RNA sculpting by the primordial Helix-clasp-Helix–Strand-Loop (HcH–SL) motif enforces chemical recognition enabling diverse KH domain functions

Received for publication, November 11, 2024, and in revised form, February 26, 2025 Published, Papers in Press, April 2, 2025,
https://doi.org/10.1016/j.jbc.2025.108474

John A. Tainer[1,2,3,*] **and Susan E. Tsutakawa**[3,*]

From the [1]*Department of Molecular and Cellular Oncology, and* [2]*Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA;* [3]*Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California, USA*

Reviewed by members of the JBC Editorial Board. Edited by Alex Toker

In all domains of life, the ancient K homology (KH) domain superfamily is central to RNA processes including splicing, transcription, posttranscriptional gene regulation, signaling, and translation. Proteins with 1 to 15 KH domains bind single-strand (ss) RNA or DNA with base sequence specificity. Here, we examine over 40 KH domain experimental structures in complex with nucleic acid (NA) and define a novel Helix-clasp-Helix–Strand-Loop (HcH–SL) NA recognition motif binding 4 to 5 nucleotides using 10 to 18 residues. HcH–SL includes and extends the Gly-X-X-Gly (GXXG) signature sequence "clasp" that brings together two helices as an ∼90° helical corner. The first helix primarily provides side chain interactions to unstack and sculpt 2 to 3 bases on the 5′ end for recognition of sequence and chemistry. The clasp and second helix amino dipole recognize a central phosphodiester. Following the helical corner, a beta strand and its loop extension recognize the two 3′ nucleotides, primarily through main chain interactions. The HcH–SL structural motif forms a right-handed triangle and concave functional interface for NA interaction that unexpectedly splays four bound nucleotides into conformations matching RNA recognition motif (RRM) bound RNA structures. Evolutionary analyses and its ability to recognize base sequence and chemistry make HcH–SL a primordial NA binding motif distinguished by its binding mode from other NA structural recognition motifs: helix-turn-helix, helix-hairpin-helix, and beta strand RRM motifs. Combined results explain its vulnerability as a viral hijacking target and how mutations and expression defects lead to diverse diseases spanning cancer, cardiovascular, fragile X syndrome, neurodevelopmental disorders, and paraneoplastic disease.

The K homology (KH) domain (Pfam PF07650) was first identified in two conserved regions in human heterogeneous nuclear ribonucleoprotein K (hnRNPK), a protein described as "tenaciously" binding cytidine-rich sequences in pre-mRNA (1). Since then over 233,000 proteins that contain at least one KH domain, have been identified in all three domains of life (2) (Fig. 1). Its prevalence throughout the tree of life suggests that the KH sequence is an ancient protein domain with fundamental significance for cells. Proteins can have one, two, or as many as 15 flexibly or rigidly tethered KH domain modules (reviewed in (3)).

Most, but not all KH domains, bind specific sequences in single stranded (ss) RNA and/or DNA nucleic acid (NA). A compilation of sequences bound by KH domains show a general preference for sequences rich in pyrimidines over purines (4), but there are exceptions (3, 5). Acting in splicing, transcription, and translation, KH domains can bind many types of RNA: (rRNA, tRNA, microRNA, long noncoding RNA, small nuclear RNA, small nucleolar RNA, mitochondrial RNAs, and messenger RNA (mRNA) (5–8).

Overexpression or loss-of-function of individual KH domains is associated with diseases. These include cancer, cardiovascular, fragile X syndrome, neurodevelopmental disorders, and paraneoplastic disease. There are multiple examples. The KH domain of ANKHD1 binds and regulates three tumor-suppressing microRNAs, and ANKHD1 overexpression is associated with renal cancer as well as the increased metastasis and larger tumors (7, 9). With 3 KH domains, HnRNPK has activities in pre-mRNA processing and its mutations result in neurodevelopmental disorders with 70% of missense variants clustering in the KH domains important in RNA recognition (10). HnRNPK overexpression is associated with diverse cancers, including breast, colorectal, hepatocellular, gastric, melanoma, nasopharyngeal, lung, esophageal squamous cell, oral squamous cell, and prostate cancers; its cancer connection is thought to be in part due to its long noncoding RNA binding (8). HnRNPK deficiencies reduce cyclin-dependent kinase inhibitor 1 levels that pause replication for DNA repair in a pathway with p53 tumor suppressor (11). Associated with breast and hepatocellular carcinoma and cardiovascular disease, vigilin is the largest KH domain protein with 14 to 15 KH domains. It binds hundreds of different mRNAs and tRNAs, acts in RNA shuttling, chromosomal stability, both positive and negative translation regulation, and mRNA stability (5, 12). Vigilin is linked to autism, DNA repair defects, cancer progression, and

* For correspondence: John A. Tainer, JTainer@mdanderson.org; Susan E. Tsutakawa, setsutakawa@lbl.gov.

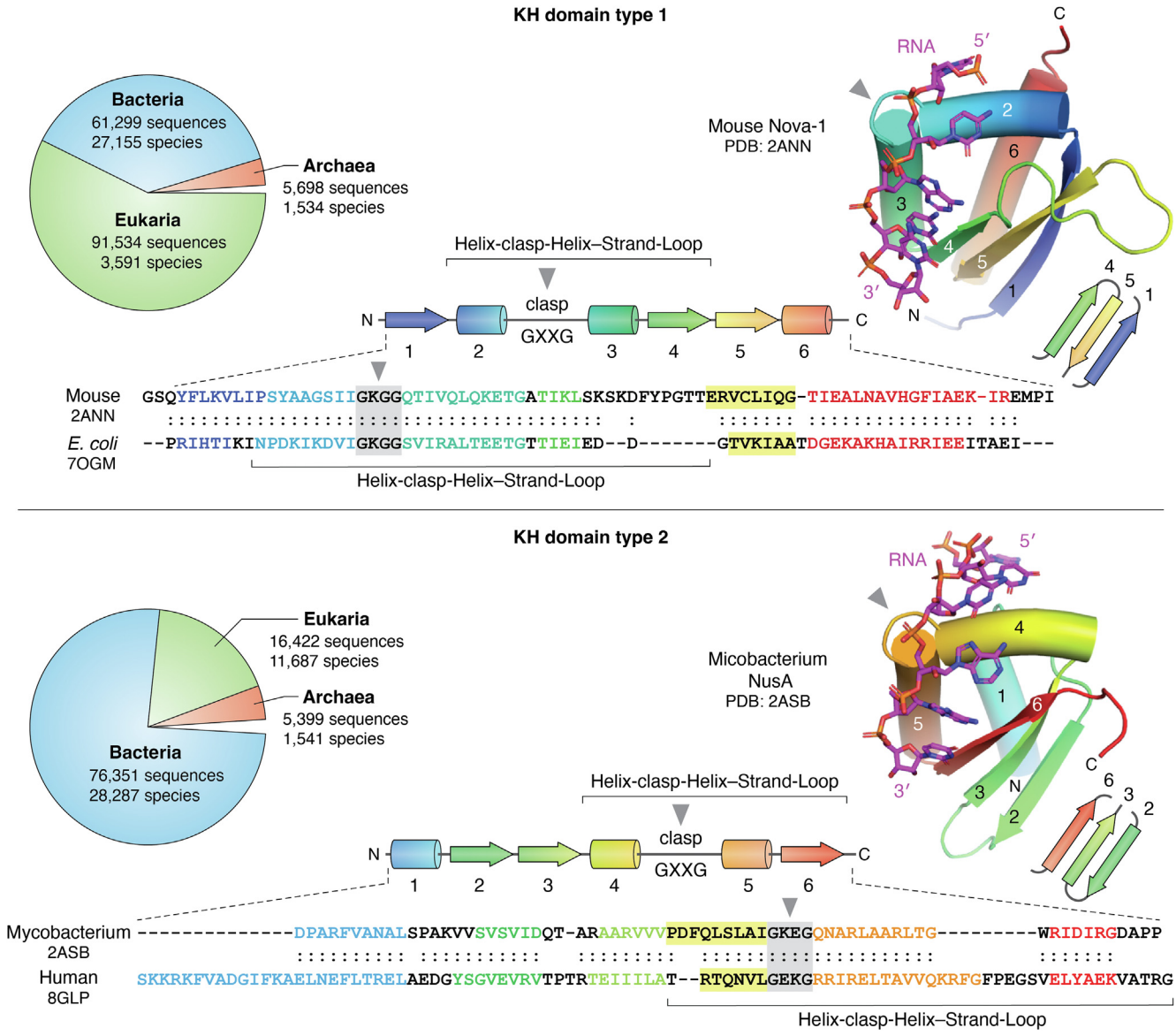

Figure 1. Type 1 and 2. KH domains are conserved in all three domains of life. *Left*: interpro-derived chart of KH domain evolutionary tree, divided based on the number of sequences. *Middle*: secondary structure of Type 1 and 2 domains, as ordered in the primary sequence with HcH–SL motif denoted. *Right*: representative Type 1 and 2 domains. Rainbow coloring within the KH domain, from the N terminus to the C terminus, highlights that overlaid secondary structure elements, are ordered differently in the primary sequence, comparing Type 1 and 2. Bottom of each panel. Structural alignment of eukaryotic and bacterial PDB structures by US-align program (146, 147). ":" represents residue pairs <5 Å. RMSDs were 1.4 and 2.4 Å respectively for depicted Type 1 and 2 structure alignment. GXXG clasp sequence is marked with an *arrow*. GXXG, Gly-X-X-Gly; K homology; HcH–SL, Helix-clasp-Helix–Strand-Loop; PDB, Protein Data Bank.

cardiovascular disease (5, 13, 14). Mutation-caused loss-of-function of FMR1 (2 KH domains) is associated with Fragile X syndrome, one of the most common genetic diseases associated with intellectual disability. A patient was identified with a rare single-base mutation T1100A, leading to a single-site mutation, Ile367Asn within the KH domain (15); this mutation did not detectably reduce the protein levels, associating a KH domain functional mutation directly to Fragile X syndrome. Aberrant splicing due to loss of sequence-specific RNA binding by a Nova-1 KH domain is thought to be related to a human paraneoplastic motor disorder called paraneoplastic opsoclonus-myoclonus ataxia (16). Notably, the KH domain's ability to bind RNA specifically and nonspecifically places it at critical points of RNA processing and regulation, making its dysfunction lead to a wide range of serious human diseases.

There are several excellent KH domain reviews on their tertiary folds, quaternary architecture, sequence specificity, and biological functions (3, 5, 6, 13, 17, 18). In contrast to existing reviews, this treatise systematically examines experimental interface structures of KH domain complexes with ssRNA or ssDNA. Our findings support and extend the key Gly-X-X-Gly (GXXG) signature sequence motif to include neighboring secondary structural elements that we observe to form a unique structural NA recognition motif. The Helix-clasp-Helix–Strand-Loop (HcH–SL) structural

motif defined here identifies a conserved triangular structural interface for binding and sculpting 4 to 5 nucleotides (nts) of ssNA for base recognition. Comparison to other helical and beta strand RNA recognition motifs (RRMs) that bind NA highlights the unique structure, binding mechanisms, and key functional implications of the HcH–SL structural motif that inform how its functions in cells remain incompletely known although its dysfunction is linked to multiple disease.

Conservation of KH domains

There are two types of KH domain. They have similar orientation of their secondary structural elements but differ in their topology leading to two distinct overall folds (19) (Fig. 1). The wide-ranging prevalence of both types in sequence databases suggests a primordial protein domain of foundational functional importance. Previous descriptions have reported that the Type 1 KH domain is primarily found in eukaryotes and the Type 2 KH domain in prokaryotes. Interpro analysis of the KH motif supports a more complex scenario (Fig. 1 pie charts) (2). In the InterPro data, both KH domain types occur frequently in prokaryotes and eukaryotes. Considering the number of species predicted to contain at least 1 KH domain, Type 2 appears more frequently than Type 1 for all domains. In terms of the number of sequences that contain a KH domain, there are more proteins in eukaryotes predicted to contain a Type 1. Nevertheless, analyses of current sequence databases do not support the notion that one type is more predominant than the other in one domain of life *versus* another.

The extraordinary frequency of Type 1 and 2 domains throughout the tree of life does support KH domains as having a primordial purpose that has remained functionally important throughout evolution. Yet, half of complete proteomes (12,171) are predicted to contain Type 1 (11,944) and/or Type 2 (11,982) out of 23,000 proteomes currently in the database (2), suggesting that either they may not be essential or are not yet identified in the proteomes missing a KH domain. Where KH domains have been recognized, more often than not, the proteome contains both types; it is rare for a proteome to contain only one type. Most proteomes contain 2 or less of Type 1 (∼80%) and/or Type 2 (∼95%). Interestingly, the human proteome is predicted to contain 143 Type 1 and 9 Type 2, while *Escherichia coli* has one Type 1 and two Type 2, consistent with higher eukaryotes having more complex regulation of their RNA. We note the absence of a detectable KH domain in viruses, surprising given the prevalence in the major domains of life.

Overall structure of the KH domain

The originally identified 45-amino-acid KH domain sequence has been expanded to 70 amino acids as the intact KH domain (1, 20). The KH domain is however a non-canonical domain in that it can be formed from two different folds (3, 18, 19). Although the three-dimensional secondary structures are strikingly similar in overall appearance, Type 1 and 2 folds differ in their topology and in the order of the secondary structure in the primary sequence. At the time of this report, we identified over 40 unique structures publicly available in the Protein Data Bank (PDB) with one or more KH domains bound to ssRNA or ssDNA, listed in Table 1 (16, 21–45). We use two of these structures, eukaryotic Nova-1

Table 1
KH domain structures in complex with nucleic acid (NA) included in this review noting the protein, KH domain type, function, and DOI

PDB	NA	Protein	KH type	Function	DOI
3R9W	RNA	aq_1994, era, era1	2	rRNA processing & biogenesis	10.1073/pnas.1017679108
6TQ0	RNA	ARC, KIAA0278	2	mRNA transport,	10.1016/j.bbrep.2021.100975
3AEV	RNA	Dim2p	1	RNA processing	10.1016/j.jmb.2010.03.055
1J4W	DNA	FBP	1	transcription	10.1038/4151051a
7YEY	RNA	GF2BP3	1	RNA transport, m6A binding	To be published
2MJH 4JVY	RNA	gld-1, T23G11.3	1	translation	10.1093/nar/gku445 10.1101/gad.216531.113
7OGM	RNA	hfq, Ecok1_42240, APECO1_2219	1	RNA transport, regulation	10.1016/j.molcel.2021.05.032
1J5K 1ZZI 1ZZJ 7CRE	DNA	hnRNP K	1	transcription	10.1093/emboj/cdf352 10.1016/j.str.2005.04.008
2N8M 8COO	RNA	IGF2BP1, VICKZ1, ZBP1	1	RNA transport, m6A binding	To be published 10.1093/nar/gkad534
7VKL	RNA	Igf2bp3, Vickz3	1	mRNA transport,	To be published
5ELS 5ELT 5EMO 5ELR	RNA	KHDRBS3, SALP, SLM2	1	splicing	10.1038/ncomms10355
4B8T	RNA	KSRP	1	decay, splicing, localization	10.1038/NSMB.2427
5WWW 5WWX	RNA	MEX3C, RKHD2, RNF194, BM-013	1	degradation	10.1074/jbc.M117.797746
4WAL 4WAN	RNA	MSL5, BBP, SF1, YLR116W, L2949	1	splicing	10.1261/rna.048942.114
2ANN 2ANR	RNA	Nova1	1	Splicing	10.1016/j.str.2011.05.002
1EC6	RNA	Nova2	1	splicing	10.1016/S0092-8674 (00)80668-6
2ASB 2ATW 5LM7 6GOV	RNA	NusA	2	transcription and termination	10.1038/sj.emboj.7600829 10.1038/nmicrobiol.2017.62 10.1016/j.molcel.2019.01.016
1ZTG	DNA	Poly(C)-binding protein-1, PCBP1	1	splicing	10.1093/nar/gks058
2PY9 2AXY 2P2R 2PQU	RNA/DNA	Poly(C)-binding protein-2, PCBP2	1	multifunctional , splicing, oxidative damage recognition	10.1261/rna.410107 10.1074/jbc.M508183200 10.1093/nar/gkm139
4AM3	RNA	PNPase	1	degradation	10.1098/RSOB.120028
4JVH	RNA	QKI, HKQ	1	stability, export, and splicing	10.1101/gad.216531.113
1K1G	RNA	Splicing factor 1	1	splicing	10.1126/science.1064719

All structural coordinates and their data are available in the Protein Data Bank under the listed PDB codes (https://www.rcsb.org/).

and bacterial NusA, to illustrate particular features of Type 1 and 2 KH domains, respectively (Fig. 1).

In common, both Type 1 and 2 KH domains have a three-stranded beta sheet with three alpha helices on one side of the sheet (Fig. 1). Most KH domains have a distinguishing signature GXXG sequence, where X's are typically basic or polar residues (46). The X's act to pin the phosphodiester backbone of an interacting NA (47). This signature sequence joins two sequential alpha helices that form a helical corner with an approximate 90° orientation to each other where GXXG acts as a structural "clasp" holding the helices to form a tight corner. These two helices and a sequential beta strand and loop extension form the primary NA binding motif, which we term the Helix-clasp-Helix–Strand-Loop (HcH–SL) motif, to contrast it with other recognized NA binding motifs: Helix-Turn-Helix (HTH), Helix-hairpin-Helix (HhH), and beta strand RRMs. Notably, the KH sequence forms a compact integral folded domain important to maintain the structure of the HcH–SL recognition motif (Fig. 1). Motif helix 1 is buttressed by either a C-terminal helix (Type 1) or an N-terminal helix (Type 2). The Strand-Loop is stabilized and positioned by an antiparallel (Type 1) or parallel (Type 2) connection to the three-stranded beta sheet (Fig. 1).

Differing in exact fold, Type 1 has the HcH–SL as an internal part in the sequence (secondary structure elements 2, 3, and 4), while Type 2 has it on the C terminal side (secondary structure elements 4, 5, and 6) (Fig. 1). Based on the number in the sequence, the Type 1 beta sheet places strands in order 1, 5, and 4, with the edge of strand 4 in the NA binding region. Type 2 beta sheet places strands in order 2, 3, and 6, with the edge of strand 6 in the NA binding region. The change in strand order from 1, 5, and 4 to 2, 3, and 6 reflects differences in KH domain fold. Having implications on the flexibility of the beta sheet (48), this fold change leads to a difference in being fully antiparallel (Type 1) to half parallel/half antiparallel (Type 2). Notably, we have redefined the structural motif. Previously, strand-helix-helix-strand (SHHS or βααβ) was defined as a minimal motif (18, 19). Yet, the first strand position switches its relative position from 1 on the edge of the sheet to 3 in the middle of the sheet and does not directly interact with NA. Because of inconsistency with its functional role, we did not include this strand in the recognition motif, but do include a loop extension from the second strand; this loop sometimes interacts with the NA bases. Thus, we redefined our previously defined motif from Helix-clasp-Helix to HcH–SL as being both the structurally and functionally conserved NA recognition motif within the KH domain (Fig. 1).

The continuous HcH–SL peptide motif defined here (28–41 residues long) is arguably a potential primordial co-factor facilitating RNA processes (Fig. 1 topology schematics). It is a motif common to distinct KH domain folds without a simple genetic relationship. In the RNA-only world preceding today's DNA world, this motif could have acted in RNA-mediated replication and translation. The poor sequence homology when comparing structural orthologs from eukaryotic and bacterial KH domains is consistent with an ancient function encoded in the motif structure (Fig. 1 sequences). This notion suggests that the HcH–SL motif, but not either of the current KH domains, is worth consideration as a possible polypeptide component of the last universal common ancestor (LUCA) cell for bacteria, archaea, and eukarya domains of life from ∼4,2 billion years ago, which remains an area of interest and controversy (49).

The HcH–SL motif forms a right triangle that is right handed

Impressively conserved between Type 1 and 2 KH domains, the HcH–SL motif conceptually forms a right triangle (Fig. 2). This triangle can usefully be considered in two parts: the HcH forms the orthogonally related sides of the triangle and the SL forming the hypotenuse (Fig. 2*A*). The angle between the two helices is ∼90°, so they form a tight helical corner that is enabled by the glycine main-chain hydrogen-bonds and protruding G-X-X-G clasp (Fig. 2*B*). Going from N terminus to C terminus in three-dimensional space, the structural elements go in a shallow right-handed spiral, with the thumb pointing in the direction of the bound NA. Importantly, these HcH–SL secondary structure elements are the primary interface for the NA interactions to KH domain proteins that include many main chain and side chain protein contacts (10–18 residues in our collective dataset) (Fig. 2*C*). For most KH domains, 4 to 5 nts of the ssNA spans the HcH–SL motif, with bases facing the same relative direction. We numbered the nucleotides from 5′ to 3′. Unexpectedly in some PDBs, 2 nts can occupy the position of nt 2, and we define them as 2A and 2B (Fig. 2*D*). There are examples of 1 to 2 nts at position 2 for both Type 1 and 2 KH domains, and this can differ even between 2 KH domains occurring in the same protein.

The ssNA phosphate backbone spans the triangle along the N-terminal end of helix-2 to the G-X-X-G clasp and C-terminal end of helix-1. Except for the 3′ base, the NA bases point into the triangle toward the Strand-Loop which creates a shallow pocket for the first two unstacked bases by closing the triangle across from the helical corner (Fig. 2, *A* and *D*). The heart of HcH–SL recognition centers on the 3 to 4 unstacked bases that essentially wrap around helix-1. Importantly, this unstacking allows the base chemistry of both ring and heteroatoms to be recognized by the HcH–SL structural elements. The mechanism for base recognition appears suitable for Strand-Loop-induced fit to the NA that would be important for increased affinities of KH domains for alkylation and oxidation base modifications (38, 50–56), as observed for a DNA repair enzyme for alkylated DNA (57). In contrast, the triangle helical corner acts as an anchor point that does not show marked changes in our comparisons of existing NA bound and free KH domain structures.

The GXXG sequence forms a key clasp between the two helices in the HcH–SL motif (Fig. 2*B*). The first Gly main chain carbonyl forms a hydrogen bond with the second Gly main chain amide. Although the "X-X" residues in-between the Gly residues are not strongly conserved, they interact with one phosphate oxygen through a main chain amide in the second X and often through their side chains. Inversely

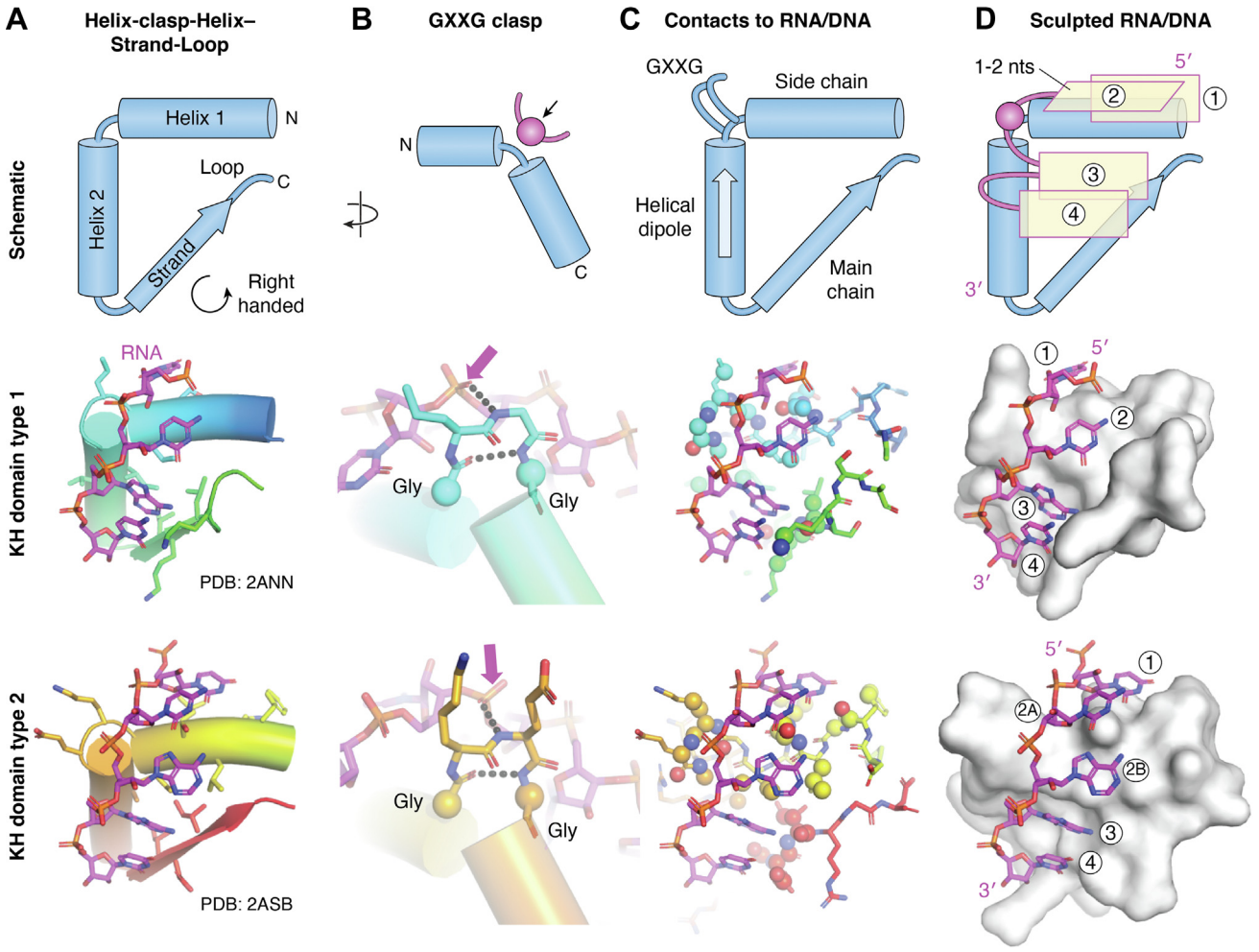

Figure 2. The HcH–SL motif is the primary region in contact with the RNA/DNA. Components are colored as in Figure 1. *A*, the secondary structure forms a *right*-handed triangle and concave surface going from the N terminus to the C terminus of the motif. The two helices are at approximately *right* angles to each other. Side chains within 4 Å of the NA are shown in *stick*. *B*, the GXXG motif resembles a clasp bringing the two helices together through main chain hydrogen bond between glycine (indicated by *sphere*). A main chain amino contacts the phosphate oxygen, at the positive dipole of the second helix. *C*, atoms within 4 Å of the RNA (*spheres*) show that the primary contacts in the first helix are through side chains, while the GXXG clasp and strand contact through main chain. *D*, forming a shallow concave NA recognition surface, the HcH–SL motif (*gray* molecular surface) sculpts the RNA, with the first helix unstacking 1 or 2 bases in respectively Type 1 and 2. Panels *A*, *C*, and *D* have the identical perspective. GXXG, Gly-X-X-Gly; HcH–SL, Helix-clasp-Helix–Strand-Loop; NA, nucleic acid.

suggesting the GXXG is required for NA interaction, domains that have KH folds but without the GXXG motif have not yet been shown to bind NA (58). As an exemplary case for significance, the single-site disease mutation for Fragile X syndrome, Ile267Asn (15) maps to the HcH–SL motif and, more specifically, to the N-terminal end of the second helix. We postulate that besides the hydrophobic to polar change noted in the structural study (47), the mutation to a helix-capping residue may mask the helical dipole. Indeed, Asn was the top preferred residue for N-capping in a helical stabilization study (59).

HcH–SL motif NA contacts

Our analyses of protein to NA contacts reveal an intriguing bias for the type of protein contacts (Fig. 2*C*). The NA on the 5′ end stretches across helix-1 of the HcH–SL motif. This first helix contacts the NA through mostly side chains, with one curiously conserved helical main chain packing against nt 2. Access to the main chain is enabled by a Gly or Ala 3 to 4 residues N-terminal to the GXXG motif. The one phosphate targeted by the main and side chains of the GXXG enabled clasp discussed above is also aligned dead center to the N-terminal helical dipole of Helix-2. We often observe N-terminal helical dipoles near the protein-NA interface (60–65). There are otherwise surprisingly few interactions between the second helix and the NA. Instead, the beta strand in the motif interacts with the 3′ end (nts 3 and 4), primarily through main chain contacts. Although it is common to think that base specificity is through side chains, the beta strand positions main chain contacts to confer base specificity. Conserved in both Type 1 and 2 KH domains, this bias for helical side chain and beta strand main chain contacts to nts is consistent with their relationship being through different evolutionary elaborations of a primordial HcH–SL peptide cofactor from the RNA world.

The HcH–SL motif sculpts ssRNA or DNA. In our structural study of the single KH domain in alkylation response

protein ASCC1, we compared structurally similar Type 1 KH domains and found a surprising splaying of the NA such that the bases were unstacked and spread out like three sides of a box (Fig. 2*D*) (66). A fourth base was stacked on one side of the box. In this more extensive analysis of Type 1 and 2 structures with bound NA, we find generally that KH domains bind 4 to 5 nts, but the bound nts can range from 11 nts (PDB: 2pqu (47)) to only 2 nts as one component of a larger binding interface (PDB: 5LM7) (Fig. 3, *A–D*). In those structures binding $>$ 4 nts, we find that the sculpted NA followed a general trend but that there could be not one but two bases on the top of the first helix, as included on our current schematic (Fig. 2*D*).

In common, both types of KH domains bind the NA in such a way as to prevent stacking or base pairing, enforcing binding specificity for ssRNA or ssDNA (Figs. 2*D* and 3, *A–C*). Notably, the interaction enables specific base recognition by main chains and side chains of residues in the HcH–SL motif. Overlaying the helical corner, the detailed position of the ssNA phosphodiester and bases can be shifted 3 Å and rotated relative to each other, not unexpected given the different sequence specificity and multiple functions in RNA processes (Fig. 3*E*).

In our finding that ASCC1 bound to a specific RNA sequence and not DNA of the same sequence (66), we postulated that KH domains may bind RNA more tightly than DNA. In the structures that we examined for this treatise, we saw that the exact position of bound NA varied, relative to the helical corner (Fig. 3, *A–C*). However, most RNA and DNA nts overlaid well onto each other, suggesting no or limited

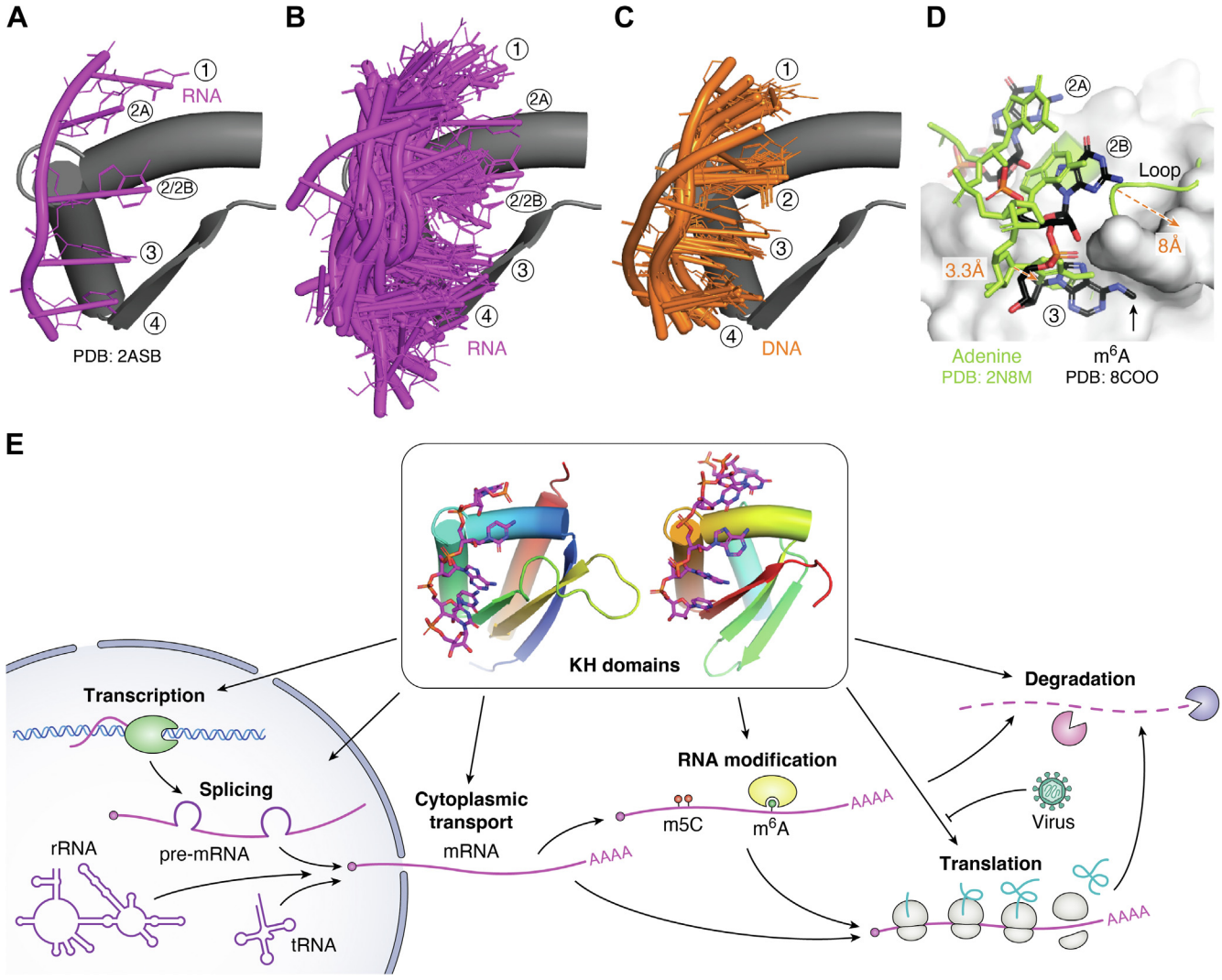

Figure 3. For a range of RNA processes, NAs are bound in similar but not identical positions by HcH–SL in different KH domains. *A–C*, overlay of KH domain HcH—SL (cartoon) reveals relative NA positions (*cartoon* and *stick*, labeled as in Fig. 2*D*). Structures overlaid to PDB:2ASB based on the helical corners. *A*, reference image of single RNA and HcH–SL from PDB:2ASB. *B*, overlay of 28 RNA structures onto 2ASB HcH–SL. RNA from PDB: 2ASB, 1EC6, 1K1G, 2ANN, 2ANR,2ATW, 2MJH, 2N8M, 2PY9M 3AEV, 3R9W, 4AM3, 4B8T, 4JVH, 4JVY, 4WAL, 4WAN, 5ELS, 5ELT, 5EMO, 5LM7, 5WWW, 5WWX, 6GOV, 6TQ0, 7OGM, 7VKL, and 8COO. *C*, overlay of 10 DNA structures onto PDB: 2ASB HcH-SL. DNA FROM PDB: 1J4W, 1J5K, 1ZTG, 1ZZI, 1ZZJ, 2AXY, 2PQU, 2P2R, 3R9W, and 7CRE. *D*, IF2BP1/IMP1 positions RNA (sticks) with an m6A at position 3 more deeply in the HcH–SL pocket (*gray* surface), compared to the same RNA with an undamaged adenine (cartoon). Only nts positions 2A, 2B, and 3 are shown for clarity. *E*, with implications for human and infectious disease, KH domain HcH–SL motifs act in many RNA processes: spanning transcription, regulation, transport, translation, and degradation and including both regulating normal processes and viral infections.

differences between how KH domains bind RNA *versus* DNA. As some KH domains are RNA-specific, such as we found in ASCC1 (66), it is likely that those KH domains have RNA-specific, 2′ hydroxyl mediated interactions. Potentially a primordial ssRNA binding role evolved to encompass new ssDNA binding functions.

Helix-clasp-Helix–Strand-Loop (HcH–SL) is a unique NA binding motif

Helices are well-known to act in DNA interactions. In the HcH–SL motif, both helices act directly in NA binding: the two helices form a nonoverlapping corner for a triangle with the helices almost orthogonal to each other. The NA extends across the N-terminal end of helix-2, the G-X-X-G clasp, and is stretched over the C-terminal end of helix-1 (Fig. 4*A*). The beta Strand-Loop hypotenuse closes the right triangle with the two-helix corner. Consequently, the HcH–SL motif forms a continuous concave interaction surface that provides interactions with NA backbone plus base sequence and chemistry (Fig. 2, *C* and *D*).

Two well-known and important NA two-helix motifs, HTH and HhH (67–71) are commonly found at interfaces with DNA, but interact with NA distinctly from the HcH–SL motif (Fig. 4, *A*–*C*). These motifs primarily interact with one of the two helices. In the HTH motif, the helices are connected by 1 to 2 turns. The helices are orthogonal but are overlapping (Fig. 4*B*). The C-terminal helix drops into the major or minor groove, distorting the double-stranded DNA helix but not breaking the stacking. In the HhH motif, the helices, connected by a hairpin turn, are overlapping with their connection at a 30° angle (69). HhH helices interact with the phosphate backbone, also not disrupting the stacking (Fig. 4*C*). Notably, one helix has an N-terminal helical dipole interaction, similar to the second helix of the HcH–SL motif. The HTH motif binding in the DNA major groove can be sequence specific, but it becomes structure-specific when binding in the DNA minor grove for DNA repair (72, 73). So the HTH in the DNA minor groove and HhH motif are generally structure-specific and not sequence-specific.

Another major RNA binding motif is the RRM (74, 75). The RRM typically binds unstacked bases near the middle of the recognized sequence by interactions with conserved aromatic residues on the central beta-sheet (Fig. 4*D*). Markedly unlike HcH–SL, RRM binds through its beta-strands (76). Both HcH–SL and RRM motifs are consequently structurally distinct binding motifs that share similarity in sequence-specific interactions as well as unstacking and splaying out of bound bases that enables recognition of diverse NA sequences and chemical modifications. Suggesting an enigmatic functional orientation, the splayed placement of 4 nts are similar between RRM-bound and HcH–SL bound its as they are

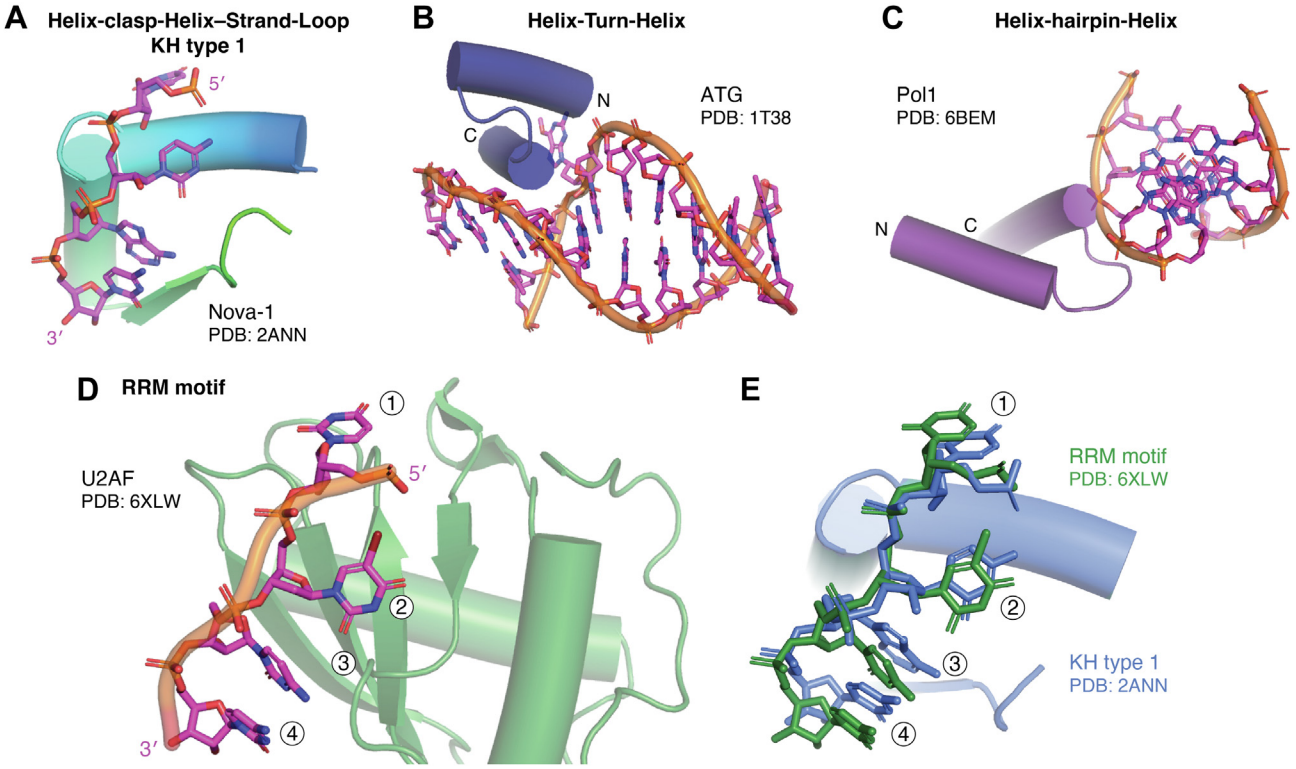

Figure 4. HcH–SL motif is distinct from other known helical and strand nucleic acid recognition motifs (nucleic acids shown as atom-colored sticks). *A*, HcH–SL motifs, as shown for type I (cartoon), sculpt the nucleic acids (stick) and enforce single strand specificity. *B*, helix-turn-helix motifs (*purple*, cartoon) recognize the DNA helical nature and insert one protein helix into the major or minor groove, distorting the DNA. The protein helices cross each other and are roughly orthogonal. *C*, helix-hairpin-helix motifs (*magenta*) bind to one strand of helical DNA. The protein helices cross each other at ∼30° angle. *D*, RNA recognition motifs employ beta strands for the major RNA-interaction interface. *E*, the first four nts bound by the RRM (*green*) are splayed in a strikingly similar orientation as the four nts bound by the HcH–SL (*blue* RNA with *blue* protein ribbons and helical cylinders). Panels (*D* and *E*) have the identical orientation. DNA shown as cartoon and stick in (*B*–*D*).

within the set of HcH–SL structures (Fig. 4*E*). In essence, the HcH–SL is a second type of RRM distinguished by its motif structure from the beta strand RRM but resulting in strikingly similar bound RNA conformation.

Emerging questions and research areas

From comparisons with other NA structural recognition motifs, the HcH–SL motif is unique both in its structure and NA interactions. Besides the well-recognized electrostatic GXXG interaction with the phosphodiester, the glycines act as a clasp to form a helical corner with the two helices (Fig. 2). Each of the other secondary structure elements provide productive NA recognition as part of the identified right-handed triangular recognition surface. Through its side chain interactions, the N-terminal helix acts as a steric wedge to splay out three-four bases. The C-terminal helix provides a basic dipole for positioning one of the phosphates. The beta strand and loop provide main chain interactions for recognition of specific base sequence and chemistry. In considering this unique structural recognition chemistry and 3 decades of research, key questions arise out of which six we note below as potential promising emerging research areas.

Why does the HcH–SL motif splay out its interacting NA?

The primordial HcH–SL motif splays out its interacting NA: this unusual open recognition arrangement certainly enables a base recognition mechanism (Fig. 2). However, it seems to be overkill for this, as the sequence can be read in double-stranded NA. Even promoting processing or protection of ssNA does not require unstacking of the bases, as can be seen in the replication protein A (RPA) and EXO5 structures with ssDNA (77–80). KH domains lack a catalytic activity. So they do not need to distort the NA or position its bonds for catalytic activity, as we have observed in many DNA repair glycosylases and nucleases (57, 60, 64, 65, 72, 73, 81–88). However, as both the phosphodiester and the bases are exposed, the KH domains could be positioning or stabilizing the substrate for partner proteins to act. We postulated as such for ASCC1, which contains an RNA phosphodiesterase domain linked to its KH domain. Certainly, the prevalence of KH domains in translation and splicing machinery would be consistent with this notion.

From a chemistry perspective, exposing the base could provide specific KH domains with a generalized mechanism for recognition of modified bases. Alkylation/methylation of bases transforms them from negative pi-stacked rings to ones that are more positively charged (89). By exposing the entire face of the base, KH domain-containing proteins could detect a change in the size and charge of the base agnostic to exact position on the base that the methylation occurs. The KH domain would then more or less strongly bind a given NA sequence. KH domains could then execute their functions, based on posttranslational modifications, such as methyl-6-adenine (m6A or 6mA), with implications for RNA regulation and cancer (90). Consistent with this hypothesis, there are KH domain-containing proteins that preferentially bind methylated RNAs. With 3 KH domains, Fragile X mental retardation protein (FMRP) binds m6A-modified and methyl-5-cytosine (m5c) RNA within particular sequence contexts impacting modulated m6A-modified mRNA levels and nuclear export (51–56). Interestingly, recognition of the trailing ssRNA overhung by FMRP's 3 KH domains may be related to its ability to bind R-loops (91), which are associated with transcription and also with base damage and modifications (92, 93). ASCC1 is associated with alkylation, as a member of an enigmatic alkylation response complex that includes the human direct alkylation reversal enzyme ALKHB3 (94–98). Its linked KH-phosphodiesterase domains may recognize alkylated bases (66, 99), although further work is needed to test this idea. Provocatively, mutant ASCC1 with shortened KH domain but full RNA ligase-like domain can lead to a severe phenotype of spinal muscular atrophy with congenital bone fractures 2 (SMABF2) (100).

One set of proteins provide exemplary interactions with methylated RNA through their KH domain. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BP1/2/3, IMP1/2/3) interact with and stabilize m6A-modified RNA in human cells through their KH domains, and this interaction is thought to act in oncogenesis (101). NMR studies revealed how the IGF2BP1/IMP1 KH domain directly recognizes methylation with an increased 5-fold affinity (38). In the structures, the m6A is bound in the deeply buried "third" position, with direct hydrophobic recognition of the methylation (Fig. 3*D*). The proposed recognition of the charge is not obvious, but the methylated base shifts ∼3 to 4 Å and is more deeply buried and has a distinct stacking with the fourth base, in comparison to a not yet published undamaged RNA with the same protein by the same authors (PDB: 2n8m). The last loop of the HcH–SL motif moves 8 Å, as can be seen by visibility of the respective loop in the nonmethylated-bound complex. Whether such deep binding requires 2 nts (2A/2B) at nt position 2 is untested, but the more loosely bound undamaged RNA structure also has 2 nts at position 2. Besides methylation, there may be multiple other modifications that alter the base chemistry, which could be detectable by how the NA is sculpted by the HcH–SL motif (102). Furthermore, as both HcH–SL and RRM motifs unstack and splay out bound bases into surprisingly similar conformations, it will be informative to comparatively examine their functional recognition of NA sequence and chemistry.

Which mechanistic roles are implied for the helix 2 N-terminal dipole and the HcH–SL mainchain in NA interfaces?

Helix 2 N-terminal dipole binds to one phosphate, typically to the nucleotide position at 2/2B (Fig. 2*C*). Unstacked on top of helix 1, this RNA base is the most rigidly exposed (Fig. 2*D*). Does the helix 2 dipole provide an anchoring interaction interface for inducing this conformation or strain? Stacking is surely the strongest noncovalent force contributing to the structure of ssNA, making unstacking an energetically costly evolutionary selection for the KH domain. It seems likely that

the helix dipole initiates NA binding through electrostatic guidance and then provides an anchor point for proximity-mediated interactions with specific base sequence and chemistry within the interface triangle.

These interactions may include potential inducible fit complementarity with the Strand-Loop elements, as seen in binding methylated RNA (101). In contrast to the HTH motif that can vary from one to two hairpin turns (helix-2turn-helix) in FEN1 superfamily enzymes (60, 103–105), we expect that the HcH–SL motif will lack clasp variants due to the clasp's critical and minimalistic role in positioning the helix corner and thus the helix 2 dipole. Although the helix dipole anchor is sequence independent, we furthermore expect that the motif will enforce sequence dependencies due to its other structural elements that directly bind individual bases. The HcH–SL motif therefore has both an anchoring phosphate backbone interaction like the HhH motif and sequence-dependent interactions like the RNA RRM motif and the HTH motif when binding to the DNA major groove. Further characterizations of HcH–SL NA complexes will test and clarify these current expectations.

To extend this helical dipole mechanism more generally to a role of main chain atoms, we also observed main chain atoms providing specific base recognition in KH domains or in positioning the RNA (Fig. 2*C*). In protein structure prediction, we typically think that side chains provide base specificity, but we saw multiple cases for main chain interaction with base-specific moieties in the Strand-Loop. We also saw main chain helix positioning of the RNA. In the Nova-2 KH domain structure, the main chain of Gly18 in the first helix of the HcH–SL motif positions the N1 of a uridine base at nt 1 (47). If main chain is providing base specificity, then noninterface residues may be positioning the interface main chain, potentially explaining the unexpected evolutionary covariance of noninterface residues (106). The observed role of main chain and helical dipoles in NA base recognition and positioning goes beyond KH domains and further analysis would test this as a generally important mechanism.

When might modular KH domains and HcH–SL functional dynamicity provide possible advantages?

Comparison of how NA is bound in the different structures of the HcH–SL motif (Fig. 3, *A*–*C*) highlights a motif that can not only be flexible in sequence specificity but in exact positioning of the NA. This is demonstrated most clearly in the IGF2BP1/IMP1/RNA complex structures shifting the RNA position ∼4 Å based only on a methylated base (Fig. 3*D*). This flexible positioning is enabled by an anchor point and a region allowing conformational changes in the HcH–SL motif (Fig. 2). The helical corner and dipole provide a stable anchor point. The Strand-Loop side can accommodate movement and consequently may allow a degree of induced complementarity—seen by an 8 Å shift of the last loop in the motif (Fig. 3*D*). These observations along with the other studies on HcH–SL interactions suggest that solution methods such as small angle X-ray scattering, NMR, and mass spectrometry that can interrogate dynamic complexes may be useful to examine HcH–SL conformations and interactions including their two tandem associations and those potentially involved in distinguishing binding to NA modifications (24, 28, 107).

Single KH domains appear tunable in their affinity for ssNA. The single HcH–SL motif in ASCC1 has specificity for RNA based on electrophoretic mobility shift analysis, but its broad bands shifting in position as a function of protein concentration implies transient interactions or dynamicity in binding (66). HnRNPK's KH3 domain HcH–SL interaction with ssDNA has a 1 μM Kd (25). Similarly, the KH domain of DDX43 helicase, which is key for the substrate specificity and unwinding processivity of the full-length helicase that is often overexpressed in cancers, has a binding affinity for ssRNA of 2.5 μM (4). So single KH domain HcH–SL binding to ssNA will likely be in the micromolar range suggesting transient interactions in cells. However, KH domain proteins like ASCC1 may have other modular interfaces that increase binding to specific substrates.

Furthermore, multiple KH domains can also increase binding affinity. For example, 2 KH domains in bacterial NusA increase ssRNA binding affinity to the nanomolar range (22). The number of modular binding sites and tandem HcH–SL motifs in a protein can therefore tune the transience of the RNA recognition as well as expand the number of bases being recognized. This multiplicative effect is similar to the four oligonucleotide-binding fold domains in replication protein A (RPA) that binds ssDNA individually with μM affinity but together bind in the pM range (108). This metastable recognition by a single HcH–SL motif makes it useful for functional dynamicity and dynamic regulatory RNA interactions. Implicated activities include biogenesis, stabilization, and processing of RNA (Fig. 3*E*). So computational studies to characterize sequence-dependent changes in dynamical behavior upon NA binding may be informative.

Functional dynamicity of the nonenzymatic KH domain is pertinent given the growing recognition in molecular cell biology that proteins function in the context of dynamic and modular molecular complexes where the regulation and activities of these assemblies is achieved by perturbations in their dynamics, as seen in transcription factor IIH (TFIIH)` results (109, 110). These complexes can furthermore act more like multifunctional molecular machines than linear pathways, as seen for alkylation damage responses, nucleotide excision repair, DNA break repair, and replication fork restart (80, 95, 105, 109, 111–118). Moreover, biological regulation typically involves rate-limiting steps that are rarely enzyme chemistry and often include conformational controls promoting assembly or disassembly (82, 110, 111, 115, 119). In this regard, the assembly and disassembly of NA complexes, mediated through modular NA binding by the nonenzymatic KH domain, has broad-based and critical activities in all domains of life (Fig. 1).

Where to look in cellular functions and how to test for the HcH–SL motif?

By unstacking bases and sculpting ssNA, KH domains go beyond the minimum needed for sequence recognition and span all processes involving RNA from transcription to degradation (Fig. 3*E*). This unusual conformational sculpting for the non-enzymatic HCH–SL motif supports its identified roles in regulation and as a potential scaffold for other enzymes. We therefore suggest that the HcH–SL regulatory motif should be considered wherever there is ssNA and a need to identify potential base modifications and/or sequence. For example, adenosine deaminases acting on RNA 1 (ADAR1) mediated A-to-I editing increases binding of KH domain-containing, RNA-binding, signal transduction-associated 1 (KHDRBS1) to thereby confer 5-fluorouracil and cisplatin (5FU+CDDP) resistance (120). Prime KH functional areas involve cell growth, maintenance, differentiation, signaling, DNA repair, gene expression and translation, and alternative splicing. For example. ribosomal protein S3 (RPS3) is a multifunctional KH domain protein that mediates selective gene regulation, DNA repair, apoptosis, and the innate immune response to bacterial infection (121, 122).

Besides their classical ssRNA interactions, KH domains are also implicated in interactions with non-B-DNA structures, which are associated with DNA instability and cancer (123–125). Notably, the three KH domains of hnRNP K unfold the c-MYC i-motif, a four-stranded DNA structure formed by two antiparallel duplexes held together by hemi-protonated cytosine-cytosine+ (C:C+) base pairs (126). KH domains in IGF2BPs, which act as m6A readers, hinder miRNA-induced RNA decay and improve expression of oncogenes (127). Ubiquitination at K213 in the first KH domain of IGF2BP3 by E3 ubiquitin ligase parkin removes its oncogenic function as an m6A reader, promoting inactivation of phosphoinositide 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) signaling pathways (128). Current KH domain disease connections include cancer, inflammation, developmental disorders, cardiovascular disease, and restricting viral replication. So, improved mechanistic understanding of KH domain roles may inform medicine and therapeutics.

Although similarly ancient iron-sulfur (Fe-S) domains show up in multiple NA proteins and participate in NA binding (110, 111, 129), we noted no examples of proteins with both a KH domain and an Fe-S. It may be that the Fe-S oxidative potential for ssNA promotes evolutionary pressure against this combination. Yet, KH domain ribosomal protein S3 surprisingly reduces the nucleotide excision repair deficiency of XPD cells through its interaction with Fe-S domain helicase XPD on UV-damaged DNA (130).

Based upon the HcH–SL motif, there are specific residues to mutate to test the functional role of a given KH domain, *e.g.* mutating the first Gly will impact the clasp and Helix-clasp-Helix geometry to abolish NA binding. In contrast, GxxG to GDDG mutation can impair nucleic acid binding without compromising the stability of the domain (66, 131–133).

Interestingly, our analyses support the notion that the HcH–SL is a quintessential conformational regulatory peptide element. Specifically, we find that the paradox of the ancient KH domain family formed from two distinct KH domain fold topologies is resolved by their separate evolution from the primordial structurally and functionally conserved structural HcH–SL NA recognition motif identified here. Despite considerable sequence variability, the bias for helical side chain and beta strand main chain contacts to nts is conserved across both Type 1 and 2 KH domains. These and other observations are consistent with the relationship of KH domains being through different evolutionary elaborations of a primordial core HcH–SL functional peptide cofactor from the RNA world. Thus, we propose that the HcH–SL structural motif, but not either of the current KH domains, is arguably a possible polypeptide component of LUCA suitable to act in the regulation, assembly, stabilization, and conformational dynamics of RNA machines.

Prokaryotic KH domain proteins are implicated in integral cellular activities spanning from cell division to protein translocation where their ability to balance dynamicity and stability in NA interactions is likely functionally important. In eukaryotes, mutation and aberrant expression of KH domain proteins is often linked to cancer and other diseases. Thus, identification and characterization of the HcH–SL recognition motif in the ancestral KH domain core may provide a new window into control of diverse dynamic functional biological networks and their regulation as fundamental to RNA biology. For example, identification of the HcH–SL motif in members of the ADP-ribosyl transferase superfamily (PARP) that catalyze ADP-ribosylation (a response to cellular stress, DNA damage, and viral infection) suggests KH domains merit further investigation for broader roles in DNA damage and stress responses (134). Interestingly, these data also suggest KH dimers and split KH domains, which are outside our focus as they lack experimental structures, but would be interesting to investigate with advanced structure prediction algorithms (135).

How good are the prospects for KH domain targeting, bioengineering, and design?

Our in-depth analysis of the HcH–SL NA recognition motif promises to benefit efforts to efficiently target specific KH domains and employ the motif in protein engineering and design. These motifs may represent unique drug targets that are, however, challenging due to their relatively shallow binding pocket. Perhaps the combination of the helix dipole anchor point and flexible Strand-Loop element will enable inhibitors by anchored plasticity approaches (119, 136, 137). Such chemical inhibitor tools would seem more practical than developing a protein mimic to bind specific HcH–SL motifs analogously to inhibitor protein mimicry to block repair of base damaged DNA (138, 139). Yet, SARS-CoV-2 NSP3 produces a unique peptide motif that binds directly to the two central KH domains of FMRPs. Furthermore, this interaction is disrupted by the I304N mutation found in a patient with fragile X syndrome suggesting that viruses do in fact

unexpectedly use protein mimicry of NA to hijack KH domain-containing FMRP proteins for efficient infection (140). Could viral infections have provided evolutionary pressure for defective HcH–SL motifs during human evolution? Testing how widespread viral hijacking of KH domains may be merits investigation.

Joining repeating structural motifs to design novel proteins can be a successful strategy (141) and designing one or a series of right-triangle NA binding motifs seems feasible with several interesting potential applications as RNA readers, biomarkers, and regulators. When the HcH–SL motif is found in multiple copies, we note that they are typically separated by only a few residues suggesting minimal needs for an extended spatial arrangement for effective ssNA recognition (35, 45). Such designed tandem constructs could, for example, identify RNA viral sequences and provide a novel regulatory component of RNA vaccines for stabilization or targeting as part of a chimeric protein. Impactful roles of KH-domain protein in pathogen resistance in plants are relatively unexplored but may offer opportunities for novel interventions (142). In archaea, N-terminal KH domains equip aCPSF1 with specific-binding capacity to function as the long-sought global transcription termination factor of archaea (143). In humans, metastasis, which is the main cause of cancer mortality, can be blocked by circular RNAs and by KH proteins that promote a positive feedback nanomedicine loop to suppress metastasis (144). We therefore expect that it will likely be useful to consider and perhaps control as well as design HcH–SL interactions in linking genotype, epigenetics, and metabolic biochemistry back to phenotype for cancer precision oncology (145).

What are the overall implications and future prospects?

The collective data presented suggest that, like Fe–S cluster motifs, the HcH–SL is a partner in the origin of life that may predate cells and the DNA world. Based on analyses of experimental structures, the HcH–SL motif is a unique primordial NA binding motif that enforces an unusual sculpting of NA conformation. Its NA interactions have general roles in recognizing RNA sequence and chemistry plus positioning or reshaping RNA structure as suitable to regulate many aspects of RNA metabolism. Its RNA binding is likely transient unless combined with other KH domains, motifs, or interfaces. As single HcH–SL motif sculpting can be energetically metastable (66), we propose that it probably often serves dynamic purposes beyond simple sequence recognition, such as to enable larger roles in the recognition of NA modifications, in forming dynamic scaffolds, or in lowering the energy barrier for handoffs to other proteins or RNA. Furthermore, the significance of the HcH–SL motif ability to recognize base modifications has largely been underappreciated. Moreover, we find that structurally disparate HcH–SL and RRM NA binding motifs, unexpectedly similarly, sculpt unusual conformations for four bound nts, suggesting their convergent evolution with functional implications. So comparative examination of their functional interactions that involve similar splayed nt conformations may advance unified insights on RNA metabolism.

Indeed, surprising connections of the KH domain and its functional recognition motif with all aspects of cell biology are increasingly being discovered including connections to DNA damage responses, translation, apoptosis, innate immunity, and hijacking for SARS-CoV-2 infectivity. As an ancient general purpose ssRNA sequence and chemistry recognition element, the HcH–SL motif offers large windows that are well worth looking into for cell biology. Examination of this motif also suggests that a useful strategy to search for potential primordial peptide motifs from LUCA may be to identify structured functional peptide motifs conserved among protein domains lacking straightforward evolutionary relationships as seen here for KH domains. Going forward, the many KH domain sequences and substantial structural database of their NA complexes promises not only an increased awareness and understanding but also the likelihood that increasing foundational mechanistic knowledge will enable the effective design and harnessing of KH domains for multiple aspects of both biology and medicine.

Acknowledgments—We thank Matthias Blum for help navigating the Interpro Database, and for discussions Naga Babu Chinnam on ASCC1, Nima Mosammaparast on RNA alkylation, Darin Jones on base chemistry, and Bohdan Schneider on RNA sugar puckers.

Author contributions—J. A. T. and S. E. T. investigation; J. A. T. and S. E. T. conceptualization; J. A. T. and S. E. T. funding acquisition; J. A. T. and S. E. T. writing–original draft; S. E. T. project administration.

Funding and additional information—This work was supported by National Cancer Institute grant NCI P01 CA092584 (J. A. T. and S. E. T.) and U.S. Department of Energy, Offices of Basic Energy Sciences and Advanced Scientific Computing Research through the Biopreparednesss Research Virtual Environment (BRaVE) under contract number DE-AC02-O5CH11231 and specifically to Taskforce 5 (DOE-BRAVET5). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: FMRP, Fragile X mental retardation protein; GXXG, Gly-X-X-Gly; HcH–SL, Helix-clasp-Helix–Strand-Loop; HhH, Helix-hairpin-Helix; HTH, Helix-Turn-Helix; hnRNPK, human heterogeneous nuclear ribonucleoprotein K; KH, K homology; LUCA, last universal common ancestor; m6A, methyl-6-adenine; NA, nucleic acid; PDB, Protein Data Bank; RRM, RNA recognition motif.

References

1. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21**, 1193–1198
2. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., *et al.* (2023) InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427

3. Nicastro, G., Taylor, I. A., and Ramos, A. (2015) KH-RNA interactions: back in the groove. *Curr. Opin. Struct. Biol.* **30**, 63–70
4. Yadav, M., Singh, R. S., Hogan, D., Vidhyasagar, V., Yang, S., Chung, I. Y. W., *et al.* (2021) The KH domain facilitates the substrate specificity and unwinding processivity of DDX43 helicase. *J. Biol. Chem.* **296**, 100085
5. Cheng, M. H., and Jansen, R. P. (2017) A jack of all trades: the RNA-binding protein vigilin. *Wiley Inter. Rev. RNA* **8**. https://doi.org/10.1002/wrna.1448
6. Hasan, M. K., and Jeannine Brady, L. (2024) Nucleic acid-binding KH domain proteins influence a spectrum of biological pathways including as part of membrane-localized complexes. *J. Struct. Biol. X* **10**, 100106
7. Fragiadaki, M., and Zeidler, M. P. (2018) Ankyrin repeat and single KH domain 1 (ANKHD1) drives renal cancer cell proliferation *via* binding to and altering a subset of miRNAs. *J. Biol. Chem.* **293**, 9570–9579
8. Wang, Z., Qiu, H., He, J., Liu, L., Xue, W., Fox, A., *et al.* (2020) The emerging roles of hnRNPK. *J. Cell Physiol.* **235**, 1995–2008
9. Mullenger, J. L., Zeidler, M. P., and Fragiadaki, M. (2023) Evaluating the molecular properties and function of ANKHD1, and its role in cancer. *Int. J. Mol. Sci.* **24**. https://doi.org/10.3390/ijms241612834
10. Gillentine, M. A., Wang, T., Hoekzema, K., Rosenfeld, J., Liu, P., Guo, H., *et al.* (2021) Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders. *Genome Med.* **13**, 63
11. Gallardo, M., Lee, H. J., Zhang, X., Bueso-Ramos, C., Pageon, L. R., McArthur, M., *et al.* (2015) hnRNP K is a Haploinsufficient tumor suppressor that regulates proliferation and differentiation programs in hematologic malignancies. *Cancer Cell* **28**, 486–499
12. Kruse, C., Willkomm, D. K., Grunweller, A., Vollbrandt, T., Sommer, S., Busch, S., *et al.* (2000) Export and transport of tRNA are coupled to a multi-protein complex. *Biochem. J.* **346**, 107–115
13. Mushtaq, A., Mir, U. S., and Altaf, M. (2023) Multifaceted functions of RNA-binding protein vigilin in gene silencing, genome stability, and autism-related disorders. *J. Biol. Chem.* **299**, 102988
14. Banday, S., Pandita, R. K., Mushtaq, A., Bacolla, A., Mir, U. S., Singh, D. K., *et al.* (2021) Autism-associated vigilin depletion impairs DNA damage repair. *Mol. Cell Biol.* **41**, e0008221
15. De Boulle, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., *et al.* (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* **3**, 31–35
16. Teplova, M., Malinina, L., Darnell, J. C., Song, J., Lu, M., Abagyan, R., *et al.* (2011) Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* **19**, 930–944
17. Adinolfi, S., Bagni, C., Castiglione Morelli, M. A., Fraternali, F., Musco, G., and Pastore, A. (1999) Novel RNA-binding motif: the KH module. *Biopolymers* **51**, 153–164
18. Valverde, R., Edwards, L., and Regan, L. (2008) Structure and function of KH domains. *FEBS J.* **275**, 2712–2726
19. Grishin, N. V. (2001) KH domain: one motif, two folds. *Nucleic Acids Res.* **29**, 638–643
20. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J., *et al.* (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85**, 237–245
21. Yang, L., Wang, C., Li, F., Zhang, J., Nayab, A., Wu, J., *et al.* (2017) The human RNA-binding protein and E3 ligase MEX-3C binds the MEX-3-recognition element (MRE) motif with high affinity. *J. Biol. Chem.* **292**, 16221–16234
22. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R., and Taylor, I. A. (2005) Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.* **24**, 3576–3587
23. Andersson, C. D., Forsgren, N., Akfur, C., Allgardsson, A., Berg, L., Engdahl, C., *et al.* (2013) Divergent structure-activity relationships of structurally similar acetylcholinesterase inhibitors. *J. Med. Chem.* **56**, 7615–7624
24. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M., and Cusack, S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**, 1055–1067
25. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21**, 3476–3485
26. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D., and Clore, G. M. (2002) Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**, 1051–1056
27. Dendooven, T., Sinha, D., Roeselova, A., Cameron, T. A., De Lay, N. R., Luisi, B. F., *et al.* (2021) A cooperative PNPase-Hfq-RNA carrier complex facilitates bacterial riboregulation. *Mol. Cell* **81**, 2901–2913.e2905
28. Du, Z., Lee, J. K., Fenn, S., Tjhen, R., Stroud, R. M., and James, T. L. (2007) X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA* **13**, 1043–1051
29. Fenn, S., Du, Z., Lee, J. K., Tjhen, R., Stroud, R. M., and James, T. L. (2007) Crystal structure of the third KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.6 A resolution. *Nucleic Acids Res.* **35**, 2651–2660
30. Feracci, M., Foot, J. N., Grellscheid, S. N., Danilenko, M., Stehle, R., Gonchar, O., *et al.* (2016) Structural basis of RNA recognition and dimerization by the STAR proteins T-STAR and Sam68. *Nat. Commun.* **7**, 10355
31. Hallin, E. I., Bramham, C. R., and Kursula, P. (2021) Structural properties and peptide ligand binding of the capsid homology domains of human Arc. *Biochem. Biophys. Rep.* **26**, 100975
32. Hardwick, S. W., Gubbey, T., Hug, I., Jenal, U., and Luisi, B. F. (2012) Crystal structure of Caulobacter crescentus polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. *Open Biol.* **2**, 120028
33. Jacewicz, A., Chico, L., Smith, P., Schwer, B., and Shuman, S. (2015) Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs. *RNA* **21**, 401–414
34. Jia, M. Z., Horita, S., Nagata, K., and Tanokura, M. (2010) An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. *J. Mol. Biol.* **398**, 774–785
35. Krupp, F., Said, N., Huang, Y. H., Loll, B., Burger, J., Mielke, T., *et al.* (2019) Structural basis for the action of an all-purpose transcription anti-termination factor. *Mol. Cell* **74**, 143–157.e145
36. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., *et al.* (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098–1102
37. Nadal, M., Anton, R., Dorca-Arevalo, J., Estebanez-Perpina, E., Tizzano, E. F., and Fuentes-Prior, P. (2023) Structure and function analysis of Sam68 and hnRNP A1 synergy in the exclusion of exon 7 from SMN2 transcripts. *Protein Sci.* **32**, e4553
38. Nicastro, G., Abis, G., Klein, P., Esteban-Serna, S., Gallagher, C., Chaves-Arquero, B., *et al.* (2023) Direct m6A recognition by IMP1 underlays an alternative model of target selection for non-canonical methyl-readers. *Nucleic Acids Res.* **51**, 8774–8786
39. Schneider, T., Hung, L. H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J., *et al.* (2019) Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat. Commun.* **10**, 2266
40. Sidiqi, M., Wilce, J. A., Vivian, J. P., Porter, C. J., Barker, A., Leedman, P. J., *et al.* (2005) Structure and RNA binding of the third KH domain of poly(C)-binding protein 1. *Nucleic Acids Res.* **33**, 1213–1221
41. Teplova, M., Hafner, M., Teplov, D., Essig, K., Tuschl, T., and Patel, D. J. (2013) Structure-function studies of STAR family Quaking proteins bound to their *in vivo* RNA target sites. *Genes Dev.* **27**, 928–940
42. Yao, J., Tu, Y., Shen, C., Zhou, Q., Xiao, H., Jia, D., *et al.* (2021) Nuclear import receptors and hnRNPK mediates nuclear import and stress granule localization of SIRLOIN. *Cell Mol. Life Sci.* **78**, 7617–7633
43. Yoga, Y. M., Traore, D. A., Sidiqi, M., Szeto, C., Pendini, N. R., Barker, A., *et al.* (2012) Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides. *Nucleic Acids Res.* **40**, 5101–5114

44. Zapf, C. W., Bloom, J. D., McBean, J. L., Dushin, R. G., Golas, J. M., Liu, H., *et al.* (2011) Discovery of a macrocyclic o-aminobenzamide Hsp90 inhibitor with heterocyclic tether that shows extended biomarker activity and *in vivo* efficacy in a mouse xenograft model. *Bioorg. Med. Chem. Lett.* **21**, 3627–3631
45. Said, N., Krupp, F., Anedchenko, E., Santos, K. F., Dybkov, O., Huang, Y. H., *et al.* (2017) Structural basis for lambdaN-dependent processive transcription antitermination. *Nat. Microbiol.* **2**, 17062
46. Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T. J., Nilges, M., *et al.* (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat. Struct. Biol.* **4**, 712–716
47. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., *et al.* (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332
48. Ho, B. K., and Curmi, P. M. (2002) Twist and shear in beta-sheets and beta-ribbons. *J. Mol. Biol.* **317**, 291–308
49. Moody, E. R. R., Alvarez-Carretero, S., Mahendrarajah, T. A., Clark, J. W., Betts, H. C., Dombrowski, N., *et al.* (2024) The nature of the last universal common ancestor and its impact on the early Earth system. *Nat. Ecol. Evol.* **8**, 1654–1666
50. Hayakawa, H., Kuwano, M., and Sekiguchi, M. (2001) Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochemistry* **40**, 9977–9982
51. Edens, B. M., Vissers, C., Su, J., Arumugam, S., Xu, Z., Shi, H., *et al.* (2019) FMRP modulates neural differentiation through m(6)A-dependent mRNA nuclear export. *Cell Rep.* **28**, 845–854.e845
52. Hsu, P. J., Shi, H., Zhu, A. C., Lu, Z., Miller, N., Edens, B. M., *et al.* (2019) The RNA-binding protein FMRP facilitates the nuclear export of N(6)-methyladenosine-containing mRNAs. *J. Biol. Chem.* **294**, 19889–19895
53. Zhang, F., Kang, Y., Wang, M., Li, Y., Xu, T., Yang, W., *et al.* (2018) Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. *Hum. Mol. Genet.* **27**, 3936–3950
54. Westmark, C. J., Maloney, B., Alisch, R. S., Sokol, D. K., and Lahiri, D. K. (2020) FMRP regulates the nuclear export of Adam9 and Psen1 mRNAs: secondary analysis of an N(6)-methyladenosine dataset. *Sci. Rep.* **10**, 10781
55. Edupuganti, R. R., Geiger, S., Lindeboom, R. G. H., Shi, H., Hsu, P. J., Lu, Z., *et al.* (2017) N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **24**, 870–878
56. Yang, H., Wang, Y., Xiang, Y., Yadav, T., Ouyang, J., Phoon, L., *et al.* (2022) FMRP promotes transcription-coupled homologous recombination *via* facilitating TET1-mediated m5C RNA modification demethylation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2116251119
57. Wilkinson, O. J., Latypov, V., Tubbs, J. L., Millington, C. L., Morita, R., Blackburn, H., *et al.* (2012) Alkyltransferase-like protein (Atl1) distinguishes alkylated guanines for DNA repair using cation-pi interactions. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18755–18760
58. Oddone, A., Lorentzen, E., Basquin, J., Gasch, A., Rybin, V., Conti, E., *et al.* (2007) Structural and biochemical characterization of the yeast exosome component Rrp40. *EMBO Rep.* **8**, 63–69
59. Doig, A. J., and Baldwin, R. L. (1995) N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci.* **4**, 1325–1336
60. Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., *et al.* (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **145**, 198–211
61. Tsutakawa, S. E., Shin, D. S., Mol, C. D., Izumi, T., Arvai, A. S., Mantha, A. K., *et al.* (2013) Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **288**, 8445–8455
62. Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., and Cramer, P. (2019) Structural basis of TFIIH activation for nucleotide excision repair. *Nat. Commun.* **10**, 2885
63. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214–5226
64. Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., *et al.* (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5083–5088
65. Dalhus, B., Nilsen, L., Korvald, H., Huffman, J., Forstrom, R. J., McMurray, C. T., *et al.* (2013) Sculpting of DNA at abasic sites by DNA glycosylase homolog mag2. *Structure* **21**, 154–166
66. Chinnam, N. B., Thapar, R., Arvai, A. S., Sarker, A. H., Soll, J. M., Paul, T., *et al.* (2024) ASCC1 structures and bioinformatics reveal a novel helix-clasp-helix RNA-binding motif linked to a two-histidine phosphodiesterase. *J. Biol. Chem.* **300**, 107368
67. Tubbs, J. L., Pegg, A. E., and Tainer, J. A. (2007) DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair (Amst)* **6**, 1100–1115
68. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108–4120
69. Batra, V. K., Oertell, K., Beard, W. A., Kashemirov, B. A., McKenna, C. E., Goodman, M. F., *et al.* (2018) Mapping functional substrate-enzyme interactions in the pol beta active site through chemical biology: structural responses to acidity modification of incoming dNTPs. *Biochemistry* **57**, 3934–3944
70. Brennan, R. G., and Matthews, B. W. (1989) The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903–1906
71. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res.* **24**, 2488–2497
72. Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., *et al.* (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **11**, 714–720
73. Tubbs, J. L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., *et al.* (2009) Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808–813
74. Query, C. C., Bentley, R. C., and Keene, J. D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**, 89–101
75. Clery, A., Blatter, M., and Allain, F. H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298
76. Maji, D., Glasser, E., Henderson, S., Galardi, J., Pulvino, M. J., Jenkins, J. L., *et al.* (2020) Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. *J. Biol. Chem.* **295**, 17148–17157
77. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181
78. Yates, L. A., Aramayo, R. J., Pokhrel, N., Caldwell, C. C., Kaplan, J. A., Perera, R. L., *et al.* (2018) A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. *Nat. Commun.* **9**, 5447
79. Fan, J., and Pavletich, N. P. (2012) Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.* **26**, 2337–2347
80. Hambarde, S., Tsai, C. L., Pandita, R. K., Bacolla, A., Maitra, A., Charaka, V., *et al.* (2021) EXO5-DNA structure and BLM interactions direct DNA resection critical for ATR-dependent replication restart. *Mol. Cell* **81**, 2989–3006.e2989
81. Tsutakawa, S. E., Lafrance-Vanasse, J., and Tainer, J. A. (2014) The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once. *DNA Repair (Amst)* **19**, 95–107
82. Tsutakawa, S. E., Thompson, M. J., Arvai, A. S., Neil, A. J., Shaw, S. J., Algasaier, S. I., *et al.* (2017) Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.* **8**, 15855
83. Shi, Y., Hellinga, H. W., and Beese, L. S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic

reactions of human exonuclease I. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 6010–6015
84. Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**, 451–456
85. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92
86. Latypov, V. F., Tubbs, J. L., Watson, A. J., Marriott, A. S., McGown, G., Thorncroft, M., *et al.* (2012) Atl1 regulates choice between global genome and transcription-coupled repair of O(6)-alkylguanines. *Mol. Cell* **47**, 50–60
87. Dalhus, B., Arvai, A. S., Rosnes, I., Olsen, O. E., Backe, P. H., Alseth, I., *et al.* (2009) Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **16**, 138–143
88. Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P., and Tainer, J. A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **98**, 397–408
89. Rutledge, L. R., Campbell-Verduyn, L. S., Hunter, K. C., and Wetmore, S. D. (2006) Characterization of nucleobase-amino acid stacking interactions utilized by a DNA repair enzyme. *J. Phys. Chem. B* **110**, 19652–19663
90. Barbieri, I., and Kouzarides, T. (2020) Role of RNA modifications in cancer. *Nat. Rev. Cancer* **20**, 303–322
91. Dettori, L. G., Torrejon, D., Chakraborty, A., Dutta, A., Mohamed, M., Papp, C., *et al.* (2021) A tale of loops and tails: the role of intrinsically disordered protein regions in R-loop recognition and phase separation. *Front. Mol. Biosci.* **8**, 691694
92. Jayakumar, S., Patel, M., Boulet, F., Aziz, H., Brooke, G. N., Tummala, H., *et al.* (2024) PSIP1/LEDGF reduces R-loops at transcription sites to maintain genome integrity. *Nat. Commun.* **15**, 361
93. Yang, H., Lachtara, E. M., Ran, X., Hopkins, J., Patel, P. S., Zhu, X., *et al.* (2023) The RNA m5C modification in R-loops as an off switch of Alt-NHEJ. *Nat. Commun.* **14**, 6114
94. Wollen, K. L., Hagen, L., Vagbo, C. B., Rabe, R., Iveland, T. S., Aas, P. A., *et al.* (2021) ALKBH3 partner ASCC3 mediates P-body formation and selective clearance of MMS-induced 1-methyladenosine and 3-methylcytosine from mRNA. *J. Transl. Med.* **19**, 287
95. Tsao, N., Brickner, J. R., Rodell, R., Ganguly, A., Wood, M., Oyeniran, C., *et al.* (2021) Aberrant RNA methylation triggers recruitment of an alkylation repair complex. *Mol. Cell* **81**, 4228–4242.e4228
96. Townley, B. A., Buerer, L., Tsao, N., Bacolla, A., Mansoori, F., Rusanov, T., *et al.* (2023) A functional link between lariat debranching enzyme and the intron-binding complex is defective in non-photosensitive trichothiodystrophy. *Mol. Cell* **83**, 2258–2275.e2211
97. Dango, S., Mosammaparast, N., Sowa, M. E., Xiong, L. J., Wu, F., Park, K., *et al.* (2011) DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. *Mol. Cell* **44**, 373–384
98. Sundheim, O., Vagbo, C. B., Bjoras, M., Sousa, M. M., Talstad, V., Aas, P. A., *et al.* (2006) Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* **25**, 3389–3397
99. Soll, J. M., Brickner, J. R., Mudge, M. C., and Mosammaparast, N. (2018) RNA ligase-like domain in activating signal cointegrator 1 complex subunit 1 (ASCC1) regulates ASCC complex function during alkylation damage. *J. Biol. Chem.* **293**, 13524–13533
100. Sharova, M., Guseva, D., Kurenkov, A., Novoselova, O., Murtazina, A., and Skoblov, M. (2022) Congenital myopathy as a new phenotype caused by two undescribed variants in ASCC1 gene. *Am. J. Med. Genet. A.* **188**, 3100–3105
101. Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., *et al.* (2018) Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20**, 285–295
102. Harcourt, E. M., Kietrys, A. M., and Kool, E. T. (2017) Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339–346
103. Finger, L. D., Atack, J. M., Tsutakawa, S., Classen, S., Tainer, J., Grasby, J., *et al.* (2012) The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell. Biochem.* **62**, 301–326
104. Grasby, J. A., Finger, L. D., Tsutakawa, S. E., Atack, J. M., and Tainer, J. A. (2012) Unpairing and gating: sequence-independent substrate recognition by FEN superfamily nucleases. *Trends Biochem. Sci.* **37**, 74–84
105. Tsutakawa, S. E., Sarker, A. H., Ng, C., Arvai, A. S., Shin, D. S., Shih, B., *et al.* (2020) Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. *Proc. Natl. Acad. Sci. U S A.* **117**, 14127–14138
106. Fukasawa, Y., and Tomii, K. (2019) Accurate classification of biological and non-biological interfaces in protein crystal structures using subtle covariation signals. *Sci. Rep.* **9**, 12603
107. Yang, K., Chen, G., Yu, F., Fang, X., Zhang, J., Zhang, Z., *et al.* (2024) Molecular mechanism of specific HLA-A mRNA recognition by the RNA-binding-protein hMEX3B to promote tumor immune escape. *Commun. Biol.* **7**, 158
108. Sibenaller, Z. A., Sorensen, B. R., and Wold, M. S. (1998) The 32- and 14-kilodalton subunits of replication protein A are responsible for species-specific interactions with single-stranded DNA. *Biochemistry* **37**, 12496–12506
109. Yu, J., Yan, C., Dodd, T., Tsai, C. L., Tainer, J. A., Tsutakawa, S. E., *et al.* (2023) Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. *Nat. Commun.* **14**, 2758
110. Yu, J., Yan, C., Paul, T., Brewer, L., Tsutakawa, S. E., Tsai, C. L., *et al.* (2024) Molecular architecture and functional dynamics of the pre-incision complex in nucleotide excision repair. *Nat. Commun.* **15**, 8511
111. Bralic, A., Tehseen, M., Sobhy, M. A., Tsai, C. L., Alhudhali, L., Yi, G., *et al.* (2023) A scanning-to-incision switch in TFIIH-XPG induced by DNA damage licenses nucleotide excision repair. *Nucleic Acids Res.* **51**, 1019–1033
112. Ye, Z., Xu, S., Shi, Y., Bacolla, A., Syed, A., Moiani, D., *et al.* (2021) GRB2 enforces homology-directed repair initiation by MRE11. *Sci. Adv.* 7. https://doi.org/10.1126/sciadv.abe9254
113. Thapar, R., Wang, J. L., Hammel, M., Ye, R., Liang, K., Sun, C., *et al.* (2021) Mechanism of efficient double-strand break repair by a long non-coding RNA. *Nucleic Acids Res.* **49**, 1199–1200
114. Syed, A., and Tainer, J. A. (2018) The MRE11-RAD50-NBS1 complex conducts the orchestration of damage signaling and outcomes to stress in DNA replication and repair. *Annu. Rev. Biochem.* **87**, 263–294
115. Hammel, M., Rosenberg, D. J., Bierma, J., Hura, G. L., Thapar, R., Lees-Miller, S. P., *et al.* (2021) Visualizing functional dynamicity in the DNA-dependent protein kinase holoenzyme DNA-PK complex by integrating SAXS with cryo-EM. *Prog. Biophys. Mol. Biol.* **163**, 74–86
116. Longo, M. A., Roy, S., Chen, Y., Tomaszowski, K. H., Arvai, A. S., Pepper, J. T., *et al.* (2023) RAD51C-XRCC3 structure and cancer patient mutations define DNA replication roles. *Nat. Commun.* **14**, 4445
117. Ye, Z., Xu, S., Shi, Y., Cheng, X., Zhang, Y., Roy, S., *et al.* (2024) GRB2 stabilizes RAD51 at reversed replication forks suppressing genomic instability and innate immunity against cancer. *Nat. Commun.* **15**, 2132
118. Tsutakawa, S. E., Tsai, C. L., Yan, C., Bralic, A., Chazin, W. J., Hamdan, S. M., *et al.* (2020) Envisioning how the prototypic molecular machine TFIIH functions in transcription initiation and DNA repair. *DNA Repair (Amst)* **96**, 102972
119. Houl, J. H., Ye, Z., Brosey, C. A., Balapiti-Modarage, L. P. F., Namjoshi, S., Bacolla, A., *et al.* (2019) Selective small molecule PARG inhibitor causes replication fork stalling and cancer cell death. *Nat. Commun.* **10**, 5654
120. Wong, T. L., Loh, J. J., Lu, S., Yan, H. H. N., Siu, H. C., Xi, R., *et al.* (2023) ADAR1-mediated RNA editing of SCD1 drives drug resistance and self-renewal in gastric cancer. *Nat. Commun.* **14**, 2861
121. Gao, X., and Hardwidge, P. R. (2011) Ribosomal protein s3: a multi-functional target of attaching/effacing bacterial pathogens. *Front. Microbiol.* **2**, 137

122. Wan, F., Anderson, D. E., Barnitz, R. A., Snow, A., Bidere, N., Zheng, L., *et al.* (2007) Ribosomal protein S3: a KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. *Cell* **131**, 927–939
123. Bacolla, A., Tainer, J. A., Vasquez, K. M., and Cooper, D. N. (2016) Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res.* **44**, 5673–5688
124. Bacolla, A., Ye, Z., Ahmed, Z., and Tainer, J. A. (2019) Cancer mutational burden is shaped by G4 DNA, replication stress and mitochondrial dysfunction. *Prog. Biophys. Mol. Biol.* **147**, 47–61
125. De-Paula, R. B., Bacolla, A., Syed, A., and Tainer, J. A. (2024) Enriched G4 forming repeats in the human genome are associated with robust well-coordinated transcription and reduced cancer transcriptome variation. *J. Biol. Chem.* **300**, 107822
126. Wu, W. Q., Zhang, X., Bai, D., Shan, S. W., and Guo, L. J. (2022) Mechanistic insights into poly(C)-binding protein hnRNP K resolving i-motif DNA secondary structures. *J. Biol. Chem.* **298**, 102670
127. Korn, S. M., Ulshofer, C. J., Schneider, T., and Schlundt, A. (2021) Structures and target RNA preferences of the RNA-binding protein family of IGF2BPs: an overview. *Structure* **29**, 787–803
128. Sun, X., Ye, G., Li, J., Shou, H., Bai, G., and Zhang, J. (2023) Parkin regulates IGF2BP3 through ubiquitination in the tumourigenesis of cervical cancer. *Clin. Transl. Med.* **13**, e1457
129. Fuss, J. O., Tsai, C. L., Ishida, J. P., and Tainer, J. A. (2015) Emerging critical roles of Fe-S clusters in DNA replication and repair. *Biochim. Biophys. Acta* **1853**, 1253–1271
130. Park, Y. J., Kim, S. H., Kim, T. S., Lee, S. M., Cho, B. S., Seo, C. I., *et al.* (2021) Ribosomal protein S3 associates with the TFIIH complex and positively regulates nucleotide excision repair. *Cell Mol. Life Sci.* **78**, 3591–3606
131. Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., *et al.* (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886
132. Ni, X., Knapp, S., and Chaikuad, A. (2020) Comparative structural analyses and nucleotide-binding characterization of the four KH domains of FUBP1. *Sci. Rep.* **10**, 13459
133. Dowdle, M. E., Park, S., Blaser Imboden, S., Fox, C. A., Houston, D. W., and Sheets, M. D. (2019) A single KH domain in Bicaudal-C links mRNA binding and translational repression functions to maternal development. *Development* **146**, dev172486
134. Saleh, H., Liloglou, T., Rigden, D. J., Parsons, J. L., and Grundy, G. J. (2024) KH-Like domains in PARP9/DTX3L and PARP14 coordinate protein-protein interactions to promote cancer cell survival. *J. Mol. Biol.* **436**, 168434
135. Suskiewicz, M. J., Munnur, D., Stromland, O., Yang, J. C., Easton, L. E., Chatrin, C., *et al.* (2023) Updated protein domain annotation of the PARP protein family sheds new light on biological function. *Nucleic Acids Res.* **51**, 8217–8236
136. Garcin, E. D., Arvai, A. S., Rosenfeld, R. J., Kroeger, M. D., Crane, B. R., Andersson, G., *et al.* (2008) Anchored plasticity opens doors for selective inhibitor design in nitric oxide synthase. *Nat. Chem. Biol.* **4**, 700–707
137. Moiani, D., Link, T. M., Brosey, C. A., Katsonis, P., Lichtarge, O., Kim, Y., *et al.* (2021) An efficient chemical screening method for structure-based inhibitors to nucleic acid enzymes targeting the DNA repair-replication interface and SARS CoV-2. *Methods Enzymol.* **661**, 407–431
138. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., *et al.* (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701–708
139. Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W., *et al.* (1999) Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. *J. Mol. Biol.* **287**, 331–346
140. Garvanska, D. H., Alvarado, R. E., Mundt, F. O., Lindqvist, R., Duel, J. K., Coscia, F., *et al.* (2024) The NSP3 protein of SARS-CoV-2 binds fragile X mental retardation proteins to disrupt UBAP2L interactions. *EMBO Rep.* **25**, 902–926
141. Brunette, T. J., Parmeggiani, F., Huang, P. S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., *et al.* (2015) Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584
142. Fabian, M., Gao, M., Zhang, X. N., Shi, J., Vrydagh, L., Kim, S. H., *et al.* (2023) The flowering time regulator FLK controls pathogen defense in Arabidopsis thaliana. *Plant Physiol.* **191**, 2461–2474
143. Li, J., Yue, L., Li, Z., Zhang, W., Zhang, B., Zhao, F., *et al.* (2021) aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. *Elife* **10**. https://doi.org/10.7554/eLife.70464
144. Shu, G., Lu, X., Pan, Y., Cen, J., Huang, K., Zhou, M., *et al.* (2023) Exosomal circSPIRE1 mediates glycosylation of E-cadherin to suppress metastasis of renal cell carcinoma. *Oncogene* **42**, 1802–1820
145. Das, C., Bhattacharya, A., Adhikari, S., Mondal, A., Mondal, P., Adhikary, S., *et al.* (2024) A prismatic view of the epigenetic-metabolic regulatory axis in breast cancer therapy resistance. *Oncogene* **43**, 1727–1741
146. Zhang, C., and Pyle, A. M. (2022) A unified approach to sequential and non-sequential structure alignment of proteins, RNAs, and DNAs. *iScience* **25**, 105218
147. Zhang, C., Shine, M., Pyle, A. M., and Zhang, Y. (2022) US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115