Integrative structural analysis of NF45–NF90 heterodimers reveals architectural rearrangements and oligomerization on binding dsRNA

Sophie Winterbourne[1], Uma Jayachandran[1], Juan Zou[2], Juri Rappsilber [2,3], Sander Granneman[1,4], Atlanta G. Cook [1,*]

[1]Institute of Quantitative Biology, Biochemistry and Biotechnology, Max Born Crescent, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom
[2]Institute of Cell Biology, Max Born Crescent, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom
[3]Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany
[4]Centre for Engineering Biology, Max Born Crescent, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

[*]To whom correspondence should be addressed. Email: atlanta.cook@ed.ac.uk

Abstract

Complexes of nuclear factors 45 and 90 (NF45–NF90) play a multitude of roles in co- and post-transcriptional RNA processing, including regulating adenosine-to-inosine editing, cassette exon and back splicing, and splicing fidelity. NF45–NF90 complexes recognize double-stranded RNA (dsRNA) and, in human cells, primarily interact with Alu inverted repeats (AluIRs) that are commonly inserted into introns and other noncoding RNA regions. Intronic AluIRs of ∼300 bp can regulate splicing outcomes, such as generation of circular RNAs. We examined domain reorganization of NF45–NF90 domains on dsRNAs exceeding 50 bp to gain insight into its RNA recognition properties on longer dsRNAs. Using a combination of phylogenetic analysis, solution methods (including small angle X-ray scattering and quantitative cross-linking mass spectrometry), machine learning, and negative stain electron microscopy, we generated a model of NF45–NF90 complex formation on dsRNA. Our data reveal that different interactions of NF45–NF90 complexes allow these proteins to coat long stretches of dsRNA. This property of the NF45–NF90 complex has important implications for how long, nuclear dsRNAs are recognized in the nucleus and how this might promote (co)-regulation of specific RNA splicing and editing events that shape the mammalian transcriptome.

Graphical abstract
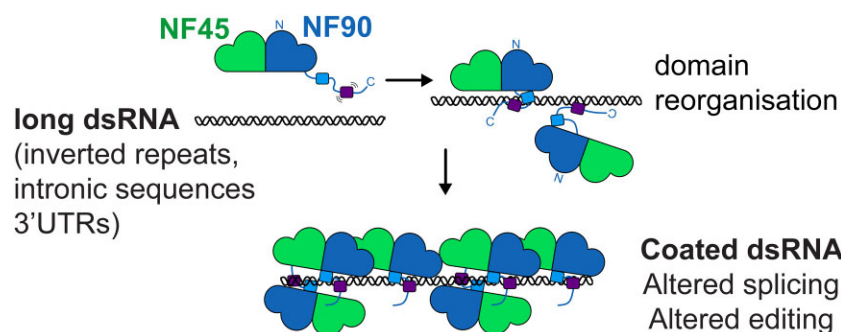
Introduction

Nuclear factors 45 and 90 (NF45 and NF90), also known as interleukin enhancer binding factors, ILF2 and ILF3, are essential proteins in mammalian cells that are involved in many post-transcriptional processes. This includes alternative splicing [1], circular RNA (circRNA) generation by back splicing [2], microRNA biogenesis [3–5], adenosine-to-inosine (A-to-I) editing [6, 7], ribosome biogenesis [8–10], RNA decay [11], and control of translation [12]. Despite the multitude of contributions of NF45–NF90 complexes in post-transcriptional processes, there is currently no clear molecular model for their functions in cells.

The NF45–NF90 complex binds to double-stranded RNA (dsRNA) and ultraviolet cross-linking analysis in human cells has shown that NF90 primarily associates with Alu elements, particularly in introns and 3′ untranslated regions [7, 9]. Alu elements are abundant in primate genomes, making up ∼10% of the human genome, with ∼50% of Alu elements mapping

to intronic regions [13]. As NF90 cross-links to both forward and reverse Alu elements in cells, and Alu inverted repeats (AluIRs) can form extended stretches of ~300-bp dsRNA, it is likely that the primary association with transcripts is via AluIRs [2, 7]. A large subset of circRNA biogenesis events is driven by complementary AluIR sequences in adjacent introns [13] and NF90 is hypothesized to stabilize these sequences to promote circRNA formation [2]. AluIRs are also promiscuously edited in the nucleus by adenine deaminases acting on RNA (ADARs), with >90% of editing sites mapping to Alu sequences [14]. Knockdown of NF45–NF90 increases A-to-I editing, suggesting these complexes may compete with ADARs for their substrates [6, 7].

NF90, and its splice variant NF110, are made up of three folded domains. At the N-terminus is a 'domain associated with zinc fingers' (DZF) domain, which was identified computationally and noted to co-occur with dsRNA-binding domains (dsRBDs) or zinc fingers [15]. This domain heterodimerizes with a homologous domain in NF45 [16]. The DZF domain is followed by two dsRBDs (dsRBD1 and dsRBD2) that are connected by a 50-amino acid linker. NF90 and NF110 splice variants have a segment of low complexity sequence after dsRBD2. These two isoforms differ in the C-terminal region after residue 688, which generates a short sequence in NF90 and a longer, low complexity sequence in NF110 [17, 18]. The two splice isoforms differ in their subnuclear localization and may have isoform-specific roles [10, 19, 20].

We previously showed, through structural studies, that the DZF domains in NF90 and NF45 are pseudoenzymes, i.e. the domain fold belongs to the family of nucleotidyltransferases, but key catalytic residues have been lost [16]. This family includes enzymes such as poly(A) polymerase and key enzymes of the innate immune response including oligoadenylate synthases (OAS) and cyclic GMP–AMP synthase (cGAS) [21]. OAS and cGAS have a common mechanism of activation whereby association of the nucleotidyltransferase domain with dsRNA or dsDNA, respectively, allosterically activates their catalytic activity [22–26, 27]. Binding to dsRNA by NF45–NF90 complexes is driven by the two dsRBDs; however, the presence of the DZF domain heterodimer increases affinity more than 10-fold, suggesting that this domain contributes to dsRNA binding, perhaps analogously to OAS and cGAS enzymes [28]. Importantly, the DZF domain heterodimer alone does not show appreciable RNA-binding activity [16].

While our previous studies provided insight into recognition of short segments of dsRNA by the tandem dsRBDs of NF90 [28], how RNA is recognized in the context of the longer NF45–NF90 complex is not known. Furthermore, how the conformation of multidomain NF45–NF90 complexes alters on RNA binding is likely to be important for its molecular function, yet this is not well understood. Here, we use an integrative structural biology approach, combining deep phylogenetic analysis, machine learning, solution-based biophysical methods, and electron microscopy (EM), to provide insights into architectural alterations in the NF45–NF90 complex on binding to dsRNA. These studies suggest that NF45–NF90 complexes can coat long segments of dsRNA, both by sandwiching the dsRNA between complexes and through lateral interactions along the dsRNA. This property suggests a unifying molecular mechanism for NF45–NF90 function *in vivo*.

Materials and methods

Phylogenetic and missense mutation analysis of NF110, NF90, and NF45

Orthologous sequences were initially collected using the orthologous matrix browser (OMA; www.omabrowser.org) selecting OMA groups 1190 687 (168 sequences, ILF2/NF45) and 1198 895 (131 sequences, ILF3/NF90) [29]. These groups were supplemented with additional sequences from PSI-BLAST [30] and iteratively using MAFFT [31] with visual inspection in AliView [32] to remove incomplete or low-quality sequences.

Alignments were used as input files for ConSurf conservation analysis using full-length AlphaFold2 models of all proteins [33]. Conservation scores were extracted from the resulting .pdb files using Conservation_score_converter.py. Population missense mutation datasets were downloaded from gnomAD v2.1.1, filtered to remove pathogenic variants, and formatted for 1D plotting [34]. A modified version of Plot Protein [35] was used to generate figures combining tracks of conservation score and missense mutations plotted per residue for each protein. Domain annotations and secondary structure elements were based on solved structures [16, 28].

NF45–NF90 protein complex expression and purification

Mouse NF90$_{long}$ (residues 1 to 591) and full-length human NF45 were expressed separately in *Escherichia coli* BL21(*DE3*) cells (Novagen) containing pRIPL plasmids (Agilent) encoding transfer RNAs for rare codons, in 2XTY media. NF90$_{long}$ was expressed as a non-cleavable C-terminally $6\times$His-tagged protein and NF45 as a cleavable N-terminal GST-tagged protein. Cells were induced at $20°C$ overnight with 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cell pellets were co-lysed together in lysis buffer [40 mM Tris–HCl (pH 7.5), 500 mM NaCl, 1 mM dithionthreitol (DTT)] with 100 mM Pefabloc (Roche) and protease inhibitor cocktail (Roche cOmplete EDTA-free) using a cell disruptor (Constant Systems). The tagged proteins were extracted as a complex from clarified lysates by binding to GSH resin (Cytiva), packed into a column and eluted with lysis buffer supplemented with 20 mM reduced glutathione (pH 7.5). The eluted protein complex was dialysed overnight in 40 mM Tris–HCl (pH 8.0), 200 mM NaCl, 10 mM imidazole (pH 8.0), and 1 mM β-mercaptoethanol. Rhinovirus 3C protease was included to allow simultaneous cleavage of the GST tag from NF45. The dialysed protein complex was then bound to Ni^{2+}-NTA resin (Sigma) and eluted over a 0.01–0.5 M imidazole gradient. Eluted protein complex was dialysed overnight into 40 mM Tris–HCl (pH 7.5), 100 mM NaCl, and 1 mM DTT. Heparin Sepharose chromatography (Cytiva) was used to remove nucleic acid contamination with sample eluted using a gradient from 0.01–1 M NaCl. Finally, the protein complex was purified using size exclusion chromatography (SEC; Superdex 200; Cytiva) in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT.

Small angle X-ray scattering

The top and bottom RNA strands for 25, 36, and 54 bp (Supplementary Table S1) were annealed for 5 min at $95°C$

and cooled overnight to form dsRNA. NF45–NF90 protein was complexed with different lengths of dsRNA as follows:

2:1 and 4:1 molar ratio for NF45–NF90 protein:25-bp dsRNA and protein:36-bp dsRNA; 2:1, 4:1, and 6:1 molar ratio for NF45–NF90 protein:54-bp dsRNA. The samples were then incubated on ice for 30 min before transferring to a sample loader for injection on to the SEC column.

SEC coupled to small angle X-ray scattering (SAXS) experiments (SEC–SAXS) were carried out at Diamond Light Source on the B21 beamline [36]. Samples were loaded onto a S200 increase 3.2/200 (Cytiva) column in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT. A total of 60 µl of 5.5 mg/ml sample was loaded onto the column at a flow rate of 0.075 ml/min for NF45–NF90 complex mixed with dsRNA. Again, 60 µl of 7.75–10 mg/ml was loaded onto the column at a flow rate of 0.1 ml/min for samples of NF45–NF90 constructs. Measurements were taken using an exposure time of 0.005 s. ScÅtter IV (https://bl1231.als.lbl.gov/scatter/) was used for solvent subtraction and basic analyses. Elution profiles were extracted in CHROMIXS [37]. *Ab initio* bead models were generated with GASBORI for protein samples and DAMMIF for samples consisting of protein and RNA [38, 39]. GASBORI was run within the ATSAS 3.2.1 software suite, while DAMMIF was launched from within PRIMUS [40]. SAXS envelopes were generated in PyMOL, and then models of NF45–NF90 and A-form 25, 36, or 54-bp dsRNA (generated in COOT) were fitted within the envelopes [41, 42].

Models of NF45–NF90$_{long}$ samples, NF45$_{DZF}$–NF90$_{DZF}$, and NF90$_{dsRBDs}$ were fitted against associated SAXS data using MultiFoXS (https://modbase.compbio.ucsf.edu/multifoxs/) [43]. Models submitted to MultiFoXS were generated by overlaying the crystal structures 4AT7 and 5DV7 with AlphaFold2 models of NF90 (AF-Q9Z1X4) and NF45 (AF-Q9CXY6) in COOT [16, 28, 41, 44]. Unsolved sections of the crystal structures were filled in by merging parts of the AlphaFold2 models into the structures, retaining only relevant residues in coordinate files for each construct. The coordinates were then minimized in Chimera 1.16 to avoid steric clashes [45]. SAXS details are given in Supplementary Table S2.

Mass photometry

Data collection was carried out on the Refeyn TwoMP instrument. A drop of immersion oil was placed on the lens, then a coverslip with a silicon gasket was placed on top. Contrast-to-mass calibration was carried out using bovine serum albumin. A total of 18.5 µl of 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT was added to the silicon gasket, a clean area of the coverslip was located, and the focus was set. A total of 1.5 µl of protein solution at 500 nM was added and mixed, to make up a final volume of 20 µl (final concentration 37.5 nM). Videos (30 s) were recorded using AcquireMP software followed by processing and analysis, including fitting Gaussian functions to identify peaks, using the DiscoverMP software (Refeyn Ltd).

Multi-angle light scattering

A total of 100 µl of purified NF45–NF90$_{long}$ at a concentration of 0.36 mg/ml was injected into a Superdex 200 increase column (Cytiva) in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT. The SEC column was coupled to a Viscotek V3580 refractive index unit and Viscotek SEC–MALS 20 (Malvern Instruments) to measure the molecular mass of the protein complex and its angular dependence of light scattering using OMNISEC software (Malvern Panalytical Ltd.).

NF45–NF90$_{long}$ protein–RNA complex formation for cross-linking analysis

The 18-bp 2′-fluorinated dsRNA (a kind gift from Frank Rigo, Ionis Pharmaceuticals; Supplementary Table S1) was used for titrating cross-linker concentrations with NF45–NF90$_{long}$. RNA strands were annealed for 5 min at 95°C and cooled down gradually overnight to form 18mer dsRNA. Purified NF45–NF90$_{long}$ was complexed with 18mer dsRNA in a molar ratio of 1:1 and 2:1 by incubating the protein and RNA on ice for 30 min and passing through SEC (Superdex 200, Cytiva) in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT. The protein–RNA complex from the main peak was used for cross-linking titrating assay.

A 25-bp GC-rich RNA sequence was used for quantitative cross-linking (Supplementary Table S1). RNA strands were annealed for 5 min at 95°C and cooled down gradually overnight to form the 25-bp dsRNA. Purified NF45–NF90$_{long}$ was complexed with 25-bp dsRNA in a molar ratio of 2:1 by incubating the protein and RNA on ice for 30 min and passing through SEC (Superdex 200, Cytiva) in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT. The protein–RNA complex from the main peak was used for quantitative cross-linking mass spectrometry (CLMS).

Sample preparation for quantitative CLMS

Triplicate samples of 10 µg of NF45–NF90$_{long}$ protein complex and NF45–NF90$_{long}$–25-bp GC-rich (Supplementary Table S1) dsRNA complex (2:1 molar ratio) were cross-linked with 30 µg 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; Thermo Fisher Scientific) [EDC:protein = 3:1 (w/w)] in the presence of 66 µg N-hydroxysulphosuccinimide (Thermo Fisher Scientific) on ice, in the dark for 90 min. EDC is a zero-length chemical cross-linker capable of covalently linking primary amines of lysine and the protein N-terminus and, to a lesser extent, hydroxyl groups of serine, threonine, and tyrosine with carboxyl groups of aspartate or glutamate. The cross-linking reaction was quenched using a final concentration of 100 mM Tris–HCl (pH 8.0). Sodium dodecyl sulfate gel-loading dye containing 50 mM DTT was then added to the cross-linked sample. A total of 6 µg of cross-linked sample was loaded per lane and separated for 5 min on a 4%–12% Bis–Tris gel (Invitrogen) using 3-(N-morpholino)propanesulfonic acid (MOPS) buffer. The gel was stained with Instant Blue (Expedeon). Stained bands were excised and washed with 50 mM ammonium bicarbonate and 100% acetonitrile (ACN) to remove the stain. The gel pieces were reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide for 20 min at room temperature. Samples were treated using 13 ng/µl trypsin (Promega) overnight at 37°C [46]. The sample was acidified with 0.1% trifluoroacetic acid (TFA). Digested peptides were loaded on stage-tips containing am Empore Disk C18 filter conditioned with methanol and equilibrated with 0.1% TFA [47]. The bound peptides were treated with 80% ACN/0.1% TFA prior to analysis by liquid chromatography mass spectrometry (LC–MS).

Quantitative CLMS data collection and analysis

LC–MS/MS analysis was performed using Orbitrap Fusion Lumos (Thermo Fisher Scientific) with a 'high/high' acquisition strategy. The peptide separation was carried out on an EASY-Spray column (50 cm × 75 μm i.d., PepMap C18, 2 μm particles, and 100 Å pore size; Thermo Fisher Scientific). Mobile phase A consisted of water and 0.1% (v/v) formic acid. Mobile phase B consisted of 80% (v/v) ACN and 0.1% (v/v) formic acid. Peptides were loaded at a flow rate of 0.3 μl/min and eluted at 0.2 μl/min using a linear gradient going from 2% mobile phase B to 40% mobile phase B over 109 min (each sample was analysed twice), followed by a linear increase from 40% to 95% mobile phase B in 11 min. The eluted peptides were directly introduced into the mass spectrometer. MS data were acquired in the data-dependent mode with 3-s acquisition cycle. Precursor spectra were recorded in the Orbitrap with a resolution of 120 000 and m/z range of 300–1700. The ions with a precursor charge state between 3+ and 8+ were isolated with a window size of 1.6 m/z and fragmented using high-energy collision dissociation with collision energy 30. The fragmentation spectra were recorded in the Orbitrap with a resolution of 15 000. Dynamic exclusion was enabled with single repeat count and 60-s exclusion duration. The mass spectrometric raw files were processed into peak lists using ProteoWizard version 3.0 [48], and cross-linked peptides were matched to spectra using Xi software version 1.6.751 [49] with in-search assignment of monoisotopic peaks [50]. Search parameters were MS accuracy, 3 ppm; MS/MS accuracy, 10 ppm; enzyme, trypsin; cross-linker, EDC; max missed cleavages, four; missing mono-isotopic peaks, two; fixed modification, carbamidomethylation on cysteine; variable modifications, oxidation on methionine; fragments, b and y ions with loss of H_2O, NH_3, and CH_3SOH.

Label-free quantitation on MS1 level was performed using Skyline version 19.1 [51]. Autovalidated cross-linked peptides were introduced as an .ssl file following the standard format for custom spectral libraries in Skyline. In the .ssl file, an entry is generated for each cross-linking feature. A cross-linking feature is defined as a unique peptide spectrum match for a cross-linked peptide with differences in charge state, linkage sites, or modification [52]. Data from Skyline were exported into a .csv file for further processing in Excel and categorised as: control enriched (cross-linking pairs intensity higher in the control group); RNA enriched (cross-linking pairs intensity higher in the control group); and common (cross-linking pairs intensity similar in both groups).

In vitro transcription

pRRG260-Cyp1A1 was a gift from Phillip Newmark (Addgene; plasmid #99081; http://n2t.net/addgene:99081; RRID: Addgene_99081) [53]. The length of the Cyp1A1 insert was reduced by site-directed mutagenesis by designing primers so that 310 and 410 bp lengths of dsRNA could be generated by *in vitro* transcription (Supplementary Table S1). A total of 50 nM of the pRRG260 Cyp1A1 plasmid mutants containing 310 and 410 bp was linearized using BamHI-HF (NEB) to create a run-off transcription template. *In vitro* transcription was performed using T7 RNA polymerase and nucleoside triphosphates. Pyrophosphatase was added to reduce inorganic pyrophosphate precipitation during the reaction. After 3 h incubation at 37°C, DNase I was added and incubated further at 37°C for 30 min to remove the template. The *in vitro* tran-

scribed dsRNA was purified by passing through a PD spintrap G-25 column with a bed volume 0.6 ml (GE).

Negative stain EM

NF45–NF90$_{long}$ protein was complexed with different lengths of *in vitro*-transcribed Cyp1A1 dsRNA (310 and 410 bp) using a 18:1 molar ratio for NF45–NF90 protein:310-bp dsRNA and a 24:1 molar ratio for NF45–NF90 protein:410-bp dsRNA. The protein and RNA were incubated for 45 min on ice. Protein–RNA samples of 4 μl (10 μg/ml) were spotted on glow discharged grids (carbon support film—400 mesh grid, TAAB laboratories) by incubating for 2 min. The grids were then washed twice with 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM DTT and stained with 2% uranyl acetate for 2 min. Images were taken at 40× magnification using JEOL JEM-1400Plus TEM. Representative images were collected on a GATAN OneView camera.

pyRBDome analysis

To generate RNA/ligand-binding site predictions for NF90 and NF45, the AlphaFold2 structures (AF-Q12906-F1 and AF-Q12905-F1, respectively) were submitted to the pyRBDome pipeline [54]. Predictions were downloaded for the prediction algorithms BindUp [55], FTMap [56], RNABindRPlus [57], DisoRDPbind [58], and PST-PRNA [59]. The eXtreme Gradient Boosting ensemble models used the prediction results to calculate the probability of RNA binding for each amino acid in the proteins. These scores were added to the B-factor column of the associated .pdb file and visualized using PyMOL [42].

Results

Population variant analysis supports functional importance of DZF domains

To better understand the functional contribution of domains and motifs in the DZF domain proteins, we undertook an extensive phylogenetic analysis of NF45 and the two splice isoforms of the ILF3 gene, NF110 and NF90 (Fig. 1). NF90 differs from NF110 by a second splice isoform that alters the C-terminal region after residue 687 in the reference sequence (Fig. 1A). Both NF90 and NF110 have additional isoforms that include a four amino acid sequence (NVKQ) inserted after residue 516. We assembled multiple sequence alignments of both ILF3 isoforms and used the alignments, along with AlphaFold2 models from the AlphaFold database (Supplementary Fig. S1), to calculate per-residue conservation scores in ConSurf [60] (Supplementary Fig. S1A). These scores were displayed along the primary structure of all proteins (Fig. 1). AlphaFold2 models showed high to very high pLDDT scores for conserved, folded domains (Supplementary Fig. S1B). Positioned aligned error matrices are consistent with DZF and dsRBD domains forming independent folding units connected with flexible linkers (Supplementary Fig. S1C).

We also considered population missense variants to gain additional data for domain-level functional importance [34]. The gnomAD database contains genomic and exomic data from both healthy adults and patient data reported in ClinVar [61–63]. Based on the 'probability loss-of-function observed over expected upper bound' metric in gnomAD, the ILF2 and ILF3 genes are under purifying selection (0.25 and 0.1,
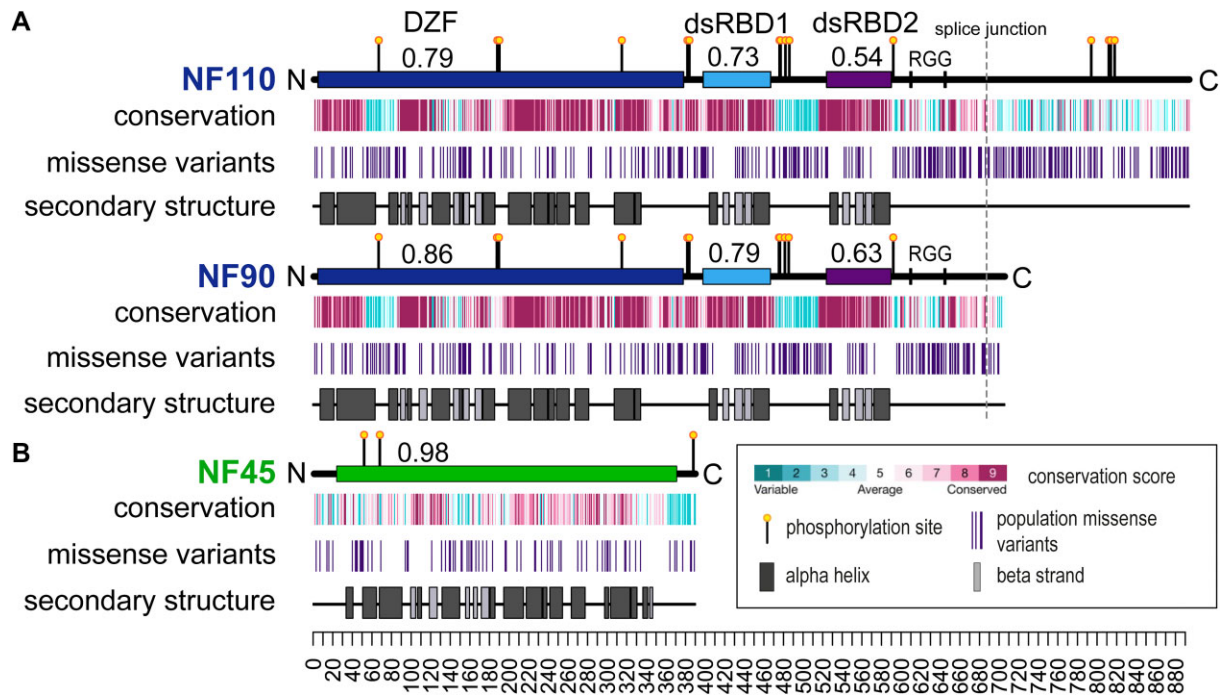
Figure 1. Folded domains and functional sites are conserved and depleted of benign population variants. (**A**) Primary structures of NF110 and NF90. Folded domains are indicated by rectangles along a horizontal line representing the length of the protein from N to C terminus. Fractional values associated with domains are missense depletion per domain over depletion over the whole protein (V_d/V_p) ratios. Lollipops are phosphorylation sites, RGG sequences are marked with a vertical line. Three tracks show conservation (from ConSurf, using a gradient of 9 conservations scores, as indicated in the key); the positions of population missense variants extracted from gnomAD (vertical lines); and positions of secondary structure elements as observed in published structures. (**B**) Similar analysis with equivalent tracks for NF45, which has only one domain.

respectively) [61]. This agrees with DepMap (https://depmap.org/portal) annotations of genes ILF2 and ILF3 as common essential genes [64]. Consequently, patterns of benign missense variants of these proteins across the human population are likely to be informative on functionally important domains [65]. We generated tracks for population missense variants, after removing likely pathogenic and ClinVar mutations, and compared these with the conservation scores on the primary structures of the DZF proteins (Fig. 1) [34]. Domains and motifs with high functional importance are expected to be both conserved and locally depleted of missense variants. Calculated depletion scores (V_d/V_p ratios), where scores <1 indicate localized depletion within a domain, show that the DZF domain is depleted of missense variants ($V_d/V_p = 0.79$–0.86) compared with the rest of the protein [34]. The dsRBD domains are both highly conserved and depleted of missense variants (V_d/V_p range 0.54–0.79; Fig. 1), consistent with their role in RNA recognition.

The variant and conservation analyses further revealed that low complexity and/or natively unstructured regions are less well conserved than structured domains. For example, residues 57–88 in NF110/NF90 have low conservation and correspond to a region of the protein that could not be modelled in the crystal structure [16]. Similarly, the connecting sequence between dsRBD1 and dsRBD2 of NF90 (Fig. 1A) and a glutamate-rich sequence at the C-terminus of NF45 (Fig. 1B) had low conservation scores; these also correspond to protein sequences that could not be modelled in crystallographic data [16, 28]. As expected, these segments are correspondingly enriched in population missense variants. Overall, the patterns of conservation and population variation indicate that both the DZF domains and the dsRBD domains make significant contributions to protein function.

NF45–NF90$_{long}$ shows a compact architecture in solution

As NF45–NF90$_{long}$ is a multidomain complex, understanding its architecture requires a combination of methods. We used SEC–SAXS to characterize the solution structure of the complex in comparison with fragments of the complex that had previously been crystallized: NF45$_{DZF}$–NF90$_{DZF}$ [16] and NF90$_{dsRBDs}$ [28] (Fig. 2, Supplementary Fig. S2, and Supplementary Table S2). NF45$_{DZF}$–NF90$_{DZF}$ separated as a single peak on SEC and showed no appreciable low-angle scatter from aggregates by SAXS (Fig. 2B and C). Scattering curves from the peak of the SEC curve (Supplementary Fig. S2A) were averaged and real space $P(r)$ functions were calculated (Fig 2D). The maximum dimension (D_{max}) derived from the NF45$_{DZF}$–NF90$_{DZF}$ $P(r)$ curve was an excellent match for the longest distance measured in the X-ray crystal structure (126 versus 124 Å, respectively; Fig. 2E). This is consistent with previous observations that the NF45$_{DZF}$–NF90$_{DZF}$ heterodimer is monomeric in solution [16]. Dimensionless Kratky analysis (Supplementary Fig. S2C) [66, 67] and *de novo* bead modelling, calculated using GASBORI (Fig. 2E) [39], indicates that the NF45$_{DZF}$–NF90$_{DZF}$ dimer is an elongated rigid structure, consistent with the crystal structure (Fig. 2E). A similar analysis for NF90$_{dsRBDs}$ (Fig. 2A), showed near-ideal behaviour by SEC–SAXS (Fig. 2B and C and Supplementary Fig. S2A). The D_{max} value derived from the $P(r)$ function for NF90$_{dsRBDs}$ is 126 Å (Fig. 2D). This is consistent with the dsRBDs being
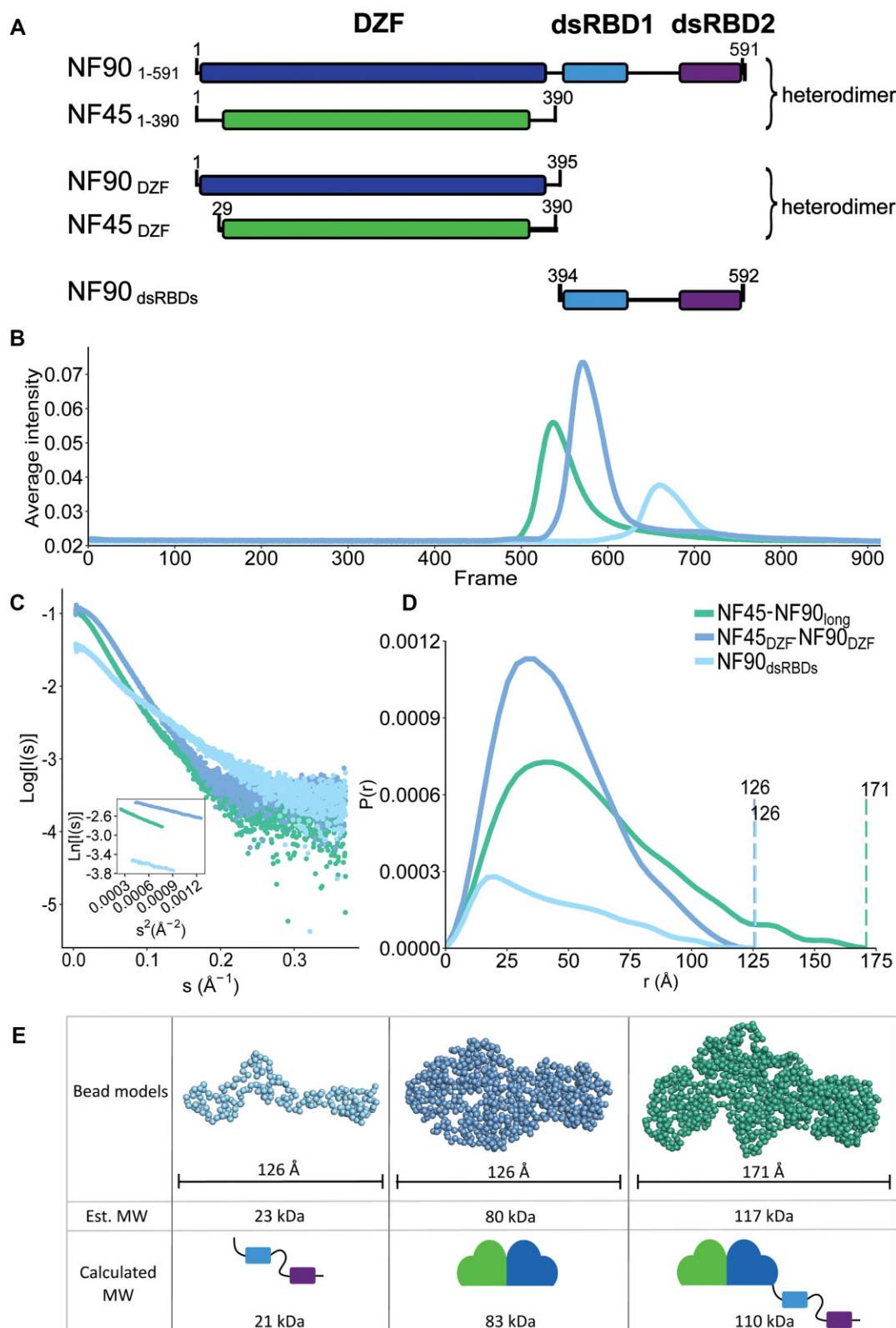
Figure 2. NF45–NF90 complexes in solution show compaction of domains. (**A**) Overview of constructs used for structural analyses. (**B**) SEC–SAXS profiles of three constructs. (**C**) Scattering curves and Guinier analysis (inset) of samples from panel (B). (**D**) $P(r)$ functions derived from panel (C). (**E**) Bead models for NF45–NF90 constructs with D_{max} and calculated MW values (from amino acid composition) or estimated MW from SAXS analysis.

connected by a flexible linker (Fig. 1A). This model is further supported by Kratky analysis, which rises at higher sR_g values, compared with $NF45_{DZF}$–$NF90_{DZF}$ (Supplementary Fig. S2C) [66, 67], and a *de novo* bead model (Fig. 2E), consistent with a dynamic structure.

To understand how domain conformations compare in the context of $NF45$–$NF90_{long}$, we also characterized this complex using SEC–SAXS and SEC coupled to multi-angle light scattering (SEC–MALS; Fig. 2A and B and Supplementary Fig. S2A and B). Scattering curves (Fig. 2C) were averaged from images from the middle of the SEC profile (Supplementary Fig. S2A). The $P(r)$ function derived from these data gave a D_{max} of 171 Å. Given that $NF45_{DZF}$–$NF90_{DZF}$ is 124 Å at its longest point and that the D_{max} of $NF90_{dsRBDs}$ is 126 Å, this suggests that $NF45$–$NF90_{long}$ is a more compact structure than expected by the combination of the three domains. This is further supported by *de novo* bead models of $NF45$–$NF90_{long}$ (Fig. 2E) and Kratky analysis, which shows that $NF45$–$NF90_{long}$ has a profile that is more closely related to $NF45_{DZF}$–$NF90_{DZF}$ than $NF90_{dsRBDs}$, i.e. has limited dynamics. Of the three samples characterized by SAXS, the $NF45$–$NF90_{long}$ complex has the least agreement between the estimated versus the calculated molecular mass (122 versus 110 kDa, respectively). A similar discrepancy is observed with SEC–MALS, where the measured molecular weight (MW) is 128 kDa (Supplementary Fig. S2B). The slightly higher-than-expected MW for $NF45$–$NF90_{long}$ could be explained by a small contribution of dimerization of $NF45$–$NF90_{long}$ complexes.

To further explore the conformational space accessible to NF45–NF90 complexes, we used MultiFoXS [43] to estimate the contributions of different states of $NF45$–$NF90_{long}$ in solution. All-atom models were generated for each sample, based on our prior crystal structures and AlphaFold2 models [44]. We searched for the best agreement with the experimental data described by the fewest possible models (Supplementary Fig. S2D–G). Fitting $NF45$–$NF90_{long}$ models to two independent datasets produced good fits with four or five different conformations ($\chi 2 = 1.18$–1.23; Supplementary Fig. S2D and E). Both sets of models included substantial contributions from structures where dsRBD1 is close to the core NF45–NF90 heterodimerization domain, leading to a compact conformation. The $NF45_{DZF}$–$NF90_{DZF}$ was well described with a single model ($\chi^2 = 1.23$; Supplementary Fig. S2F), consistent with the crystal structure. As expected from *de novo* bead modelling and Kratky analysis, $NF90_{dsRBDs}$ showed major contributions from four distinct conformations, driven by changes in linker sequence connecting the two dsRBDs ($\chi^2 = 1.04$ of fit back calculated to SAXS curve; Supplementary Fig. S2G) [39]. The conformational analysis further indicates that the $NF90_{dsRBDs}$ construct accesses a wider range of conformations than the equivalent residues in a context where the DZF heterodimerization domain is present.

$NF45$–$NF90_{long}$ interactions on dsRNA show steric exclusion

We previously showed that $NF90_{dsRBDs}$ binds to dsRNA of a minimal length of 18 bp and that two molecules of protein typically bind to RNA [28]. Moreover, our previous studies also showed that $NF45$–$NF90_{long}$ complexes have a ~10-fold higher affinity for dsRNA than $NF90_{dsRBDs}$, suggesting that the DZF heterodimerization domain contributes to this in-

teraction. To further explore binding of $NF45$–$NF90_{long}$ on dsRNA, we carried out a series of SEC–SAXS measurements of $NF45$–$NF90_{long}$ with dsRNA lengths of 25, 36, and 54 bp. Comparing SEC–SAXS profiles showed that $NF45$–$NF90_{long}$ alone elutes later than RNA-bound complexes, with the 54-bp dsRNA-associated complexes eluting earliest (Fig. 3A). These observations are consistent with the formation of higher order oligomers of $NF45$–$NF90_{long}$ on dsRNA. Notably, some samples showed a shift to earlier elution volumes when higher protein:RNA molar ratios were used. For example, 4:1 $NF45$–$NF90_{long}$:25-bp dsRNA elutes in two distinct peaks, while 2:1 $NF45$–$NF90_{long}$:25-bp dsRNA shifted intensity to the earlier peak (Fig. 3A). This suggests that in the 4:1 protein:RNA samples, the RNA-binding sites are saturated and excess protein elutes later.

Early SEC peak fractions (Fig. 3A and Supplementary Fig. S3A) were used to generate SAXS curves and Guinier analyses were performed (Supplementary Fig. S3B). Details of analyses are given in Supplementary Table S2. Based on these data, $P(r)$ functions were generated to extract D_{max} values for each sample (Fig. 3B) and Kratky plots were generated (Supplementary Fig. S3C). Kratky analyses indicate that all protein and protein:RNA complexes are consistent with compact multi-domain structures, further suggesting that $NF45$–$NF90_{long}$ complexes become more ordered on binding dsRNA. Consistent with the SEC profiles, the D_{max} values increased with increasing length of dsRNA and higher protein:RNA molar ratios. The molecular mass of each complex was estimated, and DAMMIF bead models were generated for each sample (Fig. 3B and Supplementary Fig. S3D) [38]. These data suggest that, as expected, the number of $NF45$–$NF90_{long}$ dimers bound to dsRNA increases with increasing length. However, while the minimal binding site is 18 bp, we did not observe >4 protein complexes per dsRNA on 54 bp. SAXS samples were characterized in parallel by mass photometry (Supplementary Fig. S4) [68], which gives the population of different complexes found in the sample. Mass photometry experiments are necessarily carried out in the range of 10–50 nM protein (i.e. ~1000× lower than samples used for SEC–SAXS), such that the equilibrium for these interactions shifts towards dissociation. Nevertheless, $NF45$–$NF90_{long}$ complexes on 54-bp dsRNA showed distinct stoichiometries of up to three proteins per dsRNA (Supplementary Fig. S4). This suggests that pairs of $NF45$–$NF90_{long}$ complexes binding to dsRNA impose a steric constraint that extends the binding site to around 19–26 bp (Supplementary Fig. S3D).

$NF45$–$NF90_{long}$ domains are rearranged on RNA binding

To further assess the effect of RNA binding on the architecture of $NF45$–$NF90_{long}$ complexes, we used CLMS with EDC, a zero-length cross-linker (Fig. 4). Titrating EDC concentration with protein in the absence and presence of an 18-bp dsRNA, gave different patterns of bands retarded on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (Supplementary Fig. S5A) [28]. For protein-only samples, EDC generated two main species, one likely representing single $NF45$–$NF90_{long}$ complexes, and an upper band likely representing interacting $NF45$–$NF90_{long}$ complexes. Similar band patterns were present when dsRNA was bound, along with further upper bands indicating formation of larger oligomers
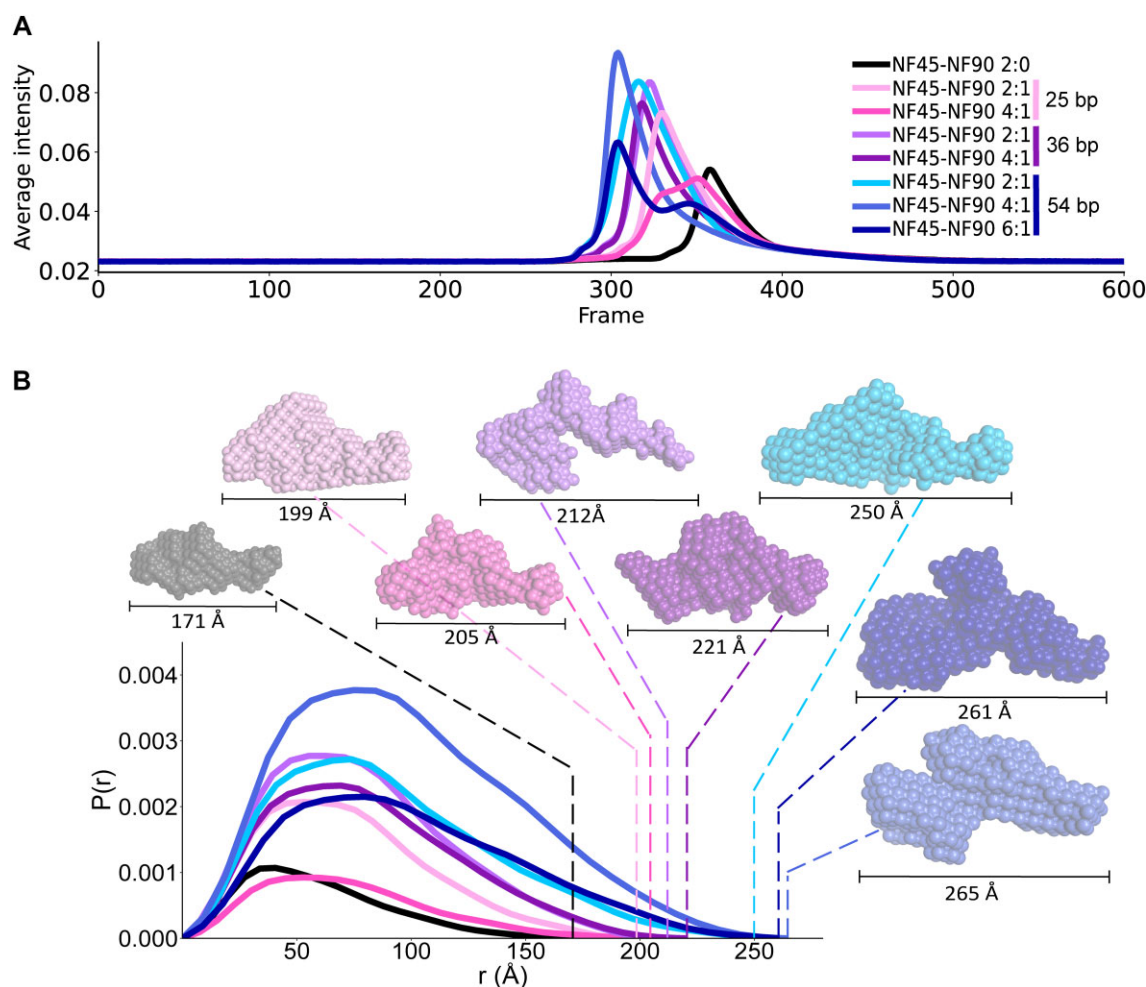
Figure 3. Solution analysis of NF45–NF90 binding to dsRNA of increasing lengths. Eight SAXS measurements were carried out with no dsRNA or with 25-, 36-, and 54-bp RNA at increasing molar ratios of protein:dsRNA. (**A**) Intensity profiles of samples as eluted from SEC. Sample identities for all graphs are given in the inset key. (**B**) Real space profiles of all SAXS samples with dotted lines showing calculated D_{max} on the curve. Associated bead models are shown with D_{max} values indicated under each model.

on RNA. Retardation patterns for NF45–NF90$_{long}$ in 1:1 protein:RNA mixtures gave a less homogeneous distribution of bands (Supplementary Fig. S5A). Consequently, we generated EDC cross-link maps for NF45–NF90$_{long}$ alone and with 18-bp dsRNA in a 2:1 protein to dsRNA ratio (Fig. 4A and B and Supplementary Fig. S5A).

Clusters of cross-links differ between samples of NF45–NF90$_{long}$ without and with dsRNA (compare Fig. 4A and B), indicating that structural rearrangements have occurred. For example, a cluster of cross-links emanating from residue 577 in NF90 shows fewer links when RNA is bound. Similarly, cross-links between the C terminus of NF45 and dsRBD1 and dsRBD2 of NF90 in the RNA-free sample (Fig. 4A) are not evident when RNA is bound (Fig. 4B). The changes observed between the unbound and RNA bound samples likely reflect conformational rearrangements of dsRBD1 and dsRBD2 on binding RNA.

Quantification of cross-links reveals RNA-dependent conformational changes

The cross-linking maps likely include contributions from unbound NF45–NF90$_{long}$ complexes and various stoichiome-

tries of RNA-bound states that are in equilibrium. To focus on differences specific to the unbound and RNA-bound states of NF45–NF90$_{long}$, we repeated EDC CLMS measurements with three independent samples of unbound and RNA-bound protein using a longer, 25-bp GC-rich dsRNA. We then quantified fold changes in peak area between samples. The cross-linking maps were similar to those observed using 18-bp dsRNA (Supplementary Fig. S5B). We then quantified and visualized cross-linked peptides that show a >2-fold increase or reduction between the two states (Fig. 4C and D and Supplementary Table S3), which are a fraction of the total cross-links identified across these samples (Supplementary Fig. S5B–D). The pattern of changes is consistent with changes seen with non-quantified data. For example, cross-links at the C termini of both NF45 and NF90$_{long}$ are reduced in intensity when RNA is bound (Fig. 4C and D).

The data imply an altered conformation of the linker region (residues 377–390) between NF90 DZF domain and dsRBD1 on RNA binding. These include residues D377, E379, and E380 which lose interactions with K454 and T452, both on dsRBD1 (Fig. 4C). Similarly, several cross-links with residues D377, E379, E380, K381, and K392 in this linker show dramatically decreased intramolecular interaction with loop
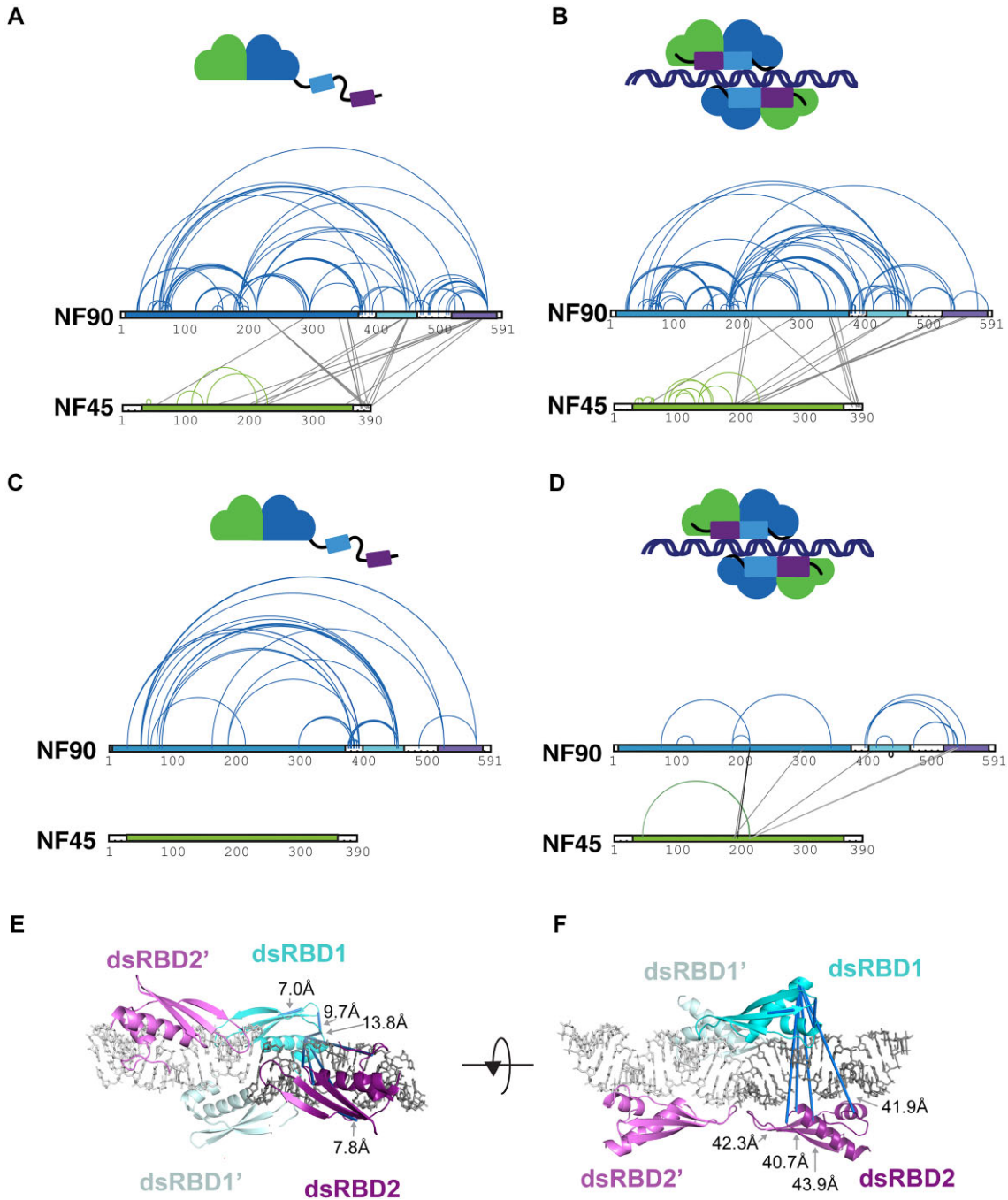
Figure 4. NF45–NF90 complexes undergo conformational change on binding to dsRNA. CLMS on NF45–NF90 complexes in solution without dsRNA (**A**) and with dsRNA (**B**). Intramolecular cross-links within NF90 and NF45 (arcs) are distinguished from intermolecular cross-links (lines). Domain limits are shown as in Fig. 1. Quantitative differences in cross-linking comparing free NF45–NF90 (**C**) and dsRNA-bound NF45–NF90 (**D**). Cross-links with at least two-fold differences are shown while unquantified cross-links are not displayed. (**E**) Mapping of cross-links on to NF90$_{dsRBDs}$. The asymmetric unit of the crystal structure is shown along with the closest symmetry equivalent of the NF90–dsRNA complex. DsRBD domains are cyan and purple with symmetry mates in lighter colours. dsRNA is shown in grey and as sticks. Shortest possible cross-links are shown in dark blue and Cα–Cα distances <24 Å are annotated. (**F**) A rotated view of the same structure is shown. Cα–Cα distances >24 Å are annotated.

57–88 (residues E52, E75, and E83; Fig. 4C and D) when RNA is bound. The linker region is highly conserved and contains a positively charged nuclear localization signal. It is possible that the reorientation of dsRBD1, or direct interactions with dsRNA, or both, contribute to the loss of cross-links on RNA binding. As several residues in this loop have more than one cross-linking partner, these lost interactions likely represent a shift to fewer conformational states.

Both CLMS datasets are consistent with a higher conformational freedom of the dsRBDs, particularly dsRBD2, in the absence of RNA. This is evidenced by the starkly altered pattern of interactions between the dsRBD domains and the DZF domain of NF90 when RNA is bound (Fig. 4C and D). Moreover, new inter-dsRBD interactions appear on binding RNA (D398^{NF90} to K526^{NF90}, K535^{NF90}, and K553^{NF90}; and E472^{NF90} to K540^{NF90} and Y541^{NF90}), as do new interac-

tions with the DZF domain of NF45 ($K413^{NF90}$, $K535^{NF90}$, and $K540^{NF90}$ to $E212^{NF45}$ and/or $E213^{NF45}$; Fig. 4D). There is a corresponding loss of the intramolecular interactions of $K454^{NF90}$ with residues E51, E61, E75, E83, and E187 on NF90 that are observed in the unbound state. Note that residues 57–88 of NF90 were not resolved in the crystal structure [16] and these residues lie in an unstructured, poorly conserved loop of NF90 (Fig. 1A). This further suggests ordering of dsRBD1 on binding to RNA.

Our previous structural studies of $NF90_{dsRBDs}$ on a dsRNA substrate showed a positioning of dsRBD1 and dsRBD2 on either side of the RNA helix [28]. Moreover, a symmetry operation showed that an additional dsRBD1 and dsRBD2 could bind in proximity on a continuous dsRNA helix (Fig. 4E). To test whether the organization of dsRBDs observed in the crystal structure match that in NF45–$NF90_{long}$ with RNA in solution, we examined the distances of cross-links in these structural models, searching for the shortest possible distances (Fig. 4E). While cross-links within individual dsRBDs are all short (Cα–Cα 7.0–13.8 Å) and in the expected range for EDC cross-links (<24 Å), the remaining distances violate the upper limit of expected cross-link size and are ~40–44 Å in this model. These cross-links connect residues within folded domains that are placed on opposing surfaces of the dsRNA in the solved structure (Fig. 4F) and these distances cannot be explained by local conformational flexibility of the protein chain. This analysis shows that the relative orientation of dsRBDs in the NF45–$NF90_{long}$ complex on RNA is likely to be different to that observed in the previous crystal structures.

Oligomers of NF45–$NF90_{long}$ are observed on dsRNA

Among the new cross-links observed in the RNA-bound state (Fig. 4D), three cross-links show >2-fold increase in the quantified data, namely $D194^{NF45}$–$K214^{NF90}$, $E190^{NF45}$–$K214^{NF90}$, and $K186^{NF45}$–$E298^{NF90}$ (Fig. 4D and Supplementary Fig. S5C). Mapping these cross-links onto the structure of $NF45_{DZF}$–$NF90_{DZF}$ heterodimers shows that these links are not possible in the context of a single NF45–NF90 complex as the distances are much longer (67–71 Å) than is possible with the EDC cross-linking chemistry (Fig. 5A). However, these cross-links could be rationalized if they occur between NF45–NF90 heterodimers.

Lateral interactions between NF45–NF90 complexes could potentially provide an extended surface for binding dsRNA. While previous proteomic approaches identified peptides derived from DZF domains that are likely to directly contact RNA [69], we currently have no direct data to show where RNA binds on the surface of the DZF heterodimerization domain. However, this structure has large areas of conserved residues that correlate with regions of positive charge (Supplementary Fig. S6A). To predict likely RNA-binding surfaces, we analysed $NF45_{DZF}$–$NF90_{DZF}$ models with pyRBDome [54], which generates ensemble predictions of protein–RNA interactions based on machine learning approaches. Models of the DZF heterodimerization domain showed substantial areas of the protein with a high likelihood of RNA binding (Fig. 5B and Supplementary Fig. S6A). Surfaces predicted to bind RNA overlap with areas of high conservation on the DZF heterodimerization domain (Fig. 5B and Supplementary Fig. S6A). We modelled the potential RNA-binding surface by placing two $NF45_{DZF}$–$NF90_{DZF}$

complexes laterally; this generated a continuous binding surface with high likelihood of RNA binding (Fig. 5C and Supplementary Fig. S6B). When queried against the cross-linking data, the excessively long cross-links within a DZF domain heterodimer (Fig. 5A) become short cross-links between heterodimers (Fig. 5D) and are well within the distance cutoff of the cross-linker. This lateral binding model provides good agreement with all the cross-linking data of the RNA-bound state (Supplementary Fig. S6C). While the orientation of the RNA in this model is one of several possible poses, the footprint of a single $NF45_{DZF}$–$NF90_{DZF}$ heterodimer on the RNA approximates 26 bp (Fig. 5C and Supplementary Fig. S6D). This agrees well with SAXS analysis (Fig. 3). Moreover, the model provides sufficient space for a second NF45–NF90 heterodimer to bind the opposite surface of the dsRNA, consistent with a 2:1 protein:RNA-binding preference.

Collectively, the data and modelling suggest that NF45–NF90 complexes could form oligomers on extended lengths of dsRNA. To generate longer stretches of dsRNA, we used plasmids with opposing T7 promoters and varied the length of the transcribed sequences. Linearized plasmids were transcribed *in vitro* to generate dsRNA. RNA was then mixed with NF45–$NF90_{long}$ complexes and examined by negative stain EM (Fig. 5E). We observed 'beads on a string' structures in micrographs where protein–RNA complexes were characterized (Fig. 5E–G) that were absent from micrographs where only protein or RNA was stained (Fig. 5E). Measurement of the lengths and widths of these 'beads' gave dimensions of ~160–170 Å for interparticle distances and a similar length dimension for each particle, which is of the same magnitude as measured for single NF45–$NF90_{long}$ in solution (Fig. 2E). Notably, the width of the particles averages around 130 Å (Fig. 5F). Measuring the width of the $NF45_{DZF}$–$NF90_{DZF}$ crystal structure gives dimensions of 45–65 Å. This suggests that the 'beads' are likely to encompass two NF45–$NF90_{long}$ complexes with dsRNA (20 Å diameter) sandwiched between them (Fig. 5H). The negative stain EM confirms that NF45–$NF90_{long}$ complexes can form oligomeric structures that coat extended stretches of dsRNA.

Discussion

The NF45–NF90 complex has many described roles in post-transcriptional regulation in mammals, including regulation of RNA A-to-I editing and control of cassette exon and circular RNA splicing [1, 2, 6, 7, 70]. In humans, the majority of NF45–NF90 is associated with AluIRs [2, 7, 9] that are both the targets of A-to-I editing by ADAR1 [71] and are highly enriched in introns [13]. However, we still lack a molecular understanding of how this protein complex interacts with dsRNA and how such recognition events change the fates of individual RNA species. Prior work showed that dsRNA binding activity is dependent on the dsRBDs of NF90, yet the presence of the heterodimerized $NF45_{DZF}$–$NF90_{DZF}$ domains also contributes to RNA binding. Here, we explore the impact of this larger domain on the architecture of the complex in solution without and with dsRNA.

Deep phylogenetic and sequence variant analysis showed that the DZF domains contribute substantially to NF45 and NF90 function (Fig. 1). Moreover, this analysis highlights several natively unstructured regions of NF110/NF90 that are conserved, indicating a contribution to function. While these contributions are not explored here, these two RGG motifs
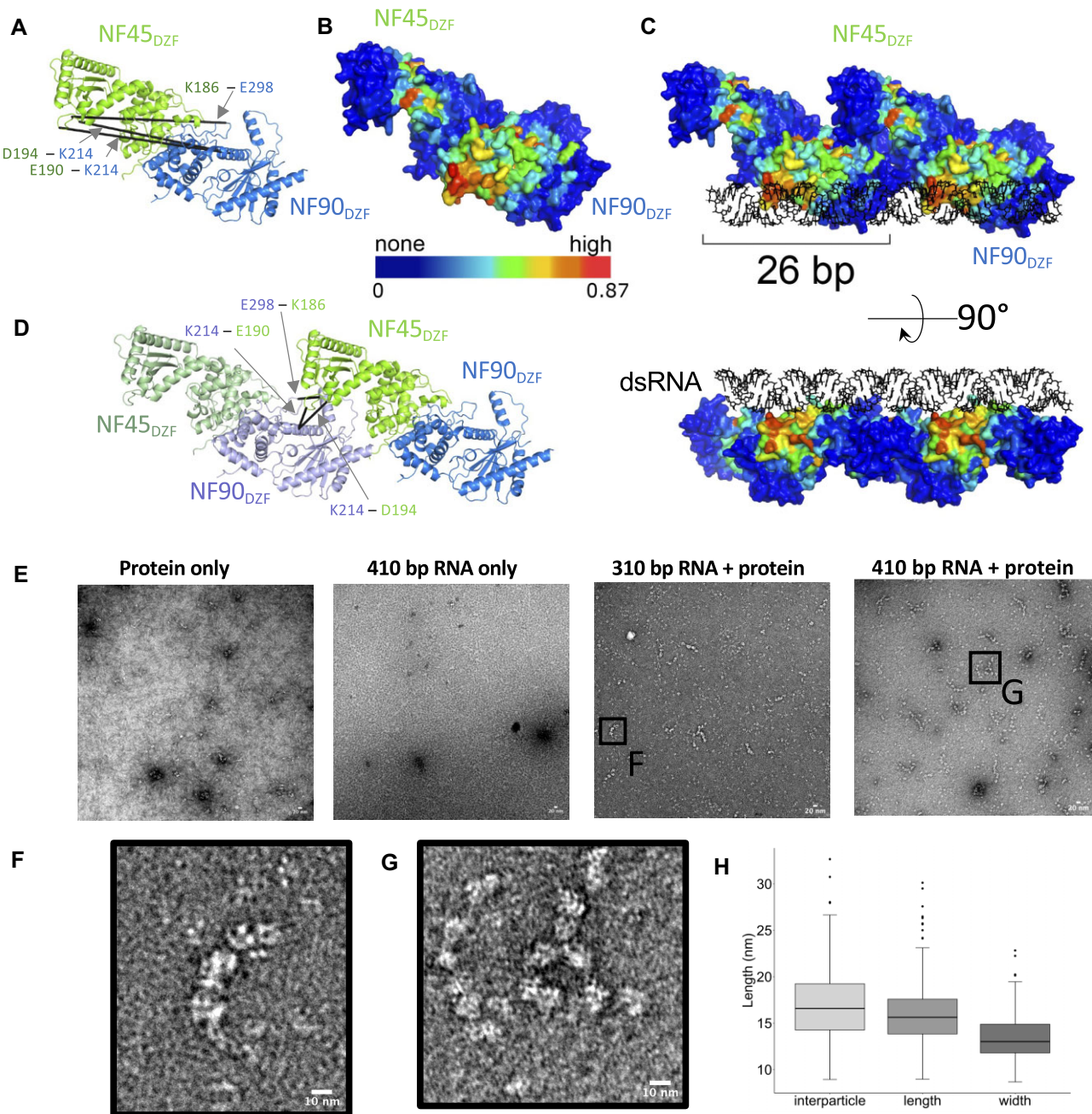
Figure 5. NF45–NF90$_{long}$ complexes oligomerize on long stretches of dsRNA. (**A**) RNA-dependent cross-links between NF90$_{DZF}$ and NF45$_{DZF}$ displayed on a single heterodimer are longer than the distance constraint for EDC cross-linkers. (**B**) Surface of an NF45$_{DZF}$–NF90$_{DZF}$ heterodimer showing a gradient of likelihood of RNA binding, based on pyRBDome analyses (none-to-high, value range is given below gradient bar). (**C**) Combination of two NF45$_{DZF}$–NF90$_{DZF}$ heterodimers into a higher order oligomer model, showing a possible binding site for dsRNA. (**D**) Re-analysis of cross-links in panel (A) in the context of an open-ended oligomer model. (**E**) Negative stain electron micrograph of NF45–NF90$_{long}$ protein alone; dsRNA with a maximum length of 410 bp; 10 mg/ml NF45–NF90$_{long}$ mixed with dsRNA with a maximum length of 310 bp; and 10 mg/ml NF45–NF90$_{long}$ mixed with dsRNA with a maximum length of 410 bp. (**F**) Zoomed-in image of box in panel (E). (**G**) Zoomed-in images of box indicated in panel (E). (**H**) Quantitation of interparticle, length, and width measurements from negative stain micrographs.

and a nearby RGRGRGRG sequence may contribute to RNA binding. The latter sequence is a substrate for PRMT5 symmetric methylation [72]. Downstream of the RG repeat motif is a region of conservation and variant depletion shared in NF110 and NF90 isoforms (residues 660–680), suggestive of a short linear motif that could mediate protein–protein interactions. NF110 has further sites of localized conservation with benign missense variant depletion, around phosphoryla-

tion sites on residues 810, 812, and 816 and a further conserved motif at 845–850. These differences in motif distribution could reflect previously observed differences in subnuclear location and functions of NF110 and NF90 [10, 19, 20].

Extensive characterization of the solution properties of NF45–NF90$_{long}$ in comparison with two previously characterized fragments suggests that the longer complex is relatively

compact in solution (Supplementary Fig. S2 and Fig. 2). Quantitative CLMS data indicate that this compact structure undergoes a substantial domain reorganization on binding dsRNA. New interactions are observed between dsRBDs, as well as between individual dsRBDs and the DZF domains (Fig. 4). Notably, the conserved linker sequence between $NF90_{DZF}$ and $NF90_{dsRBD1}$ has strongly reduced cross-linking when dsRNA binds. This sequence overlaps with the nuclear localization signal (residues 371–389), which may account for its conservation, yet it may also be an important contributor to domain reorganization.

The binding of pairs of $NF45–NF90_{long}$ complexes on dsRNA also creates a steric footprint on the dsRNA, such that around 26 bp are excluded by the binding event (Fig. 3). Analysis of RNA-dependent cross-links indicated that the heterodimerization domain likely engages in lateral contacts when bound on dsRNA. This suggested that NF45–NF90 complexes could oligomerise on dsRNA, a model further supported by a composite prediction of likely RNA-binding sites that fall primarily on $NF90_{DZF}$ (Fig. 5). When lateral interactions are modelled, the outcome is a continuous, flat RNA-binding surface that covers 26 bp of RNA, in good agreement with SAXS measurements (Fig. 3). Lateral interactions promote open-ended complexes or oligomers. We observed that $NF45–NF90_{long}$ complexes can form 'beads-on-a-string' structures on long dsRNAs (Fig. 5E and F). The dimensions of the 'beads' suggest that they are pairs of complexes bound to dsRNA, i.e. that they can coat the surface of the RNA (Fig. 5G).

That NF45–NF90 complexes can coat dsRNA suggests many possibilities of how NF45–NF90 complexes could regulate RNA fate. For example, in the context of A-to-I editing, coating of AluIRs could limit access of ADAR1 to its nuclear substrates, thus maintaining the low levels of editing observed at individual sites [71]. This would explain why loss of NF90 leads to a substantial increase in editing sites in human cells [6, 7]. In the context of cassette splicing, coating of long dsRNA structures by NF45–NF90 in introns could stabilize specific pre-messenger RNA (pre-mRNA) secondary and tertiary structures. These stabilized sites could regulate access of the splicing machinery to splicing donor/acceptor sites, analogous to splicing regulation by polypyrimidine tract binding protein [73–75]. This would explain why loss of NF90 causes both alterations of splicing outcomes and splicing fidelity [1] and how NF90 promotes formation of circular RNAs [2].

Coating of dsRNA by NF45–NF90 is reminiscent of other splicing regulators, such as heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), which can spread along ssRNA sequences from one high affinity site [76]. Spreading of hnRNP A1 requires the presence of a glycine- and tyrosine-rich, intrinsically disordered region (IDR); similar IDRs in other hnRNP proteins were found to promote co-assemblies of these proteins on pre-mRNA [77]. While NF90 and NF110 contain similar IDRs at their C termini, these were not present in the constructs used in our analysis. It is possible that they could further promote the formation NF45–NF90 complexes on dsRNA, perhaps leading to cooperative effects, and this merits future study.

The interaction of NF45–NF90 DZF domains with dsRNA is also reminiscent of the interactions of OAS and cGAS with double-stranded nucleic acids [22–24, 26, 78]. Evolutionarily, cGAS and OAS are older than NF90, with homologs found in *Drosophila* and early metazoans [79, 80]. In vertebrates, these enzymes recognize and respond to nucleic acids associated with viral infection. NF90, in its complex with NF45, may have emerged in vertebrates in response to a different kind of assault on cells, namely the expansion of elements like 7SL, B, and Alu. By virtue of their high abundance, Alu elements contribute to extensive regions of dsRNA in the human transcriptome. NF45–NF90 complexes may have emerged to sequester these potentially problematic RNA structures in the nucleus. However, it is also possible that NF45–NF90 proteins have subsequently co-evolved to use AluIRs to shape co- and post-transcriptional events.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Data availability

SAXS data are deposited in the SASDB; accession numbers are SASDUA5, SASDUB5, SASDUC5, SASDUD5, SASDUE5, SASDUF5, SASDUG5, SASDUH5, SASDUJ5, SASDUK5, and SASDUL5. MS data are deposited in PRIDE with accession number PXD053010. Data and code used to generate figures can be found at https://git.ecdf.ed.ac.uk/cooklab/nf45-nf90-oligomers.

References

1. Haque N, Will A, Cook AG *et al*. A network of DZF proteins controls alternative splicing regulation and fidelity. *Nucleic Acids Res* 2023;**51**:6411–29. https://doi.org/10.1093/nar/gkad351
2. Li X, Liu CX, Xue W *et al*. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* 2017;**67**:214–27. https://doi.org/10.1016/j.molcel.2017.05.023
3. Grasso G, Akkawi C, Franckhauser C *et al*. NF90 interacts with components of RISC and modulates association of Ago2 with mRNA. *BMC Biol* 2022;**20**:194. https://doi.org/10.1186/s12915-022-01384-2
4. Grasso G, Higuchi T, Mac V *et al*. NF90 modulates processing of a subset of human pri-miRNAs. *Nucleic Acids Res* 2020;**48**:6874–88. https://doi.org/10.1093/nar/gkaa386
5. Shang R, Kretov DA, Adamson SI *et al*. Regulated dicing of pre-mir-144 via reshaping of its terminal loop. *Nucleic Acids Res* 2022;**50**:7637–54. https://doi.org/10.1093/nar/gkac568
6. Freund EC, Sapiro AL, Li Q *et al*. Unbiased identification of trans regulators of ADAR and A-to-I RNA editing. *Cell Rep* 2020;**31**:107656. https://doi.org/10.1016/j.celrep.2020.107656
7. Quinones-Valdez G, Tran SS, Jun HI *et al*. Regulation of RNA editing by RNA-binding proteins in human cells. *Commun Biol* 2019;**2**:19. https://doi.org/10.1038/s42003-018-0271-8
8. Tafforeau L, Zorbas C, Langhendries JL *et al*. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. *Mol Cell* 2013;**51**:539–51. https://doi.org/10.1016/j.molcel.2013.08.011
9. Van Nostrand EL, Pratt GA, Yee BA *et al*. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol* 2020;**21**:90. https://doi.org/10.1186/s13059-020-01982-9
10. Wandrey F, Montellese C, Koos K *et al*. The NF45/NF90 heterodimer contributes to the biogenesis of 60S ribosomal subunits and influences nucleolar morphology. *Mol Cell Biol* 2015;**35**:3491–503. https://doi.org/10.1128/MCB.00306-15
11. Nourreddine S, Lavoie G, Paradis J *et al*. NF45 and NF90 regulate mitotic gene expression by competing with Staufen-mediated mRNA decay. *Cell Rep* 2020;**31**:107660. https://doi.org/10.1016/j.celrep.2020.107660
12. Watson SF, Bellora N, Macias S. ILF3 contributes to the establishment of the antiviral type I interferon program. *Nucleic Acids Res* 2020;**48**:116–29.
13. Zhang XO, Wang HB, Zhang Y *et al*. Complementary sequence-mediated exon circularization. *Cell* 2014;**159**:134–47. https://doi.org/10.1016/j.cell.2014.09.001
14. Levanon EY, Eisenberg E, Yelin R *et al*. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* 2004;**22**:1001–5. https://doi.org/10.1038/nbt996
15. Doerks T, Copley RR, Schultz J *et al*. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* 2002;**12**:47–56. https://doi.org/10.1101/gr.203201
16. Wolkowicz UM, Cook AG. NF45 dimerizes with NF90, zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res* 2012;**40**:9356–68. https://doi.org/10.1093/nar/gks696
17. Castella S, Bernard R, Corno M *et al*. Ilf3 and NF90 functions in RNA biology. *WIREs RNA* 2015;**6**:243–56. https://doi.org/10.1002/wrna.1270
18. Duchange N, Pidoux J, Camus E *et al*. Alternative splicing in the human interleukin enhancer binding factor 3 (ILF3) gene. *Gene* 2000;**261**:345–53. https://doi.org/10.1016/S0378-1119(00)00495-9
19. Damianov A, Ying Y, Lin CH *et al*. Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* 2016;**165**:606–19. https://doi.org/10.1016/j.cell.2016.03.040
20. Viranaicken W, Gasmi L, Chaumet A *et al*. L-Ilf3 and L-NF90 traffic to the nucleolus granular component: alternatively-spliced exon 3 encodes a nucleolar localization motif. *PLoS One* 2011;**6**:e22296. https://doi.org/10.1371/journal.pone.0022296
21. Kuchta K, Knizewski L, Wyrwicz LS *et al*. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res* 2009;**37**:7701–14. https://doi.org/10.1093/nar/gkp854
22. Civril F, Deimling T, de Oliveira Mann CC *et al*. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* 2013;**498**:332–7. https://doi.org/10.1038/nature12305
23. Donovan J, Dufner M, Korennykh A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc Natl Acad Sci USA* 2013;**110**:1652–7. https://doi.org/10.1073/pnas.1218528110
24. Donovan J, Whitney G, Rath S *et al*. Structural mechanism of sensing long dsRNA via a noncatalytic domain in human oligoadenylate synthetase 3. *Proc Natl Acad Sci USA* 2015;**112**:3949–54. https://doi.org/10.1073/pnas.1419409112
25. Hartmann R, Justesen J, Sarkar SN *et al*. Crystal structure of the 2′-specific and double-stranded RNA-activated interferon-induced antiviral protein 2′-5′-oligoadenylate synthetase. *Mol Cell* 2003;**12**:1173–85. https://doi.org/10.1016/S1097-2765(03)00433-7
26. Lohofener J, Steinke N, Kay-Fedorov P *et al*. The activation mechanism of 2′-5′-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure* 2015;**23**:851–62. https://doi.org/10.1016/j.str.2015.03.012
27. Zhang X, Wu J, Du F *et al*. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep* 2014;**6**:421–30. https://doi.org/10.1016/j.celrep.2014.01.003
28. Jayachandran U, Grey H, Cook AG. Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. *Nucleic Acids Res* 2016;**44**:1924–36. https://doi.org/10.1093/nar/gkv1508
29. Altenhoff AM, Train CM, Gilbert KJ *et al*. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* 2021;**49**:D373–9. https://doi.org/10.1093/nar/gkaa1007
30. Altschul SF, Madden TL, Schaffer AA *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402. https://doi.org/10.1093/nar/25.17.3389
31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010
32. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;**30**:3276–8. https://doi.org/10.1093/bioinformatics/btu531
33. Varadi M, Bertoni D, Magana P *et al*. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* 2024;**52**:D368–75. https://doi.org/10.1093/nar/gkad1011
34. Deak G, Cook AG. Missense variants reveal functional insights into the human ARID family of gene regulators. *J Mol Biol* 2022;**434**:167529. https://doi.org/10.1016/j.jmb.2022.167529
35. Turner T. Plot Protein: visualization of mutations. *J Clin Bioinforma* 2013;**3**:14. https://doi.org/10.1186/2043-9113-3-14
36. Cowieson NP, Edwards-Gayle CJC, Inoue K *et al*. Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light

Source. *J Synchrotron Radiat* 2020;**27**:1438–46. https://doi.org/10.1107/S1600577520009960

37. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* 2018;**34**:1944–6. https://doi.org/10.1093/bioinformatics/btx846

38. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 2009;**42**:342–6. https://doi.org/10.1107/S0021889809000338

39. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 2001;**80**:2946–53. https://doi.org/10.1016/S0006-3495(01)76260-1

40. Manalastas-Cantos K, Konarev PV, Hajizadeh NR *et al*. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* 2021;**54**:343–55. https://doi.org/10.1107/S1600576720013412

41. Emsley P, Lohkamp B, Scott WG *et al*. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:486–501. https://doi.org/10.1107/S0907444910007493

42. Schrodinger, LLC. The Pymol Molecular Graphics System. Version 2.5.5.

43. Schneidman-Duhovny D, Hammel M, Tainer JA *et al*. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 2016;**44**:W424–9. https://doi.org/10.1093/nar/gkw389

44. Jumper J, Evans R, Pritzel A *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2

45. Pettersen EF, Goddard TD, Huang CC *et al*. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12. https://doi.org/10.1002/jcc.20084

46. Maiolica A, Cittaro D, Borsotti D *et al*. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol Cell Proteomics* 2007;**6**:2200–11. https://doi.org/10.1074/mcp.M700274-MCP200

47. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2007;**2**:1896–906. https://doi.org/10.1038/nprot.2007.261

48. Kessner D, Chambers M, Burke R *et al*. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 2008;**24**:2534–6. https://doi.org/10.1093/bioinformatics/btn323

49. Mendes ML, Fischer L, Chen ZA *et al*. An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* 2019;**15**:e8994. https://doi.org/10.15252/msb.20198994

50. Lenz S, Giese SH, Fischer L *et al*. In-search assignment of monoisotopic peaks improves the identification of cross-linked peptides. *J Proteome Res* 2018;**17**:3923–31. https://doi.org/10.1021/acs.jproteome.8b00600

51. MacLean B, Tomazela DM, Shulman N *et al*. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;**26**:966–8. https://doi.org/10.1093/bioinformatics/btq054

52. Muller F, Fischer L, Chen ZA *et al*. On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J Am Soc Mass Spectrom* 2018;**29**:405–12. https://doi.org/10.1007/s13361-017-1837-2

53. Roberts-Galbraith RH, Newmark PA. Follistatin antagonizes activin signaling and acts with notum to direct planarian head regeneration. *Proc Natl Acad Sci USA* 2013;**110**:1363–8. https://doi.org/10.1073/pnas.1214053110

54. Chu LC, Christopoulou N, McCaughan H *et al*. pyRBDome: a comprehensive computational platform for enhancing RNA-binding proteome data. *Life Sci Alliance* 2024;7:e202402787. https://doi.org/10.26508/lsa.202402787

55. Paz I, Kligun E, Bengad B *et al*. BindUP: a web server for non-homology-based prediction of DNA and RNA binding proteins. *Nucleic Acids Res* 2016;**44**:W568–74. https://doi.org/10.1093/nar/gkw454

56. Kozakov D, Grove LE, Hall DR *et al*. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* 2015;**10**:733–55. https://doi.org/10.1038/nprot.2015.043

57. Walia RR, Xue LC, Wilkins K *et al*. RNABindRPlus: a predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins. *PLoS One* 2014;**9**:e97725. https://doi.org/10.1371/journal.pone.0097725

58. Peng Z, Wang C, Uversky VN *et al*. Prediction of disordered RNA, DNA, and protein binding regions using DisoRDPbind. *Methods Mol Biol* 2017;**1484**:187–203. https://doi.org/10.1007/978-1-4939-6406-2_14

59. Li P, Liu ZP. PST-PRNA: prediction of RNA-binding sites using protein surface topography and deep learning. *Bioinformatics* 2022;**38**:2162–8. https://doi.org/10.1093/bioinformatics/btac078

60. Yariv B, Yariv E, Kessel A *et al*. Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. *Protein Sci* 2023;**32**:e4582. https://doi.org/10.1002/pro.4582

61. Karczewski KJ, Francioli LC, Tiao G *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43. https://doi.org/10.1038/s41586-020-2308-7

62. Landrum MJ, Chitipiralla S, Brown GR *et al*. ClinVar: improvements to accessing data. *Nucleic Acids Res* 2020;**48**:D835–44. https://doi.org/10.1093/nar/gkz972

63. Landrum MJ, Lee JM, Riley GR *et al*. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucl Acids Res* 2014;**42**:D980–5. https://doi.org/10.1093/nar/gkt1113

64. Tsherniak A, Vazquez F, Montgomery PG *et al*. Defining a cancer dependency map. *Cell* 2017;**170**:564–76. https://doi.org/10.1016/j.cell.2017.06.010

65. Iqbal S, Perez-Palma E, Jespersen JB *et al*. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. *Proc Natl Acad Sci USA* 2020;**117**:28201–11. https://doi.org/10.1073/pnas.2002660117

66. Putnam CD, Hammel M, Hura GL *et al*. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quart Rev Biophys* 2007;**40**:191–285. https://doi.org/10.1017/S0033583507004635

67. Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod–Debye law. *Biopolymers* 2011;**95**:559–71. https://doi.org/10.1002/bip.21638

68. Young G, Hundt N, Cole D *et al*. Quantitative mass imaging of single biological macromolecules. *Science* 2018;**360**:423–7. https://doi.org/10.1126/science.aar5839

69. Castello A, Fischer B, Frese CK *et al*. Comprehensive identification of RNA-binding domains in Human cells. *Mol Cell* 2016;**63**:696–710. https://doi.org/10.1016/j.molcel.2016.06.029

70. Surka C, Jin L, Mbong N *et al*. CC-90009, a novel cereblon E3 ligase modulator, targets acute myeloid leukemia blasts and leukemia stem cells. *Blood* 2021;**137**:661–77. https://doi.org/10.1182/blood.2020008676

71. Bazak L, Levanon EY, Eisenberg E. Genome-wide analysis of Alu editability. *Nucleic Acids Res* 2014;**42**:6876–84. https://doi.org/10.1093/nar/gku414

72. Cox J, Esser LM, Judt M *et al*. NF90/NFAR (nuclear factors associated with dsRNA)—a new methylation substrate of the PRMT5-WD45-RioK1 complex. *Biol Chem* 2022;**403**:907–15. https://doi.org/10.1515/hsz-2022-0136

73. Amir-Ahmady B, Boutz PL, Markovtsov V *et al.* Exon repression by polypyrimidine tract binding protein. *RNA* 2005;**11**:699–716. https://doi.org/10.1261/rna.2250405

74. Auweter SD, Allain FH. Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol Life Sci* 2008;**65**:516–27. https://doi.org/10.1007/s00018-007-7378-2

75. Liu H, Zhang W, Reed RB *et al.* Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein. *RNA* 2002;**8**:137–49. https://doi.org/10.1017/S1355838202015029

76. Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol* 2009;**29**:5620–31. https://doi.org/10.1128/MCB.01678-08

77. Gueroussov S, Weatheritt RJ, O'Hanlon D *et al.* Regulatory expansion in mammals of multivalent hnRNP assemblies that globally control alternative splicing. *Cell* 2017;**170**:324–39. https://doi.org/10.1016/j.cell.2017.06.037

78. Hornung V, Hartmann R, Ablasser A *et al.* OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat Rev Immunol* 2014;**14**:521–8. https://doi.org/10.1038/nri3719

79. Cai H, Imler JL. cGAS-STING: insight on the evolution of a primordial antiviral signaling cassette. *Fac Rev* 2021;**10**:54. https://doi.org/10.12703/r/10-54

80. Wiens M, Kuusksalu A, Kelve M *et al.* Origin of the interferon-inducible (2′-5′)oligoadenylate synthetases: cloning of the (2′-5′)oligoadenylate synthetase from the marine sponge *Geodia cydonium*. *FEBS Lett* 1999;**462**:12–8. https://doi.org/10.1016/S0014-5793(99)01478-7