in W1 did not affect the dsDNA unwinding activity of this enzyme [31]. Although W1 appears dispensable for dsDNA unwinding, it has been shown that the RQC region of *Ec* RecQ, which contains the WHD, binds forked and G4 DNA substrates with a similar affinity as the full-length enzyme, indicating its role in recognition of complex DNA structures [33].

No atomic structure of the motor core and RQC regions of BLM is available. However, recent biochemical studies showed that the deletion of the entire WHD left the dsDNA

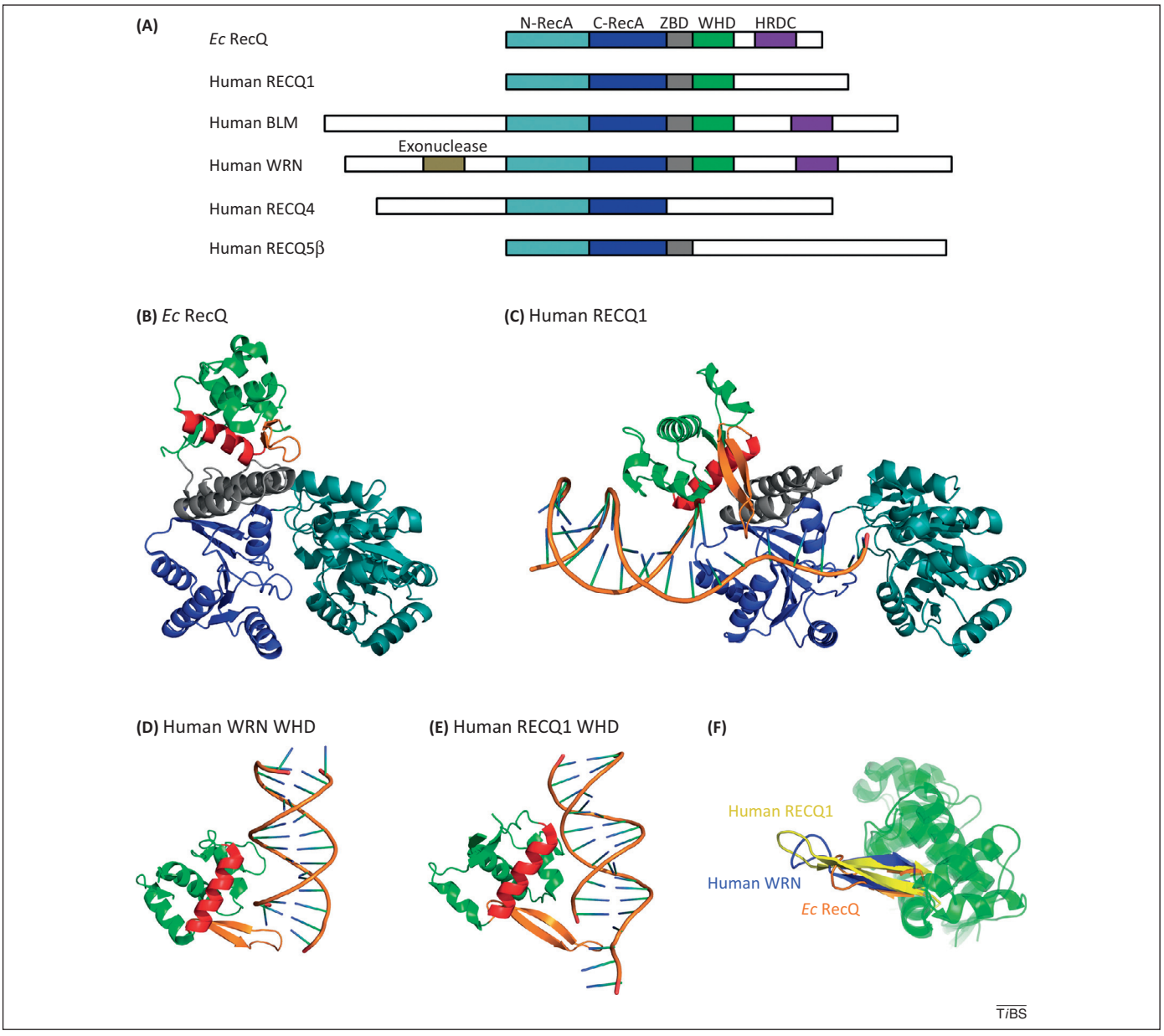

Figure 2. Winged helix domain (WHD) in RecQ-family helicases. **(A)** Schematic representation of the domain structure of *Escherichia coli* (*Ec*) and human RecQ helicases. Crystal structures of **(B)** truncated *Ec* RecQ in the absence of DNA (PDB ID: 1OYW [42]) and **(C)** truncated human RECQ1 bound to 3′-tailed double-stranded (ds) DNA (PDB ID: 2WWY [31]). Domains are colored as in (A). In B–E, the recognition helix is colored red, and W1 with adjacent β-strands is colored orange. Crystal structures of theWHD of **(D)** human WRN bound to blunt-ended dsDNA (PDB ID: 3AAF [32]), and **(E)** human RECQ1 bound to 3′-tailed dsDNA (PDB ID: 2WWY [31]). **(F)** Superimposition of WHD structures of *Ec* RecQ (PDB ID: 1OYW [42], with the W1 hairpin colored orange), human RECQ1 (PDB ID: 2WWY [31], W1 hairpin in yellow), and human WRN (PDB ID: 3AAF [32], W1 hairpin in blue). (For clarity, the WHD is shown in different orientations in panels (D, E) and (F)).

unwinding activity of the enzyme unaffected [43]. Although a BLM core construct lacking the WHD was capable of unwinding even complex multistranded DNA structures (D-loop-like substrates), it has not been assessed whether the domain deletion affects the mechanism and/or efficiency of processing complex DNA structures [43]. Similar to that observed for *Ec* RecQ, the BLM RQC region binds various DNA substrates with similar affinity as the full-length enzyme [33].

The functional significance of the WHD is also reflected in the enzymatic properties of human RecQ helicase paralogs naturally lacking this domain: RECQ4 and RECQ5β (the only splice form of RECQ5 expressed as a protein) can unwind dsDNA substrates only weakly and under specific conditions (Table 2) [44–46].

Taken together, the available data allow the delineation of two subgroups of RecQ helicase enzymes based on the functional necessity of the WHD (and specifically, W1) in dsDNA unwinding: it plays an essential role in human WRN and RECQ1, whereas it is dispensable in the basic dsDNA unwinding mechanism of *Ec* RecQ and human BLM (Table 2).

Synthesis of the available structural information on RecQ helicase enzymes highlights two additional intriguing aspects of the function and structural dynamics of the WHD. First, the essential versus nonessential nature of W1 in DNA separation appears to correlate with the length of the W1 hairpin: the functionally essential β-hairpin is longer in WRN and RECQ1 than the nonessential one in *Ec* RecQ (Figure 2F, Table 2). Second, the WHD adopts two

Table 2. Involvement of the WHD in DNA unwinding by RecQ helicases

	Ec RecQ	Human RECQ1	Human BLM	Human WRN	Human RECQ4	Human RECQ5β
WHD present	yes	yes	yes	yes	no	no
Essential role of W1 in unwinding	no [31]	yes [31]	no [43]	yes [32,34]		
Unwinding of DNA substrates[a]						
Forked (splayed-arm)	++ [67,68]	++ [30,31,69]	++ [43,70]	++ [70]	+ [45] – [44]	+ [46]
Holliday junctions	++ [67]	++ [30,31,69]	++ [70,71]	++ [70]	– [44]	+ [46]
Displacement loops	++ [67]	++[69] +[72]	++ [71]	++ [73]	– [44]	n.a.
G-quadruplexes	++ [33,68]	– [72]	++ [70]	++ [70]	– [44]	n.a.

[a]++, strong activity; +, weak or conditional activity; –, no detectable activity; n.a., no data available.

strikingly different orientations relative to the RecA ZBD helicase core in *Ec* RecQ and human RECQ1 crystal structures (Figure 2B,C). The orientation appears to be unaffected by the binding of ligands to the enzymes (ATPγS nucleotide analog in *Ec* RecQ, and ADP or DNA in RECQ1). The fundamental question that remains to be resolved is whether the different WHD orientations reflect conformational transitions occurring during the working cycle of RecQ helicases, or they are specific to the individual RecQ helicase homolog and do not form sequentially during the working cycle of any one specific enzyme.

Emerging role of the WHD as a branch-recognizing structure

Synthesis of the available data on the DNA unwinding activities of RecQ helicases on DNA substrates with different structures (Table 2) reveals that all investigated RecQ helicase enzymes possessing a WHD can efficiently unwind forked (splayed-arm), HJ, D-loop, and G4 substrates (with the single exception that RECQ1 cannot unwind G4). By contrast, RecQ helicases naturally lacking the WHD exhibit weak or no unwinding activities. Most DNA structures investigated contain some form of branch or strand junction elements, therefore, it is plausible to hypothesize that, besides its direct strand-separating role in some RecQ family members, the WHD plays a more general role in the recognition of complex multistranded DNA structures. A further intriguing hypothesis is that the WHD might be an important determinant of the functional specificity of different human RecQ helicase paralogs. It should also be noted that the DNA binding properties and substrate specificities of various RecQ helicases are also influenced by the helicase and RNaseD C-terminal (HRDC) domain [47,48], therefore the WHD would not be the only determinant of these properties.

Indications for the branch-recognizing function of helicase WHDs have also been obtained from the hexameric RuvB helicase, which forms part of the RuvABC complex that catalyzes branch migration and resolution of HJ DNA recombination intermediates. In each RuvB monomer, a C-terminal WHD is appended to the AAA+-type motor domain [49]. Mutational analysis of *Thermus thermophilus* RuvB has shown that its WHD has a crucial role in HJ recognition, albeit the isolated WHD is unable to bind DNA [50]. Structural modeling suggests that the WHDs of the RuvB hexamer interact with the dsDNA arms of the HJ structure [51]. The electrostatic properties of the recognition helix and W1 suggest a HNF3γ-like but sequence-independent dsDNA-binding mode for the RuvB WHD.

Taken together, structural and functional studies of the WHD revealed a 'bulldozer' (DNA strand separating) role for the W1 wing of this domain in human RecQ helicase family members RECQ1 and WRN (but not in *Ec* RecQ and human BLM), whereas the WHD generally appears to confer specificity for branched DNA structures in both RecQ-family and RuvB helicases.

WHDs also mediate PPIs

Besides its versatile functions brought about by interactions with nucleic acids, the WHD is also involved in various inter- and intramolecular PPIs (Table 1). The two subunits of the heterodimeric E2F4–DP2 transcription factor assemble through the interaction of the recognition helix of each subunit with the H1 and recognition helices of the other [8]. One of the WHDs of restriction endonuclease flavobacterium okeanokoites (Fok)I makes only a few contacts with DNA, and is assumed to play more dominant roles in PPIs [52]. Similarly, the WHD of the 32-kDa subunit of the heterotrimeric human replication protein A (RPA) provides a binding interface for other proteins, whereas it lacks interaction with DNA [53].

WHD structures were recently determined for members of the Ski2 family (SF2) of DNA {Hel308 [41], Holliday-junction migration (Hjm) [54]} and RNA {Brr2 [55], superkiller protein (Ski)2 [56,57], mRNA transport regulator (Mtr)4 [58]} helicases. Hel308 is a DNA repair helicase present in archaea and eukaryotes [59]. Residues of the recognition helix of Hel308 WHD are not conserved, questioning its role in DNA binding. Indeed, substitution of recognition helix residues had an insignificant effect on the dsDNA unwinding activity of *Archaeoglobus fulgidus* (*Af*) Hel308 [41]. In the DNA-bound structure of this enzyme, no part of the WHD interacts with the bound DNA substrate. Instead, the WHD interacts with one of the two tandem RecA-type motor domains, positioning it toward the so-called ratchet domain. This arrangement appears to be important for the formation of a ring-like structure of the Hel308 monomer around DNA. An additional role of the Hel308 WHD in intermolecular PPIs is suggested by the finding that an intact WHD is necessary for the interaction of Hel308 with the Holliday junction cleavage (Hjc) nuclease [59].

Ski2 RNA helicases exhibit additional intriguing structural properties. *Saccharomyces cerevisiae* (*Sc*) Brr2, functioning in RNA spliceosome activation and disassembly, is unique in possessing two entire helicase cores with attached WHDs [55]. In the case of *Sc* Ski2 and *Sc* Mtr4, which are components of cytosolic and nuclear exosome complexes, respectively, two other domains are inserted within the WHD, resulting in the disruption of the conserved WH fold [57,58]. Little is known about the functions of Ski2 RNA helicase WHDs; they are generally postulated to contribute to the maintenance of the overall protein structure by intramolecular PPIs, but their role in RNA binding and intermolecular PPIs has not been ruled out.

Recent experimentally determined or homology-modeled structures of WHD-containing DEAH/RHA RNA helicase family members {spliceosomal *Sc* pre-mRNA splicing protein (Prp)22 [60] and *Sc* Prp43 [61], and ribosomal EAH-box protein (DHX)29 [62]} generally resemble that of *Af* Hel308 [41,59]. The function of the WHDs of these helicases is yet to be determined.

The WHD also mediates PPIs in various RecQ helicases. Besides its roles in interacting with DNA substrates, the WHD of RECQ1 has been demonstrated to mediate homodimerization of the enzyme [30]. The WHD of *Ec* RecQ has also been shown to participate in PPIs between the helicase and the single-stranded binding protein (SSB) [63]. In addition to DNA-binding properties, the WHD of WRN has been shown to mediate PPIs with a variety of binding partners [64]. BLM has also been shown to interact with Flap endonuclease (FEN)-1 through its RQC region containing the WHD [65]. The ability of the WHD to interact with a wide spectrum of proteins raises the idea that the *in vivo* functional specificities of RecQ helicases that otherwise show similar DNA substrate specificity profiles (e.g., BLM and WRN) are pinpointed by their different PPIs, which may be governed at least partially by their WHDs.

Concluding remarks

The examples discussed in this review represent only a subset of identified WHD-containing proteins. However, it is apparent that the WHD is an extremely versatile protein fold that exerts a multitude of functions based on its ability to form specific interactions with nucleic acids and proteins in different structural arrangements (Table 1). In the majority of studied cases, WHD–nucleic acid interactions are dominantly mediated by the recognition helix and/or the W1 wing (Figure 1). Besides their roles as recognition 'keys' in various transcription factors and nucleic-acid-modifying proteins, WHDs attached to the motor cores of certain helicases have acquired the role of a DNA-strand-separating wedge (Figure 2C–F, Table 2). An additional, perhaps more general role of WHDs in helicases appears to be the targeting of enzymes to specific DNA branches and strand joints to aid the efficient processing of intermediates of recombination and other DNA metabolic processes (Table 2). Additional systematic studies comparing the binding affinities and unwinding efficiencies of native and WHD-modified enzymes on various DNA structures will be necessary to clarify this emerging role of WHDs (Box 1).

Taken together, the available data indicate that WHDs have adapted to exploit practically the full spectrum of nucleic acid structural features for molecular recognition (Table 1). Beyond the extensive current knowledge of protein–nucleic acid interactions governed by DNA/RNA base sequence, the versatility of WHDs suggests that they may become prominent players in broadening our knowledge of the structural principles of nucleic acid recognition and modification (Box 1). Research in this area may greatly contribute to a submolecular-level understanding of genetic processes and aid future attempts to design proteins with desired nucleic acid interaction profiles.

Box 1. Outstanding questions

- How does the WHD contribute to the recognition of branched DNA structures and strand junctions?
- Does the WHD undergo significant conformational rearrangements during the working cycle of WHD-containing enzymes?
- How do the interactions of the WHD with DNA and other proteins contribute to the functional diversity of RecQ helicases?

Acknowledgments

Work in M.K.'s laboratory is supported by the Human Frontier Science Program (RGY0072/2010 to M.K.) and the 'Momentum' Program of the Hungarian Academy of Sciences (LP2011-006/2011 to M.K.). M.K. is a Bolyai Fellow of the Hungarian Academy of Sciences. M.G. is a Marie Curie Fellow. Some of the ideas expressed in this work were published earlier [66].

References

1 Aravind, L. *et al.* (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* 29, 231–262
2 Clark, K.L. *et al.* (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412–420
3 Gajiwala, K.S. and Burley, S.K. (2000) Winged helix proteins. *Curr. Opin. Struct. Biol.* 10, 110–116
4 Gajiwala, K.S. *et al.* (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921
5 Passner, J.M. and Steitz, T.A. (1997) The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2843–2847
6 Teplova, M. *et al.* (2006) Structural basis for recognition and sequestration of UUU(OH) 3′ temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell* 21, 75–85
7 Benayoun, B.A. *et al.* (2011) Forkhead transcription factors: key players in health and disease. *Trends Genet.* 27, 224–232
8 Zheng, N. *et al.* (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* 13, 666–674
9 Zhang, A.P.P. *et al.* (2010) Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* 466, 883–886
10 Sharrocks, A.D. (2001) The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* 2, 827–837
11 Siwek, W. *et al.* (2012) Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res.* 40, 7563–7572
12 Daniels, D.S. *et al.* (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* 11, 714–720
13 Wang, G. and Vasquez, K.M. (2007) Z-DNA, an active element in the genome. *Front. Biosci.* 12, 4424–4438
14 Schwartz, T. *et al.* (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* 284, 1841–1845
15 Placido, D. *et al.* (2007) A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* 15, 395–404
16 Schwartz, T. *et al.* (2001) Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* 8, 761–765
17 Kahmann, J.D. *et al.* (2004) The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2712–2717

18 Rothenburg, S. *et al.* (2005) A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1602–1607
19 Goodman, R.A. *et al.* (2012) ADAR proteins: structure and catalytic mechanism. *Curr. Top. Microbiol. Immunol.* 353, 1–33
20 Athanasiadis, A. *et al.* (2005) The crystal structure of the Zbeta domain of the RNA-editing enzyme ADAR1 reveals distinct conserved surfaces among Z-domains. *J. Mol. Biol.* 351, 496–507
21 Takaoka, A. *et al.* (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* 448, 501–505
22 Kim, K. *et al.* (2011) Solution structure of the Zbeta domain of human DNA-dependent activator of IFN-regulatory factors and its binding modes to B- and Z-DNAs. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6921–6926
23 Ha, S.C. *et al.* (2008) The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20671–20676
24 Yoshizawa, S. *et al.* (2005) Structural basis for mRNA recognition by elongation factor SelB. *Nat. Struct. Mol. Biol.* 12, 198–203
25 Soler, N. *et al.* (2007) Structural insight into a molecular switch in tandem winged-helix motifs from elongation factor SelB. *J. Mol. Biol.* 370, 728–741
26 Wolin, S.L. and Cedervall, T. (2002) The La protein. *Annu. Rev. Biochem.* 71, 375–403
27 Alfano, C. *et al.* (2004) Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat. Struct. Mol. Biol.* 11, 323–329
28 Bernstein, K.A. *et al.* (2010) The RecQ DNA helicases in DNA repair. *Annu. Rev. Genet.* 44, 393–417
29 Larsen, N.B. and Hickson, I.D. (2013) RecQ helicases: conserved guardians of genomic integrity. *Adv. Exp. Med. Biol.* 767, 161–184
30 Lucic, B. *et al.* (2011) A prominent β-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation. *Nucleic Acids Res.* 39, 1703–1717
31 Pike, A.C.W. *et al.* (2009) Structure of the human RECQ1 helicase reveals a putative strand-separation pin. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1039–1044
32 Kitano, K. *et al.* (2010) Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure* 18, 177–187
33 Huber, M.D. *et al.* (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.* 358, 1071–1080
34 Tadokoro, T. *et al.* (2012) DNA binding residues in the RQC domain of Werner protein are critical for its catalytic activities. *Aging* 4, 417–429
35 Hu, J.S. *et al.* (2005) Solution structure of a multifunctional DNA- and protein-binding motif of human Werner syndrome protein. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18379–18384
36 von Kobbe, C. *et al.* (2003) Werner syndrome protein contains three structure-specific DNA binding domains. *J. Biol. Chem.* 278, 52997–53006
37 Velankar, S.S. *et al.* (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97, 75–84
38 Lee, J.Y. and Yang, W. (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* 127, 1349–1360
39 Korolev, S. *et al.* (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90, 635–647
40 Saikrishnan, K. *et al.* (2009) Mechanistic basis of 5′-3′ translocation in SF1B helicases. *Cell* 137, 849–859
41 Büttner, K. *et al.* (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* 14, 647–652
42 Bernstein, D.A. *et al.* (2003) High-resolution structure of the *E. coli* RecQ helicase catalytic core. *EMBO J.* 22, 4910–4921
43 Gyimesi, M. *et al.* (2012) Complex activities of the human Bloom′s syndrome helicase are encoded in a core region comprising the RecA and Zn-binding domains. *Nucleic Acids Res.* 40, 3952–3963
44 Macris, M.A. *et al.* (2006) Biochemical characterization of the RECQ4 protein, mutated in Rothmund-Thomson syndrome. *DNA Repair (Amst.)* 5, 172–180
45 Xu, X. and Liu, Y. (2009) Dual DNA unwinding activities of the Rothmund-Thomson syndrome protein, RECQ4. *EMBO J.* 28, 568–577
46 Garcia, P.L. *et al.* (2004) Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *EMBO J.* 23, 2882–2891
47 Kim, Y.M. and Choi, B.S. (2010) Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. *Nucleic Acids Res.* 38, 7764–7777
48 Kitano, K. *et al.* (2007) Crystal structure of the HRDC domain of human Werner syndrome protein, WRN. *J. Biol. Chem.* 282, 2717–2728
49 Yamada, K. *et al.* (2004) Three-dimensional structural views of branch migration and resolution in DNA homologous recombination. *Curr. Opin. Struct. Biol.* 14, 130–137
50 Ohnishi, T. *et al.* (2005) Structure-function analysis of the three domains of RuvB DNA motor protein. *J. Biol. Chem.* 280, 30504–30510
51 Yamada, K. *et al.* (2002) Crystal structure of the RuvA-RuvB complex: a structural basis for the Holliday junction migrating motor machinery. *Mol. Cell* 10, 671–681
52 Wah, D.A. *et al.* (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature* 388, 97–100
53 Deng, X. *et al.* (2007) Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation. *J. Mol. Biol.* 374, 865–876
54 Oyama, T. *et al.* (2009) Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm. *BMC Struct. Biol.* 9, 2
55 Zhang, L. *et al.* (2009) Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat. Struct. Mol. Biol.* 16, 731–739
56 Zhang, X. *et al.* (2008) Crystal structure of an archaeal Ski2p-like protein from *Pyrococcus horikoshii* OT3. *Protein Sci.* 17, 136–145
57 Halbach, F. *et al.* (2012) The crystal structure of *S. cerevisiae* Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* 18, 124–134
58 Jackson, R.N. *et al.* (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J.* 29, 2205–2216
59 Woodman, I.L. and Bolt, E.L. (2011) Winged helix domains with unknown function in Hel308 and related helicases. *Biochem. Soc. Trans.* 39, 140–144
60 Kudlinzki, D. *et al.* (2012) Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. *Biol. Chem.* 393, 1131–1140
61 Walbott, H. *et al.* (2010) Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J.* 29, 2194–2204
62 Dhote, V. *et al.* (2012) Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3150–E3159
63 Shereda, R.D. *et al.* (2009) Identification of the SSB binding site on *E. coli* RecQ reveals a conserved surface for binding SSB's C terminus. *J. Mol. Biol.* 386, 612–625
64 Lee, J.W. *et al.* (2005) Pathways and functions of the Werner syndrome protein. *Mech. Ageing Dev.* 126, 79–86
65 Sharma, S. *et al.* (2004) Stimulation of flap endonuclease-1 by the Bloom′s syndrome protein. *J. Biol. Chem.* 279, 9847–9856
66 Harami, G. *et al.* (2012) A kulcstól a bulldózerig: szárnyas hélix domének funkcionális adaptációja DNS-kötő fehérjékben. *Biokémia* 36, 30–42
67 Harmon, F.G. and Kowalczykowski, S.C. (1998) RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.* 12, 1134–1144
68 Wu, X. and Maizels, N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res.* 29, 1765–1771
69 Sharma, S. *et al.* (2005) Biochemical analysis of the DNA unwinding and strand annealing activities catalyzed by human RECQ1. *J. Biol. Chem.* 280, 28072–28084
70 Mohaghegh, P. *et al.* (2001) The Bloom′s and Werner′s syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* 29, 2843–2849
71 Bachrati, C.Z. *et al.* (2006) Mobile D-loops are a preferred substrate for the Bloom′s syndrome helicase. *Nucleic Acids Res.* 34, 2269–2279
72 Popuri, V. *et al.* (2008) The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J. Biol. Chem.* 283, 17766–17776
73 Opresko, P.L. *et al.* (2009) The Werner syndrome helicase/exonuclease processes mobile D-loops through branch migration and degradation. *PLoS ONE* 4, e4825