



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# Identification of novel interferon responsive protein partners of human leukocyte antigen A (HLA-A) using cross-linking mass spectrometry (CLMS) approach

### Citation for published version:

Singh, A, Padariya, M, Faktor, J, Kote, S, Mikac, S, Dziadosz, A, Lam, TW, Brydon, J, Wear, MA, Ball, KL, Hupp, T, Sznarkowska, A, Vojtesek, B & Kalathiya, U 2022, 'Identification of novel interferon responsive protein partners of human leukocyte antigen A (HLA-A) using cross-linking mass spectrometry (CLMS) approach', *Scientific Reports*, vol. 12, no. 1, 19422. <https://doi.org/10.1038/s41598-022-21393-z>

### Digital Object Identifier (DOI):

[10.1038/s41598-022-21393-z](https://doi.org/10.1038/s41598-022-21393-z)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Scientific Reports

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





OPEN

## Identification of novel interferon responsive protein partners of human leukocyte antigen A (HLA-A) using cross-linking mass spectrometry (CLMS) approach

Ashita Singh<sup>1,2</sup>, Monikaben Padariya<sup>3</sup>, Jakub Faktor<sup>3</sup>, Sachin Kote<sup>3</sup>, Sara Mikac<sup>3</sup>, Alicja Dziadosz<sup>3</sup>, Tak W. Lam<sup>1</sup>, Jack Brydon<sup>1</sup>, Martin A. Wear<sup>5</sup>, Kathryn L. Ball<sup>1</sup>, Ted Hupp<sup>1,3</sup>, Alicja Sznarkowska<sup>3</sup>, Borek Wojtesek<sup>4✉</sup> & Umesh Kalathiya<sup>3✉</sup>

The interferon signalling system elicits a robust cytokine response against a wide range of environmental pathogenic and internal pathological signals, leading to induction of a subset of interferon-induced proteins. We applied DSS (disuccinimidyl suberate) mediated cross-linking mass spectrometry (CLMS) to capture novel protein–protein interactions within the realm of interferon induced proteins. In addition to the expected interferon-induced proteins, we identified novel inter- and intra-molecular cross-linked adducts for the canonical interferon induced proteins, such as MX1, USP18, OAS3, and STAT1. We focused on orthogonal validation of a cohort of novel interferon-induced protein networks formed by the HLA-A protein (H2BFS-HLA-A-HMGA1) using co-immunoprecipitation assay, and further investigated them by molecular dynamics simulation. Conformational dynamics of the simulated protein complexes revealed several interaction sites that mirrored the interactions identified in the CLMS findings. Together, we showcase a proof-of-principle CLMS study to identify novel interferon-induced signaling complexes and anticipate broader use of CLMS to identify novel protein interaction dynamics within the tumour microenvironment.

Prior to the onset of the adaptive immune response, the innate host defense system establishes an antimicrobial response, which is mediated by a family of secreted  $\alpha$ -helical cytokines known as interferons (IFN). Type I-IFN species IFN $\alpha$  and IFN $\beta$  activate cellular responses, including antiviral, pro-apoptotic, pro-inflammatory, and anti-proliferative states. There are 13 IFN $\alpha$  subtypes known in humans, all of which are clustered on chromosome 9<sup>1</sup>. Surprisingly, only IFN $\alpha$ 2 has been explored for use in clinical settings. Lately, emphasis on the study of additional IFN $\alpha$  subtypes has been highlighted. A recent study identified IFN $\alpha$ 14 as one of the most effective subtypes for restricting the replication of both hepatitis B virus<sup>2</sup> and HIV-1<sup>3,4</sup> when compared to the canonical subtype IFN $\alpha$ 2.

It has been well established that activated type I-IFN receptor complex (IFNAR1 and IFNAR2) triggers a signal transduction cascade mediated by Janus kinases TYK2 and JAK1<sup>5,6</sup>. These Janus kinases phosphorylate signal transducers and activators of transcription proteins (STAT1 and STAT2) on tyrosine residues to initiate SH2 domain-mediated heterodimerization<sup>6</sup>. Subsequently, IRF9 binds the STAT heterodimer to form a trimeric IFN-stimulated gene factor 3 (ISGF3) complex that translocates to the nucleus, inducing transcription of over 2000 interferon-stimulated genes (ISGs)<sup>5–8</sup>.

The ISGs form the backbone of the innate immune system, particularly in response to viral attack. As a first line of defense against viral infections, cells rapidly deploy an extensive interaction of cellular proteins possessing a wide range of biological activities. Among these proteins are pattern-recognition receptors, signalling molecules, transcription factors and proteins with direct antiviral functions along with negative regulators of

<sup>1</sup>Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XR, Scotland, UK. <sup>2</sup>Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic. <sup>3</sup>International Centre for Cancer Vaccine Science, University of Gdansk, ul. Kładki 24, 80-822 Gdansk, Poland. <sup>4</sup>RECAMO, Masaryk Memorial Cancer Institute, Zlutykopec 7, 65653 Brno, Czech Republic. <sup>5</sup>School of Biological Sciences, Institute of Structural and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK. ✉email: wojtesek@mou.cz; umesh.kalathiya@ug.edu.pl

the immune response<sup>9</sup>. Most of the knowledge on ISG activities has been gained from functional screens, either using overexpression screens<sup>10,11</sup> or gene suppression techniques (siRNA, RNAi, and CRISPR)<sup>12,13</sup> in which single ISGs are expressed or repressed, and their activity is tested against different viruses. Even though these studies identified antiviral properties for individual ISGs, underlying molecular mechanisms behind each one largely remain unknown. It is widely accepted that many proteins interact with one or more cellular factors to achieve full activity; hence, either ISGs interact directly with each other, or their interactions are mediated by cellular proteins. For instance, a recent study using photo-crosslinking proteomics identified an ATPase, VCP/p97 as the primary interaction partner of IFITM3 and its inhibition leads to defective IFITM3 lysosomal sorting, turnover, and co-traffic with virus particles<sup>14</sup>. Using immunoprecipitation we have identified VAPA, a vesicle associated protein, as an interaction partner of IFITM1/2/3, which mediates cholesterol-mediated viral maturation and this has been supported by another study using yeast two-hybrid system<sup>15,16</sup>.

A fundamental biological process involved in the suppression of infection and malignant transformation is antigen presentation, which is mediated by major histocompatibility complex (MHC) molecules. Peptides (8–12 amino acids long) originating from the degraded proteins, prematurely terminated or misfolded proteins are loaded on to the MHC-I heterodimer (comprised of MHC-I heavy chain and light chain named  $\beta$ -2-microglobulin;  $\beta$ 2M)<sup>17,18</sup>. The resulting stable MHC-I trimers are transported to the cell surface to present an intracellular peptide to CD8+ T cells (cytotoxic T cells)<sup>17</sup>. T cells recognize and eliminate these pathogen and tumor-specific antigen bearing cells. As in consequence, pathogens and tumor cells often downregulate the antigen presentation process to evade immune surveillance. Moreover, MHC-I is downregulated in 40–90% of human tumors and often correlates with a worse prognosis<sup>19</sup>.

Genes involved in response to pathogens are required to rapidly switch between resting-state and active transcription state. Therefore, several cellular proteins are assumed to be involved in coping with the high demand of IFNs within a short timeframe, including remodeling and modification of promoter chromatin<sup>20,21</sup>. Most studies have focused on identification of protein partners of individual ISGs in the presence of IFN(s). Several proteomics and transcriptomic studies on model cell systems have illuminated the impact of IFNs on the cellular landscape. Yet, despite growing knowledge on the dynamic changes induced by IFNs, we still know relatively little about the engagement of ISGs. When considering the complexity and time-dependent dynamics of the interferon signalling, two questions arise: (i) if multiprotein complexes involved in rapid signalling can be stabilized and captured, and (ii) could these interactions be mapped in three-dimensional space?

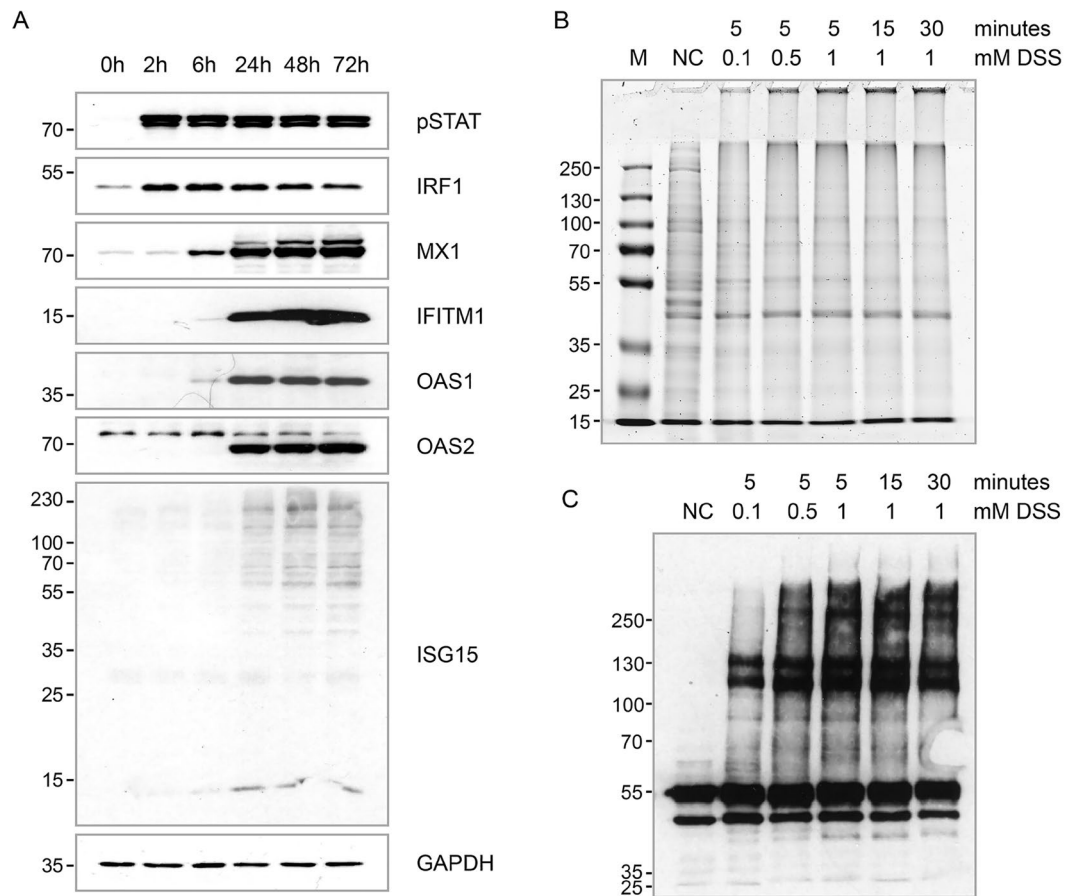
To address these questions, we implemented disuccinimidyl suberate (DSS)-mediated chemical cross-linking combined with mass spectrometry (CLMS) to investigate IFN $\alpha$ -induced protein interaction networks and their dynamics. The DSS adds covalent bonds between proximal residues of proteins and/or protein complexes in vivo. Subsequent MS analysis reveals the specific cross-linking sites, which reflect spatial proximity of regions within a particular protein, called intra-links, or of subunits in a protein complex, called inter-links. Using this approach, we have identified several novel protein–protein complexes as well as multiprotein interaction networks induced by interferon exposure. By further validation of a subset of these novel interactions, we demonstrate H2BFS (histone H2B type F-S; hereafter mentioned as H2B) and MDN1 as binding partners of HLA-A.

## Results

**In-situ CLMS to study interferon stimulated proteins.** Flo-1 cells are one of the most established in vitro Esophageal Adenocarcinoma models as they recapitulate key characteristics of tumours in the esophageal tube<sup>22,23</sup>. However, not all tumours are immunogenic, and to determine if Flo-1 cells show response to interferon treatment, we treated Flo-1 cells with 10 ng/ml IFN $\alpha$  for up to 72 h. Flo-1 cells showed early induction of pSTAT1 and IRF1 starting at 2 h after treatment, which was sustained over the 72 h time-course with a time-dependent decrease in IRF1 steady state levels (Fig. 1A). The ISGs (MX1, IFITM1, OAS1/2, and ISG15) were found to be highly induced after 6 h, mimicking classic intermediate and late responses to IFN $\alpha$  (Fig. 1A). These data together indicate that this cell model can be used to study the interferon response.

To capture the protein interaction landscape in-situ, we used DSS, a widely used cross-linking reagent, for its high membrane permeability and relatively short reaction time. The short reaction time helps to prevent formation of the large, cross-linked protein aggregates thereby maintaining the stability of the cross-linker. To determine the optimum concentration of the DSS and to avoid over cross-linking, we first treated the cells with 5, 2.5 and 1 mM DSS for 5, 10, 5 and 30 min each, and analysed the lysates by Coomassie staining SDS-PAGE (data not shown). The cell lysate appeared to be highly cross-linked at the lowest concentration and shortest time point. DSS was therefore titrated to 1, 0.5 and 0.1 mM for 5 min (Fig. 1B). Optimal cross-linking was observed with 0.5 mM DSS for 5 min and these conditions were selected for IFN $\alpha$  treated cells. Additionally, Fig. 1C represents an immunoblot probed with p53 (DO-1) antibody to assess the degree of protein cross-linking.

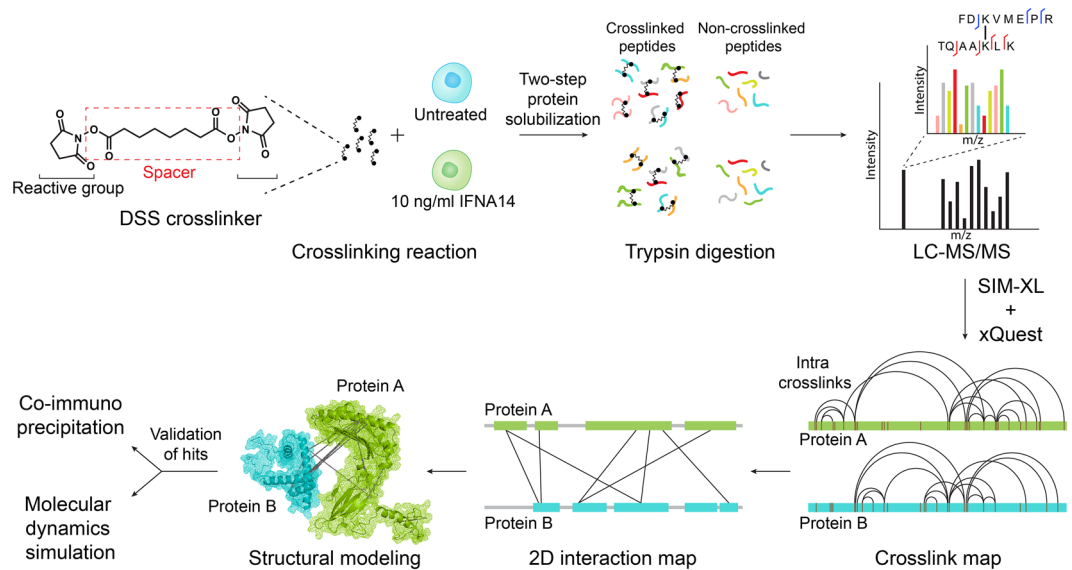
The Flo-1 cells were treated with 10 ng/ml IFN $\alpha$  for 24 h prior to cross-linker addition. Cross-linked cells were subsequently lysed using a two-step protein solubilization method and proteins were processed by the FASP method (Fig. 2)<sup>24,25</sup>. Cross-linked tryptic peptides were analyzed by mass spectrometry (Fig. 2). Next, the MS/MS spectra were aligned to protein sequences and quantitative analysis was carried out using MaxQuant<sup>26,27</sup>. The cross-linked peptides were identified from the obtained spectra using the SIM-XL program and individual linkages were merged into a complex network using an open-source computational software pipeline xQuest<sup>28</sup> along with SIM-XL<sup>29</sup> (Fig. 2). SIM-XL identifies protein–protein interaction, intra-links and mono-links in either simple or complex protein mixtures and provides scripts to visualize the interactions in the protein structure. In addition, it ranks each cross-link as an ID score depending on the quality of the MS/MS spectra<sup>29</sup>. Several high-confidence protein–protein interactions and complexes were identified, and a cohort of the novel interactions were further investigated using co-immunoprecipitation and the conformational changes of the complexes were studied using molecular dynamics (MD) simulation (Fig. 2)<sup>30,31</sup>.



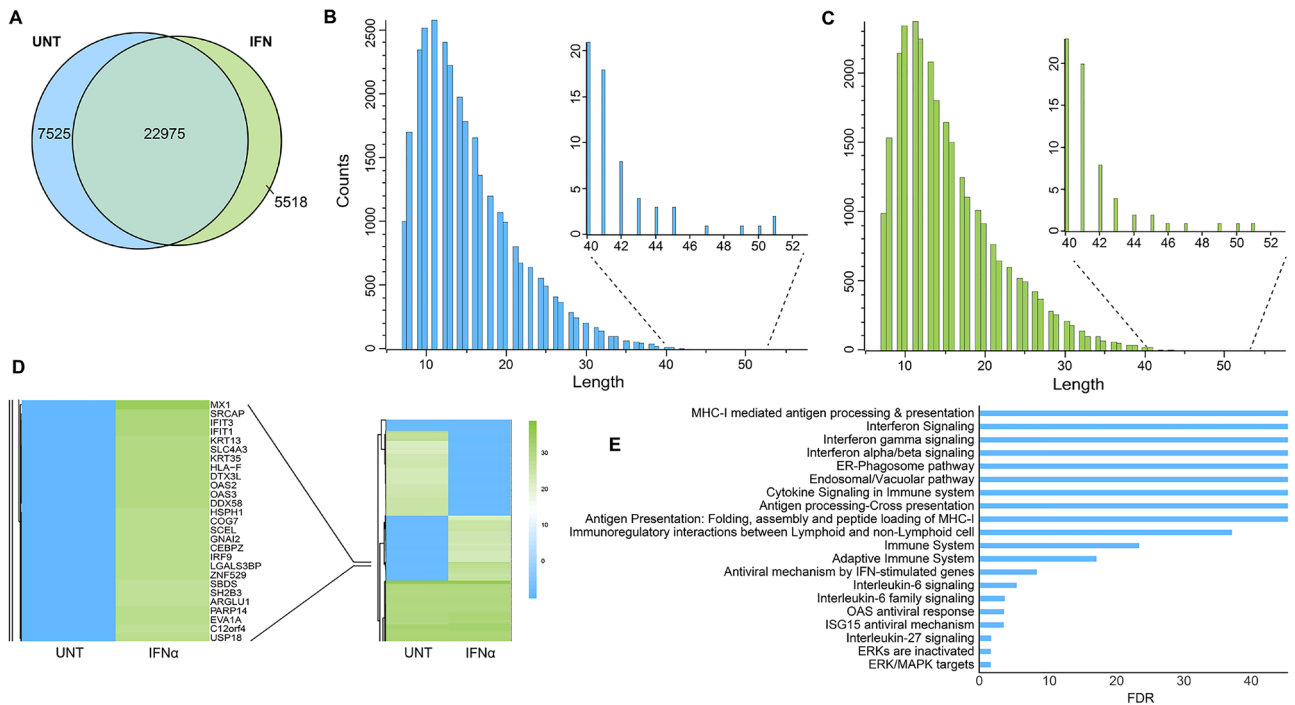
**Figure 1.** Differential protein expression response in Flo-1 cells following IFN $\alpha$  treatment. **(A)** Analysis of protein expression in Flo-1 cells following treatment with 10 ng/ml IFN $\alpha$  for 2, 6, 24, 48, and 72 h by immunoblotting using the indicated antibodies against ISGs. **(B)** Coomassie blue stained SDS-PAGE gel of whole cell extract after cross-linking with DSS for indicated time and concentration. **(C)** Representative immunoblot probed with p53 (DO-1) antibody of the same samples to assess the degree of protein cross-linking.

**Identification of cross-linked interferon induced proteins.** A total of ~30,500 and ~28,500 peptides were detected in the unstimulated and IFN $\alpha$  stimulated samples, respectively (Supplementary Table S1, Fig. 3A) using MaxQuant. Peptide length distribution for both conditions showed a higher proportion of larger peptides that suggests the presence of cross-linked peptides (Fig. 3B,C). Moreover, in the IFN $\alpha$  treated samples, a higher proportion of larger peptides were present in the range of 40–55 (Fig. 3C). Mapping proteins, against log<sub>2</sub> intensities, showed classic interferon-stimulated proteins as the most enriched compared to untreated samples, this included MX1, IFIT1/3, OAS2/3, DDX58, and HLA-F (Fig. 3D). Pathway analysis of proteins that were enriched more than threefold in response to IFN $\alpha$  treatment using Reactome pathway database showed MHC-I mediated antigen presentation and processing as the most dominant pathway (Fig. 3E). Consistent with earlier reports, OAS and ISG15 mediated antiviral response as well as IFN $\alpha$ / $\beta$  and cytokine signalling were among the upregulated pathways. Further, lysine and serine specific cross-links of proteins were identified from the initially obtained MS/MS spectra using SIM-XL. A recent study has reported 104 ISGs by conducting a meta-analysis of single ISG overexpression studies performed in 5 cell types, covering 20 viruses from 9 virus classes<sup>9</sup>. However, to overcome the computational limitation of screening a big dataset, we started with a smaller dataset and explored possible interactions between the IRDS gene list reported in Padariya et al.<sup>28</sup> out of which, the majority are ISGs.

**Identification of a novel interferon-stimulated protein network based on in-situ cross-link.** Interferon-mediated stimulation of ISGs is well documented, but at the molecular level how these proteins culminate in a wide range of biological functions is poorly understood. We looked at the high-confidence protein interactions between known ISGs. Interestingly, we identified a network involving MX1, USP18, ROBO1, OAS3, and STAT1 proteins that form a large complex in response to IFN $\alpha$  treatment (Fig. 4, Table S2)<sup>32–34</sup>. Most importantly, these interactions were detected in all three IFN $\alpha$  treated replicates, and were undetectable in untreated samples suggesting they form specifically in response to IFN $\alpha$  treatment. STAT1 is known to transcriptionally regulate expression of these ISGs, however, its interaction with ISGs at the protein level hasn't been studied. The crystal structure of STAT1 reveals that its coiled-coil domain (CCD) is not involved

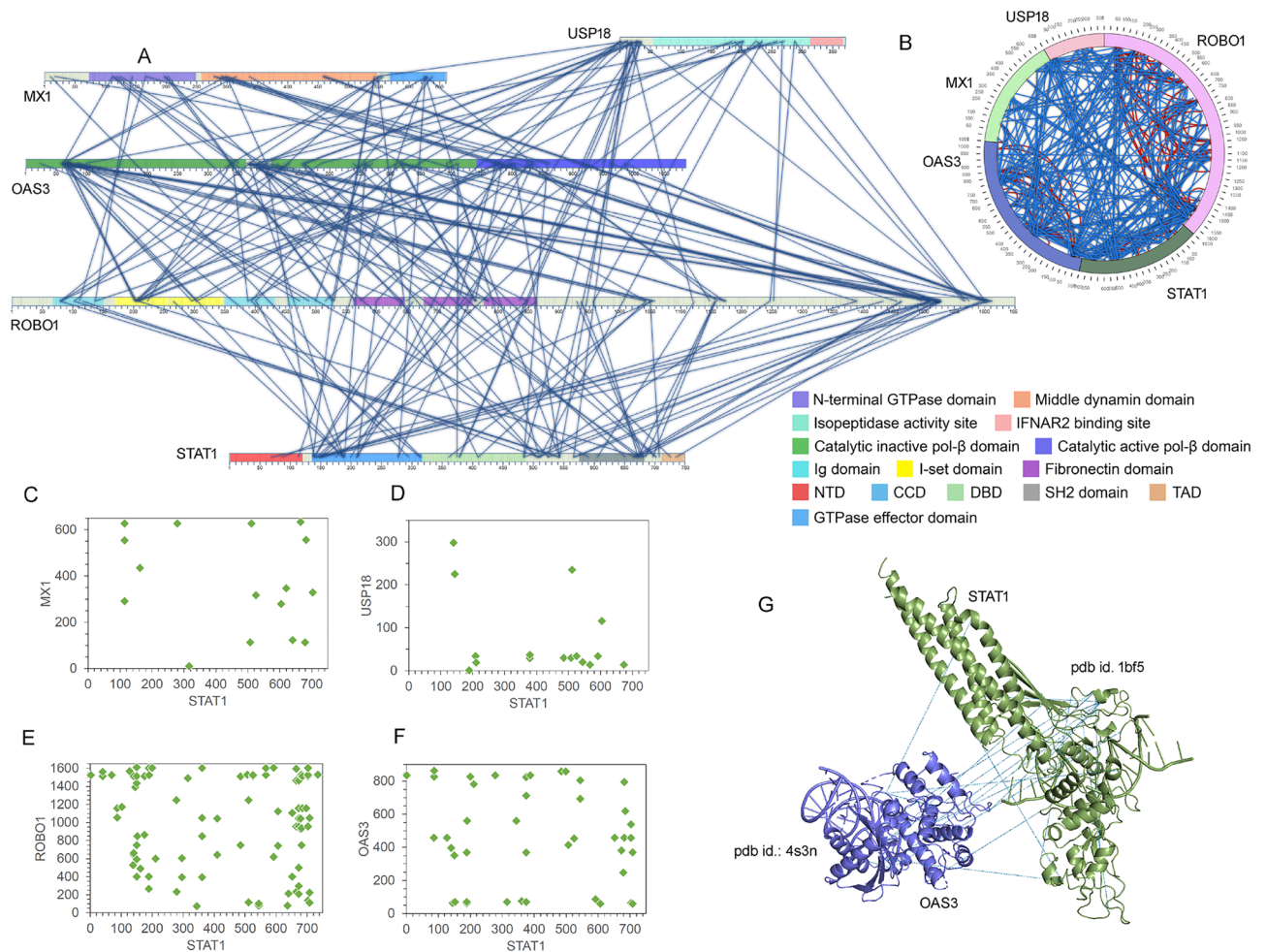


**Figure 2.** Schematic overview of the CLMS approach. Flo-1 cells were treated with 10 ng/ml of IFN $\alpha$  for 24 h and the proteins were cross-linked in-situ using DSS followed by cell lysis and trypsin digestion. The cross-linked samples were analyzed using an Orbitrap mass spectrometer and peptide precursors were further selected for fragmentation during LC-MS/MS. Two linked peptides were identified from the obtained spectra using the Spectrum Identification Machine for Cross-Linked Peptides (SIM-XL) program and all the linkages were merged into a complex network using computational pipelines. Low confidence interactions were filtered out based on false discovery rate (FDR) estimation. A few of the novel high-confidence protein–protein interactions were further validated using co-immunoprecipitation, and the conformational changes of the complex were studied using molecular dynamics (MD) simulation.



**Figure 3.** Identification of differentially expressed cross-linked proteins in response to IFN $\alpha$  (data retrieved from MaxQuant). **(A)** Venn diagram representing common and exclusive peptide counts identified in the IFN $\alpha$  treated and untreated Flo-1 samples. Peptide length distribution from the untreated **(B)** and IFN $\alpha$  **(C)** treated cross-linked samples. **(D)** Heatmap representing log<sub>2</sub>(LFQ intensities) between untreated and IFN $\alpha$ 14 treated Flo-1 cells. The left panel depicts the most upregulated proteins in presence of IFN $\alpha$ . **(E)** A bar plot representing top 20 enriched pathways following IFN $\alpha$  treatment. IFN $\alpha$  responsive upregulated proteins over fourfold change were analyzed by the Reactome pathway database.





**Figure 4.** Protein–protein intermolecular network of ISGs identified in the IFN $\alpha$  treated cross-linked cells. **(A)** 2D interactive map (generated in SIM-XL program<sup>29</sup>) showing the protein–protein interactions, the lines represent intermolecular interactions (cross-link score cut-off was set to 3.5). Domains of different identities are labeled with their respective colors<sup>32</sup>: MX1 domains; Dynamin\_N (73–249), Dynamin\_M (259–547), and GED (569–660). OAS3 domains; OAS1\_C (160–344), OAS1\_C (559–745), NTP\_transf\_2 (780–872), and OAS1\_C (903–108). ROBO1 domains; Ig\_3 (67–151), I-set (170–258), I-set (262–347), Ig\_3 (350–432), Ig\_3 (454–529), fn3 (562–646), fn3 (678–758), and fn3 (777–864). STAT1 domains; STAT\_int (2–120), STAT\_alpha (143–309), STAT\_bind (321–458), SH2 (573–657), and STAT1\_TAZ2bind (715–739). **(B)** The circular viewer of the cross-linked proteins (MX1, USP18, OAS3, ROBO1, and STAT1), with identified inter- and intra-interactions colored in blue and red, respectively. The cross-link score cut-off was set to 3.5. Dot plots representing the interacting sites of STAT1 with MX1 **(C)**, USP18 **(D)**, ROBO1 **(E)** and OAS3 **(F)**, with interaction site K or S between two peptides. In the plots, the cross-link score cut-off was set to 3.0. **(G)** Different interactions sites between STAT1 and DI domain of OAS3 overlaid over their protein structure in the PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.); STAT1 (pdb id.: 1bf5<sup>33</sup>) and OAS3 (pdb id.: 4s3n<sup>34</sup>) program.

in interactions with DNA or with the protomer when it forms dimer<sup>35</sup>. These  $\alpha$ -helices form a coiled-coil structure that provides a predominantly hydrophilic surface area for interactions to take place<sup>35</sup>. In our CLMS data, we observed that most of the interactions with STAT1 are either in the CCD, linker domain or the SH2 domain prior to the C-terminal tail segment (residues 700–708) (Fig. 4A). A previous study reported that USP18 bound to the CCD and DNA binding domains (DBD) of STAT2 and was recruited to the type I IFN receptor subunit IFNAR2 to mediate suppression of type-I IFN signaling<sup>24</sup>. Our data also indicate that the catalytic domain of USP18 interacts with DBD of STAT1 (Fig. 4A,D) suggesting that STAT1 and STAT2 both may have a role in recruiting USP18 to IFNAR2.

There are two USP18 isoforms described in humans, full-length protein, which is mainly located in the nucleus, and an isoform lacking the N-terminal domain, USP18-sf that is evenly distributed in the cytoplasm and nucleus<sup>36</sup>. In addition, the N-terminus has been predicted to be unstructured and is not required for isopeptidase activity or ISG15 binding<sup>37</sup>. Most of the interactions identified in our study are situated in the N-terminus of the protein which suggests that these interactions involve full-length USP18 (Fig. 4A,D), and therefore, have a high probability of occurring in the nucleus. Moreover, our data also implies that the N-terminus is used exclusively for protein–protein interactions. The IFNAR2 binding site is located between residues 312–368 and it is interesting

to note that none of the proteins in the complex bind to this region (Fig. 4A)<sup>37,38</sup>. Together the data suggest that the IFNAR2 binding region is used exclusively by the receptor protein. Additionally, only OAS3 and ROBO1 were found to be associated with both the N-terminus and the domain before the IFNAR2 binding site (Fig. 4A).

ROBO1 belongs to the immunoglobulin (Ig) superfamily of transmembrane signalling molecules and consists of five Ig and three fibronectin (Fn) domains in the extracellular region. These extracellular domains are followed by a membrane proximal region and a single transmembrane helix<sup>39</sup>. An unstructured intracellular region lies at the C-terminus, containing conserved sequence motifs that mediate the binding of effector proteins<sup>39</sup>. The region stretching from amino acids ~ 1100 to 1600 is mostly disordered. We found that MX1 interacted through ROBO1 via Ig, Fn, and intracellular domains while most of the interactions with STAT1 were between its CCD, linker domain, and the C-terminal of ROBO1 (Fig. 4A,E). On the other hand, interaction with DI, DIII, and the linker region of OAS3 was dispersed throughout the ROBO1 protein (Fig. 4A).

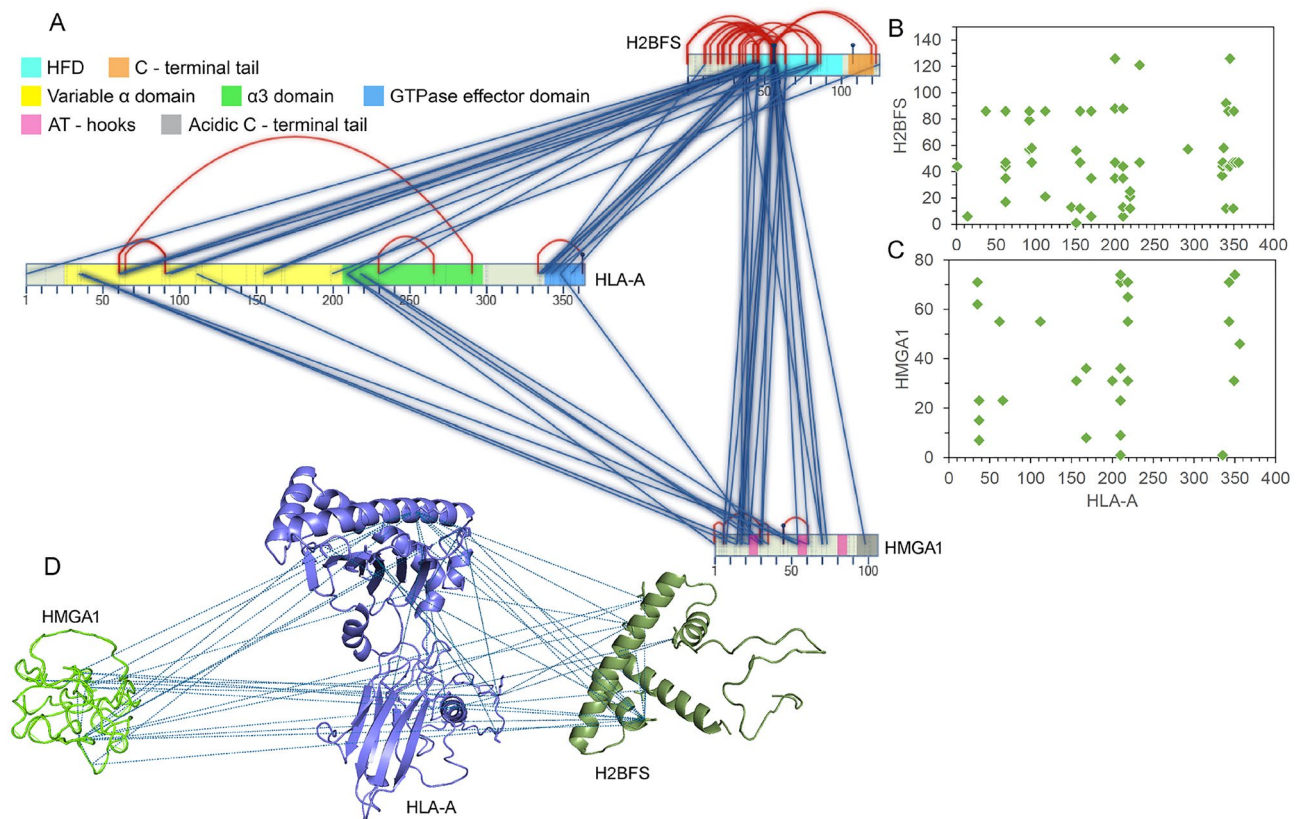
The oligoadenylate synthetase (OAS) family of proteins sense and bind to intracellular double-stranded RNA (dsRNA), undergo conformational change, and synthesize 2',5'-linked oligoadenylates (2–5 As)<sup>40</sup>. Out of the three OASs, OAS3 has been found to display higher affinity for dsRNA and to synthesize minimal 2–5 As which can activate RNase L, and thereby, restrict viral replication<sup>41</sup>. The OAS family consists of polymerase beta (pol-β)-like nucleotidyl transferase domains. A previous study showed that the catalytic activity of the C-terminal domain (DIII) is dependent on the dsRNA-binding domain (DI) that is essential for activation of OAS3<sup>42</sup>. We observed that the DI and DII domain of OAS3 interacted with the CCD and a small linker region between SH2 and TAD of STAT1 (Fig. 4A,F). Overlay of different cross-linked sites over the protein structures shows interaction between the β-sheets and loops of the STAT1 DBD with the exposed pocket or cavity formed by residues 60–75 in the DI domain of OAS3 (Fig. 4G). Orientation of the protein in the complex also showed that none of the interactions with OAS3 interfered with the DNA binding ability of its DI domain (Fig. S1A). In addition, the N-terminal GTPase domain of MX1 interacts extensively with both DI and DIII domains of OAS3 (Fig. 4A). We also observed an interaction between OAS1 and MX1 in all three IFNα treated replicates, where the only domain of OAS1, which is also catalytically active, interacts with all the three domains of MX1 (Fig. S2A,B).

MX proteins are part of large dynamin-like GTPase family that contains a N-terminal GTPase domain that binds and hydrolyses GTP, a self-assembly-mediating middle domain, and a C-terminal leucine zipper (LZ) domain which acts as a GTPase effector domain<sup>25,43</sup>. MX1 associates with subunits of the viral polymerase to block viral gene transcription<sup>43</sup>. A previously reported yeast two-hybrid screen revealed that MX1 bound to PIAS1 inhibits STAT1-mediated gene activation by blocking the DNA binding activity, and also has SUMO E3-ligase activity<sup>44,45</sup>. Here we demonstrate that MX1 binds to STAT1 (Fig. 4C,D), however, how this interaction affects STAT1-mediated gene activation in response to IFNα needs further investigation. Additionally, we also found MX1 interacting with IFIT3 and DDX60 in all three IFNα treated replicates (Fig. S2C).

DDX60, an IFN-inducible cytoplasmic helicase, has been previously reported playing a role in RIG-I-independent viral RNA degradation<sup>46</sup>. It interacts with RIG-I and activates its signalling in a ligand-specific manner<sup>46</sup>. DDX60 consists of a DEXD/H-Box helicase domain which binds to viral RNA and DNA, and a C-terminal helicase domain<sup>47</sup>. Most of its interactions with MX1 and IFIT3 were within the long N- and C-terminal regions with no typical domains or motifs (Fig. S2E,F). However, MX1 was also linked with the DEXD/H-Box helicase domain (Fig. S2E). IFIT family proteins have distinguished tandem copies of helix-turn-helix motifs called tetratricopeptide repeats (TPRs). IFIT3 was found to be a positive modulator of RIG-I signaling, and hence, a component of MAVS complex<sup>48</sup>. Together, our data suggests that IFIT3 and DDX60 interact with each other mainly in the region between TPR 3–6 of IFIT3, and may have a role in RIG-I/MAVS signalling (Fig. S2F).

**Antigen presenting MHC-I molecule and its interaction network.** Considering proteome-wide screening is computationally demanding, we next screened the entire human UniProt database for one of the IFNα treated replicates. We found some high-confidence interaction networks for HLA-A in that replicate. The pathway analysis for identified proteins from MS/MS spectra revealed MHC-I based antigen processing and presentation as the dominant pathway induced by interferon (Fig. 3D). Therefore, we focused on exploring high-confidence protein interactions of MHC-I molecules across all the cross-linked samples. HLA consists of α1, α2, and α3 domains and a light chain, β2 microglobulin (β2m) being a constant protein partner<sup>49</sup>. HLA is unstable in the absence of peptide ligand following its assembly in the endoplasmic reticulum<sup>50</sup>. The peptide binding groove is formed by α1 and α2 domains that are highly polymorphic and unstructured in the peptide-free form, and a α3 domain that is comparatively less polymorphic<sup>51</sup>. In the presence of IFNα, we found two HLA-A complexes: one that interacts with HMGA1 and H2B (Fig. 5, Table S3) and another where it interacts with MDN1, LRCH4 along with H2B (Fig. 6).

Apart from maintaining genome integrity, histone H2B is involved in the transcriptional regulation. H2B protein is composed of a central histone-fold domain (HFD) formed by three alpha helices separated by loops, and a C-terminal tails<sup>41,52</sup>. The majority of interactions with H2B were in the α1-helix that mediates trimerization with the HFD heterodimer (Fig. 5A,B). Even though lysines are involved in DNA-binding, some of the lysines are also sites for alternative acetylation or methylation. For instance, the residues K43, K46 and K57 from the H2B are not involved in direct DNA-binding but are targets of different post-transcriptional modifications<sup>53</sup>. Similarly, K44, K47 and K57 residues in H2B may have an alternative role in presence of IFNα, which includes interacting with other proteins (Fig. 5A,B). Moreover, extrachromosomal histone H2B activates immune responses in various cell types acting as a cytosolic sensor to detect double-stranded DNA (dsDNA) fragments derived from infectious agents or damaged cells<sup>54</sup>. Depletion of H2B suppressed IFN-β production and STAT1 phosphorylation in presence of DNA viruses<sup>54</sup>. H2B is also known to travel in and out of the nucleus more rapidly than other core histones<sup>54</sup>. H2B interaction with MDN1 and with LRCH4 were also observed in individual untreated samples.



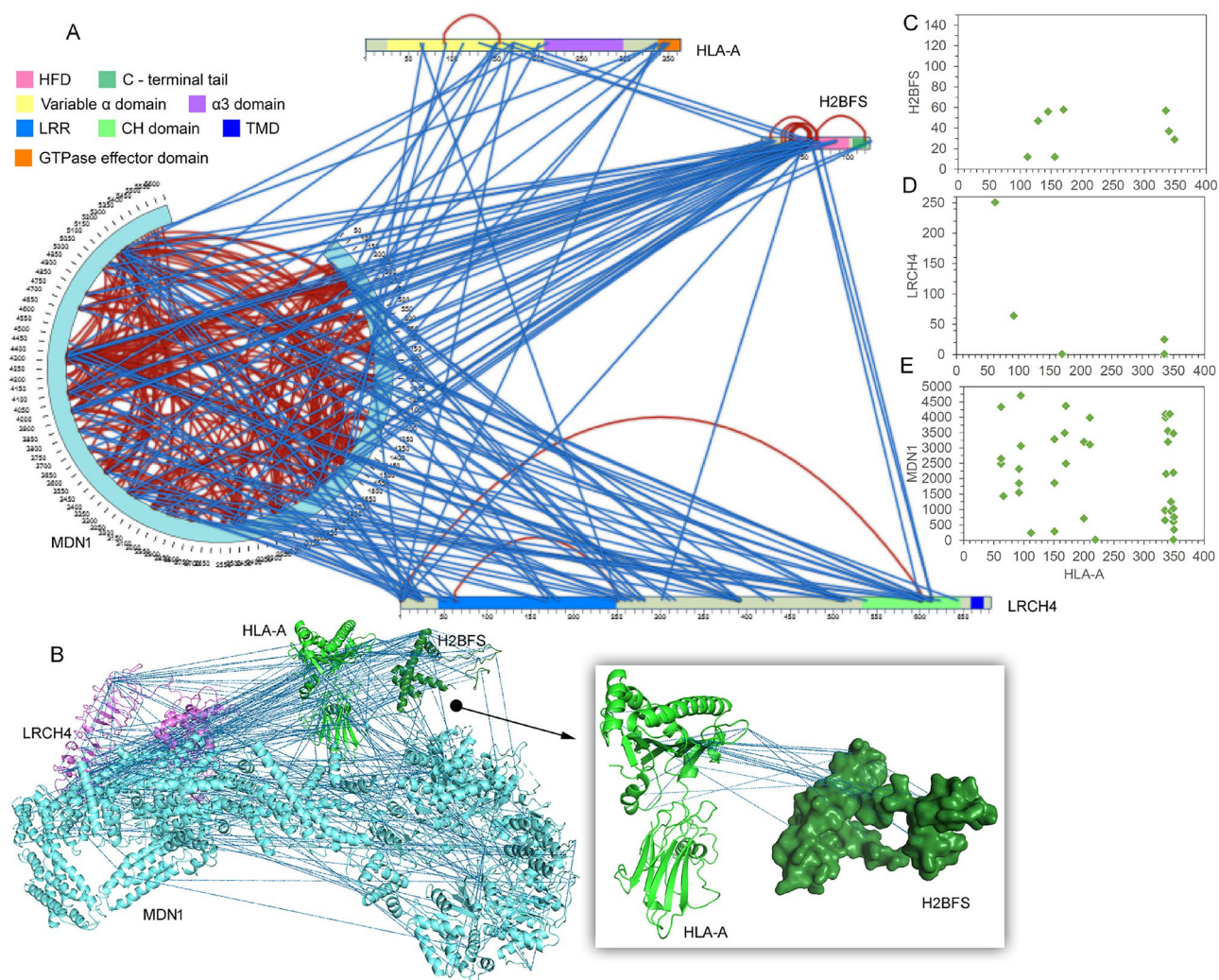
**Figure 5.** The IFN $\alpha$  induced interaction network of HLA-A with H2B (H2BFS) and HMGA1. **(A)** 2D map (generated in SIM-XL program<sup>29</sup>) depicting different types of interaction within the H2B-HLA-A-HMGA1 complex; inter-link (blue), intra-link (red) and mono-link (black). Domains of different identities are labeled with their respective colors<sup>32</sup>: H2B (histone; 2–102) and MHC-I (MHC\_1; 25–203, C1-set; 210–290, and MHC\_I\_C; 337–364). The cross-link score cut-off was set to 3.5. Dot plots representing interaction sites of HLA-A with H2B **(B)** and HMGA1 **(C)**, with interaction site K or S between two peptides. In the plots, the cross-link score cut-off was set to 3.0. **(D)** Inter-links between proteins displayed over the H2B, HLA-A, and HMGA1 protein structures in the PyMOL program. These structures were modeled using the Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2>), and the template structures used for the proteins H2B, HLA-A, and HMGA1 were 1kx5<sup>52</sup>, 1kj3<sup>49</sup>, and 2eze<sup>55</sup>, respectively.

We found HLA-A interacting with H2B in all three IFN $\alpha$  treated samples and one untreated replicate. This data reflects the role of H2B in alternative physiological functions, independent of transcriptional regulation.

HMGA1 (High Mobility Group AT-Hook 1), a small nuclear protein enriched in disorder-promoting amino acids, was identified in complex with HLA-A. It has an acidic C-terminal tail and three differentially spaced DBDs, called AT-hooks as they bind to minor grooves of AT-rich regions in dsDNA<sup>55,56</sup>. This binding induces bending or straightening of DNA allowing the access of canonical transcription factors to their consensus sequences. The C-terminal tail is assumed to be involved in protein–protein interaction and recruitment of transcription factors since a C-terminal deletion mutant fails to initiate transcription<sup>57</sup>. Moreover, this domain contains several conserved phosphorylation sites that are known kinase substrates<sup>58</sup>. We observed the interactions of HLA-A and H2B with HMGA1 outside the C-terminal domain, suggesting the C-terminal domain is mainly used for transcription factor binding (Fig. 5A,C). HMGA proteins compete with histone H1 for binding to linker DNA, thereby resulting in increased accessibility<sup>57</sup>. Similarly, it is plausible that HMGA interacts with histone H2B along linker DNA while competing with histone H1. HMGB1 induces HLA-A, -B, and -C expression in dendritic cells leading to their activation<sup>59</sup>, nevertheless, an interaction between HMGs and HLAs hasn't been reported previously. We found that HMGA1 interacts with the  $\alpha 1$  and  $\alpha 3$  domains of HLA-A with most of the interaction outside its 3 DBDs (Fig. 5A,C). In our hands, HLA-A was found to be localized in the nucleus (data not shown), and given that H2B and HMGA1 also reside in the nucleus, there is a high probability that this interaction occurs in the nucleus. The specific adducts measured between H2B, HLA-A, and HMGA1 are described in Fig. 5D.

The majority of HLA-A interactions with other proteins are localized at its  $\alpha 1$  and  $\alpha 2$  domains as well as within the disordered C-terminal domain (Fig. 6). In one of these examples, we found that HLA-A interacts with the disordered N-terminal tail of LRCH4 (Fig. 6A,D). LRCH4 regulates TLR4 activation and cytokine induction by LPS, and therefore, regulates the innate immune responses<sup>60,61</sup>. It is a membrane protein with nine Leucine-rich repeats (LRRs) and calponin homology (CH) motif in its ectodomain, followed by a transmembrane domain (TMD)<sup>60,62</sup>. The CH domain is reported to mediate protein–protein interactions<sup>60</sup>. A stretch of around 300

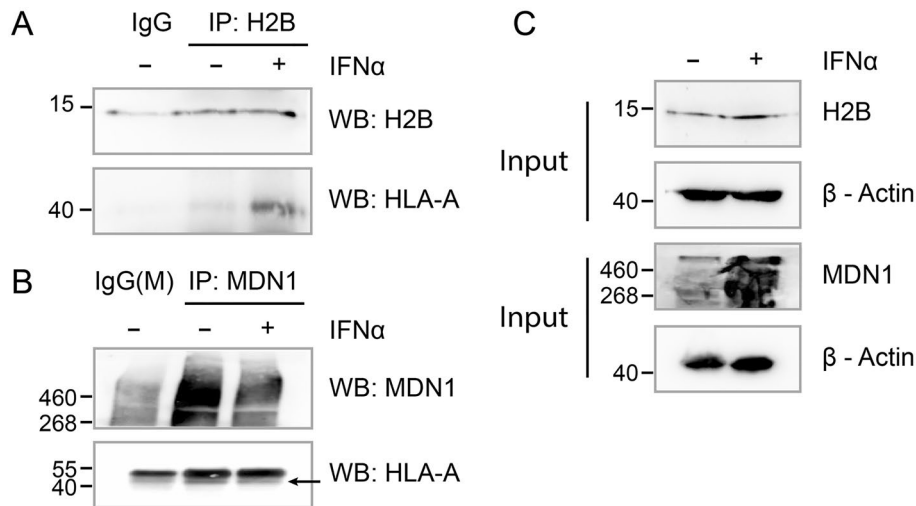




**Figure 6.** The IFN $\alpha$  induced interaction network of HLA-A with H2B (H2BFS), MDN1, and LRCH4. **(A)** The intramolecular (red) and intermolecular (blue) cross-links represented in the 2D interactive map (generated in SIM-XL program<sup>29</sup>), MDN1 is presented in the circular view. The cross-link score cut-off was set to 3.5. Domains of different identities are labeled with their respective colors<sup>32</sup>: H2B (histone; 2–102), MHC-I (MHC\_1; 25–203, C1-set; 210–290, and MHC\_I\_C; 337–364), and LRCH4 (LRR\_8 (68–126), LRR\_8 (137–194), and CH (535–641)). **(B)** Inter-link between proteins displayed over the H2B, HLA-A, LRCH4, and MDN1 protein structures in the PyMOL program. These structures were modeled using the Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2>), and the template structures for the proteins H2B, HLA-A, LRCH4, and MDN1 were 1kx5<sup>52</sup>, 1kj3<sup>49</sup>, 6hlu<sup>62</sup>, and 6i26<sup>65</sup>, respectively. Dot plots showing the interaction sites K or S of HLA-A with H2BFS **(C)**, LRCH4 **(D)**, and MDN1 **(E)**. For the plots, the cross-link score cut-off was set to 3.0.

amino acids between the LRR and CH domain is relatively accessible but disordered. In line with the function of disordered regions as mediators of protein–protein networks and vesicle trafficking<sup>63</sup>, we found most of the protein interactions in the disordered region. Interaction with MDN1 was dispersed throughout the protein's length including LRR1, LRR6, CH domain, and disordered region while H2B was bound mostly with CH domain (Fig. 6A,B). It is noteworthy that none of the interactions involved the TMD which shows the specificity of the CLMS method (Fig. 6A,B).

MDN1 was also identified as a part of the HLA-A protein network (Fig. 6A). It belongs to the AAA protein family (ATPase associated with various activities). Its identical N-terminal AAA domains orchestrate into hexameric rings and remove assembly factors from ribosomal 60S subunit<sup>64</sup>. Cryo-EM studies in yeast reveal that AAA domains are followed by six non-equivalent AAA domains linked in a single polypeptide which appears similar to dynein<sup>64–66</sup>. Further, a stretch of Asp/Glu-rich region is followed by a MIDAS (metal ion-dependent adhesion site) domain. Owing to MDN1's large size (~5600 amino acids) and its limited homology to well-studied proteins, not a lot is known about its structure and function in humans. We identified HLA-A, H2B, and LRCH4 as binding partners of MDN1 and their orientation as a protein complex was revealed in PyMol (Fig. 6A,B). These three proteins interact with AAA domains, dynein-like linker domain and probable MIDAS domain of MDN1. In a previous report, affinity purification of the bait proteins identified MDN1 as the protein



**Figure 7.** HLA-A co-purifies with H2B and MDN1. Representative immunoblotting of endogenous H2B (A) and MDN1 (B) immunoprecipitated from IFN $\alpha$  treated Flo-1 cells and probed with the indicated antibodies. Mouse and rabbit IgGs were used as negative control. (C) The relative amount (Input) of the different antigens were depicted by immunoblots probed against indicated antibodies and  $\beta$ -actin was used as a loading control.

associated with histone H2B<sup>67</sup>. Moreover, a recent study has also reported interaction between MDN and HLA-B using affinity-purification mass spectrometry in HCT116 cells which supports our findings<sup>68</sup>. Identification of this complex in IFN $\alpha$  treated samples implies that MDN1 has a role in interferon signalling.

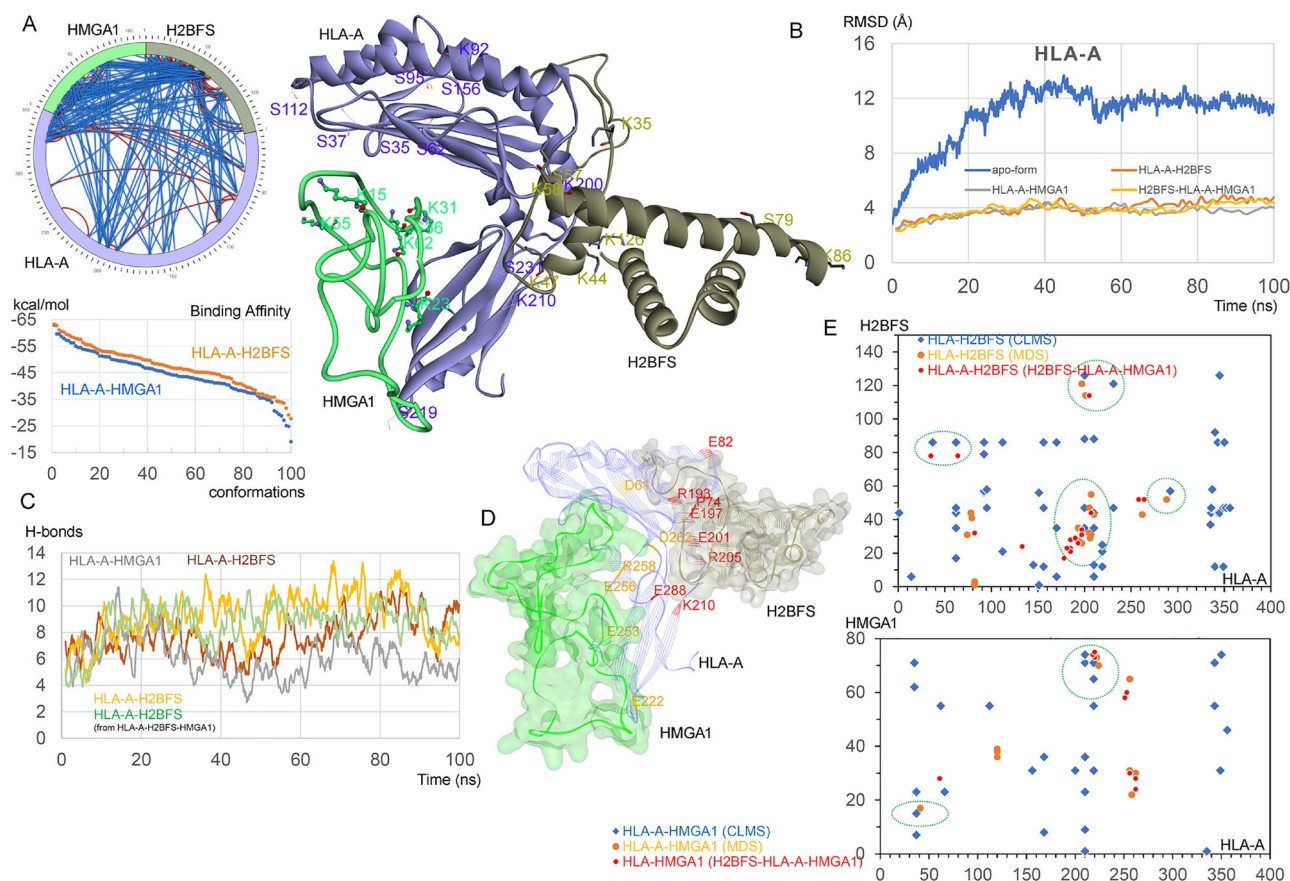
As HLA genes are highly polymorphic, we extracted sequencing reads mapping to HLA-A, -B and -C from the RNA-seq data of the Flo-1 cells (data not shown). Peptide sequences corresponding to sequencing reads showed a significant difference among HLA-A, -B and -C in the regions where the cross-linked peptides reside in HLA-A (Fig. S3). Moreover, we did not observe protein-protein cross-links for the HLA-B/C molecules with either of the H2B/HMGA1/MDN1/LRCH4 proteins. This suggests that the protein interaction found between HLA-A, MDN1, LRCH1, and HMGA1 is specific to HLA-A. Additionally, the proteomics analysis of non-cross-linked samples (Table S4) suggests that HLA-A is enriched with higher sequence coverage compared to that of the HLA-B or HLA-C. The peptides identified for HLA-A have high intensities in both IFN $\alpha$  treated and untreated samples.

**Validation of novel HLA-A binding proteins.** To ensure that the interactions identified here are not due to non-specific cross-linking of two proteins in close spatial proximity, we further validated two novel interactors of HLA-A by performing co-immunoprecipitation assay. Interactions of HLA-A with endogenous MDN1 and H2B were detected in IFN $\alpha$  treated and untreated Flo-1 cells (Fig. 7, Fig. S4). We confirmed that HLA-A was captured with H2B in immunoprecipitates, and this association was induced by IFN $\alpha$  treatment as HLA-A was absent in immunoprecipitated samples from untreated cells (Fig. 7A). However, our data demonstrate that IFN $\alpha$  differentially regulates HLA-A binding to H2B and MDN1. IFN $\alpha$  induced association between H2B and HLA-A but decreased its binding with MDN1. We found that MDN1 associates with HLA-A in control samples, while the addition of IFN $\alpha$  reduces this interaction irrespective of MDN1 induction by IFN $\alpha$  (Fig. 7B,C). Moreover, immunoprecipitation of HLA-A captured H2B in A549 cells (Fig. S4) suggesting this interaction is not cell type-dependent. Together, these results confirm interferon-mediated interaction of HLA-A with H2B and MDN1.

**The conformational dynamics of the HLA-A binding proteins.** Structural properties of one of the high-confidence cross-linked networks induced by interferon—H2B-HLA-A-HMGA1 was investigated. We took advantage of molecular dynamics simulation as an alternative approach to understand the conformational dynamics of the proteins involved in this complex (Fig. 8). Findings from CLMS data suggest the probability of different conformations between H2B, HLA-A, and HMGA1 proteins. Therefore, the following potential complexes were simulated in the solvent environment: H2B-HLA-A, HMGA1-HLA-A, and H2B-HLA-A-HMGA1. Initial protein-protein docking screening using the MOE package (Molecular Operating Environment; Chemical Computing Group Inc., Montreal, QC, Canada) proposed different possible conformations between these proteins (Fig. 8A). Visualization of docked protein complexes revealed several interactions and possible conformations (Figs. 5A, 8). As such, one of the possible conformations is represented in Fig. 8A (with labeled cross-links), which was further evaluated with the MD simulation pipeline. In addition, the binding energies of H2B or HMGA1 with HLA-A highlight that H2B has higher affinity with HLA-A (Fig. 8A).

Stability of the HLA-A molecule over time (root-mean-square deviation; RMSDs or root-mean-square fluctuations; RMSFs) suggests that the presence of H2B or HMGA1 protein in the complex stabilizes HLA-A (Fig. 8B, Fig. S5). The HMGA1 protein binding closely to the B2M site of HLA-A, induces stability in the HLA-A amino acids in both the complexes the HLA-A-HMGA1 or H2B-HLA-A-HMGA1 (Fig. 8B, Fig. S5). Particularly, HLA-A residues ~60–90 and ~180–210 were found exhibiting lesser flexibility in the presence of H2B (Fig. 8B). Both





**Figure 8.** Conformational dynamics of the possible network between H2B (H2BFS)-HLA-A, HMGA1-HLA-A and H2B-HLA-A-HMGA1 complexes. **(A)** The left panel represents a 2D map (generated in SIM-XL program<sup>29</sup>) of intramolecular (red) and intermolecular (blue) cross-links (cross-link score cut-off was set to 3.5). In addition, the identified cross-link residues were labeled on the H2B, HLA-A, and HMGA1 protein structures. The binding conformations of these proteins were retrieved by the docking pipeline implemented in the MOE package. The bottom left panel shows different possible conformations of the H2B-HLA-A and HMGA1-HLA-A complexes, varying in their protein–protein binding affinity (GBVI/WSA dG; kcal/mol). **(B)** The root-mean-square deviation (RMSDs) of atomic positions (excluding hydrogen atoms) of each protein structure. **(C)** Intermolecular protein–protein hydrogen-bond interactions from different simulated complexes, a particular interaction lasting  $\geq 10$  ns were considered. The h-bond donor–acceptor atom cutoff distance was set to 3.5 Å, and the donor–H–acceptor angle cutoff was set to  $\geq 160^\circ$ – $180^\circ$ . **(D)** Labeled residues forming protein–protein interactions from HLA-A with respective partners with occupancy  $\geq 20$  ns, retrieved from the HLA-A-H2B and HLA-A-HMGA1 simulated complex. The protein structure represents averaged structures from 100 ns MDS. **(E)** Differently identified cross-links between the HLA-A-H2B and HLA-A-HMGA1 complexes based on the interaction site K or S between two peptides, compared with the interactions traced from 100 ns MS simulations of the H2B-HLA-A/HMGA1-HLA-A/H2B-HLA-A-HMGA1 complexes. The cross-link score cut-off was set to 3.0, and particular interactions from MDS having occupancy  $\geq 10$  ns were considered. Protein structures were visualized using the BIOVIA Discovery Studio (Dassault Systèmes, BIOVIA Corp., San Diego, CA, USA) and Molecular Operating Environment (MOE; Chemical Computing Group Inc., Montreal, QC, Canada) packages.

H2B and HMGA1 display better binding with the HLA-A in the H2B-HLA-A-HMGA1 complex, compared to HLA-A binding to H2B or HMGA1 alone (Fig. 8C,D; Table S5). Residues involved in hydrogen bond formation (MD simulation high occupancy  $\geq 10$  ns) coincide with CLMS cross-link (K or S residues) interaction sites in the complex which suggests high confidence in interactions identified through the CLMS method (Fig. 8E). HLA-A residues between  $\sim 190$ – $210$  and  $\sim 200$ – $220$  amino acids were found binding H2B and HMGA1, respectively, in the CLMS and MD simulation (Fig. 8E).

## Discussion

The protein–protein interactions form into dynamic structural networks that mediate intracellular communication in response to some stimulus. Because many proteomics approaches define changes in total steady-state protein levels, protein–protein interaction dynamics need additional tools to capture binding interfaces, CLMS is one such tool. The interferon signaling system constitutes a cytokine network that allows cells to respond to a range of environmental pathogenic and internal pathological signals, eventually leading to induction of a subset

of interferon-induced proteins. We applied CLMS to determine whether novel protein–protein interactions can be defined within the set of interferon induced proteins. Global protein cross-linking analysis in the interferon responsive Flo-1 cell model was used to trap protein complexes. Tryptic peptides were recovered from non-cross-linked and cross-linked cells allowing for peptide counts, pathway enrichment, and distribution of peptide length plus defined LFQ intensities. Classic interferon induced proteins were identified as a positive internal control, whilst novel inter- and intra- molecular cross-linked adducts were observed for canonical interferon induced proteins, such as MX1, UP18, OAS3, and STAT1. Different structural features and interactions residing in functional regions were investigated.

Interactions between HLA-A, MDN1, and H2B were detected with immunoblotting in IFN $\alpha$  treated and untreated Flo-1 and A549 cells. Our findings highlight that HLA-A forms a complex with H2B in an IFN $\alpha$ -dependent manner. Our work presents an interesting avenue to further explore colocalisation of these two complexes. It will also be interesting to deploy the CLMS approach on a panel of cell lines to identify cell type-independent interferon-mediated protein interactions. We finally took advantage of MD simulations as an alternative approach to understanding the conformational dynamics of the proteins involved in the H2BFS-HLA-A-HMGA1 complex where both intra- and inter-molecular cross-link interactions were traced. Findings from CLMS data suggest the probability of different conformations between H2BFS, HLA-A, and HMGA1 proteins. Different possible conformations between these docked protein complexes revealed several interactions similar to those observed in the CLMS datasets. One of the main advantages of our approach is that it can easily identify interactors of highly polymorphic genes such as HLAs, therefore, it would be interesting to explore HLA-haplotype specific protein interactions which are otherwise difficult to study. Together, our data show that CLMS can be used to expand our knowledge on the interferon induced signalling networks and form a foundation for studying more complex cell–cell systems within the tumour microenvironment.

## Methods

**Cell culture and cross-linking reaction using DSS.** Flo-1 cells were obtained from ATCC and maintained in DMEM (Gibco) supplemented with 1% penicillin/streptomycin (Invitrogen), 10% fetal bovine serum (Gibco) and incubated at 37 °C with 5% CO<sub>2</sub>. Cells were grown to 70–80% confluence prior to treatment with IFN $\alpha$ 14 (produced by Edinburgh Protein Production Facility). All other chemicals and reagents were obtained from Sigma Aldrich unless indicated otherwise.

The Flo-1 cells were cultured in a 6-well plate and the next day cells were treated with 10 ng/ml IFN $\alpha$ 14 for 24 h until ~80% confluency. The cells were washed thrice with PBS and were cross-linked for 5 min at 37 °C using freshly prepared DSS (Thermo Fisher Scientific) (dissolved in DMSO) in PBS to achieve 0.5 mM final concentration. The DSS cross-linking reaction was replaced with PBS and the residual DSS was quenched by addition of 20 mM Tris (pH 8.0) in PBS for 15 min at 37 °C. The cells were harvested by scraping and collected in a low-binding tube (Axygen).

**Two steps cross-linked protein solubilization.** Cell pellets were lysed with 300  $\mu$ l urea lysis buffer (8 M Urea, 0.1 M Tris, pH 8.5) at room temperature for 30 min with occasional vortex. All the centrifugation steps were performed at 14,000 $\times$ g at 8 °C. The lysates were centrifuged for 10 min, and the supernatant was transferred into a new tube. The remaining transparent pellet was solubilized with 150  $\mu$ l second lysis buffer (2 M Urea, 2% (w/v) SDS (sodium dodecyl sulphate)) for 30 min or longer until uniformly aqueous solution was obtained. The lysates were centrifuged for 20 min, and the supernatant was pooled with lysate retrieved in the previous step. Protein concentration was assessed using the Micro BCA assay (Thermo Fisher Scientific) according to the manufacturer's instructions for the microplate procedure. Samples were snap frozen in liquid nitrogen and stored at –80 °C.

**Peptide generation using FASP.** Approximately 100  $\mu$ g of soluble cross-linked protein was processed using a modified filter-aided sample preparation (FASP) protocol as described in Wisniewski et al.<sup>69</sup> Briefly, cross-linked proteins were mixed with 200  $\mu$ l of Urea buffer (8 M Urea in 0.1 M Tris pH 8.5), vortexed, and split into two halves. All the centrifugation steps were performed at 14,000 $\times$ g at 25 °C. The first half of cross-linked protein lysate was transferred to a 10 kDa cut-off Microcon Centrifugal Filter Unit with Ultracel-10 membrane (Merck) followed by filter centrifugation for 25 min. Next, the second half of protein was added to the filter, and the same procedure was repeated. Protein reduction was performed by adding 100  $\mu$ l of 17 mM Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP) in the urea buffer. The reduction reaction was held on a thermomixer at 37 °C for 30 min with agitation at 600 rpm. Further, the column was centrifuged and the reduced cross-linked proteins were alkylated using 100  $\mu$ l of 50 mM iodoacetamide in the urea buffer. Alkylation reaction was carried out in the dark for 20 min at RT. The column was centrifuged, and the column walls were rinsed three times with 100  $\mu$ l of urea buffer, followed by centrifugation. The same procedure was performed three times with 100  $\mu$ l of 100 mM ammonium bicarbonate. The collection tube was replaced with a new one prior to tryptic digestion. Digestion buffer comprising 50 mM ammonium bicarbonate was added along with 1  $\mu$ l of trypsin diluted in the trypsin buffer (Promega). Trypsin to protein ratio was kept at approximately 1:33, and the digestion reaction was incubated overnight at 37 °C in a humidified chamber. Cross-linked peptides were eluted from the filter by centrifugation for 25 min. Peptide recovery was enhanced by adding 50  $\mu$ l of 0.5 M NaCl to the filter, followed by centrifugation for 25 min.

**Peptide desalting.** C18 Micro Spin Columns (Harvard Apparatus) were used to desalt cross-linked tryptic peptides using a protocol described in Bouchal et al.<sup>70</sup> with minor changes. Briefly, C18 spin columns were activated by three washes of 0.1% formic acid (FA) in acetonitrile (AcN) (Merck) and two washes of 0.1% FA.



The column was hydrated for 15 min by 0.1% FA. The sample was loaded into the spin column and washed three times with 0.1% FA. Desalted peptides were eluted by a step gradient using 50%, 80%, and 100% AcN in 0.1% FA consecutively. Samples were dried in SpeedVac Concentrator plus (Eppendorf) until no residual liquid was present. Eluted peptides were dissolved in 100  $\mu$ l of 0.08% trifluoroacetic acid in 2.5% AcN and concentration was measured on NanoDrop 2000 (Thermo Scientific). Approximately 1  $\mu$ g of cross-linked peptides from each sample was injected into the LC-MS/MS system.

**Peptide separation and data-dependent mass spectrometry acquisition.** Cross-linked peptides were separated on UltiMate 3000 RSLCnano LC System (Thermo Scientific) connected to Orbitrap Exploris 480 Mass Spectrometer (Thermo Scientific). Cross-linked peptides were concentrated on  $\mu$ -precolumn C18 trap cartridges 300  $\mu$ m inner diameter (ID) and 5 mm length packed with C18 PepMap100 sorbent with PepMap 5  $\mu$ m sorbent (Thermo Scientific). Loading pump flow was set to 5  $\mu$ l/min of 0.08% trifluoroacetic acid in 2.5% AcN. Cross-linked peptides were separated on 75  $\mu$ m ID and 150 mm long fused-silica analytical column packed with PepMap 2  $\mu$ m sorbent (Thermo Scientific). The mobile phase A and B were composed of 0.1% FA in water and 0.1% FA in AcN respectively. The gradient started at 2.5% B linearly increasing up to 40% B in 90 min, followed by a linear increase up to 90% B in the next 2 min. Mobile phase composition remained at 90% B for 10 min, followed by a linear decrease to 2.5% B in 2 min. Column equilibration before the next run was performed at 2.5% B for 8 min. Cross-linked peptides eluted from the analytical column were ionized in a nano-electrospray ion source (NSI) and were introduced into Exploris 480 mass spectrometer (Thermo Scientific).

The Orbitrap Exploris 480 mass spectrometer was operated in positive data-dependent mode. Full-scan was operated in profile mode with 120,000 resolution, and the range was set from m/z 350 Th to m/z 2000 Th. The normalized AGC target was set to 300% with 50 ms maximum injection time. Monoisotopic peak detection was set to Peptide. A setting to relax restrictions if too few precursors are found was set to true. Minimum precursor ion intensity was set to 5.0e3 and precursor charge states up to +8 were included in the experiment.

Cycle time data-dependent mode was set to 2.5 s between master scans. The dynamic mass exclusion was set to 20 s after the first precursor ion fragmentation. The precursor isolation window was set to 2 Th. Normalized collision energy type with fixed collision energy mode was selected in data-dependent MS/MS scan. The collision energy was set to 30%. Orbitrap resolution was set to 15,000, and the AGC target was set to 100%. Custom maximum injection time setting was set to 60 ms.

**Database search and analysis.** Prior to tracing protein–protein networks in the cross-linked samples, we processed the raw files to identify peptides/proteins traced in the samples using the MaxQuant package (version 1.6.12.0)<sup>26,27</sup>. In addition, similar proteomics analysis was performed for non-cross-linked IFN $\alpha$  treated and untreated Flo-1 samples. The MS/MS data was searched against human UniProt ([www.uniprot.org](http://www.uniprot.org)) database (downloaded on 12.08.2020 consisting of 75,093 entries) using the built-in andromeda search engine<sup>27</sup>. The search was performed without specifying enzyme specificity and variable modifications for deamidation (N, Q) and oxidation (M). The precursor mass tolerance was set to 20 ppm and product ions were set at 0.02 Da. The initial and the maximum mass deviation was set to 10 ppm. Maximum mass of a peptide was set to 4600 Da and sequence similarity was set between 7 and 25 amino acids (aa). Further statistical analysis was performed using the Perseus program (version 1.6.10.45). The protein abundances were computed by normalized spectral protein intensity (LFQ intensity; label-free quantification)<sup>27</sup>, and the intensity values were Log<sub>2</sub> transformed. Hierarchical clustering of the proteins identified against their peptide intensities were mapped using pheatmap (v1.0.12) package in R (v 4.1.2). Pathway enrichment analysis using the Reactome pathway database was carried out for the proteins that were upregulated more than fourfold with IFN $\alpha$  treatment compared to untreated samples.

**Identification of cross-links and protein structure modeling.** Identification of Lysine (K) or Serine (S) specific chemical cross-linking of protein complexes traced using LC-MS/MS were performed using the Spectrum Identification Machine for Cross-Linked Peptides (SIM-XL)<sup>29</sup>. First, possible interactions were searched between the Interferon (IFN)-related DNA damage resistant signature (IRDS) genes using a dataset of IRDS proteins described in Padariya et al.<sup>28</sup>. Screening the whole human UniProt for all the conditions and replicates is computationally very demanding, therefore, the entire human UniProt ([www.uniprot.org](http://www.uniprot.org)) database (downloaded on 12.08.2020 consisting of 75,093 entries) was screened for one of the IFN $\alpha$  treated replicate to obtain high-confidence interactions. These high-confidence interactions obtained were taken forward and screened in all the replicates and conditions.

In SIM-XL, the cross-linker (XL) was set to DSS, the XL mass shift and modification mass shift was set to 138.06 and 156.07, respectively. The following cross-link reaction sites were considered: KK, KS, and KN-TERM, with no reporter ions. Both precursor and fragment ppmms were set to 20, and the Xrea threshold to 0.15. Trypsin enzyme was considered fully specific, and the Higher-energy C-trap dissociation (HCD) fragmentation method was implemented. The dynamic DB reduction XCorr threshold and the dynamic DB reduction minimum number of peptides were set to 2.5 and 2, respectively. Other parameters were the following: single isotopic possibilities and peaks matched cutoff, 4 minimum AA residues per chain and intra-link maximum charge, and 3 maximum missed cleavages. Obtained cross-linked 2D maps were analyzed in (SIM-XL), as well as xQuest<sup>28</sup> graphical representations were considered to plot 2D maps. The protein–protein cross-links over the protein structure were represented in PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Using the homology modeling principles and implementing the “Hidden Markov method” the protein model structures were generated using the Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2>)<sup>11</sup>. Phyre2 generates model structures considering the sequence alignments with the known protein structures. For the proteins H2BFS, HLA-A, HMGA1, LRCH4, and MDN1 the template structures were 1kx5<sup>52</sup>, 1kj3<sup>49</sup>, 2eze<sup>55</sup>, 6hlu<sup>62</sup>, and 6i26<sup>65</sup>. In

addition, AlphaFold<sup>71</sup> structures for MX1, UBP18, and ROBO1 were considered. The BIOVIA Discovery Studio Visualizer (Dassault Systèmes, BIOVIA, San Diego, CA, USA) as well as the Molecular Operating Environment (MOE; Chemical Computing Group Inc., Montreal, QC, Canada) packages were used for protein structure visualization.

**Molecular dynamics simulations of the protein–protein network.** Protein–protein docking screen was performed using the pipelines implemented in the Molecular Operating Environment (MOE; Chemical Computing Group Inc., Montreal, QC, Canada) package applying the CHARMM27 forcefield<sup>30</sup>. Individual screens between H2BFS-HLA-A and HMGA1-HLA-A were ranked based on the Generalized Born/Volume Integral (GBVI/WSA dG; kcal/mol) score<sup>72</sup>. Docking conformations corresponding to the CLMS binding sites were further investigated using the molecular dynamics simulation approach (MDS). Applying the CHARMM27 forcefield in the GROMACS 4.6.5<sup>31</sup> package, the following 6 different systems were investigated: H2BFS (apo-form), HLA-A (apo-form), HMGA1 (apo-form), H2BFS-HLA-A, HMGA1-HLA-A, and H2BFS-HLA-A-HMGA1. MD simulation was performed using simple point charge (SPC) water molecules<sup>31</sup> and Na<sup>+</sup> and Cl<sup>−</sup> counter ions were added in the periodic boundary conditions imposed in the simulation box. Initially, each system was minimized for 50,000 steps of the steepest descent algorithm (temperature 300 K; V-rescale thermostat<sup>31</sup> and pressure 1 bar; Parrinello-Rahman barostat<sup>31</sup>). Further, the production run was performed for 100 ns (leapfrog integrator<sup>31</sup>). MDS data were visualized in VMD (Visual Molecular Dynamics) tool<sup>73</sup> and BIOVIA Discovery Studio (Dassault Systèmes, BIOVIA Corp., San Diego, CA, USA).

**SDS-PAGE and Coomassie staining.** The Flo-1 cells were washed twice with cold PBS and harvested by scraping. Cell pellets were lysed with urea lysis buffer (7 M urea, 25 mM HEPES pH 7.5, 25 mM NaCl, 5 mM DTT and 0.05% (V/V) Triton X-100, 1× protease inhibitor cocktail). Protein concentration of lysed samples was quantified using Protein Assay Dye Reagent (Bio-Rad), according to the manufacturer's instructions. Proteins were resolved on 8% polyacrylamide gels and stained with Coomassie stain solution (50% water, 40% methanol, 10% acetic acid, 0.1% Coomassie brilliant blue; CBB) for 30 min. Later, the gel was destain overnight in the stain solution excluding CBB.

**Immunoblotting and co-immunoprecipitation.** For immunoblotting, the samples were processed for SDS-PAGE and protein samples were resolved on either 8% polyacrylamide gels or on Biorad 4–15% pre-cast gel (for Co-IP samples) and transferred onto 0.2 μm nitrocellulose membranes (Amersham Protran, GE Healthcare). Immunoblots were blocked with 5% skimmed milk powder in PBST (PBS + 0.1% Tween-20) and then probed with primary antibodies against MDN1, H2BFS, HLA-A (rabbit: Thermo Fisher, PA5-29911 and mouse: Origene, TA813378), and p53 (DO-1, produced in-house) at 1:500 dilution. The secondary antibody used was horseradish peroxidase (HRP) conjugated rabbit anti-mouse (Dako, 1:1000) or HRP-conjugated goat anti-rabbit/mouse IgG (Jackson ImmunoResearch, 1:3000) for Co-IP. Immunoblots were processed by chemiluminescent substrate (Clarity Max™ Western ECL Substrate, BIO-RAD).

Either Flo-1 or A549 cells were cultured on 10 cm plates and next day the cells were treated with 10 ng/ml IFNα14 for 24 h. Cells were scraped on ice, lysed with the 400 μl of ice-cold lysis buffer (150 mM NaCl, 25 mM Tris-HCl, 0.5% Triton x-100; pH 7.5) with protease inhibitors and centrifuged at 12,000×g for 15 min. 140 μg of lysate was used for immunoprecipitation. Lysates were pre-cleared with 5 μl protein G magnetic beads (Thermo Fisher) for 30 min at 4 °C. Further, the pre-cleared lysates were incubated overnight with 3 μg of antibodies against MDN1 (Thermo Fisher, cat. No. A304-739A), histone H2BFS (Thermo Fisher, cat. No. MA5-31410) or unspecific IgG at 4 °C followed by 30 min binding to the Protein G beads (20 μl/sample). Beads were then washed three times in PBS and proteins were eluted in a 2× SDS-loading buffer at 50 °C for 10 min followed by immunoblotting.

## Data availability

All data generated or analysed during this study are included in this published article and its Supplementary Information files.

Received: 29 March 2022; Accepted: 27 September 2022

Published online: 12 November 2022

## References

- Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386 (2005).
- Chen, J. *et al.* Functional comparison of interferon-α subtypes reveals potent hepatitis B virus suppression by a concerted action of interferon-α and interferon-γ signaling. *Hepatology* **73**, 486–502 (2021).
- Harper, M. S. *et al.* Interferon-α subtypes in an ex vivo model of acute HIV-1 infection: Expression, potency and effector mechanisms. *PLoS Pathog.* **11**, e1005254 (2015).
- Lavender, K. J. *et al.* Interferon alpha subtype-specific suppression of HIV-1 infection in vivo. *J. Virol.* **90**, 6001–6013 (2016).
- de Weerd, N. A., Samarajiva, S. A. & Hertzog, P. J. Type I interferon receptors: Biochemistry and biological functions. *J. Biol. Chem.* **282**, 20053–20057 (2007).
- Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-stimulated genes: A complex web of host defenses. *Annu. Rev. Immunol.* **32**, 513–545 (2014).
- Borden, E. C. *et al.* Interferons at age 50: Past, current and future impact on biomedicine. *Nat. Rev. Drug Discov.* **6**, 975–990 (2007).
- Cheon, H., Borden, E. C. & Stark, G. R. Interferons and their stimulated genes in the tumor microenvironment. *Semin. Oncol.* **41**, 156–173 (2014).
- Hubel, P. *et al.* A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nat. Immunol.* **20**, 493–502 (2019).

10. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
11. Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and type II interferon-induced antiviral factors. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4239–4244 (2012).
12. Poddar, S., Hyde, J. L., Gorman, M. J., Farzan, M. & Diamond, M. S. The interferon-stimulated gene IFITM3 restricts infection and pathogenesis of arthritogenic and encephalitic alphaviruses. *J. Virol.* **90**, 8780–8794 (2016).
13. John, S. P. *et al.* IFIT1 exerts opposing regulatory effects on the inflammatory and interferon gene programs in LPS-activated human macrophages. *Cell Rep.* **25**, 95–106.e6 (2018).
14. Wu, X. *et al.* Site-specific photo-crosslinking proteomics reveal regulation of IFITM3 trafficking and turnover by VCP/p97 ATPase. *Cell Chem. Biol.* **27**, 571–585 (2020).
15. Gómez-Herranz, M. *et al.* The effects of IFITM1 and IFITM3 gene deletion on IFN $\gamma$  stimulated protein synthesis. *Cell. Signal.* **60**, 39–56 (2019).
16. Amini-Bavil-Olyae, S. *et al.* The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. *Cell Host Microbe* **13**, 452–464 (2013).
17. Neeffjes, J., Jongmsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
18. Jongmsma, M. L. M., Neeffjes, J. & Spaapen, R. M. Playing hide and seek: Tumor cells in control of MHC class I antigen presentation. *Mol. Immunol.* **136**, 36–44 (2021).
19. Cornel, A. M., Mimpfen, I. L. & Nierkens, S. MHC class I downregulation in cancer: Underlying mechanisms and potential targets for cancer immunotherapy. *Cancers (Basel)* **12**, 1760 (2020).
20. Au-Yeung, N. & Horvath, C. M. Transcriptional and chromatin regulation in interferon and innate antiviral gene expression. *Cytokine Growth Factor Rev.* **44**, 11–17 (2018).
21. Ni, Z. *et al.* Apical role for BRG1 in cytokine-induced promoter assembly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14611–14616 (2005).
22. Tratnjek, L. *et al.* Standardization of esophageal adenocarcinoma in vitro model and its applicability for model drug testing. *Sci. Rep.* **11**, 6664 (2021).
23. Contino, G. *et al.* Whole-genome sequencing of nine esophageal adenocarcinoma cell lines. *F1000Research* **5**, 1336 (2016).
24. Arimoto, K.-I. *et al.* STAT2 is an essential adaptor in USP18-mediated suppression of type I interferon signaling. *Nat. Struct. Mol. Biol.* **24**, 279–289 (2017).
25. Haller, O. & Kochs, G. Huzan MxA protein: An interferon-induced dynamin-like GTPase with broad antiviral activity. *J. Interferon Cytokine Res.* **31**, 79–87 (2011).
26. Cox, J. *et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* **4**, 698–705 (2009).
27. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
28. Padariya, M. *et al.* Functional interfaces, biological pathways, and regulations of interferon-related DNA damage resistance signature (IRDS) genes. *Biomolecules* **11**, 622 (2021).
29. Lima, D. B. *et al.* Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins. *Nat. Protoc.* **13**, 431–458 (2018).
30. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
31. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
32. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
33. Chen, X. *et al.* Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**, 827–839 (1998).
34. Donovan, J., Whitney, G., Rath, S. & Korennykh, A. Structural mechanism of sensing long dsRNA via a noncatalytic domain in human oligoadenylate synthetase 3. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3949–3954 (2015).
35. Grönholm, J. *et al.* Structure-function analysis indicates that sumoylation modulates DNA-binding activity of STAT1. *BMC Biochem.* **13**, 20 (2012).
36. Honke, N., Shaabani, N., Zhang, D.-E., Hardt, C. & Lang, K. S. Multiple functions of USP18. *Cell Death Dis.* **7**, e2444 (2016).
37. Basters, A. *et al.* Structural basis of the specificity of USP18 toward ISG15. *Nat. Struct. Mol. Biol.* **24**, 270–278 (2017).
38. Weichselbaum, R. R. *et al.* An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18490–18495 (2008).
39. Aleksandrova, N. *et al.* Robo1 forms a compact dimer-of-dimers assembly. *Structure* **26**, 320–328.e4 (2018).
40. Ibsen, M. S. *et al.* The 2'-5'-oligoadenylate synthetase 3 enzyme potentially synthesizes the 2'-5'-oligoadenylates required for RNase L activation. *J. Virol.* **88**, 14222–14231 (2014).
41. Li, Y. *et al.* Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2241–2246 (2016).
42. Lee, W.-B. *et al.* OAS1 and OAS3 negatively regulate the expression of chemokines and interferon-responsive genes in human macrophages. *BMB Rep.* **52**, 133–138 (2019).
43. Sadler, A. J. & Williams, B. R. G. Interferon-inducible antiviral effectors. *Nat. Rev. Immunol.* **8**, 559–568 (2008).
44. Brantis-de-Carvalho, C. E. *et al.* MxA interacts with and is modified by the SUMOylation machinery. *Exp. Cell Res.* **330**, 151–163 (2015).
45. Liu, B. *et al.* Inhibition of Stat1-mediated gene activation by PIAS1. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10626–10631 (1998).
46. Oshiumi, H. *et al.* DDX60 is involved in RIG-I-dependent and independent antiviral responses, and its function is attenuated by virus-induced EGFR activation. *Cell Rep.* **11**, 1193–1207 (2015).
47. Miyashita, M., Oshiumi, H., Matsumoto, M. & Seya, T. DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol. Cell. Biol.* **31**, 3802–3819 (2011).
48. Liu, X.-Y., Chen, W., Wei, B., Shan, Y.-F. & Wang, C. IFN-induced TPR protein IFIT3 potentiates antiviral signaling by bridging MAVS and TBK1. *J. Immunol.* **187**, 2559–2568 (2011).
49. Reiser, J. B. *et al.* A T cell receptor CDR3 $\beta$  loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity* **16**, 345–354 (2002).
50. Kurimoto, E. *et al.* Structural and functional mosaic nature of MHC class I molecules in their peptide-free form. *Mol. Immunol.* **55**, 393–399 (2013).
51. Serçinoğlu, O. & Ozbek, P. Sequence-structure-function relationships in class I MHC: A local frustration perspective. *PLoS ONE* **15**, e0232849 (2020).
52. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
53. Gnesutta, N., Nardini, M. & Mantovani, R. The H2A/H2B-like histone-fold domain proteins at the crossroad between chromatin and different DNA metabolisms. *Transcription* **4**, 114–119 (2013).
54. Kobiyama, K. *et al.* Role of extrachromosomal histone H2B on recognition of DNA viruses and cell damage. *Front. Genet.* **4**, 91 (2013).

55. Huth, J. R. *et al.* The solution structure of an HMGI(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657–665 (1997).
56. Sgarra, R. *et al.* Interaction proteomics of the HMGA chromatin architectural factors. *Proteomics* **8**, 4721–4732 (2008).
57. Ozturk, N., Singh, I., Mehta, A., Braun, T. & Barreto, G. HMGA proteins as modulators of chromatin structure during transcriptional activation. *Front. Cell Dev. Biol.* **2**, 5 (2014).
58. Xu, M. *et al.* Core promoter-selective function of HMGA1 and mediator in Initiator-dependent transcription. *Genes Dev.* **25**, 2513–2524 (2011).
59. Yang, D. *et al.* High mobility group box-1 protein induces the migration and activation of human dendritic cells and acts as an alarmin. *J. Leukoc. Biol.* **81**, 59–66 (2007).
60. Aloor, J. J. *et al.* Leucine-rich repeats and calponin homology containing 4 (Lrch4) regulates the innate immune response. *J. Biol. Chem.* **294**, 1997–2008 (2019).
61. Cornish, J., Chamberlain, S. G., Owen, D. & Mott, H. R. Intrinsically disordered proteins and membranes: A marriage of convenience for cell signalling? *Biochem. Soc. Trans.* **48**, 2669–2689 (2020).
62. Deyaert, E. *et al.* Structure and nucleotide-induced conformational dynamics of the *Chlorobium tepidum* Roco protein. *Biochem. J.* **476**, 51–66 (2019).
63. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
64. Chen, Z. *et al.* Structural insights into Mdn1, an essential AAA protein required for ribosome biogenesis. *Cell* **175**, 822–834 (2018).
65. Sosnowski, P. *et al.* The CryoEM structure of the *Saccharomyces cerevisiae* ribosome maturation factor Rea1. *Elife* **7**, 39163 (2018).
66. Garbarino, J. E. & Gibbons, I. R. Expression and genomic analysis of midasin, a novel and highly conserved AAA protein distantly related to dynein. *BMC Genomics* **3**, 18 (2002).
67. Gilmore, J. M. *et al.* WDR76 co-localizes with heterochromatin related proteins and rapidly responds to DNA damage. *PLoS ONE* **11**, e0155492 (2016).
68. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022–3040.e28 (2021).
69. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
70. Bouchal, P. *et al.* Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J. Proteome Res.* **8**, 362–373 (2009).
71. Varadi, M. *et al.* AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
72. Padariya, M. *et al.* Structural determinants of peptide-dependent TAP1-TAP2 transit passage targeted by viral proteins and altered by cancer-associated mutations. *Comput. Struct. Biotechnol. J.* **19**, 5072–5091 (2021).
73. Humphrey, W., Dalke, A. & Schulten, K. V. M. D. Visual molecular dynamics. *J. Mol. Graph.* **14**(33–8), 27–28 (1996).

## Acknowledgements

The APC was funded by Grant 2020/36/C/NZ2/00108 (The National Science Centre, Krakow, Poland/Narodowe Centrum Nauki, Krakow, Poland). U.K. is supported by Grant 2020/36/C/NZ2/00108 from The National Science Centre, Poland (Narodowe Centrum Nauki, Krakow, Poland). B.V. was supported by the Czech Science Foundation (GACR 22-02940S), the European Regional Development Fund—Project ENOCH (No. CZ.02.1.01/0.0/0.0/16\_019/0000868) and by the Ministry of Health Development of Research Organization, MH CZ—DRO (MMCI, 00209805). The International Centre for Cancer Vaccine Science (Fundacja na rzecz Nauki Polskiej: MAB/3/2017) project is carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. Authors would also like to thank the PL-Grid Infrastructure, Poland for providing their hardware and software resources.

## Author contributions

A.S., M.P., J.F., T.H., A.S., B.V., and U.K. designed experiments and analyzed data. U.K., A.S., M.P., and J.F., prepared figures and wrote the original draft of the manuscript. U.K. and B.V. have contributed to the design of the project. U.K. supervised the project and performed data analysis. A.S., M.P., J.F., S.K., S.M., A.D., T.L., J.B., M.W., K.B., T.H., A.S., B.V., and U.K. contributed to the text of the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21393-z>.

**Correspondence** and requests for materials should be addressed to B.V. or U.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022