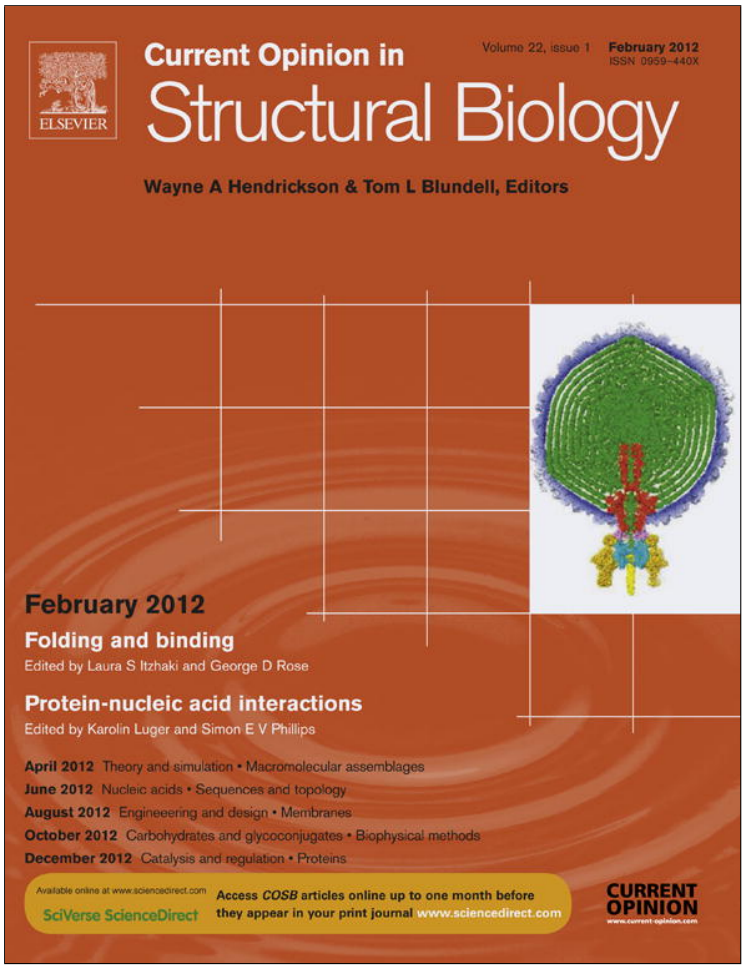


Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Nucleic acid recognition by tandem helical repeats

Emily H Rubinson and Brandt F Eichman

Protein domains constructed from tandem α-helical repeats have until recently been primarily associated with protein scaffolds or RNA recognition. Recent crystal structures of human mitochondrial termination factor MTERF1 and *Bacillus cereus* alkylpurine DNA glycosylase AlkD bound to DNA revealed two new superhelical tandem repeat architectures capable of wrapping around the double helix in unique ways. Unlike DNA sequence recognition motifs that rely mainly on major groove read-out, MTERF and ALK motifs locate target sequences and aberrant nucleotides within DNA by resculpting the double-helix through extensive backbone contacts. Comparisons between MTERF and ALK repeats, together with recent advances in ssRNA recognition by Pumilio/FBF (PUF) domains, provide new insights into the fundamental principles of protein–nucleic acid recognition.

Address
Department of Biological Sciences and Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

Corresponding author: Eichman, Brandt F (brandt.eichman@vanderbilt.edu)

Current Opinion in Structural Biology 2012, **22**:101–109

This review comes from a themed issue on
Protein-nucleic acid interactions
Edited by Simon Phillips and Karolin Luger

Available online 9th December 2011

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2011.11.005

Introduction

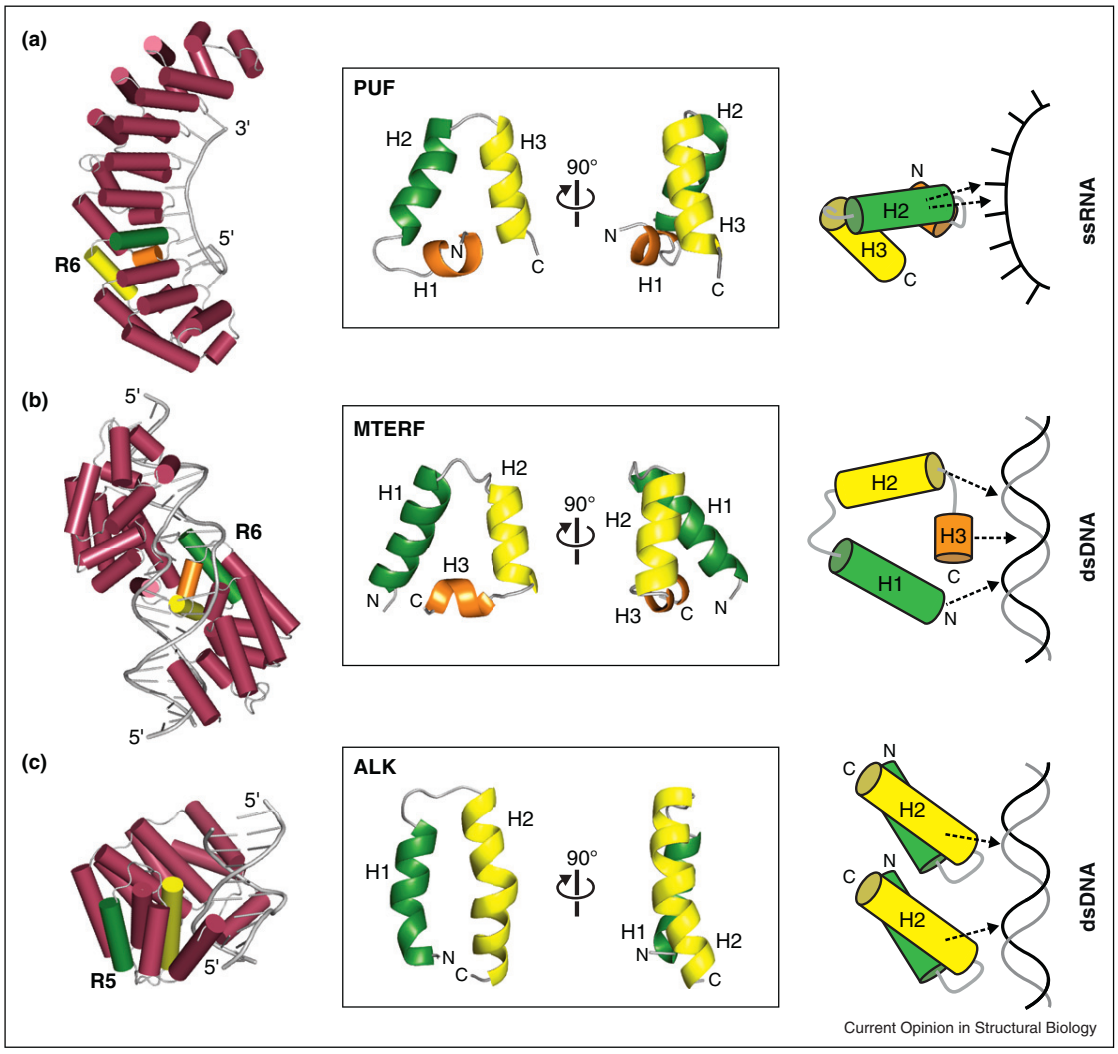
Domains constructed from tandem repeating α-helical motifs are found in all kingdoms of life and serve a variety of functions. Each repeat spans ~25–45 residues and consists of 2–3 helices with a conserved hydrophobic core, and stack in parallel arrays of 3–25 repeats to form extended superhelical or C-shaped structures. Five structural classes of tandem α-helical repeats have been defined: tetratricopeptide (TPR), leucine-rich variant (LRV), ankyrin (ANK), armadillo (ARM), and Huntington/Elongation/A subunit/Target-of-rapamycin (HEAT) repeats (Supplementary Figure S1) (previously reviewed in [1–3]). These motifs are often associated with protein-binding and scaffolding functions, with specificity for a particular ligand dictated by the particular repeat stacking arrangement [4]. For example, superhelical arrays of HEAT repeats are often found in nuclear transport and protein synthesis assemblies that bind their protein cargo within the inner channel of the superhelix [5–7].

ARM and HEAT repeats in particular are emerging as important nucleic acid binding motifs. ARM repeats consist of a two-turn α-helix (H1) perpendicular to two longer antiparallel helices (H2 and H3) that stack to form a right-handed superhelix [8,9]. HEAT motifs are pairs of antiparallel α-helices, with a kinked helix H1 that mimics ARM helices H1 and H2 (Figure S1) [2,10]. The HEAT repeats of Ro autoantigen bind misfolded, small RNAs, and the ARM-like repeat protein Rcd-1 can bind nucleic acids in general [11,12]. The best documented nucleic acid binding helical repeat is the ARM-like PUF motif, which binds ssRNA with exquisite sequence specificity (recently reviewed in [13,14]). In addition to RNA recognition, HEAT repeats form expansive domains in chromatin-remodeling proteins, including the DNA-damage response protein kinases ATM, ATR, and DNA-PK [15,16]. Cryo-EM and a recent crystal structure of DNA-PKcs implicate the HEAT architecture in mediating interactions with both Ku-80 and DNA [17••,18••]. Very recently, crystal structures of human mitochondrial transcription termination factor MTERF1 and bacterial DNA repair protein AlkD provided details for two new ARM-like and HEAT-like motifs in complex with DNA. Here, we compare the different mechanisms of double-stranded DNA and single-stranded RNA recognition by PUF, MTERF and ALK repeats.

PUF repeats

Drosophila melanogaster PUMILIO (PUM) and *Caenorhabditis elegans fem-3* binding factor (FBF), both homologs of human PUMILIO1 (PUM1), are the founding members of the PUF family of translational repressors [13,19,20]. These proteins bind specific sequences in the 3′ untranslated regions of mRNA and subsequently disrupt translation or stimulate degradation of the transcripts. PUF proteins contain eight α-helical repeats, each composed of approximately 36 residues that form a triangular 3-helix bundle similar to ARM repeats (Figures 1a and 2a) [21,22]. Unlike superhelical ARM domains, however, PUFs stack into a crescent shape with the H2 helices lining the concave surface (Figure 1a). This surface forms a binding platform for single-stranded RNA, in which side chains at defined positions of each PUF repeat form alternating hydrogen bonding and stacking interactions with the nucleobases [23] (Figure 2b). Amino acid positions 12 and 16 make sequence specific contacts to the Watson–Crick or Hoogsteen edges of each base, and residue 13 intercalates between the bases (Figures 2c and d and S3a) [23,24•,25••,26•,27].

Figure 1



The three nucleic acid binding helical repeats—PUF, MTERF, and ALK. Crystal structures of PUM1-ssRNA (PDB ID code IM8Y) **(a)**, MTERF1-dsDNA (PDB ID code 3MVA) **(b)**, and AlkD-dsDNA (PDB ID code 3JXZ) **(c)** are shown on the left with one helical repeat colored by helix (orange, green, and yellow). The middle panel shows two orthogonal views of the repeating unit. The schematic on the right illustrates the contacts between each motif and ssRNA (a) or dsDNA (b) and (c).

Structures of several PUF proteins bound to target and non-target RNAs reveal a 'code' for RNA sequence recognition [14,28]. PUF proteins recognize two conserved RNA regions within an 8-nucleotide sequence. C-terminal repeats 6–8 contact a 5-UGU triplet (Figure 2c) and repeats 2–3 recognize an internal AU dinucleotide (Figure 2d) [23,25••]. Selectivity for a particular sequence arises from hydrogen bonds to the bases outside of the UGU and AU motifs, with adenine recognized by cysteine and glutamine, guanine by serine and glutamate, uracil by asparagine and glutamine, and cytosine by arginine and serine [23,25••,29,30•,31•]. Protein–RNA stacking is also important to RNA binding affinity and specificity [32]. Mutating either hydrogen-bonding or stacking residues of a given repeat can alter sequence specificity [29,30•,32].

Variations in the PUF scaffold allow for different RNA binding modes. Typically, each of the eight PUF repeats contacts a single base [23]. Structures of *C. elegans* FBF-2 and yeast Puf4p, however, show that these proteins can accommodate 9-nucleotide RNA sequences as a result of decreased curvature in the protein (Figure 2e), which may be attributed to an extended helix and loop in a centrally

located repeat [25••,26•]. In order to maintain contacts with UGU and AU consensus sequences, one or more nucleotides are flipped away from the protein (Figure 2f). Similarly, a single flipped nucleotide was observed in structures of human PUM1 bound to noncognate 9-mer RNAs, in this case without distortion to the protein itself [24•]. Finally, yeast Puf3 binds a 10-nucleotide sequence with two additional cytosines immediately 5′ to the conserved UGU sequence. To accommodate the 5′-terminal cytosine, which is necessary for gene regulation, into a specialized binding pocket at the C-terminus, the penultimate cytosine is flipped out into the solvent [27].

MTERF repeats

Recent structures from the MTERF family of mitochondrial gene transcription and replication regulators [33] have revealed a new motif capable of binding nucleic acids. MTERF1, a transcription termination factor specific for a 28-nucleotide region near the 3′-end of the 16S rRNA gene [34,35], and MTERF3, a negative regulator of transcription that interacts with the mitochondrial promoter region with no apparent sequence specificity [36], are both constructed from ~30-residue helical repeats that resemble a left-handed ARM motif [37••,38••,39•,40]. Each so-called MTERF repeat consists of two antiparallel α-helices (H1 and H2) followed by a short 3_{10} or α-helix (H3) perpendicular to H1 (Figure 1b). Conserved proline residues between helices H2 and H3 impart a left-handed crossover; these prolines are not present in the right-handed ARM and PUF motifs (Figure S3) [39•,40]. The left-handed MTERF repeats stack together to form a right-handed superhelical protein with H3 helices lining the inside, positively charged surface (Figure 3a and S2a).

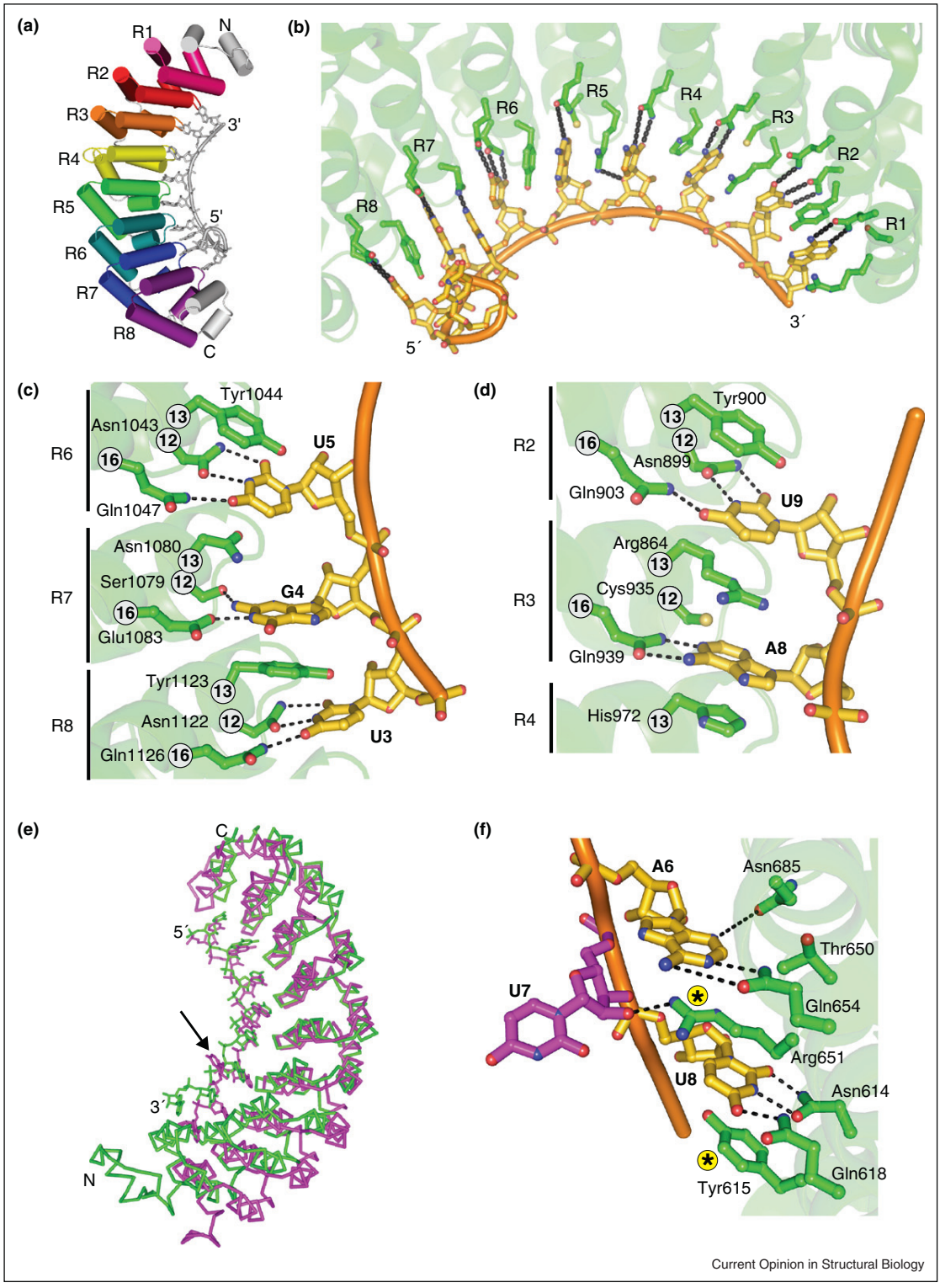
Structures of MTERF1 bound to duplex DNA containing the termination sequence were determined independently by two groups and illustrate how the MTERF architecture anchors itself at the target sequence by wrapping around the DNA major groove and partially denatures the duplex [37••,38••]. The structure determined by Yakubovskaya *et al.* shows a 22-base pair, contiguous DNA helix running through the central channel of the MTERF superhelix (Figure 3b) [37••]. The nine MTERF repeats make a complete turn around the DNA with a pitch of ~70 Å and a footprint of 22 nucleotides. Each repeat binds to the major groove side of DNA primarily through electrostatic interactions, in which the N-terminus of H1 and the C-terminus of H2 contact the backbones from opposing strands and the H3 helix lies in the groove (Figure 1b). The bound DNA is mainly in a B-form conformation, except for a slight 25° bend and a partially unwound 3-nucleotide segment at the center of the duplex. In the Jiménez-Menéndez *et al.* structure, which was determined using a 15-base pair DNA sequence [38••], two symmetry-related DNA molecules are bound at the termini of the superhelix in an identical position as the B-form segments in the continuous DNA structure, leaving a gap that corresponds to the unwound central region (Figure 3c).

In the unwound region of the DNA, three nucleotides are flipped out of the duplex and stacked against Arg162, Phe243, and Tyr288 (RFY) side chains (Figure 3d) [37••]. A structure of an RFY → AAA triple mutant, which eliminates the stacking interactions with extrahelical DNA, showed the same DNA denaturation as the wild-type protein even though the three nucleotides remain stacked in the duplex. Thus, base flipping is not a requirement for helical distortion by MTERF repeats. Rather, base flipping by MTERF1 was shown to be important for stable binding to its specific recognition sequence and consequently for termination activity [37••]. Sequence recognition by MTERF1 is largely determined by five arginine residues that contact guanines in the termination sequence (Figure 3e). Mutation of these residues affects DNA binding to varying degrees, and all are essential for transcriptional termination activity [37••].

ALK repeats

Bacillus cereus AlkC and AlkD are recently discovered DNA glycosylase enzymes that catalyze the excision of positively charged alkylated purine bases from DNA [41]. DNA glycosylases initiate the base excision repair pathway by locating chemically modified nucleobases within DNA, followed by hydrolysis of the N-glycosidic bond to create an abasic site that is processed by other enzymes in the pathway that synthesize a new patch of DNA. Unlike other glycosylases, AlkD is constructed from pairs of antiparallel helices that stack into a short left-handed solenoid (Figures 1c, 4a and S2b), and defines a new structural superfamily of DNA repair enzymes [42•,43•]. Despite similarity to HEAT and TPR motifs (Figure S1), the repeats of AlkD, which we term ALK repeats to distinguish them from their protein scaffolding cousins, are unique in structure and binding properties. ALK motifs, defined by AlkD repeats 2–6, span ~35 residues and contain a slight right-handed twist between helices H1 and H2 (Figures 1c, 4a, S2b). The H1 helix is shorter and straighter than the corresponding helix in HEAT repeats. The H2 helices line the concave surface of the protein and each contributes several basic residues that mediate interaction with dsDNA (Figure 1c and S3) [44••]. The protein makes a half turn around the DNA duplex with a footprint of 12 nucleotides, with all repeats but one (R2) directly contacting the DNA (Figure 4a,b). Depending on their position along the helical axis, ALK repeats contact the backbone of one strand or the other and are thus less specific than MTERF motifs, which contact both strands and the major groove from every repeat (Figure 1b and c). Repeat 1 is unique in that helices H1 and H2 have a left-handed twist and are separated by an inserted helix that helps to anchor the

Figure 2



PUF repeat interactions with ssRNA. (a)–(d) Crystal structures of hsPUM homology domain bound to NRE2-10 RNA (PDB ID code 1M8Y). **(a)** The overall protein architecture is colored by PUF repeat, with ssRNA in silver. **(b)** Protein residues at positions 12, 13, and 16 in each repeat that interact with RNA bases are shown as green sticks. RNA is colored gold, and hydrogen bonds are shown as dashed lines. **(c)** Close-up of UGU recognition by PUF repeats 6, 7, and 8, colored as in panel b. The numbers in gray circles denote canonical amino acid positions 12, 13, and 16 within each repeat. **(d)** AU recognition by

protein to the DNA, and interestingly contributes the only minor groove and base-specific contact from Tyr 27 (Figure 4 and S2 and S3).

Structures of AlkD bound to chemically modified DNA revealed an intriguing mechanism for detecting DNA damage that relies on helical distortion rather than direct interaction with the lesion [44••]. Three AlkD-DNA complexes were determined using short oligonucleotides (10–12 base pairs) containing 3-deaza-3-methyladenine (3d3mA, substrate analog), tetrahydrofuran (THF, product analog), and an inhibitory G·T mismatch (Figure 4b–e). The duplex arms flanking the lesions are essentially B-form with varying degrees of perturbation to the modified base pairs. The 3d3mA·T and G·T base pairs are sheared with the opposite thymine rotated into the minor groove toward the protein (Figure 4c). In the product complex, however, both THF and the opposite thymine are completely flipped out of the duplex, with the THF facing the solvent and the thymine abutted against the protein surface (Figure 4e). As a consequence of the bulge, the duplex collapses to maintain base stacking and creates a distortion in the DNA backbone relative to the substrate structure. In all three structures, the backbone around the extrahelical thymine on the undamaged strand is buried deep in a conserved cleft at the center of the protein and stabilized by a series of electrostatic and van der Waals contacts (Figure 4c and e). The protein does not change conformation upon binding DNA, whereas the conformation of the G·T mispair bound to AlkD was significantly different than a free G·T wobble pair observed in DNA alone. Thus, this cleft resculpts the DNA backbone of the undamaged strand in order to trap the protein against a destabilized base pair. Mutation of residues within this cleft decreased DNA binding activity and catalytic activity [44••].

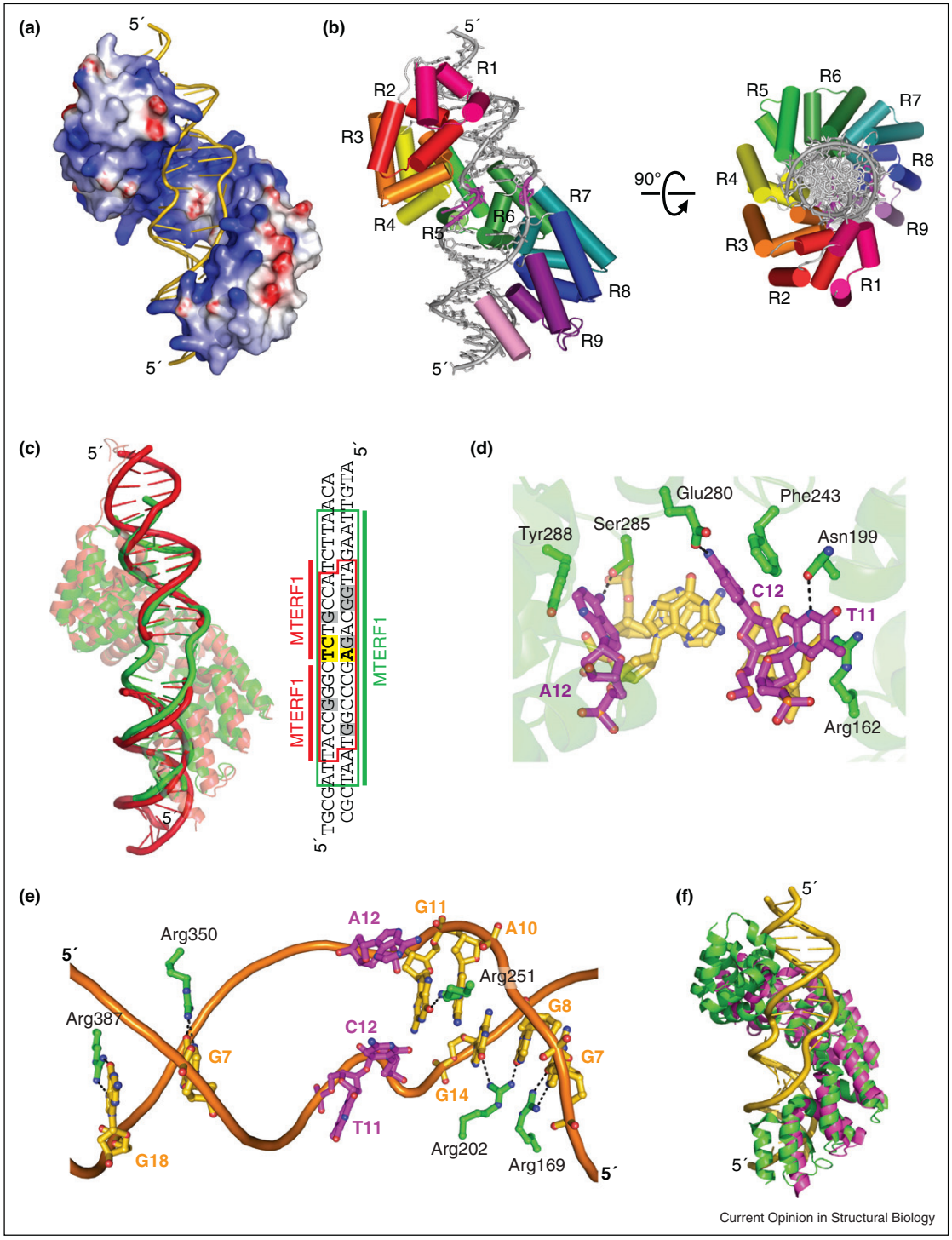
Locating specific targets by direct read-out and nucleic acid resculpting

PUF, MTERF, and ALK domains employ unique mechanisms to locate their target nucleotide sequences. PUF proteins are the most specific with a one-to-one correspondence between each repeat and a single base, wherein a single mutation to a base-binding or stacking residue can alter RNA specificity (Figure 2). The extensive network of nucleobase-side chain intercalation and hydrogen bonding helps explain the low nM binding affinities between PUF domains with their cognate RNA sequences [29]. MTERF1, on the other hand, recognizes a 28-base pair sequence with a K_d of 0.2 μM using a combination of sequence specific and non-specific interactions [37••]. Each repeat contacts both strands over a 5–6 base pair segment and penetrates the major groove in a manner reminiscent of tandem zinc finger modules. The collective MTERF assembly imposes a significant distortion to the DNA duplex, which exposes the major groove to sequence specific interactions from five conserved arginine residues (Figure 3e). The role of each of the five arginine residues in determining MTERF1 specificity for its termination sequence remains to be determined, although some are likely to contribute more to general DNA binding since superposition of MTERF3 onto the MTERF1-DNA structure (Figure 3f) reveals that the positions of three of these arginines (169, 202, and 251) at the N-terminal half of MTERF1 are conserved in MTERF3 (Arg181, His213, Arg257). AlkD is the least specific of the three motifs and relies predominantly on electrostatic contacts with the phosphate backbone for DNA binding (K_d = 2 μM) [43•,44••]. The lack of specific contacts to AlkD's damaged base targets, as well as the distortion imposed on the DNA backbone, strongly suggests that the protein probes for thermodynamically weak points in the DNA that result from alterations in base stacking or base pairing within mismatched or modified bases [44••].

Base flipping is a critical element of RNA and DNA binding specificity among PUF, MTERF, and AlkD proteins. Several PUF proteins flip one or more nucleotides into the solvent as a means of enhancing specificity or providing a recognition element for a partner protein [25••]. Both MTERF1 and AlkD promote extrahelical DNA bases on opposing strands. This base flipping mechanism helps stabilize the protein–DNA complexes and, consequently, is important for the respective biological functions of both proteins [37••,44••]. Stable binding by MTERF1 presumably interferes with the transcription elongation machinery, and base flipping by AlkD holds the damaged base captive in an environment that facilitates catalysis. Unlike other base-flipping enzymes, neither protein intercalates residues into the duplex in order to stabilize the flipped nucleotides. In MTERF1, the unwound duplex is maintained by base stacking side-chains and the extensive binding surface created by nine MTERF repeats, while in the AlkD-DNA structures the duplex collapses on itself in order to preserve stacking interactions. Further work is required to determine whether MTERF and ALK domains recognize pre-distorted DNA or impose torsional strain on the DNA from their superhelical protein scaffolds.

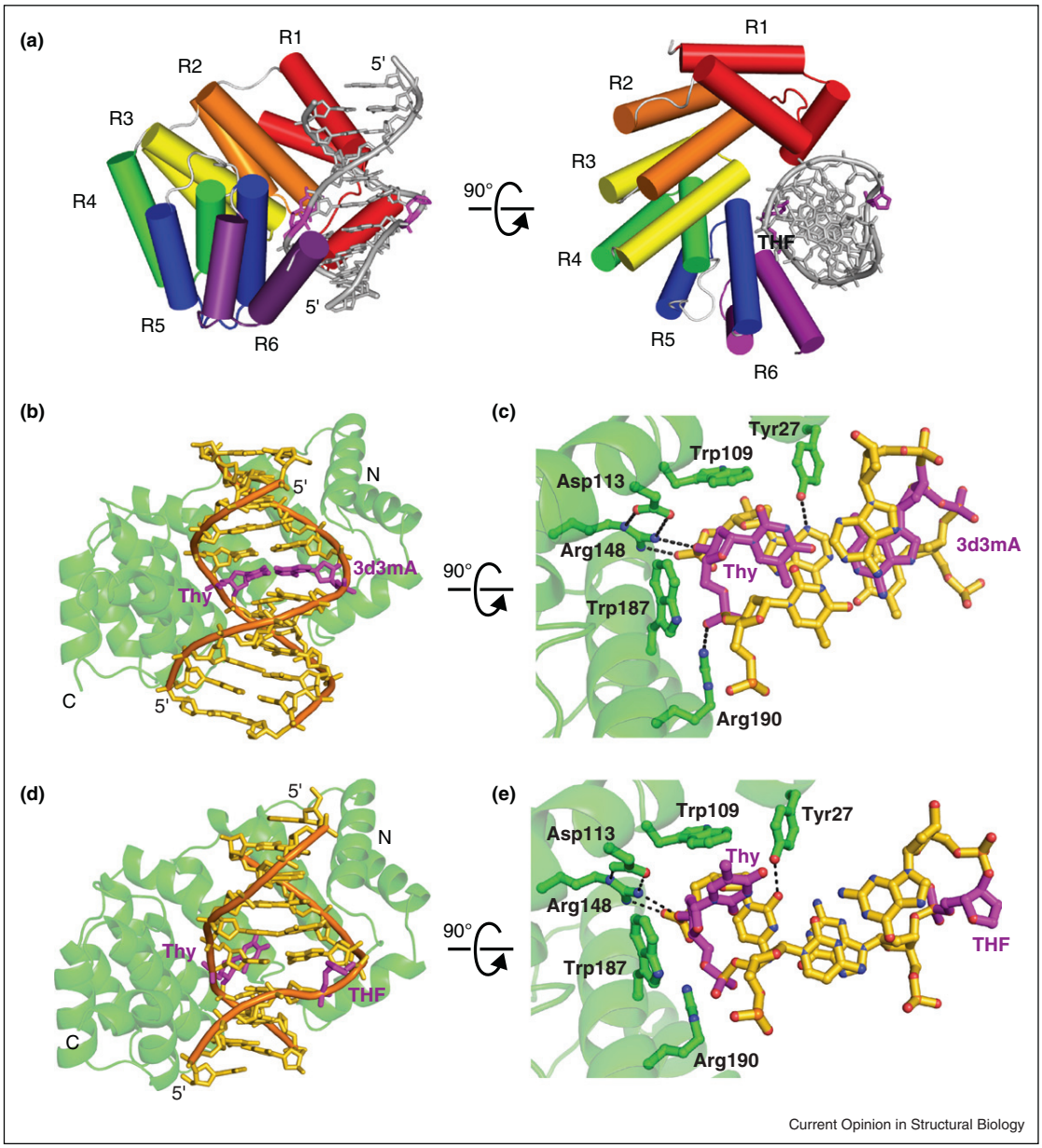
(Figure 2 Legend Continued) PUF repeats 2 and 3, annotated the same as panel c. (e)–(f) Puf4p bound to 9-mer *HO* RNA (PDB ID code 3BX2) [26•]. **(e)** Superposition of Puf4p:*HO* RNA (magenta) and hsPUM:NRE2-10 RNA (green). The protein is rendered as a Cα trace and the RNA is shown as sticks. The extended curvature between repeats 3 and 4 of Puf4p allows it to bind a nine-nucleotide RNA sequence with an extrahelical base (U7, highlighted with an arrow) [26•]. **(f)** Detailed view of the flipped out uracil (magenta) in the Puf4p structure. RNA intercalating side chains are highlighted with yellow asterisks.

Figure 3



DNA binding by MTERF repeats. Each panel shows the crystal structure of human MTERF1 bound to the termination sequence (PDB ID code 3MVA) [37••]. **(a)** Electrostatic surface potential (blue, positive; red, negative) of MTERF1 showing a high degree of positive charge inside the protein superhelix. **(b)** Orthogonal views of MTERF1 (colored by repeat) wrapped around the DNA duplex (silver). Nucleotides flipped out of the DNA are colored magenta. **(c)** Superposition of the Yakubovskaya *et al*. (PDB ID code 3MVA, green) [37••] and Jiménez-Menéndez *et al*. (PDB ID code 3N7Q, red) [38••] MTERF1 structures, showing the correspondence between the continuous 22-bp DNA (green tubes) and two symmetry-related 15-bp DNA (red tubes). The termination sequence recognized by MTERF1 is shown to the right, with oligonucleotides used for crystallization of the respective structures outlined in green and red. The protein footprint in each structure is denoted by the bold vertical lines. The flipped nucleotides are highlighted

Figure 4



Crystal structures of AlkD bound to various forms of modified DNA. **(a)** Orthogonal views of AlkD (colored by repeat) wrapped around a DNA duplex (silver) containing a THF-T bulge (magenta) (PDB ID code 3JXZ). **(b)** and **(c)** Orthogonal views of AlkD (green) bound to 3d3mA·T-DNA (gold, PDB ID code 3JX7). The modified 3d3mA·T base pair is colored magenta. Hydrogen bonds are shown as dashed lines. **(d)** and **(e)** The same views of AlkD in complex with abasic THF·T-DNA.

(Figure 3 Legend Continued) yellow and the guanine residues recognized by the protein are shaded gray. **(d)** Details of the unwound region of DNA in the MTERF1 complex. Extrahelical nucleobases are magenta, other DNA is gold, and side chains important for stabilization of the extrahelical nucleotides are green. Hydrogen bonds are shown as dashed lines. **(e)** Sequence recognition by five arginine residues (green sticks), which form specific contacts with guanines (gold sticks) in the termination sequence. Flipped bases are magenta. **(f)** Superposition of MTERF1-DNA (PDB ID code 3MVA, green) and unliganded MTERF3 (PDB ID code 3OPG, magenta) [39•]. MTERF1 and MTERF3 contain 9 and 7 MTERF repeats, respectively.

Conclusions

The recent advances in RNA sequence recognition by PUF and the discovery of DNA binding by MTERF and ALK motifs open the door for engineering new classes of proteins specific for both RNA and DNA sequences. Indeed, PUF engineering has already begun [14,30•,31•], and successful design of proteins that interact with specific protein/peptide binding partners based on TPR, ANK, LRR/LRV, ARM, and HEAT motifs has flourished in the past ten years [3,45–49]. In addition, the unique characteristics of PUF, MTERF, and ALK motifs may aid in the discovery of nucleic binding elements within protein domains known to contain tandem helical arrays.

Acknowledgements

The authors thank Traci Hall for comments on the manuscript. Work on DNA glycosylases in the author's laboratory is supported by grants from the National Science Foundation (MCB-1122098), the National Institutes of Health (R01 ES019625), and the American Cancer Society (RSG-07-063-01-GMC). E.H.R. is supported by the Vanderbilt Training Program in Environmental Toxicology (NIH T32 ES07028).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.sbi. 2011.11.005.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Groves MR, Barford D: **Topological characteristics of helical repeat proteins**. *Curr Opin Struct Biol* 1999, **9**:383-389.
2. Andrade MA, Perez-Iratxeta C, Ponting CP: **Protein repeats: structures, functions, and evolution**. *J Struct Biol* 2001, **134**:117-131.
3. Main ER, Jackson SE, Regan L: **The folding and design of repeat proteins: reaching a consensus**. *Curr Opin Struct Biol* 2003, **13**:482-489.
4. Grove TZ, Cortajarena AL, Regan L: **Ligand binding by repeat proteins: natural and designed**. *Curr Opin Struct Biol* 2008, **18**:507-515.
5. Chook YM, Blobel G: **Structure of the nuclear transport complex karyopherin-beta2-Ran × GppNHp**. *Nature* 1999, **399**:230-237.
6. Cingolani G, Petosa C, Weis K, Muller CW: **Structure of importin-beta bound to the IBB domain of importin-alpha**. *Nature* 1999, **399**:221-229.
7. Lee SJ, Matsuura Y, Liu SM, Stewart M: **Structural basis for nuclear import complex dissociation by RanGTP**. *Nature* 2005, **435**:693-696.
8. Huber AH, Nelson WJ, Weis WI: **Three-dimensional structure of the armadillo repeat region of beta-catenin**. *Cell* 1997, **90**:871-882.
9. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P: **Comparison of ARM and HEAT protein repeats**. *J Mol Biol* 2001, **309**:1-18.
10. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D: **The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs**. *Cell* 1999, **96**:99-110.
11. Garces RG, Gillon W, Pai EF: **Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties**. *Protein Sci* 2007, **16**:176-188.
12. Stein AJ, Fuchs G, Fu C, Wolin SL, Reinisch KM: **Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity**. *Cell* 2005, **121**:529-539.
13. Kaymak E, Wee LM, Ryder SP: **Structure and function of nematode RNA-binding proteins**. *Curr Opin Struct Biol* 2010, **20**:305-312.
14. Lu G, Dolgner SJ, Hall TM: **Understanding and engineering RNA sequence specificity of PUF proteins**. *Curr Opin Struct Biol* 2009, **19**:110-115.
15. Neuwald AF, Hirano T: **HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions**. *Genome Res* 2000, **10**:1445-1452.
16. Perry J, Kleckner N: **The ATRs, ATMs, and TORs are giant HEAT repeat proteins**. *Cell* 2003, **112**:151-155.
17. •• Williams DR, Lee KJ, Shi J, Chen DJ, Stewart PL: **Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding**. *Structure* 2008, **16**:468-477.
The 7 Å resolution EM analysis of DNA-PKcs revealed a HEAT-repeat architecture and provided a model for DNA binding to an α-helical segment at the core of the protein.
18. •• Sibanda BL, Chirgadze DY, Blundell TL: **Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats**. *Nature* 2010, **463**:118-121.
The first crystal structure of DNA-PKcs provided a clear confirmation that HEAT repeat motifs constitute the majority of the protein, including the putative DNA binding domain within the central channel.
19. Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP: **A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line**. *Nature* 1997, **390**:477-484.
20. Zamore PD, Williamson JR, Lehmann R: **The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins**. *RNA* 1997, **3**:1421-1433.
21. Edwards TA, Pyle SE, Wharton RP, Aggarwal AK: **Structure of Pumilio reveals similarity between RNA and peptide binding motifs**. *Cell* 2001, **105**:281-289.
22. Wang X, Zamore PD, Hall TM: **Crystal structure of a Pumilio homology domain**. *Mol Cell* 2001, **7**:855-865.
23. Wang X, McLachlan J, Zamore PD, Hall TM: **Modular recognition of RNA by a human pumilio-homology domain**. *Cell* 2002, **110**:501-512.
24. • Gupta YK, Nair DT, Wharton RP, Aggarwal AK: **Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity**. *Structure* 2008, **16**:549-557.
This work further illustrated how PUF repeats use base flipping to accommodate extra nucleotides within the recognition sequence, and revealed the plasticity of the PUF architecture.
25. •• Wang Y, Opperman L, Wickens M, Hall TM: **Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein**. *Proc Natl Acad Sci U S A* 2009, **106**:20186-20191.
This comprehensive structural analysis of an archetypal PUF protein, *C. elegans* FBF, in complex with six different RNA sequences, provides general rules for RNA recognition by the PUF proteins.
26. • Miller MT, Higgin JJ, Hall TM: **Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p**. *Nat Struct Mol Biol* 2008, **15**:397-402.
Structures of Puf4p in the absence and presence of RNA showed how variations in the PUF architecture allow for binding to a nine nucleotide RNA sequence, in which the extra base is extruded away from the protein.
27. Zhu D, Stumpf CR, Krahn JM, Wickens M: **Hall TM: a 5′ cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs**. *Proc Natl Acad Sci U S A* 2009, **106**:20192-20197.

28. Chen Y, Varani G: **Finding the missing code of RNA recognition by PUF proteins**. *Chem Biol* 2011, **18**:821-823.

29. Cheong CG, Hall TM: **Engineering RNA sequence specificity of Pumilio repeats**. *Proc Natl Acad Sci U S A* 2006, **103**:13635-13639.

30. • Filipovska A, Razif MF, Nygard KK, Rackham O: **A universal code for RNA recognition by PUF proteins**. *Nat Chem Biol* 2011, **7**:425-427.

One of two recent reports describing the use of a yeast 3-hybrid system to engineer PUF proteins to recognize cytosine in addition to adenine, guanine, and uracil.

31. • Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TM, Wang Z: **Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains**. *J Biol Chem* 2011, **286**:26732-26742.

This work expands PUF-base read-out to include cytosine, which completes recognition of all four bases and thus allows for the potential engineering of any RNA sequence by PUF proteins.

32. Koh YY, Wang Y, Qiu C, Opperman L, Gross L, Tanaka Hall TM, Wickens M: **Stacking interactions in PUF–RNA complexes**. *Rna* 2011, **17**:718-727.

33. Roberti M, Polosa PL, Bruni F, Manzari C, Deceglie S, Gadaleta MN, Cantatore P: **The MTERF family proteins: mitochondrial transcription regulators and beyond**. *Biochim Biophys Acta* 2009, **1787**:303-311.

34. Kruse B, Narasimhan N, Attardi G: **Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination**. *Cell* 1989, **58**:391-397.

35. Daga A, Micol V, Hess D, Aebersold R, Attardi G: **Molecular characterization of the transcription termination factor from human mitochondria**. *J Biol Chem* 1993, **268**:8123-8130.

36. Park CB, Asin-Cayuela J, Camara Y, Shi Y, Pellegrini M, Gaspari M, Wibom R, Hultenby K, Erdjument-Bromage H, Tempst P *et al.*: **MTERF3 is a negative regulator of mammalian mtDNA transcription**. *Cell* 2007, **130**:273-285.

37. •• Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, Garcia-Diaz M: **Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription**. *Cell* 2010, **141**:982-993.

This work provided a clear structural and biochemical basis for MTERF recognition of the complete mitochondrial transcription termination sequence, through a combination of sequence-specific contacts and non-specific helical distortion.

38. •• Jimenez-Menendez N, Fernandez-Millan P, Rubio-Cosials A, Arnan C, Montoya J, Jacobs HT, Bernado P, Coll M, Uson I, Sola M: **Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat**. *Nat Struct Mol Biol* 2010, **17**:891-893.

Crystal structure of MTERF1 bound to DNA containing 15-bp of the termination sequence provided a second illustration for how the MTERF architecture anchors itself against dsDNA protein contains nine left-handed helical tandem repeats similar to ARM repeats.

39. • Spahr H, Samuelsson T, Hallberg BM, Gustafsson CM: **Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain**. *Biochem Biophys Res Commun* 2010, **397**:386-390.

The crystal structure of MTERF3 was the first reported description of the MTERF architecture, and is the only example of this architecture in the absence of DNA.

40. Kim DU, Cho SG, Kim KL, Cho HS: **Human MTERF3 crystal structure forms a left-handed superhelix**. *Mol Cells* 2011.

41. Alseth I, Rognes T, Lindback T, Solberg I, Robertsen K, Kristiansen KI, Mainieri D, Lillehagen L, Kolsto AB, Bjoras M: **A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD**. *Mol Microbiol* 2006, **59**:1602-1609.

42. • Dalhus B, Helle IH, Backe PH, Alseth I, Rognes T, Bjoras M, Laerdahl JK: **Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats**. *Nucleic Acids Res* 2007, **35**:2451-2459.

Comparative sequence analysis and homology modeling identified the HEAT-like architecture of AlkD and implicated several residues in base excision and protection of cells from DNA alkylation damage.

43. • Rubinson EH, Metz AH, O'Quin J, Eichman BF: **A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD**. *J Mol Biol* 2008, **381**:13-23.

The crystal structure of unliganded AlkD provided details for the ALK repeats and mutagenesis confirmed the role of key residues in DNA binding and DNA glycosylase activity.

44. •• Rubinson EH, Gowda AS, Spratt TE, Gold B, Eichman BF: **An unprecedented nucleic acid capture mechanism for excision of DNA damage**. *Nature* 2010, **468**:406-411.

Crystal structures of AlkD bound to DNA containing alkylated, mismatched and abasic nucleotides provided a structural basis by which ALK repeats restructure the DNA duplex to pinpoint damage in a different manner observed for other base excision repair proteins.

45. Stumpp MT, Forrer P, Binz HK, Pluckthun A: **Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family**. *J Mol Biol* 2003, **332**:471-487.

46. Forrer P, Binz HK, Stumpp MT, Pluckthun A: **Consensus design of repeat proteins**. *Chembiochem* 2004, **5**:183-189.

47. Hausrath AC, Goriely A: **Repeat protein architectures predicted by a continuum representation of fold space**. *Protein Sci* 2006, **15**:753-760.

48. Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, Zerbe O, Caflisch A, Pluckthun A: **Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core**. *J Mol Biol* 2008, **376**:1282-1304.

49. Wetzel SK, Ewald C, Settanni G, Jurt S, Pluckthun A, Zerbe O: **Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR**. *J Mol Biol* 2010, **402**:241-258.

Supporting Information

Nucleic Acid Recognition by Tandem Helical Repeats

Emily H. Rubinson and Brandt F. Eichman

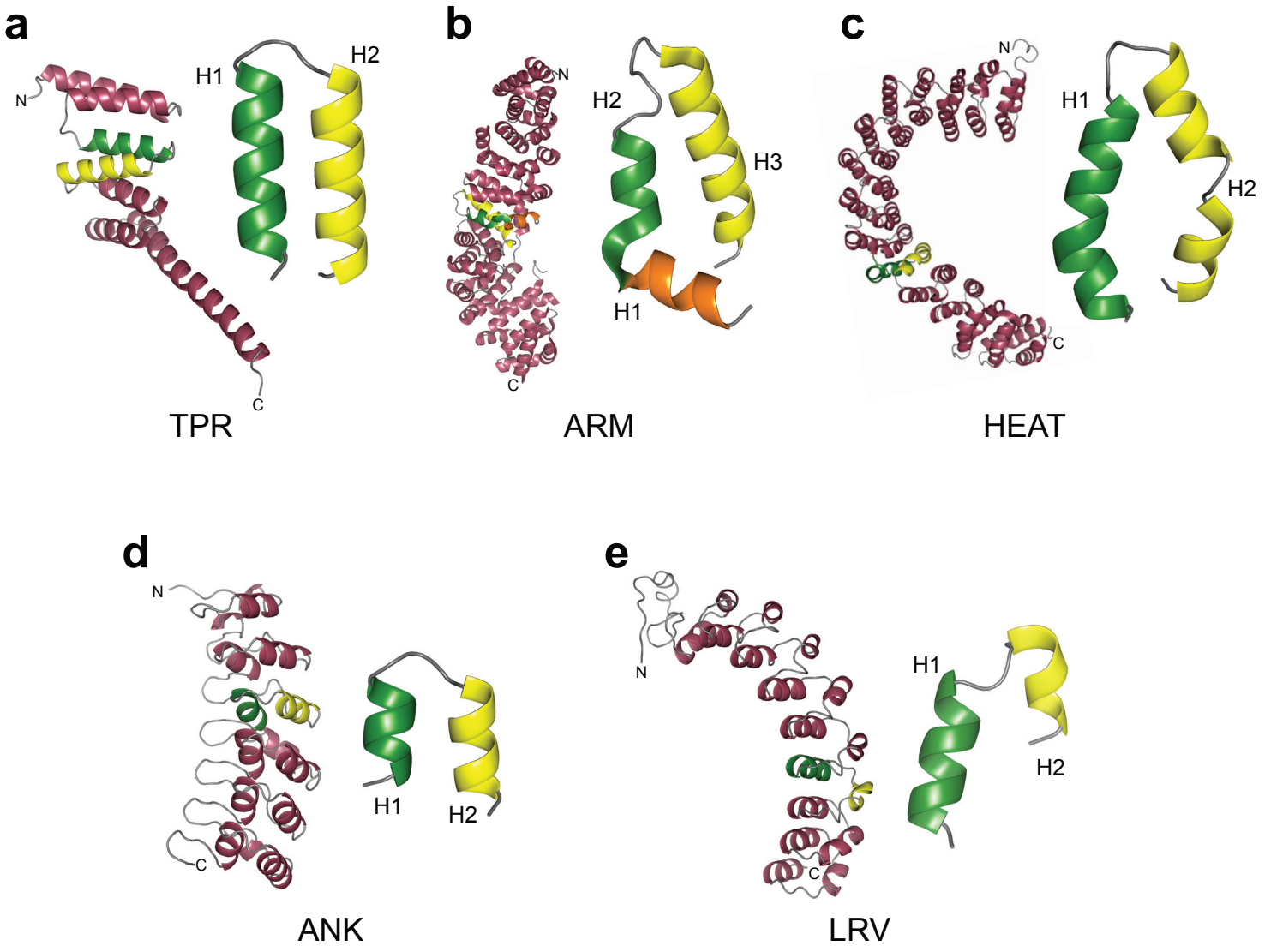


Figure S1. Tandem helical repeat proteins. (**a**) Ribbon representations of five classical repeat structures and their corresponding individual repeating unit colored by helix: tetratricopeptide repeats (TPR) from protein phosphatase 5 (PDB ID code 1A17), armadillo (ARM) repeats from β-catenin (PDB ID code 1JDH), HEAT repeats from protein phosphatase 2A (PP2A) (PDB ID code 1B3U), ankyrin (ANK) repeats of IκBα (PDB ID code 1NFI), and leucine-rich variant (LRV) motif (PDB ID code 1LRV).

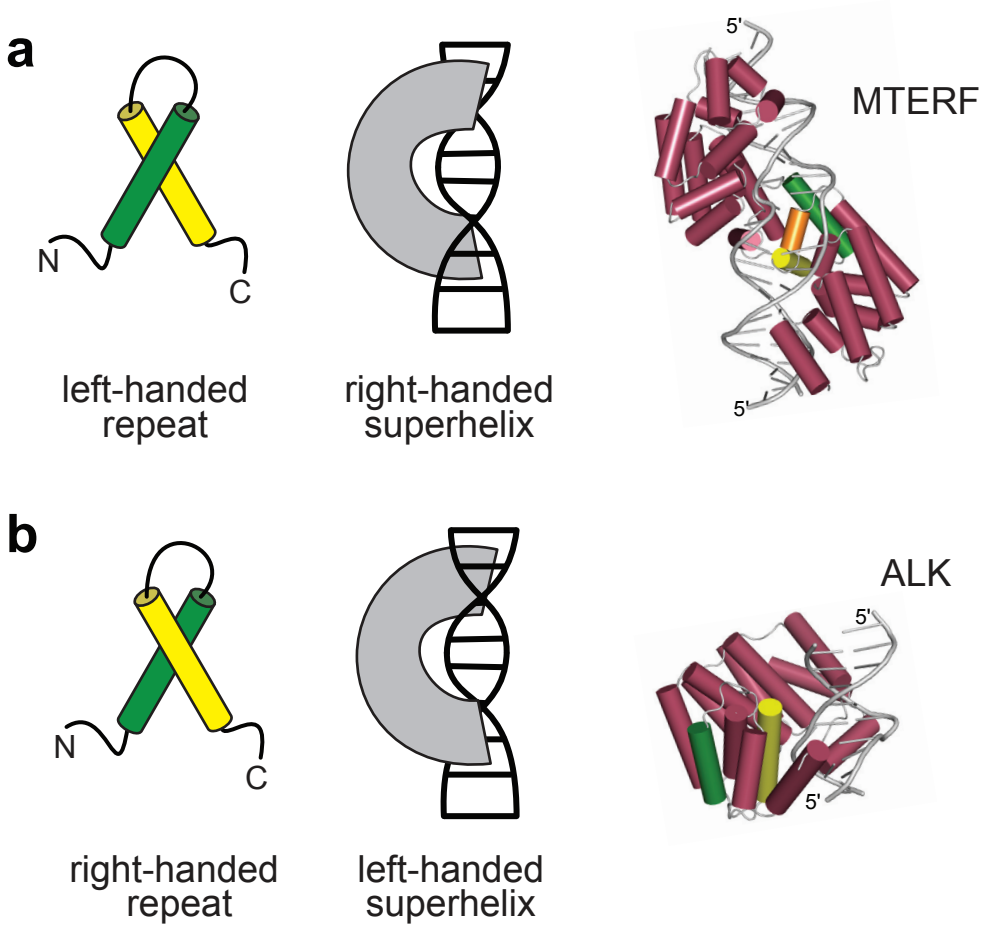


Figure S2. Definitions of handedness of pairwise helical crossovers and the superhelical protein architectures that they typically form. (**a**) MTERF repeats are left-handed arrangements of helices H1 (green) and H2 (yellow) and stack to form right-handed superhelical proteins (grey). (**b**) Right-handed ALK motifs form left-handed superhelical proteins in a similar manner to most TPR, ARM, HEAT, ANK, and LRV repeats.

Figure S3. Structures of individual nucleic acid binding helical repeats. (**a**) PUF repeats of the homology domain of human Pumilio1 (HsPUM-HD) (PDB ID code 1M8Y), (**b**) MTERF repeats of MTERF1 (PDB ID code 3MVA), (**c**) MTERF repeats of MTERF3 (PDB ID code 3M66) and (**d**) ALK motifs of AlkD (PDB ID code 3JXZ). On the left are cylindrical representations of each protein colored by repeat. Each repeat is shown individually as ribbons on the top of each panel. On the bottom are structure-based sequence alignments of individual repeats. Positions 12, 13, and 16 in PUF repeats are labeled. Hydrophobic residues that line the interface between helices are highlighted gray, residues that contact the phosphate-backbone are highlighted yellow, base binding residues, including the sequence specific arginines in MTERF1 and putative arginine counterparts in MTERF3, are highlighted green. Specific prolines that create key turns in the repeats to define the third helix of MTERF1, hPUM-HD, and MTERF3 are highlighted blue. Catalytic residues in AlkD and base stacking residues in MTERF1 and hPUM-HD are in boldface, and positively charged residues that line the concave clefts of the proteins are boxed.

Figure S3

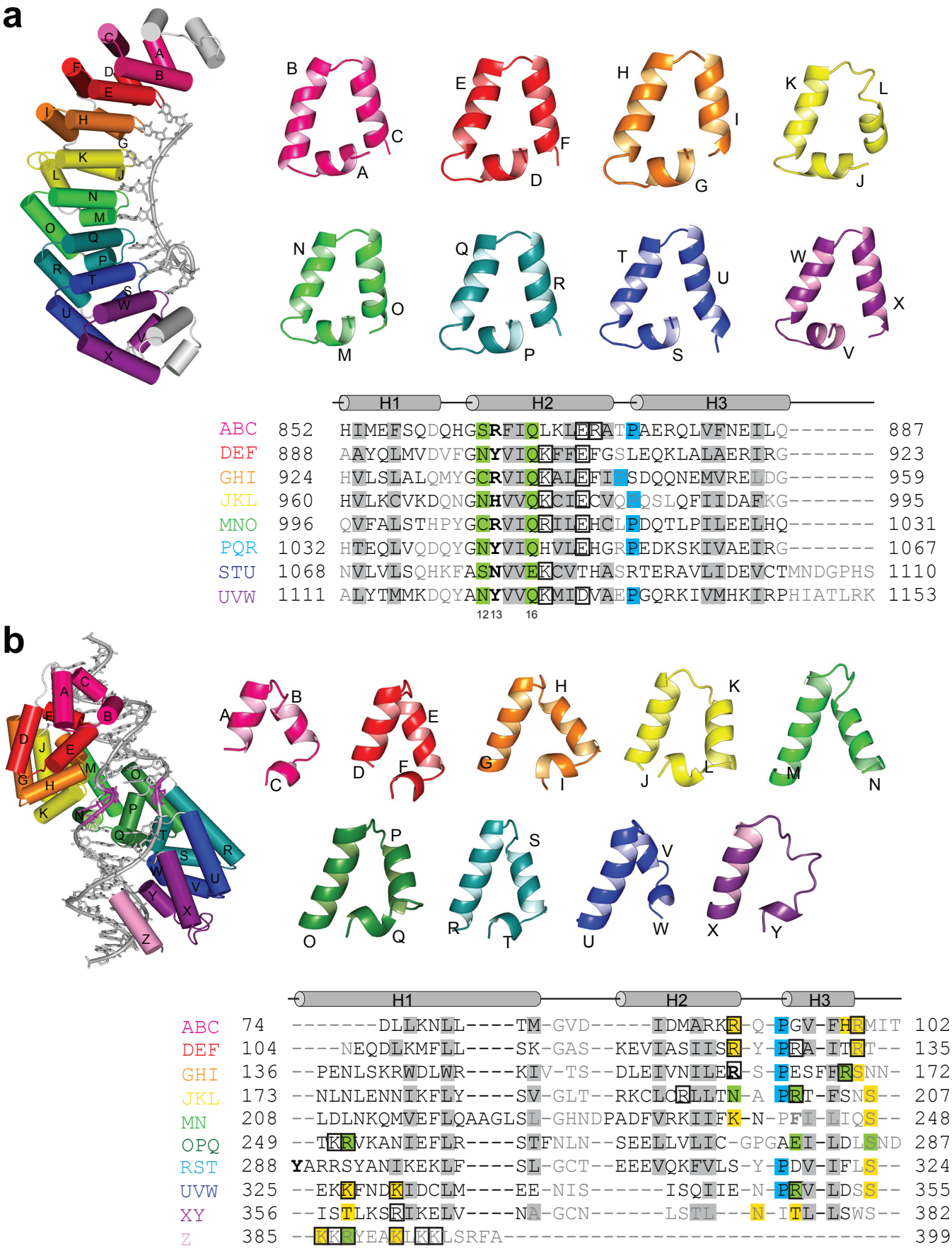
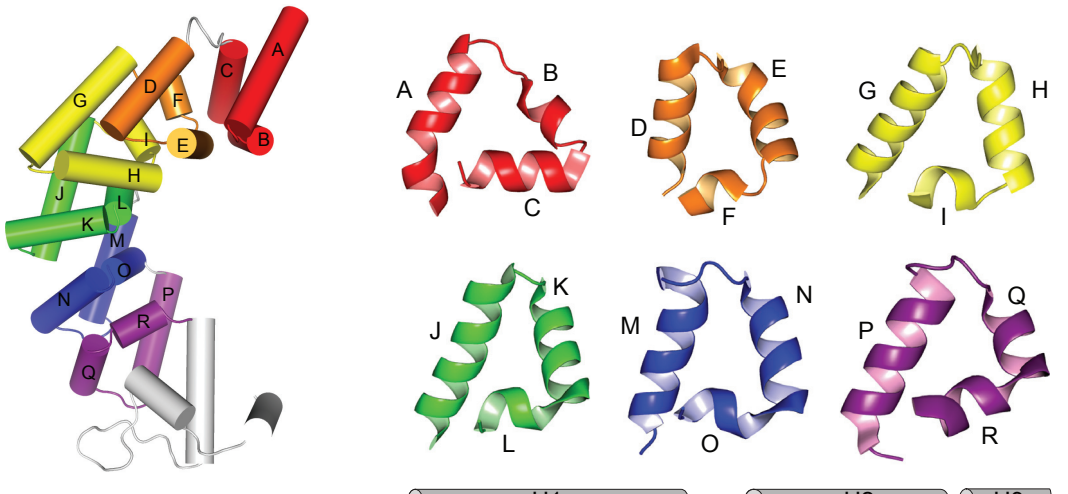

a
b
H1
H2
H3
ABC 852 HIMEFSQDQHGSRFIQLKLERATPAERQLVFNEILQ------- 887
DEF 888 AAYQLMVDVFGNYVIQKFFEFGSLEQKLALAERIRG------- 923
GHI 924 HVLSLALQMYGCRVIQKALEFIFSDQQNEMVRELDG------- 959
JKL 960 HVLKCVKDQNGNHVVQKCIECVQPQSLQFIIDAFKG------- 995
MNO 996 QVFALSTHPYGCRVIQRILEHCLPDQTLPILEELHQ------- 1031
PQR 1032 HTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRG------- 1067
STU 1068 NVLVLSQHKFASNVVEKCVTHASRTERAVLIDEVCTMNDGPHS 1110
UVW 1111 ALYTMMKDQYANYVVQKMIDVAEPGQRKIVMHKIRPHIATLRK 1153
1213 16
H1
H2
H3
ABC 74 -------DLLKNLL----TM-GVD-----IDMARKR-Q-PGV-FHRMIT 102
DEF 104 ----NEQDLKMFLL----SK-GAS--KEVIASIISR-Y-PRA-ITRT-- 135
GHI 136 --PENLSKRWDLWR----KIV-TS--DLEIVNILER-S-PESFFRSNN- 172
JKL 173 --NLNLENNIKFLY----SV-GLT--RKCLQRLLTN-A PRT-FSNS-- 207
MN 208 --LDLNKQMVEFLQAAGLSL-GHNDPADFVRKIIFK-N-PFI-LIQS-- 248
OPQ 249 --TKRVKANIEFLR----STFNLN--SEELLVLIC-GPGAEI-LDLSND 287
RST 288 YARRSYANIKEKLF----SL-GCT--EEEVQKFVLS-Y-PDV-IFLS-- 324
UVW 325 --EKKFNDKIDCLM----EE-NIS------ISQIIE-N-PRV-LDSS-- 355
XY 356 --ISTLKSRIKELV----NA-GCN------LSTL---N-ITL-LSWS-- 382
Z 385 --KKKYEAKLKKLSRFA---------- 399

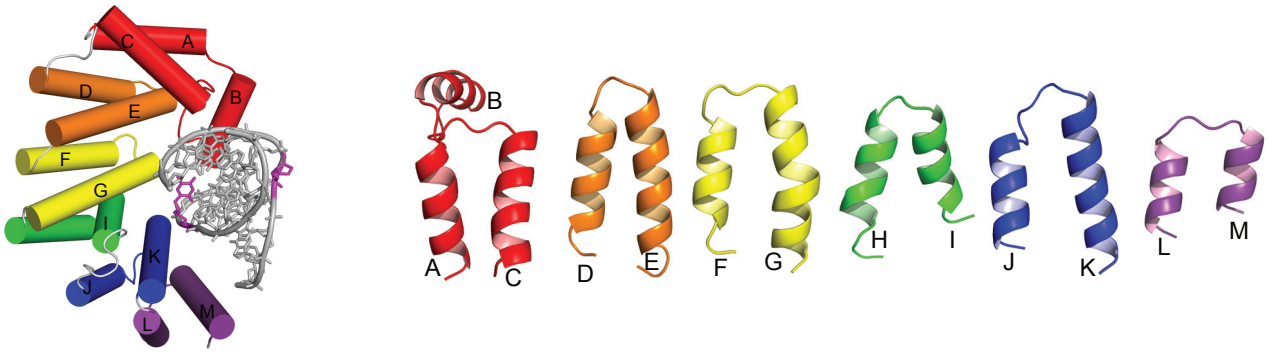
Figure S3 cont.

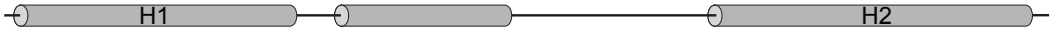
c




```
ABC 148 DYVDHSETLQKLVL-LGVDLSKIEKHPEAANLLLRLD 183
DEF 184 FEKDIKQMLLFLKD-VGIEDNQLGAFLTKNHAIFSE- 218
GHI 219 DLENLKTRVAYLHS-KNFSKADVAQMVRKAPFLLNF- 253
JKL 254 SVERLDNRLGFFQKELELSVKKTRDLVRLPRLLTGS 290
MNO 291 -LEPVKENMKVYRLELGFKHNEIQHMITRIPKMLTA- 325
PQR 326 NKMKLTETFDFVHN--VMSIPHHI--IVKFPQVFNT- 357
```

d






```
ABC   3 ---PFVKALQEHFTAHQNPEKAEPMARYMKNHFLFLGIQTPERRQLLKDVIQIH--- 53
DE   58 -----QKDFQIIIRELWD-------------LPE-----REFQAAALDIMQKMKKH- 90
FG   93 ------ETHIPFLEELIV-------------TKSW----WDSVDSIVPTFLGDIFLK 126
HI  128 -----PELISAYIPKWIAS------------DN------IWLQRAAIL-FQL----- 155
JK  163 -----EELLFWIIGQL---------------HSSKE---FFIQKAIGWVLREYAKT- 195
LM  197 -----PDVVWEYVQN----------------NELA----PLSKREAIKHIKQN---- 224
```